



Improving Metabolite Identification with Chemoinformatics

Julio E. Peironcely → @peyron

PhD student at Leiden University and TNO

June 9th 2011
ICCS 2011

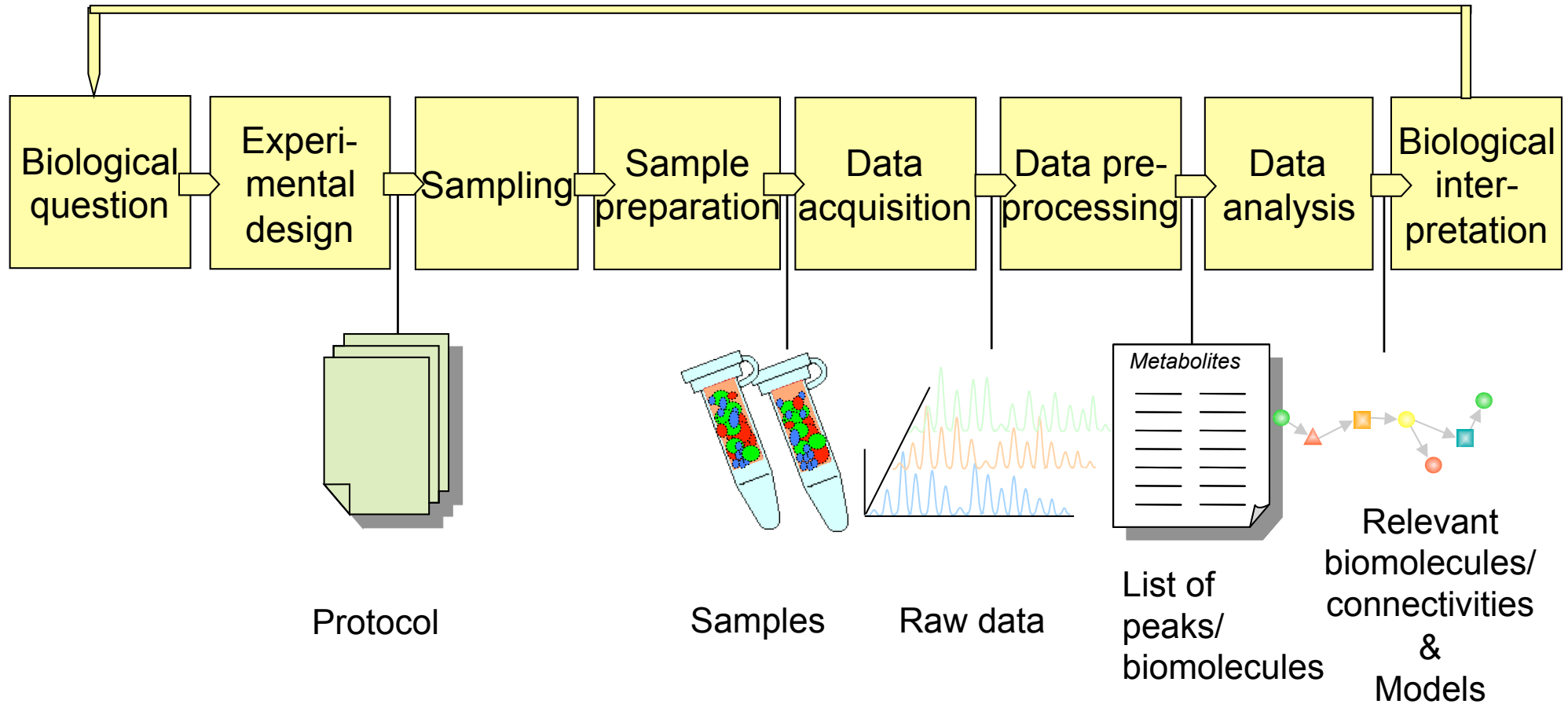


Netherlands
Metabolomics Centre

Metabolome
all low molecular weight molecules
(metabolites) in cells, body fluids,
tissues, etc.

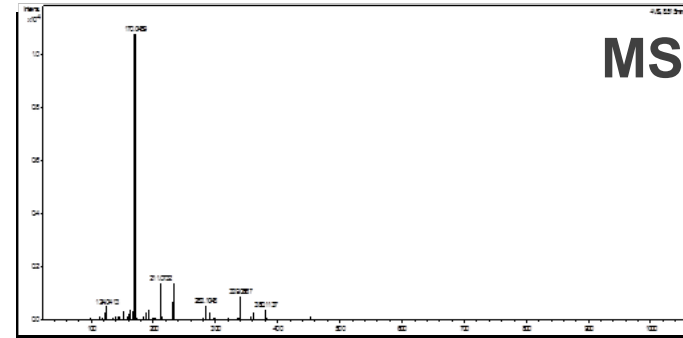
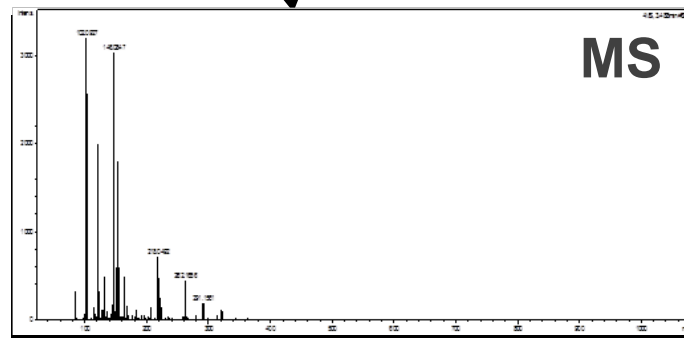
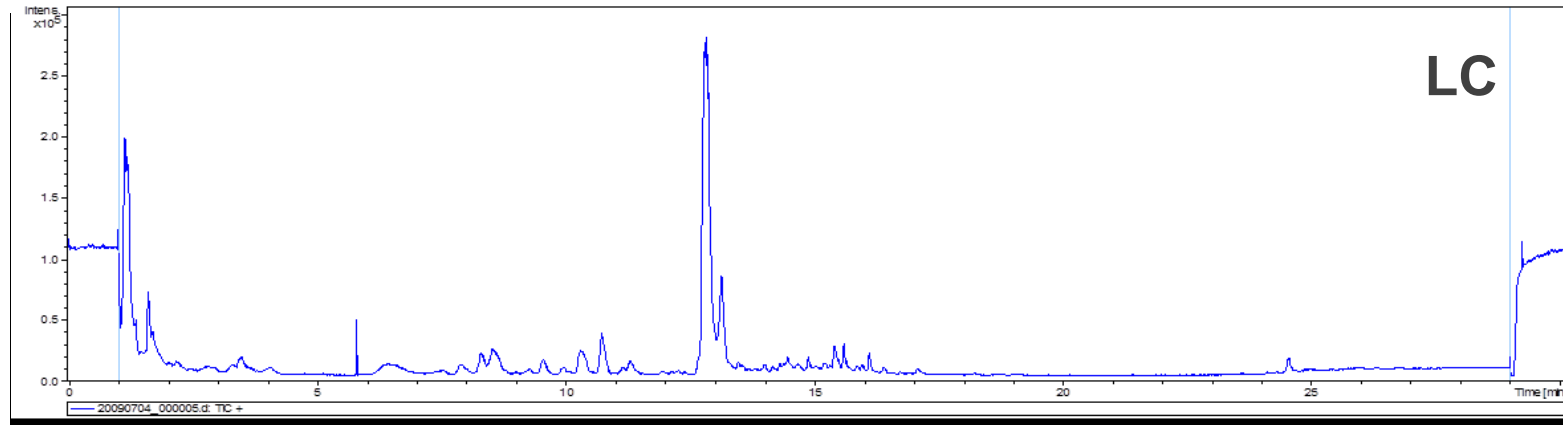
Metabolomics
the quantitative and qualitative
analysis of all metabolites in
samples of cells, body fluids,
tissues, etc.

Metabolomics



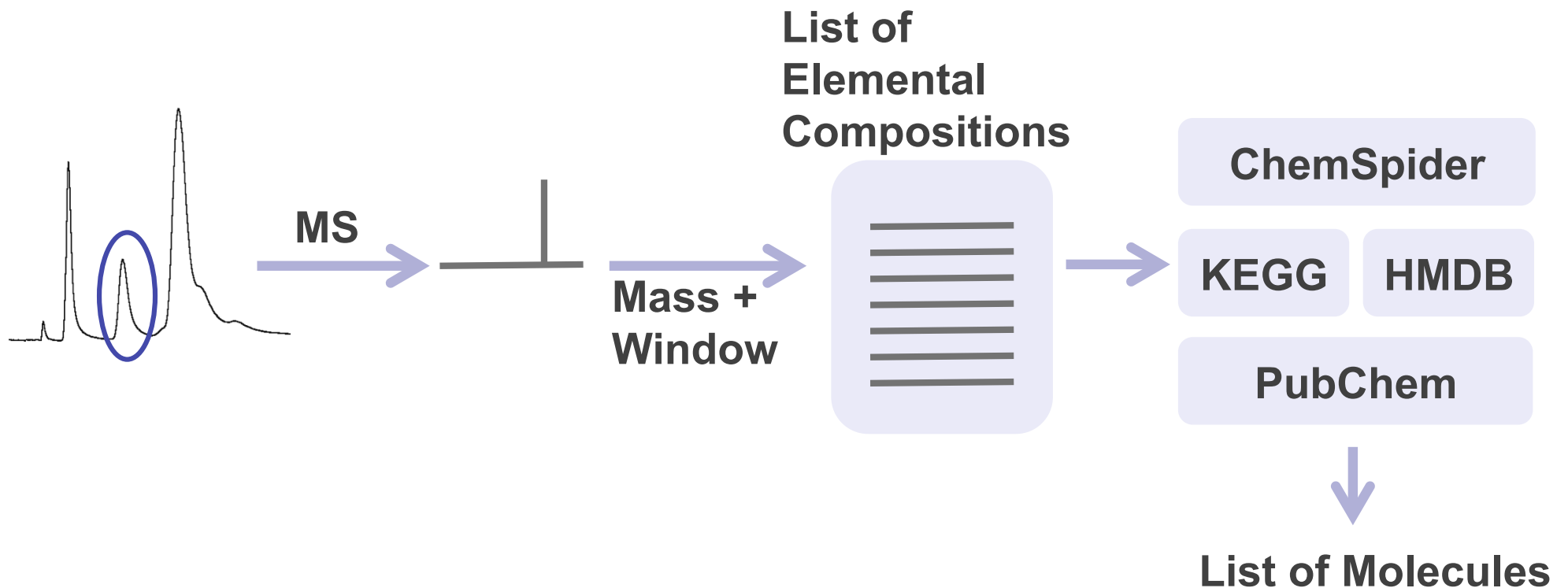
LC-MS and Metabolite Identification

RPLC-microTOF (Bruker)



peak = $m/z@rt$ = metabolite

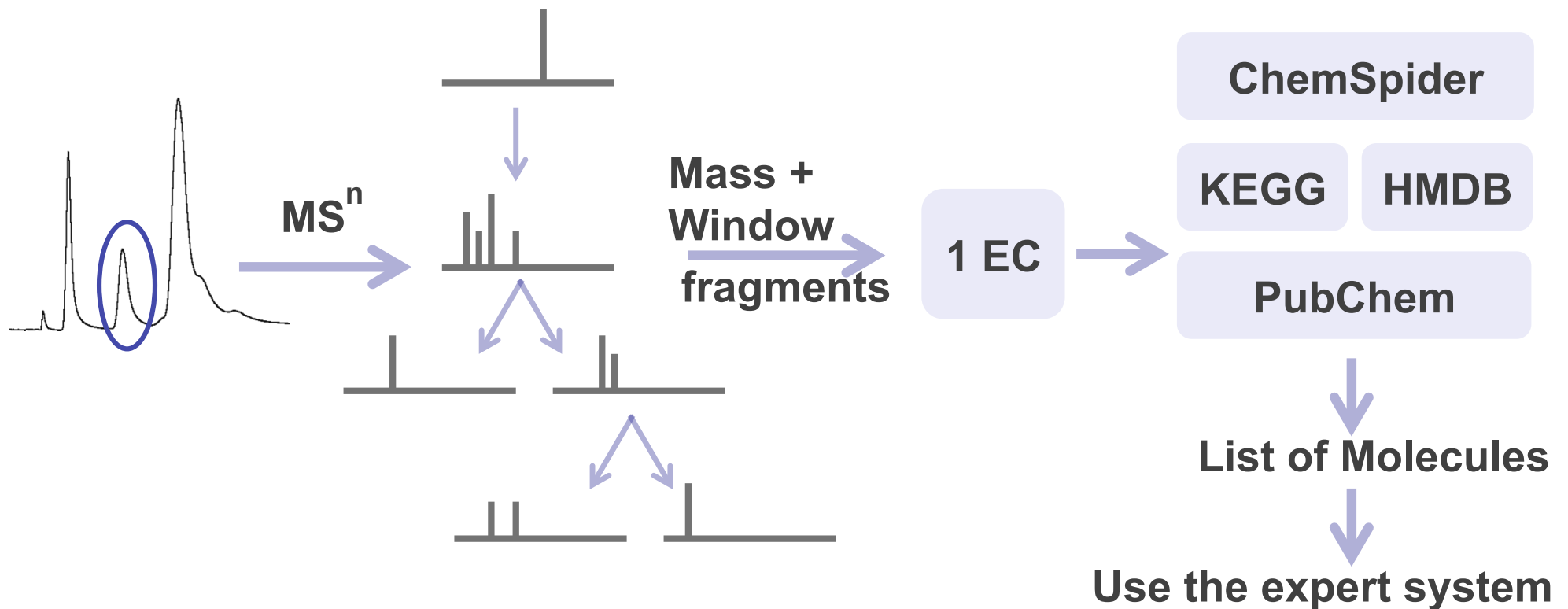
When all you have is a mass and a “window”



Measured Mass + Mass Window = Multiple Elemental Compositions

When you have MSⁿ

Measured Mass + Mass Window + Fragments
=
Single Elemental Composition



The expert system



Dr. Ronnie van Doorn



Dr. Albert Tas



Bottlenecks in Metabolite Identification

Many metabolites in LC-MS not identified

HighRes MS can obtain 1 EC = many structures

No tools for automatic identification

Takes long for the expert to identify metabolites

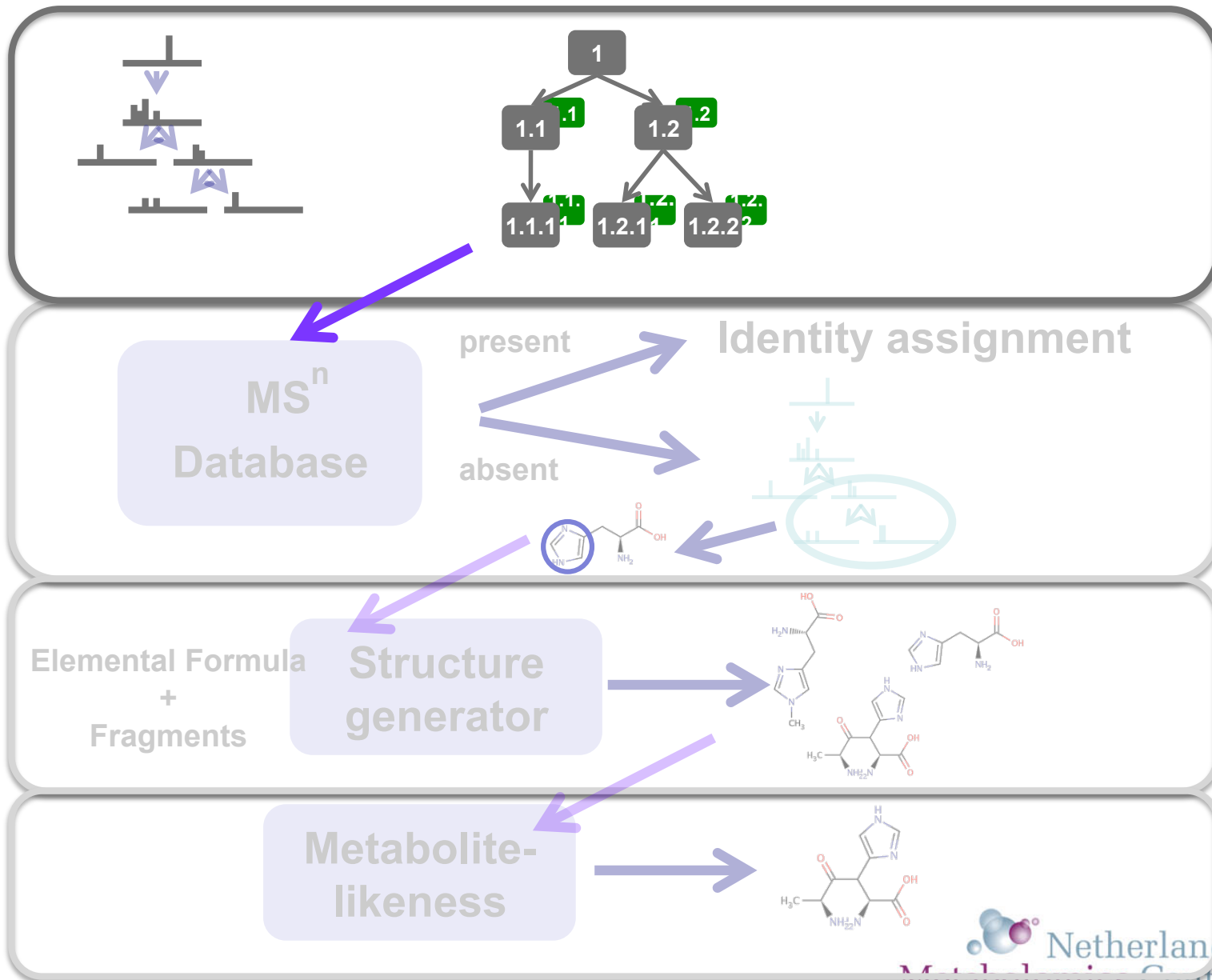
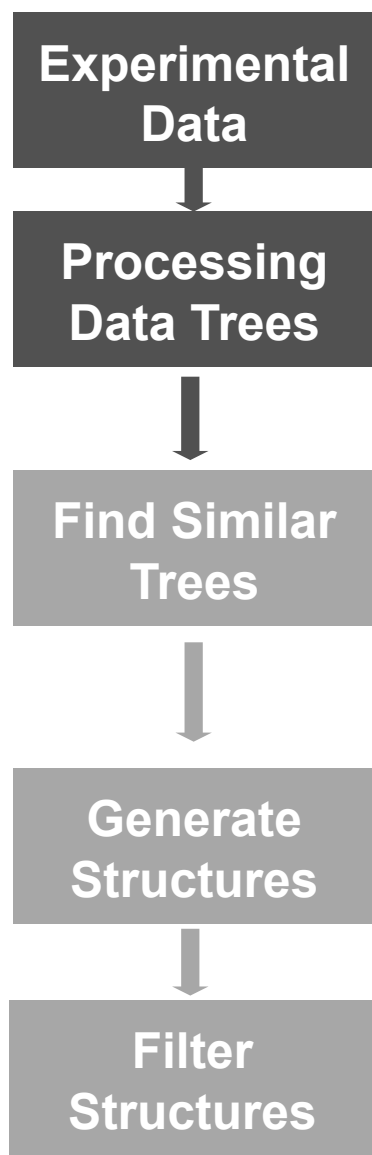
Challenges

Analytical methods

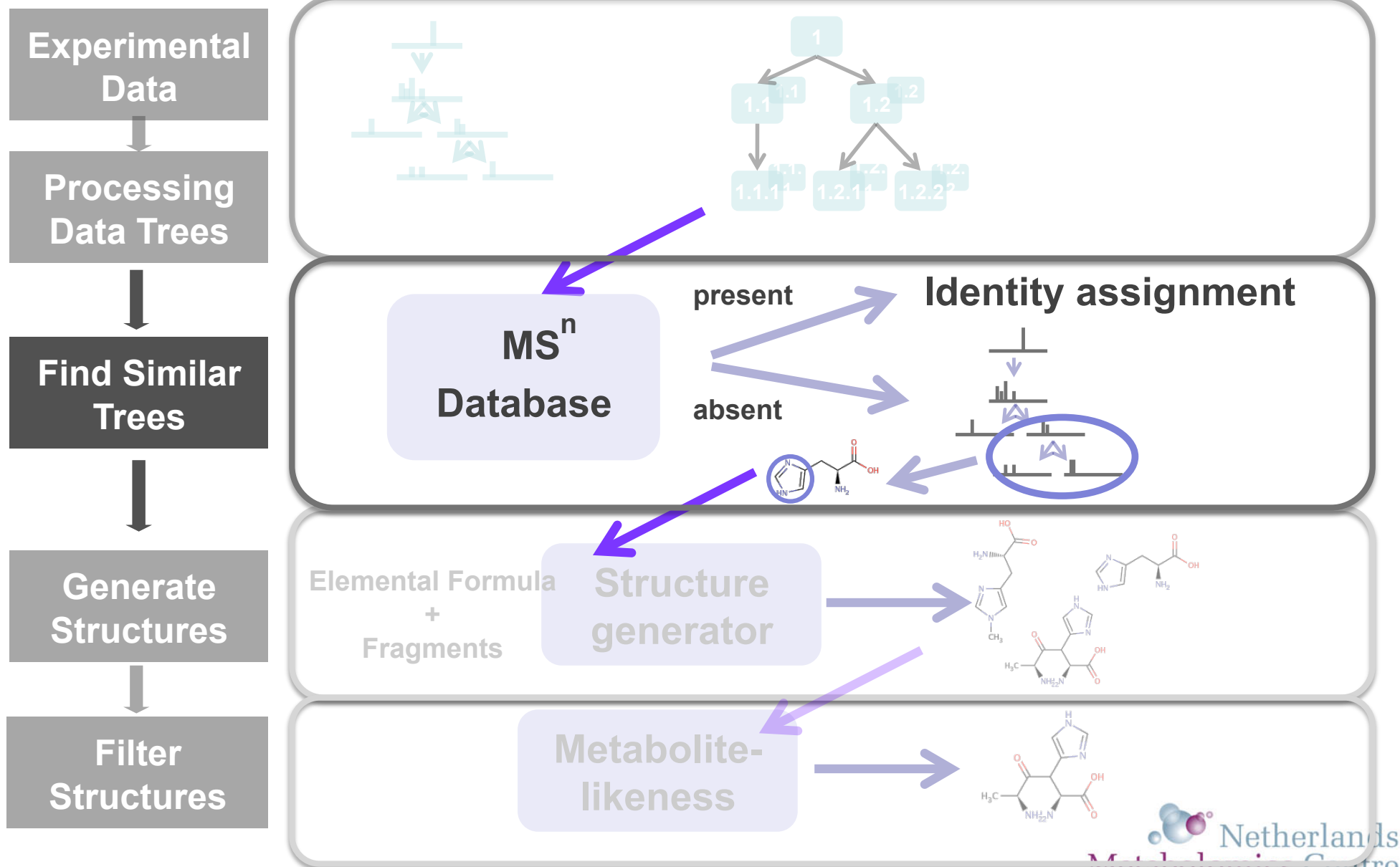
Software

Databases

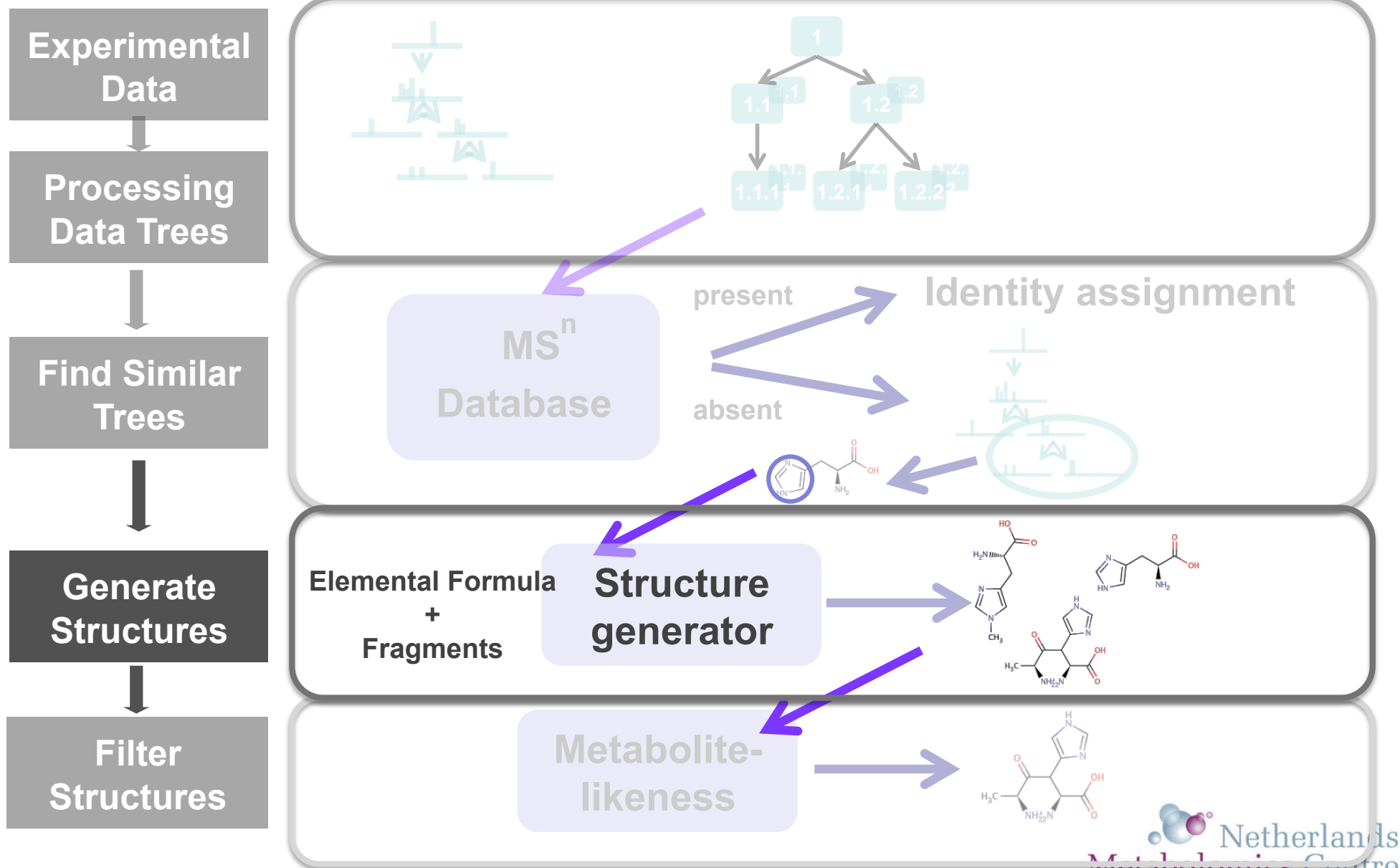
Our approach



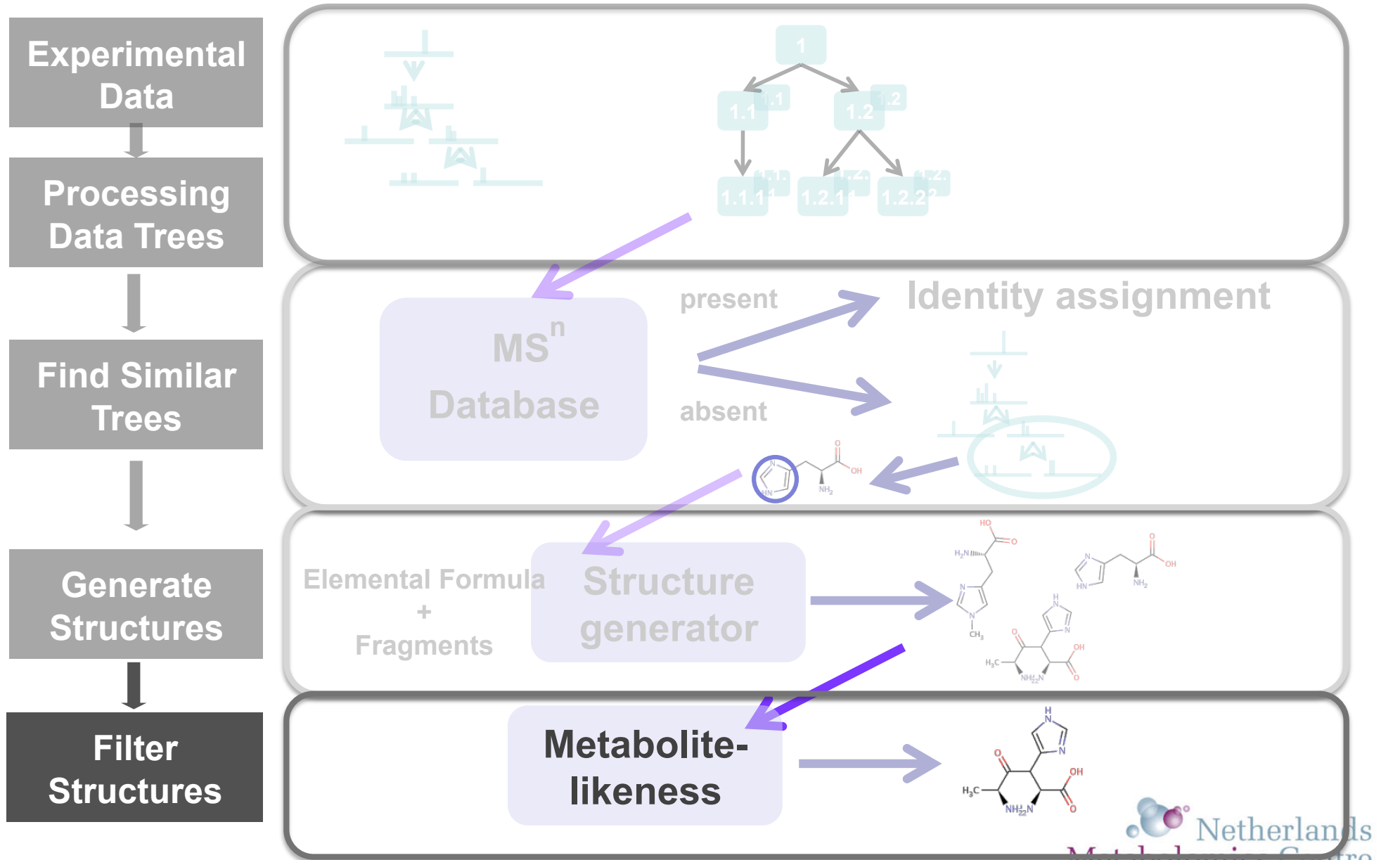
Our approach



Our approach

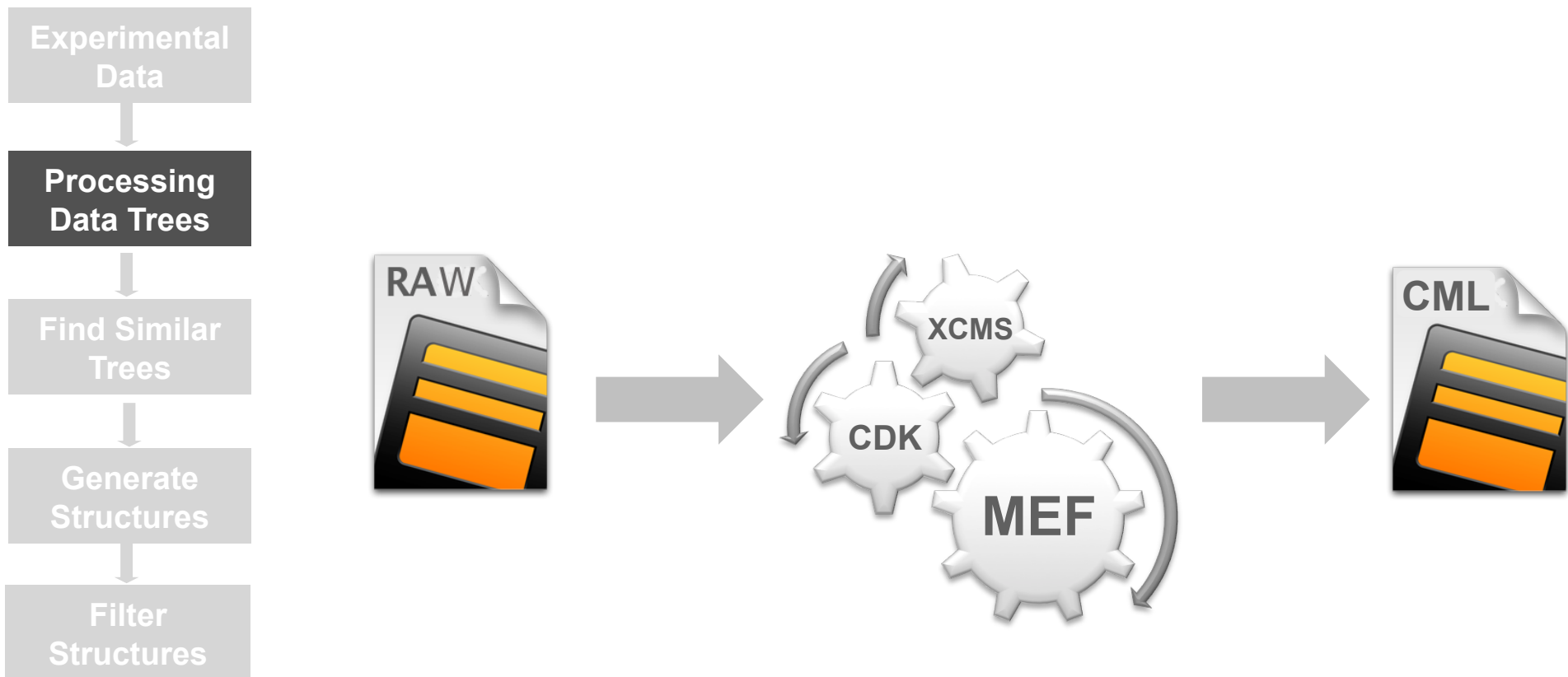


Our approach



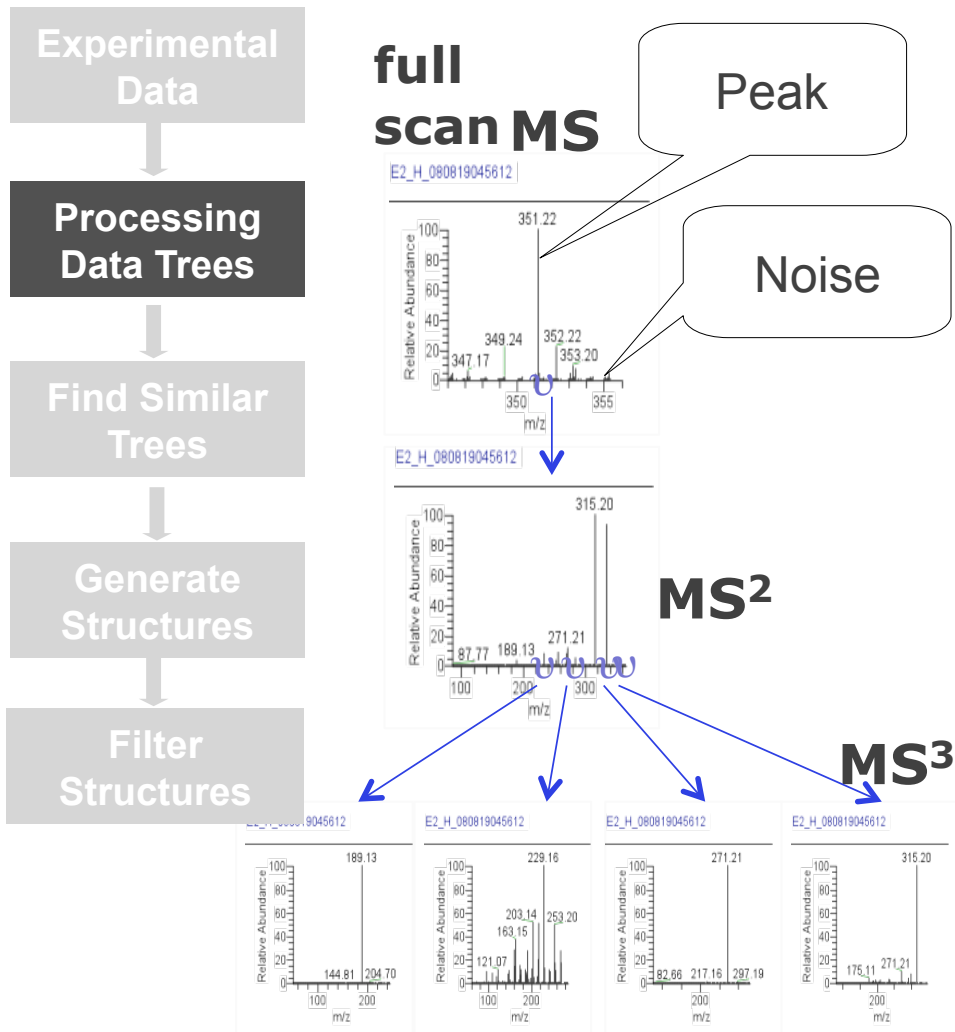
From data to information

MEF: spectral to fragmentation tree

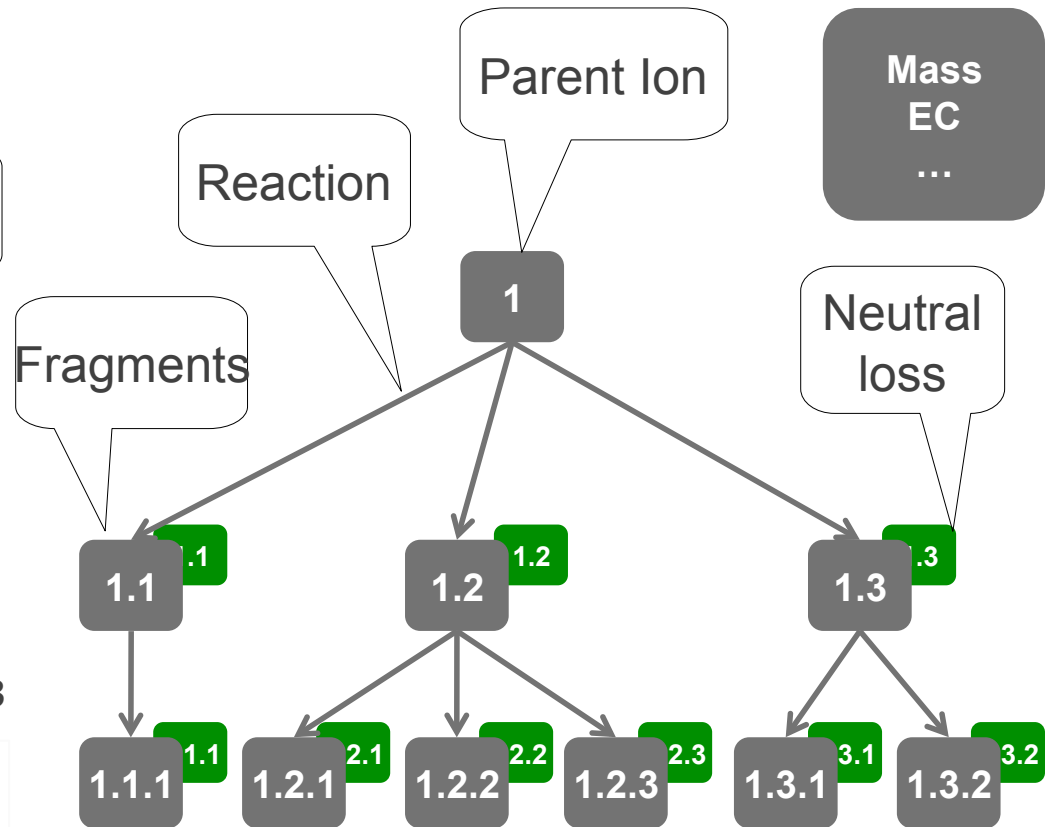


Rojas-Cherto et al. *Bioinformatics* (submitted)

MEF: spectral to fragmentation tree



Spectral Tree



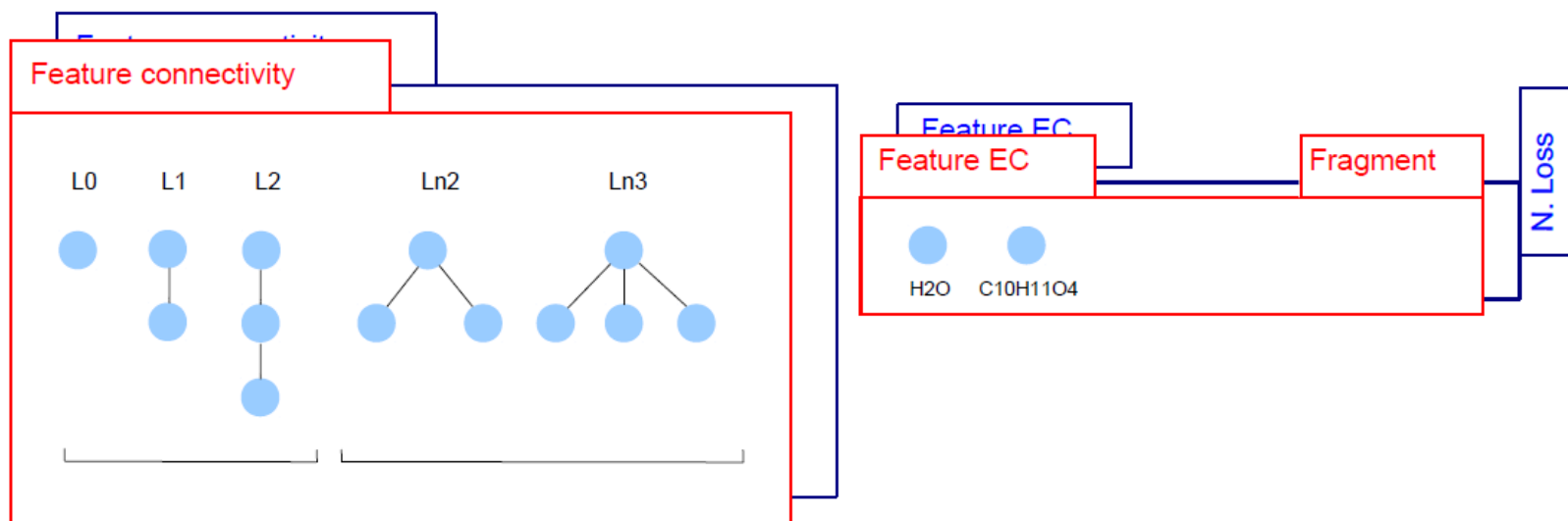
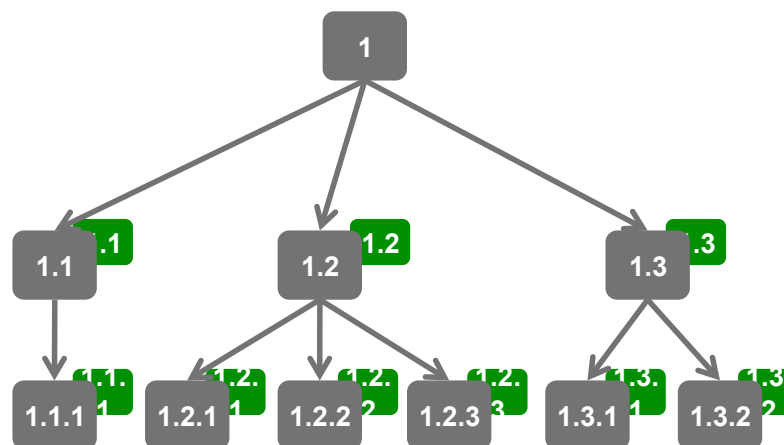
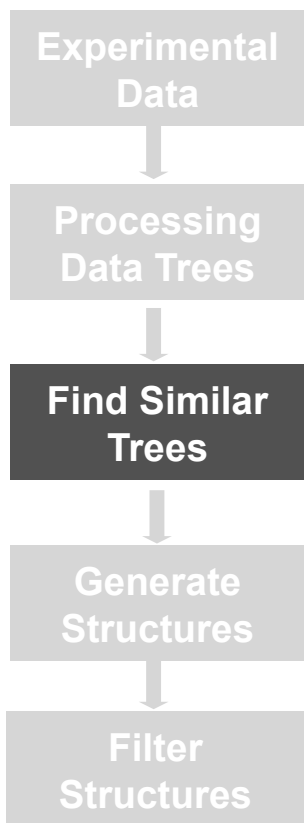
Fragmentation Tree

Rojas-Cherto et al. *Bioinformatics* (submitted)



**Extracting more information from
your data**

Fragmentation tree fingerprints



Poster 59

Miguel Rojas-Cherto, Leiden University



June 9th 2011
ICCS 2011

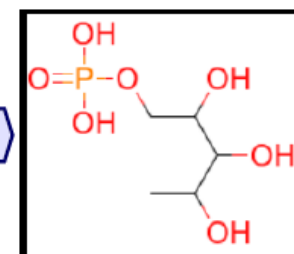
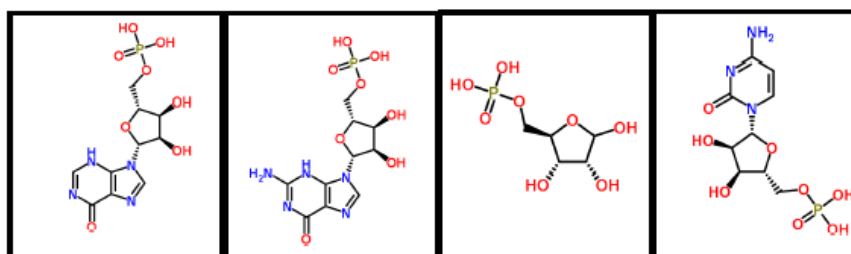
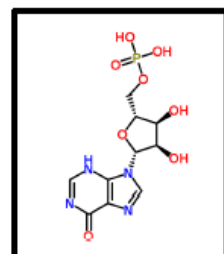
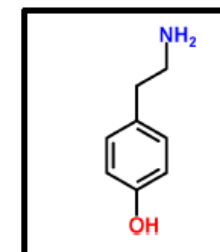
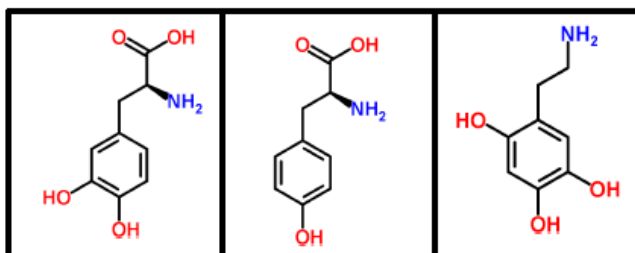
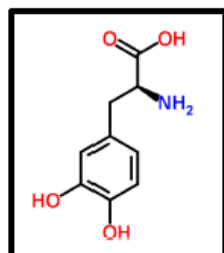
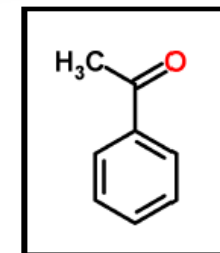
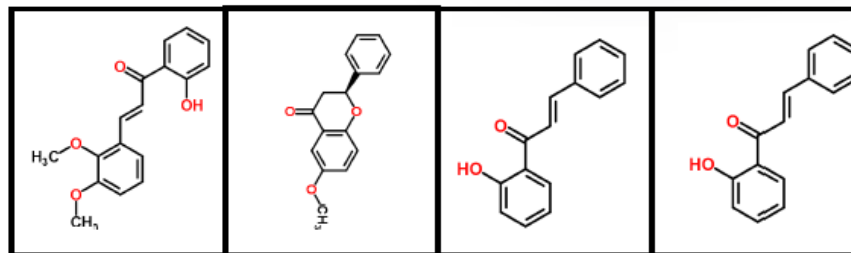
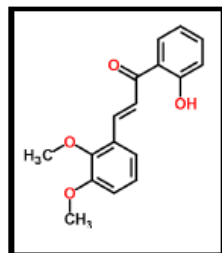
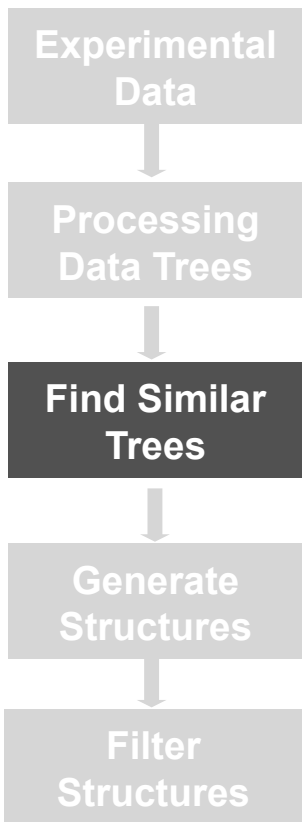
Julio E. Peironcely

Fragmentation tree fingerprints results

“unknown”

Most similar trees

MCS



Poster 59

Miguel Rojas-Cherto, Leiden University

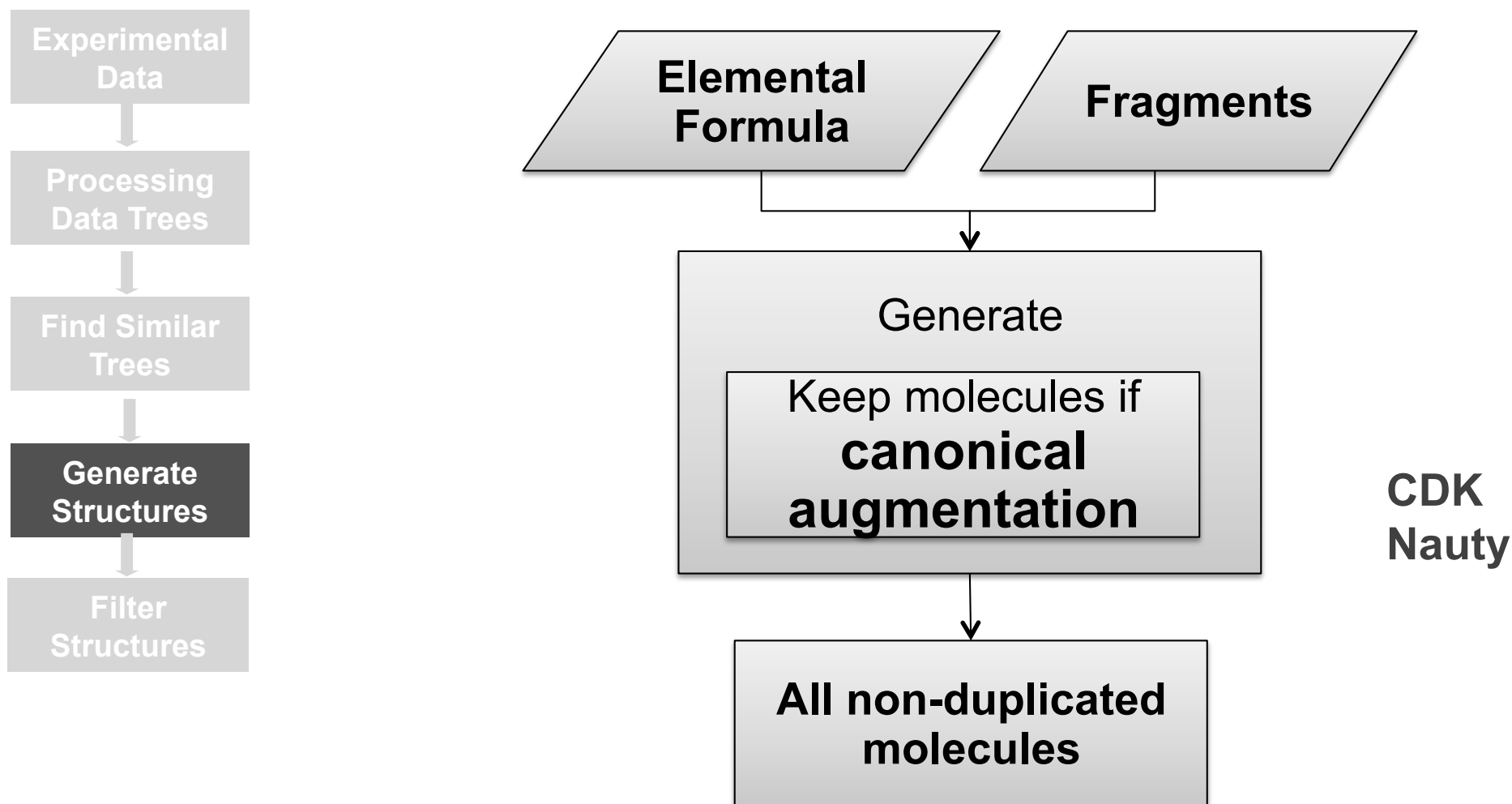
Netherlands
Metabolomics Centre

June 9th 2011
ICCS 2011

Julio E. Peironcely

De-novo identification

Structure Generator



CDK
Nauty

Poster 38

In collaboration with Jean-Loup Faulon, Evry University

 Netherlands
Metabolomics Centre

June 9th 2011
ICCS 2011 Julio E. Peironcely

Structure Generator Results

MOLGEN
same # of
molecules

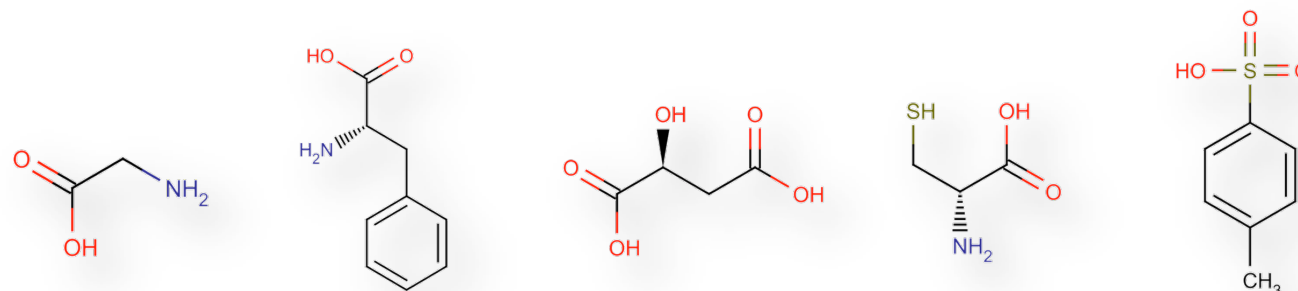
Experimental
Data



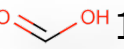

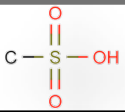




Processing
Data Trees

Find Similar
Trees

Generate
Structures

Filter
Structures



	Glycine	Phenylalanine	Malic acid	D-Cysteine	p-Cresol sulfate
Elemental Composition	C2H5NO2	C9H11NO2	C4H6O5	C3H7NO2S	C7H8O3S
# Output Molecules	84	277,810,163	8,070	3,838	10,203,389
1 Fragment	 6	 4,037,499	 1,601	 100	 19,940
2 Fragments		 93,137			 948
3 Fragments		 584			
Poster 38		 278			

In collaboration with Jean-Loup Faulon, Evry University

 Netherlands
Metabolomics Centre

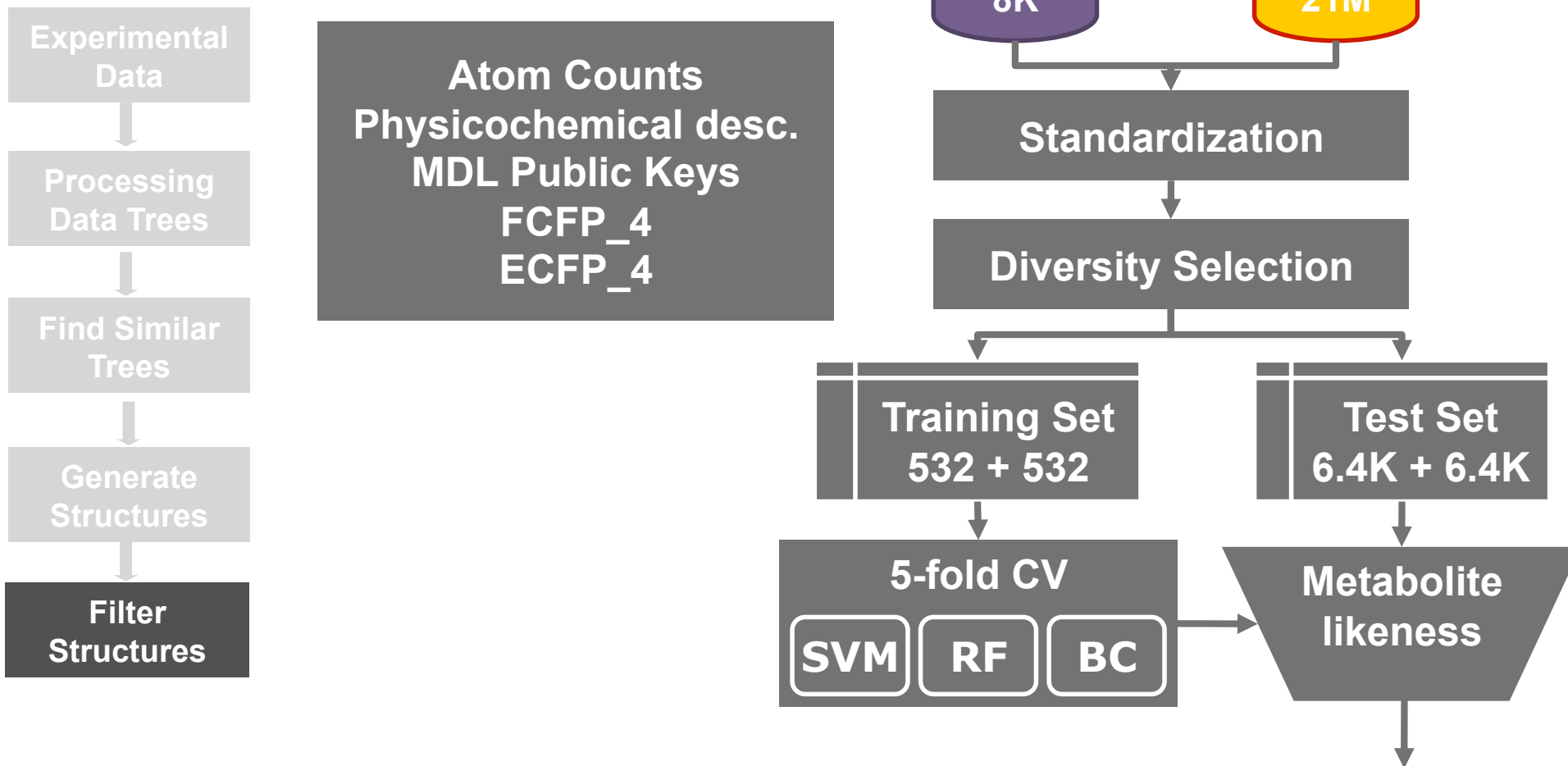
June 9th 2011

ICCS 2011

Julio E. Peironcely

Lots of candidates structures

Metabolite-likeness

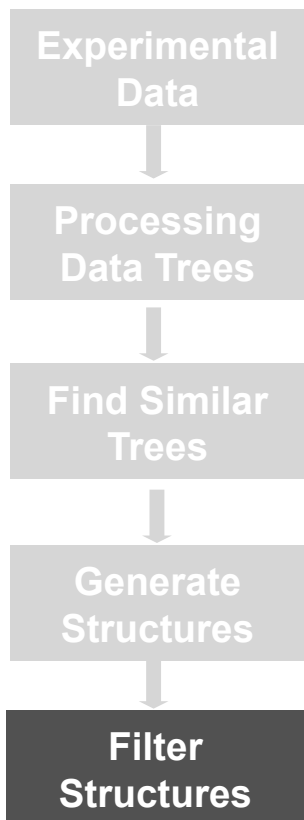


In collaboration with Andreas Bender, Cambridge Univ.

 Netherlands
Metabolomics Centre

June 9th 2011
ICCS 2011 Julio E. Peironcely

Metabolite-likeness



Atom Counts
 Physicochemical desc.
 MDL Public Keys
 FCFP_4
 ECFP_4



Standardization

Diversity Selection

Training Set
 532 + 532

Test Set
 6.4K + 6.4K



Metabolite
 likeness

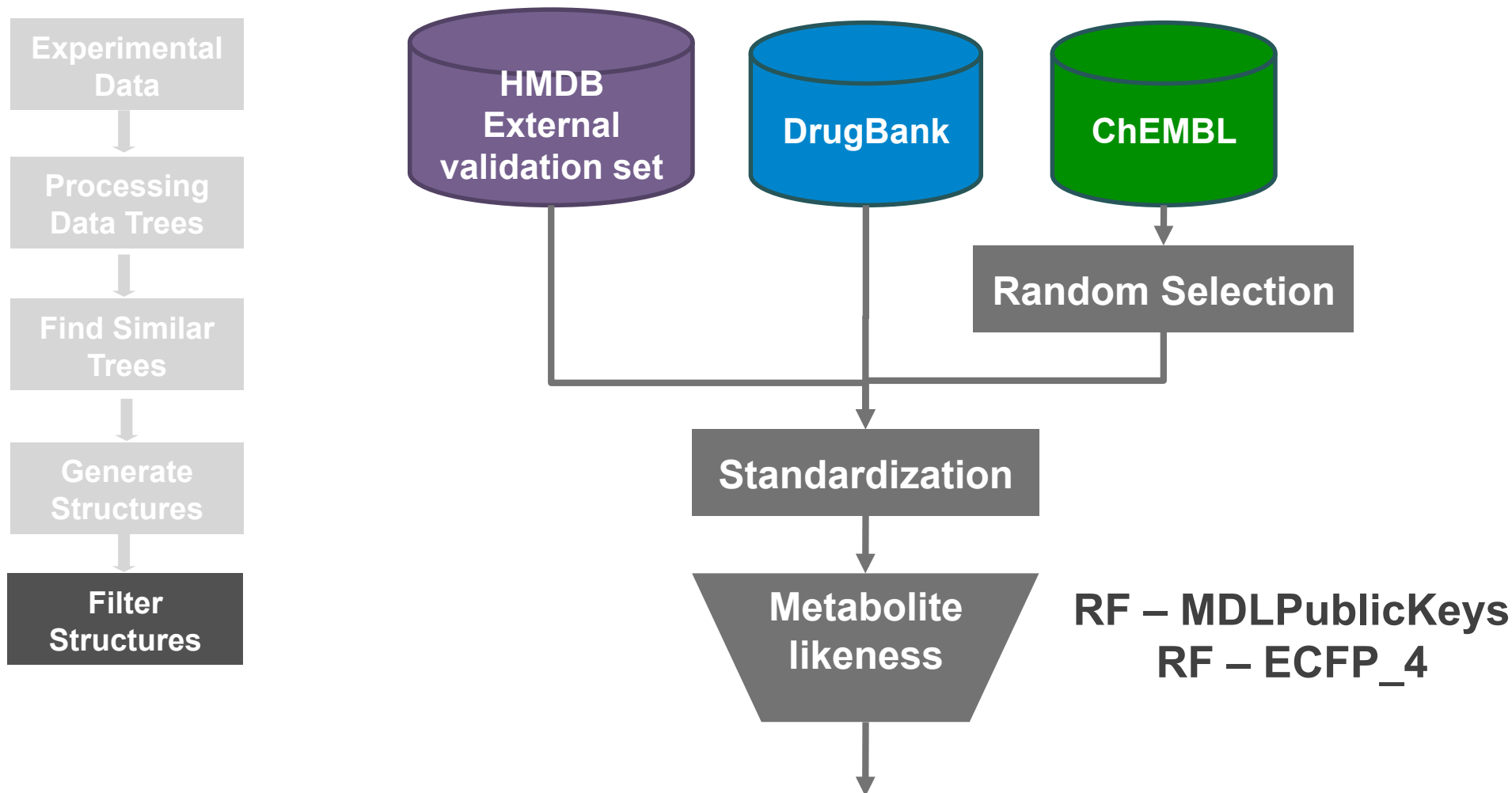
1 st RF – MDLPublicKeys		
Sensitivity	Specificity	AUC
99.84%	87.52%	99.20%

2 nd RF – ECFP_4		
Sensitivity	Specificity	AUC
99.77%	86.36%	99%

In collaboration with Andreas Bender, Cambridge Univ.



Metabolite-likeness, external validation



In collaboration with Andreas Bender, Cambridge Univ.

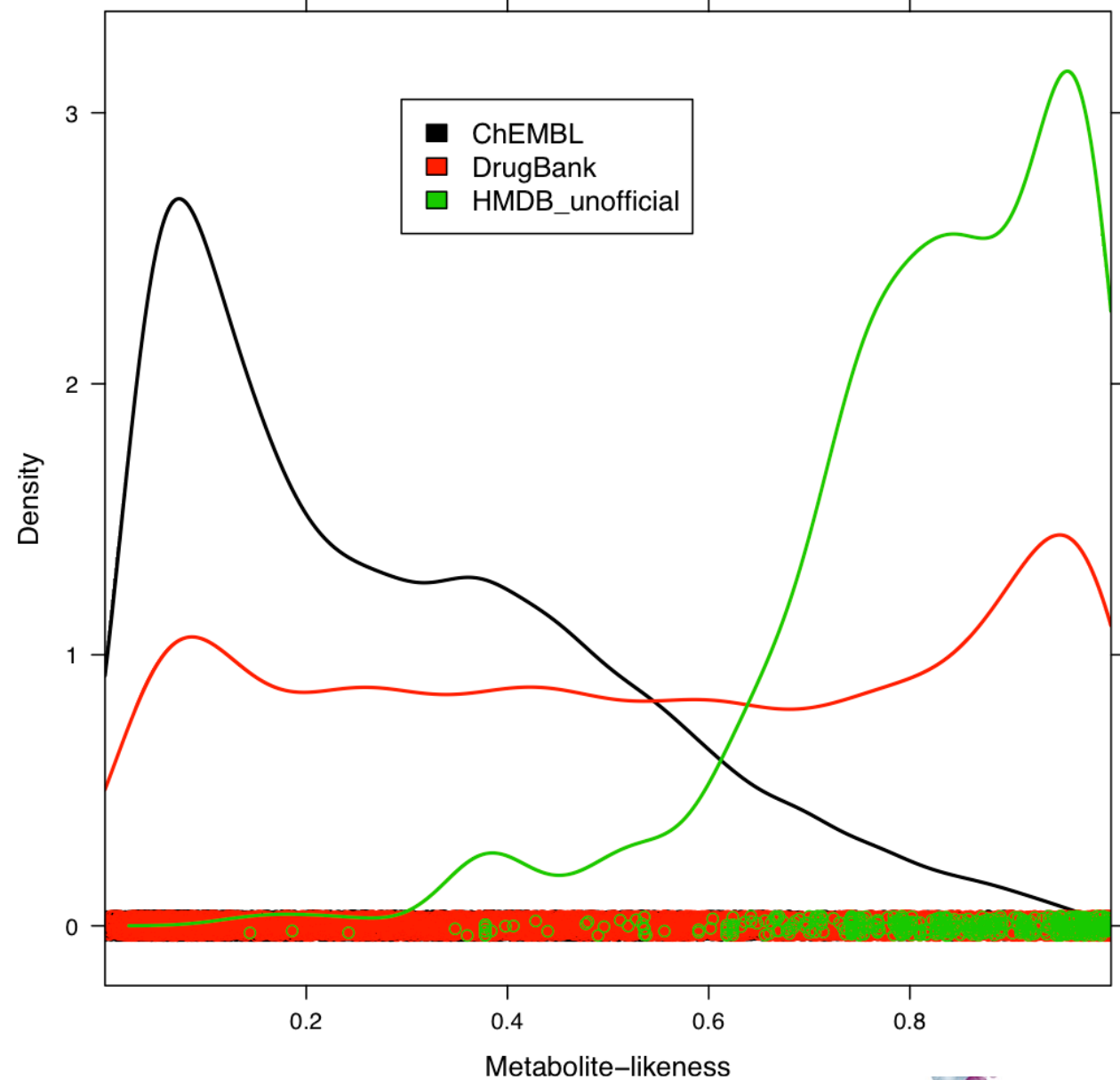
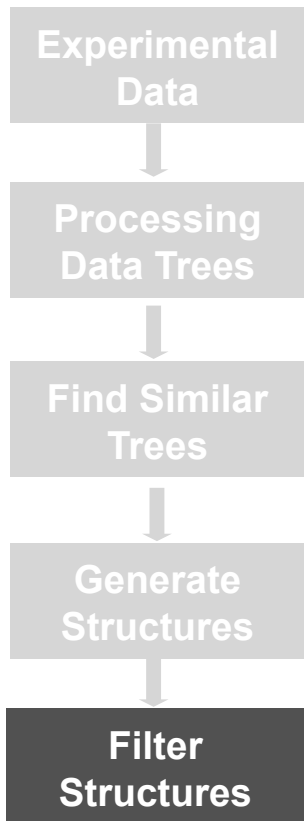
 Netherlands
Metabolomics Centre

June 9th 2011

ICCS 2011

Julio E. Peironcely

Metabolite-likeness, external validation



In collaboration with Andreas Bender, Cambridge Univ.

Netherlands
Metabolomics Centre

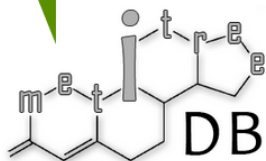
June 9th 2011
ICCS 2011

Julio E. Peironcely

**From information to
knowledge**

www.MetiTree.nl

still in beta. **not** suited for production work.



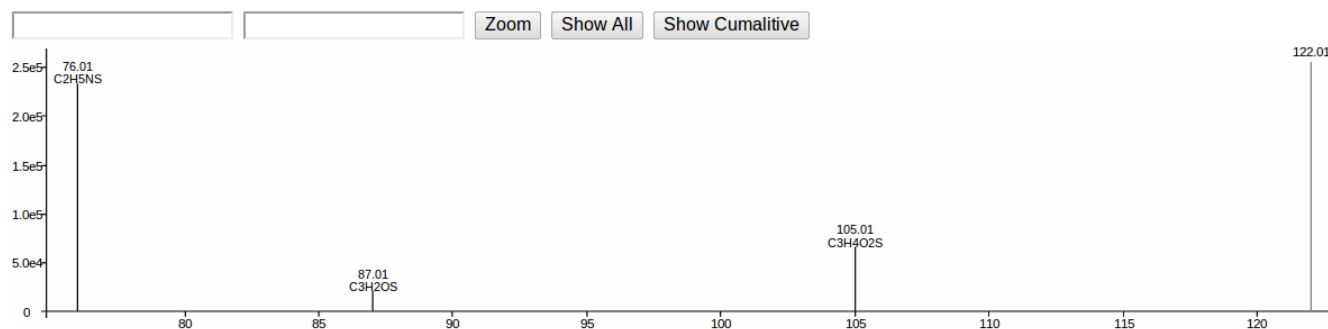
menu

msn data
process
database(s)
query by file
query by text

configuration
usergroup(s)
member(s)
api (test page)
database viewer

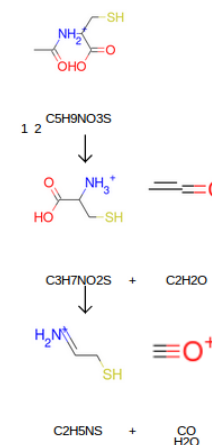
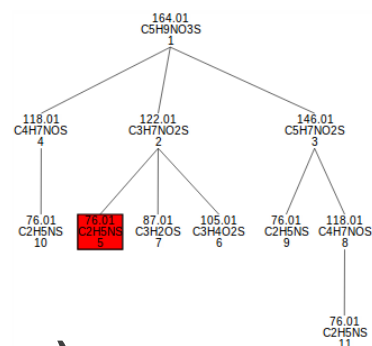
home > viewer > index

<<< BACK: Post JSON formatted Tree Structure



welcome ICCS 2011 :: NMC (Netherlands Metabolomics Centre)

Group Database
Upload and organize your own trees
Visualize trees
Find similar trees
(to be added: Struct. Gen, Met-likeness)



June 9th 2011
ICCS 2011

Julio E. Peironcely

Conclusions

Chemoinformatics plays a crucial role in the metabolite identification pipeline

Now it is the time to challenge this pipeline with real cases

Expert is still needed

Acknowledgements



Leiden University
Miguel Rojas-Cherto

Piotr Kasper
Michael van Vliet
Theo Reijmers
Rob Vreeken
Ronnie van Doorn
Thomas Hankemeier



**Wageningen University/
PRI**

Egon Willighangen
Justin van der Hooff
Ric de Vos
Jacques Vervoort

TNO Quality of Life

Leon Coulier
Albert Tas



University of Cambridge
Andreas Bender

Evry University
Jean-Loup Faulon
Davide Fichera



**HMP University of
Alberta**

David Wishart
Ying (Edison) Dong



 **Netherlands
Metabolomics Centre**