Supporting Information

Effect of non-canonical amino acids on protein-carbohydrate interactions: Structure, dynamics and carbohydrate affinity of a lectin engineered with fluorinated tryptophan analogs.

Felix Tobola[†,‡], Mickael Lelimousin[§], Annabelle Varrot[§], Emilie Gillon[§], Barbara Darnhofer[†,‖,⊥], Ola Blixt[∇], Ruth Birner-Gruenberger[†,‖,⊥], Anne Imberty*[,§], and Birgit Wiltschi*[,†]

† Austrian Centre of Industrial Biotechnology, Petersgasse 14, 8010 Graz, Austria

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

§ Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

‖ Research Unit of Functional Proteomics and Metabolomics, Institute of Pathology, Medical University of Graz, Stiftingtalstrasse 24, 8010 Graz, Austria

⊥ Omics Center Graz, BioTechMed-Graz, Stiftingtalstrasse 24, 8010 Graz, Austria

∇ Department of Chemistry, Chemical Biology, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

DOI: 10.1021/acschembio.XXX

TABLE OF CONTENTS

Supplementary Tables

Table S1 DNA and amino acid (AA) sequences of RSL and primers used in this study.

Name	Type	Sequence (DNA [5'→3'] or amino acids for protein)
rsl	gene	ATGAGCAGTGTTCAGACCGCAGCCACCAGCTGGGGTACCGTTCCGAGCATTCGTGTGT ATACAGCAAATAACGGTAAAATTACCGAACGTTGTTGGGATGGCAAAGGTTGGTATACCG GTGCATTTAATGAACCGGGCGATAACGTTAGCGTGACCTCTTGGCTGGTTGGTAGCGCA ATTCATATCCGTGTGTATGCTAGCACCGGCACCACGACCACGGAATGGTGTTGGGATGG TAATGGCTGGACCAAAGGTGCATATACCGCAACTAATTAATGA
RSL	protein	MSSVQTAATS**W**GTVPSIRVYTANNGKITERC**W**DGKG**W**YTGAFNEPGDNVSVTS**W**LVGSAIHI RVYASTGTTTTE**W**C**W**DGNG**W**TKGAYTATN [a]
pBP645	primer	TATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAAT TAACTATGAGCAGTG
pBP643	primer	CTGGATCTATCAACAGGAGTCCAAGCTCAGCTAATTAAGCTTGGATCCTCATTAATTAGT TGCGG

[a] Tryptophan residues are shown in **bold**. Those located in the ligand binding site are highlighted in **red**.

Table S2 Quantification of FW incorporation into RSL. The RSL monomer contains seven Trp residues. The degree of labeling is given in percentages of the total protein. The results of a single experiment are shown.

Variant	7/7 W labeled (%)	6/7 W labeled (%)	5/7 W labeled (%)
RSL[4FW]	85	14	1
RSL[5FW]	86	14	0
RSL[7FW]	84	15	1

Table S3 Affinity constants and thermodynamics parameters obtained by ITC.

Protein	Ligand	Kd (μM)	-ΔG (kcal mol^{-1})	-ΔH (kcal mol^{-1})	TΔS (kcal mol^{-1})
RSL	αMeFuc	1.21 ± 0.04	8.08	8.72 ± 0.04	-0.62 ± 0.06
	HType2	6.36 ± 0.44	7.09	8.52 ± 0.01	-1.42 ± 0.03
	Le^X	32.5 ± 4.9	6.13	6.01 ± 0.09	0.13 ± 0.18
RSL[4FW]	αMeFuc	1.73 ± 0.41	7.88	8.90 ± 0.08	-1.01 ± 0.22
	HType2	8.18 ± 0.99	6.95	7.85 ± 0.06	-0.89 ± 0.01
	Le^X	67.8 ± 6.8	5.69	6.16 ± 0.04	-0.46 ± 0.10
RSL[5FW]	αMeFuc	0.889 ± 0.004	8.26	8.95 ± 0.04	-0.68 ± 0.04
	HType2	6.06 ± 0.27	7.12	8.48 ± 0.07	-1.35 ± 0.05
	Le^X	52.4 ± 2.6	5.84	6.96 ± 0.05	-1.11 ± 0.08
RSL[7FW]	αMeFuc	1.11 ± 0.02	8.13	8.78 ± 0.04	-0.65 ± 0.05
	HType2	4.35 ± 0.35	7.32	8.06 ± 0.01	-0.91 ± 0.04
	Le^X	52.1 ± 0.8	5.85	6.76 ± 0.01	-0.91 ± 0.02

Table S4 Data collection and refinement statistics for fluorinated RSL co-crystallized with Le^X tetrasaccharide.

Complex	RSL[7FW]/LeX	RSL[5FW]/LeX	RSL[4FW]/LeX
Crystallization conditions	Morpheus 1-8: 60 mM divalent, 100 mM buffer pH 7.5, 12.5% PEG 1K, 12.5% PEG 3350, 12.5% MPD	Midas 1-30: 40% 5/4 PO/OH, 100 mM MES pH 6.0	Morpheus 1-9: 60 mM divalent, 100 mM buffer pH 7.5, 20% PEG 550 MME, 10 M PEG 20K
Beamline	Fip-BM30A	Proxima 1	Proxima 1
Spacegroup	F23	F23	F23
Unit cell: *a*, *b*, *c* (Å) α, β, γ (°)	129.15 129.15 129.15 90.00 90.00 90.00	129.88 129.88 129.88 90.00 90.00 90.00	130.46 130.46 130.46 90.00 90.00 90.00
Wavelength (Å)	0.9799	0.93221	0.93221
Resolution (outer shell), (Å)	32.29-1.15 (1.17-1.15)	45.92-1.28 (1.30-1.28)	19.90-1.35 (1.37-1.35)
Measured / Unique reflections	1348596 / 63033	416999 / 46835	273755 / 40182
R_{merge}	0.066 (0.641)	0.041 (0.456)	0.0052 (0.164)
Mean I (σI)	33.2 (5.6)	21.2 (3.9)	22.6 (8.7)
Completeness (%)	99.9 (100)	99.9 (99.0)	99.5 (98.0)
Average multiplicity	21.4 (20.6)	8.9 (8.1)	6.8 (6.6)
CC1/2	1.000 (0.962)	1.000 (0.944)	0.999 (0.984)
R_{cryst} / R_{free}	10.4 / 12.9	12.2 / 14.8	12.6 / 15.6
No. reflections / free reflections	59892 / 3140	44461 / 2368	18130 / 2052
RMSD Bond length(Å)	0.0149	0.0141	0.0145
RMSD Bond angles (°)	1.963	2.011	1.803
RMSD Chiral ($Å^3$)	0.105	0.105	0.107
No atoms (A/B)			
Protein	711/776	707/723	719/700
Ligand	99/75	72/83	72/89
Waters	135/142	81/70	144/143
B factors ($Å^2$)			
Protein	9.6/10.1	4.6/4.3	13.1/13.8
Ligand	14.4/13.7	10.6/8.4	22.8/21.1
Waters	22.8/23.9	28.3/27.7	19.6/17.6
Ramachandran plot:			
Allowed (%)	100	100	100
Favored (%)	97.4	98.9	96.6
Outliers (%)	0	0	0
PDB code	5O7U	5O7V	5O7W

Table S5 Conformation of the Le^X tetrasaccharide in each binding site. When alternative conformations were observed, only the one with stronger electron density is indicated.

Crystal	Binding site		Conformation Le^X	αFuc1-3GlcNAc		βGal1-4GlcNAc		βGlcNAc1-3Gal	
				Φ [a]	Ψ	Φ	Ψ	Φ	Ψ
RSL[7FW]	intramonomeric	chain A	open I [b] (trisaccharide)	-90	-72	-93	70	-	-
		chain B	(disaccharide)	-68	141	-	-	-	-
	intermonomeric	chain A	open III	-94	123	-38	82	-82	104
		chain B	open III	-96	147	-54	96	-85	104
RSL[5FW]	intramonomeric	chain A	(disaccharide)	-92	-72	-	-	-	-
		chain B	open I (trisaccharide)	-99	-67	-97	61	-	-
	intermonomeric	chain A	open III	-93	143	-49	90	-61	121
		chain B	open III	-94	141	-50	92	-64	115
RSL[4FW]	intramonomeric	chain A	(disaccharide)	-99	-69	-	-	-	-
		chain B	glycerol	-	-	-	-	-	-
	intermonomeric	chain A	open III	-92	123	-42	83	-102	76
		chain B	open III	-93	124	-49	88	-82	104

[a] torsion angles defined as Φ = Θ (O5-C1-O1-Cx) and Ψ = Θ (C1-O1-Cx-C(x+1))

[b] Nomenclature of conformation as described previously.[1]

Table S6 Atomic charges used for Amber parameterization of the fluorinated tryptophan residues. The fluorinated indole rings were energy-minimized by performing Hartree Fock (HF) calculations using a 6-31G(d) basis set. The final electron densities were used to compute atomic charges, following the RESP definition.[2]

	TRP	4FW	5FW	7FW
CG	-0.14150	-0.16320	-0.16320	-0.16070
CD1	-0.16380	-0.14000	-0.13860	-0.15460
HD1	0.20620	0.19440	0.19800	0.20700
NE1	-0.34180	-0.33360	-0.35280	-0.34530
CE2	0.13800	0.11410	0.17080	0.11410
HZ2 (7F)	0.15720	0.16090	0.17930	-0.19450 (7F)
CH2	-0.11340	-0.20500	-0.15350	-0.15680
CZ3	-0.19720	-0.22440	0.19220	-0.27920
HZ3 (5F)	0.14470	0.15850	-0.19970 (5F)	0.17710
HE3 (4F)	0.17000	-0.18490 (4F)	0.17270	0.15460

Supplementary Figures

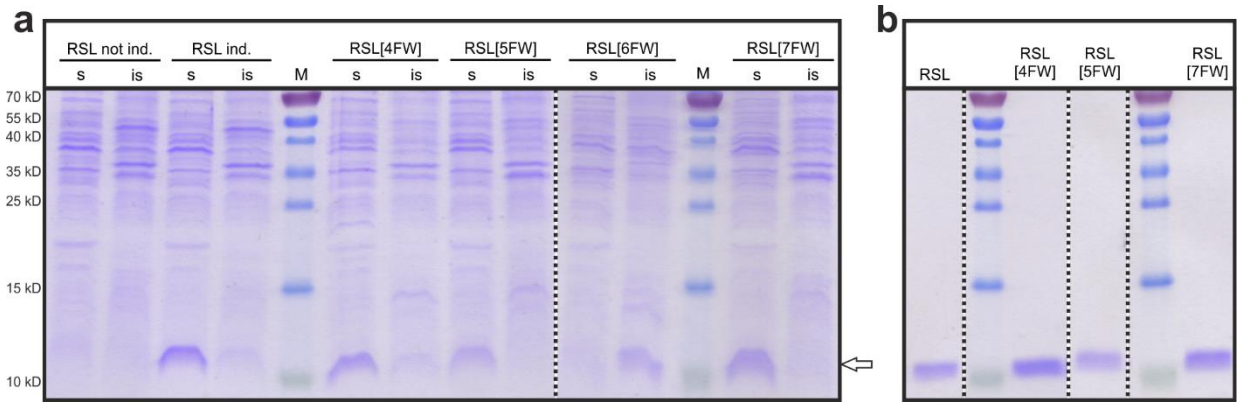

Figure S1. Expression (a) and purification (b) of fluorotryptophan-containing RSL variants. a) SDS-PAGE analysis of the soluble (s) and insoluble (is) lysate fractions of *E. coli* cultures expressing RSL and RSL[FW]-variants (indicated by the arrow). RSL was expressed only in the presence of inducer (RSL ind.), while no target protein expression was observed in its absence (not. ind.). 4-, 5-, and 7-fluorotryptophan containing variants were expressed at levels comparable to the parent protein. The incorporation of 6-fluorotryptophan (RSL[6FW]) drove the protein into insolubility. b) SDS-PAGE analysis of the purified RSL and RSL[FW]-variants. All proteins are present in very high purity and no contaminating bands are visible. The figures were assembled from different gel-images, which are visualized and separated by dotted lines.

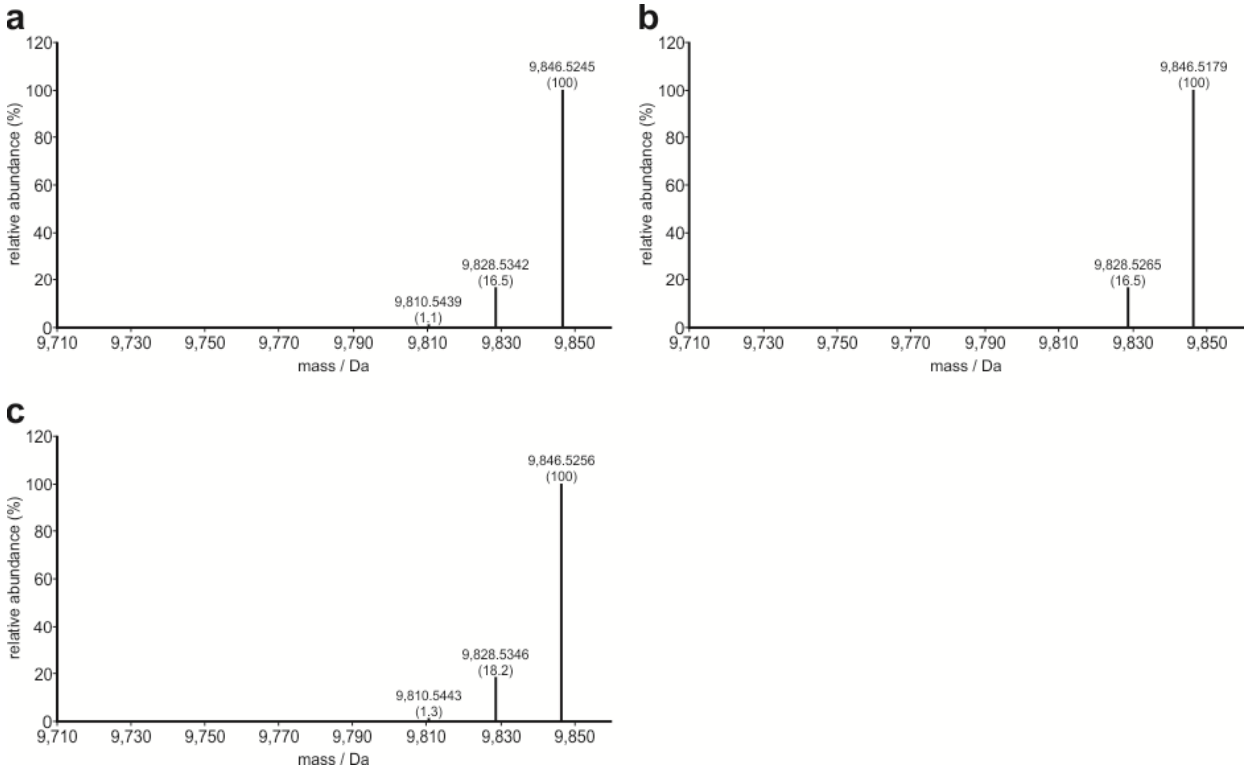

Figure S2. Mass analysis of the fluorinated RSL variants. Deconvoluted spectra of monoisotopic masses of (a) RSL[4FW], (b) RSL[5FW] and (c) RSL[7FW] are shown. Calculated masses are: 9846.5212 Da, 9828.5306 Da and 9810.5400 Da for fully labelled RSL, RSL with 6 out of 7 W exchanged against FW and with 5 out of 7 W exchanged against FW, respectively. The found monoisotopic masses are shown on top of the peaks. The peak intensities of the fully labeled variants were set to 100% and relative abundances (in %) of partially labeled variants are shown in parentheses.

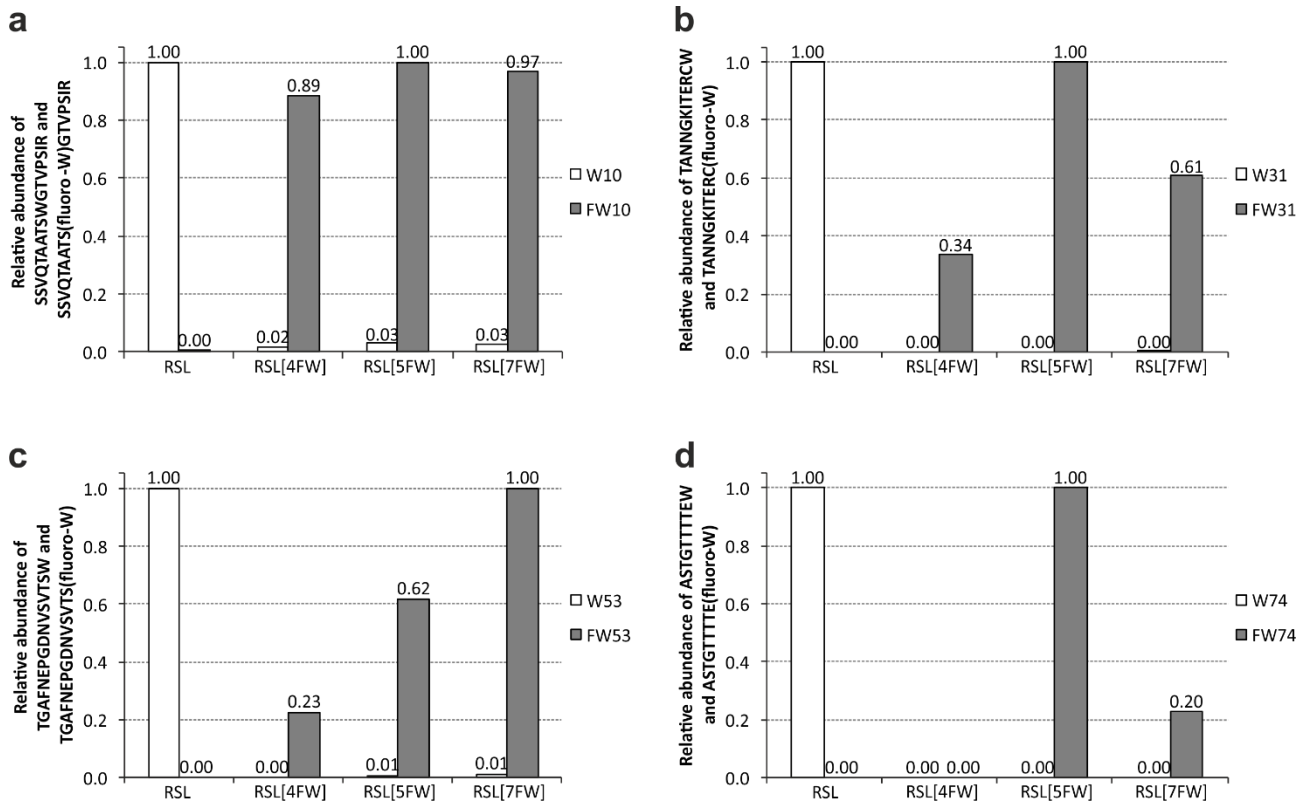

Figure S3. Quantification of the incorporation of FW at individual positions. RSL[4FW], RSL[5FW] and RSL[7FW] and wild-type RSL were subjected to trypsin or chymotrypsin digests to ensure high sequence coverage, the proteolytic digests were separated by nano-HPLC and analyzed by tandem mass spectrometry. Four Trp residues and their respective analogs could be relatively quantified (W10, W31, W53, W74) because they were covered in well ionizing peptides with a single Trp in their sequence. Only minimal residual unlabeled Trp was found in the labeled protein analogs at these four positions (<3% of wild-type).

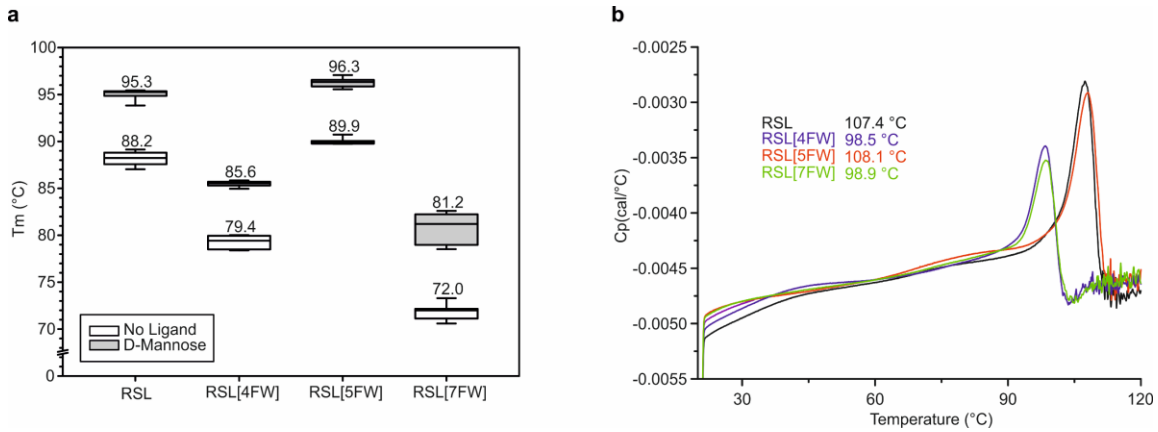

Figure S4. a) Box plot of melting temperatures (T_m) of RSL and fluorinated variants in the presence (grey boxes) or absence (white boxes) of 8 μM D-mannose. The median of the melting temperatures is indicated each. All samples show significant differences in melting temperatures (Student t-test P <0.001). Box plots were calculated from triplicate measurements performed on three different days. b) Melting temperatures (°C; shown as insets) of RSL and fluorinated variants in the presence of the high-affinity ligand αMeFuc as determined by differential scanning calorimetry. Samples were prepared in 20 mM Tris/Cl, 100 mM NaCl, 1 mM αMeFuc, pH 7.5, with concentrations of 1.5 mg mL^{-1} for RSL and RSL[5FW] and 1.0 mg mL^{-1} for RSL[4FW] and RSL[7FW]. Data was acquired on a VP-DSC (MicroCal, Inc., Northampton, MA) and analyzed with the MicroCal Origin software (VP-DSC version). Measurements were performed at 30 psi pressure and a scan rate of 1 °C min^{-1}.

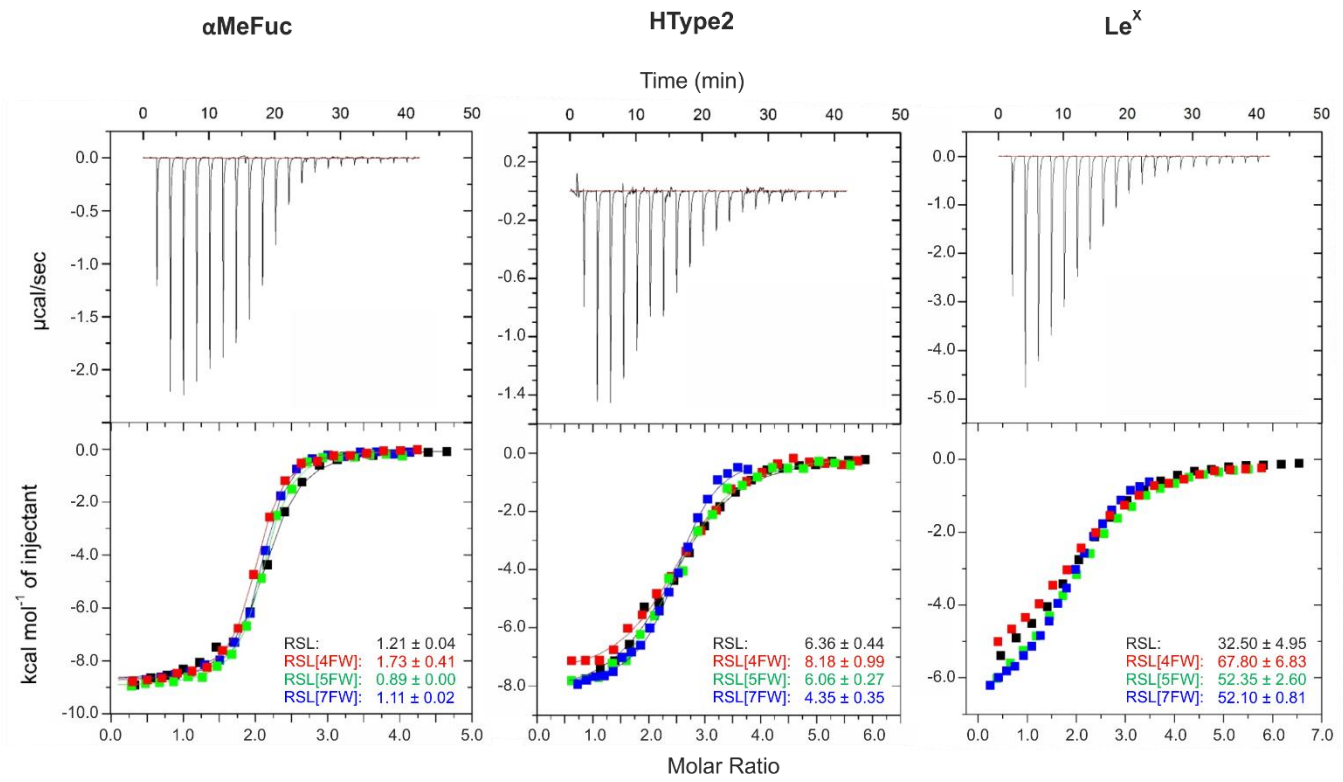

Figure S5. ITC data of the interaction of RSL and the three fluorinated variants with methyl α-L-fucopyranoside (αMeFuc) and the two tetrasaccharides, blood group H type 2 antigen (HType2: Fucα1-2Galβ1-4GlcNAcβ1-3Gal) and lewis X (Le^X: Galβ1-4[Fucα1-3]GlcNAcβ1-3Gal) (bottom panel). Thermograms of the parent protein are shown in the top panel. K_d values (μM; ± standard deviation) for the interaction of RSL and the ligands are shown as insets.

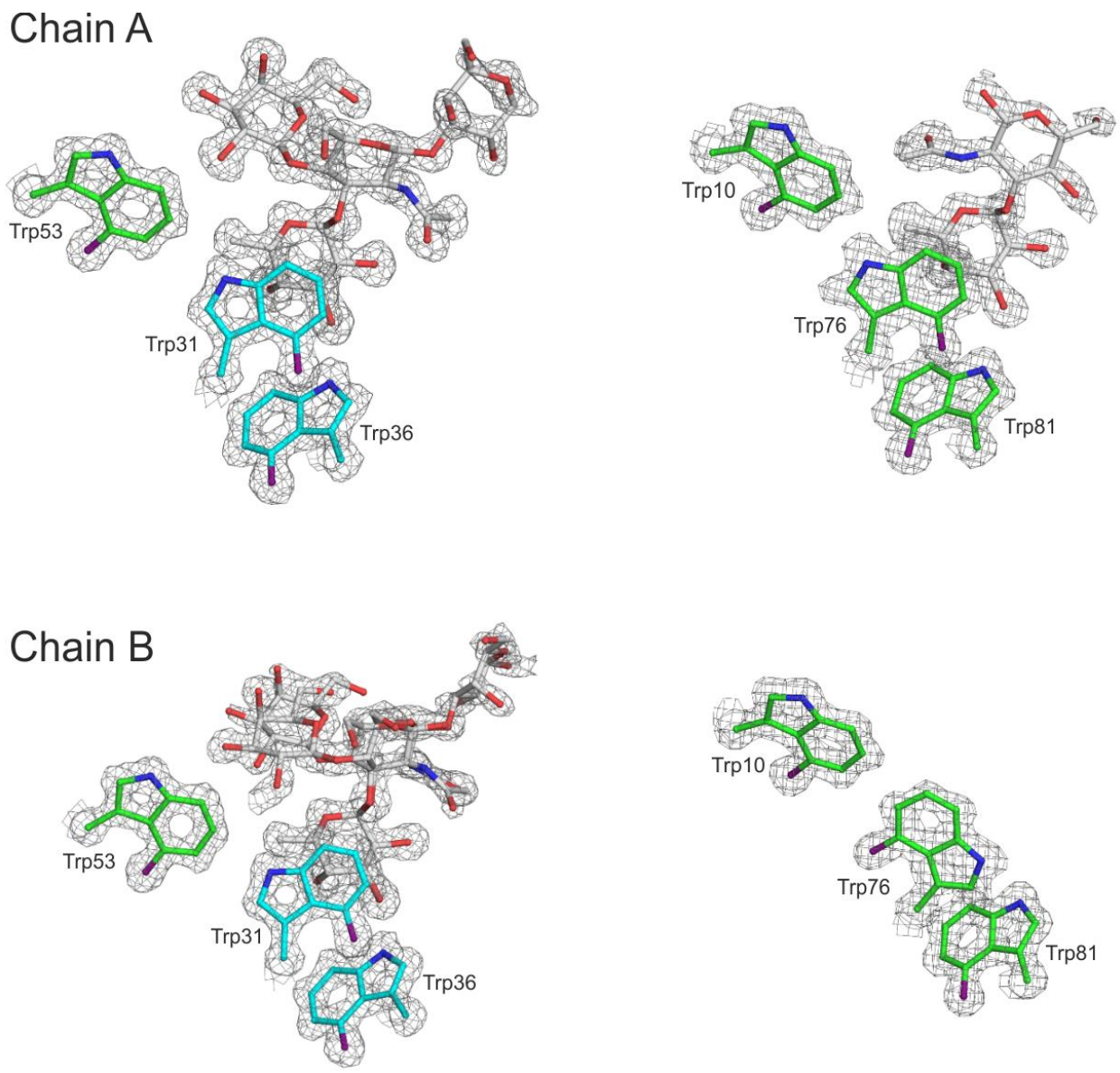

Figure S6. Electron density in the binding sites of the variant lectins RSL[4FW] complexed with the tetrasaccharide Le^x (Galβ1-4[Fucα1-3]GlcNAcβ1-3Gal). The intermonomeric binding site is represented in the left panel and the intramonomeric binding site in the right panel for both chains A and B. Maximum likelihood weighed 2mFo-DFc maps contoured at 1σ corresponding to 0.67 $Å^3$ are presented.

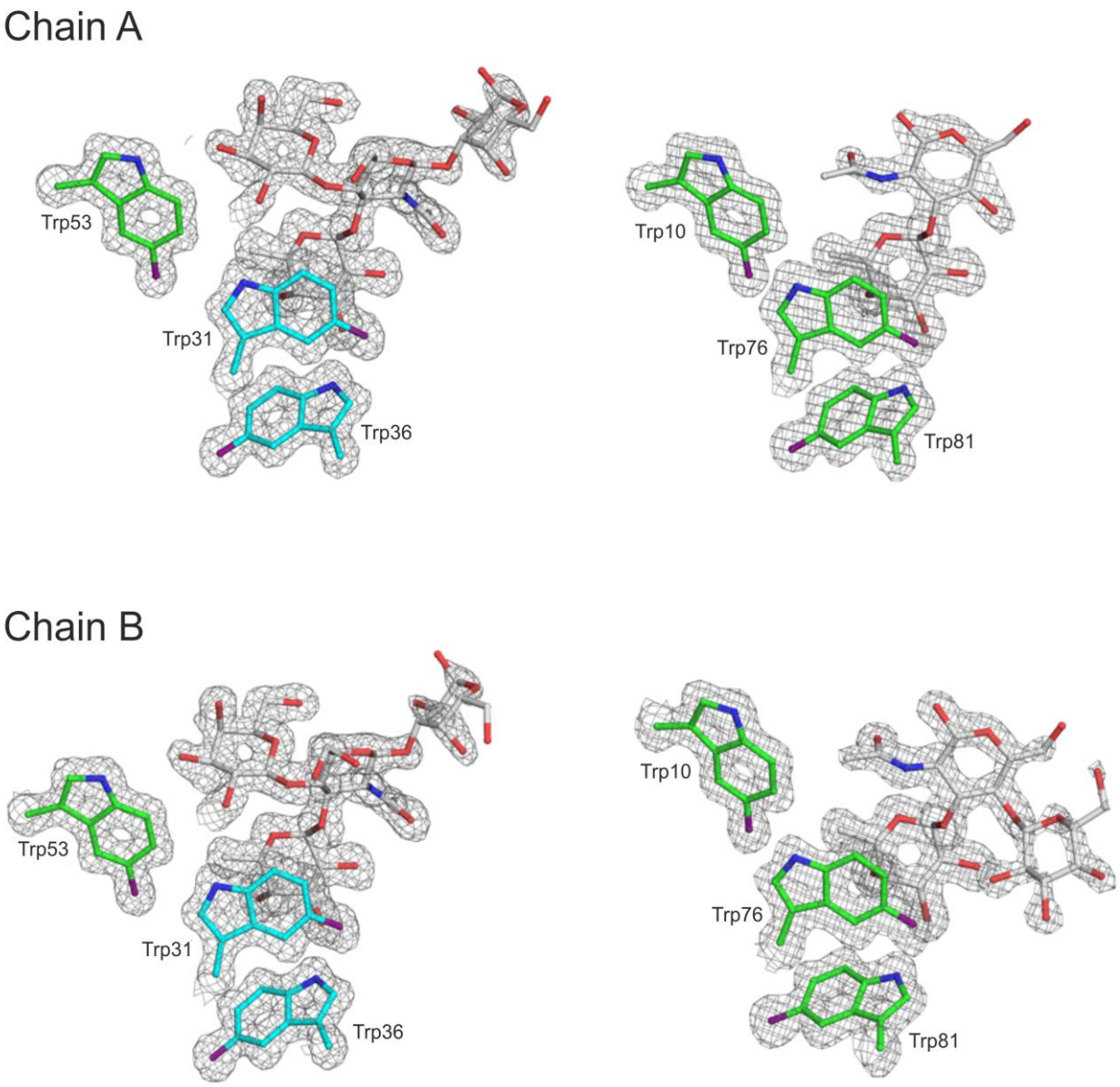

Figure S7. Electron density in the binding sites of the variant lectins RSL[5FW] complexed with the tetrasaccharide Le^x (Galβ1-4[Fucα1-3]GlcNAcβ1-3Gal). The intermonomeric binding site is represented in the left panel and the intramonomeric binding site in the right panel for both chains A and B. Maximum likelihood weighed 2mFo-DFc maps contoured at 1σ corresponding to 0.47 $Å^3$ are presented.

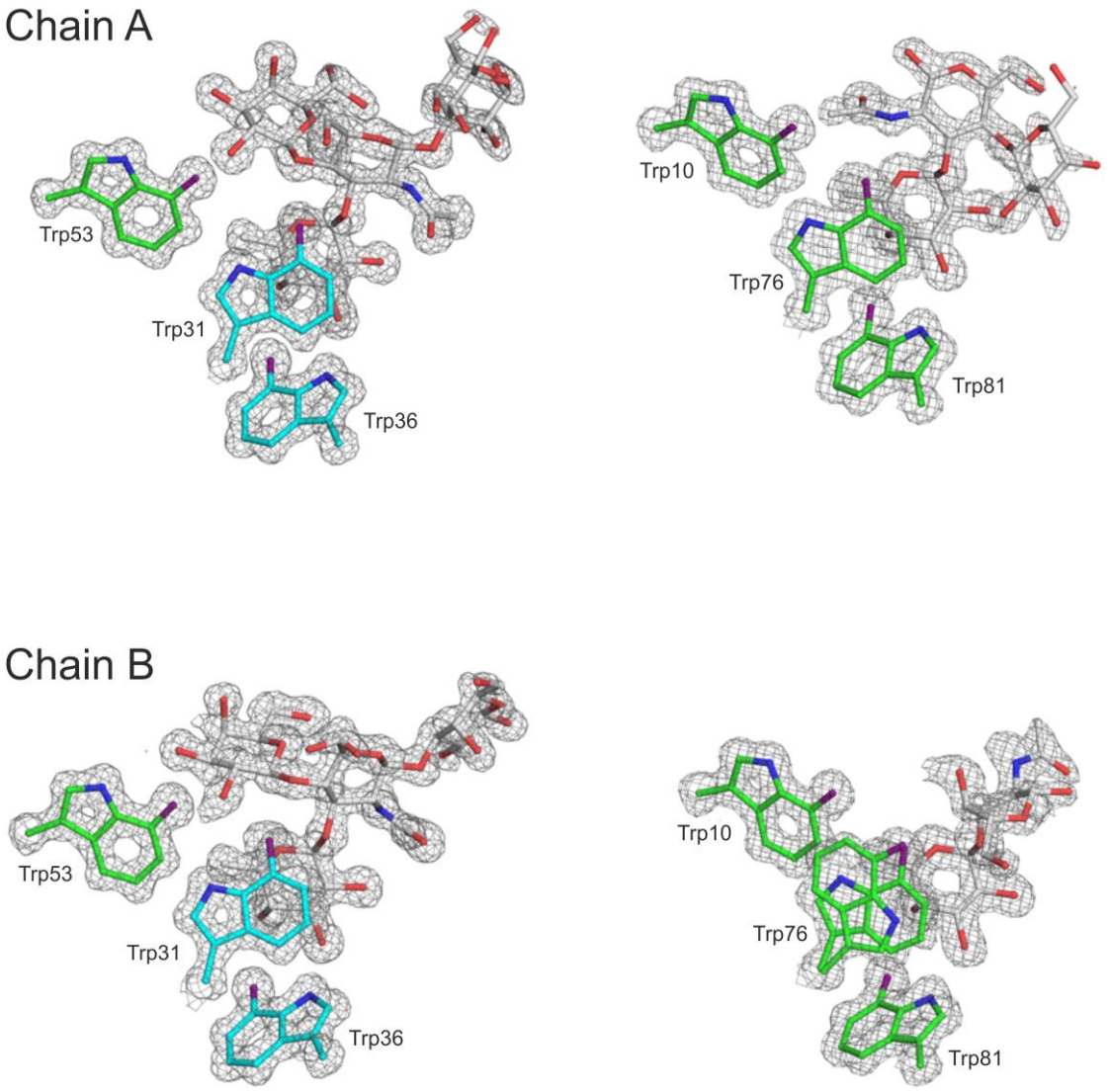

Figure S8. Electron density in the binding sites of the variant lectins RSL[7FW] complexed with the tetrasaccharide Le^x (Galβ1-4[Fucα1-3]GlcNAcβ1-3Gal). The intermonomeric binding site is represented in the left panel and the intramonomeric binding site in the right panel for both chains A and B. Maximum likelihood weighed 2mFo-DFc maps contoured at 1σ corresponding to 0.56 $Å^3$ are presented.

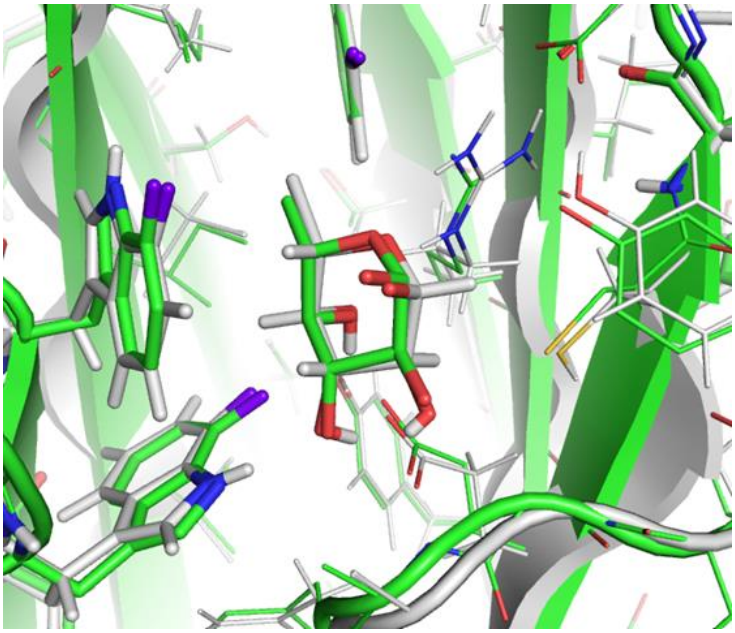

Figure S9. Superimposition of X-ray crystal structure (green) and averaged one from MD calculations in the intermonomeric binding site of fucose in RSL[7FW].

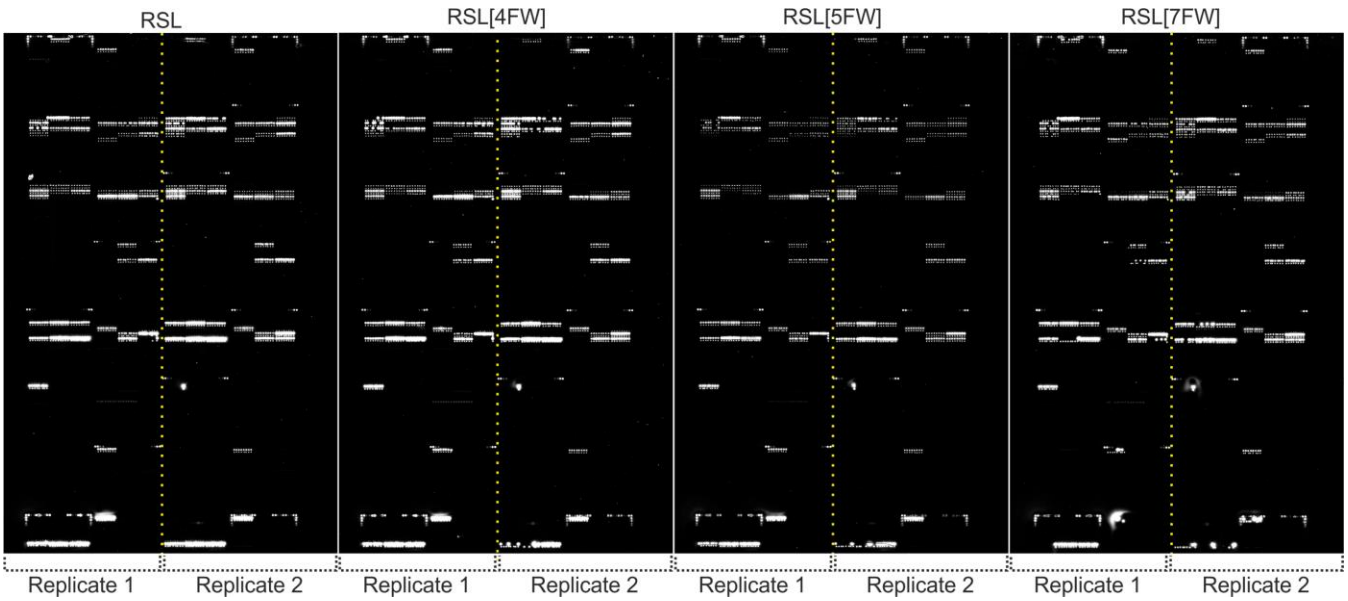

Figure S10. Image analysis of glycan microarrays performed with RSL and the fluorotryptophan-containing variants. Scans of fluorescently labelled lectins are shown. Each microarray is separated in two replicates (separated by a yellow dotted line). For details of the 317 glycans, see Supplemental information of Frederiksen, *et al.*[3].

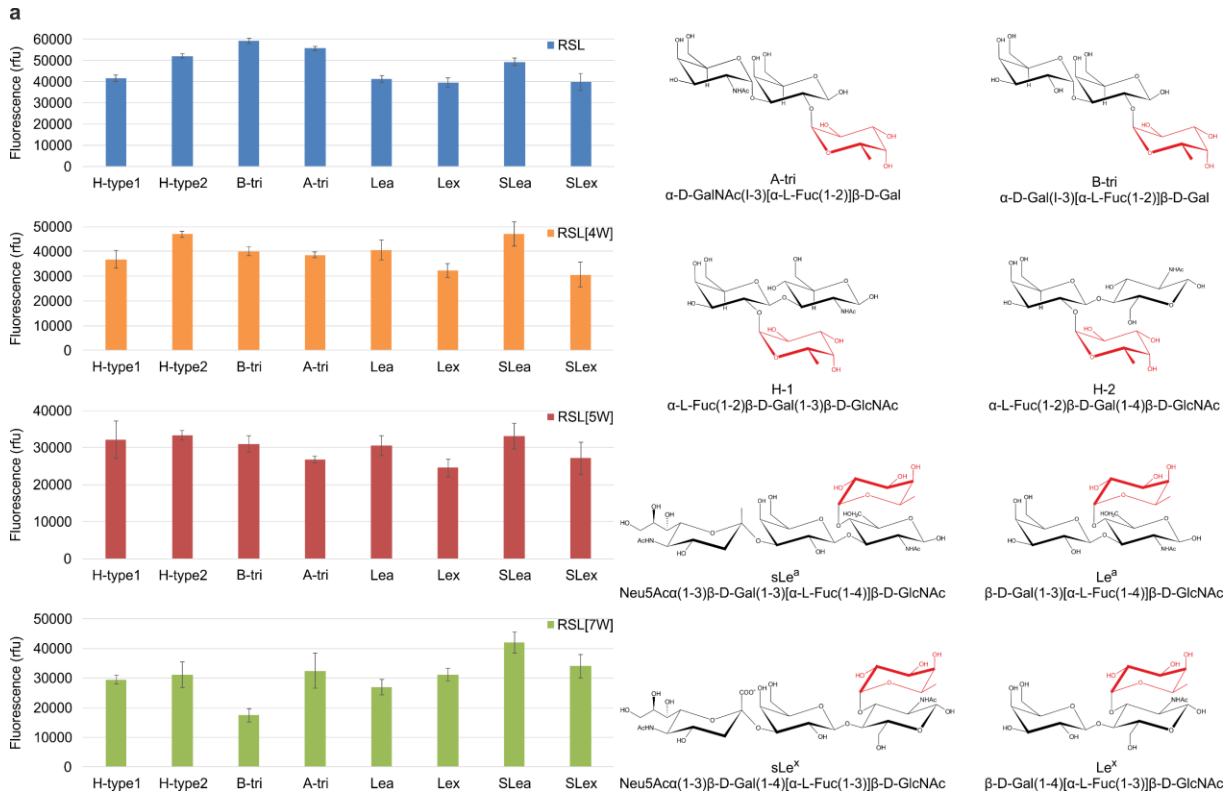

Figure S11. Interaction of the RSL parent a) and the variant proteins RSL[4W] b), RSL[5W] c), and RSL[7W] d) with the eight selected oligosaccharides represented in panel e). Data were collected from the glycan microarray (replicate 2 at lower concentrated glycans), shown in Figure S10

References

(1) Topin, J., Lelimousin, M., Arnaud, J., Audfray, A., Perez, S., Varrot, A., and Imberty, A. (2016) The hidden conformation of Lewis x, a human histo-blood group antigen, is a determinant for recognition by pathogen lectins, *ACS Chem. Biol. 11*, 2011-2020.

(2) Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem. 97*, 10269-10280.

(3) Frederiksen, R. F., Yoshimura, Y., Storgaard, B. G., Paspaliari, D. K., Petersen, B. O., Chen, K., Larsen, T., Duus, J. O., Ingmer, H., Bovin, N. V., Westerlind, U., Blixt, O., Palcic, M. M., and Leisner, J. J. (2015) A diverse range of bacterial and eukaryotic chitinases hydrolyzes the LacNAc (Galbeta1-4GlcNAc) and LacdiNAc (GalNAcbeta1-4GlcNAc) motifs found on vertebrate and insect cells, *J. Biol. Chem. 290*, 5354-5366.