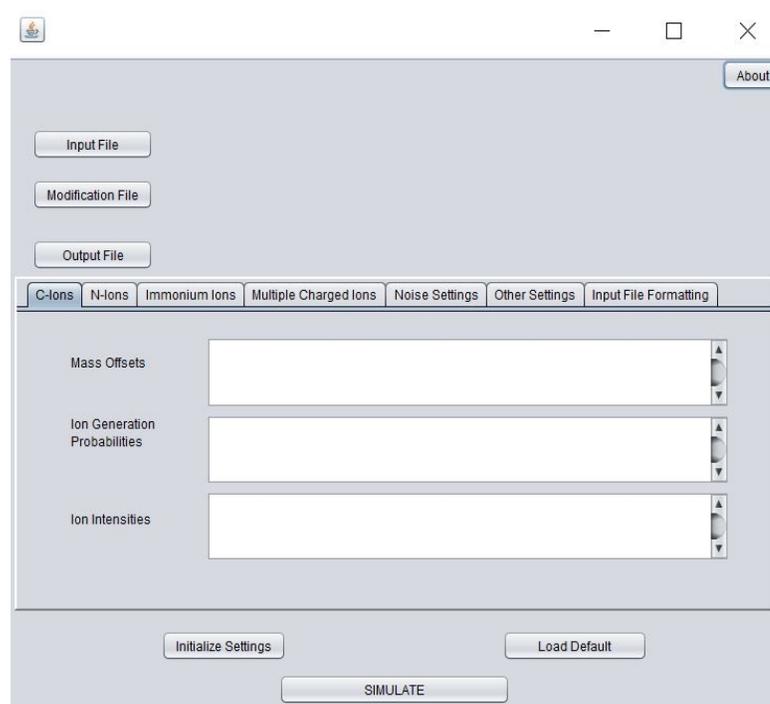


MaSS-Simulator

Setting up Input Files

- Select input file containing a list of peptides. Users can either add the modification manually (in the format mentioned on next page) or use the file formatting option available with MaSS-Simulator GUI. This option 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 few sample input files have been provided with the package.
- Select the *modification.ptm* file containing details of post-translational modifications to be simulated. A sample modification file has been provided with the package. For correct formatting of *modification.ptm* file please refer to next page.
- Select or provide an output file. Simulated spectra will be generated in .ms2 format.



Formatting the *modifications.ptm* file

modifications.ptm file provides the details of the post-translational modifications which are required by the MaSS-Simulator in case the user wants certain modifications to be simulated. 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:

$$\mathbf{X} = \mathbf{X} + \mathbf{c}$$

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:

$$\mathbf{C} = \mathbf{C} + 57.02$$

To add variable modifications to the *modifications.ptm file* equation of following format can be used:

$$\mathbf{X}[\mathbf{c}] = \mathbf{X} + \mathbf{c}$$

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

$$\mathbf{M}[15.99] = \mathbf{M} + 15.99$$

For a correctly formatted *input file* these modifications will be simulated. For instance if peptide **QMLMP** has an oxidation on first occurrence of M, in the input file 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.

Setting Ion Generation Probabilities and Intensities

- 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 Figure 2. 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.

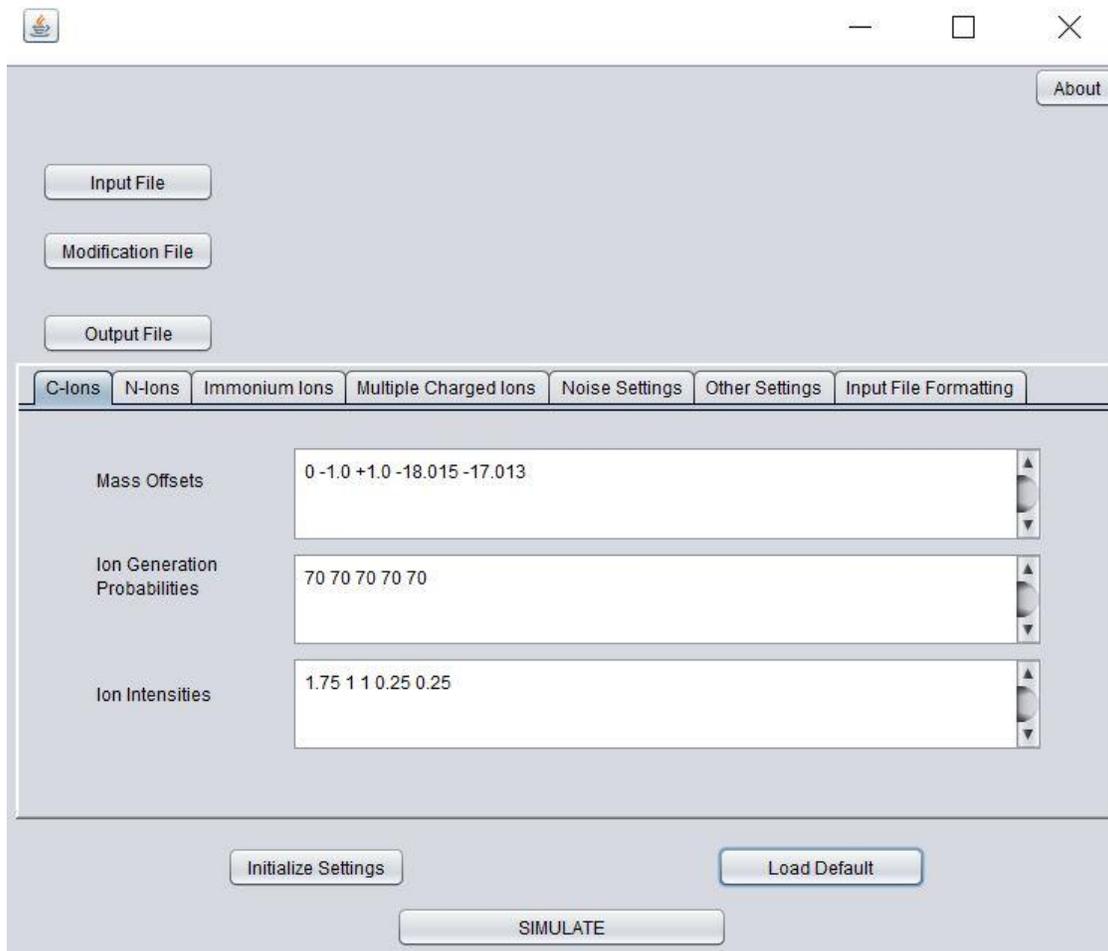


Figure 2

Common Ions and the mass-offsets are shown in Table 1.

Table 1

| C-Ion Type | Mass offset |
|-------------------|-------------|
| y- | 0 |
| y-H2O | -18.015 |
| y-NH3 | -17.013 |
| N-Ion Type | |
| b- | 0 |
| b-H2O | -18.015 |
| b-NH3 | -17.013 |
| a- | -28.01 |
| a-H2O | -45.023 |
| a-NH3 | -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 Figure 3 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 3 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.

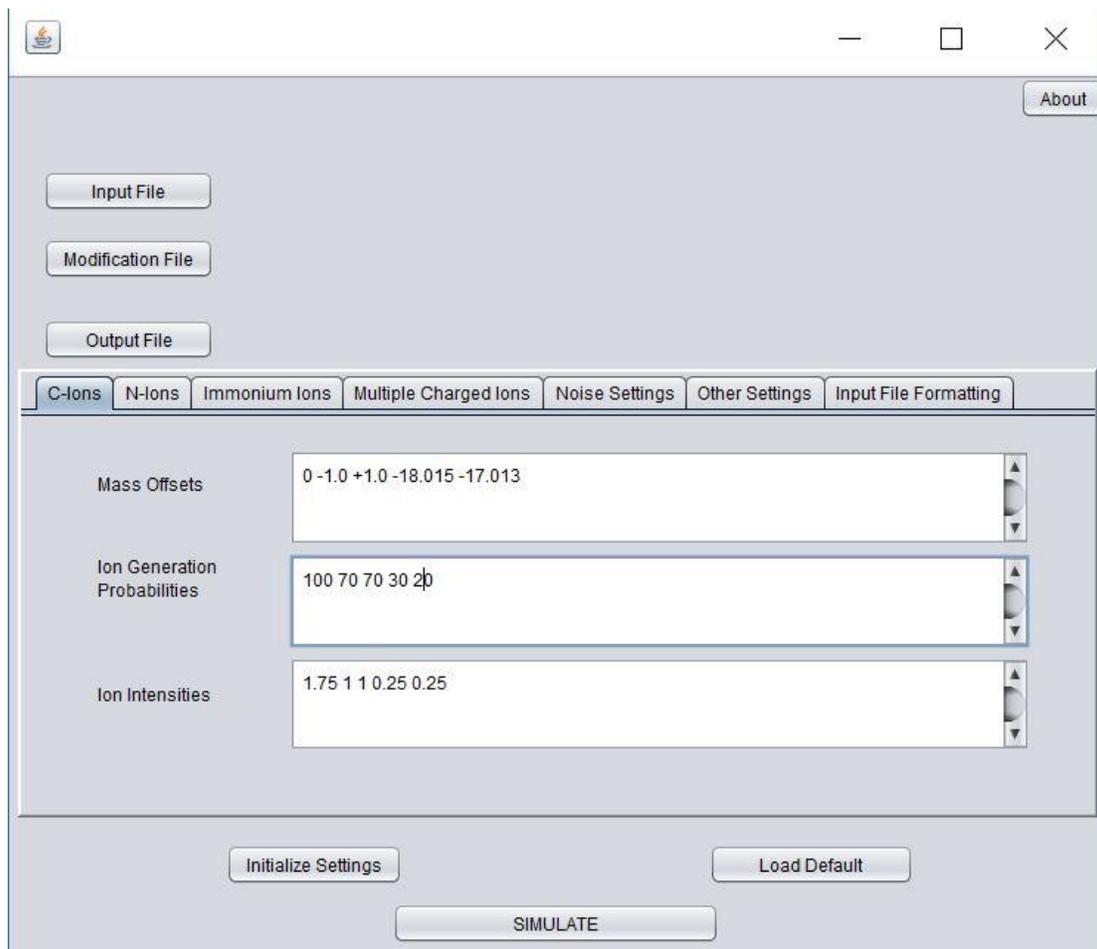
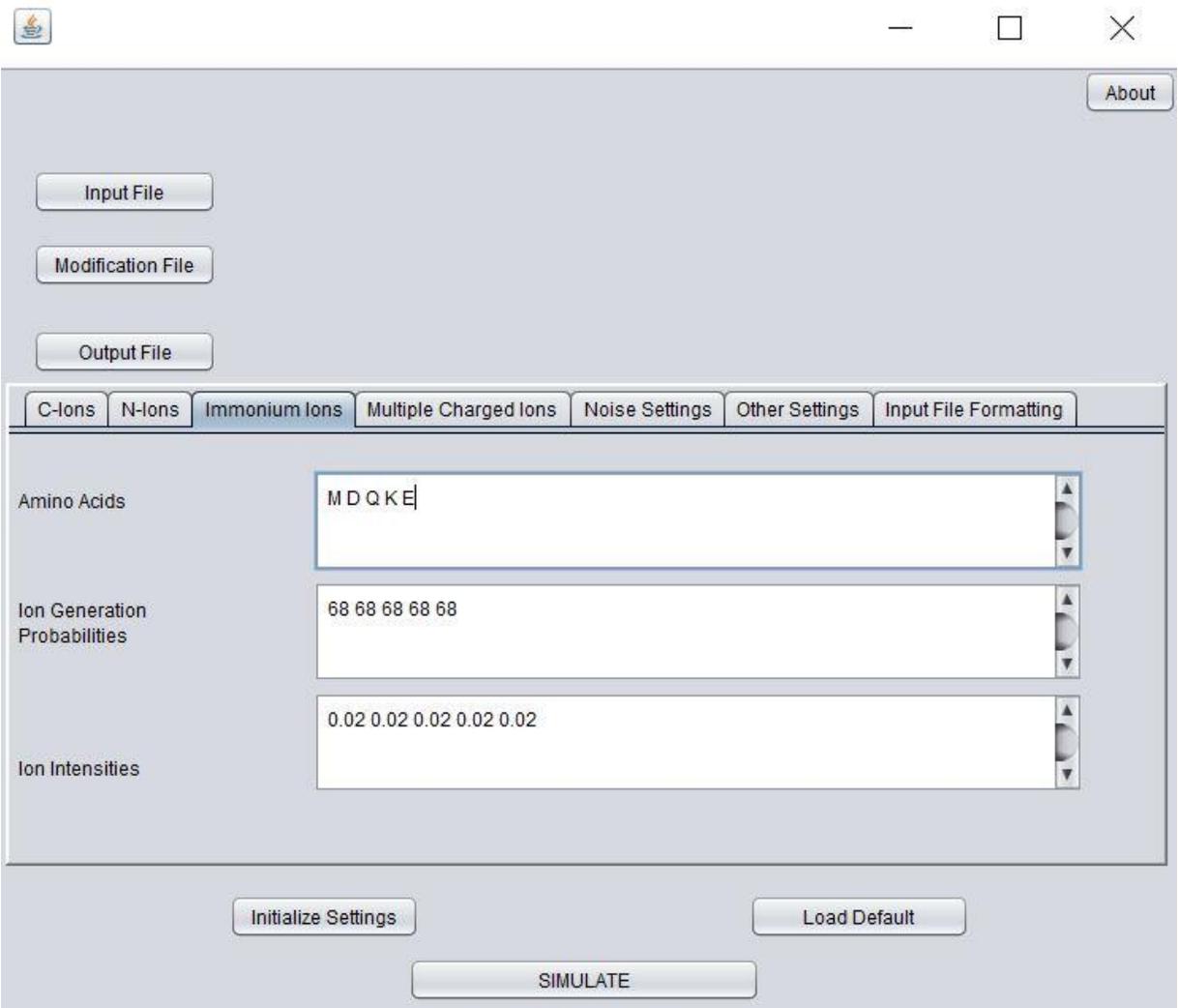


Figure 3

Immonium Ions

- Immonium ion configuration follows the same format except for first field of “Amino Acids”. Here all those amino acids can be listed for which Immonium Ions need to be generated. These must be listed using the single lettered code for the amino acids as shown in the Figure 4. 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 Ions" with a standard Windows-style title bar (minimize, maximize, close buttons) and an "About" button in the top right corner. The window contains several input fields and buttons:

- Three buttons for file selection: "Input File", "Modification File", and "Output File".
- A tabbed interface with tabs for "C-Ions", "N-Ions", "Immonium Ions" (selected), "Multiple Charged Ions", "Noise Settings", "Other Settings", and "Input File Formatting".
- Under the "Immonium Ions" tab, there are three text input fields:
 - "Amino Acids" containing the text "M D Q K E".
 - "Ion Generation Probabilities" containing the text "68 68 68 68 68".
 - "Ion Intensities" containing 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 large "SIMULATE" button.

Figure 4

Multiple Charged Ions

- To generate multiple charged variants of b- and y-ions a list of charges greater than +1 needs to be entered 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 5, 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

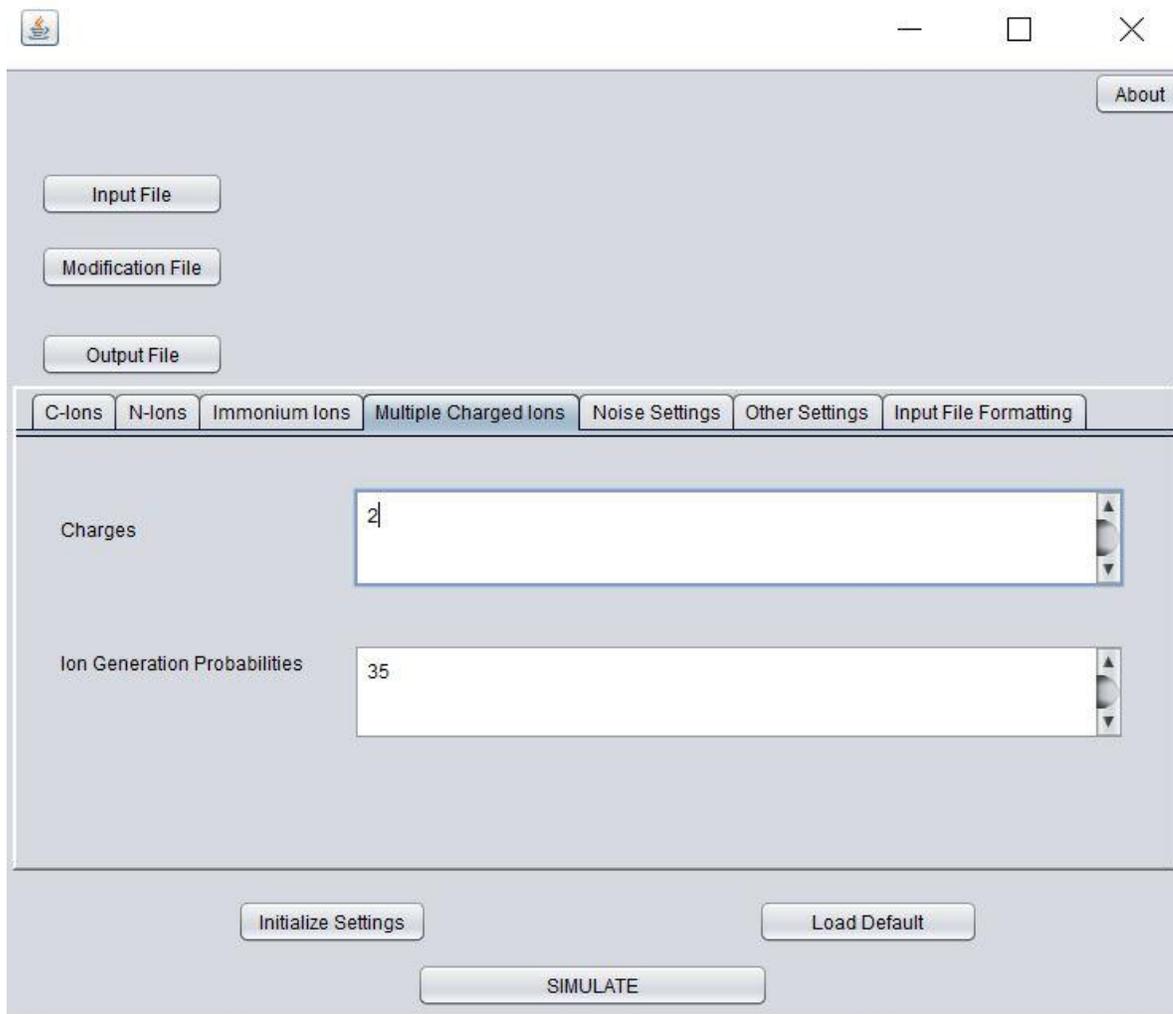


Figure 5

Controlling Noise Content

- 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 Figure 6, 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 6 intensities will be between 0.5 and 2.25 (multiplied with a factor of 1000).

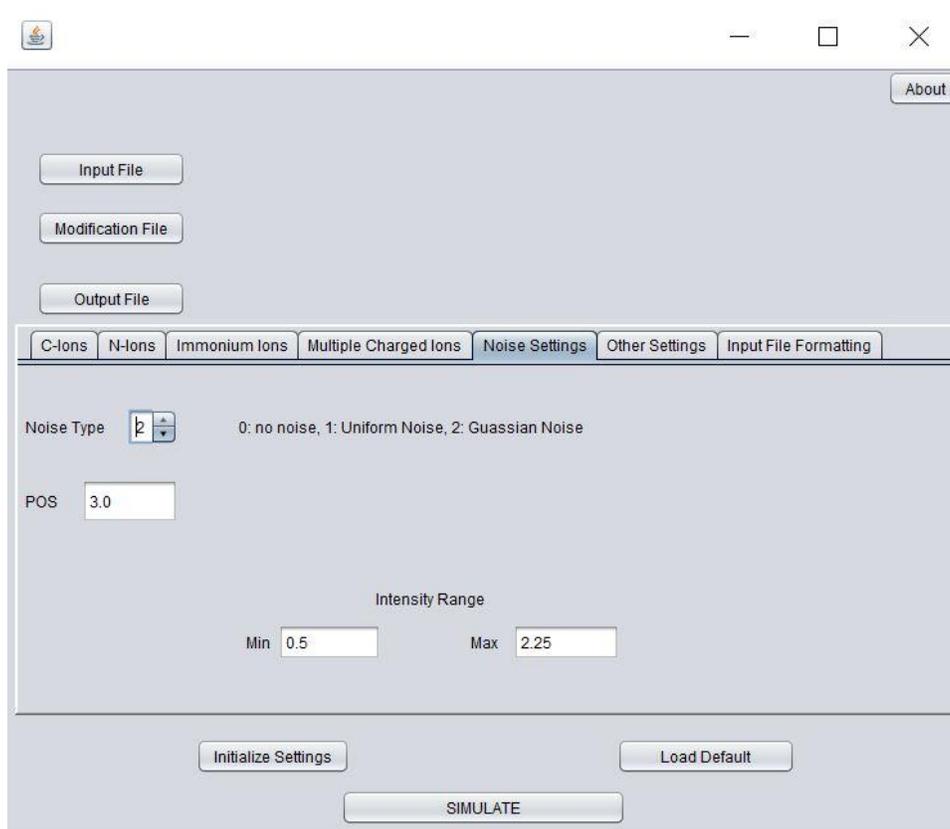
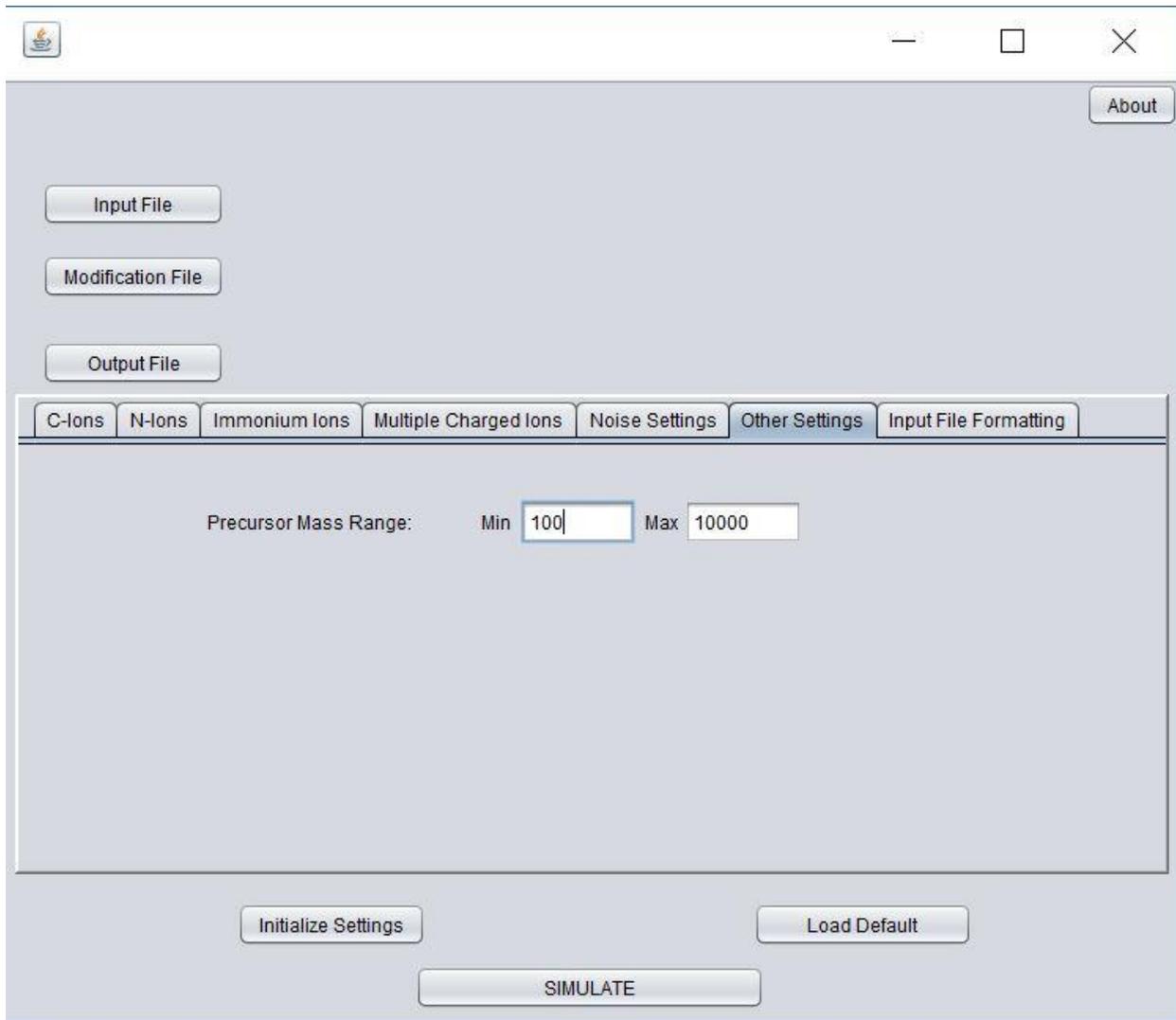


Figure 6

Defining a Mass Range for Simulation

- To simulate spectra for peptides within a given range of mass, Precursor Mass Range can be provided. For instance, in Figure 7, shows the mass range from 100 to 10,000Da.



The image 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. On the left side, there are three buttons: "Input File", "Modification File", and "Output File". Below these is a horizontal tabbed menu with the following 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. In the main area of the window, the text "Precursor Mass Range:" is followed by "Min" and a text input field containing "100", then "Max" and a text input field containing "10000". At the bottom of the window, there are three buttons: "Initialize Settings", "Load Default", and a large "SIMULATE" button.

Figure 7

Inserting Peptide Modifications Using Input File Formatting Tool

- For users with an input file containing list of peptides (un-modified) and want to add modifications using MaSS-Simulator's integrated file formatting tool. They can make use of the "Input File Formatting" Tab.
- Modifications which the user wants to add, should be listed in the ***modifications.ptm*** (check above for format of ***modifications.ptm***) file.
- By default, the check box will be checked, and both the fields below will be disabled. In this case the simulator will simulate the peptides in the input file without adding any changes to the input file.
- To add PTMs to the input file using file formatting tool. 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 8.
- 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.
- If a peptide is chosen at random for modifications, the tool will add modifications at first **x** instances of the amino acids to be modified. Where **x** is the maximum number of modifications allowed per peptide.

Note: If you decide to use the file formatting tool for adding modifications to input file, please make sure the input file consists only of un-modified peptides (don't add any modifications manually).

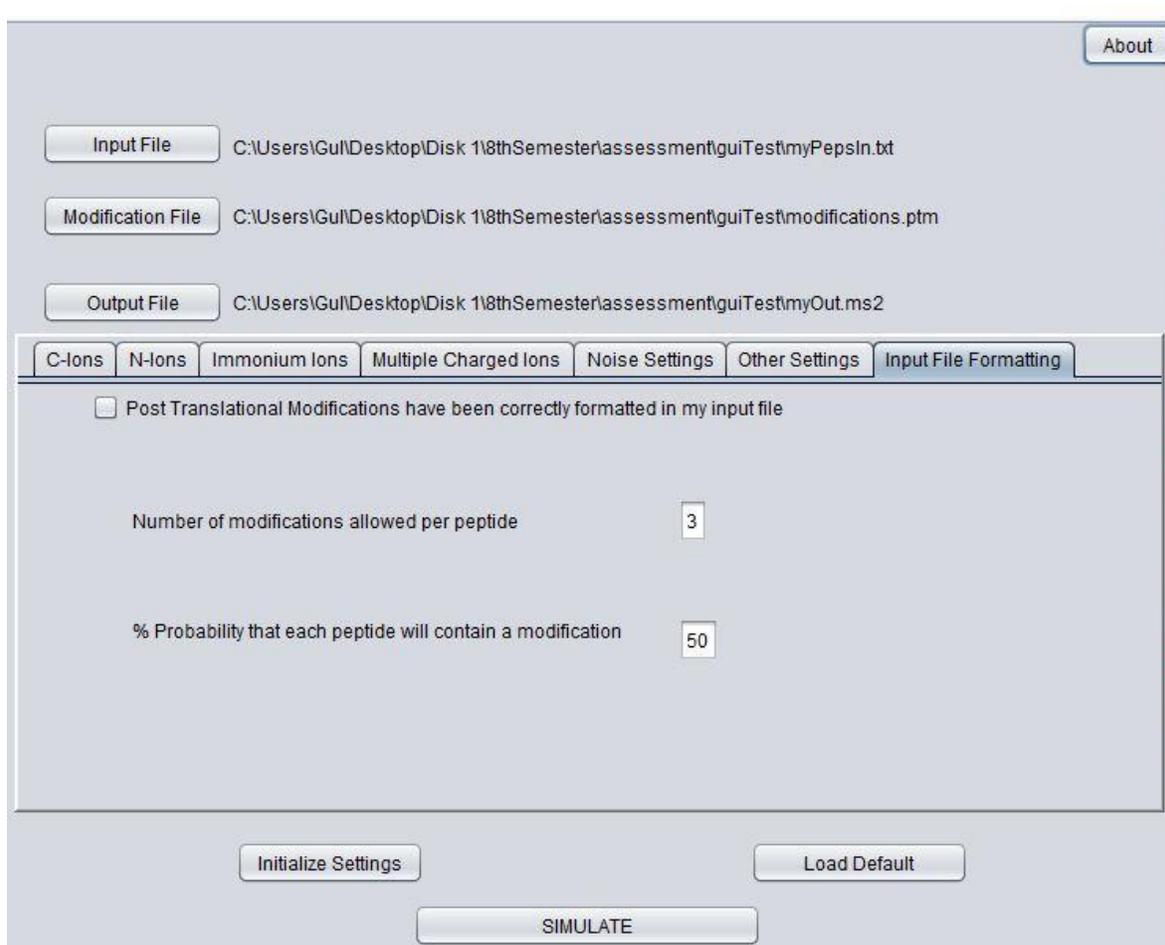


Figure 8

Running the Simulation

- 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. Please refer to Figure 9.
- Along with the output file a ground truth file (*peptides.rst*) will be generated in the same directory as the executable *jar* file.

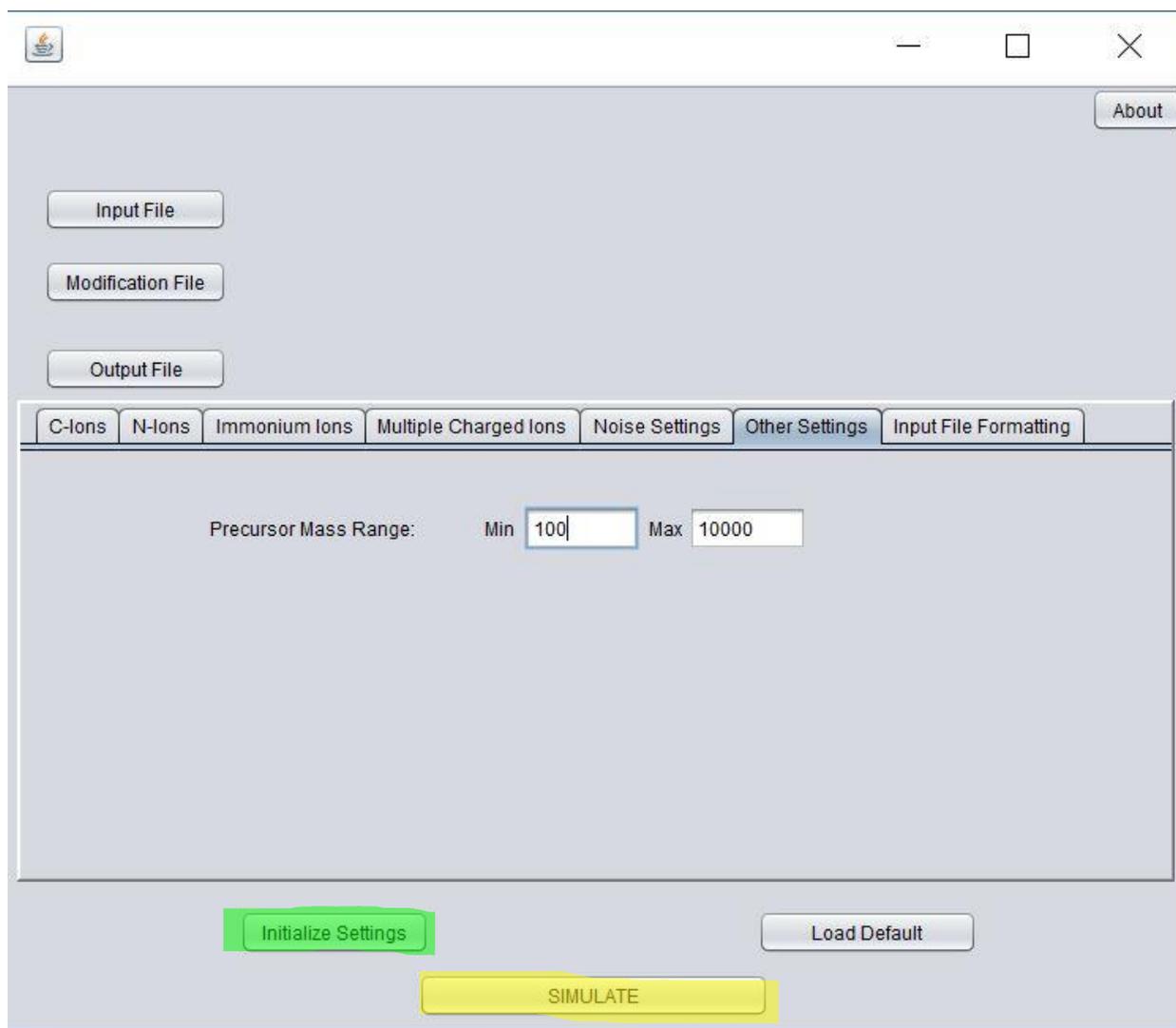


Figure 9

- **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