WAGENINGEN UNIVERSITY
WAGENINGEN UR

netherlands eScience center

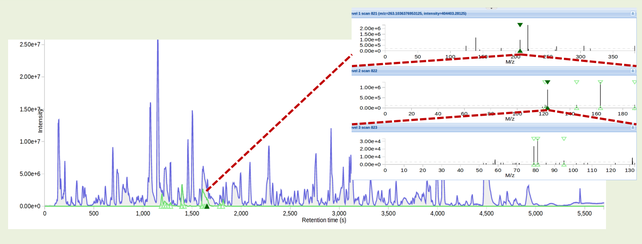
Automatic metabolite annotation in complex LC-MS$^{(n \geq 2)}$ data using MAGMa

Lars Ridder*, Justin J.J. van der Hooft, Stefan Verhoeven, Ric C. H. de Vos, Raoul J. Bino, Jacques Vervoort

*Laboratory of Biochemistry, Wageningen University, The Netherlands (lars.ridder@wur.nl)

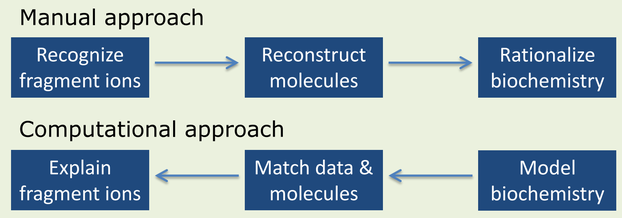
Problem

- The manual annotation of unknown compounds in complex LC-MS^n datasets is time-consuming and requires specific knowledge of the detected compound classes and their fragmentation patterns in the mass spectrometer.



Objective: develop algorithms and tools (MAGMa) to

- automatically interpret multistage MS^n spectral trees based on substructures of candidate molecules
- systematically process untargeted LC-MS^n datasets for comprehensive compound annotation
- predict candidate molecules not present in chemical databases



Methods

- Recursively split molecules by removing atoms
- Bitwise representation of substructures (fast processing)
- Substructure penalty score depends on disconnected bonds:

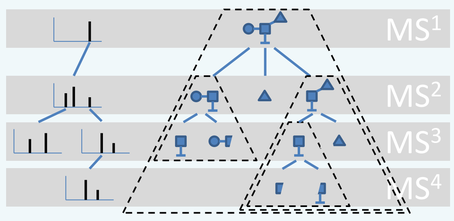
$$Substructure\ score = \sum_{b \in \underset{bonds}{disconn}} h_b p_b$$

Bond type	Value
Single bond	p=1
Double bond	p=2
Triple or aromatic bond	p=3
Carbon-carbon bonds	h=2
Bonds involving non-carbon atoms	h=1

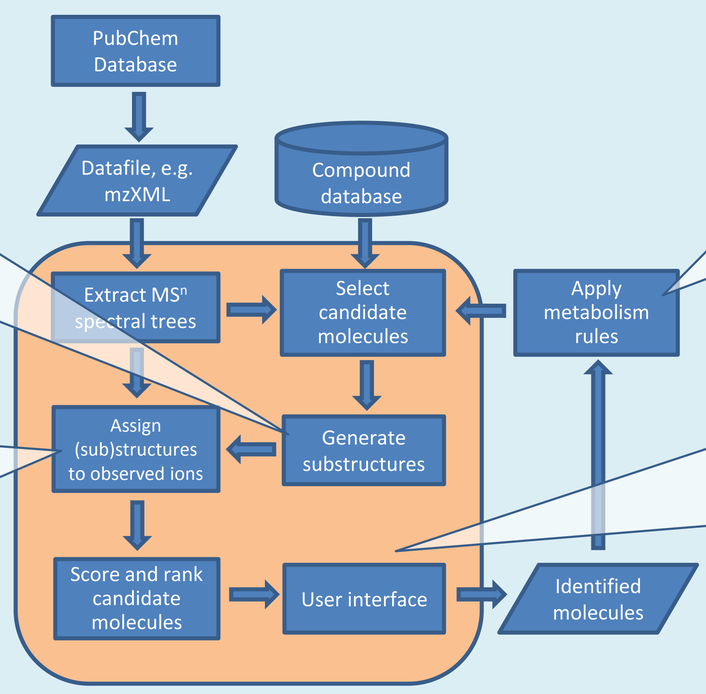
- Example: buspirone substructures versus expert (literature)
- Rearrangements are not included

m/z	Assigned substructure	score	Literature
122.0712		2	
150.1024		2	
152.1067		3	
222.1479		1	

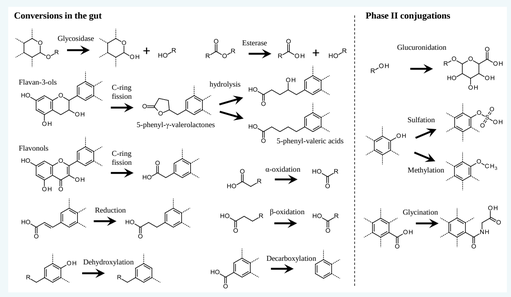
- Recursive algorithm matches multistage MS^n data with hierarchical trees of *in silico* generated substructures
- Substructure assignments at each MS level take the assignments of the precursor as well as subsequent fragmentations into account
- Total score is √(intensity) weighted average of substructure scores



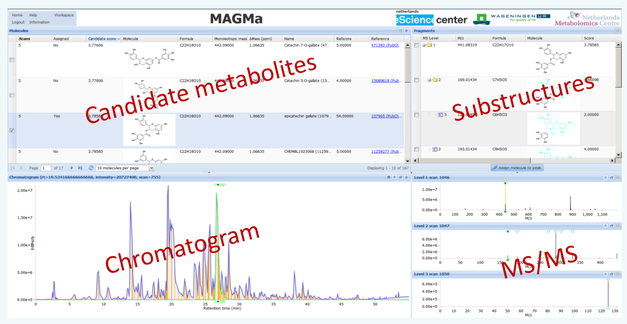
Schematic overview of MAGMa



- *In silico* reaction rules for gut metabolism and human phase I and II biotransformations are encoded as SMIRKS
- Rules are applied following a multistage "scenario" involving different types of subsequent metabolism



- Four interactive panels: candidates, chromatogram, MS^n spectra and substructures:
- Select a scan and a candidate to show the fragment annotation



Evaluation of candidate ranking

- MS/MS of 100 drugs, ramp 10-50 eV
  - On average 248 candidates per dataset

ranking statistics	selected spectra		merged spectra	
	median rank	3rd quartile rank	median rank	3rd quartile rank
MAGMa, *NBD* = 2	4	35	4	17
MAGMa, *NBD* = 3	3	17.5	3	11
MAGMa, *NBD* = 4	3	14.5	3	9
Hill et al. [b]	4[a]	17.5 [a]		
MetFrag[a]			4.5 [a]	11.75 [a]

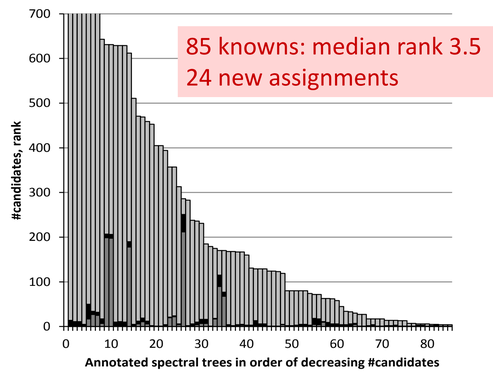
[a] Wolf et al. *BMC Bioinformatics* **2010**, 11, 148. [b] Hill et al. *Anal. Chem* **2008**, 80, 5574

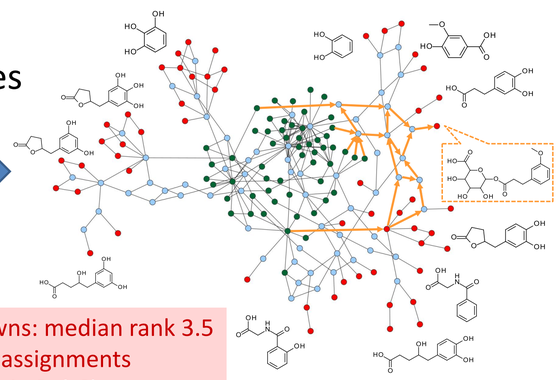
- CASMI 2013

#	Compound	# candidates	rank
1	Feruloyl tyramine	1084[a]	1
2	Feruloyl putrescine	631[a]	3
3	N2-Acetyl glutaminyl leucinamide	370	17
4	Dihydrochalcone	825	78
5	Isoprothiolane	350	2
6	Phosphatidyl-6-acetyl-glucose	7	1
7	Cinnamtannin A3	17	1
8	Prodelphinidin C2	1	1
9	Chlorpyrifos	113	1
10	VAL-HIS-LEU-THR-PRO-VAL-GLU-LYS	20	1
11	Demethoxycurcumin	906[a,b]	6
11	Demethoxycurcumin (tautomer 1)	906[a,b]	4
12	Baicalein	813[b]	271
13	EST; Aloxistatin	207	42
14	Tetrahydroalstonine	1583[a]	5
15	2-(Perfluorooctyl)ethanol	720[b]	2
16	Ofloxacin	998[a,b]	18

[a] Candidates were filter by refscore > 5.
[b] MAGMa results generated retrospectively.

Application: urinary metabolites of compounds in green tea





71 knowns: median rank 3.5
26 new assignments
77% not in PubChem

Conclusions

- MAGMa succesfully prioritizes correct candidate molecules based on (multistage) MS^n spectral data, and automatically assigns relevant substructures to multiple levels of MS fragments.
- Application to untargeted LC-MS^n profile of green tea assisted putative identification of new compounds.
- The combination with *in silico* biotransformation lead to annotation of novel urinary metabolites.
- MAGMa makes chemical interpretation of LC-MS^n data more systematic and faster.

References

- Van der Hooft et al. *Anal. Chem.* **2011,** *83*, 409
- Ridder et al. *Rapid Commun. Mass Spectrom.* **2012,** *26*, 2461.
- Van der Hooft et al. *Anal. Chem.* **2012,** *84*, 7263
- Ridder et al. *Anal. Chem.* **2013,** *85*, 6033.
- Ridder et al. *Anal. Chem.* **2014***,* **DOI:** 10.1021/ac403875b