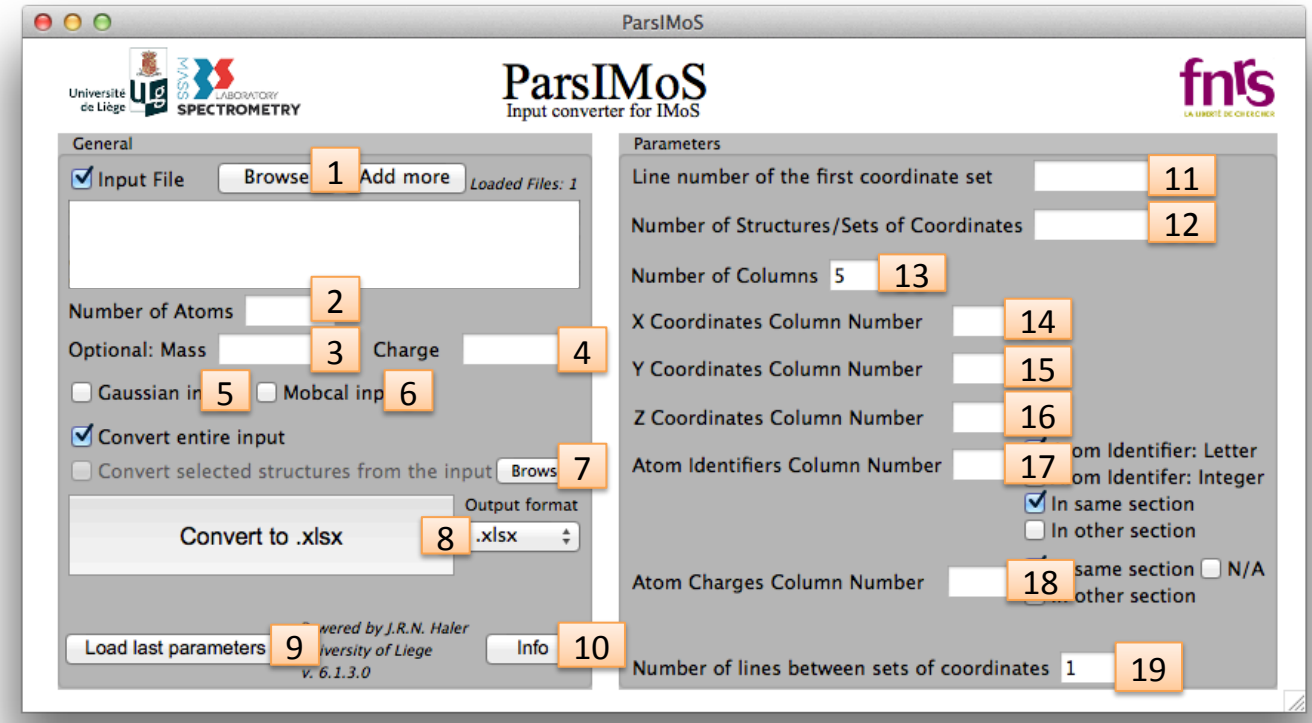
ParsIMoS Tutorial

Input converter for IMoS



*Annotated default view of **ParsIMoS v. 6.1.3.0***

1. Browse for input file(s). You can load multiple input files if they have all to be parsed the same way, using the same parameters (except for MOBCAL inputs, see below). If the button is clicked a second time after loading first inputs, only the newly loaded inputs chosen from the new file-browser prompt will be taken into account for parsing.
   The input format is not restricted to a format. If, however, parsing should fail with a given input format, try converting your file into a *.txt* format. Special characters may cause problems.
   Once you have provided first inputs, you can add additional inputs with the '*Add more*' button. The path of these new inputs can then differ from the other input file paths.
   The loaded files and file paths will show in the scrollable window below the *Browse* button, proceeded by the comment '*IN n*'. '*n*' represents the counter of loaded inputs.
   All selected inputs can be cleared from the list by unchecking the *Input File* checkbox.
   Note that you can add special characters as *Delimiters* to the library file '*Lib_Delimiters.par*'. For further details on where to find and how to encode the '*Lib_Delimiters.par*' file, refer to the section ***ParsIMoS** Modifying libraries*.

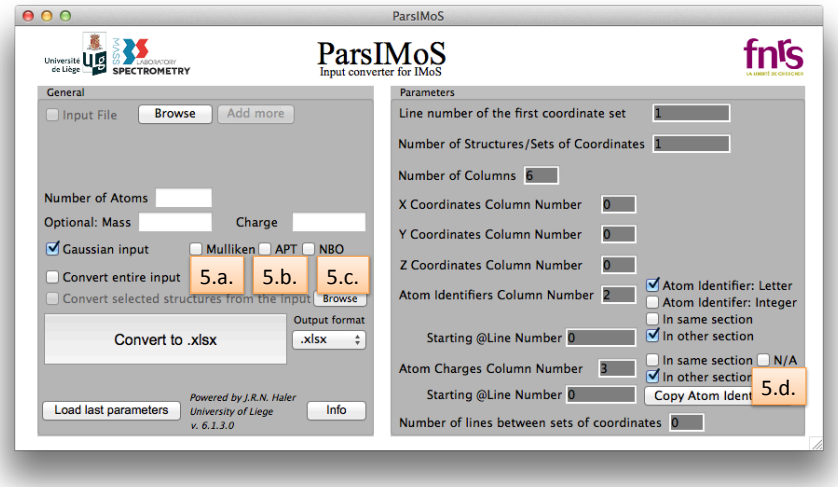
2. You have to provide the number of atoms of your molecule/ion (except for MOBCAL inputs, see below). **ParsIMoS** calculates the positions of the different sets of coordinates in the input file according to the number of atoms.

3. You can provide the mass of the molecule/ion. This parameter is optional. If the field was left *blank* or if *0* was encoded, **ParsIMoS** will calculate the mass according to the encoded atoms (using rounded integer values).
   The rounded integer values can be read and modified in the library file '*Lib_Atoms_VDW.par*'. Be aware of the automatic mass calculation for unknown atoms. They will be encoded using *400*. **ParsIMoS** will display a warning when encoding such an atom.
   Note that you can modify the library file '*Lib_Atoms_VDW.par*'. For further details on where to find and how to encode the file, refer to the section ***ParsIMoS** Modifying libraries*.

4. You can provide the charge of the ion. This parameter is optional. The charge will, however, not be calculated. If it is not provided, *0* is encoded.

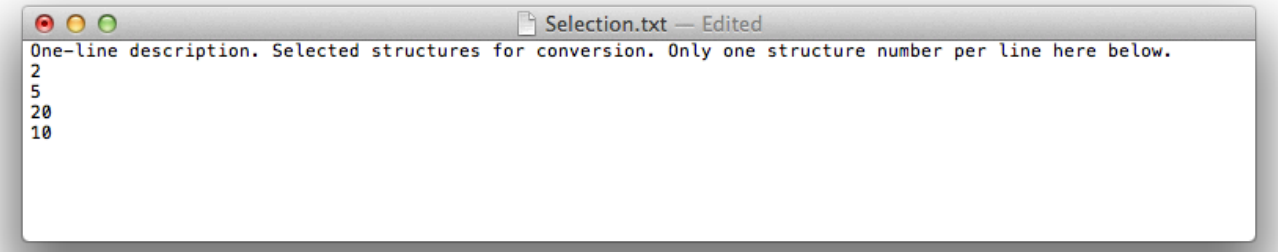
5. A shortcut for GAUSSIAN inputs. No adjustments have to be performed to the GUI selections. Only the number of atoms and, optionally, the mass and charge, have to be provided.
   If parsing multiple GAUSSIAN inputs, all parameters have to be identical for all inputs.
   When choosing GAUSSIAN input parsing, different charge calculation options will appear and you can choose between parsing:
   a. Mulliken charges
   b. APT charges
   c. NBO charges
   d. No charges '*N/A*'

   ParsIMoS scans the GAUSSIAN *.log* files for precise search terms according to the chosen charge descriptions. Depending on the software version of GAUSSIAN, these search terms might differ and **ParsIMoS** will not find the charges. The atom identifiers are read with the charge descriptions and might as well not work. Please refer to the section ***ParsIMoS** help for GAUSSIAN*.



6. A shortcut for MOBCAL inputs. No adjustments have to be performed to the GUI choices apart from adding optionally the mass and the charge. You can choose to parse the entire input or only selected structures (see hereafter). You may as well choose not to parse the charges by checking 5.d. (see above).
   MOBCAL inputs already contain information on the number of atoms and the number of sets of coordinates/structures. You can hence encode multiple MOBCAL inputs with differing number of atoms. Only MOBCAL inputs may be parsed together at once for different molecules/ions. You can choose to parse the entire MOBCAL input or only a selection of structures, as explained hereafter.

7. You can either choose to parse entire inputs, meaning multiple structures present from an input or you can choose parsing only selected structures of your input.
   When several structures are present, each structure will be encoded on a new sheet in your *.xlsx* (or different format) output file.
   If you do not want to parse all the structures from your input file, you can load a *.txt* file using the '*Browse*' button (annotated as 7). The file should contain the structure numbers you want to parse. You are free to put one line of comments in your selection file; the first line will not be read. Please provide only one structure number per line (see example below). If parsing multiple inputs, the structure selection is applied to all inputs.
   The loaded file will be shown in the scrollable window together with the input files.
   The sheets in the *.xlsx* (or different format) output file of the parsed selected structures will be named after the selected structures. In the example below, the sheets would be named *struct2*, *struct5*, *struct20* and *struct10*.



8. Button to begin parsing the input(s) into the chosen *.xls*, *.xlsx*, *.ods* or *.csv* output(s).

9. You can reload the parameters of your last parsing. The GUI parameters are saved in the file '*Last_parameters.par*' once you hit the '*Convert*' button. The file is saved in your program installation files of **ParsIMoS**.
   If you have recurring input formats, you can make copies of this file once you have found your parsing parameters and paste & replace it in the program installation files whenever you need to parse inputs with identical parsing parameters.

10. Info button. Quick tips on how to encode the parameters and on how to edit libraries. Contains as well warnings to avoid problems.
    You can find there as well information on the licensing of **ParsIMoS** and contact information.

11. The number of the line of the first set of coordinates/first structure. Count the lines in your input file, beginning with the very first line of the input file as line#1.

12. The total number of sets of coordinates/structures present in you input file. If you only want to parse the *n first* structures, you may put *n* here instead of the total number of structures. Otherwise, you can also use a selection of input structures to be parsed, as described above.

13. The total number of columns is thought to be a control for parsing. Begin counting with the first column of the section of your input where the coordinates are described, as column#1.
    If all information is in adjacent columns with no spare column, **ParsIMoS** yields a warning if the number of columns is surpassed. This indicates that a delimiter was not correctly detected and/or that more columns were read than the number of columns you read in your input.
    If you have to parse the atom identifiers or the charges in different sections of your input, these warnings only apply for the coordinate columns and they may be more difficult to be triggered, even if errors occurred during parsing.

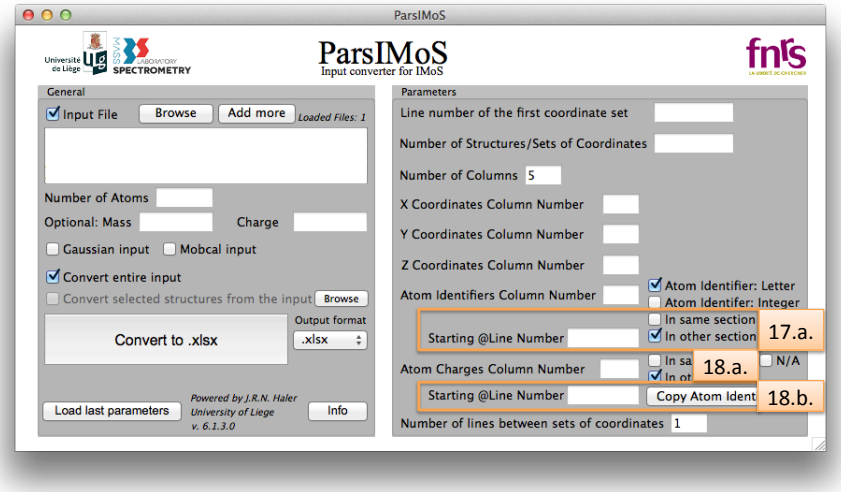
14. The column number from which to parse the x-coordinates. Begin counting with the first column you read in your input file section as column#1.

15. The column number from which to parse the y-coordinates. Begin counting with the first column you read in your input file section as column#1.

16. The column number from which to parse the z-coordinates. Begin counting with the first column you read in your input file section as column#1.

17. The column number from which to parse the atom identifiers. Begin counting with the first column you read in your input file section as column#1.
    Atom identifiers can be represented by either letters (C, O, N, S…) or integer values (12, 16, 14, 32…) in your input file. Tick accordingly the correct box in the GUI. They, however, may not be represented by real values (15.999…).
    The rounded integer values can be read and modified in the library file '*Lib_Atoms_VDW.par*'. Unknown atoms will be encoded as *Other* and *400*. **ParsIMoS** will display a warning when encoding such an atom.
    By default, the atom identifiers are parsed in a different column but from the same lines as the (x,y,z) coordinates.
    Atom identifiers may as well be read in different sections of your input (see below, 17.a.). They would then be in columns with different line numbers than the (x,y,z) coordinates. Count the lines in your input file, beginning with the very first line of the input file as line#1.

Note that you can modify the library file '*Lib_Atoms_VDW.par*'. For further details on where to find and how to encode the file, refer to the section ***ParsIMoS*** *Modifying libraries*.



18. The column number from which to parse the charges of the atoms. Begin counting with the first column you read in your input file section as column#1. You can choose to not parse the atom charges by choosing '*N/A*' (see 18.a. above).
    By default, the atom charges are parsed in a different column but from the same lines as the (x,y,z) coordinates.
    Atom charges may as well be read in different sections of your input (see 18.b. above). They would then be in columns with different line numbers than the (x,y,z) coordinates. Count the lines in your input file, beginning with the very first line of the input file as line#1.
    If atom charges are to be found in the same lines as their associated parsed atom identifiers, a button can quickly copy the line number from the atom identifiers to the atom charges.

19. The number of *blank* or comment lines between the sets of coordinates. **ParsIMoS** calculates the line numbers of the next structures to be parsed and needs this information. The lines, which will be jumped, thus have to be identical throughout the whole input file.

ParsIMoS General advice

Input converter for IMoS

Always count the LINES and COLUMNS beginning at the very first line of the input file or the first column you read as line/column #1. **ParsIMoS** takes care of any necessary conversions.

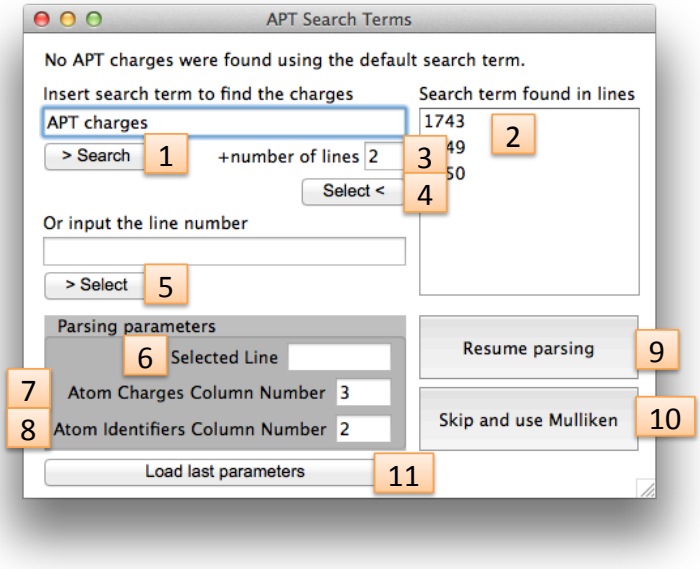
CALCIUM and ARGON are both defined as having the atom identifier *40*. **ParsIMoS** reads the library starting at the beginning of the library when checking for atom identifier matches (when encoded as integer values). An atom labeled *40* will thus be associated by default to Calcium, not Argon.
If you need to parse Argon when having integer values in your input as atom identifiers, please change manually your output files or change the library '*Lib_Atoms_VDW.par*'. You could put Argon in a line preceding Calcium. Please refer to the section ***ParsIMoS*** *Modifying libraries* in order to correctly manipulate the libraries.

ParsIMoS help for GAUSSIAN

Input converter for IMoS

ParsIMoS might encounter problems when encoding the charges and/or atom identifiers from GAUSSIAN *.log* files. **ParsIMoS** scans the GAUSSIAN *.log* files for precise search terms according to the chosen charge descriptions. Depending on the software version of GAUSSIAN, these search terms might differ from the default search terms encoded in **ParsIMoS**.
The atom identifiers are read as letter descriptors (C, O, N, S…) with the charge descriptions (see ***ParsIMoS*** *Tutorial 17*) and might as well not work. They are not read with the coordinates.

If **ParsIMoS** is unable to find the chosen charge description using the encoded search terms, a modal window gives the opportunity to encode a different search term. The window changes slightly whether the errors occur during APT and NBO parsing or during Mulliken parsing.



1. Write the new search term(s) to find the charge description and click '*Search*'. **ParsIMoS** searches through the input file to find all line numbers where the search term(s) occur(s). These line numbers are shown as a list in item *2*.

2. The line numbers, where **ParsIMoS** found the new search terms, are listed here. Choose the correct line by clicking.

3. If you have chosen the correct line number of the occurrence of your search term, you need to give the number of lines **ParsIMoS** has to skip in order to begin parsing the charges and the atom identifiers. The default value for APT and Mulliken charges is 2.

4. If you have chosen the correct line number of the occurrence of your search term and the number of lines to be skipped before beginning parsing, click '*Select*' to finish your line number editing.

5. If you do not find any (specific) search terms for parsing your charges (and atom identifiers), you can also directly input the line number from where **ParsIMoS** should begin parsing. Finish your editing by clicking '*Select*'.

6. The final line number where **ParsIMoS** will begin parsing is shown here.

7. The column number from which to parse the charges of the atoms. Begin counting with the first column you read in your input file section as column#1. The default value from the main **ParsIMoS** window is encoded by default.

8. The column number from which to parse the atom identifiers of the atoms. Begin counting with the first column you read in your input file section as column#1. The default value from the main **ParsIMoS** window is encoded by default.

9. Resume parsing of the charges and atom identifiers using the newly defined line number and charge/atom identifier columns.

10. Optional button which is only shown for APT and NBO re-parameterizations. You can skip defining new line number and charge/atom identifier columns to directly encode Mulliken charges.

11. You can reload the parameters of your last search term(s). The GUI parameters are saved in the file '*Last_parameters_searchterm.par*' once you hit the '*Resume parsing*' button. The file is saved in your program installation files of **ParsIMoS**.
    If you have recurring new search terms, you can make copies of this file once you have found your parsing parameters and paste & replace it in the program installation files whenever you need to parse inputs with identical parsing parameters.

ParsIMoS Modifying libraries

Input converter for IMoS

The libraries are loaded upon launching **ParsIMoS**. If you modify any library, you need to restart **ParsIMoS**.

ParsIMoS uses two libraries: '*Lib_Atoms_VDW.par*' and '*Lib_Delimiters.par*'. They are to be found in the installation files/bundle of **ParsIMoS**.

'*Lib_Atoms_VDW.par*' contains the descriptors of the atoms: the integer atom identifiers, the letter atom identifier and the Van der Waals radii of the atoms. The VDW radii are encoded for **IMoS** calculations.
All parameters may be modified. The library is TAB-delimited between each descriptor, following the order described above (integer – letter(s) – VDW). Each atom is described in a new line.
As mentioned in the section ***ParsIMoS*** *General advice*, be aware of identical integer atom identifiers. The first occurring integer atom identifier in the library will determine the parsed atom identifier.
The last encoded atom identifier in the library is reserved for unknown atoms (*400 – Other – 2*). Substituting this last line will prevent **ParsIMoS** from working properly.
The library file contains two lines of description at the beginning of the file. Do not modify this number of lines.

'*Lib_Delimiters.par*' contains the delimiters **ParsIMoS** reads for parsing special characters in your input files.
You can add or modify delimiters. Please only write one delimiter per line. Do not add delimiters as *char(9)*, *chr(9)*, *#09*…. They will not be read as delimiters but as '*char(9)*', '*chr(9)*' and '#09' instead. You should add your delimiter characters directly as e.g. '*é*' or '±' to the library.
The file contains one line of description at the beginning of the file. Do not modify this number of lines.

ParsIMoS License & contact info

Input converter for IMoS

Program developer.
Jean R. N. Haler (FRIA, F.R.S.-FNRS)
jean.haler@ulg.ac.be

parsimos@outlook.com

Additional contact information.
Christopher Kune
c.kune@ulg.ac.be
Dr. Johann Far
johann.far@ulg.ac.be
Prof. Edwin De Pauw
e.depauw@ulg.ac.be