

GLYMMER

GLYMMER™

A software platform for glycosylation modeling and data analysis

Table of Contents

GLYMMER™	1
A software platform for glycosylation modeling and data analysis.....	1
GLYMMER™	3
A software platform for glycosylation modeling and data analysis.....	3
Functionality of GLYMMER™ system.....	3
Key GLYMMER™ features.....	4
GLYMMER™ framework	4
Glycan representation scheme.....	4
Reaction rules.....	6
Reaction kinetics.....	8
Model structure	9
GLYMMER™ software	9
Overview.....	9
Generating the model	9
Solving the model	13
Analyzing and summarizing model results	14
Matching model results to experimental data or to glycan profile targets.....	16
Some useful tools	18
User-callable functions and subroutines.....	18
Ribbon item macros	19
GLYMMER™ Example.....	21
Model-predicted glycan substructures.....	21
Comparison of model calculated synthetic mass spectra with measured spectra for low passage and high passage human LNCaP cells for part of the mass range. Annotation is completely automatic.	21
Model-predicted glycan types.....	23

GLYMMER™

A software platform for glycosylation modeling and data analysis

Functionality of GLYMMER™ system

GLYMMER™ (GLYcosylation Metabolic Model of Enzyme Reactions) is a realistic and experimentally validated glycosylation analysis software suite with strong predictive capabilities to analyze, integrate, and interpret complex sets of glycan structure and gene expression data. The system is aimed at identification of disease specific glycans from various types of glycan structural data and to develop strategies for the control of glycan structures on recombinant therapeutic glycoproteins.

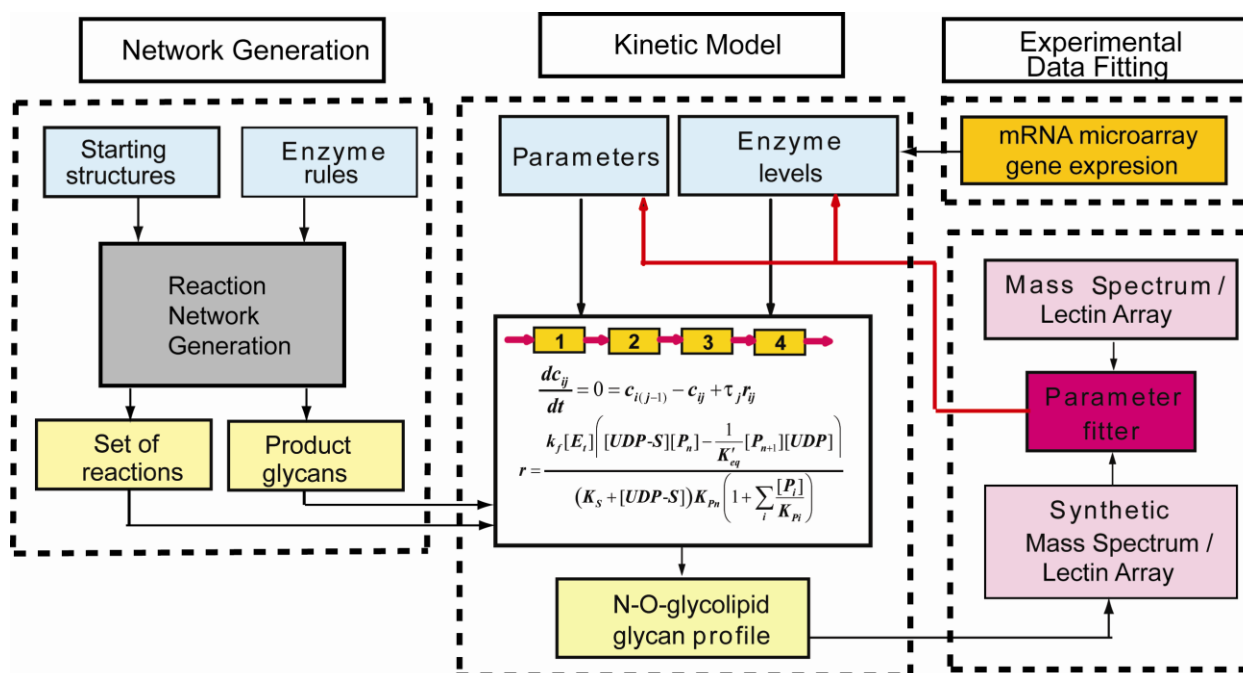


Figure 1. Principle of the GLYMMER™ software suite. The scheme shows the connection between gene expression data obtained from microarrays, cellular processing enzyme activities and glycan structure data. (From Bennun SV, Yarema KJ, Betenbaugh MJ, Krambeck FJ. 2013. Integration of the Transcriptome and Glycome for Identification of Glycan Cell Signatures. PLoS One PLoS Comput Biol 9(1): e1002813. doi:10.1371/journal.pcbi.1002813.)

The system uses a fundamental metabolic reaction model of cellular glycosylation processes to directly relate cellular enzyme activities and properties to the abundances of thousands of glycan structures. Since there are far fewer enzymes than there are glycan structures the model provides a succinct summary and interpretation of a variety of experimental measurements related to glycan structure, such as mass spectrometry and lectin arrays. The model also provides a bridge between enzyme gene expression levels, as measured by microarray experiments, and glycan structure.

For example, by adjusting the model intracellular enzyme activities to bring model-calculated mass spectra into agreement with a number of experimental mass spectra, the model provides an estimate of the the enzyme activities corresponding to each of the spectra as well as automatic estimates of the abundances of thousands of glycan isomers for each of the mass

spectra. This level of detail is not normally obtainable from mass spectra alone. From this information, we can deduce potential markers of normal and diseased cells in which the enzymatic activities, and hence the glycan profiles, have changed. Other structurally sensitive data, such as lectin array measurements, can be simultaneously incorporated in the analysis to strengthen the predictions. If microarray expression data are also available, we can confirm the consistency of the expression and glycan structure data. This information may be useful in determining new diagnostics as well as potential treatment methods. The model can also be used to predict changes in enzyme activities in producer cells required to optimize desirable glycan distributions for improved product quality of valuable recombinant glycoproteins.

Key GLYMMER™ features

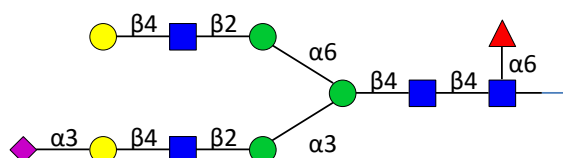
- Relates detailed glycan structure profiles to intracellular enzyme concentrations based on fundamental reaction kinetic principles.
- Incorporates algorithms to automatically annotate mass spectral data with specific glycan structural information based on adjusting cellular enzyme activities in a detailed kinetic model of glycosylation to match the data. Other structural measurements, such as lectin blot arrays, can also be analyzed, either separately or simultaneously in combination with the mass spectra.
- Provides a software tool for relating gene expression data of the glycosylation enzymes to enzyme activities and uses this information to predict resulting glycan structures.
- Serves as a software platform for deriving enzymatic activities and glycan structures for different tissues in normal and disease states in order to identify new biomarkers for disease, assays for disease, and possible targets for disease therapy.
- Includes software to predict enzyme activities in producer cells required to optimize desirable glycan distributions for improved product quality of valuable therapeutics.

GLYMMER™ framework

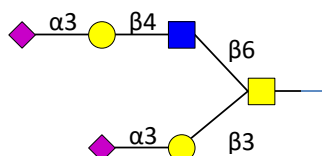
Glycan representation scheme

In the GLYMMER™ model framework, glycan structures are expressed using a condensed version of IUPAC linear formulas (McNaught 1996) with some minor modifications. The first modification is to order the branches at a branch point based just on the branch locants (the carbon atom numbers of the attachment points of each branch to the branching sugar group) without regard to the lengths of the branches (branch length is used by the IUPAC scheme). In addition the sugar abbreviations have been replaced with the shorter abbreviations of the Linear Code (Banin, et al. 2002), but we have not used the branch ordering scheme of the Linear Code (which orders branches based on the type of sugar attached). The sugar abbreviations consist of one or two uppercase letters for each sugar, such as M for D-mannose or GN for N-acetylglucosamine. Linkages between the sugar groups are designated as lower case a or b, for alpha or beta, followed by the carbon number of the attachment site. The carbon number linked in the attaching sugar is not specified since it is generally the same for each type of sugar in naturally occurring glycans. Thus GNb2M indicates a $\beta(1,2)$ linkage from N-acetylglucosamine to D-mannose. The root of the glycan is indicated by a semicolon, optionally followed by an abbreviation for the attachment site of the glycan to a polypeptide chain. This scheme provides linear formulas that are general, are easily readable by humans, are unique for each glycan structure, and allows the models to apply to N-glycans, O-glycans and glycolipids. Graphical representations of the structures within GLYMMER™ utilize the nomenclature from "[Essentials of Glycobiology](#)". Some examples:

NNa3Ab4GNb2Ma3(Ab4GNb2Ma6)Mb4GNb4(Fa6)GN;Asn



NNa3Ab3(NNa3Ab4GNb6)AN;Thr



A complete list of the sugar codes currently incorporated in GLYMMER™ is given in Table 1. More codes can easily be added by the user.

Table 1. Monosaccharide codes and symbols used in GLYMMER™.

Name	Kcodes	Code	Symbol
D-Glucose	Glc	G	
D-Galactose	Gal	A	
N-Acetylglucosamine	GlcNAc	GN	
N-Acetylgalactosamine	GalNAc	AN	
D-Mannose	Man	M	
N-Acetylneuraminic acid	Neu5Ac	NN	
N-Glycoloylneuraminic acid	Neu5Gc	NG	
Neuraminic acid	Neu	N	
Keto-deoxy-nonulonic acid	Kdn	K	
Keto-deoxy-octulonic acid	Kdo	W	
D-Galacturonic acid	GalA	L	
D-Ioduronic acid	Ido	I	
L-Rhamnose	Rha	H	
L-Fucose	LFuc	F	
D-Xylose	Xyl	X	
D-Ribose	Rib	B	
L-Arabinose	Ara	R	
D-Glucuronic acid	GlcA	U	
D-Allose	All	O	
D-Apiose	Api	P	
D-Fructose	Fru	E	
Sulfate		S	
Phosphate		PH	

Reaction rules

Using the condensed formulas, enzyme specificities are denoted using a string substitution methodology. The basic idea is to express the function of an enzyme by one or more reaction rules. Each reaction rule includes a pair of character strings. The first, or substrate substring, is the formula representation of the substructure of a glycan that the enzyme can act upon. The second, or product substring, is the formula representation of the substructure that the substrate substructure is transformed into by the enzyme action. Additional constraints on the substrate that are required for the enzyme action are expressed as a set of logical tests for the presence or absence of other substrings in the complete substrate formula. As an example consider the two reaction rules for the enzyme mannosidase I shown in Table 2.

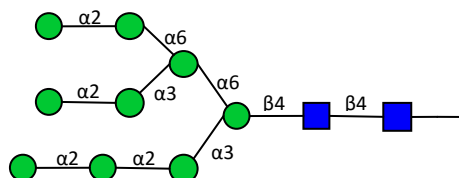
Table 2. Reaction rules for mannosidase I.

Based on (Herscovics 2001). See Table 3 for explanation of rule symbols.

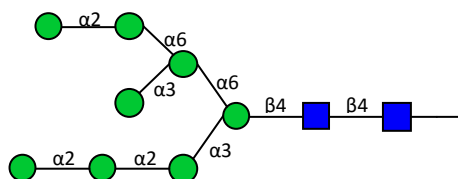
Enzyme	Substrate	Product	Constraint	Cosubstrate	Coproduct
ManI	(Ma2Ma	(Ma	$\sim^*2\text{Ma}3(\dots\text{Ma}6)\text{Ma}6$ & $\sim\text{Ga}3$	water	mannose
ManI	(Ma3(Ma2Ma3(Ma6)Ma6)	(Ma3(Ma3(Ma6)Ma6)	$\sim\text{Ga}3$	water	mannose

The rules will be applied to two starting structures:

(a) Ma2Ma2Ma3(Ma2Ma3(Ma2Ma6)Ma6)Mb4GNb4GN;



(b) Ma2Ma2Ma3(Ma3(Ma2Ma6)Ma6)Mb4GNb4GN;



In applying reaction rules to a candidate structure the algorithm first encloses the formula of the structure in parentheses. This ensures that the terminal residue of any branch begins with a left parenthesis. The first substrate rule of Table 2, “(Ma2Ma”, specifies that a potential substrate glycan must contain a terminal mannose residue connected via an $\alpha(1,2)$ linkage to another α -mannose residue for the rule to operate. Referring to the two structures above, it can be seen that structure (a) contains three such substrings, and structure (b) contains two.

The corresponding product rule, “(Ma”, specifies that the action of the rule converts the “(Ma2Ma” substructure to “(Ma”, thus removing the terminal mannose residue. The constraint

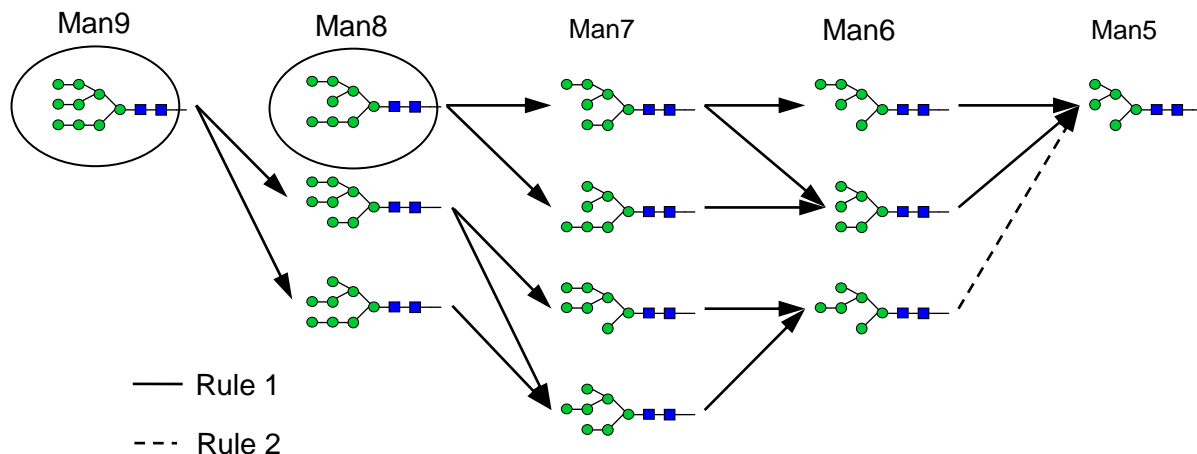
substring in Table 2 specifies further conditions that the potential substrate must satisfy for the rule to operate. The rule uses some abbreviations that have been useful in expressing reaction rules. The abbreviations used are defined in Table 3. The first part of the constraint substring of the first rule in Table 3, “~*2Ma3(...Ma6)Ma6” states that there is one particular mannose that will not react by this rule, the terminal mannose on the middle branch of structure (a). The “~” indicates the logical “not” operator and the “&” the logical “and”. The “*” indicates the reaction site on the substrate and the “...” any valid ligand. The second part of the constraint substring of the first rule states that the rule will not operate if there are any glucose residues in the structure. The second rule in Table 2 indicates that once all the other $\alpha(1,2)$ mannose residues have been removed the nonreacting $\alpha(1,2)$ residue will finally react.

Table 3. Codes for rules.

Symbol	Meaning	String expression
...	Ligand	Any string (possibly empty) with all parentheses matched.
_	Continuation	Any string (possibly empty) where every "(" is matched with a following ")"
	Possible branch point	Empty string or ")" or "(...)" or ")(...)"
*	Reaction site	Position of first difference between product and substrate strings
Gnbis	Bisecting Gn	Ma3(GNb4)(...Ma6)Mb4
#	Number of	
~	Logical not	
&	Logical and	
or	Logical or	

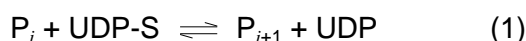
Applying the rules for each enzyme in turn to the starting structures (a) and (b) generates a set of reactions and new product glycan structures shown in Figure 1. For example, applying the first rule of Table 2 to the 9-mannose starting structure (a) gives the two structures in the second column of Figure 1 that are shown as products from the single structure in the first column of the figure. This process is then repeated using the new product structures as potential substrates to generate more reactions and structures. For example the third column of Figure 1 results from applying the first mannosidase I rule of Table 2 to the three structures of the second column of the figure. The second rule of Table 2 is applied to the last structure of the fourth column and results in the structure of the fifth column. The process of generating structures is repeated until either the process terminates, as it does in Figure 1 where no new structures can be generated, or until some other criterion is met, such as a maximum mass cutoff.

Figure 1. Reaction network generated by mannosidase I rules of Table 2. This is the same network as Herscovics (2001).



Reaction kinetics

Consider the glycosylation of a glycan P_j with a monosaccharide S catalyzed by an enzyme E . Assuming that the donor cosubstrate is $UDP-S$, the overall reaction is shown in equation (1):



Assuming that the product P_{i+1} competes for the same enzyme site as the substrate P_i , that the donor cosubstrate $UDP-S$ occupies a second site on the enzyme, and that the reaction is reversible, the Michaelis-Menten kinetic equation takes the form shown in equation (2).

$$r = \frac{k_f [E_t] \left([UDP-S][P_i] - \frac{1}{K'_{eq}} [P_{i+1}][UDP] \right)}{K_{mi} (K_{md} + [UDP-S]) \left(1 + \sum_j \frac{[P_j]}{K_{mj}} \right)}$$

$$K'_{eq} = \frac{k_f K_{m,i+1}}{k_r K_{md} K_{mi}} = \frac{[P_{i+1}]^* [UDP]^*}{[P_i]^* [UDP-S]^*} \quad (2)$$

Here k_f and k_r are the forward and reverse rate coefficients, K_{mi} and K_{md} are the dissociation constants for the substrate and donor cosubstrate, and K'_{eq} is the apparent equilibrium constant for the overall reaction. The symbols $[]^*$ in equation (2) denote equilibrium concentrations. The subscript “j” in the summation in the denominator is taken over all the substrates that compete for the same enzyme. A derivation for equation (2) is given in the KB2005 model (Krambeck and Betenbaugh 2005).

The values of the kinetic parameters k_f , K_m and K_{md} for a given enzyme can vary significantly for different substrates. This was accommodated by selecting base values for these parameters for each reaction rule and adding a set of structure-dependent adjustment rules. Examples of reaction parameters and adjustment rules for the mannosidase I enzyme are given in Table 4.

Table 4. Sample reaction rate parameters and adjustment rules for mannosidase I.

Parameters	$k_f(\text{min}^{-1})$	$K_m(\mu\text{M})$	$K_{md}(\mu\text{M})$
------------	------------------------	--------------------	-----------------------

Rule 1	1924	100	0
Rule 2	100	5000	0

Adjustments

Test	Multipliers		
#M=9	1	6.55	1
#M=8	1	4.04	1
#M=7	1	1.69	1
#M=6	1	1.00	1
Ma2Ma3(...Ma6)Ma6	1	8.62	1

Model structure

The glycosylation chemical reactions take place in the Golgi compartments of a cell. These are modeled as a series of four well-mixed reactors at steady-state. These represent the cis-, medial- and trans- Golgi cisternae and the trans-Golgi network. More compartments can easily be added if desired, for example to represent the endoplasmic reticulum. The compartments are characterized by the residence time of the glycan substrates in each compartment and the enzyme and sugar-nucleotide concentrations, that can differ for each compartment. The inputs to the first compartment are the concentrations of the glycans entering from the previous organelle. The concentrations of the excreted glycans are assumed to be those of the last compartment in the series.

GLYMMER™ software

Overview

The GLYMMER™ software is accessed through Microsoft Excel and consists of an add-in workbook "Glymmex.xlam", together with various dll libraries. The main GLYMMER™ features are accessed from ribbon buttons. The system includes a variety of functions and subroutines that can also be included in user worksheets or user-written VBA code. The ribbon buttons are contained in a number of groups under the GLYMMER tab. The macros are set up to access data in the active worksheet. The inputs are entered in the active worksheet in the form of various tables. The program searches the active worksheet for various keywords that locate the input tables. The input tables start with a row of column headings starting just below the keyword for the table. The first blank cell encountered in the header row of the table defines its width and the first blank cell in the first column of the table defines its height.

Generating the model

All the inputs required to generate the reaction network and the Golgi compartment model are specified on a worksheet. With this worksheet active, the reaction network is generated by selecting the "GLYMMER" tab on the ribbon and clicking the "Network" button in the "Build" group. The Golgi compartment model is generated by selecting the "Model" button of the same group. Inputs required on the active worksheet are listed below:

Rule table

The keyword for the rule table is "Rule table" (case-sensitive). The rule table lists the reaction rules as rows with the following column headings (in any order):

Index	The index of a rule is an arbitrary integer to identify the rule. These need not be in order.
Enzyme	An arbitrary identifier for the enzyme that catalyzes the reaction.
Substrate	The substring of a glycan formula required for the enzyme to act on the glycan.
Product	The modified substring that describes how the substrate substring is changed.
Constraint	Constraints on the substrate structure required for the reaction to occur.
Cosubstrate	A metabolite, usually a nucleotide-sugar donor, that reacts with the substrate.
Coproduct	A metabolite that is produced by the reaction.
kf	The forward rate coefficient for the Michaelis-Menten reaction kinetics.
Km	The Michaelis-Menten dissociation constant for the enzyme-substrate complex.
Kmd	The Michaelis-Menten dissociation constant for the enzyme-cosubstrate complex.

The kinetic parameters are required for network pruning if it is specified, and for generating the Golgi compartment model. Other columns can be included in the rule table and will be ignored by the program.

Starting structures

A table of starting structures is located by the keyword "Start". The table contains the following column headings:

Component	Formula for a starting glycan.
Conc, μM	The starting concentration of the glycan in micromoles per liter. This is needed for network pruning or for generating the Golgi compartment model.

Adjustment factors

The keyword for the adjustment factor table is "Adjustment factors". The columns of the table are as follows:

Index	Refers to the rule index from the rules table that the adjustment applies to.
Rule	The test to see if the adjustment rule applies to a given substrate.
kf	If the adjustment rule applies the forward rate coefficient is multiplied by this factor.
Km	If the adjustment rule applies the Michaelis-Menten substrate dissociation constant is multiplied by this factor.
Kmd	If the adjustment rule applies the Michaelis-Menten cosubstrate dissociation constant is multiplied by this factor.

Compartment residence times

The keyword for this table is "Compartment residence time, min". The headings are "Comp 1", "Comp 2", etc. and the next row lists the residence times of the compartments. The number of compartments in the model can be changed by just including more or fewer columns in this table.

Enzyme distribution

The keyword for this table is "Enzyme distribution". The columns are:

Enzyme	The identifier for an enzyme.
Conc, μM	Total enzyme concentration in the compartments.
Comp 1, ...	Fraction of enzyme concentration in compartment 1, ...

Fixed concentrations

The keyword for this table is "Fixed concentrations". The columns are:

Component	Identifier of component.
Conc, μM	Concentration of component in Golgi compartments in micromoles per liter. As discussed later these concentrations can be specified separately for each compartment if desired.

Other parameters

A list of other parameters is located by the keyword "Parameters". This consists of a two-column table without headers. The first column lists the parameters and the second column contains the parameter values:

Model name	(optional) This is a short name to identify the model. If supplied it will be appended to the name "Components" for the components sheet to the name "Reactions" for the reactions sheet and to the name "Calc" for the model calculation sheet.
Drop tolerance	A nonzero value for this parameter selects the network pruning option. This option uses an approximate method to identify lower abundance glycans and eliminate them from the network. This option can significantly reduce the size of the model but can sometimes eliminate important glycans. The higher this tolerance is set the more glycans will be dropped. Setting the value to zero or leaving the parameter out will retain all possible glycan structures for the given set of rules.
Maximum passes	Sets the maximum number of generations to test each newly generated set of glycan structures for more possible reactions.
Cutoff mass	Sets the maximum glycan mass to be included in the model. This is calculated as the monoisotopic mass of permethylated ions.
Glycan type	"N" for N-glycans and "O" for O-glycans. Needed to calculate mass spectrum ion masses (MSMass). Defaults to "N".

Network output

The results of network generation are contained in two worksheets "Components" and "Reactions". If worksheets of these names are already present in the active workbook the contents of the worksheets will be cleared and the new output will be written to them. If not the sheets will be newly created.

The "Components" worksheet contains a "Components" table listing all the chemical components in the model, including all the glycan structures and the cosubstrate and coproduct metabolites. The table also lists various properties of the glycans. The worksheet also shows the number of total components "ncomp" and the number of components with assumed fixed

composition "ncomp0". The worksheet also includes a list of the enzymes used. The columns of the "Components" table are as follows:

Index	The index number of the components from 1 to ncomp
Component	The identifier (if a metabolite) or the structural formula (if a glycan) of each component in the model.
Codes	A succinct code for each glycan giving the number of antennae, the number of repeat lactosamine groups and the number of sialic acid groups.
Saccs	The monosaccharide composition of each glycan.
Formula	The chemical formula of the glycan.
MSFormula	The chemical formula of the glycan when permethylated and sodiated for mass spectroscopy.
MSMwt	The molecular weight of the permethylated and sodiated glycan.
MSMass	The monoisotopic mass of the permethylated and sodiated glycan.
FormID	An identifier for each unique chemical formula.

The "Reactions" worksheet contains a reaction table showing all the generated chemical reactions. It also includes a list of the kinetic parameters and the adjustment rules in the model. The kinetic parameters and adjustment table parameters are changeable by the user without regenerating the network.

The columns of the reactions table are as follows:

Rule	Indicates which reaction rule generated the reaction.
Adjust	Indicates which adjustment rules apply to the reaction.
Enzyme	The enzyme for the reaction.
Substrate	The substrate for the reaction.
Cosubstrate	The cosubstrate for the reaction
Product	The product of the reaction.
Coproduct	The coproduct of the reaction.
kf	The adjusted forward reaction rate coefficient.
Km	The adjusted Michaelis-Menten dissociation constant for the substrate.
Kd	The adjusted Michaelis-Menten dissociation constant for the cosubstrate.

Model generation output

The results of the model generation are given on a worksheet labeled "Calc". This sheet includes the inputs and outputs for solving individual cases of the model. A variety of input and output tables on the "Calc" sheet are described below:

Compartment residence times

The keyword for this table is "Compartment residence time, min". The headings are "Comp 1", "Comp 2", etc. and the next row lists the residence times of the compartments. The compartment residence times may be changed in this table but not the number of compartments.

Enzyme concentrations

The keyword for this table is " Enzyme concentrations, μM ". The entries in this table are calculated by worksheet calculations from other entries on the sheet and should not be changed. To adjust these values change the entries under the heading "Total, μM " and "Enzyme fractions" instead.

Enzyme fractions

The keyword for this table is "Enzyme fractions". The "Total, μM " column is multiplied by the fractions shown for each compartment to get the enzyme concentrations in the previous table. This has turned out to be a convenient way to adjust enzyme activities without changing the distribution of the enzymes between compartments. The individual concentrations in the previous table are the ones actually used in the calculation, however.

Compartment concentrations

The keyword for this table is "Concentrations, μM ". It shows the concentrations of all components in each compartment. The last compartment, labeled "Comp Out", is the set of excreted concentrations.

Equation structure

The items of interest in this table are the columns "Variables", which is a list of all the model variables to be solved for, and "Residuals", which are the residual errors in the equations to be solved.

Other parameters

The "Parameters" table on the "Calc" sheet includes the following items:

Model name	(optional) Identifier for model.
Components table	Name of "Components" worksheet (may be changed)
Reactions table	Name of "Reactions" worksheet (may be changed)
Number of compartments	Cannot be changed on this sheet.
Total oligosaccharide concentration	Can be changed.
Sparsity table	Name of sparsity worksheet. This can be changed but Sparsity status should then be set to False.
Sparsity status	Sparsity table has been generated for this model, True or False.
Maximum iterations	Number of iterations allowed to solve model equations.

Solving the model

SetupReact

The "SetupReact" macro in the "Calculate" group reads in all the parameter values needed to solve the model, starting with the "Calc" sheet active. The name of the active sheet is irrelevant but it must include all the information on the created "Calc" sheet. Different sheet names can also be used for the "Reactions" and "Components" sheets by changing the specified "Components table" and "Reactions table" on the active sheet. This macro will only need to be executed once after the workbook is loaded for multiple runs of the model solution.

Solve

The "Solve" macro in the "Calculate" group will adjust the glycan concentrations to solve all the model equations. Multiple cases can be run for various values of parameters included on the "Calc" sheet, such as enzyme concentrations, compartment residence times, input concentrations for the first compartment, etc. If kinetic parameters or adjustment rules are changed on the "Reactions" sheet for a particular case the "SetupReact" macro must be rerun before running the "Solve" macro.

Analyzing and summarizing model results

The raw results of a GLYMMER™ model calculation consists of the abundances of thousands of individual glycan structures. We obviously need simple methods of boiling down this very detailed information into the glycan properties we are actually interested in. The GLYMMER™ platform includes a very flexible tool for summarizing a model result into a set of abundances of various glycan types or substructures. The method relies on expressing the glycan features we are interested in in terms of the same thpe of substructure codes we have used to express enzyme specificity. Table 5 gives some sample codes for summarizing some common features. The "Count" keyword used in these codes handles substructures that can occur more than once in a single glycan.

Table 5. Sample codes for summarizing N-glycan model results.

Description	Code
High mannose	#GN=2
Hybrid	#GN>2 & #M>3
Monoantennary	GNb2Ma3 (Ma6)Mb4
Biantennary	GNb2Ma3 (...GNb2Ma6)Mb4
Triantennary1	GNb2(...GNb4)Ma3 (...GNb2Ma6)Mb4
Triantennary2	GNb2Ma3 (...GNb2(...GNb6)Ma6)Mb4
Tetraantennary	GNb2(...GNb4)Ma3 (...GNb2(...GNb6)Ma6)Mb4
Bisected	Ma3(GNb4)(...Ma6)Mb4
Nonsialylated	#NN=0
Monosialylated	#NN=1
Bisialylated	#NN=2
Trisialylated	#NN=3
Tetrasialylated	#NN=4
Lactosamine repeats	Count Ab4 GNb3
Lactosamine roots	Count GNb3
Terminal mannose	Count (M

Terminal GlcNAc	Count (GN
Terminal β -galactose	Count (Ab
Terminal α -galactose	Count (Aa
Terminal sialic acid	Count (NN
Lactosamine Branch	Count GNb3(...GNb6)Ab
Core Fucose	Count GNb4(Fa6)GN
Branch fucose	Count Fa3
Lex	Count (Fa3(Ab4)GN
SLex	Count (Fa3(NNa3Ab4)GN
VIM-2	Count (Fa3(NNa3Ab4GNb3Ab4)GN

To utilize this method include a table such as Table 5 in the active sheet under the keyword "Summary". Also include the keyword "Calculation sheet" in a cell and the name of the "Calc" sheet in the adjacent cell to the right. The macro "Summary" in the group "Results" will then apply the codes to all the structures in the model and calculate the abundances for model results from the "Calc" sheet. The results will be written to the active sheet in the column just to the right of the "Code" column.

The "Summary" macro can also do the calculation for multiple cases at once using a "Results" sheet for input instead of a "Calc" sheet. Use the keyword "Results sheet" rather than "Calculation sheet" and specify a sheet containing the result of the "Results" macro in the "Results" group.

The "Results" macro gathers the results from a "Saved" sheet. "Saved" sheets are generated automatically when the "Fit" macro (described below) is executed. The output of the macro is inserted into the "Results" sheet and includes details about the mass spectra of the glycans. The "Results" sheet can also be used for input to the "Summary" macro to do multiple cases at once.

Inputs to the "Results" macro are specified under the keyword "Parameters". The inputs are "compsheet", the name of the "Components" sheet, "savedsheet", the name of the "Saved" sheet, and "calc", the name of the "Calc" sheet. The keyword "Components" must also be included in the first column of the sheet to indicate where the output of the macro will be inserted on the "Results" sheet.

Generation of synthetic mass spectra

Synthetic mass spectra, which consist of the intensities of all the individual isotopic masses for all the glycan components included in the model, are calculated using the "Setup" and "MS Syn" macros in the "Mass Spectra" group. The active "MassSpecCalc" sheet for these macros contains the following labels with entries in the cells to the immediate right of the labels:

Mass resolution	When individual isotopic peak masses differ by less than this amount they are merged into a single peak with the weighted average mass. A value of 0.1 is usually used.
-----------------	---

Mass probability tolerance	Probabilities of isotopic masses are based on the natural abundances of each element's isotopes. When the probability of a particular isotopic peak falls below this value the peak is dropped from the spectrum. A value of 1E-6 is usually used.
Components sheet	Name of the "Components" sheet that includes all the glycan structures in the model.
Calculation sheet	Name of the "Calc" sheet. The output glycan concentrations in this sheet are used to calculate the intensities of the mass peaks.
Maximum mass	A cutoff value for the calculated spectrum.

The "Setup" macro copies the FormID and mass spec chemical formulas from the components sheet and links the output glycan concentrations, normalized to 100%, from the specified "Calc" sheet to the "Amounts" column. The "MS Syn" macro calculates the synthetic mass spectrum and inserts it into the "Mass" and "Intensity" columns. Also shown under the "Projected" keyword are the total of the calculated isotopic peaks for each monoisotopic mass under the headings "MMassID", "MMass" and "MAmtsMod".

Matching model results to experimental data or to glycan profile targets

The GLYMMER™ platform includes a variety of tools that allow the user to adjust model parameters, such as enzyme or nucleotide-sugar concentrations, to bring the model into agreement with experimental measurements or to meet desired glycan property targets. The same tools can be used to easily set up series of cases to determine the impact of modifications of the underlying kinetics on desired outcome criteria. The only requirement for the type of measurement or property that can be used as a target is that the target values can be derived from the calculated concentrations of glycans from the model. Any combination of target types can be used simultaneously to specify the desired matching. The "Summary" macro and the "MS Syn" macro described above can be used to calculate desired targets.

The inputs for the matching or fitting process are included on a "Fit Data" sheet, which can be named as desired. Two macros, "One Pass" and "Fit" are included in the "Fit Data" group of the GLYMMER™ ribbon tab. The "Fit Data" sheet must be active for these macros to operate. The "Fit Data" sheet contains a variety of tables:

Settings	A list of settings for the fitting program. Currently there is just one setting called "calcsheet". This is the name of the "Calc" worksheet for the model to be matched to the data or solved for a series of cases.
Post-solution script	A list of macros to be run each time after the model is solved. This feature can be used to derive properties of the model-calculated glycan profile for matching to experimental data. The script table contains three columns. The "Worksheet" column specifies the name of a worksheet to be activated before calling the macro; the "Macro" column gives the name of the macro to be executed; and the "Source" column indicates whether the macro is contained in the GLYMMER add-in ("Addin") or the active workbook ("Local").
Inputs	Under the heading "Inputs" there is a single column of formulas pointing to cells in the "Calc" sheet or the "Reactions" sheet. These formulas must be of the form "=sheet!cell". These entries specify which input variables are to be set for each solution of the model. Labels may be placed to the left of the "Inputs" column to guide the user, but these labels are ignored by the program.

Cases	Under the heading "Cases" there are one or more columns of numbers that specify the values the input variables are to be set to for one or more cases. The number of cases are set by the number of columns provided under the "Cases" heading.
Outputs	Under the "Outputs" heading there is another list of formulas that point to the output variables that will be matched to measured or specified targets. The formulas are arbitrary, but usually just point to cells on other sheets. Labels may be placed to the left of the "Outputs" column to guide the user, but these labels are ignored by the program.
Calculated	The columns under the "Calculated" heading are automatically filled in with the values from the "Outputs" column each for each case.
Data	For each case there is a column under the "Data" heading. These columns specify the target values for each of the variables in the "Output" column. Blank entries in the "Data" columns are excluded from the matching.
Errors	Under the heading "Errors" there is a column showing the differences between the nonblank cells in the "Data" table and the corresponding calculated values in the "Calculated" table. In addition the "Errors" column can include additional criteria to be added into the objective function for minimizing the errors. The differences for all the nonempty data cells and other criteria are collected in one column.
RMS	Under the heading "RMS" there is a single number equal to the root-mean-square of the "Errors" column.
Fitting parameters	<p>Under the heading "Fitting parameters" there is a table specifying which variables are to be adjusted to bring the model into agreement with the targets. The "Parameters" column in the table contains formulas which point to the adjusted variables using the "sheet!cell" format. These cells can be on the "Calc" or "Reactions" sheets, but it is often more convenient to point to cells in the "Cases" table. This allows a single variable to take on different values for different cases rather than the same value for all cases.</p> <p>The "Fitting parameters" table also includes columns labeled "Lower" and "Upper" to set bounds on the adjusted parameters. Reference values for the parameters can be set in the "Reference" column. The deviation of the adjusted parameter from its reference value is multiplied by the number in the "Weight" column and appended to the "Errors" list. This avoids variables that have very little effect on the target variables from taking on huge values to improve the fit. Enzyme concentrations are often assigned a reference value of zero to bias the adjustments away from enzymes that aren't needed to match measured targets.</p>

One Pass

The "One Pass" macro in the "Fit Data" group solves all of the cases on the "Fit Data" sheet and saves the solutions for th cases on a "Saved" sheet. Previous solutions on the "Saved" sheet are replaced. If the "Fit Data" sheet contains just one case the solution is instead added to a "SavedAll" sheet as a column with a heading of the "Fit Data" sheet name. A previous solution on the "SavedAll" sheet with the same name is replaced. These saved solutions can subsequently be used with the "Results" macro to produce tables of results.

Fit

The "Fit" macro adjusts the selected input variables selected in the "Fitting parameters" table to match the model results to the selected targets in the "Data" table. Solutions are saved in the "Saved" or "SavedAll" sheets as the fitting progresses.

Some useful tools

The "Tools" group on the "GLYMMER" ribbon contains a variety of useful macros:

Draw structure

The "Draw structure" macro draws a glycan structure from a selected condensed formula or group of condensed formulas. The selection must be a single cell or a group of vertically adjacent cells on any worksheet. The symbols used for the monosaccharides are contained in an add-in worksheet called "Codes". This worksheet can be overridden by including a modified worksheet with the same name in the active workbook. (See "Get sheet" just below.) Adding new monosaccharide codes to the "Codes" sheet allows them to be included in models.

Run script

The "Run script" macro will execute a selected script. A script consists of a three-column table with the columns labeled "Worksheet", "Macro" and "Source". Each row of the table is an operation where a worksheet in the user's workbook named in the "Worksheet" column is activated and a macro named in the "Macro" column is executed. The "Source" specifies if the macro is contained in the GLYMMER™ add-in ("Addin") or the users workbook ("Local"). The script table is first selected and then the button is clicked.

Get sheet

There are a variety of worksheets included in the add-in that are normally invisible. These contain a variety of information the user may wish to see or possibly to change. The "Get sheet" macro displays a drop-down box with a list of worksheets in the add-in. When the sheet selected in the drop-down list is changed the macro copies it to the active workbook after the current active sheet. The worksheets in the add-in are not editable by the user, but in some cases the copied worksheet overrides the add-in worksheet of the same name.

User-callable functions and subroutines

The GLYMMER add-in includes a number of functions and subroutines that can be called directly by the user within worksheet cell formulas or within user-written VBA macros. These are described below.

AsialoFormL

result = AsialoformL(struct)

struct: String. Linear formula of a glycan.

result: String. Code for the asialylated glycan using a nomenclature proposed by Schacter(1989).

CodeFormL

result = CodeFormL(struct)

struct: String. Linear formula of a glycan.

result: String. Summary code for the glycan "struct" in the form "AiLjSk", where i is the number of antennae, j is the number of lactosamine groups, and k is the number of sialic acid groups.

Sacform

result = Sacform(struct)

struct: String. Linear formula of a glycan.

result: String. The number of each monosaccharide symbol that appears in the structure "struct", using the format "sym1n1sym2n2sym3n3...". The monosaccharide symbols are in alphabetical order.

CSForm

result = CSForm(scform, Optional gtype)

scform: String. Saccharide formula of a glycan as calculated by function Sacform; gtype: type of glycan, where "N" = N-glycan, "O" = O-glycan. Default = "N".

result: String. Chemical formula of the glycan using standard atomic symbols in the format "Sym1n1Sym2n2Sym3n3...". The atomic symbols are in alphabetical order.

CSPForm

result = CSPForm(scform, Optional gtype)

scform: String. Saccharide formula of a glycan as calculated by function Sacform;

gtype: String. Type of derivatized (for mass spec) glycan, where "N" = permethylated N-glycan, "O" = permethylated O-glycan, "TI" = toluidinated N- or O- glycan. Default = "N".

result: String. Chemical formula of the derivatized glycan using standard atomic symbols in the format "Sym1n1Sym2n2Sym3n3...". The atomic symbols are in alphabetical order.

MolMass

result = MolMass(cform)

cform: String. Chemical formula of a molecule or fragment.

result: Double. Monoisotopic mass of molecule or fragment.

MolWt

result = MolWt(cform)

cform: String. Chemical formula of a molecule or fragment.

result: Double. Molecular weight of molecule or fragment.

Ribbon item macros

Group	Item	Macro
Build	Network	SetupNetL
	Model	SetupModelL
Calculate	SetupReact	SetupReact
	Solve	SparseSolve

Fit Data	One Pass	OneRun
	Fit	FitParm
Results	Results	Results
	Summary	Summary
Mass Spectra	Setup	SetupMassSpecSynX
	MS Syn	CalcMSSyn
	Process Raw	MassSpecSynX
	MS Plot	PlotFullData
	Quick MS	MassSpecX
	Plot Fit MS	PlotData
Tools	Draw Structure	DrawOneStructure
	Run Script	RunScript


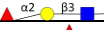


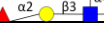
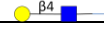
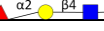



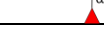
GLYMMER™ Example

Example: Prostate cancer stages

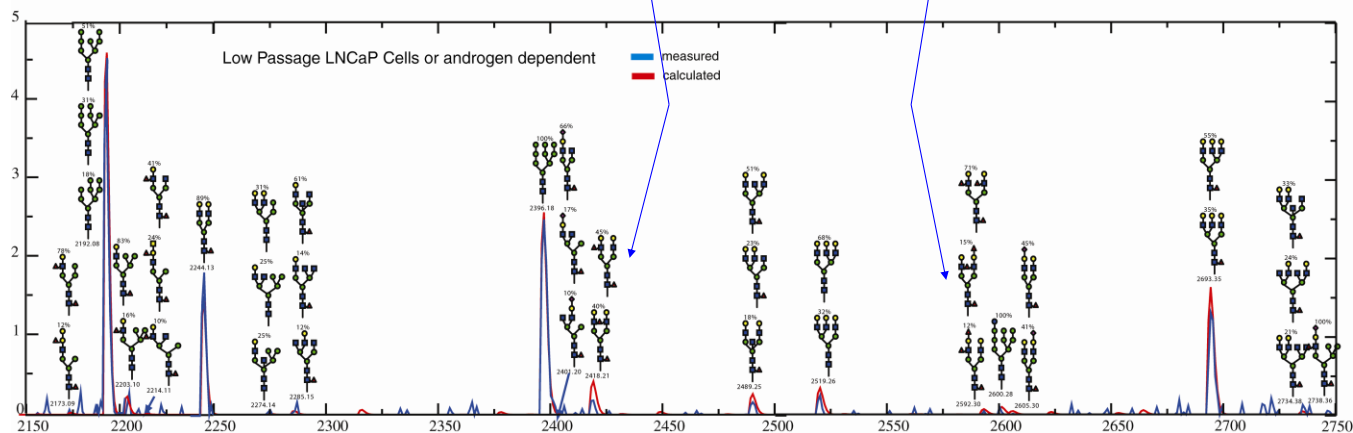
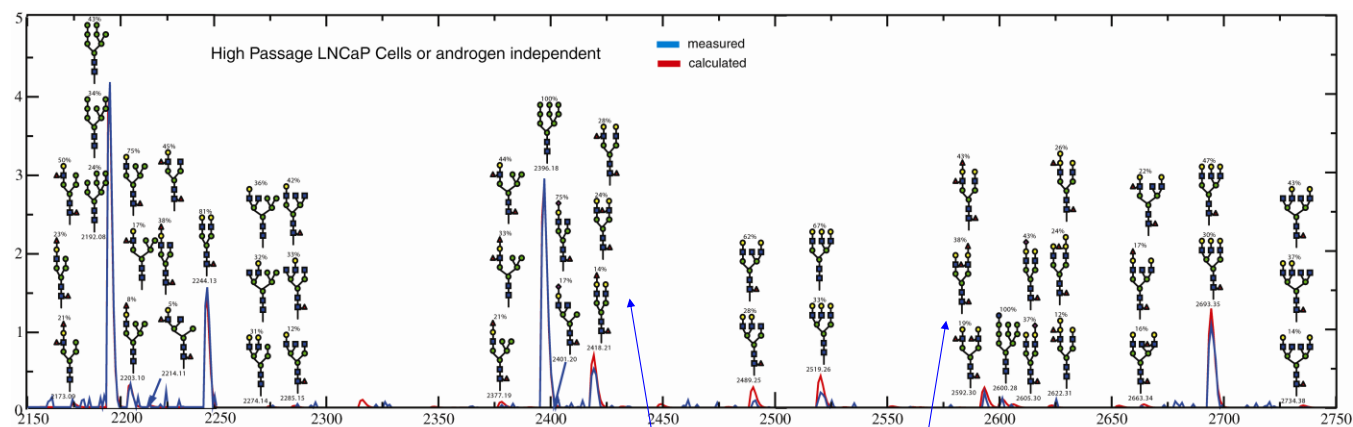
The model was applied to a set of glycan MS data and gene expression microarray data developed by PiWan Cheng's group for low and high passage LNCaP prostate cancer cells, which correspond to androgen-dependent and the more metastatic androgen independent cell stages. The enzyme activities in the model were adjusted to bring model-calculated synthetic mass spectra into agreement with experimental N-glycan mass spectra for both cell types, resulting in both an automatic identification of the structures corresponding to the mass spectral peaks and estimated enzyme activity profiles for the two cell types.

Model-predicted glycan substructures

Based on the model-predicted structures we see that the high passage cells have significantly higher levels of H and Le^y epitopes, which are known biomarkers for prostate cancer.














Glycan moiety	Structure	Key Enzymes	Low passage Abundance (%)	High passage Abundance (%)
Type I chain		b3GalT	22.34	18.43
H(I)		FucTH	2.17	4.06
Le ^a		FucTLe	0	0
Le ^b		FucTLe	0	0
Type II chain		b4GalT	105.83	88.34
H(II)		FucTH	2.67	4.66
Le ^x		FucTLe a3FucT	13.39	15.45
Le ^y		FucTLe a3FucT	1.60	3.57
A		GalNAcTA	0.06	0.16
B		GalTB	0	0
SLe ^x			0.04	0.04

Comparison of model calculated synthetic mass spectra with measured spectra for low passage and high passage human LNCaP cells for part of the mass range. Annotation is completely automatic.



Model-predicted glycan types

By providing essentially complete annotations of the mass spectra the model allows quantification of glycan types and substructures not normally available from conventional annotation techniques. Here we see that the more metastatic high passage cells have higher levels of hybrid glycans and lower levels of complex glycans, especially tetraantennary.

Glycan Type	Representative Structure	Key Enzyme	Low passage Abundance (% of total)	High passage Abundance (% of total)
High mannose		ManI	48.55	50.44
M9			8.44	9.03
M8			14.12	12.93
M7			9.93	8.80
M6			12.13	14.15
M5			3.94	5.54
Hybrid			6.99	9.20
M5		GnTI	3.07	4.34
M4		ManII	1.54	2.18
M3		ManII	2.37	2.68
Complex			44.46	40.36
Biantennary		GnTII	10.60	12.07
Triantennary(1)		GnTIV	4.35	4.04
Triantennary(2)		GnTV	5.85	6.22
Tetraantennary		GnTIV, GnTV	23.67	18.03
Bisected		GnTIII	1.04	0.17
Core Fucosylated		a6FucT	41.45	34.10
Lactosamine repeat		iGnT	0.82	0.57
Lactosamine branch		IGnT	0.16	0.14
Sialylated		a3SiaT, a6SiaT	0.42	0.31