



GlySpace: Toward a Collaborative Glycoinformatics Community

GlyGen

The University of Georgia

Will York
Rene Ranzinger
Rupali Mahadik
Tatiana Williamson
Gaurav Agarwal

The George Washington University

Raja Mazumder
Robel Kahsay
Jeet Vora
Rahi Navelkar
Reza Mousavi
Nagarajan Pattabiraman

Georgetown University

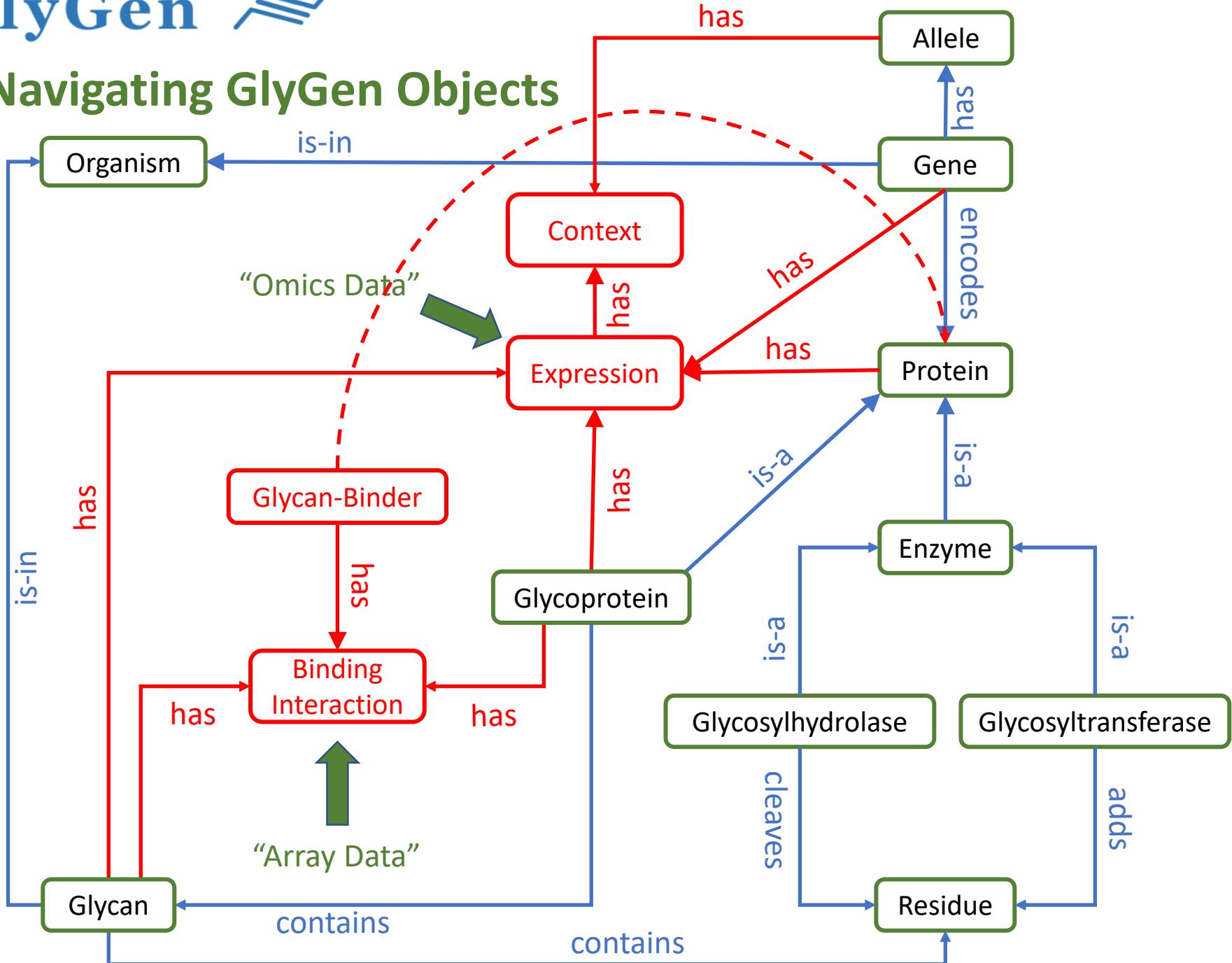
Radoslav Goldman
Darren Natale
Karen Ross



National Institutes of Health
Office of Strategic Coordination - The Common Fund



Navigating GlyGen Objects

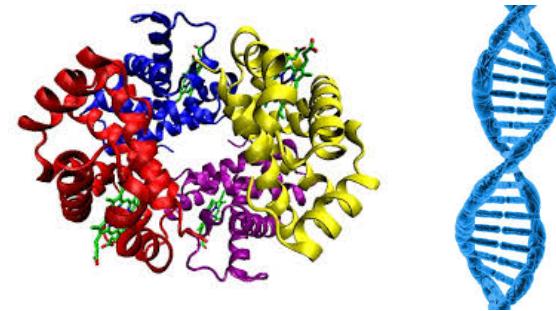


Data in GlyGen

Data is categorized into 3 data categories

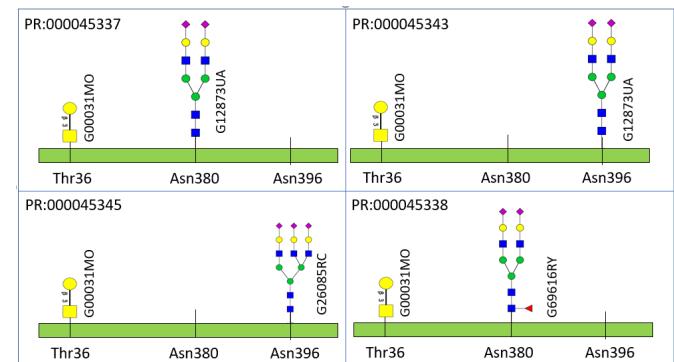
Protein Centric

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. Examples include pathways, Gene Ontology, localization, etc.



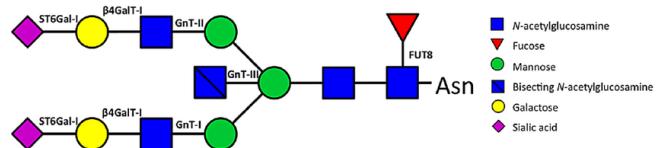
Glycoprotein (proteoform) Centric

Data types and info)rmation of all the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spliced RNA transcripts and post-translational modifications



Glycan Centric

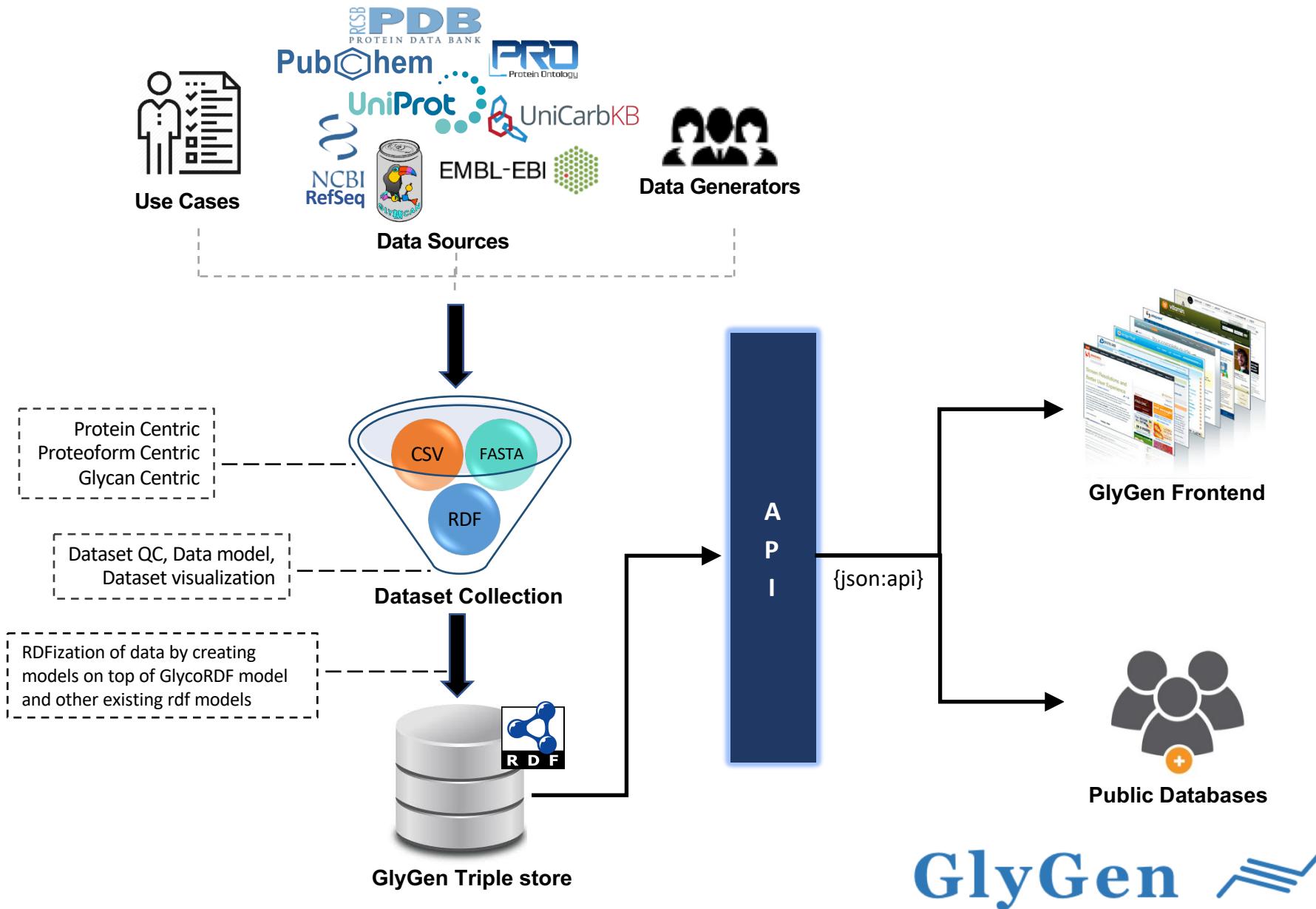
Glycan-centric data include data types and information that is about specific glycans.



Currently focus on two species: **Human** and **Mouse**



GlyGen Data integration workflow





(Maria Martin)



18 Datasets - Idmapping* cross-references* glycosyltransferase* glycohydrolase*,
citations* functions* pathway* sequences * structure* (* Human, Mouse)

21538 Canonical Proteins (Human)

72299 Isoforms (Human)

1202 Glycoproteins (Experimental)
3317 Glycoproteins (Predicted)
4519 Glycosylated Proteins
17033 Glycosylation sites

25490 Canonical Proteins (Mouse)

35263 Isoforms (Mouse)

500 Glycoproteins (Experimental)
3200 Glycoproteins (Predicted)
3800 Glycosylated Proteins
14172 Glycosylation sites



(Matthew Campbell)



SPARQL endpoint
<http://137.92.56.159:40935/>



Glycosylation Sites (UniCarbKB)

110 Human Proteins
(323 sites)
736 GlyTouCan AC
200 proteins with sites
~650 next update



(Kiyoko Aoki-Kinoshita
Nathan Edwards)



GlyTouCan Repository



(Nathan Edwards)

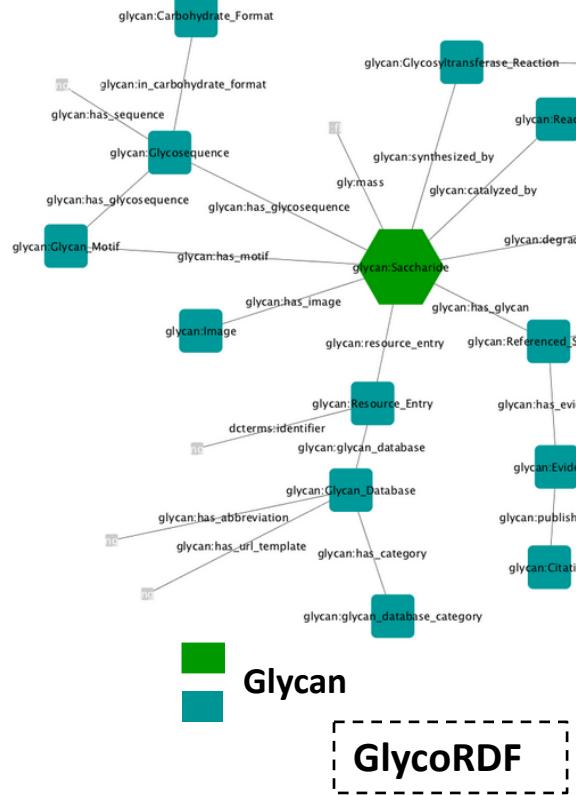


Glycan Accessions, sequences
images

3554 GlyTouCan Ac
(Human)

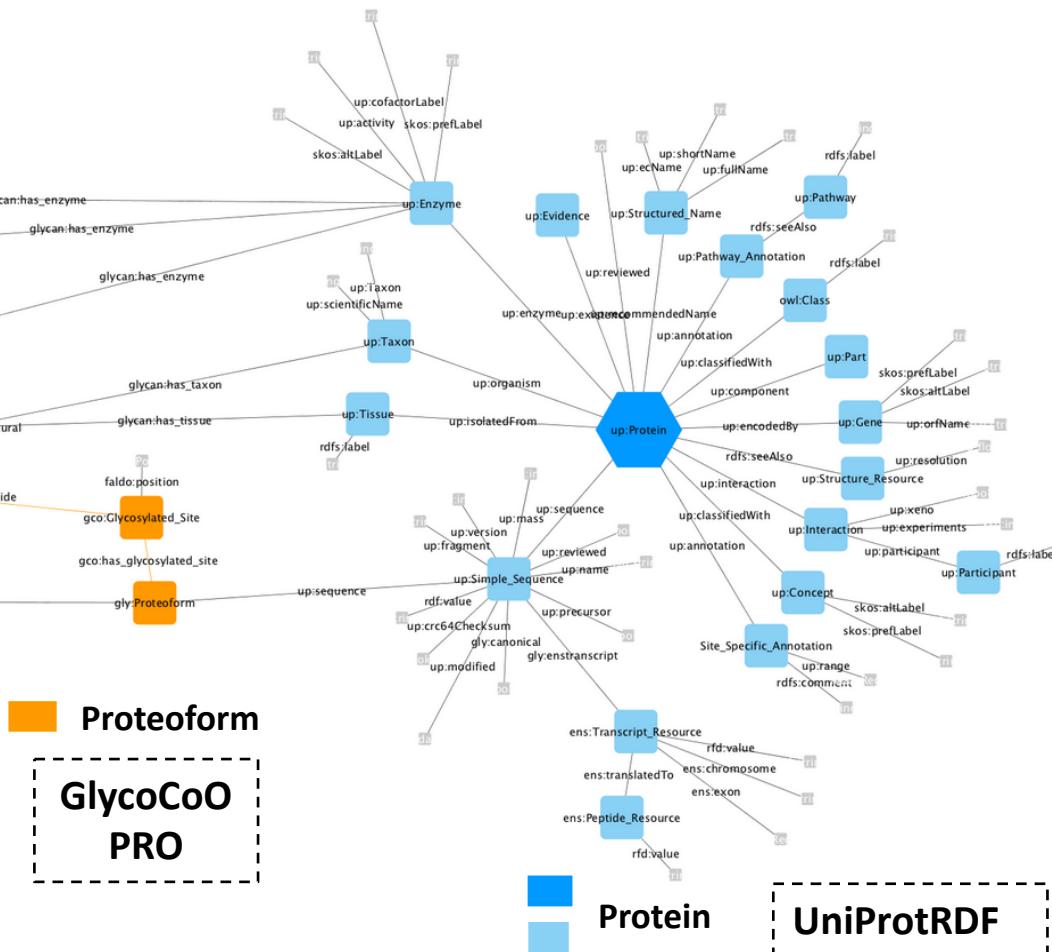
774 GlyTouCan Ac
(Mouse)

GlyGen Data model (reduced view)



Namespaces

`up:` <http://www.uniprot.org/core/>
`rdf:` <http://www.w3.org/1999/02/22-rdf-syntax-ns>
`rdfs:` <http://www.w3.org/2000/01/rdf-schema>
`skos:` <http://www.w3.org/2004/02/skos/core>



`faldo:` <http://biohackathon.org/resource/faldo>
`ens:` <http://rdf.ebi.ac.uk/resource/ensembl>
`gco:` <http://purl.jp/bio/12/glyco/conjugate>
`gly:` <http://glygen-vm-prd.biochemistry.gwu.edu/ontology/>



Sophistication Enables Simplicity

The User Interface Becomes Simpler as Technology Advances

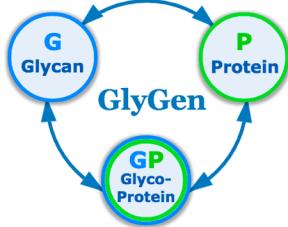


A mass spectrometer in use at NIH in 1975



ThermoFisher ProLab Benchtop Mass Spectrometer (2018)

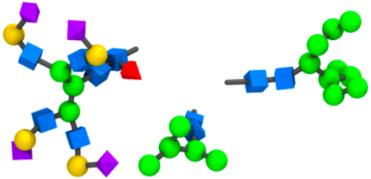
- **UI/UX** analysis and preliminary design
- Query generation and formatting
 - Type-ahead support for text fields
 - Name spaces for textual input
 - Widgets to facilitate query generation
- Search interfaces
 - Glycan Search Page
 - Protein Search Page
 - Glycoprotein Search Page
- Predefined “Quick Searches” for high-priority use cases
- “Try Me” demonstration queries that require **no** user input
- Display infrastructure for search results
- Display interfaces
 - Glycan List Page
 - Protein List Page
- **My GlyGen** - user privacy and query history



GlyGen

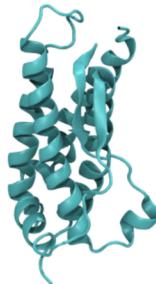
Computational and Informatics Resources for Glycoscience Research

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 webportal allows exploring this data and to perform unique searches that cannot be executed in any of the integrated databases alone.



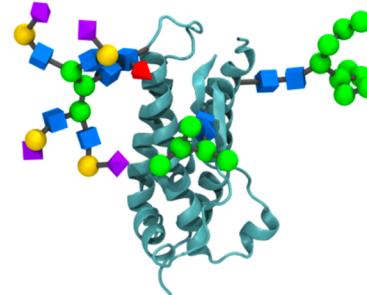
Glycan

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

[EXPLORE](#)

Protein

Search for proteins structures based on their sequence and their annotations.

[EXPLORE](#)

Glycoprotein

Search for glycoconjugates based on the proteins or glycan structures involved.

[EXPLORE](#)

Your Opinion Matters



Please provide feedback and suggestion to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

GlyGen Statistics

GlyGen's Database Statistics.

Human

Proteins	21538
Glycoproteins	4679
Glycans	5239

Mouse

Proteins	25490
Glycoproteins	3816
Glycans	2406

Explore Other Resources





Use Cases for GlyGen Version 1

1. What are the enzymes involved in the biosynthesis of **glycan X** in human/mouse?
2. Which glycans might have been synthesized in human/mouse using **enzyme X**?
3. Which proteins have been shown to bear **glycan X** and which site is this glycan attached to?
4. What are the orthologs of **glycoprotein X** in different species?
5. What are the functions of **protein X**?
6. What are the gene locations of the enzymes involved in the biosynthesis of **glycan X** in human/mouse?
7. What are the **glycosyltransferases** in human/mouse?
8. What are the **glycohydrolases** in human/mouse?
9. What are the **reported or predicted** glycosylated proteins in human/mouse?
10. Which glycosyltransferases are known to be involved in **disease X**?

Quick Search – Predefined Queries

GlyGen  [HOME](#) [EXPLORE](#) [QUICK SEARCH](#) [TRY ME](#) [ABOUT](#) [MORE](#) [MY GLYGEN](#)

SEARCH BY GLYCAN

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 gene locations of the enzymes involved in the biosynthesis of glycan X in human? 

SEARCH BY PROTEIN

What are the orthologues of protein X in different species? 

What are the functions of protein X? 

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

SEARCH BY SPECIES

What are the glycosyltransferases in species X? 

What are the glycohydrolases in species X? 

What are the reported or predicted glycosylated proteins in species X? 

SEARCH BY DISEASE

Which glycosyltransferases are known to be involved in disease X? 

GlyGen 

HOME EXPLORE QUICK SEARCH TRY ME ABOUT MORE  MY GLYGEN

SEARCH BY GLYCAN

What are the enzymes involved in the biosynthesis of glycan X in human?

Glycan X: G00279YF Example: G00279YF

Which proteins have been shown to bear glycan X and which site is this glycan attached to? 

What are the gene locations of the enzymes involved in the biosynthesis of glycan X in human? 

SEARCH BY PROTEIN

What are the orthologues of protein X in different species? 

What are the functions of protein X? 

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

SEARCH BY SPECIES

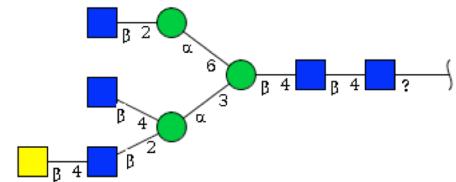
What are the glycosyltransferases in species X? 

What are the glycohydrolases in species X? 

What are the reported or predicted glycosylated proteins in species X? 

SEARCH BY DISEASE

Which glycosyltransferases are known to be involved in disease X? 



Quick Search – Predefined Queries



Page

Prev

1

Next

Records per page

25

Sort By

UniProt Name

ASC

You Found 9 results of protein

UniProt Accession	Gene Name	UniProt Name	Native Mass (Da)	Organism	RefSeq Name	RefSeq Accession
Q9H553-1	ALG2	16-mannosyltransferase ALG2	47092	Homo sapiens	alpha-1,3/1,6-mannosyltransferase ALG2	NP_149078.1
P26572-1	MGAT1	Alpha-13-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	50878	Homo sapiens	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase isoform X1	XP_016864976.1
Q9UM21-1	MGAT4A	Alpha-13-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	61544	Homo sapiens	alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A isoform 1	NP_036346.1
Q9UQ53-1	MGAT4B	Alpha-13-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B	63198	Homo sapiens	alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B isoform 2	NP_463459.1
Q9UBM8-1	MGAT4C	Alpha-13-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C	56061	Homo sapiens	alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C isoform X4	XP_016874639.1
Q10469-1	MGAT2	Alpha-16-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	51550	Homo sapiens	alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	NP_002399.1
Q9NP73-1	ALG13	Putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase ALG13	126056	Homo sapiens	putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase ALG13 isoform 2	NP_060936.1
Q6GMV1-1	ALG1L	Putative glycosyltransferase ALG1-like	21132	Homo sapiens	putative glycosyltransferase ALG1-like isoform 1	NP_001182152.1
Q96F25-1	ALG14	UDP-N-acetylglucosamine transferase subunit ALG14 homolog	24151	Homo sapiens	UDP-N-acetylglucosamine transferase subunit ALG14 homolog isoform 1	NP_659425.1

Page

Prev

1

Next

Records per page

25

Sort By

UniProt Name

ASC

You Found 9 results of protein



For Data



Attribution 4.0 International (CC BY 4.0)

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.
- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

For Software/Source code



The GNU General Public License v3.0

Free as in Freedom

- Freedom to use the software for any purpose,
- Freedom to change the software to suit your needs,
- Freedom to share the software with your friends and neighbors, and
- Freedom to share the changes you make.

Our Data-Sharing Plan Facilitates Collaboration!



Acknowledgements



University of Georgia

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

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

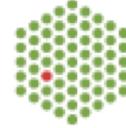
The Griffith University Matthew Campbell

Imperial College London Ten Feizi

Macquarie University Nicki Packer

NIH-NCI Jefferey Gildersleeve

NIH Grant - U01 GM125267-01



(Maria Martin)



18 Datasets - Idmapping* cross-references* glycosyltransferase* glycohydrolase*,
citations* functions* pathway* sequences * structure* (* Human, Mouse)

21538 Canonical Proteins (Human)

72299 Isoforms (Human)

1202 Glycoproteins (Experimental)
3317 Glycoproteins (Predicted)
4519 Glycosylated Proteins
17033 Glycosylation sites

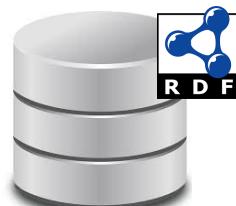
25490 Canonical Proteins (Mouse)

35263 Isoforms (Mouse)

500 Glycoproteins (Experimental)
3200 Glycoproteins (Predicted)
3800 Glycosylated Proteins
14172 Glycosylation sites



(Matthew Campbell)



SPARQL endpoint
<http://137.92.56.159:40935/>



Glycosylation Sites (UniCarbKB)

110 Human Proteins
(323 sites)
736 GlyTouCan AC
200 proteins with sites
~650 next update



GlyTouCan Repository

(Kiyoko Aoki-Kinoshita
Nathan Edwards)



(Nathan Edwards)



Glycan Accessions, sequences
images

3554 GlyTouCan Ac
(Human)

774 GlyTouCan Ac
(Mouse)

Navigating GlyGen Objects

