PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems

Jan Stanstrup[1,*], **Steffen Neumann**[2], **Urška Vrhovšek**[1]

[1]*Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige (TN), Italy.*

[2]*Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Halle, Germany.*

*Email: jan.stanstrup@fmach.it

Download me!
goo.gl/Vx57BC

The objectives of PredRet are to

- ❑ build a community-driven database of retention times (RTs) of compounds.
- ❑ create projection models between the RT of compounds in different chromatographic systems (CS).
- ❑ use these models to predict the RT of compounds not experimentally determined in different CSs.
- ❑ compare predicted and experimental RTs to pin-point possibly erroneous compound annotations.

Background

Databases of experimental LC-MS data have been developed with great success with regards to compound fragmentation and these databases have recently been used by automated tools to assist compound identification.

But utilizing *only* the fragmentation means disregarding one, equally important, half of the information in *LC*-MS.

We therefore build a database of compounds' RTs and use this database to predict the RT of compounds in systems where they have not been experimentally determined.

Conclusions

- ✓ Users can upload RTs of compounds in their own systems. Predicted RTs of other compounds can then be downloaded.
- ✓ Average prediction error is at current 0.13 min (2.6%).
- ✓ Community support will expand the database and lead to greater coverage and accuracy.
- ✓ The tool is open source and available at www.predret.org.

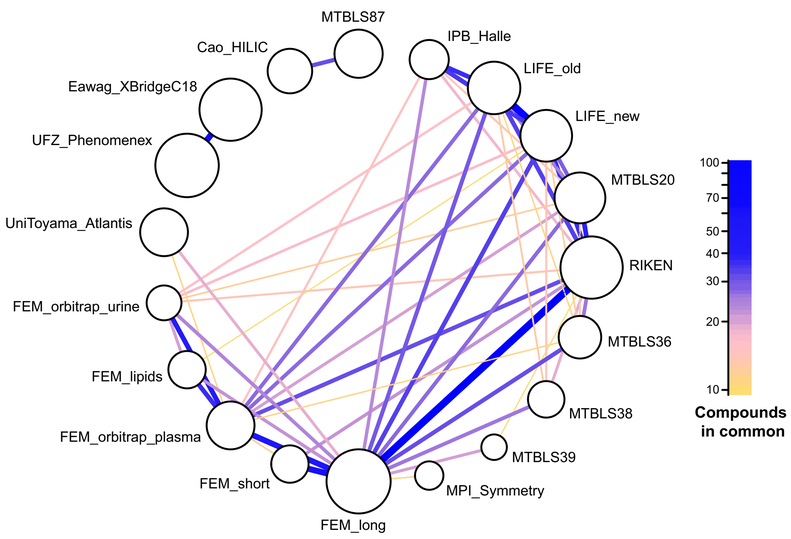

Figure 2. Compound coverage overlap. The lines connecting the CSs show the extent of overlap and go from thin orange to thick blue.

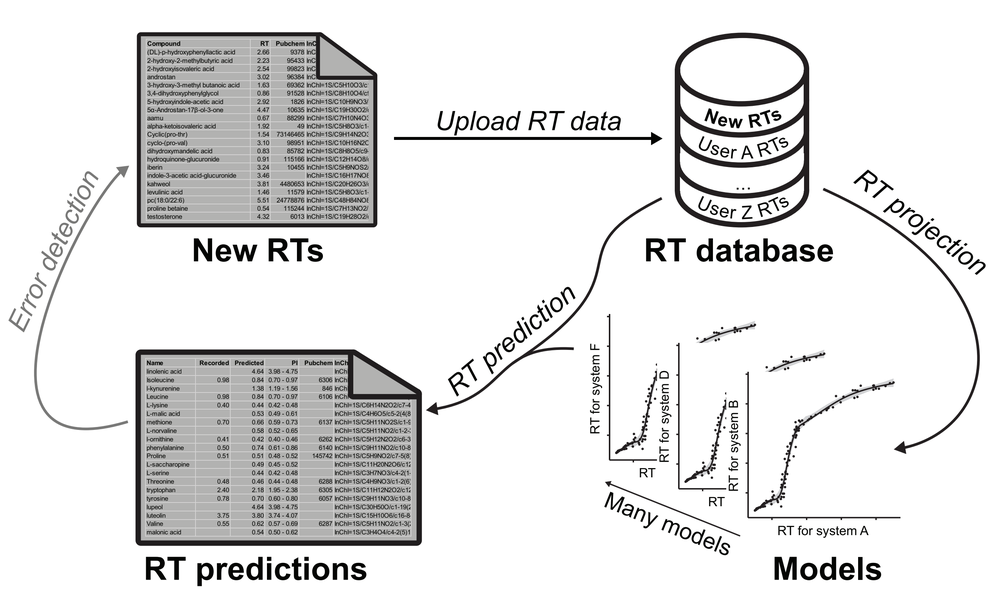

Figure 1. PredRet workflow. First data is uploaded by the user, then models are build and RTs predicted. Finally predicted RTs are reported.

Results

Building models between all CSs in the database allows the prediction of the RT of compounds in CSs where they have not been experimentally determined. The accuracy with which RTs can be predicted in a given CS depends mainly on the extend of compound coverage overlap with the other CSs in the database.

With the current small database (3,300 entries across 23 systems) it was possible to predict up to 400 RTs with a median errors between 0.01 (0.13%) and 0.28 min (3.7%) depending on the system.

Contrary to other prediction systems PredRet provides prediction intervals (PI) for each prediction. The median width of the PIs were between 0.08 min (2.6%) and 1.86 min (17%) depending on the CS.

The predictions can not only be used to obtain data for additional annotation of untargeted datasets. By comparing predicted RTs to those reported errors in the reported data can also be pin-pointed.

Methods

The prediction tool is made available as a web application at www.predret.org. The user can upload a spreadsheet with RTs of compounds measured in their own CS along with molecular identifiers such as PubChem CIDs or InChIs.

PredRet will then map the RT of compounds between CSs by building models between RTs determined in two different CSs. This model is then used to predict the RT of compounds where the RT is known in one CS, but not in the other. Building these models between all CSs in the database thus allows predicting the RT of a high number of compounds in CSs where they have not been experimentally determined.