Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Polypeptide GalNAc-Ts: from redundancy to specificity

Matilde de las Rivas[1,5], Erandi Lira-Navarrete[2,5], Thomas A Gerken[3] and Ramon Hurtado-Guerrero[1,4]

Mucin-type *O*-glycosylation is a post-translational modification (PTM) that is predicted to occur in more than the 80% of the proteins that pass through the Golgi apparatus. This PTM is initiated by a family of polypeptide GalNAc-transferases (GalNAc-Ts) that modify Ser and Thr residues of proteins through the addition of a GalNAc moiety. These enzymes are type II membrane proteins that consist of a Golgi luminal catalytic domain connected by a flexible linker to a ricin type lectin domain. Together, both domains account for the different glycosylation preferences observed among isoenzymes. Although it is well accepted that most of the family members share some degree of redundancy toward their protein and glycoprotein substrates, it has been recently found that several GalNAc-Ts also possess activity toward specific targets. Despite the high similarity between isoenzymes, structural differences have recently been reported that are key to understanding the molecular basis of both their redundancy and specificity. The present review focuses on the molecular aspects of the protein substrate recognition and the different glycosylation preferences of these enzymes, which in turn will serve as a roadmap to the rational design of specific modulators of mucin-type *O*-glycosylation.

Addresses
[1] BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus, Rio Ebro, Edificio I+D, Zaragoza, Spain
[2] Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen, Denmark
[3] Departments of Biochemistry, Chemistry and Pediatrics Case Western Reserve University, Cleveland, OH, USA
[4] Fundación ARAID, 50018, Zaragoza, Spain

Corresponding authors: Gerken, Thomas A. (txg2@cwru.edu), Hurtado-Guerrero, Ramon (rhurtado@bifi.es)
[5] These authors have contributed equally.

Current Opinion in Structural Biology 2019, **56**:xx–yy
This review comes from a themed issue on **Carbohydrates**
Edited by **Anna Panchenko** and **Monika Fuxreiter**

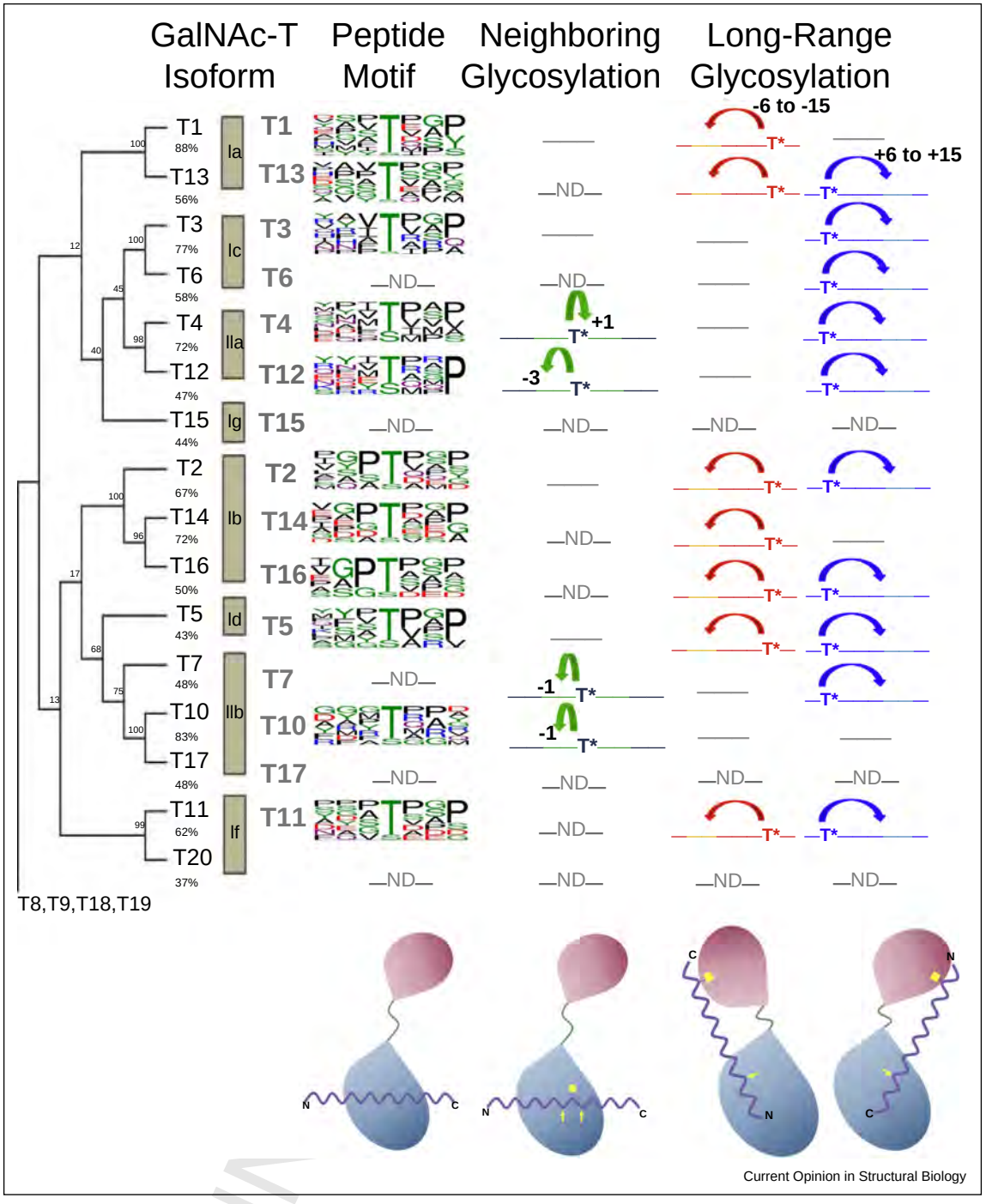
https://doi.org/10.1016/j.sbi.2018.12.007

Introduction

Polypeptide GalNAc-transferases (GalNAc-Ts) are a family of Golgi resident enzymes (20 in humans) that transfer a GalNAc moiety from UDP-GalNAc onto Ser or Thr residues of their protein substrates. This process results in the synthesis of the Tn antigen (GalNAcα1-*O*-Ser/Thr), which can be further elongated by the action of subsequent glycosyltransferases (GTs) [1•,2]. Historically, this modification is known as mucin-type *O*-glycosylation (henceforth *O*-glycosylation) as these glycans are abundant (>50% by weight) in mucins. As the GalNAc-Ts initiate and thus define sites of *O*-glycosylation in densely *O*-glycosylated proteins such as mucins, these enzymes must possess a range of properties in order to properly glycosylate their targets [2,3]. It is now known that the GalNAc-T isoenzymes have different glycosylation preferences that allow them to be classified as: a) glycopeptide/peptide-preferring isoforms (e.g. GalNAc-T1 and GalNAc-T2); b) (glyco) peptide-preferring isoenzymes (e.g. GalNAc-T4); and c) strict glycopeptide-preferring isoenzymes (e.g. GalNAc-T7 and GalNAc-T10) [4••]. This distinction is based on their activity against substrates lacking or containing one or more prior GalNAc-*O*-Ser/Thr moieties which allows them to be classified into early, intermediate or late GTs, representing a range in activities against naked peptides/proteins to already highly glycosylated peptides/proteins (e.g. GalNAc-T2, GalNAc-T4 and GalNAc-T10 are early, intermediate and late GTs, respectively) [1•,4••,5] (See Figure 1).

The glycopeptide activities of the GalNAc-Ts have been further classified into two classes, based on their short-range (or neighboring) and long-range (or remote) glycosylating capabilities. The short-range glycosylation preferences account for the glycosylation of glycopeptide substrates where the sugar moiety is bound to the catalytic domain (thus glycosylating 1–3 residues from the sugar), while the long-range glycosylation preferences comprise glycopeptide sugar binding to the lectin domain which subsequently directs distant acceptor sites (6 to ~ 17 residues away) onto the catalytic domain for glycosylation as depicted in Figure 1 [1•,6••]. It has been found that both the long-range and short-range glycopeptide preferences can operate in an N-terminal or C-terminal direction depending on isoenzyme and furthermore some isoenzymes possess both the long-range and short-range glycopeptide activities (Figure 1). Together, these properties explain how a highly coordinated repertoire of GalNAc-Ts are capable of readily generating multiple closely spaced Tn antigens, as occurs in mucins, as well as glycosylating proteins containing only a few acceptor

Figure 1



Summary of the known peptide and glycopeptide specificities of the GalNAc-T family.
Phylogenetic tree of the GalNAc-Ts showing their 1) random peptide derived peptide substrate motifs as Sequence Logos [7••,23••,32••,46] 2) neighboring prior glycosylation preferences (1–3 residues) due to catalytic domain interactions [4••], and 3) their long-range prior glycosylation preferences (∼6 to ∼17 residues) due to lectin domain binding [23••,40••]. Note that '-ND-' stands for not determined while '---' indicates no or weak activity. Also note that GalNAc-T8, GalNAc-T9, GalNAc-T18 and GalNAc-T19 exhibit nearly undetectable activities against most substrates [47] and have not been well characterized. The models at the bottom show the different substrate binding modes that lead to the indicated specificities. The catalytic and lectin domains are shown as oval-shaped figures in blue and red, respectively. (Glyco)peptides are indicated in purple while the yellow squares denote the position of prior GalNAc moieties in the glycopeptides. Arrows indicate the position of GalNAc transfer to the acceptor site.

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

sites. These combined properties must also be involved in the targeting of specific substrates.

The fact that most isoenzymes of this family are capable of glycosylating common acceptor substrates, particularly those containing the (Thr/Ser)ProXPro motif (where 'X' usually stands for a small hydrophobic residue; see Figure 1), suggests that they may also serve redundant functions [4••,7••]. Paradoxically, in recent years an incremental number of reports have demonstrated that several GalNAc-T isoenzymes are highly specific for certain protein substrates, as identified by the Simple Cell (SC) approach developed by the Clausen group [8••,9]. Using this strategy, ApoC-III was identified as a specific substrate of GalNAc-T2 [9] and GalNAc-T11 was reported to be specifically involved in glycosylation of the peptide linkers between class A repeats of the LDL receptor family [10,11]. One common theme found for isoenzyme-specific glycosylation is that such a glycosylation can interfere with the proprotein processing of neighboring sites thus controlling such a processing [12]. Two of the most studied examples are GalNAc-T3 and its substrate, the fibroblast growth factor 23 (FGF23) [13], as well as GalNAc-T2 and angiopoietin-like Protein 3 (ANGPTL3) [14], where miss-regulation of *O*-glycosylation can lead to familial tumoral calcinosis and dyslipidemia, respectively [13,14]. In addition, the aberrant expression or mutation of several GalNAc-T isoenzymes and the overexpression of the Tn-antigen is directly associated with many cancers [15,16•,17], the mechanisms of which are still not understood, although Tn–Tn self-association may play a role [16•]. Finally, the GalNAc-Ts are involved or implicated in many other biological functions including development, receptor trafficking and modulation and protein secretion, which are beyond the scope of this review [18].

How this family of GTs can glycosylate multiple common sites in some proteins and at the same time be highly isoenzyme-specific for sites in other proteins remains unanswered. Herein, we review the recent advances that have begun to unravel the substrate-recognition mechanism of several of the most representative isoenzymes, as well as presenting the structural and kinetic basis for both their overlapping and selective activities.

Structural similarities between GalNAc-Ts isoenzymes

To obtain a complete understanding of the mechanism underlying the substrate specificity of this family of enzymes, crystal structures of GalNAc-T isoenzymes have been solved, either in the apo form or in complex with (glyco)peptide substrates and products. To date, the following structures have been reported (see Figure 2a and b): a) *Mus musculus* GalNAc-T1 with Mn^{2+} (*Mm*GalNAc-T1) (PDB entry 1XHB) [19•]; b) human GalNAc-T2 (*Hs*GalNAc-T2) complexed with UDP-Mn^{2+} (PDB entry 2FFV), a 'naked' peptide (PDB entry 2FFU) [20•] and three glycopeptides (PDB entries 5AJP, 5AJO and 5AJN) [21••]; c) *Hs*GalNAc-T10 complexed with UDP-Mn^{2+} and Ser-GalNAc (PDB entries 2D7I and 2D7R) [22•]; d) *Hs*GalNAc-T4 complexed with UDP, Mn^{2+} and a glycopeptide (PDB entry 5NQA) [6••]; and e) the recently reported structures of *Hs*GalNAc-T4 complexed with UDP, Mn^{2+} and a diglycopeptide (PDB entry 6H0B) [23••] and two splice variants of the fly PGANT9A/B with UDP-Mn^{2+} and a peptide substrate (PDB entries 6E4Q and 6E4R; Figure 2b) [24••].

These structures all show an N-terminal catalytic domain adopting the typical GT-A fold, characterized by two abutting Rossmann-like folds which is linked by a short flexible linker to a C-terminal ricin-like lectin domain [6••,19•,20•,21••,22•] (Figure 2a). The lectin domain, a unique structural feature only present in this family of eukaryotic GTs, has a β-trefoil fold built from three repeat units (α, β and γ) that are potentially capable to bind a GalNAc moiety [25,26•]. It should be noted that these repeats are not necessarily all active binders based on their sequence motif [4••] and by experiment [7••,27].

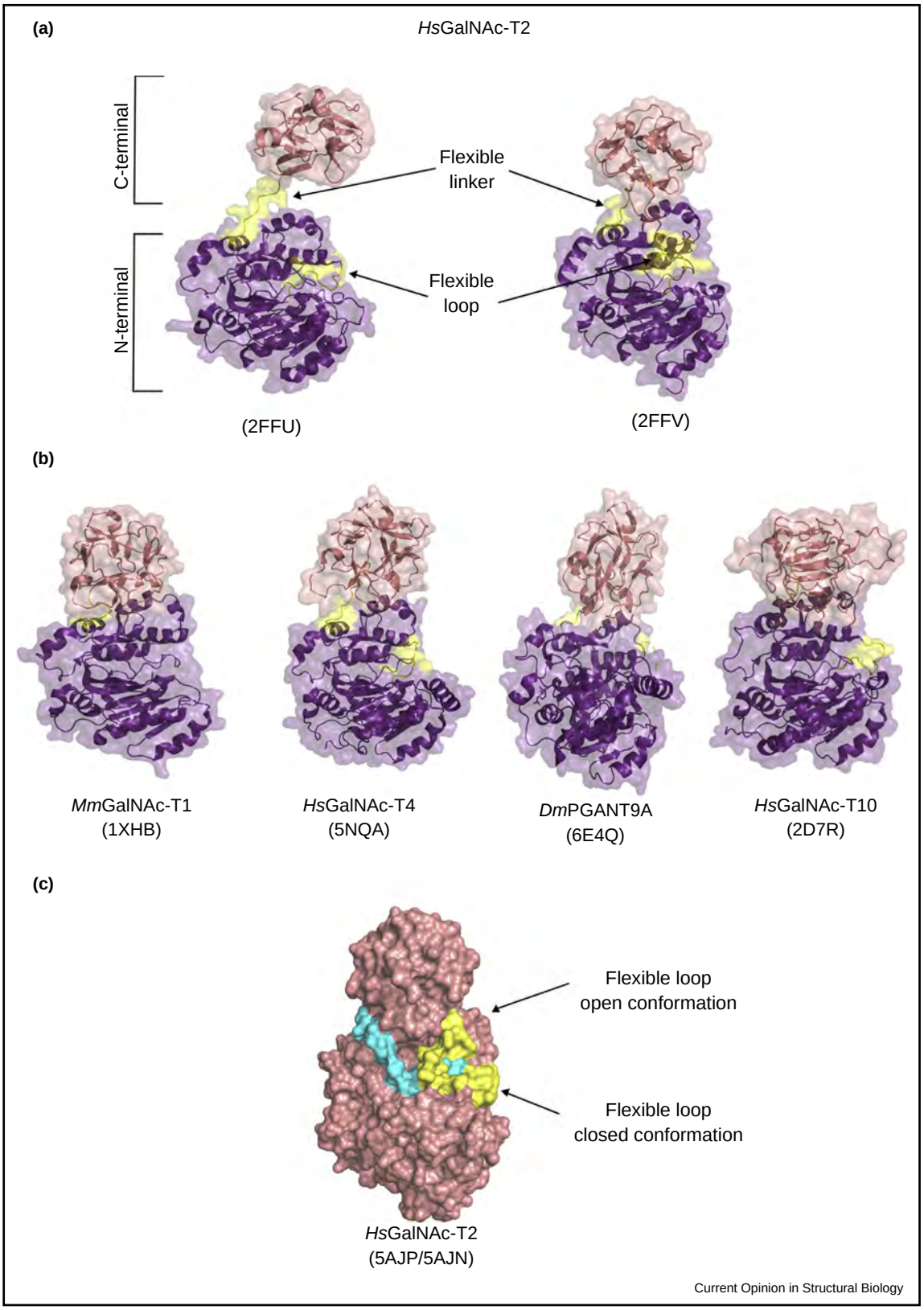
GalNAc-T catalytic domain: UDP-GalNAc binding site and flexible loop

The first crystal structure of a GalNAc-T was for *Mm*GalNAc-T1 [19•]. It provided the initial picture of the active site for which subsequent GalNAc-T structures highly resemble, particularly in the architecture of the critical and conserved Mn^{+2} binding site (formed by Asp209, His211 (the DXH motif) and His344). This structure also showed that the catalytic and lectin domains were closely associated [19•] (Figure 2b). Subsequent crystal structures of GalNAc-T2 complexed to both UDP-Mn^{2+} and to UDP-Mn^{2+} and the EA2 peptide (AspSerThrThrProAlaProThrThrLys) [20•], together with the GalNAc-T10 crystal structure complexed with hydrolyzed UDP-GalNAc and Mn^{2+} [22•], further defined the GT-A fold active site residues that tethered the uridine diphosphate of the UDP-GalNAc donor substrate [20•,22•]. In addition, the structure of GalNAc-T10 catalytic domain revealed the GalNAc moiety bound in the UDP-GalNAc-binding pocket [22•,28]. These structures, as well as the crystal structures of the first pre-Michaelis and Michaelis complexes of *Hs*GalNAc-T2 [29••], were of fundamental importance for defining models of the dynamics of GalNAc-T2 during its catalytic cycle, which consists of an ordered bi–bi kinetic mechanism [29••,30,31]. In addition, the Michaelis complex revealed that these GTs follow a front-face S_Ni-type reaction mechanism [29••].

Structures of GalNAc-Ts, both with and without bound peptide substrate, have revealed a dynamic flexible loop at the surface of the catalytic domain substrate binding site as an important structural feature of the GalNAc-Ts [20•,21••,29••]. This flexible loop, formed by residues

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

Figure 2



Cartoon and surface representation of GalNAc-Ts structures.
Catalytic and lectin domains are shown in purple and salmon respectively, while flexible loop and linker are displayed in yellow. **(a)** Extended (left

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

Arg362 to Ser373 in *Hs*GalNAc-T2, is able to adopt either a closed conformation to form a lid over UPD, rendering the enzyme in an active form, or an open conformation in which the loop folds back to expose UDP to the bulk solvent (inactive form; Figure 2c) [20•,29••]. This interconversion has recently been shown to be dependent of the presence of UDP-GalNAc in *Hs*GalNAc-T2, where this sugar nucleotide stabilizes the closed conformation and consequently allows the binding of the substrate peptide [32••]. The importance of this flexible loop in catalysis is exemplified by the molecular basis of the inactivation of the *Hs*GalNAc-T2-Phe104Ser mutant, which is linked to low levels of high density lipoprotein cholesterol [32••,33]. It was found that Phe104 controls the inactive-to-active transition of the flexible loop due to its hydrophobic interaction with Ala151/Ile256/Val360, as well as a CH-π interaction with the side-chain of Arg362 located in the flexible loop [32••]. The hydrophilic Phe104Ser mutation fails to lock the flexible loop in its active form, thus impeding peptide substrate binding and the failure to glycosylate its targeted peptide substrates (i.e. ApoC-III and ANGPTL3). This results in low levels of HDL [32••,33,34].

GalNAc-T catalytic domain: peptide-binding site

It is noteworthy that several GalNAc-T crystals soaked or cocrystallized with (glyco)peptides show indeterminate/disordered structures for the substrate bound to the catalytic domain [6••,24••]. Nevertheless, structural information of the GalNAc-T-peptide acceptor recognition could be inferred from the series of structures of (glyco)peptides bound to the *Hs*GalNAc-T2 isoenzyme [21••] and very recently the diglycopeptide bound to *Hs*GalNAc-T4 [23••]. In these latter structures the interactions between the transferase and its acceptor substrates are dissimilar, suggesting differences between isoenzymes at the peptide binding groove level. However, this could also be due to the different (glyco)peptides used for both isoenzymes. The *Hs*GalNAc-T2 structures revealed that the EA2 peptide bound in a shallow cleft on the surface of *Hs*GalNAc-T2, being recognized by hydrophobic interactions and to a lesser extent hydrogen bond interactions [20•] (see Figure 3a). It was also observed that the methyl group of the acceptor Thr residue was embedded within a hydrophobic pocket, providing a plausible explanation of why most GalNAc-T isoenzymes prefer to glycosylate Thr over Ser acceptor residues [20•,21••,35] (Figure 3a). Several other crystal structures of *Hs*GalNAc-T2 in complex with UDP-Mn^{2+} and glycopeptides also showed that the glycopeptides acted as bridges between the catalytic and lectin domains, where the latter bound the glycopeptide GalNAc [21••]. In these structures, UDP and the glycopeptides were bound to an adaptable sugar-nucleotide binding site, with the flexible loop adopting either open or closed conformations (Figure 2c). Interestingly, the binding of a mono-glycopeptide to GalNAc-T4 revealed peptide GalNAc binding at the lectin domain but no observable peptide electron density in its catalytic domain [6••] while recently a homologous diglycopeptide showed a well-resolved peptide bound to the catalytic domain in a closed conformation due to GalNAc-T4's neighboring glycopeptide binding activity (discussed below) [23••] (Figure 3b). Interestingly, in the GalNAc-T4 structure, the portion of the peptide spanning the catalytic and lectin domains was found disordered [21••,23••].

At the level of the peptide-binding groove, it was further observed that three highly conserved aromatic residues (namely Phe361, Phe280 and Trp282 in GalNAc-T2), interact with the (Thr/Ser)-Pro-X-Pro substrate sequence [20•]. Thus far the (Thr/Ser)ProXPro sequence is the only substrate consensus motif remotely conserved among most GalNAc-Ts (Figures 1, 3a and b) [1•,21••]. Indeed, all isoenzymes that experimentally display this (Thr/Ser) ProXPro preference possess the homologus Phe and Trp residues [4••,7••,23••] including GalNAc-T4 and GalNAc-T12. GalNAc-T7 and GalNAc-T10, which lack these conserved residues and do not exhibit the (Thr/Ser)ProXPro preference, instead display strong neighbouring glycosylation preferences at the +1 position relative to the acceptor (i.e. (Thr/Ser)(Thr*/Ser*), where * = -*O*-GalNAc) [4••,28]. These latter two isoenzymes are, therefore, expected to contain a GalNAc binding site in place of the ProXPro binding site found in the other isoenzymes. Presently, the structural and molecular bases for the neighboring glycosylation preferences of GalNAc-T7 and GalNAc-T10 remain to be determined. Thus, the near lack of a conserved substrate consensus motif together with their active site flexibility points to the versatility of these enzymes, allowing them to sculpt their binding sites to accommodate a wide range of acceptor substrates.

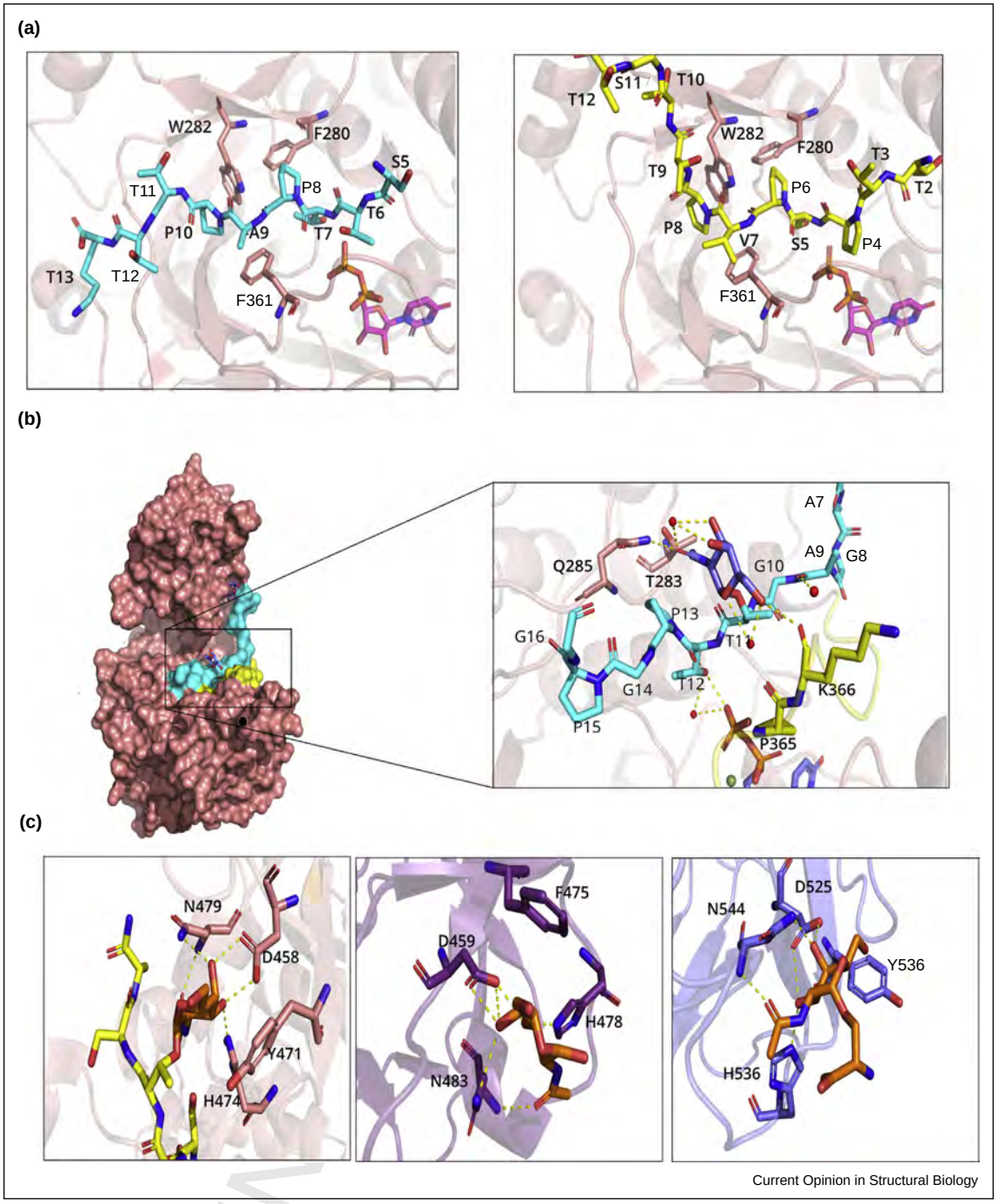
GalNAc-T catalytic domain: glycopeptide-binding site

Until very recently there were no structures describing how the so-called neighbouring glycosylation activity of the GalNAc-Ts could be accommodated. The recent report of a diglycopeptide (GlyAlaThr*3GlyAlaGlyAlaGlyAlaGlyThr*11Thr12ProGlyProGly, where Thr* = Thr-*O*-

(Figure 2 Legend Continued) panel) and compact (right panel) forms of monomeric *Hs*GalNAc-T2. **(b)** Cartoon and surface representation of *Mm*GalNAc-T1, *Hs*GalNAc-T4, *Dm*PGANT9A and *Hs*GalNAc-T10. **(c)** Surface representation of the *Hs*GalNAc-T2-UDP-MUC5AC-13 complex. The overall structure is shown in salmon, monoglycopeptide MUC5AC-13 and the flexible loop are depicted in cyan and yellow, respectively. The flexible loop of the enzyme is shown in its closed and open conformations.

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

Figure 3



Interactions between GalNAc-Ts and their substrates.
(a) Close-up view of the catalytic domain of *Hs*GalNAc-T2 with two different peptides, EA2 (cyan sticks; left panel) and glycopeptide MUC5AC-13 (GlyThrThrProSerProValProThrThrSerThrThr*SerAlaPro) (yellow sticks; right panel), which are similarly recognized by *Hs*GalNAc-T2 through a hydrophobic patch. UDP is depicted as sticks with magenta carbon atoms and Mn^{2+} is shown as a purple sphere. (**b**) On the left panel, surface representation of the *Hs*GalNAc-T4-UDP-Diglycopeptide 6 (GlyAlaThr*3GlyAlaGlyAlaGlyAlaGlyThr*11Thr12ProGlyProGly) complex. Peptide backbone is depicted in cyan with the two GalNAc groups as blue and red sticks; the enzyme flexible loop is shown in its closed conformation in yellow. On the right panel, close-up view of the main interactions between *Hs*GalNAc-T4 catalytic domain glycopeptide binding-site and the GalNAc group on T^*_{11} of diglycopeptide 6. The GalNAc-T4 residues forming the peptide-binding site are depicted in salmon and yellow and the glycopeptide is depicted in cyan, with the GalNAc groups shown as blue and red sticks. Mn^{2+} and water molecules are depicted as green and red spheres, respectively, and hydrogen bonds appears as dotted yellow lines. Please note that we only show water-mediated interactions in which

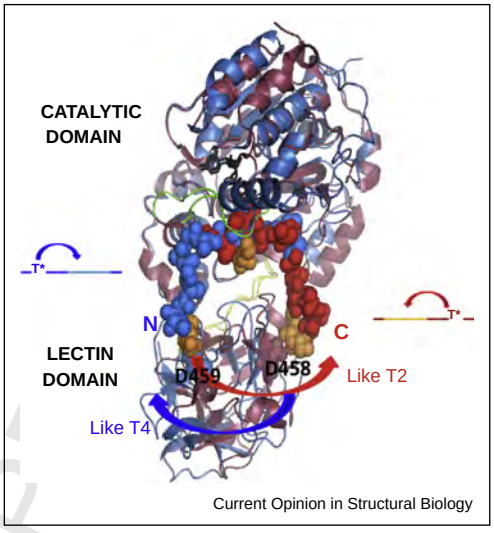
Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

GalNAc) bound to both the lectin and catalytic domains of *Hs*GalNAc-T4 now reveals how this occurs at least in one isoenzyme [23••] (Figure 3b). In this structure the GalNAc of Thr*$_{11}$ is shown tethered by hydrogen bond and hydrophobic interactions to the side chains of Thr283 and Gln285 and the back bone of Lys366 at the surface of the catalytic domain. Importantly, these residues are not conserved among other GalNAc-T isoenzymes and there is no discernible cleft or pocket for the binding of the GalNAc (Figure 3b). Such GalNAc binding presents the adjacent Thr$_{12}$ into the correct orientation to accept GalNAc from the UDP-GalNAc donor. Kinetic studies on a series of glycopeptide substrates further confirmed that the neighbouring GalNAc binding at the catalytic domain was weaker than the remote GalNAc binding of Thr*$_3$ to the lectin domain and further revealed substrate inhibition kinetics on the diglycopeptide, presumably due to competitive binding of the two Thr*'s of the substrate at the lectin domain [23••]. This work is of additional significance as the individual GalNAc-T4 remote and neighboring glycopeptide activities, and both together, could be eliminated or greatly reduced by selective mutagenesis.

GalNAc-Ts lectin domain

The GalNAc-T-glycopeptide recognition at the lectin domain is more easily compared among isoenzymes, as there are crystal structures of *Hs*GalNAc-T2, *Hs*GalNAc-T4 and *Hs*GalNAc-T10 complexed with Ser-*O*-GalNAc as well as longer Thr-*O*-GalNAc containing glycopeptides [6••,21••,22•,23••] (Figure 3c). The GalNAc-T1 lectin domain contains two known functional GalNAc-binding sites out of the possible three (i.e. the α and β subdomains) [36], whereas the GalNAc-T2, GalNAc-T4, and GalNAc-T10 lectin domains contain only one known active site (i.e. α-, α- and β-respectively) [4••]. The first structure of a glycopeptide bound to the lectin domain, that is, Ser-*O*-GalNAc bound to *Hs*GalNAc-T10, revealed the sugar moiety bound to the β-site interacting through several hydrogen bonds (including residues Asp525, Asn544, Tyr536) and one CH-π interaction (His539) (Figure 3c). Subsequent structures of GalNAc-T2 complexed with longer glycopeptides showed no discernible interactions with the peptide backbone of the lectin domain level, while the GalNAc moiety interacted exclusively with residues in the α-subdomain binding site by similar interactions as described for GalNAc-T10 (i.e. via residues Asp458/Asn479/Tyr471 and His474). These residues are conserved in nearly all isoenzymes (Figure 3c) [21••,22•]. Similarly, binding interactions of the peptide GalNAc residue to the lectin α-domain of GalNAc-T4 were recently reported [6••,23••] (Figure 3c); however, a large difference in the orientation of the lectin domains of GalNAc-T4 and GalNAc-T2 relative their catalytic domains was observed. As discussed in the sections below, these differences readily explain the origins of their different long range N-prior or C-prior glycosylation preferences (see Figures 1 and 4).

Figure 4



Superposition of *Hs*GalNAc-T2 and *Hs*GalNAc-T4. Superimposed cartoon representations of *Hs*GalNAc-T2-UDP-MUC5AC-13 glycopeptide (GlyThrThrProSerProValProThrThrSerThrThr*SerAlaPro) complex depicted in red and *Hs*GalNAc-T4-UDP-diglycopeptide 6 (GlyAlaThr*3GlyAlaGlyAlaGlyAlaGlyThr*11Thr12ProGlyProGly) complex depicted in blue. The MUC5AC-13, diglycopeptide 6 and GalNAc moieties are shown in red, blue, and orange atoms, respectively. The arrows indicate the direction of the long-range glycosylation preference of each enzyme, based on the orientation of their respective lectin domains with respect their catalytic domains. Note that the critical Asp residues of the lectin domain GalNAc-binding sites are indicated for clarification purposes.

Earlier work had suggested that the lectin domain of GalNAc-Ts could likely influence substrate specificity by steric hindrance that would depend on the size of the amino acid side chains of the glycopeptide substrate [37], while it has also been suggested that the lectin domains of some GalNAc-Ts could form hetero-dimers and/or homo-dimers that could also alter their specificity [38]. The recent crystal structures of the fly PGANT9-A and PGANT9-B lectin domain splice variants now offers intriguing evidence for something like the former [24••]. In this case, a loop on the lectin domain that protrudes toward the catalytic domain peptide binding site differs in charge between the splice variants. These charge

(Figure 3 Legend Continued) only the water molecule act as a bridge between the residues. **(c)** Main interactions between GalNAc-T2, GalNAc-T4 and GalNAc-T10 isoenzymes lectin domain (shown as salmon, purple and slate, respectively) and the GalNAc moiety (shown as sticks with orange carbon atoms).

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

differences correlate with their activities toward highly charged substrates, thus suggesting that at least electrostatic interactions, if not direct peptide substrate binding, of the lectin domain can significantly influence transferase activity [24••]. Biologically, these splice variances are used to properly glycosylate different secretory mucins, whose incomplete glycosylation is shown to alter secretory granule morphology [24••]. Concurrently, structural and molecular dynamics studies on GalNAc-T4 bound to a diglycopeptide have revealed a flexible loop on its lectin domain that can approach the GalNAc residue of catalytic domain bound glycopeptide [23••]. Mutagenesis of this loop was shown to alter the kinetic properties of GalNAc-T4 against both peptide and glycopeptide substrates thus again confirming that additional features of the lectin domain beyond glycan binding will likely play roles in substrate selection of these transferases.

The flexible linker and its role in the remote glycosylation preferences of the GalNAc-Ts

The catalytic and lectin domains of all of the GalNAc-Ts (except for T20 that lacks the lectin domain) are connected by a linker sequence whose length and sequence varies among isoenzymes [6••,19•,20•,21••,22•]. Comparing linkers, the N-terminal regions are more conserved while the C-terminal regions are less conserved [6••]. Previous studies have attributed the relative positioning of the catalytic and lectin domains to the nature of the linker sequence [21••,39], thus the more stretched-out linker of GalNAc-T10 [22•] results in fewer interactions between both domains compared to the more closely spaced domains in GalNAc-T1 [19•] (Figure 2b). This suggested that linker flexibility could function to control the relative orientation of lectin and catalytic domains, therefore, modulating the selection of new GalNAc-modification sites in previously glycosylated substrates [4••,39,40••].

One of the largest questions in the field has been how these enzymes differentially recognize remote prior glycosylation sites in an N-terminal or C-terminal direction. A recent work on *Hs*GalNAc-T2 and *Hs*GalNAc-T4 shows that their flexible linkers display both interdomain rotation and interdomain translational-like motion which could be responsible of their different long range glycopeptide preferences [6••]. The crystal structure of *Hs*GalNAc-T4 with glycopeptide bound to the lectin domain [6••] revealed that its GalNAc-binding site was located on the opposite side of the lectin domain when compared to the homologous site in *Hs*GalNAc-T2 (Figure 4). These different positions of the lectin domain (Figure 4), readily account how GalNAc-T4 promotes the opposite long-range glycosylation preference compared to GalNAc-T2 and other isoenzymes [1•,4••] (see Figure 1). That this rotation is caused by the nature of the flexible linker was supported by molecular dynamics simulations, site-directed mutagenesis, and kinetics experiments [6••]. Indeed, the glycopeptide kinetics of GalNAc-T2 chimeras containing a GalNAc-T3 or GalNAc-T4 flexible linker and a series of flexible linker mutants, demonstrated that its long-range glycosylation preference could be modulated and even reversed simply by modifying its linker [6••]. This suggests that the flexible linker plays a major role in dictating each isoenzyme's long-range glycosylation preference by altering the lectin domain's orientation relative to its catalytic domain [6••]. All together, these findings showed for the first time how a structural feature that is neither in the active site nor in the lectin domain GalNAc-binding site is capable of modifying the activity and the glycosylation preferences of these isoenzymes.

Final remarks

That the GalNAc-Ts are associated with numerous human diseases including cancer [14,15,16•,41] clearly justifies the importance of unravelling the molecular basis that lie beneath their substrate recognition, ranging from redundant overlapping sites [42] to highly specific targets. Here, we have briefly summarized the most important advances at structural level of this family of enzymes that begin to reveal the molecular origins of their unique peptide and glycopeptide specificities. However, additional structures of these isoenzymes in complex with both their redundant and specific (glyco)peptide substrates will be necessary for a thorough mechanistic understanding of their promiscuity, specificity, and distinct glycosylation preferences. In particular, much more needs to be understood regarding their short-range glycosylation preferences as we currently have only one example describing such GalNAc-T-(glyco)peptide recognition. Hence, it is of utmost importance to continue studying this complex family of enzymes to fully understand how they selectively recognize their targets in multiple signaling pathways. Such studies will in turn facilitate the development of GalNAc-T modulators and inhibitors that would certainly be useful for the treatment of many diseases [11,13,16•,43,44]. Finally, one cannot discard the potential for Nature organizing the GalNAc-T's in a cell according to their isoenzyme class (e.g. early, intermediate and late GTs) utilizing their different glycosylation preferences to produce the vast repertoire of glycosylation sites observed *in vitro*. Such organization is clearly present as the retrograde introduction of GalNAc-Ts into the ER (the so called GALA pathway) has been shown to manifestly alter the patterns of *O*-glycosylation and may play a role in cancer [26•,41]. However, this pathway is currently under an intense debate in the Glycobiology community hence its importance has yet to be fully understood [45].

Acknowledgements

The authors would like to acknowledge the support of the National Institutes of Health (grant R01 GM113534 to T. A. Gerken), ARAID, MEC (grants CTQ2013-44367-C2-2-P and BFU2016-75633-P to R. Hurtado-Guerrero), and the DGA (E34_R17). The research leading to these results has also received funding from the FP7 (2007–2013) under BioStruct-X (Grant agreement 283570 and BIOSTRUCTX_5186).

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA: **Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family**. *Glycobiology* 2011, **22**:736-756.

This work provides an overview of the GalNAc-Ts family members, as well as a classification of its isoenzymes according to their glycosylation preferences.

2. Hollingsworth MA, Swanson BJ: **Mucins in cancer: protection and control of the cell surface**. *Nat Rev Cancer* 2004, **4**:45-60.

3. Kufe DW: **Mucins in cancer: function, prognosis and therapy**. *Nat Rev Cancer* 2009, **9**:874-885.

4. •• Revoredo L, Wang S, Bennett EP, Clausen H, Moremen KW, Jarvis DL, Ten Hagen KG, Tabak LA, Gerken TA: **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* 2015, **26**:360-376.

This work digs into the glycosylation preferences of the different GalNAc-Ts isoenzymes and demonstrate that catalytic and lectin domains have unique functions and work in concert to direct glycosylation. The authors propose a more exhaustive re-classification of the family members.

5. Pratt MR, Hang HC, Ten Hagen KG, Rarick J, Gerken TA, Tabak LA, Bertozzi CR: **Deconvoluting the functions of polypeptide *N*-α-acetylgalactosaminyltransferase family members by glycopeptide substrate profiling**. *Chem Biol* 2004, **11**:1009-1016.

6. •• De Las Rivas M, Lira-Navarrete E, Daniel EJP, Compañón I, Coelho H, Diniz A, Jiménez-Barbero J, Peregrina JM, Clausen H, Corzana F *et al.*: **The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences.**. *Nat Commun* 2017, **8**:1-11.

This work provides for the first time the structural basis for the distinct long-range glycosylation preference of the GalNAc-Ts, and reveals the importance of the flexible linker in fine-tuning their glycosylation preference.

7. •• Gerken TA, Jamison O, Perrine CL, Collette JC, Moinova H, Ravi L, Markowitz SD, Shen W, Patel H, Tabak LA: **Emerging paradigms for the initiation of mucin-type protein *O*-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases**. *J Biol Chem* 2011, **286**:14493-14507.

In this article, the authors makes a thorough study into the GalNAc-Ts substrate glycosylation preferences, and conclude that each isoenzyme may be uniquely sensitive to sequence and overall charge.

8. •• Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB, Clausen H: **Mining the *O*-glycoproteome using zinc-finger nuclease–glycoengineered simplecell lines**. *Nat Methods* 2011, **8**:977-982.

This article describes for the first time the use of zinc-finger nuclease (ZFN) gene targeting as a tool to facilitate analyses of important functions of protein glycosylation.

9. Schjoldager KT, Vakhrushev SY, Kong Y, Steentoft C, Nudelman AS, Pedersen NB, Wandall HH, Mandel U, Bennett EP, Levery SB, Clausen H: **Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered simplecells**. *Proc Nat Acad Sci U S A* 2012, **109**:9893-9898.

10. Pedersen NB, Wang S, Narimatsu Y, Yang Z, Halim A, Schjoldager KT, Madsen TD, Seidah NG, Bennett EP, Levery SB, Clausen H: **Low-density lipoprotein receptor class A repeats are *O*-glycosylated in linker regions**. *J Biol Chem* 2014 http://dx.doi.org/10.1074/jbc.M113.545053.

11. Wang S, Mao Y, Narimatsu Y, Ye Z, Tian W, Goth CK, Lira-Navarrete E, Pedersen NB, Benito-Vicente A, Martin C *et al.*: **Site-specific *O*-glycosylation of members of the low-density lipoprotein receptor superfamily enhances ligand interactions**. *J Biol Chem* 2018, **293**:7408-7422.

12. Schjoldager KT, Vester-Christensen MB, Goth CK, Petersen TN, Brunak S, Bennett EP, Levery SB, Clausen H: **A systematic study of site-specific GalNAc-type *O*-glycosylation modulating proprotein convertase processing.**. *J Biol Chem* 2011, **286**:40122-40132.

13. Kato K, Jeanneau C, Tarp MA, Benet-Pagès A, Lorenz-Depiereux B, Bennett EP, Mandel U, Strom TM, Clausen H: **Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis secretion of fibroblast growth factor 23 requires *O*-glycosylation.**. *J Biol Chem* 2006, **281**:18370-18377.

14. Schjoldager KT, Vester-Christensen MB, Bennett EP, Levery SB, Schwientek T, Yin W, Blixt O, Clausen H: ***O*-Glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3-possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids**. *J Biol Chem* 2010 http://dx.doi.org/10.1074/jbc.M110.156950.

15. Fu C, Zhao H, Wang Y, Cai H, Xiao Y, Zeng Y, Chen H: **Tumor-associated antigens: Tn antigen, STn antigen, and T antigen**. *HLA* 2016, **88**:275-286.

16. • Sletmoen M, Gerken TA, Stokke BT, Burchell J, Brewer CF: **Tn and STn are members of a family of carbohydrate tumor antigens that possess carbohydrate–carbohydrate interactions.**. *Glycobiology* 2018, **28**:437-442.

This interesting work provides a model for the possible roles of Tn and STn antigens in cancer and other diseases.

17. Radhakrishnan P, Dabelsteen S, Madsen FB, Francavilla C, Kopp KL, Steentoft C, Vakhrushev SY, Olsen JV, Hansen L, Bennett EP *et al.*: **Immature truncated *O*-glycophenotype of cancer directly induces oncogenic features**. *Proc Nat Acad Sci U S A* 2014, **111**:4066-4075.

18. Pearce OMT: **Cancer glycan epitopes: biosynthesis, structure and function**. *Glycobiology* 2018, **28**:670-696.

19. • Fritz TA, Hurley JH, Trinh L-B, Shiloach J, Tabak LA: **The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc: polypeptide α-*N*-acetylgalactosaminyltransferase-1**. *Proc Nat Acad Sci U S A* 2004, **101**:15307-15312.

This work provides the first crystal structure of a member of GalNAc-Ts family: *Mm*GalNAc-T1.

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

In this article the authors solve the first ternary complex of *Hs*GalNAc-T2 in complex with UDP and an acceptor peptide. This resolves partly the long-standing question of how GalNAc-Ts recognise 'naked' peptide substrates.

21. •• Lira-Navarrete E, de Las Rivas M, Compañón I, Pallarés MC, Kong Y, Iglesias-Fernández J, Bernardes GJ, Peregrina JM, Rovira C, Bernadó P *et al.*: **Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein o-glycosylation**. *Nat Commun* 2015, **6**:1-11.

This article describes the first ternary complex between a GalNAc-T (GalNAc-T2) and a glycopeptide substrate and UDP-Mn^{+2}, and shows the influence of the mechanical properties of the flexible linker between the catalytic and the lectin domains to account for its observed glycosylation profile.

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

This work provides the first crystal structure of *Hs*GalNAc-T10 in complex with Ser-GalNAc that reveals for the first time how the lectin domain recognises the GalNAc moiety of glycopeptides.

23. •• de las Rivas M, Paul Daniel EJ, Coelho H, Lira-Navarrete E, Raich L, Compañón I, Diniz A, Lagartera L, Jlménez-Barbero J, Clausen H *et al.*: **Structural and mechanistic insights into the catalytic-domain-mediated short range glycosylation preferences of GalNAc-T4**. *ACS Cent Sci* 2018, **4**:1274-1290.

This recent work provides the first structural basis for the short-range glycosylation preferences of a GalNAc-T, the GalNAc-T4, and describes a

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007

flexible loop protruding from the lectin domain that seems to be essential for the optimal activity of the catalytic domain.

24. Ji S, Samara NL, Revoredo L, Zhang L, Tran DT, Muirhead K, •• Tabak LA, Ten Hagen KG: **A molecular switch orchestrates enzyme specificity and secretory granule morphology**. *Nat Commun* 2018, **9**:1-13.

In this work, the authors describe a mechanism for dictating substrate specifity within the *O*-glycosyltransferase enzyme family, based on the charge of a loop that controlles the acces to the active site.

25. Bourne Y, Henrissat B: **Glycoside hydrolases and glycosyltransferases: families and functional modules**. *Curr Opin Struct Biol* 2001, **11**:593-600.

26. Gill DJ, Clausen H, Bard F: **Location, location, location: new • insights into *O*- GalNAc protein glycosylation**. *Trends Cell Biol* 2011, **21**:149-158.

This review analyzes the evidences and the importance of retrograde mislocalization of GalNAc-Ts to the Endoplasmic Reticulum as a mechanism for the overexpression of the Tn antigen leading to and found in many cancers.

27. Hassan H, Bennett EP, Mandel V, Hollingsworth MA, Clausen H: **Control of mucin-type *O*-glycosylation: *O*-glycan occupancy is directed by substrate specificities of polypeptide GalNAc-transferases**. *Carbohydr Chem Biol* 2000:273-292.

28. Perrine CL, Ganguli A, Wu P, Bertozzi CR, Fritz TA, Raman J, Tabak LA, Gerken TA: **The glycopeptide preferring polypeptide-GalNAc transferase-10 (ppGalNAc-T10), involved in mucin type-*O*-glycosylation, has a unique GalNAc-*O*-Ser/Thr binding site in its catalytic domain not found in pp GalNAc-T1 or -T2**. *J Biol Chem* 2009 http://dx.doi.org/10.1074/jbc.M109.017236.

29. Lira-Navarrete E, Iglesias-Fernández J, Zandberg WF, •• Compañón I, Kong Y, Corzana F, Pinto BM, Clausen H, Peregrina JM, Vocadlo DJ *et al.*: **Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase-2**. *Angew Chem Int Ed* 2014, **53**:8206-8210.

This article reports structural snapshots of the retaining *Hs*GalNAc-T2, supporting a front face S_Ni mechanism.

30. Gómez H, Rojas R, Patel D, Tabak LA, Lluch JM, Masgrau L: **A computational and experimental study of *O*-glycosylation. Catalysis by human UDP-GalNAc polypeptide: GalNAc-transferase-2**. *Organ Biomol Chem* 2014, **12**:2645-2655.

31. Trnka T, Kozmon S, Tvaroška I, Koča J: **Stepwise catalytic mechanism via short-lived intermediate inferred from combined qm/mm merp and pes calculations on retaining glycosyltransferase pp GalNAc-T2**. *PLoS Comput Biol* 2015, **11**: e1004061.

32. de las Rivas M, Coelho H, Diniz A, Lira-Navarrete E, Compañón I, •• Jiménez-Barbero J, Schjoldager KT, Bennett EP, Vakhrushev SY, Clausen H *et al.*: **Structural analysis of a GalNAc-T2 mutant reveals an induced-fit catalytic mechanism for GalNAc-Ts**. *Chem A Eur J* 2018 http://dx.doi.org/10.1002/chem.201800701.

In this article the authors decipher the molecular basis for the *Hs*GalNAc-T2 F104S mutant inactivation and reveal that GalNAc-T2 adopts an UDP-GalNAc-dependent induced-fit mechanism.

33. Khetarpal SA, Schjoldager KT, Christoffersen C, Raghavan A, Edmondson AC, Reutter HM, Ahmed B, Ouazzani R, Peloso GM, Vitali C *et al.*: **Loss of function of GalNAc-T2 lowers high-density lipoproteins in humans, nonhuman primates, and rodents.**. *Cell Metab* 2016, **24**:234-245.

34. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ *et al.*: **Biological, clinical and population relevance of 95 loci for blood lipids**. *Nature* 2010, **466**:707-713.

35. Madariaga D, Martínez-Sáez N, Somovilla VJ, García-García L, Berbis MÁ, Valero-Gónzalez J, Martín-Santamaría S, Hurtado-Guerrero R, Asensio JL, Jiménez-Barbero J: **Serine versus threonine glycosylation with α-*O*-GalNAc: unexpected selectivity in their molecular recognition with lectins**. *Chem A Eur J* 2014, **20**:12616-12627.

36. Tenno M, Saeki A, Kezdy FJ, Elhammer AP, Kurosaka A: **The lectin domain of UDP-GalNAc:Polypeptide *N*-acetylgalactosaminyltransferase-1 is involved in *O*-Glycosylation of a polypeptide with multiple acceptor sites**. *J Biol Chem* 2002, **277**:47088-47096.

37. Pedersen JW, Schjoldager KT, Meldal M, Holmer AP, Blixt O, Clo E, Levery SB, Clausen H, Wandall HH: **Lectin domains of polypeptide GalNAc -transferases exhibit glycopeptide binding specificity**. *J Biol Chem* 2011 http://dx.doi.org/10.1074/jbc.M111.273722.

38. Lorenz V, Ditamo Y, Cejas RB, Carrizo ME, Bennett EP, Clausen H, Nores GA, Irazoqui FJ: **Extrinsic functions of lectin domains in on-acetylgalactosamine glycan biosynthesis**. *J Biol Chem* 2016 http://dx.doi.org/10.1074/jbc.M116.740795.

39. Raman J, Fritz TA, Gerken TA, Jamison O, Live D, Lu M, Tabak LA: **The catalytic and lectin domains of UDP- GalNAc: polypeptide alpha-*N*-acetylgalactosaminyltransferase function in concert to direct glycosylation site selection**. *J Biol Chem* 2008.

40. Gerken TA, Revoredo L, Thome JJ, Tabak LA, Vester-•• Christensen MB, Clausen H, Gahlay GK, Jarvis DL, Johnson RW, Moniz HA, Moremen K: **The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (pp GalNAc-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* 2013, **288**:19900-19914.

First systematic study revealing the different N- and C- terminal long-range prior glycosylation preferences of the GalNAc-Ts, which showed the importance of prior glycosylation on the glycosylation of FGF23 by GalNAc-T3.

41. Gill DJ, Tham KM, Chia J, Wang SC, Steentoft C, Clausen H, Bard-Chapeau EA, Bard FA: **Initiation of GalNAc-type *O*-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness**. *Proc Nat Acad Sci U S A* 2013, **110**:3152-3161.

42. Kong Y, Joshi HJ, Schjoldager KT, Madsen TD, Gerken TA, Vester-Christensen MB, Wandall HH, Bennett EP, Levery SB, Vakhrushev SY, Clausen H: **Probing polypeptide GalNAc-transferase isoform substrate specificities by *in vitro* analysis**. *Glycobiology* 2014, **25**:55-65.

43. Ghirardello M, de las Rivas M, Lacetera A, Delso I, Lira-Navarrete E, Tejero T, Martín-Santamaría S, Hurtado-Guerrero R, Merino P: **Glycomimetics targeting glycosyltransferases: synthetic, computational and structural studies of less-polar conjugates**. *Chem A Eur J* 2016, **22**:7215-7224.

44. Liu F, Xu K, Xu Z, de las Rivas M, Li X, Lu J, Delso I, Merino P, Hurtado-Guerrero R, Zhang Y: **The small molecule luteolin inhibits *N*-acetyl-α-galactosaminyltransferases and reduces mucin-type *O*-glycosylation of amyloid precursor protein**. *J Biol Chem* 2017 http://dx.doi.org/10.1074/jbc.M117.814202.

45. Herbomel GG, Rojas RE, Tran DT, Ajinkya M, Beck L, Tabak LA: **The GalNAc-T activation pathway (GALA) is not a general mechanism for regulating mucin-type *O*-glycosylation.**. *PLoS one* 2017, **12**:e0179241.

46. Festari MF, Trajtenberg F, Berois N, Pantano S, Revoredo L, Kong Y, Solari-Saquieres P, Narimatsu Y, Freire T, Bay S *et al.*: **Revisiting the human polypeptide GalNAc-T1 and T13 paralogs**. *Glycobiology* 2017, **27**:140-153.

47. Li X, Wang J, Li W, Xu Y, Shao D, Xie Y, Xie W, Kubota T, Narimatsu H, Zhang Y: **Characterization of ppGalNAc-T18, a member of the vertebrate-specific Y subfamily of UDP-*N*-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferases**. *Glycobiology* 2012, **22**:602-615.

Please cite this article in press as: de las Rivas M, et al.: Polypeptide GalNAc-Ts: from redundancy to specificity, Curr Opin Struct Biol (2019), https://doi.org/10.1016/j.sbi.2018.12.007