

Using inference on Semantic Web data to enrich the data in GlyCosmos

Kiyoko F. Aoki-Kinoshita

Glycan and Life Systems Integration Center, Soka University

What is the Semantic Web?

- Also known as Linked Open Data

Legend

Cross Domain

Geography

Government

Life Sciences

Linguistics

Media

Publications

Social Networking

User Generated

Legend

Cross Domain

Geography

Government

Life Sciences

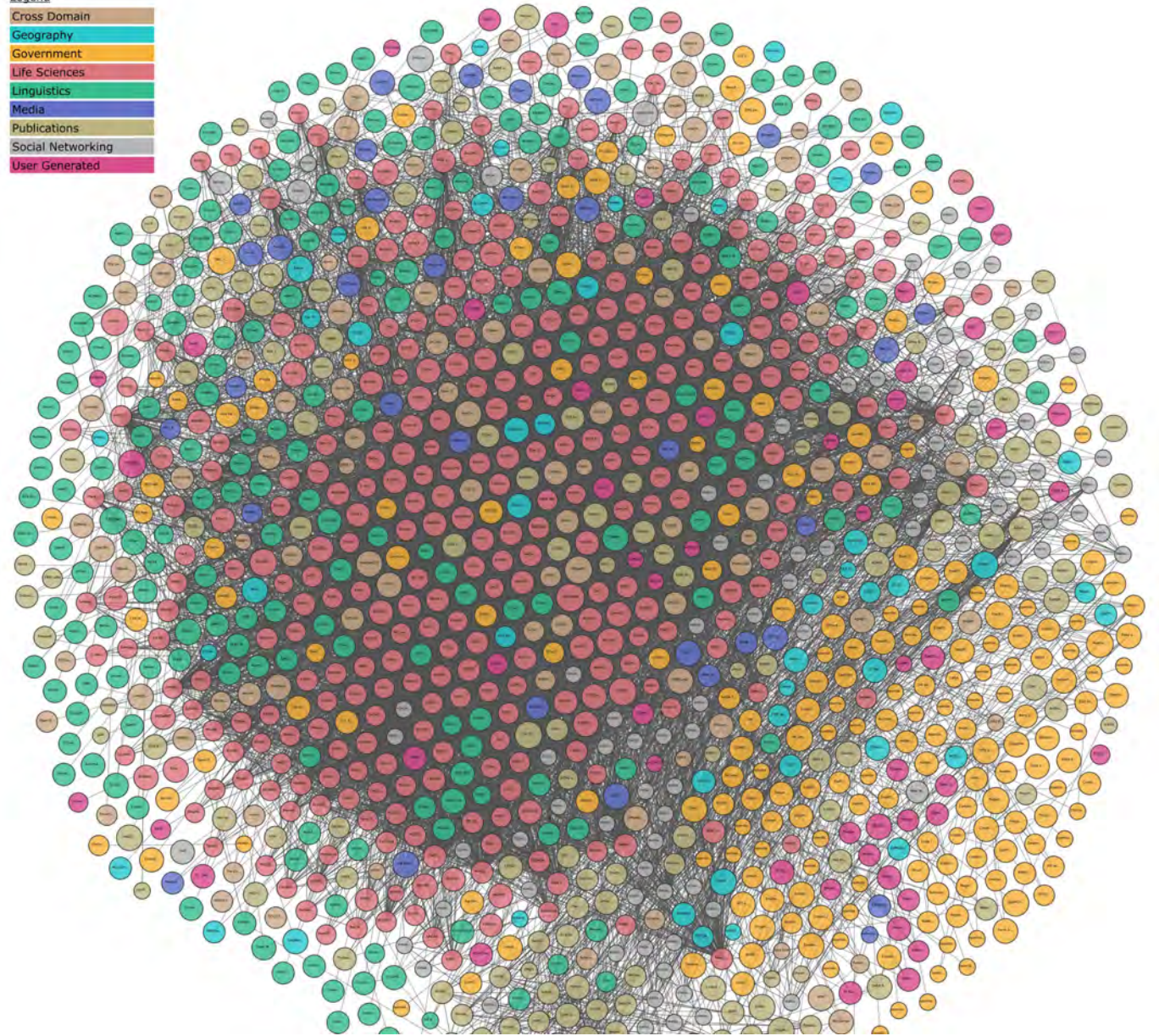
Linguistics

Media

Publications

Social Networking

User Generated

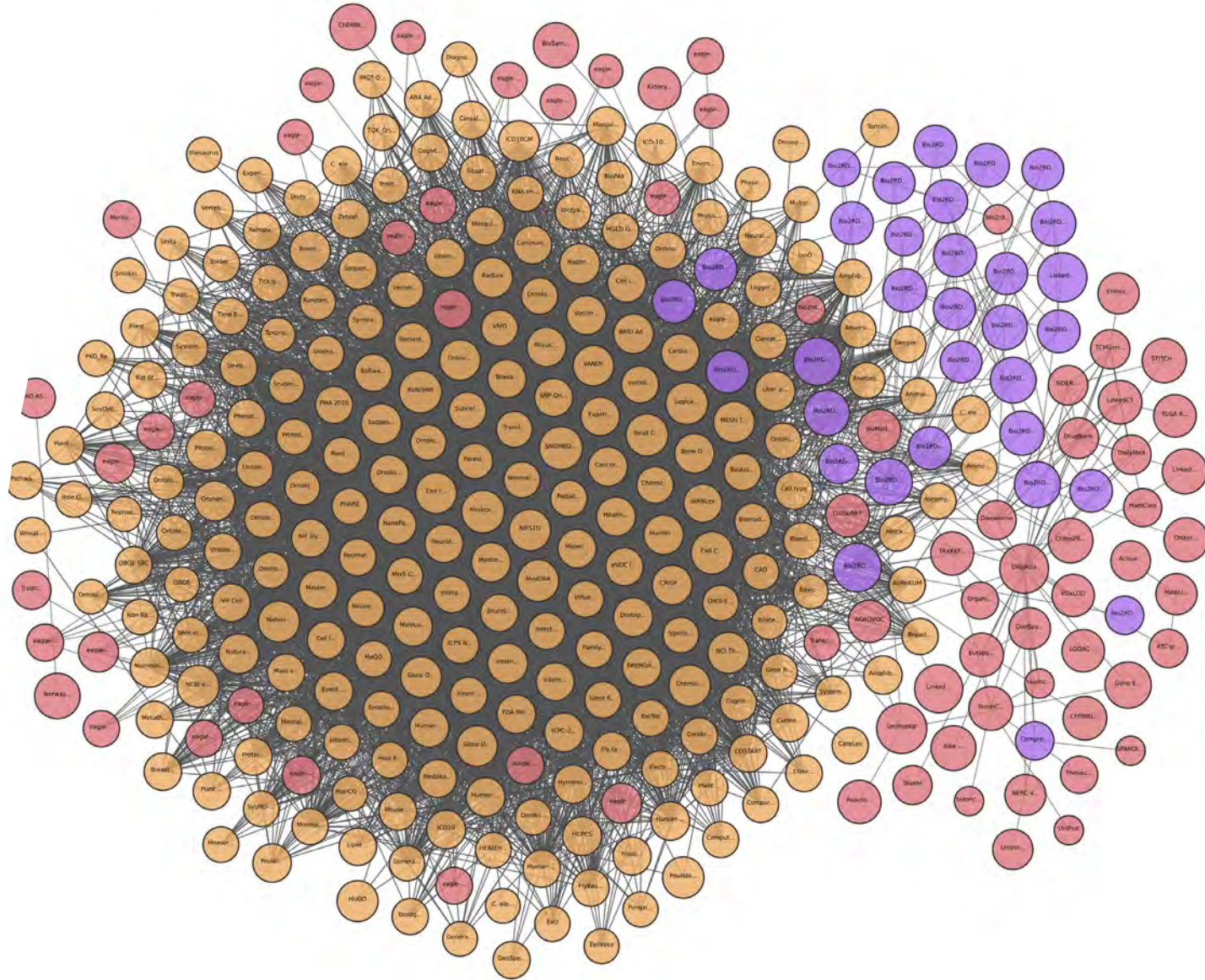


Life Sciences LOD

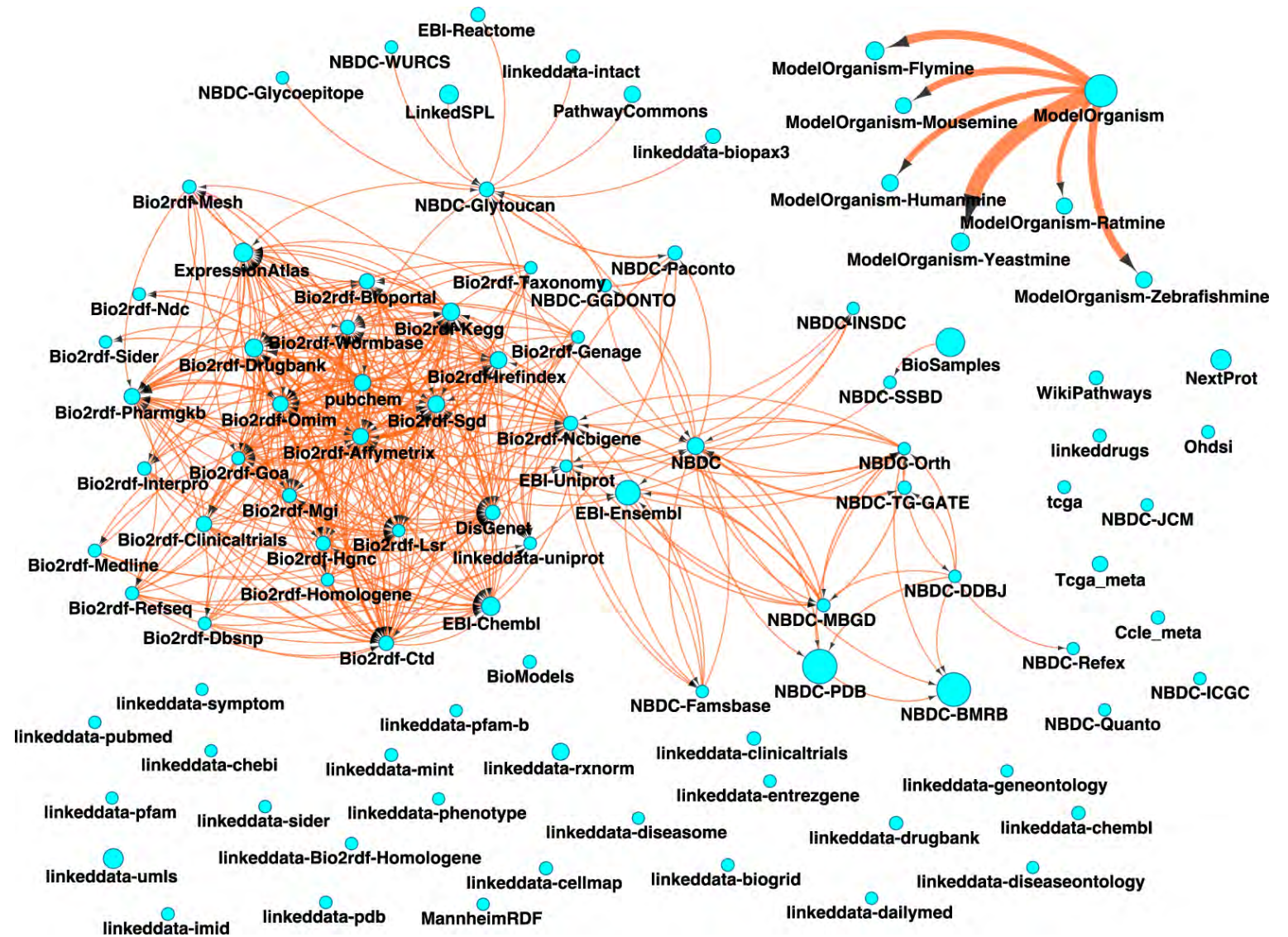
BioPortal

Bio2RDF

Other



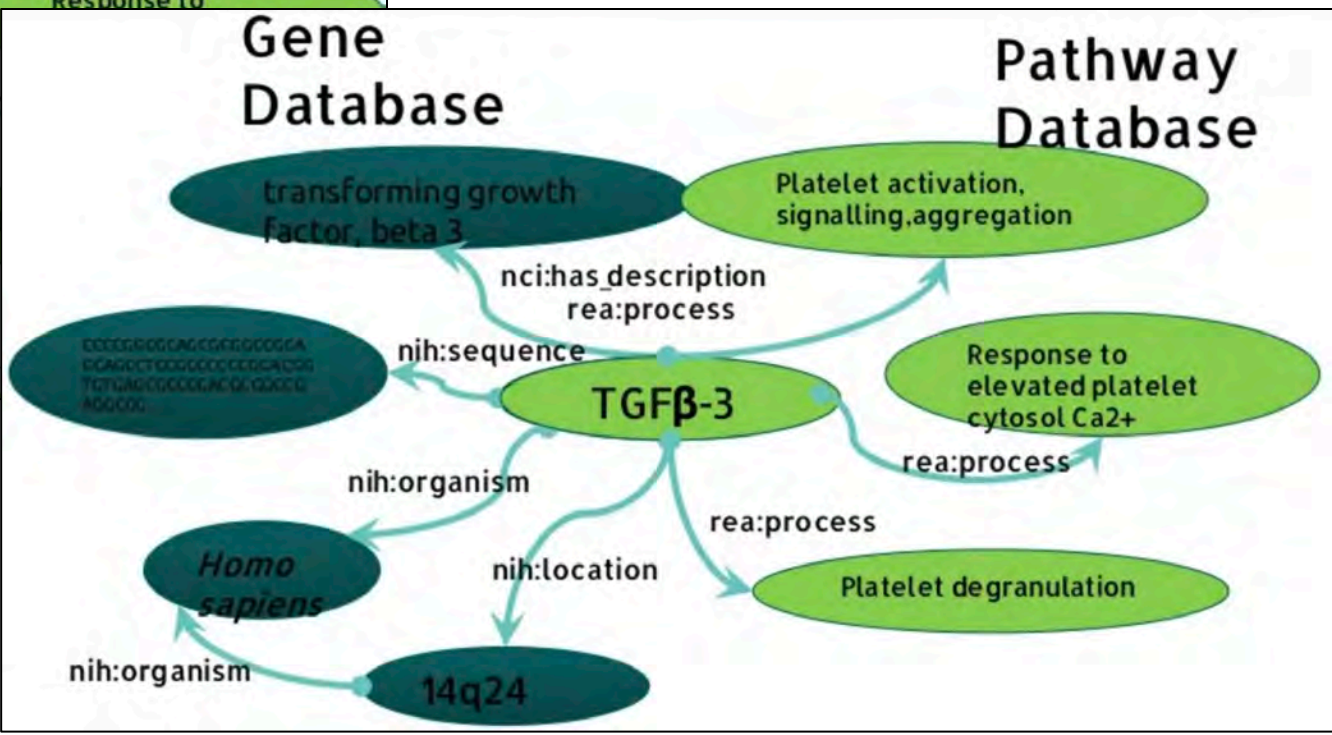
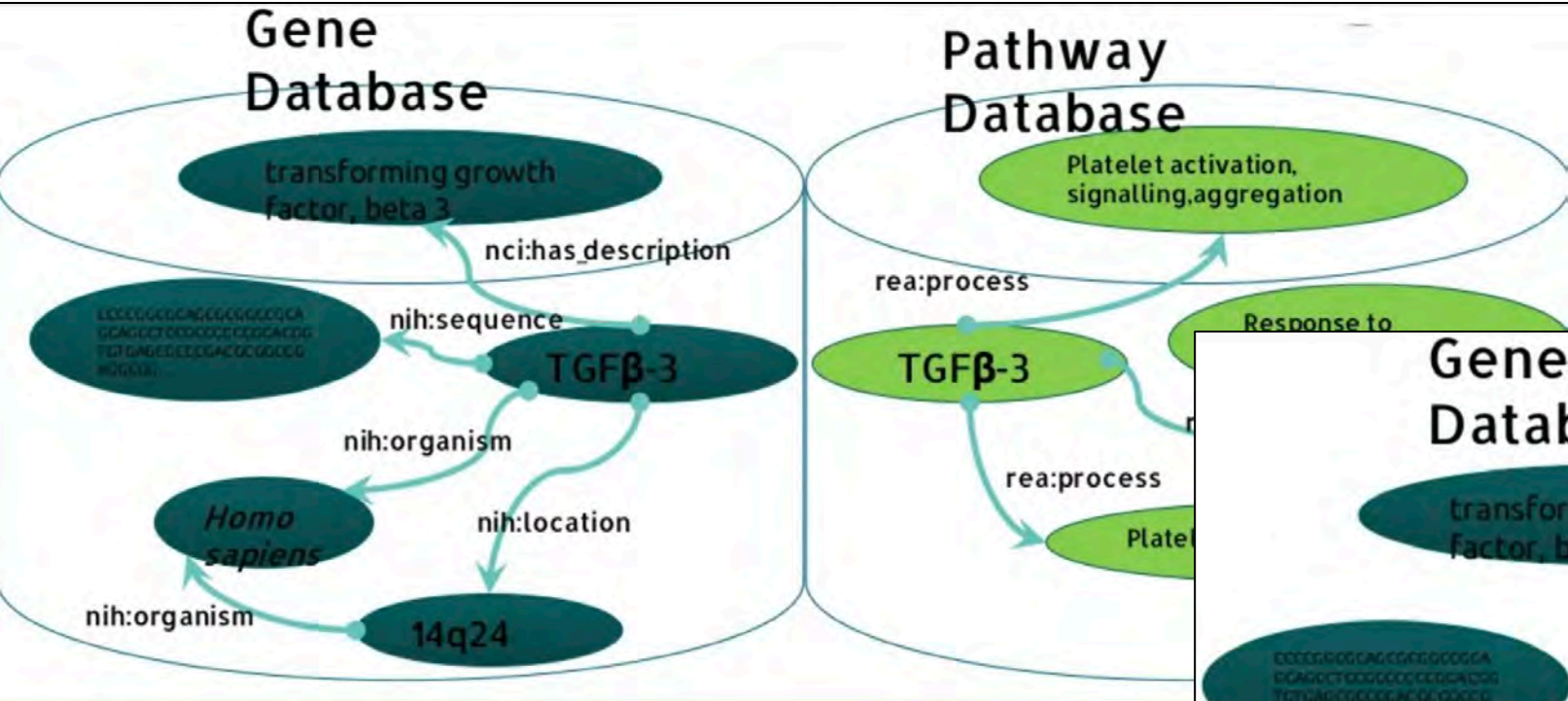
Life Sciences LOD Meta- analysis results



Advantages of SemWeb

- The Web is used as the foundation of a truly de-centralized database.
 - No need to set up or manage a central database
 - Anyone can join the cloud
- Truly open and shareable data
 - Easy for re-use and mash-up
 - Easy for cross-domain, cross-discipline use
- Single format for all data
 - Easy for data processing

Example of RDB vs. RDF



Inferencing on the SemWeb

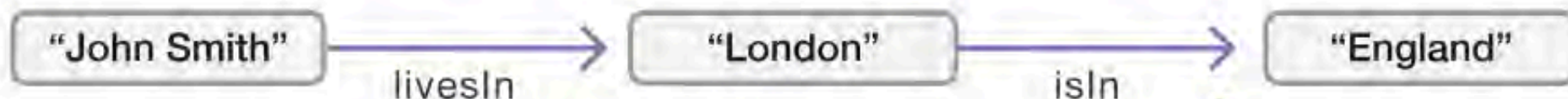
- Semantic inference or inference over the Semantic Web is a process by which new **data** is added to a dataset, created from the existing data.
 - No extra data has to be collected to produce new insights.
 - These insights come in the form of new **relationships**, providing connections in the data that were previously unobserved.
- Inference relies on two tools, **ontologies** and **rules**. The former describes the **structural model** of the data, how its layered into classes and sub-classes etc., and the latter dictates the laws which the data must obey.
 - An ontology will tell you that the countries in a continent are a subset of all the countries, and a rule will tell you that a city in a country must also be in that country's continent.

The Semantic Web: triples and graphs

DATA is stored in Triples, expressed as:

Subject	:	Predicate	:	Object
John Smith		livesIn		London
London		isIn		England

QUERY with SPARQL, gives us a simple look up... and more! Find people who live in (a place that's in) England



The GlyCosmos Gly

- **Submissions (Repositories)**
 - GlyTouCan (glycans)
 - GlycoPOST (raw MS data)
 - UniCarb-DR (annotated glycomics data)
 - GlyComb (glycoconjugates; currently only glycopeptides)
- **Data Resources**
 - Glycomes
 - Glycans
 - Glycoconjugates
 - Glycogenes
 - Diseases
 - Pathways
 - etc.
- **Tools**
 - Glycan drawing tools
 - GlycoMaple
- **Standards**

<https://glycosmos.org>

Version 3.0 released August 8, 2022

The screenshot shows the GlyCosmos Portal homepage. At the top, there is a navigation bar with the logo and links for Submissions, Resources, Tools, Standards, About, Help, Feedback, and a search bar. The main content area is divided into several sections:

- Submissions:** Lists GlyTouCan (Glycan Structure Repository), GlyComb (beta) (Glycoconjugate Repository), GlycoPOST (Glycomic MS Repository), and UniCarb-DR (Glycomic MS Repository).
- Resources:** Lists Glycogenes, Glycoproteins, Lectins, Glycolipids, Glycans, Pathways, Diseases, and Organisms.
- Tools:** Lists SugarDrawer, GlycanBuilder, Glycan Converters, GlycoMaple, GALAXY, MCAW, and GlycoSim.
- Search Tools:** Lists Glycan Search, Publication Search, Cross Search, and Programmatic Access.

The central area features a "Welcome to GlyCosmos!!" message and a "Cross Search" section with a search bar and a "Search" button. Below this are "Glycan Search" and "Publication Search" sections, each with a search bar and a "Search" button. The "Data Resources" section is titled "Main Data Resources" and includes:

- Glycogenes:** Integrated list of glycan-related genes, including glycosyltransferases, hydrolases, etc.
- Glycoproteins:** Integrated list of glycoproteins extracted from UniProt and annotated with glycosylation data from GlyGen and GlyConnect. For each entry, information such as glycosylation site, glycans, diseases, 3D structures, and pathway information are available.
- Lectins:** Protein entries annotated as carbohydrate-binding in UniProt. Glycosylation site information, along with glycan-binding patterns are also integrated where available.
- Glycolipids:** Glycolipid entries from the LIPID MAPS Structure Database (LMSD).
- Glycans:** The list of validated glycans extracted from GlyTouCan, updated weekly.
- Pathways:** Pathways from Reactome containing glycoproteins as annotated in UniProt. Search by pathway name, species, and protein name.
- Diseases:** List of diseases involving glycan related genes. The information of each database of Glyco-Disease Genes Database (GDGDB), DisGeNET, and Alliance of Genome Resources is integrated into one list.
- Organisms:** List of all species in GlyCosmos Resources.

At the bottom right, there are buttons for "Release Statistics" and "Download". On the right side of the page, there is a "News" section with a "Tweets by @GlyCosmos" feed. The feed includes a tweet from GlyCosmos about a beta release update and a retweet from the C14CB research group about a new GlySpace Alliance website.

Standards

- Ontologies
- Nidations

Other Information

Release Statistics

Downloads

GlyCosmos Lectins

Database	Last Updated
MCAW-DB	July 10, 2019
UniLectin	November 20, 2020
UniProt	July 20, 2022

GlyCosmos Pathways

Database	Last Updated
Reactome	February 8, 2022 (Version 79)
UniProt	February 8, 2022

Glyco-Disease Genes Database (GGDDB)

Database	Last Updated
Glyco-Disease Genes Database (GGDDB)	January 25, 2017

GlycoEpitope

Database	Last Updated
GlycoEpitope	November 18, 2015

GlycoGene Database (GGDB)

Database	Last Updated
GlycoGene Database (GGDB)	January 26, 2018

GlycoMaple

Database	Last Updated

GlycoNAVI-Glycans

Database	Last Updated
Carbohydrate Database (PACDB)	Pathogen Adherence to Carbohydrate Database (PACDB)
	June 1, 2016

GlycoNAVI-Lectins

Database	Last Updated
Plant GARDEN	Plant GARDEN
	May 28, 2020

GlycoNAVI-Proteins

Database	Last Updated
SugarBind	SugarBind
	December 14, 2021

GlycoProtDB

Database	Last Updated

GlycomeAtlas

Database	Last Updated

LIPID MAPS Gene/Proteome Database

Database	Last Updated

LM-GlycomeAtlas Lectin Frontier DataBase (LfDB)

Database	Last Updated

MatrixDB

Database	Last Updated

PSICQUIC

Database	Last Updated

Pathogen Adherence

Database	Last Updated

To see the latest features, please visit our beta site.

Home License Logo Policies Contact: support@glycosmos.org

International Collaboration

GlyCosmos is a member of the *GlySpace Alliance* together with GlyGen and Glycomics@EXPASY.



Acknowledgements

Supported by JST NBDC Grant Number JPMJND2204



Partly supported by NIH Common Fund Grant #1U01GM125267-01



This work is licensed under a Creative Commons Attribution 4.0 International License

Metadata for Data Resources

Alliance of Genome Resources

Database	Last Updated
Alliance of Genome Resources	February 1, 2021

CAZy

Database	Last Updated
CAZy	May 9, 2022
UniProt	May 9, 2022

DisGeNET

Database	Last Updated
DisGeNET	June 29, 2021

FlyGlycoDB

Database	Last Updated
FlyGlycoDB	September 1, 2018

GALAXY

Database	Last Updated
GALAXY	March 31, 2022

GlyCosmos Diseases

Database	Last Updated
Alliance of Genome Resources	February 1, 2021
DisGeNET	June 29, 2021
Glyco-Disease Genes Database (GGDDB)	January 25, 2017

GlyCosmos Glycans

Database	Last Updated
GlyYouCan	September 12, 2022

GlyCosmos Glycogenes

Database	Last Updated
FlyGlycoDB	September 1, 2018
Glyco-Disease Genes Database (GGDDB)	January 25, 2017
GlycoGene Database (GGDB)	January 26, 2018
KEGG BRITE	July 19, 2022
LIPID MAPS Gene/Proteome Database	June 24, 2019
Plant GARDEN	May 28, 2020
UniProt	July 12, 2022

GlyCosmos Glycolipids

Database	Last Updated
CHEBI	February 3, 2022
KEGG	February 3, 2022
LIPID MAPS Structure Database	February 3, 2022
LipidBank	February 3, 2022
PubChem	February 3, 2022

GlyCosmos Glycoproteins

Database	Last Updated
GlyConnect	February 2, 2022
GlycoProtDB	October 1, 2021
GlyGen	February 2, 2022
Human Protein Atlas	September 5, 2019 (Protein Atlas version 19)
MCAW-DB	July 10, 2019
Reactome	February 8, 2021 (Version 79)

Submissions

GlyYouCan
Glycan Structure Repository

GlyComb (beta)
Glycoconjugate Repository

GlycoPOST
Glycomic MS Repository

UniCarb-DR
Glycomic MS Repository

Resources

Glycogenes

Glycoproteins

Lectins

Glycolipids

Glycans

Pathways

Diseases

Organisms

All Resources

Genes / Proteins / Lipids

Glycans / Glycoconjugates

Glycomes

Pathways / Interactions /

Diseases / Organisms

Tools

SugarDrawer

GlycanBuilder

Glycan Converters

GlycoMaple

GALAXY

MCAW

GlycoSim

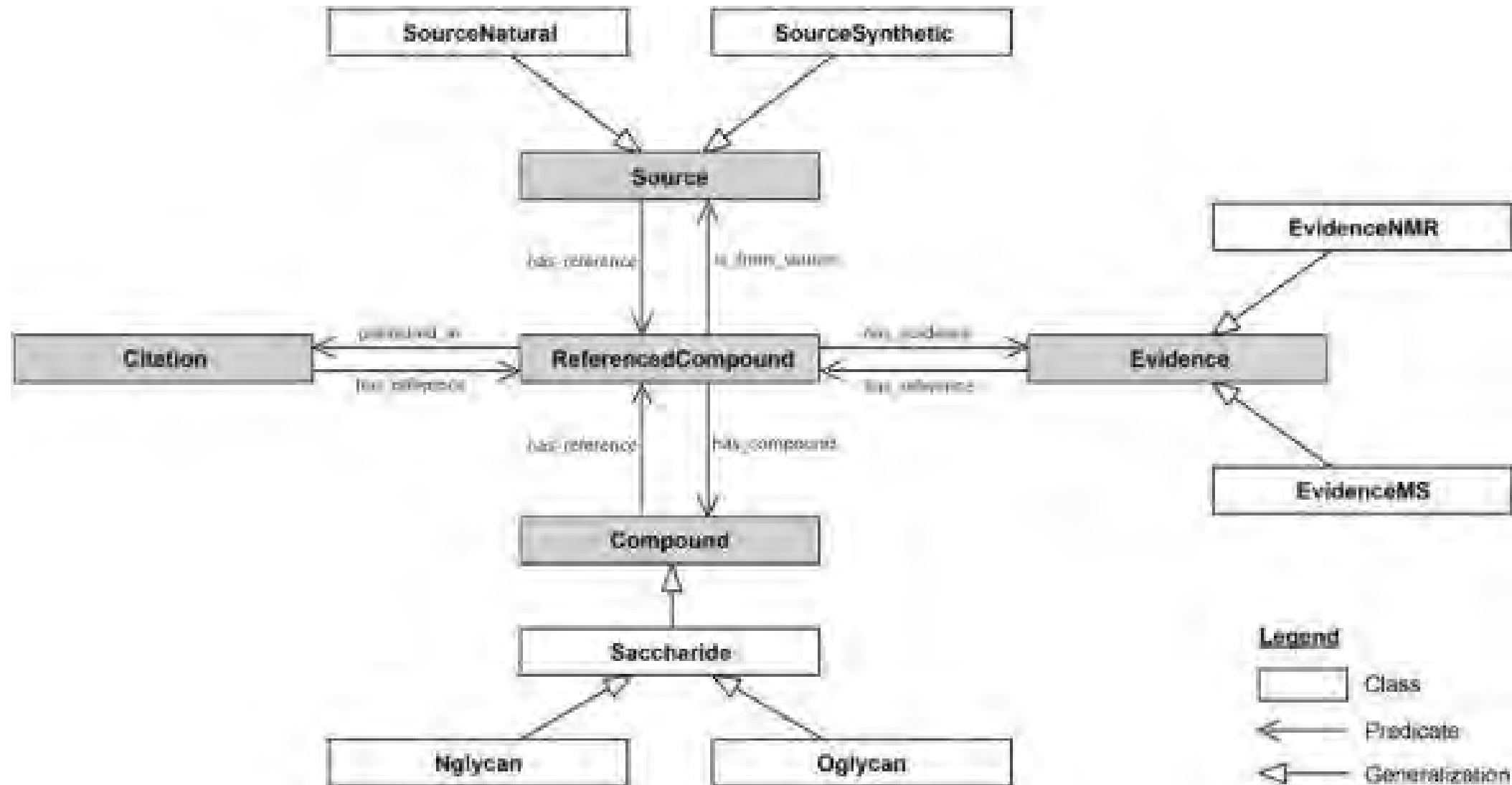
Search Tools

Glycan Search

Publication Search

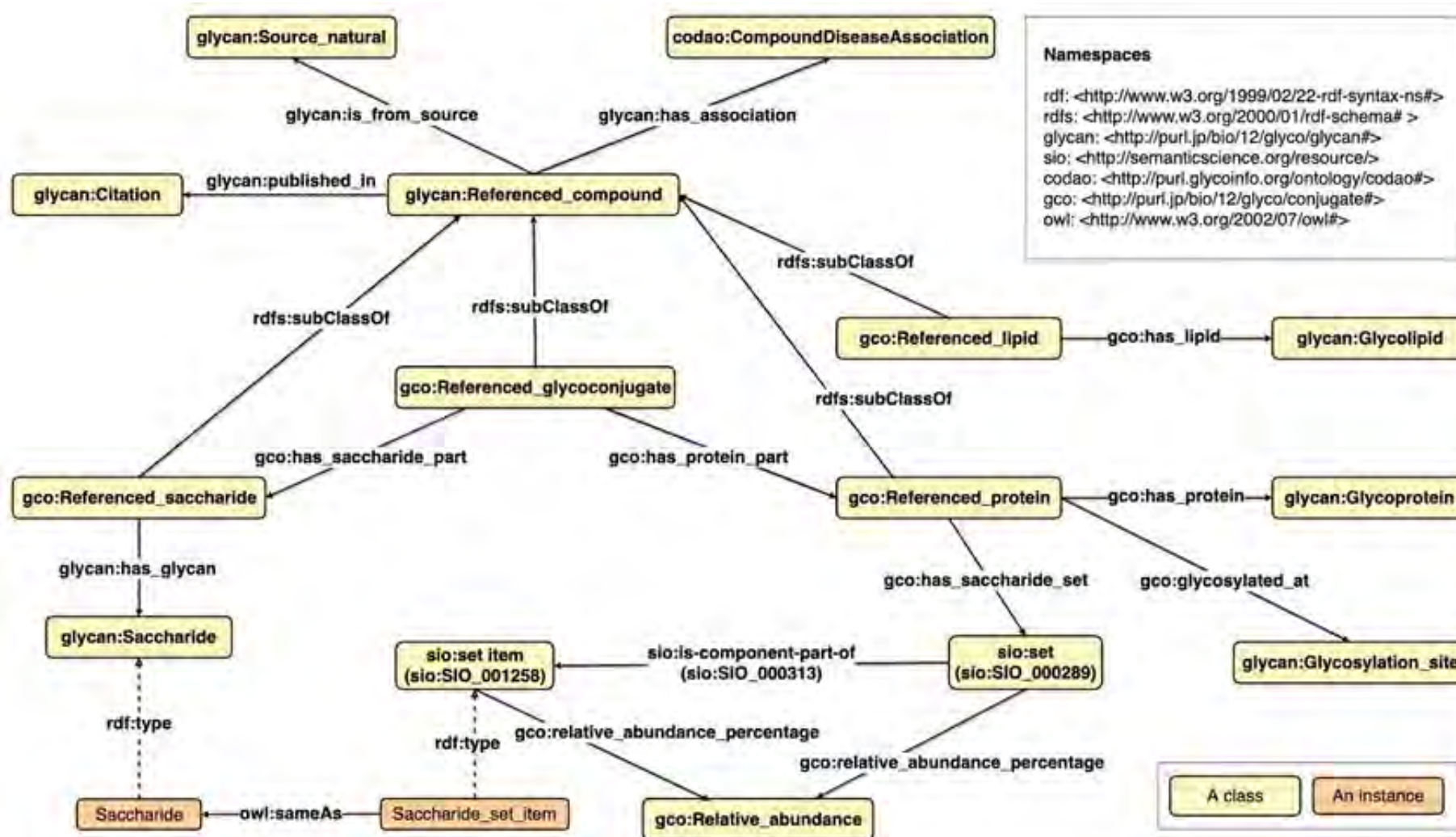
Cross Search

GlycoRDF (glycan ontology)



GlycoCoO (Glycoconjugate Ontology)

<https://github.com/glycoinfo/GlycoCoO>





GlyCosmos Pathways

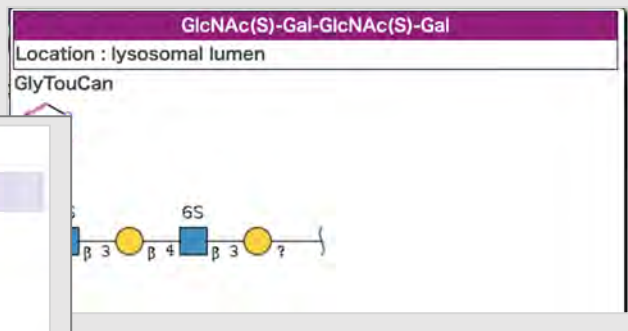
R-HSA-2022857

Pathway Name: Keratan sulfate
Species: Homo sapiens

Glycosylation Sites

#	Position	Description	PubMed ID	Glycan
1	26	N-linked (GlcNAc...) asparagine		
2	247	N-linked (GlcNAc...) asparagine	22128166	$\triangleright\text{O}_1$
			24737316	$\triangleright\text{O}_1$
3	464	N-linked (GlcNAc...) asparagine	G05724UK	$\triangleright\text{O}_1$
			G05724UK	$\triangleright\text{O}_1$
			G06110VR	$\triangleright\text{O}_1$
			G10773YW	$\triangleright\text{O}_1$
			G11870QZ	$\triangleright\text{O}_1$
			G11911BT	$\triangleright\text{O}_1$
			G20210JR	$\triangleright\text{O}_1$
			G23294PN	$\triangleright\text{O}_1$
			16263699	$\triangleright\text{O}_1$
			19159218	$\triangleright\text{O}_1$
			22128166	$\triangleright\text{O}_1$
			24737316	$\triangleright\text{O}_1$
G23984SE	$\triangleright\text{O}_1$			
G39188ZX	$\triangleright\text{O}_1$			
G51640EO	$\triangleright\text{O}_1$			
G66621EA	$\triangleright\text{O}_1$			
G72735IY	$\triangleright\text{O}_1$			
G74724QE	$\triangleright\text{O}_1$			

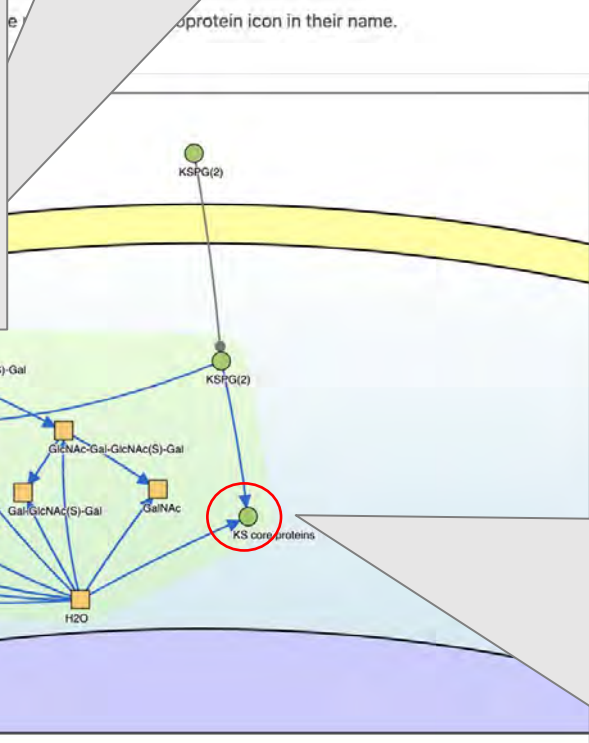
【Pathway
Reactome



Lectin Frontier DataBase (LfDB)

Top 3 high affinity glycans and the actual measurement ($V-V_0$) values by an automated frontal affinity chromatography with fluorescence detection from LfDB.

GlyYouCan	$V-V_0$
G65070LE	76.458
G67280SN	35.628
G55601IV	13.412



Location : lysosomal lumen

UniProtID ProteinName

O60938	Keratanase
P16112	Aggrecan core protein
P20774	Mimecan
P51884	Lumican
P51888	Prolargin
Q06828	Fibromodulin







BGAL

UniProtID ProteinName

P16278	Beta-galactosidase
Q6UWU2	Beta-galactosidase-1-like protein

New Tools section in GlyCosmos

Tools

-  SugarDrawer
-  GlycanBuilder
-  Glycan Converters
-  GlycoMaple
-  GALAXY [↗](#)
-  MCAW [↗](#)
-  GlycoSim [↗](#)

- Drawing tools to generate SNFG images
- Text format conversion tools for glycan text representations
- GlycoMaple: visualizing glycome expression data on glycan pathways
- GALAXY: 3D HPLC data
- MCAW: multiple glycan alignment
- GlycoSim: glycan biosynthesis simulation

Please input .csv file (Gene ID, TPM value). Do you need [sample](#) files?

Choose File No file chosen

Submit

Complex: Min

Overlap: Ave

If two or more tabs are open in the diagram, two datasets can be compared using the Compare button below.

Compare

LLO biosynthesis and OST

N-glycan processing and branching

Complex capping N-glycans/O-glycans/GSLs

Biosynthesis of GPI-APs

Biosynthesis of mucin-type O-glycans

Biosynthesis of O-Fuc/O-Glc/Col-Gal/O-GlcNAc/C-Man

Biosynthesis of O-Man

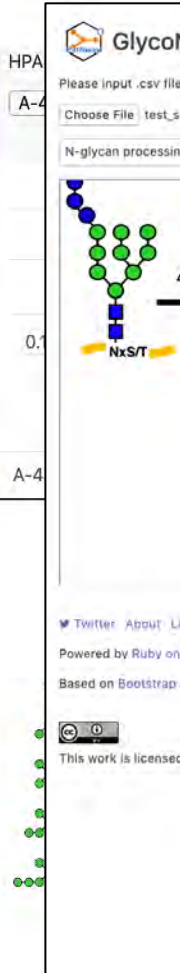
Biosynthesis of GAG

Biosynthesis of Heparan sulfate (HS)

Biosynthesis of Chondroitin sulfate (CS) and dermatan sulfate (DS)

Biosynthesis of Keratan sulfate (KS)

Biosynthesis of



default

A-4

Download Image

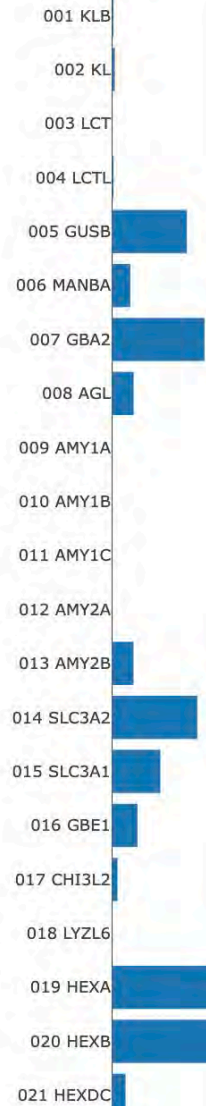
Legend

Glycosylhydrolase

Carbohydrate Binding Module

Download Image

Legend

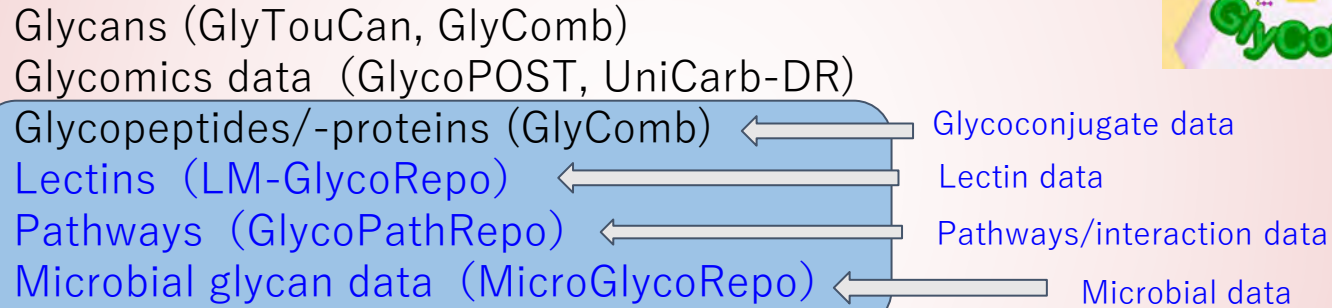


colon

The new GlyCosmos project to supplement and enhance glyco-data

Repository expansion

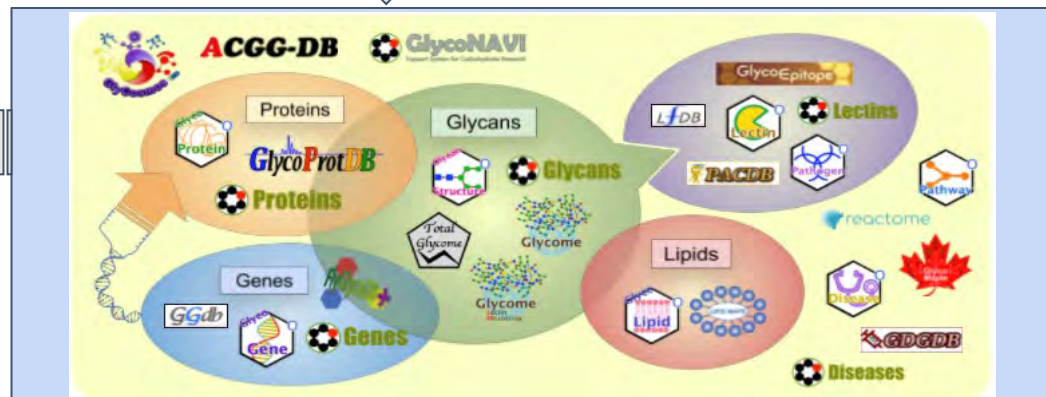
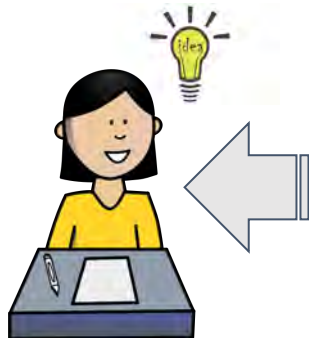
MIRAGE
WIBACE



Easier access to data for understanding glycan function!

- Semi-automatic curation system
- Annotation jamborees

Data Resource expansion



All linked in the [Semantic Web](#)

Glycan Pathway Repository

Pathway information Input

Basic Information of Glycosylation Pathway

Title

Description

Pathway Category

Species

Tissue

Cell Type

Backbone

Related Disease

Pathway Category

- classic metabolic pathway
- signaling pathway
- regulatory pathway
- disease pathway
- drug pathway

Tissue List

- animal
- organism form
- plant
- other source
- fungus

Pathway ontology

NCBI Taxon

Brenda tissue ontology

Cell ontology

UniProt protein

Disease ontology

Reaction Information Input

Add of New Reaction

Reaction Number

Reactant Glycan

Enzyme

Sugar Nucleotide

Product Glycan

Cellular Localization

- GlyToucan ID
- IUPACcondensed
- IUPACextended
- Linear code

EC enzyme

Text of Glycan

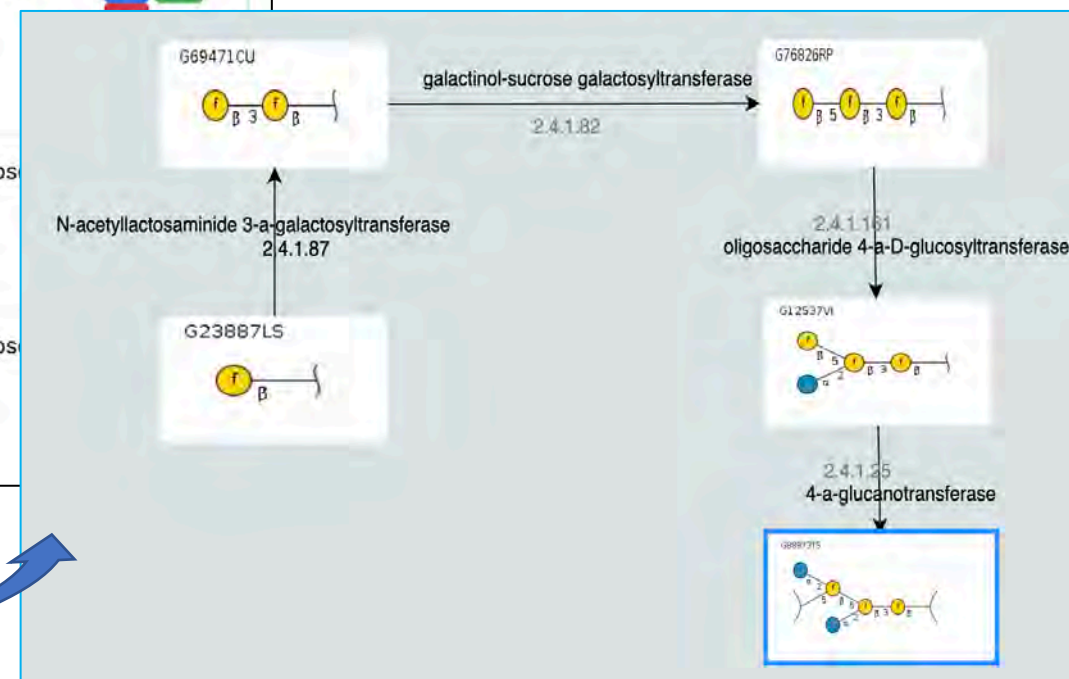
Cellular location

➤ Glycan Pathway Repository: Uploaded results

• Confirmation of input reaction steps

Reaction Table						
RXN ID	Reactant	Enzyme	Sugar Nucleotide	Product	Cell Location	Action
1	 G49108TO	rhamnogalacturonan 1 rhamnosyltransferase	L-Rhamnose	 G16062FD	cytosol	
2	 G16062FD	rhamnogalacturonan 1 rhamnosyltransferase	L-Rhamnose	 G27187IX	cytosol	
3	 G27187IX	rhamnogalacturonan 1 rhamnosyltransferase	L-Rhamnose	 G22722WA	cytosol	
4	 G22722WA	protein O-mannose b-1,4-N-acetylglucosaminyltransferase	D-Mannosamine	 G87777BW	cytosol	

➤ The visualized pathway information



Upload

- [Home](#)
- [Search](#)
- [About](#)
- [Downloads](#)
- [SPARQL](#)
- [Feedback](#)

Welcome to MicroGlycoDB!!

MicroGlycoDB is a new database of glycan-related information in microorganisms using the **Semantic Web** technologies, in order to elucidate various biological processes that occur in vivo in these organisms.

[Learn More](#)

- Organisms**
- [Bifidobacterium](#)
 - [Campylobacter](#)
 - [Cryptococcus](#)
 - [Mycobacterium](#)

- Glycan drawers**
- [CSDB/SNFG structure editor](#)
 - [DrawGlycan-SNFG](#)

- Other resources**
- [UniProt](#)
 - [Rhea](#)
 - [KEGG](#)
 - [ChEBI](#)
 - [PubChem](#)
 - [Cazy](#)
 - [CDD](#)

-  **Organisms**
Please click each icon.
- **Bifidobacterium**
 - *Bifidobacterium bifidum*
 - *Bifidobacterium longum*
 - **Campylobacter**
 - *Campylobacter jejuni*
 - **Cryptococcus**
 - *Cryptococcus neoformans*
 - **Mycobacterium**
 - *Mycobacterium abscessus*
 - *Mycobacterium tuberculosis*

 [Downloads](#)



Soka University
International Collaborative
Research Grant

- CCRC, U. of Georgia, USA
- U. of Lille
- Others

MicroGlycoDB prototype
<https://microglycodb.alpha.glycosmos.org/>

Name	Affiliation
Kazuhiro Aoki	Complex Carb. Res. Ctr., Univ. of Georgia
Christine M. Szymanski	Complex Carb. Res. Ctr., Univ. of Georgia
Yann Guerardel	CNRS UGSF, Univ. Lille
Louis-David Leclercq	CNRS UGSF, Univ. Lille
Tamara L. Doering	Dept. of Mol. Microbio., Washington Univ.
Thomas Hurtaux	Dept. of Mol. Microbio., Washington Univ.
Kiyotaka Fujita	Grad. Sch. of Agricultural Sci., Kagoshima Univ.
Takane Katayama	Grad. Sch. of Biostud., Kyoto Univ.
Toshihiko Katoh	Grad. Sch. of Biostud., Kyoto Univ.

- Home
- Search
- About
- Downloads
- SPARQL
- Feedback

Organisms

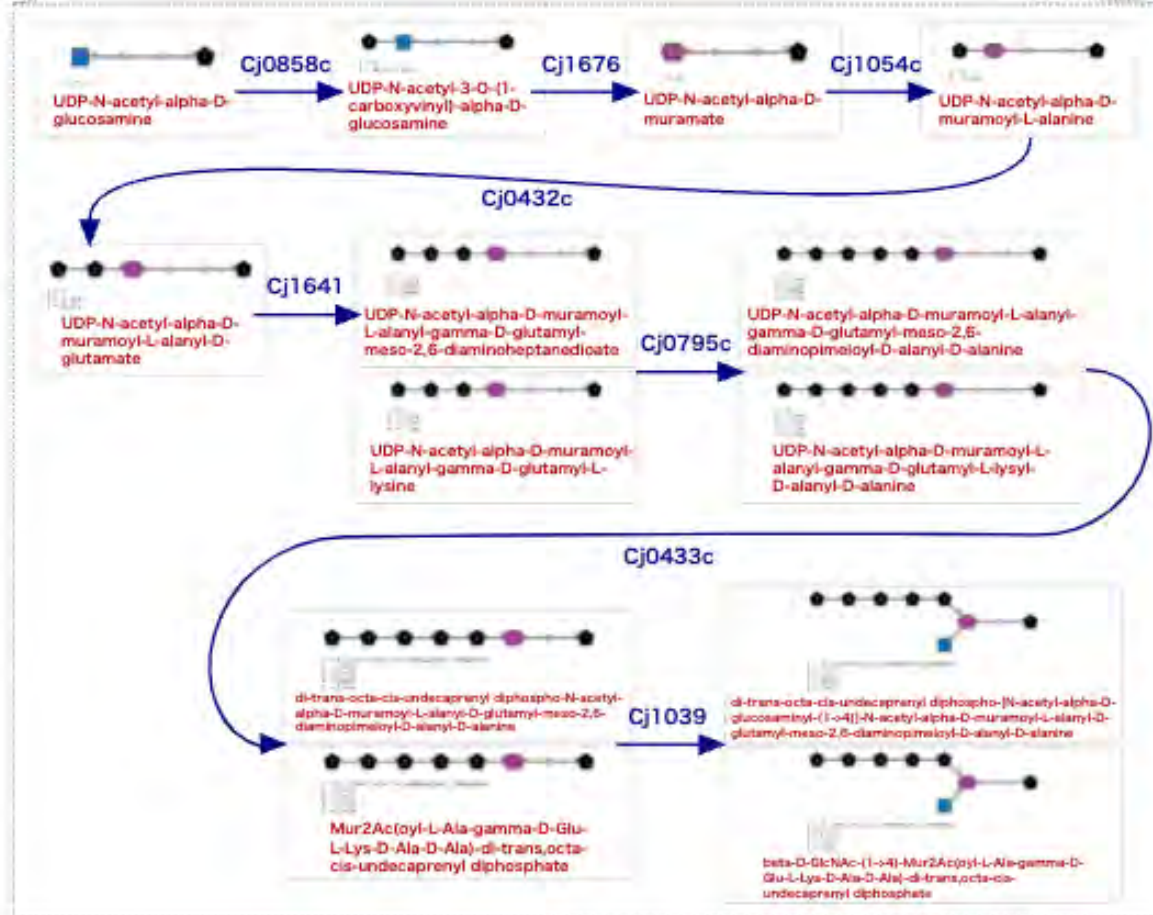
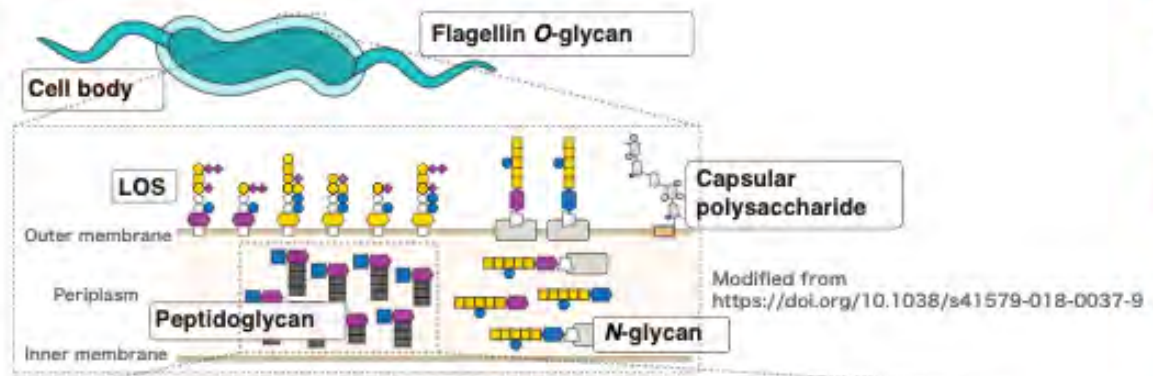
- Bifidobacterium*
- Campylobacter*
- Cryptococcus*
- Mycobacterium*

Glycan drawers

- CSDB/SNFG structure editor
- DrawGlycan-SNFG

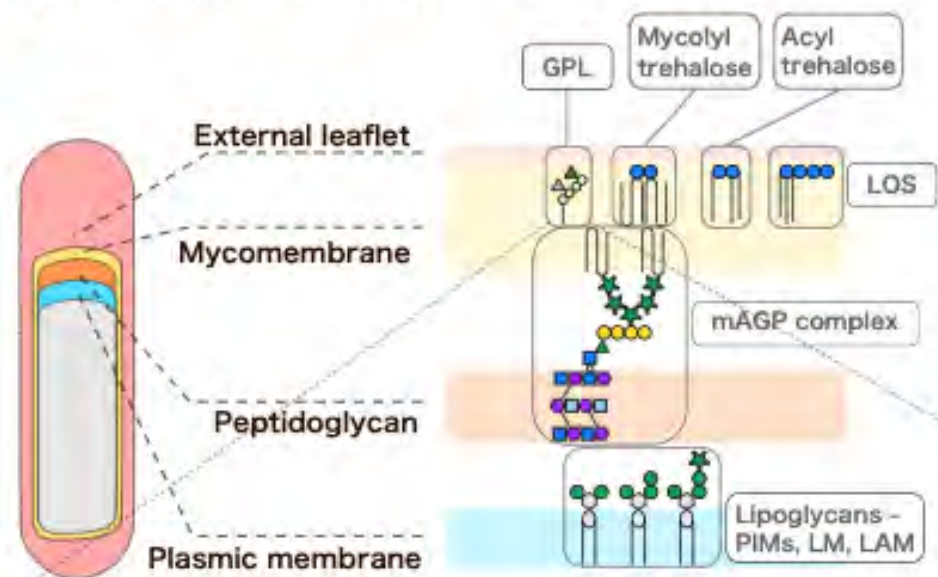
Other resources

- UniProt
- Rhea
- KEGG
- ChEBI
- PubChem
- Cazy
- CDD



Gene ID	Gene No.	Gene Name	NOTES - phase variation etc
microglycodb_gene_00186	Cj1039	<i>murG</i>	UDP-diphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase, leads to the formation of Lipid II
microglycodb_gene_00187	Cj1676	<i>murB</i>	Catalyzes the reduction of UDP N-acetylglucosamine enolpyruvate to form UDP-N-acetylmuramate in peptidoglycan biosynthesis
microglycodb_gene_00188	Cj1054c	<i>murC</i>	Catalyzes the formation of UDP-N-acetylmuramoyl-L-alanine from UDP-N-acetylmuramate and L-alanine in peptidoglycan synthesis
microglycodb_gene_00189	Cj0432c	<i>murD</i>	DP-N-acetylmuramoylalanine-D-glutamate ligase
microglycodb_gene_00190	Cj0433c	<i>mraY</i>	First step of the lipid cycle reactions in the biosynthesis of the cell wall peptidoglycan, leads to the formation of Lipid I
microglycodb_gene_00191	Cj0858c	<i>murA</i>	Enolpyruvyl to UDP-N-acetylglucosamine as a component of cell wall formation
microglycodb_gene_00192	Cj1641	<i>murE</i>	Involved in cell wall formation; peptidoglycan synthesis; cytoplasmic enzyme; catalyzes the addition of meso-diaminopimelic acid to the nucleotide precursor UDP-N-aceylmuramoyl-l-alanyl-d-glutamate
microglycodb_gene_00193	Cj0795c	<i>murF</i>	UDP-N-acetylmuramoyl-tripeptide
microglycodb_gene_00194	Cj0798c	<i>ddl</i>	D-alanine-D-alanine ligase
microglycodb_gene_00195	Cj1652c	<i>murI</i>	Converts L-glutamate to D-glutamate, a component of peptidoglycan
			Cell division / peptidoglycan biosynthesis

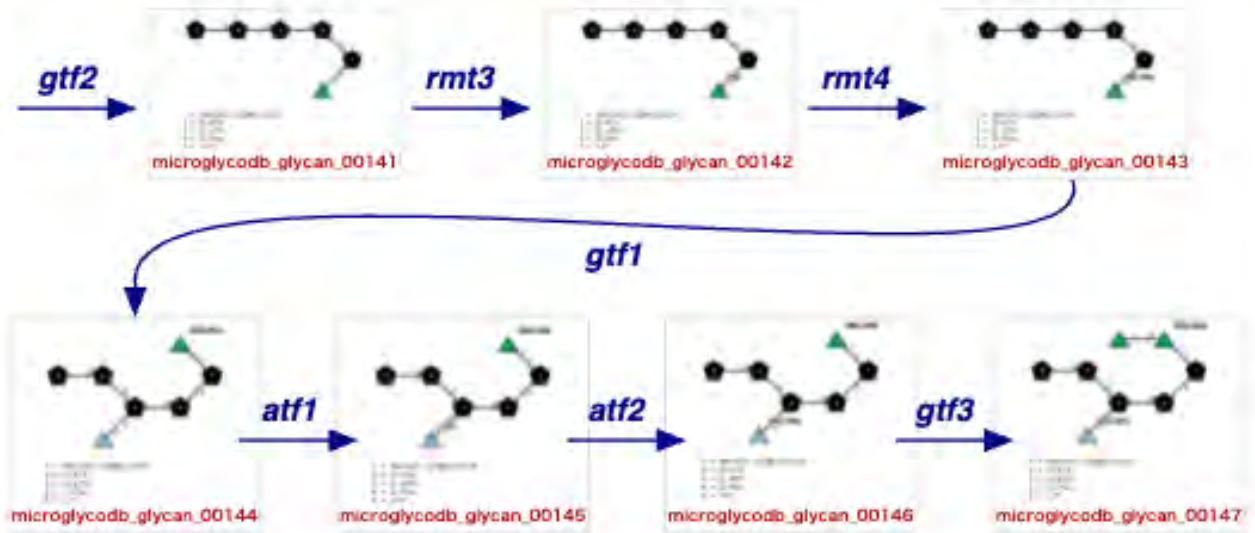
Mycobacterium abscessus



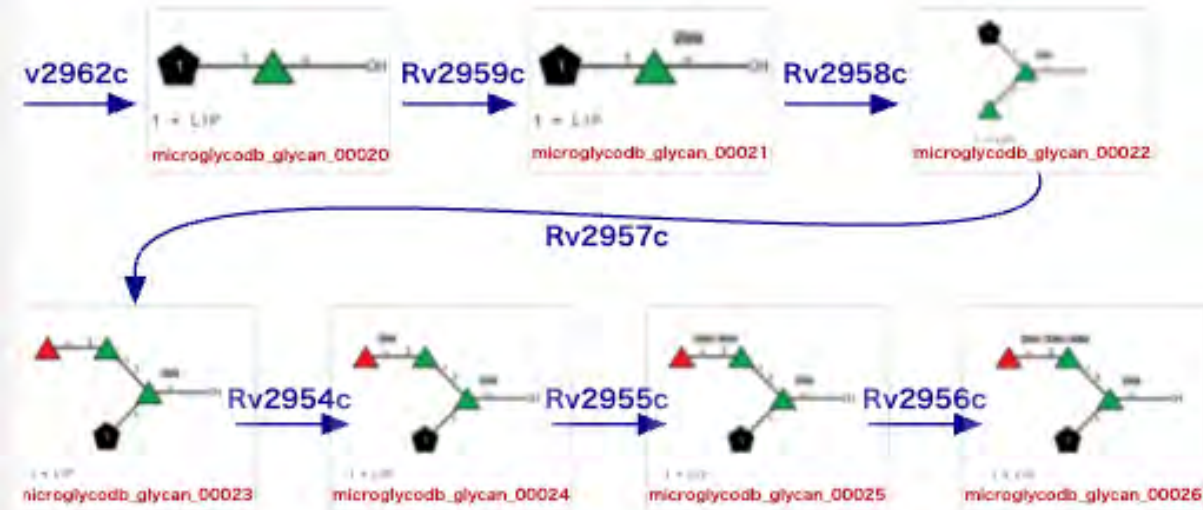
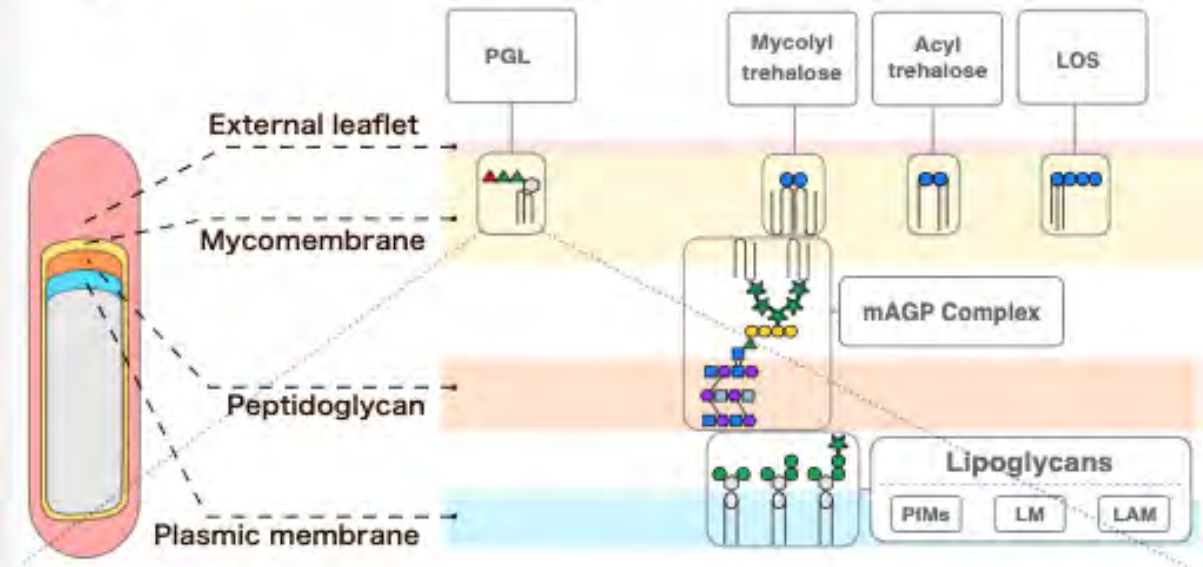
Genes Glycans

GPL

Gene ID	Gene No.	Gene Name	Protein Name
microglycodb_gene_00291	MAB_4104	<i>gtf2</i>	Putative glycosyltransferase GtfB
microglycodb_gene_00292	MAB_4105c	<i>rmt3</i>	Methyltransferase MtfD
microglycodb_gene_00293	MAB_4108c	<i>rmt4</i>	Methyltransferase MtfB
microglycodb_gene_00294	MAB_4107c	<i>gtf1</i>	Glycosyltransferase GtfA
microglycodb_gene_00295	MAB_4106c	<i>atf1</i>	Acetyltransferase
microglycodb_gene_00296	MAB_4110c	<i>atf2</i>	Probable acetyltransferase AtfA
microglycodb_gene_00297	MAB_4112c	<i>gtf3</i>	Putative glycosyltransferase GtfA



Mycobacterium tuberculosis



Genes

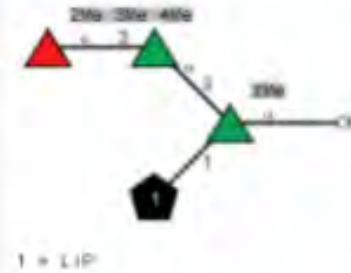
Glycans

All data (26)	PGL (7)	Mycolyltrehalose (0)	Acyltrehalose (0)	LOS (0)	mAGP (0)	PIMs,LM,LAM (19)
---------------	---------	----------------------	-------------------	---------	----------	------------------

Glycan ID

microglycodb_glycan_00026

Structure



CSDB Linear

LIP(1-1)[Me(1-2)[Me(1-4),Me(1-3)]aLFucp(1-3)aLRhap(1-3),Me(1-2)]aLRhap

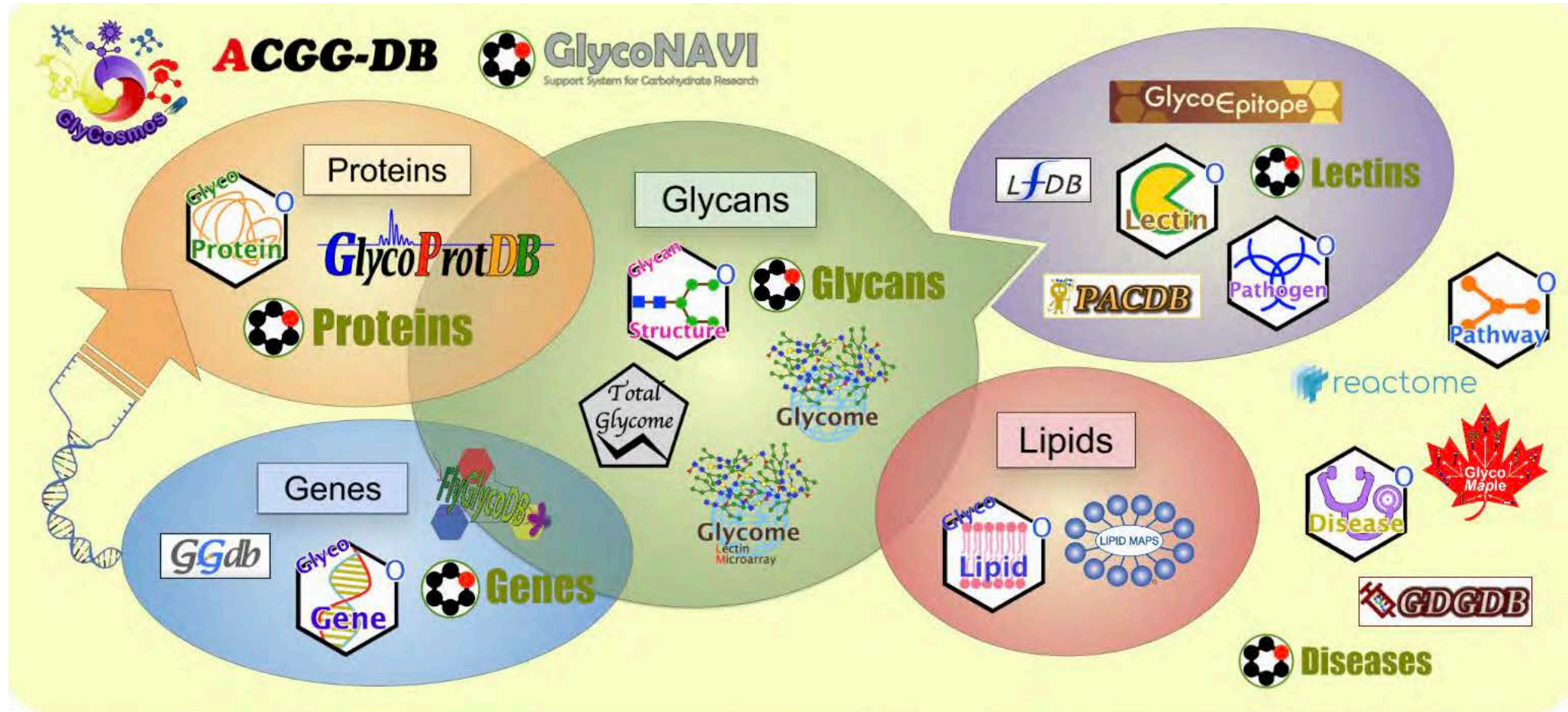
Modification/variation

Add α 1-3 L-Fuc

▼ Please select a file type from below ▼

Download!

Data in GlyCosmos annotated with taxonomy



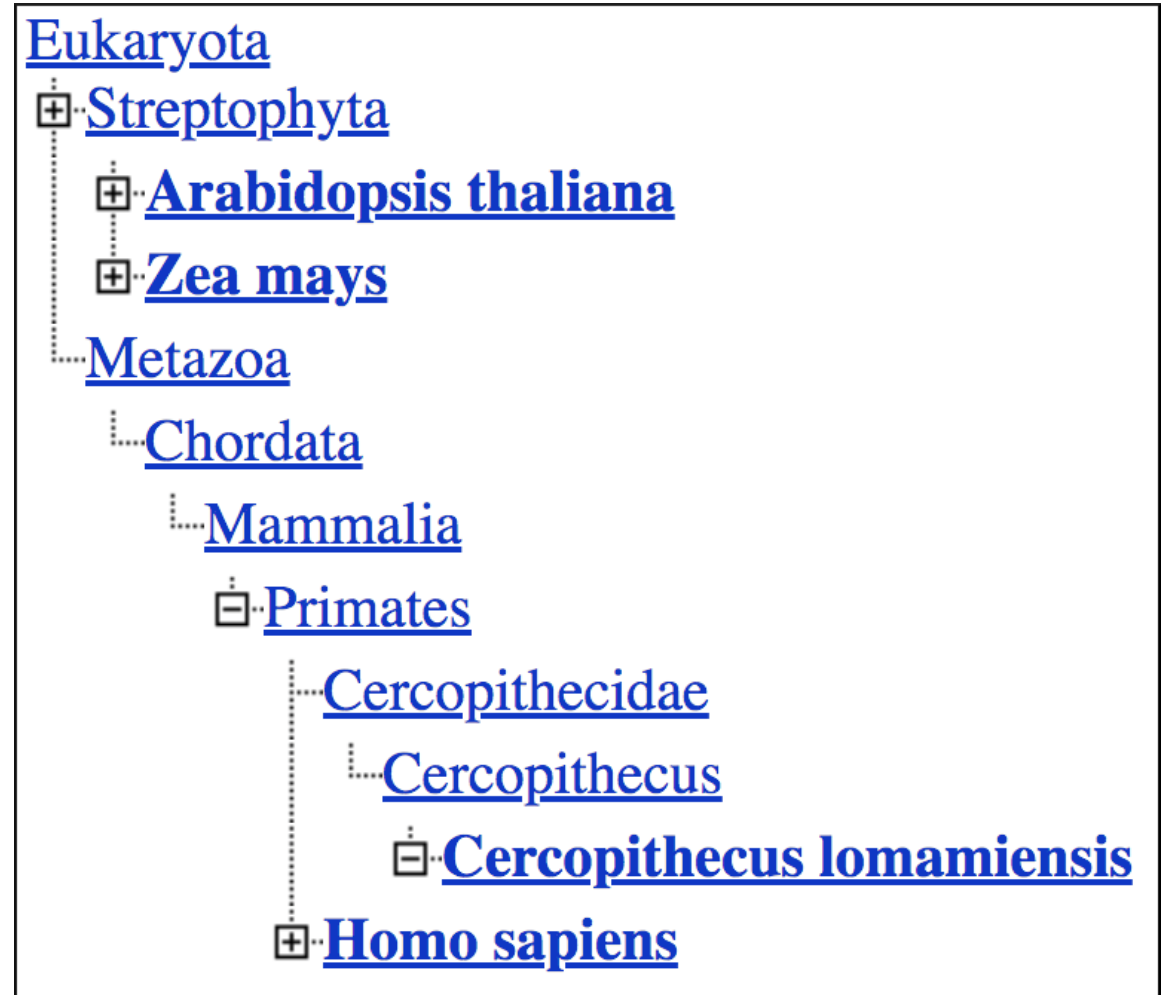
Organism	Superkingdom	Rank	Glycogenes Data	Glycans Data	Glycoproteins Data	Lectins Data	Pathways Data
Influenza A virus (A/New York/35/2014(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Nevada/14/2012(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/swine/Tianjin/01/2004(H1N1))	Viruses	no rank	0	0	1	0	0
Human immunodeficiency virus type 1 group M subtype C (isolate ETH2220)	Viruses	no rank	0	0	1	0	0
Kryptolebias marmoratus	Eukaryota	Species	18	0	11	0	0
Influenza A virus (A/Pennsylvania/03/2009(H1N1))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/New York/560/1997(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/duck/Jiangxi/27810/2013(mixed))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Texas/JMM_25/2012(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Boston/YGA_01027/2012(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Alaska/29/2014(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Singapore/H2009.471C/2009(H3N2))	Viruses	no rank	0	0	1	0	0
Pempheris schomburgkii	Eukaryota	Species	0	0	1	0	0
Influenza A virus (A/swine/Mexico/Qro35/2010(H1N1))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Minnesota/26/2015(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/TayNguyen/TN46/2004(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Colorado/UR06-0206/2007(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/HaNoi/ISBM16/2005(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Boston/97/2009(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/New York/313/1998(H3N2))	Viruses	no rank	0	0	1	0	0
Influenza A virus (A/Singapore/H2010 619/2010(H3N2))	Viruses	no rank	0	0	1	0	0

Aspergillus Genus

Organism ▲	Rank	Glycogenes Data	Glycans Data	Glycoproteins Data	Lectins Data	Pathways Data
Aspergillus	Genus	0	10	0	0	0
Aspergillus aculeatus	Species	0	0	12	1	0
Aspergillus awamori	Species	0	14	13	0	0
Aspergillus burnettii	Species	0	0	3	0	0
Aspergillus calidoustus	Species	0	0	1	0	0
Aspergillus desertorum	Species	0	0	1	0	0
Aspergillus ficuum	Species	0	0	4	0	0
Aspergillus flavipes	Species	0	0	2	0	0
Aspergillus flavus	Species	0	0	2	0	0
Aspergillus japonicus	Species	0	0	3	0	0
Aspergillus kawachii	Species	0	0	2	0	0
Aspergillus niger	Species	0	29	61	3	0
Aspergillus oryzae	Species	0	16	5	1	0
Aspergillus parasiticus	Species	0	0	1	0	0
Aspergillus phoenicis	Species	0	4	4	0	0
Aspergillus rugulosus	Species	0	0	7	0	0
Aspergillus sojae	Species	0	0	2	0	0
Aspergillus sp. (strain MF297-2)	Species	0	0	6	0	0
Aspergillus terreus	Species	0	0	6	0	0
Aspergillus tubingensis	Species	0	0	7	0	0
Aspergillus usamii	Species	0	0	2	0	0
Aspergillus ustus	Species	0	0	2	0	0

Setting up inferencing for taxonomies

1. Create an inference graph containing full taxonomic tree from NCBI
2. Create a rule set to refer to the inference graph
3. Reference the inference graph during queries:
 - e.g. Search for all glycans having taxonomy ID = 5052 (Aspergillus)









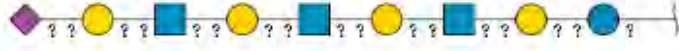
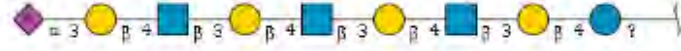


Results of searching for glycans in Aspergillus

http://rdf.glycoinfo.org/glycan/G68668TB	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G59365XO	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G68593OS	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G82314XU	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G92694DA	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G92890KE	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G93578EF	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G97859JV	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G99019LG	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G40198OD	"Aspergillus"
http://rdf.glycoinfo.org/glycan/G50119EM	"Aspergillus phoenicis"
http://rdf.glycoinfo.org/glycan/G70651XZ	"Aspergillus phoenicis"
http://rdf.glycoinfo.org/glycan/G03652TR	"Aspergillus phoenicis"
http://rdf.glycoinfo.org/glycan/G01260RY	"Aspergillus phoenicis"
http://rdf.glycoinfo.org/glycan/G09724ZC	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G70323CJ	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G81315DD	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G83582BK	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G27838JR	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G33711PF	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G67221GY	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G69382ZH	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G19747OA	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G37216UM	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G40915GX	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G42227JK	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G63024BG	"Aspergillus awamori"
http://rdf.glycoinfo.org/glycan/G79920MY	"Aspergillus awamori"

- A query to search just for Aspergillus genus glycans was made
- Successfully retrieved all glycans under the Aspergillus genus including all species!
- No data needed to be added for each of the species within this genus
- Query only needed one additional reference to the inference graph

Future plans : Subsumption, archetype and substructure search using inference

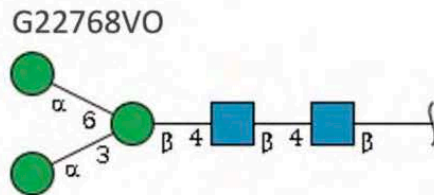
Mass	Base composition	Subsumed compositions		
1728.6082159	 <p>G68190MT</p> <p>5x  ?</p> <p>3x  ?</p> <p> ?</p>	<p>Monosaccharide composition</p> <p>4x  ?</p> <p> ?</p> <p>3x  ?</p> <p> ?</p> <p>G94020CX</p>	<p>Glycosidic topology</p> <p></p> <p>G47220PE</p>	<p>Linkage defined structure</p> <p></p> <p>G30764XF</p>

“archetype” concept

N-glycan chitobiose core structures:

- blue: IUPAC sequence
- green: WURCS sequence
- red: GlycoCT of GlcNAc at the reducing end

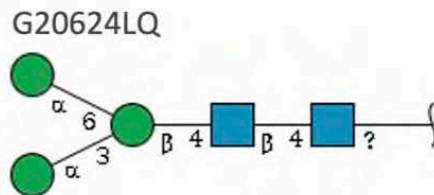
- Chemically modified glycans that could be considered biologically the same.
- The archetype concept will introduce the idea of an “all-encompassing” glycan to represent all of these.
- Once implemented, only the archetype glycan needs to be searched, and links to the other modified glycans made available at a lower level.



Man(a1-3)[Man(a1-6)]Man(b1-4)GlcNAc(b1-4)GlcNAc(b1-

WURCS=2.0/3,5,4/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-1-2-3-3/a4-b1_b4-c1_c3-d1_c6-e1

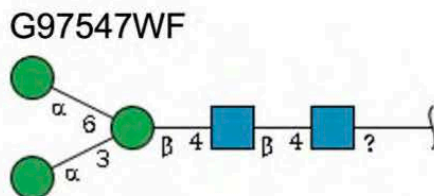
1b:b-dglc-HEX-1:5



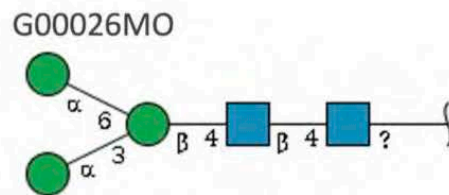
Man(a1-3)[Man(a1-6)]Man(b1-4)GlcNAc(b1-4)GlcNAc(?1-

WURCS=2.0/4,5,4/[a2122h-1x_1-5_2*NCC/3=O][a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-2-3-4-4/a4-b1_b4-c1_c3-d1_c6-e1

1b:x-dglc-HEX-1:5



WURCS=2.0/4,5,4/[a2122h-1x_1-?_2*NCC/3=O][a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-2-3-4-4/a4-b1_b4-c1_c3-d1_c6-e1



Man(a1-3)[Man(a1-6)]Man(b1-4)GlcNAc(b1-4)GlcNAc

WURCS=2.0/4,5,4/[u2122h_2*NCC/3=O][a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-2-3-4-4/a4-b1_b4-c1_c3-d1_c6-e1

archetype

Summary and Ongoing work

- GlyCosmos has been renewed for another five years, during which new **repositories** for glycan pathways, microbial glycans and glycozymes, and lectin microarray data will be developed.
- These repositories will then be processed to supplement the Data Resources in GlyCosmos.
- Currently developing a glycosylation simulation tool GlycoSim and GSS as well as a database of models that have been shown to successfully simulate known data so that the estimated parameters could be used for simulations of modified versions of these models.

Acknowledgements

Soka University

Masaaki Shiota
Tamiko Ono
Sunmyoung Lee
Yotsuba Hattori
Achille Zappa
Winnie Hsiung
Haruko Kitakaze

AIST

Atsushi Kuno
Noriaki Fujita
Kiyohiko Angata
Chiaki Nagai-Okatani
Daisuke Shinmachi



Niigata University

Shujiro Okuda
Yushi Takahashi

The Noguchi Institute

Issaku Yamada
Shinichiro Tsuchiya
Masaaki Matsubara
Masaaki Shiota
Tamiko Ono
Haruko Kitakaze

Funding agencies:

JST Database Coordination and
Integration Program, JST MIRAI
Program and NIH

Name	Affiliation
Kazuhiro Aoki	Complex Carb. Res. Ctr., Univ. of Georgia
Christine M. Szymanski	Complex Carb. Res. Ctr., Univ. of Georgia
Yann Guerardel	CNRS UGSF, Univ. Lille
Louis-David Leclercq	CNRS UGSF, Univ. Lille
Tamara L. Doering	Dept. of Mol. Microbio., Washington Univ.
Thomas Hurtaux	Dept. of Mol. Microbio., Washington Univ.
Kiyotaka Fujita	Grad. Sch. of Agricultural Sci., Kagoshima Univ.
Takane Katayama	Grad. Sch. of Biostud., Kyoto Univ.
Toshihiko Kato	Grad. Sch. of Biostud., Kyoto Univ.