Compound Design Process - To A Better Start

Aurora D. Costache¹, Anna Tomin², Ákos Tarcsey², András Stráczy², Gábor Imre², Ivan Solt²
ChemAxon LLC, 1 Broadway, Cambridge MA, 02142 USA
ChemAxon Kft. Záhony u.7 H-1031 Budapest

ABSTRACT: Compound and formulation design is an information demanding activity, since all relevant knowledge is to be accessible within a single space and requires synchronized application of computational models to assist decision making on synthesis candidates. Our study aims to evaluate a software platform coping with this complexity (Marvin Live). The tool provides central management of innovative ideas and helps triage them based on predicted properties and available knowledge collected from a variety of sources. The calculated properties span physico-chemical descriptors, combined metrics, 3D overlay and modeling results, with the ability to check if they are compliant with local and international regulations. Use cases of rapid freedom to operate analysis by using various tools from ChemAxon (JChem For Office plugins, Scaffold Based Enumerations, Mad fast Similarity Search) of exemplified structures from patents (SureCHEMBL, ~17M entries) and opened access documents and databases in real time will be shown to ensure that designers can seamlessly exploit the chemical space around their ideas. The poster will walk through an example compound design cycle to obtain statistical results regarding performance as well as to demonstrate the suitability of the calculations.

1. Characterization of a Reference Compound

The dataset was analyzed via the following workflow using ChemAxon tools:

- a) Content was downloaded from SkinSensDB public source [1].
- b) The dataset was analyzed using ChemAxon's Standardizer [4] – Fig. 1.
- c) SkinSensDB structures were used as a starting point for new compound design, analysis and SAR studies. See schematic workflow on Fig 1.

2. Compound Design

Fig 2 – Idea generation and tracing available information for compounds of interest

3. SAR Analysis

To further characterize available information about these compounds, Marvin Live platform, which offers an agnostic service to ship a set of applications for idea generation, assessment and management into a single space tailored for chemist needs, was used. Marvin Live offers both individual (private) idea repositories and collaborative (public) spaces [5].

In Fig 2 a Marvin Live workflow is illustrated to perform idea generation and track information available from different public databases like, CHEMBL and SureCHEMBL using the same scaffold shown below.

- Scaffold based enumeration was applied to generate a small combinatorial library via subtitution variations, resulting in molecules having the same core structures. The possible combinations based on the fragment decomposition illustrated in Marvin JS canvas, could be tracked underneath of it.
- Compounds can be picked and different plugins can be activated to display physico-chemical properties and their distributions (MW vs logP scatter plot), as well as available activities listed in public databases.

4. Similarity search

Mad Fast Similarity Search (MFSS) [6]engine was applied to query large available databases in real-time. The SureCHEMBL database was preprocessed 1024 bit Chemical Hashed Fingerprints for the ~ 17 million molecules. Loading these into a single virtual machine for in-memory search required ~ 24 cores and 64GB RAM. The hyper-optimized fingerprint comparison method, Madfast Similarity Search was accessed through a Marvin Live plugin, to automatically run similarity searches on the selected molecule with a maximum hit count of 20 sorted by similarity – Fig 4.

Fig 4 – Snapshot of a Mad Fast Similarity Search within SureCHEMBL database (~17 mil compounds)

Conclusions

Our workflow reveals that compound design can be fostered by providing instant access to extra large compound collections, knowledge bases and predicted properties. Therefore, it facilitates widening the scope of the chemical space and helps escape an availability bias that would constrain the series of ideas and increase the likelihood of project failure due to insufficient freedom to operate. The rich standardization (Standardizer) visualization/analysis (JChem For Office), searching (MFSS engine) and design and idea management (Marvin Live) tools provided by ChemAxon help prioritizing the compound selection and decision making in compound design processes.

References