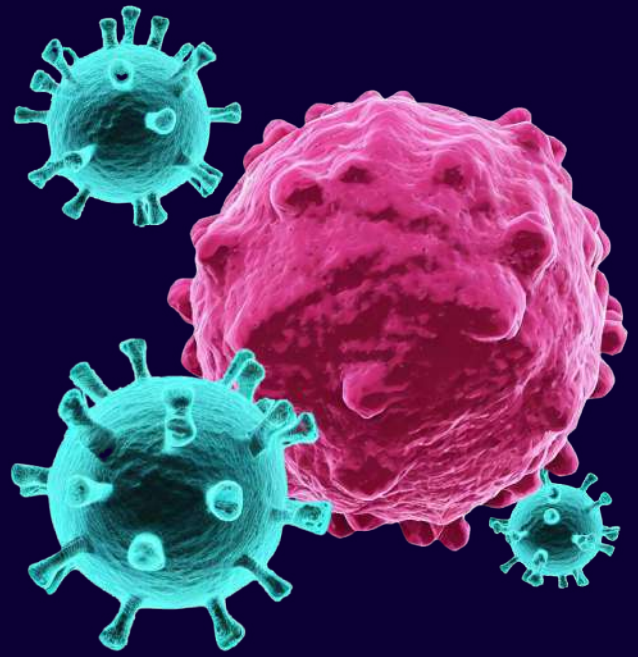
DeepOrigin

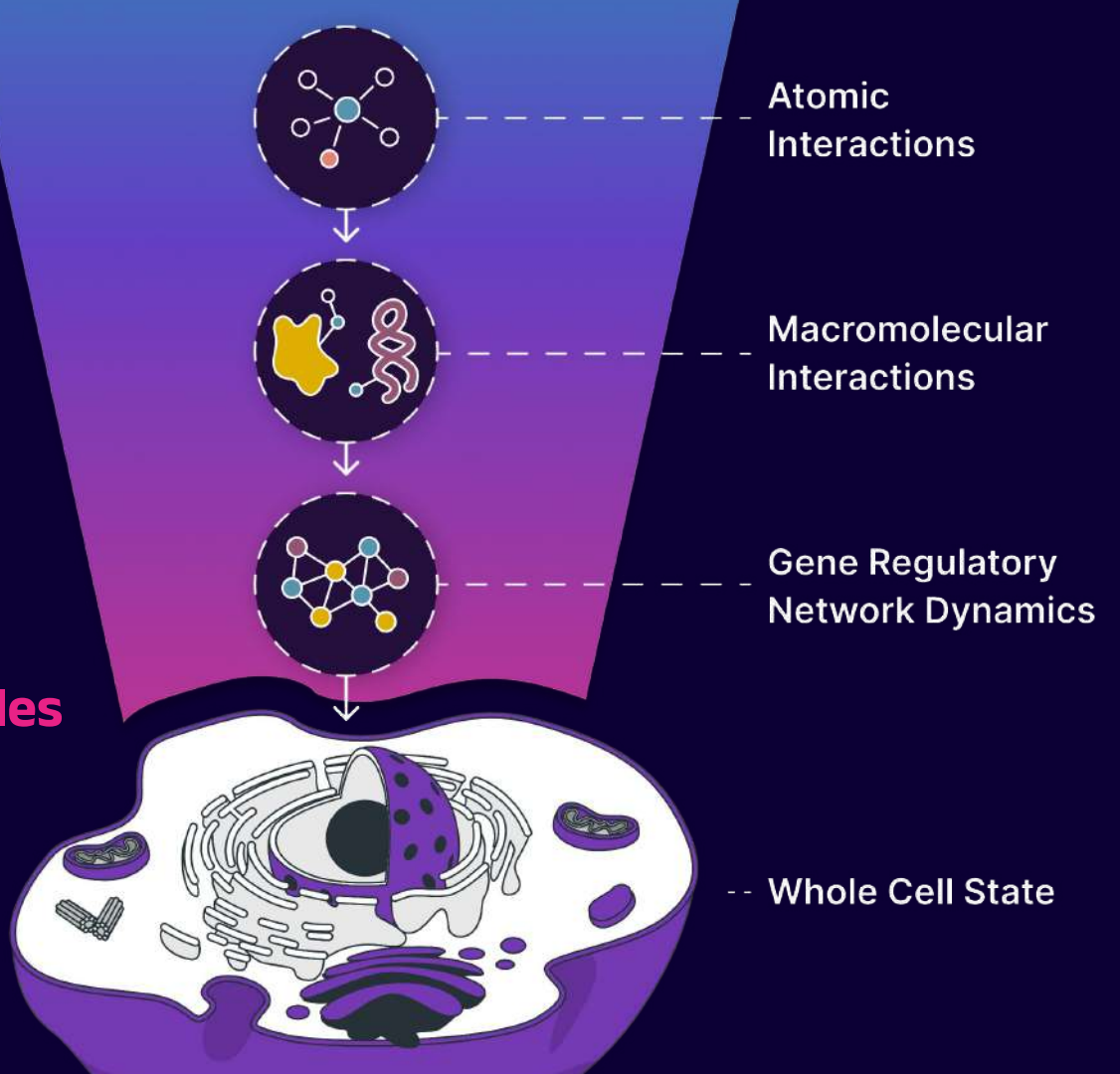
Overcoming the accuracy-generalization tradeoff in computational drug discovery



Dr. Garegin Papoian, CSO
DO Webinar, 9/2/2026

Our mission

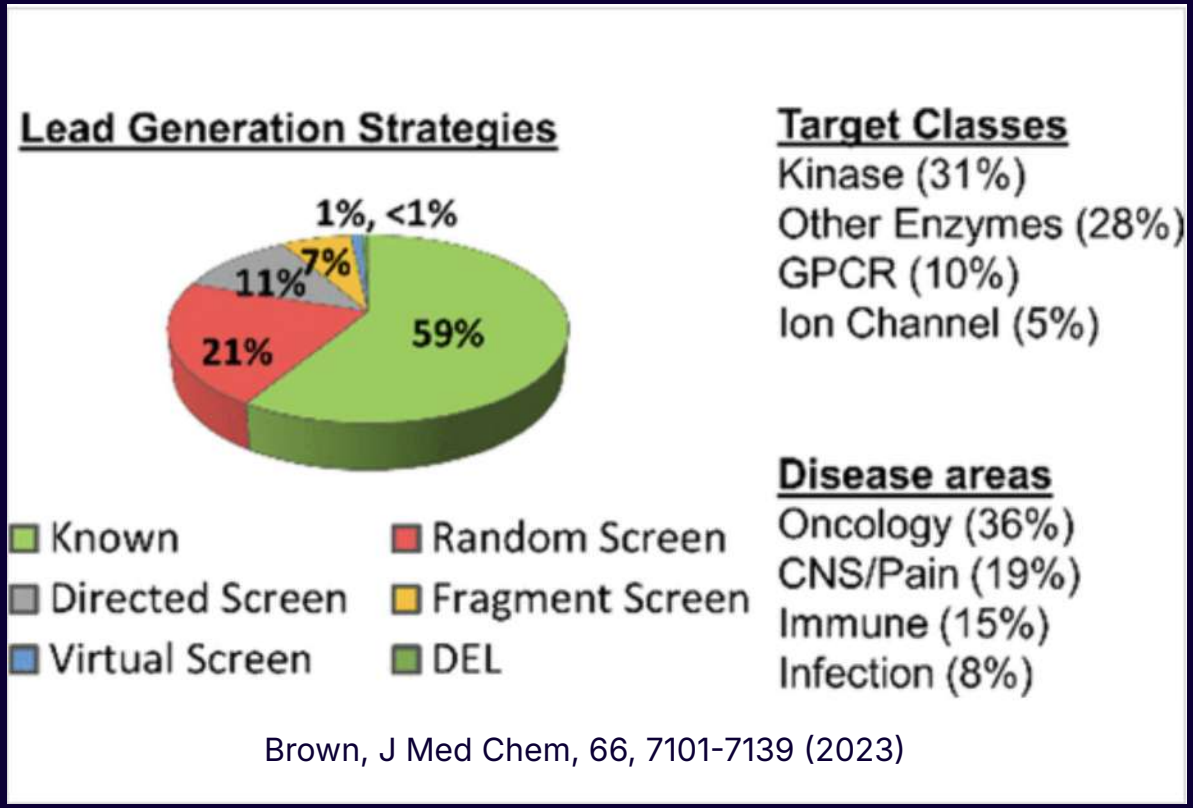
Build a virtual cell to de-risk drug discovery and development across scales



DeepOrigin

Improving both generalization and accuracy of docking and screening methods

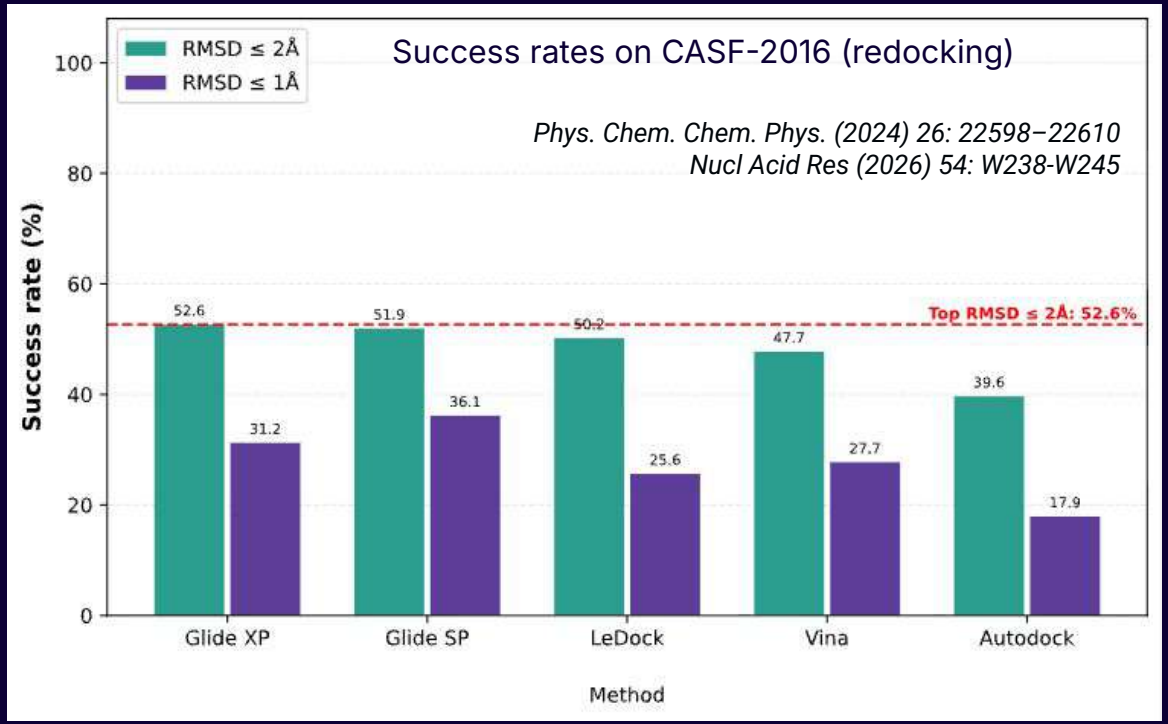
Virtual screening is not commonly used in drug discovery



Brown, J Med Chem, 66, 7101-7139 (2023)

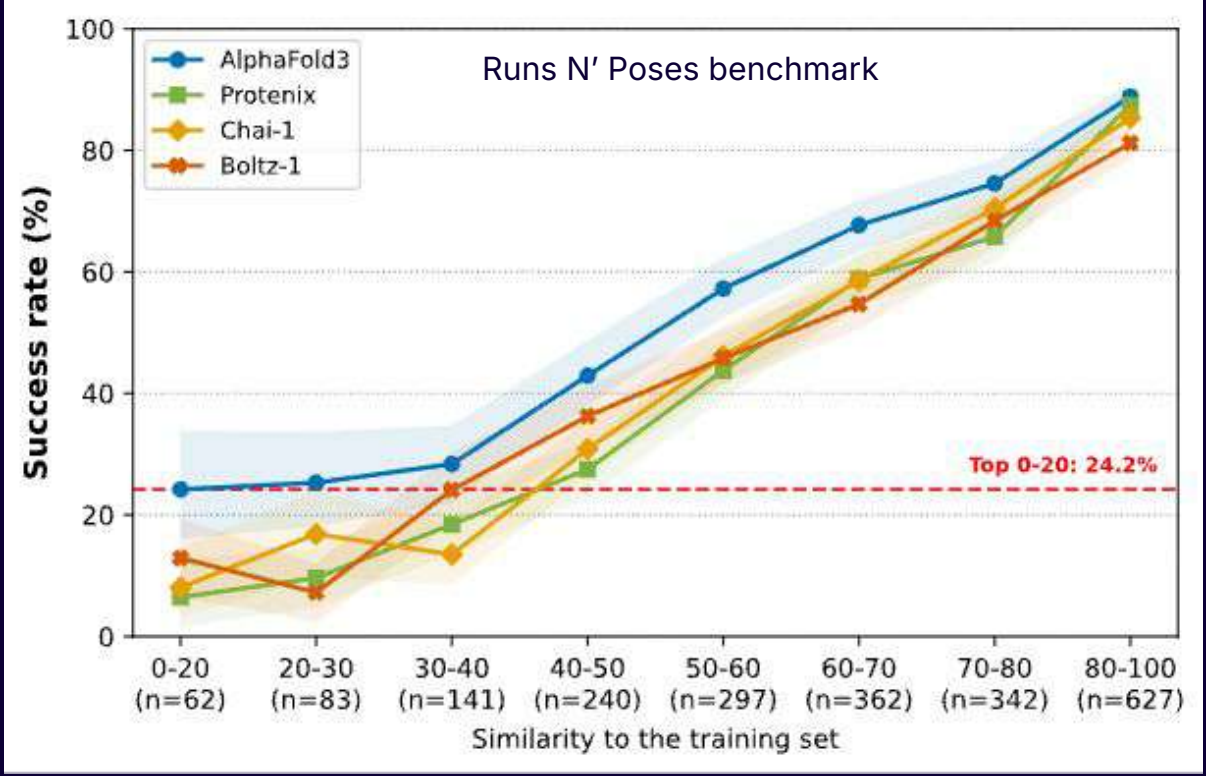
- Only about 1% of lead generation has been attributed to virtual screening
- Virtual screening has contributed to relatively few clinical candidates relative to its prominence in the methods literature
- **Why?**
  - **Persistent inaccuracies in:**
    - Docking - protein–ligand structure prediction
    - Scoring - estimating binding affinity

Classical docking methods seem to have an accuracy ceiling



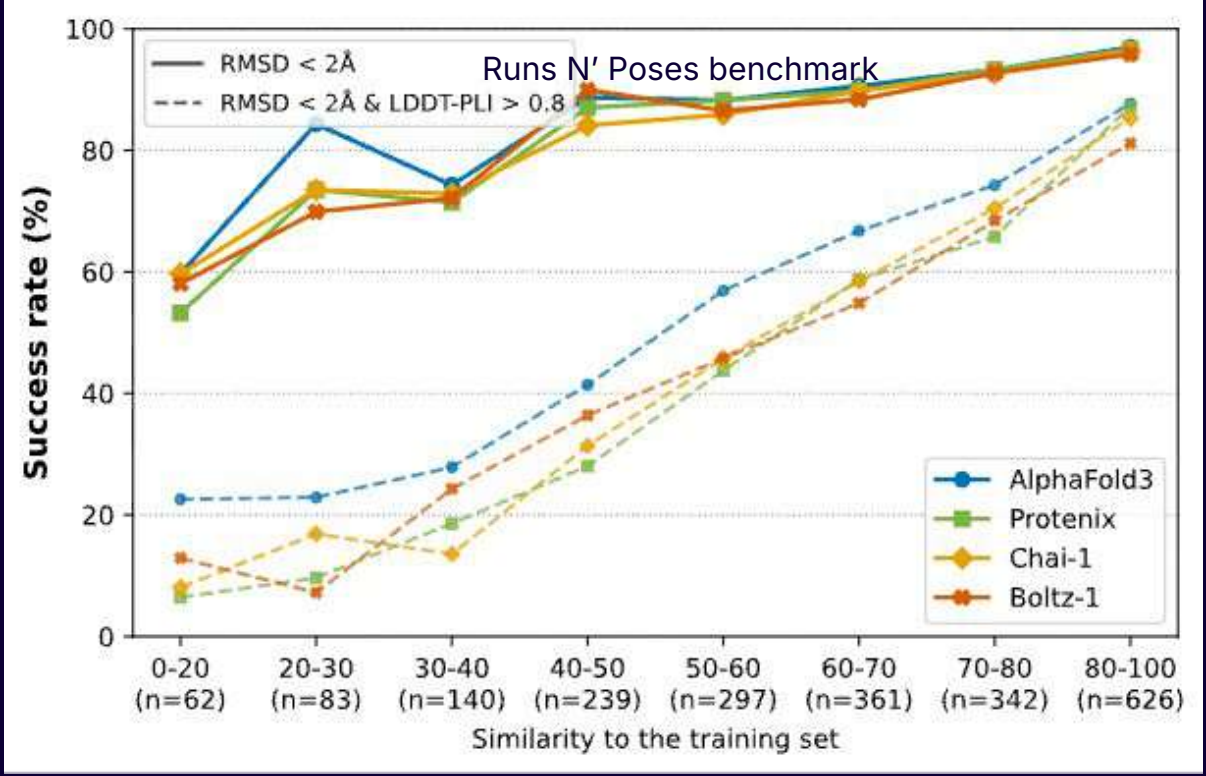
- Pros:
  - Fast
  - Parsimonious models - generalize well
  - Physics based - interpretable
  - Multiple mature methods, tens of thousands of papers
- Cons:
  - Interaction terms are too simplistic, often pairwise only
  - Too much reliance on geometric shape complementarity
  - Ignore idiosyncratic electronic structures of highly chemical diverse molecules, polarization and orbital effects

Cofolding methods tend to over-memorize



- Difficult to evaluate methods' generalization
- Persistent problems with data leakage from train to test
- Time based splits can be opaque
- Success rates of these models fall steeply with decreasing similarity to the training set, from roughly 80% to below 25%

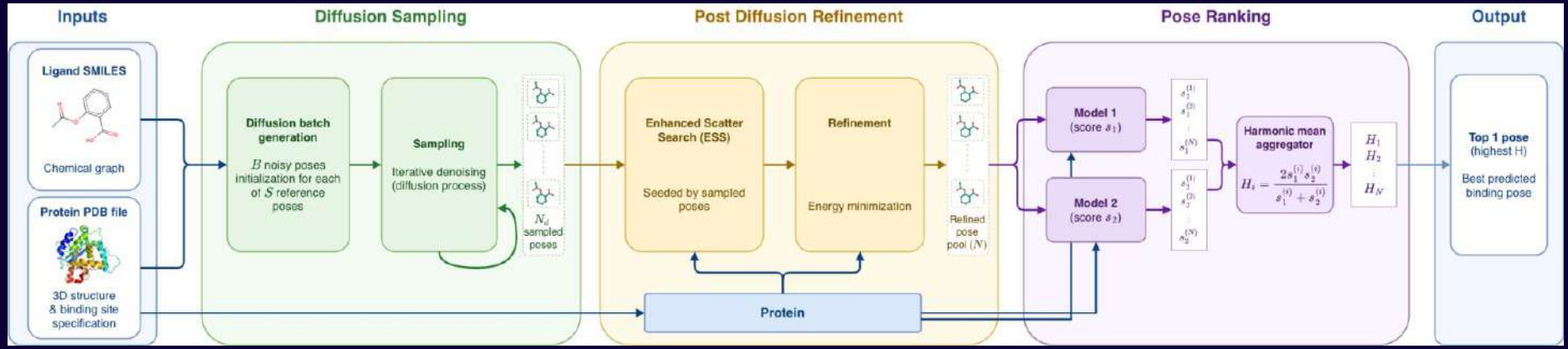
Cofolding methods specifically struggle with docking



- Cofolding methods predict the pocket structure quite accurately
- The right half of this figure is particularly clear:
  - “Self-docking” performance deteriorates sharply as complexes become less similar to the training data

Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

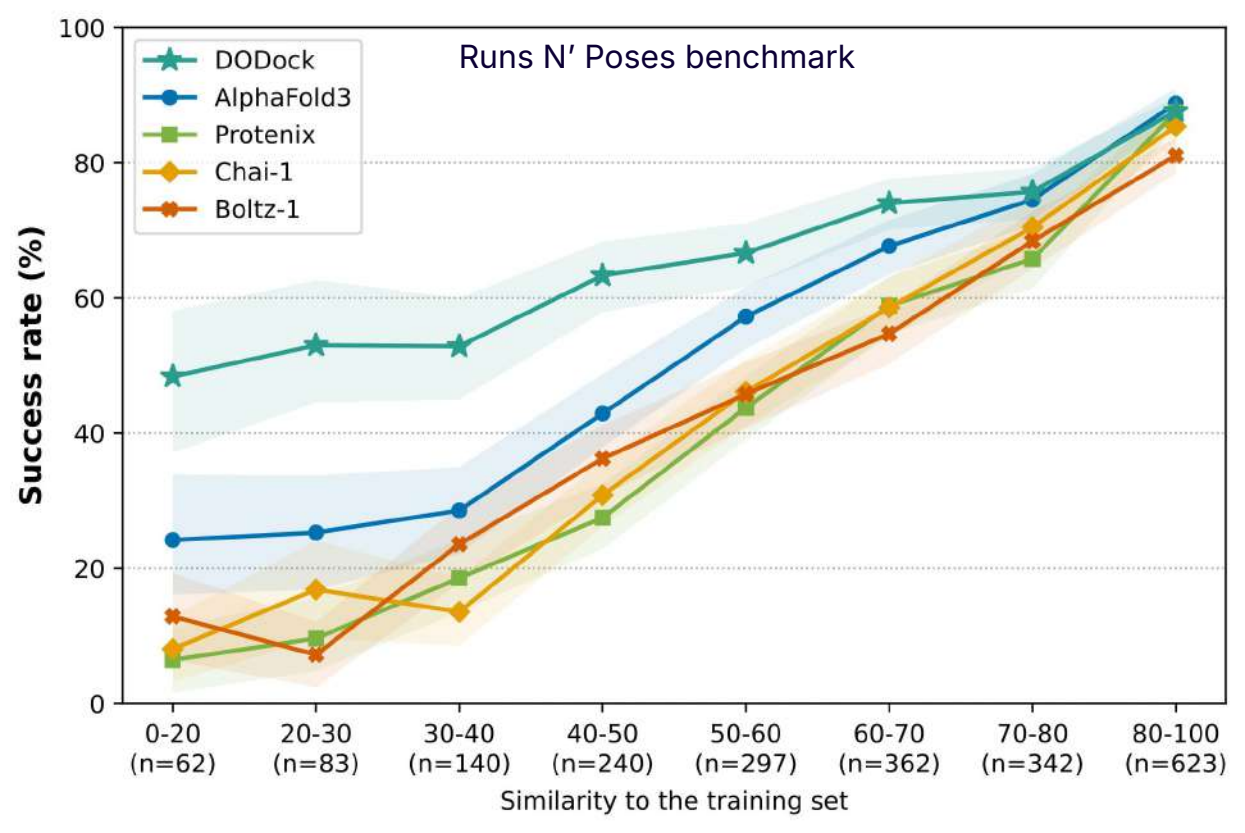
The architecture of the DODock model



- Diffusion model first proposes candidate poses, supplying broad exploration of pose space
- A physics-based search then refines them against an interpretable energy function that does not overfit to training-set similarity
  - A parsimonious model derived by postulating a binding funnel
- A score network built on atomic environment vectors and rigid-body-invariant point-pair features

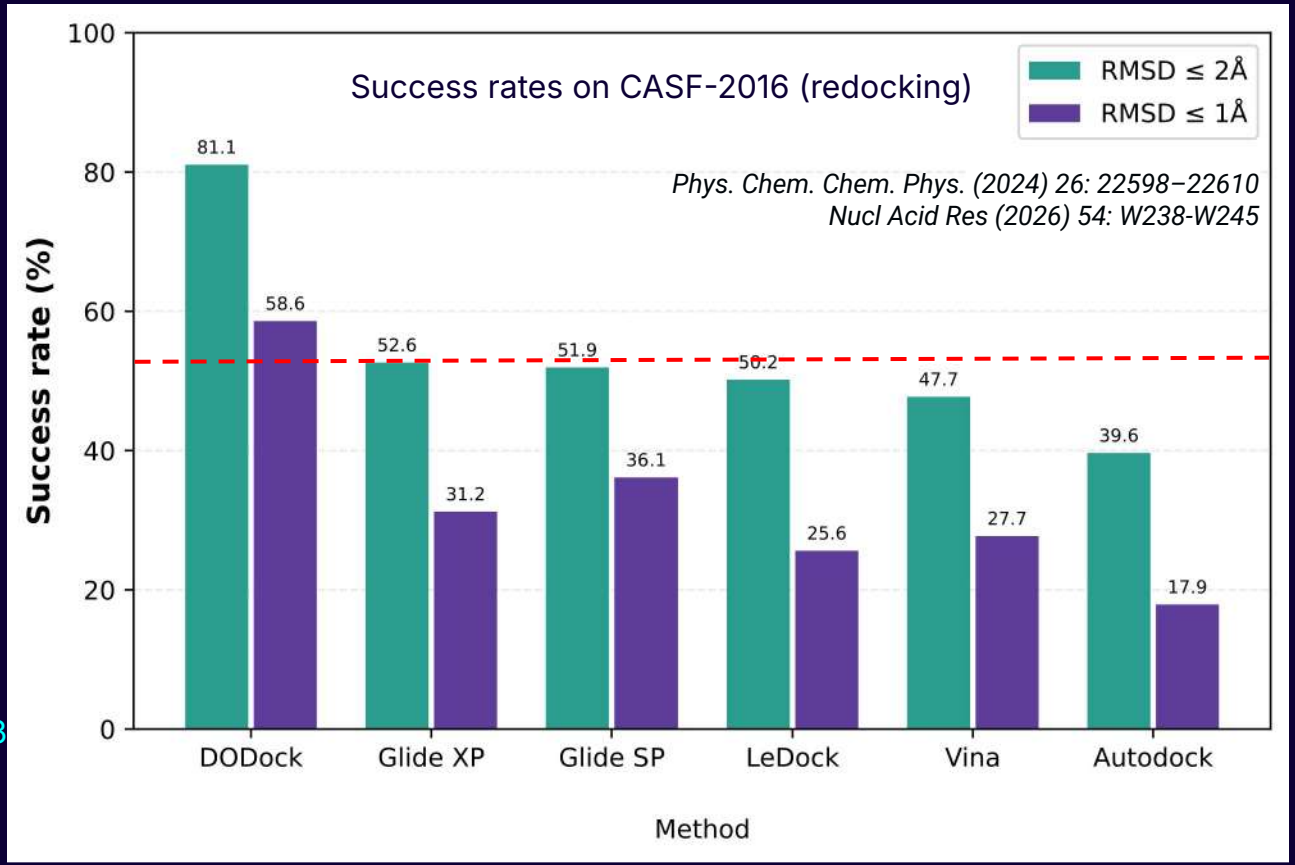
Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

DODocking shows better generalization



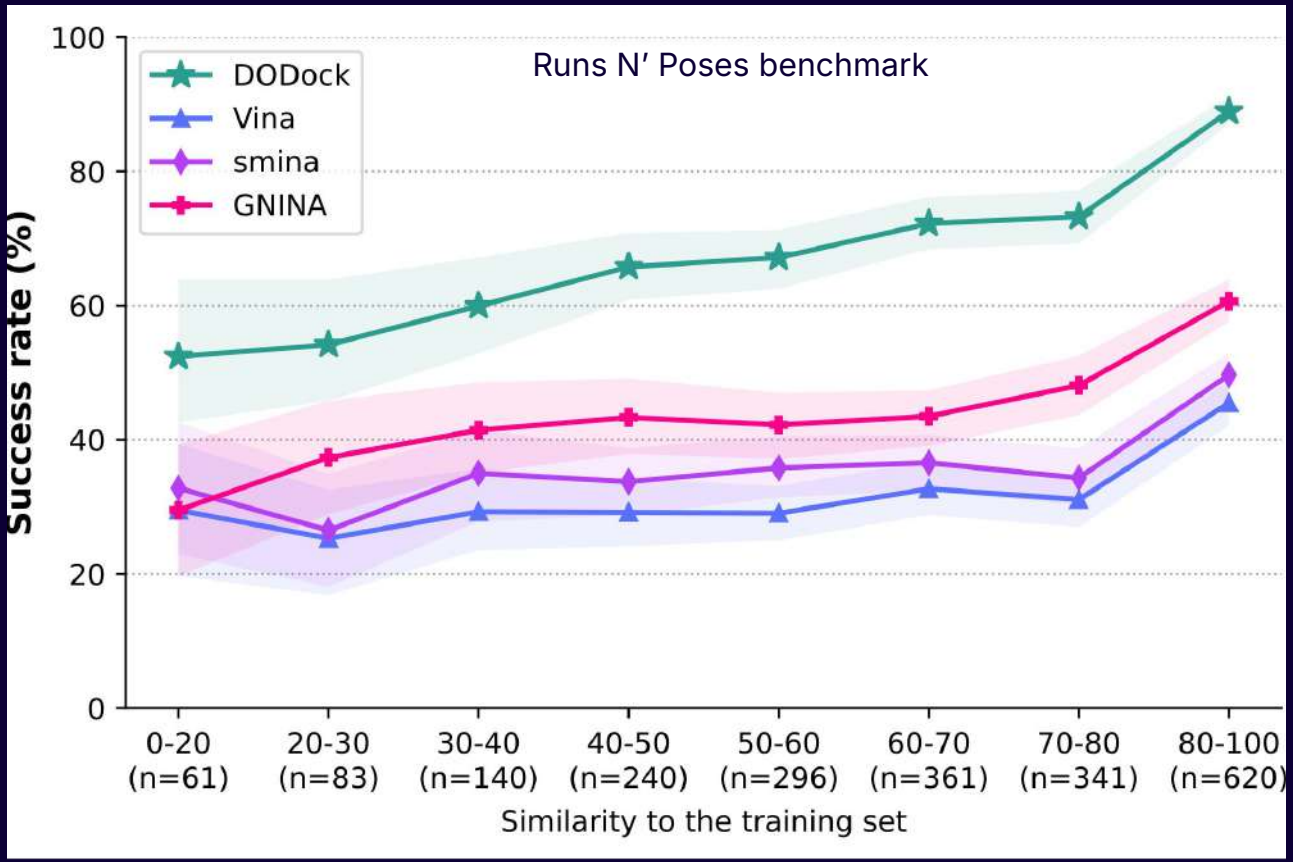
Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

Docking accuracy ceiling can be significantly improved



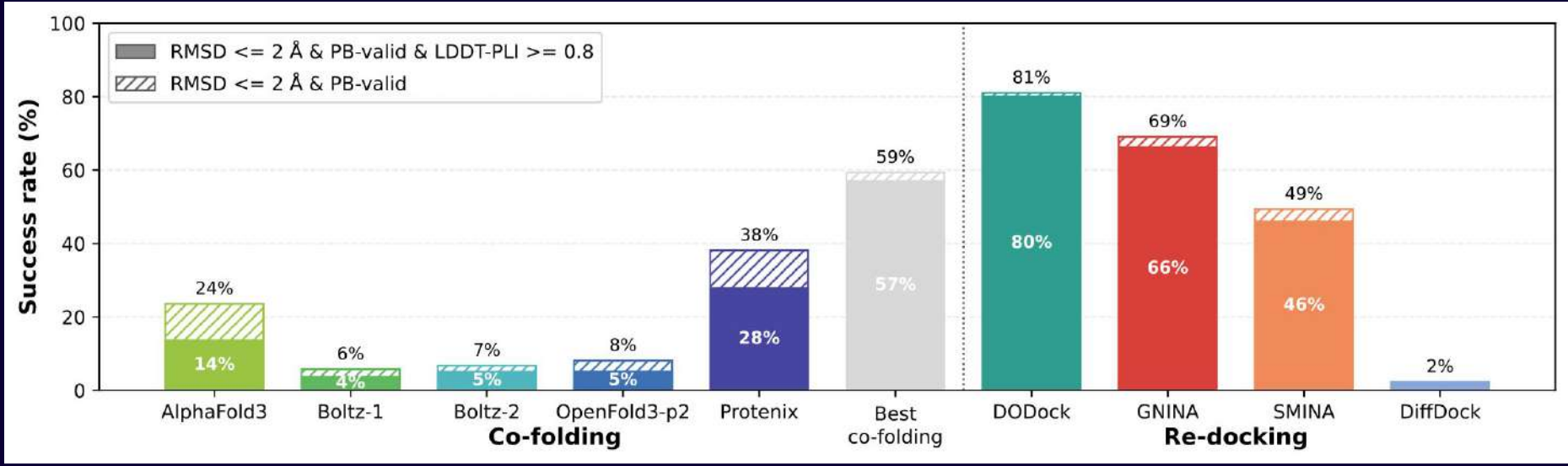
Petrosyan, G. et al.
bioRxiv
https://doi.org/10.64898/2026.08.03.742480
(2026)

Docking accuracy ceiling can be significantly improved



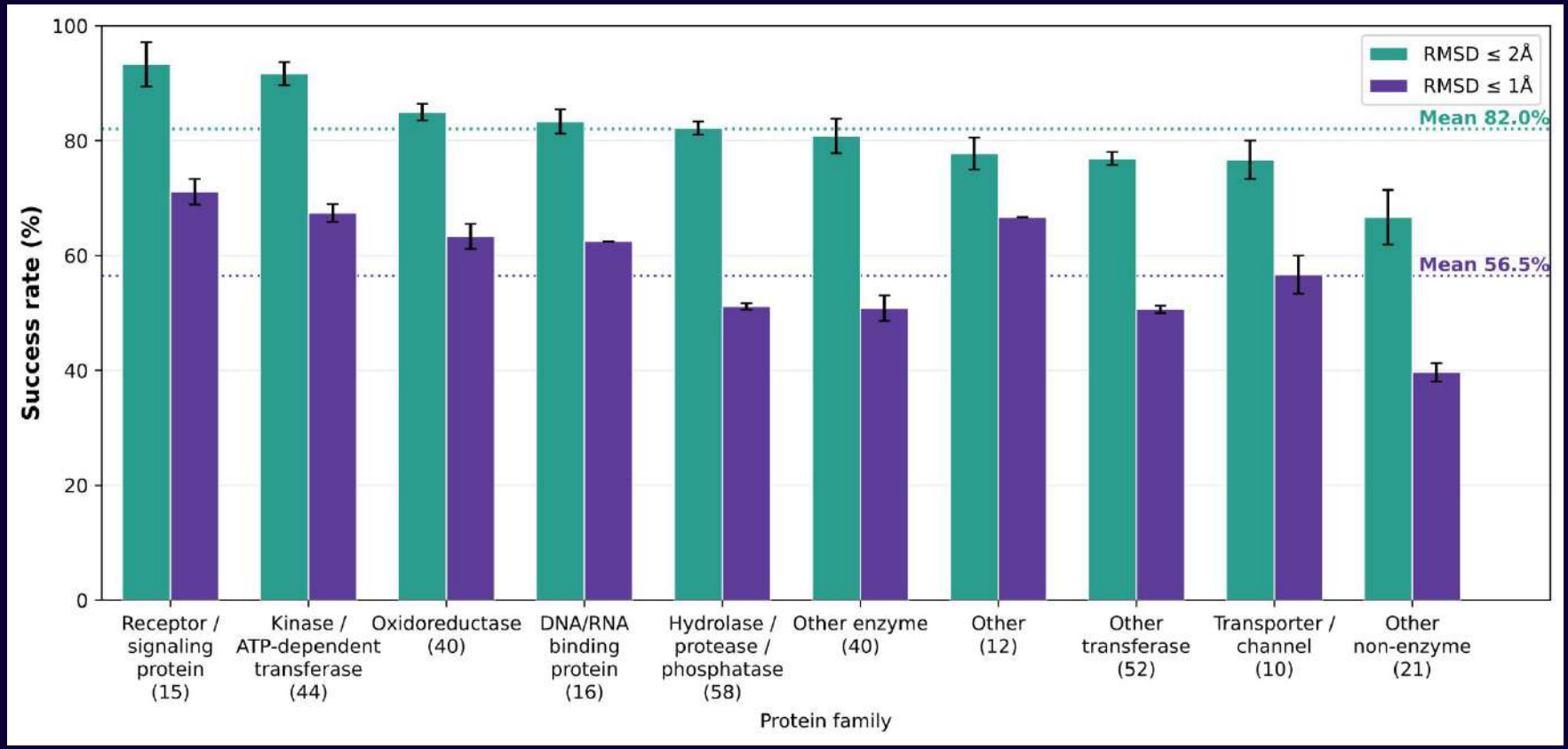
Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

Success rates on the OpenBind benchmark

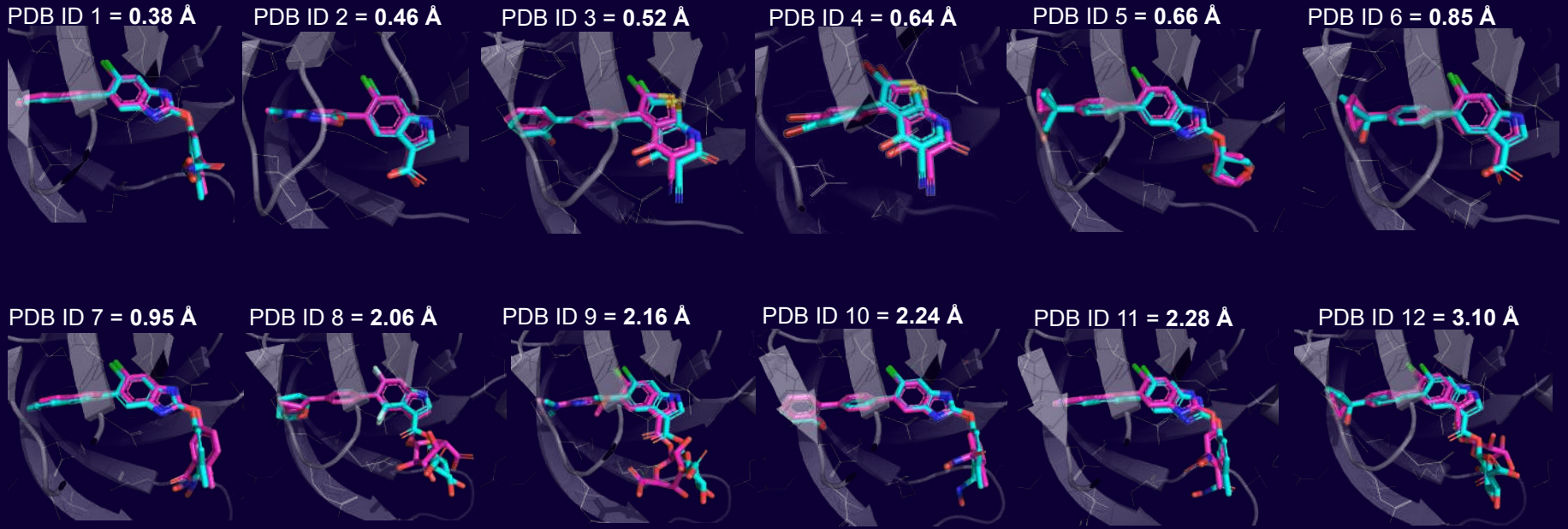


Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

DODock performance on PoseBusters benchmark binned by protein family

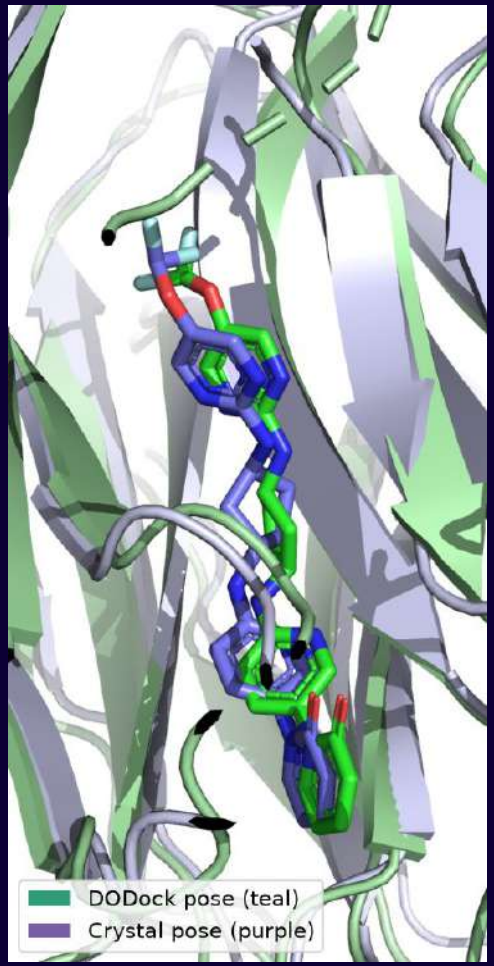


DO Docking reproduces crystal poses for AMPK



RMSD

Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

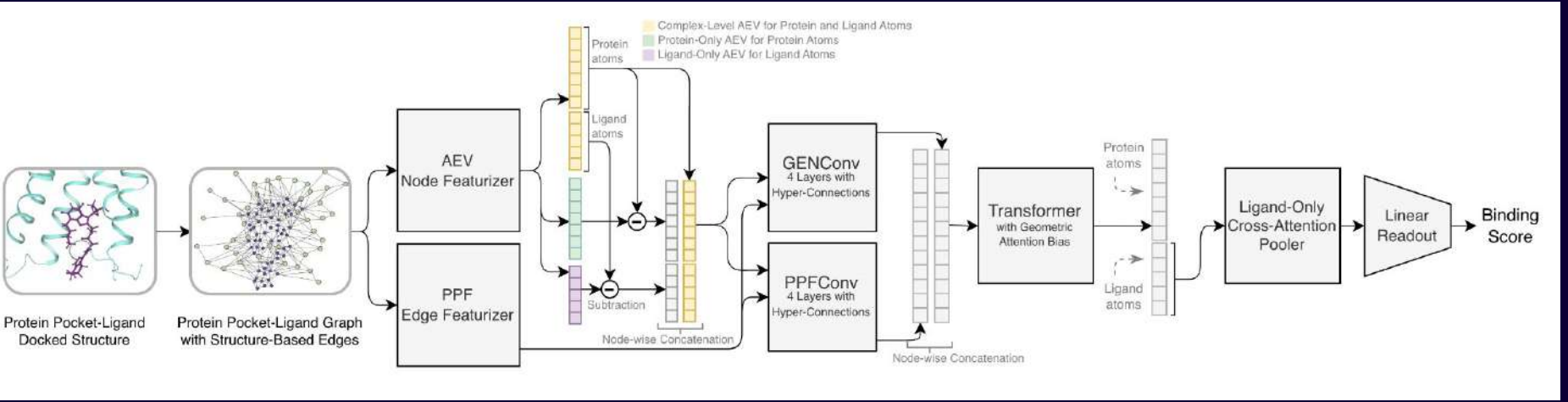


Prospective docking of AZD-0780 to PCSK9

- PCSK9 is an important pharmaceutical target, a central regulator of plasma LDL-cholesterol
- An oral macrocycle inhibitor was recently approved, hailed as a breakhough
- Astrazeneca has developed AZD–0780 as a small molecule inhibitor
- No crystal structure was available a year ago
- We made a docking prediction, then decided to pursue crystal structure determination, which was successful
- **Our docked pose agreed with the crystallographic pose to within 1.2 A RMSD accuracy**

Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

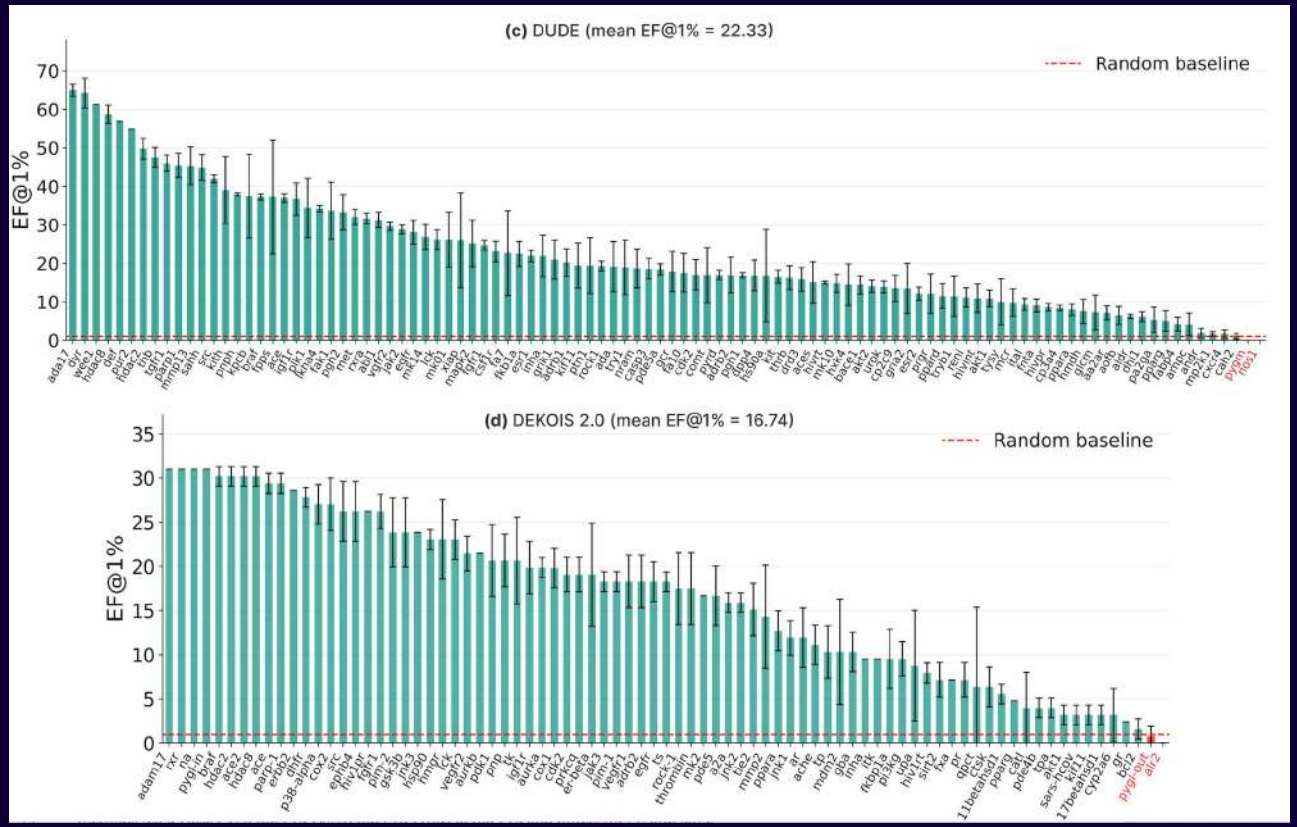
The architecture of DOScore model



- Atomic environments are encoded with atomic environment vectors and processed by graph neural network
- Edges are defined by spatial proximity rather than covalent bonds, capturing non-bonded and long-range interactions
- Hyper-connections to counter over-smoothing

Petrosyan, G. et al. bioRxiv
https://doi.org/10.64898/2026.08.03.742480 (2026)

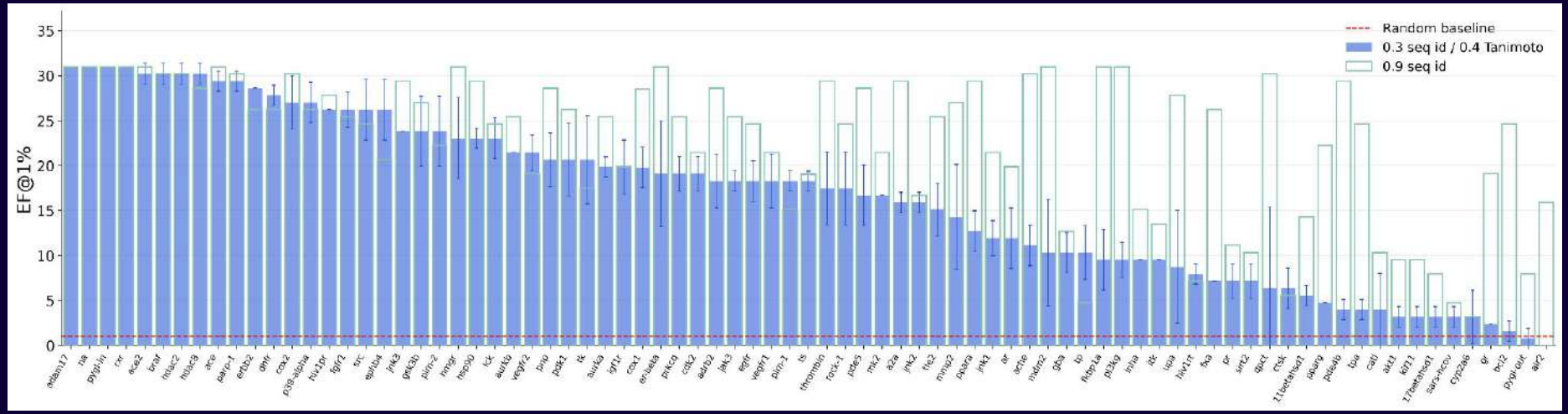
Success rates on the DUDE and DEKOIS 2.0 benchmarks



- EF1% - enrichment factor in top 1% ranked compounds
  - 1.0 - random
- DUDE
  - Mean EF1% ~ 22x
- DEKOIS 2.0
  - Mean EF1% ~ 17x

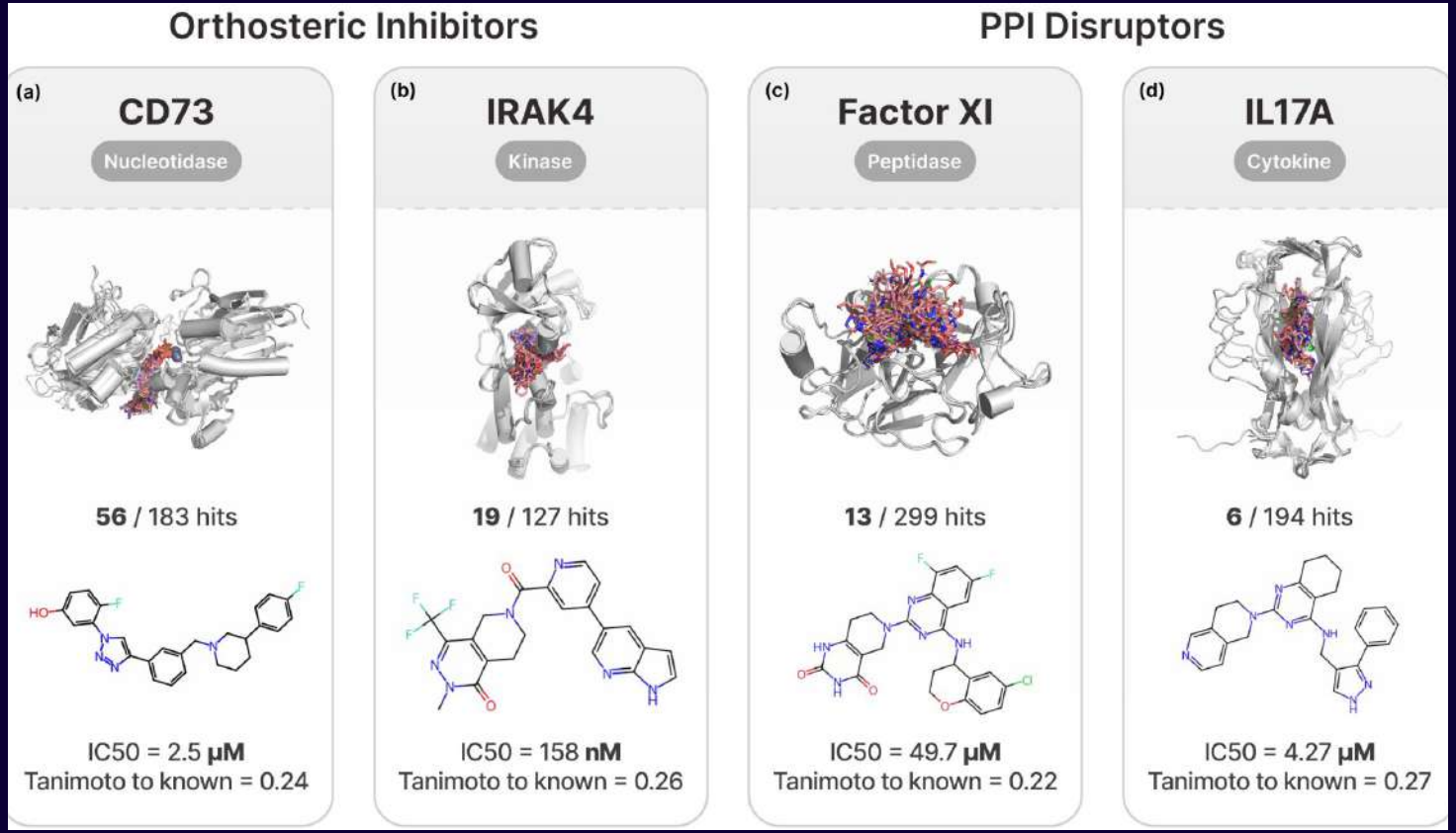
Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

Training data similarity significantly boosts benchmarks

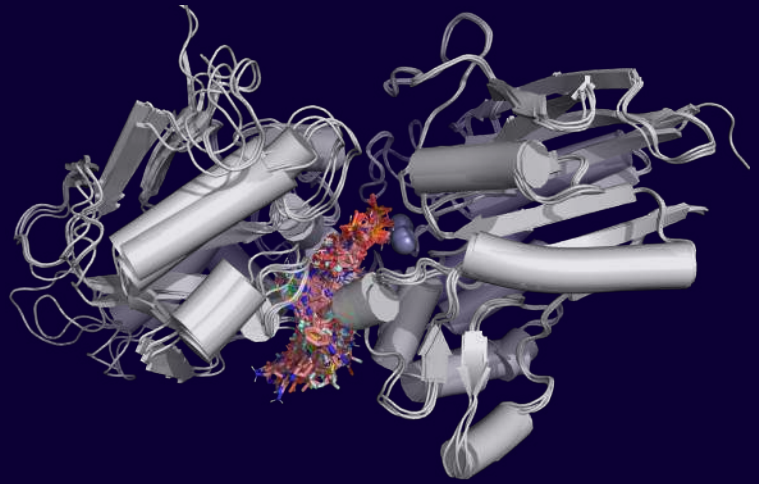


Petrosyan, G. et al. bioRxiv https://doi.org/10.64898/2026.08.03.742480 (2026)

Prospective virtual screening identifies novel inhibitors across diverse target classes

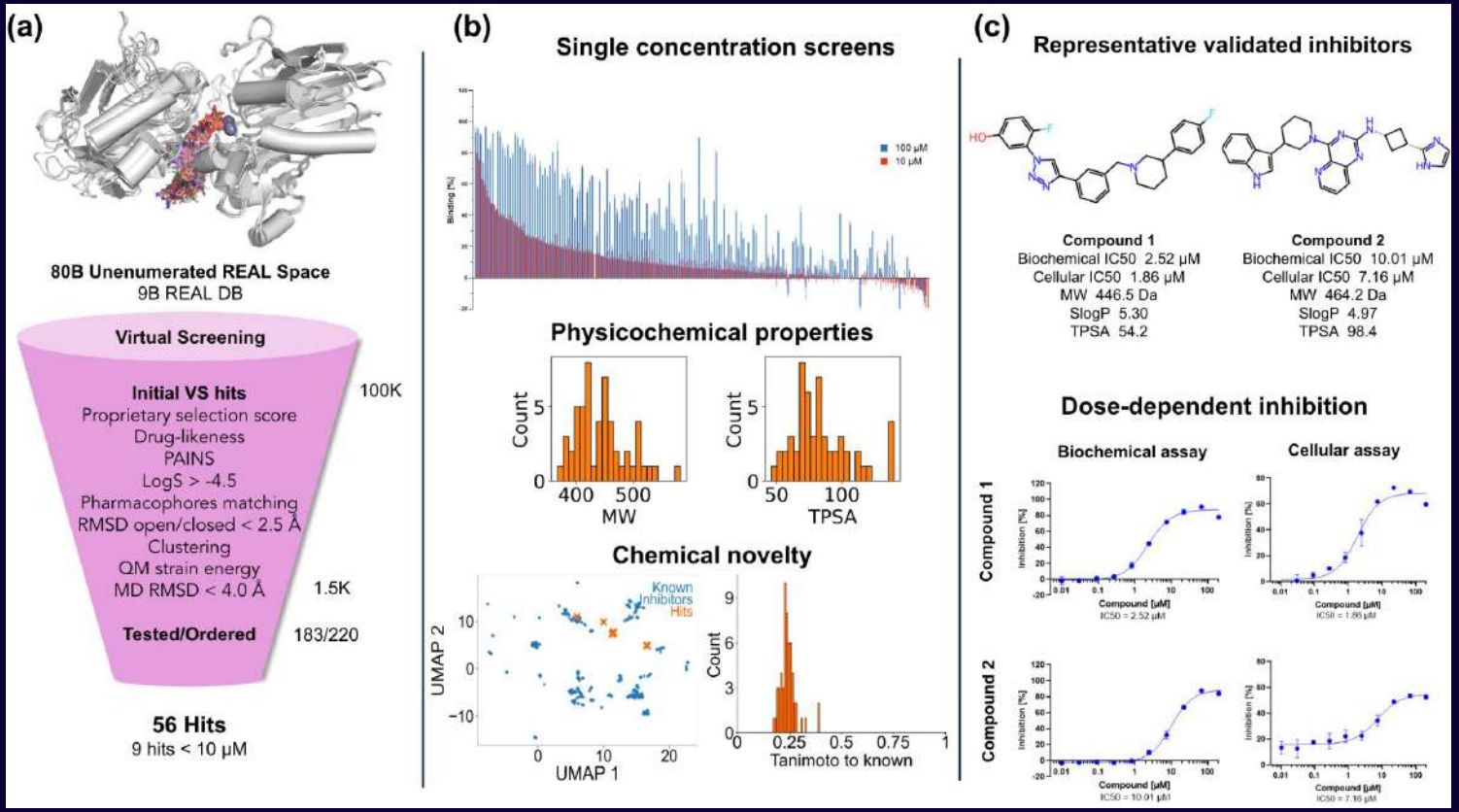


High hit rates and broad chemotype diversity achieved in prospective discovery on CD73



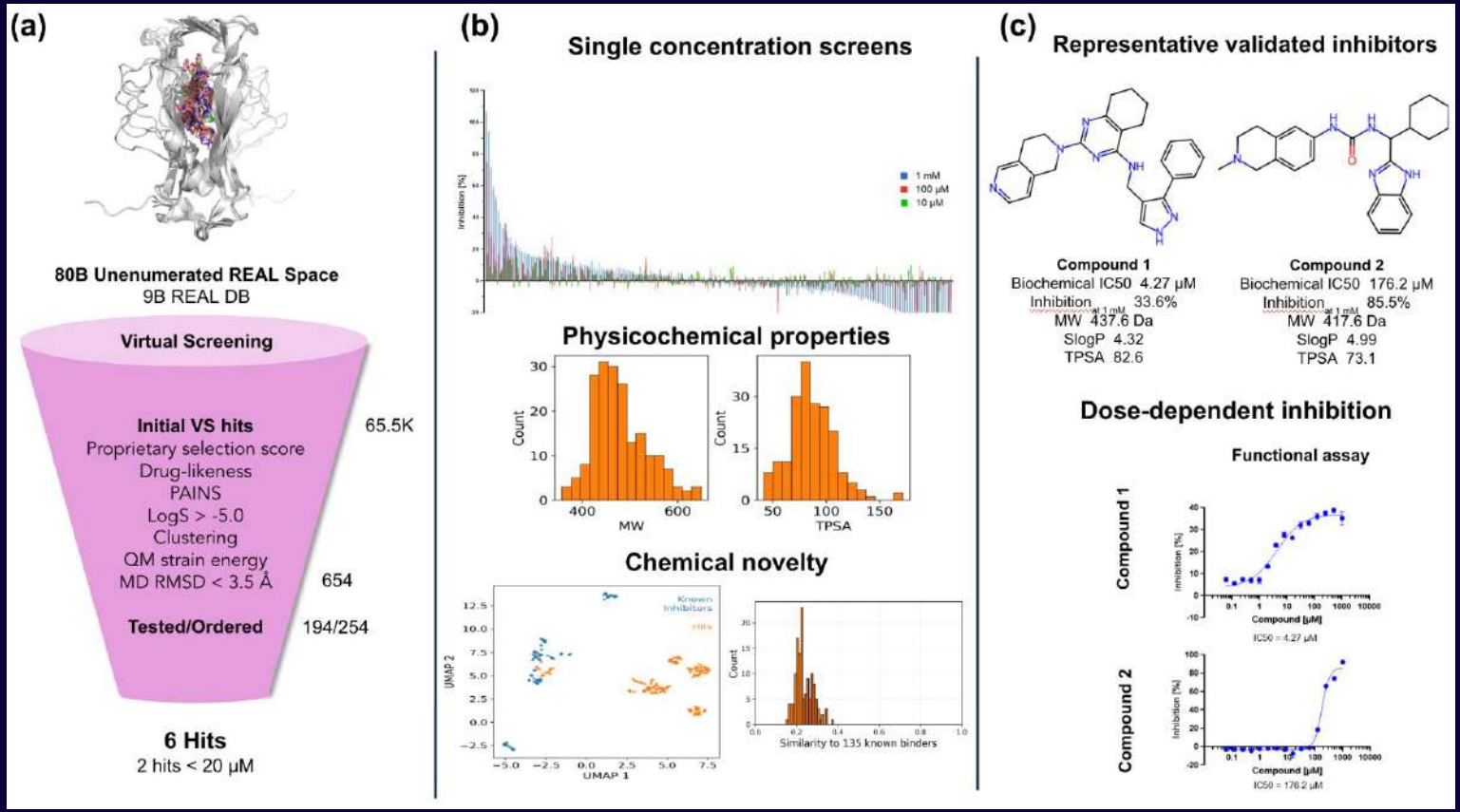
- CD73 (5'-nucleotidase NT5E) converts AMP to adenosine modulating immune suppression and tumor metabolism.
- Elevated soluble CD73 levels correlate with poor prognosis across multiple tumor types.
- Most CD73 inhibitors are stable nucleotide analogs, with limited selectivity and potency.
- Non-nucleotide inhibitor scaffolds are very promising, but prior campaigns have been largely unsuccessful.
- **We have screened CD73 against ~9B and ~71B ultra-large libraries as part of our internal validation project.**

Prospective virtual screening of novel CD73 inhibitors

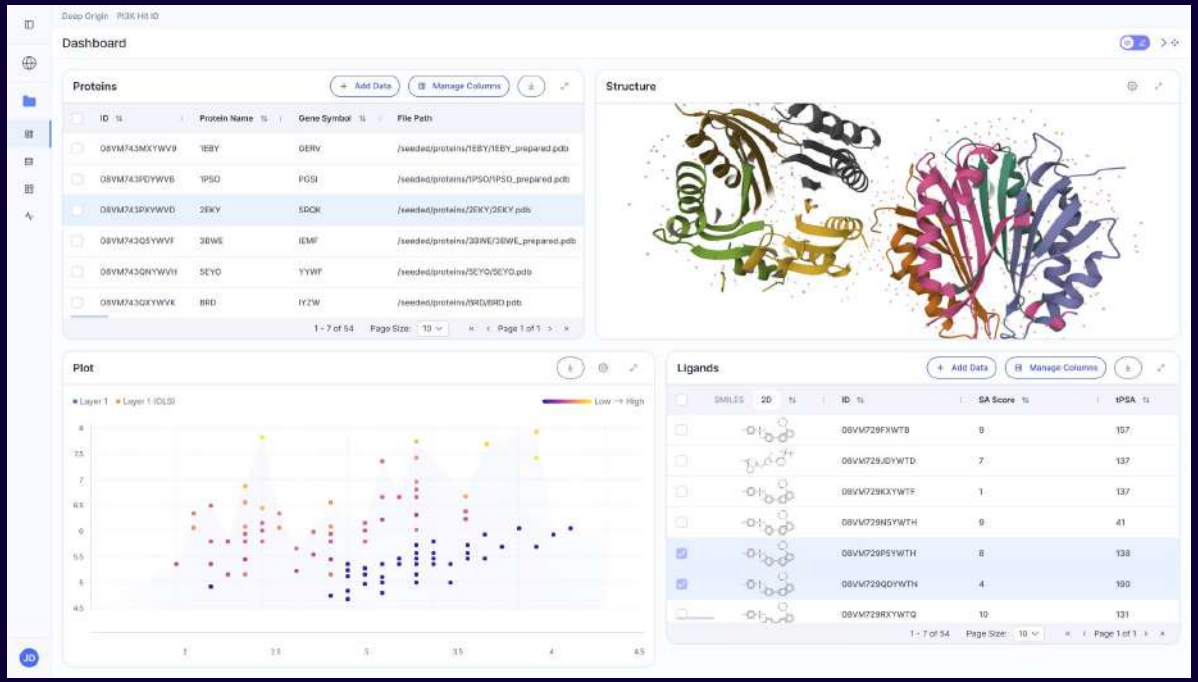


- ~30% hit rate
- Versus 0.3% hit rate in a previous screening (Atomwise AIMS Program)
- Diverse hits dissimilar to previously known binders

IL17: Discovering allosteric PPI disruptors



DO Studio - The new interface for Drug Discovery

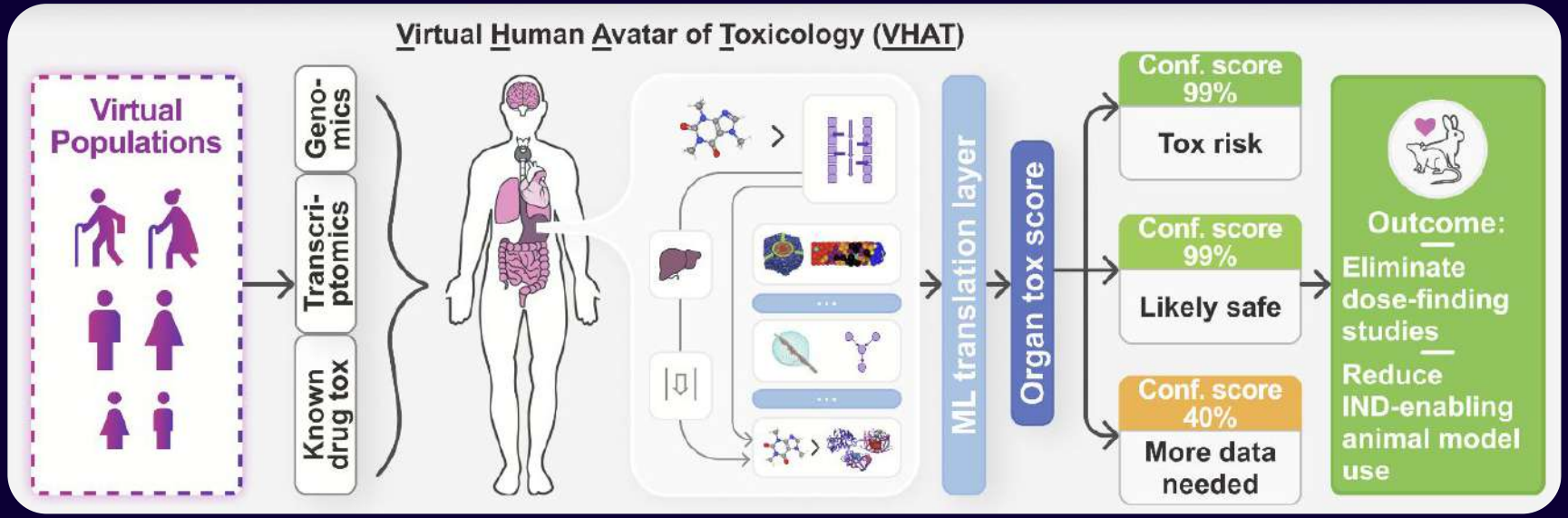


DO Studio Dashboard

Integrated data management, cloud compute, and molecular modeling tools:
* Pocket Finder
* Docking
* FEP
* QSAR & Enumeration
* Molecular Properties
* Custom Project-aware dashboards
* Streamlined analysis components

All accessed via the web!

Funded with >$30 M from ARPA-H to build a virtual human



We unite expertise across bio and tech

We focus on technology, so our partners can make therapies faster.

We have ~130 employees across the globe with expertise in physics, chemistry, biology, tech, and drug development.

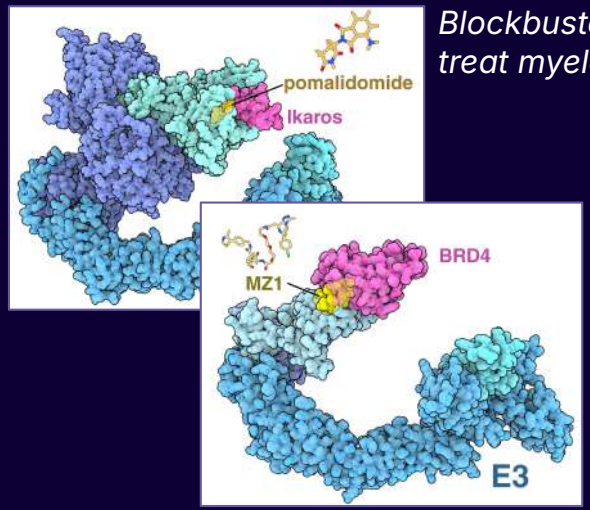
DeepOrigin

DO Molecular Glue Technology

Enormous therapeutic potential

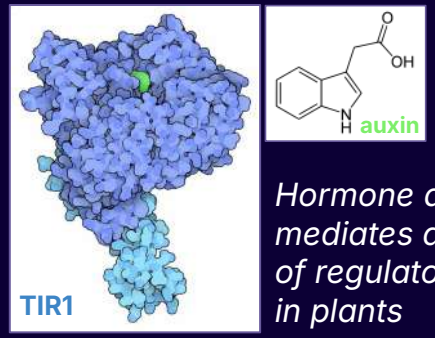
Chemical space differentiated by small molecule structure and mechanism of action

Molecular glue degraders & PROTACs



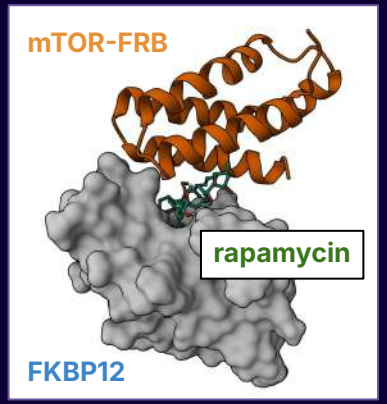
Blockbuster IMiDs treat myeloma

Target POI for proteasomal degradation via ubiquitination through interaction with an E3 ligase



Hormone auxin mediates degradation of regulatory proteins in plants

Molecular glue stabilizers



Immunosuppressant rapamycin inhibits mTOR signalling

RIPTACs, AUTACs, KineTACs

Facilitate interaction in a specific niche:

- *engage tumor-selective POI to inhibit pan-essential effector*
- *target POI for autophagy degradation*

Design of new proximity therapeutics is difficult

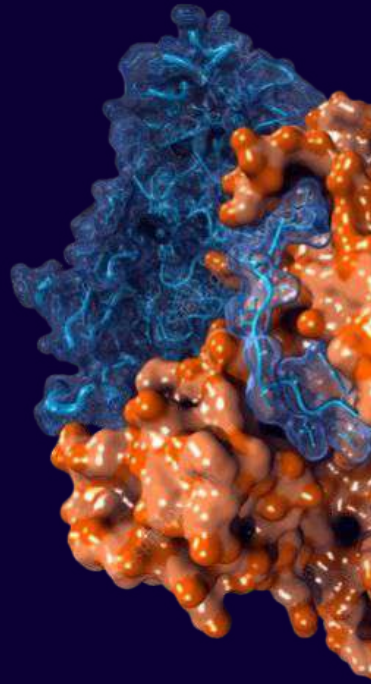
Most molecular glues today share a common scaffold - limiting potential

The majority of ligands are structural analogs of a handful of scaffolds discovered by serendipity or found in nature

Only target/effector proteins known to interact with existing scaffolds are accessible

Shared scaffolds result in IP constraints and shared resistance liability across programs

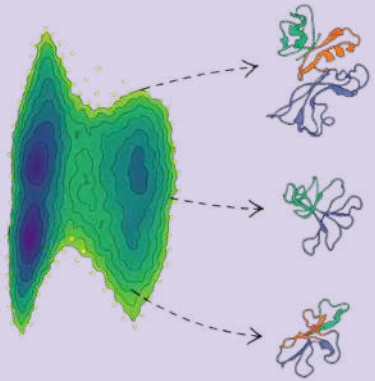
Applicability of AI and rigid docking methods tends to break down in the novel target and chemical space



We need an approach that can discover new molecules beyond the known scaffolds

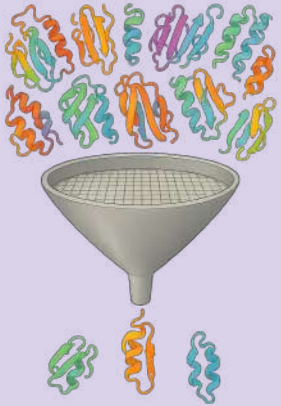
DO molecular glue pipeline

E3 ligase –Target Interaction Metastable States



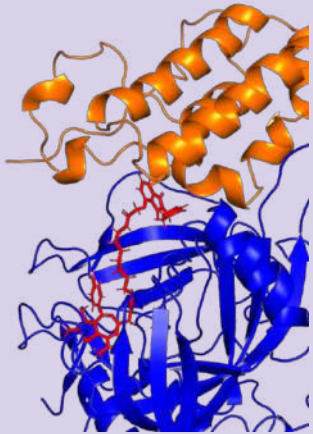
- Run binding simulations and find metastable states

Conformation selection



- Filter and cluster to select useful structures for design

PROTAC/Mol glue docking



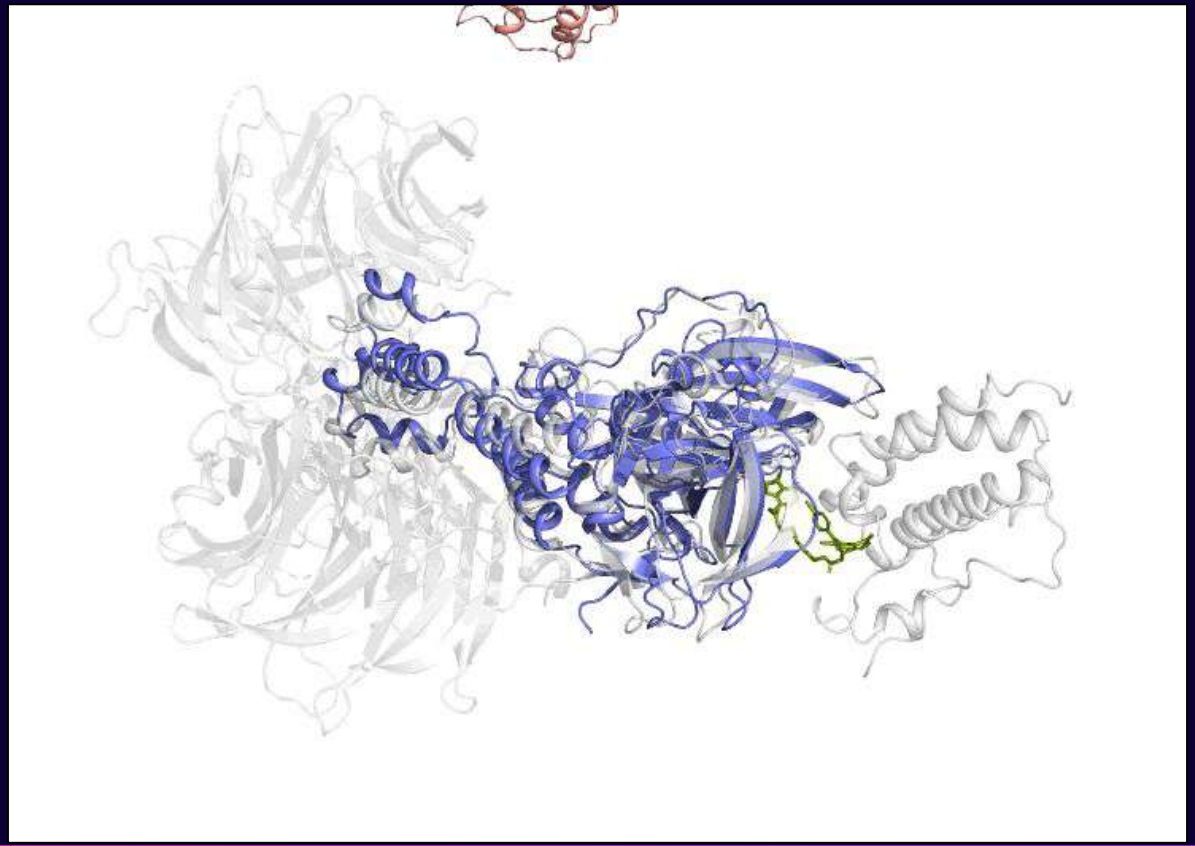
- Run short all atom MD
- Dock mol glues

PROTAC/Mol glue scoring

- Score each complex based on stability and probability of ubiquitination

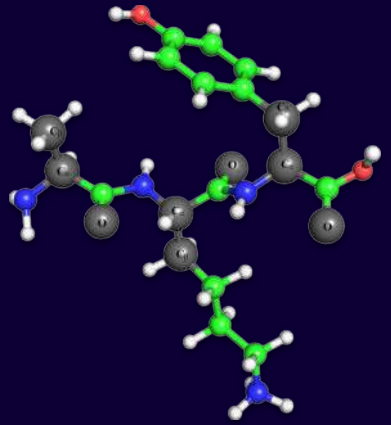
DO-AWSEM binding simulation of an E3 ligase and target protein

CRBN:BRD4^{BD1}

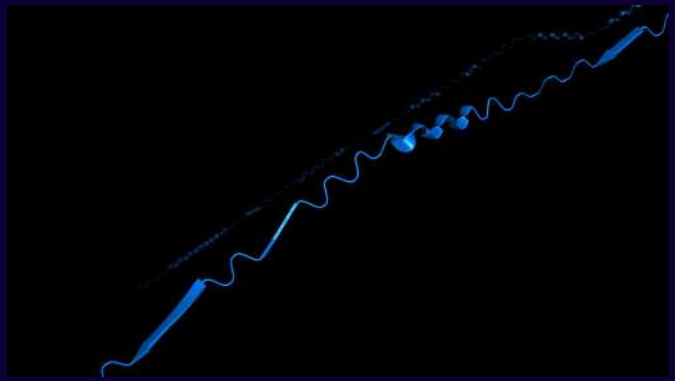


Ligand dBET23 not included in the simulation

AWSEM – Associative Memory, Water Mediated, Structure and Energy Model



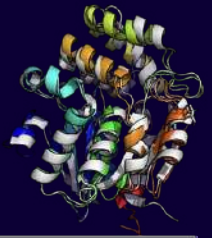
- Developed in 2012 by Papoian and Wolynes groups based on Energy Landscape Theory.
- Each amino-acid is represented by three beads placed at the positions of C_α, C_β, O
- Simple representation combined with physics-based differentiable energy function and ML parametrization enables efficient and accurate modeling of protein dynamics



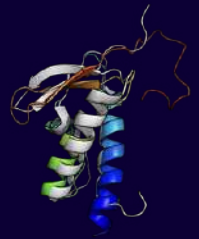
Human PIF1 Helicase (PDB: 6hph)

✏ Davtyan, et al. **J. Chem Phys B**, 116, (**2012**), 1709–1715

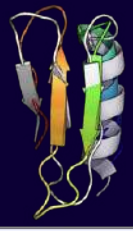
Accurate structure and binding prediction using AWSEM framework was previously demonstrated



T0951(PDB: 5Z82)
RMSD: 2.5 Å

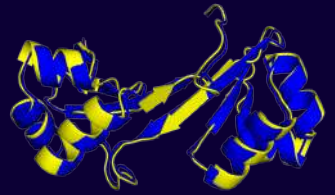


T0958(PDB: 6BTC)
RMSD: 2.7 Å



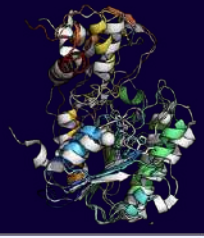
T0955(PDB: 5W9F)
RMSD: 3.0 Å

Arc repressor
1ARR N=106

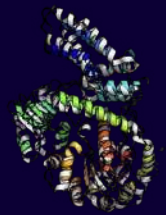


Lambda Cro repressor
1COP N=132

Factor for inversion stimulation
1F36 N=178

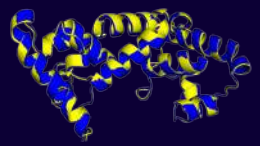


T0965(PDB: 6D2V)
RMSD: 3.1 Å

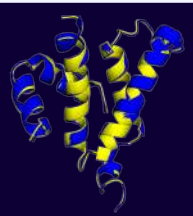


T0966(PDB: 5W6L)
RMSD: 3.4 Å

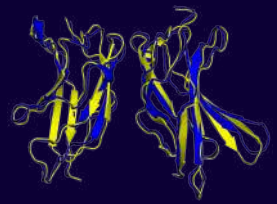
T0971(PDB: 6D34)
RMSD: 1.8 Å



Lambda repressor
1LMB N=179



KIX-PKID
1KDX N=109



NFkB P50/P65
1VKX N=207

Predicted
Crystal;

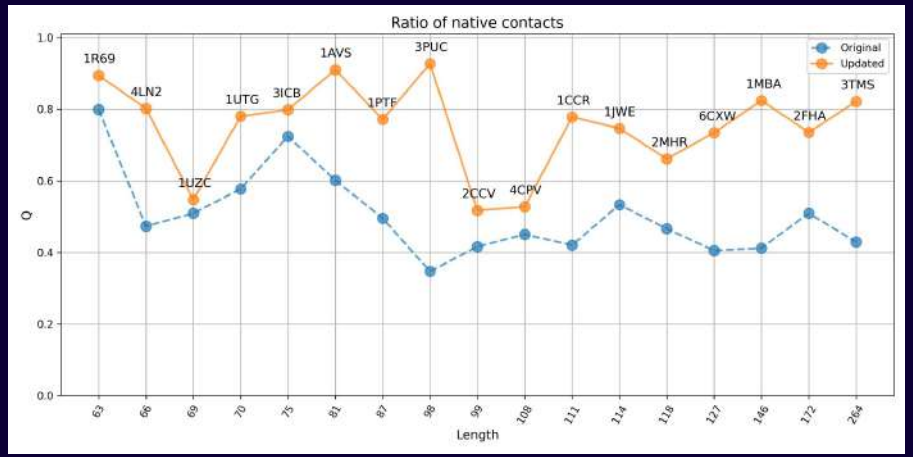
✏ Jin, Chen, Chen, Bueno, Lu, Schaefer, Lin, Onuchic, Wolynes, **J Chem Theory Comput**, 16, (**2020**), 3977-3988

Predicted
Crystal structure

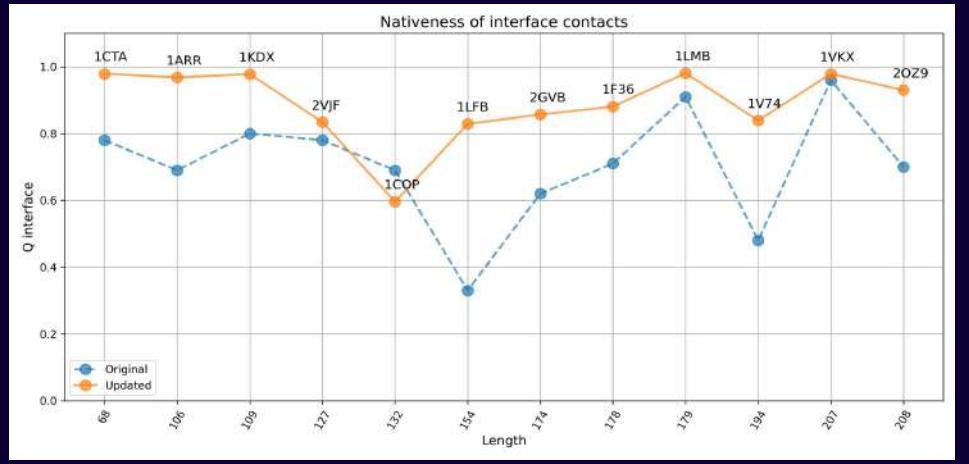
✏ Zheng, Schafer, Davtyan, Papoian, Wolynes, **PNAS**, 109, (**2012**), 19244-19249

DO-AWSEM shows significantly better results both for folding and binding compared with prior models

Folding



Binding

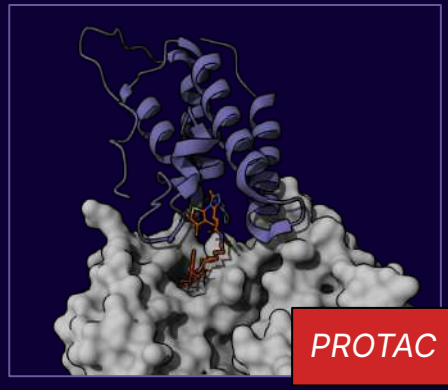


Benchmarking on well studied systems

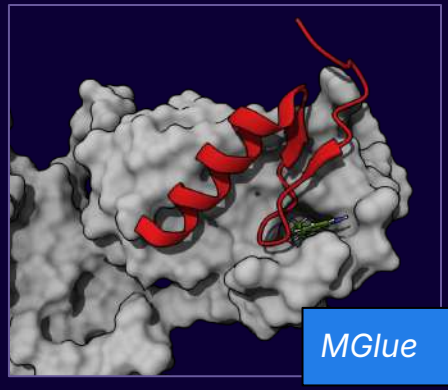
Well studied, medium sized baseline systems with known crystal structure reference

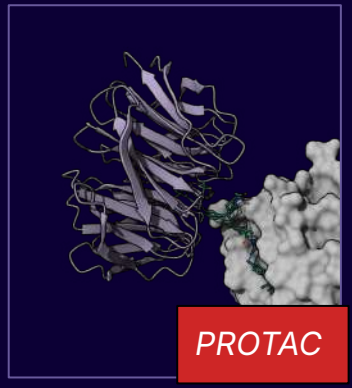
CRBN:BRD4

conformationally stable PPI and complex



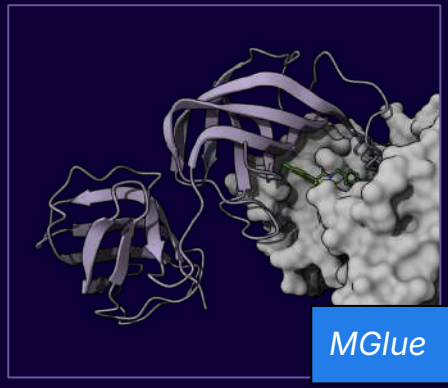
CRBN:IKZF1 “Ikaros”
— the classic IMiD target





VHL:WDR5

transient PPI and variety of complex structures

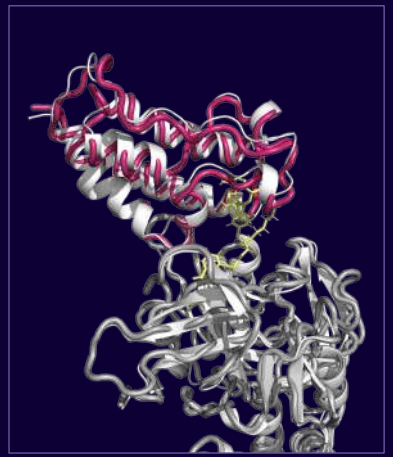


CRBN:GSPT1
— IMiD and non-IMiD space

Sampling recovers known protein-protein interfaces

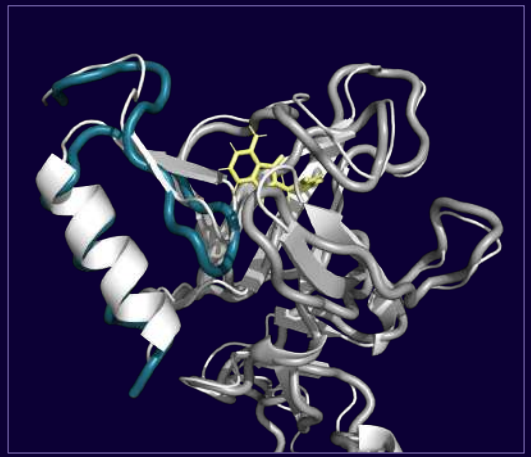
CRBN:BRD4

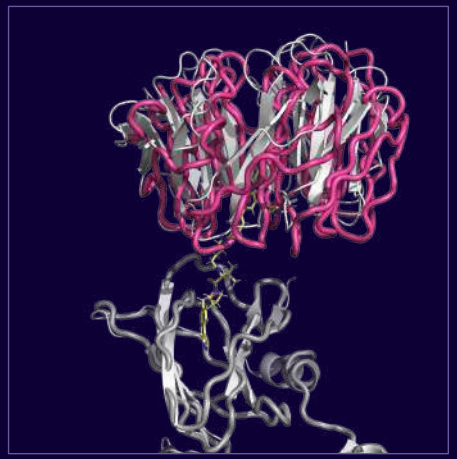
DockQ = 0.73
RMSD = 1.4 Å



CRBN:IKZF1

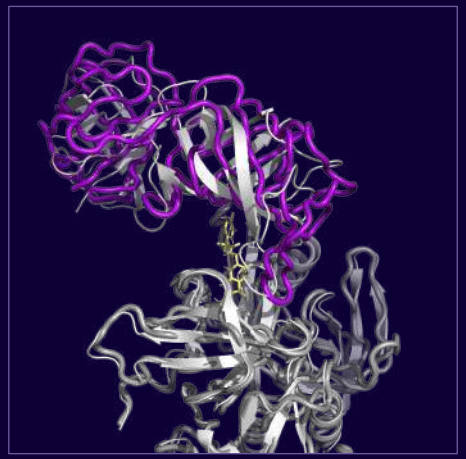
DockQ = 0.83
RMSD = 1.9 Å





VHL:WDR5

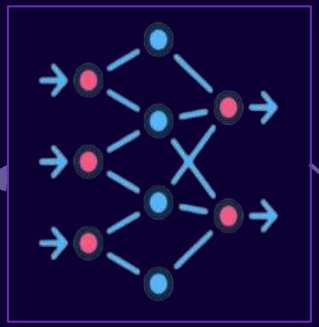
DockQ = 0.93
RMSD = 3.3 Å



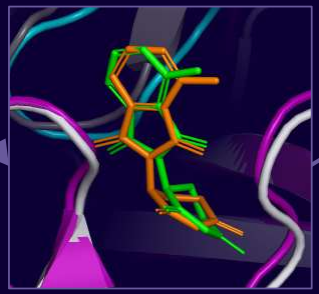
CRBN:GSPT1

DockQ = 0.66
RMSD = 3.6 Å

Our scoring, filtering and ligand docking approach allows for reconstruction of known ternary complexes



ML scoring

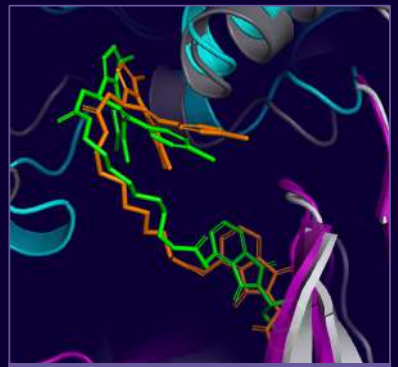


PROTAC / MG docking

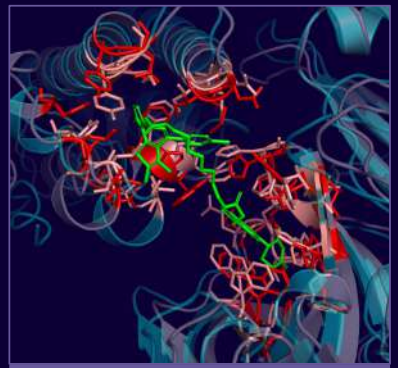
Aggregated score

Top 10 conformations by aggregated score

CRBN:BRD4

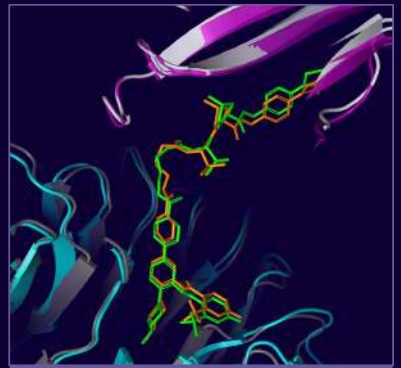


Ligand RMSD **2.26 Å**

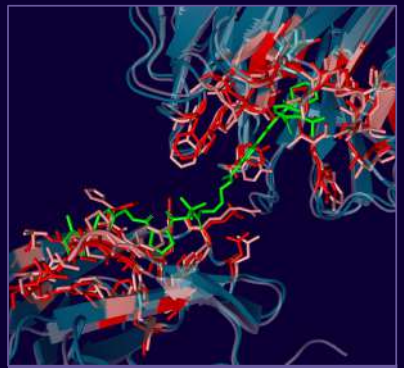


Pocket RMSD **2 Å**

VHL:WDR5



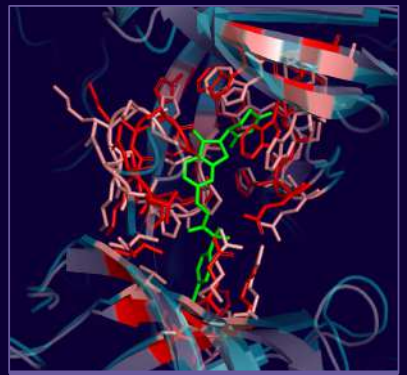
Ligand RMSD **0.83 Å**



Pocket RMSD **0.8 Å**

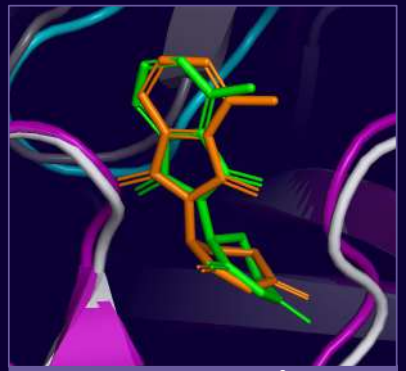
CRBN:GSPT1

Ligand RMSD **2.6 Å**

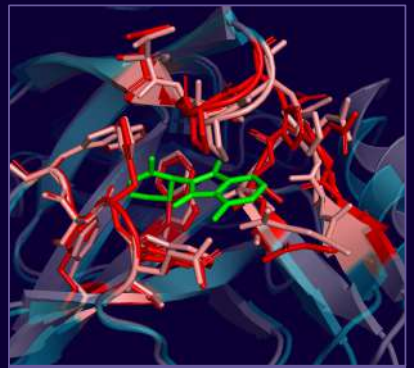


Pocket RMSD **1.65 Å**

CRBN:IKZF1



Ligand RMSD **1 Å**

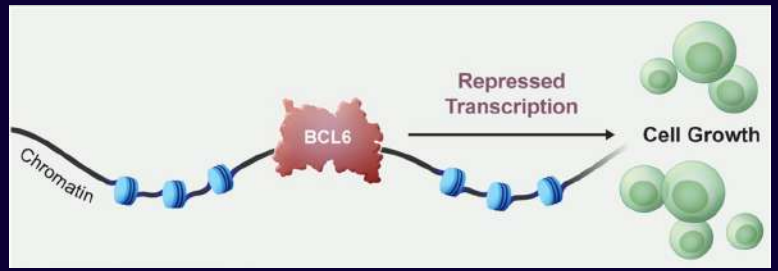


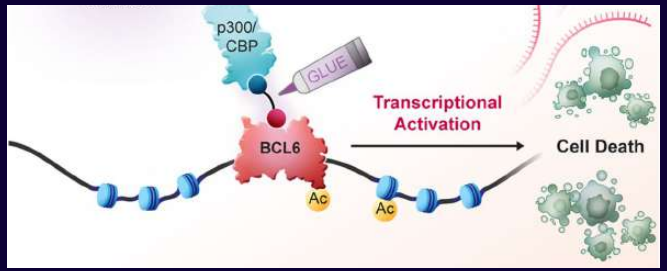
Pocket RMSD **0.8 Å**

Application — turning a foe into a friend

A bivalent molecular glue linking lysine acetyltransferases to oncogene-induced cell death
Gray Lab (Stanford), UT MD Anderson Cancer Center, Deep Origin, Cell, July 20 2026

- Diffuse large B-cell lymphoma (DLBCL) is most common non-Hodgkin lymphoma, which is driven by BCL6 deregulation in 40-60% of cases
- BCL6 is a transcriptional repressor that binds to DNA and suppresses the expression of genes involved in apoptosis and cell-cycle arrest, helping malignant B cells survive and multiply
- p300/CBP are lysine acetyltransferases (KATs) that switch genes on through histone acetylation — the same activity that sustains oncogenic enhancers in DLBCL
- Bringing p300/CBP into close proximity to BCL6 flips a repressor into an activator — and kills the lymphoma cell.

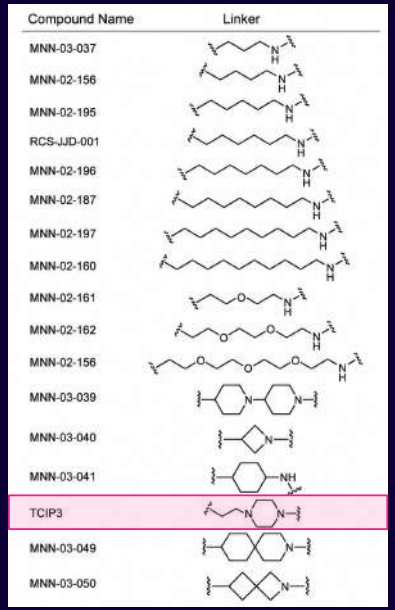


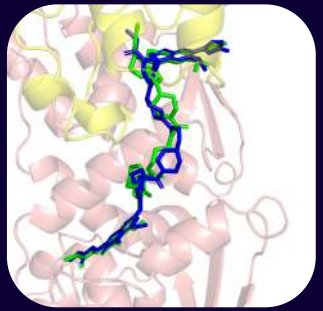


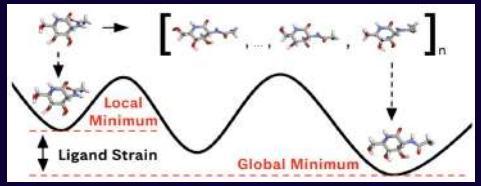
Our *in silico* pipeline helped to identify TCIP3 — a sub-nanomolar glue that clears B-cell lymphoma tumors in vivo

A bivalent molecular glue linking lysine acetyltransferases to oncogene-induced cell death

Gray Lab (Stanford), UT MD Anderson Cancer Center, Deep Origin, Cell, July 20 2026

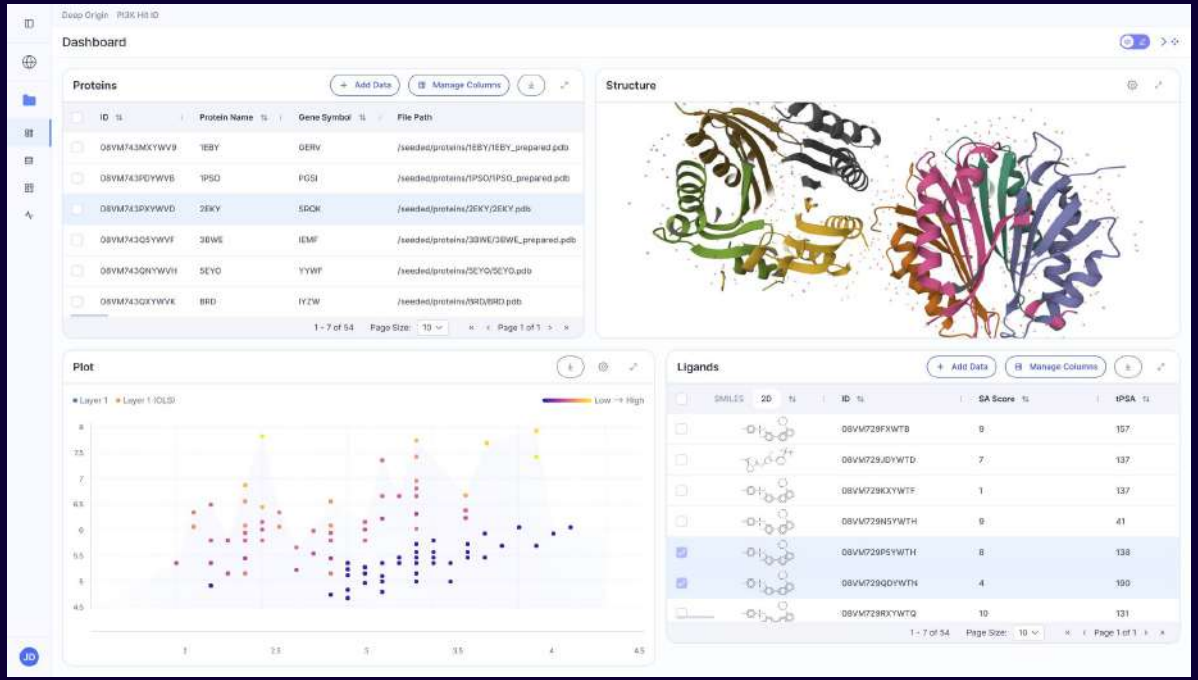






TCIP3 outperformed MNN-02-155 in the K422 cells, inhibited SUDHL5 proliferation with an IC_{50} of 0.80 nM, and drove near-complete tumor clearance by day 11 in CDX mice

DO Studio - The new interface for Drug Discovery

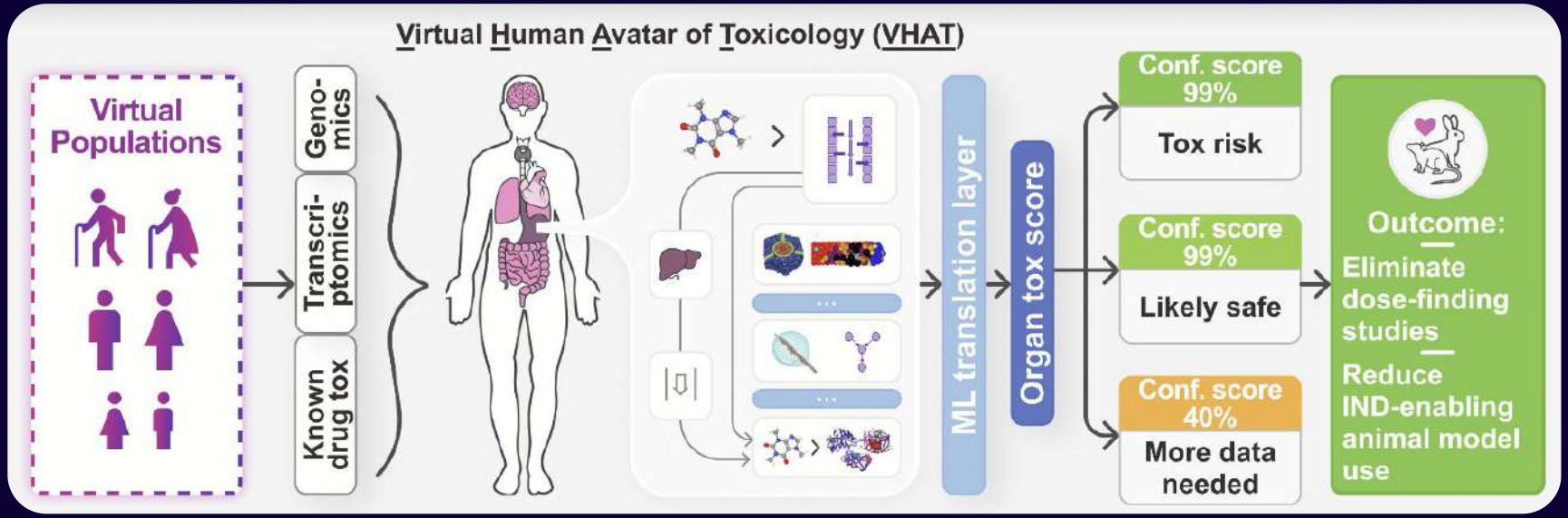


DO Studio Dashboard

Integrated data management, cloud compute, and molecular modeling tools:
* Pocket Finder
* Docking
* FEP
* QSAR & Enumeration
* Molecular Properties
* Custom Project-aware dashboards
* Streamlined analysis components

All accessed via the web!

Funded with >$30 M from ARPA-H to build a virtual human



We unite expertise across bio and tech

We focus on technology, so our partners can make therapies faster.

We have ~130 employees across the globe with expertise in physics, chemistry, biology, tech, and drug development.

Michael Antonov
Co-Founder and CEO

Garegin Papoian, PhD
Co-Founder and CSO

Matt Shlosberg
COO

Ashot Papoyan, PhD
COO-Scientific Development

Henk Campher
Head of Revenue

Andrew Mochalskyy
CTO

Natalie Ma, PhD
CBO

Tigran Abramyan, PhD
Partnerships

Grigor Arakelov, PhD
Partnerships

Elin Barnes
Program Management

Garik Petrosyan, PhD

Oliver Purcell, PhD
Cellular Simulations

Max Ratnikov, PhD
Product

Hayk Saribekyan, PhD
Scientific Development

Aram Davtyan, PhD

oculus

AbCellera

UNIVERSITY OF MARYLAND

Atomwise

strateos

NOVARTIS

+O+US

Google

TetraGenetics

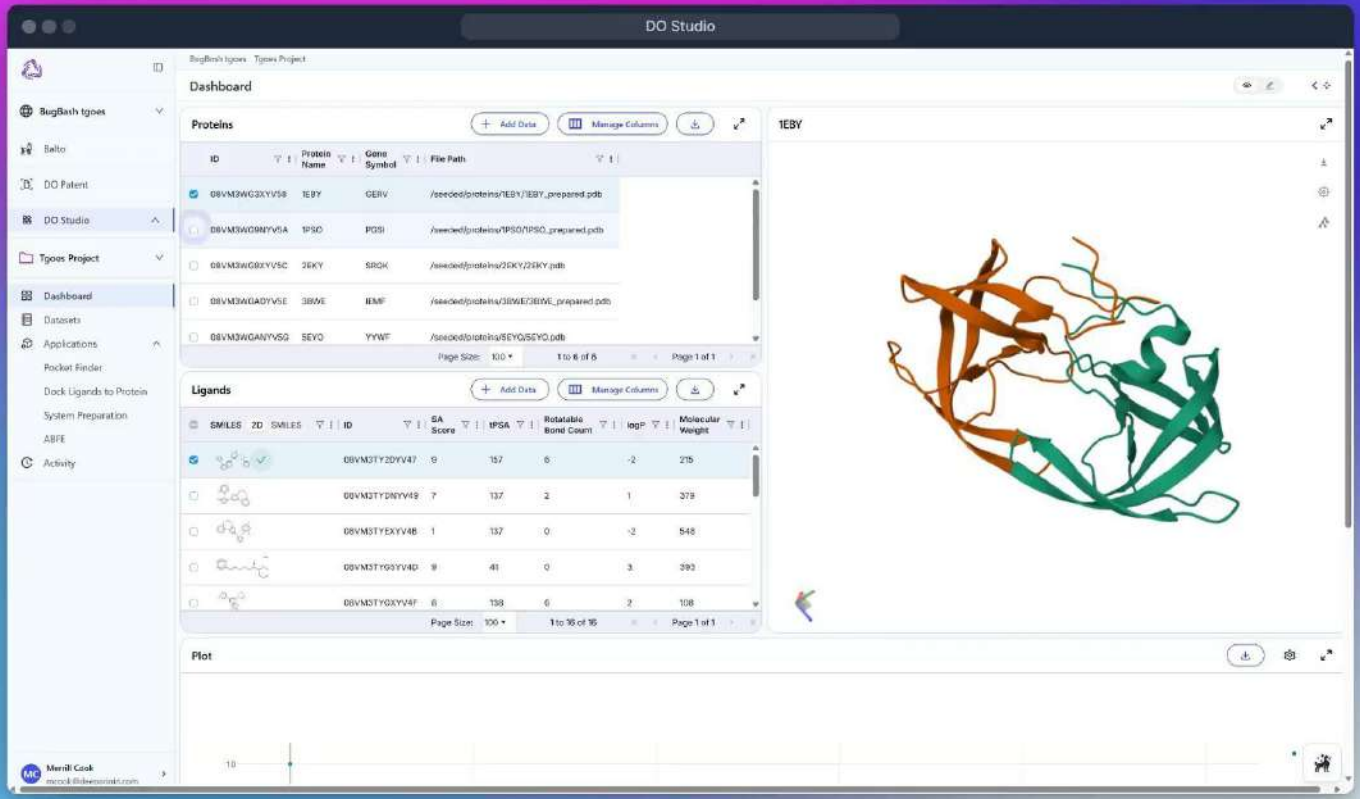
engine BIOSCIENCES

CLEARVIEW Healthcare Partners

Yale University

MIT

salesforce


DO Studio
Dashboard
BugBash tgoes
Belto
DO Patent
DO Studio
Tgoes Project
Dashboard
Datasets
Applications
Pocket Finder
Dock Ligands to Protein
System Preparation
ABFE
Activity
Merrill Cook
Proteins
Add Data
Manage Columns
ID
Protein Name
Gene Symbol
File Path
08VM3WG3XYV58 1EBY GERV /seeded/proteins/1EBY/1EBY_prepared.pdb
08VM3WG9NYV5A 1PSO PGSI /seeded/proteins/1PSO/1PSO_prepared.pdb
08VM3WG8XYV5C 2EKY SROK /seeded/proteins/2EKY/2EKY.pdb
08VM3WGADYV5E 3BWE IEMF /seeded/proteins/3BWE/3BWE_prepared.pdb
08VM3WGANYV5G 5EYQ YYWF /seeded/proteins/5EYQ/5EYQ.pdb
Page Size: 100
1 to 6 of 6
Page 1 of 1
Ligands
Add Data
Manage Columns
SMILES
2D SMILES
ID
SA Score
tPSA
Rotatable Bond Count
logP
Molecular Weight
08VM3TY2DYV47 9 157 6 -2 215
08VM3TYDNYV49 7 137 2 1 379
08VM3TYEXYV4B 1 137 0 -2 548
08VM3TY93YV4D 9 41 0 3 393
08VM3TYGXYV4F 6 138 6 2 108
Page Size: 100
1 to 16 of 16
Page 1 of 1
1EBY
Plot
10