

Oligator

user manual - version 1.0.1

Virginie Lollier
contacts-bibs@inrae.fr

October 2020

Contents

| | | |
|----------|---|-----------|
| 1 | Oligosaccharide features | 3 |
| 1.1 | Topology design | 3 |
| 1.2 | Structure notation | 4 |
| 1.2.1 | Import a structure | 5 |
| 1.2.2 | Export a structure | 6 |
| 2 | Ose features | 7 |
| 2.1 | Structure description of oses | 7 |
| 2.1.1 | View controls | 7 |
| 2.1.2 | Cycle forms | 7 |
| 2.1.3 | Carbon features | 8 |
| 2.2 | Chemical substitutions | 8 |
| 2.2.1 | List | 9 |
| 2.2.2 | Definition | 10 |
| 2.2.3 | Glycan treatment | 11 |
| 3 | Theoretical spectrum production | 13 |
| 3.1 | Theoretical m/z | 13 |
| 3.2 | Theoretical intensities | 14 |
| 3.3 | Export | 15 |

Preamble

Oligator tool is composed of 2 categories of windows. The first one is the drawing space where elements constituting the oligosaccharide are viewed and assembled. Hydroxyl groups and hydrogens bound to monosaccharide carbons are not represented on the drawing space but implicitly considered.

The second one is a series of formular windows to refine the description of the structure and the resulting mass spectrum. The drawing space forms the main window whereas formular windows are displayed as pop-ups.

All computed masses correspond to monoisotopic masses of atoms. Charges are not expressed in the structure definition as the mass spectrum is built by selecting ionization mode of the neutral form of the drawn structure.

1 Oligosaccharide features

1.1 Topology design

An user can quickly draw the skeleton of an oligosaccharide with few controls. From the menu bar, the "Design" choice is used to define default ose type and default osidic linkage.

The add button, figure 1(a), of the drawing bar adds the image corresponding to the default ose type on the drawing space. An user can therefore add multiple units of the same type and then change the default value for a remaining set of units to add. The default ose can be either defined according to the cycle type or chosen in a name list of 14 standard monosaccharides.



Figure 1: Drawing bar buttons.

The remove button, figure 1(b), deletes all selected units (see also section 2.1.1) from the topology either linked or not to the polymer. The "del" key press from keyboard has the same effect. The "Clear all" option from the "Design" menu removes all units from the drawing space.

The grid button, figure 1(c), aligns connected units and disposes them into a grid according to their carbon bindings. It has no effect on free units without osidic bond.

The glycan sketching can be speed up when repeated motifs exist by copy and paste a selection of connected items, using context menu options popped-up by right mouse button clicking (figure 2). Copies are placed according to the cursor position of paste operation. Binding of the motif to the molecule is achieved by selecting and connecting the pair of oses implied in glycosidic bond.

The binding and unbinding of oses are managed by right-clicking a selection of units. The bind action connect carbons defined by default (from

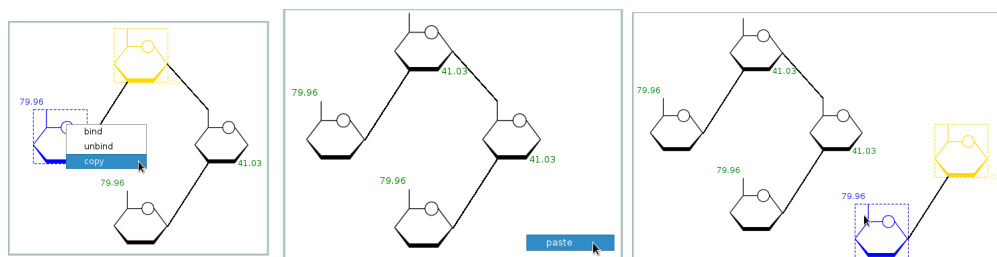


Figure 2: Copy and paste a selection.

"Design" menu). This assignment can be modified by double clicking the edge. Available carbons to connect are C#1 (for aldoses) or C#2 (for cetoses) at reducing end of oses and free carbons at other locations. Free carbons are those without chemical substitution and not yet implied in any osidic binding (figure 3). The reducing end of the oligosaccharide is on the right side of the draw space.

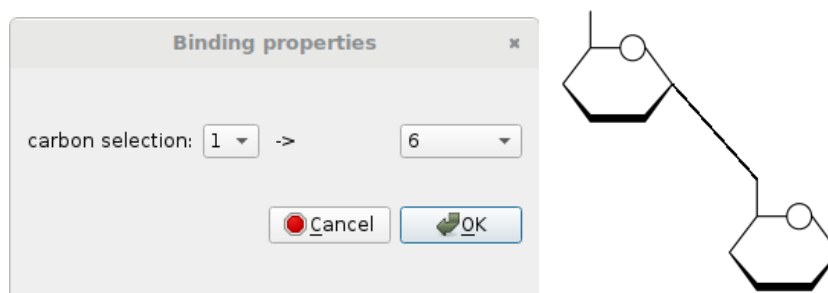


Figure 3: Carbons selection of an osidic binding.

1.2 Structure notation

From the "Notation" option of the menu bar, user can import and export text files containing the chemical description of an oligosaccharide structure in the form of linear notations.

The WURCS notation (Web3 unique representation of carbohydrate structures) is specific to glycans and compliant with the repository GlyTouCan. Oligator manages only the WURCS notations without ambiguity on atom composition and connections (i.e. atoms implied in rings and glycosidic bonds), but accepts potential lack of stereoisomerism description.

The SMILES notation (Simplified molecular-input line-entry system) is

unspecific to a type of compound and can be used in various chemical structure editors.

1.2.1 Import a structure

Oligator reads only the version 2 of the WURCS notation and SMILES from different sources (e.g. PubChem repository, ChemSketch editor). Paste notation from clipboard is unavailable.

When the notation is parsed, the glycan described by both notations is displayed on the drawing space and its detail is accessible by double-clicking its image. In the presence of a substituent unknown to Oligator, the resulting mass delta is related to a default naming and numbering: ukn#.

The figure 4 displays the GnGnX glycan from both the WURCS and SMILES notations extracted from the PubChem database (compound id : 57339199).

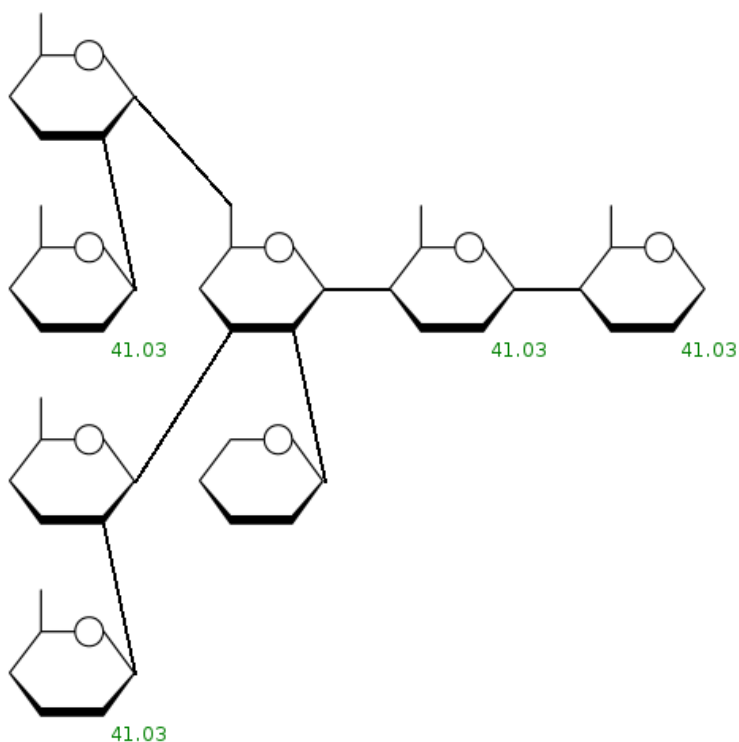


Figure 4: GnGnX compound from WURCS and SMILES notations.

1.2.2 Export a structure

Structures can be exported as both notations SMILES and WURCS except when the oligosaccharide carries cyclic substituent or anhydric bond. In that case, only SMILES notation export is possible. Implementation of WURCS notation follows the rules published by Tanaka, Kenichi *et al.*.

Implementation of SMILES follows the OpenSMILES specifications. Nested branches are used during the parsing of the molecule that always starts from its reducing end (e.g. "C1(C(C(C(C(O1)CO)O)O)O)O" for hexopyranose). Aromaticity is represented in the Kekule form (C1=CC=CC=C1). The lower-case letters notation is not implemented in the current version of Oligator.

2 Ose features

2.1 Structure description of oses

Oses are represented through 4 elementary images: hexofuranose, hexopyranose, pentofuranose and pentopyranose.

2.1.1 View controls

Within the draw space, ose controls are accessed through the selection of images. The selection is viewed as dot lines around ose image. There are 3 ways to select oses:

- by clicking on a single image,
- by drawing a rectangular area around a group of images,
- or by pressing the ctrl key + clicking on images one by one.

The selected images can be further dragged and dropped together across the drawing space. Clicking outside any image deselects all oses. Double-clicking on an image triggers the pop-up of ose detail window (figure 5).

| Ose description | | |
|---------------------|----------------|---|
| Number of carbons : | 6 | Ring start : 1 |
| | | Form : pyrane |
| Isomers: | Substituents : | |
| C1 | L | hydroxy |
| C2 | D | hydroxy |
| C3 | L | hydroxy |
| C4 | D | hydroxy |
| C5 | D | hydroxy |
| C6 | | sulfate |
| b-dGlc | | |
| | | <input type="checkbox"/> 3,6-anhydro bond |
| | | Cancel OK |

Figure 5: Customization of ose structure

2.1.2 Cycle forms

An ose is displayed as images derived from the Haworth description with implicate OH groups (figure 6). As a consequence, the stereo-isomerism of carbons is not directly displayed (no incidence on molecule mass). However, when it is assigned (using primer oses from the default list, described in

subsection 1.1, or defined from ose detail), a selected ose is highlighted in color according to the epimer affiliation.

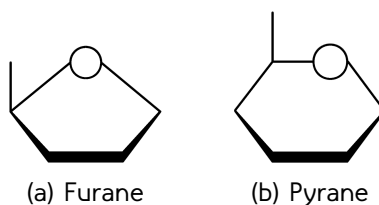


Figure 6: Display of cyclic forms of oses.

Cyclic form can be changed from the detail window (figure 5). By the same way, the total number of carbons constituting the monosaccharide and the carbon starting the cycle can also be changed. A vertical line on the left of the carbon list indicates which ones are in the cycle.

2.1.3 Carbon features

For each chiral carbon, the stereochemistry can be assigned according to Fisher projection where D indicates that OH group is on the right of the carbon and L, on its left. For each free carbon, i.e. unrelated to cycle closure or any osidic bond, a substituent can be added. Non-substituted ose has hydroxyl on each carbon.

2.2 Chemical substitutions

Numbers at carbon locations indicate chemical substitutions of OH groups resulting in mass gain (green) or loss (red) on the molecule (figure 7).

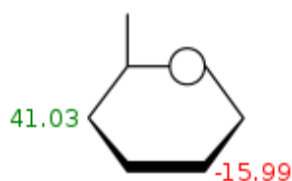


Figure 7: Ose displaying desoxydation of carbon #2 and O-acetylation of carbon #4.

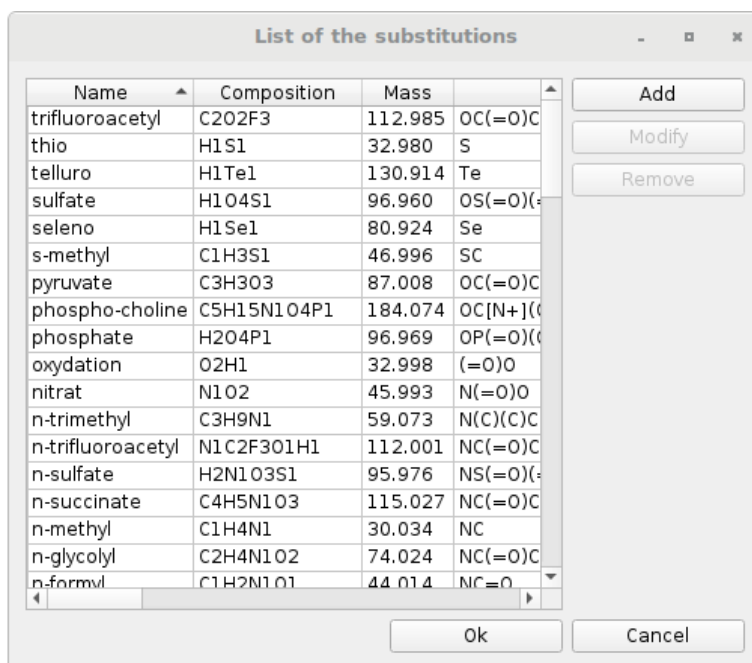
2.2.1 List

A substituent list is available from the menu and can be edited by user during the session from the menu option named "Substituents".

This list is a manual extraction of monosaccharidDB with the addition of ferulic acid (found in plants). Anhydro bridge found in algae is a particular ose modification that is presented in Oligator as a check box in ose detail rather than as a substituent of the list.

Displayed as a table where column names are "Name", "Composition", "Mass", "SMILES" and "Linker" (figure 8), the list can be ordered according to one of these columns. The user can add, edit or remove any line of the table.

The substituent list is stored with the Oligator code as a json file in the Ressources directory. Although not recommended, this format enables text edition to store durably additional substituents.



| Name | Composition | Mass | SMILES |
|-------------------|-------------|---------|-------------------------|
| trifluoroacetyl | C2O2F3 | 112.985 | OC(=O)C(F)(F)F |
| thio | H1S1 | 32.980 | S |
| telluro | H1Te1 | 130.914 | Te |
| sulfate | H1O4S1 | 96.960 | OS(=O)(=O)O |
| seleno | H1Se1 | 80.924 | Se |
| s-methyl | C1H3S1 | 46.996 | SC |
| pyruvate | C3H3O3 | 87.008 | OC(=O)C(=O)O |
| phospho-choline | C5H15N1O4P1 | 184.074 | OC[N+](C)(C)COP(=O)(O)O |
| phosphate | H2O4P1 | 96.969 | OP(=O)(O)O |
| oxydation | O2H1 | 32.998 | (=O)O |
| nitrat | N1O2 | 45.993 | N(=O)O |
| n-trimethyl | C3H9N1 | 59.073 | N(C)(C)C |
| n-trifluoroacetyl | N1C2F3O1H1 | 112.001 | NC(=O)C(F)(F)F |
| n-sulfate | H2N1O3S1 | 95.976 | NS(=O)(=O)O |
| n-succinate | C4H5N1O3 | 115.027 | NC(=O)CC(=O)O |
| n-methyl | C1H4N1 | 30.034 | NC |
| n-glycolyl | C2H4N1O2 | 74.024 | NC(=O)CO |
| n-formyl | C1H2N1O1 | 44.014 | NC=O |

Figure 8: Substituent list

2.2.2 Definition

When adding or editing a substituent, the user has to fill in the **canonical SMILES notation** beginning by the atom linker to use carbon (figure 9). This notation can be collected from the pubchem database, or wikipedia web site or generated from molecular editors like JChemPaint, ChemSketch... In order to help the user to select the atom linker, the graph of atom is displayed according to the SMILES notation filled. Any atom can be the use carbon linker according to its valence and is highlighted in green on the graph when binding is possible, in red otherwise. The SMILES notation is reordered according to the user selection. The "refresh view" button rearranges the nodes of the graph and can be used when overlaps occur.

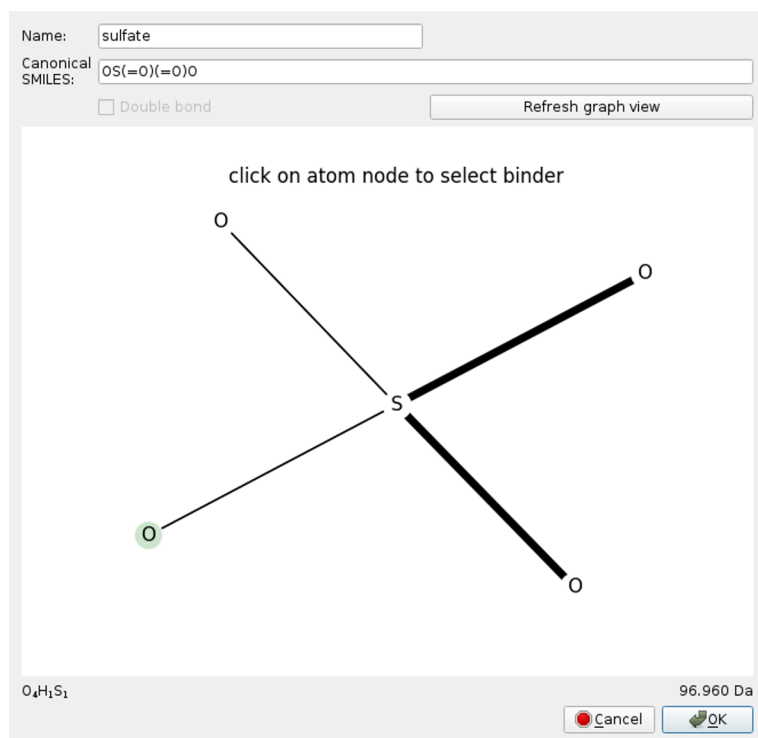


Figure 9: Substituent edition

The chemical name if the substituent is optional and randomly produced when missing. The atom composition is displayed at the left bottom of the window, minus one Hydrogen for use binding (or 2 for a double bond) and the monoisotopic mass derived from this composition is displayed at the right bottom. The mass displayed on the drawing space of the oligosaccharide corresponds to the difference between this substituent mass and OH mass.

2.2.3 Glycan treatment

It is possible from the menu bar to substitute in a single step all free OH groups of oses by a selected substituent. Substituents carrying free OH already assigned on any oses, are not modified. Users has to define additional substituents derived from their OH substitution(s).

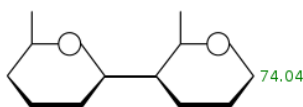


Figure 10: di-saccharide with a glycerol on its reducing end.

As an example, the per-methylation of a built glycan having a glycerol present on its reducing end (figure 10), implies to create a per-methylated glycerol in the substituent list in addition to the glycan treatment by methyl (figures 11 and 12).

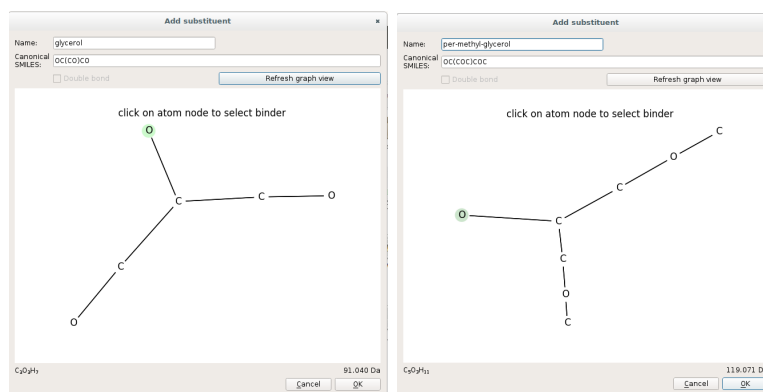


Figure 11: Substituent definition of glycerol and its methylated version.

The inverse step is possible of removing all substitutions of a selected substituent name from the glycan. Again, this does not affect other substituents on glycan. From the previous exemple, user has to manually change the substituant on the glycan reducing end by glycerol.

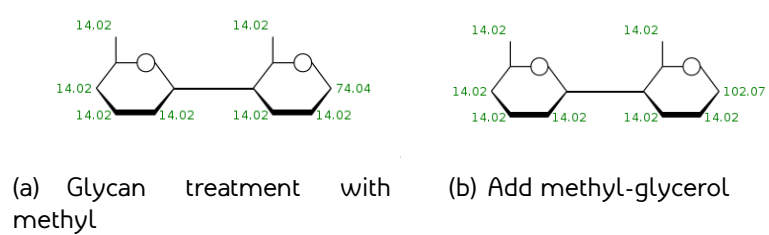


Figure 12: Per-methylation of a substituted di-saccharide (glycerol on its reducing end).

3 Theoretical spectrum production

The specification of mass and intensities is made in the "MS/MS spectrum preview" option from the "MS/MS" menu bar (figure 13). By default, the theoretical spectrum is produced from the model of a CID+ experiment. Any changes in the preview window will affect the export of the spectrum.

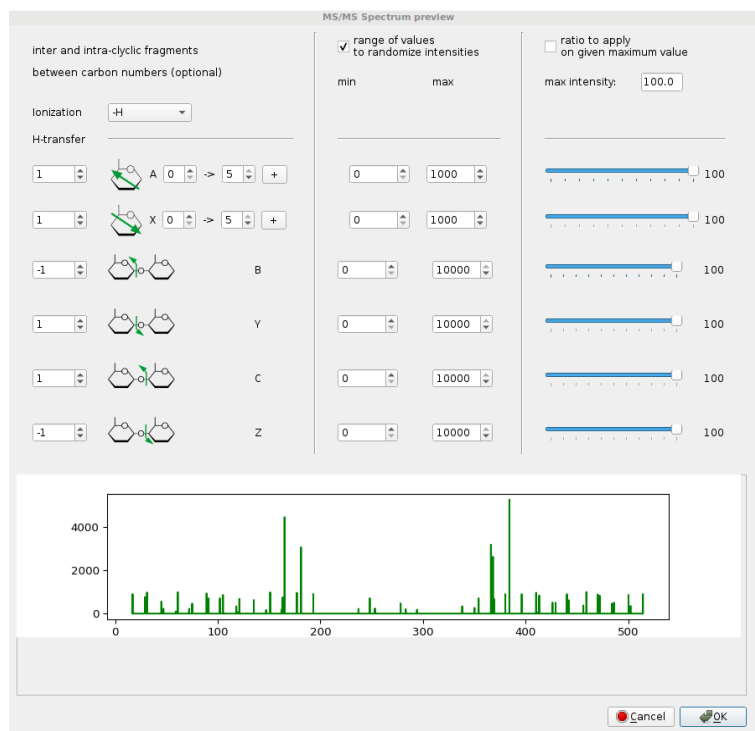


Figure 13: MS/MS spectrum preview.

3.1 Theoretical m/z

The spectrum produced from a theoretical MS/MS fragmentation of oligosaccharide structure relies on the standard nomenclature published by Domon & Costello. Ion types coming from this nomenclature identifies the location of fragmentation between oses (ions Y and Z for the reducing end side, B and C on the other side) or within cycles (ion X for reducing end side, A, for other side). Branching "numbering" with greek letters and quotes is also achieved according to this standard.

Oligator handles singly charged precursor only. User can select +H or -H ionization and alkali cationization (figure 14).

Rearrangements of fragments leading to H-transfers are assigned by default according to the fragmentation pathway described by Domon & Costello. User can adjust the number of H transferred if needed. For these types, parameters can be defined for all fragments occurring between oxygen and carbon#5 or for some specific pairs of carbons ("plus" button). The "– >" symbol means "from carbon i to carbon j", whereas the "-" letter means between carbons i and j.

inter and intra-cyclic fragments
between carbon numbers (optional)

ionization +H

H-transfer

| | | | | | | | | | | |
|----|--|---|---|---|---|---|---|---|---|---|
| 1 | | A | 0 | - | 0 | + | 5 | - | 5 | + |
| 1 | | X | 0 | - | 0 | + | 5 | - | 5 | + |
| -1 | | B | | | | | | | | |
| 1 | | Y | | | | | | | | |
| 1 | | C | | | | | | | | |
| -1 | | Z | | | | | | | | |

Figure 14: Parameters influencing the m/z values of ion types.

Masses are computed according to masses of monoisotopic atoms most frequent in the nature.

3.2 Theoretical intensities

Specifying the distribution of intensities is a way to describe the probability of observing ion types. As for example, ions produced by CID experiments at low dissociation energies displays less A and X ions than other types. As for masses, the intensities can be defined for each ion type and each of sub-types coming from the multiple combinations of carbon pairs from internal ose fragmentation.

There are 2 modes to define how intensities are assigned (figure 15):

- randomly between min and max bounds,
- or relatively to a maximum value.

The interface is divided into two columns. The left column is titled 'range of values to randomize intensities' and contains a 'min' and 'max' header. Below this are six rows, each with a 'min' input field (all set to 0) and a 'max' input field (values of 1000, 1000, 10000, 10000, 10000, and 10000). The right column is titled 'ratio to apply on given maximum value' and contains a 'max intensity' input field (set to 100.0) and six horizontal sliders, each with a range from 0 to 100.

Figure 15: Parameters influencing the distribution of intensities.

3.3 Export

The exported spectrum is a simple text file including the SMILES notation of the oligosaccharide and ionization mode as headers followed by the list of ions. Each ion line contains its m/z value, intensity and annotation in a LaTeX syntax. These values are separated by a tabulation.