

OPEN

Structural and functional role of disulphide bonds and substrate binding residues of the human beta-galactoside alpha-2,3-sialyltransferase 1 (hST3Gal1)

Maria Elena Ortiz-Soto¹, Sabine Reising¹, Andreas Schlosser² & Jürgen Seibel^{1*}

Overexpression of hST3Gal1 leads to hypersialylation of cell-surface glycoconjugates, a cancer-associated condition that promotes cell growth, migration and invasion. Upregulation of this enzyme in ovarian cancer is linked to cancer progression and metastasis, contributing also to chemotherapy resistance. Strategies for preventing metastasis include the inhibition of hST3Gal1, which demands structure-based studies on its strict regioselectivity and substrate/donor preference. Herein we describe the contribution of various residues constituting donor CMP-Neu5Ac and acceptor Gal β 1-3GalNAc-R binding sites to catalysis. Removal of hydrogen bonds and/or stacking interactions among substrates and residues Y191, Y230, N147, S148 and N170 affected the enzyme's activity to a different extent, revealing the fine control needed for an optimal catalytic performance. To gain further understanding of the correlation among structure, activity and stability, the *in vitro* role of hST3Gal1 disulphide bonds was analysed. As expected, disruption of the Glycosyltransferase family 29 (GT29) invariant bond C142-C281, as well as the ST3Gal1 subfamily conserved disulphide C61-C139 inactivates the enzyme. While disulphide C59-C64 is not essential for function, its absence reduces the activity (k_{cat}) for donor and acceptor substrates to about 67 and 72%, respectively, and diminishes the enzyme's melting temperature (T_m) by 7°C.

Glycans of different complexity that decorate the cell surface of organisms of the three domains of life are involved in a variety of physiological and pathological events¹⁻³. Glycoconjugates containing terminal sialic acids participate in biological recognition processes such as system embryogenesis, cell trafficking, transmembrane signalling, immunological regulation, tumor invasion and host-pathogen interactions^{2,4-6}. In mammals the transfer of sialic acid from donor CMP-Sia is catalysed by twenty sialyltransferases classified into the Glycosyltransferase family 29 (GT29) (<http://www.cazy.org/>)^{7,8}. GT29 enzymes are Type II membrane proteins confined to the Golgi apparatus. They present a transmembrane domain (16–20 amino acids) for Golgi retention, which is flanked by a short N-terminal cytoplasmic tail and a C-terminal stem region followed by the catalytic domain⁹. Sialyltransferases vary in terms of their substrate specificity, tissue distribution, and biochemical parameters⁸. They are classified into subfamilies ST3Gal1-6, ST6Gal1-2, ST6GalNAc1-6 and ST8Sia1-6, based on their regioselectivity for the transferred sialic acid and their acceptor specificity^{10,11}. ST3Gal and ST6Gal enzymes form α 2-3- and α 2-6-linkages, respectively, on oligosaccharide chains containing a terminal galactose. ST6GalNAc and ST8Sia generate α 2-6- and α 2-8-bonds, respectively. Proteins in the former subfamily transfer sialic acid onto terminal GalNAc residues and enzymes from the latter recognise acceptors carrying an external sialic acid unit. Sialyltransferases are inverting enzymes regarding the configuration of the anomeric centre of sialic acid upon transfer, which occurs via a direct displacement S_N2 -like mechanism^{12,13}. Members within each subfamily show a typical pattern of conserved cysteine residues involved in the formation of disulphide bonds that are important for proper folding and activity^{9,14}. Multiple sequence alignments of the catalytic domain reveal highly conserved

¹Institut für Organische Chemie, Universität Würzburg, Am Hubland, 97074, Würzburg, Germany. ²Rudolf-Virchow-Zentrum für Experimentelle Biomedizin, Universität Würzburg, Josef-Schneider Str. 2, Haus D15, 97080, Würzburg, Germany. *email: seibel@chemie.uni-wuerzburg

regions denoted as sialylmotifs L (long), S (short), III and VS (very short). Mutational studies performed during the last decades have contributed to a better understanding of the role of residues within the sialylmotifs^{12,15,16}. Residues contributing to substrate binding are localized in the sialylmotifs L and S, while amino acids from sialylmotifs III and VS participate in catalytic activity^{13,17}. Crystal structures of sialyltransferases from the four subfamilies exhibit a single Rossmann-like (GT-A variant 2) fold consisting of seven central β -strands flanked by multiple α -helices^{13,15,17–19}. Differences in the helical and loop segments that include various insertion and deletion sequences are presumably responsible for the rigorous linkage and/or acceptor specificities displayed by sialyltransferases from family GT29.

Even minor changes in glycosylation provoke anomalous cellular processes. Hypersialylation for instance correlates with cancer stage and disease prognosis^{20,21}. High expression levels of hST3Gal1 and other sialyltransferases are observed in different cancer types; increased α -2,3 sialylated core 1 O-glycans are common in breast, bladder and colon cancer²⁰. Overexpression of hST3Gal1 is also observed in ovarian, prostatic and pancreatic cancer⁸. Sialyltransferases and other enzymes that participate in the synthesis and processing of sialic acids are therefore considered as potential targets for drug development, and inhibition of sialyltransferases has been proposed as a strategy to prevent metastasis of different types of cancer⁸. The elucidation of protein-ligand molecular interactions with donor and acceptor molecules are significant components of inhibitor design. Although the three-dimensional structure of hST3Gal1 has not been solved, the structure of its closely related porcine homolog ST3Gal1 (pST3Gal1) to which it shares 85% identity was determined in the apo-form, as a binary complex with acceptor Gal β 1,3GalNAc α -PhNO₂ and as a ternary complex with the acceptor and CMP¹³. The latter was used as a template for modelling hST3Gal1 and to investigate the role of various residues in the donor and acceptor binding sites that were not previously explored. The effect of altering residues N147, S148 and N170 which are located at the donor binding site, and residue V315, which is adjacent to the catalytic H316 is described. The contribution of residues Y191 and Y230 to catalysis was also determined and compared to corresponding mutations reported in pST3Gal1¹². ST3Gal1 enzymes possess three disulphide bonds, with the pair formed by C142 and C281 being completely conserved among all members of family GT29. In the present work we explored the implication of disulphide bonds in activity and protein stability. While native disulphides C61-C139 and C142-C281 are necessary for activity, the bond formed between C59 and C64 is not absolutely required for proper folding and activity, yet it contributes to the enzyme thermal stability in *in vitro* assays.

Results and Discussion

Homology modelling of hST3Gal1 and selection of variants in donor and acceptor binding sites.

Two homology models of hST3Gal1 based on the apo- and substrate-bound crystal structures of pST3Gal1 (PDB codes 2wml and 2wnb, respectively)¹³ were generated employing the SWISS-MODEL server²². Structural alignment of models and templates performed with UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera>) provided global RMSD (root-mean-square deviation) values of 0.073 and 0.063 Å for the models based on the apo- and substrate-bound structures, respectively.

As mentioned above, the structure of pST3Gal1 displays seven β -strands flanked by 12 α -helices, with the donor and acceptor binding sites localized in a cleft formed by the β -core and some α -helices and loops of variable length¹³. Residues constituting hST3Gal1 substrate binding sites show a clear spatial alignment with those in the porcine homolog in both models (Supplementary Information, Fig. S1). On this basis we assumed that changes induced in the side chain of some amino acids upon substrate binding are similar in both enzymes and hence employed the model based on pST3Gal1 bound-structure (2wnb) to identify target residues for mutagenesis (Fig. 1 and Supplementary Information, Fig. S1).

Enzymes from the four sialyltransferase subfamilies, namely ST3Gal1-6, ST6GalNAc1-6, ST8Sia1-6 and ST6Gal1-2 possess low overall sequence identity when compared with one another. They present however conserved regions known as sialylmotifs L, S, III and VS^{23,24}, that include some of the residues forming the substrates binding sites and a pair of conserved cysteines. Sialylmotif L is constituted by 53 amino acids²⁵, spanning in hST3Gal1-6 from the invariant C139 to Y191, while sialylmotif S includes 24 amino acids (from P267 to G290) and contains the other conserved cysteine (C281). Sialylmotif III is only 4 amino acids long (²⁹⁹H-Y-Y/f/w/h-E/D/g/q/n/s/t/k³⁰²) and contains an invariant histidine (H299) and a conserved tyrosine (Y300). The sialylmotif VS comprises 6 residues (³¹⁶H-X-X-X-E³²¹), starting with the catalytic base H316 and ending with a fully conserved glutamic acid. Sialylmotif sequences for subfamily ST3Gal1-6 are shown in Supplementary Information, Fig. S2.

Residues equivalent to Q105, Y191, Y230, Y266, T269, W297, H299, F310, R311, K312 and the catalytic base H316 were previously mutated in pST3Gal1¹² (Fig. 1A). Likewise, various mutations were reported for hST3Gal1, albeit only variants H299A/Y and H316A were explored within the acceptor/donor binding sites²³ (Fig. 1A). Exchange of H299, Y300 and H316 by alanine rendered hST3Gal1 variants inactive, while mutants W301F, E302Q, and E321Q retained 25–80% of the wild type activity²³.

Residues Y191, H299 and catalytic base H316 belong to the sialylmotifs L, III and VS, respectively. Y266 precedes sialylmotif S, while T269 is part of it. Here we investigated the effect of replacing residues N147, S148, N170, Y191, Y230, and V315 on hST3Gal1 activity. All residues except the latter interact with donor and/or acceptor substrates (Fig. 1B). Residues N147, S148, N170 and Y191 belong to the sialylmotif L.

V315 is replaced by either tyrosine or tryptophan in enzymes from the subfamily ST6Gal1-2, which transfer sialic acid to the C6-hydroxyl group of the terminal galactose unit of glycoconjugates and produce α 2-6-linkages. Aromatic stacking between Y369 (the residue found at the equivalent position to V315 in hST6Gal1) with the nonpolar face of the terminal galactosyl moiety of the carbohydrate acceptor possibly contributes to the regioselectivity of ST6Gal enzymes (Supplementary Information, Fig. S3). This and other interactions mostly involving tyrosine residues, place the galactose C6-hydroxyl group close to the N ϵ atom of the catalytic base histidine and the C2 position of Neu5Ac^{15,17}. Mutation of the equivalent tyrosine residue by alanine in rat ST6Gal1 (Y366)

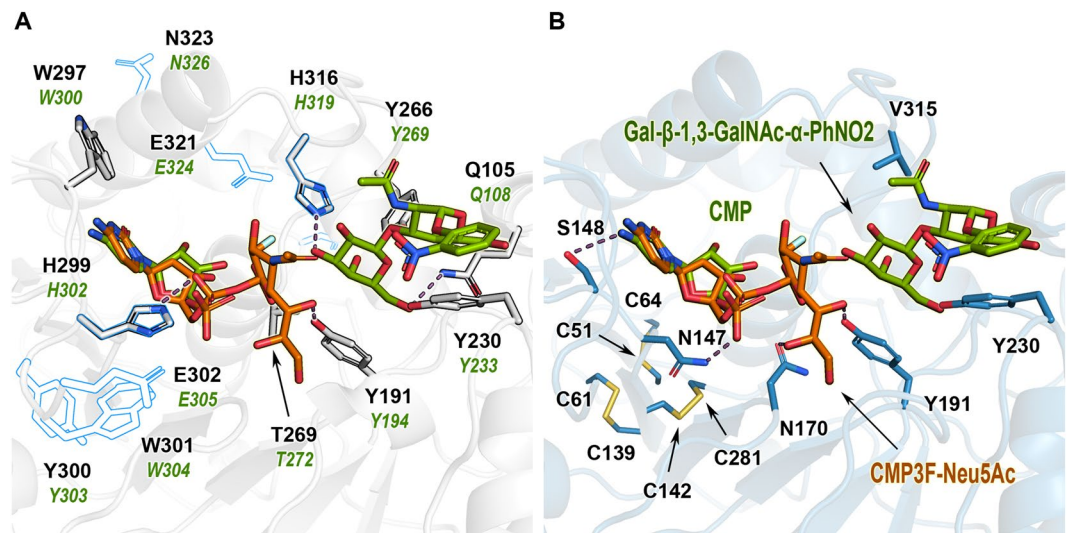


Figure 1. A view of the active site of human and porcine ST3Gal1 enzymes. **(A)** Cartoon representation of pST3Gal1 (PDB 2wnb). Positions previously mutated in pST3Gal1¹² and hST3Gal1²³ are shown as white sticks and as blue outlines, respectively. In the latter enzyme, only H316 and H299 are part of the active/binding site. Amino acid sequence numbering for porcine and human enzymes is in green and black font, respectively. **(B)** Cartoon representation of modelled hST3Gal1. Positions selected for mutagenesis of hST3Gal1 in this paper are displayed as blue sticks. CMP and acceptor Gal- β -1,3-GalNAc- α -PhNO₂ from pST3Gal1 (PDB 2wnb, green sticks) as well as CMP3F-Neu5Ac from the sialyltransferase CstII of *Campylobacter jejuni* (PDB 1ro7, orange sticks) are shown in the active site. Possible interactions between binding site residues and substrates (<4 Å) are displayed as dashed lines. Disulphide bonds are located >16 Å away from donor and acceptor binding sites. Structures of CstII and pST3Gal1 as well as the model of hST3Gal1 were aligned with UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera>).

renders the enzyme inactive¹⁵. We analysed the replacement of V315 by alanine as well as tyrosine and analysed the effect of the mutations on the enzyme's activity and regioselectivity.

Another structural feature that differentiates GT29 enzymes is the subfamily-specific conservation of disulphide bonds. Only the disulphide bond formed between C142-C281 is conserved throughout the GT29 family. Enzymes from the subgroups ST3Gal-1, -2 and -6 contain 3 disulphide bonds in the catalytic domain, while ST3Gal-3 and -5 can only form two bonds.

We analysed the significance of the three disulphide bonds of hST3Gal1 in activity and stability. For this purpose, cysteine pairs involved in disulphide bonds were replaced by serine in the double variants C59S-C64S, C61S-C139S and C142S-C281S.

A 3DM superfamily database (Bio-Product) containing originally 3642 sequences of annotated, putative and predicted GT29 sialyltransferases from different organisms was utilized to analyse the conservation of residues involved in contacts with the donor and acceptor molecules, after eliminating redundancy and sequences with less than 300 amino acids. The 3DM superfamily comprises sequences with a common fold and uses a structural core alignment to generate a phylogenetic tree, thereby forming groups of sequences that are classified as subfamilies²⁶. Each group contains sequences that are at least 80% identical at the core (aligned) positions²⁷. Only 271 and 43 non-redundant sequences belonging to the subfamily ST3Gal1-6 and the subgroup ST3Gal1, respectively, were employed for analyses herein.

An N-terminally truncated version of hST3Gal1 (Δ 45) lacking the transmembrane region was fused to the maltose binding protein (MBP) to improve its solubility and was co-expressed in *E. coli* Origami2 (DE3) with chaperone/foldases sulphhydryl oxidase and disulphide isomerase C, as reported previously²⁸. Both chaperone/foldases assist in the proper formation of disulphide bonds²⁹. Variants of hST3Gal1 were constructed using the wild-type gene cloned in the vector pMAL-C5x as a template and enzymes were expressed and purified by immobilized metal affinity chromatography.

Circular dichroism spectroscopy and kinetic characterization of variants. The integrity of the secondary structure of MBP-fused hST3Gal1 variants was explored by circular dichroism prior to their kinetic analysis. All variants (except N170A, for which not enough protein was obtained for structural analysis), showed a similar CD spectrum to the wild-type hST3Gal1 (Supplementary Information, Fig. S4). This excludes the possibility of mutation-induced secondary structure modification. The CD spectrum of the wild-type enzyme that appears in Supplementary Information, Fig. S4 was reported earlier²⁸. Because the appendage/cleavage of MBP does not modify the acceptor specificity nor the kinetic behaviour of porcine and human ST3Gal1^{12,13,28}, further characterization of all variants was performed with the fusion constructs.

hST3Gal1 is an inverting enzyme that catalyses the transfer of sialic acids onto a terminal galactose in the core 1 oligosaccharide motif Gal- β -1,3-GalNAc-R of Ser/Thr O-glycans and ganglio-series glycolipids to form

	Donor CMP-Neu5Ac			Acceptor Gal- β -1,3-GalNAc- α -O-Bn		
	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{mM}^{-1} \text{min}^{-1}$)	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{mM}^{-1} \text{min}^{-1}$)
ST3Gal1 ^a	106 (± 7)	201 (± 5)	1896	26 (± 1)	189 (± 2)	7269
N147S (150)	668 (± 159) ^b	75 (± 9)	112	211 (± 25)	48 (± 2)	227
S148A (151)	481 (± 55) ^b	219 (± 11)	455	105 (± 12)	139 (± 5)	1324
N170A (173)	184 (± 29)	37 (± 2)	201	96 (± 16)	39 (± 2)	406
Y191A (Y194)	876 (± 206) ^b	38 (± 5)	43	444 (± 88) ^b	28 (± 2)	63
Y230A (233)	205 (± 18)	150 (± 4)	732	854 (± 105) ^b	178 (± 11)	208
Y230F (233)	176 (± 19)	162 (± 6)	920	352 (± 26)	208 (± 5)	591
V315A (318)	180 (± 23)	63 (± 2)	350	1388 (± 254) ^b	145 (± 17)	105
V315Y (318)	271 (± 51)	8 (± 1)	28	549 (± 78) ^b	12 (± 1)	22
C59S/C64S						
(62–67)	121 (± 13)	135 (± 4)	1116	20 (± 2)	137 (± 3)	6850

Table 1. Kinetic parameters of MBP-fused hST3Gal1 and variants. Numbering corresponding to pST3Gal1 is given in parenthesis. Mean values and standard deviations of at least two replicates are shown. ^aValues for wild-type MBP-ST3Gal1 were taken from²⁸. ^bFor these variants saturation was not observed at the highest donor and acceptor concentrations employed (1.2 and 1.4 mM, respectively) (Supplementary Information, Figs. S5 and S6).

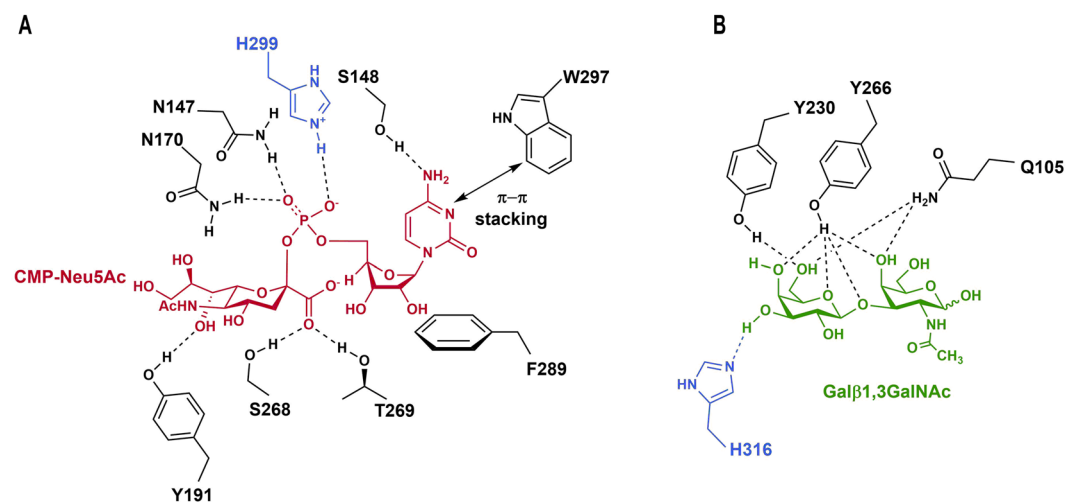


Figure 2. Residues involved in the recognition of donor CMP-Neu5Ac (A) and acceptor Gal β 1,3GalNAc (B) by ST3Gal1 enzymes^{12,13}. The catalytic acid H299 and catalytic base H316 are highlighted in blue.

α 2,3 linkages^{30,31}. Activity and kinetic parameters of wild-type hST3Gal1 and its variants were determined via a multi-enzymatic assay based on the catalysed release of CMP from donor CMP-Neu5Ac, which is coupled to the oxidation of NADH. Gal- β -1,3-GalNAc- α -O-Bn was employed as acceptor of the Neu5Ac moiety. The effect of replacing selected residues on the activity of hST3Gal1 (Table 1) is discussed in later sections.

Influence of mutations in the binding of donor and acceptor substrates. The homology model of hST3Gal1 using pST3Gal1 substrate bound structure (2wnb) as template proved a close overlapping of residues creating the binding sites of both enzymes (Supplementary Information, Fig. S1). Therefore, this model was employed to investigate the distances of selected residues to the donor and acceptor substrates and the possible effect of mutations on hST3Gal1 activity.

Because GT29 enzymes share CMP-Neu5Ac as a common donor, the side chains involved in its binding are highly conserved. As previously observed for pST3Gal1^{12,13} and based on the hST3Gal1 model, residue W297 forms an aromatic stacking interaction with the cytidine moiety while S148 stabilizes the position of CMP by hydrogen bonding. The hydrophobic side of ribose interacts with F289 and further binding is achieved by hydrogen bonds between N147 and N170 with CMP phosphate group (Fig. 2). The side chain conformation of N173 (N170 in hST3Gal1) observed in the CMP bound crystal structures of pST3Gal1 and in hST3Gal1 model is transposed in the structures of other sialyltransferases, for example ST6Gal1 (4js2), ST8SiaIII (5bo6) and the sialyltransferase CstII from *Campylobacter jejuni* (1ro7). In those structures asparagine can establish hydrogen bonds with the donor via its side chain nitrogen. This conformation depicted in Fig. 2 is most likely adopted by N170 during catalysis in order to neutralize the negative charge of the leaving group CMP^{12,17}. H299 acts as a

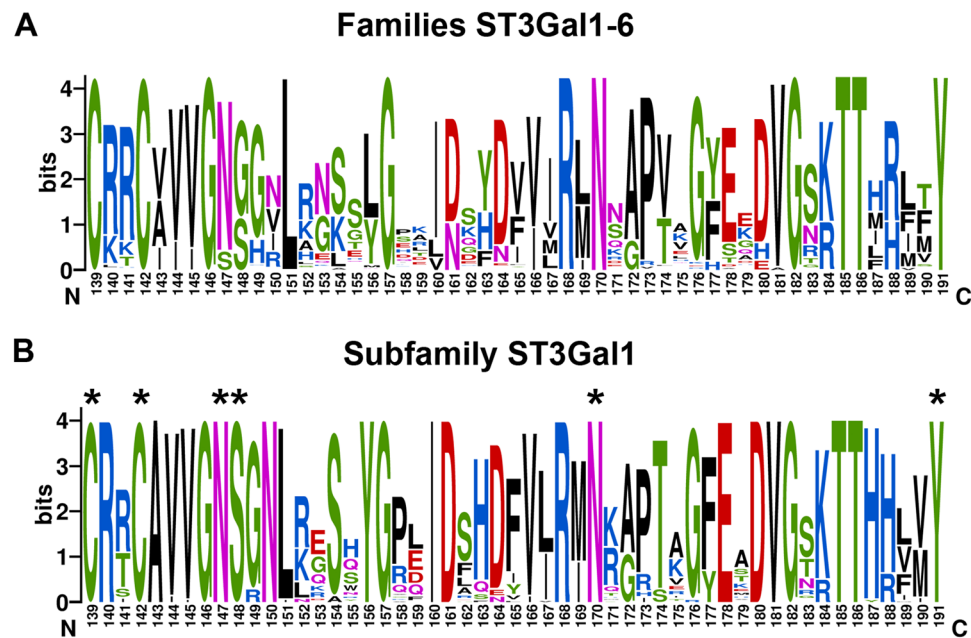


Figure 3. Conservation of sialylmotif L residues in the subfamily ST3Gal1-6 (A) and the subgroup ST3Gal1 (B). hST3Gal1 numbering is used. An asterisk indicates the residues that were mutated in this work. The graphical representation of sequence conservation was generated with WebLogo⁴⁷ using 271 and 43 non redundant sequences from ST3Gal1-6 and ST3Gal1 enzymes, respectively.

hydrogen bond donor and most importantly as the catalytic acid, concomitantly activating phosphate by enhancing its quality as a leaving group^{12,13}. Neu5Ac is engaged in polar interactions with Y191, S268 and T269 (Fig. 2).

In contrast to the conserved binding motif of the Neu5Ac donor, the substrate recognition site varies among sialyltransferases. This is expected due to the differing specificity for substrates and the formation of distinct glycosidic bonds. The substrate Gal- β -1,3-GalNAc-R is recognized by residues Q105 and Y266, which coordinate the C4-OH group of the GalNAc moiety. Additional hydrogen bonds are formed between the galactose C6-OH group and residues Y230 and Q105¹². Residue Y266 furthermore contributes to the organization of the GalNAc moiety by polar interactions with the C4-OH group and the acetal function (Fig. 2). H316 contributes crucially to binding as well as catalysis of sialic acid transfer as it acts as a hydrogen bond acceptor and as the catalytic base, thus activating the galactose C3-hydroxyl group for the nucleophilic attack and stabilizing the transition state. In this S_N2 like mechanism H299 and H316 act as acid and base catalysts, respectively (Fig. 2 and Supplementary Information, Fig. S7)^{12,13}. Both side chains stabilize the transition state, subsequently leading to a one step inverting reaction and forming an $\alpha,2,3$ bond after release of the CMP leaving group.

Residues N147 and S148 interact exclusively with the donor substrate and are invariant in all characterized and predicted ST3Gal1 enzymes (Figs. 2 and 3). A strong hydrogen bond (2.7 Å) occurs between the δNH_2 group of N147 and the phosphate group of CMP (Fig. 2 and Supplementary Fig. S8). CMP binds in the same conformation in the chemically different environment of all sialyltransferases with available crystal structure¹⁷; however, in most sequences from ST3Gal5 and ST6Gal1 a serine is found replacing asparagine at the position equivalent to N147. Although with a shorter side chain, serine is in principle able to coordinate the binding of CMP via its hydroxyl group. Substitution of asparagine by serine in variant N147S affected k_{cat} for donor and acceptor and increased the K_M for both substrates, showing no saturation at the assayed donor concentrations (Table 1 and Supplementary Fig. S5) and indicating that serine is unable to stabilize the donor binding in this enzyme.

In the subfamily ST3Gal1-6 only serine and glycine can be found at the position corresponding to S148 in hST3Gal1. While ST3Gal1 and ST3Gal2 enzymes have a serine, sialyltransferases from subgroups ST3Gal3-6 possess a glycine instead (Figs. 3 and 4). The exchange of serine by alanine eliminates the possibility to form a hydrogen bond between this residue and the amino group of the cytidine moiety. This change has a small effect on the k_{cat} for donor and acceptor, but affects K_M for CMP-Neu5Ac and Gal- β -1,3-GalNAc- α -O-Bn by 4.5- and 4-fold, respectively. Once again, saturation for the donor could not be observed at the measured CMP-Neu5Ac concentrations.

Residue N170 interacts with the phosphate group of CMP and can form a hydrogen bond with the C8-OH group of Neu5Ac, as observed from the structural alignment of porcine (2wnb) and human (model) ST3Gal1 with the sialyltransferase CstII from *Campylobacter jejuni* bound to CMP3F-Neu5Ac (PDB 1ro7) (Fig. 1 and Supplementary Fig. S8). N170 is entirely conserved through family ST3Gal1-6 (Fig. 3). Expression and purification of variant N170A were rather inefficient and unsurprisingly, substitution of N170 by alanine had a great impact on k_{cat} for donor and acceptor molecules, as well as on K_M .

The contribution of residues Y191 and Y230 to catalysis was previously analysed in corresponding variants Y194F and Y233A from pST3Gal1¹². Y191 is fully conserved in subfamily ST3Gal1-6 and its phenolic hydroxyl group forms a hydrogen bond with the C7-OH group of Neu5Ac and stacking interactions with the

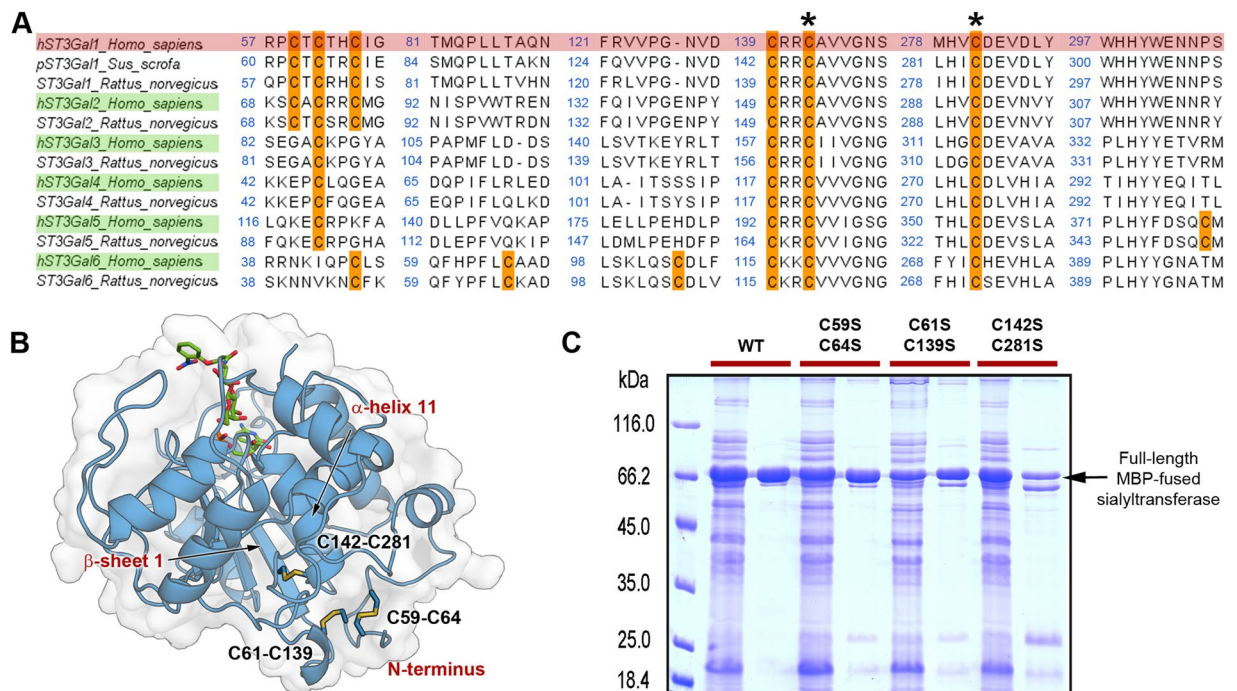


Figure 4. Position of disulphide bonds in subfamily hST3Gal1-6 and expression of hST3Gal1 disulphide bond variants. **(A)** Conservation of cysteine residues in subfamily ST3Gal1-6. Two asterisks indicate the conserved cysteines through the superfamily GT29. UniProt ID for sequences of human, porcine and rat sialyltransferases are shown in the next order: Q11201, Q02745, Q6H8N0, Q16842, Q11205, Q11203, Q2734, Q11206, P61131, Q9UNP4, Q68G12, Q9Y274 and P61943. **(B)** Location of disulphide bridges in hST3Gal1 (model). CMP and acceptor Gal- β -1,3-GalNAc- α -PhNO₂ from PDB 2wnb are displayed to show the location of the active site. **(C)** Reducing SDS-PAGE (12% acrylamide) showing hST3Gal1 wild-type and disulphide bond variants in soluble cleared extracts and after purification by immobilized metal affinity chromatography. The band observed below the full length sialyltransferase corresponds to a slightly degraded hST3Gal1 with the C-terminal His-tag missing (Supplementary Information, Figs. S11 and 12).

galactosyl-moiety of the acceptor (Figs. 1 and 2). Substitution of Y194 by phenylalanine in pST3Gal1 decreases the variant's activity on the donor by around 15-fold as a consequence of the removal of said hydrogen bond, leading to a reduction in the binding transition state stabilization of 1.7 kcal mol⁻¹¹². A more drastic effect was observed in this work for the modification Y191A in hST3Gal1 (Table 1), in particular for the acceptor. The exchange by alanine prompted naturally a severe decrease in k_{cat} , since the stacking interaction with galactose is removed. The reduction in the transition state stabilization corresponds to 2.3 and 2.9 kcal mol⁻¹ for donor and acceptor, respectively (Supplementary Information, Table S1). Substrate saturation was not achieved for the variant, neither for the donor nor for the acceptor.

Residue Y230 stacks with the *N*-acetylgalactosamine moiety of the acceptor and forms a hydrogen bond with the galactose C6-OH group (Figs. 1 and 2). Stacking is preserved when Y230 is exchanged by phenylalanine and thus only moderate changes are observed in the kinetic parameters of variant Y230F for CMP-Neu5Ac when compared to the wild-type activity (Table 1). Conversely, the loss of the hydrogen bond between galactose and the hydroxyl group of tyrosine reduces the catalytic efficiency for the acceptor by approximately 12-fold. Removal of both, hydrogen bonding and stacking interactions in variant Y230A causes a 35-fold activity reduction on the acceptor, a result comparable to a similar mutation (Y233A) in pST3Gal1, where a decrease in transition state stabilization of 1.2 and 2.1 kcal mol⁻¹ for donor and acceptor was observed¹². These results are in agreement with the values obtained for hST3Gal1 variant Y230A, as the removal of hydrogen bonding and stacking interactions causes a decrease in the transition state stabilization of 0.6 kcal mol⁻¹ for the donor and 2.2 kcal mol⁻¹ for the acceptor (Supplementary Information, Table S1).

β -galactoside α -2,6-sialyltransferase 1 (ST6Gal1) transfers Neu5Ac to the C6-hydroxyl group of a disaccharide motif containing a terminal galactose. Structurally ST3Gal1 and ST6Gal1 share a central core of about 200 amino acids, but differ in the surrounding loops, including an extra α -helix in ST6Gal1 that inverts its C-terminal β -strand with regard to ST3Gal1^{13,17} (Supplementary Information, Fig. S3). A clear difference between these two enzymes is the network of interactions that provide an optimal binding and orientation for the acceptor substrates, thus contributing to their regioselectivity. Residue Y369 is part of said network of contacts in enzyme ST6Gal1. Residue V315 from hST3Gal1 occupies a similar position to Y369 (Supplementary Information, Fig. S3). The mutation V315A affects drastically the functionality of hST3Gal1, decreasing the activity for donor and acceptor by 18- and 70-fold, respectively, when compared to the wild-type enzyme. The mutation most likely impacts the fine tuning of the catalytic base H316 regarding to the substrates, demonstrating that small changes in H316 vicinity have a significant effect on activity. As expected, the exchange of valine by tyrosine had a greater impact on activity. The mutation possibly introduces a steric

hindrance between V315Y and Y266 (a residue interacting with the GalNAc unit), or directly interferes with the optimal binding of Gal- β -1,3-GalNAc- α -O-Bn by reducing the size of the binding cavity. Both possibilities are in agreement with the low k_{cat}/K_M values obtained for this variant (Table 1) and are supported by the lack of substrate saturation resulting in a high K_M for the acceptor.

The regioselectivity of variant V315Y (the formation of an α -2,3 linkage in the product) remained unaltered (Supplementary Information, Fig. S9). As the wild-type enzyme, V315Y was unable to use *N*-acetylglucosamine (GlcNAc) efficiently as acceptor to produce either 3- or 6-sialyl-LacNAc (Supplementary Information, Fig. S9). Although the LacNAc motif is the best acceptor for ST6Gal enzymes, the absence of the axial C4-OH group in substrates containing a GlcNAc- instead of the GalNAc-moiety is responsible for the disruption of important contacts between ST3Gal1 enzymes and these kind of acceptors, resulting in poor activities (500-fold lower k_{cat}/K_M values)¹².

Up to now, a switch in regioselectivity from α -2,3 to α -2,6 has only been achieved for bacterial sialyltransferases, which belong to GT80. Such change occurred by introducing a couple of amino acid substitutions that prompted the binding of the acceptor in a favourable conformation to receive the sialic acid at the galactose C6-OH group in lactose and LacNAc acceptors^{32,33}. In human sialyltransferases substrate binding and regioselectivity seem to be controlled by structural elements that greatly differ in ST6Gal1 and ST3Gal1 enzymes, such as a large N-terminal extension in the catalytic domain of the former that is absent in the latter proteins¹⁷ (Supplementary Information, Fig. S3). Thus, a switch in the regioselectivity of GT29 sialyltransferases is predicted to necessitate a major acceptor binding site remodelling.

Contribution of disulphide bonds to activity and stability. The role of family- and subfamily-conserved disulphide bonds has been thoroughly analysed for ST6Gal1^{14,34,35}. *In vivo* experiments demonstrated that mutations of cysteines 181 and 332 (corresponding to C142 and C281 in hST3Gal1) abrogates the activity of hST6Gal1 and induces protein misfolding, impeding therefore the transport of the enzyme from the endoplasmic reticulum to the Golgi apparatus³⁴. While essential for activity, mutation of the disulphide formed between C353 and C364 does not affect the protein localization. Interruption of the bond C142-C406 preserves both activity and localization³⁵. Studies regarding the implication of ST3Gal1 disulphide bonds in activity, stability and cellular localization *in vivo* are missing.

We analysed the *in vitro* effect of suppressing each one of the three disulphide bonds of hST3Gal1 by exchanging cysteine by serine residues in double variants C59S-C64S, C61S-C139S and C142S-C281S. hST3Gal1 and hST3Gal2 enzymes share an amino acid identity of 49.3% and the same pattern and spatial location of disulphide bonds, based on multiple alignments and modelling of hST3Gal2 with pST3Gal1 as template (not shown). ST3Gal1 and ST3Gal2 have as well overlapping acceptor specificity on O-linked glycans, although ST3Gal2 is primarily responsible for terminal sialylation of gangliosides GD1a and GT1b^{11,36}. ST3Gal3 and ST3Gal4 enzymes also display the same disulphide arrangement, while ST3Gal5 and ST3Gal6 enzymes show a unique disulphide distribution when compared to other members of the ST3Gal subfamily (Fig. 4A). In agreement with previous studies, the removal of cysteine residues C142 and C281 (variant C142S-C281S) abolishes the enzyme's activity (Supplementary Information, Fig. S10). C142 is located in the β -sheet 1 in the sialylmotif L and C281 is found in a loop that precedes the α -helix 11 in the sialylmotif S¹³ (Fig. 4B). These structural elements bring together the scaffold that generates the CMP-Neu5Ac binding site. According to the structural model, disulphide C61-C139 is found in a disordered loop at the hST3Gal1 N-terminal region, spatially close to C59S-C64S. Activity of variant C61-C139 could not be detected neither using the cleared lysate nor purified protein (Supplementary Information, Fig. S10). Cysteine 139 is located in a loop that antecedes the β -sheet 1, thus disulphide C61-C139 probably contributes to the stabilization of this β -sheet in relation to α -helix 11.

Finally, removal of disulphide C59S-C64S has a lesser impact on activity, as the variant retains around 70% of the wild type activity (k_{cat}) for donor and acceptor (Table 1).

We investigated whether the removal of a cysteine pair would affect the formation of the remaining disulphides by submitting the MBP-fusion proteins hST3Gal1 and variants C59S-C64S and C61-C139 to NanoLC-MS/MS analysis (Supplementary Information, Figs. S13–15). Cysteine residues not involved in disulphide bonds were labelled with *N*-ethylmaleimide before sample reduction with TCEP (tris (2-carboxyethyl) phosphine). After reduction, cysteines were alkylated with NEM-D5. The wild-type sialyltransferase and both variants show similar profiles regarding to the presence/absence of free cysteines in differential cysteine labelling experiments, indicating that probably the exclusion of one disulphide bridge does not influence the occurrence of the other two native cysteine pairs.

Finally, the contribution of the intra molecular bond C59-C64 to the protein thermal stability was also examined. The melting temperature (T_m) of the MBP-fused wild-type enzyme and variant C59S-C64S was analysed via differential scanning fluorimetry³⁷. Two thermal denaturation curves were clearly identified for solubility enhancer MBP and the sialyltransferases. The absence of disulphide C59-C64 reduces the T_m of variant C59S-C64S by 7 °C (Fig. 5).

Non fused MBP displays a higher T_m than that observed in fusions with the sialyltransferases because it is stabilized by substrate binding. While MBP-fused sialyltransferases were purified via a hexa-histidine-tag, non-fused MBP was purified over an amylose resin and eluted with maltose, which remains bound to the protein. Binding of maltose to MBP was reported to raise the T_m in the range of 8–15 °C, depending on the buffer conditions³⁸. In this work we observed an increase of 6.3 °C (Fig. 5). To the best of our knowledge, data regarding the thermal stability of heterologously expressed non-glycosylated mammalian sialyltransferases are not available.

Conclusions

Sialic acids play multiple biological roles in mammals. For instance, they are involved in immunology, hereditary diseases and cancer progression^{21,39}. The beta-galactoside alpha-2,3-sialyltransferase 1 (ST3Gal1) is upregulated in cancer and thus it is central to a strategy focused on the synthesis of inhibitors against hypersialylation of proteins and lipids to prevent metastasis and overcome resistance to chemotherapy²¹. Structural and biochemical

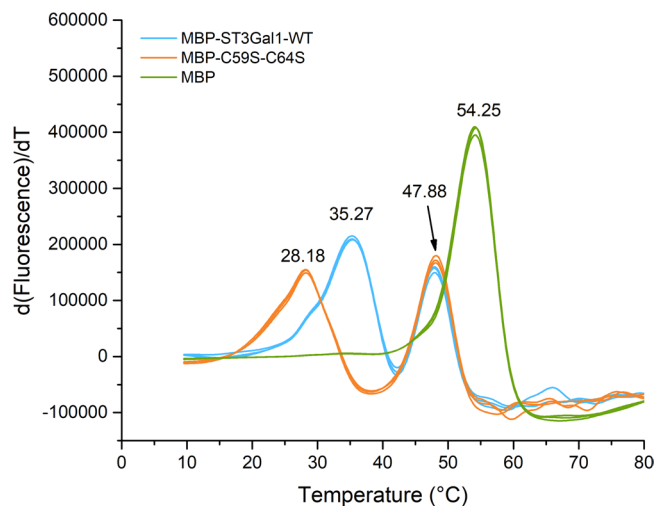


Figure 5. Melting temperature (T_m) of maltose binding protein and MBP-fused hST3Gal1-WT and variant C59S-C64S. T_m values ($^{\circ}\text{C}$) corresponding to each protein are displayed.

studies on this and other sialyltransferases are needed in order to assist the rational design of potent and selective inhibitors.

Here we exchanged various residues of hST3Gal1 that form the binding sites for donor CMP-Neu5Ac and the acceptor motif Gal- β -1,3-GalNAc-R to analyse their influence on catalysis. Residues were selected employing the crystallographic structure of pST3Gal1 as a template for structural modelling of its human homolog hST3Gal1¹³. Binding sites for both substrates are constituted by the catalytic base H316 and amino acids Q105, Y191, Y230, Y266, T269, W297, H299, F310, R311, K312, whose contribution to catalysis was earlier reported for pST3Gal1¹². We explored the effect of mutating residues Y191 and Y230 in hST3Gal1, confirming the importance of the hydrogen bonds provided by their phenolic hydroxyl groups as well as the stacking interactions for an optimal donor and/or substrate binding. The outcome of mutating amino acids N147, S148 and N170, which interact exclusively with the donor was also analysed, observing that activities toward both donor and acceptor are drastically decreased for variants N147S and N170A.

The involvement of subfamily-specific disulphide bonds in the proper folding and activity of sialyltransferases ST6Gal1, ST8Sia1, ST8Sia4 and ST8Sia6 is well documented^{14,34,40–42}; however, information regarding ST3Gal1 enzymes is scarce. Here it was demonstrated that disulphides C61-C139 and C142-C281 (which is entirely conserved in GT29 enzymes) are essential for activity of hST3Gal1. Although disulphide C59-C64 is not strictly needed for activity, its removal has a detrimental impact on the protein thermal stability.

Materials and Methods

Materials. Nucleoside Monophosphate Kinase (NMPK) was obtained from Roche (Germany). CMP-Neu5Ac disodium salt and benzyl 2-acetamido-2-deoxy-3-*O*-(β -D-galactopyranosyl)- α -D-galactopyranoside (Gal- β -1,3-GalNAc- α -*O*-Bn) were acquired from Carbosynth (UK). Lactate dehydrogenase/pyruvate kinase solution (PK/LDH) and asialofetuin were obtained from SIGMA (Germany). NADH and phosphoenolpyruvic acid monopotassium salt (PEP) were purchased from Alfa AppliChem (Germany) and Aesar (Germany), respectively. Electrophoresis and other reagents were from Merck and SIGMA.

Cloning, expression and purification of hST3Gal1 and variants. Synthetic gene encoding hST3Gal1 (Uniprot code Q11201) lacking the transmembrane region (starting in E45) and with optimized codon usage for expression in *E. coli* was obtained from GeneArt (Life technologies, Germany). The gene includes the restriction sites *Nde*I and *Bam*HI at the 3' and 5' ends respectively and was cloned into the vector pMAL-c5X (New England Biolabs, USA). This plasmid was used as a template to generate single and double mutants employing the Quikchange lightning and Quikchange lightning Multi site-directed mutagenesis kits (Stratagene, USA), respectively. Variant C59S-C64S was generated by inverse mutagenesis. Primers used to generate the variants are shown in Table 2 (SIGMA, Germany).

E. coli Origami2 (DE3) (Novagen, Germany) carrying plasmid pMJS10 (kindly provided by Prof. Ruddock, University Oulu, Finland) was transformed with pMAL-c5X carrying the wild type hST3Gal1 or its variants. pMJS10 contains the genes that encode for the enzymes sulphhydryl oxidase and disulphide isomerase (Erv1p and DsbC)²⁹. Sialyltransferases were expressed as N-terminus MBP- and C-terminus His-tagged-proteins as previously reported²⁸. Briefly, 5 mL overnight cultures started from one colony of freshly transformed *E. coli* were used to inoculate 0.5 L LB broth with 12.5 $\mu\text{g mL}^{-1}$ of tetracycline, 200 $\mu\text{g mL}^{-1}$ of ampicillin and 30 $\mu\text{g mL}^{-1}$ of chloramphenicol added. Expression of Erv1p/DsbC was induced at 30 $^{\circ}\text{C}$ at an OD_{600} of 0.3–0.4 by adding L-(+)-arabinose to a final concentration of 0.3%. Temperature was decreased to 17 $^{\circ}\text{C}$ when the cultures reached an OD_{600} of 0.5–0.6 and the expression of sialyltransferases was induced during 22 h with IPTG to a final concentration of 0.1 mM. Cells were harvested by centrifugation and the pellet was resuspended in 8 mL lysis buffer (25 mM NaH_2PO_4 , 25 mM Na_2HPO_4 , 0.3 M NaCl and 20 mM imidazole, pH 8) and frozen at -20°C .

Position in pST3Gal1	Mutation in hST3Gal1 ^a	Primers sequence (5' to 3')
N150	N147S	Forward: GCTGTAGTAGGCTCTAGCGGGAATCTG
		Reverse: CAGATTCCTCCGCTAGAGCCTACTACAGC
S151	S148A	Forward: GTAGTAGGCAATGCAGGGAATCTGCGTG
		Reverse: CACGCAGATTCCTGCATTGCCTACTAC
N173	N170A	Forward: GTCCTGCGTATGGCAAAGCCCCGACTG
		Reverse: CAGTCGGGGCTTTTGCCATACGCAGGAC
Y194	Y191A	Forward: CCATCATCTGGTTGCGCCTGAGAGCTTTC
		Reverse: GAAAGCTCTCAGGCGCAACCAGATGATGG
Y233	Y230A	Forward: CATCTCACACACTGCGATCCCTGTGCCGG
		Reverse: CCGGCACAGGGATCGCAGTGTGTGAGATG
	Y230F	Forward: CATCTCACACACTTTTATCCCTGTGCCGG
		Reverse: CCGGCACAGGATAAAAGTGTGTGAGATG
V318	V315A	Forward: CGTAAAACCGCGCGCACGATGCTG
		Reverse: CAGCATCGTGCGCCGGTTTACG
	V315Y	Forward: CTTTTCGTAAAACCGGCTATCACGATGCTGATTTTC
		Reverse: GAAATCAGCATCGTGATAGCCGGTTTACGAAAAG
C62-C67	C59S-C64S ^b	Forward: CCTTCTACGTGTACTIONACTCTATTGGTCAGCGTAAACTG
		Reverse: ACGGTGTTTAAATCAGACGTTTCAGGTTCTCGCTCAG
C64-C142	C61S-C139S	Forward C61S: CCGTCCTTGTACGTCTACTACTGTATTGGTC
		Forward C139S: CGTTCTGTGGGTTCCCGCCGTTGTGC
C145-C284	C142S-C281S	Forward C142S: GGTGCGCCGTTCTGCTGTAGTAGGC
		Forward C281S: CTCCATGCATGTCTCTGACGAGGTGGATCTG

Table 2. Primers for site directed mutagenesis. ^aNumbers in model are based on the full-length sequence. ^bWhole plasmid amplification was performed by inverse PCR with phosphorylated primers; the forward primer contains the double mutation.

Thawed cells were disrupted by sonication. The lysate was centrifuged at 4 °C and 16,000 g for 30 minutes and the cleared supernatant was filtered, diluted to 4 mg mL⁻¹ total protein and loaded onto a gravity column containing His-Select nickel affinity gel (SIGMA, Germany) previously equilibrated with lysis buffer (25 mM NaH₂PO₄, 25 mM Na₂HPO₄, 0.3 M NaCl and 20 mM imidazole, pH 8). After 1 h incubation at 4 °C, the column was washed with 20 mL of lysis buffer added with 10% glycerol. The protein was eluted in a single step with 2.5 mL elution buffer (25 mM NaH₂PO₄, 25 mM Na₂HPO₄, 0.3 M NaCl and 250 mM imidazole, pH 8). Purified proteins were buffer exchanged against 50 mM HEPES, pH 7.5 or 50 mM Soerensen's buffer and concentrated on Vivaspin-500 ultrafiltration tubes with a MWCO of 50,000 Da (Sartorius, Germany). Expression and purification of sialyltransferases were analysed by SDS-PAGE loading crude lysates, cleared supernatants and purified proteins onto 12% acrylamide gels.

Protein modelling. Two homology models of hST3Gal1 were generated with the SWISS-MODEL server²² using the pST3Gal1 apo- (PDB 2wml) and substrate-bound (2wnb) structures as templates. Alignment of templates and hST3Gal1 models was performed with Chimera (<http://www.rbvi.ucsf.edu/chimera>), reporting a Q-score of around 0.96 for both models. Identical structures would have a Q-score of 1.0. RMSD values for mutagenized residues N147, S148, N170, Y191, Y230 and V315 are shown in Supplementary Information, Fig. S1.

Activity assay. A continuous coupled spectrophotometric assay was employed to determine hST3Gal1 initial rates on donor and acceptor substrates. One unit is defined as the amount of protein that transfers 1 μmol of Neu5Ac from CMP-Neu5Ac to Gal-β-1,3-GalNAc-α-O-Bn or asialofetuin per min at 37 °C, in 50 mM HEPES, pH 7.4. The multi-enzymatic assay is based on NADH oxidation via the coupled reactions of sialyltransferase, NMPK, PK and LDH⁴³. The concentration of the acceptor Gal-β-1,3-GalNAc-α-O-Bn was varied from 0.005 to 1.4 mM, while keeping the donor (CMP-Neu5Ac) concentration constant to 0.7 mM. Additionally, the concentration of the donor was modified from 0.01 to 1.2 mM while maintaining the concentration of Gal-β-1,3-GalNAc-α-O-Bn to 1 mM. Kinetic parameters were obtained at 37 °C using a GENios plate reader (Tecan, Switzerland). 100 μL reactions were performed on a 384-well microtiter plates containing 50 mM HEPES, pH 7.4, 0.7 mM PEP, 0.27 or 0.29 mM NADH, 2 mM ATP, 50 mM KCl, 10 mM MnCl₂, 10 mM MgCl₂, 1 g L⁻¹ BSA, 15 mU of NMPK, 8 U of PK, 12 U of LDH and variable concentrations of donor and acceptor. Microtiter plates were centrifuged at 1000 g during 30 s and reactions were incubated at 37 °C for 22 min to deplete the CMP present in the donor solution and NDPs from the ATP and NMPK solutions¹². Incubation of the reaction components before adding hST3Gal1 is also required to obtain the rate of CMP-Neu5Ac spontaneous hydrolysis. After this period of incubation, 2 μL of sialyltransferase solution was added, the microtiter plate was centrifuged at 1000 g for 30 s and the change in absorbance (340 nm) was measured during 42 minutes at 37 °C. An extinction coefficient for NADH of 6220 M⁻¹ cm⁻¹ and a path length of 0.82 cm were considered for initial rates calculations. At least two replicates were performed for each substrate/acceptor concentration.

K_M and k_{cat} values were obtained in OriginPro (OriginLab) by using nonlinear least squares and the Michaelis-Menten equation from fitting a set of rate versus substrate concentration values. Mean values and standard deviations (\pm SD) were calculated from these measurements.

The impact of mutations on the global catalytic efficiency in terms of the transition state destabilization was calculated using the following equation that incorporates kinetics and thermodynamics parameters^{12,44}:

$$\Delta\Delta G^\ddagger = 2.303 \text{ RT} \log(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{mutant}$$

Synthesis of sialyl-products and analysis by HPAEC-PAD. Samples containing cleared cell lysates with expressed sialyltransferases (50 ng total protein μL^{-1}) or purified enzymes (1 ng μL^{-1}), Gal- β -1,3-GalNAc- α -O-Bn (0.5 or 2.0 mM) or LacNAc (2.0 mM) and CMP-Neu5Ac were prepared in 20.0 mM MOPS, pH 7.5 in 85.0 μL final volumes. For variant C142S-C281S, the double amount of total (cleared extract) or purified protein was used. Reactions were incubated at 37 °C and samples were taken after 1 or 4 h reaction time. Substrate consumption and synthesis of sialosides were analysed by High-Performance Anion Exchange Chromatography (HPAEC) with a Dionex ICS-5000 + SP system utilizing a CarboPac PA10 column (250 \times 2 mm) at a flow of 250 $\mu\text{L} \text{ min}^{-1}$. 100 mM NaOH (A), 100 mM NaOH containing 1 M NaOAc (B) and 250 mM NaOH (C) were used as eluents. Acceptor molecules and sialosides were resolved using a multistep gradient programmed as follows: 0–5 min 100% A, 5–20 min 0–20% B, 20–21 min 20–45% B, 21–28 min 45% B, 28–35 min 100% C, 35–45 min 100% A. Acceptors LacNAc and Gal- β -1,3-GalNAc- α -O-Bn and sialosides 6-sialyl-LacNAc and 3'-sialyl-Gal- β -1,3-GalNAc- α -O-Bn were identified using appropriate standards.

Circular dichroism spectroscopy. CD spectra of the fusion hST3Gal1 wild type and variants were recorded with a JASCO model J-810 spectropolarimeter. The measurements were performed at 20 °C employing a protein concentration of 0.8–1.2 μM in 50 mM Soerensen's buffer, pH 6.5. CD spectra were recorded over the range of 190 to 280 nm with a response of 0.5 s, a data pitch of 0.1 nm and a scanning speed of 100 nm min^{-1} . Signal to noise ratio was reduced by averaging 10 accumulations. The spectrum of Soerensen's buffer was subtracted.

Differential scanning fluorimetry (DSF). 100 μL samples containing a final concentration of 2 μM protein (ST3Gal1-wild-type and variant C59S-C64S fused with the maltose binding protein (MBP), as well as MBP alone), 50 mM HEPES buffer pH 7.4, 100 mM NaCl and 10 \times Sypro Orange (Thermo Fisher, Germany) were prepared in 0.5 mL Eppendorf tubes. Control experiments without protein were also included. Aliquots of 20 μL of each sample were applied 3 times on a 96 well plate that was later sealed with a transparent adhesive film and centrifuged at 1000 g for 1 minute. Samples were maintained in ice during preparation. The well plate was placed in a QuantStudio™ 5 Real-Time PCR System (ThermoFisher Scientific, Germany) and changes in fluorescence over a temperature range of 10–95 °C were recorded using the ROX filter and a temperature ramp of 0.02 °C s^{-1} .

Fluorescence data were processed with the Protein Thermal Shift™ Software to obtain melting temperature values.

The experiments were performed with two different batches of purified proteins to demonstrate the presence of homogenous protein preparations.

Differential cysteine labelling. *N*-ethylmaleimide (NEM) labelling was performed according to the protocol described by McDonagh *et al.*,⁴⁵ with minor modifications. Briefly, 15 μL 200 mM NEM (Sigma) in 100 mM HEPES, pH 7.4, was added to 5 μg 15 μL^{-1} protein in HEPES and stored overnight at 4 °C. SDS was added to a final concentration of 1%. After 2 h at 37 °C protein precipitation was performed according to Wessel and Flügge with chloroform/methanol⁴⁶. The precipitated protein was washed 2 times with 400 μL methanol:chloroform (3:1, v/v). The protein pellet was dissolved in 10 μL 1% SDS in HEPES and heated to 95 °C for 5 min. Reduction was performed with 1.1 μL 500 mM Bond-Breaker TCEP (Thermo Scientific) for 5 min at 95 °C. Reduced cysteine residues at this state were alkylated with 10 μL 200 mM NEM-D5 (Euriso-Top) in HEPES for 2 h at room temperature.

Gel electrophoresis and in-gel digestion. NuPAGE LDS sample buffer (Life Technologies) was added and separation was performed on NuPAGE Novex 4–12% Bis-Tris gels (Life Technologies) with MOPS buffer according to manufacturer's instructions. Gels were washed three times for 5 min with water, stained for 45 min with Simply Blue™ Safe Stain (Life Technologies) and washed with water for 1 h.

Gel bands were excised and destained with 30% acetonitrile in 0.1 M NH_4HCO_3 (pH 8), shrunk with 100% acetonitrile, and dried in a vacuum concentrator (Concentrator 5301, Eppendorf, Germany). Digests were performed with 0.1 μg trypsin or elastase per gel band overnight at 37 °C in 0.1 M NH_4HCO_3 (pH 8). After removing the supernatant, peptides were extracted from the gel slices with 5% formic acid, and extracted peptides were pooled with the supernatant.

NanoLC-MS/MS analysis and MS data analysis. NanoLC-MS/MS analyses were performed on a LTQ-Orbitrap Velos Pro (Thermo Scientific) equipped with a PicoView Ion Source (New Objective) and coupled to an EASY-nLC 1000 (Thermo Scientific). Peptides were loaded on capillary columns (PicoFrit, 30 cm \times 150 μm ID, New Objective) self-packed with ReproSil-Pur 120 C18-AQ, 1.9 μm (Dr. Maisch) and separated with a 30-minute linear gradient from 3% to 30% acetonitrile and 0.1% formic acid and a flow rate of 500 nL/min. MS scans were acquired in the Orbitrap analyser with a resolution of 30,000 at m/z 400, MS/MS scans were acquired in the Orbitrap analyser with a resolution of 7,500 at m/z 400 using HCD fragmentation with 30% normalized collision energy. A TOP5 data-dependent MS/MS method was used; dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 30 seconds; singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with AGC target

a value of 1e6 for MS scans and 5e4 for MS/MS scans. Lock mass option was applied for internal calibration in all runs using background ions from protonated decamethylcyclotrasiloxane (m/z 371.10124).

Database search was performed against a small in-house database containing the protein sequences of interest with PEAKS 10 software (Bioinformatics solution Inc.) with the following parameters: peptide mass tolerance: 10 ppm, MS/MS mass tolerance: 0.015 Da, enzyme: “none”; variable modifications: Acetylation (Protein N-term), Pyro-glu from Q, *N*-ethylmaleimide on cysteines, D5 *N*-ethylmaleimide on cysteines. Results were filtered to 1% PSM-FDR by target-decoy approach.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 7 March 2019; Accepted: 11 November 2019;

Published online: 29 November 2019

References

- Lengeler, K. B., Tielker, D. & Ernst, J. F. Protein-O-mannosyltransferases in virulence and development. *Cell Mol Life Sci* **65**, 528–544 (2008).
- Li, Y. H. & Chen, X. Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl Microbiol Biot* **94**, 887–905, <https://doi.org/10.1007/s00253-012-4040-1> (2012).
- Varki, A. Sialic acids in human health and disease. *Trends Mol Med* **14**, 351–360, <https://doi.org/10.1016/j.molmed.2008.06.002> (2008).
- Zhao, Y. Y. *et al.* Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci* **99**, 1304–1310 (2008).
- Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255–266, <https://doi.org/10.1038/Nri2056> (2007).
- Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nat Rev Immunol* **8**, 874–887, <https://doi.org/10.1038/Nri2417> (2008).
- Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henriksas, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495, <https://doi.org/10.1093/Nar/Gkt1178> (2014).
- Szabo, R. & Skropeta, D. Advancement of sialyltransferase inhibitors: Therapeutic challenges and opportunities. *Med Res Rev* **37**, 219–270, <https://doi.org/10.1002/med.21407> (2017).
- Paulson, J. C. & Colley, K. J. Glycosyltransferases. *Structure, localization, and control of cell type-specific glycosylation*. *J Biol Chem* **264**, 17615–17618 (1989).
- Tsuiji, S., Datta, A. K. & Paulson, J. C. Systematic nomenclature for sialyltransferases. *Glycobiology* **6**, R5–R7 (1996).
- Audry, M. *et al.* Current trends in the structure-activity relationships of sialyltransferases. *Glycobiology* **21**, 716–726, <https://doi.org/10.1093/glycob/cwq189> (2011).
- Rakic, B. *et al.* Structure-based mutagenic analysis of mechanism and substrate specificity in mammalian glycosyltransferases: Porcine ST3Gal-I. *Glycobiology* **23**, 536–545, <https://doi.org/10.1093/glycob/cwt001> (2013).
- Rao, F. V. *et al.* Structural insight into mammalian sialyltransferases. *Nat Struct Mol Biol* **16**, 1186–1188, http://www.nature.com/nsmb/journal/v16/n11/supinfo/nsmb.1685_S1.html (2009).
- Hirano, Y. *et al.* Disulphide linkage in mouse ST6Gal-I: determination of linkage positions and mutant analysis. *J Biochem* **151**, 197–203, <https://doi.org/10.1093/jb/Mvr133> (2012).
- Meng, L. *et al.* Enzymatic basis for N-Glycan sialylation: structure of rat α 2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem* **288**, 34680–34698, <https://doi.org/10.1074/jbc.M113.519041> (2013).
- Datta, A. K., Sinha, A. & Paulson, J. C. Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates. *J Biol Chem* **273**, 9608–9614, <https://doi.org/10.1074/jbc.273.16.9608> (1998).
- Kuhn, B. *et al.* The structure of human $[\alpha]$ -2,6-sialyltransferase reveals the binding mode of complex glycans. *Acta Cryst. D* **69**, 1826–1838, <https://doi.org/10.1107/S0907444913015412> (2013).
- Volkers, G. *et al.* Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nat Struct Mol Biol* **22**, 627–635, <https://doi.org/10.1038/nsmb.3060> (2015).
- Moremen, K. W. *et al.* Expression system for structural and functional studies of human glycosylation enzymes. *Nat Chem Biol* **14**, 156–162, <https://doi.org/10.1038/Nchembio.2539> (2018).
- Rodrigues, E. & Macauley, M. S. Hypersialylation in Cancer: Modulation of Inflammation and Therapeutic Opportunities. *Cancers* **10**, <https://doi.org/10.3390/cancers10060207> (2018).
- Wu, X. *et al.* Sialyltransferase ST3GAL1 promotes cell migration, invasion, and TGF- β 1-induced EMT and confers paclitaxel resistance in ovarian cancer. *Cell Death Dis* **9**, 1102, <https://doi.org/10.1038/s41419-018-1101-0> (2018).
- Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303, <https://doi.org/10.1093/nar/gky427> (2018).
- Jeanneau, C. *et al.* Structure-function analysis of the human sialyltransferase ST3Gal I: Role of N-glycosylation and a novel conserved sialylmotif. *J Biol Chem* **279**, 13461–13468, <https://doi.org/10.1074/jbc.M311764200> (2004).
- Datta, A. K. & Paulson, J. C. Sialylmotifs of sialyltransferases. *Indian J Biochem Bio* **34**, 157–165 (1997).
- Harduin-Lepers, A. Comprehensive analysis of sialyltransferases in vertebrate genomes. *Glycobiol Insights* **2**, 29–61, <https://doi.org/10.4137/GBI.S3123> (2010).
- Kuipers, R. K. *et al.* 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **78**, 2101–2113, <https://doi.org/10.1002/Prot.22725> (2010).
- Kourist, R. *et al.* The α/β -hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *ChemBioChem* **11**, 1635–1643, <https://doi.org/10.1002/cbic.201000213> (2010).
- Ortiz-Soto, M. E. & Seibel, J. Expression of functional human sialyltransferases ST3Gal1 and ST6Gal1 in *Escherichia coli*. *Plos One* **11**, <https://doi.org/10.1371/journal.pone.0155410> (2016).
- Nguyen, V. D. *et al.* Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E-coli. *Microb Cell Fact* **10**, <https://doi.org/10.1186/1475-2859-10-1> (2011).
- Harduin-Lepers, A., Mollicone, R., Delannoy, P. & Oriol, R. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology* **15**, 805–817, <https://doi.org/10.1093/glycob/cwi063> (2005).
- Lee-Sundlov, M. M. *et al.* Circulating blood and platelets supply glycosyltransferases that enable extrinsic extracellular glycosylation. *Glycobiology* **27**, 188–198, <https://doi.org/10.1093/glycob/cww108> (2017).
- McArthur, J. B., Yu, H., Zeng, J. & Chen, X. Converting Pasteurella multocida α 2-3-sialyltransferase 1 (PmST1) to a regioselective α 2-6-sialyltransferase by saturation mutagenesis and regioselective screening. *Org Biomol Chem* **15**, 1700–1709, <https://doi.org/10.1039/c6ob02702d> (2017).

33. Schmolzer, K. *et al.* Complete switch from alpha-2,3- to alpha-2,6-regioselectivity in *Pasteurella dagmatis* beta-D-galactoside sialyltransferase by active-site redesign. *Chem Commun* **51**, 3083–3086, <https://doi.org/10.1039/c4cc09772f> (2015).
34. Datta, A. K., Chammas, R. & Paulson, J. C. Conserved cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond. *J Biol Chem* **276**, 15200–15207, <https://doi.org/10.1074/jbc.M010542200> (2001).
35. Qian, R., Chen, C. & Colley, K. J. Location and mechanism of alpha 2,6-sialyltransferase dimer formation - Role of cysteine residues in enzyme dimerization, localization, activity and processing. *J Biol Chem* **276**, 28641–28649, <https://doi.org/10.1074/jbc.M103664200> (2001).
36. Sturgill, E. R. *et al.* Biosynthesis of the major brain gangliosides GD1a and GT1b. *Glycobiology* **22**, 1289–1301, <https://doi.org/10.1093/glycob/cws103> (2012).
37. Semisotnov, G. V. *et al.* Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119–128, <https://doi.org/10.1002/bip.360310111> (1991).
38. Novokhatny, V. & Ingham, K. Thermodynamics of maltose binding protein unfolding. *Protein Sci* **6**, 141–146 (1997).
39. Schauer, R. & Kamerling, J. P. Exploration of the sialic acid world. *Adv Carbohydr Chem Biochem* **75**, 1–213, <https://doi.org/10.1016/bs.accb.2018.09.001> (2018).
40. Angata, K., Yen, T. Y., El-Battari, A., Macher, B. A. & Fukuda, M. Unique disulfide bond structures found in ST8Sia IV polysialyltransferase are required for its activity. *J Biol Chem* **276**, 15369–15377, <https://doi.org/10.1074/jbc.M100576200> (2001).
41. Datta, A. K. & Paulson, J. C. Conserved cysteines of sialylmotifs of sialyltransferases are essential for enzyme activity. *Glycobiology* **6**, 1101–1101 (1996).
42. Takashima, S., Matsumoto, T., Tsujimoto, M. & Tsuji, S. Effects of amino acid substitutions in the sialylmotifs on molecular expression and enzymatic activities of alpha 2,8-sialyltransferases ST8Sia-I and ST8Sia-VI. *Glycobiology* **23**, 603–612, <https://doi.org/10.1093/glycob/cwt002> (2013).
43. Gosselin, S., Alhussaini, M., Streiff, M. B., Takabayashi, K. & Palcic, M. M. A Continuous spectrophotometric assay for glycosyltransferases. *Anal Biochem* **220**, 92–97, <https://doi.org/10.1006/abio.1994.1303> (1994).
44. Fersht, A. R. *et al.* Hydrogen-bonding and biological specificity analyzed by protein engineering. *Nature* **314**, 235–238, <https://doi.org/10.1038/314235a0> (1985).
45. McDonagh, B., Sakellariou, G. K., Smith, N. T., Brownridge, P. & Jackson, M. J. Differential cysteine labeling and global label-free proteomics reveals an altered metabolic state in skeletal muscle aging. *J Proteome Res* **13**, 5008–5021, <https://doi.org/10.1021/pr5006394> (2014).
46. Wessel, D. & Flugge, U. I. A method for the quantitative recovery of protein in dilute-solution in the presence of detergents and lipids. *Anal Biochem* **138**, 141–143, [https://doi.org/10.1016/0003-2697\(84\)90782-6](https://doi.org/10.1016/0003-2697(84)90782-6) (1984).
47. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res* **14**, 1188–1190, <https://doi.org/10.1101/gr.849004> (2004).

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft TRR SFB 225 Biofabrikation (Nr. 397988206, subproject B05) is acknowledged. We are grateful to Stephanie Lamer for Mass spectrometric analysis and Professors Caroline Kisker and Hermann Schindelin for granting us access to the JASCO model J-810 spectropolarimeter.

Author contributions

J.S. and M.E.O.S. conceived the project and designed experiments. M.E.O.S. and S.R. performed experiments. A.S. conducted nanoLC-MS/MS experiments and analyses. J.S. and M.E.O.S. analysed the data. M.E.O.S. wrote the paper. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-54384-8>.

Correspondence and requests for materials should be addressed to J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019