



GlyComb: a novel glycoconjugate data repository that bridges glycomics and proteomics

Description

Protein glycosylation is a general post-translational modification involved in cell membrane formation, is crucial to dictate proper conformation of many membrane proteins, retaining stability on some secreted glycoproteins, and playing a role in cell–cell adhesion. Therefore, the analysis of post-translational modifications of proteins, including glycosylation, is essential to understand the exact functions and interactions of each protein molecule. To conduct large-scale analyses more efficiently, it is essential to promote the accumulation, sharing, and reuse of experimental and analytical data in accordance with the FAIR (Findability, Accessibility, Interoperability, and Re-usability) data principles. However, there was no FAIR data repository for storing and sharing information on glycoconjugates, including glycopeptides and glycoproteins, in a standardised format.

The authors have developed GlyComb (<https://glycomb.glycosmos.org>) as a new standardized data repository for glycoconjugate data. GlyComb can assign a unique identifier to a set of glycosylation information associated with a specific peptide sequence or UniProt ID. By standardising glycosylation data using GlyComb identifiers and coordinating with existing web resources such as GlyYouCan and GlycoPOST, a comprehensive system for data submission and sharing between researchers can be established. This article describes how GlyComb can integrate the variety of glycoconjugate data already registered in existing data repositories to gain a better understanding of the available glycopeptides and glycoproteins and their glycosylation patterns. We also explain how this system can be used as a basis for a better understanding of glycan function.

(a) Example Glycopeptide Data

Paste your data to the text area below:

```

DLTLANTTLQTCR  8  HexNAc4(Hex7)HexAc2
DLTLANTTLQTCR  8  HexNAc4(Hex7)Fuc2
DLTLANTTLQTCR  8  HexNAc2(Hex8)
DLTLANTTLQTCR  8  HexNAc2(Hex9)
VANSSEAPPPHNVSTLLTSAGAR  3  G2S8Q3AS  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  3  G2736KV  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  3  G3643ZY  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  3  G5064TK  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  23  G61299CK  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  23  G9454FU  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  23  G262134FA  Q8E7H1
VANSSEAPPPHNVSTLLTSAGAR  23  G23432EQ  Q8E7H1

```

(b) Example Glycopeptide Data

Choose your file to submit (TSV/Excel/MS file format):

Select a file... datasets.xlsx

Choose the worksheet to read the data:

GlyComb

Choose the start row number to read the data (without headers):

Row 1

Choose the column for amino acid sequences:

Column A

Choose the column for glycosylation sites:

Column B

Choose the column for glycan structures:

Column C

Choose the column for additional data:

Column D

(c) No error found.

Status	User input	Amino acid sequence	Glycosylation site	Glycan structure	Additional data
✓	DLTLANTTLQTCR HexNAc2(Hex8)	DLTLANTTLQTCR	8	HexNAc2(Hex8)	NA
✓	DLTLANTTLQTCR HexNAc2(Hex9)	DLTLANTTLQTCR	8	HexNAc2(Hex9)	NA
✓	DLTLANTTLQTCR HexNAc4(Hex7)Fuc2	DLTLANTTLQTCR	8	HexNAc4(Hex7)Fuc2	NA
✓	DLTLANTTLQTCR HexNAc4(Hex7)HexAc2	DLTLANTTLQTCR	8	HexNAc4(Hex7)HexAc2	NA

(d) All your requests were accepted successfully.

Your submission numbers are:

- 0ee6030-5b01-4218-971d-988b6f887b1
- 85ae4082-a5b6-4303-9bc2-6ce65a034caa

(e) Confirm

Please paste the submission numbers you want to make public below line by line. All of the specified submissions will be opened by clicking the "I understand and wish to proceed" button. **This action is irreversible.**

```

0ee6030-5b01-4218-971d-988b6f887b1
85ae4082-a5b6-4303-9bc2-6ce65a034caa
b1df3e00-0a10-4581-895d-f4e0b9b84fc3
4fa76116-7d3c-4e55-924e-0fd9fc733d2f
21f5cb2f-459a-4eaf-8b4c-1ce05a419c18
a5228db7-5584-45fa-ae88-95c520528334
82228da2-4a87-49e2-a7db-856d64a49590

```

Workflow for registering and making glycopeptide entries publicly available in GlyComb. Researchers access the entry registration screen (<https://glycomb.glycosmos.org/registration>) after logging into GlyComb with their Google account. They can either (a) copy and paste the entries they want to register from the clipboard into the text area on the screen, or (b) select an MS Excel file or TSV file to upload. When uploading a file, they can choose the worksheet to be read or select the columns to be read. (c) A confirmation screen to verify that the submission content is correct will be displayed. (d) GlyComb displays a submission number instead of the GlyComb ID, which is the accession number when the submission of an input entry is completed. Researchers can later make their registered entries open to the public by using these submission numbers. (e) A confirmation screen to make each registered entry available to the public (https://glycomb.glycosmos.org/user_profile). Through batch processing on the GlyComb server, they are assigned unique GlyComb IDs within a few hours after the input entries are submitted and each entry is ready to be published. By

entering the submission numbers generated when the glycopeptide or glycoprotein entries were submitted, one per line, and clicking the button at the bottom of the pop-up window, researchers can make multiple GlyComb entries they have submitted open to the public at once.

Category

1. News