Integration of Deep Docking into Virtual Screening Workflow. The CACHE-1 Study Case

Artem Cherkasov[1], Evgeny Gutkin[3], Filipp Gusev[3,4], Francesco Gentile[1,2], Fuqiang Ban[1], S Benjamin Koby[3], Chamali Narangoda[3], Maria G Kurnikova[3], Olexandr Isayev[3,4]

[1]Vancouver Prostate Centre, University of British Columbia, 2660, Oak street, Vancouver, BC, V6H1A3, Canada.
[2]Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON, Canada`
[3]Department of Chemistry, Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[4]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

With the recent explosion of chemical libraries into- and beyond billion molecules scales, the use of ML methodologies become an integral part of virtual screening (VS). Thus, the DL-based approach Deep Docking (DD) has been pioneered in 2019 to significantly accelerate structure-based VS by docking only a subset of a chemical library, while iteratively synchronizing with ligand-based prediction of the remaining docking scores [1].
When benchmarked against conventional docking, Deep Docking demonstrated hundreds-to-thousands fold hit enrichment without significant loss of potential target binders. Thus, the use of DD, that relies on computationally inexpensive protein-independent 2D chemical fingerprints, makes it particularly suited for screening ultra-large chemical libraries.
On the other hand, ML acceleration and active learning enables implementation of stringent consensus VS protocols by engaging more scoring methods and enabling hit-selection strategies with minimized human intervention [2].
Thus, we have recently utilized DD component in VS screening pipeline developed to participate in the Critical Assessment of Computational Hit-Finding Experiments (CACHE) Challenge. The DD-accelerated Glide protocol was used to rapidly score LRRK2 WDR domain (drug target implicated in Parkinson's) against 4.1 billion ligands from Enamine REAL database. The DD selection of ~800 molecules was followed by absolute binding free energy (ABFE) calculations, in turn, relying on automated molecular dynamics based thermodynamic integration (MD TI) [3]. As the result, we selected 76 ligands for experimental validation, which revealed a significant number of actives ranking our prediction on the top of CACHE-1 competition (tied for the 1st place among 23 finalists).
Thus, the use of consensus VS protocols relying on ML-accelerated methodologies and active learning appear very effective for screening ultra-large chemical libraries and rigorous protein-ligand binding affinity estimation leveraging modern computational resources.

Bibliography:

[1] F. Gentile et al, ACS Cent. Sci. 6 (2020), 939–949
[2] F. Gentile et al, Chemical science 12 (2021), 15960-15974
[3] E. Gutkin et al, chemrxiv. (2024), doi.org/10.26434/chemrxiv-2023-lnzvr