BioNumerics Tutorial:

Import and preprocessing of raw spectrum data

1 Aim

Comprehensive tools for the import of spectrum data, both raw spectrum data as processed spectrum data are incorporated into BioNumerics. In this tutorial we will focus on the import and preprocessing of raw spectrum data as mass-intensity list. Processing tools include resampling, background substraction, smoothing, peak detection etc.

2 Sample data

As an exercise, we will import a set of MALDI-TOF spectra from different isolates and species. This set can be downloaded from the Applied Maths website: go to http://www.applied-maths.com/download/sample-data and click on "Demo raw spectra". When the download is complete, unzip the file.

3 Preparing a sample database

3.1 Creating a new database

For this example we will be importing the raw spectra to a database with levels, so first we need to define a database with levels.

1. Double-click on the BioNumerics icon () on the desktop.
2. In the *BioNumerics Startup* window, press the  button to enter the *New database* wizard.
3. Enter a database name, e.g. "Demo spectra".
4. Click <***Next***> and then click <***Finish***>.

A new dialog box pops up, asking where the data should be stored.

The ***Create new*** option creates a new relational database on the local computer. Note that BioNumerics can only create new databases in SQLite, MS SQL Server Express (if installed with an earlier version of the software) or MS Access.

5. Leave the default option ***Create new*** enabled to create a new database and press <***Next***>.

The next window asks which database engine should be used for storing data:

- ***Use default SQLite***: The default SQLite option is a file-based database system with a theoretical storage limit of 32 TB.
- ***MS SQL Server Express® based***: Uses the MS SQL Server Express BioNumerics instance, which has a storage limit of 10 GB. This option is only available if the MS SQL Server Express database engine was installed previously with an earlier (7.0 - 7.5) BioNumerics version.

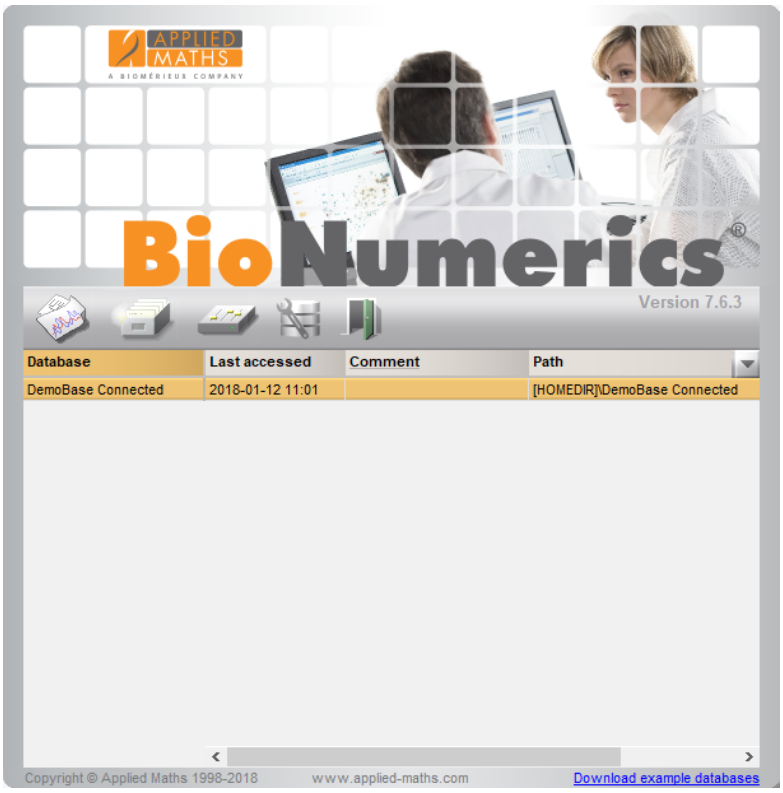

Figure 1: The *BioNumerics Startup* window.

- ***Use MS Access® based***: Uses the Microsoft Jet Engine (shipped with all Windows operating systems), which has a storage limit of 2 GB. This option is not compatible with 64-bit versions of BioNumerics.

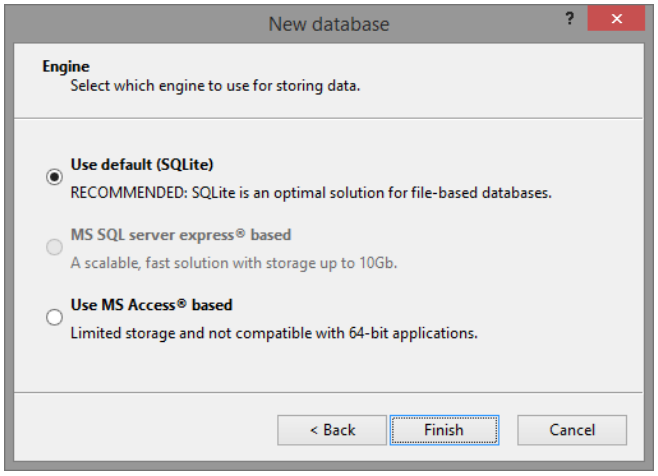

Figure 2: The *Database engine* wizard page.

6. Leave the default option enabled and press <***Finish***>.

The *Plugins* dialog box pops up which allows you to install additional functionality. This dialog can be called at any time from the *Main* window with ***File*** > ***Install / remove plugins...*** ( ).

7. Press <***Proceed***> to start BioNumerics.

The *Main* window opens with an empty database.

3.2 Creating a spectrum type experiment

Before importing spectrum data, we will first create a spectrum experiment type.

8. In the *Main* window, click on  in the toolbar of the *Experiment types* panel and select ***Spectrum type*** from the list. Press <***OK***>.

9. Enter a name, for example **Maldi**, leave the units for the horizontal and vertical axis at their defaults and press <***Next***>.

Three predefined preprocessing templates are included and a short description of the selected template can be found in the panel on the right: ***Preprocessing (Default)***, ***Preprocessing (Relaxed)***, and ***Preprocessing (Strict)***. The settings of each template can be changed in the *Spectrum Preprocessing* window (see 5) and saved to the database.

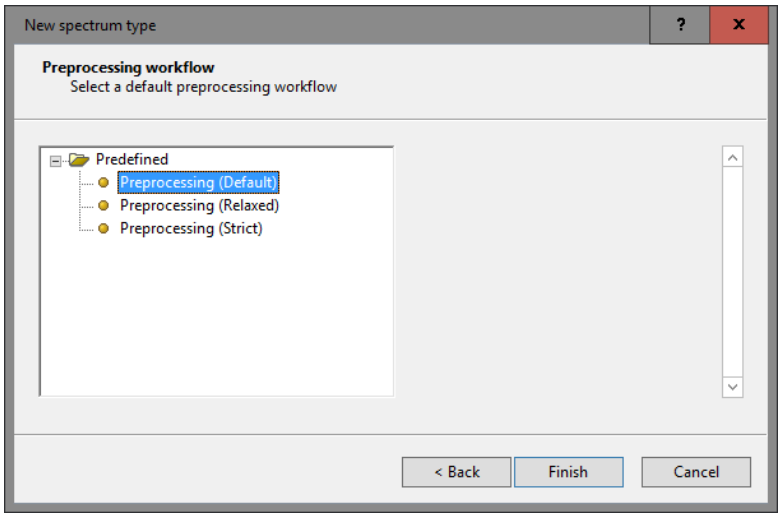

Figure 3: Preprocessing template.

10. Select the template *Preprocessing (strict)* and press <***Finish***> to complete the creation of the new spectrum type experiment.

The *Experiment types* panel now lists the spectrum type **Maldi** (see Figure 4).

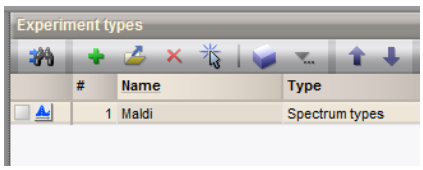

Figure 4: The *Experiment types* panel.

3.3 Creating levels

Next, we need to create the correct levels in the database. Database levels are schematically visualized in the *Database design* panel. This panel appears by default as a tab behind the *Database entries* panel.

When working with levels, it is easier to switch between levels if the *Database design* panel is docked above the *Database entries* panel:

11. Click on the *Database design tab* and - while keeping the mouse button pressed - drag it upwards in the *Database entries* panel. Drop the floating panel on the top part of the docking guide that appears.

The *Database design* panel is now shown above the *Database entries* panel.

12. In the *Database design* panel, click on ”All levels” and select ***Database*** > ***Levels*** > ***Add new level...*** (). This will show the *Level information* dialog box. Fill in **Species** and press <***OK***>.

13. Next, click on the new level ”Species” and select ***Database*** > ***Levels*** > ***Add new level...*** (), fill in **Isolate** and press <***OK***>. Finally, click on the new level ”Isolate” and select ***Database*** > ***Levels*** > ***Add new level...*** (), fill in **Raw spectra** and confirm.

After making the levels, the structure of the database should look like Figure 5.

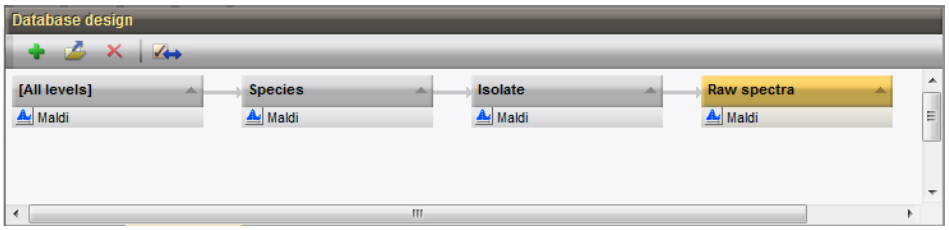

Figure 5: Database design after creating levels.

3.4 Creating information fields

Level-specific information fields can be created during the import of the spectra (see 4), during import of entry information from an external Excel or text file (see tutorial: ”Adding entry information”) or manually with the menu-items.

To manually add level specific information fields with the menu-items follow these steps:

- Click on the correct level in the *Database design* panel.
- Click on the *Entry fields* tab in the right upper corner in the *Main* window, and select ***Edit*** > ***Create new object...*** () in the *Entry fields* panel.
- Enter a name for the new information field and press <***OK***>.

Information in entry fields can be imported during import of the spectra when the information is present in the file names (see 4) or from an external Excel or text file (see tutorial: ”Adding entry information”).

4 Importing data

When importing data (descriptive information and/or experimental data) in a database with levels, it is important to highlight the corresponding level first. Typically, information will most often be imported at the deepest child level (i.e. **Raw spectra** in our example database).

1. Make sure the ’Raw spectra’ level is highlighted, select ***File*** > ***Import...*** (, **Ctrl+I**), highlight ***Spectrum type data*** > ***Import spectrum data*** and press <***Import***>.

2. Browse to the folder, select all files in this folder and press <***Open***> and <***Next***>.

3. In the *Import template* wizard page page of the wizard, press <***Create new***>.

The only source of information available in the newly created import template is the file name.

4. Double-click on the only row available in the grid or press <***Edit Destination***>.

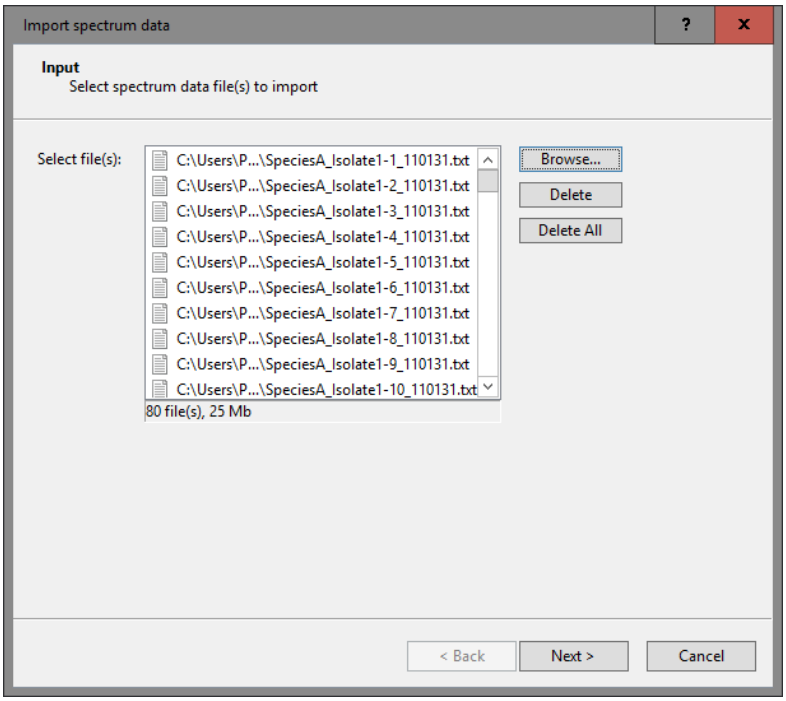

Figure 6: Select files.

5. Select **Raw spectra key** in the *Edit data destination* dialog box (see Figure 7).

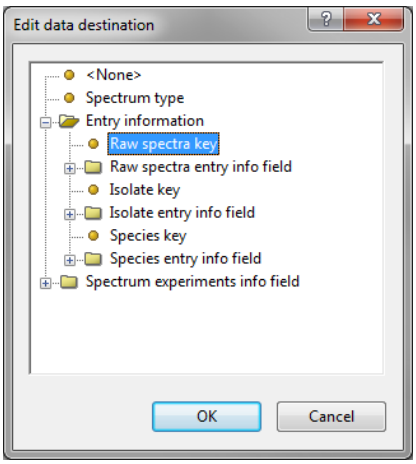

Figure 7: Select the destination

6. Press <***OK***>.

7. Visualize the advanced options for the *Import template* dialog box by clicking on the check box next to ***Show advanced options***.

8. Press <***Add rule***> to open the *Add data conversion rule* wizard.

9. In the first page of the *Add data conversion rule* wizard select ***File*** > ***Name*** (Figure 8) and press <***Next***>.

10. In the second page of the *Add data conversion rule* wizard, select **Isolate key** and press <***Next***> (see Figure 9).

11. In the *Data parsing* dialog box, fill in following data parsing string: "[DATA]-*". Press the <***Preview***> button (see Figure 10).

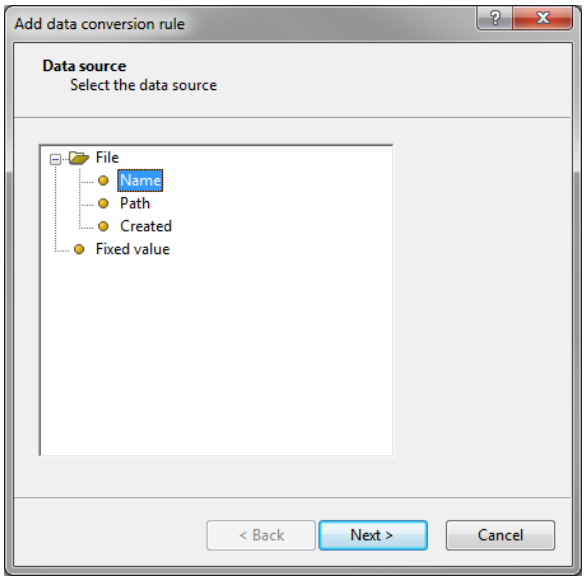

Figure 8: Adding a new rule to the import template.

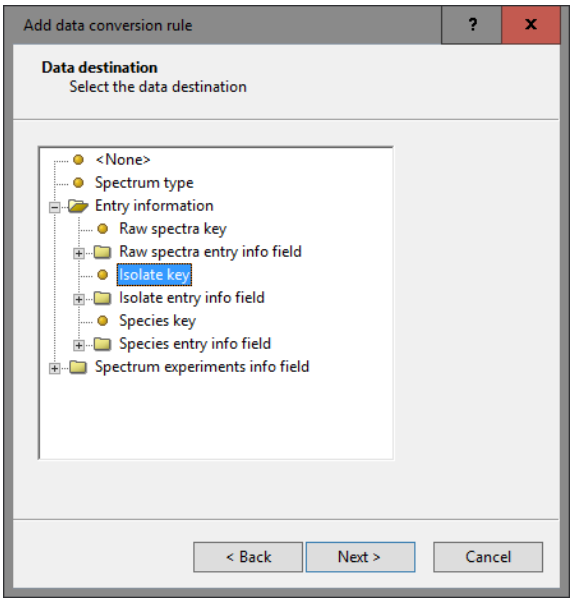

Figure 9: Isolate key.

12. When the information is parsed correctly press <***Next***> and <***Finish***>.

13. Press <***Add rule***> again to open the *Add data conversion rule* wizard.

14. In the first page of the *Add data conversion rule* wizard select ***File*** > ***Name*** and press <***Next***>.

15. In the second page of the *Add data conversion rule* wizard, select **Species key** and press <***Next***> (see Figure 11).

16. In the *Data parsing* dialog box, fill in following data parsing string: "[DATA]_*". Press the <***Preview***> button (see Figure 12).

17. When the information is parsed correctly press <***Next***> and <***Finish***>.

18. Press <***Add rule***> again to open the *Add data conversion rule* wizard.

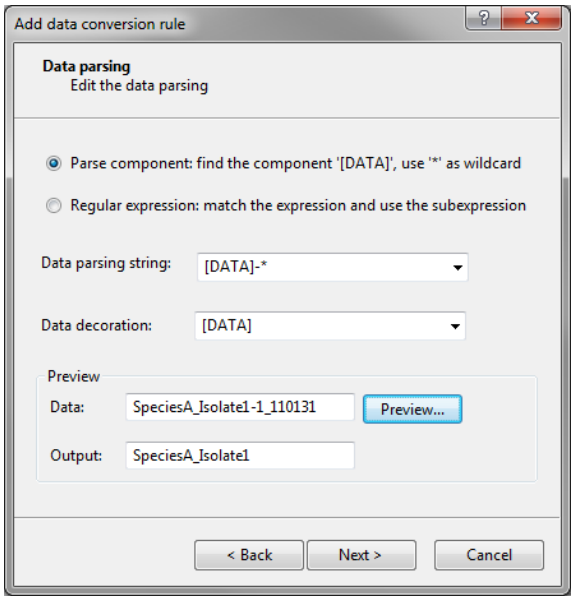

Figure 10: Parsing string.

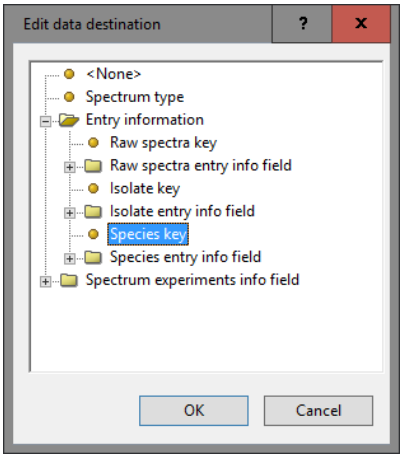

Figure 11: Species key.

19. In the first page of the *Add data conversion rule* wizard select ***File*** > ***Name*** and press <***Next***>.

20. In the second page of the *Add data conversion rule* wizard, select **Create new** under ***Raw sepctra entry info field*** and press <***Next***> (see Figure 13).

21. In the *Data parsing* dialog box, fill in following data parsing string: "*_*_[DATA]". Press the <***Preview***> button (see Figure 14).

22. When the information is parsed correctly press <***Next***>, specify a name for the new field (e.g. ***Date***) and press <***Finish***>.

The grid is updated.

23. In the *Import template* dialog box, press <***Preview***> and verify the preview of the import (see Figure 15).

24. Name the import template (e.g. "Import to levels") and optionally give it a description. Press <***OK***>.

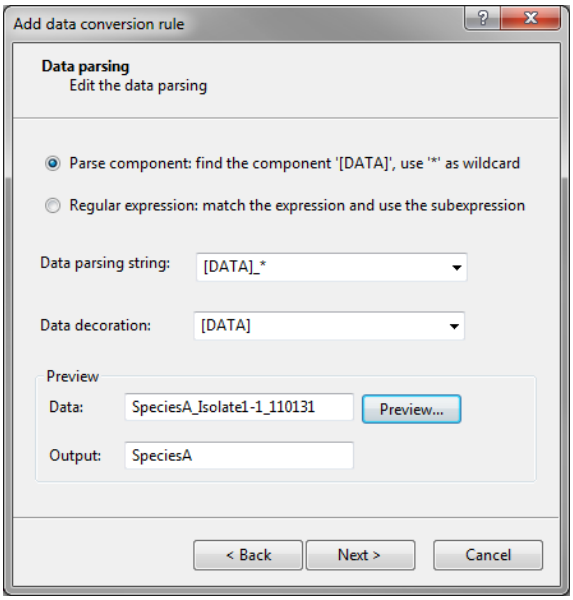

Figure 12: Species key.

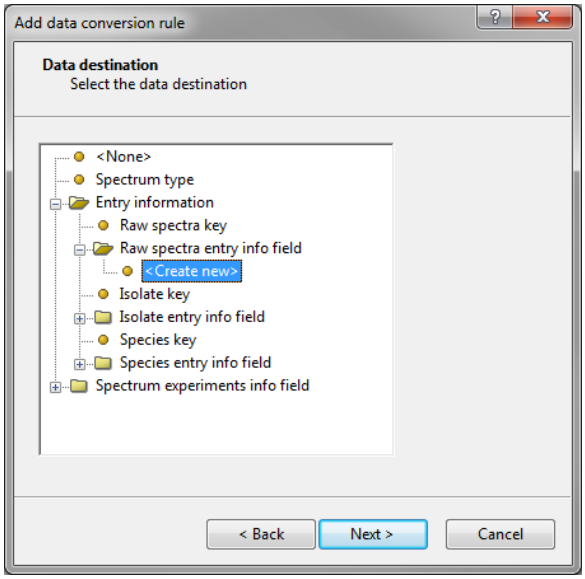

Figure 13: Create new information field.

25. With the new import template highlighted, and the **Maldi** experiment selected press <***Next***> to go the next step where an overview of the actions that will be performed during the import is displayed (see Figure 16).

26. Press <***Next***>, select **Preprocessing template (strict)** as ***Preprocessing*** option, leave all other settings at default and press <***Finish***>.

The raw spectra will be imported. Depending on the performance of your computer, this may take a few minutes.

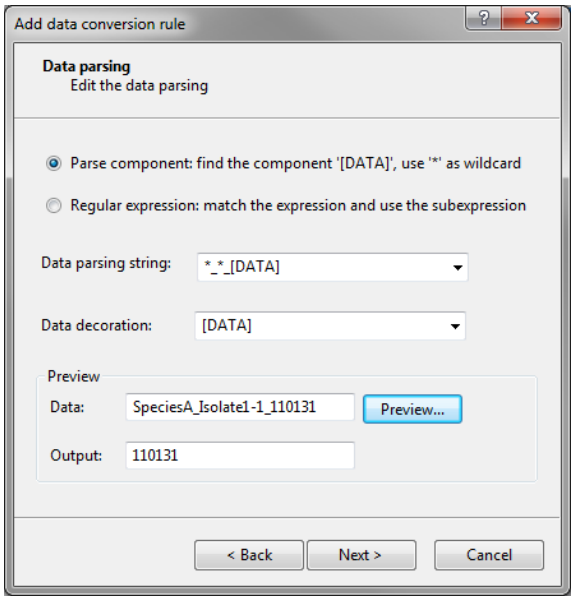

Figure 14: Parsing string.

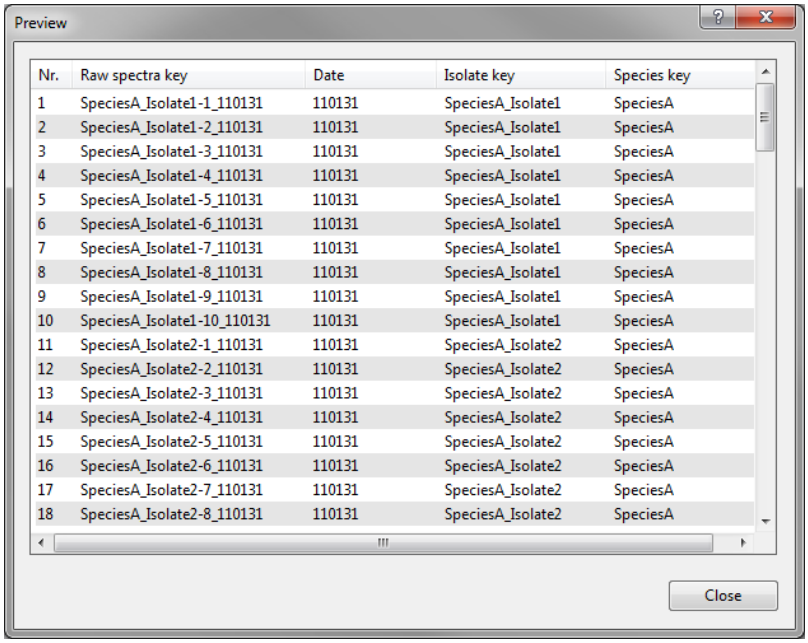

Figure 15: Preview.

5 Preprocessing of spectra

Upon completion of the import, the *Spectrum Preprocessing* window will open automatically. If it does not open automatically, you can select all entries on the lowest level "Raw spectra", and use ***Analysis*** > ***Spectrum types*** > ***Open preprocessing window...*** to open the *Spectrum Preprocessing* window.

All the imported spectra are loaded in the *Spectrum Preprocessing* window with the preprocessing template selected in the last step of the *Import spectrum data* wizard (see Figure 17). Because there is a slight difference in the range of the different spectra, we will be adding a trimming step to this template.

1. With the **Import & Resample** step of the preprocessing template highlighted, select ***Workflow*** > ***Show flow chart...***. This will open the *Workflow* window for this step of the template.

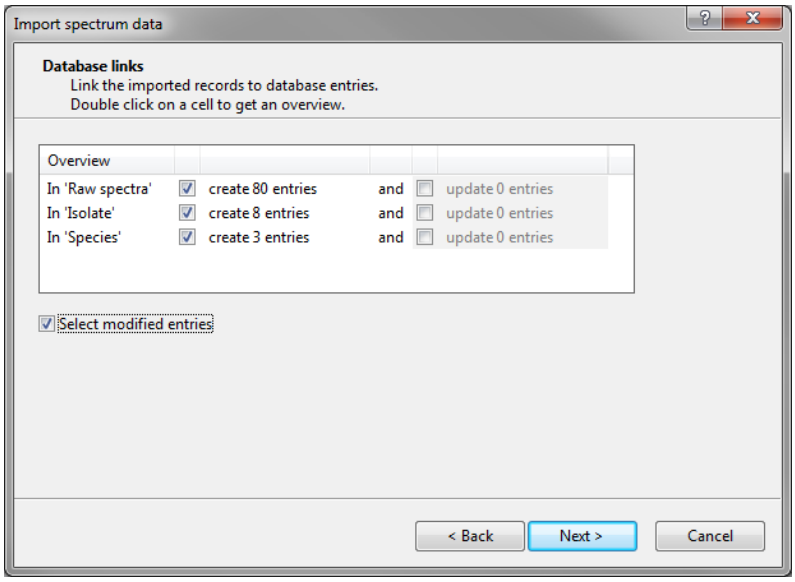

Figure 16: Import actions.

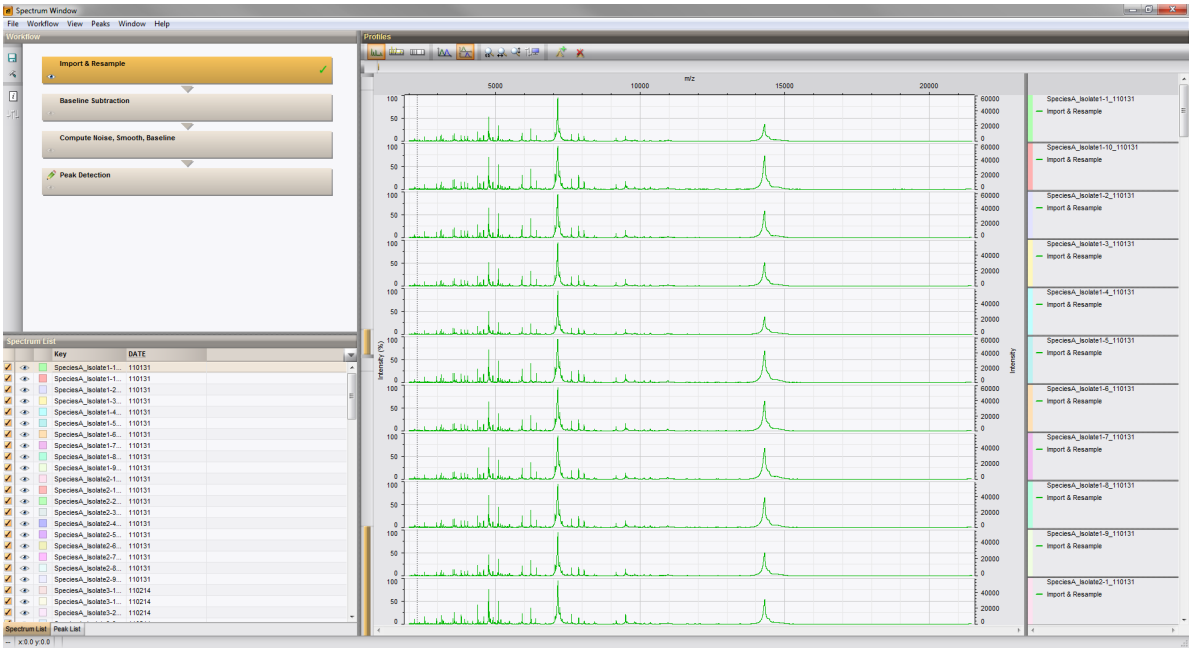

Figure 17: Preprocessing window after import of spectra.

2. In the left panel, highlight the operator ***Processing*** > ***X-axis*** > ***Trimming*** and select ***File*** > ***Add operator...*** ().

3. In the resulting dialog box (see Figure 18), fill in **2000** as X-axis minimum, do not use an upper limit and press <***Next***>. The trimming operator will be added to the end of the workflow, resulting in the workflow shown in Figure 19. After the operator has been added, close the *Workflow* window.

4. Select ***File*** > ***Save workflow as template...***, enter a name for the new template (e.g. "Strict with trimming") and press <***OK***>. This template will now be available for future import and preprocessing of spectra.

5. Press the last step of the preprocessing template, <***Peak Detection***> to execute the entire preprocessing workflow. Leave the limit for the signal to noise ratio at default and press <***Next***>. Depending on the performance of your computer, the execution might take a few minutes.

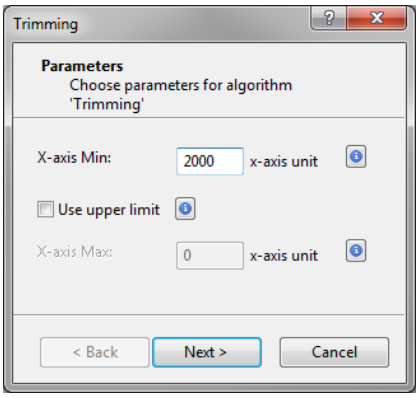

Figure 18: Parameters of the trimming operator.

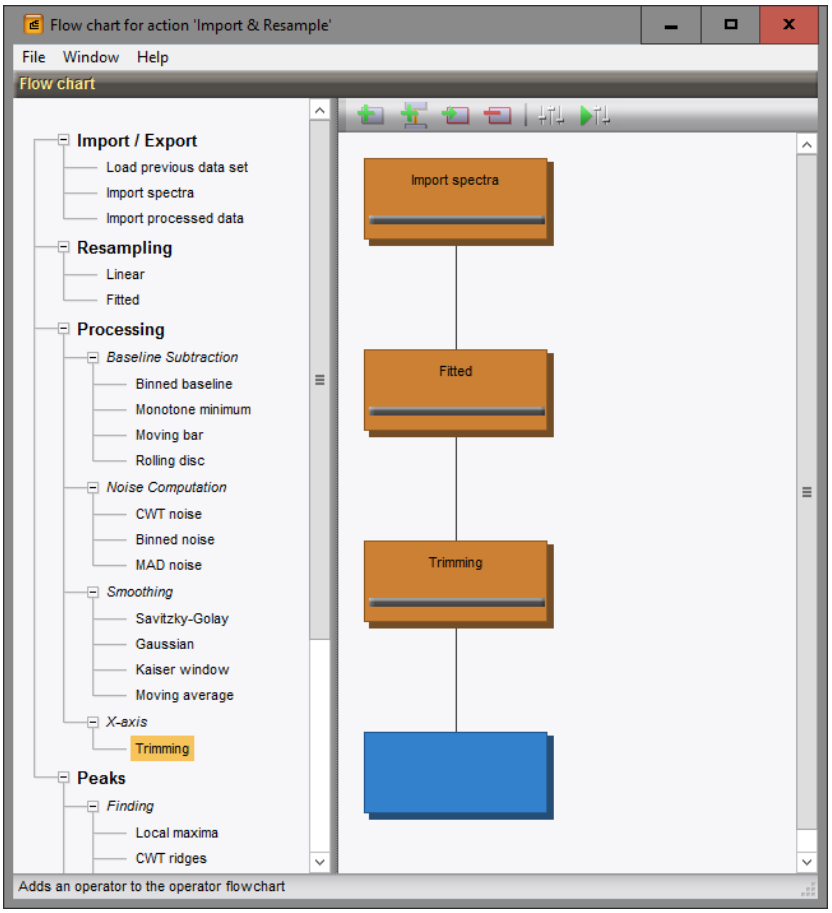

Figure 19: Final workflow with trimming operator.

6. When the spectra have been preprocessed, save the results by selecting ***File*** > ***Save spectrum data*** ( ). Close the *Spectrum Preprocessing* window.

The preprocessed spectra are now available in the database and further analysis, such as creating summary spectra or comparing the spectra, can be performed.