



The following guidelines were developed as a revision to those formulated previously (Paris guidelines) at a workshop held in Philadelphia on May 30<sup>th</sup> of this year and attended by about 25 people with various interests and areas of expertise. It is anticipated that these guidelines will be adopted by the Journal around January 1st 2010.

The editors of MCP welcome comments and criticisms from the public, which may be emailed to: [mcpeditor@asmb.org](mailto:mcpeditor@asmb.org)

The identification of proteins or peptides is generally accomplished by peptide sequencing (MS/MS analyses), peptide mass fingerprinting (PMF), or some combination of both. Both methods typically depend on acquiring mass spectra, conversion of data into a format for searching, followed by interpretation via matching the data against a sequence database or spectral library with an appropriate search engine. Many parameters are common to the two approaches and some are unique to each. Similarly, post translational modifications (PTMs) can affect protein identifications; their determination/localization may also be the objective of the experiment. Finally, quantification by isotope-dependent or isotope-free methods may be an additional aspect of the data set. The following guidelines describe the information required by the journal for articles dealing with mass spectrometric analyses designed for protein, peptide or PTM identification and their quantification. **Manuscripts that use generated datasets for software or algorithm development may provide only the parameters in section I below providing they do NOT report the identities of the peptides or proteins used. If identifications or sequences are listed, the requirements of Section II (and III, if germane) must be met.**

## **I. SEARCH PARAMETERS AND ACCEPTANCE CRITERIA**

The following supporting information should be included in the Experimental section of the manuscript for either MS/MS or PMF analyses (recognizing that some software do not explicitly provide some of these parameters):

- Peak Lists: The method and/or program (including version number and/or date) used to create the "peak lists" from the original data and the parameters used in the creation of this peak list, particularly any processing which might affect the quality of the subsequent database search. Examples include smoothing, any signal-to-noise thresholding, charge states assignment or de-isotoping, etc. In cases where additional customized processing of the collections of peak lists has been performed, e.g. clustering or filtering, the method and/or program (including version number) should be referenced.
- Search Engine: The name and version (or release date) of all programs used for database searching must be provided.
- Sequence Database or Spectral Library: The name and version (or release date) of all sequence database(s) or spectral libraries used must be listed. If a database or library was compiled in-house, a complete description of the source of the sequences or spectra is required. The number of entries actually searched from each database or library should be included. If the database or library used



---

**DRAFT: Revised Publication Guidelines for Documenting the Identification of Peptides, Proteins, and Post-Translational Modifications by Mass Spectrometry**

is very small (<1000 entries) or excludes common contaminants, justification must be specifically provided since this may generate misleading assignments and an inaccurate false discovery rate estimate.

- Enzyme specificity: A description of all enzymes used to generate peptides, including the number of missed and non-specific cleavages (e.g. semi-tryptic) permitted, must be listed.
- Fixed modification(s): A list of all modifications considered (including residue specificity) must be given.
- Variable modifications: A list of all modifications considered (including residue specificity) must be given.
- Mass tolerance for precursor ions.
- Mass tolerance for fragment ions (not required for PMF data).
- Known contaminants excluded (particularly for PMF data): All omitted peaks from pre-designated contaminants (or if any of these fragments are used for calibration) must be identified.
- Threshold score/expectation value: Criteria used for accepting *individual* spectra should be stated along with a justification.
- False Discovery Rates at Peptide and Protein levels: For large scale experiments, the results of any additional statistical analyses that estimate a measure of identification certainty for the dataset, or allow a determination of the false discovery rate, e.g., the results of decoy searches or other computational approaches.

## II. PROTEIN AND PEPTIDE IDENTIFICATION

The information below for each protein and peptide sequence identified should be specified in the Results (or Supplemental) section. If the identifications are presented only at the peptide level, then protein level information may be omitted.

- All peptide sequences assigned: A list (in one or more Tables), noting any deviation from the expected enzyme cleavage specificity, must be provided..
- Precursor charge, mass/charge, and mass error observed: These parameters should be listed for each peptide assignment in the same table.
- All modifications observed.
- Protein accession number and sequence database or spectral library source;
- Count of the number of distinct peptides assigned to each protein: To compute this number multiple matches to peptides with the same primary sequence count as one, even if they represent different precursor charge states or modification states.
- Protein sequence % coverage: This value should be expressed as the number of amino acids spanned by the assigned peptides divided by the sequence length X 100.



---

**DRAFT: Revised Publication Guidelines for Documenting the Identification of Peptides, Proteins, and Post-Translational Modifications by Mass Spectrometry**

- Number of matched and unmatched masses: For PMF data, the total number of masses for each peptide peaks, both, divided between matched and unmatched, should be listed in the identification table.
- Score(s): The relative relevant score (depending on the software used) and any associated statistical information obtained for searches conducted must be provided for each peptide.
- **For all proteins identified on the basis of ONE OR TWO unique peptide spectra, (a protein mass fingerprint spectrum counts as one spectrum), the ability to view annotated spectra for these identifications must be made available.** This can be achieved in one of three ways:
  - Submission of spectra and search results to a public results repository that is equipped with a spectral viewer (e.g. PRIDE, Peptidome etc). This information will appear as a hyperlink in the published article.
  - Submission (with the manuscript) of spectra and search results in a file format that allows visualization of the spectra using a freely-available viewer.
  - Submission (with the manuscript) of annotated spectra in an 'office' or PDF format.

**Note: Files submitted through the online manuscript submission process must be less than 100 Mb in size. If files are greater than 100 Mb in size, the journal recommends depositing the file in a suitable repository, such as Tranche [<https://proteomecommons.org/index.jsp>], then supplying the hash (from Tranche) (or other identification code) in the manuscript and in the cover letter accompanying the manuscript submission. Posting of results on the author's website as the sole source of this data does not satisfy this requirement, as the ability to anonymously access the data is necessary for the review process.**

### III. POST-TRANSLATIONAL MODIFICATIONS

Studies focusing on posttranslational modifications require specialized methodology and documentation to assign the presence and the site(s) of modification. No current MS data analysis software is infallible in the automatic assignment of modification sites in peptides, and these analyses are particularly error prone when multiple possible sites within a peptide are being utilized. For these reasons, additional documentation supporting assignment of PTMs is required. In addition to the tabular presentation(s) of the data described in guideline II:

- The site(s) of modification within each peptide sequence must be clearly presented.
- An indication of the certainty of localization for each PTM: The manner in which the modification was located (by computation or manually) and a description of the software used, if any..
- A justification for any localization score threshold employed.



---

**DRAFT: Revised Publication Guidelines for Documenting the Identification of Peptides, Proteins, and Post-Translational Modifications by Mass Spectrometry**

- **Ambiguous assignments:** Peptides containing ambiguous PTM site localizations must be listed in a **separate table** from those with unambiguous site localizations. In cases where there are multiple modification sites and at least one is ambiguous, then these peptides should be listed with the ambiguous assignments. Ambiguous assignments **must** be clearly labeled as such.

Examples of ambiguities include:

- Modified peptides in which one or more modification sites are ambiguous.
  - Instances where the peptide sequence is repeated in the same protein so the specific modification site cannot be assigned.
  - Instances in which the same peptide is repeated in multiple proteins, e.g. splice variants (See also Section IV).
  - Isobaric modifications (e.g., acetylation vsvs. trimethylation, phosphorylation vsvs. sulfonation etc), where the possibilities may not be distinguished. Examples of methods able to distinguish between these include mass spectrometric approaches such as accurate mass determination, observation of signature fragment ions (e.g. m/z 79 vsvs. m/z 80 in negative ion mode for assignment of phosphorylation over sulfonation), or biological or chemical strategies.
- **Annotated, mass labeled spectra:** Spectra for **ALL** modified peptides must be either submitted to a public repository or accompany the manuscript as described in guideline II.

#### **IV. PROTEIN INFERENCE FROM PEPTIDE ASSIGNMENTS**

Because experimental strategies based on proteolytic digestion of protein mixtures result in the loss of connectivity between peptides and their protein precursors, assignment of peptide sequences can result in a combination of two possible outcomes; distinct peptides that map to only one protein sequence or shared peptides that map to more than one protein sequence. Consequently, authors are expected to assemble peptides into proteins and protein groups containing shared peptides while adhering to principles of parsimony, i.e., describe the minimum set of protein groups that adequately account for all distinct peptides observed. In addition to the tabular presentation(s) of the data described in guideline II, authors are required to:

- When reporting a protein group, accession numbers or identifiers should be provided for all proteins that were combined into the group. Authors should explain and be able to justify cases where a single protein from a protein group has been singled out or when asserting that more than one indistinguishable member of a protein group is actually present.
- When reporting a summary list of peptides belonging to each protein group, peptides shared among multiple proteins and those unique to a specific protein should be clearly indicated.



---

***DRAFT: Revised Publication Guidelines for Documenting the Identification of Peptides, Proteins, and Post-Translational Modifications by Mass Spectrometry***

- If proteins are identified from a different species than the one being studied, the circumstances should be mentioned and justified. For example, identification of a mouse or human protein in a hamster study.

## **V. QUANTIFICATION**

Manuscripts presenting quantitative proteomic results must provide significant information about the experimental method and design, as well as the employed data analysis methods.

The following information must be provided:

- All relevant quantification data (as part of the peptide and protein identification tables), along with a description of how the raw data was processed to produce these measurements.
- A description of how the analytical reliability of measurements was validated using technical replicates and statistical methods. Citation of standard methods or specialized software may be used. However, it is essential to demonstrate that the data contained in the manuscript actually conforms to the same models.
- A description of how the biological reliability of measurements was validated using biological replicates, statistical methods, independent experiments, etc. Studies based on a single biological experiment are generally not acceptable (except as a dataset to test bioinformatic systems). If a biological replicate from the same source cannot be performed (e.g. patient sample), a large enough number of similar biological samples must be performed in order to enable sound conclusions.
- A description of the treatment of relevant systematic error effects such as interference from overlapping precursor ions, incomplete isotope labeling, bias correction for pipetting error, etc.
- A description of the treatment of random error issues such as outlier rejection and the categorical exclusion of data by thresholds; for example, based on signal to noise or minimum ion counts.
- Proper estimates of uncertainty and the methods used for the error analysis. Quantification of many proteins or peptides generally results in the need to use some form of multiple hypothesis testing correction. Whenever possible, confidence in protein quantification should be provided for each individual protein rather than the global dataset. Any conclusions drawn or hypothesis generated from the quantitative data in the manuscript must be in concert with the determined estimate of uncertainty.
- If a component is not being identified by database searching in a particular experiment, assurance of the identity of the analyte being measured and the specificity with which it is measured must be provided. This particularly applies to intensity-based methods such as SELDI, SRM/MRM and AMT (accurate mass and time tag).
- A description of the way multiple isoforms in a protein group were quantified..



---

**DRAFT: Revised Publication Guidelines for Documenting the Identification of Peptides, Proteins, and Post-Translational Modifications by Mass Spectrometry**

- For spectral counting measurements, in addition to the above guidelines, additional details should be provided such as whether numbers of peptides or spectra were counted, whether modified peptides, semi-tryptic peptides or shared peptides were counted, and whether or not dynamic exclusion was used, etc.

## **VI. RAW DATA SUBMISSION**

If a manuscript is accepted by the journal, all mass spectra contributing to the described work must be deposited in electronic form by the time of publication at a publicly accessible site that is independent of the authors' control. Submission of all mass spectrometric output files in the original instrument vendor file format is the preferred and most direct means of meeting this requirement. Data conversion to an open format such as mzML or mzXML is encouraged when providing instrument vendor formatted files is impractical or software capable of reading the file format is not widely available. In all cases, the spectra are expected to be provided in a form prior to any processing that might affect the quality of subsequent interpretation as described in the peak list guideline (See section I). **Requests for exemptions (or delays) from this requirement must be made in writing to one of the co-editors at the time of submission.** Upon acceptance of their manuscript (and by the time of publication), authors should provide a URL and password, if appropriate, for accessing the data. This will be listed as part of the published article.

E.gE.g.: The data associated with this manuscript may be downloaded from the ProteomeCommons.org Tranche system: <<http://www.proteomecommons.org/data-downloader.jsp?fileName=90MaGKV4KHKHOyOvNGSXxtDhAEObJA3KbZap6ruHxvUF Dk%2BvOFyhawX%2BhSQa%2Bxa/KvG6oQCYON4nsZ/uDw55FfNDAU0AAAAAAAMLW==>>>.