SUPPLEMENTARY FIGURES S1-S18

New carbohydrate binding domains identified by phage display based functional metagenomic screens of human gut microbiota

Akil Akhtar[1,3], Madhu Lata[1,2], Sonali Sunsunwal[1,4], Amit Yadav[1,2,4], Kajal[1,2,4], Srikrishna Subramanian[1,2] and T.N.C. Ramya[1,2,*]

[1]CSIR- Institute of Microbial Technology, Sector 39-A, Chandigarh 160036, INDIA

[2]Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh 201002, INDIA

[3]Currently at Emory Vaccine Center, Emory University, Atlanta, GA, United States of America

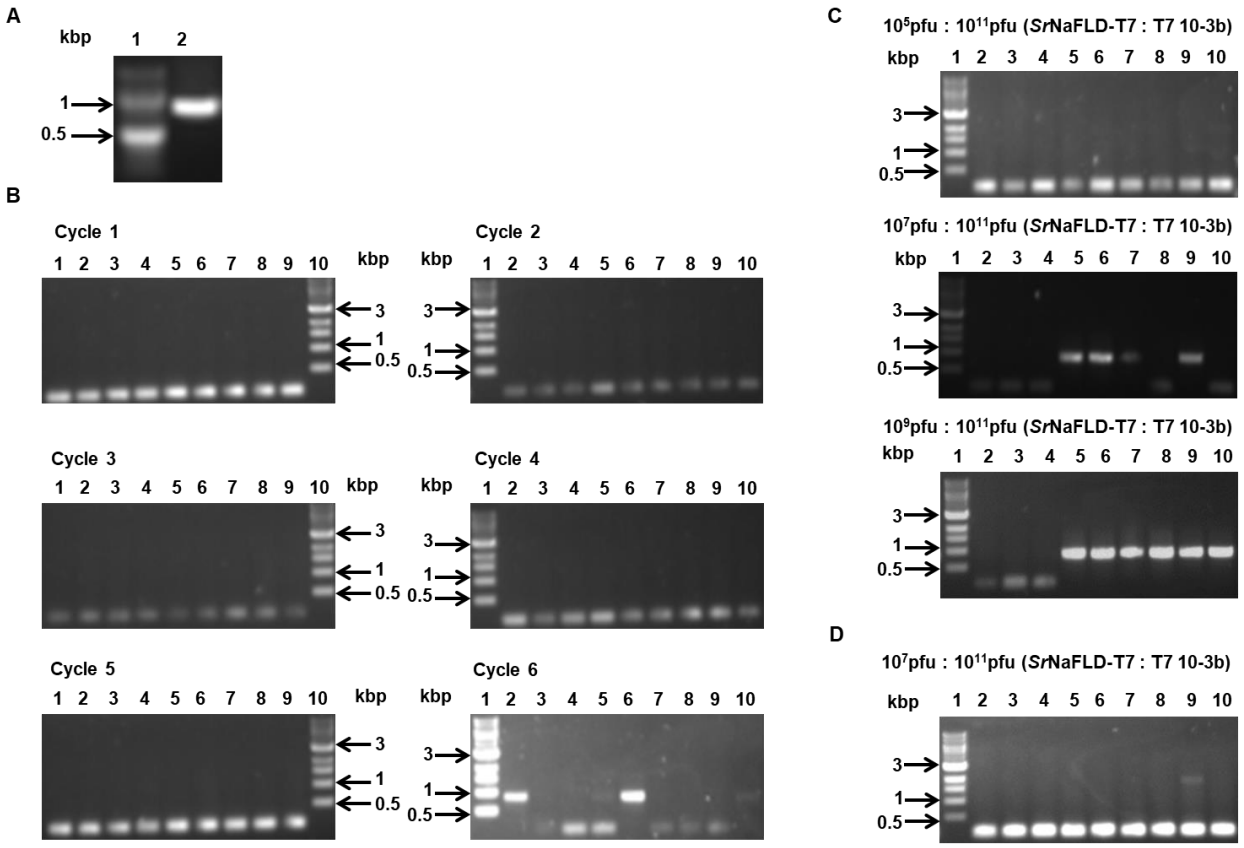
[4]These authors contributed equally to the work and should be considered as joint third co-authors

[*]Correspondence to be addressed to

T.N.C. Ramya, CSIR- Institute of Microbial Technology, Sector 39-A, Chandigarh 160036, INDIA.
Tel: 91-172-2880243; E-mail: ramya@imtech.res.in

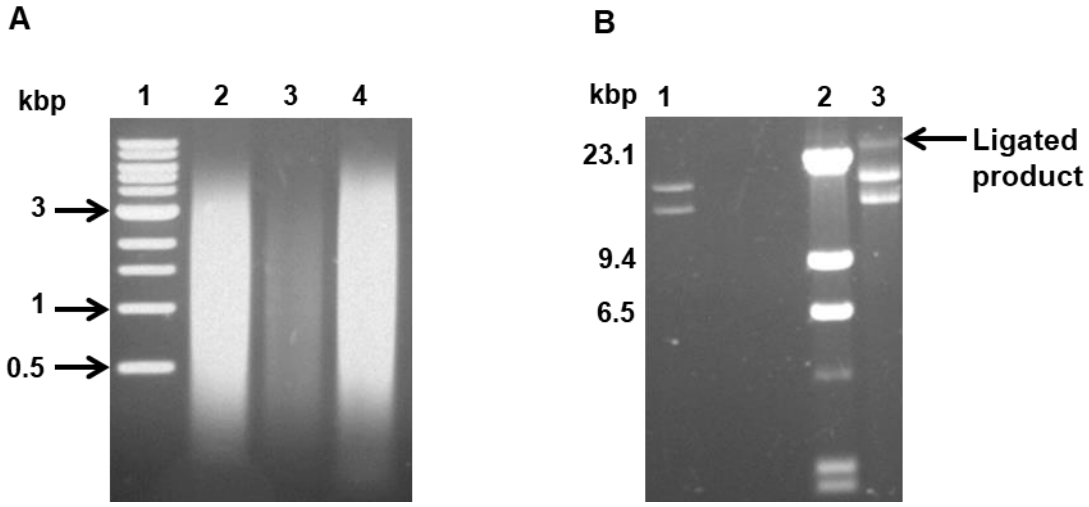
Running Title: Carbohydrate binding domains from a metagenomic human fecal phage display library

Figure S1



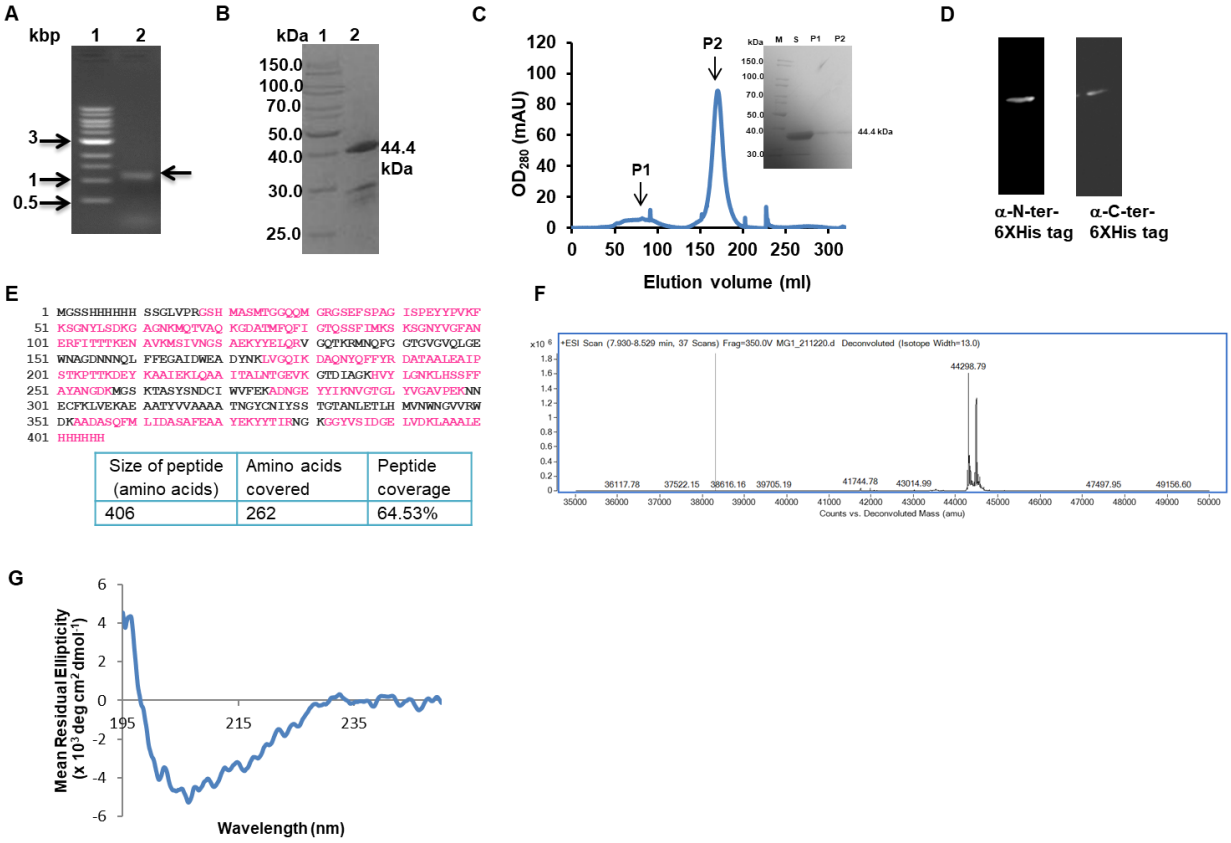
Generation of *Sr*NaFLD recombinant phages and optimization of screening protocol. (A) Agarose gel electrophoresis indicating the amplicon in *Sr*NaFLD recombinant T7 phage. Lane 1: 1 kbp DNA marker (NEB). Lane 2: PCR amplicon amplified from the *Sr*NaFLD T7 phage DNA. (B) Agarose gel electrophoresis of PCR amplicons indicating the presence or absence of the *Sr*NaFLD DNA insert in phages extracted from nine plaques randomly selected following each of the six cycles of the biopanning procedure performed. The biopanning was initiated using *Sr*NaFLD recombinant phages and non-recombinant T7 phages in a ratio $1x10^{3}$: $1x10^{11}$ (*Sr*NaFLD-T7 phage: T7Select 10-3b phage). Lane 10 in gels of cycles 1, 3 and 5, and lane 1 in gels of cycles 2, 4, and 6 have 1 kbp DNA molecular marker. DNA bands corresponding to *Sr*NaFLD DNA insert (~900 bp) are observed in phage plaques following biopanning in cycle 6. (C) Agarose gel electrophoresis indicating the results of the fourth round of the biopanning of *Sr*NaFLD recombinant phages and non-recombinant T7 phages in different ratios $1x10^{5}$: $1x10^{11}$, $1x10^{7}$: $1x10^{11}$, $1x10^{9}$: $1x10^{11}$ (*Sr*NaFLD-T7 phage: T7 10-3b phage). Lane 1: 1 kbp DNA marker. Lanes 2-10: PCR amplicons of the phages extracted from 9 different plaques after biopanning. (D) Agarose gel electrophoresis indicating the results of the biopanning of *Sr*NaFLD recombinant phages and non-recombinant T7 phages in ratio $1x10^{7}$: $1x10^{11}$ (*Sr*NaFLD-T7 phage: T7 10-3b phage) with 1%SDS used as the elution agent. Lane 1: 1 kbp DNA marker. Lanes 2-10: PCR products of the phages extracted from 9 different plaques after 6 rounds of biopanning.

Figure S2



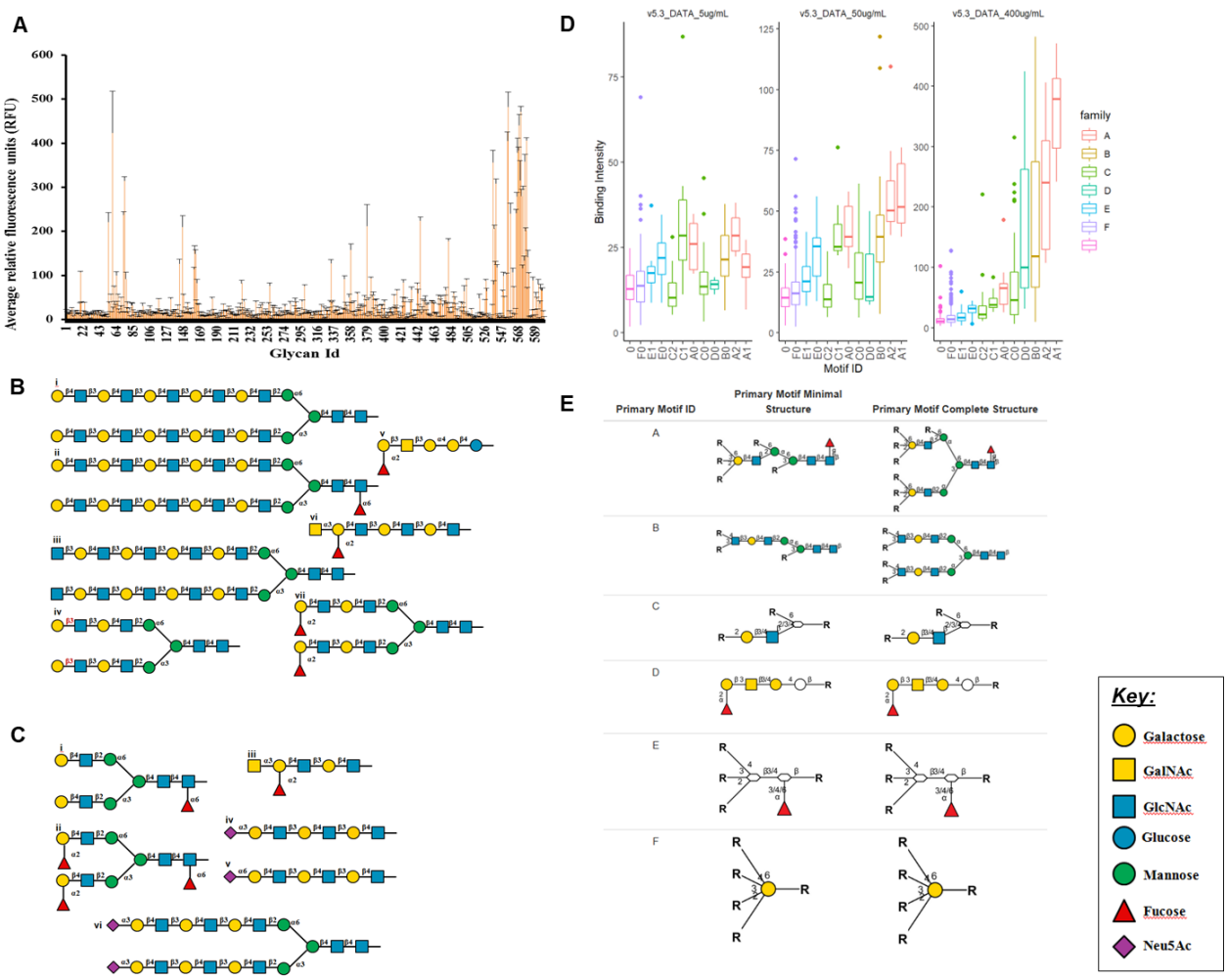
Construction of metagenomic phage display library. (A) Agarose gel electrophoresis indicating fragmentation of metagenomic DNA. Lane 1: 1 kbp DNA marker from NEB. Lanes 2-4: Fragments of 500 to 300 base pairs of the metagenomic DNA isolated from different kits produced by sonication. (B) Agarose gel electrophoresis indicating ligation of metagenomic fragments into T7Select 10-3b vector. Lane 1: SmaI (NEB) digested T7Select 10-3b vector DNA. Lane 2: λ DNA-HindIII Digest (marker DNA) from NEB. Lane 3: Ligation reaction mix containing a band >23.1 kbp confirming the blunt end ligation of the fragmented metagenomic DNA into the T7Select 10-3b vector.

Figure S3



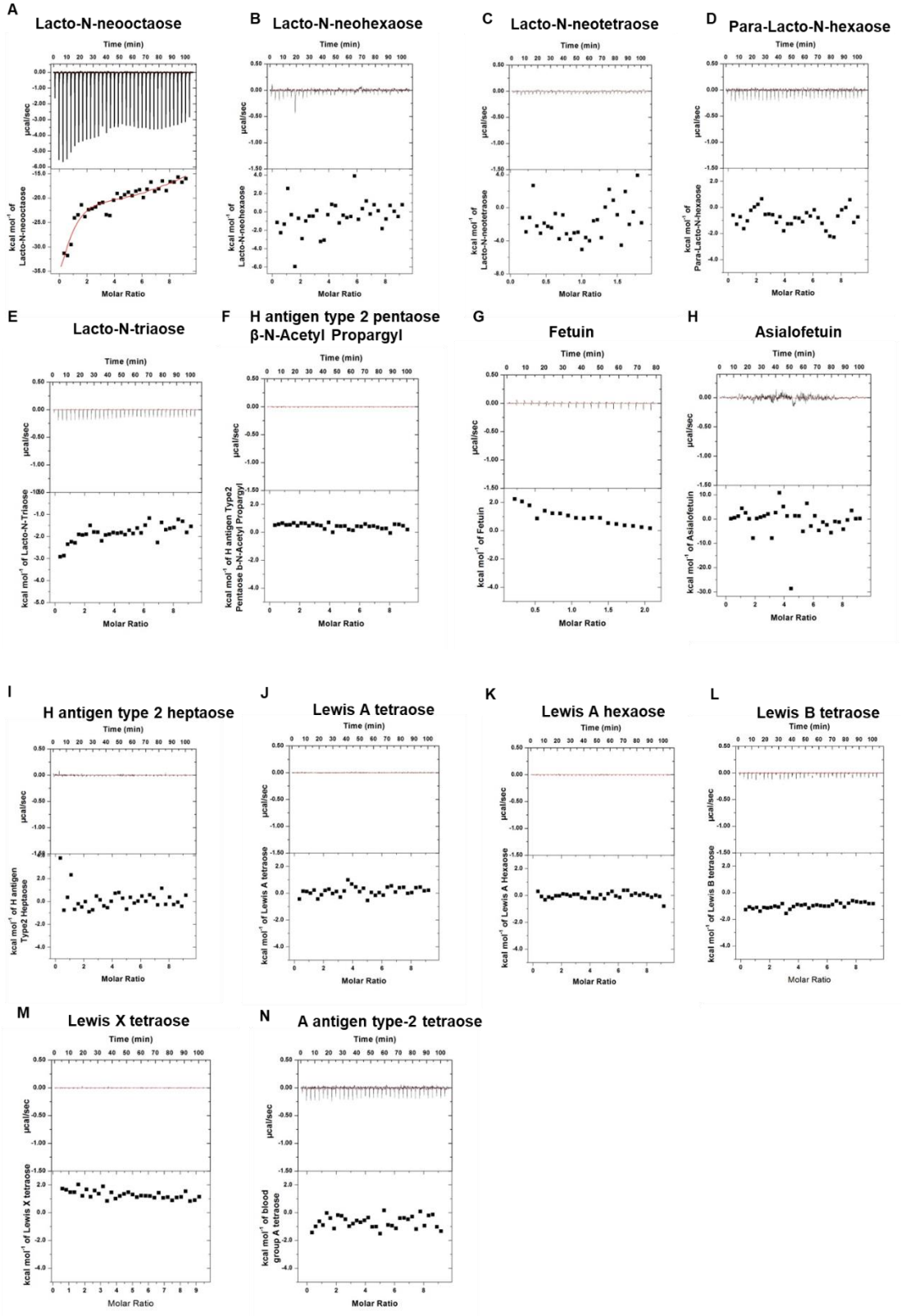
Cloning, expression, and purification of MG1. (A) PCR amplification. Lane 1: DNA marker, Lane 2: PCR amplicon of MG1 clone. (B) Coomassie stained SDS-PAGE. Lane 1: Molecular mass marker, Lane 2: Purified recombinant MG1 protein. (C) Size Exclusion Chromatogram of protein using Hiprep 26/60 Sephacryl S-200 High Resolution column with eluted peak fractions P1 and P2, and input protein sample S analyzed by SDS-PAGE in inset. (D) Western analysis of recombinant MG1 protein using N-terminal anti-His antibody and C-terminal anti-His antibody. (E) Sequence coverage obtained with MS/MS spectrometry of recombinant MG1 protein followed by targeted sequence search; the peptide sequence coverage is shown in pink color. Results of a single mass spectrometry run. (F) Deconvoluted spectrum showing the intact masses of recombinant MG1 protein using Agilent G6550A Quadrupole Time of Flight (Q-TOF) mass spectrometer. (G) Circular dichroism spectrum of recombinant MG1 protein.

Figure S4

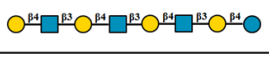
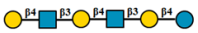
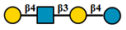
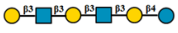
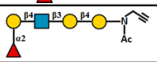
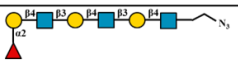
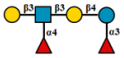


Glycan binding by purified recombinant MG1. (A) Bar plot showing the glycan binding profile of MG1. Glycan array analysis was performed on CFG microarray v5.3 with 400 µg/ml MG1. Results of a single glycan array experiment. Spots with highest and lowest fluorescence intensity (of the eight spots for each glycan) were removed. Error bars represent standard deviations. (B) Glycans found to be bound by MG1. Results of Glycopattern tool using three glycan array datasets of the same protein preparation at 5, 50 and 400 µg/ml. (i-ii) Glycans with terminal Galactose (the top two binders); (iii) Glycan with terminal GlcNAc; (iv) Glycan having type-1 extension unit; (v) Glycan with type-4 H antigen; (vi) Glycan with type-2 A antigen; (vii) Glycan with type-2 H antigen. (C) Glycans found not to bind to MG1. Results of Glycopattern tool using three glycan array datasets of the same protein preparation at 5, 50 and 400 µg/ml. (i-iii) Glycans with short LacNAc chains. (iv-vi) Glycans having terminal acidic residues. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (Varki, A., et al. 2015). (D) Concentration dependent binding of MG1 to glycan motifs identified by Motif Finder. Results of Motif Finder analysis using three glycan array datasets of the same protein preparation at 5, 50 and 400 µg/ml. (E) List of primary glycan motifs identified by Motif Finder. Results of Motif Finder analysis using three glycan array datasets of the same protein preparation at 5, 50 and 400 µg/ml.

Figure S5


A
Lacto-N-neooctaose
B
Lacto-N-neohexaose
C
Lacto-N-neotetraose
D
Para-Lacto-N-hexaose
E
Lacto-N-triaose
F
H antigen type 2 pentaose
β-N-Acetyl Propargyl
G
Fetuin
H
Asialofetuin
I
H antigen type 2 heptaose
J
Lewis A tetraose
K
Lewis A hexaose
L
Lewis B tetraose
M
Lewis X tetraose
N
A antigen type-2 tetraose
Time (min)
µcal/sec
Molar Ratio

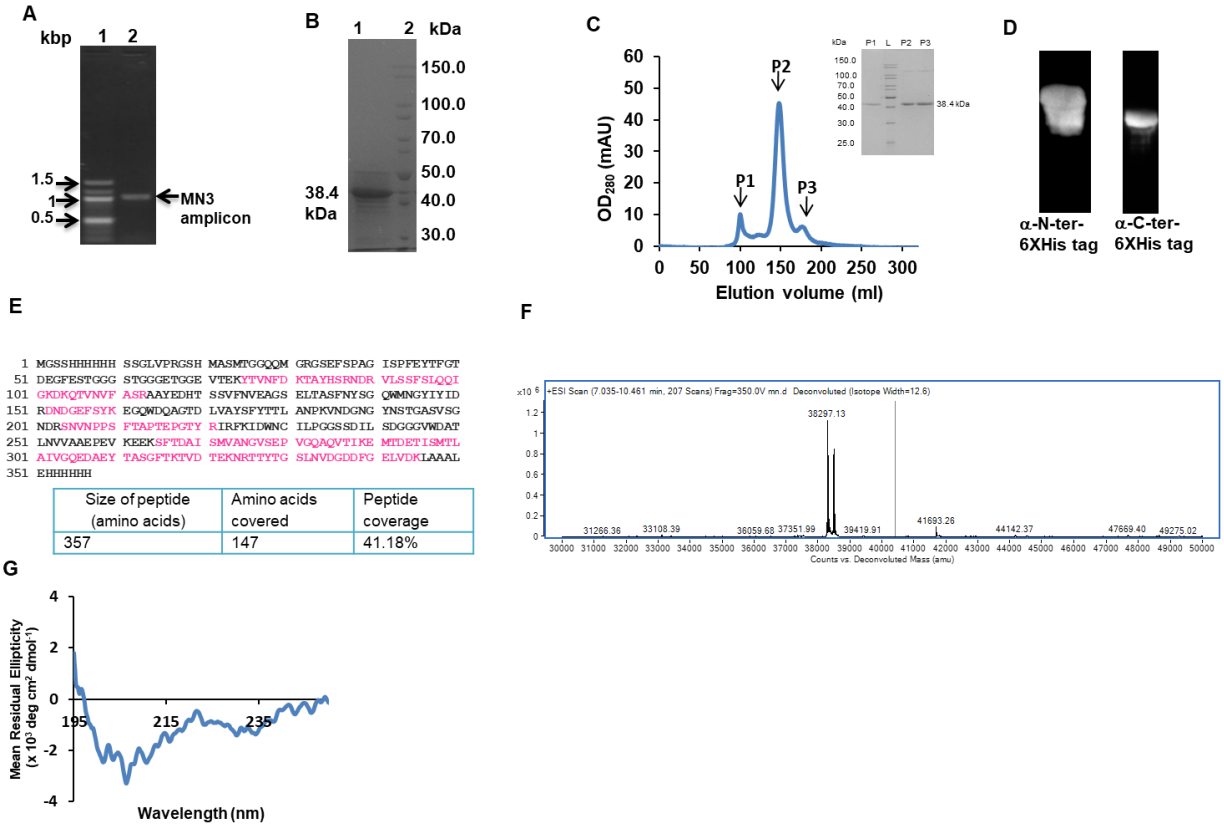
O

Ligand	Structure of Ligand	Protein concentration (mM)	Ligand concentration (mM)	Conclusion
Gly023 (Lacto-N-neooctaose)		0.03	1.5	Inconclusive (High heats were observed in control reaction when Gly023 was injected into buffer in the cell)
Gly022 (Lacto-N-neohexaose)		0.03	1.5	Binding not detected
Gly021 (Lacto-N-neotetraose)		0.05	2.5	Binding not detected
Gly012 (Para-Lacto-N-hexaose)		0.03	1.5	Binding not detected
Gly011 (Lacto-N-triaose) *		0.01	0.5	Binding not detected
Gly035-2 (A antigen type-2 tetraose)		0.03	1.5	Binding not detected
Gly033-2NPR (H antigen pentaose type-2-β-N-Acetyl Propargyl) *		0.01	0.5	Binding not detected
Fetuin	Mainly triantennary oligosaccharides	0.015	0.5	Binding not detected
Asialofetuin	Mainly triantennary oligosaccharides without terminal sialic acid residues	0.015	0.5	Binding not detected
TE135 (H antigen type-2 heptaose azido ethyl) *		0.01	0.5	Binding not detected
Gly054 (Lewis A tetraose) *		0.01	0.5	Binding not detected
Gly055 (Lacto-N-difucohexaose II) *		0.01	0.5	Binding not detected
Gly045 (Lewis B tetraose) *		0.01	0.5	Binding not detected
Gly050 (Lewis X tetraose) *		0.01	0.5	Binding not detected

Temperature of binding assays was 20 °C except for * where the temperature was 30 °C

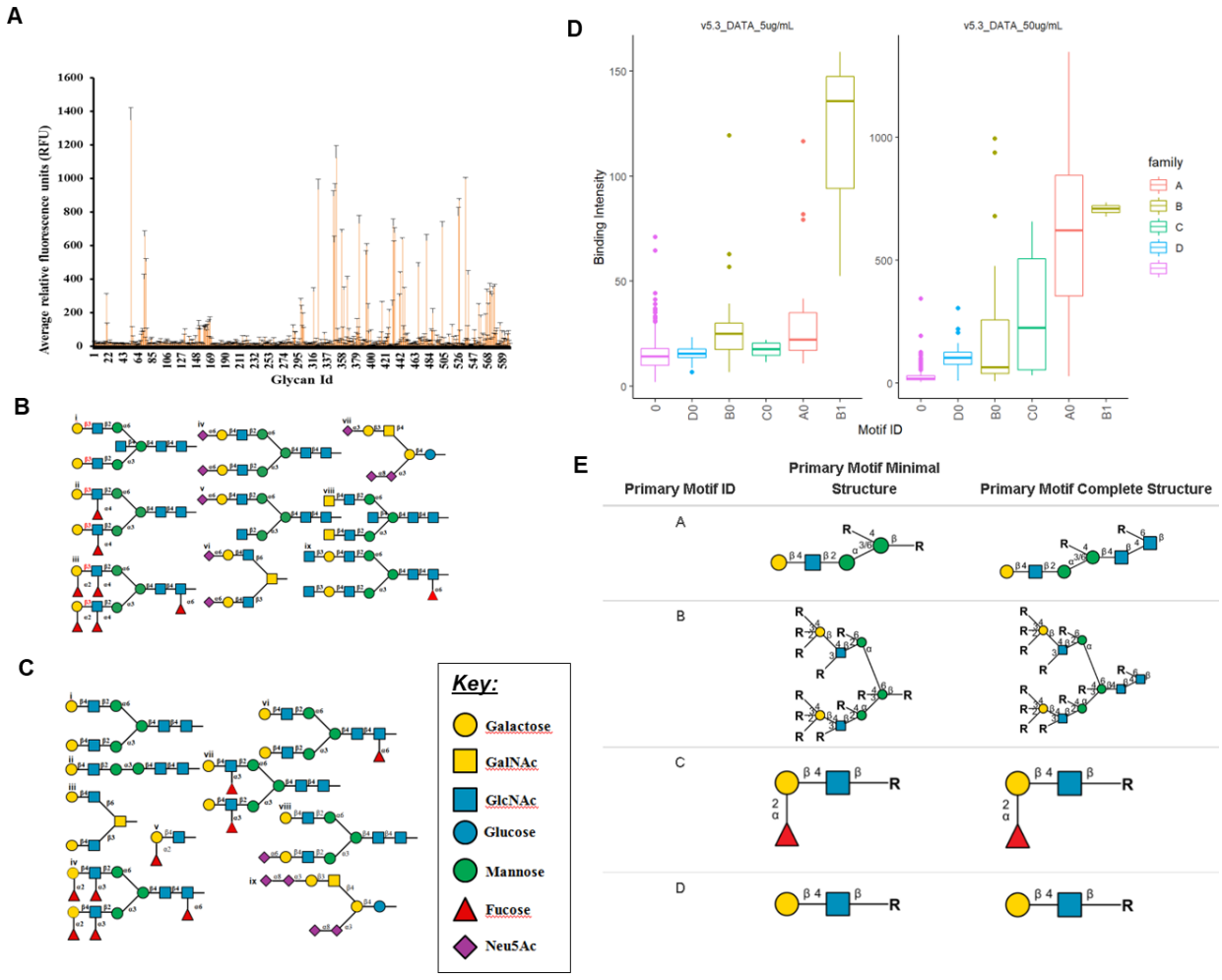
Isothermal calorimetry of MG1. (A-N) Data for titrations of MG1 (loaded in the cell) with various glycans (loaded in the syringe) tabulated (O) as per the conditions mentioned. All data shown are results of single ITC runs. Two control titrations were performed for each titration – glycan injected into buffer (in the cell), and buffer injected into protein (in the cell). The first injection was removed for all the plots. The control titration in which buffer was injected into protein solution was used as the control, and subtracted from the data prior to plotting, except for the titration of MG1 with Lacto-N-triaose, where the control titration in which sugar was injected into buffer solution was used to subtract the background.

Figure S6



Cloning, expression, and purification of MN3. (A) PCR amplification. Lane 1: DNA marker, Lane 2: PCR amplicon of MN3 clone. (B) Coomassie stained SDS-PAGE. Lane 1: Molecular mass marker, Lane 2: Purified recombinant MN3 protein. (C) Size Exclusion Chromatogram of protein using Hiprep 26/60 Sephacryl S-200 High Resolution column with eluted peak fractions P1, P2, and P3 analyzed by SDS-PAGE in inset. (D) Western analysis of recombinant MN3 protein using N-terminal anti-His antibody and C-terminal anti-His antibody. (E) Sequence coverage obtained with MS/MS spectrometry of recombinant MN3 protein followed by targeted sequence search; the peptide sequence coverage is shown in pink color. Results of a single mass spectrometry run. (F) Deconvoluted spectrum showing the intact masses of recombinant MN3 protein using Agilent G6550A Quadrupole Time of Flight (Q-TOF) mass spectrometer. (G) Circular dichroism spectrum of recombinant MN3 protein.

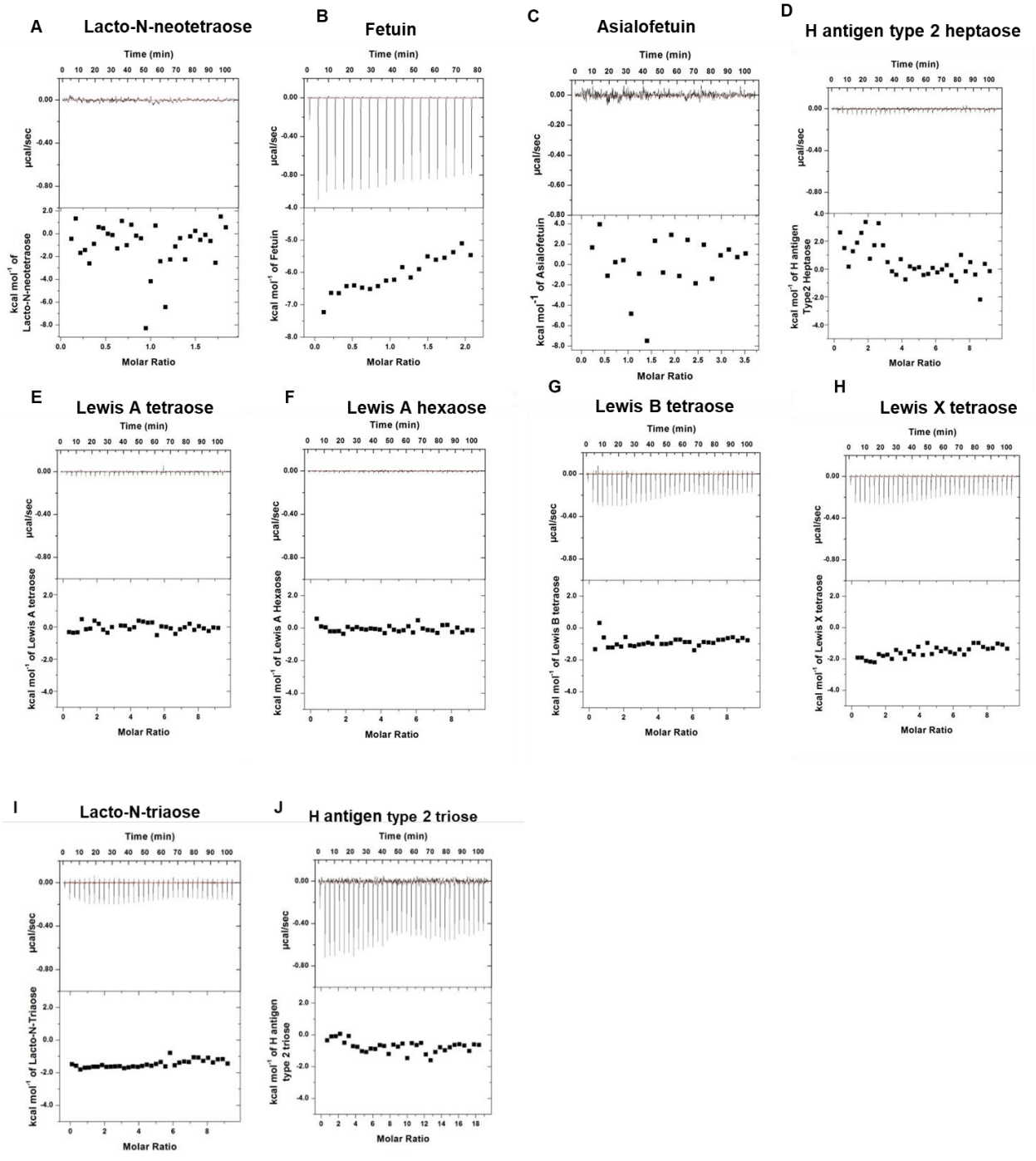
Figure S7



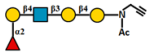
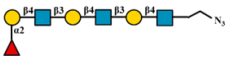
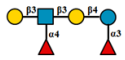
Glycan binding by purified recombinant MN3. (A) Bar plot showing the glycan binding profile of MN3. Glycan array analysis was performed on CFG microarray v5.3 with 50 µg/ml MN3. Results of a single glycan array experiment. Spots with highest and lowest fluorescence intensity (of the eight spots for each glycan) were removed. Error bars represent standard deviations. (B) Glycans found to be bound by MN3. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (i-iii) Glycans with LacNAc unit with galactose in terminal position(s); (iv-vii) Glycans with fucose residues; (iv) Glycan with core fucose; (v) Glycan with type-2 H antigen; (v) Glycan with Lewis X antigen; (vi) Glycan with Lewis Y antigen; (viii) Bi-antennary glycan with one antenna having LacNAc and terminal Galactose and other antenna with terminal sialic acid; (ix) Glycan with terminal disialic acid sequences with α2-8 linkages. (C) Glycans found not to bind to MN3. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (i-iii) Glycans with type-1 glycan unit; (iv-vi) Glycans with terminal acidic residues or GlcNAc and no galactose; (vii) Glycan with only one α2-8 linkage in the bi-antennary structure; (viii) Glycan with terminal galactose replaced by GalNAc; (ix) Glycan with terminal GlcNac. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (Varki, A., et al. 2015). (D) Concentration dependent binding of MN3 to glycan motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (E) List of primary glycan

motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml.

Figure S8


A
Lacto-N-neotetraose
B
Fetuin
C
Asialofetuin
D
H antigen type 2 heptaose
E
Lewis A tetraose
F
Lewis A hexaose
G
Lewis B tetraose
H
Lewis X tetraose
I
Lacto-N-triaose
J
H antigen type 2 triose
Time (min)
µcal/sec
Molar Ratio
kcal mol-1 of Lacto-N-neotetraose
kcal mol-1 of Fetuin
kcal mol-1 of Asialofetuin
kcal mol-1 of H antigen Type2 Heptaose
kcal mol-1 of Lewis A tetraose
kcal mol-1 of Lewis A Hexaose
kcal mol-1 of Lewis B tetraose
kcal mol-1 of Lewis X tetraose
kcal mol-1 of Lacto-N-Triaose
kcal mol-1 of H antigen type 2 triose

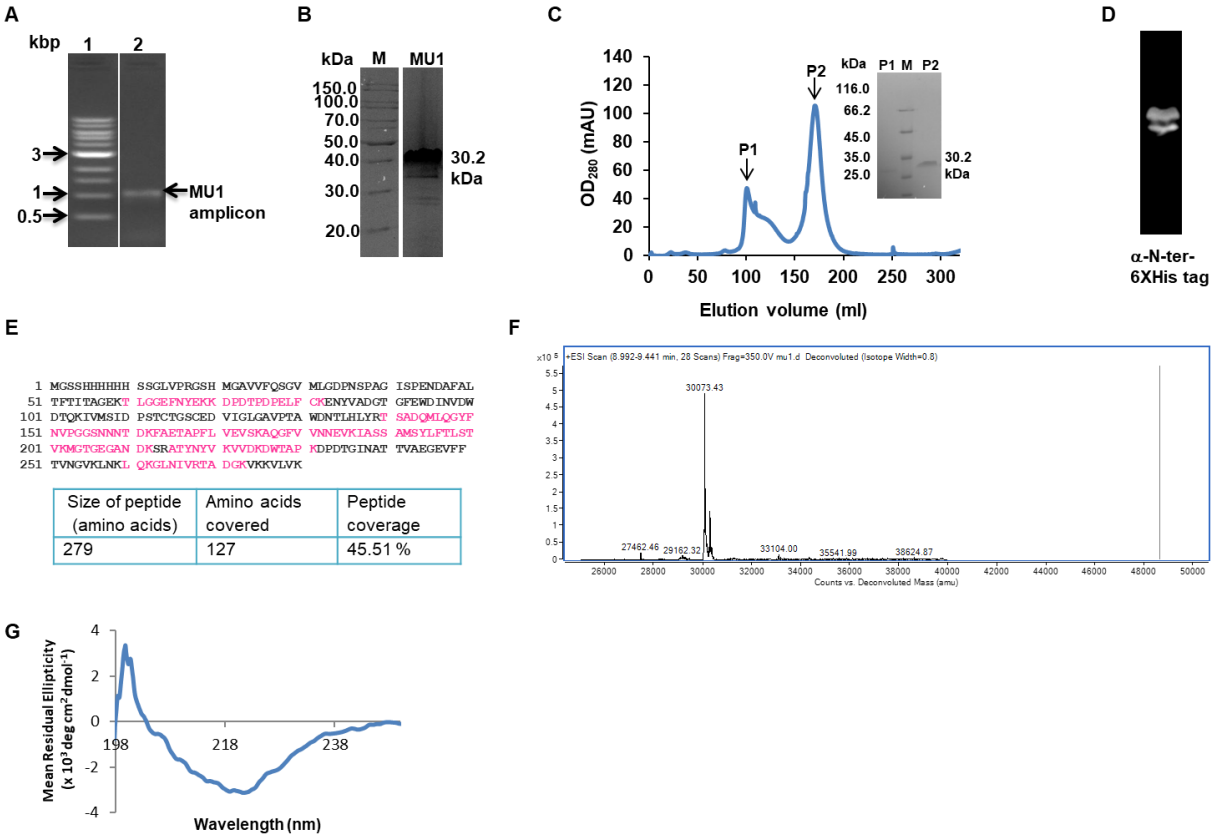
K

Ligand	Structure of Ligand	Protein concentration (mM)	Ligand concentration (mM)	Conclusion
Gly021 (Lacto-N-neotetraose)		0.05	2.5	Binding not detected
Gly011 (Lacto-N-triaose) *		0.01	0.5	Binding not detected
Gly033-2NPR (H antigen pentaose type-2-β-N-Acetyl Propargyl)		0.05	2.5	**Binding**
Gly031-2 (H Antigen type-2 triaose)		0.03	3	Binding not detected
Fetuin	Mainly triantennary oligosaccharides	0.05	0.5	Binding not detected
Asialofetuin	Mainly triantennary oligosaccharides without terminal sialic acid residues	0.015	0.5	Binding not detected
TE135 (H antigen type-2 heptaose azido ethyl) *		0.01	0.5	Binding not detected
Gly054 (Lewis A tetraose) *		0.01	0.5	Binding not detected
Gly055 (Lacto-N-difucohexaose II) *		0.01	0.5	Binding not detected
Gly045 (Lewis B tetraose) *		0.01	0.5	Binding not detected
Gly050 (Lewis X tetraose) *		0.01	0.5	Binding not detected

Temperature of binding assays was 20 °C except for * where the temperature was 30 °C

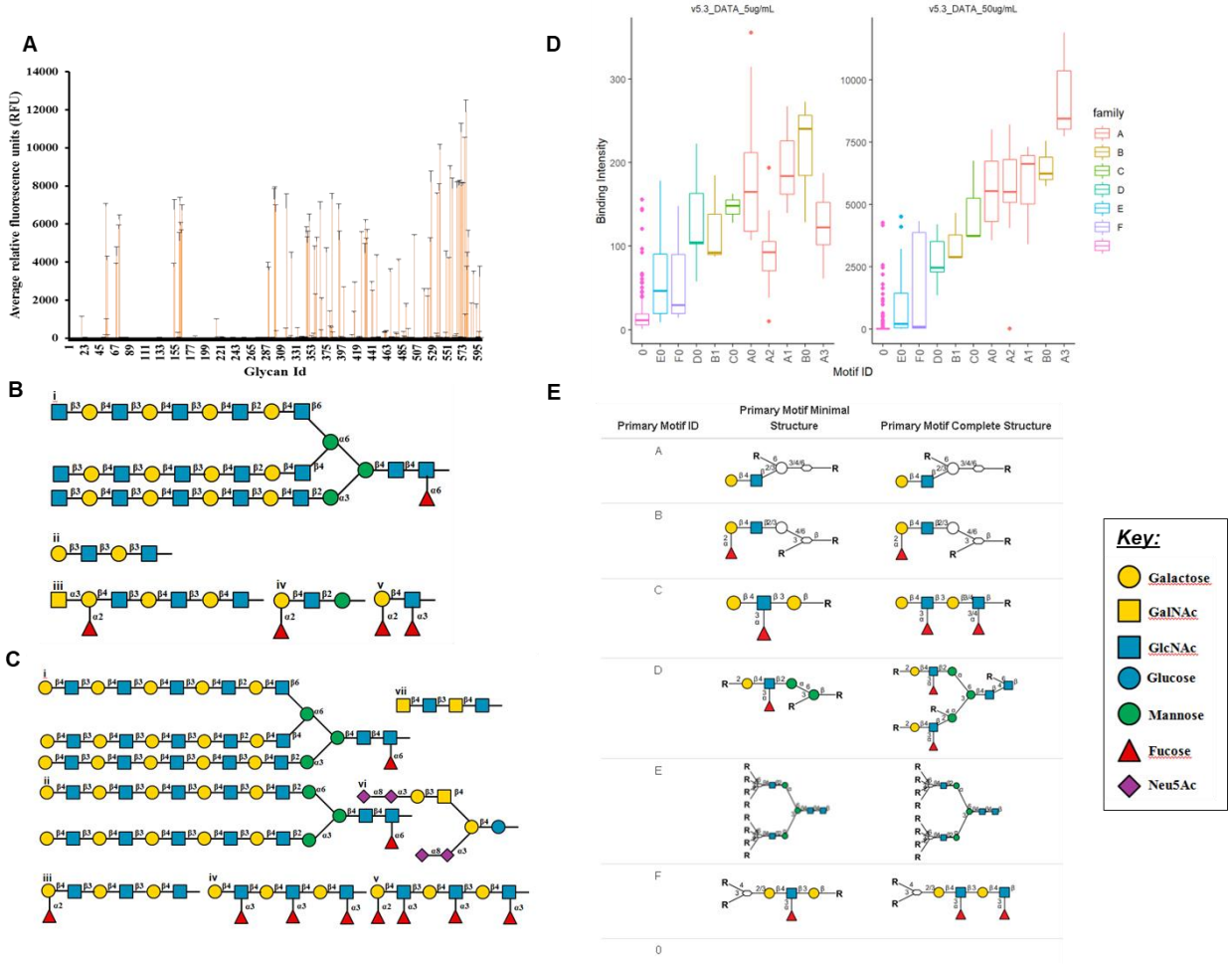
Isothermal calorimetry of MN3. (A-J) Data for titrations of MN3 (loaded in the cell) with various glycans (loaded in the syringe) tabulated (K) as per the conditions mentioned. All data shown are results of single ITC runs. Two control titrations were performed for each titration – glycan injected into buffer (in the cell), and buffer injected into protein (in the cell). The first injection was removed for all the plots. The control titration in which buffer was injected into protein solution was used as the control. For the titration of MN3 with Lewis A tetraose, the data point corresponding to injection 13 was removed.

Figure S9



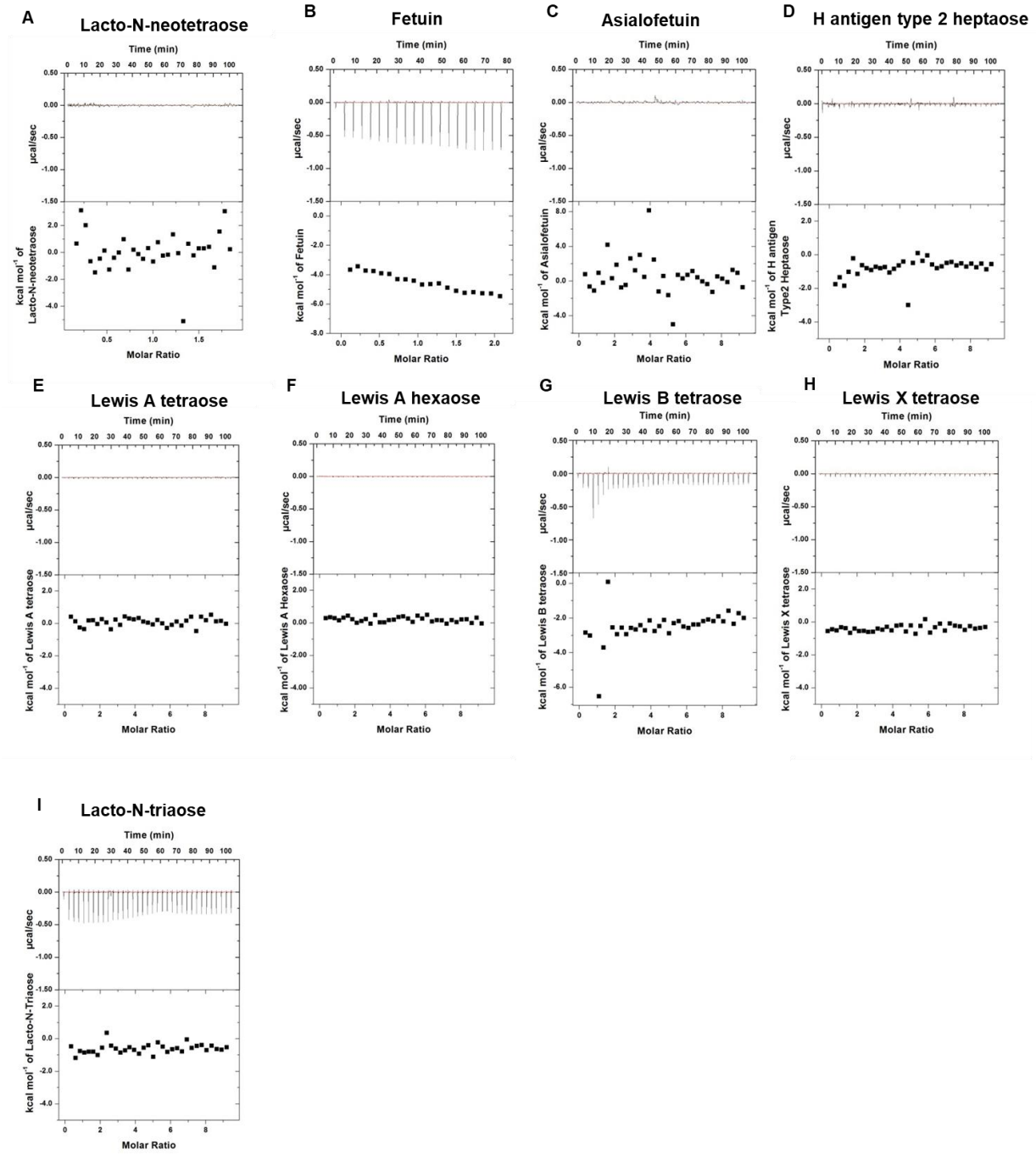
Cloning, expression, and purification of MU1. (A) PCR amplification. Lane 1: DNA marker, Lane 2: PCR amplicon of MU1 clone. (B) Coomassie stained SDS-PAGE. Lane 1: Molecular mass marker, Lane 2: Purified recombinant MU1 protein. (C) Size Exclusion Chromatogram of protein using Hiprep 26/60 Sephacryl S-200 High Resolution column with eluted peak fractions P1 and P2 analyzed by SDS-PAGE in inset. (D) Western analysis of recombinant MU1 protein using N-terminal anti-His antibody and C-terminal anti-His antibody. (E) Sequence coverage obtained with MS/MS spectrometry of recombinant MU1 protein followed by targeted sequence search; the peptide sequence coverage is shown in pink color. Results of a single mass spectrometry run. (F) Deconvoluted spectrum showing the intact masses of recombinant MU1 protein using Agilent G6550A Quadrupole Time of Flight (Q-TOF) mass spectrometer. (G) Circular dichroism spectrum of recombinant MU1 protein.

Figure S10

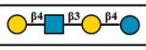
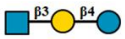
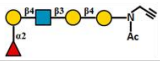
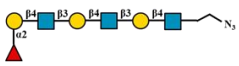
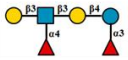


Glycan binding by purified recombinant MU1. (A) Bar plot showing the glycan binding profile of MU1. Glycan array analysis was performed on CFG microarray v5.3 with 50 µg/ml MU1. Results of a single glycan array experiment. Spots with highest and lowest fluorescence intensity (of the eight spots for each glycan) were removed. Error bars represent standard deviations. (B) Glycans found to be bound by MU1. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (i-ii) Glycans with terminal galactose residue(s); top two binders of MU1; they are also core fucosylated; (iii-iv) Glycans with fucose residue(s); (iii) Type-2 H antigen, (iv) Lewis-X antigen, (v) Lewis-Y antigen; (vi) Glycan with terminal disialic acid sequences with α2-8 linkages; (vii) Glycan with LacDiNAc units. (C) Glycans found not to bind to MU1. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (i) Glycans with terminal GlcNAc residue; (ii) Glycans with type-1 glycan units. (iii) Glycan with terminal type-2 A antigen. (iv-v) Fucose containing glycans with short LacNAc chains. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (Varki, A., et al. 2015). (D) Concentration dependent binding of MU1 to glycan motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (E) List of primary glycan motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml.

Figure S11


A
Lacto-N-neotetraose
B
Fetuin
C
Asialofetuin
D
H antigen type 2 heptaose
E
Lewis A tetraose
F
Lewis A hexaose
G
Lewis B tetraose
H
Lewis X tetraose
I
Lacto-N-triaose
Time (min)
µcal/sec
Molar Ratio
kcal mol-1 of Lacto-N-neotetraose
kcal mol-1 of Fetuin
kcal mol-1 of Asialofetuin
kcal mol-1 of H antigen Type2 Heptaose
kcal mol-1 of Lewis A tetraose
kcal mol-1 of Lewis A Hexaose
kcal mol-1 of Lewis B tetraose
kcal mol-1 of Lewis X tetraose
kcal mol-1 of Lacto-N-Triaose

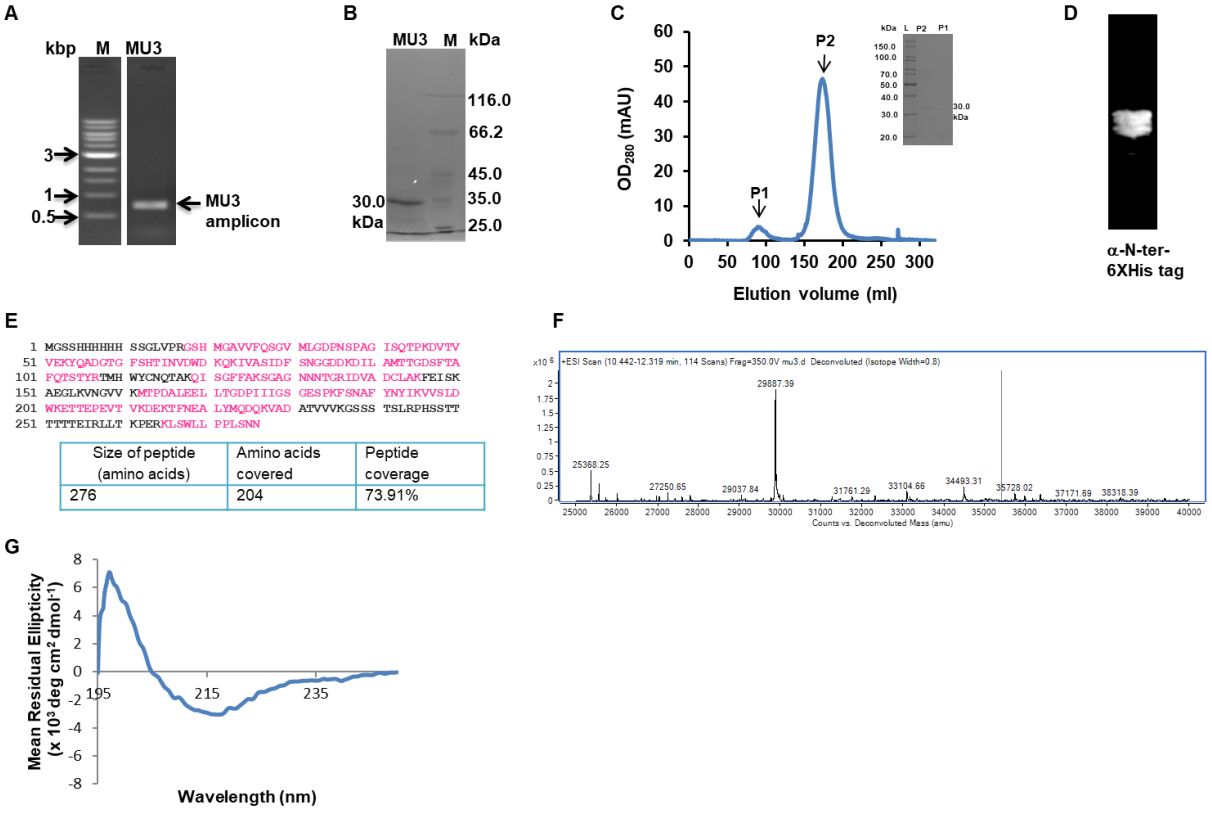
J

Ligand	Structure of Ligand	Protein concentration (mM)	Ligand concentration (mM)	Conclusion
Gly021 (Lacto-N-neotetraose)		0.05	2.5	Binding not detected
Gly011 (Lacto-N-triaose) *		0.01	0.5	Binding not detected
Gly033-2NPR (H antigen pentaose type-2-β-N-Acetyl Propargyl)		0.05	2.5	**Binding**
Fetuin	Mainly triantennary oligosaccharides	0.05	0.5	Binding not detected
Asialofetuin	Mainly triantennary oligosaccharides without terminal sialic acid residues	0.015	0.5	Binding not detected
TE135 (H antigen type-2 heptaose azido ethyl) *		0.01	0.5	Binding not detected
Gly054 (Lewis A tetraose) *		0.01	0.5	Binding not detected
Gly055 (Lacto-N-difucohexaose II) *		0.01	0.5	Binding not detected
Gly045 (Lewis B tetraose) *		0.01	0.5	Binding not detected
Gly050 (Lewis X tetraose) *		0.01	0.5	Binding not detected

Temperature of binding assays was 20 ^{o}C except for * where the temperature was 30 ^{o}C

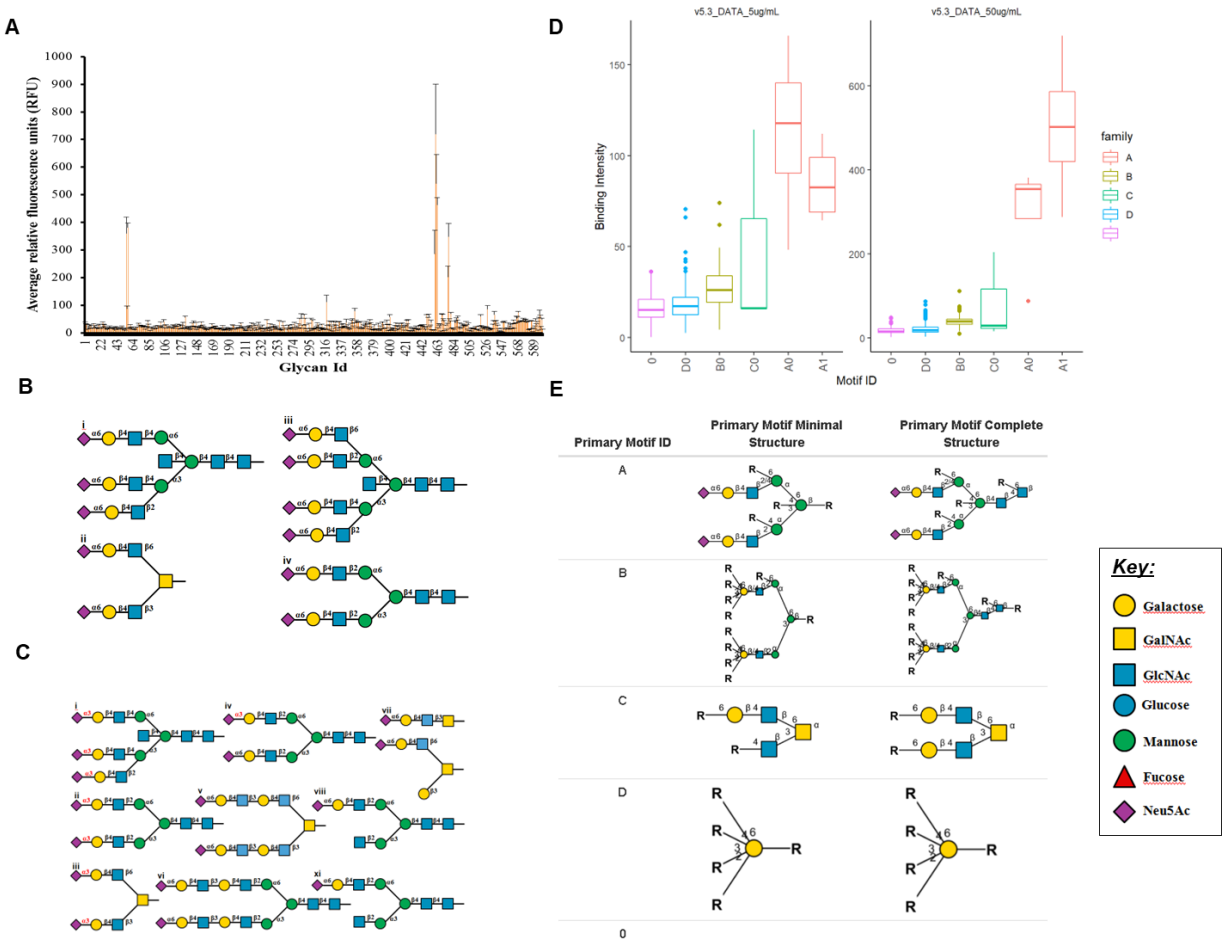
Isothermal calorimetry of MU1. (A-I) Data for titrations of MU1 (loaded in the cell) with various glycans (loaded in the syringe) tabulated (J) as per the conditions mentioned. All data shown are results of single ITC runs. Two control titrations were performed for each titration – glycan injected into buffer (in the cell), and buffer injected into protein (in the cell). The first injection was removed for all the plots. The control titration in which buffer was injected into protein solution was used as the control.

Figure S12



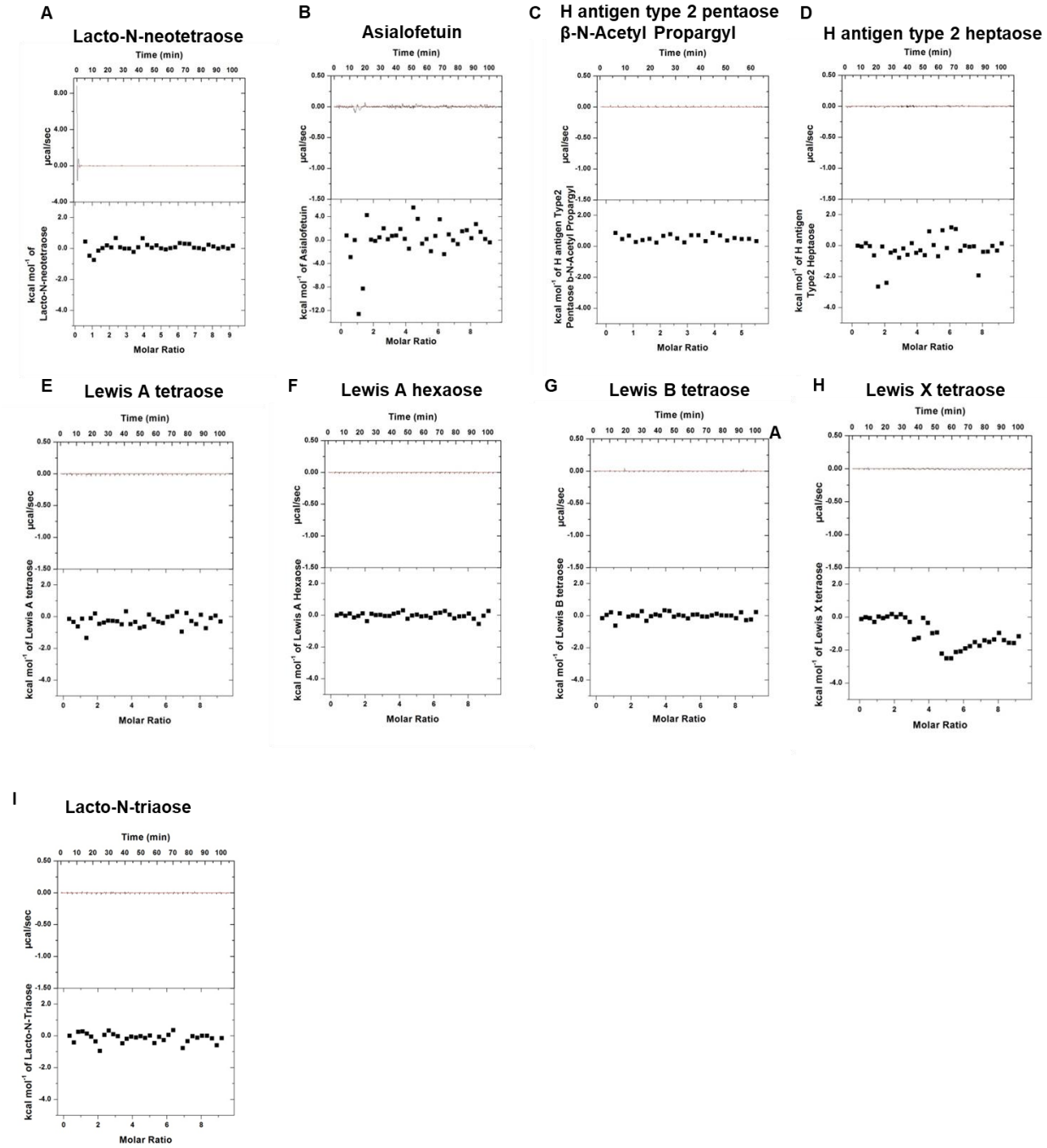
Cloning, expression, and purification of MU3. (A) PCR amplification. Lane 1: DNA marker, Lane 2: PCR amplicon of MU3 clone. (B) Coomassie stained SDS-PAGE. Lane 1: Molecular mass marker, Lane 2: Purified recombinant MU3 protein. (C) Size Exclusion Chromatogram of protein using Hiprep 26/60 Sephacryl S-200 High Resolution column with eluted peak fractions P1 and P2 analyzed by SDS-PAGE in inset. (D) Western analysis of recombinant MU3 protein using N-terminal anti-His antibody and C-terminal anti-His antibody; (E) Sequence coverage obtained with MS/MS spectrometry of recombinant MU3 protein followed by targeted sequence search; the peptide sequence coverage is shown in pink color. Results of a single mass spectrometry run. (F) Deconvoluted spectrum showing the intact masses of recombinant MU3 protein using Agilent G6550A Quadrupole Time of Flight (Q-TOF) mass spectrometer. (G) Circular dichroism spectrum of recombinant MU3 protein.

Figure S13

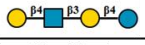
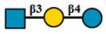
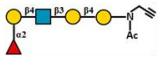
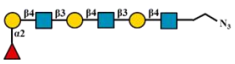
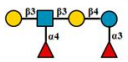


Glycan binding by purified recombinant MU3. (A) Bar plot showing the glycan binding profile of MU3. Glycan array analysis was performed on CFG microarray v5.3 with 50 µg/ml MU3. Results of a single glycan array experiment. Spots with highest and lowest fluorescence intensity (of the eight spots for each glycan) were removed. Error bars represent standard deviations. (B) Glycans found to be bound by MU3. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (i-iv) All the glycans have terminal sialic acid residue(s) linked via α2-3 linkage; (C) Glycans found not to bind to MU3. Results of Glycopattern tool using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (v-x) Glycans with terminal sialic acid residues linked via α2-6 linkage; (v-vi) Biantennary glycans with two LacNAc units on each arm; (vii) Glycans without more than one antenna. (viii-x) Glycans without sialic acid in α2-6 linkage in one of the two arms. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (Varki, A., et al. 2015). (D) Concentration dependent binding of MU3 to glycan motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml. (E) List of primary glycan motifs identified by Motif Finder. Results of Motif Finder analysis using two glycan array datasets of the same protein preparation at 5 and 50 µg/ml.

Figure S14


A
Lacto-N-neotetraose
B
Asialofetuin
C
H antigen type 2 pentaose
β-N-Acetyl Propargyl
D
H antigen type 2 heptaose
E
Lewis A tetraose
F
Lewis A hexaose
G
Lewis B tetraose
H
Lewis X tetraose
I
Lacto-N-triaose
Time (min)
µcal/sec
Molar Ratio
kcal mol-1 of Lacto-N-neotetraose
kcal mol-1 of Asialofetuin
kcal mol-1 of H antigen Type2 Pentaose b-N-Acetyl Propargyl
kcal mol-1 of H antigen Type2 Heptaose
kcal mol-1 of Lewis A tetraose
kcal mol-1 of Lewis A Hexaose
kcal mol-1 of Lewis B tetraose
kcal mol-1 of Lewis X tetraose
kcal mol-1 of Lacto-N-Triaose

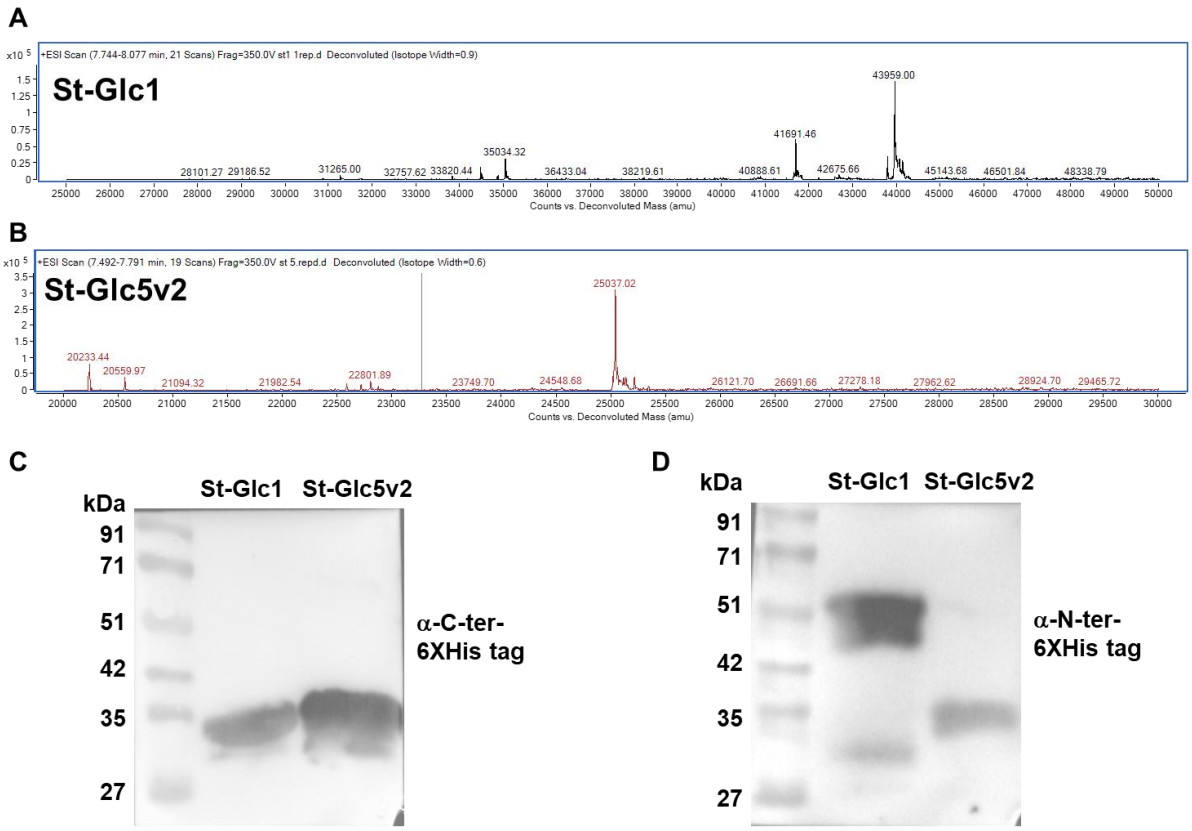
J

Ligand	Structure of Ligand	Protein concentration (mM)	Ligand concentration (mM)	Conclusion
Gly021 (Lacto-N-neotetraose)		0.05	2.5	Binding not detected
Gly011 (Lacto-N-triaose) *		0.051	0.5	Binding not detected
Gly033-2NPR (H antigen pentaose type-2-β-N-Acetyl Propargyl) *		0.01	0.5	Binding not detected
Fetuin	Mainly triantennary oligosaccharides	0.05	0.5	**Binding**
Asialofetuin	Mainly triantennary oligosaccharides without terminal sialic acid residues	0.01	0.5	Binding not detected
TE135 (H antigen type-2 heptaose azido ethyl) *		0.01	0.5	Binding not detected
Gly054 (Lewis A tetraose) *		0.01	0.5	Binding not detected
Gly055 (Lacto-N-difucohexaose II) *		0.01	0.5	Binding not detected
Gly045 (Lewis B tetraose) *		0.01	0.5	Binding not detected
Gly050 (Lewis X tetraose) *		0.01	0.5	Binding not detected

Temperature of binding assays was 20 °C except for * where the temperature was 30 °C

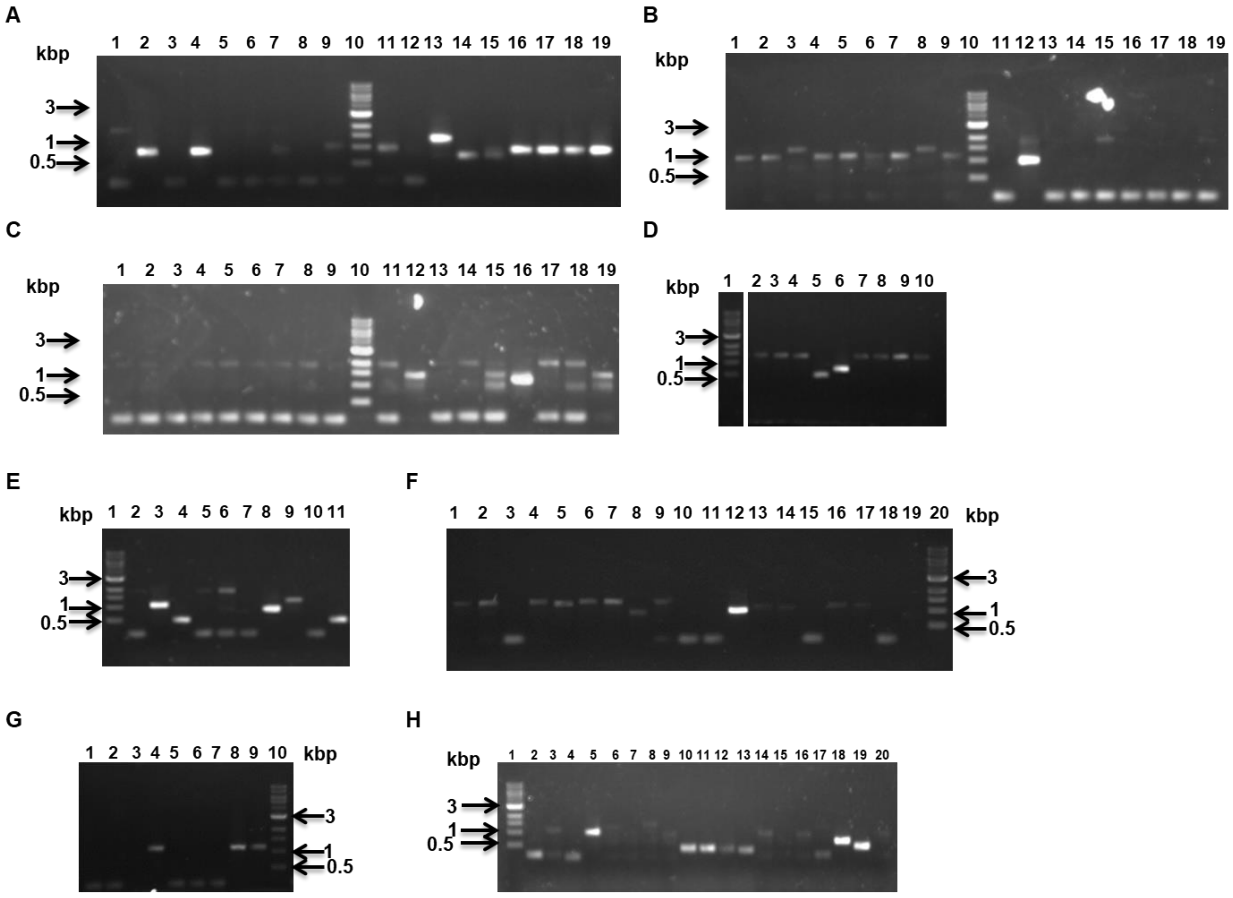
Isothermal calorimetry of MU3. (A-I) Data for titrations of MU3 (loaded in the cell) with various glycans (loaded in the syringe) tabulated (J) as per the conditions mentioned. All data shown are results of single ITC runs. Two control titrations were performed for each titration – glycan injected into buffer (in the cell), and buffer injected into protein (in the cell). The control titration in which buffer was injected into protein solution was used as the control, and subtracted from the data prior to plotting, except for the titration of MU3 with Lacto-N-triaose, where the control titration in which sugar was injected into buffer solution was used to subtract the background. The first injection was removed for all the plots. For the titration of MU3 with Lewis B tetraose, the data point corresponding to injection 7 was removed.

Figure S15



Recombinant St-Glc1 and St-Glc5v2 proteins. (A) Intact mass analysis of St-Glc1 by ESI-mass spectrometry. (B) Intact mass analysis of St-Glc5v2 by ESI-mass spectrometry. (C) Western analysis of St-Glc1 and St-Glc5v2 by anti-C-ter-6XHis tag antibody. (D) Western analysis of St-Glc1 and St-Glc5v2 by anti-N-ter-6XHis tag antibody.

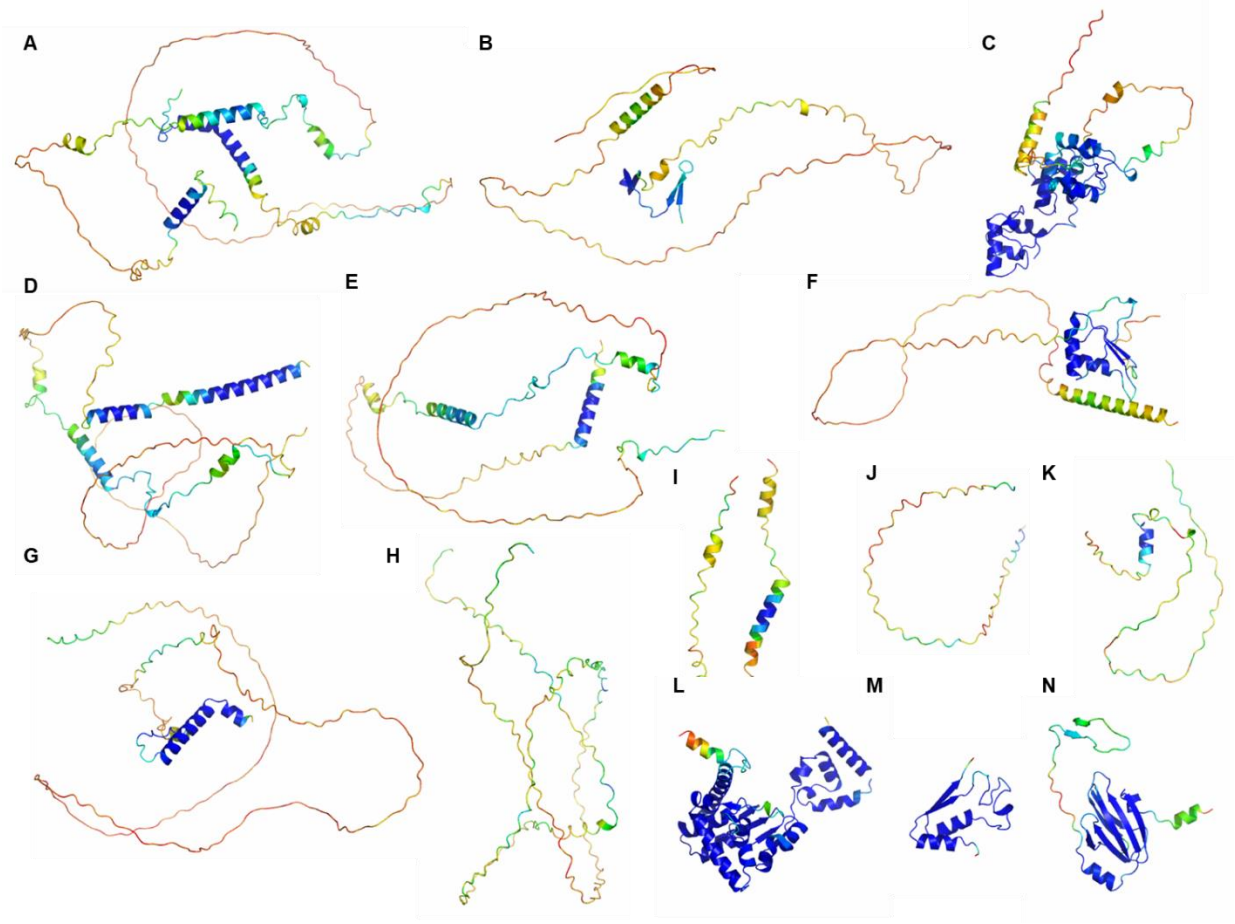
Figure S16



Identification of phage clones following screening against various carbohydrates. (A-J) Agarose gel electrophoresis of PCR amplicons indicating the presence or absence of any recombinant insert following biopanning of the metagenomic phage display library against various carbohydrates. (A) Biopanning against *Staphylococcus aureus* peptidoglycan and elution with GlcNAc or *Staphylococcus aureus* peptidoglycan. Results of a single screening experiment. Lanes 1-9: PCR amplicons of clones StapPG-GlcNAc1 to StapPG-GlcNAc9. Lane 10: 1 kbp DNA marker. Lanes 11-19: PCR amplicons of clones StapPG-StapPG1 to StapPG-StapPG9. (B) Biopanning against *Methanobacterium sp.* peptidoglycan and elution with GlcNAc or *Methanobacterium sp.* peptidoglycan. Results of a single screening experiment. Lanes 1-9: PCR amplicons of clones MethPG-GlcNAc1 to MethPG-GlcNAc9. Lane 10: 1 kbp DNA marker. Lanes 11-19: PCR amplicons of clones MethPG-MethPG1 to MethPG-MethPG9. (C) Biopanning against inulin and elution with D-fructose or D-glucose. Results of a single screening experiment. Lanes 1-9: PCR amplicons of clones Iln-Fru1 to Iln-Fru9. Lane 10: 1 kbp DNA marker. Lanes 11-19: PCR amplicons of clones Iln-Glc1 to Iln-Glc9. (D) Biopanning against inulin and elution with inulin. Results of a single screening experiment. Lane 1: 1 kbp DNA marker. Lanes 2-9: PCR amplicons of clones Iln-Iln1 to Iln-Iln9. (E) Biopanning against arabinogalactan and elution with D-galactose. Results of a single screening experiment. Lane 1: 1 kbp DNA marker. Lanes 2-10: PCR amplicons

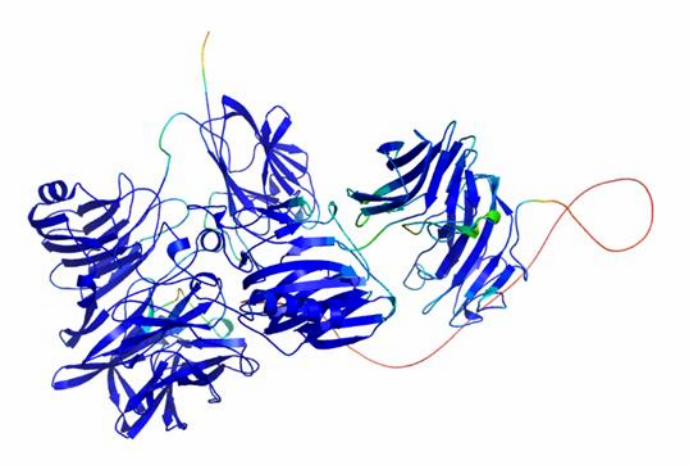
of clones AG-Gal1 to AG-Gal9. (F) Biopanning against arabinogalactan and elution with L-arabinose or arabinogalactan. Results of a single screening experiment. Lanes 1-9: PCR amplicons of clones AG-Ara1 to AG-Ara9. Lanes 10-19: PCR amplicons of clones AG-AG1 to AG-AG9. Lane 20: 1 kbp DNA marker. (G) Biopanning against PAA-b-D-Galactose and elution with D-galactose. Results of a single screening experiment. Lanes 1-9: PCR amplicons of clones PBGal-Gal1 to PBGal-Gal9. Lane 10: 1 kbp DNA marker. (H) Biopanning against PAA-b-D-GlcNAc and elution with D-Glucose. Results of a single screening experiment. Lane 1: 1 kbp DNA marker. Lanes 2-20: PCR amplicons of clones PBGlcNAc-Glc1 to PBGlcNAc-Glc9.

Figure S17



AlphaFold2 predicted structural models. (A) AG-Gal2. (B) Iln-Glc6. (C) PBGal-Gal8. (D) MethPG-MethPG8. (E) StapPG-GlcNAc6. (F) StapPG-StapPG6. (G) StapPG-StapPG7. (H) PBGlcNAc-GlcNAc4F1. (I) PBGlcNAc-GlcNAc8. (J) PBGlcNAc-GlcNAc9F1. (K) PBGlcNAc-GlcNAc17F1. (L) PBGlcNAc-GlcNAc4F2. (M) PBGlcNAc-GlcNAc9F2. (N) PBGlcNAc-GlcNAc17F2.

Figure S18



AlphaFold2 predicted structural model of *Paenibacillus macerans* (*Bacillus macerans*) protein (GenBank: AAG47946.1) with CBM38 and GH32 cycloinulooligosaccharide fructanotransferase domains.