The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences

Matilde de las Rivas, Erandi Lira-Navarrete, Earnest James Paul Daniel, Ismael Compañón, Helena Coelho, Ana Diniz, Jesús Jiménez-Barbero, Jesús M. Peregrina, Henrik Clausen, Francisco Corzana, Filipa Marcelo, Gonzalo Jiménez-Osés, Thomas A. Gerken & Ramon Hurtado-Guerrero

The polypeptide GalNAc-transferases (GalNAc-Ts), that initiate mucin-type O-glycosylation, consist of a catalytic and a lectin domain connected by a flexible linker. In addition to recognizing polypeptide sequence, the GalNAc-Ts exhibit unique long-range N- and/or C-terminal prior glycosylation (GalNAc-O-Ser/Thr) preferences modulated by the lectin domain. Here we report studies on GalNAc-T4 that reveal the origins of its unique N-terminal long-range glycopeptide specificity, which is the opposite of GalNAc-T2. The GalNAc-T4 structure bound to a monoglycopeptide shows that the GalNAc-binding site of its lectin domain is rotated relative to the homologous GalNAc-T2 structure, explaining their different long-range preferences. Kinetics and molecular dynamics simulations on several GalNAc-T2 flexible linker constructs show altered remote prior glycosylation preferences, confirming that the flexible linker dictates the rotation of the lectin domain, thus modulating the GalNAc-Ts’ long-range preferences. This work for the first time provides the structural basis for the different remote prior glycosylation preferences of the GalNAc-Ts.

DOI: 10.1038/s41467-017-02006-0
Polyepitope N-acetylgalactosaminyltrrasferases (GalNAc-Ts) are higher eukaryotic-retaining glycosyltransferases (GTs) that transfer a GalNAc moiety from uridine diphosphate N-acetylgalactosamine (UDP-GalNAc) onto Ser/Thr residues of proteins. This family of enzymes, which initiates the most common type of O-glycosylation in metazoans, plays profound roles for the structure, stability, and function of proteins. GalNAc-O-glycosylation is one of the most abundant types of protein glycosylation in eukaryotes that is chiefly found in the densely clustered heavily glycosylated domains of mucins and mucin-domain-containing glycoproteins, hence the name, mucin-type O-glycosylation. However, it is also clear that many other proteins contain isolated sites of GalNAc-O-glycosylation, where more than 5000 human proteins trafficking the secretory pathway have been identified to date containing one or more mucin-type O-glycans.

The initiation of mucin-type O-glycosylation by the GalNAc-Ts is one of the most complex regulated types of protein glycosylation. The GalNAc-Ts comprise a family of up to 20 iso-enzymes possessing a range of kinetic properties and expression patterns. These enzymes orchestrate with high fidelity, the initial patterns of O-glycosylation on diverse proteins, including the high-density regions in mucins, where 30–50% of the amino acids may be glycosylated. The GalNAc-Ts are unique among metazoan GT enzymes because in addition to their N-terminal catalytic domain adopting a GT-A fold (family 27 in the CAZy database), they possess a C-terminal GalNAc-binding lectin domain with a β-trefoil fold (classified as a carbohydrate-binding module 13 in the CAZy database), which provides additional functions to these enzymes (see below for further details). Both domains are linked through a short flexible linker whose motion has been suggested to be responsible for the dynamic conformational landscape of these enzymes.

GalNAc-Ts can be classified according to their different glycosylation capacity on peptide and glycopeptide substrates. Three distinctive glycosylation modes have been reported: catalytic domain-dependent glycosylation on “naked” peptides (Fig. 1a) and on glycopeptides (Fig. 1b), hereafter termed short-range glycosylation; and remote lectin domain-dependent glycosylation on glycopeptides (Fig. 1c), termed long-range glycosylation. Adjacent and proximal sites relative to an existing GalNAc glycosite are glycosylated in a catalytic domain-dependent manner (Fig. 1), with GalNAc-T being part of a small number of GalNAc-Ts (including GalNAc-T7 and T10) that will also glycosylate contiguous sites. Distant sites can be glycosylated in a lectin domain-dependent manner (Fig. 1) with individual GalNAc-Ts having distinct preferences on remote sites located N or C terminus from a prior GalNAc glycosite.

For example, GalNAc-T1/T2/T14 tend to glycosylate preferentially N-terminal sites remote from prior C-terminal GalNAc glycosities in glycopeptides, whereas GalNAc-T3/T4/T6/T12 exhibit the opposite preference (Fig. 1c). Furthermore, some GalNAc-Ts, including GalNAc-T5/T13/T16, appear to exhibit both orientational preferences for long-range glycosylation. Previous structural work with GalNAc-T2 in complex with “naked” peptides and glycopeptides has provided insight into the catalytic domain-dependent reaction, and also demonstrated how the GalNAc-T2 lectin domain guides catalysis to N-terminal acceptor sites very distant from a prior C-terminal GalNAc glycosite. However, the molecular basis of how other GalNAc-Ts, such as GalNAc-T3/T4/T6/T12, achieve the opposite long-range glycosylation preferences remains unclear.

The GalNAc-T4 isoform is of particular interest as it is the only isoform capable of glycosylating two out of the five acceptor sites in the partially glycosylated MUC1 mucin tandem repeat in a lectin domain-dependent manner. The density of glycosylation together with the structure of O-glycans in the mucin repeat regions is important features for the cancer-glycoforms of MUC1 being exploited for diagnostic as well as therapeutic purposes. Generally, GalNAc-T4 glycosylates very few isolated glycosylation sites in “naked” peptide acceptors, although it readily glycosylates the important Thr57 in PSGL-1 that is the essential O-glycan required for P-selectin-mediated leukocyte trafficking. We report herein a multidisciplinary approach on GalNAc-T4, combined with the characterization of different chimeras and mutants of GalNAc-T2, which begins to reveal the molecular basis of the long-range glycosylation preferences of the GalNAc-Ts on glycopeptide acceptors substrates.

Fig. 1 Modes of O-glycosylation found for the GalNAc-Ts. a-c These panels denote the three distinct modes of O-glycosylation performed by GalNAc-Ts: the neighboring glycosylation activity by the catalytic domain in a and b, and long-range lectin domain-mediated glycosylation activity in c. In c, two plausible models are suggested that might explain the different long-distance glycosylation preferences of GTs such as GalNAc-T3/T4/T6/T12. Oval-shaped figures in blue and green depict the lectin and catalytic domains, respectively. Peptides are indicated in blue and the black hexagon-shaped figure denotes the position of prior GalNAc moieties in the glycopeptides. Arrows indicate the positions of acceptor sites.
**Results**

Models explaining the long-range glycosylation preferences. Based on previous results, we can infer two possible models (see Fig. 1c) that might provide a plausible explanation for the GalNAc-T3/T4/T6/T12 long-range glycosylation preference in contrast to the one found for enzymes such as GalNAc-T1/T2/T14. In model 1, the rotation of the lectin domain GalNAc-binding site is required, thereby maintaining the same orientation of the peptide on the catalytic domain. In model 2, the lectin domain GalNAc-binding site would adopt the same orientation as found in GalNAc-T2 while the glycopeptide adopts an inverse orientation on the catalytic domain, an assumption unlikely to happen providing the high similarity of the peptide-binding groove among GalNAc-Ts.

The presence of STD-NMR signals for peptide 1 revealed that the “naked” peptide was very poorly recognized by GalNAc-T4 in agreement with the low glycosylation observed during the kinetics assays. For both monoglycopeptides 2 and 3 relatively weak STD-NMR signals were observed for the Ala and Thr methyl protons, suggesting weak recognition of the peptide by GalNAc-T4. However, the O-GalNAc glycosylated Thr of the monoglycopeptides 2 and 3 receive a higher STD enhancement. Remarkably, the sugar protons of the GalNAc moiety of the monoglycopeptides 2 and 3 (region 3.5 a 4.2 ppm of the spectrum together with the singlet at 2 ppm) showed unambiguous and more intense STD-NMR signals (Fig. 2b and Supplementary Fig. 1). In addition, the STD-NMR-derived epitope mapping of the GalNAc moiety is identical for both monoglycopeptides 2 and control α-methyl-GalNAc. In fact, for both compounds the highest STD response corresponded to the H2 proton of the GalNAc moiety followed by a modest STD of protons H4, H3, and N-acetyl group, suggesting similar lectin-binding modes for these molecules. In contrast, monoglycopeptide 3 presents a somewhat different GalNAc STD-NMR-binding pattern, where H2, H4, and the N-acetyl group of the GalNAc moiety have the highest STD-NMR response closely followed by the H3 proton. This suggests that the monoglycopeptide 3 has a distinct binding mode of interaction with the enzyme, which is different to the one found for monoglycopeptide 2 and α-methyl-GalNAc control. From our kinetic and STD-NMR results, we can infer that the unique binding of the GalNAc residue of monoglycopeptide 3 to the lectin domain, facilitates the correct orientation of the peptide acceptor onto the catalytic domain for optimal C-terminal glycosylation. Thus, even though monoglycopeptide 2 shows GalNAc binding to the lectin domain, its N-terminal acceptor site must be directed away from the catalytic domain, leading to the inability of GalNAc-T4 to glycosylate this substrate. Hence, the correct location of a prior glycosite in a monoglycopeptide, with respect to potential acceptor sites, is essential for an optimal interaction of the peptide within the catalytic domain, which in turn is critical for an efficient catalysis and specificity.

Crystal structure of GalNAc-T4. To further elucidate the molecular basis of GalNAc-T4’s monoglycopeptide glycosylation preference and rule out one of the two proposed models (Fig. 1),

![Figure 2](image_url)
we obtained triclinic crystals of GalNAc-T4 that were subsequently soaked with monoglycopeptide 3 and UDP and MnCl₂. The resulting crystal allowed us to solve the corresponding structure at a high resolution (1.90 Å) (Supplementary Table 1). Within the asymmetric unit, two independent GalNAc-T4 molecules were present. The crystal structure shows a compact structure with the typical GT-A fold and the lectin domain located at the N- and C-terminal regions, respectively (Fig. 3a). A sequence alignment between GalNAc-T2 and T4 displays a sequence identity of ~40% and a high resemblance between these two enzymes at the secondary structure level (Supplementary Fig. 2). In fact, superposition analysis of both enzymes shows that both catalytic domains are more structurally similar (root-mean-square deviation (RMSD) of 1.23 Å) than the lectin domains (RMSD of 1.82 Å) (Fig. 3b and Supplementary Fig. 3). A global superposition of both enzymes renders a larger RMSD of 2.12 Å, which suggests that a major shift has occurred between the orientation of the lectin domains (see below for a further discussion; Fig. 3b). As expected and accordingly to model 1, the GalNAc-binding site in GalNAc-T4 is located on the right side of the lectin domain, which requires a rotation of 28° with respect to the homologous site in GalNAc-T2 (Fig. 3b and Supplementary Fig. 4). To calculate the angle of rotation, the Cαs of Asp458, Trp282, and Asp459 were used. Note that the angle of rotation could be also determined using the Cαs of Asp458, Trp282, and Asp459. In this case, the rotation angle was 30°, which is very close to 28°. This is due to Trp282 of GalNAc-T2 and Trp286 of GalNAc-T4 adopt almost identical positions in the catalytic domain (Fig. 3b). This different position of the GalNAc-T4 GalNAc-binding site, consistent with the STD-NMR results, may explain why GalNAc-T4 promotes a distinct type of long-range glycosylation, consisting of the glycosylation of acceptor sites located at the C-terminal from prior GalNAc glycosites in glycopeptides. This also suggests that the location of the GalNAc-binding site of the lectin domain is coupled with the long-range glycosylation preferences of these enzymes.

**Lectin domain-binding site of GalNAc-T4.** Even though the crystals were soaked with a high concentration of peptide 3, UDP
and MnCl₂, we could only visualize well-defined density for the GalNAc-O-Thr moiety of monoglycopeptide 3 bound to the lectin α-subdomain GalNAc-binding site (Fig. 3a, c). Note that GalNAc-T3 lectin domains contain three potential carbohydrate-binding subdomains named α, β, and γ; however, in most GalNAc-Ts only one of them is functional.15, 16, 18, 22, 23. The almost absence of density for the backbone of monoglycopeptide 3, is consistent with the STD-NMR results (showing no significant STD with the acceptor peptide residues) and with the weak binding affinity of this peptide (Kₐ estimated in the mM range), determined by surface plasmon resonance (SPR) experiments in the presence of UDP and MnCl₂ (Supplementary Fig. 5). The discrepancies between GalNAc-T4’s high-affinity Kₐ obtained from our kinetics studies (Fig. 2a) versus the poor Kₐ values from the binding studies can be attributed to the fact that the enzyme kinetics was performed with UDP-GalNAc. It is known that this donor substrate stabilizes the flexible loop of the catalytic domain active site in a closed conformation, thus completing the formation of the peptide-binding groove and leading to an active enzyme.4,12 Thus, our observations in the absence of UDP-GalNAc represent the weak binding of peptide substrate to the open flexible loop conformation of the enzyme.

A closer inspection at the lectin domain GalNAc-binding site depicts that the GalNAc moiety was tethered by conserved residues, including Asp459, Asn483, Phe475, and His478, of the lectin α-subdomain GalNAc-binding site (equivalent residues in GalNAc-T2 are Asp458, Asn479, Tyr471, and His474; Fig. 3d). It is important to note that only the lectin α-subdomain GalNAc-binding site has been shown to be functional in both GalNAc-T2 and GalNAc-T4 lectin domains.4,18 Most of the interactions are through hydrogen bonds in both enzymes, although they also share CH-π interactions between the GalNAc moiety and Phe475/Tyr471. The importance of the Asp residue in recognition of the GalNAc moiety and its role in the lectin domain-mediated glycosylation has been previously reported.4,18 Furthermore, the crystal structure explained the GalNAc STD-NMR-derived epitope mapping. In particular, the structure shows that H4 and H3 of the GalNAc moiety are in close contact with Phe475 and His478, which is further supported by the STD response (92% for H4 and 72% for H3) observed for these protons in monoglycopeptide 3.

Molecular docking and dynamics simulations. To address how the lectin domain guides the delivery of potential peptide acceptor sites to the catalytic domain, we performed molecular docking and dynamics (MD) simulations on GalNAc-T4 in complex with UDP/Mn²⁺ and monoglycopeptide 3 (see “Methods” section). The calculations suggest that although the peptide bound at the catalytic domain was highly dynamic, the two potential peptide acceptor Thr residues were in close contact to the β-phosphate of UDP during the simulated timeframe (200 ns), thereby accounting for the glycosylation of this peptide (Fig. 3e, f, and Supplementary Fig. 6 and Supplementary Movie 1). A comparison with the GalNAc-T2 structure in complex with MUC5AC-13 and UDP/Mn²⁺ clearly provides evidence of the two different orientations of the lectin domains and supports our conclusion based on the coupling of the location of the GalNAc-binding site with the observed long-distance glycosylation preference (Fig. 3e, f).

The flexible linker dictates the lectin domain rotation. Previous studies suggested that the flexible linker allowed for the interdomain translational-like motion in these enzymes.4 We therefore rationalized that different flexible linkers found in these GTs might also be behind the rotation capacity of the lectin domain (Supplementary Fig. 7). To test this hypothesis, we initially performed 500 ns MD simulations in water solution of native GalNAc-T2 and GalNAc-T4. Simulations clearly showed that the lectin domains of GalNAc-T2 and GalNAc-T4 apo enzymes do not rotate in the considered timescale (Supplementary Movies 2 and 3). This suggests that the orientation of the GalNAc-binding site with respect to the catalytic domain may remain relatively fixed in a fully folded transferase. We then performed MD simulations on chimeras 1 and 2 that corresponded to GalNAc-T2 with two different lengths of the flexible linker from GalNAc-T3 (Table 2). In agreement with our hypothesis, we observed the prompt rotation of the lectin domain of the GalNAc-T2 chimera 2 toward a position similar to that in GalNAc-T4 (Fig. 4a, Supplementary Fig. 8, and Supplementary Movie 4). This stepwise motion was completed in ~30 ns and involved sequential extension of the flexible linker, rotation of the lectin domain around the Z and Y axes, and finally compression of the linker (Fig. 4a). Thus, initially the GalNAc-T3 linker was artificially forced to adopt the compact conformation of the GalNAc-T2 native linker, but early on in the simulation (after ~10 ns), the linker quickly recovered its more native extended conformation and as a consequence, the lectin domain springs and rotates sidewise to finally re-assemble in a structure similar to that of our GalNAc-T4. This motion is likely smoothed by the absence of strong interdomain

| Table 2 Sequences of the chimeras and mutant GalNAc-Ts used in this study |
|-----------------------------|-----------------------------|
| **Transfere**               | **Sequence of flexible linker** |
| GalNAc-T2                   | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T2-t₄Flexible linker | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T2-t₄Flexible linker | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| (Chimera 2)                 | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T2 (double mutant. R₄₃₈A-D₄₄₄A) | P₄₃₅ELRVPDHQDIAGF₄₄₇ |
| GalNAc-T3                   | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T4                   | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T4-t₂Flexible linker | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| (Chimera 4)                 | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T4-t₂Flexible linker-A | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| (Chimera 6)                 | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| GalNAc-T3-t₂Flexible linker | P₄₃₅ELRVPDHQDIAGF₄₄₇         |
| (Chimera 5)                 | P₄₃₅ELRVPDHQDIAGF₄₄₇         |

Above are the sequences of the native transferases and linker domain mutants and chimeras utilized in this work. Only the sequences of the linkers that are exchanged are shown. Shown names in (bold) are referred to in the text and figures. Note that some of the exchanged linker regions also include part of the succeeding β-strand of GalNAc-T2 and GalNAc-T4. This is due to the boundaries of flexible linkers are not well defined at the C terminus. The additional Pro in the flexible linker of GalNAc-T3/14, that is not present in GalNAc-T2, is underlined.
interactions in GalNAc-T2, which do take place in GalNAc-T4 (see for instance the persistent salt bridge between Arg397 and Glu487 in Supplementary Movie 3). Thus, the intrinsic conformational preference(s) of a short, unfolded fragment of ~13 residues, was found to determine the relative orientation of two very large protein domains (see also Supplementary Movies 6 and 7; these movies show the intrinsic different conformational preferences of the GalNAc-T2 (folded) and GalNAc-T3 (extended) flexible linkers by performing 500 ns MD simulations of these isolated fragments).

**Kinetics characterization of the chimeras.** To validate our MD simulations predictions, we expressed and purified chimeras 1–3 and performed a series of partial and full kinetic analysis against our peptide substrates (Fig. 4b, c) (also note that chimeras 4–6...
could not be expressed in our hands, see “Methods” section and Table 2. Remarkably, the GalNAc-T2 chimeras rendered glycosylation preferences similar to GalNAc-T3 and GalNAc-T4,15, 16 (preferring monoglycopeptide 3 over monoglycopeptide 2, by approximately fourfold, which is significantly different from native GalNAc-T2) despite comprising most of the GalNAc-T2 architecture (Fig. 4b, c and Supplementary Fig. 9). Interestingly, GalNAc-T2’s preference for these substrates varied with peptide substrate concentration, ranging from a high preference for monoglycopeptide 2 at low substrate concentrations to nearly equal preferences for both monoglycopeptides 2 and 3 at higher substrate concentrations (Fig. 4c). These differences arise from monoglycopeptide 2’s ~fourfold lower $K_m$ compared to monoglycopeptide 3, thus giving an over threefold increase in catalytic efficiency for monoglycopeptide 2 over monoglycopeptide 3 (Fig. 4c and Supplementary Table 2). These findings readily explain the previously reported differences in GalNAc-T2’s long-range glycosylation preferences.15, 17 This type of behavior was not observed in the kinetics of GalNAc-T4 or chimera 3 (note that detailed kinetics were not performed on chimera 1 or 2). Further analysis revealed that the changes in specificity for chimeras 1–3 were chiefly due to a decrease in the observed specific activity (chimera 1 and 2), and changes in $V_{\text{max}}$ and $K_m$ (chimera 3) against monoglycopeptide 2 compared to that of monoglycopeptide 3 whose specific activity and kinetic parameters remained relative constant (Fig. 4b, c, Supplementary Table 2, and Supplementary Fig. 9). It is also worth noting that the specific activities (and detailed kinetics when obtained) for chimeras 1–3 toward the “naked” peptide 1 are quite similar to those of GalNAc-T2, indicating that the GalNAc-T3/T4 flexible linkers in these chimeras did not affect the overall architecture of the GalNAc-T2’s active site (Fig. 4b, c, Supplementary Table 2, and Supplementary Fig. 9). Finally, we evaluated which Thr residues of the substrates were preferably glycosylated by Edman amino acid sequencing, finding that the Thr residue adjacent to the PXP motif was glycosylated by chimeras 1–3, as would be expected for the PXP motif (Supplementary Fig. 10). Little to no glycosylation was observed at the proximal Thr.

Together, our results suggest that the glycosylation preferences found for the GalNAc-T2 chimeras can be reasonably explained by the rotation of the lectin domain as predicted by the MD simulations of chimera 2. Thus, the lectin domains of these chimeras must begin to adopt an equivalent position to that of GalNAc-T4 to account for their altered long-range glycosylation preferences. We conclude that the flexible linker modulates the rotation of the lectin domain, which subsequently determines the orientation of the functional GalNAc-binding site of the lectin domain with respect to the catalytic domain. This modulation of the lectin domain relative to the catalytic domain likely determines the distinct long-range glycosylation preferences for all of the GalNAc-T isoforms. Our studies below further examine the role of the linker in dictating these long-range preferences.

Site-directed mutagenesis of the GalNAc-T2 and chimera 2. Next, we explored whether specific particular residues within the flexible linkers might be responsible for the lectin domain orientation. A multiple alignment of the flexible linkers (Supplementary Fig. 7) shows that GalNAc-T3/T4/T6/T12, which share the same long-distance glycosylation preferences15, 16, contain an additional Pro that is not present in GalNAc-Ts encompassing the opposed glycosylation preferences such as GalNAc-T1/T2/T14. This Pro in the flexible linkers of GalNAc-T3 (P_244EVYPDNLNPPPS_266 underlined) and GalNAc-T4 (P_243NLHPEDRPGWH_265 underlined) is located at the end of the loop with respect to the flexible linker of GalNAc-T2 (P_243ELRVPDHDQDIAP_264). Considering that Pro residues are abundant in protein turns, we hypothesized that this additional Pro might be, at least in part, responsible of the different orientation found crystallographically for the GalNAc-T4 lectin domain, and computationally for chimera 2. To explore this hypothesis, we mutated this Pro (in the GalNAc-T3 flexible linker sequence of the GalNAc-T2 chimera 2) to Ala-producing chimera 2_P503A with the expectation that it would reverse its preference (Table 2). This chimera, however, showed an identical glycosylation preference as the starting chimera 2 (Fig. 4b and Supplementary Fig 9), rather than reverting to the expected wild-type GalNAc-T2 preference. This suggests that more complex events within the flexible linker might take place to explain the interplay between the flexible linker dynamics and its coupling with the lectin domain rotation. Also note that this mutated chimera displays a very low-specific activity compared to GalNAc-T2 (~3–10%) (Supplementary Fig. 9). This may suggest that the architecture of the catalytic site of this chimera may have been altered or perhaps blocked by an altered orientation of its lectin domain.

Next, we evaluated in detail the interactions between residues of the flexible linker in the compact and extended crystal structures of GalNAc-T2.14 Two major interactions, namely a salt bridge between Arg438 and Asp444, and a CH–π interaction between Pro440 and Phe447, appeared to be important to maintain the compact, folded structure of its flexible linker (Supplementary Fig. 11). We reasoned that these interactions might be important for fixing the GalNAc-T2 lectin domain orientation with respect to its catalytic domain. To test this hypothesis, we generated the double mutant R438A-D444A and the triple mutant R438A-D444A-F447A (Table 2). Our assumption was that the disruption of these interactions might lead to more “flexible” linkers, which in turn could cause some degree of rotation of the lectin domain toward the orientation observed in GalNAc-T4. MD simulations showed that, whereas the lectin domain did not rotate in the appropriate direction in the double mutant R438A-D444A (Supplementary Movie 8), it did undergo complete rotation in the triple mutant R438A-D444A-F447A toward the orientation observed in chimera 2 and GalNAc-T4, although such motion required a much longer simulation time (complete rotation observed after 160–450 ns, Supplementary Movie 9).

Fig. 4 Characterization of the GalNAc-T2 chimeras and mutants. a MD simulations in explicit water of chimera 2 (500 ns total simulation time; water molecules were removed for clarity). The 28 ns time-lapse snapshots show the dynamic events occurring during the lectin domain reorientation observed for this chimera. The GalNAc-T2 (far left) and GalNAc-T4 (far right) crystal structures are shown as references for the initial and final states. All structures are illustrated in a surface view with two different orientations. The lectin and catalytic domains, and the flexible linker of GalNAc-T2 and GalNAc-T4 are depicted as yellow/purple blue, orange/white gray, and red/deep teal, respectively. The colors used for the chimera 2 and GalNAc-T2 are the same except for the flexible linker of the former, which is pale green. Flexible linkers are also shown at the bottom in a cartoon- and sticks-like view. b Glycosylation time course plots of GalNAc-T2 and the GalNAc-T2 chimeras and mutants against (glyco)peptides 1–3 at substrate concentration of 1.4 mM. The obtained specific activities and selected substrate activity ratios are given in Supplementary Fig. 9. c Complete glycosylation kinetics (initial specific activity versus substrate concentration) for GalNAc-T2, the GalNAc-T2-triple mutant and the GalNAc-T2 chimera 3 against substrate (glyco)peptides 1–3 (top panels) with plots of the obtained Michaelis–Menten kinetic parameters, $K_m$, $V_{\text{max}}$, catalytic efficiency ($V_{\text{max}}/K_m$), and catalytic efficiency ratios (monoglycopeptide 2 over 3) (bottom panels). Kinetic parameter values are summarized in Supplementary Table 2.

ARTICLE
These results were validated by their observed substrate preferences (obtained for both mutants; Fig. 4b and Supplementary Fig. 9) and by a kinetic analysis of the triple mutant R438A-D444A-F447A (Fig. 4c and Supplementary Table 2). Whereas the double mutant gave nearly equal monoglycopeptide preferences as observed for GalNAc-T2 (at high substrate concentration), the triple mutant showed a 1.5–2-fold preference for monoglycopeptide 3 over monoglycopeptide 2, resembling the preference of GalNAc-T4 (Fig. 4b,c, Supplementary Table 2, and Supplementary Fig. 9). As with chimeras 1–3, the change in preferences observed with the GalNAc-T2 triple mutant was due to a decrease in $V_{\text{max}}$ and increase in $K_m$ for monoglycopeptide 2 rather than due to a significant change in the kinetic properties of monoglycopeptide 3 (Fig. 4b,c, Supplementary Table 2, and Supplementary Fig. 9). The specific activity (and kinetic constants, when obtained) of these linker mutants against peptide substrate is nearly identical, indicating that for these transfases, the flexible linker that provides the molecular basis for the long-range glycosylation specificity of this family of enzymes by altering the relative orientations of the catalytic and lectin domains. In addition, our results suggest that specific interactions within the flexible linker are sufficient to modulate this orientation thereby dictating each isoforms long-range glycosylation preferences.

**Discussion**

Most types of protein O-glycosylation in higher eukaryotes (O-Man, O-Xyl, O-Glc, O-GlCNac, and O-Fuc) are initiated by only one or two GT isoenzymes. However, these types of protein O-glycosylation are only found on limited proteins and sequence motifs. On the contrary, Nature has imposed a large family of GalNAc-Ts isoforms to account for mucin-type O-glycosylation. Initially, these GTs were considered redundant isoenzymes but recent findings clearly demonstrate that individual GalNAc-Ts isoforms serve unique biological functions in health and disease. For example, deficiency in GALNT3 and GALNT2 causes familial tumoral calcinosis and dyslipidemia, respectively. How such transferase specificity may be regulated is best illustrated by GalNAc-T3, which is the only isoform capable of glycosylating Thr178 at a proprotein convertase (PC) processing site (RHRTR$^{179}$) of FGF23, an important regulator of phosphate homeostasis. However, glycosylation of Thr178 by GalNAc-T3 requires prior O-glycosylation of residue Thr171 of FGF23, demonstrating that GalNAc-T3 utilizes its long-range N-terminal preference for prior GalNAc glycosylation to target or amplify the glycosylation of Thr178. Thus, the combination of local peptide sequence and remote prior glycosylation is used by these transfases to ensure the fidelity of isoform-specific glycosylation. With the recent quantitative differential O-glycoproteomics studies, it is now clear that many individual GalNAc-T isoforms serve unique non-redundant contributions to glycosylation of the proteome, presumably employing similar strategies as that of GalNAc-T3 and FGF23. The long-range lectin domain-mediated functions of GalNAc-Ts are likely instrumental for the coordinated glycosylation of additional diverse sequence motifs and in particularly high-density acceptor sequence motifs such as in mucins. Therefore, it is likely that the large number of GalNAc-T isoforms and their lectin domains evolved to recognize diverse peptide sequence motifs as well as to differentially recognize prior sites of GalNAc glycosylation. This would ensure the fidelity of glycosylation in both low- and high-density glycosylated regions by sequential orchestrated glycosylation reactions, coordinated by both the specificity of the catalytic domain as well as the long-range recognition of the lectin domain, that interacts with the prior attached GalNAc residue. Therefore, to fully understand the biological functions of these enzymes and to elucidate their roles in disease, we must understand the molecular basis, and mechanisms therein, which dictate and modulate their complex long- and short-range substrate specificity.

Here we have addressed the molecular mechanisms behind the unique long-range lectin domain-dependent glycosylation functions of GalNAc-Ts and have provided a model explaining how these transfases differentially recognize such N- and C-terminal long-range glycosylation preferency. We previously reported that the location of the lectin α-subdomain GalNAc-binding site, relative to the catalytic domain, was key to explaining how GalNAc-T2 delivered N-terminal acceptor sites of C-terminal monoglycopeptides to the catalytic domain in a lectin domain-dependent manner. We also demonstrated that the flexible linker, located ~28–36Å from the active site of GalNAc-T2, and ~18–25Å from GalNAc-T2’s lectin α-subdomain was in charge of the conformational heterogeneity of GalNAc-T2. This was clearly due to the translational capacity of the lectin domain with respect to the catalytic domain, which was further modulated in the presence of substrates. Note also that in all of the GalNAc-T2 structures reported, irrespectively of whether they are compact or extended structures, the lectin domain is positioned such that its α-subdomain GalNAc-binding site is always located on the left side of the lectin domain. However, it has remained a conundrum how other GalNAc-Ts (i.e., GalNAc-T3, T4, T12, etc.) presented the opposite long-range glycosylation preference of GalNAc-T2 and others. Our multidisciplinary approach presented herein has revealed additional conformational properties of the short flexible linker that provide the molecular basis for the distinct long-range glycosylation preferences found for these enzymes. Our crystal structure of GalNAc-T4 revealed that its lectin domain is positioned such that its lectin α-subdomain GalNAc-binding site is now located on the right side of the lectin domain, in agreement with our proposed Model 1 (Figs. 1c and 3). This is fundamental to explaining why GalNAc-T4 and similar behaving GalNAc-Ts prefer to glycosylate C-terminal acceptor sites of N-terminal glycosylated glycopeptides. Our STD-NMR studies together with the transfase kinetic studies suggested that the orientation of the lectin domain-bound GalNAc moiety was fundamentally important for the correct orientation and presentation of the acceptor residues to the catalytic site, thereby leading to more efficient catalysis. Our MD simulations on multiple chimera transfases together with their kinetic analysis, further provide compelling evidence that the flexible linkers not only allow translational motion to the lectin domain, but also rotational motion. Both types of motions likely differ between the GalNAc-T isoforms and therefore will contribute to the different glycosylation preferences found among these enzymes. All together, our studies point to the dissimilar flexible linkers as major contributors to the distinct long-range glycosylation preferences of these transfases (Supplementary Fig. 7). It is further likely that some flexible linkers will have conformational properties in solution that allow two or more conformational states of the lectin domain relative to the catalytic domain thus accounting for the dual long-range preferences observed for GalNAc-T2 (and its chimeras and mutants) as well as the several GalNAc-Ts that display nearly equal N- and C-terminal long-range glycosylation preferences.
such conformational flexibility in our MD trajectories of GalNAc-T2 (and some chimeras and mutants) may suggest that the activation energy for this interconversion may be very high and/or that our simulations were not sufficiently long enough.

In conclusion, we have provided for the first time the molecular basis for the distinctive long-range glycosylation preferences of the GalNAc-Ts, which is based on a very small flexible linker that provides rotational and translational capacity to the lectin domain. This work further exemplifies how a structural feature, very distant from both the active site and the lectin domain GalNAc-binding site, is capable of tuning the activity and specificity of these biologically important multidomain enzymes.

Methods

Cloning and purification of human GalNAc-T4/chimeras. The DNA sequence encoding amino acid residues 36–578 of the human GalNAc-T4, defined as galnact4, was codon optimized and synthesized by GenScript (USA) for expression in Pichia 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 encoding a histidine tag, a stop codon and a recognition sequence for SacII, was cloned into the pUC57 vector (GenScript). Following transfection with XhoI and SacII, the construct was subcloned into the protein expression vector pPICZaA, resulting in the expression plasmid pPICZaA-galnact4. Subsequently, the plasmid was linearized by SacI-HF and transformed into SMD1168. Transformants were selected and colonies were grown into the protein expression vector pPICZα nact4

UDP, 2 mM MnCl2, and 5 mM MUC5AC-3,134 (GTT*PSPVPTTSTT*SAP) in 25

ments at 18 °C by mixing 0.5

Crystallization

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-02006-0 ARTICLE

Surface plasmon resonance experiments. SPR experiments were performed at 25 °C with a Biacore X-100 apparatus (Biacore AB) in 25 mM Tris buffer, 1 mM DTT, 100 mM NaCl, 100 mM MnCl2, pH 7.2. The IgG (100 μg/ml) was immobilized by aminoethylcarbodiimide (EDC). GalNAc-T4 was immobilized on a flow cell 2 by injection of a 100 μg/ml protein solution diluted with 10 mM sodium acetate buffer with a flow rate of 10 μl/min for 7 min followed by an injection of 130 μl etha-

nolamine to block any remaining activated groups on the surface. The level of immobilization reached was about 8000 RU. Flow cell 1, used as reference, was blocked with ethanolamine at the same conditions of flow cell 2 without immo-

bilation of protein. Affinity experiments were made using a series of different concentrations of peptide 3 in the range of 0.01–5 μM with a flow rate of 30 μl/min during 100 s. Each injection was followed by a 100 s injection of running buffer (dissociation phase). No regeneration steps were performed between injections. Response data were collected at real time and analyzed with the Biacore X-100 Evaluation software and plotted as response shift versus analyte concentration.

We could not determine a Kd value due to binding saturation was not achieved.

Synthesis of peptides. Peptides were synthesized via stepwise microwave-assisted solid-phase peptide synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). The glycosylated amino acid building blocks (2.0 equiv) were synthesized as described in the literature and manually coupled using HBTU, while the other Fmoc amino acids (5.0 equiv) were auto-

matically coupled using oxyma pure/DIC. The N-acetyl groups of (AcO)3GalNAc were removed in a mixture of N,N-diisopropylethylamine (DIPEA) and 1 M HCl in 1:2 ratio of DMF/Water for 1 h, then were released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS 2.5%, H2O 2.5%, followed by precipitation with cold diethyl ether. Finally, they were purified by HPLC using a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 20 μl/min. Peptides were further subjected to Edman amino acid sequencing on a Shimadzu PPSQ-53A peptide sequencer prior to use.

Peptide 1: HPLC Rt = 11.53 min (Grad: water 0.1% TFA/acetoniotn (95:5) → (85:15), 13 min, λ = 212 nm). HRMS ESI + (m/z) calcd. for C56H93N18O24 [M + 2 H]+ 1401.6541, found 1401.6540.

Peptide 2: HPLC Rt = 11.75 min (Grad: water 0.1% TFA/acetoniotn (95:5) → (85:15:4), 15 min. λ = 212 nm). HRMS ESI + (m/z) calcd. for C56H93N18O24 [M + H]+ 1415.6761, found 1415.6739.

Peptide 3: HPLC Rt = 13.72 min (Grad: water 0.1% TFA/acetoniotn (95:5) → (86:14), 15 min, λ = 212 nm). HRMS ESI + (m/z) calcd. for C56H93N18O24 [M + H]+ 1401.6605, found 1401.6546.

NMR experiments. All NMR experiments were recorded on a Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe head. The 1H NMR resonances of the peptides 1–3 were completely assigned through standard 2D-TOCSY (30 and 80 ms mixing time) and 2D-NOESY experiments (400 ms mixing time). Solution conditions used for the NMR characterization studies were 1–3 (pH 7.5 at 25 °C) and 1:130 mM NaCl, and 1 mM DTT, uncoupled pH 7.4. The assignments were accomplished either at 278 or 298 K. The resonance of 2,2,3,3-tetrafluoroe-trimethylsilylpropionic acid (TSP) was used as a chemical shift reference in the 1H NMR experiments (8 TSP = 0 ppm). Peak lists for the 2D-TOCSY and 2D-NOESY spectra were generated by interactive peak picking using the computer aided resonance assignment (CARA) software.

Samples for STD experiments were prepared in deuterated water, 7.5 mM NaCl, and 1 mM DTT, uncoupled pH 7.4. STD-NMR experiments were performed at 298 K in the presence of 75 μM UDP, 75 μM MUC5AC-3-13 with ~880 μM peptide (or GalNAc-O-Me) and 13.5 mM GalNAc-T4 giving a molar ratio of 65:1 peptide:GalNAc-T4.

The STD-NMR spectra were acquired with 1920 transients in a matrix with 64 k data points in 12 in a spectral window of 12335.53 Hz centered at 2819.65 Hz. An excitation sculpting module with gradients was employed to suppress the water resonances (90°). Selective saturation of the proton resonances (on resonance spectrum) was performed by irradiating at ~1 ppm using a series of Eurbp2.1000-

shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation time of
Molecular docking. Two different docking calculations were conducted to generate the ternary complex. First, UDP was docked into GaINAc-T4 with the aid of AutoDock Vina 1.1.234. The predicted binding energies ranged from −6.4 to −5.2 kcal/mol. The AutoGrid grid point spacing was set at 0.375 Å, center coordinates of the grid box were 5.4, −2.4, −7.2 (x, y, z), and number of grid points in xz plane was 76, 40, 42, respectively. All allowed torsional bonds were considered rotatable. The 3D structure of the docked ligand to the protein with the lowest-binding energy was used for further calculations (see below).

In a second simulation, glycopeptide T(o-O-GalNAc)GAGAGAGTPGPG was docked into GaINAc-T4. The AutoGrid grid point spacing was set at 0.375 Å, center coordinates of the grid box were 0.0, −9.7, 9.4 (x, y, z), and number of grid points in xz plane was 86, 60, 90, respectively. All allowed torsional bonds were considered rotatable, except the bonds involved in the glycosidic linkage, which were fixed to $\phi=64^\circ$, $\psi=110^\circ$. The 3D structure of the docked ligand to the protein with lowest-binding energy and with the GalNAc unit located in the lectin domain region, was then used for further calculations.

The combination of both structures was used to build up the ternary complex, which was subjected to MD simulations (see below).

MD simulations. The starting coordinates for the complex between glycopeptide 3, UDP/Mn$^{2+}$, and GaINAc-T4 were generated combining the X-ray structure of GaINAc-T4 solved in this manuscript and the crystal structure of the activated form of GaINAc-T2 complex with UDP/Mn$^{2+}$ and MUC5AC-13 (PDB entry 5AJP). The GalNAc unit of the glycopeptide 3 was located by superimposition to the one solved in the X-ray structure of GaINAc-T4, and UDP was located by superimposition to UDP in GaINAc-T2 and T4. The peptide backbone of bound 3 was situated so that the Thr residues were laying on the proximity of the UDP moiety. The starting coordinates for chimera 2 were generated by superimposing the X-ray structures of GaINAc-T2 (PDB entry 5AJP) and GaINAc-T4. Residues 435–446 of GaINAc-T2 and GaINAc-T4 were combined and replaced by residues 494–506 (PEVYVYDPLVNP) of GaINAc-T3 using PyMol (http://www.pymol.org). Residues Asn$^{207}$Pro were maintained in GaINAc-T2 to facilitate the linkage between its catalytic and lectin domains. The double (R438A-D444A) and triple mutant (R438A-D444A-F447A) were generated using PyMol. The same protocol was used for other computational chimeras.  

Force field parameters for the substrates were generated with the antechamber module of Amber14 using a combination of GLYCAM05 parameters for the GalNAc unit and the general Amber force field (GAFF) for GDP, with partial charges set to fit the electrostatic potential generated with \( \text{HF}/6-31 \text{G}^\ast \) by RESP. The charges are calculated according to the Merz–Singh–Kollman scheme using Gaussian 09\textsuperscript{37}. Each protein was immersed in a truncated octahedral box with a 10 Å buffer of TIP3P water molecules and neutralized by adding explicit counter ions (Na$^+$, Cl$^-$). All subsequent simulations were performed using the hEsb force field\textsuperscript{37}. 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 was used to control and equilibrate 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. An 8-Å cutoff was applied to Lennard–Jones and electrostatic interactions. Each system was equilibrated for 2 ns with a 2-Å stepwise increase at a constant volume and temperature of 300 K. Production trajectories were then run for additional 100–500 ns under the same simulation conditions.

Transferrase assays and kinetics. Specific activity determinations: GaINAc-T glycosylation reactions against (glyco)peptides 1–3 were performed with 73 mM UDP-GalNAc (determined by OD$_{280}$), 1 mM radiolabeled UDP-[3H]-GalNAc (~6 × 10$^8$ DPM/mole, American Radiolabeled Chemicals Inc.), and 1.4 mM (−1.25 mg/mL) of (glyco)peptide (from 4 mM stock in 1 mM TRIS, pH 7.0) and varying concentrations (0.01 to 0.5 µM) of transferrase (determined by OD$_{234}$, giving a final reaction volumes of 20 µL in 100 µL capped Eppendorf tubes. Reactions were incubated at 37°C in a thermostated microplate shaker (Taitec Microincubator M-36) and quenched at the appropriate reaction time by the addition of 20 µL of 250 mM EDTA and placed on ice prior to final workup. Typically, for each transferrase, glycosylation reactions were performed at time points with all three (glyco)peptide substrates (including a no peptide control) using the same batch of transferrase and the same UDP-[3H]-GalNAc stock. Before performing time course experiments, the relative activity of each transferrase was determined by trial and error against all three peptides to determine the optimal transferase concentration or incubation time that gave no more than 10% peptide glycosylation against the best substrate. Final time course experiments typically consisted of incubation times of 10, 20, and 30, 60, and 90 or 120 min (depending on activity) and were performed at least twice. Most transfers were also characterized at different enzyme concentrations (0.01–0.03 and 0.2–0.5 µM) and the results combined. After quenching the reaction and diluting to 4 mL, see UDP and non-hydrolyzed UDP-GalNAc were removed by passage over a Dowex 1 × 8 anion 350, 700W resin (~3 mL column). Total UDP-[3H]-GalNAc utilization (transfer to peptide substrate and transfer to water, i.e. hydrolysis) was determined by difference after scintillation counting (Beckman LS5801 scintillation counter) 1/20 of the sample before and after passage over Dowex 1 × 8. The extent of [3H]-GaINAc transfer to peptide and the extent of [3H]-UDP-GalNAc hydrolyzed was determined by gel filtration for the longest reaction time point, as previously described\textsuperscript{35}. Example gel filtration chromatograms are given in Supplementary Fig. 12. Significant hydrolysis was only observed for GaINAc-T4, which was corrected for in determining its specific activity. Specific activity was obtained by calculating at each individual data time point (of the combined time course plots, Fig. 4b), an individual-specific activity, which was then averaged to obtain the overall average specific activity (and standard deviation). Specific activity is reported as nmol of glycopeptide product formed per min, pmol of enzyme. Substrate activity ratios (i.e., substrate ratios 2/1, 3/1, and 3/2) were also obtained at each individual time point and averaged (with standard deviation). These data are tabulated in Supplementary Fig. 9.

Note that the Microsoft Excel least squares fitting function LINEST was used to obtain the linear plots shown in Fig. 4b while fixing the y intercept to zero. Note that the specific activity calculated from the least square slope differed slightly (less than 10%) from that determined from the average of the individual time points.

Detailed enzyme kinetics on GaINAc-T4, GaINAc-T2, the GaINAc-T2 triple mutant and GaINAc-T2 chimera 3: Glycosylation reactions were performed as described above using a fixed enzyme concentration (0.06 µM for GaINAc-T4, and 0.02 µM for GaINAc-T2 and its linker constructs) with varying peptide concentration (typically 46, 93, 186, 373, 750, and 2000 µM). Incubation times (10–90 min) were chosen so that no more than 30% of the UDP-GaINAc donor was depleted while typically giving less than 10% peptide glycosylation, after correction for UDP-GaINAc hydrolysis described above. Specific activities were obtained from one to two individual incubation time points performed in two–three separate experiments. These individual-specific activity values were used to calculate the kinetic constants of $K_v$ and $V_{max}$ using nonlinear Michaelis–Menten fitting program in GraphPad Prism 7.03.

Determination of site of glycosylation. The determination of substrate glycosylation sites was performed by Edman amino acid sequencing (Applied Biosystems Procise 494 peptide sequencer) of the G10 isolated [3H]-GaINAc glycosylated peptides as previously described\textsuperscript{9,10}, where each cycle was collected on a fraction collector and scintillation counted (Beckman LS5801 scintillation counter) for [3H]-GaINAc content. Although there is commonly sample-to-sample variability in the [3H]-GaINAc content loaded on the sequencer (due sample losses and different instrument specificities), the observed sites of incorporation were found to be identical between different experiments with the same transferrase and substrate. The presence of [3H]-GaINAc lag after a peak of [3H]-GalNAc incorporation is commonly observed in these determinations, which is due to the poor extraction from the sample filter of the glycosylated-PTH residues compared with the standard amino acid PTH derivatives\textsuperscript{9, 10}.

Data availability. The coordinate and structure factor for GaINAc-T4 in complex with monoglycopeptide 3 has been deposited in the Worldwide Protein Data Bank (wwPDB) with the accession code SNQA. All relevant data are available from the authors on reasonable request.

Received: 8 April 2017 Accepted: 1 November 2017
Published online: 05 December 2017
References

1. Bennett, E. P. et al. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22, 736–756 (2012).
2. Gill, D. J., Clausen, H. & Bard, F. Location, location, location: new insights into O-GalNAc protein glycosylation. Trends Cell Biol. 21, 149–158 (2011).
3. Hurtado-Guerrero, R. Recent structural and mechanistic insights into protein O-GalNAc transferase activity. Biochim. Biophys. Acta 1820, 2079–2094 (2012).
4. Lira-Navarrete, E. et al. Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. Nat. Commun. 6, 6937 (2015).
5. Tabak, L. A. The role of mucin-type O-glycans in eukaryotic development. Semin. Cell Dev. Biol. 21, 616–621 (2010).
6. Schjoldager, K., Clausen, H. & Tabak, L. A. Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. Biochim. Biophys. Acta 1820, 2079–2094 (2012).
7. Schjoldager, K. T. et al. Deconstruction of O-glycosylation-GalNAc-T isoforms direct distinct downstream of the O-glycoproteome. EMBO Rep. 16, 1713–1722 (2015).
8. Hanson, R. L. & Hollingsworth, M. A. Functional consequences of differential O-glycosylation of MUC1, MUC4, and MUC16 (downstream effects on signalizing). Biomolecules 6, E34 (2016).
9. 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).
10. Fritz, T. A., Hurley, J. H., Trinh, L. B., Shioloa, J. & Tabak, L. A. The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. Proc. Natl Acad. Sci. USA 101, 15307–15312 (2004).
11. Fritz, T. A., Ramaz, J. & Tabak, L. A. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. J. Biol. Chem. 281, 8613–8619 (2006).
12. Lira-Navarrete, E. et al. Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. Angew. Chem. Int. Ed. Engl. 53, 8206–8210 (2014).
13. Kubota, T. et al. Structural basis of carbohydrate transfer activity by human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase (ppGalNAc-T10). J. Mol. Biol. 359, 708–727 (2006).
14. Wandall, H. H. et al. The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. Glycobiology 17, 374–387 (2007).
15. Gerken, T. A. et al. The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc-Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-glycosylation. J. Biol. Chem. 288, 19990–19994 (2013).
16. Revoredo, L. et al. Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. Glycobiology 26, 360–376 (2016).
17. Raman, J. et al. The catalytic and lectin domains of UDP-GalNAc-polypeptide alpha-N-Acetylgalactosaminyltransferase function in concert to direct glycosylation site selection. J. Biol. Chem. 283, 22942–22951 (2008).
18. Hassan, H. et al. The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. J. Biol. Chem. 275, 38197–38205 (2000).
19. Posey, A. D. Jr et al. Engineered CAR T cells targeting the cancer-associated Toxofrorn of the membrane mucin MUC1 control adenocarcinoma. Immunity 44, 1444–1454 (2016).
20. McEver, R. P. & Cummings, R. D. Role of PSGL-1 binding to selectins in leukocyte recruitment. J. Clin. Invest. 100, 597–5103 (1997).
21. Bennett, E. P. et al. Cloning of a human UDP-N-acetyl-alpha-D-galactosaminylpolypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. J. Biol. Chem. 273, 30472–30481 (1998).
22. Kato, K. et al. Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. J. Biol. Chem. 281, 18570–18577 (2006).
23. Pedersen, J. W. et al. Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. J. Biol. Chem. 286, 32684–32696 (2011).
24. Topaz, O. et al. Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. Nat. Genet. 36, 579–581 (2004).
25. Khetarpal, S. A. et al. Loss of function of GALNT2 lowers high-density lipoproteins in humans, nonhuman primates, and rodents. Cell Metab. 24, 234–245 (2016).