## MaSS-Simulator Configuration Manual

This manual explains different features offered by the MaSS-Simulator and explains how these features can be programmed using the configuration file.

### Installing and Running MaSS-Simulator

MaSS-Simulator has been developed using Java programing language and can be run on any operating system with **openJDK version 1.8**. Following is a step by step process to download and run MaSS-Simulator in Linux environment:

- Download the MaSS-Simulator source code and java class files from https://github.com/pcdslab/MaSS-Simulator.
- The downloaded directory contains *params.txt* and *modifications.ptm* files in addition to the source code and class files. Details of the files contained in the download package can be found in the readMe.txt file (included in the download package).
- To run the MaSS-Simulator with default parameters and default modifications following command can be executed:
    - **Java SimSpec 0 0 samplePeptides.txt sampleOut.ms2**
- The above command will read peptides listed in the samplePeptides.txt and generate MS/MS spectra based on the configurations listed in *params.txt* file and modifications listed in *modifications.ptm* file. Simulated spectra will be stored in *sampleOut.ms2* file while the ground truth data will be in a file named *peptides.rst.*
- To run MaSS-Simulator with user defined parameters and modifications following command can be executed:
    - **Java SimSpec 1 1 samplePeptides.txt sampleOut.ms2 myParams.txt myMods.ptm**
- The above command will read peptides listed in the samplePeptides.txt and generate MS/MS spectra based on the configurations listed in *myParams.txt* file and modifications listed in *myMods.ptm* file. Simulated spectra will be stored in *sampleOut.ms2* file while the ground truth data will be in a file named *peptides.rst*

## Configuring the parameters file

In the following sections details about configuring the parameter file are given. For all the below sections we will use the default parameters file as an example. Following this example, users can write their own parameters files for their experiments.

### C- and N-terminal Ion Generation

To list the C terminal ions which the user desires to be generated, mass offset value for that ion with respect to y-ions should be listed. All the desired ions should be listed as their mass offset values with spaces separating them, this list of offset masses should be preceded by the keyword “C-Ions”. For instance, in the Figure 1 highlighted line shows the C-terminal ions which are going to be generated. 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. A list of frequently occurring ions and their offset values are given in Table.1.

Similarly, second highlighted line in Figure 1 shows the list of B-terminal ions which will be generated. An offset of 0 means b-ion series will be added to the simulation while the offset of -28 means the a-ion series will be generated.

### Controlling the Ion Generation Probabilities (IGP)

IGP values for C and N-terminal ions can be entered by listing the keywords “C-Probabilities” and “N-Probabilities” respectively. IGP values for ion series should be given in the same order as that of the ion-offsets using a space in between the values. For example, in Figure 2 IGP value for y-ion series is 100% while 35% each for the isotopic ions and 65 and 50% for the neutral water and ammonia losses respectively. For N-terminal ions the b-series has IGP value of 80% while for mass offset -46 (a-NH3) it is 35%.

### Intensities for C and N-terminal Ions

Intensities for C and N-terminal Ions can be added using the keywords “C-Intensities” and “N-Intensities” 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. For N-

terminal ions the b-series has intensity value of 1.5 while for mass offset -46 (a-NH3) it is 0.25. All the intensity values are multiplied by a factor of 1000.

### Adding Immonium Ions to Simulation

Immonium ions can be added using the keyword of "Immonium-Ions" and following it with the single lettered code for the amino acid. IGP values and intensity values for immonium ions can be added using the keywords "Immonium-Probabilities" and "Immonium-Intensities" as shown in Figure 4. Intensities and IGP values for Immonium ions should be listed in the same order as their amino acid's code.

### Adding Ions with more than charges greater than +1

In spectra which have a precursor Ion of charge greater than +2, usually ions with charges greater than +1 appear. To make that happen, keyword "Max-Charge" can be used followed by a list of positive charges for which Ions need to be generated. IGP values for these ions can be listed using the keyword "Prob-Charge". For example, in Figure 5, the highlighted lines show that 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

### Controlling the Noise Content of Spectra

To set the sound to noise ratio (SNR), keyword "SNR" can be used, followed by the value of SNR of choice. To select the type of noise distribution keyword "Noise" is used, followed by 0,1 or 2. Where 0 means no noise, 1 selects uniform distribution and 2 selects gaussian distribution.

To adjust the intensities of noise peaks, keyword "Noise-Intensities" can be used, followed by the low range and then the upper range. For instance, in Figure 6, the highlighted lines show that an SNR value of 2.8 is being used to generate a gaussian distribution of noise peaks with their intensities uniformly distributed between 0.5 and 2.25 (these values are multiplied with a factor of 1000).

### Filtering Peptides based on masses

To simulate spectra for peptides within a given range of mass, the keyword "Mass-Range" can be used, followed by the lower range and then the upper range separated by a space in between. For instance, in Figure 7 the highlighted line shows the mass range from 100 to 10,000Da.

### Considering Peptide Modifications

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. 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 (based upon sign) 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.

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 35 35 35 35 35
N-Probabilities 35 35 35 35 35 35 35 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 1: Adding C and N terminal Ions*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 100 35 35 65 50
N-Probabilities 80 35 35 65 55 70 25 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 2: Adding IGP values for C and N terminal ions*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 35 35 35 35 35
N-Probabilities 35 35 35 35 35 35 35 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 3: Intensity values for C and N terminal Ions*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 35 35 35 35 35
N-Probabilities 35 35 35 35 35 35 35 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 4: Adding Immonium Ions*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 35 35 35 35 35
N-Probabilities 35 35 35 35 35 35 35 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 5: Adding Multiply charged Ions*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 35 35 35 35 35
N-Probabilities 35 35 35 35 35 35 35 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 6: adjusting noise peaks in the simulation*

```
C-Ions 0 -1.0 +1.0 -18.015 -17.013
N-Ions 0 -1.0 +1.0 -18.015 -17.013 -28.01 -45.023 -46.025
C-Probabilities 100 35 35 65 50
N-Probabilities 80 35 35 65 55 70 25 35
C-Intensities 1.75 1 1 0.25 0.25
N-Intensities 1.5 1 1 0.75 0.75 0.25 0.25 0.25
Immonium-Ions M D Q K E
Immonium-Probabilities 68 68 68 68 68
Immonium-Intensities 0.02 0.02 0.02 0.02 0.02
Max-Charge 2
Prob-Charge 35
SNR 2.8
Noise 2
Noise-Intensities 0.5 2.25
Mass-Range 100 10000
```

*Figure 7: Mass Filter*

| **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 |

Table 1