The Journal of Open Source Software

RIAssigner: A package for gas chromatographic retention index calculation

Helge Hecht [1¶], **Maksym Skoryk** [1,2], **Martin Čech** [1], **and Elliott James Price** [1]

1 RECETOX, Faculty of Science, Masaryk University, Kotlářská 2, Brno, Czech Republic **2** Institute of Computer Science, Masaryk University, Brno, Czech Republic ¶ Corresponding author

DOI: 10.21105/joss.04337

Software

- Review
- Repository
- Archive

Editor: David Hagan

Reviewers:

- @yguitton
- @davidegraff

Submitted: 24 February 2022
Published: 24 July 2022

Summary

RIAssigner is a software package for the computation of gas chromatographic (GC) retention indices (RI). The package uses matchms (Huber et al., 2020) and pandas (The pandas development team, 2020) for data I/O and supports the `.msp` as well as tabular (`.csv` & `.tsv`) formats, among others. It supports multiple keywords identifying the retention time (RT) and RI information and handling SI units for RT. The RI can be computed using non-isothermal Kováts retention-indexing (from temperature programming, using the definition of van Den Dool & Kratz (1963)) or cubic spline interpolation (Halang et al., 1978) based on a reference dataset containing RT & RI. The MIT-licensed package is hosted via bioconda (Grüning et al., 2018) and is also accessible to users as a Galaxy tool (Jalili et al., 2020; Spectrometric Data Processing and Analysis & Institute of Computer Science, 2022).

Statement of need

Compounds can be characterized by their retention behavior or elution time from a chromatographic column, under specified conditions. Analyte retention behavior is a function of physicochemical properties and elution time varies with chromatographic conditions. In gas chromatography, the retention index of n-alkanes is solely dependent on number of carbon atoms (similarly, the retention indices for other homologous series can depend solely on number of functional groups) and so unlike retention time, retention index provides a direct relationship to chemical structure (Peng, 2010).

Therefore, RI is only subject to very small deviations when using a column with similar separation properties. This allows comparison of data coming from samples analyzed under different analytical conditions e.g., columns of different length or different temperature gradient. An example use case is illustrated in Figure 1. It can therefore be used to improve identification of unknown target compounds when employed alongside spectral similarity in spectral library matching based identification of unknowns (Halket et al., 1999; Strehmel et al., 2008).

To leverage the RI in open-source identification workflows, a package providing computation methods as well as data handling is crucial.

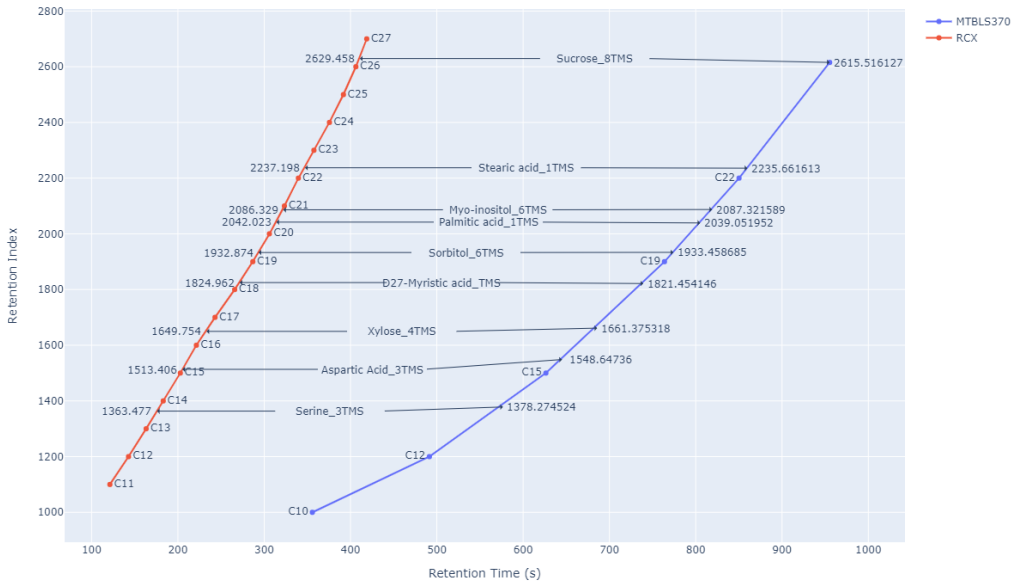

Figure 1: Example mapping of RI between two experiments with differing chromatographic setup. The markers denote the positions of reference compounds while the arrows indicate the RT and RI values of chemical compounds measured as standards via Price et al. (2021) and identified in the study conducted by Weidt et al. (2016). The RI for Sucrose has been extrapolated from the reference compounds.

State of the field

RI computation is contained in the most widely used GUI applications such as MS-DIAL (Tsugawa et al., 2015) and MZmine2 (Pluskal et al., 2010), the Galaxy tool metaMS (Wehrens et al., 2014) and the Python package CoreMS (Corilo et al., 2021). However, there is no standalone package which provides support for various computation methods based on homologous series (e.g., alkanes (Kováts, 1958), fatty acid methyl esters (FAMEs) (Kind et al., 2009)) or the virtual carbon number (Harangi, 2003). Additionally, existing tools expect input data in a fixed format and only perform RI computation as an intermediate step without reporting the actual values with the output.

To address aforementioned issues, we developed RIAssigner: a lightweight Python package that supports multiple computation methods and data formats and is built on an expandable architecture, consequently closing the gap towards modular annotation workflows. It can be integrated into file-based workflows by supporting various open standards or linked directly via its API into more complex Python applications.

Authors' Contributions

HH wrote the manuscript and developed the software. MS contributed to the software. MČ contributed via code reviews and implementation guidance. EJP provided conceptual oversight and contributed to the manuscript.

Acknowledgements

Authors thank to Research Infrastructure RECETOX RI (No LM2018121) financed by the Ministry of Education, Youth and Sports, and OP RDE project CETOCOEN EXCELLENCE (No CZ.02.1.01/0.0/0.0/17_043/0009632) for supportive background. EJP was supported from OP RDE - Project "MSCAfellow4@MUNI" (No. CZ.02.2.69/0.0/0.0/20_079/0017045). This project was supported from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857560. This publication reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains

References

Corilo, Y. E., Kew, W. R., & McCue, L. A. (2021). *EMSL-computing/CoreMS: CoreMS 1.0.0* (Version v1.0.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.4641553

Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., Valieris, R., & Köster, J. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nature Methods*, *15*(7), 475–476. https://doi.org/10.1038/s41592-018-0046-7

Halang, W. A., Langlais, R., & Kugler, E. (1978). Cubic Spline Interpolation for the Calculation of Retention Indices in Temperature-Programmed Gas-Liquid Chromatography. *Analytical Chemistry*, *50*(13), 1829–1832. https://doi.org/10.1021/ac50035a026

Halket, J. M., Przyborowska, A., Stein, S. E., Mallard, W. G., Down, S., & Chalmers, R. A. (1999). Deconvolution gas chromatography/mass spectrometry of urinary organic acids - potential for pattern recognition and automated identification of metabolic disorders. *Rapid Communications in Mass Spectrometry*, *13*(4), 279–284. https://doi.org/10.1002/(SICI)1097-0231(19990228)13:4%3C279::AID-RCM478%3E3.0.CO;2-I

Harangi, J. (2003). Retention index calculation without n-alkanes—the virtual carbon number. *Journal of Chromatography A*, *993*(1-2), 187–195. https://doi.org/10.1016/S0021-9673(03)00320-0

Huber, F., Verhoeven, S., Meijer, C., Spreeuw, H., Castilla, E., Geng, C., Hooft, J. van der, Rogers, S., Belloum, A., Diblen, F., & Spaaks, J. (2020). Matchms - processing and similarity evaluation of mass spectrometry data. *Journal of Open Source Software*, *5*(52), 2411. https://doi.org/10.21105/joss.02411

Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J., & Nekrutenko, A. (2020). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, *48*(W1), W395–W402. https://doi.org/10.1093/nar/gkaa434

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., & Fiehn, O. (2009). FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Analytical Chemistry*, *81*(24), 10038–10048. https://doi.org/10.1021/ac9019522

Kováts, E. (1958). Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica Acta*, *41*(7), 1915–1932. https://doi.org/10.1002/hlca.19580410703

Peng, C. T. T. (2010). Prediction of retention indices. VI: Isothermal and temperature-programmed retention indices, methylene value, functionality constant, electronic and steric effects. *Journal of Chromatography A*, *1217*(23), 3683–3694. https://doi.org/10.1016/j.chroma.2010.02.005

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*(1), 395. https://doi.org/10.1186/1471-2105-11-395

Price, E. J., Coufalíková, K., Jbebli, A., Palát, J., Koudelka, Š., & Klánová, J. (2021). *RECETOX metabolome HR-[EI+]-MS library* (Version 1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5483565

Spectrometric Data Processing and Analysis, & Institute of Computer Science. (2022). *RECETOX/galaxytools: Release v0.2.0* (Version v0.2.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6035335

Strehmel, N., Hummel, J., Erban, A., Strassburg, K., & Kopka, J. (2008). Retention index thresholds for compound matching in GC–MS metabolite profiling. *Journal of Chromatography B*, *871*(2), 182–190. https://doi.org/10.1016/j.jchromb.2008.04.042

The pandas development team. (2020). *Pandas-dev/pandas: pandas* (latest) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3509134

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., & Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, *12*(6), 523–526. https://doi.org/10.1038/nmeth.3393

van Den Dool, H., & Kratz, P. Dec. (1963). A generalization of the retention index system including linear temperature programmed gas—liquid partition chromatography. *Journal of Chromatography A*, *11*(3), 463–471. https://doi.org/10.1016/S0021-9673(01)80947-X

Wehrens, R., Weingart, G., & Mattivi, F. (2014). metaMS: An open-source pipeline for GC–MS-based untargeted metabolomics. *Journal of Chromatography B*, *966*, 109–116. https://doi.org/10.1016/j.jchromb.2014.02.051

Weidt, S., Haggarty, J., Kean, R., Cojocariu, C. I., Silcock, P. J., Rajendran, R., Ramage, G., & Burgess, K. E. V. (2016). A novel targeted/untargeted GC-Orbitrap metabolomics methodology applied to Candida albicans and Staphylococcus aureus biofilms. *Metabolomics*, *12*(12), 189. https://doi.org/10.1007/s11306-016-1134-2