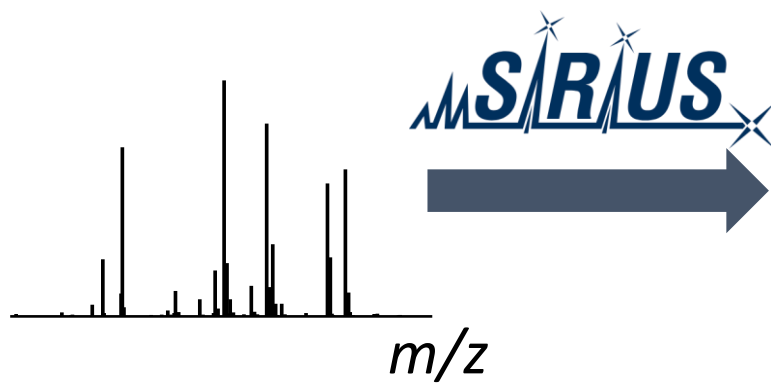


MS2Tox Reloaded: Discovering Endocrine Disruptors

Yvonne Kreutzer, Ida Rahu, Ulf Norinder, Anneli Kruve

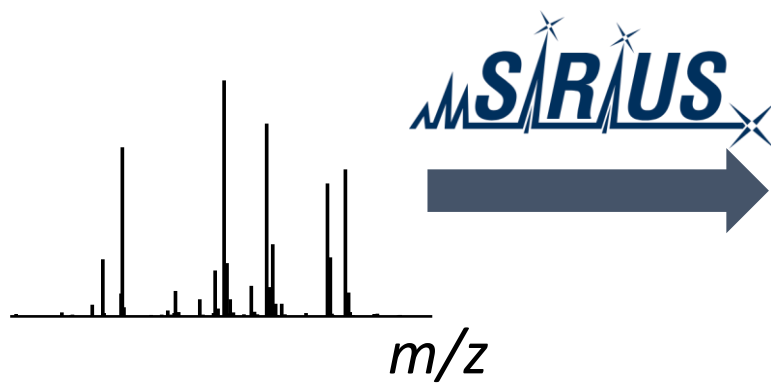


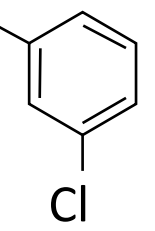
yvonne.kreutzer@su.se



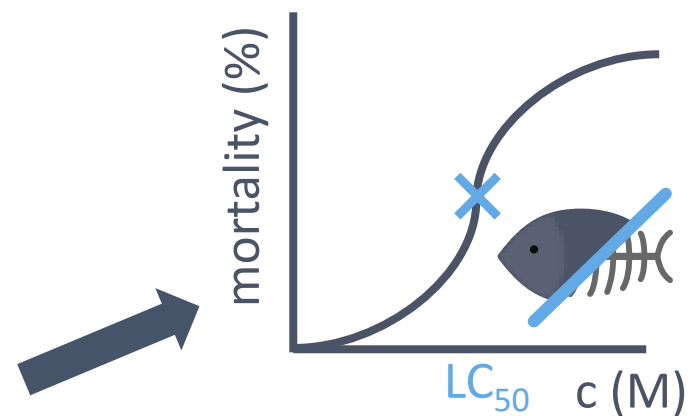
C — N	0.20
<chem>Clc1ccc(Cl)cc1</chem>	0.64
C — Cl	0.81
...	...

Fingerprints

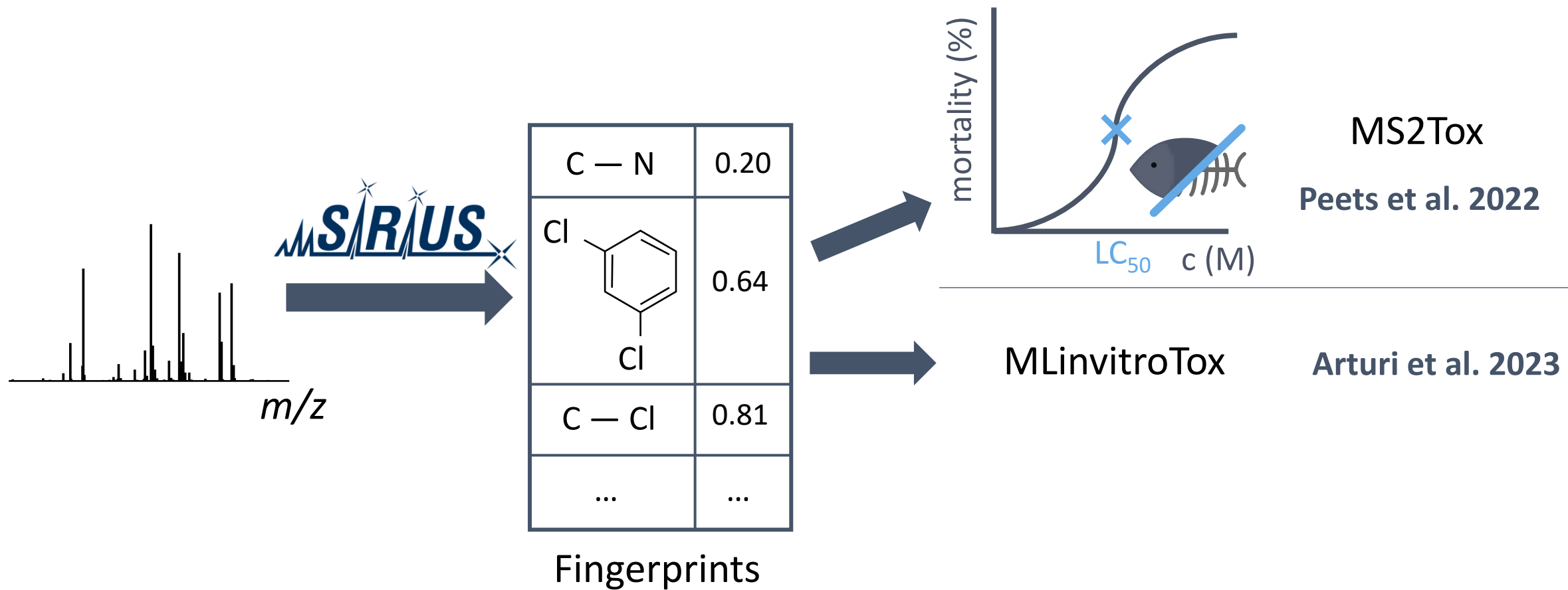


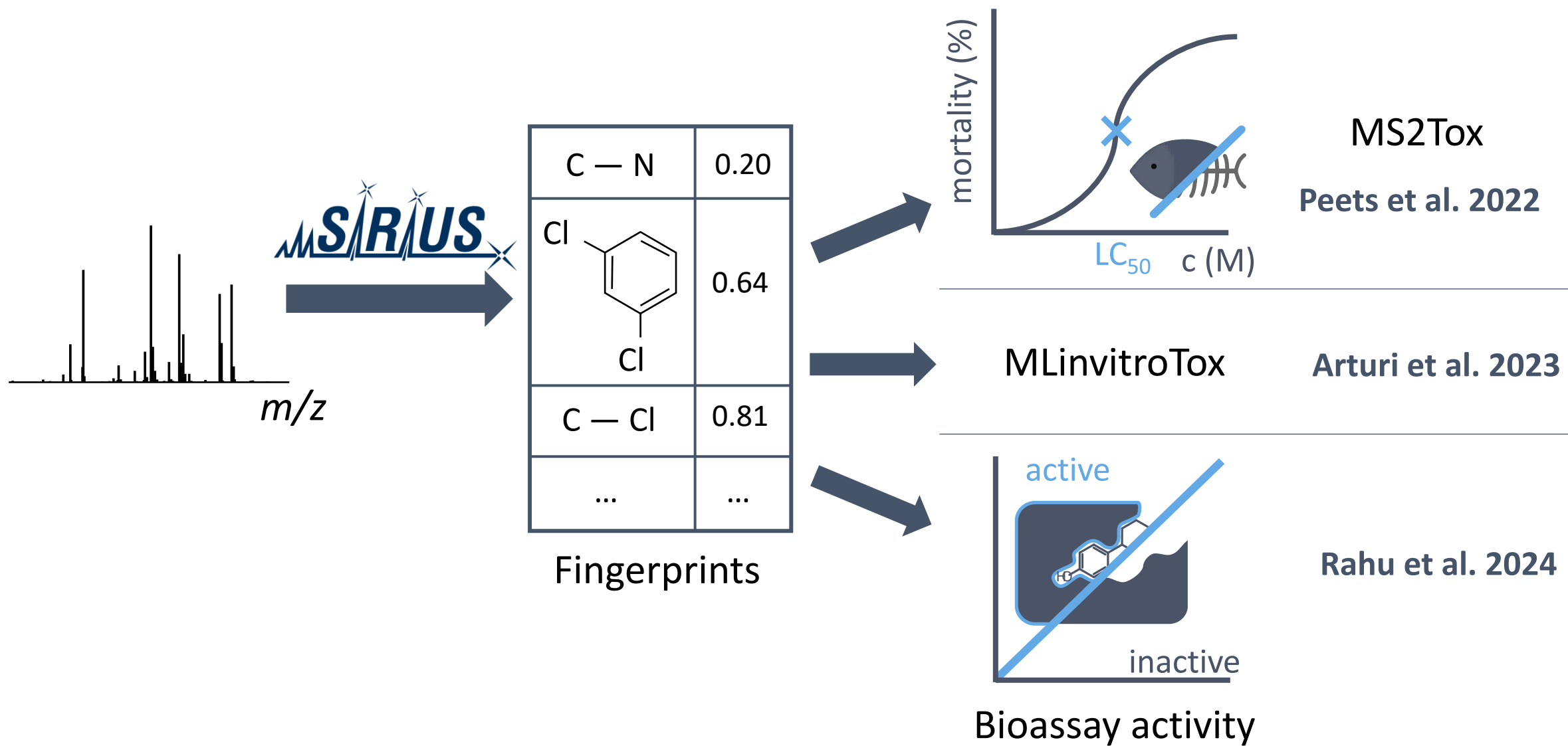
C — N	0.20
<chem>Clc1ccc(Cl)cc1</chem> 	0.64
C — Cl	0.81
...	...

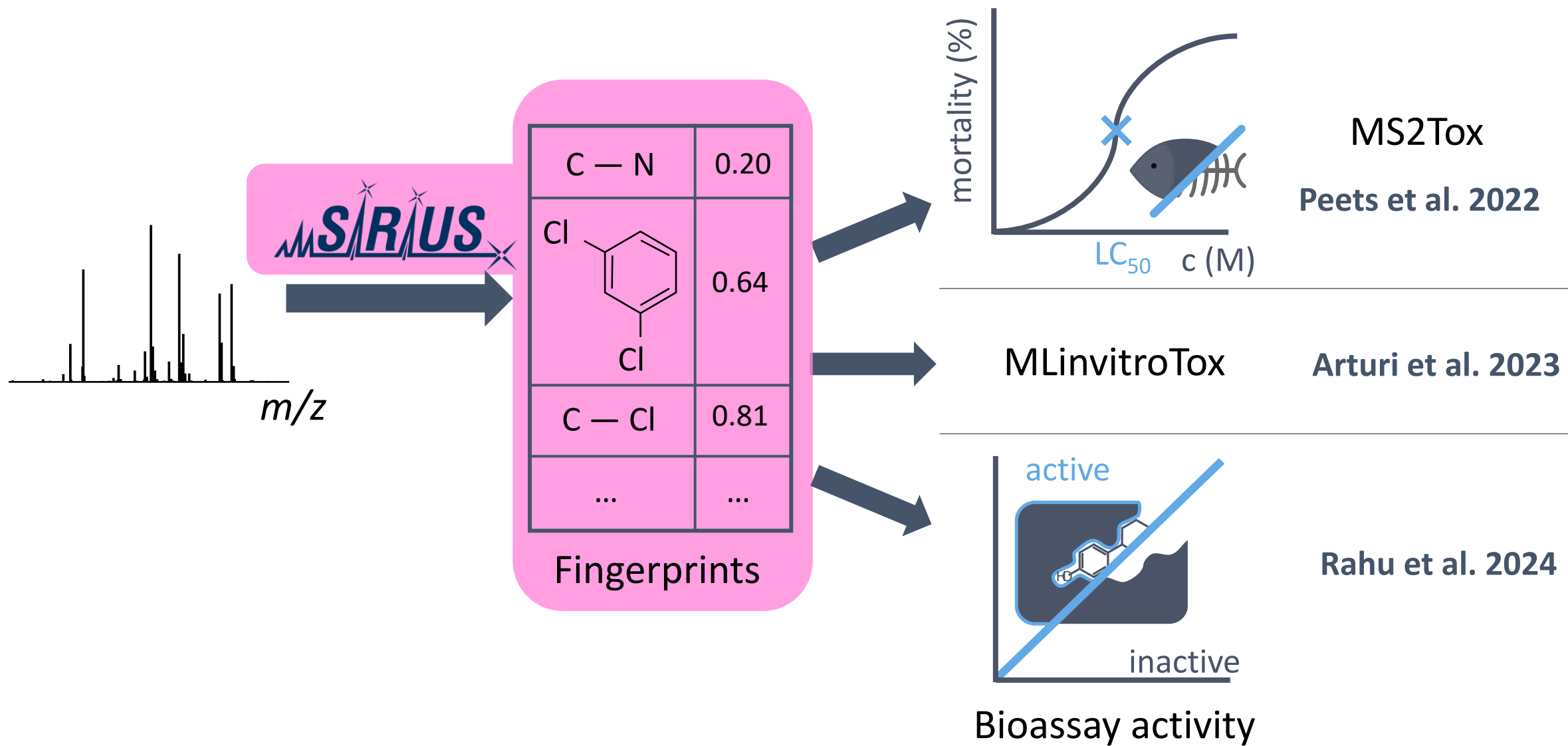
Fingerprints

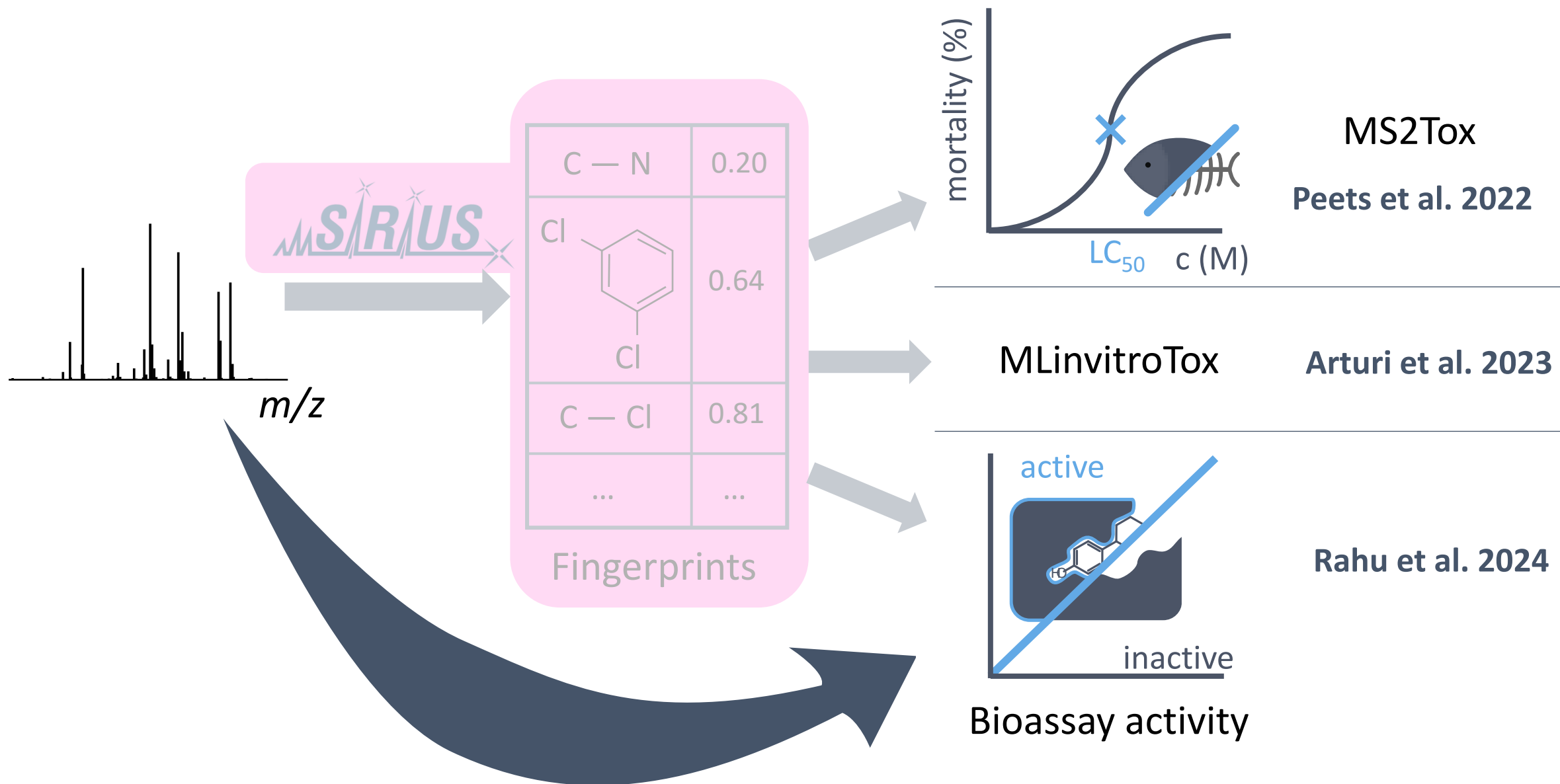


MS2Tox
Peets et al. 2022









Aim



Development of alternative approaches



Molecular networking



Conformal predictions



Compare alternative approaches against Fingerprint approach



Application on wastewater samples

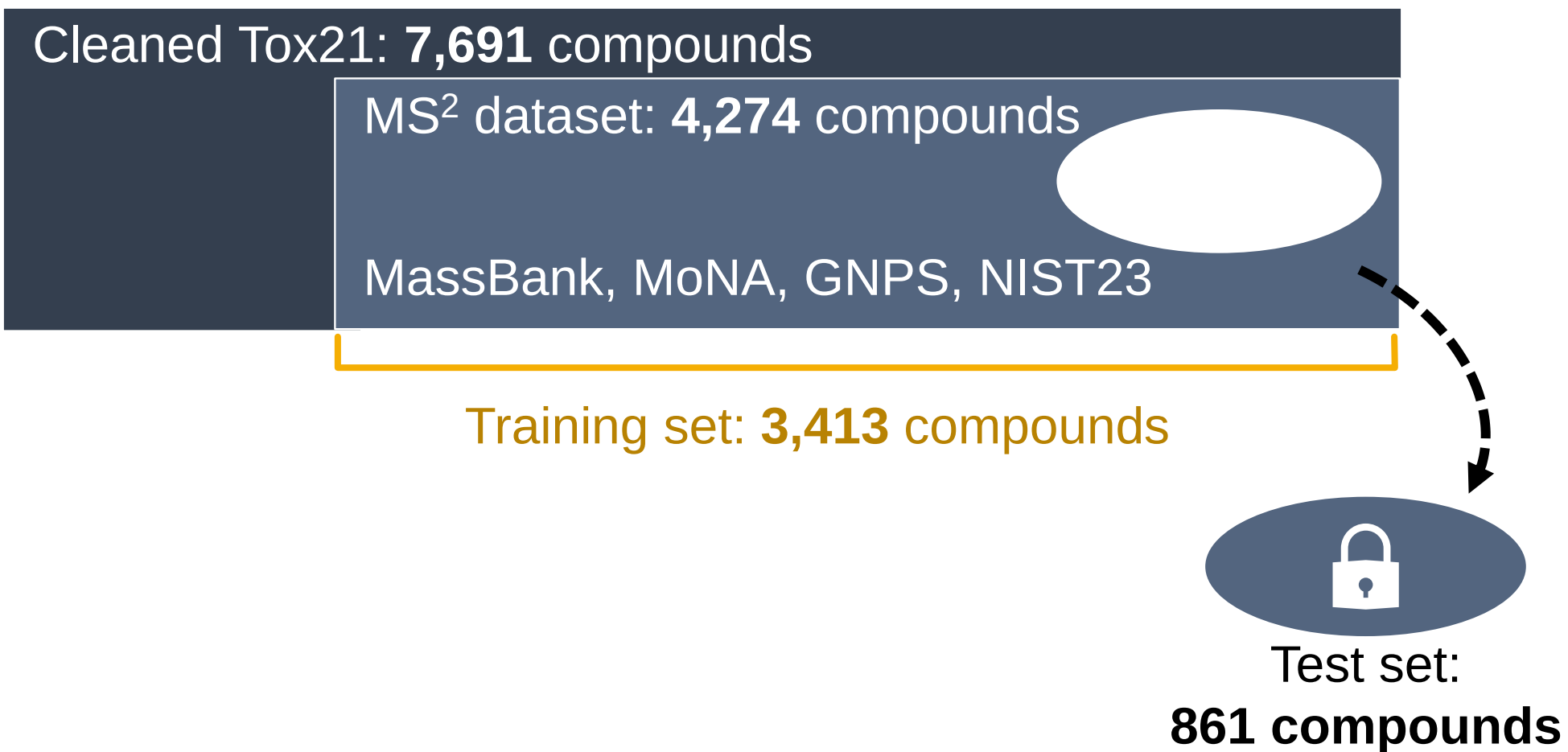
Data

Cleaned Tox21: **7,691** compounds

MS² dataset: **4,274** compounds

MassBank, MoNA, GNPS, NIST23

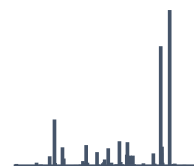
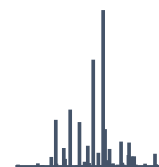
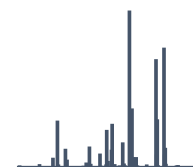
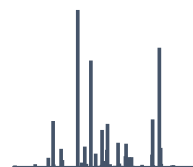
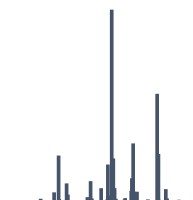
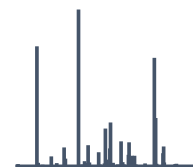
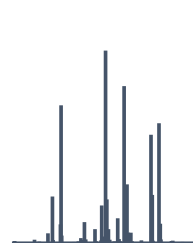
Data



1



Alternative Approach: Molecular Networking

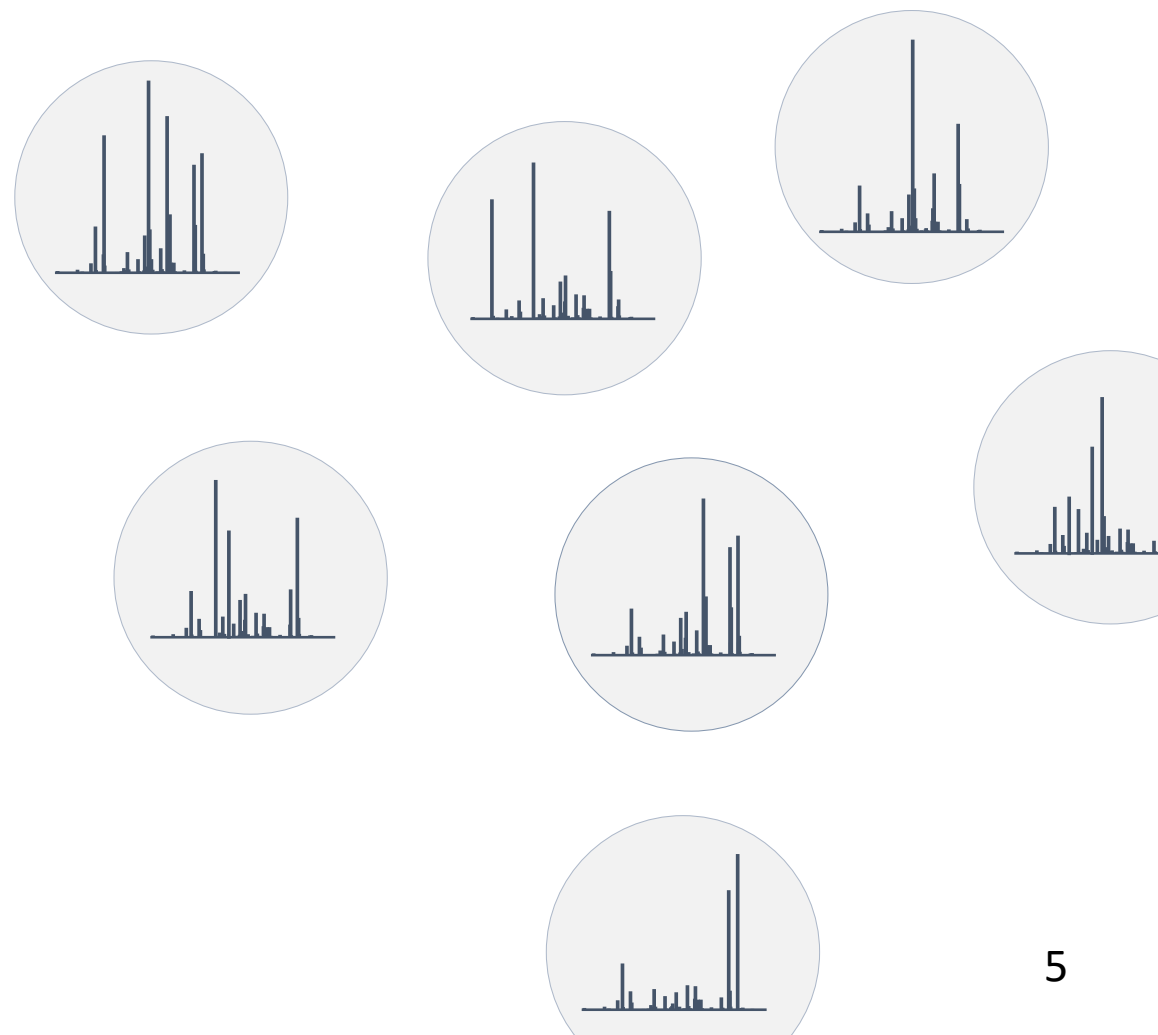


1



Alternative Approach: Molecular Networking

Nodes = MS² spectra



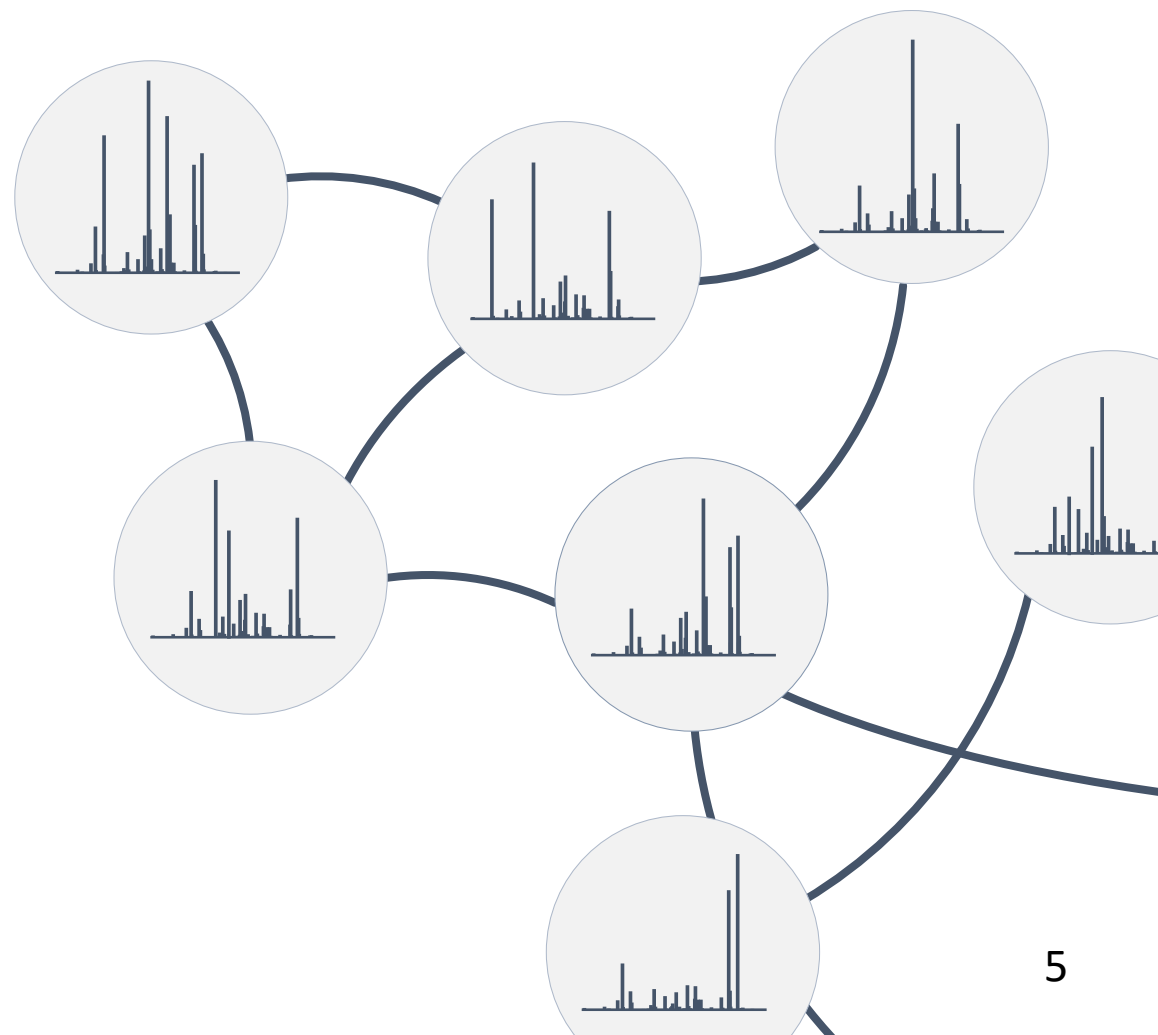
1



Alternative Approach: Molecular Networking

Nodes = MS² spectra

Edges = spectra similarity





Alternative Approach: Molecular Networking

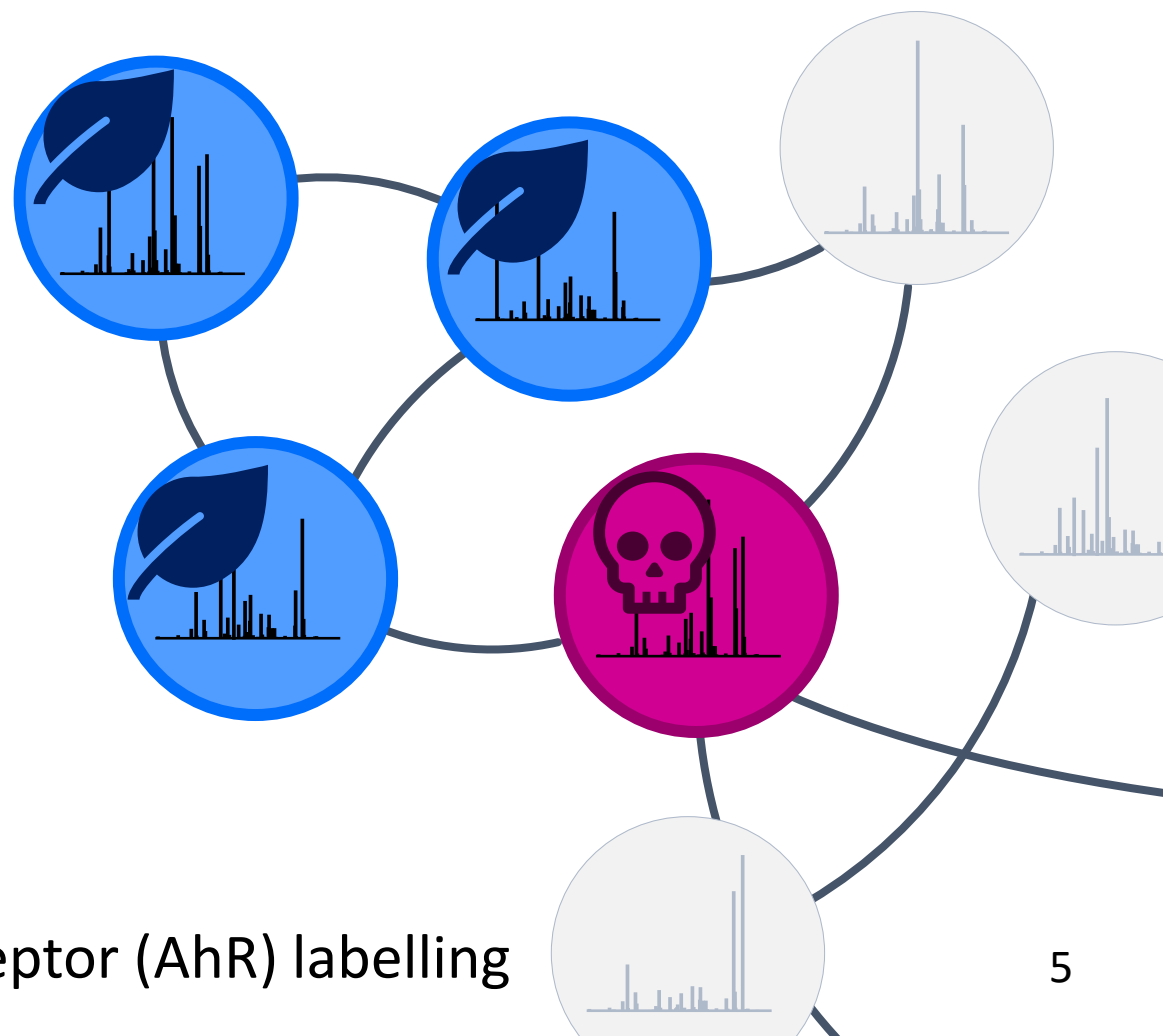
Nodes = MS² spectra

Edges = spectra similarity

● Active

● Inactive

● Inconclusive



Aryl hydrocarbon receptor (AhR) labelling

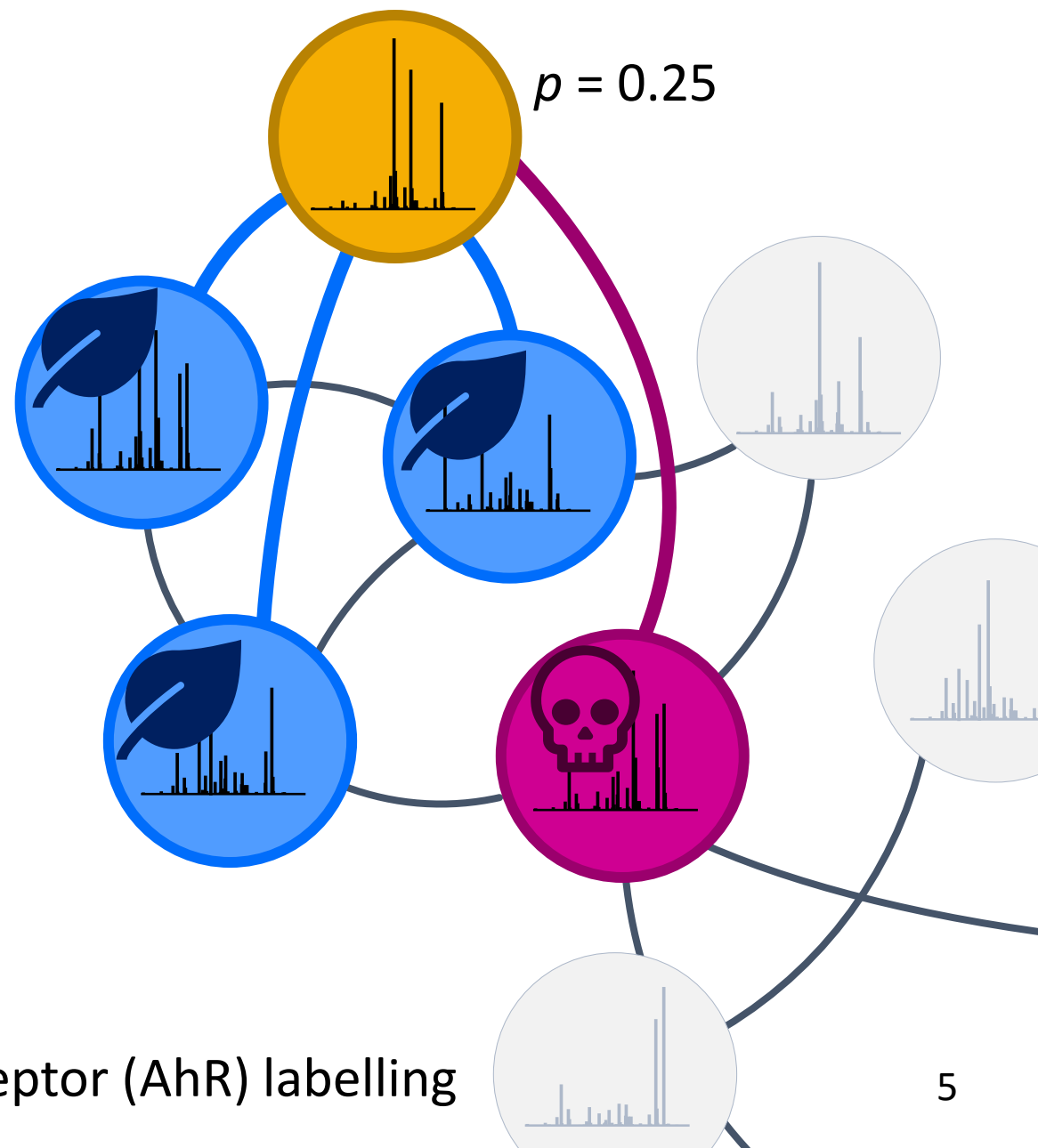


Alternative Approach: Molecular Networking

Nodes = MS² spectra

Edges = spectral similarity

- Active
- Inactive
- Inconclusive
- Unknown



Aryl hydrocarbon receptor (AhR) labelling

1



Alternative Approach: Molecular Networking

Spectral similarity metrics

Cosine similarity

M cosine similarity

MS2DeepScore

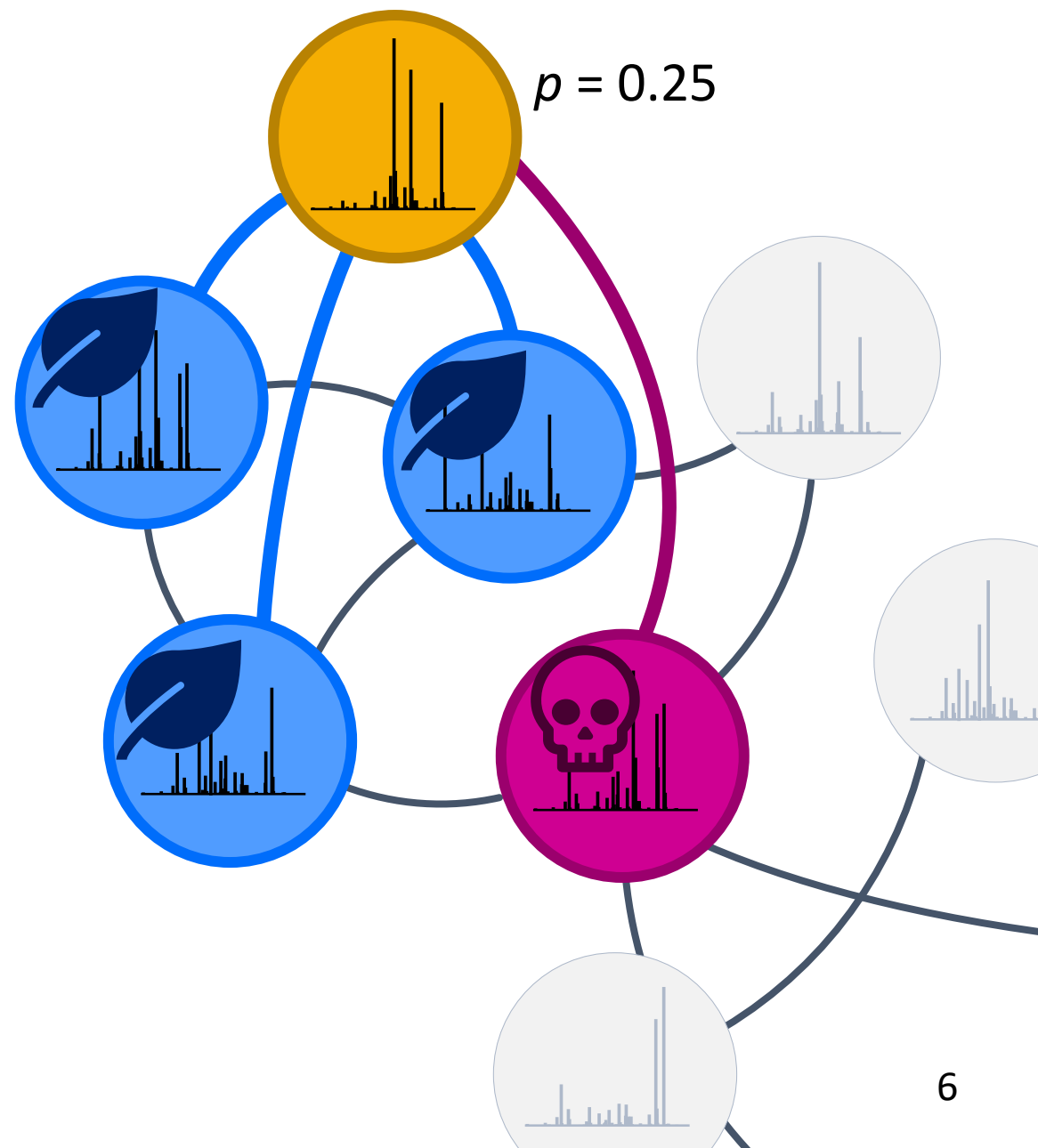
Similarity thresholds

0.0, 0.1, ..., 0.9

Voting scheme

Majority

Weighted



1



Alternative Approach: Molecular Networking

Spectral similarity metrics

Cosine similarity

Modified cosine similarity

MS2DeepScore

Similarity thresholds

0.0, 0.1, ..., 0.5, ..., 0.9

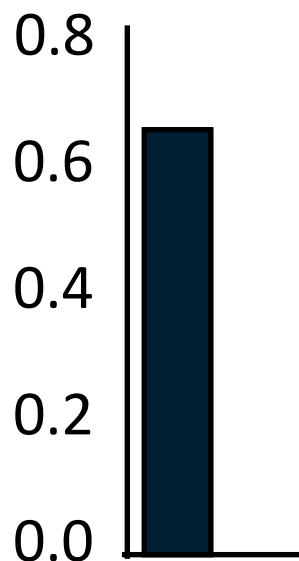
Voting scheme

Majority

Weighted

AhR

$FPR_{TPR=0.9}$





Alternative Approach: Molecular Networking

Spectral similarity metrics

Cosine similarity

Modified cosine similarity

MS2DeepScore

Similarity thresholds

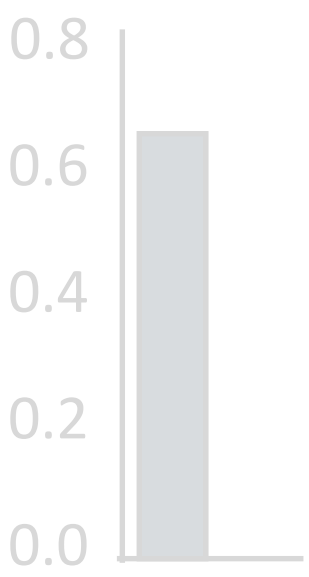
0.0, 0.1, ..., 0.5, ..., 0.9

Voting scheme

Majority

Weighted

AhR
FPR_{TPR=0.9}



		true label	
		active	non-active
prediction	active	TP	FP
	non-active	FN	TN

1



Alternative Approach: Molecular Networking

Spectral similarity metrics

Cosine similarity

Modified cosine similarity

MS2DeepScore

Similarity thresholds

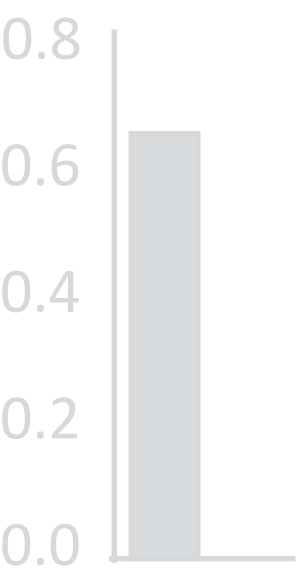
0.0, 0.1, ..., 0.5, ..., 0.9

Voting scheme

Majority

Weighted

AhR
 $FPR_{TPR=0.9}$



		true label	
		active	non-active
prediction	active	TP	FP
	non-active	FN	TN

1



Alternative Approach: Molecular Networking

Spectral similarity metrics

Cosine similarity

Modified cosine similarity

MS2DeepScore

Similarity thresholds

0.0, 0.1, ..., 0.5, ..., 0.9

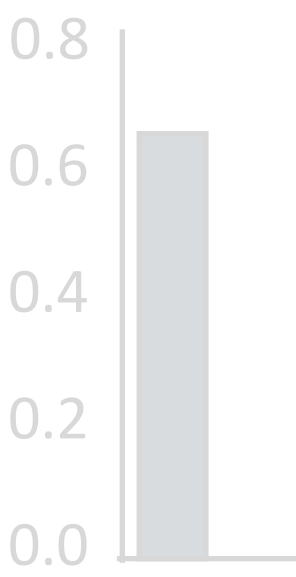
Voting scheme

Majority

Weighted

AhR

FPR_{TPR=0.9}



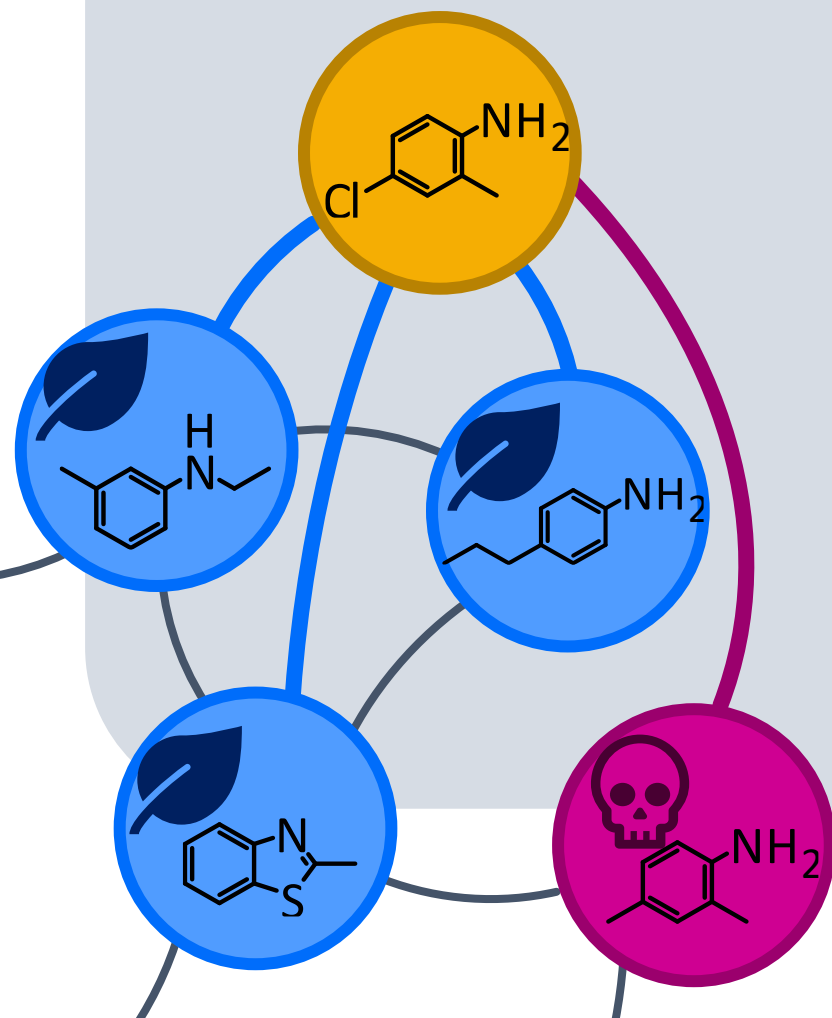
		true label	
		active	non-active
prediction	active	TP	FP
	non-active	FN	TN

1



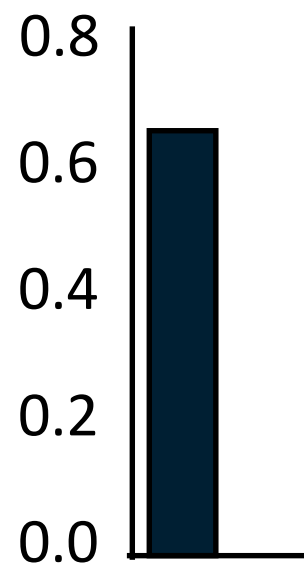
Alternative Approach: Molecular Networking

Structural similarity reference



AhR

$\text{FPR}_{\text{TPR}=0.9}$

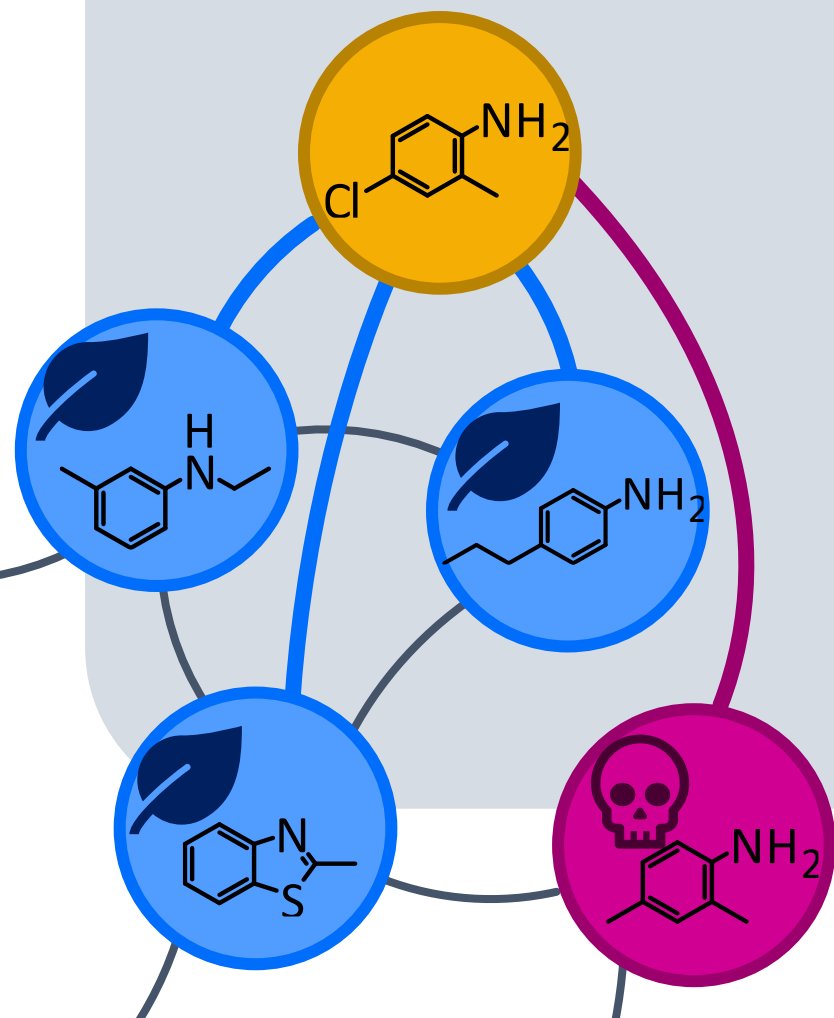


1



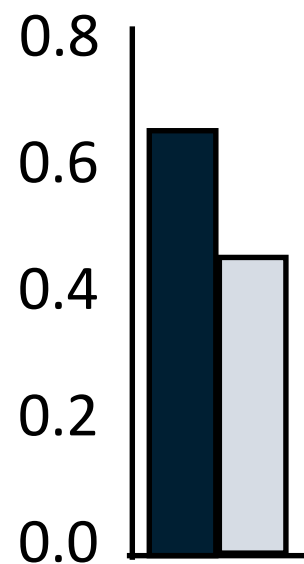
Alternative Approach: Molecular Networking

Structural similarity reference



AhR

$FPR_{TPR=0.9}$

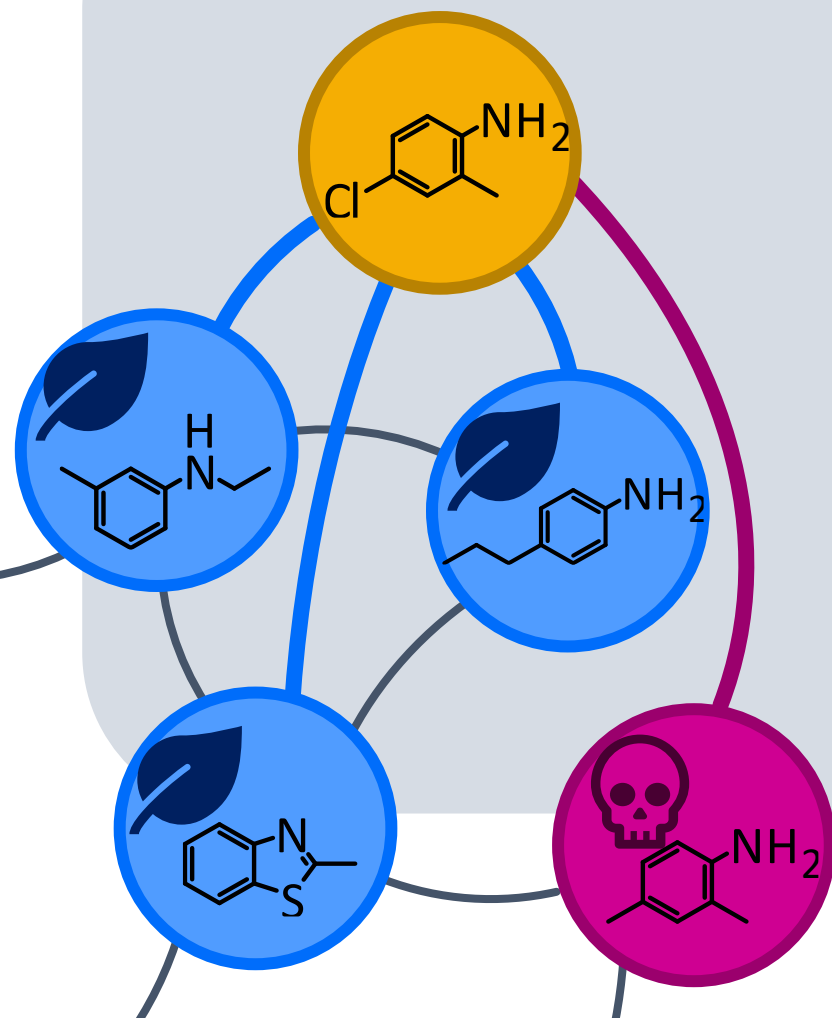


1



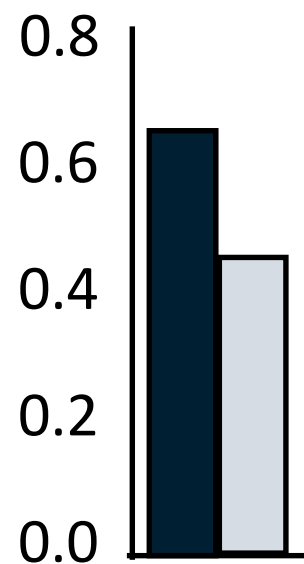
Alternative Approach: Molecular Networking

Structural similarity reference



AhR

$\text{FPR}_{\text{TPR}=0.9}$



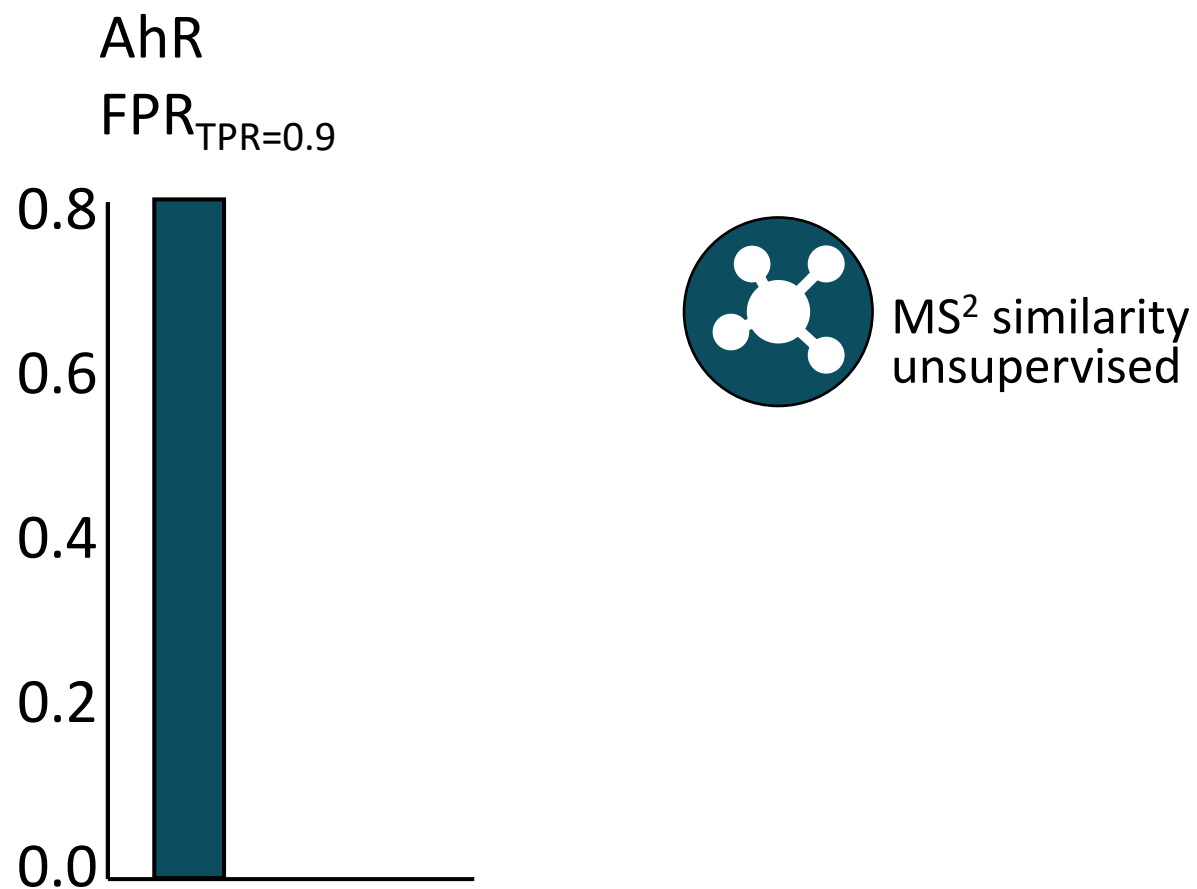
Limited structural
information in MS^2
spectra
(e.g. low spectral
quality)



Approach Comparison



Test set

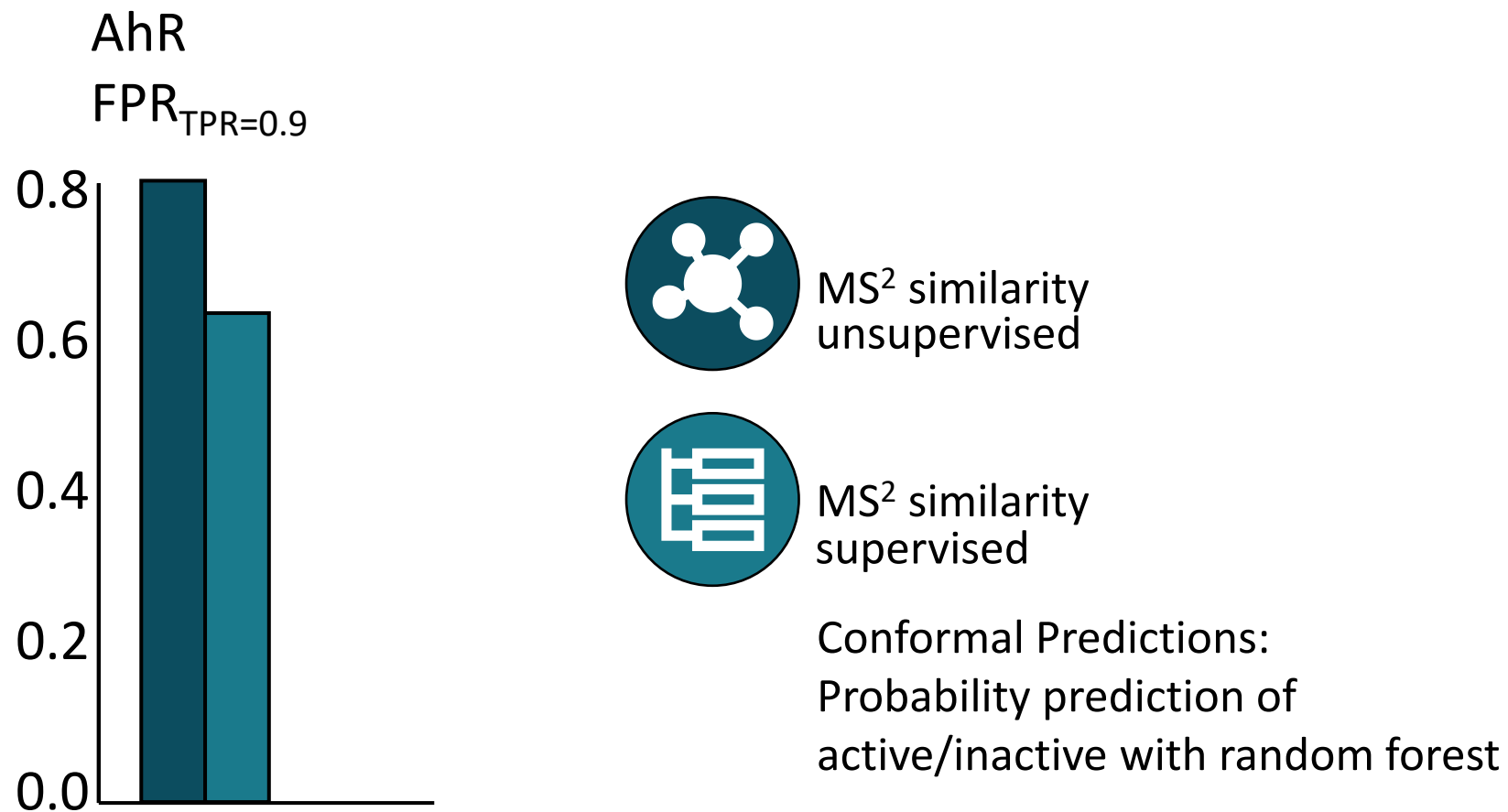




Approach Comparison



Test set

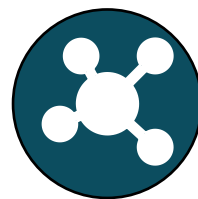
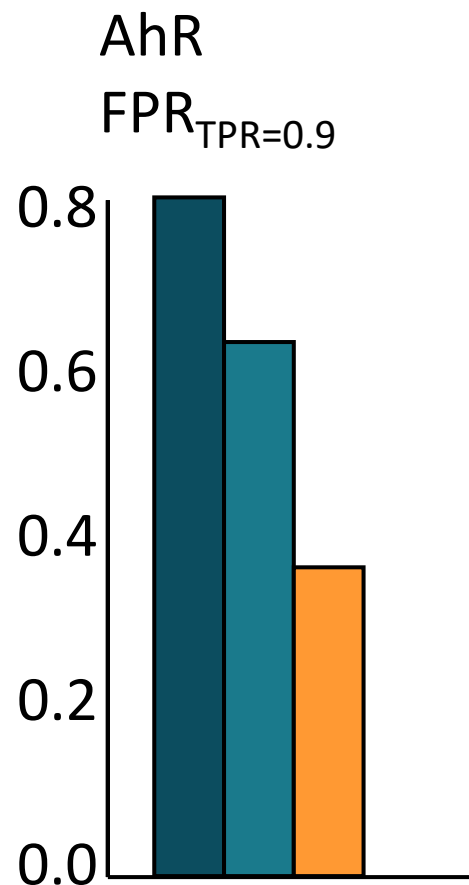




Approach Comparison



Test set



MS² similarity
unsupervised



MS² similarity
supervised



Fingerprints
supervised



Fingerprint models
achieve lowest
 $FPR_{TPR=0.9}$

3 Wastewater Application



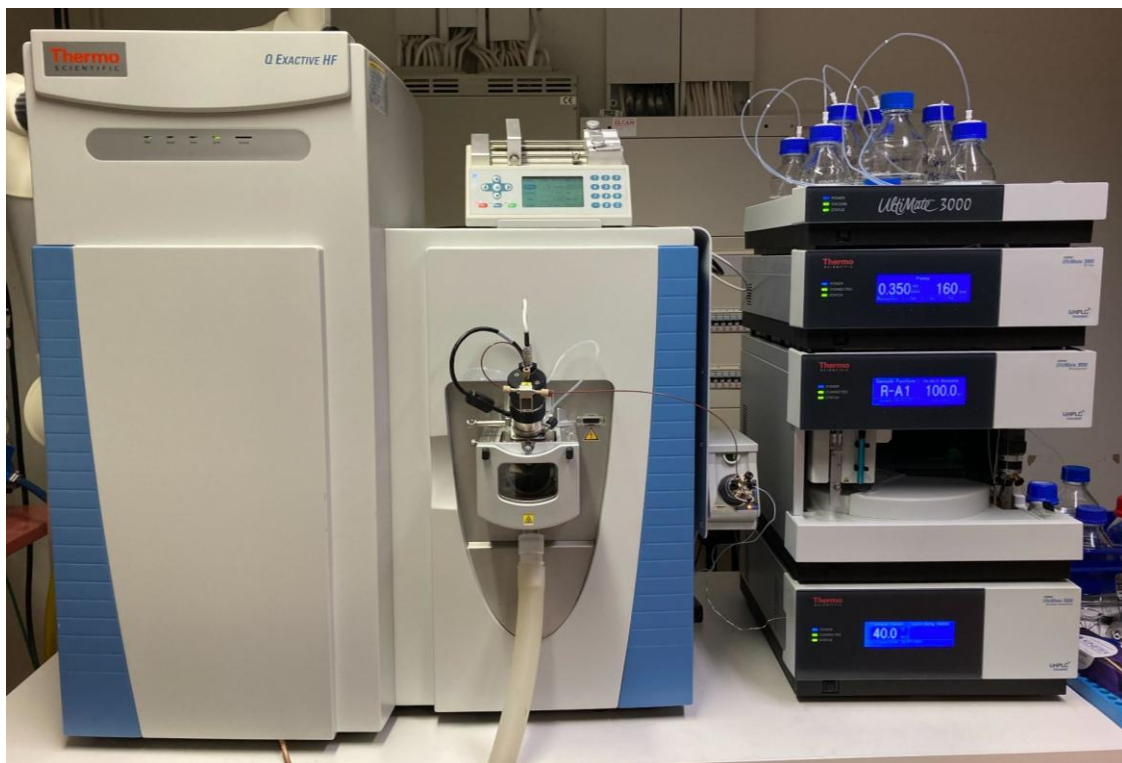
Influent



Effluent



MS-DIAL

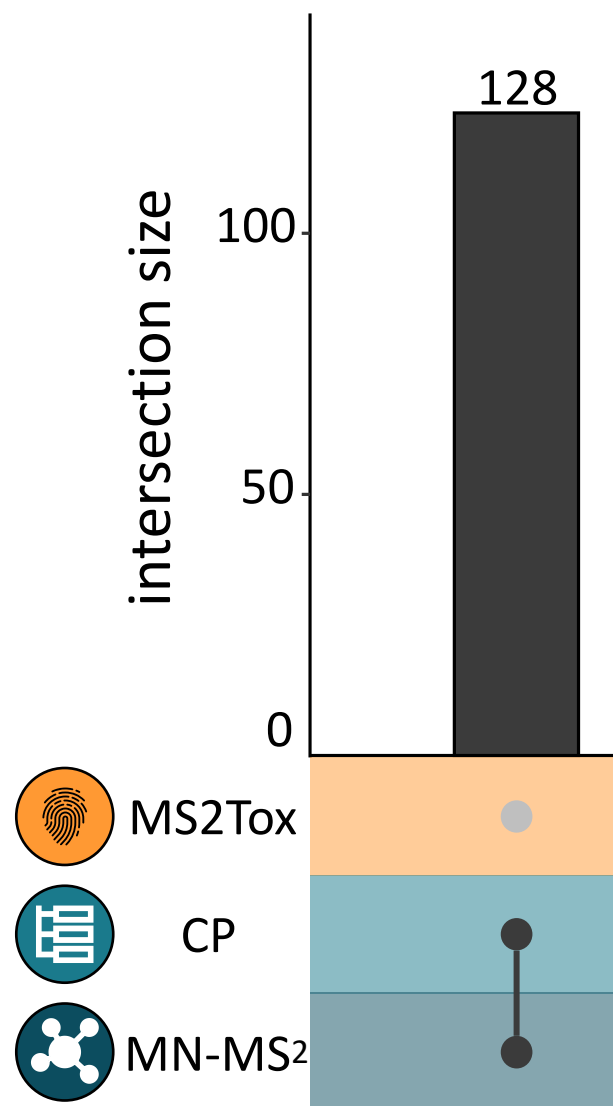


4,375 features

193 potential
persistent compounds or
transformation products

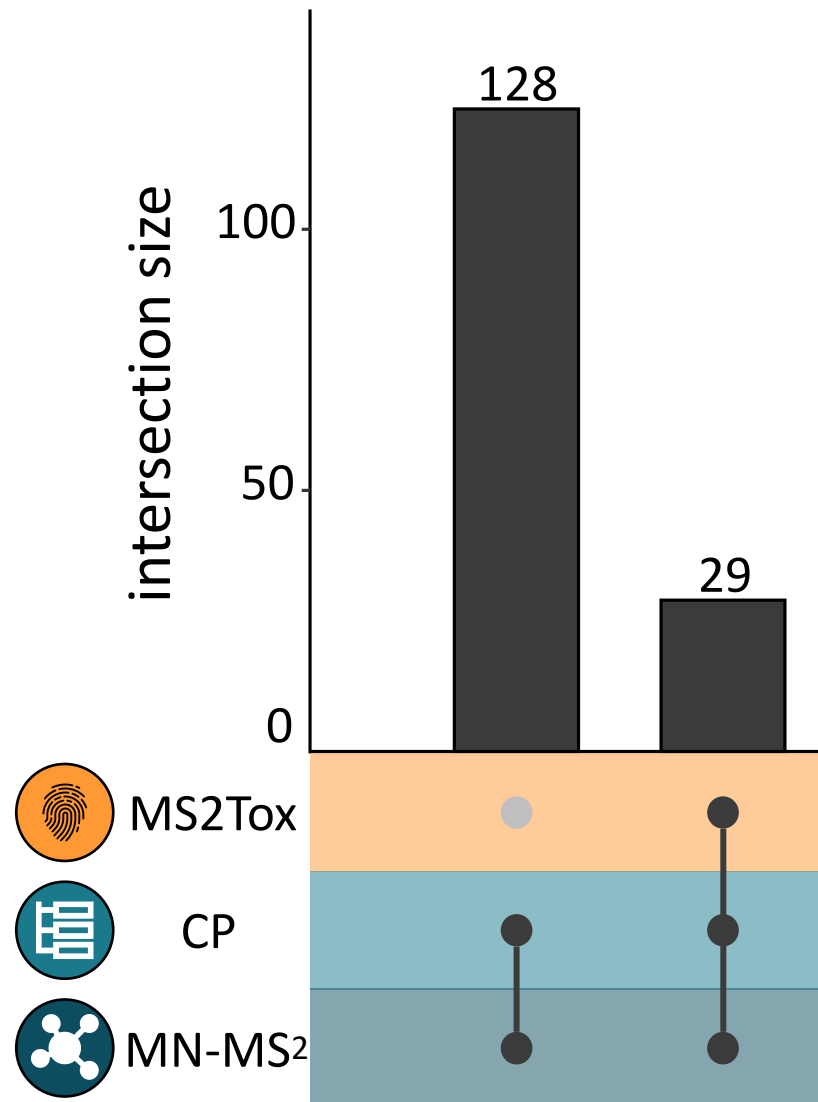


Wastewater Application



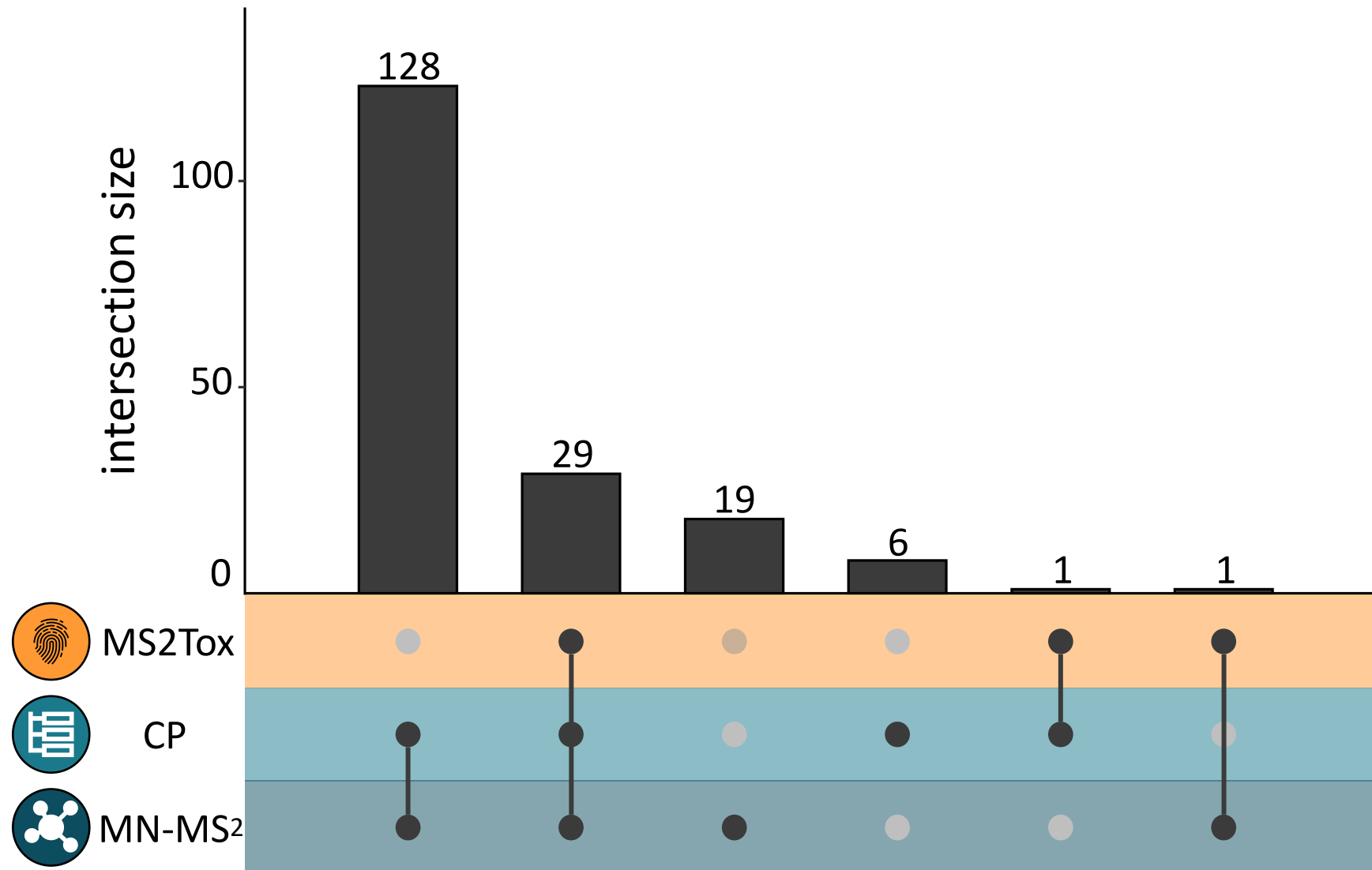


Wastewater Application



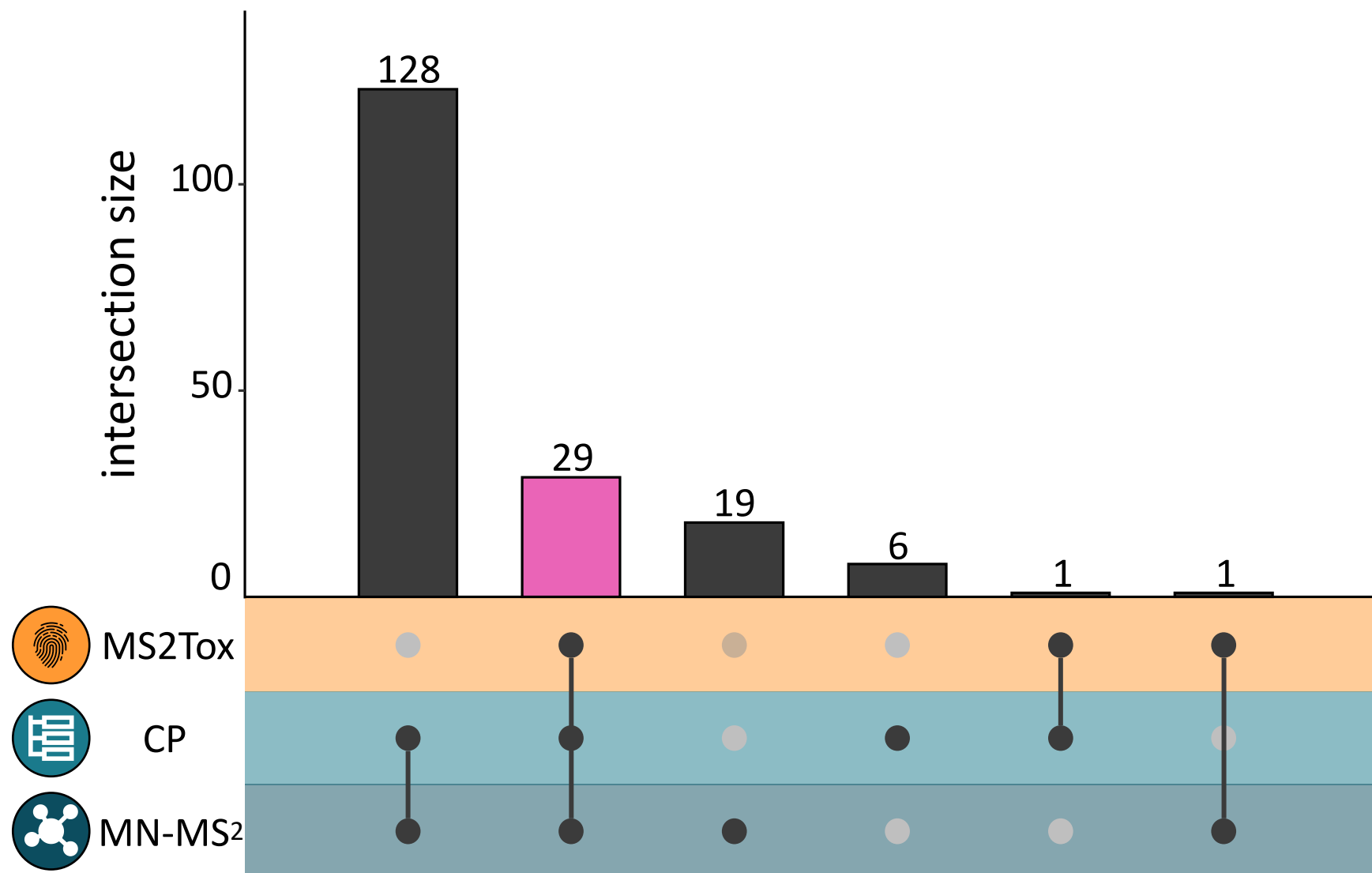


Wastewater Application

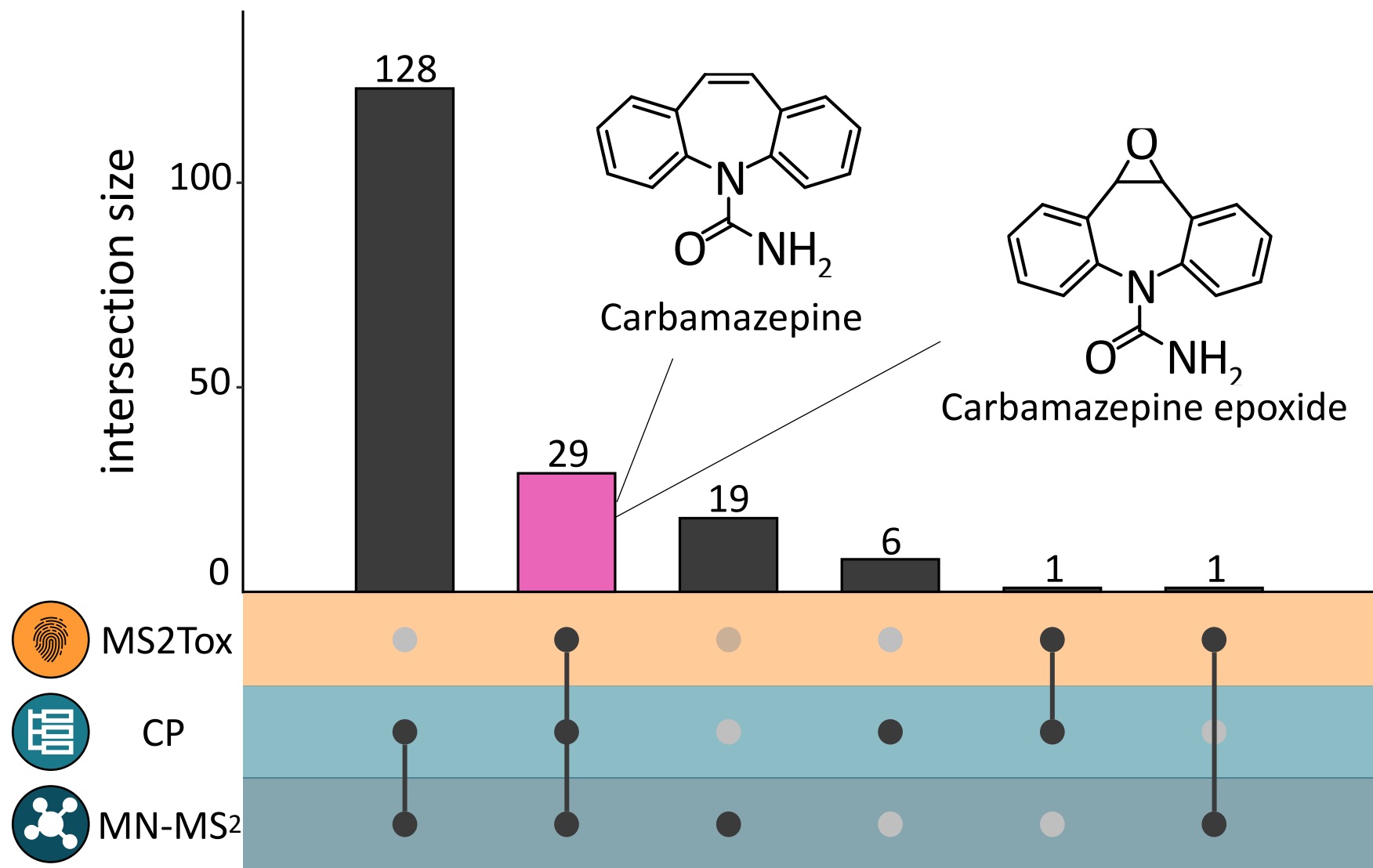




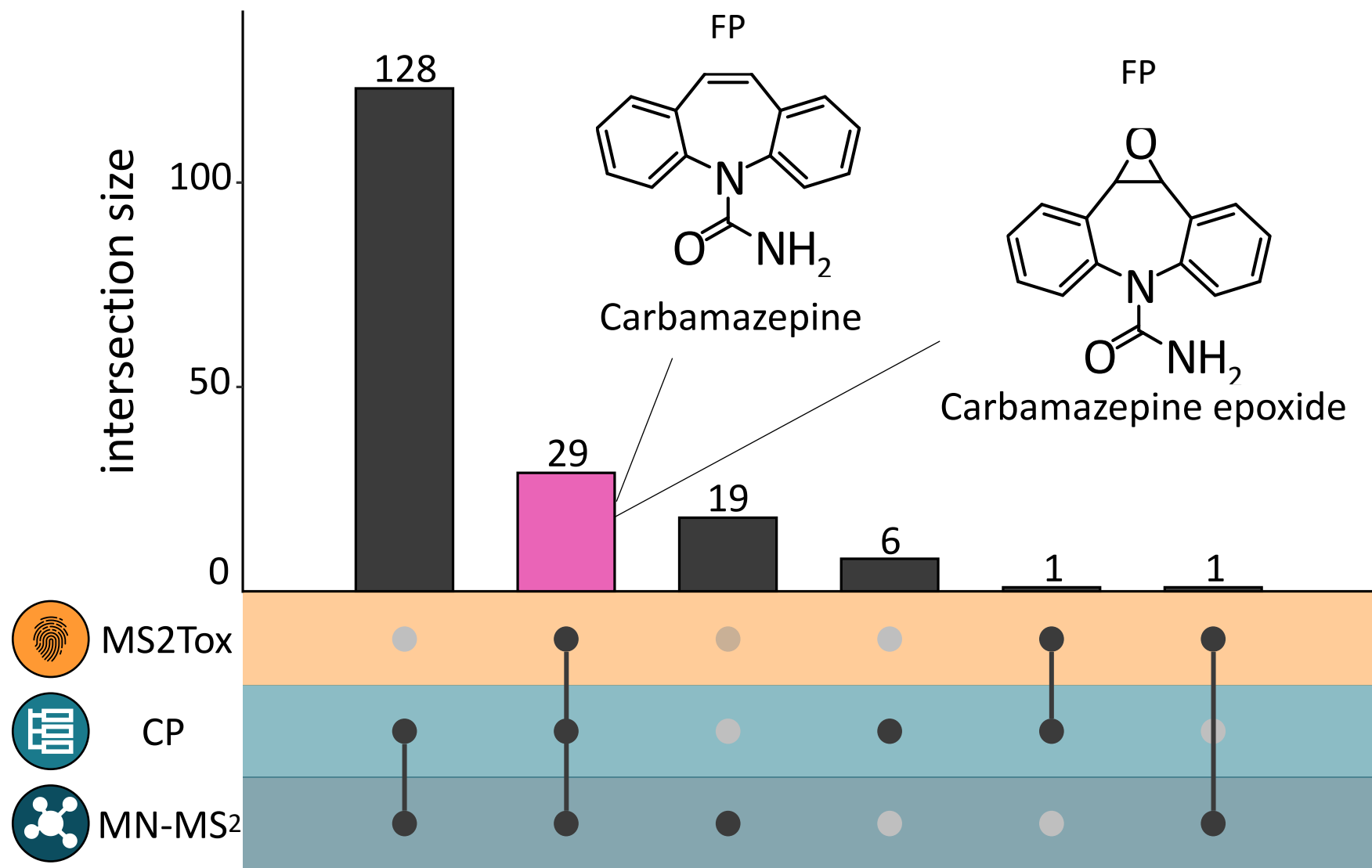
Wastewater Application



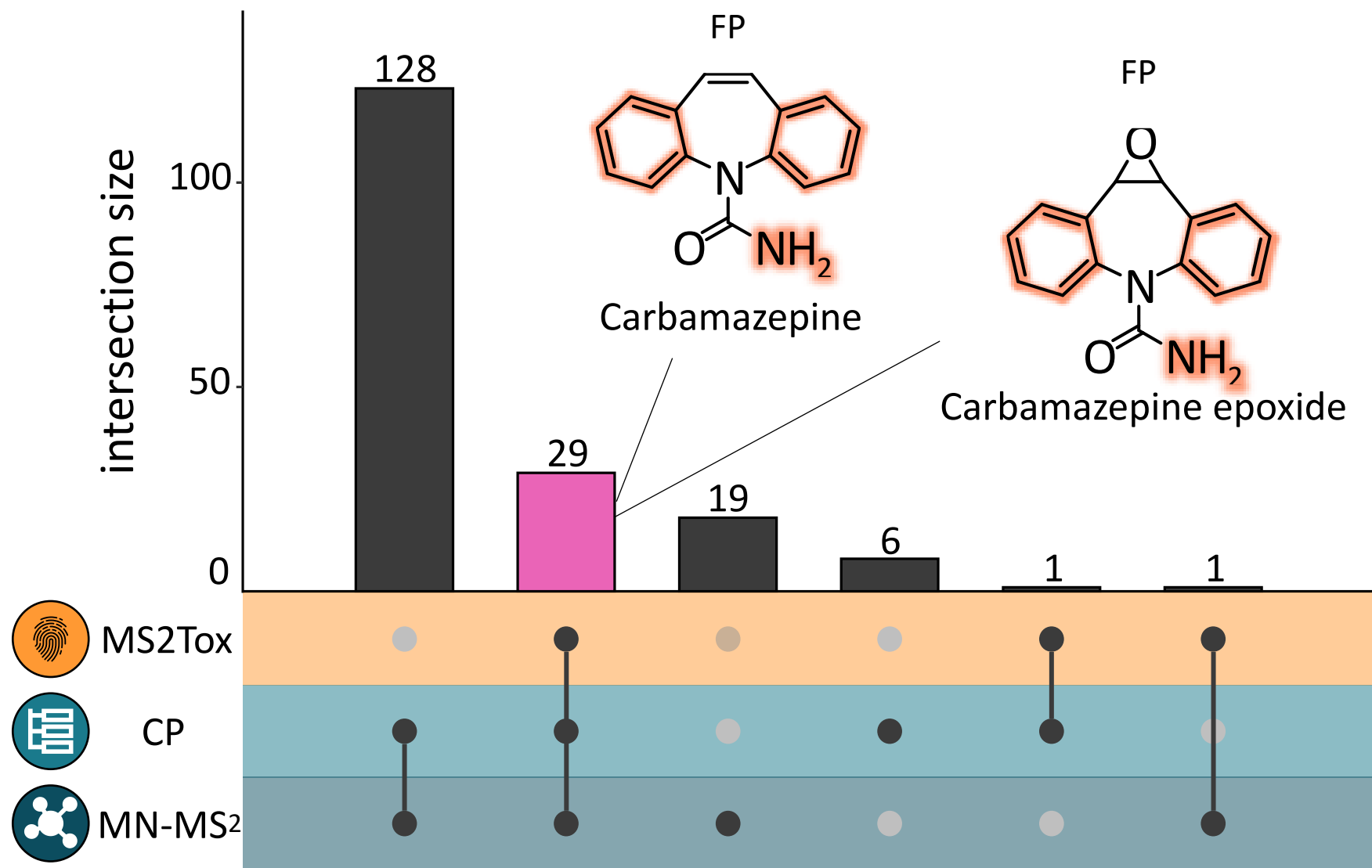
Wastewater Application



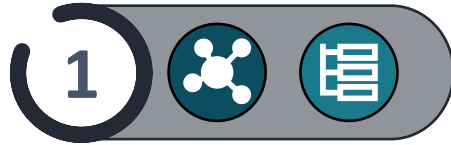
Wastewater Application



Wastewater Application



Conclusions



Development of 2 alternative approaches



Fingerprint-based models achieved lowest $FPR_{TPR=0.9}$



Use of complementary approaches to help
narrowing down features of interest

Acknowledgements



Stockholm
University

Kreuz lab



Ida Rahu



Ulf Norinder



VR 3R
2022-01353



yvonne.kreutzer@su.se