Data-Driven Exploration of Kinase Inhibitor Space:

AI-Assisted, Structure-Based, and Fragment-Guided Discovery

Andrea Volkamer, Data Driven Drug Design, Saarland University, Saarbrücken, GER

Despite the availability of over 6,000 kinase structures and over 70 approved small molecule inhibitors, challenges such as selectivity, resistance, and the vastness of unexplored chemical space persist in kinase drug discovery. Addressing these issues requires data-driven strategies that efficiently combine structural insights, fragment-based design, and machine learning.

To support training and reproducibility in this rapidly evolving field, we first highlight recent updates to our open-source TeachOpenCADD platform, with a dedicated focus on kinases and molecular deep learning [1]. The new teaching material provides FAIR, hands-on pipelines that bridge classical structure-based drug design with modern deep learning methods, lowering the barrier for education and method adoption in data-driven kinase research.
Second, we introduce guided docking as a data generation strategy for structure-based machine learning [2]. By integrating prior structural knowledge into docking workflows, this approach enables the generation of high-quality, task-specific training data and substantially improves deep learning performance for kinase affinity prediction.
Third, we present fragment-guided exploration of kinase chemical space based on KinFragLib [3], a kinase-focused, subpocket-annotated fragment library derived from co-crystallized complexes. To ensure tractability and chemical relevance, CustomKinFragLib [4] applies medicinal chemistry-informed filtering, drastically reducing combinatorial complexity while preserving diversity. The refined fragment space is explored using a subpocket-aware, template-based docking pipeline [5], enabling the systematic generation and prioritization of novel, synthesizable kinase inhibitors. In a PKA case study, 11 candidates were synthesized and tested, with 4 compounds reducing kinase activity below 50% and 2 below 10% at 100 µM.
Finally, we integrate fragment-based active learning to further accelerate chemical space exploration. Surrogate models iteratively approximate docking scores and guide fragment additions, deletions, and exchanges, balancing exploration and exploitation while maintaining molecular diversity. Applied to PKA kinase, we observed an improvement in binding affinity of the generated ligands over each active learning cycle, while retaining a high molecular diversity among the generated ligands. This kinase-agnostic strategy enables efficient navigation of vast combinatorial spaces with reduced computational costs.
Together, these strategies highlight the synergy of fragment-based design, structural knowledge, and machine learning in accelerating the discovery of novel, diverse, and synthesizable kinase inhibitors.

References:

[1] M. Backenköhler, *et al., ChemRxiv*, **2023**, Doi: 10.26434/chemrxiv-2023-kz1pb
[2] M. Backenköhler, *et al.*, *JCIM*, **2024**, 64, 10, 4009–20, Doi: 10.1021/acs.jcim.4c00055
[3] S. Dominique, *et al.*, *JCIM*, **2020**, 60(12), 6081-94. Doi: 10.1021/acs.jcim.0c00839
[4] PL. Kramer, *et al.*, *ACS Omega,* **2026**; Doi: 10.1021/acsomega.5c12231
[5] K. Buchthal, PL. Kramer, *et al., ChemRxiv*, **2025**; Doi:10.26434/chemrxiv-2025-f9ctg