Supplemental Information

Malectin - a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein *N*-glycosylation

Thomas Schallus, Christine Jaeckh, Krisztina Féher, Angelina S. Palma, Yan Liu, Jeremy C. Simpson, Mukram Mackeen, Gunter Stier, Toby J. Gibson, Ten Feizi, Tomas Pieler, and Claudia Muhle-Goll

Supplemental Material and Methods

Sequence analysis of malectin and malectin-like proteins

In the *Xenopus laevis* malectin sequence UniProtKB/TrEMBL accession number Q6INX3 and the human homologue with gene name KIAA0152 and accession Q14165, the SMART server (Letunic et al., 2006) identified a signal peptide using the SignalP algorithm (Emanuelsson et al., 2007) and the transmembrane domain using TMHMM2 (Sonnhammer et al., 1998). Metazoan homologues were retrieved with NCBI Blast (hhtp://www.ncbi.nlm.nih.gov/BLAST/) searching the NR non-redundant database, then aligned with Clustal X (Chenna et al., 2003) and edited to tidy the gaps with SeaView (Galtier et al., 1996).

The homologous domain in plant RLKs was identified with NCBI PSI-BLAST searching the NR non-redundant database. The first iteration used the X. laevis malectin globular domain (residues 29-211). The second iteration was seeded with the set of metazoan sequences (all E-values < e-30). The third iteration included the ciliate malectin-like proteins and retrieved many plant receptor-like kinases (RLKs) with E-values spanning e-9 – e-5. Addition of any plant RLKs at all into a fourth iteration essentially collects a very large number of plant homologues generated by the genome projects (and is complicated by the fact that agricultural plants are polyploids). Reciprocal PSI-BLAST searches with the plant domain will retrieve the animal sequences, but the returned list becomes very large and unwieldy, due to all these plant sequences. Bacterial carbohydrate-binding proteins are retrieved with good E-values in the PSI-BLAST searches but, after it was found that they added noise to the eukaryotic searches, they were excluded. The plant RLKs were first aligned together and then the malectin-like domains excised and aligned to the animal domains using Clustal X profile alignment. A composite alignment is shown in supplementary Figure S1A and indicates a few well-conserved residues shared by plants and animals. These raise the possibility of a partially conserved function. However, the aromatic residues in the glucose disaccharide-binding pocket (Fig. 4 C and D) are not conserved.

Online resources were examined for any hint that the human KIAA0152 gene was associated with inherited disease. Neither OMIM (Hamosh et al., 2005) nor Gene2Disease (Perez-Iratxeta et al., 2002) contained information on the gene. A literature survey with 12q24.31 (the malectin locus) revealed that the lifestyle-affected, often interlinked conditions of obesity and type II diabetes have both been mapped to 12q24 (Mahtani et al., 1996, Wilson et al., 2006), sometimes to 12q24.31 (reviewed in Florez et al., 2003). The 12q24 gene for the transcription factor HNF1alpha is responsible for an early onset variant of diabetes termed MODY3, but is excluded from the NIDDM2 variant (OMIM:601407) also located at 12q24. The NIDDM2 gene has not been identified, suggesting that the causative gene (or genes) is non-obvious, and may possibly be one of the orphan genes within this region of chromosome12.

cDNA library screen and whole mount in situ hybridization

The original *X. laevis* malectin full-length cDNA clone was isolated from an adult pancreas cDNA library (Afelik et al., 2004). Aiming at the identification of new marker genes expressed in the pancreas, 192 clones from this λ ZAP Express phage cDNA library (Stratagene) were randomly picked and plasmids used for antisense RNA generation to study their spatial distribution in *Xenopus laevis* embryos by whole mount in situ hybridization (WMISH). One of them (originally termed p150) was further investigated in the context of pancreas organogenesis. It has been renamed malectin in this study. The spatial distribution of malectin transcripts in *X.laevis* embryos was determined by WMISH following standard procedures (Harland 1991, with modifications according to Hollemann et al., 1998). For this purpose, the original clone obtained from the cDNA library, pBK-CMV-p150, was linearised with BamHI and digoxigenin (Roche) labeled antisense RNA was transcribed with T7 RNA-polymerase. Transcripts were purified using the *illustraTM RNAspin Mini Kit* (GE healthcare). Embryos were cleared in 50% formamide/ 5xSSC/ 1% H₂O₂ before imaging.

Expression constructs for structure determination and interaction studies

The open reading frame of malectin was subcloned by PCR from the full-length cDNA obtained in the pancreatic library screen (Afelik et al. 2004) into the pCS2+ eukaryotic expression vector. The malectin construct (globular segment, AA 27-213) was cloned into a modified pET-24d vector containing a N-terminal His₆-tag fused to a Z-tag removable through cleavage with TEV protease for NMR-related studies. It was expressed in *E. coli* BL21 [DE3] and purified on Ni-NTA-agarose with a subsequent cleavage of the fusion tag by TEV-protease and a second purification step on Ni-NTA-agarose.

To determine whether the recombinant protein was monomeric at the high concentrations used in the NMR experiments, we performed an analytical run on a gel filtration column and included 2% maltose in the buffer. Under these conditions the protein eluted at the expected molecular weight. ¹H-T2 measurements showed that the free protein is also monomeric.

Saturation transfer difference (STD) experiments

All sugars for NMR studies were purchased from Sigma-Aldrich with exception of mannobiose (Man β 1-4Man), which was purchased from Megazyme, and nigerose which was purchased from COSMO BIO CO., LTD.

The spectra were measured using a pulse sequence in which the difference between the on- and offresonance experiments was created by phase cycling (Mayer and Meyer, 1999). Presaturation of the protein NMR signals was performed using a train of selective Gaussian pulse of a duration of 49 ms field-strength of 75 Hz each and separated by a short delay of 1ms.

Oligosaccharides and oligosaccharide probes and microarray analyses

Laminarin oligosaccharides, di-, tri- and tetrasaccharides were from Dextra (Reading, UK), penta and hexasaccharides (Megazyme,Wicklow, Ireland), and heptasaccharide (Seikagaku America, Falmouth, MA). Nigerose was from Wako Chemicals (Neuss, Germany). Cellobiose, kojibiose and maltooligosaccharides, di- to heptasaccharides were from Sigma. Tri- to hexasaccharides of cellulose were separated by gel filtration chromatography from a cellooligosaccharide mixture (Vlabs purchased via Dextra). Partial depolymerizations of pustulan (Calbiochem) and dextran (Pharmacia) were performed by acetolysis and acid hydrolysis, respectively, and their fragments, di- to heptasaccharides, were fractionated by gel filtration (Palma et al., 2006). The molecular masses of the main components of oligosaccharide fractions from gel filtration were corroborated by MALDI-MS. The high mannose N-glycan, Man₇GlcNAc₂ was a gift from Dr VE Piskarev (Moscow) and Man₉GlcNAc₂ was from Dextra. Glc₃Man₇(D1)GlcNAc was recovered from Glc₃Man₇(D1)GlcNAc₂ that was prepared for NMR residual dipolar coupling measurements using filamentous bacteriophages (Pf1). Glc₃Man₇(D1)GlcNAc₂ was isolated from the glycoprotein HIV-IIIB gp120 secreted by CHO cells treated with the α -glucosidase inhibitor N-butyl deoxynojirimycin, as described previously (Petrescu et al., 1997). The loss of one GlcNAc unit from the original material was attributed to residual endoglycosidase activity present in the bacteriophages. The Glc₂Man₇(D1)GlcNAc was obtained from the Glc₃Man₇(D1)GlcNAc isomer [(one of three isomers D1, D2 and D3 (Petrescu et al., 1997)] by digestion with α -glucosidase I at 37°C until completion (Alonzi et al., 2007). Glc₁Man₉GlcNAc₂ was isolated from hen egg yolk IgY using a similar procedure to that previously reported by Ohta et al. (1991). These oligosaccharides and the glucooligosaccharides were converted to oxime-linked neoglycolipids (AO-NGLs) and their molecular masses were corroborated by MALDI-MS as described (Liu et al., 2007). The molecular ions for the AO-NGLs for Glc₃Man₇(D1)GlcNAc, and Glc₁Man₉GlcNAc₂ were exclusively mz 2559.3 and 2762.3, respectively; and for Glc₂Man₇(D1)GlcNAc mz 2396.8 with minor ions as for the parent glycan, Glc₃Man₇(D1)GlcNAc.

Microarrays of 335 lipid-linked oligosaccharide probes, NGL and glycolipids (supplemental Table S6), were robotically generated as described (Palma et al., 2006). Unless otherwise stated the NGLs were prepared by reductive amination (Chai et al., 2003).

Microarray analyses with his-tagged malectin were performed essentially as described (Palma et al., 2006), except that the protein was pre-complexed with mouse monoclonal anti-poly-histidine (Ab1) and biotinylated anti-mouse IgG antibodies (Ab2) (both from Sigma) in a ratio of 1:3:3 (by weight). In brief, the malectin-antibody complexes were prepared by pre-incubating Ab1 with Ab2 for 15 min at ambient temperature, followed by addition of malectin and incubation for a further 15 min. The malectin-antibody complexes were diluted in casein (Pierce) containing 1% (w/v) bovine serum albumin (Sigma) and 10 mM CaCl₂, to give a final malectin concentration of 20, 5, 1 or 0.5 μ g/ml. Microarray analysis with biotinylated Concanavalin A (ConA) from Vector Laboratories (at 10 μ g/ml) was performed as described (Liu et al., 2007). Binding was detected with Alexa Fluor-647-labelled streptavidin (Molecular Probes) and imaging was as described (Palma et al., 2006). Data analysis was performed with a dedicated software (Mark P. Stoll of the Glycosciences Laboratory, unpublished).

Supplemental References

Afelik, S., Chen, Y., and Pieler, T. (2004). Pancreatic protein disulfide isomerase (XPDIp) is an early marker for the exocrine lineage of the developing pancreas in Xenopus laevis embryos. Gene Expr Patterns 4, 71-76.

Alonzi, D.S., Neville, D.C., Lachmann, R.H., Dwek, R.A., and Butters, T.D. (2007). Glucosylated free oligosaccharides are biomarkers of endoplasmic reticulum α -glucosidase inhibition. Biochem. J. 409, 571-580.

Chai, W., Stoll, M.S., Galustian, C., Lawson, A.M. and Feizi, T. (2003). Neoglycolipid technology - deciphering information content of glycome. Methods Enzymol. *362*, 160-195.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. *31*, 3497-3500.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2, 953-971.

Florez, J.C., Hirschhorn, J., and Altshuler, D. (2003). The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annu Rev Genomics Hum Genet. *4*, 257-291.

Galtier, N., Gouy, M., and Gautier, C. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. *12*, 543-548.

Harland, R.M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. Methods Cell. Biol. *36*, 685-695.

Hollemann, T., Chen, Y., Grunz, H. and Pieler, T. (1998) Regionalized metabolic activity establishes boundaries of retinoic acid signalling. EMBO J. *17*, 7361-7372.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. *33* (Database issue), D514-517.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. *34*, D257-260.

Liu, Y., Feizi, T., Campanero-Rhodes, M.A., Childs, R.A., Zhang, Y., Mulloy, B., Evans, P.G., Osborn, H.M., Otto, D., Crocker, P.R., and Chai, W. (2007). Neoglycolipid probes prepared via oxime ligation for microarray analysis of oligosaccharide-protein interactions. Chem Biol. *14*, 847-859.

Mahtani, M.M., Widén, E., Lehto, M., Thomas, J., McCarthy, M., Brayer, J., Bryant, B., Chan, G., Daly, M., Forsblom, C., Kanninen, T., Kirby, A., Kruglyak, L., Munnelly, K., Parkkonen, M., Reeve-Daly, M.P., Weaver, A., Brettin. T., Duyk, G., Lander, E.S., and Groop, L.C. (1996). Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. Nat Genet. *14*, 90-94.

Mayer, M., and Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectra. Angew. Chem. Int. Ed. *38*, 1784-1788.

Niehrs, C., Steinbeisser, H., and DeRobertis, E.M. (1994) Mesodermal patterning by a gradient of the vertebrate homeobox gene goosecoid. Science 263, 817-820.

Ohta, M., Hamako, J., Yamamoto, S., Hatta, H., Kim, M., Yamamoto, T., Oka, S., Mizuochi, T., and Matsuura, F. (1991) Structures of asparagine-linked oligosaccharides from hen egg-yolk antibody (IgY). Occurence of unusual glucosylated oligo-mannose type oligosaccharides in a mature glycoprotein. Glycoconj. J. 8, 400-413.

Palma, A.S., Feizi, T., Zhang, Y., Stoll, M.S., Lawson, A.M., Díaz-Rodríguez, E., Campanero-Rhodes, M.A., Costa, J., Gordon, S., Brown, G.D., and Chai, W. (2006). Ligands for the beta-glucan receptor, Dectin-1, assigned using "designer" microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. J. Biol. Chem. 281, 5771-5779.

Perez-Iratxeta, C., Bork, P., and Andrade, M.A. (2002). Association of genes to genetically inherited diseases using data mining. Nat Genet. *31*, 316-319.

Petrescu, A.J., Butters, T.D., Reinkensmeier, G., Petrescu, S., Platt, F.M., Dwek, R.A., and Wormald, M.R. (1997). The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. EMBO J. *16*, 4302-4310.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. *6*, 175-182.

Wilson, S.G., Adam, G., Langdown, M., Reneland, R., Braun, A., Andrew, T., Surdulescu, G.L., Norberg, M., Dudbridge, F., Reed, P.W., Sambrook, P.N., Kleyn, P.W., and Spector, T.D. (2006). Linkage and potential association of obesity-related phenotypes with two genes on chromosome 12q24 in a female dizygous twin cohort. Eur J Hum Genet. *14*, 340-348.

Supplemental results Fig S1

А										
		>. →			60 + ^{>}	50	\sim	× .×	100	
OCTIVO VINT			30: :		DUGD CD				. * 100	
Q61NX3_XENLA	A ADKVIWAV	NAGGESHVDVHG	THYRKDPLEG		DVGRASD		LRSNPE	-DOVLYOTERINED	-SFGYDIPIKE-	EGETVLVLKF
Q14105_HUMAN	DECUTWAN	NAGGEARVDVHG	TUPPYDDIEC		DUCDACDI	CMELP	I DCMDE	DOT NOTERVIER	TEGILVPIKE-	ECDYVIVIVE
OPCT01 MOUCH	P DECUTWAN	NACCEAUVDUUC			DUCDACD			DOTIVOTERVINE	TECYEUDURE	- FODYVIVIVIVE
OBCIJI_MOUSI		NACCDENTDLNC			-VOTACEN			DEVI VENERVUMT	TROYDUPSDO	DODYALTMER
OSDCB4 SCH.I	HGEVVWAV	NCCCPKHTDSHC	VEVENDELN		TGTASD	CRSEAT		DYTLYOTERVHND		DGEVILTIKE
0950G1 CAREI	KNRVVAAV	NCCCPDALCAYC	TEVSADTSE		DOVSSOF	GMOVSE	NNAEVD	DMETVOTERWSKE	-SESVDUPTSE-	DEEVVIILKE
0610S1 CAEBE	RDRVVAAT	NCGGPEALGSYG	TTYAADYHD		-EGOASD		NNAEND	DVETVOTERWSKE	SEDVEVEVO	DGEYVITLKE
07YY74 CRYPI	HAEVTYAU	NCGGPRYESKSEN	TLYEEDNGYN		GGISTDS	GKOLST	FPYVE	DEVYLSERVETD	BTLOYMUNUNKI	TPOVETTVLKE
09FXF2 ARATH	RYSSCLHW	NCGG-SDMYVKEKK	KELYEGDGNVE-GGA	AKYFLKPDANWGES	STODEMDI	NNFONTRET	FVPASN	OSDLYKSARTAPV	SLTYFHACLE	NGNYTTNLDF
022579 ARATH	I RYSSCLHV	NCGG-SDMYVKEKK	KELYEGDGNVE-GGA	AKYFLKPDANWGFS	STGDFMDI	NNFONTRFT	FVPASN	-OSDLYKSARIAPV	-SLTYFHACLE-	-NGNYTINLDF
Q8RWZ9 ARATH	I RYSSCLHV	NCGG-SDMYVKEKK	KELYEGDGNVE-GGA	AKYFLK <mark>P</mark> DANWGFS	STGDFMDI	NNFONTRFTN	FVPASN	-OSDLYKSARIAPV	-SLTYFHACLE-	-NGNYTINLDF
Q9FXE9 ARATH	I KYQRF <mark>LH</mark> I	NCGG-EEVSIRNSL	KITYQTDNSRQ-TNA	ASNQQFDYWGVS	NTGDFTDI	NSDHDEYYTS	MLTLSGD	-YPDLYKTARRSAL	-SLVYYAFCLE-	-NGNYNVKLHF
Q9FXF0_ARATH	I NYSRSLHI	NCGG-PDVTIENSRC	RFLYEGDNYGLTGSA	INYYRKNWGYS	NTGDFMDI	ATTEDTYT	SS-ESAVSAK	-Y <mark>PDLYQ</mark> NARRSPL	-SLAYYAFCFE-	-NGSYNVKLHF
Q9C6G4_ARATH	I LDSRSLHI	NCGG-PDVTIENSRO	RFLYEGDNYGLTGSA	TNYYRKNWGYS	NTGDFMDI	D-AITEDTYT	/SS-ESAVSAK	-Y <mark>PDLYQ</mark> NARRSPL	-SLAYYAFCFE-	-NGSYNVKLHF
Q9FXF1_ARATH	H -YNRSLHI	NCGG-PDVTIENSRO	RFLYEGDNYGLTGSA	TNYYGKNWGFS	NTGDFMDI	D-AITEDTYT	/SS-ES <mark>A</mark> VSAK	-Y <mark>PDLYQ</mark> NARRSPL	-SLAYFAICFE-	-NGSYNVKLHF
Q9LH71_ARATH	I SAFY <mark>GLH</mark> I	NCGG-NEITSNE	-TKYDADTWDT	PGYYDSK-NGWVSS	NTGNFLDI	DRTNNGKSK	-SNSSELKITNSS	IDFRLYTQARLSAI	-SLTYQALCLG-	-KGNYTVNLHF
Q9MAG1_ARATH	I IAFNALHI	NCGG-DEMSING	-TIYESDKYDRL	ESWYESR-NGWFSN	INVGVFVDI	-KHVPERVT]	ESNSSELNV	V DF<mark>GLY</mark>TQARISA I	-SLTYYALCLE-	-NGNYNVNLHF
Q9LPG0_ARATH	I AHHSSLFI	NCGG-NRLKVDK	-DEYADDLNKRGA	STFSSVS-ERWGYS	SSGAWLGI	IDGATYLI	TDTFNLINE	ST <mark>PEYY</mark> KTARLASQ	-SLKYYGLCMR-	-RGSYKVQLYF
Q9LPF9_ARATH	I <mark>A</mark> KQSS <mark>LFI</mark>	NCGG-SRLKIGK	-D <mark>TY</mark> TDDLNSRGQ	STFSSVS-ERWGYS	SS <mark>GVW</mark> LGH	(EDA <mark>GYL</mark>	TDRFNLING	ST <mark>PEYY</mark> KTARLSPQ	-SLKYYGLCLR-	-RGSYKLQLHF
Q6ZKW6_ORYS	A <mark>P</mark> EYYS FAV	DCGSNKSTRGSDN	-TLYETDAQNIGA	A <mark>SYYVS</mark> DNARW <mark>G</mark> VS	SVGKFNE7	ASNGSYAIN	(SP-QQFQSAL	-NSELFQTARMSPS	-SLRYYGIGLE-	
Q7F106_ORYSA	A <mark>P</mark> EYYS FAV	DCGSNKSMKGSDN	-TIYEVDAANLGV	ASYYVTRNTRWGVS	NVGIFNDA	ASSRNY <mark>VI</mark> N	ISS-QQFQNTL	-DSELFQTARMSPS	-SLRYYGLGLE-	NGNYSVKLQF
Q6ZKW7_ORYSA	A PKYYSFAV	DCGSNRSIRVSDN	-TMYELDSTNLGD	SSYYVTSQTRWGVS	NVGKLFQ7	A <mark>P</mark> NDSKIIF	ISG-EKIQNAV	-DSELFQTARMSPS	-SLRYYGLGLE-	-NGNYTVLLKF
Q6ATA6_ORYS7	A PRYYSFAV	DCGSNGSTRGSDD	-TIYEADPTNLGA	ATYYVTGQTRWGVS	SVGNYFQF	REDAKNII	SS-QNFQNVV	-HSELFQTARMSPS	-SLRYYGLGLE-	-NGNYTVLLQF
Q7XTP6_ORYS	A PQSSSFAV	DCGSNRLISASDN	-LRYOTDDASLGP	ASYSVTGAPTWGVS	NVGKFVDF	APNGSYIIY	SS-RQFQNTL	-DSELFQTSRMSPS	-SLRYYGIGLE-	-NGNYTVTLQF
Q7XTP4_ORYSA	POSSSFAU	NCGSNRF1SGSDN	-LRYETDDVNLQA	ASYNVIGAPTWGVE	NVGKEMDA	APNGNYIII	SS-ROFOHTL	-DSELFLISRMSPS	-SLRYYGIGLE-	-NGNYTVTLQF
Q/XTP5_ORYSA	PRTASFAV	NCGG-PLISGSDN	-LRYOSDEVNLGD	ASYYITGEPTWGVS	TVGREMDA	ASNGGYTTF	KSS-ROFONTL	-DSEMFONTRISAS	-SLRYYGIGLE-	-NGNYTVTLQF
Q/XTR9_ORYSA	PRSASFAV	DCGGSRTISGSDN	-AMYQADNANLGA	ASYYVAGTPTWGVS	TTGREMDI	PNGSYIII	SS-ROFDRTL	-DSGLFQTARMSPS	-SLRYYGIGLE-	-NGNYTVTLQF
QSQL86_ORISA	A PEISSPAV		FILESDGANL-QG	ASIIVIRPVRWGVE	NTGKF IM	-EPNRSIIII	TS-NOFNATL	-DSELFQTARTSPS	-SLRIIGIGLK-	NGRIIVALKE
Q95GT9_ARATI	I GLISDESI	NCGG PEARSVIG	ALFEREDEDF	ABFFVSAGQRWAAS	SVGLEAGE	SINNIIIA	SQ-SQFVNTL	DOELFOGARLOAS	GI DANGI CI E	NCONCUMUCE
Q9SGUI_ARATI	I GVIENFEV	NCGG-RDIRSSSG	-ALIERDEGALGP	ATFFVSKTQRWAVE	NAGPE LOS	NSNQIIAI	SA-TOPANTS	-DSELFUSARLSAS	-STKIIGTE-	-NGGISVIVQF
	-		<u> </u>	~ _		<u> </u>	<u> </u>			
	*. 1	20 .: *.:	130	. 150	. 16	0:	170 18	šo 🛄 . : .	200	
Q6INX3_XENLA	A AEVYFAQS	QQKVFD	RVN-GHTVVKDLDIF	DRVGHSTAHDEI	IPISIKK	KLSVQGE	STFTGKLSVEF	V <mark>KGYY</mark> D <mark>NP</mark> KVCALF	IMKGTADDV	
Q14165_HUMAN	I <mark>AEV</mark> YFAQS	QQKVFD	RLN-GHVVVKDLDIF	DR <mark>VG</mark> HST <mark>AH</mark> DEI	IPMSIRK	KLSVQGE	STFTGKLYIEF	V <mark>KGYY</mark> D <mark>NP</mark> KVCALY	IMAGTVDDV	
Q5FVQ4_RA1	r <mark>AEV</mark> Y <mark>F</mark> AQS	QQKVFD	RLN-GHVVVKDLDIF	DR <mark>VG</mark> HST <mark>AH</mark> DEI	IPMSIRK	KLSVQGEV	/STFTGKLYIEF	V <mark>KGYY</mark> D <mark>NP</mark> KVCALY	IMAGTVDDV	
Q8CI91_MOUSE	E AEVYFAQS	QQKVFD	RLN-GHVVVKDLDIF	DR <mark>VG</mark> HSTAHDEI	IPMSIRK	KLSVRGE	STFTGKLYIEF	V <mark>KGYY</mark> D <mark>NP</mark> KVCALY	ILAGTVDDV	
Q8SZL6_DROME	E CEVYFDA <mark>F</mark>	QKKVFD	LLNRKHTVVRQLDIY	NE <mark>VG</mark> R <mark>G</mark> SAHDEI	VYFKINN	GRLNYEGE	SDVRNGR-LRLDF	I <mark>KGAL</mark> D <mark>NP</mark> KINAFA	LLK <mark>G</mark> DISQ <mark>V</mark>	
Q5DCB4_SCHJA	A S <mark>EVWF</mark> TEI	YQKVFD	YVQHTIPIVQNLDIF	GQVGFATAYDFH	LQFKIK-I	DRVLFIGDS-P	AA <mark>PIDADH-FTVDF</mark>	I <mark>KTQFDNP</mark> KINAIV	VTK <mark>G</mark> TIEDV	
Q95QG1_CAEEI	J SEVYFQKA	GEKIFNI	RIN-SHLAVKNLDIF	DAAGG-RGFAHDVY	I PVVIKGS	SSISVSGHS	SREYK <mark>GKVIIEL</mark> :	S <mark>KGPH</mark> DNPKLNGYA	ILRGTVEDL	
Q61QS1_CAEBI	R SEVYFORV	QEKVFN	RIN-SYTAVKNLDIY	EAAGG-RGYAHDIY	IPLVIKDI	KKISVNGHF	KKDYKGKIVIEF	A <mark>KGPH</mark> DNPKVNGFA	VLRGKVEDL	
Q7YY74_CRYP	7 SEIHFKE	GKKVFS]	AVG-NVIFKOSFDIY	KEVGFGVPMEEN	IECTFDGI	ENISLNGMNI	QGYSKEEKLLILA	MFKQEDNPKINAIV	VYKGSKDEI	
Q9FXFZ_ARATE	A AFIRFTNL	ENYNRLGRRLFD	YIQ-EKLVAKDENIM	DEAKGAQTPIIKPI	T-AYVINI	IFLTIRLS	WAGK-GTTRIP	TRGVYG-PIISAIS	I VSDSKPCE	
0225/9_ARATE	A AFIRFTNL	ENINRLGRRLFD	YIQ-EKLVARDENIM	DEAKGAQTPIIKPI	T-AYVINF	IFLTIRLS	WAGK-GTTRIP	TRGVIG-PIISAIS	I VSDSKPCE	
QORWZ9_ARATI	I METOFOD		TIQ-EKLVARDENIM	UEANGAQTPIIKPI	AIVENI	IF LTIRLS	WAGK GTTRIP	VRGVIG-PIISAIS		
OGEVED ARATI		E	YTO-CKLTWEDESTP	FFANGTUKEVTRET	N-TTUTN			OPCVVC-SLTSAVS	UCDCCFCFC	
O9C6C4 ARATI	ARTORSDU	P	VTO-CKLTWEDESTR	FRANGTHERVIRE			WACK CTMTTP	ORGYVG-SLISAVS	VCDSSESEC	
OPEXEL ARATE	ARTORSDE	E-PESRIAKRVENT	YVO-CKLTWEDESTR	EFANGTHERVIKE	N-TTVTD	JTLETELY	WACK-CTTTP	KRONVG-SLISATS	VCPTPVOTH	
O9LH71 ARATE	ARTMENER	NMYSNLGRRYFDI	YVO-GKREVKDENTV	DEAKGVGKAVVKK	P-VMVTNO	KLETBLO	WAGK-GTOATP	VRGVYG-PLISAVS	VDPDFTPPK	
09MAG1 ARATH	ARTMENGN	INNYOSLGRREEDI	YTO-BKLEVKDENTA	KEAKDVGNVVTKT	P-VETKDO	KLETRLY	WAGE-GTTVTP	KERVYG-PLISATS	VDSSVNPSP	
09LPG0 ARATH	AETMESNE	OTYSSLGRBLEDI	YVO-GILLERDENTA	ORAGGVGKPFLROV	DEVOVNG	TLETHLK	WTGK-GTNVTP	TROVYG-PLISATT	VTPNFKVDT	
09LPF9 ARATH	AETMESNE	OTENSLGRRIED	YVO-GNLLERDENIA	ERAGGVGKPFIROI	DGVOVNGS	STLEIHLO	WTGK-GTNVIP	TRGVYG-PLISAIT	TTPNFKVDT	
Q6ZKW6_ORYSA	A AEFVYPNS	LTSNSIGRRVFD	YVQ-GELKEKNFNIR	KMAGGKSLIAVNKF	TATVSKI	IFLEIHLF	WAGK-GTCCIP	TQGHYG-PTISALS	VTPNFIPTV	
Q7F106_ORYSA	A AEFAYPDS	KTWE <mark>STGRRIFD</mark>	YVQ-GVLKE <mark>KNFDI</mark> R	KA <mark>VGG</mark> KSFT <mark>AV</mark> NKI	YNTIVSKN	IFLEIHLF	WAGK-GTCCIP	TQGYYG-PMISALS	VTP	
Q6ZKW7_ORYSA	A AELGFPDI	PTWQSLGRRFFD	YIQ-GELKEKDFNIR	KM <mark>AGG</mark> KSFT <mark>AV</mark> YKS	YTTTVSKI	IFLEIHLF	<mark>WAGK-GT</mark> CCIP	IQGYYG-PLISALS	IT <mark>P</mark> NFS <mark>P</mark> TV	
Q6ATA6_ORYS	A AEFAFPDS	QTWLSLGRRVFD	YIQ-GALKE <mark>KDFDI</mark> K	KT <mark>AGG</mark> KSFR <mark>VV</mark> NRS	FMVTVSKI	IF <mark>LEI</mark> HLF	WAGK-GTDAIP	I <mark>KGYYG-P</mark> MISALR	VT <mark>P</mark> NFT <mark>P</mark> TV	
Q7XTP6_ORYS	A AEFGIEDI	QTWKSLGRRVFD]	YLQ-GERQE <mark>KNFDI</mark> R	KA <mark>AG</mark> DKSYT <mark>VV</mark> KKS	YKVPVTK1	IFLEIHLF	WAGK-GTCCIP	GQGYYG-PTISALS	VT <mark>P</mark> ADFT <mark>PT</mark>	
Q7XTP4_ORYS	A AEFGIEDI	QSWKSLGRRVFD	YVQ-GERKE <mark>KNFDI</mark> R	KT <mark>AG</mark> DKSYT <mark>VV</mark> KKÇ	YKVPVTK1	IFLEIHLF	WAGK-GTCCIP	TQGYYG-PTISALS	VI <mark>P</mark> ADFT <mark>PT</mark>	
Q7XTP5_ORYSA	A AEFGFEDI	QSWKSLGRRVFD	YLQ-GERKEQNFDIR	ka <mark>ag</mark> dksyt <mark>vv</mark> krs	YKVPVTKI	IFVEIHLF	WAGK-GTCCIP	TQDNYG-PSISALS	LI <mark>P</mark> ADFT <mark>P</mark> T	
Q7XTR9_ORYS	A AEVDFPDV	QSWRSRGRRIFD	YIQ-GERKEQNFDIR	KA <mark>AGG</mark> KSFT <mark>VV</mark> KKÇ	YVVPVTKI	IFLEIHLF	WAGK-GTCCIPI	HQGYYG-PAISALS	ATP-NFIPT	
Q5QL86_ORYSA	A AEI-FPDG	QIWQSMGRRIFD]	YIQ-GERKEQDFDIK	KY <mark>A</mark> NEKSNT <mark>PV</mark> ERQ	YFT D VT NI	IFMEIHLF	WAGK-GTCCIP	TL <mark>GFYG-P</mark> SISALS	VSFSGDPGL	
Q9SGT9_ARATH	H AEIQILGS	TSTTWKGLGRRRFD]	YVQ-GRLVEKDFDVR	RT <mark>AG</mark> DSTVRAVQRV	YKANVSEN	HLEVHLF	WAGK-GTCCIP	IQGAYG-PLISAVS	ATPDFTPTV	
Q9SGU1_ARATH	I AEIQIQGS	NTWKSLGRRIFD	YVQ-GKLVEKDFDMQ	KAANGSSIRVIQRV	YKANVSEN	IYLEVHLF	WAGK-GTCCIP	AQGTYG-PLVSAIS	ATPDFIPTV	

В



Figure S1: Sequence alignment of animal malectin proteins vs. *Arabidopsis thaliana* and *Oryza sativa* **RLK** proteins. (A) Alignment of representative arabidopsis and rice RLK-malectin-like domains with the malectin core domain of *X. laevis*, humans, rat and mouse. Sequences are labelled with UniProtKB/TrEMBL accession numbers. (XENLA = *Xenopus laevis*, DROME = *Drosophila melanogaster*, SCHJA = *Schistosoma japonicum*, CAEEL = *Caenorhabditis elegans*, CRYPV =

Cryptospridium parvum, ARATH = *Arabidopsis thaliana*, ORYSA = *Oryza sativa*). The aromatic residues and the aspartate that in *X. laevis* mediate interactions with the glucose residues (Fig.4D) are marked by red crosses, and are not conserved in plants. (B) Domain topologies of plant and animal proteins that contain the malectin core domain. The two topologies among plant RLKs are shown. Labels: SP - signal peptide; TM- transmembrane helix; LRR - leucine rich repeat: ST-Kinase – serine/threonine receptor-like kinase.



Figure S2: Malectin binding to glucose disaccharides studied by isothermal titration calorimetry. Kojibiose (A), nigerose (B) and maltose (C). The raw data are shown in the upper panel, and the integrated heat data, corrected for dilution, are shown in the lower panel.

ITC measurements were carried out using a VP-ITC Mircocal clorimeter (Mircocal, Northhampton, MA, USA) in 20mM phosphate buffer (pH 6.8), 150mM KCl and 1mM TCEP. A typical titration consisted of injecting 10 μ l of the sugar into the malectin sample, at time intervals of 5min, to ensure that the titration peak returned to the baseline.



Figure S3: NOEs between malectin and nigerose. (A) Part of a ¹³C-edited half-filtered-NOESY experiment (mixing time 150 ms) showing intermolecular NOEs between malectin and nigerose. (B) Structure of α -nigerose.



Figure S4: Microarray analyses of the interactions of malectin with Glc₁-, Glc₂- and Glc₃-high mannose *N*-glycans and gluco-oligosaccharide probes. The oligosaccharide probes were printed as duplicate spots and binding was assayed with malectin at 20, 5, 1 and 0.5 µg/ml (panels A to D, respectively). Numerical scores are shown for the binding signals [means of duplicate values at 2 and 7 fmol/spot, (blue and red bars, respectively) with error bars]. At a malectin concentration of 20 µg/ml, the binding signals for the Glc₂-high mannose *N*-glycan probe, both at 2 and 7 fmol, were too high to be accurately quantified (asterisk in A) and were annotated as >> 50000 in Table 1. Other oligosaccharide probes tested included the glucose disaccharides kojibiose (Glc α 1-2Glc) nigerose (Glc α 1-3Glc), maltodextrins (Glc α 1-4Glc, dp 2-7); and oligosaccharides from dextran (isomalto) (Glc α 1-6Glc, dp 2-7); laminarin (Glc β 1-3Glc, dp 2-7); cellulose (Glc β 1-4Glc, dp 2-6); and pustulan (Glc β 1-6Glc, dp 2-7). Abbreviations G₃N, G₂N and G₁N designate Glc₃Man₇(D1)GlcNAc, Glc₂Man₇(D1)GlcNAc and Glc₁Man₉GlcNAc₂ *N*-glycan probes, respectively; dp, degree of polymerization of the gluco-oligomers.

Table S1: Structural statistics of malectin

	<sa>^a</sa>	<sa<sup>watref>^a</sa<sup>
A. number of structural restraints		
All NOE (unambiguous/ambigious)	4426/0	4426/0
Intraresidual	1320	1320
Sequential (i-j =1)	869	869
Medium-range $(1 \le i-j \le 4)$	456	456
Long-range (i-j >4)	1781	1781
Dihedral angles $\phi \psi$	99/99	99/99
Hbonds	66	66
B. rmsd (Å) from experimental restraints ^b		
All distance restraints	0.018 ± 0.000	0.025±0.001
Dihedral angles ^c	1.42±0.085	1.65±0.10
C. coordinate precision (Å) ^d		
N,C ^α ,C [′]	1.04±0.19	1.12±0.17
All heavy atoms	1.94±0.28	1.96±0.26
D. structural quality ^e		
Bad contacts	12.2±3.8	2.1±1.7
Ramachandran plot (%)		
Most favored region	61.1±3.0	68.8±1.7
Additionally allowed region	32.0±3.5	24.3±2.0
Generously allowed	4.8±1.9	4.6±1.9
Disallowed	2.0±1.2	2.4±0.7

^a \leq SA> is an ensemble of ten lowest-energy solution structures (out of 100 calculated) of malectin (AA28-201) before water-refinement, \leq SA^{watref}> is the \leq SA> ensemble after refinement in a shell of water (Linge et al., Methods Enzymol. 339,71-90, 2001). The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25 kcal mol⁻¹ Å⁻⁴ using "PROLSQ" van der Waals radii; Rms deviations for bond length, bond angles and improper dihedral angles are 0.0023 (±0.0001) Å, 0.426 (±0.012)° and 0.365 (±0.015)° before and 0.0055 (±0.0002) Å, 0.782 (±0.043)° and 2.15 (±0.21)° after water-refinement. 1kcal=4.18kJ.

^bDistance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol⁻¹ Å⁻². No distance restraint in the $\langle SA \rangle^a$ was violated by more than 0.2 Å.

^cDihedral angle restraints derived from TALOS (Cornilescu et al., J. Biomol. NMR *13*, 289-302, 1999) were applied to ϕ, ψ backbone angles using energy constants of 200 kcal mol⁻¹ rad⁻². ^dCoordinate precision is given as the Cartesian coordinate r.m.s. deviation of the ten lowest-energy structures in the

^dCoordinate precision is given as the Cartesian coordinate r.m.s. deviation of the ten lowest-energy structures in the NMR ensemble with respect to their mean structure for residue 28 to 201.

^eStructural quality was analyzed using PROCHECK.

Table S2: Domains having the highest structural similarity to malectin. The most similar structures in the PDB data bank identified by a search with MSD-Fold (www.ebi.ac.uk/msd-srv/ssm/) were the structures of the carbohydrate-binding modules (CBMs) of bacterial glycosylhydrolases (representatives of different families are shown) and the carbohydrate recognition domain of a ubiquitin ligase, Fbs1,. Their reported ligands are also shown.

domain/PDB	z-score	RMSD(Å)	seq-ID(%)	ligand	species
CBM22 (1dyo)	7.3	3.2	9	Xylan	Clostridium thermocellum
Fbs1 CRD (1umh)	7.2	2.9	7	Chitobiose	Mus musculus
CBM22 (1h6y)	7.2	3.2	10	Xylan	Clostridium thermocellum
CBM22 (1h6x)	7.2	3.3	9	Xylan	Clostridium thermocellum
CBM35 (2bgp)	6.6	2.8	14	Mannan	Clostridium japonocus
CBM6 (1uz0)	6.3	3.7	11	Cellubiose	Clostridium mixtus
CBM27 (1pmj)	6.1	3.3	12	Mannan	Clostridium sacharolyticum
CBM27 (1pmh)	6.0	3.3	10	Mannan	Clostridium sacharolyticum
CBM15 (1gny)	6.0	3.4	12	Xylan	Pseudomonas cellulosa
CBM4 (1k45)	6.0	3.0	9	Xylan	Rhodothermus marinus

PDB = protein data bank accession code, seq-ID = overall sequence identity

Carbohydrate	binding to malectin
D-sialic acid	_
L-fucose	_
D-glucose	_
D-N-acetyl-glucosamine	_
D-galactose	_
D-mannose	_
fructose	_
D-N-acetyl-galactosamine	_
lactose (Galβ1-4Glc)	_
sucrose (Glca1-2Fruc)	_
chitobiose (GlcNAcβ1-4GlcNAc)	_
trehalose (Glcβ1-1Glc)	_
mannobiose (Manα1-3Man)	_
mannobiose (Manβ1-4Man)	_
3'-siayllactose (NeuAcα2,3Galβ1-4Glc)	_
maltose (Glca1-4Glc)	++
maltotriose (Glca1-4Glca1-4Glc)	++
maltotetraose (Glca1-4Glca1-4Glca1-4Glc)	++
$maltoheptaose~(Glc\alpha 1-4Glc\alpha $	Gle) ++
kojibiose (Glca1-2Glc)	+
nigerose (Glca1-3Glc)	++
isomaltose (Glcα1-6Glc)	++
cellobiose (Glcβ1-4Glc)	+

Table S3: Carbohydrate-malectin interaction screen by NMR. Interactions were determined through ¹⁵N-chemical shift perturbation experiments. Relative affinities of oligosaccharides bound were determined through STD experiments.

-

an interaction was not observed in chemical shift perturbation experiments;

+ an interaction was observed in chemical shift perturbation experiments, but the STD signals were weak

++ an interaction was observed in chemical shift perturbation experiments and the STD signals were strong

Table S4: Chemical shift assignment of the disaccharides investigated by STD. The reducing end glucose rings exist in two isomeric forms in equilibrium, α and β according to the chirality of the anomeric centre. In the case of the disaccharides maltose, cellobiose, and isomaltose, the proton resonances of the non-reducing end glucose ring are fully overlapping, whereas the frequencies of the protons of the reducing end glucose ring are clearly separated. For nigerose and kojibiose, protons of the non-reducing end glucose also display two sets of resonances for the two isomers.

	Maltose				Cellobiose				
	α		β	β		α		β	
	$^{1}\mathrm{H}$	¹³ C	¹ H	¹³ C	$^{1}\mathrm{H}$	¹³ C	$^{1}\mathrm{H}$	¹³ C	
1	5.382	99.69	5.382	99.69	4.485	102.72	4.485	102.72	
2	3.551	71.81	3.551	71.81	3.291	73.34	3.291	73.34	
3	3.651	72.95	3.651	72.95	3.481	75.64	3.481	75.64	
4	3.389	69.42	3.389	69.42	3.463	76.13	3.463	76.13	
5	3.693	72.77	3.693	72.77	3.393	69.60	3.393	69.60	
6	3.825	60.592	3.825	60.592	3.892	60.71	3.892	60.71	
	3.732		3.732		3.709		3.709		
1'	5.198	92.02	4.622	95.88	5.197	91.91	4.636	95.91	
2'	3.537	71.42	3.243	74.12	3.549	71.36	3.245	74.03	
3'	3.939	73.38	3.738	76.30	3.795	71.45	3.620	78.72	
4'	3.617	76.97	3.615	76.76	3.615	78.84	3.597	74.44	
5'	3.911	70.06	3.567	74.69	3.921	70.29	3.570	74.94	
6'	3.799	60.71	3.881	60.85	3.848	60.04	3.929	60.19	
			3.736				3.782		

	Nigerose				Kojibiose			
	α		β		α		β	
1	5.361	99.05	5.347	98.99	5.091	96.43	5.381	97.87
2	3.561	71.65	3.554	71.60	3.549	71.54	3.538	71.58
3	3.753	72.86	3.734	72.91	3.783	72.9	3.744	72.94
4	3.435	69.41	3.453	69.25	3.450	69.56	3.456	69.47
5	4.000	71.71	4.007	71.65	3.939	71.95	4.027	71.78
6	3.816	60.36	3.797	60.19	3.835	60.48	3.800	60.41
	3.765				3.775			
1'	5.226	92.19	4.656	95.93	5.433	89.57	4.791	96.38
2'	3.626	70.07	3.327	72.80	3.630	76.04	3.378	78.68
3'	3.841	79.58	3.630	82.15	3.810	71.37	3.565	74.59
4'	3.626	70.07	3.640	70.02	3.453	69.72	3.422	69.89
5'	3.843	71.16	3.567	75.59	3.855	71.45	3.456	75.87
6'	3.837	60.37	3.878	60.54	3.840	60.72	3.890	60.94
	3.747		3.711		3.762		3.710	

	Isomaltose			
	α		β	
1	4.932	98.07	4.932	98.07
2	3.531	71.63	3.531	71.63
3	3.713	73.25	3.713	73.25
4	3.405	69.62	3.405	69.62
5	3.710	71.91	3.710	71.91
6	3.826	60.59	3.826	60.59
	3.741		3.741	
1'	5.218	92.37	4.650	96.25

	Isomaltos	se		
2'	3.527	71.57	3.228	74.19
3'	3.685	73.19	3.459	76.14
4'	3.493	69.68	3.494	69.50
5'	3.986	70.14	3.616	74.44
6'	3.973	65.88	3.932	65.83
	3.675		3.733	

Table S5: Structural statistics of the malectin-nigerose complex

<SA^{nigerose}>^a

All NOE (unambiguous/ambigious) Protein to protein	4458/0 4427
Protein to protein	4427
Protein to sugar	31
Intraresidual	1320
Sequential (i-j =1)	869
Medium-range $(1 < i-j \le 4)$	456
Long-range (i-j >4)	1782
Dihedral angles $\phi\psi$	99/99
Hbonds	66
B. rmsd (Å) from experimental restraints ^b	
All distance restraints	0.033±00.3
Dihedral angles ^c	1.92±0.08
C. coordinate precision (Å) ^d	
N,C ^α ,C [′]	0.92±0.16
All heavy atoms	1.77±0.18
D. structural quality ^e	
Bad contacts	14.0±2.6
Ramachandran plot (%)	
Most favored region	63.8±2.2
Additionally allowed region	28.8±2.6
Generously allowed	4.8±2.0
Disallowed	2.7±0.8

^a <SA^{nigerose}> is an ensemble of ten lowest-energy solution structures (out of 100 calculated) of the malectin nigerose complex (AA28-201) before water-refinement. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25 kcal mol⁻¹ Å⁻⁴ using "PROLSQ" van der Waals radii; Rms deviations for bond length, bond angles and improper dihedral angles are 0.0023 (±0.0001) Å, 0.426 (±0.012)° and 0.365 (±0.015)° before water-refinement. 1kcal=4.18kJ.

^bDistance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol⁻¹Å⁻².

^cDihedral angle restraints derived from TALOS (Cornilescu et al., J. Biomol. NMR *13*, 289-302, 1999) were applied to ϕ, ψ backbone angles using energy constants of 200 kcal mol⁻¹ rad⁻². ^dCoordinate precision is given as the Cartesian coordinate r.m.s. deviation of the ten lowest-energy structures in the

^dCoordinate precision is given as the Cartesian coordinate r.m.s. deviation of the ten lowest-energy structures in the NMR ensemble with respect to their mean structure for residue 28 to 201.

^eStructural quality was analyzed using PROCHECK.

Table S6: Oligosaccharide probes* included in the microarray and the binding signals (means at 7 fmol) with malectin at 5μ g/ml.

Pos**	Probe Name	Sequence	Signal at
			7 fmol
	Mammalian-type		
1	Glucocerebrosides	Glc8-Cer	5
2	Lac	GalB-4Glc-DH	12
3	Lac-AO	GalB-4Glc-AO	16
4	Lactocerebrosides	GalB-4GlcB-Cer	11
5	LacNAc	GalB-4GlcNAc-DH	10
6	Galß-3GalNAc	GalB-3GalNAc-DH	1
7	Galß-6GalNAc	GalB-6GalNAc-DH	10
8	GalNAca-3Galß-4Glc	GalNAca-3GalB-4Glc-DH	11
9	Gala-4Galß-4GlcNAc	Gala-4GalB-4GlcNAc-DH	2
10	Ceramide trihexoside	Gala-4GalB-4GlcB-Cer	<1
11	Globoside (P-antigen)	GalNAcB-3Gala-4GalB-4GlcB-Cer	3
12	Forssmann glycolipid	GalNAca-3GalNAcB-3Gala-4GalB-4GlcB-Cer	<1
13	H-Di	Fuca-2Gal-DH	31
		Galα-3Gal-DH	
14	B-Tri	Fucα-2	5
-			
15	A Tri	GalNAca-3Gal-DH	-1
15	A-111	Fuca-2	
16	Lex-Tri	GalB-4GlcNAc-DH	9
		Fuca-3	
		GalB-3GlcNAc-DH	
17	Lea-Tri	Fuca-4	<1
18	GM4	NeuAca-JGalβ-Cer	<1
19	NeuAcα-(3')Lac	NeuAca-3Gal&-4Glc-DH	31
20	NeuAcα-(6')Lac	NeuAca-6Galb-4Glc-DH	10
21	NeuAcß-(3')Lac	NeuAcS-3GalS-4Glc-DH	7
22	NeuAcß-(6')Lac	NeuAcB-6GalB-4Glc-DH	25
23	Neuα-(3')Lac	Neua-3GalB-4Glc-DH	10
24	Neuα-(6')Lac	Neua-6GalB-4Glc-DH	21
25	Neu4,5Ac-(3')Lac	Neu4, 5Aca-3Galb-4Glc-DH	<1
26	Neu5,9Ac-(6')LN	Neu5,9Aca-6GalB-4GlcNAc-DH	9
27	GM3	NeuAca-3GalB-4GlcB-Cer	<1
28	GM3(Gc)	NeuGca-3GalB-4GlcB-Cer	23
29	NeuAca-(6')LN	NeuAca-6Gal8-4GlcNAc-DH	<1
30	Haematoside	NeuAca-3GalB-4GlcB-Cer	<1
		NeuAca-3Galb-3GlcNAc-DH	
31	SA(3')-Lea-Tri	Fuca-4	<1
22	CD3	Newber 20-18 401-8 0er	-1
22	Sulfatide		×1
33			<1
34	005-1		15
35	GOF-19		15
30	SM3	SU-3GAID-46TC-DH	4
37	SU(3')-LN	SU-3GALB-4GLCNAC-DH	5
38	GSC-210	SU-JGICAB-JGALB-Cer42	6
30	SU(3') Lev-Tri	SU-3Galb-4GlcNAc-DH	-1
55	50(0)-Lex-111	Fuca-3	

40	SU(3')-Lea-Tri	SU-3Gal8-3GlcNAc-DH	9
41	GSC-150	SU-3GalB-4GlcB-C30 Fuca-3	3
42	GSC-432	3-deoxy, 3-carboxymethy1-GalB-4GlcB-C30	15
43	GSC-430	3-deoxy,3-carboxymethy1-Gal&-3Glc&-C30	1
44	GSC-260	3-deoxy,3-carboxymethy1-Gal8-4Glc8-C30	6
45	GSC-209	GlcAB-3GalB-Cer42	<1
46	Glc	Glc-DH	7
47	Gal	Gal-DH	2
48	Gal-AO	Gal-AO	15
49	LacNAc-AO	GalB-4GlcNAc-AO	11
50	NeuAca-(3')LN	NeuAca-3GalB-4GlcNAc-DH	<1
51	NeuAca-(3')LN-AO	NeuAca-3Galb-4GlcNAc-AO	<1
52	Lea-Tri-AO	Gal&-3GlcNAc-AO Fuca-4	<1
53	Lex-Tri-(Me)AO	Gal8-4GlcNAc-(Me)-AO Fucα-3	<1
54	Neu4,5Ac-(3')Lac-AO	Neu4,5Aca-3GalB-4Glc-AO	<1
55	LacN(1-3)-AO	Gal8-3GlcNAc-AO	<1
56	NeuAcα-(3')Lac-AO	NeuAca-3Gal8-4Glc-AO	7
57	NeuAcα-(6')Lac-AO	NeuAca-6Gal8-4Glc-AO	<1
58	NeuAcß-(3')Lac-AO	NeuAcs-3Gals-4Glc-AO	29
59	NeuAcß-(6')Lac-AO	NeuAcs-6Gals-4Glc-AO	5
60	Neuα-(3')Lac-AO	Neua-3GalB-4Glc-AO	19
61	Neuα-(6')Lac-AO	Neua-6GalB-4Glc-AO	8
62	Lex-Tri-AO	GalB-4GlcNAc-AO	<1
63	Galactocerebrosides	GalB-Cer	5
64	GSC-426	3-deoxy-3Ac-GalB-C30	15
65	LNT	GalB-3GlcNAcB-3GalB-4Glc-DH	10
66	LNnT	GalB-4GlcNAcB-3GalB-4Glc-DH	14
67	Paragloboside	GalB-4GlcNAcB-3GalB-4GlcB-Cer	<1
68	B-like pentaosylceramide	Gala-3GalB-4GlcNAcB-3GalB-4GlcB-Cer	<1
69	Klaus glycolipid	GalB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer	5
70	LNFP-I	Fuca-2GalB-3GlcNAcB-3GalB-4Glc-DH	20
71	LNFP-III	GalB-4GlcNAcB-3GalB-4Glc-DH Fucα-3	39
72	LNFP-II	GalB-3GlcNAcB-3GalB-4Glc-DH Fucα-4	20
73	A-Hexa	GalNAcα-3GalB-3GlcNAcB-3GalB-4Glc-DH Fucα-2	8
74	A-Hepta	GalNAca-3GalB-3GlcNAcB-3GalB-4Glc-DH Fuca-2 Fuca-4	30

		-	
75	LNnDFH-V	GalB-4GlcNAcB-3GalB-4Glc-DH Fuca-3 Fuca-2	3
76	LNDFH-I	Fuca-2GalB-3GlcNAcB-3GalB-4Glc-DH	<1
77	LNnDFH-I	Fuca-2Gal&-4GlcNAc&-3Gal&-4Glc-DH Fuca-3	<1
78	LNnDFH-II	GalB-4GlcNAcB-3GalB-4Glc-DH Fuca-3 Fuca-3	<1
79	LNDFH-II	GalB-3GlcNAcB-3GalB-4Glc-DH Fuca-4 Fuca-3	<1
80	LNnTFH-I	Fuca-2GalB-4GlcNAcB-3GalB-4Glc-DH Fuca-3 Fuca-2	6
81	LNTFH-I	Fucα-2GalB-3GlcNAcB-3GalB-4Glc-DH Fucα-4 Fucα-2	14
82	Leb-hexaosylceramide	Fuca-2Gal&-3GlcNAc&-3Gal&-4Glc&-Cer Fuca-4	<1
83	B-hexaosylceramide	Gala-3GalB-4GlcNAcB-3GalB-4GlcB-Cer Fuca-2	22
84	LSTa	NeuAca-3GalB-3GlcNAcB-3GalB-4Glc-DH	24
85	GSC-31	NeuAca-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36	9
86	Sialylparagloboside	Neu2co-3Ce18-4C1cN2c8-3Ce18-4C1c8-Cer	<
87	SA(3')-LNFP-II	NeuAcα-3GalB-3GlcNAcB-3GalB-4Glc-DH Fucα-4	<1
88	SA(3')-LNFP-III	NeuAcα-3Gal&-4GlcNAc&-3Gal&-4Glc-DH Fucα-3	<1
89	GSC-64	NeuAca-3Gal&-4GlcNAc&-3Gal&-4Glc&-Cer36 Fuca-3	<1
90	GSC-472	Neua-3Gal&-4GlcNAc&-3Gal&-4Glc&-Cer36 Fuca-3	<1
91	SA(3/6)LNFP-I	NeuAca-3/6GalB-3GlcNAcB-3GalB-4Glc-DH Fuca-2	<1
92	GSC-97	Neuλcα-6Galß-4GlcNAcß-3Galß-4Glc-Cer36 Fucα-3	<1
93	LSTb	Galβ-3GlcNAcβ-3Galβ-4Glc-DH NeuAcα-6	<1
94	DSLNT	NeuAcα-3Gal&-3GlcNAc&-3Gal&-4Glc-DH NeuAcα-6	1
95	SU(3')-Tri	SU-3Gal8-4GlcNAc8-3Gal-DH	<1
96	GSC-190	SU-3GlcAB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer42	3
97	SU(3')-LNFP-II	SU-3Gal8-3GlcNAc8-4Gal8-4Glc-DH Fucα-4	<1
98	SU(3')-LNFP-III	SU-3Gal8-4GlcNAc8-3Gal8-4Glc-DH Fuca-3	9

99	SU(6')-LNFP-II	SU-6Gal&-3GlcNAc&-3Gal&-4Glc-DH	3
100	SU(6')-LNFP-III	SU-6Gal&-4GlcNAc&-3Gal&-4Glc-DH Fuca-3	7
101	SU(3',6)-LNFP-III	SU-6 SU-3Gal&-4Glc-DH Fuca-3	4
102	GSC-268	SU-6 NeuAca-3GalB-4GlcNAcB-3GalB-4Glc-Cer36 Fuca-3	5
103	GSC-268 deNAc	SU-6 Neua-3GalB-4GlcNAcB-3GalB-4Glc-Cer36 Fuca-3	2
104	GSC-269	SU-6 NeuAca-3GalB-4GlcA-Cer36 Fuca-3	<1
105	GSC-406	SU-6 Neua-3Galß-4Glcß-Cer36 Fuca-3	15
106	GSC-270	SU-6 SU-6 NeuAcα-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36 Fucα-3	<1
107	GSC-189	GlcAB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer42	16
108	GSC-207	GlcAB-3GalB-4GlcNAcB-3GalB-4GlcB-C30	2
109	Led-II pentaosylceramide	Fucα-2Galß-3GlcNAcß-3Galß-4Glcß-CerA	<1
110	Led-I pentaosylceramide	Fuca-2GalB-3GlcNAcB-3GalB-4GlcB-CerB	<1
111	LNFP-III-AO	GalB-4GlcNAcB-3GalB-4Glc-AO Fuca-3	<1
112	GSC-31	NeuAca-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36	15
113	LSTc	NeuAca-6Galβ4-GlcNAcβ3-Galβ4-Glc-DH	3
114	GSC-208	SU-3GlcAB-3GalB-4GlcNAcB-3GalB-4GlcB-C30	3
115	GSC-97	NeuAca-6GalB-4GlcNAcB-3GalB-4GlcB-Cer36 Fuca-3	5
116	DLNN	GlcNAcB-3GalB-4Glc-DH	6
117	GSC-105	NeuAca-3GalB-4GlcNAcB-3GalB-Cer36 Fuca-3	<1
118	GSC-516B	Neuα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36 SU-6	6
119	pLNnH	GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4Glc-DH	<1
120	LNnH	GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-4GlcNAcB-3	3



128	B-like pentaeicosaosylceramide	Gal&-4GlcNAc&-6 Gal&-4GlcNAc&-3 Gal&-4GlcNAc&-3 Gal&-4GlcNAc&-3	2
129	pLNFH-IV	Galß-3GlcNAcβ-3Galß-4GlcNAcβ-3Galß-4Glc-DH	<1
130	DFpLNH-II	GalB-3GlcNAcB-3GalB-4GlcNAcB-3GalB-4Glc-DH Fuca-4 Fuca-3	9
131	MFLNH-III	Fucα-3 Galb-4GlcNAcb-6 Galb-4Glc-DH Galb-3GlcNAcb-3	<1
132	MFLNnH(a)	Fucα-3 GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-4GlcNAcB-3	8
133	DFLNH(a)	GalB-4GlcNAcB-6 Fuca-3 GalB-4Glc-DH Fuca-2GalB-3GlcNAcB-3	23
134	DFLNH(c)	GalB-4GlcNAcB-6 GalB-4Glc-DH Fuca-2GalB-3GlcNAcB-3 Fuca-4	16
135	DFLNnH	Fucα-3 GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-4GlcNAcB-3 Fucα-3	<1
136	DFLNH(b)	Fucα-3 GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-3GlcNAcB-3 Fucα-4	9

137	TFLNH	Fucα-3 GalB-4GlcNAcB-6 GalB-4Glc-DH Fucα-2GalB-3GlcNAcB-3 Fucα-4	8
138	TFpLNH-I	Fuca-2Gal&-3GlcNAc&-3Gal&-4GlcNAc&-3Gal&-4Glc-DH Fuca-4 Fuca-3	11
139	MFilno-IV	Fuca-3 Galß-3GlcNAcß-3Galß-4GlcNAcß-6 Galß-4Glc-DH Galß-3GlcNAcß-3	20
140	MFLND	Fucα-3 Gal8-4GlcNAc8-6 Gal8-3GlcNAc8-3 Gal8-3GlcNAc8-3 Gal8-3GlcNAc8-3 Gal8-3GlcNAc8-3 Gal8-3GlcNAc8-3	7
141	B-III dodecaosylceramide	Galα-3Galß-4GlcNAcß-6 Fucα-2 Galß-4GlcNAcß-3Galß-4Glcß-Cer Galα-3Galß-4GlcNAcß-3 Fucα-2	12
142	B-IV tetradecaosylceramide	Gala-3GalB-4GlcNAcB-6 Fuca-2 GalB-4GlcNAcB-3GalB-4GlcB-Cer Gala-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3 Fuca-2	<1
143	MSLNH	NeuAca-6Gal&-4GlcNAc&-6 Gal&-4Glc-DH Gal&-3GlcNAc&-3	<1
144	MSLNnH-I	Gal&-4GlcNAc&-6 Gal&-4Glc-DH NeuAca-6Gal&-3GlcNAc&-3	<1
145	MSMFLNH	Fuca-3 Gal&-4GlcNAc&-6 Gal&-4Glc-DH NeuAca-3Gal&-3GlcNAc&-3	<1
146	MFMSLNnH	Galß-4GlcNAcß-6 Fucα-3 Galß-4Glc-DH NeuAcα-6Galß-3GlcNAcß-3	<1

147	DSLNnH	Neuλcα-6GalB-4GlcNλcB-6 GalB-4Glc-DH Neuλcα-6GalB-4GlcNλcB-3	<1
148	GSC-219	SU-3GlcAB-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36	2
149	C4U	Neuλcα-3Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH SU-6 SU-6 SU-6	3
150	FucC4U	Fuca-3 NeuAca-3Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH SU-6 SU-6 SU-6	<1
151	GSC-216	GlcAB-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer42	8
152	TFiLNO	Fucα-4 Fucα-3 GalB-3GlcNAcB-3GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-3GlcNAcB-3 μ Fucα-4	6
153	pLNH	GalB-3GlcNAcB-3GalB-4GlcNAcB-3GalB-4Glc-DH	<1
154	LNH	GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-3GlcNAcB-3	2
155	iLNO	GalB-3GlcNAcB-3GalB-4GlcNAcB-6 GalB-4Glc-DH GalB-3GlcNAcB-3	<1
156	l-octaosylceramide	GalB-4GlcNAcB-6 GalB-4GlcNAcB-3GalB-4GlcB-Cer GalB-4GlcNAcB-3	<1
157	B-like decaosylceramide	Galα-3Galß-4GlcNAcß-6 Galß-4GlcNAcß-3Galß-4Glcß-Cer Galα-3Galß-4GlcNAcß-3	8
158	GSC-217	SU-3GlcAB-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer42	10
159	GSC-220	NeuAca-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36	19
160	GSC-221	NeuAca-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36 Fuca-3	3
161	GSC-218	GlcAB-3GalB-4GlcNAcB-3GalB-4GlcNAcB-3GalB-4GlcB-Cer36	<1
162	Glc3Man7(D1)GN1-AO	Manα-6 Manα-3Manα-6 Manβ-4GlcNAc-AO Glcα-2Glcα-3Glcα-3Manα-2Manα-2	1077

2	6
7	υ

163	Glc2Man7(D1)GN1-AO	Manα-6 Manα-3Manα-6 Manβ-4GlcNAc-AO Glcα-3Glcα-3Manα-2Manα-2 Manα-3	46454
164	Glc1Man9GN2-AO	Manα-2Manα-6 Manα-6 Manα-2Manα-3 Manβ-4GlcNAcβ-4GlcNAc-AO Glcα-3Manα-2Manα-2Manα-3	182
165	Man9GN2-AO	Manα-2Manα-6 Manα-2Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-AO Manα-2Manα-2Manα-3	<1
166	Man8(D1D3)GN2	Manα-2Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-2Manα-2Manα-3	3
167	Man7(D1)GN2	Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-2Manα-2Manα-3	<1
168	Man7(D1)GN2-AO	Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-AO Manα-2Manα-2Manα-3	17
169	Man7(D3)GN2	Manα-2Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-2Manα-3	<1
170	Man6GN2	Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-2Manα-3	<1
171	Man5GN2	Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3	<1

172	Man4aGN2	Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3	<1
173	Man4bGN2	Manα-6 Manα-3Manα-6 Manß-4GlcNAcß-4GlcNAc-DH	<1
174	Man3FGN2	Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3	<1
175	Man3GN2	Manα-6 Manβ-4GlcNAcB-4GlcNAc-DH Manα-3	11
176	Man2aGN2	Mana-6Manß-4GlcNAc8-4GlcNAc-DH	25
177	Man2GN1	Mana-3ManB-4GlcNAc-DH	<1
178	Man1GN1	ManB-4GlcNAc-DH	1
179	N1	GalB-4GlcNAcB-2Manα-6 Fucα-6 ManB-4GlcNAcB-4GlcNAc-DH Manα-3	5
180	N2	Manα-6 Manß-4GlcNAcß-4GlcNAc-DH Galß-4GlcNAcß-2Manα-3	1
181	N4	Galß-4GlcNAcß-2Manα-6 Manß-4GlcNAcß-4GlcNAc-DH Manα-3	1
182	N3	GlcNAcB-2Manα-6 +GalB-4 ManB-4GlcNAcB-4GlcNAc-DH GlcNAcB-2Manα-3	3
183	NGA2	GlcNAcB-2Manα-6 ManB-4GlcNAcB-4GlcNAc-DH GlcNAcB-2Manα-3	<1
184	NGA2B	GlcNAcB-2Manα-6 GlcNAcB-4GlcNAcB-4GlcNAc-DH GlcNAcB-2Manα-3	10
185	NA2	GalB-4GlcNAcB-2Manα-6 ManB-4GlcNAcB-4GlcNAc-DH GalB-4GlcNAcB-2Manα-3	<1
186	NGA2F	GlcNAcB-2Mana-6 Fuca-6 ManB-4GlcNAcB-4GlcNAc-DH	<1

GlcNAcβ-2Manα-3

187	NA2F	GalB-4GlcNAcB-2Manα-6 Fucα-6 ManB-4GlcNAcB-4GlcNAc-DH GalB-4GlcNAc6-2Manα-3	<1
188	NA2F-AO	Galb-4GlcNAcb-2Manα-6 Fucα-6 Manb-4GlcNAcb-4GlcNAc-AO Galb-4GlcNAcb-2Manα-3	<1
189	NA2FB	GalB-4GlcNAcB-2Manα-6 Fucα-6 GlcNAcB-4ManB-4GlcNAcB-4GlcNAc-DH GalB-4GlcNAcB-2Manα-3	<1
190	A2(2-6)	NeuAca-6Gal&-4GlcNAc&-2Mana-6 Man&-4GlcNAc&-4GlcNAc-DH NeuAca-6Gal&-4GlcNAc&-2Mana-3	1
191	A2F(2-3)	NeuAca-3Gal&-4GlcNAc&-2Mana-6 Fuca-6 Man&-4GlcNAc&-4GlcNAc-DH NeuAca-3Gal&-4GlcNAc&-2Mana-3	<1
192	AGP-Bi-Ac2	NeuAcα-Galß-4GlcNAcß-2Manα-6 Manß-4GlcNAcß-4GlcNAc-DH NeuAcα-Galß-4GlcNAcß-2Manα-3	<1
193	AGP-Bi-AcGc	Galß-4GlcNAcβ-2Manα-6 +NeuGc Manβ-4GlcNAcβ-4GlcNAc-DH +NeuAc Galß-4GlcNAcβ-2Manα-3	<1
194	AGP-Bi-Gc2	NeuGcα-Galb-4GlcNAcb-2Manα-6 Manb-4GlcNAcb-4GlcNAc-DH NeuGcα-Galb-4GlcNAcb-2Manα-3	<1
195	NGA3B	GlcNAcB-2Manα-6 GlcNAcB-4Manß-4GlcNAcB-4GlcNAc-DH GlcNAcB-4Manα-3 GlcNAcB-2	<1
196	NA3-Lex	Galß-4GlcNAcß-2Manα-6 Manß-4GlcNAcß-4GlcNAc-DH +Fucα-3 Galß-4GlcNAcß-4Manα-3 Galß-4GlcNAcß-2	<1
197	NA3	GalB-4GlcNAcB-2Manα-6 ManB-4GlcNAcB-4GlcNAcB-4GlcNAc-DH GalB-4GlcNAcB-4Manα-3 GalB-4GlcNAcB-2	<1

			1
198	A3	NeuAcα-3GalB-4GlcNAcB-2Manα-6 ManB-4GlcNAcB-4GlcNAcB-4GlcNAc-DH NeuAcα-3GalB-4GlcNAcB-4Manα-3 NeuAcα-6GalB-4GlcNAcB-2	<1
199	NGA4	GlcNAcB-6 GlcNAcB-2Mana-6 ManB-4GlcNAcB-4GlcNAc-DH GlcNAcB-4Mana-3 GlcNAcB-2	<1
200	NA4	GalB-4GlcNAcB-6 GalB-4GlcNAcB-2Manα-6 ManB-4GlcNAcB-4GlcNAcB-4GlcNAc-DH GalB-4GlcNAcB-4Manα-3 GalB-4GlcNAcB-2	10
201	NGA5B	GlcNAcB-2 GlcNAcB-4Mana~6 GlcNAcB-6 GlcNAcB-4ManB-4GlcNAcB-4GlcNAc-DH GlcNAcB-4Mana~3 GlcNAcB-2	6
202	GN2-Man5BGN2	Manα-6 Manα-3Manα-6 GlcNAcB-4ManB-4GlcNAcB-4GlcNAc-DH GlcNAcB-2Manα-3	<1
203	Fuc-GlcNAc	Fuca-6GlcNAc-DH	8
204	Man3(α3,α6)	Manα-6Man-DH Manα-3	<1
205	Man5(α3,α6)	Manα-6Man-DH Manα-3 Manα-3	3
206	Man2(α3)	Mana-3Man-DH	<1
207	Man2(α2)	Manα-2Man-DH	<1
208	SA2(α8)	NeuAca-SNeuAc-DH	6
209	SA3(uo)	Neuhou-oneuhou-oneuhou-dh Neuhou-8Neuhou-8Neuhou-DH	10
210	SA5(a8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-BNeuAca-DH***	<1
212	SA6(α8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-DH***	<1
213	SA7(α8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-DH***	<1
214	SA8(α8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-8NeuAca-8NeuAc-DH***	3
215	SA9(α8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-8NeuAca-8NeuAc-8NeuAca-DH***	7
216	SA10(a8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-DH***	<1
217	SA11(α8)	NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAca-8NeuAc-8NeuAca-8NeuAac-8NeuAca-8NeuAca-8NeuAca-8NeuAc-DH***	3

218	NeuAc	NeuAc-DH	<1
219	NeuAc-AO	NeuAc-AO	<1
220	Asialo-GM2	GalNAcβ-4Galβ-4Glcβ-Cer	10
221	GalNAc-Tetra	GalB-3GalNAcB-4GalB-4Glc-DH	<1
222	Asialo-GM1	Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer	3
223	GM1	Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer NeuAca-3	<1
224	GM1(Gc)	Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer NeuGcα-3	<1
225	GM1-penta	Galβ-3GalNAcβ-4Galβ-4Glc-DH NeuAcα-3	<1
226	GM1(Gc)-penta	Galβ-3GalNAcβ-4Galβ-4Glc-DH NeuGcα-3	<1
227	GM2	GalNAcβ-4Galβ-4Glcβ-Cer NeuAcα-3	<1
228	GD1a	NeuAca-3GalB-3GalNAcB-4GalB-4GlcB-Cer NeuAca-3	<1
229	GalNAc-GD1a(Ac,Gc)	GalNAcB-4GalB-3GalNAcB-4GalB-4GlcB-Cer NeuGca-3 NeuAca-3 and/or GalNAcB-4GalB-3GalNAcB-4GalB-4GlcB-Cer NeuAca-3 NeuGca-3	<1
230	GD1b	GalB-3GalNAcB-4GalB-4GlcB-Cer NeuAca-8NeuAca-3	<1
231	GD2	GalNAcB-4GalB-4GlcB-Cer NeuAca-8NeuAca-3	<1
232	GT1a	NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer NeuAcα-3	<1
233	GT1b	NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer NeuAcα-8NeuAcα-3	<1
234	GQ1b	NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer NeuAcα-8NeuAcα-3	<1
235	SM1a	GalB-3GalNAcB-4GalB-4Glc-DH	<1
236	SB2	SU-3GalNAcB-4GalB-4Glc-DH	<1
237	SB1a	SU-3Gal&-3GalNAc&-4Gal&-4Glc-DH	<1
238	GD1a-hexa	NeuAcα-3GalB-3GalNAcB-4Glb-4Glc-DH NeuAcα-3	<1
239	GM1b	NeuAca-3GalB-3GalNAcB-4GalB-4GlcB-Cer***	9
240	SM2	GalNAcB-4GalB-4GlcB-Cer SU-3	10
241	A8/1	GlcNAca-4Galβ-OX	<1

242	A15/3	GlcNAca-4Galβ-3Galβ-OX	<1
243	B13/b	GalNAcβ-4Galβ-3GlcNAcβ-0X KDNα-3	3
244	DSL	NeuAca-3GalB-3GalNAc-DH	<1
245	A15/1	SU-6GlcNAc8-OY	<1
246	A8/2	SU-6 Fuca-3GlcNAcB-0Y	3
247	B13/a-AO	GlcAβ-3Galβ-3GlcNAcβ-OX-AO	18
248	Notch-1	Fuca-Thr-DH	<1
249	Notch-2	GlcNAc _b -3Fuca-Thr-DH	<1
250	Notch-3	Galβ-3GlcNAcβ-3Fucα-Thr-DH	10
251	B12/3	GalNAcβ-4Galβ-OX NeuGcα-3	70
252	Man-Ser	Mana-Ser-DH	15
253	Man-Thr	Mana-Thr-DH	4
254	GalNAc-Thr	GalNAca-Thr-DH	<1
255	GalNAc-Ser	GalNAca-Ser-DH	7
256	Man-Thr-Succ	Manα-Thr-Succ-DH	<1
257	Man-Ser-Succ	Mana-Ser-Succ-DH	8
258	CSA-4	AUA-3GalNAc6-4GlcA8-3GalNAc-DH*** SU-4 SU-4	7
259	CSA-14	AUA-3GalNAcB-4GlcAB-3GalNAcB-4GlcAB-3GalNAcB-4GlcAB-3GalNAcB-4GlcAB- I I I SU-4 SU-4 SU-4 3GalNAcB-4GlcAB-3GalNAcB-4GlcAB-3GalNAcB-4GlcAB- I SU-4 SU-4 SU-4 3GalNAcB-4GlcAB-3GalNAcD-DH*** I I SU-4 SU-4 SU-4	6
260	CSB-4	ΔUA-3GalNAc8-4IdoAα-3GalNAc-DH*** SU-4 SU-4	<1
261	CSB-14	ΔUA-3GalNAcB-4IdoAα-3GalNAcB-4IdoAα-3GalNAcB-4IdoAα-3GalNAcB-4IdoAα-3GalNAcB-4IdoAα- SU-4 SU-4 SU-4 SU-4 SU-4 3GalNAcB-4IdoAα-3GalNAc-DH*** SU-4 SU-4 SU-4	14
262	CSC-4	ΔUA-3GalNAcB-4GlcAB-3GalNAc-DH*** SU-6 SU-6	5
263	CSC-14	ΔUA-3GalNAc8-4GlcA8-3GalNAc8-4GlcA8-3GalNAc8-4GlcA8-3GalNAc8-4GlcA8- SU-6 SU-6 SU-6 SU-6 3GalNAc8-4GlcA8-3GalNAc-DH*** SU-6 SU-6	<1
264	HA-4	GlcAb-3GlcNAcb-4GlcAb-3GlcNAc-DH	5
265	HA-14	GlcAB-3GlcNAcB-4GlcAB-3GlcNAcB-4GlcAB-3GlcNAcB-4GlcAB-3GlcNAcB-4GlcAB-3GlcNAcB-4GlcAB- 3GlcNAcB-4GlcAB-3GlcNAc-DH***	14
266	HEP-Di-IS	ΔUA-4G1cNS-DH*** SU-2 6-SU	12
267	HS-8	ΔUA-4GlcNAcα-4HexAB-4GlcNAcα-4HexAB-4GlcNAcα-4HexAB-4aMan-DH***	14
268	HEP-Di-IS-AO	ΔUA-4GlcNS-AO SU-2 6-SU	<1

	Homo-oligomers		
269	Glc2(a2)-AO	Glca-2Glc-AO	<1
270	Glc2(a3)-AO	Glca-3Glc-AO	555
271	Glc2(a4)-AO	Glca-4Glc-AO	507
272	Glc3(a4)-AO	Glca-4Glca-4Glc-AO	30
273	Glc4(a4)-AO	Glca-4Glca-4Glca-4Glc-AO	95
274	Malto-5-AO	Glca-4Glca-4Glca-4Glca-4Glc-AO	28
275	Malto-6-AO	Glca-4Glca-4Glca-4Glca-4Glc-AO	140
276	Malto-7-AO	Glca-4Glca-4Glca-4Glca-4Glca-4Glca-AO	492
277	Glc2(a6)-AO	Glca-6Glc-AO	42
278	Glc3(a6)-AO	Glca-6Glca-6Glc-AO	34
279	Glc4(a6)-AO	Glca-6Glca-6Glca-6Glc-AO***	933
280	Glc5(a6)-AO	Glca-6Glca-6Glca-6Glc-AO***	3805
281	Glc6(a6)-AO	Glca-6Glca-6Glca-6Glca-6Glc-AO***	5793
282	Glc7(a6)-AO	Glca-6Glca-6Glca-6Glca-6Glca-6Glc-AO***	458
283	Glc2(ß3)-AO	Glc8-3Glc-AO	<1
284	Lam-3-AO	Glc8-3Glc8-3Glc-AO	<1
285	Lam-4-AO	GlcB-3GlcB-3GlcB-3Glc-AO	36
286	Lam-5-AO	GlcB-3GlcB-3GlcB-3GlcB-3Glc-AO	54
287	Lam-6-AO	GlcB-3GlcB-3GlcB-3GlcB-3GlcB-3Glc-AO***	285
288	Lam-7-AO	Glc&-3Glc&-3Glc&-3Glc&-3Glc&-3Glc&-3Glc-AO***	210
289	Glc2(ß4)-AO	Glc8-4Glc-AO	<1
290	Cello-3-AO	GlcB-4GlcB-4Glc-AO	<1
291	Cello-4-AO	GlcB-4GlcB-4GlcB-4Glc-AO	<1
292	Cello-5-AO	Glc8-4Glc8-4Glc8-4Glc8-4Glc-AO	38
293	Cello-6-AO	GlcB-4GlcB-4GlcB-4GlcB-4GlcB-4Glc-AO	22
294	Glc2(ß6)-AO	Glc8-6Glc-AO	<1
295	Pust-3-AO	Glc8-6Glc8-6Glc-AO	27
296	Pust-4-AO	GlcB-6GlcB-6GlcB-6Glc-AO***	<1
297	Pust-5-AO	GlcB-6GlcB-6GlcB-6GlcB-6Glc***	48
298	Pust-6-AO	GlcB-6GlcB-6GlcB-6GlcB-6GlcB-6Glc-AO***	238
299	Pust-7-AO	Glc8-6Glc8-6Glc8-6Glc8-6Glc8-6Glc8-6Glc-AO***	244
300	GalNAca-3GalNAc	GalNAca-3GalNAc-DH	<1
301	GN2	GlcNAc8-4GlcNAc-DH	<1
302	GN3	GlcNAc8-4GlcNAc8-4GlcNAc-DH	13
303	Man2(α6)	Mana-6Man-DH	22
304	Man5(ß4)	Manß-4Manß-4Manß-4Man-DH	16
305	Man4(ß4)	Manß-4Manß-4Man-DH	<1
306	Man6(ß4)	Manß-4Manß-4Manß-4Manß-4Manß-4Man-DH	20
307	Ara6(α5)	Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH	19
308	Ara7(α5)	Araα-5Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH	7
309	Xyl5(ß4)	XylB-4XylB-4XylB-4Xyl-DH	13
	Miscellaneous		
310	Man3FXylGN2	Manα-6 Xylβ-2Manα-4GlcNAcβ-4GlcNAc-DH Manα-3 Fucα-3	<1
311	Man3XylGN2	Manα-6 XylB-2ManB-4GlcNAcB-4GlcNAc-DH Manα-3	2
312	Galα-4Glc-AO	Gala-4Glc-AO	27

313	Xyl3Glc4	GlcB-4GlcB-4GlcB-4Glc-DH Xyla-6 Xyla-6 Xyla-6	17
314	Gal4(α3,B4,α3)	Gala-3GalB-4Gala-3Gal-DH	<1
315	SU-Tyr	SU-Tyr-DH	8
316	Carra-Hexa-3S	aGalα-3GalB-4aGalα-3GalB-4aGalα-3Gal-DH SU-4 SU-4 SU-4	<1
317	Carra-Hexa-4S	aGala-3Gal8-4aGala-3Gal8-4aGala-3Gal-DH SU-4 SU-2 SU-4 SU-4	<1
318	Carra-Octa-4S	aGala-3Gal8-4aGala-3Gal8-4aGala-3Gal-DH SU-4 SU-4 SU-4 SU-4	<1
319	Carra-Tetra-2S	aGalα-3Gal8-4aGalα-3Gal-DH SU-4 SU-4	<1
320	Carra-Tetra-1S	aGalα-3Gal&-4aGalα-3Gal-DH SU-4	<1
321	Glc-AO	Glc-AO	13
322	GalNAc	GalNAc-DH	16
323	GalNAc-AO	GalNAc-AO	7
324	(6P)-Man	P-6Man-DH	13
325	(6P)-Man-AO	P-6Man-AO	15
326	Man	Man-DH	21
327	Man-AO	Man-AO	33
328	Fuc	Fuc-DH	15
329	Fuc-AO	Fuc-AO	62
330	Rha	Rha-DH	<1
331	Rha-AO	Rha-AO	11
332	GN	GlcNAc-DH	8
333	GN-AO	GlcNAc-AO	25
334	(6P)-Man5	P-6Mana-3Mana-3Mana-2Man-DH	2
335	Glc4(a6,a4,a4)	Glca-6Glca-4Glca-4Glc-DH	40

* The oligosaccharide probes are all lipid-linked, and are from the collection assembled in the course of research in Glycosciences Laboratory. DH, designates NGLs prepared from reducing oligosaccharides by reductive amination with the amino lipid, 1,2-dihexadecyl-*sn*-glycero-3-phosphoethanolamine (DHPE); OX and OY designate, respectively, the C1-4 fragment and the C5-6 fragments of GalNAcol of reduced oligosaccharides after mild periodate oxidation followed by reductive amination with DHPE (Chai et al., Methods Enzymol. 362, 160-195, 2003); AO, NGLs prepared from reducing oligosaccharides by oxime ligation with an aminooxy (AO) functionalized DHPE (Liu et al., Chem. Biol. 14, 847–859, 2007); Cer, natural glycolipids with various ceramide moieties; CerA and CerB denote different natural ceramides; Cer36 and Cer42, synthetic glycolipids with ceramide having a total of 36 and 42 carbon atoms, respectively; C30, a synthetic lipid [2-(tetradecyl)hexadecanol] with 30 carbon atoms. ΔUA, 4,5-unsaturated hexuronic acid; aMan, 2,5-anhydro-mannose; aGal, 3,6-anhydro-galactose.

** Pos, position in screening microarray.

*** Major component.