# Fragment-based de novo design and searching for hit molecules in ultra-large chemical libraries

Polishchuk Pavel1, Minibaeva Guzel1, Ivanova Aleksandra1, Kutlushina Alina1

1 Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Hněvotínská 1333/5, Olomouc, Czech Republic

A significant limitation of fragment-based structure generation is the poor synthetic accessibility of the structures produced. The previously established CReM framework [1] offers a potential solution to this issue. By integrating this framework with molecular docking, pharmacophore modeling, or machine learning, we have developed a suite of tools capable of addressing various tasks, including de novo design, hit generation, hit/lead optimization, and scaffold hopping. Nevertheless, the structures generated may still be syntheticaly infeasible. To address this issue, we proposed a pipeline that facilitates the rapid identification of hit molecules within ultra-large libraries of synthetically accessible compounds. The fundamental concept involves generating molecules de novo, selecting the most promising candidates, and utilizing these candidates for similarity searches within an ultra-large library. We validated this protocol during the first CACHE challenge, which aimed to identify binders for the WD40 domain of the LRRK2 kinase, a target that previously lacked known binders and for which only the X-ray structure of the domain was available. We employed the proposed strategy and designed promising hits de novo using the CReM-dock tool [2], which were subsequently used to search for similar molecules in Enamine REAL Space, containing approximately 23 billion structures at that time. As a result, out of 82 synthesized compounds, eight exhibited binding affinity with Kd values ranging from 25 to 117 µM, thereby confirming the efficacy of the proposed protocol. The outcomes of all top-performing teams have recently been published in a collaborative article [3]. The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through INTER-EXCELLENCE II LUAUS23262, the e-INFRA CZ (ID:90140, ID:90254), ELIXIR-CZ (LM2018131, LM2023055), CZ-OPENSCREEN (LM2018130, LM2023052) grants.

[1] Polishchuk, P. *J. Cheminf.* **2020**, 12 (1), 28.

[2] Minibaeva, G. et al. *ChemRxiv* **2024** - <https://doi.org/10.26434/chemrxiv-2024-fpzqb-v3>.

[3] Li, F. et al. *J. Chem. Inf. Model.* **2024**, 64 (22), 8521-8536.