

## Title: Towards lattice-based interactive pharmacophore exploration

Etienne Lehembre<sup>a</sup>, Jean-Luc Lamotte<sup>b</sup>, Abdelkader Ouali<sup>a</sup>, Christophe Couronne<sup>a</sup>, Bruno Crémilleux<sup>a</sup>, Alban Lepailleur<sup>b</sup>, Jean-Luc Manguin<sup>a</sup>, Ronan Bureau<sup>b</sup>, Albrecht Zimmermann<sup>a</sup>, Bertrand Cuissart<sup>a</sup>

<sup>a</sup> GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France

<sup>b</sup> Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

In medicinal chemistry, a pharmacophore denominates a spatial arrangement of chemical features shared by a group of molecules which is responsible for a favourable interaction with binding site of the studied target. We have recently designed a method that automatically computes pharmacophores from a large data set of molecules without any prior supervised selection of a small subset of molecules [1]. The connections between the computed pharmacophores provide a hierarchical organization: the pharmacophore network. The latter is structured by pharmacophores' size and contains a large number of them. The current work aims at developing a structure which a drug designer can use to support his analysis without having to repeatedly mine pharmacophores. For this purpose, we enhance the knowledge on the pharmacophore network organization by taking into account parent-children relations and/or grouping the pharmacophores into equivalence classes, i.e. sets of pharmacophores occurring in exactly the same molecules. Additionally, each pharmacophore is annotated with one or several quality measures (e.g. confidence or growth rate measurements), information that will be exploited later-on.

The lattice is built and organized in layers which contain pharmacophores with the same number of pharmacophoric features, as well as their extensions (molecules in which they appear). Layers are linked by establishing parent-children relationships based on a sub-graph relationship. The lattice can then be used to apply different filtering and selection algorithms, e.g. based on user-specified quality measures, and visualizing the effect of this application on each layer. In order not to overwhelm the expert, the visualization is condensed by clustering nodes in each layer using the above-mentioned equivalence classes. We will also use the lattice to identify *Pharmacophore Activity Deltas* (PAD). PADs are syntactic pharmacophores families, i.e. pharmacophores linked via the afore-mentioned parent-child relationships exhibiting a minimum amount of syntactic similarity, yet members of which show a significant difference between the values of their quality measures.

With a view towards Interactive Pattern Mining [2], we also use the lattice to simulate the effects of users or functions interacting with the search space during the search itself. Each non-explored branch in the lattice represents a pharmacophore and its descendants which would not have been explored during the interactive mining, allowing users to understand the repercussions of their interaction. This can be considered a training step for users, who do, after all, make up an indispensable ingredient of interactive pattern mining.

[1] Métivier J-P., Cuissart B. Bureau R. and 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] Van Leeuwen, Matthijs. "Interactive data exploration using pattern mining." *Interactive knowledge discovery and data mining in biomedical informatics*. Springer, Berlin, Heidelberg, 2014. 169-182.