3 - Compensating for variations in the mass defect

The mass defect

The mass defect is the difference between the monoisotopic and the integer mass value of a given amino acid residue. For example, with a monoisotopic mass of 71.03711 Da for Alanine, the mass defect is 0.03711 Da or 523 ppm.

The mass defect of the 20 amino acid residues in ppm:

- Ala: 523 ppm
- Arg: 648 ppm
- Asp: 299 ppm
- Cys: 567 ppm
- Glu: 457 ppm
- Gln: 376 ppm
- His: 430 ppm
- Ile: 428 ppm
- Lys: 413 ppm
- Met: 370 ppm
- Phe: 465 ppm
- Pro: 254 ppm
- Ser: 269 ppm
- Thr: 256 ppm
- Trp: 426 ppm
- Tyr: 388 ppm

The average mass defect is 494 ppm.}

Calibrating on the mass defect

An initial attempt to calibrate on the mass defect, i.e. performing a 'straight' multipoint linear calibration on the deviation from the average mass defect, yielded a reasonable mass precision in the order of 50-80 ppm. This was better than no calibration, but prompted an investigation into the nature of the mass defect.

As the most common enzyme by far is tryptic digestion, we used the entire Swiss Prot database (164,000 proteins) that were digested with tryptic chymotrypsin (20 residues) as a standard with 0.03711 Da mass defect calibration. This method is common in the proteomics field of MALDI mass spectrometry.

The average mass defect is 494 ppm. In silico and the peptide mass values from a tryptic digest against a list of known contaminations, contaminating peptides can easily be identified and removed. At the same time you perform a multipoint calibration leading to improved precision. These features leaves you with a much wider precision in the actual PMS search. Furthermore, the program helps you detect new contaminants either in a given dataset or after general use.

A novel development of the program enables the calibration using the mass defect which results in the calibration of mass spectra that we otherwise were not able to calibrate. In general the final precision is in the 30-50 ppm region.

4 - Multipoint and mass defect calibration

**The program**

Protein mass searching/fingerprinting (PMS) is one of the most common methods for identifying proteins in proteomics. The method relies on identifying proteins based on the mass values of the peptides generated by enzymatic digestion, typically trypsin digestion. Very advanced search programs are available today, making identification of proteins by PMS using tryptic MALDI-TOF MS data quite straightforward.

Two criteria are essential for a successful identification: A high number of (significant) mass values and high mass precision. In the current software we have tried to address both problems, with most emphasis on improving the precision. In a recent database the PMS data can be improved by removing non-significant values, i.e. remove peptide contamination. 2-D gel separated peptides are typically contaminated by keratin and trypsin autodigest peptides, but other project specific contaminants may be present.

Including these peaks in the peptide mass search will obviously make the search less precise and consequently they have to be excluded from the peptide mass list before using the list as input for the protein identification.

Calibration is usually performed as a two- or three-point calibration (either internal or external) of the mass spectra and is used in most cases the obtained accuracy is sufficient for identification of the protein in question. However, performing a multipoint internal calibration will improve the precision of the final dataset. Furthermore, the peaks used for calibration may be missing, of poor quality or the auto envelope may overlap with other peaks, making internal calibration next to impossible.

To address these problems we have created a small, user-friendly Windows-based program, PeakErazor, using a simple concept. By comparing all mass values from a tryptic digest against a list of known contaminants, contaminating peptides can easily be identified and removed. At the same time you perform a multipoint calibration leading to improved precision. These features leaves you with a much wider precision in the actual PMS search. Furthermore, the program helps you detect new contaminants either in a given dataset or after general use.

A novel development of the program enables the calibration using the mass defect which results in the calibration of mass spectra that we otherwise were not able to calibrate. In general the final precision is in the 30-50 ppm region.

**The mass defect**

The mass defect is the difference between the monoisotopic and the integer mass value of a given amino acid residue, i.e. with a monoisotopic mass of 71.03711 Da a mass defect of 0.03711 Da or 523 ppm.

The mass defect of the 20 amino acid residues in ppm:

- Ala: 523 ppm
- Arg: 648 ppm
- Asp: 299 ppm
- Cys: 567 ppm
- Glu: 457 ppm
- Gln: 376 ppm
- His: 430 ppm
- Ile: 428 ppm
- Lys: 413 ppm
- Met: 370 ppm
- Phe: 465 ppm
- Pro: 254 ppm
- Ser: 269 ppm
- Thr: 256 ppm
- Trp: 426 ppm
- Tyr: 388 ppm

The average mass defect is 494 ppm.}

5 - Isolated peptides

A major problem when analyzing MALDI-TOF separated peptides is that your only choice for calibration is either an external calibration or calibrating on the matrix peaks. However, in many cases the mass defect calibration can also be used here – it is mainly a case of sacrificing the precision of the very (inter) internationally M011 mass spectra.

A. The initial mass spectrum of an isolated peptide peak at m/z 1815. Not being microfuged the spectrum shows a number of adduct ions. The precision is still quite poor at 602 ppm.

B. Decreasing the signal to noise ratio to 2.5 ppm shows a large number of peaks, most of which are peptide related, but of very low intensity. This peptide mass list is copied into PeakErazor.

C. The mass defect calibration map shows a distribution both above and below the x-axis. This necessitates a manual calibration, performed by selecting ‘Manual’ followed by a click at both ends of the low mass ‘line of deviations’ as shown.

D. A following ‘Auto’ calibration on the measured mass values shows a substantial improvement in the precision.

E. After switching the graph display to ‘Mass defect view’, the same raw mass list is pasted into PeakErazor. The mass defect for each peptide is plotted relative to the mass and can be seen to form a diffuse sloping line.

F. Performing an linear multipoint calibration on the mass defect with removal of ‘outliers’ (<125 ppm) yields a diffuse scattering of points centered around the x-axis. The mass list is now calibrated and can be used for PMS.

6 - Conclusion

The current program depends on a combination of the human brain to pick out visual trends in a graph followed by a linear multipoint calibration. In the ‘standards’ calibration option of calibrating on contaminations a 20-50% improvement in the precision of the mass list can usually be achieved. In addition, contaminant mass values are removed from the search list, thus improving the search as the mass list gets smaller.

The program saves the mass search data as either removed or included in the search, and after several hundred spectra have been analyzed, the program enables you to identify system specific contaminants. Another function allows you to identify common values in a smaller dataset (i.e. contaminants specific to a particular gel or when analyzing isoforms/alleles).

The most recent addition to the program, mass defect calibration, comes into its own in the cases where no internal calibrants or contaminations can be calibrated on the data. By taking advantage of the fact that peptide mass values are not evenly distributed, but most values fall within ±125 ppm regions it is possible to calibrate any peptide containing mass spectrum given that a sufficient number of peptides are present. As a side effect, most non-peptide mass values can be removed from the list, as they will have a mass defect larger than 125 ppm.

The mass defect calibration does not work if the spectrum is heavily contaminated with non-peptide peaks (i.e. matrix, detergent or adduct ions). These cases are usually immediately recognized, as the spread of the mass defect becomes larger than ±125 ppm and it is possible to perform a ‘stable’ calibration. Few non-peptide mass values are easily removed and recalibration is just as problematic. Furthermore, the higher the number of peptides the more accurate the calibration. In practice you need ~10 mass values for a proper calibration. In a few cases (in our experience <5%) the method fails and we have to resort to external calibration.

Reloading the theoretical peptide mass list of the target protein enables you to identify normally assigned peptides (larger than average deviation) and helps to locate potentially modified peptides.

8 - Where to find PeakErazor

The PeakErazor program can be downloaded for free from http://welcome.to/gpmaw.

Please go to the ‘download’ section of the web site.

If you have any problems or have suggestions for improvements, please contact the author on php@bmb.sdu.dk.