USING STATISTICAL MODELS TO IDENTIFY FACTORS THAT INFLUENCE IONS FORMED IN A TANDEM MASS SPECTROMETER AND PREDICT ABUNDANCE OF THESE IONS

Ms Sheila Barton[1], Prof John Whittaker[2], Prof Sylvia Richardson[3], Dr David Perkins[4]

[1] University of Southampton Clinical Trials Unit, UK
[2]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK
[3] Division of Epidemiology, Public Health and Primary Care, Imperial College, London, UK
[4]Cancer Research UK Clinical Centre, University of Leeds, UK

Proteomic technologies play an important role in drug discovery diagnostics and medicine as they provide the link between genes, protein and disease. The classic workflow in proteomics consists of separation of proteins followed by identification of the individual proteins using mass spectrometry. If a protein proves difficult to identify it can be subjected to further mass spectrometry called tandem mass spectrometry or MS/MS.
Peptides or proteins fragment in the mass spectrometer to produce ions composed of shorter chains of amino acids. The intensity or height of a particular peak for one type of ion is a measure of how much of that ion has been formed. Different types of ions are produced according to the bond broken on the original peptide during fragmentation. In this study data from about five thousand peptides were subjected to mass spectrometry and the resultant spectra studied in order to discover factors influencing the intensity of the ion peaks.
Regression models were fitted using suitably transformed values of peak intensity as the outcome variable. Non linearity between some of the explanatory variables and the peak intensity was dealt with by the use of regression splines. Standard regression models do not take into account the fact that many ions can originate from the same peptide. Therefore mixed models were also fitted with a variable representing the peptide sequence as a random effect.
Several factors in the models were found to have a highly significant influence ($p<0.00001$) on the intensity of ions formed. These include the actual mass of the ion formed after fragmentation as well as the percentage of the original peptide mass. The composition of the fragmenting peptide was also found to be important: for example amino acids either side of the fragmentation site can greatly influence the intensity of ions produced.
In conclusion the models formulated provide useful information about the fragmentation process and allow prediction of the amount of various types of ion formed if the sequence of the peptide is known. Incorporating these models into algorithms for peptide identification should give improved performance, by using information on peak intensity which is currently ignored.