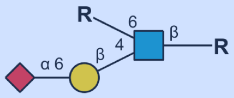
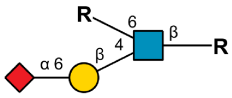
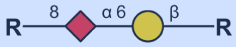
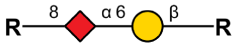
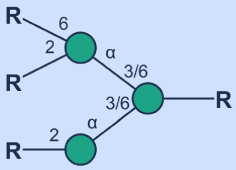
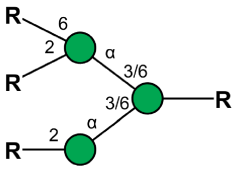
SNA

Automated Analysis report for SNA evaluated at 10 ug/mL

List of Primary SNA Motifs

Primary Motifs are the main families of glycan binding. These are given a letter identifier.

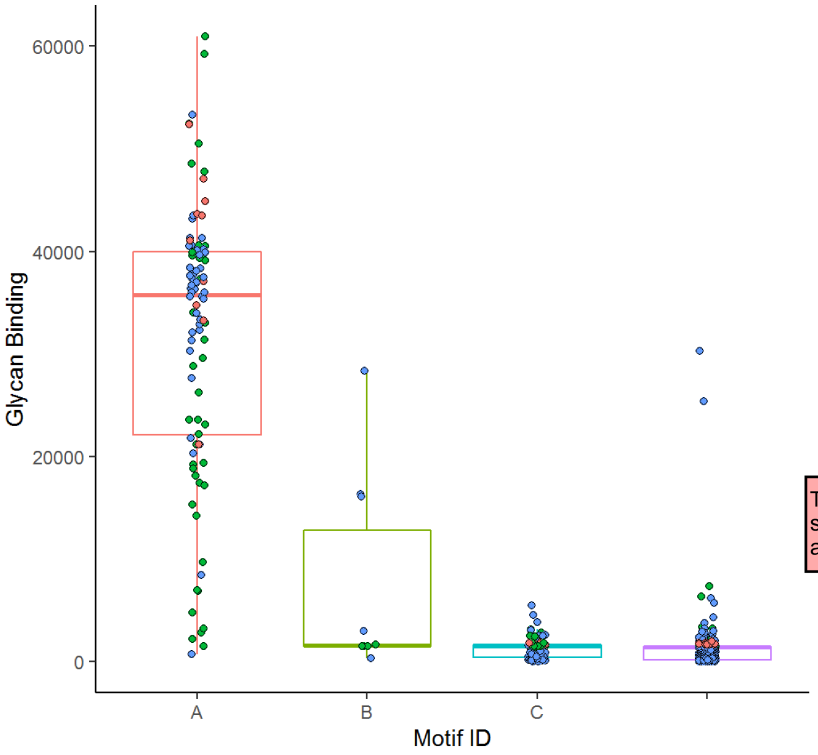
Primary Motif ID	Primary Motif Minimal Structure	Primary Motif Complete Structure	Relative Binding	Number of Glycans
A			1.00	92
B			0.23	10
C			0.04	202
0			0.04	941

Minimal and complete motif structures match the same glycans on the array, these motifs are the least detailed and most detailed respectively. Here they are the same, indicating the array tested the motifs thoroughly!

Relative binding is the average binding of glycans with the motif at the given concentration (10ug/mL). These are normalized so that the highest binding motif is “1".

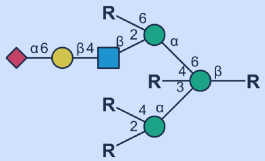
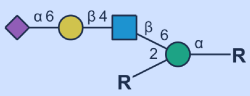
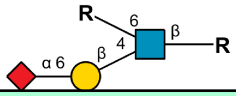
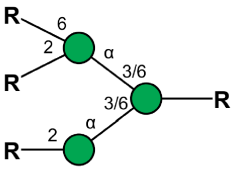
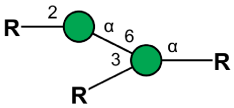
Minimal and complete m… the array. Com… excluded in the minimal … ties and subsitu… tested by the array, they…

Boxplot of Primary Motifs



This boxplot presents the data summarized in the first table, with a separate box for each primary motif. When there are multiple arrays analyzed together these points are distinguished by different colors.

List of Fine-Specificity SNA Motifs

Motif ID	Nearest Common Name (Accuracy%**)	Motif Graphic Structure	Relative Binding	Number of Glycans	P-Value***
A4*	a6 Sialyl Type 2 LacNAc N-Glycan a6 Core Fucose (100%)		1.00	5	<0.001
A3	Tetraantennary N-Glycan (98%)		0.99	5	<0.001
A2	a6 Sialyl Type 2 LacNAc (94%)		0.76	22	<0.001
A0*	a6 Sialyl Type 2 LacNAc (97%)		0.72	54	<0.001
A1	a6 Sialyl Type 2 LacNAc Lacto Glycosphingolipid (100%)			6	<0.001
B0	a6 Sialyl Galactose (93%)		0.15	10	<0.001
C0*	N-Glycan (91%)		0.00	179	0.997
C1	Terminal Mannose (98%)		0.00	23	1.000
0	Non-Binders (100%)		0.00	941	NA

Fine specificity motifs are the variations on primary motifs which have unique degrees of binding. Fine specificity motifs are identified by a letter corresponding to the primary motif and a number

Motifs denoted with an asterisk (*) next to the ID indicate that the motif captures the remaining glycans which lack the more specific motifs. ie. Motif "A0*" matches glycans with the primary motif "A" but not motifs A1-A4. See model structure for more details.

Common names are generated by attempting to match known motifs to the generated motif. The accuracy of these names is given in parentheses.

Some motifs may have a red ID, this indicates the motif is determined to be a non-binder motif, likely to exclude certain glycans from binding motifs.

P-value shown here is the result of Dunnet's test for multiple comparisons. Each test compares motifs to the non-binder group.

Man GalNAc Neu5Gc Monosaccharide ~SKIP~
Gal Fuc Sia Hydrogen or Monosaccharide R-

See Symbol Nomenclature for Glycans (SNFG) for complete key: https://www.ncbi.nlm.nih.gov/glycans/snfg.html (https://www.ncbi.nlm.nih.gov/glycans/snfg.html)

*Motif indicates the remaining glycans not matched by motifs which are a subset. Motif definition needs to be taken in the context of the model.

**Accuracy describes the consistency between common-name definition of the motif and the formal, text-based definition of the motif, in

terms of percent agreement in the glycans containing the two motifs. Common Name label definitions given here (http://carbogrove.org/MotifLabels.php).

***P-Value refers to difference from Non-Binders with multiple testing correction (Dunnet's Test)

Motifs with a red motif ID fail to show a logistic response to protein concentration in the range of concentrations analyzed. These motifs may be nonbinding motifs (motifs which define nonbinding exceptions) or simply fail to fit. Nonbinding motifs are determined based on concentration dependent response when available or the average binding of non-motif glycans otherwise.

Boxplot of Fine-Specificity Motifs

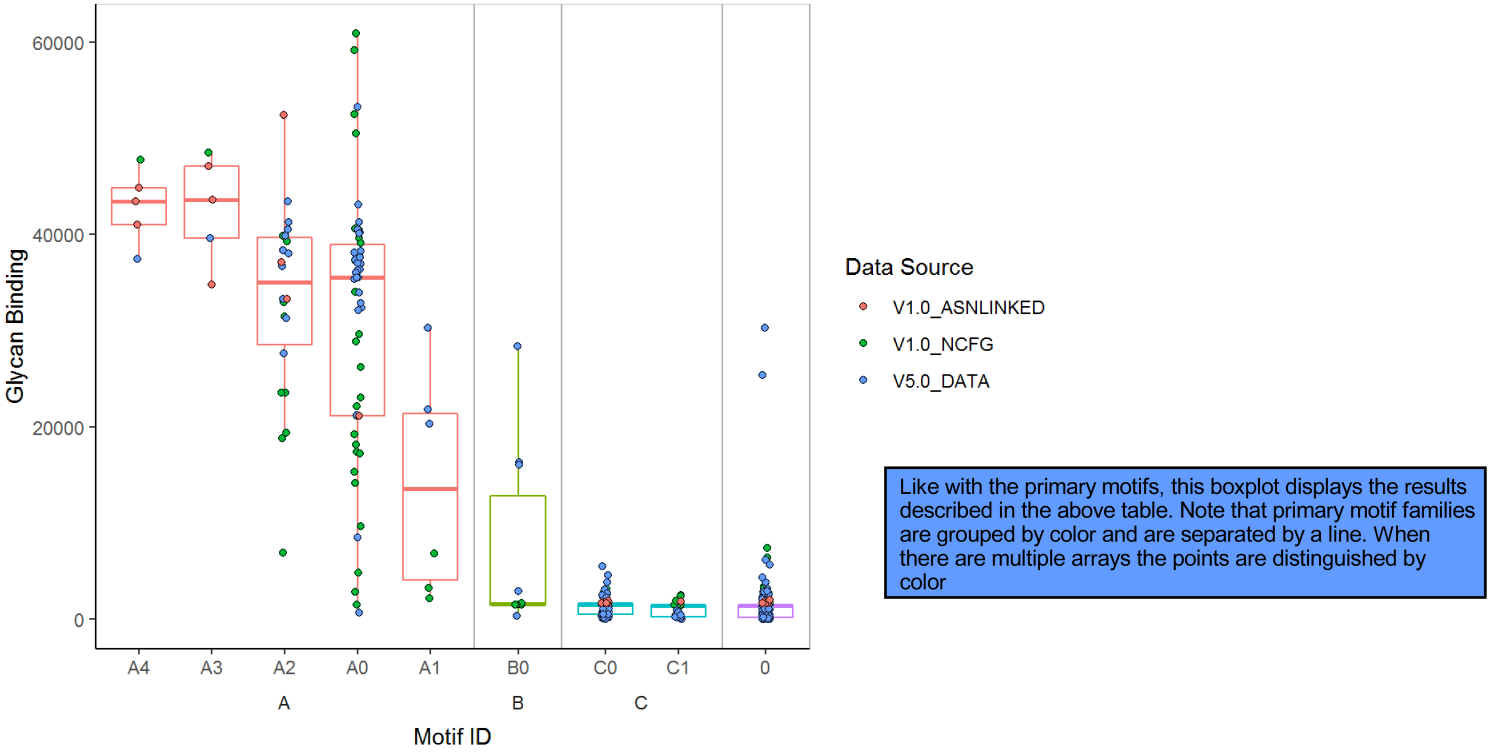


Figure 1. Glycan binding grouped by motif and motif family. Individual glycans are given as points on the plot.

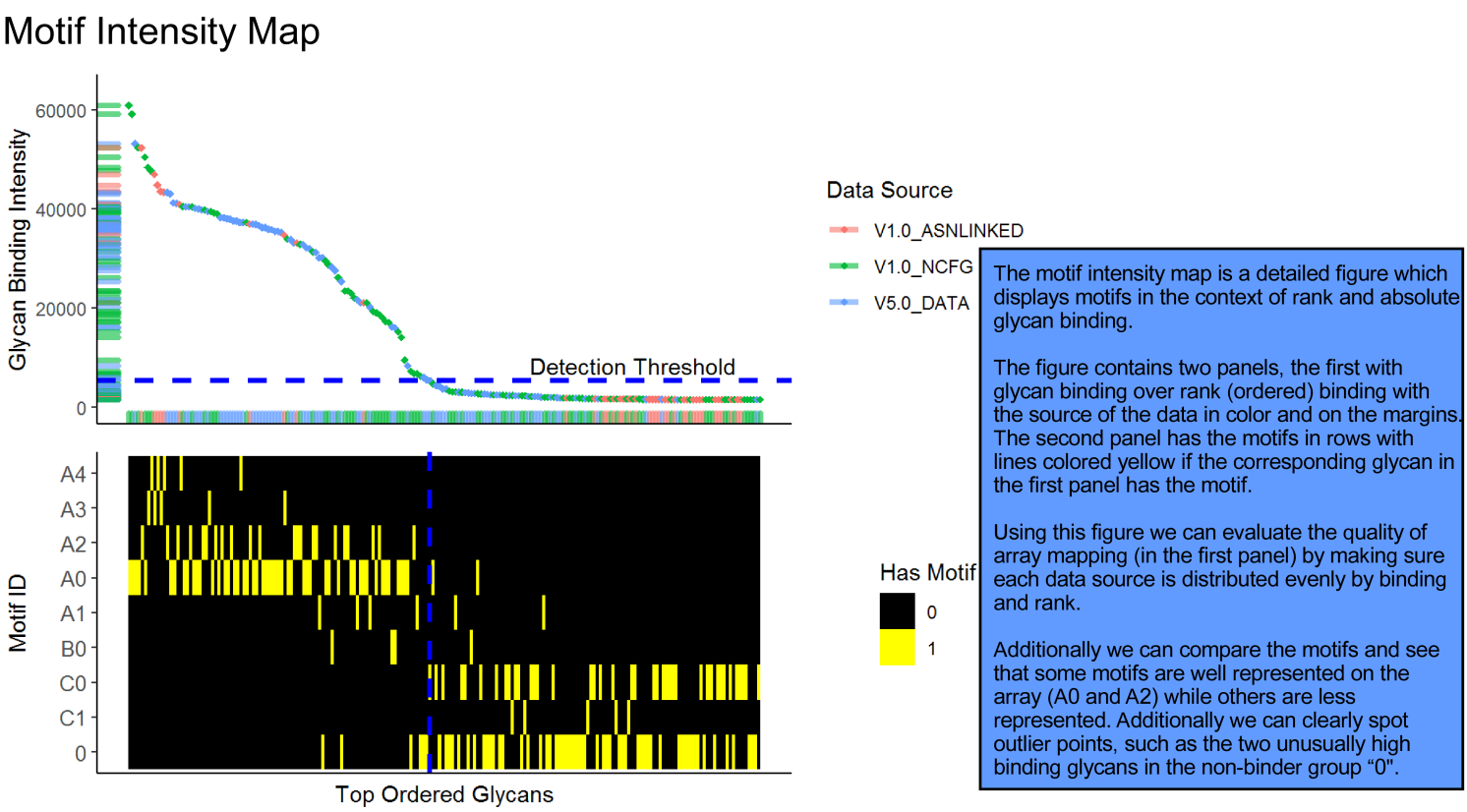
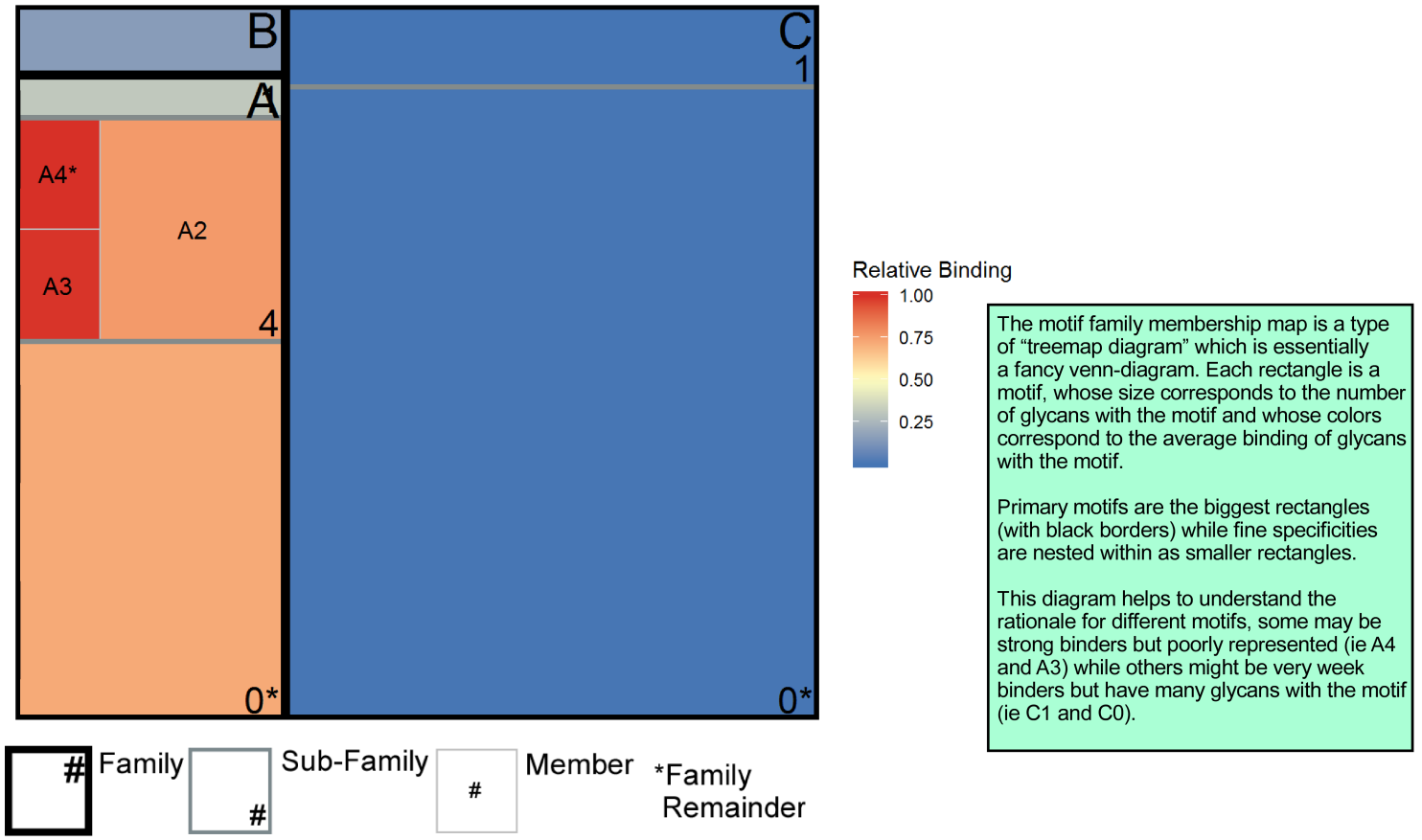


Figure 2. Glycan intensity and motif distribution plot. The top half of the plot presents the observed glycan binding intensity of various glycans used in the array over their rank binding intensity; only the top glycans are shown. The second plot indicates the position of glycans containing the various motifs in the top plot with a yellow tick.

Motif Family Membership Map



*Motif indicates the remaining glycans not matched by motifs which are a subset. Motif definition needs to be taken in the context of the

model.

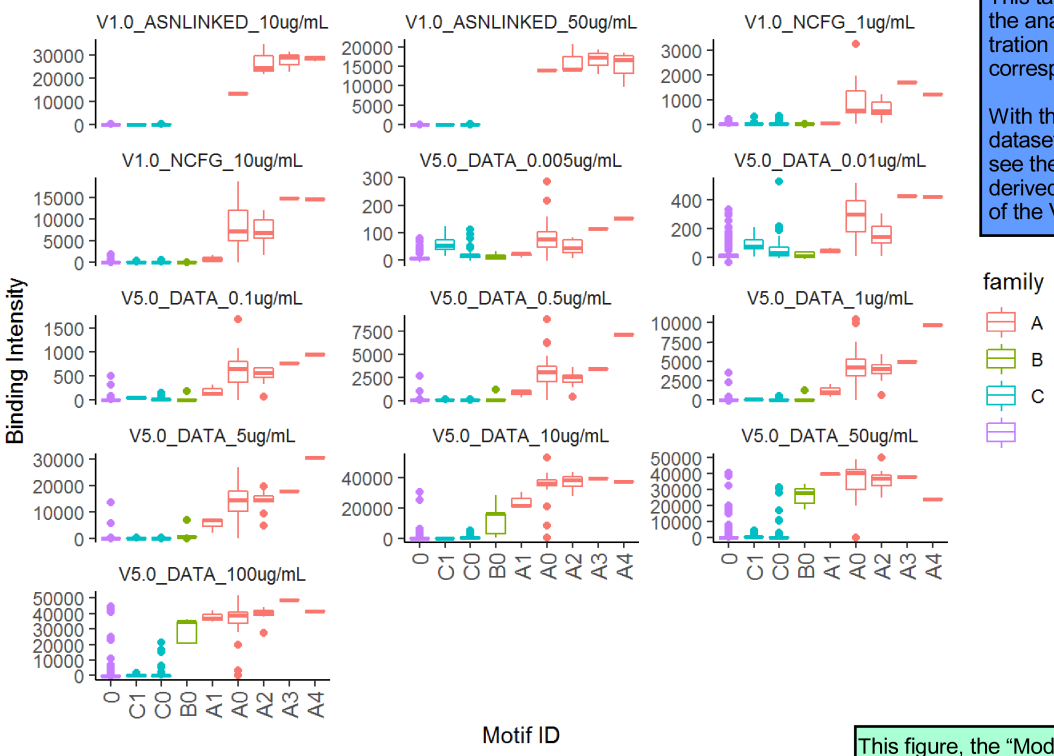
Figure 3. Treemap of glycan binding grouped by motif and family structure. The model structure can be represented as nested boxes where box size is proportional to the number of glycans with t g of glycans with the motif. Only three layers of data splitting are included he

Detailed Model Breakdown

Motif Glycan Examples:

Key:

All Concentration Plots:

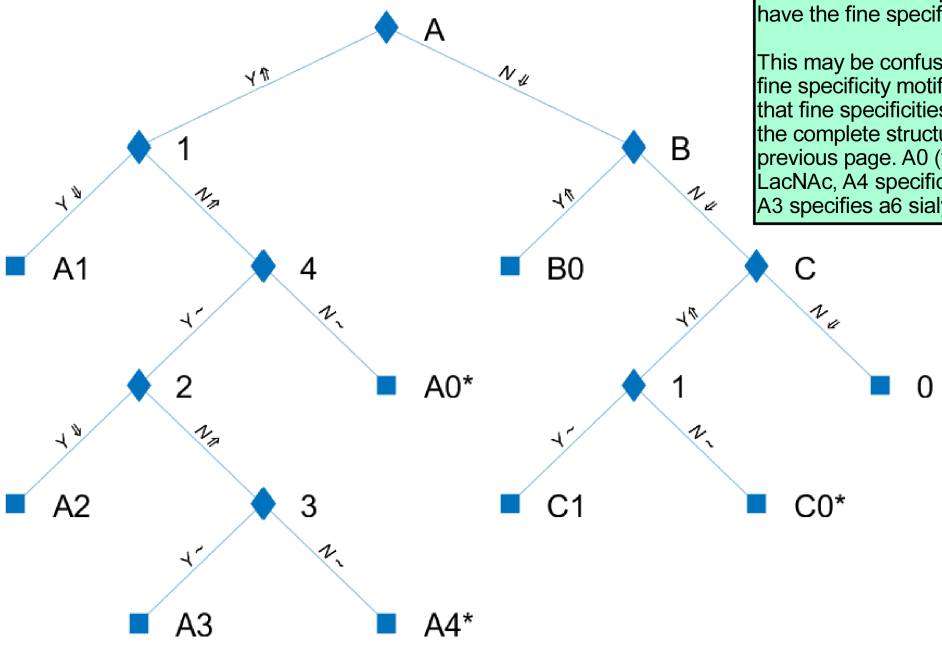


This table briefly shows all the data that went into the analysis with the array source and concentration given at the top of each boxplot. Colors correspond to the primary motif families as before.

With this figure it is possible to identify from which datasets support for a given motif arises. Here we see the support for the weak motifs C1 and C0 is derived from the highest and lowest concentrations of the V5.0_DATA

Figure 4. Boxplots of glycan binding grouped by motif for each dataset in the model. M
(for the selected concentration) and colored by family.

Model Structure:



This figure, the “Model Structure” gives the decision tree made by the analysis. This tree is used to uniquely assign glycans into different motif groups. Assignment into groups (leaves of the tree) is done by following the tree from top to bottom, travelling left if the glycan has the motif and right if not. Further the relative binding of glycans in each is denoted with arrows.

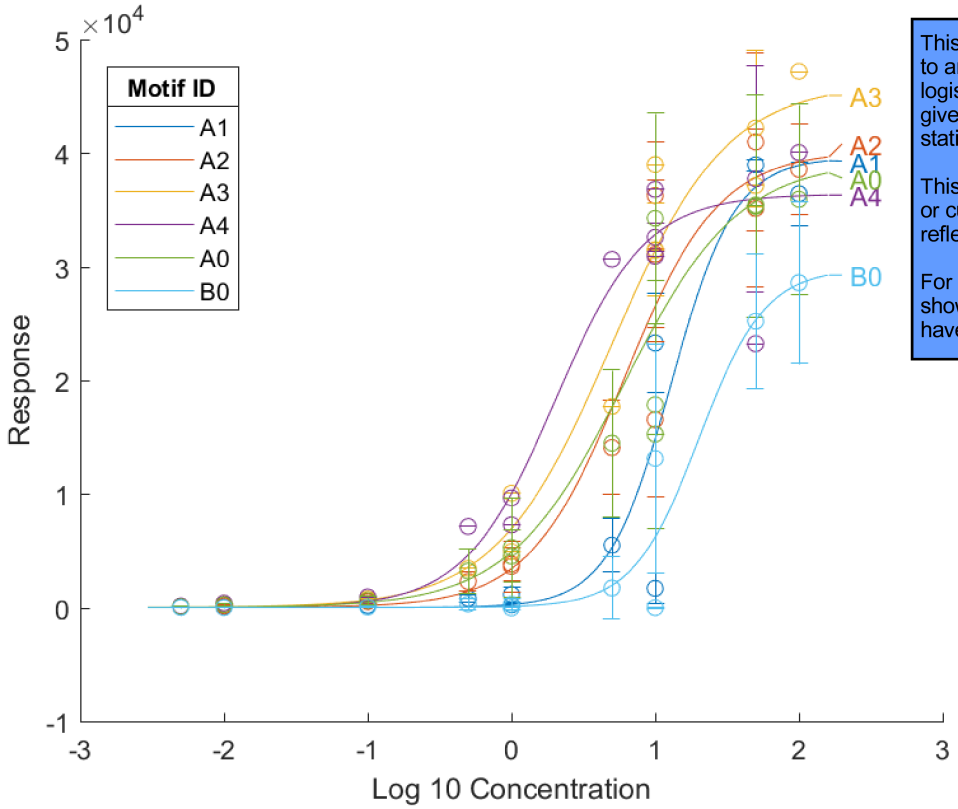
For example: Glycans with the motif “A3" are first evaluated for the primary motif “A”. If they have that they must not have the fine specificity motif A1, they must have fine specificity motif A4 and not have fine specificity motif A2. Then finally they must have the fine specificity motif A3 for which they are named.

This may be confusing at first glance, but if we consider that fine specificity motifs are a subset of primary motifs we can see that fine specificities simply add additional constraints. Look at the complete structures of motifs A0, A4, and A3 on the previous page. A0 (the primary motif) specifies a6 Sialyl LacNAc, A4 specificies a6 sialyl lacNAc on N-glycans, and A3 specifies a6 sialyl lacNAc on 6' tri-antennary N-glycans.

*Motif matches the remaining glycans not matched by earlier motifs in the model.

Figure 5. Tree representation of the regression tree model trained on array data. Data flows through the tree (top-down) and is split by the various motifs. The motif used the split the data at each point has the id “family+split number” except when further split. In the case of futher splits the id of the motif used to split the data is denoted with an asterisk.

Curve Fitting:



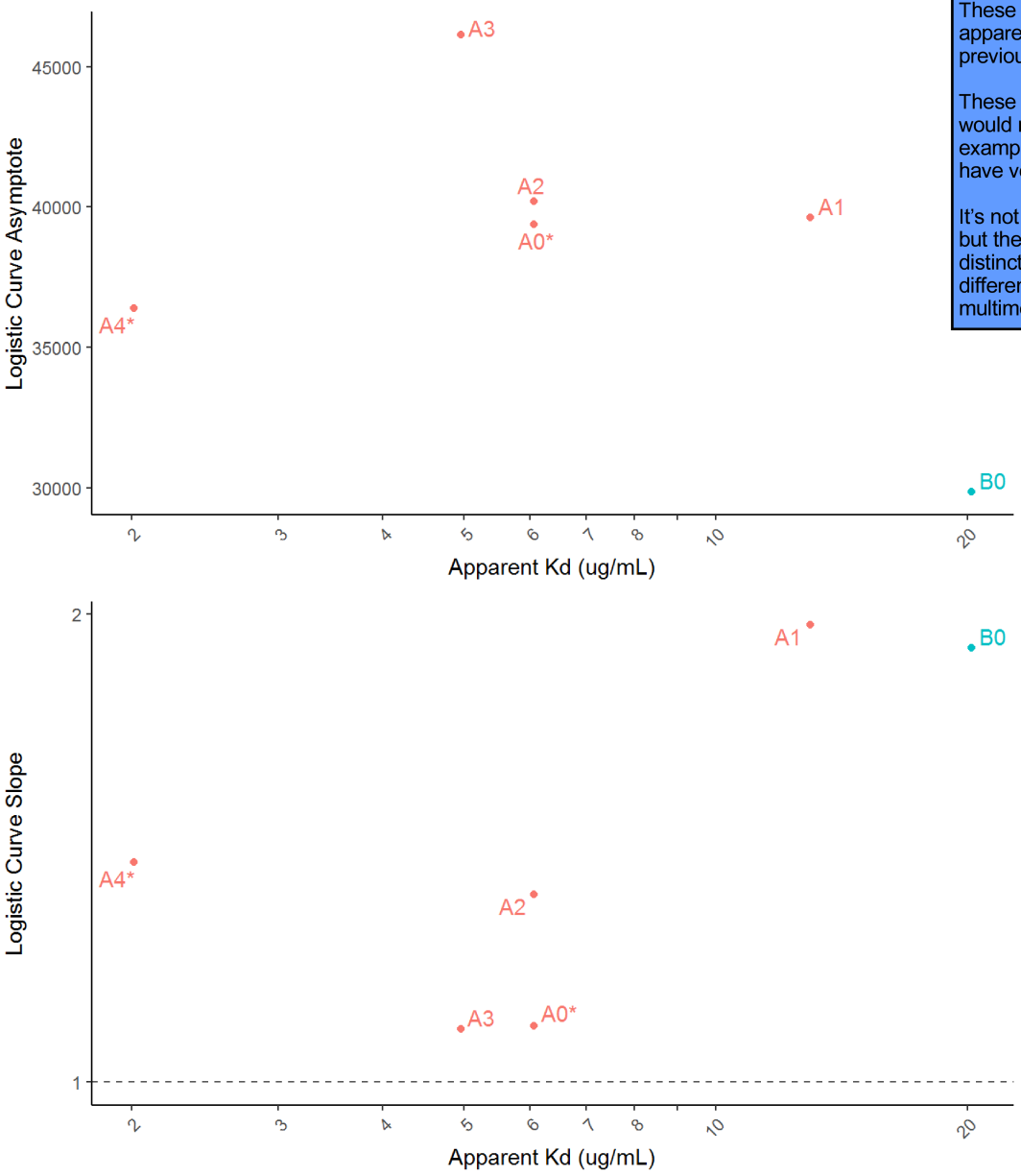
This figure summarizes the curve fitting procedure applied to analyses with more than 2 concentrations. By fitting logistic curves to the average binding of glycans with a given motif, it is possible to calculate “motif-level” binding statistics. Plotted here are those logistic curves.

This plot in particular is best for identifying outlier curves, or curves where apparent Kd might not be accurately reflected in the observed binding.

For example the motif A4 has the best (lowest) Kd as shown here, but at the higher concentrations it does not have the highest binding.

Figure 6. Logistic curves fit to average motif binding. Curves are fit with as many parameters as possible given the data. All curves are based on the logistic curve with a fixed intercept of 0. Nonbinding motifs are excluded from the plot.

Motif Curve Parameter Scatterplots

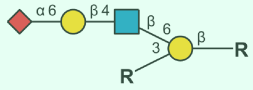
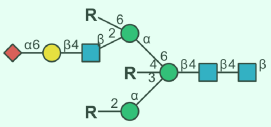
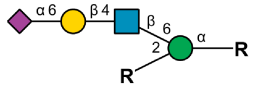


These figures give scatterplots of the measured apparent binding properties for the curves in the previous figure.

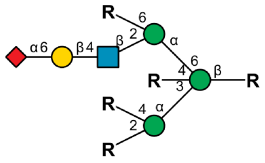
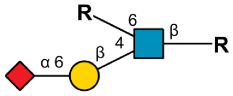
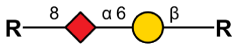
These plots are best used to identify trends which would not be clear from the curves alone. For example we can see that motifs A3, A2, and A0 have very similar binding properties.

It's not possible to prove from microarray data alone but these clusters of binding curves may represent distinct modes of glycan binding, perhaps with different binding sites or binding due to multimerization.

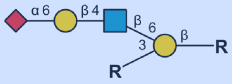
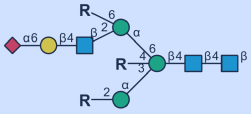
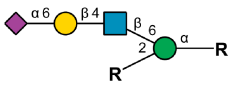
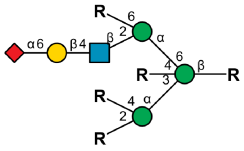
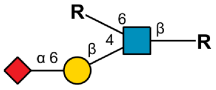
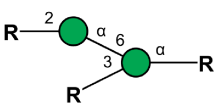
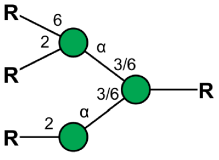
Figure 7. Scatterplots of motif curve parameters. Plotting the asymptote of the curve and the slope of the curve over the log 10 Kd of the motif allows for critical assesment and comparison of motif curves. Note that the apparent Kd of the motif is the concentration at which the curve is half of the asymptote. A standard logistic slope of 1 is marked in a dashed line for reference. Nonbinding motifs are excluded from the plot. The highest affinity motifs which show the best concentration dependent response are in the top left of the top plot. Curves which have very low slopes may also indicate nonspecific binding while higher slopes may indicate higher order binding.

Motif ID	Motif Graphic Structure	Estimated Kd (ug/mL)	Curve Asymptote	Hill Slope
A1		12.971	39622	1.969
A2		6.056	40200	1.320
A3		4.953	46140	1.082

This table simply lists the binding properties of the curves for each of the motifs.

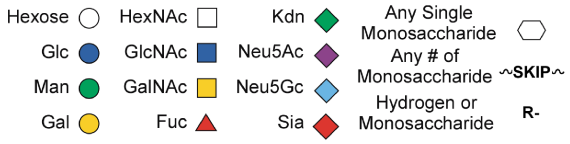
Motif ID	Motif Graphic Structure	Estimated Kd (ug/mL)	Curve Asymptote	Hill Slope
A4*		2.013	36400	1.384
A0*		6.061	39375	1.086
B0		20.230	29845	1.902

Motif Text Structures:

Motif ID	Motif Graphic	Motif Text
A1		<5f6f8f>SiaA2-6<2f3f4f>GalB1-4<3f6f>GlcNAcB1-6<2f4f>GalB
A2		<5f6f8f>SiaA2-6<2f3f4f>GalB1-4<3f6f>GlcNAcB1-2<3f4f>ManA1-6(<3f4f6f>ManA1-3)<2f>ManB1-4<3f6f>GlcNAcB1-4<1f3f6f>GlcNAcB
A3		<6f8f>Neu5AcA2-6<2f3f4f>GalB1-4<3f6f>GlcNAcB1-6<3f4f>ManA
A4*		<5f6f8f>SiaA2-6<2f3
A0*		<5f6f8f>SiaA2-6<2f3f4f>GalB1-4<3f>GlcNAcB
B0		<5f6f>SiaA2-6<2f3f4f>GalB
C1		<3f4f6f>ManA1-6<2f4f>ManA
C0*		<3f4f6f>ManA1-<3or6>(<3f4f>ManA1-<3or6>)<2f4f>Man?
0		Non-Binders

This table simply lists the text form of each of the motifs with the corresponding structures. The details of the motif language is given in the 2017 ACS MotifFinder paper: doi.org/10.1021/acs.analchem.7b04293

Key:


Hexose
HexNAc
Kdn
Any Single Monosaccharide
Glc
GlcNAc
Neu5Ac
Any # of Monosaccharide
~SKIP~
Man
GalNAc
Neu5Gc
Hydrogen or Monosaccharide
R-
Gal
Fuc
Sia