Accelerating Drug Discovery Through Machine Learning in Chemical and Biological Space

Machine learning (ML) is now everywhere in drug discovery. Yet despite the excitement, many questions remain about where these methods truly add value and where they fall short. In this talk, we will discuss how ML can help explore chemical and biological space, while arguing that its true potential lies when combined with physics-based methods and experimental validation.

First, we will present DockBox2, a graph neural network framework that learns from ensembles of protein-ligand conformations rather than individual docking or crystallographic poses. By combining information from multiple binding modes, the model improves pose prediction, binding affinity prediction, and virtual screening performance compared with traditional and ML docking workflows[1]. We will then show how this approach can be applied to real drug discovery projects, including the identification of novel inhibitors of the dengue virus envelope protein targeting a previously unexplored pocket, with activity against all four dengue serotypes.

We will then move beyond small molecules and introduce a multimodal machine learning platform developed for antibody-drug conjugates (ADCs). By integrating ADC chemical structure with transcriptomic and proteomic characteristics of cancer cells, the model can predict ADC activity across diverse tumor types and payload classes. Prospective validation across previously unexplored ADC-cell line combinations demonstrated that combining chemistry with biological context substantially improves predictive power[2].

Together, these examples highlight both the opportunities and limitations of ML across diverse therapeutic modalities. Rather than replacing experiments or molecular modeling, ML is emerging as a practical tool for guiding exploration and making better use of increasingly large chemical and biological datasets.

References

1. Thaingtamtanha, T., Preto, J. & Gentile, F. Pose ensemble graph neural networks to improve docking performances. *Chem. Sci.* **16**, 19876–19887 (2025).

2. Mslati, H. *et al.* Interconnecting ADC Structure with Tumor Cell Biology with Multimodal Learning. 2026.05.22.727320 Preprint at https://doi.org/10.64898/2026.05.22.727320 (2026).