Computational Automation of Pharmacophore Model Optimization

Katharina Munk[1], Veronika Temml[2], Susanne Schaller[1], Stephan Winkler[1], Thierry Langer[3], Daniela Schuster[2]

[1]Bioinformatics Research Group, University of Applied Sciences Upper Austria, Campus Hagenberg, Austria;
[2]Paracelsus Medical University, Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Salzburg, Austria;
[3]University of Vienna, Department of Pharmaceutical Sciences, Faculty of Life Sciences, Vienna, Austria

Pharmacophore modeling and virtual screening play a central role in modern drug discovery by enabling the identification of promising therapeutic candidates. The quality of the underlying pharmacophore models is therefore critical for achieving accurate and meaningful screening results. In practice, researchers often iteratively adjust and refine pharmacophore models to improve their predictive performance. However, the manual refinement of pharmacophore models to identify a higher number of active candidates while minimizing inactive ones remains time-intensive and inefficient. Manually optimizing multiple models for a single drug target can take several weeks, slowing down the overall drug-development process.
Therefore, we introduce a computational, fully automated approach to streamline the manual optimization process by developing algorithmic solutions that replicate expert-driven optimization strategies. Implemented in Python as an autonomous command-line workflow, the system interfaces directly with LigandScout [1], eliminating the need for manual intervention and ensuring consistent, reproducible refinement of pharmacophore models. Starting from the standard model, over feature tolerance adaptations to increase the number of actives hits, to precise placement of exclusion volumes to decrease the number of inactive hits, the automated workflow covers all optimization steps. The resulting pipeline reduces researcher workload while maintaining high model quality, effectively accelerating the optimization stage of drug discovery.
Initial results show that the optimization of a single model has a runtime of around four hours, when using an active library with approx. 780 compounds and an inactive library with approx. 260 compounds. Additional tests, using 2000 inactives and 50 actives, led to a runtime of around 8-10 hours for the optimization of one model. This means that optimizing multiple models for a single drug target would take only a few days, compared to several weeks when the optimization is performed manually. Additionally, preliminary analyses of automatically optimized models indicate similar or even better quality than the manually optimized models.
In the future several extensions are planned, including the incorporation of heuristic strategies, aim to further improve performance and broaden applicability to large-scale virtual screening campaigns, ultimately supporting faster and more resource-efficient identification of novel drug candidates.

[1] Wolber, G., & Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. Journal of chemical information and modeling, 45(1), 160-169.