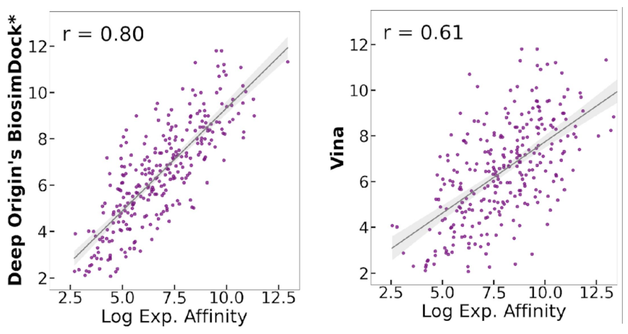
Accelerating Drug Discovery with Physics-Informed Machine Learning

G. Petrosyan, V. Altunyan, K. Smbatyan, T. Ghukasyan, G. Arakelov, V. Arakelov, A. Davtyan, H. Saribekyan, L. Tsidilkovski, T. Abramyan, M. Ratnikov, P. Janczyk, N. Ma, A. Papoyan, G. Papoian

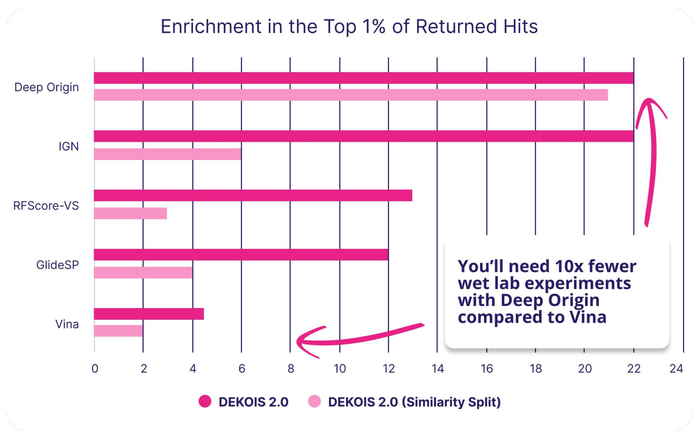
Despite substantial advances in ML for drug discovery, many difficult targets remain beyond reach. Virtual screening of ultra-large libraries holds promise, but docking remains largely unresolved for both pose prediction and binding-affinity estimation. Deep Origin's physics-informed ML docking advances both objectives: in retrospective benchmarks against experimental structures and binding data, it improves pose accuracy and affinity rank-ordering versus established methods. In prospective assays, we observe dramatically higher enrichment than current state-of-the-art approaches.

We have developed a computational drug-discovery pipeline that integrates proprietary docking, chemical-property prediction, virtual screening, and molecular-dynamics tools. To broaden accessibility, the pipeline is available via a code-based Python API and a chat-based natural-language interface. Together, these capabilities provide accurate, predictive computational chemistry to medicinal chemists and support accelerated discovery and development of small-molecule therapeutics.

DODock outperforms other models on predicting small molecule binding affinities on the PDBbind core dataset

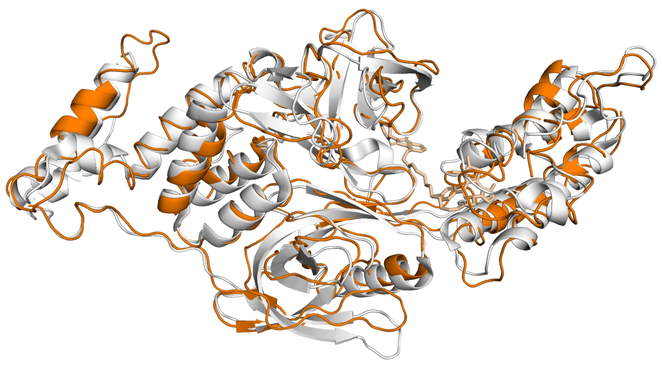


Correlation between the docking scores (absolute values) and log experimental binding affinity for Deep Origin DODock trained only on protein sequences with 30% or less homology and ligands with 0.5 or less Tanimoto similarity versus test set, and Autodock Vina. The dataset is the PDBbind core set (285 protein-ligand complexes with measured dose-dependent experimental affinities).*



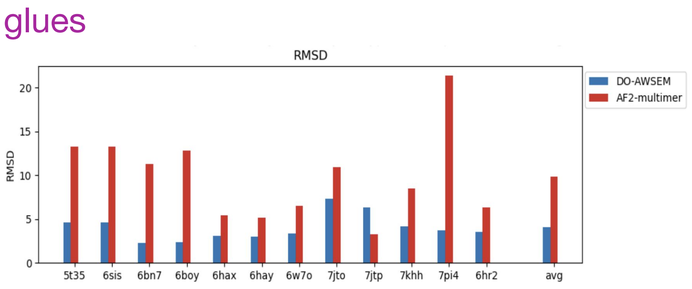
DODock outperforms many popular docking tools in the ability to identify true binders from decoys. The DEKOIS 2.0 dataset contains 80 target proteins with true binders and decoy molecules that are similar in physical and chemical properties, but do not bind the target protein. Dark pink bars are for all data, while light pink bars (Similarity Split) are the results after filtering out targets and ligands that are similar to training data. This enrichment increases the ability to find true binders, providing more freedom to optimize and increasing likelihood of novel patent space.

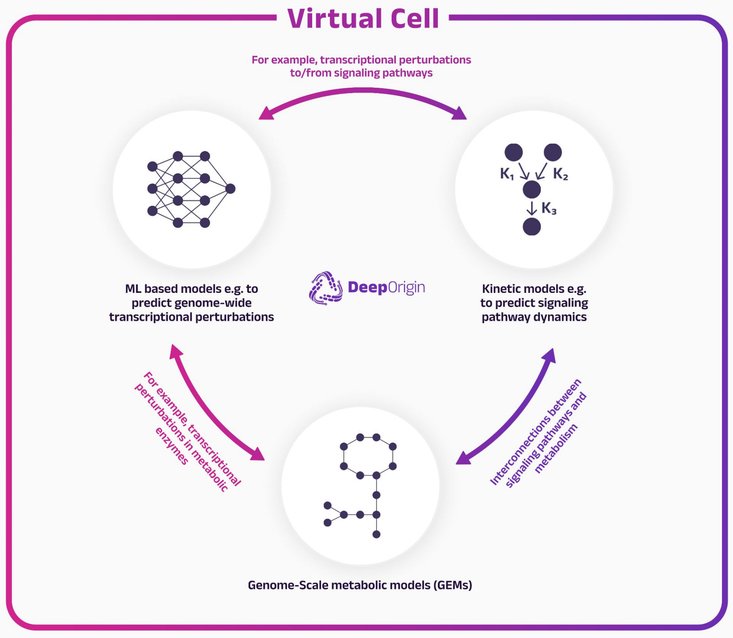
DO-AWSEM binding prediction between E3 ligase and a target protein, when the PROTAC ligand is absent!



Overlap with native complex of Cereblon and BRD4BD1 mediated by dBET6 PROTAC
PDB ID: 6BOY. RMSD: 1.5 Å. The native structure is shown in white and prediction in orange.

Modeling conformations of PROTACs and glues

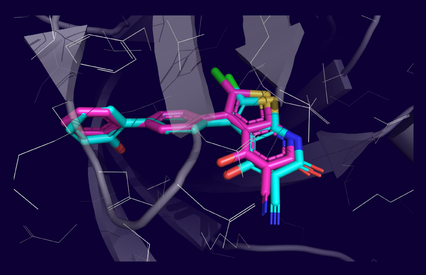




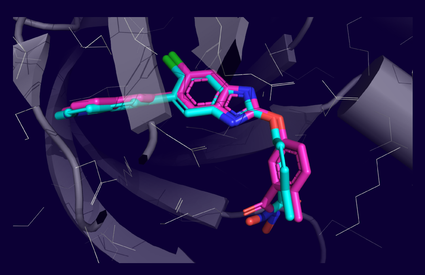
Different modeling formalisms connected together through compositional frameworks, e.g. Vivarium

DODock exhibits stronger out-of-distribution generalization than AF3 and peers

Runs N' Poses benchmark,P. Škrinjar, J. Eberhardt, et al. "Have protein-ligand cofolding methods moved beyond memorisation?" doi: https://doi.org/10.1101/2025.02.03.636309

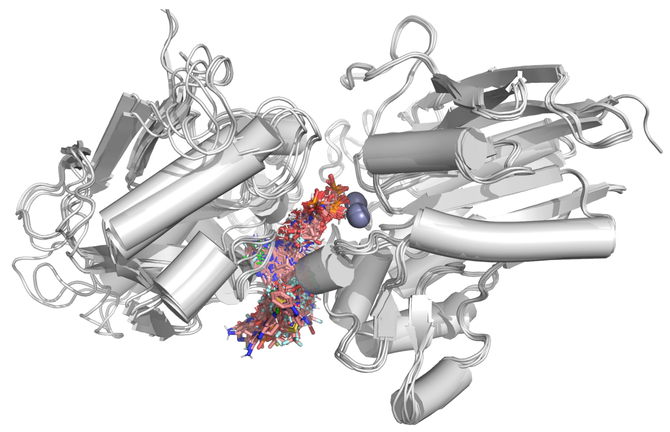


4QFR_32J: 0.5 Å



6C9J_R34: 1.0 Å

High hit rates and broad chemotype diversity achieved in prospective discovery on difficult targets

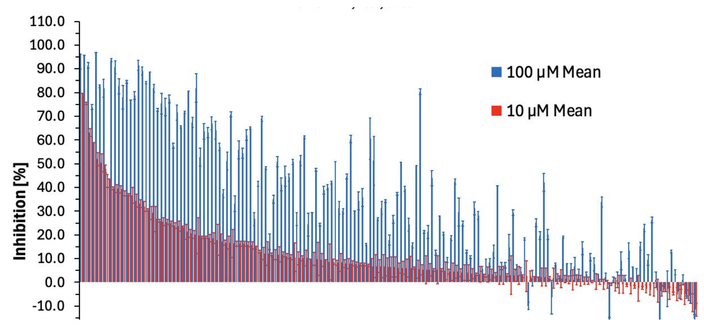


CD73 (5'-nucleotidase NT5E). Cell-surface enzyme converting AMP to adenosine; modulates immune suppression and tumor metabolism. Widely expressed in cancers – found on tumor cells, stromal and endothelial cells, and infiltrating immune cells. Elevated soluble CD73 levels correlate with poor prognosis across multiple tumor types.

Most CD73 inhibitors are stable nucleotide analogs, commonly showing limited selectivity and potency at high AMP concentrations.

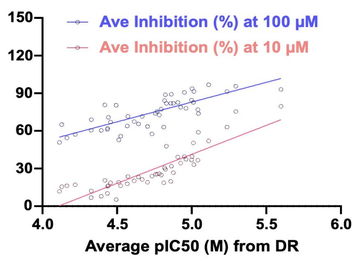
Non-nucleotide inhibitor scaffolds are very promising, but prior virtual screening campaigns have been largely unsuccessful.

~9B REAL DB +
~71B Unenumerated REAL Space
+ Reinforcement learning
160 diverse candidates experimentally tested

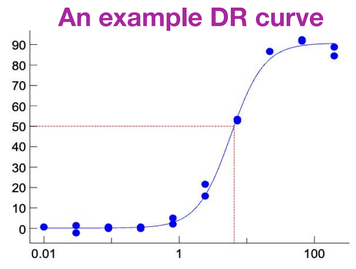


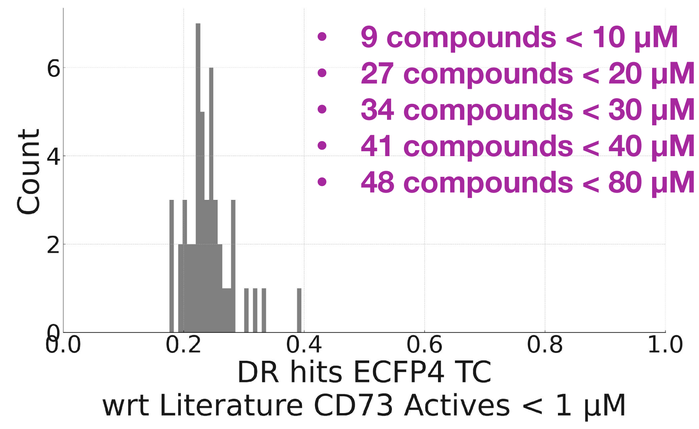
A biochemical CD73 enzyme inhibition assay using AMP as substrate to measure IC_{50} values for test compounds via a 12-point dilution series with known CD73 inhibitors as controls.

Correlation of SD with DR



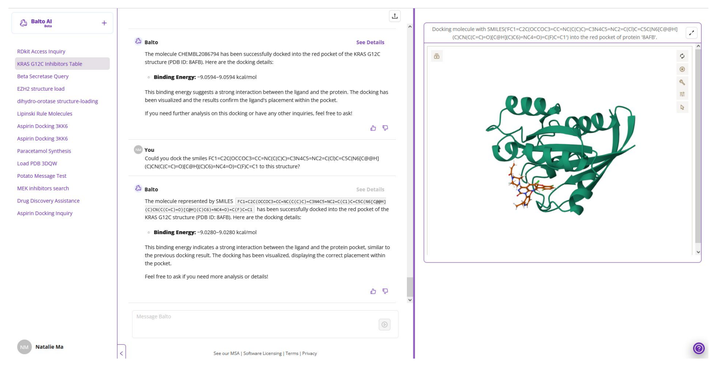
An example DR curve





- **9 compounds < 10 µM**
- **27 compounds < 20 µM**
- **34 compounds < 30 µM**
- **41 compounds < 40 µM**
- **48 compounds < 80 µM**

Balto: the first AI Assistant in Drug Discovery



Balto provides a chat-based interface to load protein structures, identify binding pockets, create structures and predict properties of molecules, and dock molecules to pockets. Balto can also do research, summarize publications, analyze images, and search the web for answers.

DeepOrigin
Dr. Garegin Papoian (CSO), gpapoian@deeporigin.com
6000 Shoreline Ct. Ste 325, South San Francisco, CA 94080

Interested in collaborating?
Emails us here!
Dr. Natalie Ma: nma@deeporigin.com