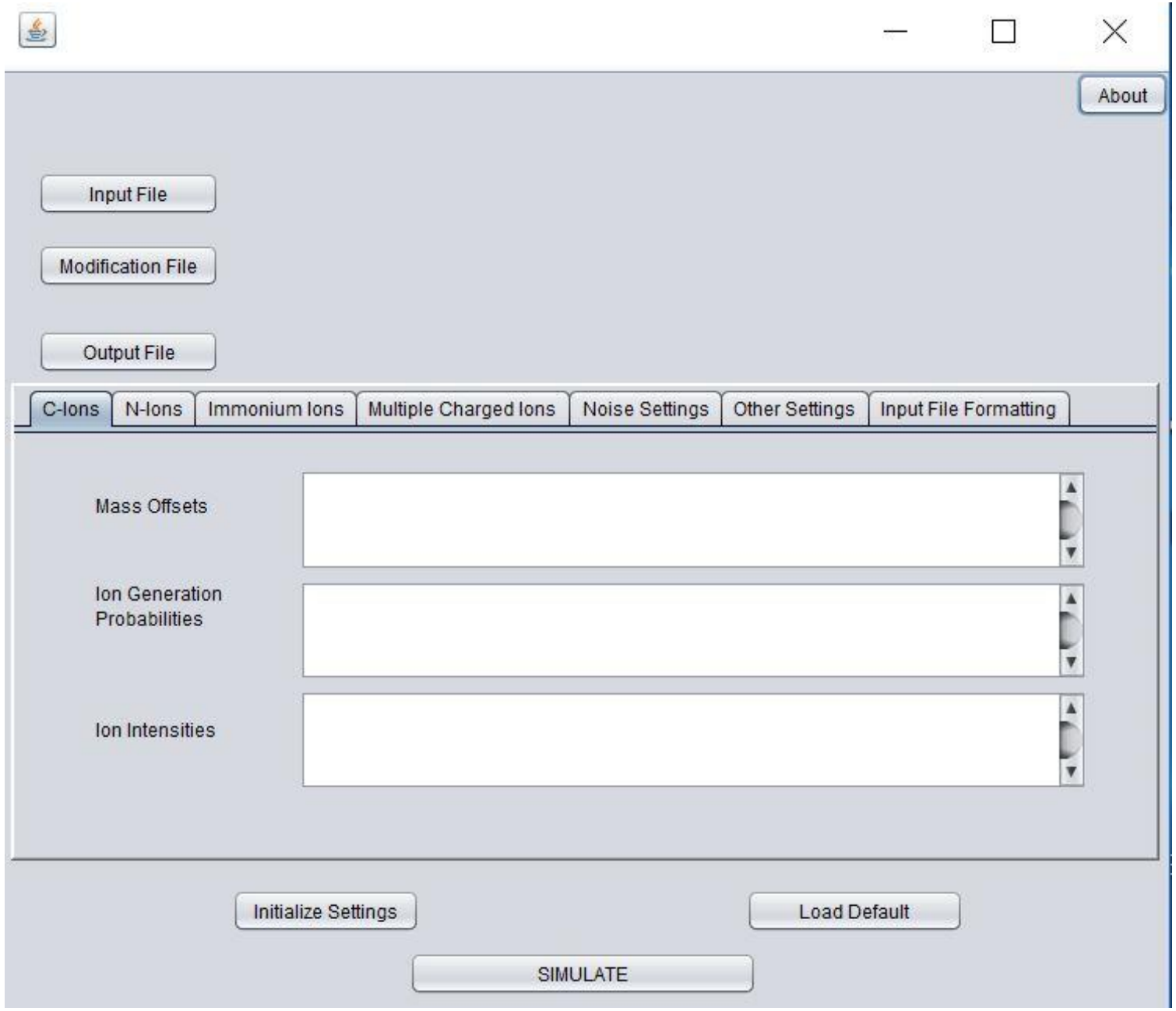
MaSS-Simulator

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.



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**.

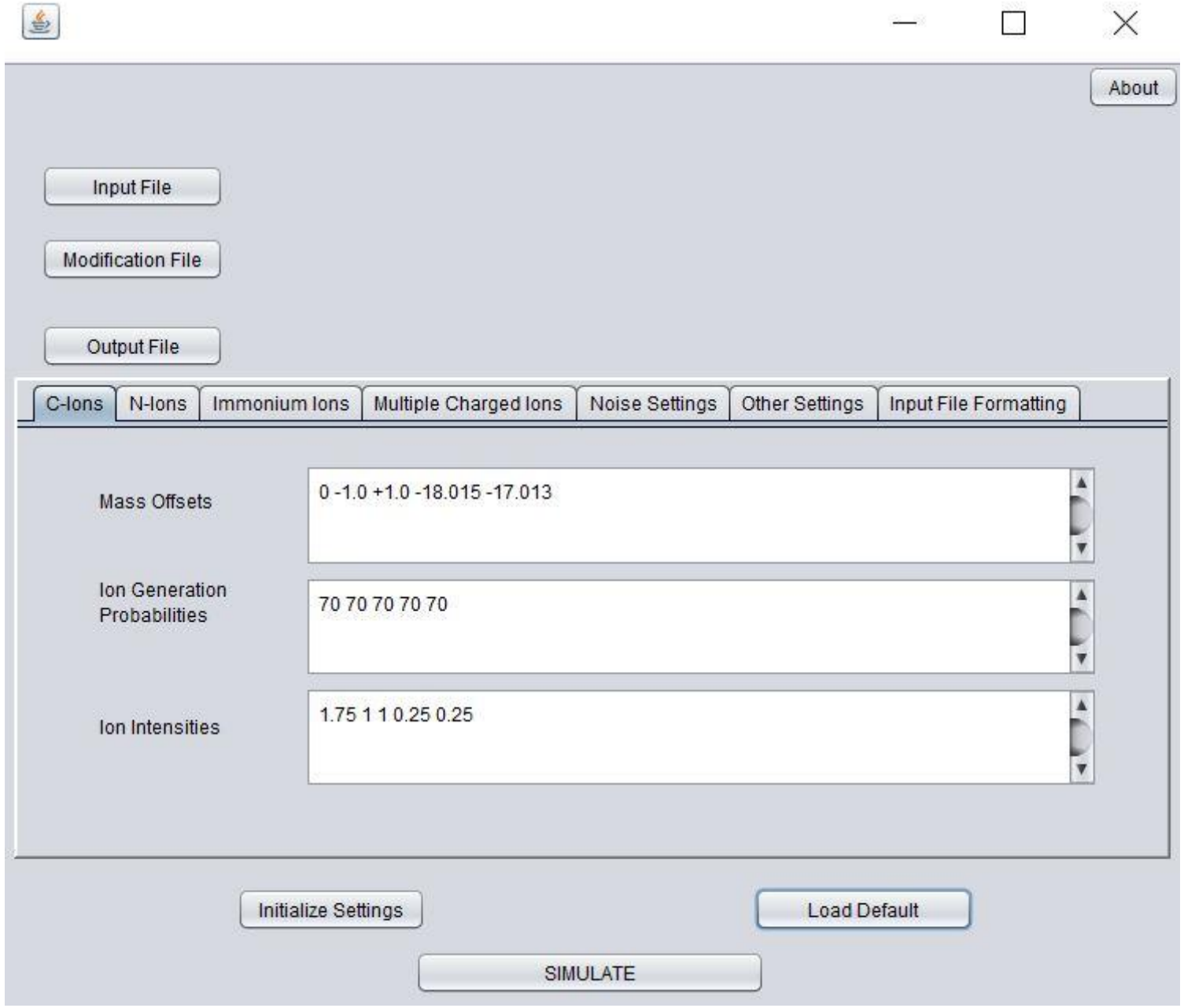
To consider the **static modifications at the Termini**, users can simply use the following equations:

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.

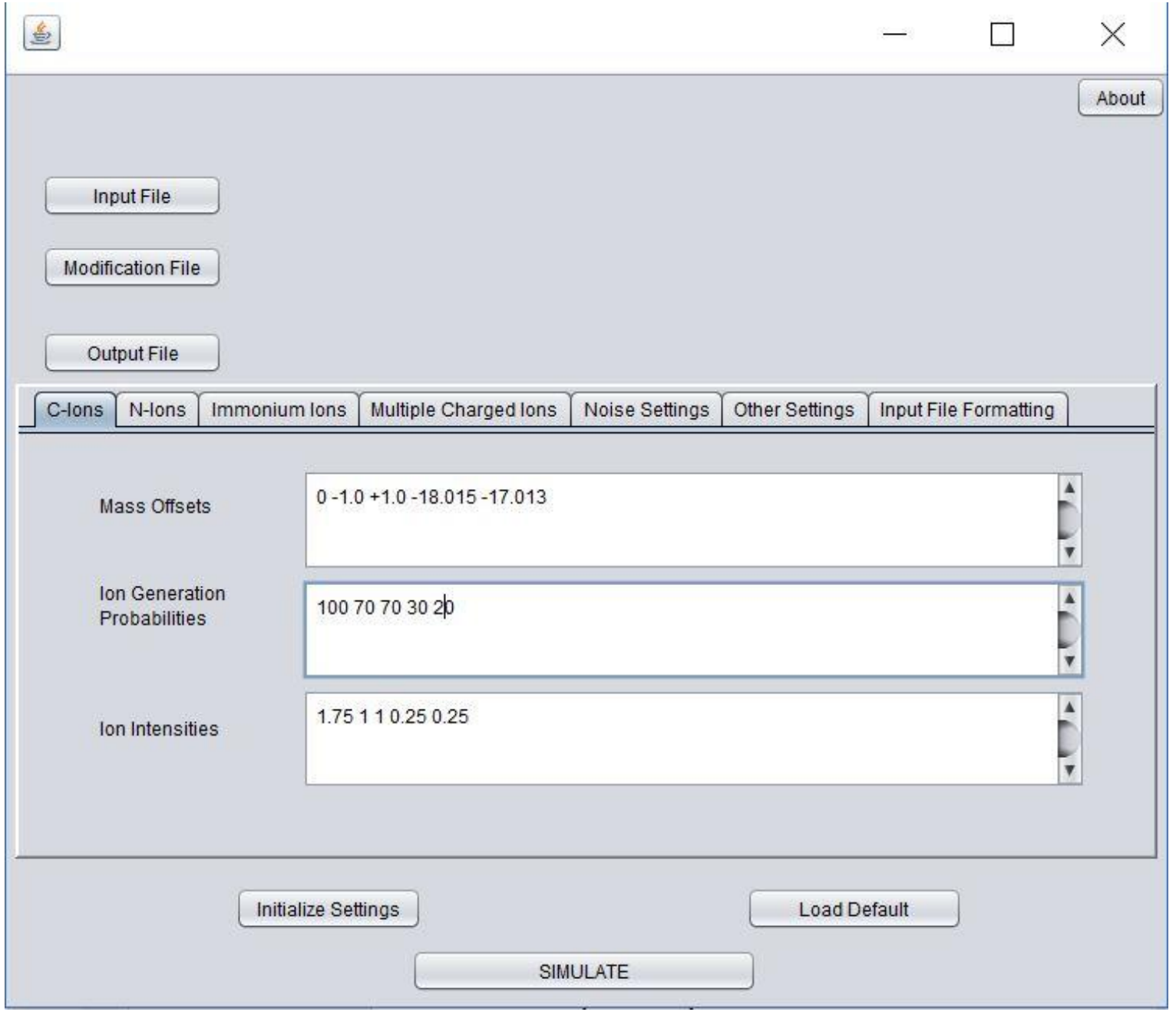
- To setup the simulator with default settings, user can press the Load Default button. Otherwise all the parameters need to be specified manually.
- To list the type of C/N-terminal ions to be simulated, users can list the mass-offsets for each ion in the “Mass Offsets” space in *C/N-Ions* tab. Table 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.



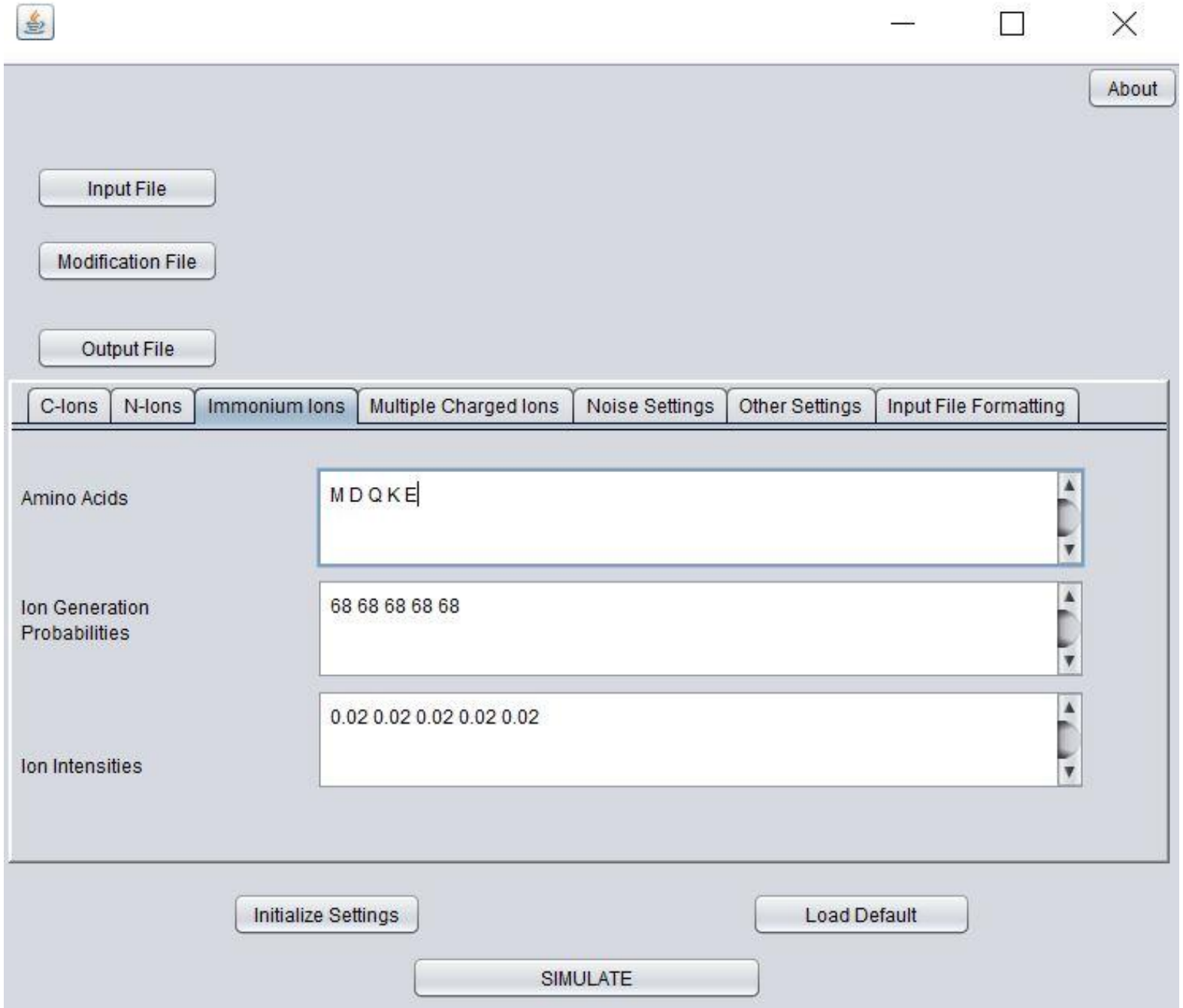
Common Ions and their Mass-Offsets

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

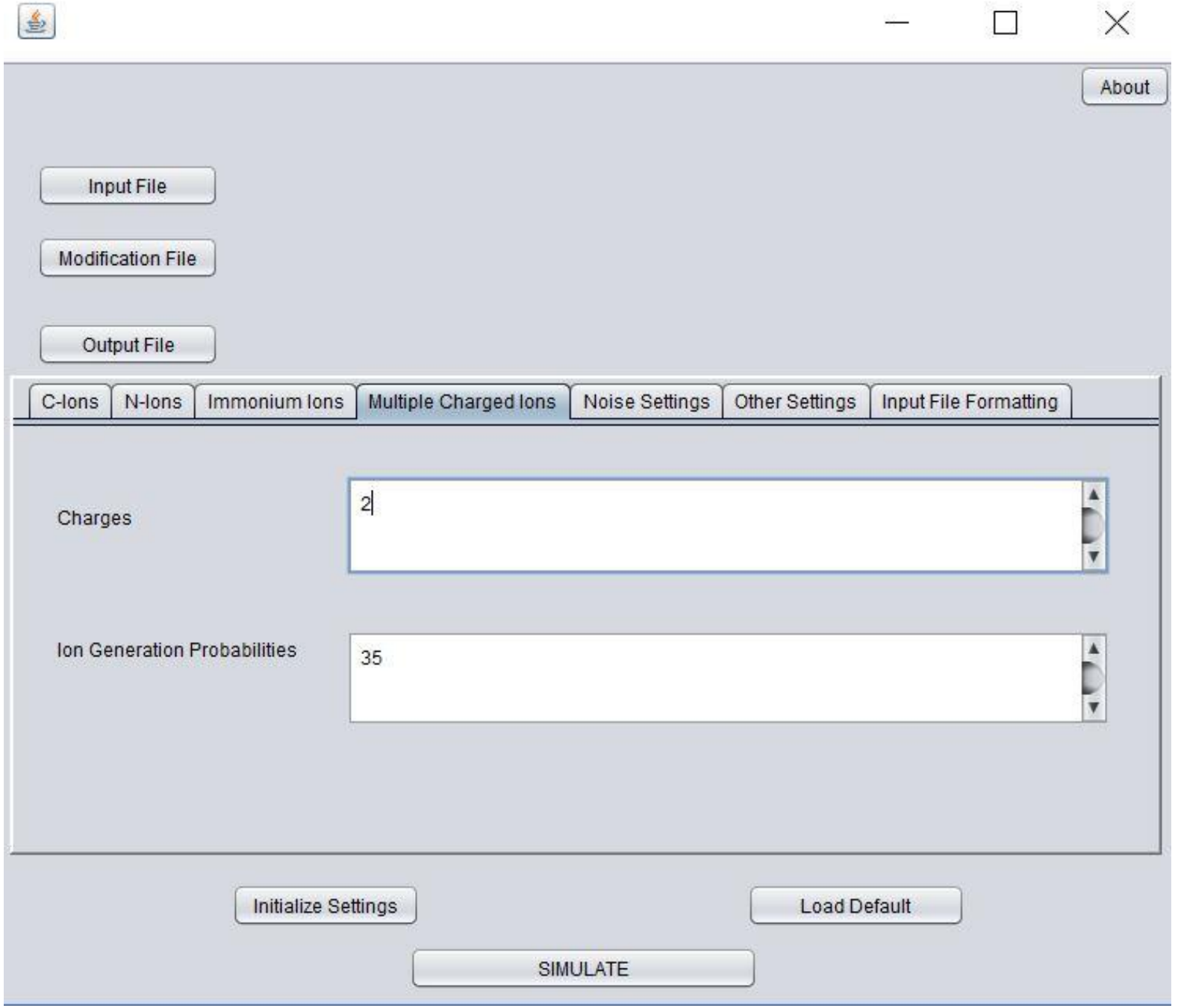
- Ion Generation Probabilities (IGP) determine the possibility of each ion to be generated. IGP values for C and N-terminal ions should be given in the same order as that of the mass-offsets using a space in between the values. For example, in on the right IGP value for y-ion series is 100% while 70% 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.



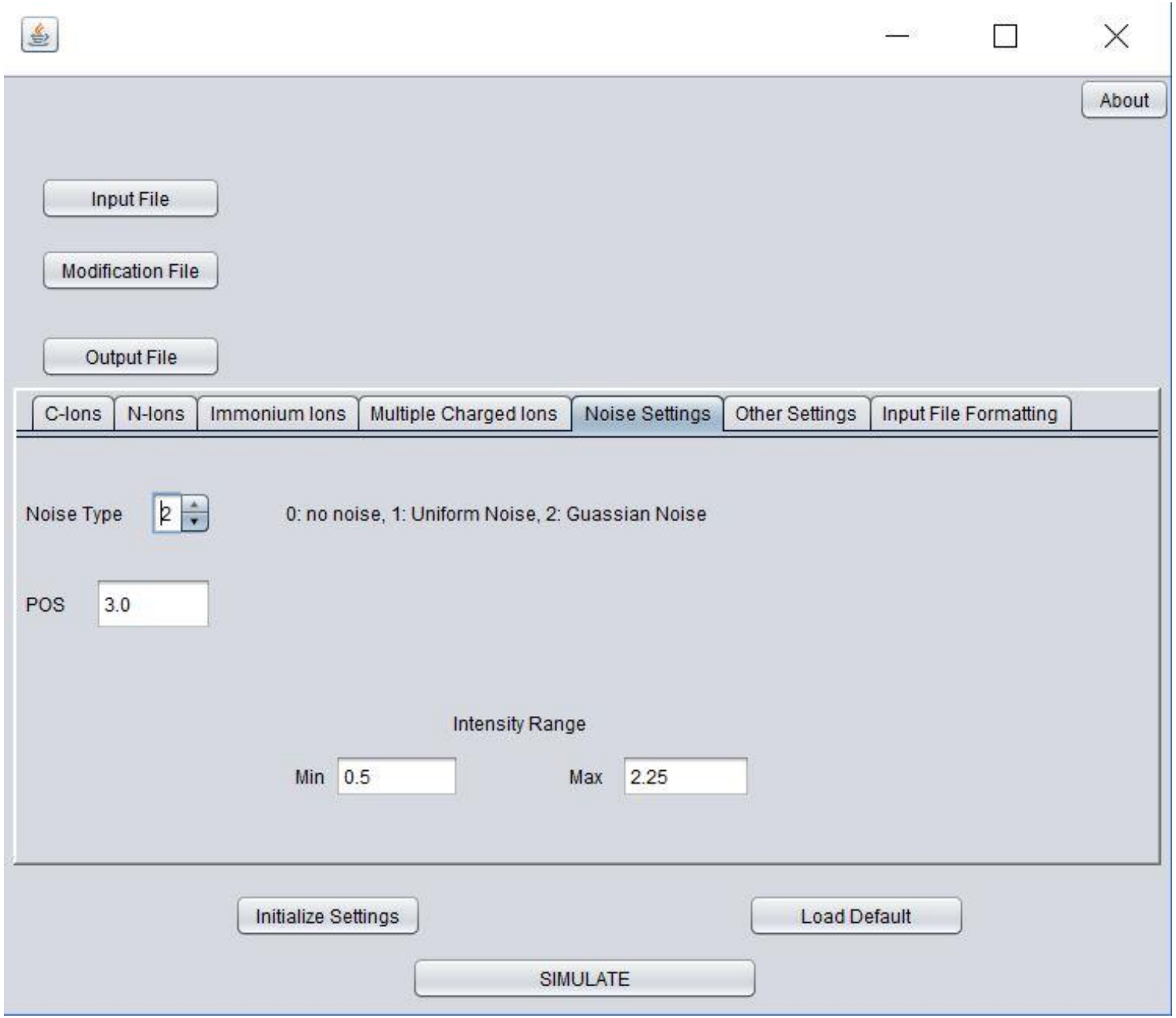
- Immonium ion configuration follows the same format with the exception of first field of “Amino Acids”. Here all those amino acids can be listed for which Immonium Ions need to be generated. These have to be listed using the single lettered code for the amino acids as shown in the Figure on right. IGP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



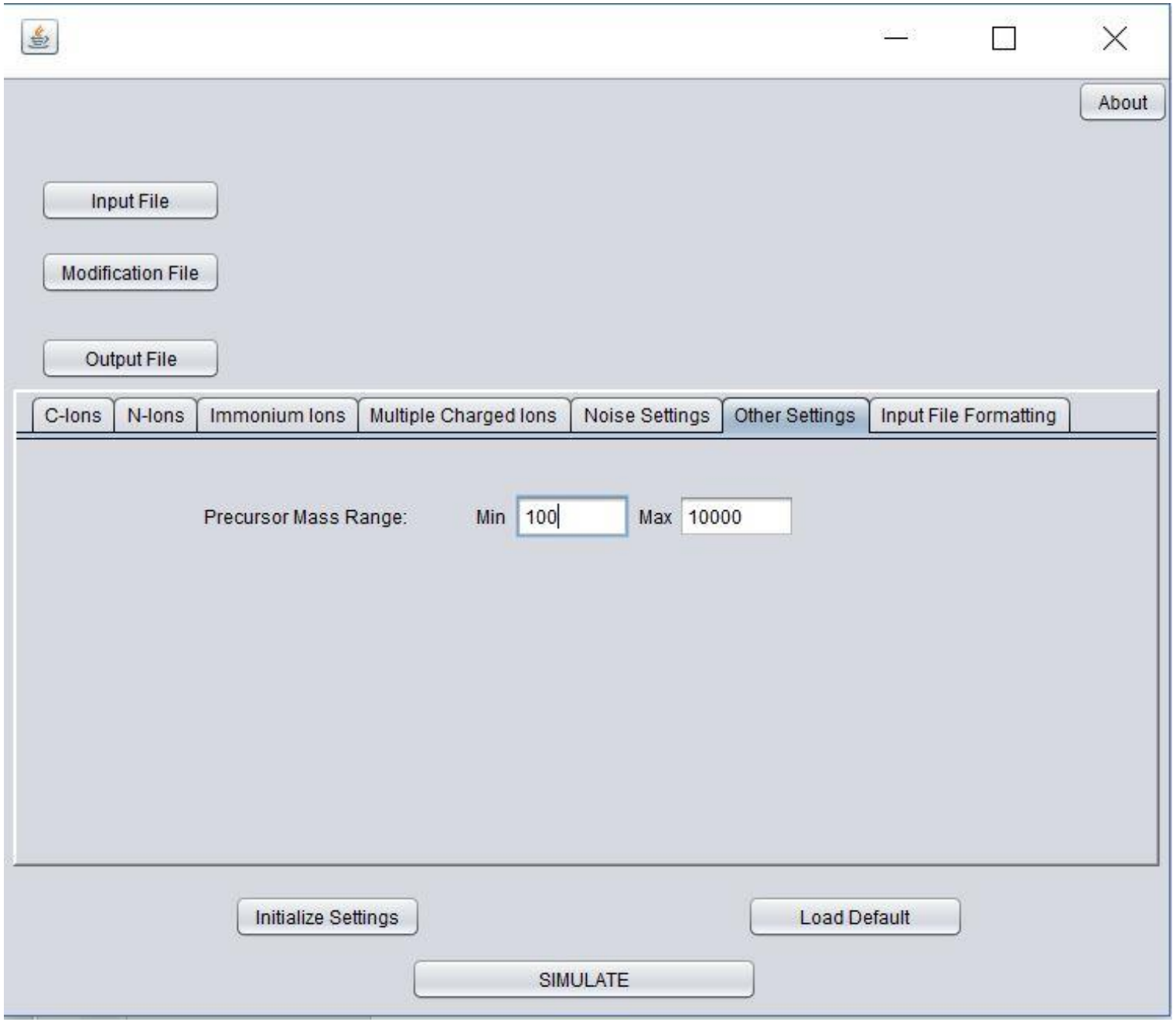
- To generate multiple charged variants of b- and y-ions a list of > +1 charges needs to be listed in “Charges” field.
- IGP values for these multiply charged ions can be listed in the field below in the same sequence as those of the charges.
- For example, in Figure on the right, ions with +2 charges 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



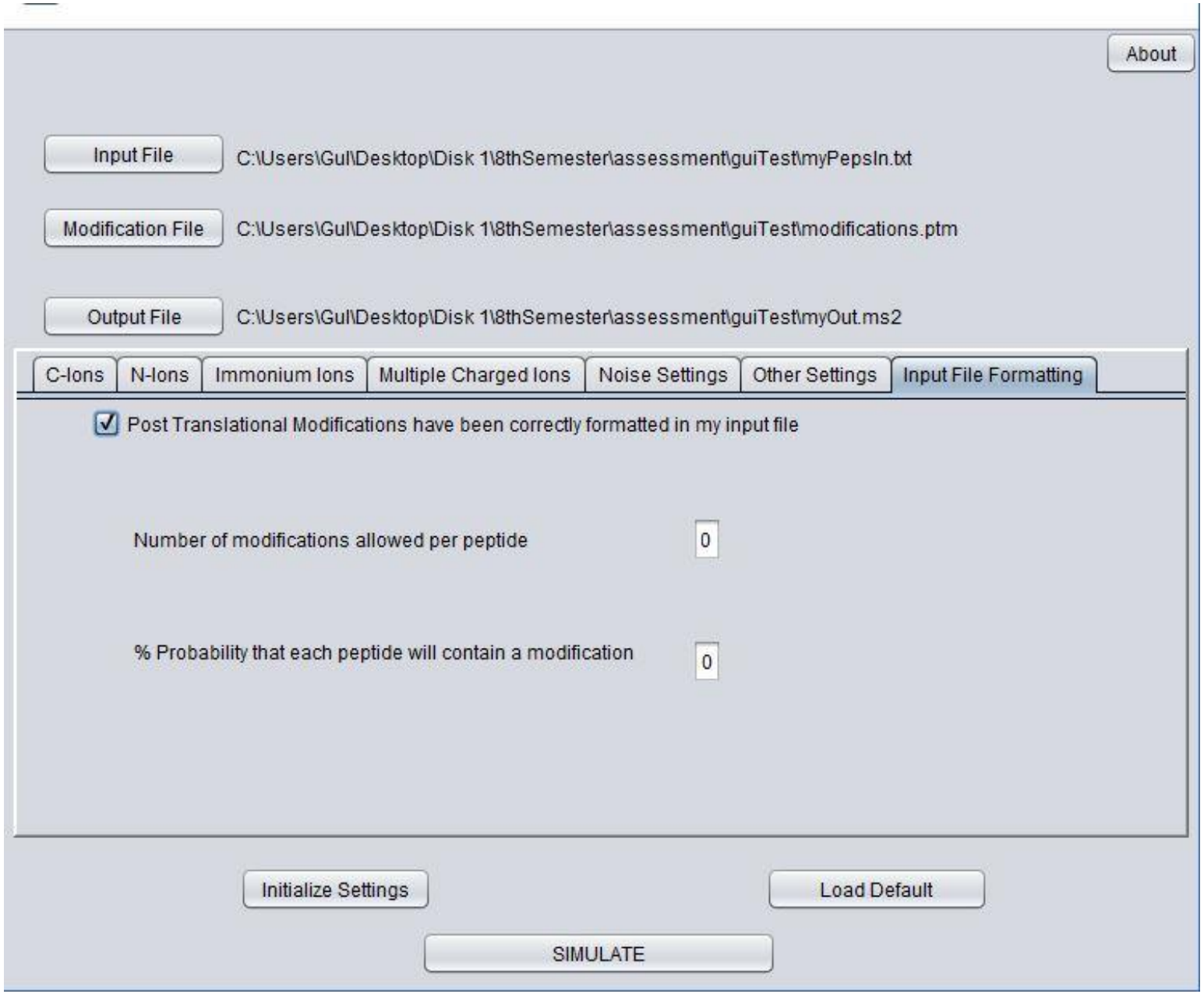
- Percentage of Sound (POS) value can entered in the POS field to control the noise content. A higher value would mean smaller noise. For instance in the figure on right, 3.0 means only 3% peaks in a spectrum would be sound (peaks on interest) and remaining peaks would be noise.
- To select the type 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).



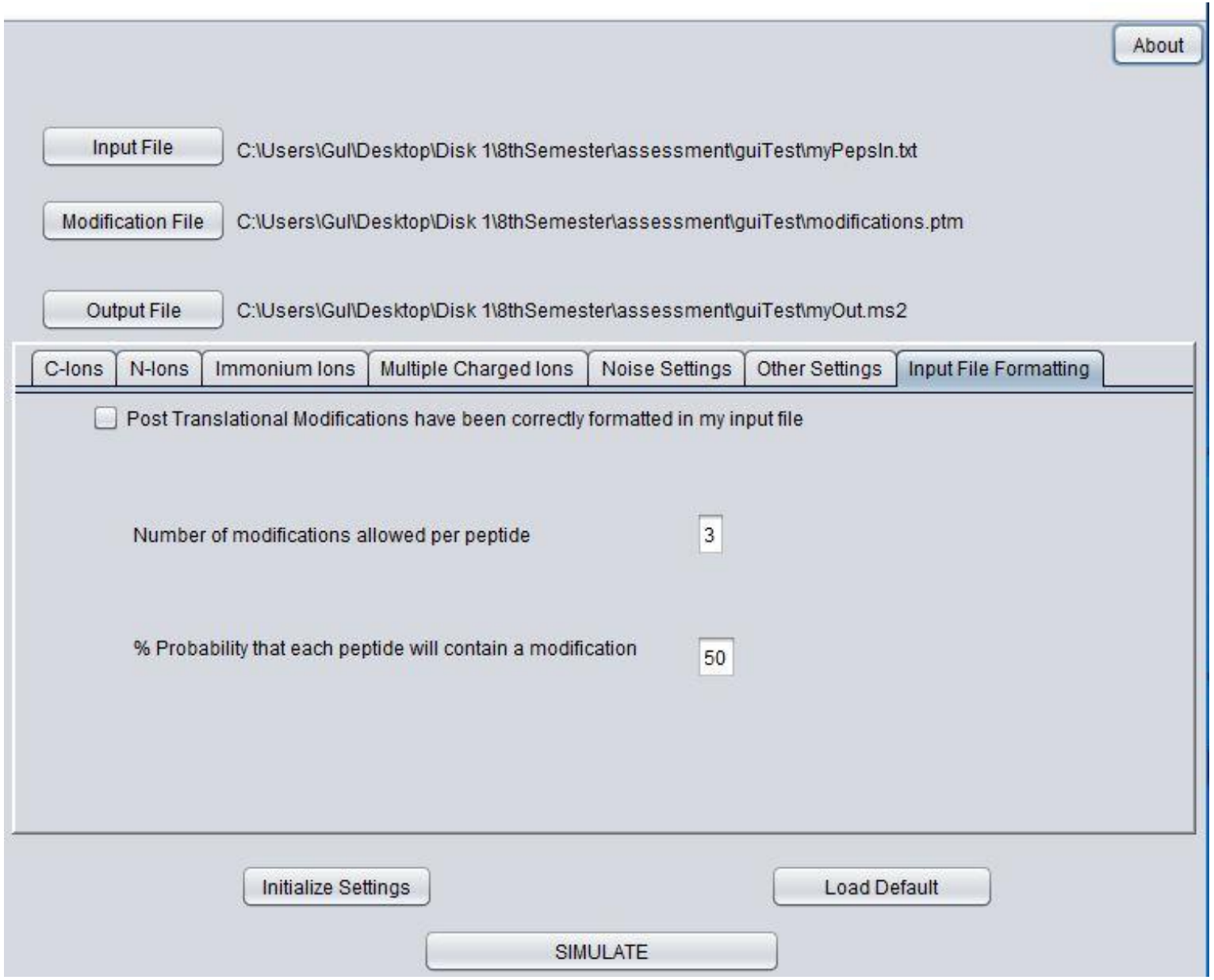
- To simulate spectra for peptides within a given range of mass, Precursor Mass Range can be provided. For instance, in Figure on right, shows the mass range from 100 to 10,000Da.



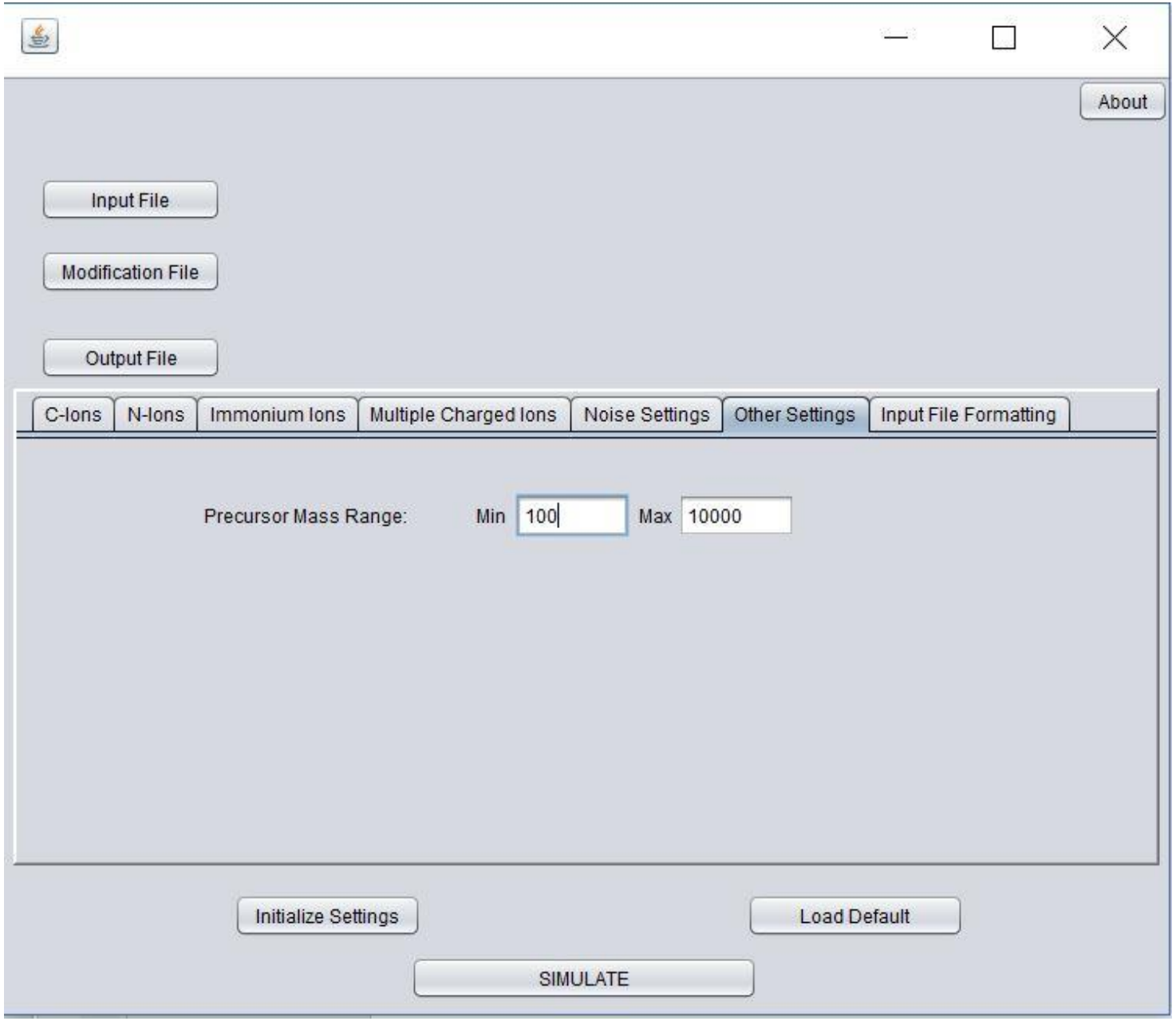
- For users who have an input file containing list of peptides (un-modified) and want to add modifications. Can make use of the “Input File Formatting” Tab.
- By default the check box will checked and both the fields below will be disabled. In this case the simulator will simulate the un-modified peptides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



- 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.



- Once all the paramters 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.



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

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