Machine learning tools can pinpoint high-risk water pollutants

Helen Sepman
Helen.Sepman@aces.su.se

Kruve Lab

Stockholm University

Introduction

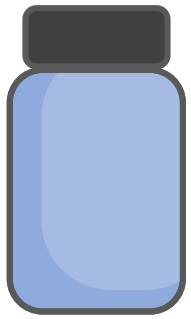
Introduction

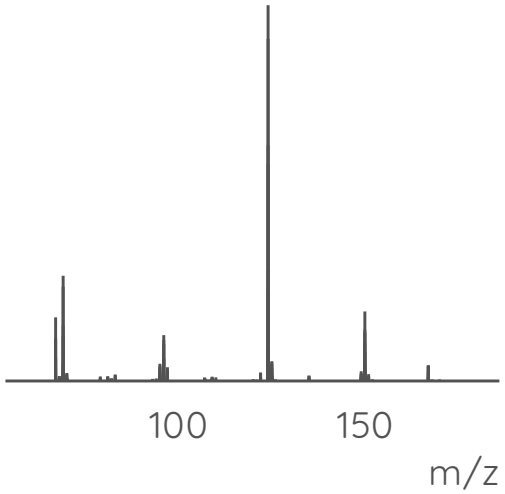
Research questions

How to pinpoint high risk chemicals?

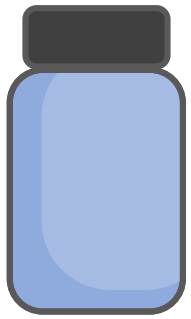
How to acquire high quality spectra for high risk chemicals?

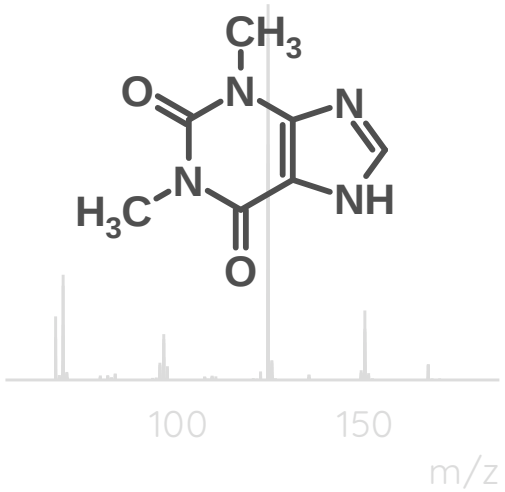
Assessing the risk of a detected chemical





Assessing the risk of a detected chemical

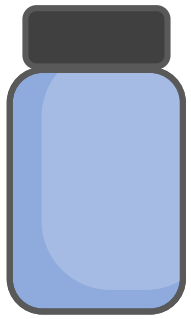


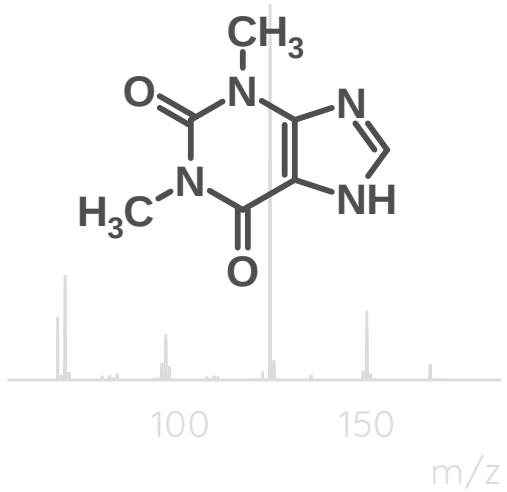


How much of this chemical is present in the environment?

How much of this chemical would start to cause adverse outcomes?

Assessing the risk of a detected chemical





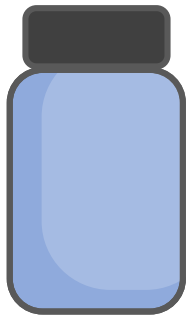
$C_{environmental} >> C_{toxic}$

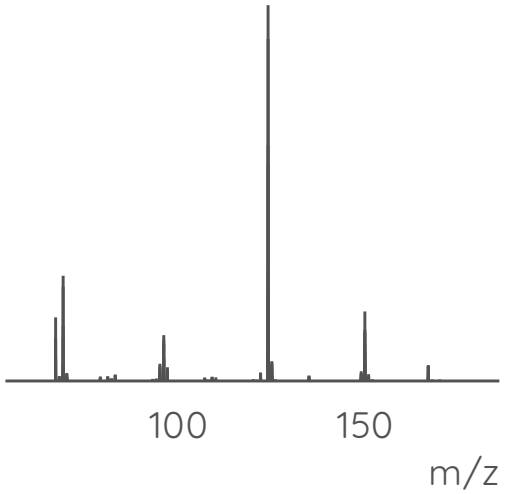
$C_{environmental} << C_{toxic}$

How much of this chemical is present in the environment?

How much of this chemical would start to cause adverse outcomes?

Assessing the risk of a detected chemical





$C_{environmental} >> C_{toxic}$

$C_{environmental} << C_{toxic}$

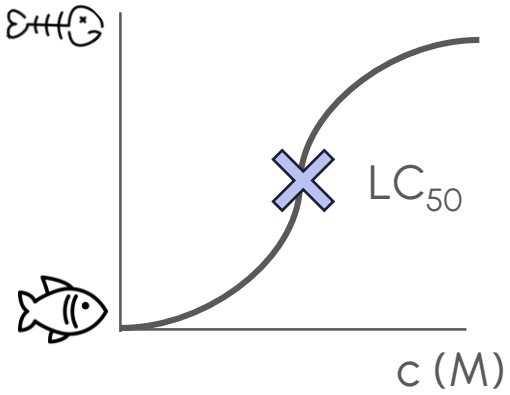
How much of this chemical is present in the environment?

How much of this chemical would start to cause adverse outcomes?

Machine learning tools for risk estimation

Machine learning tools for risk estimation

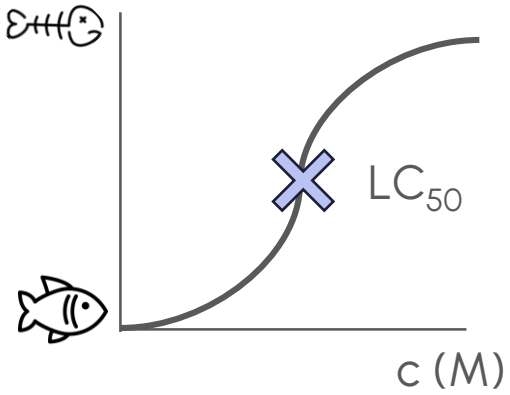
Lethal concentration, 50% (LC_{50})



Machine learning tools for risk estimation

MS2Tox

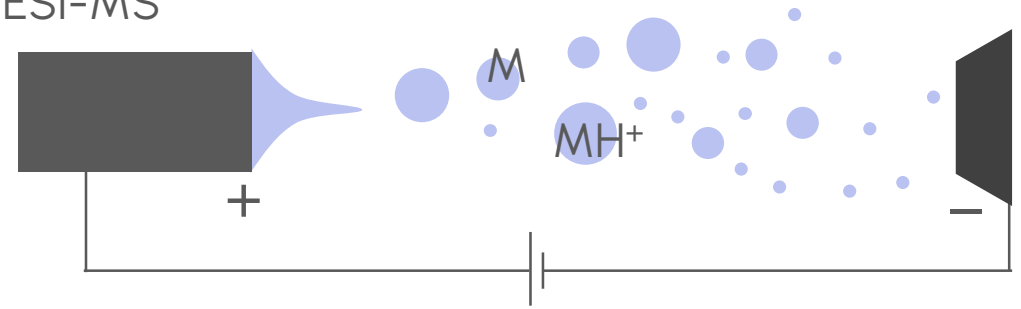
Lethal concentration, 50% (LC_{50})



MS2Quant

Ionization efficiency (IE)

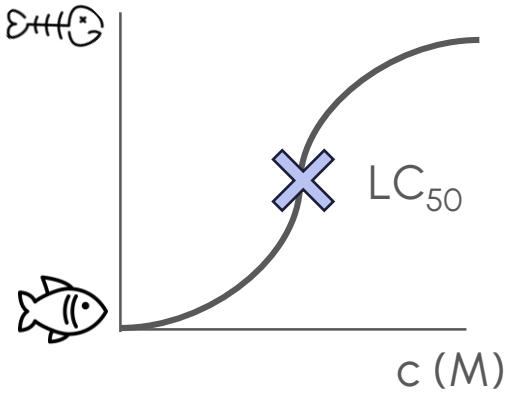
LC-ESI-MS



Machine learning tools for risk estimation

MS2Tox

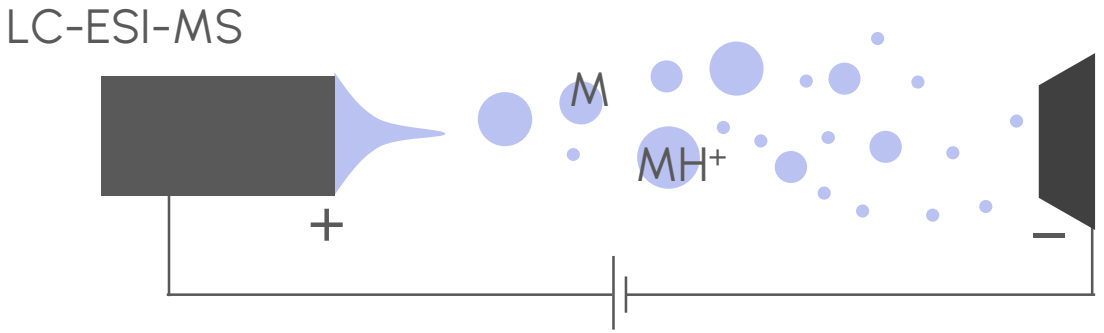
Lethal concentration, 50% (LC_{50})



$c_{toxic} \sim LC_{50}$

MS2Quant

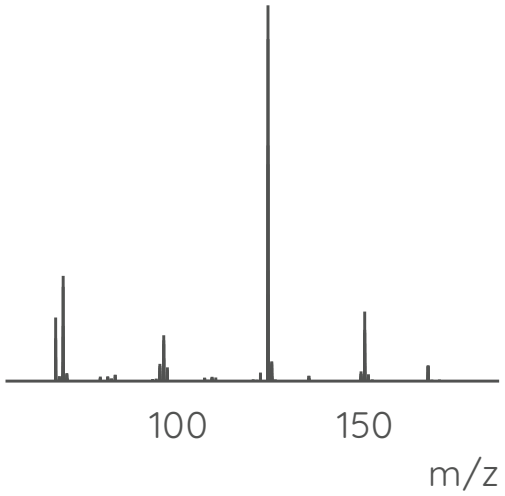
Ionization efficiency (IE)



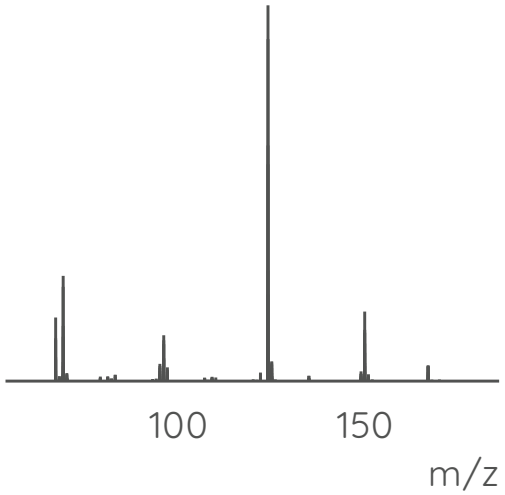
$c_{environmental} \sim \frac{Intensity}{IE}$

$$PriorityScore = \frac{\frac{Intensity}{IE}}{LC_{50}}$$

Machine learning tools for risk estimation



Machine learning tools for risk estimation

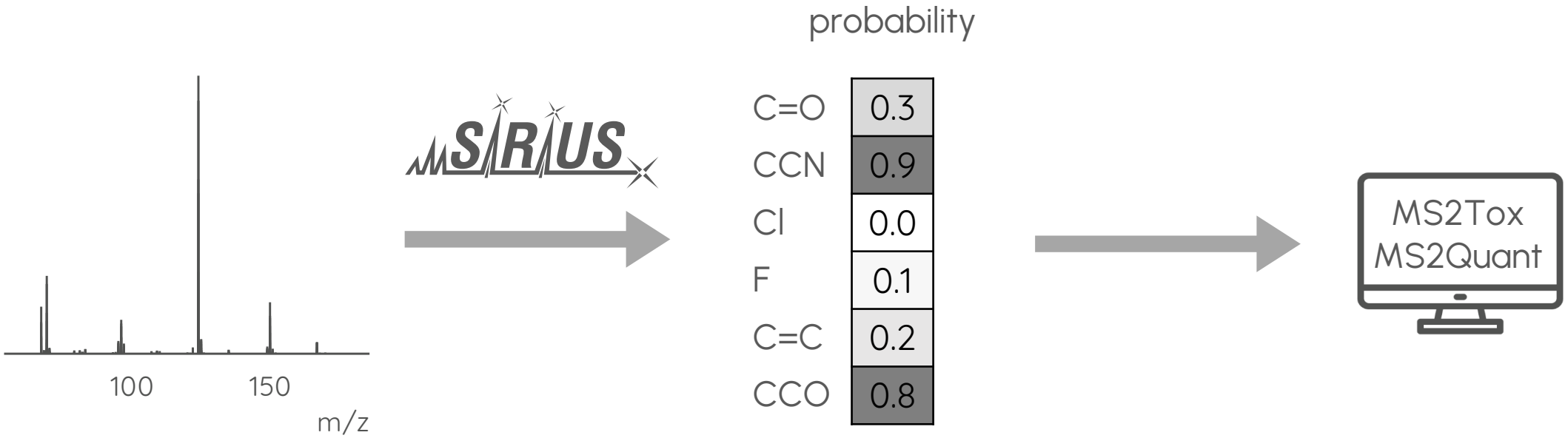


probability

C=O	0.3
CCN	0.9
Cl	0.0
F	0.1
C=C	0.2
CCO	0.8

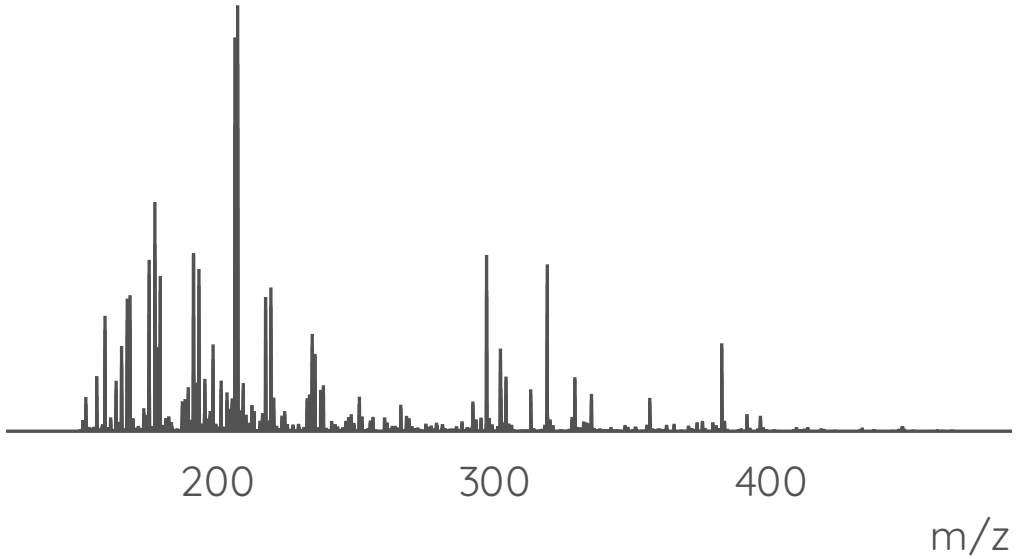
Dührkop et al., Nat. Methods, 2019, 16, 299–302

Machine learning tools for risk estimation



Dührkop et al., Nat. Methods, 2019, 16, 299–302

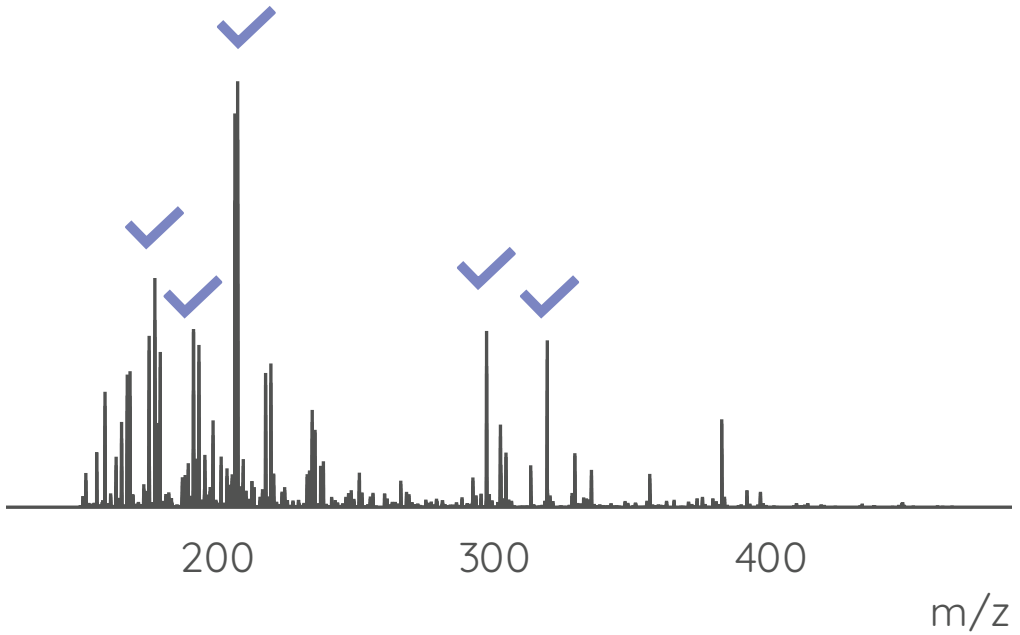
Data aqcuisition approaches



Data aqcuisition approaches

Top5

Top 5 highest
Intenstiy peaks



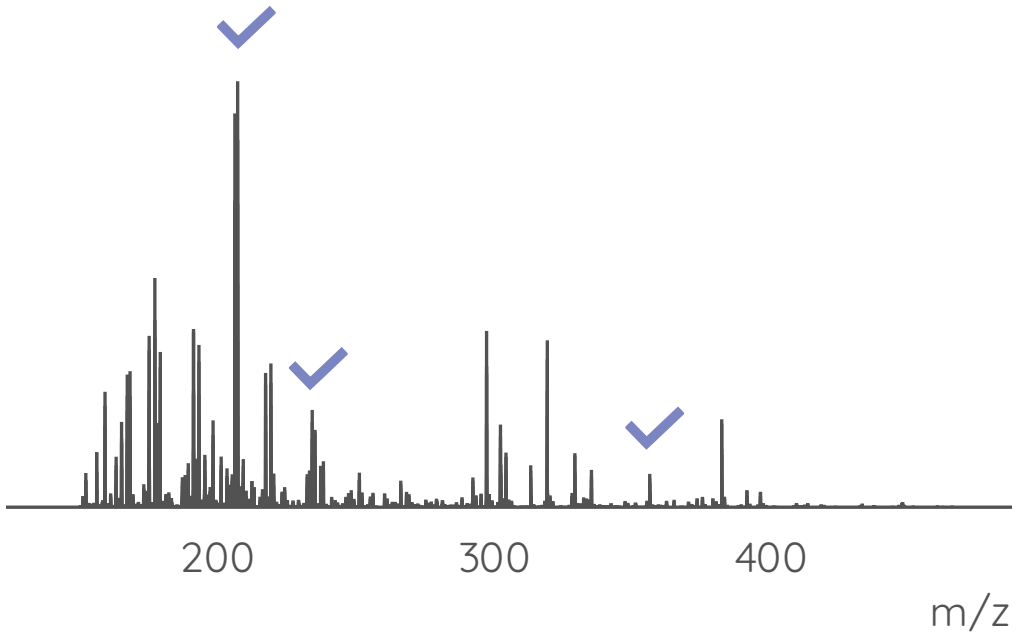
Data aqcuisition approaches

Top5

SWS

Surface water suspect list as inclusion list

523 compound masses



Tian et al., Environ. Sci. Technol. Lett., 2023, 10, 40−45.

Data aqcuisition approaches

Top5

SWS

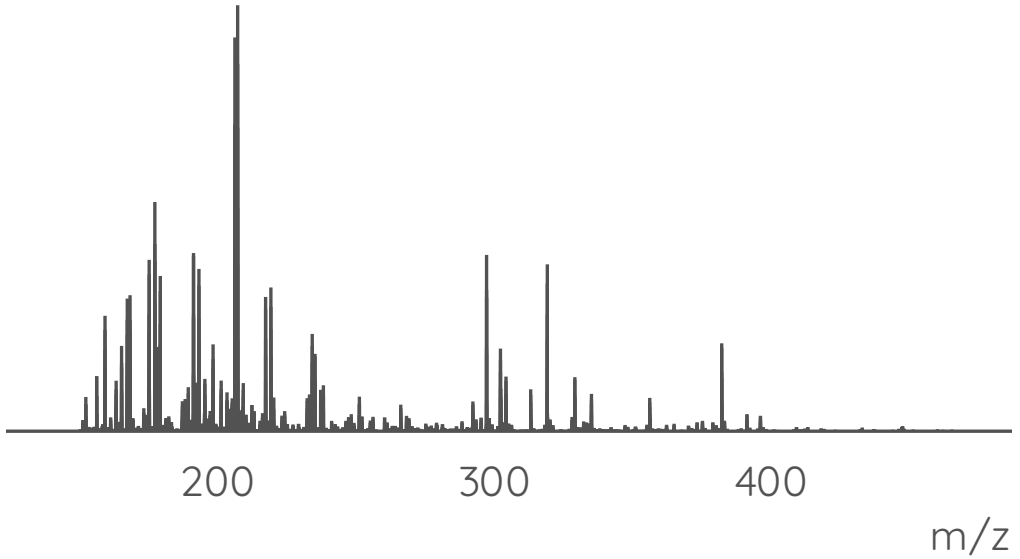
PCL1000

1000 high risk chemicals from PubChemLite

PubChemLite

SMILES	*m/z*
...	...
...	...
...	...
...	...
...	...

~450 K chemicals



Data aqcuisition approaches

Top5

SWS

PCL1000

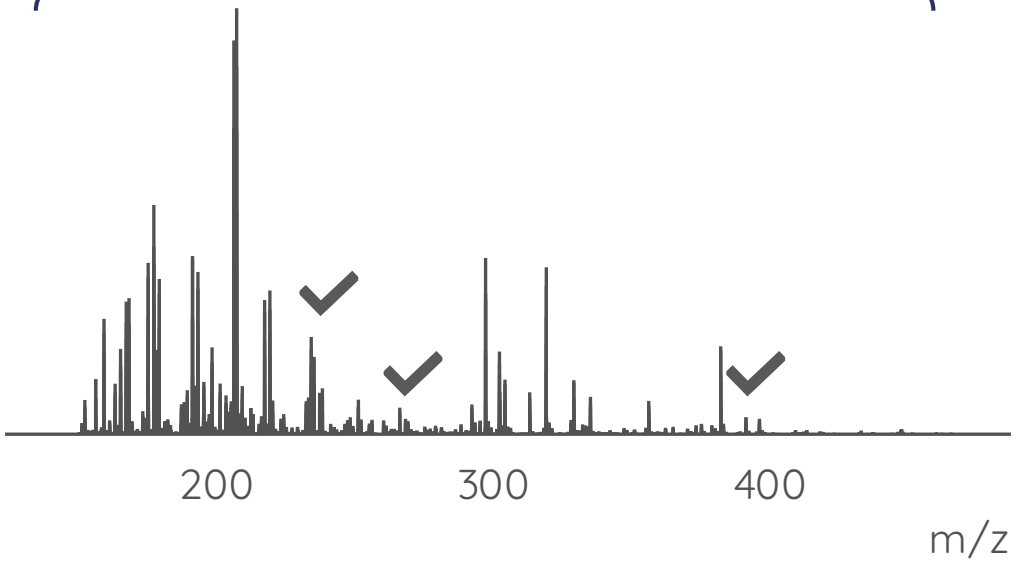
1000 high risk chemicals from PubChemLite

PubChemLite

SMILES	*m/z*
...	...
...	...
...	...
...	...
...	...

~450 K chemicals

Match



Data aqcuisition approaches

Top5

SWS

PCL1000

1000 high risk chemicals from PubChemLite

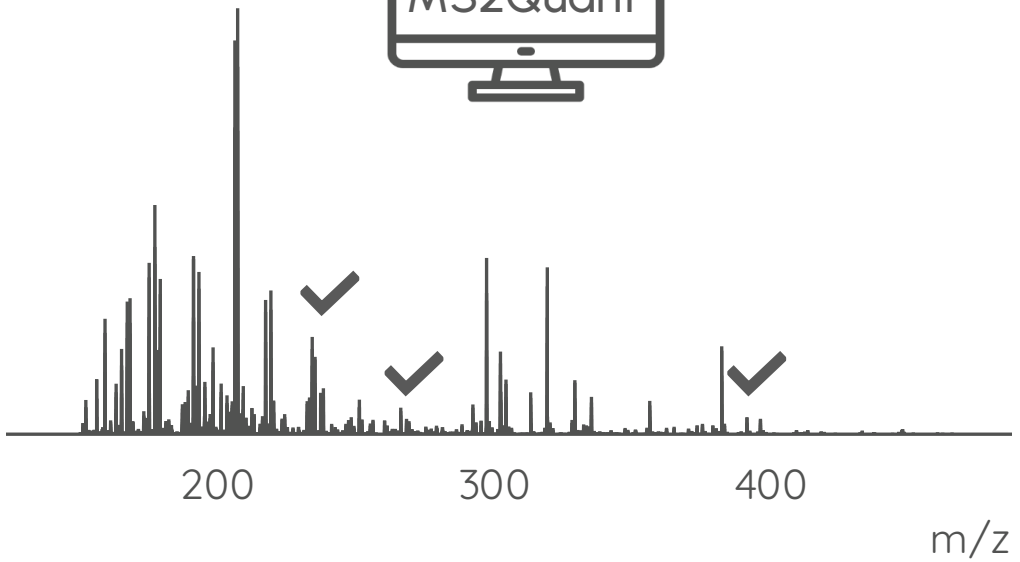
PubChemLite

SMILES	*m/z*
...	...
...	...
...	...
...	...
...	...

~450 K chemicals

Match	IE	LC_{50}
✓	...	...
✗		
✓	...	...
✓	...	...
✗		

MS2Tox
MS2Quant



Schymanski et al., J. Cheminformatics, 2021, 13, 19

Data aqcuisition approaches

Top5

SWS

PCL1000

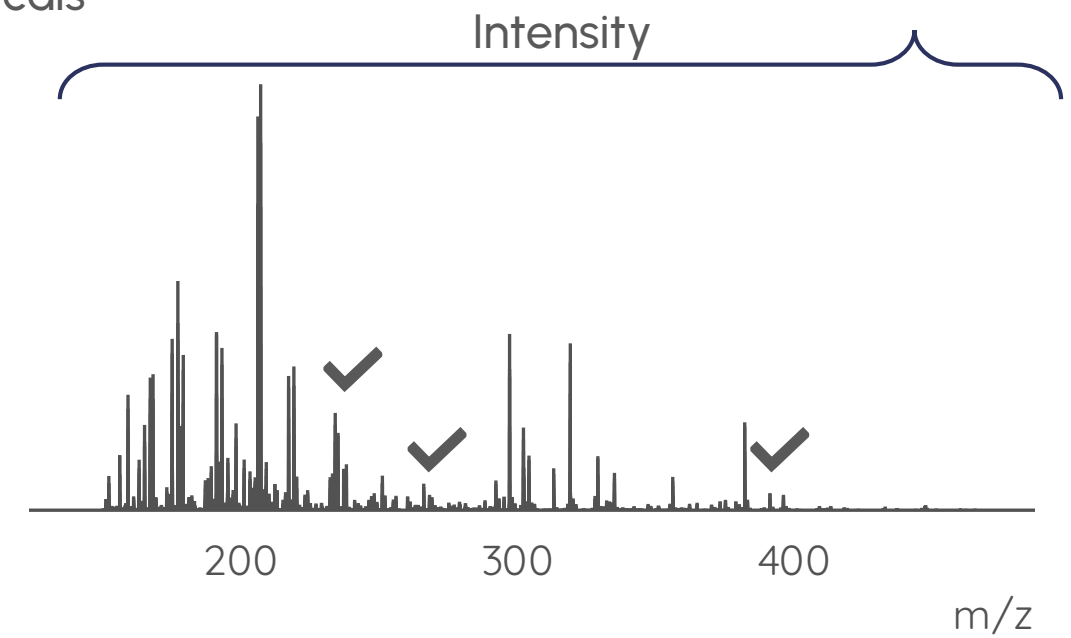
1000 high risk chemicals from PubChemLite

PubChemLite

SMILES	*m/z*	Match	IE	LC_{50}
...	...	✓	...	...
...	...	✗		
...	...	✓	...	...
...	...	✓		
...	...	✗	...	...

~450 K chemicals

$$\text{PriorityScore} = \frac{\frac{\text{Intensity}}{\text{IE}}}{LC_{50}}$$



Data aqcuisition approaches

Top5

SWS

PCL1000

1000 high risk chemicals from PubChemLite

PubChemLite

SMILES	*m/z*
...	...
...	...
...	...
...	...
...	...

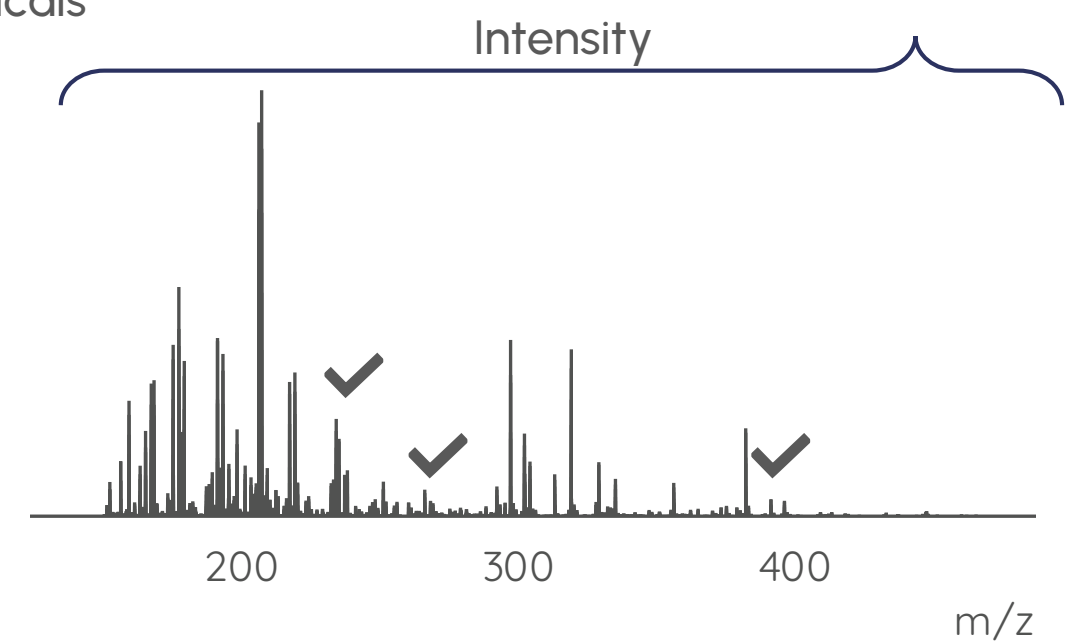
~450 K chemicals

Match

IE ... ...

LC_{50} ... ...

PriorityScore



Data aqcuisition approaches

Top5

SWS

PCL1000

1000 high risk chemicals from PubChemLite

PubChemLite

SMILES	*m/z*
...	...
...	...
...	...
...	...
...	...

~450 K chemicals

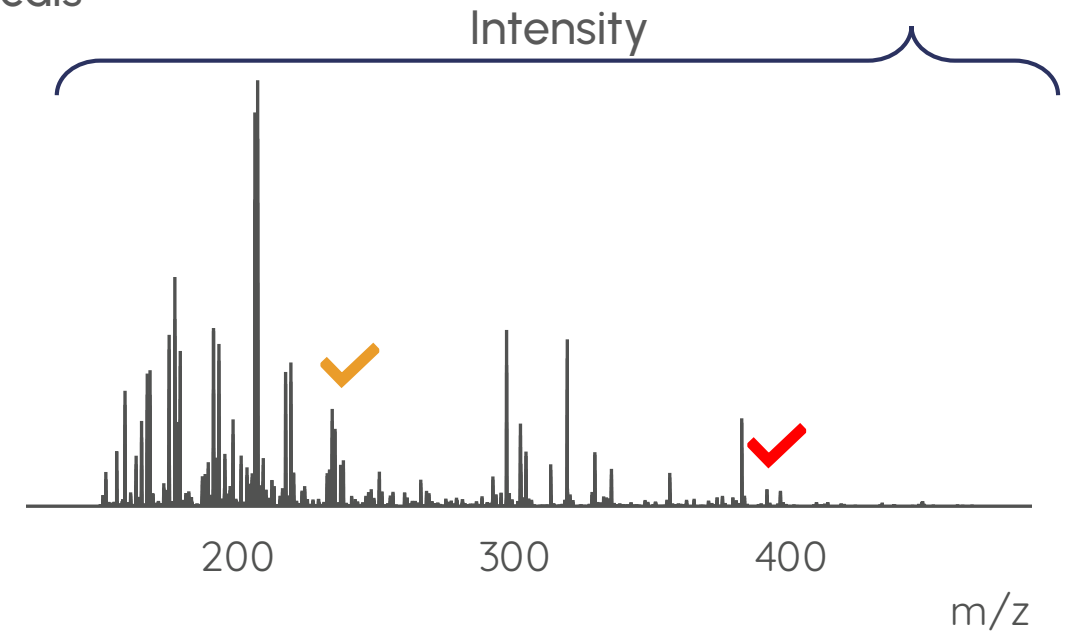
Match

IE

LC_{50}

PriorityScore

1000 chemicals



Data aqcuisition approaches

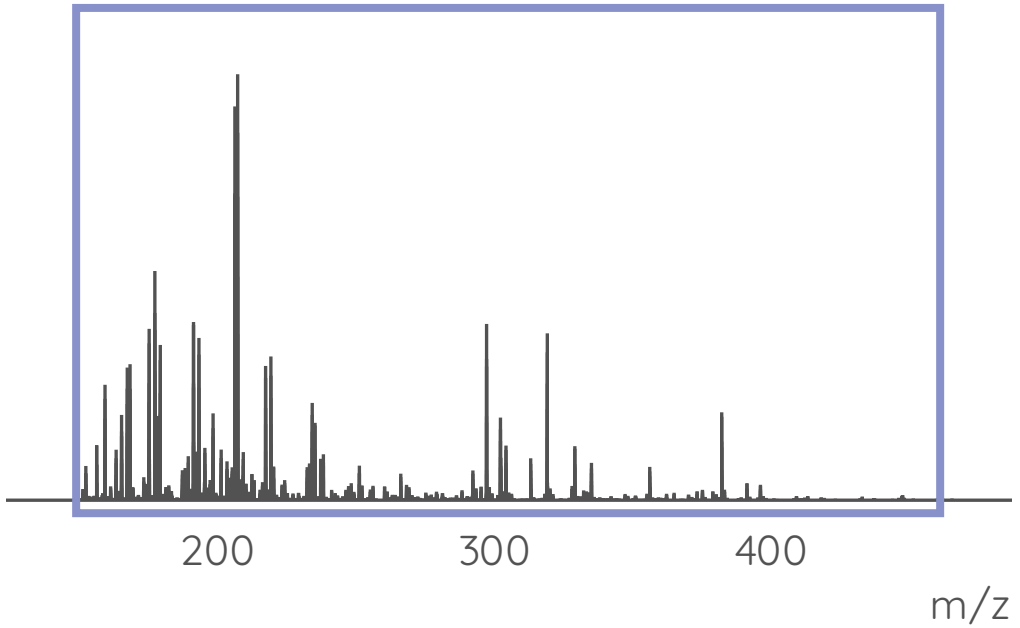
Top5

SWS

PCL1000

DIA

Data independent acqusition

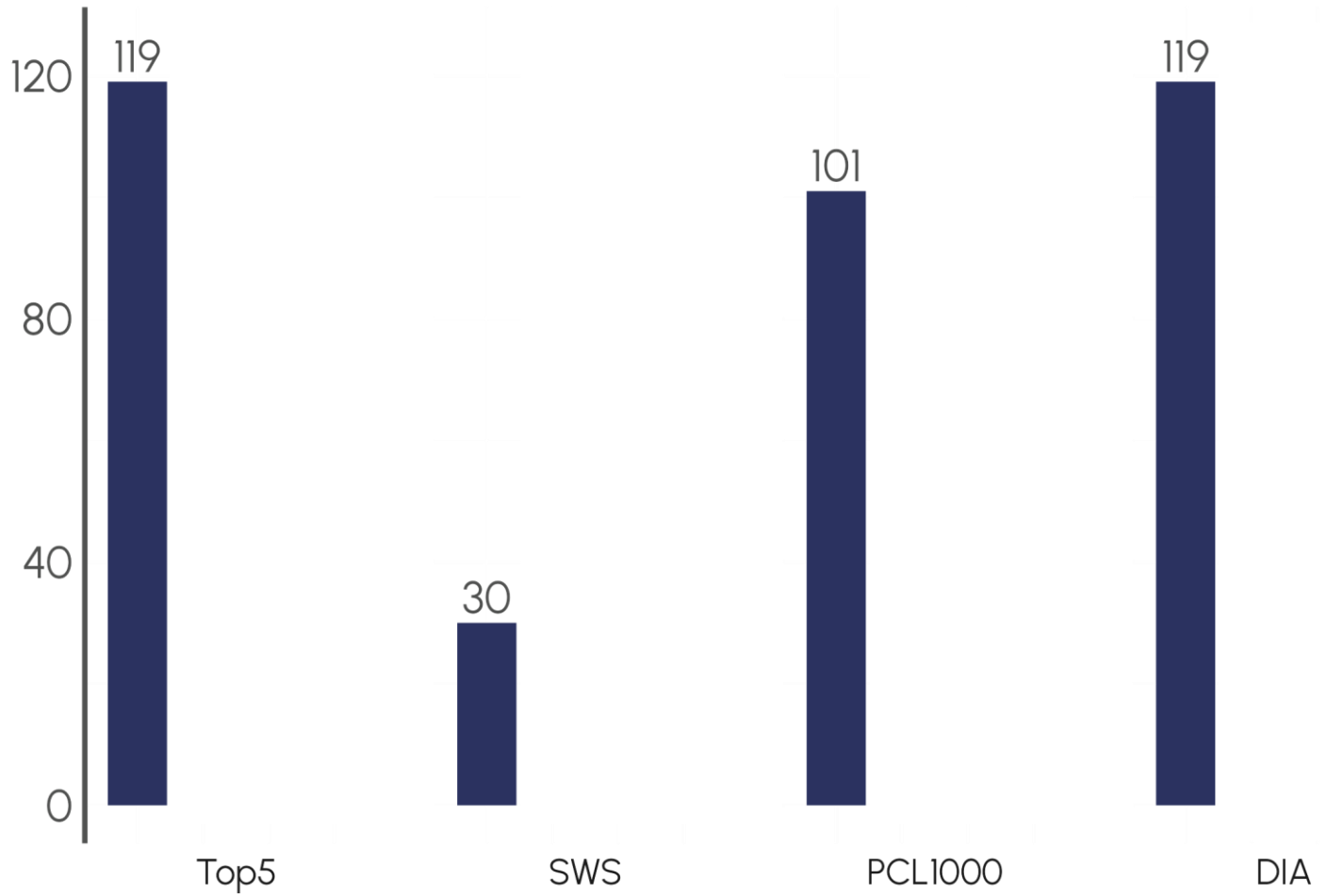


Comparison of methods - standards

Standard mix
119 chemicals

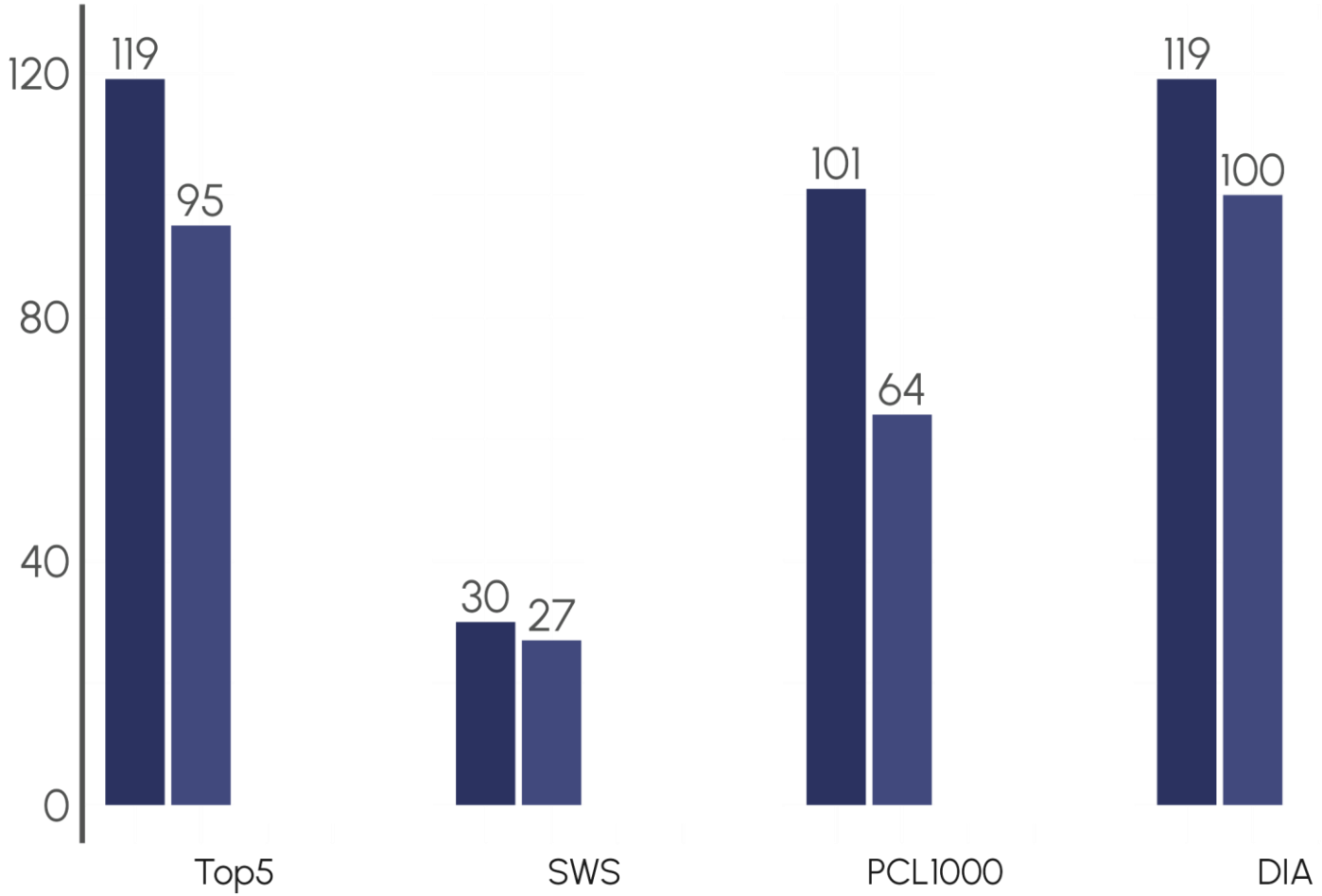
Comparison of methods - standards

Standard mix
119 chemicals



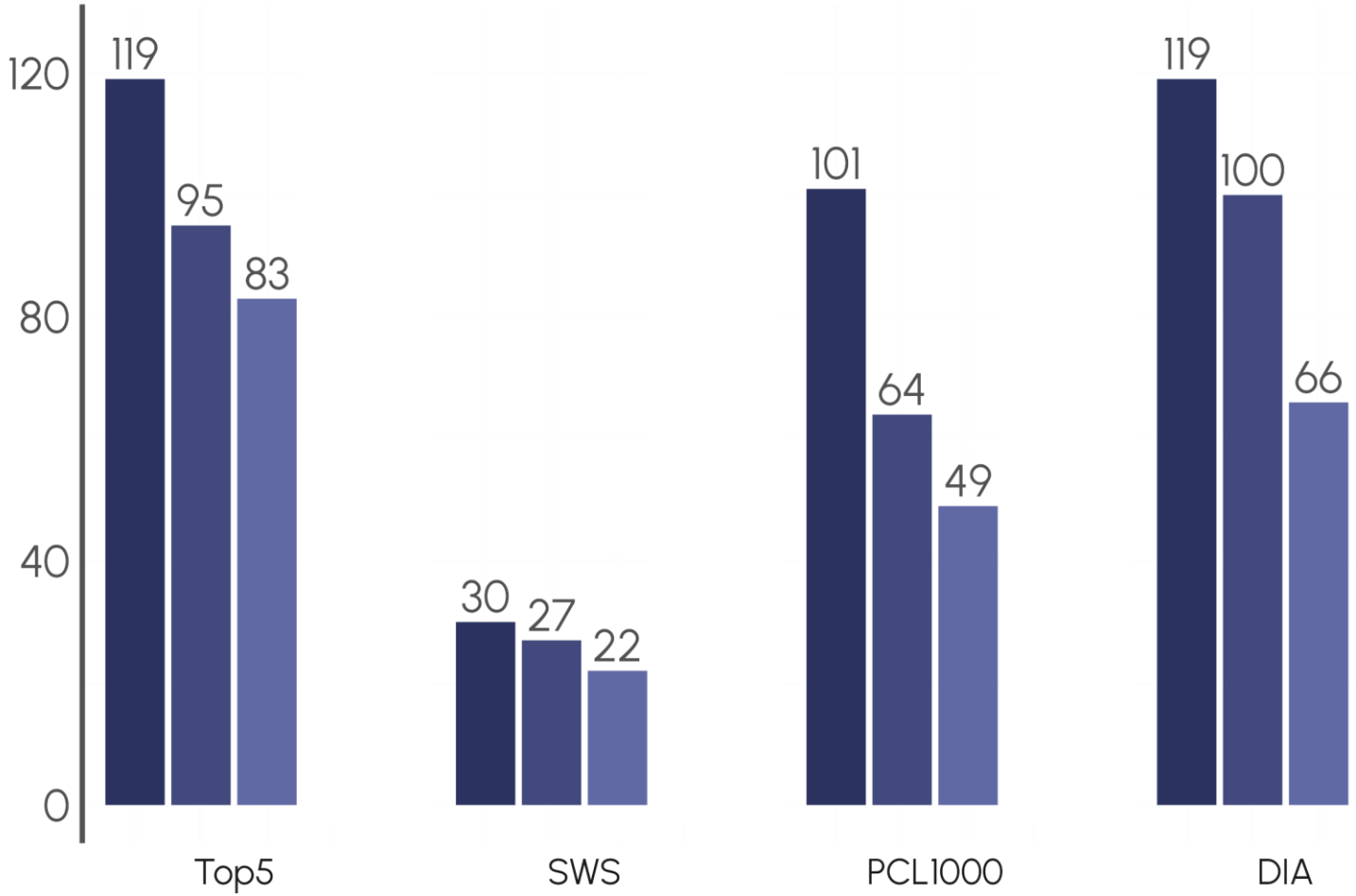
Comparison of methods - standards

Standard mix
119 chemicals



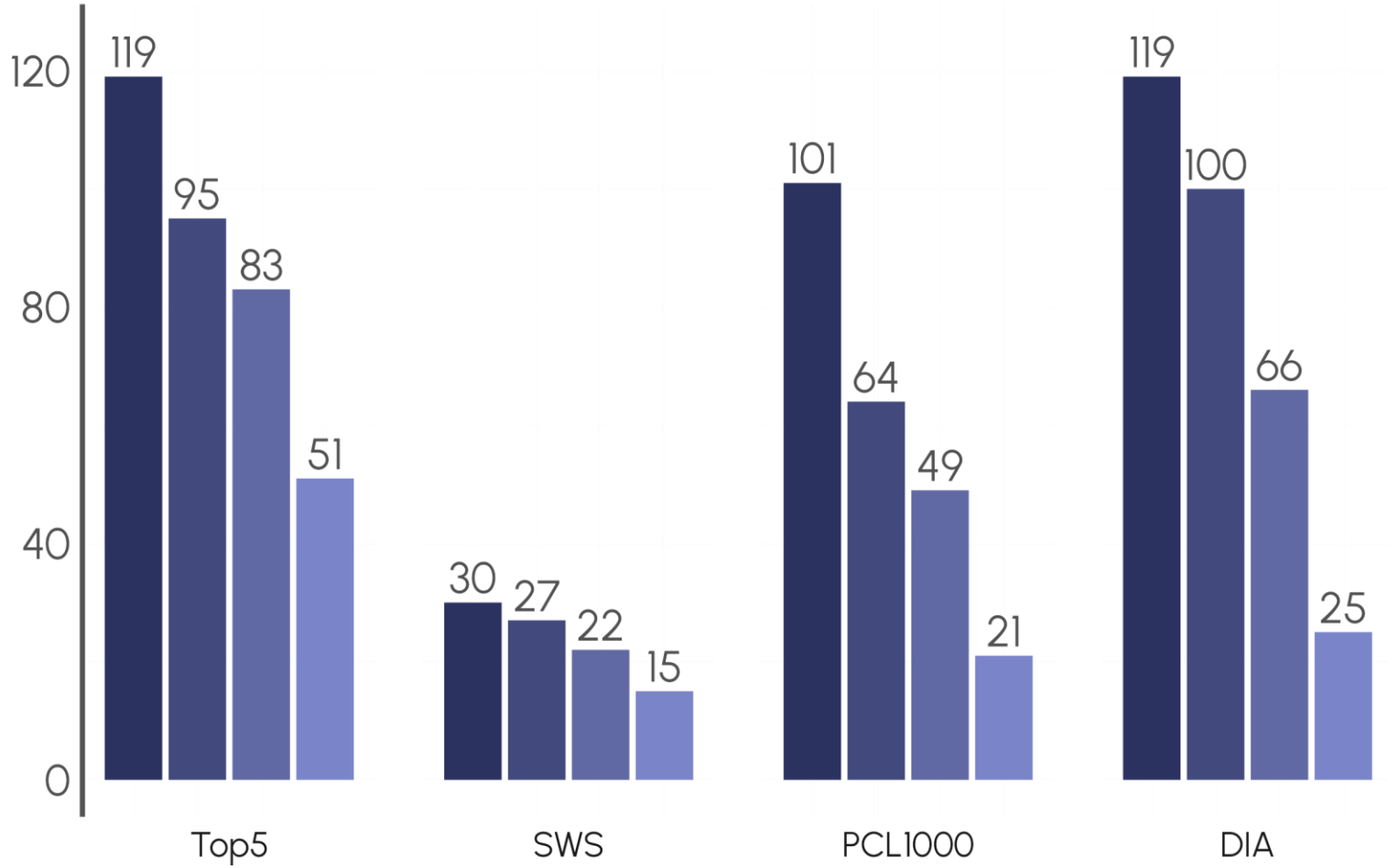
Comparison of methods - standards

Standard mix
119 chemicals



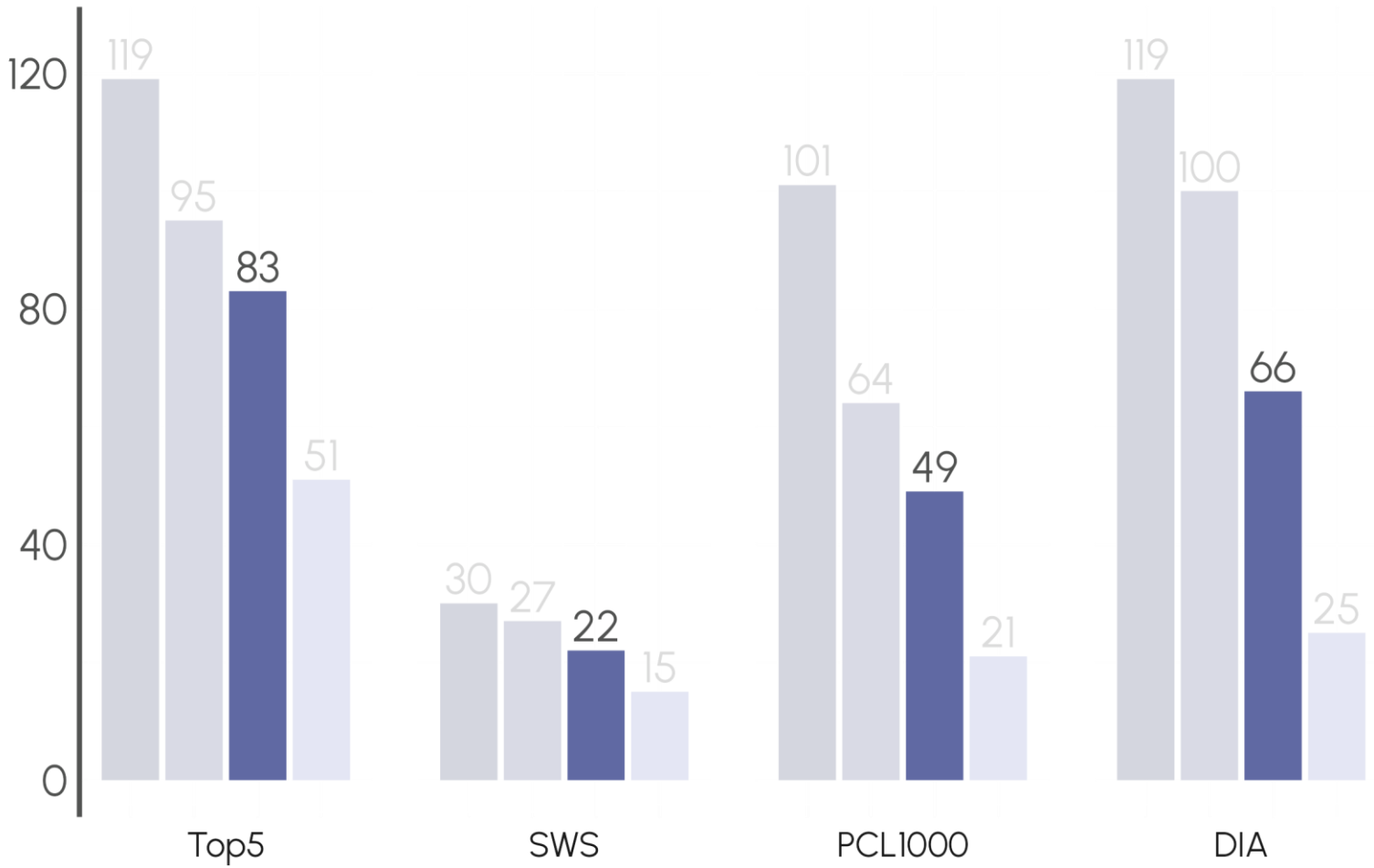
Comparison of methods - standards

Standard mix
119 chemicals



Comparison of methods - standards

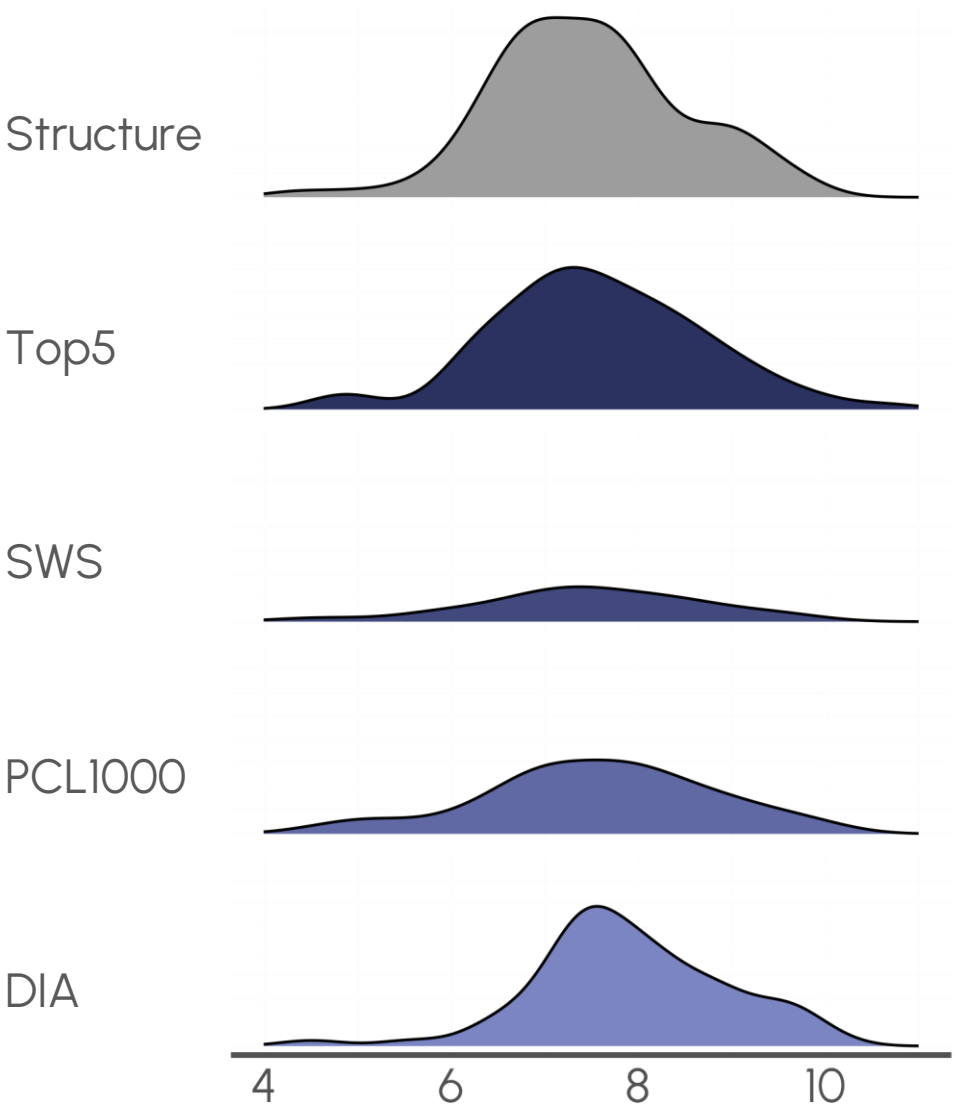
Standard mix
119 chemicals



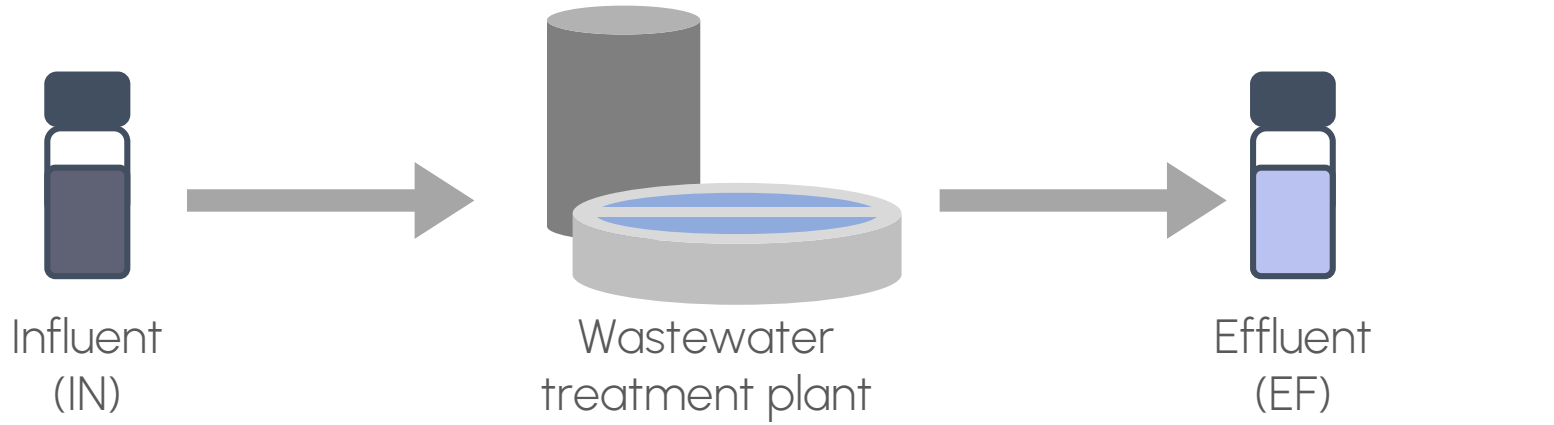
Comparison of methods - standards

Standard mix
119 chemicals

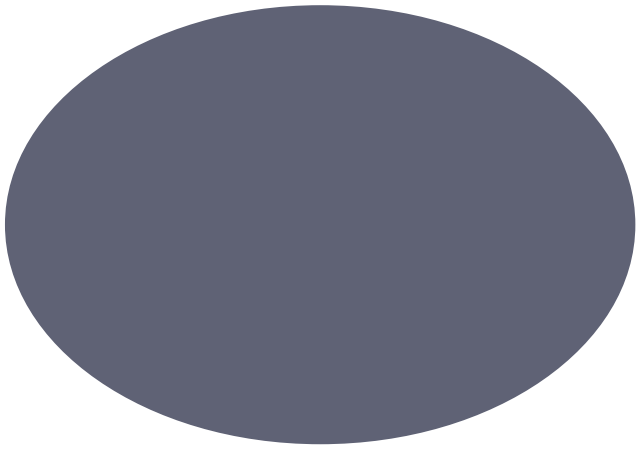
$$\text{PriorityScore} = \frac{\frac{\text{Intensity}}{\text{IE}}}{\text{LC}_{50}}$$



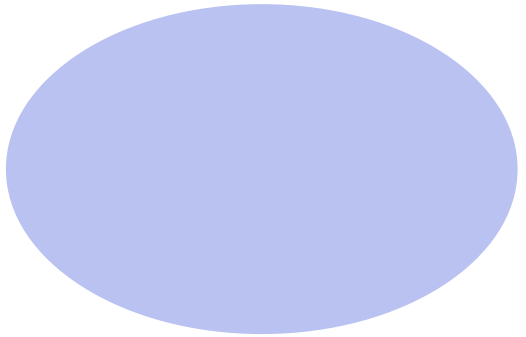
Comparison of methods - wastewater



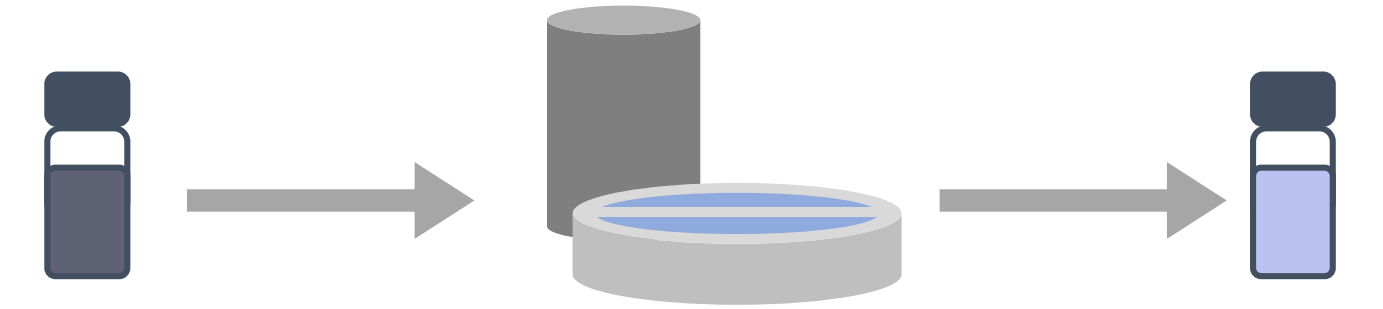
12,029



6,006



Comparison of methods - wastewater



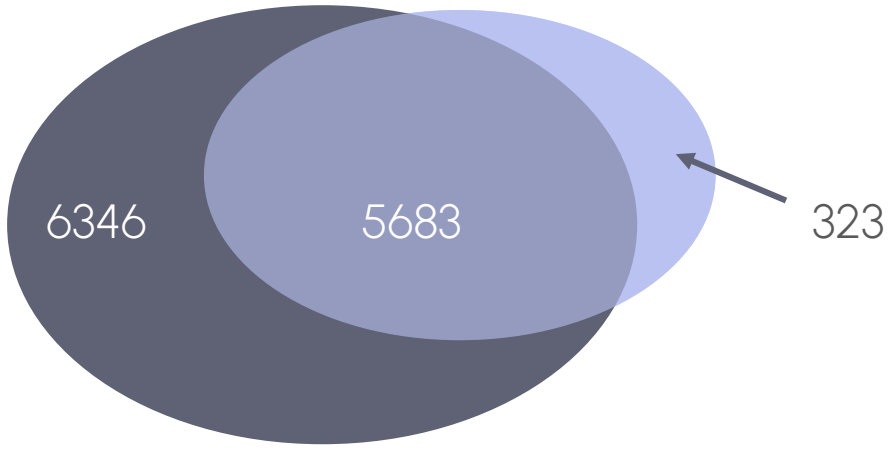
Influent (IN)

Wastewater treatment plant

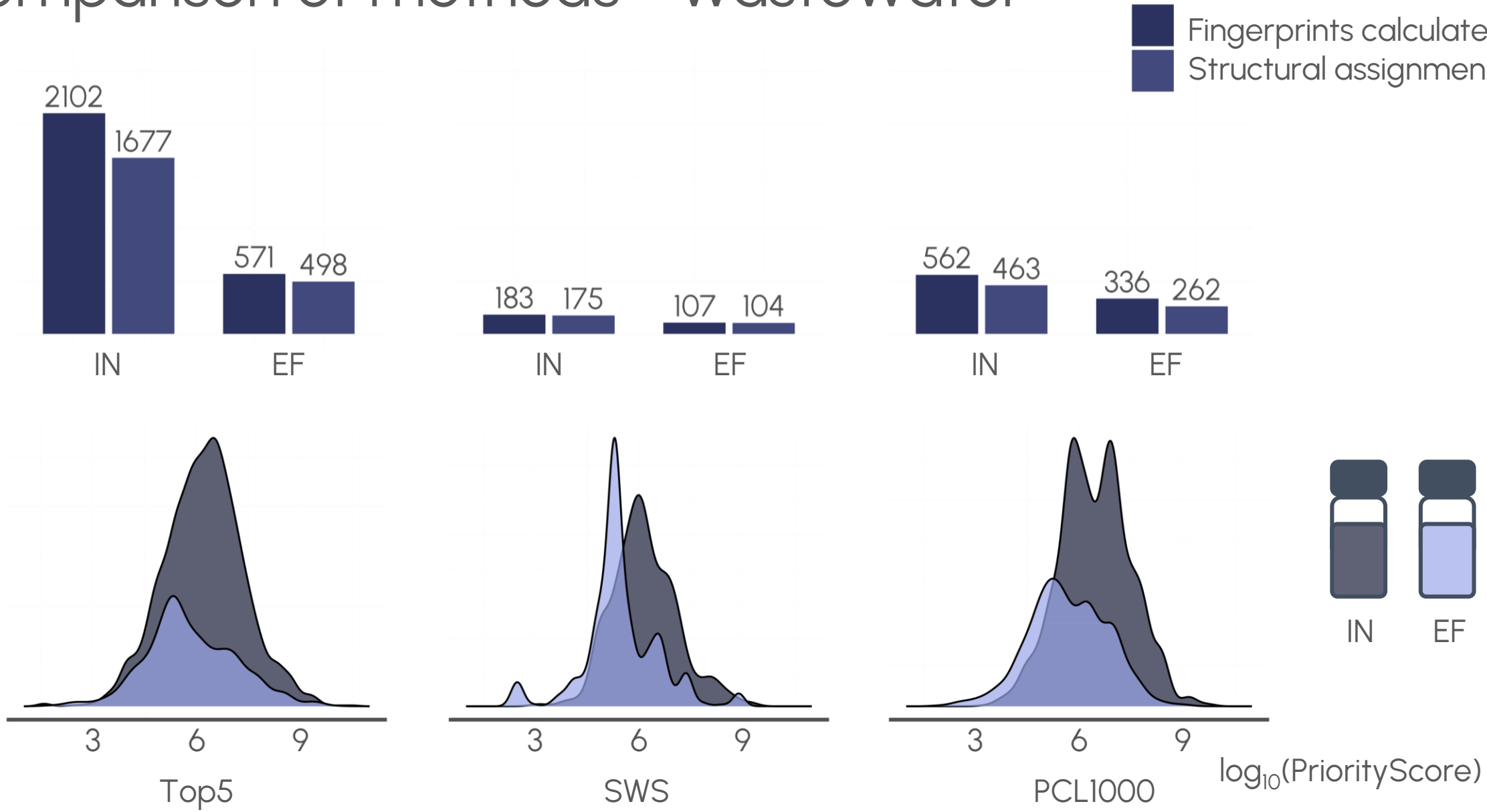
Effluent (EF)

12,029

6,006

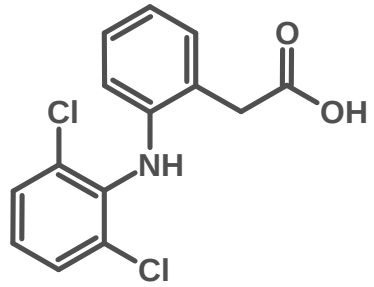


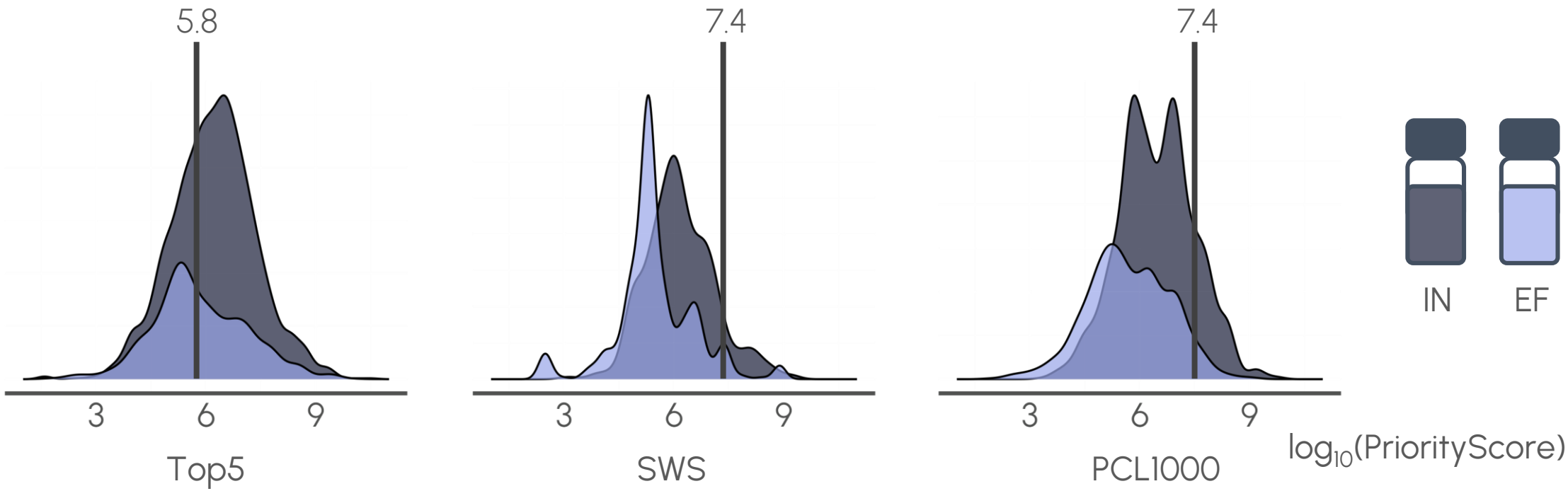
Comparison of methods - wastewater



Comparison of methods - wastewater

Diclofenac

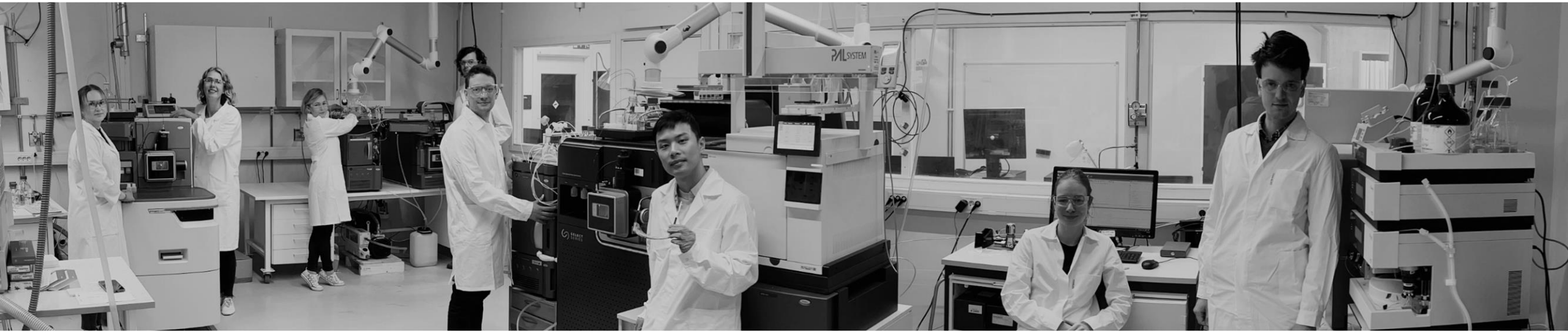




Preliminary conclusions and future prospects

- Chemicals triggered by different approaches cover similar range of priority scores
- Acquisition methods are complementary
- Confirmation of detected chemicals

Acknowledgements



Lisa Jonsson, Louise Malm, Pilleriin Peets, Malte Posselt, Michael McLachlan, Matthew MacLeod, Magnus Breitholtz, Jonathan Martin, Anneli Kruve

Depmartment of Materials and Environmental Chemistry and Department of Environmental Science in Stockholm University

FORMAS RapMixTox #2021-01511

Helen Sepman Helen.Sepman@aces.su.se

Kruvelab.com