The IMS Library: an in-House Collection Based on Chemical Diversity, Scaffold Novelty and Synthetic Accessibility

Teodora Djikic-Stojsic[1], Guillaume Bret[1], Gaëlle Blond[1], Nicolas Girard[1], Clotilde Le Guen[1], Claire Marsol[1], Martine Schmitt[1], Séverine Schneider[1], Frederic Bihel[1], Dominique Bonnet[1], Michaela Gulea[1], and Esther Kellenberger[1]

1. *Laboratoire d'Innovation Thérapeutique, UMR7200 CNRS-Université de Strasbourg, 74 route du Rhin, Illkirch-Graffenstaden 67400, France*

The initial phase of rational drug discovery process is, to a large extent, based on screening (both HTS and VS) a collection of chemicals (library), searching for a hit, that would bind to a target of interest. It is commonly believed that the wider the chemical space is covered by the library, the better the chance that a hit compound would be found.[1] However, even though obtainable chemical spaces are expanding rapidly, containing billions of possible compounds, the usage of large data sets remains challenging. Working with such a large chemical space presents a constant compromise between cost, size, and time.[2,3] Accordingly, diversity, novelty and drug-likeliness of the library are presenting themselves as more important than its size, in order to secure a better coverage, and to explore new areas of chemical space.[4,5]

Thanks to joint efforts of chemoinformaticians and medicinal chemist, we have developed two original and diverse, in-house, ready-to-use screening libraries.

Essential chemical library (eIMS) was created from the collection of previously synthetized compounds. It comprises of 578 compounds, with 80% purity, which are plated and ready for HTS. Diverse compounds were chosen by clustering, using molecular fingerprints, while their originality was determined by comparison to our in-house database of purchasable, drug-like compounds, collected from 25 trusted suppliers.

Virtual chemical library (vIMS) consists of 783.000 virtual compounds, derived from the most interesting and original scaffolds from the Essential library. Novel compounds were generated by adding different substituents on the specified connection points of the central scaffold, following the rules determined by medicinal chemists, to ensure their synthetic accessibility. Compounds were further filtered based on their chemical properties and substructures, to remove reactive and/or peculiar structures. Finally, we compared our virtual library to the Enamine REAL Space[6], and concluded that around 90% of generated compounds could not be found there. Additional advantage of this library is the fact that it is based on well-known synthetic pathways and expert knowledge of medicinal chemists, which secures the easy synthesis of novel derivatives and facilitates hit-to-lead optimization.

Bibliography

(1) Saldívar-González, F. I.; Huerta-García, C. S.; Medina-Franco, J. L. Chemoinformatics-Based Enumeration of Chemical Libraries: A Tutorial. *Journal of Cheminformatics* **2020**, *12* (1), 64. https://doi.org/10.1186/s13321-020-00466-z.

(2) Tingle, B. I.; Tang, K. G.; Castanon, M.; Gutierrez, J. J.; Khurelbaatar, M.; Dandarchuluun, C.; Moroz, Y. S.; Irwin, J. J. ZINC-22─A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *J. Chem. Inf. Model.* **2023**, *63* (4), 1166–1176. https://doi.org/10.1021/acs.jcim.2c01253.

(3) Perebyinis, M.; Rognan, D. Overlap of On-Demand Ultra-Large Combinatorial Spaces with On-the-Shelf Drug-like Libraries. *Molecular Informatics* **2023**, *42* (1), 2200163. https://doi.org/10.1002/minf.202200163.

(4) Nahar, L.; Sarker, S. D. Chapter 5 - Application of Computation in Building Dereplicated Phytochemical Libraries. In *Computational Phytochemistry*; Sarker, S. D., Nahar, L., Eds.; Elsevier, 2018; pp 141–163. https://doi.org/10.1016/B978-0-12-812364-5.00005-5.

(5) Lyu, J.; Irwin, J. J.; Shoichet, B. K. Modeling the Expansion of Virtual Screening Libraries. *Nat Chem Biol* **2023**, 1–7. https://doi.org/10.1038/s41589-022-01234-w.

(6) *REAL Space - Enamine*. https://enamine.net/compound-collections/real-compounds/real-space-navigator (accessed 2023-05-02).