Supplement A: MIACE: Minimum Information About a Complexome profiling Experiment

Suggested (meta-)data standards for the reporting of complexome profiling experiments

Version 1.0

Joeri van Strien[1], Hans-Peter Braun[2], Alfredo Cabrera-Orefice[1], Jyoti S. Choudhary[3], Felix Evers[4], Erika Fernandez-Vizarra[5], Sergio Guerrero-Castillo[6], Taco W.A. Kooij[4], Petra Páleníková[5], Mercedes Pardo[3], Uwe Schulte[7,8], Cristina Ugalde[9,10], Ilka Wittig[11], Lars Wöhlbrand[12], Ulrich Brandt[13], Susanne Arnold[13], Martijn A. Huynen[1]*

1. Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
2. Institute of Plant Genetics, Leibniz Universität Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany
3. Functional Proteomics, The Institute of Cancer Research, London SW7 3RP, UK
4. Medical Microbiology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
5. MRC Mitochondrial Biology Unit, University of Cambridge, UK
6. University Children's Research@Kinder-UKE, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany
7. Institute of Physiology, Faculty of Medicine, University of Freiburg, 79104 Freiburg, Germany
8. Center for Biological Signalling Studies (BIOSS) and Center for Integrative Signalling Studies (CIBSS), 79104 Freiburg, Germany
9. Hospital 12 de Octubre Research Institute, Madrid 28041, Spain
10. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), U723, Madrid, Spain
11. Functional Proteomics, Medical School, Goethe-University, 60590 Frankfurt am Main, Germany
12. General and Molecular Microbiology, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University of Oldenburg, Oldenburg, Germany
13. Radboud Center for Mitochondrial Medicine, Radboud University Medical Center, Nijmegen

* corresponding author

A proposal of (meta-)data standards for the reporting of complexome profiling experiments

introduction

Complexome profiling is an emerging 'omics approach that interrogates protein complex composition, assembly and dynamics [1,2,11–20,3,21,4–10]. This approach combines biochemical methods to separate protein complexes present in a biological sample into a series of fractions, with mass spectrometry to determine the protein content in each of the separated fractions, to then combine this information and create a protein distribution profile. The generated data inherently contains comprehensive information on a large collection of multiprotein assemblies present in the studied system. It should be noted that the published complexome profiling studies have generally

focused on a small subset of the protein complexes detected in the collected datasets. Thus, the potential for reusability of these data is very high, e.g. to uncover the comigration of proteins with each other in independent experiments. Nevertheless, a standard for its reporting, storing and annotation has not yet been defined, limiting the optimal use of this valuable and growing information resource.

The establishment of standards for reporting of a certain experiment through a minimum information document has proven to be very effective, greatly improving the availability and (re-)usability of the results of, for instance, microarray experiments, high-throughput sequencing, and various proteomics approaches [22–24]. Therefore, we propose the MIACE standard, the Minimum Information About a Complexome profiling Experiment. MIACE is created in order to facilitate efficient interpretation, sharing and reuse of complexome profiling data. It consists of the following seven components, each describing a set of core parameters required for the description and interpretation of a complexome profiling experiment:

1. General information
2. Sample specification
3. Protein separation method(s)
4. Sample processing and digestion
5. Mass spectrometry
6. Protein identification and quantification
7. Data processing and evaluation

Several aspects of these components are unique to a complexome profiling experiment, i.e. the separation of the native complexes into discrete fractions, and the calibration of apparent complex mass corresponding to these fractions with the help of known soluble and/or membrane bound reference complexes. Others (protein separation, mass spectrometry, protein identification/quantification and data processing/evaluation) are at least in part already addressed in modules of MIAPE, the Minimum Information About a Proteomics Experiment (www.psidev.info/miape)[22]. Where appropriate, we refer to these MIAPE components, such as gel electrophoresis, column chromatography, mass spectrometry, mass spectrometry informatics[25–28], for a description of the minimum information required. This document combines all information required for reporting a complete complexome profiling experiment.

1. General information

This section contains a general description of the experiment, which is the complete set of related complexome profiling samples that together address a certain scientific question. An experiment typically (but not necessarily) refers to the complete collection of complexome profiling data published in a single paper. It includes:

a) information about the author, affiliation(s), contact information.
b) short, single sentence title of the experiment.
c) longer general description of the experimental design and setup.
d) links to other related resources, reference to a publication if applicable.

e) relationship between all samples and associated data files as well as any information about relationships between samples. For example, which samples originate from the same batch or gel. This is best represented as a table, containing all samples, data and their relations.

2. Sample specification

This section contains a detailed description of all samples comprising the experiment. It should include:

a) the species origin(s) of the sample.
b) the strain, cell type and/or tissue type.
c) if applicable, the specific cell compartment or type of protein complex that has been isolated or for which the sample was enriched.
d) any experimental /parameters, which could for example be a treatment, a (disease) condition.
e) if applicable, any isotope label/mass tag or other chemical derivatization performed.
f) any additional information that is deemed relevant by the author/submitter, like for example sex, age or developmental stage.

3. Protein separation method(s)

This section should contain a description of the method used for separating protein complexes under native conditions prior to protein identification by mass spectrometry. This can be achieved by various methods, and new methods will likely be developed in the future. These methods include, but are not limited to, separation by gel electrophoresis, size exclusion chromatography, field-flow fractionation and sucrose density gradient ultracentrifugation. Detailed explanations on the information required for the description of gel electrophoresis and column chromatography are present in the respective MIAPE components (www.psidev.info/miape) [25,26]. In general, a description of the protein separation method should include the following:

a) a description of the specific method(s) of protein separation.
b) a description of or reference to the protocol(s) used.
c) the specific conditions or parameters under which the separation took place.
d) the number of fractions into which the protein complexes are separated, and if applicable, the molecular mass range that is covered by these fractions. This can be specified per sample if samples have varying numbers of fractions or mass ranges.
e) any other information deemed necessary to accurately and fully describe the method of protein separation.
f) if technically feasible, a visualization of the separation result, like an image of the resulting gel with visible bands in the case of native gel electrophoresis, or UV traces in the case of chromatography.

4. Sample processing and digestion

This section should contain a description of any additional processing steps the samples have undergone, along with a description of the digestion process to prepare samples for mass spectrometry. This section should include:

a) the protease(s) used for digestion.
b) description of or reference to the digestion protocol used.
c) description of or reference to protocols of any other processing steps or treatments that were performed on the samples prior to the mass spectrometry analysis.

5. Mass spectrometry

This section contains descriptions of the instruments and settings used for liquid chromatography (LC) coupled with mass spectrometry (MS). For a complete description of the information required to report a mass spectrometry experiment, refer to the column chromatography and mass spectrometry components of MIAPE (www.psidev.info/miape)[26,29]. The main aspects are briefly restated here:

a) chromatography column and LC instrument details.
b) MS instrument type and details.
c) any customizations or parameter settings of the LC-MS instruments.
d) description of the ion source, and the relevant parameters.
e) a description of the analyzer, the type and its relevant parameters.

6. Protein identification and quantification

This section contains the mass spectrometry output, the processing of primary (i.e. LC-MS and MS/MS raw) data, and peptide and protein assignment and abundance determination in each sample. For a complete description of the information required to report the output of an analyzed mass spectrometry experiment, refer to the mass spectrometry informatics component of MIAPE (www.psidev.info/miape)[28]. The main aspects are briefly restated here:

a) if feasible, (a link to) 'raw' mass spectrometry output data, to enable re-analysis with different tools/algorithms.
b) peptide and protein identification including the used software, search database and parameters, allowed modifications, mass calibration, m/z tolerance, scores and error estimates (FDR).
c) peptide-level abundance determination (software for feature extraction, m/z and retention time alignment and assignment to peptides).
d) protein abundance determination including a description of the processing.
e) if applicable, the final data after any additional processing steps.
f) the quantification metric in which protein abundances are expressed, for example iBAQ, LFQ or intensity.

To these parameters we add a proposed file format for the storage of protein abundances of a complexome profiling experiment. The protein abundance data should be made available in the 'CompTab' format, which is specified in the associated MIACE CompTab document. These data should provide the original protein abundances, before normalization or alignment, and represent the protein abundances from a single replicate. The table should include the protein length, the number of identified peptides and the peptide coverage of the protein. Additionally, the apparent protein complex mass range for each fraction should be indicated, if applicable.

7. Data processing and evaluation

This section contains the description of any additional data processing steps performed on the data after computation of protein abundances, such as the clustering method used to reconstruct protein complexes, as well as any normalization, alignment, mass calibration etc. Additionally, this section contains a description of the quality evaluation of the experimental results. This should include the following:

   a) tools and algorithms used for post-processing steps, with versions and all relevant settings and parameters.
   b) the reference proteins/complexes and regression used for the apparent molecular mass calibration.
   c) if available, any measures of the resolution of complex separation, as well as measures of quality of the protein identification and quantitation (false positive and negative assignment rates, dynamic range of abundance etc.).

Intended use

The MIACE standard is meant to be used as a minimal descriptor of a complexome profiling experiment. Authors are encouraged to supply additional information as they deem necessary for a full description of the reported experiment. Data archives should include the seven components of MIACE as a requirement, but allow for reporting of additional information. Use of controlled vocabulary in the annotation of experiments will further improve the potential and ease of data reuse. An additional goal of this document is to prompt the development of software and data archives compliant with MIACE, which will further facilitate adoption of this standard. Additionally, this minimum information could be required and checked by journals and reviewers when complexome profiling experiments are reported or published.

Conclusion

Community-wide adoption of this standard will greatly benefit research using and reusing complexome profiling data. It will make interpretation and verification of analysis results easier, and will enable large-scale meta-analysis and reuse of (published) complexome profiling data to better quantify the association between proteins into complexes.

References

[1]   H. Heide, L. Bleier, M. Steger, J. Ackermann, S. Dröse, B. Schwamb, M. Zörnig, A.S. Reichert, I. Koch, I. Wittig, Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex, Cell Metab. 16 (2012) 538–549.

[2]   H.J.C.T. Wessels, R.O. Vogel, R.N. Lightowlers, J.N. Spelbrink, R.J. Rodenburg, L.P. van den Heuvel, A.J. van Gool, J. Gloerich, J.A.M. Smeitink, L.G. Nijtmans, Analysis of 953 human proteins from a mitochondrial HEK293 fraction by complexome profiling, PLoS One. 8 (2013) e68340.

[3]   S. Guerrero-Castillo, F. Baertling, D. Kownatzki, H.J. Wessels, S. Arnold, U. Brandt, L. Nijtmans, The assembly pathway of mitochondrial respiratory chain complex I, Cell Metab. 25 (2017) 128–139.

[4]   J. Van Strien, S. Guerrero-Castillo, I.A. Chatzispyrou, R.H. Houtkooper, U. Brandt, M.A. Huynen, COmplexome Profiling ALignment (COPAL) reveals remodeling of mitochondrial protein complexes in Barth syndrome, Bioinformatics. 35 (2019). https://doi.org/10.1093/bioinformatics/btz025.

[5]   L. Wöhlbrand, H.S. Ruppersberg, C. Feenders, B. Blasius, H. Braun, R. Rabus, Analysis of membrane–protein complexes of the marine sulfate reducer Desulfobacula toluolica Tol2 by 1D blue native-PAGE complexome profiling and 2D blue native-/SDS-PAGE, Proteomics. 16 (2016) 973–988.

[6]   T. Wai, S. Saita, H. Nolte, S. Müller, T. König, R. Richter-Dennerlein, H. Sprenger, J. Madrenas, M. Mühlmeister, U. Brandt, The membrane scaffold SLP2 anchors a proteolytic hub in mitochondria containing PARL and the i-AAA protease YME1L, EMBO Rep. 17 (2016) 1844–1856.

[7]   C.S. Müller, W. Bildl, A. Haupt, L. Ellenrieder, T. Becker, C. Hunte, B. Fakler, U. Schulte, Cryo-slicing blue native-mass spectrometry (csBN-MS), a novel technology for high resolution complexome profiling, Mol. Cell. Proteomics. 15 (2016) 669–681.

[8]   S. Vidoni, M.E. Harbour, S. Guerrero-Castillo, A. Signes, S. Ding, I.M. Fearnley, R.W. Taylor, V. Tiranti, S. Arnold, E. Fernandez-Vizarra, MR-1S interacts with PET100 and PET117 in module-based assembly of human cytochrome c oxidase, Cell Rep. 18 (2017) 1727–1738.

[9]   C. Hillier, M. Pardo Calvo, L. Yu, E. Bushell, T. Sanderson, T. Metcalf, C. Herd, B. Anar, J.C. Rayner, O. Billker, Landscape of the Plasmodium Interactome, (2019).

[10]   H. Angerer, S. Schönborn, J. Gorka, U. Bahr, M. Karas, I. Wittig, J. Heidler, J. Hoffmann, N. Morgner, V. Zickermann, Acyl modification and binding of mitochondrial ACP to multiprotein complexes, Biochim. Biophys. Acta (BBA)-Molecular Cell Res. 1864 (2017) 1913–1920.

[11]   J. Senkler, M. Senkler, H. Eubel, T. Hildebrandt, C. Lengwenus, P. Schertl, M. Schwarzländer, S. Wagner, I. Wittig, H. Braun, The mitochondrial complexome of Arabidopsis thaliana, Plant J. 89 (2017) 1079–1092.

[12]   J. Senkler, N. Rugen, H. Eubel, J. Hegermann, H.-P. Braun, Absence of complex I implicates rearrangement of the respiratory chain in European mistletoe, Curr. Biol. 28 (2018) 1606–1613.

[13]   N. Rugen, H. Straube, L.E. Franken, H.-P. Braun, H. Eubel, Complexome profiling reveals association of PPR proteins with ribosomes in the mitochondria of plants, Mol. Cell. Proteomics. (2019) mcp-RA119.

[14]   P.C. Havugimana, G.T. Hart, T. Nepusz, H. Yang, A.L. Turinsky, Z. Li, P.I. Wang, D.R. Boutz, V. Fong, S. Phanse, A census of human soluble protein complexes, Cell. 150 (2012) 1068–1081.

[15]   M. Larance, K.J. Kirkwood, M. Tinti, A.B. Murillo, M.A.J. Ferguson, A.I. Lamond, Global membrane protein interactome analysis using in vivo crosslinking and mass spectrometry-based protein correlation profiling, Mol. Cell. Proteomics. 15 (2016) 2476–2490.

[16]   A.R. Kristensen, J. Gsponer, L.J. Foster, A high-throughput approach for measuring temporal changes in the interactome, Nat. Methods. 9 (2012) 907.

[17]   P.L. Kastritis, F.J. O'Reilly, T. Bock, Y. Li, M.Z. Rogon, K. Buczak, N. Romanov, M.J. Betts, K.H. Bui, W.J. Hagen, Capturing protein communities by structural proteomics in a thermophilic eukaryote, Mol. Syst. Biol. 13 (2017).

[18]   L. Van Haute, A.G. Hendrick, A.R. D'Souza, C.A. Powell, P. Rebelo-Guiomar, M.E. Harbour, S. Ding, I.M. Fearnley, B. Andrews, M. Minczuk, METTL15 introduces N4-methylcytidine into human mitochondrial 12S rRNA and is required for mitoribosome biogenesis, Nucleic Acids Res. 47 (2019) 10267–10281.

[19]   T. Lobo-Jarne, R. Pérez-Pérez, F. Fontanesi, A. Timón-Gómez, I. Wittig, A. Peñas, P. Serrano-Lorenzo, I. García-Consuegra, J. Arenas, M.A. Martín, Multiple pathways coordinate assembly of human mitochondrial complex IV and stabilization of respiratory supercomplexes, EMBO J. (2020) e103912.

[20]   C.L. Alston, J. Heidler, M.G. Dibley, L.S. Kremer, L.S. Taylor, C. Fratter, C.E. French, R.I.C. Glasgow, R.G. Feichtinger, I. Delon, Bi-allelic Mutations in NDUFA6 Establish Its Role in Early-Onset Isolated Mitochondrial Complex I Deficiency, Am. J. Hum. Genet. 103 (2018) 592–601.

[21]   C.L. Alston, M.T. Veling, J. Heidler, L.S. Taylor, J.T. Alaimo, A.Y. Sung, L. He, S. Hopton, A. Broomfield, J. Pavaine, Pathogenic Bi-allelic Mutations in NDUFAF8 Cause Leigh Syndrome with an Isolated Complex I Deficiency, Am. J. Hum. Genet. 106 (2020) 92–101.

[22]   C.F. Taylor, N.W. Paton, K.S. Lilley, P.-A. Binz, R.K. Julian Jr, A.R. Jones, W. Zhu, R. Apweiler, R. Aebersold, E.W. Deutsch, The minimum information about a proteomics experiment (MIAPE), Nat. Biotechnol. 25 (2007) 887.

[23]   A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, Nat. Genet. 29 (2001) 365.

[24]   A. Brazma, Minimum information about a microarray experiment (MIAME)–successes, failures, challenges, Sci. World J. 9 (2009) 420–423.

[25]   F. Gibson, L. Anderson, G. Babnigg, M. Baker, M. Berth, P.-A. Binz, A. Borthwick, P. Cash, B.W. Day, D.B. Friedman, Guidelines for reporting the use of gel electrophoresis in proteomics, Nat. Biotechnol. 26 (2008) 863.

[26]   A.R. Jones, K. Carrol, D. Knight, K. MacLellan, P.J. Domann, C. Legido-Quigley, L. Huang, L. Smallshaw, H. Mirzaei, J. Shofstahl, Guidelines for reporting the use of column chromatography in proteomics, Nat. Biotechnol. 28 (2010) 654–655.

[27]   S. Martínez-Bartolomé, E.W. Deutsch, P.-A. Binz, A.R. Jones, M. Eisenacher, G. Mayer, A. Campos, F. Canals, J.-J. Bech-Serra, M. Carrascal, Guidelines for reporting quantitative mass spectrometry based experiments in proteomics, J. Proteomics. 95 (2013) 84–88.

[28]   P.-A. Binz, R. Barkovich, R.C. Beavis, D. Creasy, D.M. Horn, R.K. Julian Jr, S.L. Seymour, C.F. Taylor, Y. Vandenbrouck, Guidelines for reporting the use of mass spectrometry informatics in proteomics, Nat. Biotechnol. 26 (2008) 862–863.

[29]   C.F. Taylor, P.-A. Binz, R. Aebersold, M. Affolter, R. Barkovich, E.W. Deutsch, D.M. Horn, A. Hühmer, M. Kussmann, K. Lilley, Guidelines for reporting the use of mass spectrometry in proteomics, Nat. Biotechnol. 26 (2008) 860–861.

The CompTab file format specification

We propose CompTab (COMPlexome profiling TABular data format), a data format for the storage of protein abundance measurements of complexome profiling data. It is meant to be used alongside MIACE, with the goal to standardize the format in which complexome profiling data is shared and stored. It aims to be a clear and simple, yet explicit format that can be used and created by researchers and software alike. The proposed data format stores protein abundance values, as well as the following information about the identified proteins: protein length, number of unique peptides assigned to the protein and peptide coverage of the protein.
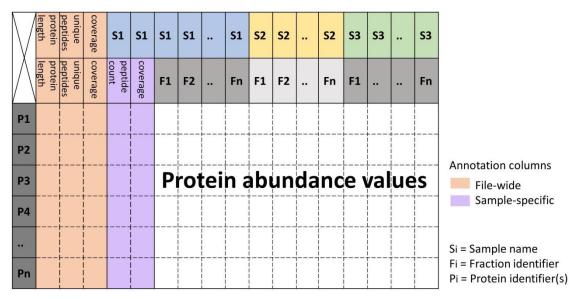

Figure 1. schematic representation of the CompTab format data structure.

File structure

The data is stored as a plain text file, with tab-separated values. Two header rows at the top specify the sample as well as the fraction identifiers. The first column contains protein identifiers. Each row can contain a single identifier, or in the case of protein groups, multiple comma-separated protein identifiers. If the sample fractions separate the protein complexes on their mass, the fractions are to be ordered so that lowest mass is in the leftmost column, and the heaviest fraction is in the rightmost. The columns with protein abundance profiles should contain numeric values only.

Identification details

In addition to the columns containing protein abundances, three types of additional columns are required, containing information about the identified proteins. Any of these annotation columns can refer to the entire set of samples in the comptab file ("File-wide"). In this case each column should occur once in the file. Alternatively, in case there is sample-specific information about the identification of proteins, these annotation columns can be repeated for each sample ("Sample-specific"). If the top header row of these columns specifies a sample, they are associated with this specific sample. If both header rows contain the column name (i.e. 'coverage', 'unique peptides' etc.) these columns refer to all samples in the CompTab file. The position or order of these columns in the file does not matter, as long as the headers are correct. The 'coverage' column provides the percent of the protein sequence that is covered by peptides. Values in the coverage column should be

numeric only, without characters like ' %'. The 'unique peptides' column contains the number of different peptides assigned to the protein using whole numbers. The 'protein length' column contains the amino acid length of the protein sequence, which should be presented as integers.