



HAL
open science

Engineering the ligand specificity of the human galectin-1 by incorporation of tryptophan analogs

Felix Tobola, Martin Lepšík, Syeda Rehana Zia, Hakon Leffler, Ulf J Nilsson, Ola Blixt, Anne Imberty, Birgit Wiltschi

► **To cite this version:**

Felix Tobola, Martin Lepšík, Syeda Rehana Zia, Hakon Leffler, Ulf J Nilsson, et al.. Engineering the ligand specificity of the human galectin-1 by incorporation of tryptophan analogs. *ChemBioChem*, 2022, 23 (5), pp.e202100593. 10.1002/cbic.202100593 . hal-03517502

HAL Id: hal-03517502

<https://hal.science/hal-03517502>

Submitted on 7 Jan 2022

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

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

Engineering the ligand specificity of the human galectin-1 by incorporation of tryptophan analogs

Felix Tobola^{a,b}, Martin Lepšik^{c#}, Syeda Rehana Zia^d, Hakon Leffler^e, Ulf J. Nilsson^f, Ola Blixt^g, Anne Imberty^c and Birgit Wiltschi^{a*}

^a Austrian Centre of Industrial Biotechnology Petersgasse 14, A-8010 Graz, Austria

^b Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

^c Université Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

^d Department of Chemistry, University of Karachi, Karachi, Pakistan

^e Department of Laboratory Medicine, Section MIG, Lund University BMC-C1228b; Klinikgatan 28, 221 84 Lund, Sweden

^f Centre for Analysis and Synthesis, Department of Chemistry Lund University, POB 124 Box 124, 221 00 Lund, Sweden

^g Department of Biotechnology and Biomedicine Technical University of Denmark Søtofts Plads, 2800 Kgs. Lyngby, Denmark

[#] Present Address: Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nam. 2, 166 10, Prague 6, Czechia

*email: E-mail: birgit.wiltschi@acib.at

Keywords: Lectin, carbohydrates, glycoconjugate, glycomimetics, LecA

Abstract

Galectin-1 is a β -galactoside-binding lectin with manifold biological functions. A single tryptophan residue (W68) in its carbohydrate binding site plays a major role in ligand binding and is highly conserved among galectins. To fine tune galectin-1 specificity, we introduced several non-canonical tryptophan analogs at this position of human galectin-1 and analyzed the resulting variants using glycan microarrays. Two variants containing 7-azatryptophan and 7-fluorotryptophan showed a reduced affinity for 3'-sulfated oligosaccharides. Their interaction with different ligands was further analyzed by fluorescence polarization competition assay. Using molecular modeling we provide structural clues that the change in affinities comes from modulated interactions and solvation patterns. Thus, we show that the introduction of subtle atomic mutations in the ligand binding site of galectin-1 is an attractive approach for fine-tuning its interactions with different ligands.

Introduction

Lectins are proteins that specifically bind to complex carbohydrates without enzymatically modifying them. They are ubiquitously found in all domains of life and find applications as biological tools, such as the purification of glycoproteins, detection of tumor markers, as biosensors or in high-throughput microarrays and blotting applications.^[1-3] Classic lectin research relies on identification of new specificity from natural sources, but with the advent of modern protein engineering, the new field of 'synthetic glycobiology' emerged.^[4,5]

Researchers started to engineer artificial lectins not found in nature by modifying binding sites, architectures and topologies of the carbohydrate binding domain, or assembling them with other modules of interest. While considerable progress has been made, the engineering of lectins by a synthetic biology approach is still in its infancy.^[6-8] In particular, the application of non-canonical amino acids (ncAAs) for lectin engineering is a promising strategy.

While Nature uses almost exclusively 20 canonical amino acids prescribed by the genetic code for the biosynthesis of proteins, there are hundreds of ncAAs known.^[9] By their admission to ribosomal translation, new chemistries can be introduced into target-proteins to engineer them in a diverse and/or subtle fashion that is complementary to classical mutagenesis approaches.^[10,11] However, the effects provoked by the introduction of new chemical groups into proteins are still difficult to predict. Therefore, exhaustive research on and in-depth analysis of synthetic proteins is indispensable to expand our ability to engineer proteins with chemistries beyond those prescribed in the genetic code. Attempts to utilize non-canonical amino acids in lectins are still to be developed, although preliminary studies are promising.^[12,13]

Galectins are a family of structurally related carbohydrate-binding proteins ubiquitously found in mammalian cells and tissues,^[14] which share an affinity for β -galactosides.^[15] Galectin-1 (Gal-1) was the first characterized galectin,^[16] and it displays preference for the terminal *N*-acetylglucosamine (LacNAc) motif.^[17,18] Gal-1 is expressed in many vertebrate and invertebrate organisms, occurs free in the cytoplasm and extracellularly^[19,20] and is involved

in many biological functions, including cell adhesion, and regulation of adaptive immunity.^[21,22] Gal-1 contains several cysteine residues, which confer sensitivity to oxidation. Cysteine oxidation inactivates Gal-1 by changes in conformation and multimerization.^[23,24] The more stable mutant CSGal-1, in which all six cysteines are replaced by serines, is often used for analytical studies.^[25]

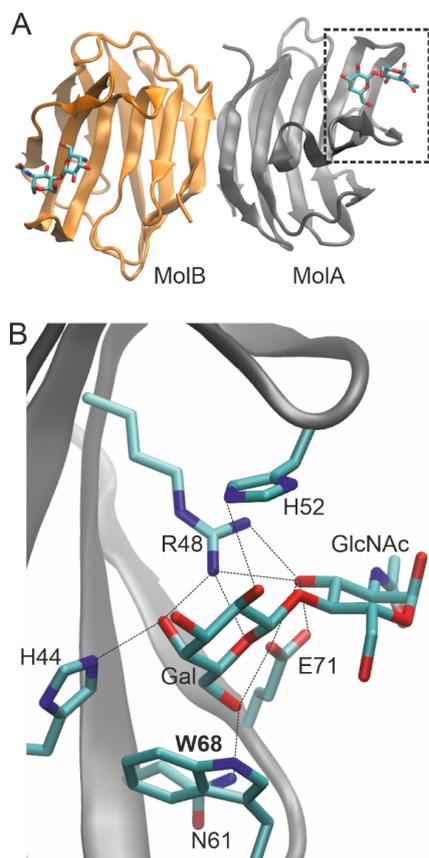


Figure 1: A) Crystal structure of the C2S mutant of the human galectin-1 complexed with LacNAc (PDB: 1W6P^[26]). The two monomers are shown as grey and orange cartoons and the bound LacNAc ligands as sticks. The binding site in one monomer is outlined by the dashed box. B) Close up of the ligand binding site in one monomer. Residues involved in ligand-binding are shown as sticks. Putative hydrogen bonds are shown as dashed lines. In this study, the sole tryptophan residue W68 (bold) was exchanged against non-canonical analogs.

Several X-ray and NMR structures are available of Gal-1 and CSGal-1,^[25-29] demonstrating that lactose/LacNAc binding occurs *via* several direct hydrogen bonds established to amino acid residues in the carbohydrate binding site (H44, R48, H52, N61, E71) (Figure 1). Water networks are also important players in the carbohydrate recognition.^[30] Furthermore, as classically observed in protein-carbohydrate interaction,^[31,32] the ligand is stabilized by CH- π interactions between the Gal ring and the aromatic ring of tryptophan W68. Several lines of experimental evidence (NMR, UV resonance Raman spectroscopy) and molecular dynamics simulations agree on the crucial role of W68 in stabilizing the Gal moiety in complex with Gal-1.^[29,33,34] The importance of this aromatic residue for positioning and stabilizing the ligand in the binding site is also evident from the fact that it is conserved in all known galectins.^[34] Indeed, the capability of Gal-1 to bind lactose is reduced when the single tryptophan is mutated to amino acids with a smaller aromatic ring (tyrosine, phenylalanine) or aliphatic side chain (leucine)^[24,34,35] due to decrease of CH- π interaction strength.

We demonstrated previously that changes in single atoms in tryptophan residues in carbohydrate binding sites can alter the specificity in more subtle manner than classical

mutagenesis, modifying the lectin-glycan interaction without abolishing it.^[12] Since the Gal-1 sequence contains only a single tryptophan residue in the ligand binding site, it appears to be an optimal target for applying a synthetic biology approach to introduce subtle atomic mutations at the tryptophan residue to fine-tune the Gal-1-ligand-interactions.

Here, we introduced a set of non-canonical tryptophan analogs into CSGal-1 and analyzed the effects on ligand specificity by a glycan microarray, then further quantified observed changes by a fluorescence polarization assay and provided a structural and dynamical view using molecular simulations. We found that introducing atomic changes in the important tryptophan residue is an appealing strategy to reduce the affinity for specific sulfated ligands without substantially altering the interaction with other carbohydrates.

Results and Discussion

To introduce the desired atomic mutations, we exchanged the single tryptophan residue W68 (corresponds to W77 in the CSGal-1 amino acid sequence used in this study, which contained an N-terminal hexahistidine-tag; see Table S5) in the binding site of CSGal-1 with twelve non-canonical analogs (Figure S1A). We applied a tryptophan-auxotrophic *E. coli* strain for the supplementation-based incorporation (SPI) of the analogs.^[36] Since the CSGal-1 sequence contains only a single W68, the exchange was site-specific. By this strategy, we avoided the drawbacks in product yields often associated with the incorporation of ncAAs at in-frame amber codons with orthogonal translation systems.^[37] However, as discussed further down, possibly incomplete labelling with ncAAs is a drawback of this method. We observed CSGal-1 expression in the presence of all analogs, albeit at varying levels. The expression was most efficient with canonical tryptophan and of comparable levels with the fluorinated analogs. The hydroxyl-, methyl- and amine-derivatives yielded considerably lower expression levels. CSGal-1 was efficiently expressed in the presence of 7-azatryptophan (7AzaW) but not the other aza-analogs. The incorporation of all tryptophan analogs except of 4-fluorotryptophan (4FW) resulted in soluble variant proteins (Figure S1A), that were all purified by immobilized metal affinity chromatography (Figure S1B). Mass analysis of the purified proteins revealed that 4-, 5- and 6-azatryptophan as well as 1-methyltryptophan were not incorporated into the protein, it still contained tryptophan. In contrast, 7AzaW was successfully incorporated into the CSGal-1. However, the azatryptophan-containing variant only had a small mass difference in comparison to the parent protein (0.865 Da), which was at the sensitivity limit of the LC-ESI-MS. A potential contamination of the 7AzaW variant with parent protein cannot be excluded entirely. The fluorotryptophans were efficiently incorporated with only little contamination with parent protein (17%, 7% and 8% contamination with WT for 5-, 6- and 7-fluorotryptophan, respectively), which matches as reported with another lectin.^[12] The hydroxytryptophans were less well incorporated and showed a contamination with parent protein between 21% and 42% (Table S1).

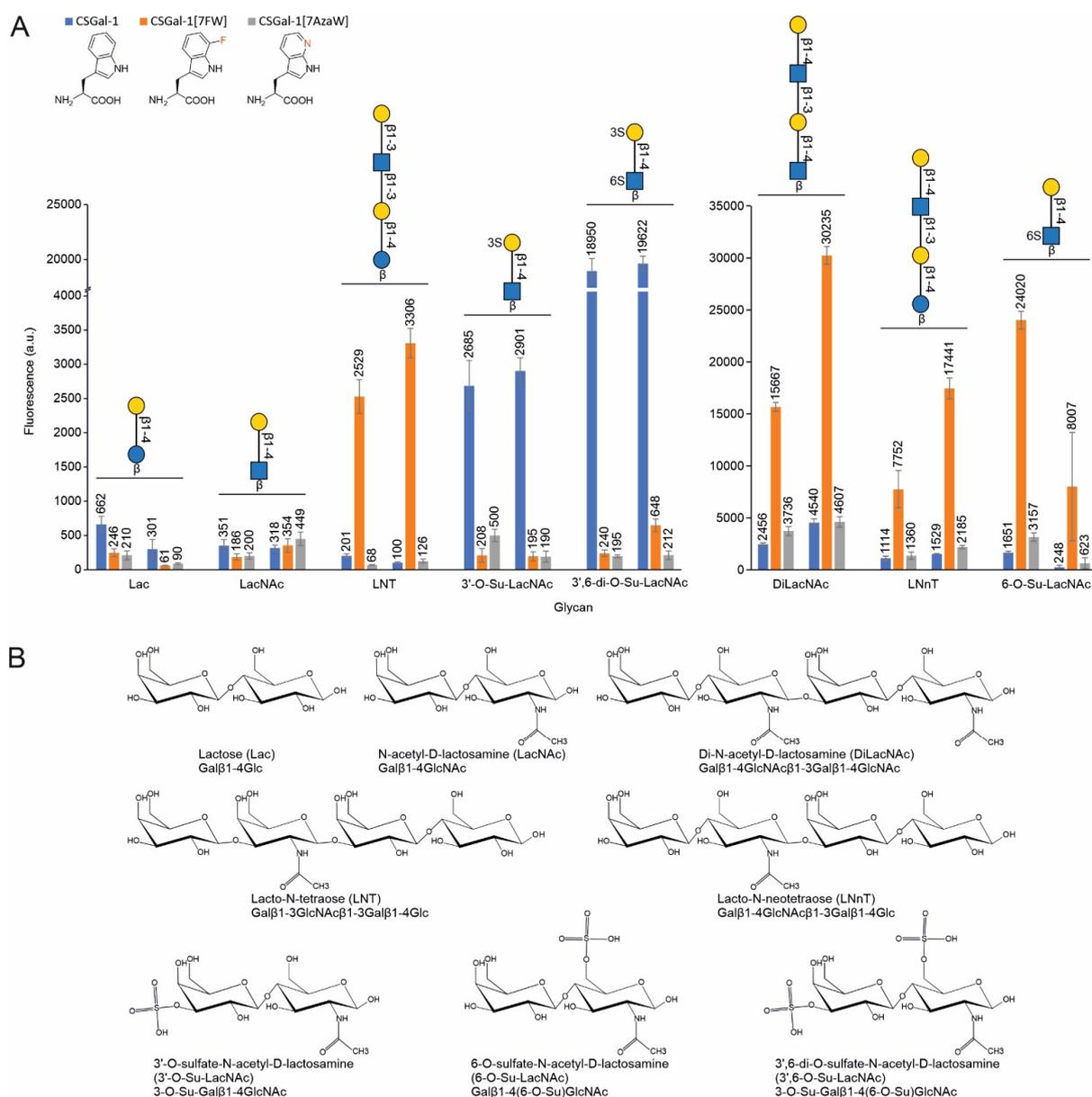


Figure 2: Glycan microarray analysis and glycan structures. A) The fluorescence values of quantified microarray data are shown for eight different ligands (two replicates each). Note the differently scaled y-axes for the results with high (left panel) and low affinity ligands (right panel). Exact values can be found in Table S2 and in the raw data in the supplementary excel-file. Tryptophan analogs introduced in CSGal-1, as well as symbolic glycan representations, are shown as insets. B) The structures, names and abbreviations of the carbohydrates used in this study are shown..

To assess possible modifications in ligand specificity, the synthetic variants containing 4- and 5-hydroxytryptophan, 4-aminotryptophan, 5-, 6- and 7-fluorotryptophan and 7-azatryptophan as well as the parent protein were screened against a glycan microarray containing more than 300 carbohydrates.^[38]

All variants actively bound to the glycan microarray. As expected, the overall binding pattern was similar to that of the parent protein. However, selected variants bound the glycans with considerably changed specificity compared to the CSGal-1 parent protein. CSGal-1[5FW]

showed reduced binding to all glycans (see raw data in Figure S2 and Figure S3). The differences were particularly pronounced with the Gal α 1-3Gal-terminated ligands Galili tri, Galili tetra and with LacNAc-terminated glycans (Figure S3). Only Neu5Ac-terminated ligands (comp3 and comp7 in Figure S3) were bound significantly compared to the others. CSGal-1[6FW] showed a better binding to the glycan microarray chip than the 5FW variant (Figure S2). No drastic changes in specificity were observed compared to the parent protein (Figure S4) except a reduced affinity towards most glycans. However, some ligands were bound equally well or even slightly better, without drastic differences to CSGal-1. CSGal-1[7FW] showed a marked affinity gain or -loss for specific glycans compared to the parent protein (Figure S5). The affinity for many LacNAc-terminated ligands was improved. However, when these ligands carried a sulfate-group at the 3'-O of galactose, the affinity was considerably reduced (Figure 2). Furthermore, we observed an increased affinity for some Gal β 1-3GlcNAc-terminated ligands, such as H type I, and Gal α 1-3Gal-terminated ligands (Galili tri and tetra), and a reduced binding to Neu5Ac-terminated glycans (comp 3, -7 and -11). Similar to CSGal-1[7FW], CSGal-1[7AzaW] showed a considerably reduced binding to 3'-O-sulfated LacNAc as compared to the parent protein (Figure 2 and Figure S6). At the same time, the binding to LacNAc was unchanged. Some Gal α 1-3Gal-terminated ligands were bound slightly better, but the difference to the parent protein was not as pronounced as for the 7FW-containing variant. Apart from some Neu5Ac-terminated high-affinity ligands, where a drastically reduced binding was observed, the binding to other carbohydrates on the array was comparable with that of the parent protein. CSGal-1[4NH2W] showed many small differences to the parent protein, while overall binding to glycans on the array was comparable to that of the parent protein (Figure S7). In general, some low-affinity ligands were bound better, while some high-affinity ligands were bound less well. Again, a markedly reduced binding to 3'-O-sulfated LacNAc was observed, while the binding to LacNAc was unaltered.

The most striking differences in ligand specificity were the increased interaction of the 7FW-containing variant for terminating LacNAc and the drastically reduced interaction of the 7FW- and 7AzaW-containing variants with the 3'-O-sulfated LacNAc-terminated carbohydrates, namely 3'-O-Su-LacNAc and 3',6-di-O-Su-LacNAc (Figure 2).

To verify and quantify this finding, we examined the affinity of CSGal-1, CSGal-1[7FW] and CSGal-1[7AzaW] towards these and other ligands shown in Figure 2B by a fluorescence polarization assay. We first confirmed that the binding of the fluorescein-labelled high affinity probe^[39] was not drastically affected by the atomic mutation on the modified galectin, and the optimal lectin/probe concentrations were determined (Figure S8). In the competition assay, the displacement of the fluorescein-labelled probe from the ligand binding site of the galectins by the analyzed glycans (Figure 2B) was evaluated. From several measurements with different concentrations of competing ligand (Figure S9), the K_d of the ligands for the galectins was determined (Figure 3). The competition assays with the eight ligands validated the results generated by the glycan microarray.

CSGal-1 showed an increased affinity towards the 3'-O-Su-LacNAc in comparison to the non-sulfated compound. The decrease in K_d value from $34.9 \pm 6.7 \mu\text{M}$ to $10.5 \pm 2.8 \mu\text{M}$ is in accordance with a previous report.^[40]

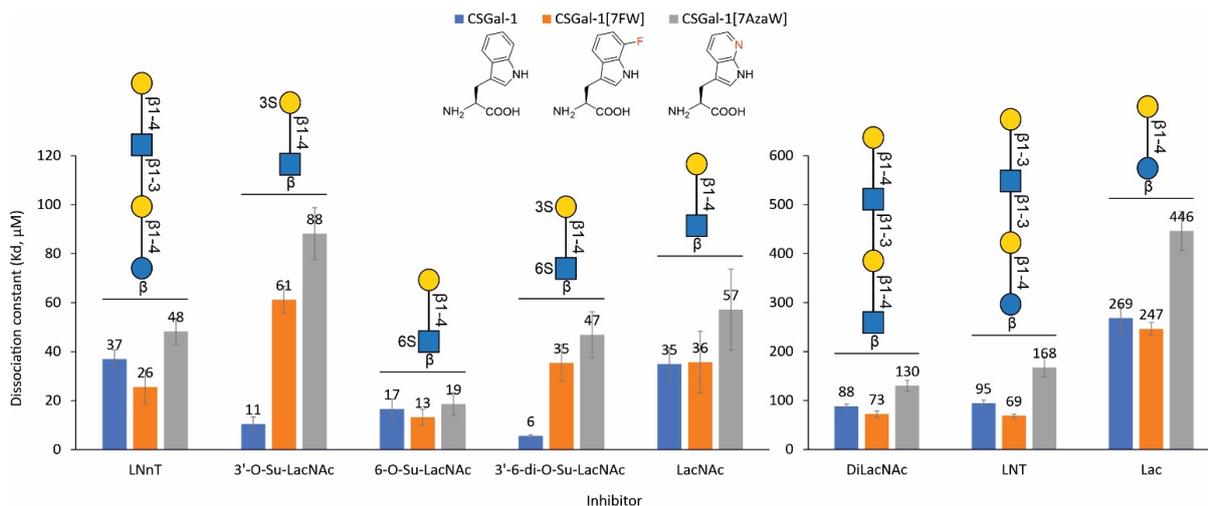


Figure 3: Affinity of CSGal-1 and variants towards different glycans, as determined by a fluorescence polarization competition assay. K_d average (shown above bars) and SD were calculated from 4 to 6 single point measurements showing between 15-85% inhibition (for the inhibition curves and calculated K_d values see Figure S8 and Table S3, respectively). Note the differently scaled y-axes for the results with high (left panel) and low affinity ligands (right panel). Tryptophan analogs introduced in CSGal-1, as well as symbolic glycan representations, are shown as insets.

CSGal-1[7FW] showed a comparable binding behavior as the parent lectin in terms of ligand preferences, except for 3'-O-Su-LacNAc and 3',6-di-O-Su-LacNAc. For the latter two ligands, the affinities were about 5.5-fold lower than for CSGal-1 (K_d values of $61.2 \pm 5.5 \mu\text{M}$ and $35.4 \pm 7.4 \mu\text{M}$ versus $10.5 \pm 2.8 \mu\text{M}$ and $5.6 \pm 0.5 \mu\text{M}$, respectively). At the same time the interaction with the other tested ligands remained unaltered or even slightly improved. The substitution of tryptophan with 7-azatryptophan in general resulted in decreased affinities for all tested ligands. In comparison to CSGal-1[7FW], the differences in affinity in relation to the parent protein were more complex with the 7AzaW-variant. While the K_d values for 6-O-Su-LacNAc and lacto-N-neotetraose (LNnT, Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc) were similar to those of CSGal-1, the affinity for the other ligands was reduced by a factor of at least 1.5. For 3'-O-Su-LacNAc and 3',6-di-O-Su-LacNAc, the K_d values were even 8-times higher than for CSGal-1 (K_d values of $88.2 \pm 10.6 \mu\text{M}$ and $46.9 \pm 9.4 \mu\text{M}$ versus $10.5 \pm 2.8 \mu\text{M}$ and $5.6 \pm 0.5 \mu\text{M}$, respectively).

To rationalize the observed binding preferences of the CSGal-1 variants for the ligands (Figure 3.), molecular modelling and simulations were conducted. For simplicity, we focused on the 3'-O-Su-LacNAc disaccharide, since it presents the largest affinity differences for the three CSGal-1 variants, and LacNAc for comparison. Coupled with three CSGal-1 variants, this summed up to six systems studied. Specific parameterization consistent with the used protein force field was necessary to describe the non-canonical amino acids 7FW and 7AzaW. The method description and all the parameters are given in the Supporting Information.

To validate the protocol, we ran 1 μs molecular dynamics (MD) simulations of dimeric Gal-1 in complex with LacNAc and compared it to the original X-ray structure. The overall protein structure as well as the details of the protein-ligand hydrogen bonding were very well maintained (Table S6.). Similarly, the CH- π stacking interaction between the Gal moiety and W68 was maintained, although the distances defining it were slightly increased with respect to the X-ray structure (Table S7). A comparable behavior had been previously reported and was

ascribed to the imperfections of the force field.^[33] Water networks bridging the Gal moiety to Gal-1 residues W68, K63, H52, S29 and N33 of the extended carbohydrate-binding site and even to the mobile A1 and D123 residues were observed (Figure 4A).

During the 1 μ s MD simulation of the complex between Gal-1 and 3'-O-Su-LacNAc, the sulfate group was wedged between residues W68, H44 and V31 but no direct H-bonding was observed. The W68 stacking interactions with 3'-O-sulfate-Gal were similar to those of LacNAc, although the X-ray structures showed shorter distances for the wild-type Gal-1/LacNAc complex (Table S7, Table S8). Water-mediated H-bonds between 3'-O-sulfate-Gal and Gal-1 residues were similar as in the Gal-1/LacNAc complex (Figure 4B) but presumably stronger due to negative charge and polarization of the sulfate, the phenomenon which is not described in classical fixed-charge force fields ^[41]. This effect could cause the three-fold stronger binding of the 3'-O-Su-LacNAc to Gal-1 as compared to LacNAc (cf. Figure)

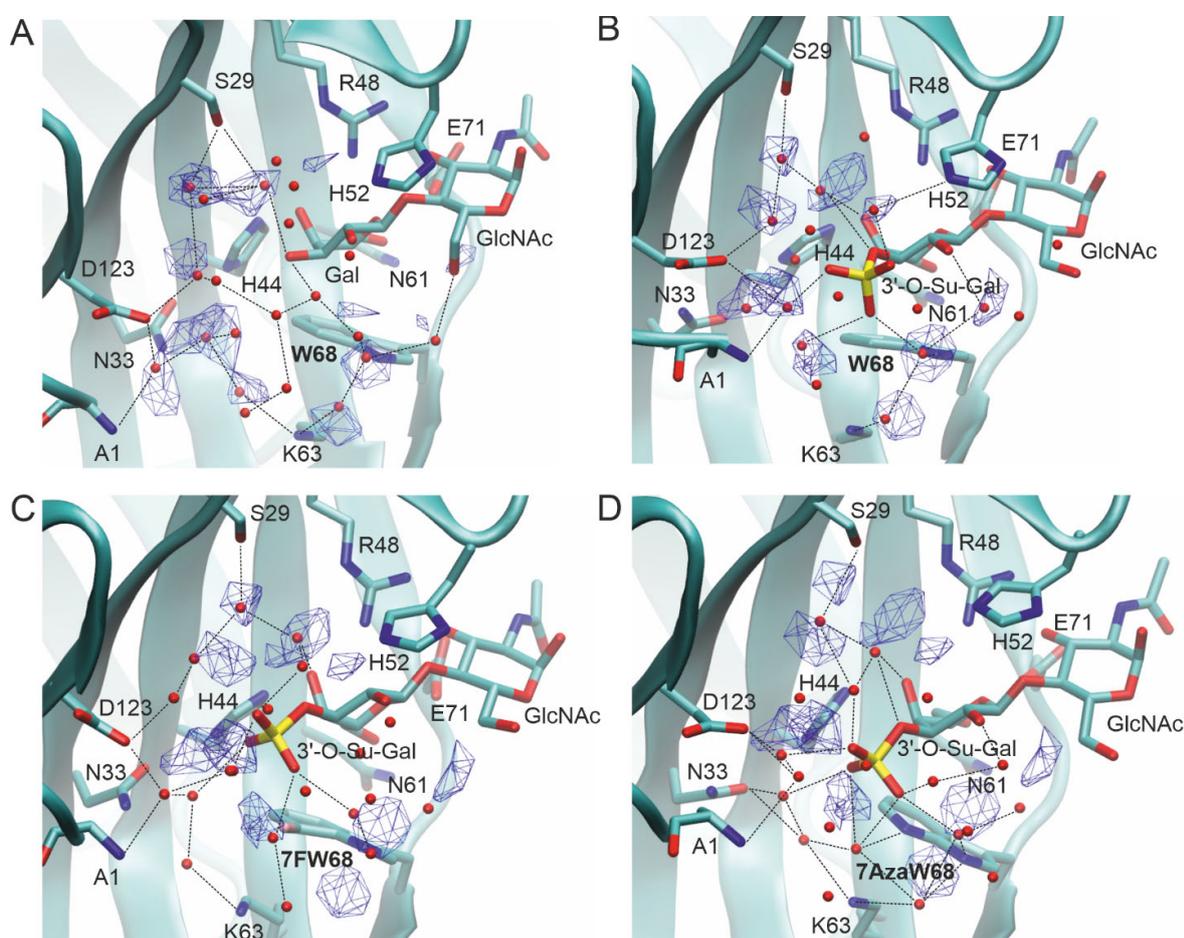


Figure 4: Snapshots from MD simulations of the A) wild-type Gal-1 in complex with LacNAc B) wild-type Gal-1 with 3'-O-Su-Gal- β -1-3-GlcNAc, C) Gal-1 [7FW] with 3'-O-Su-Gal- β -1-3-GlcNAc and D) with Gal-1 [7AzaW] with 3'-O-Su-Gal- β -1-3-GlcNAc. Water oxygen occupancies are shown as blue meshes. Figure rendered with VMD, ver. 1.9.3.^[42]

We evaluated the effects of 7FW and 7AzaW analogs in position 68 of Gal-1 complexes with LacNAc and 3'-O-Su-LacNAc from 1 μ s MD simulations. The stacking distances between (sulfated) galactose and residue 68 become slightly larger (i.e. weaker) in the order

W68/LacNAc ~ 7FW68/LacNAc ~ W68/3'-O-Su-LacNAc < 7AzaW68/LacNAc ~ 7FW68/3'-O-Su-LacNAc < 7AzaW68/3'-O-Su-LacNAc (Table S7, Table S8). In 7FW68/3'-O-sulfate-Gal and 7AzaW68/3'-O-sulfate-Gal complexes this weakened stacking is due to electrostatic repulsion between S-O and F-C or N moieties that are only partially screened by surrounding water networks (Figure 4C, D). Overall, with the exception of the Gal-1/LacNAc and Gal-1/3'-O-Su-LacNAc pair mentioned above, the stacking distance criterion describes the changes in affinity (Figure 3, left panel).

Conclusions

Recently, we have shown that introducing atomic mutations on the tryptophan residues involved in ligand binding in the *Ralstonia solanacearum* lectin can modulate its interaction with different ligands.^[12] However, the effects of fluorination of the three tryptophan residues present in the ligand binding sites interfered with each other. Here, we have chosen a lectin with only one tryptophan residue in the ligand binding site, which plays a key-role in the interaction with carbohydrates. Remarkably, only the affinity for the sulfated ligands 3'-O-Su-LacNAc and 3',6-di-O-Su-LacNAc was substantially weakened by the introduction of the additional fluorine or nitrogen atoms. The effect of reduced affinity of the CSGal-1[7FW] and CSGal-1[7AzaW] variants for the sulfated ligands might be even more pronounced than indicated by the results: The CSGal-1[7FW] protein preparation contained approx. 8% unlabeled parent protein, which very likely affects the affinity measurements yet cannot be removed by standard purification procedures. The removal of the parent protein impurity from the CSGal-1[7AzaW] preparation is not viable as its physico-chemical properties very closely resemble those of the 7FW variant. Due to the impurity of the assessed synthetic proteins with parent protein the effect of reduced affinity for the sulfated ligands might be even more pronounced than visible from the present results.

To explain the effect of 3'-O-sulfation of LacNAc on binding to Gal-1 and its synthetic analogs, we used MD simulations. They hinted at the importance of W68/Gal stacking as well as water-mediated H-bonding at the protein-ligand interface, which are both modulated upon changes in the protein and/or ligand. Solvation effects around W68 in Gal-1 were studied previously by UV resonance Raman spectroscopy and molecular dynamics simulation.^[43] The effects of fluorinated tryptophan residues on ligand-receptor interactions might be even more complex than the previously observed decreased aromaticity of the tryptophan indole ring and the introduction of additional hydrogen bonds with fluorine as weak acceptor.^[12] Quantification of some of these phenomena in galectin-3 was carried out recently by quantum mechanical methods.^[44,45]

In summary, the introduction of atomic mutations in the ligand binding site of galectins is a promising strategy for modifying the interaction with different carbohydrates. In particular, targeting the important tryptophan residue, which cannot be mutated satisfyingly by classical approaches, might be an attractive strategy to fine-tune the ligand specificity of other galectins as well, since the tryptophan residue is highly conserved among them.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF): I 1708-B22 (ERA-SynBio). The COMET center: acib: Next Generation Bioproduction is funded by BMK, BMDW, SFG,

Standortagentur Tirol, Government of Lower Austria und Vienna Business Agency in the framework of COMET - Competence Centers for Excellent Technologies. The COMET-Funding Program is managed by the Austrian Research Promotion Agency FFG. AI acknowledges support from the ANR PIA Glyco@Alps (ANR-15-IDEX-02) and Labex Arcane/CBH-EUR-GS (ANR-17-EURE-0003). M.L. has received funding for this project from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.795605, E.U. Part of the computations presented in this paper were performed using the Dahu platform of the CIMENT infrastructure which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), France and the Equip@Meso project (reference ANR-10-EQPX-29-01). The work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), E.U., with the support of the EC Research Innovation Action under the H2020 Programme; in particular, ML gratefully acknowledges the computer resources and technical support provided by EPCC at the University of Edinburgh, Scotland.

Material and Methods - see supplemental information below

References

- [1] P. Bojarová, V. Křen, *Biomater. Sci.* **2016**, *4*, 1142-1160.
- [2] X. Dan, W. Liu, T. B. Ng, *Med. Res. Rev.* **2016**, *36*, 221-247.
- [3] J. Hirabayashi, M. Yamada, A. Kuno, H. Tateno, *Chem. Soc. Rev.* **2013**, *42*, 4443-4458.
- [4] W. B. Turnbull, A. Imberty, O. Blixt, *Interface Focus* **2019**, *9*, 20190004.
- [5] W. Kightlinger, K. F. Warfel, M. P. DeLisa, M. C. Jewett, *ACS Synth. Biol.* **2020**, *9*, 1534-1562.
- [6] J. Arnaud, A. Audfray, A. Imberty, *Chem. Soc. Rev.* **2013**, *42*, 4798-4813.
- [7] J. Hirabayashi, R. Arai, *Interface Focus* **2019**, *9*, 20180068.
- [8] S. Notova, F. Bonnardel, F. Lisacek, A. Varrot, A. Imberty, *Curr. Opin. Struct. Biol.* **2020**, *62*, 39-47.
- [9] I. Wagner, H. Musso, *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 816-828.
- [10] A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **2015**, *6*, 50-69.
- [11] O. Vargas-Rodriguez, A. Sevostyanova, D. Söll, A. Crnković, *Curr. Opin. Chem. Biol.* **2018**, *46*, 115-122.
- [12] F. Tobola, M. Lelimosin, A. Varrot, E. Gillon, B. Darnhofer, O. Blixt, R. Birner-Gruenberger, A. Imberty, B. Wiltschi, *ACS Chem. Biol.* **2018**, *13*, 2211-2219.
- [13] E. Shanina, E. Siebs, H. Zhang, D. Varon Silva, I. Joachim, A. Titz, C. Rademacher, *Glycobiology* **2021**, *31*, 159-165.
- [14] H. J. Allen, H. Ahmed, K. L. Matta, *Glycoconj. J.* **1998**, *15*, 691-695.
- [15] S. H. Barondes, V. Castronovo, D. N. W. Cooper, R. D. Cummings, K. Drickamer, T. Felzi, M. A. Gitt, J. Hirabayashi, C. Hughes, K.-i. Kasai, H. Leffler, F.-T. Liu, R. Lotan, A. M. Mercurio, M.

- Monsigny, S. Pillai, F. Poirer, A. Raz, P. W. J. Rigby, J. M. Rini, J. L. Wang, *Cell* **1994**, *76*, 597-598.
- [16] V. I. Teichberg, I. Silman, D. D. Beitsch, G. Resheff, *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 1383-1387.
- [17] N. Ahmad, H. J. Gabius, S. Sabesan, S. Oscarson, C. F. Brewer, *Glycobiology* **2004**, *14*, 817-825.
- [18] S. R. Stowell, C. M. Arthur, P. Mehta, K. A. Slanina, O. Blixt, H. Leffler, D. F. Smith, R. D. Cummings, *J. Biol. Chem.* **2008**, *283*, 10109-10123.
- [19] A. Leppänen, S. Stowell, O. Blixt, R. D. Cummings, *J. Biol. Chem.* **2005**, *280*, 5549-5562.
- [20] F. Cedeno-Laurent, C. J. Dimitroff, *Clin. Immunol.* **2012**, *142*, 107-116.
- [21] C. M. Arthur, M. D. Baruffi, R. D. Cummings, S. R. Stowell, *Methods Mol. Biol.* **2015**, *1207*, 1-35.
- [22] I. Camby, M. Le Mercier, F. Lefranc, R. Kiss, *Glycobiology* **2006**, *16*, 137R-157R.
- [23] M. Cho, R. D. Cummings, *J. Biol. Chem.* **1995**, *270*, 5207-5212.
- [24] J. Hirabayashi, K. Kasai, *J. Biol. Chem.* **1991**, *266*, 23648-23653.
- [25] N. Nishi, A. Abe, J. Iwaki, H. Yoshida, A. Itoh, H. Shoji, S. Kamitori, J. Hirabayashi, T. Nakamura, *Glycobiology* **2008**, *18*, 1065-1073.
- [26] M. F. López-Lucendo, D. Solís, S. André, J. Hirabayashi, K.-i. Kasai, H. Kaltner, H.-J. Gabius, A. Romero, *J. Mol. Biol.* **2004**, *343*, 957-970.
- [27] N. Bertleff-Zieschang, J. Bechold, C. Grimm, M. Reutlinger, P. Schneider, G. Schneider, J. Seibel, *ChemBioChem* **2017**, *18*, 1477-1481.
- [28] T. J. Hsieh, H. Y. Lin, Z. Tu, B. S. Huang, S. C. Wu, C. H. Lin, *PLoS One* **2015**, *10*, e0125946.
- [29] I. V. Nesmelova, E. Ermakova, V. A. Daragan, M. Pang, M. Menéndez, L. Lagartera, D. Solís, L. G. Baum, K. H. Mayo, *J. Mol. Biol.* **2010**, *397*, 1209-1230.
- [30] J. M. Romero, M. Trujillo, D. A. Estrin, G. A. Rabinovich, S. Di Lella, *Glycobiology* **2016**, *26*, 1317-1327.
- [31] L. L. Kiessling, R. C. Diehl, *ACS Chem. Biol.* **2021**, *16*, 1884-1893.
- [32] J. L. Asensio, A. Arda, F. J. Canada, J. Jimenez-Barbero, *Acc. Chem. Res.* **2013**, *46*, 946-954.
- [33] M. G. Ford, T. Weimar, T. Kohli, R. J. Woods, *Proteins* **2003**, *53*, 229-240.
- [34] C. Meynier, F. Guerlesquin, P. Roche, *J. Biomol. Struct. Dyn.* **2009**, *27*, 49-58.
- [35] W. M. Abbott, T. Feizi, *J. Biol. Chem.* **1991**, *266*, 5552-5557.
- [36] B. Wiltschi in *Synthetic Biology* (Eds.: A. Glieder, C. Kubicek, D. Mattanovich, B. Wiltschi, M. Sauer), Springer, Cham, **2016**, pp. 143-209.
- [37] F. Tobola, E. Sylvander, C. Gafko, B. Wiltschi, *Interface Focus* **2019**, *9*, 20180072.
- [38] R. F. Frederiksen, Y. Yoshimura, B. G. Storgaard, D. K. Paspaliari, B. O. Petersen, K. Chen, T. Larsen, J. Ø. Duus, H. Ingmer, N. V. Bovin, U. Westerlind, O. Blixt, M. M. Palcic, J. J. Leisner, *J. Biol. Chem.* **2015**, *290*, 5354-5366.

- [39] K. Peterson, P. M. Collins, X. Huang, B. Kahl-Knutsson, S. Essén, F. R. Zetterberg, S. Oredsson, H. Leffler, H. Blanchard, U. J. Nilsson, *RSC Adv.* **2018**, *8*, 24913-24922.
- [40] H. J. Allen, H. Ahmed, K. L. Matta, *Glycoconj. J.* **1998**, *15*, 691-695.
- [41] L. Pegado, O. Marsalek, P. Jungwirth, E. Wernersson, *Phys. Chem. Chem. Phys.* **2012**, *14*, 10248-10257.
- [42] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33-38.
- [43] S. Di Lella, L. Ma, J. C. Díaz Ricci, G. A. Rabinovich, S. A. Asher, R. M. S. Álvarez, *Biochemistry* **2009**, *48*, 786-791.
- [44] R. Kumar, K. Peterson, M. Misini Ignjatović, H. Leffler, U. Ryde, U. J. Nilsson, D. T. Logan, *Org. Biomol. Chem.* **2019**, *17*, 1081-1089.
- [45] R. Kumar, M. M. Ignjatović, K. Peterson, M. Olsson, H. Leffler, U. Ryde, U. J. Nilsson, D. T. Logan, *ChemMedChem* **2019**, *14*, 1528-1536.

Supporting Information

Experimental procedures

Protein expression, purification and sample preparation

CSGal-1 and variants were expressed using SPI and a tryptophan auxotrophic *E. coli* strain carrying the pQE80L_6H-CSGal-1 plasmid^[1], as described previously.^[2] In short, 500 mL of M9 medium (47.76 mM Na₂HPO₄, 22.04 mM KH₂PO₄, 8.56 mM NaCl, 18.69 mM NH₄Cl, 22 mM α-D-glucose, 1 mM MgSO₄, 0.1 mM CaCl₂, 8.63 μM FeSO₄, 3.55 μM MnSO₄, 2.49 μM AlCl₃, 1.84 μM CoCl₂, 0.42 μM ZnSO₄, 0.5 μM Na₂MoO₄, 0.35 μM CuCl₂, 0.49 μM H₃BO₃) supplemented with 1% (w/v) casamino acids (BD Biosciences, San Jose, CA), 35 μM L-tryptophan, and 100 μg mL⁻¹ ampicillin were incubated with the auxotrophic *E. coli* strain carrying the expression plasmid to an attenuation at 600 nm (D₆₀₀) of 0.1 and incubated at 37 °C and 120 rpm. After tryptophan- and growth depletion at a D₆₀₀ of around 3, which resulted in stalled growth, the cells were starved for an additional hour before 1 mM indole (Sigma-Aldrich, St. Louis, MO) or indole analog (4-fluoroindole, Tokyo Chemical Industry Europe, Zwijndrecht, Belgium; 5-, 6-, and 7-fluoroindole and 4-azaindole, Molekula, Newcastle Upon Tyne, U.K.; 4- and 5-hydroxyindole and 4-aminoindole, Abcr, Karlsruhe, Germany; 5- and 6-azaindole, Alfa Aesar, Heysham, U.K.; 7-azaindole and 1-methylindole, Sigma-Aldrich) was added together with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG, Sigma-Aldrich). Cells were incubated at 28 °C, 115 rpm overnight (approx. 16 h) and subsequently harvested by centrifugation (20 min at 8000 xg and 4 °C) and stored as pellets at -20 °C until further processing.

Purification was facilitated by immobilized metal ion affinity chromatography *via* a hexahistidine-tag present at the N-terminus of CSGal-1 (Table S5). Therefore, cell pellets were resuspended in 25 mL lysis buffer (50 mM NaPi, 150 mM NaCl, 10 mM imidazole, pH 7.2) and disrupted by sonication. The lysate was centrifuged for 30 min at 20000 xg and 4 °C. Subsequently, the supernatant was loaded onto a 1.5 mL Ni-NTA agarose (Cube Biotech, Monheim, Germany) column equilibrated with 60 mL of lysis buffer and unbound protein was removed by washing with 15 mL wash buffer (50 mM NaPi, 150 mM NaCl, 30 mM imidazole, pH 7.2). The target proteins were eluted by applying 10 times 1 mL of elution buffer (50 mM NaPi, 150 mM NaCl, 300 mM imidazole, pH 7.2). Eluted fractions containing the target proteins (determined by protein concentration measurements and SDS-PAGE) were pooled, filtered through 0.2 μM PES syringe filters and subsequently buffer-exchanged against phosphate-buffered saline (PBS; 9.55 mM Na₂HPO₄, 136.89 mM NaCl, 2.68 mM KCl and 1.47 mM KH₂PO₄) using a HiPrep 26/10 column (GE Healthcare, Uppsala, Sweden). Small aliquots (50 μL to 300 μL) of the protein solution were lyophilized and stored at 4 °C until use. Purified protein titers in mg L⁻¹ of cell culture were: CSGal-1: 48 mg L⁻¹; CSGal-1[5FW]: 21 mg L⁻¹; CSGal-1[6FW]: 20 mg L⁻¹; CSGal-1[7FW]: 32 mg L⁻¹; CSGal-1[4OHW]: 7 mg L⁻¹; CSGal-1[5OHW]: 10 mg L⁻¹; CSGal-1[4AzaW]: 8 mg L⁻¹; CSGal-1[5AzaW]: 4 mg L⁻¹; CSGal-1[6AzaW]: 4 mg L⁻¹; CSGal-1[7AzaW]: 29 mg L⁻¹; CSGal-1[1MeW]: 9 mg L⁻¹; CSGal-1[4NH2W]: 10 mg L⁻¹.

For experiments, protein samples were reconstituted in PBS pH 7.2 by the addition of doubly distilled H₂O (ddH₂O) to lyophilized powder. The solution was centrifuged for 5 min at 7200 xg at room temperature (RT) and the protein concentration of the supernatant was determined by its absorbance at 280 nm (NanoDrop 2000 spectrophotometer, Thermo Fisher Scientific, Waltham, MA) applying an extinction coefficient of 8.48 x 10⁴ M⁻¹ cm⁻¹. Subsequently protein

solutions were diluted in PLI-P Buffer (6.5 mM Na₂HPO₄, 1.5 mM KH₂PO₄, 500 mM NaCl, 3 mM KCl, 1% (w/v) BSA, 1% (v/v) Triton-X-100, pH 7.4) or in PBS with 0.1 μM BSA (10 mM Na₂HPO₄, 10 mM NaH₂PO₄, 138 mM NaCl, 2.7 mM KCl, 0.1 μM BSA, pH 7.4) for the glycan microarray and fluorescence polarization assays, respectively. The following molecular weights were used for the calculation of the molarity of protein solution: CSGal-1, 15586.32 g mol⁻¹; CSGal-1[7FW], 15604.31 g mol⁻¹; CSGal-1[7AzaW], 15587.19 g mol⁻¹.

Intact protein mass analysis by HPLC ESI-MS

A final protein concentration of 10 ng μl⁻¹ was prepared in water containing 5% (v/v) acetonitrile (ACN) and 0.1% (v/v) trifluoroacetic acid. Protein species were separated by nano-HPLC (Dionex Ultimate 3000) equipped with a Pepswift precolumn (monolithic, 5 x 0.2 mm) and an Acclaim ProSwift RP-4H column (monolithic, 100 μm x 25 cm) (all Thermo Fisher Scientific, Vienna, Austria). Approximately 1 μl of protein sample was injected and concentrated on the enrichment column for 2 min at a flow rate of 5 μl min⁻¹ with 0.1% (v/v) formic acid as an isocratic solvent. Separation was carried out on the nanocolumn at a flow rate of 1 μl min⁻¹ using the following gradient, where solvent A was 0.3% (v/v) formic acid in water and solvent B was ACN containing 0.3% (v/v) formic acid: 0-2 min: 5% B; 2-17 min: 4-60% B; 17-20 min: 60% B; 20-20.1 min: 60-5% B; 20.1-29 min: 5% B. The maXis II ETD mass spectrometer (Bruker, Bremen, Germany) was operated with the captive spray source in positive mode with the following settings: mass range: 300-3000 m/z, 1 Hz, source voltage 1.6 kV and dry gas flow 3 L min⁻¹ at 180 °C. The protein mass spectra were deconvoluted by the data analysis software using the MaxEnt2 algorithm. The following main parameters were applied: charge carrier, H⁺; m/z range, minimum 800 to maximum 2000; minimum instrument resolving power was set to 50000. For peak detection, SNAP algorithms with the following parameters were used: quality factor threshold 0.9, S/N threshold 2 and maximum charge state of 12.

Glycan microarray assay

The glycan microarrays (mammalian printed array version 3.0 from the consortium of functional glycomics^[3]) contained 317 carbohydrates (see supporting excel file) and were printed as described previously.^[4] For details on the amine-functionalized spacers used for immobilization of the glycans on the N-hydroxysuccinimide-activated glass slides by amine-coupling, see the supporting excel file and previous publications.^[5,6] The microarray slides were blocked with blocking buffer (50 mM ethanolamide in 50 mM borate buffer, pH 8.0) for one hour, subsequently rinsed with ddH₂O and spin-dried. The lectins were diluted to 25 μg mL⁻¹ in PLI-P buffer, applied on the microarray slides and incubated for one hour at RT in a humidified chamber with gentle agitation. Subsequently, the microarray glass slides were carefully rinsed with PLI-P buffer and spin-dried, before these were incubated for one hour with anti-his-tag-antibody (MAB050, R&D Systems, Minneapolis, MN) at 4 μg mL⁻¹ at RT in a humidified chamber with gentle agitation. After rinsing the microarray slides with PLI-P buffer, these were incubated for one hour with a cyanine-3-labelled goat anti-mouse IgG antibody (Sigma-Aldrich) as secondary antibody at 4 μg mL⁻¹ in a humidified chamber at RT with gentle agitation, shielded from light. Subsequently, the slides were rinsed with PLI-P buffer, spin-dried and fluorescence measurements were performed using the ScanArray 4000 Microarray Analysis System (PerkinElmer, Waltham, MA). Fluorescence intensities were quantified using the ScanArray Express Microarray Analysis System 4.0 (PerkinElmer) and the data was further analyzed with Microsoft Excel (Microsoft Corporation, Redmond, WA).

Fluorescence polarization experiments

Binding curve with fluorescein-labelled high-affinity-probe. The determination of probe properties was carried out as described previously.^[7] A fixed concentration (4 nM) of the fluorescein-labelled probe 25 (3,3'-Dideoxy-3-(fluorescein-5-yl-carbonylamino)-3'-[4-(thiazol-2-yl)-1H-1,2,3-triazol-1-yl]-1,1'-sulfanediyl-di-β-D-galactopyranoside)^[7] in PBS containing 0.1 μM BSA was mixed with a range of galectin-1 dilutions ranging from 0.3 nM to 20 μM in the same buffer in a final volume of 160 μL at RT, in black 96 well plates (FluoroNunc, Thermo

Fisher). Fluorescence polarization was measured using a PheraStarFS plate reader with software PHERAstar Mars version 2.10 R3 (BMG, Offenburg, Germany) and fluorescence anisotropy of probe 25 measured with excitation at 485 nm and emission at 520 nm. K_d and SEM values were determined in GraphPad Prism as previously described.^[8]

Competitive fluorescence polarization assays with different ligands to determine binding affinities. The determination of K_d values for the interaction of CSGal-1 and variants towards different ligands was carried out as described previously.^[7] A fixed concentration (4 nM) of probe 25 and galectin-1 (CSGal-1, 60 nM; CSGal-1[7FW], 100 nM and CSGal-1[7AzaW], 200 nM) in PBS containing 0.1 μ M BSA was mixed with a range of ligand dilutions between 0.25 μ M and 5 mM of lactose, 0.1 μ M and 2 mM of DiLacNAc, LNT and LNnT (all from Elicityl, Crolles, France) and between 0.46 μ M and 1 mM of 3'-, 3',6- and 6-O-Su-LacNAc (all from GlycoNZ, Auckland, New Zealand) in the same buffer in a final volume of 160 μ L at RT. Measurements were carried out as described above. K_d average, SD and SEM were calculated from 4 to 5 single point measurements showing between 15 and 85% inhibition as previously described.^[8] For more information on the values used for the calculations see Table S4 and Figure S9.

Molecular modelling and simulations

X-ray structures. The crystal structure of the Gal-1 C2S mutant (referred to as Gal-1) with bound LacNAc (PDB: 1W6P, resolution of 1.8 Å)^[9] was selected because it captures a local change in the 122–125 loop with respect to the wild-type. Other cysteine-to-serine mutations in the CSGal-1 protein used in the experiment should induce changes neither in the overall structure, nor in the carbohydrate-binding site. The Gal-1/3'-O-Su-LacNAc complex was modelled from 1W6P for consistency. For structural validations, the related complex of Gal-1/3'-O-Su-Gal- β -1-3-GlcNAc (PDB: 4Y22, unpublished) was used.

Parametrization. GLYCAM parameters for the sulfated carbohydrates were obtained from the glycam.org web page. Parameters for the 7FW and 7AzaW residues were obtained by exchanging/deleting the atoms in PyMol (Molecular Graphics System, Version 1.7.6.3, Schrödinger, LLC), capping the N- and C-termini with acetyl and hydrogen, respectively, and optimizing using the DFT-D3/COSMO method. The potential protonation of the pyridine of 7AzaW was excluded based on the predicted pKa of 3.3 using Chemicalize tool from ChemAxon. Partial atomic charges were derived by the RESP procedure at the HF/6-31G* level.^[10] The capping groups were subsequently deleted and the excess charge was dispersed over all the remaining atoms to restore neutrality. The 7FW and 7AzaW residues were assigned AMBER ff99SB force field parameters.^[11] These parameters for nonstandard residues are given in the Mendeleev Repository.^[12]

Molecular dynamics simulations and analysis. Gal-1 was modeled as dimer but interactions were analyzed in chain A only because the electron densities of the ligands in the 1W6P X-ray structure^[9] were weak in chain B as inspected by WinCoot, ver. 0.8.9.2.^[13] Both histidine residues, H44 and H52, were treated as monoprotonated at H δ to maintain hydrogen bonding patterns. AMBER ff14SB force field^[11] was used for the protein and GLYCAM-06j^[14] for the carbohydrates. The systems were immersed in a box of explicit TIP3P water molecules, neutralized with 0.15 M NaCl, and relaxed stepwise following the published protocol.^[15] Production runs of 1 μ s ensued. Protein-carbohydrate H-bonds were listed only when they exceeded 10% occupancy. W68 stacking was characterized in analogy to reference^[16] as a set of distances between the center of pyrrole and benzene rings of W68 and C3-C6 atoms of Gal. Water networks were calculated from a 10 ns MD simulation with protein and ligand frozen and visualized using the VOLMAP tool. All the analyses were performed with the Cpptraj program.^[17]

Author contributions

F.T., H.L., O.B., U.N. and B.W. designed experiments; F.T. M.L. S.R.Z., performed experiments; F.T., M.L., A.I and B.W. wrote the manuscript.

Supplementary files

Supplementary Figures

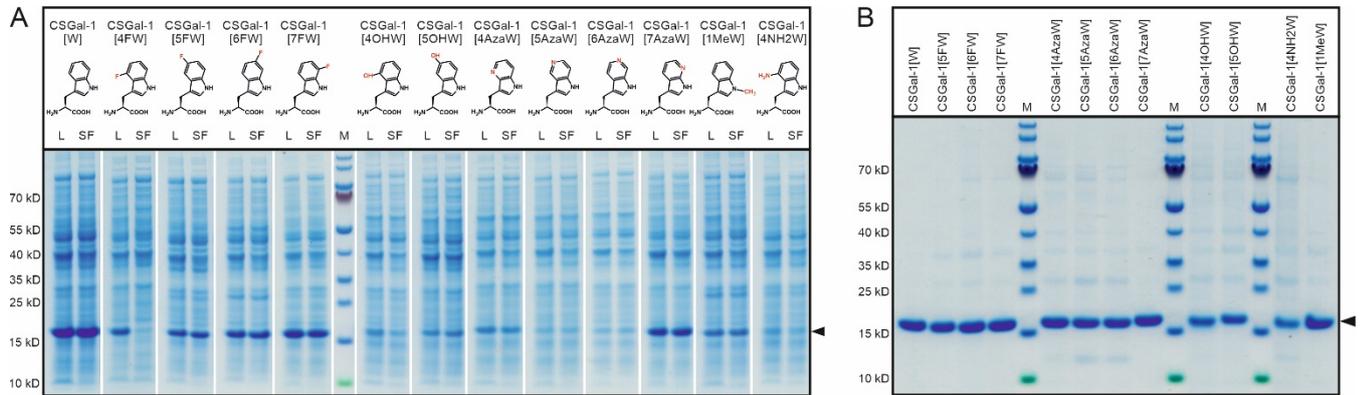


Figure S1. A) SDS-PAGE analysis of CSGal-1 expression in the presence of tryptophan or different tryptophan analogs. Structures of tryptophan and analogs are shown, and atomic mutations are highlighted in red. The protein bands corresponding to CSGal-1 or synthetic variants (MW CSGal-1 15.6 kDa) are indicated by a black arrow. L, total cell lysate; SF, soluble protein fraction; M, molecular size marker. The figure was composed from individual SDS-gels as indicated by the white spaces. B) SDS-PAGE analysis of IMAC-purified CSGal-1 and tryptophan-analog-containing variants.

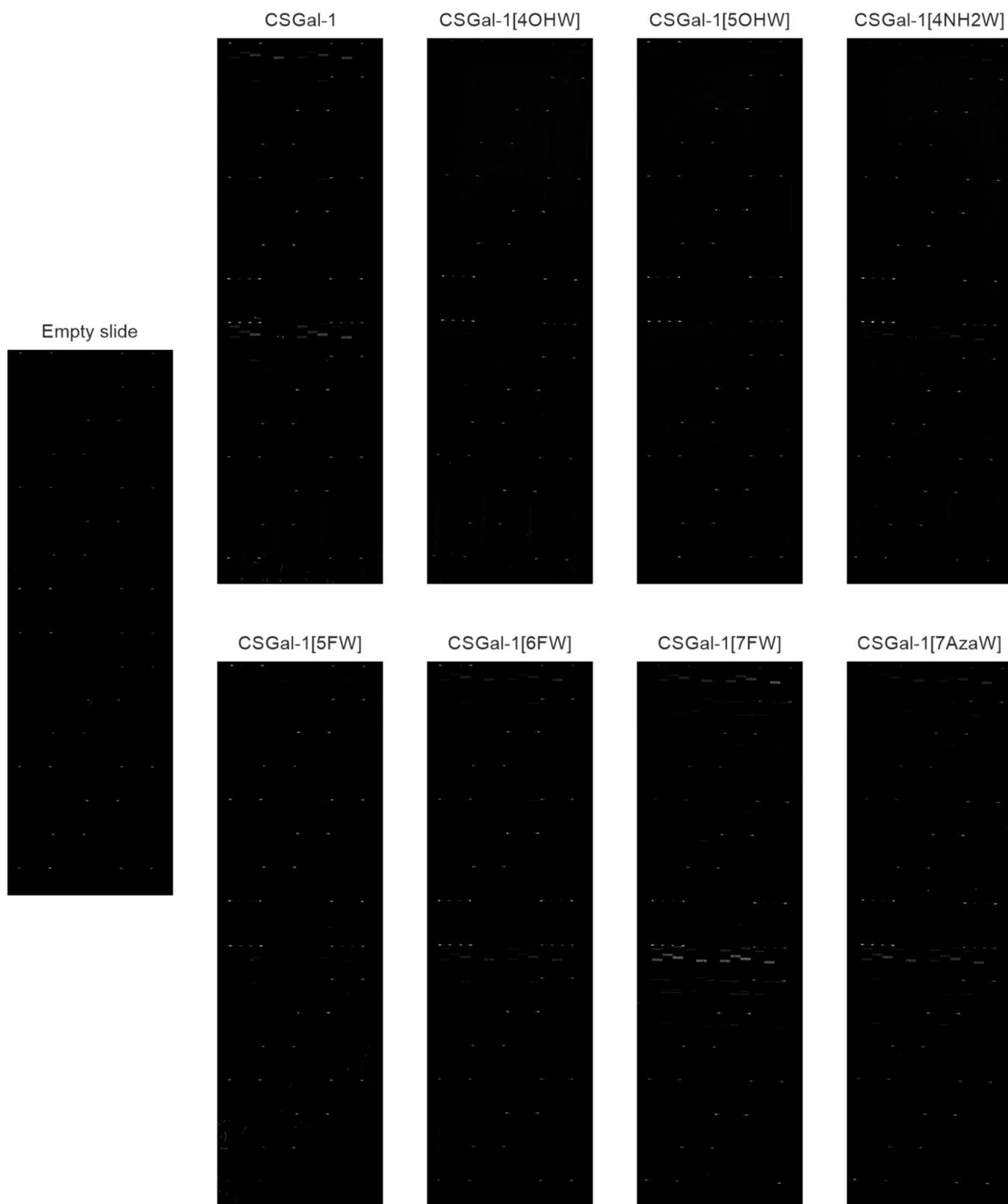


Figure S2. Scan images of glycan microarray slides with fluorescently labelled CSGal-1 and variants. Shown are whole slides containing two replicates of the glycan microarray with 317 glycans each. For details of the glycans, see Supplemental information of Frederiksen, *et al.*^[4]

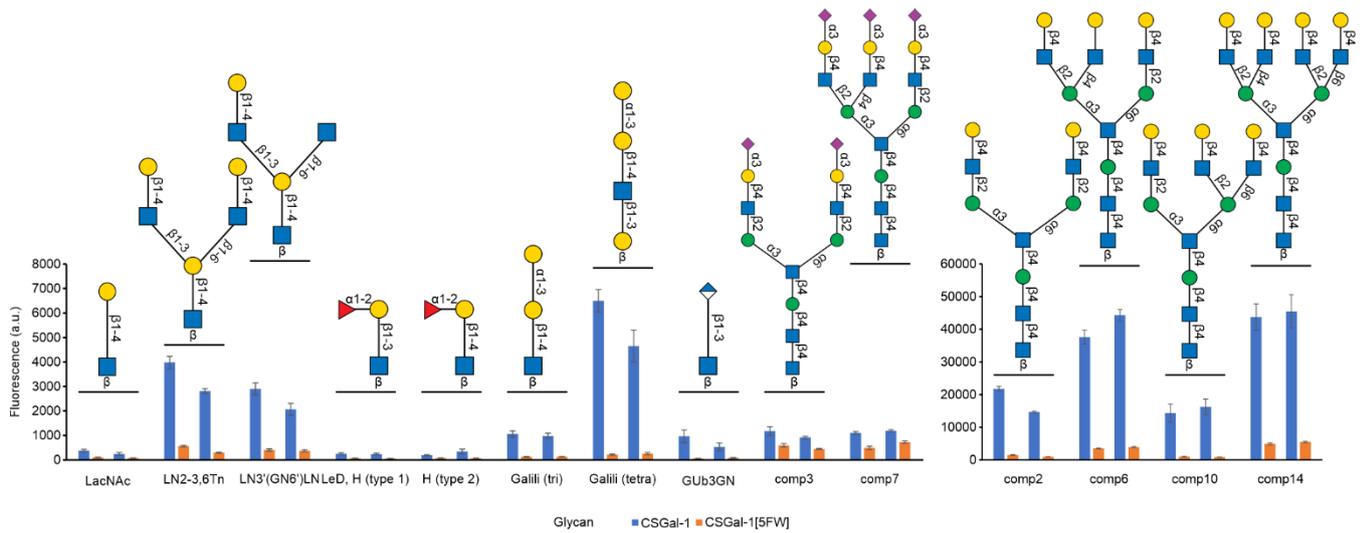


Figure S3. Comparison of glycan microarray fluorescence values of CSGal-1 parent protein and the variant protein containing 5-fluorotryptophan (CSGal-1[5FW]). LacNAc and selected glycans possessing pronounced differences between variant and parent protein are shown. Symbolic representations of glycan structures are shown for each carbohydrate. Bars represent mean values of 8 replicates with SEM in duplicate. Glycan symbols:^[18] were drawn with GlycoGlyph^[19] and adapted in CorelDraw (version 20.1.0.708, Corel Corporation): circle blue, Glc; circle green, Man; circle yellow, Gal; diamond blue divided, GlcA; diamond purple, Neu5Ac; square blue, GlcNAc; square yellow, GalNAc; triangle red, Fuc.

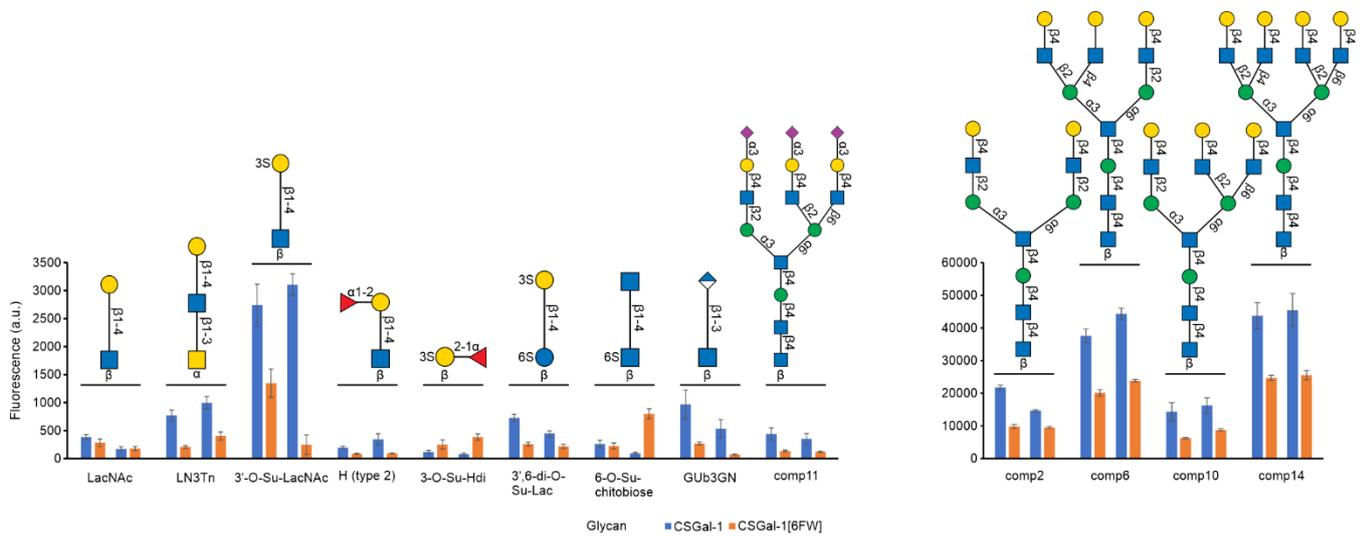


Figure S4. Comparison of glycan microarray fluorescence values of CSGal-1 parent protein and the variant protein containing 6-fluorotryptophan (CSGal-1[6FW]). LacNAc and selected glycans possessing pronounced differences between variant and parent protein are shown. Symbolic representations of glycan structures are shown for each carbohydrate. For definition of glycan symbols see legend to Figure S3.

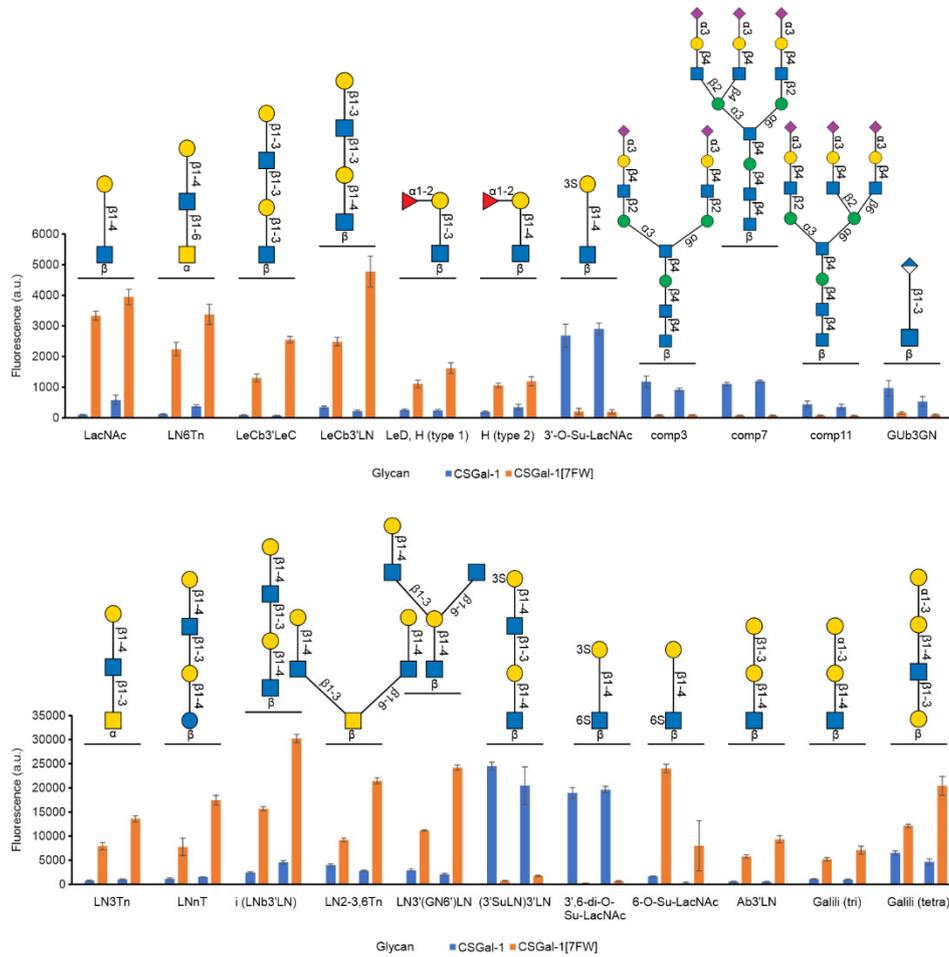


Figure S5. Comparison of glycan microarray fluorescence values of CSGal-1 parent protein and the variant protein containing 7-fluorotryptophan (CSGal-1[7FW]). LacNAc and selected glycans possessing pronounced differences between variant and parent protein are shown. Symbolic representations of glycan structures are shown for each carbohydrate. Bars represent mean values of 8 replicates with SEM in duplicate. Glycan symbols defined in legend to Figure S3.

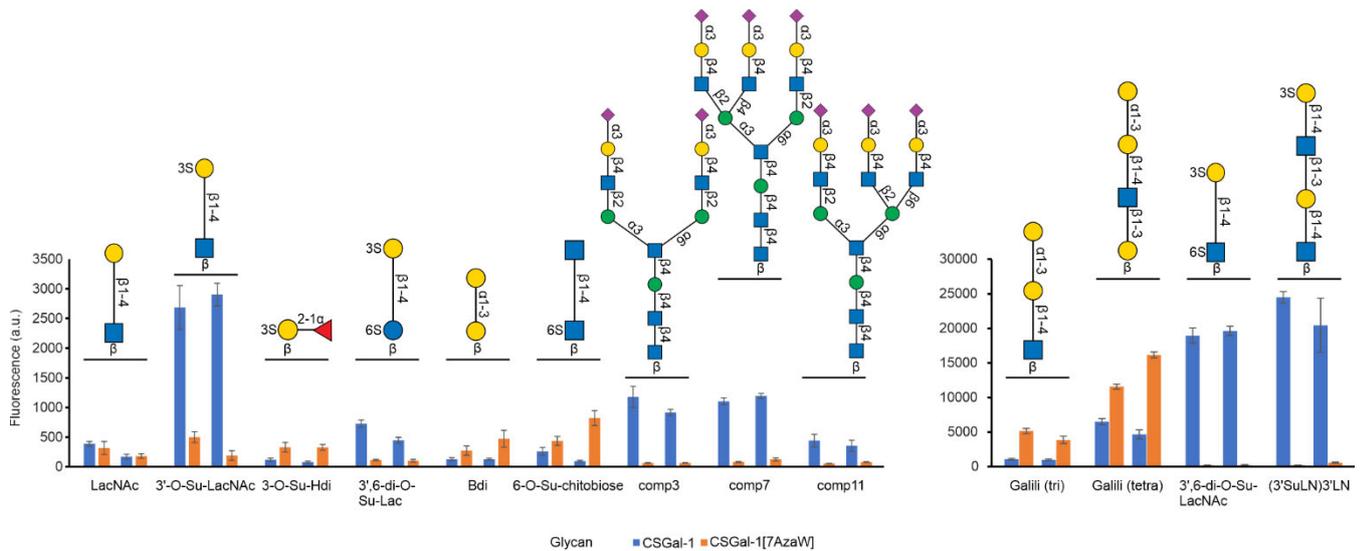


Figure S6. Comparison of glycan microarray fluorescence values of CSGal-1 parent protein and the variant protein containing 7-azatryptophan (CSGal-1[7AzaW]). LacNAc and selected glycans possessing pronounced differences between variant and parent protein are shown. Symbolic representations of glycan structures are shown for each carbohydrate. Bars represent mean values of 8 replicates with SEM in duplicate. Glycan symbols defined in legend to Figure S3.

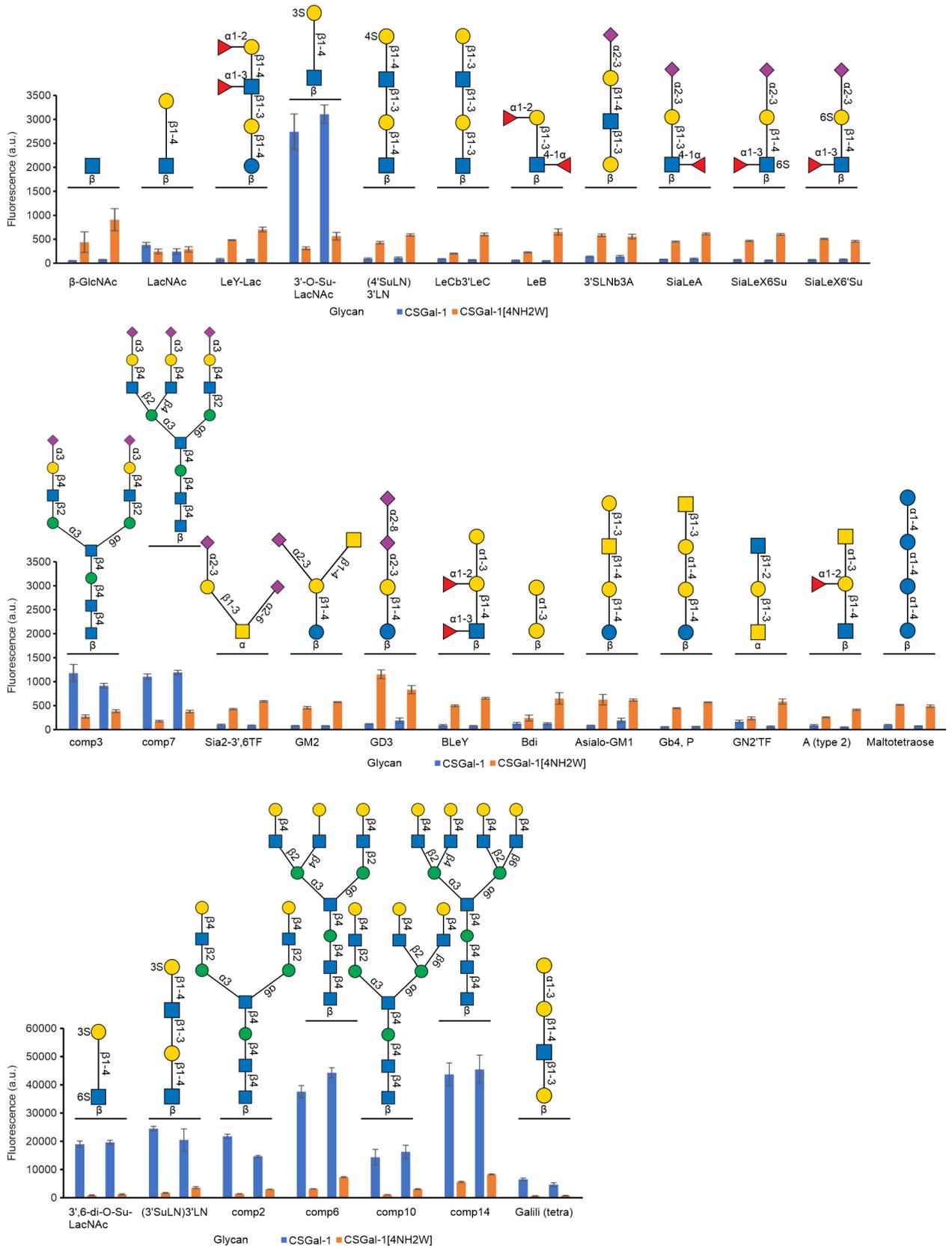
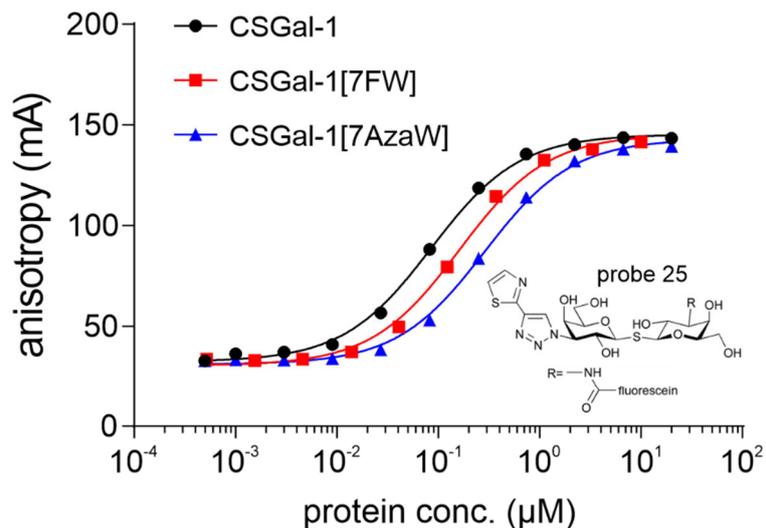


Figure S7. Comparison of glycan microarray fluorescence values of CSGal-1 parent protein and the variant protein containing 4-aminotryptophan (CSGal-1[4NH2W]). LacNAc and selected glycans possessing pronounced differences between variant and parent protein are shown. Symbolic representations of glycan structures are shown for each carbohydrate. Bars represent mean values of 8 replicates with SEM in duplicate. Glycan symbols defined in legend to Figure S3.



	CSGal-1	CSGal-1[7FW]	CSGal-1[7AzaW]
A_{\max}	145.4	145.4	143.1
K_d (μM)	0.0853	0.1615	0.2868

Figure S8. Binding of fluorescein-labelled probe 25^[7] to CSGal-1 and variants containing 7-fluorotryptophan (CSGal-1[7FW]) or 7-azatryptophan (CSGal-1[7AzaW]). The binding curve was measured to find optimal lectin/probe concentrations for the subsequent competition assays (Figure S9). Calculated anisotropy values were plotted against the protein concentrations and the maximum difference in anisotropy between ligand-bound and ligand-free state (A_{\max}) and dissociation constant (K_d) values were calculated. Compared to the CSGal-1 parent protein, CSGal-1[7FW] and CSGal-1[7AzaW] showed a reduced, but still very well binding of the fluorescein-labelled probe. The calculated K_d for CSGal-1 (85 nM) was close to value described earlier for the native Gal-1 (65 nM).^[7] The structure of the fluorescein-labelled probe is shown as inset.

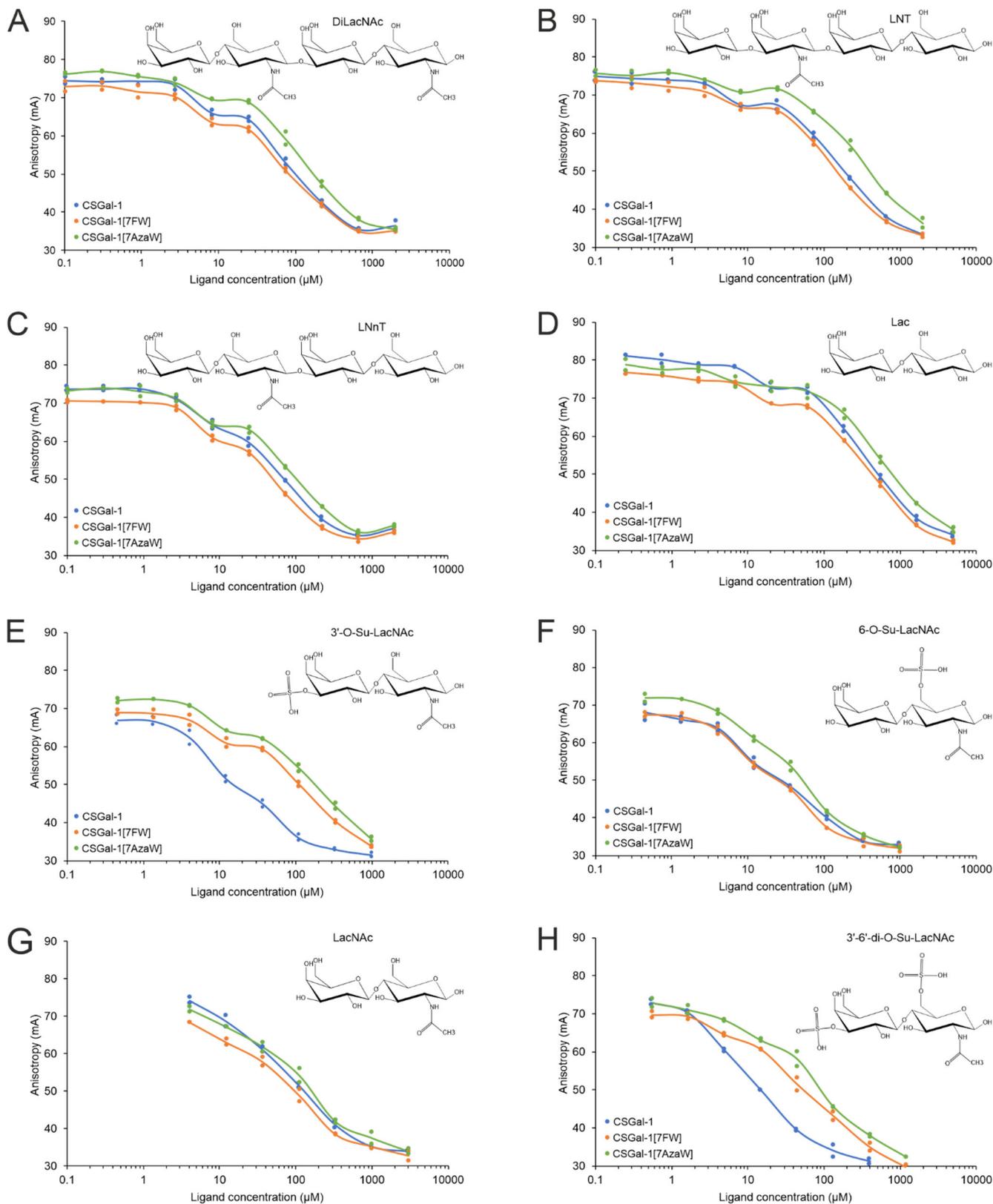


Figure S9. Inhibition curves of CSGal-1, CSGal-1[7FW] and CSGal-1[7AzaW] with different ligands (A-H). The curves were generated from the data shown in Table S4. The duplicate measurements for each inhibitor concentration (points) and their mean (lines) are shown. A fixed concentration (4 nM) of fluorescein-labelled probe 25 and galectin-1 (CSGal-1: 60 nM; CSGal-1[7FW]: 100 nM and CSGal-1[7AzaW]: 200 nM) was mixed with a range of ligand dilutions ranging from 0.25 μM to 5 mM for lactose, from 0.1 μM to 2 mM for DiLacNac, LNT and LNnT and from 0.46 μM to 1 mM for 3'- and 6-O-Su-LacNac and 3',6-di-O-Su-LacNac. Please note that not all inhibition curves could be followed to a plateau at high concentrations, which owes to the limited accessibility of these ligands. Calculated anisotropy values were plotted against the inhibitor (ligand) concentrations.

Supplementary Tables

Table S1. Intact mass analysis of CSGal-1-variants by HPLC ESI-MS. The relative abundance of parent- (WT) and synthetic protein (variant), calculated based on peak intensity, is given in percentage. While no mass shift in comparison to wild type was detected for the 4-, 5- and 6-azatryptophan-containing variants, the expected mass for the 7-azatryptophan containing-variant was detected. The azatryptophan-containing variants only had small mass differences in comparison to the parent protein (0.865 Da), which challenged the sensitivity of the applied method. Therefore, no concrete statement on the incorporation efficiency of these variants (in brackets) could be made and the incorporation was judged as successful (100%) or not (0%).

ncAA introduced ^[a]	Variant (%)	WT (%)
W	0.0	100.0
5FW	83.5	16.5
6FW	93.5	6.5
7FW	91.6	8.4
(4AzaW)	0.0	100.0)
(5AzaW)	0.0	100.0)
(6AzaW)	0.0	100.0)
(7AzaW)	100.0	0.0)
4OHW	58.5	41.5
5OHW	79.2	20.8
4NH2W	75.1	24.9
1MeW	0.0	100.0

[a] W - tryptophan; FW - fluorotryptophan; AzaW - azatryptophan; OHW - hydroxytryptophan; NH2W - aminotryptophan; MeW – methyltryptophan

Table S2. The fluorescence mean values of quantified microarray data from duplicate measurements with 8 technical replicates each are shown for 8 different ligands. For the 8 technical replicates, the standard error of mean (SEM) is given.

Ligand ^[a]	CSGal-1[W]		CSGal-1[7FW]		CSGal-1[7AzaW]	
	Mean	SEM	Mean	SEM	Mean	SEM
Lac	662	117	246	58	210	65
	301	138	61	3	90	12
LacNAc	351	88	186	46	200	45
	318	42	354	97	449	99
LNT	201	30	2529	247	68	5
	100	15	3306	216	126	29
LNnT	1114	201	7752	1797	1360	339
	1529	27	17441	996	2185	156
DiLacNAc	2456	146	15667	422	3736	434
	4540	373	30235	861	4607	509
3',6-di-O-Su-LacNAc	18950	1114	240	47	195	20
	19622	679	648	92	212	62
6-O-Su-LacNAc	1651	133	24020	860	3157	409
	248	194	8007	5205	623	539
3'-O-SuLacNAc	2685	370	208	100	500	90
	2901	192	195	68	190	81

[a] Carbohydrate structures are shown in Figure 2.

Table S3. Calculated K_d values for the interaction of CSGal-1 and variants with eight different ligands measured by fluorescence polarization assay. The tail denotes the range (percentage of inhibition) in which the measurements had to fall to be included in the calculations. E.g. a tail of 25% included measurements showing between 25 % and 75 % inhibition. The number of points used for each calculation can be inferred from Table S4.

Ligand ^[a]	Protein	Tail	K_d ^[b]	SD ^[c]	SEM ^[d]
LacNAc	CSGal-1		34.9	6.7	3.4
	CSGal-1[7FW]	29.0%	35.7	12.6	5.7
	CSGal-1[7AzaW]		57.2	16.5	8.3
DiLacNAc	CSGal-1		88.4	4.5	2.2
	CSGal-1[7FW]	27.0%	72.5	6.3	3.2
	CSGal-1[7AzaW]		130.3	10.9	5.4
LNT	CSGal-1		94.7	6.2	3.1
	CSGal-1[7FW]	25.0%	68.9	3.4	1.7
	CSGal-1[7AzaW]		167.5	19.3	9.7
LNnT	CSGal-1		37.0	3.7	1.9
	CSGal-1[7FW]	25.0%	25.6	6.8	3.1
	CSGal-1[7AzaW]		48.3	5.4	2.4
Lac	CSGal-1		268.5	30.6	15.3
	CSGal-1[7FW]	20.0%	246.5	12.6	6.3
	CSGal-1[7AzaW]		446.1	39.7	19.9
3'-O-Su-LacNAc	CSGal-1		10.5	2.8	1.4
	CSGal-1[7FW]	28.5%	61.2	5.5	2.8
	CSGal-1[7AzaW]		88.2	10.6	5.3
6-O-Su-LacNAc	CSGal-1		16.6	4.0	2.0
	CSGal-1[7FW]	30.0%	13.2	3.2	1.6
	CSGal-1[7AzaW]		18.6	4.6	2.3
3'-6-di-O-Su-LacNAc	CSGal-1		5.6	0.5	0.3
	CSGal-1[7FW]	25.0%	35.4	7.4	3.7
	CSGal-1[7AzaW]		46.9	9.4	4.7

[a] Carbohydrate structures are shown in Figure 2.; [b] Dissociation constant; [c] Standard deviation; [d] Standard error of the mean

Table S4. Anisotropy values of duplicate measurements from the competition assays with different ligands. The values shown in bold were used for K_d calculations (see Table S3 and Figure 3).

Ligand / Variant	Inhibitor concentration / Anisotropy										
Lac	Inhibitor concentration (μ M)										
	5000	1666.67	555.56	185.19	61.73	20.58	6.86	2.29	0.76	0.25	
	CSGal-1	34.5	38.9	49.7	61.1	71.2	73.8	78.0	78.7	79.0	81.1
		33.4	37.7	48.5	62.5	72.7	71.8	78.3	79.0	81.3	81.2
	CSGal-1[7FW]	32.5	36.4	47.8	58.9	68.1	68.5	74.3	74.2	75.8	77.1
		31.9	36.6	46.8	58.6	67.3	68.4	73.3	75.1	76.0	76.3
	CSGal-1[7AzaW]	34.7	42.3	52.9	64.6	69.9	71.7	75.6	78.1	76.6	77.2
		36.0	42.2	54.6	66.9	73.4	74.2	72.8	76.9	78.2	80.2
	DiLacNAc	Inhibitor concentration (μ M)									
		2000	666.67	222.22	74.07	24.69	8.23	2.74	0.91	0.30	0.10
		CSGal-1	37.7	35.6	42.1	52.2	63.8	65.3	72.0	73.2	73.8
35.4			35.6	43.0	53.9	64.8	66.7	74.7	75.5	74.7	75.4
CSGal-1[7FW]		34.7	34.8	41.3	50.5	61.0	62.6	69.5	69.9	71.9	71.5
		35.6	35.2	42.3	51.3	62.2	64.3	70.8	73.0	74.1	74.2
CSGal-1[7AzaW]		35.1	37.9	46.7	57.5	68.4	69.4	73.0	75.3	77.0	76.0
		35.9	38.3	47.9	61.0	69.2	69.7	74.9	75.9	76.8	76.4

Ligand / Variant	Inhibitor concentration / Anisotropy									
LNT	Inhibitor concentration (µM)									
	2000	666.67	222.22	74.07	24.69	8.23	2.74	0.91	0.30	0.10
CSGal-1	33.0	38.0	48.2	58.5	66.2	67.0	72.7	73.9	72.7	73.6
	33.4	37.9	47.9	59.9	68.3	67.5	72.9	73.8	75.8	76.0
CSGal-1[7FW]	32.5	36.9	45.4	56.9	65.4	65.9	69.6	71.1	71.8	73.4
	33.4	36.4	45.6	57.9	66.3	67.5	72.0	73.2	74.4	74.0
CSGal-1[7AzaW]	35.0	44.0	55.4	65.8	71.3	70.5	74.0	75.5	74.0	74.8
	37.5	44.2	58.0	65.4	72.0	71.0	16.6 ^[a]	76.3	76.3	76.6
LNnT	Inhibitor concentration (µM)									
	2000	666.67	222.22	74.07	24.69	8.23	2.74	0.91	0.30	0.10
CSGal-1	37.8	35.4	39.9	49.7	58.6	63.1	71.5	73.2	74.3	73.2
	36.4	35.0	39.0	49.5	60.6	65.4	70.5	74.5	73.3	74.3
CSGal-1[7FW]	35.7	33.5	36.9	45.8	57.3	60.0	68.0	70.1	70.3	70.0
	36.5	35.1	37.6	46.3	56.4	61.5	69.3	70.0	70.4	70.9
CSGal-1[7AzaW]	38.1	35.7	42.6	53.3	62.1	65.1	70.3	71.6	73.8	73.8
	37.4	36.4	42.9	53.0	63.8	63.8	72.1	74.5	74.1	72.7
3'-O-Su-LacNAc	Inhibitor concentration (µM)									
	1000	333.33	111.11	37.04	12.35	4.12	1.37	0.46		
CSGal-1	32.1	33.2	35.1	44.0	50.6	60.4	65.6	65.8		
	30.8	32.6	36.7	45.8	52.2	64.0	67.7	68.1		
CSGal-1[7FW]	33.4	40.0	49.4	59.0	59.8	65.6	67.7	68.4		
	33.9	40.6	50.6	59.5	62.0	68.2	69.7	69.6		
CSGal-1[7AzaW]	35.1	43.6	53.4	61.9	64.2	70.7	72.3	71.5		
	36.2	45.1	55.2	62.2	64.2	70.6	72.5	72.7		
6-O-Su-LacNAc	Inhibitor concentration (µM)									
	1000	333.33	111.11	37.04	12.35	4.12	1.37	0.46		
CSGal-1	32.5	34.0	39.5	48.0	53.1	62.7	65.4	65.7		
	33.1	33.5	40.3	48.4	55.8	64.8	66.7	70.2		
CSGal-1[7FW]	31.0	32.3	37.3	47.8	53.4	62.2	66.1	66.9		
	32.5	34.5	37.3	47.1	54.6	64.3	67.7	68.0		
CSGal-1[7AzaW]	31.9	35.6	41.3	52.5	60.3	67.5	71.5	70.8		
	32.2	35.1	41.8	54.7	61.6	68.8	71.5	72.9		
LacNAc	Inhibitor concentration (µM)									
	3000	1000.00	333.30	111.10	37.10	12.30	4.10			
CSGal-1	33.8	34.6	41.6	51.9	61.7	67.1	74.9			
	33.9	35.6	40.1	50.8	60.2	70.0	73.3			
CSGal-1[7FW]	31.4	34.5	38.5	47.1	56.6	62.2	68.4			
	33.9	35.7	38.2	50.5	58.9	63.9	68.4			
CSGal-1[7AzaW]	33.2	35.7	42.2	56.0	63.0	67.1	71.1			
	34.5	39.1	41.3	52.3	60.5	67.1	72.5			
3'-6-di-O-Su-LacNAc	Inhibitor concentration (µM)									
	1200	400.00	133.30	44.40	14.80	4.94	1.65	0.55		
CSGal-1		31.7	32.3	39.5	49.8	60.0	69.9	72.3		
		30.5	35.5	39.0	49.9	60.7	70.6	73.6		
CSGal-1[7FW]	30.2	33.9	42.0	53.2	60.5	64.9	70.2	68.9		
	29.0	36.0	44.2	49.7	60.8	64.1	68.4	70.6		
CSGal-1[7AzaW]	32.3	37.6	45.6	56.2	63.4	68.1	69.9	71.8		
	32.3	38.3	45.4	60.0	62.7	68.6	72.2	74.1		

[a] This value was excluded from calculations

Table S5. DNA and amino acid sequence of the cysteine-less galectin-1 (CSGal-1) used in this study.

Name	Type	Sequence (DNA [5'→3'] or amino acids (N-terminal→C-terminal) for protein)
CSGal-1	DNA	ATGCATCACCATCACCATCACGGATCGGCCAGCGGCCTTGTAGCCAGTAACTTGAACCTAAAGCCGGGAGAGAGTTTGGCGGTACGTGGCGAA GTTGCTCCCGACGCCAAATCCTTTGTGCTTAACTGGGAAAAGACTCCAATAACCTGAGCTTGCACTTCAACCCCTCGTTTCAATGCACACGGG GACGCAAACACTATTGTGTCAAATAGTAAAGATGGTGGTGCCTGGGGCACGGAGCAACGTGAGGCCGTGTTCCCTTTCCAACCGGGGAGTGTG GCTGAAGTAAGCATTACTTTCGACCAAGCCAACCTGACAGTTAAGTTGCCCGATGGATATGAGTTCAAATTTCCGAATCGCTTAAACTTGGAA CGCATCAACTACATGGCCCGGACGGTGACTTTAAGATTAAGTCTGTGCCTTCGATTGATAA
CSGal-1	Protein ^[a]	MHHHHHGSASGLVASNLNLKPGESLRVRGEVAPDAKSFVLNLGKDSNNLSLHFNPRFNAHGDANTIVSNSKDGGAWGTEQR EAVFPFQPGSVAEVSITFDQANLTVKLPDGYEFKPNRLNLEAINYMAADGDFKIKSVAFD

[a] The introduced hexahistidine-tag including a glycine-serine-linker is shown in blue.

Table S6. Protein-carbohydrate H-bonds for Gal-1/LacNAc complex in X-ray (PDB: 1W6P) and 1 μ s MD

Protein	Carbohydrate	X-ray distance (Å)	Avg dist MD (Å)	MD Occupancy (%)
H44:NE2	Gal:O4	2.7	2.8	99.4
R48:NH2	Gal:O4	3.0	3.0	93.6
R48:NH2	Gal:O5	2.9	3.0	81.1
R48:NH1	GlcNAc:O3	2.7	2.8	100.0
R48:NH2	GlcNAc:O3	3.0	3.2	89.8
R48:NH2	GlcNAc:O4	3.6	3.3	72.2
H52: NE2	Gal:O2	3.3	3.0	59.5
N61:ND2	Gal:O6	2.9	3.0	100.0
E71:OE2	GlcNAc:O3	2.6	2.7	100.0
E71:OE1	GlcNAc:O3	3.3	3.0	65.1
E71:OE2	Gal:O6	2.7	2.7	98.9

Table S7. Protein-carbohydrate stacking at position W68 in Gal-1 in complex with LacNAc. X-ray (PDB: 1W6P^[9]) data are compared to 1 μ s MD simulations of wildtype and the indicated variants. For pyrrole (Pyr) and benzene (Ben) rings, their centroids were used.

Protein	Carbohydrate	X-ray distance (Å)	Average distance MD (Å)		
		1W6P	Gal-1 WT[W68] / LacNAc	Gal-1[7FW] / LacNAc	Gal-1[7AzaW] / LacNAc
W68: Pyr	Gal:C3	4.8	5.5	5.4	5.7
W68: Pyr	Gal:C5	4.0	4.5	4.5	4.8
W68: Ben	Gal:C4	4.1	4.3	4.3	4.4
W68: Ben	Gal:C6	4.2	4.1	4.0	4.2
Average		4.3	4.6	4.6	4.8

Table S8. Protein-carbohydrate stacking at position W68 in Gal-1 in complex with 3'-O-Su-LacNAc. X-ray (PDB: 4Y22, chain B; unpublished) data were compared to 1 μ s MD simulations of the wildtype and the indicated variants. 3'-O-Su-Gal abbreviated as 3LAC; for pyrrole (Pyr) and benzene (Ben) rings, their centroids were used.

Protein	Carbohydrate	X-ray distance (Å)	Average distance MD (Å)		
		4Y22/B	Gal-1 WT[W68] / 3LAC	Gal-1[7FW] / 3LAC	Gal-1[7AzaW] / 3LAC
W68: Pyr	3LAC:C3	5.7	5.6	5.8	5.9
W68: Pyr	3LAC:C5	4.6	4.6	4.8	4.9
W68: Ben	3LAC:C4	4.5	4.3	4.5	4.5
W68: Ben	3LAC:C6	4.0	4.0	4.1	4.2
Average		4.7	4.6	4.8	4.9

Supplementary References

- [1] F. Tobola, E. Sylvander, C. Gafko, B. Wiltschi, *Interface Focus* **2019**, *9*, 20180072.
- [2] F. Tobola, M. Lelimosin, A. Varrot, E. Gillon, B. Darnhofer, O. Blixt, R. Birner-Gruenberger, A. Imberty, B. Wiltschi, *ACS Chem. Biol.* **2018**, *13*, 2211-2219.
- [3] CFG Functional Glycomics Gateway (2010) <http://www.functionalglycomics.org/> (accessed 13/12/2021)
- [4] R. F. Frederiksen, Y. Yoshimura, B. G. Storgaard, D. K. Paspaliari, B. O. Petersen, K. Chen, T. Larsen, J. Ø. Duus, H. Ingmer, N. V. Bovin, U. Westerlind, O. Blixt, M. M. Palcic, J. J. Leisner, *J. Biol. Chem.* **2015**, *290*, 5354-5366.
- [5] O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong, J. C. Paulson, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17033-17038.
- [6] N. Shilova, M. Navakouski, N. Khasbiullina, O. Blixt, N. Bovin, *Glycoconj. J.* **2012**, *29*, 87-91.
- [7] K. Peterson, P. M. Collins, X. Huang, B. Kahl-Knutsson, S. Essén, F. R. Zetterberg, S. Oredsson, H. Leffler, H. Blanchard, U. J. Nilsson, *RSC Adv.* **2018**, *8*, 24913-24922.
- [8] P. Sörme, B. Kahl-Knutsson, M. Huflejt, U. J. Nilsson, H. Leffler, *Anal. Biochem.* **2004**, *334*, 36-47.
- [9] M. F. López-Lucendo, D. Solís, S. André, J. Hirabayashi, K.-i. Kasai, H. Kaltner, H.-J. Gabius, A. Romero, *J. Mol. Biol.* **2004**, *343*, 957-970.
- [10] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269-10280.
- [11] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.
- [12] M. Lepsik, Force-Field Parameters for Tryptophan Analogues (2021) Mendeley Data, V2; doi: 10.17632/tbxs9j8pv.2
- [13] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486-501.
- [14] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622-655.
- [15] P. Srb, M. Svoboda, L. Benda, M. Lepšík, J. Tarábek, V. Šícha, B. Grüner, K. Grantz-Šašková, J. Brynda, P. Řezáčová, J. Konvalinka, V. Veverka, *Phys. Chem. Chem. Phys.* **2019**, *21*, 5661-5673.
- [16] M. Wimmerová, S. Kozmon, I. Nečasová, S. K. Mishra, J. Komárek, J. Koča, *PLoS One* **2012**, *7*, e46032.
- [17] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
- [18] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. G. Vliegthart, T. Lütke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, *Glycobiology* **2015**, *25*, 1323-1324.
- [19] A. Y. Mehta, R. D. Cummings, *Bioinformatics* **2020**, *36*, 3613-3614.