

MIAPE: Mass Spectrometry Quantification

Salvador Martínez-Bartolomé[1,12], Eric W. Deutsch[2], Pierre-Alain Binz[3], Andrew R. Jones[4], Martin Eisenacher[5], Gerhard Mayer[5], Alex Campos[6,12], Francesc Canals[7,12], Joan-Josep Bech-Serra[7,12], Montserrat Carrascal[8,12], Marina Gay[8,12], Alberto Paradela[1,12], Rosana Navajas[1,12], Miguel Marcilla[1,12], María Luisa Hernández[9,12], María Dolores Gutiérrez-Blázquez[9,12], Luis Felipe Clemente Velarde[9,12], Kerman Aloria[10,12], Jabier Beaskoetxea[11,12], J. Alberto Medina-Aunon[1,12], Juan P. Albar[1,12]

[1] Proteomics Facility, Centro Nacional de Biotecnología - CSIC, Madrid, Spain.

[2] Institute for Systems Biology, 1441 N 34th Street, Seattle, WA 98103, USA.

[3] Swiss Institute of Bioinformatics, Rue Michel-Servet 1, CH-1211 Geneva 4, Switzerland.

[4] Institute of Integrative Biology, University of Liverpool, Liverpool, UK.

[5] Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany.

[6] Proteomics Platform, Barcelona Science Park, Barcelona, Spain.

[7] Proteomics Laboratory, Vall d'Hebron Institute of Oncology, Vall d'Hebron University Hospital, Barcelona, Spain.

[8] CSIC/UAB Proteomics Laboratory, Instituto de Investigaciones Biomédicas de Barcelona-Consejo Superior de Investigaciones Científicas/Institut d'investigacions Biomèdiques August Pi i Sunyer (IIBB-CSIC/IDIBAPS), Bellaterra, Spain.

[9] Proteomics Unit. Universidad Complutense de Madrid- Parque Científico de Madrid. Madrid, Spain.

[10] Proteomics Core Facility-SGIKER, University of the Basque Country, UPV/EHU, 48940 Leioa Spain.

[11] Department of Biochemistry and Molecular Biology, University of the Basque Country, UPV/EHU, 48940 Leioa Spain.

[12] ProteoRed Consortium, Spanish National Institute of Proteomics, Madrid, Spain.

Author contact information:

Salvador Martínez de Bartolomé Izquierdo

Proteomics Facility, Centro Nacional de Biotecnología - CSIC, Madrid, Spain.

ProteoRed Consortium, Spanish National Institute of Proteomics, Madrid, Spain.

smartinez@proteored.org

MIAPE: Mass Spectrometry Quantification

Version 1.0, 24th October, 2012, Final.

This module identifies the minimum information required to report the use of quantification techniques in a proteomics experiment, sufficient to support both the effective interpretation and assessment of the data and the potential recreation of the results of the data analysis.

Introduction

This document is one of a collection of technology-specific modules that together constitute the Minimum Information about a Proteomics Experiment (MIAPE) reporting guidelines produced by the Proteomics Standards Initiative. MIAPE is structured around a parent document that lays out the principles to which the individual reporting guidelines adhere. In brief, a MIAPE module represents the minimum information that should be reported about a data set or an experimental process, to allow a reader to interpret and critically evaluate the conclusions reached, and to support their experimental corroboration. In practice a MIAPE module comprises a checklist of information that should be provided (for example about the protocols employed) when a data set is submitted to a public repository or when an experimental step is reported in a scientific publication (for instance in the materials and methods section). The MIAPE modules specify neither the format that information should be transferred in, nor the structure of the repository/text. However, PSI is not developing the MIAPE modules in isolation; several compatible data exchange standards are now well established and supported by public databases and data processing software in proteomics (for details see the PSI website <http://www.psidev.info>).

Mass spectrometry is already a well-established protein identification tool and recent methodological and technological developments have also made possible the extraction of quantitative data of protein abundance in large-scale studies. Several strategies for absolute and relative quantitative proteomics and the statistical assessment of quantifications are possible, each having specific measurements and therefore, different data analysis workflows.

The guidelines for Mass Spectrometry Quantification allow the description of a wide range of quantitative approaches, including labelled and label-free techniques and also targeted approaches such as Selected Reaction Monitoring (SRM).

These reporting guidelines cover the description of experimental design (as far as important for quantification) and samples, input data, quantification algorithm and resulting quantitative and statistical data. They do not explicitly cover sample preparation, but do require some minimal description for each sample and the labelling protocol (if applicable). They do not include any acquisition parameter or transition list (in case of SRM approaches), since these should be included in the corresponding MIAPE Mass Spectrometry module.

Sections from earlier versions of other MIAPE modules dealing with quantification have been removed to avoid overlap (e.g. from MIAPE-MS and MIAPE-MSI). Reporting of quantification in gel experiments is still described in MIAPE-GI (gel informatics).

Items falling outside the scope of this module may be captured in complementary modules, which can be obtained from the website (<http://www.psidev.info/miape>). Note that subsequent versions of this document may have altered scope, as will almost certainly be the case for all the MIAPE modules.

The following section, detailing the reporting guidelines for the peptide and protein quantification analysis and their statistics, is subdivided as follows:

1. General features; the quantitative approach.
2. Experimental design and sample description; description of samples and assays (a run together with a labelling feature), group structure (resulting from experimental conditions and their values) and replicate structure.

3. Input data; description or reference to the dataset used for quantitative analysis.
4. Protocol; software and methods applied in the quantitative analysis (including transformation functions, aggregation functions, and statistical calculations).
5. Resulting data; actual resulting quantification values at the feature, peptide and/or protein level, plus units where appropriate.

Reporting guidelines for the peptide and protein quantification analysis.

1. General features

- 1.1. Experiment identifier or name
- 1.2. Responsible person or role
- 1.3. Quantitative approach

2. Experimental design and sample description

- 2.1. Experimental design
 - 2.1.1. Groups
 - 2.1.2. Biological and technical replicates
- 2.2. Sample / Assay description
 - 2.2.1. Labelling protocol (if applicable)
 - 2.2.2. Sample description:
 - 2.2.2.1. Sample name
 - 2.2.2.2. Sample amount
 - 2.2.2.3. Sample labelling with assay definition, i.e. MS run / data set together with reporting ion mass, reagent or isotope labelled amino acid
 - 2.2.2.4. Replicates and/or groups
 - 2.2.3. Isotopic correction coefficients
 - 2.2.4. Internal references

3. Input data

Description and reference of the dataset used for quantitative analysis (no actual values).

- 3.1. Input data type
- 3.2. Input data format
- 3.3. Input data merging
- 3.4. Availability of the input data

4. Protocol

Description of the software and methods applied in the quantitative analysis (including transformation functions, aggregation functions and statistical calculations).

- 4.1. Quantification software name, version and manufacturer

- 4.2. Description of the selection and/or matching method of features, together with the description of the method of the primary extracted quantification values determination for each feature and/or peptide.
- 4.3. Confidence filter of features or peptides prior to quantification
- 4.4. Description of data calculation and transformation methods
 - 4.4.1. Missing values imputation and outliers removal
 - 4.4.2. Quantification values calculation and / or ratio determination from the primary extracted quantification values
 - 4.4.3. Replicate aggregation
 - 4.4.4. Normalization
 - 4.4.5. Inference protocol for calculating protein quantification values from peptide quantification values
 - 4.4.6. Protocol specific corrections
- 4.5. Description of methods for (statistical) estimation of correctness
- 4.6. Calibration curves of standards

5. Resulting data

The actual quantification values resulting from your quantification software together with their estimated confidence

- 5.1. Quantification values at feature and/or at peptide level:
 - 5.1.1. Primary extracted quantification values for each feature, with their statistical estimation of correctness
 - 5.1.2. Quantification values for each peptide as a result of the aggregation of the values of the previous section (5.1.1), with their statistical estimation of correctness
- 5.2. Quantification values at protein level:
 - 5.2.1. Basic / raw quantification values with statistical estimation of correctness
 - 5.2.2. Transformed / aggregated / combined quantification values of the proteins at group level, with their statistical estimation of correctness

Summary

The MIAPE: Mass Spectrometry Quantification minimum reporting guidelines for the peptide and protein quantification analysis specify that a significant number of details have to be captured for quantitative analysis, such as label-free, ¹⁸O-labelling, AQUA, SILAC, iTRAQ, TMT, ICAT, ICPL, ¹⁴N/¹⁵N or SRM approaches. However, it is clear that providing the information required by this document will enable the effective interpretation and assessment of quantitative data and potentially, support experimental corroboration. Much of the information required herein may already be stored in an electronic format, or exportable from the analysis software; we anticipate further automation of this process.

These guidelines will evolve. To contribute, or to track the process to remain 'MIAPE-compliant', browse to the website at <http://psidev.info>

Appendix One. The MIAPE: Mass Spectrometry Quantification glossary of required items

<i>Classification</i>	<i>Definition</i>
1. General features — <i>Global descriptors of the experiment</i>	
1.1 Experiment identifier or name	Descriptive name and/or an identifier assigned to the experiment (may be assigned later by a public repository). Include keywords indicating which disease, model... was studied.
1.2 Responsible person or role	The (stable) primary contact person for this data set; this could be the experimenter, lab head, line manager etc. Where responsibility rests with an institutional role (<i>e.g., one of a number of duty officers</i>) rather than a person, give the official name of the role rather than any one person. In all cases give affiliation including e-mail address and stable contact information.
1.3 Quantitative approach	Indicate the quantitative approach deployed in the experiment (<i>e.g., iTRAQ-4plex, SILAC, Label-free identity-based, feature- based, SRM, AQUA, ¹⁴N/¹⁵N</i>).
2. Experimental design and sample description — <i>2.1 Experimental design (to reference in section 2.2.2.4)</i>	
2.1.1 Groups	Description of the experimental sample grouping including group names (if appropriate). A group is a sample set that corresponds to a certain experimental condition with its value (i.e. condition “gender” with value “female” is group “gender_female”).
2.1.2 Biological and technical replicates	Description of the biological and/or technical replicate structure (if appropriate). This is also a group structure like above, but hierarchically ordered. In the case of technical replicates, describe the kind of technical replicates: <i>e.g., cell lysis replicates: samples are replicated prior to cell lysis step, digestion replicates: samples are replicated prior to digestion step, LC-MS/MS replicates: samples are replicated prior to LC-MS/MS runs.</i>
2. Experimental design and sample description — <i>2.2 Sample / Assay description</i>	
2.2.1 Labelling protocol (if applicable)	In the case of label-based quantification methods, describe the labelling protocol, including the labelling level: if labelling occurs at the element, amino acid, peptide (terminus) or protein (terminus) level.
2.2.2 Sample description	For each sample, provide, if applicable, the information required in 2.2.2.1, 2.2.2.2, 2.2.2.3 and 2.2.2.4:
2.2.2.1 Sample name	Unique name for the sample to be referred to elsewhere in the document.

2.2.2.2 Sample amount	Sample amount (including the appropriate units).
2.2.2.3 Sample labelling with assay definition, i.e. MS run / data set together with reporting ion mass, reagent or isotope labelled amino acid	In the case of label-based quantification methods, indicate which label is associated with which sample. If appropriate, provide the mass of the reporter ion (e.g. <i>iTRAQ 4-plex : 114</i>), the reagent (e.g., <i>ICPL-¹²C₆</i>) or the isotope labeled amino acid(s) used in this sample (e.g., <i>¹³C₆-Lysine and ¹³C₆-Arginine</i>) or the isotope-labeled element (e.g. <i>¹⁵N</i>).
2.2.2.4 Replicates and/or groups	Replicates and/or groups from sections 2.1.1 and 2.1.2 that the sample belongs to. One sample / assay is usually part of several replicates / groups, e.g. part of technical replicate X AND biological replicate Y, or part of "gender_male" AND "age_50-60".
2.2.3 Isotopic correction coefficients	In case of iTRAQ/TMT like labelling, provide the isotopic correction coefficients. In the case of other isotopic labels, percent enrichment, such as 98% <i>¹⁵N</i> .
2.2.4 Internal references	List of internal references used and their amount. Also state their specific purpose such as quantification, normalization or alignment.
3. Input data — Description and reference of the dataset used for quantitative analysis: type, format and availability of the data. No actual values are requested here.	
3.1 Input data type	Description of the type of the source data used for the quantitative analysis: Zoom scan, full MS scan, tandem MS, SRM chromatogram, etc...
3.2 Input data format	Specify the input data format for the quantitation analysis, either post-processing formats such as <i>mzIdentML</i> , <i>protXML</i> , <i>pepXML</i> , <i>dat</i> , or if needed provide the original format such as <i>Thermo raw</i> , <i>AB Sciex wiff</i> , <i>mzData</i> , <i>mzML</i> , <i>mzXML</i> , <i>mgf</i> . For XML format data, provide the schema version number.
3.3 Input data merging	In case of fractionation has been applied to the samples prior to the acquisition of MS data, specify which type of fractionation has been performed (e.g. <i>chromatography-based</i> , <i>gel-based</i> , etc.) and how many <u>fractions/bands</u> have been generated for each sample. State whether input data coming from each fractionation unit is merged in a single dataset prior to quantification analysis or not. In that case, describe the software or algorithm used for this purpose (name, manufacturer and version) or refer to the one described below in section 4.1.
3.4 Availability of the input data	Specify the location of source data. Provide either a URI or the location of the files, or their availability.

4. Protocol – <i>Description of the software and methods applied in the quantitative analysis (including transformation functions, aggregation functions and statistical calculations).</i>	
4.1 Quantification software name, version and manufacturer	For each software (or library) used in the quantitative analysis, provide the trade name together with the version name or number according to the original code and functionality of the software. In case of in-house developed tools provide a brief description of it together with a reference if available. Brief description of the selected analysis pipeline if it is customizable from the software.
4.2 Description of the selection and/or matching method of features, together with the description of the method of the primary extracted quantification values determination for each feature and/or peptide	Description of the methods together with its relevant parameters used in the selection and/or matching of features: <i>e.g., feature detection (m/z and retention time windows, applied filters by charge state or by signal to noise, etc...), pairs finding, chromatogram alignment, the iTRAQ reporter integration method (most confident centroid, most intense centroid, centroid sum over 20ppm window, etc...),</i> or any other method performed to select and/or match features prior to their quantification. Also describe how the primary extracted quantification value for each feature and/or peptide is calculated (<i>e.g., volume measurement, max peak intensity, etc...</i>).
4.3 Confidence filter of features or peptides prior to quantification	Describe any filter threshold applied to the features (<i>e.g. a threshold filter over the “seed score” in OpenMS software</i>) or to the identification matches prior to quantification of features, such as <i>FDR, FWER, score, probability thresholds, etc...</i> State if applicable in which level has been applied (<i>PSM, peptide or protein level</i>). This includes the description of how the threshold has been determined (<i>e.g. arbitrary, decoy strategy, binomial distribution, etc...</i>). Also state whether quantitation restricted to unique (i.e. non-shared) peptides or whether some features/peptides have been excluded from the quantification (<i>e.g. containing a non-quantitative modification like oxidized methionine, etc...</i>). If the later filter is described in a MIAPE Mass Spectrometry Informatics document, provide a reference to section 1.4 in which the reference is.
4.4 Description of data calculation and transformation methods	Provide, if applicable, the information in 4.4.1 to 4.4.6. Also state at which level (feature, peptide and/or protein) the data calculation/transformation methods are performed and in which order (mathematical formulas are allowed).
4.4.1 Missing values imputation and outliers removal	Describe the missing values imputation and outlier removal methods. Describe (if appropriate) what does a '0' value means, that is, whether it means absent or just below the level of detection, etc...
4.4.2 Quantification values calculation and / or ratio determination from the primary extracted quantification values	Describe how quantification values and / or ratios have been calculated from values described in section 4.2, <i>e.g., aggregation of all features per peptide (average, geometric average, summing, ANOVA analysis), description of the enumerator and denominator of the ratio, etc...</i>
4.4.3 Replicate aggregation	Describe the calculation method aggregating / combining the values from peptide or experimental replicates and/or groups (<i>e.g., average, geometric average, weighting, etc. ...</i>), as well as any data transformation (<i>e.g., logarithmization, multiplication, division, etc...</i>).

4.4.4 Normalization	Description of any normalization procedure, e.g. ' <i>median peptide ratio for all measurements forced to 1</i> ' or ' <i>each sample spiked with equal amount of peptide X or protein Y, as per 2.2.4</i> '.
4.4.5 Protein quantification values calculation and / or ratio determination from the peptide quantification values	Describe how protein quantification values and / or ratios have been inferred from peptide quantification values (especially in case of shared peptides). Specify additional criteria like: " <i>take into account only proteins with at least two identified peptides</i> " used in order to filter PSMs before rolling up them into protein abundance.
4.4.6 Protocol specific corrections	Describe any protocol specific corrections like: " <i>Arg-Pro conversion</i> ".
4.5 Description of methods for (statistical) estimation of correctness	Describe how are calculated the values that describe the error potential of given / calculated quantification values (e.g., <i>standard deviation, standard error, confidence interval, super-median of the experiment, standard error for the fit, standard error for the calculated ratio, coefficient of variation, p-value, FDR, FWER, etc...</i>). Also describe any applied threshold or filter based in these statistical values (e.g., <i>p-values <= 0.05</i>). Also describe if any threshold value has been used to consider that a feature/peptide/protein is over-expressed, and why that value has been selected.
4.6 Calibration curves of standards	If internal references are used (described in section 2.2.4), provide the resulting calibration curves.
5. Resulting data — <i>Provide the actual quantification values resulting from your quantification software together with their estimated confidence. Depending of the quantification technique or even of the quantification software, only some of the following items could be satisfied (e.g., for spectral counting, only quantification values at protein level can be provided)</i>	
5.1 <i>Quantification values at peptide and/or feature level: Actual quantification values achieved for each peptide and/or, in case of feature-based quantification, for the corresponding features (mapped back from each peptide), together with their estimated confidence. Report each quantification value together with the appropriate information to identify the feature or peptide that is quantified (peptide sequence or peptide sequence plus the charge state, mass, retention time, etc...).</i>	
5.1.1 Primary extracted quantification values for each feature (e.g. area, height, etc.), with their statistical estimation of correctness	The calculated primary extracted quantification values for each feature and also values representing their confidence (see above 4.5). The values can be provided either by a URI or the location of the resulting data files.
5.1.2 Quantification values for each peptide as a result of the aggregation of the values of the previous section (5.1.1), with their statistical estimation of correctness	The transformed / aggregated / combined quantification values for each peptide, calculated from the quantification values from the previous section (5.1.1) (e.g. <i>medians, averages, standard deviations, super-medians, fold changes, p-values etc...</i>). Also provide the values representing their statistical estimation of confidence (see above 4.5). The values can be provided either by a URI or the location of the resulting data files, or their availability.
5.2 <i>Quantification values at protein level: Actual quantification values achieved for each protein and for each protein ambiguity group, together with the confidence in the quantification value. Report each quantification value together with the appropriate information to identify the protein, such as the protein accessions.</i>	

5.2.1 Basic / raw quantification values with statistical estimation of correctness	The calculated quantification value for each protein and for each protein ambiguity group and also values representing their confidence (see above 4.5). The values can be provided either by a URI or the location of the resulting data files.
5.2.2 Transformed / aggregated / combined quantification values of the proteins at group level, with their statistical estimation of correctness	The transformed / aggregated / combined quantification values for each protein, calculated from the quantification values above (<i>e.g. medians, averages, standard deviations, super-medians, fold changes, p-values etc...</i>), at group level. Also provide the values representing their statistical estimation of confidence (see above 4.5). The values can be provided either by a URI or the location of the resulting data files, or their availability.

Term glossary

Feature: Any primary extracted quantification measurement.

SRM: Selected Reaction Monitoring.

Replicates: experiment to be reproduced or repeated in order to test the variability of observed values.

Technical replicates: Replicates that share the same sample; i.e. the measurements are repeated. The technical variability is tested.

Biological replicates: Replicate where different samples are used for testing the biological variability between the selected samples.

Transformation functions: Any mathematical transformation of the data, such as logarithmizations, multiplications, divisions, etc...

Aggregation functions: Any mathematical operation applied to values from different replicates in order to obtain a single representative value. They can be averages, geometric averages, weighting, etc...

FDR: False Discovery Rate: the “expected” proportion of incorrect assignments among the accepted assignments.

FWER: Family-wise Error Rate: probability of making one or more false discoveries, or type I errors among all the hypotheses when performing multiple pairwise tests.

MIAPE Quant examples

MIAPE Quant examples files are available in the HUPO-psi web page at: http://psidev.info/sites/default/files/files/MIAPE_Quant_v1.0_examples.rar

The following table shows which quantification technique is represented by each example:

File name	Quantification technique	Institution where the example comes from
MIAPE_Quant_v1.0_ICPL_CNB.docx	ICPL	CNB
MIAPE_Quant_v1.0_ICPL_VH.doc	ICPL	HUVH
MIAPE_Quant_v1.0_iTRAQ_UCM-PCM.doc	iTRAQ	UCM-PCM
MIAPE_Quant_v1.0_iTRAQ_LPSCICUAB.doc	iTRAQ	LP-CSIC/UAB

MIAPE_Quant_v1.0_LabelFree_PCB.doc	Accurate Mass and Time Label-free	PCB
MIAPE_Quant_v1.0_SILAC_CNB.doc	SILAC	CNB
MIAPE_Quant_v1.0_SILAC_MLHS.doc	SILAC	UCM-PCM
MIAPE_Quant_v1.0_SILAC_VH.doc	SILAC	HUVH
MIAPE_Quant_v1.0_SRM_LPCSICUAB.doc	SRM	LP-CSIC/UAB
MIAPE_Quant_v1.0_spectralCount_MPC.xls	Label-free spectral counting	MPC

Institutions

CNB: National Center for Biotechnology, ProteoRed, Madrid, Spain

HUVH: Vall d'Hebron University Hospital, Barcelona, ProteoRed, Spain

UCM-PCM: Universidad Complutense de Madrid- Parque Científico de Madrid, Madrid, ProteoRed, Spain

LP-CSIC/UAB: Proteomics Laboratory, Instituto de Investigaciones Biomédicas de Barcelona - Consejo Superior de Investigaciones Científicas, Barcelona, ProteoRed, Spain

PCB: Proteomics Platform, Barcelona Science Park, Barcelona, ProteoRed, Spain

MPC: Medizinisches Proteom-Center, Ruhr-Universitaet Bochum, Bochum, Germany