

Dear author,

Please note that changes made in the online proofing system will be added to the article before publication but are not reflected in this PDF.

We also ask that this file not be used for submitting corrections.



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Current Opinion in  
Structural Biology

# Polypeptide GalNAc-Ts: from redundancy to specificity

Matilde de las Rivas<sup>1,5</sup>, Erandi Lira-Navarrete<sup>2,5</sup>,  
Thomas A Gerken<sup>3</sup> and Ramon Hurtado-Guerrero<sup>1,4</sup>

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

<sup>1</sup> BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus, Rio Ebro, Edificio I+D, Zaragoza, Spain

<sup>2</sup> Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup> Departments of Biochemistry, Chemistry and Pediatrics Case Western Reserve University, Cleveland, OH, USA

<sup>4</sup> Fundación ARAID, 50018, Zaragoza, Spain

Corresponding authors: Gerken, Thomas A. ([txg2@cwru.edu](mailto:txg2@cwru.edu)), Hurtado-Guerrero, Ramon ([rhurtado@bifi.es](mailto:rhurtado@bifi.es))

<sup>5</sup> 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>

0959-440X/© 2018 Elsevier Ltd. All rights reserved.

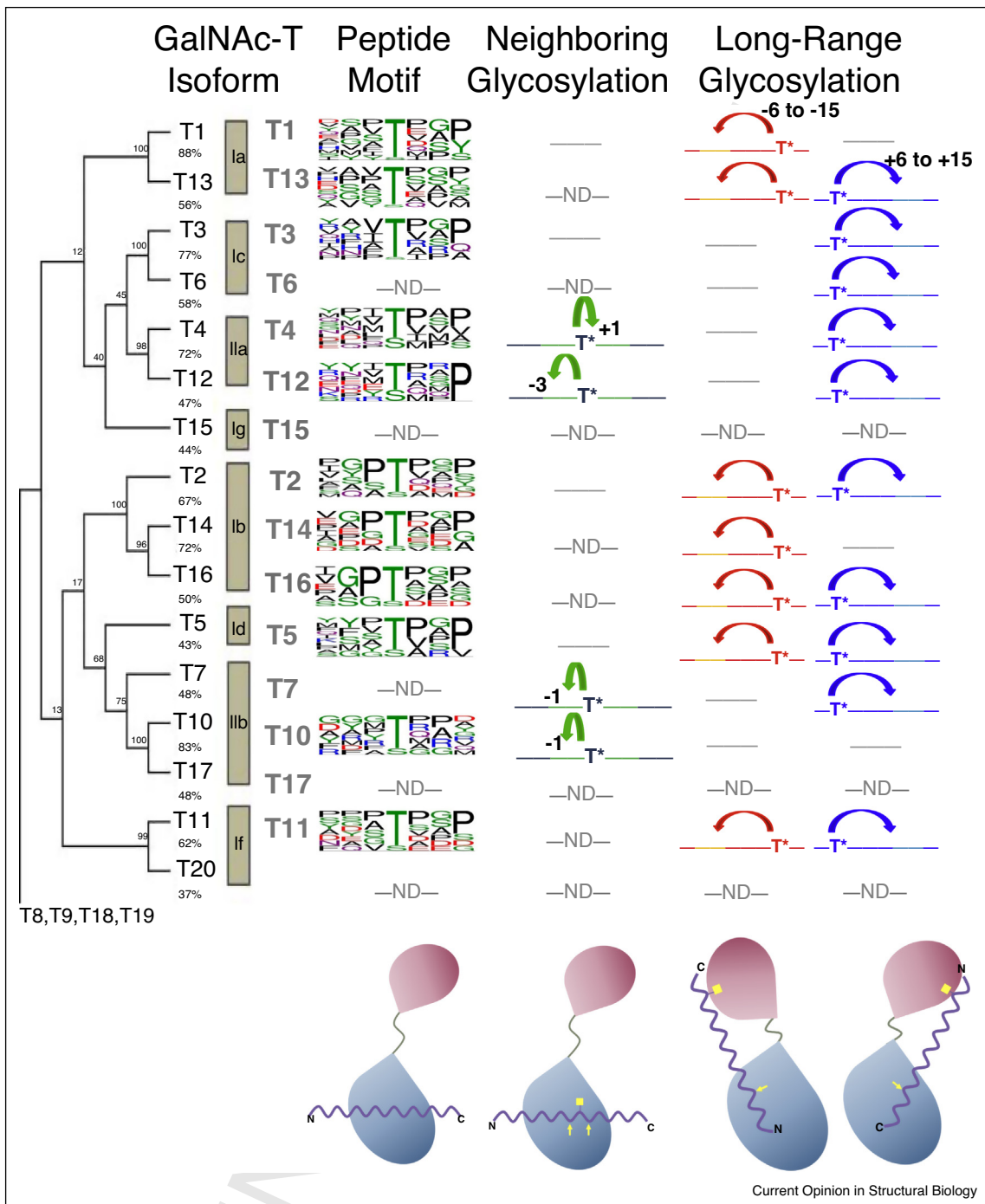
## Introduction

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 (GalNAc $\alpha$ 1-*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

## 2 Carbohydrates

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.

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

90 The fact that most isoenzymes of this family are capable  
91 of glycosylating common acceptor substrates, particularly  
92 those containing the (Thr/Ser)ProXPro motif (where 'X'  
93 usually stands for a small hydrophobic residue; see Fig-  
94 ure 1), suggests that they may also serve redundant  
95 functions [4<sup>••</sup>,7<sup>••</sup>]. Paradoxically, in recent years an incre-  
96 mental number of reports have demonstrated that several  
97 GalNAc-T isoenzymes are highly specific for certain  
98 protein substrates, as identified by the Simple Cell  
99 (SC) approach developed by the Clausen group [8<sup>••</sup>,9].  
100 Using this strategy, ApoC-III was identified as a specific  
101 substrate of GalNAc-T2 [9] and GalNAc-T11 was  
102 reported to be specifically involved in glycosylation of  
103 the peptide linkers between class A repeats of the LDL  
104 receptor family [10,11]. One common theme found for  
105 isoenzyme-specific glycosylation is that such a glycosyla-  
106 tion can interfere with the proprotein processing of  
107 neighboring sites thus controlling such a processing  
108 [12]. Two of the most studied examples are GalNAc-  
109 T3 and its substrate, the fibroblast growth factor 23  
110 (FGF23) [13], as well as GalNAc-T2 and angiopoietin-  
111 like Protein 3 (ANGPTL3) [14], where miss-regulation of  
112 *O*-glycosylation can lead to familial tumoral calcinosis and  
113 dyslipidemia, respectively [13,14]. In addition, the aber-  
114 rant expression or mutation of several GalNAc-T isoen-  
115 zymes and the overexpression of the Tn-antigen is  
116 directly associated with many cancers [15,16<sup>•</sup>,17], the  
117 mechanisms of which are still not understood, although  
118 Tn–Tn self-association may play a role [16<sup>•</sup>]. Finally, the  
119 GalNAc-Ts are involved or implicated in many other  
120 biological functions including development, receptor traf-  
121 ficking and modulation and protein secretion, which are  
122 beyond the scope of this review [18].

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

### 130 Structural similarities between GalNAc-Ts 131 isoenzymes

132 To obtain a complete understanding of the mechanism  
133 underlying the substrate specificity of this family of  
134 enzymes, crystal structures of GalNAc-T isoenzymes  
135 have been solved, either in the apo form or in complex  
136 with (glyco)peptide substrates and products. To date, the  
137 following structures have been reported (see Figure 2a  
138 and b): a) *Mus musculus* GalNAc-T1 with Mn<sup>2+</sup> (*Mm*Gal-  
139 NAc-T1) (PDB entry 1XHB) [19<sup>•</sup>]; b) human GalNAc-  
140 T2 (*Hs*GalNAc-T2) complexed with UDP-Mn<sup>2+</sup> (PDB

141 entry 2FFV), a 'naked' peptide (PDB entry 2FFU) [20<sup>•</sup>]  
142 and three glycopeptides (PDB entries 5AJP, 5AJO and  
143 5AJN) [21<sup>••</sup>]; c) *Hs*GalNAc-T10 complexed with UDP-  
144 Mn<sup>2+</sup> and Ser-GalNAc (PDB entries 2D7I and 2D7R)  
145 [22<sup>•</sup>]; d) *Hs*GalNAc-T4 complexed with UDP, Mn<sup>2+</sup> and  
146 a glycopeptide (PDB entry 5NQA) [6<sup>••</sup>]; and e) the  
147 recently reported structures of *Hs*GalNAc-T4 complexed  
148 with UDP, Mn<sup>2+</sup> and a diglycopeptide (PDB entry 6H0B)  
149 [23<sup>••</sup>] and two splice variants of the fly PGANT9A/B with  
150 UDP-Mn<sup>2+</sup> and a peptide substrate (PDB entries 6E4Q  
151 and 6E4R; Figure 2b) [24<sup>••</sup>].

152 These structures all show an N-terminal catalytic domain  
153 adopting the typical GT-A fold, characterized by two  
154 abutting Rossmann-like folds which is linked by a short  
155 flexible linker to a C-terminal ricin-like lectin domain  
156 [6<sup>••</sup>,19<sup>•</sup>,20<sup>•</sup>,21<sup>••</sup>,22<sup>•</sup>] (Figure 2a). The lectin domain, a  
157 unique structural feature only present in this family of  
158 eukaryotic GTs, has a  $\beta$ -trefoil fold built from three  
159 repeat units ( $\alpha$ ,  $\beta$  and  $\gamma$ ) that are potentially capable to  
160 bind a GalNAc moiety [25,26<sup>•</sup>]. It should be noted that  
161 these repeats are not necessarily all active binders based  
162 on their sequence motif [4<sup>••</sup>] and by experiment [7<sup>••</sup>,27].

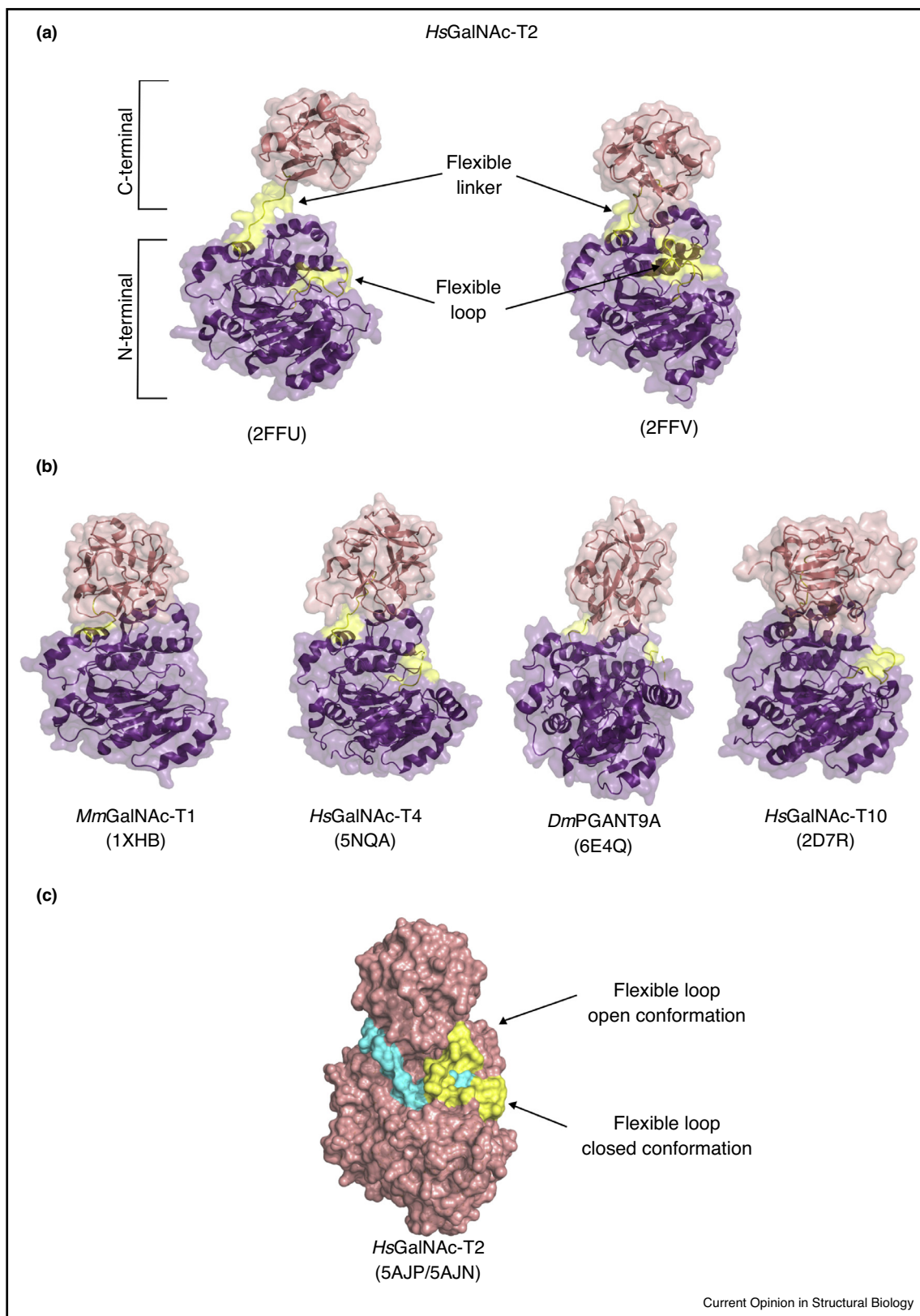
### 163 GalNAc-T catalytic domain: UDP-GalNAc 164 binding site and flexible loop

165 The first crystal structure of a GalNAc-T was for *Mm*Gal-  
166 NAc-T1 [19<sup>•</sup>]. It provided the initial picture of the active  
167 site for which subsequent GalNAc-T structures highly  
168 resemble, particularly in the architecture of the critical  
169 and conserved Mn<sup>2+</sup> binding site (formed by Asp209,  
170 His211 (the DXH motif) and His344). This structure also  
171 showed that the catalytic and lectin domains were closely  
172 associated [19<sup>•</sup>] (Figure 2b). Subsequent crystal struc-  
173 tures of GalNAc-T2 complexed to both UDP-Mn<sup>2+</sup> and  
174 to UDP-Mn<sup>2+</sup> and the EA2 peptide (AspSerThrThrProAla-  
175 ProThrThrLys) [20<sup>•</sup>], together with the GalNAc-T10  
176 crystal structure complexed with hydrolyzed UDP-Gal-  
177 NAc and Mn<sup>2+</sup> [22<sup>•</sup>], further defined the GT-A fold active  
178 site residues that tethered the uridine diphosphate of the  
179 UDP-GalNAc donor substrate [20<sup>•</sup>,22<sup>•</sup>]. In addition, the  
180 structure of GalNAc-T10 catalytic domain revealed the  
181 GalNAc moiety bound in the UDP-GalNAc-binding  
182 pocket [22<sup>•</sup>,28]. These structures, as well as the crystal  
183 structures of the first pre-Michaelis and Michaelis com-  
184 plexes of *Hs*GalNAc-T2 [29<sup>••</sup>], were of fundamental  
185 importance for defining models of the dynamics of Gal-  
186 NAc-T2 during its catalytic cycle, which consists of an  
187 ordered bi–bi kinetic mechanism [29<sup>••</sup>,30,31]. In addi-  
188 tion, the Michaelis complex revealed that these GTs  
189 follow a front-face S<sub>N</sub>i-type reaction mechanism [29<sup>••</sup>].

190 Structures of GalNAc-Ts, both with and without bound  
191 peptide substrate, have revealed a dynamic flexible loop  
192 at the surface of the catalytic domain substrate binding  
193 site as an important structural feature of the GalNAc-Ts  
194 [20<sup>•</sup>,21<sup>••</sup>,29<sup>••</sup>]. This flexible loop, formed by residues

4 Carbohydrates

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

195 Arg362 to Ser373 in *HsGalNAc-T2*, is able to adopt either  
 196 a closed conformation to form a lid over UDP, rendering  
 197 the enzyme in an active form, or an open conformation in  
 198 which the loop folds back to expose UDP to the bulk  
 199 solvent (inactive form; Figure 2c) [20<sup>•</sup>,29<sup>••</sup>]. This inter-  
 200 conversion has recently been shown to be dependent of  
 201 the presence of UDP-GalNAc in *HsGalNAc-T2*, where  
 202 this sugar nucleotide stabilizes the closed conformation  
 203 and consequently allows the binding of the substrate  
 204 peptide [32<sup>••</sup>]. The importance of this flexible loop in  
 205 catalysis is exemplified by the molecular basis of the  
 206 inactivation of the *HsGalNAc-T2*-Phe104Ser mutant,  
 207 which is linked to low levels of high density lipoprotein  
 208 cholesterol [32<sup>••</sup>,33]. It was found that Phe104 controls  
 209 the inactive-to-active transition of the flexible loop due to  
 210 its hydrophobic interaction with Ala151/Ile256/Val360, as  
 211 well as a CH- $\pi$  interaction with the side-chain of Arg362  
 212 located in the flexible loop [32<sup>••</sup>]. The hydrophilic  
 213 Phe104Ser mutation fails to lock the flexible loop in its  
 214 active form, thus impeding peptide substrate binding and  
 215 the failure to glycosylate its targeted peptide substrates (i.  
 216 e. ApoC-III and ANGPTL3). This results in low levels of  
 217 HDL [32<sup>••</sup>,33,34].

### 218 GalNAc-T catalytic domain: peptide-binding 219 site

220 It is noteworthy that several GalNAc-T crystals soaked or  
 221 cocrystallized with (glyco)peptides show indeterminate/  
 222 disordered structures for the substrate bound to the  
 223 catalytic domain [6<sup>••</sup>,24<sup>••</sup>]. Nevertheless, structural infor-  
 224 mation of the GalNAc-T-peptide acceptor recognition  
 225 could be inferred from the series of structures of (glyco)  
 226 peptides bound to the *HsGalNAc-T2* isoenzyme [21<sup>••</sup>]  
 227 and very recently the diglycopeptide bound to *HsGal-*  
 228 *NAc-T4* [23<sup>••</sup>]. In these latter structures the interactions  
 229 between the transferase and its acceptor substrates are  
 230 dissimilar, suggesting differences between isoenzymes at  
 231 the peptide binding groove level. However, this could  
 232 also be due to the different (glyco)peptides used for both  
 233 isoenzymes. The *HsGalNAc-T2* structures revealed that  
 234 the EA2 peptide bound in a shallow cleft on the surface of  
 235 *HsGalNAc-T2*, being recognized by hydrophobic interac-  
 236 tions and to a lesser extent hydrogen bond interactions  
 237 [20<sup>•</sup>] (see Figure 3a). It was also observed that the methyl  
 238 group of the acceptor Thr residue was embedded within a  
 239 hydrophobic pocket, providing a plausible explanation of  
 240 why most GalNAc-T isoenzymes prefer to glycosylate  
 241 Thr over Ser acceptor residues [20<sup>•</sup>,21<sup>••</sup>,35] (Figure 3a).  
 242 Several other crystal structures of *HsGalNAc-T2* in com-  
 243 plex with UDP-Mn<sup>2+</sup> and glycopeptides also showed that  
 the glycopeptides acted as bridges between the catalytic

244 and lectin domains, where the latter bound the glycopep-  
 245 tide GalNAc [21<sup>••</sup>]. In these structures, UDP and the  
 246 glycopeptides were bound to an adaptable sugar-nucleo-  
 247 tide binding site, with the flexible loop adopting either  
 248 open or closed conformations (Figure 2c). Interestingly,  
 249 the binding of a mono-glycopeptide to GalNAc-T4  
 250 revealed peptide GalNAc binding at the lectin domain  
 251 but no observable peptide electron density in its catalytic  
 252 domain [6<sup>••</sup>] while recently a homologous diglycopeptide  
 253 showed a well-resolved peptide bound to the catalytic  
 254 domain in a closed conformation due to GalNAc-T4's  
 255 neighboring glycopeptide binding activity (discussed  
 256 below) [23<sup>••</sup>] (Figure 3b). Interestingly, in the Gal-  
 257 *NAc-T4* structure, the portion of the peptide spanning  
 258 the catalytic and lectin domains was found disordered  
 259 [21<sup>••</sup>,23<sup>••</sup>].

260 At the level of the peptide-binding groove, it was further  
 261 observed that three highly conserved aromatic residues  
 262 (namely Phe361, Phe280 and Trp282 in GalNAc-T2),  
 263 interact with the (Thr/Ser)-Pro-X-Pro substrate sequence  
 264 [20<sup>•</sup>]. Thus far the (Thr/Ser)ProXPro sequence is the only  
 265 substrate consensus motif remotely conserved among  
 266 most GalNAc-Ts (Figures 1, 3a and b) [1<sup>•</sup>,21<sup>••</sup>]. Indeed,  
 267 all isoenzymes that experimentally display this (Thr/Ser)  
 268 ProXPro preference possess the homologous Phe and Trp  
 269 residues [4<sup>••</sup>,7<sup>••</sup>,23<sup>••</sup>] including GalNAc-T4 and Gal-  
 270 *NAc-T12*. GalNAc-T7 and GalNAc-T10, which lack  
 271 these conserved residues and do not exhibit the (Thr/  
 272 Ser)ProXPro preference, instead display strong neigh-  
 273 bouring glycosylation preferences at the +1 position rela-  
 274 tive to the acceptor (i.e. (Thr/Ser)(Thr\*/Ser\*), where  
 275 \* = -O-GalNAc) [4<sup>••</sup>,28]. These latter two isoenzymes  
 276 are, therefore, expected to contain a GalNAc binding  
 277 site in place of the ProXPro binding site found in the  
 278 other isoenzymes. Presently, the structural and molecular  
 279 bases for the neighboring glycosylation preferences of  
 280 GalNAc-T7 and GalNAc-T10 remain to be determined.  
 281 Thus, the near lack of a conserved substrate consensus  
 282 motif together with their active site flexibility points to  
 283 the versatility of these enzymes, allowing them to sculpt  
 284 their binding sites to accommodate a wide range of  
 285 acceptor substrates.

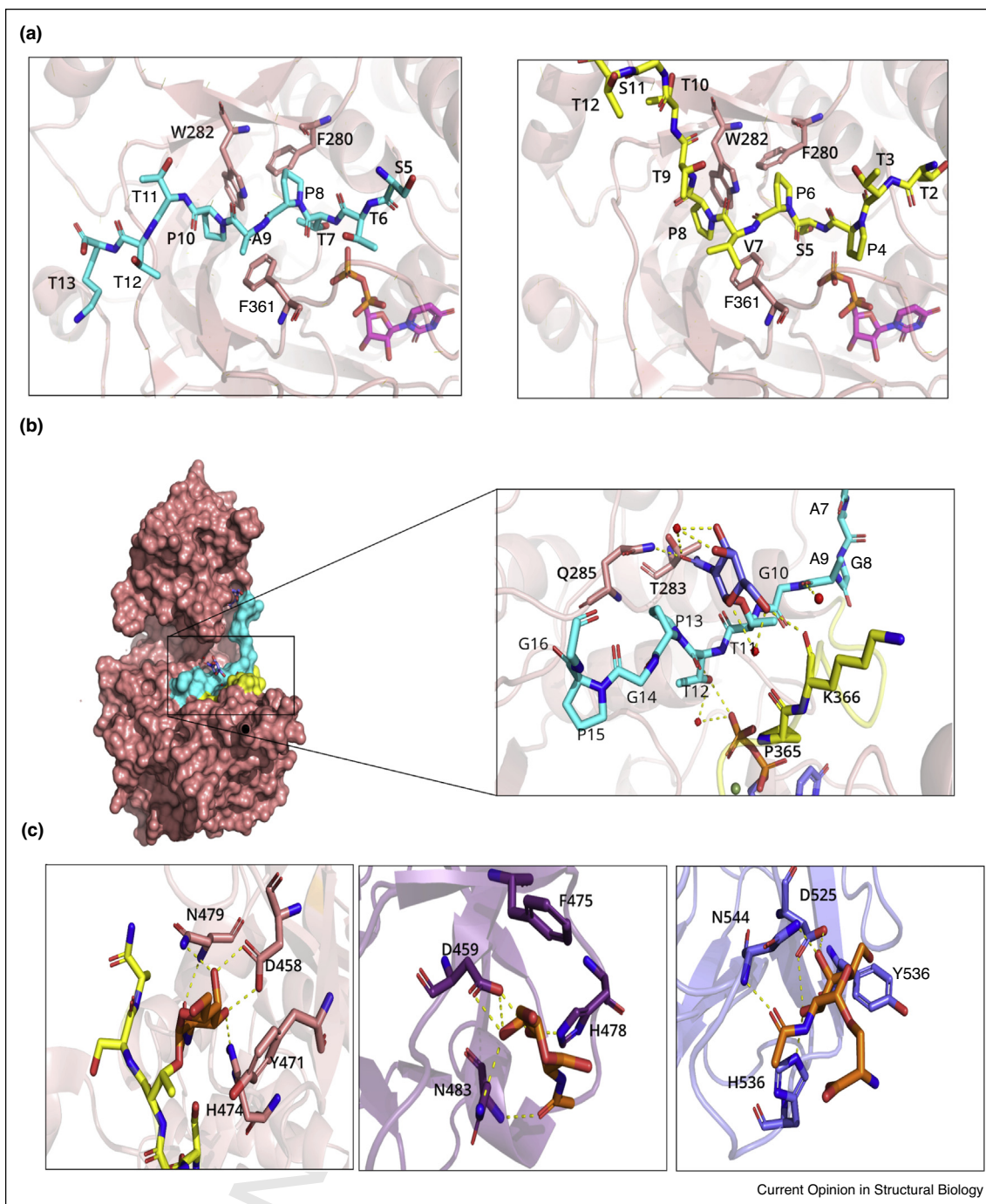
### 286 GalNAc-T catalytic domain: glycopeptide- 287 binding site

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

(Figure 2 Legend Continued) panel) and compact (right panel) forms of monomeric *HsGalNAc-T2*. (b) Cartoon and surface representation of *MmGalNAc-T1*, *HsGalNAc-T4*, *DmPGANT9A* and *HsGalNAc-T10*. (c) Surface representation of the *HsGalNAc-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.

## 6 Carbohydrates

Figure 3



Interactions between GalNAc-Ts and their substrates.

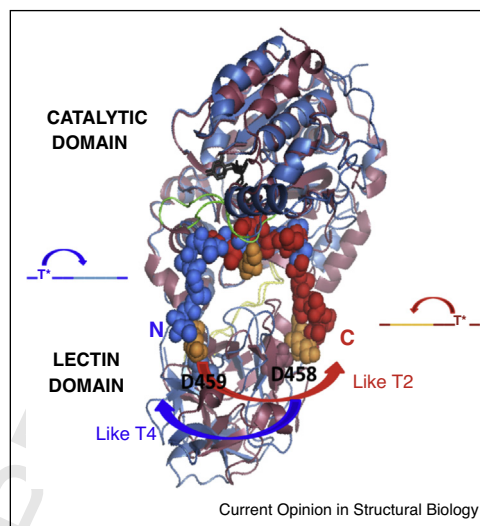
**(a)** Close-up view of the catalytic domain of *HsGalNAc-T2* with two different peptides, EA2 (cyan sticks; left panel) and glycopeptide MUC5AC-13 (GlyThrThrProSerProValProThrThrSerThrThr<sup>+</sup>SerAlaPro) (yellow sticks; right panel), which are similarly recognized by *HsGalNAc-T2* through a hydrophobic 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 *HsGalNAc-T4*-UDP-Diglycopeptide 6 (GlyAlaThr<sup>+</sup>3GlyAlaGlyAlaGlyAlaGlyThr<sup>+</sup>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 *HsGalNAc-T4* catalytic domain glycopeptide binding-site and the GalNAc group on T<sup>+11</sup> 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

293 GalNAc) bound to both the lectin and catalytic domains of  
 294 *HsGalNAc-T4* now reveals how this occurs at least in one  
 295 isoenzyme [23\*\*] (Figure 3b). In this structure the GalNAc  
 296 of Thr\*<sub>11</sub> is shown tethered by hydrogen bond and hydrophobic  
 297 interactions to the side chains of Thr283 and Gln285  
 298 and the back bone of Lys366 at the surface of the catalytic  
 299 domain. Importantly, these residues are not conserved  
 300 among other GalNAc-T isoenzymes and there is no discernible  
 301 cleft or pocket for the binding of the GalNAc (Figure 3b). Such  
 302 GalNAc binding presents the adjacent Thr<sub>12</sub> into the correct  
 303 orientation to accept GalNAc from the UDP-GalNAc donor. Kinetic  
 304 studies on a series of glycopeptide substrates further confirmed  
 305 that the neighbouring GalNAc binding at the catalytic domain  
 306 was weaker than the remote GalNAc binding of Thr\*<sub>3</sub> to the  
 307 lectin domain and further revealed substrate inhibition  
 308 kinetics on the diglycopeptide, presumably due to competitive  
 309 binding of the two Thr\*'s of the substrate at the lectin domain  
 310 [23\*\*]. This work is of additional significance as the individual  
 311 GalNAc-T4 remote and neighboring glycopeptide activities, and  
 312 both together, could be eliminated or greatly reduced by selective  
 313 mutagenesis.

### GalNAc-Ts lectin domain

314 The GalNAc-T-glycopeptide recognition at the lectin  
 315 domain is more easily compared among isoenzymes, as there are  
 316 crystal structures of *HsGalNAc-T2*, *HsGalNAc-T4* and *HsGalNAc-T10*  
 317 complexed with Ser-*O*-GalNAc as well as longer Thr-*O*-GalNAc  
 318 containing glycopeptides [6\*\*,21\*\*,22\*,23\*\*] (Figure 3c). The  
 319 GalNAc-T1 lectin domain contains two known functional GalNAc-  
 320 binding sites out of the possible three (i.e. the  $\alpha$  and  $\beta$  sub-  
 321 domains) [36], whereas the GalNAc-T2, GalNAc-T4, and  
 322 GalNAc-T10 lectin domains contain only one known active  
 323 site (i.e.  $\alpha$ -,  $\alpha$ - and  $\beta$ -respectively) [4\*\*]. The first  
 324 structure of a glycopeptide bound to the lectin domain, that  
 325 is, Ser-*O*-GalNAc bound to *HsGalNAc-T10*, revealed the  
 326 sugar moiety bound to the  $\beta$ -site interacting through several  
 327 hydrogen bonds (including residues Asp525, Asn544, Tyr536) and  
 328 one CH- $\pi$  interaction (His539) (Figure 3c). Subsequent  
 329 structures of GalNAc-T2 complexed with longer glycopeptides  
 330 showed no discernible interactions with the peptide backbone of  
 331 the lectin domain level, while the GalNAc moiety interacted  
 332 exclusively with residues in the  $\alpha$ -subdomain binding site by  
 333 similar interactions as described for GalNAc-T10 (i.e. via  
 334 residues Asp458/Asn479/Tyr471 and His474). These residues  
 335 are conserved in nearly all isoenzymes (Figure 3c) [21\*\*,22\*].  
 336 Similarly, binding interactions of the peptide GalNAc residue  
 337 to the lectin  $\alpha$ -domain of GalNAc-T4 were recently reported  
 338 [6\*\*,23\*\*] (Figure 3c); however, a large difference in the  
 339 orientation of the lectin domains of  
 340 GalNAc-T4 and GalNAc-T2 relative their catalytic domains was  
 341 observed. As discussed in the sections below, these differences  
 342 readily explain the origins of their different long range N-prior  
 343 or C-prior glycosylation preferences (see Figures 1 and 4).  
 344 Earlier work had suggested that the lectin domain of GalNAc-Ts  
 345 could likely influence substrate specificity by steric hindrance  
 346 that would depend on the size of the amino acid side chains of  
 347 the glycopeptide substrate [37], while it has also been  
 348 suggested that the lectin domains of some GalNAc-Ts could  
 349 form hetero-dimers and/or homo-dimers that could also alter  
 350 their specificity [38]. The recent crystal structures of the fly  
 351 PGANT9-A and PGANT9-B lectin domain splice variants now  
 352 offers intriguing evidence for something like the former [24\*\*].  
 353 In this case, a loop on the lectin domain that protrudes toward  
 354 the catalytic domain peptide binding site differs in charge  
 355 between the splice variants. These charge  
 356 differences

Figure 4



Superposition of *HsGalNAc-T2* and *HsGalNAc-T4*. Superimposed cartoon representations of *HsGalNAc-T2*-UDP-MUC5AC-13 glycopeptide (GlyThrThrProSerProValProThrThrSerThrThr\*SerAlaPro) complex depicted in red and *HsGalNAc-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.

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).

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).



## 8 Carbohydrates

361 differences correlate with their activities toward highly  
 362 charged substrates, thus suggesting that at least electro-  
 363 static interactions, if not direct peptide substrate binding,  
 364 of the lectin domain can significantly influence transfer-  
 365 ase activity [24\*\*]. Biologically, these splice variances are  
 366 used to properly glycosylate different secretory mucins,  
 367 whose incomplete glycosylation is shown to alter secre-  
 368 tory granule morphology [24\*\*]. Concurrently, structural  
 369 and molecular dynamics studies on GalNAc-T4 bound to  
 370 a diglycopeptide have revealed a flexible loop on its lectin  
 371 domain that can approach the GalNAc residue of catalytic  
 372 domain bound glycopeptide [23\*\*]. Mutagenesis of this  
 373 loop was shown to alter the kinetic properties of GalNAc-  
 374 T4 against both peptide and glycopeptide substrates thus  
 375 again confirming that additional features of the lectin  
 376 domain beyond glycan binding will likely play roles in  
 377 substrate selection of these transferases.

### 378 The flexible linker and its role in the remote 379 glycosylation preferences of the GalNAc-Ts

380 The catalytic and lectin domains of all of the GalNAc-Ts  
 381 (except for T20 that lacks the lectin domain) are connected  
 382 by a linker sequence whose length and sequence varies  
 383 among isoenzymes [6\*\*,19\*,20\*,21\*\*,22\*]. Comparing link-  
 384 ers, the N-terminal regions are more conserved while the  
 385 C-terminal regions are less conserved [6\*\*]. Previous stud-  
 386 ies have attributed the relative positioning of the catalytic  
 387 and lectin domains to the nature of the linker sequence  
 388 [21\*\*,39], thus the more stretched-out linker of GalNAc-  
 389 T10 [22\*] results in fewer interactions between both  
 390 domains compared to the more closely spaced domains  
 391 in GalNAc-T1 [19\*] (Figure 2b). This suggested that linker  
 392 flexibility could function to control the relative orientation  
 393 of lectin and catalytic domains, therefore, modulating the  
 394 selection of new GalNAc-modification sites in previously  
 395 glycosylated substrates [4\*\*,39,40\*\*].

396 One of the largest questions in the field has been how  
 397 these enzymes differentially recognize remote prior gly-  
 398 cosylation sites in an N-terminal or C-terminal direction.  
 399 A recent work on *Hs*GalNAc-T2 and *Hs*GalNAc-T4  
 400 shows that their flexible linkers display both interdomain  
 401 rotation and interdomain translational-like motion which  
 402 could be responsible of their different long range glyco-  
 403 peptide preferences [6\*\*]. The crystal structure of *Hs*Gal-  
 404 NAc-T4 with glycopeptide bound to the lectin domain  
 405 [6\*\*] revealed that its GalNAc-binding site was located on  
 406 the opposite side of the lectin domain when compared to  
 407 the homologous site in *Hs*GalNAc-T2 (Figure 4). These  
 408 different positions of the lectin domain (Figure 4), readily  
 409 account how GalNAc-T4 promotes the opposite long-  
 410 range glycosylation preference compared to GalNAc-T2  
 411 and other isoenzymes [1\*,4\*\*] (see Figure 1). That this  
 412 rotation is caused by the nature of the flexible linker was  
 413 supported by molecular dynamics simulations, site-  
 414 directed mutagenesis, and kinetics experiments [6\*\*].  
 415 Indeed, the glycopeptide kinetics of GalNAc-T2

416 chimeras containing a GalNAc-T3 or GalNAc-T4 flexible  
 417 linker and a series of flexible linker mutants, demon-  
 418 strated that its long-range glycosylation preference could  
 419 be modulated and even reversed simply by modifying its  
 420 linker [6\*\*]. This suggests that the flexible linker plays a  
 421 major role in dictating each isoenzyme's long-range gly-  
 422 cosylation preference by altering the lectin domain's  
 423 orientation relative to its catalytic domain [6\*\*]. All  
 424 together, these findings showed for the first time how a  
 425 structural feature that is neither in the active site nor in  
 426 the lectin domain GalNAc-binding site is capable of  
 427 modifying the activity and the glycosylation preferences  
 428 of these isoenzymes.

### 429 Final remarks

430 That the GalNAc-Ts are associated with numerous human  
 431 diseases including cancer [14,15,16\*,41] clearly justifies the  
 432 importance of unravelling the molecular basis that lie  
 433 beneath their substrate recognition, ranging from redun-  
 434 dant overlapping sites [42] to highly specific targets. Here,  
 435 we have briefly summarized the most important advances  
 436 at structural level of this family of enzymes that begin to  
 437 reveal the molecular origins of their unique peptide and  
 438 glycopeptide specificities. However, additional structures  
 439 of these isoenzymes in complex with both their redundant  
 440 and specific (glyco)peptide substrates will be necessary for  
 441 a thorough mechanistic understanding of their promiscuity,  
 442 specificity, and distinct glycosylation preferences. In par-  
 443 ticular, much more needs to be understood regarding their  
 444 short-range glycosylation preferences as we currently have  
 445 only one example describing such GalNAc-T-(glyco)pep-  
 446 tide recognition. Hence, it is of utmost importance to  
 447 continue studying this complex family of enzymes to fully  
 448 understand how they selectively recognize their targets in  
 449 multiple signaling pathways. Such studies will in turn  
 450 facilitate the development of GalNAc-T modulators and  
 451 inhibitors that would certainly be useful for the treatment of  
 452 many diseases [11,13,16\*,43,44]. Finally, one cannot dis-  
 453 card the potential for Nature organizing the GalNAc-T's in  
 454 a cell according to their isoenzyme class (e.g. early, inter-  
 455 mediate and late GTs) utilizing their different glycosyla-  
 456 tion preferences to produce the vast repertoire of glycosyl-  
 457 ation sites observed *in vitro*. Such organization is clearly  
 458 present as the retrograde introduction of GalNAc-Ts into  
 459 the ER (the so called GALA pathway) has been shown to  
 460 manifestly alter the patterns of *O*-glycosylation and may  
 461 play a role in cancer [26\*,41]. However, this pathway is  
 462 currently under an intense debate in the Glycobiology  
 463 community hence its importance has yet to be fully under-  
 464 stood [45].

### 465 Acknowledgements

466 The authors would like to acknowledge the support of the National  
 467 Institutes of Health (grant R01 GM113534 to T. A. Gerken), ARAID, MEC  
 468 (grants CTQ2013-44367-C2-2-P and BFU2016-75633-P to R. Hurtado-  
 469 Guerrero), and the DGA (E34\_R17). The research leading to these results  
 470 has also received funding from the FP7 (2007–2013) under BioStruct-X  
 471 (Grant agreement 283570 and BIOSTRUCTX\_5186).

## References and recommended reading

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

- of special interest
- of outstanding interest

- 472 1. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA:  
473 • **Control of mucin-type O-glycosylation: a classification of the**  
474 **polypeptide GalNAc-transferase gene family.** *Glycobiology*  
2011, **22**:736-756.
- 475 This work provides an overview of the GalNAc-Ts family members, as well  
476 as a classification of its isoenzymes according to their glycosylation  
477 preferences.
- 481 2. Hollingsworth MA, Swanson BJ: **Mucins in cancer: protection**  
482 **and control of the cell surface.** *Nat Rev Cancer* 2004, **4**:45-60.
- 483 3. Kufe DW: **Mucins in cancer: function, prognosis and therapy.**  
484 *Nat Rev Cancer* 2009, **9**:874-885.
- 485 4. Revoredo L, Wang S, Bennett EP, Clausen H, Moremen KW,  
486 •• Jarvis DL, Ten Hagen KG, Tabak LA, Gerken TA: **Mucin-type O-**  
487 **glycopeptide substrate recognition that varies among**  
488 **members of the polypeptide GalNAc transferase family.**  
489 *Glycobiology* 2015, **26**:360-376.
- 490 This work digs into the glycosylation preferences of the different GalNAc-  
491 Ts isoenzymes and demonstrate that catalytic and lectin domains have  
492 unique functions and work in concert to direct glycosylation. The authors  
propose a more exhaustive re-classification of the family members.
- 493 5. Pratt MR, Hang HC, Ten Hagen KG, Rarick J, Gerken TA,  
494 Tabak LA, Bertozzi CR: **Deconvoluting the functions of**  
495 **polypeptide N- $\alpha$ -acetylglucosaminyltransferase family**  
496 **members by glycopeptide substrate profiling.** *Chem Biol* 2004,  
**11**:1009-1016.
- 497 6. De Las Rivas M, Lira-Navarrete E, Daniel EJP, Compañón I,  
498 •• Coelho H, Diniz A, Jiménez-Barbero J, Peregrina JM, Clausen H,  
499 Corzana F et al.: **The interdomain flexible linker of the**  
500 **polypeptide GalNAc transferases dictates their long-range**  
501 **glycosylation preferences.** *Nat Commun* 2017, **8**:1-11.
- 502 This work provides for the first time the structural basis for the distinct  
503 long-range glycosylation preference of the GalNAc-Ts, and reveals the  
504 importance of the flexible linker in fine-tuning their glycosylation  
505 preference.
- 506 7. Gerken TA, Jamison O, Perrine CL, Collette JC, Moinova H, Ravi L,  
507 •• Markowitz SD, Shen W, Patel H, Tabak LA: **Emerging paradigms**  
508 **for the initiation of mucin-type protein O-glycosylation by the**  
**polypeptide GalNAc transferase family of**  
**glycosyltransferases.** *J Biol Chem* 2011, **286**:14493-14507.
- 509 In this article, the authors makes a thorough study into the GalNAc-Ts  
510 substrate glycosylation preferences, and conclude that each isoenzyme  
511 may be uniquely sensitive to sequence and overall charge.
- 512 8. Steentoft C, Vakhrushev SY, Vester-Christensen MB,  
513 •• Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H,  
514 Lavery SB, Clausen H: **Mining the O-glycoproteome using zinc-**  
515 **finger nuclease-glycoengineered simplecell lines.** *Nat*  
516 *Methods* 2011, **8**:977-982.
- 517 This article describes for the first time the use of zinc-finger nuclease  
518 (ZFN) gene targeting as a tool to facilitate analyses of important functions  
519 of protein glycosylation.
- 520 9. Schjoldager KT, Vakhrushev SY, Kong Y, Steentoft C,  
521 Nudelman AS, Pedersen NB, Wandall HH, Mandel U, Bennett EP,  
522 Lavery SB, Clausen H: **Probing isoform-specific functions of**  
523 **polypeptide GalNAc-transferases using zinc finger nuclease**  
524 **glycoengineered simplecells.** *Proc Natl Acad Sci U S A* 2012,  
**109**:9893-9898.
- 525 10. Pedersen NB, Wang S, Narimatsu Y, Yang Z, Halim A,  
526 Schjoldager KT, Madsen TD, Seidah NG, Bennett EP, Lavery SB,  
527 Clausen H: **Low-density lipoprotein receptor class A repeats**  
528 **are O-glycosylated in linker regions.** *J Biol Chem* 2014 <http://dx.doi.org/10.1074/jbc.M113.545053>.
- 529 11. Wang S, Mao Y, Narimatsu Y, Ye Z, Tian W, Goth CK, Lira-  
530 Navarrete E, Pedersen NB, Benito-Vicente A, Martin C et al.: **Site-**  
531 **specific O-glycosylation of members of the low-density**  
532 **lipoprotein receptor superfamily enhances ligand**  
533 **interactions.** *J Biol Chem* 2018, **293**:7408-7422.
- 534 12. Schjoldager KT, Vester-Christensen MB, Goth CK, Petersen TN,  
535 Brunak S, Bennett EP, Lavery SB, Clausen H: **A systematic study**  
536 **of site-specific GalNAc-type O-glycosylation modulating**  
537 **proprotein convertase processing.** *J Biol Chem* 2011,  
538 **286**:40122-40132.
- 539 13. Kato K, Jeanneau C, Tarp MA, Benet-Pagès A, Lorenz-  
540 Depiereux B, Bennett EP, Mandel U, Strom TM, Clausen H:  
541 **Polypeptide GalNAc-transferase T3 and familial tumoral**  
542 **calcinosis secretion of fibroblast growth factor 23 requires O-**  
543 **glycosylation.** *J Biol Chem* 2006, **281**:18370-18377.
- 544 14. Schjoldager KT, Vester-Christensen MB, Bennett EP, Lavery SB,  
545 Schwientek T, Yin W, Blixt O, Clausen H: **O-Glycosylation**  
546 **modulates proprotein convertase activation of angiotensin-**  
547 **like protein 3-possible role of polypeptide GalNAc-**  
548 **transferase-2 in regulation of concentrations of plasma**  
549 **lipids.** *J Biol Chem* 2010 <http://dx.doi.org/10.1074/jbc.M110.156950>.
- 550 15. Fu C, Zhao H, Wang Y, Cai H, Xiao Y, Zeng Y, Chen H: **Tumor-**  
551 **associated antigens: Tn antigen, STn antigen, and T antigen.**  
552 *HLA* 2016, **88**:275-286.
- 553 16. Sletmoen M, Gerken TA, Stokke BT, Burchell J, Brewer CF: **Tn and**  
554 **STn are members of a family of carbohydrate tumor antigens**  
555 **that possess carbohydrate-carbohydrate interactions.**  
556 *Glycobiology* 2018, **28**:437-442.
- 557 This interesting work provides a model for the possible roles of Tn and  
558 STn antigens in cancer and other diseases.
- 559 17. Radhakrishnan P, Dabelsteen S, Madsen FB, Francavilla C,  
560 Kopp KL, Steentoft C, Vakhrushev SY, Olsen JV, Hansen L,  
561 Bennett EP et al.: **Immature truncated O-glycophenotype of**  
562 **cancer directly induces oncogenic features.** *Proc Natl Acad Sci*  
563 *U S A* 2014, **111**:4066-4075.
- 564 18. Pearce OMT: **Cancer glycan epitopes: biosynthesis, structure**  
565 **and function.** *Glycobiology* 2018, **28**:670-696.
- 566 19. Fritz TA, Hurley JH, Trinh L-B, Shiloach J, Tabak LA: **The**  
567 **beginnings of mucin biosynthesis: the crystal structure of**  
568 **UDP-GalNAc: polypeptide  $\alpha$ -N-**  
569 **acetylglucosaminyltransferase-1.** *Proc Natl Acad Sci U S A*  
570 2004, **101**:15307-15312.
- 571 This work provides the first crystal structure of a member of GalNAc-Ts  
572 family: *MmGalNAc-T1*.
- 573 20. Fritz TA, Raman J, Tabak LA: **Dynamic association between the**  
574 **catalytic and lectin domains of human UDP-GalNAc:**  
575 **polypeptide  $\alpha$ -N-acetylglucosaminyltransferase-2.** *J Biol*  
576 *Chem* 2006, **281**:8613-8619.
- 577 In this article the authors solve the first ternary complex of *HsGalNAc-T2*  
578 in complex with UDP and an acceptor peptide. This resolves partly the  
579 long-standing question of how GalNAc-Ts recognise 'naked' peptide  
580 substrates.
- 581 21. Lira-Navarrete E, de Las Rivas M, Compañón I, Pallarés MC,  
582 •• Kong Y, Iglesias-Fernández J, Bernardes GJ, Peregrina JM,  
583 Rovira C, Bernadó P et al.: **Dynamic interplay between catalytic**  
584 **and lectin domains of GalNAc-transferases modulates protein**  
585 **o-glycosylation.** *Nat Commun* 2015, **6**:1-11.
- 586 This article describes the first ternary complex between a GalNAc-T  
587 (GalNAc-T2) and a glycopeptide substrate and UDP-Mn<sup>2+</sup>, and shows  
588 the influence of the mechanical properties of the flexible linker between  
589 the catalytic and the lectin domains to account for its observed glyco-  
590 sylation profile.
- 591 22. Kubota T, Shiba T, Sugioka S, Furukawa S, Sawaki H, Kato R,  
592 •• Wakatsuki S, Narimatsu H: **Structural basis of carbohydrate**  
593 **transfer activity by human UDP-GalNAc: polypeptide  $\alpha$ -N-**  
594 **acetylglucosaminyltransferase (pp-GalNAc-T10).** *J Mol Biol*  
595 2006, **359**:708-727.
- 596 This work provides the first crystal structure of *HsGalNAc-T10* in complex  
597 with Ser-GalNAc that reveals for the first time how the lectin domain  
598 recognises the GalNAc moiety of glycopeptides.
- 599 23. de las Rivas M, Paul Daniel EJ, Coelho H, Lira-Navarrete E,  
600 •• Raich L, Compañón I, Diniz A, Lagartera L, Jiménez-Barbero J,  
601 Clausen H et al.: **Structural and mechanistic insights**  
602 **into the catalytic-domain-mediated short range**  
603 **glycosylation preferences of GalNAc-T4.** *ACS Cent Sci* 2018,  
604 **4**:1274-1290.
- 605 This recent work provides the first structural basis for the short-range  
606 glycosylation preferences of a GalNAc-T, the GalNAc-T4, and describes a

## 10 Carbohydrates

- 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 specificity within the O-glycosyltransferase enzyme family, based on the charge of a loop that controls the access 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 ppGalNAc-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 HsGalNAc-T2, supporting a front face S<sub>N</sub>i 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 ppGalNAc-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 HsGalNAc-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-González 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, Lavery 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 (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* 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 Natl 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, Lavery 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-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferases.** *Glycobiology* 2012, **22**:602–615.