A Method for Auto-calibrating MS-MS Spectra for Peptide Analytes.

Alexis J Polley, Steve P Thompson, Victor C Parr. Scientific Analysis Instruments, Manchester, England

Aim

To improve the mass measurement calibration of peptide fragments resulting from post source decay (PSD) in tandem time of flight instruments.

If improvements can be made then the following benefits would result:

- Increase the sensitivity of peptide map fingerprint (PMF) experiments.
- More accurate *de novo* sequencing experiments.

Introduction

Conventionally in MS-MS, an external calibration must be used for the mass measurement of the peptide fragments. Unfortunately, the benefits, in terms of improved mass accuracy, of using an internal calibrant is not available since it is impossible to arrange for the unknown peptide to be accompanied a known calibrant of similar mass.

An alternative scheme is proposed that exploits the known mass sufficiency function (MSF). The MSF dictates the probability of a given accurate mass from a peptide's nominal mass - see Figure 1 and [1]. This work was originally motivated by the observation that the instrumental PSD fragment mass assignment errors are often much greater than the difference between the correct fragment mass and the nearest mass in the MSF - Figure 2 shows an example of this.

It is important to note that the accurate mass of a peptide's PSD fragments is not generally measured since at present modern MALDI ToF-ToF instrument's precursor ion selectors do not have sufficient resolution to select a single isotopic peak without compromising sensitivity. This means that the second mass analyser accepts more than one precursor ion mass which in turn leads to an extremely complex accurate mass calibration. Under these circumstances there are advantages in using the calculated average mass, accordingly the MSF is extended to apply to average masses (AMSF).

In summary this paper:

- Uses the Mass Sufficiency Function to assist the calibration of MS-MS data.
- Uses average masses to make this approach suitable for precursor selection by low resolution time ion gate.

Algorithm

The "auto-calibration" algorithm combines the externally calibrated masses, M_m with two linear parameters (slope, k and offset, c) and compares the result to the nearest most abundant mass in the AMSF. The squared difference is summed to produce a cost function C(c,k). The MSF applied to accurate mass is defined by Equation 1.

$$MSF(m) = 0.00048.m$$

Equation 1. Where m is the nominal mass in Da.

The typical mass distribution zone (MDZ) of 95% of possible amino acid combinations W(m) is defined in Equation 2. Masses outside of this region are in the zone of non-typical peptide masses.

$$W(m) = 0.19 + m/10000$$

Equation 2.

A plot of the MSF at nominal mass 775 Da can be seen in Figure 1.

The ASMF can be developed by producing a weighted average of the MSF using the abundance of the i^{th} isotope of the chemical composition of the "average peptide". This "average peptide" is defined as one constituted of an equal number of each of the 20 common amino acids. The AMSF(m) is defined in Equation 3.

$$AMSF(m) = \sum_{i=0}^{n} A(i)MSF(m-i)$$

Equation 3. Where n is the number of isotopes of the average peptide of mass m and A(i) is the abundance of that isotope.

A valid external calibration on a modern instrument can guarantee mass accuracy of better than 0.5 Da in a range along the mass scale adjacent to the precursor ion. Peaks in this range are selected for the optimisation of k and c by minimising the cost function defined in Equation 4.

$$C(m,k) = \sum_{j=0}^{p} (Min(AMSF(m) - (km_j + c)))^2$$

Equation 4. Where p is the number of measured masses (of error less than 0.5 Da).

Constraining the Optimisation of k and c

The values for k and c **must be constrained**:

- There are an **infinite** number of local minima in C(m,k).

This in turn means:

- Computation complexity can be unacceptable.
- The risk exists that we significantly **degrade** the mass assignment accuracy (>1.0Da): all but a handful of the local minima result in **incorrect** mass assignments.

The constraints are obtained by exploiting the fact that the original masses are assigned to within 0.5 Da. This equality is defined in Equation 5.

$$\left|((km_j + c) - m_j)\right| < 0.5(Da)$$

Equation 5.

The other constraint imposed is that the masses, corrected by c and k, should not lie in the zone on non-typical peptide masses (see Figure 2). This is defined in Equation 6.

$$\left|(Min(AMSF(m) - k(m_j + c))\right| < W(m)$$

Equation 6.

In summary, the values of k and c are constrained using:

- Mass accuracy of the input data (Equation 5).
- The mass distribution zone of the mass sufficiency function (Equation 6).

Method

A MALDI source tandem time of flight mass spectrometer equipped with harmonic field reflectron was used to measure the mass of fragments resulting from PSD of the following four peptides YYYYYY, DRVYIHPF, ADSGEGDFLAEGGGVR and DLFPDIIQDR. The peptides ADSG...GVR and DLFPDIIQDR were sulphonated using the reagent from Keough *et al* [2] this promotes Y-ion only fragmentation and leads to MS-MS spectra that are not only considerably more "human readable" but also more discriminative in the context of both PMF and *de novo* sequencing experiments. All PSD masses pertaining to the sub 0.5 Da region were selected for correction using the algorithm described above.

Results and Discussion

Two example annotated spectra are shown in Figure 3 and Figure 4 these illustrate the quality of data acquired.
The mean squared error (MSE) in Da for each of the peptide's fragments before and after calibration correction is shown in Table 1.

	ADSG...GVR	DRVYIHPF	DLFPDIIQDR	YYYYYY
Uncorrected	0.08506	0.02722	0.11779	0.06781
Corrected	0.05209	0.01910	Not solved	Not solved

Table 1

In two cases the mass correction algorithm improves the MSE by a significant margin. Here the correlation between the theoretical masses and those produced by the AMSF has been successfully exploited to reduce measurement error.

It is interesting that in the other cases a solution could not be reached.

Why can a solution sometimes not be found?

This possibility exists because of the constraints imposed on the linear parameters by Equation 5 and Equation 6. To illustrate this with an example, the peak mass assignments, their associated correct masses and the uncorrected nearest mass in the AMSF for the sulphonated peptide DLFPDIIQDR are shown in Table 2.

Measured Mass	Theoretical Mass (Error)	Min{AMSF} (Error)
1232.8	1232.37 (0.39)	1232.38 (0.40)
1117.7	1117.29(0.41)	1117.23 (0.35)
1004.3	1004.13 (0.16)	1004.08 (0.05)
856.6	856.95 (-0.35)	856.90 (-0.40)

Table 2

Considering the constraints and the data in Table 2 the reason for the algorithm returning in a not solved condition can be seen. Notice that the algorithm is likely to fail to find a solution if the following conditions are true:

- If the original mass measurements are in the zone of non typical peptide masses: in the example: Min{AMSF(1117)} 0.35 is greater than W(m) =0.29 from Equation 6.
- The errors are not a "good fit" to a linear function.

Although it is disappointing that a mass assignment improvement cannot always be found it does illustrate the **robustness** of the algorithm under adverse conditions. Clearly to risk a mass correction that led to a mass assignment error of greater than 1 Da would be totally unacceptable in this context.

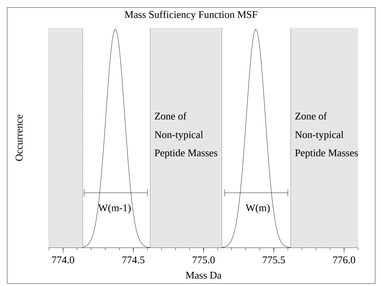

Figure 1 Mass Sufficiency Function

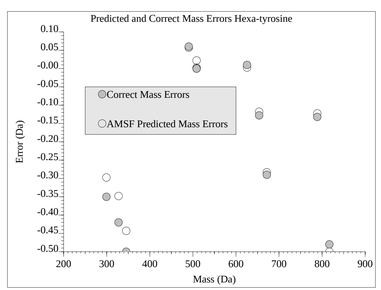

Figure 2 Mass Assignment Errors and AMSF Assignment Errors Showing the Correlation between Theoretical Mass Assignment Error and AMSF Predicted Errors in the Context of Instrumental Error

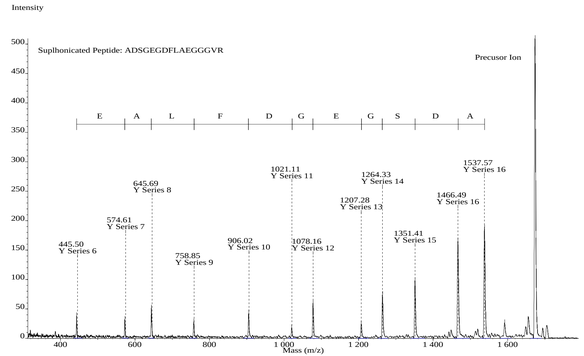

Figure 3 PSD Mass Spectrum Sulphonated ADSGEGDFLAEGGGVR

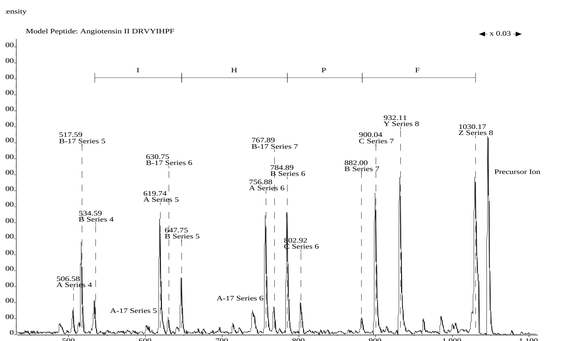

Figure 4 PSD Mass Spectrum DRVYIHPF

Conclusion

This paper presents an algorithm for improving the mass assignment accuracy of PSD experiments by exploiting the known mass sufficiency function. The algorithm has demonstrated significant improvements in mass measurement accuracy in some cases. At other times it has shown itself to be robust in that some input mass lists have been "rejected" as no solution can be found within the constraints presented here. These two facts mean that in the future it is worth refining the parameter constraints so that solution can more often be found. Furthermore it is a simple step to extend this technique to work in the time domain and would therefore be more likely to correct for calibration errors. Again more solutions should be found and this should be accompanied by a better mass accuracy optimisation.

[1] Mann, Possible Peptide Masses, Proceedings of the 43rd ASMS Atlanta 1995.
[2] Keough T, Lacey MP, Youguuist RS(2000) Rapid Comm. in Mass Spectrom. 14:2348-2356