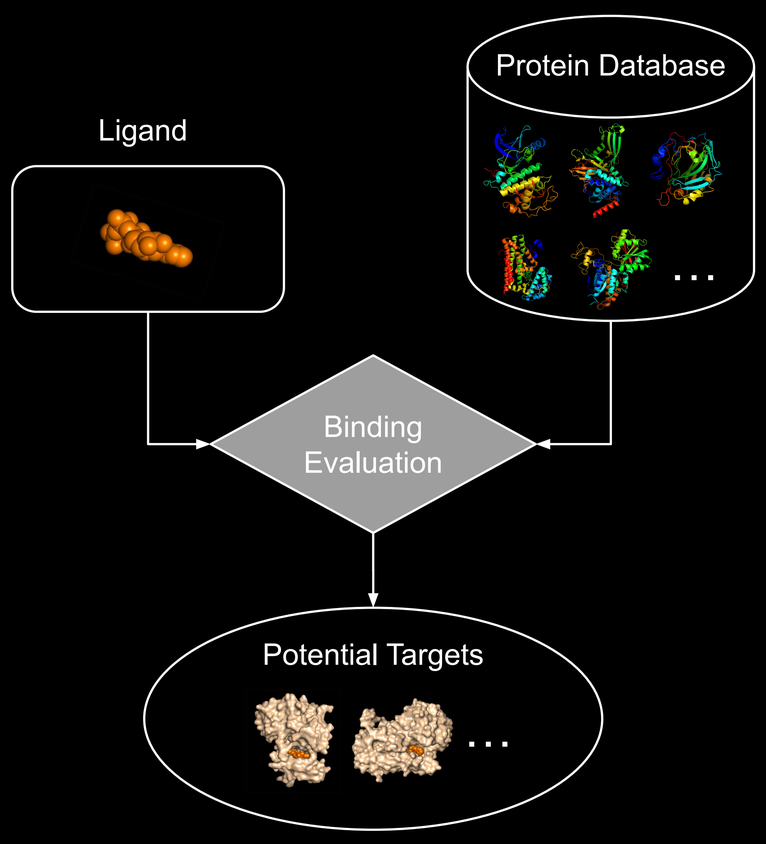
Integrated AI and physics-based methodologies

for in-Silico Target Identification

Anatoly Buchin[#], Jordan Crivelli-Decker[#], Benjamin Shields[#], Mary Pitman[#], Ly Le[#], Valentin Senicourt[#], Zane Beckwith[#], Lucia Vina Lopez[#] , Andrea Bortolato[#], Romelia Salomon Ferrer[#*]
[#]SandboxAQ, Palo Alto, California, USA. [*]Corresponding author: romelia.salomon@sandboxaq.com Presenter: anatoly.buchin@sandboxaq.com

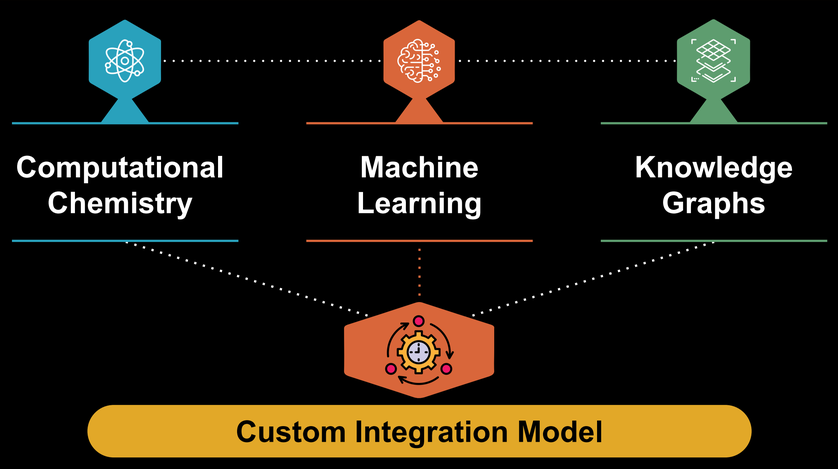
The Problem of Inverse Screening



- In silico target identification often reliant on omics data, lacks comprehensive benchmarks for methods independent of such data.
- Many available models focus on predicting likelihood of binding
- Phenotypic drug screening require complex, long and expensive experimental process to determine the precise protein target or targets responsible for the observed phenotype.
- SandboxAQ developed an innovative target ID platform that integrates AI protein structure prediction, binding affinity predictions, knowledge graph, and physics based methodologies

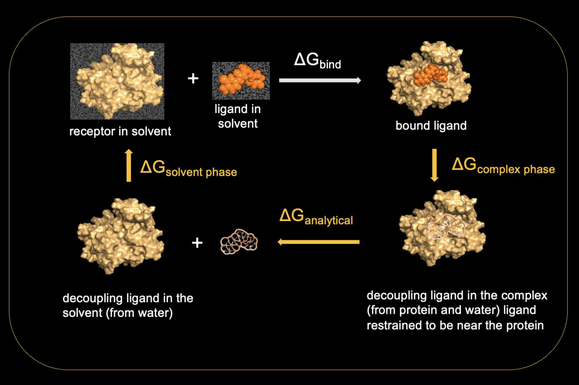
SandboxAQ's innovative platform

3 Orthogonal methods to provide unique views of target interaction



Systematic Multi-step Computational Workflow

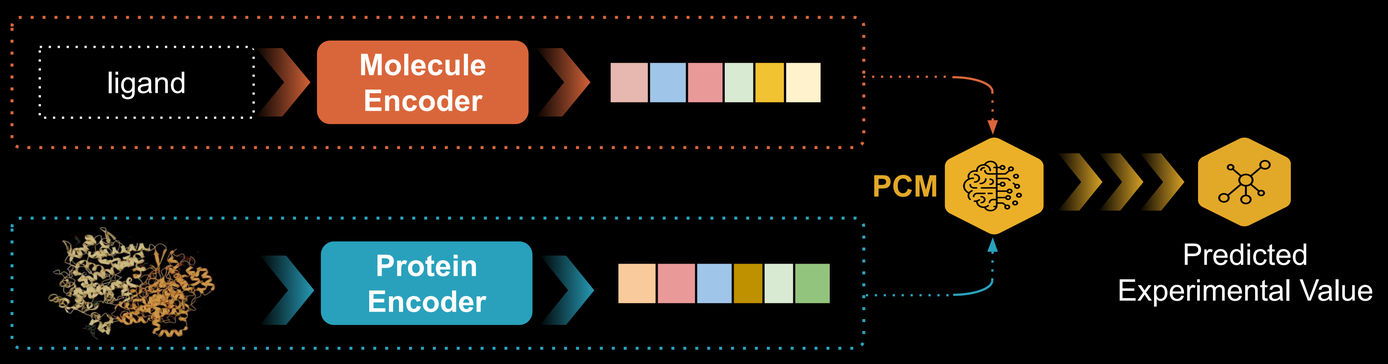
1. Protein Structure Curation: Retrieve, preprocess, and optimize the protein structure.
2. Physics, informatics, and machine learning for binding site identification: Systematically detect prepare and assess binding poses
3. Pose Selection & molecular dynamics refinement: Highlight top protein-ligand conformations
4. AQFEP Calculation: Perform alchemical free energy perturbation to estimate binding affinities.
5. Data Analysis: Binding affinity scoring per protein



AQFEP, a rigorous physics-based approach to predict ligand binding affinity, is an absolute free energy perturbation method based on the double decoupling approach and provides an optimal compromise between speed & ranking accuracy

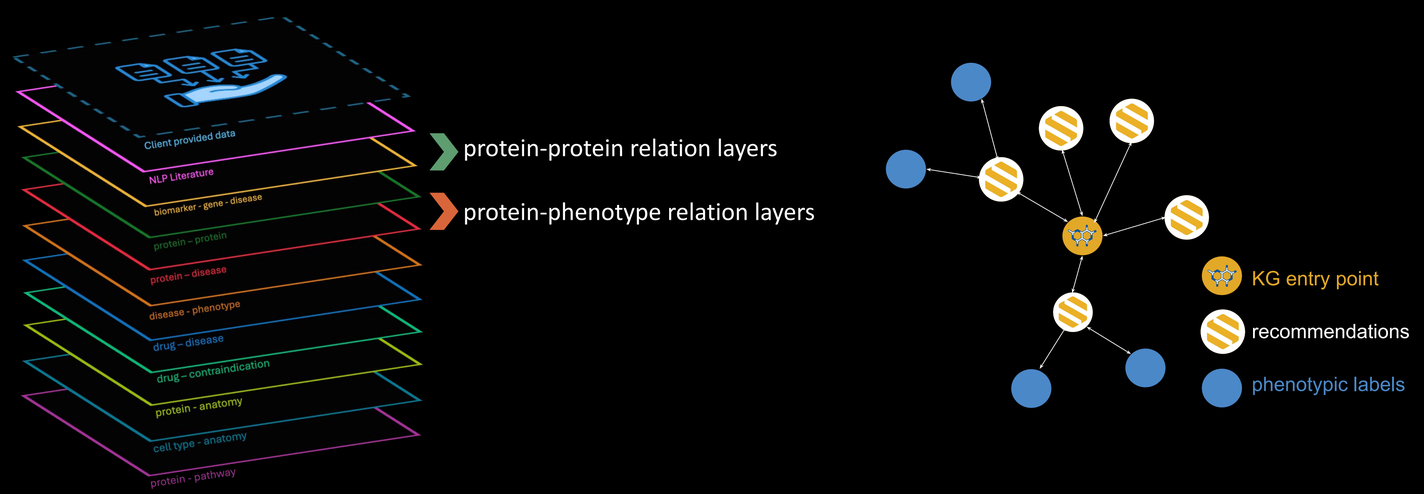
Proteochemometric Model (PCM)

- Orthogonal machine learning method trained on curated experimental datasets to predict various properties (e.g. IC50)
- Blind to the binding mode during training and cannot distinguish between pockets, no MOA
- High throughput, without need for protein-ligand conformation

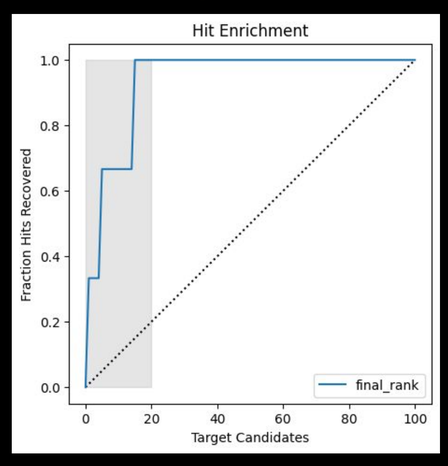


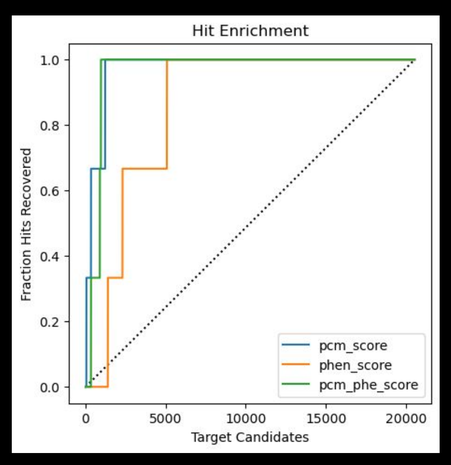
Knowledge Graph (KG)

- Knowledge Graph incorporates many "layers" of relational data
- Enable expansion and extraction of entities of interest (e.g. proteins, phenotypes) based on direct association or similarity to a knowledge graph entry point



Our Target ID Technology Accurately Ranks Targets





Target-ligand pairs were previously missed by proteome integral solubility alteration (PISA) and photoaffinity labeling (PAL).

SandboxAQ's target ID technology accurately identified targets for two different ligands:

- Inverse-screened 100 targets against ligand A and correctly ranked the true targets at #1, 5 and 15
- Inverse-screened 50 targets against ligand B and ranked the true target at #1

protein ID	rank by pcm_score	rank by phen_score	rank by pcm_phe_score
True Target 1	335	5077	956
True Target 2	48	2302	333
True Target 3	1220	1386	889

SandboxAQ Inverse platform finds true targets from whole proteome rapidly and accurately

- Whole proteome screening using proteochemometric model (PCM) enriches for hit targets in the top 10% before any fine tuning
- Phenotypic analysis using indication as therapeutic hypothesis via knowledge graph further improves ranking performance for challenging targets (e.g. Target 3)

Impacts and Applications

- SandboxAQ developed a novel approach that integrates AI protein structure prediction, binding affinity predictions, knowledge graphs and physics based methodologies to enable fast and accurate target identification across large numbers of therapeutic molecules.
- The inclusion of diverse data types and knowledge graph technology enhances the accuracy and bridge the gap between ligand-target binding and observed phenotypes.
- Serves as starting points for other applications such as drug repurposing, toxicity and selectivity prediction

References

1. *Bryant, P., Pozzati, G., & Elofsson, A. (2022). Nature communications, 13(1), 1265.*
2. *Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). arXiv preprint arXiv:2210.01776.*
3. *Qiao, Z., Nie, W., Vahdat, A., Miller III, T. F., & Anandkumar, A. (2024). Nature Machine Intelligence, 6(2), 195-208.*
4. *Aggarwal, R., Gupta, A., Chelur, V., Jawahar, C. V., & Priyakumar, U. D. (2021). Journal of Chemical Information and Modeling, 62(21), 5069-5079.*
5. *McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., ... & Koes, D. R. (2021). Journal of cheminformatics, 13(1), 43.*
6. *Crivelli-Decker, J. E., Beckwith, Z., Tom, G., Le, L., Khuttan, S., Salomon-Ferrer, R., ... & Bortolato, A. (2023). Journal of Chemical Theory and Computation.*