



**HAL**  
open science

## **N-Glycan on the Non-Consensus N-X-C Glycosylation Site Impacts Activity, Stability, and Localization of the Sd Synthase B4GALNT2.**

Virginie Coge, Dorothee Vicogne, Celine Schulz, Lucie Portier, Giulia Venturi, Jerome de Ruyck, Mathieu Decloquement, Marc Lensink, Guillaume Brysbaert, Fabio Dall'Olio, et al.

### ► To cite this version:

Virginie Coge, Dorothee Vicogne, Celine Schulz, Lucie Portier, Giulia Venturi, et al.. N-Glycan on the Non-Consensus N-X-C Glycosylation Site Impacts Activity, Stability, and Localization of the Sd Synthase B4GALNT2.. International Journal of Molecular Sciences, 2023, International journal of molecular sciences, 24. hal-04257639

**HAL Id: hal-04257639**

**<https://hal.univ-lille.fr/hal-04257639>**

Submitted on 25 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Article

# N-Glycan on the Non-Consensus N-X-C Glycosylation Site Impacts Activity, Stability, and Localization of the Sd<sup>a</sup> Synthase B4GALNT2

Virginie Coge<sup>1</sup>, Dorothée Vicogne<sup>1</sup> , Céline Schulz<sup>1</sup> , Lucie Portier<sup>1</sup>, Giulia Venturi<sup>2</sup>, Jérôme de Ruyck<sup>1</sup> , Mathieu Decloquement<sup>1</sup> , Marc F. Lensink<sup>1</sup> , Guillaume Brysbaert<sup>1</sup> , Fabio Dall'Olio<sup>2</sup> , Sophie Groux-Degroote<sup>1</sup> and Anne Harduin-Lepers<sup>1,\*</sup>

<sup>1</sup> CNRS, UMR 8576-UGSF-Unité de Glycobiologie Structurale et Fonctionnelle, Université de Lille, F-59000 Lille, France

<sup>2</sup> Department of Medical and Surgical Sciences (DIMEC), University of Bologna, General Pathology Building, Via San Giacomo 14, 40126 Bologna, Italy

\* Correspondence: anne.harduin-lepers@univ-lille.fr; Tel.: +33-320-33-62-46

**Abstract:** The Sd<sup>a</sup> carbohydrate epitope and its biosynthetic B4GALNT2 enzyme are expressed in the healthy colon and down-regulated to variable extents in colon cancer. The human *B4GALNT2* gene drives the expression of a long and a short protein isoform (LF-B4GALNT2 and SF-B4GALNT2) sharing identical transmembrane and luminal domains. Both isoforms are trans-Golgi proteins and the LF-B4GALNT2 also localizes to post-Golgi vesicles thanks to its extended cytoplasmic tail. Control mechanisms underpinning Sd<sup>a</sup> and B4GALNT2 expression in the gastrointestinal tract are complex and not fully understood. This study reveals the existence of two unusual *N*-glycosylation sites in B4GALNT2 luminal domain. The first atypical N-X-C site is evolutionarily conserved and occupied by a complex-type *N*-glycan. We explored the influence of this *N*-glycan using site-directed mutagenesis and showed that each mutant had a slightly decreased expression level, impaired stability, and reduced enzyme activity. Furthermore, we observed that the mutant SF-B4GALNT2 was partially mislocalized in the endoplasmic reticulum, whereas the mutant LF-B4GALNT2 was still localized in the Golgi and post-Golgi vesicles. Lastly, we showed that the formation of homodimers was drastically impaired in the two mutated isoforms. An AlphaFold2 model of the LF-B4GALNT2 dimer with an *N*-glycan on each monomer corroborated these findings and suggested that *N*-glycosylation of each B4GALNT2 isoform controlled their biological activity.

**Keywords:** B4GALNT2; dimer; glycosyltransferase; *N*-glycan; unusual N-X-C glycosylation site



**Citation:** Coge, V.; Vicogne, D.; Schulz, C.; Portier, L.; Venturi, G.; de Ruyck, J.; Decloquement, M.; Lensink, M.F.; Brysbaert, G.; Dall'Olio, F.; et al. *N*-Glycan on the Non-Consensus N-X-C Glycosylation Site Impacts Activity, Stability, and Localization of the Sd<sup>a</sup> Synthase B4GALNT2. *Int. J. Mol. Sci.* **2023**, *24*, 4139. <https://doi.org/10.3390/ijms24044139>

Received: 29 November 2022

Revised: 10 February 2023

Accepted: 16 February 2023

Published: 18 February 2023

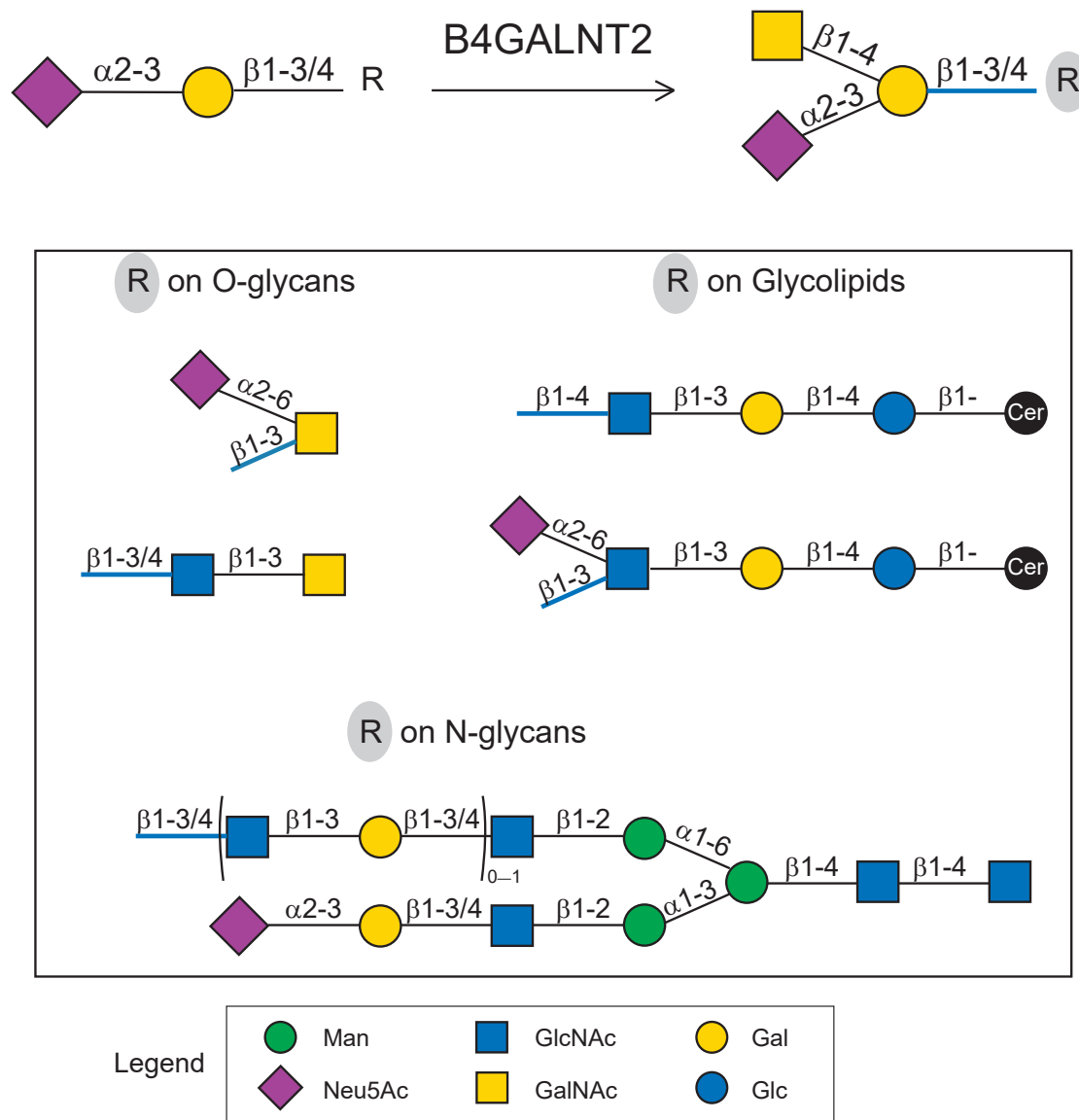


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Sd<sup>a</sup> carbohydrate epitope (GalNAcβ1-4[Neu5Acα2-3]Galβ1-) is a histo-blood group antigen found on the erythrocytes of 90% of Caucasians [1,2] defining the blood group Sid [3,4]. It is also described in human adult tissues on *O*- and *N*-glycans of proteins and on glycolipids (Figure 1), as reviewed recently [3,5]. In particular, Sd<sup>a</sup> carried by mucins *O*-glycans is a major structural feature in the mucus layer of the healthy human descending colon that disappears in cancer colon to the benefit of the sialyl Lewis x epitope (sLex, Neu5Acα2-3Galβ1-4[Fucα1-3]GlcNAc) [3,6]. It is also a tumor-associated carbohydrate antigen (TACA) found on glycolipids and glycoproteins in the prostate cancer renal cell carcinoma cell lines known as RM2 epitope [7,8]. A wide range of effects are attributed to Sd<sup>a</sup> according to its environment, as reviewed in [5]. Among these, Sd<sup>a</sup> inhibits metastasis in gastro-intestinal cancer cells and the stemness-associated malignant phenotype in cancerous colon cells [5,9,10], prevents the development of muscular dystrophy at the neuro-muscular junction [11,12], regulates clearance of the Von Willebrand factor [13], modulates the lytic

function of cytotoxic T lymphocytes [14], impacts gut microbiota in mice [15], and regulates embryo attachment in a mouse model [16].



**Figure 1.** Schematic illustrating  $Sd^a$  biosynthesis.  $B4GALNT2$  catalyzes the transfer of a *N*-acetylgalactosamine (GalNAc) residue onto sialylated disaccharides (Neu5Ac $\alpha$ 2,3Gal $\beta$ 1,3/4) found either on *O*-glycans, glycolipids, or *N*-glycans. The possible R moieties (circled in grey) are represented in the box below the reaction. Glycans are drawn according to the Symbol Nomenclature for Glycans (SNFG) guidelines [17] with Drawglycan-SNFG [18], as indicated in the legend. The non-glycan ceramide moiety is represented here by a black-filled circle.

The  $Sd^a$ /Cad synthase (i.e.,  $\beta$ 1,4-*N*-acetylgalactosaminyltransferase II,  $\beta$ 1,4-GalNAcT-II,  $B4GALNT2$ ) catalyzing the transfer of an *N*-acetylgalactosamine residue from UDP-GalNAc to the sialylated disaccharide Neu5Ac $\alpha$ 2,3Gal $\beta$ 1,3/4-R (Figure 1) is encoded by the  $B4GALNT2$  gene. The human gene maps to chromosome 17 and contains 11 exons [19,20].  $B4GALNT2$  drives the expression of two major transcripts differing in their 5' ends as a result of the use of two alternative first exons named exon 1S and exon 1L [19,20]. It was shown that in the gastrointestinal tract that exon 1S is predominantly used [6], and expression of this transcript requires ETS1 combined with DMTF1 and to a lesser extent SP1, both transcription factors (TF). However, these TF are not sufficient to explain the

tissue-specific expression of B4GALNT2, suggesting the existence of additional regulatory elements [3,21,22]. The two transcripts were shown to encode two Sd<sup>a</sup>/Cad synthase protein isoforms with different-length cytoplasmic tails, which are designated short form (SF) and long form (LF) [6]. Both isoforms stably expressed in colon cancer cells LS174T lead to the formation of Sd<sup>a</sup> antigen [23]. Recent studies have shown that the two isoforms have a Golgi localization and that the unusual, extended cytoplasmic tail of LF-isoform conferred an additional subcellular location of LF-B4GALNT2 in dynamic post-Golgi vesicles [24].

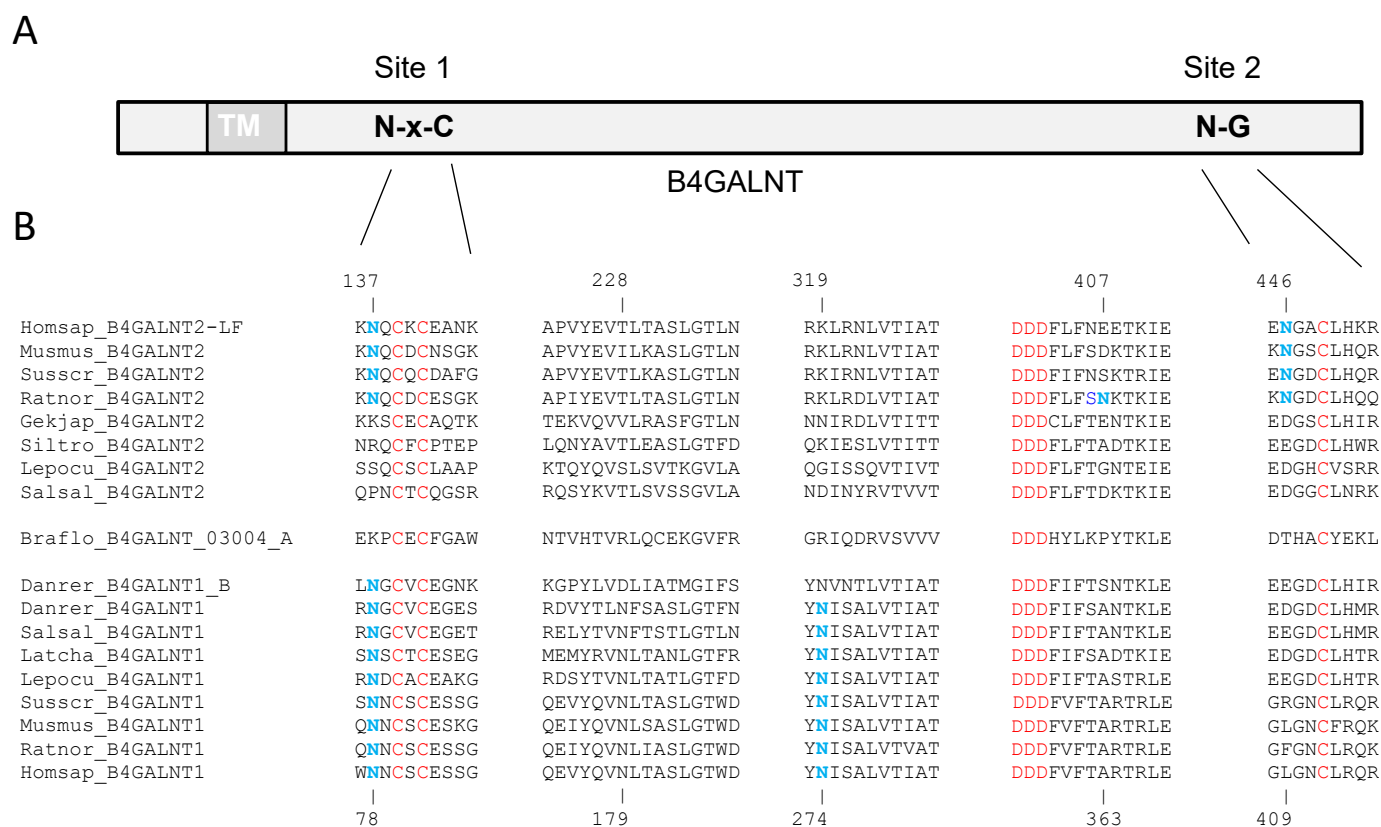
Although substantial achievements have been made towards elucidating the molecular mechanisms underpinning Sd<sup>a</sup> expression, biosynthesis of this carbohydrate epitope and regulation of its expression in the human tissues appear to be complex and remain poorly understood [3]. Asparagine-linked glycosylation (*N*-glycosylation) of proteins is a widely distributed post-translational modification (PTM) of proteins in eukaryotic cells. It is a process initiated in the endoplasmic reticulum (ER) on the nascent proteins that occurs predominantly at the sequons N-X-T/S, where T = Thr = threonine, S = Ser = serine, and X can be any amino acid residue except proline (Pro, P), which is also known as classical or canonical motif N-[!P]-S/T-[!P]. *N*-glycosylation is a dynamic process playing roles in protein folding, maturation, secretion, and intracellular trafficking and also in disease progression [25]. Most glycosyltransferases involved in glycans construction of glycoproteins and glycolipids are themselves glycoproteins. These past years, several studies showed that elimination of their *N*-glycans could affect their activity, as recently reviewed [26]. In contrast to the mouse B4GALNT2 protein [27], sequence analysis of the two human B4GALNT2 polypeptides using the sequence-based predictor NetNGlyc [28] failed to predict the potential *N*-glycosylation site [20]. In addition, the human paralogue B4GALNT1 (also known as GM2/GD2 synthase) of B4GALNT2 in the CAZy family GT12 was shown to possess three *N*-glycosylation sites, and their removal resulted in a significant decrease in the enzyme activity [29]. Although 96.5% of *N*-glycosylation sites of mammalian proteins are found at a canonical sequon, minor and less predictable *N*-glycosylation sites such as N-X-C (1.3%), N-G (0.5%), or N-X-V (0.4%) are possibly used; they are referred to as unusual or atypical glycosylation sites [30–33].

In this study, we reveal the existence of two potential *N*-glycosylation-sites in the human B4GALNT2 sequence. We show that the unusual N-X-C site (site 1) in the stem region is highly conserved during evolution in both vertebrate B4GALNT1 and mammalian B4GALNT2. This site is occupied by a complex-type *N*-glycan on mouse, rat, and human colonic B4GALNT2 and on the two recombinant isoforms of human B4GALNT2 produced in HeLa cells. To clarify the role of this *N*-glycan on human B4GALNT2 isoform localization, dimerization, and activity, we generated mutant protein isoforms lacking this glycosylation site. We show that this *N*-glycan slightly affects enzyme activity for both isoforms. Furthermore, our data suggest that it is essential for homodimers formation and for maintaining active conformation of the enzyme, as corroborated with a model of the glycosylated B4GALNT2 obtained with AlphaFold2.

## 2. Results

### 2.1. Vertebrate B4GALNT2 Sequences Contain Unusual and Conserved *N*-Glycosylation Site(s)

No canonical *N*-glycosylation sites could be predicted using NetNGlyc in the human B4GALNT2 [20] or in the mouse B4GALNT2, whereas a typical sequon N-X-T/S could be predicted only in the rat B4GALNT2 sequence (Supplementary Figure S1). However, manual inspection of the human, mouse, and rat B4GALNT2 sequences suggested the existence of two unusual *N*-glycosylation sites N-X-C (site 1) and N-G (site 2), respectively, in the stem region and in the catalytic domain of the mammalian B4GALNT2 sequences. In the human B4GALNT2 isoforms, the sites 1 and 2 are positioned on N77 and N386 in the SF-B4GALNT2 and on N137 and N446 in the LF-B4GALNT2 (Figure 2).



**Figure 2.** Conservation of the predicted *N*-glycosylation sites in vertebrates B4GALNT sequences of CAZy GT12. **(A)** Vertebrate B4GALNT sequences. Localization of the *N*-glycosylation sites N-X-C (site 1) and N-G (site 2) in vertebrate B4GALNT2 sequences are indicated in the schematic above. TM, transmembrane domain. Multiple sequence alignments were carried out with ClustalW using 136 full-length B4GALNT sequences (36 B4GALNT1, 70 B4GALNT2, and 20 B4GALNT1/2). **(B)** The MSA of B4GALNT sequences from the mammals *Mus musculus* B4GALNT2 (Musmus\_B4GALNT2, BC139166) and B4GALNT1 (Musmus\_B4GALNT1, NM\_001478), *Rattus norvegicus* B4GALNT2 (Ratnor\_B4GALNT2, XM\_039087862) and B4GALNT1 (Ratnor\_B4GALNT1, D17809), *Sus scrofa* B4GALNT2 (susscr\_B4GALNT2, NM\_001244330) and B4GALNT1 (Susscr\_B4GALNT1, AK237665), the squamata *Gekko japonicus* B4GALNT2 (Gekjap\_B4GALNT2, XM\_01509212) and B4GLNT1 (Gekjap\_B4GALNT1, XM\_015423486), the amphibian *Silurana tropicalis* B4GALNT2 (Siltro\_B4GALNT2, XM\_002935708) and B4GALNT1 (Siltro\_B4GALNT1, XM\_012956468), the non-teleostean fish *Lepisosteus oculatus* B4GALNT2 (Lepocu\_B4GALNT2, XM\_015362122) and B4GALNT1 (Lepocu\_B4GALNT1, XM\_006629420), the teleostean fish *Salmo salar* B4GALNT2 (Salsal\_B4GALNT2, XM\_01404566) and B4GALNT1 (Salsal\_B4GALNT1, BT058789), and one invertebrate B4GALNT sequence of amphioxus *Branchiostoma floridae* (Braflo\_B4GALNT\_03004\_A, XM\_002601302). Position of each predicted *N*-glycosylation site in the B4GALNT2 sequences is indicated above the alignment for the LF-B4GALNT2 sequence, and the conserved asparagine (N) appears in bold blue characters. In addition, the conserved *N*-glycosylation sites (N in blue) and positions described for the human B4GALNT1 sequence [29] are indicated below. Finally, the other conserved features of the B4GALNT sequences (i.e., cysteine residues (C) and D-X-D motif) are indicated in red characters.

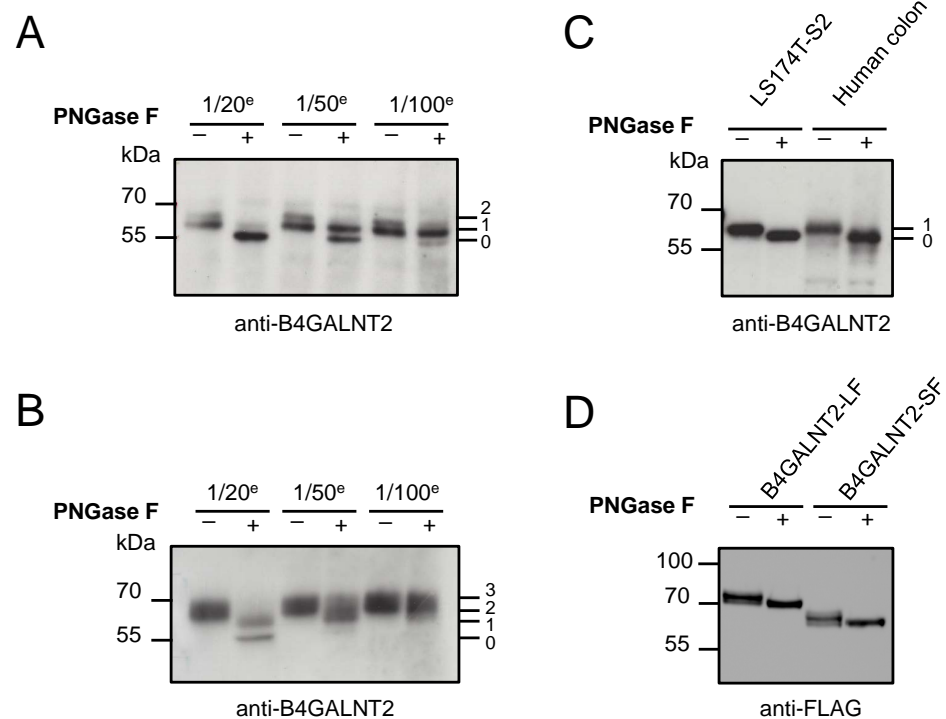
Conservation of these predicted *N*-glycosylation sites in B4GALNT-related sequences was analyzed during vertebrate evolution. Firstly, we identified 187 B4GALNT-related sequences (61 B4GALNT1, 94 B4GALNT2, and 32 invertebrates B4GALNT1/2) in public databases at NCBI using a BLAST search approach and carried out multiple sequence alignments (MSA). The MSA of B4GALNT sequences from the mammals *Mus musculus*, *Rattus norvegicus*, *Sus scrofa*, the squamata *Gekko japonicus*, the amphibian *Silurana tropicalis*, the non-teleostean fish *Lepisosteus oculatus*, the teleostean fish *Salmo salar*, the amphioxus

*Branchiostoma floridae*, and position of each predicted *N*-glycosylation site are shown in Figure 2. The predicted *N*-glycosylation sites in mammalian B4GALNT2 sequences were more or less conserved in B4GALNT2 sequences during vertebrate evolution and in the GM2 synthase (B4GALNT1). The first atypical *N*-glycosylation site N-X-C (site 1) is conserved in amniotes B4GALNT2 and in all B4GALNT1 sequences but not conserved in fish and frog B4GALNT2 sequences or in invertebrates B4GALNT1/2 sequences (Figure 2). The canonical sequon N-[!P]-S/T-[!P] found in the catalytic domain of rat B4GALNT2 sequences nearby the D-X-D motif is not conserved in the majority of vertebrate species or in B4GALNT1 sequences (Figure 2). The unusual *N*-glycosylation site N-G (site 2) is conserved in amniotes but lost in amphibian and fish B4GALNT2 sequences and is not conserved in B4GALNT1 sequences. Three *N*-glycosylation sites have been reported in the human B4GALNT1 sequence (N79, N179, and N274) [29]. The first site N79 described in this study is localized in a sequon (N-C-S) and conserved only in the mammalian B4GALNT1 sequences, whereas the two others are conserved in nearly all the vertebrate B4GALNT1 sequences (Figure 2).

## 2.2. *N*-Glycosylation of B4GALNT2 in Mammals

We next checked the occupancy of these predicted B4GALNT2 *N*-glycosylation sites *in vivo* in protein extracts from mouse, rat, and human colonic mucosa (Figure 3). We used serial dilutions of the endoglycosidic *N*-Glycosidase F (PNGase F) enzyme on colon protein extracts and ran SDS-PAGE. After transfer onto nitrocellulose membrane, the various glycosylated isoforms of the B4GALNT2 could be visualized with a specific anti-B4GALNT2 antibody directed towards their shared stem domain. As shown in Figure 3, three *N*-glycoforms (with no, one, or two *N*-glycans) of the B4GALNT2 were detected in mouse colon, supporting the existence of two *N*-glycans on the mouse enzyme *in vivo* (Figure 3A). Similarly, four *N*-glycoforms were observed on the rat B4GALNT2 (Figure 3B), supporting the presence of three *N*-glycans on the rat enzyme in colon. Only two *N*-glycoforms were observed on the SF-B4GALNT2 expressed in human colon and in the stably transfected LS174T-S2 cells [23] (Figure 3C), further suggesting the existence of a unique *N*-glycan on human B4GALNT2.

As a first step towards deciphering the role of this *N*-glycan, we constructed FLAG-tagged, full-length SF-B4GALNT2 and LF-B4GALNT2 in the p3×FLAG-CMV-10 expression vector, and recombinant proteins were expressed in HeLa cells. Partial digestion with PNGase F and Western blotting (WB) with an anti-FLAG antibody indicated the presence of one *N*-glycan on each recombinant protein, as shown previously *in vivo* (Figure 3D). Additional digestion assays of these recombinant proteins with Endo H indicated each recombinant B4GALNT2 isoform carried both Endo H-sensitive *N*-glycans (i.e., high-mannose/hybrid type) as found in ER and PNGase F-sensitive *N*-glycans (i.e., complex type) as found in the Golgi apparatus (Supplementary Figure S2). To determine which *N*-glycosylation site was occupied, C-terminus deletion constructs of the SF-B4GALNT2 ( $\Delta$ site 2 ( $\Delta$ 262–506),  $\Delta$ site 2 ( $\Delta$ 97–506), and  $\Delta$ site 1&2 ( $\Delta$ 80–506)) were expressed in HeLa cells, and WB showed that the *N*-glycan is positioned on site 1 (Supplementary Figure S3).

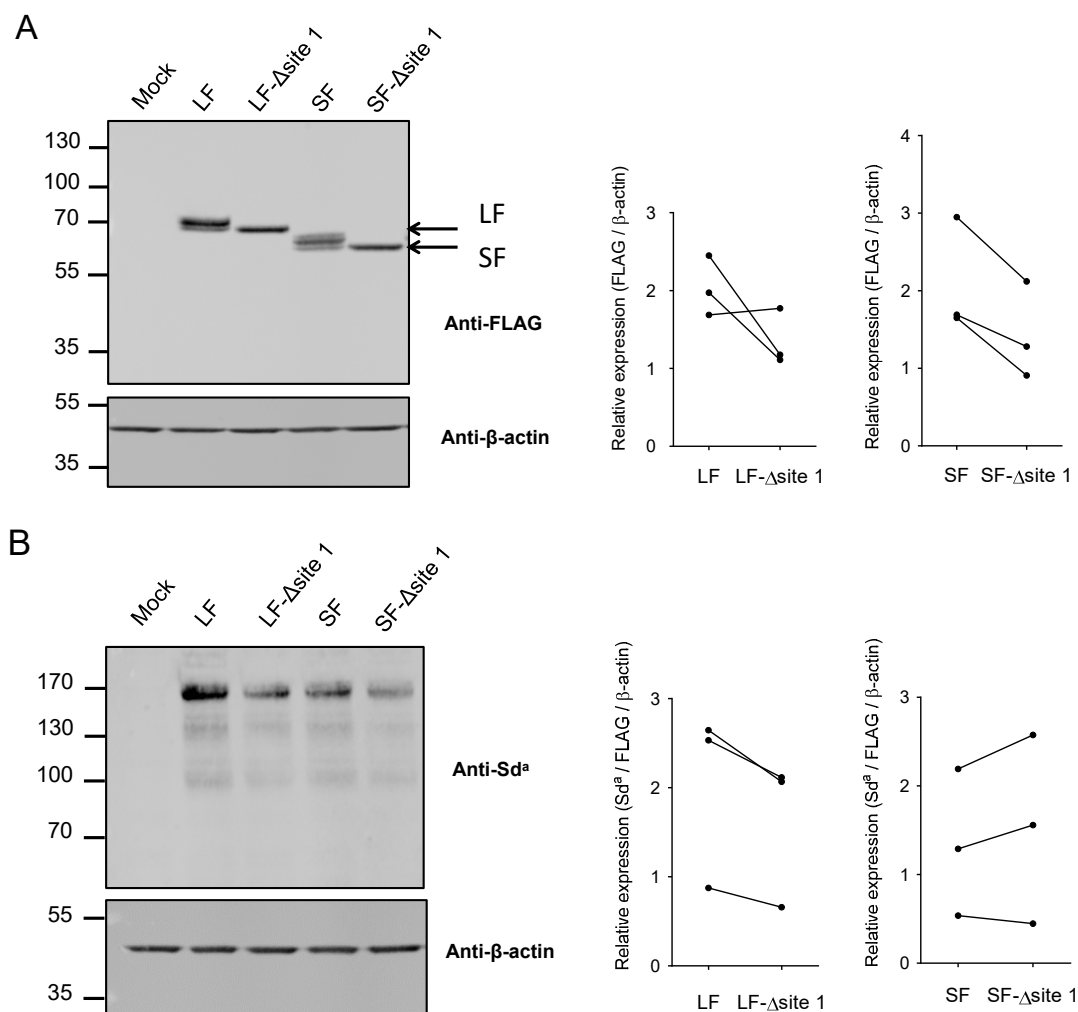


**Figure 3.** Partial PNGase F treatment on B4GALNT2. Serial dilutions (1/20<sup>e</sup>, 1/50<sup>e</sup>, or 1/100<sup>e</sup>) of PNGase F were used on mouse (A) or on rat (B) or on human (C) colon extracts. After SDS-PAGE and transfer on nitrocellulose membranes, immunoblotting with the anti-B4GALNT2 antibody reveals the presence in vivo of two, three, or one *N*-glycans on mouse, rat, and human B4GALNT2, respectively. PNGase F digestion of proteins from LS174T cells stably transfected with the SF-B4GALNT2 (LS174T-S2) and immunoblotting show a similar profile of glycosylation of the human B4GALNT2 (C). Finally, a representative immunoblot carried out with an anti-FLAG antibody shows a similar effect of PNGase F digestion on the LF-B4GALNT2 and SF-B4GALNT2 FLAG-tagged recombinant proteins transiently expressed in HeLa cells (D).

### 2.3. Impact of *N*-Glycan on B4GALNT2 Activity and Sd<sup>a</sup> Synthesis

To gain insights into the role of this unique *N*-glycan in the human enzyme, we used site-directed mutagenesis substituting asparagine with glutamine residues N77Q in SF-B4GALNT2 and N137Q in LF-B4GALNT2, respectively, thereby eliminating the *N*-glycan on *N*-glycosylation site 1. Each construct was transiently transfected in HeLa cells in three individual experiments, and expression of each isoform was assessed by WB with the anti-FLAG antibody and normalized to  $\beta$ -actin. As illustrated in Figure 4A for one representative experiment (left panel), the absence of *N*-glycan slightly affected relative B4GALNT2 protein expression for both isoforms. Expression of the Sd<sup>a</sup> antigen on glycoproteins in transfected HeLa cells was assessed by WB using the anti-Sd<sup>a</sup> antibody KM694 (Figure 4B) and normalized to enzyme and  $\beta$ -actin expression. Sd<sup>a</sup> expression was found to be slightly but not significantly reduced only in cells transfected with the mutated LF-B4GALNT2 form (Figure 4B).

We also assessed the enzymatic activity of each full-length recombinant protein produced in COS-7 using in vitro assays. Interestingly, normalization of enzymatic activity for relative B4GALNT2 protein expression revealed a clear tendency to lower activity for mutant forms, which was more pronounced and statistically significant for the short form (Figure 5).

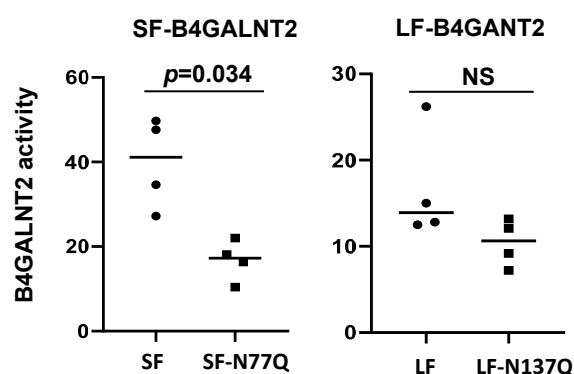


**Figure 4.** Effect of asparagine mutations (SF-Δsite1 and LF-Δsite 1) on B4GALNT2 production and activity in transfected HeLa cells. HeLa cells were transiently transfected with either an empty vector (Mock), the pFLAG-LF-B4GALNT2 (LF), and the pFLAG-SF-B4GALNT2 (SF) and their mutated counterparts pFLAG-LF-B4GALNT2-N137Q (LF-Δsite 1) and pFLAG-SF-B4GALNT2-N77Q (SF-Δsite 1). Total cell lysates were prepared and subjected to SDS-PAGE, and WB was performed with the anti-FLAG, anti-Sd<sup>a</sup>, and anti-β-actin antibody. Expression of the recombinant B4GALNT2 is shown 24 h after transfection (**A**). The Sd<sup>a</sup> profile obtained after 24 h transfection is shown in (**B**). WB shown on the left side of the figure is one representative dataset of three distinct transfection experiments. ImageJ was used to compare the density of bands detected on WB. On the right side are represented quantifications of the expression level of B4GALNT2 isoforms normalized with β-actin and of Sd<sup>a</sup> normalized relative to B4GALNT2/β-actin.

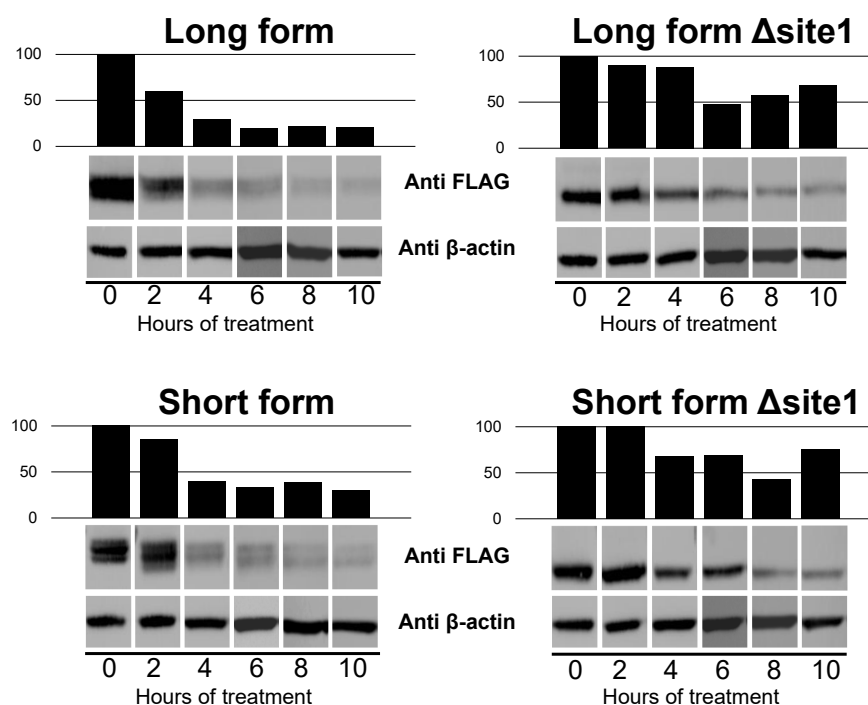
#### 2.4. Impact of N-Glycan on Dimerization of B4GALNT2 Isoforms

We first assessed the impact of this unique N-glycan on protein stability. For that purpose, both the native SF-B4GALNT2 and LF-B4GALNT2 and their corresponding glycosylation mutants were expressed in HeLa cells in the presence of the protein synthesis inhibitor anisomycin for 2 to 10 h. As shown in Figure 6, our data indicated that N-glycosylation of the two B4GALNT2 isoforms had an impact on protein stability. For all time points, the mutated LF-B4GALNT2 and SF-B4GALNT2 forms were more expressed than their wild-type counterparts ( $p = 0.0023$  and  $0.02$ , respectively). These data strongly suggest that the N-glycan reduces the B4GALNT2 isoforms half-life. We also confirmed that with or without N-glycosylation, each B4GALNT2 isoforms transiently expressed in HeLa cells were not secreted (data not shown).



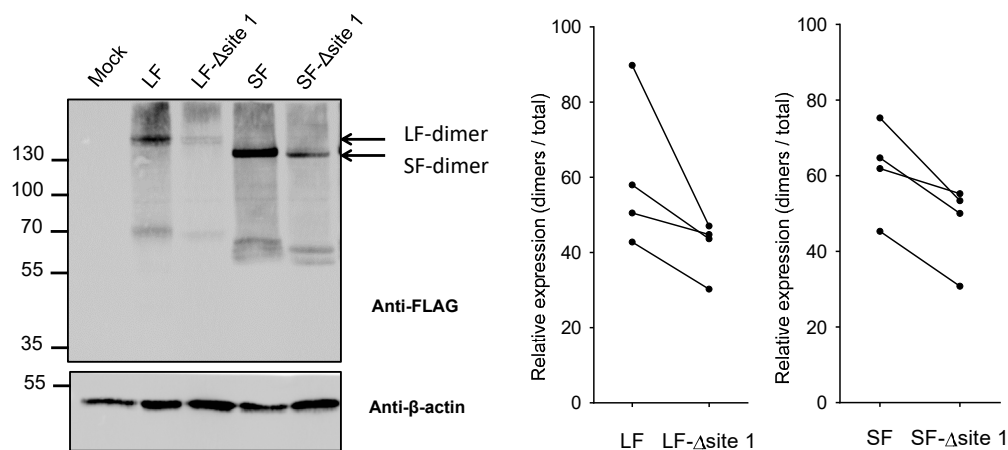


**Figure 5.** Normalized B4GALNT2 activity. COS-7 cells were transiently transfected with either an empty vector (Mock), the pFLAG-LF-B4GALNT2 (LF), and the pFLAG-SF-B4GALNT2 (SF) and their mutated counterparts pFLAG-LF-B4GALNT2-N137Q (LF- $\Delta$ site 1) and pFLAG-SF-B4GALNT2-N77Q (SF- $\Delta$ site 1). Total cell lysates were prepared and subjected to SDS-PAGE, and WB was performed with the anti-B4GALNT2 and  $\beta$ -actin antibody for quantification. Total cell lysates were used as an enzyme source for B4GALNT2 enzymatic assays using UDP- $^{3}\text{H}$ GalNAc and fetuin or asialofetuin as acceptors, as described previously [19]. The SF- $\Delta$ site 1 enzyme shows significantly reduced enzymatic activity compared to the SF isoform (**left side**), whereas the LF- $\Delta$ site 1 shows not significantly (NS) reduced activity compared to LF isoform (**right side**). Normalized data of B4GALNT2 activity in four experiments from three distinct transfections with statistical analysis (*t*-test for paired samples) are presented here; NS = not significant.



**Figure 6.** Effect of *N*-glycosylation site mutation on B4GALNT2 stability. HeLa cells were transiently transfected with pFLAG-LF-B4GALNT2 (LF), pFLAG-SF-B4GALNT2 (SF), and their mutated glyco-variants pFLAG-LF-B4GALNT2-N137Q (LF- $\Delta$ site1) and pFLAG-SF-B4GALNT2-N77Q (SF- $\Delta$ site1). Twenty-four hours after transfection, cells were cultured in DMEM supplemented with 10% FBS and incubated with 5  $\mu\text{g}/\text{mL}$  anisomycin for 2 to 10 h. After treatment, total cell lysates were prepared and subjected to SDS-PAGE, and WB was performed with the indicated antibodies. Graphs represent the intensity of the anti-FLAG signals normalized for the respective anti- $\beta$ -actin signals. Statistical analysis by Student's *t*-test for paired samples indicated that the intensity of the  $\Delta$ site1 signals was statistically higher than that of the non-mutated form for both the long form ( $p = 0.0023$ ) and the short form ( $p = 0.02$ ).

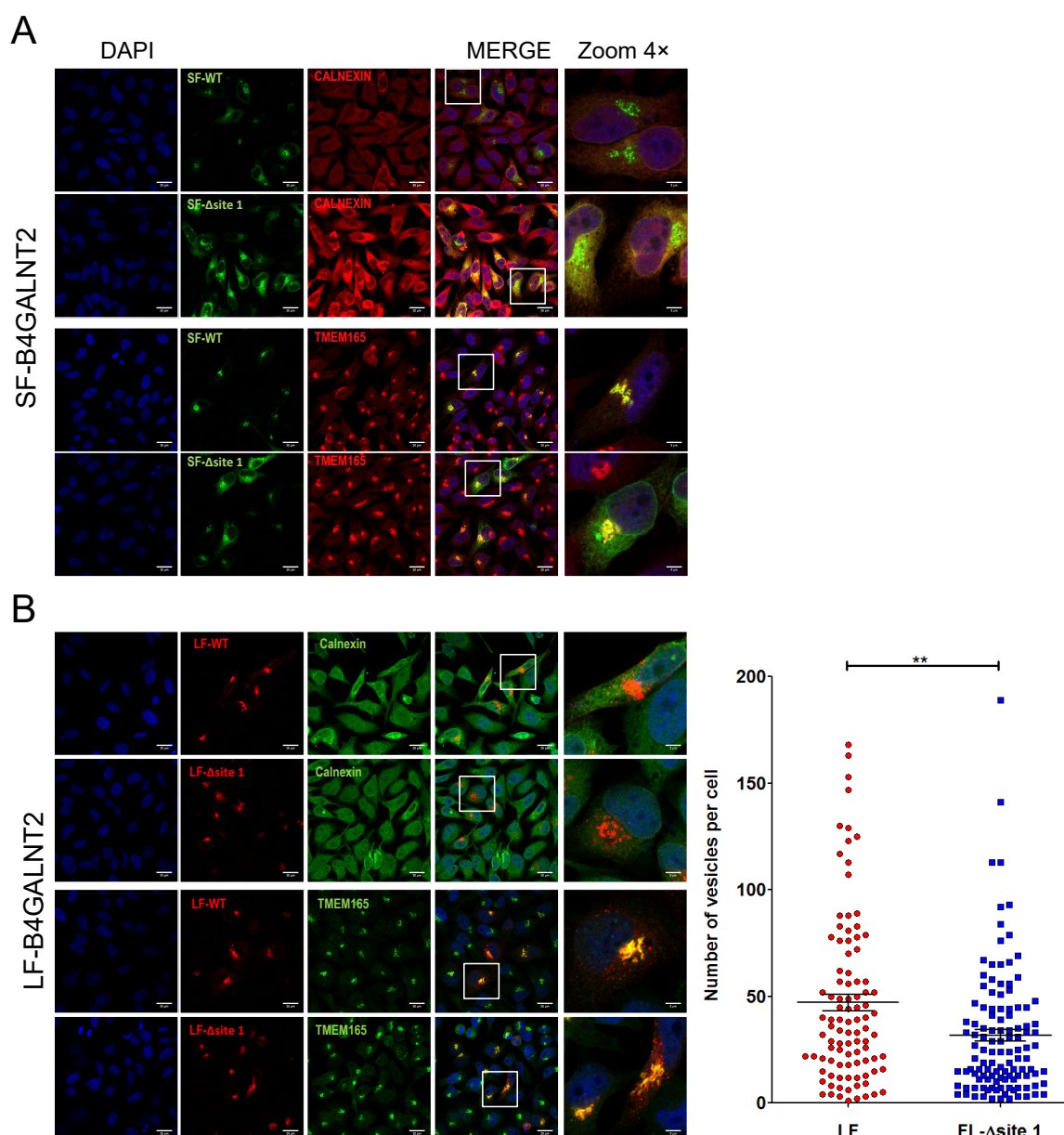
As glycosyltransferases potentially form enzymatically active homo- and heterodimers, we next investigated oligomerization of each B4GALNT2 glycovariant expressed in HeLa cells on non-reducing SDS-PAGE and WB. As illustrated in Figure 7 for one representative experiment (left panel), our data consistently show that eliminating the *N*-glycan decreased protein production and drastically impacted the relative formation of dimers for the LF-B4GALNT2 isoform and to a lesser extent for the SF-B4GALNT2 (Figure 7, right panel).



**Figure 7.** Effect of asparagine mutations (SF- $\Delta$ site 1 and LF- $\Delta$ site 1) on B4GALNT2 dimers formation. HeLa cells were transiently transfected with either an empty vector (Mock), the pFLAG-LF-B4GALNT2 (LF), and the pFLAG-SF-B4GALNT2 (SF) and their mutated counterparts pFLAG-LF-B4GALNT2-N137Q (LF- $\Delta$ site 1) and pFLAG-SF-B4GALNT2-N77Q (SF- $\Delta$ site 1). Total cell lysates were prepared and subjected to non-reducing SDS-PAGE, and WB was performed with the indicated antibodies. The figure shows a representative immunoblot of the B4GALNT2 dimers profile obtained after 24 h transfection. ImageJ was used to compare the density of bands detected on WB. On the right side are represented the relative expression level of B4GALNT2 isoforms dimers relative to dimers + monomers of B4GALNT2 isoforms/ $\beta$ -actin.

### 2.5. Subcellular Localization of B4GALNT2 Isoforms

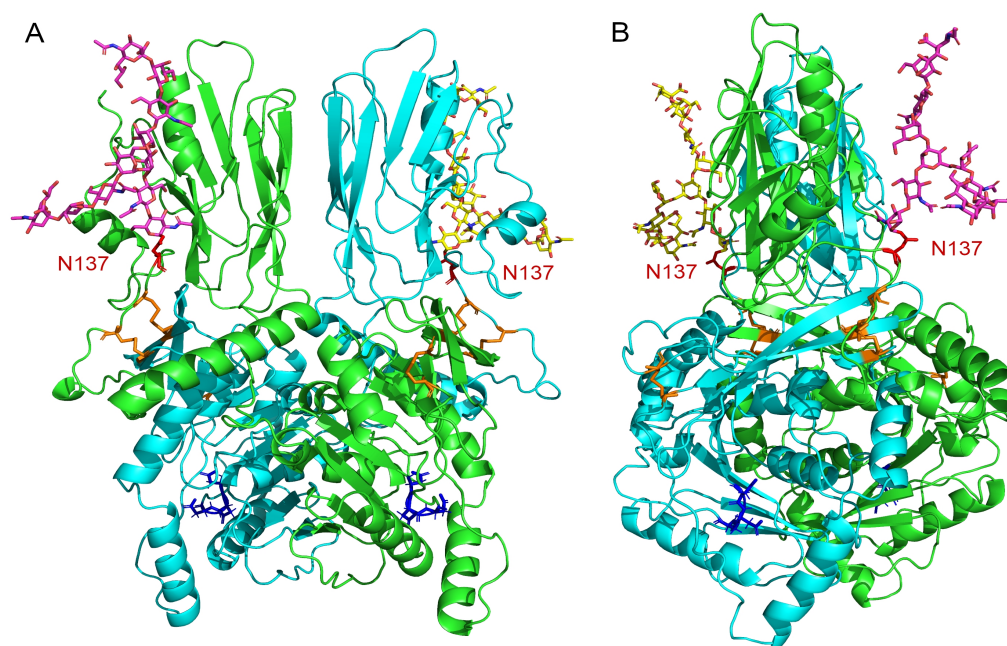
It was previously shown that the two B4GALNT2 isoforms localized in the Golgi and that LF-B4GALNT2 showed additional post-Golgi vesicle localization because of its extended cytoplasmic tail [24]. To assess the impact of *N*-glycan of B4GALNT2 onto its subcellular localization, we used fusion full-length B4GALNT2 C-terminal fluorescently tagged proteins. The SF-B4GALNT2 glycovariants were coupled with green fluorescent protein (eGFP), while the LF-B4GALNT2 glycovariants were coupled with mCherry. We then performed immunostaining of HeLa cells overexpressing the various SF- and LF-B4GALNT2 fluorescent glycovariants with the Golgi and ER markers TMEM165 [34] and Calnexin, respectively, and analyzed subcellular localization of each protein using confocal microscopy (Figure 8). As previously reported, SF-B4GALNT2 co-localized with the Golgi marker TMEM165 [24]. By contrast, the SF-B4GALNT2 glyco-mutant displayed a characteristic pattern of cytoplasmic expression, which included but was not limited to Golgi. In fact, this pattern of expression largely colocalized with the ER marker Calnexin (Figure 8A). Interestingly, the expression of Calnexin, which is a molecular chaperone involved in the unfolded protein response, appeared to be increased in cells expressing this mutant. These data indicate that the absence of *N*-glycan profoundly alters the intracellular distribution of SF-B4GALNT2 and suggest that the accumulation of the unglycosylated form induced Calnexin overexpression. Surprisingly, the LF-B4GALNT2 glyco-mutant was mostly not affected by the absence of *N*-glycan, as it was found in the Golgi membranes and in post-Golgi vesicles, as described previously for LF-B4GALNT2 [24]. However, vesicles number counted per cell was significantly decreased in the non-glycosylated LF-B4GALNT2-transfected cells (Figure 8B).



**Figure 8.** Subcellular localization of fluorescently labeled full-length B4GALNT2 glyco-variants in transfected HeLa cells using confocal microscopy. The SF-B4GALNT2 and its glyco-mutant SF-Δsite 1 were tagged with eGFP (green), and the LF-B4GALNT2 and its glyco-mutant LF-Δsite 1 were tagged with mCherry (red). Each tagged construct was transfected individually in HeLa cells. Their co-localization with ER and Golgi compartment markers Calnexin and TMEM165 (red in (A) and green in (B)) was examined by confocal microscopy. As previously reported [24], the SF-B4GALNT2 (A) and LF-B4GALNT2 (B) mostly co-localize with TMEM165 in the Golgi membranes, and LF-B4GALNT2 showed an additional localization in post-Golgi vesicles. The glyco-mutant SF-Δsite 1 (A) colocalized partially with Calnexin and TMEM165, whereas the LF-Δsite 1 (B) showed mostly the same subcellular localization as its non-mutated counterpart. The total number of red vesicles distinct from Golgi compartments was counted per cell and quantified using the ICY software (version ICY: 2.3.0.0, <http://icy.bioimageanalysis.org>, accessed on 5 November 2021), as previously described [24]. Data are plotted in whiskers plots, where means and SEMs are depicted as shown on the right side, panel B; they indicate a little but significant decrease in vesicles number in the LF-Δsite 1 transfected cells, as assessed by the Kruskal–Wallis non-parametric test with selected comparison using Dunn’s post hoc test. Scale bars: 5 μm; the white squares indicated the position of the 4× zoom image shown on the right side; \*\* represent the level of significance and correspond to  $p = 0.0019$ .

### 2.6. Model of the B4GALNT2 Structure

To gain insights into the potential role of this *N*-glycan on B4GALNT2 dimerization from a 3D-structure representation of B4GALNT2, we built a model of the dimer exhibiting the *N*-glycan on site 1. For that purpose, we focused on the core protein shared by the LF- and SF-B4GALNT2 isoforms, removing the first 111 residues of LF-B4GALNT2 encompassing the cytoplasmic tail, a flexible domain with respect to the rest of the protein (see DynaMine backbone flexibility prediction of Supplementary Figure S4). Since this *N*-terminal part shows the worse confidence in the structural model of the AlphaFold database [35] for the *Homo sapiens* LF-B4GALNT2 (Supplementary Figure S4B), we built a model of the dimer with AlphaFold2 for the 112–566 sequence. We observed that the site 1 asparagine residues of both monomers are very accessible in the model: relative accessibility is 84.1% for chain A (ND2 accessibility = 58.311 Å<sup>2</sup>) and 82.6% for chain B (ND2 accessibility = 58.168 Å<sup>2</sup>), leaving a large space for a modification of this amino acid residue. We then further grafted a complex-type *N*-glycan on site 1 of both monomers (see Section 4 for details), which is shown in Figure 9. Even if the *N*-glycans are big, they are not long enough to block the access to the catalytic pocket. Interestingly, the asparagine residue bearing the *N*-glycan is located in a region that is not so flexible (Supplementary Figure S4) and is close to disulfide bridges, which limits the movements of the *N*-glycan to its own motions. Measurements in the structure show that the size of the *N*-glycan is inferior to 40 Å (~37 Å), while the distance between the N137 and the entry of the catalytic pocket is superior to 40 Å (~43 Å). Therefore, the two *N*-glycans in the dimer cannot sterically inhibit the enzymatic activity, which is in accordance with our experimental results.



**Figure 9.** Three-dimensional model representation of B4GALNT2 dimer generated with AlphaFold2. (A) The two monomers are in green and cyan, and the *N*-glycosylation site 1 is depicted in red on both monomers, with a complex-type *N*-glycan depicted in yellow and pink (structure of the *N*-glycan described in Section 4); inter- and intra-monomeric disulfide bridges are colored orange, and the D-X-D motif in the catalytic cleft are shown in blue; (B) is a 90° rotation on the *y*-axis of (A).

### 3. Discussion

Bioinformatics approaches for *N*-glycosylation sites prediction are based on recognition of the sequon N-X-T/S. However, the presence of such a typical sequence is neither necessary nor sufficient for *N*-glycosylation. Although 96.5% of *N*-glycosylation sites of mammalian proteins are found at this canonical site, minor and less predictable *N*-

glycosylation sites in glycoproteins such as cysteine-proximal acceptor sites N-X-C (1.3%) or N-G (0.5%) or N-X-V (0.4%) have been previously reported in a mouse *N*-glycoproteome study and are referred to as unusual glycosylation sites [33]. In this study, additional potential *N*-glycosylation sites in the mammalian B4GALNT2 sequences were identified that could not be predicted earlier [20,27]. Two sites, site 1 and site 2, were identified in the human SF- and LF-B4GALNT2 sequences and in the mouse B4GALNT2, and an additional typical *N*-glycosylation site was identified in the rat B4GALNT2 sequence (Figure 2). Analysis of the evolutionary conservation of these *N*-glycosylation sites in animal B4GALNT-related sequences (e.g., B4GALNT1 and B4GALNT2 sequences of the CAZy family GT12) revealed that the unusual site 1 was likely present in the B4GALNT ancestor since it is observed after vertebrate radiation in both B4GALNT1 and B4GALNT2 sequences. It was highly conserved in all the vertebrate B4GALNT1 sequences and was lost in the B4GALNT2 sequences in the fishes and amphibians branches during vertebrate evolution (Figure 2), further suggesting a solvent-accessible and functionally constrained site [36]. The unusual *N*-glycosylation site 2 N-G was not conserved in all B4GALNT-related sequences; in particular, it is absent in all the vertebrate B4GALNT1 sequences (Figure 2). This observation further suggested that this site was acquired in the common ancestor of mammals (Figure 2) and may have resulted in functional innovation for the mammalian B4GALNT2 since it was fixed during mammalian evolution. Located at the interface between catalytic and stem domains, this *N*-glycan could play an important role in the dimerization of B4GALNT2. Another *N*-glycosylation site was predicted in a canonical sequon N-[!P]-S/T-[!P] in the catalytic domain of the rat B4GALNT2, which is nearby the D-X-D motif (Figure 2). This site is not conserved in the majority of vertebrate species with the notable exception of ferret and elephant B4GALNT2 sequences (data not shown). Furthermore, it could not be predicted in primate, mouse, amphibian, or fish B4GALNT2 sequences, and this position is also not conserved in B4GALNT1 sequences (Figure 2). The gain of this site restricted to rat B4GALNT2 may affect the structure and molecular function of the rat enzyme, therefore conferring novel traits to this species.

A previous study reported the existence of three *N*-glycosylation sites in the human B4GALNT1 sequence (N79, N179, and N274) [29]. The first-described *N*-glycosylation site N79 is found in a typical sequon N-X-T/S and is conserved only in the mammalian B4GALNT1 sequences. It is juxtaposed to the highly conserved and unusual site 1 described in this study (N-N-C-T/S), where N78 could be *N*-glycosylated. The two other *N*-glycosylation sites are conserved in most vertebrate B4GALNT1 sequences but not in the B4GALNT2 sequences (Figure 2).

We then investigated whether these predicted *N*-glycosylation sites in human B4GALNT2 could be occupied *in vivo*. Protein extracts from human, mouse, and rat large intestine were treated with serial dilutions of PNGase F and run on SDS-PAGE. Immunoblotting with an anti-B4GALNT2 antibody revealed two, three, or four bands with a ~2.5 kDa molecular mass shift likely corresponding to the various glycoforms of B4GALNT2 in human, mouse, and rat colon, respectively (Figure 3). To determine whether all potential *N*-glycosylation sites could be used, various FLAG- and GFP-tagged recombinant proteins truncated or not in their catalytic domain were produced in HeLa cells. Our data demonstrated that the two human B4GALNT2 isoforms harbor a unique *N*-glycan located on the N-X-C site 1 (Figure 3 and Supplementary Figure S3). As frequently observed in other glycosyltransferases such as ST6GAL1 [37], the site 2 in human B4GALNT2 remained unglycosylated *in vivo* although the SF-B4GALNT2 isoform produced in HeLa cells sometimes shows little amounts of a second glycoform with two *N*-glycans (Figures 3, 4 and 7). Interestingly, the mouse B4GALNT2 that shares the same conserved *N*-glycosylation sites 1 and 2 appears to have a higher amount of the glycoform with two *N*-glycans in colon (Figure 3). In the mouse sequence, site 2 is a conventional sequon (N-G-S) and consequently likely occupied with an *N*-glycan compared to the rat sequence (N-G-D) or in the human sequence (N-G-A). Finally, further digestion assays using PNGase F and Endo H indi-

cated the presence of a complex-type *N*-glycan on human B4GALNT2 site 1 (Figure 3 and Supplementary Figure S2).

It has long been known that *N*-glycosylation of proteins is a dynamic process involved in their folding, maturation, and trafficking in the secretory pathway [25]. Most of the resident Golgi glycosyltransferases involved in the last glycosylation steps of proteins and lipids are *N*-glycosylproteins themselves, and their *N*-glycan(s) are often associated with their enzymatic activity although the effects vary from protein to protein [26]. In the sialyltransferase family, for instance, the rat ST6GAL1 *N*-glycosylation on N146 and N158 is not an absolute requirement for  $\alpha$ 2,6-sialylation in vivo [37]. Similarly, for the human ST3GAL1, none of the *N*-glycans found on N27, N79, N114, N201, and N323 are essential to the enzymatic activity assayed in vitro [38,39]. However, the *N*-glycan found on the evolutionary conserved N211 glycosylation site of human ST3GAL2 dramatically affects in vitro activity of the enzyme, likely through influencing substrates recognition or enzyme folding [40]. In this study, we used biochemical methods and site-directed mutagenesis to evaluate the contribution of the *N*-glycan of the evolutionary conserved site 1 to the biological function of the two human B4GALNT2 isoforms. Elimination of this *N*-glycan in each isoform resulted in a moderate effect on the amount of each protein and Sd<sup>a</sup> epitope formed, as assessed by WB and illustrated in Figure 4. However, normalized enzymatic activity assessed in a cell-free system was significantly lower for the unglycosylated short form (Figure 5). A similar tendency was observed, although it was less pronounced, for the long form (Figure 5).

To gain insights into the molecular mechanisms responsible for the decreased activity, we also investigated subcellular localizations, secretion, and potential dimerization of each B4GALNT2 isoform. Several studies have highlighted the crucial role of *N*-glycans in correct subcellular localizations and/or in facilitating and stabilizing proper protein folding [26], such as for the murine ST3GAL5 and chicken and murine ST8SIA1, which were unable to exit the ER once deglycosylated [41–43]. Interestingly, in contrast to the human B4GALNT1, whose intracellular localization was not affected by the absence of *N*-glycans [29], we showed in this study that the absence of the *N*-glycan on the SF-B4GALNT2 impaired its exit from the ER and Golgi localization (Figure 8A). Concomitantly, in cells expressing the mutated short form, we observed an overexpression of the molecular chaperone Calnexin, which could be indicative of the activation of the unfolded protein response. This possibility, which requires further investigations to be demonstrated, would indicate that without *N*-glycosylation, the short form fails to adopt the proper folding and to pass the cellular quality control mechanisms. On the other hand, the absence of the *N*-glycan on the LF-B4GALNT2 had almost no effect on its subcellular localization (Figure 8B). This later observation can be explained by the existence of a strong signal in the extended cytoplasmic tail driving Golgi targeting and post-Golgi sorting of the LF-B4GALNT2, which we previously described [3,24].

Most Golgi-glycosyltransferases are known to assemble into homodimers and heterodimers in vivo [44–47]. Previous protein chemistry experiments and mass spectrometry analysis have shown that all cysteine residues of the B4GALNT1 were involved in intra- and inter-disulfide bonds responsible for the formation of homodimers in an antiparallel orientation [48]. These cysteine positions are highly conserved in the human parologue B4GALNT2 and are likely engaged in intra- and inter-disulfide bonds, fitting perfectly with the proposed dimer model (Figure 9). Because of the cysteine-rich context found for site 1, we also assessed dimer formation and the impact of *N*-glycan loss on the dimer formation. We showed that most of the full-length recombinant B4GALNT2 isoforms produced in HeLa cells are dimerized, whereas the lack of *N*-glycan at site 1 largely decreased homodimer formation (Figure 7). This data further suggested the crucial role of this *N*-glycan in dimerization, proper folding and enzyme activity of each B4GALNT2 isoform.

To gain insights into the structure/function relationships of B4GALNT2, we used Alphafold2 to produce a model of the LF-B4GALNT2 dimer deleted of the first 111 amino acid residues, and we grafted two *N*-glycans: one on site 1 of each monomer (Figure 9). The

scores given by the AlphaFold2 modelling software (i.e., confidence score called ipTM + pTM) were very high, and the five models produced were highly homogeneous. However, the first-ranked model was slightly different, exhibiting only intra-monomeric disulfide bridges, while the four others showed inter- and intra-monomeric disulfide bridges, as experimentally shown for B4GALNT1 [48]. We selected one of the later for the addition of an *N*-glycan on the site 1 of each monomer, which is located in a region that is not flexible and readily accessible to the oligosaccharyltransferase (OST) for core *N*-glycan transfer (Supplementary Figure S4). As observed on this *N*-glycosylated B4GALNT2 dimer, the *N*-glycan, even if long and flexible, cannot physically block the substrate's access to the catalytic pocket of the enzyme. In fact, the presence of the *N*-glycan not only does not hamper enzyme activity but, on the contrary, slightly increases it. It is interesting to put into perspective here the work of Gilmore's laboratory showing that mammalian cells likely use two mechanisms for the *N*-glycosylation of proteins. The first one involves the OST complex SEC61-STT3A, which mediates co-translational glycosylation, while the second one involves the STT3B complex acting in a posttranslocational mode on a subset of cysteine-rich *N*-glycosylation sites skipped by STT3A [49,50]. In light of this work, dimerization of B4GALNT2 likely occurs concomitantly to *N*-glycosylation of site 1 in the ER. Our model shows that dimerization and *N*-glycosylation are both possible at the same time because no sterical hindrance can be observed, with the dimerization leaving space for the activity of the STT3B complex and vice versa.

Unlike the biosynthesis of DNA, RNA, and proteins, which are deterministic processes, glycosylation is intrinsically stochastic, being the product of the cooperative and competitive interaction of glycosyltransferases, glycosidases, and other enzymes organized along the secretory pathway. In this study highlighting the control of the glycosyltransferase B4GALNT2 by its own *N*-glycosylation on an atypical *N*-glycosylation site, we provide detailed evidence of how glycosylation controls itself, further enlightening the extreme complexity of the process.

## 4. Materials and Methods

### 4.1. Bioinformatic Analysis of B4GALNT Protein Sequences

The human B4GALNT1 (NM\_001478) and B4GALNT2 (AJ517770 (SF) and AJ517771 (LF)) protein sequences were used as seed sequences for the identification of 187 B4GALNT homologues in the NCBI database using BLAST. B4GALNT sequences from the mammals *Mus musculus* (NM\_001478 and BC139166), *Rattus norvegicus* (D17809 and XM\_039087862), *Sus scrofa* (AK237665 and NM\_001244330), the squamata *Gekko japonicus* (XM\_015423486 and XM\_01509212), the amphibian *Silurana tropicalis* (XM\_012956468 and XM\_002935708), the non-teleostean fish *Lepisosteus oculatus* (XM\_006629420 and XM\_015362122), the teleostean fish *Salmo salar* (BT058789 and XM\_01404566), and the amphioxus *Branchiostoma floridae* (XM\_002601302) were selected for multiple sequence alignment using Clustal W at PRABI and *N*-glycosylation sites prediction. The online predictors of *N*-glycosylation sites in proteins NetNGlyc was used (<https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0>, accessed on 29 April 2022) [51].

### 4.2. Antibodies and Reagents

Anisomycin, the polyclonal rabbit anti-B4GALNT2 antibody (HPA015721), the monoclonal mouse anti-GFP (G1546), the monoclonal mouse anti-Flag M2 (F3165), the polyclonal rabbit anti-TMEM165, and the mouse anti- $\beta$ -actin antibodies were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). Anti-Sd<sup>a</sup> (KM694) antibody was a kind gift from Dr. Taeko Dohi and Dr. Akiko Furuya (Biologics Research Laboratories, Research Division, Kyowa Hakko Kirin Co., Tokyo, Japan). The polyclonal goat anti-mouse horseradish peroxidase-conjugated IgG was purchased from Agilent Technologies (Santa Clara, CA, USA). The polyclonal goat anti-mouse horseradish peroxidase-conjugated IgM was purchased from Fisher Scientific (Illkirch, France). The polyclonal goat anti-rabbit horseradish peroxidase-conjugated IgG was purchased from Sigma-Aldrich (Saint Quentin Fallavier,

France). The secondary antibodies Alexa fluor<sup>®</sup> 488 anti-rabbit and Alexa fluor<sup>®</sup> 568 anti-rabbit were from Life Technologies (Fisher Scientific, Illkirch Graffenstaden, France). The polyclonal rabbit anti-Calnexin was purchased from Enzo life technologies (Villeurbanne, France). The GFP expression vector pF<sub>x</sub> was a kind gift of Dr. Jack Röhrer (University of Zurich), the expression vectors pEGFP-N1 and pmCherry-N1 were from Clontech, whereas p3×FLAG-CMV10 was from Sigma.

#### 4.3. Plasmid Constructions of Human Long and Short B4GALNT2 Isoforms and Mutants

Full-length SF- and LF-B4GALNT2 cDNAs (AJ517770 and AJ517771) were obtained previously [20]. Full-length SF- and LF-B4GALNT2 were amplified by PCR using oligonucleotide primers containing *NotI-XbaI* restriction sites for directed ligation in a *NotI-XbaI* p3×FLAG-CMV10 digested vector (Supplementary Table S1). For the production of fluorescent chimera proteins, full-length SF- and LF-B4GALNT2 were amplified by PCR using oligonucleotide primers containing *NheI-BglI* or *NheI-BamHI* restriction sites, respectively, for directed ligation in *NheI-BglI* or *NheI-BamHI* pEGFP or pmCherry digested vectors (Table S1). The three truncated SF-B4GALNT2 constructs ( $\Delta$ site 1 $\Delta$ 262–506,  $\Delta$ site 2 $\Delta$ 97–506, and  $\Delta$ site 1&2 $\Delta$ 80–506) deleted of the last 144, 409, and 426 amino acids in the catalytic domain were amplified by PCR using oligonucleotide primers containing *NheI-BglIII* (SF-B4GALNT2 $\Delta$ 262–506, SF-B4GALNT2 $\Delta$ 97–506) or *NheI-BamHI* (SF-B4GALNT2 $\Delta$ 80–506) restriction sites for directed ligation in a *NheI-BglIII* or *NheI-BamHI* pF<sub>x</sub> or pYFP digested vector (Table S1), as previously reported [24].

The p.N77Q and p.N137Q mutations in the SF- and LF-B4GALNT2 were obtained as follows: Two PCR amplifications (PCR1 and PCR2) were carried out with couples of oligonucleotide primers, one of which was carrying the codon mutation (underlined), i.e., the forward primer GCCCGCCTGGCATTATGC and the reverse primer GCTTCA-CATTTGCACTGCTGTTTCGGGAACAGCCAG (PCR 1) and the forward primer CTGGCT-GTTCCCGAAACAGCAGTGCAAATGTGAAGC and the reverse primer GCTTCTGGCT-GTCATCAGCCAC (PCR2), shown in Table S1. The two amplicons from PCR1 and PCR2 were mixed and subjected to a third PCR with primers GCCCGCCTGGCATTATGC and GCTTCTGGCTGTCATCAGCCAC. The resulting cDNA fragment was digested by *XhoI* and *BspEI* present in the B4GALNT2 cDNA from either side of the mutation and introduced by fragment replacement in the p3×FLAG-CMV10 containing SF- or LF-B4GALNT2 digested with *XhoI* and *BspEI*. Similarly, these mutations (p.N77Q or p.N137Q) were introduced in the pEGFP-SF-B4GALNT2 or in pmCherry-LF-B4GALNT2 constructs by fragment replacement using restriction sites *HindIII* and *EcoRI*.

#### 4.4. Cell Culture and Transfections

HeLa cells (ATCC CCL-2, LGC Standards SARL, Molsheim, France) and COS-7 cells (ATCC CRL-1651) were routinely grown in Dulbecco's modified eagle's medium (DMEM) (Biowest, Nuaille, France) supplemented with 10% fetal bovine serum (Biowest, Nuaille, France) and maintained at 37 °C in a humidity-saturated 5% CO<sub>2</sub> atmosphere.

The transfections of HeLa and COS-7 cells were performed using Lipofectamine 2000 transfection reagent (ThermoFisher Scientific, Waltham, MA, USA). Cells were grown in six-well plates until they reached 70–80% confluency, washed twice with UltraMEM (Lonza, Basel, Switzerland), and then transfected for 24 h with 2 µg of plasmid DNA and 4 µL of lipofectamine. After 4 h, the transfections were stopped by removal of the transfection mixture and the addition of fresh culture medium. When used, anisomycin (Sigma) was added 24 h after transfection for 2 to 10 h at the final concentration of 5 µg/mL.

The cell culture media were harvested 24 h post transfection. After methanol/chloroform precipitation, protein pellets were suspended in RIPA buffer (Tris/HCl 50 mM pH 7.9, NaCl 120 mM, NP 40 0.5%, EDTA 1 mM, Na<sub>3</sub>VO<sub>4</sub> 1 mM, NaF 5 mM) supplemented with a protease cocktail inhibitor (Roche Diagnostics, Penzberg, Germany). SDS-PAGE was performed on 4–12% Bis-Tris gels (Fisher Scientific, Illkirch, France), and the membrane was blotted using the polyclonal rabbit anti-B4GALNT2 antibody.



#### 4.5. Western Blot Analysis

Cells were scraped in Dulbecco's Phosphate Buffer (DPBS, Lonza) and then centrifuged at 6000 rpm, 4 °C for 10 min. Supernatant was discarded, and cells were then suspended in RIPA buffer supplemented with a protease cocktail inhibitor, as described above. Cell lysis was performed by passing the cells several times through a syringe with a 26 G needle. Cells were centrifuged for 30 min, 4 °C, at 14,000 rpm. The protein concentration contained in the supernatant was estimated with the micro BCA Protein Assay Kit (Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Then, 10 µg of total protein lysate were mixed with NuPAGE Lithium Dodecyl Sulfate (LDS) Sample Buffer (Fisher Scientific, Waltham, MA, USA) pH 8.4 supplemented with 4% β-mercaptoethanol (Sigma-Aldrich, Saint Louis, MO, USA). Samples were heated 5 min at 95 °C and then separated on 4–12% Bis-Tris gels (Fisher Scientific, Waltham, USA) and transferred to nitrocellulose membrane Hybond ECL (GE Healthcare, Little Chalfont, UK). Membranes were blocked using TBS containing 0.05% Tween 20 (TBS-T) and 5% (*w/v*) non-fat dried milk or a Blocking Reagent (Sigma-Aldrich, Saint Louis, MO, USA) for at least 1 h at RT. Primary antibodies mouse anti-FLAG, mouse anti-β-actin, mouse anti-GFP, and mouse anti-Sd<sup>a</sup> were incubated overnight at 4 °C in TBS-T and 5% (*w/v*) non-fat dried milk or TBS-T at respectively 1:1000, 1:10,000, 1:3000, and 1:500 dilution. After three TBS-T washes, membranes were then incubated with the appropriate peroxidase-conjugated secondary antibody (goat anti-mouse IgG from Agilent or goat anti-mouse IgM from Fisher Scientific, used at a dilution of 1:10,000 or 1:20,000 in TBS-T and 5% (*w/v*) non-fat dried milk or TBS-T) for 1 h at RT. After five TBS-T washes, blots were developed using enhanced chemiluminescence (West Pico Plus, ThermoScientific). The images were acquired using a CCD camera (Fusion Solo, Vilbert Lourmat, Collégien, France).

#### 4.6. Total Proteins Preparation from Various Mammalian Gastrointestinal Tissues and Immunochemical Analyses

Total proteins extraction from frozen pieces of mouse or rat or human gastrointestinal tissues was achieved essentially as previously described [52] except for sodium deoxycholate. Protein concentration was determined with the Micro BCA Protein Assay Reagent kit (Biorad, Marnes-La-Coquette, France). First, 25 µg of total protein extract and 7 µg of controls LS 174T protein were boiled for 5 min for reducing Laemmli sample buffer and then resolved by SDS-PAGE on 4–12% minigels (Life Technologies, Fisher Scientific, Illkirch Graffenstaden, France). Proteins were transferred onto a nitrocellulose membrane (200 mA, 2 h).

For B4GALNT2 detection, a blocking step was performed using PBS/5% non-fat dried milk/0.05% Tween 20 overnight at 4 °C for subsequent incubation with 1:1000 of the anti-B4GALNT2 polyclonal antibody performed for 16 h at 4 °C in the same buffer. After 3 PBS/0.05% Tween 20 (PBS-T) washes, membranes were incubated with 1:10,000 horseradish peroxidase conjugated goat anti-rabbit antibody (Sigma Aldrich) in PBS-T and 5% (*w/v*) non-fat dried milk for 1 h at room temperature (RT). Membranes were washed three times for 10 min in PBS-T, and detection was achieved using enhanced chemiluminescence (SuperSignal West Femto Chemiluminescent Substrate, Amersham Biosciences, Little Chalfont, UK).

For β-actin detection, a blocking step was performed using PBS/5% non-fat dried milk/0.05% Tween 20 for 1 h at RT for subsequent incubation with 1:1000 of the monoclonal anti β-actin antibody (Sigma Aldrich). After three washing steps in PBS-T, membranes were incubated for 1 h at RT with 1:10,000 horseradish peroxidase conjugated goat anti-mouse antibody (Sigma Aldrich) in PBS-T and 5% (*w/v*) non-fat dried milk. Membranes were washed three times for 10 min in PBS-T, and detection was achieved using enhanced chemiluminescence (ECL 2 Western Blotting Substrate, Amersham Biosciences, Little Chalfont, UK).

#### 4.7. PNGase F and Endo H Treatments

To examine the *N*-glycosylation modification, 50 or 10 µg of total protein extracts from frozen pieces of tissues or from transiently transfected cells, respectively, was mixed with 10× Glycoprotein Denaturing Buffer (New England Biolabs, Ipswich, MA, USA) and heated 10 min at 100 °C. Ten percent of NP 40 and 10× GlycoBuffer 2 (New England Biolabs, Ipswich, MA, USA) was added then completed with H<sub>2</sub>O and 500,000 U/mL of *N*-Glycosidase F (PNGase F) (New England Biolabs, Ipswich, MA, USA) or diluted PNGase F to make a 20 µL total reaction volume. Samples were incubated 1 h at 37 °C. Then, PNGase-F-treated and non-treated samples were dissolved in reducing NuPage LDS Sample Buffer and resolved by SDS-PAGE on 4–12% Bis-Tris gels (Fisher scientific, Illkirch, France). Detection of B4GALNT2 and β-actin was performed as previously described.

To examine *N*-glycan maturation, 10 µg of total protein lysate from transiently transfected HeLa cells were mixed with 10× Glycoprotein Denaturing Buffer (New England Biolabs, Ipswich, MA, USA) and heated 10 min at 100 °C. 10× GlycoBuffer 3 (New England Biolabs, Ipswich, MA, USA) was added and then completed with H<sub>2</sub>O and 500,000 U/mL of Endoglycosidase H (Endo H) (New England Biolabs, Ipswich, MA, USA) to make a 20 µL total reaction volume. Samples were incubated 1 h at 37 °C. Then, Endo-H-treated and non-treated samples were dissolved in reducing NuPage LDS Sample Buffer and resolved by SDS-PAGE on 4–12% Bis-Tris gels (Fisher Scientific, Illkirch, France). Detection of B4GALNT2 and β-actin was performed as previously described.

#### 4.8. B4GALNT2 Enzyme Activity

COS-7 transfected with B4GALNT2 constructs as described above was harvested and homogenized in distilled water. B4GALNT2 activity of the homogenates was assessed as previously described [19] by the difference between incorporation of radioactive GalNAc on fetuin and asialofetuin.

Immunofluorescence localization of mutated B4GALNT2 was performed using confocal microscopy HeLa cells seeded on 12 mm round glass coverslips and grown for 24 h. Cells were then transfected with the different constructs tagged with eGFP or mCherry, as described in Table S1, using Lipofectamine 2000 (Thermo Fisher Scientific Bioscience). The immunostaining experiments were performed 24 h after transfection following a protocol described previously [24]. Briefly, after three washes with PBS, cells were fixed with paraformaldehyde solution at 3.6% in PBS for 20 min at RT, then permeabilized 10 min at RT in PBS containing 0.5% of Triton X-100. The saturation steps and incubation of antibodies were performed in blocking buffer (PBS containing 2% (*w/v*) BSA, 2% (*v/v*) FBS, and 0.2% (*w/v*) gelatin). Nuclei were stained with Dapi for 15 min in PBS before mounting coverslips with mowiol for observation using fluorescent confocal microscopy.

Different primary antibodies against markers of the subcellular compartments were used for the immunofluorescence assays: Anti-Calnexin (Enzo Life Sciences, New York, NY, USA) and anti-TMEM165 (Sigma) were, respectively, diluted at 1:100 and 1:300 in blocking buffer. The secondary antibodies, Alexa fluor<sup>®</sup> 488 anti-rabbit and Alexa fluor<sup>®</sup> 568 anti-rabbit, were diluted at 1:600 in blocking buffer. The subcellular localization of the different fluorescent proteins and the immunostaining were detected through an inverted Zeiss LSM700 (Oberkochen, Germany) confocal microscope with a 40× oil immersion objective. Data were collected using ZEN PRO 2.1 software (Zeiss) and analyzed with FIJI-WIN64 and ICY free software (Version 2.3.0.0). The vesicles were quantified using the ICY software (<http://icy.bioimageanalysis.org>, accessed on 5 November 2021) in the same manner as previously described in [24].

#### 4.9. Flexibility Prediction

A flexibility prediction of the full-length LF-B4GALNT2 (isoform 1 of the Uniprot ID Q8NHY0) was performed with the DynaMine tool [53,54]. This tool gives a  $S^2$  score that can be interpreted for flexibility prediction as flexible for  $S^2 < 0.69$ , rigid for  $S^2 > 0.8$ , and context-dependent in between.

#### 4.10. Modeling of the B4GALNT2 Dimeric Structure

A model of the B4GALNT2 dimer was produced from the long protein isoform sequence between residues 112 to 566 (isoform 1 of the Uniprot ID Q8NHY0) with a locally installed version of the AlphaFold v2.1 software [55–57]. This version of AlphaFold allows building models of monomers as well as multimers. The options used were the following:

```
-is_prokaryote_list=false  
-max_template_date=2021-11-17  
-model_preset=multimer  
-db_preset=full_dbs
```

Five models were generated with iptm+ptm scores between 0.90 and 0.92, which were all predicted as very good, and superimposing these models shows they are very similar. The first model was slightly different from the four others in that it did not exhibit the interchain disulfide bridges. Although the four models exhibited very similar predicted aligned error matrices, the third-ranked model with a very good matrix and a more homogeneous pLDDT graph compared to the second ranked model was chosen. This model was further used for the addition of a complex type *N*-glycan on the N137.

The accessibility (in Å<sup>2</sup>) of the two N137 of the dimeric model was computed with Naccess v2.1.1 [58]. The following complex-type *N*-glycan was added to the N137 residue of both monomers in the dimeric model: Neu5Ac $\alpha$ 2,3Gal $\beta$ 1,4GlcNAc $\beta$ 1,2Man $\alpha$ 1,3(Neu5Ac $\alpha$ 2,3Gal $\beta$ 1,4GlcNAc $\beta$ 1,2Man $\alpha$ 1,6)Man $\beta$ 1,4GlcNAc $\beta$ 1,4GlcNAc $\beta$ 1-

A 3D structure of the *N*-glycan was generated and minimized with the GLYCAM web carbohydrate builder [59]. The *N*-glycan was then manually grafted on the N137 of both monomers. A minimization of the model bearing the *N*-glycans was finally performed with YASARA [60].

The model was rendered with Pymol 2.3.0 [61], which was used for the production of all images of the structures of this manuscript.

#### 4.11. Statistical Methodology

Data were plotted as scatter plots, and statistical analyses were performed using GRAPHPAD PRISM 5.0 software (GraphPad Software Inc., La Jolla, CA, USA) and were compared using a nonparametric Mann–Whitney test. Values were considered significantly different (\*), with  $p < 0.05$  (\*\*):  $p < 0.001$ .

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24044139/s1>.

**Author Contributions:** Conceptualization, A.H.-L., F.D., M.F.L., G.B. and S.G.-D.; methodology and investigation, V.C., D.V., C.S., L.P., G.V., J.d.R., M.D. and G.B.; data curation and validation, A.H.-L., F.D., G.B., and S.G.-D.; writing—original draft preparation, A.H.-L.; writing—review and editing, A.H.-L., S.G.-D., G.B. and F.D.; supervision, A.H.-L., S.G.-D., F.D. and M.F.L.; project administration and funding acquisition, A.H.-L. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was funded in part by the Centre National de la Recherche Scientifique (CNRS), la région Hauts de France (program Arcir dynamique) and the University of Lille. The authors are indebted to la Ligue contre le cancer 2016 and the Agence Nationale de la Recherche (ANR) for their financial support (ANR-2010-BLAN-120401 GalFish).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article or Supplementary Materials.

**Acknowledgments:** The authors are very grateful to the Bioimaging Center of Lille (BICel) for confocal imaging. They acknowledge the excellent technical assistance of master students and Anais Barry. The contribution of the COST Action CA18103-INNOGLY supported by the European Cooperation in Science and Technology (COST) is greatly acknowledged.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Macvie, S.I.; Morton, J.A.; Pickles, M.M. The reactions and inheritance of a new blood group antigen, Sda. *Vox Sang.* **1967**, *13*, 485–492. [\[CrossRef\]](#)
2. Renton, P.H.; Howell, P.; IKIN, E.W.; Giles, C.M.; Goldsmith, K.L.G. Anti-Sda, a New Blood Group Antibody. *Vox Sang.* **1967**, *13*, 493–501. [\[CrossRef\]](#)
3. Groux-Degroote, S.; Vicogne, D.; Cogez, V.; Schulz, C.; Harduin-Lepers, A. B4GALNT2 Controls Sd<sup>a</sup> and SLe<sup>x</sup> Antigen Biosynthesis in Healthy and Cancer Human Colon. *Chembiochem* **2021**, *22*, 3381–3390. [\[CrossRef\]](#)
4. Stenfelt, L.; Hellberg, Å.; Möller, M.; Thornton, N.; Larson, G.; Olsson, M.L. Missense mutations in the C-terminal portion of the B4GALNT2-encoded glycosyltransferase underlying the Sd(a<sup>-</sup>) phenotype. *Biochem. Biophys. Rep.* **2019**, *19*, 100659. [\[CrossRef\]](#)
5. Dall'Olio, F.; Malagolini, N.; Chiricolo, M.; Trinchera, M.; Harduin-Lepers, A. The expanding roles of the Sd<sup>a</sup>/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim. Biophys. Acta* **2014**, *1840*, 443–453. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Groux-Degroote, S.; Wavelet, C.; Krzewinski-Recchi, M.A.; Portier, L.; Mortuaire, M.; Mihalache, A.; Trinchera, M.; Delannoy, P.; Malagolini, N.; Chiricolo, M.; et al. B4GALNT2 gene expression controls the biosynthesis of Sd<sup>a</sup> and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **2014**, *53*, 442–449. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Ito, A.; Levery, S.B.; Saito, S.; Satoh, M.; Hakomori, S. A novel ganglioside isolated from renal cell carcinoma. *J. Biol. Chem.* **2001**, *276*, 16695–16703. [\[CrossRef\]](#)
8. Tsuchida, A.; Senda, M.; Ito, A.; Saito, S.; Kiso, M.; Ando, T.; Harduin-Lepers, A.; Matsuda, A.; Furukawa, K.; Furukawa, K. Roles of GalNAc-disialyl Lactotetraosyl Antigens in Renal Cancer Cells. *Sci. Rep.* **2018**, *8*, 7017. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Kawamura, Y.I.; Kawashima, R.; Fukunaga, R.; Hirai, K.; Toyama-Sorimachi, N.; Tokuhara, M.; Shimizu, T.; Dohi, T. Introduction of Sd<sup>a</sup> carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res.* **2005**, *65*, 6220–6227. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Pucci, M.; Gomes Ferreira, I.; Malagolini, N.; Ferracin, M.; Dall'Olio, F. The Sd<sup>a</sup> Synthase B4GALNT2 Reduces Malignancy and Stemness in Colon Cancer Cell Lines Independently of Sialyl Lewis X Inhibition. *Int. J. Mol. Sci.* **2020**, *21*, 6558. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Martin, P.T.; Xu, R.; Rodino-Klapac, L.R.; Oglesbay, E.; Camboni, M.; Montgomery, C.L.; Shontz, K.; Chicoine, L.G.; Clark, K.R.; Sahenk, Z.; et al. Overexpression of Galgt2 in skeletal muscle prevents injury resulting from eccentric contractions in both mdx and wild-type mice. *Am. J. Physiol. Cell Physiol.* **2009**, *296*, C476–C488. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Xu, R.; DeVries, S.; Camboni, M.; Martin, P.T. Overexpression of Galgt2 reduces dystrophic pathology in the skeletal muscles of alpha sarcoglycan-deficient mice. *Am. J. Pathol.* **2009**, *175*, 235–247. [\[CrossRef\]](#)
13. Mohlke, K.L.; Purkayastha, A.A.; Westrick, R.J.; Smith, P.L.; Petryniak, B.; Lowe, J.B.; Ginsburg, D. Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell* **1999**, *96*, 111–120. [\[CrossRef\]](#)
14. Lefrancois, L.; Puddington, L.; Machamer, C.E.; Bevan, M.J. Acquisition of cytotoxic T lymphocyte-specific carbohydrate differentiation antigens. *J. Exp. Med.* **1985**, *162*, 1275–1293. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Staubach, F.; Kunzel, S.; Baines, A.C.; Yee, A.; McGee, B.M.; Backhed, F.; Baines, J.F.; Johnsen, J.M. Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. *ISME J.* **2012**, *6*, 1345–1355. [\[CrossRef\]](#)
16. Li, P.T.; Liao, C.J.; Yu, L.C.; Wu, W.G.; Chu, S.T. Localization of B4GALNT2 and its role in mouse embryo attachment. *Fertil. Steril.* **2012**, *97*, 1206–1212. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Varki, A.; Cummings, R.D.; Aebi, M.; Packer, N.H.; Seeberger, P.H.; Esko, J.D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; et al. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, *25*, 1323–1324. [\[CrossRef\]](#)
18. Cheng, K.; Zhou, Y.; Neelamegham, S. DrawGlycan-SNFG: A robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* **2017**, *27*, 200–205. [\[CrossRef\]](#)
19. Lo Presti, L.; Cabuy, E.; Chiricolo, M.; Dall'Olio, F. Molecular cloning of the human beta1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd<sup>a</sup> histo-blood group antigen: The sequence predicts a very long cytoplasmic domain. *J. Biochem.* **2003**, *134*, 675–682. [\[CrossRef\]](#)
20. Montiel, M.D.; Krzewinski-Recchi, M.A.; Delannoy, P.; Harduin-Lepers, A. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acα2-3Galβ-Rβ1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sd<sup>a</sup>/Cad antigen: Evidence for an unusual extended cytoplasmic domain. *Biochem. J.* **2003**, *373*, 369–379. [\[CrossRef\]](#)
21. Pucci, M.; Malagolini, N.; Dall'Olio, F. Glycosyltransferase B4GALNT2 as a Predictor of Good Prognosis in Colon Cancer: Lessons from Databases. *Int. J. Mol. Sci.* **2021**, *22*, 4331. [\[CrossRef\]](#)
22. Wavelet-Vermuse, C.; Groux-Degroote, S.; Vicogne, D.; Cogez, V.; Venturi, G.; Trinchera, M.; Brysbaert, G.; Krzewinski-Recchi, M.A.; Hadj Bachir, E.; Schulz, C.; et al. Analysis of the proximal promoter of the human colon-specific B4GALNT2 (Sd<sup>a</sup> synthase) gene: B4GALNT2 is transcriptionally regulated by ETS1. *Biochim. Biophys. Acta Gene Regul. Mech.* **2021**, *1864*, 194747. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Malagolini, N.; Santini, D.; Chiricolo, M.; Dall'Olio, F. Biosynthesis and expression of the Sd<sup>a</sup> and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology* **2007**, *17*, 688–697. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Groux-Degroote, S.; Schulz, C.; Cogez, V.; Noel, M.; Portier, L.; Vicogne, D.; Solorzano, C.; Dall'Olio, F.; Steenackers, A.; Mortuaire, M.; et al. The extended cytoplasmic tail of the human B4GALNT2 is critical for its Golgi targeting and post-Golgi sorting. *FEBS J.* **2018**, *285*, 3442–3463. [[CrossRef](#)]
25. Rosnoblet, C.; Peanne, R.; Legrand, D.; Foulquier, F. Glycosylation disorders of membrane trafficking. *Glycoconj. J.* **2013**, *30*, 23–31. [[CrossRef](#)] [[PubMed](#)]
26. Mikolajczyk, K.; Kaczmarek, R.; Czerwinski, M. How glycosylation affects glycosylation: The role of *N*-glycans in glycosyltransferase activity. *Glycobiology* **2020**, *30*, 941–969. [[CrossRef](#)]
27. Smith, P.L.; Lowe, J.B. Molecular cloning of a murine *N*-acetylgalactosamine transferase cDNA that determines expression of the T lymphocyte-specific CT oligosaccharide differentiation antigen. *J. Biol. Chem.* **1994**, *269*, 15162–15171. [[CrossRef](#)]
28. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. In Proceedings of the Pacific Symposium on Biocomputing 2002, Kauai, HI, USA, 3–7 January 2002; pp. 310–322.
29. Haraguchi, M.; Yamashiro, S.; Furukawa, K.; Takamiya, K.; Shiku, H.; Furukawa, K. The effects of the site-directed removal of *N*-glycosylation sites from  $\beta$ -1,4-*N*-acetylgalactosaminyltransferase on its function. *Biochem. J.* **1995**, *312*, 273–280. [[CrossRef](#)]
30. Faid, V.; Denguir, N.; Chapuis, V.; Bihoreau, N.; Chevreux, G. Site-specific *N*-glycosylation analysis of human factor XI: Identification of a noncanonical NXC glycosite. *Proteomics* **2014**, *14*, 2460–2470. [[CrossRef](#)]
31. Lowenthal, M.S.; Davis, K.S.; Formolo, T.; Kilpatrick, L.E.; Phinney, K.W. Identification of Novel *N*-Glycosylation Sites at Noncanonical Protein Consensus Motifs. *J. Proteome Res.* **2016**, *15*, 2087–2101. [[CrossRef](#)]
32. Sun, S.; Zhang, H. Identification and validation of atypical *N*-glycosylation sites. *Anal. Chem.* **2015**, *87*, 11948–11951. [[CrossRef](#)]
33. Zielinska, D.F.; Gnad, F.; Wisniewski, J.R.; Mann, M. Precision mapping of an in vivo *N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010**, *141*, 897–907. [[CrossRef](#)]
34. Foulquier, F.; Amyere, M.; Jaeken, J.; Zeevaert, R.; Schollen, E.; Race, V.; Bammens, R.; Morelle, W.; Rosnoblet, C.; Legrand, D.; et al. TMEM165 deficiency causes a congenital disorder of glycosylation. *Am. J. Hum. Genet.* **2012**, *91*, 15–26. [[CrossRef](#)]
35. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic. Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]
36. Park, C.; Zhang, J. Genome-wide evolutionary conservation of *N*-glycosylation sites. *Mol. Biol. Evol.* **2011**, *28*, 2351–2357. [[CrossRef](#)] [[PubMed](#)]
37. Chen, C.; Colley, K.J. Minimal structural and glycosylation requirements for ST6Gal I activity and trafficking. *Glycobiology* **2000**, *10*, 531–583. [[CrossRef](#)] [[PubMed](#)]
38. Jeanneau, C.; Chazalet, V.; Auge, C.; Soumpasis, D.M.; Harduin-Lepers, A.; Delannoy, P.; Imbert, A.; Breton, C. Structure-function analysis of the human sialyltransferase ST3Gal I: Role of *N*-glycosylation and a novel conserved sialylmotif. *J. Biol. Chem.* **2004**, *279*, 13461–13468. [[CrossRef](#)] [[PubMed](#)]
39. Vallejo-Ruiz, V.; Haque, R.; Mir, A.M.; Schwientek, T.; Mandel, U.; Cacan, R.; Delannoy, P.; Harduin-Lepers, A. Delineation of the minimal catalytic domain of human Gal $\beta$ 1-3GalNAc  $\alpha$ 2,3-sialyltransferase (hST3Gal I). *Biochim. Biophys. Acta* **2001**, *1549*, 161–173. [[CrossRef](#)] [[PubMed](#)]
40. Ruggiero, F.M.; Vilcaes, A.A.; Iglesias-Bartolome, R.; Daniotti, J.L. Critical role of evolutionarily conserved glycosylation at Asn211 in the intracellular trafficking and activity of sialyltransferase ST3Gal-II. *Biochem. J.* **2015**, *469*, 83–95. [[CrossRef](#)] [[PubMed](#)]
41. Bieberich, E.; Tencmnao, T.; Kapitonov, D.; Yu, R.K. Effect of *N*-glycosylation on turnover and subcellular distribution of *N*-acetylgalactosaminyltransferase I and sialyltransferase II in neuroblastoma cells. *J. Neurochem.* **2000**, *74*, 2359–2364. [[CrossRef](#)] [[PubMed](#)]
42. Martina, J.A.; Daniotti, J.L.; Maccioni, H.J. Influence of *N*-glycosylation and *N*-glycan trimming on the activity and intracellular traffic of GD3 synthase. *J. Biol. Chem.* **1998**, *273*, 3725–3731. [[CrossRef](#)]
43. Uemura, S.; Kurose, T.; Suzuki, T.; Yoshida, S.; Ito, M.; Saito, M.; Horiuchi, M.; Inagaki, F.; Igarashi, Y.; Inokuchi, J. Substitution of the *N*-glycan function in glycosyltransferases by specific amino acids: ST3Gal-V as a model enzyme. *Glycobiology* **2006**, *16*, 258–270. [[CrossRef](#)] [[PubMed](#)]
44. El-Battari, A.; Prorok, M.; Angata, K.; Mathieu, S.; Zerfaoui, M.; Ong, E.; Suzuki, M.; Lombardo, D.; Fukuda, M. Different glycosyltransferases are differentially processed for secretion, dimerization, and autoglycosylation. *Glycobiology* **2003**, *13*, 941–953. [[CrossRef](#)]
45. Hassinen, A.; Kellokumpu, S. Organizational interplay of Golgi *N*-glycosyltransferases involves organelle microenvironment-dependent transitions between enzyme homo- and heteromers. *J. Biol. Chem.* **2014**, *289*, 26937–26948. [[CrossRef](#)] [[PubMed](#)]
46. Hassinen, A.; Rivinoja, A.; Kaupilla, A.; Kellokumpu, S. Golgi *N*-glycosyltransferases form both homo- and heterodimeric enzyme complexes in live cells. *J. Biol. Chem.* **2010**, *285*, 17771–17777. [[CrossRef](#)] [[PubMed](#)]
47. Khoder-Agha, F.; Harrus, D.; Brysbaert, G.; Lensink, M.F.; Harduin-Lepers, A.; Glumoff, T.; Kellokumpu, S. Assembly of B4GALT1/ST6GAL1 heteromers in the Golgi membranes involves lateral interactions via highly charged surface domains. *J. Biol. Chem.* **2019**, *294*, 14383–14393. [[CrossRef](#)]
48. Li, J.; Yen, T.Y.; Allende, M.L.; Joshi, R.K.; Cai, J.; Pierce, W.M.; Jaskiewicz, E.; Darling, D.S.; Macher, B.A.; Young, W.W., Jr. Disulfide bonds of GM2 synthase homodimers. Antiparallel orientation of the catalytic domains. *J. Biol. Chem.* **2000**, *275*, 41476–41486. [[CrossRef](#)]

49. Cherepanova, N.A.; Shrimal, S.; Gilmore, R. Oxidoreductase activity is necessary for *N*-glycosylation of cysteine-proximal acceptor sites in glycoproteins. *J. Cell Biol.* **2014**, *206*, 525–539. [[CrossRef](#)]
50. Cherepanova, N.A.; Venev, S.V.; Leszyk, J.D.; Shaffer, S.A.; Gilmore, R. Quantitative glycoproteomics reveals new classes of STT3A- and STT3B-dependent *N*-glycosylation sites. *J. Cell Biol.* **2019**, *218*, 2782–2796. [[CrossRef](#)]
51. Blom, N.; Sicheritz-Ponten, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633–1649. [[CrossRef](#)]
52. Colomb, F.; Krzewinski-Recchi, M.A.; El Machhour, F.; Mensier, E.; Jaillard, S.; Steenackers, A.; Harduin-Lepers, A.; Lafitte, J.J.; Delannoy, P.; Groux-Degroote, S. TNF regulates sialyl-Lewisx and 6-sulfo-sialyl-Lewisx expression in human lung through up-regulation of ST3GAL4 transcript isoform BX. *Biochimie* **2012**, *94*, 2045–2053. [[CrossRef](#)] [[PubMed](#)]
53. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W.F. From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* **2013**, *4*, 2741. [[CrossRef](#)] [[PubMed](#)]
54. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W.F. The DynaMine webserver: Predicting protein dynamics from sequence. *Nucleic. Acids Res.* **2014**, *42*, W264–W270. [[CrossRef](#)] [[PubMed](#)]
55. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **2021**. [[CrossRef](#)]
56. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
57. Schrödinger, LLC. *The PyMOL Molecular Graphics System*, version 2.3.0; 2010. Available online: <https://pymol.org/2/support.html?#citing> (accessed on 23 December 2021).
58. Hubbard, S.J.; Thornton, J.M. *Naccess, Computer Program*, version 2.1.1; Department of Biochemistry and Molecular Biology, University College: London, UK, 1993.
59. Woods Group, R. GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia: Athens, GA, USA. Available online: <http://legacy.glycam.org> (accessed on 1 January 2020).
60. Krieger, E.; Vriend, G. YASARA View—Molecular graphics for all devices—From smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981–2982. [[CrossRef](#)]
61. Schrödinger, LLC. *The PyMOL Molecular Graphics System*, version 1.8; 2015. Available online: <https://pymol.org/2/> (accessed on 23 December 2021).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.