CHEMOINFORMATICS:

Pharmacophore graph

Nhat-Vinh Vo‡ Bertrand Cuissart‡ Ronan Bureau†

‡ GREYC – FR CNRS UMR 6072
† CERMN – UPRES EA 4258 – FR CNRS 3038 INC3M
firstname.lastname@unicaen.fr

April 20, 2018

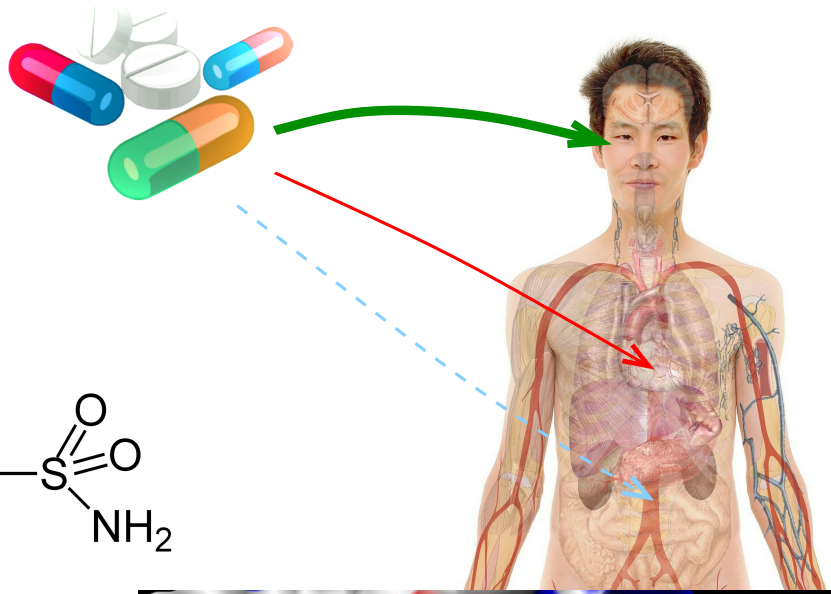
AGAC
Analyse de Graphe Appliquée à la Chémoinformatique
Financial support from the Normandy region

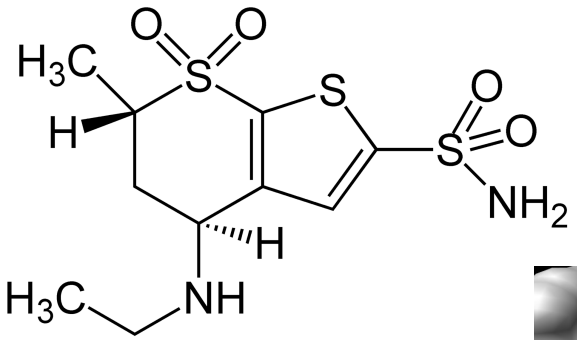
Plan

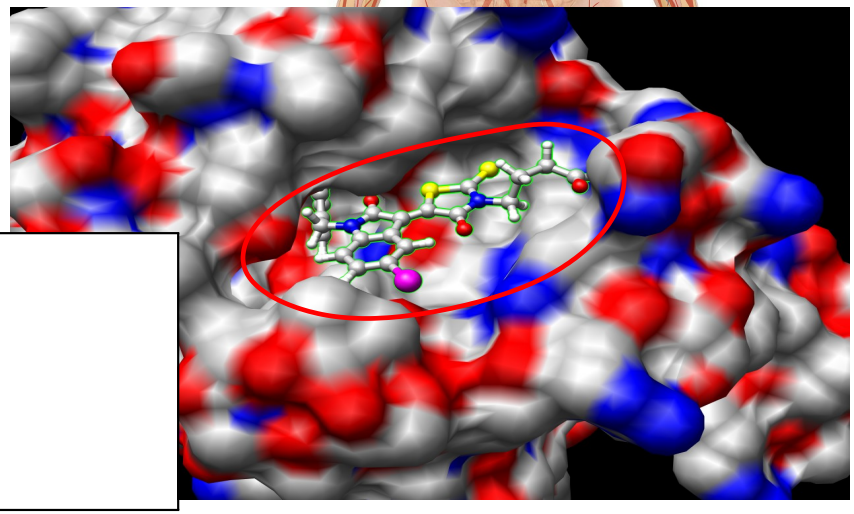
- Data environment for chemoinformatics
  - Protein-Ligand interactions
  - ChEMBL
- Pharmacophore graph
  - Notations
  - Norns
- Graph edit distance
  - Similarity between two graphs
  - Mining the graph edit path

Data environment for chemoinformatics

- Protein-Ligand interactions







- Objective :
  - Maximize the therapeutic effect,
  - Limit all side effects,
  - Avoid any adverse effect.

Data environment for chemoinformatics

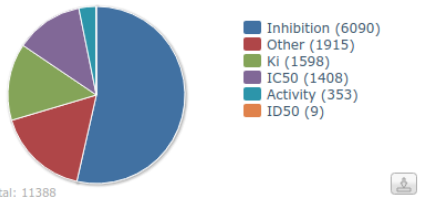
- "ChEMBL or ChEMBLdb is a manually curated chemical database of bioactive molecules with drug-like properties" (Wikipedia) *(https://www.ebi.ac.uk/chembl/)*

Consequence : a lot of experimental measurements available

Example: ABL on 12 Apr 2018

Target Associated Bioactivities

ChEMBL Activity Types for Target CHEMBL1862



ChEMBL Statistics

- DB: ChEMBL_23
- Targets: 11,538
- Compound records: 2,101,843
- Distinct compounds: 1,735,442
- Activities: 14,675,320
- Publications: 67,722
- Release Notes

Target Name and Classification

Target ID	CHEMBL1862
Target Type	SINGLE PROTEIN
Preferred Name	Tyrosine-protein kinase ABL
Synonyms	ABL | ABL1 | Abelson murine leukemia viral oncogene homolog 1 | Abelson tyrosine-protein kinase 1 | JTK7 | Proto-oncogene c-Abl | Tyrosine-protein kinase ABL1 | p150
Organism	Homo sapiens
Species Group	No
Protein Target Classification	• enzyme > kinase > protein kinase > tk protein kinase group > tyrosine protein kinase abl family

Target Components

Component Description	Relationship	Accession
Tyrosine-protein kinase ABL1	SINGLE PROTEIN	P00519

Target Relations

ChEMBL ID	Pref Name	Target Type
CHEMBL2096618	Bcr/Abl fusion protein	CHIMERIC PROTEIN
CHEMBL2111414	Tyrosine-protein kinase ABL	PROTEIN FAMILY

Approved Drugs and Clinical Candidates

ChEMBL ID	Name	Mechanism of Action	Max Phase	References
CHEMBL288441	BOSUTINIB	Tyrosine-protein kinase ABL inhibitor	4	Expert
CHEMBL1421	DASATINIB	Tyrosine-protein kinase ABL inhibitor	4	DailyMed

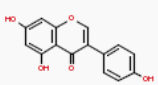
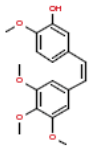
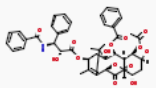
Data environment for chemoinformatics

ChEMBL Bioactivity Search Results: 1409

Please select....

10 records per page

Show / hide columns

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	pChEMBL Value	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
 CHEMBL44	270.24	IC50	=	10000	nM	5	B	In vitro inhibition of the v-abl tyrosine kinase activity in A431 membranes using angiotensin II as phosphate acceptor as substrate	Scientific Literature		SINGLE PROTEIN	Tyrosine-protein kinase ABL	Homo sapiens	J. Med. Chem., (1995) 38:13:2441
 CHEMBL67	316.35	IC50	>	10000	nM		B	Inhibition of Abl1 (unknown origin) assessed as incorporation of [32P] gamma-ATP in to myelin basic protein after 30 mins by autoradiographic analysis	Scientific Literature	Homo sapiens	SINGLE PROTEIN	Tyrosine-protein kinase ABL	Homo sapiens	J. Med. Chem., (2014) 57:11:4598
Chiral  CHEMBL428647	853.91	IC50	>	10000	nM		B	Inhibition of Abl1 (unknown origin) assessed as incorporation of [32P] gamma-ATP in to myelin basic protein after 30 mins by autoradiographic analysis	Scientific Literature	Homo sapiens	SINGLE PROTEIN	Tyrosine-protein kinase ABL	Homo sapiens	J. Med. Chem., (2014) 57:11:4598

- Data environment for chemoinformatics
- **Reduced pharmacophore graph**
  - **Notations**
  - **Norns**
- Graph edit distance

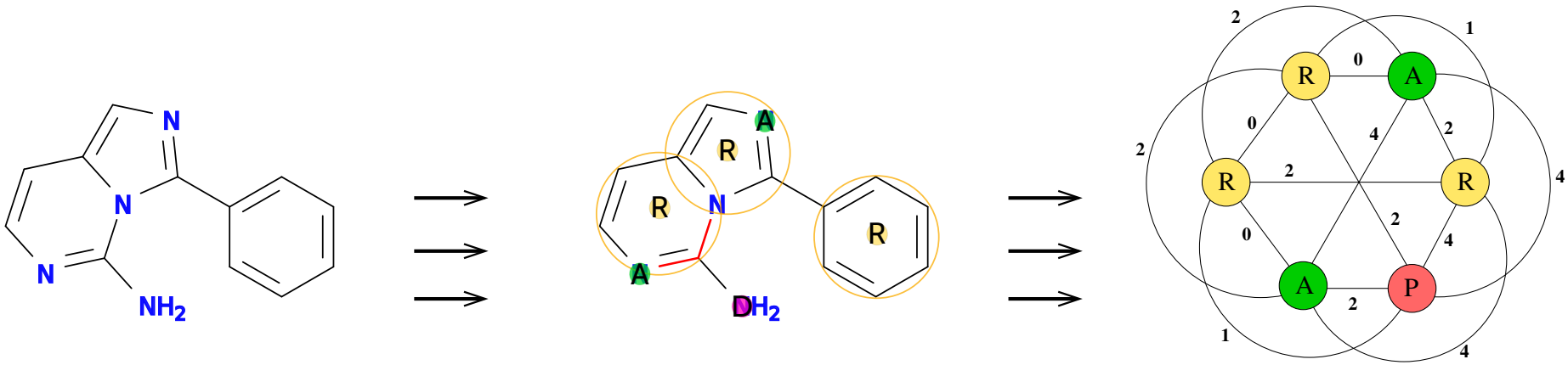
Pharmacophore graph

An abstracted representation of a molecule

Pharmacophore Graph includes the following features

- Hydrogen Bond **A**cceptor, Hydrogen Bond **D**onor, aromatic **R**ing, **H**ydrophobic area, **P**ositively ionizable group, **N**egatively ionizable group[a]
- an adjustable part of the input.

[a] OpenBabel, N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, J. Cheminformatics 2011, 3, 33.



Skeletal formula | Feature occurrences | Pharmacophore graph

The Pharmacophore graph (of a molecule)

- A vertex: an occurrence of a feature
- An edge: the distance between two features

Pharmacophore graph

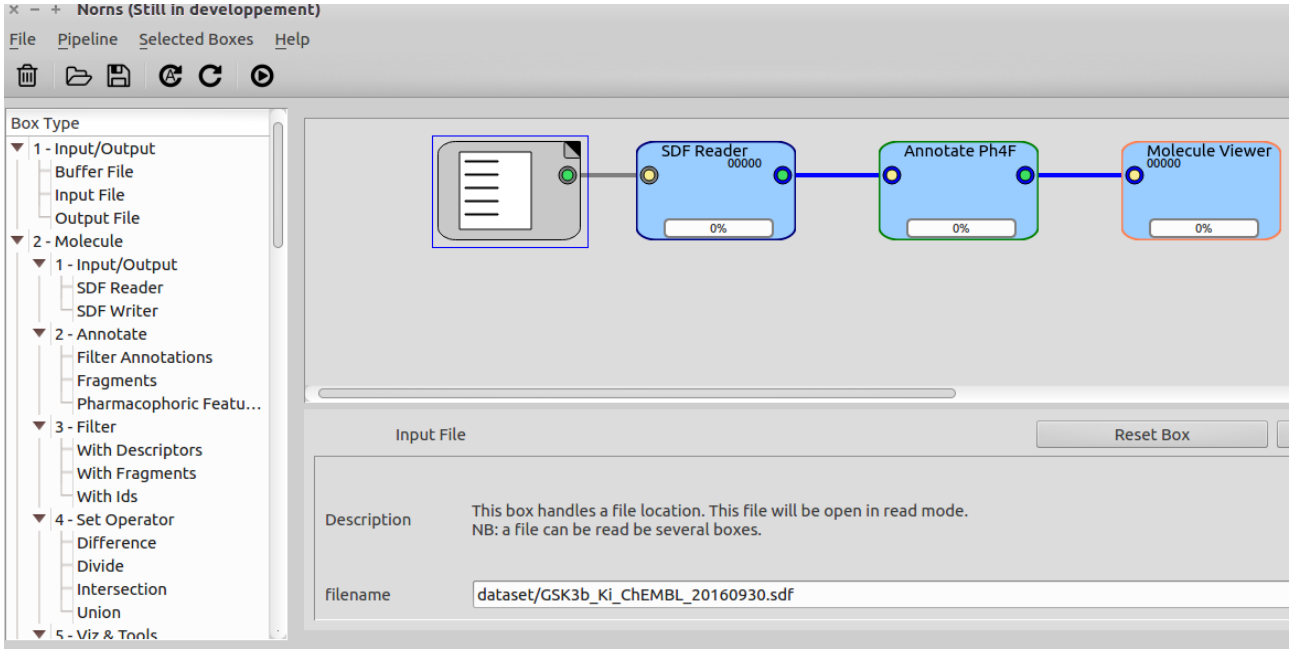
Norns

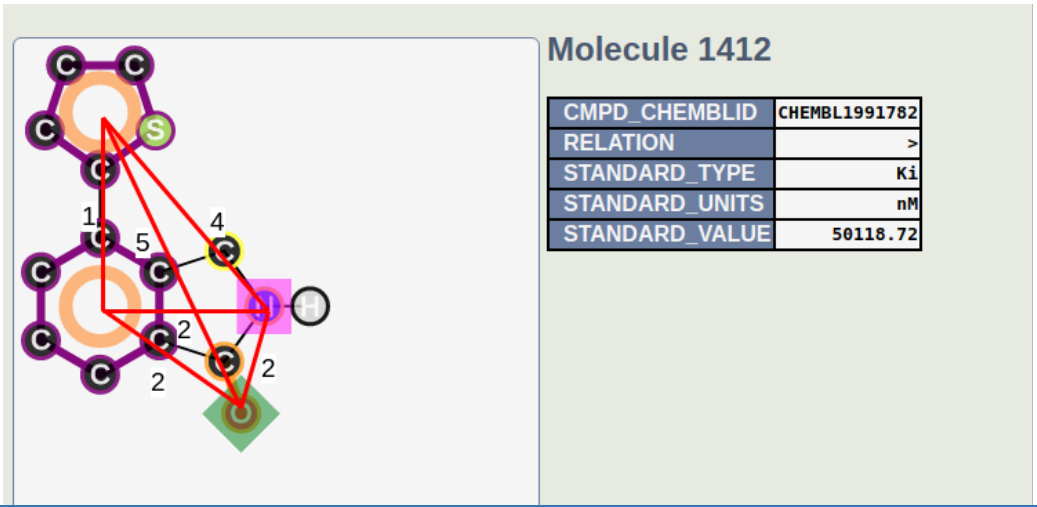
- a GUI to express workflows in chemoinformatics,
- 60 configurable boxes that include:
  - **the reading/writing of molecules from/into files**,
  - **molecular representations**,
  - **pharcophoric features mining**,
  - **pharmacophore manipulations**,
  - and **classifiers**
- implementation of complex processes by connecting boxes,
- result analysis through dynamic interactive web pages.
- Available at : *https://chemoinfo.greyc.fr/2017_metivier/MiniNorns.tgz*

The Pharmacophore Network: a Computational Method for Exploring Structure-Acti
Jean-Philippe Métivier, Bertrand Cuissart, Ronan Bureau, Alban Lepailleur
(J Med Chem. 2018 Apr 12. doi: 10.1021/acs.jmedchem.7b01890)

Pharmacophore graph

Norns





- Data environment for chemoinformatics
- Pharmacophore graph
- **Graph edit distance**
  - ***Similarity between two graphs***
  - ***Mining the graph edit path***

Graph edit distance

Graph edit distance

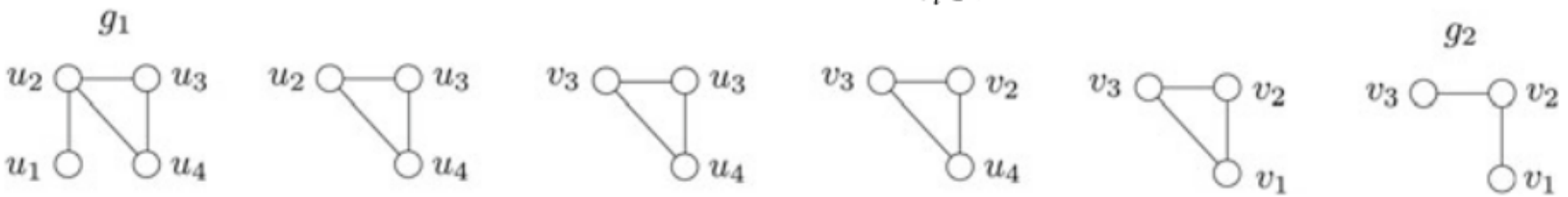
- measures distances between two graphs g_1 and g_2 by the amount of distortion that is needed to transform g_1 into g_2.[a]

[a] Graph Edit Distance, K. Riesen, Structural Pattern Recognition with Graph Edit Distance 2015, p29-45.

Definition 2.1 (*Edit Path*) A set $\{e_1, \ldots, e_k\}$ of k edit operations e_i that transform g_1 completely into g_2 is called a *(complete) edit path* $\lambda(g_1, g_2)$ between g_1 and g_2. A *partial edit path*, i.e., a subset of $\{e_1, \ldots, e_k\}$, edits proper subsets of nodes and/or edges of the underlying graphs.

Definition 2.2 (*Graph Edit Distance*) Let $g_1 = (V_1, E_1, \mu_1, \nu_1)$ be the source and $g_2 = (V_2, E_2, \mu_2, \nu_2)$ the target graph. The *graph edit distance* $d_{\lambda_{\min}}(g_1, g_2)$, or $d_{\lambda_{\min}}$ for short, between g_1 and g_2 is defined by

$$d_{\lambda_{\min}}(g_1, g_2) = \min_{\lambda \in \Upsilon(g_1, g_2)} \sum_{e_i \in \lambda} c(e_i), \tag{2.1}$$



Graph edit distance

Mining the graph edit path

- each graph edit path corresponding to the edition's performance
- group of graph edit paths leading to the same tendency of edition
  - ⟶ Is there any operation leading to the significant performance?
  - ⟶ Is there any series of operations leading to the significant performance?

Mining the graph edit path in chemoinformatics

- ⟶ Is there any operation on a ligand leading to the better affinity?
- ⟶ Is there any series of operations on a ligand leading to the better affinity?

Summary

- Data environment for chemoinformatics
  - Protein-Ligand interactions
  - ChEMBL
- Pharmacophore graph
  - Notations
  - Norns
- Graph edit distance
  - Similarity between two graphs
  - Mining the graph edit path