

Intuitive Structure-Activity Analysis for Medicinal Chemists



Jan Dreher^{*#}, Martin Stahl[#] and Daniel Stoffler[§]

[#] Medicinal Chemistry, Pharma Research & Early Development Informatics, F. Hoffmann-La Roche Ltd

[§] Cheminformatics and Statistics, Pharma Research & Early Development Informatics, F. Hoffmann-La Roche Ltd

* Corresponding author e-mail: jan.dreher@roche.com

1 – Introduction

For a medicinal chemist, the association between chemical structure information and multiple property and assay data is vital. However, identification of structure-activity relationship (SAR) patterns is a complex process, as observed trends are typically not statistically significant. Rather, individual observations lead to hypotheses and rational design is accompanied by guess work and exploration based on synthetic feasibility. Furthermore, most scientists work with form- or table-based views that hide most of the data and give static views that make interactive work difficult. The goal of this project is to provide interaction methods enabling Roche scientists to engage in real-time data exploration, fully utilizing all generated data to identify any relevant trends.

2 – The Tool in Action

The main analysis takes place in the so-called workspace of the tool. Starting from any selected compound, the user might ask for modifications in a particular area. Hits can be dropped in the workspace, where they can be compared to further molecules. In addition to a manual arrangement of all structures, they can be sorted according to predefined properties.

The screenshot illustrates the tool's workspace with several key features highlighted by blue callouts:

- Automatic 2D alignment using a common substructure (MCS or SSS):** Shows a series of molecules aligned to a common substructure.
- Ask for modifications and place queries:** A dialog box for defining search queries based on structural modifications.
- Select compounds for close-up analysis in the workspace:** A list of molecules with a selected one for detailed view.
- Highlight MCS or substructure:** A callout pointing to a common substructure within a molecule.
- Highlight differences:** A callout pointing to differences between two molecules.
- Arrange hits around query:** A callout pointing to a query molecule with its related hits arranged around it.
- Thrombin selectivity:** A callout pointing to a molecule with a note about its selectivity.
- Scatter compounds:** A callout pointing to a collection of scattered molecules.
- Annotate compounds and relationships:** A callout pointing to a molecule with annotations like "3 x increase" and "10 x increase".
- Display search hits:** A callout pointing to a grid of search results.

3 – Matched Molecular Pairs

A matched molecular pair (MMP) is defined as two molecules that differ from each other by a small, specified change at one or more specified locations and that share a large, identical structural feature.

Specific MMPs can be searched in two different ways:

1. By selecting a substructure of interest (see query 1), MMPs can be searched where this substructure is replaced. The resulting pairs of molecules can be sorted by the size change and property delta.
2. Starting from one molecule (see query 2), the structure is fragmented according to published definitions. Next, the database is searched for molecules that differ from the query only in the respective fragments.

This section provides examples of MMPs and activity cliffs:

- Query 1:** List of MMPs where chlorobenzene is involved. Shows pairs of molecules with their IDs and property deltas.
- Query 2:** Chemical neighborhood of query 2. Shows molecules related to a specific query.
- Easy detection of activity cliffs:** Compares two molecules with similar structures but different activity.
Left: K_i (FacVIIa): 60 nM
Right: K_i (FacVIIa): 0.9 μ M

4 – Discussion and Outlook

The innovation of this project lies in the way the system is built and in the way it is used. It will change how chemists, project teams, management teams or seminar audiences view and discuss project data. It will help to re-focus scientists' attention and depart from the canonical notion of "statistical significance", which has led to much unproductive work and is unfortunately still the basic concept in most published articles on computational SAR analysis. Ultimately, this new system will enable medicinal chemists to get a clear overview of the series of chemical structures they are working on, and it will support their thought processes to develop and optimize compound series.