

# User Manual

A software for data processing of xCGE-LIF  
technology

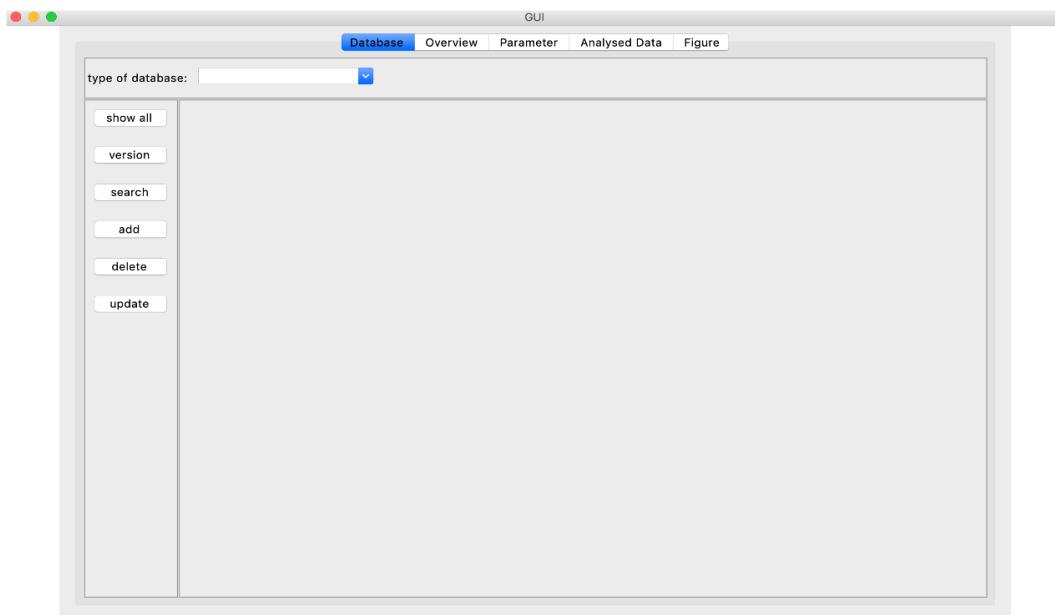
# 1 Structure of the Software

This chapter introduces the overall structure and the functions of the software.

The software consists of five main pages: Database (Chapter 3.1), Overview (Chapter 3.2), Parameter (Chapter 3.3), Analysed Data (Chapter 3.4), Figure (Chapter 3.5).

## 1.1 Page of Database

This page is meant to assist researchers in updating the database on a irregular basis. It performs fundamental database operations such as inserting, removing, updating, and selecting data. Figure 3.1 depicts the main interface of Database.



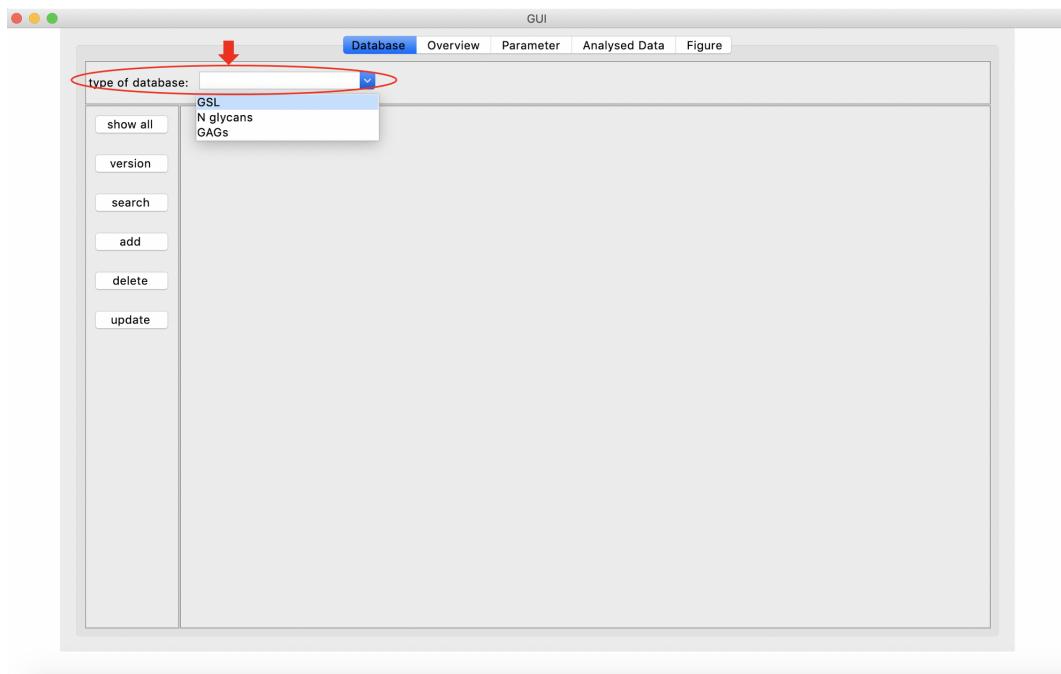
**Figure 1.1:** Main screen of the "Database" page in the software

The following chapters introduce the checkbox "type of database" (Chapter 3.1.1), button "show all" (Chapter 3.1.2), button "version" (Chapter 3.1.3), button "search" (Chapter 3.1.4), button "add" (Chapter 3.1.5), button "delete" (Chapter 3.1.6), and "update" (Chapter 3.1.7).

### 1.1.1 Checkbox – “type of database”

GSL, N-glycans, and GAGs (according to Prof. Dr. Büttner’s demands) are three separate tables in the database produced for different analysis fields. Each record in the database contains the essential information about one glycan. “name\_of\_glycan”, “MTU”, “date”, “flag” are among the most fundamental pieces of information.

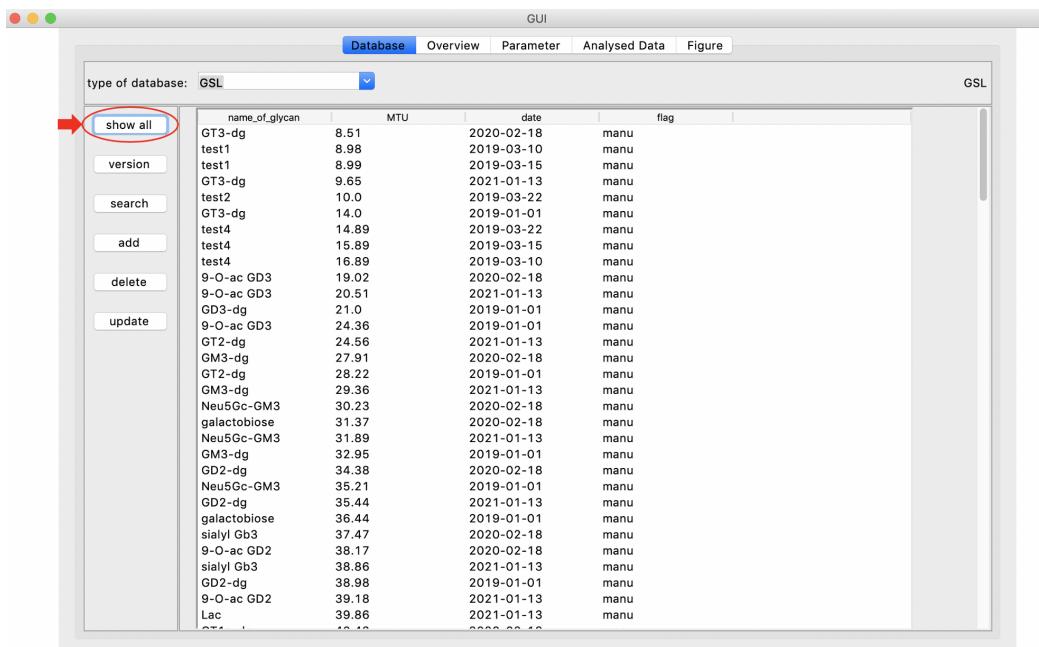
- ▶ “name\_of\_glycan”: string type, indicating the name of the glycan
- ▶ “MTU”: float type, indicating the migration time unit of the glycan
- ▶ “date”: datetime type, indicating the time/date of measurement when the record is stored
- ▶ “flag”: enumerate type, which has manually(manu) and automatically(auto), manu means the record is added manually, auto means the data is added after software calculation.



**Figure 1.2:** On the "Database" page, select the type of database.

### 1.1.2 Button – “show all”

This button brings up a list of all the records stored under the currently selected database as shown in Figure 3.3.



Database				
Overview Parameter Analysed Data Figure				
type of database: GSL				
<b>show all</b>	name_of_glycan	MTU	date	flag
GT3-dg	8.51	2020-02-18	manu	
test1	8.98	2019-03-10	manu	
test1	8.99	2019-03-15	manu	
GT3-dg	9.65	2021-01-13	manu	
test2	10.0	2019-03-22	manu	
GT3-dg	14.0	2019-01-01	manu	
test4	14.89	2019-03-22	manu	
test4	15.89	2019-03-15	manu	
test4	16.89	2019-03-10	manu	
9-O-ac GD3	19.02	2020-02-18	manu	
9-O-ac GD3	20.51	2021-01-13	manu	
GD3-dg	21.0	2019-01-01	manu	
9-O-ac GD3	24.36	2019-01-01	manu	
GT2-dg	24.56	2021-01-13	manu	
GM3-dg	27.91	2020-02-18	manu	
GT2-dg	28.22	2019-01-01	manu	
GM3-dg	29.36	2021-01-13	manu	
Neu5Gc-GM3	30.23	2020-02-18	manu	
galactobiose	31.37	2020-02-18	manu	
Neu5Gc-GM3	31.89	2021-01-13	manu	
GM3-dg	32.95	2019-01-01	manu	
GD2-dg	34.38	2020-02-18	manu	
Neu5Gc-GM3	35.21	2019-01-01	manu	
GD2-dg	35.44	2021-01-13	manu	
galactobiose	36.44	2019-01-01	manu	
sialyl Gb3	37.47	2020-02-18	manu	
9-O-ac GD2	38.17	2020-02-18	manu	
sialyl Gb3	38.86	2021-01-13	manu	
GD2-dg	38.98	2019-01-01	manu	
9-O-ac GD2	39.18	2021-01-13	manu	
Lac	39.86	2021-01-13	manu	
...	...	...	...	...

**Figure 1.3:** Function of the "show all" button on the "Database" page.  
Click this button to display all records stored under the currently selected database.

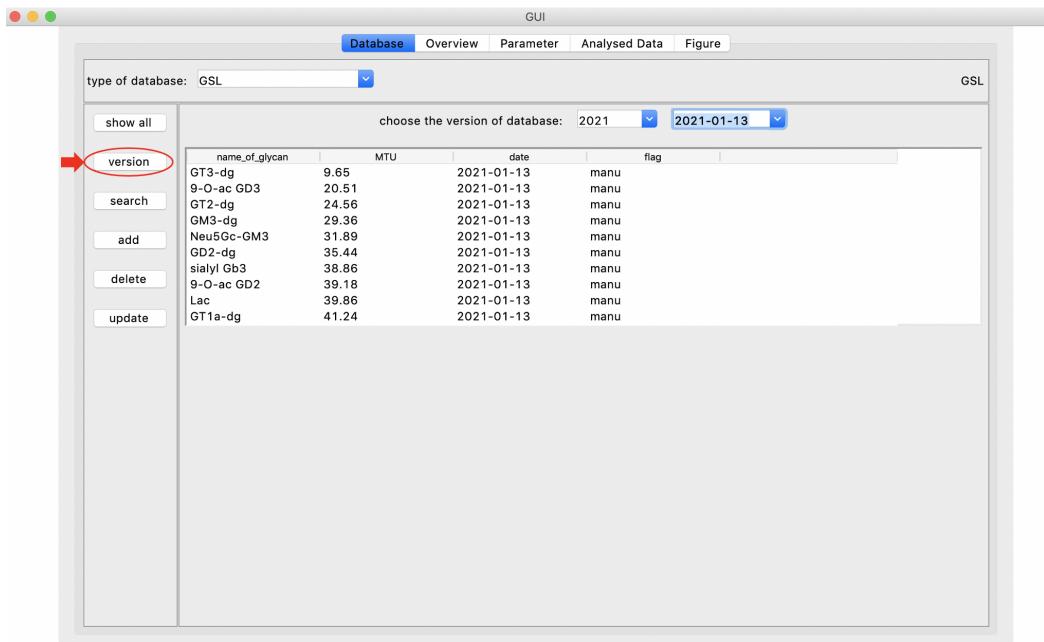
### 1.1.3 Button – “version”

As stated in Chapter 2.3, the database has established the corresponding version for the known glycan measured in each specific period. Only a specific version of the data information will be used to update and label the sample during the data analysis process. Information about the different versions of the data can be found here.

As shown in Figure 3.4, select "year-month-day" to get the measurement data at the appropriate time. If the glycan has a record saved at the selected version, the record will be displayed; if the glycan does not have a record stored at the selected version, its MTU will be displayed as "nan" and the flag will be displayed as "None". Figure 3.5 depicts the condition (Since the storage format of MTU is float, and the storage format of flag is String, the final result will be different when the value is set to null, which are "nan" and "None" respectively).

### 1.1.4 Button – “search”

The "search" button is used to search the database for information. "name\_of\_glycan", "date", and "flag" can be used individually or in combination to search. Figure 3.6 depicts the interface after clicking the "Search" button.



**Figure 1.4:** Function of the "version" button on the "Database" page.

This button displays the records stored at different times in the currently selected database.

name_of_glycan	MTU	date	flag
globoA	224.97	2021-01-13	manu
A type 2 hexa	229.75	2021-01-13	manu
fucosyl Lc6	267.29	2021-01-13	manu
Man6	279.8	2021-01-13	manu
galili antigen hepta	292.88	2021-01-13	manu
GD3-dg	nan	2021-01-13	None
GM1a-dg	nan	2021-01-13	None
GM2-dg analogue	nan	2021-01-13	None

**Figure 1.5:** The results of the "Database" page with and without stored record when a glycan is under the currently selected "version".

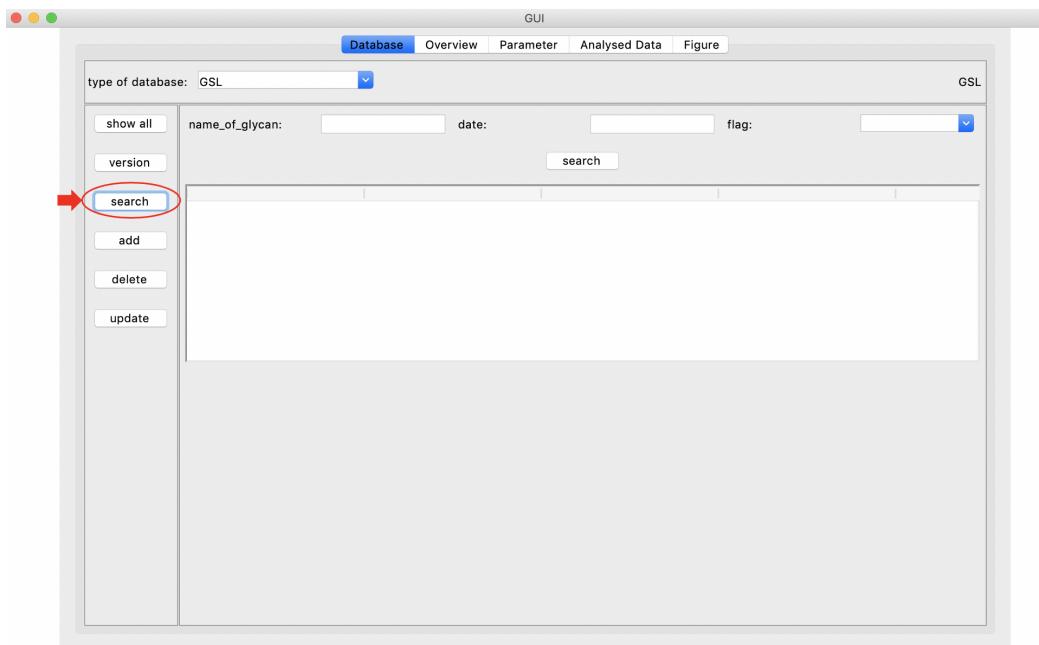
### 1.1.5 Button – “add”

This button aids in the process of entering records into the database. Enter the "name\_of\_glycan" (string type), the corresponding MTU value (floating point type), add date (datetime type), and flag (enumeration type) on the page. As in the Figure 3.7.

### 1.1.6 Button – “delete”

This button is used to delete one or more records from the database.

- a. Query information.



**Figure 1.6:** Function of the "search" button on the "Database" page.

- b. Select a record (click a record) or multiple records (click the left mouse button while hitting the “command” key to achieve multiple selection) based on the query results.
- c. Click the "confirm to delete" button to delete the selected records from the database.

(See in Figure3.8)

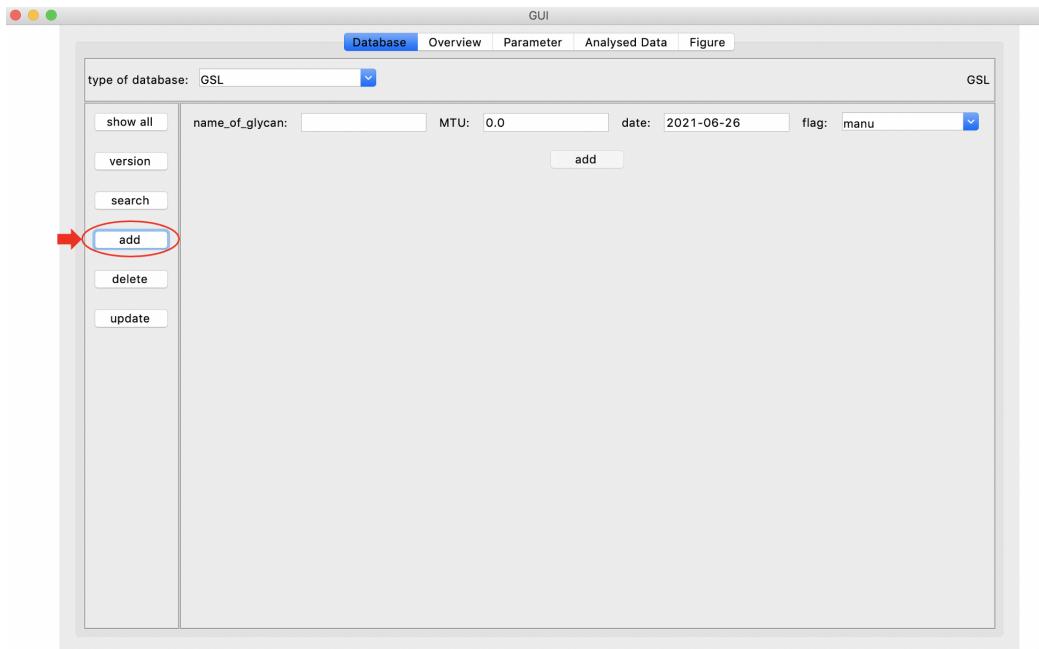
### 1.1.7 Button – “update”

When researchers measure the known glycans in the database from time to time, they need to update the measurement data into the database. This button helps users to add/update records in batches into the database. Figure 3.9 is the basic interface of the update function.

Researchers need to update the measurement data into the database when they measure the known glycans in the database . This button used to add/update records in batches into the database. Figure 3.9 is the basic interface of the update function.

It is divided into two parts. The upper part is for importing data and the lower part is for updating data.

Requirements for the format of the input file: The input file needs to be in \*.txt format, where each line represents a record and each column is a description of the record, which must include

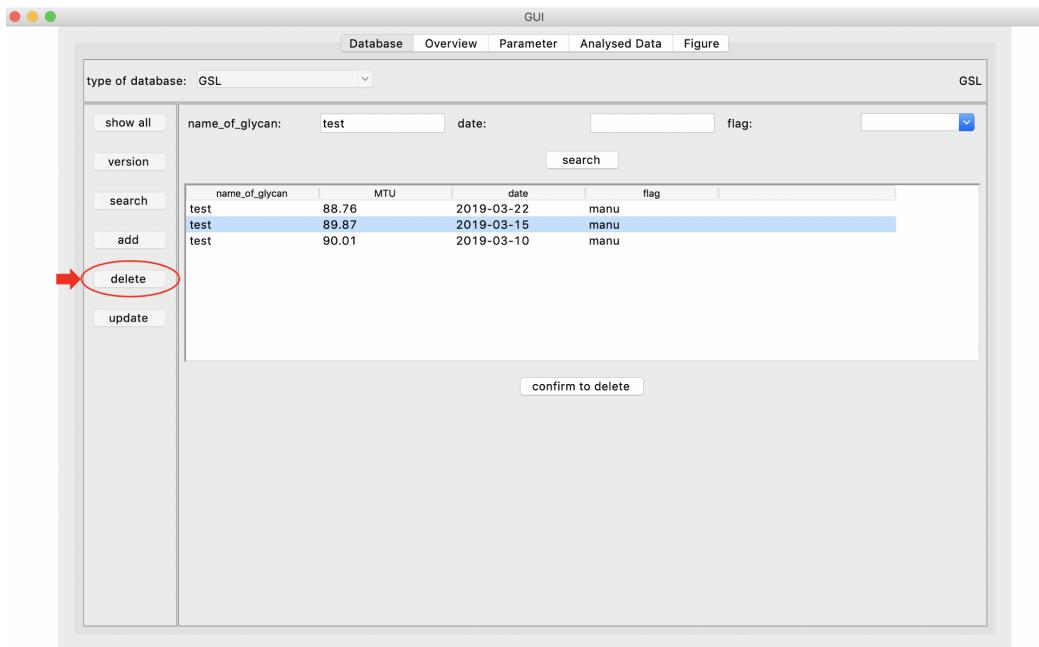


**Figure 1.7:** Function of the "add" button on the "Database" page.

"Dye/Sample Peak", " Sample File Name", "Size", "Height" . "Dye/Sample Peak" indicates the color and serial number of the peek, " Sample File Name" indicates the sample name (glycan's name), "Size" indicates the MTU value (can be empty), "Height" indicates intensity. The sample text is shown in Figure 3.10.

The operations required by the user are as follows,

- a. Click on the "input" button and select a \*.txt file in the computer to import (①),
- b. Select the color of channel in the selection box (②),
- c. Show the selected data (③), and the software will extract the records from the import file that contain the "Size" value and are in the color selected by the user.
- d. Click on the "match" button to match the records in the imported file with the records in the database (④), the specific matching method in the software is as follows,
  - d1. Iterate through every glycan that has appeared in the records extracted according to the color and size values (Name the current glycan as: glycan\_temporary)
  - d2. If glycan\_temporary has a record in the database, then the new\_flag for that glycan is marked as "No"; Find the most recent record of glycan\_temporary in the database (the MTU value of the record is noted as latest\_MTU); Find the size of the



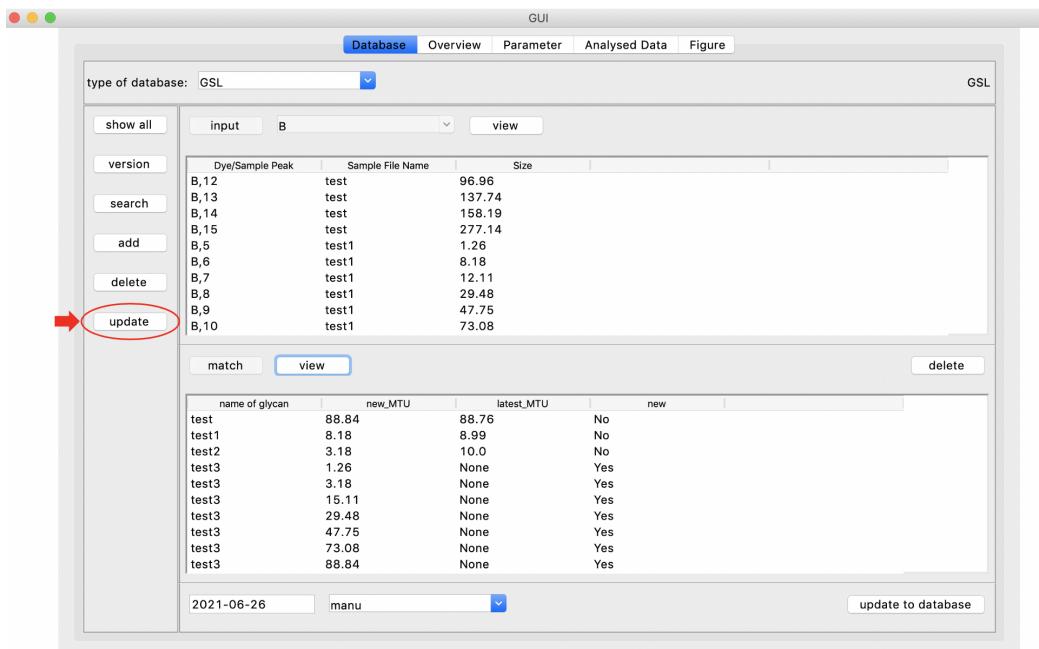
**Figure 1.8:** Function of the "delete" button on the "Database" page.

Deletes one or more records from the database.

glycan\_temporary with the smallest difference from the latest\_MTU and use it as the new\_MTU of the glycan\_temporary; The name, new\_MTU, latest\_MTU and flag of the glycan\_temporary form a new record of the glycan. As in Figure 3.11, record-d2

d3. If glycan\_temporary has no record in the database, the new\_flag of the glycan is marked as "Yes"; Find all the records related to glycan\_temporary among the extracted records and keep them. As shown in Figure 3.11, records-d3

- e. Click the "view" button to show all the new records generated after matching (⑤)
- f. New records are subjectively observed. If the difference between new\_MTU and latest\_MTU in the new record is too obviously, the record can be deleted (by clicking the record, and then clicking the "delete" button). For new glycans users can choose to keep a record and delete the extra records(⑥).
- g. After the delete operation is completed, "date", "flag" need to be confirmed. Finally, click "update to database" to update all records in the table into the database (⑦).



**Figure 1.9:** Function of the "update" button on the "Database" page.  
Updating some records in the database.

## 1.2 Page of Overview

This page handles the import of raw data files as well as the extraction of data from raw data. As discussed in Chapter 2.2, the original data must undergo a data extraction stage in order to retrieve valuable data information for further data analysis. Reading the raw data file, selecting the reference sample, selecting the samples to be analysed, selecting the channel color, selecting the replicate groups, and so on are among the operations provided.

### 1.2.1 Layout of the Page

The "Overview" page is organized into three primary modules, one on the left, one in the middle, and one on the right. As illustrated in Figure 3.12, the three modules are raw data read-in, data extraction, and replicate group selection.

### 1.2.2 Read-in of Raw Data

Select the file to import by clicking on the "input" button, then confirm the selection by clicking on the "ok" button.

The raw data file format is detailed in Chapter 2.1.

Dye/Sample	Peak	Sample	File Name	Size	Height	Area	Data Point
B,1	test	21	183	1420			
B,2	test	2825	23920	1455			
B,3	test	22	188	1527			
B,4	test	181	2835	2898			
B,5	test	1.26	46	589	2951		
B,6	test	3.18	36	572	2980		
B,7	test	15.11	29	458	3009		
B,8	test	29.48	3175	52688	3075		
B,9	test	47.75	142	4107	3652		
B,10	test	73.08	30	709	4034		
B,11	test	88.84	27	438	4179		
B,12	test	96.96	92	2228	4393		
B,13	test	137.74	3428	93001	5053		
B,14	test	158.19	185	5062	5360		
B,15	test	277.14	114	3361	7376		
0,1	test	18	277	2053			
0,2	test	24	369	2107			
0,3	test	26	322	2116			
0,8	test	2.79	62	636	2974		
0,9	test	5.04	105	1554	3008		
0,10	test	16.23	128	1375	3026		
0,11	test	27.29	10	42	3042		
0,12	test	40.38	10	39	3541		
0,13	test	75.03	11	58	3611		
0,14	test	87.61	1002	9347	3650		
0,15	test	94.07	39	269	4049		
0,16	test	108.07	39	269	4049		
0,17	test	164.07	39	269	4049		
G,1	test	18	277	2053			
G,2	test	24	369	2107			
G,3	test	26	322	2116			
G,8	test	2.79	62	636	2974		
G,9	test	5.04	105	1554	3008		
G,10	test	16.23	128	1375	3026		
G,11	test	27.29	10	42	3042		
G,12	test	40.38	10	39	3541		
G,13	test	75.03	11	58	3611		
G,14	test	87.61	1002	9347	3650		
G,15	test	94.07	39	269	4049		
G,16	test	108.07	39	269	4049		
G,17	test	164.07	39	269	4049		
B,1	test1	21	183	1420			
B,2	test1	2825	23920	1455			
B,3	test1	22	188	1527			
B,4	test1	181	2835	2898			
B,5	test1	1.26	46	589	2951		
B,6	test1	8.18	36	572	2980		

**Figure 1.10:** Template of the format of the import file in the update database.

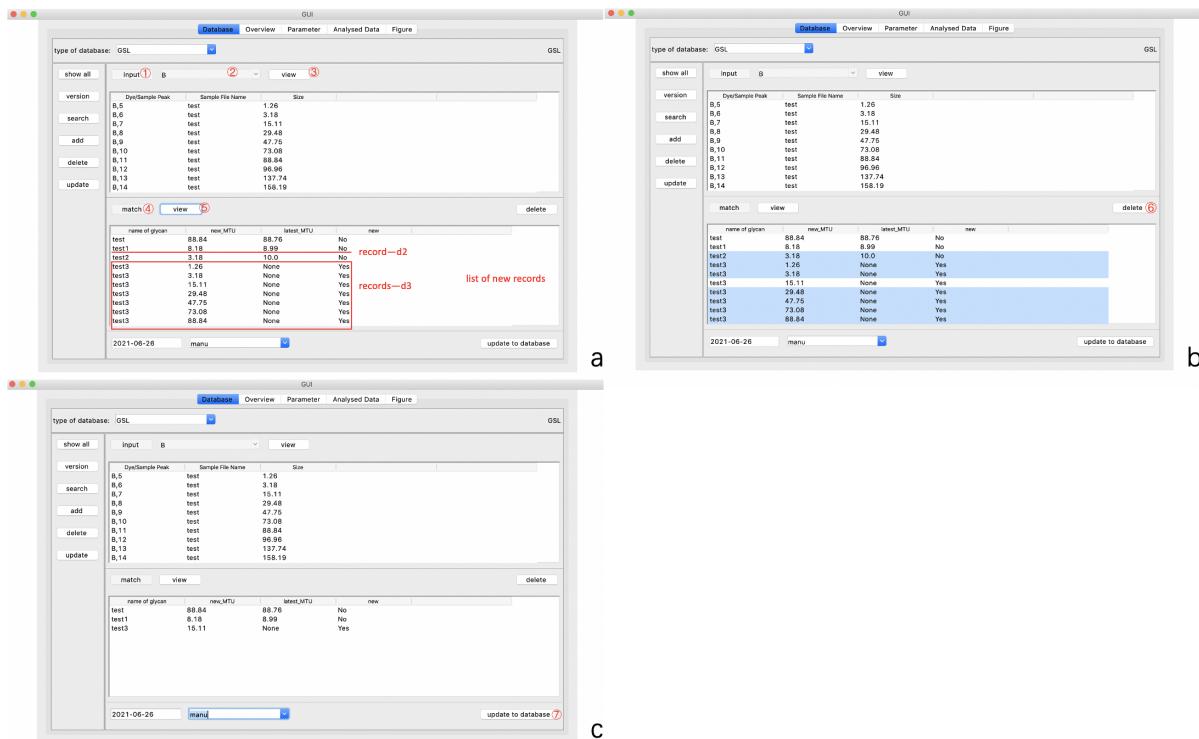
### 1.2.3 Data Extraction

The reference sample, samples for analysis, and channel color should all be selected in the data extraction module, as illustrated in Figure 3.13.

#### a. reference sample

The reference sample is the set of samples used to update the database records during the experiment. The records in the database are the results of the previous measurement at a certain time. In order to better match the MTU in the database when annotating samples in this experiment, researchers typically add several known glycans to the referencewell in the reference sample, (typically three), so that the software can use the differences between the measured and stored values of the MTU of these glycans to update the entire database. Then, the variation between different runs can be effectively minimized.

If there is no reference sample in the experiment, the reference sample can be set to "None," and the update database step in the data processing will not update the database in the future. That is, the updated database is identical to the database for the selected version.



**Figure 1.11:** Diagram of the "update" function on the "Database" page.

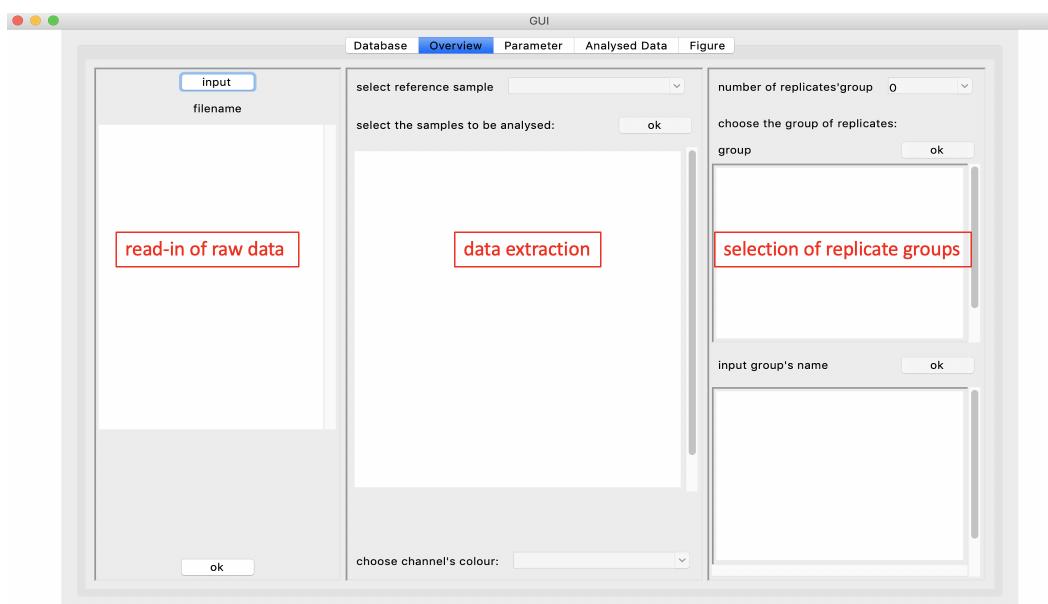
In Figure a, the user imports the data file ①, selects the color of channel ②, shows the extracted data ③, and matches the data in the file with the records in the database ④ and shows the extracted data ⑤. In Figure b, the user filters the matched records and removes the data that do not needed ⑥. In Figure c, the user stores all the remaining records into the database ⑦.

## b. samples for analysis

Generally, the reference sample is only used as the basis for database update and will not be involved in the subsequent data processing. Researchers can also choose the samples they want to compare according to their demands.

## c. color of the channel

In the raw data or electropherograms, each sample has different color channels, for example, B (for blue), O (for orange), and so on. During the experiment, only one channel with a specific color in each sample represents the data information of the experimental object, that is, APTS-labelled glycans. The other color channels represent other standards added due to some experimental needs. For example, in order to calibrate the migration times observed in electropherograms, Dye Size Standard will be added to the sample, and there will be a color channel in the sample that represents the measured information of Dye Size Standard.



**Figure 1.12:** Layout of "Overview" page.

The page is organized into three modules, read-in of raw data, data extraction, and selection of replicate groups.

Therefore, only the channel information of the color corresponding to the APTS-labeled glycan is utilized.

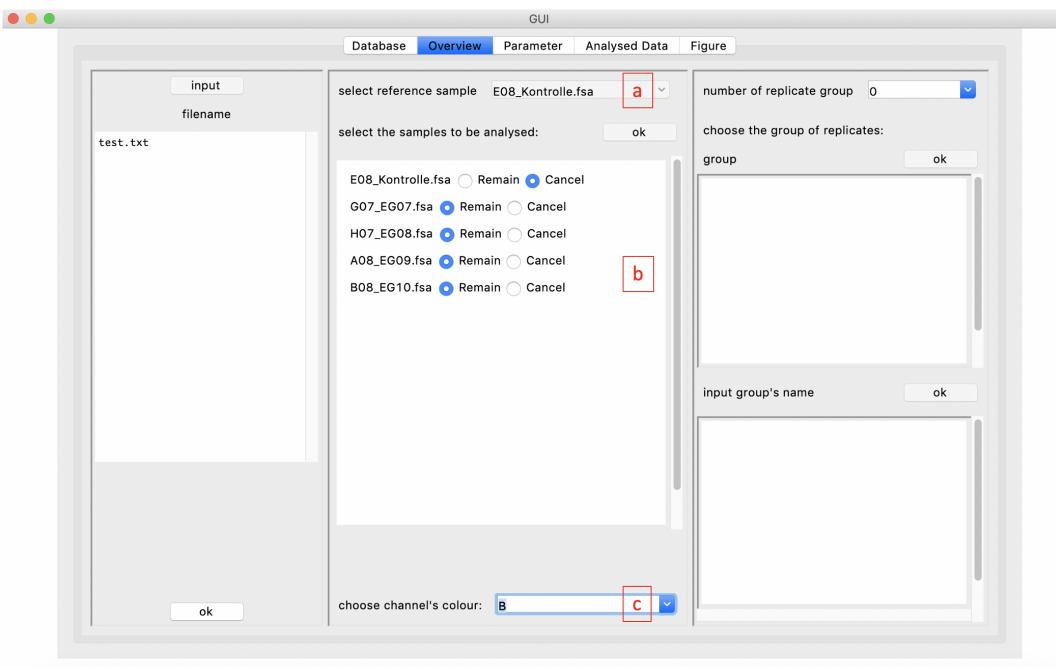
#### 1.2.4 Selection of Replicate Groups

In biological settings, different conditions are often compared with each other. Usually the difference is small, and the researcher needs replicates for each condition to see whether the differences are statistically significant. The best type of replicates is the so-called "biological replicates", in which different biological samples with the same condition are analysed individually.

Therefore, in order to compare the differences under different conditions more accurately, while comparing different conditions, replicate groups are set. In each group, the same conditions are repeated, and different groups represent different conditions. In this way, while comparing different groups, the mean and variance of replicates in the group can be calculated.

The specific selection process is as follows:

- a. Select the number of replicate groups,
- b. Select the samples in each group, and click the "ok" button to confirm after each group is selected,



**Figure 1.13:** Diagram of data extraction in the "Overview" page.

- Select the reference sample, b. Select the samples to be analysed, c. Select the color of the channel.
- After all groups are selected, name each replicate group, and click the "ok" button to confirm after entering all names.

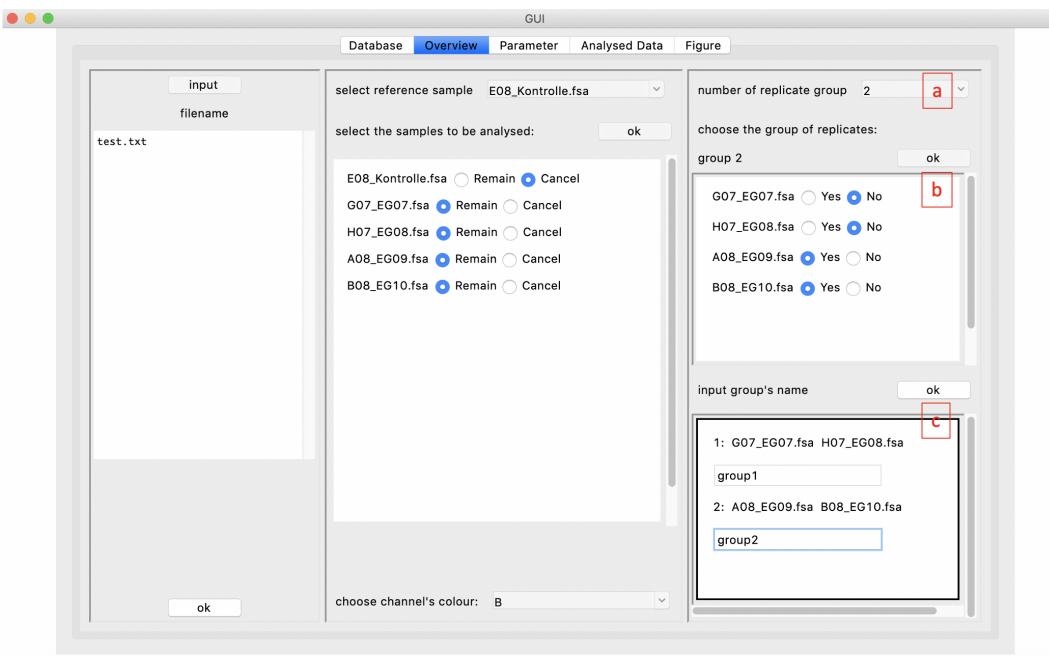
As in the Figure 3.14.

## 1.3 Page of Parameter

Some settings used in data processing will be confirmed on this page. It is mainly concerned with the selection of database and version, glycans in referencewell, analysis methods, etc. The page layout is shown in the Figure 3.15.

### 1.3.1 Selection of Database and Version

The analysis object must be checked before each measurement. Currently, the software performs GSL-derived glycan analysis on samples (a database has been created), which means that samples are annotated with GSL-derived glycans from the database. The software also creates a database for N-glycans and GAGs, although no data records are yet available. However, multiple versions of information are frequently maintained in the database due to the irregular measurement of the known glycans MTU value in the database. For example, the GSL-derived glycan database stores



**Figure 1.14:** The diagram of the data extraction operation in the "Overview" page.

a. Select the number of replicate groups, b. Select the samples for each replicate groups, c. name each sample.

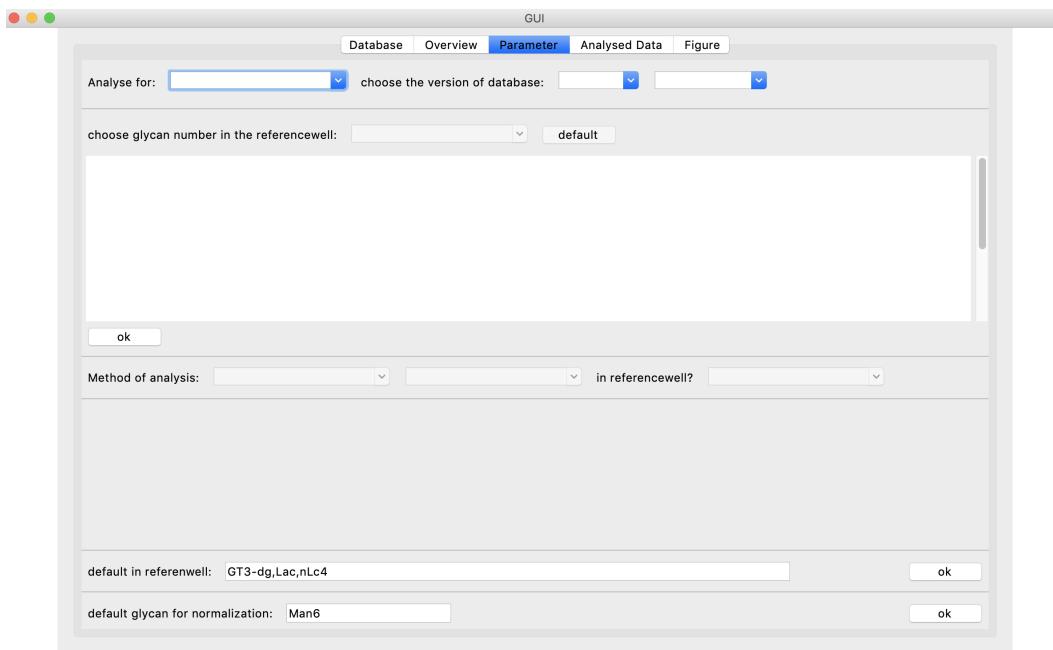
data measured at three times of 2019, 2020.02, 18, and 2021.01.13. So, select a version of the recorded data by first selecting the year and then the appropriate month-day (Figure 3.16). Then update the database based on the data of this version (Chapter 2.3), and the updated database can be used for the final sample annotating (Chapter 2.7).

### 1.3.2 Selection of Glycans in Referencewell

The reference sample described in Chapter 3.2.2 is the sample set added to update the database records. Since it is not possible to measure all the known glycans in the database at the same time when measuring samples, the researchers chose a few typical glycans and added them to the referencewell as a sample that was measured alongside other samples. Use the difference between the measured value and the stored value of the representative glycans to update the remaining known glycans in the database. The presence of glycans in the referencewell should be verified here.

Specific process (Figure 3.17):

- Select the number of glycans in the referencewell,
- Select glycans in the referencewell,
- Click the "ok" button.



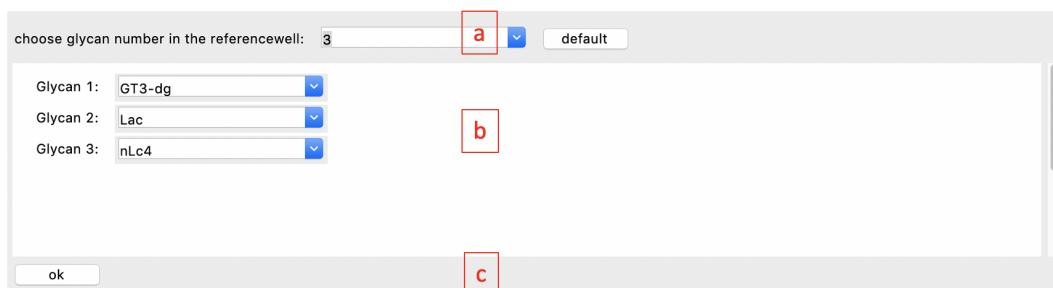
**Figure 1.15:** Layout of the "Parameter" page



**Figure 1.16:** In "Parameter" page, select the database and version

Since researchers regularly use some glycans in the experiments, So the “default” button is set here, click the "default" button, and the software will automatically select the number of glycans in the referencewell and glycans in the referencewell that have been set in advance.

The default values for glycans in the referencewell can also be set in the "Parameter" page. As shown in the Figure 3.18, enter the commonly used names of glycans in the referencewell into the input box, names are separated by commas, and click the “ok” button, the default settings can be saved.



**Figure 1.17:** In the "Parameter" page, selection of glycans in referencewell.  
a. Select the number of glycans in the referencewell, b. Select the glycan type, c. Confirm

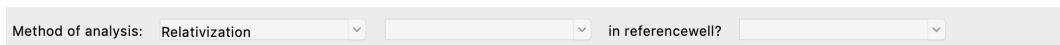


**Figure 1.18:** Set the default values for glycans in the referencewell in the Parameter screen.

### 1.3.3 Selection of Analysis Method

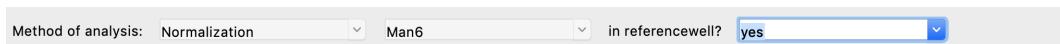
It is essential to ascertain whether a relative or a normalized analysis should be performed during data processing (Chapter 2.3).

If "Relativization" has been chosen, there is no need to make subsequent selections. As shown in the Figure 3.19.



**Figure 1.19:** Select "Relativization" as the analysis method in the "Parameter" screen.

If "Normalizing" is chosen, the standard glycan for normalization must also be chosen (the specific explanation is given in Chapter 2.3). Since the standard glycan of normalization is also added into the reference sample, the MTU of this measurement and the MTU value in the database are known. The standard glycan can also be chosen as the reference glycan and participate in the "Update the database" procedure (Figure 3.20).



**Figure 1.20:** Select "Normalization" as the analysis method in the "Parameter" page.

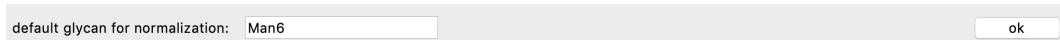
Because a certain glycan is frequently used as a standard in experiments, the normalization standard glycan can also be set to its default value. Enter the name of the standard glycan of normalization into the input box and click the "ok" button to set the default value, as illustrated in Figure 3.21.

## 1.4 Page of Analysed Data

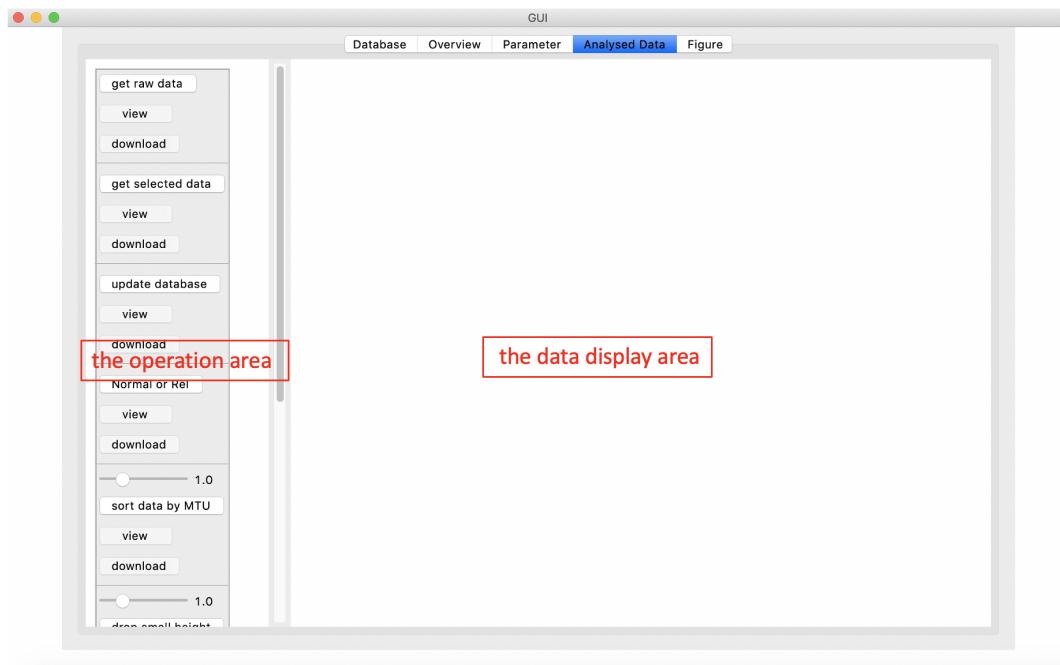
The data processing procedures will be completed step by step on this page of the software, following the approach shown in Figure 2.1, with the exact algorithm in Chapter 2.

The page is separated into two sections: the operation area on the left and the data display area on the right. Figure 3.22 depicts the layout of this page.

In each step, the button with the step's name is used to perform data processing operations; the "view" button is used to display the results; the "download" button is used to store the processed data in an excel file.



**Figure 1.21:** Set the default value for the standard glycan of normalization in the “Parameter” page.



**Figure 1.22:** Layout of the "Analysed Data" page

There are also other parameters that need to be set in the steps "sort data based on MTUs," "drop small height," and "annotate."

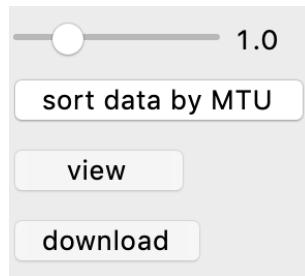
#### *Step of “sort data based on MTUs”*

"Sort data based on MTUs" is the so-called "aligning migration time". In order to compare different samples, the peaks with similar MTU values in different samples are grouped together. The average value of MTUs in this group is used as the mean migration time unit of this group.

There is a criterion for discriminating between similar MTUs. If the difference between two MTUs is less than a threshold value, they are considered similar. This threshold can be adjusted. See Figure 3.23. The range of the scale bar is 0.0-5.0, the precision is 0.1, the step is 0.1, and the default value is 1.0.

#### *Step of “drop small height”*

Deleting smaller peaks. The determination of which values are smaller is determined by the threshold value. The threshold can be adjusted. The range of the scale bar is 0.0-5.0, the precision is 0.1, the step is 0.1, and the default value is 1.0. For more information on how to drop small height ,



**Figure 1.23:** The operation box for “sort data based on MTUs” in the "Analysed Data" page

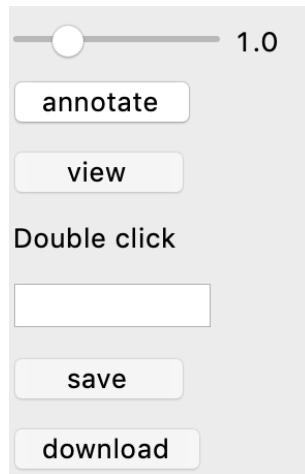
see Chapter 2.6.

#### *Step of “annotate”*

Following the above steps, the samples are annotated with known glycan structures from the updated database. The thresholds for the annotation process should be defined here (Chapter 2.6). The known glycan structure will be assigned to a peak if the difference between the mean MTU of a peak and the MTU of a known glycan in the database is less than this threshold. The range of the scale bar is 0.0-5.0, the precision is 0.1, the step is 0.1, and the default value is 1.0.

For some studies, it is impossible for certain specific glycans to be present in the sample. In this case, the annotation needs to be changed manually.

If the annotation needs to be artificially modified after the final data processing, double-click the row with the left mouse button and re-enter the changed name in the input box shown in Figure 3.24. If an annotation needs to be deleted, do not need to enter anything in the input box. Click the "save" button to save the annotation, and then click the "view" button to see the changed annotation.



**Figure 1.24:** The operation box for “annotate” in the "Analysed Data" page

- ▶ Scale Bar – Set the threshold value
- ▶ Button – “annotate”: Processing the data
- ▶ Button – “view”: Display the results of processing
- ▶ Input Box – Enter the annotation to be changed
- ▶ Button – “save”: Save the contents of the input box
- ▶ Button – “download”: Store data in excel

## 1.5 Page of Figure

This page generates bar graphs of the data processing results. There are two sorts of bar graphs that can be created: bar graphs for samples and bar graphs for replicate groups.

The basic page layout consists of two parts: the operation area and the figure display area. As shown in Figure 3.25.

In the operation area, researchers can change some attributes of the bar graph according to their needs. For example, they can set whether or not the annotation is required; set the rotation angle of the annotation in order to prevent the glycan’s names from overlapping ; change the color of the bar graph according to the their preference. The range of MTUs, the length of the x-axis and y-axis, and other parameters can also be changed to observe more detailed information.

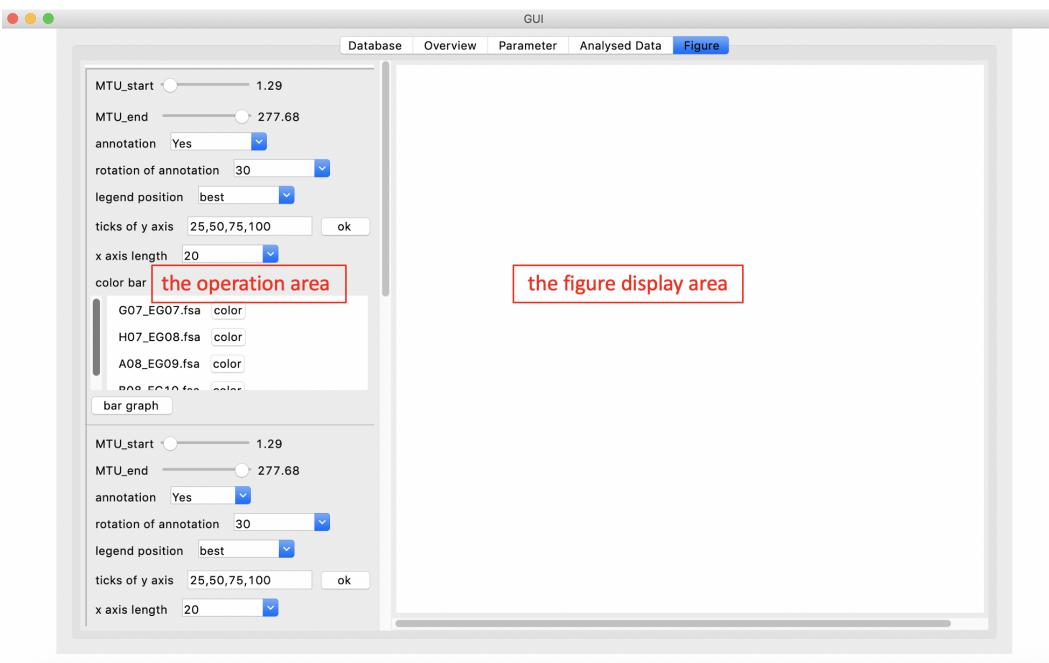
### 1.5.1 Bar Graph for Samples

"Bar Graph for Samples" shows the difference of nRFUs or RI between different samples annotated with known glycans.

The operating area of the bar graph for samples is shown in Figure 3.26.

Selections for users:

- ▶ MTU range. Note: MTU\_end > MTU\_start, scale bar range [min\_MTU, max\_MTU]
- ▶ Whether to display annotation on the bar graph. Candidate selections: [Yes, No]
- ▶ If annotation is displayed, the rotation angle of annotation on the bar graph. Candidate selections: [0, 20, 30, 45, 60, 70, 90].



**Figure 1.25:** Layout of the "Figure" page.

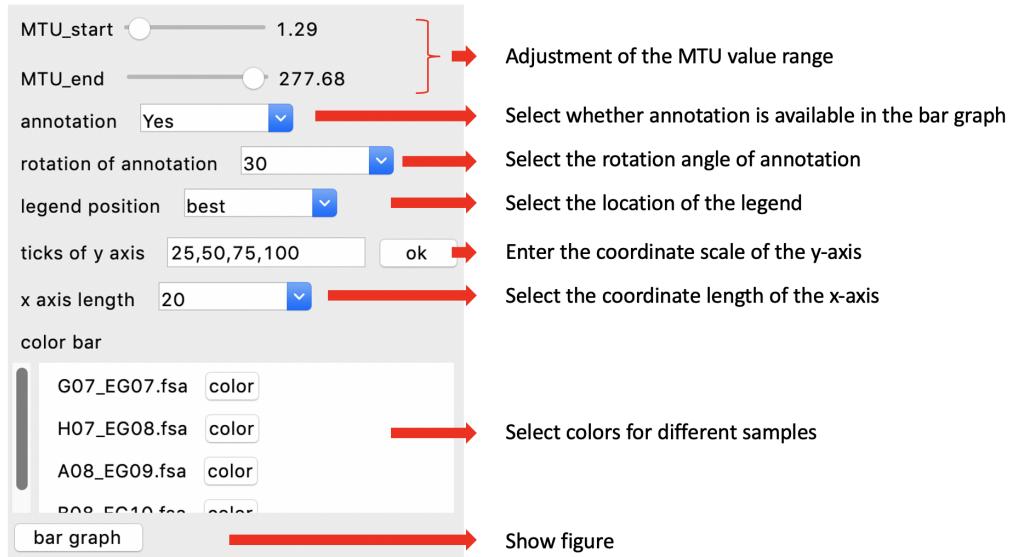
On the left is the operation area, and on the right is the figure display area.

- ▶ The position of Legend. Candidate selections: ["none", "best", "upper right", "upper center", "upper left", "center right", "center left", "lower right", "lower center", "lower left", "outer upper", "outer lower"]
- ▶ The coordinate scale of the y-axis. Enter the y-axis scale into the input box, separated by commas.
- ▶ The length of the x-axis. Candidate selections: ["normal", "20", "30", "40", "50"]
- ▶ Color for each sample's bar. If not selected, the bar graph will be generated by using the system default colors. If the user needs to customize the colors, the selection needs to be made for all samples.

The above parameters can be adjusted, and after each selection, the "bar graph" button needs to be clicked to generate a new bar graph.

### 1.5.2 Bar Graph for Replicate Groups

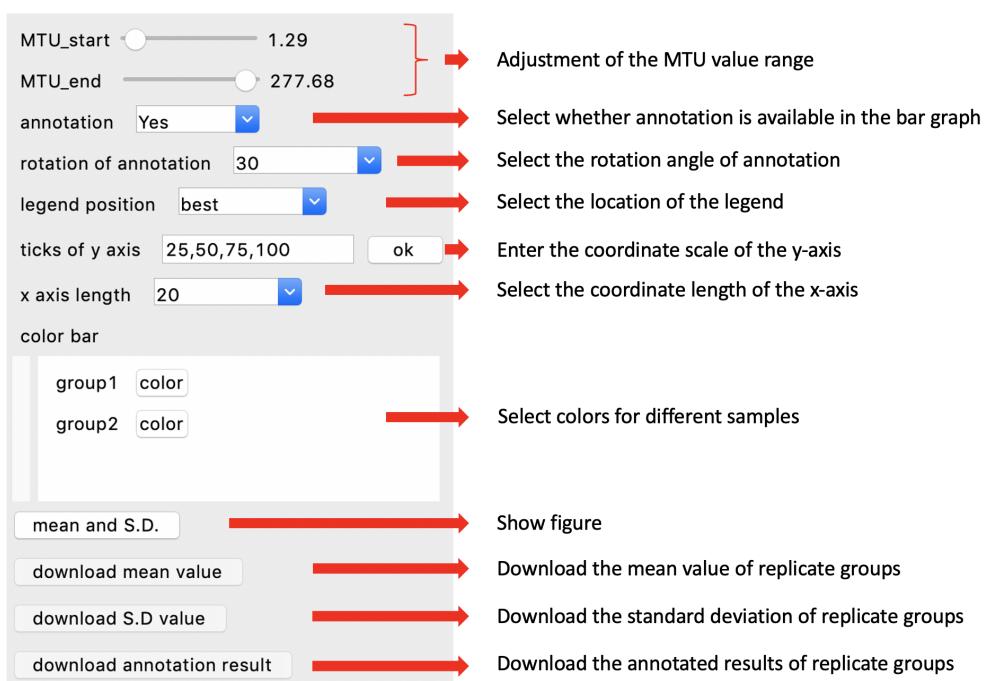
The "Bar Graph for Replicate Groups" calculates the mean value of nRFUs or RI of samples in each group to demonstrate the difference between groups. The standard deviations of the replicate samples in each group are also calculated and displayed in the bar graph to observe the experiment's accuracy.



**Figure 1.26:** Bar graph for samples with detailed explanation of the operating area functions

Figure 3.27 depicts the operation area of the bar graph for replicate groups. In addition to the same functionality as "Bar Graph for Samples," this area downloads the duplicate groups' mean and standard deviation.

- ▶ Button – “download mean value”: Download the mean value of replicate groups
- ▶ Button – “download S.D. value”: Download the standard deviation of replicate groups
- ▶ Button – “download annotation result”: Download the annotated result of replicate groups



**Figure 1.27:** Bar graph for replicate groups with detailed explanation of the functions of the operating area

## 2 Functions and Results of the Software

In this chapter, a dataset will be utilized to demonstrate the software's functionality.

### 2.1 Data Preparation

This experiment uses xCGE-LIF to analyse glycosphingolipid(GSL) glycosylation to identify the cell surface markers of human induced pluripotent stem cells (iPSCs).

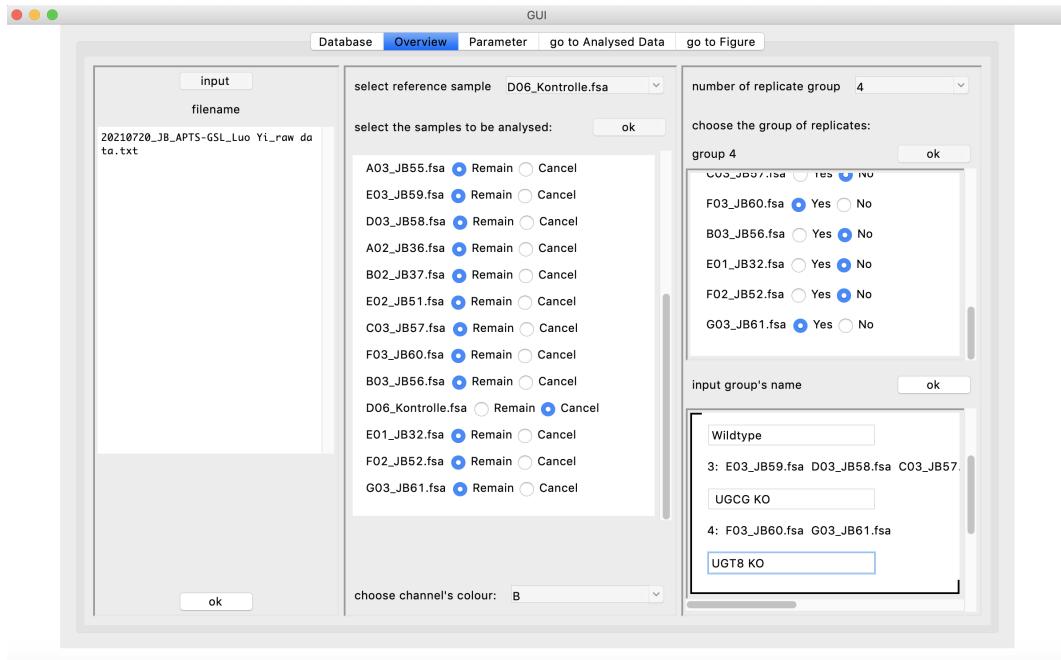
All samples come from four different conditional factors, they are control (no cells, only enzyme was used), wildtype iPSC Phoenix, UGCG KO in iPSC Phoenix, and UGT8 KO in iPSC Phoenix. Among them, the "control" samples are just to show that during preparation process (GSL extraction out of cell pellet, CGase digestion, APTS labeling, HILIC purification) some GSL are incorporated, so these signals can be regarded as contaminants. "wildtype iPSC Phoenix" is a wildtype cell line. "UGCG KO in iPSC Phoenix" is one cell line in which the enzyme UGCG is knocked out by CRISPR/Cas9, "UGT8 KO in iPSC Phoenix" is one cell line in which the enzyme UGT8 is knocked out. Both Knock Outs dramatically influence glycosphingolipid synthesis.

Biological replicates were used to measure variation in the experiment so that statistical tests can be applied to evaluate differences between the tested conditions. The samples were grouped as the following: "control" group has JB28, JB29, JB30, JB51, JB52; "Wildtype iPSC Phoenix" group has JB37, JB38, JB55, JB56; "UGCG KO in iPSC Phoenix" group has JB57, JB58, JB59; "UGT8 KO in iPSC Phoenix" group has JB60, JB61.

In addition to the samples mentioned above, there is a set of samples that includes a reference well, which includes glycan standards for a database update and an internal standard for normalization, was measured simultaneously. GT3-dg, Lac, and nLc4 are the glycans utilized to update the database, while Man6 is the internal standard.

## 2.2 Selection of Parameters

Select the corresponding options on the Overview page and Parameter page according to the data information. As shown in the Figure 2.1 and 2.2.



**Figure 2.1:** The parameter setting of the Overview page

If the data requires relative analysis, select “Relativization” in Method of analysis.

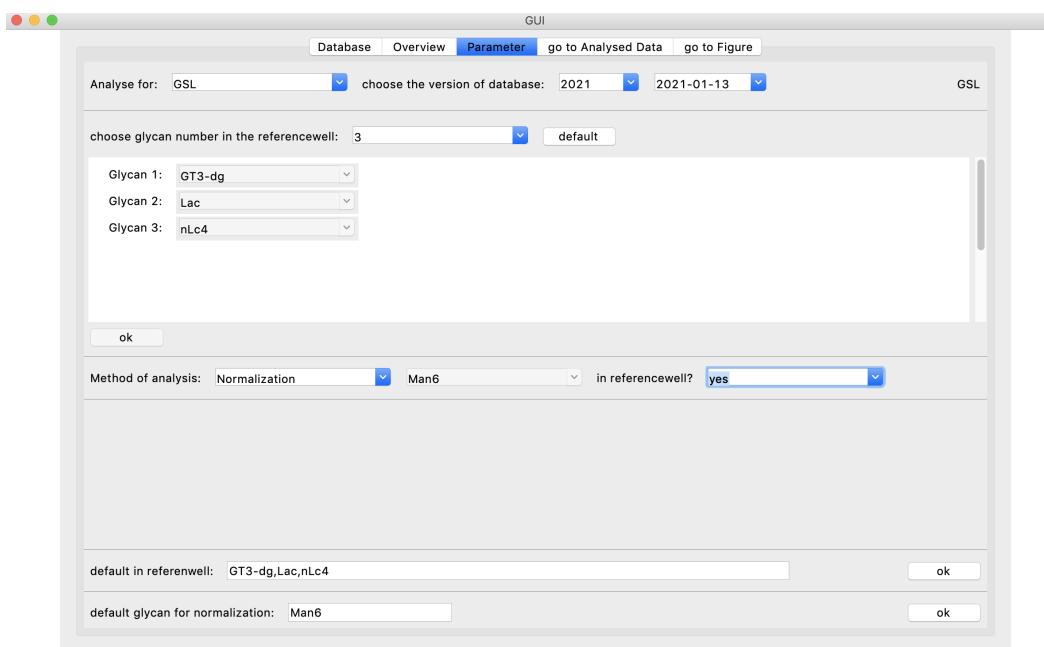
Set the threshold for "sort data by MTU" to 1; the threshold for "drop small height" to 2; and the threshold for "annotate" to 1 in process of data analysis. As shown in the Figure 2.3.

## 2.3 Data analysis between samples

### Purpose

In the glycoprofiling, the xCGE-LIF technology can be used to analyse the components of a single sample, i.e., use the known glycans in the database to annotate the sample components by matching the migration time of APTS-labelled glycans detected in the sample.

As for comparative glycomics, the glycome is influenced by both genetic and environmental factors. Set different influencing factors to different samples, and use xCGE-LIF technology to detect these samples together, the difference in the influence of different factors on glycome can be compared. Depending on the equipment of the used sequencer, up to 96 samples can be measured



**Figure 2.2:** The parameter setting of the Parameter page

simultaneously in one xCGE-LIF run. Thereby sample specific migration times and intensities of APTS-labeled glycans are detected. By processing the data, as described in Chapter 2, not only the composition of each sample individually is defined but also a comparison between many different factors can be calculated.

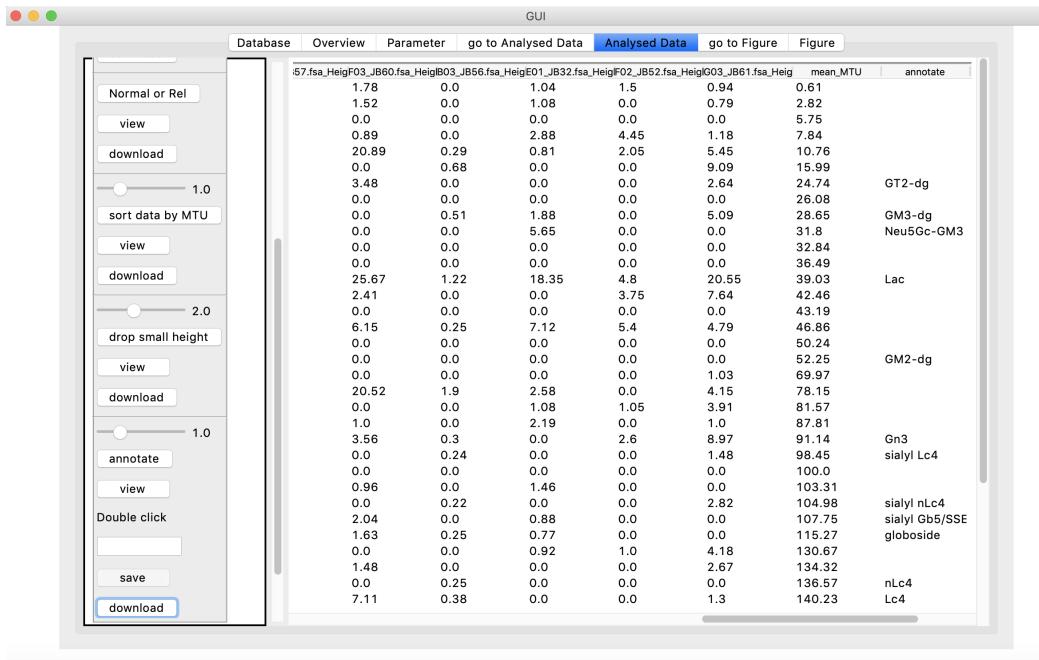
### 2.3.1 Variation of normalized signal intensity with aligned MTU

#### *Purpose*

For inter-sample quantitative comparison of signal intensities, a defined amount of APTS labelled internal standard will be spiked in to each sample. The migration time of the internal standard should be different to that of the target derived glycans and therefore the spikein of internal standard does not interfere with derived glycan analysis. Thus, the obtained signal of internal standard could be used for normalization of peak intensities. The height of the internal standard is set to 1 normalized relative fluorescence units (nRFUs). By putting together standardised intensities from different samples at the same MTUs, the relative signal intensities of individual peaks in different samples could be visualized (quantitative comparison of signal intensities).

#### *Output*

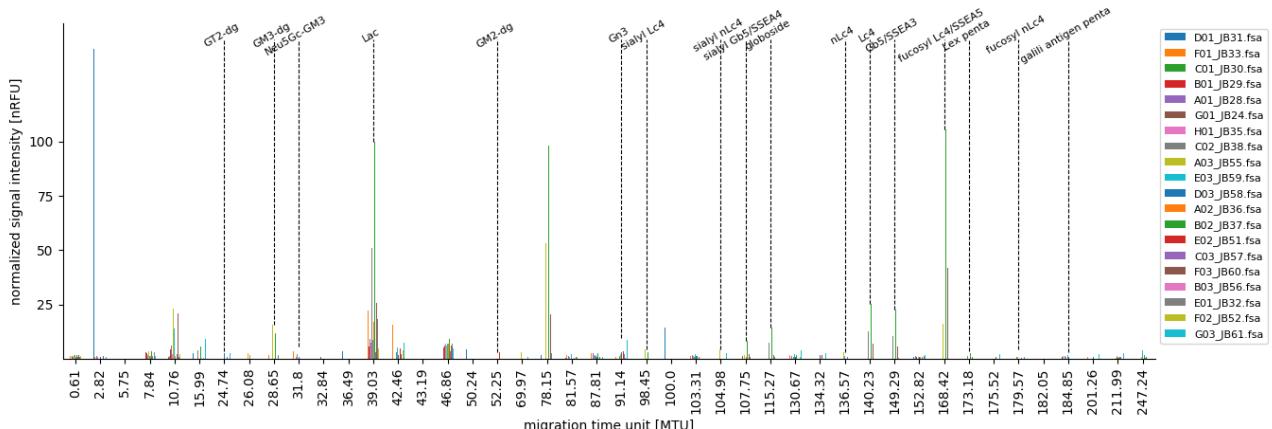
The software has two outputs for data analysis. **One** is a file in excel format, which contains the data annotated by known glycans in the database after normalized analysis of each group of



**Figure 2.3:** The parameter setting of the Analysed-data page for the comparison of normalized signal intensity between samples

samples; **the other** is a bar graph, which visualizes the data. It can more intuitively reflect the differences between samples.

- Bar graph: See Figure 2.4
- Excel file with analysed data: See Figure 2.5



**Figure 2.4:** The bar graph for the comparison of normalized signal intensity between samples

	K	L	M	N	O	P	Q	R	S	T	U	V	W													
1	JB59.fsa	He	JB58.fsa	He	JB36.fsa	He	JB37.fsa	He	JB51.fsa	He	JB57.fsa	He	JB60.fsa	He	JB56.fsa	He	JB32.fsa	He	JB52.fsa	He	JB61.fsa	He	mean	MTU	annotate	
2	1.6	1.24	0.77	1.6	2	1	1.78	0	1.04	1.5	0.94	0.61														
3	0	1.09	0	0	0	0.78	1.52	0	1.08	0	0.79	2.82														
4	0	0	0	0	0	0	0	0	0	0	0	5.75														
5	1.44	0	1.25	3.35	8.33	1.28	0.89	0	2.88	4.45	1.18	7.84														
6	14.04	4.79	1.15	0	13.05	2.3	20.89	0.29	0.81	2.05	5.45	10.76														
7	0	0	5.7	0	0	0	0.68	0	0	0	9.09	15.99														
8	0	2.47	0	0	0	0.76	3.48	0	0	0	2.64	24.74	GT2-dg													
9	0	0	1.92	0	0	0	0	0	0	0	0	26.08														
10	0	0	0	11.65	0	0	0	0.51	1.88	0	5.09	28.65	GM3-dg													
11	0	0.79	0	0	0	0.57	0	0	5.65	0	0	31.8	Neu5Gc-GM3													
12	0	0	0	3.8	0	0	0	0	0	0	0	32.84														
13	0	0	0	0	0	0	0	0	0	0	0	36.49														
14	8.64	3.82	17.23	99.8	3.33	3.2	25.67	1.22	18.35	4.8	20.55	39.03	Lac													
15	5.36	1.71	0	0	4.67	0.93	2.41	0	0	3.75	7.64	42.46														
16	0	0	0	0	0	0	0	0	0	0	0	43.19														
17	7.08	4.97	7.17	9	4.62	3.33	6.15	0.25	7.12	5.4	4.79	46.86														
18	0	0	0	0	0	0	0	0	0	0	0	50.24														
19	0	0	0	0	3.29	0	0	0	0	0	0	52.25	GM2-dg													
20	0	0	0	0	0	0	0	0	0	0	0	69.97														
21	0	0	0	98.1	0	0	20.52	1.9	2.58	0	4.15	78.15														
22	2.4	0	0.83	0	0	0.54	0	0	1.08	1.05	3.91	81.57														
23	1.88	0.79	1.69	2.75	0	0.69	1	0	2.19	0	1	87.81														
24	9.8	3.24	0	0	3.33	2.26	3.56	0.3	0	2.6	8.97	91.14	Gn3													
25	0	0	2.95	0	0	0	0.24	0	0	1.48	98.45	sialyl Lc4														
26	0	0	0	0	0	0	0	0	0	0	0	100														
27	2.36	1.21	1.33	1.4	0	0.8	0.96	0	1.46	0	0	103.31														
28	0	0	0	0	0	0	0.22	0	0	2.82	104.98	sialyl nLc4														
29	0	0	0.9	8.5	0	0	2.04	0	0.88	0	0	107.75	sialyl Gb5/SSEA4													
30	0	0	0	14.45	0	0	1.63	0.25	0.77	0	0	115.27	globoside													
31	2.28	0.85	0.58	1.85	0	0.67	0	0	0.92	1	4.18	130.67														
32	3.04	1.76	0	0	1.95	0	1.48	0	0	0	2.67	134.32														
33	1.32	0	0	0	0	0.7	0	0.25	0	0	0	136.57	nLc4													
34	0	0	0	25.25	0	0	7.11	0.38	0	0	1.3	140.23	Lc4													
35	0	0	0	22.55	0	0	5.59	0.33	0	0	0	149.29	Gb5/SSEA3													
36	2.32	1.03	0.94	0	0	0.78	0	0	1.27	0	1.73	152.82														
37	0	0	0	105.65	0	0	41.7	1	0	0	4.55	168.42	fucosyl Lc4/SSEA5													
38	0	0	0	2.75	0	0	1.04	0	0	0	0	173.18	Lex penta													
39	4.76	0	0	0.95	0.63	0	0	0	0	0	2.27	175.52														
40	4.36	0	0	0	0	0.54	0	0	0	0	0.82	179.57	fucosyl nLc4													
41	0	0	0	0	0	0	0	0	0	0	0	182.05														
42	1.88	0.91	0.44	0	0	0.54	0	0	0	0	3.12	184.85	galili antigen penta													
43	2.32	1.12	0.65	0	0	0.89	0	0	0	0	2.15	201.26														
44	4.88	1.06	0.48	0	0.95	0.76	1.26	0	0	0	2.64	211.99														
45	3.92	0.71	0	1.6	0	0.48	0.96	0	0	0	2.03	247.24														

**Figure 2.5:** Excel file for the comparison of normalized signal intensity between samples

### 2.3.2 Variation of relative signal intensity with aligned MTU

#### Purpose

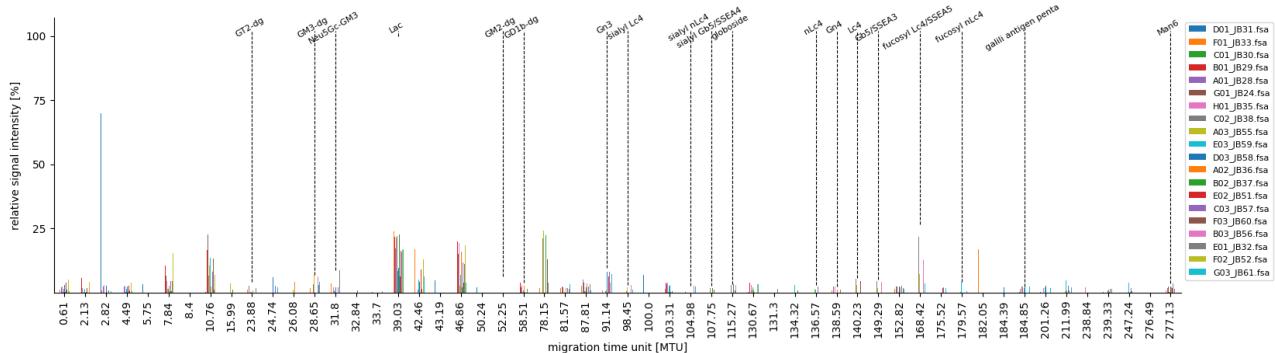
For determination of relative signal intensities (RI) within an electropherogram, the heights of all peaks in the specified range of MTUs are summed up and the height of individual peaks are calculated as percentage referring to that summarized peaks heights. These relative signal intensities of individual peaks could then be compared between different samples (quantitative comparison of relative peak intensities).

#### Output

The software has two outputs for data analysis. **One** is a file in excel format, which contains the data annotated by known glycans in the database after relative analysis of each group of samples; **the other** is a bar graph, which visualizes the data. It can more intuitively reflect the differences between samples.

- Bar graph: See Figure 2.6

- Excel file with analysed data: See Figure 2.7



**Figure 2.6:** The bar graph for the comparison of relative signal intensity between samples

## 2.4 Data analysis between replicate groups

### *Purpose*

In biological settings, different conditions are often compared with each other (for example, no tumor tissue and tumor tissue, health and disease, or wild type and gene knockout...). Usually the difference is small and the researcher needs replicates for each condition to see whether the differences are statistically significant. The best type of replicates is the so-called "biological replicates", in which different biological samples with the same condition are analysed individually, not just multiple analyses of the same sample (this would be "technical replicates"). For example, it is better to compare 5 people with tumors to 5 healthy people, rather than just analyzing one person's tumor 5 times.

For replicate groups, not only the quantitative comparison of signal intensities or the quantitative comparison of relative peak intensities between the groups, but also the mean and standard deviation of replicates in each group must be considered.

### 2.4.1 Variation of normalized signal intensity with aligned MTU

### *Purpose*

As with Data analysis between samples, data analysis between replicate groups also requires the study of differences between groups of normalized signal intensity with aligned MTU.

### *Output*

	K	L	M	N	O	P	Q	R	S	T	U	V	W	
	JB59.fsa	JB58.fsa	JB36.fsa	JB37.fsa	JB51.fsa	JB57.fsa	JB60.fsa	JB56.fsa	JB32.fsa	JB52.fsa	JB61.fsa	mean_MTU	annotate	
1														
2	1.57	3.11	1.72	0.36	3.85	3.58	1.13	0	1.62	5.17	0.78	0.61		
3	1.53	0	0	0.24	1.92	0	0	0	0	4.31	0	2.13		
4	0	2.74	0	0	0	2.78	0.97	0	1.68	0	0.65	2.82		
5	1.53	2.59	1.44	0.24	2.66	2.98	0.87	0	1.62	3.97	0.78	4.49		
6	0	0	0	0	0	0	0	0	0	0	0	5.75		
7	1.41	0	2.79	0.76	16.04	4.57	0.57	0	4.5	15.34	0.98	7.84		
8	0	4.07	0	0	0	0	0	0	0	0	0.93	8.4		
9	13.78	12.07	2.56	0	25.11	8.21	13.31	3.74	1.26	7.07	4.52	10.76		
10	0	0	0	1.3	0	0	0	8.67	0	0	7.54	15.99		
11	1.65	0	0.98	0.26	0	0	0	0	1.74	0	0	23.88	GT2-dg	
12	0	6.22	0	0	0	2.72	2.22	0	0	0	0	2.19	24.74	
13	0	0	4.28	0	0	0	0	0	0	0	0	0	26.08	
14	0	0	0	2.66	0	0	0	6.5	2.94	0	4.22	28.65	GM3-dg	
15	0	2	0	0	0	2.05	0	0	8.82	0	0	31.8	Neu5Gc-GM3	
16	0	0	0	0.87	0	0	0	0	0	0	0	32.84		
17	0	0	2.88	0	0	0	0	0	0	0	0	0.5	33.7	
18	8.48	9.63	38.47	22.75	6.42	11.46	16.36	15.67	28.63	16.55	17.04	39.03	Lac	
19	5.26	4.3	0	0	8.98	3.31	1.53	0	0	12.93	6.33	42.46		
20	0	0	0	0	0	0	0	0	0	0	0	43.19		
21	6.95	12.52	16	2.05	8.89	11.92	3.92	3.25	11.1	18.62	3.97	46.86		
22	0	0	0	0	0	0	0	0	0	0	0	50.24		
23	0	0	0	0	6.32	0	0	0	0	0	0	0	52.25	GM2-dg
24	0.94	0	2.65	0.24	0	0	0	0	1.56	0	0.93	58.51	GD1b-dg	
25	0	0	0	22.36	0	0	13.08	24.33	4.02	0	3.44	78.15		
26	2.35	0	1.86	0	0	1.92	0	0	1.68	3.62	3.24	81.57		
27	1.84	2	3.77	0.63	0	2.45	0.64	0	3.42	0	0.83	87.81		
28	9.62	8.15	0	0	6.42	8.08	2.27	3.84	0	8.97	7.44	91.14	Gn3	
29	0	0	0	0.67	0	0	0	3.05	0	0	1.23	98.45	sialyl Lc4	
30	0	0	0	0	0	0	0	0	0	0	0	100		
31	2.32	3.04	2.98	0.32	0	2.85	0.61	0	2.28	0	0	103.31		
32	0	0	0	0	0	0	0	2.76	0	0	2.34	104.98	sialyl nLc4	
33	0	0	2	1.94	0	0	1.3	0	1.38	0	0	107.75	sialyl Gb5/SSEA4	
34	0	0	0	3.29	0	0	1.04	3.15	1.2	0	0	115.27	globoside	
35	2.24	2.15	1.3	0.42	0	2.38	0	0	1.44	3.45	3.47	130.67		
36	0	0	0	0	0	0	0	0	1.56	0	0	131.3		
37	2.98	4.44	0	0	3.76	0	0.94	0	0	0	2.21	134.32		
38	1.3	0	0	0	0	2.52	0	3.15	0	0	0	136.57	nLc4	
39	0	0	1.26	0	0	0	0	0	1.26	0	0	138.59	Gn4	
40	0	0	0	5.75	0	0	4.53	4.83	0	0	1.08	140.23	Lc4	
41	0	0	0	5.14	0	0	3.56	4.24	0	0	0	149.29	Gb5/SSEA3	
42	2.28	2.59	2.09	0	0	2.78	0	0	1.98	0	1.43	152.82		
43	0	0	0	24.08	0	0	26.58	12.81	0	0	3.77	168.42	fucosyl Lc4/SSEA5	
44	4.67	0	0	0	1.83	2.25	0	0	0	0	1.88	175.52		
45	4.28	0	0	0	0	1.92	0	0	0	0	0.68	179.57	fucosyl nLc4	

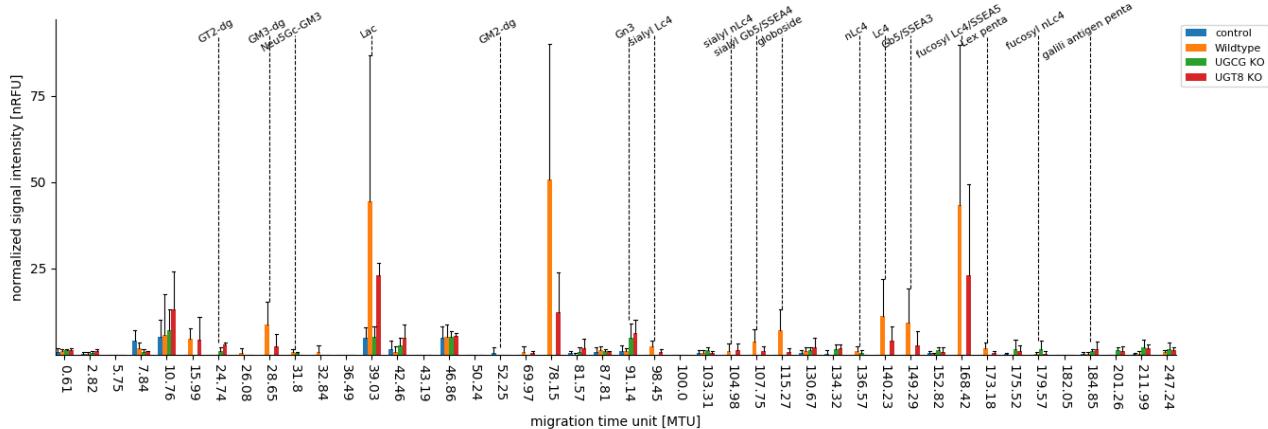
**Figure 2.7:** Excel file for the comparison of relative signal intensity between samples

The software has three outputs for this function. **First** is a bar graph between groups, which shows the difference between the mean and standard deviation of the normalized signal intensity of each group of aligned MTUs, and is annotated with known glycans in the database; **second** is an excel file that contains the mean of the normalized signal intensity of each group; **third** is an excel file that contains the standard deviation of the normalized signal intensity of each group.

- ▶ Bar graph between groups: See Figure 2.8
- ▶ Excel file with mean of the normalized signal intensity of each groups: See Figure 2.9
- ▶ Excel file with standard deviation of the normalized signal intensity of each groups: See Figure 2.10

#### 2.4.2 Variation of relative signal intensity with aligned MTU

##### *Purpose*



**Figure 2.8:** The bar graph for the comparison of normalized signal intensity between groups

In the absence of an internal standard or for the determination of relative signal intensities between replicate groups, it is possible to calculate the relative signal intensity instead of the normalized signal intensity.

### Output

The software has three outputs for this function. **First** is a bar graph between groups, which shows the difference between the mean and standard deviation of the relative signal intensity of each group of aligned MTUs, and is annotated with known glycans in the database; **second** is an excel file that contains the mean of the relative signal intensity of each group; **third** is an excel file that contains the standard deviation of the relative signal intensity of each group.

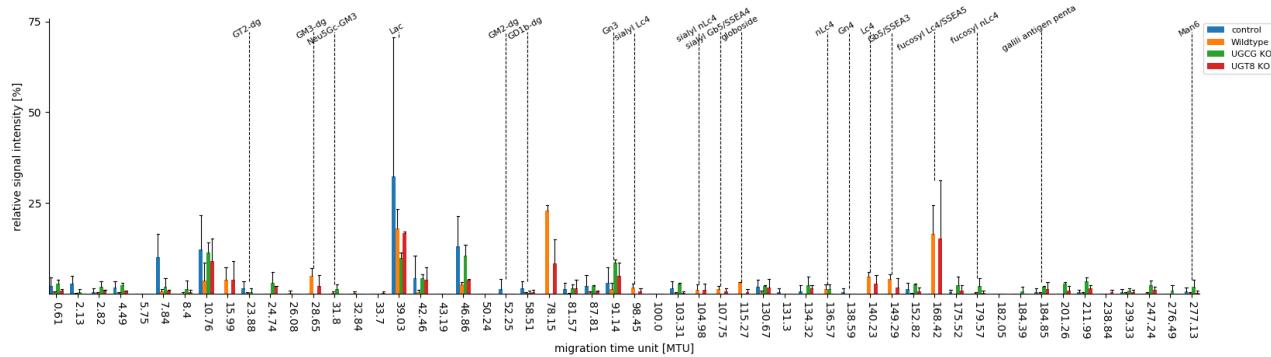
- Bar graph between groups: See Figure 2.11
- Excel file with mean of the relative signal intensity of each groups: See Figure 2.12
- Excel file with standard deviation of the relative signal intensity of each groups: See Figure 2.13

	A	B	C	D	E	F
	mean_MTU	control	Wildtype	UGCG KO	UGT8 KO	
2	0	0.61	0.94	1.0225	1.28	1.36
3	1	2.82	0.25	0.2875	0.62333333	1.155
4	2	5.75	0	0	0	0
5	3	7.84	4.046	1.9875	0.90666667	1.035
6	4	10.76	5.2	5.8875	7.04333333	13.17
7	5	15.99	0	4.595	0	4.545
8	6	24.74	0	0	1.07666667	3.06
9	7	26.08	0	0.685	0	0
10	8	28.65	0	8.875	0	2.545
11	9	31.8	0	0.72	0.45333333	0
12	10	32.84	0	0.95	0	0
13	11	36.49	0	0	0	0
14	12	39.03	4.846	44.545	5.22	23.11
15	13	42.46	1.684	0.9725	2.66666667	5.025
16	14	43.19	0	0	0	0
17	15	46.86	4.974	5.21	5.12666667	5.47
18	16	50.24	0	0	0	0
19	17	52.25	0.658	0	0	0
20	18	69.97	0	0.815	0	0.515
21	19	78.15	0	50.635	0	12.335
22	20	81.57	0.48	0.1575	0.98	1.955
23	21	87.81	0.85	1.33	1.12	1
24	22	91.14	1.186	0.9875	5.1	6.265
25	23	98.45	0	2.395	0	0.74
26	24	100	0	0	0	0
27	25	103.31	0.56	0.575	1.45666667	0.48
28	26	104.98	0	1.13	0	1.41
29	27	107.75	0	3.7575	0	1.02
30	28	115.27	0	7.28	0	0.815
31	29	130.67	0.68	1.2275	1.26666667	2.09
32	30	134.32	0.39	0	1.6	2.075
33	31	136.57	0	1.05	0.67333333	0
34	32	140.23	0	11.2525	0	4.205
35	33	149.29	0	9.47	0	2.795
36	34	152.82	0.47	0.2175	1.37666667	0.865
37	35	168.42	0	43.3225	0	23.125
38	36	173.18	0	1.925	0	0.52
39	37	175.52	0.19	0	1.79666667	1.135
40	38	179.57	0	0.4075	1.63333333	0.41
41	39	182.05	0	0	0	0
42	40	184.85	0.26	0.325	1.11	1.56
43	41	201.26	0	0	1.44333333	1.075
44	42	211.99	0.19	0.435	2.23333333	1.95
45	43	247.24	0	0.7325	1.70333333	1.495

**Figure 2.9:** Excel file with mean of the normalized signal intensity of each groups

	A	B	C	D	E	F
	mean_MTU	control	Wildtype	UGCG KO	UGT8 KO	
2	0	0.61	0.90443352	0.75632334	0.30199338	0.5939697
3	1	2.82	0.55901699	0.575	0.56163452	0.51618795
4	2	5.75	0	0	0	0
5	3	7.84	3.01882924	1.72999759	0.78926126	0.20506097
6	4	10.76	5.0082432	11.5824735	6.18587369	10.9177287
7	5	15.99	0	3.18083322	0	6.42760064
8	6	24.74	0	0	1.26508234	0.5939697
9	7	26.08	0	1.37	0	0
10	8	28.65	0	6.40967238	0	3.59917352
11	9	31.8	0	0.9997333	0.40771722	0
12	10	32.84	0	1.9	0	0
13	11	36.49	0	0	0	0
14	12	39.03	3.04796325	42.0550603	2.9779859	3.62038672
15	13	42.46	2.32874	1.41528266	2.36487491	3.69816847
16	14	43.19	0	0	0	0
17	15	46.86	3.36390844	3.64835123	1.87990248	0.96166522
18	16	50.24	0	0	0	0
19	17	52.25	1.47133273	0	0	0
20	18	69.97	0	1.63	0	0.72831998
21	19	78.15	0	39.3184507	0	11.575338
22	20	81.57	0.66577023	0.315	1.25904726	2.76478751
23	21	87.81	1.24498996	1.12927705	0.66007575	0
24	22	91.14	1.6443783	1.04910676	4.09970731	3.82544769
25	23	98.45	0	1.8472051	0	1.04651804
26	24	100	0	0	0	0
27	25	103.31	0.79561297	0.6946222	0.80872327	0.67882251
28	26	104.98	0	2.11587649	0	1.99404112
29	27	107.75	0	3.66495452	0	1.44249783
30	28	115.27	0	5.79778118	0	1.15258405
31	29	130.67	0.63007936	1.01414578	0.88217534	2.95570635
32	30	134.32	0.87206651	0	1.52630272	0.84145707
33	31	136.57	0	1.42849571	0.66040392	0
34	32	140.23	0	10.586058	0	4.1082904
35	33	149.29	0	9.71026605	0	3.95272691
36	34	152.82	0.6553625	0.435	0.82645831	1.22329473
37	35	168.42	0	46.4563999	0	26.2690169
38	36	173.18	0	1.6322752	0	0.73539105
39	37	175.52	0.42485292	0	2.58558182	1.60513239
40	38	179.57	0	0.47981767	2.37674848	0.57982756
41	39	182.05	0	0	0	0
42	40	184.85	0.58137767	0.65	0.69202601	2.20617316
43	41	201.26	0	0	0.76787586	1.52027958
44	42	211.99	0.42485292	0.55883808	2.29698353	0.97580736
45	43	247.24	0	0.67148467	1.92313113	0.75660426

**Figure 2.10:** Excel file with standard deviation of the normalized signal intensity of each groups



**Figure 2.11:** The bar graph for the comparison of relative signal intensity between groups

	A	B	C	D	E	F
1	mean_MTU	control	Wldtype	UGCG KO	UGT8 KO	
2	0	0.61	2.258	0.3675	2.753333333	0.958
3	1	2.13	2.764	0.14	0.51	0
4	2	2.82	0.472	0.13	1.84	0.87
5	3	4.49	1.798	0.285	2.36666667	0.829
6	4	5.75	0	0	0	0
7	5	7.84	10.126	0.705	1.993333333	0.775
8	6	8.4	0	0.1175	1.35666667	0.465
9	7	10.76	12.18	3.5625	11.35333333	8.915
10	8	15.99	0	3.8275	0	3.77
11	9	23.88	1.412	0.2525	0.55	0
12	10	24.74	0	0	2.98	2.209
13	11	26.08	0	0.31	0	0
14	12	28.65	0	4.8925	0	2.11
15	13	31.8	0	0.315	1.35	0
16	14	32.84	0	0.2175	0	0
17	15	33.7	0	0	0	0.25
18	16	39.03	32.406	18.08	9.85666667	16.3
19	17	42.46	4.382	0.425	4.29	3.93
20	18	43.19	0	0	0	0
21	19	46.86	13.042	2.61	10.46333333	3.945
22	20	50.24	0	0	0	0
23	21	52.25	1.264	0	0	0
24	22	58.51	1.452	0.2225	0.31333333	0.465
25	23	78.15	0	23.0225	0	8.26
26	24	81.57	1.234	0.07	1.423333333	1.62
27	25	87.81	2.14	0.4425	2.09666667	0.735
28	26	91.14	3.078	1.365	8.61666667	4.858
29	27	98.45	0	1.6425	0	0.615
30	28	100	0	0	0	0
31	29	103.31	1.45	0.1775	2.73666667	0.309
32	30	104.98	0	1.1775	0	1.17
33	31	107.75	0	1.2	0	0.65
34	32	115.27	0	3.205	0	0.52
35	33	130.67	1.99	0.4475	2.25666667	1.735
36	34	131.3	0.472	0	0	0
37	35	134.32	0.752	0	2.473333333	1.579
38	36	136.57	0	1.23	1.27333333	0
39	37	138.59	0.492	0	0	0
40	38	140.23	0	4.775	0	2.809
41	39	149.29	0	3.99	0	1.76
42	40	152.82	1.246	0.0975	2.55	0.715
43	41	168.42	0	16.51	0	15.175
44	42	175.52	0.366	0	2.30666667	0.94
45	43	179.57	0	0.18	2.06666667	0.34

**Figure 2.12:** Excel file with mean of the relative signal intensity of each groups

	A	B	C	D	E	F
		mean_MTU	control	Wildtype	UGCG KO	UGT8 KO
1						
2	0	0.61	2.30279613	0.29432125	1.0513959	0.24748737
3	1	2.13	2.24310722	0.16492423	0.88334591	0
4	2	2.82	1.05542409	0.26	1.59361225	0.22627417
5	3	4.49	1.74943419	0.26451213	0.75035547	0.06363961
6	4	5.75	0	0	0	0
7	5	7.84	6.46802752	0.66795708	2.34017806	0.28991378
8	6	8.4	0	0.235	2.3498156	0.65760931
9	7	10.76	9.51824564	4.95587446	2.8533197	6.21546861
10	8	15.99	0	3.40004289	0	5.33158513
11	9	23.88	1.99581312	0.18246004	0.95262794	0
12	10	24.74	0	0	3.11814047	0.0212132
13	11	26.08	0	0.62	0	0
14	12	28.65	0	2.19276652	0	2.98399062
15	13	31.8	0	0.43401229	1.16940156	0
16	14	32.84	0	0.435	0	0
17	15	33.7	0	0	0	0.35355339
18	16	39.03	38.1986509	5.22771461	1.50287502	0.48083261
19	17	42.46	6.16067529	0.61305247	0.97503846	3.39411255
20	18	43.19	0	0	0	0
21	19	46.86	8.48399493	0.52383203	3.05738995	0.03535534
22	20	50.24	0	0	0	0
23	21	52.25	2.82638992	0	0	0
24	22	58.51	1.99487343	0.15413738	0.54270925	0.65760931
25	23	78.15	0	1.46586436	0	6.81650937
26	24	81.57	1.73155422	0.14	1.25125271	2.29102597
27	25	87.81	2.93249893	0.29915158	0.31628047	0.13435029
28	26	91.14	4.31007193	1.70398552	0.86961677	3.65574206
29	27	98.45	0	1.1371712	0	0.86974134
30	28	100	0	0	0	0
31	29	103.31	2.00706253	0.20694202	0.37313983	0.43133514
32	30	104.98	0	1.39929447	0	1.65462987
33	31	107.75	0	0.94077982	0	0.91923882
34	32	115.27	0	0.07234178	0	0.73539105
35	33	130.67	1.90231438	0.41835989	0.11590226	2.45366053
36	34	131.3	1.05542409	0	0	0
37	35	134.32	1.68152312	0	2.26294793	0.89802561
38	36	136.57	0	1.41286942	1.26021162	0
39	37	138.59	1.10014544	0	0	0
40	38	140.23	0	1.20610945	0	2.4395184
41	39	149.29	0	1.40904223	0	2.51730014
42	40	152.82	1.75230705	0.195	0.25238859	1.0111627
43	41	168.42	0	7.87586609	0	16.1291057
44	42	175.52	0.81840088	0	2.33551565	1.32936075
45	43	179.57	0	0.21354157	2.14376616	0.48083261

**Figure 2.13:** Excel file with standard deviation of the relative signal intensity of each groups

# Appendix

## A Instruction of Software Development

### A.1 Scripting Language

The software is developed in python. Considering that Python is open source and free, Python's technology stack can cover web development, data analysis, data mining, background development and many other aspects, and Python can flexibly use a large number of third-party libraries, such as tkinter for GUI and numpy, pandas for scientific calculation. In addition, as an interpretation language, Python has cross-platform characteristics, and its code is very readable and maintainable, so it is a better choice to use Python to develop.

The python libraries used in development are:

- ▶ tkinter for GUI
- ▶ numpy, pandas for data analysis
- ▶ matplotlib for plotting

### A.2 Script Files and Configuration Files

- ▶ GSL\_database.csv: GSL database, storing information about GSL-derived glycans.
- ▶ N\_glycans\_database.csv: N-glycans database, storing information about N-glycans-derived glycans.
- ▶ GAGs\_database.csv: GAGs database, storing information about GAGs-derived glycans.
- ▶ database.py: A script file that extracts or stores data from the database to the GUI.
- ▶ frame\_database.py: Script file for the "Database" page in the software (Chapter 3.1).

- ▶ frame\_overview.py: Script file for the "Overview" page in the software (Chapter 3.2).
- ▶ frame\_parameter.py: The script file for the "Parameter" page in the software (Chapter 3.3).
- ▶ frame\_analysed\_data.py: Script file for the "Analyzed Data" page in the software (Chapter 3.4).
- ▶ frame\_figure.py: The "Figure" page script file in the software (Chapter 3.5).
- ▶ all.py: Main script file of the software.
- ▶ default\_in\_referencewell.txt: used to store the user-defined default glycans in referencewell in the "Analyzed Data" page.
- ▶ default\_normalization.txt: used to store the user-defined standard glycan for normalization in the "Analyzed Data" page.

## B Instruction of Installation

The script file of the software has been packaged into an exe file (executable file) with pyinstaller, currently there is a version for Windows and a version for Mac OS, users can download the exe file and load the software directly, but also need to download several configuration files together.

### B.1 Installation for Windows

Relevant files:

- ▶ all.exe: Users can use the software by double-clicking on this file with the left mouse button.
- ▶ configuration files:
  - ▶ GSL\_database.csv: GSL database, storing information about GSL-derived glycans.
  - ▶ N\_glycans\_database.csv: N-glycans database, storing information about N-glycans-derived glycans.
  - ▶ GAGs\_database.csv: GAGs database, storing information about GAGs-derived glycans.
- ▶ default\_in\_referencewell.txt: used to store the user-defined default glycans in referencewell in the "Analyzed Data" page.
- ▶ default\_normalization.txt: used to store the user-defined standard glycan for normalization in the "Analyzed Data" page.

## B.2 Installation for Mac OS

Relevant files:

- ▶ all.exe: Users can use the software by double-clicking on this file with the left mouse button.
- ▶ configuration files:
  - ▶ GSL\_database.csv: GSL database, storing information about GSL-derived glycans.
  - ▶ N\_glycans\_database.csv: N-glycans database, storing information about N-glycans-derived glycans.
  - ▶ GAGs\_database.csv: GAGs database, storing information about GAGs-derived glycans.
- ▶ default\_in\_referencewell.txt: used to store the user-defined default glycans in referencewell in the "Analyzed Data" page.
- ▶ default\_normalization.txt: used to store the user-defined standard glycan for normalization in the "Analyzed Data" page.

## B.3 Installation for Linux

Currently there is no exe file packaged for Linux, Linux computers can create a project in python and load the script file and configuration file (in Chapter 4.1.2) in order to use the software.

Additional python libraries that need to be downloaded are:

- ▶ numpy
- ▶ pandas
- ▶ matplotlib

The code is in the webpage <https://github.com/LuoYi8989/Software-for-Data-Analysis-of-xCGE-LIF.git>