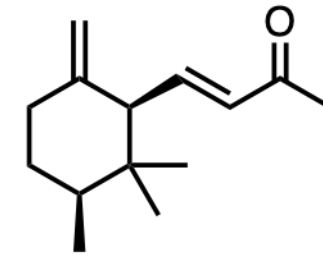


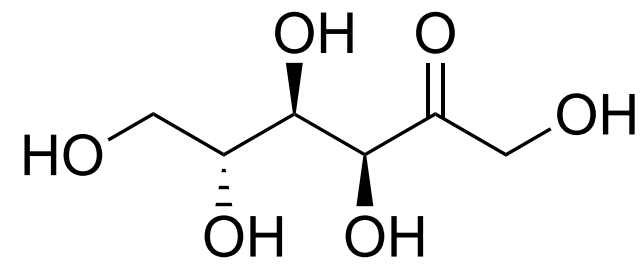
# Prior knowledge as a compass guiding untargeted metabolomics

**Jena Centre for Bioinformatics (JCB)  
Ernst-Abbe-Platz 2**

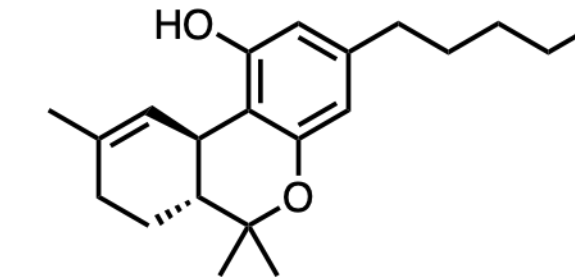
# Chemical analysis



Volatiles

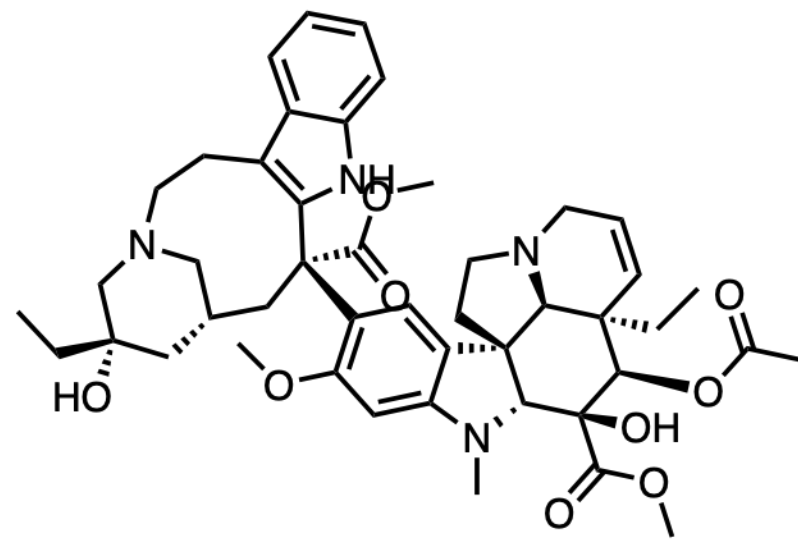


Sugars

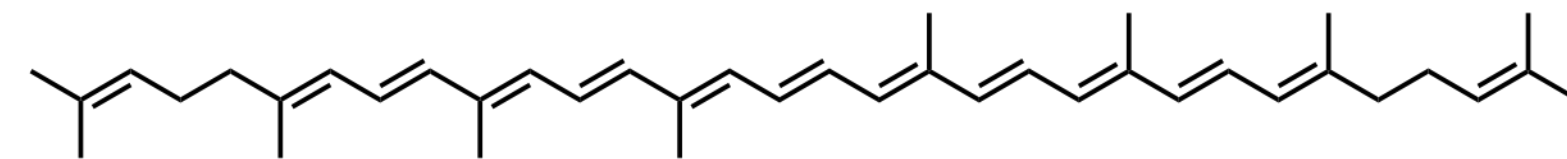


Exposome

Sample

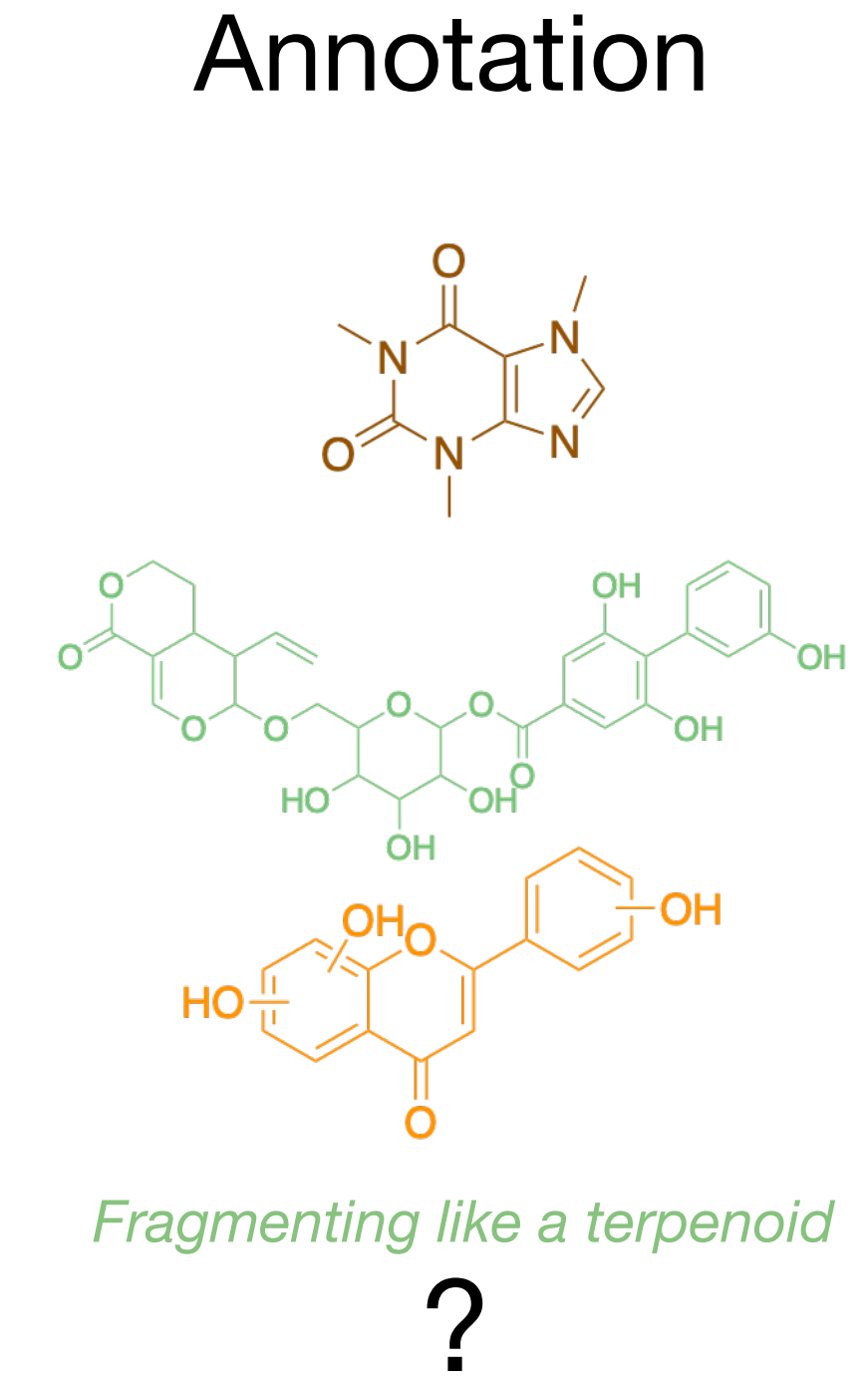
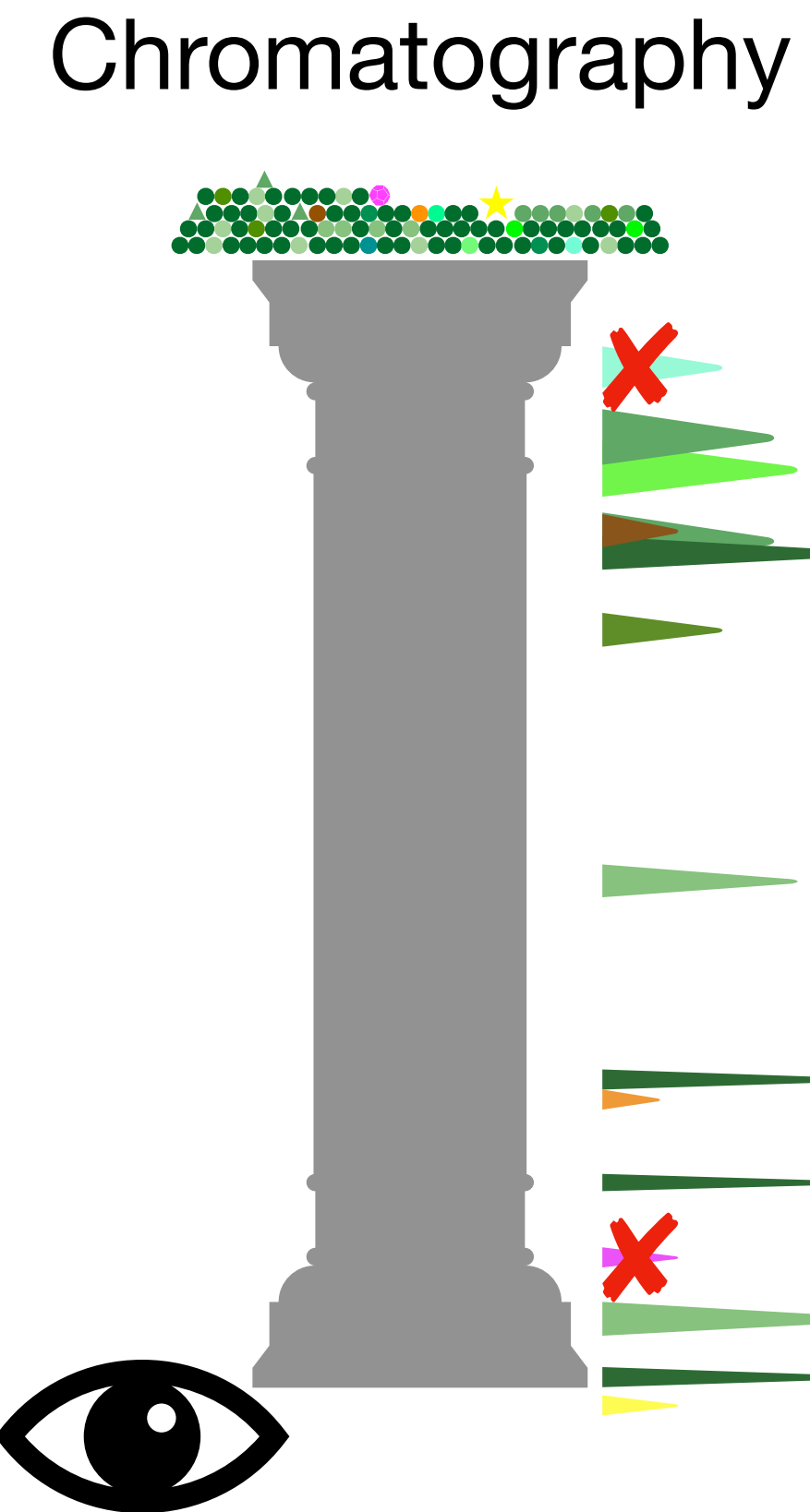
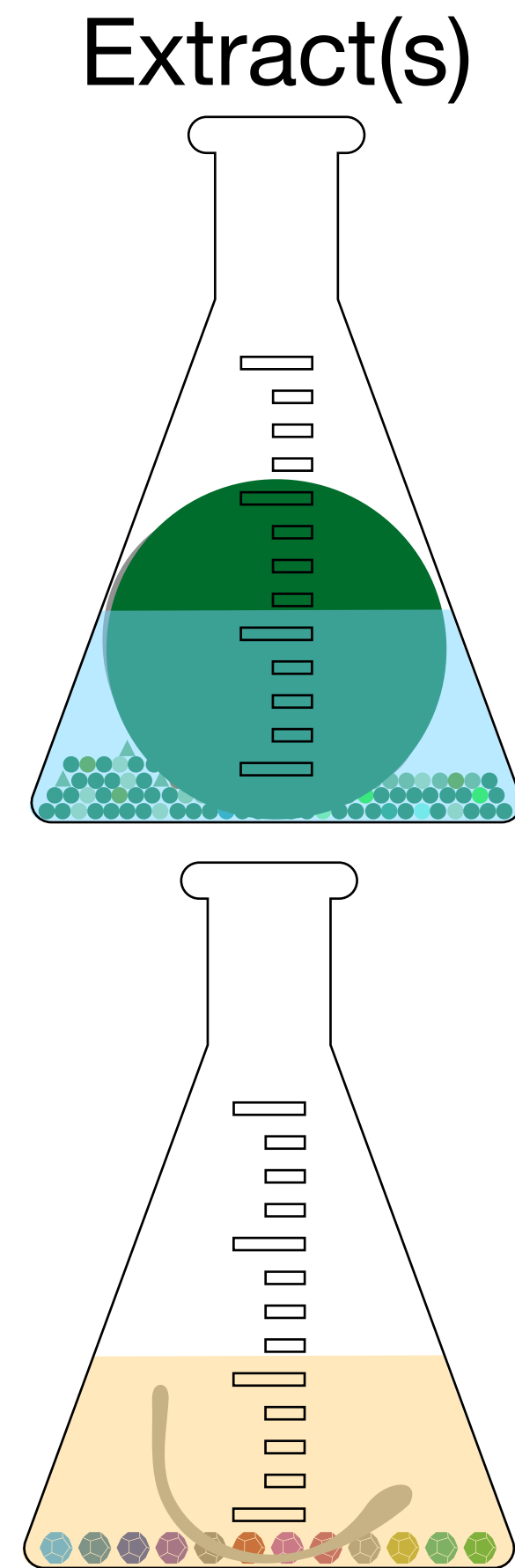
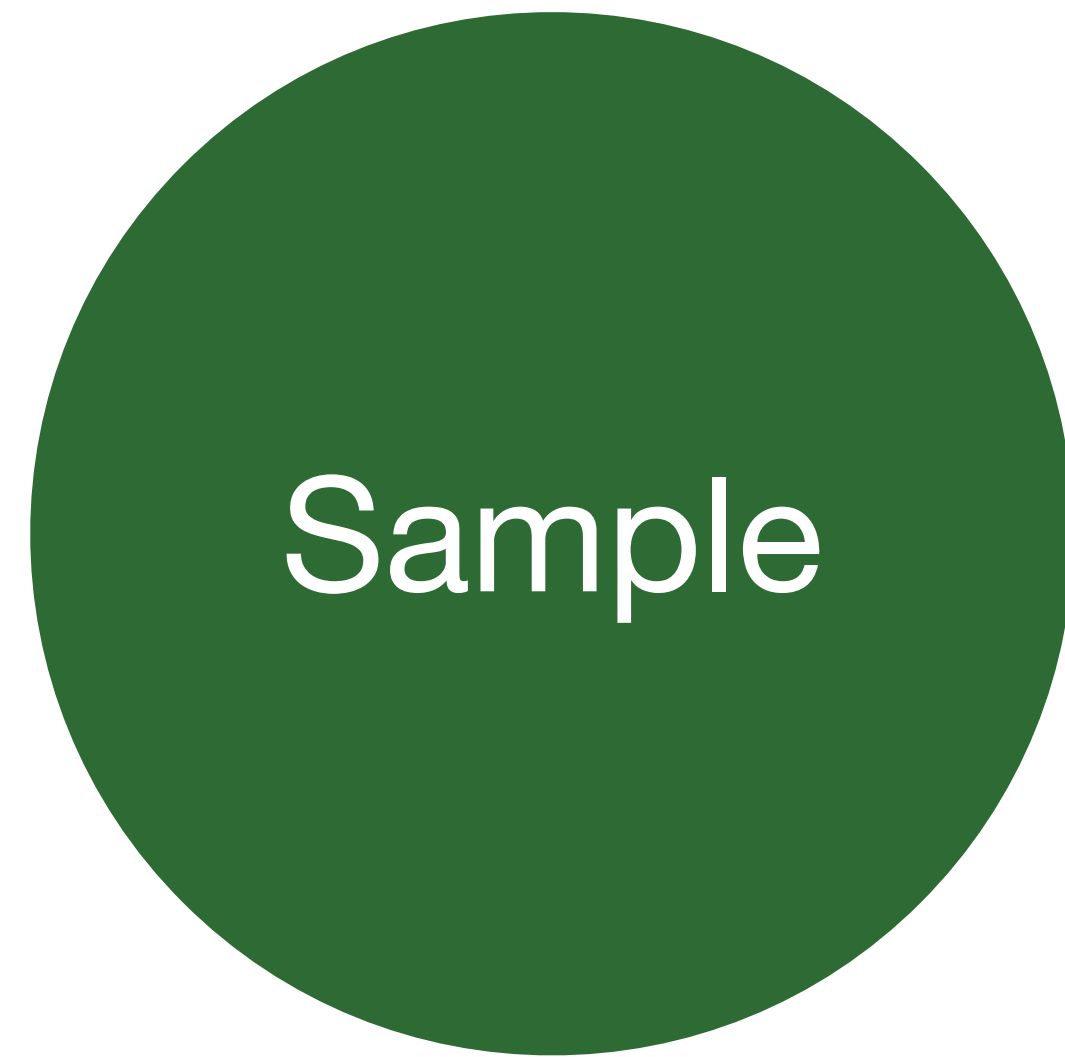


Natural products

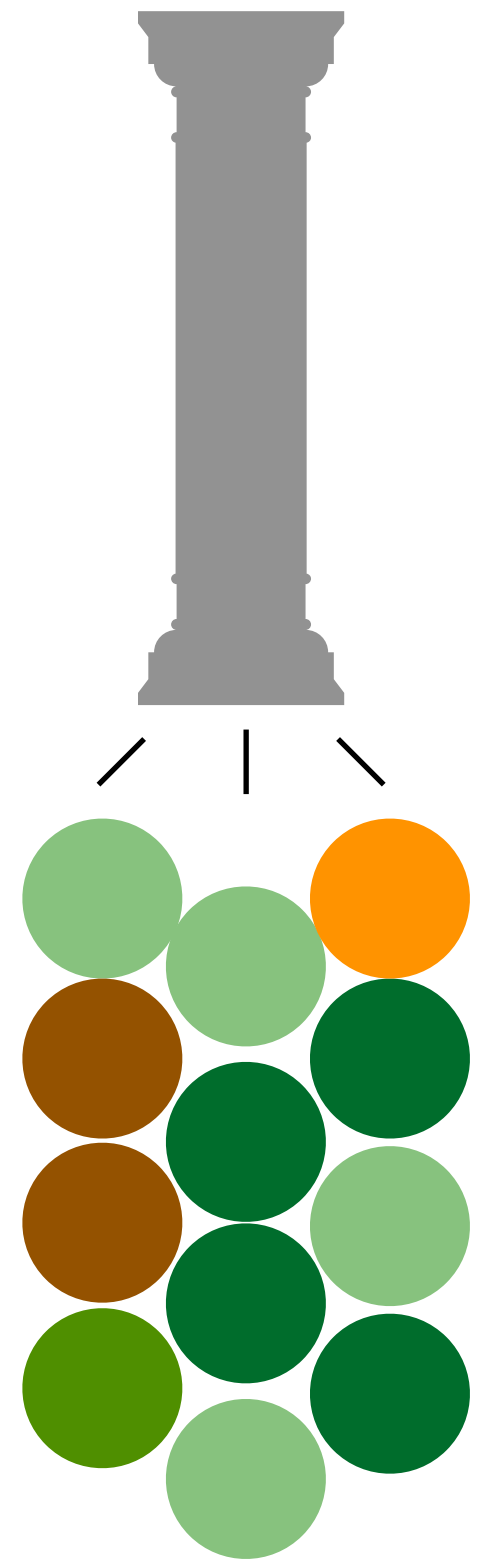


Lipids

# Chemical analysis

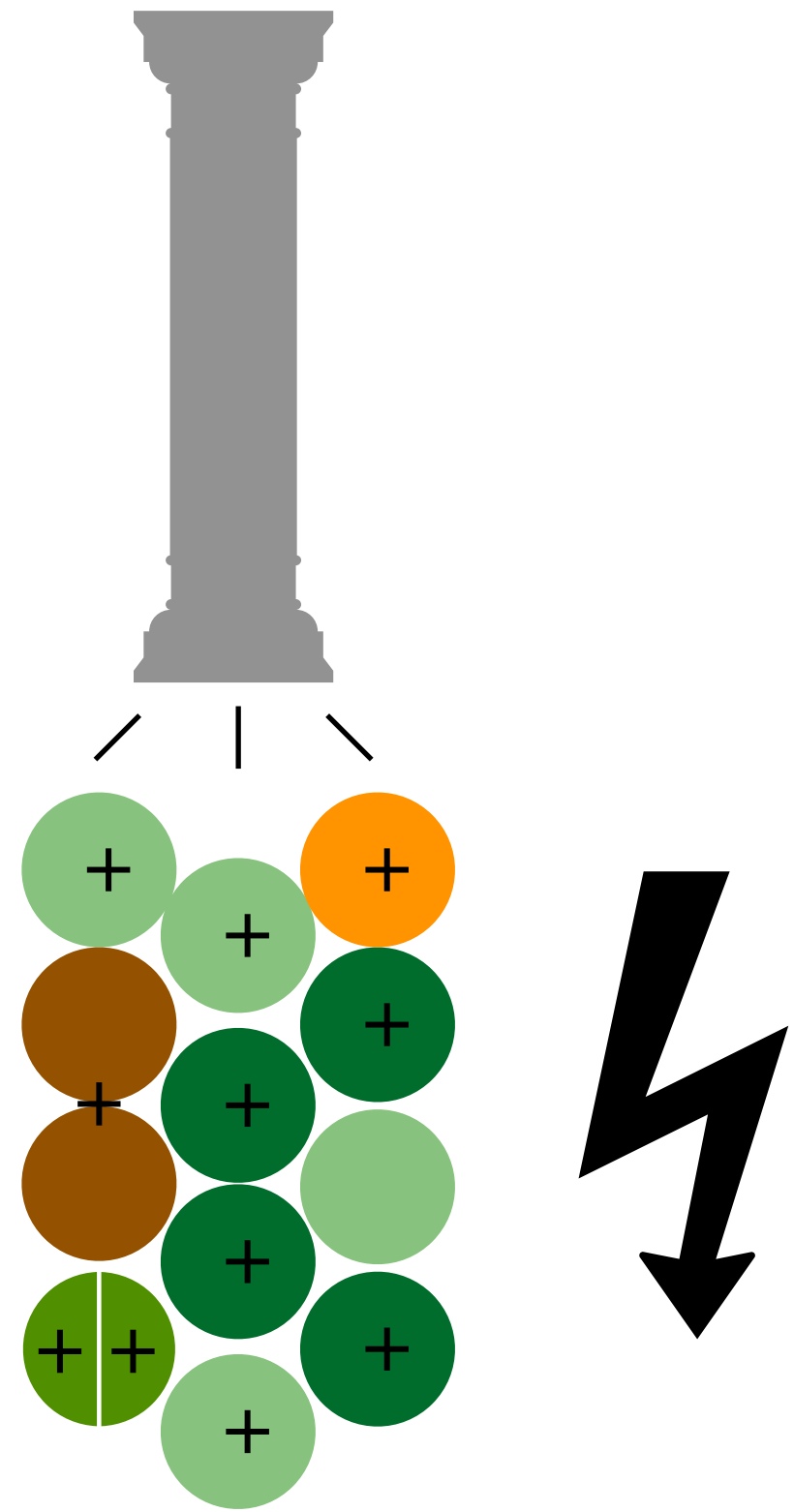


# Electrospray ionization mass spectrometry

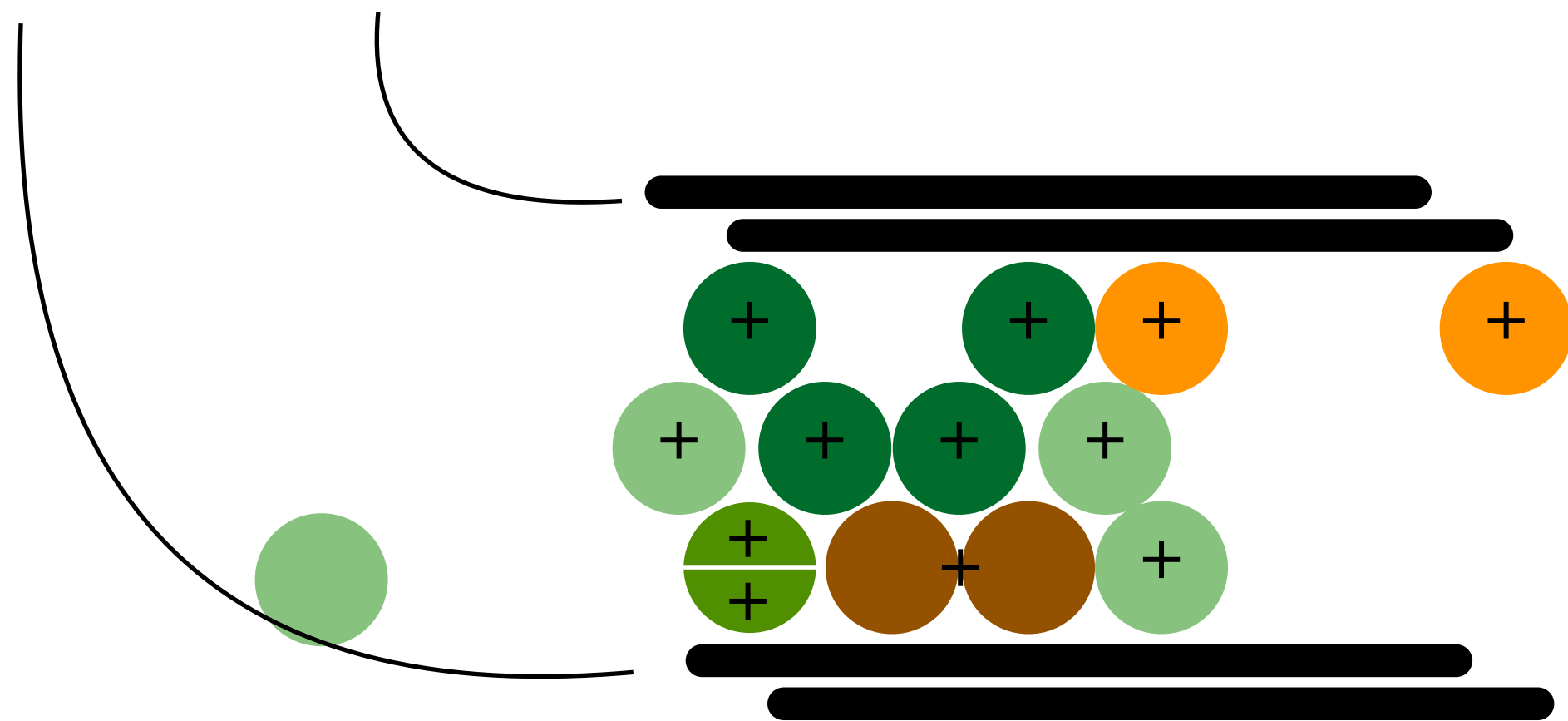




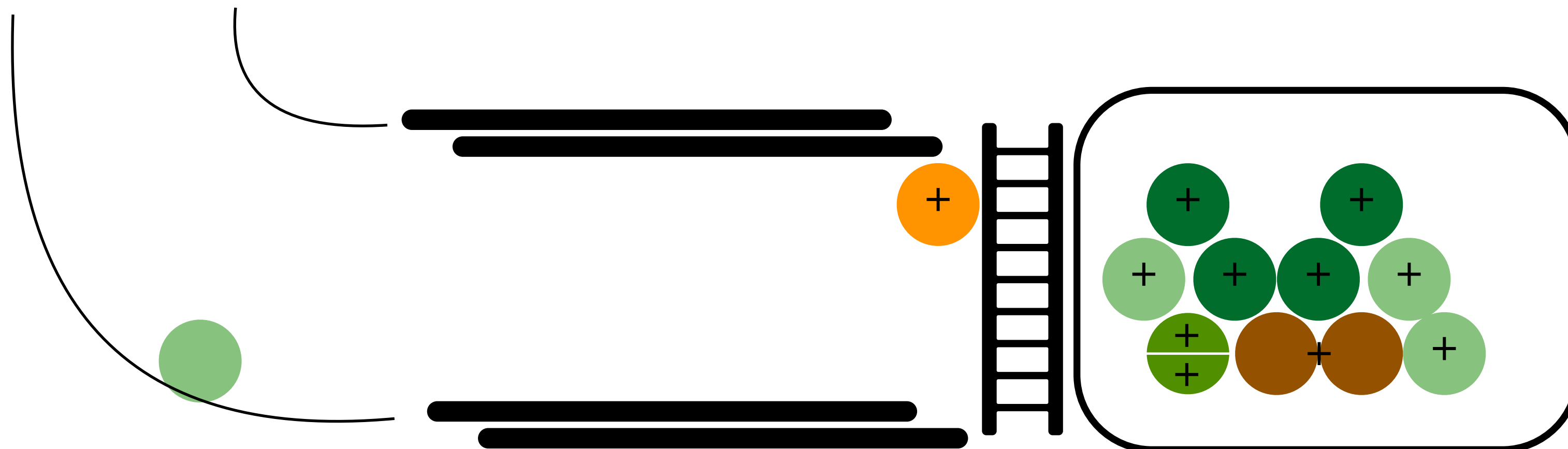
# Electrospray ionization mass spectrometry



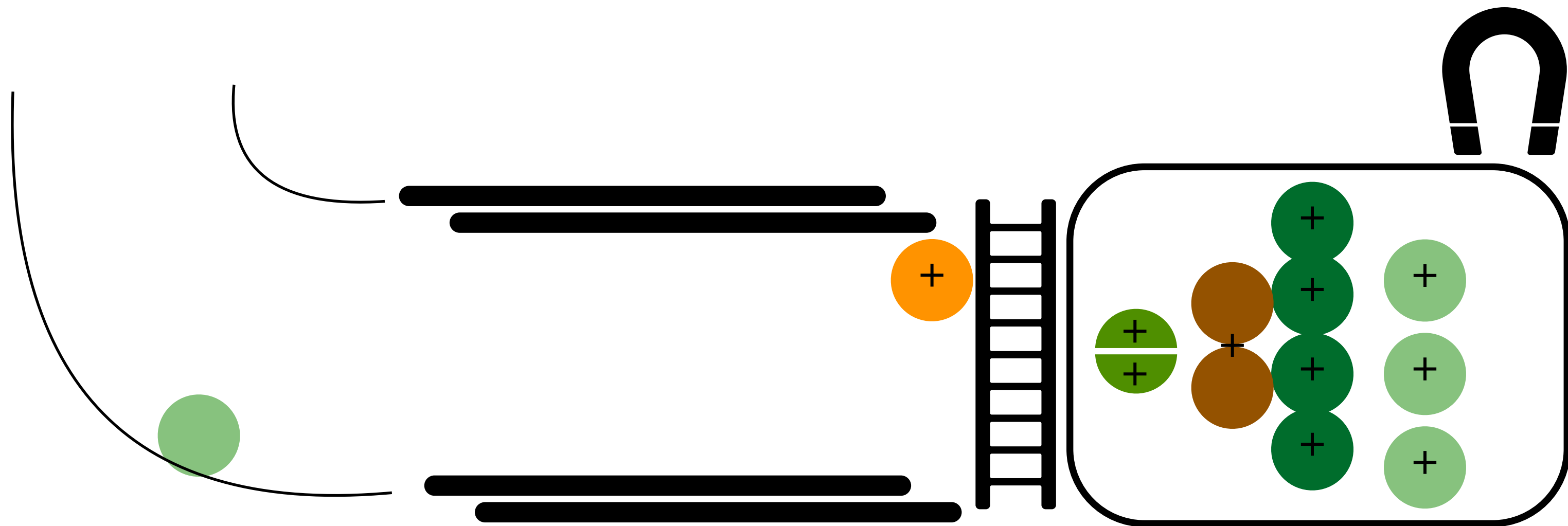
# Electrospray ionization mass spectrometry



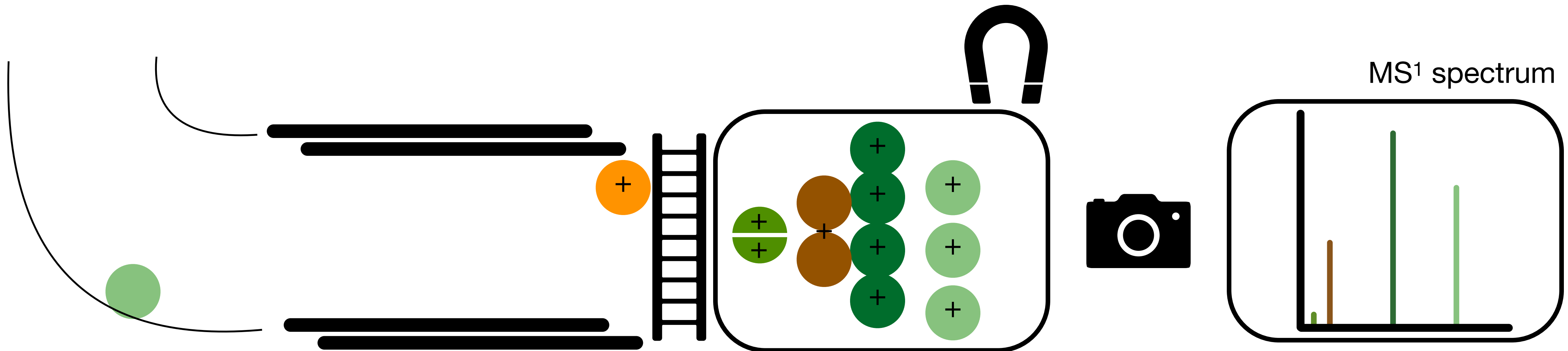
# Electrospray ionization mass spectrometry



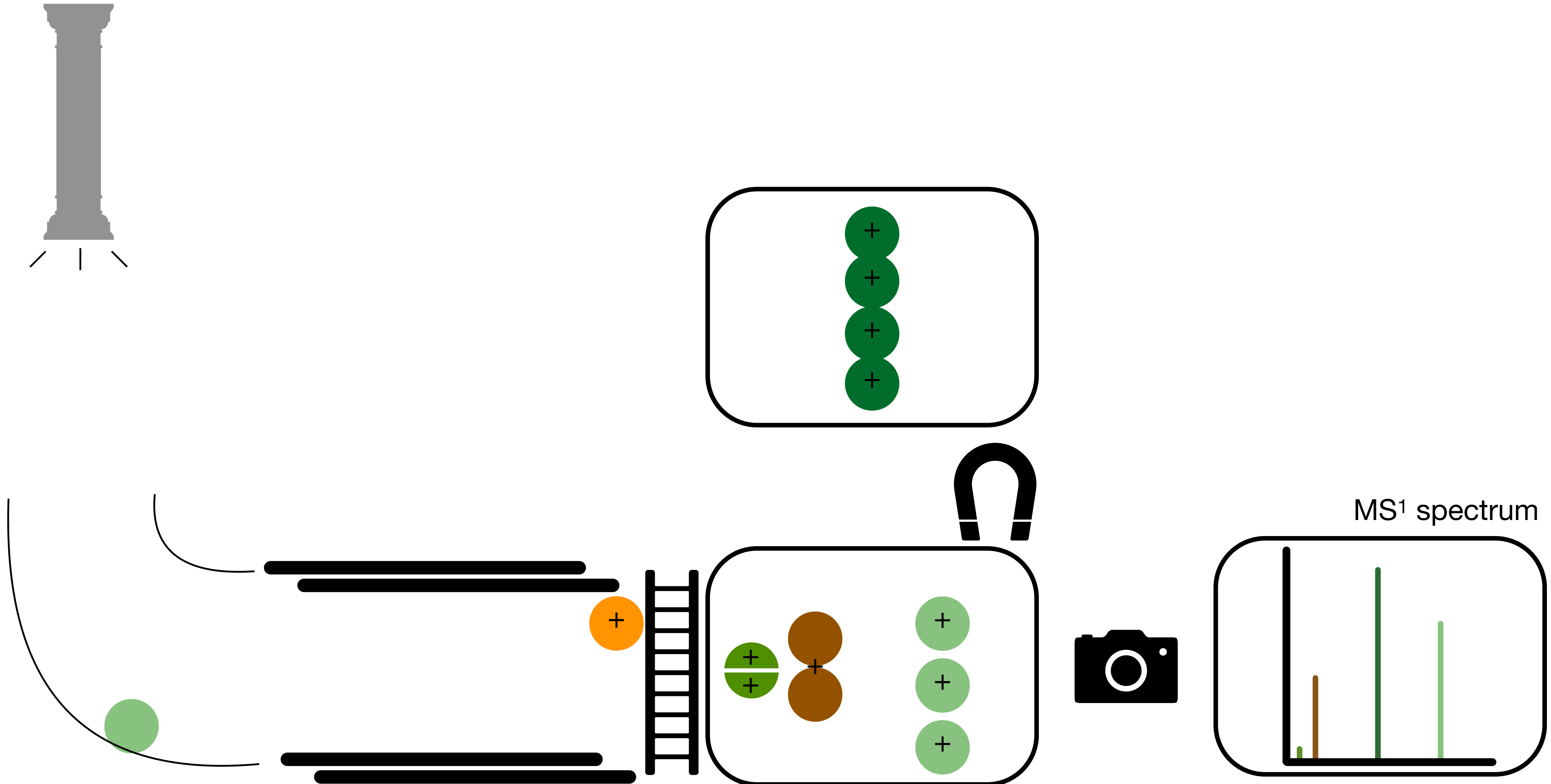
# Electrospray ionization mass spectrometry



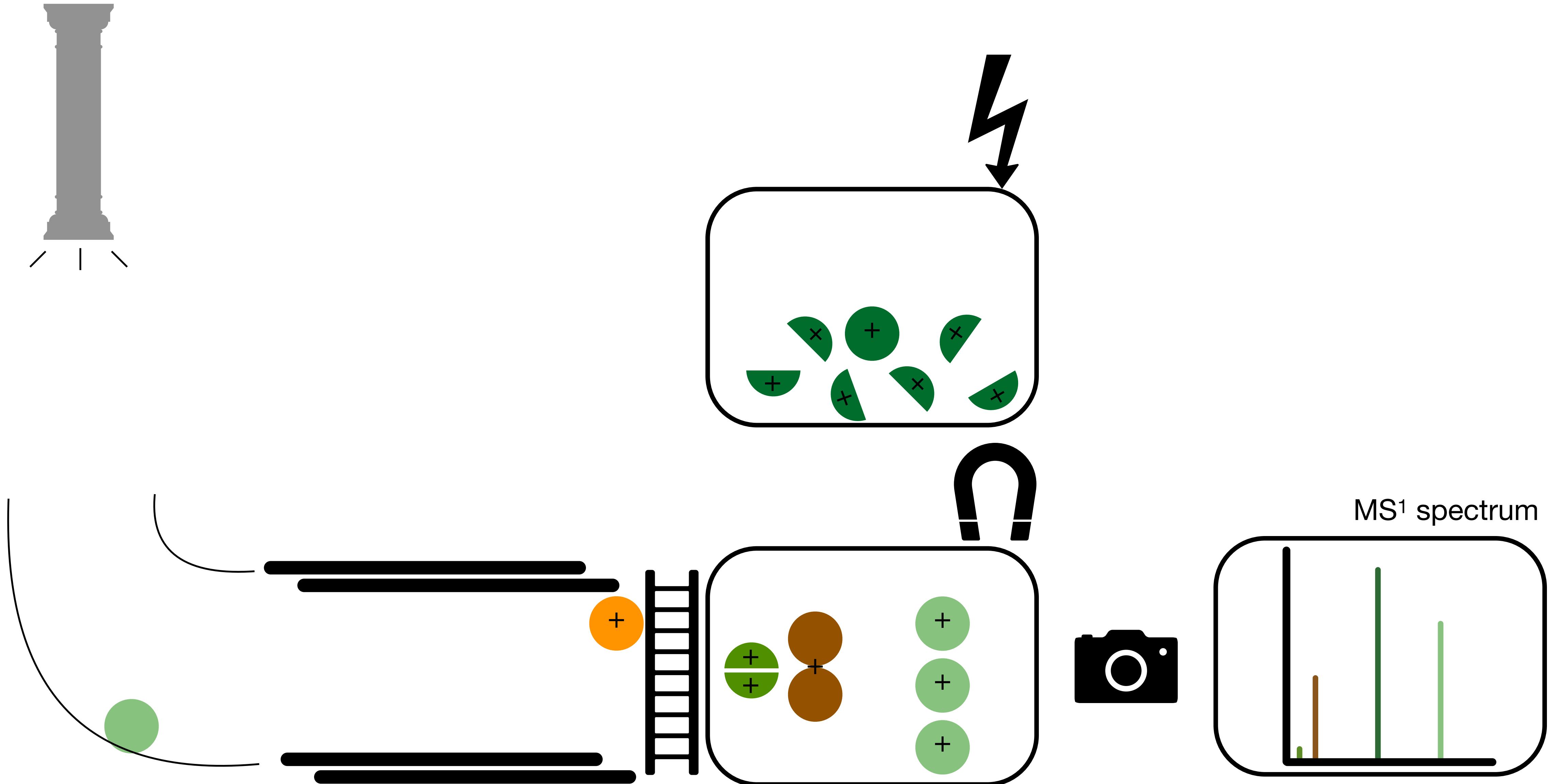
# Electrospray ionization mass spectrometry



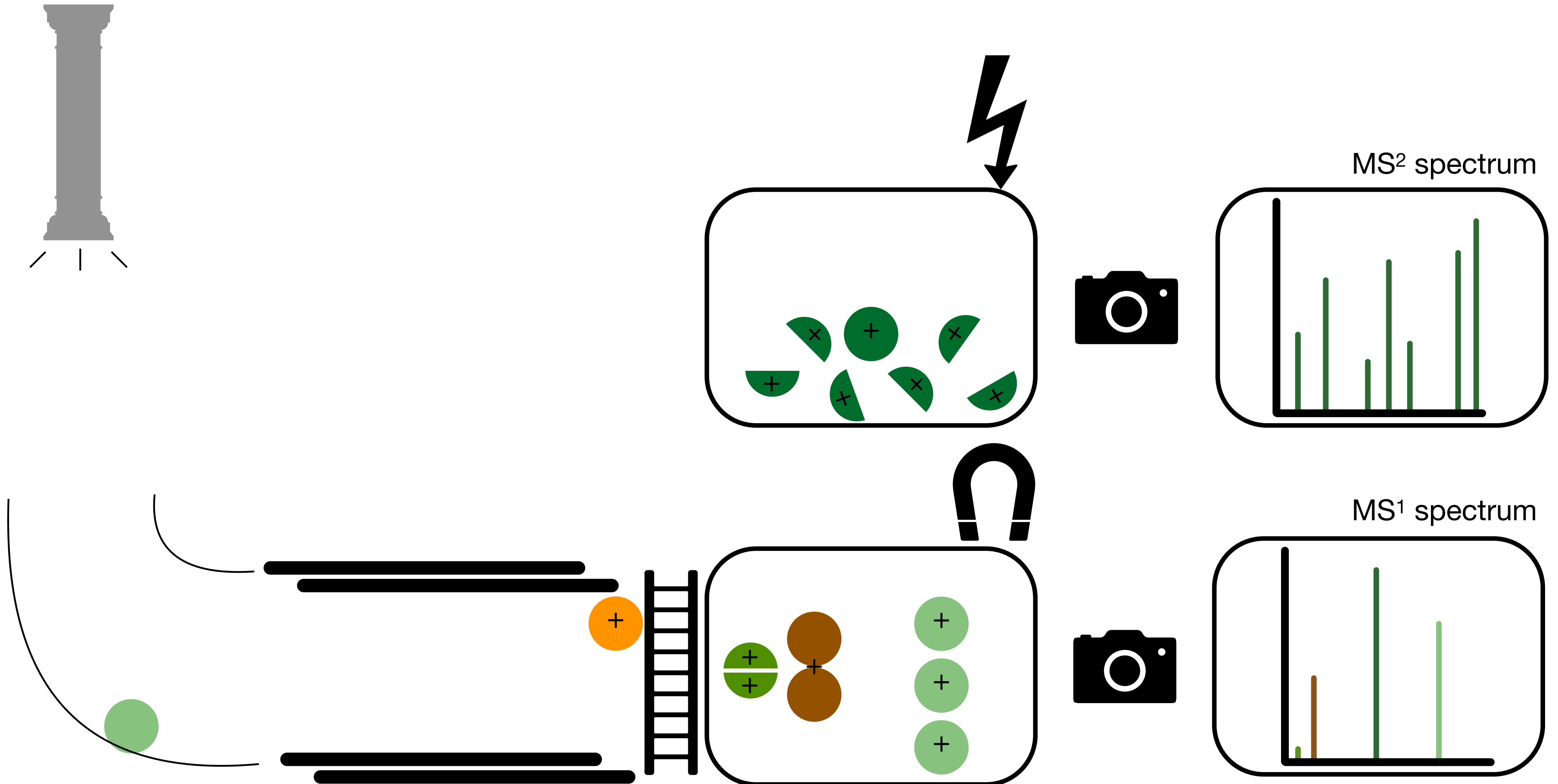
# Electrospray ionization mass spectrometry



# Electrospray ionization mass spectrometry



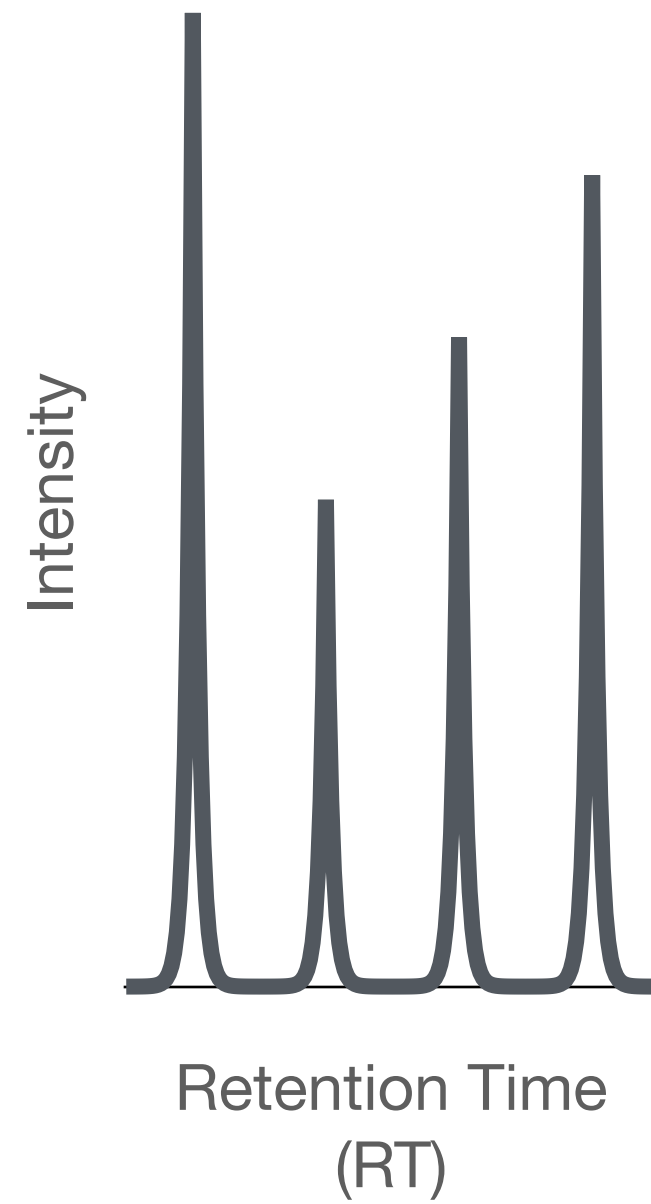
# Electrospray ionization mass spectrometry



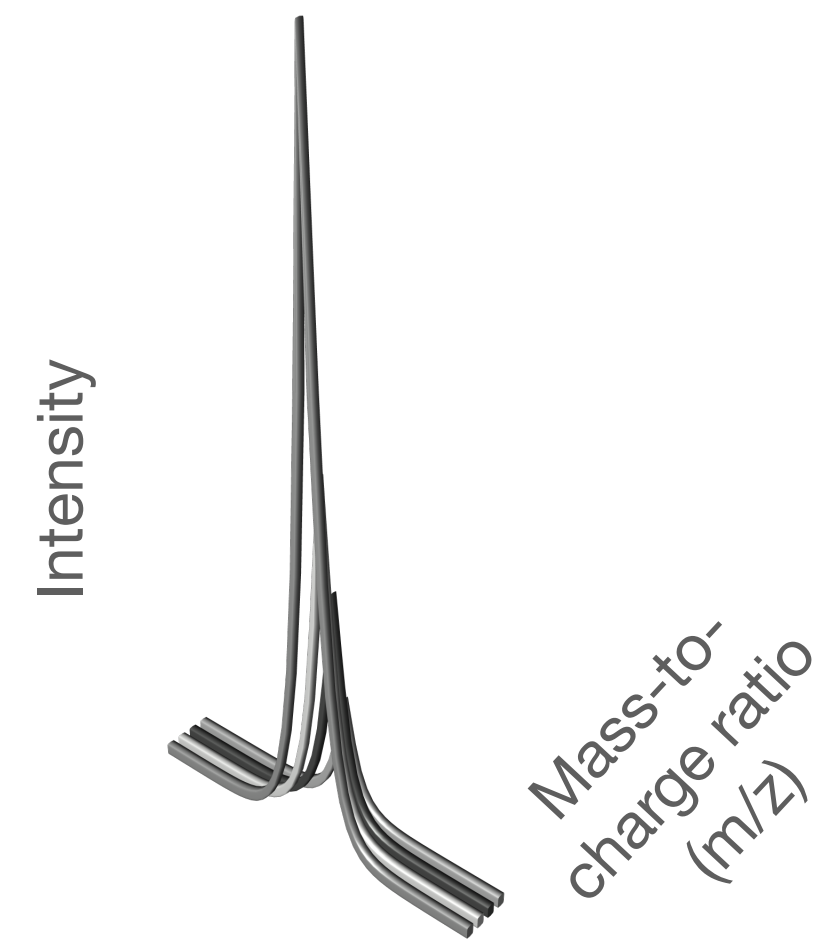


# Glossary

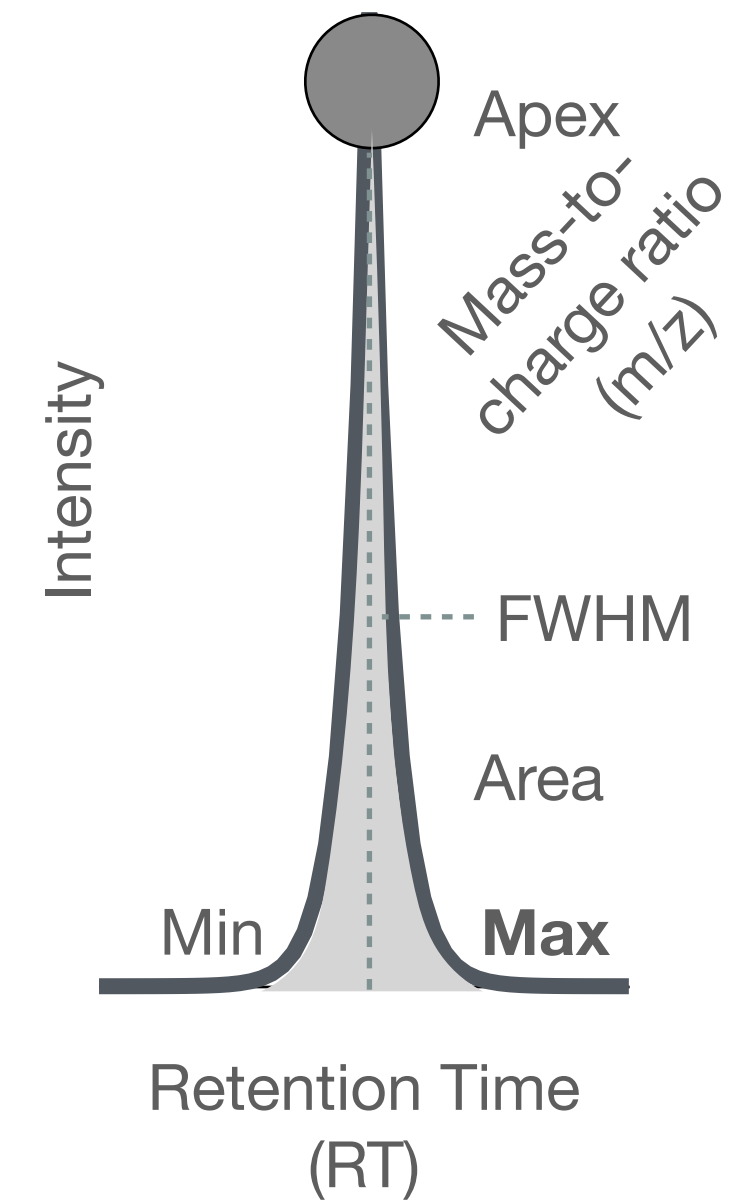
Chromatogram:  
Intensity =  $f(\text{RT})$



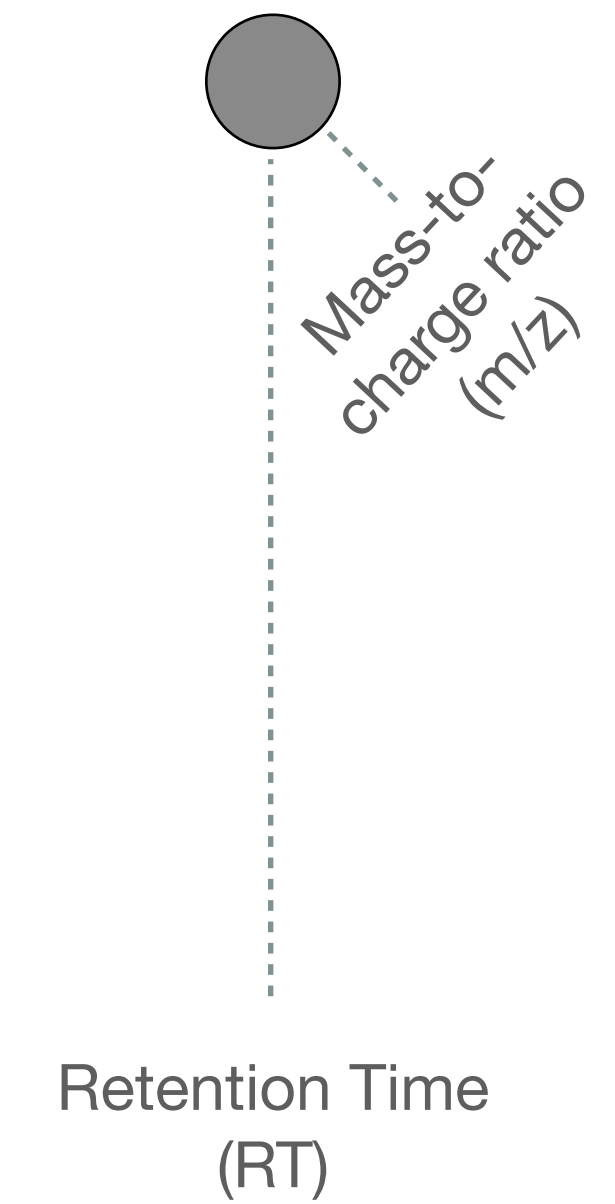
Spectrum:  
Intensity =  $f(m/z)$



Peak:  
Parameters with  
intensity, ( $m/z$ ) and RT



Feature:  
RT @  $m/z$  pointer

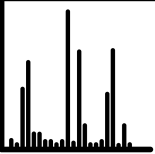


Feature list (or table):

ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

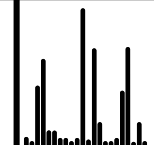
# Metabolite annotation

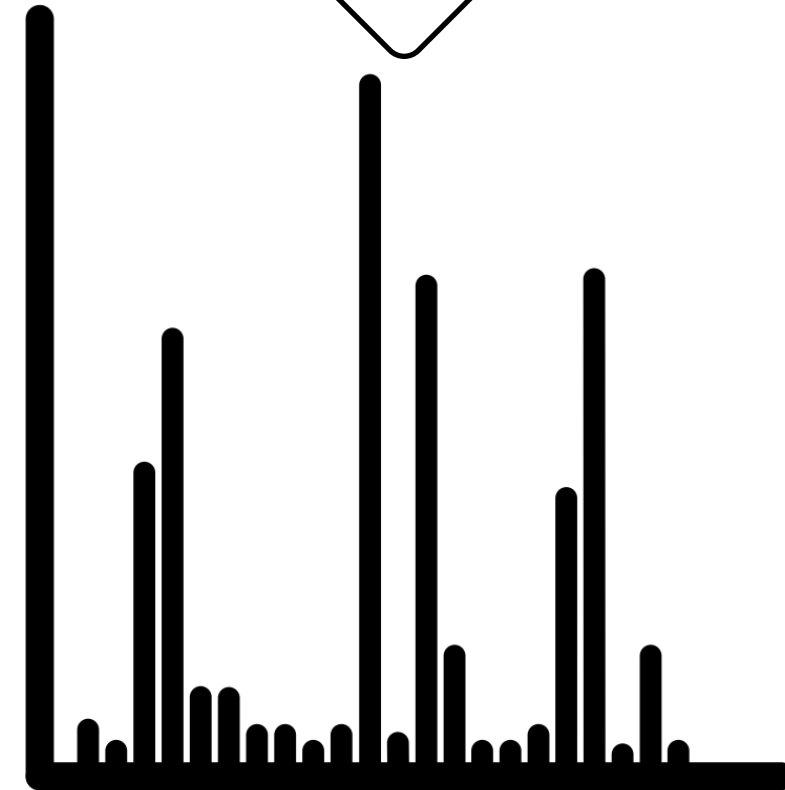
Feature list (or table):

ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

# Metabolite annotation

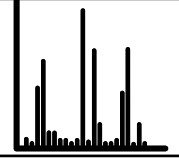
Feature list (or table):

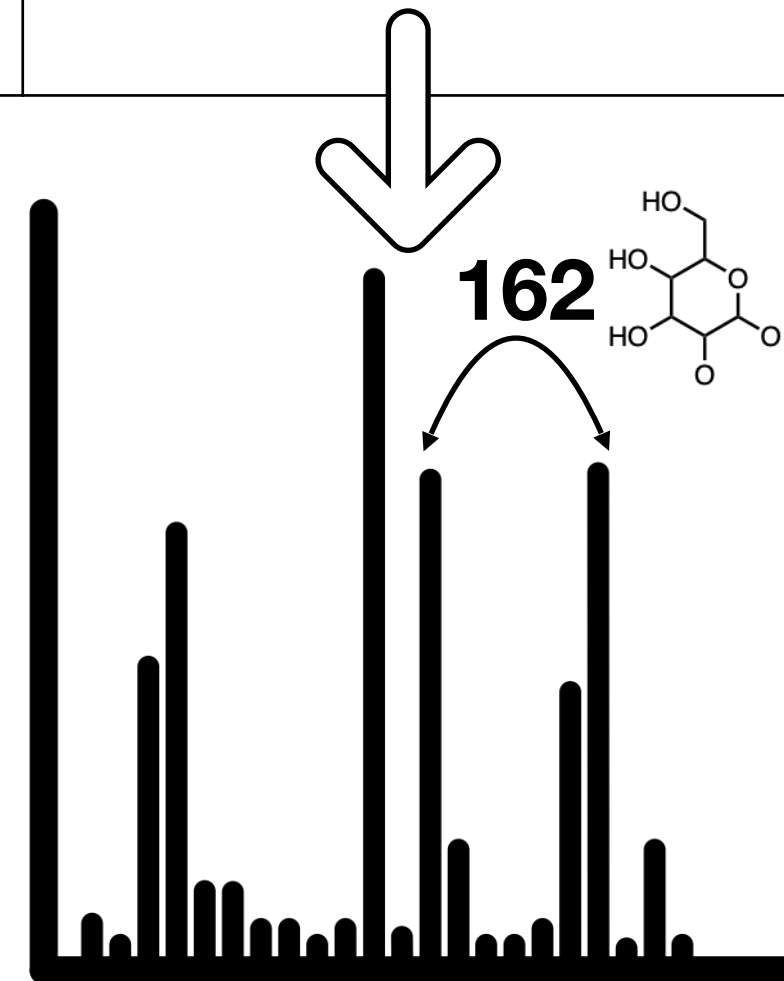
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

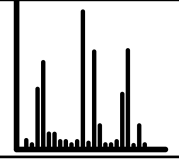
Feature list (or table):

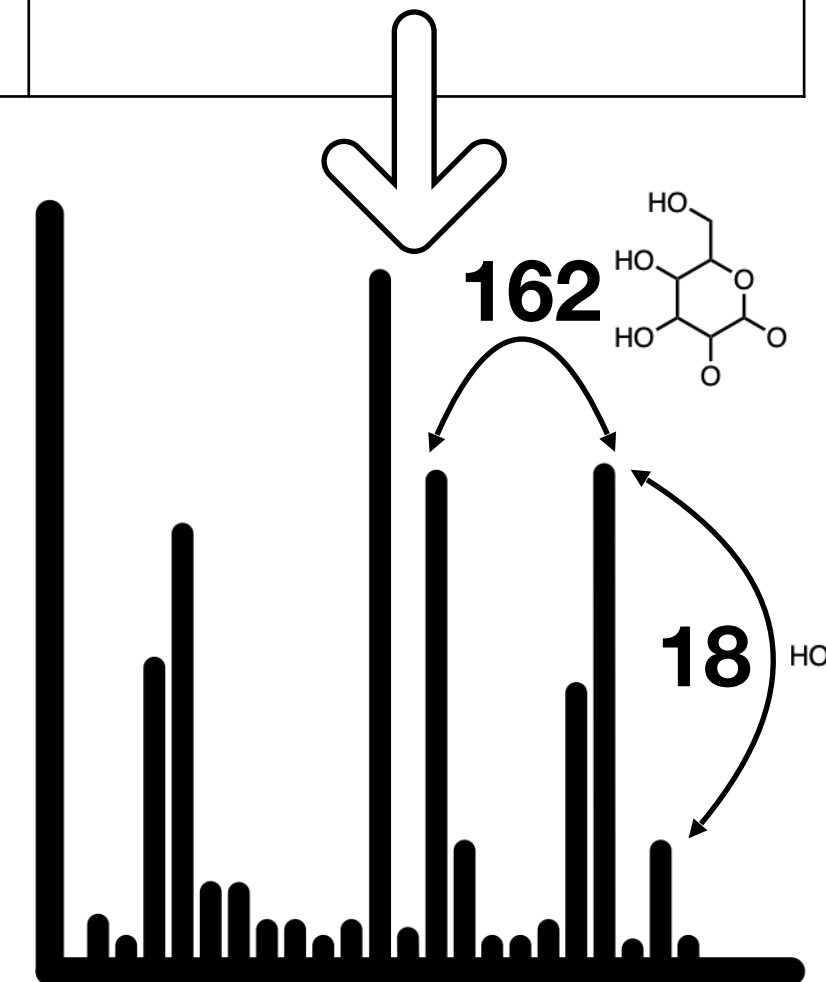
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

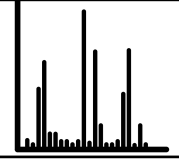
Feature list (or table):

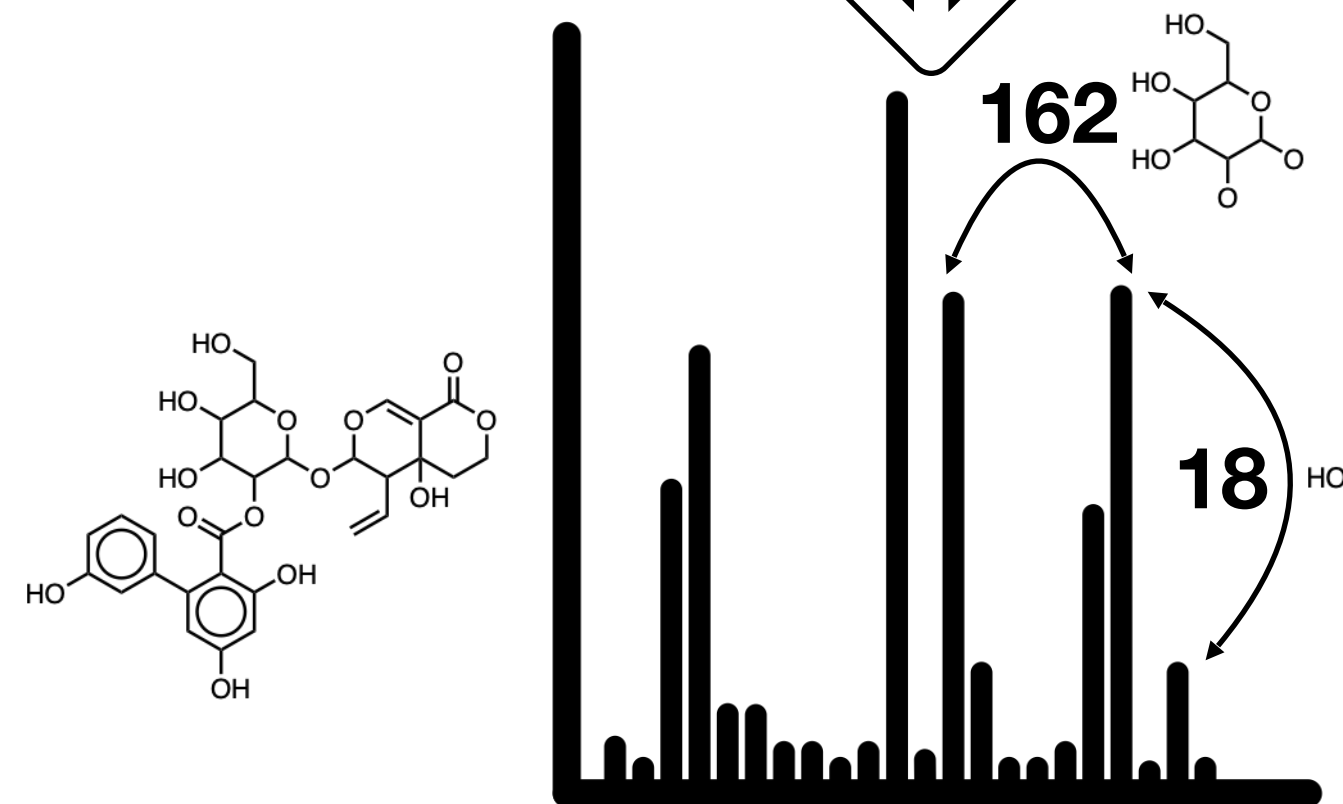
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

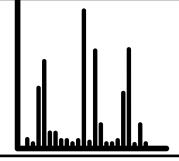
Feature list (or table):

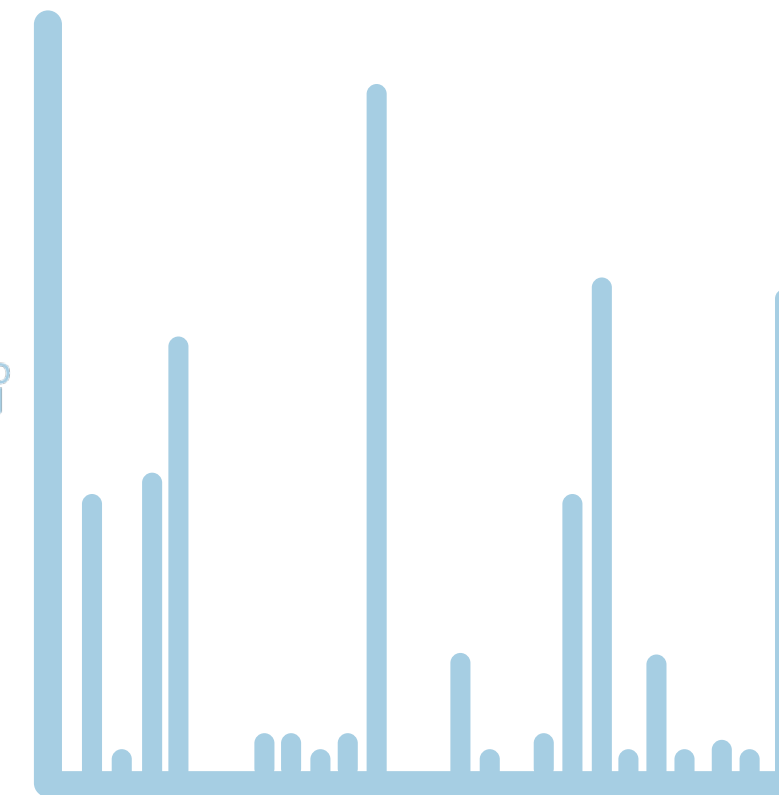
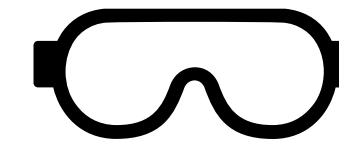
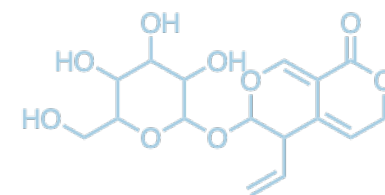
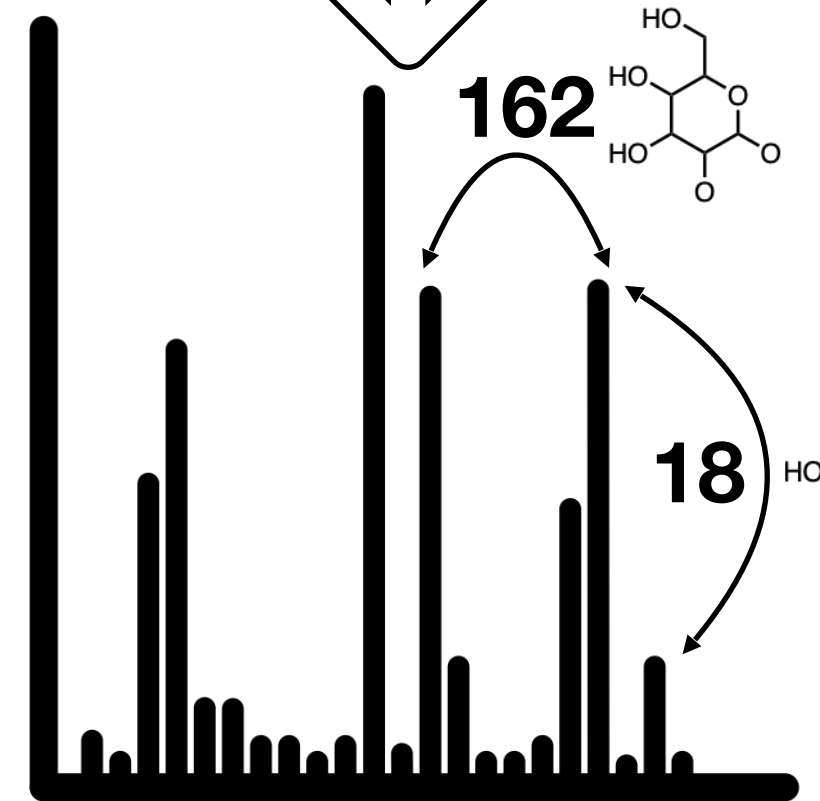
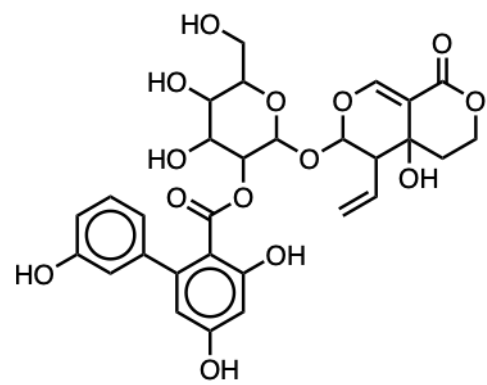
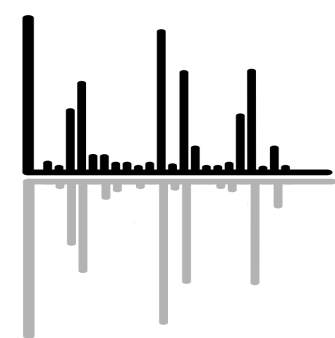
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



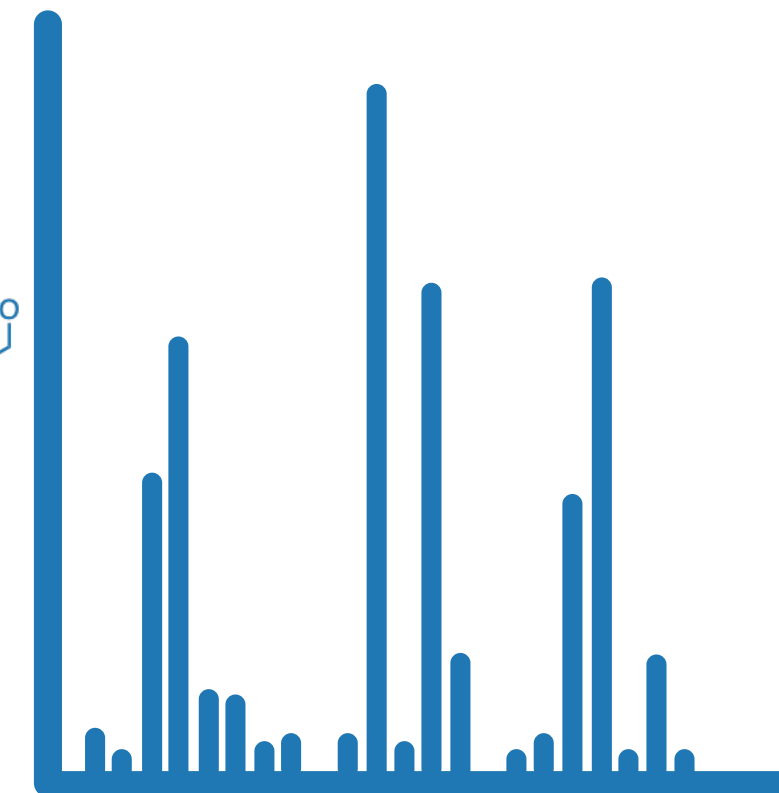
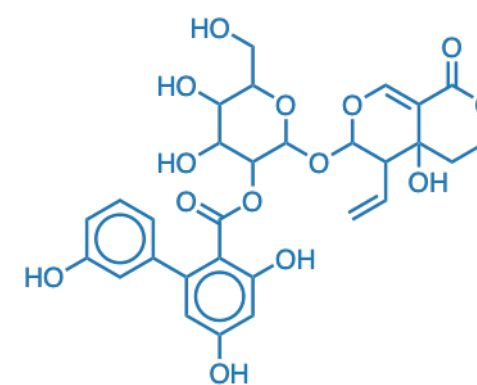
# Metabolite annotation

Feature list (or table):

ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

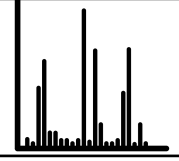


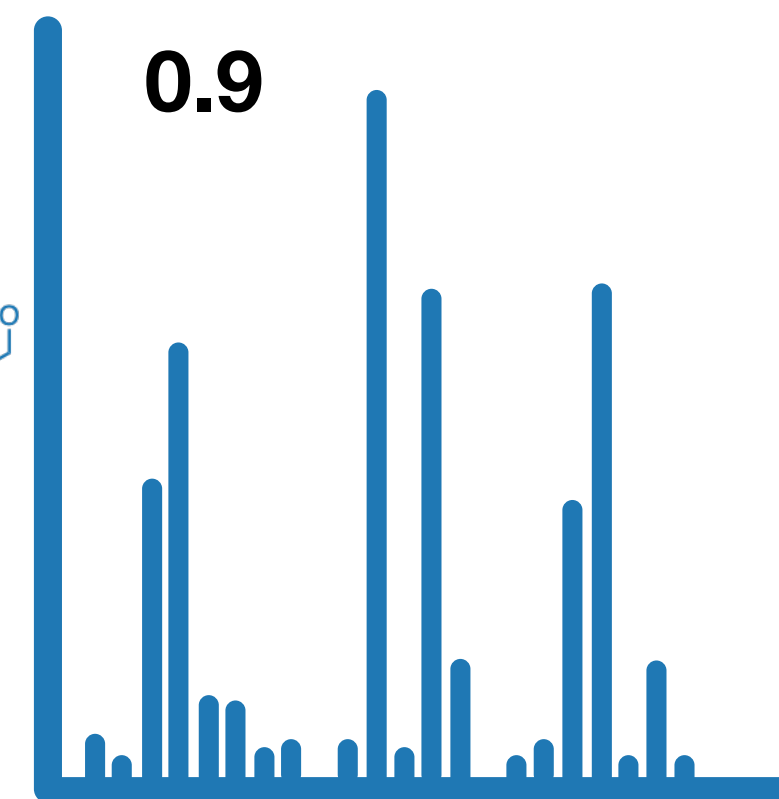
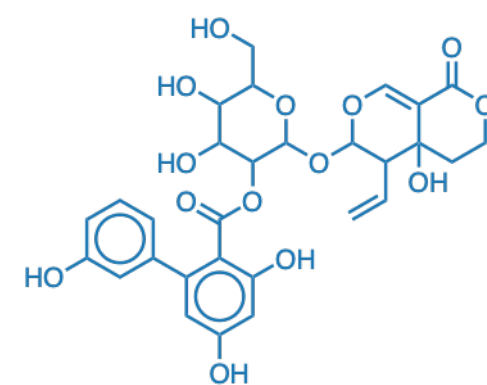
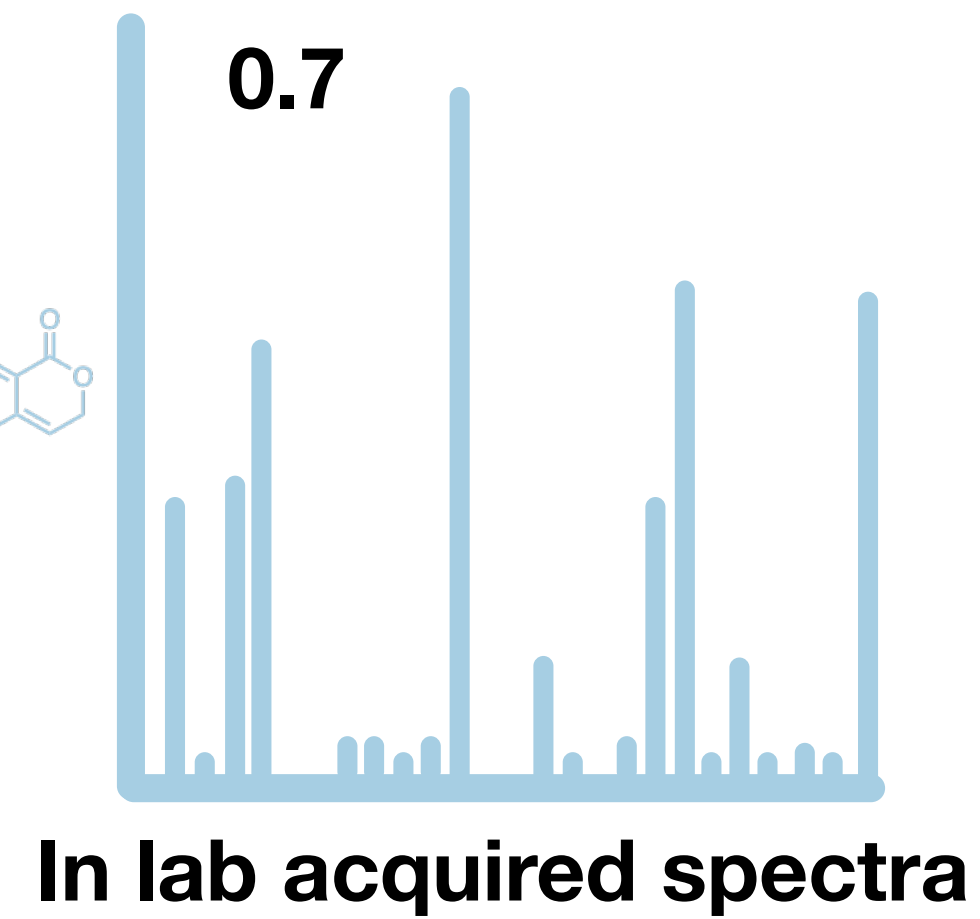
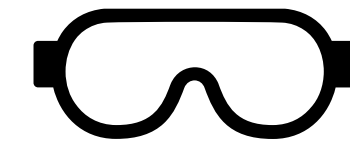
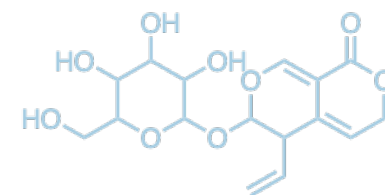
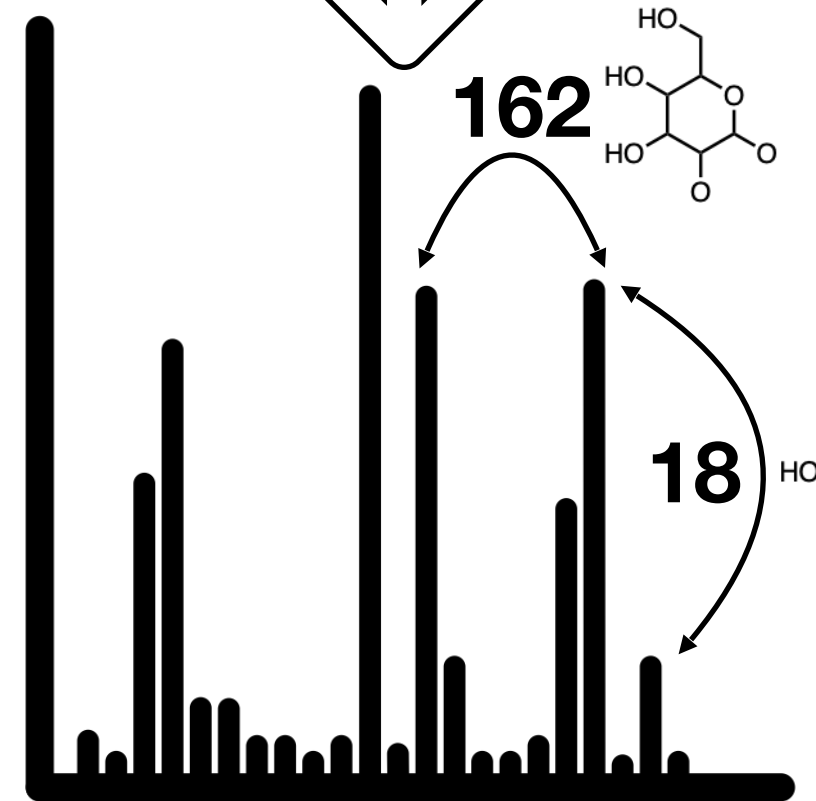
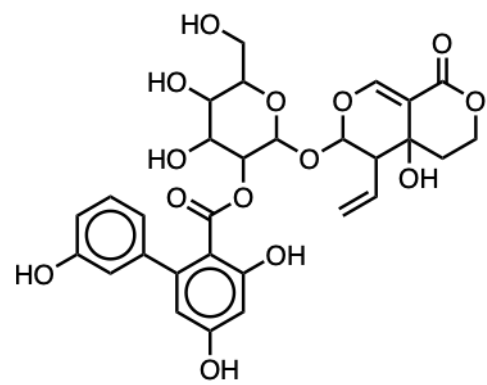
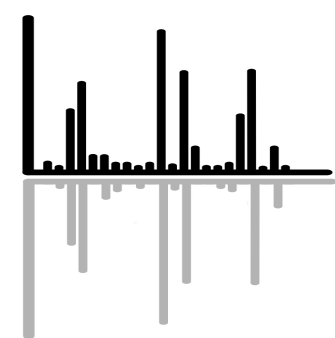
In lab acquired spectra



# Metabolite annotation

Feature list (or table):

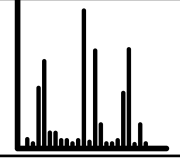
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

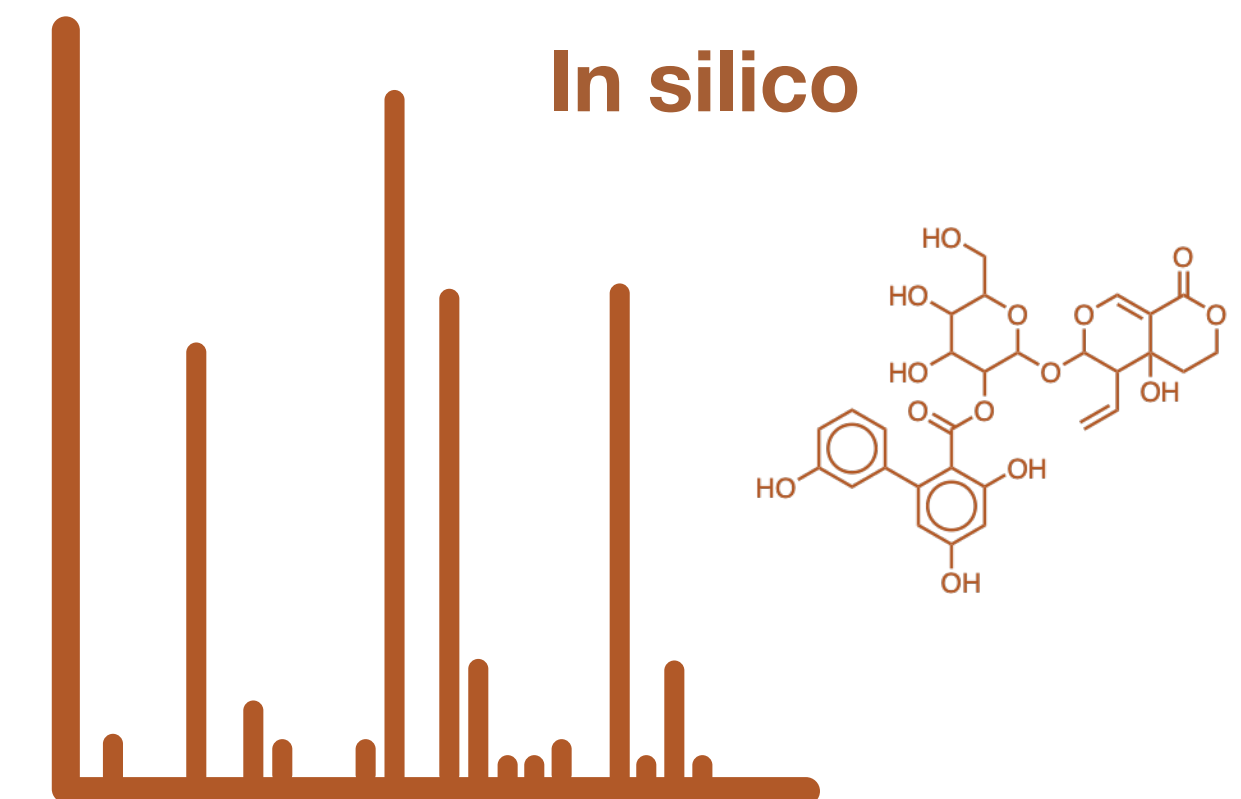
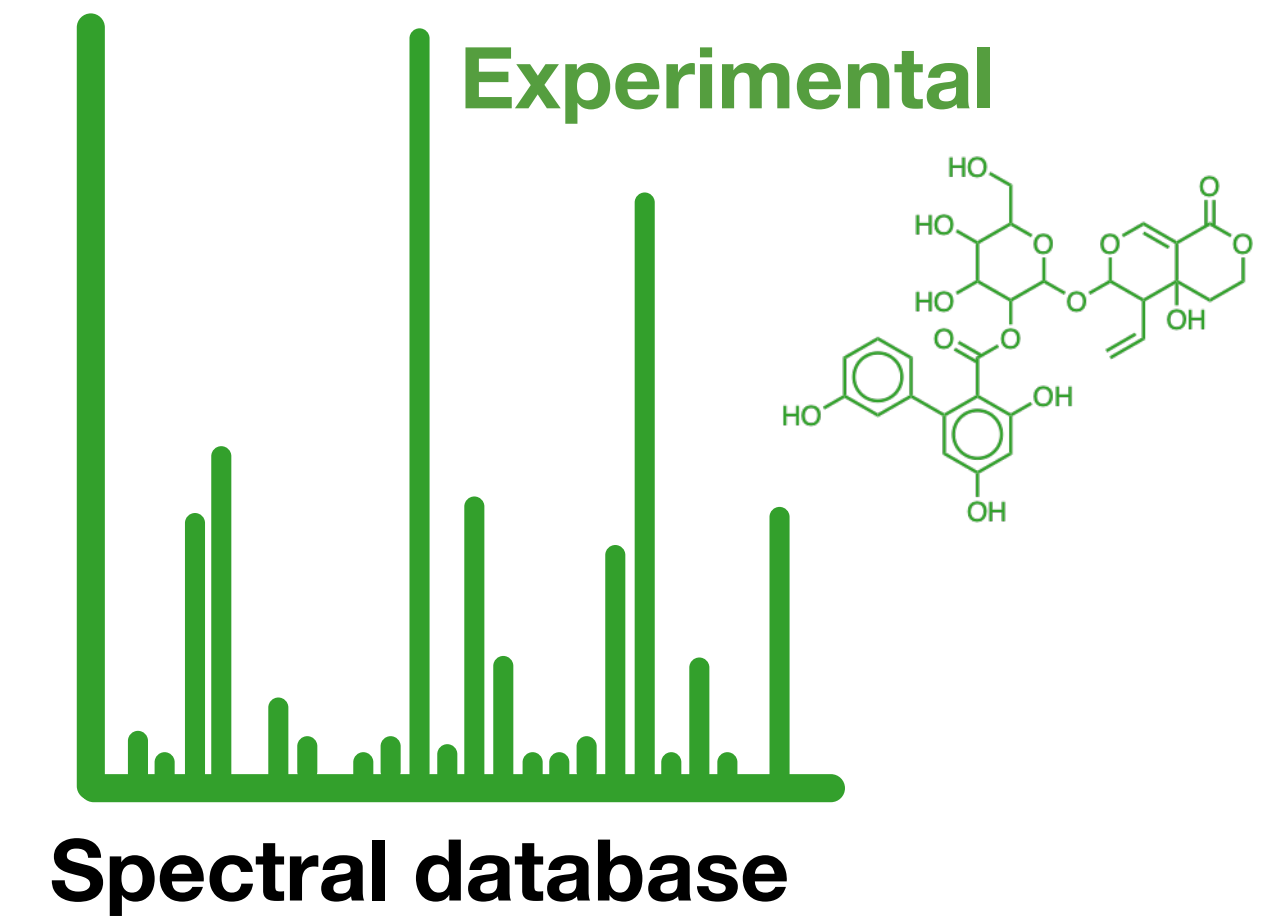
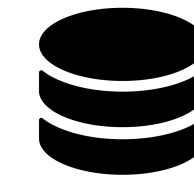
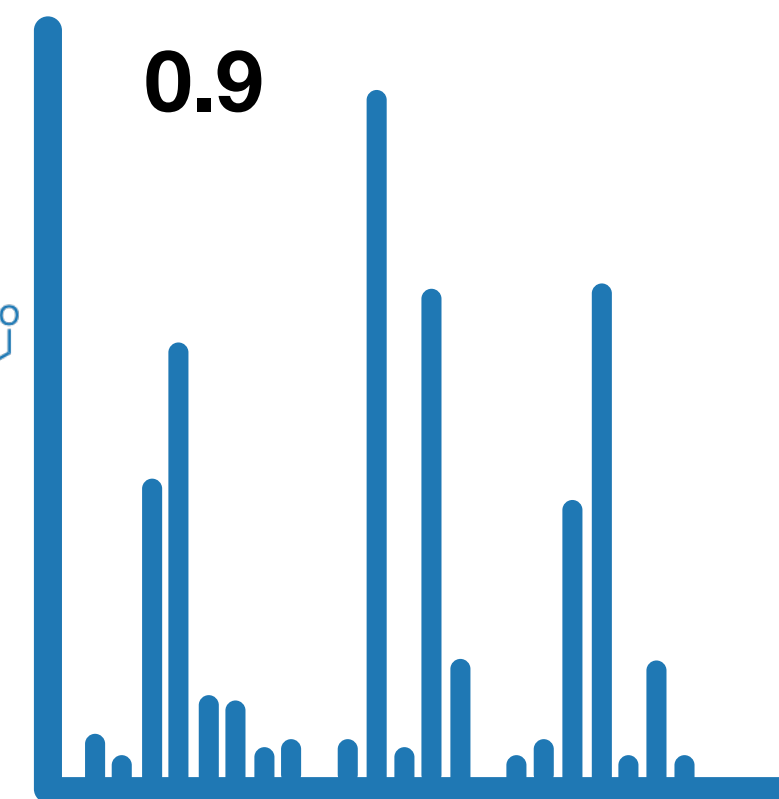
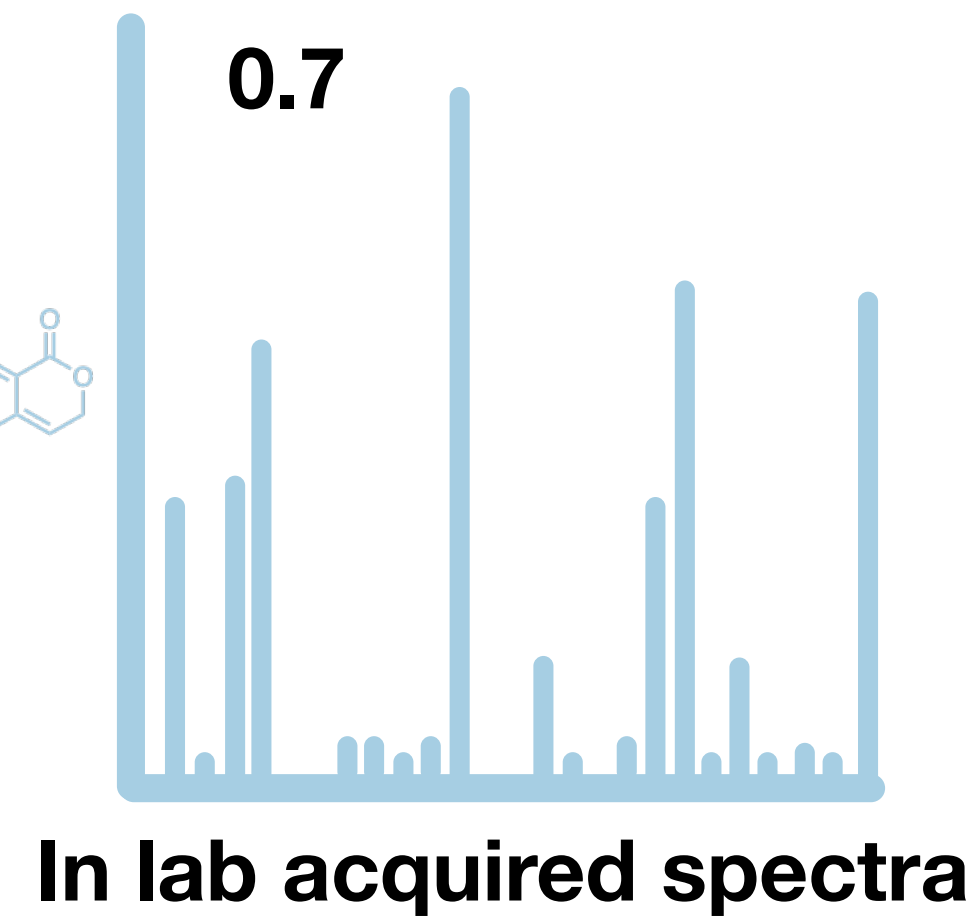
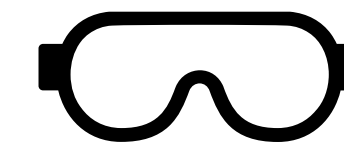
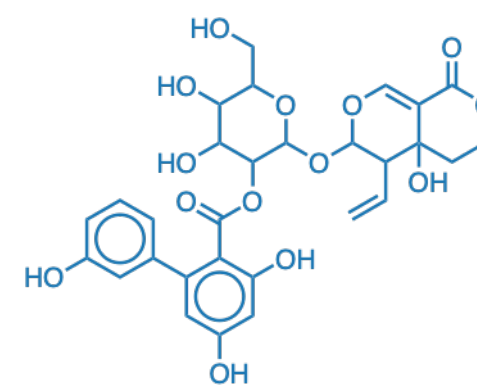
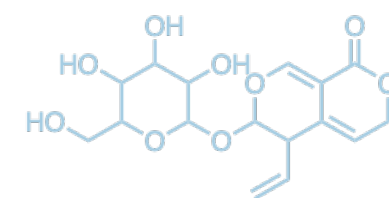
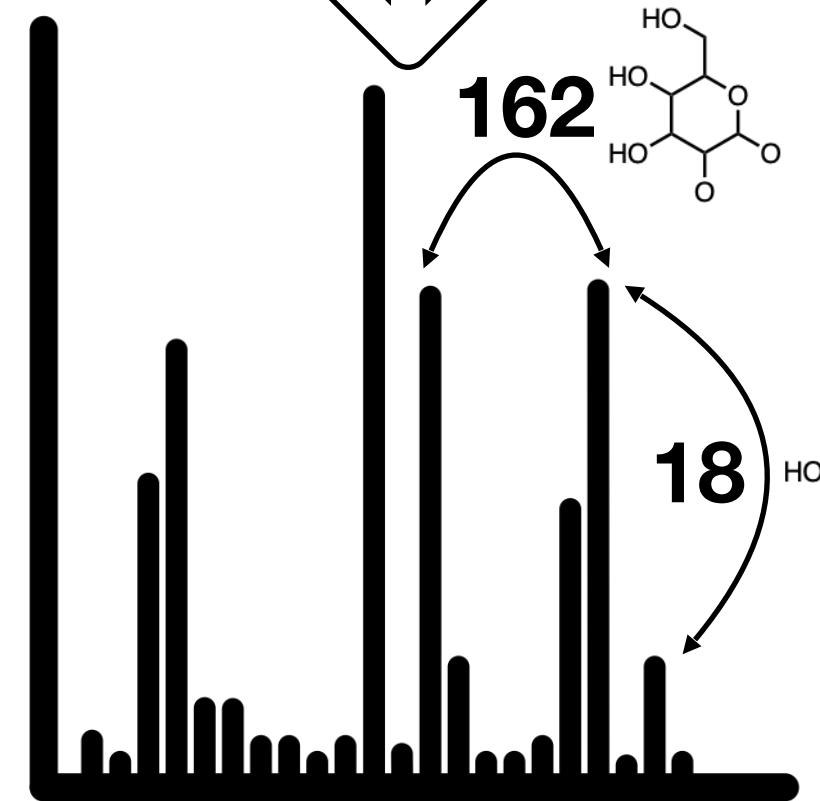
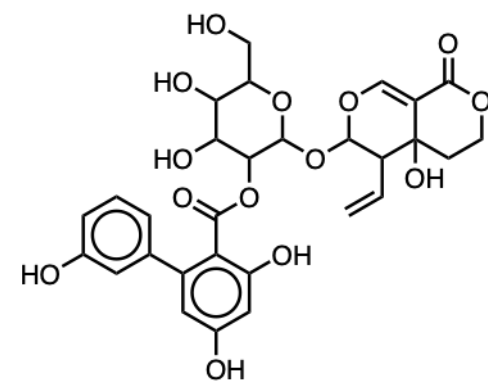
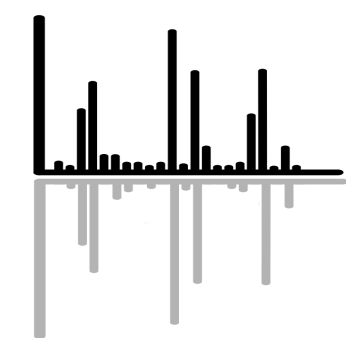




# Metabolite annotation

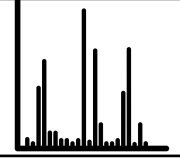
Feature list (or table):

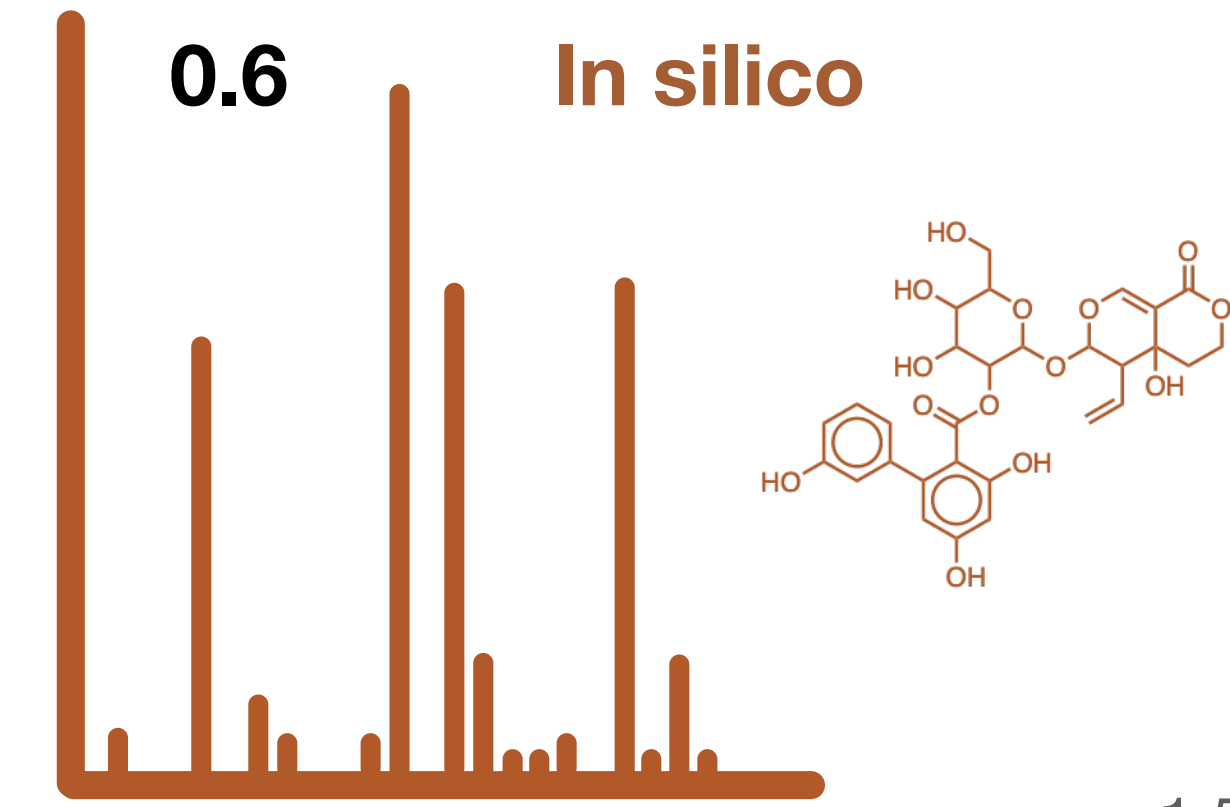
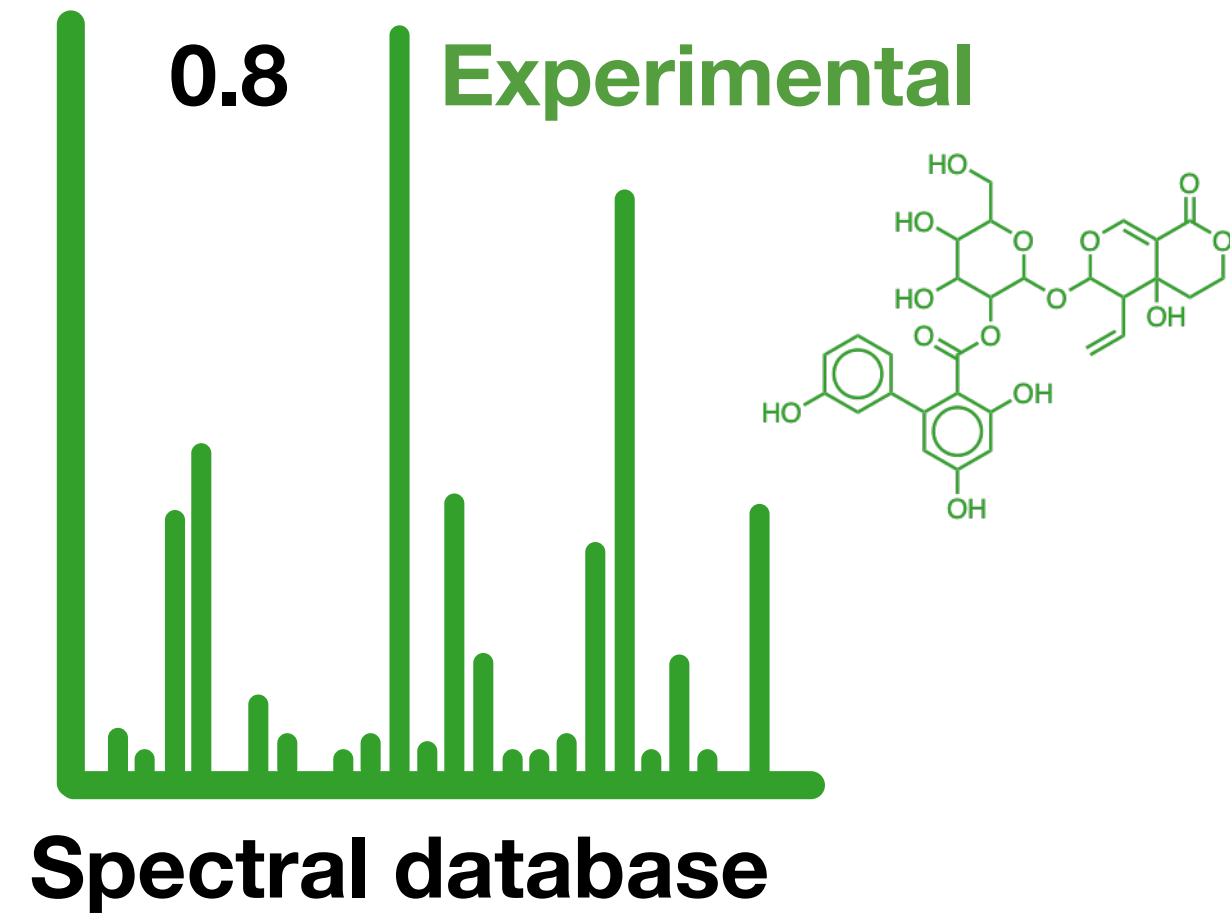
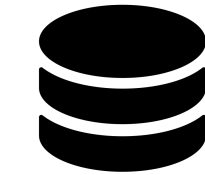
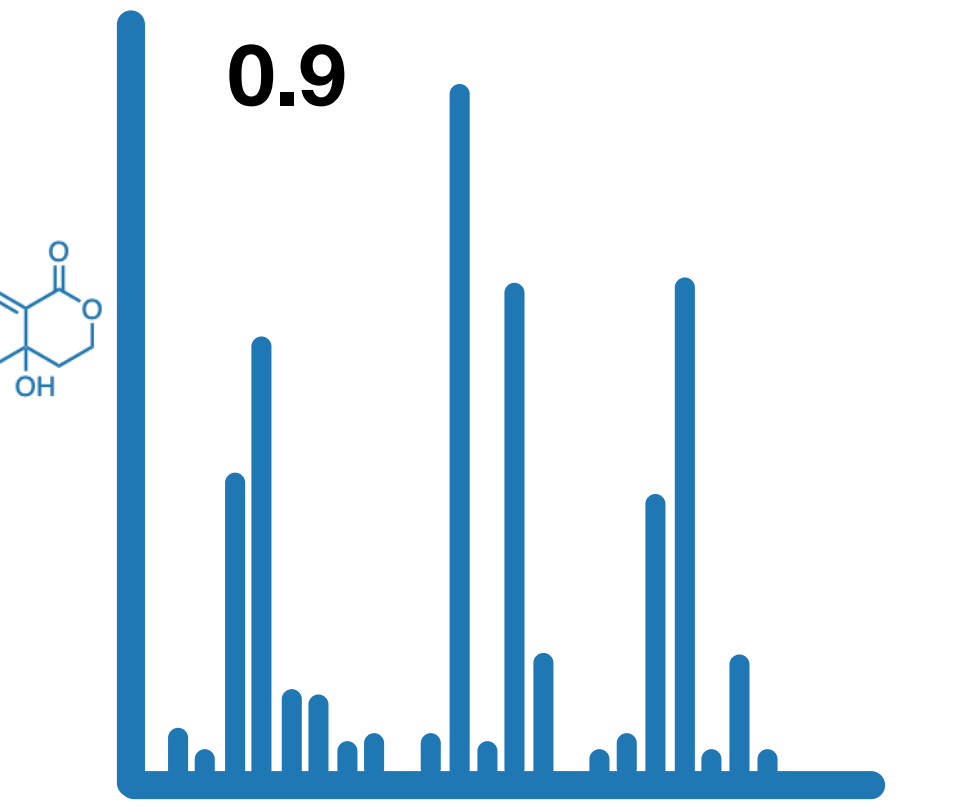
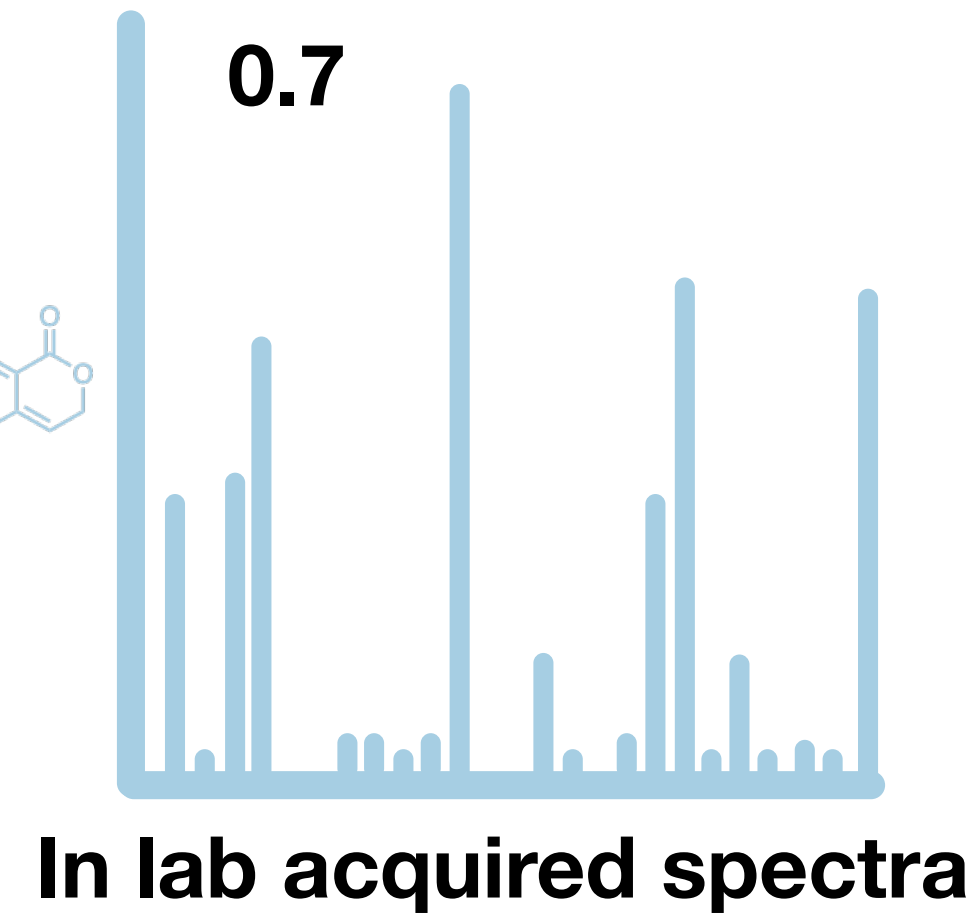
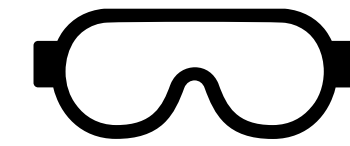
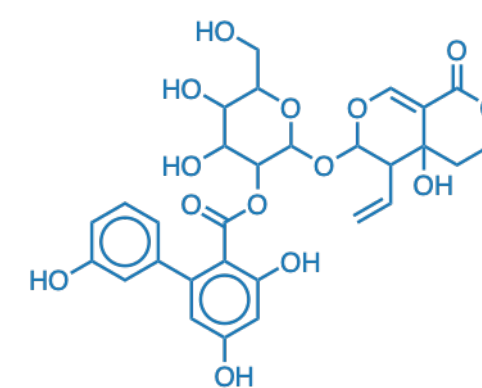
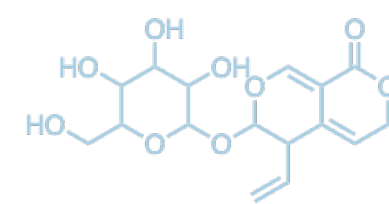
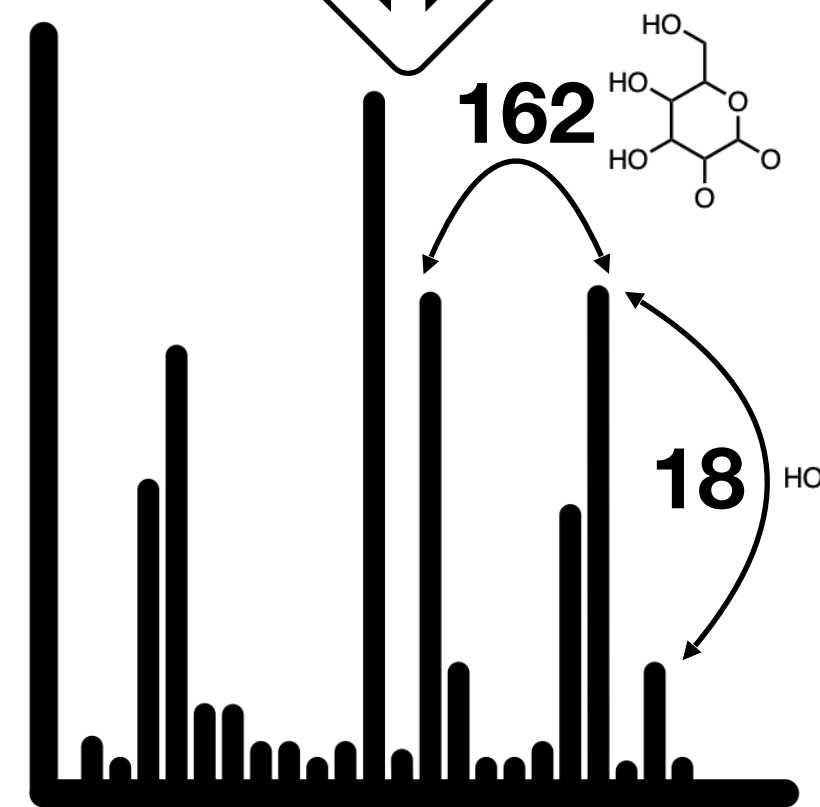
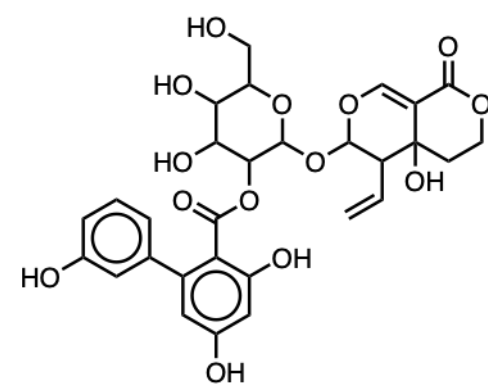
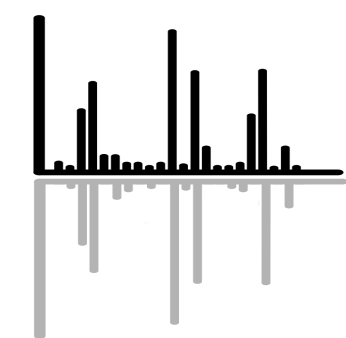
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



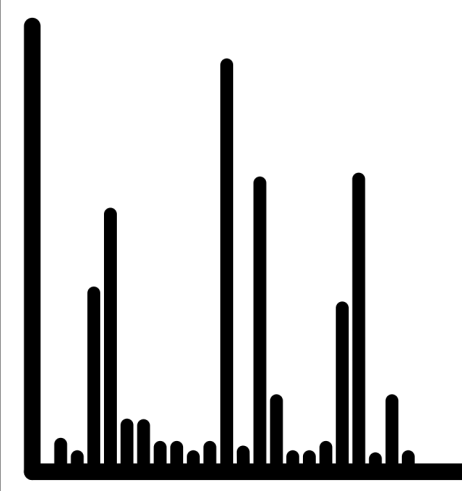
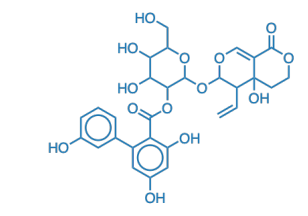
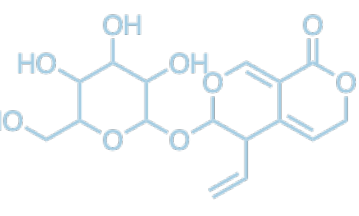
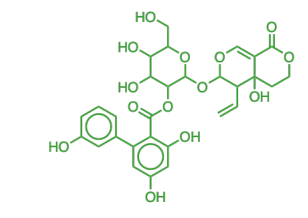
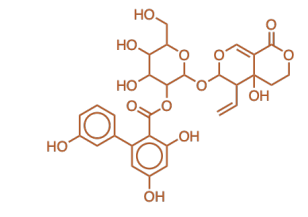
# Metabolite annotation

Feature list (or table):

ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

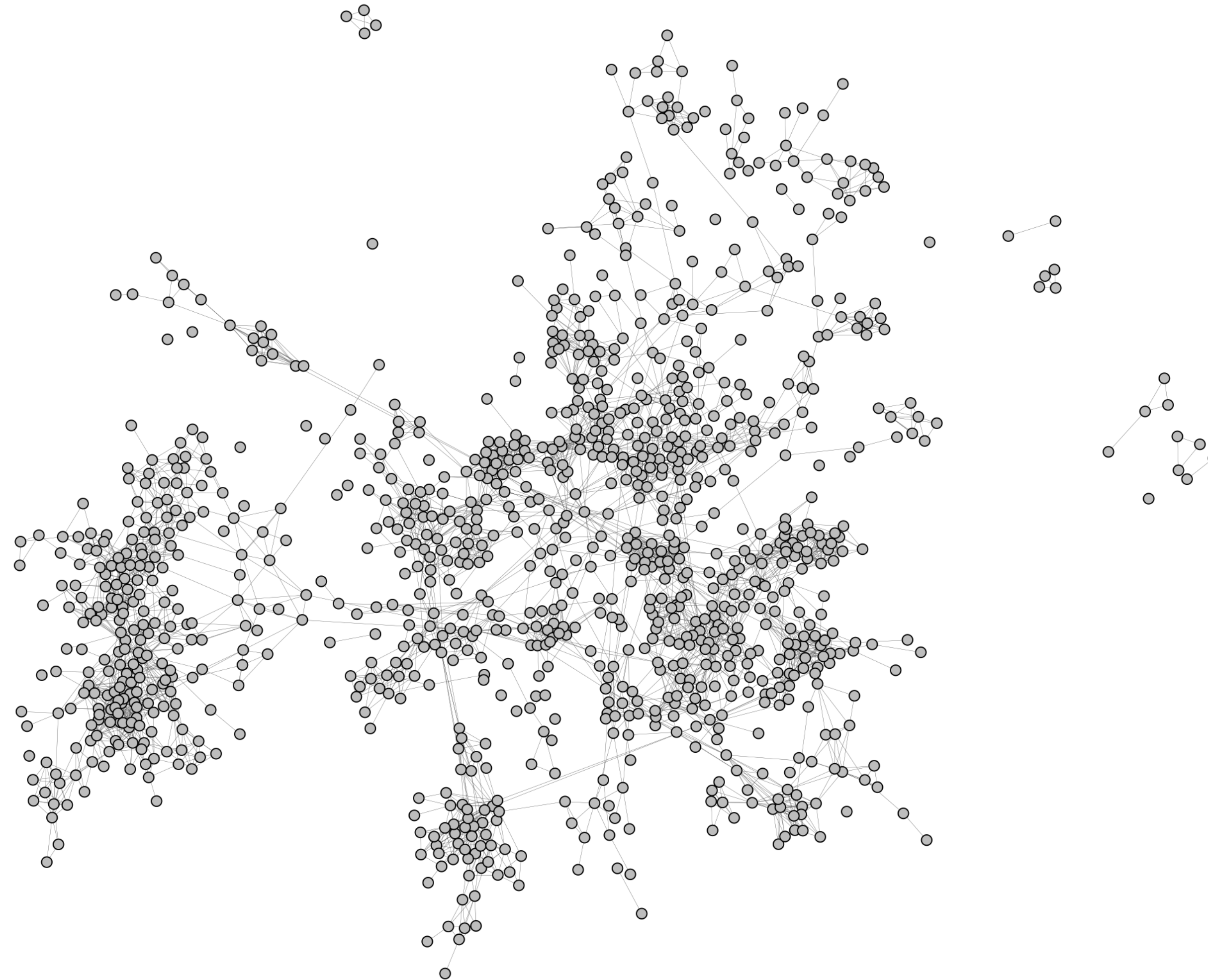


# Metabolite annotation

Feature ID	Spectrum	Candidate structure	Score $S_1$	Initial rank
1			0.9	1
			0.8	2
			0.7	3
			0.6	4
...	...	...	...	...

# Metabolite annotation

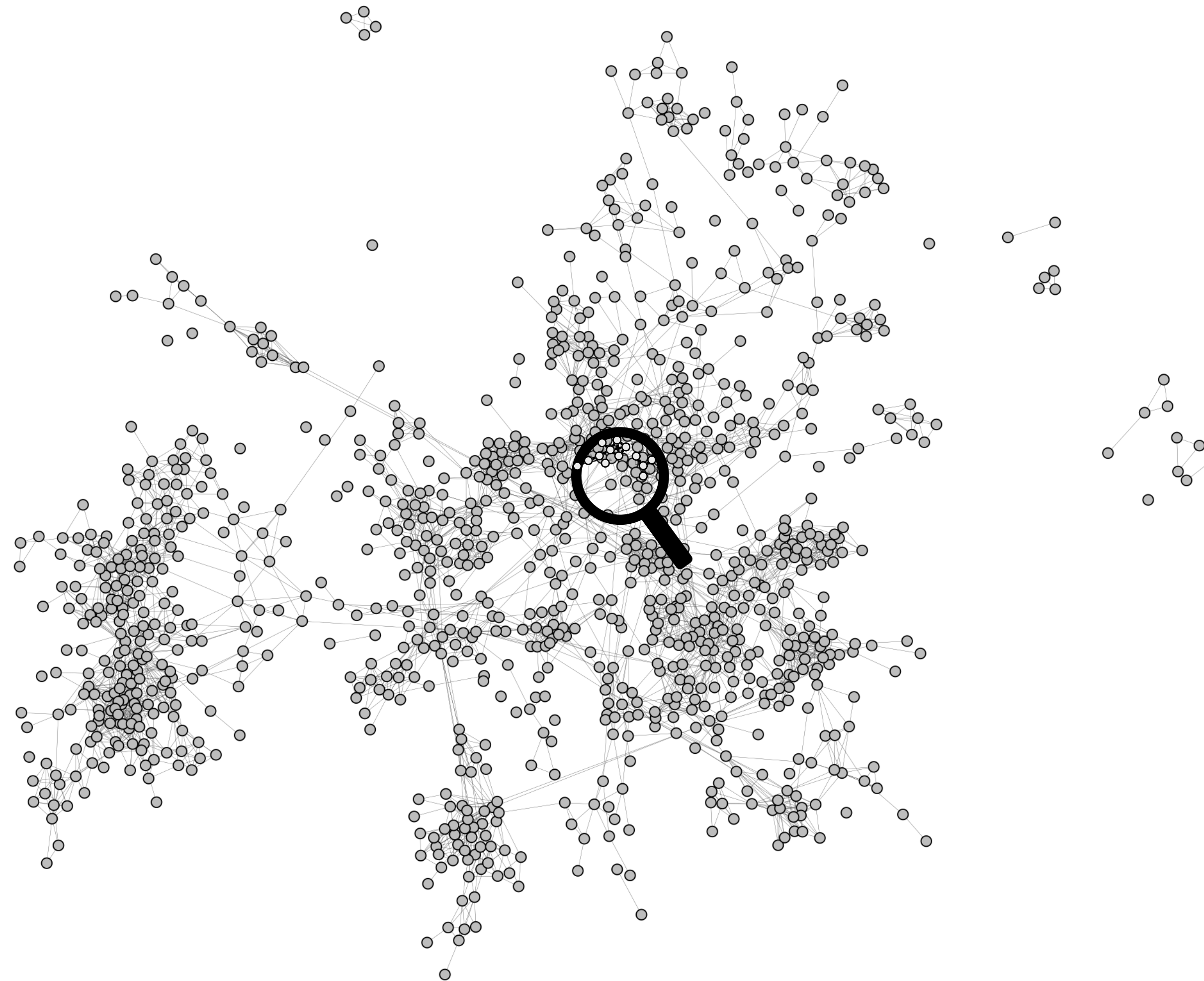
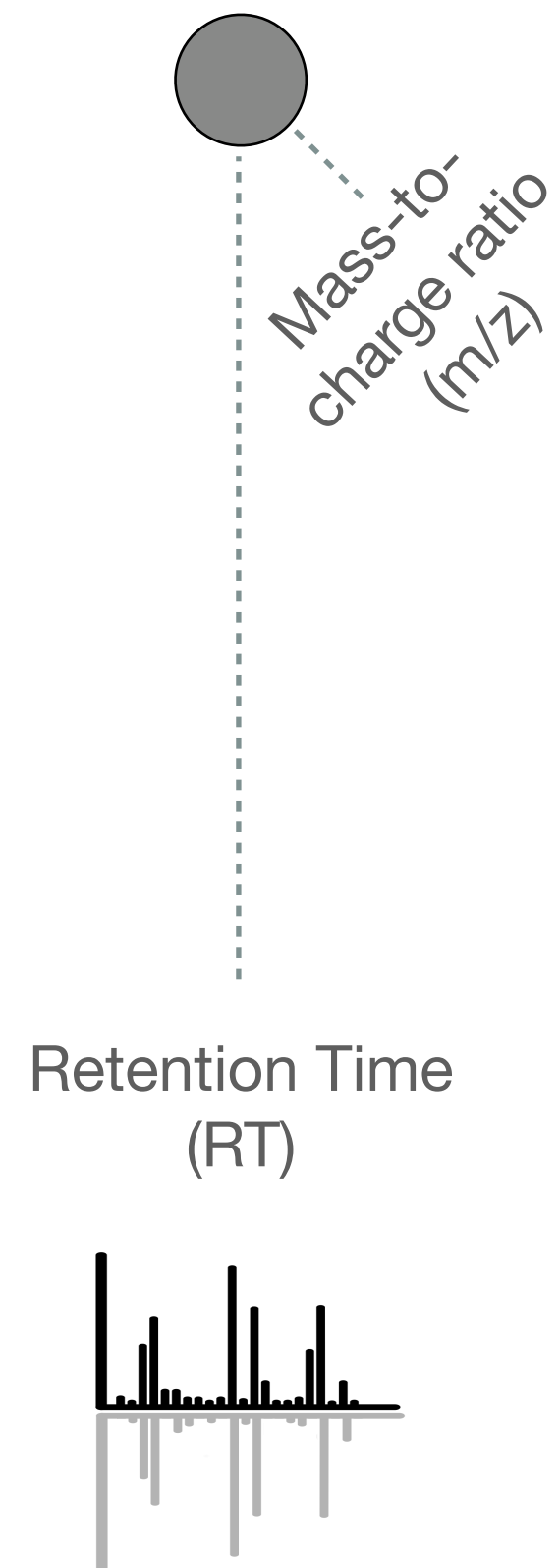
Feature:  
RT @  $m/z$  pointer



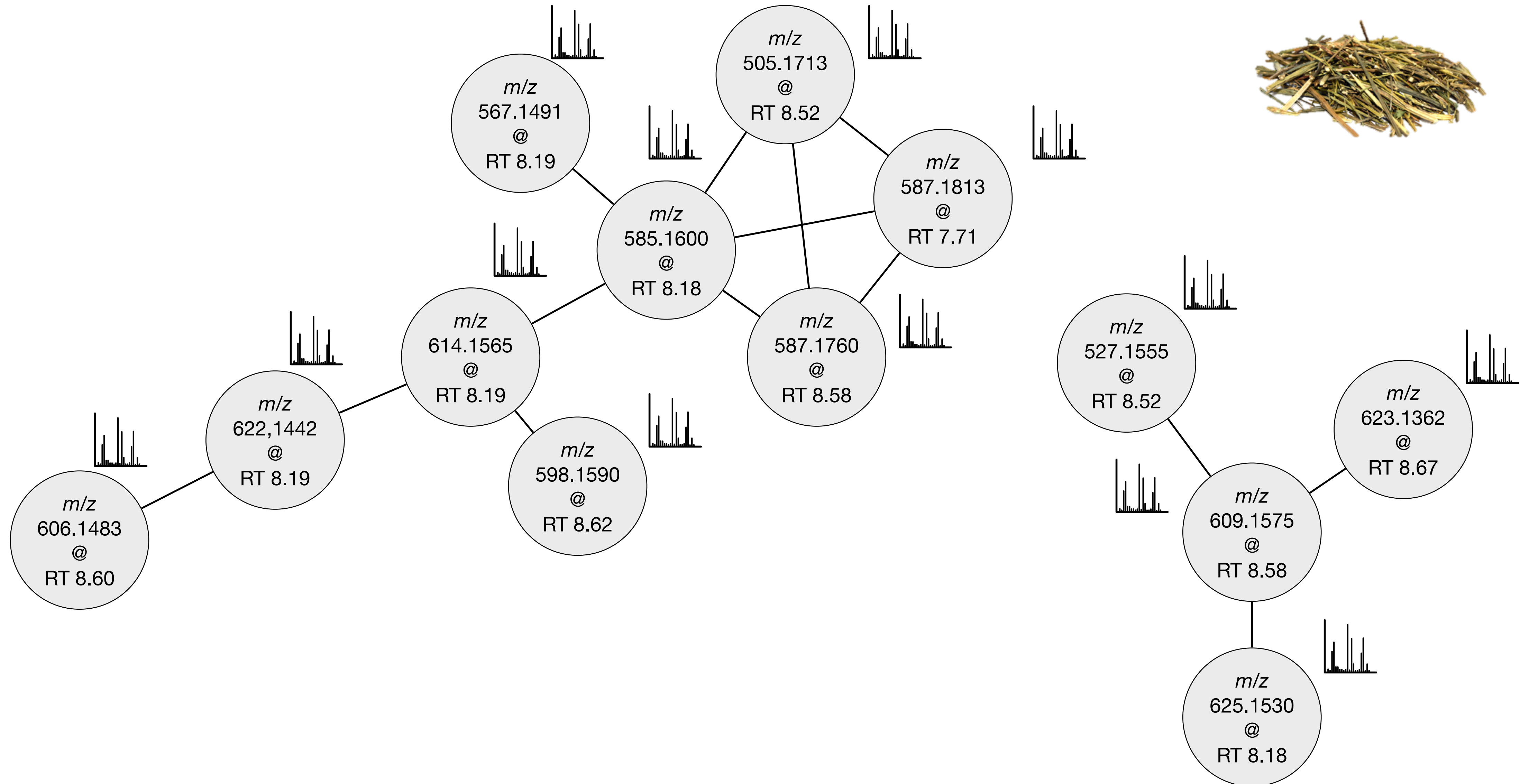


# Metabolite annotation

Feature:  
RT @  $m/z$  pointer

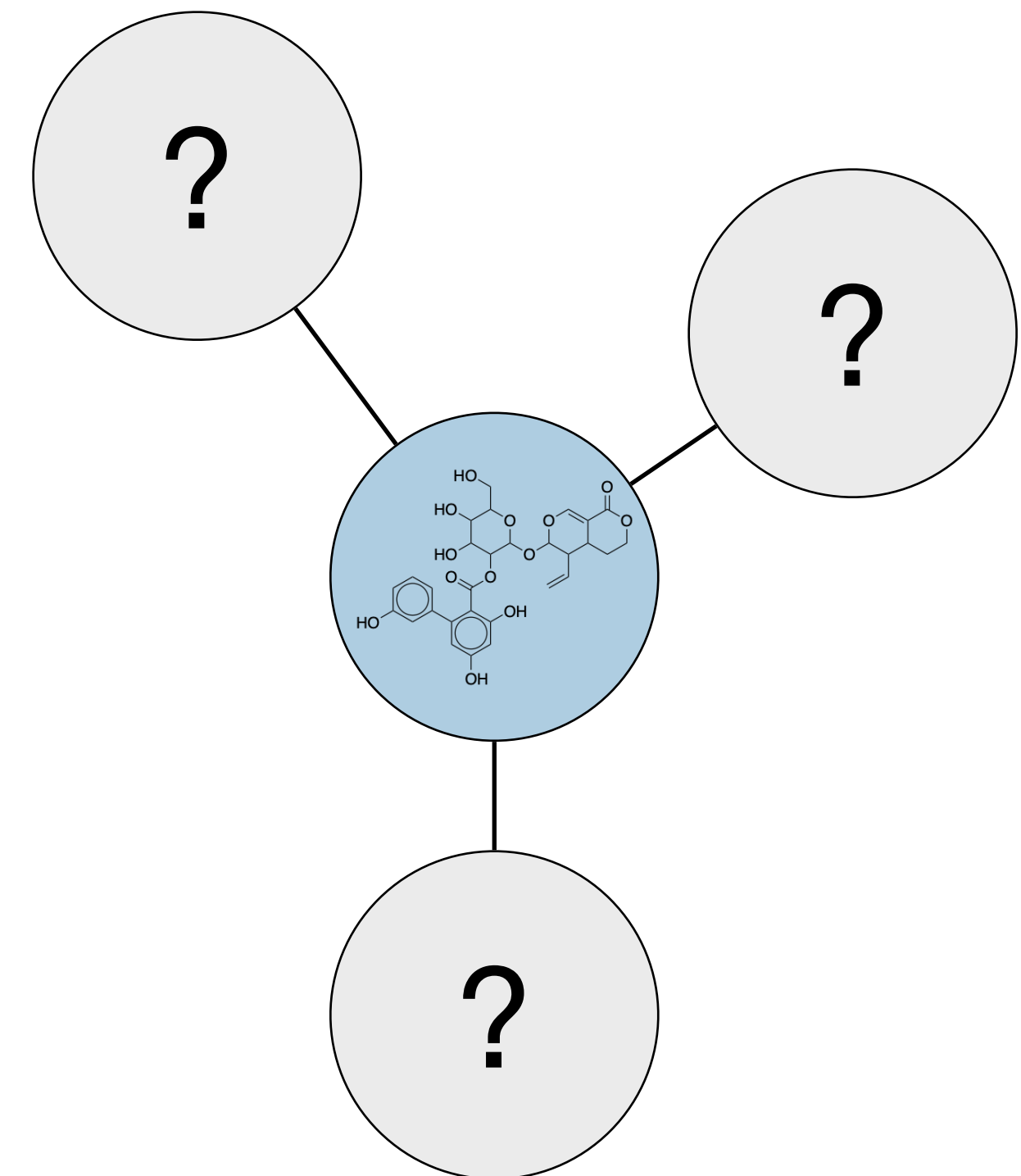
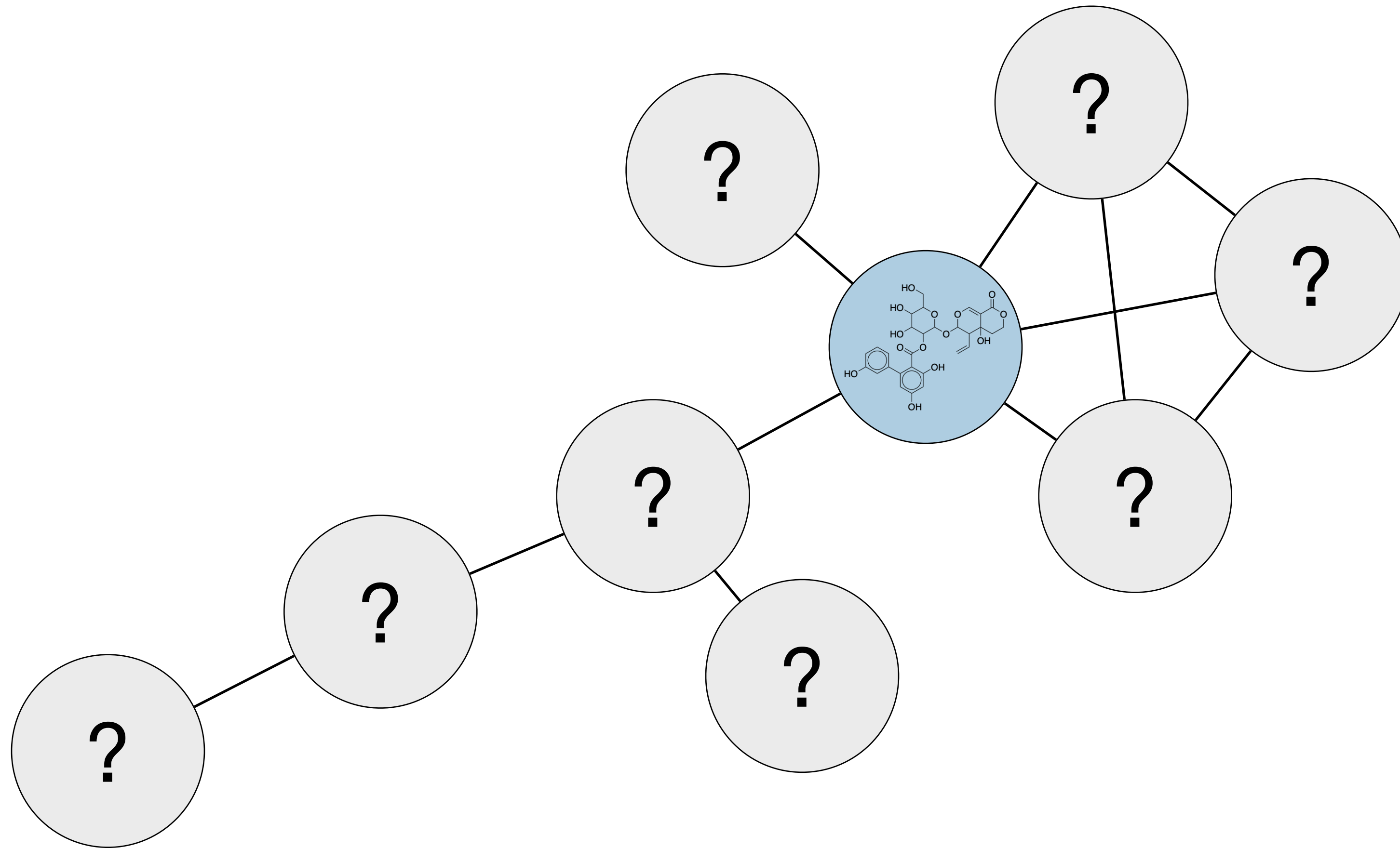


# Metabolite annotation



# Metabolite annotation

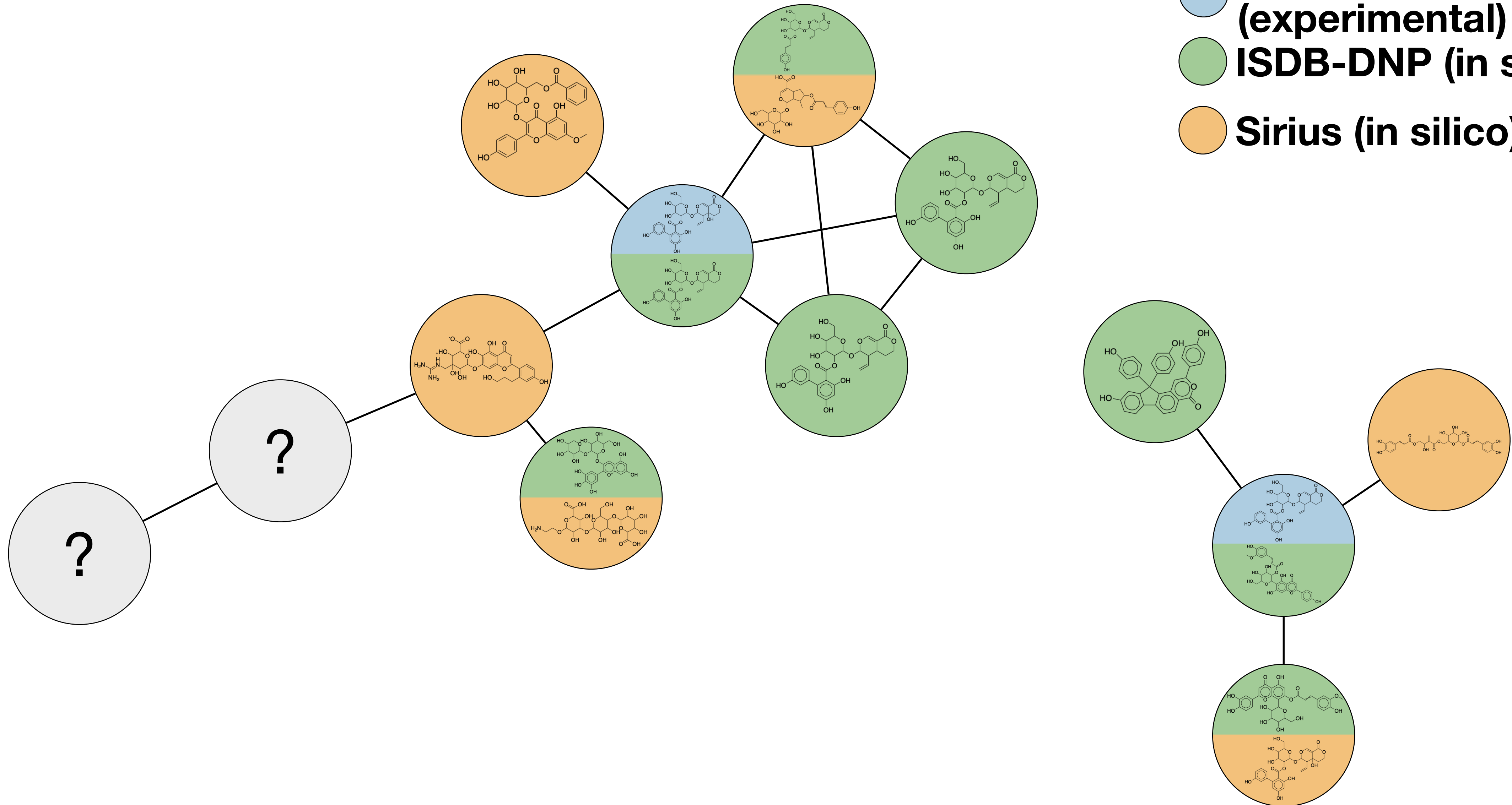
● GNPS  
(experimental)



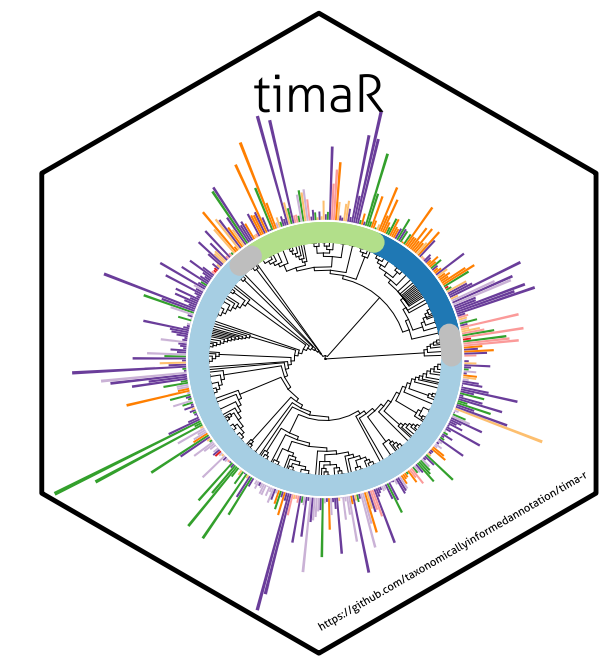


# Metabolite annotation

- GNPS (experimental)
- ISDB-DNP (in silico)
- Sirius (in silico)







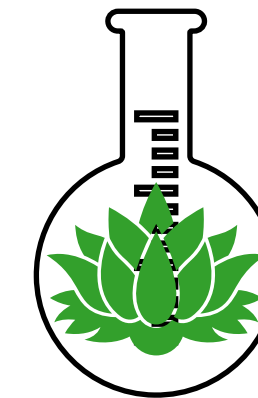
# Taxonomy as prior

- Strongly informative
- Orthogonal
- Highly interoperable
- Long history, well defined, used to handle both identifiers and ambiguity

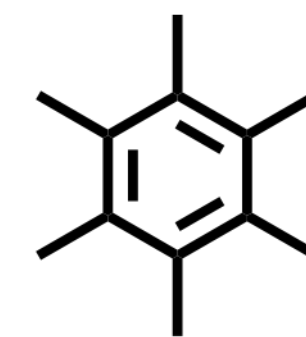
# Taxonomically Informed Metabolite Annotation

---

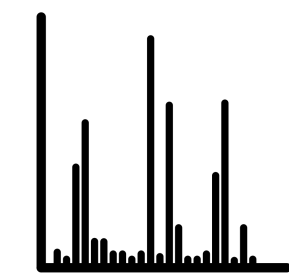
## *Metabolomics*



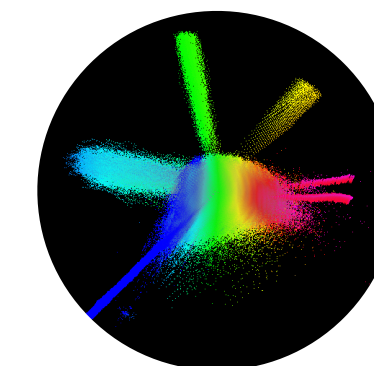
Natural Extract



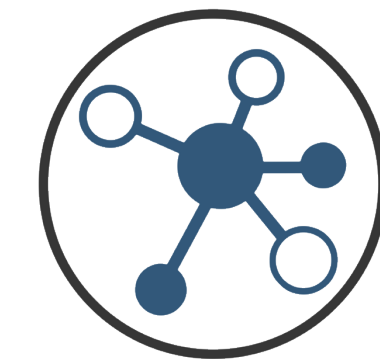
Structure



MS<sup>2</sup>  
spectrum



Chemical  
Space



Molecular  
Network

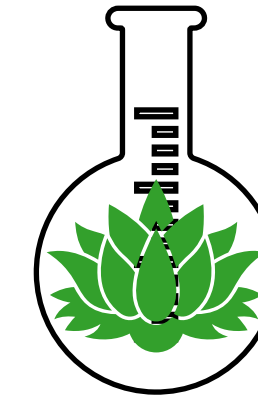
# Taxonomically Informed Metabolite Annotation

*Taxonomy*

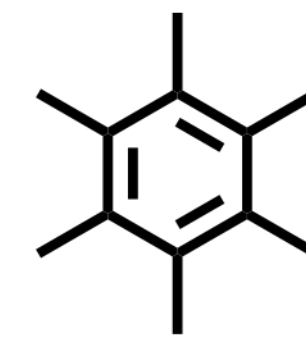


Biological organism

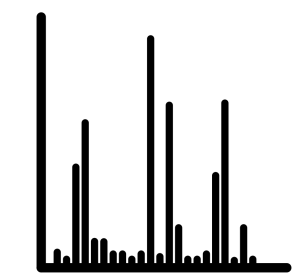
*Metabolomics*



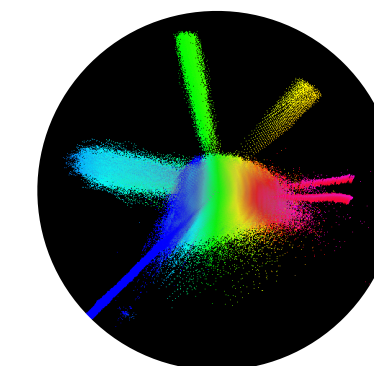
Natural Extract



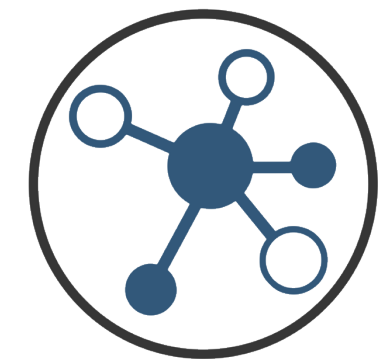
Structure



MS<sup>2</sup>  
spectrum



Chemical  
Space

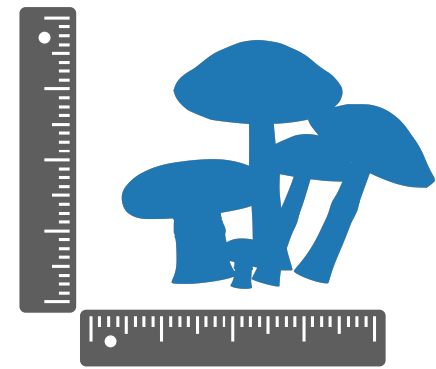


Molecular  
Network

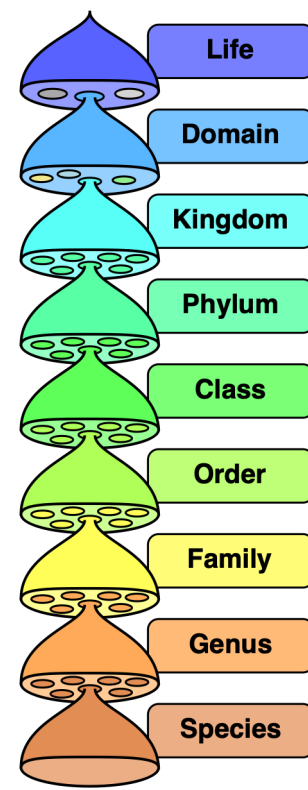
## Taxonomy



Biological organism

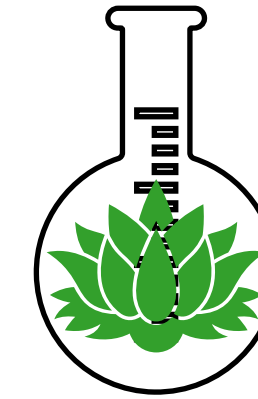


Morphological traits

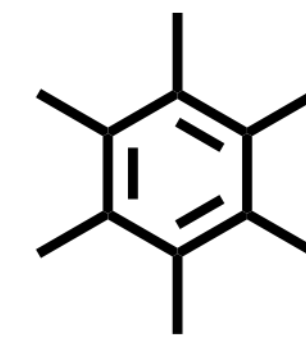


Taxonomy

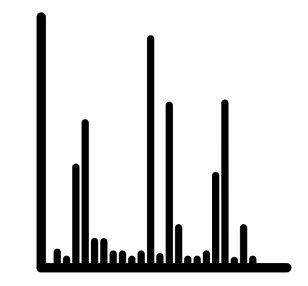
## Metabolomics



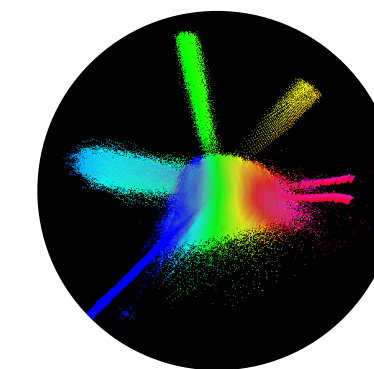
Natural Extract



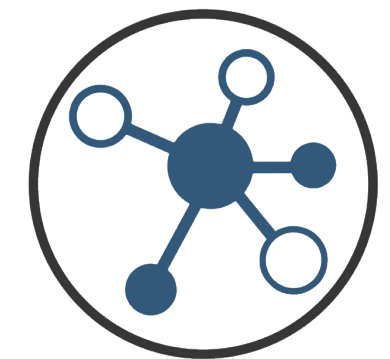
Structure



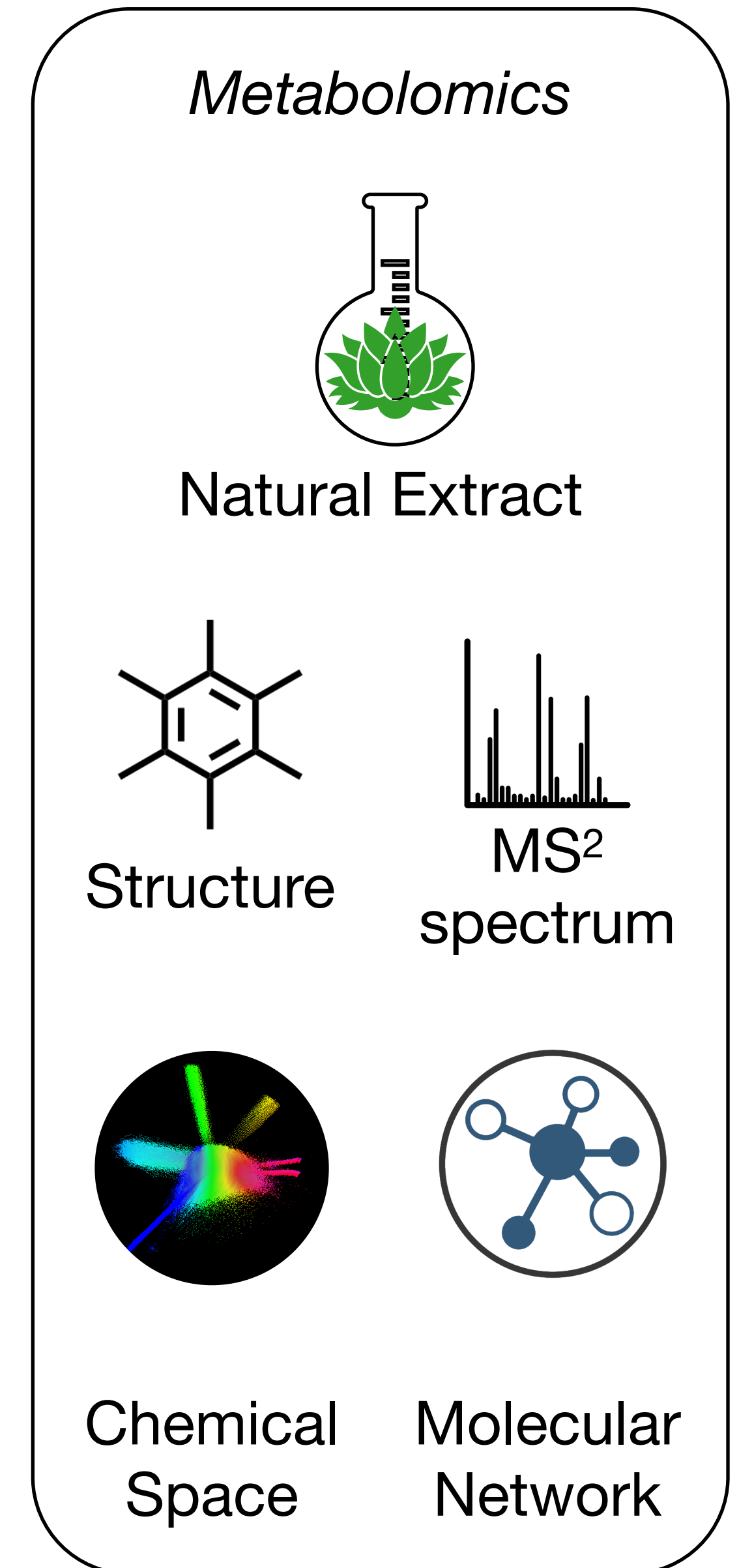
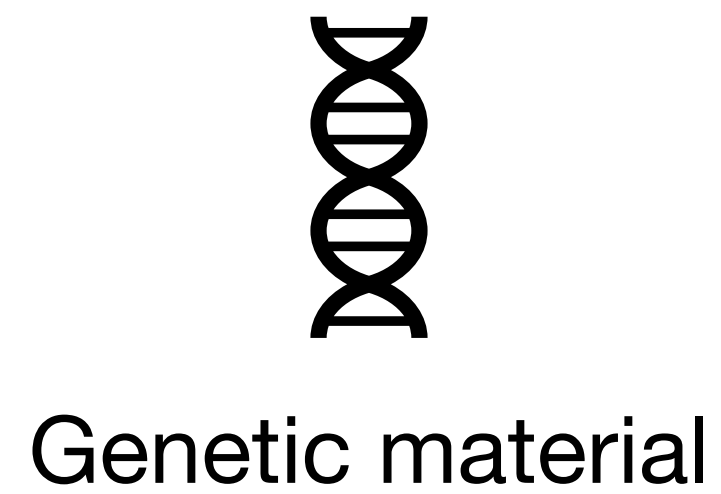
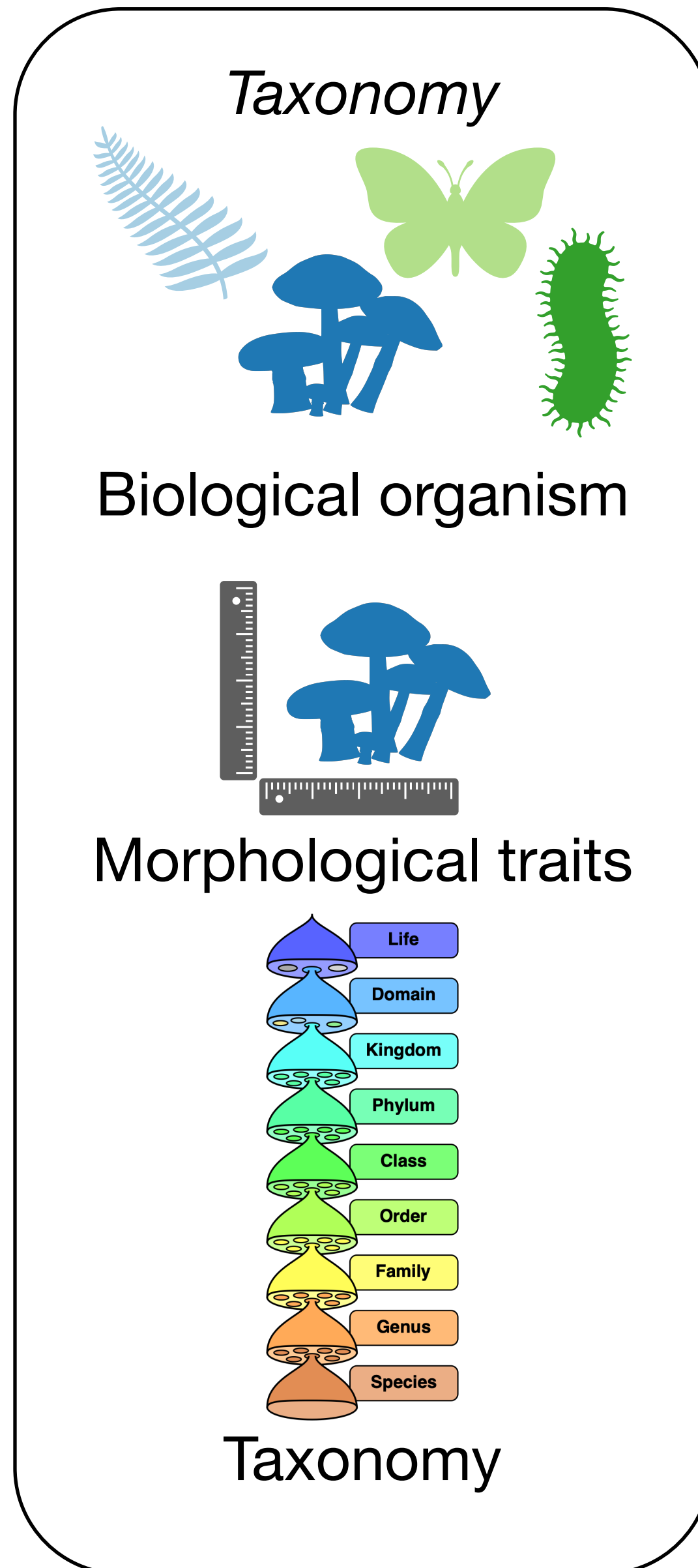
MS<sup>2</sup>  
spectrum



Chemical  
Space



Molecular  
Network



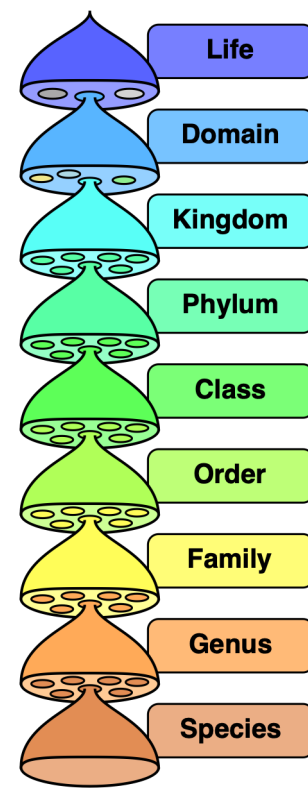
## Taxonomy



Biological organism

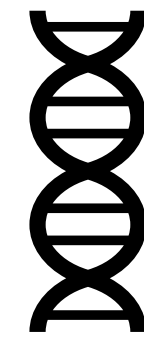


Morphological traits



Taxonomy

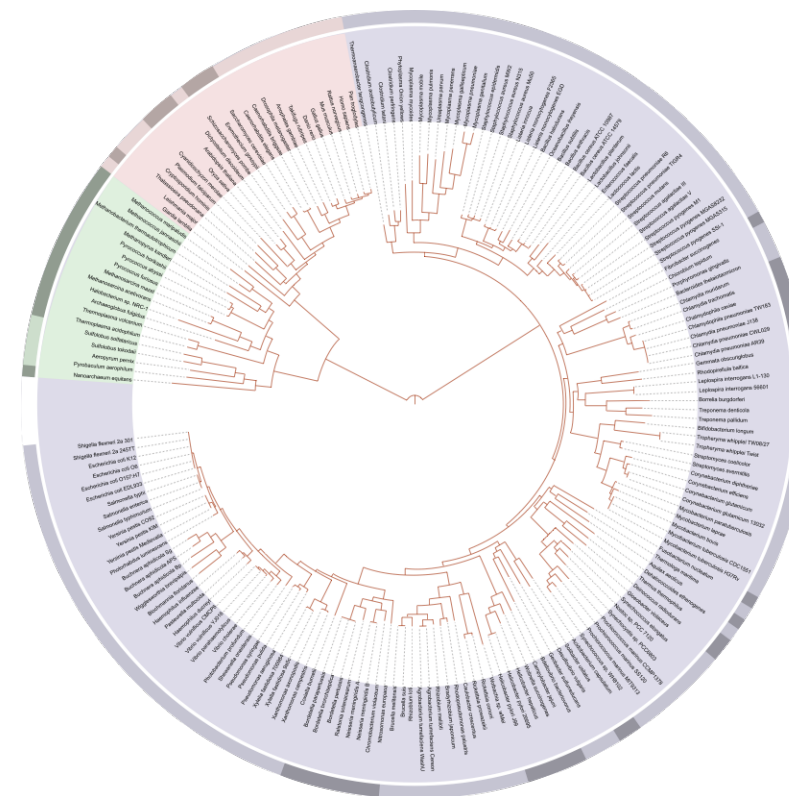
## Genomics



Genetic material

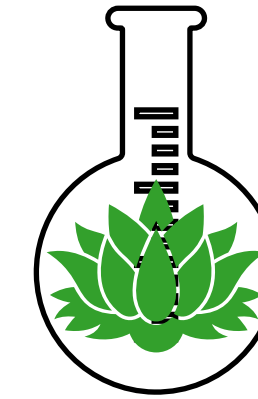
ATGCAAGCTTA  
CGTACCTAGGC

DNA sequence

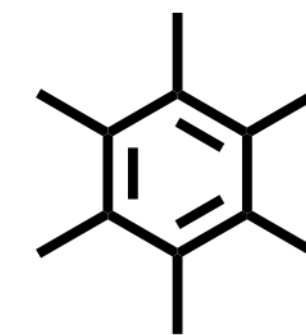


Phylogeny

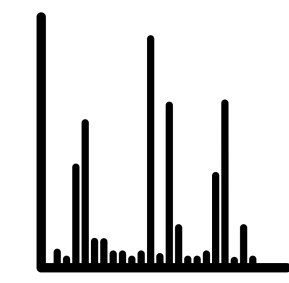
## Metabolomics



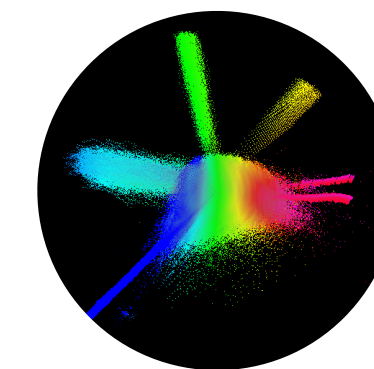
Natural Extract



Structure



MS<sup>2</sup>  
spectrum



Chemical  
Space




Molecular  
Network




# Taxonomically Informed Metabolite Annotation

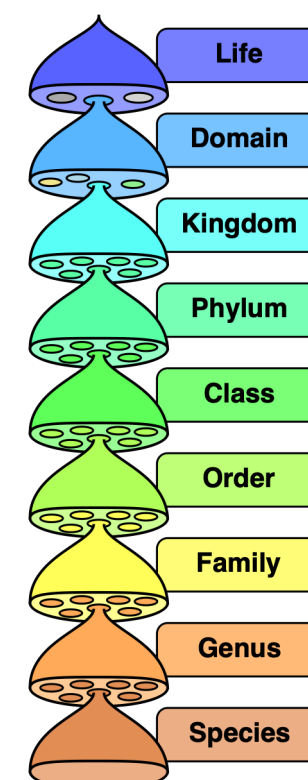
*Taxonomy*



Biological organism




Morphological traits



Taxonomy

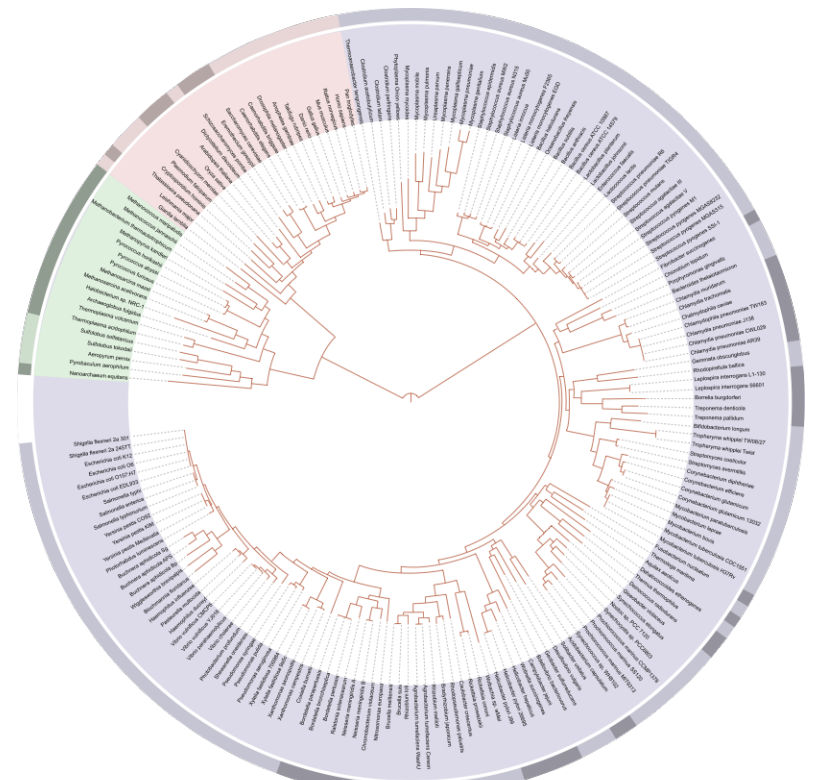
*Genomics*



Genetic material

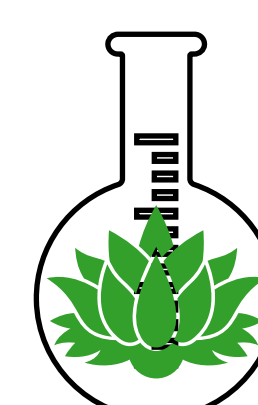
ATGCAAGCTTA  
CGTACCTAGGC

DNA sequence

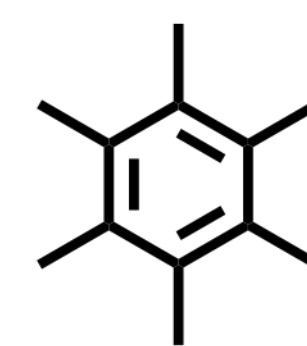


Phylogeny

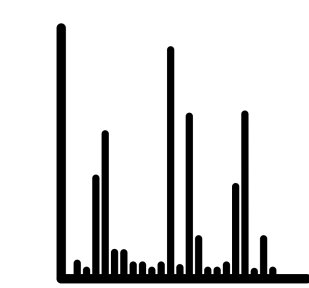
*Metabolomics*



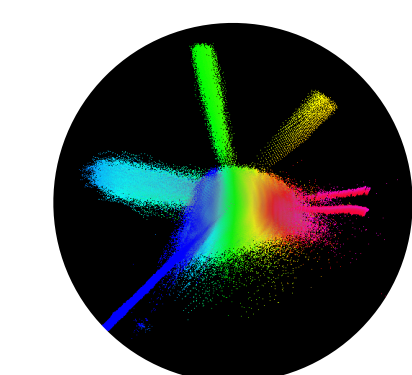
Natural Extract



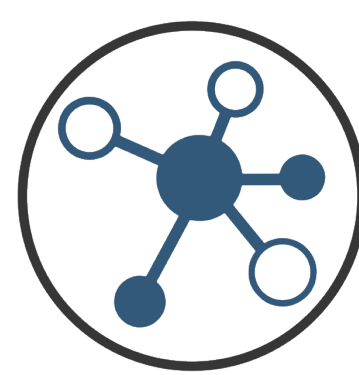
Structure



MS<sup>2</sup> spectrum

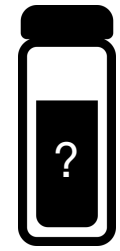


Chemical Space



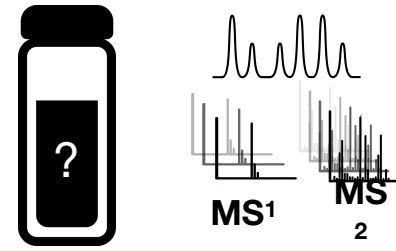
Molecular Network

# Candidates re-ranking

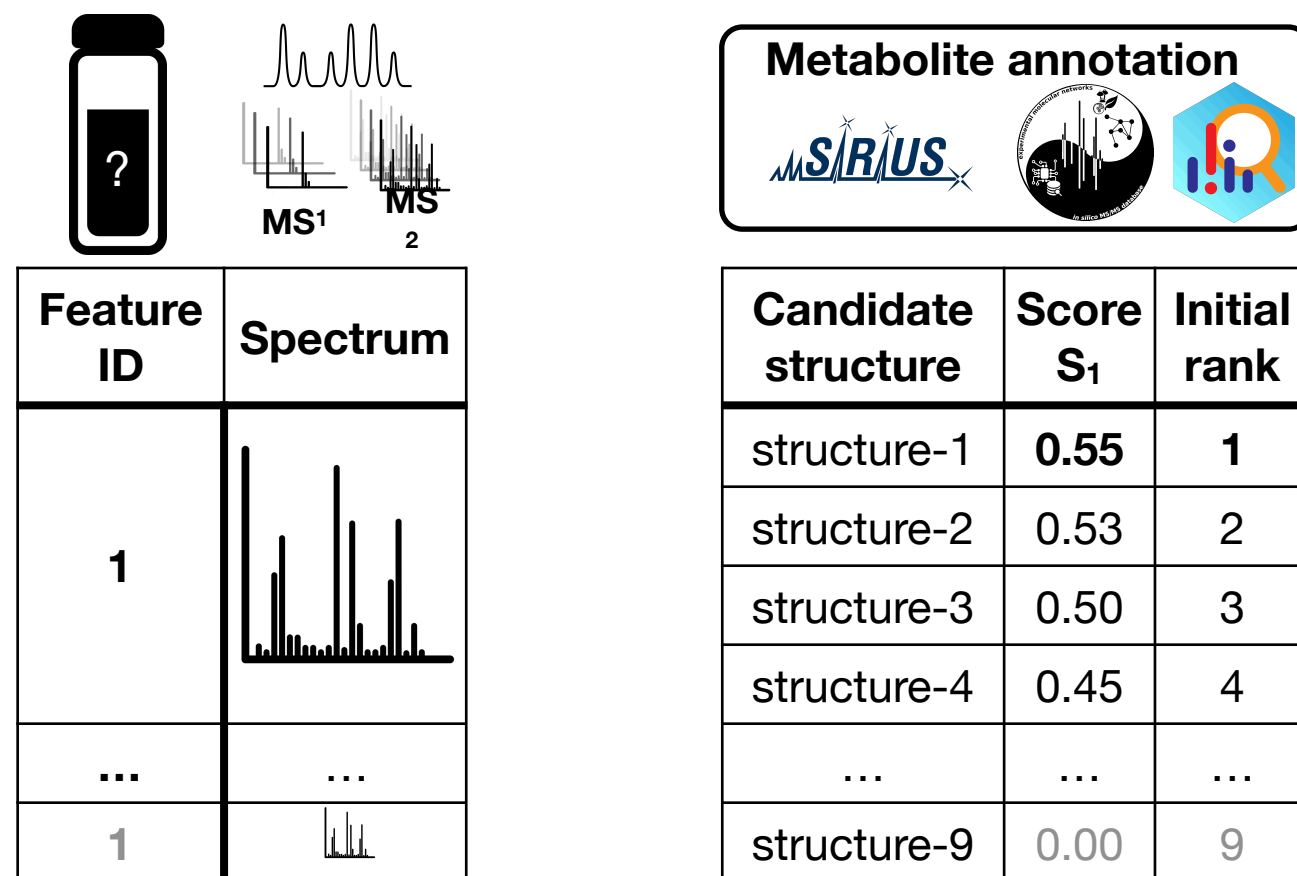




# Candidates re-ranking



# Candidates re-ranking



# Candidates re-ranking



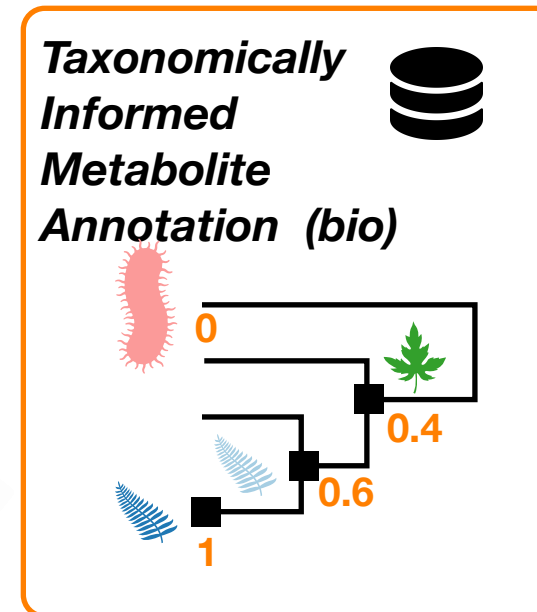
Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9

# Candidates re-ranking



Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9

re-ranking



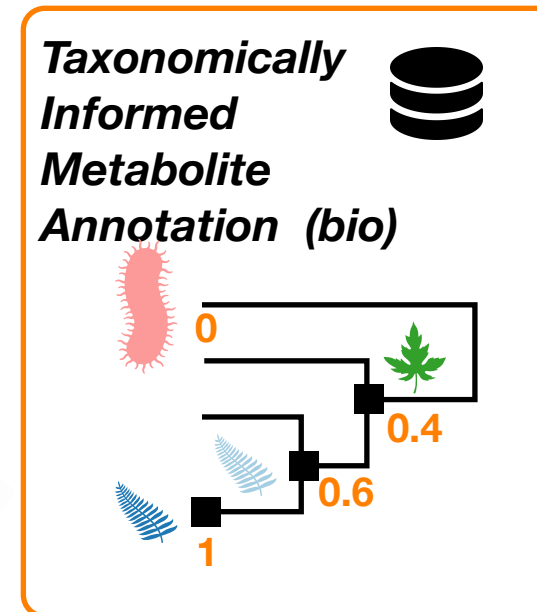
Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank	Candidate biological source	Score $S_2$	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1
			structure-1	0.55	1		0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3
			structure-9	0.00	9		0.60	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5

# Candidates re-ranking



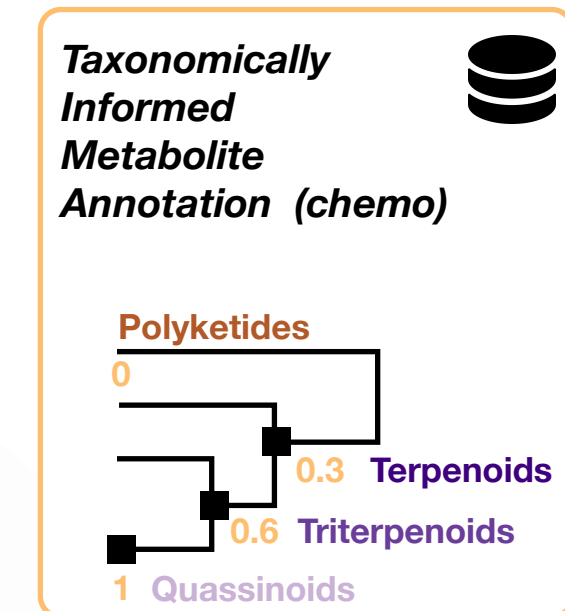
Feature ID	Spectrum	Biological source	Candidate structure	Score S <sub>1</sub>	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9

re-ranking



Feature ID	Spectrum	Biological source	Candidate structure	Score S <sub>1</sub>	Initial rank	Candidate biological source	Score S <sub>2</sub>	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1
			structure-1	0.55	1		0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3
			structure-9	0.00	9		0.60	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5

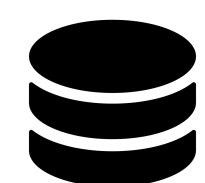
re-ranking



Feature ID	Spectrum	Biological source	Candidate structure	Score S <sub>1</sub>	Initial rank	Candidate biological source	Score S <sub>2</sub>	Combined score	2nd rank	Attributed chemical class	Candidate chemical class	Score S <sub>3</sub>	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1	Quassinoid	Quassinoid	1.00	0.82	1
			structure-1	0.55	1		0.60	0.58	2		Triterpenoids	0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3		Terpenoids	0.30	0.41	3
			structure-9	0.00	9		0.60	0.30	4		Terpenoids	0.30	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5	...	Polyketides	0.00	0.17	5

# Benchmark

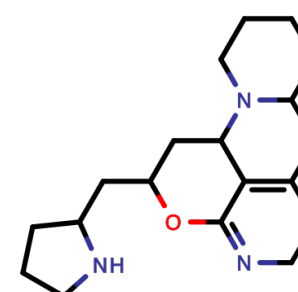
## EXPERIMENTAL SPECTRA



GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

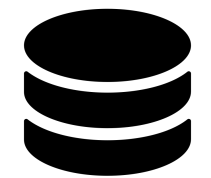
66,646 experimental spectra

## STRUCTURES





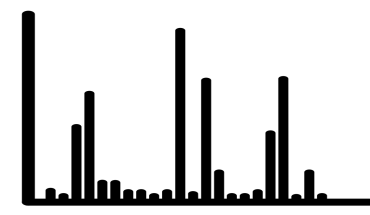
# Benchmark

## EXPERIMENTAL SPECTRA

 GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

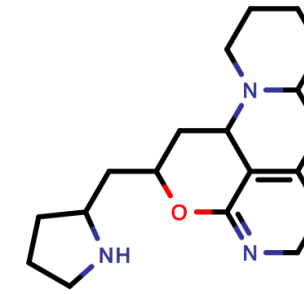
66,646 experimental spectra

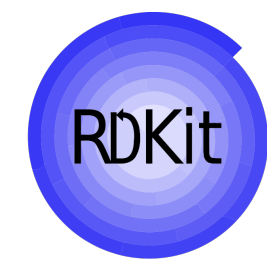
 python™  
 jupyter  
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

## STRUCTURES



 RDKit  
chemical translation  
and sanitization

Cleaned  
2D structures





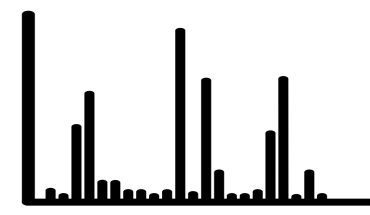
# Benchmark

## EXPERIMENTAL SPECTRA

 GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

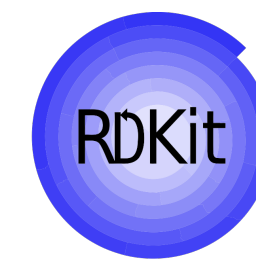
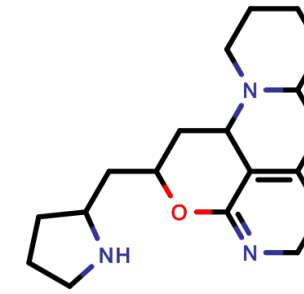
66,646 experimental spectra

 python™  
 jupyter  
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

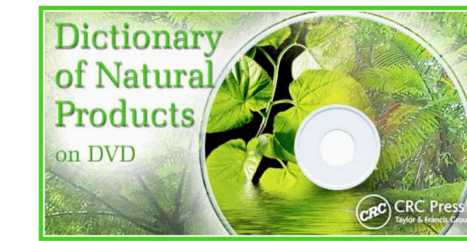
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



“Alkaloid from *Brunfelsia hopeana*”







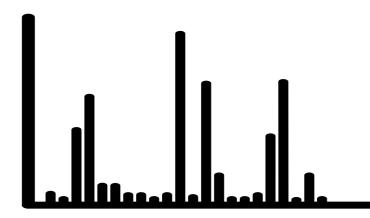
# Benchmark

## EXPERIMENTAL SPECTRA

 GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

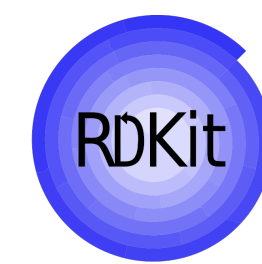
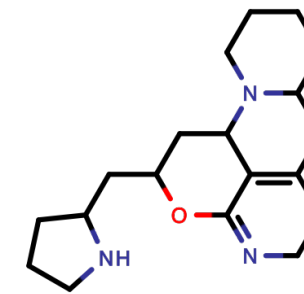
66,646 experimental spectra

 python™  
 jupyter  
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

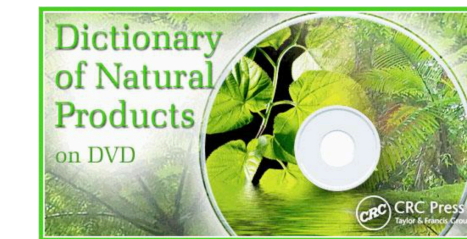
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



“Alkaloid from *Brunfelsia hopeana*”



text recognition  
matching and  
resolving against the  
Catalogue of Life



Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

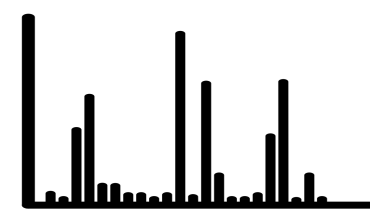
# Benchmark

## EXPERIMENTAL SPECTRA

 GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

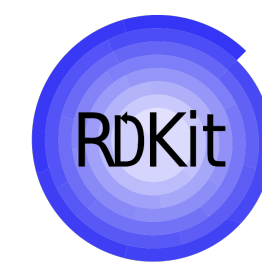
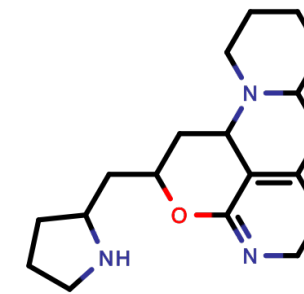
66,646 experimental spectra

 python™  
 jupyter  
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

## STRUCTURES

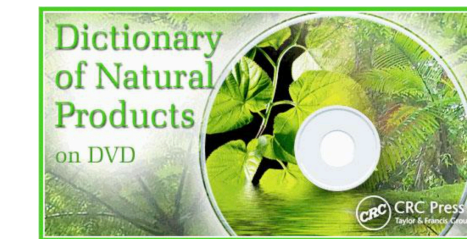


chemical translation  
and sanitization

Cleaned  
2D structures

2107 unique entries

## BIOLOGICAL SOURCES



“Alkaloid from *Brunfelsia hopeana*”



text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Evaluation

■ ISDB-DNP

■ Sirius

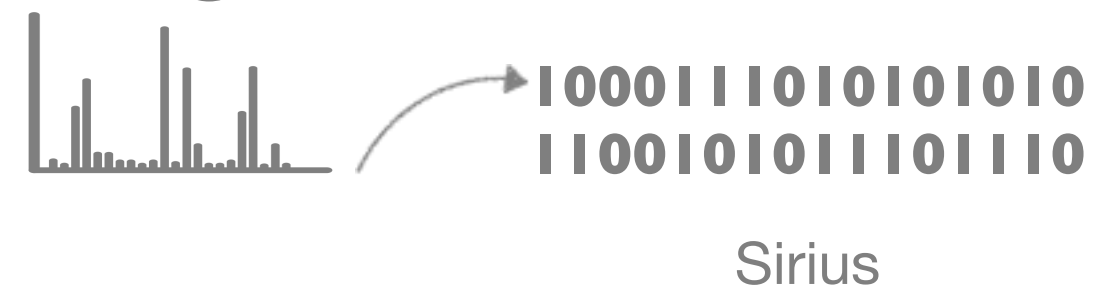
■ MS-Finder

# Evaluation

## Spectral similarity



## Fingerprint similarity



■ ISDB-DNP

■ Sirius

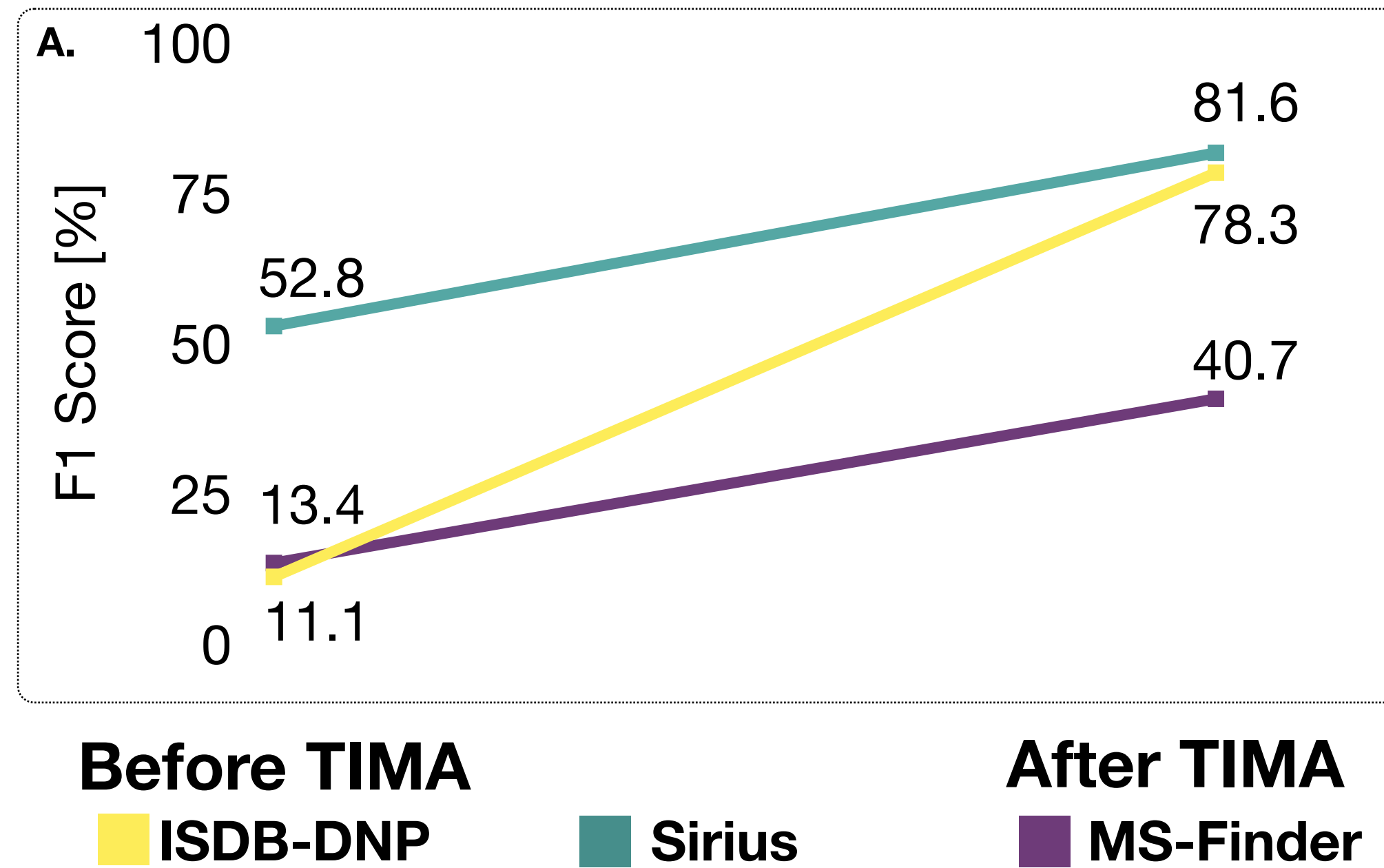
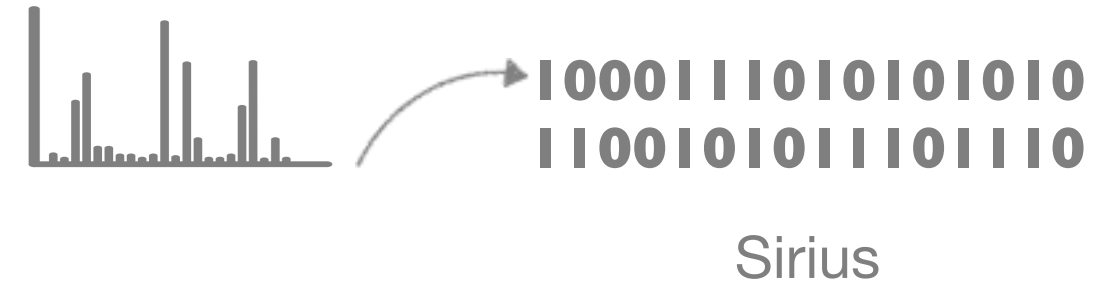
■ MS-Finder

# Evaluation

Spectral similarity



Fingerprint similarity

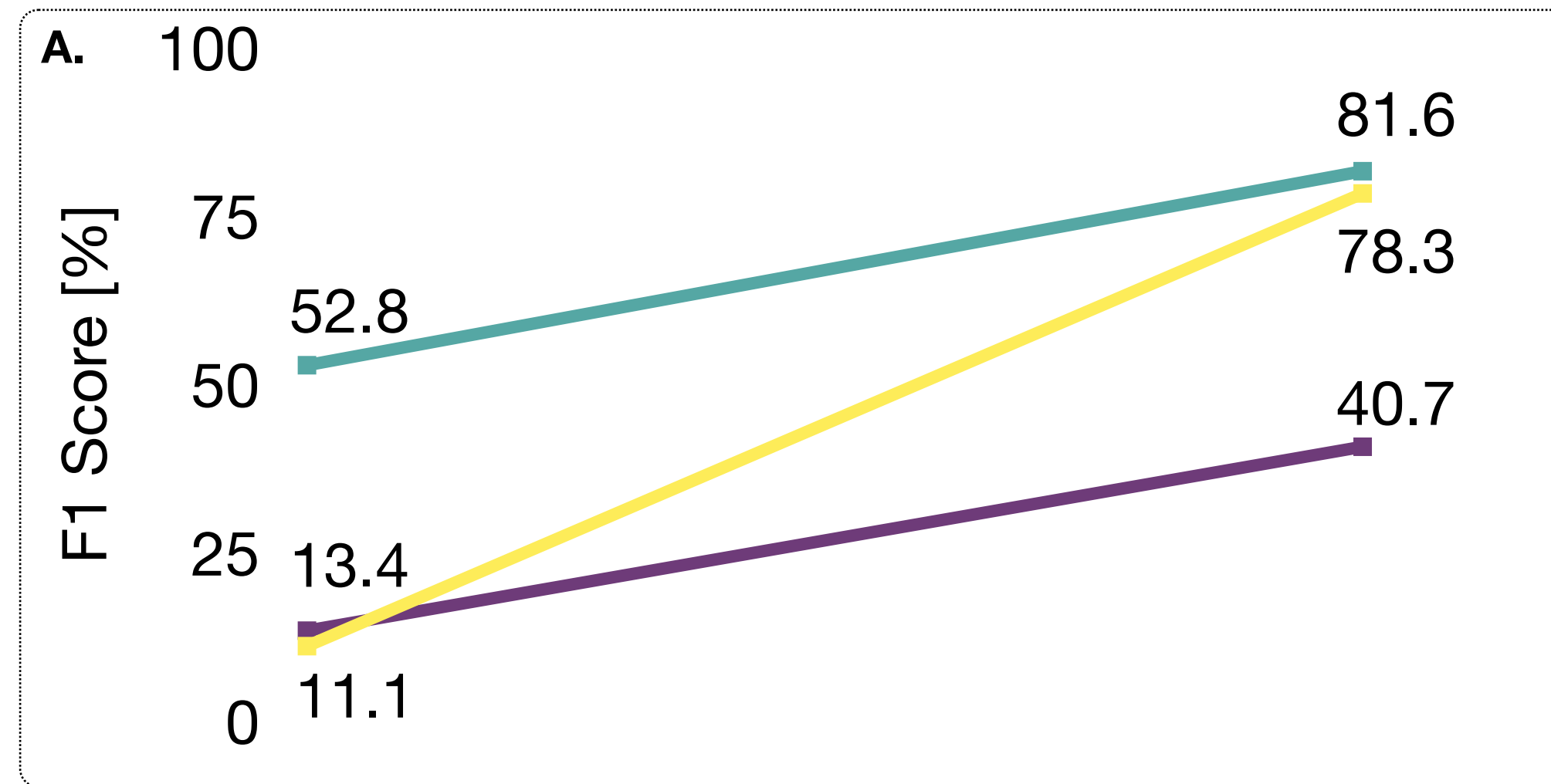
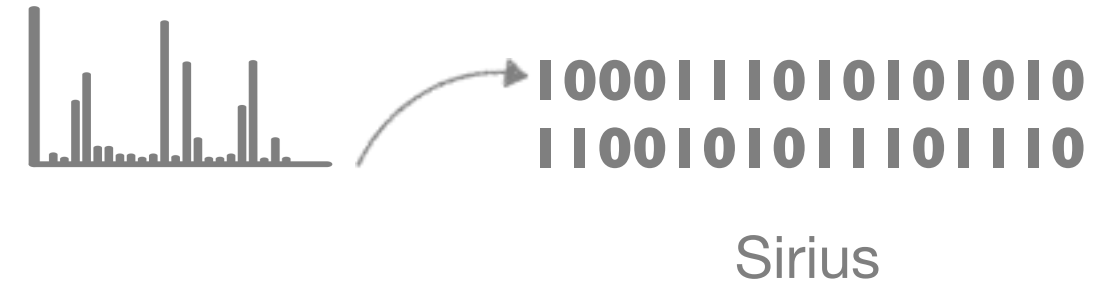


# Evaluation

Spectral similarity



Fingerprint similarity



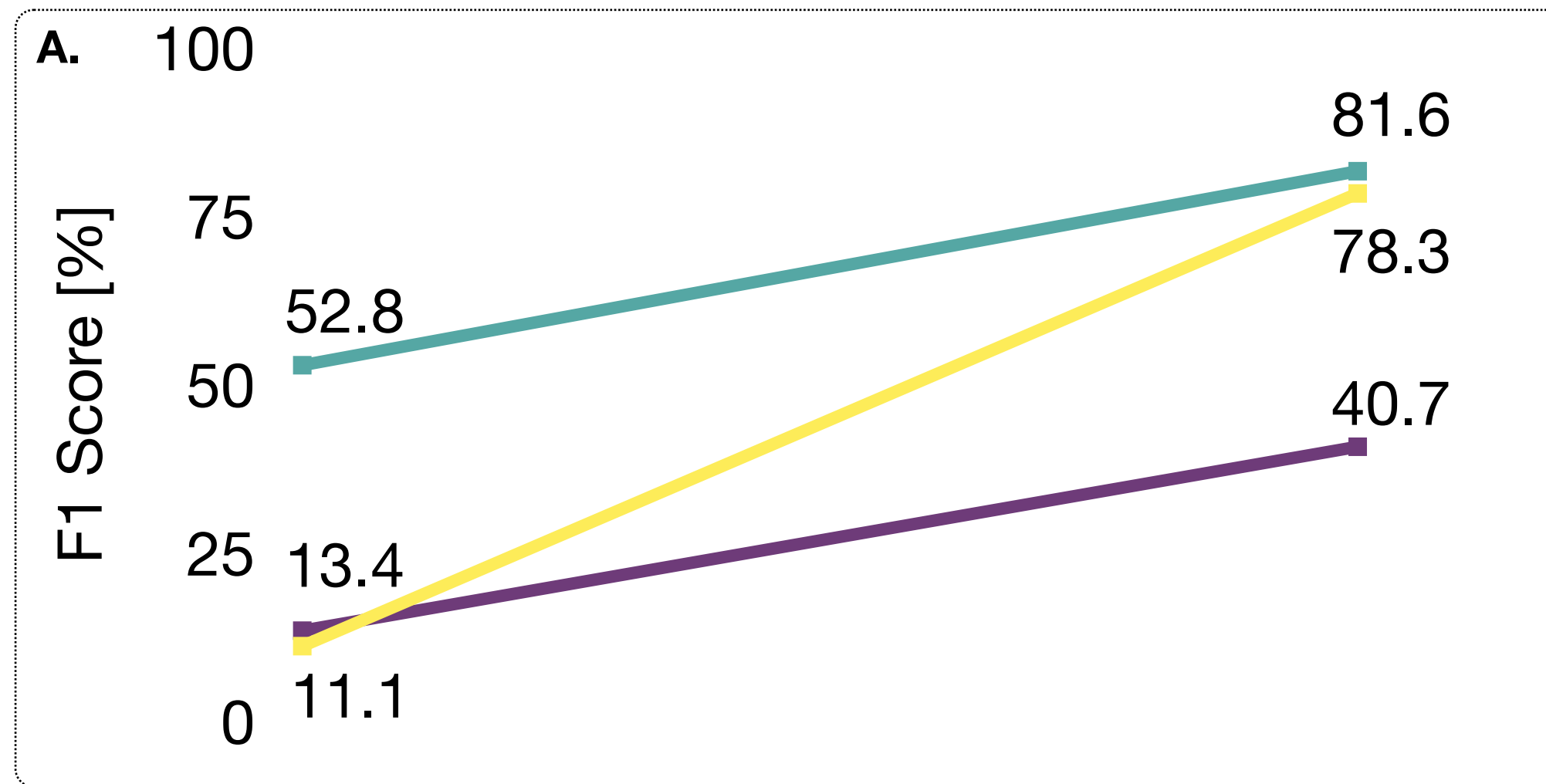
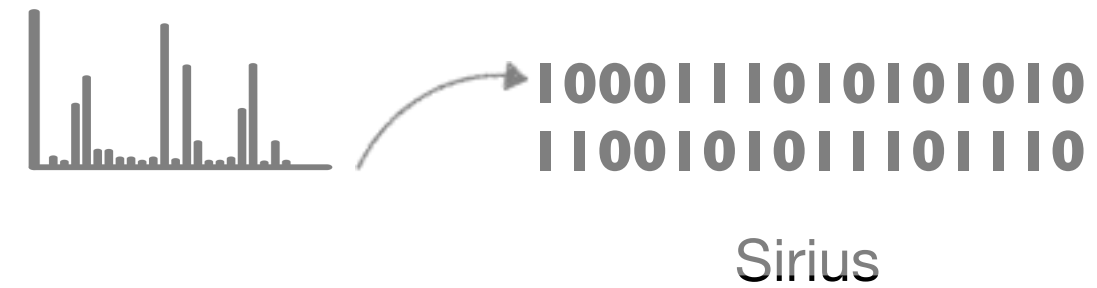
**F1 score  $\approx$  correct candidates at rank 1**

# Evaluation

Spectral similarity



Fingerprint similarity



**F1 score  $\approx$  correct candidates at rank 1**

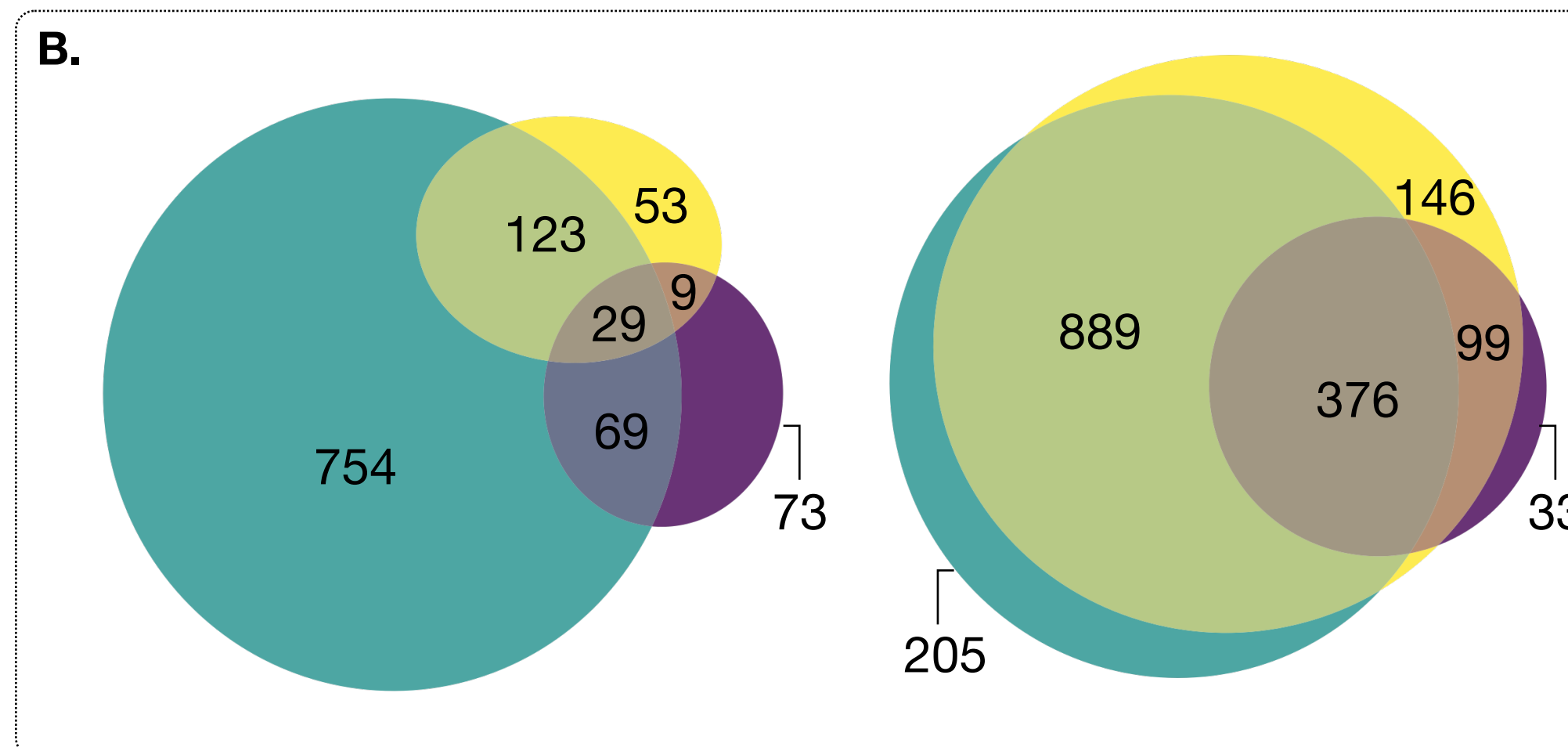
**Before TIMA**

ISDB-DNP

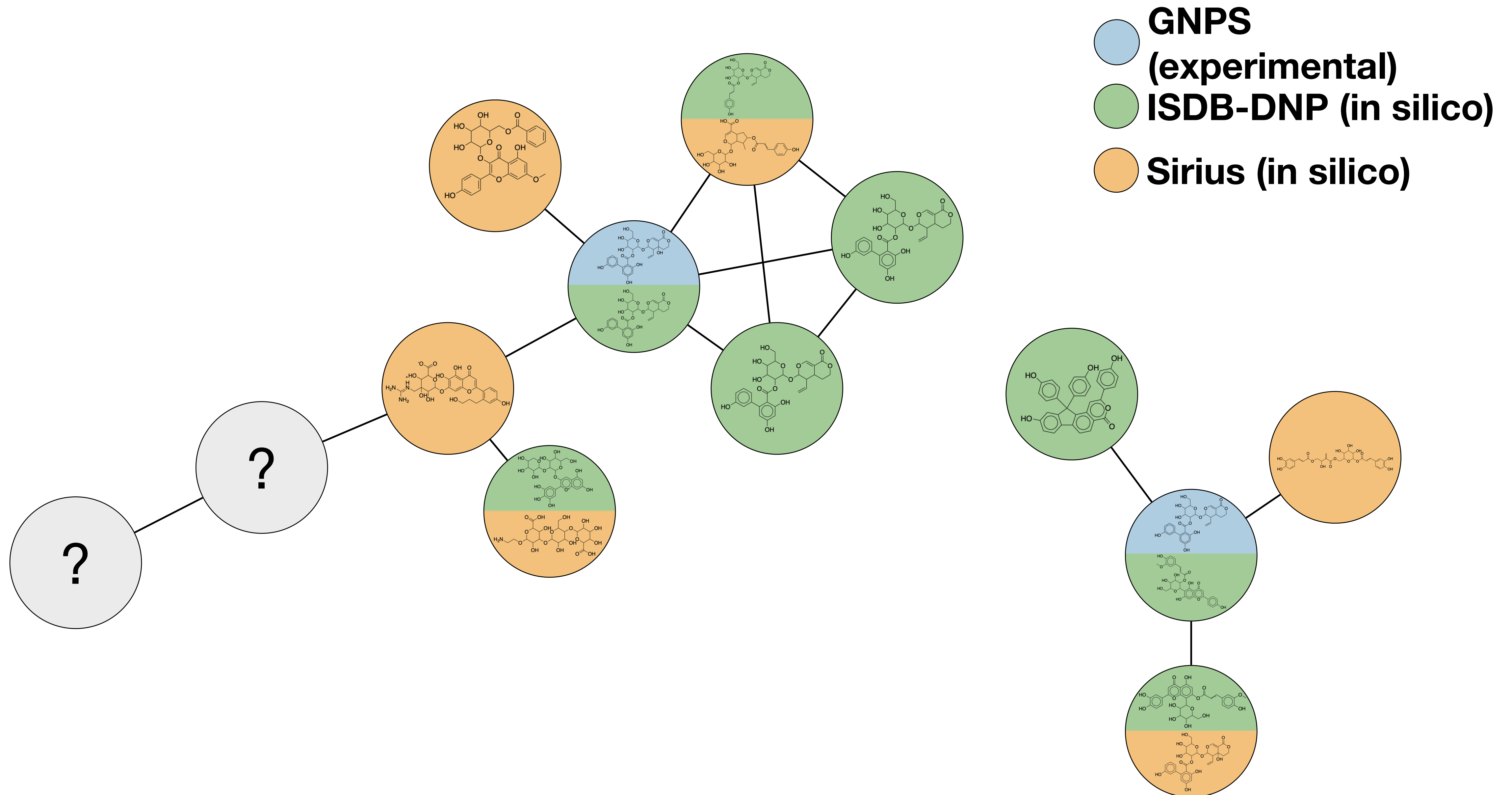
Sirius

**After TIMA**

MS-Finder

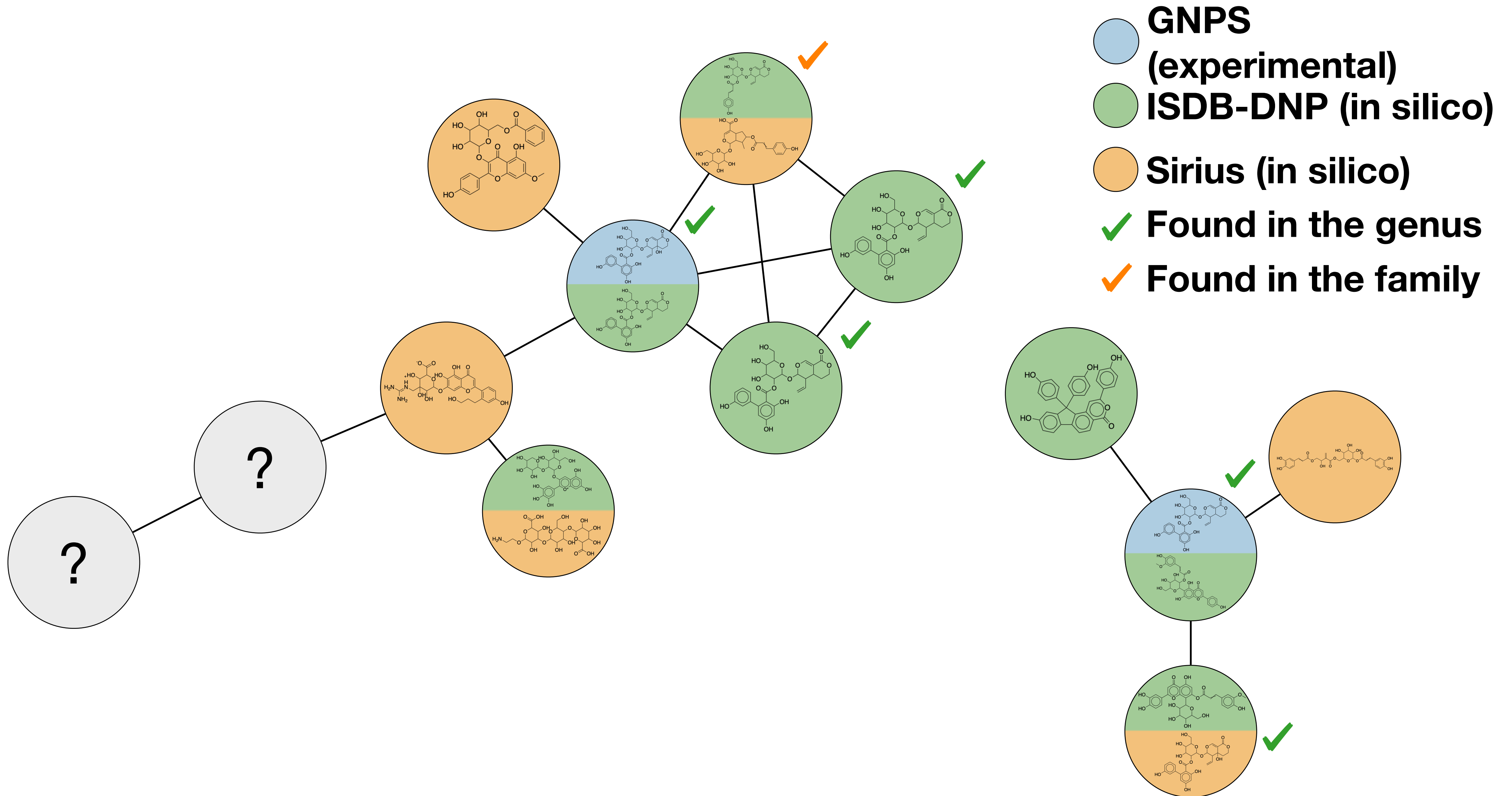


# Taxonomically Informed Metabolite Annotation

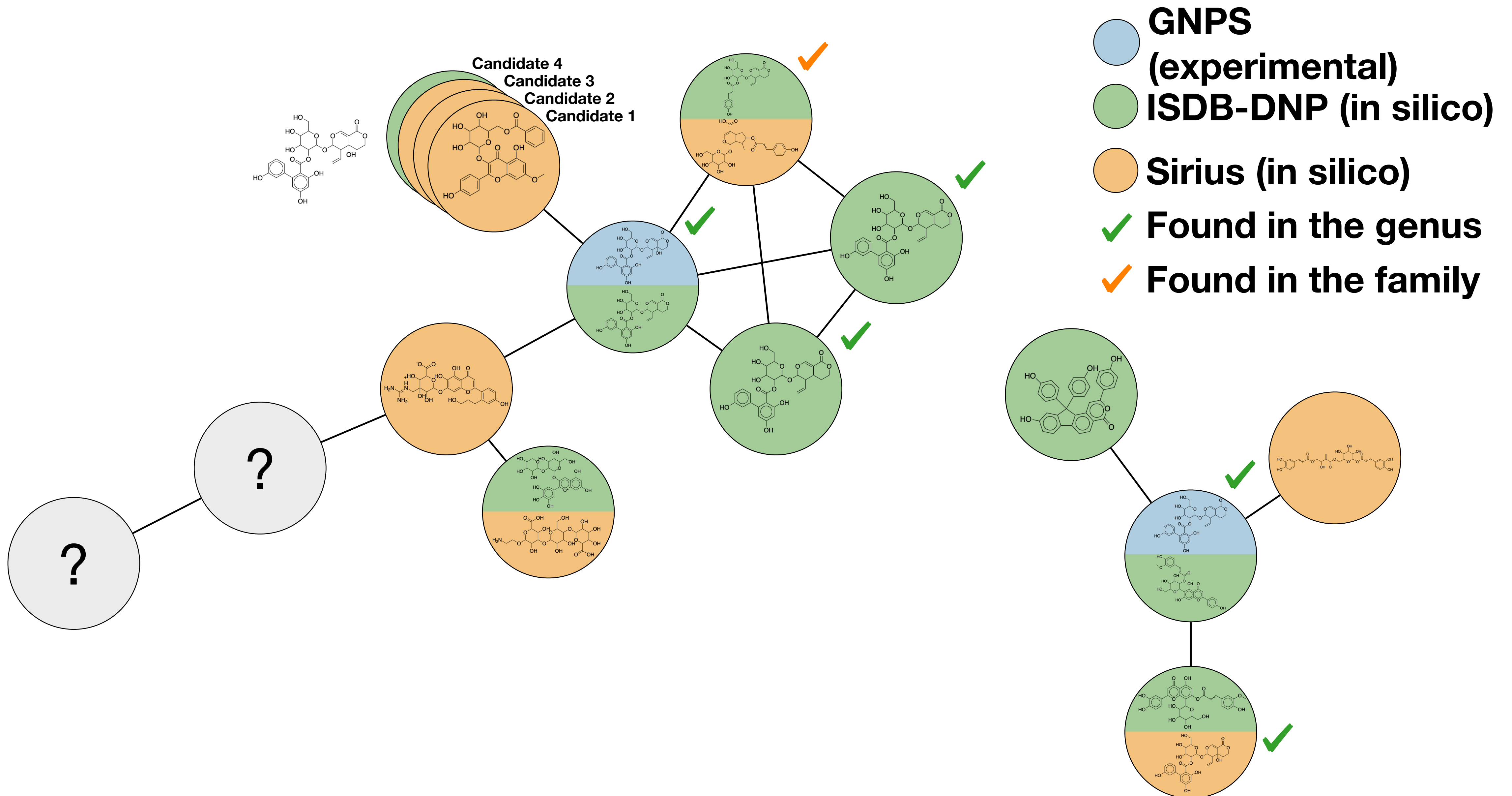




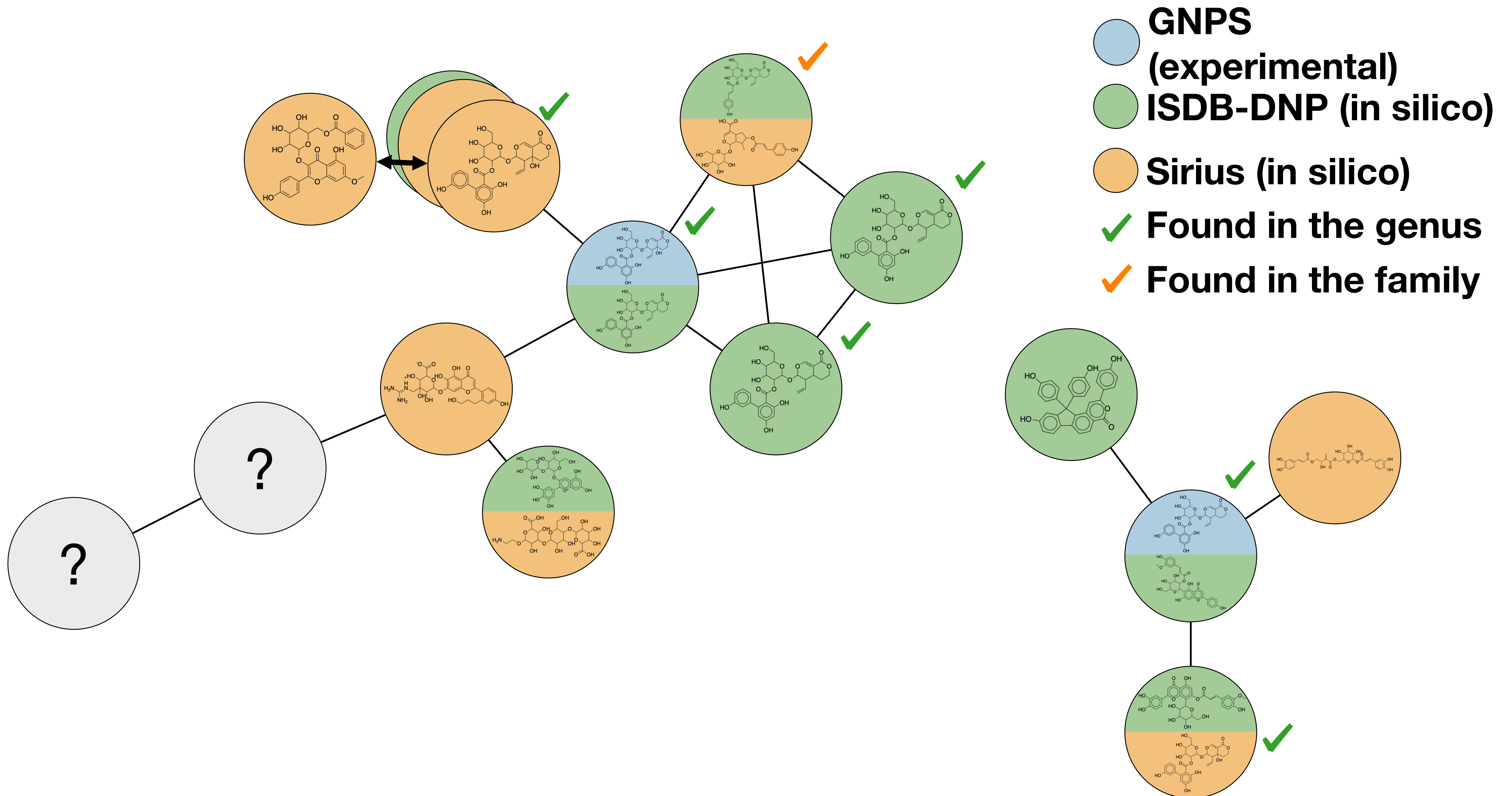
# Taxonomically Informed Metabolite Annotation



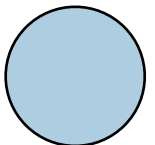
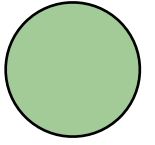
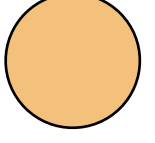


# Taxonomically Informed Metabolite Annotation

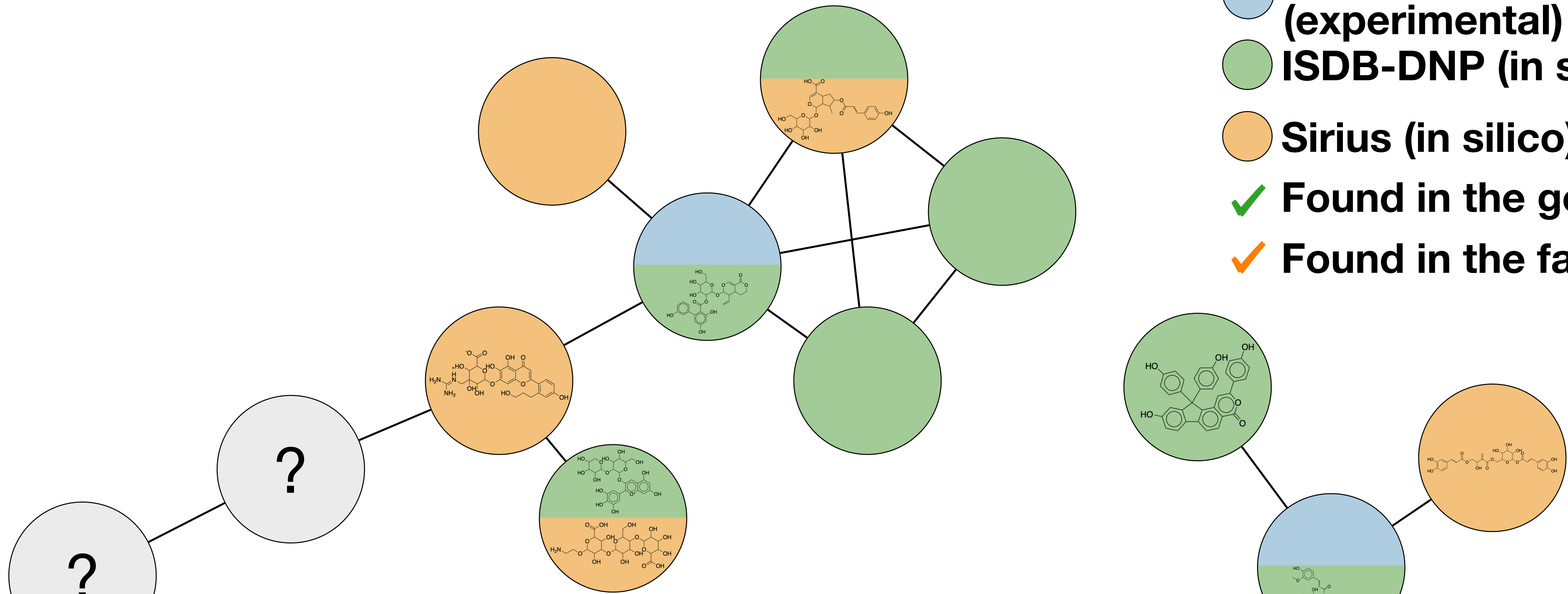


# Taxonomically Informed Metabolite Annotation

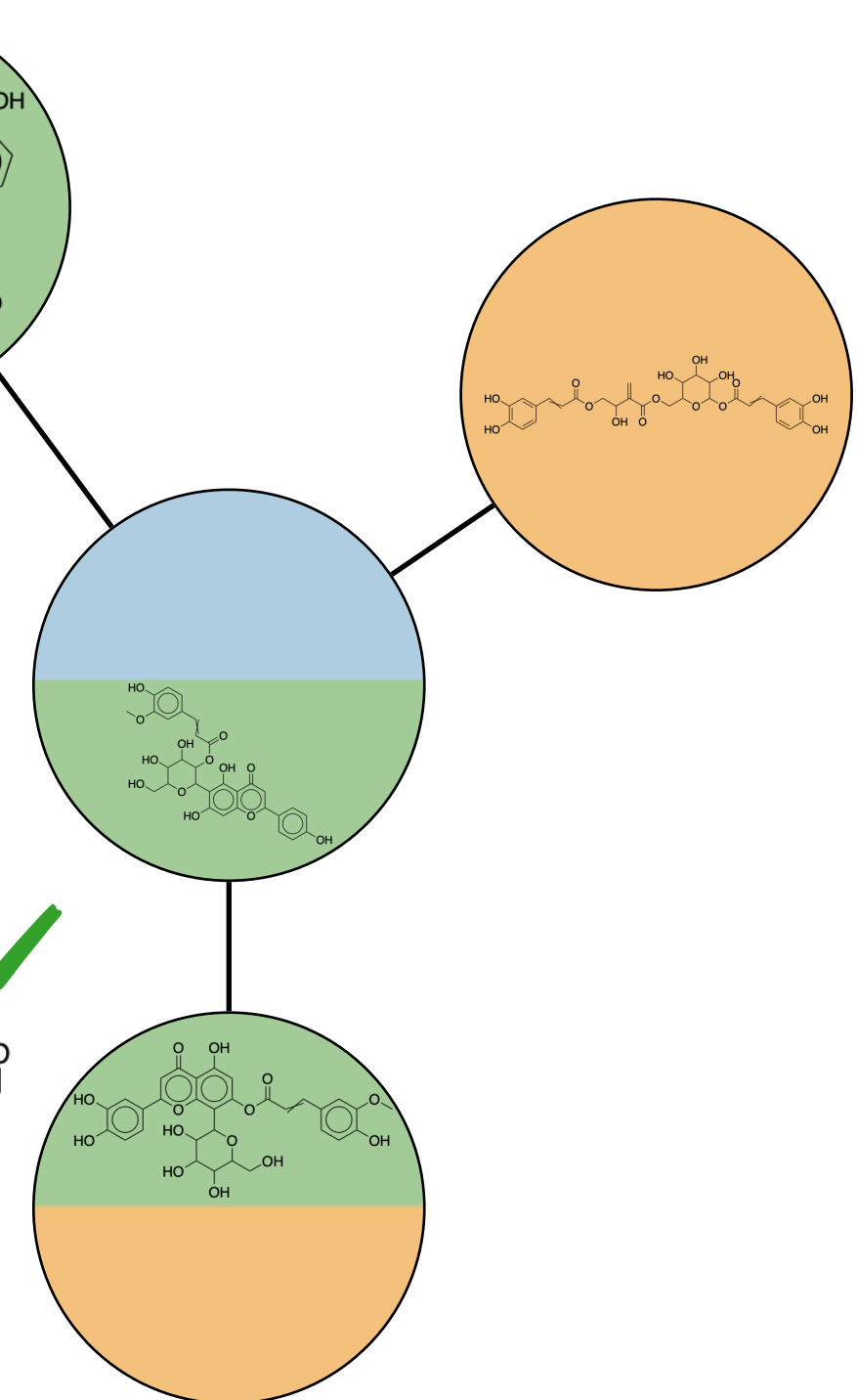
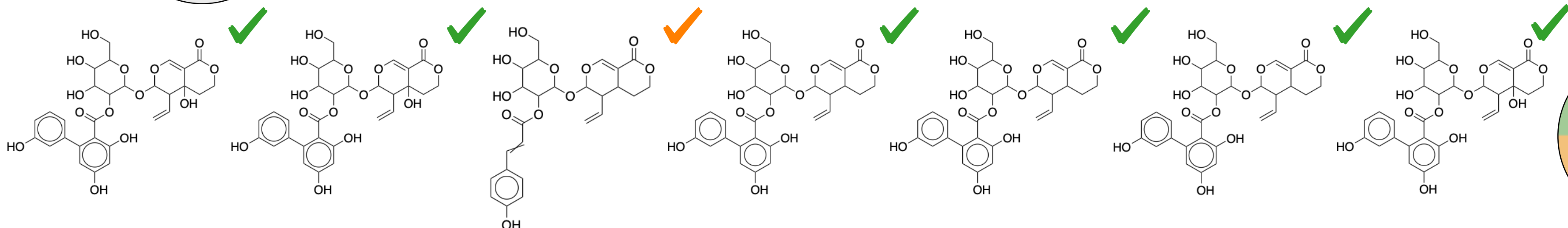


# Taxonomically Informed Metabolite Annotation

-  **GNPS (experimental)**
-  **ISDB-DNP (in silico)**
-  **Sirius (in silico)**
-  **Found in the genus**
-  **Found in the family**


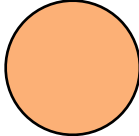
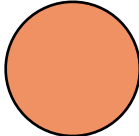
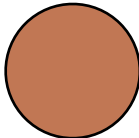
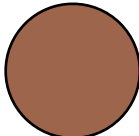
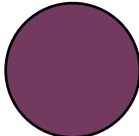


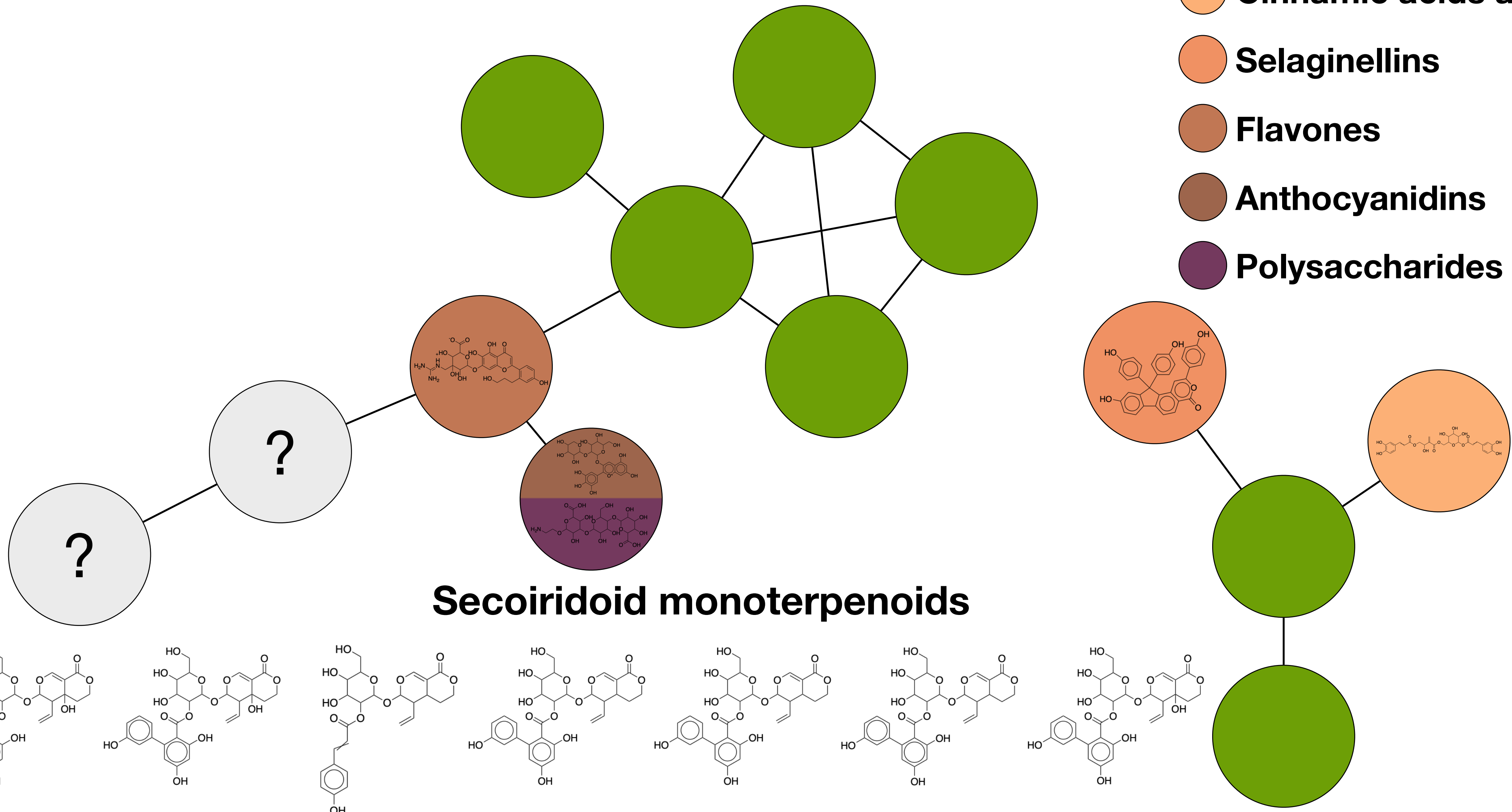
## Secoiridoid monoterpenoids





# Taxonomically Informed Metabolite Annotation

-  **Secoiridoid monoterpenoids**
-  **Cinnamic acids and derivative**
-  **Selaginellins**
-  **Flavones**
-  **Anthocyanidins**
-  **Polysaccharides**



# Taxonomically Informed Metabolite Annotation

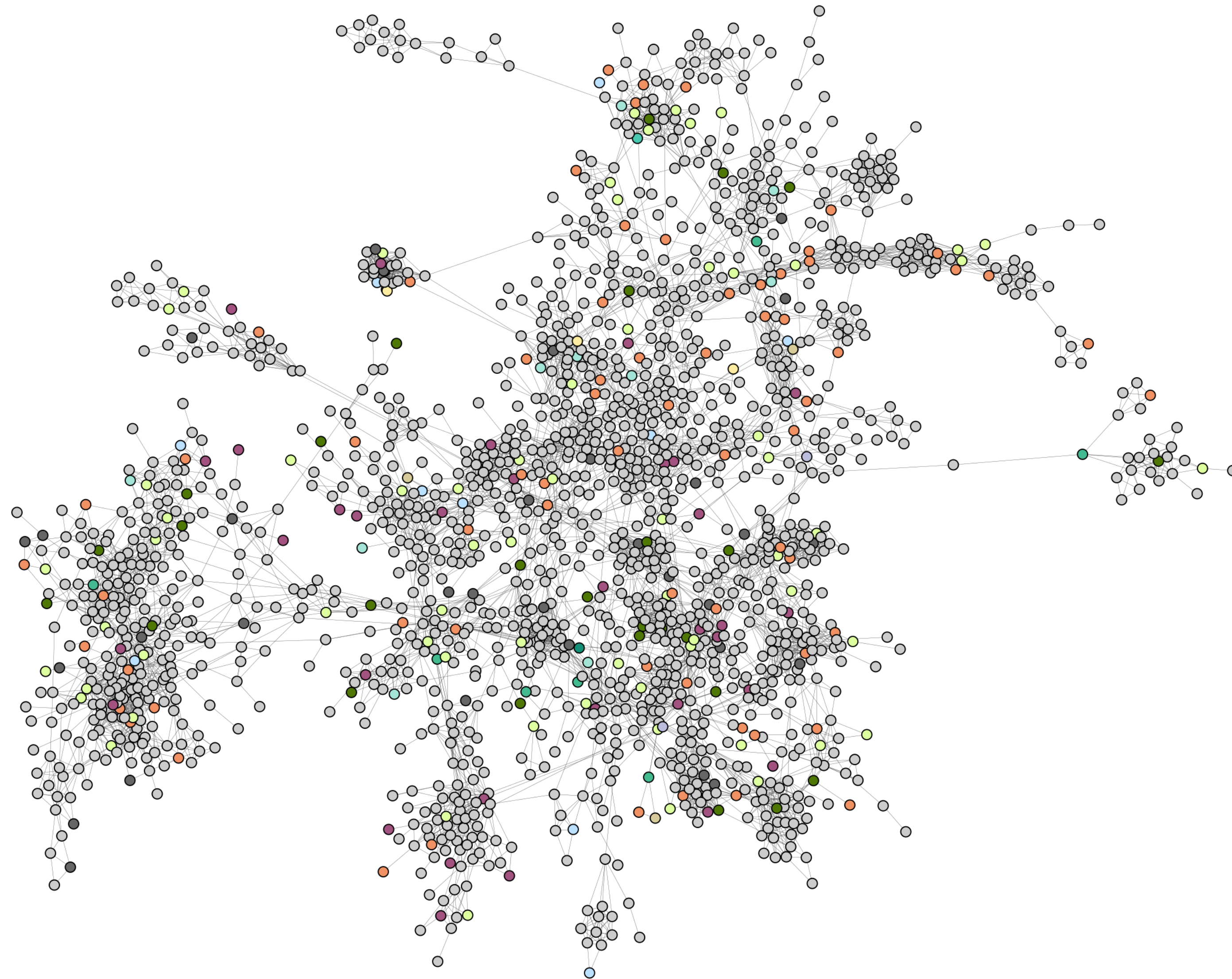
---





# Taxonomically Informed Metabolite Annotation

---




# Establishment of an Open Knowledge Base for Natural Products Research






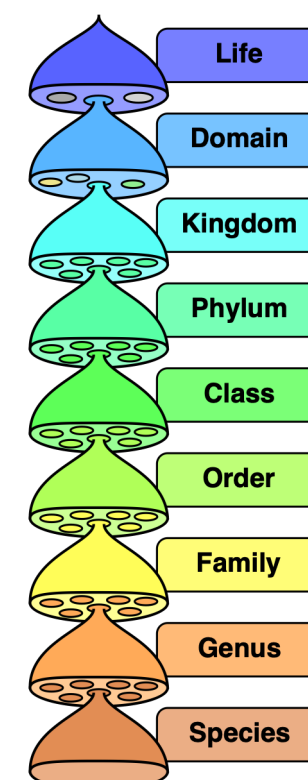
*Taxonomy*



Biological organism




Morphological traits

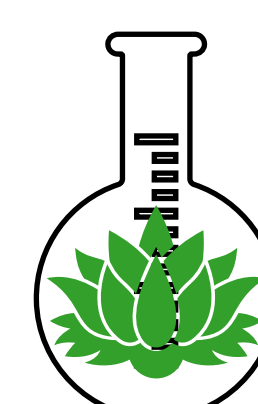


Taxonomy

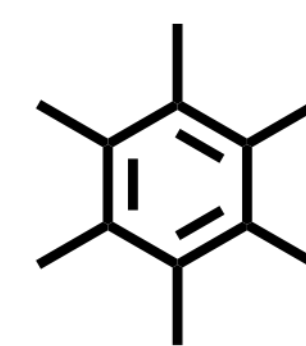
# LOTUS



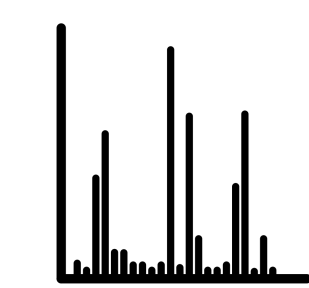
*Metabolomics*



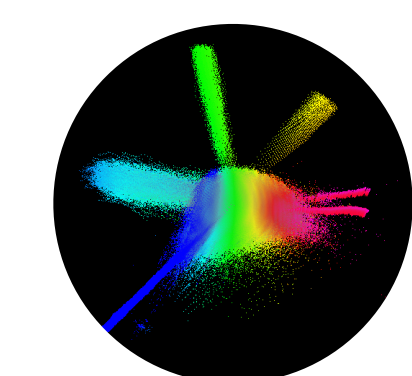
Natural Extract



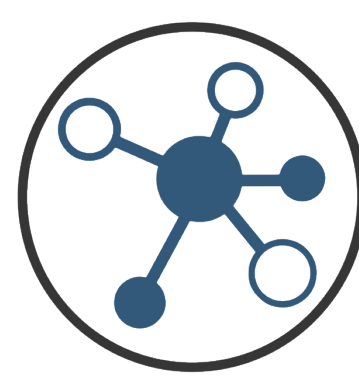
Structure



MS<sup>2</sup> spectrum

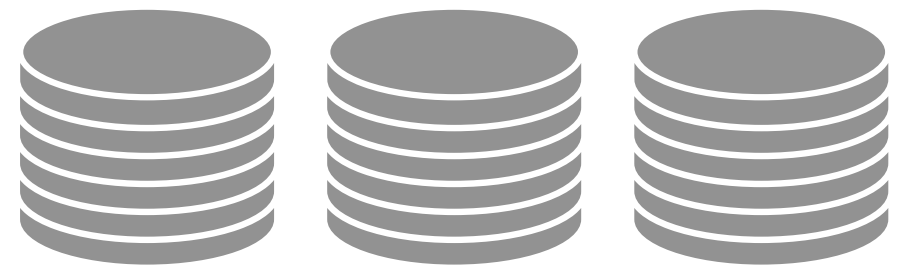


Chemical Space



Molecular Network

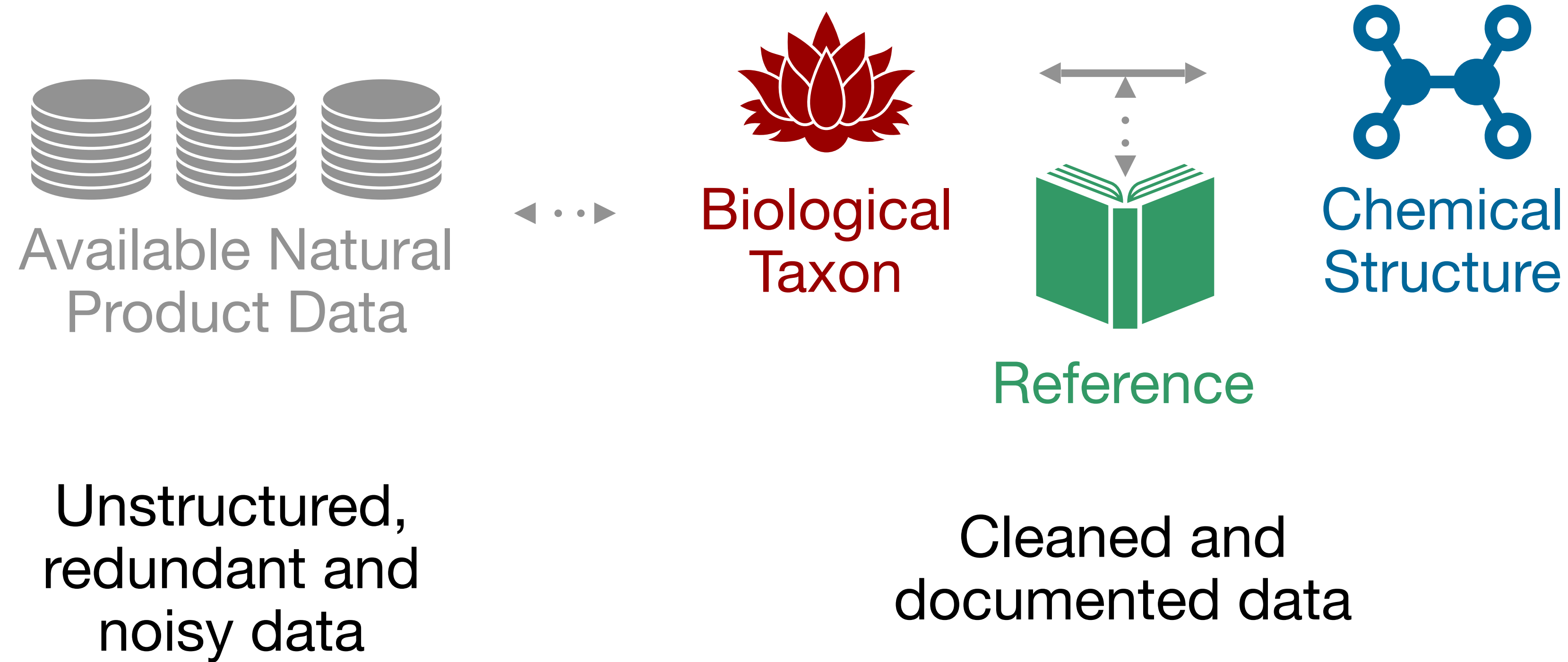
# The initiative



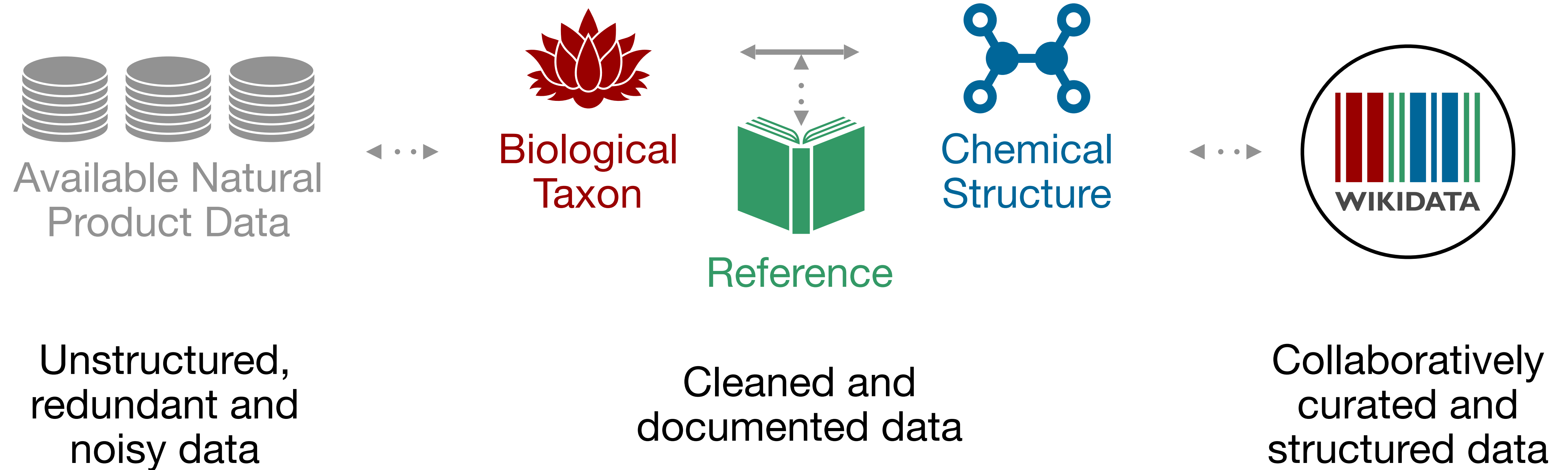
Available Natural  
Product Data

Unstructured,  
redundant and  
noisy data

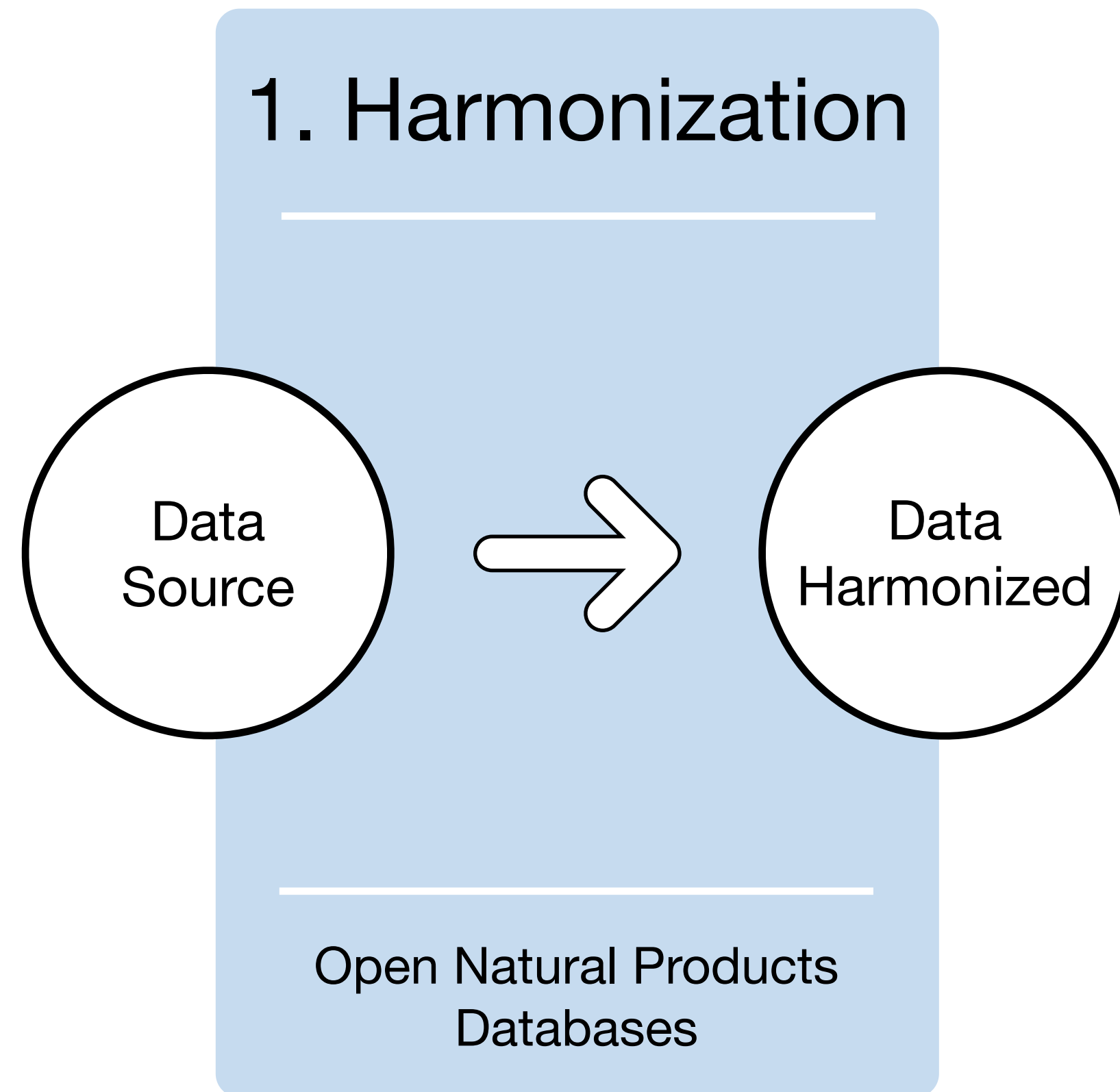
# The initiative



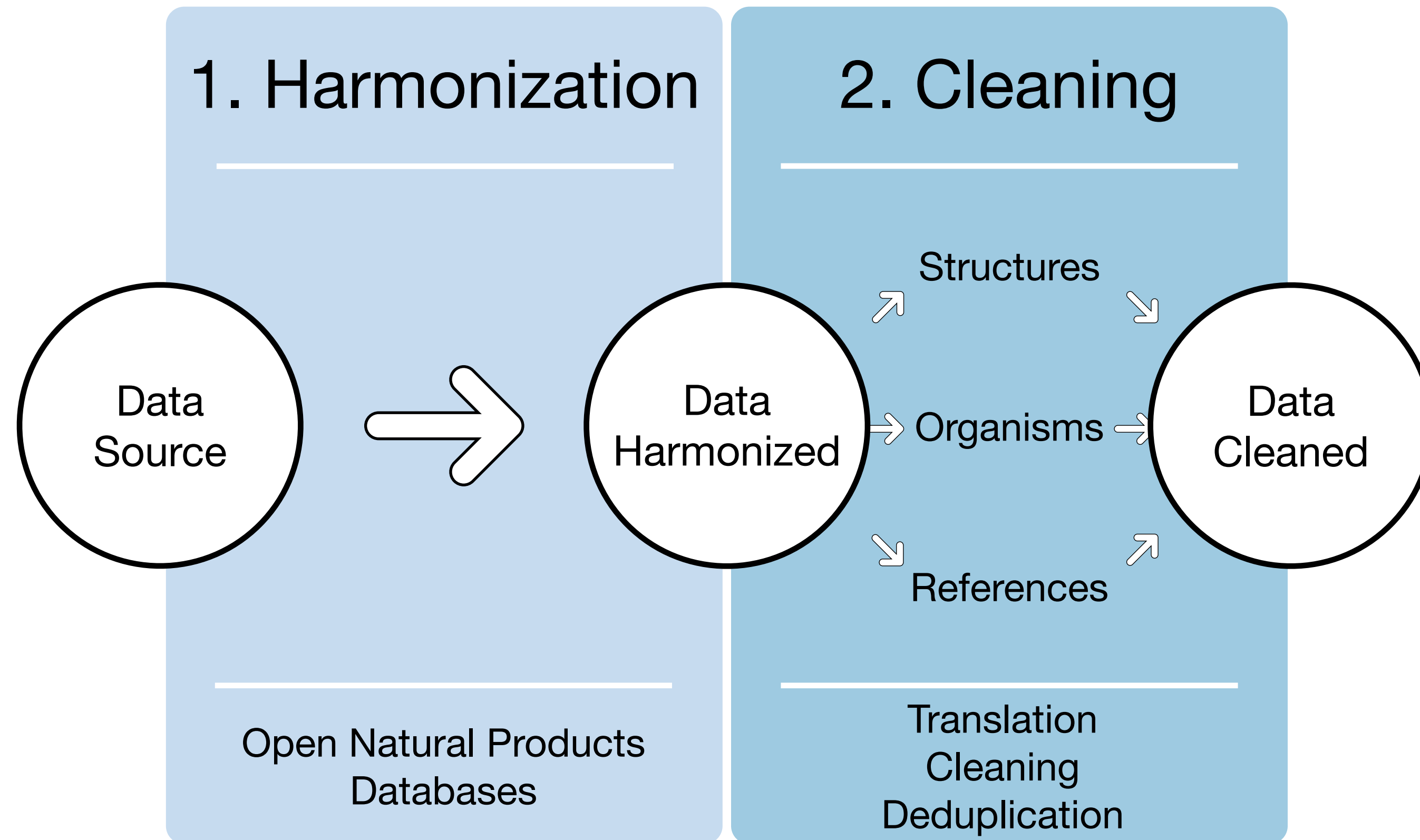
# The initiative



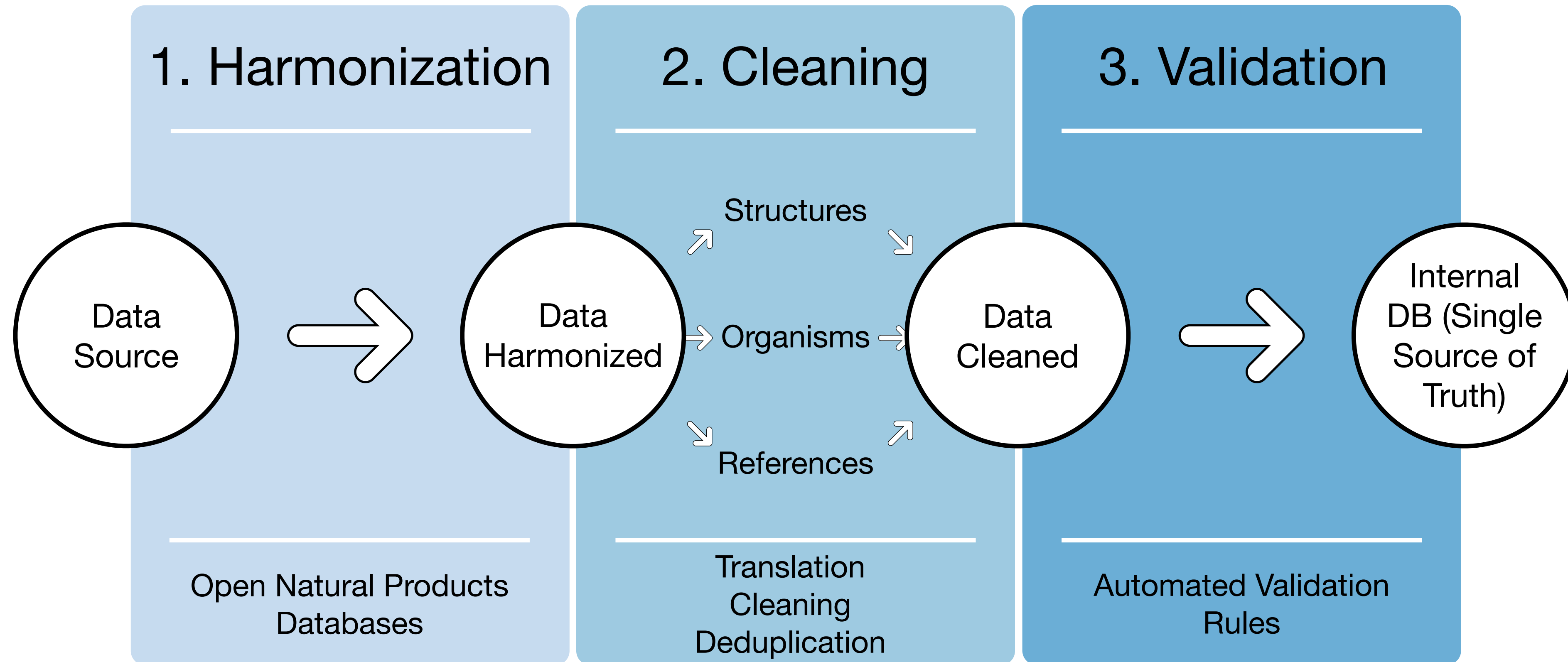
# The initiative - How?



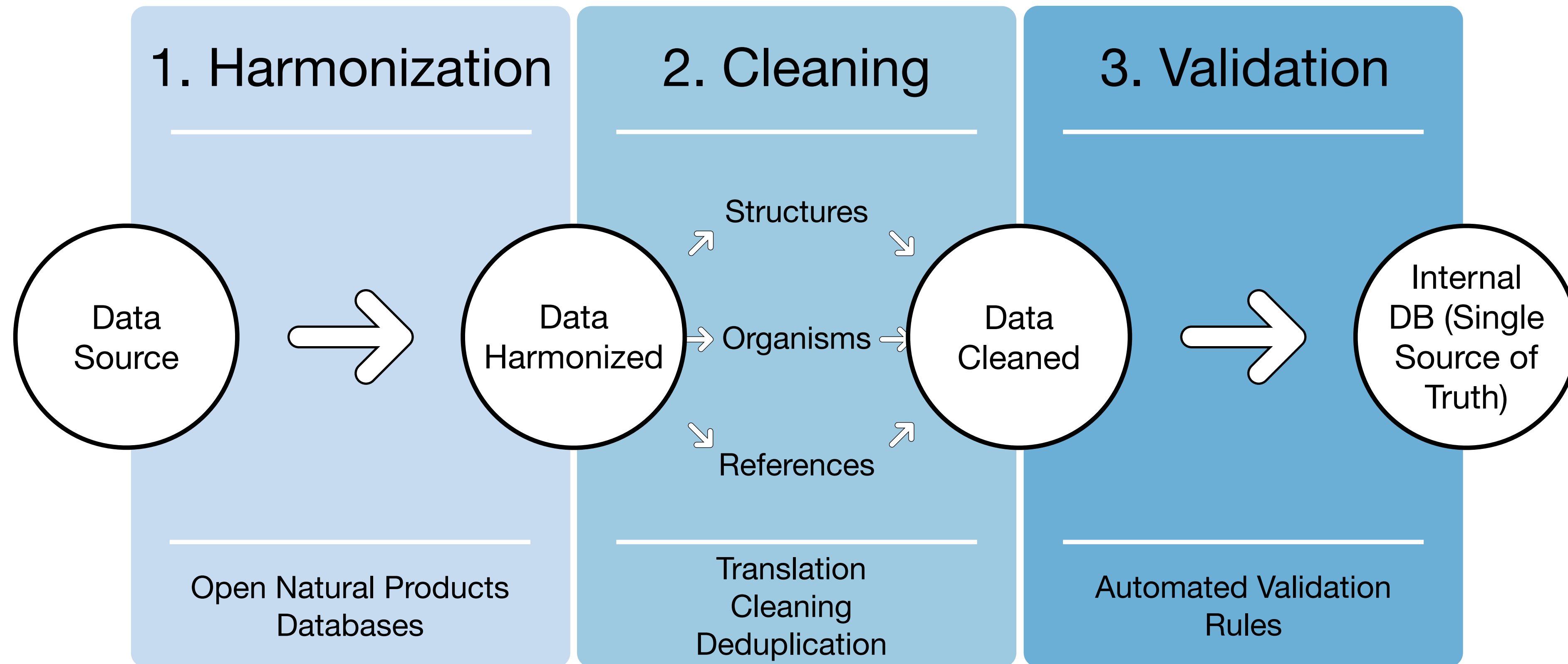
# The initiative - How?



# The initiative - How?



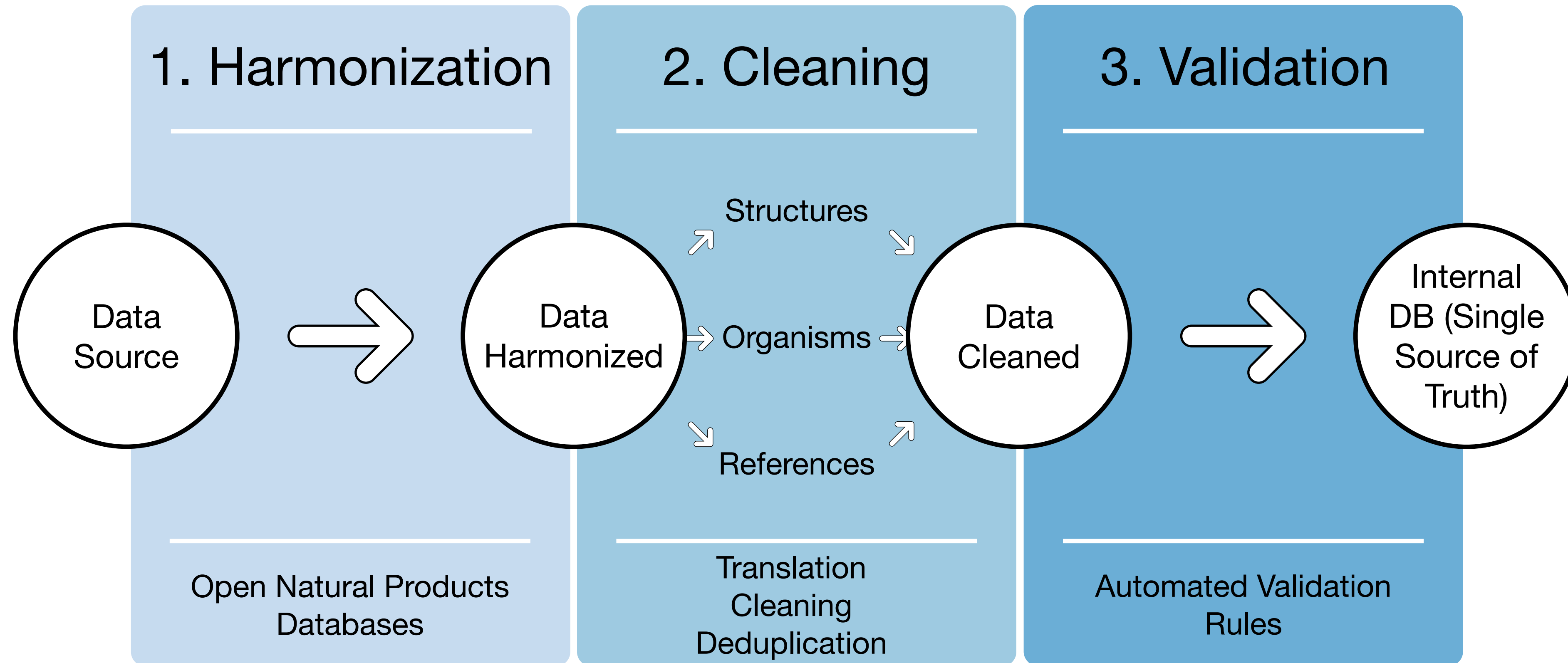
# The initiative - How?



	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of <i>Cyathocalyx zeylanica</i> CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).

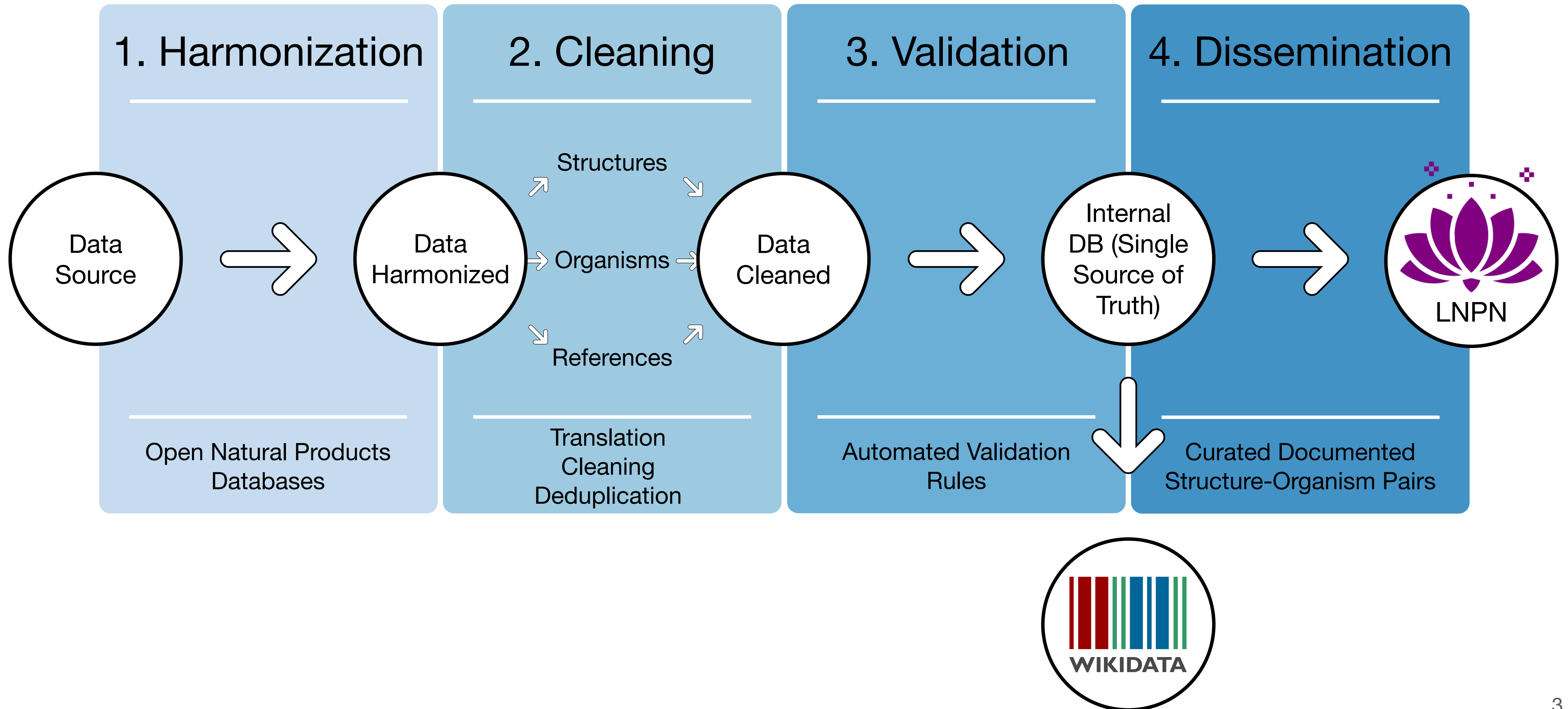


# The initiative - How?

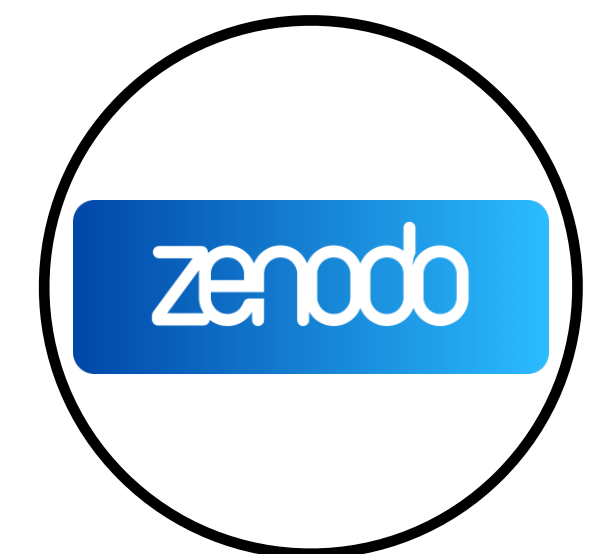
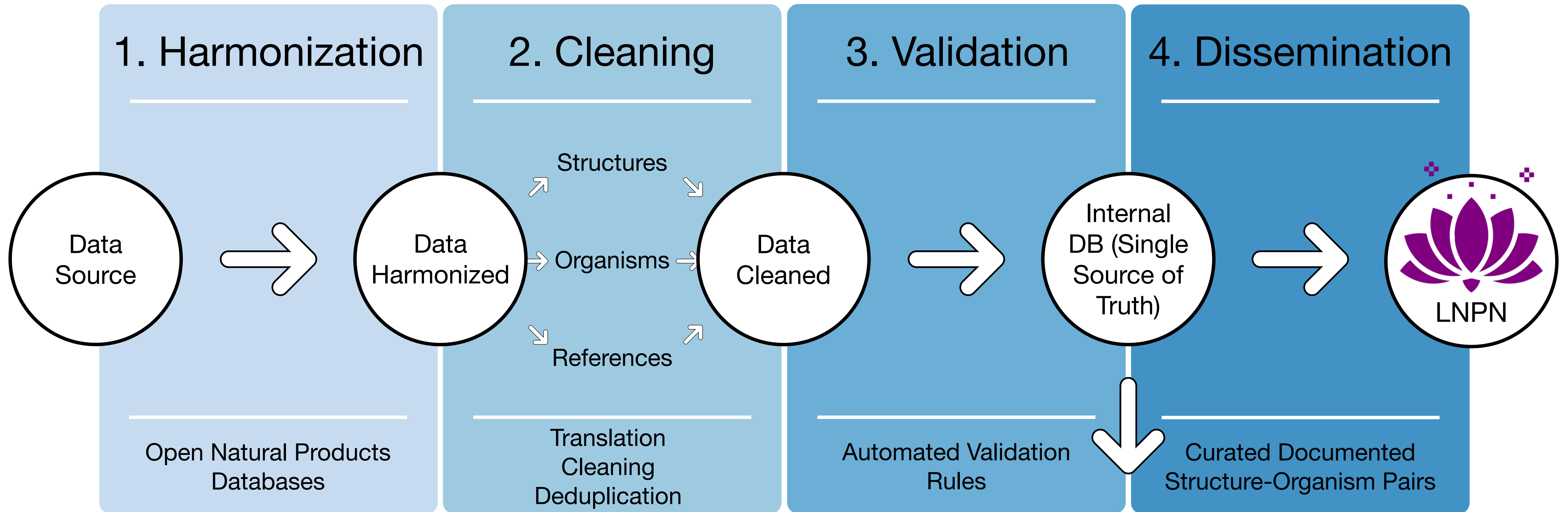


	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of <i>Cyathocalyx zeylanica</i> CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).
After curation	VFIIVOHWCNHINZ-UHFFFAOYSA-N	<i>Cyathocalyx zeylanicus</i>	10.1021/NP50117A020

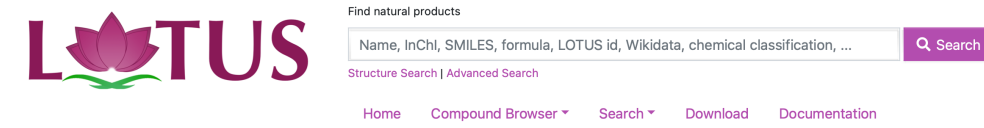
# The initiative - Where?



# The initiative - Where?



# The initiative - Where?



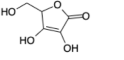
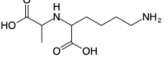
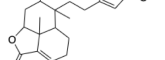
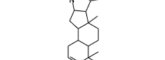
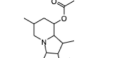
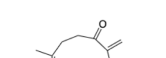
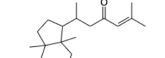
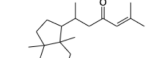
Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

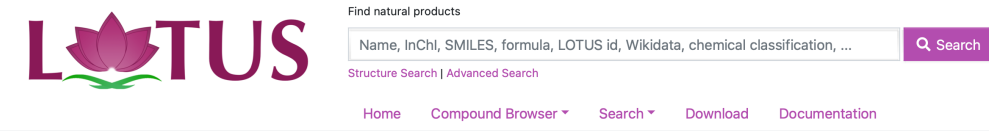
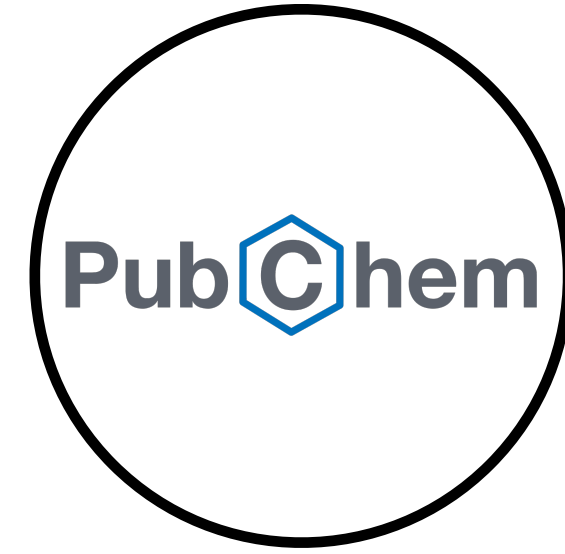
Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

 <p><b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one Mol. formula C<sub>5</sub>H<sub>6</sub>O<sub>5</sub> Mol. weight 146.1 Temp. LOTUS id LTS0249032</p>	 <p><b>Q27102265</b> Lysopine Mol. formula C<sub>20</sub>H<sub>35</sub>N<sub>2</sub>O<sub>4</sub> Mol. weight 353.52 Temp. LOTUS id LTS0160430</p>	 <p><b>Q105387202</b> (1r,8s,9r,10s,12r)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0<sup>1,10</sup>]dodec-4-en-3-one Mol. formula C<sub>20</sub>H<sub>28</sub>O<sub>4</sub> Mol. weight 330.42 Temp. LOTUS id LTS0145658</p>	 <p><b>Q105387201</b> (1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-en-18-yl acetate Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 458.67 Temp. LOTUS id LTS0165716</p>
 <p><b>Q105387200</b> 7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-en-18-yl acetate Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0157374</p>	 <p><b>LTS0044532</b> (4z,8e)-4,7,7-trimethyl-11-methylidenecycloundeca-4,8-dien-1-one Mol. formula C<sub>19</sub>H<sub>28</sub>O Mol. weight 274.44 Temp. LOTUS id LTS0044532</p>	 <p><b>Q105387197</b> 6-(4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-15-yl)-2-methylhept-2-en-4-one Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0237245</p>	 <p><b>Q105387196</b> (6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r)-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-15-yl]-2-methylhept-2-en-4-one Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0074160</p>

# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

<p><b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one</p> <p>Mol. formula C<sub>5</sub>H<sub>6</sub>O<sub>5</sub> Mol. weight 146.1 Temp. LOTUS id L750249032</p>	<p><b>Q27102265</b> Lysopine</p> <p>Mol. formula C<sub>9</sub>H<sub>18</sub>N<sub>2</sub>O<sub>4</sub> Mol. weight 218.25 Temp. LOTUS id L750160430</p>	<p><b>Q105387202</b> (1r,8s,9r,10s,12r)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0<sup>1,4</sup>]dodec-4-en-3-one</p> <p>Mol. formula C<sub>20</sub>H<sub>28</sub>O<sub>4</sub> Mol. weight 330.42 Temp. LOTUS id L750145658</p>	<p><b>Q105387201</b> (1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-tetracos-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 456.57 Temp. LOTUS id L750165716</p>
<p><b>Q105387200</b> 7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>.0<sup>1,10</sup>]-tetracos-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L750157374</p>	<p><b>L750044532</b> (4z,8e)-4,7,7-trimethyl-11-methylidenecycloundeca-4,8-dien-1-one</p> <p>Mol. formula C<sub>19</sub>H<sub>28</sub>O Mol. weight 274.44 Temp. LOTUS id L750044532</p>	<p><b>Q105387197</b> 6-(4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,9</sup>.0<sup>1,9</sup>.0<sup>1,9</sup>.0<sup>1,9</sup>]-15-yl)-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L750237245</p>	<p><b>Q105387196</b> (6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r,16r)-4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,9</sup>.0<sup>1,9</sup>.0<sup>1,9</sup>.0<sup>1,9</sup>]-15-yl]-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L750074160</p>

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records (e.g. phenylpropanoates, or Gene Ontology: DNA repair). [More...](#)

Select classification Search

**LOTUS - the natural products occurrence database** K

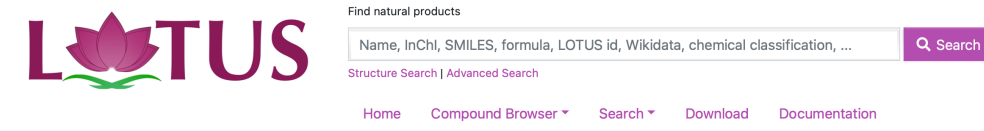
Classification description (from **LOTUS - the natural products occurrence database**)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 [More...](#)

Data type counts to display Display zero count nodes?

**None** **Compound** **Taxonomy** **Yes** **No**



# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

 <b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one Mol. formula C <sub>5</sub> H <sub>6</sub> O <sub>5</sub> Mol. weight 146.1 Temp. LOTUS id LTS0249032	 <b>Q27102265</b> Lysopine Mol. formula C <sub>9</sub> H <sub>18</sub> N <sub>2</sub> O <sub>4</sub> Mol. weight 218.25 Temp. LOTUS id LTS0160430	 <b>Q105387202</b> (1r,8s,9r,10s,12r)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0 <sup>1,4</sup> ]dodec-4-en-3-one Mol. formula C <sub>20</sub> H <sub>28</sub> O <sub>4</sub> Mol. weight 330.42 Temp. LOTUS id LTS0145658	 <b>Q105387201</b> (1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>2,11</sup> .0 <sup>3,12</sup> .0 <sup>4,13</sup> ]tetracos-4-en-18-yl acetate Mol. formula C <sub>29</sub> H <sub>45</sub> N <sub>3</sub> O <sub>3</sub> Mol. weight 458.67 Temp. LOTUS id LTS0165716
 <b>Q105387200</b> 7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>2,11</sup> .0 <sup>3,12</sup> .0 <sup>4,13</sup> ]tetracos-4-en-18-yl acetate Mol. formula C <sub>29</sub> H <sub>45</sub> N <sub>3</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0157374	 <b>LTS0044532</b> (4z,8e)-4,7,7-trimethyl-11-methylidenecycloundeca-4,8-dien-1-one Mol. formula C <sub>19</sub> H <sub>28</sub> O Mol. weight 218.34 Temp. LOTUS id LTS0044532	 <b>Q105387197</b> 6-(4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,9</sup> .0 <sup>2,10</sup> .0 <sup>3,11</sup> .0 <sup>4,12</sup> ]-15-yl)-2-methylhept-2-en-4-one Mol. formula C <sub>30</sub> H <sub>48</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0237245	 <b>Q105387196</b> (6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r)-7,7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,9</sup> .0 <sup>2,10</sup> .0 <sup>3,11</sup> .0 <sup>4,12</sup> ]-15-yl]-2-methylhept-2-en-4-one Mol. formula C <sub>30</sub> H <sub>48</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id LTS0074160

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records (phenylpropionates, or Gene Ontology: DNA repair). [More...](#)

Select classification Search

**LOTUS - the natural products occurrence database** K

Classification description (from **LOTUS - the natural products occurrence database**)  
 Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
 Created on 09/10/2022 09:05:48 [More...](#)

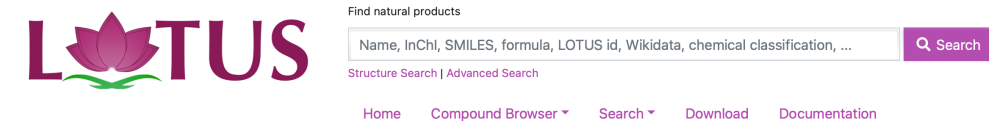
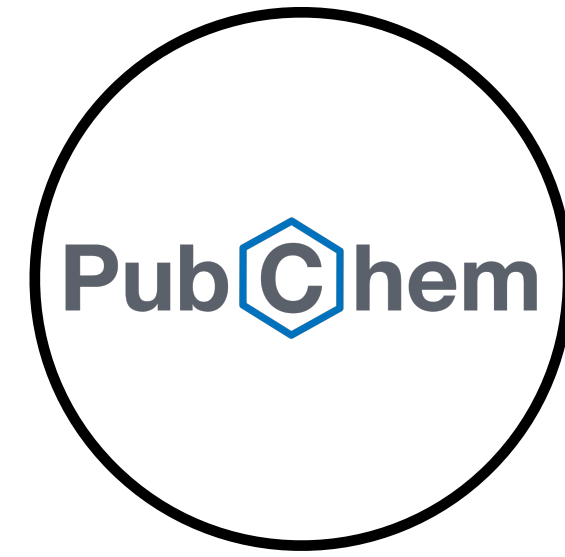
Data type counts to display Display zero count nodes?

**None** **Compound** **Taxonomy** **Yes** **No**

## Browse LOTUS - the natural products occurrence database Tree

- ▼ LOTUS Tree ? 214,481
- ▼ Biological Tree 192,579
- ▶ Archaea 72
- ▶ Bacteria 17,467
- ▶ Eukaryota 175,387

# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

<p><b>Q105387204</b></p> <p>3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one</p> <p>Mol. formula C<sub>5</sub>H<sub>6</sub>O<sub>5</sub></p> <p>Mol. weight 146.1</p> <p>Temp. LOTUS id L7S0249032</p>	<p><b>Q27102265</b></p> <p>Lysopine</p> <p>Mol. formula C<sub>9</sub>H<sub>18</sub>N<sub>2</sub>O<sub>4</sub></p> <p>Mol. weight 218.25</p> <p>Temp. LOTUS id L7S0160430</p>	<p><b>Q105387202</b></p> <p>(1r,8s,9r,10s,12z)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0<sup>1,2</sup>]dodec-4-en-3-one</p> <p>Mol. formula C<sub>20</sub>H<sub>28</sub>O<sub>4</sub></p> <p>Mol. weight 330.42</p> <p>Temp. LOTUS id L7S0145658</p>	<p><b>Q105387201</b></p> <p>(1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>4,9</sup>.0<sup>11,12</sup>.0<sup>13,14</sup>.0<sup>15,16</sup>]octacos-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub></p> <p>Mol. weight 458.67</p> <p>Temp. LOTUS id L7S0165716</p>
<p><b>Q105387200</b></p> <p>7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>4,9</sup>.0<sup>11,12</sup>.0<sup>13,14</sup>.0<sup>15,16</sup>]octacos-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub></p> <p>Mol. weight 456.7</p> <p>Temp. LOTUS id L7S0157374</p>	<p><b>L7S0044532</b></p> <p>(4z,8e)-4,7,7-trimethyl-11-methylidenecycloundeca-4,8-dien-1-one</p> <p>Mol. formula C<sub>19</sub>H<sub>28</sub>O</p> <p>Mol. weight 278.44</p> <p>Temp. LOTUS id L7S0044532</p>	<p><b>Q105387197</b></p> <p>6-(4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>11,12</sup>.0<sup>13,14</sup>.0<sup>15,16</sup>]-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub></p> <p>Mol. weight 456.7</p> <p>Temp. LOTUS id L7S0237245</p>	<p><b>Q105387196</b></p> <p>(6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r,16r)-4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>11,12</sup>.0<sup>13,14</sup>.0<sup>15,16</sup>]-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub></p> <p>Mol. weight 456.7</p> <p>Temp. LOTUS id L7S0074160</p>

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records (phenylpropanoates, or Gene Ontology: DNA repair). [More...](#)

Select classification Search

**LOTUS - the natural products occurrence database** K

Classification description (from LOTUS - the natural products occurrence database)

Biological and chemical tree provided by the LOTUS (natural products occurrence database)

Created on 09/10/2022 09:05:48 [More...](#)

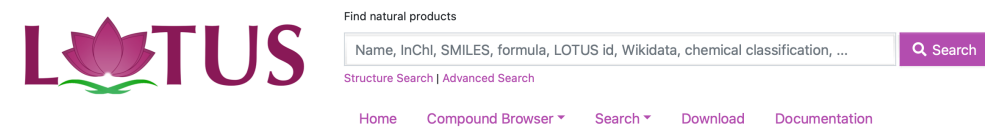
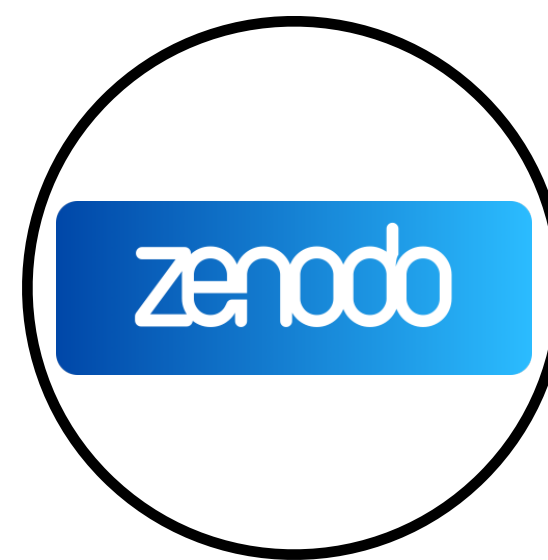
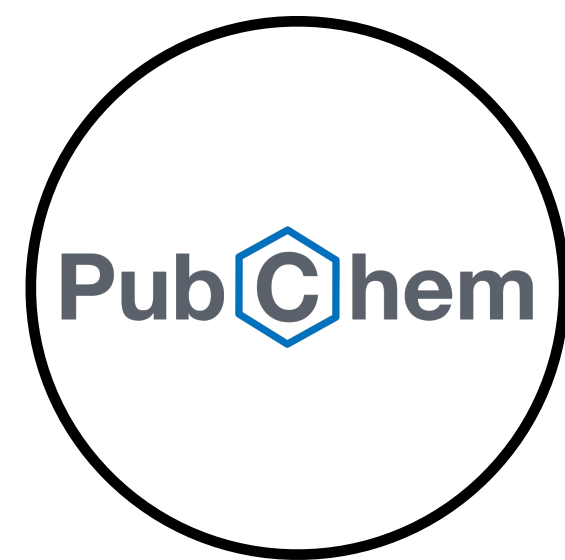
Data type counts to display Display zero count nodes?

**None** **Compound** **Taxonomy** **Yes** **No**

## Browse LOTUS - the natural products occurrence database Tree

- ▼ LOTUS Tree ? 214,481
- ▼ Biological Tree 192,579
  - ▶ Archaea 72
  - ▶ Bacteria 17,467
  - ▶ Eukaryota 175,387
- ▼ Chemical Tree 214,481
  - ▶ Alkaloids 28,065
  - ▶ Amino acids and Peptides 13,012
  - ▶ Carbohydrates 2,144
  - ▶ Fatty acids 10,266
  - ▶ Polyketides 27,981
  - ▶ Shikimates and Phenylpropanoids 42,177
  - ▶ Terpenoids 97,499
  - ▶ Not classified 7,328

# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

 <b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one Mol. formula C <sub>5</sub> H <sub>6</sub> O <sub>5</sub> Mol. weight 146.1 Temp. LOTUS id L7S0249032	 <b>Q27102265</b> Lysopine Mol. formula C <sub>20</sub> H <sub>35</sub> N <sub>2</sub> O <sub>4</sub> Mol. weight 353.52 Temp. LOTUS id L7S0160430	 <b>Q105387202</b> (1r,8s,9r,10s,12r)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0 <sup>1,4</sup> ]dodec-4-en-3-one Mol. formula C <sub>20</sub> H <sub>26</sub> O <sub>4</sub> Mol. weight 330.42 Temp. LOTUS id L7S0145658	 <b>Q105387201</b> (1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>2,9</sup> .0 <sup>3,12</sup> ]tetracos-4-en-18-yl acetate Mol. formula C <sub>29</sub> H <sub>45</sub> N <sub>3</sub> O <sub>3</sub> Mol. weight 458.67 Temp. LOTUS id L7S0165716
 <b>Q105387200</b> 7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>2,9</sup> .0 <sup>3,12</sup> ]tetracos-4-en-18-yl acetate Mol. formula C <sub>29</sub> H <sub>45</sub> N <sub>3</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L7S0157374	 <b>L7S0044532</b> (4z,8e)-4,7,7-trimethyl-11-methylenecycloundeca-4,8-dien-1-one Mol. formula C <sub>19</sub> H <sub>28</sub> O Mol. weight 272.44 Temp. LOTUS id L7S0044532	 <b>Q105387197</b> (6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r,16r)-4,6-dihydroxy-7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,9</sup> .0 <sup>2,10</sup> .0 <sup>3,11</sup> ]octadecan-15-yl]-2-methylhept-2-en-4-one Mol. formula C <sub>30</sub> H <sub>48</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L7S0237245	 <b>Q105387196</b> (6r)-6-[(1s,3s,4s,6s,8s,11s,12s,15r,16r)-4,6-dihydroxy-7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,9</sup> .0 <sup>2,10</sup> .0 <sup>3,11</sup> ]octadecan-15-yl]-2-methylhept-2-en-4-one Mol. formula C <sub>30</sub> H <sub>48</sub> O <sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L7S0074160

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records (phenylpropanoates, or Gene Ontology: DNA repair). [More...](#)

Select classification Search

**LOTUS - the natural products occurrence database**

Classification description (from **LOTUS - the natural products occurrence database**)  
 Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
 Created on 09/10/2022 09:05:48 [More...](#)

Data type counts to display Display zero count nodes?

**None** **Compound** **Taxonomy** **Yes** **No**

## Browse LOTUS - the natural products occurrence database Tree

- LOTUS Tree 214,481
  - Biological Tree 192,579
    - Archaea 72
    - Bacteria 17,467
    - Eukaryota 175,387
  - Chemical Tree 214,481
    - Alkaloids 28,065
    - Amino acids and Peptides 13,012
    - Carbohydrates 2,144
    - Fatty acids 10,266
    - Polyketides 27,981
    - Shikimates and Phenylpropanoids 42,177
    - Terpenoids 97,499
    - Not classified 7,328

The LOTUS Initiative

Recent uploads

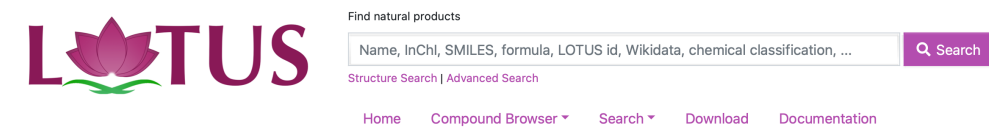
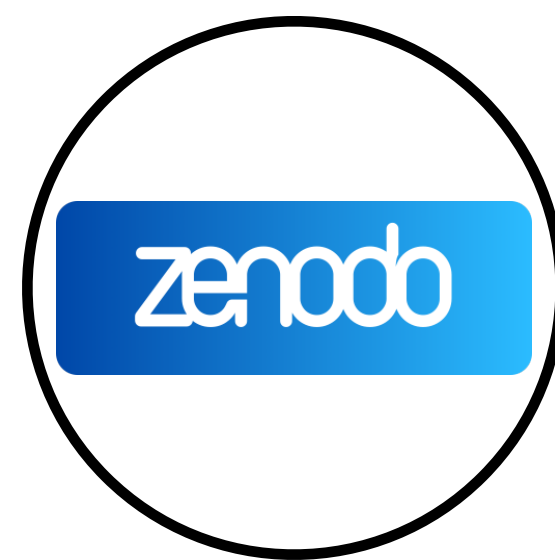
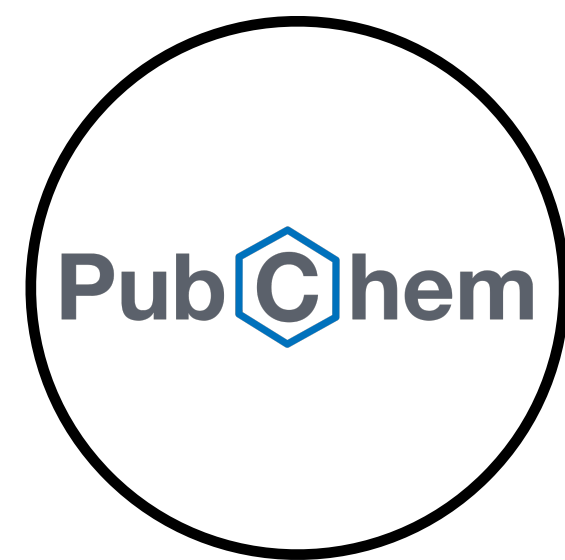
- The LOTUS Initiative for Open Natural Products Research: TMAP**  
 Rutz, Adriano, Gaudry, Arnaud  
 TMAP of the compounds present on Wikidata curated in the frame of the LOTUS initiative.  
<https://doi.org/10.7554/eLife.70780>  
 Uploaded on September 16, 2022  
 3 more version(s) exist for this record
- The LOTUS Initiative for Open Natural Products Research: frozen dataset**  
 Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie  
 Dataset used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
 Uploaded on September 16, 2022  
 5 more version(s) exist for this record
- The LOTUS Initiative for Open Natural Products Research: waste to recycle**  
 Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie  
 Dataset not uploaded to Wikidata. Generated in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
 Shared for further curation.  
 Uploaded on September 16, 2022  
 4 more version(s) exist for this record
- The LOTUS Initiative for Open Natural Products Research: biological and chemical trees**  
 Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie  
 Biological and chemical trees made from frozen metadata (10.5281/zenodo.5794106) (for example, for PubChem)  
 Uploaded on September 16, 2022  
 6 more version(s) exist for this record
- The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata)**  
 Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie  
 Dataset present on Wikidata used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
 Uploaded on September 16, 2022  
 8 more version(s) exist for this record
- The LOTUS Initiative for Open Natural Products Research: metadata**  
 Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie  
 Metadata of each of the three objects (structures, organisms, references) used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
 Uploaded on September 16, 2022  
 3 more version(s) exist for this record

Want your upload to appear in this community?

- Click the button above to upload a record directly to this community.
- To add one of your existing records to the community, edit the record, add this community under the "Communities" section, save, and finally publish.
- The community curator will then be notified to either accept or reject your upload (see community curation policy below).
- If your upload is rejected by the curator, it will still be available on Zenodo, just not in this community.



# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

1 2 3 4 5 6 7 8 9 10 ... 11521

<p><b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5H-furan-2-one</p> <p>Mol. formula C<sub>5</sub>H<sub>6</sub>O<sub>5</sub> Mol. weight 146.1 Temp. LOTUS id L7S0249032</p>	<p><b>Q27102265</b> Lysopine</p> <p>Mol. formula C<sub>20</sub>H<sub>35</sub>N<sub>2</sub>O<sub>4</sub> Mol. weight 357.54 Temp. LOTUS id L7S0160430</p>	<p><b>Q105387202</b> (1<i>r</i>,8<i>s</i>,9<i>r</i>,10<i>s</i>,12<i>r</i>)-9,10,12-trimethyl-9-[2-(5-oxo-2H-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0<sup>1,2</sup>]dodec-4-en-3-one</p> <p>Mol. formula C<sub>20</sub>H<sub>26</sub>O<sub>4</sub> Mol. weight 330.42 Temp. LOTUS id L7S0145658</p>	<p><b>Q105387201</b> (1<i>r</i>,2<i>s</i>,7<i>s</i>,10<i>r</i>,11<i>s</i>,14<i>s</i>,15<i>r</i>,16<i>s</i>,17<i>s</i>,18<i>s</i>,20<i>s</i>,23<i>s</i>)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>2,11</sup>.0<sup>3,12</sup>.0<sup>4,9</sup>.0<sup>5,8</sup>]tridecane-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 455.67 Temp. LOTUS id L7S0185716</p>
<p><b>Q105387200</b> 7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[12.10.0.0<sup>1,10</sup>.0<sup>2,11</sup>.0<sup>3,12</sup>.0<sup>4,9</sup>.0<sup>5,8</sup>]tridecane-4-en-18-yl acetate</p> <p>Mol. formula C<sub>29</sub>H<sub>45</sub>N<sub>3</sub>O<sub>3</sub> Mol. weight 455.67 Temp. LOTUS id L7S0157374</p>	<p><b>L7S0044532</b> (4<i>z</i>,8<i>e</i>)-4,7,7-trimethyl-11-methylidenecycloundeca-4,8-dien-1-one</p> <p>Mol. formula C<sub>19</sub>H<sub>28</sub>O Mol. weight 272.44 Temp. LOTUS id L7S0044532</p>	<p><b>Q105387197</b> 6-(4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>2,11</sup>.0<sup>3,12</sup>.0<sup>4,9</sup>]-15-yl)-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L7S0237245</p>	<p><b>Q105387196</b> (6<i>r</i>)-6-[(1<i>s</i>,3<i>s</i>,4<i>s</i>,6<i>s</i>,8<i>s</i>,11<i>s</i>,12<i>s</i>,15<i>r</i>,16<i>r</i>)-4,6-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0<sup>1,10</sup>.0<sup>2,11</sup>.0<sup>3,12</sup>.0<sup>4,9</sup>]-15-yl]-2-methylhept-2-en-4-one</p> <p>Mol. formula C<sub>30</sub>H<sub>48</sub>O<sub>3</sub> Mol. weight 456.7 Temp. LOTUS id L7S0074160</p>

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records (phenylpropanoates, or Gene Ontology: DNA repair). [More...](#)

Select classification: **LOTUS - the natural products occurrence database** Search

Classification description (from LOTUS - the natural products occurrence database): Biological and chemical tree provided by the LOTUS (natural products occurrence database) Created on 09/10/2022 09:05:48 [More...](#)

Data type counts to display: **None** **Compound** **Taxonomy** Display zero count nodes? **Yes** **No**

## Browse LOTUS - the natural products occurrence database Tree

- LOTUS Tree **214,481**
  - Biological Tree **192,579**
    - Archaea **72**
    - Bacteria **17,467**
    - Eukaryota **175,387**
  - Chemical Tree **214,481**
    - Alkaloids **28,065**
    - Amino acids and Peptides **13,012**
    - Carbohydrates **2,144**
    - Fatty acids **10,266**
    - Polyketides **27,981**
    - Shikimates and Phenylpropanoids **42,177**
    - Terpenoids **97,499**
    - Not classified **7,328**

zenodo The LOTUS Initiative

Recent uploads

- September 16, 2022 (v4) **Figure** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: TMAP  
Rutz, Adriano, Gaudry, Arnaud.  
TMAP of the compounds present on Wikidata curated in the frame of the LOTUS initiative: <https://doi.org/10.7554/eLife.70780>  
Uploaded on September 16, 2022  
3 more version(s) exist for this record
- September 16, 2022 (v6) **Dataset** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: frozen dataset  
Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie.  
Dataset used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
Uploaded on September 16, 2022  
5 more version(s) exist for this record
- September 16, 2022 (v5) **Dataset** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: waste to recycle  
Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie.  
Dataset not uploaded to Wikidata. Generated in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
Shared for further curation.  
Uploaded on September 16, 2022  
4 more version(s) exist for this record
- September 16, 2022 (v7) **Dataset** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: biological and chemical trees  
Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie.  
Biological and chemical trees made from frozen metadata (10.5281/zenodo.5794106) (for example, for PubChem)  
Uploaded on September 16, 2022  
4 more version(s) exist for this record
- September 16, 2022 (v9) **Dataset** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata)  
Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie.  
Dataset present on Wikidata used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
Uploaded on September 16, 2022  
8 more version(s) exist for this record
- September 16, 2022 (v8) **Dataset** **Open Access**  
The LOTUS Initiative for Open Natural Products Research: metadata  
Rutz, Adriano, Bission, Jonathan, Allard, Pierre-Marie.  
Metadata of each of the three subjects (structures, organisms, references) used in the frame of the LOTUS Initiative: <https://doi.org/10.7554/eLife.70780>  
Uploaded on September 16, 2022  
3 more version(s) exist for this record

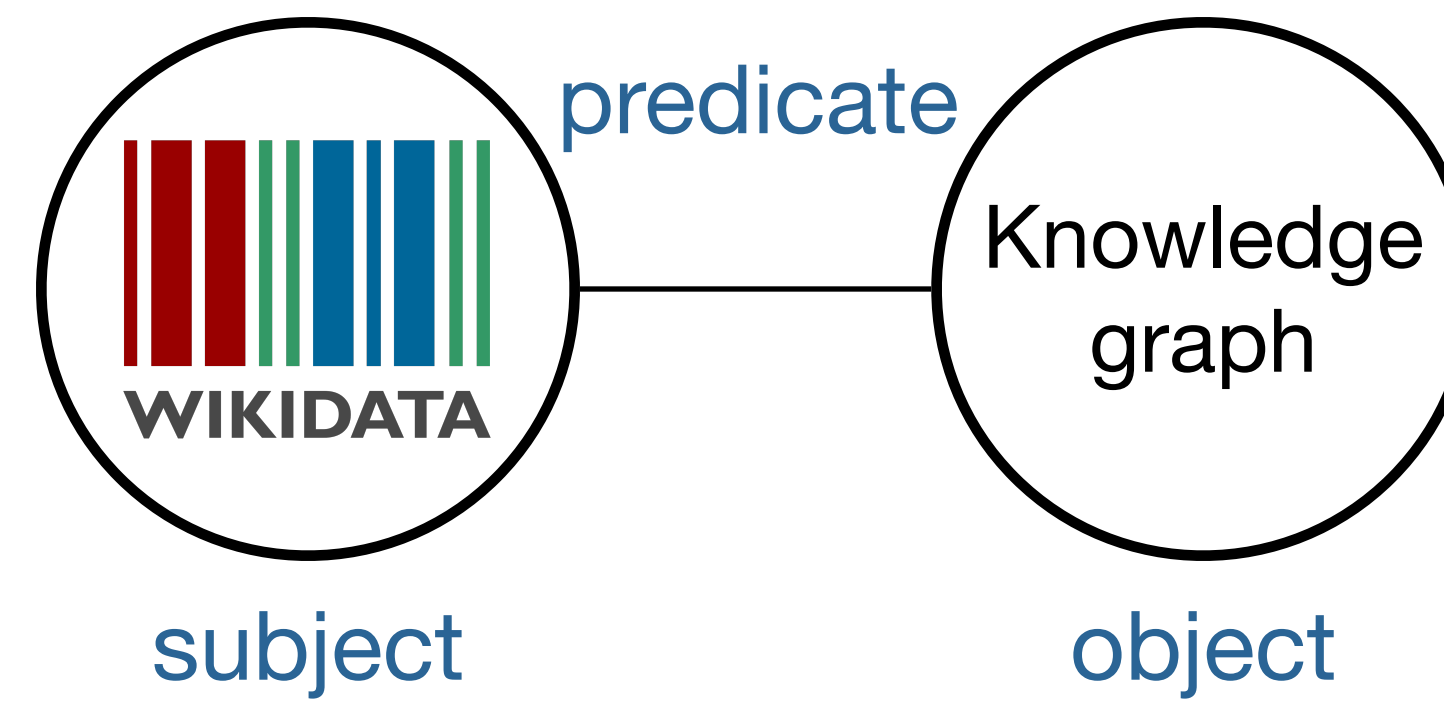
## Extraction

**SELECT** ?desired items  
**WHERE** {  
 ?VALUES ?classes {  
   Chemical compound  
 }  
 ?item instance of ?classes.  
 }  
 ?item found in taxon ?stmt.  
 ?stmt found in taxon ?taxon.  
**OPTIONAL** {  
   ?stmt occurrence ?ref.  
   ?ref stated in ?art.  
 }  
**SERVICE** wikibase:label {  
 language  
 "[AUTO\_LANGUAGE],en". }  
 }

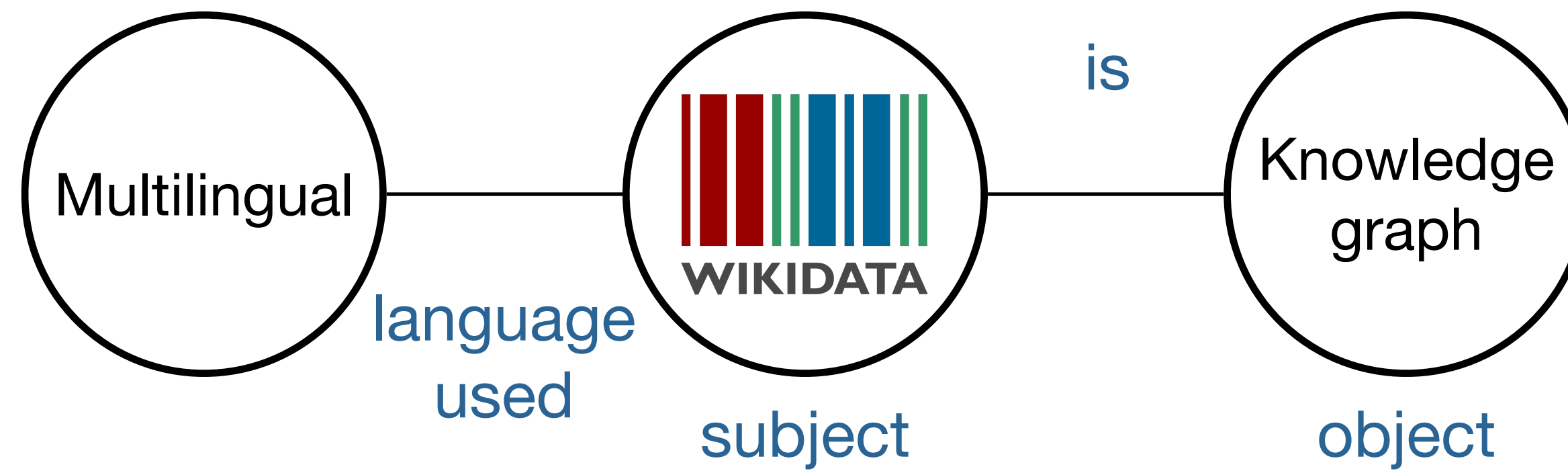
# The initiative - Wikidata



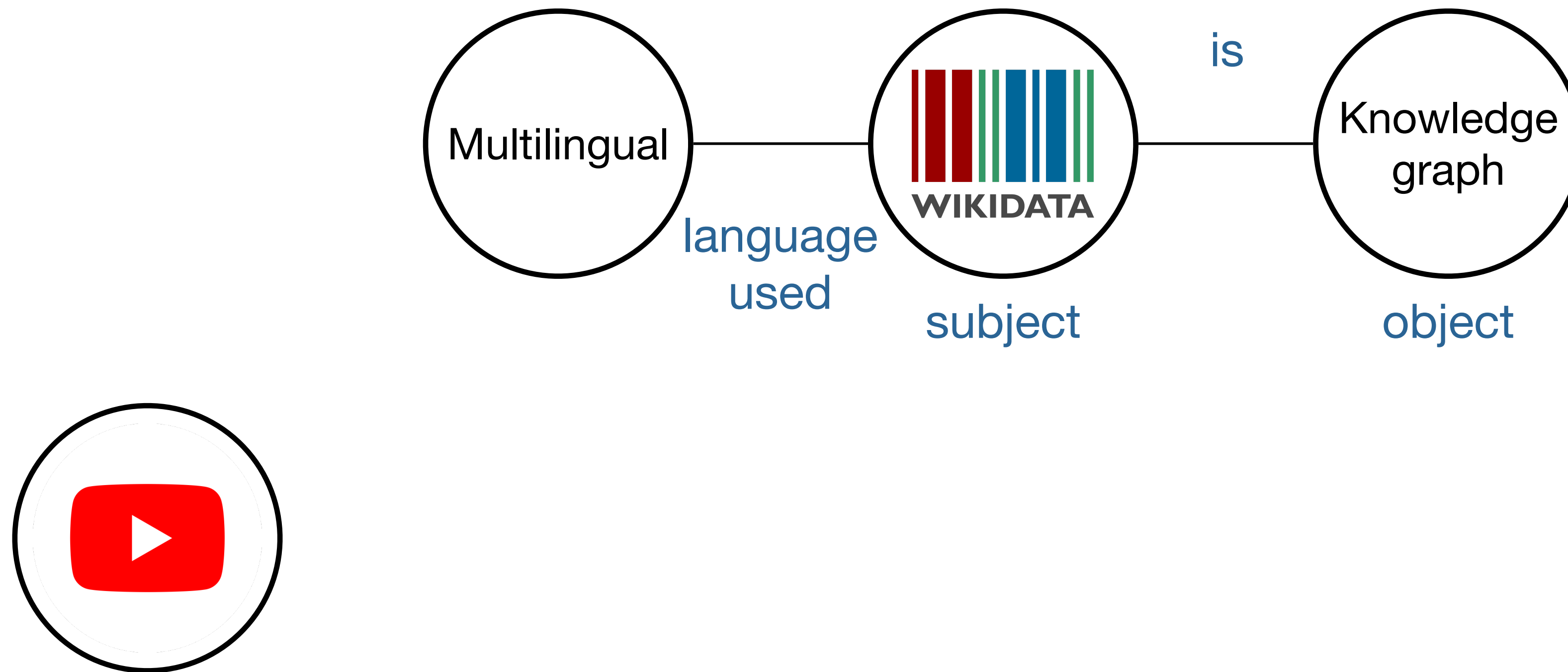
# The initiative - Wikidata



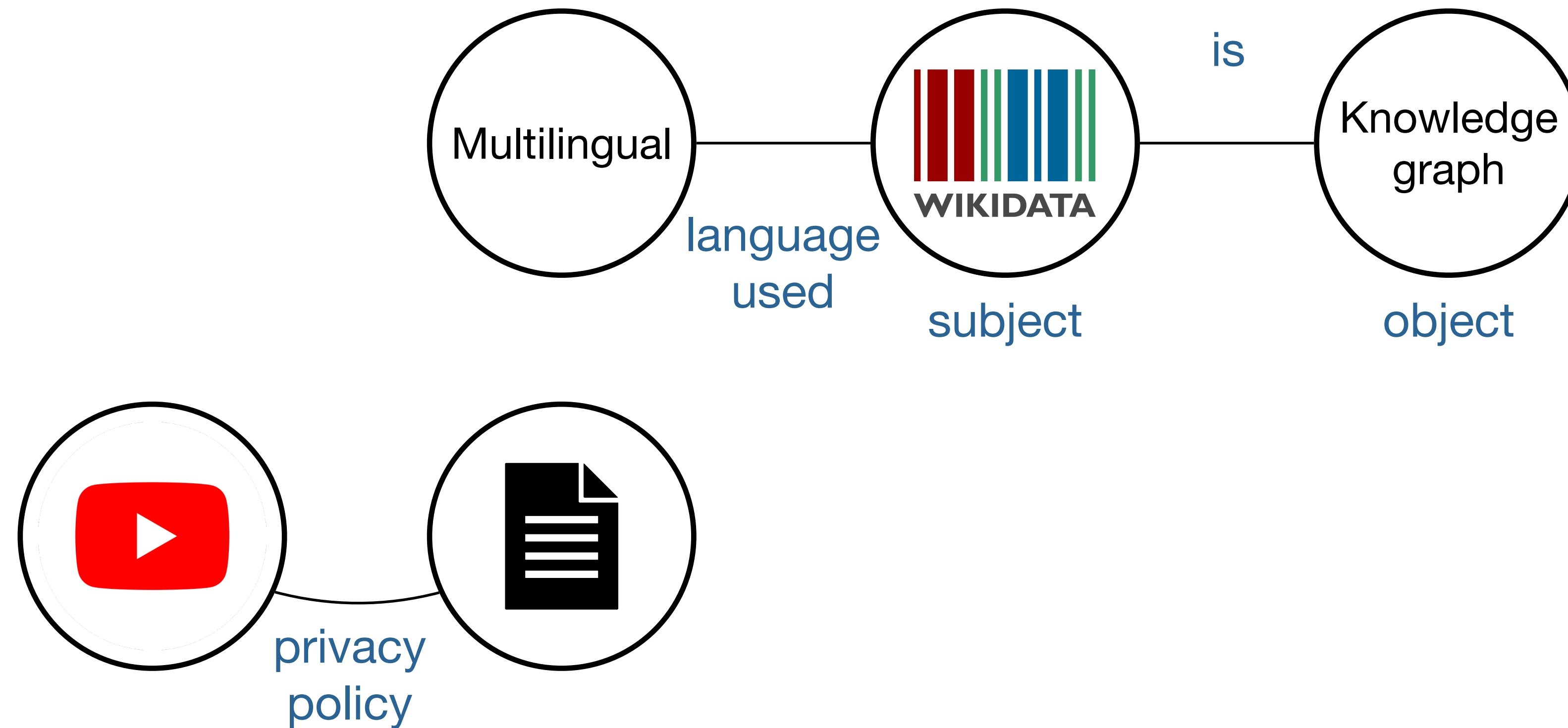
# The initiative - Wikidata



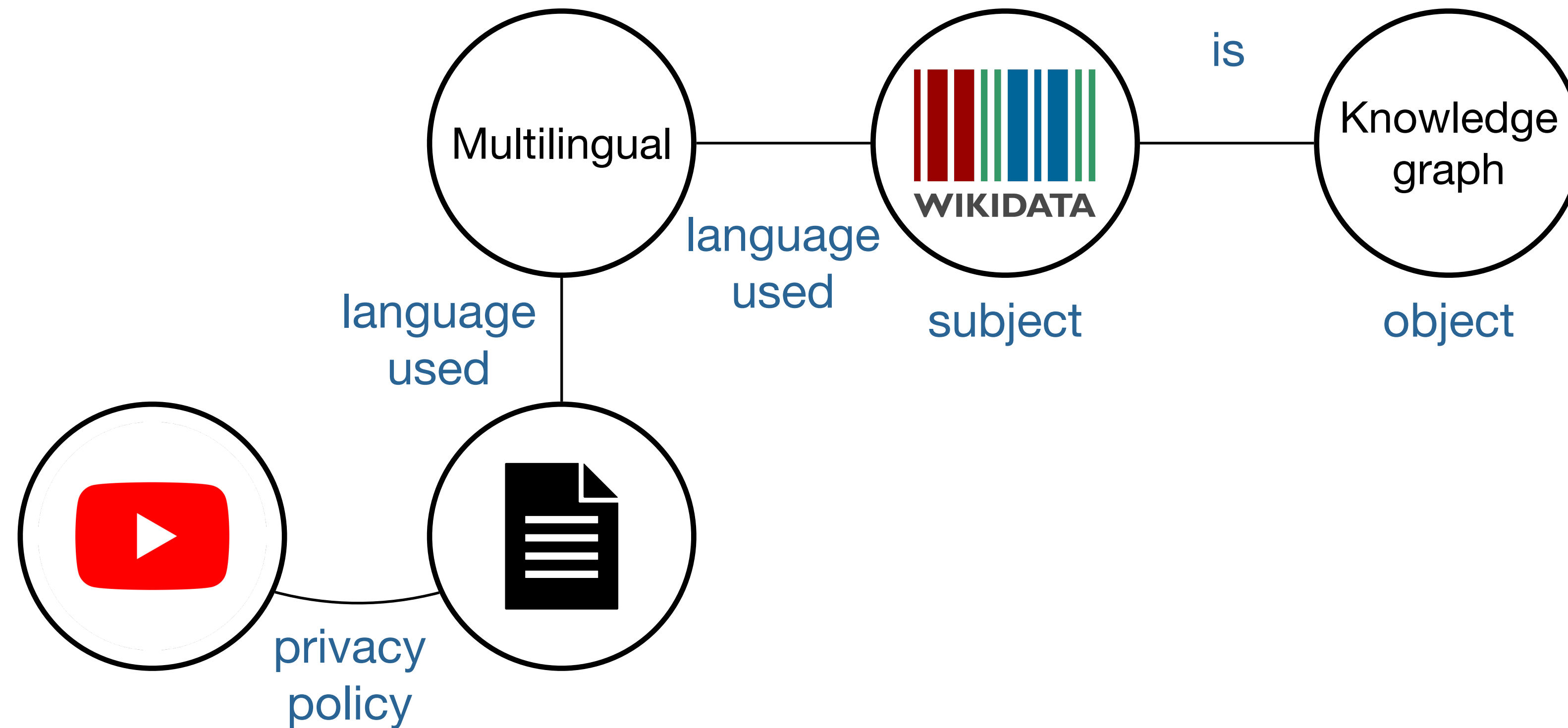
# The initiative - Wikidata



# The initiative - Wikidata



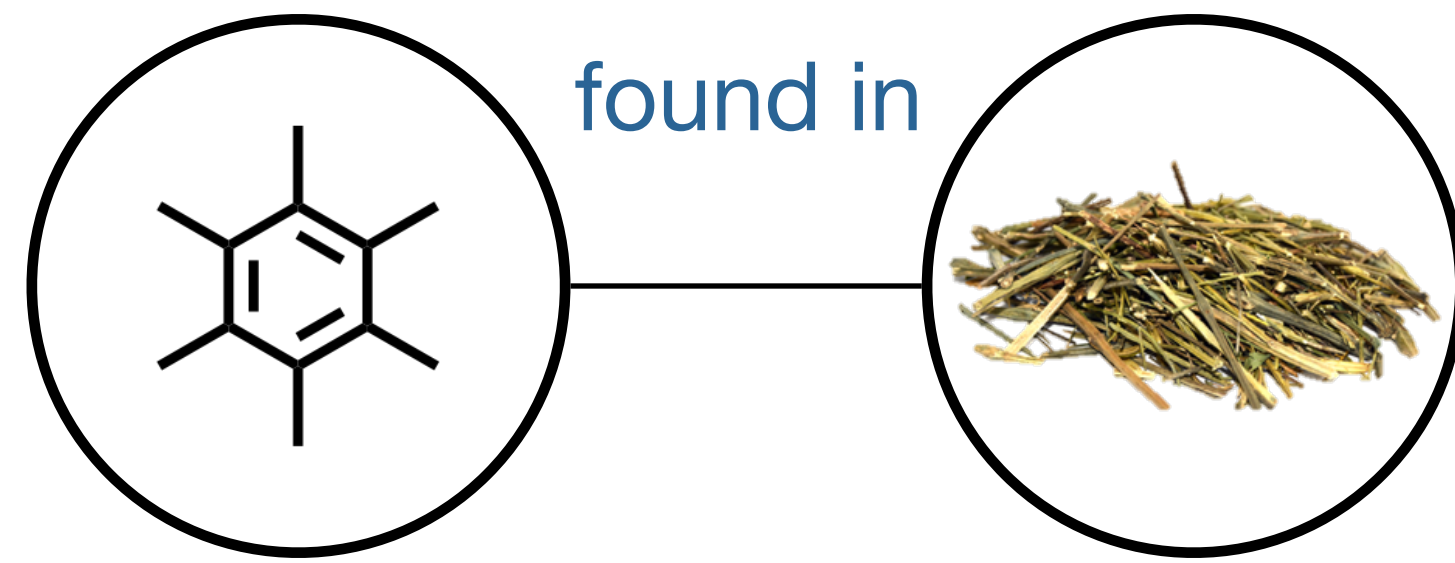
# The initiative - Wikidata





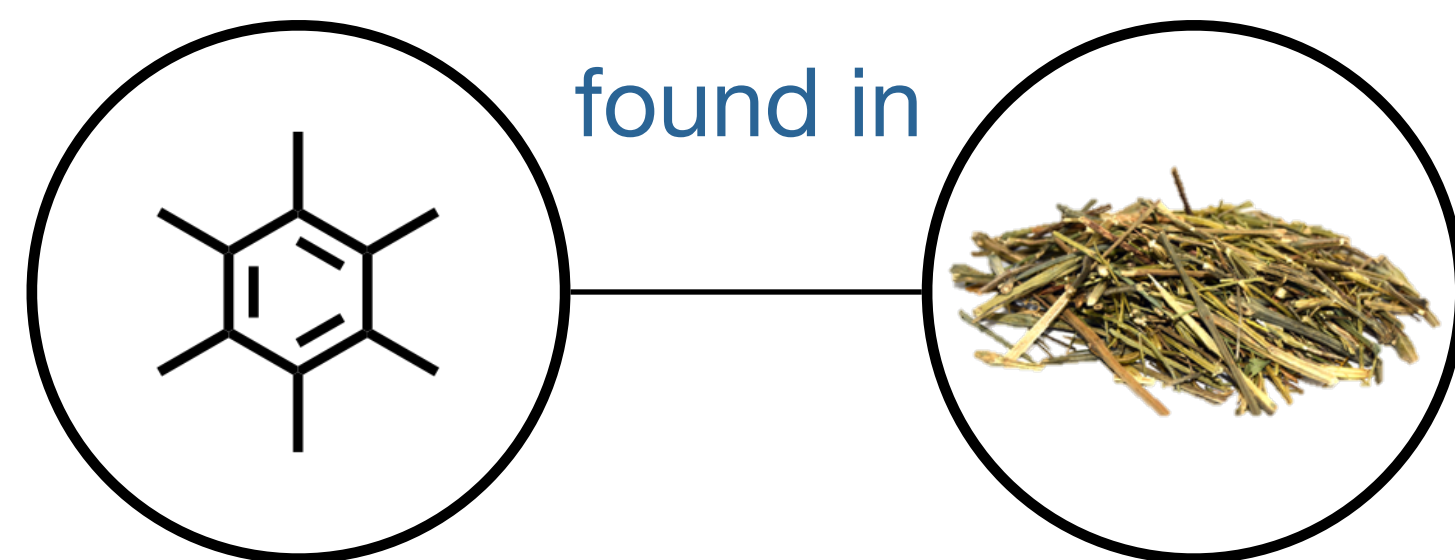
# The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Swertia chirayita*? »



# The initiative - SPARQL

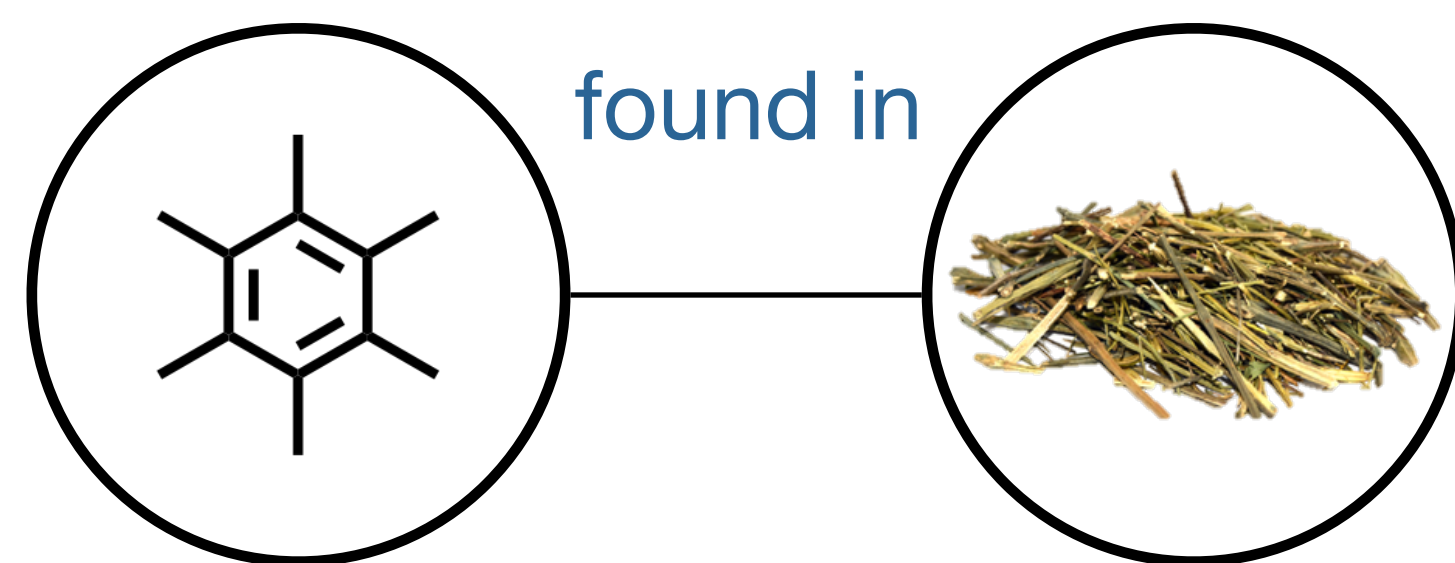
« Hey Wiki, what are the compounds found in *Swertia chirayita*? »



```
SELECT DISTINCT ?structure ?structure_smiles ?  
structure_inchikey  
WHERE {  
  VALUES ?organism {  
    wd:Q21318003 # Swertia chirayita  
  }  
  ?organism_child (wdt:P171★) ?organism.  
  ?structure wdt:P233 ?structure_smiles;  
  wdt:P235 ?structure_inchikey;  
  (p:P703/ps:P703) ?organism_child.  
}
```

# The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Swertia chirayita*? »



```
SELECT DISTINCT ?structure ?structure_smiles ?
structure_inchikey
WHERE {
  VALUES ?organism {
    wd:Q21318003 # Swertia chirayita
  }
  ?organism_child (wdt:P171★) ?organism.
  ?structure wdt:P233 ?structure_smiles;
  wdt:P235 ?structure_inchikey;
  (p:P703/ps:P703) ?organism_child.
}
```

42 results in 0.4 second

row	structure	structure_smile s	structure_inchikey
1	<a href="#">wd:Q105165035</a>		MIJYXULNPSFWEK-ZZAAMMQTSA-N
...	...	...	...
42	<a href="#">wd:Q1074417</a>		AEDDIBAIWPIIBD-ZJKJAXBQSA-N

# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

*« Hey Wiki, what are the compounds found in Gentianales? »*

10,620 results in 3.3 seconds

# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

*« Hey Wiki, what are the compounds found in Gentianales? »*

10,620 results in 3.3 seconds

*« Sorry Wiki, I forgot, how many species belong to Gentianales? »*

43,611 results in 3.5 seconds

# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

*« Sorry Wiki, I forgot, how many species belong to Gentianales? »*

43,611 results in 3.5 seconds

*« Hey Wiki, what are the compounds found in Gentianales? »*

10,620 results in 3.3 seconds

*« Ok Wiki, among the compounds found in the Gentianales how many were already described as bitter? »*

112 results in 2.7 seconds



# The initiative - SPARQL

*« Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in Swertia chirayita, and how many? »*

# The initiative - SPARQL

*« Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in Swertia chirayita, and how many? »*

```
PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#>
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT
  ?taxon
  ?taxon_name
  (COUNT(DISTINCT ?compound) AS ?count)
WHERE {
  SERVICE idsm:wikidata {
    SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {
      VALUES ?organism {
        wd:Q21318003 # Swertia chirayita
      }
      VALUES ?taste {
        wd:Q1517187 # Bitterness
      }
      ?organism_child (wdt:P171*) ?organism.
      ?structure wdt:P233 ?structure_smiles;
        wdt:P235 ?structure_inchikey;
        wdt:P1552 ?taste;
        (p:P703/ps:P703) ?organism_child.
    }
    ?compound sachem:similarCompoundSearch _:b40.
    _:b40 sachem:query ?structure_smiles;
      sachem:cutoff "0.9"^^xsd:double.
  }
  hint:Prior hint:runFirst "true"^^xsd:boolean.
  ?compound wdt:P703 ?taxon.
  ?taxon wdt:P225 ?taxon_name.
}
GROUP BY ?taxon ?taxon_name
ORDER BY DESC (?count)
```

# The initiative - SPARQL

« *Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in Swertia chirayita, and how many?»*

14 results in 2.1 seconds

row	taxon	taxon_name	count
1	<a href="#">wd:Q162579</a>	<i>Gentiana purpurea</i>	3
2	<a href="#">wd:Q11255805</a>	<i>Swertia japonica</i>	3
...	...	...	...
5	<a href="#">wd:Q158572</a>	<i>Gentiana lutea</i>	2
6	<a href="#">wd:Q13859874</a>	<i>Gentianella nitida</i>	2
...	...	...	...
14	<a href="#">wd:Q1074417</a>	<i>Swertia mileensis</i>	1


```

PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#>
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT
  ?taxon
  ?taxon_name
  (COUNT(DISTINCT ?compound) AS ?count)
WHERE {
  SERVICE idsm:wikidata {
    SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {
      VALUES ?organism {
        wd:Q21318003 # Swertia chirayita
      }
      VALUES ?taste {
        wd:Q1517187 # Bitterness
      }
      ?organism_child (wdt:P171*) ?organism.
      ?structure wdt:P233 ?structure_smiles;
        wdt:P235 ?structure_inchikey;
        wdt:P1552 ?taste;
        (p:P703/ps:P703) ?organism_child.
    }
    ?compound sachem:similarCompoundSearch _:b40.
    _:b40 sachem:query ?structure_smiles;
      sachem:cutoff "0.9"^^xsd:double.
  }
  hint:Prior hint:runFirst "true"^^xsd:boolean.
  ?compound wdt:P703 ?taxon.
  ?taxon wdt:P225 ?taxon_name.
}
GROUP BY ?taxon ?taxon_name
ORDER BY DESC (?count)
    
```

# The initiative - Community curation

[English](#)
[Not logged in](#)
[Talk](#)
[Contributions](#)
[Create account](#)
[Log in](#)

[Read](#)
[View history](#)



[Main page](#)  
[Community portal](#)  
[Project chat](#)  
[Create a new Item](#)  
[Recent changes](#)  
[Random Item](#)  
[Query Service](#)  
[Nearby](#)  
[Help](#)  
[Donate](#)

[Lexicographical data](#)  
[Create a new Lexeme](#)  
[Recent changes](#)  
[Random Lexeme](#)

[Tools](#)  
[What links here](#)  
[Related changes](#)  
[Special pages](#)  
[Permanent link](#)  
[Page information](#)  
[Concept URI](#)  
[Cite this page](#)

Item
Discussion

## hyperelodione D (Q116482353)

chemical compound  [edit](#)

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

### Statements

**instance of** chemical compound

0 references

**mass** 548.797 dalton

1 reference


based on heuristic
inferred from SMILES

**chemical formula** C<sub>36</sub>H<sub>52</sub>O<sub>4</sub>

1 reference

based on heuristic
inferred from SMILES

# The initiative - Community curation



WIKIDATA

[English](#) [Not logged in](#) [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Item [Discussion](#)

[Read](#) [View history](#)

---

## hyperelodione D (Q116482353)

chemical compound [edit](#)

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

[canonical SMILES](#)

C/C(C)=C\CC/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(C)(C)O2

[reason for deprecated rank](#)
[structural reassignment](#)

1 reference

C/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(CC/C=C(C)/C)(C)O2

[reason for preferred rank](#)
[structural reassignment](#)

1 reference

[stated in](#) [Cytotoxic polyprenylated phloroglucinol derivatives from Hypericum elodeoides Choisy modulating the transactivation of RXRα](#)

[stated in](#) [Bioinspired Total Synthesis of Erectones A and B, and the Revised Structure of Hyperelodione D](#)

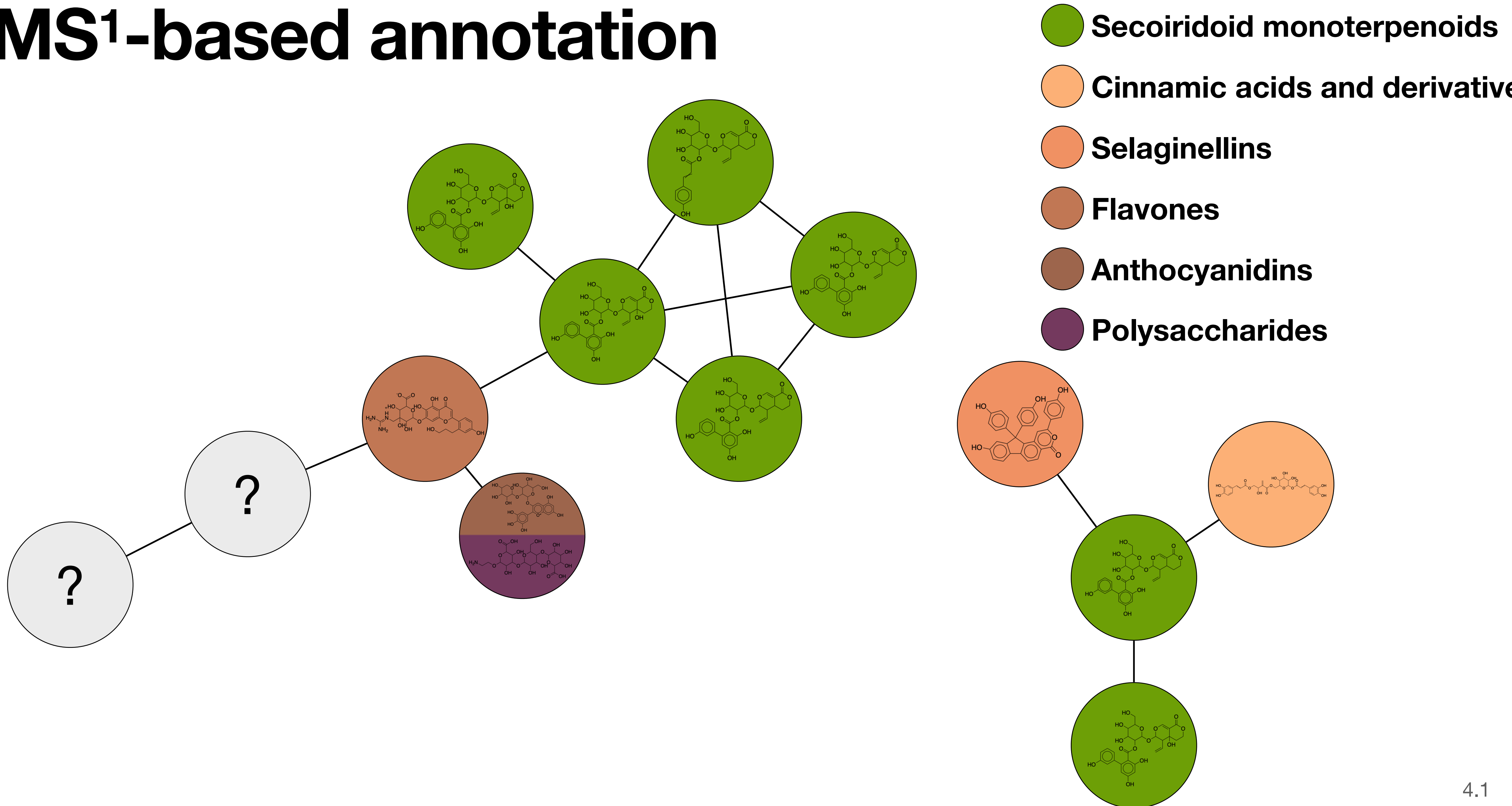
[Main page](#) [Community portal](#) [Project chat](#) [Create a new Item](#) [Recent changes](#) [Random Item](#)

# The initiative - back to metabolites


*« But Wiki, can this help me with my metabolite annotation? »*

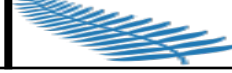


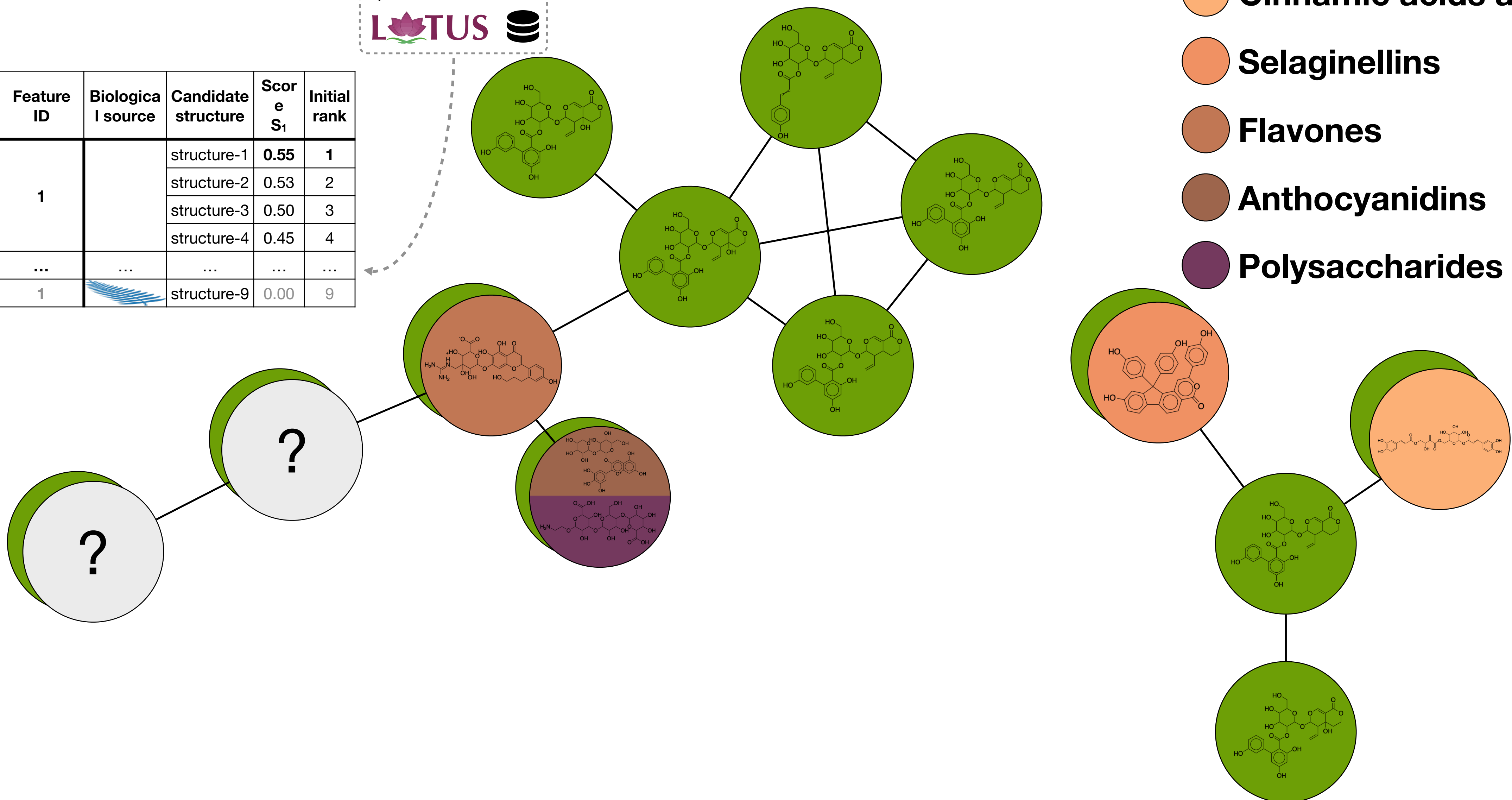
# MS<sup>1</sup>-based annotation



# MS<sup>1</sup>-based annotation

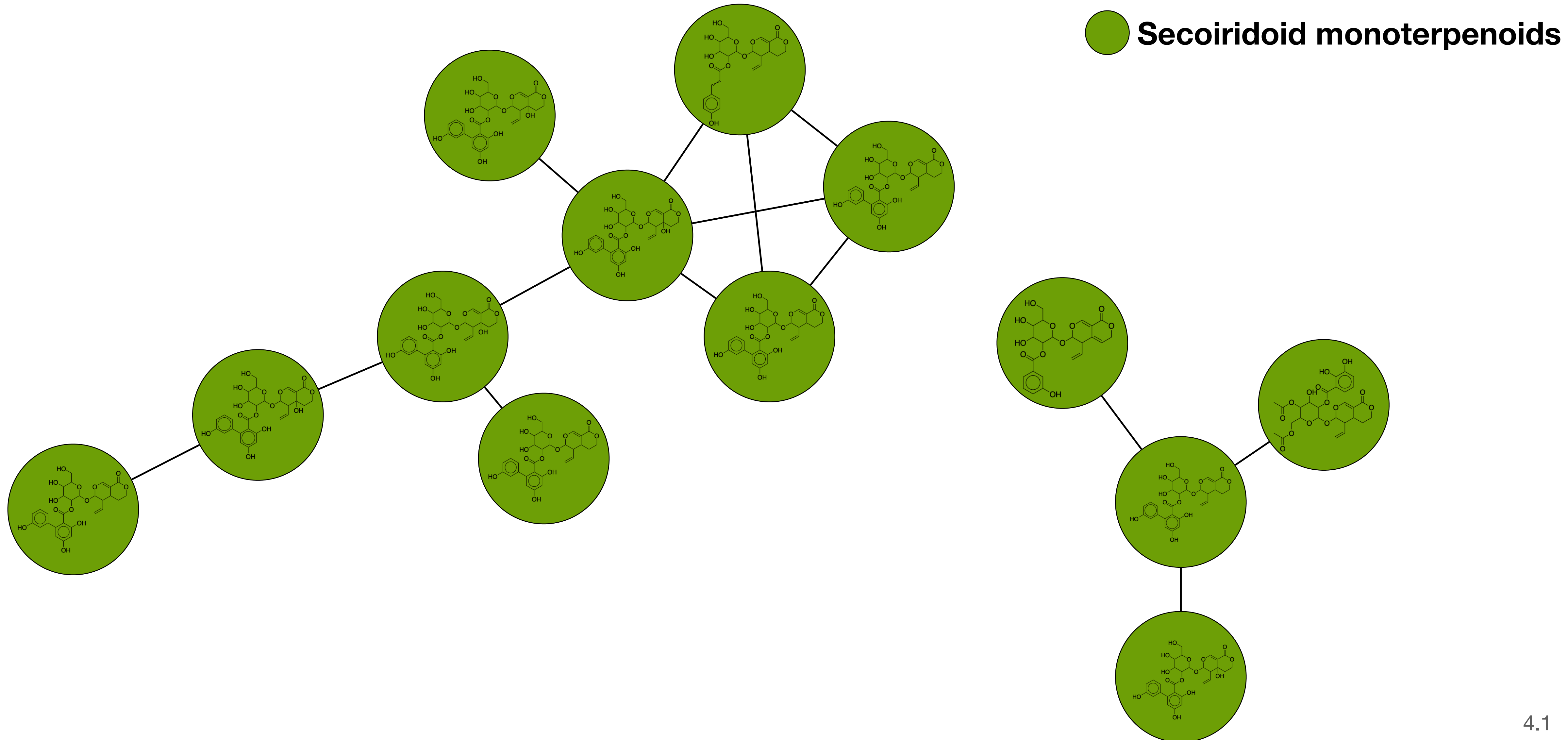
Optional MS<sup>1</sup>-based annotation  


Feature ID	Biological source	Candidate structure	Score S <sub>1</sub>	Initial rank
1		structure-1	0.55	1
		structure-2	0.53	2
		structure-3	0.50	3
		structure-4	0.45	4
...	...	...	...	...
1		structure-9	0.00	9

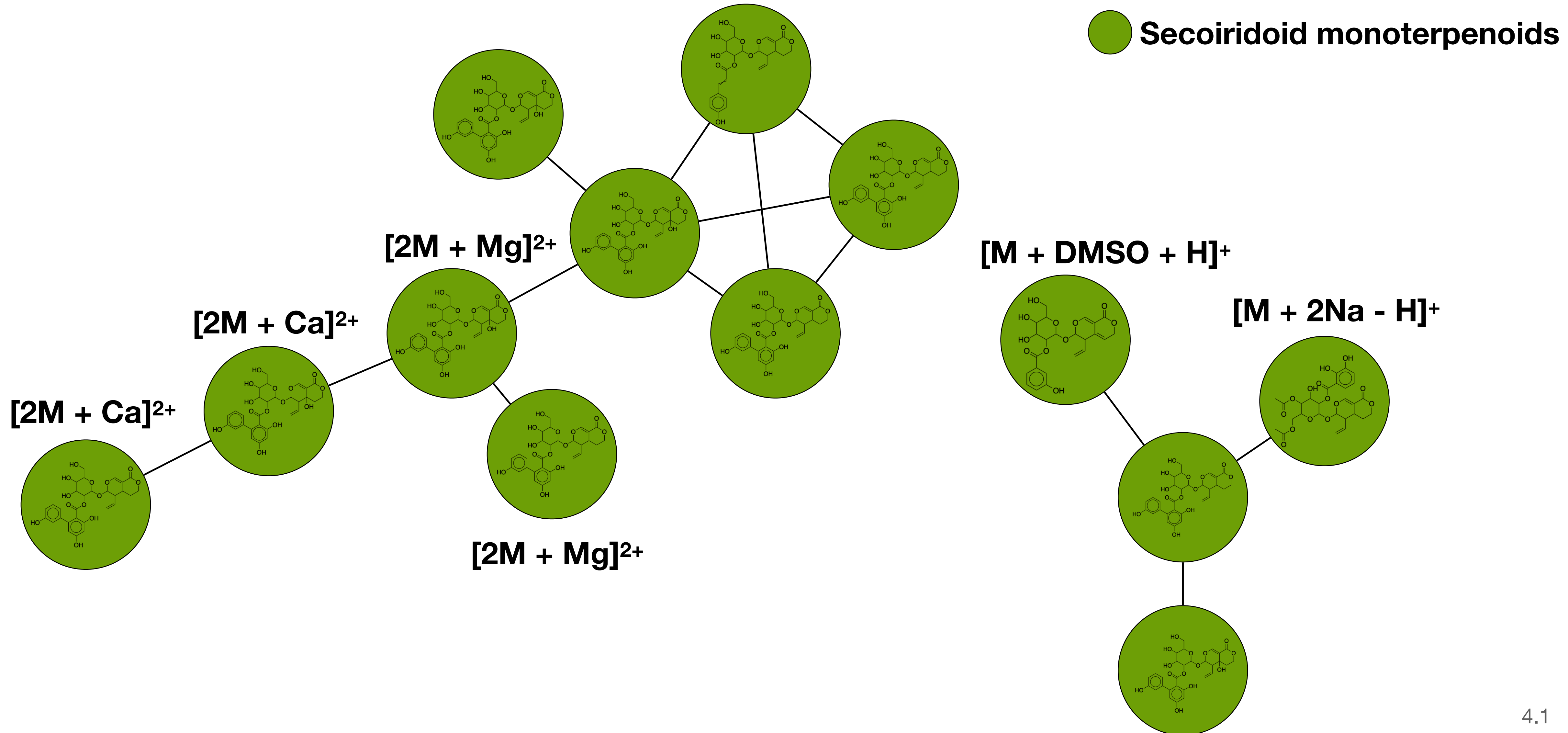


- Secoiridoid monoterpene
- Cinnamic acids and derivative
- Selaginellins
- Flavones
- Anthocyanidins
- Polysaccharides

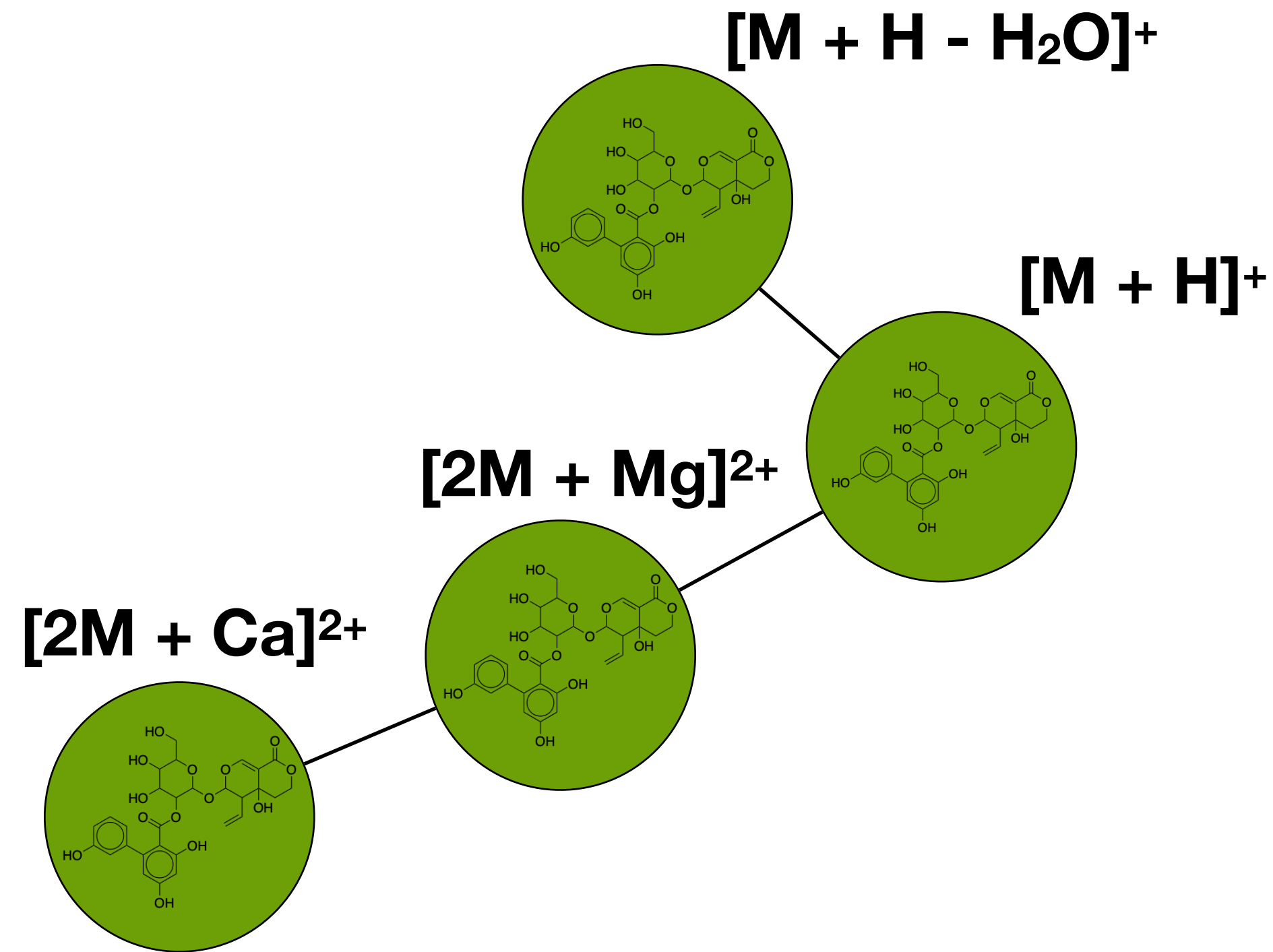
# MS<sup>1</sup>-based annotation



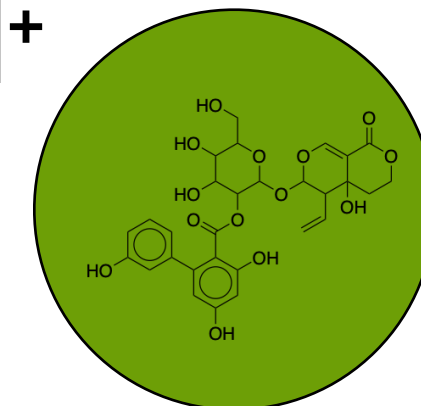
# MS<sup>1</sup>-based annotation



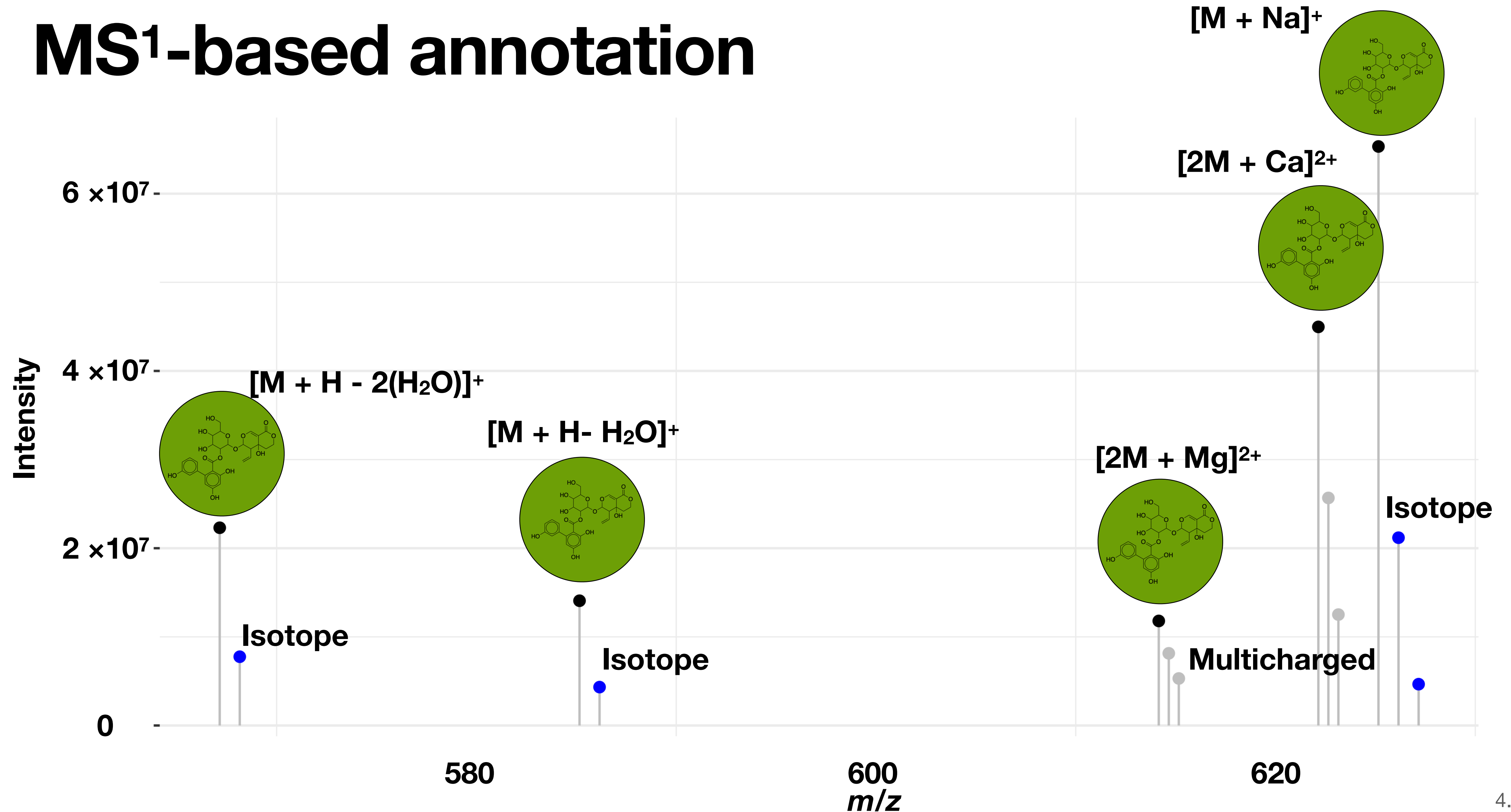
# MS<sup>1</sup>-based annotation



**[M + Na]<sup>+</sup>**

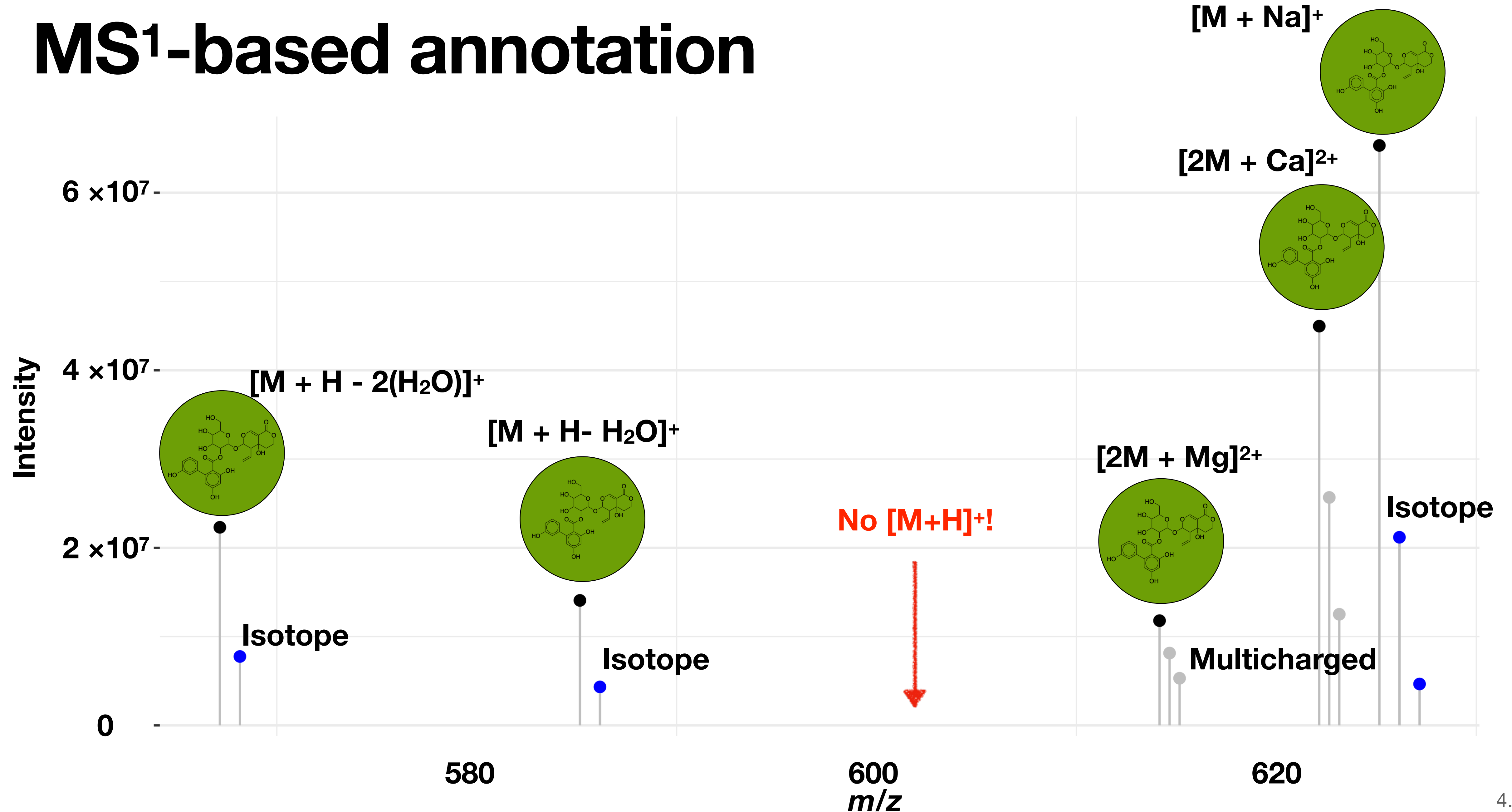


# MS<sup>1</sup>-based annotation



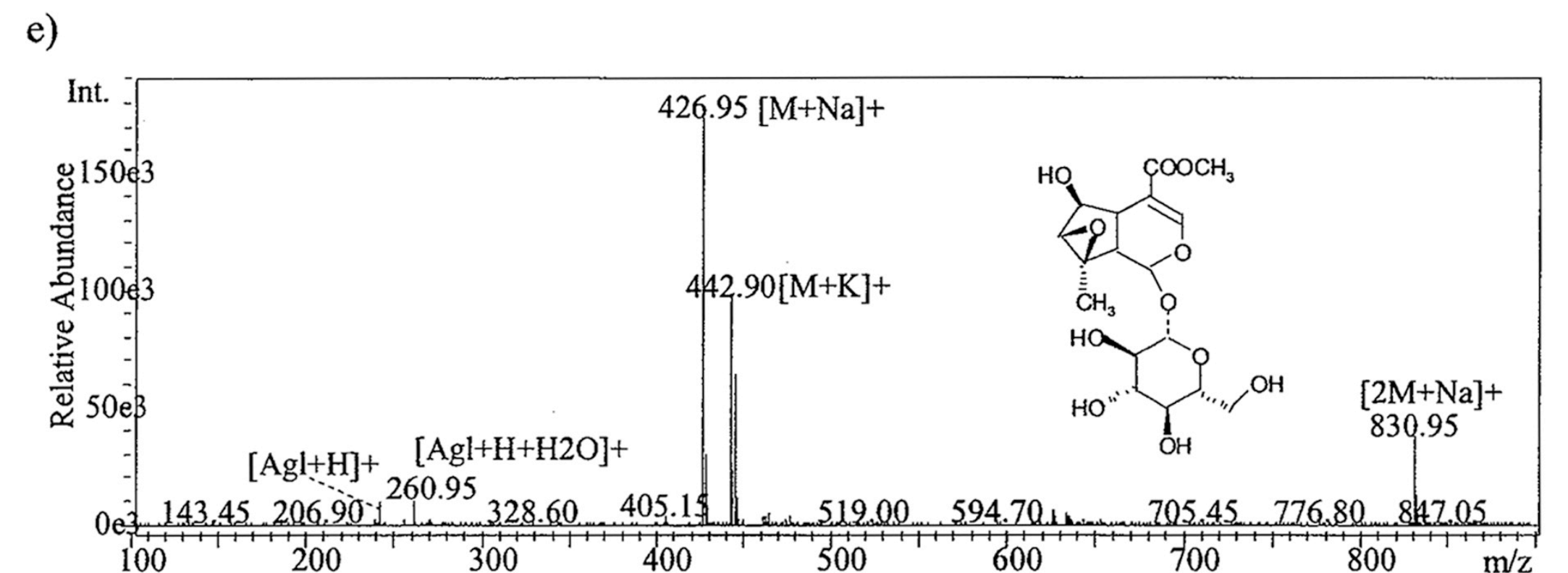
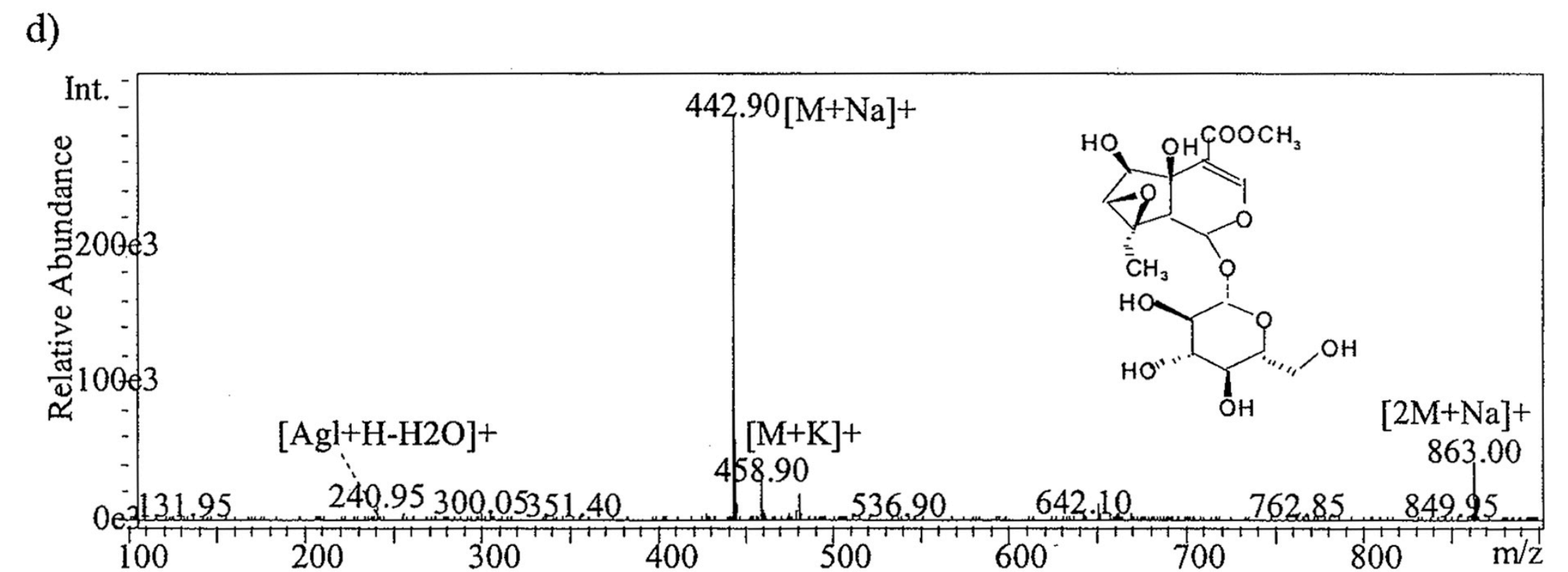
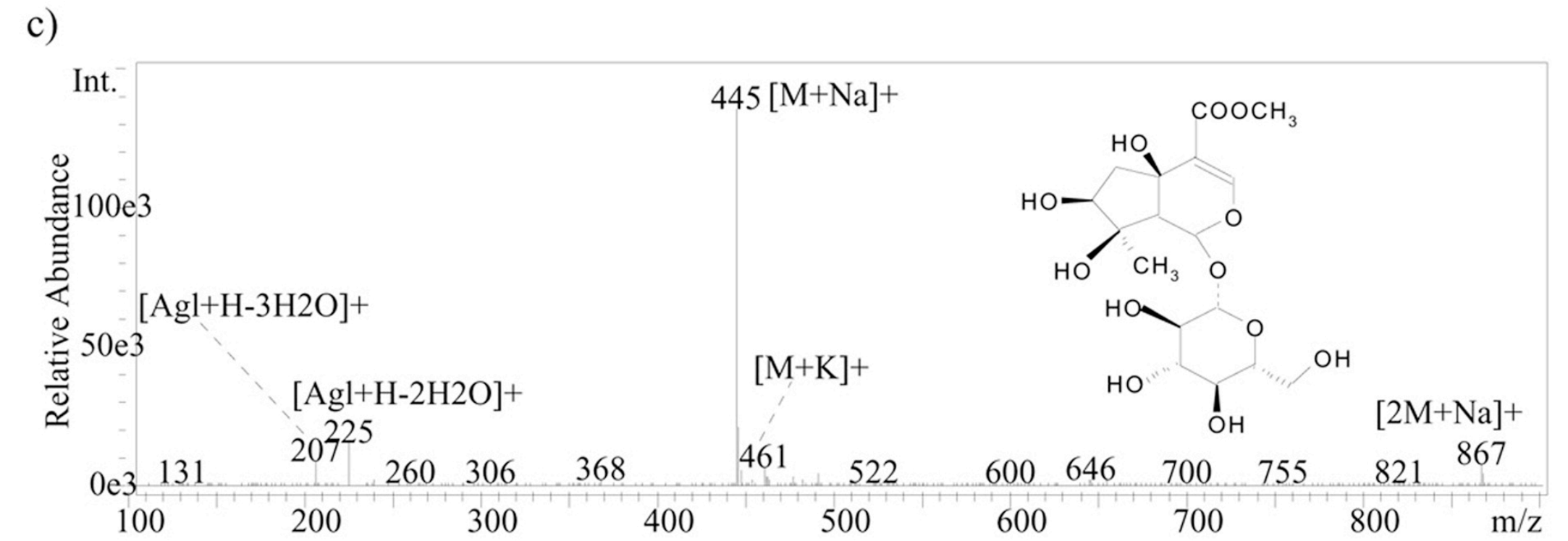
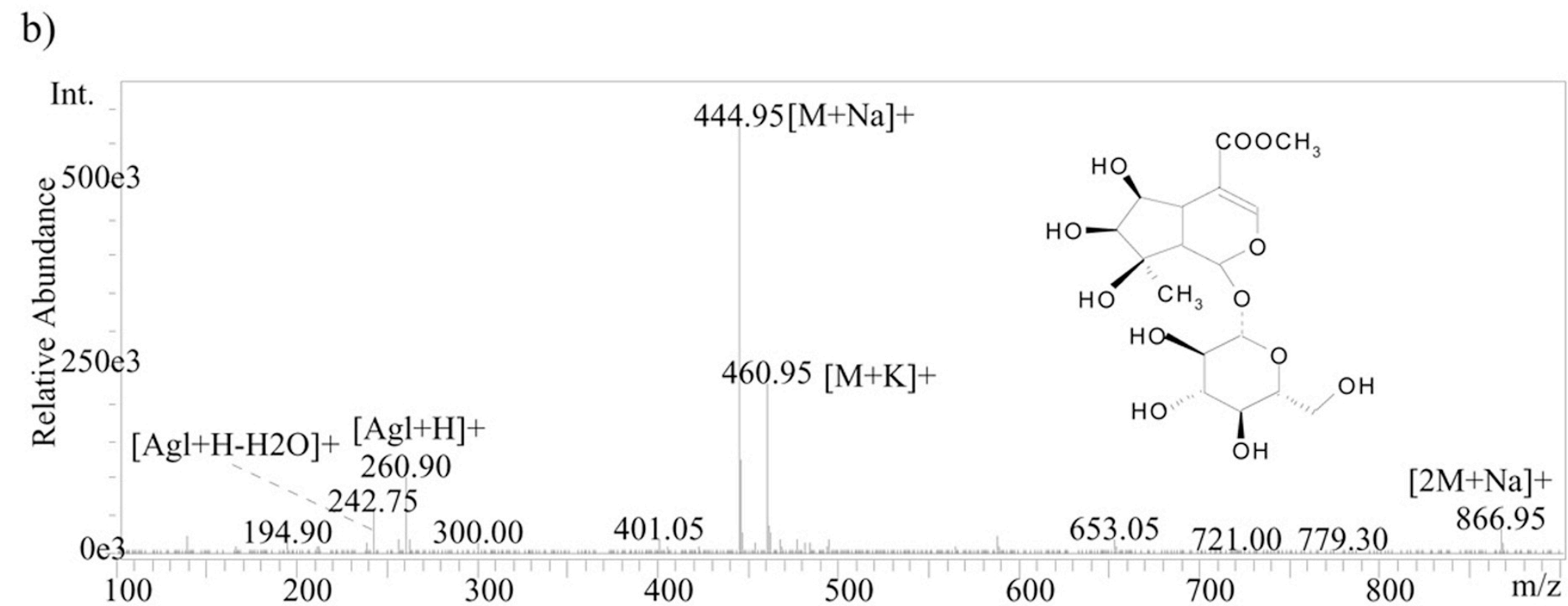
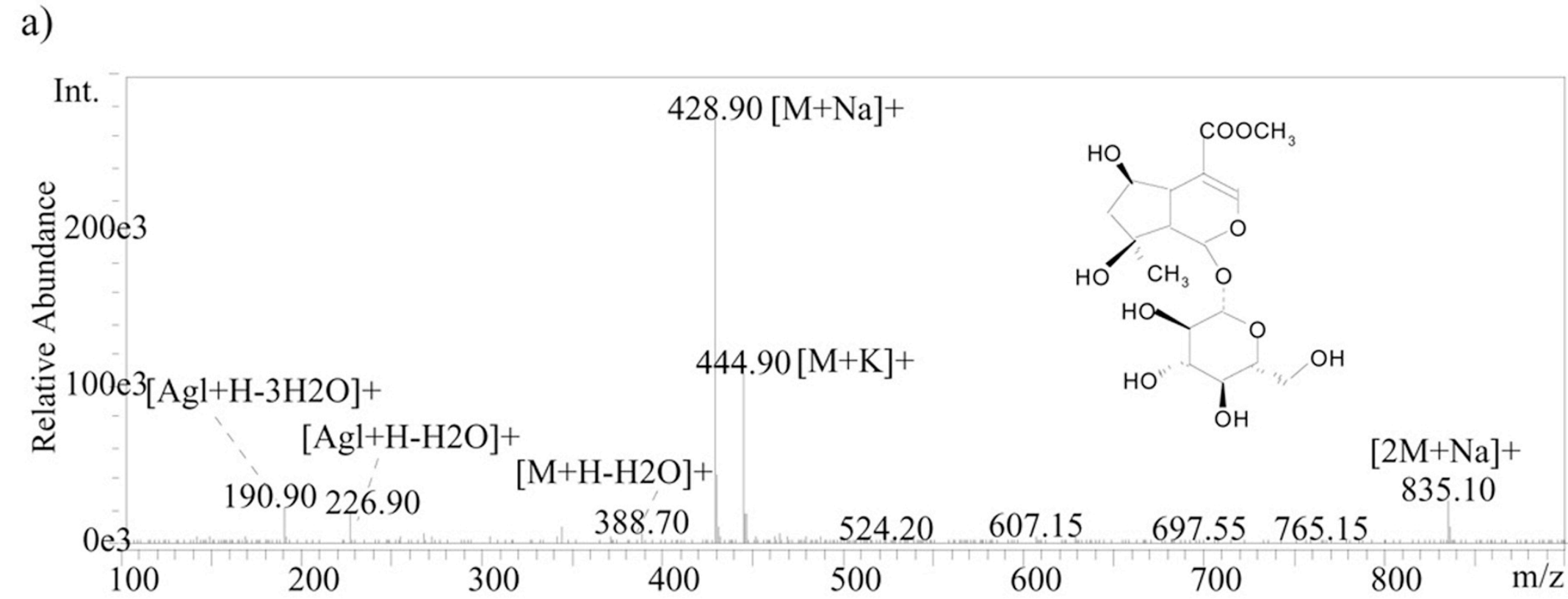


# MS<sup>1</sup>-based annotation



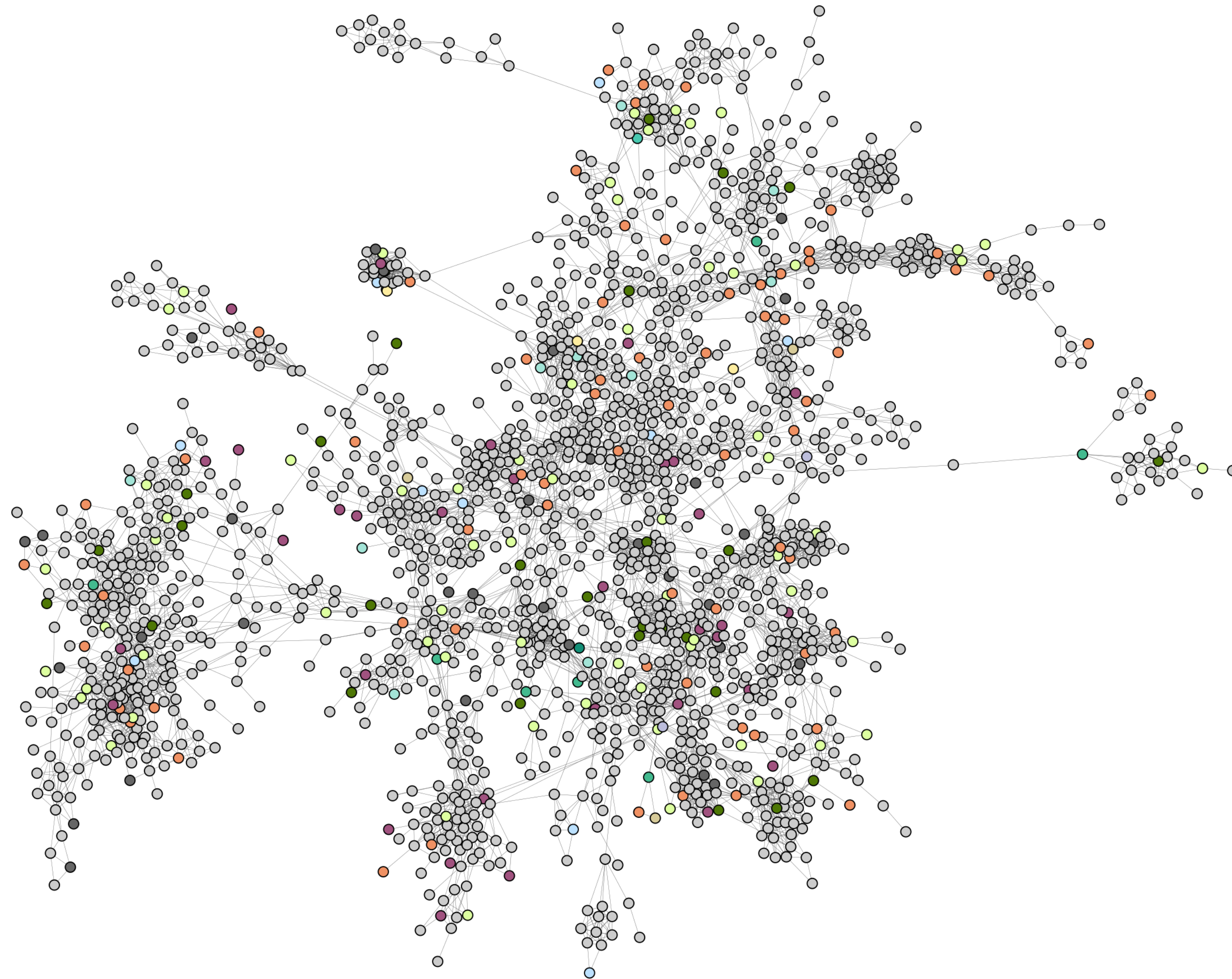


# MS<sup>1</sup>-based annotation



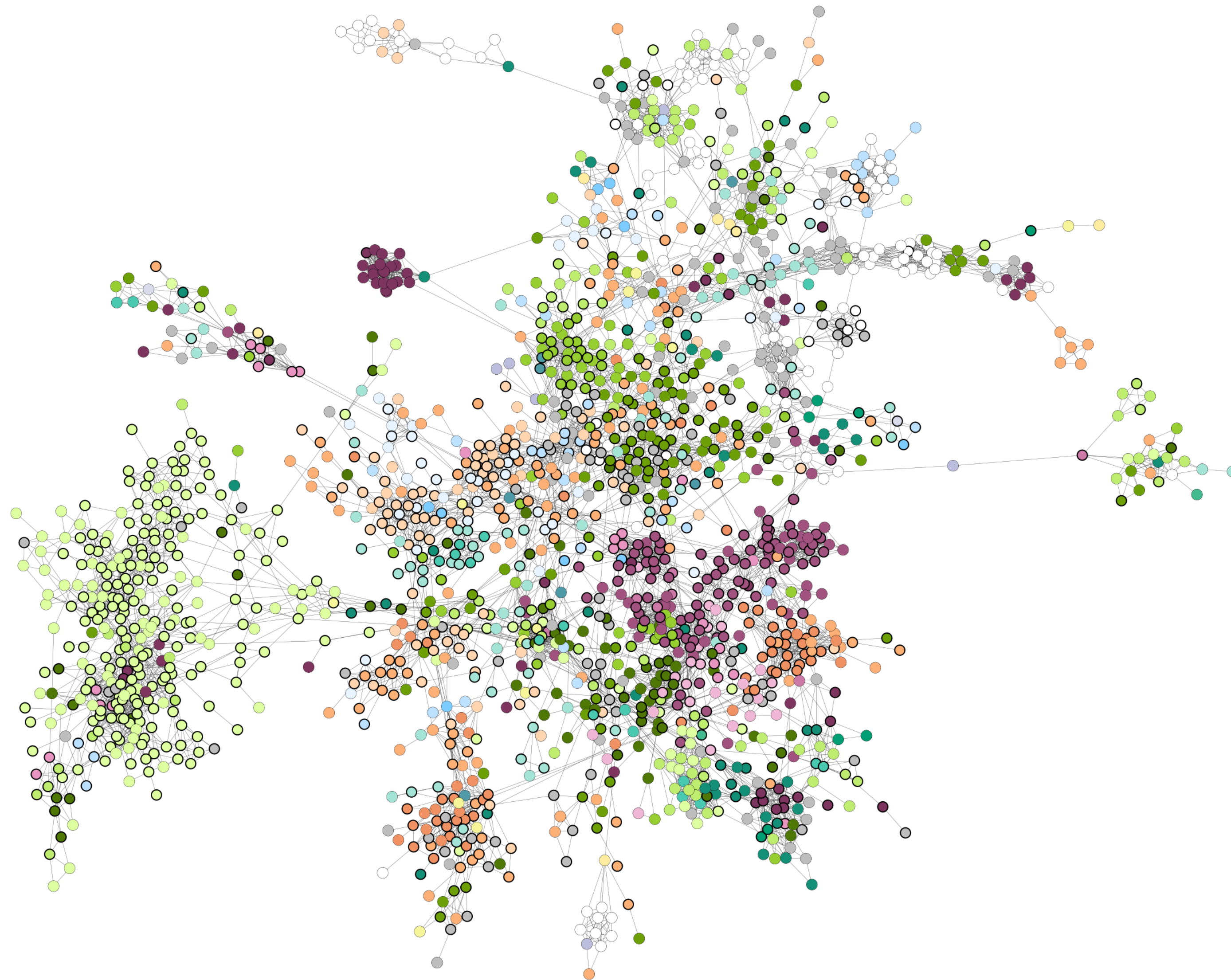
<https://doi.org/10.1080/10739140500311288>

# Previous annotation





# Improved annotation



# Benchmark

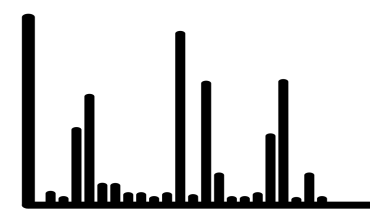
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>

public and third parties libraries

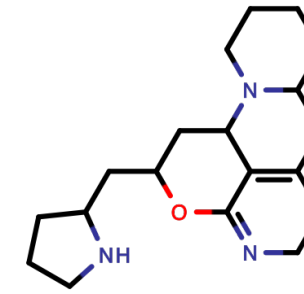
66,646 experimental spectra

at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

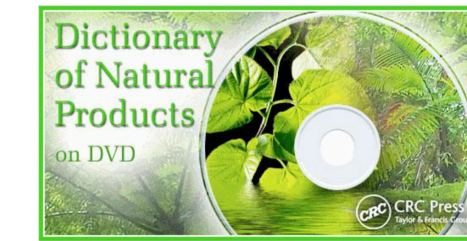
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



“Alkaloid from *Brunfelsia hopeana*”



[http://  
globalnames.org](http://globalnames.org)

text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

2107 unique entries



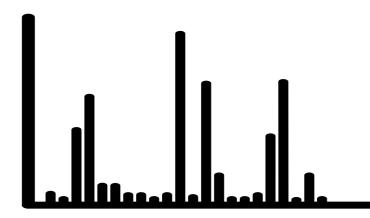
# Benchmark - v2

## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra

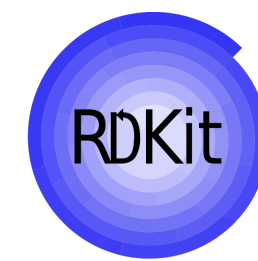
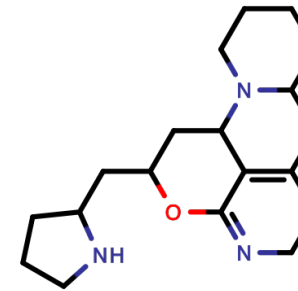
210,400

at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra

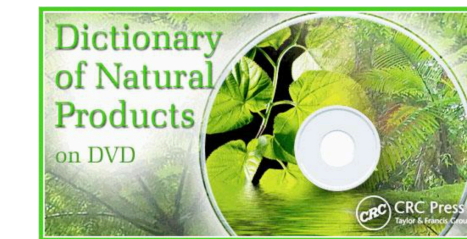
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



<http://globalnames.org>

text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

2107 unique entries

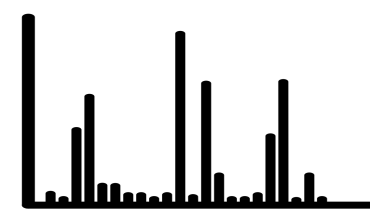
# Benchmark - v2

## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra

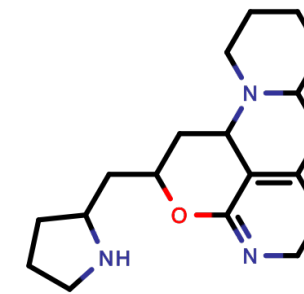
210,400

at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
~~[M+H]<sup>+</sup> adduct filter~~



40,138 cleaned spectra

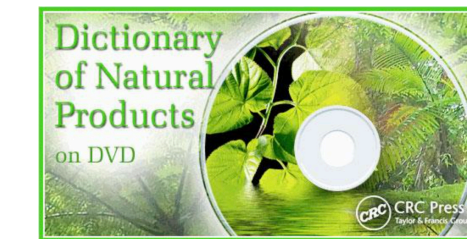
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



<http://globalnames.org>

text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

2107 unique entries

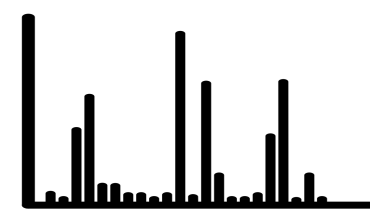
# Benchmark - v2

## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra

210,400

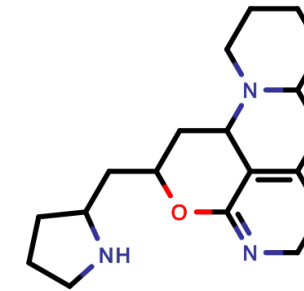
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
~~[M+H]<sup>+</sup> adduct filter~~



~~40,138~~ cleaned spectra

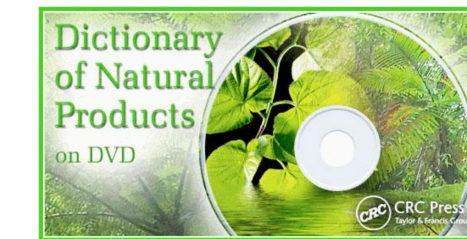
184,698

## STRUCTURES



chemical translation  
and sanitization

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



<http://globalnames.org>

text recognition  
matching and  
resolving against the  
Catalogue of Life

Cleaned  
2D structures

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

2107 unique entries



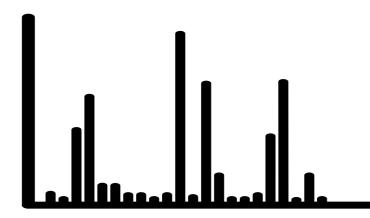
# Benchmark - v2

## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
66,646 experimental spectra

210,400

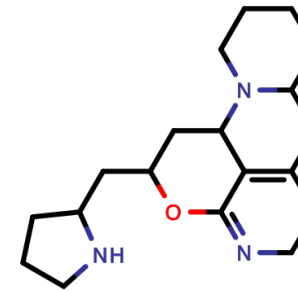
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filter



40,138 cleaned spectra

184,698

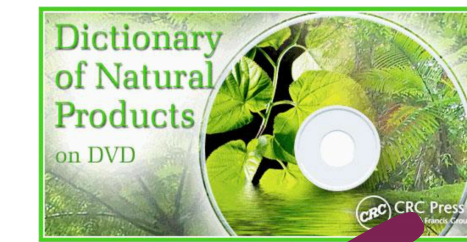
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"

recognition  
matching and  
resolving against the  
Catalogue of Life  
<http://balnames.org>

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

2107 unique entries

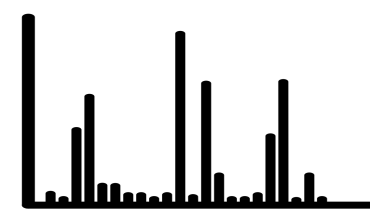
# Benchmark - v2

## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
66,646 experimental spectra

210,400

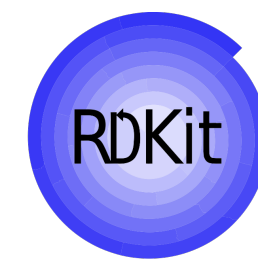
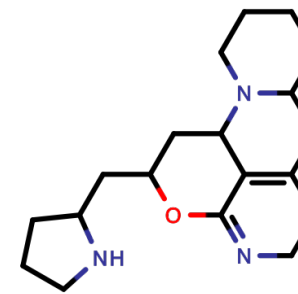
at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filter



40,138 cleaned spectra

184,698

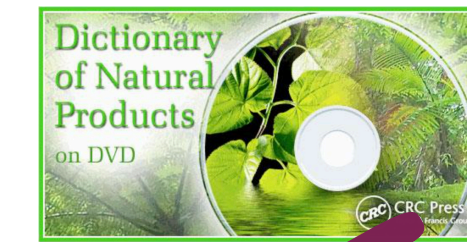
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



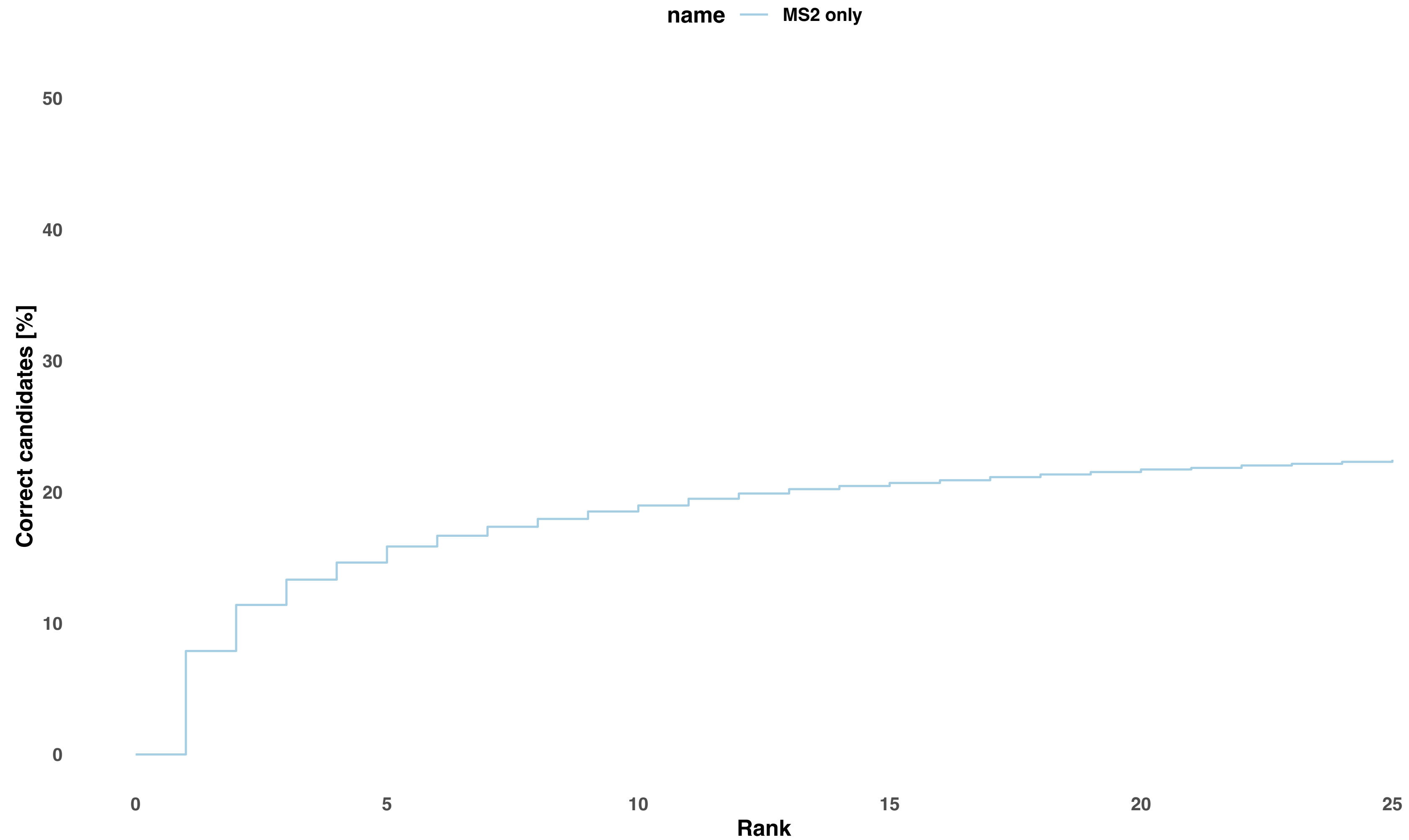
"Alkaloid from *Brunfelsia hopeana*"

recognition  
matching and  
resolving against the  
Catalogue of Life  
<http://balnames.org>

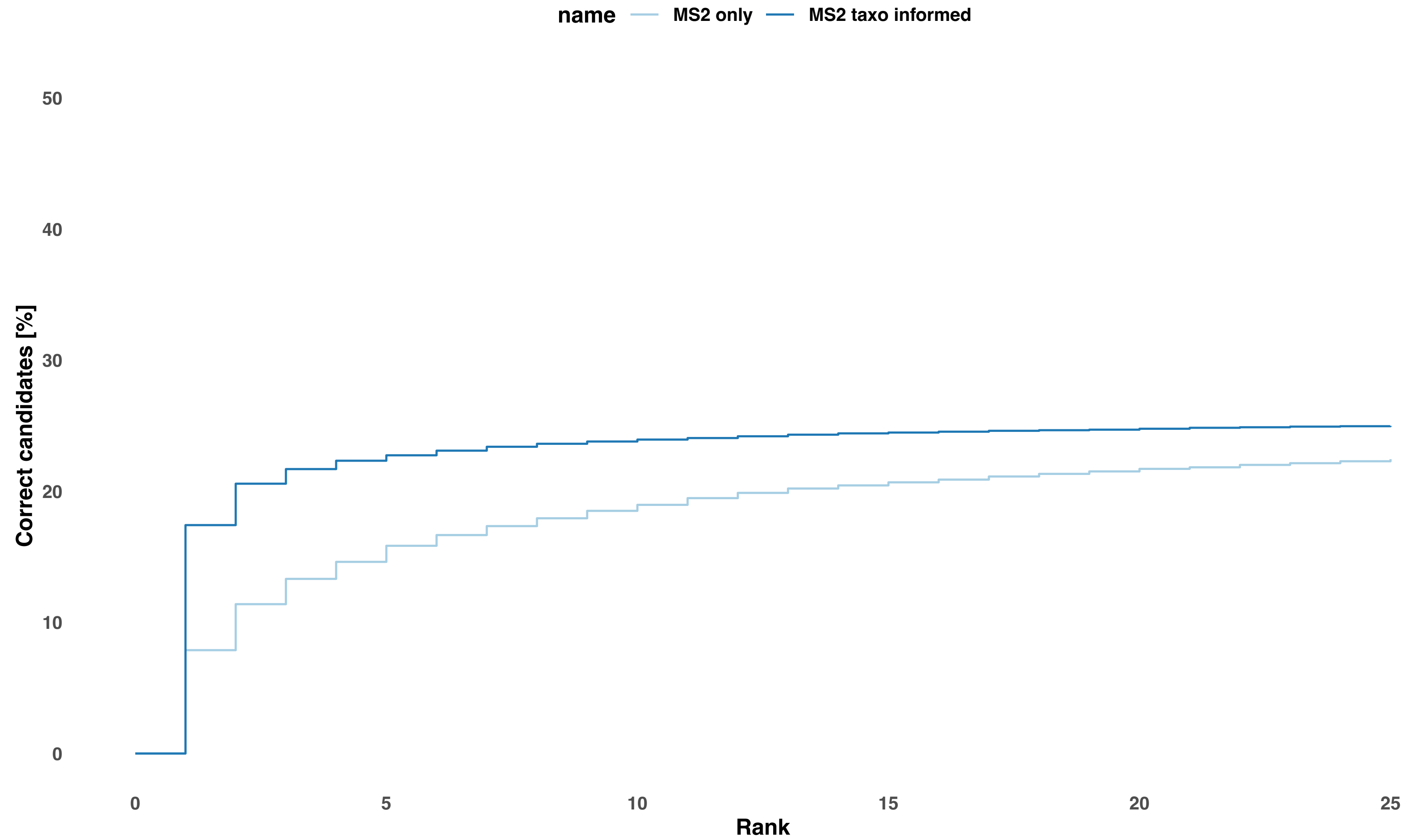
Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

22,251 spectra (12,482 structures) in positive mode  
7,472 spectra (6,587 structures) in negative mode

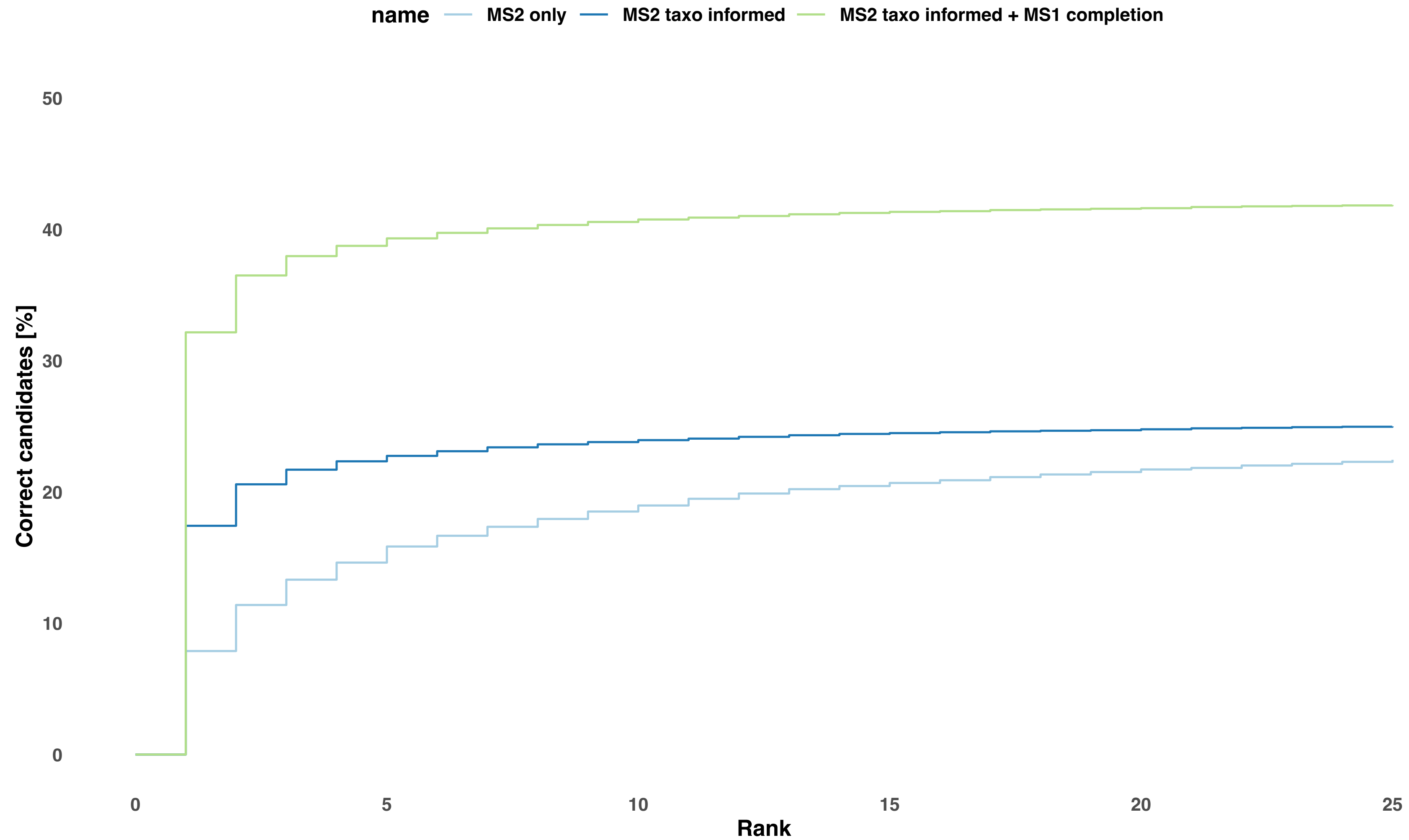
# Benchmark - v2



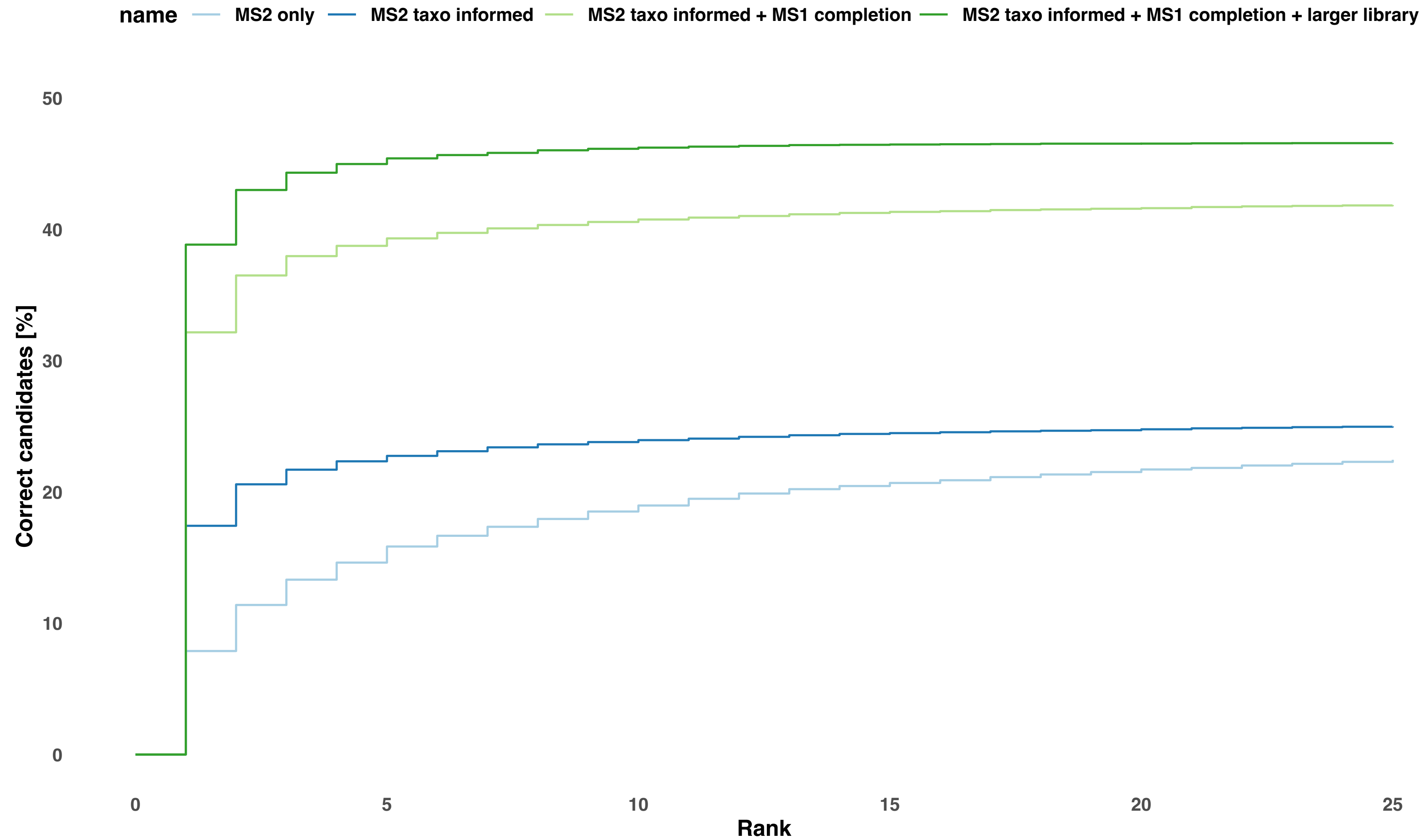
# Benchmark - v2



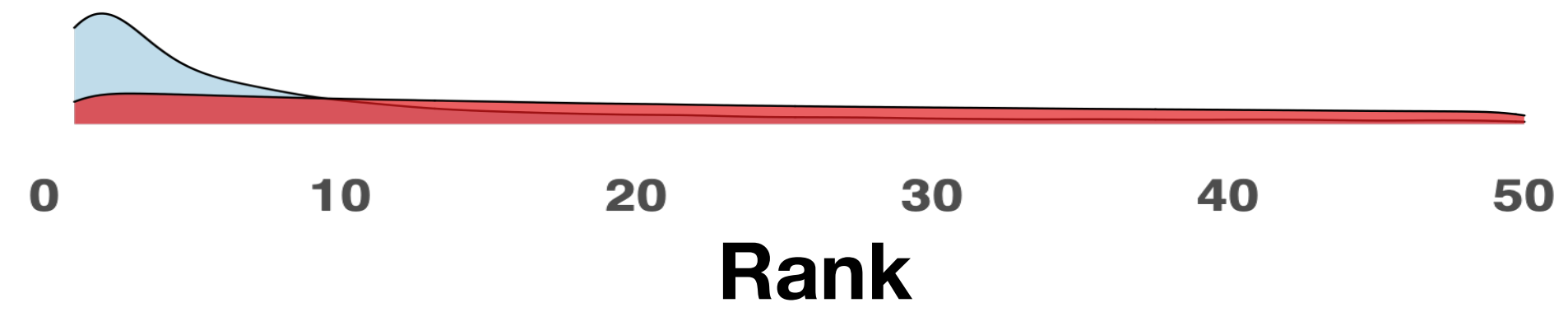
# Benchmark - v2



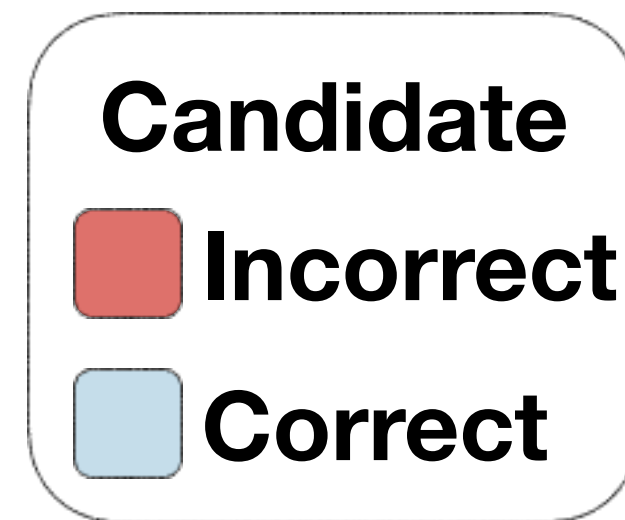
# Benchmark - v2



# Benchmark - v2

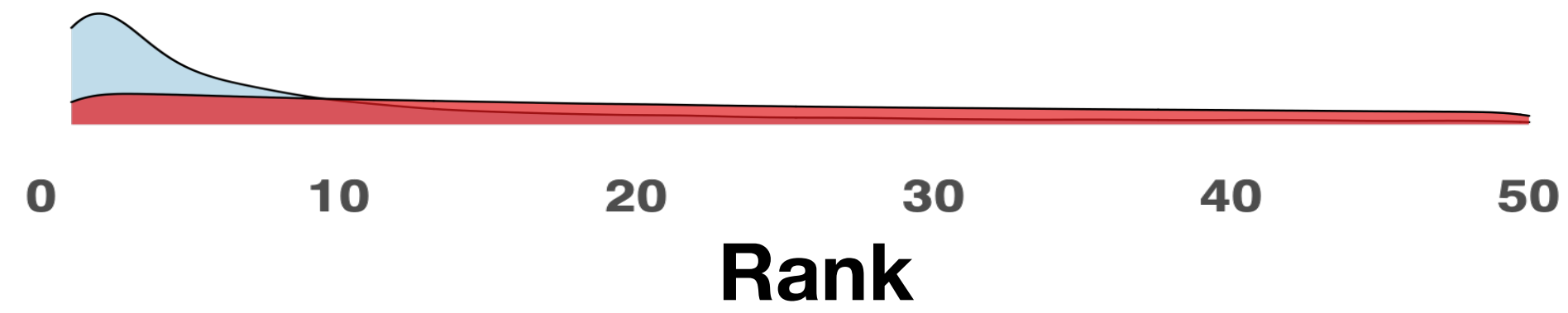


**Before TIMA**

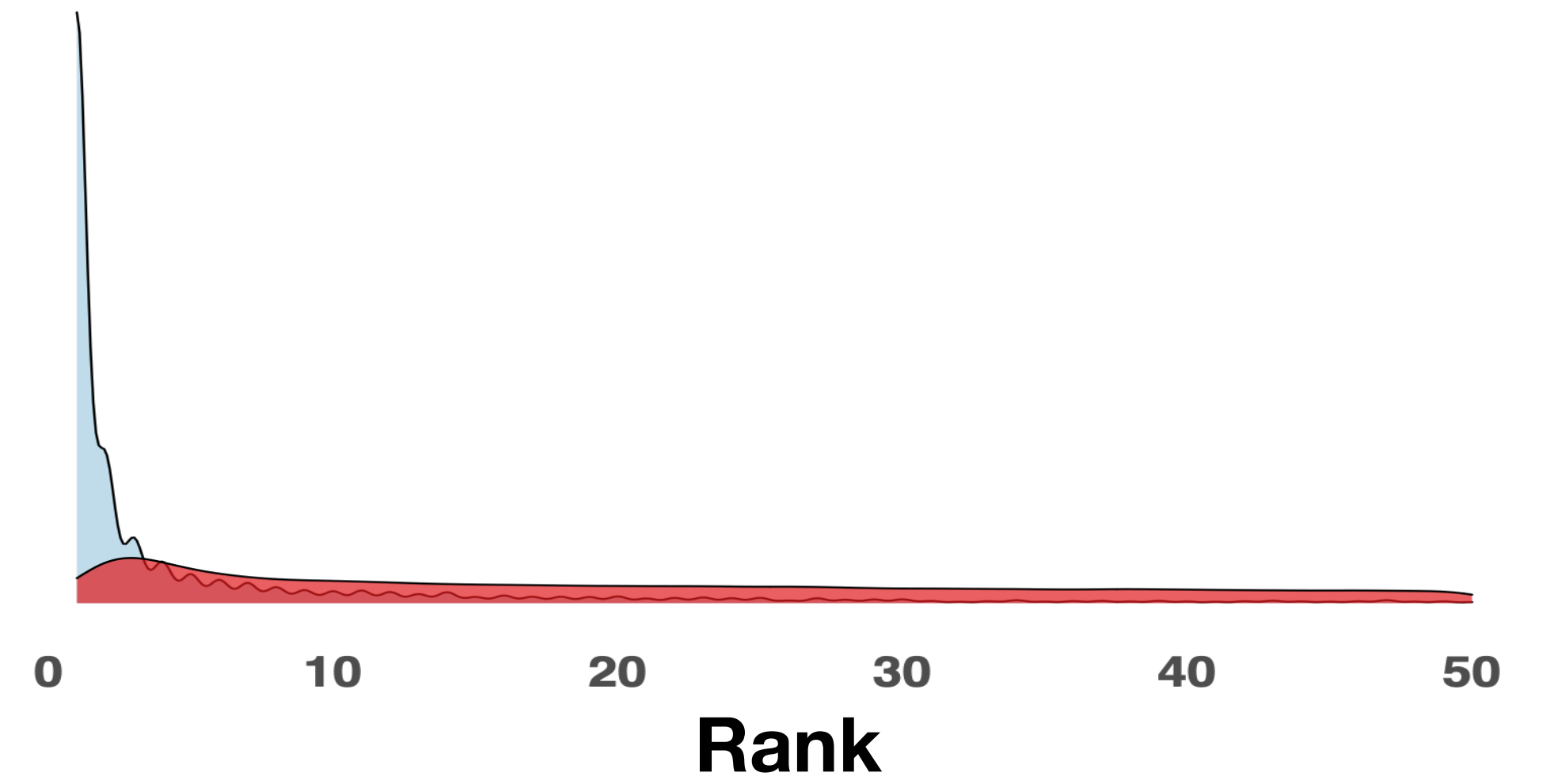
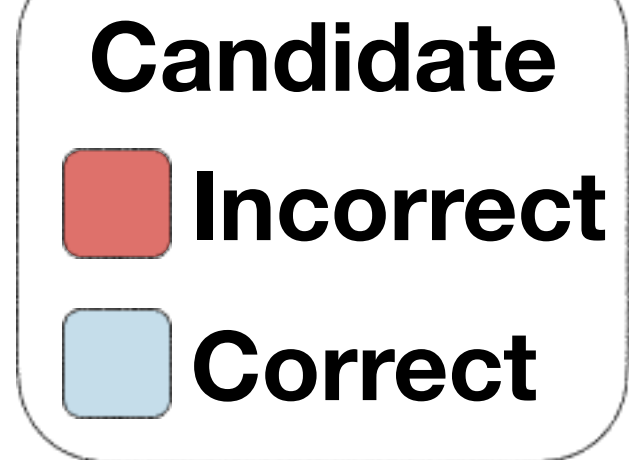




# Benchmark - v2

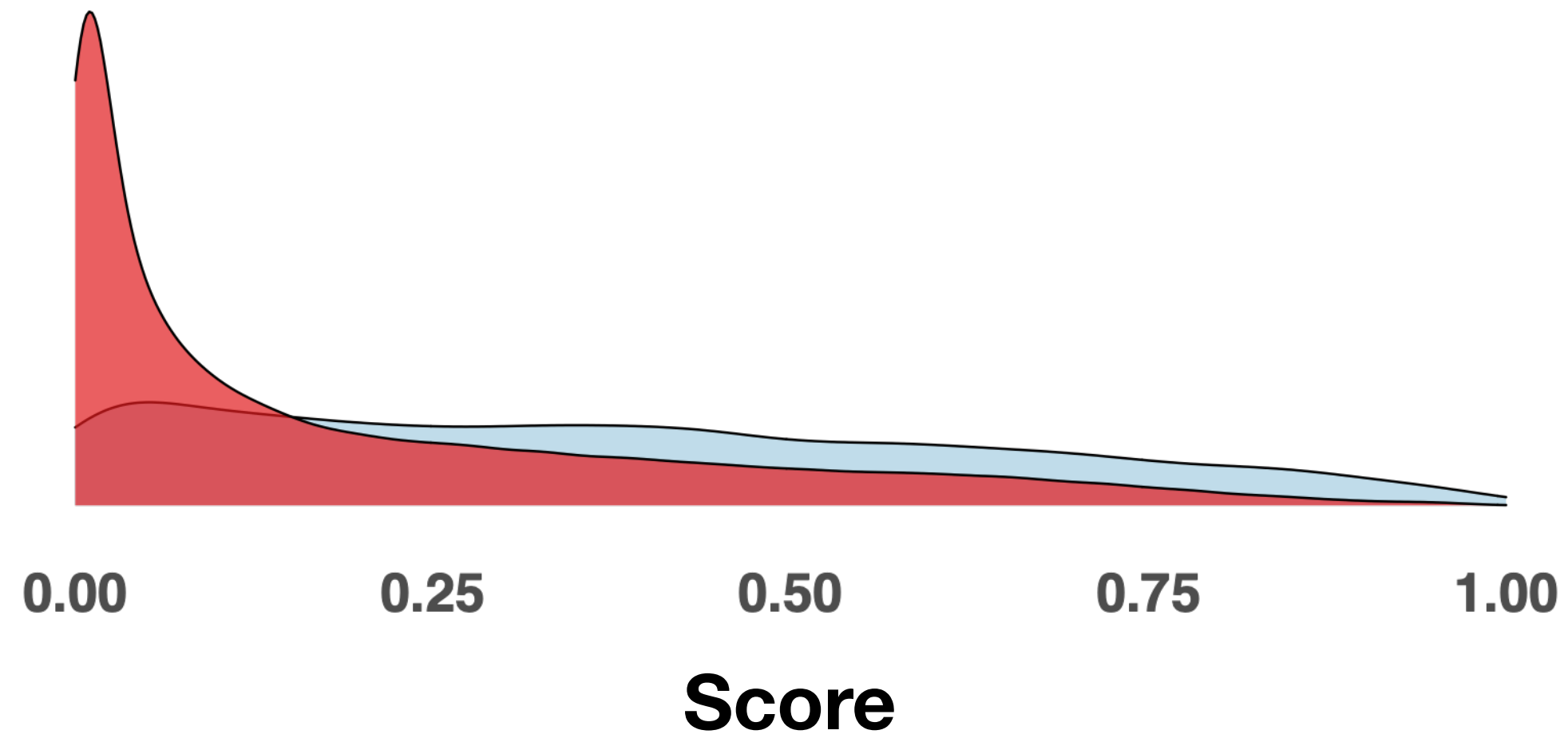
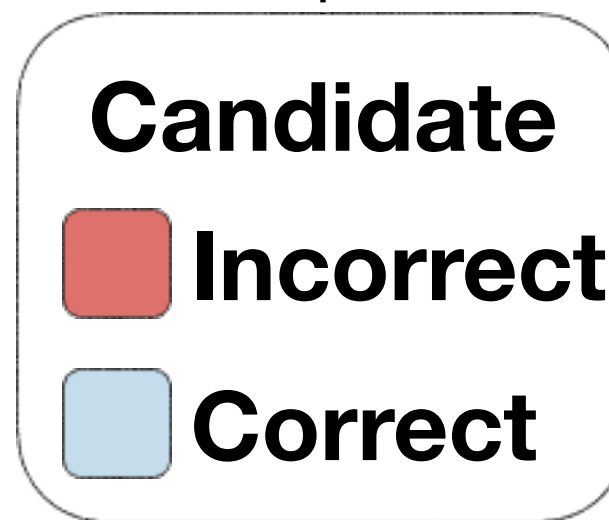
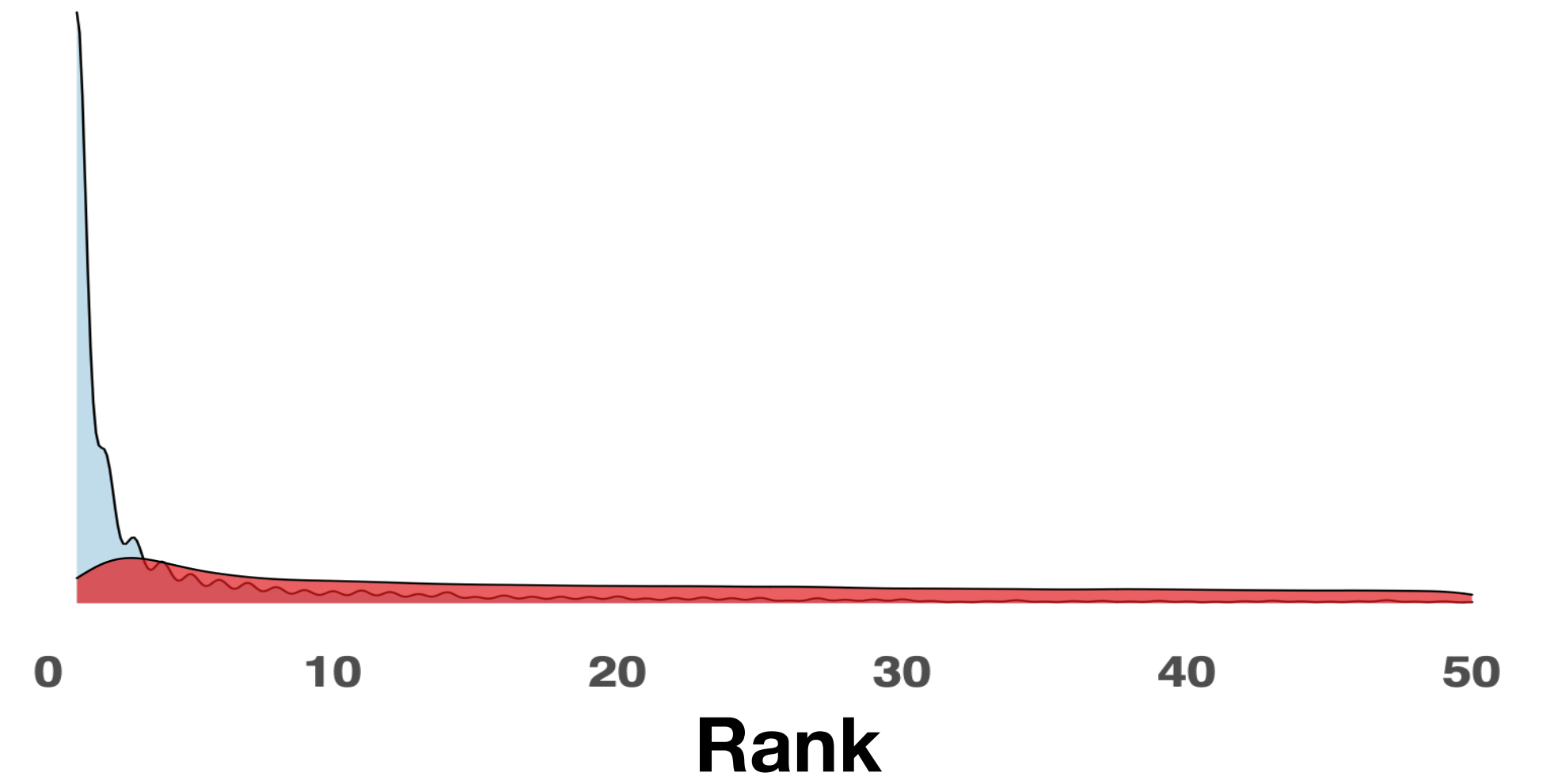
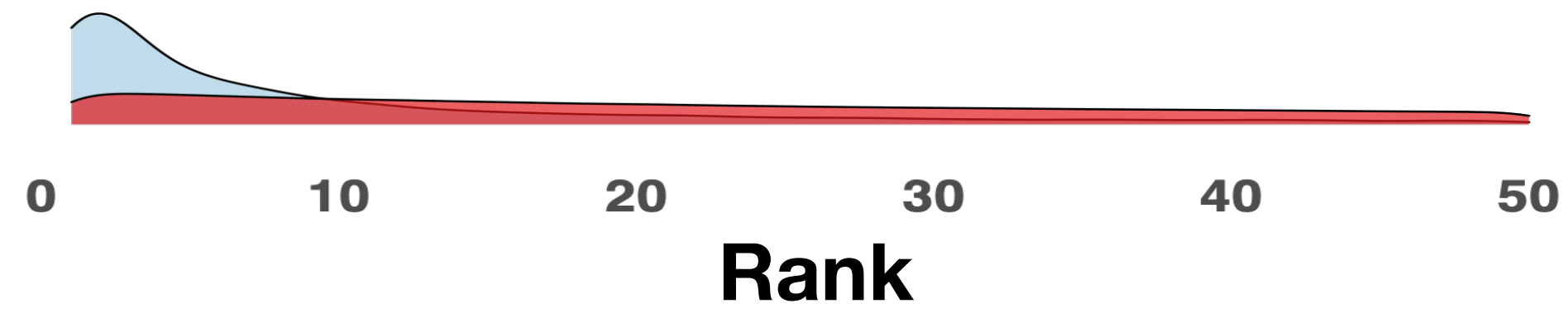


Before TIMA

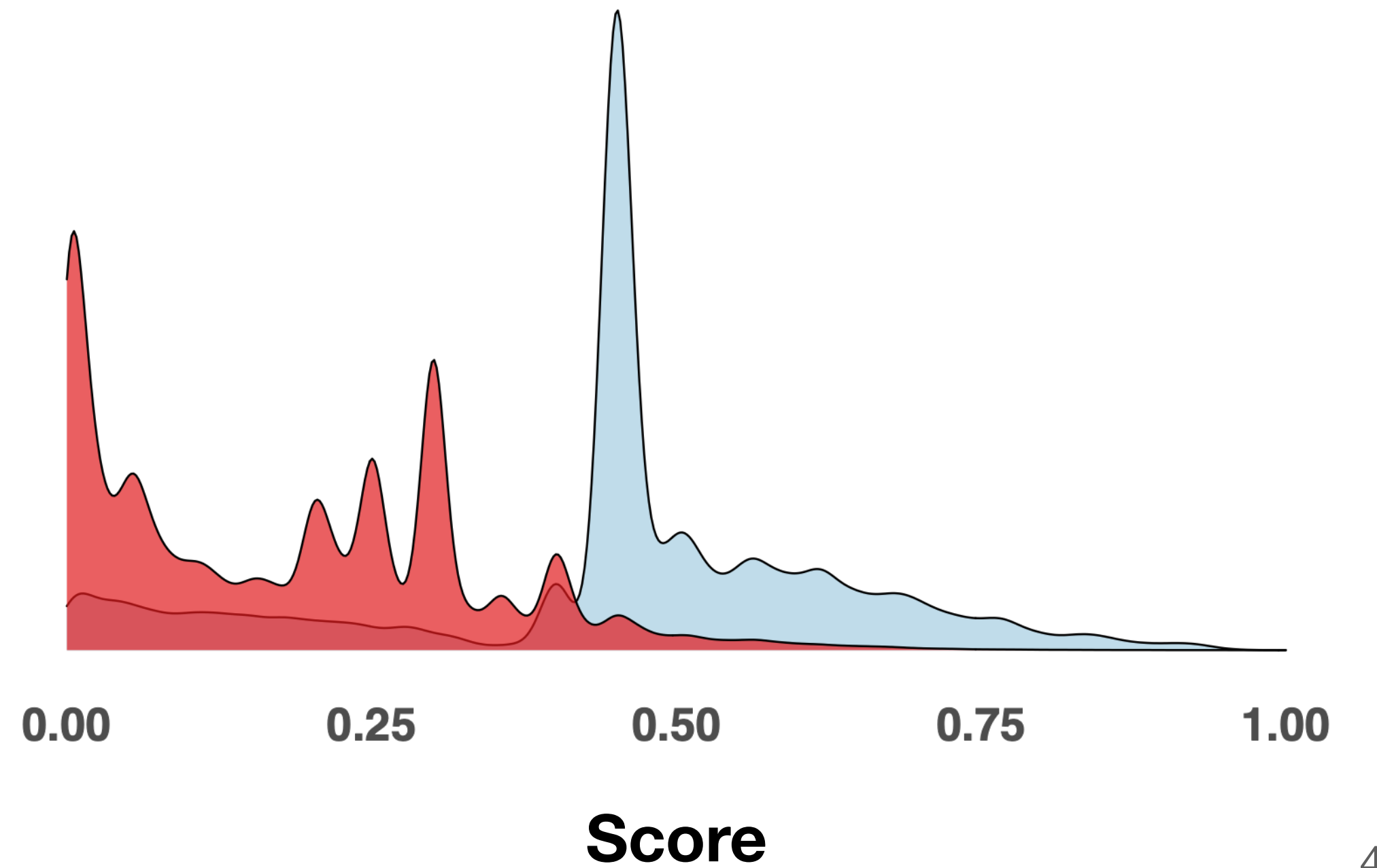
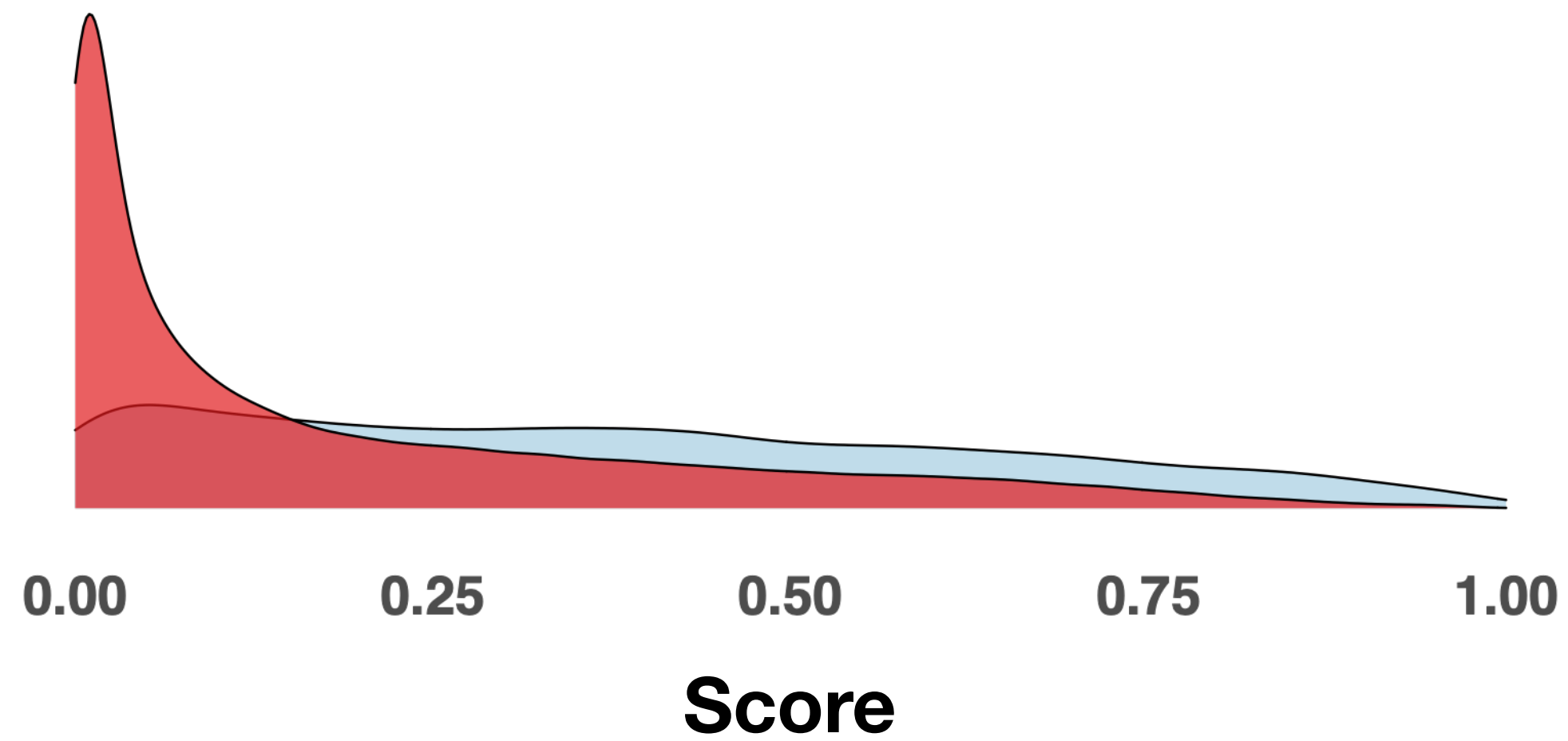
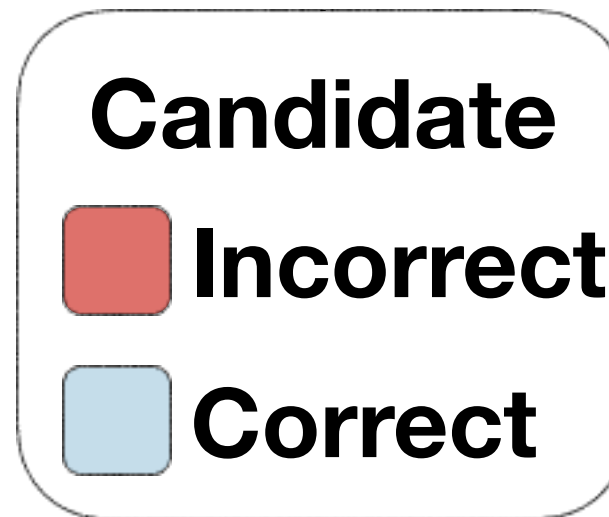
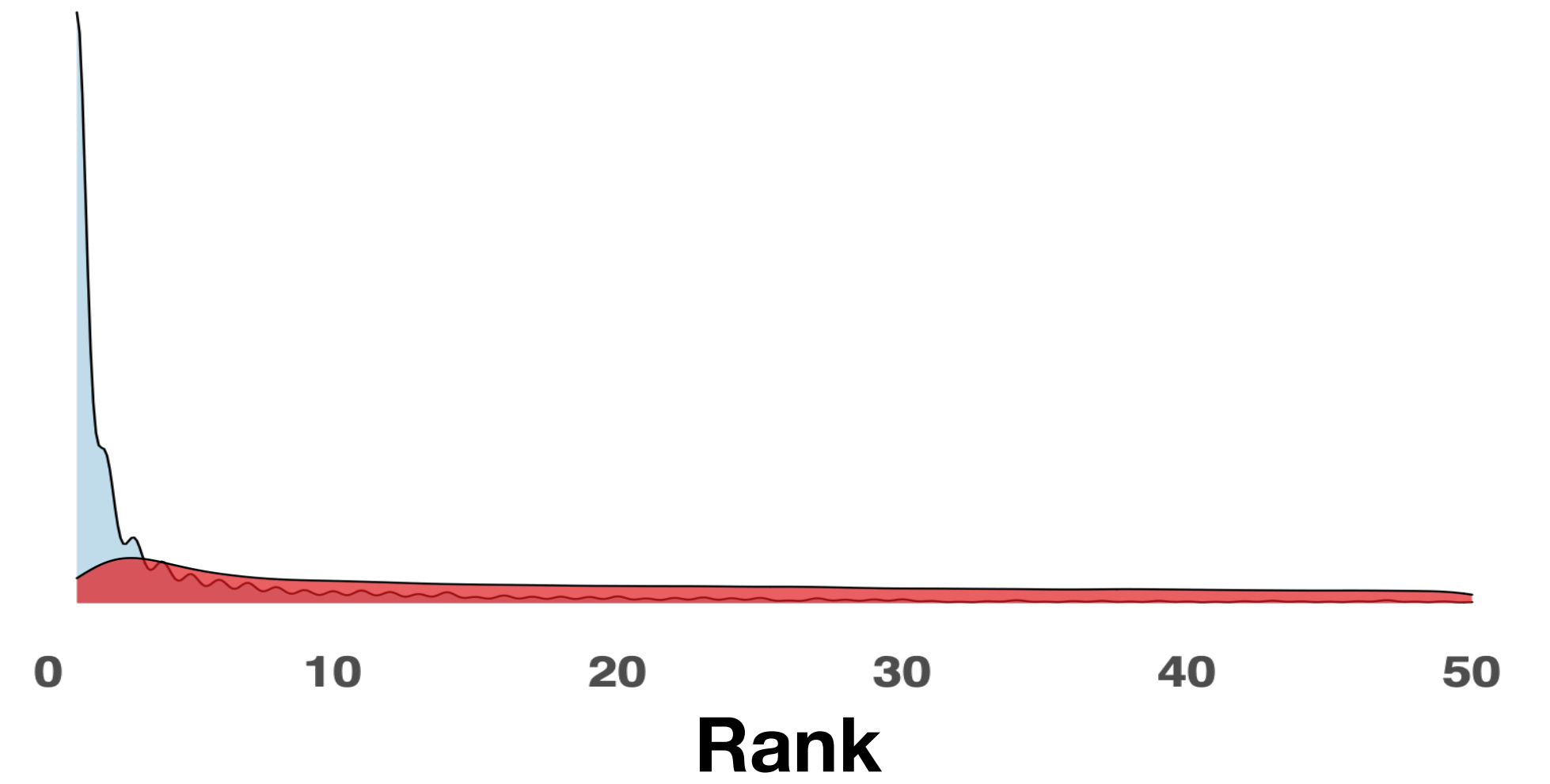
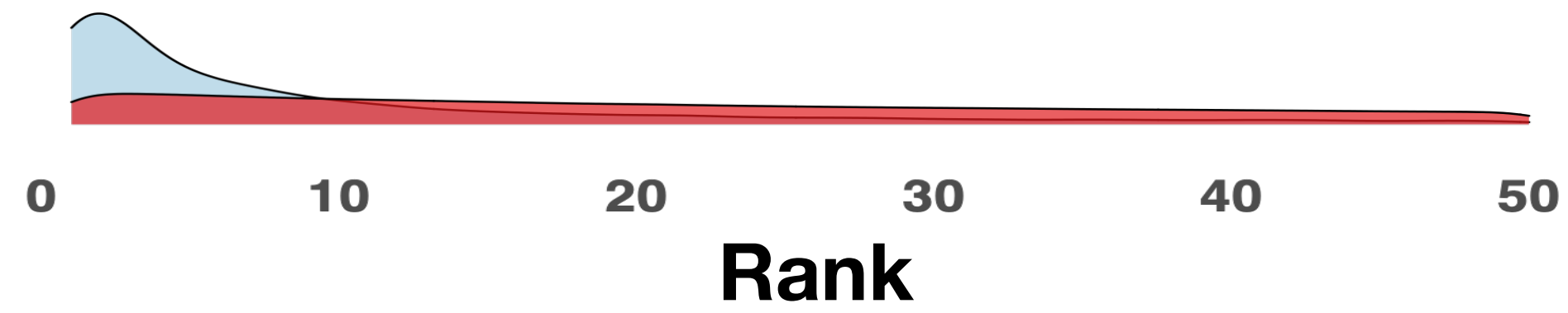


After TIMA

# Benchmark - v2

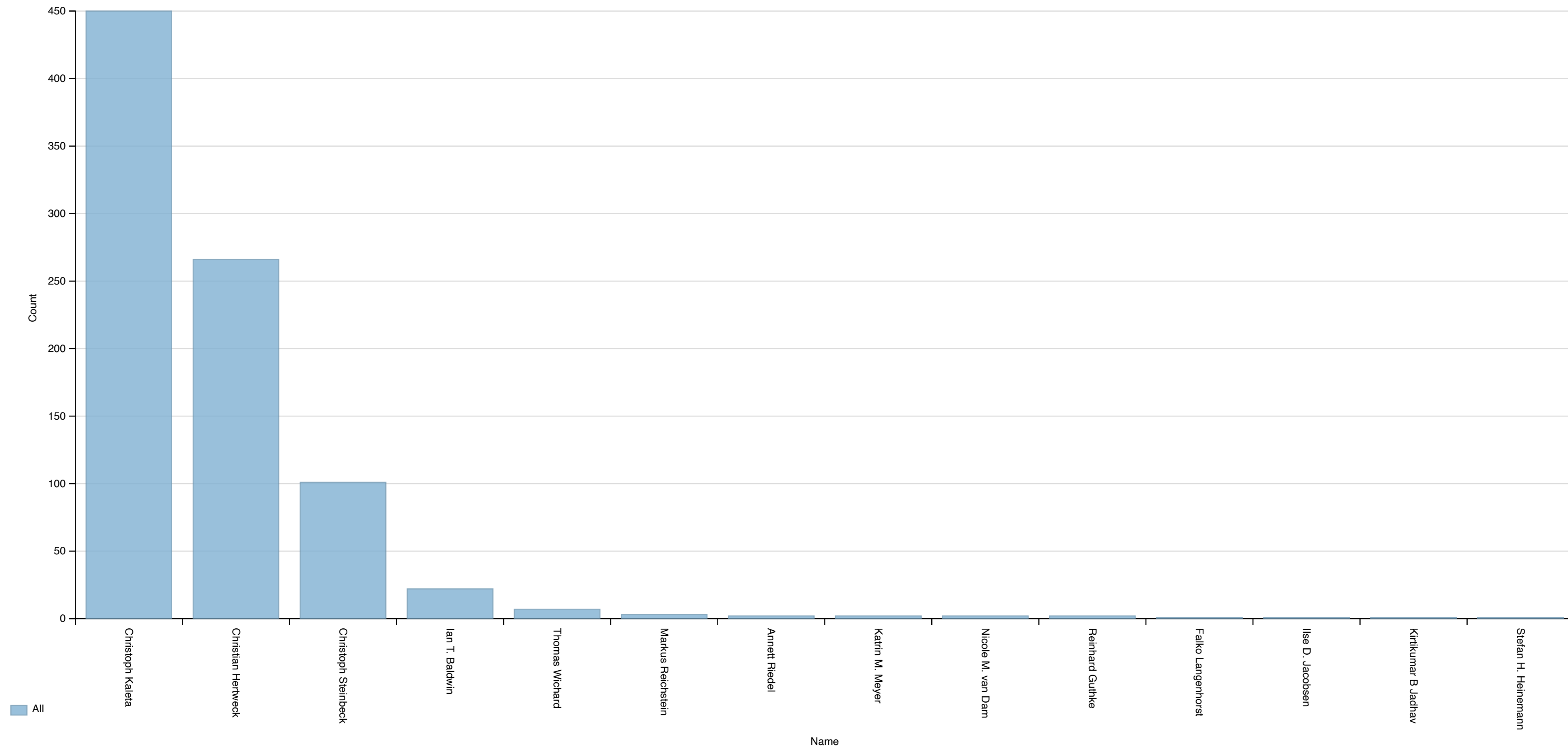


# Benchmark - v2



# From metabolites to hypotheses (and back)

« Hey Wiki, who can I contact if I want to talk about Natural Products chemistry in Jena? »



# From metabolites to hypotheses (and back)

« *Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same?* »



Contents lists available at [ScienceDirect](#)

Microbiological Research

journal homepage: [www.elsevier.com/locate/micres](http://www.elsevier.com/locate/micres)



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose<sup>a,\*,1</sup>, Anjisha Maharshi<sup>a,2</sup>, Bhavanath Jha<sup>a,b,\*\*</sup>

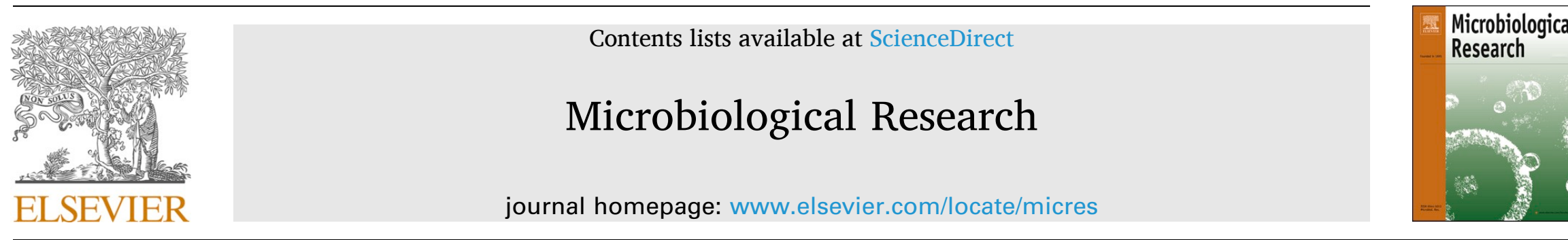
<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same? »



*Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?*

Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose<sup>a,\*,1</sup>, Anjisha Maharshi<sup>a,2</sup>, Bhavanath Jha<sup>a,b,\*\*</sup>

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India  
<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

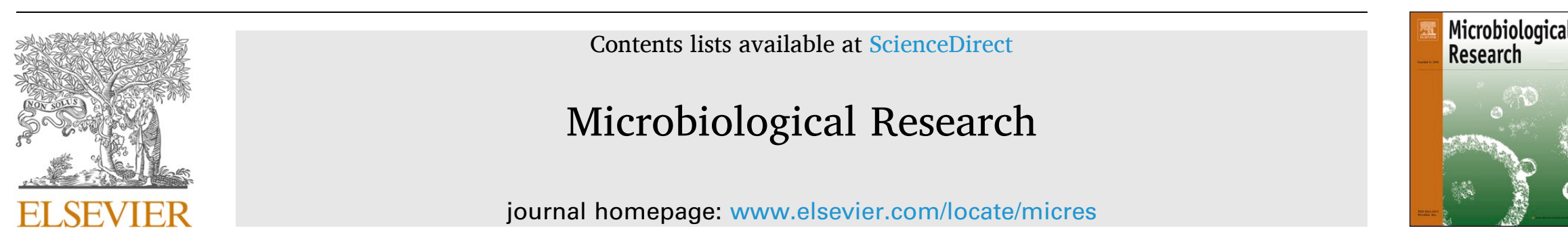


<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>



# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same? »



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose<sup>a,\*,1</sup>, Anjisha Maharshi<sup>a,2</sup>, Bhavanath Jha<sup>a,b,\*\*</sup>

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India  
<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

**Table 1**  
 Novel bioactive compounds isolated from Actinobacteria during last five years (2014-2019).

Organism	Compound	Bioactivity <sup>#</sup>	Reference
<i>Arthrobacter</i> sp. PGVB1	Arthroamide	f (anti-quorum sensing)	J Nat Prod. 2015;78:2827–2831
<i>Actinoallomurus</i> sp. ID145113 and 145,206	Paramagnetoquinones A–C (1–3)	(1/2) a; (3) a	J Nat Prod. 2017;80:819–827
<i>Actinoalloteichus hymeniacidonis</i> 179DD-027	Dokdolipid B	c	Mar Drugs 2019;17:237
<i>Actinomadura atramentaria</i> NBRC 14,695**	Cinnamycin B	a	J Ind Microbiol Biotechnol. 2016;43:1159–65
<i>Actinomadura</i> sp. BCC 35,430*	Actinomadurone	b	Tetrahedron Lett. 2017;58:3223–3225
<i>Actinomadura</i> sp. K13–0306	Sagamilactam	c, f (antiparasitic)	J Antibiot. 2016;69:818–824
<i>Actinomadura</i> sp. K4S16	Nonthmicin (1) & ecteinamycin (2)	(1,2) c	Org Lett. 2017;19:1406–1409
<i>Actinomadura</i> sp. KC191*	Actinomadurol	a	J Nat Prod. 2016;79:1886–1890
<i>Actinosynnema pretiosum</i> HGF052::asm18**	Actinosynneptide A (1) & B (2)	(1,2) c	Appl Microbiol Biotechnol. 2016;101:2273–2279
<i>Amycolatopsis</i> sp. IRD-009	Pradimicin-IRD	a, c	Natural Product Res. 2019;33:1713–1720

Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?

508 results in 4.8 seconds

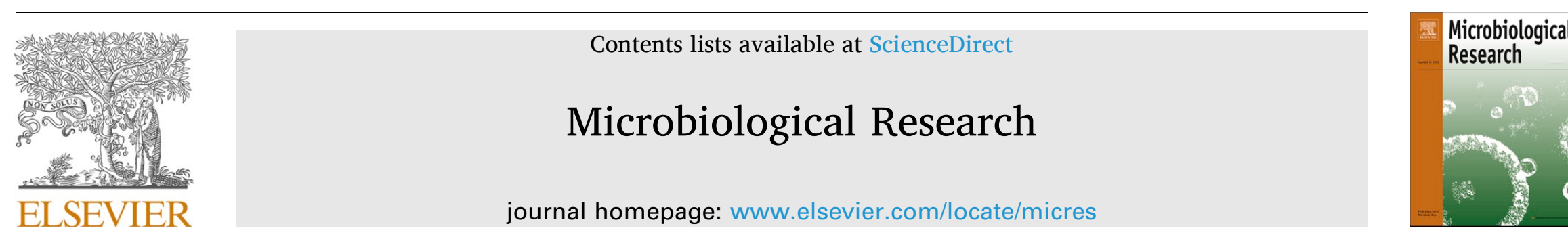
Taxon name	Structure	Activity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic   immunosuppressive drug	<a href="https://doi.org/10.1007/S11306-016-1025-6">10.1007/S11306-016-1025-6</a>
<i>Saccharopolyspora erythraea</i>		Antibiotic   protein synthesis inhibitors   gastrointestinal agent	<a href="https://doi.org/10.1021/ACSSYNBIO.8B00372">10.1021/ACSSYNBIO.8B00372</a>
<i>Streptomyces venezuelae</i>		Antibiotic   protein synthesis inhibitors	<a href="https://doi.org/10.1016/J.JBIOTEC.2014.01.028">10.1016/J.JBIOTEC.2014.01.028</a>

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>



# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same? »



## Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose<sup>a,\*,1</sup>, Anjisha Maharshi<sup>a,2</sup>, Bh...

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, Marg, Bhavnagar, Gujarat, 364002, India  
<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

**Table 1**  
 Novel bioactive compounds isolated from Actinobacteria during last five years (2014-2019).

Organism	Compound	Bioactivity <sup>#</sup>	Reference
<i>Arthrobacter</i> sp. PGVB1	Arthroamide	f (anti-quorum sensing)	J Nat Prod. 2015;78:2827–2831
<i>Actinoallomurus</i> sp. ID145113 and 145,206	Paramagnetoquinones A–C (1–3)	(1/2) a; (3) a	J Nat Prod. 2017;80:819–827
<i>Actinoalloteichus hymeniacidonis</i> 179DD-027	Dokdolipid B	c	Mar Drugs 2019;17:237
<i>Actinomadura atramentaria</i> NBRC 14,695**	Cinnamycin B	a	J Ind Microbiol Biotechnol. 2016;43:1159–65
<i>Actinomadura</i> sp. BCC 35,430*	Actinomadurone	b	Tetrahedron Lett. 2017;58:3223–3225
<i>Actinomadura</i> sp. K13–0306	Sagamilactam	c, f (antiparasitic)	J Antibiot. 2016;69:818–824
<i>Actinomadura</i> sp. K4S16	Nonthmicin (1) & ecteinamycin (2)	(1,2) c	Org Lett. 2017;19:1406–1409
<i>Actinomadura</i> sp. KC191*	Actinomadurol	a	J Nat Prod. 2016;79:1886–1890
<i>Actinosynnema pretiosum</i> HGF052::asm18**	Actinosynneptide A (1) & B (2)	(1,2) c	Appl Microbiol Biotechnol. 2016;101:2273–2279
<i>Amycolatopsis</i> sp. IRD-009	Pradimicin-IRD	a, c	Natural Product Res. 2019;33:1713–1720

Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?

208 results in 4.8 seconds

Taxon name	Structure	Activity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic   immunosuppressive drug	<a href="https://doi.org/10.1007/S11306-016-1025-6">10.1007/S11306-016-1025-6</a>
<i>Saccharopolyspora erythraea</i>		Antibiotic   protein synthesis inhibitors   gastrointestinal agent	<a href="https://doi.org/10.1021/ACSSYNBIO.8B00372">10.1021/ACSSYNBIO.8B00372</a>
<i>Streptomyces venezuelae</i>		Antibiotic   protein synthesis inhibitors	<a href="https://doi.org/10.1016/J.JBIOTEC.2014.01.028">10.1016/J.JBIOTEC.2014.01.028</a>

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

# From metabolites to hypotheses (and back)


« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

Parent taxon name	Count
<i>Streptomyces</i>	412
<i>Streptomycetaceae</i>	206
<i>Apis</i>	42
<i>Penicillium</i>	25
<i>Torrubiella</i>	25
<i>Micromonospora</i>	22
<i>Saccharopolyspora</i>	21
<i>Kitasatospora</i>	21
<i>Fusarium</i>	21
<i>Albifimbria</i>	21
<i>Aspergillus</i>	20

<https://w.wiki/3HMA>

# From metabolites to hypotheses (and back)

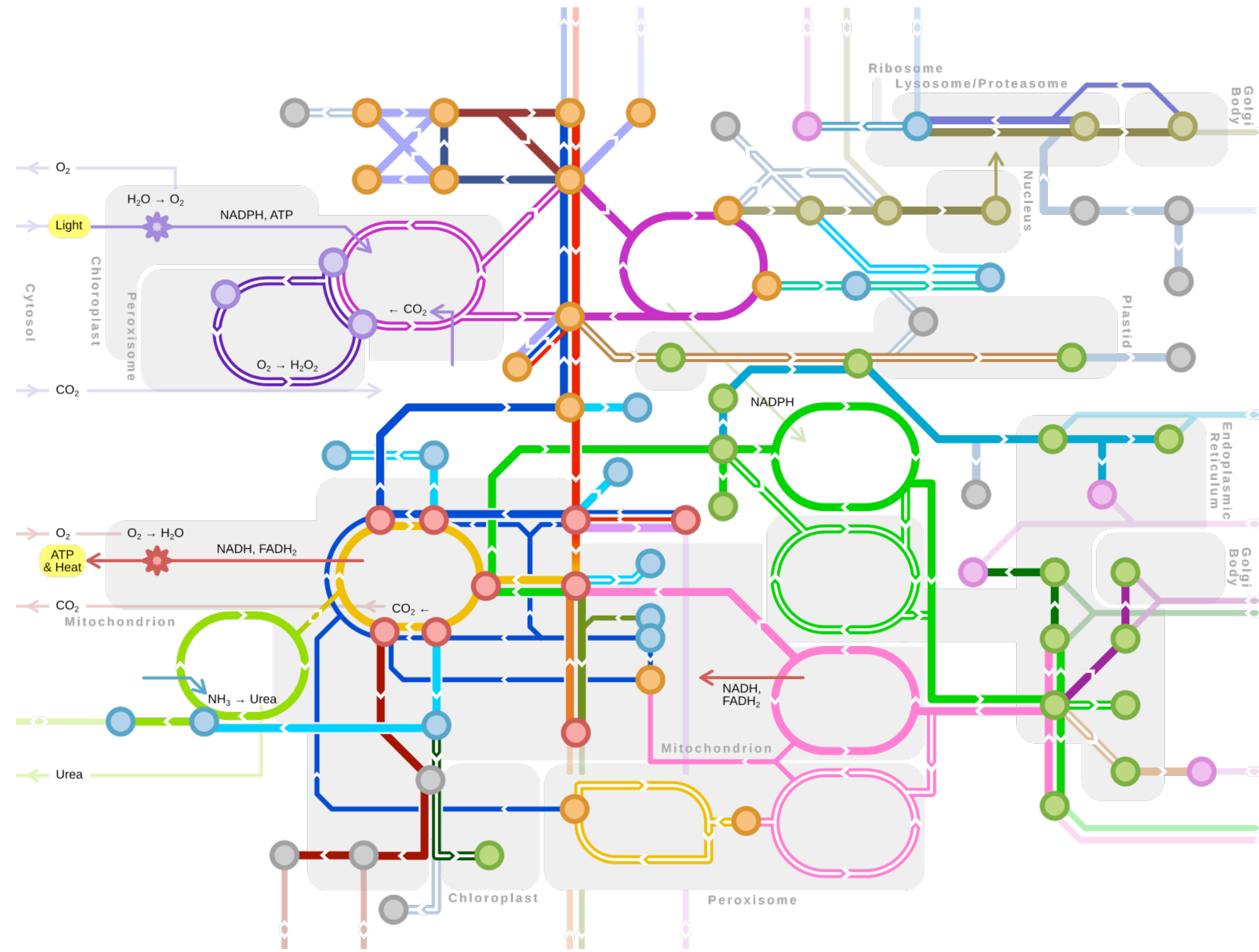
« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

Parent taxon name	Count
<i>Streptomyces</i>	412
<i>Streptomycetaceae</i>	206
<i>Apis</i> 	42
<i>Penicillium</i>	25
<i>Torrubiella</i>	25
<i>Micromonospora</i>	22
<i>Saccharopolyspora</i>	21
<i>Kitasatospora</i>	21
<i>Fusarium</i>	21
<i>Albifimbria</i>	21
<i>Aspergillus</i>	20

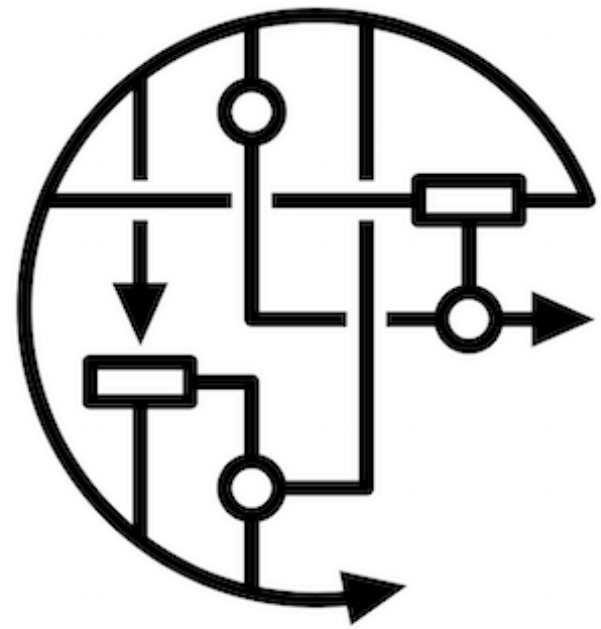
<https://w.wiki/3HMA>



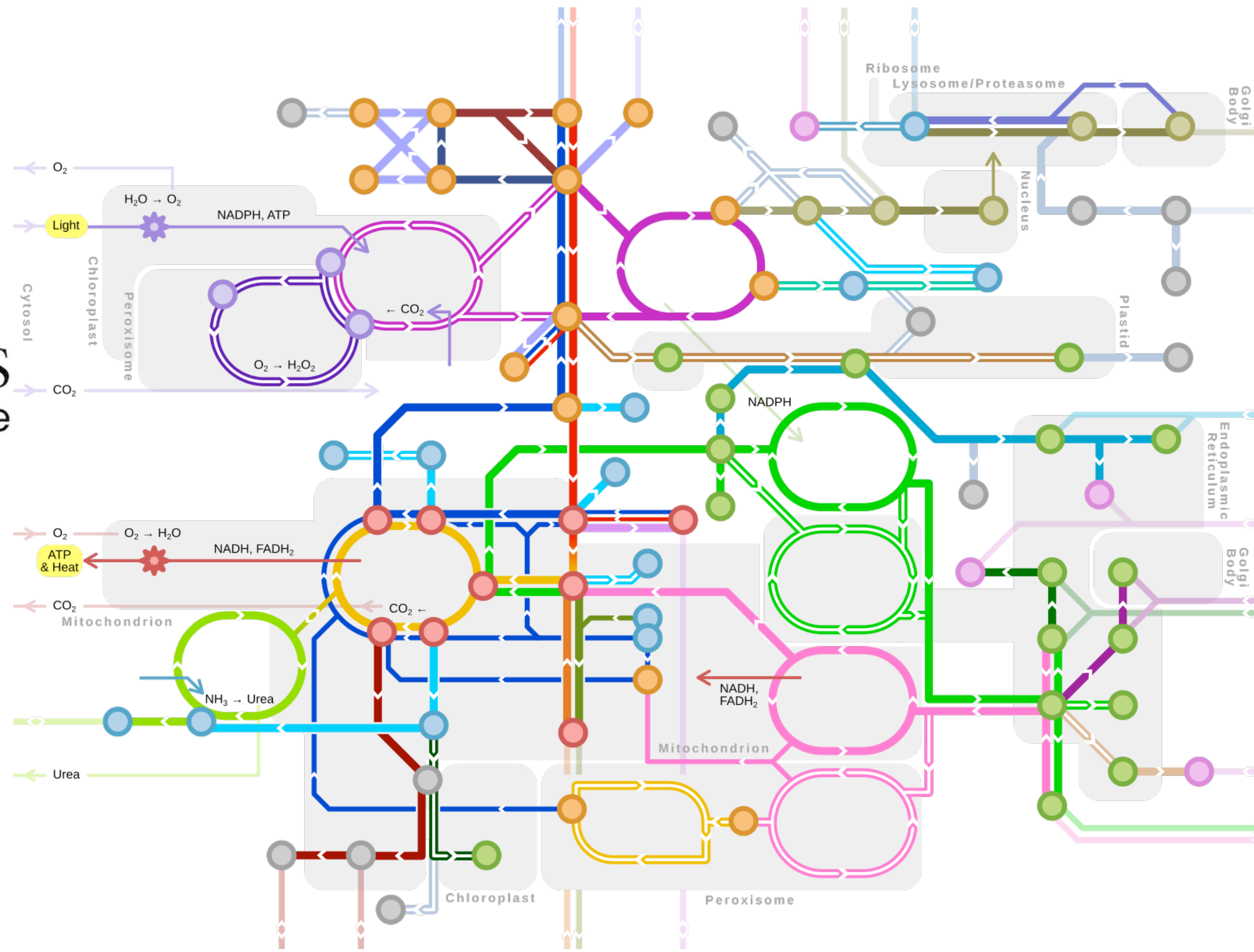
# From metabolites to hypotheses (and back)



# From metabolites to hypotheses (and back)



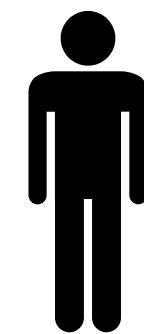
WIKIPATHWAYS  
Pathways for the People



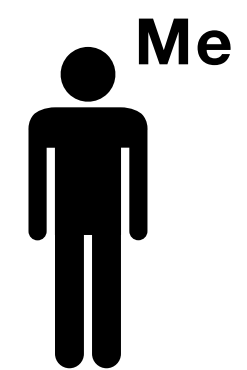
« Hey Wiki, are the metabolites I annotated known to interact with the cholesterol biosynthesis pathway? »



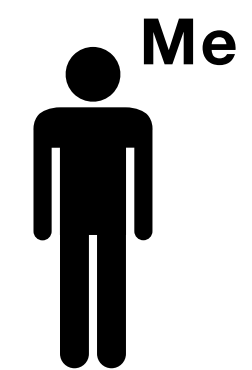
# Sharing is caring



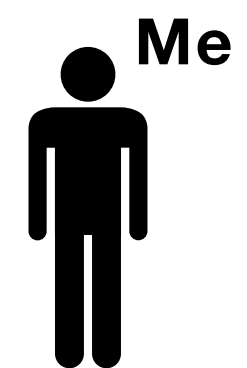
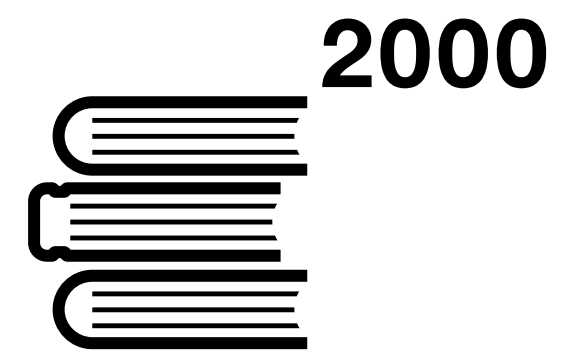
# Sharing is caring



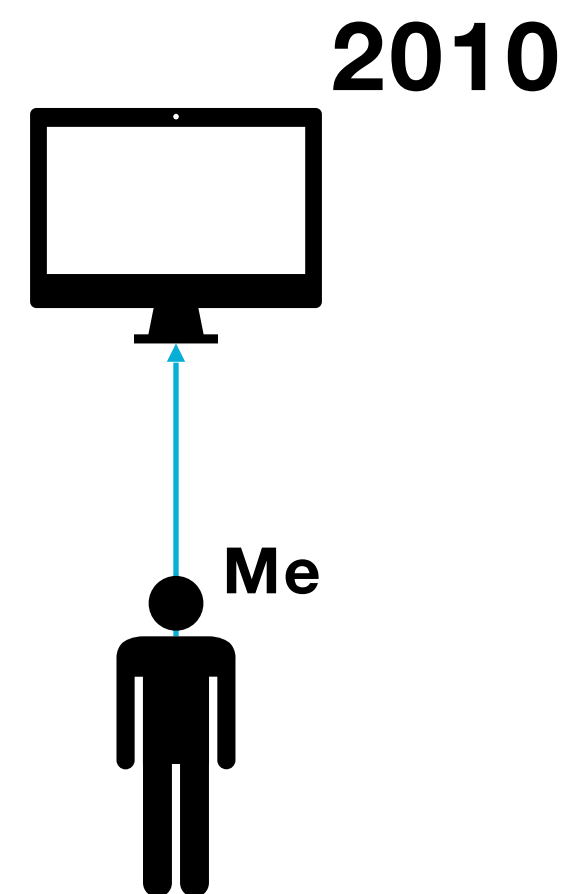
# Sharing is caring



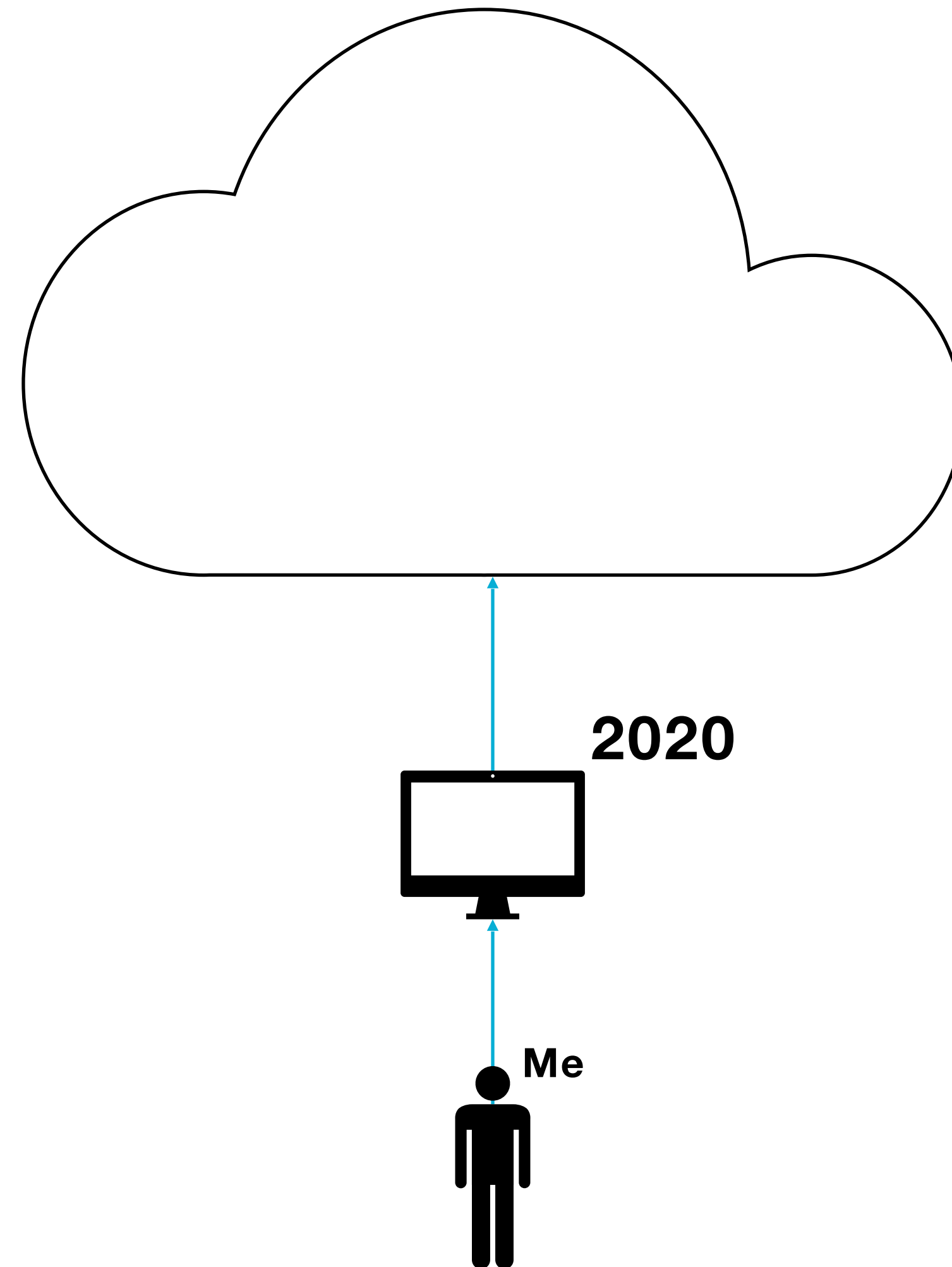
# Sharing is caring



# Sharing is caring

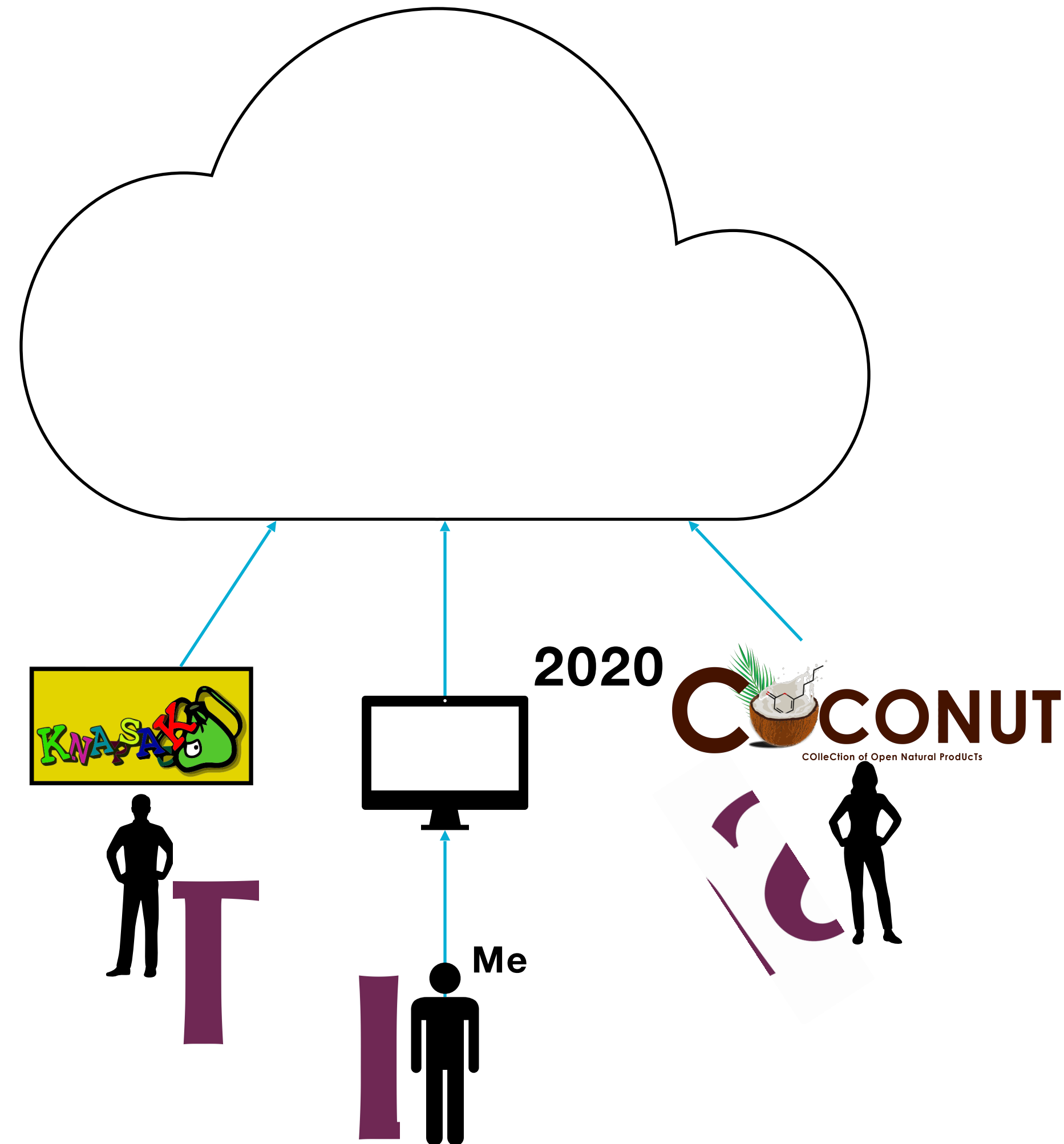


# Sharing is caring

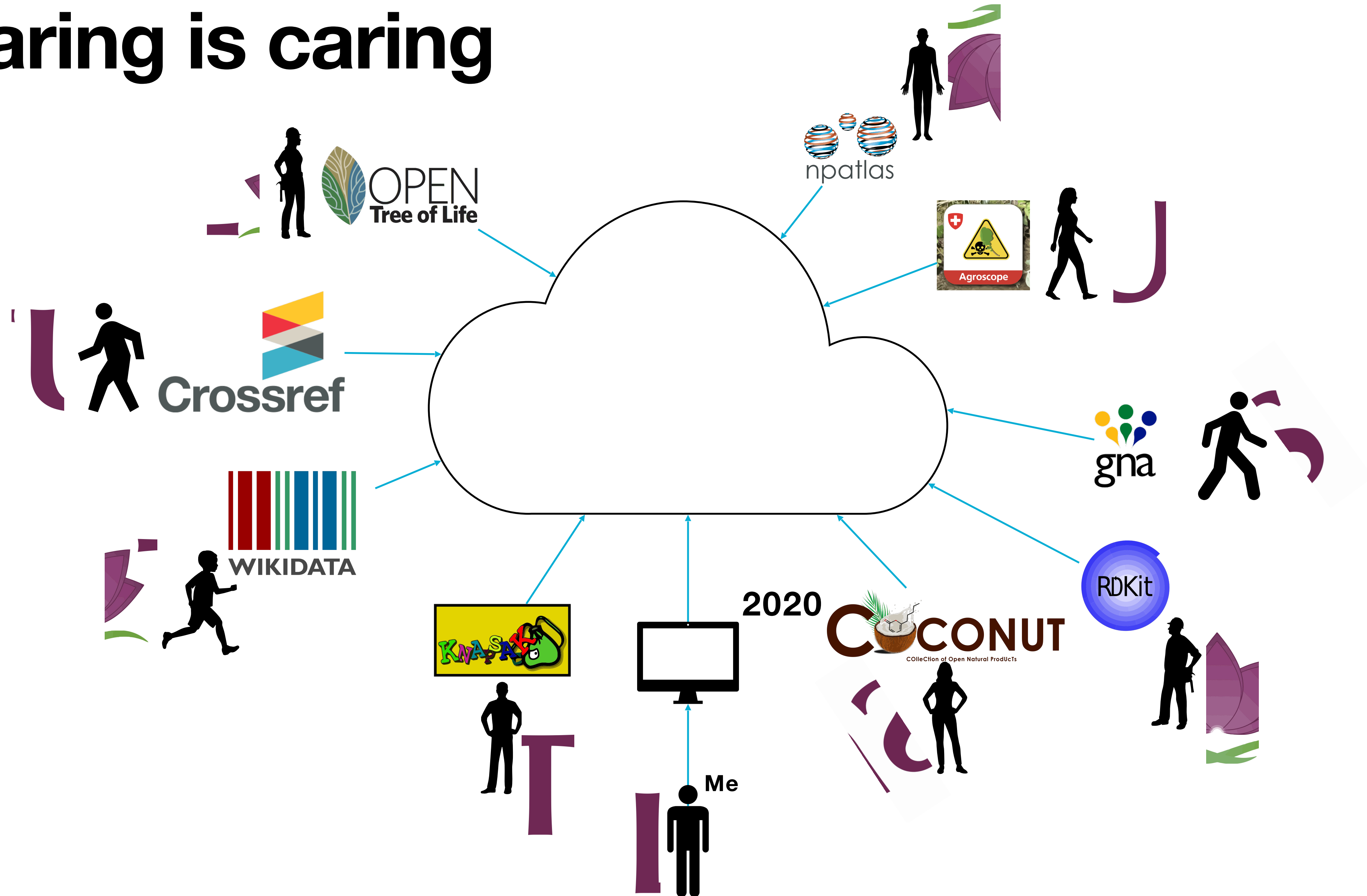




# Sharing is caring



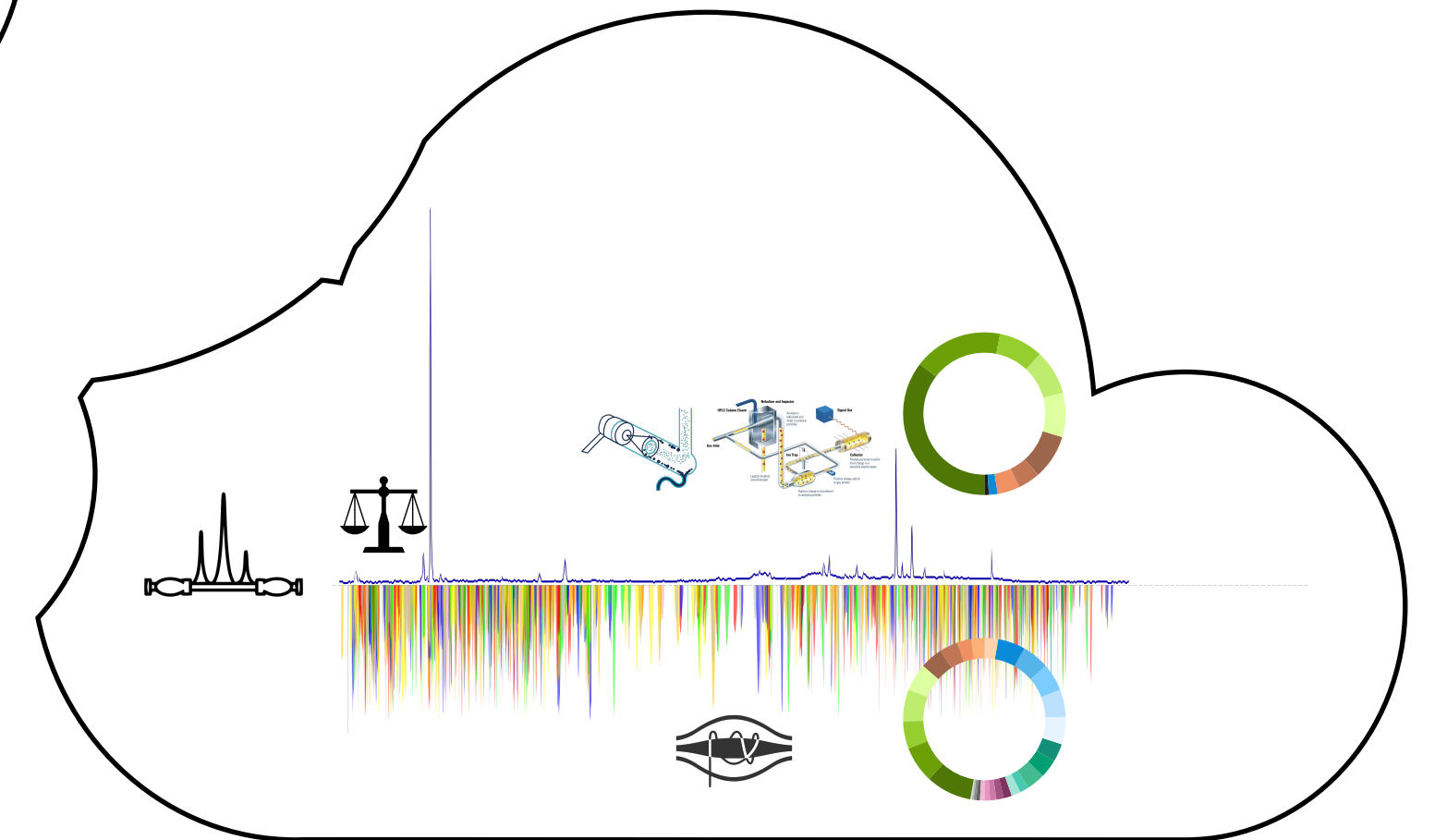
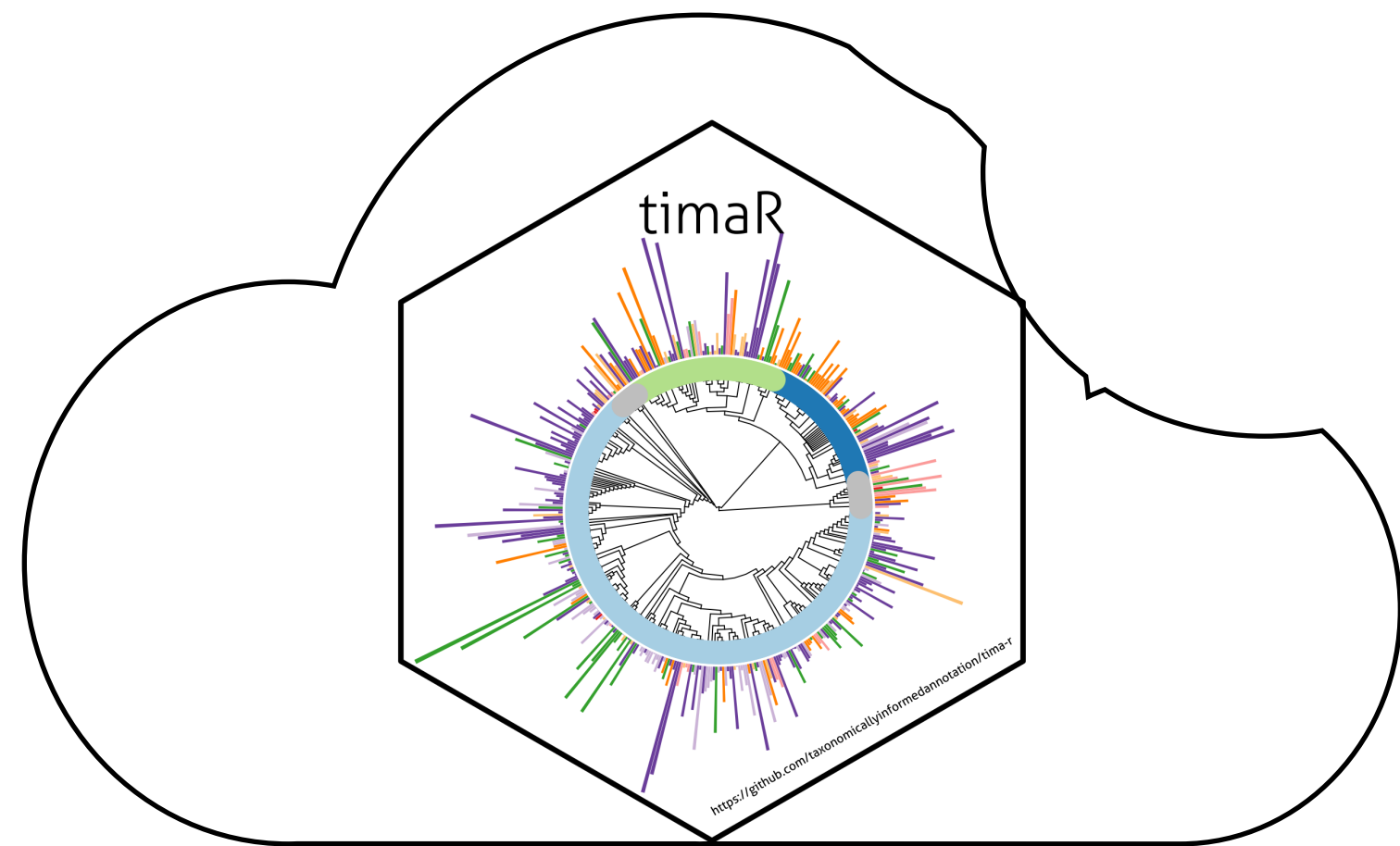
# Sharing is caring



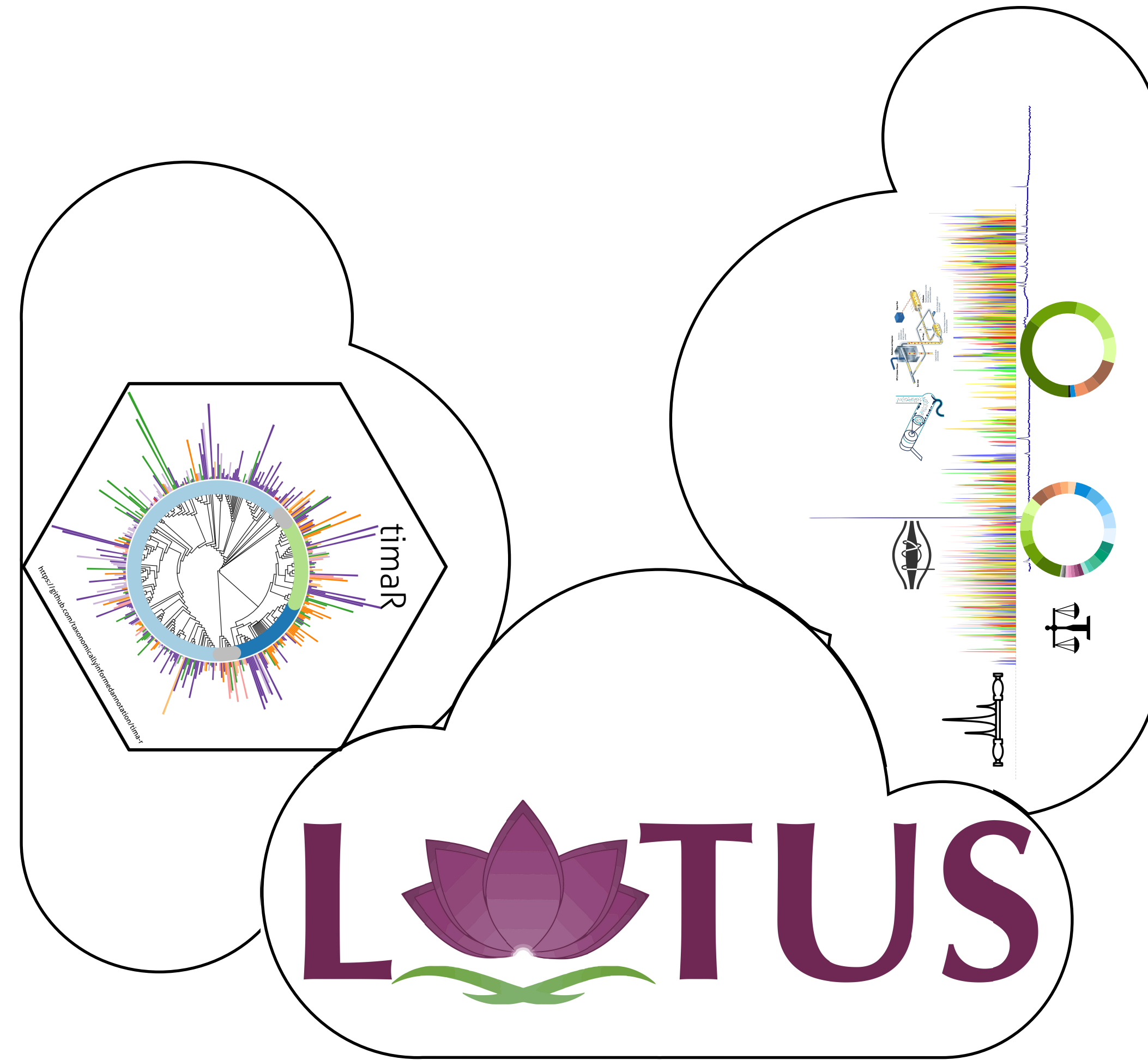
# Swiss knife for metabolomics



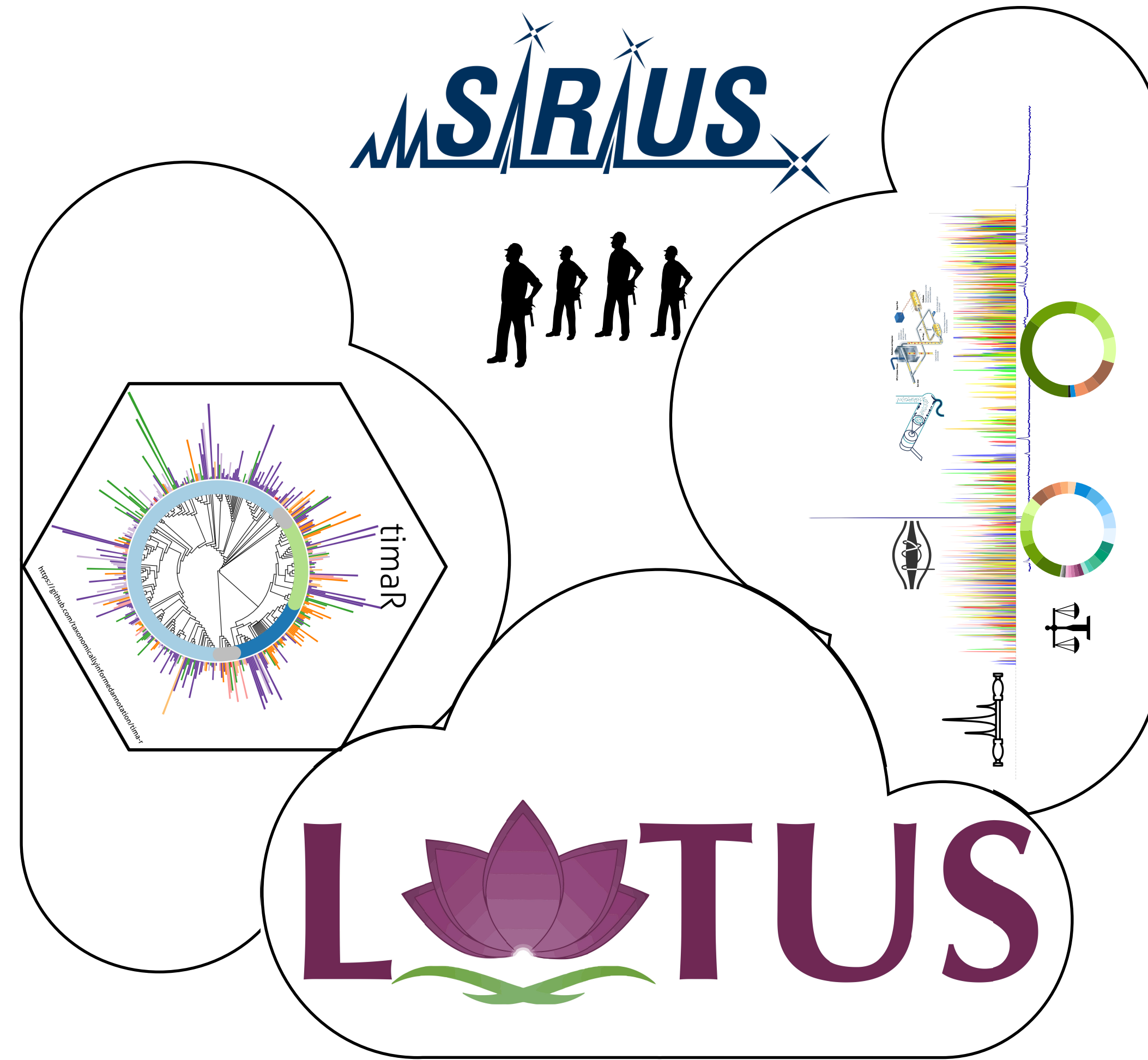
# Swiss knife for metabolomics



# Swiss knife for metabolomics



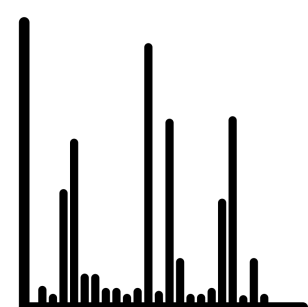
# Swiss knife for metabolomics





# Future research on the chemistry of Life

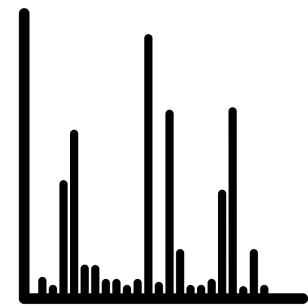
Smarter acquisition at high throughput



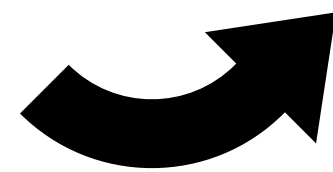
« Hey Wiki, have we  
seen this spectrum  
already? »

# Future research on the chemistry of Life

## Smarter acquisition at high throughput



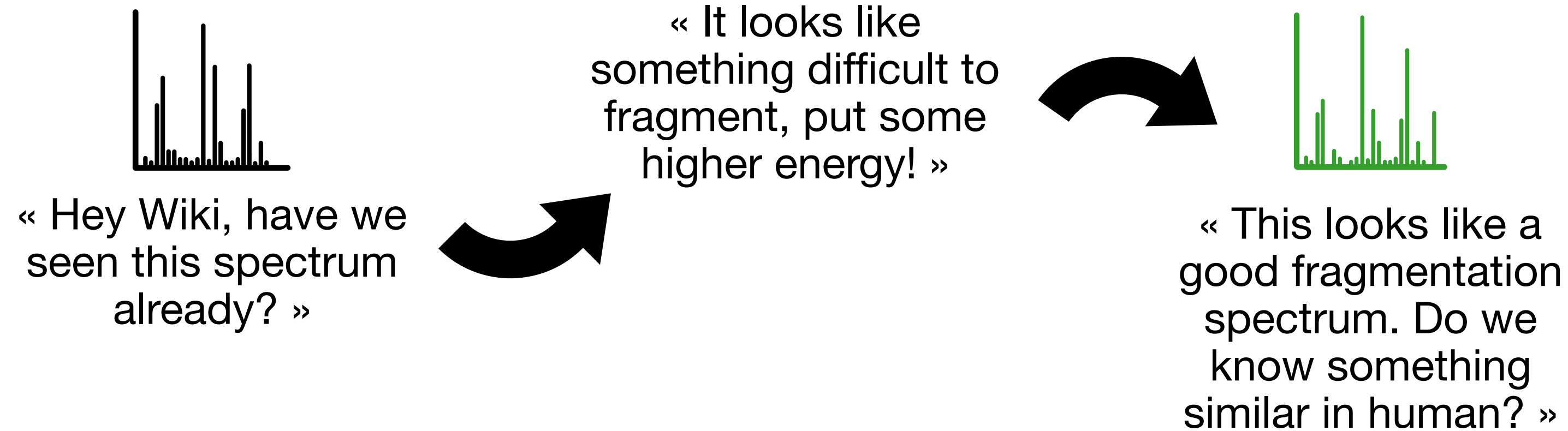
« Hey Wiki, have we seen this spectrum already? »



« It looks like something difficult to fragment, put some higher energy! »

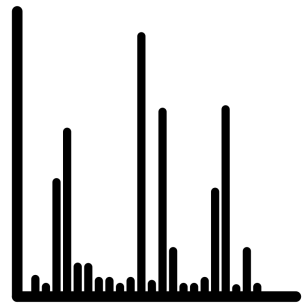
# Future research on the chemistry of Life

## Smarter acquisition at high throughput



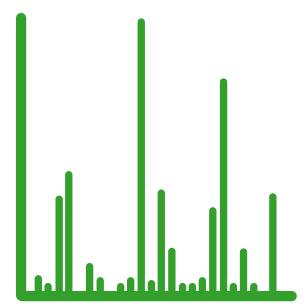
# Future research on the chemistry of Life

## Smarter acquisition at high throughput



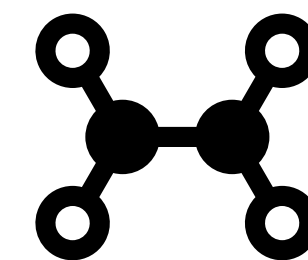
« Hey Wiki, have we seen this spectrum already? »

« It looks like something difficult to fragment, put some higher energy! »

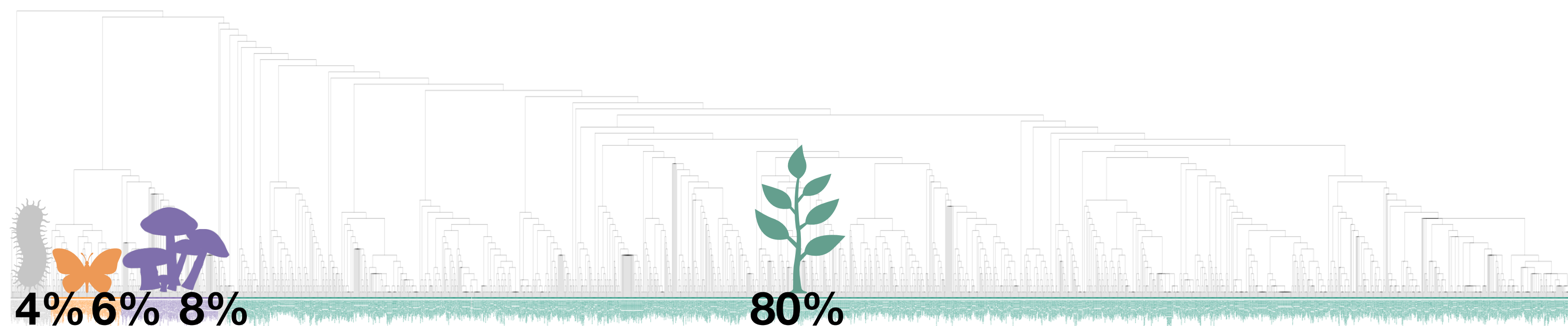


« This looks like a good fragmentation spectrum. Do we know something similar in human? »

« The closest match is a doping agent that was reported to interact with other detected metabolites also present in the sample. Maybe a new analog? »



# Future research on the chemistry of Life

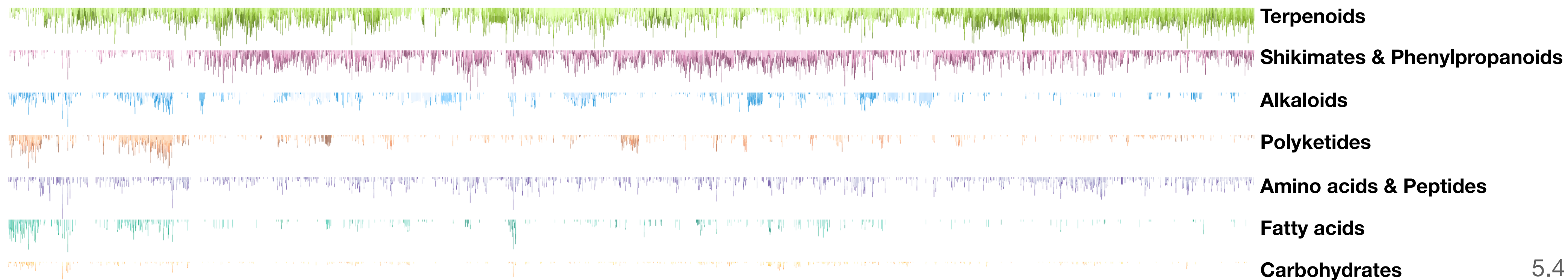


7,848 species  
≥ 25 structures

32,907 species

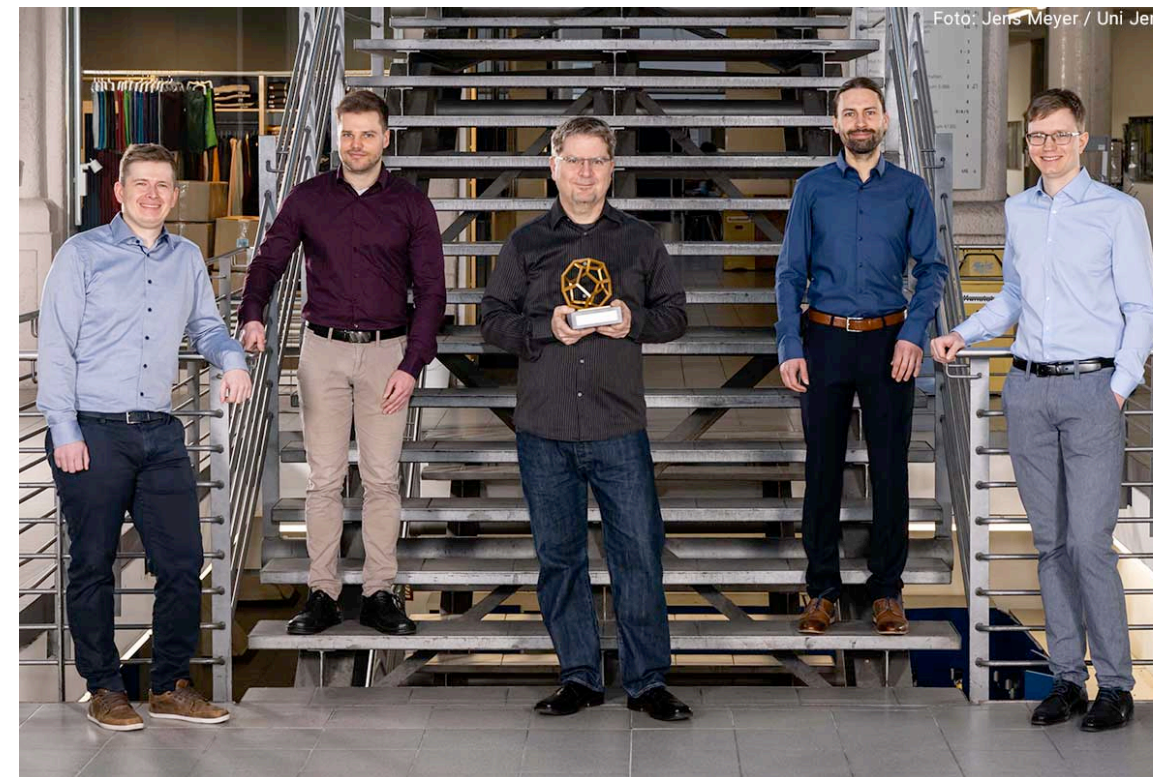
2,956,117 species  
on Wikidata

1% of known biodiversity





# Acknowledgements

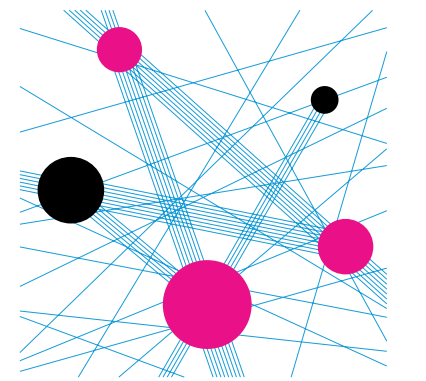


**ETH** zürich  
D BIOL

Institute of Molecular Systems Biology



Prof. Nicola Zamboni



**Contact**



[rutz@imsb.biol.ethz.ch](mailto:rutz@imsb.biol.ethz.ch)