InvolvD ANR-20-CE23-0023

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

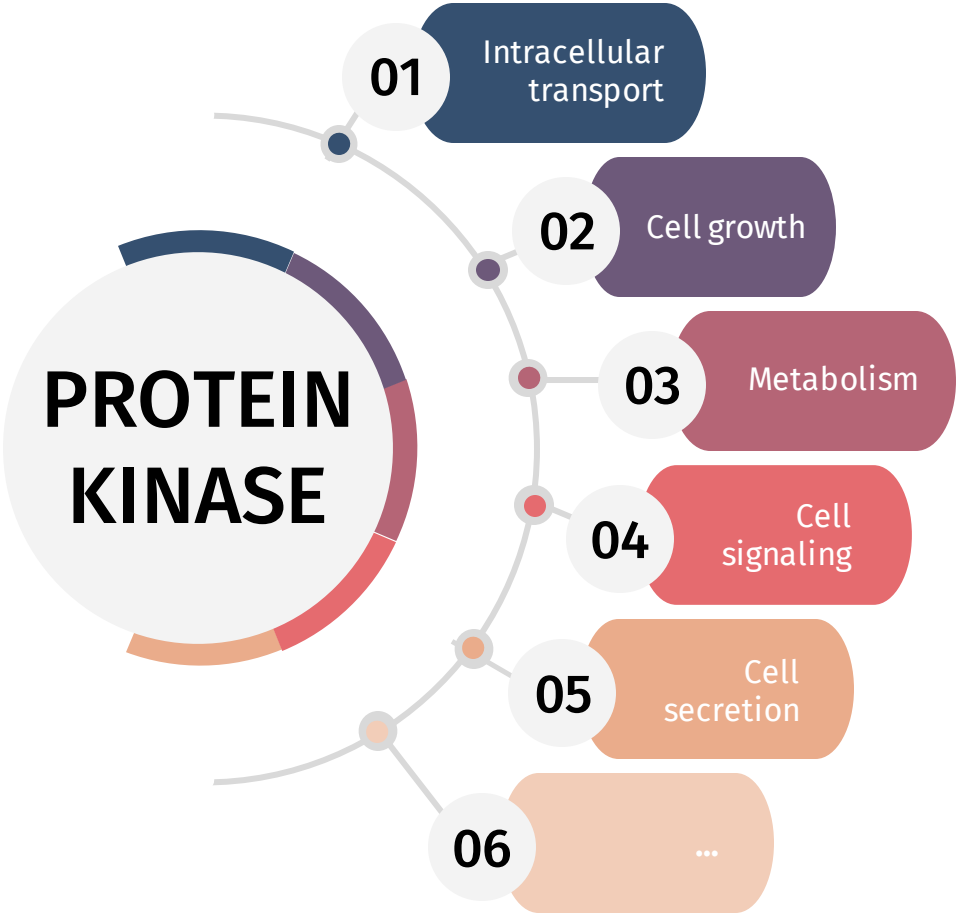
Towards DAG-based interactive pharmacophore exploration:

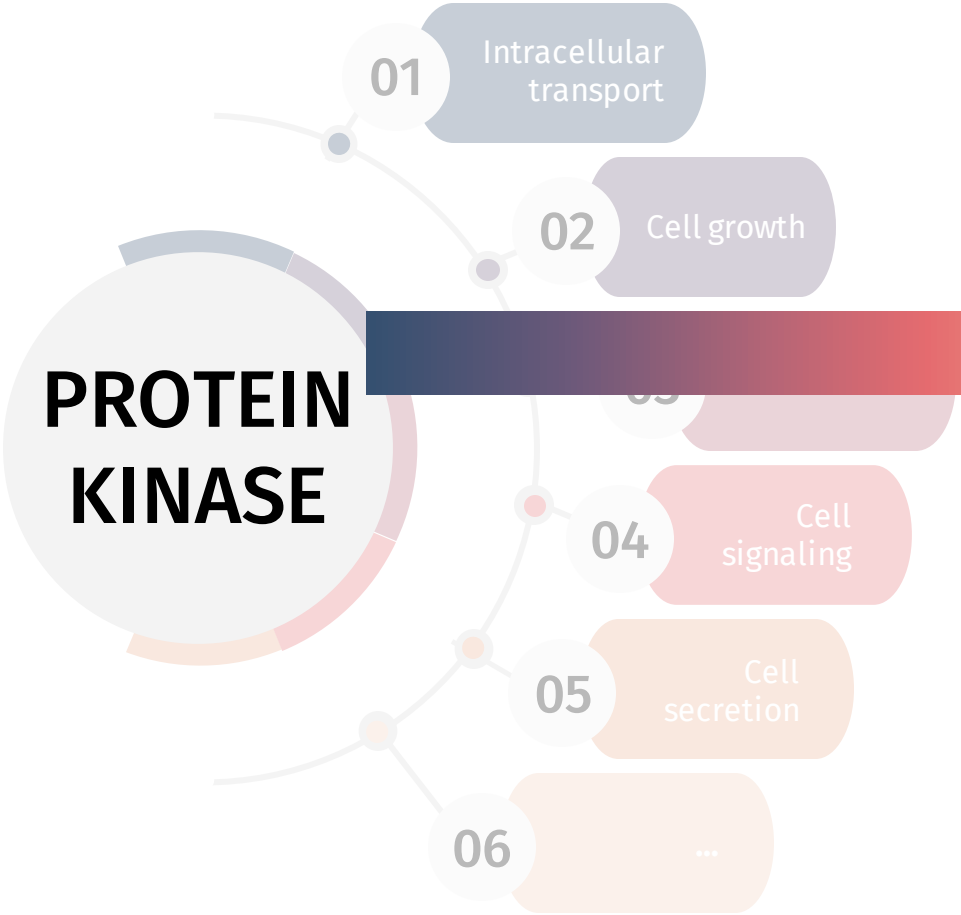
Application to the BCR-ABL ligand set

Towards DAG-based interactive pharmacophore exploration:

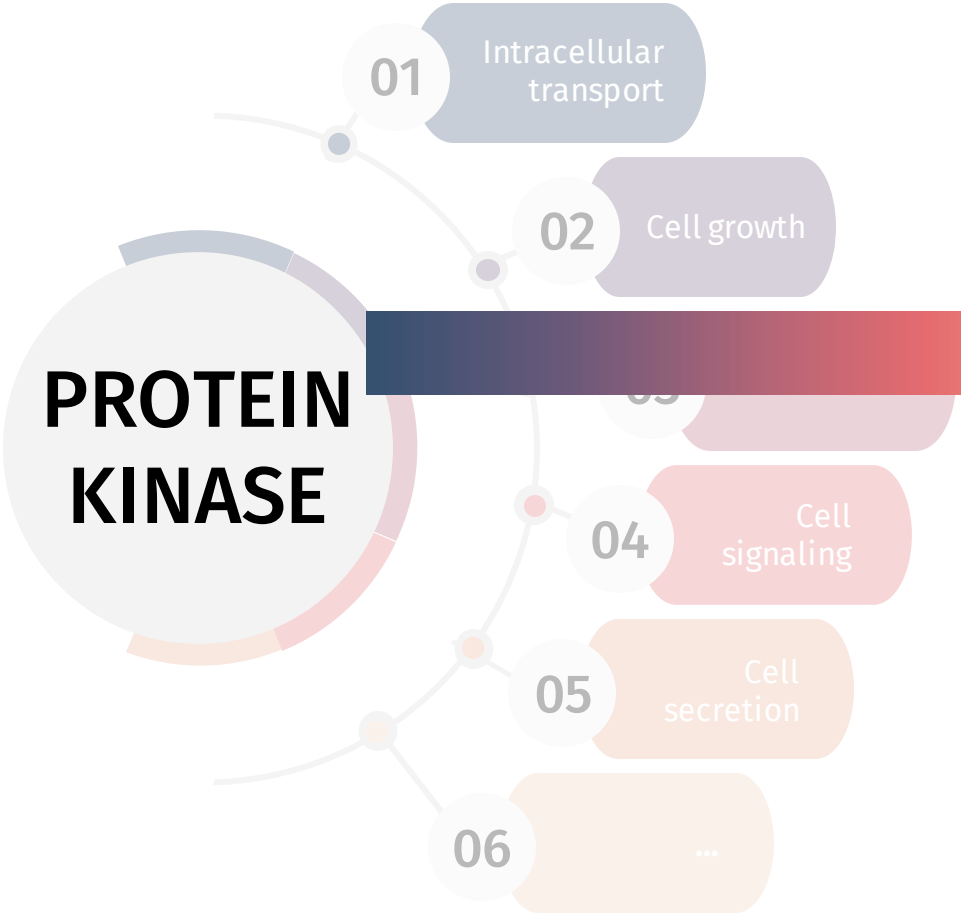
Application to the BCR-ABL ligand set

PROTEIN KINASE


PROTEIN KINASE
01
Intracellular transport
02
Cell growth
03
Metabolism
04
Cell signaling
05
Cell secretion
06
...



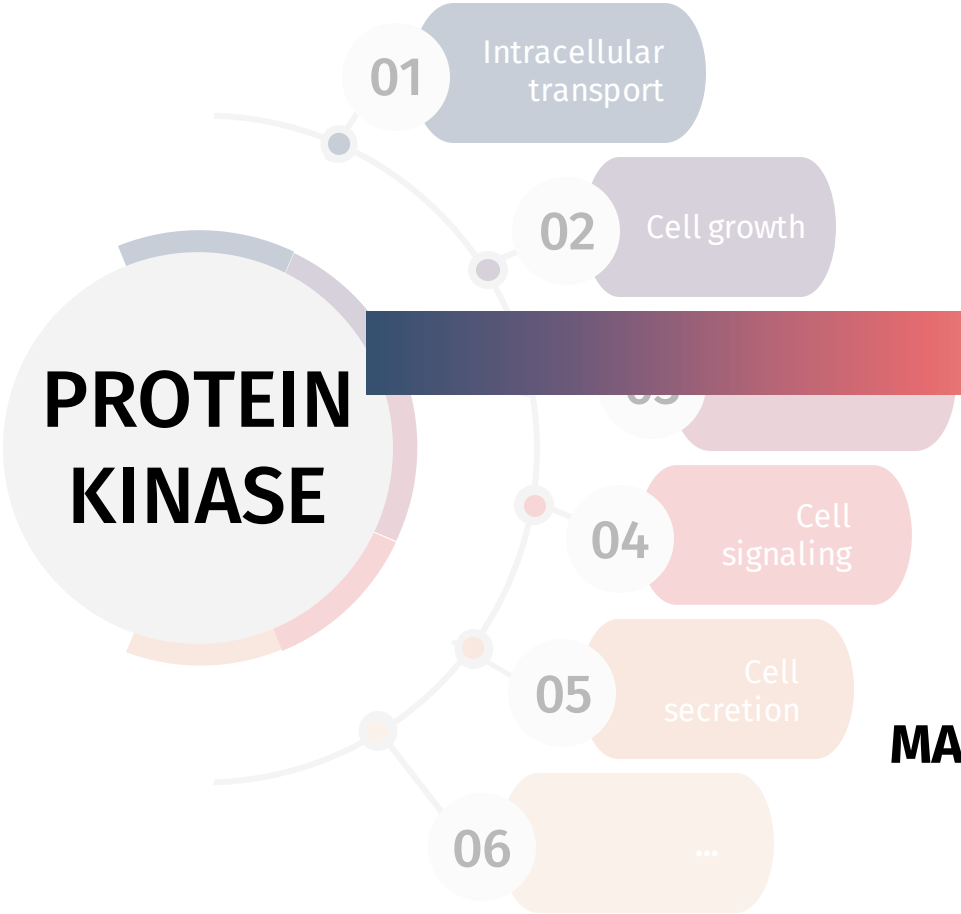
Pathological context



Pathological context

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

...



Pathological context

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

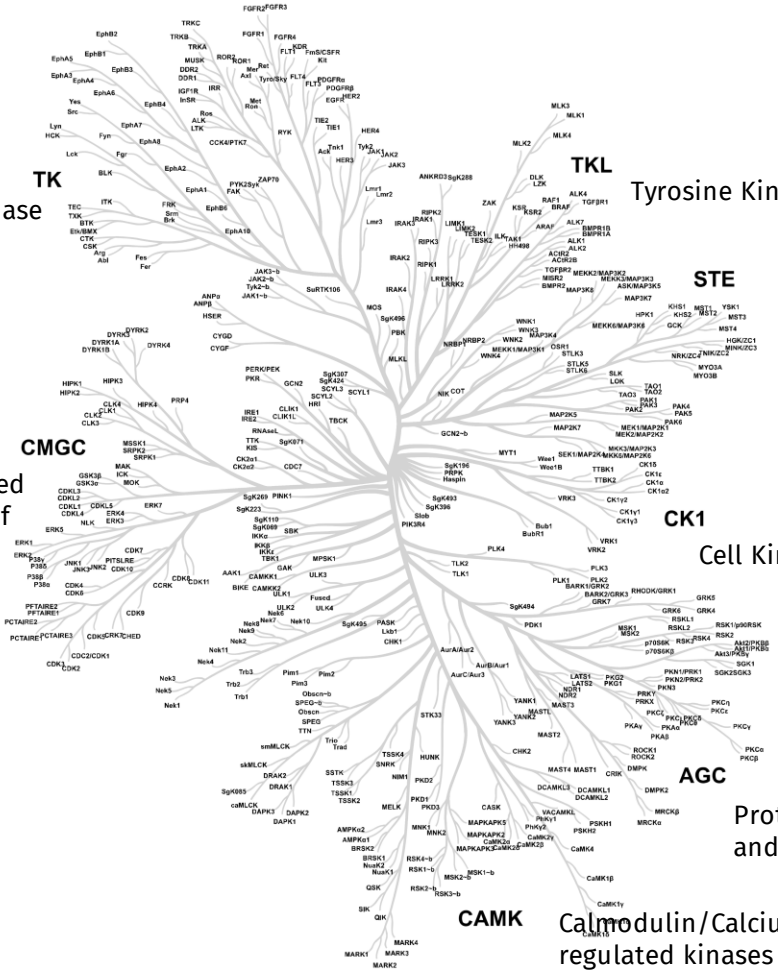
...

MAJOR Therapeutic interest

for kinase inhibitors

KINOME TREE

PROTEIN KINASE

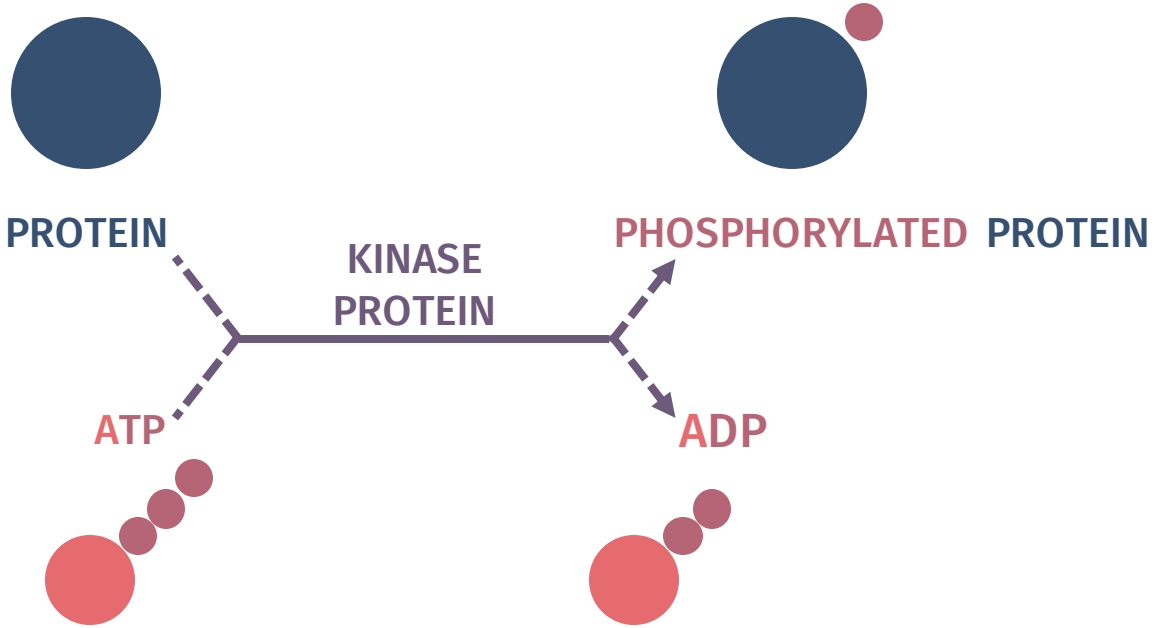


Representation made with CORAL web application ; Metz *et al.*, Cell Systems, 2018, 7 (3) ; doi.org/10.1016/j.cels.2018.07.001

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

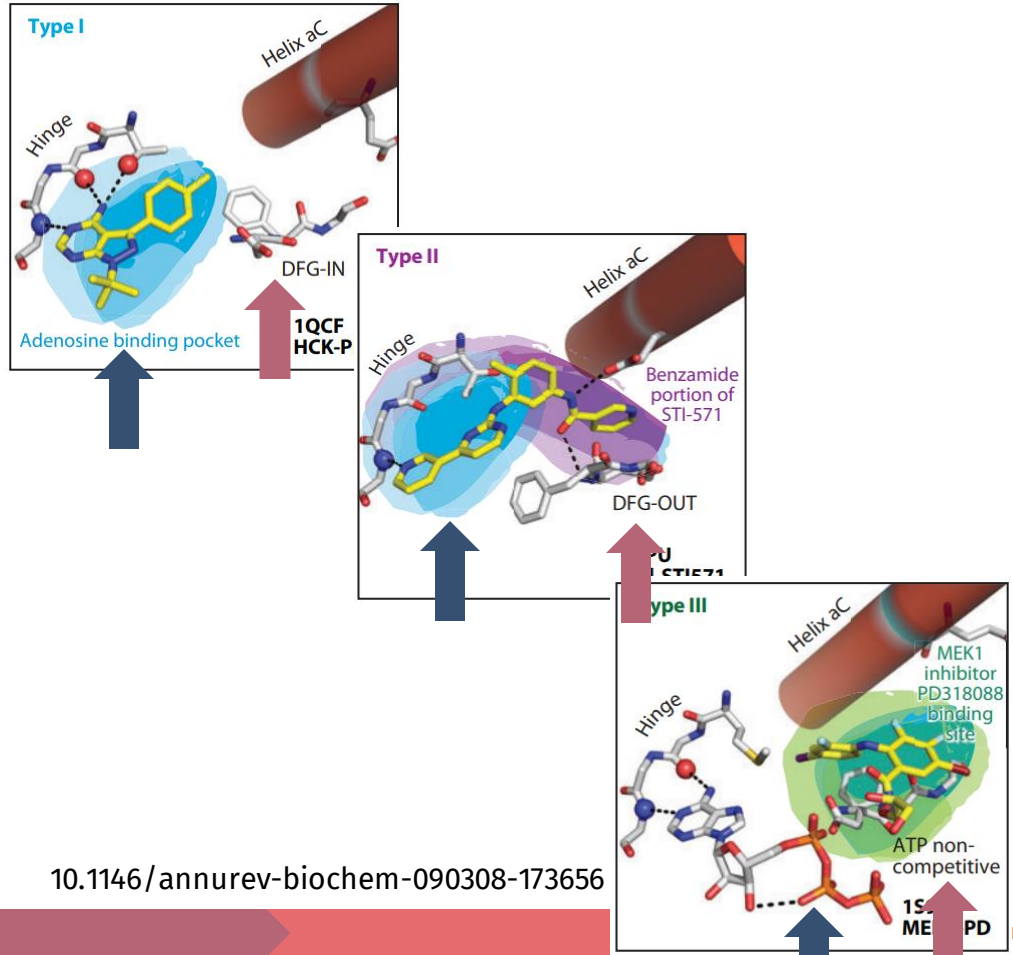
- ATP competitive inhibition
- DFG *in*

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 II

- ATP competitive inhibition
- DFG *out*

PKI – TYPE III

- Allosteric inhibitors
- Binding site adjacent to ATP binding site

PKI – TYPE IV

- Allosteric inhibitors
- Various sites of action

PKI – TYPE V

- Bivalent inhibitors
- ATP binding site + adjacent pocket

PKI – TYPE VI

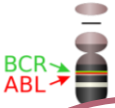
- Covalent binding with the protein (CYS, LYS or TYR)

●●●

BCR-ABL protein

BCR-ABL protein

Chromosome 22



Philadelphia chromosome

Chromosome 9

BCR gene

BCR = *Breakpoint Cluster Region*

BCR-ABL1 gene

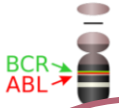
Fusion protein

ABL1 gene

ABL = *Abelson*

BCR-ABL protein

Chromosome 22



Mutation
t(9;22) (q34;q11)

Philadelphia chromosome

Chromosome 9

BCR gene

BCR =
Breakpoint Cluster Region

BCR-ABL1 gene

Fusion protein

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

CHRONIC MYELOID LEUKAEMIA (CML)

ABL1 gene

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

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

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

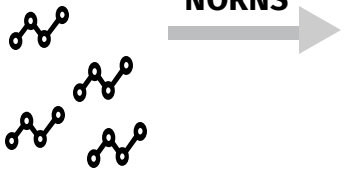
Métivier *et al.*, J Med Chem, 2018, 61

NORNS PROCESS

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

NORNS PROCESS

A



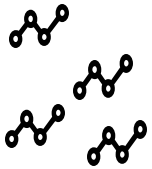
Molecular dataset

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

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

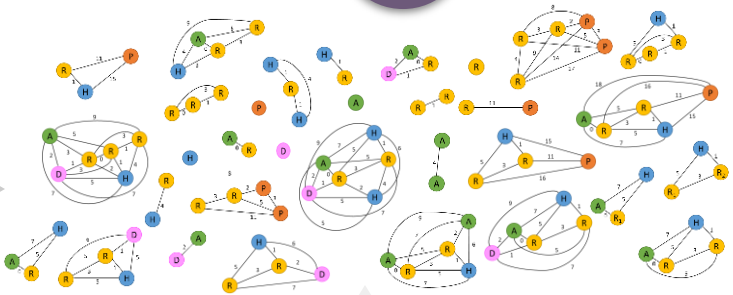
A



Molecular dataset

NORNS

B



Computation of the combinations of pharmacophoric features (2D pharmacophore)

Pharmacophore order
Number of motifs carried by the pharmacophore

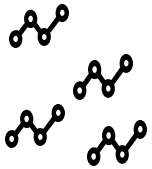
Frequency threshold
According to the number of molecules covered by the pharmacophore (support size)

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

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

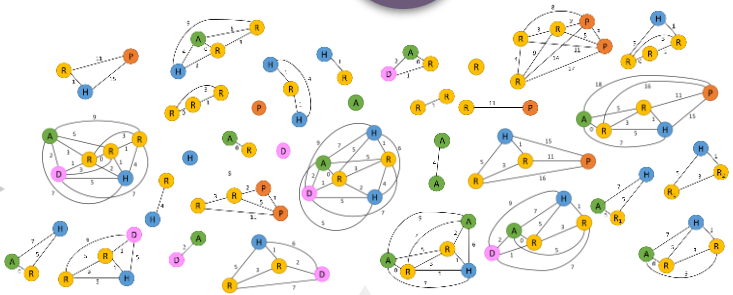
A



Molecular dataset

NORNS

B



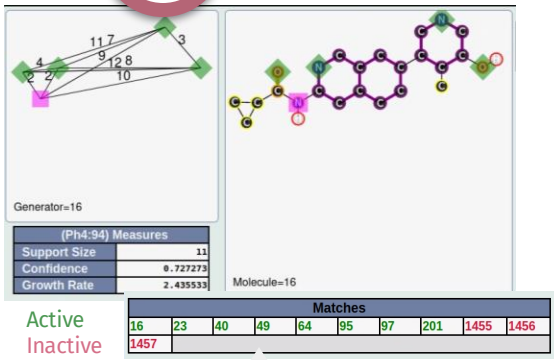
Computation of the combinations of pharmacophoric features (2D pharmacophore)

Pharmacophore order
Number of motifs carried by the pharmacophore

Frequency threshold
According to the number of molecules covered by the pharmacophore (support size)

NORNS

C

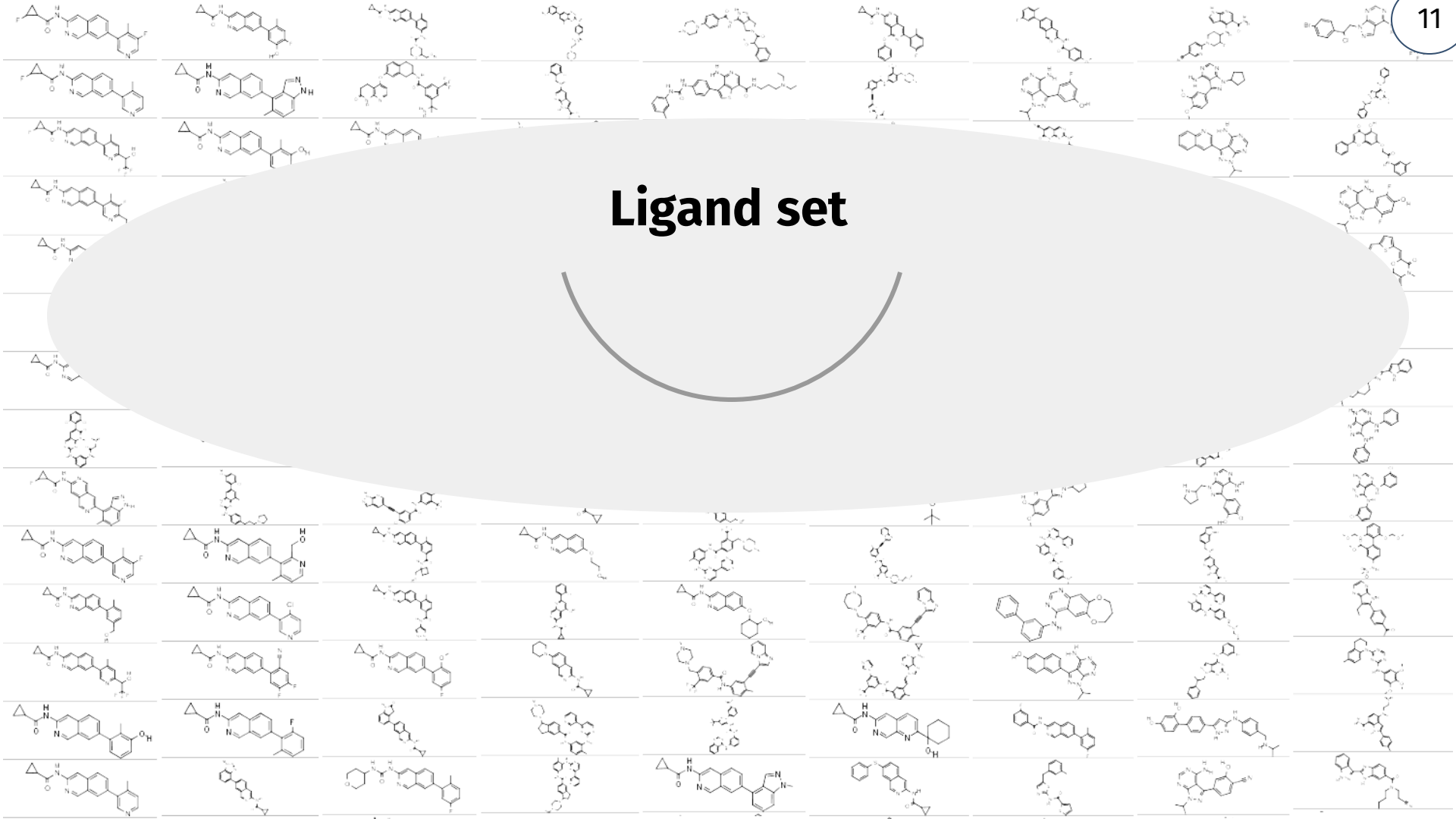


Growth rate
Ratio between fit frequencies of a pharmacophore within the 2 groups (active *vs* inactive)

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

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)

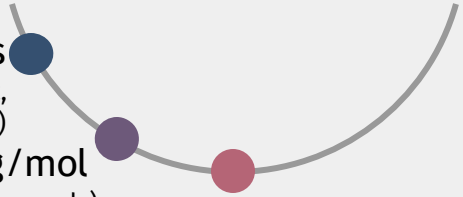
MW ≤ 800 g/mol
(15 255 records)

Ligand set

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

MW ≤ 800 g/mol
(15 255 records)

K_i or IC_{50}



Ligand set

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

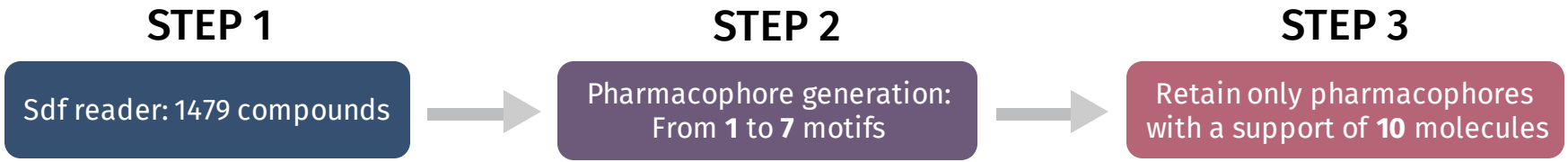
MW ≤ 800 g/mol
(15 255 records)

K_i or IC_{50}

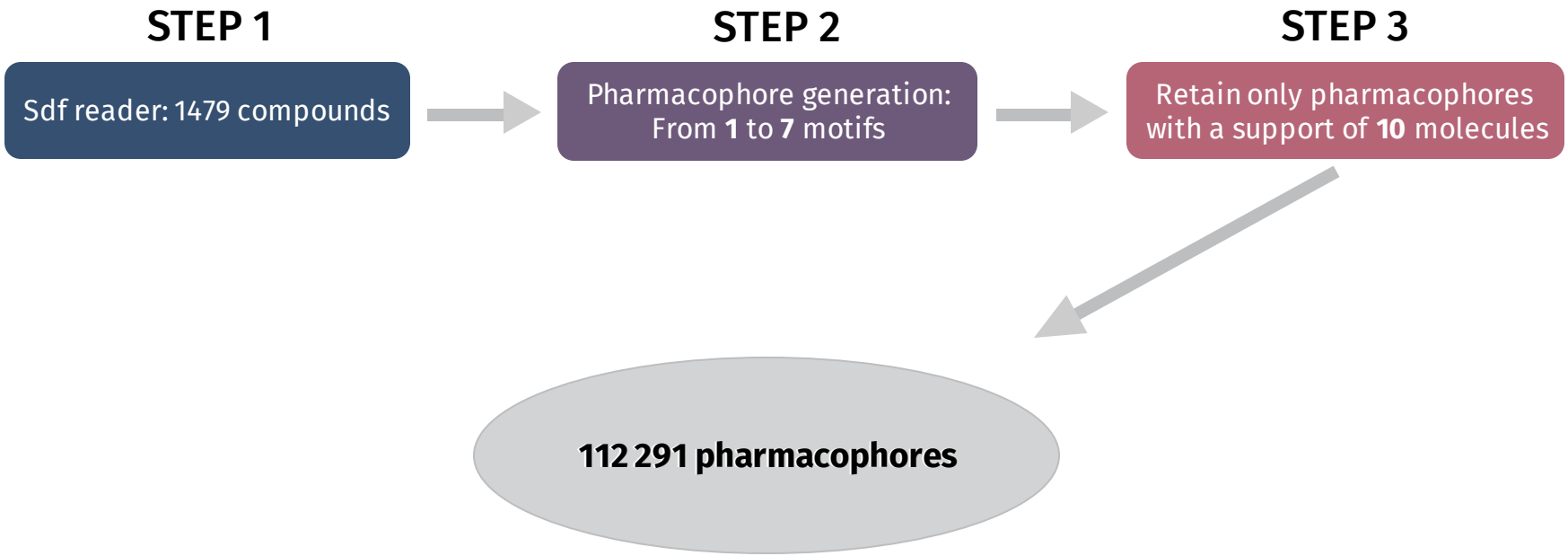
1479 compounds

773 of which with K_i or IC_{50} ≤ 100 nM

NORNS pipeline

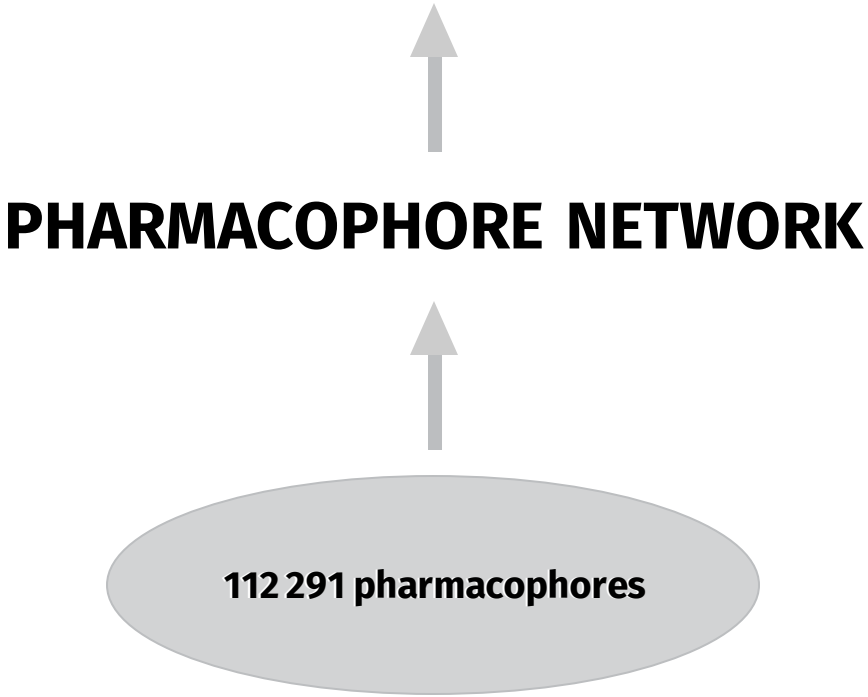


NORNS pipeline



112 291 pharmacophores

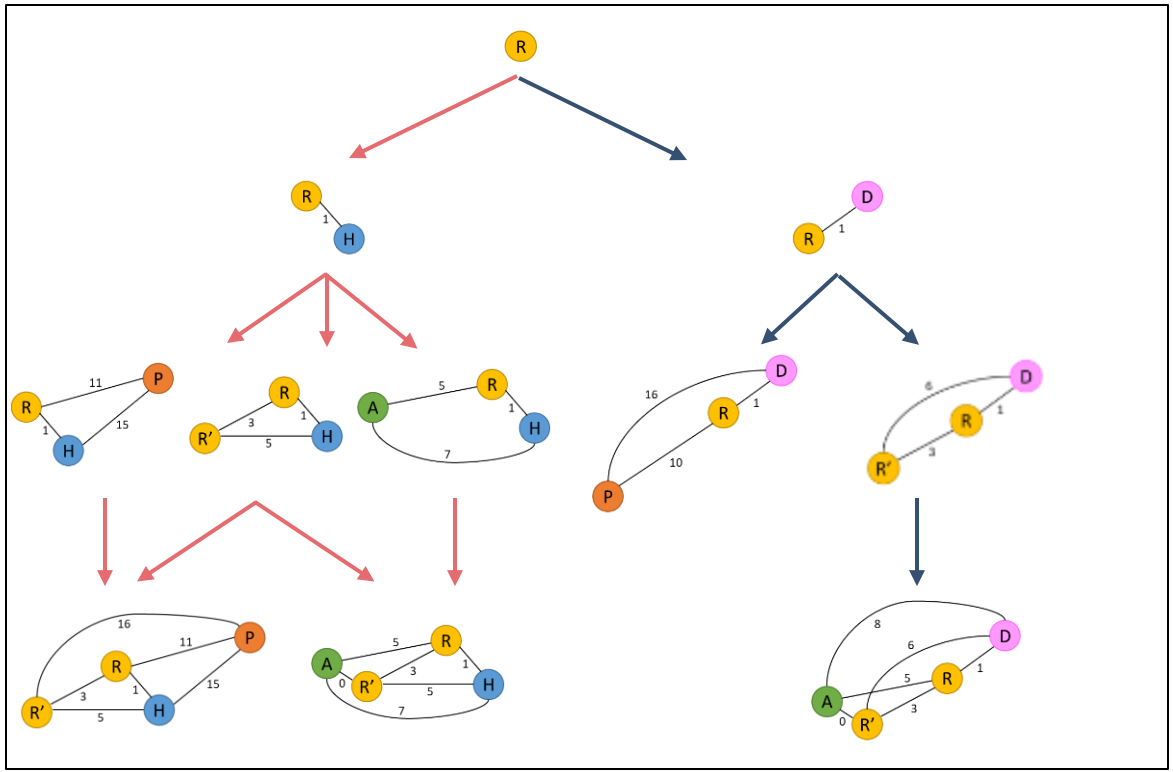
Directed **A**cyclic **G**raph (DAG)



Directed Acyclic Graph (DAG)

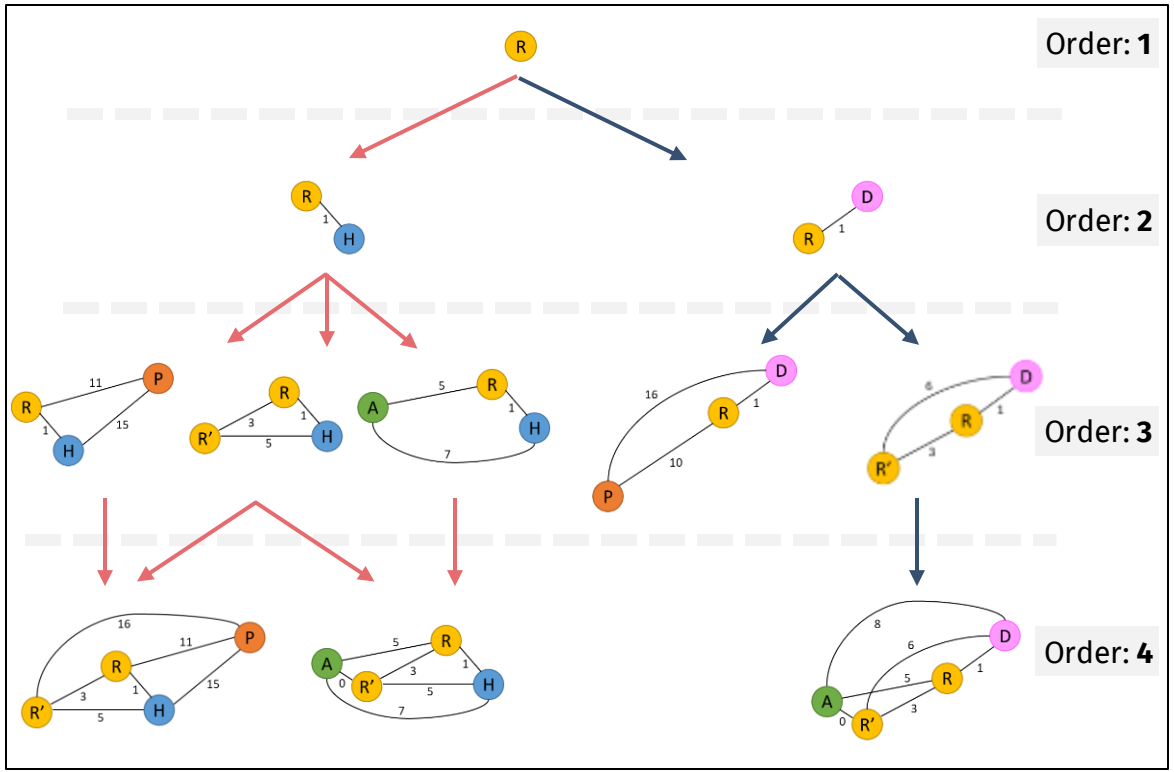
- H Hydrophobic group
- R Aromatic ring
- A Hydrogen-bond acceptor
- D Hydrogen-bond donor
- P Positively-ionizable group

Directed Acyclic Graph (DAG)



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

Directed Acyclic Graph (DAG)

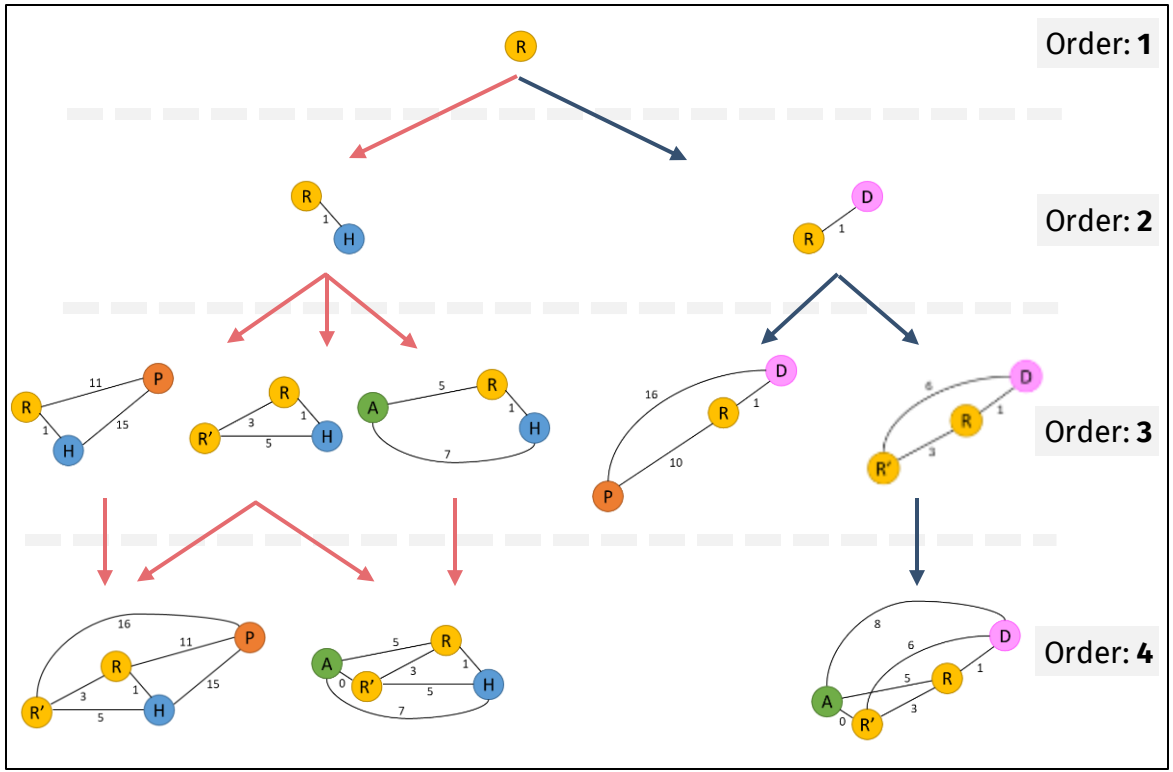


From top to bottom:

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

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

Directed Acyclic Graph (DAG)



From top to bottom:

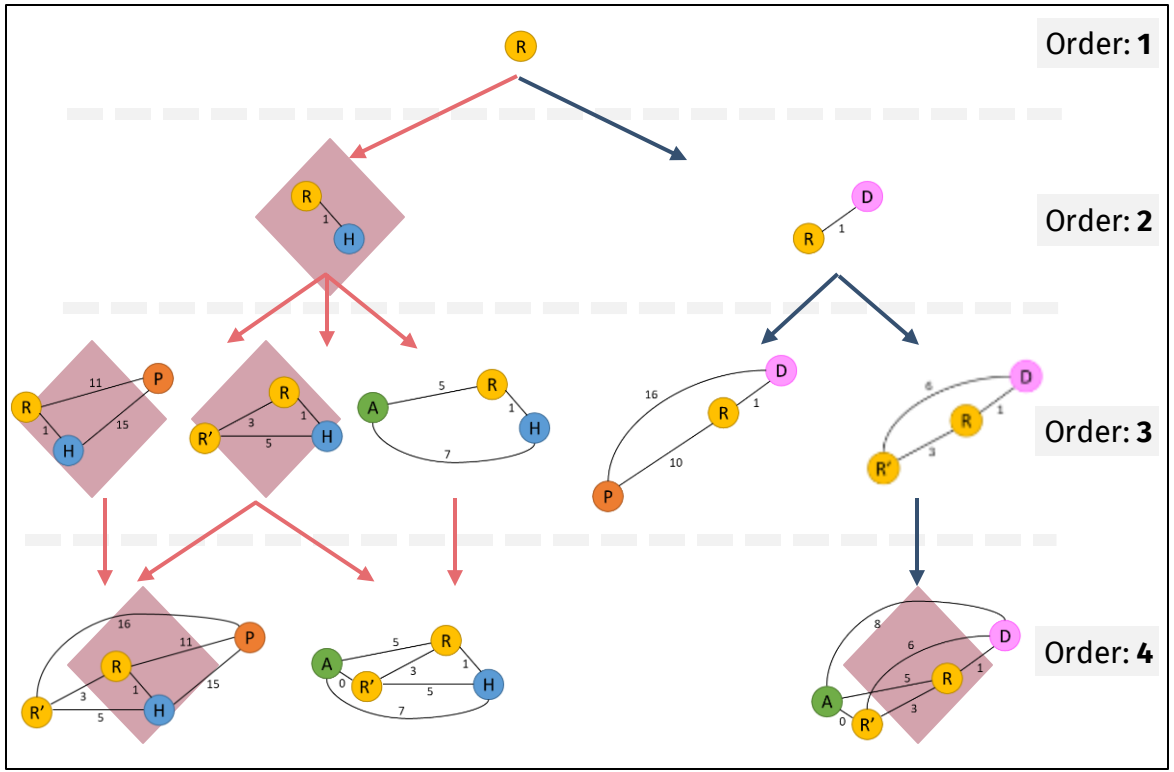
Two pharmacophores are linked if one is included in the other.

Considering molecule subsets:

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

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

Directed Acyclic Graph (DAG)



From top to bottom:

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

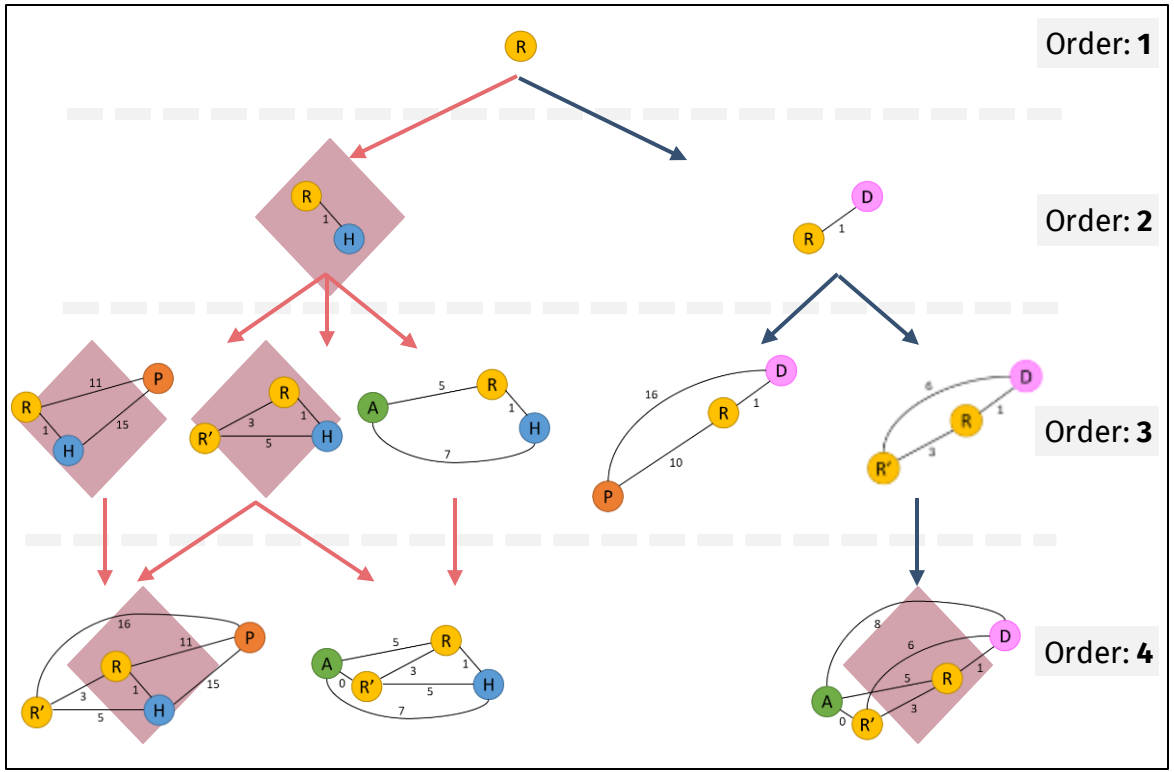
Considering molecule subsets:

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

Molecule subset 1

- 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.

Considering molecule subsets:

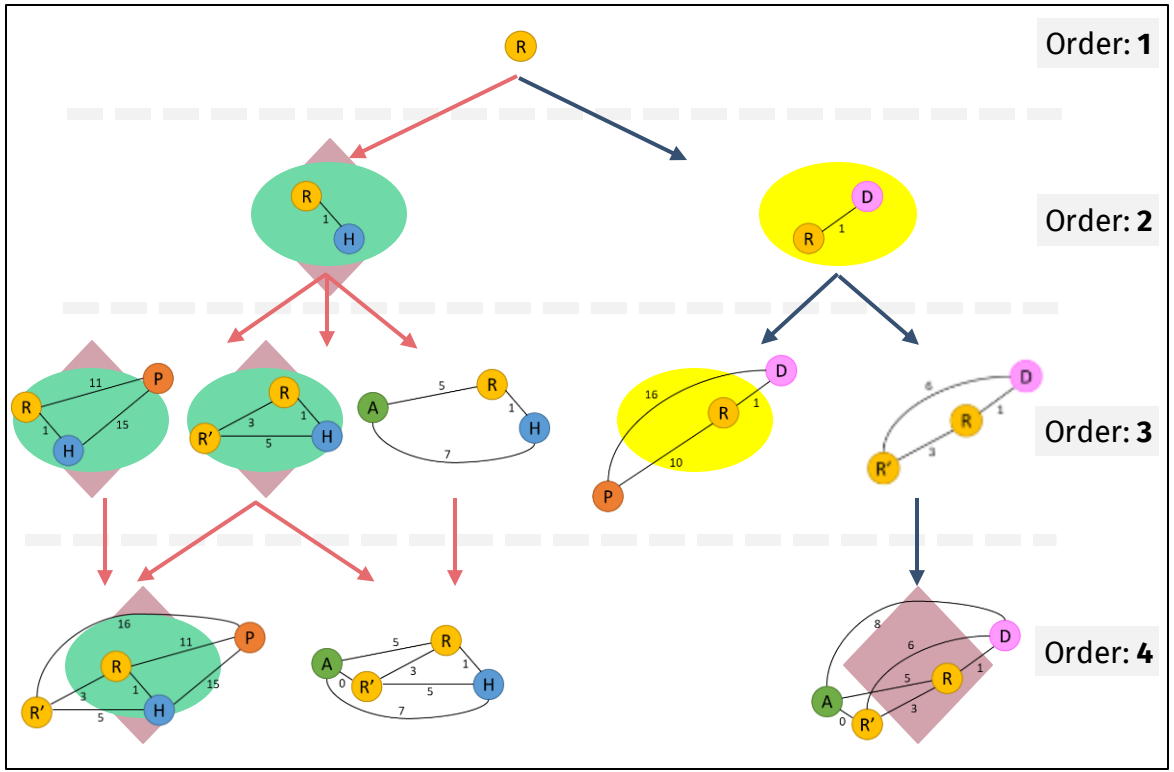
Distinct pharmacophores covering the exact same molecule subset form a General Equivalent Class (GEC).

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

Molecule subset 1

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

Directed Acyclic Graph (DAG)



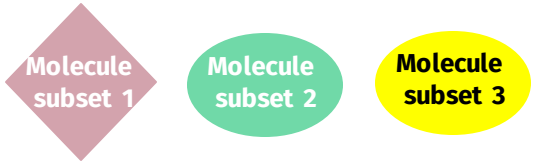
From top to bottom:

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

Considering molecule subsets:

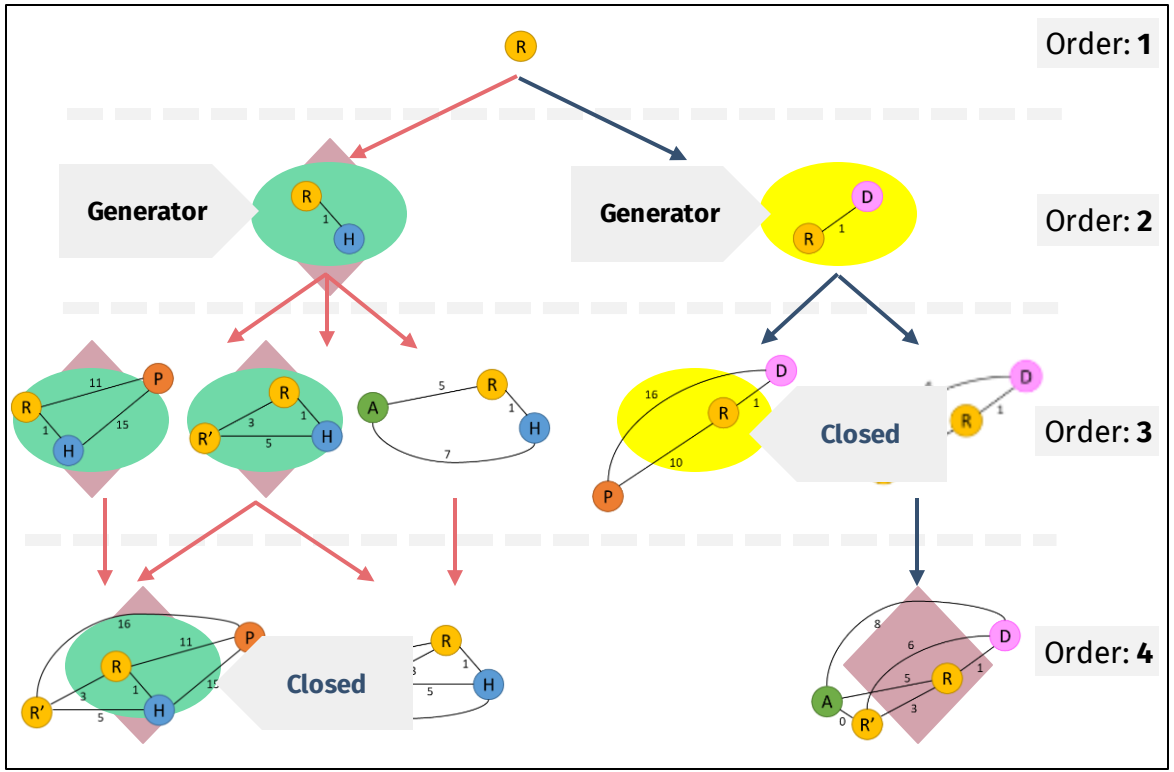
Distinct pharmacophores covering the exact same molecule subset form a General Equivalent Class (GEC).

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



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

Directed Acyclic Graph (DAG)



⇒ **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).

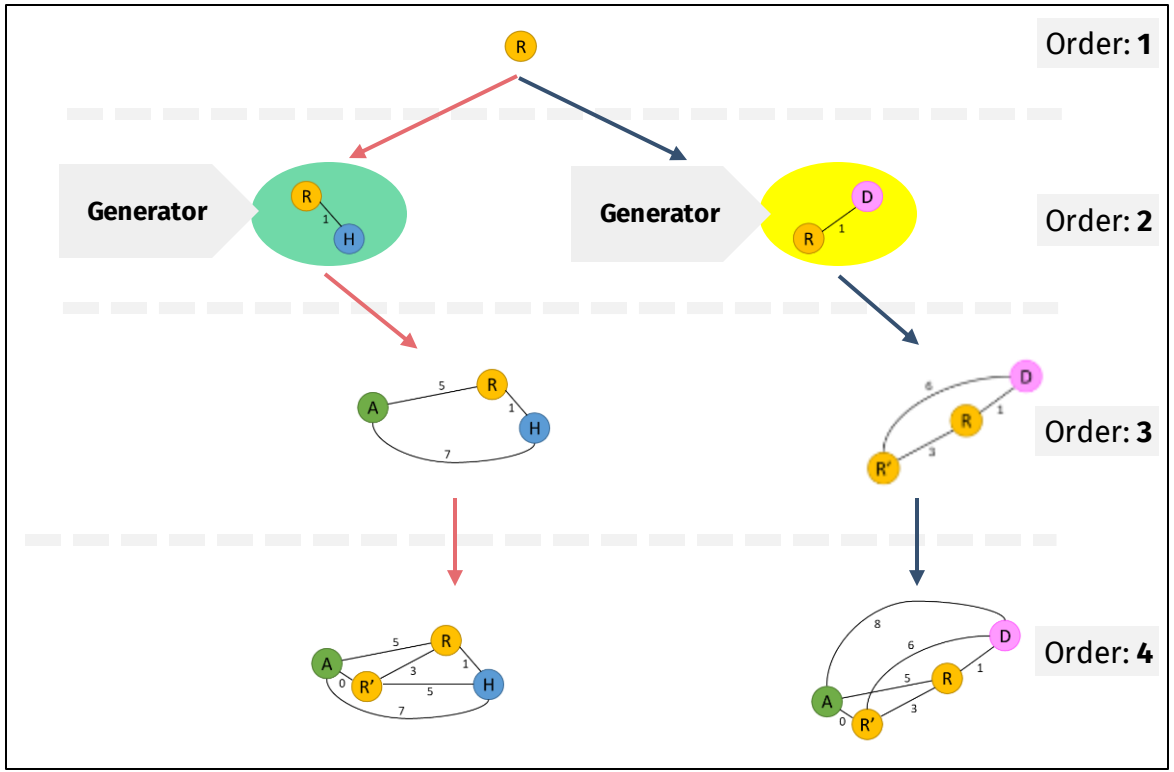
Molecule subset 1

Molecule subset 2

Molecule subset 3

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

Directed Acyclic Graph (DAG)



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

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

⇒ Consideration of the greatest pharmacophoric description : Closed

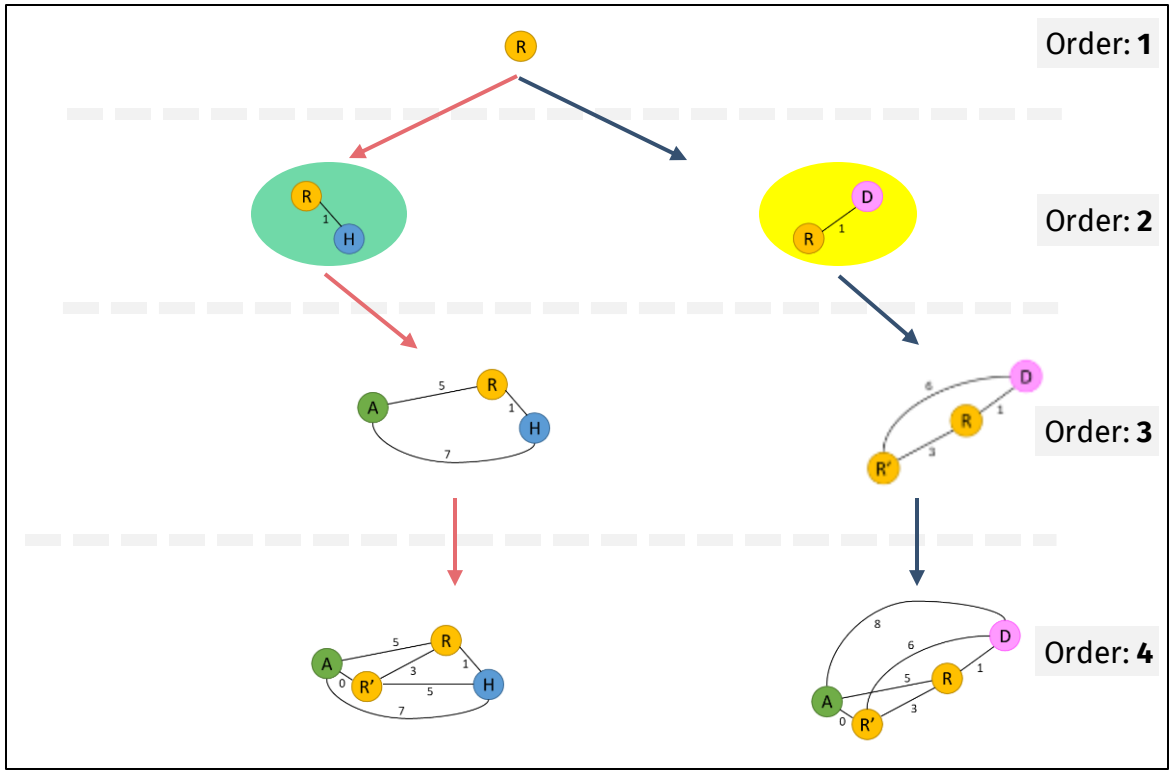
Distinct pharmacophores covering the exact same molecule subset and sharing a family relationship form a Structured Equivalent Class (SEC).

Molecule subset 2

Molecule subset 3

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

Directed Acyclic Graph (DAG)

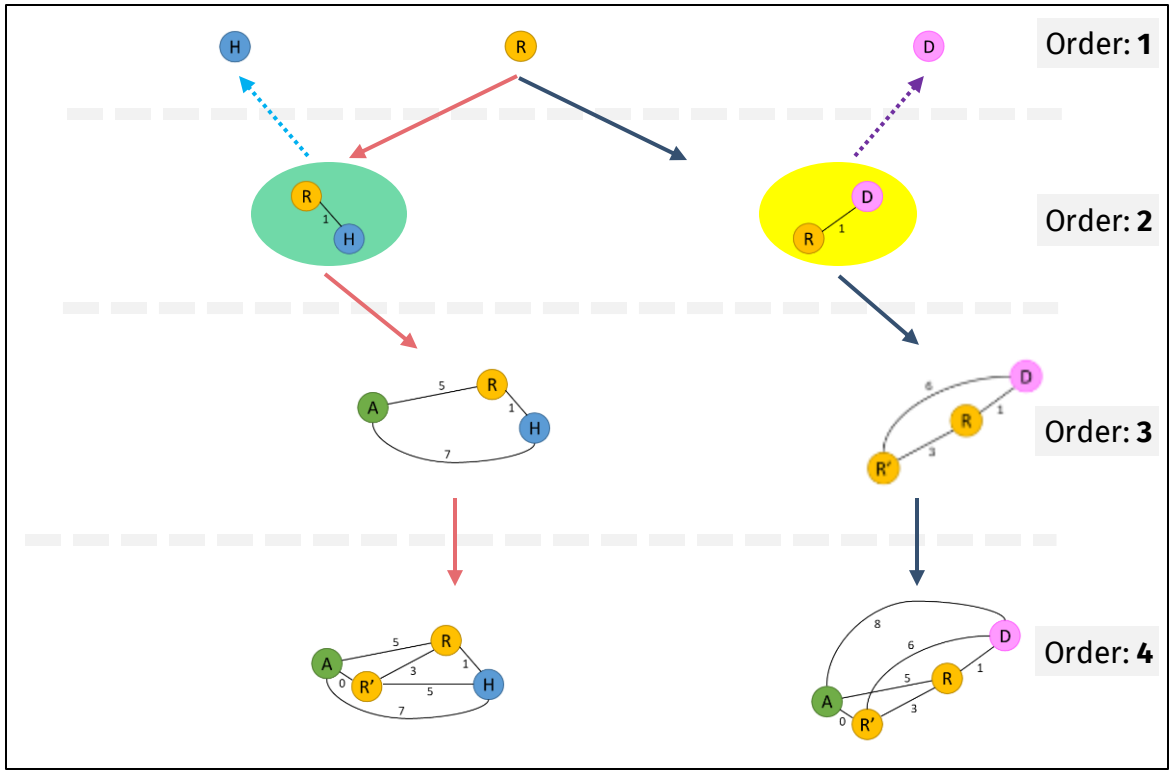


Siblings concept

⇒ 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

Directed Acyclic Graph (DAG)



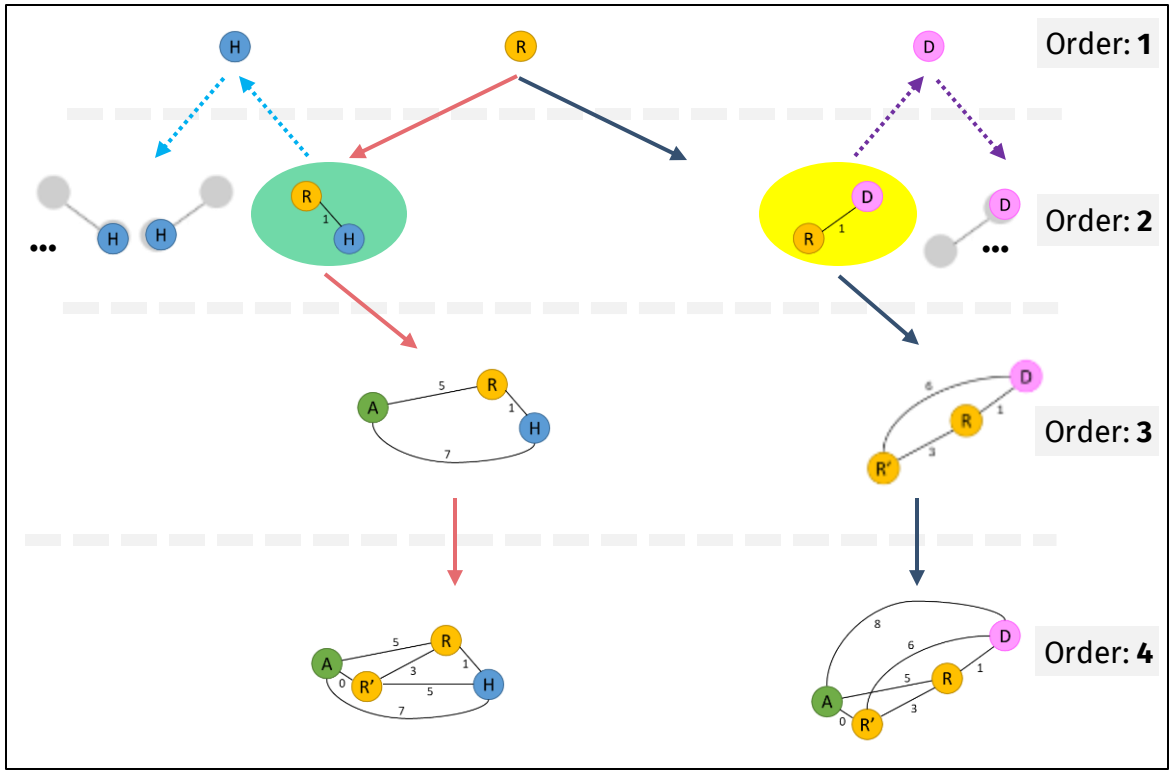
⇒ From a **relational DAG** to a **SEC-clusterized relational diagram**

Siblings concept

Consideration of pharmacophore *ancestors* (*parents*)

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**

Siblings concept

Consideration of pharmacophore *ancestors* (*parents*)

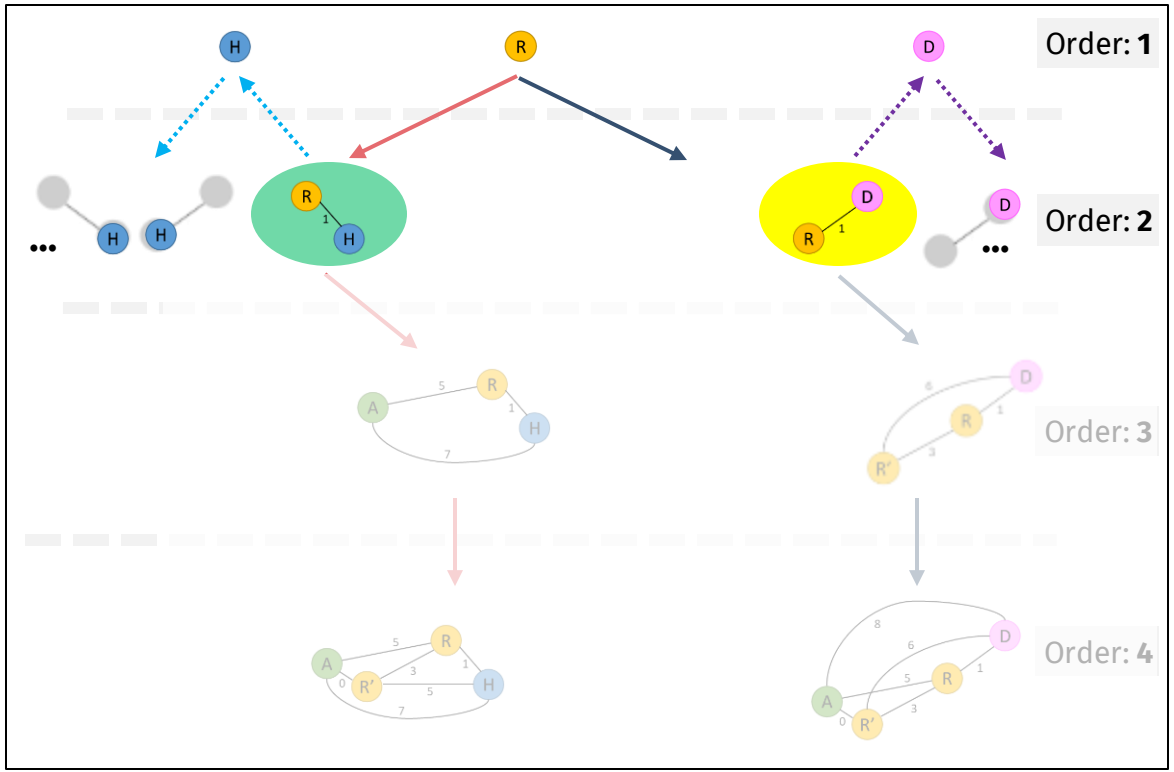
↓

Listing of all pharmacophore *successors* (*children*)

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

SEC = **S**tructured **E**quivalent **C**lass

Directed Acyclic Graph (DAG)



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

Siblings concept

Consideration of pharmacophore *ancestors* (*parents*)

↓

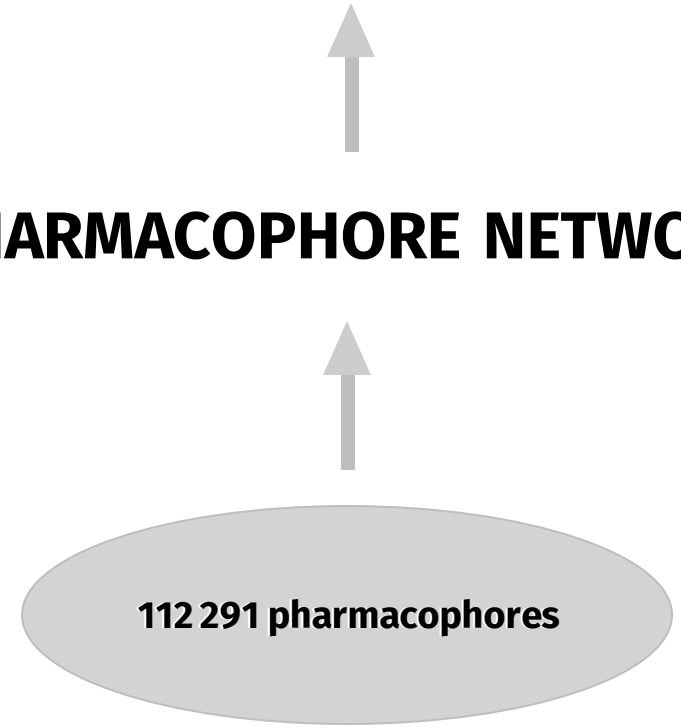
Listing of all pharmacophore *successors* (*children*)

Same order

SEC

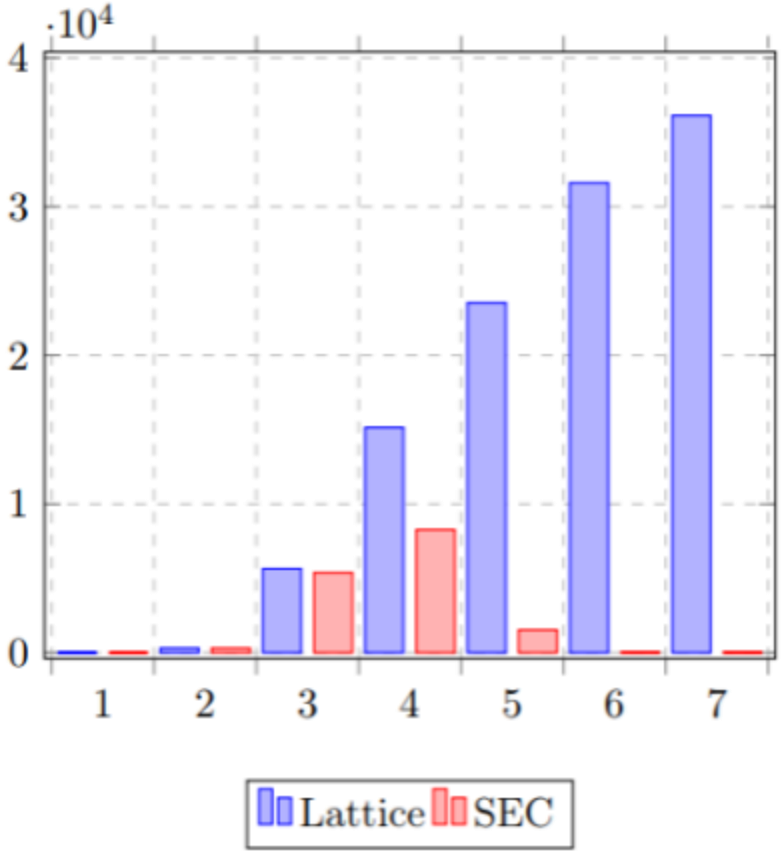
NORNS
LATTICE

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



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

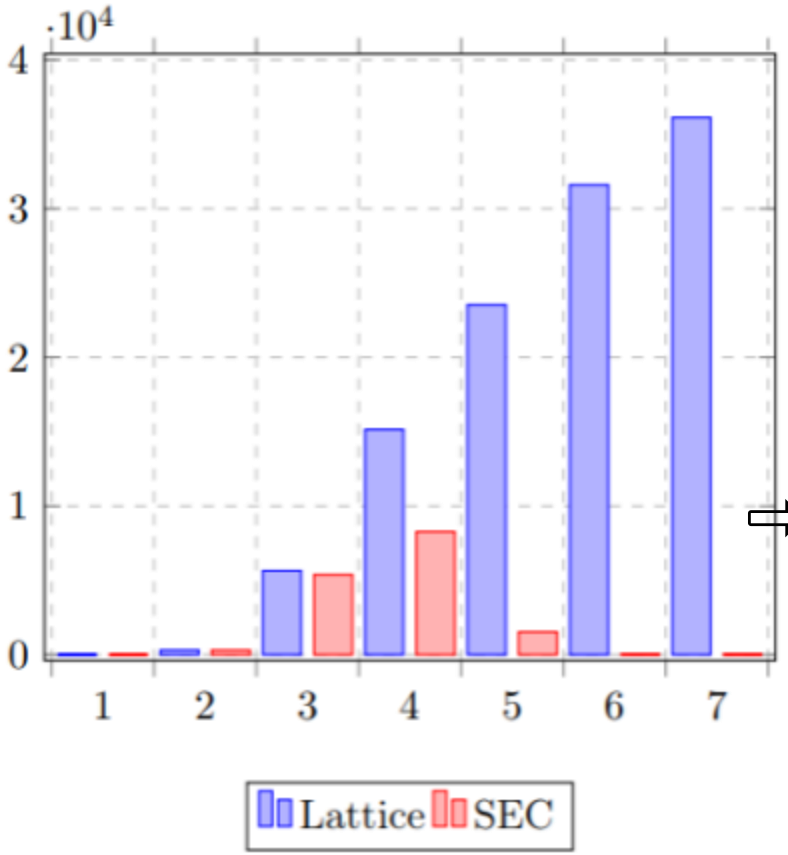
NORNS
LATTICE



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

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

NORNS
LATTICE



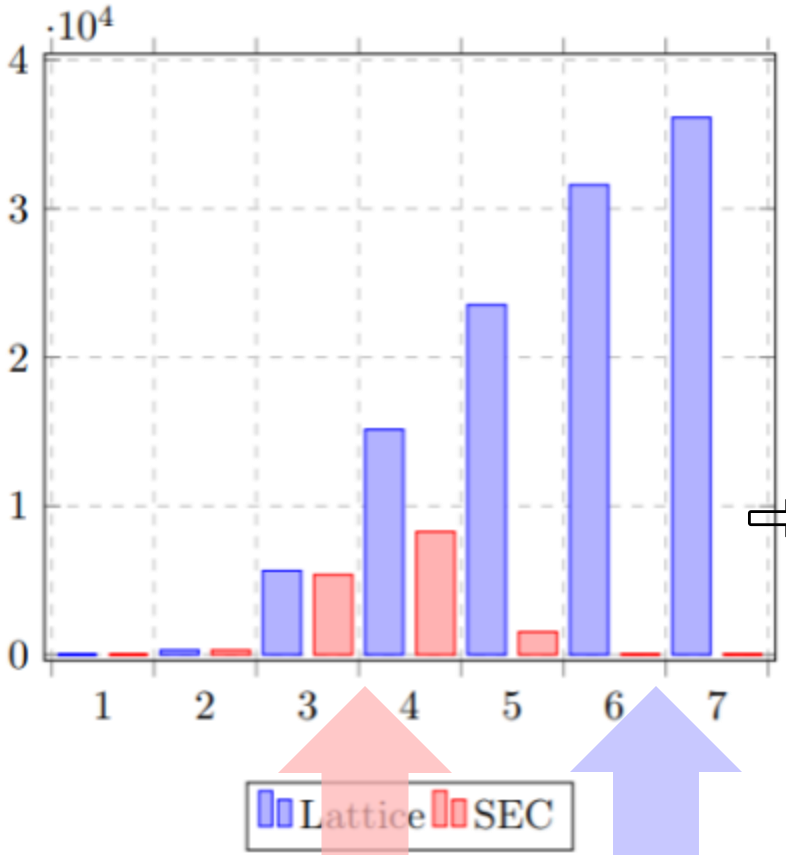
Distinct pharmacophores bearing the exact same molecule subset and sharing a family relationship form a Structured Equivalent Class (SEC).

112 291 pharmacophores → **15 477 SEC (generators)**

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

Directed **A**cyclic **G**raph (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**

Directed Acyclic Graph (DAG)

NORNS LATTICE

LOOKING FOR REMARKABLE PHARMACOPHORES

15 477 SEC

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**cyclic **G**raph (DAG)

NORNS
LATTICE

LOOKING FOR REMARKABLE PHARMACOPHORES

15 477 SEC

...

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)

NORNS
LATTICE

FROM SEC TO PAD ?

15 477 SEC

SEC = **S**tructured **E**quivalent **C**lass
PAD = Pharmacophore Activity Delta

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

NORNS
LATTICE

FROM SEC TO PAD ?

15 477 SEC

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

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

SEC = **S**tructured **E**quivalent **C**lass
PAD = Pharmacophore Activity Delta

Directed **A**cyclic **G**raph (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 = **S**tructured **E**quivalent **C**lass
PAD = Pharmacophore Activity Delta

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

Directed Acyclic Graph (DAG)

NORNS LATTICE

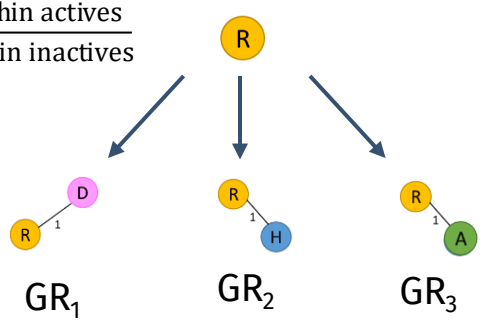
FROM SEC TO PAD ?

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 = **S**tructured **E**quivalent **C**lass
PAD = Pharmacophore Activity Delta

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

Directed Acyclic Graph (DAG)

NORNS *LATTICE*

FROM SEC TO PAD ?

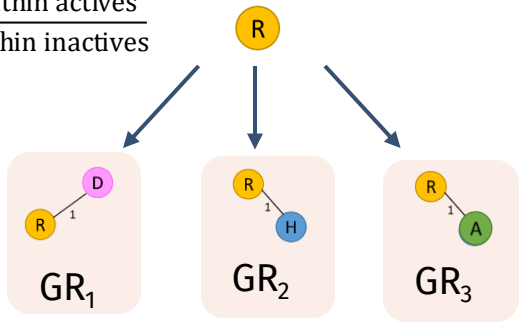
15 477 SEC

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 $|GR_i - \mu| \geq 2$ SD then **i** is a **PAD**

SEC = **S**tructured **E**quivalent **C**lass
PAD = Pharmacophore Activity Delta

6 essential pharmacophoric motifs
- Hydrophobic group
- Aromatic cycle
- Hydrogen-bond donor
- Hydrogen-bond acceptor
- Negatively-ionizable function
- Positively-ionizable function

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

NORNS *LATTICE*

FROM SEC TO PAD ?

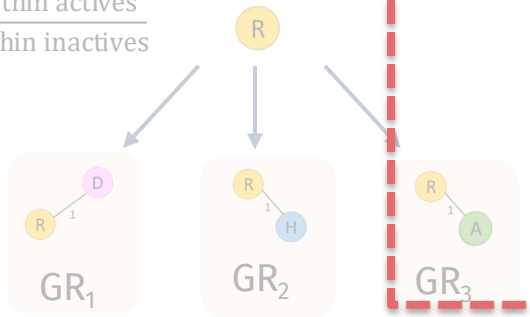
15 477 SEC

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 - μ ≥ 2 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)

NORNS
LATTICE

FROM SEC TO PAD ?

32 PAD

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

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

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

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

FROM SEC TO PAD ?

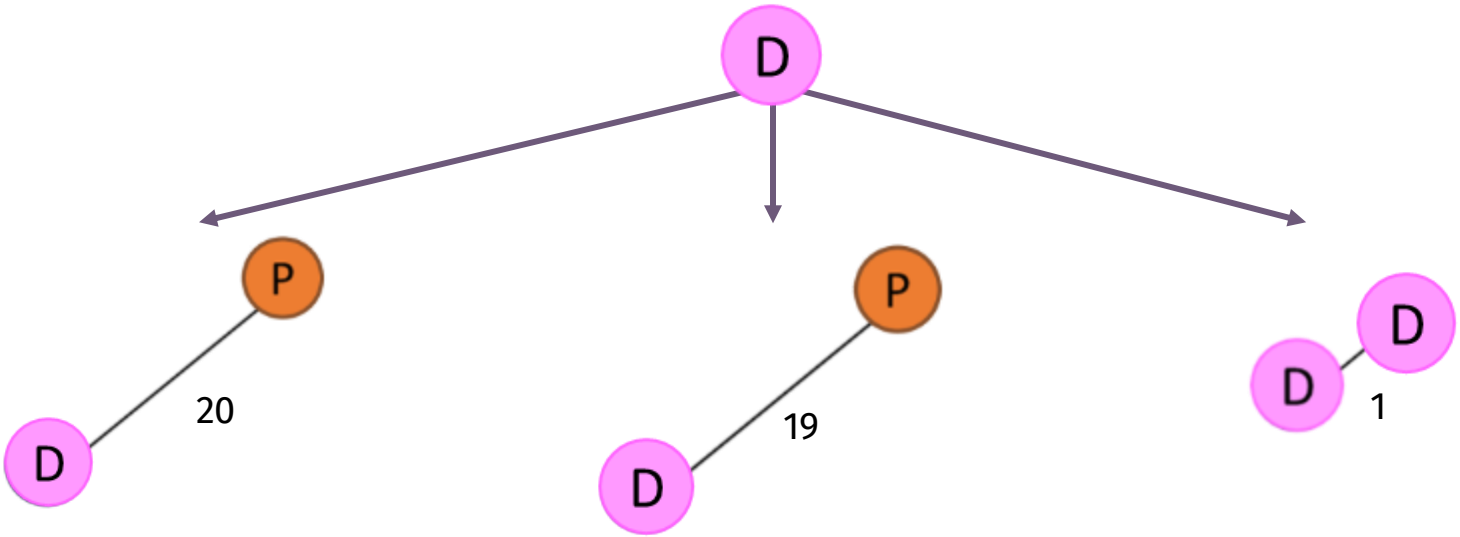
One example in O2 PADs

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

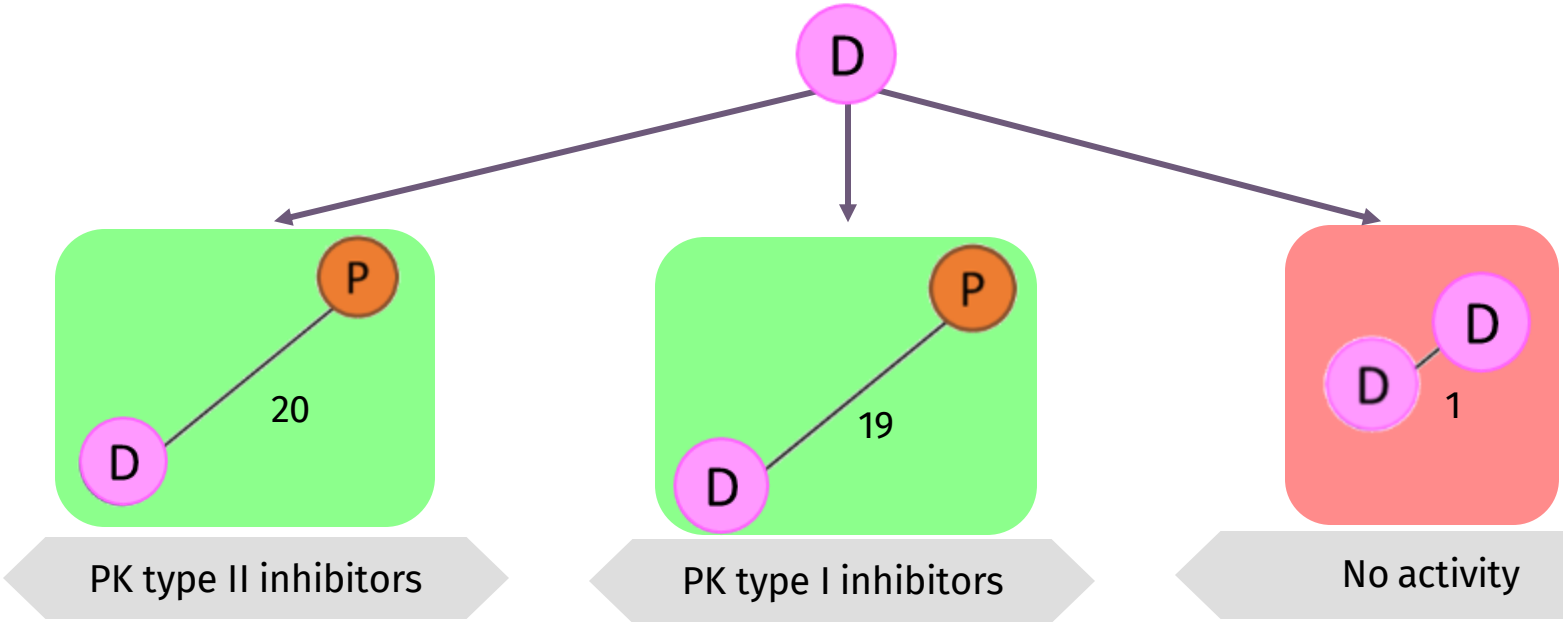
6 essential pharmacophoric motifs
- Hydrophobic group
- Aromatic cycle
- Hydrogen-bond donor
- Hydrogen-bond acceptor
- Negatively-ionizable function
- Positively-ionizable function

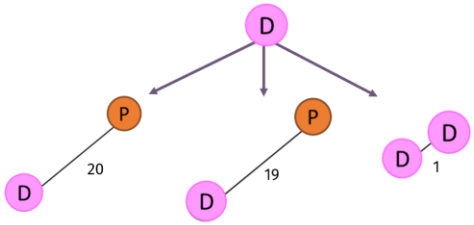
Active

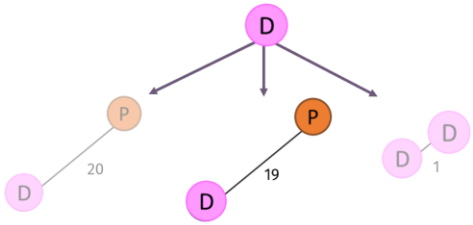
Inactive

FROM SEC TO PAD ?

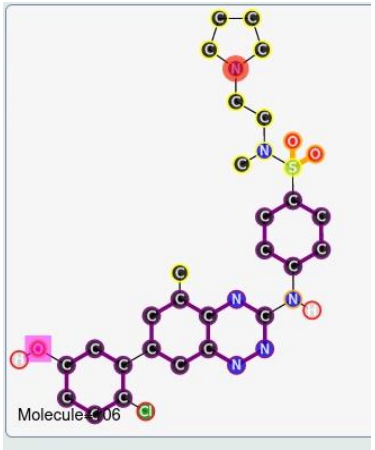
One example in O2 PADs




D
P
P
D
D
D
D
20
19
1



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

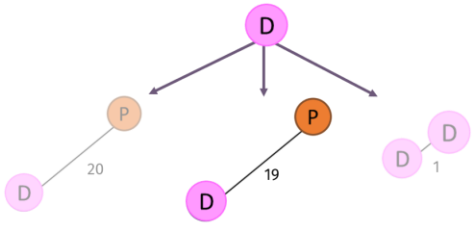


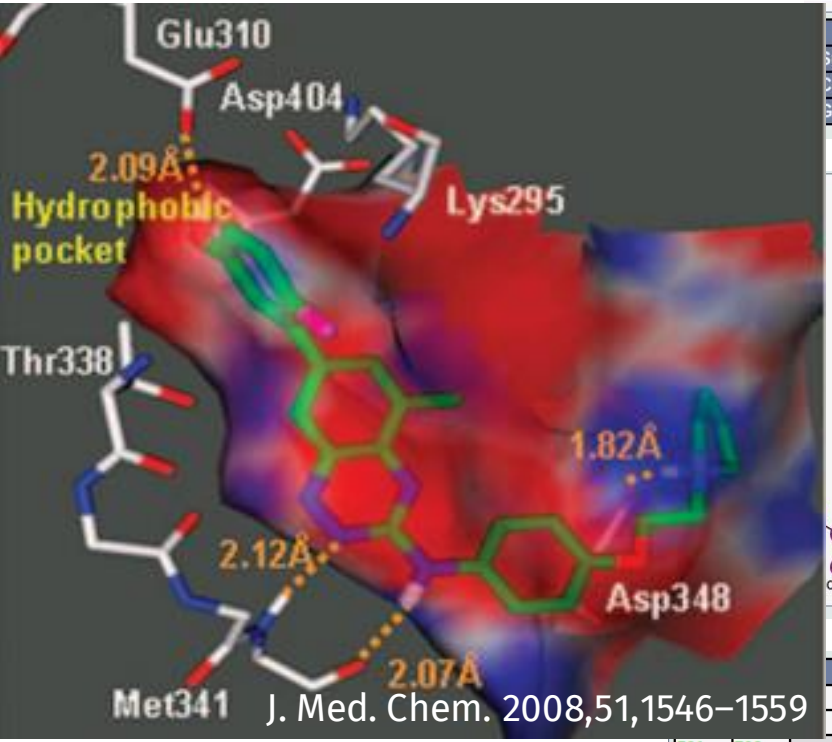
Alkaline moiety able to interact with an acid residue

Phenylbenzotriazinamine linker able to angle properly the two terminal moities

Constant phenol in terminal position

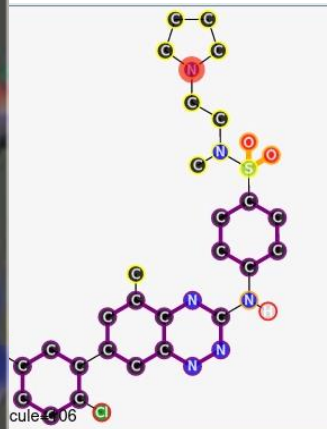
Matches									
106	132	155	244	254	266	267	327	331	345
355	357	358	381	406	423	482	512	567	583
591	702								





J. Med. Chem. 2008,51,1546–1559

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

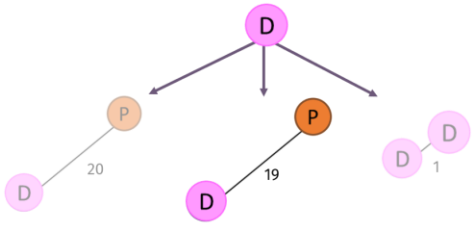


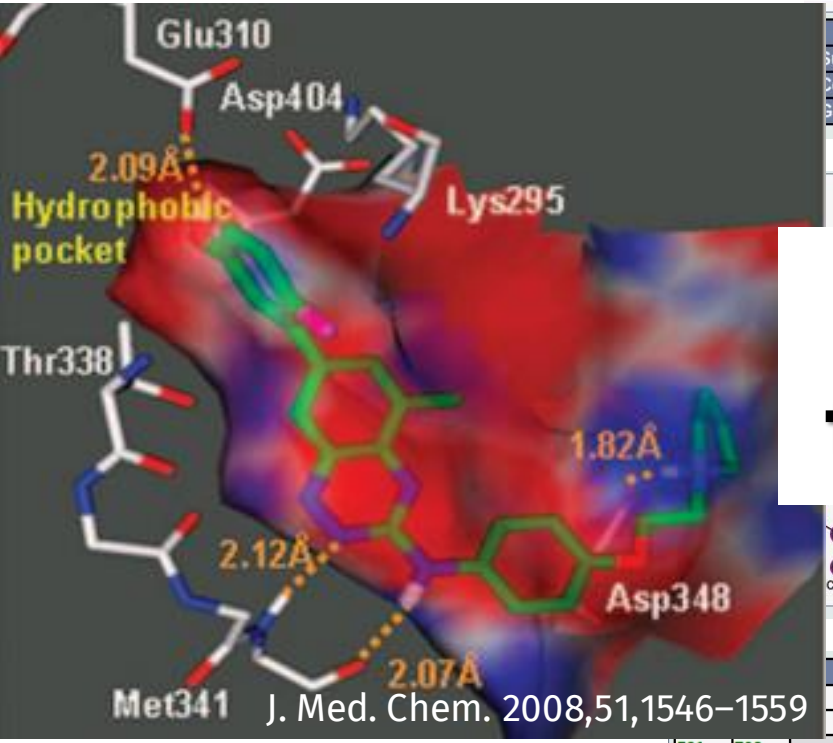
Matches							
244	254	266	267	327	331	345	
381	406	423	482	512	567	583	
591	702						

Alkaline moiety able to interact with an acid residue

Phenylbenzotriazinamine linker able to angle properly the two terminal moities

Constant phenol in terminal position





J. Med. Chem. 2008,51,1546–1559

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

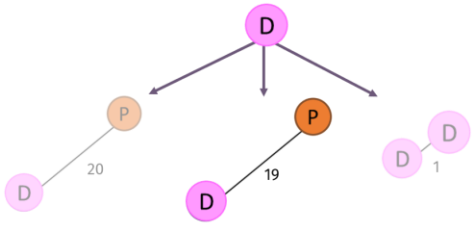
PKI type I

Matches							
244	254	266	267	327	331	345	
381	406	423	482	512	567	583	
591	702						

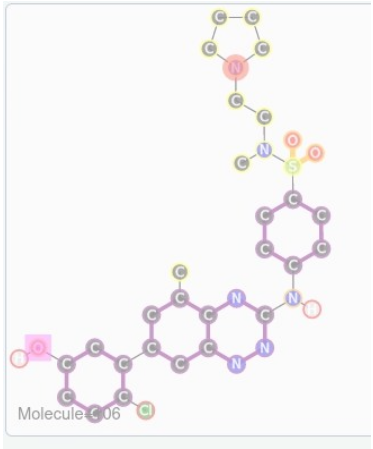
Alkaline moiety able to interact with an acid residue

Phenylbenzotriazinamine linker able to angle properly the two terminal moities

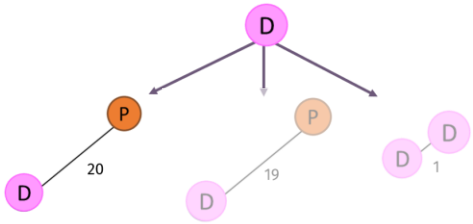
Constant phenol in terminal position



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

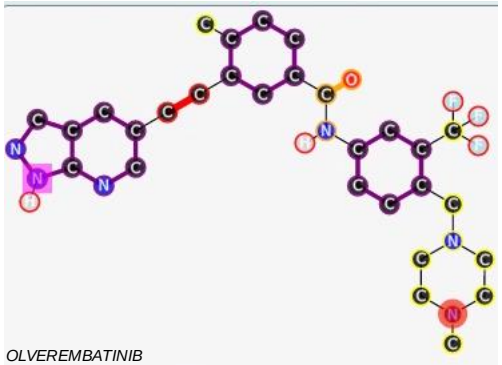


Matches									
106	132	155	244	254	266	267	327	331	345
355	357	358	381	406	423	482	512	567	583
591	702								

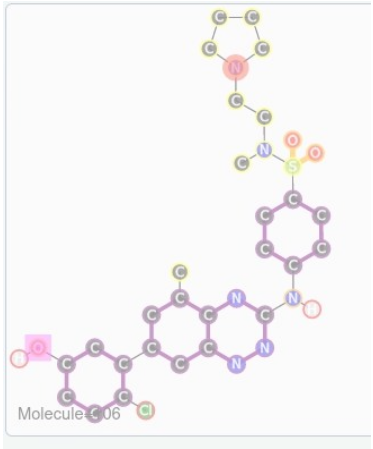


(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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



OLVEREMBATINIB



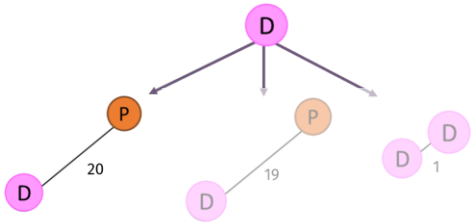
Pyrrazolopyridine able to interact with hydrophilic residues

Constant central benzene for hydrophobic contacts

Alkaline moiety able to interact with an acid residue

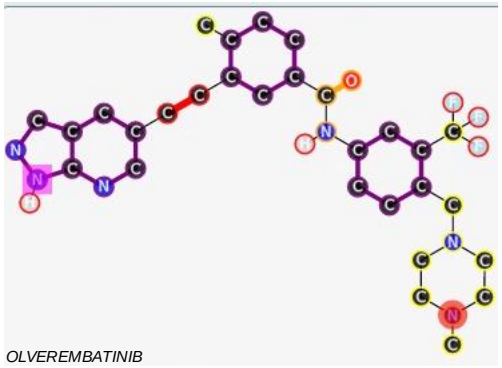
Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

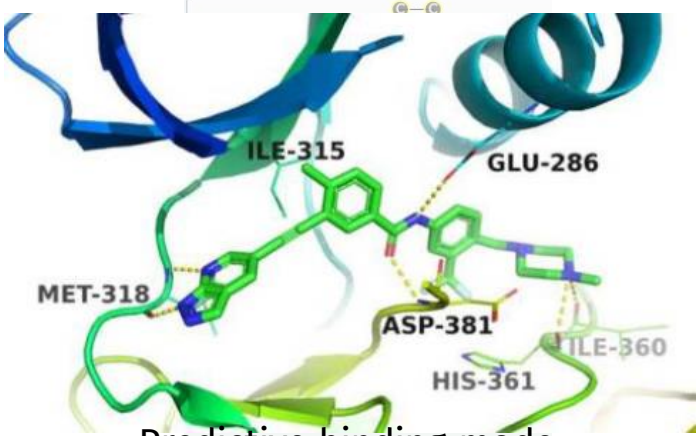
Matches									
106	132	155	244	254	266	267	327	331	345
355	357	358	381	406	423	482	512	567	583
591	702								



(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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


OLVEREMBATINIB



Predictive binding mode

https://pubs.acs.org/doi/pdf/10.1021/jm301581y

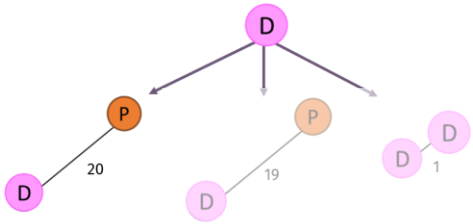
Pyrrazolopyridine able to interact with hydrophilic residues

Constant central benzene for hydrophobic contacts

Alkaline moiety able to interact with an acid residue

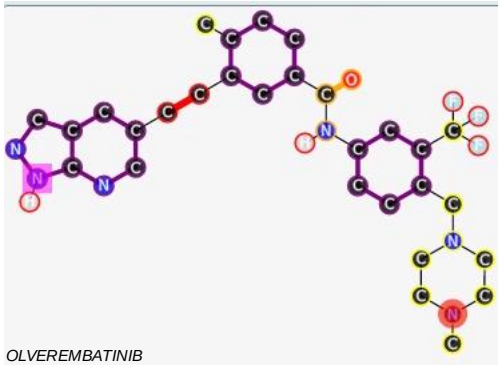
Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

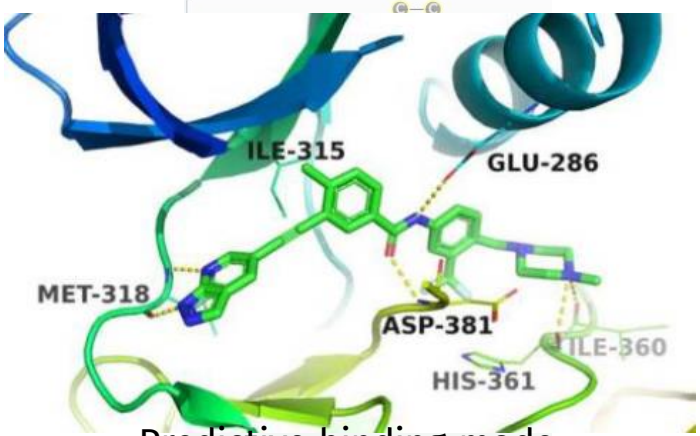


(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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



OLVEREMBATINIB



Predictive binding mode

https://pubs.acs.org/doi/pdf/10.1021/jm301581y

Pyrrazolopyridine able to i hilic r

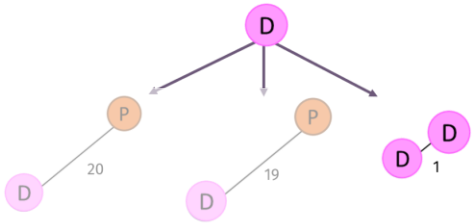
PKI type II

nzene for ts

Alkaline moiety able to interact with an acid residue

Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

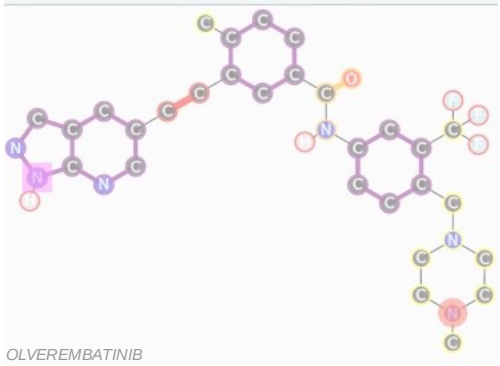
106	132	155	244	254	266	267	327	331	345
355	357	358	381	406	423	482	512	567	583
591	702								

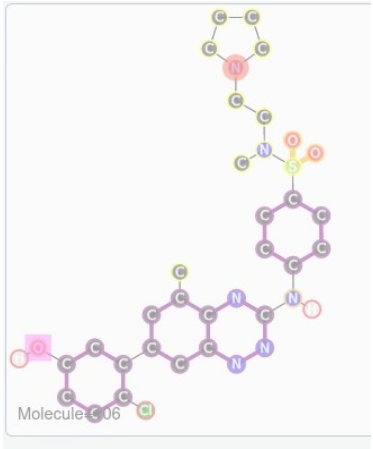


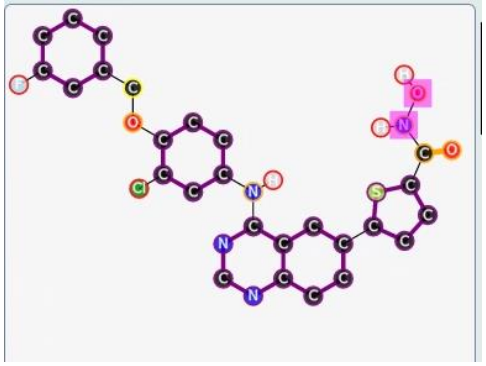
(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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

(Ph4:3) Measures	
Support Size	22
Confidence	0
Growth Rate	0


OLVEREMBATINIB

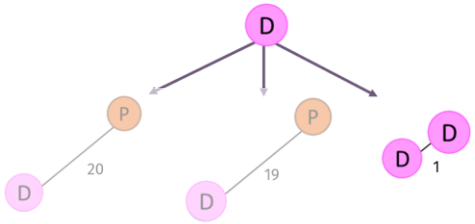

Molecule 406



Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

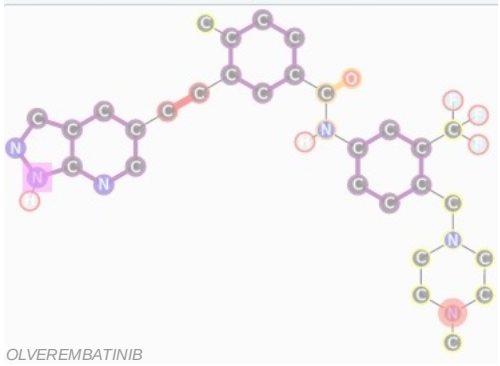
Matches									
885	886	937	1026	1028	1029	1032	1035	1062	1111
1119	1131	1159	1169	1170	1262	1283	1325	1353	1408
1431	1448								

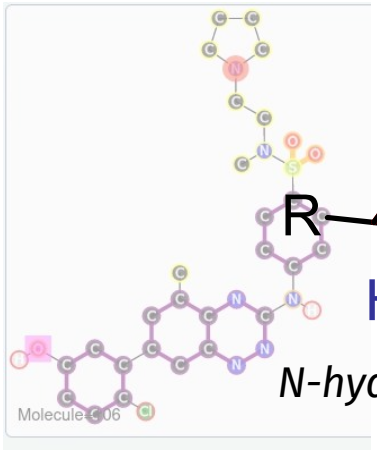


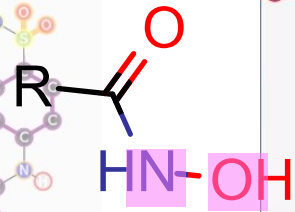
(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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

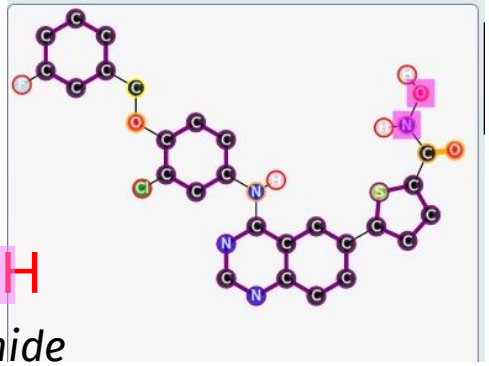
(Ph4:3) Measures	
Support Size	22
Confidence	0
Growth Rate	0


OLVEREMBATINIB





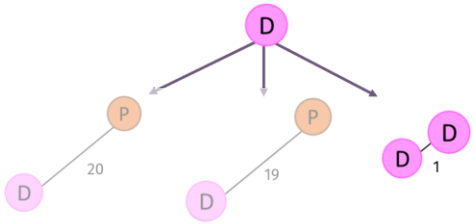
N-hydroxyamide



Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

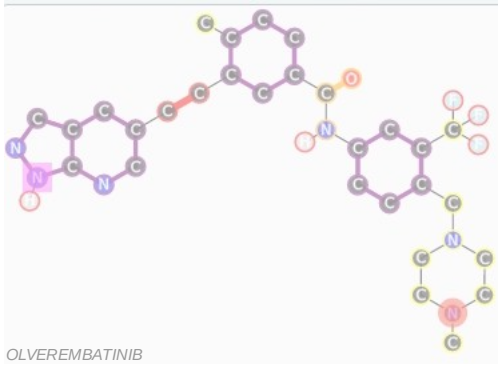
Matches									
885	886	937	1026	1028	1029	1032	1035	1062	1111
1119	1131	1159	1169	1170	1262	1283	1325	1353	1408
1431	1448								

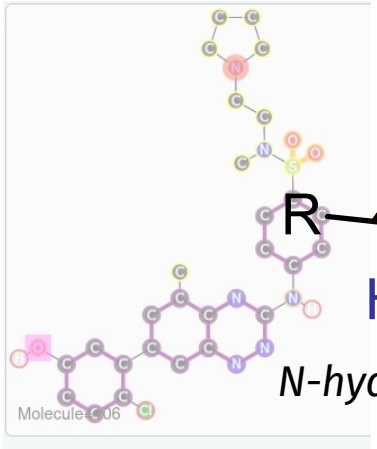


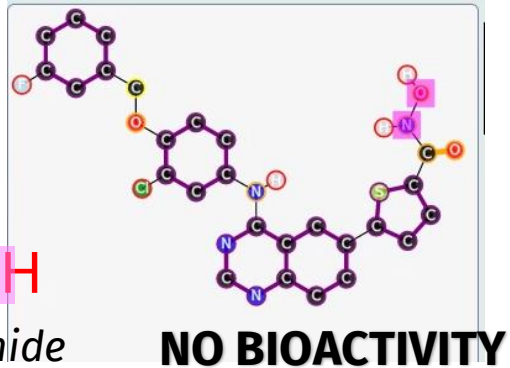
(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

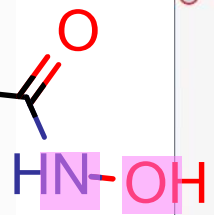
(Ph4:0) Measures	
Support Size	22
Confidence	1
Growth Rate	Infinity

(Ph4:3) Measures	
Support Size	22
Confidence	0
Growth Rate	0


OLVEREMBATINIB







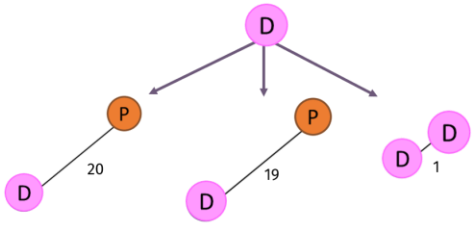
N-hydroxyamide

NO BIOACTIVITY TOWARDS BCR-ABL

Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

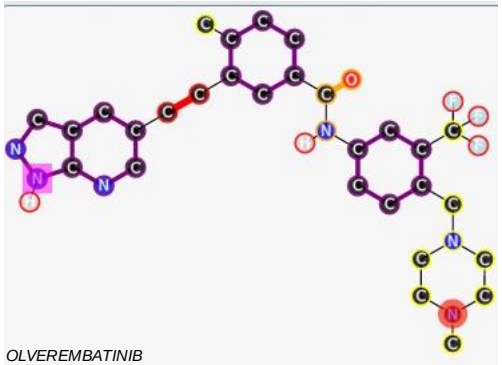
Matches									
885	886	937	1026	1028	1029	1032	1035	1062	1111
1119	1131	1159	1169	1170	1262	1283	1325	1353	1408
1431	1448								



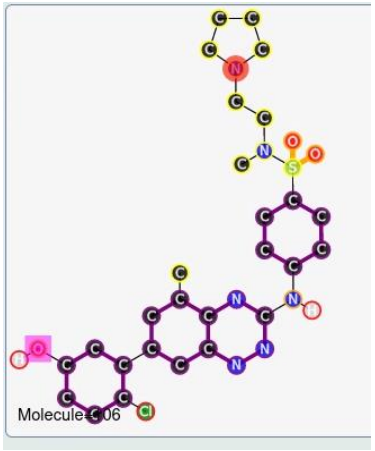
(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

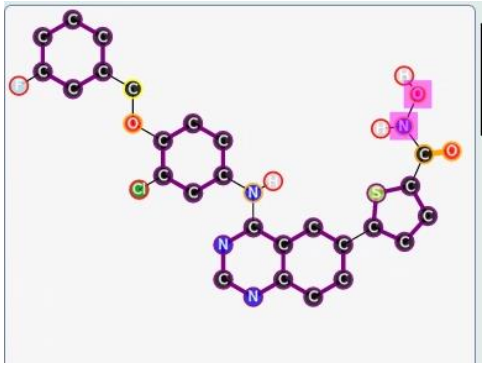
(Ph4:0) Measures	
Support Size	22
Confidence	1
Growth Rate	Infinity

(Ph4:3) Measures	
Support Size	22
Confidence	0
Growth Rate	0



OLVEREMBATINIB

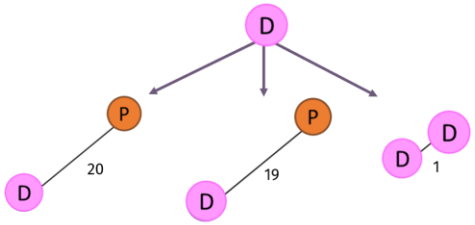




Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

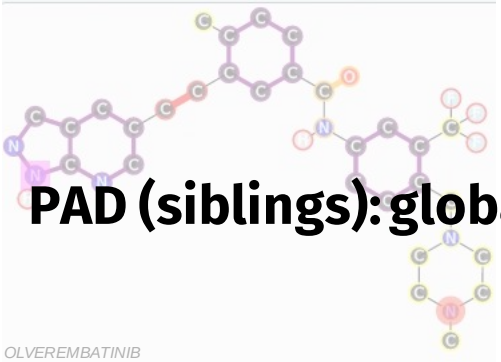
Matches									
885	886	937	1026	1028	1029	1032	1035	1062	1111
1119	1131	1159	1169	1170	1262	1283	1325	1353	1408
1431	1448								

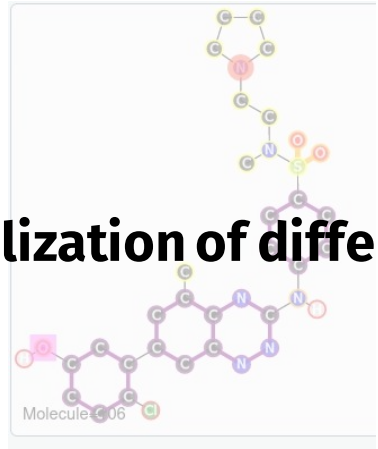


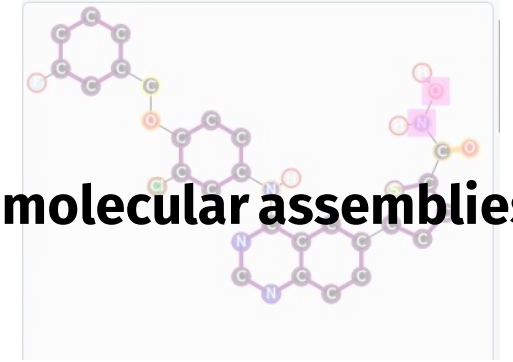
(Ph4:1) Measures	
Support Size	32
Confidence	1
Growth Rate	Infinity

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

(Ph4:3) Measures	
Support Size	22
Confidence	0
Growth Rate	0







PAD (siblings): global visualization of differents molecular assemblies

Matches									
54	61	84	85	111	133	138	145	159	166
168	179	207	308	310	328	360	362	380	384
395	423	487	520	522	523	529	573	590	653
670	688	688							

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

Matches									
885	886	937	1026	1028	1029	1032	1035	1062	1111
1119	1131	1159	1169	1170	1262	1283	1325	1353	1408
1431	1448								

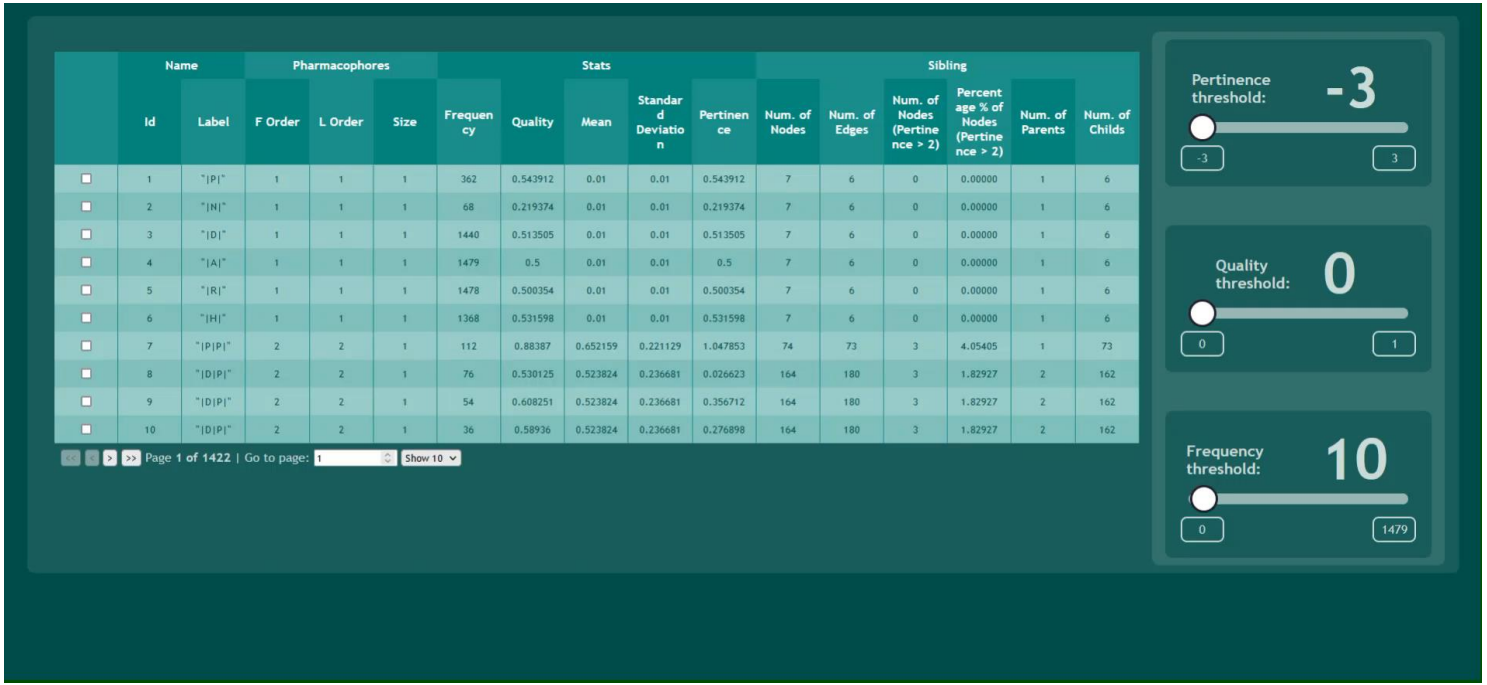
LABORATOIRE BORDELAIS DE RECHERCHE EN INFORMATIQUE
LaBRI

Directed **A**cyclic **G**raph (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



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

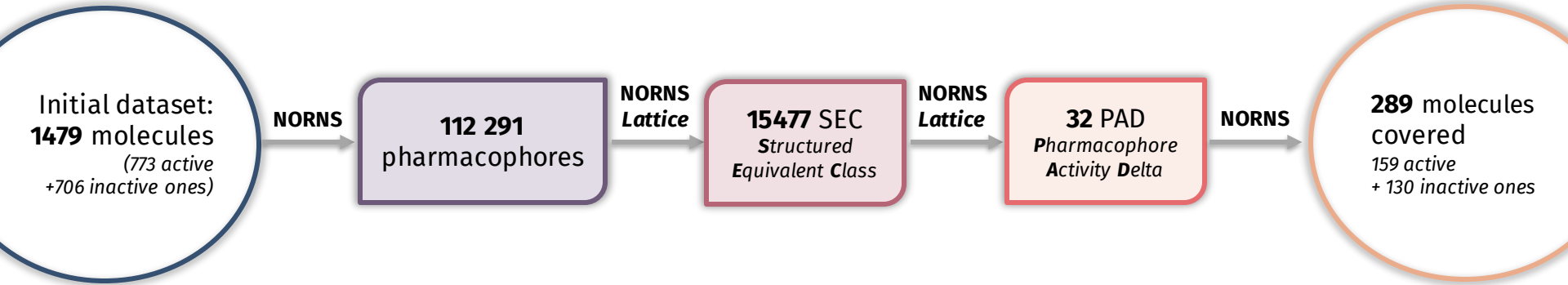
Summary

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

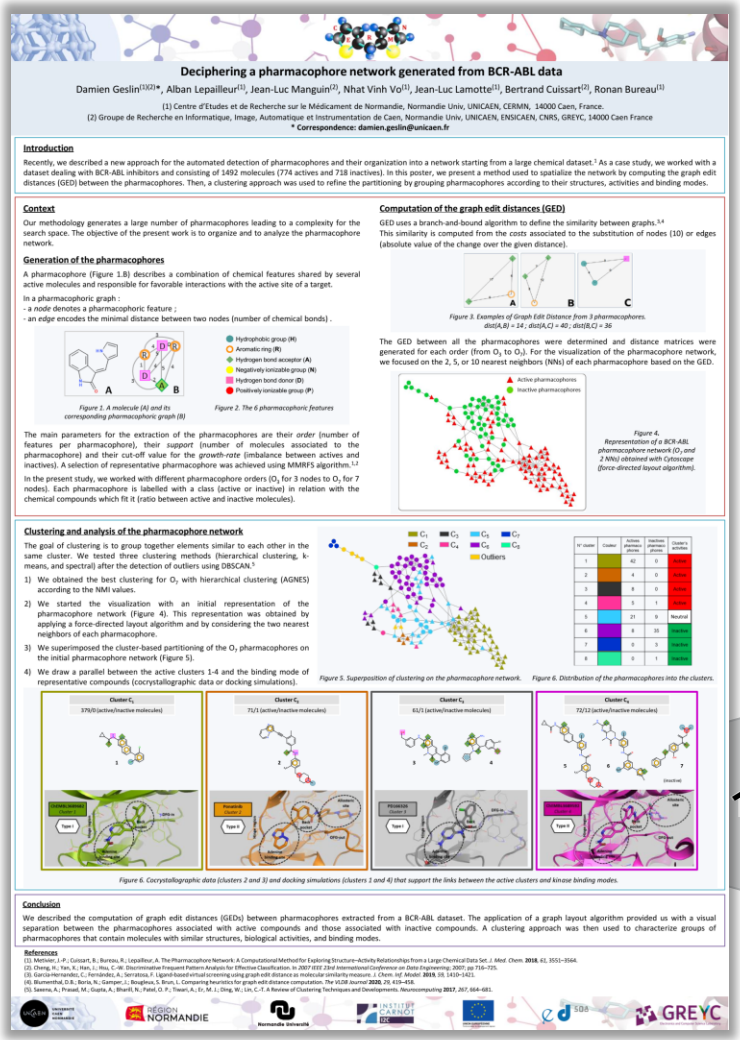


112 291 pharmacophores

What else ?

112 291 pharmacophores

What else ?



Deciphering a pharmacophore network generated from BCR-ABL data

Damien Geslin[1][2]*, Alban Lepailleur[1], Jean-Luc Manguin[2], Nhat Vinh Vo[1], Jean-Luc Lamotte[1], Bertrand Cuissart[2], Ronan Bureau[1]

(1) Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 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

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 inactives). 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) .

Figure 1. A molecule (A) and its corresponding pharmacophoric 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 cut-off value for the *growth-rate* (imbalance between actives and inactives). A selection of representative pharmacophore was achieved using MMRFS algorithm.[1,2]

In the present study, we worked with different pharmacophore orders (O_3 for 3 nodes to O_7 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).

Computation of the graph edit distances (GED)

GED uses a branch-and-bound algorithm to define the similarity between graphs.[3,4]
This similarity is computed from the *costs* associated to the substitution of nodes (10) or edges (absolute value of the change over the given distance).

Figure 3. Examples of Graph Edit Distance from 3 pharmacophores.
dist(A,B) = 14 ; dist(A,C) = 40 ; dist(B,C) = 36

The GED between all the pharmacophores were determined and distance matrices were generated for each order (from O_3 to O_7). For the visualization of the pharmacophore network, we focused on the 2, 5, or 10 nearest neighbors (NNs) of each pharmacophore based on the GED.

Figure 4. Representation of a BCR-ABL pharmacophore network (O_3 and 2 NNs) obtained with Cytoscape (force-directed layout algorithm).

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 DBSCAN.[5]

1) We obtained the best clustering for O_3 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 O_3 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 (cocrystallographic data or docking simulations).

Figure 5. Superposition of clustering on the pharmacophore network.

N° cluster	Cluster	Actives pharmaco. clusters	Inactives pharmaco. clusters	Cluster's activities
1		42	0	Active
2		4	0	Active
3		8	0	Active
4		5	1	Active
5		21	9	Neutral
6		8	16	Inactive
7		0	3	Inactive
8		0	1	Inactive

Figure 6. Distribution of the pharmacophores into the clusters.

Figure 6. Cocrystallographic 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

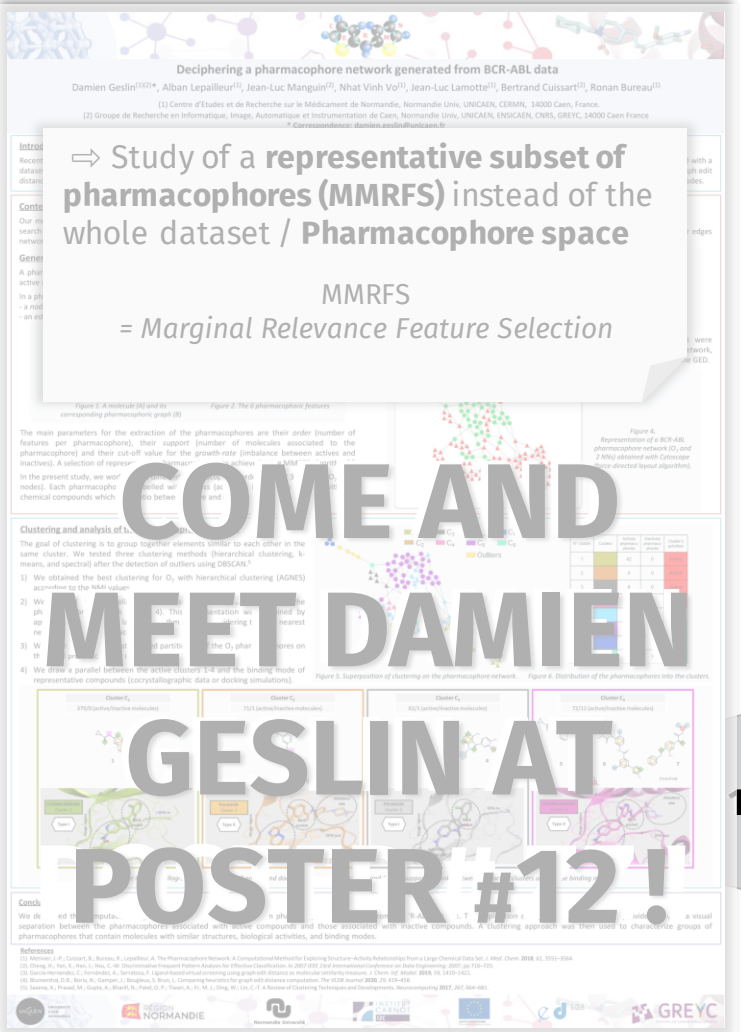
We described the computation of graph edit distances (GEDs) between pharmacophores extracted from a BCR-ABL dataset. The application of a graph layout algorithm provided us with a visual separation between the pharmacophores associated with active compounds and those associated with inactive compounds.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

References
(1). Mehlser, J.-P.; Cuissart, 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*, 3551–3564.
(2). Cheng, H.; Yan, X.; Han, J.; Hsu, C.-W. Discriminative Frequent Pattern Analysis for Effective Classification. In *2007 IEEE 23rd International Conference on Data Engineering*; 2007; pp 716–725.
(3). García-Hernández, C.; Fernández, A.; Serratosa, F. Ligand-based virtual screening using graph edit distance as molecular similarity measure. *J. Chem. Inf. Model.* **2019**, *59*, 1410–1421.
(4). Blumenthal, D.B.; Boria, N.; Gamper, J.; Bougleux, S. Brun, L. Comparing heuristics for graph edit distance computation. *The VLDB Journal* **2020**, *29*, 419–458.
(5). Saxena, A.; Prasad, M.; Gupta, A.; Bharill, N.; Patel, O. P.; Tiwari, A.; Er, M. J.; Ding, W.; Lin, C.-T. A Review of Clustering Techniques and Developments. *Neurocomputing* **2017**, *267*, 664–681.

112 291 pharmacophores

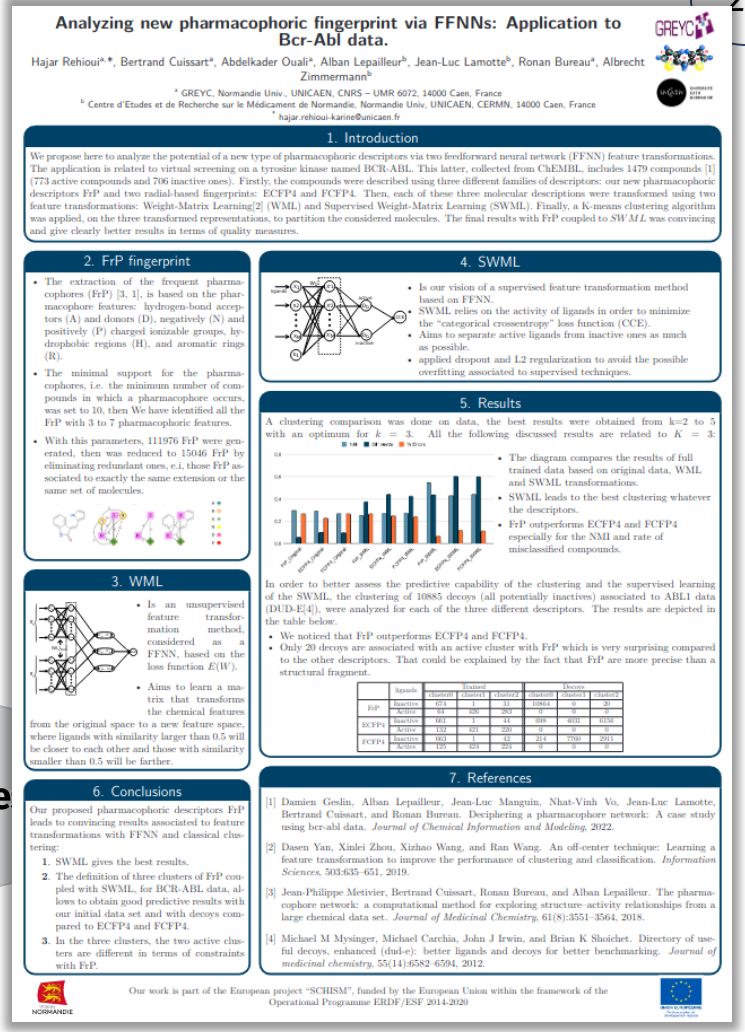
What else ?

112 291 pharmacophores



What else ?

112 291 pharmacophore



What else ?

112 291 pharmacophore

Towards DAG-based interactive pharmacophore exploration

Johanna GIOVANNINI[1#], Etienne LEHEMBRE[2#], Jean-Luc LAMOTTE[2], Abdelkader OUALI[2], Alban LEPAILLEUR[1], Albrecht ZIMMERMANN[3], Bertrand CUISSART[3], Ronan BUREAU[1] and Bruno CREMILLEUX[3]

(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 (c), 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.

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

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[4] (MW ≤ 800 g/mol): 1479 molecule with either Ki or IC50 information towards ABL1 target affinity. 773 were considered as active compounds (Ki or IC50 ≤ 100 nM).

ChEMBL 1479 compounds 773 active towards BCR-ABL

Using our in-house software Norns (*2.0 version*), we generated 112,291 pharmacophores from order 1 to order 7 with a minimal support of 10 compounds.[5,6]

NORNS 112 291 pharmacophores

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).

NORNS LATTICE $GR_{Normalized}$ 32 PAD $GR_{Classic}$ 394 PAD

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.

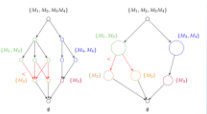


Fig 2. *From a relational directed acyclic graph (DAG) to a clusterize relationnal diagram.*

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**).

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**).

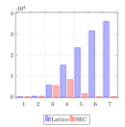


Fig 3. *Number of pharmacophores per layers (in blue) and number of EC per layer (in red).*

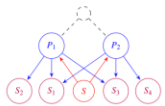


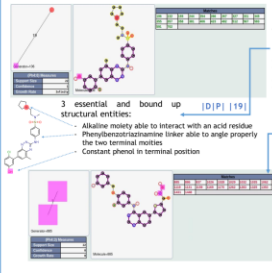
Fig 4. *From parents Pi (in blue) of a starting EC S (in red), we extract the EC siblings Si (in purple) in order to study S regarding the Si.*

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 allow us to study pharmacophores which deviate from their siblings and might explain biological activity.

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:



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.

References:
[1] T. Langer, *et al.*, Pharmacophores and Pharmacophore Searches, John Wiley & Sons, 2006. [2] D. Horvath, Chemoinformatics Approaches Virtual Screen., 2008. [3] G. Schneider, W. Neidhart *et al.*, Angew. Chem. Int. Ed. 1999, 38, 2894-2896. [4] ebi.ac.uk/chembl/target_report_card/CHEMBL1862/ [5] Métivier *et al.*, J. Med. Chem. 2018, 61. [6] Goslin *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

InvolvD ANR-20-CE23-0023