

Protein Identification by Database Searching

John S. Cottrell

Matrix Science Ltd.

Protein identification and characterisation by database searching is a well-established technique, generating in excess of 1000 peer reviewed publications last year.

The procedure is non-trivial because real mass spectra are not ideal transformations of molecular sequences. A real mass spectrum has finite resolution and finite signal to noise ratio, mass spectrometers have limited mass accuracy, enzymes do not cleave proteins with perfect fidelity, and peptide ions rarely dissociate into complete fragment ion series.

This talk is intended to provide an introduction to the subject. It will include an overview of the most commonly used methods, the software packages that are available, a discussion of the statistical nature of the technique, and some practical tips.

Suggested reading:

1. Aebersold, R. and Mann, M., Mass spectrometry-based proteomics, *Nature* 422 198-207 (2003)
2. Ashcroft, A. E., Protein and peptide identification: the role of mass spectrometry in proteomics, *Natural Product Reports* 20 202-215 (2003)
3. Baldwin, M. A., Protein identification by mass spectrometry - Issues to be considered, *Mol. & Cellular Proteomics* 3 1-9 (2004)
4. Elias, J. E., *et al.*, Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations, *Nature Methods* 2 667-675 (2005)
5. Henzel, W. J., *et al.*, Protein identification: The origins of peptide mass fingerprinting, *Journal of the American Society For Mass Spectrometry* 14 931-942 (2003)
6. Jensen, O. N., Interpreting the protein language using proteomics, *Nature Reviews Molecular Cell Biology* 7 391-403 (2006)
7. Johnson, R. S., *et al.*, Informatics for protein identification by mass spectrometry, *Methods* 35 223-36 (2005)
8. Lin, D., *et al.*, Large-scale protein identification using mass spectrometry, *Biochimica Et Biophysica Acta-Proteins and Proteomics* 1646 1-10 (2003)
9. Mann, M., *et al.*, Analysis of proteins and proteomes by mass spectrometry, *Ann. Rev. of Biochem.* 70 437-473 (2001)
10. Sadygov, R. G., *et al.*, Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book, *Nature Methods* 1 195-202 (2004)