




Data and text mining

TIMSCONVERT: a workflow to convert trapped ion mobility data to open data formats

Gordon T. Luu ¹, Michael A. Freitas², Itzel Lizama-Chamu¹,
Catherine S. McCaughey¹, Laura M. Sanchez ¹ and Mingxun Wang ^{3,*}

¹Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, CA, 95064, USA, ²Department of Cancer Biology and Genetics, Ohio State University, Columbus, OH, 43210, USA and ³Department of Computer Science and Engineering, University of California Riverside, Riverside, CA 92521, USA

*To whom correspondence should be addressed.

Associate Editor: Olga Vitek

Received on December 17, 2021; revised on May 10, 2022; editorial decision on June 18, 2022; accepted on June 26, 2022

Abstract

Motivation: Advances in mass spectrometry have led to the development of mass spectrometers with ion mobility spectrometry capabilities and dual-source instrumentation; however, the current software ecosystem lacks interoperability with downstream data analysis using open-source software and pipelines.

Results: Here, we present TIMSCONVERT, a data conversion high-throughput workflow from timsTOF Pro/fleX mass spectrometer raw data files to mzML and imzML formats that incorporates ion mobility data while maintaining compatibility with data analysis tools. We showcase several examples using data acquired across different experiments and acquisition modalities on the timsTOF fleX MS.

Availability and implementation: TIMSCONVERT and its documentation can be found at <https://github.com/gtluu/timsconvert> and is available as a standalone command-line interface tool for Windows and Linux, NextFlow workflow and online in the Global Natural Products Social (GNPS) platform.

Contact: mingxun.wang@cs.ucr.edu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1. Introduction

In recent years, ion mobility spectrometry (IMS) has been integrated into instruments configured for liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Silveira *et al.*, 2017). Examples include the timsTOF Pro (Bruker Daltonics) featuring an electrospray ionization (ESI) source and timsTOF fleX, which is a dual source instrument configured with ESI and matrix-assisted laser desorption/ionization (MALDI) sources coupled to a trapped ion mobility spectrometer (TIMS) and hybrid quadrupole-time-of-flight (qTOF) mass analyzer. IMS allows for the separation of ions based on their mobility in a carrier buffer gas; a common application includes the separation of isobars (compounds with the same nominal mass but different chemical formula) and isomers (compounds with the same chemical formula but different 3D configuration) in biological samples. The inclusion of IMS online separation increases the dimensionality of the data produced during acquisition, as previously discussed (Willems *et al.*, 2021). Therefore, the introduction of more advanced instrumentation has also been accompanied by new data formats.

The dual-source capabilities of the timsTOF fleX allow for a wide range of new acquisition modes including the following: (i) LC-TIMS-MS/MS, (ii) MALDI-TIMS-qTOF dried droplet (DD) and (iii) MALDI-TIMS-qTOF mass spectrometry imaging (MSI). The flexibility

of the timsTOF fleX has resulted in many different data formats ([Supplementary material](#) NOTE Bruker timsTOF Data Formats). While there are scattered solutions to convert (i) ([Supplementary Table S1](#)) in a low throughput fashion—current software solutions critically are incompatible with widely used data analysis pipelines already in existence [e.g. Global Natural Products Social (GNPS) molecular networking] (Wang *et al.*, 2016). Furthermore, the field lacks open-source software to convert MALDI DD and MSI data from the timsTOF fleX to mzML (ii) and imzML (iii), respectively. In mirroring the success of open software for other instruments, e.g. Hulstaert *et al.* (2020) in ThermoRawFileParser, we have developed TIMSCONVERT which aims for high-throughput conversion of all acquisition modes to their respective open formats on the Bruker timsTOF fleX MS. Importantly, to our knowledge, this is the first open-source workflow capable of converting MALDI data originating from the timsTOF ([Supplementary Table S1](#)). TIMSCONVERT aims to meet the needs of Bruker timsTOF users and ensure data compatibility with downstream analyses.

2. Implementation

TIMSCONVERT has been designed to take any raw data format (i.e. BAF, TSF and TDF) from the timsTOF Pro and fleX series of

instruments. The acquisition mode is detected automatically without the need for user input, and data are converted into an appropriate open data format through the use of proprietary and open-source libraries. Below, we describe the conversion process for each acquisition mode.

2.1 Conversion of LC-MS/MS data

To convert LC-MS/MS data, we use the Baf2Sql (Bruker Daltonics) library to parse data from BAF files to the widely used mzML format. A complete list of parameters and descriptions is available at <https://github.com/gtluu/timsconvert#parameters>. All spectra are written to mzML files using the psims API (Klein and Zaia, 2019).

2.2. Conversion of LC-TIMS-MS/MS data

The conversion of LC-TIMS-MS/MS data from TDF to mzML in data-dependent and data-independent acquisition parallel accumulation-serial fragmentation (ddapASEF and diaPASEF) experiments (Meier *et al.*, 2018) is provided by the tdf2mzml tool (utilizing the TDF-SDK from Bruker Daltonics). We encode ion mobility information in the mzML file format as illustrated in Supplementary Figure S1. Notably, we include the option to export ion mobility arrays ($1/K_0$) in MS^1 spectra. In ddapASEF runs, multiple MS/MS spectra for a given precursor ion are often found in separate frames. Therefore, to achieve a data format with a more traditional data structure compatible with downstream analyses, these precursor ion spectra from multiple frames are binned and merged when exporting MS/MS spectra. Ion mobility values ($1/K_0$) are exported for the selected precursor ions, and if the charge state is available, the collisional cross-section area is included as well. All spectra are written to mzML files using the psims API (Klein and Zaia, 2019).

2.3 Conversion of MALDI-qTOF and MALDI-TIMS-qTOF DD data

TIMSCONVERT supports the conversion of MALDI-qTOF (TSF format) and MALDI-TIMS-qTOF (TDF format) DD data to mzML using the TDF-SDK. Users can specify whether to optionally merge spectra from different spots on the MALDI plate into a single file. Spectra are written using the psims API (Klein and Zaia, 2019).

2.4 Conversion of MALDI-qTOF and MALDI-TIMS-qTOF MSI data

Support for MALDI-qTOF (TSF format) and MALDI-TIMS-qTOF (TDF format) MSI data to imzML is included in TIMSCONVERT using the TDF-SDK. MSI data exported to imzML include ion mobility data that is compatible with downstream imaging software, e.g. Cardinal MSI. imzML files are written using a modified version of the pyimzML package (<https://github.com/gtluu/pyimzML>).

3. Availability

3.1 TIMSCONVERT availability

TIMSCONVERT was developed in Python 3.7 and is as a stand-alone command-line interface tool in Windows and Linux, a Nextflow workflow, or online at <https://proteomics2.ucsd.edu/ProteoSAFe/index.jsp?params={%22workflow%22%3A%20%22TIMSCONVERT%22%7D>.

3.2 Github source

All source code and documentation for usage can be found at <https://github.com/gtluu/timsconvert>. Source code and documentation for tdf2mzml can be found at <https://github.com/mafreitas/tdf2mzml>. Source code for the modified pyimzML package can be found at <https://github.com/gtluu/pyimzML/>.

4 Conclusion

In this article, we present TIMSCONVERT: a high-throughput workflow to convert a wide array of data formats generated by the timsTOF Pro and fleX instruments; specifically, TIMSCONVERT incorporates ion mobility data into mzML and imzML while maintaining downstream compatibility with existing tools. Included are a variety of real-world examples using TIMSCONVERT and their analysis in downstream tools, including but not limited to GNPS molecular networking (Wang *et al.*, 2016) and Cardinal MSI (Bemis *et al.*, 2015) (Supplementary material NOTE Use Cases 1 and 3). With the growing usage of ion mobility and the capability of existing open-source data formats to house it, we envision TIMSCONVERT to play a key role in increasing the usability of ion mobility mass spectrometry data.

Funding

This work was supported by the National Institute of General Medical Sciences Award Number [R01GM125943] (LMS); the National Cancer Institute Award Number [R01CA240423] (LMS) of the National Institutes of Health, by the National Science Foundation [2128044] (LMS); and UC Santa Cruz Startup funds (LMS).

Conflict of interest: M.W. is a co-founder of Ometa Labs LLC.

Data availability

All data are incorporated into the article and its online supplementary material.

References

- Bemis, K.D. *et al.* (2015) Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. *Bioinformatics*, **31**, 2418–2420.
- Hulstaert, N. *et al.* (2020) ThermoRawFileParser: modular, scalable, and cross-platform RAW file conversion. *J. Proteome Res.*, **19**, 537–542.
- Klein, J. and Zaia, J. (2019) psims - a declarative writer for mzML and mzIdentML for python. *Mol. Cell. Proteomics*, **18**, 571–575.
- Meier, F. *et al.* (2018) Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics*, **17**, 2534–2545.
- Silveira, J.A. *et al.* (2017) Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry. *Int. J. Mass Spectrom.*, **413**, 168–175.
- Wang, M. *et al.* (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.*, **34**, 828–837.
- Willems, S. *et al.* (2021) AlphaTims: indexing trapped ion mobility spectrometry-TOF data for fast and easy accession and visualization. *Mol. Cell. Proteomics*, **20**, 100149.