Tim Leefmann[1], **Stephan Frickenhaus**[1], **Boris P. Koch**[1]

[1]*Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Germany*

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

A fast library based formula search approach for high-resolution mass spectrometry

Introduction

In the evaluation of high-resolution mass spectrometric data a considerable amount of time and computational power can be spent on matching molecular formulas to the neutral mass of measured ions. During the evaluation of multiple samples using the classical combinatory approach based on molecular building blocks and nested loops, the time consuming step of calculating the molecular mass may be repeated for the same molecular formula multiple times. Here we present a new formula assignment algorithm that is based on prebuilt molecular formula libraries and thus avoids repetitive calculations of molecular formulas.

Library based algorithm

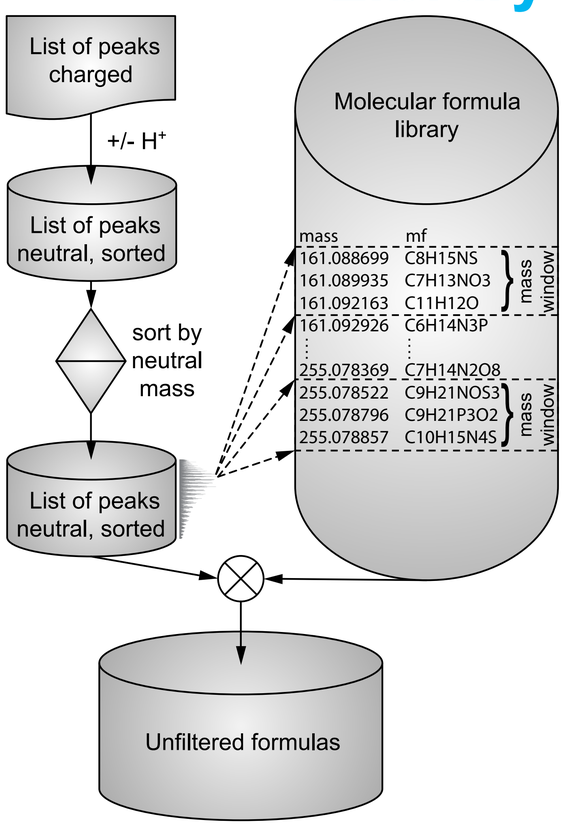

Figure 1. Flow chart illustrating the formula matching algorithm.

The formula assignment algorithm

- was coded in R[1]
- uses prebuilt, static molecular formula libraries
- builds on comparison of sorted peaklists with sorted libraries in the data.table[2] format
- was performance-tested on a common workstation (Win10 64bit, i5-6500 3.20 Ghz, 8GB RAM, SATA HDD) with 50 samples of marine dissolved organic matter extracts comprising 413,547 peaks

Performance

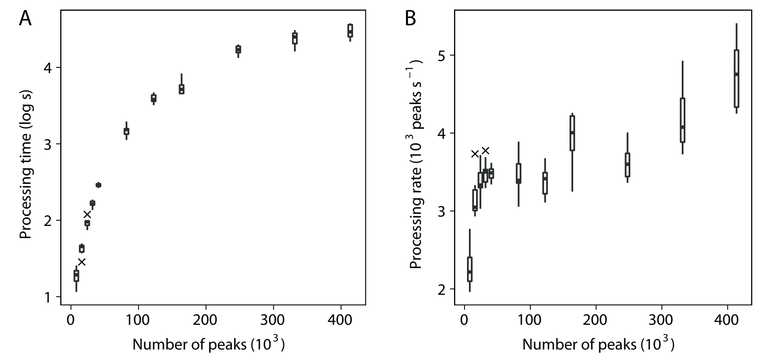

Figure 2. Formula matching algorithm benchmark. Standard boxplots (n=10) of the processing time (A) and the processing rate (B), respectively, vs the number of peaks supplied.

- the assignment rate increases with the length of the supplied peaklist
- the assignment rates reached 4,745 peaks s^{-1}
- a set of 50 samples with 413,547 peaks was processed in Ø 88 s

Integration into UltraMassExplorer

The formula assignment algorithm forms the basis of the web-application UltraMassExplorer (UME). The graphical user interface of UME builds on R Shiny[3] and allows for the easy integration of the R based algorithm. UME provides the user with

- a complete data pipeline for high-resolution mass data comprising
  - the formula assignment algorithm
  - advanced filter functions
  - linkage to the PubChem data base for searching compounds corresponding to molecular formulas
  - export of data, metadata and publication-quality graphics
- the capability for swift reanalysis of complete datasets
- an interactive data evaluation experience through the on-the-fly display of filter effects
- a transparent, open-access source code in R allowing for straightforward improvement and extension of UME

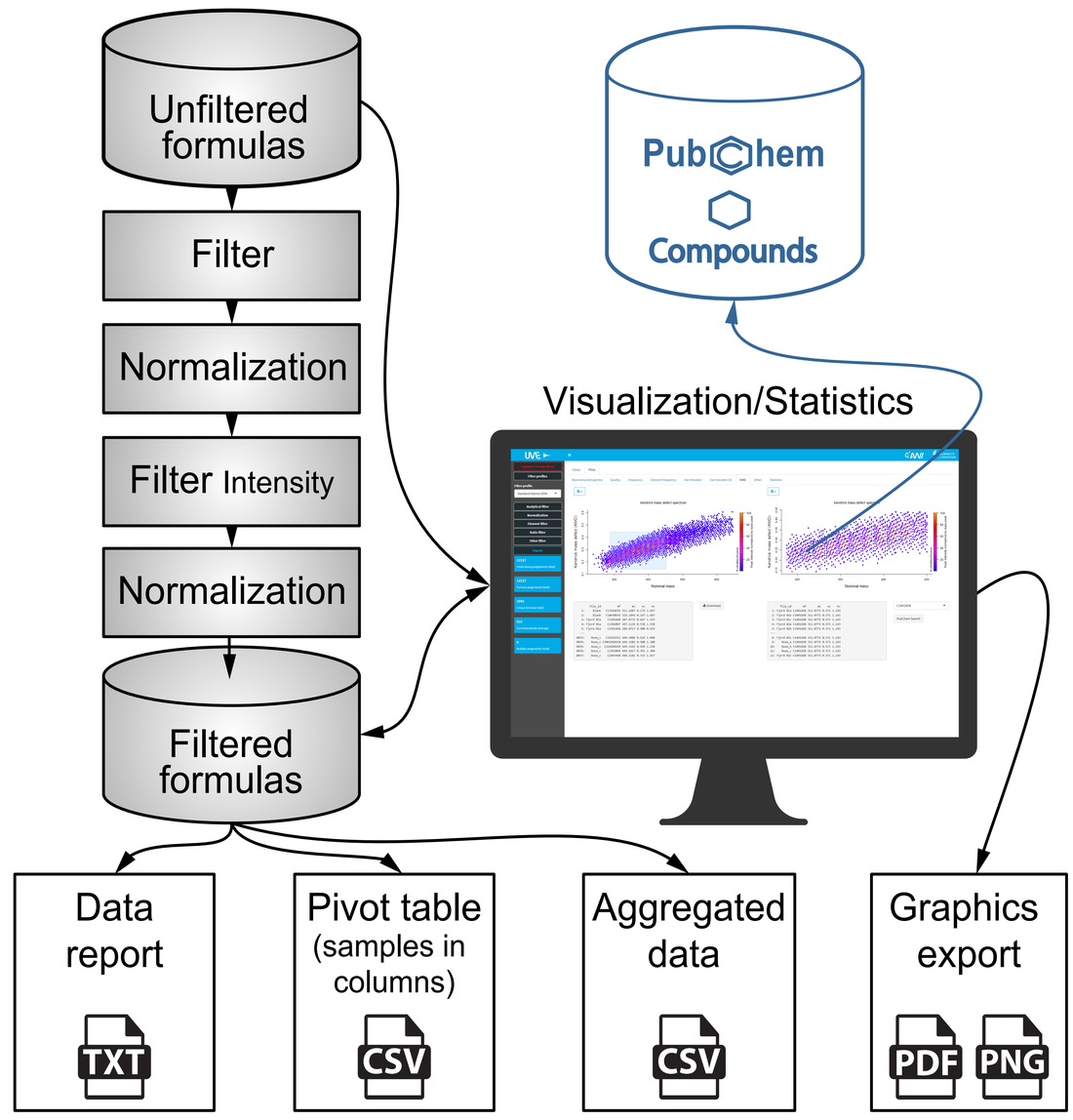

Figure 3. Flow chart illustrating the processes in UME from the display of unfiltered results to the export of final results and report generation.

References

1. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
2. Matt Dowle and Arun Srinivasan (2017). data.table: Extension of `data.frame`. R package version 1.10.4-2. https://CRAN.R-project.org/package=data.table
3. Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web Application Framework for R. R package version 1.0.5. https://CRAN.R-project.org/package=shiny

Acknowledgements

TL and BPK were kindly supported by the strategy fund of the Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research (AWI; project title: “Inorganics in organics: Chemical and biological controls on micronutrient and carbon fluxes in the polar ocean”). The authors thank Christian Schaefer-Neth, Angelo Steinbach and Jörg Matthes from AWI for setting up the Docker server for UME and helpful inputs on developing a gateway site for UME. Ana Macario and Sylke Wohlrab helped building the website. Olaf Graetz is acknowledged for Piwik integration and Jens Michael Schlüter for securing safe access to UME from outside the institute. Oliver Lechtenfeld, Jana Geuer and Marianna Lucio provided valuable comments on a previous version of UME.

Website

Tutorial

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

HELMHOLTZ GEMEINSCHAFT

BREMERHAVEN

Am Handelshafen 12
27570 Bremerhaven
Telefon 0471 4831-0
www.awi.de