Minimum Reporting Guidelines for Proteomics Released by the Proteomics Standards Initiative

Andrew R. Jones‡ and Sandra Orchard§

The Human Proteome Organisation Proteomics Standards Initiative (HUPO-PSI)¹ has recently published the Minimum Information About a Proteomics Experiment (MIAPE) specification. MIAPE is a set of documents aimed at formalizing the information or metadata that should be reported about a proteomics experiment, for example when it is published in a journal or a data set is deposited in a public data base. MIAPE is divided into a parent document that lays down the principles and development process (1) and a series of modules, generally one per technology. The first four modules have recently been published for molecular interactions (2), mass spectrometry (3), mass spectrometry informatics (e.g. data analyses for peptide and protein identification) (4), and gel electrophoresis (5), and several additional modules are close to finalization, describing column chromatography, gel image informatics, and capillary electrophoresis. Each module comprises a checklist of items that should be reported about how the experiment was performed and the data types that should be provided. The reported information should allow the conclusions reached by the study investigators to be critically evaluated and should contain sufficient details such that protocols could be re-applied by other laboratories. Carr et al. (6) and Wilkins et al. (7) have also examined the issue of improving the quality and reliability of peptide and protein identifications, and their requirements have been taken on board in the development of MIAPE. The consistent reporting of technology-independent factors is also important, such as sample descriptors and phenotypes, and they are being discussed in the forum of the Minimum Information for Biological and Biochemical Investigations; a cross-technology consortium that includes PSI (8).

In common with other experimental approaches, the results of a proteomics study are dependent on both biological and technical variability. Where proteomics differs from other experimental approaches is the sheer range of methods available for protein or peptide separation, mass spectrometry, and data analysis; there is little consensus on the best approach for any stage of a workflow. As such, without a mechanism for the consistent reporting of metadata, it is highly challenging to assess the quality of results, and teasing apart the causes of biological or technical variation between different studies is not possible. In practical terms, to be “MIAPE compliant” requires the following steps to be followed. When publishing a journal article or sending a data set to a public repository, the experimentalist should download the relevant technology modules from the PSI website. The recommended information can be provided in a number of ways depending on the context. It may be appropriate to report the protocols employed within the Materials and Methods section of the article and provide a link to data sets, stored in a public data base. While this would be minimally MIAPE compliant, it would not assist data base users wishing to query for particular details. In other contexts, metadata can be included directly in the relevant data format, such that it can be queried and reviewed directly alongside results in a data base. It is anticipated that the level of detail requested by MIAPE Mass Spectrometry and MIAPE Mass Spectrometry Informatics will be exported automatically by the instrument and analysis software, for instance using PSI's mzML format for spectral data and the upcoming AnalysisXML format for peptide and protein identifications.

MIAPE is not intended to place an unnecessary burden on experimentalists. Past experience has shown that by formalizing the set of requirements and providing data standards increases the availability of free open source software for recording and managing the large quantities of data. In the absence of standards, experimentalists are left with the problem of developing ad hoc solutions or purchasing expensive software that may be difficult to tailor for new tasks or deviations from standard approaches. It is not an issue that can be safely ignored by laboratory scientists since data sharing is now a major part of the landscape of omics science. Proteomics experiments are expensive to perform, and in academic research, they are generally funded by research councils, government, or charitable organizations. Funders do not want to see potentially highly valuable data sets sitting in storage in laboratories, or worse, irretrievable after a researcher moves on to a new organization. As one example, the United Kingdom's Biotechnology and Biological Sciences Research Council has a data sharing policy. Funded proposals must guarantee to make data sets publicly accessible and resources associated with data sharing can be costed into applications. MIAPE can be seen as an attempt to derive maximal value from data. If poorly annotated data sets are

¹ The abbreviations used are: HUPO-PSI, human proteome organisation proteomics standards initiative; MIAPE, minimum information about a proteomics experiment.
deposited in data bases, this may fulfil the basic requirement of a journal or research council for data to be made available, since they will rarely have the resources to check individual files for the detail therein. MIAPE provides a framework so that the quality of data annotation can be checked automatically; for instance, allowing software to be created that check for compliance.

Considerable quantities of proteome data are already available. As of August 2008, PRIDE contains 10,422,704 spectra, PeptideAtlas 7,711,914 spectra, and GPM contains 9,184,251 peptide identifications; although many of the early depositions in these repositories do not contain sufficient metadata to fully comprehend how the experiment has been performed, limiting the potential uses of the data. We will undoubtedly see a rapid improvement in the quality of data annotation in these repositories, allowing new biological findings to be made beyond the original study objectives, for genome annotation or for improving data analysis algorithms.

The MIAPE modules will continue to evolve as new experimental techniques are devised, and they are dependent on input from experimentalists about the level of detail that should be reported. We would like to encourage feedback on the modules, for example through the PSI’s mailing lists or by attending a PSI workshop or meeting.

§ To whom correspondence should be addressed: Ph.: 44-1223-494 675; Fax: 44-1223-494 468; E-mail: orchard@ebi.ac.uk.

REFERENCES