2023 GlycoMIP Annual Meeting

Knowledge Sharing

Sanket A. Deshmukh
Chair, Cyber Resources Committee

March 06, 2023

GlycoMIP is funded by the National Science Foundation
through cooperative agreement DMR-1933525.

➢ Data Sharing Management: Cyber Resources Committee

Faculty

Sanket Deshmukh (VT)
Committee Chair

Maren Roman (VT)
GlycoMIP Director

Rob Woods (UGA)
GlycoMIP Associate Director

Zhenyu James Kong
(VT)

Pengyu Hong
(Brandeis)

GlycoData Developers

Postdocs

Fangxi Wang
(VT) [0.5 FTE]

Lakshmi Kunche
(VT) [0.5 FTE]

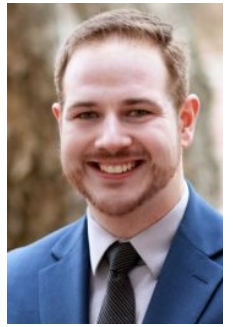
Software Engineer

Hunter Irving
(VT) [0.2 FTE]

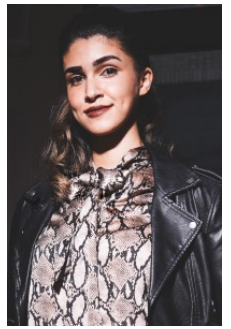
Scientific Staff

Lachele Foley
(UGA)

Graduate Student



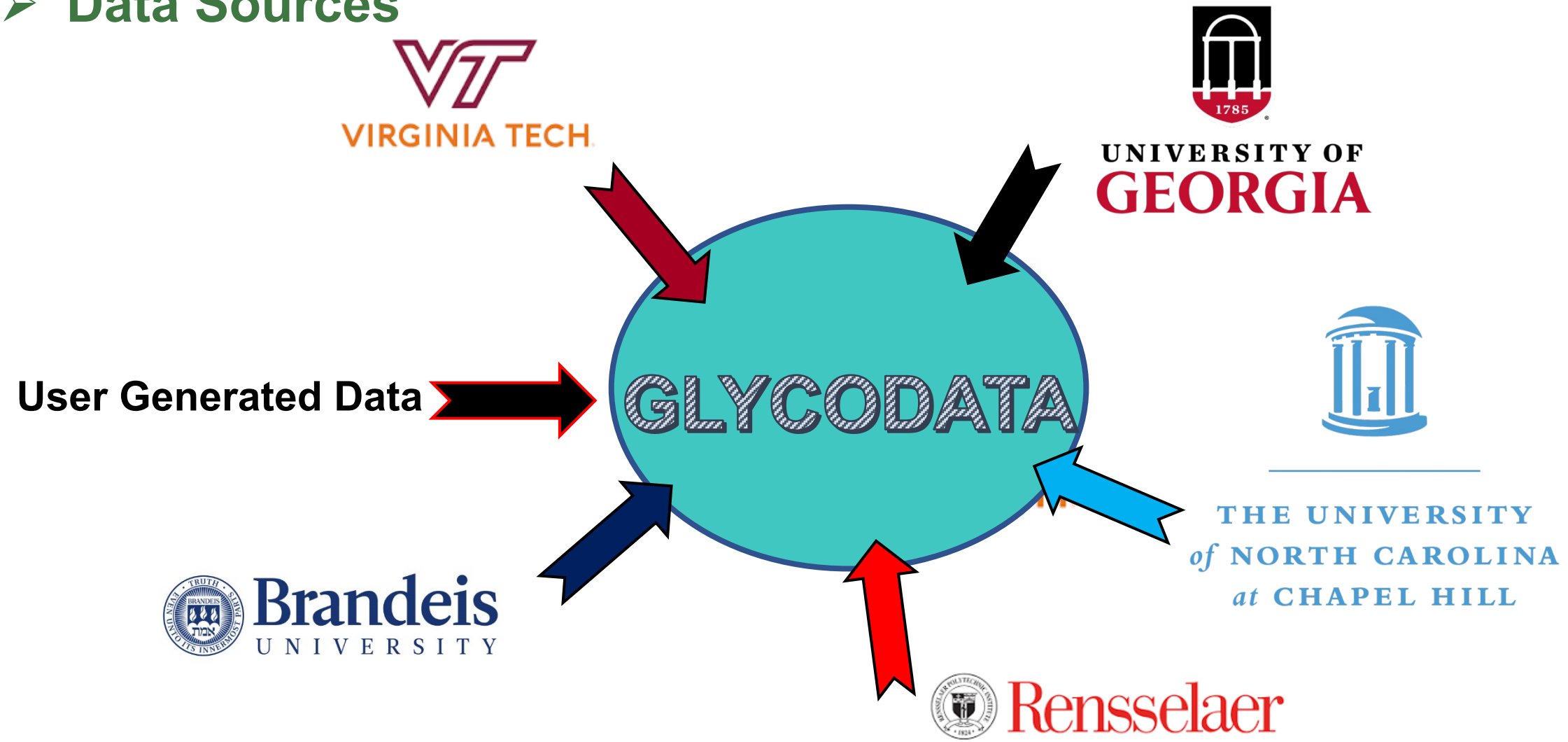
David Montgomery
(UGA)



Meade Maftouni
(VT)

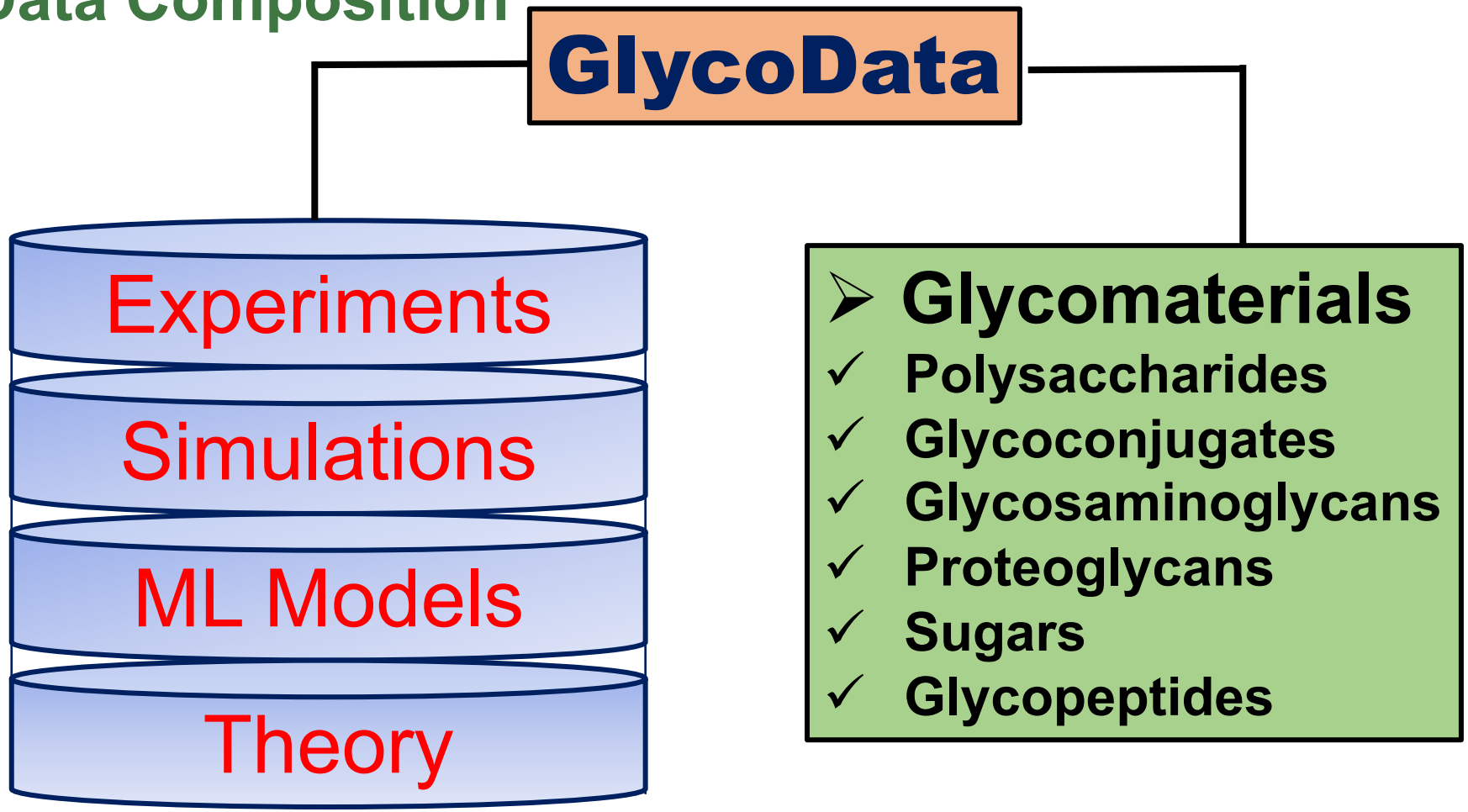
- The committee meets once every month to discuss progress.
- Chair and GlycoData Developers meet once a week to discuss work plan.

➢ Data Sources



- Data generated will contain structured and unstructured data obtained from experiments and simulations.
- ***The vision of GlycoData is to enable researchers within and outside the GlycoMIP community to make both structured and unstructured data available for a worldwide scientific community.***

➢ GlycoData Composition



❖ Based on the current ongoing research within GlycoMIP and the literature, we envision categories of materials based on different types of glycans.

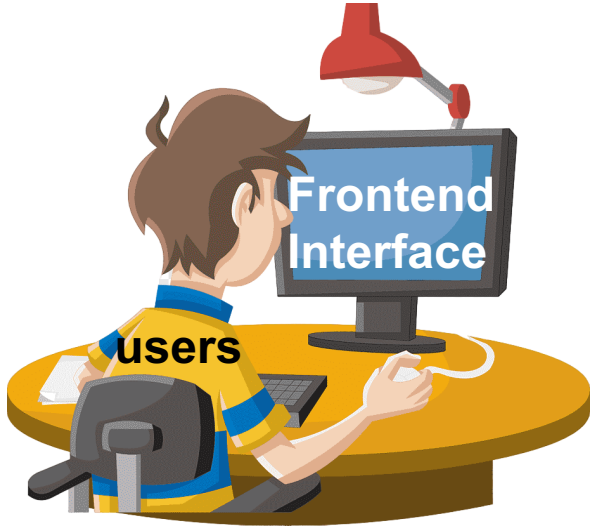
➢ GlycoData Architecture

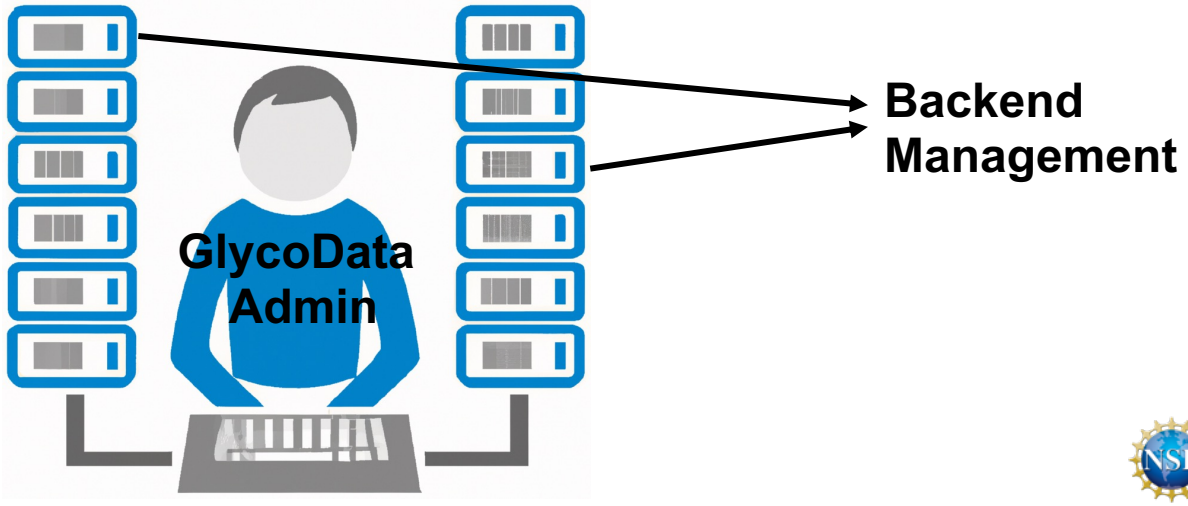
❖ The data management architecture of the platform consists of:

1. The **frontend web interface** (*user interface*) for data access, including data searching, uploading, and downloading.

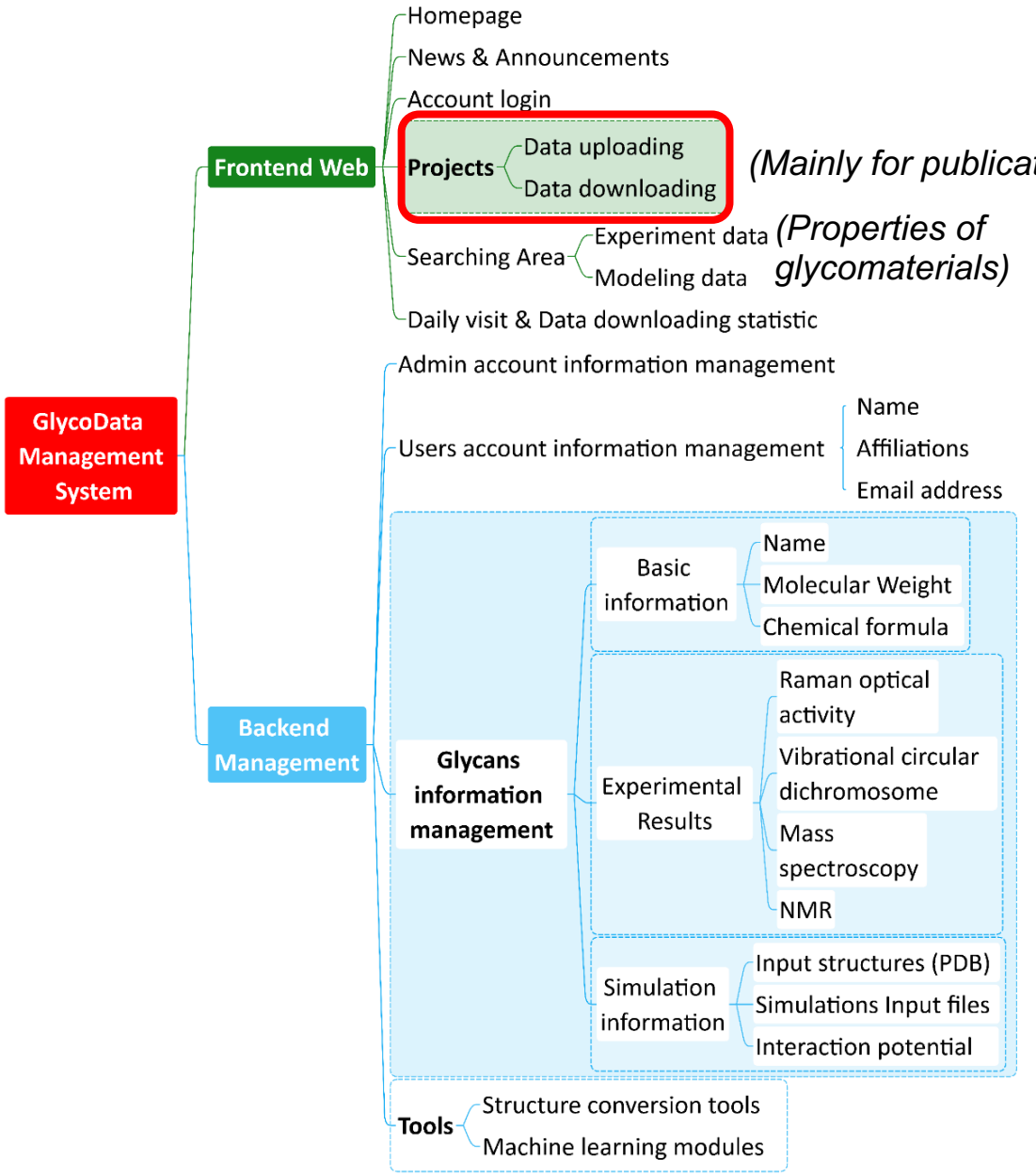
2. The **backend management** for data collection, storage, and management, including admin interface.

3. **Scientific tools and modules** for data conversion, analysis, etc.





GlycoData Configuration



Support system and frameworks

Operation System
Linux

Frontend & Backend Framework
Django

Database Management System
MySQL

Web Server
Nginx

Toolkits
Django REST framework
Bootstrap

Deploy in ARC Virtual Machine at VT

➢ Access To The Data

User

Parallel requests

NGINX

Communicate via Unix socket

gunicorn

MySQL® + django

Invokes a callable object

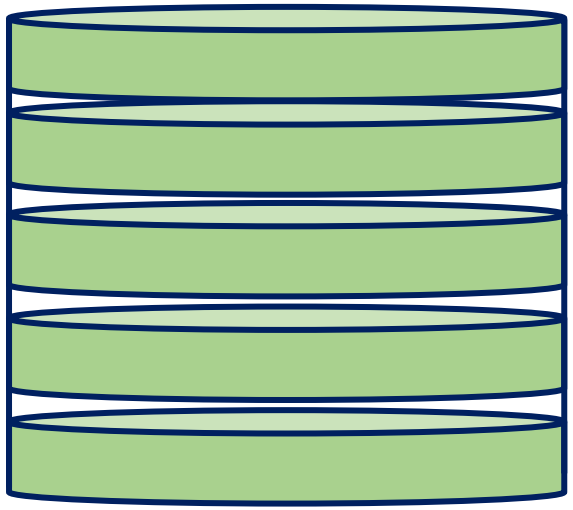
- ❖ Nginx listens to requests from the browsers through port 80/443 and transfers it to gunicorn socket.
- ❖ Gunicorn acts as an interface between django and NGINX to relay the information about requested search.

➢ Benefits to The GlycoData Users

Registration only after Admin Approval and login with 2-factor authentication
(A secured data sharing platform)

Database

User



- ❖ Images and raw experimental data in text format
- ❖ Simulation scripts and data
- ❖ ML models and tools
- ❖ Scientific tools for data conversion

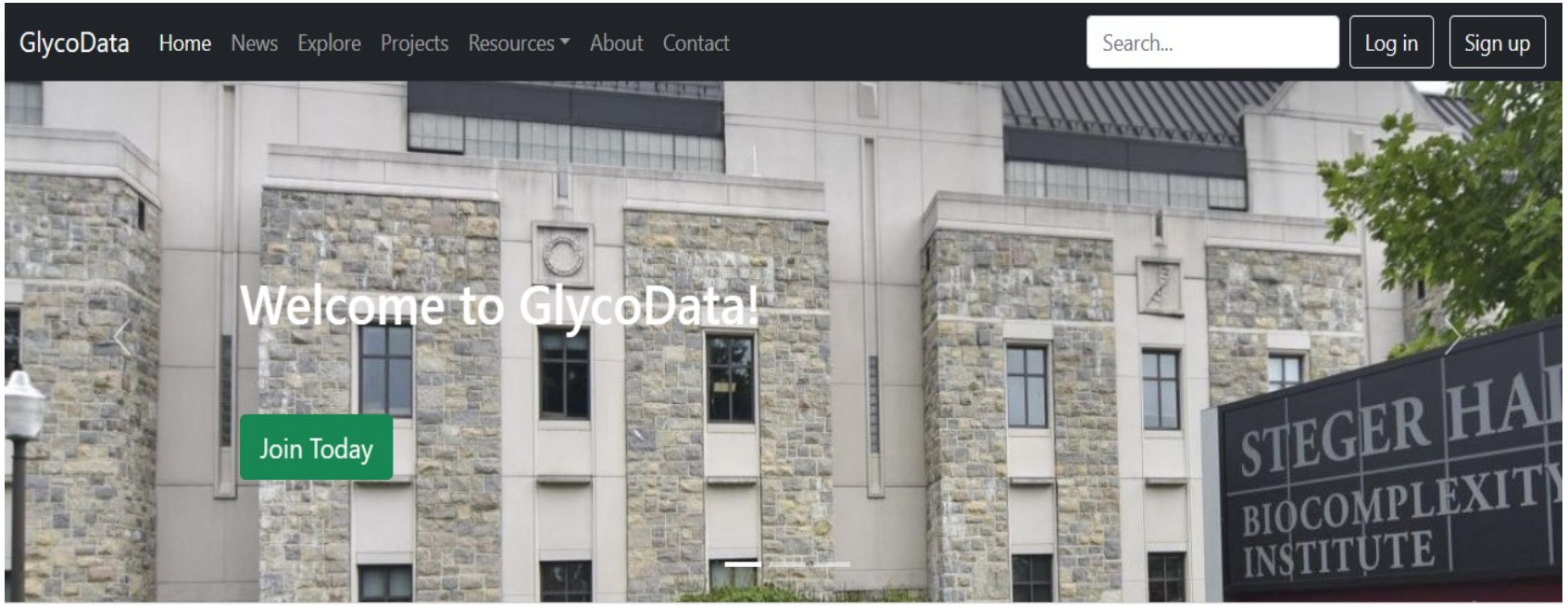
Database can supply info for up to 100K users.

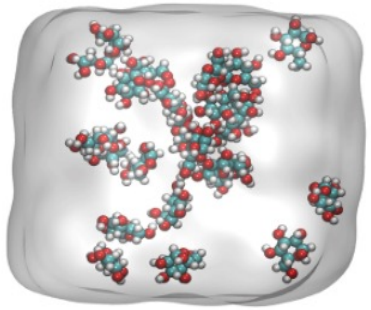
➢ Improves research visibility and promotes collaborative research by uploading/downloading data.

➢ GlycoData Homepage (*glycodata.org*)

User interface elements

- ❖ Navigation menus
  - Explore
  - Projects
  - Resources
- ❖ Search bars
- ❖ Login and Account Management
- ❖ Contact information and links to GlycoMIP social media





News

1. GlycoMIP Annual Meeting 2023, Athens, University of Georgia, March 5th - 7th, 2023
2. Upcoming Event GlycoMIP Scientists on Screen.
3. Save the Date for GlycoMIP Summer School 2023 @ Virginia Tech, June 12th - 16th, 2023
4. Save the Date! GlycoMIP Summer School 2023 @ University of Georgia, September 11th -13th , 2023

Security and Privacy Measures

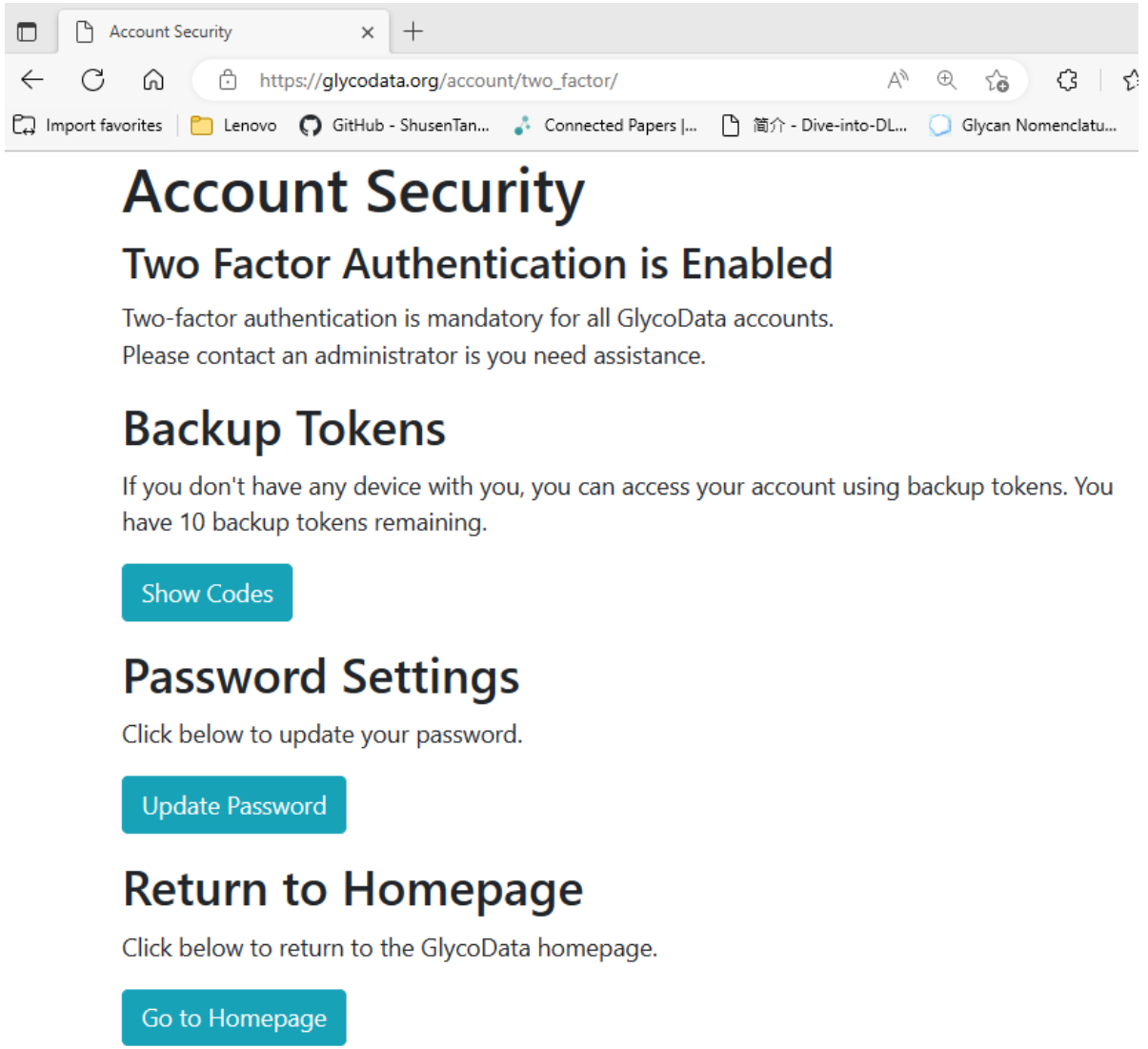
Network Security:

- Hypertext Transfer Protocol Secure (HTTPS) is enabled to protect sensitive data. Visiting **Glycodata** via HTTPS, the web browser establishes a secure connection with the **Glycodata** server, and all data exchanged between the two is encrypted.

User Account Security:

- Two Factor Authentication(2FA) is required for signup users to obtain full experience of GlycoData. The 2FA is also required by the server provider (VT ARC).

- Google Authenticator, Duo Mobile or other 2FA authenticator can be used. For users' convenience, Backup Tokens are provided for 2FA as well.



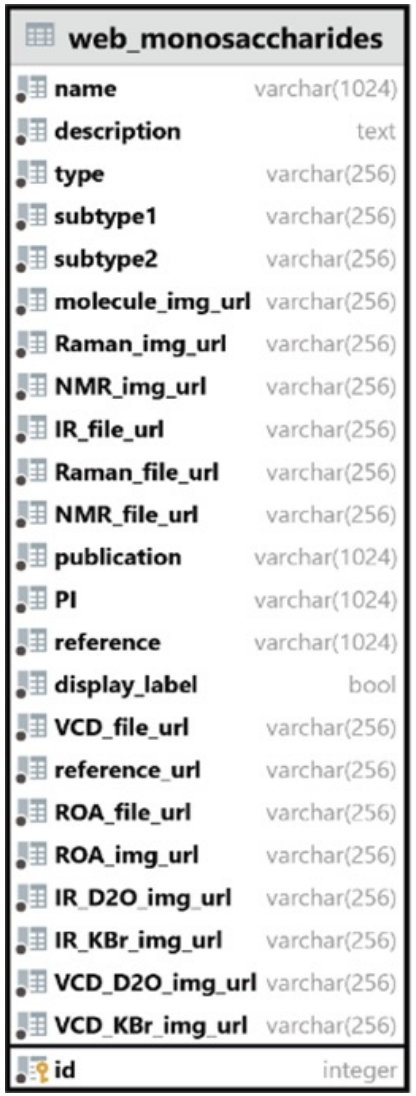
➢ Data management and storage systems

Currently, **GlycoData** platform consists of three databases:

1. Projects Database
2. Monosaccharides Database
3. Polysaccharides Database

Projects Database

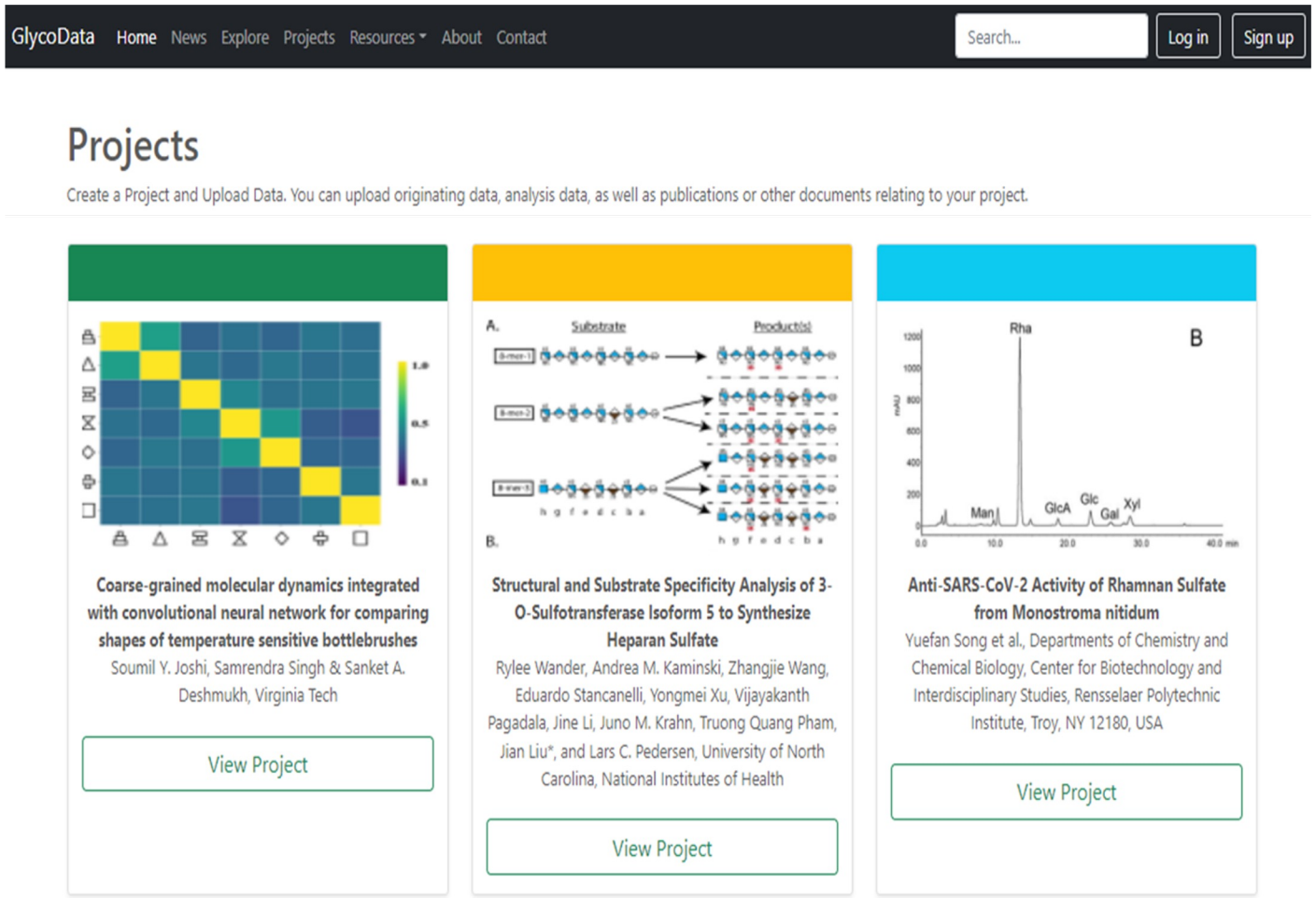
- ❖ Users can create projects that link to their published work or create a new, open-access project. A project description is required and a graphical abstract is recommended.
- ❖ To accommodate projects with large datasets, users can also share a link to raw data housed in Google Drive or on GitHub.

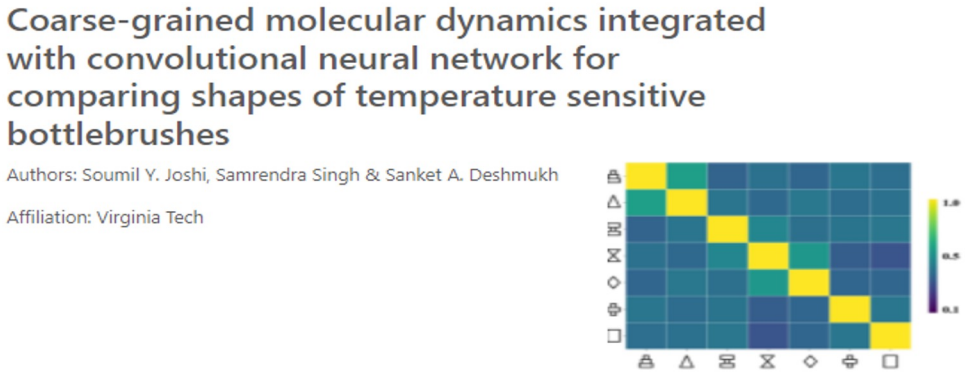


➢ We will constantly update our table schemas based on the available data and build up the relationships between tables.

Project Database

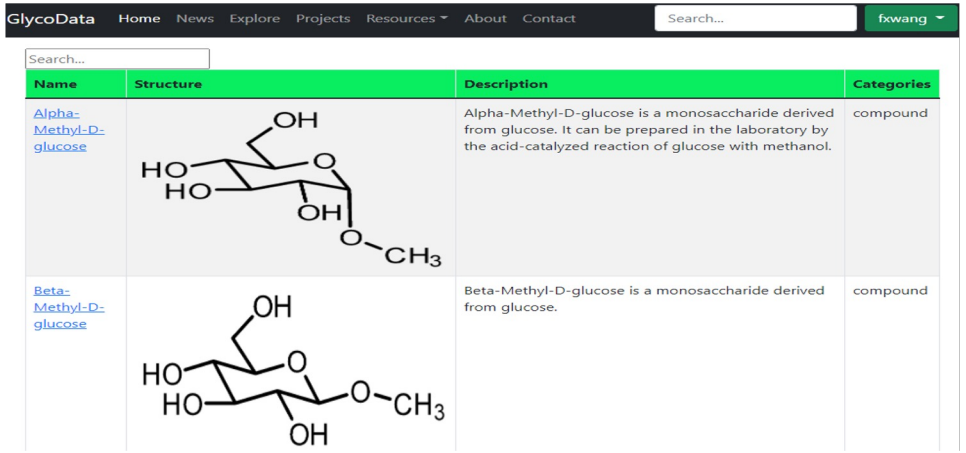
- Color schema is used to define the different type of projects, e.g In-house project (Green).

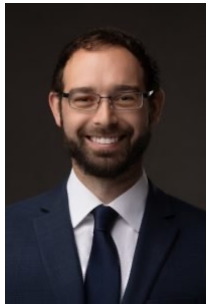




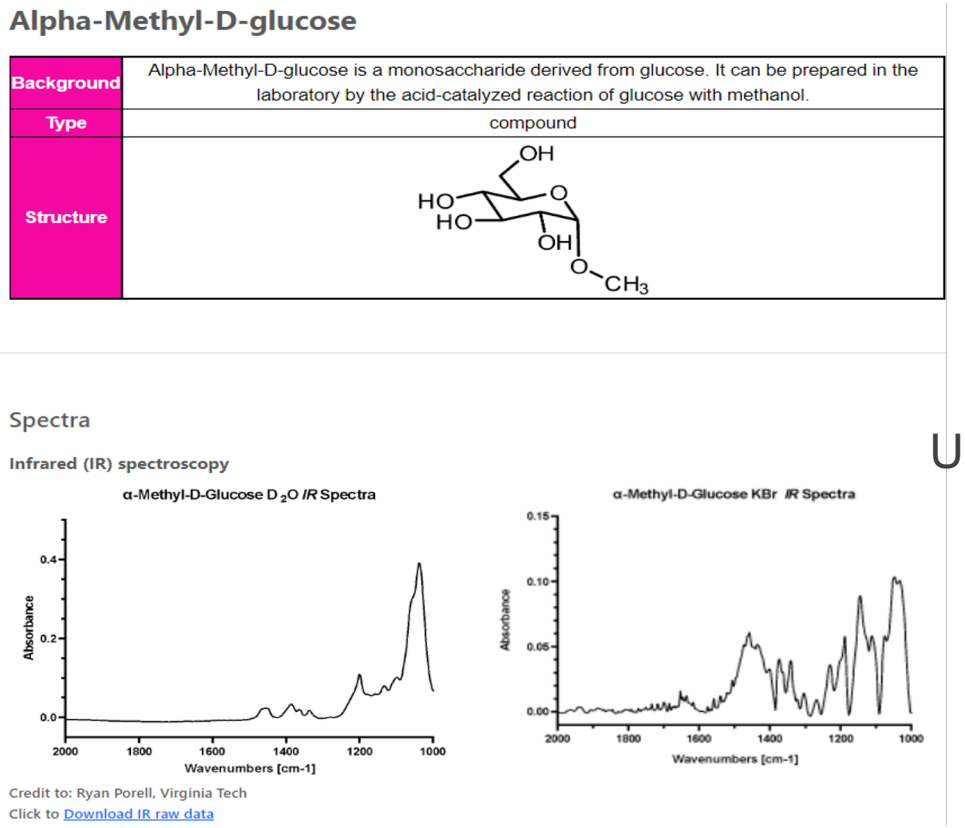
Monosaccharides Database

- The experimental data of each monosaccharide characterized using GlycoMIP facilities is provided, including :
  - Infrared (IR)
  - Vibrational Circular Dichroism (VCD)
  - Raman Optical Activity (ROA)
  - Nuclear Magnetic Resonance (NMR)
- This experimental raw data can be download ONLY by registered users.
- This data will be formatted such that it will hold considerable value for those who wish to utilize machine learning approaches in their projects.





Ryan Porell (VT)



Unpublished Data

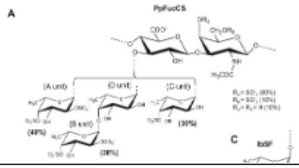
➢ Polysaccharides Database

- ❖ Currently, this database stores data from published, in-house GlycoMIP projects.
- ❖ We are reaching out to the Principal Investigators and Teams to obtain their experimental and/or computational raw data.
- ❖ The database table schema we use to catalog and present this information will evolve based on the data formats.

Search...

Name	Reference	P.I
Fucosylated chondroitin sulfate	R. Dwivedi et al., J. Biol. Chem. 297, 101207 (2021).	Robert J. Linhardt
Rhamnan sulfate	Y. Song et al., Mar. Drugs. 19 (2021)	Robert J. Linhardt
Heparan Sulfate	E. Ahat et al., Cell. Mol. Life Sci. 79, 199 (2022).	Robert J. Linhardt
Chondroitin Sulfate	E. Ahat et al., Cell. Mol. Life Sci. 79, 199 (2022).	Robert J. Linhardt
Pentosan polysulfate	F. Zhang et al., Pharmaceuticals . 15 (2022)	Robert J. Linhardt
Mucopolysaccharide polysulfate	F. Zhang et al., Pharmaceuticals . 15 (2022)	Robert J. Linhardt

Fucosylated chondroitin sulfate

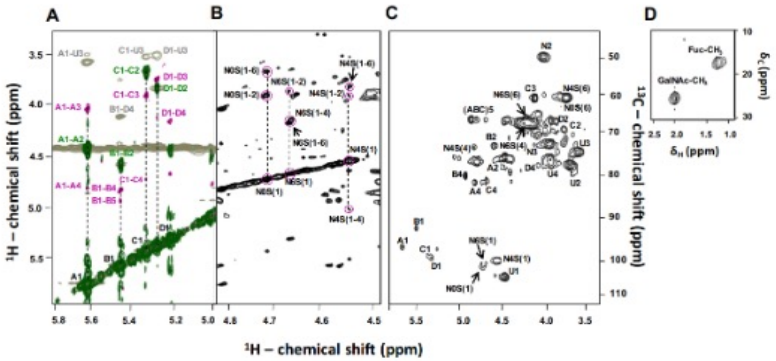
Background	Pentacta pygmaea (PpFucCS) was extracted from sea cucumber and it's structure was analyzed using NMR.PpFucCS exhibits strong binding to S-protein RBD compared to heparin.
Type	Glycosaminoglycans
Source	Pentacta pygmaea(Sea cucumber)
Weight	10-60 kDa
Backbone/Unit	{→3)-β-GalNAcX-(1→4)-β-GlcA-[(3→1)Y]-(1→} where X = 4S(80%), 6S(10%) or nonsulfated(10%), Y = αFuc2,4S (40%), α-Fuc2,4S-(1→4)-α-Fuc(30%)
Structure	

Properties

IC_{50} calculations for SARS-CoV-2: 0.0279 mg/l

Spectra

NMR



References

Structural and Kinetic Analyses of Holothurian Sulfated Glycans Suggest Potential Treatment for SARS-CoV-2 Infection

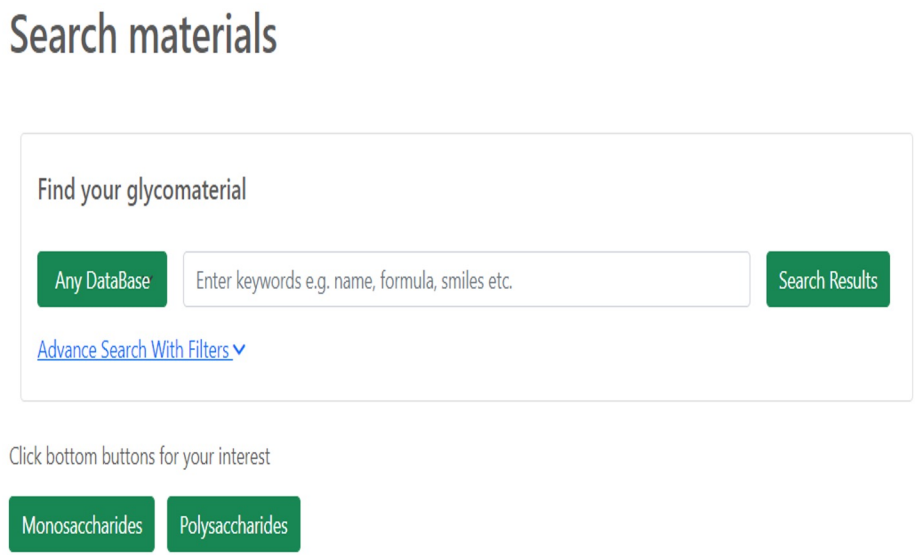
➢ Search Functionality and Data Retrieval

Searching

- ❖ Users are able to search glycomaterials with a Keyword search, advanced search, and filtering options.
- ❖ Drop menu allows user to search glycomaterials in specific database.

Data retrieval and tracking

- ❖ We plan to provide APIs to users for easy access and downloading the data.
- ❖ The number of downloads for glycomaterials data will be tracked to helps provide data to users.



Links to Useful Resources

Resources List

Below are useful databases for glycans and related glycomaterial

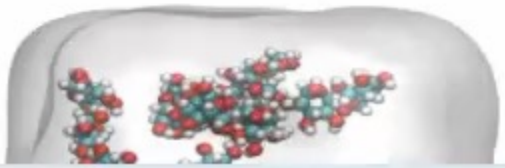
Resource Name	Description
GlyGen: Computational and Informatics Resources for Glycoscience	GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.
GlyTouCan	GlyTouCan is the international glycan structure repository.
GLYCAM-Web	GLYCAM-Web is produced by the research group of Professor Robert J. Woods in the Complex Carbohydrate Research Center at the University of Georgia in Athens GA. It¡¯s purpose is to facilitate quality computer modeling of carbohydrates.
Glycosciences.de	The Glycosciences.de web portal provides databases and tools to support glycobiology and glycomics research.
KEGG GLYCAN Database	KEGG GLYCAN is a collection of experimentally determined glycan structures, containing all unique structures taken from CarbBank, structures entered from recent publications, and structures present in KEGG pathways.
Carbohydrate Structure Database (CSDB)	The Carbohydrate Structure Database is a free curated repository storing various data on glycans of bacterial, fungal and plant origins.
UniCarb-DB	UniCarb-DB is one of the largest experimental glycomic MS databases
Glycan Library from Glycosciences Laboratory	A list of lipid-linked, neoglycolipids (NGLs) or glycosylceramides oligosaccharide probes from the collection assembled in the course of research in the Glycosciences Laboratory.
Carbohydrate-Active enZYmes Database (CAZy)	The CAZy database describes the families of structurally-related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds.
Expasy resource portal	Expasy is the bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics
The Human Metabolome Database (HMDB)	The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body.
The O-GlcNAc Database	The O-GlcNAc database is where to match your protein of interest or your dataset to get O-GlcNAcylation information

Home-GlycoData

127.0.0.1:8000

GitHub - ShusenTan... | Connected Papers |... | 简介 - Dive-into-DL... | Glycan Nomenclatu... | MDTraj — MDTraj 1.... | 菜鸟教程 - 学的不... | Scikit-learn cheat s... | Virtual Machine Env... | Backup | 3D Collection of Bin...

GlycoData Home News Explore Projects Resources About Contact Search... Log in Sign up



News

1. GlycoMIP Annual Meeting 2023, Athens, University of Georgia, March 5th - 7th, 2023
2. Upcoming Event GlycoMIP Scientists on Screen.
3. Save the Date for GlycoMIP Summer School 2023 @ Virginia Tech, June 12th - 16th,

8:40 PM
2/26/2023

➢ Ongoing work

- We are closely engaging with the GlycoMIP community to gather data and update existing or create new tables for different categories of glycomaterials.

- The simulation tools, machine learning scripts, and analysis tools are being developed.

- We are also building a file management module for internal data sharing.

- We welcome feedback on the GlycoDATA platform.

Thank you.