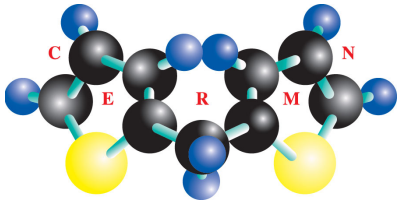

COMPUTING SIMILARITY BETWEEN PHARMACOPHORES

USING GRAPH EDIT DISTANCES

Nhat-Vinh Vo[‡] Bertrand Cuissart[‡] Ronan Bureau[†] Alban Lepailleur[†]

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

May 3, 2019

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

Plan

- Context
- Objective and Principle
- Implementation
  - Preliminary implementation
  - Deviation
- Conclusion and Perspective.

Context

Terminologies:

- Ligand: an organic molecule that binds with a biomolecule
- Target: a part of the organism to which a ligand binds, producing a change of behaviour. Proteins is a common class of targets.
- Pharmacophore: a part of a molecular structure that is responsible for a particular biological or pharmacological interaction that it undergoes.

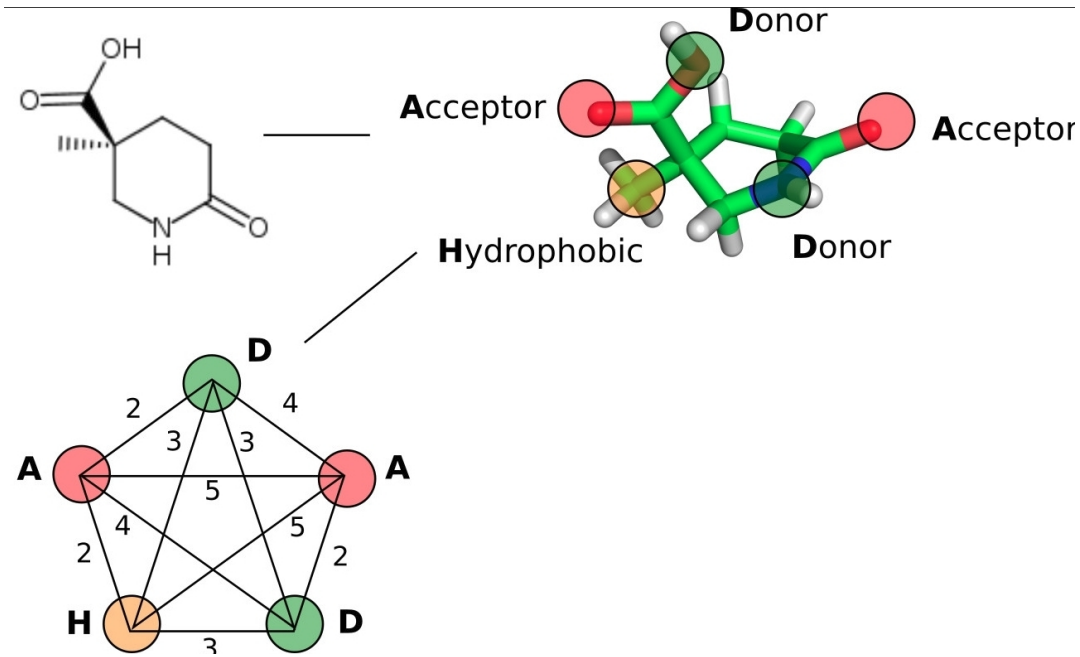

Figure : A ligand-pharmacophore model [Kutlushina et al., 2018]

Context

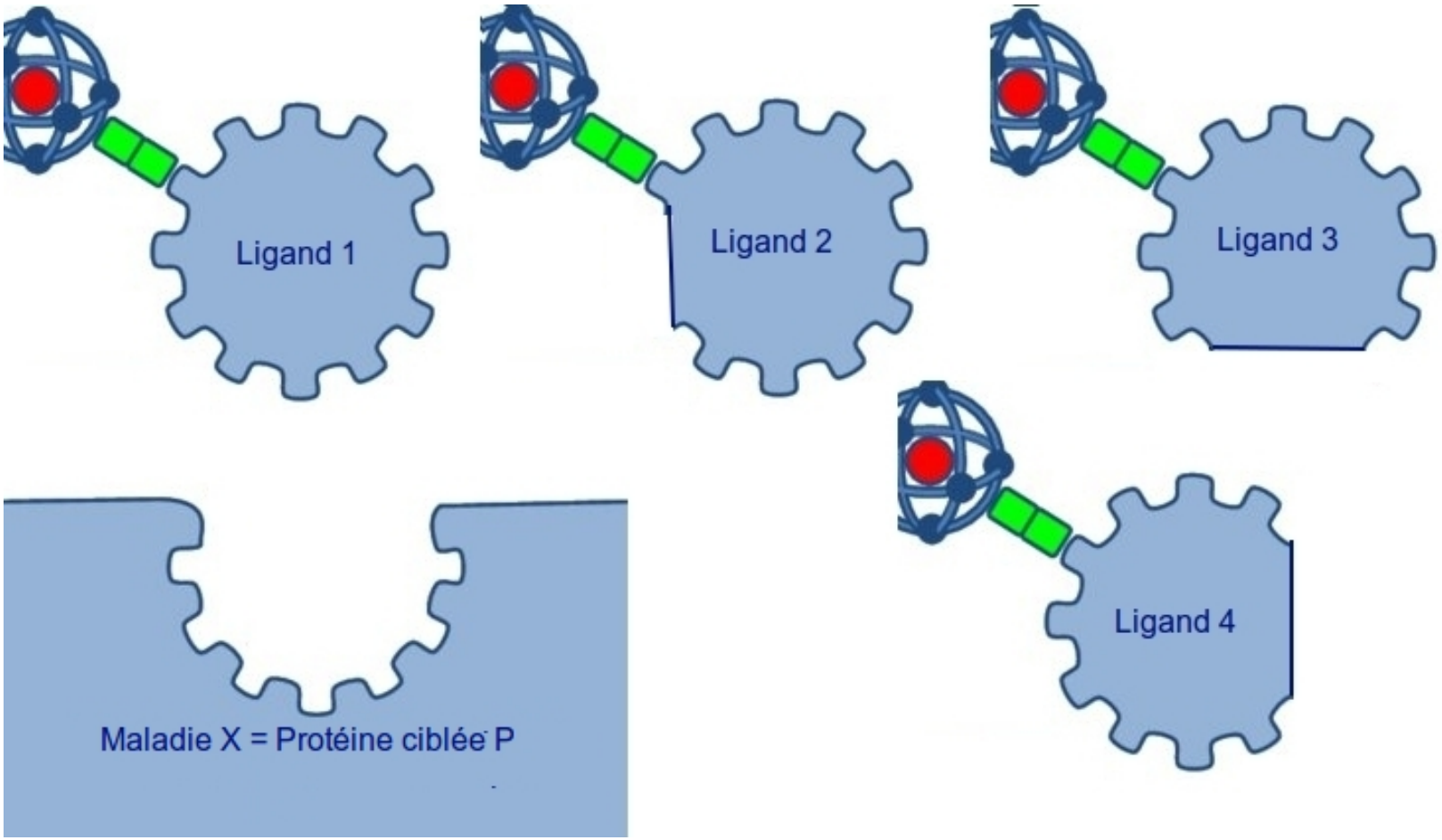

Figure : Ligand-Target Model.

Objective

Given a set of studied ligands and their biological profiles/bioactivity (to a target)

$\longrightarrow$ Predict the potential bioactivity of ligand of which the biological profile is currently unknown.

Principle

- Compute similarity between studied ligands
- Elaborate a mechanism for a supervised classification
- Apply the supervised classification to the candidate ligand of which the bioactivity has not been known yet.

Implementation

- Preliminary implementation
- Deviation.

Preliminary implementation

- Compute similarity between ligands using Graph Edit Distance (GED)
- Compute Graph Edit Distance using a Branch-and-Bound procedure *[Abu-Aisheh et al., 2017]*
- Apply GED to 1492 ligands (from 8 to 13 vertex) from a ligand set for ABL-target

  $\longrightarrow$ It exists instances without any optimality within 10 computation minutes

  $\longrightarrow$ Size of graph should be reduced by using sub-parts of ligands which are responsible for pharmacological interactions.

Preliminary implementation

Example: H-H-H-4-3-3 vs. A-H-H-H-4-3-3-4-7-8

$\longrightarrow$ Assumption: good cost configuration

$\longrightarrow$ 3 Node substitutions and 1 Node insertion: $3! \times 4 = 24$ possibilities

- 1 Node substitution: H $\rightarrow$ A
- 1 Node insertion: H
- 3 Edge insertions: A-H (3); H-H (7); H-H (8)
- 3 Edge substitutions: H-H (4) $\rightarrow$ A-H (4); H-H (3) $\rightarrow$ H-H (4); H-H (3) $\rightarrow$ A-H (3).

Deviation

- Apply **Norns** (by projet Minomics) to the dataset to extract a space of pharmacophores which satisfy the following constraints *[Métivier et al., 2018]*:
  - Size of graph: from 1 to 6 vertex
  - Support/Frequency: at least 10
  - Growth Rate: used to classify pharmacophores
  - With and without MMRFS algorithm

  $\longrightarrow$ Without MMRFS (S_1): 87175 pharmacophores

  $\longrightarrow$ With MMRFS (S_2): 137 pharmacophores
- Edit Cost Configuration:
  - Node Deletion/Insertion: 6
  - Node Substitution: 10
  - Edge Deletion/Insertion: 0
  - Edge Substitution: *difference*(edge1,edge2)

  $\longrightarrow$ Computation time for GED: 0.5 second

Representation processing

- From viewpoint of pharmacophores in S_1:
  - Compute GED between 2 pharmacophores in S_2 and all pharmacophores in S_1
  - Represent the nearest pharmacophores in S_1 to the 2 pharmacophores in S_2

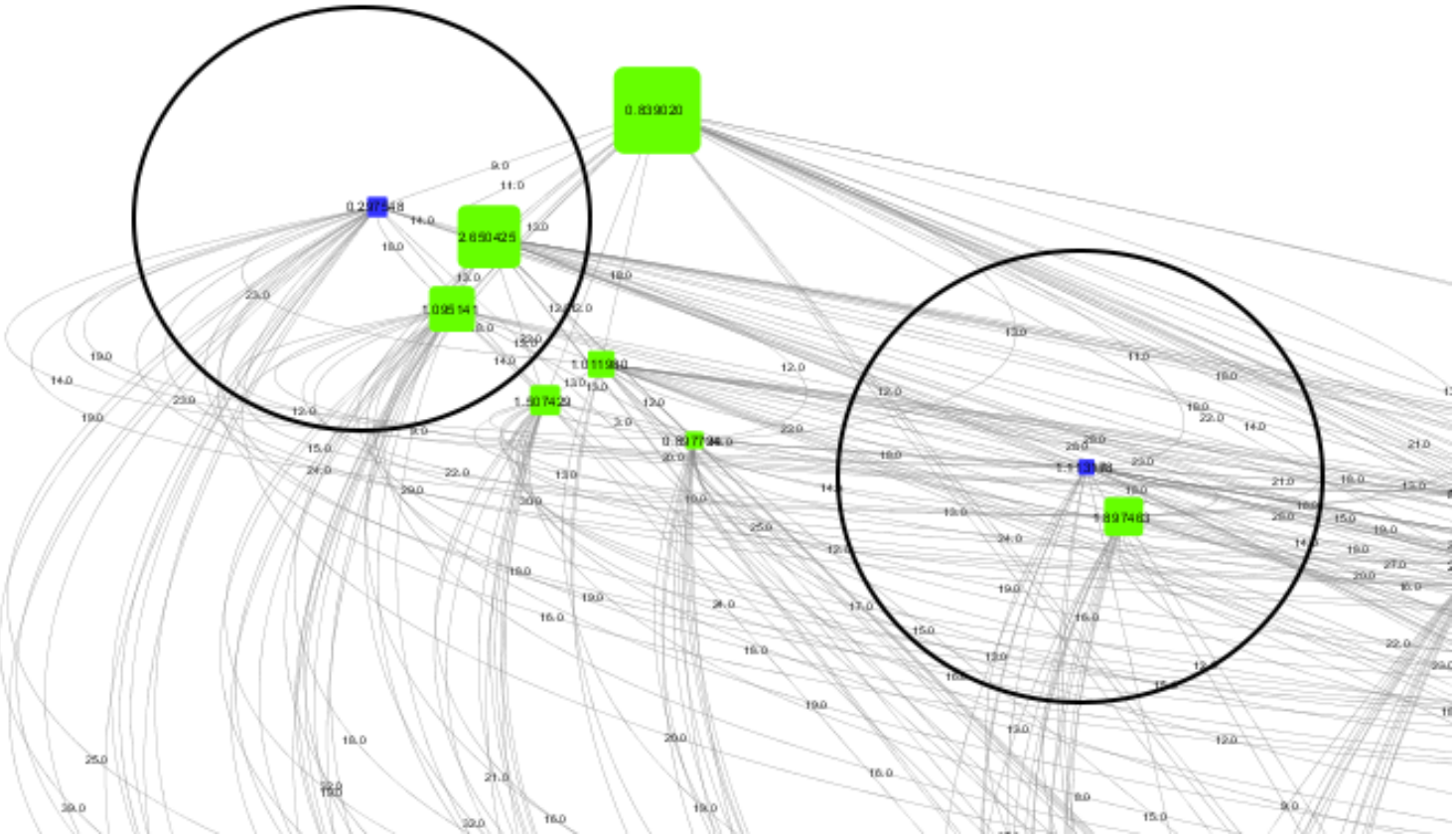

Figure : An example of representation processing usind GED.

Representation processing

- From viewpoint of each candidate ligand with currently unknown biological profile:
  - Compute GED between all pharmacophores in S_2 belonging to the ligand and all pharmacophores in S_1
  - Predict the bioactivity of the candidate ligand using profiles of the nearest pharmacophores.

Conclusion and Perspective

- Use GED as similarity between ligands and between pharmacophores
- Apply a Branch-and-Bound procedure to GED computation

  $\longrightarrow$ Is there any better way for an exact solution?
- Use the nearest neighbours to represent pharmacophores and to predict the bioactivity of the candidate ligand

  $\longrightarrow$ Is there any better way for a prediction with relational values?