MACHINE LEARNING METHODOLOGIES AND THE FUTURE OF DRUG DISCOVERY

Rachelle Bienstock

XXXI Symposium on Bioinformatics and Computer-Aided Drug Discovery

Monday, October 20,2025

Increasing application of AI/ML Methods in the Drug Discovery Space…

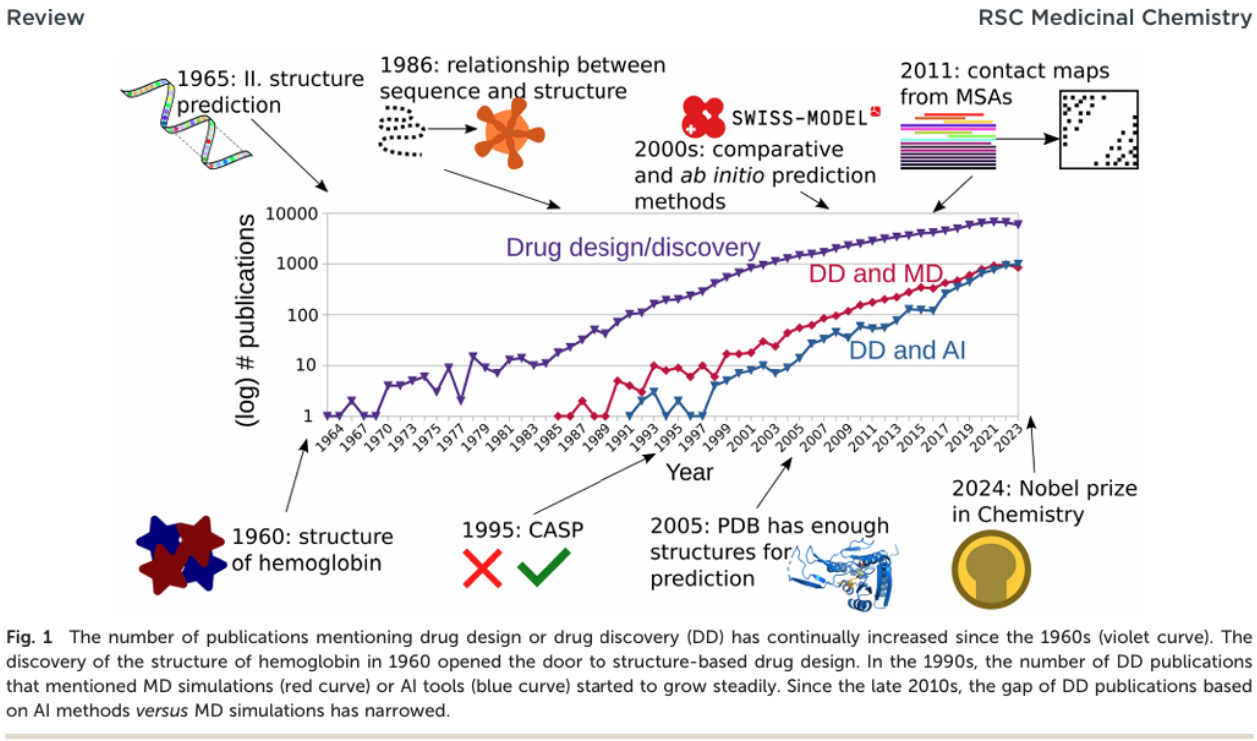


Fig. 1 The number of publications mentioning drug design or drug discovery (DD) has continually increased since the 1960s (violet curve). The discovery of the structure of hemoglobin in 1960 opened the door to structure-based drug design. In the 1990s, the number of DD publications that mentioned MD simulations (red curve) or AI tools (blue curve) started to grow steadily. Since the late 2010s, the gap of DD publications based on AI methods *versus* MD simulations has narrowed.

- With the first solved x-ray crystal structure (hemoglobin 1960)-structure based drug design was born.

- The red curve shows the growing role of molecular dynamics (i.e. physics based computational methods) in drug discovery

- The blue shows the growth of AI/ML computational methods basically now reaching comparable usage as physics based (molecular dynamics) methods

RSC Med. Chem.,2025,16, 149, The physics-AI dialogue in drug design

Big Pharma Collaborations and Data Pooling: OpenFold3 and TuneLab

- Astex Pharmaceuticals, Bristol Myers Squibb and Takeda have agreed to pool data to support work on an artificial intelligence model, joining AbbVie and Johnson & Johnson to contribute to the Federated OpenFold3 Initiative, supporting the Columbia University lab of Mohammed AlQuraishi, Ph.D., developing OpenFold3.
- Biopharma companies are collectively sitting on a vast trove of data. Pooling resources to build a bigger, more diverse data set could theoretically yield drug discovery models that are beyond what any one company could build in isolation.
- DeepMind's AI spinout Isomorphic announced two drug discovery deals, with Eli Lilly and Novartis.
- TuneLab is a collaborative platform created to offer access to AI/ML tools leveraging Lilly's own drug discovery models. Lilly TuneLab: Our AI/ML models at your fingertips https://tunelab.lilly.com/
- Google to Launch Open AI Models for Drug Discovery-TxGemma. TxGemma understands the structures of therapeutic entities, -small molecules, and proteins. Researchers can ask TX Gemma to help predict the properties of potential new therapeutics. https://developers.googleblog.com/en/introducing-txgemma-open-models-improving-therapeutics-development/

https://www.insideprecisionmedicine.com/topics/precision-medicine/google-to-launch-open-ai-models-for-drug-discovery/

https://www.fiercebiotech.com/biotech/bms-and-takeda-dive-ai-data-pool-joining-peers-collaborative-push-unfold-future?utm_medium=email&utm_source=nl&utm_campaign=LS-NL-FierceBiotech&oly_enc_id=1016C9521889B3V

Impact of AI/ML on Drug Discovery

- The Impact of ML on drug discovery has been significant in the past several years
- AlphaFold3 , RF diffusion methods, graphical neural networks and generative AI techniques have impacted every stage of the drug discovery process- target identification, hit finding and lead optimization
- In addition to big pharma, several companies which specifically focus on developing AI/ML methods across the entire drug discovery process have arisen (e.g. Recursion, Insilico Medicine, Insitro, Atomwise)
- However, some claims regarding AI/ML methods and their applications are exaggerated, and there are some caveats regarding these new methods and their applications.

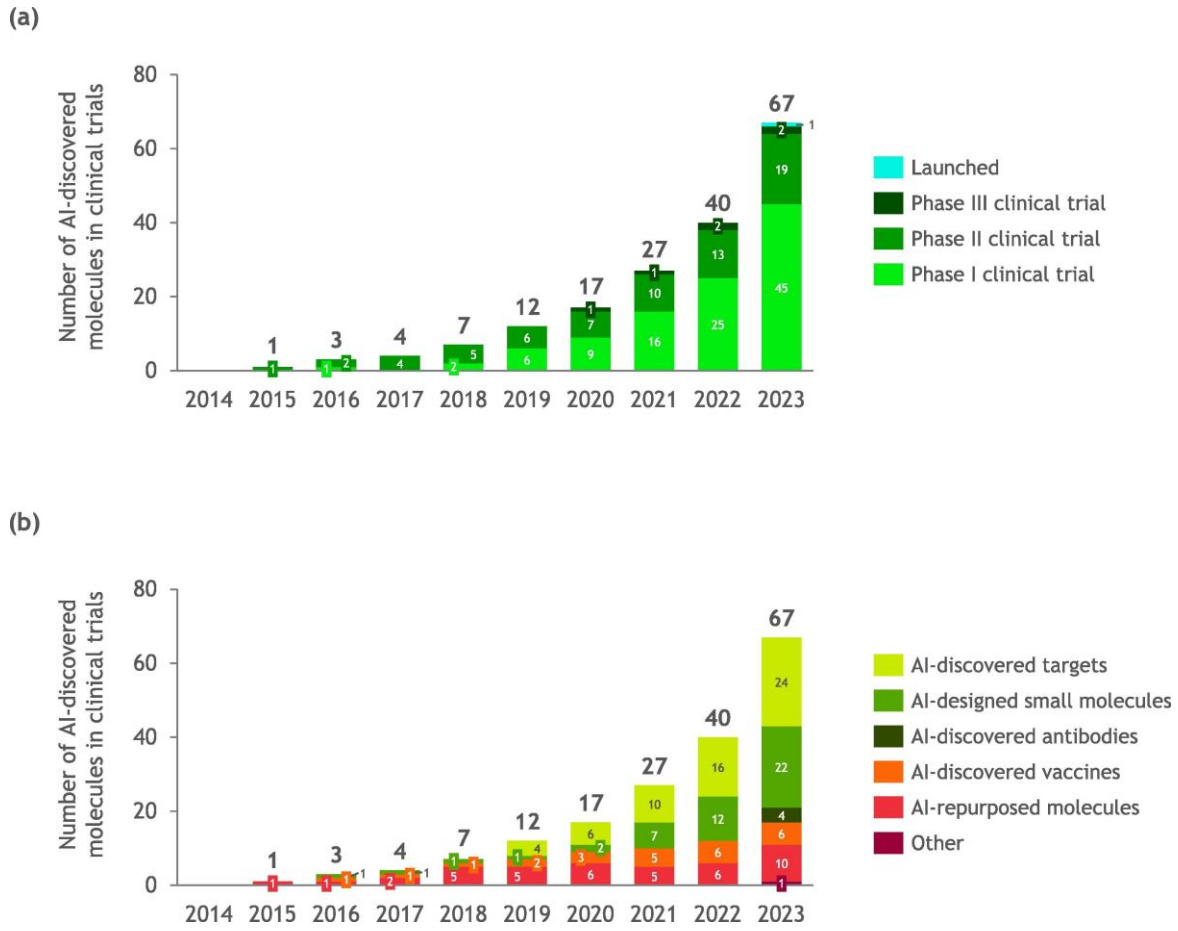
Where have ML algorithms had a significant impact in early drug discovery?

- Predicting Properties ; ADMET (absorption, distribution, metabolism, toxicity) properties
- Hit identification (database searching methods) small molecule ligand/chemical identification-neural networks, generative chemistry; AI enabled vHTS
- Target Identification (sequence LLM methods) and mechanism of action-Target/Protein Modeling and Structure Prediction; OpenFold, AlphaFold2,3, Boltz-1
- Docking and Co-Folding (AlphaFold3, Boltz-2)
- Drug Design and Optimization (including macromolecules and new molecular entities)
- ML and deep learning (DL) active and reinforced learning, are enhancing binding affinity prediction (FEP)

How are ML techniques having real and significant impact in the drug discovery space?

- How much significant improvement in identifying new chemical entities and exploring chemical space more efficiently with greater diversity?
- Machine Learning methods have impacted representation of small molecules, virtual high throughput screens, docking/ co-folding “AlphaFold3”
- How can AI/ML methods complement physics-based methods like absolute and relative free energy perturbation methods, MMGBSA, and molecular dynamics studies?
- Active learning methods, such as active learning FEP combine QSAR and FEP learning workflows; augmenting AI in structure-based drug design by feeding back scoring in AI workflows and data imputation.
- But…what is the success of these AI discovered molecules down the road in clinical trials?

Why ML Methods are of interest for drug discovery?



MKP Jayatunga,Drug Discovery Today, 29, (6), June 2024, 104009

- Bring Down Costs…
  - $2.3 billion average R and D cost to develop drug from discovery to product…(2)
- Time…
  - New chemical entity small molecule drugs typically take 4-6 years to discover
  - 10+ years average time to bring a new drug to market (5)

(2) “Unleash AI’s potential: Measuring the return from pharmaceutical innovation – 14th edition,” Deloitte, April 2024

- 5 “Research and Development in the Pharmaceutical Industry,” Congressional Budget Office, April 2021

- AI discovered small molecule growing exponentially- since 2015 (article in DDT) 75 molecules developed using ML methods have entered the clinic of which 67 were in clinical trials (2023 data)

AI Drugs in Clinical Trials….so far-very small sample size…

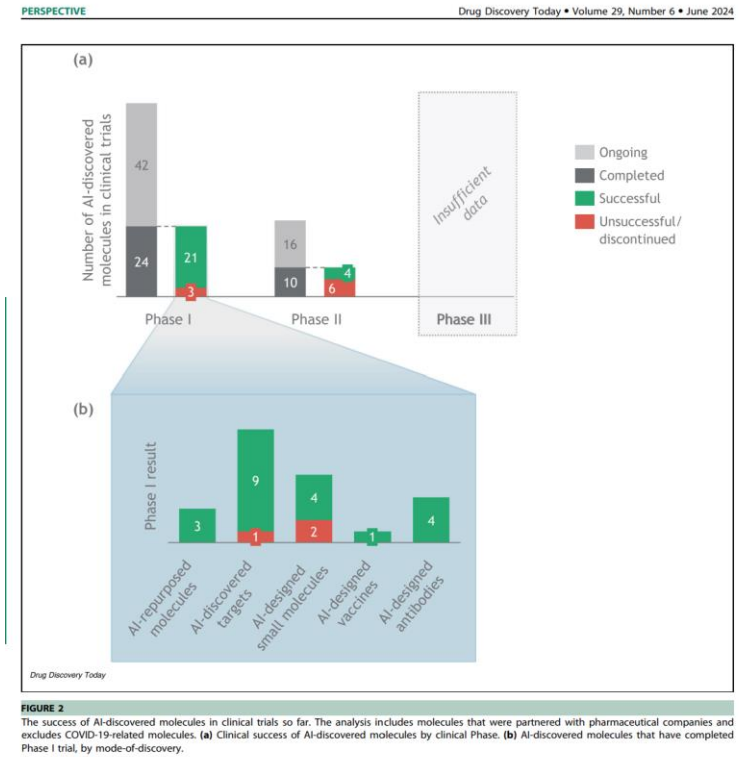
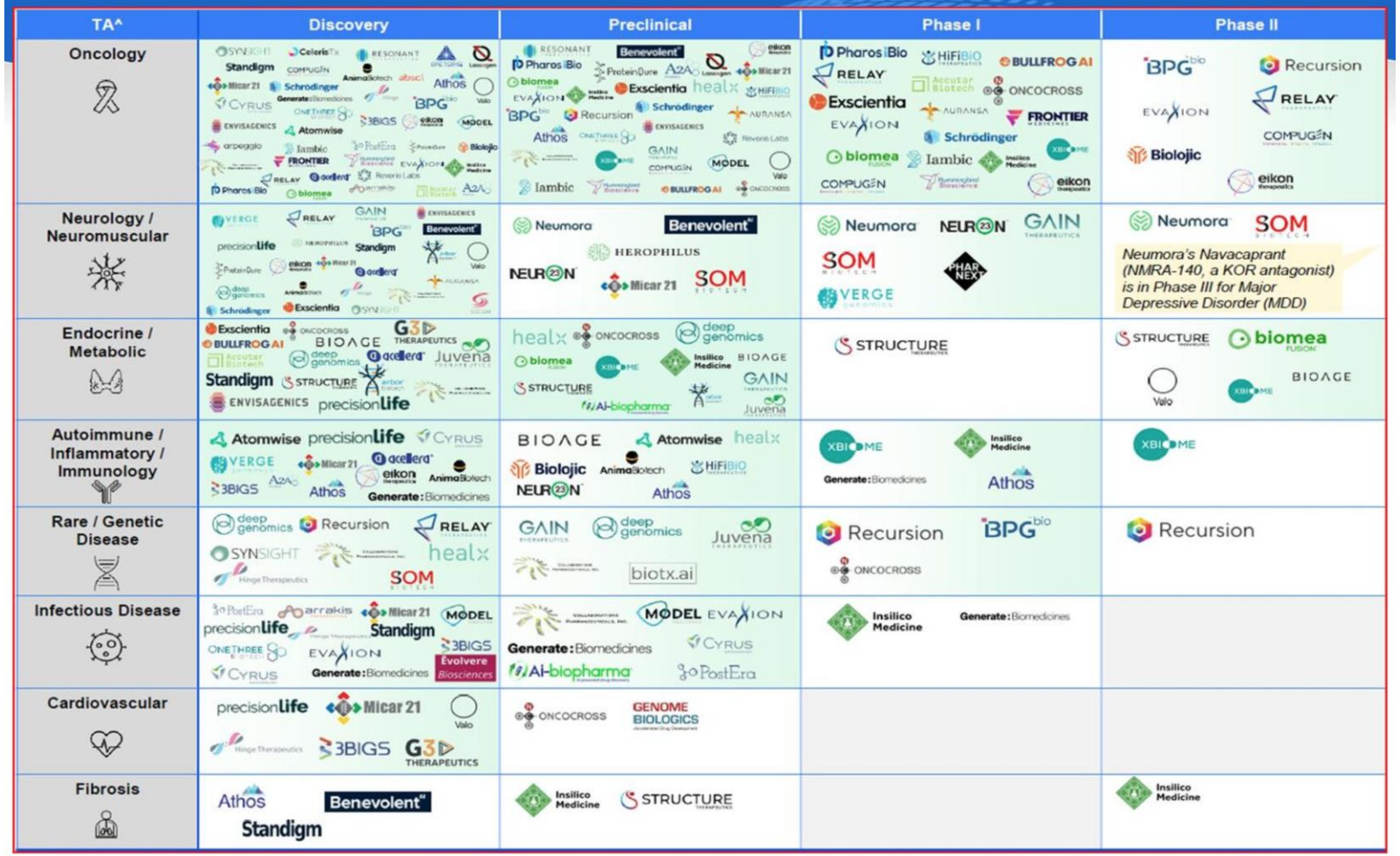


FIGURE 2
The success of AI-discovered molecules in clinical trials so far. The analysis includes molecules that were partnered with pharmaceutical companies and excludes COVID-19-related molecules. **(a)** Clinical success of AI-discovered molecules by clinical Phase. **(b)** AI-discovered molecules that have completed Phase I trial, by mode-of-discovery.

- As of 2023 December, 24, AI-discovered molecules had completed Phase I trials, of which 21 were successful
- This suggests a success rate of 80–90%, which is substantially higher than historical industry averages that range from 40% to 55– 65%.
- In Phase II trials, AI-discovered molecules drop in success …to 40%,
- Phase II typically involves the proof of a biological or mechanistic concept; this might suggest that AI algorithms can identify disease-relevant targets and pathways but are less successful in some of the more subtle aspects of drug design.

Madura KP Jayatunga 1 , Margaret Ayers 1 , Lotte Bruens 2 , Dhruv Jayanth 3 , Christoph Meier, DDT
https://doi.org/10.1016/j.drudis.2024.104009

AIDDC pipelines 2024 snapshot

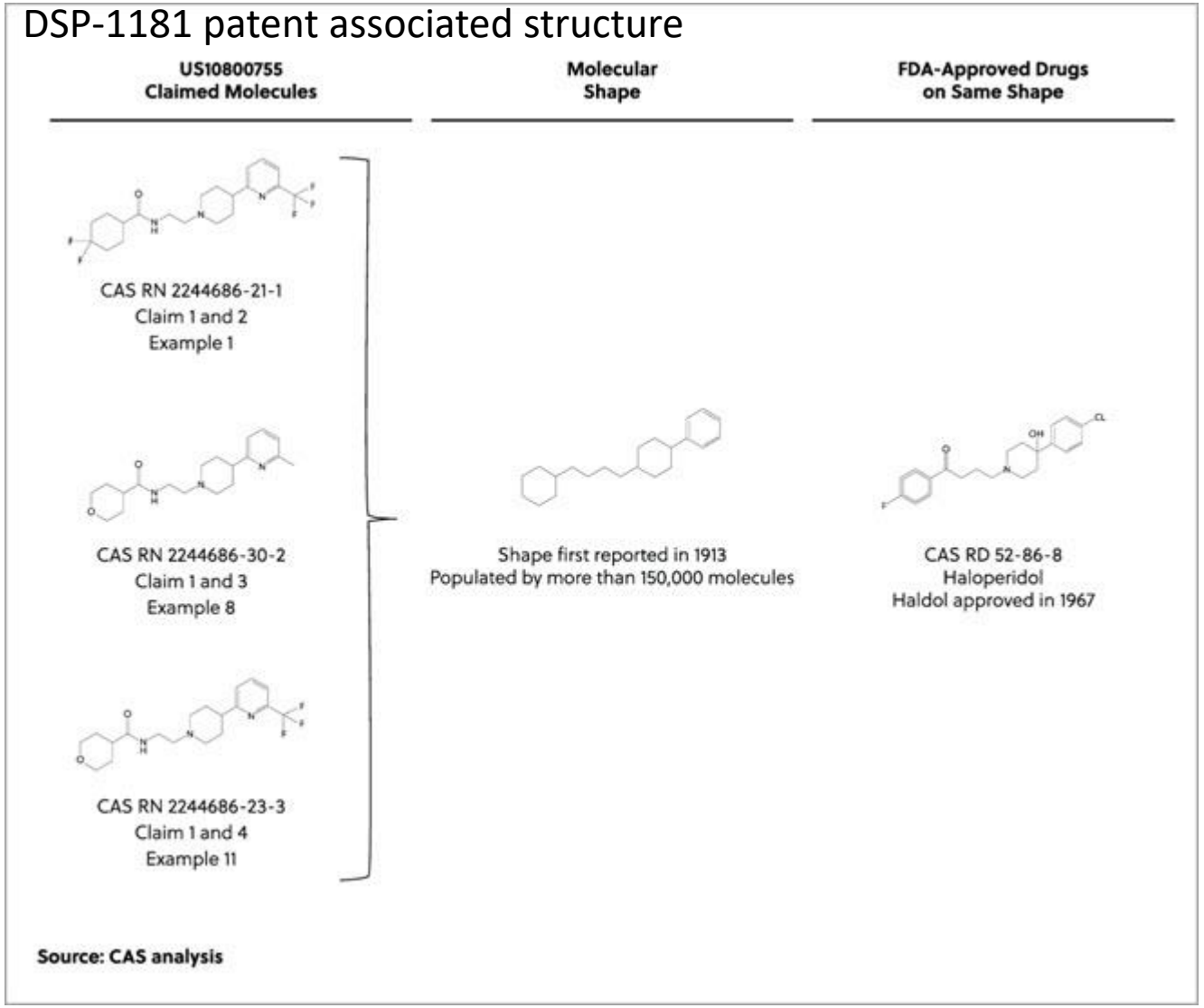


Summary from a slide presented by Dr. Christopher Southan

DSP-1181; Serotonin 5-HT1a Receptor Agonist

- Advertised as first AI medicine…developed by Exscientia(Recursion) and Sumitomo Dainippon Pharma
- DSP-1181 development discontinued after Phase I, -the acceleration of discovery timelines by AI does not guarantee clinical success.
- Selection of 5-HT1a as a target for OCD is not new and challenging
- Discontinuation Reason: The trial was stopped because the drug was found to be "insufficiently novel" and too similar in structure to the existing antipsychotic drug haloperidol.

DSP-1181 patent associated structure

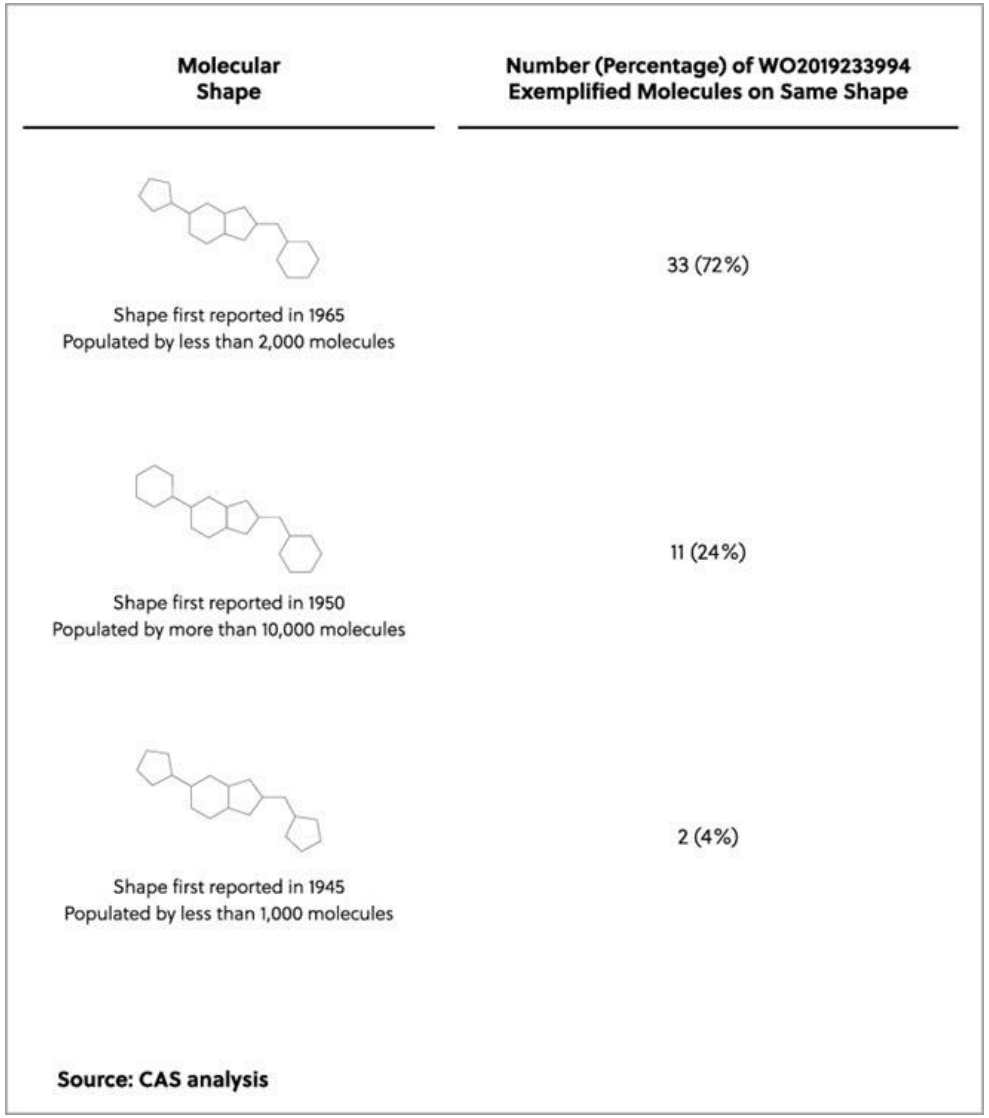


https://www.science.org/content/blog-post/another-ai-generated-drug
https://www.cas.org/resources/cas-insights/ai-drug-discovery-assessing-the-first-ai-designed-drug-candidates-to-go-into-human-clinical-trials

EXS-21546, adenosine A2a receptor antagonist

- Exscientia(Recursion) and Evotec Phase 1 clinical trail of AI discovered drug EXS-21546, adenosine A2a receptor antagonist
- Once again, scaffold similar to FDA approved drugs (Janssen) reported in earlier patents

EXS21546 candidate discontinued

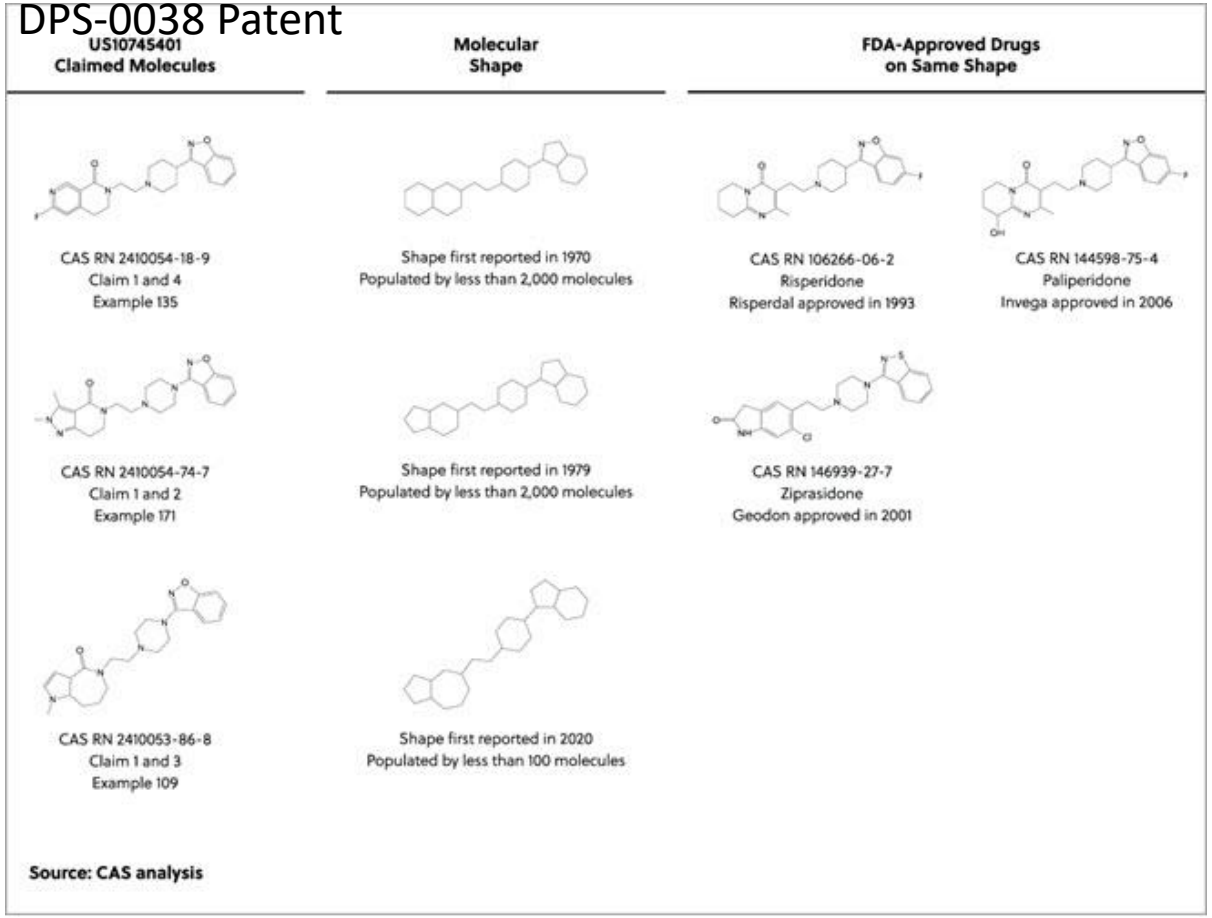


https://www.cas.org/resources/cas-insights/ai-drug-discovery-assessing-the-first-ai-designed-drug-candidates-to-go-into-human-clinical-trials

DSP-0038 dual 5-HT1a receptor agonist and 5-HT2a receptor antagonist

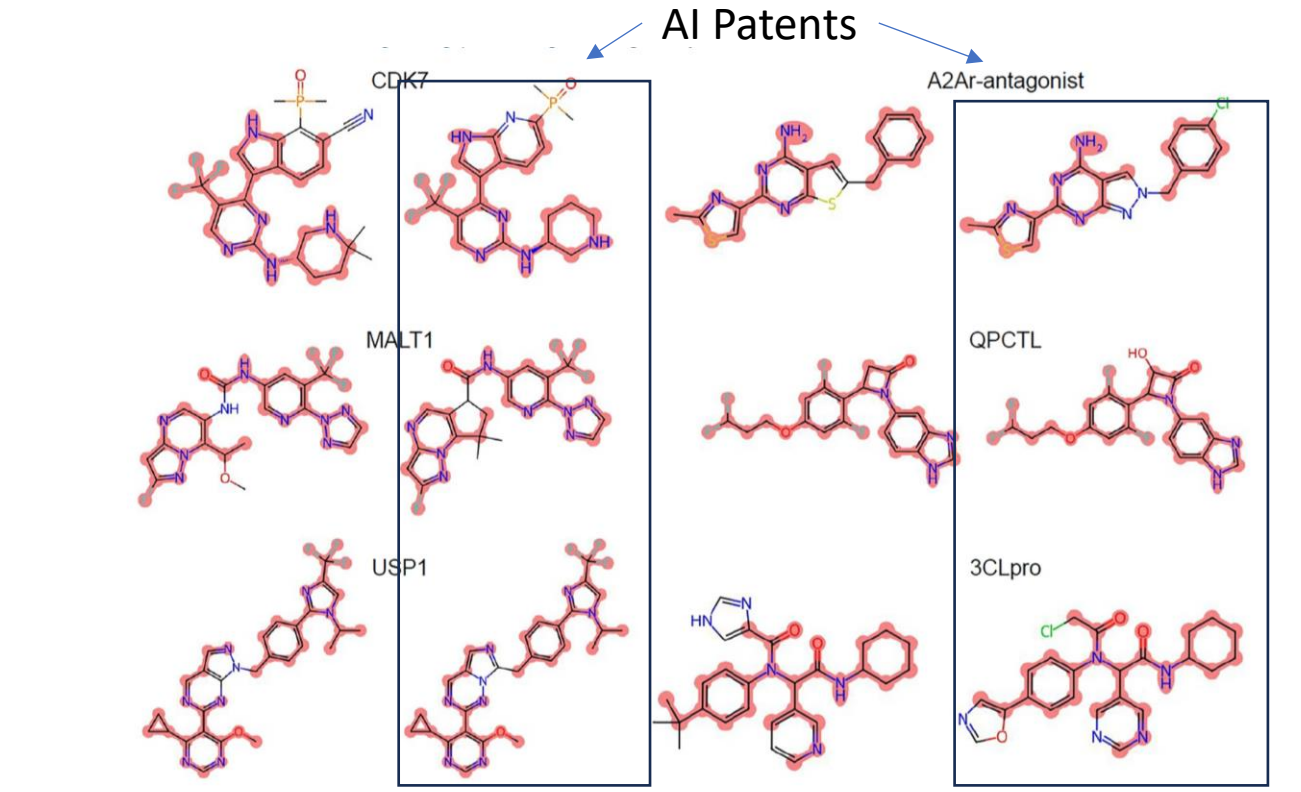
- For Alzheimer's psychosis:Phase I study between Exscientia (Recursion) and Sumitomo Dainippon Pharma.
- Shape is shared with other FDA approved antipsychotics and serotonin receptor agonists/antagonist
- Designing dual selective active molecules is a challenge with traditional drug discovery methods.
- However, there is still the challenge of novel scaffold discovery
- As of late 2025, DSP-0038 is in Phase 1 clinical trials.

DPS-0038 Patent



https://www.sumitomo-pharma.com/rd/pipeline_new-medicine/pipeline.html

ML patented molecules(right) lack of novelty



Pairs of representative compounds extracted from AI patents (right) and corresponding prior patents (left) for clinical-stage programs (CDK7[92,93], A2Ar-antagonist[94,95], MALT1[96,97], QPCTL[98,99], USP1[100,101], and 3CLpro[102,103]). The identical atoms between the chemical structures are highlighted in red.

The Atomwise AIMS Program. AI is a viable alternative to high throughput screening: a 318-target study. *Sci Rep* **14**, 7526 (2024). https://doi.org/10.1038/s41598-024-54655-z

Rentosertib (INS018-055), the first AI-generated drug to enter **Phase 2a trials (Insilico Medicine)**

- Few novel AI designed drugs have advanced in clinical trials (i.e. to Phase 2, 3 trials)
- **Rentosertib** (formerly ISM001-055), a first-in-class AI-generated small-molecule inhibitor of Traf2- and Nck-interacting kinase (TNIK) , a first-in-class target , identified as a critical regulator of idiopathic pulmonary fibrosis (IPF) discovered using a generative AI approach
- INS018-055 small molecule inhibitor designed to treat idiopathic pulmonary fibrosis.
- The biological target(TNIK) and INS018-055 were identified using Insilico Medicine commercial AI platforms, PandaOmics (target identification) and Chemistry42 (molecule generation) to generate inhibitor candidates

A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial *Nature Medicine* **volume 31**, pages2602–2610 (2025) (Insilico Medicine)

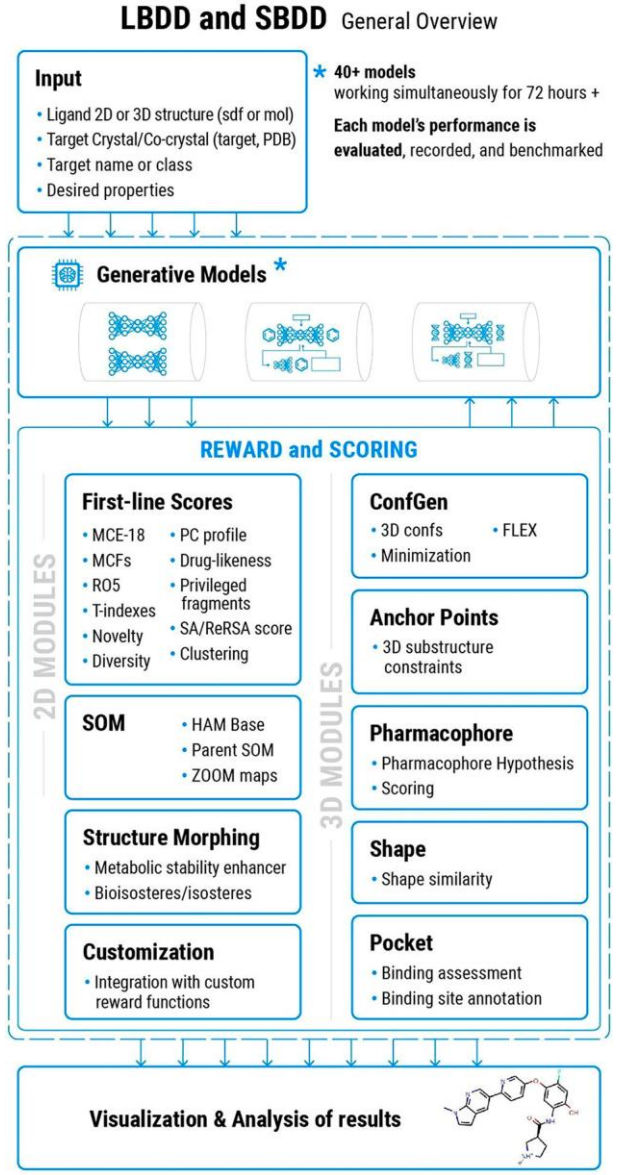
Chemistry 42 General Features

- Ligand based and structure-based design
- Ligand base design input: Input 2D or 3D ligand input as sdf , SMILES or sketcher (pharmacophore can be included)
- Structure based design: structure of a protein target, either in the *apo* format or in complex with a ligand, must be uploaded to the platform as a prepared pdb file. One can pick either the pocket around the ligand (ligand binding site) or select one from the set of alternative pockets indicated by the *Pocket Scanner Module*. Anchor points- 3D privileged scaffold
- The Chemistry42 platform is commercially available to the public (https://chemistry42.com). Parts of the platform, such as the GENTRL algorithm, (generative tensorial reinforcement learning)are available online https://github.com/insilicomedicine/GENTRL. Data for training the models is constructed from publicly available sources such as ChEMBL (https://www.ebi.ac.uk/chembl/).

Chemistry 42 Platform (Commercial- Insilico Medicine)

- Begin with data and properties of structures
- Platform contains 40+ generative model functions which are run in parallel to generate novel structures
  - including generative autoencoders, generative adversarial networks, flow-based approaches, evolutionary algorithms, language models, and others. These models employ different molecular representations – string-based, graph-based, and 3D-based.
- Filter generated structures
  - Medicinal chemistry filters include PAINS, reactivity, toxic unstable functional groups, synthetic accessibility, unique molecular descriptor that scores novelty in terms of sp3 complexity, drug-likeness, similarity to reference data set (novelty), how structurally diverse the generated molecules are based on the number of generated chemotypes following clustering. *Privileged Fragments* (PFs) which define structural motifs that contribute to the activity of a target or target class.
  - Special 3D filters- ConfGen produces conformational ensemble for each structure,3D descriptors, pharmacophore module to match hypothesis, shape similarity to reference molecule.
  - Pocket Module approximate binding affinity
- Multiple sets of reward and scoring modules (2D and 3D) assess generated structure properties dynamically
- Generated structure scores are then fed back into the models- learning phase
- All Generated structures are analyzed and ranked based on predicted properties, diversity and synthetic accessibility

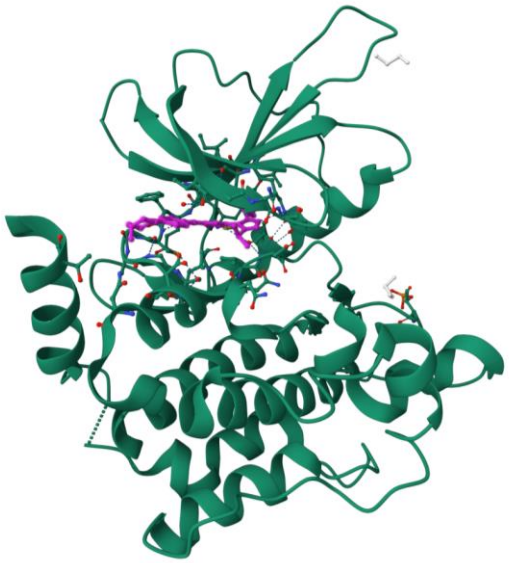
J. Chem. Inf. Model. 2023, 63, 3, 695–701

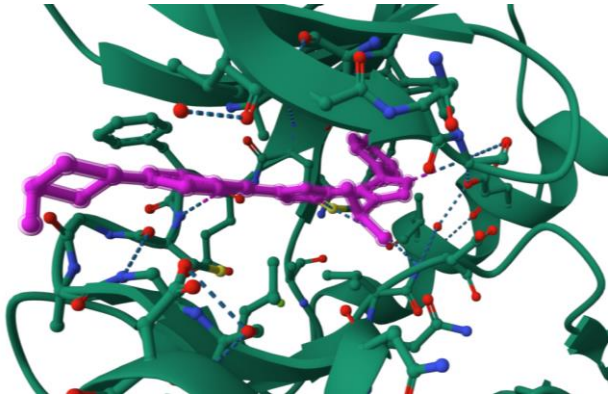


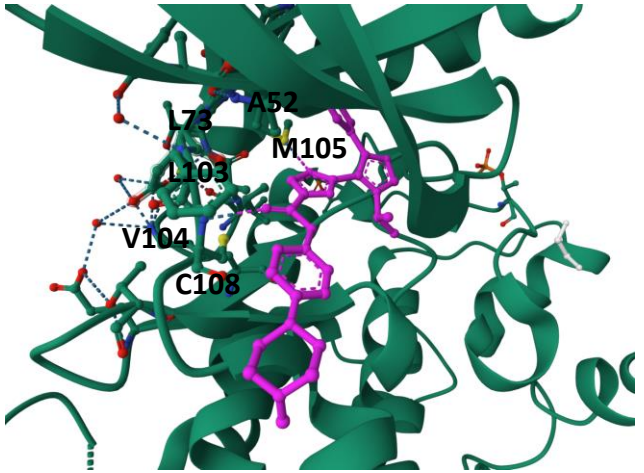
INS018-055 Traf2- and Nck-interacting kinase (TNIK) Inhibitor

- To identify TNIK inhibitors, available crystal structures of the TNIK kinase domain were used in the Chemistry42 structure-based drug-design AI workflow
- The ATP-binding site was selected as a pocket for compound generation
- AI-driven platform was configured to produce small-molecule structures capable of forming hydrogen bonds with the Cys108-NH of the TNIK hinge region.
- Additional hydrophobic pharmacophore was applied to prioritize structures bearing hydrophobic functions to deeply occupy the back cavity formed by Met105, Leu73, Leu103, Ala52 and Val104.

PDB: 8ZML

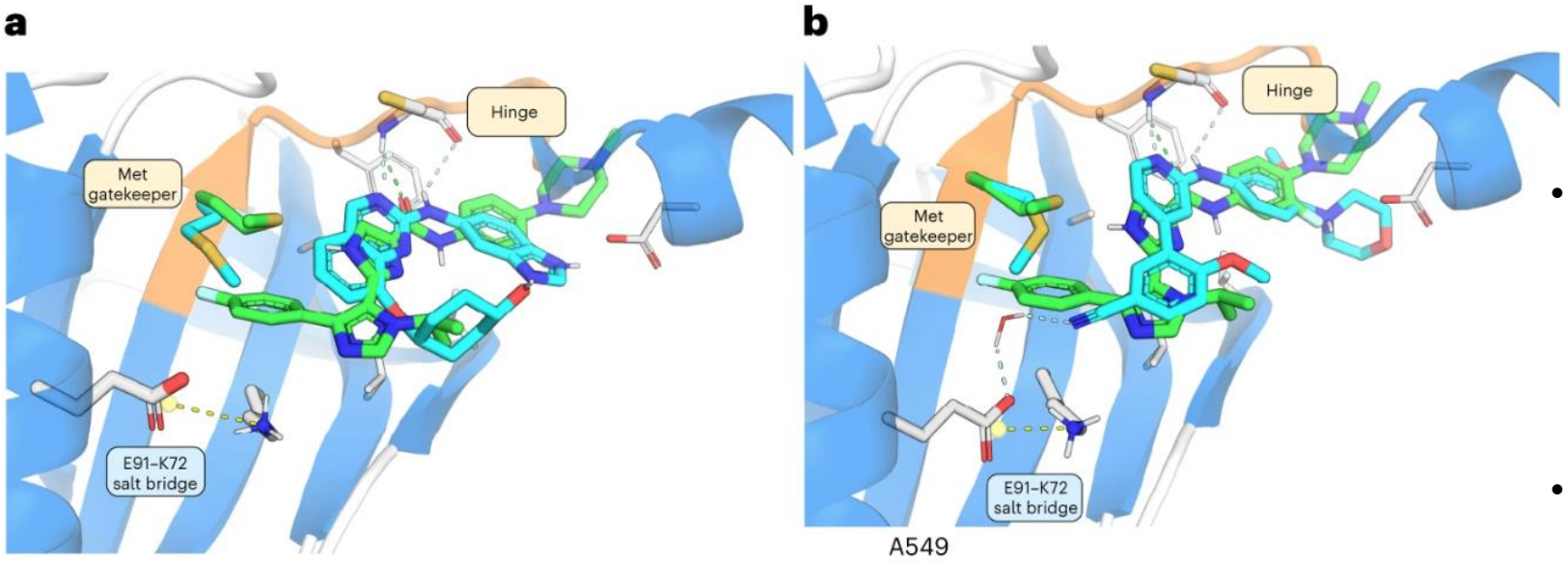






Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. Through the 'gatekeeper door': exploiting the active kinase conformation. *J. Med. Chem.* **53**, 2681–2694 (2010).

Rentosertib (INS018-055) (green) binding mode compared with previous TNIK inhibitors



Crystal structure of the NCB-0846 (cyan)-bound TNIK kinase domain (PDB 5D7A) aligned with the predicted binding mode of INS018_055 (green). **b**, Crystal structure of the compound 9 (cyan)-bound TNIK kinase domain (PDB 5AX9) aligned with the predicted binding mode of INS018_055 (green).

- AI-driven platform small-molecule structure results capable of forming hydrogen bonds with the Cys108-NH of the TNIK hinge region.
- Targeting less-conserved adjacent allosteric pockets (such as a hydrophobic back cavity close to the gatekeeper residue) in addition to the active site can achieve better selectivity of the lead compounds.

Ren, F., Aliper, A., Chen, J. *et al.* A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat Biotechnol* **43**, 63–75 (2025). https://doi.org/10.1038/s41587-024-02143-0
Fabio Zuccotto, J. Med. Chem. 2010, 53, 7, 2681–2694

Novelty of **Rentosertib** (INS018-055) Inhibitor

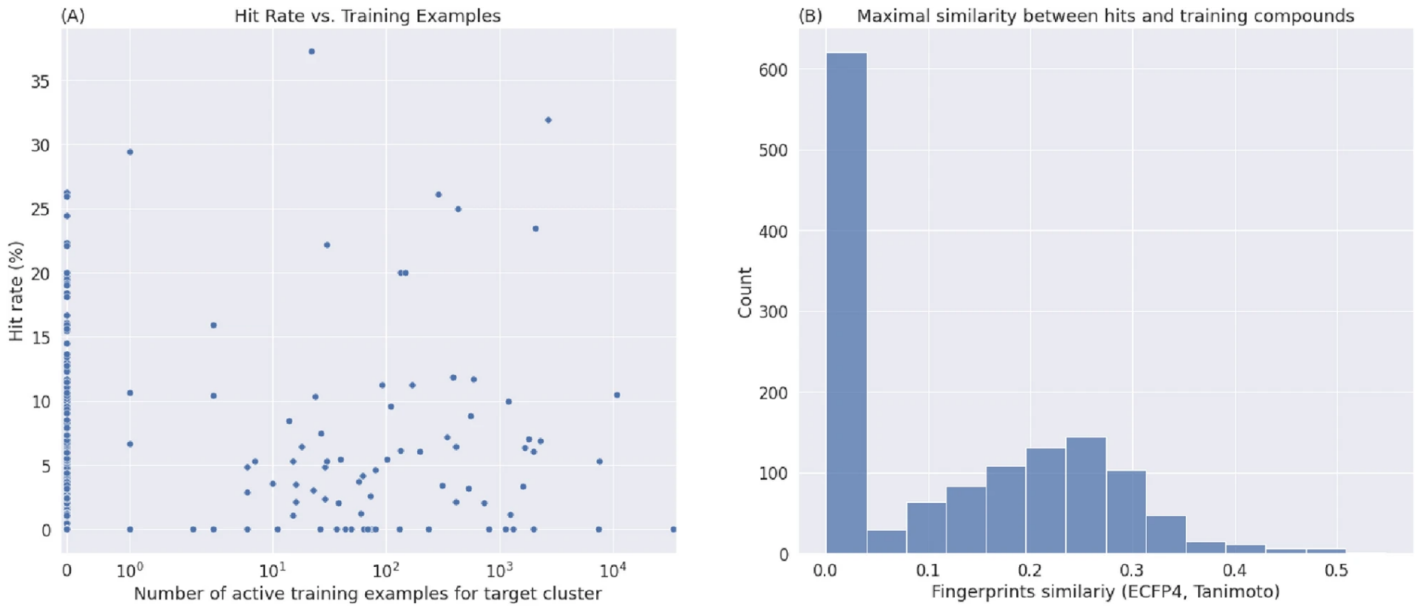
Table S3-2. Tanimoto similarity scores between INS018_055 and known TNIK inhibitors.

Molecule Name	SMILES	Tanimoto Similarity Score
Compound 21k[1]	O=C1c2ccc(-c3cnc4[nH]ccc4c3)cc2OCCN1Cc1ccc(F)cc1	0.15
NCB-0846[2]	OC1CCC(Oc2cccc3cnc(Nc4ccc5nc[nH]c5c4)nc23)CC1	0.05
Compound 16 [3]	CN1CCN(Cc2ccc(-c3cnc4[nH]cc(-c5cc(C#N)cc(NS(C)(=O)=O)c5)c4c3)cc2)CC1	0.08
Compound 8 [4]	COc1ccc(C#N)cc1-c1ccnc(Nc2ccc(N3CCOCC3)cc2)c1	0.15
PF-794 [5]	CC(C)NC(=O)c1ccc(-c2cnc(N)c(-c3ccc(C#N)cc3)c2)cc1	0.07
ON108600 [6]	O=C1Nc2cc(S(=O)(=O)Cc3c(Cl)cccc3Cl)ccc2SC1=Cc1ccc(O)c([N+](=O)[O-])c1	0.06
Compound 3 [7]	CN1CCN(c2ccc(NC(=O)c3ccc(-c4cc(Cl)ccc4Cl)o3)cc2)CC1	0.36

Atomwise-virtual high throughput screen

- Virtual high throughput AtomNet platform is a graph convolutional neural network architecture with atoms represented as vertices and pairwise distance dependent edges representing atom proximities
- They used their platform to identify novel bioactive scaffold hits for a diverse set of 235 out of 318 targets without any previously known x-ray structures or binding ligands
- Their molecular hits were novel and not similar to ones found by conventional HTS using standard libraries
- Several of their hits were first in class novel scaffold binders for their targets
- They identified hits for challenging targets- allosteric binds and protein-protein interactions
- Their ML virtual screen is multilayered and does consider a physics-based docking score (AutoDock Vina)

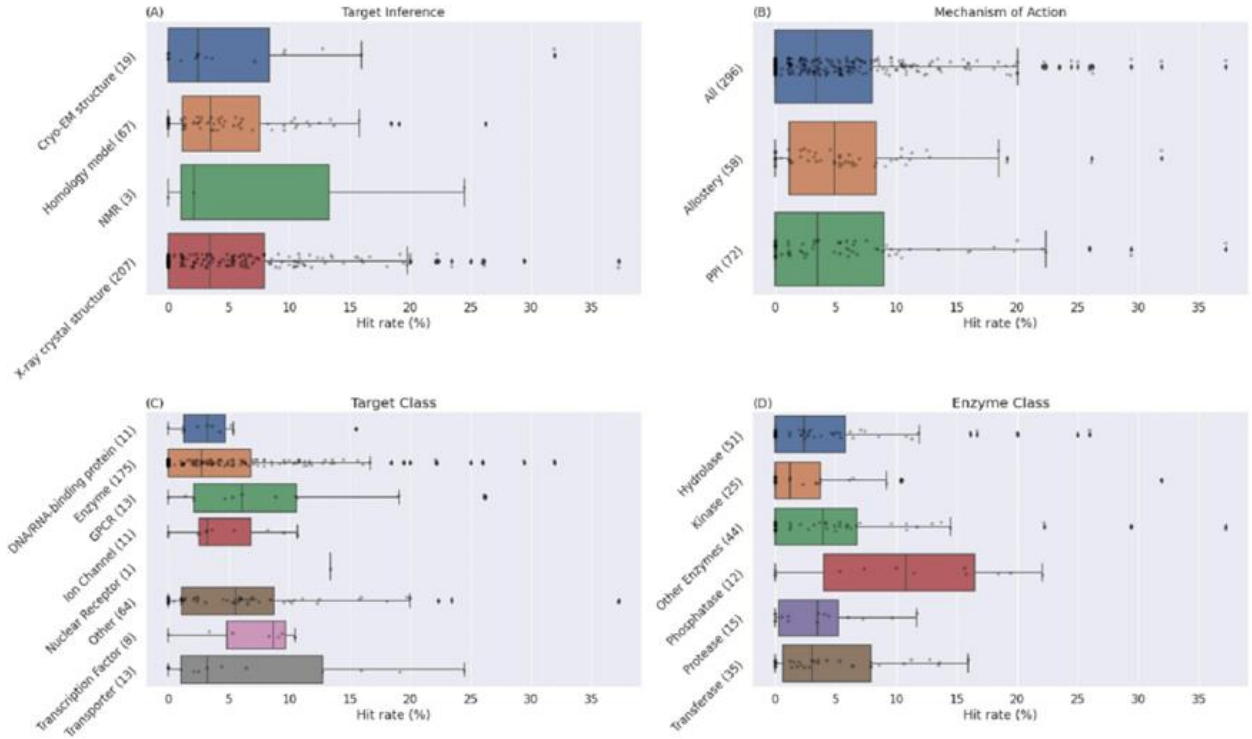
Atomwise : Novelty of Hits



(**A**) An illustration of the hit rate versus the number of training examples available to our model. Each point represents a project, with the x-axis denoting the number of active molecules in our training for the target protein or homologs and the y-axis denoting the hit rate of the project (the percentage of molecules tested in the project that were active). The model shows no dependence on the availability of on-target training examples. For 70% of the targets, the AtomNet model training data lacked any active molecules for that target or any similar targets with greater than 70% sequence identity, yet the model achieved a hit rate of 5.3% compared to 6.1% when on-target data was available. (**B**) The distribution of similarities between hits and their most-similar bioactive compounds in our training data. Our screening protocol ensures that the compounds subjected to physical testing are not similar to known active compounds or close homologs (< 0.5 Tanimoto similarity using ECFP4, 1024 bits). Because 70% of the AIMS targets had no annotated bioactivities in our training dataset, hits identified in these projects have a similarity value of zero.

The Atomwise AIMS Program. AI is a viable alternative to high throughput screening: a 318-target study. *Sci Rep* **14**, 7526 (2024). https://doi.org/10.1038/s41598-024-54655-z

Atomwise : performance diverse targets; with and without structural information



Hit rates obtained for the 296 AIMS projects. (**A**) A comparison of hit rates using X-ray crystallography, NMR, Cryo-EM, and homology for modeling the structure of the proteins. Each point represents a project with the x-axis denoting the hit rate of the project (the percentage of molecules tested in the project that were active). The number of projects of each type is given in parentheses. We observed no substantial difference in success rate between the physical and the computationally inferred models. We achieved average hit rates of 5.6%, 5.5%, and 5.1% for crystal structures, cryo-EM, and homology modeling, respectively. The number of projects using NMR structures is too small to make statistically-robust claims. (**B**) A comparison of hit rates observed for traditionally challenging target classes such as protein–protein interactions (PPI) and allosteric binding. Of the 296 projects, 72 targeted PPIs and 58 allosteric binding sites. The average hit rates were 6.4% and 5.8% for PPIs and allosteric binding, respectively. (**C**) Comparison of hit rates observed for different target classes and (**D**) enzyme classes. No protein or enzyme class falls outside the domain of applicability of the algorithm.

ML and Target (protein structure prediction)

- After many years of CASP Structure Prediction Competitions, Threading and Homology Modeling- First AlphaFold Model at CASP13
- AlphaFold developers John Jumper and Demis Hassabis shared 2024 Nobel Prize Chemistry
- AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/)
- AlphaFold relies on multiple sequence alignments to find evolutionary relationships to predict inter residue contacts-original AlphaFold used a statistical model;
- AlphaFold2 uses transformer architecture to integrate MSA and structural template information;
- AlphaFold3 uses multiple sequence alignment to find close residue close pairs; other similar models- ESMFold
- AlphaFold3 and RoseTTAFoldAll Atom use diffusion models

AlphaFold 3 (AF3) and RoseTTAFold All-Atom (RFAA)

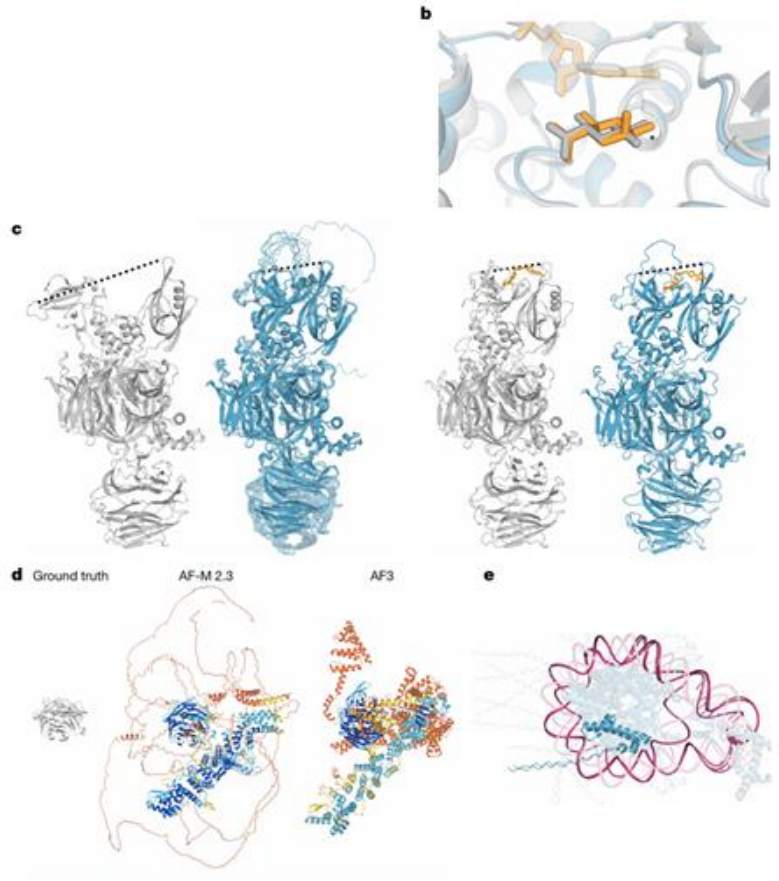
- Co-Folding: incorporating interactions with proteins, nucleic acids, and small molecules within a single predictive framework
- Diffusion-based architecture, AF3 removed stereochemical loss, amino-acid specific frames, and special handling of bonding patterns; de-emphasized protein evolutionary data
- These changes allowed AF3 to train on nearly all structural data which extended its modeling capabilities to new tasks, such as protein-ligand and protein-nucleic acid complexes.
- AF3 and RFAA Performance: In blind docking of small molecules to proteins with the PoseBusterV2 dataset, AF3 achieved an accuracy of around 81% for predicting the native pose within 2Å RMSD
- Chai-1 and Boltz-1 , AlphaFold3 comparable accuracy

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.10.615955

Buttenschoen, M., Morris, G. M. & Deane, C. M. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2024).

However, AF3 Issues…

- Limitations of AF3 with respect to stereochemistry, hallucinations, dynamics and accuracy for certain targets.
- Stereochemistry, two main classes of violations. The first is that the model outputs do not always respect chirality (Fig. 5b), despite the model receiving reference structures with correct chirality as input features
- Second class of stereochemical violations is a tendency of the model to occasionally produce overlap ping (clashing) atoms in the predictions. This sometimes manifests as extreme violations in homomers in which entire chains have been observed to overlap
- Diffusion-based AF3 model introduces the challenge of spurious structural order (hallucinations) in disordered regions
- Dynamics- conformational states not captured correctly…example, E3 ubiquitin ligases natively adopt an open conformation in an apo state and have been observed only in a closed state when bound to ligands, but AF3 exclusively predicts the closed state for both holo and apo systems



Nature, (630), 2024,Accurate structure prediction of biomolecular interactions with AlphaFold 3 https://doi.org/10.1038/s41586-024-07487-w

Do Deep Learning Models for Co-Folding Learn the Physics of Protein-Ligand Interactions: AF3, RoseTTAFold AllAtom, Boltz-1 and Chai-1

Article https://doi.org/10.1038/s41467-025-63947-5

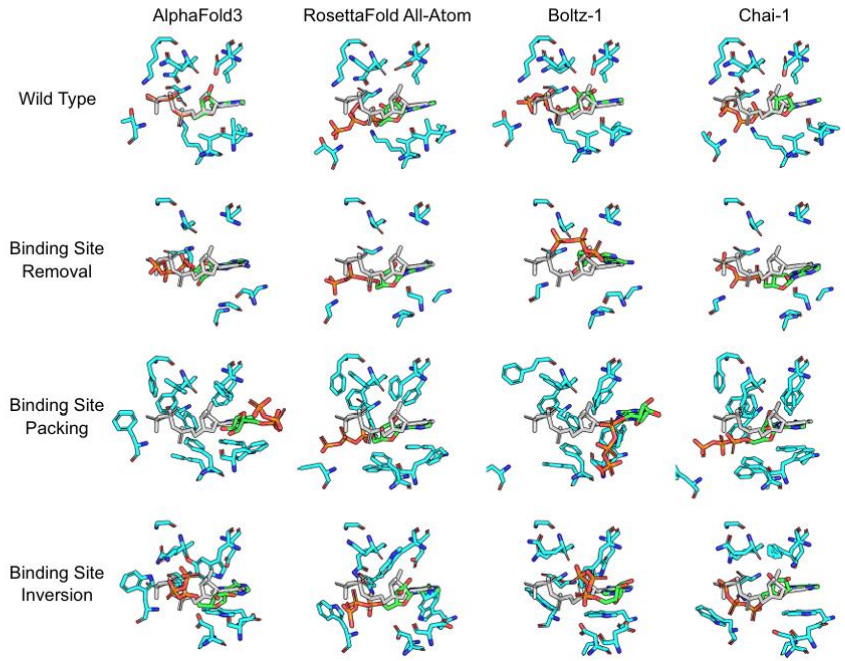


Fig. 1 | Binding site mutagenesis challenges against co-folding models using the CDK2 system (PDB: 1B38). Predicted binding-site residues are shown as cyan sticks, predicted ligand poses are shown as green sticks, and the original co-crystallized ligand pose is shown as gray sticks. The first row shows each model's prediction for the wild-type protein-ligand system prior to any modification. The remaining rows show different adversarial challenges where all binding site residues are mutated. In binding site removal, all residues are mutated to glycines effectively removing all ligand-side-chain interactions from the original system. The packing challenge mutates all residues to phenylalanine, removing all native interactions with side-chains and further occupying the pocket with bulky, hydrophobic groups. In the inversion challenge, binding site residues are mutated to residues with dissimilar properties. These mutations should annihilate the binding site and remove the majority of native protein-ligands interactions necessary for binding. However, in many cases the ligand is still predicted within the binding site and can adopt a low RMSD pose, indicating that these co-folding models are not predicting poses based on physics of interactions, but rather learning patterns in global protein structures and sequences.

- Bias toward preserving the original binding geometry, even when significant structural, chemical, and physical changes were introduced (mutated residues in binding pocket)- Why? Most of the co-folding models rely heavily on multiple sequence alignment and 3D template-based input features to make predictions
- When removing the binding site residues, or small mutations, the sequence alignment and template search will return exactly the same results as before, as they are still the closest related sequences and structures in the data set. Therefore, the MSA and template features that the network accepts as input are identical despite the mutations, leading the model to make a similar prediction
- Validation by a physics-based methods, molecular dynamics simulations, or inclusion of some experimental data (i.e. mass spec or NMR) would improve predictions

Can ML/AI predicted Structures be used for FEP? FEP and Boltz-2

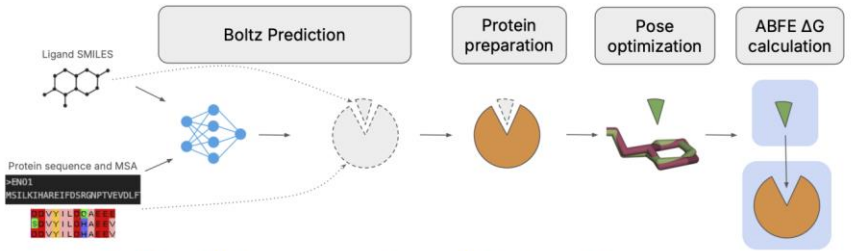


Figure 1: The Boltz-ABFE pipeline uses Boltz-1/Boltz-2 to predict the protein-ligand complex given the SMILES of the ligand and sequence(s) of the protein chain(s). Afterwards, the predicted protein receptor is further prepared for docking and molecular dynamics applications. To correct ligand chemistry errors, the known ligand is re-docked into the receptor using the predicted complex as a template. The resulting structure is then relaxed and equilibrated before performing the ABFE simulation.

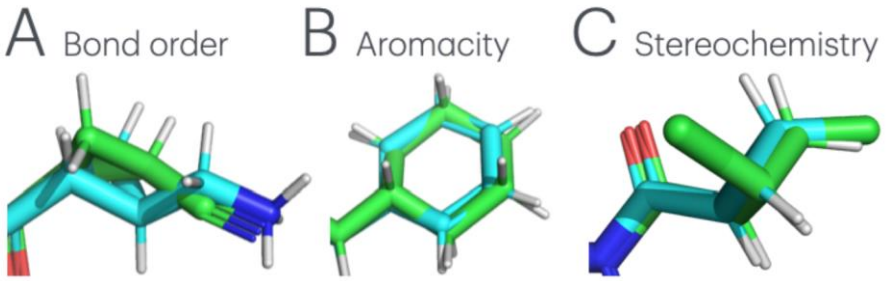


Figure 2: Common chemical inaccuracies in Boltz-1-generated ligand structures and the effects of refinement strategies. Boltz-1 occasionally models the input SMILES incorrectly, leading to incorrect bond orders (A), improper aromaticity (B), or incorrect stereochemistry (C).

- Boltz predicted protein-ligand complexes were used in ABFE to initialize simulations and could accurately estimate the free energy of binding (ΔG), (*provided that some care was taken when choosing which structure prediction is taken forward for use in MD simulations*).
- A pipeline that prepares Boltz predicted structures for MD by automating the removal of common defects in the predicted structures such as overlapping atoms , clashes, and incorrect ligand stereochemistry.
- The goal of Boltz-ABFE is to accurately predict the protein-ligand binding affinity from the compound's SMILES string and protein sequence information alone
- Boltz-ABFE, a pipeline corrects defects of predicted structures and allows to perform 15 free energy simulations without requiring experimentally-determined protein-ligand complex structures

arXiv:2508.19385v1 https://doi.org/10.48550/arXiv.2508.19385

Boltz-ABFE 4 proteins from the FEP+ benchmark: CDK2, TYK2, JNK1, and P38

- Prediction Methods:
  - **Boltz-2 without redocking (labeled ”Boltz-2” in the Figure), GREEN**
  - **Boltz-1 with redocking using POSIT (”Boltz-1+P”) DARK BLUE**
  - **Boltz-2 with redocking using POSIT (”Boltz-2+P”) PURPLE**
  - **Boltz-2 Affinity module (”Boltz-2-A”) RED**
- Results were compared against simulations starting with **crystal structures with POSIT re-docking (ORANGE)**.
- ABFE simulations starting from the crystal structure achieved the most consistent results over all the targets when considering all of the success metrics (RSME, MUE, R2 and Kendall’s τ).
- The best results for TYK2 start from Boltz1+P predicted structures.
- The ABFE results initiated from any of the Boltz predicted structures achieved satisfactory results with MUE < 1 kcal/mol on average. (MUE= mean unsigned error)
- The poorer performance of Boltz-2 can be attributed to the TYK2 protein, where the predictions have MUE’ s > 1 kcal/mol. the predicted TYK2 structures, we observed that Boltz-2 flipped a side-chain in the binding pocket compared to Boltz-1,
- Boltz-2 Affinity module also performs well on this dataset, yielding correlation metrics that are slightly better than those from the Boltz-1+P ABFE simulations.

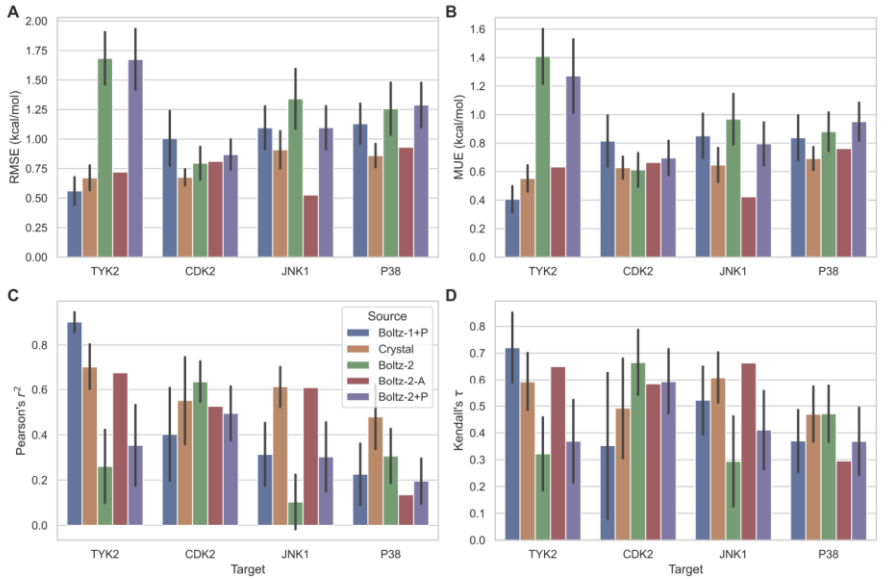


Figure 8: (A) Root mean square error (RMSE) between calculated and experimental ΔG values for four target systems (TYK2, CDK2, JNK1, and P38), using input structures generated by the Boltz-1/2 models or taken from the protein-ligand benchmark. Three replicates were performed for all sets. RMSE error bars represent the standard deviation across replicates. (B) Mean unsigned error (MUE) between calculated and experimental ΔG values. (C) Pearson’s r^2 between calculated and experimental ΔG values. (D) Kendall’s τ correlation between calculated and experimental ΔG values.

PoseBuster Docking Challenge: ML vs Physics based methods…

- PoseBuster's checks the quality of docked ligand structures using the RDKit Distance Geometry Module rules evaluating stereochemistry and inter and intramolecular measurements- bond lengths, planarity of aromatics and atom clashes.
- In the evaluation and comparison of five deep learning "AI" docking methods- DeepDock, DiffDock, EquiBind, TankBind and Uni-Mol, compared with traditional physics-based docking methods -Auto Dock Vina and CCDC Gold, **the physics-based docking methods limited the degrees of movement in the ligand to only the permissible rotatable bonds in the ligand and included penalties for protein and ligand clashes.**
- **The conclusion reached by this published study was that "no deep learning-based method yet outperforms classical docking tools". And "molecular mechanics force fields contain docking-relevant physics missing from deep learning methods"**

PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences†
Martin Buttenschoen, Garrett M. Morris and Charlotte M. Deane *: Chem. Sci., 2024, 15, 3130

In summary: ML cannot extrapolate if data is not represented in the training set…

- The analyses and predictions made by AI and ML software can only be as good as the data sets that support them.
- Even small, biologically plausible perturbations can result in significant discrepancies in predicted structures, highlighting vulnerabilities in these models. –like the residue mutations in the binding site.
- ML designed molecules are more successful in Phase I than Phase II
  - ML success could be explained by the fact that test data sets already have optimized ADME and safety profiles and so using these types of data sets for training makes Phase I outcomes more successful, but this does not transfer to Phase II and beyond…

In Summary: Deep Learning Model Problems…

- Deep learning models rely in data driven patterns
- Studies* have shown that the performance of these deep learning methods predominantly comes from their pocket finding ability and not an ability to resolve detailed molecular interactions.
- Fundamental principles of physical interactions- hydrogen bonding, electrostatic forces and steric constraints –interactions that govern stability and specificity in molecular interactions and are important for predicting biologically and functionally relevant conformation are not considered- ability to model physical interactions is crucial for drug discovery
- Deep learning models can not generalize beyond their training data set and can overfit to statistical correlation and can lead to incorrect conclusions regarding biological activity
- Researchers have shown that co-folding models largely memorize ligands from their training data and do not generalize well to unseen ligand structures+.

*Yu, Y., Lu, S., Gao, Z., Zheng, H. & Ke, G. Do deep learning models really outperform traditional approaches in molecular docking? *International Conference on Learned Representations, MLDD Workshop* (2023).
Masters, M., Mahmoud, A. H. & Lill, M. A. Pocketnet: ligand-guided pocket prediction for blind docking. *International Conference on Learned Representations, MLDD Workshop* (2023).
+Škrinjar, P., Eberhardt, J., Durairaj, J. & Schwede, T. Have protein-ligand co-folding methods moved beyond memorisation? *bioRxiv*, https://doi.org/10.1101/2025.02.03.636309 (2025)

FINAL CONCLUSIONS

“Our findings underscore the models’ (i.e. AF2, RF) limitations in generalizing effectively across diverse protein-ligand structures and highlight the necessity of integrating robust physical and chemical priors in the development of such predictive tools. The results advocate a measured reliance on deep-learning-based models for critical applications in drug discovery and protein engineering, where a deep understanding of the underlying physical and chemical properties is crucial.”

- Investigating whether deep learning models for co-folding learn the physics of protein-ligand interactions, Matthew R. Masters, Amr H. Mahmoud Markus A. Lill Nature Communications volume 16, Article number:8854(2025)

AI Does Not Make It Easy

IN THE PIPELINE:DRUG DEVELOPMENT , 18 OCT 2024, BY DEREK LOWE

https://www.science.org/content/blog-post/ai-does-not-make-it-easy