MSnbase, efficient R-based access and manipulation of raw mass spectrometry data

Laurent Gatto,[*,†] Sebastian Gibb,[‡] and Johannes Rainer[¶]

†*de Duve Institute, Université catholique de Louvain, Brussels, Belgium*

‡*Department of Anaesthesiology and Intensive Care of the University Medicine Greifswald, Germany*

¶*Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy*

E-mail: laurent.gatto@uclouvain.be

Abstract

We present version 2 of the `MSnbase` R/Bioconductor package. `MSnbase` provides infrastructure for the manipulation, processing and visualisation of mass spectrometry data. Here we present how the new *on disk* infrastructure allows the handling of hundreds on commodity hardware and present some application of the package.

Introduction

What is `MSnbase`: `MSnbase`[1] offers a solution between low level liraries such as jmzml, exclusively for the development of new applications and GUI software, that provides a limited and fixed set of functionality.

Software functionality

On disk backend

Efficient low level access: in memory vs on disk mode (using mzR), benchmarking, used in ms-based proteomics and metabolomics.

Use cases

Example applications: visualisation, ...

Discussion

To address (from guidelines):

- potential for reuse: see[2–4] for examples.
- general limitations
- system limitations
- end-user documentation
- developer documentation
- sample data
- benchmark data set
- availability
- license information
- system requirements

Collaborative development, 11 contributors since creation (see blog post).

Count packages depending on `MSnbase`.

Future developments.

The version of `MSnbase` used in this manuscritp is version 2.9.4. The main features presented here were available since version 2.0.

Acknowledgement

The authors thank the various contributors and users who have provided constructuve input and feedback that have helped, over the years, the improvement of the package. The authors declare no conflict of interest.

References

(1) Gatto, L.; Lilley, K. S. MSnbase - an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **2012**, *28*, 288–9.

(2) Wieczorek, S.; Combes, F.; Lazar, C.; Giai Gianetto, Q.; Gatto, L.; Dorffer, A.; Hesse, A. M.; CoutÃľ, Y.; Ferro, M.; Bruley, C.; Burger, T. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **2017**, *33*, 135–136.

(3) Griss, J.; Vinterhalter, G.; Schwämmle, V. IsoProt: A Complete and Reproducible Workflow To Analyze iTRAQ/TMT Experiments. *J Proteome Res* **2019**, *18*, 1751–1759.

(4) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006**, *78*, 779–87.