Title page

Delving into human α1,4-galactosyltransferase acceptor specificity: the role of enzyme dimerization

Krzysztof Mikołajczyk[1,2], Karol Wróblewski[3], Sebastian Kmiecik[3]

[2] Laboratory of Glycobiology, Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Rudolfa Weigla St. 12, 53-114 Wroclaw, Poland.

[3] Biological and Chemical Research Center, Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland.

[1] To whom correspondence should be addressed: Tel: (+48)-71-3371172; e-mail: krzysztof.mikolajczyk@hirszfeld.pl

Keywords: glycosyltransferase, dimerization, specificity, A4GALT, B4GALT

Abstract

Glycosyltransferases (GTs) exhibit precise donor and acceptor specificities, governed by intricate mechanisms, including protein assembly. Human α1,4-galactosyltransferase (A4galt), a Golgi apparatus-resident GT, catalyzes the synthesis of Gb3 glycosphingolipid (GSL) and P1 glycotope on glycoproteins (GPs), receptors for Shiga toxin type 1 (Stx1) and 2 (Stx2). These toxins are produced by enterohemorrhagic *Escherichia coli* and can lead to severe outcomes, such as hemolytic-uremic syndrome (HUS). To elucidate the molecular basis of A4galt specificity towards GSLs and GPs, we explored its interaction with GP-specific β1,4-galactosyltransferase 1 (B4galt1) and GSL-specific β1,4-galactosyltransferase isoenzymes 5 and 6 (B4galt5 and B4galt6). Using a novel NanoBiT assay, we demonstrated that A4galt could form homodimers and heterodimers with B4galt1 and B4galt5. Furthermore, by employing AlphaFold for state-of-the-art structural prediction, we analyzed the interactions and structures of these enzyme complexes. Our analysis highlighted that the A4galt-B4galt5 heterodimer exhibited the highest prediction confidence, indicating a significant role of A4galt heterodimers in determining enzyme specificity toward GSLs and GPs. These insights not only increase our knowledge about glycosylation processes but also open avenues for developing new therapeutic approaches to Shiga toxin-related diseases.

Introduction

Human α1,4-galactosyltransferase (A4galt, Gb3/CD77 synthase, P1/Pk synthase, EC 2.4.1.228) is a type II transmembrane glycosyltransferase (GT) located in the *trans*-Golgi apparatus. This enzyme consists of a C-terminal globular catalytic domain oriented towards the Golgi lumen and an N-terminal cytoplasmic domain [Mikolajczyk et al. 2021; Kaczmarek R, et al. 2014]. Topologically, human A4galt comprises a cytoplasmic domain (spanning 1-22 amino acid residues), a transmembrane domain (23-43 aa residues), which anchors the enzyme within the Golgi membrane, and a luminal domain (44-353 aa residues), containing the catalytic site (according to UniProt Q9NPC4, https://www.uniprot.org/).

In silico analysis of the human A4galt AlphaFold model revealed that the enzyme belongs to the GT-A fold-type GT and is a member of 32 families in the CAZy glycosyltransferase database (Carbohydrate Active Enzymes Database, CAZy, http://www.cazy.org/). For catalytic activity, it requires a divalent metal ion (typically Mn^{2+}) coordinated by two aspartic acid residues to form a DXD motif (D_{192}TD in human A4galt) [Mikolajczyk K, et al. 2022; Mikolajczyk K, et al. 2021]. Human A4galt occurs as a high-frequency enzyme, containing a glutamine (Q) at position 211 (p.Q211), and a rare enzyme variant (referred to as mutein) with a p.Q211E substitution (rs397514502), found in only two families worldwide [Suchanowska A, et al. 2012]. This amino acid substitution affects the enzyme acceptor specificity towards GalNAc-capped oligosaccharides, enabling the synthesis of NOR1 and NOR2 antigens (terminally with a Galα1→4GalNAc disaccharide) [Kaczmarek R, et al. 2016; Suchanowska A, et al. 2012; Kaczmarek R, et al. 2014].

For many years, human A4galt has been considered as a glycosphingolipid (GSL)-specific enzyme, responsible for transferring Gal residues into GSL acceptors, such as lactosylceramide, resulting in the formation of globotriaosylceramide (Gb3, Galα1→4Galβ1→4Glc-Cer). However, recent studies have shown that A4galt reveals a broader enzyme specificity toward glycoprotein (GP)-related acceptors, particularly N-glycans. It may produce a unique oligosaccharide motif, named P1 glycotope (consists of a Galα1→4Galβ1→4GlcNAc carbohydrate sequence), which contains a Galα1→4Galβ disaccharide, identical to the terminal disaccharide of Gb3. Both products generated by A4galt, Gb3 on GSLs and the P1 glycotope on glycoproteins, may serve as receptors for Shiga toxins (Stxs) [Szymczak-Kulus K, et al. 2021; Bereznicka A, et al. 2021; Morimoto K, et al. 2020]. Shiga toxin-producing *Escherichia coli* (STEC) strains, similar to *Shigella dysenteriae* serotype 1, are posing an increasing threat to the human population. STEC

infections lead to hemorrhagic colitis, which often progresses to hemolytic-uremic syndrome (HUS), a severe complication that usually causes acute kidney failure [Cody EM and Dixon BP, 2019; Johannes L and Römer W, 2010; Bruyand M, et al. 2018; Michael M, et al. 2022]. Stxs are produced by STEC are of two types: Stx1, which is identical to the toxin secreted by *S. dysenteriae* serotype 1, and the more distinct Stx2. Structurally, Stxs belong to the AB_5 toxin family and comprise an A catalytic subunit responsible for cytotoxic effects and a pentameric B subunit that specifically binds to cell receptors (Gb3 and P1 glycotope) [Cooling L, 2015; Johannes L and Römer W, 2010; Kim JS, et al. 2020; Szymczak-Kulus K, et al. 2021]. Stx1 exhibits a remarkable ability to utilize not only the GSL receptor but also the P1 glycotope on N-glycoproteins for cell entry and cytotoxicity, whereas Stx2 exclusively relies on GSL-related receptor [Szymczak-Kulus K, et al. 2021; Morimoto K, et al. 2020].

Herein, we proposed that human A4galt forms complexes within the Golgi apparatus, thereby orchestrating sequential glycan biosynthesis pathways utilizing GSL- and GP-based acceptors [Kellokumpu S, et al. 2016]. Preliminary data have shown that human A4galt forms homodimers [Kaczmarek, et al. 2012] and heterodimers with β1,4-galactosyltransferase 6 (B4galt6) [Takematsu H, et al. 2011]. In this study, we used the STRING v12 database to identify putative protein partners interacting with human A4galt. Then, a relatively new NanoBiT (NanoLuc® Binary Technology) Protein: Protein Interaction System was employed to assess potential dimers using CHO-Lec2 and HEK293T cells. To complement this data, we investigated the molecular underpinnings of A4galt homo- and heterodimerization using the AlphaFold enzyme model.

Results

Identification of putative protein partners for interaction with human A4galt

The specificity towards GP- and GSL-based acceptors of human A4galt may be mediated by its interaction with other glycosyltransferases, which vary in acceptor specificities. To identify potential protein partners for human A4galt interactions, we employed the STRING v12 database [Szklarczyk D, et al. 2023; Szklarczyk D, et al. 2019], considering *Criticetulus griseus* (Chinese hamster) and *Homo sapiens* as host organisms. The selection of Chinese hamsters was based on our earlier studies, in which we determined Stxs-receptor interactions using CHO-Lec2 cells, a cell line derived from Chinese hamsters [Szymczak-Kulus K, et al. 2021]. To translate these findings into human cells, we identified putative protein partners to interact with A4galt also in *H. sapiens*. We obtained 9 protein candidates in humans and 10 in

Chinese hamsters, classified into glycosyltransferases and glycosidases. For further studies, we selected the β1,4-galactosyltransferases (B4galt1-4 and B4galt5-6), because of their localization in the Golgi apparatus (similar to human A4galt) and their ability to synthesize potential Gal-terminated substrates for human A4galt.

B4galt1 is a glycosyltransferase located in the *trans*-Golgi that synthesizes β-galactosylated N-glycans, while B4galt2-4 also exhibits glycoprotein-utilizing activity, albeit to a lesser extent [Lee et al. 2001; Bydlinski et al. 2018]. B4galt5 and B4galt6 catalyze the synthesis of lactosylceramide (LacCer), a major GSL-based substrate for human A4galt. In addition, both enzymes are located in the Golgi cisternae and *trans*-Golgi network (TGN) [Rizzo et al. 2021]. The co-occurrence of *trans*-Golgi and involvement in the synthesis of Gal-terminated N-glycans and lactosylceramide (LacCer) suggests the existence of heterodimers containing A4galt and B4galt1-4 and/or B4galt5-6. Their spatial proximity in glycosylation pathways increases the likelihood of interaction, potentially contributing to the formation of supramolecular complexes involving these GTs. Such heterocomplexes may elucidate the dual specificity of human A4galt towards Gal-terminated GSL- and GP-based acceptors.

Analysis of human A4galt dimerization using NanoBiT technology

NanoBiT technology consists of expression constructs containing genes encoding the interaction protein partner fused with one of the two subunits of NanoLuc luciferase: a small SmBiT (11 amino acids), and a large LgBiT (18 kDa). Upon dimer formation between the analyzed proteins, reconstitution of the active luciferase enzyme prompts luminescence. Although it is a relatively new method, NanoBiT has already been used to identify interaction partners for glycosyltransferases, such as human B4galt4 [Shauchuk et al. 2020] and B4galt1 [Wiertelak et al. 2020].

In this study, NanoBiT technology was used to assess the formation of homo- and heterodimers by high-frequency and mutein human A4galt. Based on STRING v12 results (Figure 1), B4galt1, B4galt5, and B4galt6 were selected as potential partners for heterodimerization with A4galt. Additionally, we used two cell lines, CHO-Lec2 and HEK293T. CHO-Lec2 cells transfected with the human *A4GALT* gene were used to examine the ability of human A4galt to recognize both GSL- and GP-based acceptors [Szymczak-Kulus K, et al. 2021]. Since CHO-Lec2 cells do not contain the homolog of the *A4GALT* gene, which encodes A4galt, these cells are not primarily sensitive to Stxs. So, we decided to

validate the proposed experiments in HEK293T cells, which express human A4galt and can naturally bind Stxs.

Using NanoBiT, we tested various combinations of proteins fused with the LgBiT and SmBiT subunits (located on either the N- or C-terminus of the analyzed protein) to determine the optimal orientation for interactions (Suppl. Table 3). We found that both forms of human A4galt (high-frequency and mutein) could form homodimers in CHO-Lec2 and HEK293T cells (Figures 2 and 3). Notably, the highest average luminescence value (compared to the respective controls) was observed for the mutein in CHO-Lec2 cells (Figure 3), in contrast to HEK293T cells (Figure 2), in which the average luminescence of A4galt homodimers was similar for both the high-frequency and mutein forms.

Additionally, we observed the formation of A4galt-B4galt1 and A4galt-B4galt5 heterodimers both in CHO-Lec2 and HEK293T cells, with no detectable luminescence for the A4galt-B4galt6 pairs (Figure 2 and 3). The average luminescence measured for A4galt-B4galt1 pairs of high-frequency and mutein enzymes in CHO-Lec2 cells was 90 and 360 times higher than that for controls, respectively (Figure 3). Similarly, A4galt-B4galt1 heterodimers analyzed in HEK293T cells exhibited robust luminescence signals, approximately 60 times higher for the high-frequency enzyme and approximately 30 times higher for the mutein, compared to the control samples (Figure 2). The greatest luminescence signals were detected for A4galt-B4galt5 pairs in CHO-Lec2 cells, with approximately 1000 times greater luminescence for the mutein compared to 300 times higher signal for the high-frequency enzyme, in comparison to controls (Figure 3). These findings confirm the ability of human A4galt to form homodimers, as well as heterodimers with B4galt1 and B4galt5, highlighting the varied GSL- and GP-based acceptor specificity.

Molecular basis of human A4galt dimerization

Our *in vitro* investigations highlighted that human A4galt can form homodimers as well as heterodimers with B4galt1 and B4galt5; however, the mechanism of these PPIs remains elusive. Given the unresolved spatial structure of human A4galt, we employed the AlphaFold-Multimer tool [Evans R, et al. 2021] to predict the structures of these complexes. Models for all monomers, including A4galt, B4galt1, and B4galt5, were generated using AlphaFold [Jumper J, et al. 2021] and are available in the AlphaFold database [Varadi M, et al. 2022]. The monomer models obtained from the AlphaFold database exhibited high pLDDT scores, indicating a high level of structural reliability, with the exception of the flexible N-terminal

fragment. For our modeling endeavors, we considered predictions based on both the full enzyme sequences and those with the N-terminus removed. However, predictions of PPIs using full sequences did not yield reliable outcomes because the N-terminal protein fragment interfered with the formation of intermolecular interfaces in the dimers. Consequently, we focused solely on models with the N-terminus omitted, targeting regions with predictably stable structures for our analysis.

Thus, we attempted to predict the structures of the protein complexes, with the A4galt-B4galt5 heterodimer model emerging as the most accurate prediction (Fig. 4A). This model exhibited the highest pLDDT score, with an average of 92.24. High pLDDT values, particularly those exceeding 90, indicated precise predictions. Furthermore, the high quality of the model was supported by the Predicted Aligned Error (PAE) scores, where the average PAE for residue pairs within a 6 Å contact distance was 3.76 Å (see Fig. S1). Low PAE values, notably below 5 Å for specific residue pairs, are indicative of a reliable prediction of their spatial relationship, emphasizing the importance of PAE as a key metric for evaluating the structural integrity and precision of protein complexes.

Further analyses, facilitated by EDHB [Verma N, et al. 2021], suggested the potential formation of hydrogen bonds between A4galt and B4galt5 polypeptide chains, particularly between W244 (A4galt) and A369 (B4galt5), as well as W246 (A4galt) and Y367 (B4galt5) (Fig 5.). Further insights into additional interactions were provided by the Mapiya web-server [Badaczewska-Dawid AE, et al. 2022], which are elaborated in the Supporting Materials (Fig. S2-S3).

In addition to the A4galt-B4galt5 model, we also generated models for the homodimer of A4galt-A4galt and a heterodimer of A4galt-B4galt1. However, these complexes exhibited lower pLDDT scores compared to the A4galt-B4galt5 model, particularly within regions where the two monomers interact. Moreover, the high PAE scores for contacting residue pairs in these models suggested diminished confidence in the predicted dimer interface and overall structural configuration (Fig 4B and C).

To explore these complexes, we performed coevolution sequence analysis using EVcouplings [Hopf TA, et al. 2019] on the A4galt-B4galt5 heterodimer to identify correlated mutations. The analysis was challenging because of the low number of effective sequences, resulting in a ratio of effective sequences to protein length of 0.47 (the value below 1 suggests the limited reliability of this analysis). We identified only a few residue pairs outside the anticipated interface area, showing only a slight increase in the probability of correlated

mutations. A similar coevolution analysis for the A4galt-B4galt1 heterodimer revealed an even smaller number of sequences in the multiple sequence alignment (MSA), highlighting the difficulties in obtaining sufficient confidence in structural predictions for this complex via AlphaFold (Fig. 4B). AlphaFold typically relies on a robust MSA for precision, and the limited number of sequences in MSA poses challenges for accurate predictions.

In addition, we explored the potential active sites of the monomers using the COACH meta-server. We included B4galt1, for which the active site is known [Harrus D, et al. 2018], to test the reliability of the results. This method proved to be reliable, as it provided consistent predictions across various methods for binding pockets and known substrates associated with these proteins. Details regarding the potential interaction sites are provided in the Supporting Material (Fig. S4-S6).

Analysis of the mutual arrangement of these active sites, supported by the known structure of the B4galt1 homodimer (PDB ID: 6FWU), suggested that the active sites in the heterodimer A4galt-B4galt1 are proximal enough to facilitate substrate exchange. The AlphaFold-predicted structure showed that most of the interface between the two monomers was composed of residues that formed their respective active centers (Fig. 6A). A similar observation was made for the A4galt-B4galt5 heterodimer, however only the active site residues of A4galt contributed to the interface (Fig. 6B). The proximity between the catalytic centers of A4galt and B4galt1 indicated the potential for substrate transfer.

Discussion

Herein, we aimed to elucidate the ability of human A4galt, including its high-frequency and mutein forms, to engage in homo- and heterodimerization. Our findings indicate the propensity of human A4galt for homodimerization, but the functional implications of this phenomenon for enzyme functions require further investigation. Homodimerization is a well-documented phenomenon among various GTs with distinct acceptor specificities, such as glycoprotein-specific α1,6-fucosyltransferase 8 (Fut8) [Ihara H, et al. 2007] and glycolipid-specific β1,4-*N*-acetylgalactosaminyltransferase 1 (B4galnt1) [Giraudo CG, et al. 2001]. In addition to homodimerization, we showed that both high-frequency and mutein forms of human A4galt can form heterodimers with B4galt1 and B4galt5, exhibiting distinct acceptor specificity towards GSLs and GPs. Notably, previous research conducted by Takematsu et al. (2011) suggested a potential interaction between human A4galt and B4galt6, which was supported by the shared presence of the GOLPH3 recognition polypeptide motif in these

enzymes. It underpins the possibility of heterodimerization within the *trans*-Golgi apparatus [Takematsu et al., 2011; Rizzo R, et al. 2021; Lujan P and Campelo F, 2021].

Furthermore, various examples of GT heterodimers with varied specificities towards GSLs and GPs have been described. For instance, heterodimers such as glycoprotein-specific β1,4-galactosyltransferase 1/α2,6-sialyltransferase 1 [Hassinen A, et al. 2011] and β1,3-*N*-acetylglucosaminyltransferase 2/β1,3-*N*-acetylglucosaminyltransferase 8 [Seko A, et al. 2008], are crucial in N-glycoprotein synthesis. Additionally, the glycolipid-specific heterodimers formed by β1,4-*N*-acetylgalactosaminyltransferase 1 with β1,3-galactosyltransferase 4 [Maccioni HJ, 2007] and α2,8-sialyltransferase 8A with β1,4-*N*-acetylgalactosaminyltransferase 1 [Giraudo CG and Maccioni HJ, 2003] play pivotal roles in ganglio-series GSLs synthesis.

By utilizing AlphaFold models for A4galt, B4galt1, and B4galt5, we attempted to elucidate the molecular intricacies underlying their interactions. Analysis of the A4galt-B4galt1 heterodimer revealed spatial proximity of the active sites of both enzymes, potentially facilitating mutual substrate exchange. This observation resembles the mechanism demonstrated in human α2,8-sialyltransferase 3, which forms a homodimer close to the Golgi apparatus membrane [Volkers G, et al. 2015]. For the A4galt-B4galt5 heterodimer, we proposed an electrostatic bond-mediated enzyme interaction, similar to the mechanism previously described for homo- and heterodimer formation by β1,4-galactosyltransferase 1 and α2,6-sialyltransferase 1 [Khoder-Agha F, et al. 2019]. Both A4galt-B4galt1 and A4galt-B4galt5 heterocomplexes encourage the transfer of products from B4galt1 and B4galt5 (Gal-terminated N-glycoproteins and glycosphingolipids, respectively) to A4galt. Subsequently, A4galt synthesizes Gal-capped oligosaccharides on two different types of molecules (GSLs and GPs).

Understanding the impact of GT dimerization on acceptor specificity is essential to unravel the molecular basis of the glycosylation processes in cells. Future studies should focus on the structural determinants of acceptor substrate recognition within heterodimeric complexes and investigate how alterations in acceptor specificity may translate into functional consequences in cell physiology and pathology. Further exploration of the mechanistic details underlying the changes in acceptor specificity related to heterodimer formation is crucial for comprehensively deciphering the regulatory role of these GT complexes in shaping the cell glycome.

Material and methods

Evaluation of potential protein-protein interaction networks

The Search Tool for the Retrieval of Interacting Genes/Proteins database (STRING v12) was used to evaluate the protein-protein interactions (PPI) of human α1,4-galactosyltransferase (A4galt) with other proteins. The STRING v12 database generates a PPI network using several data sources and integrates the physical and functional basis of protein-protein interactions [Szklarczyk D, et al. 2023]. For STRING v12 analysis, we chose human A4galt as an input enzyme, and *H. sapiens* and *C. griseus* (corresponding to the cell lines HEK293T and CHO-Lec2, respectively) as target organisms. In further studies, only a putative interaction with a high level of confidence (score ≥ 0.90) was considered.

Cell cultures and stable transfection

CHO-Lec2 and HEK293T cells were obtained from the American Type Culture Collection (Rockville, MD, USA). The cells were grown and maintained in a humidified incubator with 5% CO2 at 37°C in DMEM/F12 medium (Thermo Fisher Scientific, Inc., Waltham, MA, USA) supplemented with 10% fetal bovine serum (Gibco, Inc., Waltham, MA, USA) and Pen-Strep (Gibco, Inc., Waltham, MA, USA). The culture medium was changed every second or third day, and after reaching 85–90% confluence, the cells were subcultured by treatment with trypsin (0.25% trypsin, 137 mM NaCl, 4.3 mM NaHCO3, 5.4 mM KCl, 5.6 mM glucose, 0.014 mM phenol red, 0.7 mM EDTA), harvested, centrifuged at 800 × g for 5 min, resuspended in fresh medium and seeded to new tissue culture plates.

Construction of expression vectors

The ORF sequences of human B4GALT1 (GenBank accession number NM_001378495.1), B4GALT5 (GenBank accession number NM_004776.4), and B4GALT6 (GenBank accession number NM_004775.5) were customized by Thermo Fisher Scientific and cloned into pET-TOPO plasmids. The template for human A4GALT (GenBank accession number NG_007495.2) amplification was the pCAG-A4GALT expression plasmid, containing the A4GALT gene encoding full-length human A4galt (high-frequency and mutein with p.Q211E substitution). The expression constructs for the split luciferase complementation assay were prepared according to the manufacturer's protocol (NanoLuc® Binary Technology NanoBiT™, Promega, Madison, WI, USA). Briefly, the genes encoding the protein candidates for interaction with human A4galt were cloned into pBiT1.1 and pBiT2.1 vectors

to produce fusion constructs with LgBiT and SmBiT subunits, respectively. The luciferase LgBiT and SmBiT subunits were localized at the N- or C-terminus of the analyzed protein to establish the optimal orientation for PPIs. PCR amplification, DNA sequencing, and preparation of plasmid constructs were carried out as previously described [Mikolajczyk K, et al. 2021]. Cloning of pBiT vectors was performed using XhoI and EcoRI restriction sites added to the forward and reverse primers. The sequences of all primers used in this study are listed in Suppl. Table 1.

Split luciferase complementation assay

CHO-Lec2 and HEK293T cells (2 × 104) were seeded in a complete growth medium onto a 96-well plate with white polystyrene wells and a flat transparent bottom (Corning Inc., New York, USA). 20–24 h after plating, the cells were transfected with appropriate combinations of plasmids (25 ng/well of each plasmid, 50 ng/well in total) using linear PEI 25 kDa (1 mg/mL, Polysciences Inc.) transfection reagent. 2–3 h before the measurement, the conditioned medium was replaced with the serum-free Opti-MEM medium (Life Technologies, CA, USA). Immediately before measurement, Nano-Glo® Live Cell Substrate (Promega) was added to all wells, according to the manufacturer's instructions. Cell-derived luminescence was recorded using a Clariostar Luminescence Microplate Reader (BMG LABTECH). Each tested combination was accompanied by the respective negative control comprising the analyzed GT fused with the large NanoLuc subunit combined with the HaloTag (a recombinant protein not synthesized by mammalian cells, thus not interacting with proteins of mammalian origin) that was fused with the small NanoLuc subunit (Promega).

Data analysis

For statistical analysis, one-way ANOVA with Bonferroni post-hoc test was used. All analyses were performed with GraphPad Prism (GraphPad Software, CA, USA). Statistical significance was assigned to p-value < 0.05. Only results exceeding the respective negative controls by at least 10-fold were considered indicative of an interaction, as suggested by the manufacturer.

Dimer structure prediction

Structure prediction was carried out using AlphaFold-Mulitmer [Jumper J, et al. 2021; Evans R, et al. 2021] implemented in ColabFold [Mirdita M, et al. 2022]. Multiple parameters were tested, that is, template modes (and using different templates), dropout, or the number of recycles, which produced an array of structure predictions. The final dimer models for structural analysis were selected using two criteria: (1) the highest average pLDDT score and (2) the lowest average PAE score for the residue pairs that were in contact (< 6 Å apart).

Coevolution analysis

The coevolution analysis was performed using EVcouplings [Hopf TA, et al. 2019]. The method generates multiple sequence alignments (MSA) for each of the two sequences using Jackhammer (with a Bitscore cutoff of 0.3) and then combines them into a single combined MSA. The measure of whether MSA is sufficiently deep is the number of effective sequences divided by the length of the protein, which should be above 1.

Active site prediction

The potential active sites were predicted using the COACH meta-server [Yang J, et al. 2013]. It is a consensus method that uses multiple different methods to predict potential active sites. Only methods that correctly predicted substrates known for these proteins were taken into consideration. The residues were considered as a part of the active site if the majority of methods predicted them to form the active center.

Acknowledgments

The authors are deeply grateful to prof. Marcin Czerwinski for his invaluable insights and thoughtful suggestions throughout the creation of this manuscript.

Author contributions

Krzysztof Mikolajczyk (Conceptualization [lead], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [lead], Software [supporting], Supervision [lead], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]),

Karol Wróblewski (Data curation [equal], Formal analysis [supporting], Investigation [equal], Methodology [equal], Software [lead], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), **Sebastian Kmiecik** (Conceptualization [supporting], Data curation [supporting], Formal analysis [supporting], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [supporting], Software [supporting], Supervision [supporting], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]).

Abbreviations

A4galt, α1,4-galactosyltransferase; B4galt, β1,4-galactosyltransferase; Gb3, globotriaosylceramide, Gb3Cer, CD77, P^k antigen, ceramide trihexoside, Galα1→4Galβ1→4Glc-Cer; GP, glycoprotein; GL, glycolipid; HUS, hemolytic-uremic syndrome; P1 glycotope, Galα1→4Galβ1→4GlcNAc-R; LacCer, lactosylceramide; NanoBiT, NanoLuc® Binary Technology; PPI, protein-protein interaction; STEC, Shiga toxin-producing *Escherichia coli*; Stx, Shiga toxin; Stx1B, Shiga toxin 1B subunit; Stx2B, Shiga toxin 2B subunit.

Funding

This research was funded by the National Science Centre of Poland, PRELUDIUM 20 Project 2021/41/N/NZ6/00949 (K.M). K.W. and S.K. acknowledge funding from the National Science Centre [2020/39/B/NZ2/01301].

Conflict of interest statement

None declared.

References

Badaczewska-Dawid AE, Nithin C, Wroblewski K, Kurcinski M, Kmiecik S. 2022. MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes. *Nucleic Acids Res*. 50:W474-W482.

Bereznicka A, Mikolajczyk K, Szymczak-Kulus K, Kapczynska K, Majorczyk E, Modlinska A, et al. 2021. Two Paralogous Gb3/CD77 Synthases in Birds Show Different Preferences for Their Glycoprotein and Glycosphingolipid Substrates. *Int J Mol Sci*. 22:9761.

Bruyand M, Mariani-Kurkdjian P, Gouali M, de Valk H, King LA, Le Hello S, et al. 2018. Hemolytic uremic syndrome due to Shiga toxin-producing Escherichia coli infection. *Med Mal Infect*. 48:167-174.

Bydlinski N, Maresch D, Schmieder V, Klanert G, Strasser R, Borth N. 2018. The contributions of individual galactosyltransferases to protein specific N-glycan processing in Chinese Hamster Ovary cells. *J Biotechnol*. 282:101-110.\

Cooling L. 2015. Blood Groups in Infection and Host Susceptibility. *Clin Microbiol Rev*. 28:801-70.

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. 2022. Protein complex prediction with AlphaFold-Multimer. bioRxiv doi: 10.1101/2021.10.04.463034.

Giraudo CG, Daniotti JL, Maccioni HJ. 2001. Physical and functional association of glycolipid N-acetyl-galactosaminyl and galactosyl transferases in the Golgi apparatus. *Proc Natl Acad Sci U S A*. 98:1625-30.

Giraudo CG, Maccioni HJ. 2003. Ganglioside glycosyltransferases organize in distinct multienzyme complexes in CHO-K1 cells. *J Biol Chem*. 278:40262-71.

Harrus D, Khoder-Agha F, Peltoniemi M, Hassinen A, Ruddock L, Kellokumpu S, et al. 2018. The dimeric structure of wild-type human glycosyltransferase B4GalT1. *PLoS One*. 13:e0205571.

Hassinen A, Pujol FM, Kokkonen N, Pieters C, Kihlström M, Korhonen K, et al. 2011. Functional organization of Golgi N- and O-glycosylation pathways involves pH-dependent complex formation that is impaired in cancer cells. *J Biol Chem*. 286:38329-38340.

Hopf TA, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, et al. 2019. The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics*. 35:1582-1584.

Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, et al. 2007. Crystal structure of mammalian alpha1,6-fucosyltransferase, FUT8. *Glycobiology*. 17:455-66.

Johannes L, Römer W. 2010. Shiga toxins-from cell biology to biomedical applications. *Nat Rev Microbiol*. 8:105-16.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. 596:583-589.

Kaczmarek R, Buczkowska A, Mikołajewicz K, Krotkiewski H, Czerwinski M. 2014. P1PK, GLOB, and FORS blood group systems and GLOB collection: biochemical and clinical aspects. Do we understand it all yet? *Transfus Med Rev*. 28:126-36.

Kaczmarek R, Duk M, Szymczak K, Korchagina E, Tyborowska J, Mikolajczyk K, et al. 2016. Human Gb3/CD77 synthase reveals specificity toward two or four different acceptors depending on amino acid at position 211, creating P(k), P1 and NOR blood group antigens. *Biochem Biophys Res Commun*. 470:168-174.

Kaczmarek R, Suchanowska A, Lisowska E, Czerwinski M. 2013. Gb3/CD77 synthase (α1,4-galactosyltransferase) and its variant form, NOR-synthase, exist as dimers. Conference: 38th FEBS Congress, St. Petersburg; 07/2012 At: St. Petersburg Volume: 280S1.

Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 51:D587-D592

Kellokumpu S, Hassinen A, Glumoff T. 2016. Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. *Cell Mol Life Sci*. 73:305-25.

Khoder-Agha F, Harrus D, Brysbaert G, Lensink MF, Harduin-Lepers A, Glumoff T, et al. 2019. Assembly of B4GALT1/ST6GAL1 heteromers in the Golgi membranes involves lateral interactions via highly charged surface domains. *J Biol Chem*. 294:14383-14393.

Kim JS, Lee MS, Kim JH. 2020. Recent Updates on Outbreaks of Shiga Toxin-Producing Escherichia coli and Its Potential Reservoirs. *Front Cell Infect Microbiol*. 10:273.

Lee J, Sundaram S, Shaper NL, Raju TS, Stanley P. 2001. Chinese hamster ovary (CHO) cells may express six beta 4-galactosyltransferases (beta 4GalTs). Consequences of the loss of functional beta 4GalT-1, beta 4GalT-6, or both in CHO glycosylation mutants. *J Biol Chem*. 276:13924-34.

Lujan P, Campelo F. 2021. Should I stay or should I go? Golgi membrane spatial organization for protein sorting and retention. *Arch Biochem Biophys*. 707:108921.

Maccioni HJ. 2007. Glycosylation of glycolipids in the Golgi complex. *J Neurochem*. 103 Suppl 1:81-90.

Michael M, Bagga A, Sartain SE, Smith RJH. 2022. Haemolytic uraemic syndrome. *Lancet*. 400:1722-1740.

Mikolajczyk K, Bereznicka A, Szymczak-Kulus K, Haczkiewicz-Lesniak K, Szulc B, Olczak M,et al. 2021. Missing the sweet spot: one of the two N-glycans on human Gb3/CD77 synthase is expendable. *Glycobiology*. 31:1145-1162.

Mikolajczyk K, Sikora M, Hanus C, Kaczmarek R, Czerwinski M. 2022. One of the two N-glycans on the human Gb3/CD77 synthase is essential for its activity and allosterically regulates its function. *Biochem Biophys Res Commun*. 617:36-41.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods*. 19:679-682.

Morimoto K, Suzuki N, Tanida I, Kakuta S, Furuta Y, Uchiyama Y, et al. 2020. Blood group P1 antigen-bearing glycoproteins are functional but less efficient receptors of Shiga toxin than conventional glycolipid-based receptors. J Biol Chem. 295:9490-9501.Cody EM, Dixon BP. 2019. Hemolytic Uremic Syndrome. *Pediatr Clin North Am*. 66:235-246.

Rizzo R, Russo D, Kurokawa K, Sahu P, Lombardi B, Supino D, et al. 2021. Golgi maturation-dependent glycoenzyme recycling controls glycosphingolipid biosynthesis and cell growth via GOLPH3. *EMBO J*. 40:e107238.

Seko A, Yamashita K. 2008. Activation of beta1,3-N-acetylglucosaminyltransferase-2 (beta3Gn-T2) by beta3Gn-T8. 2008. Possible involvement of beta3Gn-T8 in increasing poly-N-acetyllactosamine chains in differentiated HL-60 cells. *J Biol Chem*. 283:33094-100.

Shauchuk A, Szulc B, Maszczak-Seneczko D, Wiertelak W, Skurska E, Olczak M. 2020. N-glycosylation of the human β1,4-galactosyltransferase 4 is crucial for its activity and Golgi localization. *Glycoconj J*. 37:577-588.

Suchanowska A, Kaczmarek R, Duk M, Lukasiewicz J, Smolarek D, Majorczyk E, et al. 2012. A single point mutation in the gene encoding Gb3/CD77 synthase causes a rare inherited polyagglutination syndrome. *J Biol Chem*. 287:38220-30.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 47:D607-D613.

Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 51:D638-D646.

Szymczak-Kulus K, Weidler S, Bereznicka A, Mikolajczyk K, Kaczmarek R, Bednarz B, et al. 2021. Human Gb3/CD77 synthase produces P1 glycotope-capped N-glycans, which mediate Shiga toxin 1 but not Shiga toxin 2 cell entry. *J Biol Chem*. 296:100299.

Takematsu H, Yamamoto H, Naito-Matsui Y, Fujinawa R, Tanaka K, Okuno Y, et al. 2011. Quantitative transcriptomic profiling of branching in a glycosphingolipid biosynthetic pathway. *J Biol Chem*. 286:27214-24.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. 50:D439-D444.\

Verma N, Tao Y, Kraka E. 2021. Systematic Detection and Characterization of Hydrogen Bonding in Proteins via Local Vibrational Modes. *J Phys Chem B*. 125:2551-2565.

Volkers G, Worrall LJ, Kwan DH, Yu CC, Baumann L, Lameignere E, et al. 2015. Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nat Struct Mol Biol*. 22:627-35.

Wiertelak W, Sosicka P, Olczak M, Maszczak-Seneczko D. 2020. Analysis of homologous and heterologous interactions between UDP-galactose transporter and beta-1,4-galactosyltransferase 1 using NanoBiT. *Anal Biochem*. 593:113599.

Yang J, Roy A, Zhang Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 29:2588-95.

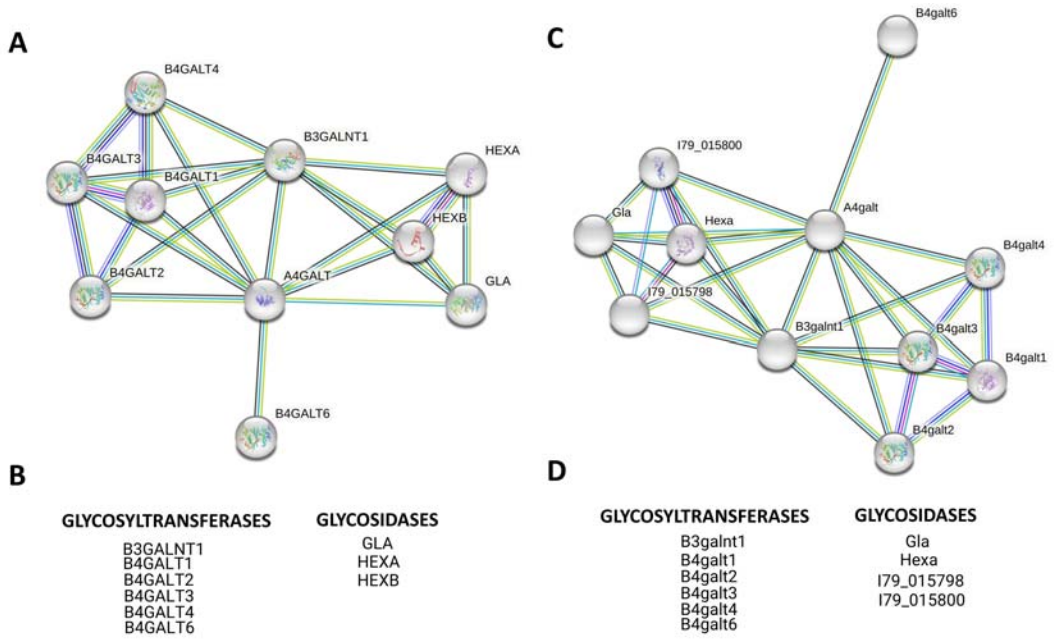

Fig. 1. Protein-protein interaction (PPI) network analysis of human A4galt in *H. sapiens* (**A**) and *C. griseus* (**C**) organisms was performed using the STRING v12 database analysis tool. Only nodes representing potential heterologous partners with the highest confidence (>□0.9) and functional correlation (glycosylation pathway-related) are visualized in the network. Colored lines between the nodes indicate the different types of interaction evidenced by gene neighborhood (green line), cooccurrence across species (blue line), experimental evidence (purple line), text mining of abstracts from literature (yellow line), databases (light blue line), co-expression in the same or others species (black line). Below the network map, the determined protein partners (belonging to two enzyme classes, glycosyltransferases and glycosidases) for human A4galt in *H. sapiens* (**B)** and *C. griseus* (**D**) were shown.

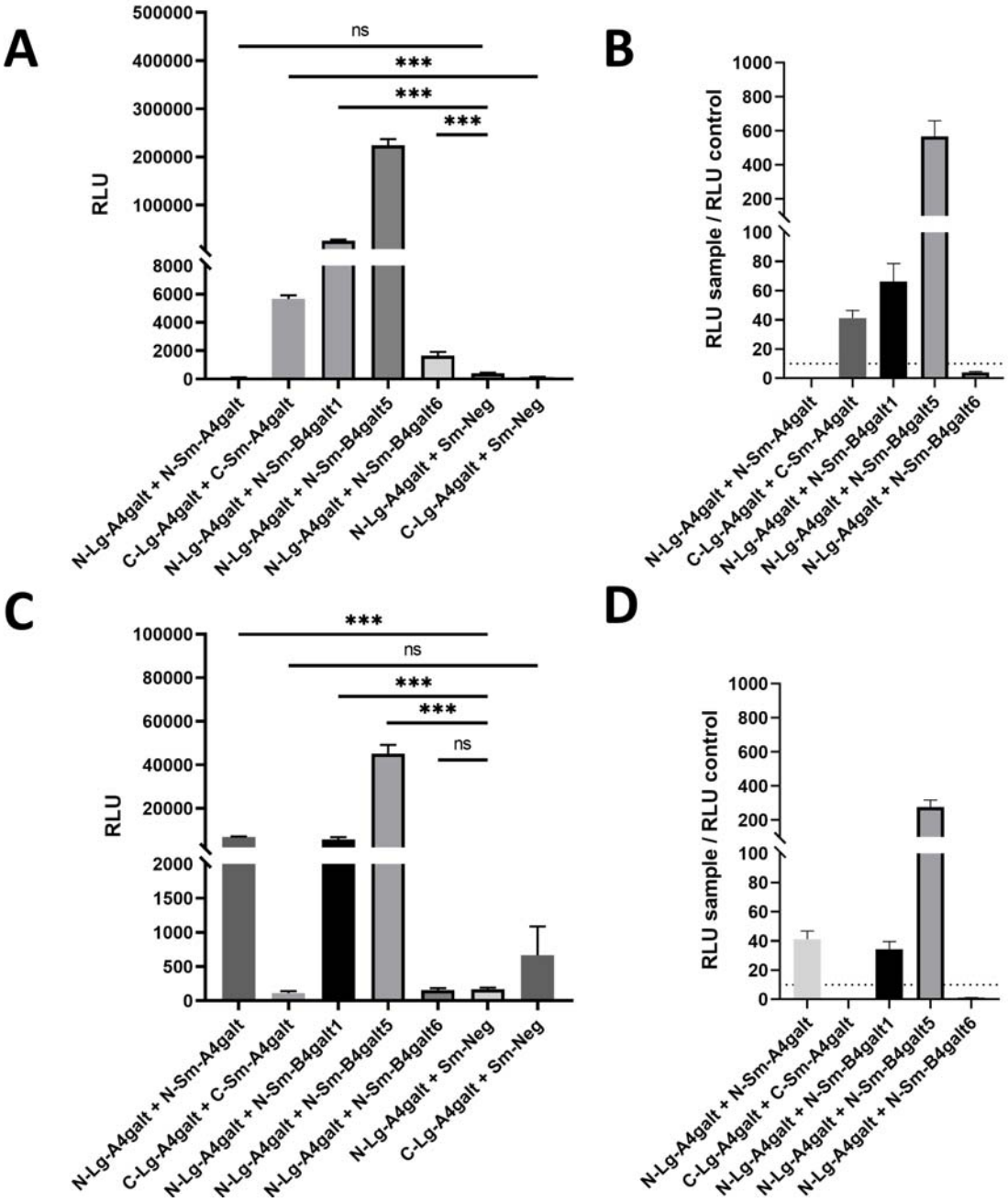

Fig. 2. NanoBiT assay was performed for high-frequency human A4galt (**A**) and its mutein form (**C**) with the potential heterologous protein partners (B4galt1, B4galt5, and B4galt6 isoenzyme) using HEK293T cells. The negative controls comprise HaloTag fused with the small NanoLuc subunit. The fold changes of luminescence were calculated by dividing the average luminescence measured for the tested combinations (Sample RLU) was divided by the average luminescence obtained for the corresponding negative controls (Control RLU) to calculate the fold changes of the interaction pairs for high-frequency human A4galt (**B**) and its

mutein (**D**). The threshold value was set at 10, according to the manufacturer. N-Lg-A4galt, human A4galt N-terminated tag with the LgBiT NanoLuc luciferase subunit; N-Lg-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme N-terminated fused with the LgBiT NanoLuc luciferase subunit; C-Lg-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme C-terminated fused with the LgBiT NanoLuc luciferase subunit; N-Sm-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme N-terminated fused with the SmBiT NanoLuc luciferase subunit; C-Sm-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme C-terminated fused with the SmBiT NanoLuc luciferase subunit.

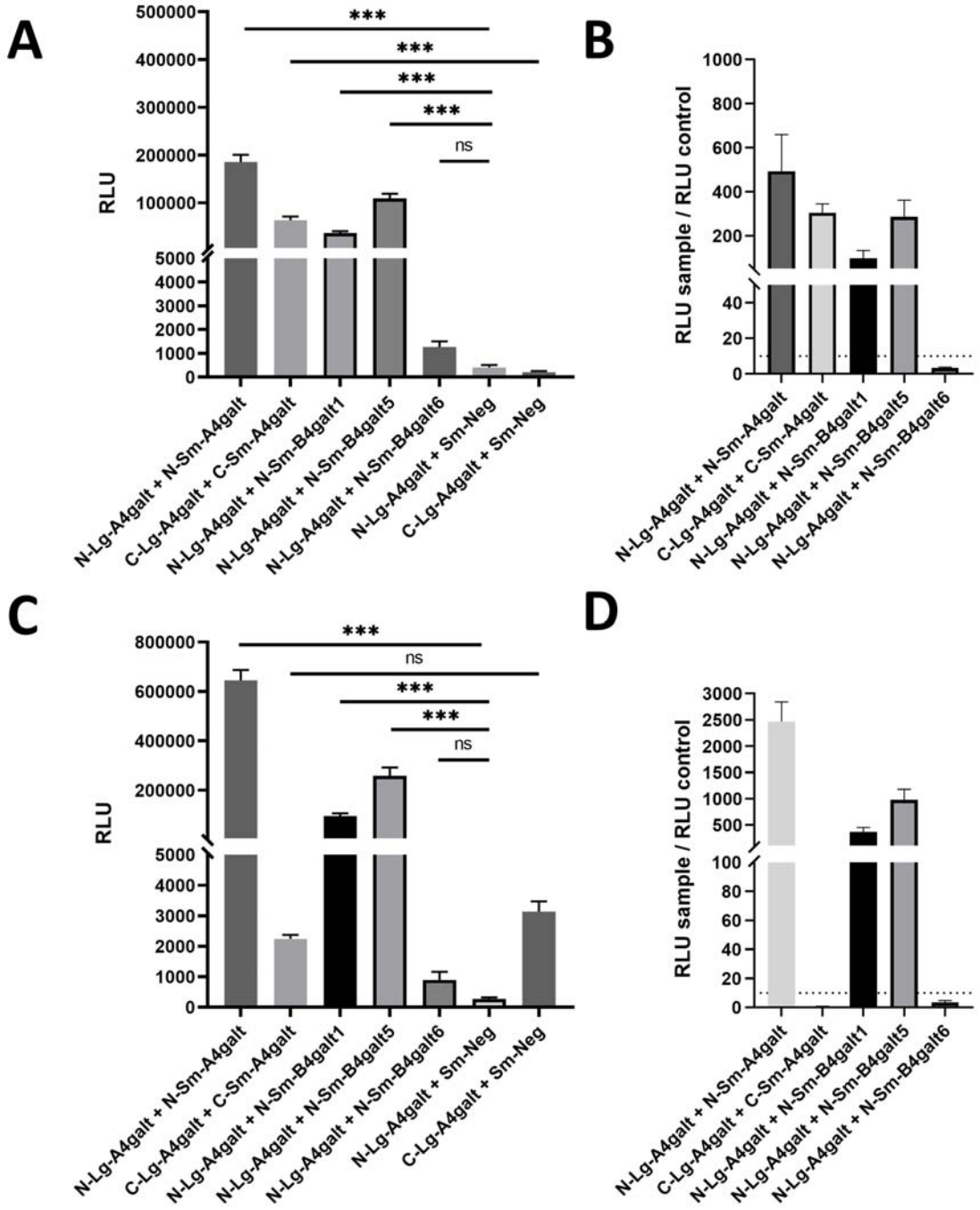

Fig. 3. NanoBiT assay performed for high-frequency human A4galt (**A**) and its mutein form (**C**) with the potential heterologous protein partners (B4galt1, B4galt5, and B4galt6 isoenzyme) using CHO-Lec2 cells. The negative controls comprise HaloTag fused with the small NanoLuc subunit. The fold changes of luminescence were calculated by dividing the average luminescence measured for the tested combinations (Sample RLU) was divided by the average luminescence obtained for the corresponding negative controls (Control RLU) to

calculate the fold changes of the interaction pairs for high-frequency human A4galt (**B**) and its mutein (**D**). The threshold value was set at 10, according to the manufacturer. N-Lg-A4galt, human α1,4-galactosyltransferase N-terminated tag with the LgBiT NanoLuc luciferase subunit; N-Lg-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme N-terminated fused with the LgBiT NanoLuc luciferase subunit; C-Lg-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme C-terminated fused with the LgBiT NanoLuc luciferase subunit; N-Sm-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme N-terminated fused with the SmBiT NanoLuc luciferase subunit; C-Sm-A4galt/B4galt1/B4galt5/B4galt6, human α1,4-galactosyltransferase/β1,4-galactosyltransferase 1/β1,4-galactosyltransferase 5 isoenzyme/β1,4-galactosyltransferase 6 isoenzyme C-terminated fused with the SmBiT NanoLuc luciferase subunit.

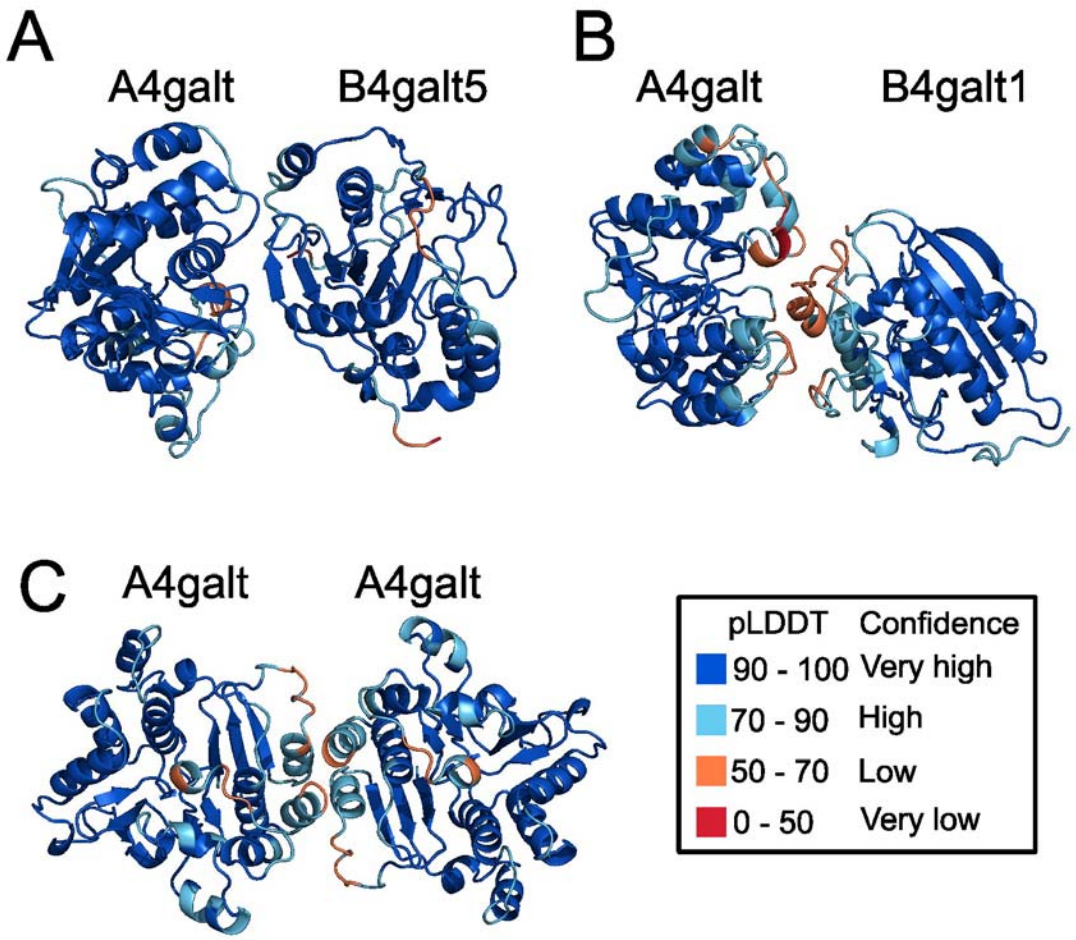

Fig. 4. Structure predictions by Alphafold-Multimer: (A) heterodimer A4galt (left) - B4galt1 (right); (B) homodimer A4galt - A4galt; (C) heterodimer A4galt (left) - B4galt5 (right). Coloring based on pLDDT confidence scores.

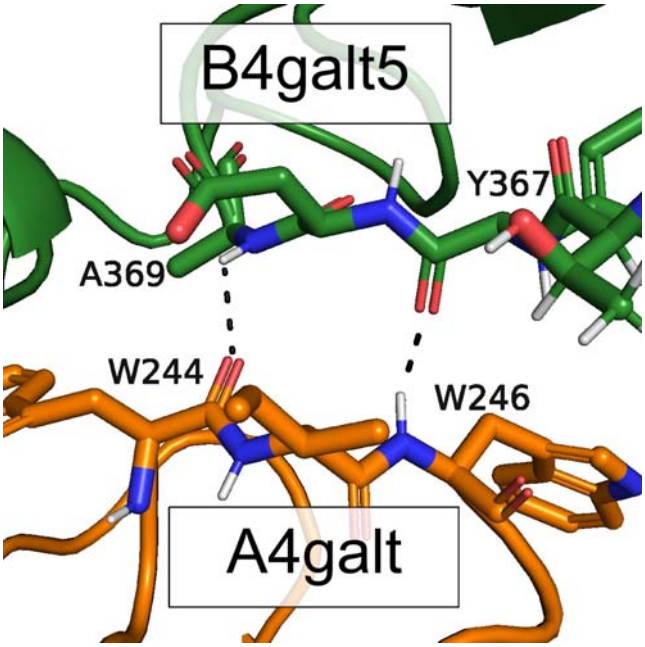

Figure 5. Potential hydrogen bonds in heterodimer A4galt (bottom, in orange) - B4galt5 (bottom, in green).

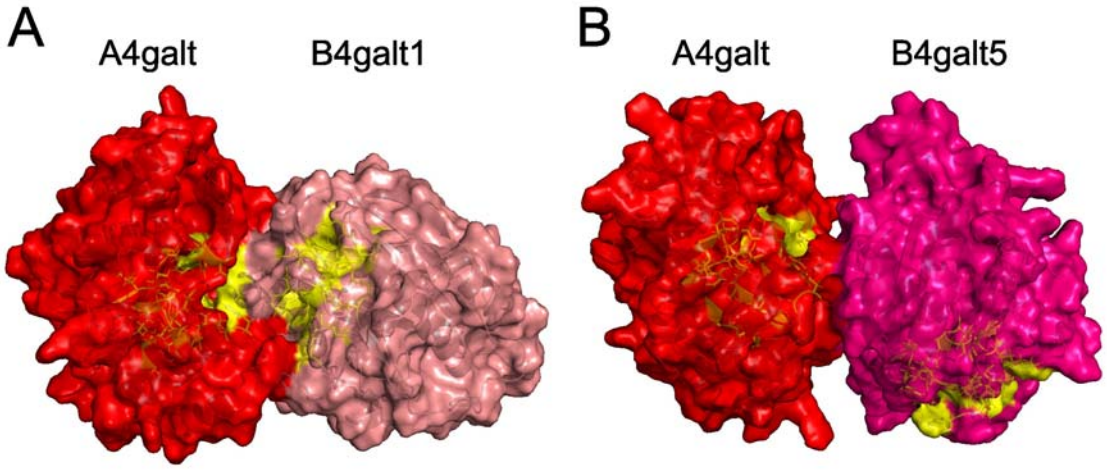

Figure 6. The mutual relationship of active sites (marked by yellow) in heterodimers (A) A4galt (left) - B4galt1 (right); (B) A4galt (left) - B4galt5 (right).