Curation and Evaluation of an In-House Virtual Compound Library: Comparative Analysis with Commercial Collections and Synthetic Feasibility Assessment

Leonard Bui, Teodora Djikic-Stojsic, Esther Kellenberger

The initial phase of the rational drug discovery process is, mainly, based on screening a collection of chemicals, searching for a hit that binds a target of interest [1]. Therefore, chemical libraries represent a useful tool in drug discovery, offering a vast variety of compounds for screening and exploration [2]. Previously, through collective efforts of the chemoinformaticians and chemists, our group has created two libraries : essential eIMS containing 578 in-stock compounds on plates ready for High Throughput Screening and a virtual library vIMS, containing 821.069 compounds generated from the scaffolds of the eIMS compounds, and decorated with substituents from a collection of R-groups. Here, we aim to focus on the validation of such a virtual library and the decoration method that was used to produce it.

Firstly, to validate the protocol of compounds generation and the collection of R-groups used for the creation of novel molecules, a comparison between virtual libraries was made to validate the method by generating new compounds from Real Space scaffolds and comparing them with the actual compounds from Real Space, to get the overlap between the libraries. Secondly, the synthetic accessibility of the compounds was determined by combining machine learning and the chemists' expertise. Specifically, around 4000 compounds from vIMS database were randomly selected and sent to the chemists who designed the scaffold and the synthetic routes for the parent compound in eIMS, for the evaluation. Compounds are classified into three categories depending on this expert-predicted synthesizability: “Yes”, “Maybe” or “No”. Based on this information, knowledge-based models were created and applied to the entire vIMS set. Analysis of the results provides avenues for improving the virtual compound design.

Bibliography :

[1] : Saldívar-González, F.I., Huerta-García, C.S., Medina-Franco, J.L. *J Cheminform* **12**, 64 (2020)
[2] : Hartenfeller, M., Eberle, M., Meier, P. *J. Chem. Inf. Model.* , **51**, 12 (2011) 3093–3098