SimGlycan: Facilitating Automated Evaluation of Glycoengineered Therapeutic Glycoproteins using MS Methods

PREMIER Biosoft

Accelerating Research in Life Sciences

Ningombam Sanjib Meitei[1,2], Rupanjan Goswami[1]

[1]PREMIER Biosoft, Indore, India; [2]PREMIER Biosoft, CA, US

Corresponding author e-mail: sanjib@premierbiosoft.com

1

INTRODUCTION

Monoclonal antibody (mAb) based therapeutics are the fastest growing class of human pharmaceuticals [1]. Variation in IgG Fc glycoforms affects the safety, clinical efficacy, and effector functions of therapeutic antibodies including antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) e.g., the terminal galactose enhances CDC activity, but fucose significantly decreases ADCC [2].

Liquid Chromatography-Mass spectrometry (LC-MS) based methods are used for analysis of therapeutic mAb Fc glycosylation profiles [3,5]. Tandem MS enables identification of monosaccharide compositions and linkage orientation of the glycans based on glycosidic and cross-ring fragment ions [6] thereby enabling the quality evaluation of the glycoengineered mAbs e.g., monitoring specific transformation and change in relative quantities of a set of glycans between glycoengineered mAb vs the mAb produced in CHO cell lines [3].

Efficient bioinformatics tools are required to facilitate evaluation of glycoengineered therapeutic mAbs using MS-based methods. We have created a database of N-glycan structures of therapeutic glycoproteins curated from published literature [1-5], developed new modules of SimGlycan software [7,8] to identify glycans, quantify them using their corresponding extracted ion chromatogram peak areas, and exporting the results into MS excel files enabling comparative evaluation of glycoengineered mAbs using MS methods.

EXPERIMENTAL METHODS

N-glycans were released by PNGase F from a genetically engineered human monoclonal antibody (mAb), Sample A and a glycoengineered mAb, Sample B both produced by mammalian cell (Chinese Hamster Ovary) suspension culture in a nutrient medium containing the antibiotic gentamicin. Using a similar published method [9], the released glycans were then labeled with RapiFluor-MS reagent (Waters). The final solution was treated for HILIC enrichment and the enriched glycan was injected onto the ACQUITY I UPLC Glycan BEH Amide column, and electro-sprayed into the Orbitrap Fusion Mass spectrometer (Thermo Fisher).

DECLARATION

The main objective of the experiments was to investigate whether the glycoengineering processes would cause any other post-translational modifications of the mAbs by setting up a null hypothesis that there is no significant difference in the molar composition of the glycoforms between the two antibodies. However, reporting the finding on the hypothesis is out of the scope for this poster. We only showed how we can automate the process of glycan identification using mass spectrometry data for the samples and perform comparative analysis between the samples thereby facilitating evaluation of glycoengineered monoclonal antibodies.

DATA ANALYSIS

SimGlycan v. 5.92 software (PREMIER Biosoft, USA), and Microsoft Excel 2016.

DATABASE CURATION

To facilitate rapid and accurate annotation of glycans, a custom database containing 71 curated compositional structures of manufactured recombinant monoclonal antibody drugs [10] was created by categorizing them into four groups namely Group 1: Acidic and Afucosylated (9 compositions); Group 2: Acidic and Fucosylated (27 compositions) 3: Neutral and Afucosylated (21 compositions), and Group 4: Neutral and Fucosylated (17 compositions) in SimGlycan server database (Table 1). This custom database was used as the MS/MS database search template.

Group 1: Acidic and Afucosylated Glycan Composition	Group 2: Acidic and Fucosylated Glycan Composition	Group 3: Neutral and Afucosylated Glycan Composition	Group 4: Neutral and Fucosylated Glycan Composition
(Gal)1 (GlcNAc)3 (Man)4 (NeuAc)1	(Ac)1 (Fuc)1 (Gal)3 (GlcNAc)5 (Man)3 (NeuAc)3	(Gal)1 (GlcNAc)3 (Man)4	(Fuc)1 (Gal)1 (GlcNAc)3 (Man)3
(Gal)1 (GlcNAc)3 (Man)5 (NeuAc)1	(Ac)1 (Fuc)1 (Gal)4 (GlcNAc)6 (Man)3 (NeuAc)4	(Gal)1 (GlcNAc)3 (Man)5	(Fuc)1 (Gal)1 (GlcNAc)4 (Man)3
(Gal)2 (GlcNAc)4 (Man)3 (NeuAc)1	(Fuc)1 (Gal)1 (GlcNAc)3 (Man)3 (NeuAc)1	(Gal)1 (GlcNAc)4 (Man)3	(Fuc)1 (Gal)1 (GlcNAc)5 (Man)3
(Gal)2 (GlcNAc)4 (Man)3 (NeuAc)1 (NeuGc)1	(Fuc)1 (Gal)1 (GlcNAc)3 (Man)4 (NeuAc)1	(Gal)2 (GlcNAc)4 (Man)3	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3
(Gal)2 (GlcNAc)4 (Man)3 (NeuAc)2	(Fuc)1 (Gal)1 (GlcNAc)4 (Man)3 (NeuAc)1	(Gal)3 (GlcNAc)5 (Man)3	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3
(Gal)3 (GlcNAc)4 (Man)3 (NeuAc)2	(Fuc)1 (Gal)1 (GlcNAc)5 (Man)3 (NeuAc)1	(Gal)4 (GlcNAc)6 (Man)3	(Fuc)1 (Gal)2 (GlcNAc)5 (Man)3
(Gal)3 (GlcNAc)5 (Man)3 (NeuAc)2	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)1	(Glc)1 (GlcNAc)2 (Man)9	(Fuc)1 (Gal)3 (GlcNAc)4 (Man)3
(Gal)3 (GlcNAc)5 (Man)3 (NeuAc)3	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)1	(Glc)2 (GlcNAc)2 (Man)9	(Fuc)1 (Gal)3 (GlcNAc)5 (Man)3
(Gal)4 (GlcNAc)6 (Man)3 (NeuAc)4	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)1 (NeuGc)1	(Glc)3 (GlcNAc)2 (Man)5	(Fuc)1 (Gal)4 (GlcNAc)4 (Man)3
	(Fuc)1 (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)2	(Glc)3 (GlcNAc)2 (Man)7	(Fuc)1 (Gal)4 (GlcNAc)6 (Man)3
	(Fuc)1 (Gal)3 (GlcNAc)5 (Man)3 (NeuAc)1	(Glc)3 (GlcNAc)2 (Man)9	(Fuc)1 (Gal)6 (GlcNAc)8 (Man)3
	(Fuc)1 (Gal)3 (GlcNAc)5 (Man)3 (NeuAc)2	(GlcNAc)2 (Man)10	(Fuc)1 (GlcNAc)2 (Man)3
	(Fuc)1 (Gal)3 (GlcNAc)5 (Man)3 (NeuAc)2 (NeuGc)1	(GlcNAc)2 (Man)3	(Fuc)1 (GlcNAc)2 (Man)5
	(Fuc)1 (Gal)3 (GlcNAc)5 (Man)3 (NeuAc)3	(GlcNAc)2 (Man)4	(Fuc)1 (GlcNAc)3 (Man)3
	(Fuc)1 (Gal)4 (GlcNAc)6 (Man)3 (NeuAc)2	(GlcNAc)2 (Man)5	(Fuc)1 (GlcNAc)3 (Man)5
	(Fuc)1 (Gal)4 (GlcNAc)6 (Man)3 (NeuAc)3	(GlcNAc)2 (Man)6	(Fuc)1 (GlcNAc)4 (Man)3
	(Fuc)1 (Gal)4 (GlcNAc)6 (Man)3 (NeuAc)3 (NeuGc)1	(GlcNAc)2 (Man)7	(Fuc)1 (GlcNAc)5 (Man)3
	(Fuc)1 (Gal)4 (GlcNAc)6 (Man)3 (NeuAc)4	(GlcNAc)2 (Man)8	
	(Fuc)1 (Gal)5 (GlcNAc)7 (Man)3 (NeuAc)3	(GlcNAc)2 (Man)9	
	(Fuc)1 (Gal)5 (GlcNAc)7 (Man)3 (NeuAc)4	(GlcNAc)3 (Man)3	
	(Fuc)1 (Gal)6 (GlcNAc)8 (Man)3 (NeuAc)3	(GlcNAc)4 (Man)3	
	(Fuc)1 (Gal)6 (GlcNAc)8 (Man)3 (NeuAc)4		
	(Fuc)1 (Gal)7 (GlcNAc)9 (Man)3 (NeuAc)4		
	(Fuc)3 (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)2		

Table 1: List of the 71 curated compositional glycan structures from CHO Cell Lines.

RESULTS AND DISCUSSION

The LC-MS data were subjected into SimGlycan's LC-MS data processing module for peak detection, peak picking, molecular feature finding, and cluster MS/MS data to LC compounds (Figure 1). Figure 2 shows the typical graphical user interfaces of SimGlycan software displaying peaklist and an extracted ion chromatogram (XIC) of a detected LC-compound.

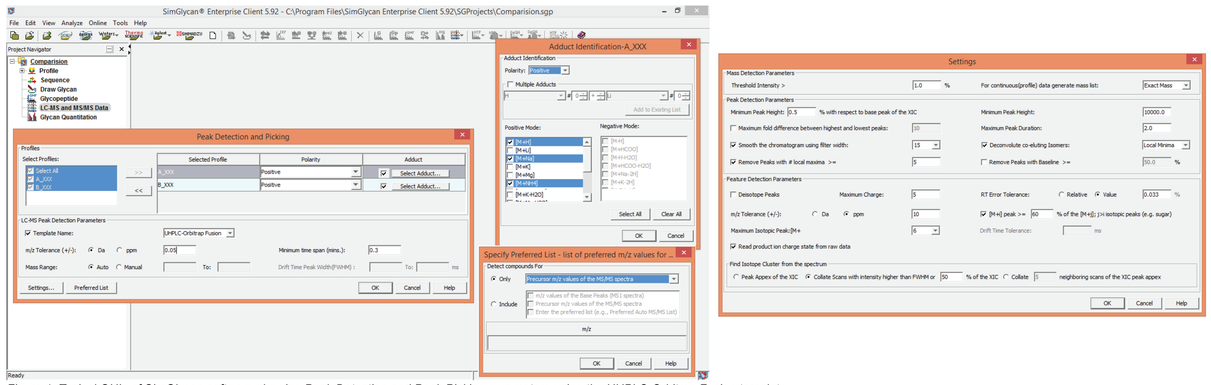


Figure 1: Typical GUIs of SimGlycan software showing Peak Detection and Peak Picking parameters using the UHPLC-Orbitrap Fusion template.

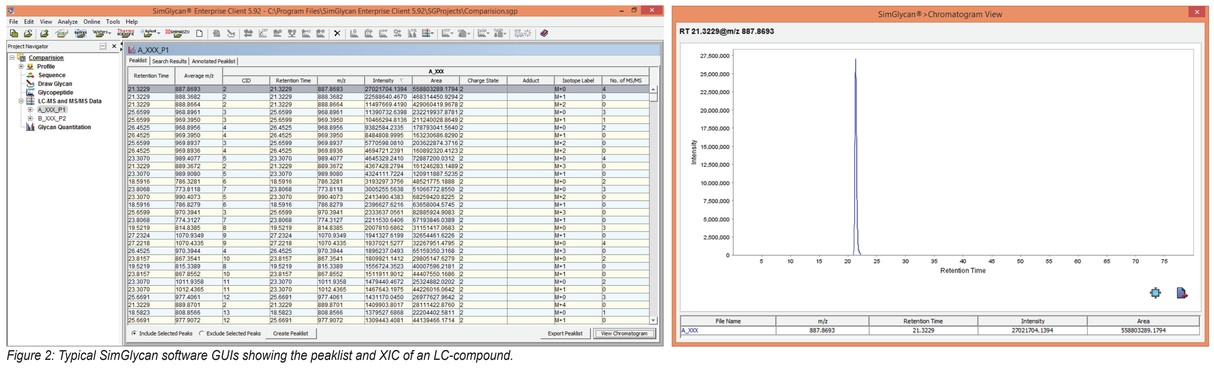


Figure 2: Typical SimGlycan software GUIs showing the peaklist and XIC of an LC-compound.

2

The MS/MS data was subjected to SimGlycan database search for glycan annotation using the parameter settings displayed in the Figure 3. The search was performed on the custom database containing 71 curated compositional structures using 10 ppm tolerance for the precursor and product ions.

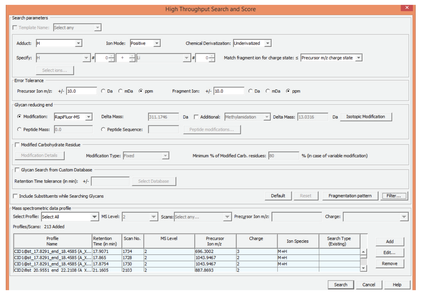


Figure 3: Typical SimGlycan software search window showing search parameters for database search.

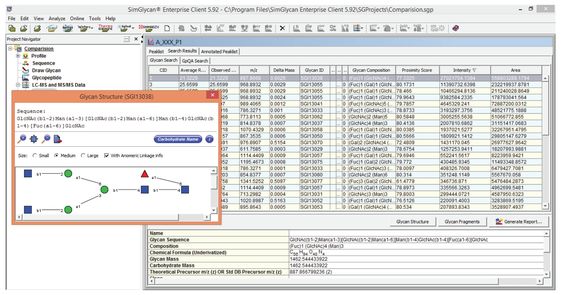


Figure 4: Typical GUIs of SimGlycan software showing identification of (Fuc)1 (GlcNAc)4 (Man)3 compositional structure as the most abundant glycan in the sample A.

The SimGlycan database search annotated 43 LC-compounds with 24 compositional structures (Figure 4) for the sample A while for the sample B, the program annotated 201 LC-compounds with 49 unique compositional structures (data not shown). Peaks in the MS/MS spectra are automatically annotated with fragment structures using three different methods namely, fragment names based on Domon and Costello fragment nomenclature [6], successive loss of monosaccharide residues, and symbolic representations of fragment structures. Finally, the results were exported into a Microsoft Excel file for further analysis (Figure 6).

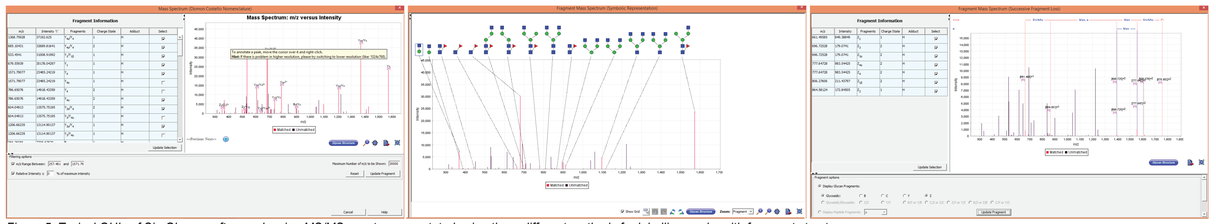
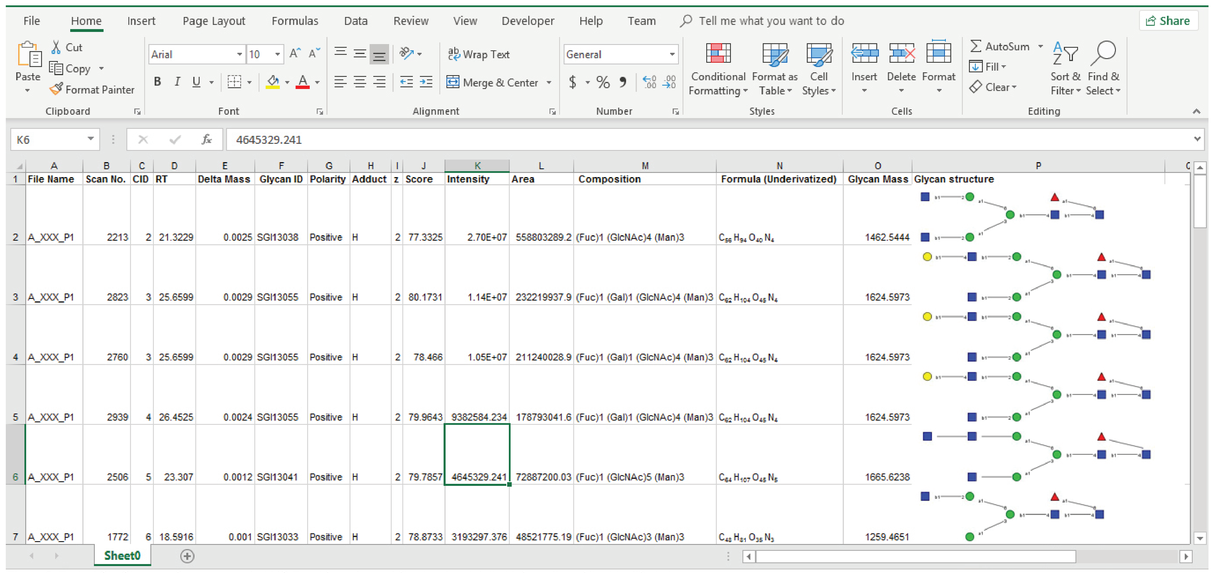


Figure 5: Typical GUIs of SimGlycan software showing MS/MS spectrum annotated using three different methods for labelling peaks with fragment structures.



	File Name	Scan No.	CID	RT	Delta Mass	Glycan ID	Polarity	Adduct	z	Score	Intensity	Area	Composition	Formula (Underivatized)	Glycan Mass	Glycan structure
2	A_XXX_P1	2213	2	21.3229	0.0025	SGI13038	Positive	H	2	77.3325	2.70E+07	558803289.2	(Fuc)1 (GlcNAc)4 (Man)3	$C_{56}H_{94}O_{40}N_4$	1462.5444	
3	A_XXX_P1	2823	3	25.6599	0.0029	SGI13055	Positive	H	2	80.1731	1.14E+07	232219937.9	(Fuc)1 (Gal)1 (GlcNAc)4 (Man)3	$C_{62}H_{104}O_{45}N_4$	1624.5973	
4	A_XXX_P1	2760	3	25.6599	0.0029	SGI13055	Positive	H	2	78.466	1.05E+07	211240028.9	(Fuc)1 (Gal)1 (GlcNAc)4 (Man)3	$C_{62}H_{104}O_{45}N_4$	1624.5973	
5	A_XXX_P1	2939	4	26.4525	0.0024	SGI13055	Positive	H	2	79.9643	9382584.234	178793041.6	(Fuc)1 (Gal)1 (GlcNAc)4 (Man)3	$C_{62}H_{104}O_{45}N_4$	1624.5973	
6	A_XXX_P1	2506	5	23.307	0.0012	SGI13041	Positive	H	2	79.7857	4645329.241	72887200.03	(Fuc)1 (GlcNAc)5 (Man)3	$C_{64}H_{107}O_{48}N_5$	1665.6238	
7	A_XXX_P1	1772	6	18.5916	0.001	SGI13033	Positive	H	2	78.8733	3193297.376	48521775.19	(Fuc)1 (GlcNAc)3 (Man)3	$C_{48}H_{81}O_{36}N_3$	1259.4651	

Figure 6: Typical SimGlycan software generated MS Excel format report file showing detailed information of the identified glycans.

DOWNSTREAM ANALYSIS

The identified compositional glycans were converted into monosaccharide groups – Hex, HexNAc, Fuc, NeuAc, and NeuGc. 49 glycans with unique group compositions were identified between the samples. Area under curve of XICs of the identified glycans were used as their measure of quantity in the samples.

DATA NORMALIZATION

The program identified 49 glycans with unique group compositions between the samples. We classify the glycans into two groups namely, Acidic: glycans, containing NeuAc or NeuGc, and Neutral: glycans, that do not contain NeuAc or NeuGc. The percentage abundance of each glycan in the sample is the ratio between the observed abundance of each glycan in a group and the sum of abundances of all the glycans in that group. The classification of glycans into two groups for data normalization was done to nullify the effect of variation in ionization efficiency between the acidic versus the neutral glycans (if there are any).

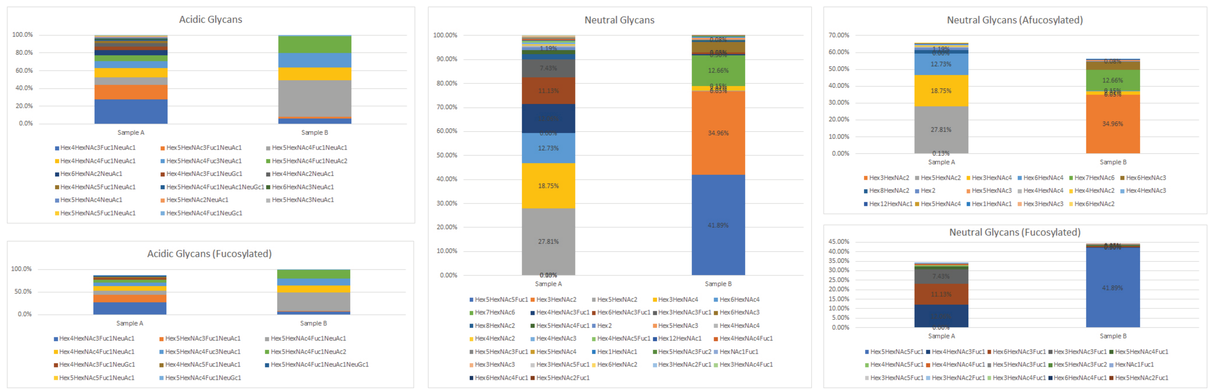


Figure 7: Bar charts generated using MS Excel showing change in molar percentage of acidic and neutral glycans in the samples.

CONCLUSION

New software modules of SimGlycan v. 5.92 provide the following key features:

1. High resolution accurate mass (HRAM) LC-MS data processing – accurate peak picking, and charge state identification.
2. Rapid and accurate identification of glycans using LC-MS templates.
3. Automated interpretation of MS/MS spectra by annotating peaks with fragment ions using three different methods namely, (a) Domon and Costello Fragment nomenclature, (b) Successive loss of monosaccharide residues, and (c) Symbolic representation of fragment ion structures.
4. Export results into MS Excel files for further downstream analysis.
5. Different data normalization methods to facilitate effective evaluation of changes in glycoforms in glycoengineered mAbs.

REFERENCES

1. Li W et al. Front. in Immun., 2017; 8, 1554
2. Mimura Y et al., Protein & Cell, 2018; 9(1): 47–62.
3. Liu R et al., mAbs, 2016; 8(2): 340–46.
4. Ratner M. Nat Biotech., 2014; 32:6-7.
5. Reusch D et al, MAbs. 2015; 7(4): 732-42.
6. Domon B Costello, C.E., Glycoconj J., 1988; 5.4: 397-09.
7. Apte A, Meitei, NS, Func Glycomics, Humana Press, 2010, 269-81.
8. Meitei NS et al., J of prot., (2015) 127: 211-22.
9. Kim, Young Hun, et al. J Cell Sci 129.1 (2016): 29-38.
10. Song, T et al, (2014), 86(12), 5661-5666.