Oligator
user manual - version 1.0.0

Virginie Lollier
contacts-bibs@inrae.fr

October 2020

Contents

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 implicitely 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.

(a) add (b) remove (c) align

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 binding and unbinding of oses are managed by right-clicking a selection of units. The bind action connect carbons defined by default (from "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 2). The reducing end of the oligosaccharide is on the right side of the draw space.

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

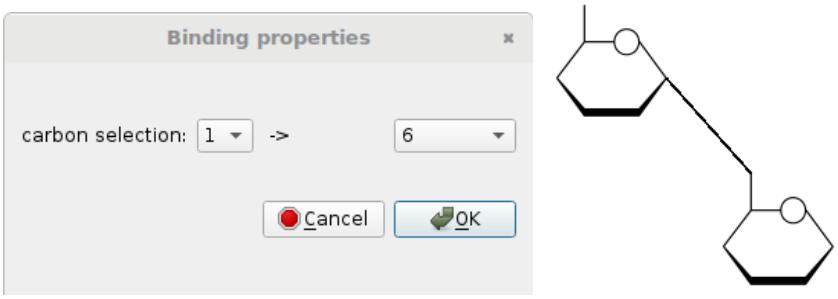


Figure 2: Carbons selection of an osidic binding.

the form of linear notations. The WURCS notation (Web3 unique representation of carbohydrate structures) is specific to glycans and compliant with the repository GlyTouCan. 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 ose 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 3 displays the GnGnX glycan from both the WURCS and SMILES notations extracted from the PubChem database (compound id : 57339199).

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

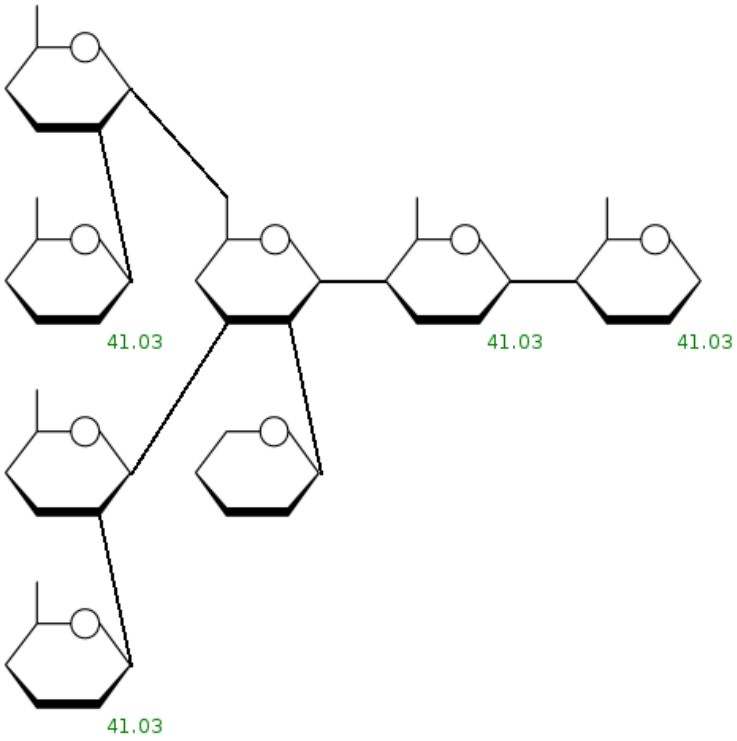


Figure 3: GnGnX compound from WURCS and SMILES notations.

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 lowercase 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 accross the drawing space. Clicking outside any image deselects all oses. Double-clicking on an image triggers the pop-up of ose detail window (figure 4).

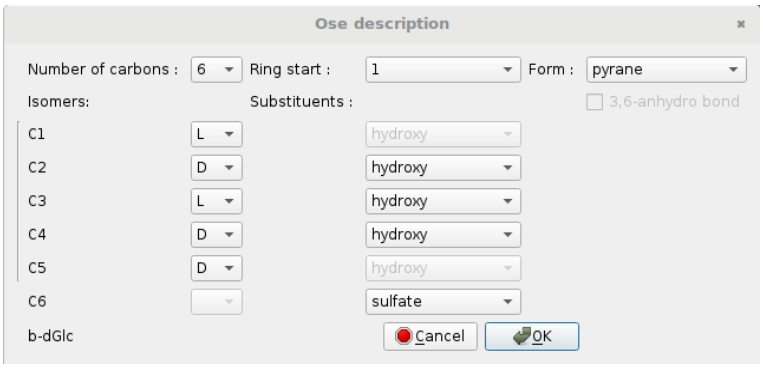
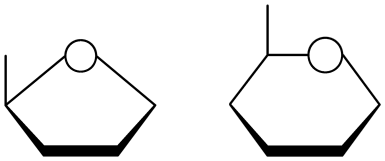


Figure 4: Customization of ose structure

2.1.2 Cycle forms

An ose is displayed in either as images derived from the Haworth description with implicite OH groups. As a consequence, the stereo-isomerism of carbons is not displayed (no incidence on molecule mass).



(a) Furane (b) Pyrane

Figure 5: Display of cyclic forms of oses.

Cyclic form can be changed from the detail window (figure 4). 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 6).

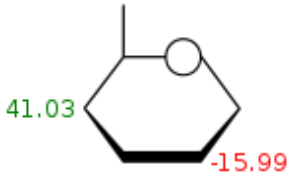


Figure 6: Ose diplaying desoxydation of carbon #2 and N-acetylation of carbon #4.

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 7), 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.

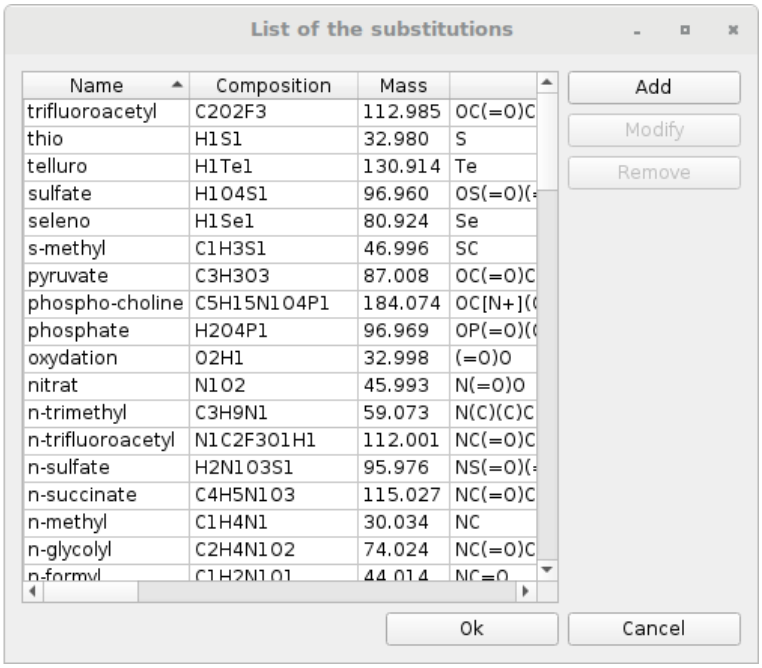


Figure 7: Substituent list

2.2.2 Definition

When adding or editing a substituent, the user has to fill in the SMILES notation beginning by the atom linker to ose carbon, or, at least, the atom composition (figure 8). The chemical name is optional and randomly produced when missing.

The list of linkers is filled according to the atom composition or to the SMILES notation. In addition, the linker "-" is also listed for the case of pair of atoms is bound to ose carbon. The formula field corresponds to the atom composition of the substituent without one Hydrogen atom removed for ose binding (2 for a double bond) and is auto-filled from the SMILES field. The mass is automatically computed from the formula and displayed as a difference to OH mass on the drawing space of the oligosaccharide.

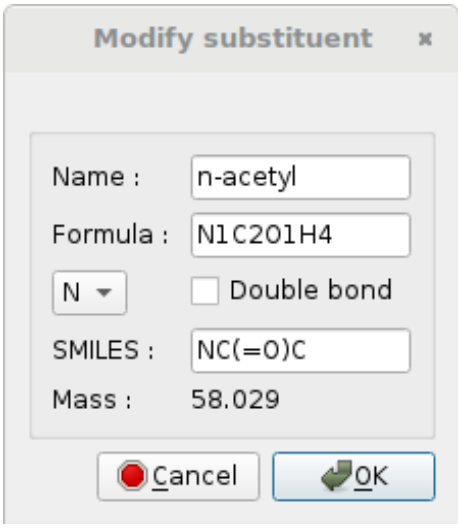


Figure 8: Substituent edition

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 9). 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.

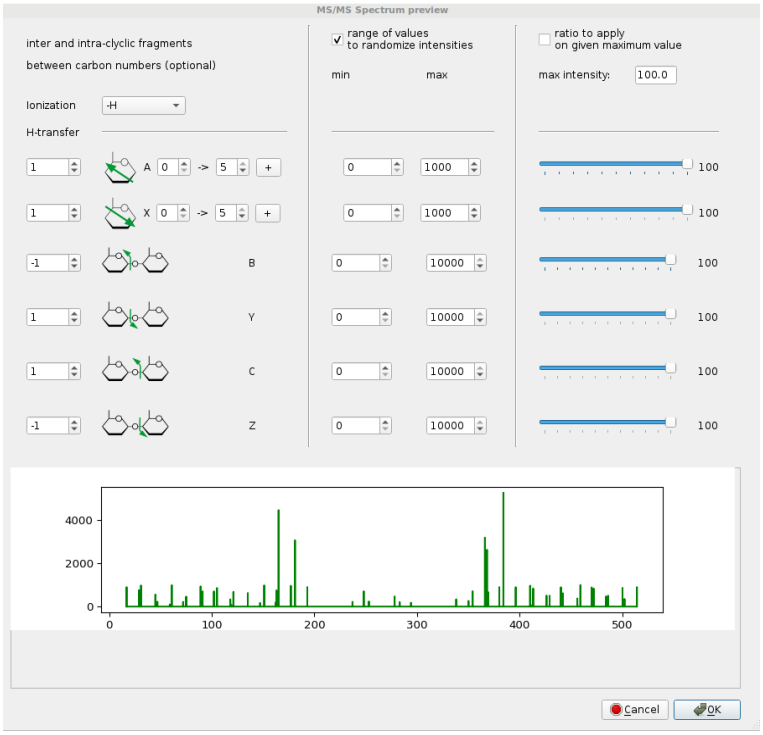


Figure 9: MS/MS spectrum preview.

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 10).

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.

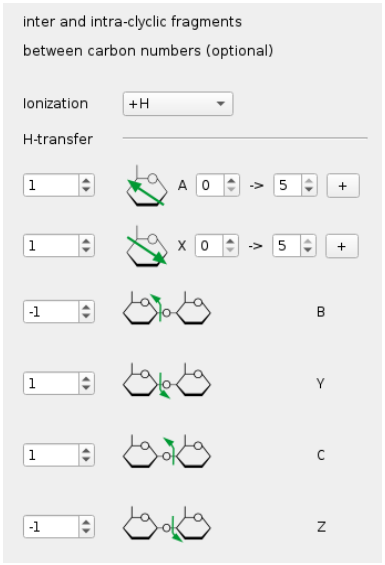


Figure 10: Parameters influencing the m/z values of ion types.

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 probablity of observing ion types. As for example, ions produced by CID experiments at low disociation 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 subtypes coming from the multiple combinations of carbon pairs from internal ose fragmentation.

There are 2 modes to define how intensities are assigned (figure 11):

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

Figure 11: 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.