



Towards DAG-based interactive pharmacophore exploration:

Application to the BCR-ABL ligand set

**Johanna GIOVANNINI Etienne LEHEMBRE Jean-Luc LAMOTTE Abdelkader OUALI Alban LEPAILLEUR
Albrecht ZIMMERMANN Bertrand CUISSART Ronan BUREAU and Bruno CREMILLEUX**

Centre d'Etudes et de Recherche sur le Médicament de Normandie (CERMN)
GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072
Université de Caen Normandie, France



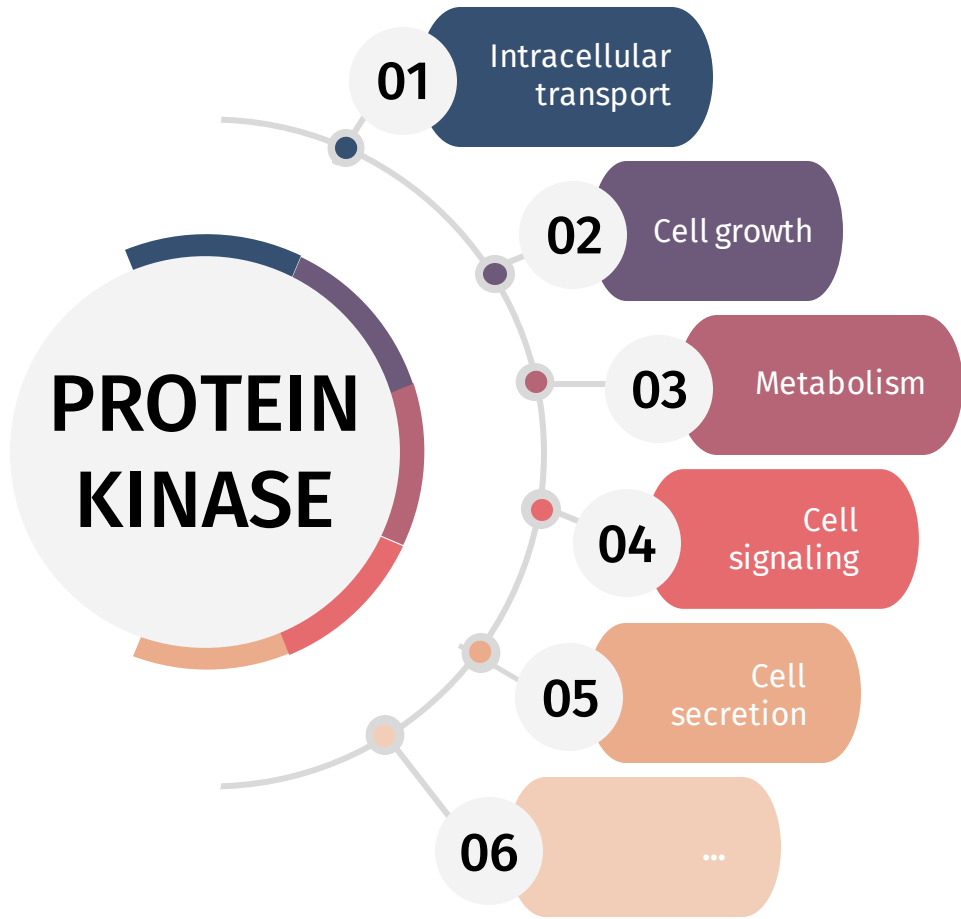
**Towards DAG-based interactive pharmacophore
exploration:
Application to the BCR-ABL ligand set**

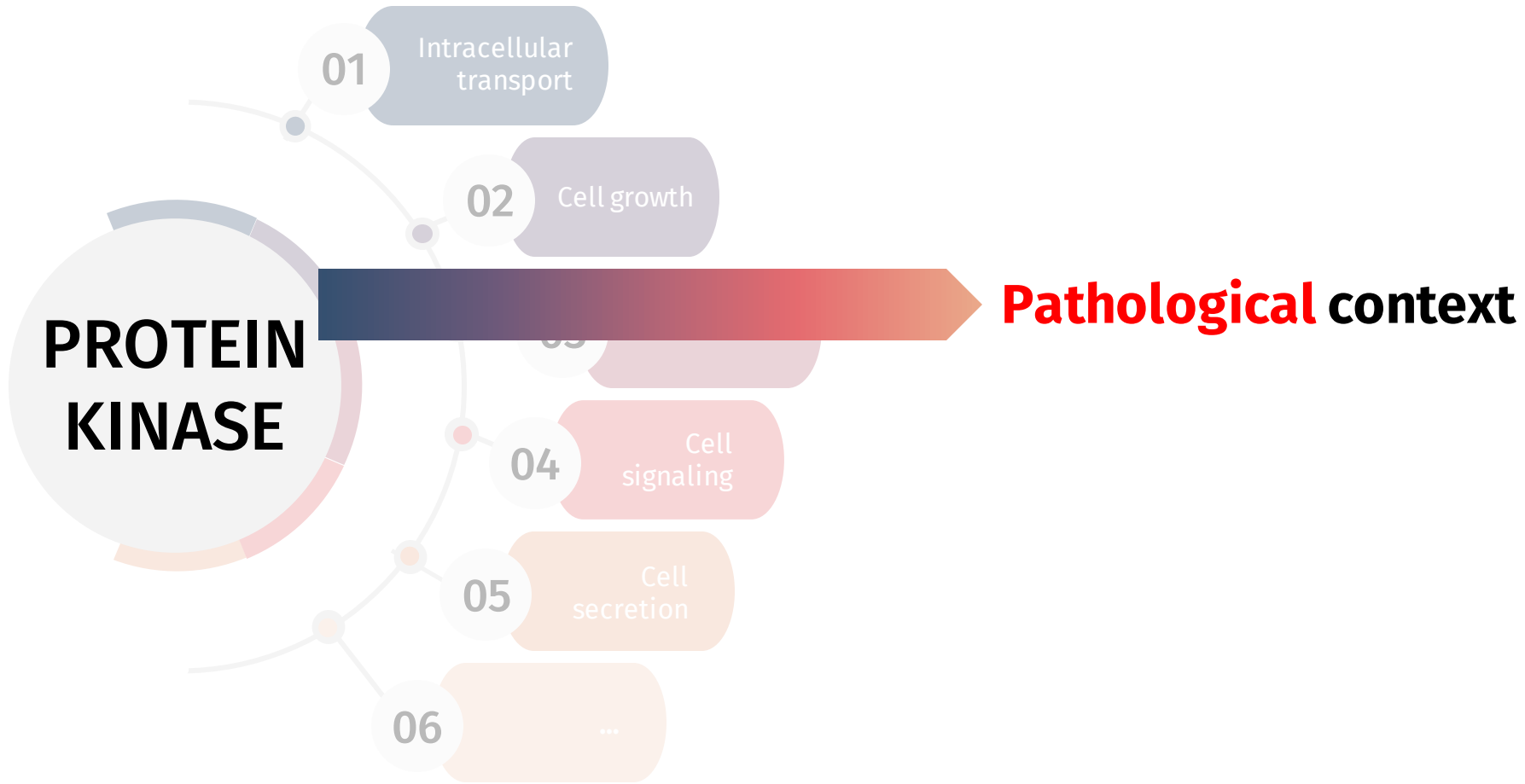


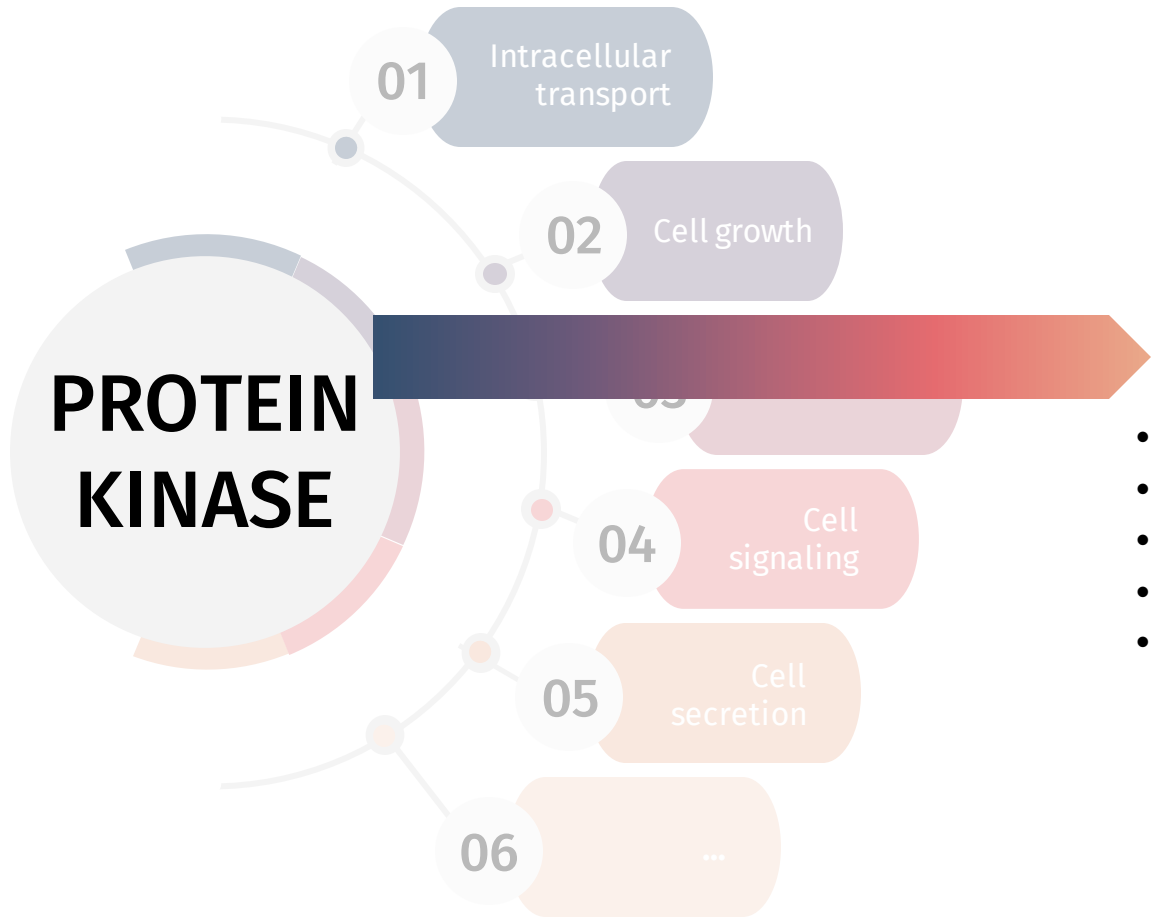
Towards DAG-based interactive pharmacophore exploration: Application to the BCR-ABL ligand set



PROTEIN KINASE



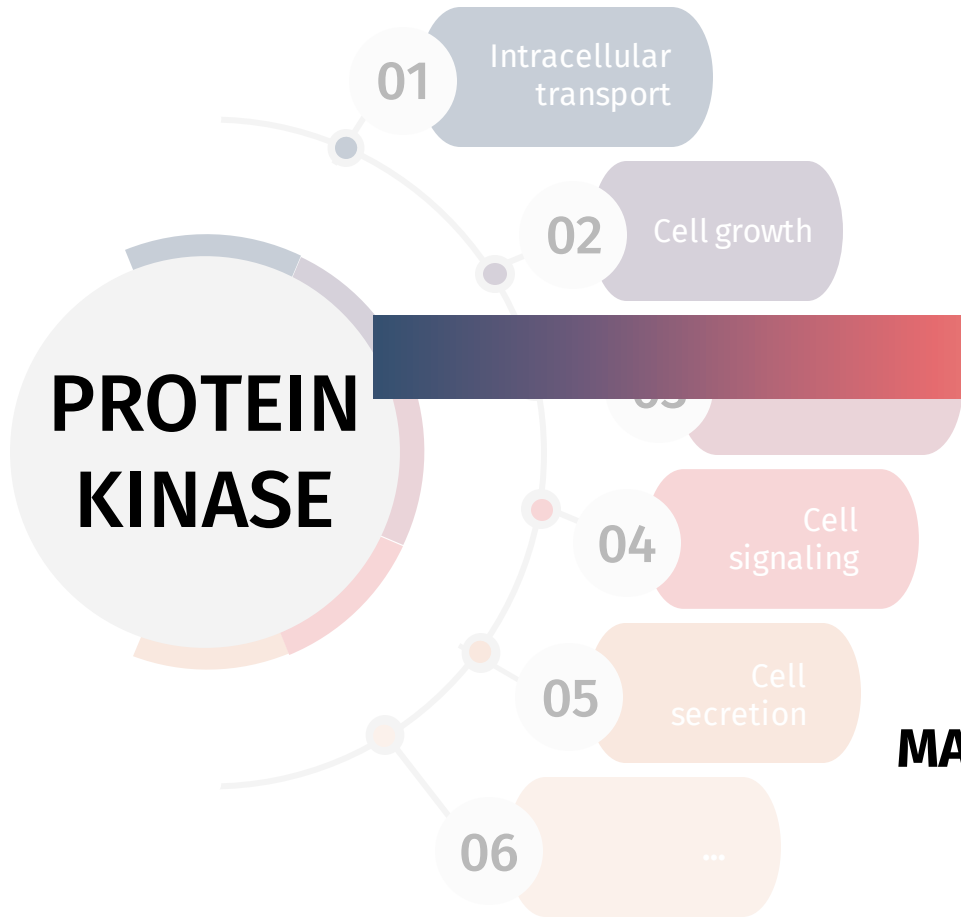




PROTEIN KINASE

Pathological context

- Cancers
- Inflammatory diseases
- Metabolic diseases (diabetes...)
- Autoimmune disorders
- Neurodegenerative diseases
- ...



PROTEIN KINASE



Pathological context

- Cancers
- Inflammatory diseases
- Metabolic diseases (diabetes...)
- Autoimmune disorders
- Neurodegenerative diseases

...

MAJOR Therapeutic interest

for kinase inhibitors

KINOME TREE

PROTEIN KINASE

TK
Tyrosine Kinase

TKL
Tyrosine Kinase-Like

STE
Homologs of the yeast STE7, STE11 and STE20 genes

CK1
Cell Kinase 1

AGC
Protein Kinase A, G, and C families

CAMK
Calmodulin/Calcium regulated kinases

CMGC
CMGC group (named after the initials of some members)

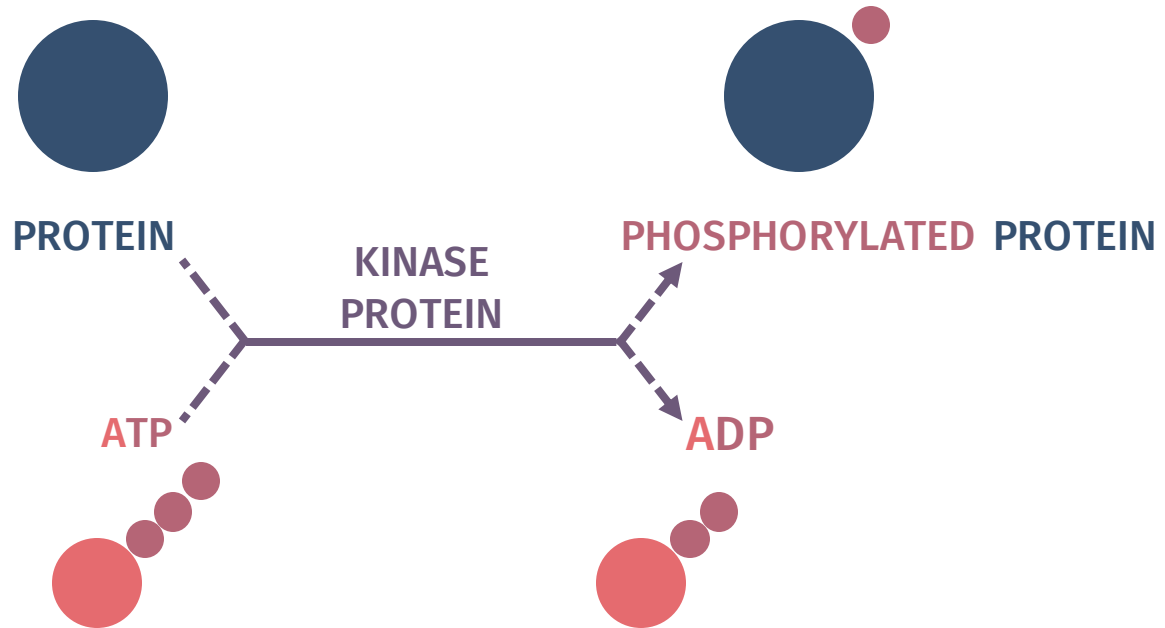
CMGC

Representation made with CORAL web application ; Metz *et al.*, Cell Systems, 2018, 7 (3) ; doi.org/10.1016/j.cels.2018.07.001

Mechanism of action

Phosphorylation of proteins on hydroxyl groups (OH)
→ Protein Activation / Deactivation via a phosphate group transfer

**PROTEIN
KINASE**



PKI: 6 different types

PROTEIN KINASE INHIBITORS (PKI)

PKI – TYPE I

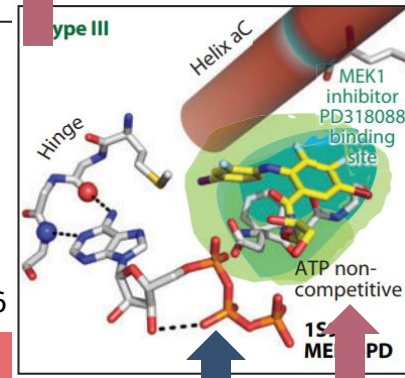
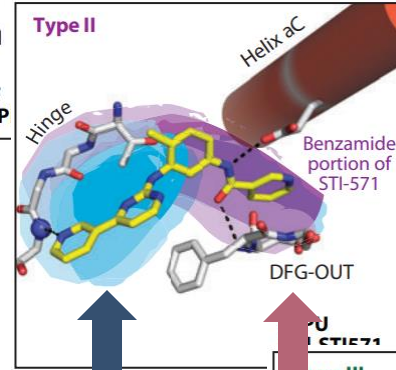
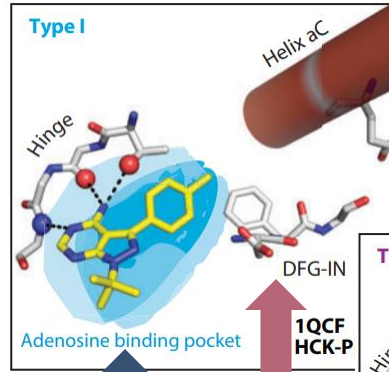
- ATP competitive inhibition
- DFG *in*

PKI – TYPE II

- ATP competitive inhibition
- DFG *out*

PKI – TYPE III

- Allosteric inhibitors
- Binding site adjacent to ATP binding site



10.1146/annurev-biochem-090308-173656

PKI: 6 different types

PROTEIN KINASE INHIBITORS (PKI)

PKI – TYPE I

- ATP competitive inhibition
- DFG *in*

PKI – TYPE IV

- Allosteric inhibitors
- Various sites of action

PKI – TYPE II

- ATP competitive inhibition
- DFG *out*

PKI – TYPE V

- Bivalent inhibitors
- ATP binding site + adjacent pocket

PKI – TYPE III

- Allosteric inhibitors
- Binding site adjacent to ATP binding site

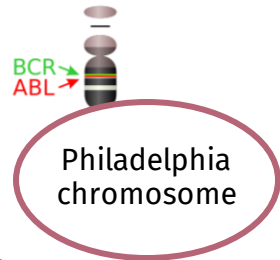
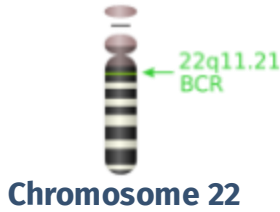
PKI – TYPE VI

- Covalent binding with the protein (CYS, LYS or TYR)



BCR-ABL protein

BCR-ABL protein



BCR gene

BCR =
Breakpoint
Cluster Region

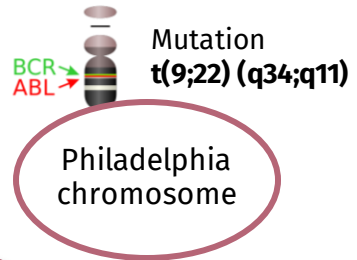
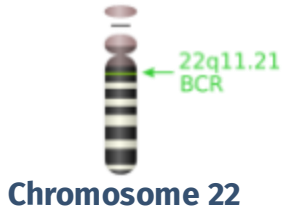
BCR-ABL1 gene

Fusion protein

ABL1 gene

ABL =
Abelson

BCR-ABL protein



BCR gene

BCR-ABL1 gene

ABL1 gene

BCR =
Breakpoint Cluster Region

Fusion protein

- intracellular protein
- deregulated tyrosine kinase activity
- stimulates blood cell division

CHRONIC MYELOID LEUKAEMIA (CML)

ABL =
Abelson



Towards DAG-based interactive pharmacophore exploration: Application to the BCR-ABL ligand set

NORNS

Métivier *et al.*, J Med Chem, 2018, 61

NORNS

PHARMACOPHORE EXPLORATION AND EVALUATION

Automatically computes pharmacophores from a large data set of molecules without any supervised selection of molecules

NORNS

PHARMACOPHORE EXPLORATION AND EVALUATION

Automatically computes pharmacophores from a large data set of molecules without any supervised selection of molecules

SCREENING OF MOLECULAR DATABASES

Query based on 2D pharmacophores

SEARCH FOR MULTI-ACTIVE MOLECULES

Discriminative capability through Emerging Pattern calculation

DEFINITION OF PHARMACOPHORE SPACE

Able to identify features occurring with high or low frequencies

NORNS PROCESS

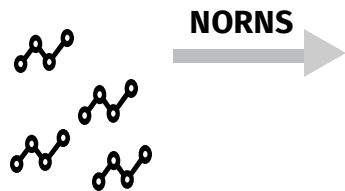
Métivier *et al.*, J Med Chem, 2018, 61

Pharmacophores



NORNS PROCESS

A

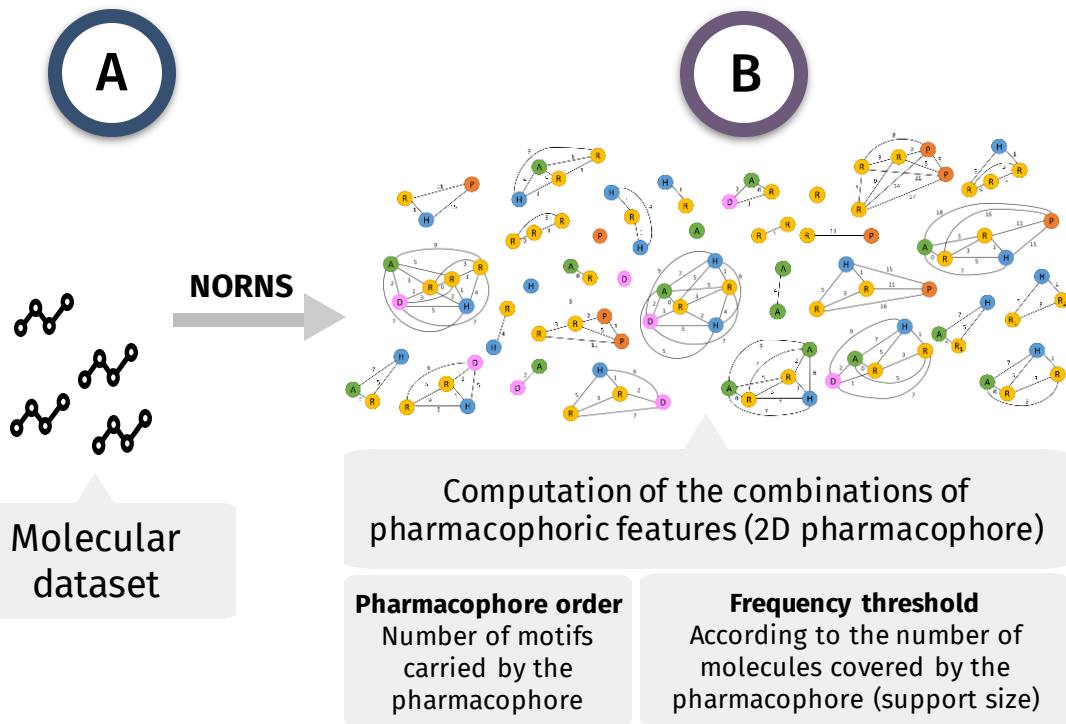


Molecular
dataset

6 essential pharmacophoric motifs

- H** Hydrophobic group
- R** Aromatic cycle
- D** Hydrogen-bond donor
- A** Hydrogen-bond acceptor
- N** Negatively-ionizable function
- P** Positively-ionizable function

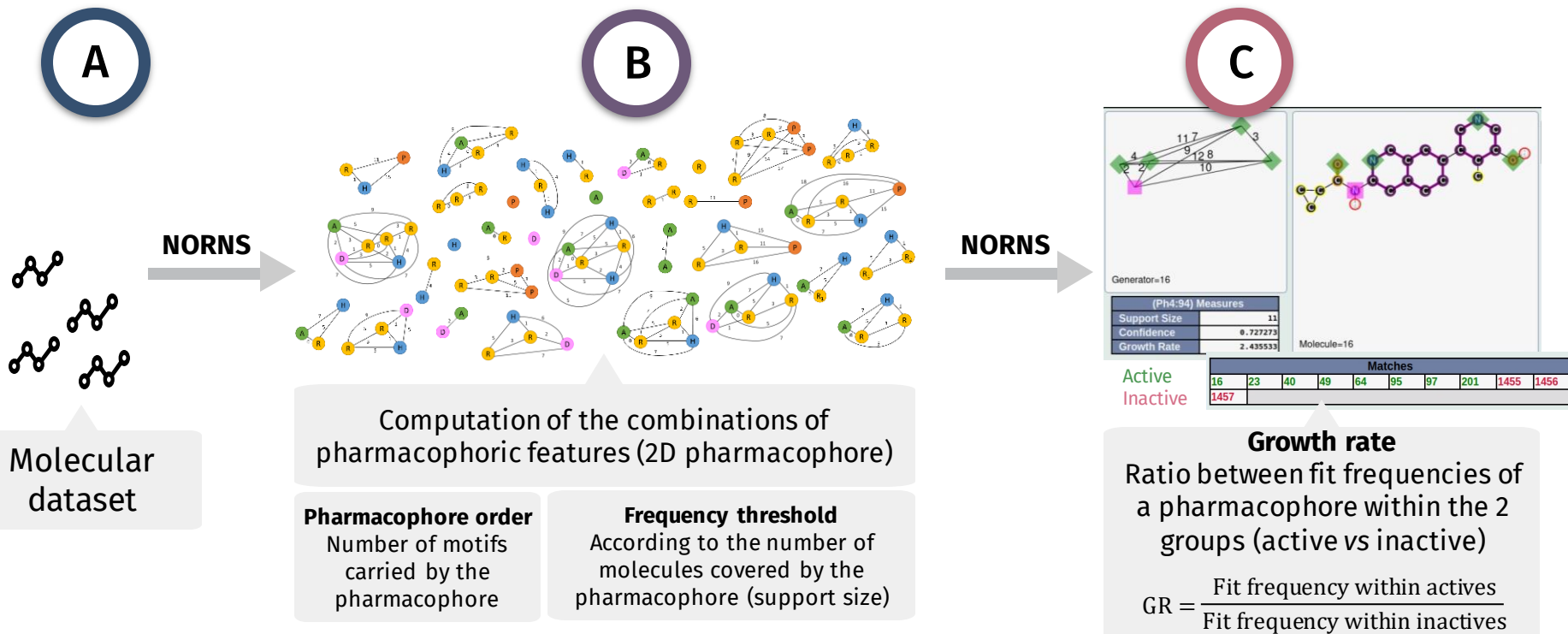
NORNS PROCESS



6 essential pharmacophoric motifs

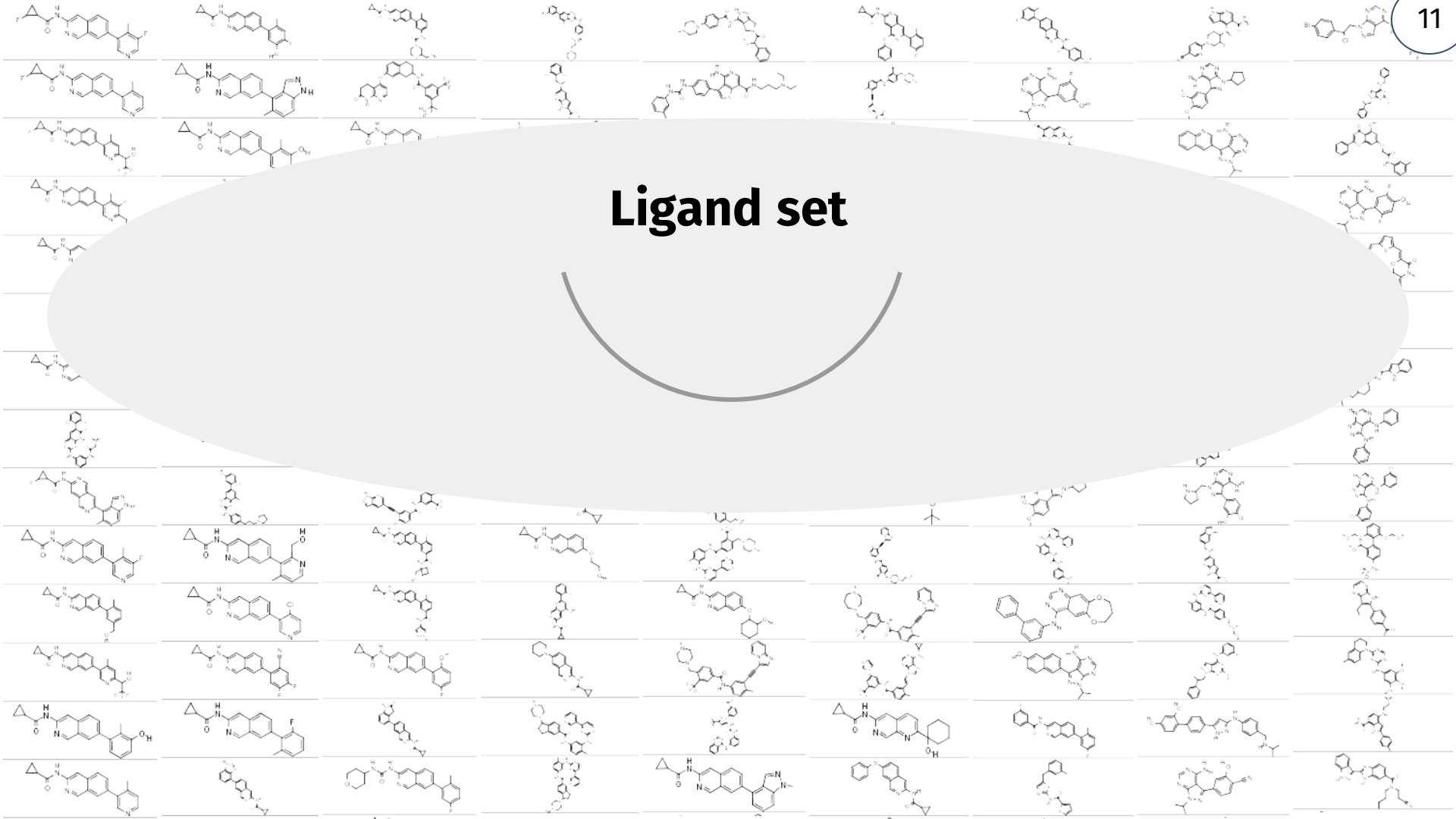
- H** Hydrophobic group
- R** Aromatic cycle
- D** Hydrogen-bond donor
- A** Hydrogen-bond acceptor
- N** Negatively-ionizable function
- P** Positively-ionizable function

NORNS PROCESS




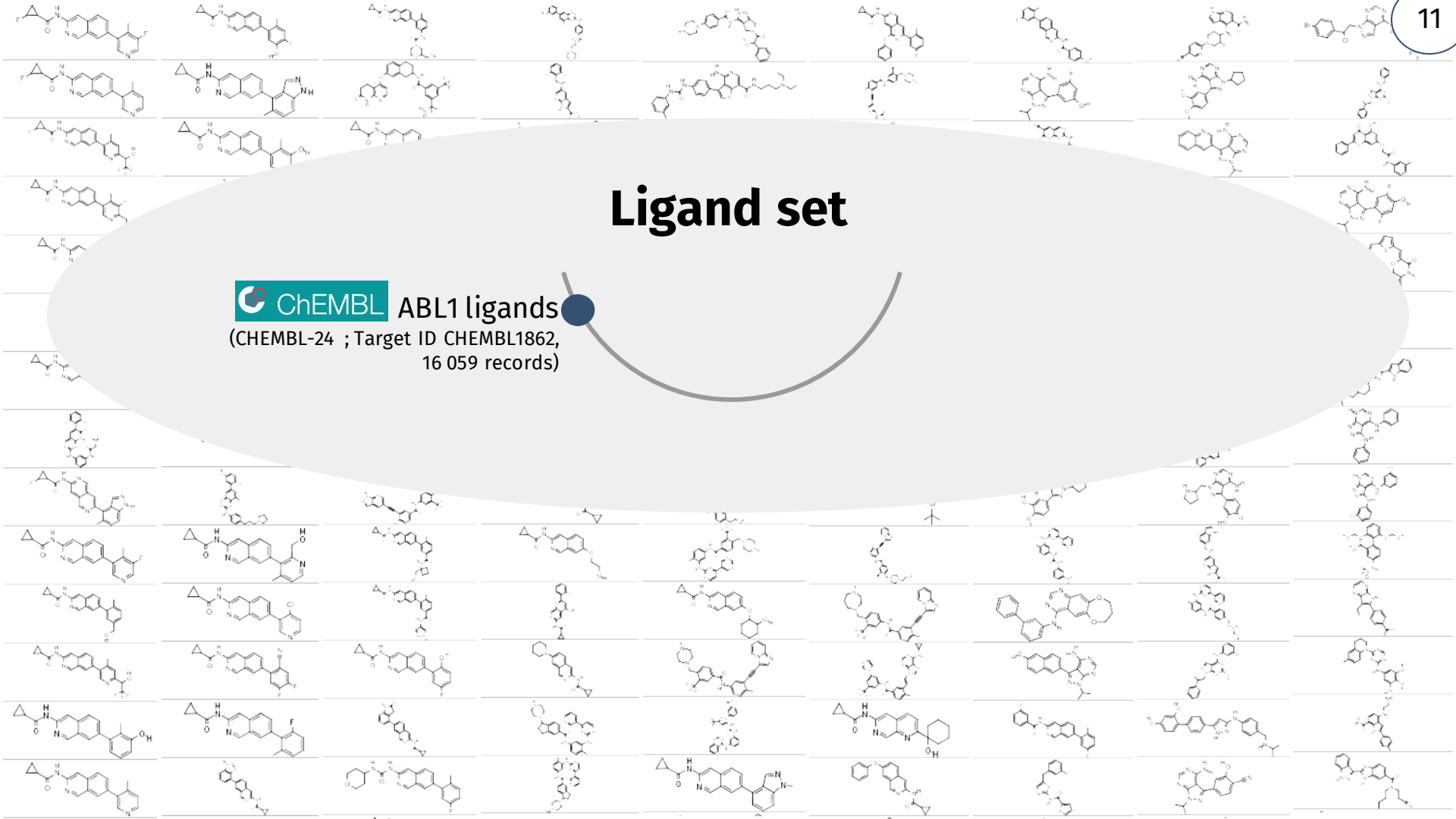
Métivier et al., J Med Chem, 2018, 61

Ligand set



Ligand set

 **ChEMBL ABL1 ligands**
(CHEMBL-24 ; Target ID CHEMBL1862,
16 059 records)



Ligand set



ChEMBL ABL1 ligands
(ChEMBL-24 ; Target ID CHEMBL1862,
16 059 records)

MW \leq 800 g/mol
(15 255 records)

Ligand set



ChEMBL ABL1 ligands
(ChEMBL-24 ; Target ID CHEMBL1862,
16 059 records)

MW \leq 800 g/mol
(15 255 records)

K_i or IC_{50}

Ligand set



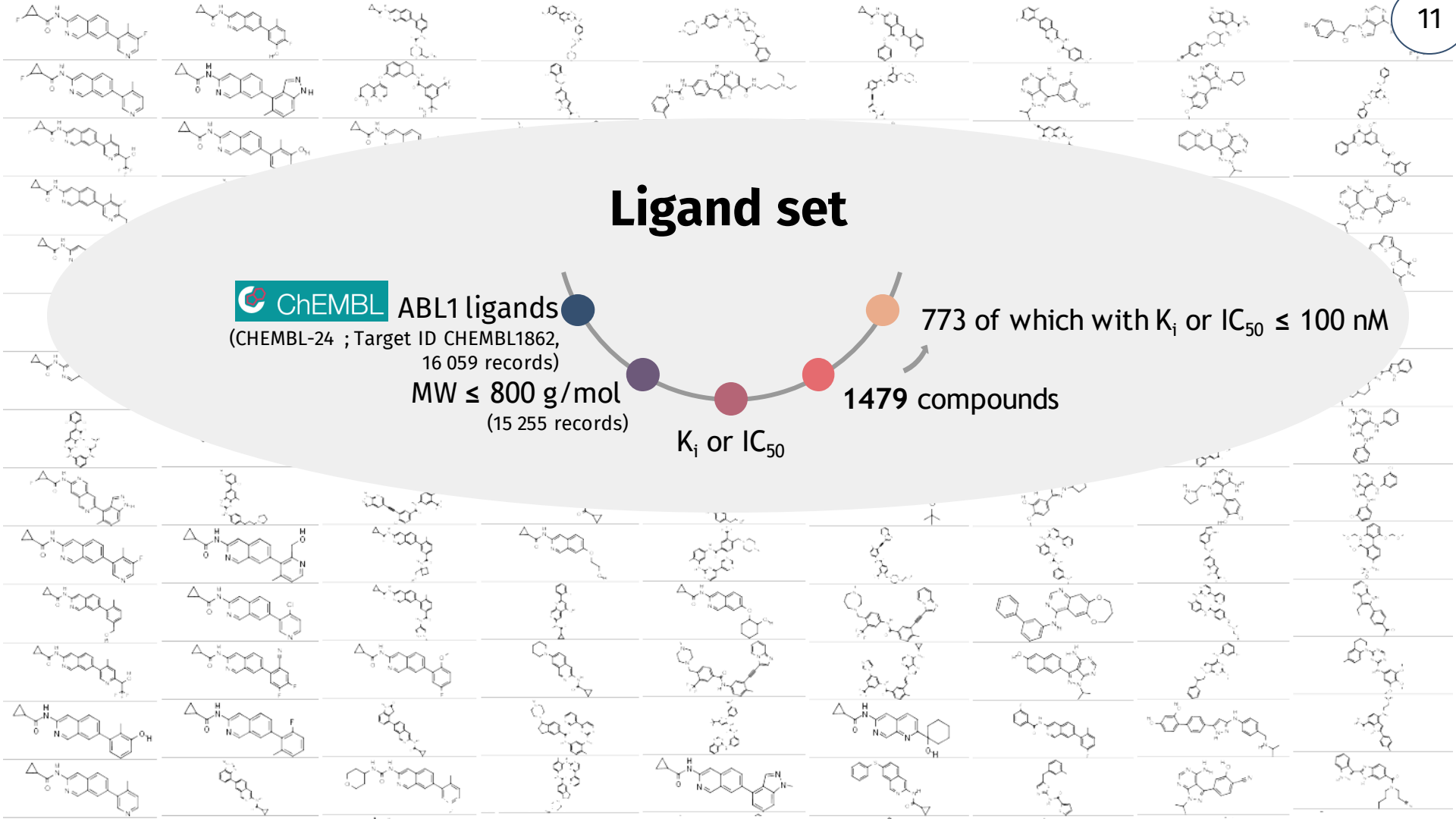
ABL1 ligands
(CHEMBL-24 ; Target ID CHEMBL1862,
16 059 records)

MW ≤ 800 g/mol
(15 255 records)

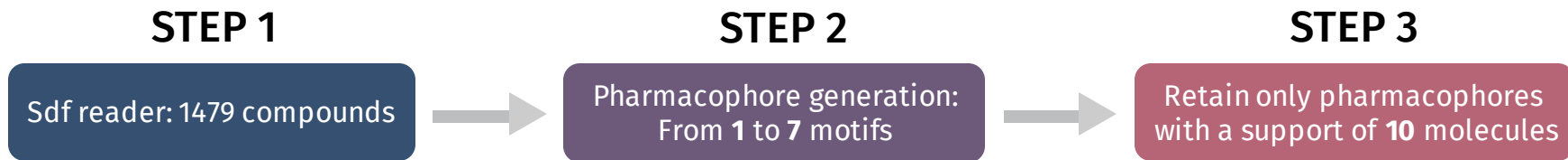
K_i or IC₅₀

1479 compounds

773 of which with K_i or IC₅₀ ≤ 100 nM

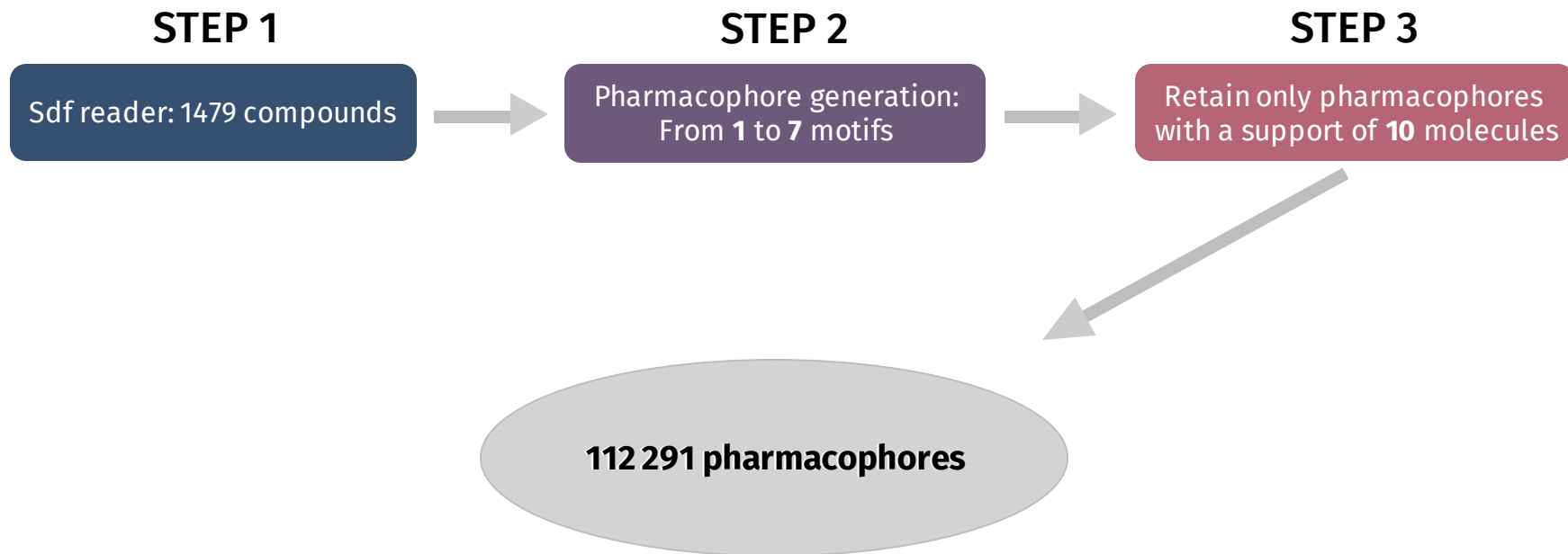


NORNS pipeline



Pharmacophores

NORNS pipeline



112 291 pharmacophores

Directed Acyclic Graph (DAG)



PHARMACOPHORE NETWORK

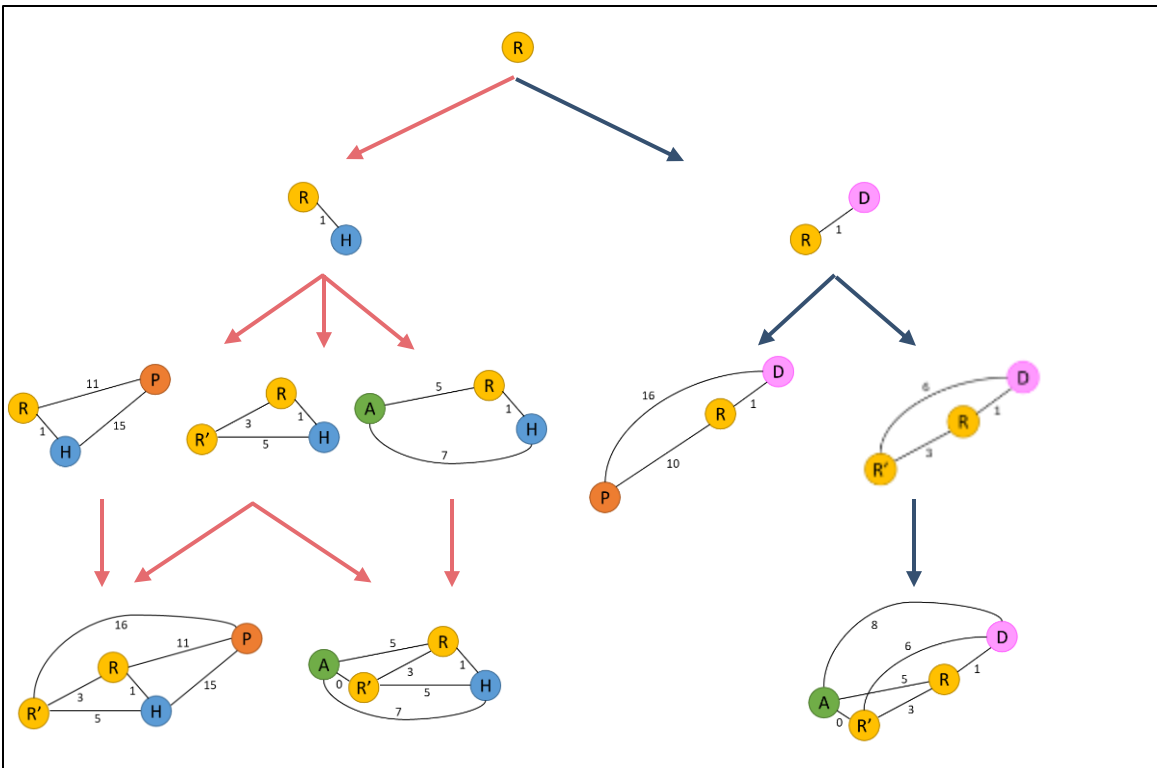


112 291 pharmacophores

Directed **A**cyclic **G**raph (DAG)

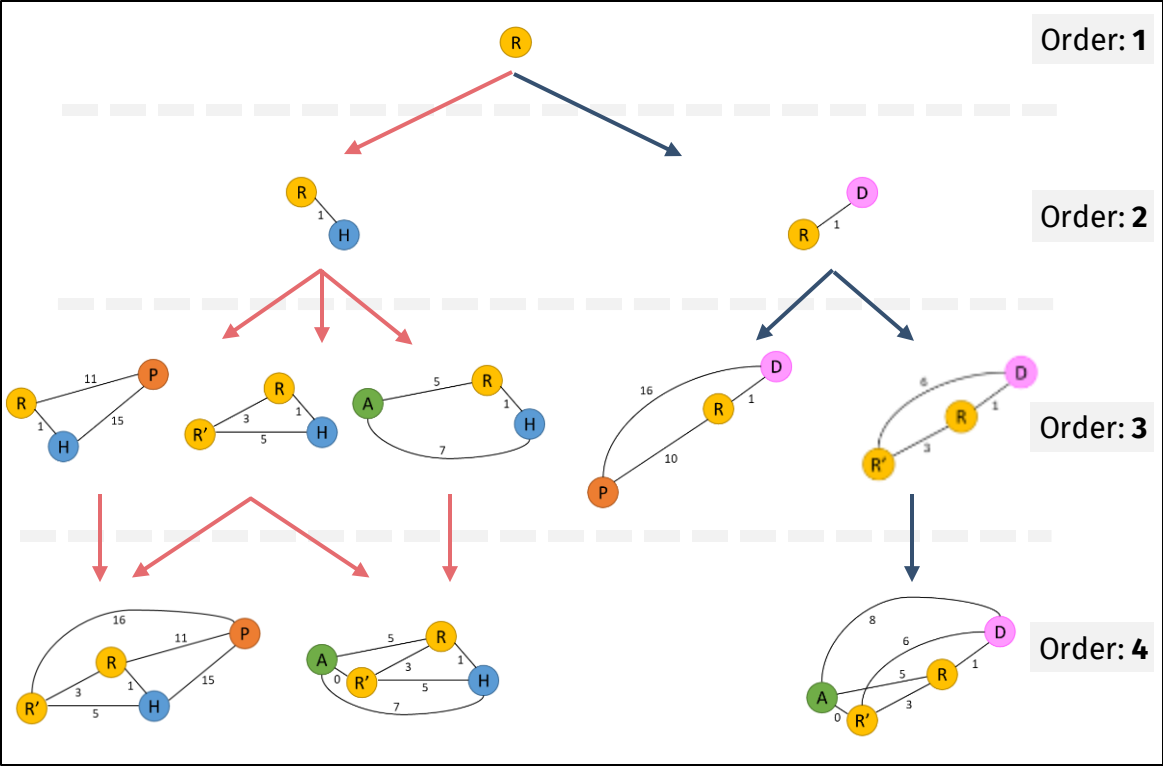
Directed Acyclic Graph (DAG)

- H Hydrophobic group
- R Aromatic ring
- A Hydrogen-bond acceptor
- D Hydrogen-bond donor
- P Positively-ionizable group



- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group

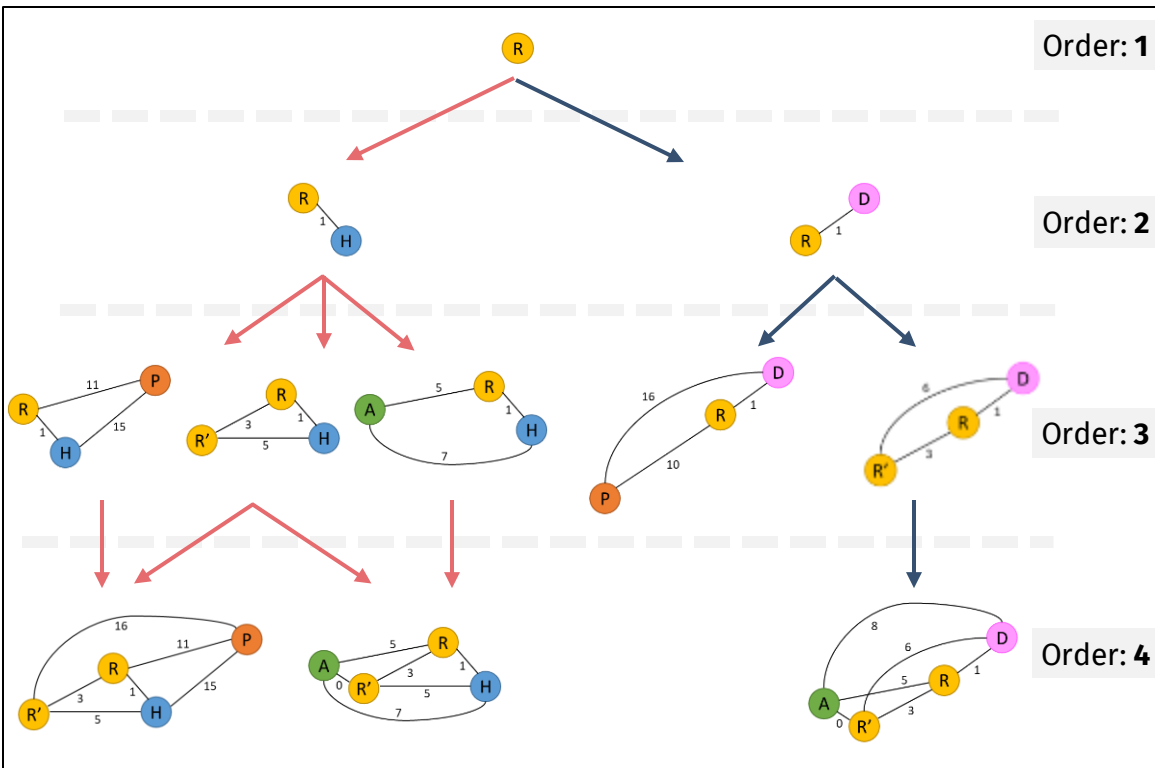
Directed Acyclic Graph (DAG)



From top to bottom:
Two pharmacophores are linked if one is included in the other.

Directed Acyclic Graph (DAG)

- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



From top to bottom:

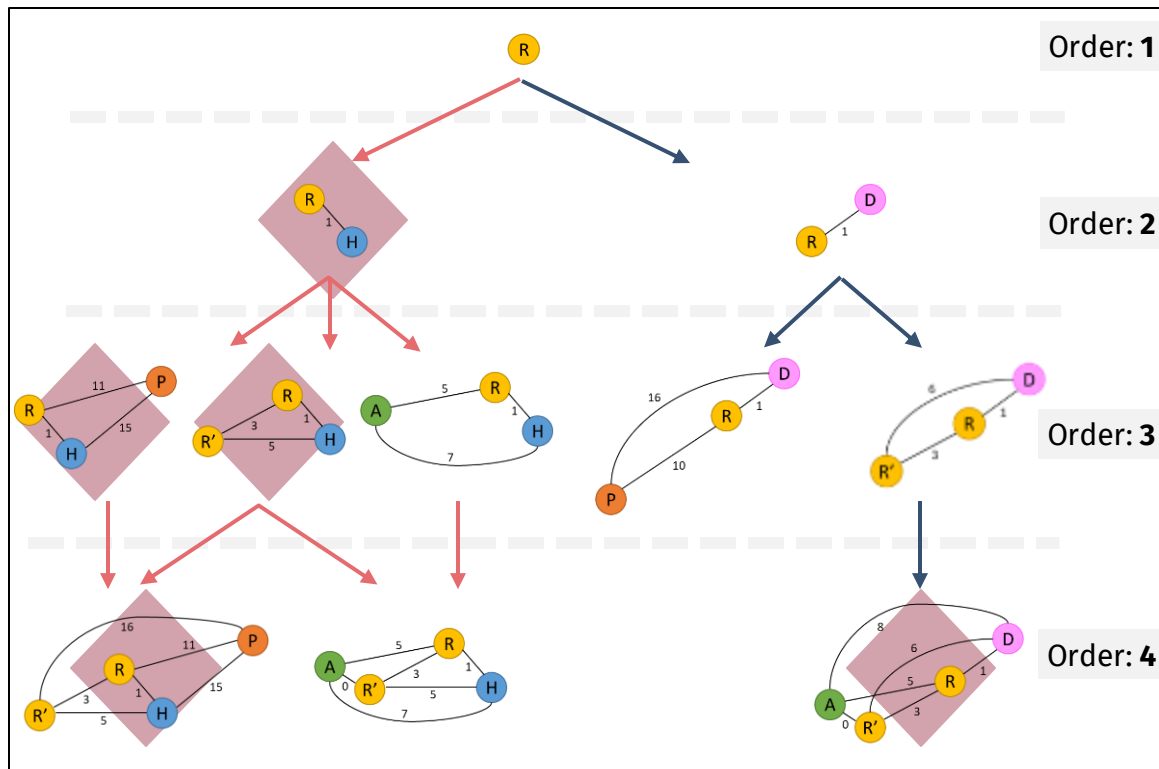
Two pharmacophores are linked if one is included in the other.

Considering molecule subsets:

Distinct pharmacophores covering the exact same molecule subset form a General Equivalent Class (GEC).

Directed Acyclic Graph (DAG)

- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



From top to bottom:

Two pharmacophores are linked if one is included in the other.

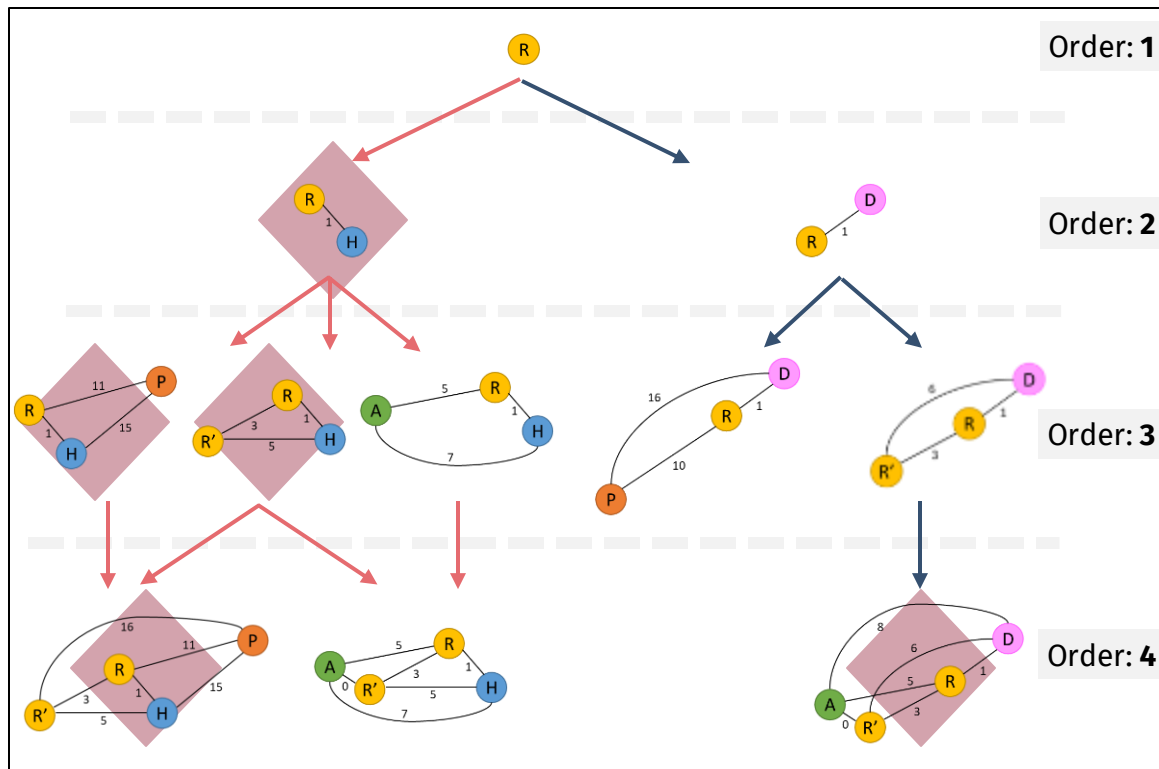
Considering molecule subsets:

Distinct pharmacophores covering the exact same molecule subset form a General Equivalent Class (GEC).

Molecule subset 1

Directed Acyclic Graph (DAG)

- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



From top to bottom:

Two pharmacophores are linked if one is included in the other.

Considering molecule subsets:

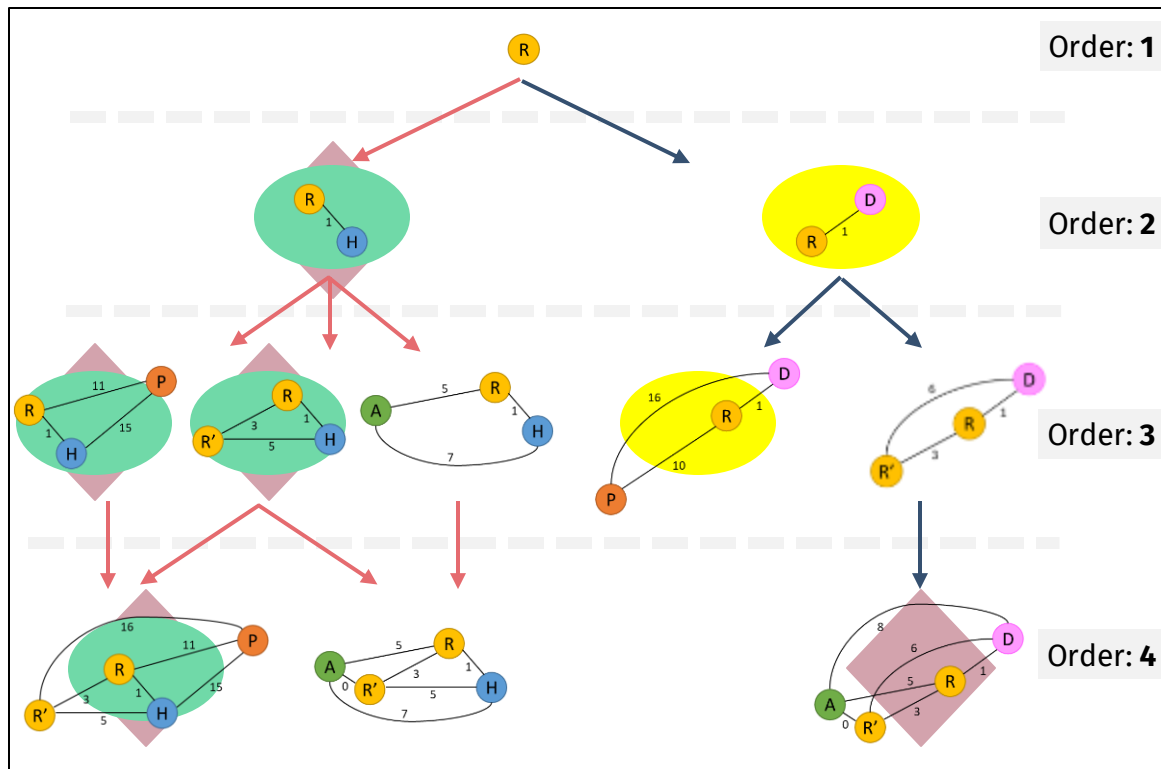
Distinct pharmacophores covering the exact same molecule subset form a **General Equivalent Class (GEC).**

Distinct pharmacophores covering the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC).**

Molecule
subset 1

Directed Acyclic Graph (DAG)

- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



From top to bottom:

Two pharmacophores are linked if one is included in the other.

Considering molecule subsets:

Distinct pharmacophores covering the exact same molecule subset form a **General Equivalent Class (GEC).**

Distinct pharmacophores covering the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC).**

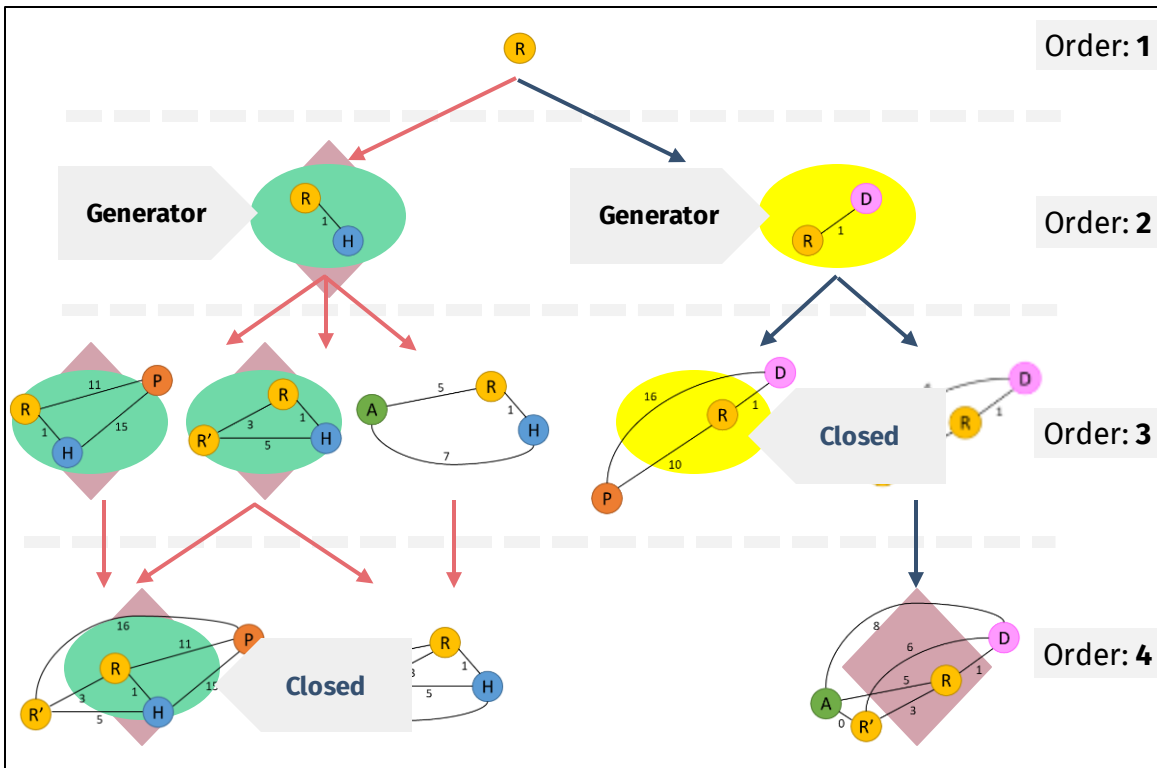
Molecule subset 1

Molecule subset 2

Molecule subset 3

Directed Acyclic Graph (DAG)

- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



⇒ Consideration of the smallest pharmacophoric description :

Generator

⇒ Consideration of the greatest pharmacophoric description :

Closed

Distinct pharmacophores covering the exact same molecule subset and sharing a family relationship form a Structured Equivalent Class (SEC).

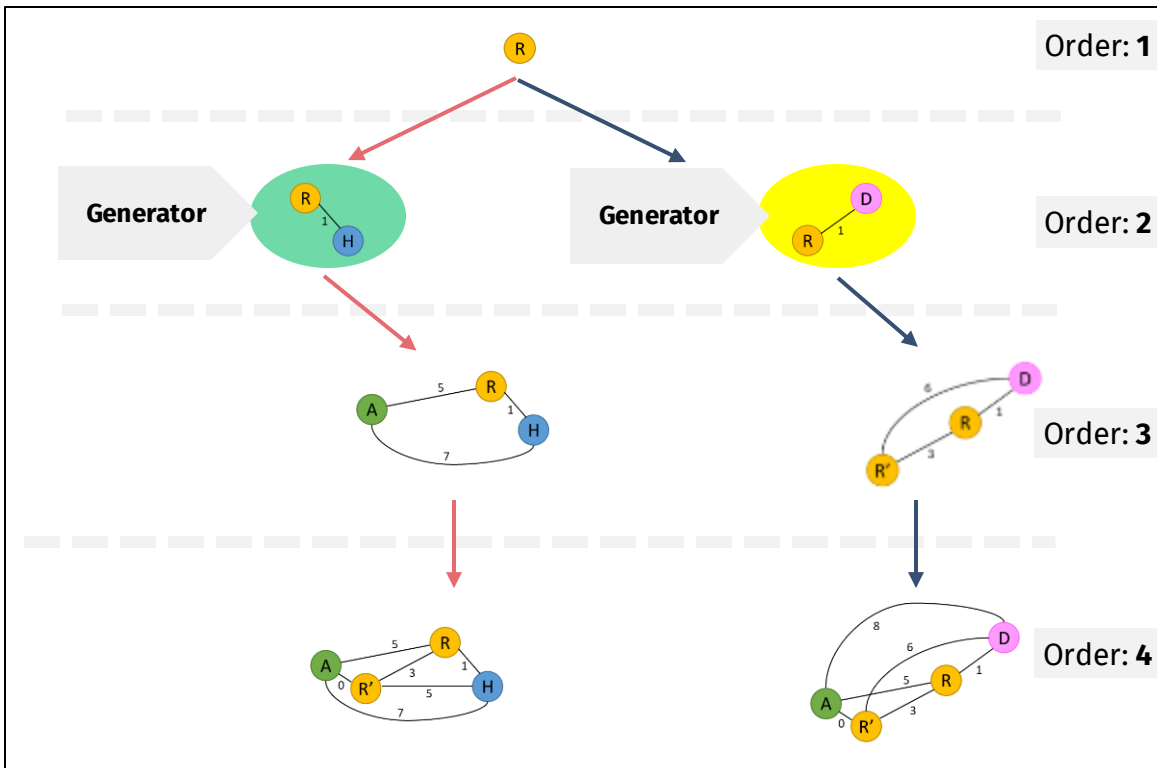
Molecule subset 1

Molecule subset 2

Molecule subset 3

Directed Acyclic Graph (DAG)

- H Hydrophobic group
- R Aromatic ring
- A Hydrogen-bond acceptor
- D Hydrogen-bond donor
- P Positively-ionizable group



⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

⇒ Consideration of the **smallest pharmacophoric description** :

Generator

⇒ Consideration of the **greatest pharmacophoric description** :
Closed

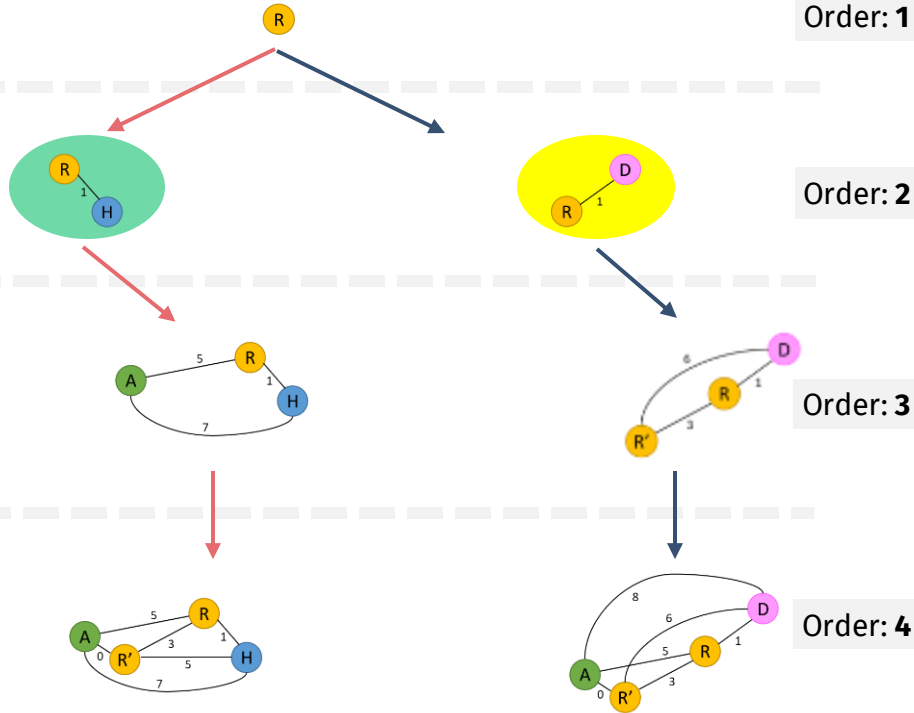
Distinct pharmacophores covering the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC)**.

Molecule
subset 2

Molecule
subset 3

Directed Acyclic Graph (DAG)

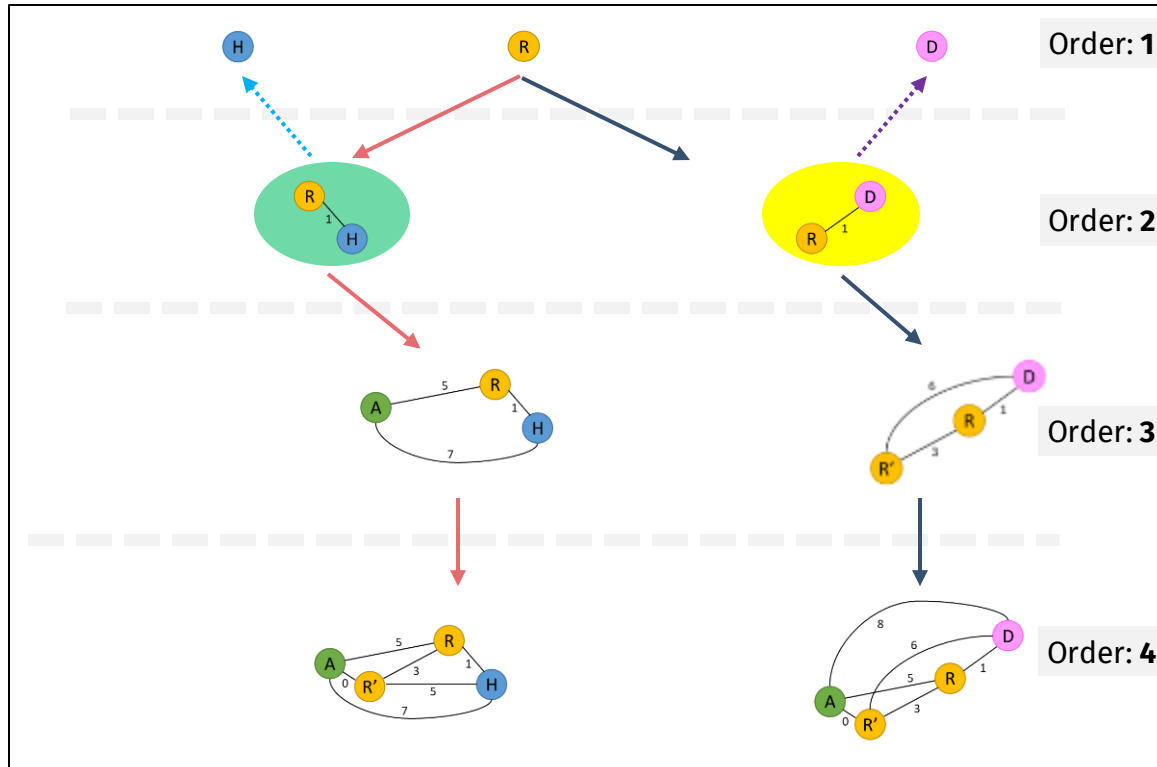
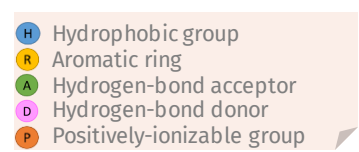
- Hydrophobic group
- Aromatic ring
- Hydrogen-bond acceptor
- Hydrogen-bond donor
- Positively-ionizable group



Siblings concept

⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

Directed Acyclic Graph (DAG)

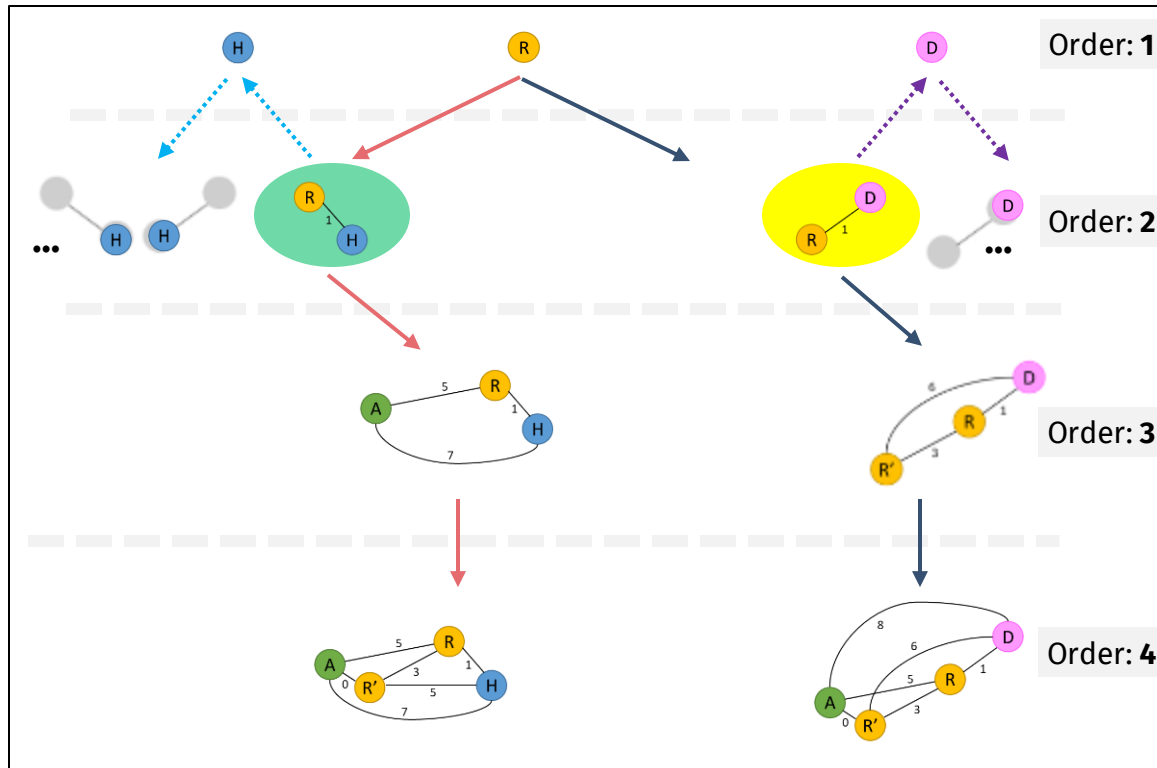
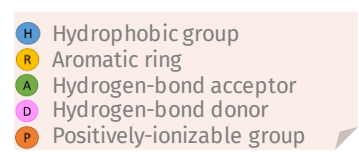


Siblings concept

Consideration of pharmacophore ancestors (parents)

⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

Directed Acyclic Graph (DAG)



Siblings concept

Consideration of pharmacophore ancestors (parents)



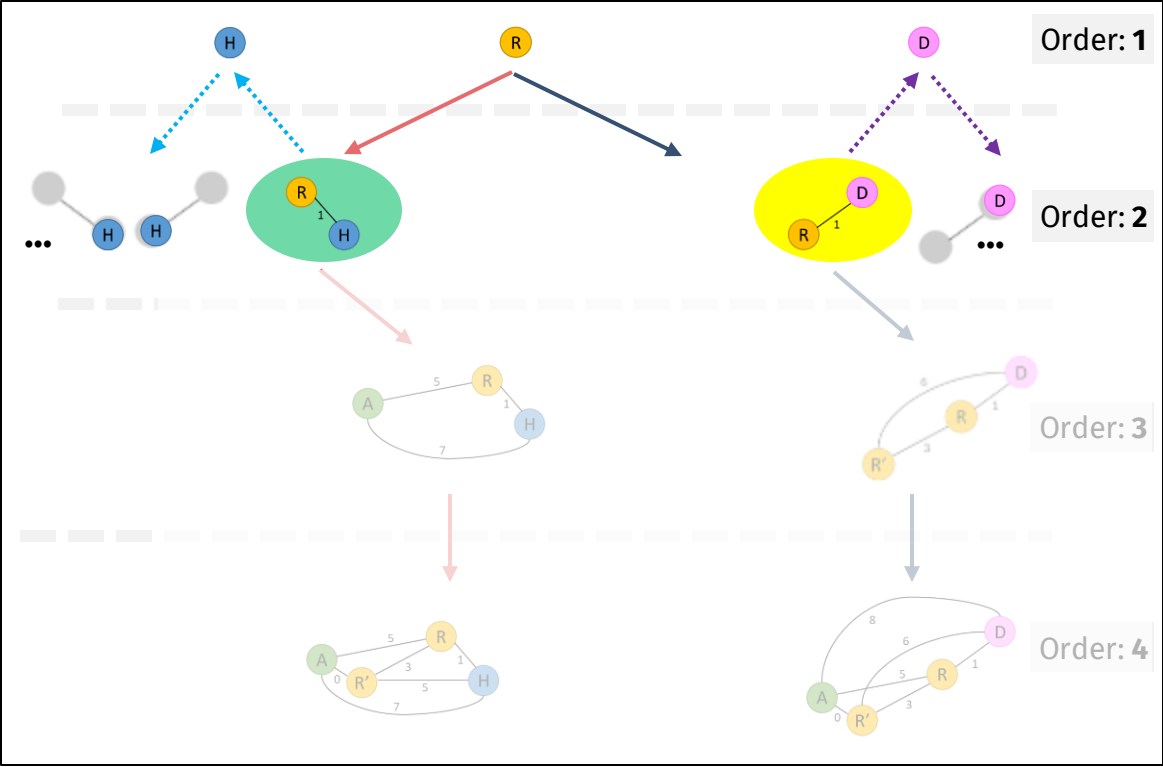
Listing of all pharmacophore successors (children)

⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

- H Hydrophobic group
- R Aromatic ring
- A Hydrogen-bond acceptor
- D Hydrogen-bond donor
- P Positively-ionizable group

SEC = **Structured Equivalent Class**

Directed **A**cyclic **G**raph (DAG)



Siblings concept
 Consideration of pharmacophore ancestors (parents)
 ↓
 Listing of all pharmacophore successors (children)

Same order

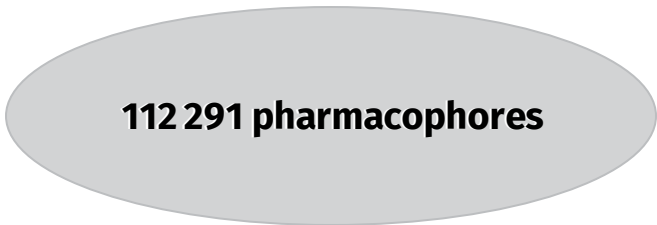
SEC

⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

Directed Acyclic Graph (DAG)



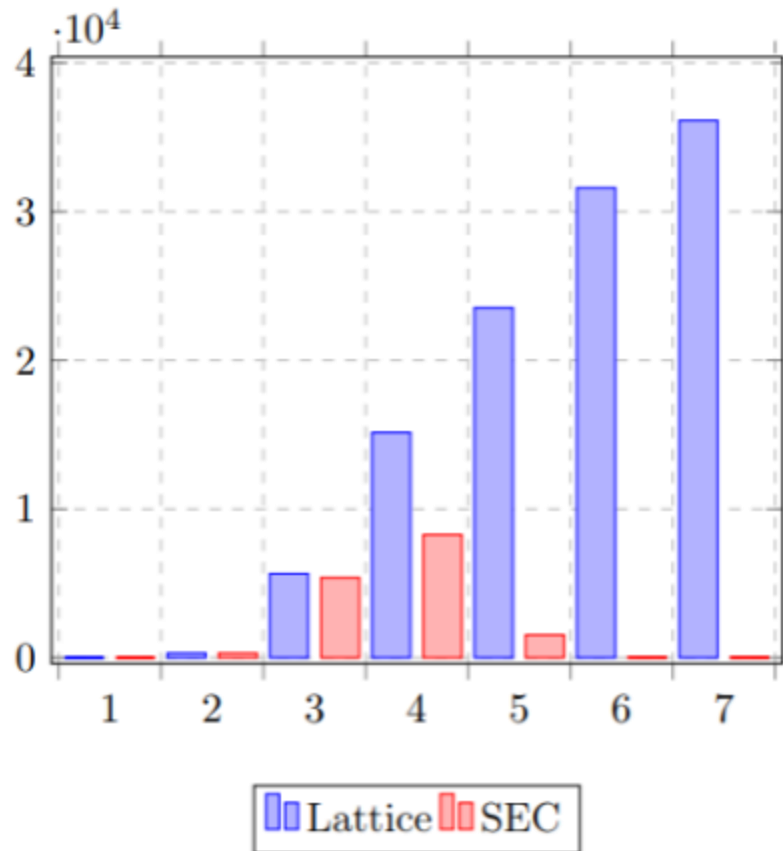
PHARMACOPHORE NETWORK



112 291 pharmacophores

Directed Acyclic Graph (DAG)

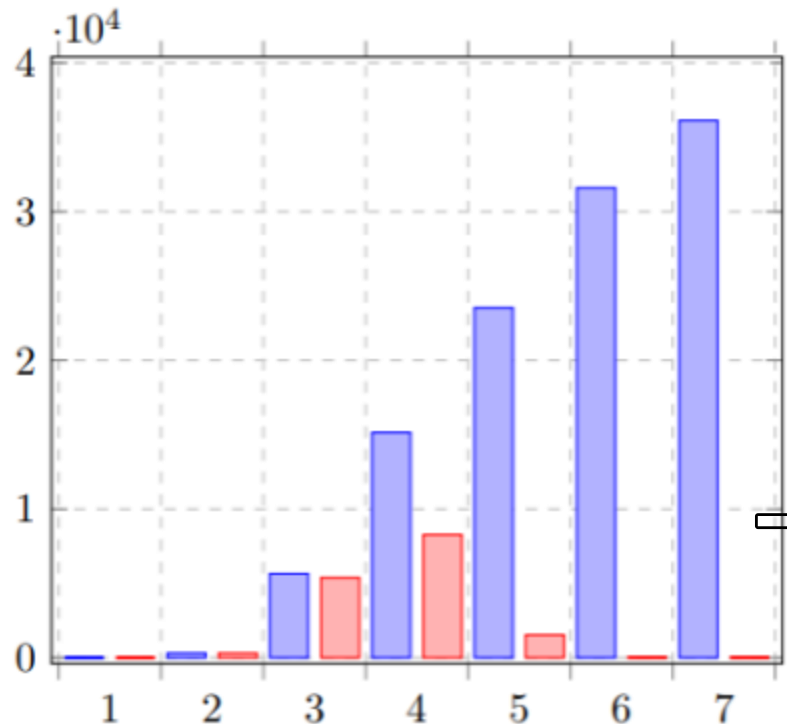
NORNS
LATTICE



Distinct pharmacophores bearing the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC)**.

Directed Acyclic Graph (DAG)

NORNS
LATTICE



Distinct pharmacophores bearing the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC)**.

112 291
pharmacophores



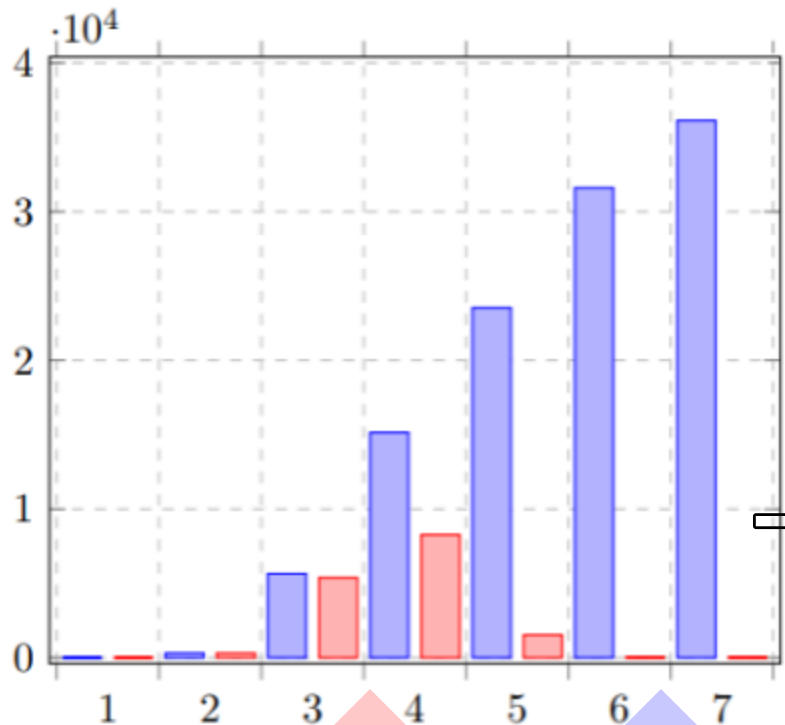
15 477 SEC
(generators)

⇒ **Decreasing the number of pharmacophore to assess**

Lattice SEC

Directed Acyclic Graph (DAG)

NORNS
LATTICE



Distinct pharmacophores bearing the exact same molecule subset and sharing a family relationship form a **Structured Equivalent Class (SEC)**.

112 291
pharmacophores



15 477 SEC
(generators)

⇒ **Decreasing the number of pharmacophore to assess**

Lattice SEC

Directed **A**cylic **G**raph (DAG)

NORNS
LATTICE

15 477 SEC

LOOKING FOR REMARKABLE PHARMACOPHORES

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

NORNS
LATTICE

15 477 SEC

**LOOKING FOR
REMARKABLE PHARMACOPHORES**

...

FROM SEC TO PAD ?

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cylic **G**raph (DAG)

NORNS
LATTICE

15 477 SEC

LOOKING FOR REMARKABLE PHARMACOPHORES

...

FROM SEC TO PAD ?

*The search for outstanding details among
pharmacophores*

Generators

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

FROM SEC TO PAD ?

NORNS
LATTICE

15 477 SEC

SEC = Structured Equivalent Class
PAD = Pharmacophore Activity Delta

Directed Acyclic Graph (DAG)

NORNS
LATTICE

FROM SEC TO PAD ?

15 477 SEC

For each SEC : Computation of the
GR *growth rate* value

$$GR = \frac{\text{Fit frequency within actives}}{\text{Fit frequency within inactives}}$$

SEC = Structured Equivalent Class
 PAD = Pharmacophore Activity Delta

Directed Acyclic Graph (DAG)

**NORNS
LATTICE**

FROM SEC TO PAD ?

15 477 SEC



For each SEC : Computation of the **GR growth rate** value →

A **PAD** is a pharmacophore which GR value differs by at least 2 standard deviations from the mean GR value over itself and its siblings.

$$GR = \frac{\text{Fit frequency within actives}}{\text{Fit frequency within inactives}}$$

SEC = Structured Equivalent Class
 PAD = Pharmacophore Activity Delta

6 essential pharmacophoric motifs

- R Hydrophobic group
- R Aromatic cycle
- D Hydrogen-bond donor
- A Hydrogen-bond acceptor
- N Negatively-ionizable function
- P Positively-ionizable function

Directed Acyclic Graph (DAG)

NORNS
LATTICE

FROM SEC TO PAD ?

15 477 SEC

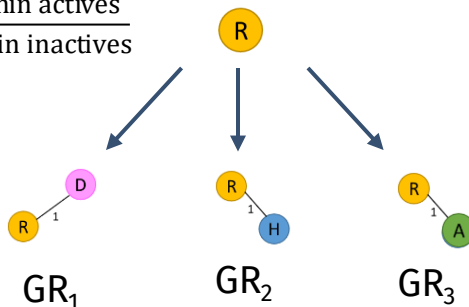
For each SEC :

Computation of the
GR growth rate value



A **PAD** is a pharmacophore which GR value differs by at least 2 standard deviations from the mean GR value over itself and its siblings.

$$GR = \frac{\text{Fit frequency within actives}}{\text{Fit frequency within inactives}}$$



SEC = Structured Equivalent Class
PAD = Pharmacophore Activity Delta

- 6 essential pharmacophoric motifs
- R Hydrophobic group
 - R Aromatic cycle
 - D Hydrogen-bond donor
 - A Hydrogen-bond acceptor
 - N Negatively-ionizable function
 - P Positively-ionizable function

Directed Acyclic Graph (DAG)

**NORNS
LATTICE**

FROM SEC TO PAD ?

15 477 SEC



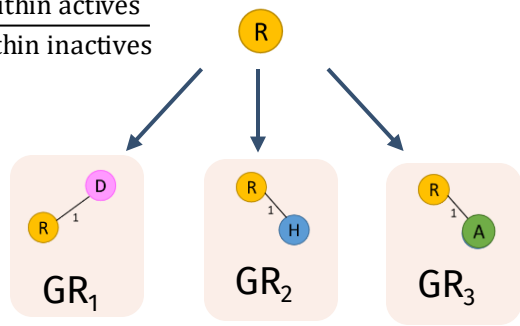
For each SEC : Computation of the **GR growth rate** value



A **PAD** is a pharmacophore which GR value differs by at least 2 standard deviations from the mean GR value over itself and its siblings.

$$GR = \frac{\text{Fit frequency within actives}}{\text{Fit frequency within inactives}}$$

SEC generator



μ mean (GR_1, GR_2, GR_3)
SD Standard deviation

If $|GR_i - \mu| \geq 2 \text{ SD}$ then **i** is a **PAD**

SEC = Structured Equivalent Class
PAD = Pharmacophore Activity Delta

- 6 essential pharmacophoric motifs
- R Hydrophobic group
 - R Aromatic cycle
 - D Hydrogen-bond donor
 - A Hydrogen-bond acceptor
 - N Negatively-ionizable function
 - P Positively-ionizable function

Directed Acyclic Graph (DAG)

**NORNS
LATTICE**

FROM SEC TO PAD ?

15 477 SEC

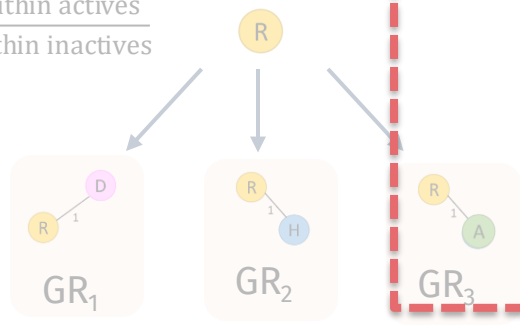
For each SEC :

Computation of the **GR growth rate** value

$$GR = \frac{\text{Fit frequency within actives}}{\text{Fit frequency within inactives}}$$

A **PAD** is a pharmacophore which GR value differs by at least 2 standard deviations from the mean GR value over itself and its siblings.

SEC generator



μ mean (GR_1, GR_2, GR_3)
SD Standard deviation

If $GR_i - \mu \geq 2 \text{ SD}$ then i is a

32 PAD

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

FROM SEC TO PAD ?

NORNS
LATTICE

32 PAD

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

FROM SEC TO PAD ?

NORNS
LATTICE

32 PAD

17 '*active*' pharmacophores
&
15 '*inactive*' pharmacophores

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

NORNS
LATTICE

FROM SEC TO PAD ?

32 PAD

17 'active' pharmacophores
&
15 'inactive' pharmacophores

02 : n=7
03 : n=21
04 : n=2
05 : n=2

SEC = **S**tructured **E**quivalent **C**lass
 PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

NORNS
LATTICE

FROM SEC TO PAD ?

32 PAD

17 'active' pharmacophores
 &
15 'inactive' pharmacophores

O2 : n=7
 O3 : n=21
 O4 : n=2
 O5 : n=2

289 molecules covered
 159 active molecules [20.5% of active molecules]
 130 inactive molecules [18.4% of inactive molecules]

Initial dataset: 1479 molecules
 (773 active+706 inactive ones)

SEC = **S**tructured **E**quivalent **C**lass
 PAD = **P**harmacophore **A**ctivity **D**elta

Directed **A**cyclic **G**raph (DAG)

NORNS
LATTICE

FROM SEC TO PAD ?

32 PAD

17 'active' pharmacophores
 &
 15 'inactive' pharmacophores

02 : n=7
03 : n=21
04 : n=2
05 : n=2

289 molecules covered
 159 active molecules [20.5% of active molecules]
 130 inactive molecules [18.4% of inactive molecules]

Initial dataset: 1479 molecules
 (773 active+706 inactive ones)

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

FROM SEC TO PAD ?

One example in **O2** PADs

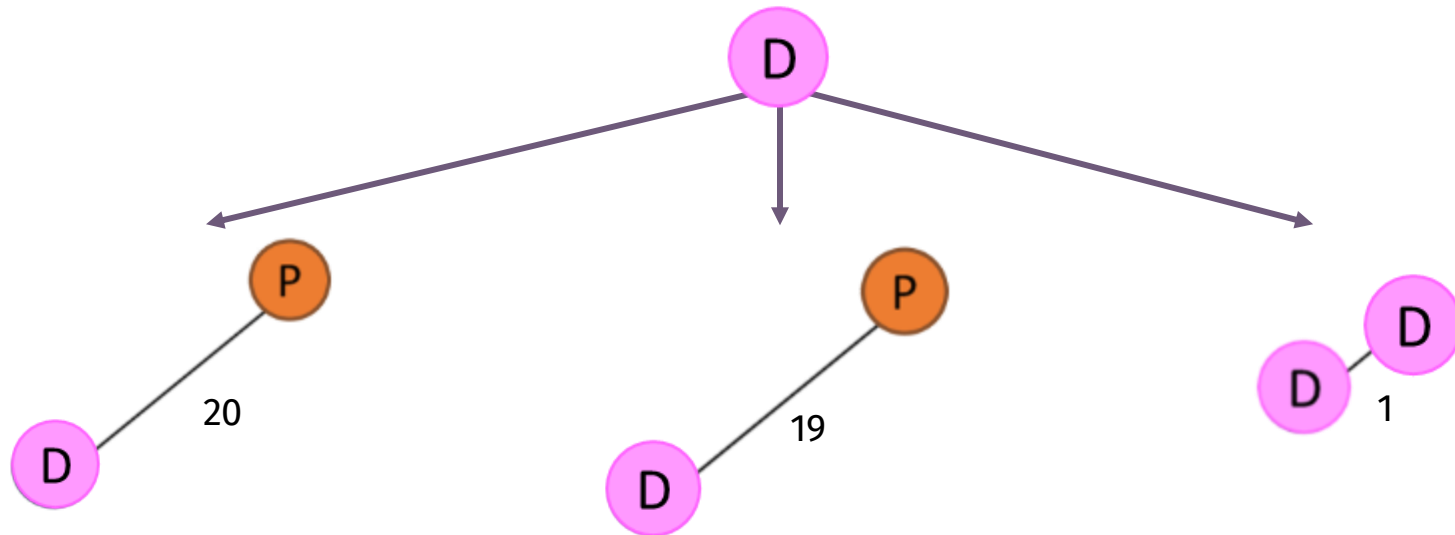
SEC = **S**tructured **E**quivalent **C**lass
 PAD = **P**harmacophore **A**ctivity **D**elta

6 essential pharmacophoric motifs

- H Hydrophobic group
- R Aromatic cycle
- D Hydrogen-bond donor
- A Hydrogen-bond acceptor
- N Negatively-ionizable function
- P Positively-ionizable function

FROM SEC TO PAD ?

One example in O₂ PADs



Order: 1

Order: 2

SEC = **S**tructured **E**quivalent **C**lass
 PAD = **P**harmacophore **A**ctivity **D**elta

6 essential pharmacophoric motifs

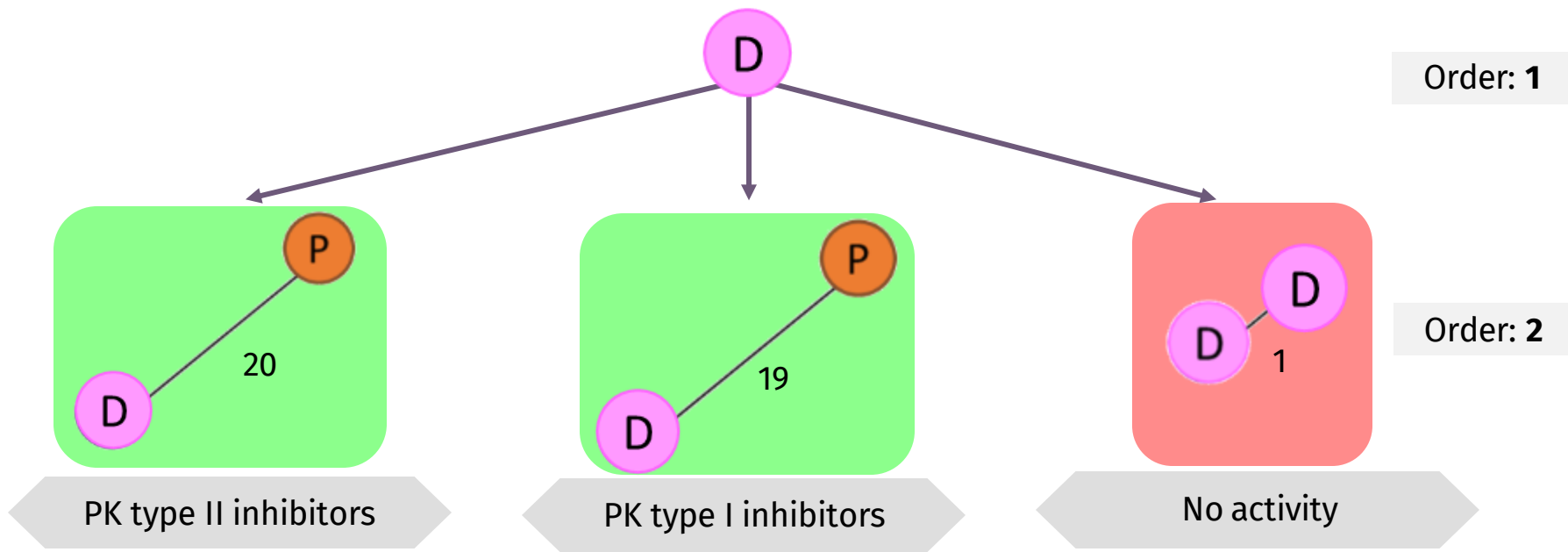
- H Hydrophobic group
- R Aromatic cycle
- D Hydrogen-bond donor
- A Hydrogen-bond acceptor
- N Negatively-ionizable function
- P Positively-ionizable function

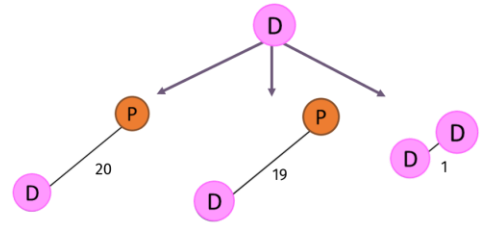
Active

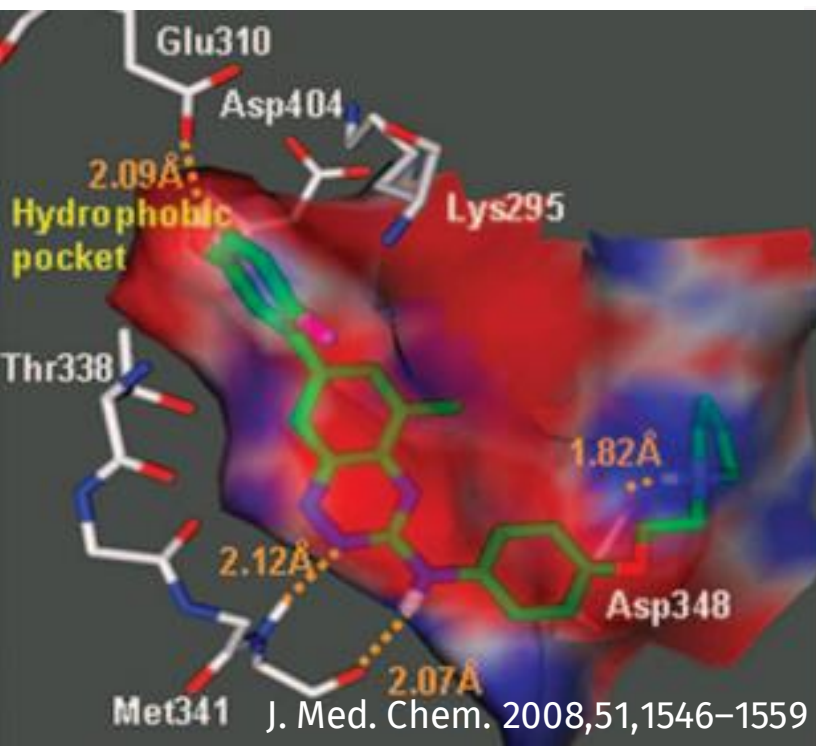
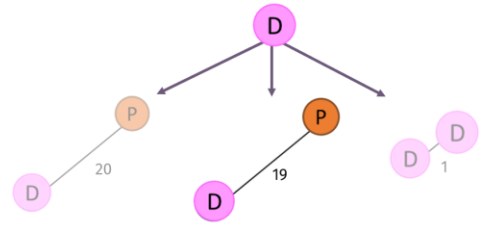
Inactive

FROM SEC TO PAD ?

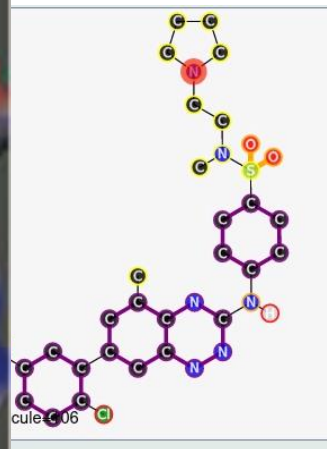
One example in O₂ PADs







(Ph4:0) Measures	
Support Size	22
Confidence	1
Growth Rate	Infinity



Alkaline moiety able to interact with an acid residue

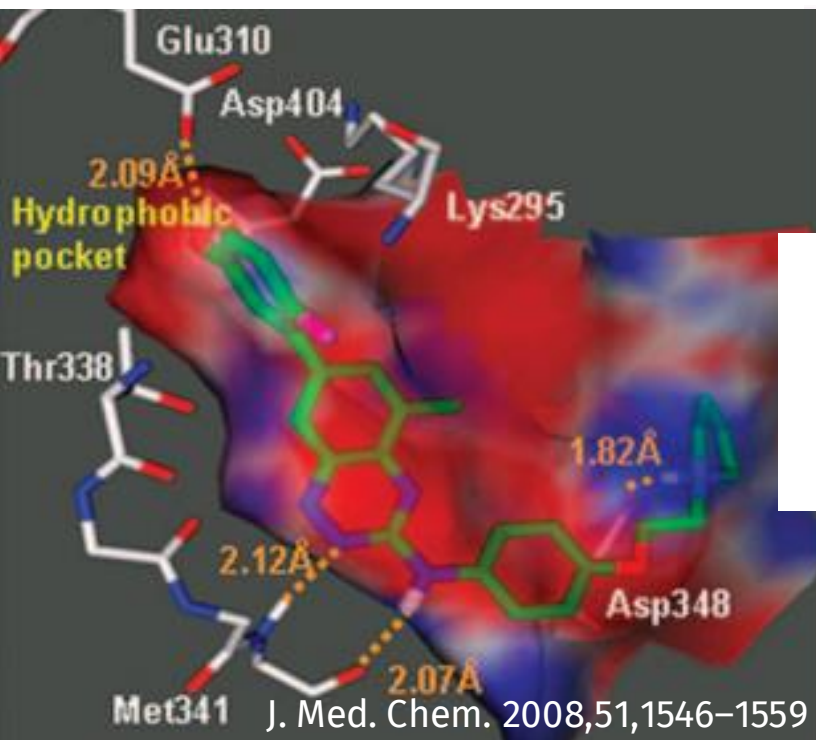
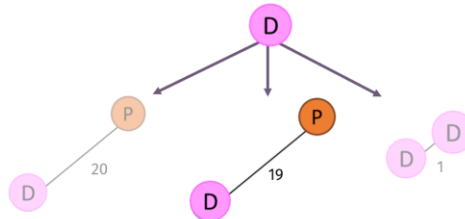
Phenylbenzotriazinamine linker able to angle properly the two terminal moieties

Constant phenol in terminal position

J. Med. Chem. 2008,51,1546–1559

Matches						
244	254	266	267	327	331	345
381	406	423	482	512	567	583

591	702
-----	-----



(Ph4:0) Measures	
Support Size	22
Confidence	1
Growth Rate	Infinity



PKI type I



Matches						
244	254	266	267	327	331	345
381	406	423	482	512	567	583

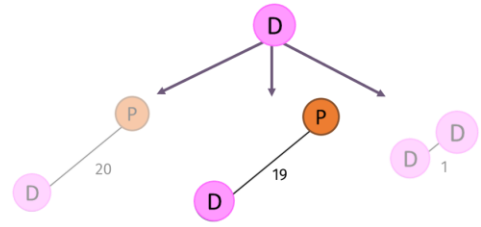
J. Med. Chem. 2008,51,1546–1559

591 702

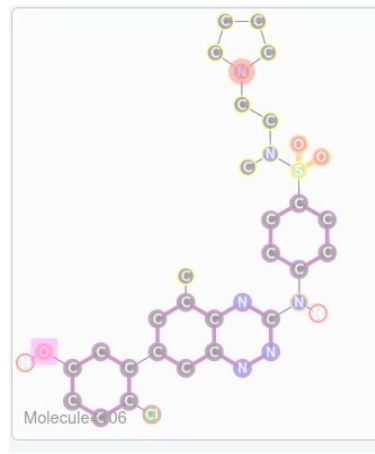
Alkaline moiety able to interact with an acid residue

Phenylbenzotriazinamine linker able to angle properly the two terminal moieties

Constant phenol in terminal position



(Ph4:0) Measures	
Support Size	22
Confidence	1
Growth Rate	Infinity



Matches									
106	132	155	244	254	266	267	327	331	345
355	357	358	381	406	423	482	512	567	583
591	702								

Directed Acyclic Graph (DAG)

To go further into structural entities and pharmacophore DAG:
An interactive visualization tools from Bordeaux University

DAVID AUBER
THEO DUPONT
THOMAS BARILLOT

THIERNO-AMADOU DIALLO
NICOLAS MAJOREL
SOUFIANE EL-KHAILI
YOHAN LEMATRE

	Name		Pharmacophores				Stats				Sibling					
	Id	Label	F Order	L Order	Size	Frequency	Quality	Mean	Standard Deviation	Pertinence	Num. of Nodes	Num. of Edges	Num. of Nodes (Pertinence > 2)	Percentage % of Nodes (Pertinence > 2)	Num. of Parents	Num. of Childs
<input type="checkbox"/>	1	" P "	1	1	1	362	0.543912	0.01	0.01	0.543912	7	6	0	0.00000	1	6
<input type="checkbox"/>	2	" N "	1	1	1	68	0.219374	0.01	0.01	0.219374	7	6	0	0.00000	1	6
<input type="checkbox"/>	3	" D "	1	1	1	1440	0.513505	0.01	0.01	0.513505	7	6	0	0.00000	1	6
<input type="checkbox"/>	4	" A "	1	1	1	1479	0.5	0.01	0.01	0.5	7	6	0	0.00000	1	6
<input type="checkbox"/>	5	" R "	1	1	1	1478	0.500354	0.01	0.01	0.500354	7	6	0	0.00000	1	6
<input type="checkbox"/>	6	" H "	1	1	1	1368	0.531598	0.01	0.01	0.531598	7	6	0	0.00000	1	6
<input type="checkbox"/>	7	" P R "	2	2	1	112	0.88387	0.652159	0.221129	1.047853	74	73	3	4.05405	1	73
<input type="checkbox"/>	8	" D R "	2	2	1	76	0.530125	0.523824	0.236681	0.026623	164	180	3	1.82927	2	162
<input type="checkbox"/>	9	" D P "	2	2	1	54	0.608251	0.523824	0.236681	0.356712	164	180	3	1.82927	2	162
<input type="checkbox"/>	10	" D P "	2	2	1	36	0.58936	0.523824	0.236681	0.276898	164	180	3	1.82927	2	162

Page 1 of 1422 | Go to page: Show 10

Pertinence threshold: **-3**

Quality threshold: **0**

Frequency threshold: **10**

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Summary

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

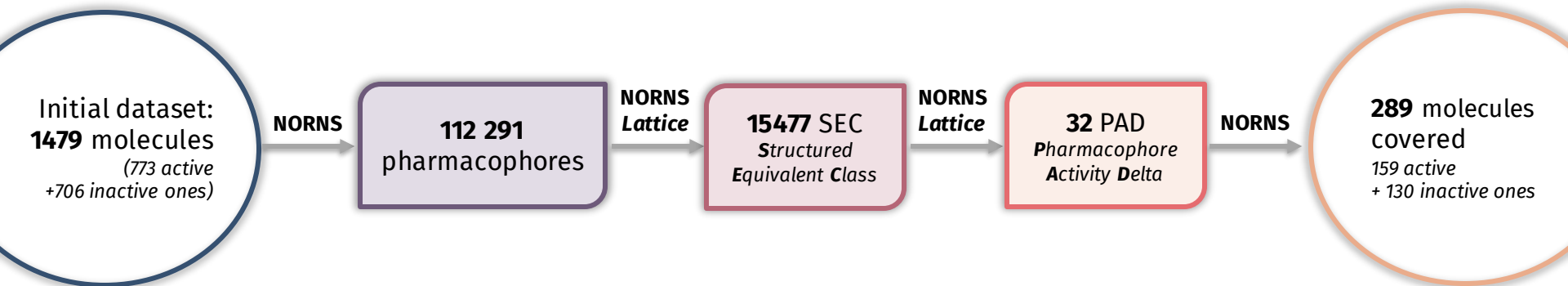
Summary

Initial dataset:
1479 molecules
(773 active
+706 inactive ones)

289 molecules
covered
159 active
+ 130 inactive ones

SEC = **S**tructured **E**quivalent **C**lass
PAD = **P**harmacophore **A**ctivity **D**elta

Summary





112 291 pharmacophores

What else ?

112 291 pharmacophores

Deciphering a pharmacophore network generated from BCR-ABL data

Damien Geslin^{1,2,3*}, Alban Lepailleur^{1,3}, Jean-Luc Mangin^{1,2}, Nhat Vinh Vo^{1,3}, Jean-Luc Lamotte^{1,3}, Bertrand Cuisant^{1,2}, Ronan Bureau^{1,3}

¹ Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ., UNICAEN, CERMA, 14000 Caen, France.

² Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ., UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen France

* Correspondence: damien.geslin@unicaen.fr

What else ?

Introduction
Recently, we described a new approach for the automated detection of pharmacophores and their organization into a network starting from a large chemical dataset.¹ As a case study, we worked with a dataset dealing with BCR-ABL inhibitors and consisting of 1492 molecules (774 actives and 718 inactive). In this poster, we present a method used to spatialize the network by computing the graph edit distances (GED) between the pharmacophores. Then, a clustering approach was used to refine the partitioning by grouping pharmacophores according to their structures, activities and binding modes.

Context
Our methodology generates a large number of pharmacophores leading to a complexity for the search space. The objective of the present work is to organize and to analyze the pharmacophore network.

Generation of the pharmacophores
A pharmacophore (Figure 1.B) describes a combination of chemical features shared by several active molecules and responsible for favorable interactions with the active site of a target. In a pharmacophoric graph:
- a node denotes a pharmacophoric feature ;
- an edge encodes the minimal distance between two nodes (number of chemical bonds).



The main parameters for the extraction of the pharmacophores are their order (number of features per pharmacophore), their support (number of molecules associated to the pharmacophore) and their cut-off value for the growth-rate (imbalance between actives and inactives). A selection of representative pharmacophore was achieved using MMARS algorithm.¹ In the present study, we worked with different pharmacophore orders (O_i for 3 nodes to O₇ for 7 nodes). Each pharmacophore is labelled with a class (active or inactive) in relation with the chemical compounds which fit it (ratio between active and inactive molecules).

Clustering and analysis of the pharmacophore network

The goal of clustering is to group together elements similar to each other in the same cluster. We tested three clustering methods (hierarchical clustering, k-means, and spectral) after the detection of outliers using ORSCAN.²

- 1) We obtained the best clustering for O₃ with Hierarchical clustering (AGNES) according to the NMI values.
- 2) We started the visualization with an initial representation of the pharmacophore network (Figure 4). This representation was obtained by applying a force-directed layout algorithm and by considering the two nearest neighbors of each pharmacophore.
- 3) We superimposed the cluster-based partitioning of the O₃ pharmacophores on the initial pharmacophore network (Figure 5).
- 4) We draw a parallel between the active clusters 1-4 and the binding mode of representative compounds (cocrytalographic data or docking simulations).

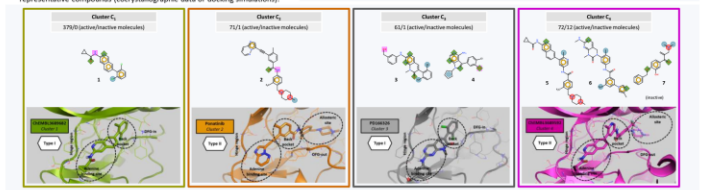


Figure 5. Superposition of clustering on the pharmacophore network.

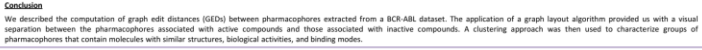
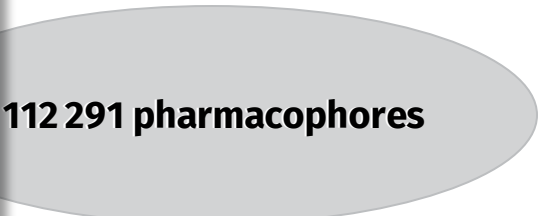


Figure 6. Cocrytalographic data (clusters 2 and 3) and docking simulations (clusters 1 and 4) that support the links between the active clusters and kinase binding modes.

Conclusion
We described the computation of graph edit distances (GEDs) between pharmacophores extracted from a BCR-ABL dataset. The application of a graph layout algorithm provided us with a visual separation between the pharmacophores associated with active compounds and those associated with inactive compounds. A clustering approach was then used to characterize groups of pharmacophores that contain molecules with similar structures, biological activities, and binding modes.

¹ Geslin, D., Cuisant, B., Bureau, R., Lepailleur, A. The Pharmacophore Network: A Computational Method for Exploring Structure-Activity Relationships from a Large Chemical Data Set. *J. Med. Chem.* **2018**, *61*, 3551-3558.
² Mouton, L.P., Geslin, D., Bureau, R., Lepailleur, A. The Pharmacophore Network: A Computational Method for Exploring Structure-Activity Relationships from a Large Chemical Data Set. *J. Med. Chem.* **2018**, *61*, 3551-3558.
³ Garcia-Morales, C., Ramirez, A., Gonzalez, F. Ligand-based virtual screening using graph edit distance as molecular similarity measure. *J. Chem. Inf. Model.* **2014**, *14*, 1410-1421.
⁴ Bourne, S.E., Riva, A., Gargner, J., Havelka, T. Computing Similarity for graph edit distance comparison. *The 10th Int. Conf. on Artificial Intelligence and Applications* **2008**, *10*, 428-432.
⁵ Seema, A., Prasad, M., Gupta, A., Bharati, N., Patel, D.P., Thant, A., Et, M.J., Ding, W., Liu, C.-T. A Review of Clustering Techniques and Developments. *Neurocomputing* **2017**, *267*, 684-681.



What else ?

Deciphering a pharmacophore network generated from BCR-ABL data

Damien Geslin^{1,2}*, Alban Lepaillerie^{1,2}, Jean-Luc Manguin^{1,2}, Nhat Vinh Vo^{1,2}, Jean-Luc Lamotte^{1,2}, Bertrand Cussart^{1,2}, Ronan Bureau^{1,2}

⁽¹⁾ Centre d'Études et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMAN, 14000 Caen, France.

⁽²⁾ Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen France

*Correspondence: damien.geslin@unicaen.fr

⇒ Study of a **representative subset of pharmacophores (MMRFS)** instead of the whole dataset / **Pharmacophore space**

MMRFS

= Marginal Relevance Feature Selection

Figure 1. A molecule (A) and its corresponding pharmacophore graph (B)

Figure 2. The 6 pharmacophoric features

The main parameters for the extraction of the pharmacophores are their order (number of features per pharmacophore), their support (number of molecules associated to the pharmacophore) and their call-off value for the growth-rate (imbalance between active and inactive). A selection of representative pharmacophores is achieved using MMRFS. In the present study, we worked with 112 291 pharmacophores. Each pharmacophore is a graph of chemical compounds which are connected by edges.

Figure 4. Representation of a BCR-ABL pharmacophore network (O₁ and O₂) obtained with ClusterG (force-directed layout algorithm).

COME AND MEET DAMIEN

Clustering and analysis of the pharmacophore network
The goal of clustering is to group together elements similar to each other in the same cluster. We tested three clustering methods (hierarchical clustering, k-means, and spectral) after the detection of outliers using ORSCAN.

- We obtained the best clustering for O₁ with hierarchical clustering (AGNES) according to the RM values.
- We used the hierarchical clustering algorithm to group the pharmacophores. This algorithm is based on the nearest neighbor method.
- We used the hierarchical clustering algorithm to group the pharmacophores on the basis of their support.
- We draw a parallel between the active clusters O₁ and the binding mode of representative compounds (cocrystallographic data or docking simulations).



GESLIN AT POSTER #12!

Conclusion
We described a computation on the pharmacophore network. In this paper, we present a visual separation between the pharmacophores associated with active compounds and those associated with inactive compounds. Clustering approach was then used to characterize groups of pharmacophores that contain molecules with similar structures, biological activities, and binding modes.

References
(1) Geslin, D., Cussart, B., Bureau, R., Lepaillerie, A. The Pharmacophore Network: A Computational Method for Exploring Structure-Activity Relationships from a Large Chemical Data Set. *J. Med. Chem.* **2018**, *61*, 3552-3564.
(2) Geslin, D., Vo, N. H., Vo, N. C. An Improved Fragment-Based Analysis for Molecular Classification in Drug Discovery. *Drug Discovery* **2019**, *10*, 118-126.
(3) Garcia-Mendez, C., Fernandez, A., Gonzalez, F. Ligand-based virtual screening and graph edit distance as molecular similarity measure. *J. Chem. Inf. Model.* **2014**, *54*, 1410-1421.
(4) Bourgeois, S. E., Kelly, C., Gaspard, J., Roussel, Y. B. C. Comparing Network to graph edit distance comparison. *The IJCN* **2009**, *30*, 474-478.
(5) Serrano, A., Prasad, M., Shalev, A., Bharath, N., Faloutsos, D. P., Tam, A. E., M. J., Ding, W., Liu, C.-T. A Review of Clustering Techniques and Developments. *Neurocomputing* **2017**, **267**, 664-681.

112 291 pharmacophores

Deciphering a pharmacophore network generated from BCR-ABL data

Damien Geslin^{1,2,3*}, Alban Lepaillier^{1,2}, Jean-Luc Manguin^{1,2}, Nhat Vinh Vo^{1,1}, Jean-Luc Lamotte^{1,1}, Bertrand Cuisart^{1,1}, Ronan Bureau^{1,1}
¹ Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ., UNICAEN, CERMAN, 14000 Caen, France.
² Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ., UNICAEN, ENICAEN, CNRS, GREYC, 14000 Caen France
*Correspondence: damien.geslin@unicaen.fr

⇒ Study of a representative subset of pharmacophores (MMRFS) instead of the whole dataset / Pharmacophore space

MMRFS

= Marginal Relevance Feature Selection

What else ?

Analyzing new pharmacophoric fingerprint via FFNNs: Application to Bcr-Abi data.

Hajar Rehioui^{1*}, Bertrand Cuisart¹, Abdelkader Ouali¹, Alban Lepaillier², Jean-Luc Lamotte³, Ronan Bureau⁴, Albrecht Zimmermann⁵
¹ GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France
² Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ., UNICAEN, CERMAN, 14000 Caen, France
³ Hajar.rehioui.karine@unicaen.fr

1. Introduction

We propose here to analyze the potential of a new type of pharmacophoric descriptors via two feedforward neural network (FFNN) feature transformations. The application is related to virtual screening on a tyrosine kinase named BCR-ABL. This latter, collected from ChEMBL, includes 1479 compounds [1] (73 active compounds and 706 inactive ones). Firstly, the compounds were described using three different families of descriptors: our new pharmacophoric descriptors FFP and two radial-basis fingerprints: ECFP4 and FCFP4. Then, each of these three molecular descriptions were transformed using two feature transformations: Weight-Matrix Learning[2] (WML) and Supervised Weight-Matrix Learning (SWML). Finally, a K-means clustering algorithm was applied, on the three transformed representations, to partition the considered molecules. The final results with FFP coupled to SWML was convincing and give clearly better results in terms of quality measures.

2. FrP fingerprint

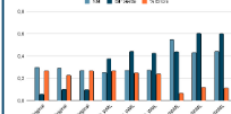
- The extraction of the frequent pharmacophores (FrP) [3, 1], is based on the pharmacophore features: hydrogen-bond acceptors (A) and donors (D), negatively (N) and positively (P) charged ionizable groups, hydrophobic regions (H), and aromatic rings (R).
- The minimal support for the pharmacophores, i.e. the minimum number of compounds in which a pharmacophore occurs, was set to 3, then we have identified all the FrP with 3 to 7 pharmacophoric features.
- With this parameters, 11976 FrP were generated, then was reduced to 15046 FrP by eliminating redundant ones, e.i. those FrP associated to exactly the same extension or the same set of molecules.

4. SWML

- Is our vision of a supervised feature transformation method based on FFNN. SWML relies on the activity of ligands in order to minimize the "categorical crossentropy" loss function (CE). Aims to separate active ligands from inactive ones as much as possible.
- applied dropout and L2 regularization to avoid the possible overfitting associated to supervised techniques.

5. Results

A clustering comparison was done on data, the best results were obtained from k=2 to 5 with an optimum for k = 3. All the following discussed results are related to K = 3.



- The diagram compares the results of full trained data based on original data, WML and SWML transformations.
- SWML leads to the best clustering whatever the descriptors.
- FrP outperforms ECFP4 and FCFP4 especially for the NMI and rate of misclassified compounds.

3. WML

- Is an unsupervised feature transformation method, considered as a FFNN, based on the low function (EW').
- Aims to learn a matrix that transforms the chemical features from the original space to a new feature space, where ligands with similarity larger than 0.5 will be closer to each other and those with similarity smaller than 0.5 will be farther.

6. Conclusions

Our proposed pharmacophoric descriptors FFP leads to convincing results associated to feature transformations with FFNN and classical clustering.

- SWML gives the best results.
- The definition of three clusters of FrP coupled with SWML for BCR-ABL data, allows to obtain good predictive results with our initial data set and with decoys compared to ECFP4 and FCFP4.
- In the three clusters, the two active clusters are different in terms of constraints with FrP.

7. References

- Damien Geslin, Alban Lepaillier, Jean-Luc Manguin, Nhat-Vinh Vo, Jean-Luc Lamotte, Bertrand Cuisart, and Ronan Bureau. Deciphering a pharmacophore network: A case study using bcr-abl data. *Journal of Chemical Information and Modeling*, 2022.
- Daen Yan, Xindei Zhou, Xihao Wang, and Ran Wang. An off-center technique: Learning a feature transformation to improve the performance of clustering and classification. *Information Sciences*, 503:635–651, 2019.
- Jean-Philippe Metivier, Bertrand Cuisart, Ronan Bureau, and Alban Lepaillier. The pharmacophore network: a computational method for exploring structure-activity relationships from a large chemical data set. *Journal of Medicinal Chemistry*, 61(8):3551–3564, 2018.
- Michael M Mysinger, Michael Cerchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582–6594, 2012.

Intro
Recent
dataset
distanc
Conte
Our m
search
network
Gene
A phar
active
In a g
- n-AB
- n-AB

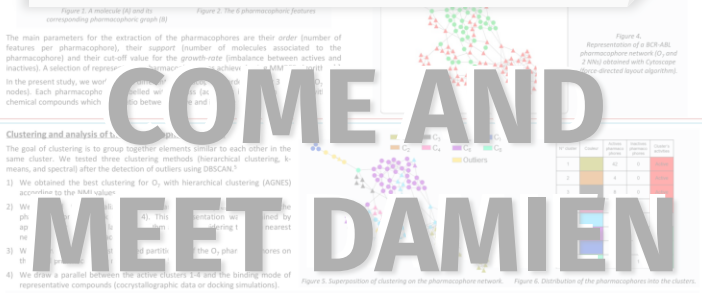


Figure 1: A molecular graph and its corresponding pharmacophore graph (B). Figure 2: The 6 pharmacophore features. Figure 3: Representation of a BCR-ABL pharmacophore network (D) and 7 WML clusters with Consensus score-directed laplacian weights. Figure 4: Clustering and analysis of the network. Figure 5: Superposition of clustering on the pharmacophore network. Figure 6: Distribution of the pharmacophores into the clusters.

112 291 pharmacophore

GESLIN AT POSTER #12!

Deciphering a pharmacophore network generated from BCR-ABL data

Damien Geslin^{1,2,3*}, Alban Lepaillier^{1,2}, Jean-Luc Manguin^{1,2}, Nhat Vinh Vo^{1,2}, Jean-Luc Lamotte^{1,2}, Bertrand Cuissart^{1,2}, Ronan Bureau^{1,3}

⁽¹⁾ Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ., UNICAEN, CERMAN, 14000 Caen, France

⁽²⁾ Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ., UNICAEN, ENICAEN, CNRS, GREYC, 14000 Caen France

*Correspondence: damien.geslin@unicaen.fr

⇒ Study of a representative subset of pharmacophores (MMRFS) instead of the whole dataset / Pharmacophore space

MMRFS

= Marginal Relevance Feature Selection

Figure 1. A molecule (R) and its corresponding pharmacophore graph (B)

Figure 2. The 8 pharmacophore features

The main parameters for the extraction of the pharmacophores are their order (number of features per pharmacophore), their support (number of molecules associated to the pharmacophore) and their call-off value for the growth-rate (imbalance between active and inactive). A selection of regression coefficients is shown in the table below.

Clustering and analysis of the pharmacophore network

The goal of clustering is to group together elements similar to each other in the same cluster. We tested three clustering methods (hierarchical clustering, k-means, and spectral) after the detection of outliers using OSCAR.

- 1) We obtained the best clustering for Q_2 with Hierarchical clustering (AGNES) according to the NMI value.
- 2) We used the hierarchical clustering to identify the most representative pharmacophore in each cluster.
- 3) We used the hierarchical clustering to identify the most representative pharmacophore in each cluster.
- 4) We draw a parallel between the active clusters 1-4 and the binding mode of representative compounds (crystallographic data or docking simulations).

Figure 3. Superposition of clustering on the pharmacophore network. Figure 4. Distribution of the pharmacophores into the clusters



We do not find significant differences in the pharmacophore space between the pharmacophores associated with active compounds and those associated with inactive compounds. A clustering approach was then used to characterize groups of binding modes that contain molecules with similar structures, biological activities, and binding modes.

10.1002/chem.201402001
11.1002/chem.201402001
12.1002/chem.201402001
13.1002/chem.201402001
14.1002/chem.201402001
15.1002/chem.201402001
16.1002/chem.201402001
17.1002/chem.201402001
18.1002/chem.201402001
19.1002/chem.201402001
20.1002/chem.201402001

What else ?

Analyzing new pharmacophoric fingerprint via FFNNs: Application to Bcr-Abl data.

Hajar Rehioui^{1*}, Bertrand Cuissart¹, Abdelkader Ouali¹, Alban Lepaillier¹, Jean-Luc Lamotte¹, Ronan Bureau¹, Albrecht Zimmermann²

¹ GREYC, Normandie Univ., UNICAEN, CNRS - UMR 6072, 14000 Caen, France

² Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ., UNICAEN, CERMAN, 14000 Caen, France

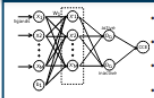
*hajar.rehioui@carnebiucaen.fr



1. Introduction

⇒ Use of new pharmacophoric descriptors and Neural Network/Clustering

- The extraction of the frequent pharmacophores (FP) [3, 1], is based on the pharmacophore features: hydrogen-bond acceptors (A) and donors (D), negatively (N) and positively (P) charged ionizable groups, hydrophobic regions (H), and aromatic rings (R).
- The minimal support for the pharmacophores, i.e. the minimum number of compounds in which a pharmacophore occurs, was set to 10, then we have identified all the FP with 3 to 7 pharmacophoric features.
- With this parameter set, 112 291 pharmacophores were generated, then we used to identify the most discriminating redundant ones, e.g. the FP related to exactly the same set of molecules or the same set of molecules.



- Is our vision of a supervised feature transformation method based on FFNN. SWML relies on the activity of ligands in order to minimize the "categorical crossentropy" loss function (CE).
- Aims to separate active ligands from inactive ones as much as possible.
- applied dropout and L2 regularization to avoid the possible overfitting associated to supervised techniques.

5. Results

A clustering comparison was done on data, the best results were obtained from k=2 to 5.

SWML results of full and partial data, WML results of full and partial data, clustering whatever the data.

FP outperforms ECFP4 and FCFP4 especially for the NMI and rate of misclassified compounds.

Only 20 decoys are associated with an active cluster FP which is very surprising compared to the other descriptors. That could be explained by the fact that FP are more precise than a structural fragment.



from the original where ligands will be closer to each other than smaller than 0.5

6. Conclusions

Our proposed pharmacophoric descriptors FP leads to a better classification of the compounds. 1. SWML results of full and partial data, WML results of full and partial data, clustering whatever the data.

7. References

- [1] Damien Geslin, Alban Lepaillier, Jean-Luc Manguin, Nhat-Vinh Vo, Jean-Luc Lamotte, Bertrand Cuissart. Deciphering a pharmacophore network: a case study. *Journal of Medicinal Chemistry*, 55(14):3551-3564, 2012.
- [2] Yan Zhang, Michael Carich, John J Irwin, and Brian K Shoichet. Mining a large chemical data set and with decoys compared to ECFP4 and FCFP4. *Journal of Medicinal Chemistry*, 55(14):3551-3564, 2012.
- [3] Michael M Mysinger, Michael Carich, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582-6594, 2012.

COME AND MEET DAMIEN GESLIN AT POSTER #12!

112 291 pharmacophore

POSTER #42!

Towards DAG-based interactive pharmacophore exploration

Johanna GIOVANNINI^{1*}, Etienne LEHEMBRE^{2*}, Jean-Luc LAMOTTE², Abdelkader OUALI¹, Alban LEPAILLEUR¹, Albrecht ZIMMERMANN¹, Bertrand CUISSART¹, Ronan BUREAU¹ and Bruno CREMILLEUX¹

⁽¹⁾ Centre d'Etudes et de Recherche sur le Médicament de Normandie, Université de Caen Normandie, 14000 Caen, France
⁽²⁾ GREYC, Normandie Univ., UNICAEN, CNRS - UMR 6072, 14000 Caen, France
 * Contributed equally

Corresponding authors: johanna.giovannini@unicaen.fr; etienne.lehembre@unicaen.fr



CONTEXT
 Pharmacophore comprehension is essential for the understanding of molecule bioactivities (therapeutic and side effects).¹ From a chemical point of view, a pharmacophore is the greatest common structural denominator associated with a group of compounds exhibiting the same biological response profile. Considering graph theory, 2D topological pharmacophores represent particular patterns present in a number of structures (See Fig. 1).^{2,3}



Fig. 1. From a chemical structure (a) to a pharmacophore graph (b), by superimposing pharmacophoric motifs and chemical structure (b).

When applied to a data set partitioned into two classes (e.g., active versus inactive molecules toward a biological receptor (for instance), emerging pattern (EP) mining can identify the pharmacophoric patterns that occur with higher frequency in one of the two classes.
 => In this study, we focus on outstanding pharmacophores among a dataset.

$$GR_R = \frac{GR}{(GR + 1)}$$

GOAL
 Apply EP mining methods to pharmacophore network examination in order to create a tool to highlight outstanding pharmacophores among a dataset.

We retrieved the ChEMBL compound dataset of BCR-ABL ligands (MW ≤ 800 g/mol): 1479 molecules with either Ki or IC50 information towards ABL1 target affinity. 773 were considered as active compounds (Ki or IC50 ≤ 100 nM).

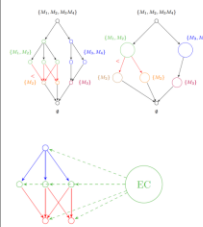
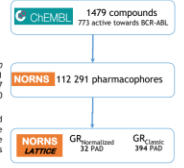


Fig. 2. From a relational directed acyclic graph (DAG) to a cluster relational diagram.

STRUCTURING
 Using the sub-graph inclusion relation of the pharmacophores we build a transitive relational diagram where two pharmacophores are linked if one is included in the other. Then we group elements which appear in the exact same set of molecules if they are linked by a path in the relation graph which only goes through vertices with the exact same set of molecules. Those groups are called Equivalence Classes (EC) (See Fig. 2).

Using our in-house software Norms (2.0 version), we generated 112,291 pharmacophores from order 1 to order 7 with a minimal support of 10 compounds.^{3,4}
 GR stands for 'Growth Rate' and corresponds to the ratio between the frequencies with which a pharmacophore fits into each of the subgroups (active vs inactive compounds).

We applied SEC condensed DAG on the pharmacophores and we retrieved normalized and classic GR values for each pharmacophore. This provided PAD, for Pharmacophore Activity Delta. A PAD is a pharmacophore whose GR value differs by at least two standard deviations from the mean GR value over itself and its siblings.

APPLICATION OF SEC EP MINING TO BCR-ABL ChEMBL COMPOUND DATASET

=> The structural analysis of molecules related to the PAD allowed us to quickly get crucial information on the initial large database.

Here are two examples:

CONSEQUENCES
 Clustering pharmacophores allows us to reduce the number of elements to evaluate (see Fig. 3). We can take an interest in generators which are elements without parents in the equivalence class or closed patterns which are elements without children in the equivalence class (respectively in blue and red in Fig. 2).

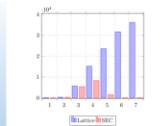


Fig. 3. Number of pharmacophores per layers (in blue) and number of EC per layer (in red).

DP19, related to active molecules towards BCR-ABL, and DP1, related to inactive ones.
 3 essential and bound up structural entities:
 - Aliphatic moiety able to interact with an acid residue
 - Phenylbenzoxazine linker able to angle properly the two terminal moieties
 - Constant phenol in terminal position
 This subset of molecules is composed of inactive ligands. They are all bearing an N-hydroxyamide entity, providing this information as a prerequisite for binding towards BCR-ABL protein.

FINDING REMARKABLE PHARMACOPHORES
 In order to identify remarkable EC we will take interest in their siblings (in purple). As they share at least one parent with our starting node (in red) they have the smallest graph/set intersection difference (see Fig. 4). It allows us to study pharmacophores which deviate from their siblings and might explain biological activity.

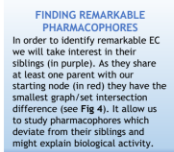
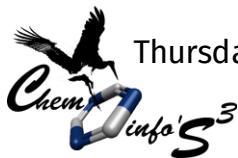


Fig. 4. From parents (Pi) (in blue) of a starting EC (in red), we extract the EC siblings (Si) (in purple) in order to study Si regarding the S.

With the structural study of only 32 PAD among the 112291 initial pharmacophore dataset, we could highlight structural entity related to only active molecules or related to only inactive molecules. Then, a bibliographical study gave us information related to how the structural entities would arrange spatially towards BCR-ABL.

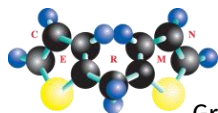
^[1] J. Klotzel, et al., Pharmacophores and Pharmacophore Spaces, John Wiley & Sons, 2009. ^[2] D. Horvath, Chemoinformatic Approaches Virtual Screen., 2008. ^[3] G. Schneider, W. Neubart et al., Angew. Chem. Int. Ed., 1999, 38, 2894-2896. ^[4] et al. chem/target_report_dag04/ChEMBL1922 ^[5] McKeown et al., J. Med. Chem., 2018, 61, 35 ^[6] Galin et al., J. Chem. Inf. Model., 2022, 62.



Thursday, June 30th

Towards DAG-based interactive pharmacophore exploration: Application to the BCR-ABL ligand set

Johanna GIOVANNINI Etienne LEHEMBRE
Jean-Luc LAMOTTE Abdelkader OUALI Alban LEPAILLEUR
Albrecht ZIMMERMANN Bertrand CUISSART
Ronan BUREAU and Bruno CREMILLEUX



Centre d'Etudes et de Recherche sur le Médicament de Normandie (CERMN)

Groupe de recherche en informatique, image et instrumentation de Caen (GREYC),
Normandie Univ., UNICAEN, CNRS – UMR 6072



Université de Caen Normandie, France

ronan.bureau@unicaen.fr

anr **UNICAEN** **GREYC** **UNIVERSITÉ DE CAEN NORMANDIE**

Towards DAG-based interactive pharmacophore exploration
Johanna GIOVANNINI¹, Etienne LEHEMBRE², Jean-Luc LAMOTTE², Abdelkader OUALI¹, Alban LEPAILLEUR¹, Albrecht ZIMMERMANN¹, Bertrand CUISSART¹, Ronan BUREAU¹ and Bruno CREMILLEUX¹

¹ Centre d'Etudes et de Recherche sur le Médicament de Normandie, Université de Caen Normandie, 14000 Caen, France
² GREYC, Normandie Univ., UNICAEN, CNRS - UMR 6072, 14000 Caen, France
* Contributed equally
Corresponding authors: johanna.giovannini@unicaen.fr; etienne.lehembre@unicaen.fr

CONTEXT
Pharmacophore comprehension is essential for the understanding of molecule bioactivities (therapeutic and side effects).¹ From a chemical point of view, a pharmacophore is the greatest

GOAL
Apply EP mining methods to pharmacophore network examination in order to create a tool to highlight outstanding pharmacophores among a dataset.

Fig 1. From a chemical structure (a) to a pharmacophore graph (b), by superimposing pharmacophoric motifs and chemical structure (b).

Fig 2. A data set partitioned into two classes (e.g., 1479 molecules toward a biological receptor binding pattern (EP) mining can identify the patterns that occur with higher frequency in one of the two classes. We focus on outstanding pharmacophores among a dataset.

$GR_{ik} = \frac{GR}{(GR + 1)}$

Fig 3. Using our in-house software Norms (2.0 version), we generated 112,291 pharmacophores from order 1 to order 7 with a minimal support of 10 compounds.^{2,3,4} GR stands for "Growth Rate" and corresponds to the ratio between the frequencies with which a pharmacophore fits into each of the subgroups (active vs inactive compounds). We applied SEC condensed DAG on the pharmacophores and we retrieved normalized and classic GR values for each pharmacophore. This provided PAD for Pharmacophore Activity Delta. A PAD is a pharmacophore whose GR value differs by at least two standard deviations from the mean GR value over itself and its siblings.

APPLICATION OF SEC EP MINING TO BCR-ABL CHEMBL COMPOUND DATASET

→ The structural analysis of molecules related to the PAD allowed us to quickly get crucial information on the initial large database. Here are two examples:

Here are presented two examples of PAD:

- DP19, related to active molecules towards BCR-ABL, and DD1, related to inactive ones.
- 3 essential and bound up structural entities:
 - Alkaline moiety able to interact with an acid residue
 - Phenylbenzoxazine linker able to angle properly the two terminal moieties
 - Constant phenol in terminal position

This subset of molecules is composed of inactive legends. They are all bearing an N-hydroxyamide entity, providing this information as a prerequisite for binding towards BCR-ABL proteins.

With this initial study (Fig 3), among 112,291 initial pharmacophores, we highlighted 32 active and 32 inactive. Then, a bibliographical study gave us information related to how the structural entities would arrange spatially towards BCR-ABL.

Fig 4. From parents D (in blue) and S (in red), we extract the siblings S₁ and S₂ in order to study S regarding the S.

Fig 5. In order to identify remarkable EC nodes, we take interest in their children. When with one of the starting nodes, it is possible to separate from their siblings and might explain biological activity.

POSTER #13

InvolveD ANR-20-CE23-0023

anr **agence nationale de la recherche**

TREPLIN CARNOT IZC

UNIVERSITÉ DE CAEN NORMANDIE

UNION EUROPÉENNE

REGION NORMANDIE

Fig 1. [1] J. Giovannini et al., Pharmacophores and Pharmacophore Spaces, John Wiley & Sons, 2009. [2] G. Herzlich, Chemoinformatics Approaches Virtual Screen., 2008. [3] G. Schneider, W. Neidhart et al., Angew. Chem., Int. Ed., 1999, 38, 2894-2896. [4] doi:10.1039/c8cc00019a