

# CHEMISTRY

## A European Journal

A Journal of



### Accepted Article

**Title:** Structural analysis of a GalNAc-T2 mutant reveals an induced-fit catalytic mechanism for GalNAc-Ts

**Authors:** Ramon Hurtado-Guerrero, Matilde de Las Rivas, Helena Coelho, Ana Diniz, Erandi Lira-Navarrete, Ismael Compañón, Jesús Jiménez-Barbero, Katrine T. Schjoldager, Eric P. Bennett, Sergey Y. Vakhrushev, Henrik Clausen, Francisco Corzana, and Filipa Marcelo

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *Chem. Eur. J.* 10.1002/chem.201800701

**Link to VoR:** <http://dx.doi.org/10.1002/chem.201800701>

Supported by  
**ACES**

WILEY-VCH

# Structural analysis of a GalNAc-T2 mutant reveals an induced-fit catalytic mechanism for GalNAc-Ts

Matilde de las Rivas,<sup>[a]</sup> Helena Coelho,<sup>[b],[c]</sup> Ana Diniz,<sup>[b]</sup> Erandi Lira-Navarrete,<sup>[d]</sup> Ismael Compañón,<sup>[e]</sup> Jesús Jiménez-Barbero,<sup>[c],[f]</sup> Katrine. T. Schjoldager,<sup>[d]</sup> Eric. P. Bennett,<sup>[d]</sup> Sergey Y. Vakhrushev,<sup>[d]</sup> Henrik Clausen,<sup>[d]</sup> Francisco Corzana,<sup>[e]</sup> Filipa Marcelo,<sup>\*,[b]</sup> and Ramon Hurtado-Guerrero<sup>\*,[a],[g]</sup>

**Abstract:** The family of polypeptide GalNAc-transferases (GalNAc-Ts) orchestrates the initiating step of mucin-type protein O-glycosylation by transfer of GalNAc moieties to serine and threonine residues in proteins. Deficiencies and dysregulation of GalNAc-T isoenzymes have been found to be related to different diseases. Recently, we have demonstrated that an inactive GalNAc-T2 mutant (F104S), which is not located at the active site, induces low levels of high-density lipoprotein cholesterol (HDL-C) in humans. Here, we have deciphered the molecular basis for F104S mutant inactivation. Saturation transfer difference NMR experiments demonstrate that the mutation induces loss of binding to peptide substrates. The analysis of the crystal structure of the F104S mutant bound to UDP-GalNAc, combined with molecular dynamics (MD) simulations, has revealed that the flexible loop is disordered and displays larger conformational changes in the mutant enzyme than in the wild-type (WT) enzyme. <sup>19</sup>F-NMR experiments reveal that the WT enzyme reaches the active state only in the presence of UDP-GalNAc, providing compelling evidences that GalNAc-T2 adopts an UDP-GalNAc-dependent induced-fit mechanism. The F104S mutation precludes the enzyme to achieve the active conformation and concomitantly to bind peptide substrates. The present study provides new insights into the catalytic mechanism of the large family of GalNAc-Ts and how these enzymes orchestrate protein O-

glycosylation.

## Introduction

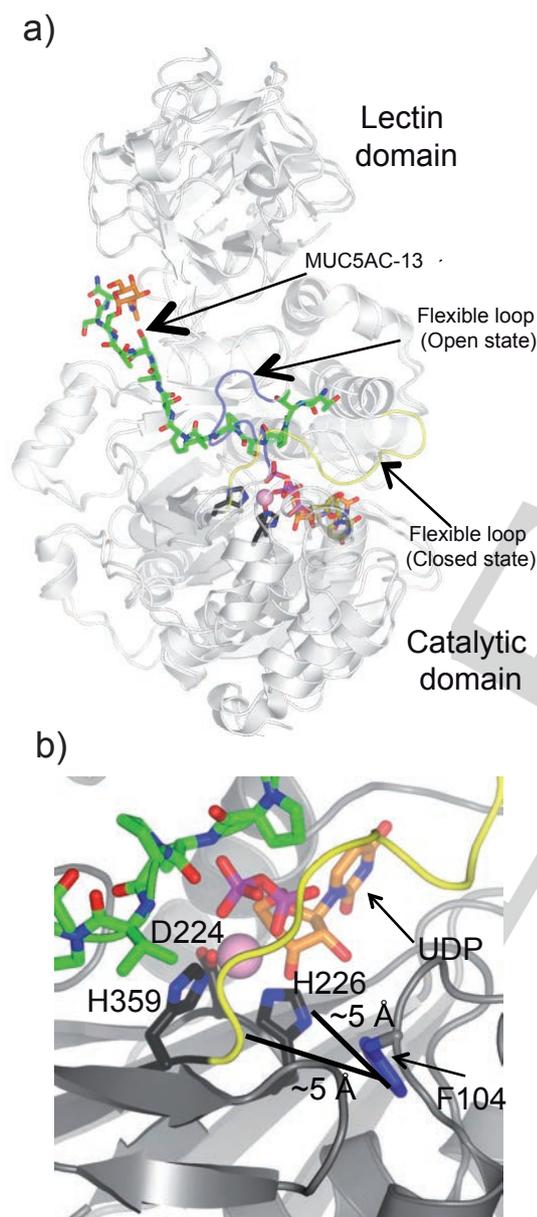
Mucin-type (GalNAc-type) O-glycosylation is the most differentially and complex regulated type of protein glycosylation, and likely the most abundant, with over 80% of all proteins predicted to be O-glycosylated.<sup>[1]</sup> This post-translational modification is initiated by a large family (20 in humans) of enzymes named polypeptide N-acetylgalactosaminyltransferases (GalNAc-Ts), which catalyze the attachment of a GalNAc residue of the substrate UDP-GalNAc to Ser/Thr residues of the peptide backbone.

Although these enzymes have a partial overlap in recognition and glycosylation of substrates, a subset of glycosylation sites is subjected to site-specific O-glycosylation by individual GalNAc-Ts. For instance, GalNAc-T3 is responsible for the site-specific glycosylation of Thr178 of FGF23, a major regulator of phosphate homeostasis<sup>[2]</sup>. This site-specific glycosylation of Thr178 prevents the inactivation of FGF23 by proprotein convertase proteases<sup>[3]</sup>, thus mutations in GalNAc-T3 leading to a decrease or a complete abolishment of its activity are known to cause familial tumoral calcinosis.<sup>[3]</sup> By a similar mechanism, we have recently demonstrated that the loss of GalNAc-T2 function lowers high-density lipoprotein cholesterol (HDL-C) levels in humans.<sup>[4]</sup> We also identified two independent deficiencies in the *GALNT2* gene, and one of these involved the homozygous missense mutation F104S<sup>[4]</sup>, which was not predicted to be deleterious based on conservation and prior structural models of the enzyme.<sup>[5]</sup> Those examples illustrate how GalNAc-T isoforms, with seemingly many roles in protein glycosylation, have in fact limited non-redundant biological functions.<sup>[3-4]</sup>

From a structural perspective, GalNAc-Ts are unique among glycosyltransferases because in addition to their N-terminal catalytic domain, they also possess a unique C-terminal lectin domain with both domains being connected by a short flexible linker (Figure 1a and Figure S1). There is also an interesting structural feature in the catalytic domains of these enzymes, a so-called flexible loop, which comprises residues Val360 to Gly372 in GalNAc-T2. This loop may adopt different conformations during the catalytic cycle, thus rendering the enzyme catalytically inactive or active.<sup>[5-6]</sup> Furthermore, the flexible loop dynamics are coupled to the key catalytic residue Trp331 mobility that can adopt 'in' (inside of the active site) and 'out' (outside of the active site) conformations, which in turn are also associated to the active or inactive states of GalNAc-T2,

- [a] Mrs. M de las Rivas, Dr. R. Hurtado-Guerrero  
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), BIFI-IQFR (CSIC) Joint Unit, Universidad de Zaragoza, 50018 Zaragoza, Spain. E-mail: rhurtado@bifi.es
- [b] Mrs. H. Coelho, Mrs. A. Diniz, Dr. F. Marcelo  
UCBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, 2829-516 Caparica, Portugal. E-mail: filipa.marcelo@fct.unl.pt
- [c] Mrs. H. Coelho, Prof. J. Jiménez-Barbero  
CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain; Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Bizkaia, Spain
- [d] Mrs. E. Lira-Navarrete, Dr. K. T. Schjoldager, Dr. E. P. Bennett, Dr. S. Y. Vakhrushev, Prof. H. Clausen  
Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen, Denmark.
- [e] Mr. I. Compañón, Dr. F. Corzana  
Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, E-26006 Logroño, Spain.
- [f] Prof. J. Jiménez-Barbero  
Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain.
- [g] Dr. R. Hurtado-Guerrero  
Fundación ARAID, 50018, Zaragoza, Spain.
- [†] Mrs. M de las Rivas and H. Coelho contributed equally to this work  
Supporting information for this article is given via a link at the end of the document.

respectively.<sup>[5-6]</sup> Crystal structures of GalNAc-Ts can be used as a platform to infer how mutations on these enzymes might lead to loss-of-function.<sup>[5-7]</sup> However, this is not always straightforward and further experimental work in solution to infer the dynamic behavior of the global system is required. The F104S mutant in GalNAc-T2 precisely illustrates this idea since Phe104 is not located at the active site, but placed at a distance of  $\sim 5$  Å from the flexible loop and the critical  $Mn^{+2}$  binding site, which is formed by the  $D_{224}xH_{226}$  residues and His359 (Figure 1b).<sup>[5]</sup> Thus, it is not obvious how mutation of Phe to Ser impairs GalNAc-T2 catalytic activity.



**Figure 1.** States of the flexible loop and location of Phe104: a) Superposition of the overall crystal structure of GalNAc-T2 in complex with UDP and the glycopeptide MUC5AC-13<sup>[5]</sup> (closed conformation shown in white color; see

PDB entry 5AJP), with the crystal structure of GalNAc-T2 in complex with UDP<sup>[7b]</sup> (open conformation also shown in white color; see PDB entry 2FFV). The GalNAc moiety of MUC5AC-13 is colored in orange. UDP and MUC5AC-13 are depicted as orange and green carbon atoms, respectively. The  $Mn^{+2}$  binding  $D_{224}xH_{226}$  motif and His359 are shown as black carbon atoms. The flexible loop is depicted in blue (open conformation) and yellow (closed conformation) for the PDB entries 2FFV and 5AJP, respectively.  $Mn^{+2}$  is shown as a pink sphere. b) Close-up view of the active site (closed conformation; PDB entry 5AJP) showing the distance of Phe104 (in blue) to the flexible loop (in yellow) and the His226 (in black). The backbone structure is shown in grey while the other structural features and amino acids are shown in the same colors as in panel a.

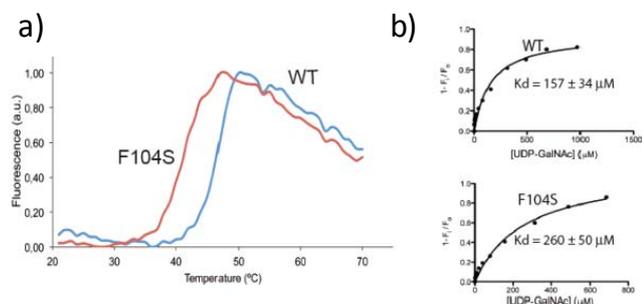
Herein, we show that the F104S mutation in GalNAc-T2 constrains the flexible loop and renders the enzyme locked in an inactive state unable to bind the acceptor peptide substrate. We have combined saturation transfer difference NMR (STD-NMR), <sup>19</sup>F-NMR spectroscopy, X-ray crystallography and molecular dynamics simulations to demonstrate that Phe104 fine-tunes the dynamic behavior of the flexible loop by establishing a CH- $\pi$  interaction with the side-chain of Arg362, which is located at the flexible loop. This interaction appears to be the key to maintain the flexible loop in the required closed conformation that leads to the active state of the enzyme. Therefore, the inability of the mutant to adopt the active state disrupts the peptide substrate recognition and prevents O-glycosylation. Finally, we have experimentally demonstrated that GalNAc-T2 adopts an UDP-GalNAc-dependent induced-fit mechanism for catalysis and suggest that the behavior of the flexible loop can be extrapolated to the rest of the members of the GalNAc-T family.

## Results and Discussion

### The GalNAc-T2 F104S mutant does not bind to peptides.

The available data based on kinetic assays of the F104S mutant cannot explain the molecular basis of the inactivation of the GalNAc-T2 mutant F104S against protein substrates.<sup>[4]</sup> Since Phe104 is not located at the active site (Figure 1b), its mutation to Ser must induce conformational changes of the residues located in the active site, affecting either the binding of the mutant to the substrates or the catalytic turnover. To rule out one of these two hypotheses, we have determined the binding of the GalNAc-T2 WT enzyme and the F104S mutant to UDP-GalNAc and a peptide substrate applying tryptophan fluorescence spectroscopy and STD-NMR experiments, respectively. In addition, we have also conducted a thermal shift-assay experiment on both enzymes to evaluate the differences in the protein stability.<sup>[6]</sup> The data in Figure 2 shows that the WT enzyme is more stable than the F104S mutant (denaturation temperatures of  $T_0$  of  $47 \pm 0.09$  °C and  $42 \pm 0.05$  °C, respectively), implying that the mutation leads to a significant decrease on the stability of the mutant (Figure 2a). The tryptophan fluorescence spectroscopy analysis of UDP-GalNAc binding revealed that both the WT enzyme and the F104S mutant bind to the donor sugar nucleotide, with the  $K_d$  of UDP-GalNAc to the mutant being 1.7-fold higher than the  $K_d$  determined for the WT enzyme (Figure 2b and Figure S2). In contrast, STD-NMR experiments recorded for the peptide

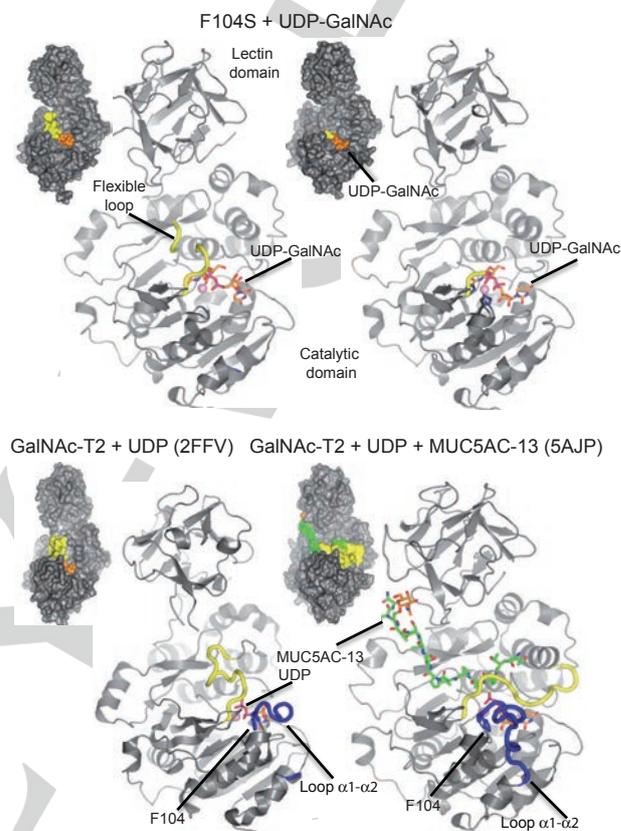
substrate MUC5AC (GTPSPVPTTSTTSA) only showed WT GalNAc-T2 binding to the MUC5AC peptide (Figure S3). No STD signal response was observed for MUC5AC in presence of the mutant F104S (Figure S3 and Table S1). This result points out that the mutant is not able to properly recognize and bind the peptide substrate, thus explaining the previously observed inactivity of the mutant towards protein substrates such as phospholipid transfer protein (PLTP) and ApoC-III<sup>[4]</sup>. In addition, the WT enzyme showed clear STD-NMR signals for the protons of Valine and Proline amino acids from the sequence "PSPVPT" (where the underlined Thr is the main acceptor site on this peptide for GalNAc-T2) than for the rest of peptide sequence (Figure S4). The identified STD-NMR epitope is consistent with the sequence preferences of GalNAc-T2, particularly having a Pro residue preceding the site of glycosylation<sup>[9]</sup>. Hence, while there are no major differences in UDP-GalNAc recognition between the WT enzyme and the F104S mutant, the dramatic differences in activity are accounted for either the inability or poor affinity of the F104S mutant to bind the peptide substrates.



**Figure 2.** Stability and binding properties of the WT enzyme and the F104S mutant: a) Thermal denaturation curves of the WT enzyme and the F104S mutant as monitored by 1-anilino-8-naphthalene sulfonate (ANS) fluorescence. b)  $K_d$  values for sugar nucleotide binding determined from the tryptophan fluorescence intensity as a function of UDP-GalNAc concentrations.

**Phe104 controls the inactive-to-active state transition.** To elucidate the molecular basis for the loss of acceptor peptide substrate binding induced by the F104S mutation, we obtained tetragonal crystals of this mutant in complex with UDP that were subsequently soaked with UDP-GalNAc. The resulting crystals allowed us to solve the structure at 2.70 Å resolution (Table S2).

Within the asymmetric unit, two molecules of the F104S mutant were present. The crystal structure shows a compact structure with the typical N-terminal GT-A fold and a C-terminal lectin domain (Figure 3).

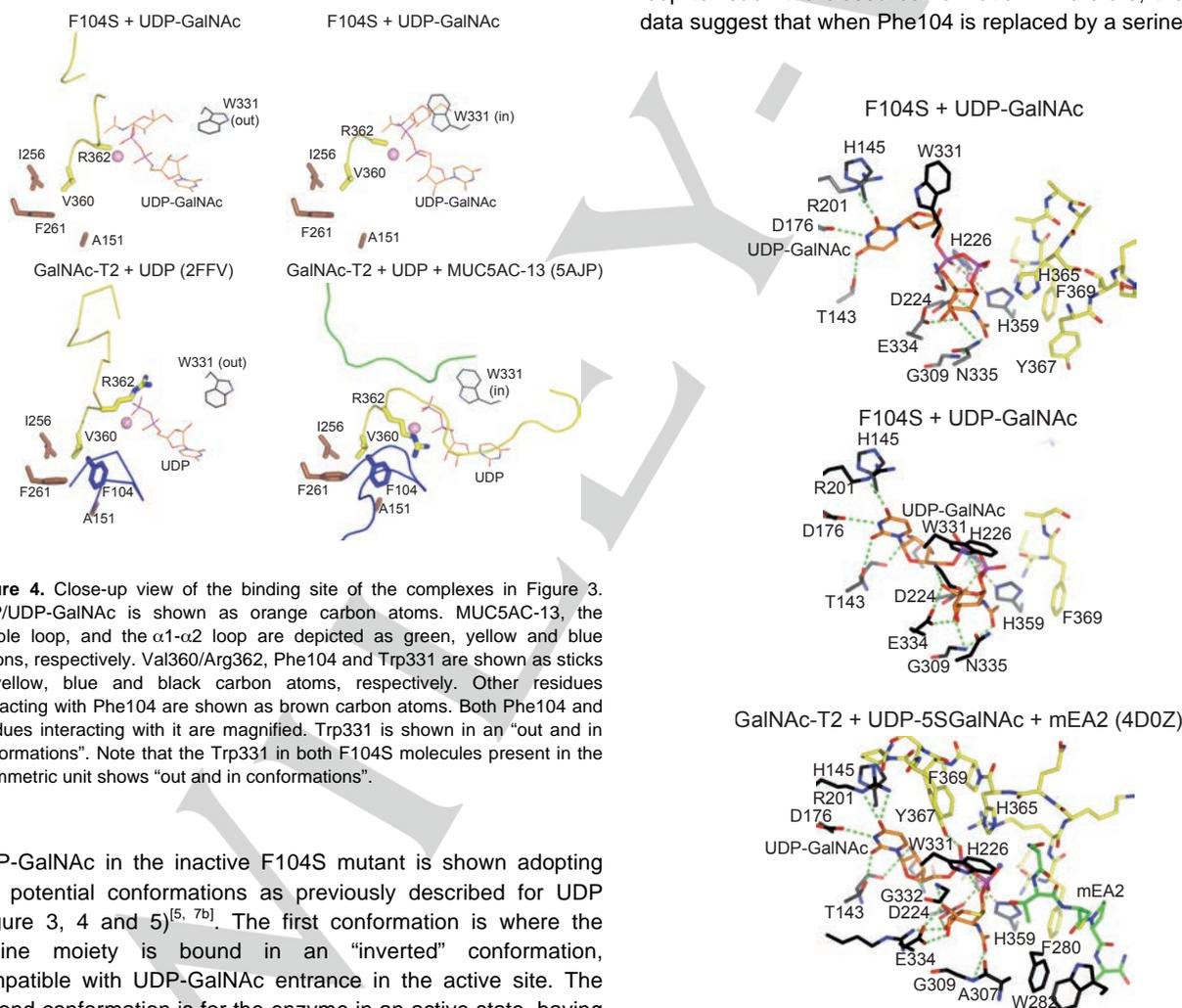


**Figure 3.** Overall crystal structure of the F104S mutant bound to UDP-GalNAc: Cartoon representation of the overall F104S mutant in complex with UDP-GalNAc and the WT enzyme either in complex with UDP<sup>[7b]</sup> (PDB entry 2FFV) or UDP/MUC5AC-13<sup>[5]</sup> (PDB entry 5AJJ). The proteins and the GalNAc moiety of MUC5AC-13 are colored in grey and orange, respectively. UDP/UDP-GalNAc is shown as orange carbon atoms. MUC5AC-13, the flexible loop, and the  $\alpha 1$ - $\alpha 2$  loop are depicted as green, yellow and blue ribbons, respectively. The Phe104 is shown as sticks in blue.  $Mn^{+2}$  is depicted as a pink sphere. (insert) Surface representation of the above complexes showing inactive and active states. Colours are the same as above. Note that we do present the overall structure of both F104S molecules because there are some differences between them. This is because the asymmetric unit contains two molecules of the F104S mutant.

The flexible loop in both F104S molecules of the asymmetric unit is highly disordered (Figure 3, 4, 5 and Figure S5), exposing the substrate UDP-GalNAc to the solvent. In these structures the flexible loop displays the open conformation that restrains the enzyme in an inactive state, resembling the structure of the WT enzyme in its inactive state (see PDB entry 2FFV; Figure 3, 4 and 5). This conformation is opposed to that presenting the active state. In the active state the flexible loop displays a closed conformation that functions to cover UDP-GalNAc or UDP from the solvent (e.g. see PDB entries 5AJP, 4D0T and 4D0Z; Figure 3). This geometrical arrangement is required for the binding of the peptide/protein substrates<sup>[5, 6c]</sup>. This closed conformation of the flexible loop is further stabilized by interactions of His365 and Phe369 with Trp331 (in-conformation) and by additional interactions with the donor substrate UDP-GalNAc (Figure 5)<sup>[6c]</sup>. In fact, the fully active state can only be reached when both the flexible loop and Trp331 adopt the closed geometry and the “in-conformation”, respectively.

although this geometry is likely related to product release (Figure 4).

A detailed analysis of the F104S mutant structure reveals a lack of density for Ser104 and for most of the residues of the loop between helical  $\alpha 1$  and  $\alpha 2$  (named hereafter  $\alpha 1$ - $\alpha 2$  loop). In the WT enzyme, Phe104 establishes hydrophobic and stacking interactions with Ala151/Ile256/Val360 and Phe261, likely stabilizing the location of both the  $\alpha 1$ - $\alpha 2$  and the flexible loops, in the inactive state (Figure 4). The disordered states of both loops in the inactive mutant might also explain why this structure is less stable than the WT enzyme (see above). In the active state of the enzyme, Phe104 also establishes a CH- $\pi$  interaction with the side-chain of Arg362 located in the flexible loop (Figure 4). This stabilizing interaction appears to be a key feature rendering the closed conformation of the flexible loop. In contrast, the polar Ser104 residue likely disrupts the interactions with the hydrophobic/aromatic residues, leading to the destabilization of the  $\alpha 1$ - $\alpha 2$  loop and the inability for the flexible loop to reach the closed conformation. Therefore, the obtained data suggest that when Phe104 is replaced by a serine residue



**Figure 4.** Close-up view of the binding site of the complexes in Figure 3. UDP/UDP-GalNAc is shown as orange carbon atoms. MUC5AC-13, the flexible loop, and the  $\alpha 1$ - $\alpha 2$  loop are depicted as green, yellow and blue ribbons, respectively. Val360/Arg362, Phe104 and Trp331 are shown as sticks in yellow, blue and black carbon atoms, respectively. Other residues interacting with Phe104 are shown as brown carbon atoms. Both Phe104 and residues interacting with it are magnified. Trp331 is shown in an “out and in conformations”. Note that the Trp331 in both F104S molecules present in the asymmetric unit shows “out and in conformations”.

UDP-GalNAc in the inactive F104S mutant is shown adopting two potential conformations as previously described for UDP (Figure 3, 4 and 5)<sup>[5, 7b]</sup>. The first conformation is where the uridine moiety is bound in an “inverted” conformation, compatible with UDP-GalNAc entrance in the active site. The second conformation is for the enzyme in an active state, having a conformation suited for catalysis (see also PDB entry 5AJP; Figure 4). It should also be noted that, the crystal structure of GalNAc-T2 in complex with UDP shown in the PDB entry 2FFV, shows the uridine moiety in the “inverted” conformation,

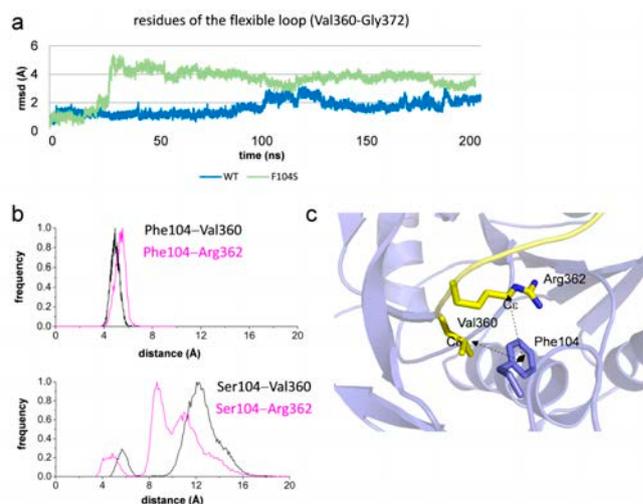
**Figure 5.** Close-up view of the sugar nucleotide and peptide binding site of the F104S-UDP-GalNAc and WT-UDP-5S-GalNAc-mEA2 complexes. Residues

and the sugar nucleotides are depicted with the same colors as shown in Figure 3 and 4. Hydrogen bond interactions are shown as dotted green lines.

(Ser104), the enzyme is not able to reach the active state, even in the presence of UDP-GalNAc due to the lack of the essential Phe104/Arg362 CH- $\pi$  interaction, which provides the additional stabilization required to induce the closed conformation of the flexible loop. Interestingly, most of the residues interacting with Phe104, Ile256/Val360/Phe261/Arg362, are highly conserved among different GalNAc-Ts (Figure S6). This fact suggests that the flexible loop behavior and the corresponding activation mechanism via the Phe104/Arg362 CH- $\pi$  interaction is conserved across most GalNAc-T isoenzymes.

Regarding the UDP-GalNAc recognition, the binding sites of both the F104S mutant and the WT enzyme in the active form show almost equivalent interactions, explaining the similar  $K_d$  values obtained for this sugar nucleotide (Figure 2b). The interactions with UDP-GalNAc are mostly based on hydrogen bonds with the surrounding protein residues, whereas the pyrophosphate is involved in the coordination of the  $Mn^{2+}$  ion, together with the residues of the  $D_{224}XH_{226}$  motif and His359 (Figure 5). Minor additional interactions are found for the WT enzyme in the active form (e.g. hydrogen bond interactions with Gly332 and Tyr367; Figure 5), which is in agreement with the slightly better affinity for UDP-GalNAc on the WT enzyme versus the mutant (1.7-fold).

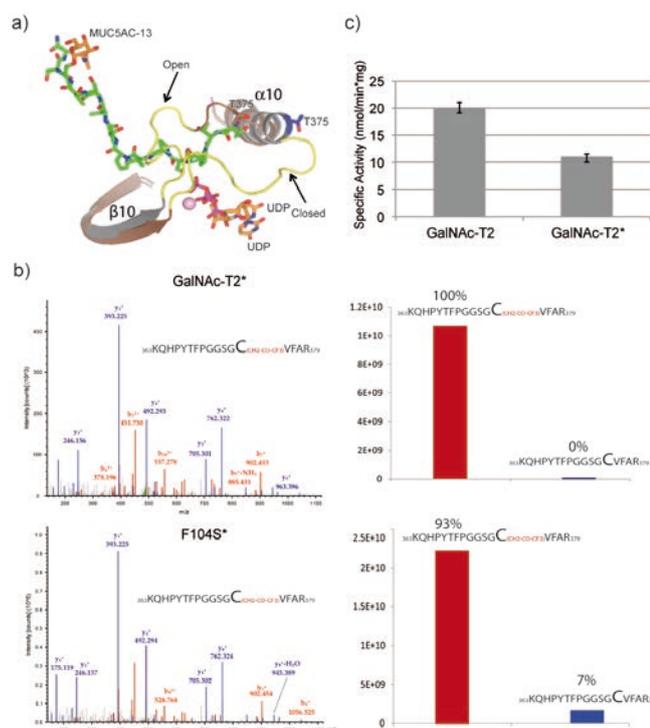
**The flexible loop in the F104S mutant displays large conformational changes.** To understand how the mutation F104S disturbs the dynamics of the flexible loop, we ran 200 ns molecular dynamics (MD) simulations in explicit water on both the WT enzyme and the F104S mutant in the presence of UDP-GalNAc. The data revealed that the flexible loop of the mutant displayed significantly larger conformational changes, with root-mean-square-deviation (RMSD) ranging between 1-5.3 Å, as compared to 1-3 Å deviations for the WT enzyme (Figure 6a). This trend was also shown in the variation of the  $C_{\alpha}$  distances between the Phe104/Ser104 residues and Val360. While a shorter distance of  $5.7 \pm 1.4$  Å was found for Phe104-Val360, highly variable and larger distances (ranging from 4.5 to 16 Å) were found for Ser104-Val360 (Figure 6b and 6c). We also determined the distance distributions for Arg362/Phe104, and Arg362/Ser104- $C_{\alpha}$  pairs (Figure 6b). The average distance of  $4.7 \pm 1.4$  Å found for the WT enzyme suggests the presence of stabilizing hydrophobic interactions between the aromatic ring of Phe104 and the side-chain of Val360. In contrast, two populations were found for the mutant: a minor population, with distances ranging between 3.5-7 Å, and a major population, with highly oscillating distances, between 7.5-17 Å (Figure 6b). These results also point out that the flexible loop and the  $\alpha 1$ - $\alpha 2$  loop are much more structured in the WT enzyme than in the mutant and that these structural entities remain at a closer distance in the WT enzyme than in the mutant. Therefore, according to the MD analysis, the closed conformation of the flexible loop in the inactive mutant would rarely form.



**Figure 6.** Molecular dynamics simulations on the WT enzyme and the F104S mutant: a) The RMSD (root-mean-square deviation) values calculated from 200 ns molecular dynamics (MD) simulations for the flexible loop indicates that this structural motif is more flexible in the mutant (green plot) than in the WT enzyme (blue plot). Note that the RMSD was obtained using the fixed coordinates of the flexible loop from the PDB entry 4DOT. b) This is also evidenced by larger distance distribution between Ser104 ( $C_{\alpha}$ ) and Val360 ( $C_{\delta}$ )/Arg362 ( $C_{\epsilon}$ ) in the mutant (bottom) compared to the corresponding distances (Phe104-Val360 and Phe104-Arg362) found in the WT enzyme (top). c) The center of mass of the  $\pi$ -electron system of Phe104 was used to calculate the distances.

**$^{19}F$  labeling of the T375C and F104S-T375C variants for NMR experiments.** To experimentally monitor the dynamics of flexible loop, we also performed  $^{19}F$ -NMR spectroscopy experiments on the WT enzyme and on the F104S mutant.  $^{19}F$  chemical shifts are highly sensitive to changes in the local conformational environment and provide a simple and well-established approach for studying biomolecular structure, conformation, and dynamics.<sup>[10]</sup> To introduce the required  $^{19}F$  label, Thr375 was mutated to the more nucleophilic Cys residue, in both the WT enzyme and the F104S mutant. This residue was selected for several reasons: i) it is exposed and accessible to be modified and WT GalNAc-T2 and the F104S mutant contain only two free buried cysteines at the active site; ii) it is fairly close to the flexible loop but sufficiently far away from the  $Mn^{2+}$  ion to minimize strong paramagnetic relaxation enhancements; and iii) the crystallographic analyses indicates that Thr375 would be sensitive to the flexible loop dynamics and may adopt different conformations depending on the geometry of the flexible loop (open or closed). It should be noted that Thr375 in the PDB entry 2FFV is 10 Å away from the same residue in PDB entry 5AJP (Figure 7a). Thus, both the WT enzyme and the F104S mutant variants containing the T375C mutation were produced (see the experimental section). The  $^{19}F$  label was subsequently introduced into the T375C and the F104S-T375C variants using 3-bromo-1,1,1-trifluoroacetone (BFA) (see the experimental section).<sup>[10]</sup> A single  $SCH_2(CO)CF_3$  adduct was obtained for each enzymes (GalNAc-T2\* and F104S\*) as shown by trypsin digestion and MS fragmentation studies (Figure 7b). While the

GalNAc-T2\* was 100% labeled, the F104S\* was only 93% modified (Figure 7b). The activity of GalNAc-T2\* was evaluated and compared with that obtained for the WT enzyme (Figure 7c and the experimental section). GalNAc-T2\* still remained active giving ~50% activity against MUC1 peptide compared to the WT enzyme.



**Figure 7.** Labeling of the T375C and F104S-T375C variants with BFA: a) Close-up view of the active sites of the PDB entries 2FFV and 5AJP, which are in an inactive and active states, respectively. The flexible loop adopting an open (2FFV) and closed conformation (5AJP) is colored in yellow. The MUC5AC-13 is depicted as sticks in green carbon atoms whereas the GalNAc-moiety is in orange carbon atoms. UDP is colored in orange. The  $\alpha$  and  $\beta$  secondary structures are brown and grey for the PDB entries 2FFV and 5AJP, respectively. Thr375 is depicted as sticks as brown (2FFV) and blue (5AJP) carbon atoms. Note that DSSP<sup>[11]</sup> extracts from the 3D structure that Thr375 in the PDB entry 2FFV is part of the flexible loop whereas in the PDB entry 5AJP is located in  $\alpha 10$ . b) MS/MS spectrum of the precursor ions at  $m/z$  654.6346,  $z=3+$  assigned to  ${}_{363}\text{KQHYPYTFPGGSGCVFAR}_{379}$  peptide containing the  ${}^{19}\text{F}$  label at the position C375 (left) and relative quantification based on the extracted ion chromatogram between peptides with and without the  ${}^{19}\text{F}$  label at the position C375 (right). The experiments were performed for the GalNAc-T2\* (top) and F104S\* (bottom). c) Specific activity of the WT enzyme and GalNAc-T2\* using the MUC1 peptide as acceptor substrate. The data represent means  $\pm$  S.D. for three independent experiments.

**${}^{19}\text{F}$  NMR experiments reveal that GalNAc-T2 adopts an UDP-GalNAc dependent induced-fit mechanism.** With both WT and mutant proteins labeled with BFA, we first screened enzymes activity against MUC5AC by  ${}^1\text{H}$ -NMR experiments confirming that only GalNAc-T2\* and not the F104S\* mutant was able to glycosylate MUC5AC (Figure S7 and S8).

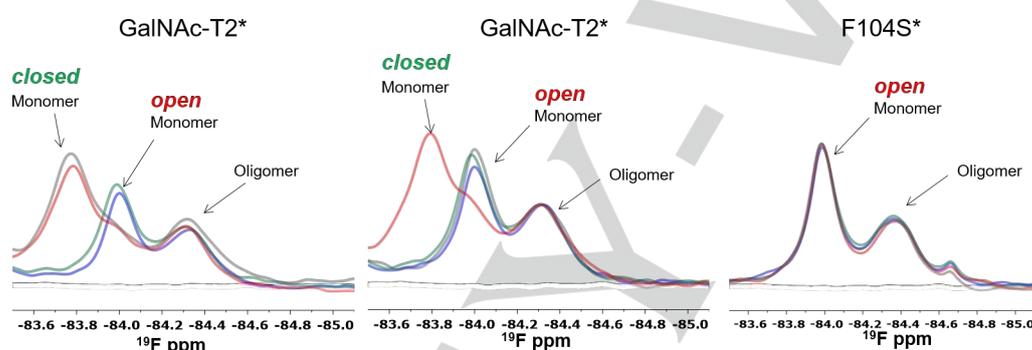
To evaluate whether the Cys375  ${}^{19}\text{F}$  label was sensitive to the dynamics of the flexible loop upon substrate binding, we firstly performed  ${}^{19}\text{F}$ -NMR experiments on GalNAcT2\*. A set of experiments using distinct combinations of UDP-GalNAc,  $\text{MnCl}_2$  and MUC5AC were used (Figure 8 left and middle panel and Figures S9 and S10 in SI). The  ${}^{19}\text{F}$ -NMR spectra of unbound GalNAcT2\* (Figure 8, left and middle panels blue tracings) showed two  ${}^{19}\text{F}$  resonances at  $\delta = -84.0$  ppm and  $\delta = -84.3$  ppm. The upfield  ${}^{19}\text{F}$  signal at  $\delta = -84.3$  ppm remained unperturbed for all combinations of added substrates, which we attribute to the oligomeric GalNAc-T2\* state. The linewidth of the upfield oligomer  ${}^{19}\text{F}$  peak ( $\delta = -84.3$  ppm) is 161 Hz, significantly larger than that observed for the monomer at  $\delta = -84.0$  ppm (111 Hz), which is consistent with a reduced relaxation time of the oligomeric enzyme<sup>[12]</sup>. We have previously demonstrated by small-angle X-ray scattering experiments that GalNAc-T2 forms oligomers that are not fully catalytically active.<sup>[5]</sup> This may explain the lack of perturbation of this particular NMR signal under different substrate combinations. After addition of UDP-GalNAc and  $\text{Mn}^{+2}$  to GalNAc-T2\*, the  ${}^{19}\text{F}$ -NMR resonance at  $\delta = -84.0$  ppm almost vanished and a new signal appeared at  $\delta = -83.7$  ppm (Figure 8, left panel and middle panel red tracing). It is remarkable that the exchange between these two  ${}^{19}\text{F}$ -NMR resonance ( $-84.0$  ppm and  $-83.7$  ppm) only occurs in the presence of UDP-GalNAc. Indeed, in the absence of UDP-GalNAc, the addition of  $\text{Mn}^{+2}$  (Figure 8, left panel and middle panel green tracing) or the peptide MUC5AC (Figure 8, left panel and middle panel grey tracing) did not induce any appreciable perturbation in the  ${}^{19}\text{F}$ -NMR spectrum of GalNAc-T2\*. This result strongly suggests that the  ${}^{19}\text{F}$ -NMR resonances at  $\delta = -84.0$  ppm and  $\delta = -83.7$  ppm correspond to the open and closed conformations of the flexible loop, respectively. These two populations are in slow exchange on the NMR chemical shift timescale and thus detected as two resonances. An additional set of experiments was also recorded in the presence of UDP and  $\text{Mn}^{+2}$  (Figure S11), giving similar results as obtained for addition of UDP-GalNAc. However, the exchange between the open and closed conformations of the flexible loop was not so pronounced compared to that in the presence of UDP-GalNAc. All together, our data show that UDP-GalNAc is absolutely required for the interconversion between the open and closed conformations of the flexible loop and that UDP-GalNAc stabilizes the closed conformation better than UDP (see Figure S9-S11). In fact, as deduced from the analysis of the crystal structure (see Figure 5 and panel corresponding to the active site of the PDB entry 4D0Z), UDP-GalNAc displays additional interactions with the enzyme through the GalNAc moiety. The fact that UDP does not completely exchange the population of the enzyme conformations towards the fully active state is consistent with its role as a reaction product after catalysis. The described  ${}^{19}\text{F}$ -NMR experiments provide a compelling evidence that GalNAc-T2 follows an induced-fit mechanism and support our previously proposed ordered bi-bi kinetic mechanism for this enzyme.<sup>[6c]</sup> The induced-fit mechanism has been also proposed for distant glycosyltransferases such as B4GALT1 and the lactose synthase by applying X-ray crystal structure studies<sup>[13]</sup>. In this mechanism, GalNAc-T2 is always in

the inactive state (open flexible loop) and only exchanges to the active state (closed flexible loop) in the presence of UDP-GalNAc/MnCl<sub>2</sub>. In contrast to the earlier hypothesis,<sup>[6c]</sup> our new data does not support the existence of an equilibrium between the inactive and active states in the absence of UDP-GalNAc/MnCl<sub>2</sub>. The uridine moiety of UDP-GalNAc might also shift its orientation from the "inverted" conformation to the typical geometry required for catalysis during the transition from the inactive to the active state of the enzyme. Only after the enzyme is in the active state, the peptide can bind, and the catalysis can take place.

#### The inactive-to-active state transition of the flexible loop is hindered in the F104S mutant.

The same set of experiments was performed for the F104S\* construct (Figure 8 right panel and Figure S12-S13). In this case,

no alterations in the <sup>19</sup>F-NMR spectrum were observed for any substrate addition. In particular, the addition of UDP-GalNAc or UDP did not induce any perturbation of the <sup>19</sup>F resonance at  $\delta = -84.0$  ppm (Figure 8 and S12-S13). This result also confirms the assignment of the peak at  $\delta = -84.0$  ppm to the open conformation of the flexible loop and highlights that this conformation is associated to the inactive form of GalNAc-T2. Furthermore, the <sup>19</sup>F-NMR experiments demonstrate that the flexible loop in F104S\* is fixed in the open conformation, precluding the binding to peptide substrates and their glycosylation. Overall, the <sup>19</sup>F-NMR experiments provide the molecular basis to explain why the F104S mutant does not recognize peptides and in turn is inactive.



**Figure 8.** <sup>19</sup>F-NMR Spectra of GalNAc-T2\* and F104S\* showing the open and closed loop conformations as a function of donor and acceptor substrate additions: (Left panel) Blue color corresponds to unliganded GalNAc-T2\*. The green color corresponds to GalNAc-T2\* in the presence of MnCl<sub>2</sub> (225  $\mu$ M). Red color corresponds to GalNAc-T2\*, MnCl<sub>2</sub> (225  $\mu$ M) and UDP-GalNAc (225  $\mu$ M). Grey color corresponds to GalNAc-T2\*, MnCl<sub>2</sub> (225  $\mu$ M), UDP-GalNAc (225  $\mu$ M) and MUC5Ac (150  $\mu$ M). (Middle panel) Blue color corresponds to unliganded GalNAc-T2\*. Grey color corresponds to GalNAc-T2\* and MUC5Ac (150  $\mu$ M). Green color corresponds to GalNAc-T2\*, MUC5Ac (150  $\mu$ M) and MnCl<sub>2</sub> (225  $\mu$ M). Red color corresponds to GalNAc-T2\* (135  $\mu$ M), MUC5Ac (150  $\mu$ M), MnCl<sub>2</sub> (225  $\mu$ M) and UDP-GalNAc (225  $\mu$ M). (Right panel) Blue color corresponds to unliganded F104S\*. Green color corresponds to F104S\*, MnCl<sub>2</sub> (130  $\mu$ M). Red color corresponds to F104S\*, MnCl<sub>2</sub> (130  $\mu$ M) and UDP-GalNAc (130  $\mu$ M). Grey color corresponds to F104S\*, MnCl<sub>2</sub> (130  $\mu$ M), UDP-GalNAc (130  $\mu$ M) and MUC5Ac (100  $\mu$ M). The concentration of GalNAc-T2\* and F104S was 135  $\mu$ M and 86  $\mu$ M, respectively. The experiments were recorded at 298K and at 600MHz (see the experimental section).

## Conclusions

Studies of naturally occurring mutations in glycosyltransferases that underlie human diseases or predispositions to diseases or other phenotypes have brought important insights into our understanding of this diverse group of enzymes.<sup>[14]</sup> In particular, studies of mutations that are not clear about how they affect protein structure and function have resulted in surprisingly new discoveries. From this perspective, our studies with the GalNAc-T2 F104S mutant provides another excellent example. Here, we provide compelling evidence that the inactive-to-active state transition of the flexible loop in the GalNAc-T2 F104S mutant is hindered. The presence of the polar residue (Ser104) causes instability of both the flexible and the  $\alpha$ 1- $\alpha$ 2 loops and prevents the flexible loop from adopting the closed conformation required for the active form of the enzyme. This result is further supported by the conservation of the Phe104, Val360 and Arg362 residues of the flexible loop in most GalNAc-Ts. We also predict that this

behavior of the flexible loop will be conserved among this family of enzymes despite their different lengths and their partially different amino acid sequences. Our study has also experimentally demonstrated from our <sup>19</sup>F NMR studies that GalNAc-T2 follows an induced-fit-mechanism in which UDP-GalNAc is absolutely required to activate the enzyme.

The finding that the GalNAc-T isoenzymes undergo an activation process induced by the donor substrate UDP-GalNAc may be an important factor in the regulation of the complex O-glycosylation process. The GalNAc-T isoenzymes not only have different, albeit partly overlapping, acceptor substrate specificities<sup>[9]</sup>, but they also exhibit rather different apparent K<sub>m</sub> values for UDP-GalNAc ranging from 10-160  $\mu$ M for GalNAc-T1, T2, T3, and T4.<sup>[15]</sup> Provided that all the GalNAc-T isoenzymes undergo similar donor-induced activation, the existing differences in affinities for the donor substrate may serve as a regulatory mechanism to ensure that the many isoenzymes act in a required order and without a premature competitive binding to

the acceptor proteins. This hypothesis is also supported by the lower  $K_m$  values for UDP-GalNAc of the first enzymes that are thought to initiate the O-glycosylation process, GalNAc-T1 and T2. These enzymes display low apparent  $K_m$  values (27 and 10  $\mu\text{M}$ , respectively), while the isoenzymes involved in the follow-up glycosylation of the partially (GalNAc) glycosylated substrates, e.g. GalNAc-T4, show a much higher apparent  $K_m$  for UDP-GalNAc (160  $\mu\text{M}$ ).<sup>[15]</sup>

The hexosamine biosynthetic pathway (HBP) regulates the cellular levels of UDP-GlcNAc and the cellular levels of UDP-GalNAc are co-regulated with UDP-GlcNAc through epimerization by the UDP-Gal/UDP-GalNAc 4-epimerase.<sup>[16]</sup> As the levels of UDP-GlcNAc are highly dynamic, HBP is regarded as a nutrient-sensing pathway.<sup>[17]</sup> Therefore, it is conceivable that the nutrient-sensing pathway may also influence O-GalNAc glycosylation via changes in UDP-GalNAc concentration.

Finally, our results have exploited the knowledge obtained from the structural analysis of a mutant associated with a known disease to obtain fundamental insights into the molecular mechanism of the GalNAc-T substrate recognition. Our multidisciplinary approach may furthermore provide the basis to treat low HDL-C levels through the design and implementation of new therapeutic strategies.

## Experimental Section

**Cloning, site-directed mutagenesis and purification.** The expression plasmid pPICZa*Agal*nact2 (K75-Q571), previously described<sup>[6c]</sup> was used as a template for introducing the following single amino-acid changes by site-directed mutagenesis as follows: T375C, F104S and F104S-T375C. Site-directed mutagenesis was carried out using the QuikChange protocol (Stratagene) as recommended by the supplier. In brief, F104S mutation was introduced into *GALNT2* codon optimized K75-Q571 ORF using the template pPICZa*Agal*nact2 (K75-Q571) and the oligo's T2F104SF (3'-TACGCAAGAAATAAGTCTAATCAGGTTGAAAGTGATAA-3') and T2F104SR (5'-TTATCACTTTCAACCTGATTAGACTTATTCTTGC GTA-3'), generating the plasmid pPICZa*Agal*nact2-F104S. Note that the position of the mutation is underlined in the above oligos. Hereafter the mutation T375C was introduced by custom based mutagenesis service from Genewiz generating pPICZa*Agal*nact2-T375C and pPICZa*Agal*nact2-F104S/T375C. The expression plasmid pPICZa*Agal*nact2 (K75-Q571), previously described<sup>[6c]</sup> was used as a template for introducing the following single amino-acid changes by site-directed mutagenesis as follows: F104S, T375C and F104S-T375C. Site-directed mutagenesis was carried out following the QuikChange protocol (Stratagene), using the Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific). All plasmids were verified by sequencing. The mutants were purified using the purification protocol of the wild-type enzyme described previously.<sup>[6c]</sup>

**Crystallization.** Crystals of the F104S mutant in complex with UDP were grown by sitting drop vapor diffusion at 18°C. Crystals were obtained by mixing 0.5  $\mu\text{l}$  of protein solution (a mix formed by 7 mg/ml F104S mutant, 5 mM UDP, 5 mM  $\text{MnCl}_2$ , 25 mM Tris pH 7.5, 0.5 mM EDTA and 1 mM TCEP-HCl) with 0.5  $\mu\text{l}$  of precipitant solution (10% PEG6000 and 100 mM Hepes pH 7) against 60  $\mu\text{l}$  of precipitant solution. The crystals were soaked within the mother liquor containing UDP-GalNAc at 50 mM for 60 minutes. Subsequently, they were cryoprotected in mother liquor

containing 25% ethylene glycol, 10% PEG6000 and 100 mM Hepes pH 7, and frozen in a nitrogen gas stream cooled to 100 K.

**Structure determination and refinement.** All data were processed and scaled using the XDS package<sup>[18]</sup> and CCP4 software,<sup>[19]</sup> relevant statistics are given in Table S2. The crystal structures were solved by molecular replacement with Phaser<sup>[19b]</sup> and using the PDB entry 4D0T as the template. Initial phases were further improved by cycles of manual model building in Coot and refinement with REFMAC5.<sup>[20]</sup> The final models were validated with PROCHECK,<sup>[21]</sup> model statistics are given in Table S2. The asymmetric unit of the tetragonal crystals contain 2 molecules of the inactive mutant in the AU (Figure 3). Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB, and see Table S2 for the pdb code).

**Labeling of the mutants T375C and F104S-T375C.** Both the T375C and the F104S-T375C variants were site-selectively modified by treatment with 3-bromo-1,1,1-trifluoroacetone (BFA) under mild conditions (5 min, phosphate buffer pH 7.0, room temperature), as previously described by Rydzik *et al.*<sup>[10]</sup>

**Solid-phase peptide synthesis.** MUC5AC was synthesized as previously described.<sup>[5]</sup>

**Tryptophan fluorescence spectroscopy.** Fluorescence spectroscopy was used to determine the dissociation constants of the WT enzyme and the F104S mutant against UDP-GalNAc. All experiments were carried out in a Cary Eclipse spectrofluorometer (Varian) at 25 °C with the WT enzyme and the F104S mutant at 1  $\mu\text{M}$ , and concentrations of UDP-GalNAc varying from 1 to 1000  $\mu\text{M}$  in 25 mM Tris, 150 mM NaCl, 1 mM  $\text{MnCl}_2$ , pH 7.5. Fluorescence emission spectra were recorded in the 300–400 nm range with an excitation wavelength of 280 nm, with slit width of 5 nm. The data analysis was performed in Prism (GraphPad software) considering a model with a single binding site (see equation (1), where  $F_0$  is the intrinsic fluorescence of the enzyme in the absence of quencher ( $Q$ ),  $F_1$  is the observed fluorescence at a given quencher concentration,  $f_a$  is the fractional degree of fluorescence and  $K_d$  is the dissociation constant.

Eq 1.

$$1 - \frac{F_1}{F_0} = \frac{f_a * [Q]}{K_d + [Q]}$$

**Activity assay** The WT GalNAc-T2 and GalNAc-T2\* activities were tested by incubating 100 nM of each enzyme with 10  $\mu\text{M}$  MUC1 peptide<sup>[22]</sup> and 20  $\mu\text{M}$  UDP-[<sup>14</sup>C]GalNAc (2000 cpms/nm), in 25 mM cacodylate pH 7.4, 10 mM  $\text{MnCl}_2$ , 0.25% Triton X-100. Excess of UDP-GalNAc was separated from the glycosylated peptide through DOWEX 1 chromatography. The product was analyzed by scintillation counting.

**Thermal shift assays.** The protein unfolding process was monitored using the environmentally sensitive dye 1-anilino-8-naphthalene sulfonate (ANS). Its quantum yield increased upon binding to hydrophobic surfaces exposed during protein unfolding and so does the fluorescence signal. The thermal-shift assay was conducted in the FluoDia T70 (Photal Otsuka Electronics). Solutions of 2  $\mu\text{M}$  WT GalNAc-T2 or the F104S mutant, 100  $\mu\text{M}$  ANS, 20 mM HEPES pH 7.0, and 200 mM NaCl final concentration and a volume of 100  $\mu\text{l}$  were added to the wells of a 96-well PCR plate. Mineral oil was last added to avoid evaporation during the experiment. Data were analysed with the Origin

software. To obtain  $T_0$ , a Boltzmann model from Origin was used to fit the fluorescence imaging data.<sup>[8]</sup>

**Molecular Dynamics simulations.** MD simulations were performed using Amber16 and the force-field ff14SB<sup>[23]</sup>. Each protein was immersed in a box with a 10 Å buffer of TIP3P<sup>[24]</sup> water molecules and neutralized by adding explicit counter ions (Cl<sup>-</sup>). A two-stage geometry optimization approach was used. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal/mol were applied to the solute, and the Andersen temperature coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages. Water molecules are treated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects are modeled using the particle-Mesh-Ewald method.<sup>[25]</sup> An 8-Å cut-off was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 200 ns under the same simulation conditions.

**nLC MS/MS and data analysis.** To preserve the trifluorinated compound at the peptide backbone, samples were not reduced and alkylated. GalNAcT\* and F104S\* were directly digested with trypsin and after StageTip (Empore 3M) desalting, submitted to nLC-MS/MS analysis. EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via nanoSpray Flex ion source to an Orbitrap Fusion MS (Thermo Scientific) were used for analysis. Briefly, the nLC was operated in a single analytical column set up using PicoFrit Emitters (New Objectives, 75 µm inner diameter) packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, 1.9-µm particle size, 19-21 cm column length). Each sample was injected onto the column and eluted in a gradient from 2 to 25 % B in 45 min at 200 nL/min (Solvent A, 100 % H<sub>2</sub>O; Solvent B, 100 % acetonitrile; both containing 0.1 % (v/v) formic acid). A precursor MS1 scan ( $m/z$  350–2,000) of intact peptides was acquired in the Orbitrap at the nominal resolution setting of 120,000, followed by Orbitrap HCD-MS2 and at the nominal resolution setting of 60,000 of the five most abundant multiply charged precursors in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 was used for triggering data-dependent fragmentation events. Activation times were 75 ms for HCD fragmentation; isolation width was 1.2 mass units, and usually 1 microscan was collected for each spectrum. Polysiloxane ions at  $m/z$  445.12003 were used as a lock mass in all runs. Data processing was performed using Proteome Discoverer 1.4 software (Thermo Scientific) using Sequest HT node as a search engine. In all cases the precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 20 mmu. Methionine oxidation and the trifluorinated compound at Ser/Thr were used as variable modifications. All spectra were searched against GalNAcT2\* and F104S\* sequences with and without mutations at the position 375.

**NMR experiments.** NMR experiments for peptide characterization and saturation transfer difference NMR experiments were recorded on a Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe head. The <sup>1</sup>H-NMR resonances of the peptide were assigned through standard 2D-TOCSY (30 and 80 ms mixing time) and 2D-NOESY experiments (400 ms mixing time). Solution conditions used for the NMR characterization studies were 0.5 mM peptide in 90:10 H<sub>2</sub>O/D<sub>2</sub>O. The assignments were accomplished at 278K. As a chemical shift reference in the <sup>1</sup>H NMR experiments we used the resonance of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) ( $\delta$  TSP = 0 ppm). Peak lists for the 2D-TOCSY and 2D-NOESY spectra were generated by

interactive peak picking using the Computer Aided Resonance Assignment (CARA) software. Table S1 contains the <sup>1</sup>H-NMR assignment for the MUC5AC peptide. Samples for STD-experiments were prepared in perdeuterated 25 mM TRIS-d11 in deuterated water, 7.5 mM NaCl and 1 mM DTT, uncorrected pH 7.4. STD-NMR experiments were performed at 298 K in the presence of 75 µM UDP, 75 µM MnCl<sub>2</sub> with a molar ratio of 40:1 peptide:protein. The STD-NMR spectra were acquired with 1920 transients in a matrix with 64k data points in  $t_2$  in a spectral window of 12335.5 Hz centered at 2820.6 Hz. An excitation sculpting module with gradients was employed to suppress the water proton signals. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at -1 ppm using a series of Eburp2.1000-shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. Proper STD-NMR control experiments were performed: i) with the ligand in absence of the protein to ensure that the ligand signals were not affected; ii) with the protein in absence of the ligand (data not shown). The peptide MUC5AC when irradiated at -1 ppm in the absence of protein showed residual saturation on the aliphatic methyl groups in the STD-NMR spectra. The STD-NMR experiments shown in this work take into account the STD control experiments. To obtain the STD-NMR-derived epitope mapping the STD-NMR total intensities were normalized with respect the highest STD-NMR response. <sup>19</sup>F-NMR experiments were recorded by 600 MHz Bruker Avance spectrometer equipped with a <sup>19</sup>F,<sup>1</sup>H SEF dual probe optimized for direct <sup>19</sup>F detection. These experiments were carried out to monitor the <sup>19</sup>F reporter (CF<sub>3</sub>)<sub>2</sub>COCH<sub>2</sub>-selectively introduced at the SH- group of the mutants T375C from both constructs of the WT GalNAc-T2 and the F104S mutant. The <sup>19</sup>F experiments of GalNAcT2\* and F104S\* constructs were recorded at distinct conditions in absence and presence of MnCl<sub>2</sub>, UDP or UDP-GalNAc and peptide. The UDP or UDP-GalNAc and MnCl<sub>2</sub> were added in an excess of 1.5 times of the protein concentration while peptide was in a 1.1:1 peptide:protein molar ratio. Higher concentrations of UDP or UDP-GalNAc (500 µM), MnCl<sub>2</sub> (500µM) and peptide (protein:peptide molar ratio 1:2.2) do not induce higher <sup>19</sup>F chemical shift perturbations. The <sup>19</sup>F-NMR experiments with proton decoupling were acquired at 298 K by acquisition of 640 scans in a matrix with 8k data points in a spectral window of 11295.2 Hz centered at (CF<sub>3</sub>)<sub>2</sub>CHOH -45174.9 Hz. The resonance of 1,1,1,3,3,3-hexafluoro-2-propanol was used as a chemical shift reference in the <sup>19</sup>F NMR experiments ( $\delta$  = -75.7 ppm). Samples for <sup>19</sup>F-NMR experiments were prepared in 25 mM TRIS-d11 in 90:10 H<sub>2</sub>O/D<sub>2</sub>O, uncorrected pH 7.5. In the <sup>19</sup>F-NMR spectra, trace amounts of the 3-Bromo-1,1,1-trifluoroacetone reagent used in protein labelling ( $\delta$  CF<sub>3</sub>COCH<sub>2</sub>Br = -83.3 ppm) were observed. In addition, trace amounts of trifluoroacetic acid (TFA), used in the synthesis of peptides, were observed in the <sup>19</sup>F-NMR spectra ( $\delta$  TFA = -75.4 ppm). To monitor the glycosylation of MUC5AC by GalNAcT2\* and F104S\* two <sup>1</sup>H-NMR spectra were recorded before and after the addition of UDP-GalNAc.

## Acknowledgements

We thank the Ministerio de Economía y Competitividad (CTQ2013-44367-C2-2-P and BFU2016-75633-P to R.H.G., CTQ2015-67727-R to F.C., CTQ2015-64597-C2-1P to J.J-B). F.M. thanks FCT-Portugal to IF project (IF/00780/2015) and UCIBIO funding UID/Multi/04378/2013 co-financed by the FEDER (POCI-01-0145-FEDER-007728). The NMR spectrometers are part of PTNMR supported by Project No 022161. H.C. thanks the Lundbeck Foundation and the Danish National Research Foundation (DNRF107). E.L.N.

acknowledges her postdoctoral EMBO fellowship ALTF 1553-2015 co-funded by the European Commission (LTFCOFUND2013, GA-2013-609409) and Marie Curie Actions. H.C. and J.J-B. thank EU for the TOLLerant project. R.H-G. thanks Agencia Aragonesa para la Investigación y Desarrollo (ARAID) and the Diputación General de Aragón (DGA, B89) for financial support. The research leading to these results has also received funding from the FP7 (2007-2013) under BioStruct-X (grant agreement N°283570 and BIOSTRUCTX\_5186). We thank synchrotron radiation source DIAMOND (Oxford) and beamline I04 (number of experiment MX10121-19). We would also like to thank to Prof. Tom Gerken for his valuable comments in our manuscript.

**Keywords:** GalNAc-T2• F104S mutant • X-ray crystallography • missense mutations • NMR

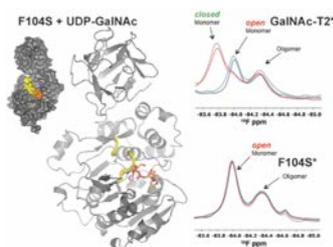
- [1] C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, R. Gupta, E. P. Bennett, U. Mandel, S. Brunak, H. H. Wandall, S. B. Lavery, H. Clausen, *Embo J* **2013**, *32*, 1478-1488.
- [2] S. Fukumoto, *Curr Opin Nephrol Hypertens* **2014**, *23*, 346-351.
- [3] K. Kato, C. Jeanneau, M. A. Tarp, A. Benet-Pages, B. Lorenz-Depiereux, E. P. Bennett, U. Mandel, T. M. Strom, H. Clausen, *J Biol Chem* **2006**, *281*, 18370-18377.
- [4] S. A. Khetarpal, K. T. Schjoldager, C. Christoffersen, A. Raghavan, A. C. Edmondson, H. M. Reutter, B. Ahmed, R. Ouazzani, G. M. Peloso, C. Vitali, W. Zhao, A. V. Somasundara, J. S. Millar, Y. Park, G. Fernando, V. Livanov, S. Choi, E. Noe, P. Patel, S. P. Ho, S. Myocardial Infarction Exome Sequencing, T. G. Kirchgessner, H. H. Wandall, L. Hansen, E. P. Bennett, S. Y. Vakhrushev, D. Saleheen, S. Kathiresan, C. D. Brown, R. Abou Jamra, E. LeGuern, H. Clausen, D. J. Rader, *Cell Metab* **2016**, *24*, 234-245.
- [5] E. Lira-Navarrete, M. de Las Rivas, I. Companon, M. C. Pallares, C. Rovira, P. Bernado, P. Bruscolini, H. Clausen, A. Lostao, F. Corzana, R. Hurtado-Guerrero, *Nat Commun* **2015**, *6*, 6937.
- [6] aM. Ghirardello, M. de Las Rivas, A. Lacetera, I. Delso, E. Lira-Navarrete, T. Tejero, S. Martin-Santamaria, R. Hurtado-Guerrero, P. Merino, *Chemistry* **2016**, *22*, 7215-7224; bR. Hurtado-Guerrero, *Biochem Soc Trans* **2016**, *44*, 61-67; cE. Lira-Navarrete, J. Iglesias-Fernandez, W. F. Zandberg, I. Companon, Y. Kong, F. Corzana, B. M. Pinto, H. Clausen, J. M. Peregrina, D. J. Vocadlo, C. Rovira, R. Hurtado-Guerrero, *Angew Chem Int Ed Engl* **2014**, *53*, 8206-8210.
- [7] aT. A. Fritz, J. H. Hurley, L. B. Trinh, J. Shiloach, L. A. Tabak, *Proc Natl Acad Sci U S A* **2004**, *101*, 15307-15312; bT. A. Fritz, J. Raman, L. A. Tabak, *J Biol Chem* **2006**, *281*, 8613-8619; cT. Kubota, T. Shiba, S. Sugioka, S. Furukawa, H. Sawaki, R. Kato, S. Wakatsuki, H. Narimatsu, *J Mol Biol* **2006**, *359*, 708-727; dM. L. Rivas, E. Lira-Navarrete, E. J. P. Daniel, I. Companon, H. Coelho, A. Diniz, J. Jimenez-Barbero, J. M. Peregrina, H. Clausen, F. Corzana, F. Marcelo, G. Jimenez-Oses, T. A. Gerken, R. Hurtado-Guerrero, *Nat Commun* **2017**, *8*, 1959.
- [8] E. Lira-Navarrete, J. Valero-Gonzalez, R. Villanueva, M. Martinez-Julvez, T. Tejero, P. Merino, S. Panjikar, R. Hurtado-Guerrero, *PLoS One* **2011**, *6*, e25365.
- [9] T. A. Gerken, O. Jamison, C. L. Perrine, J. C. Collette, H. Moinova, L. Ravi, S. D. Markowitz, W. Shen, H. Patel, L. A. Tabak, *J Biol Chem* **2011**, *286*, 14493-14507.
- [10] A. M. Rydzik, J. Brem, S. S. van Berkel, I. Pfeffer, A. Makena, T. D. Claridge, C. J. Schofield, *Angew Chem Int Ed Engl* **2014**, *53*, 3129-3133.
- [11] W. G. Touw, C. Baakman, J. Black, T. A. te Beek, E. Krieger, R. P. Joosten, G. Vriend, *Nucleic Acids Res* **2015**, *43*, D364-368.
- [12] J. M. Aramini, K. Hamilton, L. C. Ma, G. V. T. Swapna, P. G. Leonard, J. E. Ladbury, R. M. Krug, G. T. Montelione, *Structure* **2014**, *22*, 515-525.
- [13] B. Ramakrishnan, E. Boeggeman, P. K. Qasba, *Biochem Biophys Res Commun* **2002**, *291*, 1113-1118.
- [14] H. H. Freeze, H. Schachter, T. Kinoshita, in *Essentials of Glycobiology*, 3rd ed. (Eds.: A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger), Cold Spring Harbor (NY), **2015**.
- [15] E. P. Bennett, H. Hassan, U. Mandel, E. Mirgorodskaya, P. Roepstorff, J. Burchell, J. Taylor-Papadimitriou, M. A. Hollingsworth, G. Merckx, A. G. van Kessel, H. Eiberg, R. Steffensen, H. Clausen, *J Biol Chem* **1998**, *273*, 30472-30481.
- [16] aB. A. Lewis, J. A. Hanover, *J Biol Chem* **2014**, *289*, 34440-34448; bJ. B. Thoden, T. M. Wohlers, J. L. Fridovich-Keil, H. M. Holden, *J Biol Chem* **2001**, *276*, 15131-15136.
- [17] S. Marshall, V. Bacote, R. R. Traxinger, *J Biol Chem* **1991**, *266*, 4706-4712.
- [18] W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.
- [19] aActa Crystallogr D Biol Crystallogr **1994**, *50*, 760-763; bM. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 235-242.
- [20] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 355-367.
- [21] R. A. e. a. Laskowski, *J. Appl. Cryst* **1993**, *26*
- [22] H. H. Wandall, F. Irazoqui, M. A. Tarp, E. P. Bennett, U. Mandel, H. Takeuchi, K. Kato, T. Irimura, G. Suryanarayanan, M. A. Hollingsworth, H. Clausen, *Glycobiology* **2007**, *17*, 374-387.
- [23] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712-725.
- [24] T. A. Andrea, W. C. Swope, H. C. Andersen, *J. Chem. Phys.* **1983**, *79*, 4576-4458.
- [25] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.

## Entry for the Table of Contents (Please choose one layout)

Layout 1:

## FULL PAPER

**Decoding disease-causing mechanisms of missense mutations:** Here, we have studied the GalNAc-T2 F104S mutant, which causes low levels of HDL-C and predisposition to cardiovascular diseases. The mutation precludes the shifting of the flexible loop from the inactive to the active conformation, affecting the enzyme ability to bind its peptide substrates. We also demonstrate that UDP-GalNAc induces the active conformation to promote O-glycosylation, suggesting that UDP-GalNAc acts as a regulatory molecule of GalNAc-T isoenzymes.



Matilde de las Rivas, Helena Coelho, Ana Diniz, Erandi Lira-Navarrete, Ismael Compañón, Jesús Jiménez-Barbero, Katrine. T. Schjoldager, Eric. P. Bennett, Sergey Y. Vakhrushev, Henrik Clausen, Francisco Corzana, Filipa Marcelo,\* and Ramon Hurtado-Guerrero\*

Page No. – Page No.

**Structural analysis of a GalNAc-T2 mutant reveals an induced-fit catalytic mechanism for GalNAc-Ts**