

Text S1: Details of peptide-to-spectrum matching

The followings describe the detailed implementation of peptide-to-spectrum matching.

1) Removal of charge-reduced precursor ions of ETD spectra

Peaks whose m/z value corresponds to the charge-reduced precursor ions are removed with given mass tolerance.

2) Non-monoisotopic peak filtering

Obvious non-monoisotopic peaks are removed, since spectrum-to-peptide matching is performed for theoretical monoisotopic peaks. It is accomplished by examining experimental peaks within 1.5 m/z and eliminating less intense peaks in the region.

- iterate through a vector of peaks in order of m/z value.

- examine the intensities of the other peaks in the vector that fall within $\pm 1.5 m/z$.

- If the intensity of the current peak is the highest among others, retain the peak. Otherwise exclude the peak from a vector of peaks. The m/z width of filtering is user editable.

3) Selecting most intense peaks

Fragmentation spectra potentially contain noise peaks which are not derived from the selected precursor ion, and most cases they are less intense than relevant fragment ion peaks. To exclude such noises in peptide-spectrum matching and retain relevant fragment ion peaks, observed fragment ions are first filtered to select intense peaks as follows. For example, consider a peptide with molecular weight (mw) 2000. Since the average of the molecular weight of an amino acid is 110 Da, one can roughly assume that the amino acid length of the peptide can be around 2000/110. The number of most intense peaks (n) to be selected is determined by setting coefficient (c) as follows.

$$n = \frac{mw}{110} c$$

The coefficient c is user editable. The c was set to 5 in the present study.

3) Intensity normalization

The intensities of the filtered spectrum are normalized to have maximum intensity 50. This value is user editable

4) Fragment Ion species.

In the present study, b-, b- that lost H₂O, y-, y- that lost NH₃, y- that lost H₂O are taken into account for CID spectra. For ETD spectra, c'-, c'-, z'-, and z'- are taken into account. User can choose ion species to be used.

5) Charge states of theoretical fragment ions

The charge states of theoretical fragment ion peaks to be matched are determined based on the charge states of precursor ions as shown in the following table.

Charge states of the theoretical fragment ions used for peptide-spectrum matching

charge state of precursor	CID		ETD	
	b-	y-	c-	z-
2+	1+	1+	1+	1+
3+ or higher	1+, 2+	1+, 2+	1+, 2+	1+, 2+

6) Generating the combinations of multiple variable modifications

To suppress the combinatorial expansion of the number of the possible forms of modified peptides when multiple variable modifications are selected, upper limitation of the number of combinations of modifications is set. In OMSSA, the limitation of the combination of modifications can be specified with `-mm` option from the command line. The creation of combinations is performed by the function “CSearch::CreateModCombinations” in “omssa.cpp” file of OMSSA source code. Our method (MPCombi) generates the combination of variable modifications in the same way as OMSSA.

Combinations are made for each peptide by

- Increasing the number of modifications on a peptide
- Iterating through potential modification sites on peptide
- Applying potential modifications on each sites

until the number of combinations reach the limitation set.

To illustrate this procedure, here we show an example how combinations are generated in OMSSA and our method. Consider a peptide sequence “GGFMTSEKSQTP” and the following variable modifications.

PTM name	target residue
[a]	peptide N-termini
[b]	peptide N-termini
[c]	E
[d]	P

The combinations of the position of modifications are made until the number of combinations reaches the upper limitation. By setting upper limitation to 9, the following combinations will be generated

number of PTMs on peptide	location of modified site	combination	number of combinations
0		GGFMTSEKSQTP	1
1	1	G[a]GFMTSEKSQTP	2
1	1	G[b]GFMTSEKSQTP	3
1	7	GGFMTSE[c]KSQTP	4
1	12	GGFMTSEKSQTP[d]	5
2	1 and 7	G[a]GFMTSE[c]KSQTP	6
2	1 and 7	G[b]GFMTSE[c]KSQTP	7
2	1 and 12	G[a]GFMTSEKSQTP[d]	8
2	1 and 12	G[b]GFMTSEKSQTP[d]	9

Then the theoretical masses of combinations are examined, and combinations that match the observed mass with a given mass tolerance will be used for peptide-to-spectrum matching processes.