EDITORIAL

Guidelines for experimental design and data analysis of proteomic mass spectrometry-based experiments

Karin Barnouin

The end of the 1990s saw a surge in the use of mass spectrometric applications in proteomics research. A huge amount of data was being generated. But proportionally little of the acquired data was being validated biologically. As the years went by it became evident that many false positives were being presented as bona fide results. In 2004, *Molecular and Cellular Proteomics* (*MCP*) generated guidelines for proteomics data (Carr et al. 2004). The editorial that accompanies the guidelines published in *MCP* in 2004 discusses the background to the criteria for publishable MS data. Those guidelines were largely adopted by *Proteomics* and *J. Proteome Res.* in subsequent years (see author proteomics data guidelines of respective journals). In 2006, the guidelines were made official and in 2009, *MCP* had a meeting in Paris to revise and update the guidelines, rendering the criteria for publication more applicable and broadened to include quantitative data (Bradshaw et al. 2006; Chalkley et al. 2009). It is essential, that journals, such as *Amino Acids*, publishing mass spectrometric proteomics data also adhere to these guidelines. We hope that other journals who publish mass spectrometry proteomics data will follow suit.

References

Bradshaw RA, Burlingame AL, Carr S, Aebersold R (2006) Reporting protein identification data: the next generation of guidelines. Mol Cell Proteomics 5(5):787–788

K. Barnouin (✉)
London Research Institute, Cancer Research UK, London, UK
e-mail: karin.barnouin@cancer.org.uk

Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A (2004) The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data. Mol Cell Proteomics 3(6):531–533
Chalkley R, Clauser K, Carr S (2009) Updating the *MCP* proteomics publication guidelines. ASBMB Today, August 16–17

Guidelines for experimental design and data analysis of proteomic mass spectrometry-based experiments

The experimental design and MS acquisition parameters must be detailed in Methods or Supplementary sections.

1. Protein and peptide identification and characterisation

The method(s) used to generate peak lists from raw MS and MS/MS data should be described. The name and version(s) of the programmes used for database searching and search parameters should be detailed, i.e. parent and fragment ion mass tolerances, number of missed cleavages, and modifications. Any processes, such as peak smoothing, summing or deisotoping performed on the raw data that could affect its quality should be mentioned.

For each protein identified, the database accession number, the statistical measure, such as p value or MASCOT score must be given, as well as the number of peptides identified by MS/MS per protein should be listed including sequences and charge state. For peptide mass fingerprinting, the number of peptides used for the identification should equal $n > 5$ and the sequence coverage should be provided. Where possible, at least two peptides should be subjected to MS/MS fragmentation sequencing to confirm the identification. For large experimental datasets, a measure of the false positive rate, such as that obtained from a decoy database search, is required. For all

data, the methods, threshold applied and statistical analysis used to interpret MS/MS data to judge certainty of identification should be described.

For post-translational modifications, the method used to obtain the characterisation should be stated. In the case of isobaric modifications, such as phosphorylation and sulfation, mass spectrometric, biological or chemical evidence should be provided proving the correct assignment of the modification. In addition, MS/MS fragmentation spectra and peptide sequence should be provided. The spectra should at least be submitted for review purposes. Together with the spectra it should also be stated whether or not the site of modification can be unambiguously assigned.

2. Quantitative experimental design and data analysis

The experimental design must be provided with at least three biological replicates. For quantitative experiments, software for data extraction and software used for data analysis should be detailed. In addition, statistical tests must be performed and the results reported. Furthermore, any extra handling of data such as details of outlier removal methods utilised should also be provided, e.g. those based on signal to noise or statistical tests like Dixon's outlier removal test. Moreover, methods of data normalisation are to be described. Where possible, the results should be validated with the biologically relevant experiments.