

## Molecular Subgraph Mining for Analyzing Ligand Activity Classes

Julio E. Peironcely<sup>1</sup>, Eelke van der Horst<sup>1</sup>, Adriaan P. IJzerman<sup>1</sup>, Michael Emmerich<sup>2</sup>, Andreas Bender<sup>1</sup>

<sup>1</sup>Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research, Leiden, The Netherlands

<sup>2</sup>Leiden Institute for Advanced Computer Science, University of Leiden, The Netherlands

Protein sequence or overall structural similarities are often employed to categorize the similarity of receptors, but this approach might not be ideal from the chemist's perspective. It can happen that chemically similar ligands interact with proteins without any obvious sequence similarity. Relating receptors by the similarity of their ligands can provide relationships that may be missed if we only study the sequence of the targets. In this study we grouped targets by finding frequent substructures in their ligands employing different graph mining approaches. Knowing these frequencies allowed us to discover substructures that are useful to effectively separate two families, i.e. when they are frequent in one family and infrequent in the others. We used these frequencies to build a phylogenetic tree to visualize the distance at which the target families are related according to the similarities of their ligands. We analyzed activity classes that are similar from the ligand-side, despite having a small sequence similarity, and assume these similarities to be relevant in the context of drug side effect predictions. On the other hand, our study provides tools to detect which fragments increase the specificity of a ligand, reducing promiscuity and off-target interactions.