

Interactive Design and Application of MassQL Queries after Preprocessing for the Annotation of PFAS in LC-TIMS-PASEF data

Andrea Kiehne¹; Silke Bodendiek¹; Sofie Weinkouff¹; Mingxun Wang²;
Heiko Neuweiger¹; Nikolas Kessler¹

¹Bruker Daltonics GmbH & Co. KG, Bremen, Germany; ²Department of Computer Science and Engineering, University of California Riverside, Riverside, CA

Introduction

We present a new workflow for the design and application of MassQL [1] queries within an interactive data exploration software for untargeted metabolomics. The Mass Spec Query Language (MassQL) is a domain specific language meant to be a succinct way to express a query in a mass spectrometry centric fashion. Here it is applied to data after its untargeted preprocessing.

Due to the full integration into the MetaboScape® software, the workflow provides multiple tools for the identification and validation of characteristic rules for compound classes of interest. These rules can then be formalized as MassQL queries, and subsequently be applied to reveal other, potentially unknown compounds fulfilling these rules.

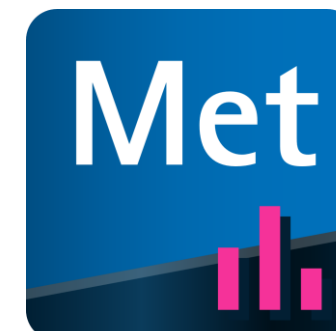
The workflow is exemplified with the annotation of previously reported PFAS compounds in drinking water and the detection of further, similar candidate molecules.

Methods

Analyses were performed using a timsTOF Pro with VIP-HESI coupled to an Elute UHPLC (Bruker Daltonics). Compounds were separated on an Intensity Solo column using a H₂O/MeOH/5 mM ammonium acetate gradient. Spectra were recorded in negative ion mode using PASEF® MS/MS. The mass range was set to 30-1000 m/z and the ion mobility range to 0.5-1.3 V/cm² with ramp times of 100 ms.

The timsTOF was calibrated for mass and mobility using a 1:1 mixture of 10 mM sodium formate and ESI Tune Mix (Agilent Technologies).

Raw data was processed using the T-ReX® 4D workflow within MetaboScape. The resulting features have been annotated with a Target List of known PFAS and additional molecular formula generation, with a ruleset tailored for highly fluorinated compounds using SmartFormula.



Design and Validation Workflow for MassQL Queries in MetaboScape

- 1 Untargeted processing of the LC-TIMS-PASEF raw data with the T-ReX® 4D algorithm.
- 2 Dereplication, annotation of known compounds with Target Lists or Spectral Libraries.
- 3 Interactive on-the-fly navigation and filtering of the Feature Table to identify reoccurring MS/MS fragments and neutral losses. Investigation of the ion mobility ranges for groups of PFASs.
- 4 Compilation of the rules derived from 3 into a MassQL query.



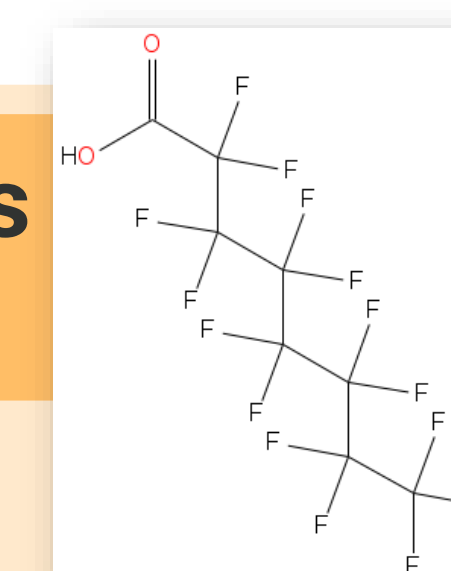
massql.gnps2.org

Alan K. Jarmusch *et al.*, 2022

- 5 Filter the Feature Table using the new MassQL query, narrowing it down to Features that meet the new criteria.
- 6 Generate molecular formulas for the remaining Features using SmartFormula. Configure SmartFormula for the targeted compound class.
- 7 Use the generated molecular formulas as a starting point to search for structural candidates in public or internal databases. Configure MetaboScape's Compound Crawler for suitable databases.
- 8 Use MetFrag *in silico* fragmentation [2] to assess whether the measured MS/MS spectrum of a Feature can be explained with one or more of the respective structural candidates.
- 9 Use CCS-Predict Pro to assess whether the measured CCS value of a Feature meets the predicted CCS value for the structure.
- 10 Annotate and repeat from 7 or 3.

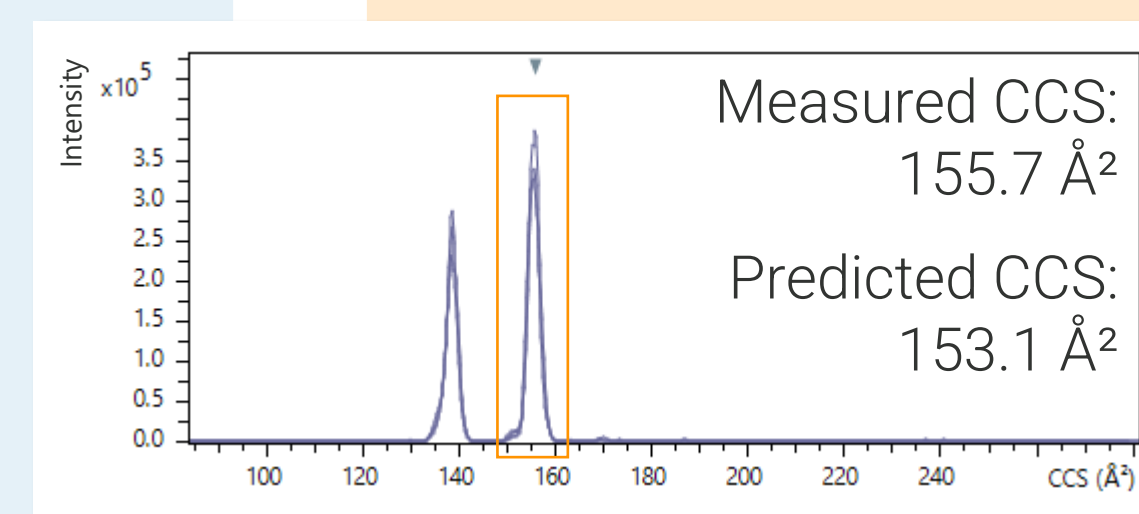
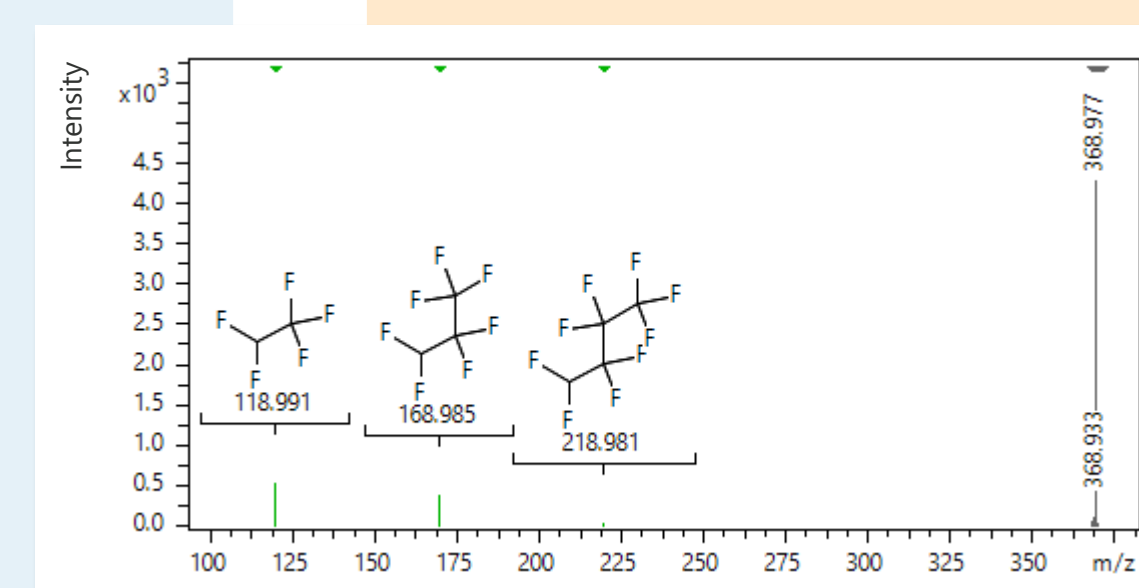
Application example to find groups of PFASs in drinking water

- 1 and 2 of the negative mode data resulted in a processed table of 975 Features. 69 Features could be annotated (48 unique) using a list of 70 known PFAS structures. Annotation confidence was supported using MetFrag *in silico* fragmentation and CCS-Predict Pro.
- 3 The MS/MS spectrum of PFOA contained fragment peaks for C₂F₅⁻, C₃F₇⁻, and C₄F₉⁻ at m/z 118.987, 168.987 and 218.986, respectively.
- 4 Created below QUERY based on the fragments described in 3.



**QUERY scaninfo(MS2DATA) WHERE
MS2PROD=118.987:TOLERANCEMZ=0.005 AND
MS2PROD=168.987:TOLERANCEMZ=0.005 AND
MS2PROD=218.986:TOLERANCEMZ=0.005 AND
MOBILITY = range(min=120,max=210)**

- 5 Above query revealed seven Features, of which two were not yet annotated.
- 6 Configured SmartFormula for elements CHFNOPS, with Seven Golden Rules [3] adapted for ratios C/H = 0.0 - 3.1 and F/C = 1.0 - 5.0.
- 7 Compound Crawler configured for the local AnalyteDB and public databases ChemSpider [4], PubChem [5] and ChEBI [6].



- 8 Assess 1-hydro-pentadecafluoroheptane (ChemSpider 9395) for m/z 368.977 using MetFrag *in silico* fragmentation: The fragments relate to the fluorinated chain common to many PFAS.
- 9 The mobilogram shows two distinct CCS peaks, of which the larger is in acceptable agreement (1.7%) with the CCS value predicted by CCS-Predict Pro.

MassQL for processed data

Originally, MassQL was designed as an expressive, precise and natural way to query raw data with suitable, predesigned queries. Here we present a workflow that allows to build upon processed data, to interactively design and validate new MassQL queries to be used and shared.

To this end, MetaboScape 2023b provides a filter for the Feature Table, that in the current version implements some of the keywords available in MassQL:

MS2PROD and **MS2NL** to query for fragments and neutral losses; with either **TOLERANCEMZ** or **TOLERANCEPPM**. **MS1MZ** or **MS2PREC** to filter for the MS1 information of a feature; and **MOBILITY** together with **range** to narrow down based on CCS information.

References

- [1] Jarmusch, A. K., et al. "A Universal Language for Finding Mass Spectrometry Data Patterns." (2022).
- [2] Ruttkies, Christoph, et al. "MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation." *Journal of cheminformatics* 8.1 (2016): 1-16.
- [3] Kind, Tobias, and Oliver Fiehn. "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry." *BMC bioinformatics* 8.1 (2007): 1-20.
- [4] www.chemspider.com
- [5] pubchem.ncbi.nlm.nih.gov
- [6] www.ebi.ac.uk/chebi/

Conclusion

- MassQL is a powerful query language, which is applicable to processed data in MetaboScape 2023b.
- To aid the design and validation of new MassQL queries, MetaboScape provides a fully integrated, interactive user interface.
- With MetFrag *in silico* fragmentation and CCS-Predict Pro for the prediction of CCS values, the applicability of new MassQL queries can be validated on-the-fly.

Untargeted Metabolomics