DANGO

version 1.0

Instructions for DANGO annotation

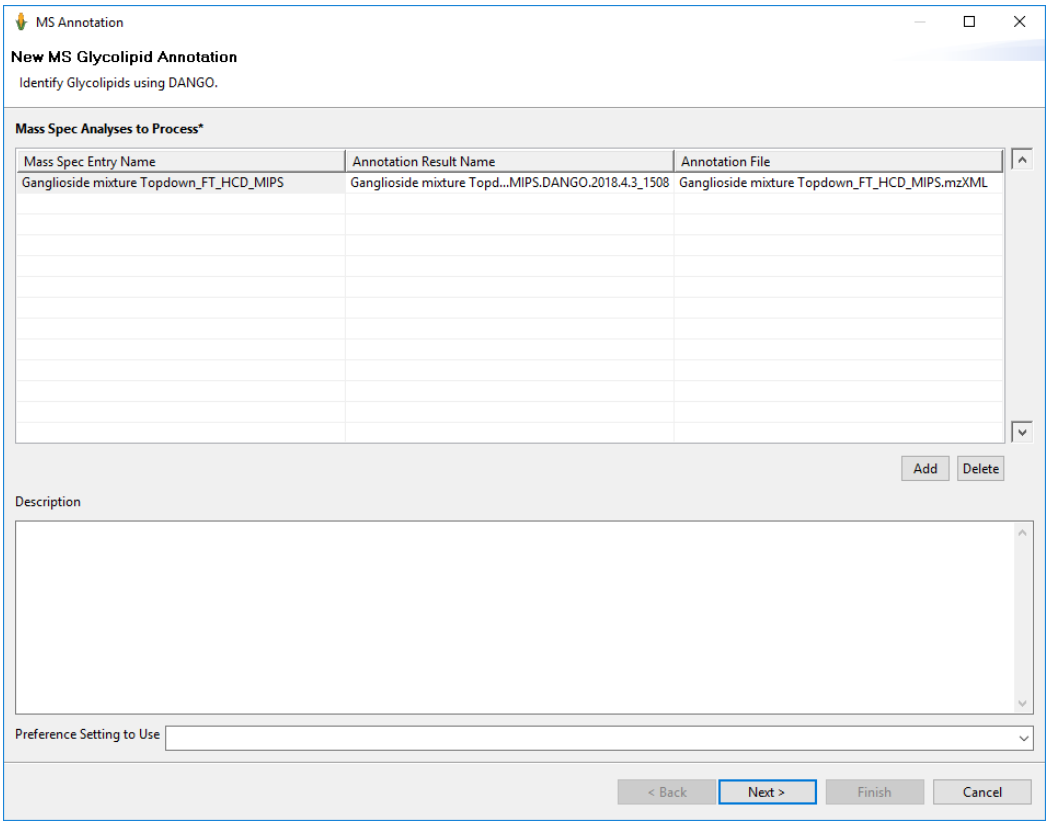
by Masaaki Matsubara

Last update: Sep 28th 2018

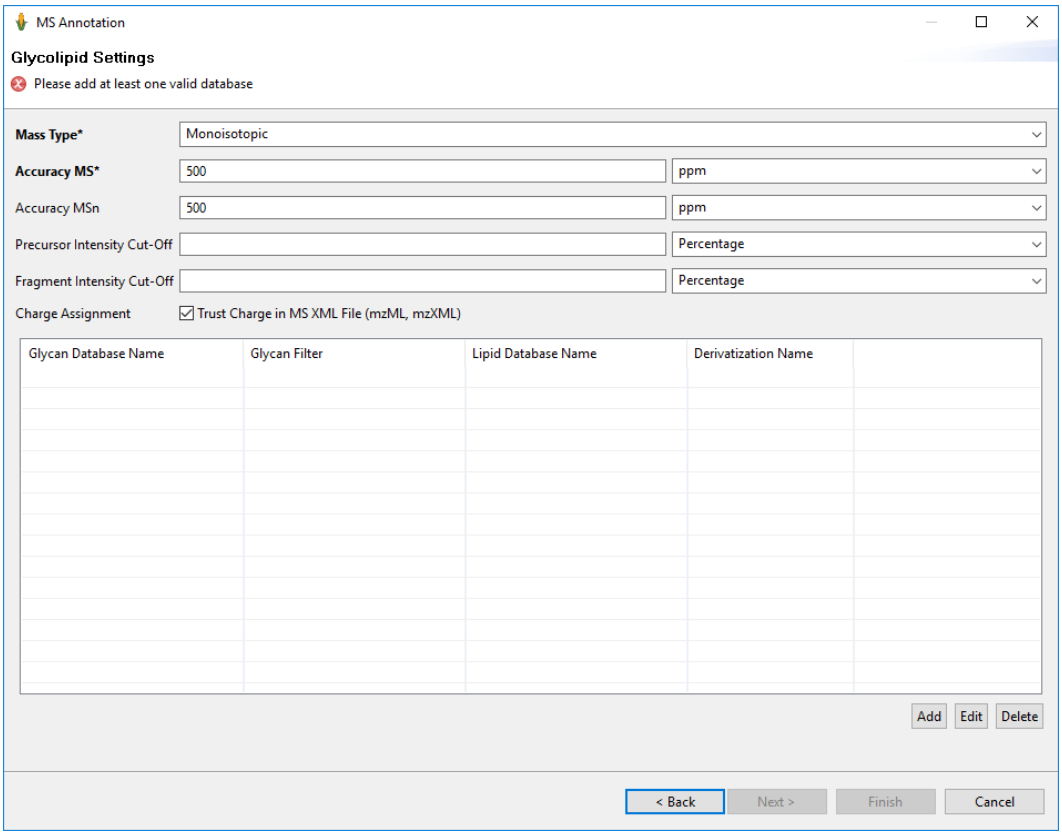
Table of Contents

1 Detailed instructions

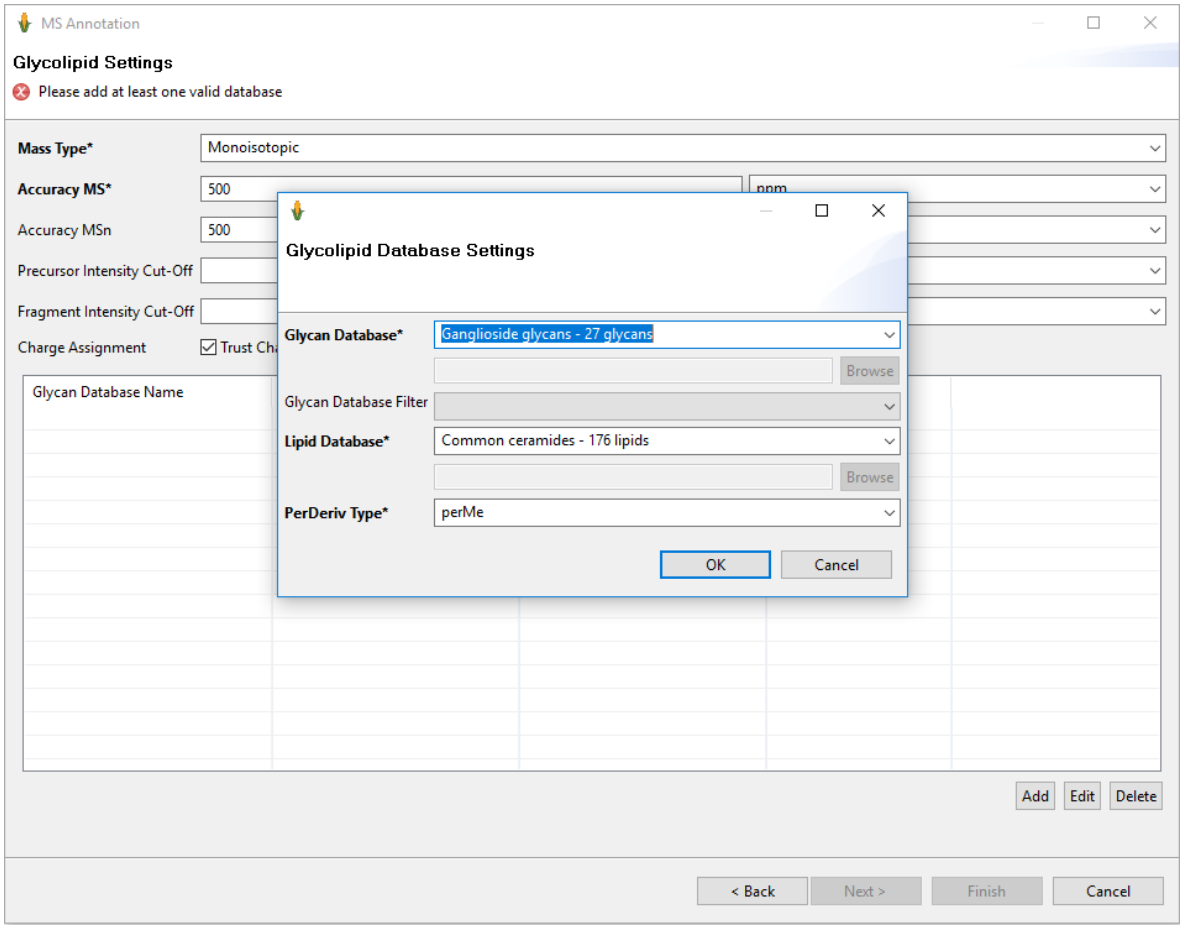
1. Right click on an MS entry or use the *File* main menu: *New* →*DANGO Annotation*



2. The MS entry that will get annotated is shown in the table in the first column. The second column shows the name of the Annotation entry that will be created below the MS entry. This name can be changed be clicking in the cell. Third column allows the user to select the annotation file to be used for annotation. The first available annotation file is selected by default.
3. If multiple MS entries shall be annotated at once
    a. Select all them before doing step (1)
    b. Use the add button in the dialog to add them to the table
4. The user can also select a preference setting to use so that the settings in the next pages can be prefilled with the values from the selected preference setting.
5. Click *Next >* button



6. In the **Glycolipid Settings** page, you can choose the following configurations
    a. **Mass Type**: Depending on the instrument either *Monoisotopic* or *Average*
    b. **Accuracy:** Can be specified either in PPM or Dalton. There are separate settings for the MS^1 (*Accuracy MS*) and MS^n scans (*Accuracy MSn*)
    c. **Intensity Cut-Off**: Allows ignoring all peaks that are smaller than the specified value. This value can be given in *Percentage* to the highest peak or as *Absolute Value.* There are separate settings for Precursor and Fragment.
    d. **Trust Charge:** Usually the instrument software will assign charge states to peaks. If this option is set only annotations with the assigned charge state will be calculated. However, sometimes the instrument assigns the wrong charge state to a peak. If this happens leave this deselected and DANGO will make annotations with all possible charge states



7. **Database Table**
    a. There must be at least one database selection in the database table. A new glycan and lipid databases can be added by clicking *Add* button under the table. The user can also *Edit* or *Delete* an existing database selection. Multiple database settings are allowed.
8. **Adding a Database selection**
    a. **Glycan Database:** Select the glycan database to be used as a glycan part of glycolipid. GRITS Toolbox has a list of predefined databases that can be used. The "-topology" databases contain structures with no linkage information and are smaller as their counterparts that have fully defined structures. This can reduce the number of redundant annotation if you do not have A/X cleavages that would allow differentiating linkage isomers. It is also possible to specify a user defined database by choosing *other* and use the *Browse* button below to select the file with the database.
    b. **Database Filter:** Database can further be reduced by applying structure based filters. If there are any predefined filters (created in preferences), the user can select one from the drop-down or select "other" to create a new filter setting.
    c. **Lipid database:** Select the lipid database to be used as a lipid part of glycolipid. GRITS Toolbox has a list of predefined lipid databases that can be used. In this version, there are three databases:

i. *Ceramide_common*: Ceramide database containing common structures. It contains 176 lipids. Their structure values are settled within the following:
    1. Sphingosine part:
        a. Carbon length: 18
        b. # of hydroxyl group: 2,3
        c. # of double bond: 0,1
    2. Fatty Acid part:
        a. Carbon length: 16-26
        b. # of hydroxyl group: 0,1
        c. # of double bond: 0,1

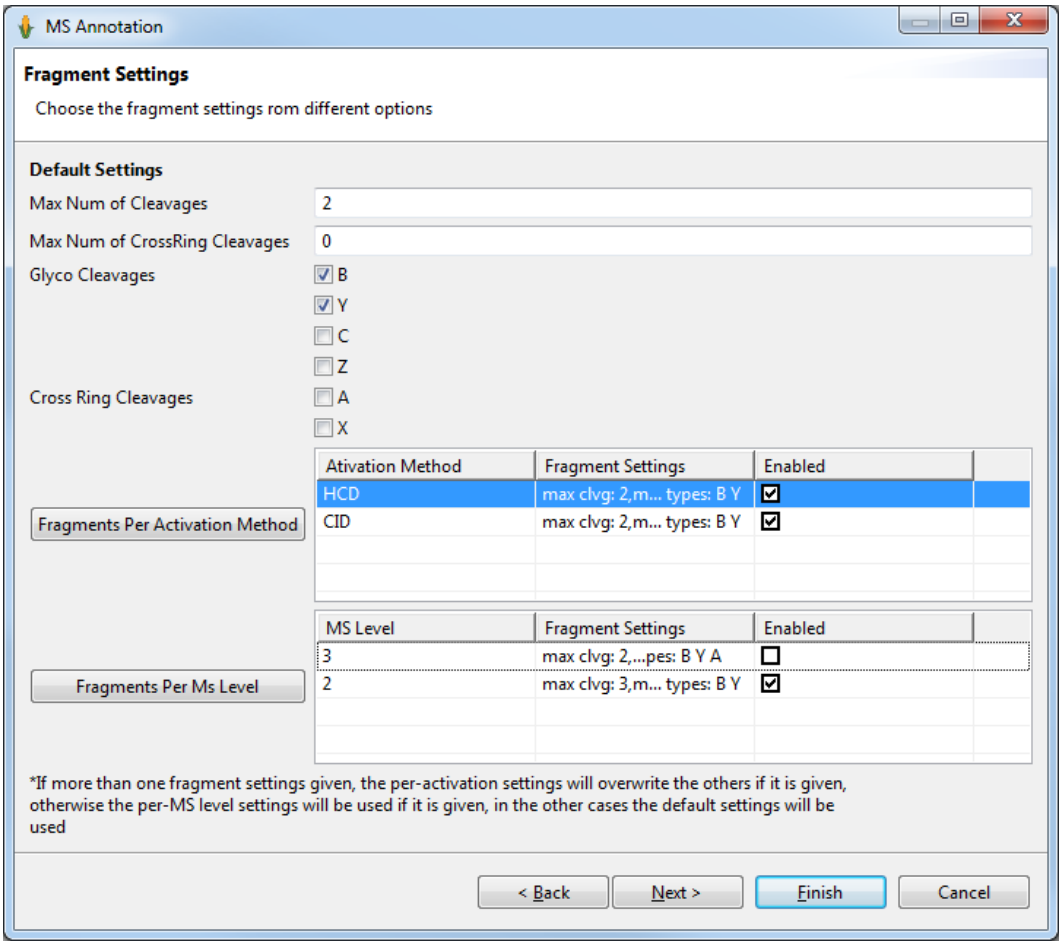
ii. *Ceramide_common_extendedLCB*: Ceramide database containing common and extended LCB structures. It contains 352 lipids.
    1. Sphingosine part:
        a. Carbon length: 18,20
        b. # of hydroxyl group: 2,3
        c. # of double bond: 0,1
    2. Fatty Acid part:
        a. Carbon length: 16-26
        b. # of hydroxyl group: 0,1
        c. # of double bond: 0,1

iii. *Ceramide_all*: Ceramide database containing all possible structures. It contains 42432 lipids.
    1. Sphingosine part:
        a. Carbon length: 14-30
        b. # of hydroxyl group: 1-4
        c. # of double bond: 0-3
    2. Fatty Acid part:
        a. Carbon length: 14-26
        b. # of hydroxyl group: 0-2
        c. # of double bond: 0-3

It is also possible to specify a user defined database by choosing *other* and use the *Browse* button below to select the file with the database. A new lipid database can be created using the *Tools* main menu: *Glycolipid Tools -> Create Lipid Database*. See chapter **2.2**.

d. **PerDeriv Type:** Type of derivatization used for the glycan. Option are:
    i. *perMe* – permethylation
    ii. *None* – no derivatization

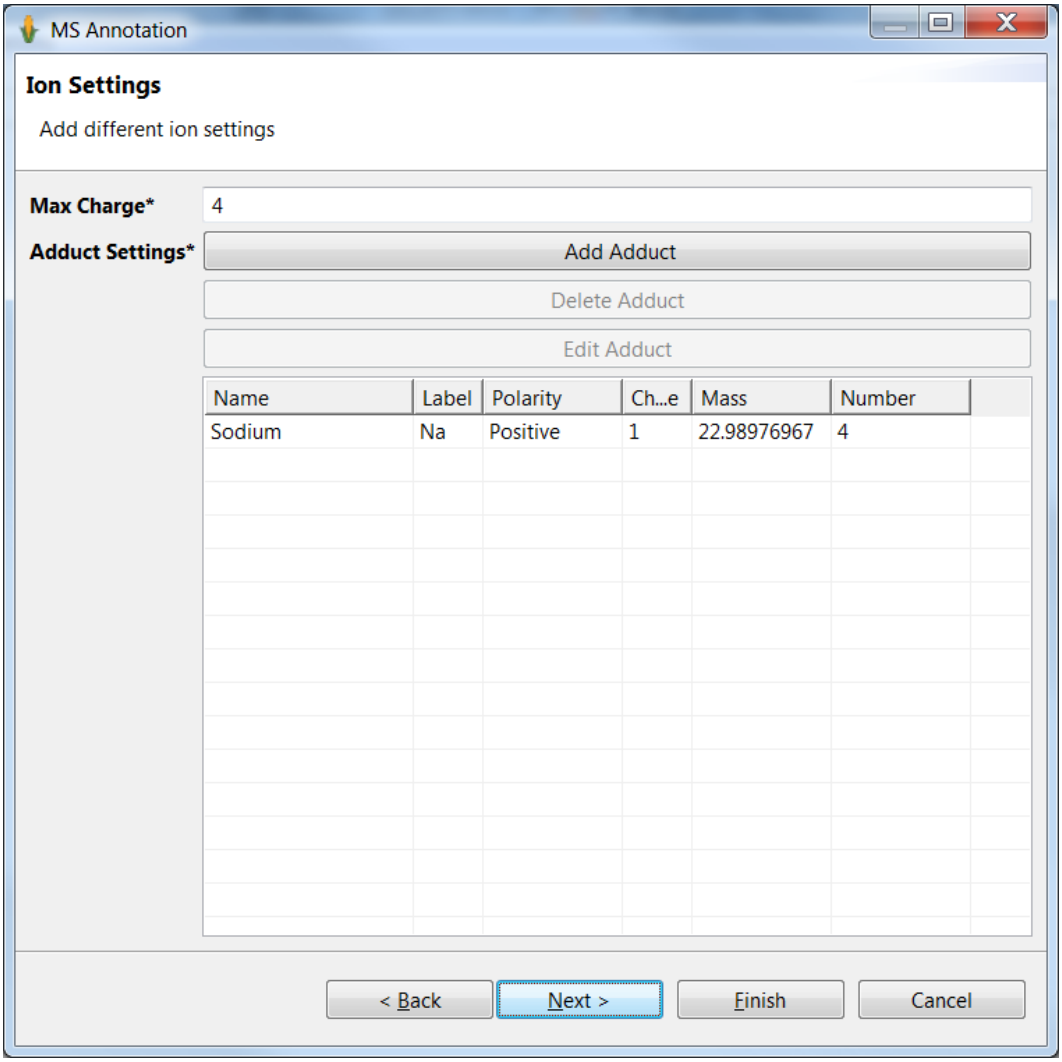
9. Click on *Next >* button



10. In the **Fragment Settings** page, you can choose the fragmentation to be used in the MS^n spectra. If you are running MS profile data these settings will not have any effect.
    a. The section **Default Settings** contains fragment configurations that will be used by default
        i. **Max Num of Cleavages**: Maximum number of glycosidic cleavages that are allowed per fragment.
        ii. **Max Num of CrossRing Cleavages**: Maximum number of crossring cleavages that are allowed per fragment. If this number is 0 no crossrings will be calculated even if they are selected below.
        iii. **Glyco Cleavages**: Type of glyosidic cleavages that are allowed. Uses Domon and Costello naming (e.g. B, C, Y, Z)
        iv. **Cross Ring Cleavages:** Type of glyosidic cleavages that are allowed. Uses Domon and Costello naming (e.g. A, B)
    b. **Fragments Per Activation Method**: Clicking the button will open a dialog that allows choosing an activation method (e.g. CID, HCD, ETD) and assigning fragment settings. Spectra that have been generated using this activation method will be annotated with the corresponding setting. For example for CID only glycosidic cleavages are allowed while for HCD also cross ring cleavages are allowed. These settings have preference over all other settings in this page.

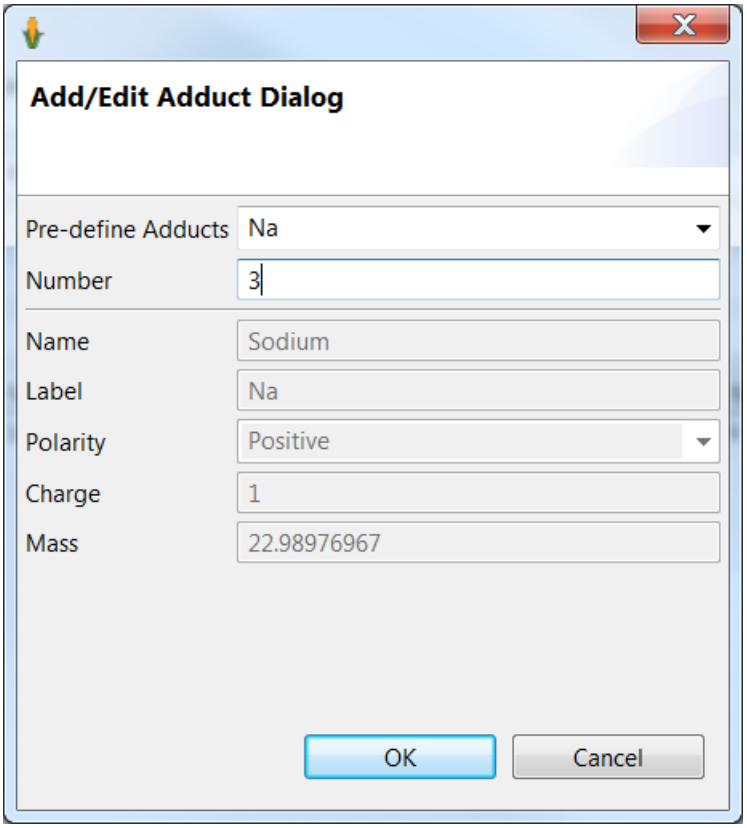
c. **Fragments Per Ms Level**: Clicking this button will open a dialog that allows choosing different fragment settings for different MS level. For example in MS2 only glycosidic cleavages are allowed while for MS3 and MS4 also cross ring cleavages are allowed.
d. *Please Note*: DANGO will first try to find fragmentation settings based on the activation method. If no match can be found it will look next if there are settings for the MS level. And only if no rule can be found there either the default settings are applied.

11. Click on *Next* >button



12. The **Ion Settings** page allows specifying the possible adducts of the peaks in the scans. The page allows also specifying mixtures, for example if a scan contains protonated and sodiated peaks within the same scan. If positive charged adducts and negative charged adducts are specified only adducts that match the polarity of a scan will be used to calculate annotations for a scan.
    a. **Max Charge**: Specifies the total maximum of charges users expect to see in the scan. This is the upper limit even if combinations of adducts would allow for more. For example, up to 4 Na+ and 4 H+ are allowed but if the total maximum is 3 no annotation with more than charge state 3 will be calculated.

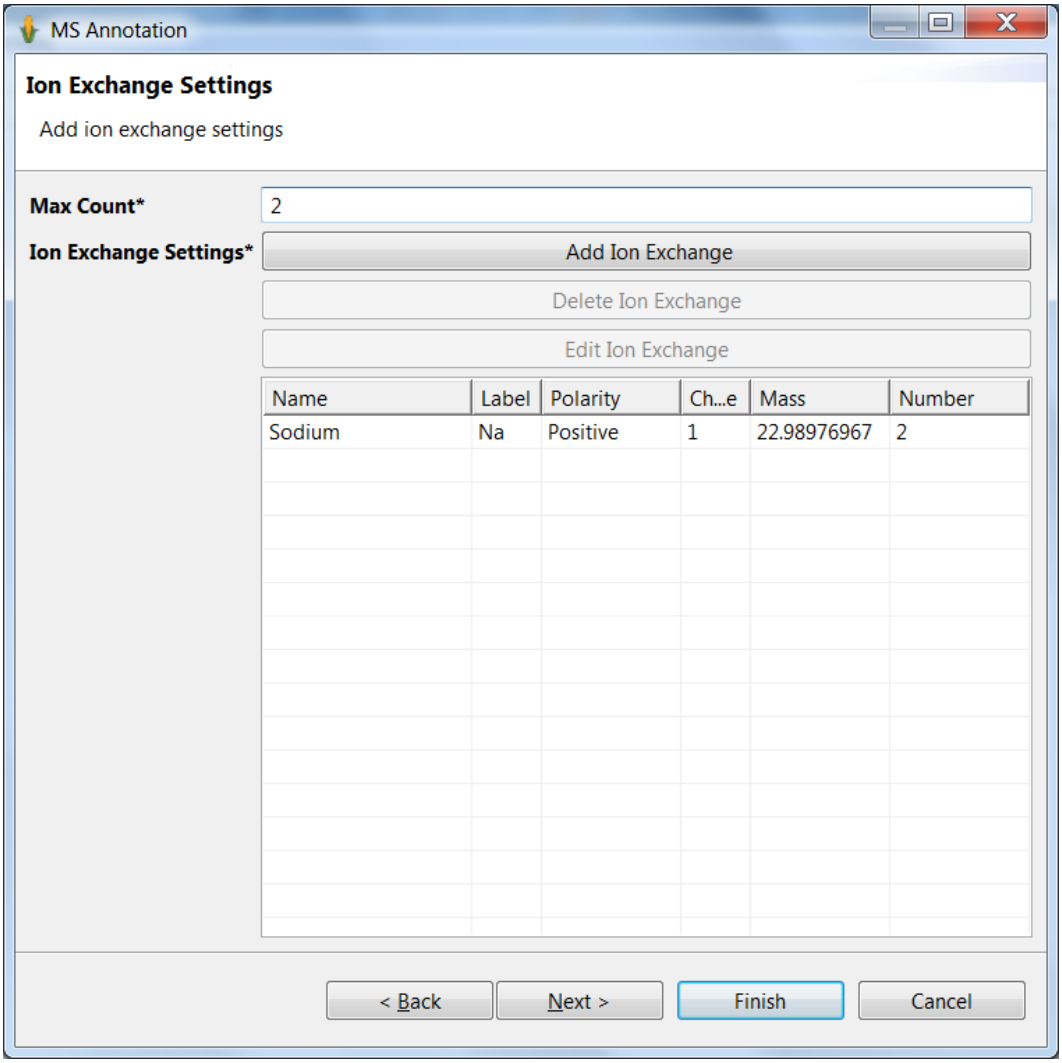
b. **Adduct Settings**: The three buttons allow adding, deleting and editing adducts in the table. For the latter two a line in the table needs to be selected.



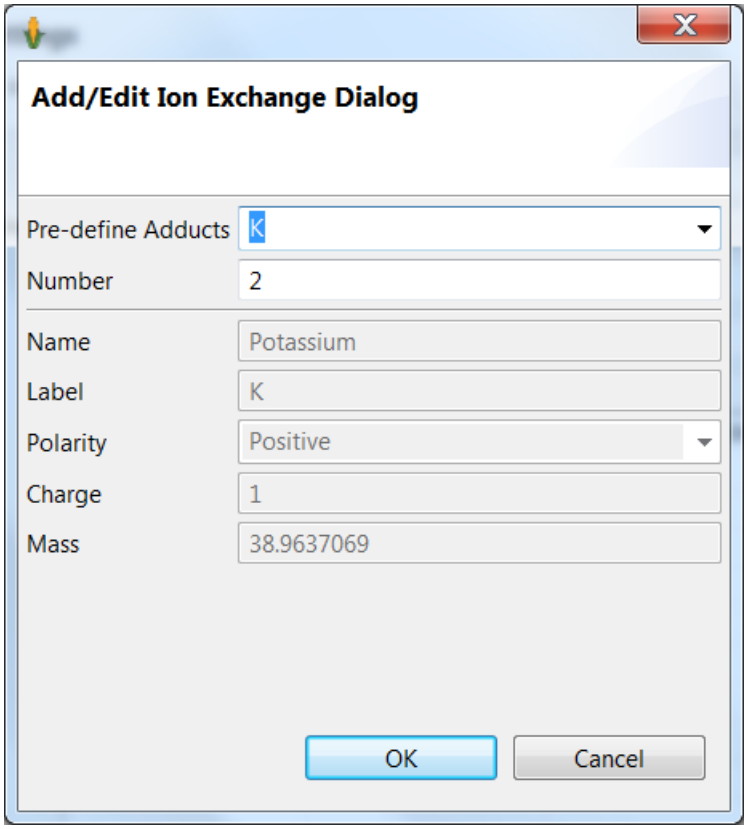
c. The **Add/Edit Adduct Dialog** will be shown if the appropriate buttons are clicked. This dialog allows specifying an adduct to be used for annotation:
    i. **Pre-defined Adducts**: The dropdown contains a list of common adducts, which are already fully defined, such as Na+, H+, or -H. Missing adducts can be added by choosing *other* which will enable the fields in the bottom part of the dialog.
    ii. **Number**: Maximum number how many of these adducts can be found for a peak. Keep in mind that this number is capped by the **Max Charge** in the **Ion Settings** page.

    The following fields become available when selecting other in the adducts field and will allow specifying user defined adducts:
    iii. **Name**: User defined name of the new adduct.
    iv. **Label**: Short name of the new adduct. This will be used in most displays to save space.
    v. **Polarity**: Positive or negative adduct. This information will be used to decide if an adduct is applicable for a scan by considering the scan polarity.
    vi. **Charge**: Number of charges of the adduct.
    vii. **Mass**: Mass of the charged adduct which will be added to the mass of the glycan.

13. Click on *Next* >button

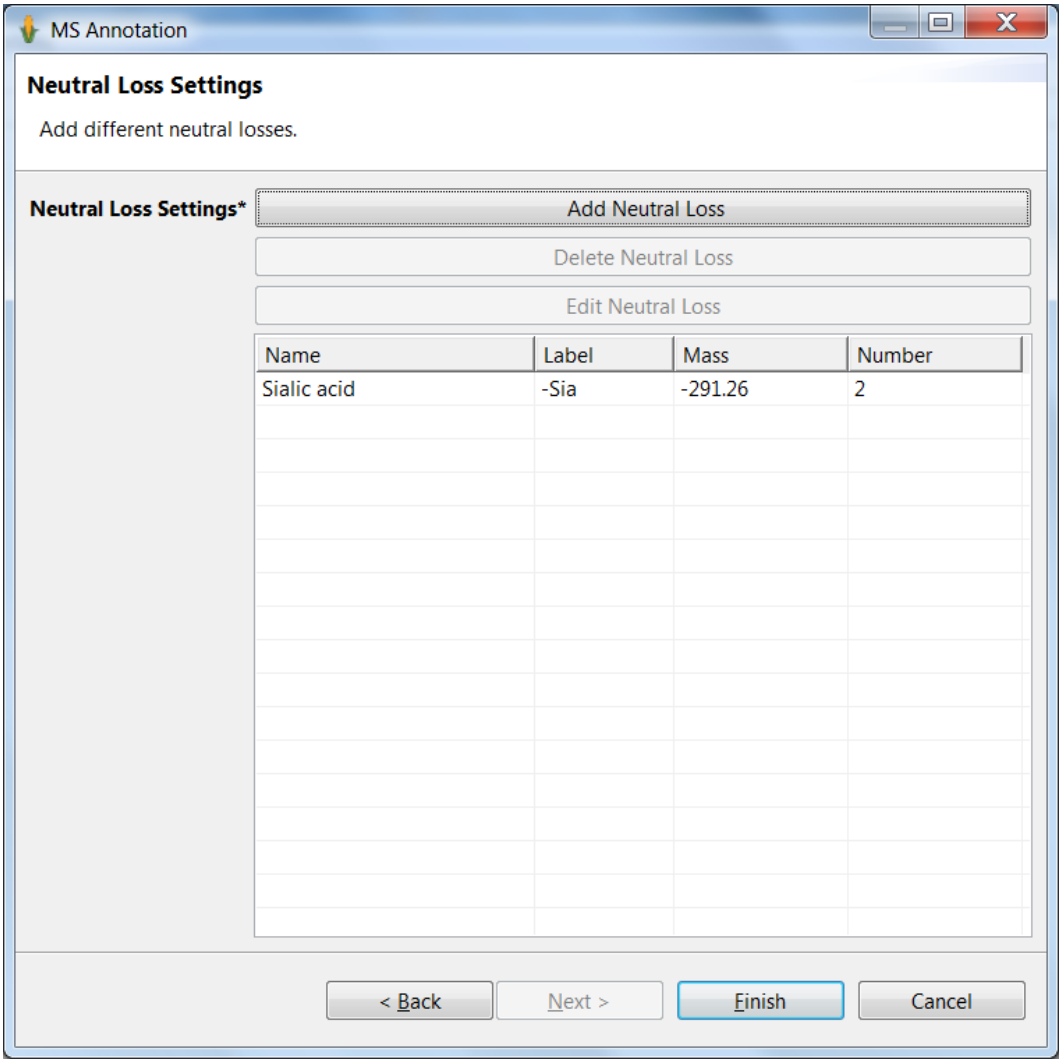


14. Sometimes it is possible that the molecule loses an H+ and this gets replaced by one of the adducts (e.g. Na+). This is called ion exchange (-H+Na) and can be configured on the **Ion Exchange Settings** page. The effect is that the mass of the molecules is increased by another adduct but the charge stage does not change. For example, for a ion exchange with sodium adducts the mass will be increase by 22 Da (-H+Na = -1+23 = 22). If multiple adducts are chosen on the previous page it is possible to have a mixture of exchanges as well.
    a. **Max Count**: Specifies the total maximum of exchange users expect to see in the scan for a peak. This is the upper limit even if combinations of exchanges would allow for more. For example, up to 2 Na exchanges and 2 Li exchanges are allowed but if the total maximum is 3 no annotation with more than 3 exchanges will be calculated.
    b. **Ion Exchange Settings**: The three buttons allow adding, deleting and editing adducts in the table. For the latter two a line in the table needs to be selected.

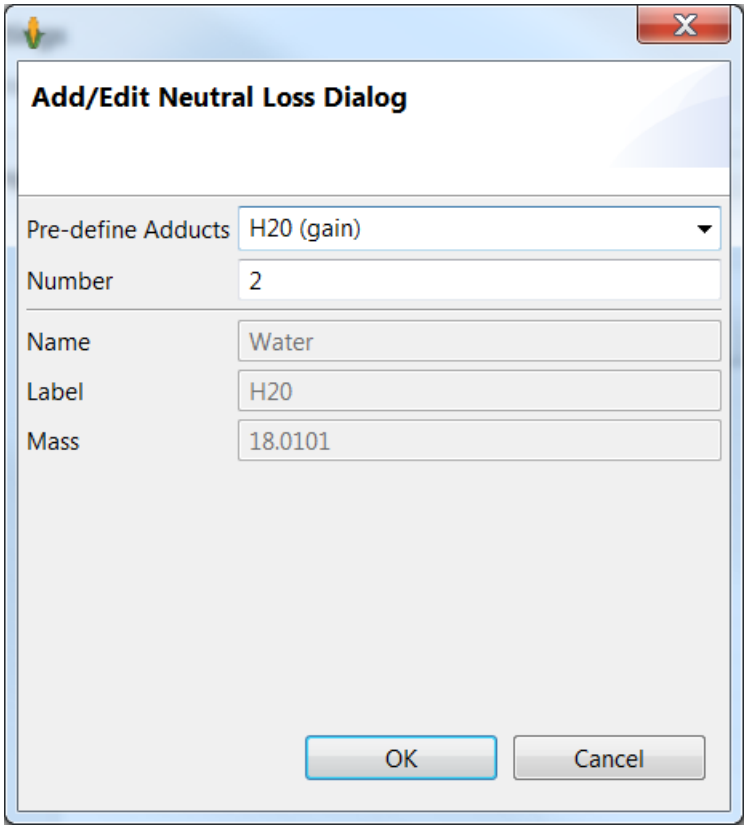


c. The **Add/Edit Ion Exchange Dialog** will be shown if the appropriate buttons are clicked. This dialog allows specifying an adduct that will be used for exchange in the annotation:
    i. **Pre-defined Adducts**: The dropdown contains a list of all adducts specified in the previous page. Information about the adduct will be shown in the lower part of the dialog.
    ii. **Number**: Maximum number how many exchanges are possible for this adduct. Keep in mind that this number is capped by the **Max Count** in the **Ion Exchange Settings** page.

15. Click on *Next* >button



16. Sometimes certain residues (e.g. Sialic acid in positive mode) or fragments of residues (e.g. H_2O for underivatized glycans) can fall of in the MS instrument. These losses can be configured using the **Neutral Loss Settings** page. This page can also be used to configure additions (neutral mass gains) of masses. It is also possible to model are effects like undermethylation for permethylated glycans which will result in the loss of 14 Dalton per side. This option is optional and can be left empty if not loss or gain is expected.
    a. **Neutral Loss Settings**: The three buttons allow adding, deleting and editing neutral losses in the table. For the latter two a line in the table needs to be selected.



b. The **Add/Edit Neutral Loss Dialog** will be shown if the appropriate buttons are clicked. This dialog allows specifying a loss or gain that will be considered in the annotation:
    i. **Pre-defined Adducts**: The dropdown contains a list of predefined loses or gains. If the desired option is not present it is possible to choose other which will enable the options in the lower part of the dialog and allow for definition of user defined gain or loses.
    ii. **Number**: Maximum number of losses or gains for this molecule. The annotation will consider that that no modification happens and then all combination till the defined number.

    If the option other is selected in Pre-defined Adducts the following fields will become available allowing for the definition of new losses or gains:
    iii. **Name**: User defined name of the new loss or gain.
    iv. **Label**: Short name of the molecule. This will be used in most displays to save space.
    v. **Mass**: Mass of the modification. A positive mass specifies a neutral mass gain and will be added to the mass of the glycan. A negative mass specifies a neutral mass loss and will be subtracted from the mass of the glycan.

17. Click on *Finish* >button to that the annotation
18. After the annotation is finished a new entry for this annotation is created below the MS entry and the annotation will be opened automatically.

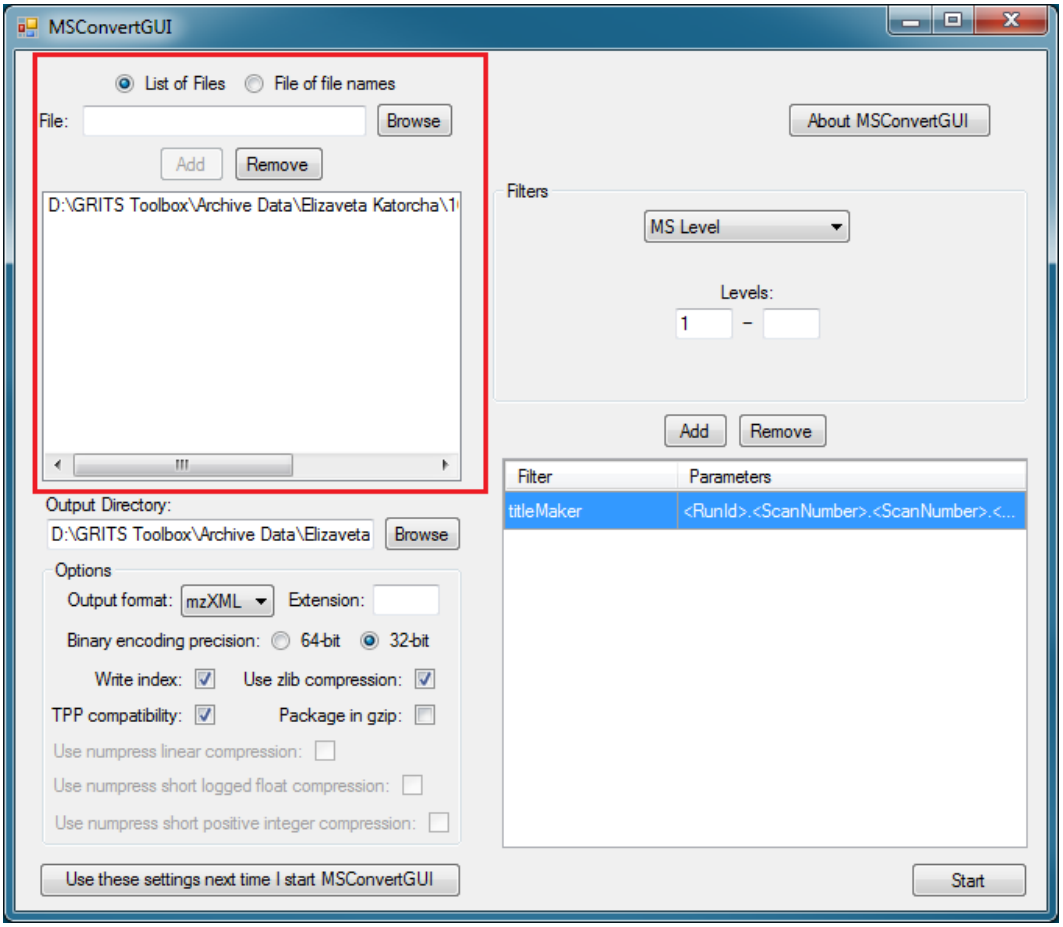
2 Questions and answers

This section contains some answers to question you may have and also shows workarounds for common problems.

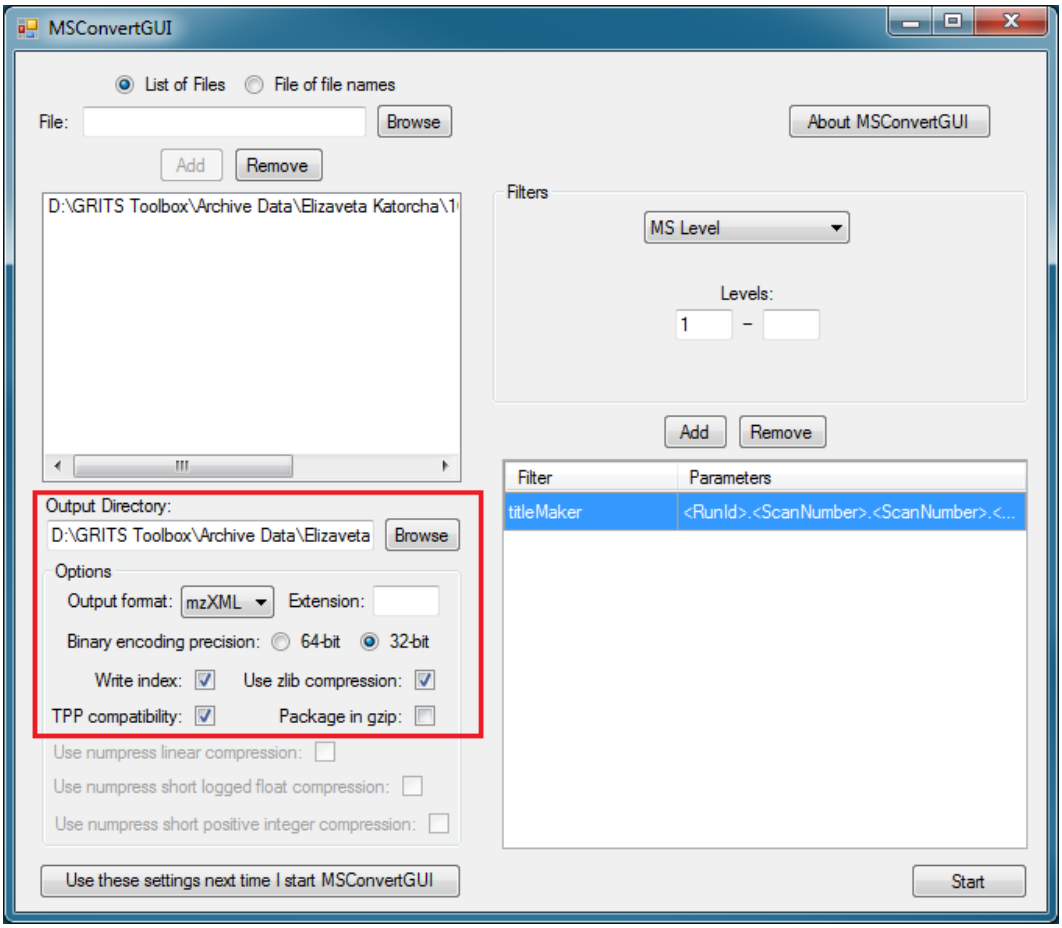
2.1 How to deal with large files

Sometimes you need to annotate large data sets (e.g. LC/MS runs) with tens of thousands of spectra. For programs low memory or older CPUs this can take forever or may even cause the program to crash. In such cases it is better to split the data file in smaller chunks and analyze them separately. This will also allow excluding sections in the data that do not have information of interested. For example if you use a LC-column and know that glycans or glycopeptides come from the column after 5 minutes and stop after 40 minutes, it is not necessary to annotated the complete 90 minutes run but just the section 5-40 minutes. To split the original file in pieces you can use MSconvert (part of the ProteoWizard software: http://proteowizard.sourceforge.net):

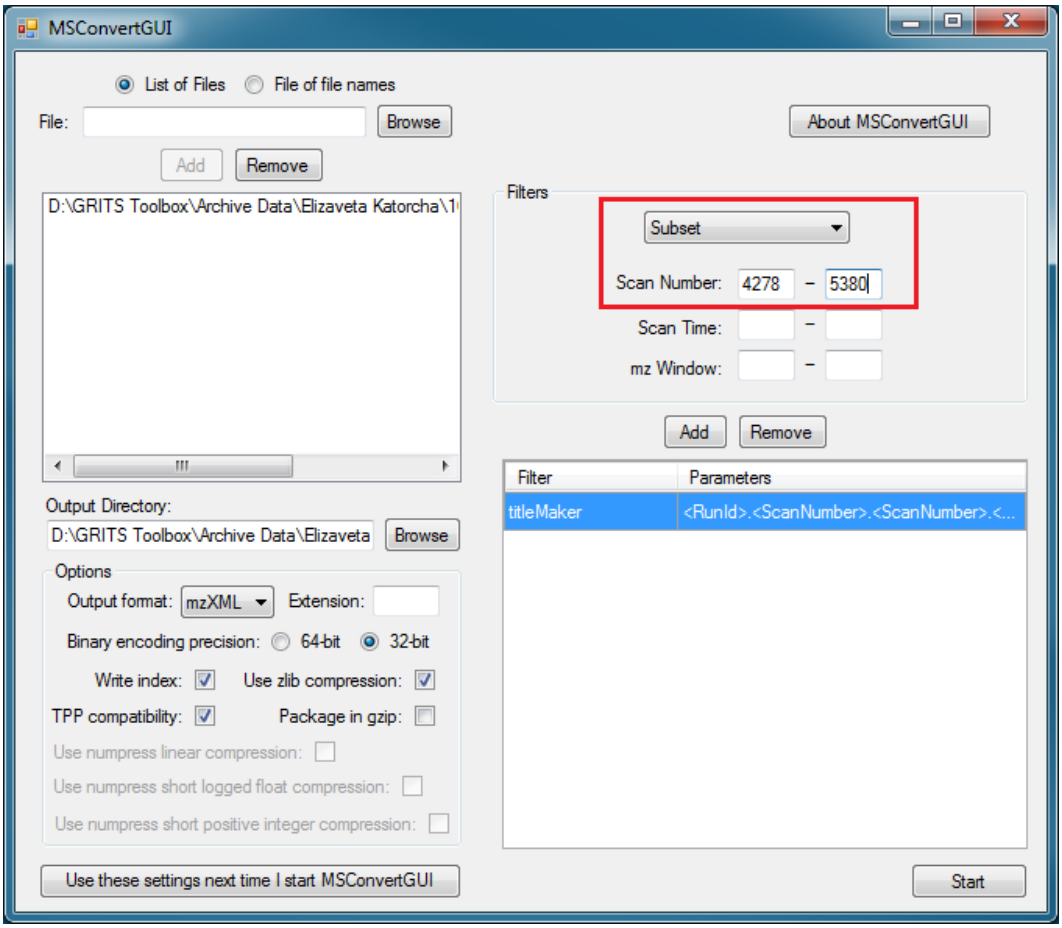
1. Load your original data file or an mzXML/mzML file to msconvert



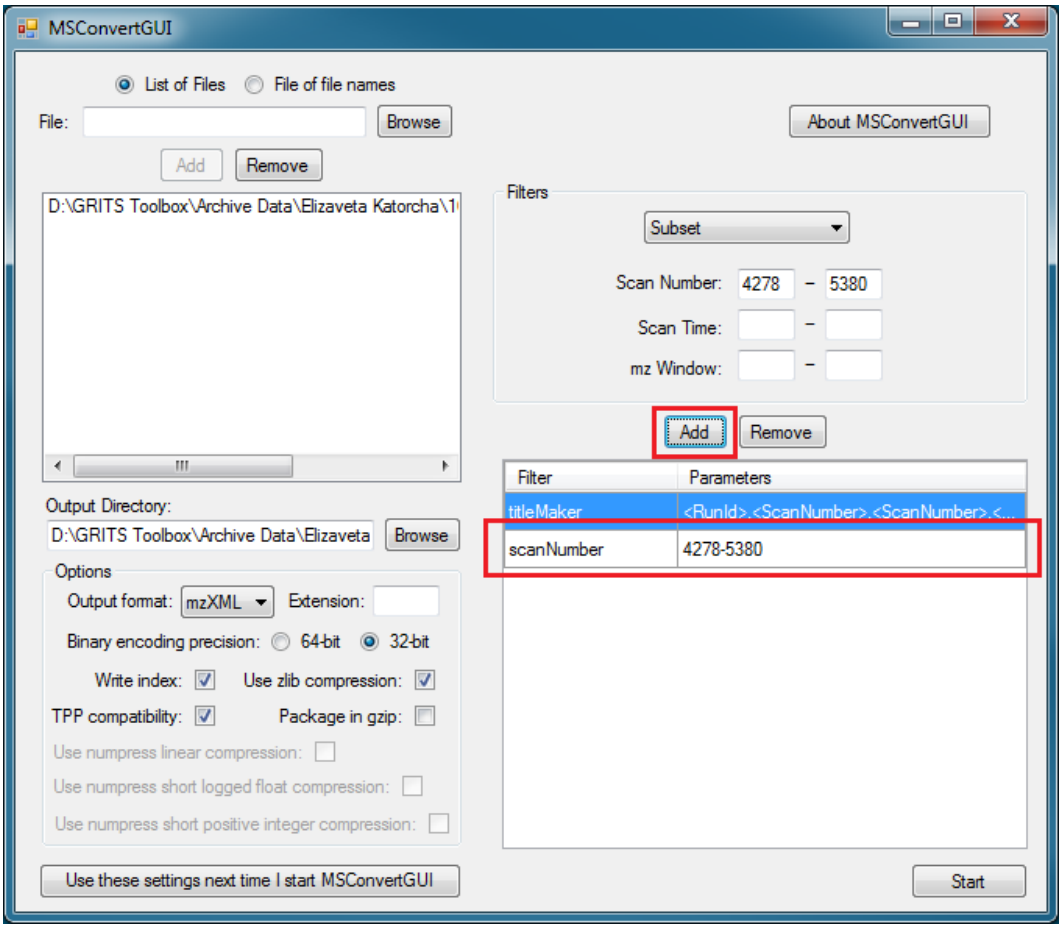
2. Configure to program to generate a new mzXML or mzML file. If you start with a mzXML/mzML file it is recommended to choose a different **Output Directory**.



3. Add a **Subset** filter to the dataset and specify the first and the last **Scan Number**. Please make sure that the start number (in the example screenshot 4278) is an MS1 scan. Otherwise GRITS may have problems processing the data if the scan hierarchy is inconsistent. To find the scan number based on the retention time open the data file in your vendor software or use SeeMS, a tool that is part of the ProteoWizard software. Go to the retention time you are interested in. If the scan at this time is a MS1 remember this number otherwise find the next MS1 before the selected scan. For TIM data you do not need to pick an MS1 scan only for MS/MS or LC/MS experiments.



4. Add the filter to the dataset by pressing the **Add** button.

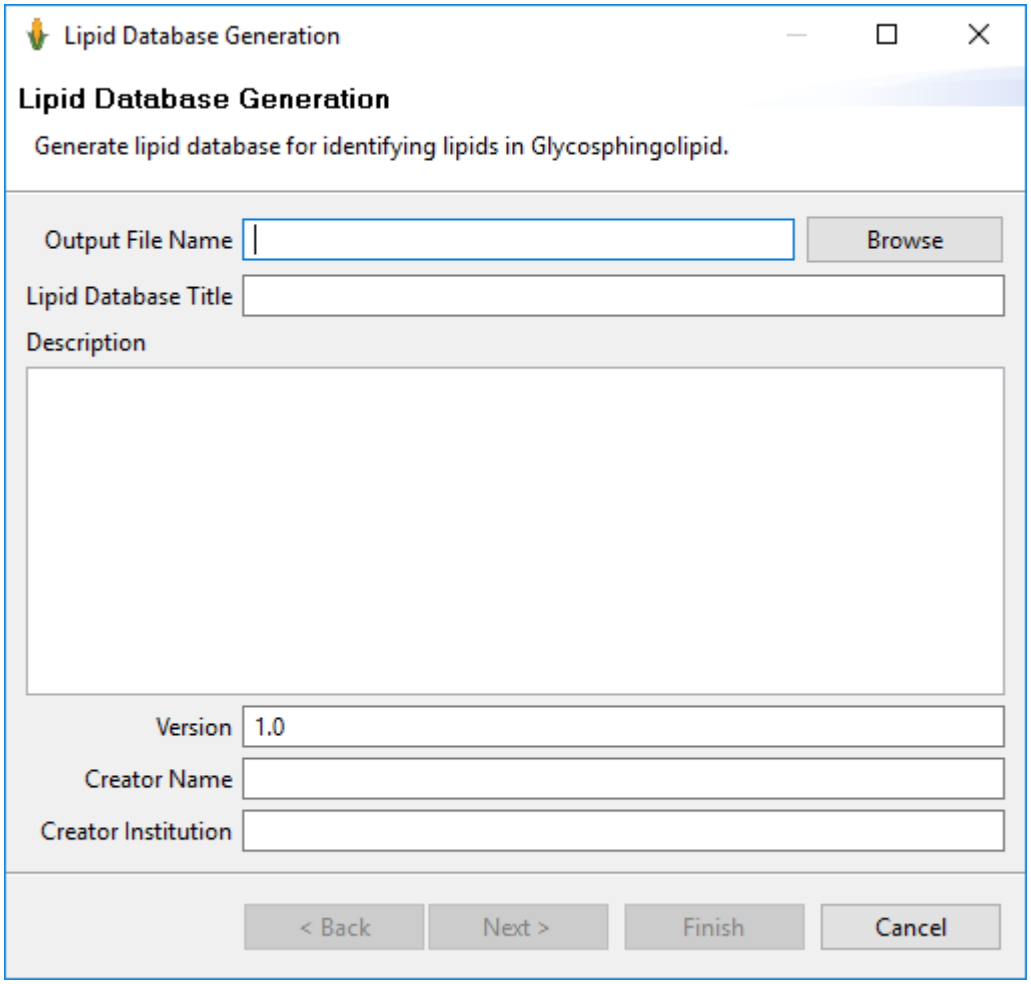


5. You can add additional filters or start the mzXML/mzML file generation by clicking the **Start** button. After the process is finished you will find the mzXML/mzML file in the specified output folder.

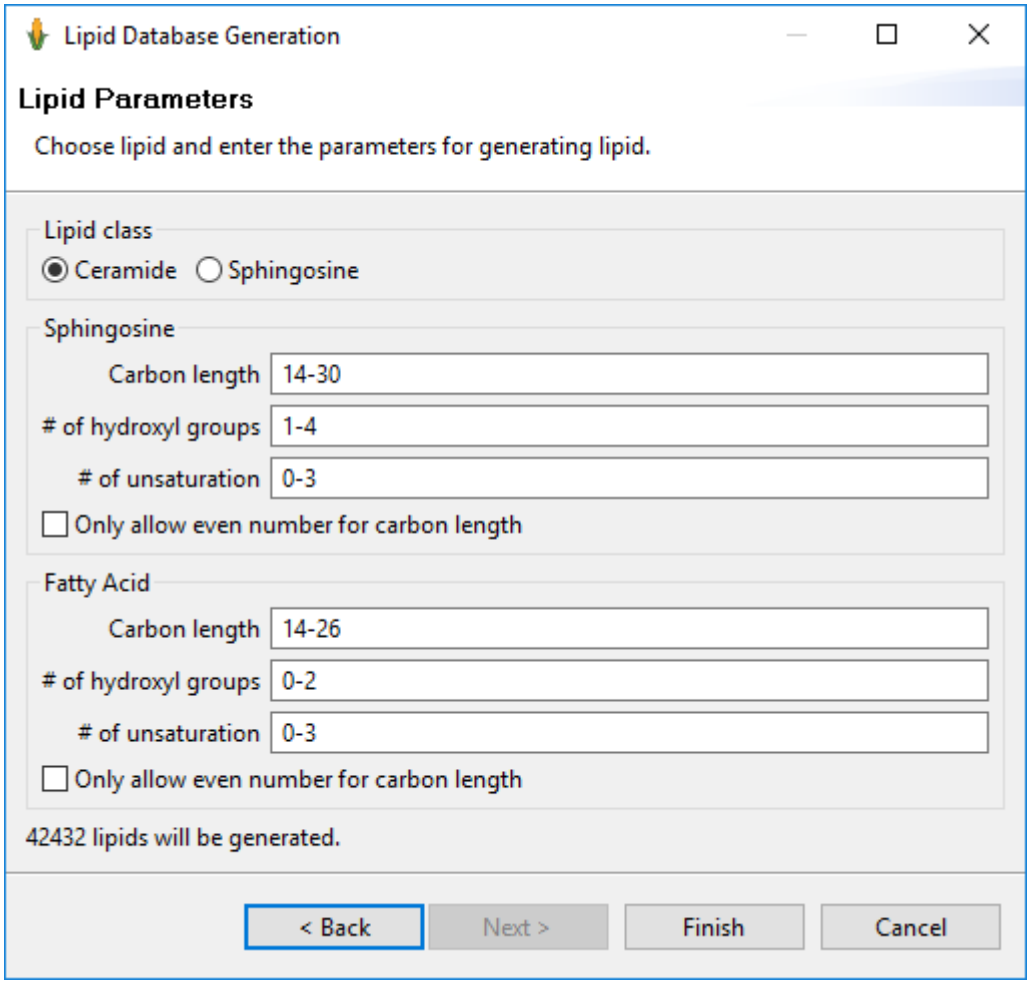
2.2 How to create a new lipid database

It is possible to specify a user defined lipid database. (See chapter 1.8.c) You can create a new lipid database using the *Tools* main menu

1. Click *Tools* main menu: *Glycolipid Tools* -> *Create Lipid Database*



2. Fill lipid database information.
    a. **Output File Name:** Mandatory. You can specify the file name to be used to save the created lipid database. It is also possible to use the *Browse* button below to use file save dialog.
    b. **Lipid Database Title:** Mandatory. Define a title of the lipid database.
    c. **Description:** Optional.
    d. **Version:** Optional.
    e. **Creator Name:** Optional.
    f. **Creator Instirution:** Optional
3. Click *Next* > button.



4. In this page, you can specify lipid type and its structure information for lipid generation. The number of the lipids to be generated is shown at the bottom.
    a. **Lipid class:** Select *Ceramide* or *Sphingosine*. Selected lipid will be generated. If you select *Sphingosine*, *Fatty Acid* box cannot be used (will be grayed out).
    b. Boxes of **Sphingosine** and **Fatty Acid:** You can specify the number(s) of structure information to be used for lipid generation. Each box corresponds to each lipid part. You can also specify the value(s) as a single number or range of numbers. For example, "14-30" means the numbers from 14 through 30. "2-5,8" means the numbers from 2 through 5 and 8. The generated lipids have all combinations of these structure values.
        i. Carbon length: Mandatory. Number of carbon length of the lipid. If the checkbox "**Only allow even number for carbon length**" is checked, odd numbers will be ignored even if specified.
        ii. # of hydroxyl groups: Number of hydroxyl groups to be contained in a generated lipid.
        iii. # of unsaturation: Number of double bonds to be contained in a generated lipid.
5. Click *Finish* button to create the lipid database.

6. After the generation is finished a new xml file and excel file for this generation are created in the specified folder. The xml file is used as a lipid database for the DANGO annotation. The excel file is just for confirming the generation results.