Structural basis for the synthesis of the core 1 structure by C1GalT1

Andrés Manuel González-Ramírez 1, Ana Sofia Grosso 2,3,7, Zhang Yang 4,7, Ismael Compañón 5,7, Helena Coelho 2,3,7, Yoshiki Narimatsu 4, Henrik Clausen 4, Filipa Marcelo 2,3, Francisco Corzana 5 & Ramon Hurtado-Guerrero 1,4,6 ✉

C1GalT1 is an essential inverting glycosyltransferase responsible for synthesizing the core 1 structure, a common precursor for mucin-type O-glycans found in many glycoproteins. To date, the structure of C1GalT1 and the details of substrate recognition and catalysis remain unknown. Through biophysical and cellular studies, including X-ray crystallography of C1GalT1 complexed to a glycopeptide, we report that C1GalT1 is an obligate GT-A fold dimer that follows a SN2 mechanism. The binding of the glycopeptides to the enzyme is mainly driven by the GalNAc moiety while the peptide sequence provides optimal kinetic and binding parameters. Interestingly, to achieve glycosylation, C1GalT1 recognizes a high-energy conformation of the α-GalNAc-Thr linkage, negligibly populated in solution. By imposing this 3D-arrangement on that fragment, characteristic of α-GalNAc-Ser peptides, C1GalT1 ensures broad glycosylation of both acceptor substrates. These findings illustrate a structural and mechanistic blueprint to explain glycosylation of multiple acceptor substrates, extending the repertoire of mechanisms adopted by glycosyltransferases.

https://doi.org/10.1038/s41467-022-29833-0

1 Institute of Biocomputation and Physics of Complex Systems, University of Zaragoza, Mariano Esquillor s/n, Campus Rio Ebro, Edificio 1-D, 50018 Zaragoza, Spain. 2 Associate Laboratory i4HB - Institute for Health and Bioeconomy, NOVA School of Science and Technology, 2829-516 Caparica, Portugal. 3 UCIBIO – Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, 2829-516 Caparica, Portugal. 4 Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. 5 Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, E-26006 Logroño, Spain. 6 Fundación ARAID, 50018 Zaragoza, Spain. 7 These authors contributed equally: Ana Sofia Grosso, Zhang Yang, Ismael Compañón, Helena Coelho. ✉ Email: francisco.corzana@unirioja.es; rhurtado@bifi.es
Normal development and angiogenesis involve preferences for the peptide sequence around the GalNAc moiety and the peptide. In addition, we unveil that C1GalT1 recognizes the staggered conformation for the α-GalNAc-Thr linkage, a high-energy conformation that is negligibly populated in solution. With this 3D-arrangement, characteristic of α-GalNAc-Ser peptides, C1GalT1 ensures broad glycosylation of both acceptor substrates.

**Results**

**Kinetics of DmC1GalT1 against glycopeptide substrates.** To perform biophysical experiments using DmC1GalT1, we designed a construct that did not contain the predicted signal sequence and the transmembrane domain, and the enzyme was secreted from HEK293 cells (residues T43-Q388; see Supplementary Fig. 1 and methods). To assess the activity of DmC1GalT1, we designed a series of glycopeptides (designated P1–P7) based on a previous study (Fig. 1). These glycopeptides contained a Gly at +1 and either a Phe or Tyr at +3, residues that clearly improved the activity, and Tyr/Phe/Pro at −3 that enhanced the activity slightly of the human C1GalT1. Glu at −1 was present in P6 to compare it with the most similar glycopeptides, P1 (Ala at −1) and P7 (Asp at −1). The other positions were occupied by a Pro at +2 and Ala at −2 because both were previously well tolerated. We also included the naked APDTRP, the APDT*RP and the APDS*RP glycopeptides for further evaluation (where * represents a GalNAc moiety bound to the underlying amino acid, either Thr or Ser). The APDTRP is an immunogenic epitope found in the tandem repeat sequence present in MUC1 and the APDT*RP, whose structure in the bound state with an antitumor antibody was recently reported, is the basis for development of several cancer vaccines, and it is a natural substrate for C1GalT1 in the context of MUC1. The use of the APDS*RP was to confirm whether the activity of C1GalT1 was better with a glycopeptide containing α-GalNAc-Thr over α-GalNAc-Ser, as previously reported.

To initiate kinetic studies, we set up first the experimental conditions using DmC1GalT1 toward UDP-Gal and APDT*RP (Fig. 1a, Supplementary Fig. 2a and Supplementary Table 1). DmC1GalT1 showed a hyperbolic profile in the presence of variable concentrations of either UDP-Gal or APDT*RP, which was also observed in the presence of the other glycopeptides (Fig. 1a and Supplementary Fig. 2b). The apparent Michaelis constants (K_m) for UDP-Gal and APDT*RP were 88 ± 8 and 195 ± 43 μM, respectively, and the V_max was ~3.5 min⁻¹ (Fig. 1b, left and middle panels, and Supplementary Table 1), a value consistent with other previously reported low values for follow-up GTs such as POMGnT1/POMGnT2 (k_cat values ranged from 7 to 1920 min⁻¹ depending on the glycopeptide sequences). As expected, DmC1GalT1 was inactive on the naked APDTRP and its activity was slightly reduced in the presence of APDS*RP. Particularly, the K_m values, k_cat values, and catalytic efficiency using APDS*RP were 1.26-, 1.85-, 2.25-fold worse than those of C1GalT1 in the presence of APDT*RP (Fig. 1b), finding that...
matches the results found with the rat enzyme, and that suggests that the methyl group of Thr is likely important for obtaining slightly better kinetic parameters. We also determined whether the GalNAc moiety behaved as an acceptor substrate. To explore that, we used the α-O-methyl-GalNAc as a substrate, which turned out to be a worse substrate than the APDT*RP, since kinetic parameters could not be determined, a finding in agreement with a previous report using similar analogues versus glycopeptides. At a saturating concentration of the glycopeptides, agreement with a previous report using similar analogues versus glycopeptides and α-O-methyl-GalNAc, versus the APDT*RP, and the enzyme did not reach saturation up to 2 mM α-O-methyl-GalNAc (Fig. 1a and Supplementary Fig. 3). These data show that the context of the peptide around the sugar moiety is key to having optimal kinetic parameters and that the GalNAc moiety is not sufficient to achieve that.

Regarding the kinetics with P1–P7 glycopeptides (Fig. 1a), we firstly determined the kinetic parameters for UDP-Gal under a saturated concentration of P4, rendering a slightly better $K_m^{app}$ and almost an identical $k_{cat}^{app}$ value for UDP-GalP4 compared to those parameters for UDP-Gal under the presence of APDT*RP (Supplementary Fig. 2 and Supplementary Table 1). The $K_m^{app}$ for P1–P7 glycopeptides were slightly better than that of the natural APDT*RP glycopeptide, ranging from 1.3- to 4-fold improvements, with P4 having the better $K_m^{app}$ (Fig. 1b, left panel). The data also suggest that the Pro at −3 is slightly better for binding than the Tyr at −2 (P4 versus P4), and that Glu or Asp at −2 are slightly worse for binding than the Ala at −2 (P6/P7 versus P1). With the $k_{cat}^{app}$ parameters, the range of values is more restricted, with P1 being the slowest substrate and P6 being the fastest (Fig. 1b, middle panel). Finally, the range of catalytic efficiency values were slightly less restricted than that of $k_{cat}^{app}$, and suggested that for the series of glycopeptides containing an acceptor glycosylated Thr, the best substrate was P4 and the worst ones were P1 and APDT*RP (Fig. 1b, right panel and Supplementary Table 1). Overall, our data suggest that the differences in the kinetic parameters between the glycopeptides are small and that not only the GalNAc moiety is important for glycosylation, but also that the peptide sequence is crucial for achieving optimal kinetic parameters, suggesting that C1GalT1 may interact directly with the peptide of the acceptor substrates. Note that saturation is not achieved in the presence of α-O-methyl-GalNAc and that only this is achieved in the presence of the different peptides within the glycopeptide substrates.

STD NMR reveals that DmC1GalT1 directly engages with the GalNAc moiety and the peptide sequence. We then performed STD NMR experiments to shed light onto the DmC1GalT1-glycopeptide interaction mode. The STD NMR experiment is a ligand-observed technique (only the 1H-NMR assignment of the ligand is required for analysis) that relies on saturation transfer, through nuclear Overhauser effect, from receptor (e.g., protein/enzyme) proton resonances to protons of a ligand (e.g., carbohydrate, glycopeptide) exchanging between a protein-bound and free state. Analysis of the STD responses allows to infer which atoms of the binding ligand are in closer contact with the receptor, and to determine the so-called STD-derived epitope mapping. We selected the α-O-methyl-GalNAc, the naked APDTRP, one of the worst substrates (APDT*RP), and one of the best substrates (P4). In the case of α-O-methyl-GalNAc, the naked and the APDTRP, two on-resonance frequencies were used, at aliphatic (~0.5 ppm) and aromatic (7 ppm) region. However, for P4 (due to the presence of the aromatic Tyr), only on-resonance frequency at ~0.5 ppm was used. First, in the presence of a ~6-fold and 7-fold excess of UDP and MnCl₂ with respect to the enzyme, we found that while the naked peptide itself did not display STD response, the α-O-methyl-GalNAc clearly presented STD enhancements, indicating the importance of GalNAc for the enzyme recognition (Supplementary Figs. 4 and 5a, and Supplementary Table 2). Next, we performed STD NMR of the glycopeptides APDT*RP and P4 in the absence and presence of UDP and MnCl₂. We conducted these experiments in the absence of the nucleotide because previously we found that other distant GT-A fold GTs such as GalNAc-T2, and the NleB/ SseK effector proteins were dependent on the presence of UDP for binding to their protein/peptide acceptor substrates, implying...
the existence of an induced-fit mechanism\textsuperscript{20,35,36}. Interestingly, in the case of GalNAc-T2, the active conformation of GalNAc-T2, characterized by the shifting of a flexible loop from an open to a closed conformation, was completely achieved in the presence of UDP-GalNAc and less in the presence of UDP\textsuperscript{20}. We also found that the GT-B fold FUT8 showed similar properties to the other two GTs, though FUT8 bound better to an N-glycan in the presence of GDP, and the nucleotide was not essential for binding to the N-glycan\textsuperscript{37,38}. In addition, NleB/SseK and FUT8 also contained flexible loops that were ordered in the presence of the nucleotide as we found for GalNAc-T2, implying that the active site adopted an active conformation once that these flexible loops bound to the sugar nucleotide\textsuperscript{20,35,36,38}. Overall, we proposed for these enzymes that the binding of the sugar nucleotide was required for binding to the acceptor substrate (optimal binding for FUT8 in the presence of GDP) and in turn for glycosylation. Herein, in the case of DmC1GalT1 and in the absence of UDP, only P4 clearly showed STD signals. However, both glycopeptides showed a clear STD response in the presence of UDP. The results suggest differences in the recognition of the two glycopeptides. In both cases, the GalNAc moiety displayed high saturation transfer indicating that it should be in closer contact with the enzyme.

GalNAc STD-derived epitope for these glycopeptides and for the \(\alpha-O\)-methyl-GalNAc were comparable, implying a similar binding orientation for GalNAc unit in all structures (Fig. 2a, Supplementary Fig. 5a, b and Supplementary Tables 2–4). However, the STD amplification factor was lower in the case of \(\alpha-O\)-methyl-GalNAc than that of the APDT\textsuperscript{*}RP or P4 (Supplementary Tables 2–4), which likely reflects the expected lower binding affinity of \(\alpha-O\)-methyl-GalNAc versus those of the glycopeptides, inferred from the kinetics experiments. Indeed, at the level of the peptide sequence, the Thr methyl group displayed a clear STD response in all cases, while the results varied for the rest of amino acids of both glycopeptides. For the APDT\textsuperscript{*}RP, modest STD enhancements were detected for the methyl of Ala1, and few protons of overlapped Pro2 and Pro6. No STD response was observed for Asp3 protons. Remarkably in the case of P4, significant STD response was found for Pro1 and Tyr7 side chains protons, either in absence or presence of UDP (Fig. 2a), suggesting that these amino acids should be in close contact with...
DMcCGaT1. These additional interactions might explain the differences in $K_a^{\text{UDP}}$ between both glycopeptides (fourfold better $K_a^{\text{UDP}}$ of P4 than that of APDTRP). Overall, the data suggest that the binding of the GaINAc moiety is the driving force for recognition, and optimal binding is reached in the presence of both the GaINAc moiety and the peptide.

**DMcCGaT1 does not show an allosteric behavior with glycopeptides.** To corroborate the different behavior between the glycopeptides in the absence and presence of UDP by STD NMR, we performed isothermal titration calorimetry (ITC) experiments. First, we determined the $K_d$ of UDP for binding to DMcCGaT1 in the presence of MnCl$_2$ ($K_d$ = 18.39 ± 4.67 μM) (Supplementary Fig. 6, and Supplementary Table 5). As expected, no binding was shown for the naked APDTRP under an excess of UDP (Supplementary Fig. 6). Then, we evaluated whether this enzyme requires UDP binding prior to binding the glycopeptides. While DMcCGaT1 only showed binding to the APDTP$^*$RP in the presence of UDP, DMcCGaT1 bound well to P4 in the presence or absence of UDP, in agreement with the results from the STD NMR experiments (Fig. 2b, c). The $K_a$s for the glycopeptides matched their $K_a^{\text{UDP}}$ and the differences found between the $K_a^{\text{UDP}}$ (-3.5-fold better $K_a$ of P4 than that of APDTRP$^*$RP). Since the APDTP$^*$RP is an unusual glycopeptide containing two charged residues (an Asp and Arg residue), we wondered whether this could be the reason for its behavior in the absence of UDP. To rule out this, we also performed ITC experiments with P7, which contains a negatively charged residue. P7 behaved similarly to P4 and bound indistinctly to the enzyme in the presence or absence of UDP (Fig. 2b, c, Supplementary Fig. 6 and Supplementary Table 5), suggesting that the Arg residue of APDTRP$^*$RP or its conformation might be behind its behavior (see further experiments below). Regarding the analysis of the thermodynamic parameters of the interaction, these were somewhat complex and difficult to interpret for the glycopeptides, impeding obtaining a meaningful conclusion (Supplementary Table 5).

Our results also imply that DMcCGaT1 does not likely follow an induced-fit mechanism as found for other GTs such as NleB1, GalNAc-T2, and FUT8, and that therefore, DMcCGaT1 does not need prior binding to the sugar nucleotides to bind its acceptor substrates.

**Architecture of the DMcCGaT1-UDP-APDTRP$^*$RP complex.** To provide atomic insights into the structure of DMcCGaT1 and its interaction with UDP-Gal-UDP and glycopeptides, we worked with a truncated version of DMcCGaT1 (residues S73-Q388) that was secreted from High Five (Hi5) cells (see Methods). The kinetic parameters of this construct were highly similar to those found for the longer construct DMcCGaT1T43-Q388 (see Supplementary Fig. 7 and Supplementary Table 1), verifying that the further truncation of N-terminal residues did not affect the kinetic properties. Crystals of the DMcCGaT1 in the presence of UDP-MnCl$_2$ and APDTP$^*$RP were obtained in the space group P2$_1$. Other attempts with other glycopeptides failed to obtain crystals. The crystal structure was obtained at 2.40 Å by molecular replacement and using the DMcCGaT1 model obtained from alpha fold 2 server (Methods, Fig. 3a, upper panel, and Supplementary Table 6). The asymmetric unit (AU) of P2$_1$ crystals contained two molecules of DMcCGaT1 that were arranged as a homodimer with each monomer adopting the typical GT-A fold (Fig. 3a). The dimeric form was confirmed by gel filtration chromatography (Supplementary Fig. 8b) and was also reported for the human C1GalT1 orthologue. The PISA server further confirmed this dimeric structure and revealed that the dimer presented a large buried surface area (7621 Å$^2$), implying a very stable and tight interface. The residues at the interface engaged in the stabilization of the dimer were located at the N-terminus loop, a$_1$, a$_2$, a$_4$, loop a$_6$, a$_7$, loop a$_7$-a$_8$, loop a$_8$-$\beta_7$, $\beta_7$, loop $\beta_7$-a$_9$, and the long and unstructured C-terminus loop (Fig. 3a, lower panel, and Supplementary Fig. 1). One of these residues, Tyr321 (336 in DMcCGaT1), highly conserved and located in a$_9$ at the interface, was found mutated to Asn, leading to thrombocytopenia and kidney disease in mice (Supplementary Fig. 1). Interestingly, Cosmc was shown to bind to residues 83–97 of the human C1GalT1 (HsC1GalT1)43, located in a$_1$, $\beta_1$ and loop $\beta_1$-a$_2$ of the DMcCGaT1 structure. One of these residues, Leu95DMcCGaT1 (Leu82HsC1GalT1) is in a$_1$ at the dimer interface (Supplementary Fig. 1), suggesting that it is likely that Cosmc is important to form the obligate dimer of C1GalT1. However, this peptide region is partly conserved within C1GalT1 found in vertebrates and invertebrates, implying that this particular peptide in DMcCGaT1 is likely not recognized by Cosmc. Nevertheless, both examples illustrate the importance of interface residues in stability and function of C1GalT1. The root-mean-square deviation (RMSD) between both molecules belonging to chain A and B in the AU is 0.24 Å on 278 equivalent Ca atoms. Hereafter we will discuss only molecule A because it contains a better-defined density for the ligands. In addition, DMcCGaT1 also contained the four conserved landmark features among GT-A GTs42: the DxD motif for metal cation interactions (Asp181-X-$\alpha$-Asp183), a “glycine-rich” loop facing the acceptor and donor sugar site located in DMcCGaT1 at loop $\beta$5-$\beta$6, an “xD” motif at the beginning of $\alpha$6 in DMcCGaT1 harboring the catalytic base (Asp255, see further experiments below), and a “C-His” residue that coordinates with the metal ion (His324) (Fig. 3b and Supplementary Fig. 1).

A close inspection of the active site of DMcCGaT1 and its comparison with other orthologs such as the human, mouse, and chicken C1GalT1 revealed that both the UDP-Gal and glycopeptide binding sites were identical (Fig. 3c, upper panel, and Supplementary Fig. 1), exemplifying that the DMcCGaT1 is an excellent model to understand the biochemical aspects of the human enzyme. An analysis of the electrostatic surface potential showed a negatively charged UDP-Gal binding site required to coordinate the Mn$^{2+}$, and moderate positively charged patches and neutral patches for binding to the peptide. In addition, the GalNAc binding site was moderately negatively and positively charged facing the central core and the acetamide group/OH$_6$, respectively (Fig. 3c, lower panel).

Regarding the structural homology of DMcCGaT1 to other described structures, the DALI server43 revealed structural homology to two galactosyltransferases, namely the dimeric human B3GNT2 (e.g., PDB entries 7HN4 and 6WMO) and the monomeric mouse Manic fringe (Mfnf; PDB entries 2JOA and 2JOB), both belonging to the CAZY31 family (Fig. 3d). Although DMcCGaT1 is very distant to B3GNT2 and Mfnf in terms of acceptor substrates, the server rendered good scores implying that they superimposed fairly well (RMSDs of ~1.7 and ~3.17 Å between DMcCGaT1 and B3GNT2, and DMcCGaT1 and Mfnf crystal structures, respectively; the superimposed residues ranged from 189 to 151 residues). Interestingly, the strong similarities between DMcCGaT1 and B3GNT2 at the overall fold were matched by the excellent superposition of the UDP and the acceptor substrates (Fig. 3e). It is worth mentioning that the GalNAc OH3 of APDTP$^*$RP and Gal OH3 of LNnT were located at almost identical positions (~0.92 Å atomic shift between the GalNAc and the Gal moieties) and close to the $\beta$-phosphate, in agreement with their role as the acceptor sites. Note that UDP in Mfnf also superimposed very well with UDP of DMcCGaT1 though the former structure was only obtained with UDP-Mn$^{2+}$.
The active site of DmC1GalT1. The DmC1GalT1 binding site is formed by the UDP-Gal and the glycopeptide binding sites (Fig. 4a). The uridine moiety of UDP establishes a CH–π interaction with Leu155 while the uracil moiety is tethered via hydrogen bonds to Glu150 and Lys158 side chains and Gly151 backbone. The ribose moiety of uridine interacts with the Asp182 side chain and Met160 backbone, and the pyrophosphate interacts with Arg152, His324 and Tyr325 side chains. The pyrophosphate group oxygen atoms, Asp181 and Asp183 of the DxD motif and His324 hexagonally coordinate Mn$^{2+}$. Unlike the intimate recognition of UDP by DmC1GalT1, APDT*RP displays fewer contacts with the enzyme (Fig. 4a), in line with our ITC data in which the binding of UDP was ~9.5-fold stronger than the binding of APDT*RP to the enzyme (Supplementary Table 5). The glycopeptide GalNAc moiety is recognized through hydrogen bonds formed between the acetamide carbonyl and Ser220 side chain, OH3 with Asp255/Tyr218 side chains, and OH6 with Tyr304 side chain. At the peptide level, the Pro2 side chain and the methyl group of Thr4 establish CH–π interactions.
with Tyr213, and Trp300/Tyr304, respectively, and the Thr4 backbone makes a hydrogen bond with Trp300 side chain. Arg5 side chain is engaged in a hydrogen bond with Arg152, and Pro6 establishes a CH–π interaction with Tyr325 (Fig. 4a). These interactions also reveal that the GalNAc moiety is more intimately recognized than the peptide and that the GalNAc moiety is only tethered through hydrogen bonds, while the peptide is engaged in hydrophobic and hydrogen bond interactions. Overall, our data align with the STD-derived epitope map, suggesting that the GalNAc moiety is the driving force for

**Fig. 4 Structural features of the active site.** a Stereo view of the active site for the DmC1GalT1-UDP-Mn2+·APDT*RP complex. The residues forming the active site are depicted as orange carbon atoms. UDP and the glycopeptide are shown as gray and green carbon atoms, respectively. The manganese atom is shown as a pink sphere. Hydrogen bond interactions are shown as dotted black lines. b Geometry of the glycosidic linkage of the glycopeptide APDT*RP in solution derived from 0.5 µs MD simulations. The dihedral angles are defined as follow, φ = O5-C1-O1-Cβ, and ψ = C1-O1-Cβ-Cα. The red circle corresponds to the conformation found for this linkage in the crystal structure of the glycopeptide bound to DmC1GalT1 in the presence of UDP. This conformation is also present in glycopeptide APDS*RP in solution (left panel). 3D view of APDS*RP in complex with DmC1GalT1 in the presence of UDP-Gal obtained from 0.5 µs MD simulations (right panel). c Free-energy map (φ, ψ) of the glycosidic dihedral angle calculated for the free peptide APDT*RP in water (see Methods) at 300 K. The contour maps are drawn with a spacing of 1 kcal/mol. Regions that were never visited by the peptide are shown in dark orange. “A” refers to the ‘eclipsed’ conformation typically found for α-GalNAc-Thr derivatives in solutions47,50. “B” refers to the ‘staggered’ conformation found for α-GalNAc-Ser derivatives in solution72. d Close-up view of the binding site region of the DmC1GalT1-UDP-Gal-APDT*RP complex showing the Asp255 as the catalytic base in the plausible S212 single-displacement reaction mechanism. Note the proximity and the orientation of the GalNAc OH3 to the anomeric carbon (3.81 Å) which is compatible with the inversion of the configuration during the reaction. e Histogram showing the relative activities of the mutants compared to the wild-type (WT) protein. All experiments were done in triplicate (n = 3 independent experiments). Error bars represent the standard deviation calculated by the GraphPad Prism fit of the data sets. Source data are provided as a Source Data file. f Proposed S212 single-displacement reaction mechanism for C1GalT1.
A high-energy conformation of the glycosidic linkage of α-GalNAc-Thr is required for the molecular recognition by DmC1GalT1. An intriguing feature inferred from the crystal structure was the presence of an energetically less favorable conformation of the glycosidic linkage displayed by α-GalNAc-Thr (Fig. 4b, left panel). This staggered conformation (with \( \psi \approx 180^\circ \)), typically found in glycosidic linkage between α-GalNAc and a serine residue, is not found in solution for α-GalNAc-Thr either in the free form\(^2\) or bound to proteins, where the eclipsed rotamer (with \( \psi \approx 120^\circ \)) is the usual form.\(^{21,23,28,35,48-51}\) We performed molecular dynamics (MD) simulations on DmC1GalT1 in complex with UDP-Gal and APDS\(^*\)RP. These calculations showed that the staggered conformation was also predicted for α-GalNAc-Ser (Fig. 4b, right panel, and Methods), implying that C1GalT1 requires the staggered conformation for the effective glycosylation of α-GalNAc-Ser and α-GalNAc-Thr. MD simulations performed for the analogous complex with APDT\(^*\)RP, where the eclipsed conformation was fixed in α-GalNAc-Thr (Supplementary Fig. 9 and Methods) showed a loss of interactions between the peptide and the protein compared to those found in the X-ray structure. Specifically, the CH-π interactions between the methyl group of Thr4 and Trp300/Tyr304 and Pro6(C8) of the glycopeptide and Tyr325 were significantly weakened due to the increased distance between the aromatic rings and the peptide. Moreover, the hydrogen bond between the carbonyl group of Thr4 and Trp300 was negligible throughout the MD simulation trajectory with constraints. As for the GalNAc moiety, the hydrogen bonding between the side chain of Tyr304 and GalNAc OH6 was lost. On the other hand, the APDS\(^*\)RP peptide has slightly worse \( K_m^{app} \) than the threonine derivative and does not have the conformational penalty that operates in the Thr-containing peptide. These results suggest that rather subtle free-energetic effects are probably guiding the binding. In this regard, the free-energy penalty associated to bring the glycosidic linkage from a ‘eclipsed’ conformation to a ‘staggered’ one was calculated to be 2.5 kcal/mol (Fig. 4c and Supplementary Fig. 10). In contrast, this conformational shift is favored by 1.9 kcal/mol in the serine derivative. (Supplementary Fig. 10). This finding likely explains why C1GalT1 has similar kinetic parameters for both glycosites, and can glycosylate either α-GalNAc-Ser or α-GalNAc-Thr indistinctly.

The inversion mechanism of C1GalT1. To get further insights into the inversion mechanism of C1GalT1, we superimposed our crystal structure with the structure on the human B3GNT2-UDP-GlcNAc complex (PDB entry 7JHL), and then the coordinates of UDP-GlcNAc were replaced by UDP-Gal. The resulting complex, DmC1GalT1-UDP-Gal-Mn\(^{2+}\)-APDT\(^*\)RP, was minimized using molecular mechanics (MM) calculation as shown in Methods (Fig. 4d). In this structure, the GalNAc OH3 was properly aligned with the protein compared to those found in the X-ray structure. In this regard, the free-energy penalty associated to bring the glycosidic linkage from a ‘eclipsed’ conformation to a ‘staggered’ one was calculated to be 2.5 kcal/mol (Fig. 4c and Supplementary Fig. 10). In contrast, this conformational shift is favored by 1.9 kcal/mol in the serine derivative. (Supplementary Fig. 10). This finding likely explains why C1GalT1 has similar kinetic parameters for both glycosites, and can glycosylate either α-GalNAc-Ser or α-GalNAc-Thr indistinctly.

In vitro and in cells activity of C1GalT1 mutants. To get insights into the role of residues of DmC1GalT1 engaged in interactions with the glycopeptide, we tested Ala mutations of Arg152, Tyr213, Tyr218, Trp300 and Tyr325 to Ala residues and the resulting mutants were characterized at in vitro level under the same conditions used for the D255A. The results showed that Y218A and W300A were inactive while R152A and Y213A/Y325A suffered a 15- and 25-fold decrease in activity with respect to the WT, respectively (Fig. 4e). We then generated the equivalent mutants of DmC1GalT1 in the HsC1GalT1 (see Supplementary Fig. 1). To evaluate the activity of these HsC1GalT1 mutants in cells, we used a HEK293T cell without capacity for producing core 1 (KO C1GALT1) and without capacity to modify the core 1 (T) O-glycan, including capacity for core 2 (KO GCNT1) and sialylation of core 1 (KO ST3GAL1/2 and ST6GALNAC2/3/4). This cell line would thus have no competitive enzymes working on the Tn O-glycan substrate or enzymes converting the T O-glycans when produced (Fig. 5a). We then installed the full coding construct of HsC1GalT1 and mutants (R140A, Y201A, Y206A, D240A, W285A and Y310A) by targeted knock-in (KI) (Supplementary Fig. 11 and Fig. 5d). The induction of core 1 (T) expression on cell surface was evaluated by flow cytometry with the anti-T monoclonal antibody (mAb) 3C9 (Fig. 5b, c). mAb 3C9 did not bind HEK293T \( ^{3Tn} \) cell but strongly bound the cells after KI of WT HsC1GalT1. KI of HsC1GalT1 mutants R140A, Y201A, Y206A and Y310A produced partial restoration of 3C9-binding with Y206A being the least effective, while KI of D240A and W285A mutants produced no binding suggesting these were completely inactive (Fig. 5c). Therefore, our results support that the D240 in HsC1GalT1 (D255 in DmC1GalT1) is the catalytic base, and the Y206 (Y218A\(_{\text{Dm} \text{C1GalT1}}\)) and W285 (W300\(_{\text{Dm} \text{C1GalT1}}\)) residues are also critical in recognition and catalysis. Overall, the results in cells with the HsC1GalT1 mutants match those found with the DmC1GalT1 mutants, validating that the DmC1GalT1 enzyme serves as a model for the human enzyme.

Putative 3D structures derived from Molecular dynamics (MD) simulations. We generated putative 3D structures for the apo form of the enzyme, as well as for the enzyme in the presence of UDP-Gal and for complexes between DmC1GalT1 and the glycopeptides APDT\(^*\)RP, APDS\(^*\)RP, P2, P4, and P7 (Fig. 6, Supplementary Figs. 12–16 and Methods). According to these calculations, the protein retains its 3D structure almost unchanged in the presence of UDP-Gal and upon the formation of the ternary complex with APDT\(^*\)RP (Fig. 6a), consistent with the lack of an induced-fit mechanism. In all complexes, the hydrogen bonds between the GalNAc moiety and the enzyme present in the X-ray structure were observed in the MD simulations, regardless of the peptide sequence (Supplementary Figs. 12–15). Moreover, the glycosidic linkage of all glycopeptides exhibited a staggered conformation which could be a mechanism used by the enzyme to glycosylate α-GalNAc-Thr and α-GalNAc-Ser residues in a similar manner (Supplementary Fig. 14). For the peptide APDS\(^*\)RP (Supplementary Fig. 16), the calculations show the absence of a CH-π interaction between Trp300 and Ser4. However, a similar interaction was observed between the hydrogen atoms of CB2 of this residue and the side chain of Tyr304. For APDT\(^*\)RP in complex with DmC1GalT1, the GalNAc and UDP-Gal showed the correct orientation, with a distance O3-GalNAc/C1-Gal \(<5.5\ \text{Å} \) throughout the entire trajectory, which is consistent with the inversion mechanism (Fig. 6b, c and Supplementary Fig. 15). In addition, the binding mode for the glycopeptide observed by MD
simulations agrees with the STD experiments described above (Supplementary Table 7). The absence of UDP-Gal does not significantly alter the interactions between the glycopeptides and the enzyme compared to the ternary complexes, except for the glycopeptide APDT*R, which agrees with the experimental results. In absence of UDP-Gal, Arg5 of the peptide interacts with Glu254, which leads to a shift of the GalNAc unit from its binding site. Indeed, some frames of the MD simulations of the binary APDTRP-DmC1GalT1 complex show a lack of hydrogen bonds between OH3 and OH4 of the sugar and Asp255 (Fig. 6d). Therefore, the occurrence of UDP-Gal in this complex may stabilize the positive charge and hinder the interaction of Arg5 with Ghu254. On the contrary, the absence of UDP-Gal may favor nonspecific interactions with the protein, explaining the absence of binding of this glycopeptide to the enzyme when UDP is not added. For glycopeptide P2, the MD simulations show three relevant interactions between the peptide fragment and the protein (Supplementary Fig. 13). A hydrogen bond between the side chain of Trp300 of the protein and the carbonyl group of Gly5 is present for about 94% of the trajectory. Moreover, the side chains of Tyr231 and Phe299 are involved in CH-π interactions with the N- and C-terminal residues of the peptide, respectively. Similarly, P7 forms a hydrogen bond between its Gly and Trp300, as well as a CH-π interaction between its N-terminal residue and Tyr213 (Supplementary Fig. 13). Finally, the simulations of P4 in complex with UDP-Gal and DmC1GalT1 indicate a highly populated hydrogen bond between Gly5 and the side chain of Trp300 (population ≈ 95%), together with stabilizing contacts between the protein and both the N- and C-terminal regions of the peptide. Also, in the case of glycopeptide P4, good agreement is observed.

Fig. 5 Flow Cytometry Analysis of the reinstallation of T glycoform with HsC1GalT1 mutants. a The O-glycosylation pathway is indicated with the name of the enzymes involved in the synthesis of O-glycan structures. Note that the non-expressed genes and the predicted basic glycan features missing in HEK293 cells are faded out based on RNA-seq analysis. The engineering strategy to develop HEK293Tn clone was performed with CIGALT1 gene KO in HEK293core1 cells (HEK293KO GONTTSTGAL2/STGALT1NAC2/3/4) followed by the individual KI of C-terminal Myc tagged HsC1GalT1 WT and 6 mutants complementary DNA. b Flow cytometry analysis with the core 1 specific monoclonal antibody 3C9 (1 to 100 diluted hybridoma supernatant) were used to evaluate the cell surface level of T or core 1 glycoform. c Bar diagrams show mean fluorescence intensities of the mAb 3C9 binding. d Immunocytochemistry analysis of single KI clones with anti-Myc-tag mAb 9E10 detecting C-terminal Myc tag of C1GalT1. Note that HEK293core1 clone (HEK293KO GONTTSTGAL2/STGALT1NAC2/3/4) has endogenous expression of HsC1GalT1 without Myc tag. Images are representative of three experiments (n = 3 independent experiments). Scale bar = 20 μm.
between the glycopeptide-protein interproton distances derived along the MD simulations in the presence of UDP-Gal and the STD responses estimated for GalNAc, Pro1, Ala2, and Thr4 (Supplementary Table 8). Transient close contacts between Ala3 or Tyr7 with protein residues were observed throughout the MD trajectory, which could also explain the STD response for these amino acids.

Discussion

The C1GalT1 is critical for the immediate elongation and processing of GalNAc-type protein O-glycosylation in most normal cells, and here we provided insights into this enzyme and its catalytic mechanism by solving the crystal structure of the Drosophila orthologue. The presence of a private chaperone has been attributed to the fact that the higher eukaryotes C1GalT1 are not N-glycosylated (the lower eukaryotes C1GalT1 are N-glycosylated[11]; see Supplementary Fig. 8a). Yet, this may not necessarily be the explanation because several human GTs lacking N-glycosites are still properly folded without the need for a chaperone[21,33]. We hypothesize here, based on our structural analysis, that Cosmc is likely important in C1GalT1 dimer interface formation in higher eukaryotes. Our results also provide an explanation for the conundrum that the first step in O-glycosylation is covered by the largest isoenzyme family catalyzing a single glycosidic linkage presumably to cover the wide variation in substrate sequences in the proteome, while the immediate next step in elongation is covered by only a single non-redundant enzyme, C1GalT1. C1GalT1 was found to have very broad acceptor substrate specificity and clearly showed the strongest interactions with the GalNAc acceptor sugar residue. Interactions of C1GalT1 with the peptide were identified with some sequence preferences, but these were shown not to be critical for activity. This suggests that the C1GalT1 can serve widely in core 1 O-glycan elongation and cover the entire spectrum of O-glycans distributed in the proteome. Clearly, C1GalT1 may have different kinetic properties for GalNAc glycosylated O-glycopeptides, but in normal cells, most if not all O-glycans are elongated to mask expression of the cancer-associated Tn structure. Exposure of Tn in cancer cells is generally not due to inactivating mutations in the C1GalT1 gene[18] and heterogeneous with both Tn and core 1 structure are found in most cancer cells[53]. Thus, reduced expression of C1GalT1 may instead lead to incomplete O-glycan elongation with preferences for O-glycan sites that are less preferred substrates for C1GalT1.

Core 1 O-glycan structures synthesis in human cells also depend on the expression and equilibrium between C1GalT1 and other GTs such as core 3 synthase (B3GnT6) and ST6GalNAc-I (Fig. 5a). While core 3 synthase adds GlcNAc onto the initial GalNAc OH3, ST6GalNAc-I transfers sialic acid onto the Tn antigen GalNAc OH6 forming the STn antigen. In most human cells, the core 1 O-glycan structure is the most abundant precursor for building complex O-glycans. However, e.g., in normal colon, the major O-glycan core structure is the core 3 structure, while interestingly goblet cells also accumulate acetylated STn intracellularly[54,55]. The core 3 synthase is up-regulated in colonic cells while the C1GalT1 is also expressed, and competition for the initial GalNAc residues attached may be in favor of core 3. An explanation for the accumulation of acetylated STn glycoforms intracellularly in goblet cells is less obvious. However, ST6GalNAc-I is selectively expressed in the colon and can compete with C1GalT1, and C1GalT1 cannot transfer to STn O-glycans. Our structural studies provide a molecular basis for why C1GalT1 cannot glycosylate the STn antigen. The sialic acid will likely clash with Tyr201 (Tyr213DmC1GalT1), Tyr206 (Tyr218DmC1GalT1) and Tyr289 (Tyr304DmC1GalT1) of HsC1GalT1 (see Fig. 4a and particularly the position of the GalNAc OH6).
Different crystal structures of initiating GTs with acceptor substrates have revealed that these enzymes employ different strategies to recognize their protein substrates\textsuperscript{23,36,56}. However, for follow-up GTs acting immediately after the first monosaccharide is attached to the protein backbone, only the crystal structure of POMGnT1 has been published, revealing that this enzyme tethers the mannose moiety through hydrogen bond interactions while the peptide is exclusively recognized by hydrophobic interactions with the enzyme\textsuperscript{77}. Herein, the integration of X-ray crystallographic data, STD NMR and molecular modeling allowed to decode the recognition of APDT\textsuperscript{3}RP by DmCiGalT1. From the visual inspection of the complex’s crystal structure, the APDT\textsuperscript{3}RP is mainly recognized through GalNAc unit by a network of H-bonds involving OH3, OH4 and OH6. This observation is complemented by the STD NMR spectra of APDT\textsuperscript{3}RP in presence of DmCiGalT1, which provides information about GalNAc aliphatic protons, and pinpoints that GalNAc protons are those in closer contact with the protein, reinforcing the conclusion that GalNAc is the main contact point to the enzyme. With respect to the peptide, both techniques indicate α-GalNAc-Thr and the methyl group of Thr are involved in the recognition, helping to stabilize the peptide by a mix of hydrophobic and hydrogen bond interactions.

Our structural studies also show the striking finding that the enzyme imposes a non-natural staggered conformation to α-GalNAc-Thr linkage that is typically found in α-GalNAc-Ser. In doing so, α-GalNAc-Thr behaves highly similar to α-GalNAc-Ser except for the Thr methyl group, whose gain in binding through interaction to neighboring active site aromatic residues might compensate for the energy penalty due to the unfavorable conformation for the α-GalNAc-Thr. This feature, which is essential to achieve glycosylation, is likely behind why this enzyme indissolently glycosylates both acceptor glycosomes. It is tempting to speculate that it might be more structurally and energetically advantageous for C1GalT1 to impose the staggered conformation to α-GalNAc-Thr, which is low populated in solution, than to adapt its active site to the main conformer found in solution for α-GalNAc-Ser (staggered conformation) and α-GalNAc-Thr (eclipsed conformation). A similar unfavorable enzyme-induced acceptor substrate conformation has been reported for FUT8, in which the enzyme also imposes a more unstable anti-ψ conformation to the core-chitobiose GlcNAc moieties of the N-glycan to achieve core-fucosylation\textsuperscript{78}. This clearly exemplifies that enzymes do not always select for more stable acceptor substrate conformations to achieve catalysis and that in cases like the C1GalT1 or FUT8, a more unstable conformation is selected for catalysis.

In summary, we propose that C1GalT1 follows the typical S\textsubscript{2}2 mechanism described for inverting GTs, and reveal the molecular basis of glycopeptide recognition. We further uncover that C1GalT1 imposes a high-energy and unfavorable conformation to α-GalNAc-Thr as a required step for glycosylation. This is a remarkable example of how GTs have implemented strategies to promote conformational changes in the acceptor substrates to achieve glycosylation.

Methods

Production of DmCiGalT1-expressing baculovirus. The DNA sequence encoding amino acid residues of the DmCiGalT1 (aa S73-Q388) was designed and synthesized (GenScript (USA)) for expression in HEK293 cells (GIBCO). The construct, containing the 5′-end a recognition sequence for Agel, and at the 3′-end a recognition sequence for KpnI, was cloned into a pHLSec plasmid that contained a sequence encoding a 6xHis tag at the 3′-end followed by a stop codon, rendering the vector pHLSec-DmCiGalT1-6His. The cloning of the construct into the pFastBac-mellitin-DmCiGalT1-6His was also performed by GenScript.

Recombinant baculovirus was produced by the method of Tn7 transposition method in DH10Bac according to the Bac-to-Bac\textsuperscript{TM} Expression System (Invitrogen).
DmcG1aT1-His and its mutants were then loaded in HiLoad 26/60 Superdex 75 Column (GE Healthcare) previously equilibrated with buffer C. Quantification of the protein was done by absorbance at 280 nm (theoretical extinction coefficient (ε280 nm values ranging between 5890 M⁻¹ cm⁻¹ and 64290 M⁻¹ cm⁻¹ depending on the protein).

Crystalization and data collection. Crystals of the DmcG1aT1 model obtained from alpha fold 2 server99. Initial phases were further improved by cycles of manual model building in Coot62 and refinement with REFMAC5. The final structure of the DmcG1aT1-373-Q388-UDP-Mn2⁺-APDTRP complex was validated with PROCHECK, model statistics are given in Supplementary Table 6. The structure was solved by molecular replacement with Phaser60 using the DmcG1aT1 model obtained from alpha fold 2 server99. The Ramachandran plot for the DmcG1aT1-373-Q388-UDP-Mn2⁺-APDTRP complex shows that 87.8%, 11.2%, 1% and 0.0% of the amino acids are in most favored, allowed, generously allowed and disallowed regions, respectively.

Isothermal titration microcalorimetry (ITC). ITC was used to characterize the interaction of DmcG1aT1-373-Q388 with UDP, APDTRP with and without UDP, APDTRP with UDP, and P4/P7 with and without UDP. All experiments were carried out in an Auto-iTC200 (Microcal, GE Healthcare) at 20 °C. The concentration of enzyme with 800 µM UDP in 25 mM TRIS pH 7.5, 150 mM NaCl and 1 mM MnCl₂. The titrations with APDTRP and UDP/P7 in the absence of enzyme were carried out at 60 µM of the enzyme with 2 mM P4/P7 in 25 mM TRIS pH 7.5, 150 mM NaCl and 1 mM MnCl₂. To determine the Kα for APDTRP, APDTRP and P4/P7 under an excess of UDP, the experiments were made in 25 mM TRIS pH 7.5, 150 mM NaCl, 1 mM UDP and 1 mM MnCl₂. The concentration of the enzyme was 60 µM for the titrations of APDTRP and P4/P7 and 50 µM for the titration with P4/P7. The concentration of the (glyco)peptides was 2 mM in all the ITC experiments. The experiments were performed in duplicate. Data integration, correction and analysis were carried out in Origin 7 (Microcal). The data fit to a one-site equilibrium-binding model. Stoichiometry (n) of binding in all cases was –1:1 except for the UDP where n = 0.4.

Kinetic analysis. Enzyme kinetics for the DmcG1aT1-373-Q388, DmcG1aT1-373-Q388 and the mutants were determined using the UDP-Glo luminescence assays (Promega). Reactions contained 500 nM of the enzymes in 25 mM Tris pH 7.5, 150 mM NaCl, 50 µM MnCl₂, 1 mg/ml BSA (bovine serum albumin) and 500 µM UDP-GlcNAc in the presence of variable concentrations of the ligands and α-O-methyl-GlcNAc. The concentrations of P1-P7 and the α-O-methyl-GlcNAc ranged from 12.5 to 500 µM and from 125 µM to 2 mM, respectively. The concentrations of APDTRP and APDTRP ranged from 1.25 to 1000 µM to get a better kinetic non-linear fit. Michaelis–Menten fitting in order to determine the kinetic parameters for UDP-GlcNAc binding using DmcG1aT1-373-Q388, we used 500 nM DmcG1aT1-373-Q388 and variable concentrations of UDP-GlcNAc (12.5 µM–1 mM) in the presence of P4 and APDTRP at a saturated concentration (250 µM and 1 mM respectively, which was approximately five-fold higher than the Kα,off value). For the mutants, the activity assay was performed using the mutants at 500 nM with 500 µM UDP-GlcNAc and 500 µM APDTRP. Reactions were incubated 30 min at 37 °C and stopped using 5 µl of UDP-detection reagent at a 1:1 ratio in a white and opaque 384-well plate. Then, the plates were incubated in the dark for 1 h at room temperature. Subsequently, the values were obtained by using a Synergy H1 (BioTek). To estimate the amount of UDP produced in the glycosyltransferase reaction, we created a UDP standard curve. The values were corrected against the UDP-GlcNAc hydrolysis and were fit to a non-linear Michaelis–Menten program in GraphPad Prism 8 software from which the Kα,off, kα,off and kα,off/Kα,off along with their standard errors were obtained. All experiments were performed in at least three replicates for the determination of the activity of the mutants that were performed in triplicate.

Solid-phase peptide synthesis (SPPS). (Glyco)peptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Fmoc-Thr[GalNAc(Ac)₃, α-D]-OH or Fmoc-Ser[GalNAc(Ac)₃, α-D]-OH (2.0 equiv) were synthesized as before100. Peptide couplings were then performed using HBTU [(2(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate), while all other Fmoc amino acids (5.0 equiv.) were automatically coupled using oxima pure/DIC (N,N-diisopropylcarbodiimide), the O-acetyl groups of GalNAc moiety were removed in a mixture of NH₄OH/H₂O/MeOH (7:3). (Glyco)peptides were then released from the resin, and all acid sensitive side chains protecting groups were simultaneously removed using TFA 95%, TIS (trisopropylsilylether) 2.5% and H₂O 2.5%, followed by precipitation with cold dethyl ether. The crude products were purified by HPLC on a Phenomenon Luna C18(2) column (10 µm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 10 ml/min.

Peptide preparation. All the peptides used in this work were dissolved at 100 mM in buffer 25 mM Tris-HCl pH 7.5. The pH of each solution was measured with pH strips and when needed adjusted to pH 7–8 through the addition of 0.1–5 µL of 2 M NaOH.

NMR experiments. All NMR experiments were recorded on a Bruker Avance III 600 MHz spectrometer equipped with a 5 mm inverse detection triple-resonance cryogenic probe head with z-gradients. The 1H-NMR resonances of the (glyco)peptide solutions were completely assigned by standard 2D-1H 1D (10–20 ms mixing time), 2D-NOESY (400 ms mixing time) and 2D 1H,13C-HSQC experiments at 283 and 298 K. The α-O-methyl-GalNAc (Carboxymethyl, MM06786) was assigned with 2D-TOCSY/NOESY and 1H,13C-HSQC experiments at 298 K. Typical concentrations were around 1 mM for the heteronuclear experiments in 25 mM Tris(1,1,1,3,3-pentamethyldimethylammonium) chloride buffer, pH 7.5, with 150 mM NaCl and 2 mM MnCl₂. Some of the STD NMR experiments were accomplished in the presence of UDP (135 µM). In the presence of 150 mM MnCl₂, strong paramagnetic relaxation enhancements are observed for UDP, which prevents the observation of UDP proton signals in the NMR spectra. However, the presence of MnCl₂ does not preclude the observation of the other signals of the ligand information we required for STD experiments. STD NMR spectra (stddiffesgp pulse sequence from Bruker pulse program library) were acquired with 1728 scans and 64 K data points, in a spectral window of 12335.53 Hz centered at 2818 Hz. Selective saturation (on resonance) was performed by irradiating at 7 and/or −0.5 ppm (depending if the ligand contains or not aromatic residues) using a series of 40 t1/2p2000-shaped (from 28 ppm to 28–22 ppm) (50 ms) or a total saturation time of 2 s, a relaxation delay of 4 s. For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. Proper control experiments were performed for each ligand in the absence of protein and residual STD signals of the methyl groups of Ala/Thr were observed. This result was taken in account (subtracted) when analyzing the STD experiment in presence of DmcG1aT1-373-Q388. Protein control experiments were also accomplished using DmcG1aT1-373-Q388 in absence of a ligand and also subtracted from STD experiment. The STD spectrum (Istd) was obtained by subtracting the on-resonance spectrum (Ioff) to the off-resonance spectrum (Ioff). The % of STD (Istd/Ioff × 100) was estimated by comparing the integral of the signal in the STD spectrum (Istd) with the signal intensities of the reference spectrum (Ioff). The STD amplification factor (STDα) was also estimated by multiplying the % STD values by the ligand excess11, which in the case of our experiments was 35 for every ligand. To determine the STD-derived epitope map the relative % of STD values were completely assigned by standard 2D-1H 1D (10–20 ms mixing time) and 2D-NOESY (400 ms mixing time) and 2D 1H,13C-HSQC experiments. A two-stage geometry optimization approach was used to find the UDP bound to DmcG1aT1-373-Q388. Molecular dynamics (MD) simulations. The crystal structure of DmcG1aT1-UDP-Mn2⁺-APDTRP was superimposed with the human B3GNT2-UDP-GlcNAc complex (PDB entry 7HL), providing the coordinates of the UDP-GlcNAc in an identical location as that found for the UDP bound to DmcG1aT1 (see Fig. 3e illustrating that B3GNT2 and DmcG1aT1 are superimposed very well). Subsequently we generated the DmcG1aT1-UDP-GlcNAc-Mn2⁺-APDTRP complex, we replaced the UDP-GlcNAc by UDP-Gal resulting in the UDP-Gal-Mn2⁺-APDTRP complex. The other complexes were generated by mutating and adding or removing the corresponding residues with PyMOL 2.5. The calculations were carried out using AMBER 20 package, which was implemented with ff14SB and GLYCAM4 force fields. Each complex was immersed in a water box with a 10 Å buffer of TIP3P water molecules. The system was neutralized by adding explicit counter ions (Na+ or Cl−). A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules, and the second stage is an unrestrained minimization of all the atoms in the simulation box, and the number of step was 10000. The final complexes were then relaxed for 5 ps using the npt ensemble of 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal mol⁻¹ were applied to the solute, and the Andersen
temperature-coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method. An 8 Å cut-off was applied to Lennard-Jones 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. Adaptively Biased Molecular Dynamics method46 implemented in AMBER 20 was used to calculate the free-energy maps for the APDT7*RP and APDS*RP glycopeptides in water at 300 K.

Cell culture. All isogenic glycoengineered HEK293 cell lines were cultured in DMEM (Sigma-Aldrich) supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich) and 2 mM GlutaMAX (Gibco) in a humidified incubator at 37 °C and 5% CO2.

CRISPR/Cas9-targeted KO in HEK293 cells. CRISPR/Cas9 KO was performed using the GlycoCRISPR resource containing validated gRNAs libraries for targeting CRISPR/Cas9-targeted KO in HEK293 cells.

Construction of C1GalT1 enzyme and site-directed mutants. The codon-optimized full coding human C1GalT1 containing a C-terminal Myc-tag was synthesized by Genewiz USA and subcloned into pEP71 vector (Addgene ID 90018) for AAV5 targeting KO. The site-directed mutagenesis was performed by Genscript with targeting the six candidate amino acid residues (R140, Y201, Y206, D240, W285, and Y310) replaced to Ala.

ZFN-mediated KO of C1GalT1 variants in HEK293T cells. For site-directed knock-in (KI) a modified OlBiGale targeted AAV5 safe harbor site KO strategy utilizing two inverted ZFN binding sites flanking the C1GalT1 variants in donor plasmids were used. KI was performed as described before for targeted KO with 1 μg of each ZFN tagged with GFP/Crein and 2 μg donor plasmid. 48 h after transfection, the 10–15% most highly expressed cell pool (KI pool) for both GFP and Crein was enriched by FACs (SONY SH800). After 1 week of culture, the bulk-sorted cells were single cell-sorted into 96-well plates. The targeted KI clones were screened by PCR using a primer pair specific for the junction area between the donor plasmid and the human AAV5I locus, as well as a primer pair flanking the targeted KI locus. An allele-specific WT PCR (forward primer: 5'-CCGATCTCCCCGAGTCTAG-3'; reverse primer: 5'-TGAGATCCCGTGAAAC-3') amplifying C1GalT1 gRNA targeting sites and were further verified by Sanger sequencing.

Flow cytometry analysis. The level of core 1 structure on cell surface was measured by flow cytometry with mouse mAb 3C9 (an in-house produced antibody) specific to core 1 glycosylation45. Cells were incubated on ice with 3C9 mAb (undiluted hybridoma supernatant which is equivalent to 1:1 dilution) for 30 min, followed by washing and incubation with Alexa Fluor 647 conjugated goat anti-mouse IgM (1 μg/mL) (Invitrogen, catalog: A21235) for 30 min. Diluting and washing was performed in PBS with 1% BSA and cells were resuspended for flow cytometry analysis (SONY SA5800). Mean fluorescence intensity of the binding of mAb 3C9 populations was quantified by FlowJo software (FlowJo LLC).

Immunocytochemistry. Cells were fixed with cold acetone for 10 min and incubated with anti Myc tag dilution (ATCC, catalogue: CRL-1729) and mAb 3C9 (undi- diluted hybridoma supernatant which is equivalent to 1:1 dilution) overnight at 4 °C, followed by secondary Alexa Fluor 594 conjugated goat anti-mouse IgM (1 μg/mL) (Invitrogen, catalog: A-21204). All samples were imaged using a Zeiss Axioskop 2 plus with an AxiosCam MR3 followed by analysis with Imager (NIH).

Received: 1 December 2021; Accepted: 31 March 2022; Published online: 03 May 2022

References

1. Hurtado-Guerrero, R. Recent structural and mechanistic insights into protein O-GalNAc glycosylation. Biochem. Soc. Trans. 44, 61–67 (2016).
2. Bennett, E. P. et al. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glyobiology 22, 736–756 (2012).
3. de Las Rivas, M., Lira-Navarrete, E., Gerken, T. A. & Hurtado-Guerrero, R. Polypeptide GalNAc-Ts from redundancy to specificity. Curr. Opin. Struct. Biol. 56, 87–96 (2019).
4. Ju, T., Otto, V. I. & Cummings, R. D. The N-glycan-switch: collaboration of core 1,3-galactosyltransferase. J. Biol. Chem. 277, 178–186 (2002).
5. Cummings, R. D. Stuck on sugars: how carbohydrates regulate cell adhesion, recognition, and signaling., Glycoconj. J. 36, 241–257 (2019).
6. Kudella, M. R., Jr., Tu, T., Heimburg-Molinaro, J. & Cummings, R. D. Simple sugars to complex disease–mucin-type O-glycans in cancer. Adv. Cancer Res. 126, 53–135 (2019).
7. Tu, T. & Cummings, R. D. A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. Proc. Natl Acad. Sci. USA 104 (2008).
8. Hanes, M. S., Lomewmen, K. W. & Cummings, R. D. Biochemical characterization of functional domains of the chaperone Cosmc. PLoS ONE 12, e0180242 (2017).
9. Yoshida, H. et al. Identification of the Drosophila core 1 beta,3-galactosyltransferase gene that synthesizes T antigen in the embryonic central nervous system and hemocytes. Glyobiology 18, 1094–1104 (2008).
10. Zeng, J. et al. Cosmc controls B cell homing. Nat. Commun. 11, 3990 (2020).
11. Zeng, J. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
12. Zeng, J. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
13. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
14. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
15. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
16. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
17. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
18. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
19. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
20. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
21. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
22. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
23. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
24. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
25. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
26. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
27. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
28. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
29. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
30. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
31. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
32. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
33. Trend, F. et al. Cosmc controls B cell homeostasis. Nat. Commun. 11, 958 (2020).
34. Trend, F. et al. Cosm...
controlled by the amino acid sequence and glycosylation of glycopeptide substrates. *EUR. J. Biochem.* **221**, 1039–1046 (1994).

26. Muller, R. et al. Characterization of mucin-type core-1 beta1-3 galactosyltransferase homologous enzymes in Drosophila melanogaster. *FEBS J.* **272**, 4295–4305 (2005).

27. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).

28. Martinez-Saez, N., Decipio, S. & Serine and Threonine O-Glycosylation Points: Implications for Molecular Recognition of the N-T Antigen by an anti-MUC1 Antibody. *Angew. Chem. Int. Ed. Engl.* **54**, 9830–9834 (2015).

29. Martinez-Saez, N., Peregrina, J. M. & Corzana, F. Principles of mucin structure: implications for the rational design of cancer vaccines derived from MUC1 glycopeptides. *Chem. Soc. Rev.* **46**, 7154–7175 (2017).

30. Nason, R. et al. Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* **12**, 4070 (2021).

31. Clement, E. M. et al. Mild POMGnT1 mutations underlie a novel limb-girdle muscular dystrophy variant. *Arch. Neurol.* **65**, 137–141 (2008).

32. Akasaka-Manya, K., Manya, H., Mizuno, M., Inazu, T. & Endo, T. Effects of length and amino acid sequence of O-mannosyl peptides on substrate specificity of protein O-linked mannos beta1,2-N-acetylgalcosaminyltransferase 1 (POMGnTI). *Biochem. Biophys. Res. Commun.* **410**, 632–636 (2011).

33. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **38**, 1784–1788 (1999).

34. Bamborough, G. N. et al. REFMACS for the refinement of macromolecular crystal structures. *Acta Cryst. D Biol. Crystallogr*. **67**, 355–367 (2011).

35. Plattner, C., Hofener, M. & Seward, N. One-pot azidochlorination of glycals. *Org. Lett.* **13**, 545–547 (2011).

36. Garcia-Garcia, A. et al. Characterization of an unprecedented lectin-glycosyltransferase-mediated synthesis of glycans. *Nat. Commun.* **7**, 1352 (2016).

37. Madariaga, D. et al. Serine versus threonine glycosylation with alpha-O-mannosyltransferase 2. *Nat. Chem. Biol.* **12**, 240–246 (2016).

38. Kuwabara, N. et al. Carbohydrate-binding domain of the POMGnTI stem region modulates O-mannosylation sites of alpha-dystroglycan. *Proc. Natl Acad. Sci. USA* **113**, 9280–9285 (2016).

39. Schatz, J. & Suppan, M. A new single-step protocol for rapid baculovirus-driven protein production in insect cells. *MBCM* **17**, 83 (2017).

40. Maier, J. A. et al. FL-658: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

41. Babin, V., Roland, C. & Sagui, C. Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **128**, 134101 (2008).

42. Narimatsu, Y. et al. A validated gRNA library for CRISPR/Cas9 targeting of the human glycosylation enzyme genome. *Glycobiology* **28**, 295–305 (2018).

43. Yang, Z. et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* **33**, 842–844 (2015).

44. Steentoft, C. et al. A validated collection of mouse monoclonal antibodies to human glycosyltransferases functioning in mucin-type O-glycosylation. *Glycobiology* **29**, 645–656 (2019).

45. Tellinghusen, J. Isothermal titration calorimetry at very low *fi* *Ann. Biochem.* **373**, 395–397 (2008).

46. Turnbull, W. B. & Daranis, A. H. On the value of c: how many affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866 (2003).

47. Corzana, F. et al. New insights into alpha-GalNAc-Ser motif: influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **128**, 14640–14648 (2006).

Acknowledgements

We thank the Diamond Light Source (Oxford, UK) synchrotron beamline I24 (experiment numberMX2209-11). We thank ARAID, the Agencia Estatal de Investigación (AEI, BFU2016-75633-P and PID2019-105451GB-I00 to R.H.-G., and RTI2018-099592- B-C21 to F.C.), Gobierno de Aragón (E34_R17 and LMP58_18 to R.H-G.) with FEDER support. A.M.G-R. thanks the Spanish Ministry of Science, Innovation and University (COFUND, BTO-13555/2016, co-funded by FEDER through COMPETE 2020, POCI and POCI and FCT through PIDDAC). A.M.G-R. thanks the Spanish Ministry of Science, Innovation and University for the FPI fellowship. The research leading to these results has also received funding from the FP7 (2007–2013) under BioStruct-X (grant agreement N°238570 and BIOSTRUCTX-5186).

Author contributions

R.H.-G. designed the crystallization construct and solved the crystal structure. A.M.G-R. performed the expression and purification of all proteins, the enzyme kinetics and ITC experiments, and crystallized the complex. A.M.G-R. refined the crystal structure. F.C. performed the molecular mechanics and MD calculations. I.C. synthesized the glyco-peptides. A.S.G. and H.Co. performed the STD NMR experiments, and crystallized the complex. A.M.G-R. restructured the crystals. A.M.G-R. designed the crystallization construct and solved the crystal structure. A.M.G-R. wrote the article with mainly the
contribution of F.M., F.C., Y.N., A.M.G.-R., and H.C. All authors read and approved the final paper.

**Competing interests**
The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-29833-0.

**Correspondence** and requests for materials should be addressed to Francisco Corzana or Ramon Hurtado-Guerrero.

**Peer review information** *Nature Communications* thanks Chris Oostenbrink, Thomas H. Peters and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2022