

GlyGen



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC



PI: Raja Mazumder

The George Washington University

PI: William York

CCRC, University of Georgia



NIH Common Fund Glycoscience Program
1U01GM125267-01 (York & Mazumder)



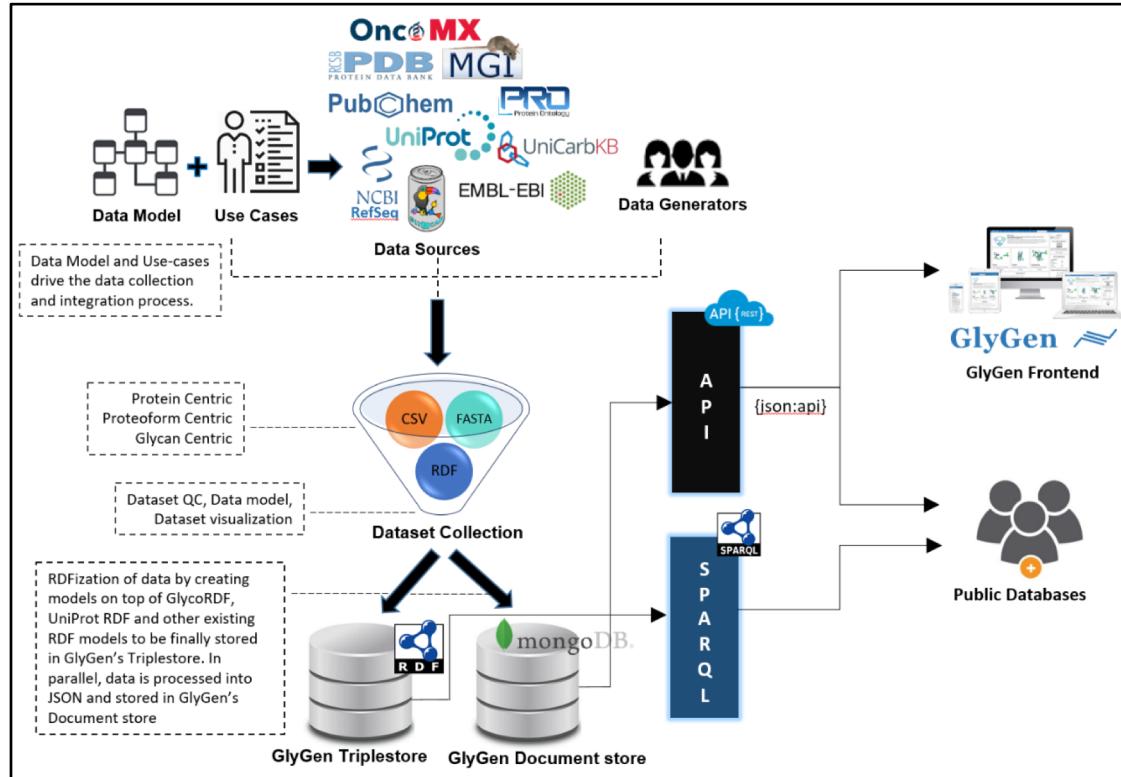
Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

GlyGen is an international project funded by The National Institutes of Health

G
O
A
L
S

- ❖ Integrating up-to-date glycobiology-related information from diverse disciplines
 - ❖ Creating an intuitive web portal to browse and search for knowledge in glycobiology
 - ❖ Developing essential new information resources, including:
 - ❖ An open, comprehensive Glycan Microarray Database
 - ❖ A Glycan Naming Ontology (GNOME) that facilitates interpretation of incomplete structural information in the context of biological function
 - ❖ An open, standardized environment for independent development
-
- E ❖ GlyGen is a cooperative, community-driven project, and an international project funded by The National Institutes of Health.
 - F ❖ An open, standardized environment for independent development and integration of additional research tools by other investigators.
 - R ❖ More than 15 investigators in four countries play key roles in the project.
 - T ❖ Two years of organized discussion and planning involving nearly 100 investigators.



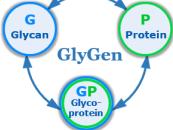
- Data model and use-cases drive GlyGen's data collection and integration.
- The data from various sources are harmonized and QC'd after collection and processed into open format datasets.
- The datasets are further processed into RDF and JSON format to be stored in **GlyGen Triple Store** and **Document Store** that can be accessed by **SPARQL endpoint** and **WS APIs**.
- GlyGen frontend retrieves data through the WS APIs whereas the Public databases and users can retrieve the data through both the and the SPARQL endpoint.



GlyGen  [HOME](#) [EXPLORE](#) [QUICK SEARCH](#) [TRY ME](#) [DATA](#) [SPARQL](#) [HELP](#) [Beta Testing](#)  [MY GLYGEN](#)

GlyGen
Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



Glycan
Search for glycan structures based on their chemical and structural properties.
[EXPLORE](#)

Protein
Search for proteins based on their sequences, accessions, and annotations.
[EXPLORE](#)

Glycoprotein
Search for glycoproteins based on the protein or glycan structures involved.
[EXPLORE](#)

Quick Search

What are the enzymes involved in the biosynthesis of [glycan X](#) in human?
 Which proteins have been shown to bear [glycan X](#) and which site is this glycan attached to?
 What are the orthologues of [protein X](#) in different species?
 Which glycans might have been synthesized in mouse using [enzyme X](#)?
 What are the glycosyltransferases in [species X](#)?
 Which glycosyltransferases are known to be involved in [disease X](#)?
[SEE ALL](#)

Try Me

What are the enzymes involved in the biosynthesis of [Man5](#) in human?
 Which proteins have been shown to bear a [bi-antennary fully sialylated N-Glycan](#) and which site is this glycan attached to?
 Which glycans might have been synthesized in mouse using [Mgt1](#)?

License | Privacy Policy | Disclaimer | Contact Us

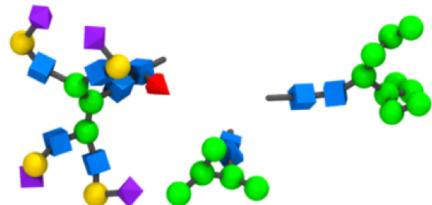
Funded by NIH Glycoscience Common Fund
Grant # 1U01GM125267 - 01



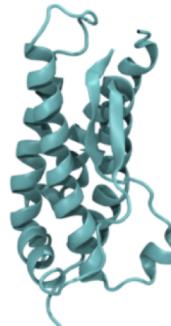
- Glycan, Protein, Glycoprotein searches by choosing intuitive simple or advanced search.
- Quick Search to answer user-specific question
- GlyGen data can be accessed from the Data page, API and SPARQL endpoint
- Helpful tutorials, data resources, tools, learning materials, social media, contact us, and feedback.
- Use of open public license for data and source code

<https://glygen.org>

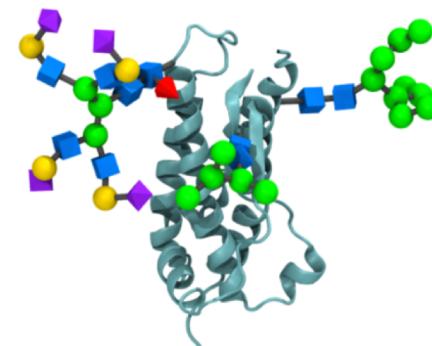
GlyGen.org

**Glycan**

Search for glycan structures based on their chemical and structural properties.

[EXPLORE](#)**Protein**

Search for proteins based on their sequence, accessions, and annotations.

[EXPLORE](#)**Glycoprotein**

Search for glycoproteins based on the protein or glycan structures involved.

[EXPLORE](#)

Glycan search - Data types and information that is about specific glycans.

Protein search - Data types and information that is about some particular protein coding gene, or that can be mapped to the canonical protein sequence representing that gene.

Glycoprotein search - Data types and information of proteins that are glycosylated at specific sites with or without glycan information.

Current Species Data in GlyGen – **Human and Mouse**



Summary of your Protein Search

Performed on: April 30th 2019, 5:40:08 pm (EST)

What are the enzymes involved in the biosynthesis of G17689DH in human?

[Update Results](#)[Modify Search](#)

** To perform the same search again using the current version of the database, click "Update Results".

Page [Prev](#) [1](#) [Next](#) Records per page [20](#) "12 Proteins were found"

UniProtKB Accession	Gene Name	UniProtKB Name	Organism	RefSeq Name	RefSeq Accession
Q96F25-1	ALG14	UDP-N-acetylglucosamine transferase subunit ALG14 homolog	Homo sapiens	UDP-N-acetylglucosamine transferase subunit ALG14 homolog isoform 2	NP_001292171.1
Q9Y274-1	ST3GAL6	Type 2 lactosamine alpha-2,3-sialyltransferase	Homo sapiens	type 2 lactosamine alpha-2,3-sialyltransferase isoform 3	NP_001258074.1
Q11206-1	ST3GAL4	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4	Homo sapiens	CMP-N-acetylneuraminate-beta-galactosamide- alpha-2,3-sialyltransferase 4 isoform 2	NP_001241686.1
Q11203-1	ST3GAL3	CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase	Homo sapiens	CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase isoform k	NP_001257388.1
Q9BT22-1	ALG1	Chitobiosyldiphosphodolichol beta-mannosyltransferase	Homo sapiens	chitobiosyldiphosphodolichol beta-mannosyltransferase isoform 2	NP_001317433.1
O60512-1	B4GALT3	Beta-1,4-galactosyltransferase 3	Homo sapiens	beta-1,4-galactosyltransferase 3	NP_001186802.1
O60909-1	B4GALT2	Beta-1,4-galactosyltransferase 2	Homo sapiens	beta-1,4-galactosyltransferase 2 isoform b	NP_001005417.1
P15291-1	B4GALT1	Beta-1,4-galactosyltransferase 1	Homo sapiens	beta-1,4-galactosyltransferase 1	NP_001488.2
Q10469-1	MGAT2	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	Homo sapiens	alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	NP_002399.1
Q9H553-1	ALG2	Alpha-1,3/1,6-mannosyltransferase ALG2	Homo sapiens	alpha-1,3/1,6-mannosyltransferase ALG2	NP_149078.1
P26572-1	MGAT1	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	Homo sapiens	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	NP_001108089.1
Q9BYC5-1	FUT8	Alpha-(1,6)-fucosyltransferase	Homo sapiens	alpha-(1,6)-fucosyltransferase isoform b	NP_004471.4

Page [Prev](#) [1](#) [Next](#) Records per page [20](#) "12 Proteins were found"

Which glycans might have been synthesized in mouse using enzyme X?

Glycan Detail Page for G17689DH

G17689DH - Bi-antennary N-Glycan Complex (reduced overview)



General

Species

Motif

Found Glycoproteins

Cross References

Biosynthetic Enzymes

Digital Sequence

General

GlyTouCan Accession: G17689DH
Monoisotopic Mass: 2,368.84 Da
Glycan Type/Subtype: N-Glycan complex

Motif

G00026MO-N-Glycan core basic G00030MO-N-Glycan complex

Species

Homo sapiens: GlycomeDB 1 UniCarbKB 3

Found Glycoproteins

Sources	Protein Name	UniProtKB Accession	Position
UniCarbKB 1	Hepatocyte growth factor	P14210-1	653
UniCarbKB 1	Hepatocyte growth factor	P14210-1	294
UniCarbKB 1	Hepatocyte growth factor	P14210-1	566

Digital Sequence

WURCS
2.0/6,12,11/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][Aad21122h-2a_2-6_5*NCC/3.....]
IUPAC
alpha-D-NeupAc-(2->3)-beta-D-Galp-(1->4)-beta-D-GlcNAc-(1->2)-alpha-D-Manp-(1->3)[alpha-D-NeupAc-(2->3)-beta-D-G.....]

General

Species

Function

Glycosylation

Sequence

Cross References

Pathway

Isoform

Orthologs

Disease

Mutation

Expression Tissue

Expression Disease

Publications

General

- Gene Name: HGF
- Gene Location: Chromosome: 7 (81,699,008 - 81,770,438)
- UniProtKB ID: HGF_HUMAN
- UniProtKB Accession: P14210-1
- UniProtKB Entry Name: Hepatocyte growth factor
- RefSeq Accession: NP_000592.3
- RefSeq Name: hepatocyte growth factor isoform 1 preproprotein
- Chemical Mass: 83,134 Da

Glycosylation

With Reported Glycans Without Reported Glycans

Sources	GlyTouCan Accession	Type	Residue	Image of Glycan Structure
UniCarbKB 1	G01543ZX	N-linked	Asn653	
PubMed 2				
UniCarbKB 1	G17689DH	N-linked	Asn653	
PubMed 2				

Protein Detail Page for HGF

Highlighting General and Glycosylation section information (reduced overview)

GlyGen.org

GlyGen

GlyGen Data page

GlyGen

HOME EXPLORE QUICK SEARCH TRY ME DATA ABOUT HELP Search... Beta Testing MY GLYGEN

Dataset Collection

Search term Search

You can search by BCO ID, name or contributor.

Total of 100 datasets (100 passed filter) Filters

Filter by categories
 Protein centric (75)
 Proteiform centric (8)
 Glycan centric (16)
 Other (2)

Filter by species
 Homo sapiens (55)
 Mus musculus (44)
 Other (2)

Filter by status
 reviewed (100)

Filter by file type
 csv (94)
 fasta (4)
 nt (2)

Protein centric GLYDS000001	Homo sapiens, CSV
Human Proteome Accessions	
uniprotkb_ca...	status...
Q9Y6C9-1	reviewed...
Q658T7-1	reviewed...
The Human Proteome dataset contains human [taxid:9606] UniProtKB canonical accessions mapped to the ... view details	

Protein centric GLYDS000002	Homo sapiens, FASTA
Human UniProtKB Canonical sequences	
The Human UniProtKB Canonical sequences dataset contains human [taxid:9606] protein fasta sequences ... view details	

Protein centric GLYDS000003	Homo sapiens, NT
Human EBI-UniProtKB NT file	
The Human EBI-UniProtKB NT file contains proteome data for human [taxid:9606] from UniProtKB and oth ... view details	

- Data page contains all the datasets collected by GlyGen.
 - The datasets can be easily searched and filtered.
 - FAQs helps to understand the data.
 - Each dataset is downloadable, has a sample view, JSON dataset BCO and a conventional readme with extensive metadata.

<https://data.glygen.org>

GlyGen.org

**University of Georgia**

Rene Ranzinger
Michael Pierce
Robert Woods
Rupali Mahadik
Tatiana Williamson
Sena Arpinar
Sanath Bhatt
Sujeet Kulkarni
Sandeep Nakarakommula

EMBL-EBI

Maria Martin
Leyla Jael Garcia Castro
Preethi Vasudev

NCBI

Kim Pruitt
Evan Bolton

The George Washington University

Raja Mazumder
Robel Kahsay
Jeet Vora
Rahi Navelkar
Reza Mousavi
Nagarajan
Pattabiraman
Xavier Holmes
Brian Fochtman

Georgetown University

Nathan Edwards
Radoslav Goldman
Darren Natale
Karen Ross
Wenjin Zhang

Harvard University
Richard Cummings**The Jackson Laboratory**
Judith Blake**Soka University**
Kiyoko Aoki-Kinoshita**The Griffith University**
Matthew Campbell**Imperial College London**
Ten Feizi**Macquarie University**
Nicki Packer**NIH-NCI**
Jefferey Gildersleeve

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC



Imperial College London



Web portal: <https://glygen.org>
Data store: <https://data.glygen.org>
WS API: <https://api.glygen.org>

Will York
Raja Mazumder

CONTACT

will@ccrc.uga.edu
mazumder@gwu.edu

GlyGen.org