

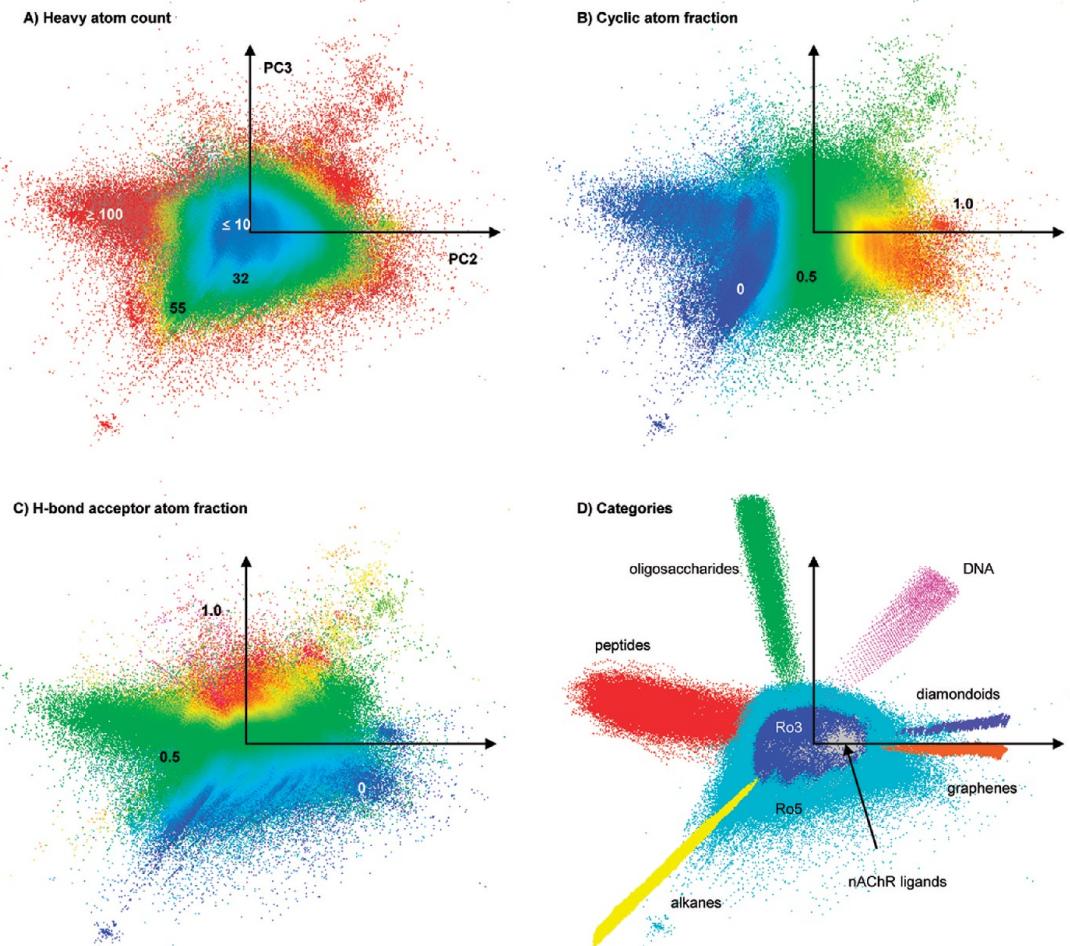


InSpectra – A Global Platform for Identifying Emerging Chemical Threats

Saer Samanipour, M. Feraud, D. van Herwerden J. W. O'Brien, J. Boelrijk, M. J. Reid, and K. V. Thomas

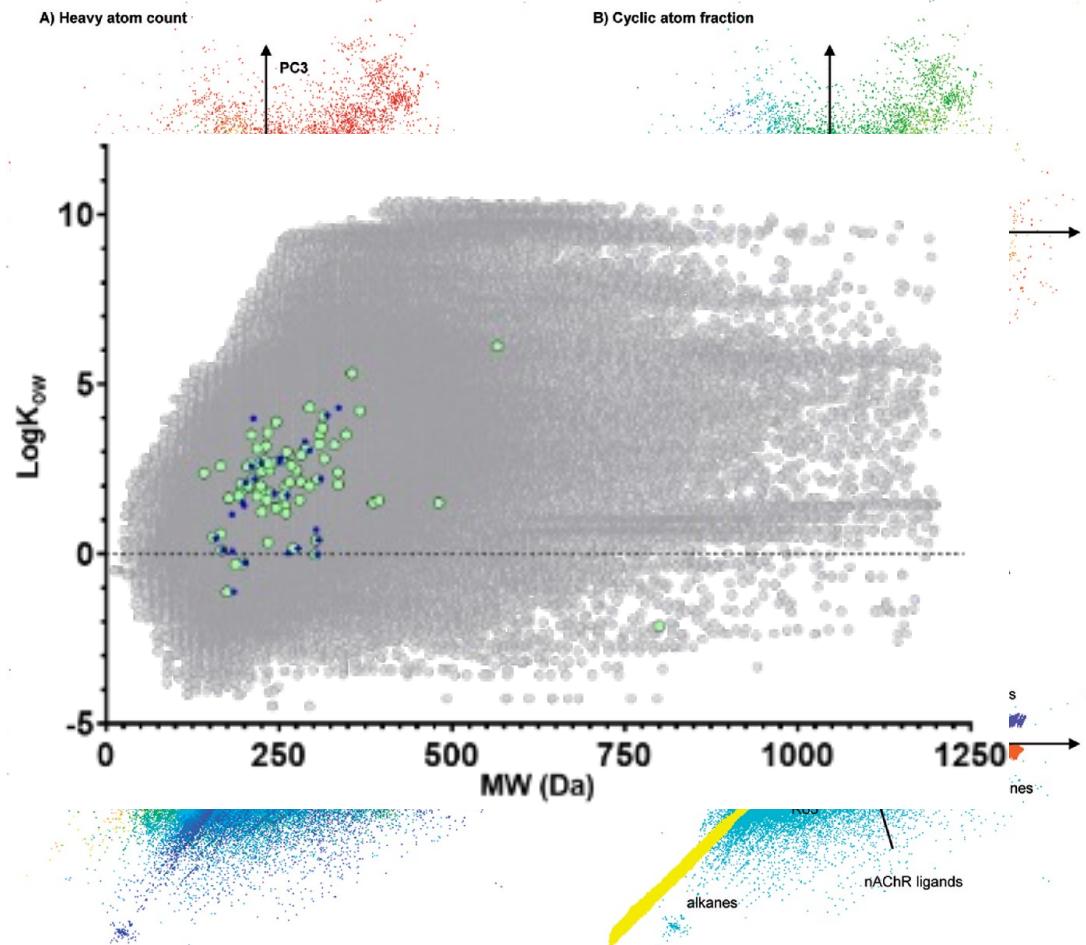


As environmental chemists, we aim at assessing and minimizing the health associated impact of chemical exposure.



- There are $> 10^{60}$ possible structures with $M_w < 500$ Da.
- The physiochemical property range is too wide.
- Technologically we cannot cover this space.

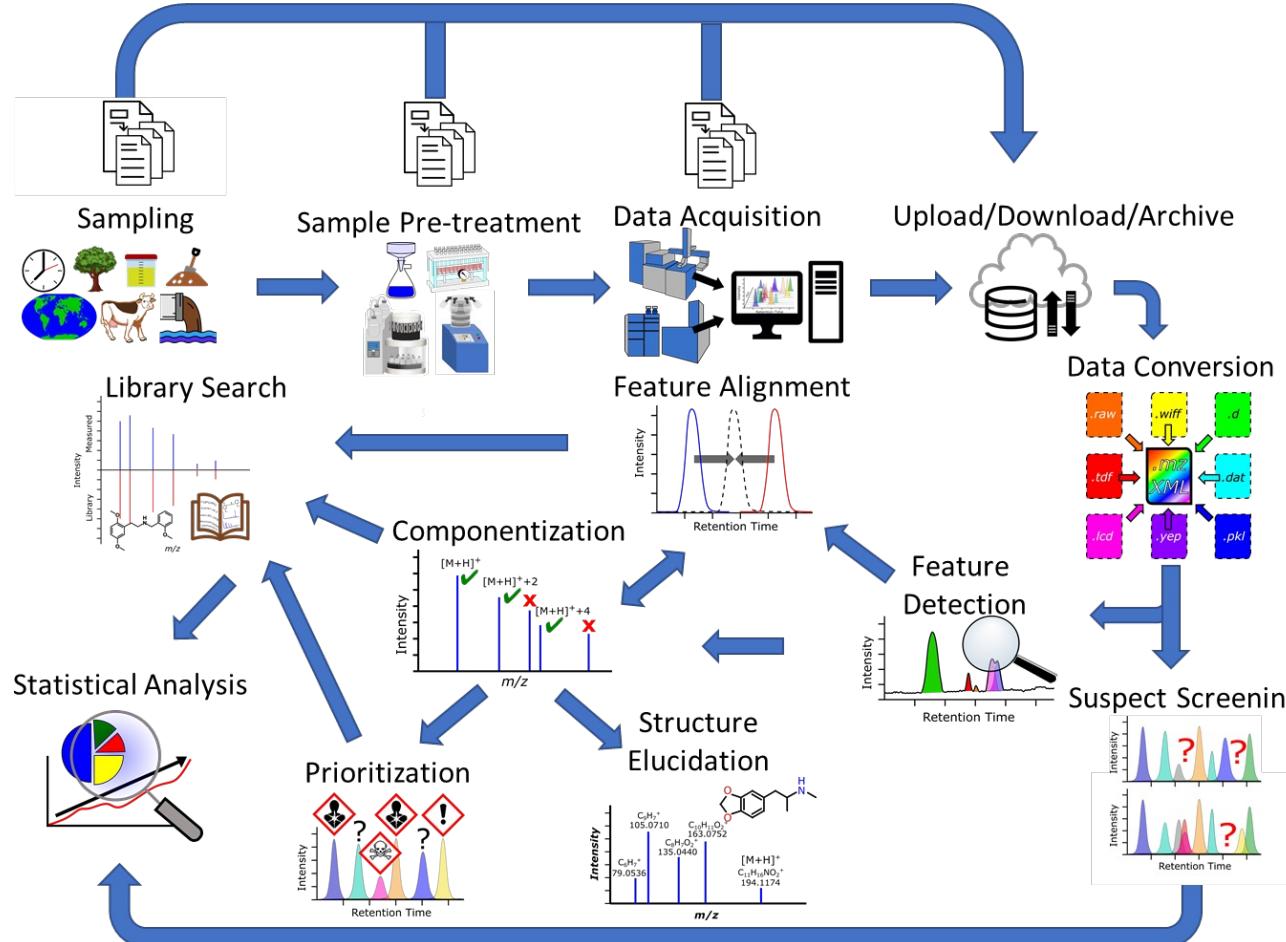




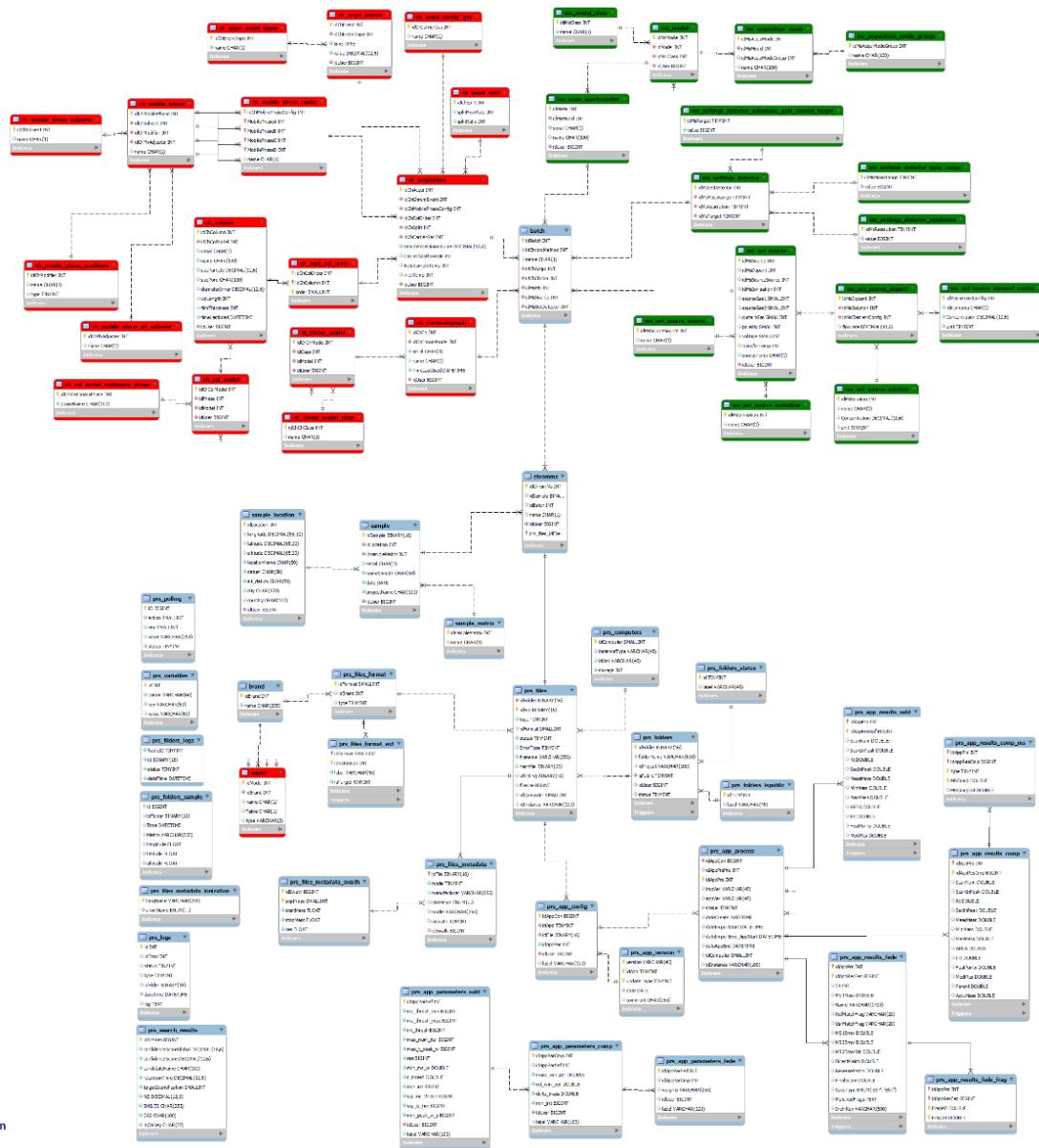
- There are more than 800k known chemicals that are actively released into the environment.
- The transformation products are ignored.
- All natural chemicals were excluded.
- We have methods for less than 1%.



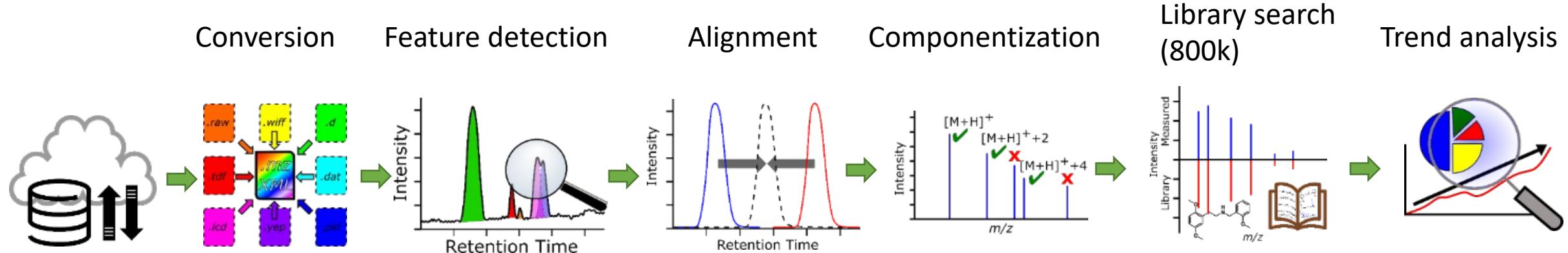
INSP_{z/w}^{cps} ECTRA



- A completely modular and FAIR platform.
- It is meant for archiving and retrospective analysis.
- It includes a spectral database of 800 k chemicals (experimental and in-silico).
- Different workflows can be set up by the user.
- It is highly scalable.

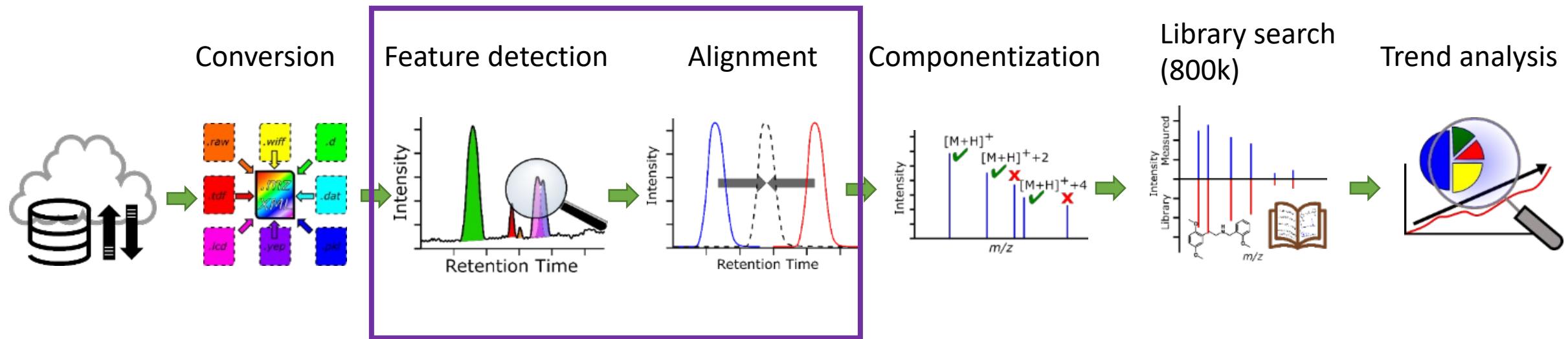


- It contains a relational database for the metadata and reports.
 - All the method parameters are also stored with the results.
 - The results can easily be queried for further analysis.
 - The same dataset can be examined multiple times in a reproducible manner.



A typical workflow consists of the raw data upload, waiting, WAITING, and then downloading the reports.

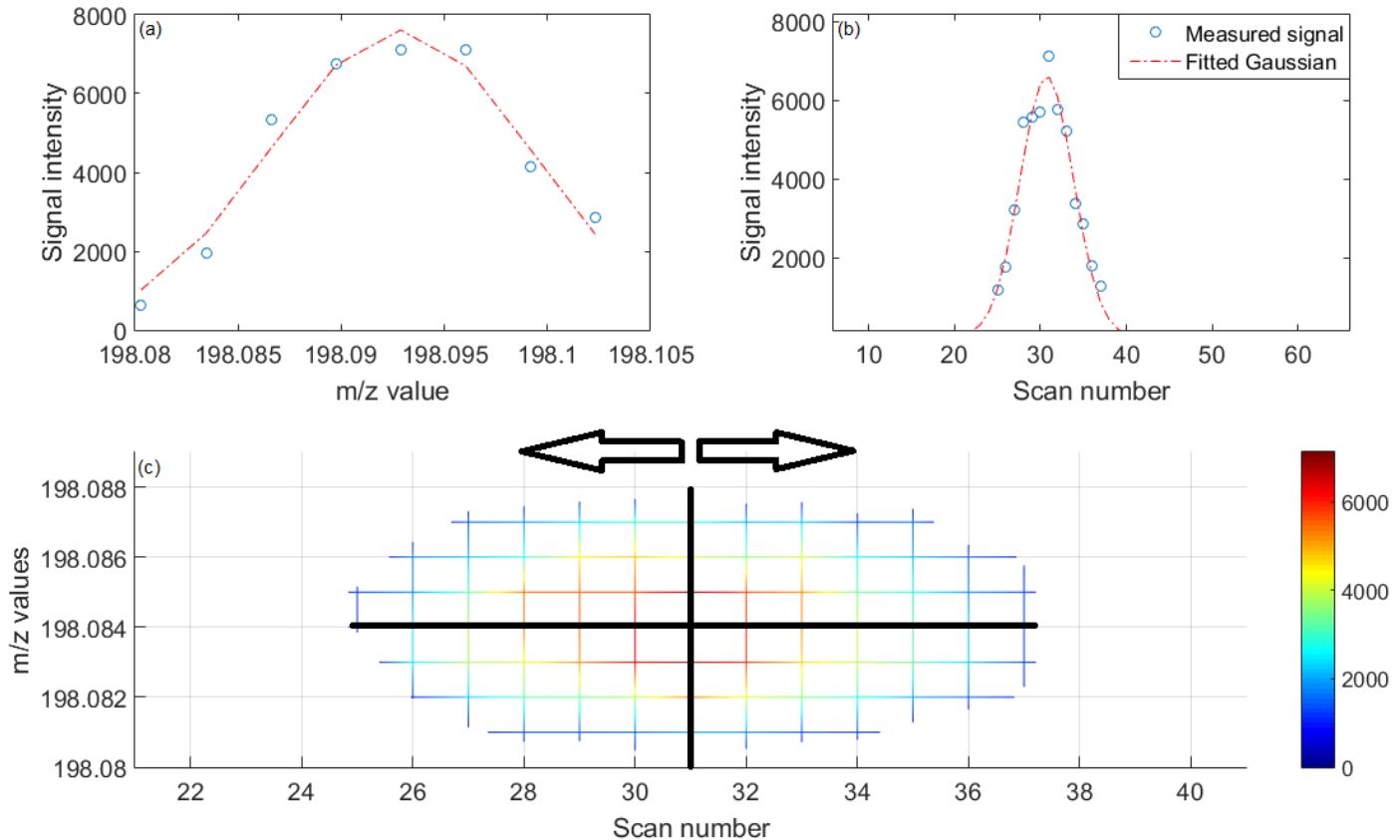




A typical workflow consists of the raw data upload, waiting, WAITING, and then downloading the reports.

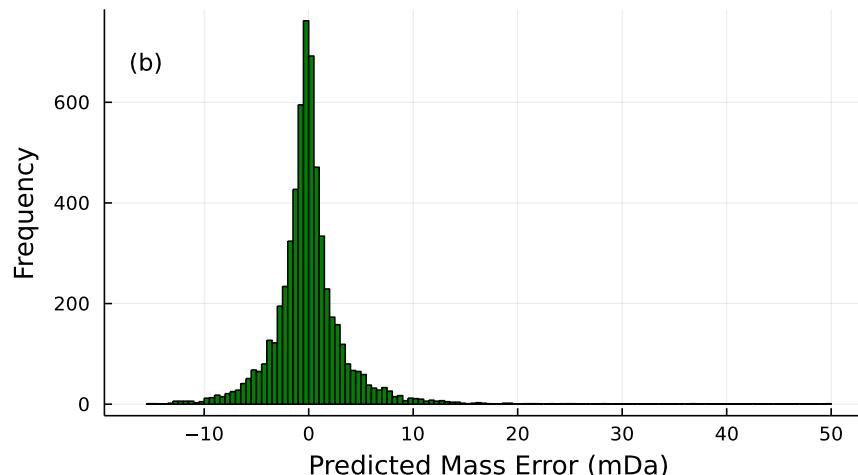
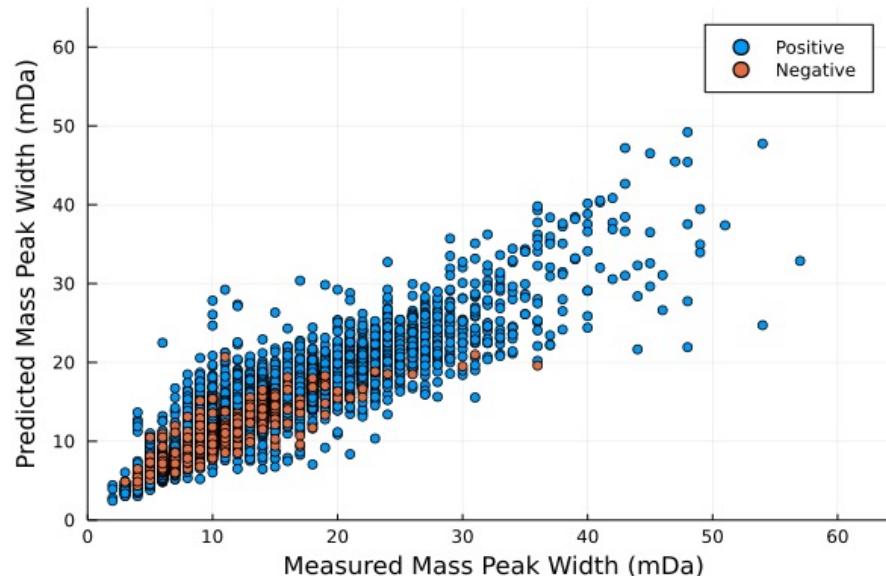


SAFD



1. Gaussian function is fitted into the data in both mass and time domain.
2. The signal that meets the fit criteria is considered a chromatographic feature.

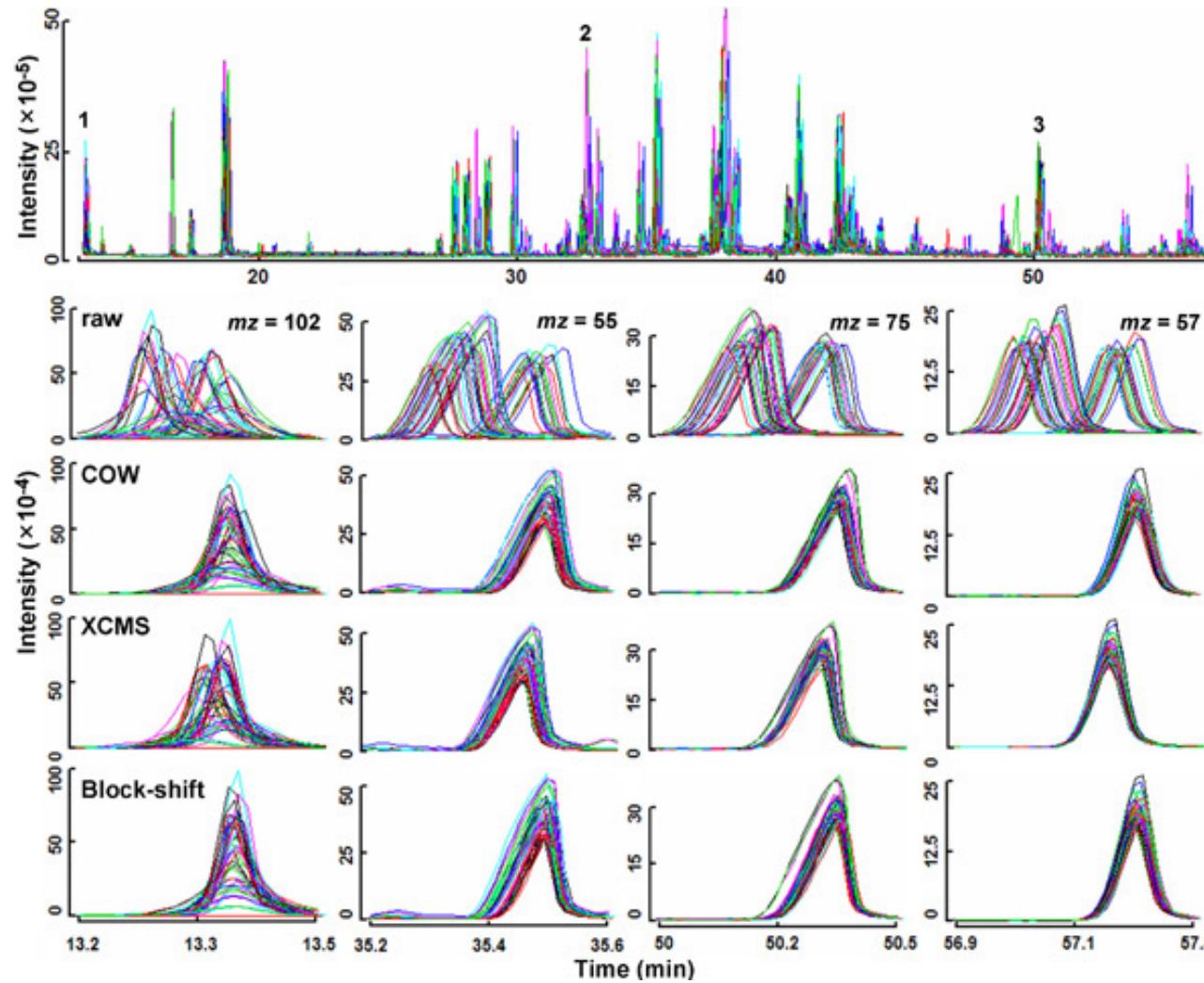




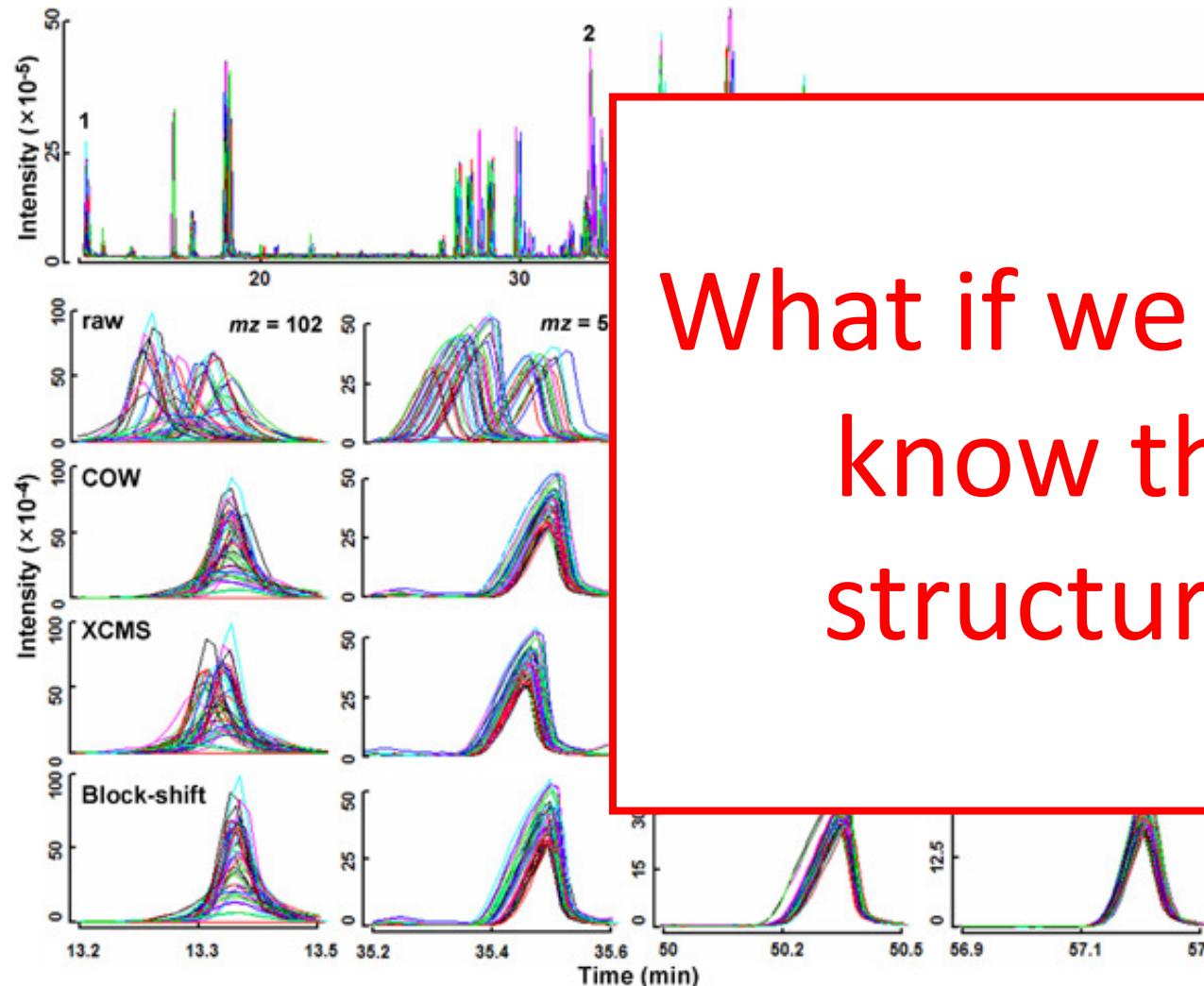
1. The model explained 85% variance in the data.
2. More than 95% of error was between pm 10 mDa.
3. The model appeared to underestimate the peak widths for peak widths larger than 30 mDa.

This is the only feature detection algorithm that can perform feature detection on both centroided and profile data.





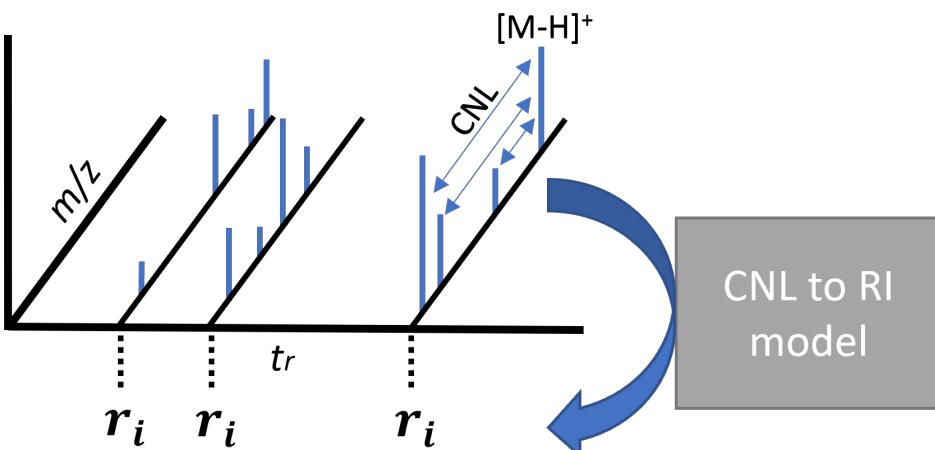
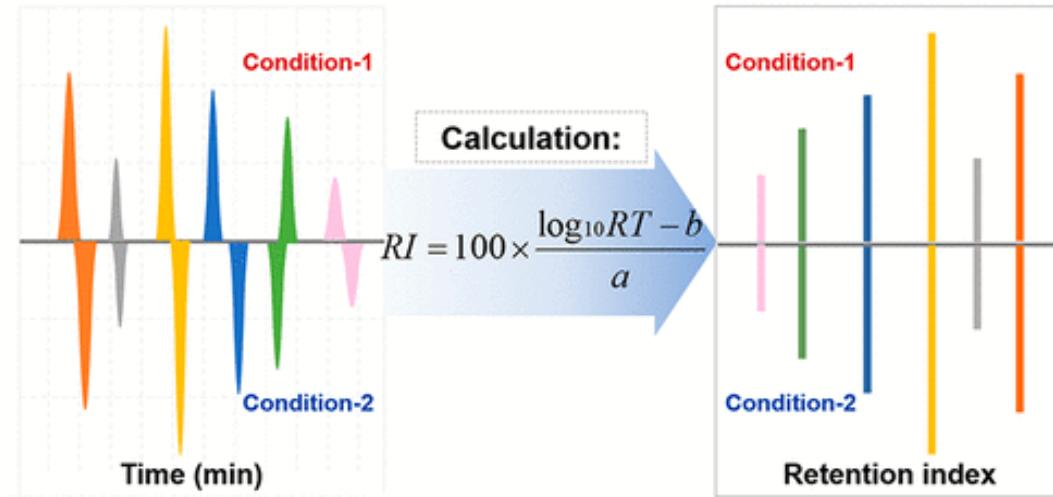
1. Current tools either are set for the chromatograms run using the same methods.
2. The chemical identity is used to align chromatograms with run with different methods.



What if we do not know their structures?

ent tools either are set for chromatograms run using same methods.

chemical identity is used to in chromatograms with run different methods.



1. The retention times are converted to retention indices (RI).
2. The mass spectra are converted to CNL spectra.
3. The RI values are predicted using the CNL spectra.

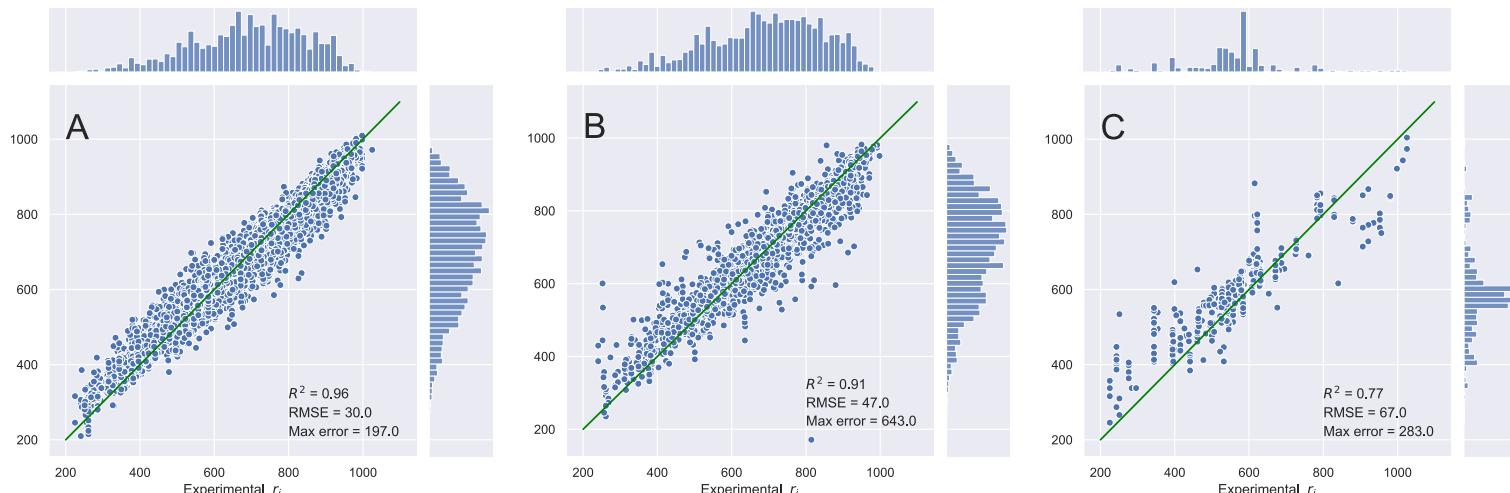


Figure 5: Parity plot of the CNL model predictions and the experimental r_i values for the training set ($n=17998$) (A) the external NORMAN test set ($n=3131$) (B) and the external amide test set ($n=604$) (C) with the coefficient of determination (R^2), root mean squared error (RMSE) and maximum error. In addition, marginal distributions of the experimental and predicted r_i are shown.

1. We were able to predict the RI values for structurally unknown chemicals only based on their CNL spectra.
2. The model can be used for the alignment of chromatograms, without the method knowledge.



- We have built an online platform to archive, analyze, and share the LC/GC-HRMS data.
- The system is modular and scalable.
- It incorporates the latest tools FAIR tools available.
- It is a collaborative project. Please get in touch, if you want to contribute.

- The ion mobility data is a must.
- We do not have a UI yet. Hopefully, it will come soon.
- More toxicity related prioritization tools will be incorporated as they are being developed.

www.emcms.info



Code (public - FAIR)





UNIVERSITY OF AMSTERDAM
Van 't Hoff Institute for Molecular Sciences

