Peak Detection Method in Mass++ --- A Protein Analysis Platform in Mass Spectrometry

Jingwen Yao[1], Satoshi Tanaka[1], Howell Parry[1], Tsuyoshi Tabata[2], Shin-ichi Utsunomiya[1], Shigeki Kajihara[1], Koichi Tanaka[1]

[1] Koichi Tanaka Laboratory of Advanced Science and Technology, SHIMADZU Corporation, Nishinokyo-kuwabaracho, Nakagyo-ku, Kyoto 604-8511, Japan
[2] Eisai Co., Ltd., Tokodai 5-1-3, Tsukuba, 300-2635, Japan

1: Introduction

Mass++[1,2] is developed in a software plug-in architecture. It is intended to provide various functionalities required by qualitative and quantitative analyses of proteins and peptides using mass spectrometry. It is free software and can be downloaded from the website: http://www.first-ms3d.jp/english/achievement/software.

With the services and plug-in tools provided by the Mass++ platform, users can also add their own software components to extend and accomplish specific analysis in combination with the elementary functions provided in the application (Figure 1). It accepts software components written in commonly used computing languages, such as C++/C, C# and so on. In this presentation, a new peak detection method developed recently is introduced, along with the way in which this function plugs into Mass++ to derive peak lists from mass spectra.

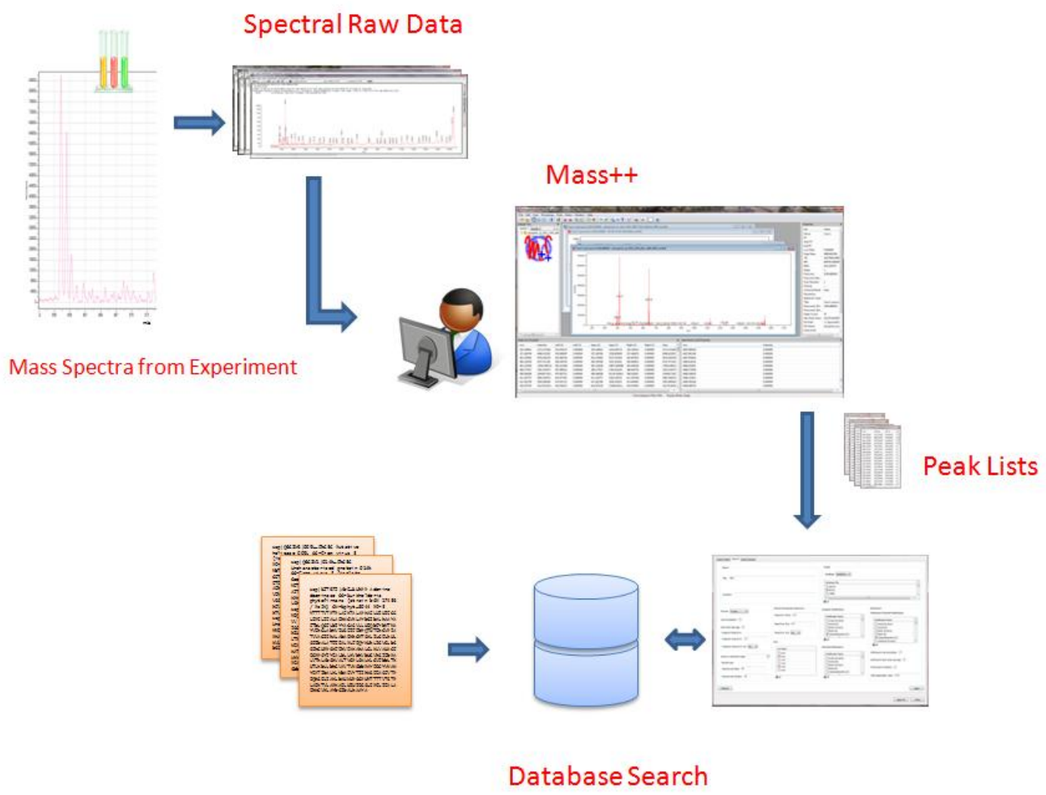


Figure 1 Protein identification from MS/MS spectra using database search in Mass++

2: Peak Detection

Peak detection is a starting point for protein analysis in mass spectrometry. The fragment ion peaks from a mass spectrometer are selected through this process. A quality peak list can usually ensure a more reliable result in the analysis.

2-1: Method

The peak detection receives spectral raw data from the input and includes several processes to get mono-isotopic ion peaks in the peak list, as shown in the flow chart in Figure 2. The raw data is firstly smoothing[3] and some pre-processing is carried out to enhance signal-to-noise ratio; and all the peaks are then detected in the spectra.

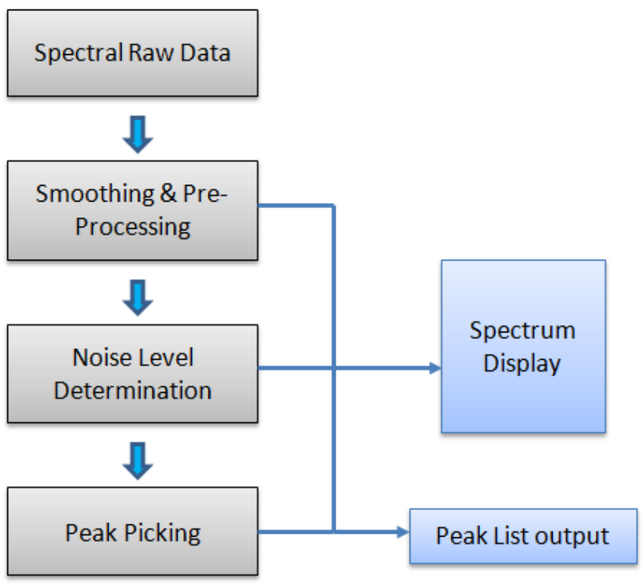


Figure 2. Flow chart for the peak detection

A new algorithm based on intensity classification is implemented in a software component to determine the noise levels in each selected mass range.

For spectral raw data, the noise peaks are assumed to fluctuate to a certain level. Strong ion peaks are significantly higher in intensity than these noise peaks. Therefore, for a given mass range, all the peak points are classified firstly. Three classes A, B and C, as illustrated in an example given in Figure 3 (b), are assigned from the classification according to their intensity values recorded in the spectra. The points in class C are considered as contributed only by noise; class A represents the ion peaks with significant heights; while the points in B are a possible mixture of ion peaks with low intensity and noise peaks. The noise level is determined from the class C.

Since the points in class A need to be selected and omitted from the determination of noise level, a Z-score method is applied to find these points as outliers from the main class C.

$$Z_i = \frac{|I_i - I_m|}{I_{sd}}$$

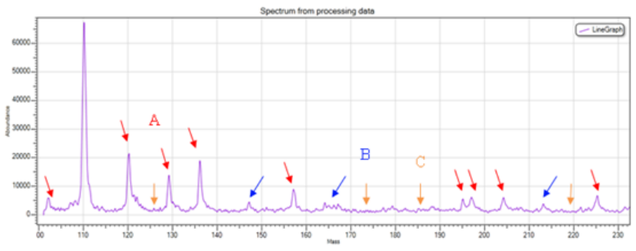
where I_m is the mean value and I_{sd} is the standard deviation calculated from all values.

Z_i value reflects how far a measured value I_i is from the mean value. A critical Z-value is deduced from training of a set of experimental data to instruct the selection.

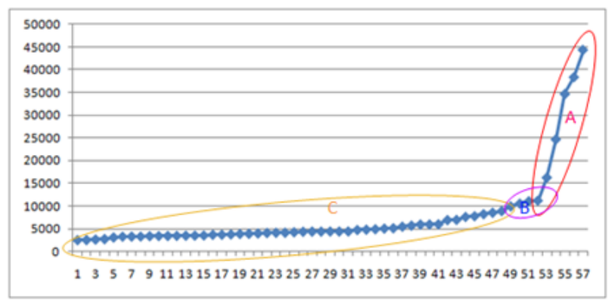
Once the noise level is determined, the ion peaks can be selected from a given signal-to-noise S/N ratio.

In peptide/protein identification from MS/MS spectra, only mono-isotopic ion peaks are useful. Therefore, an isotopic model to identify the isotopic clusters in MS and MS/MS spectra is applied to determine the selection of mono-isotopic peaks. This is usually the first peak in the found isotopic cluster in the peptide mass range. The model also considers the overlapping peaks caused possibly from two or more ions.

For ESI spectra, multiply charged ions are usually present. The model is also used to determine the charge state of the ions.



(a) Color code: red: peaks in A; blue: peaks in B; orange: peaks in C



(b) Intensity distribution in the selected mass range. The *y*-axis is intensity and *x*-axis represents the number of points in the range

Figure 3. An example of intensity classification

2-2: Plug into Mass++

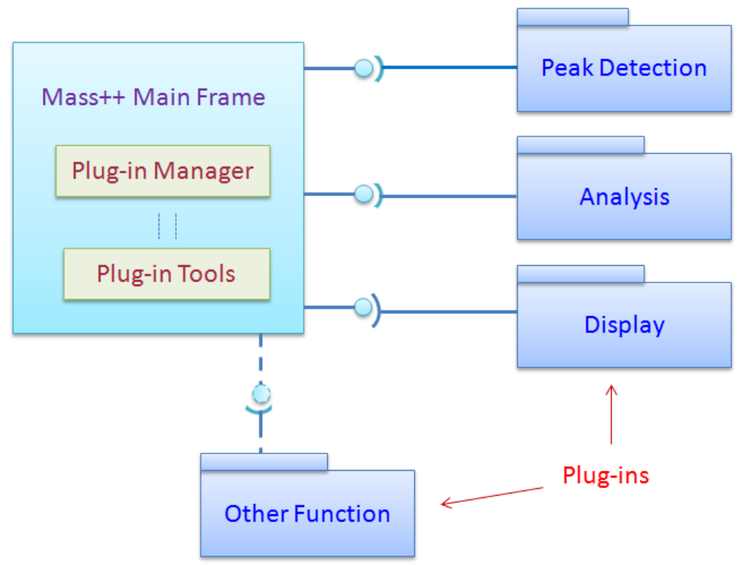


Figure 4. Mass++ plug in

The new peak detection is developed separately and named as MWD (**M**ass **W**indow **D**etection). All the functions are implemented in components of C# .NET in order to plug into Mass++ (Figure 4). Generating peak lists from input spectral raw data, MWD can be used with other plug-ins in Analysis and Display to perform an automated protein analysis. Users can select or set parameters required by the calculations through the user interface provided by the Mass++ plug-in. Any other required functionality may be added into Mass++ in a similar way.

3: Result

Table 1 shows a testing result on a set of LC-MALDI-TOF spectral data by using MWD as a peak picking tool to derive peak lists, and all peak lists were then submitted to the Mascot search engine to match proteins in the databases. For a comparison, the peak lists were also produced in parallel by a well-known commercial peak picking software (A). The protein scores obtained from the test for each sample are listed in the table for the peak lists produced from the two programs respectively. There are 23 samples from three proteins in the data set. Each data was acquired under different experimental conditions.

The column titled "Sample" in the table records the name of each data sample; "MWD" lists protein scores from this program while "Software A" gives the results from the other peak picking tool. It can be seen that the peak lists from MWD can mostly find more peptide matches and result in more reliable protein scores from the database search. The total score for these samples are 9509 and 6716 from two sets of peak lists, respectively.

Table 1. Mascot search result

Sample	MWD	Software A
091203-2_5p_BSAdigest+CHCA_B	955	689
091203-3_5p_BSAdigest+CHCA_B	890	473
091211-1_5p_BSAdigest+CHCA	1072	951
100121-1_5p_BSAdigest+CHCA_B	701	338
100121-2_1p_BSAdigest+CHCA_B	362	82
100125-1_5p_BSAdigest+CHCA_B	440	234
100125-2_5p_BSAdigest+CHCA_B	650	378
100125-3_5p_BSAdigest+CHCA_B	823	438
100127-1_5p_BSAdigest+CHCA	503	378
100127-2_5p_BSAdigest+CHCA	590	338
100128_5p_BSAdigest+CHCA_B	325	175
100203-1_5p_ADHdigest+CHCA_B	169	209
100203-2_1p_ADHdigest+CHCA_B	154	160
100203-3_5p_LZMdigest+CHCA_B	105	82
100203-4_1p_LZMdigest+CHCA_B	120	83
100203-5_1p_LZMdigest+CHCA_B	139	122
100208-1_500f_LZMdigest+CHCA_B	181	143
100217-1_5p_ADHdigest+CHCA	489	512
100217-2_1p_ADHdigest+CHCA	166	204
100217-3_500f_ADHdigest+CHCA	129	149
100222-1_5p_LZM-TrypsinDigest+CHCA	194	215
100222-2_1p_LZM-TrypsinDigest+CHCA	171	199
100222-3_500f_LZM-TrypsinDigest+CHCA	181	164
Total score	**9509**	**6716**

Reference

[1] Tanaka, S., Kajihara, S., Utsunomiya, S., Tabata, T., Aoshima, K., Oda, Y., Nihei, Y., Nishioka, T. and Tanaka, K., *59th ASMS Conference on Mass Spectrometry and Allied Topics*, 2011, Denver, USA.

[2] Parry, H., Tanaka, S. Tabata, T., Aoshima, K., Oda, Y., Nihei, Y., Nishioka, T., Utsunomiya, S., Kajihara, S. and Tanaka, K., *59th MSSJ Annual Conference on Mass Spectrometry* (2P-01), 2011, Suita, Japan

[3] Savitzky, A. and Golay, M.J.E., *Anal. Chem*., **36**(8), 1627-1639, (1964).

Acknowledgment

This research is supported by the Japan Society for the Promotion of Science (JSPS) through its "**F**unding Program for World-Leading **I**nnovative **R**&D on **S**cience and **T**echnology" (FIRST Program)