

MaConDa: a publicly accessible mass spectrometry contaminants database

Ralf J. M. Weber^{1,2}, Eva Li¹, Jonathan Bruty¹, Shan He³ and Mark R. Viant^{1,2,*}

¹School of Biosciences, ²Center for Systems Biology and ³School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK

Associate Editor: Jonathan Wren

ABSTRACT

Mass spectrometry is widely used in bioanalysis, including the fields of metabolomics and proteomics, to simultaneously measure large numbers of molecules in complex biological samples. Contaminants routinely occur within these samples, for example, originating from the solvents or plasticware. Identification of these contaminants is crucial to enable their removal before data analysis, in particular to maintain the validity of conclusions drawn from uni- and multivariate statistical analyses. Although efforts have been made to report contaminants within mass spectra, this information is fragmented and its accessibility is relatively limited. In response to the needs of the bioanalytical community, here we report the creation of an extensive manually well-annotated database of currently known small molecule contaminants.

Availability: The Mass spectrometry Contaminants Database (MaConDa) is freely available and accessible through all major browsers or by using the MaConDa web service <http://www.maconda.bham.ac.uk>.

Contact: m.viant@bham.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on May 4, 2012; revised and accepted on August 21, 2012

1 INTRODUCTION

Our understanding of biological systems has considerably improved through recent developments in mass spectrometry (MS)-based metabolomics (Dettmer *et al.*, 2007; Patti *et al.*, 2012). Continuous efforts have been made to improve the quality of metabolome measurements, including in sample preparation (Villas-Boas *et al.*, 2005), data collection (Dunn *et al.*, 2011; Weber *et al.*, 2011) and data analysis (Dunn *et al.*, 2012; Weber *et al.*, 2011). Nonetheless, sample preparation methods and MS analyses have the potential to introduce contaminants, such as plasticizers, additives and solvents (Keller *et al.*, 2008). Such contaminants of laboratory origin can obscure or even falsify biological interpretation of the data. For example, when using univariate or multivariate statistical analyses for biomarker discovery, the conclusions of that study can be fundamentally flawed if signals remain unidentified and are later discovered to be exogenous chemicals. Several analytical methods have been reported to minimize the interference caused by MS contaminants. Despite these improvements, contaminants are still a major

problem in MS experiments. Improved methods to identify and then treat contaminants appropriately are required urgently. Although in most cases identified contaminants should be eliminated from datasets, occasionally they can be beneficial, for example, when used for internal mass calibration of spectra (Scheltema *et al.*, 2008).

Here we present a manually well-annotated database of currently known MS contaminants to assist both the metabolomics and bioanalytical chemistry communities in their data processing.

2 METHODS AND IMPLEMENTATION

The information contained in Mass spectrometry Contaminants Database (MaConDa) is based on published literature (e.g. scientific papers, notes and hand books; see Supplementary Material) and data provided by several colleagues and instrument manufacturers, including the extensive resource by Keller *et al.* (2008). The raw data have been manually searched and curated, and parsed to a relational MySQL 5.0 database using Python 2.7.2 scripts. Theoretical and/or experimental data were stored for each MS contaminant, such as name, type of contaminant (e.g. plasticizer, detergent or buffer; see Table 1), empirical formula and details regarding the MS platform used (e.g. ion trap, triple quadrupole or time-of-flight MS). Theoretical mass values calculated to six decimal places and observed mass values for related ion forms (e.g. $[M+H]^+$, $[M+Na]^+$ and $[M-H]^-$) were also recorded, along with the original reference. Finally, cross references (e.g. PubChem Compound Identifier and Standard InChI code) were added for all MS contaminants when available. A user-friendly web interface to search the database has been implemented in PHP 5.3.3 JavaScript and deployed on an Apache 2.2 web server. The NuSOAP PHP library (<http://sourceforge.net/projects/nussoap/>) has been used to create a Web service, trivializing the integration of MaConDa into existing MS data analysis workflows. The Web service has been tested using a Simple Object Access Protocol (SOAP) Python client (<https://fedorahosted.org/suds/>) and Taverna (Hull *et al.*, 2006).

3 RESULTS

MaConDa contains more than 200 contaminant records detected across several MS platforms. The majority of records include theoretical as well as experimental MS data. In a few cases, experimental data were included without rigorous identification

*To whom correspondence should be addressed.

Table 1. Examples of contaminants within MaConDa and that are commonly observed in mass spectra

ID	Exact mass	Name	Formula	Type
CON00019	278.15183	Dibutyl phthalate	C ₁₆ H ₂₂ O ₄	Plasticizer
CON00053	281.27185	Oleamide	C ₁₈ H ₃₅ NO	Slip agent
CON00103	82.00308	Sodium acetate	C ₂ H ₃ O ₂ Na	Solvent
CON00121	121.07389	TRIS	C ₄ H ₁₁ NO ₃	Buffer
CON00298	189.04259	4-HCCA	C ₁₀ H ₇ NO ₃	Matrix compound

(Sumner *et al.*, 2007). The majority of experimental data reported in the literature has been collected in positive ion mode, which is reflected in the database. Also, the amount of MS/MS data for contaminants is currently rather limited. However, the database has the capability to store this type of data as more is recorded by the community. As such, and to the best of our knowledge, this is the first publicly accessible, readily searchable, readily implementable into an automated computational pipeline, readily expandable database of mass spectral contaminants.

A summary of the MaConDa features:

- Database access via SOAP web service;
- Database access via a user-friendly browser Web interface;
- Batch processing of peak lists;
- Searching of contaminants using additional ion forms;
- Exporting results into different formats (e.g. tab-delimited and CSV);
- Multiple database identifiers (e.g. PubChem Compound Identifier and Standard InChI code) for each contaminant to allow cross-referencing with other resources or databases;
- The total database is freely available in several formats (e.g. tab-delimited, CSV, XML and SQL format).

4 CONCLUSIONS

MaConDa is an extensive manually well-annotated database that provides a useful and unique resource for the MS community. Analytical techniques used in metabolomics and proteomics

are continually enhanced to improve their sensitivity. As a result, new contaminants are introduced into the experimental pipeline. Continued input of these new contaminants from the MS community and our own laboratory will enhance MaConDa as a valuable resource.

ACKNOWLEDGEMENTS

We gratefully thank our colleagues (David Watson, University of Strathclyde; John Draper, Aberystwyth University; John Langley, University of Southampton; John Newman, University of California Davis; Warwick Dunn, University of Manchester; William Griffiths, Swansea University) and instrument manufacturers (Thermo Fisher Scientific and Bruker Daltonics) who provided us with MS contaminant data. We thank Cheng Cao for his contribution to the website.

Funding: We thank both the British Heart Foundation (PG/10/036/28341) and UK Engineering and Physical Sciences Research Council (EP/J501414/1) for support, as well as the University of Birmingham's Systems Science for Health initiative.

Conflict of Interest: none declared.

REFERENCES

- Dettmer, K. *et al.* (2007) Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.*, **26**, 51–78.
- Dunn, W.B. *et al.* (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.*, **6**, 1060–1083.
- Dunn, W.B. *et al.* (2012) Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, DOI: 10.1007/s11306-012-0434-4.
- Hull, D. *et al.* (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.
- Keller, B.O. *et al.* (2008) Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta*, **627**, 71–81.
- Patti, G.J. *et al.* (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.*, **13**, 263–269.
- Scheltens, R.A. *et al.* (2008) Increasing the mass accuracy of high-resolution LC-MS data using background ions—a case study on the LTQ-Orbitrap. *Proteomics*, **8**, 4647–4656.
- Sumner, L. *et al.* (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics*, **3**, 211–221.
- Villas-Boas, S.G. *et al.* (2005) Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast*, **22**, 1155–1169.
- Weber, R.J.M. *et al.* (2011) Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification. *Anal. Chem.*, **83**, 3737–3743.