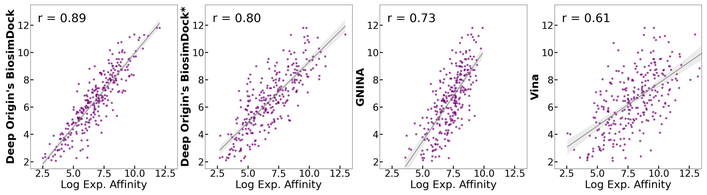
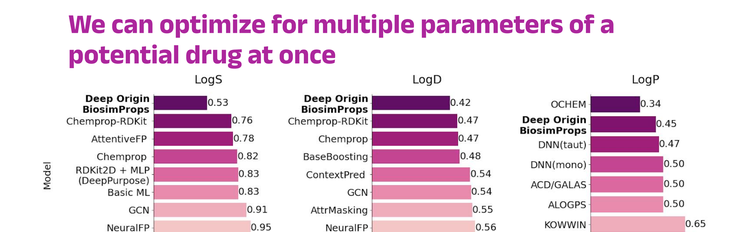
BiosimDock outperforms other models on predicting small molecule binding affinities on the PDBbind core dataset



Correlation between the docking scores (absolute values) and log experimental binding affinity for Deep Origin BiosimDock, Deep Origin BiosimDock trained only on protein sequences with 30% or less homology and ligands with 0.5 or less Tanimoto similarity versus test set, GNINA, and Autodock Vina. The dataset is the PDBbind core set (285 protein-ligand complexes with measured dose-dependent experimental affinities). [1], [2], [3], [4]*

Our model outperforms others in predicting the binding pose of ligands in the PDBbind core dataset



We outperform other models on docking accuracy, as benchmarked on the PDBbind core dataset, which contains crystal structures of 285 protein-ligand pairs. Accuracy is measured by the percentage of ligand poses predicted within 2 Å of experimental results, which is generally considered accurate for drug discovery. There are no error bars because each protein-ligand pair is assessed to have a binary 'yes' or 'no' value. [1], [2], [3], [4]

Our performance on predicting ligand binding poses by protein class, as defined by Expasy's enzyme classifications. Accuracy is measured by the percentage of ligand poses predicted to be within 2 Å of experimental results, which is generally considered accurate for drug discovery. The dataset used is PDBbind core set; there are no error bars because each protein-ligand pair is assessed to have a binary 'yes' or 'no' value. [1], [2], [3], [4]

We better identify true binders from decoys

We outperform some popular docking tools, including industry standard such as GlideSP [5], 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 [4]. Dark pink bars are for all data, while light pink bars (Similarity Split) are the results after filtering out targets and ligands from the test data similar to training data. [6], [7]

We can optimize for multiple parameters of a potential drug at once

Performance of our models for solubility in water (logS), simple octanol-water partition coefficient not accounting for molecular charge (logP), and a more realistic octanol-water partition coefficient accounting for pKa (logD). The performance is reported in mean absolute error (MAE) for logS and logD, and in root-mean-square error (RMSE) for logP, from Therapeutics Data Commons datasets. Lower numbers indicate less error, and thus greater accuracy. [8], [9], [10]

DeepOrigin

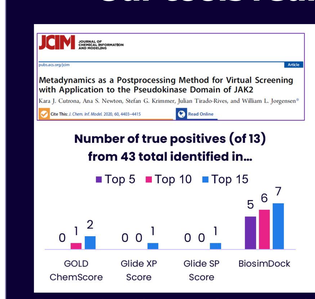
Addressing Drug Discovery Challenges with a Multiscale Molecular Modeling Pipeline

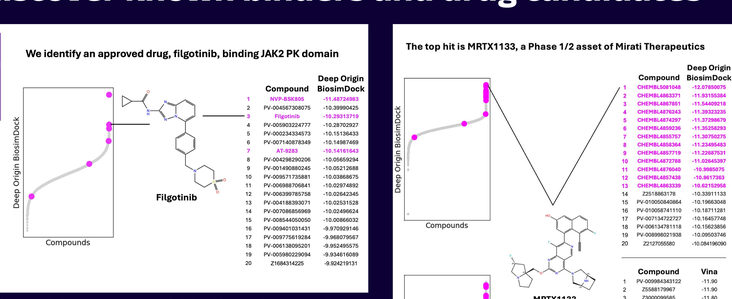
Garik Petrosyan, Hayk Saribekyan, Aram Davtyan, Tigran Abramyan, Natalie Ma, Ashot Papoyan, Jason Sunardi, Garegin Papoian

Iterating over the vast chemical search space to find the best possible drug candidates remains challenging, despite decades of *in silico* technology development in virtual screening and molecular dynamics. In particular, three challenges stand out: 1) the weak ability of virtual screening to distinguish true binders from false positives and provide accurate information on their binding affinity and conformation, leading to synthesis of 100's to 1000's of inactive compounds, 2) the difficulty in optimizing multiple chemical properties simultaneously to ensure identified compounds are suitable drug candidates, leading to time- and cost-intensive optimization campaigns, and 3) the steep learning curve of computational tools that remains a barrier-to-entry for most medicinal chemists. These challenges drive repeated cycles of *in silico* and experimental work, stretching preclinical discovery and optimization campaigns from months into years.

To address these challenges, we used physics and AI to develop a computational drug discovery pipeline comprising proprietary docking, chemical property prediction, virtual screening, and molecular dynamics, and protein dynamics tools. We demonstrate that these tools enable rediscovery and enrichment of known binders and drugs against diverse protein targets, including kinases, non-kinase nucleotide binders, receptors, and proteases. To enable broader accessibility, these tools are available with a code-based Python interface and a chat-based natural language interface. Combined, we aim to provide accurate, predictive computational chemistry tools to medicinal chemists across backgrounds and enable accelerated discovery and development of small-molecule therapeutics.

Our tools rediscover known binders and drug candidates





***Top:** Benchmarking our performance against well-known commercial docking software and identify known binders, using JAK2 pseudokinase as an example (screenshot of paper header) [11].We used the same PDB structure for JAK2 (PDB ID 4FVR) that is referenced in Cutrona et al., 2020. Of note, filgotinib is approved as a JAK1 inhibitor but demonstrates activity against JAK2 as well [12]. Schrodinger's terms of service make it difficult to benchmark directly, so we must use published retrospective data.*

***Right:** We re-discover known, experimentally-validated binders of KRAS G12D at a much higher rate than Autodock Vina. Starting library is 16 experimentally validated binders in a pool of 100,000 randomly selected druglike molecules; PDB ID 7RPZ used for docking.*

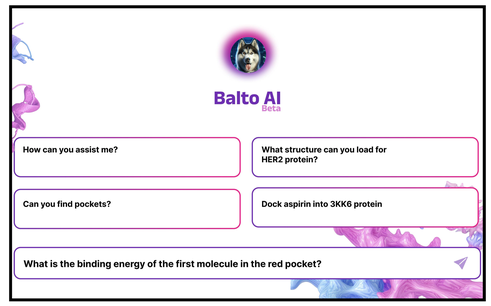
About Deep Origin

We help scientists solve disease and extend health-span by building tools that simplify R&D, simulate biology, and untangle the complexity of life. We provide our tools as software and work in partnership with others to design better therapeutics faster.

References

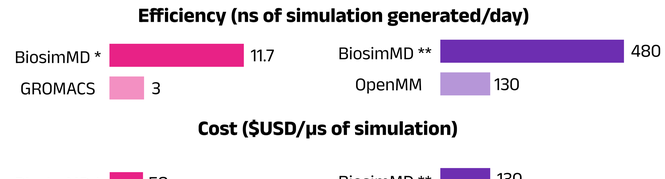
[1] http://www.pdbbind.org.cn/ [2] https://www.pharmchem.uni-tuebingen.de/dekois/ [3] Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., Wang, R., 2019. Comparative Assessment of Scoring Functions: The CASF-2016 Update. J. Chem. Inf. Model. 59, 895–913. https://doi.org/10.1021/acs.jcim.8b00545 [4] Bauer, M.R., Ibrahim, T.M., Vogel, S.M., Boeckler, F.M., 2013. Evaluation and Optimization of Virtual Screening Workflows with DEKOIS 2.0 – A Public Library of Challenging Docking Benchmark Sets. J. Chem. Inf. Model. 53, 1447–1462. https://doi.org/10.1021/ci400115b [5] Corso, G., Stärk, H., Jing, B., Barzilay, R., Jaakkola, T., 2022. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. https://doi.org/10.48550/ARXIV.2210.01776 [6] Ibrahim, T.M., Bauer, M.R., Boeckler, F.M., 2015. Applying DEKOIS 2.0 in structure-based virtual screening to probe the impact of preparation procedures and score normalization. J Cheminform 7, 21. https://doi.org/10.1186/s13321-015-0074-6 [7] Boeckler, F.M., Bauer, M.R., Ibrahim, T.M., Vogel, S.M., 2014. Use of DEKOIS 2.0 to gain insights for virtual screening. J Cheminform 6, O24. https://doi.org/10.1186/1758-2946-6-S1-O24 [8] https://tdcommons.ai/benchmark/admet_group/06logsol/ [9] Ulrich, N., Goss, K.-U., Ebert, A., 2021. Exploring the octanol–water partition coefficient dataset using deep learning techniques and data augmentation. Commun Chem 4, 90. https://doi.org/10.1038/s42004-021-00528-9 [10] https://tdcommons.ai/single_pred_tasks/adme/#lipophilicity-astrazeneca [11] Cutrona, K.J., Newton, A.S., Krimmer, S.G., Tirado-Rives, J., Jorgensen, W.L., 2020. Metadynamics as a Postprocessing Method for Virtual Screening with Application to the Pseudokinase Domain of JAK2. J. Chem. Inf. Model. 60, 4403−4415. https://doi.org/10.1021/acs.jcim.0c00276 [12] Newton, A.S., Deiana, L., Puleo, D.E., Cisneros, J.A., Cutrona, K.J., Schlessinger, J., Jorgensen, W.L., 2017. JAK2 JH2 Fluorescence Polarization Assay and Crystal Structures for Complexes with Three Small Molecules. ACS Med. Chem. Lett. 8, 614–617. https://doi.org/10.1021/acsmedchemlett.7b00154 [13] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE. Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005 Dec;26(16):1701-18. doi: 10.1002/jcc.20291 [14] Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, Pande VS. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol. 2017 Jul 26;13(7):e1005659. doi: 10.1371/journal.pcbi.1005659

Balto: the first AI Assistant for Drug Discovery



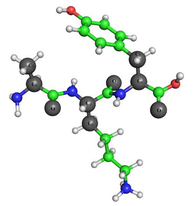
*Balto provides a **chat-based interface** to load protein structures, identify binding pockets, create structures and predict properties of molecules, and dock molecules to pockets. Balto can also summarize publications, analyze images, and search the web for answers.*

More efficient Molecular Dynamics

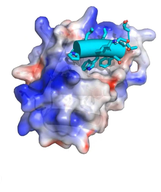


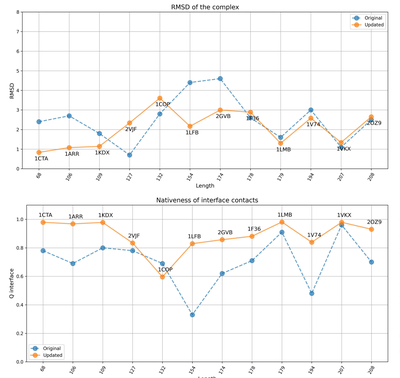
*We generate more simulations for less overall cost compared to GROMACS [13] or OpenMM [14], without the need for a GPU. AFBE and RFBE methods are also available. * On one core, Intel 3.8 GHz CPU; ** 12 simulations run in parallel, 8 cores each, to approximate OpenMM GPU.*

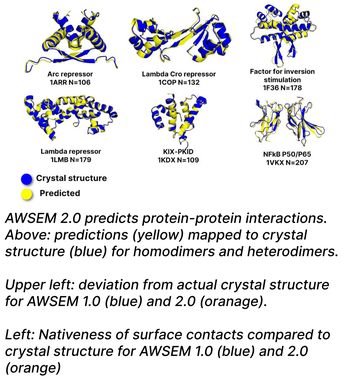
AWSEM 2.0: Protein Dynamics



- *Enables modeling of protein dynamics, protein-protein and protein-macromolecule interactions*
- *Each amino-acid is represented by three beads placed at the positions of Cα, Cβ, O*
- *The interactions are based on an energy function that balances physical terms and bioinformatics*
- *Simple yet effective representations allows for description of complex behavior of polypeptide chain*







AWSEM 2.0 predicts protein-protein interactions. Above: predictions (yellow) mapped to crystal structure (blue) for homodimers and heterodimers.

Upper left: deviation from actual crystal structure for AWSEM 1.0 (blue) and 2.0 (oranage).

Left: Nativeness of surface contacts compared to crystal structure for AWSEM 1.0 (blue) and 2.0 (orange)

Want to try out Balto?

Interested in Collaborating?

Email us here!
nma@deeporigin.com

Contact: Natalie Ma, nma@deeporigin.com
486 Cabot Road, South San Francisco, CA 94080.