

MaSS-Simulator

Muaaz Awan and Fahad Saeed

<https://saeedlab.cis.fiu.edu/>

- Select input file containing a list of peptides (modified and un-modified). MaSS-Simulator accepts the input file with modifications shown in a particular format. To facilitate the users, MaSS-Simulator can add desired modifications with required settings while ensuring correct formatting of input file. This can be done from the “***Input File Formatting***” Tab (described at the end). A sample input file has been provided with the package.
- Select the modification file for simulation. A sample modification file has been provided with the package. For correct formatting of modification file please refer to next page.
- Select or provide an output file. Simulated spectra will be generated in .ms2 format.

The screenshot shows the MaSS-Simulator application window. At the top, there are three buttons: "Input File", "Modification File", and "Output File". Below these is a tabbed interface with seven tabs: "C-Ions", "N-Ions", "Immonium Ions", "Multiple Charged Ions", "Noise Settings", "Other Settings", and "Input File Formatting". The "Input File Formatting" tab is currently selected. This tab contains three text input fields with vertical scrollbars, labeled "Mass Offsets", "Ion Generation Probabilities", and "Ion Intensities". At the bottom of the window, there are three buttons: "Initialize Settings", "Load Default", and a large "SIMULATE" button.

If the peptides listed in the input file have modifications, MaSS-Simulator can be configured to consider them. To do so, users need to edit the *modifications.ptm* file. For any **static modification** i.e. if a modification needs to be applied to all the occurrences of an amino acid then following equation should be listed in the *modifications.ptm* file:

$$X = X + c$$

Here X will be replaced by the single letter code of amino acid and c with the mass to be added. For instance, for Carbamidomethylation following equation should be added to *modifications.ptm* file:

$$C = C + 57.02$$

For **variable modifications** equation of the following type needs to be added to *modifications.ptm* file:

$$X[c] = X + c$$

For instance, if we want to add oxidation for M, we will have to add following line to the *modifications.ptm* file:

$$M[15.99] = M + 15.99$$

For the MaSS-Simulator to consider the listed variable modifications, any peptide which has that modification needs to be correctly formatted in the *input file*. For instance if peptide **QMLMP** has an oxidation on first occurrence of M, it should be written like **QM[15.99]LMP**.

To consider the **static modifications at the Termini**, users can simply use the following equations:

$$cTerm = c$$

$$nTerm = b$$

Here *c* and *b* will be the mass value added/subtracted to the C-Terminal and N-Terminal respectively. Variable modifications at the termini can be handled using the format for variable modifications on any amino acid.

- To setup the simulator with default settings, user can press the Load Default button. Otherwise all the parameters need to be specified manually.
- To list the type of C/N-terminal ions to be simulated, users can list the mass-offsets for each ion in the “Mass Offsets” space in *C/N-Ions* tab. Table on the next page provides mass-offsets for most common C-Terminal and N-Terminal ions.
- For instance, in the Figure on right. An offset of 0 means y-ion series, offsets of -1/+1 will generate isotopic ions for y-series and similarly offset of -18 and -17 will generate neutral water and ammonia losses for each y-ion.

The screenshot shows the 'C-Ions' tab of a simulator interface. At the top, there are three buttons: 'Input File', 'Modification File', and 'Output File'. Below these is a tabbed interface with the following tabs: 'C-Ions' (selected), 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings', 'Other Settings', and 'Input File Formatting'. An 'About' button is in the top right corner. The main area contains three input fields with scrollbars:

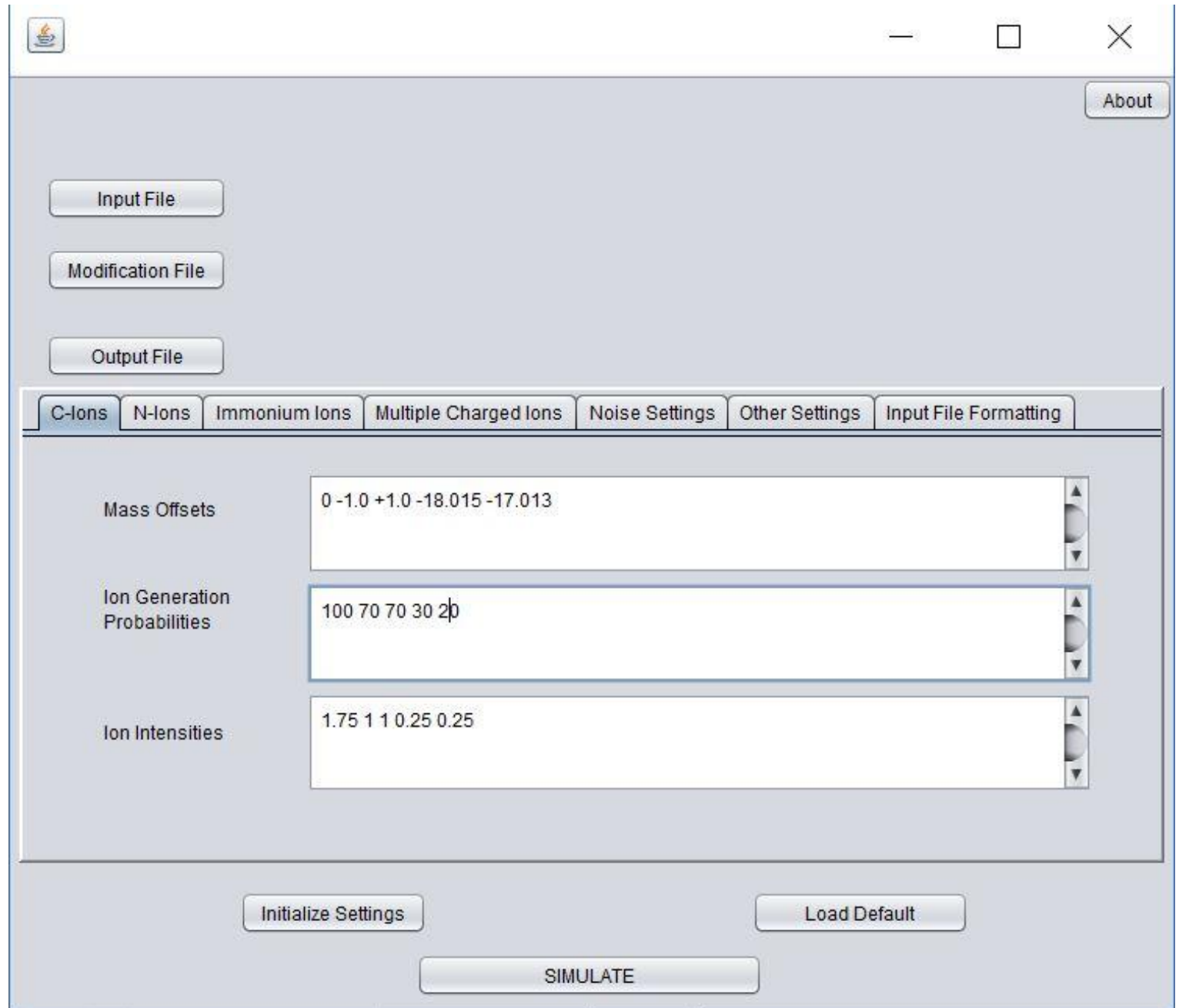
Parameter	Value
Mass Offsets	0 -1.0 +1.0 -18.015 -17.013
Ion Generation Probabilities	70 70 70 70 70
Ion Intensities	1.75 1 1 0.25 0.25

At the bottom, there are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

Common Ions and their Mass-Offsets

C-Ion Type	Mass offset
y-	0
y-H ₂ O	-18.015
y-NH ₃	-17.013
N-Ion Type	
b-	0
b-H ₂ O	-18.015
b-NH ₃	-17.013
a-	-28.01
a-H ₂ O	-45.023
a-NH ₃	-46.025

- Ion Generation Probabilities (IGP) determine the possibility of each ion to be generated. IGP values for C and N-terminal ions should be given in the same order as that of the mass-offsets using a space in between the values. For example, in on the right IGP value for y-ion series is 100% while 70% each for the isotopic ions and 30 and 20% for the neutral water and ammonia losses respectively.
- Intensity values for ion series should be given in the same order as that of the ion mass-offsets using a space in between the values. For example, in Figure on the right intensity value for y-ion series is 1.75 while 1 each for the isotopic ions and 0.25 for the neutral water and ammonia losses. All the intensity values are multiplied by a factor of 1000.



The screenshot shows a software window with a title bar containing a flame icon, a minus sign, a square, and a close button. An 'About' button is in the top right corner. Below the title bar are three buttons: 'Input File', 'Modification File', and 'Output File'. A tabbed interface follows, with tabs for 'C-Ions', 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings', 'Other Settings', and 'Input File Formatting'. The 'C-Ions' tab is selected. It contains three text input fields with scrollbars: 'Mass Offsets' with the value '0 -1.0 +1.0 -18.015 -17.013', 'Ion Generation Probabilities' with the value '100 70 70 30 20', and 'Ion Intensities' with the value '1.75 1 1 0.25 0.25'. At the bottom are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

Input File

Modification File

Output File

C-Ions N-Ions Immonium Ions Multiple Charged Ions Noise Settings Other Settings Input File Formatting

Mass Offsets 0 -1.0 +1.0 -18.015 -17.013

Ion Generation Probabilities 100 70 70 30 20

Ion Intensities 1.75 1 1 0.25 0.25

Initialize Settings Load Default

SIMULATE

- Immonium ion configuration follows the same format with the exception of first field of “Amino Acids”. Here all those amino acids can be listed for which Immonium Ions need to be generated. These have to be listed using the single lettered code for the amino acids as shown in the Figure on right. IGP values and intensity values for each immonium ion needs to be entered. These will be entered in the same sequence as the amino acids.

The screenshot shows a software window titled "Immonium Ion Configuration" with standard Windows window controls (minimize, maximize, close) in the top right corner. An "About" button is located in the top right corner of the window's content area. Below the window title bar, there are three buttons: "Input File", "Modification File", and "Output File". A tabbed interface follows, with tabs for "C-Ions", "N-Ions", "Immonium Ions" (which is currently selected), "Multiple Charged Ions", "Noise Settings", "Other Settings", and "Input File Formatting". The main content area of the "Immonium Ions" tab contains three input fields, each with a vertical scrollbar on the right. The first field, labeled "Amino Acids", contains the text "M D Q K E". The second field, labeled "Ion Generation Probabilities", contains the text "68 68 68 68 68". The third field, labeled "Ion Intensities", contains the text "0.02 0.02 0.02 0.02 0.02". At the bottom of the window, there are three buttons: "Initialize Settings", "Load Default", and a larger "SIMULATE" button.

Input File

Modification File

Output File

C-Ions N-Ions **Immonium Ions** Multiple Charged Ions Noise Settings Other Settings Input File Formatting

Amino Acids M D Q K E

Ion Generation Probabilities 68 68 68 68 68

Ion Intensities 0.02 0.02 0.02 0.02 0.02

Initialize Settings Load Default

SIMULATE

- To generate multiple charged variants of b- and y-ions a list of > +1 charges needs to be listed in “Charges” field.
- IGP values for these multiply charged ions can be listed in the field below in the same sequence as those of the charges.
- For example, in Figure on the right, ions with +2 charges will be generated with IGP value of 35.
- Multiply charged ions are generated only for b- and y-ion series (mass-offset value 0) and their intensities are also same as the b- and y-ion series

The screenshot shows a software window with a title bar containing a flame icon and standard window controls. The main area has a tabbed interface with the following tabs: C-Ions, N-Ions, Immonium Ions, Multiple Charged Ions (selected), Noise Settings, Other Settings, and Input File Formatting. Above the tabs are three buttons: Input File, Modification File, and Output File. In the 'Multiple Charged Ions' tab, there are two input fields: 'Charges' with the value '2' and 'Ion Generation Probabilities' with the value '35'. At the bottom of the window are three buttons: Initialize Settings, Load Default, and a large SIMULATE button.

- Percentage of Sound (POS) value can be entered in the POS field to control the noise content. A higher value would mean smaller noise. For instance in the figure on right, 3.0 means only 3% peaks in a spectrum would be sound (peaks of interest) and remaining peaks would be noise.
- To select the type of noise distribution, Noise Type field can be set with 0, 1 or 2. Where 0 means no noise, 1 selects uniform distribution and 2 selects gaussian distribution.
- To control the intensity range of noise peaks a minimum intensity value and a maximum intensity value needs to be provided. Intensities within the mentioned range are randomly assigned to each noise peak. For instance, in figure on right intensities will be between 0.5 and 2.25 (multiplied with a factor of 1000).

The screenshot shows a software window with a title bar containing a logo, a minimize button, a maximize button, and a close button. In the top right corner of the window is an 'About' button. Below the title bar are three buttons: 'Input File', 'Modification File', and 'Output File'. A tabbed interface follows, with tabs for 'C-Ions', 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings' (which is currently selected), 'Other Settings', and 'Input File Formatting'. The 'Noise Settings' tab contains the following controls:

- Noise Type:** A dropdown menu showing '2'. To its right is the text '0: no noise, 1: Uniform Noise, 2: Gaussian Noise'.
- POS:** A text input field containing the value '3.0'.
- Intensity Range:** Two text input fields labeled 'Min' and 'Max'. The 'Min' field contains '0.5' and the 'Max' field contains '2.25'.

At the bottom of the window are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

- To simulate spectra for peptides within a given range of mass, Precursor Mass Range can be provided. For instance, in Figure on right, shows the mass range from 100 to 10,000Da.

The screenshot shows a software window with a standard Windows-style title bar (minimize, maximize, close buttons). In the top right corner, there is an 'About' button. Below the title bar, there are three buttons: 'Input File', 'Modification File', and 'Output File'. A horizontal tab bar contains seven tabs: 'C-Ions', 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings', 'Other Settings' (which is currently selected), and 'Input File Formatting'. The main area of the window displays the 'Precursor Mass Range' settings. It includes the text 'Precursor Mass Range:' followed by 'Min' and a text box containing '100', then 'Max' and a text box containing '10000'. At the bottom of the window, there are three buttons: 'Initialize Settings', 'Load Default', and a larger 'SIMULATE' button.

- For users who have an input file containing list of peptides (un-modified) and want to add modifications. Can make use of the “Input File Formatting” Tab.
- By default the check box will be checked and both the fields below will be disabled. In this case the simulator will simulate the un-modified peptides without adding any PTMs.

The screenshot shows a software interface with a tabbed menu at the top. The tabs are: C-Ions, N-Ions, Immonium Ions, Multiple Charged Ions, Noise Settings, Other Settings, and Input File Formatting (which is currently selected). Above the tabs, there are three file selection fields: 'Input File' with the path 'C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\myPepsIn.txt', 'Modification File' with the path 'C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\modifications.ptm', and 'Output File' with the path 'C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\myOut.ms2'. An 'About' button is in the top right corner. Below the tabs, there is a checked checkbox labeled 'Post Translational Modifications have been correctly formatted in my input file'. Underneath this, there are two disabled input fields: 'Number of modifications allowed per peptide' with a value of '0', and '% Probability that each peptide will contain a modification' with a value of '0'. At the bottom of the window, there are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

Input File: C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\myPepsIn.txt

Modification File: C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\modifications.ptm

Output File: C:\Users\Gul\Desktop\Disk 1\8thSemester\assessment\guiTest\myOut.ms2

Input File Formatting

☒ Post Translational Modifications have been correctly formatted in my input file

Number of modifications allowed per peptide: 0

% Probability that each peptide will contain a modification: 0

Initialize Settings Load Default

SIMULATE

- To add PTMs to the input file in a format which is suitable for MaSS-Simulator. Users can uncheck the check box on top of “Input File Formatting” tab and enter the values in fields below. As shown in the figure on right.
- The first field indicates the maximum modifications allowed per peptide.
- The second field indicates the percentage probability that a given peptide will have modifications. For instance in the figure on right, there is a 50% chance that a given peptide will have PTMs added to it.

The screenshot shows the 'Input File Formatting' tab of the MaSS-Simulator interface. At the top right is an 'About' button. Below it are three file selection fields: 'Input File' (C:\Users\Gu\\Desktop\Disk 1\8thSemester\assessment\guiTest\myPepsIn.txt), 'Modification File' (C:\Users\Gu\\Desktop\Disk 1\8thSemester\assessment\guiTest\modifications.ptm), and 'Output File' (C:\Users\Gu\\Desktop\Disk 1\8thSemester\assessment\guiTest\myOut.ms2). A tab bar contains 'C-Ions', 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings', 'Other Settings', and 'Input File Formatting' (which is selected). Below the tabs is a checkbox labeled 'Post Translational Modifications have been correctly formatted in my input file', which is unchecked. Two input fields follow: 'Number of modifications allowed per peptide' with the value '3', and '% Probability that each peptide will contain a modification' with the value '50'. At the bottom are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

Input File: C:\Users\Gu\Desktop\Disk 1\8thSemester\assessment\guiTest\myPepsIn.txt

Modification File: C:\Users\Gu\Desktop\Disk 1\8thSemester\assessment\guiTest\modifications.ptm

Output File: C:\Users\Gu\Desktop\Disk 1\8thSemester\assessment\guiTest\myOut.ms2

Input File Formatting

☐ Post Translational Modifications have been correctly formatted in my input file

Number of modifications allowed per peptide: 3

% Probability that each peptide will contain a modification: 50

Initialize Settings Load Default

SIMULATE

- Once all the parameters have been set and input/output files have been selected. Click ***Initialize Settings*** (highlighted in green) and then click ***SIMULATE*** (highlighted in yellow) to execute.
- Along with the output file a ground truth file (***peptides.rst***) will be generated in the same directory as the executable ***jar*** file.

The screenshot shows a software window with a title bar containing a logo and standard window controls (minimize, maximize, close). In the top right corner is an 'About' button. On the left side, there are three buttons: 'Input File', 'Modification File', and 'Output File'. Below these is a horizontal tab bar with six tabs: 'C-Ions', 'N-Ions', 'Immonium Ions', 'Multiple Charged Ions', 'Noise Settings', 'Other Settings', and 'Input File Formatting'. The 'Other Settings' tab is currently selected. The main area of the window displays the 'Precursor Mass Range' with 'Min' and 'Max' labels and corresponding input fields. The 'Min' field contains the value '100' and the 'Max' field contains '10000'. At the bottom of the window, there are three buttons: 'Initialize Settings', 'Load Default', and a large 'SIMULATE' button.

Input File

Modification File

Output File

C-Ions N-Ions Immonium Ions Multiple Charged Ions Noise Settings Other Settings Input File Formatting

Precursor Mass Range: Min 100 Max 10000

Initialize Settings Load Default

SIMULATE

Please report any bugs to fsaeed@fiu.edu. Cite us using:

Muaaz G. Awan and Fahad Saeed, "MaSS-Simulator: A highly configurable simulator for generating MS/MS datasets for benchmarking of proteomics algorithms", *Wiley Proteomics*, September 2018