

AlphaFold 3–enhanced graph neural network repurposing of FDA-approved drugs against neglected tropical disease proteins

Daniel Korkin

Massachusetts Academy of Math & Science, Worcester, MA 01605

INTRODUCTION & BACKGROUND

Neglected tropical diseases (NTD):

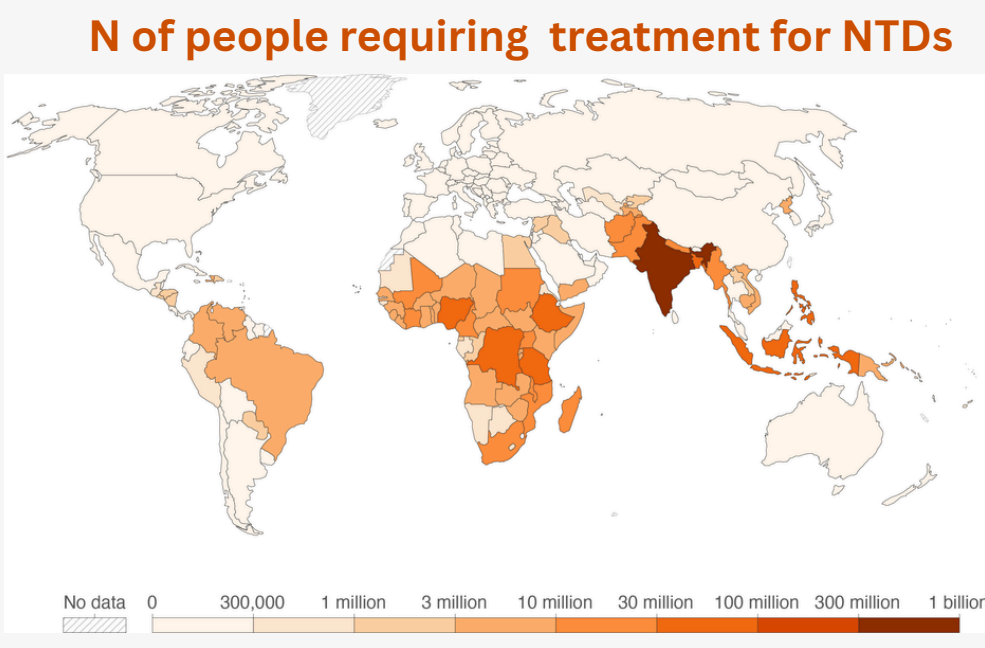
- Infections caused by various pathogens and transmitted by vectors
- Mainly affect **poor populations** in tropical and subtropical regions
- Limited** attention & funding by governments and companies

Every year due to NTDs:

- 1.5B** people require prevention or treatment
- 14.1M** years of healthy life is lost
- 120K** people die

Common NTDs:

- Toxoplasmosis
- River Blindness (Onchocerciasis)
- Snail Fever (Schistosomiasis)
- Malaria is considered together with NTDs



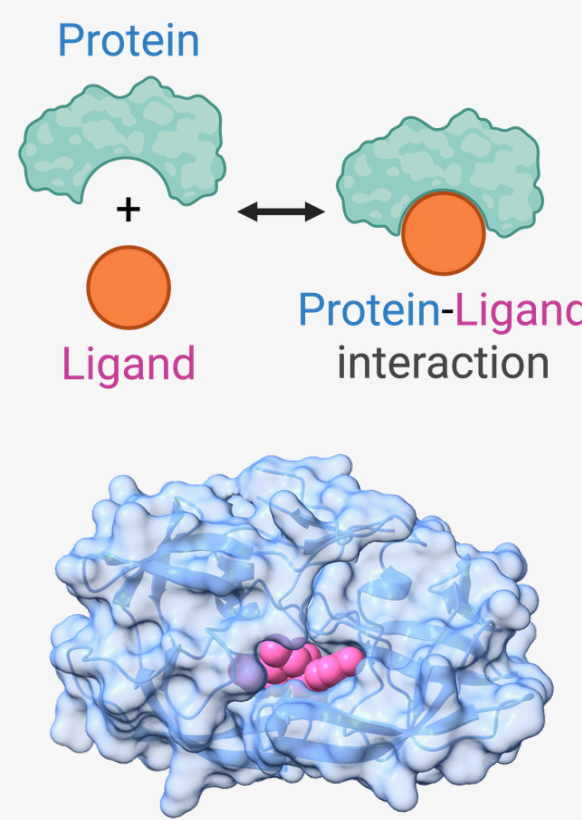
Project Goal: Develop a scalable AI/ML-driven platform that identifies **FDA-approved drugs** that can be repurposed to treat **NTDs**, accelerating drug discovery for populations overlooked by traditional drug development

How drugs work:

- Ligands** (drugs or bioactive compounds) bind to specific **targets** (usually proteins) and **change** how these molecules normally function

- Binding Affinity:** how strongly a drug binds to its target:

- Is often measured with K_d (dissociation constant)
 - Lower K_d = **stronger** binding (higher affinity)
 - Higher K_d = **weaker** binding (lower affinity)
- pK_d is the logarithmic form of K_d , $pK_d = -\log(K_d)$
 - Higher pK_d = **stronger** binding
 - Lower pK_d = **weaker** binding



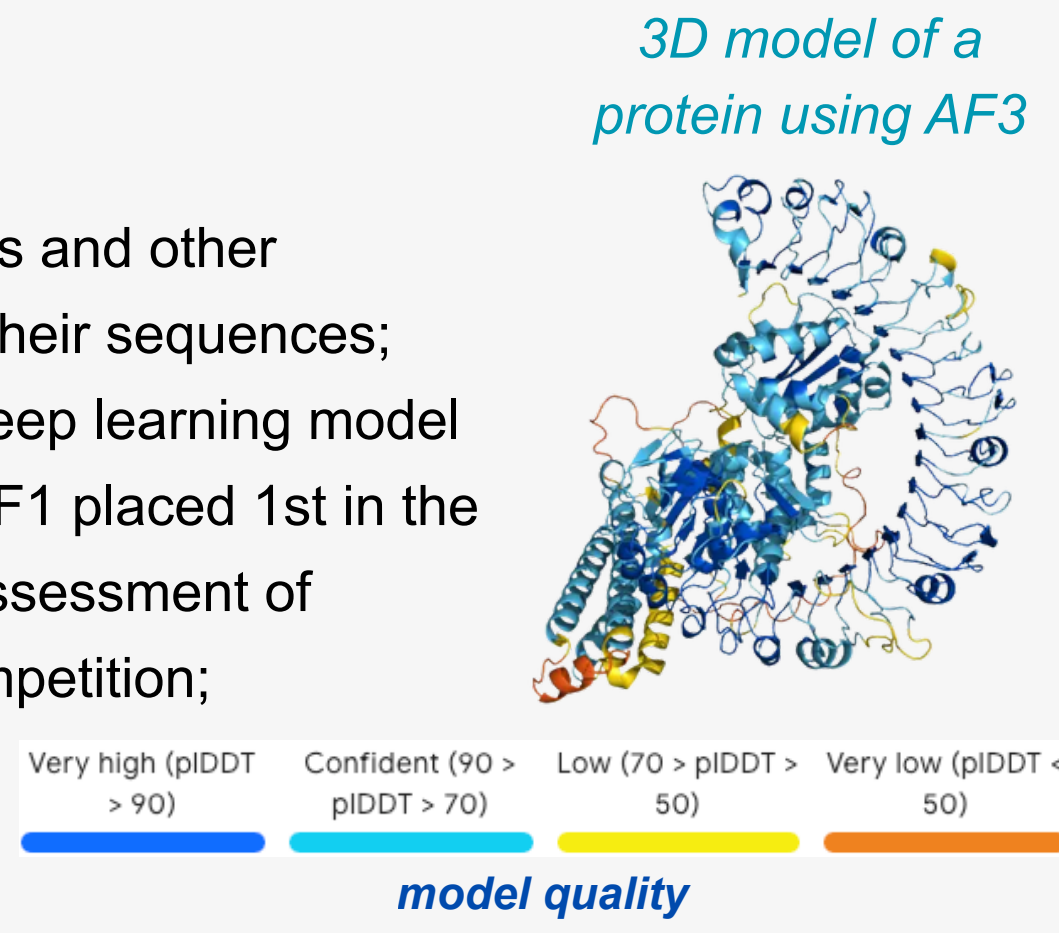
pK_d	≤ 5.0	6.0	7.0	8.0	9.0
Binding	very weak	moderate	strong	very strong	extreme
Is it a drug candidate?	Early screening hits	Reasonable starting lead	Good lead quality	Drug-like molecule	Very tight binder

Progress of AI/ML methods in life sciences and drug discovery

Deep Learning (DL) made a breakthrough in 3D protein modeling in late 2010s

AlphaFold (AF)

- AI model by Google DeepMind;
- Predicts 3D structures of proteins and other biological molecules given only their sequences;
- Uses attention neural network deep learning model
- When first introduced in 2018, AF1 placed 1st in the overall rankings of the Critical Assessment of Structure Prediction (CASP) competition;



- Scientists of AlphaFold and RosettaFold (another protein prediction method) won the Chemistry Nobel Prize in 2024

Graph Neural Networks (GNNs)

- Most recent advancement of deep learning in biological science
- 3D structure of proteins is directly encoded into graph structure of GNNs
- Allows to incorporate various biological features of macromolecules
- Current applications: protein function prediction; protein-protein interaction prediction; and protein-ligand interaction prediction and drug discovery

We can leverage DL approaches for our problem:

- 3D structural model can be obtained for any protein through DL predictions;
- GNN-based methods are promising for predicting protein-drug interactions.

MOTIVATION / SIGNIFICANCE / GOALS

Knowledge Gap: NTDs are heavily underfunded and understudied

- Funding in developed countries is focused on non-communicable, chronic, and prioritized infectious diseases

- Cancer research** yearly funding is ~\$5B

- Research for **schistosomiasis** (one of the

- biggest NTDs) annual funding is only ~\$14M

- For many tropical infectious diseases, there

- is **little or no precedent** for developing

- small-molecule drugs**

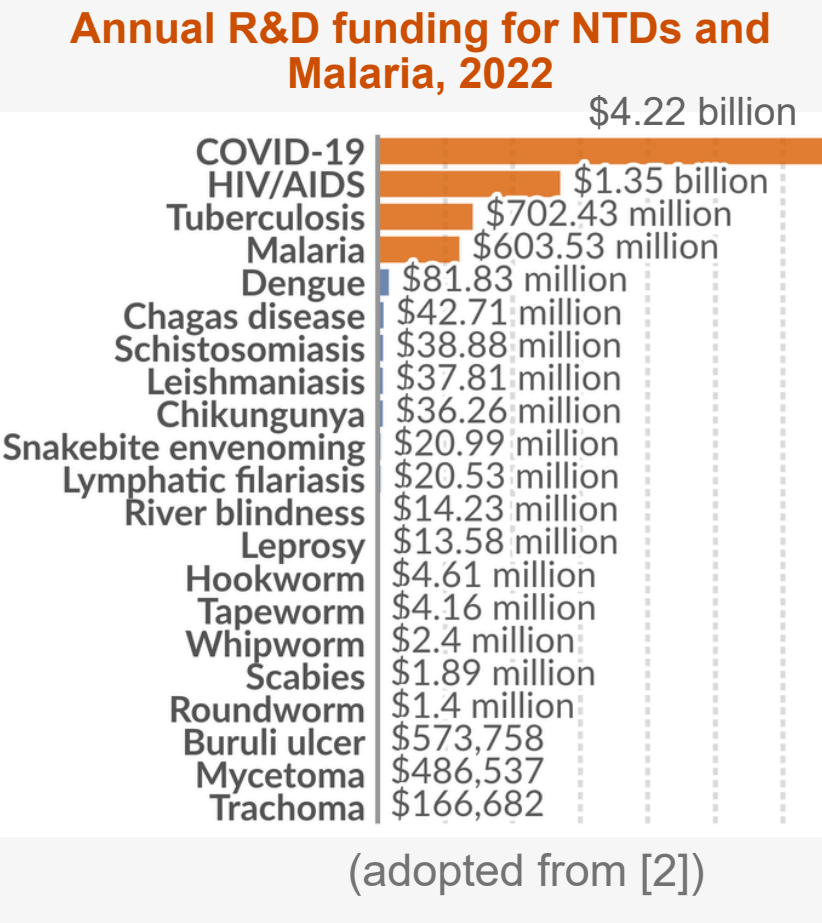
- This problem is made worse by our **limited**

- understanding** of the pathogen's biology and

- lack of well-defined **protein targets**

- There is an **urgent need** for new, safe,

- low-cost and efficacious treatments

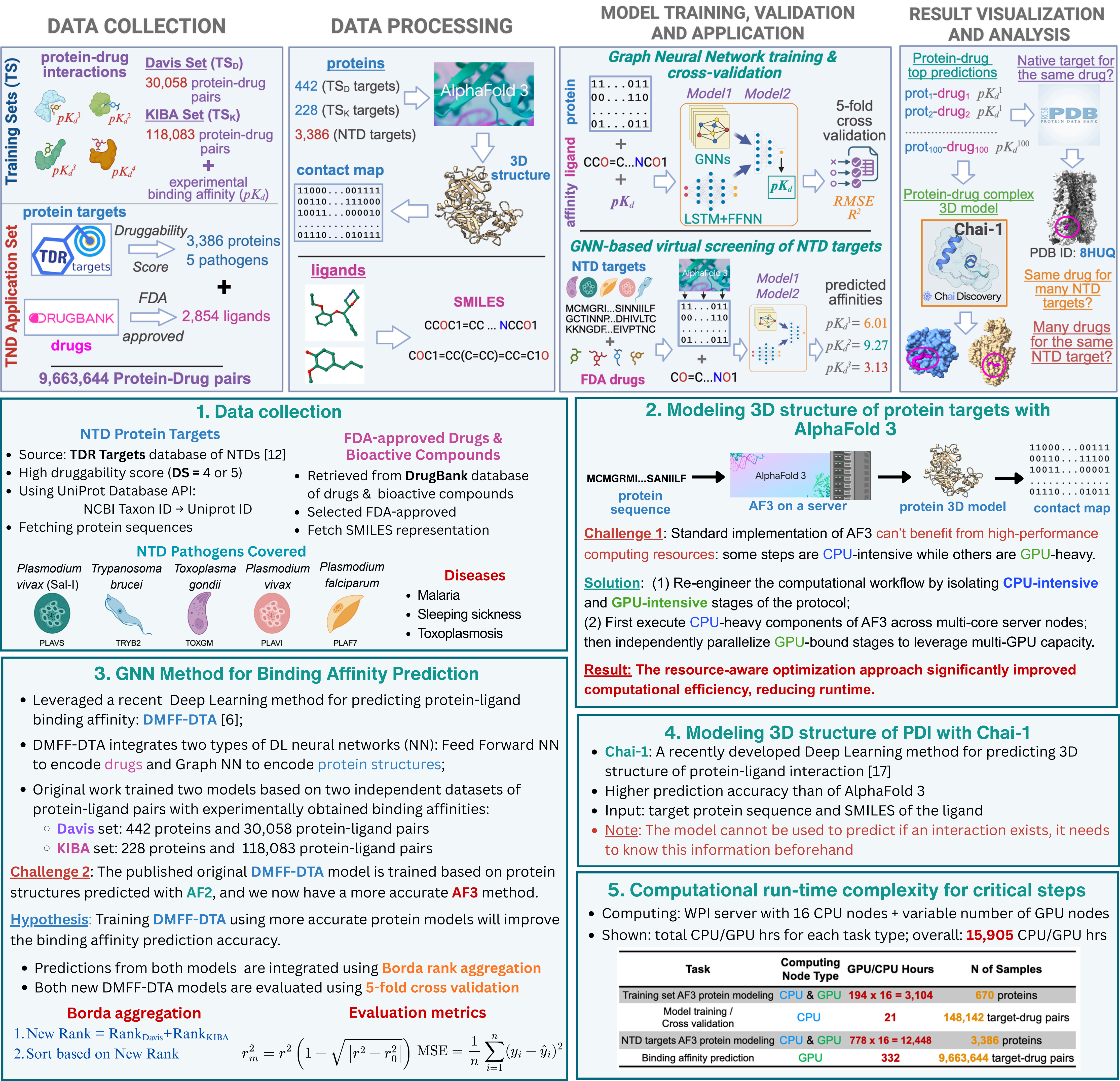


Specific Goal 1: Improve state-of-the-art GNN-based deep learning models for binding affinity by training them on an accurate set of protein 3D structures generated with AlphaFold 3 (AF3).

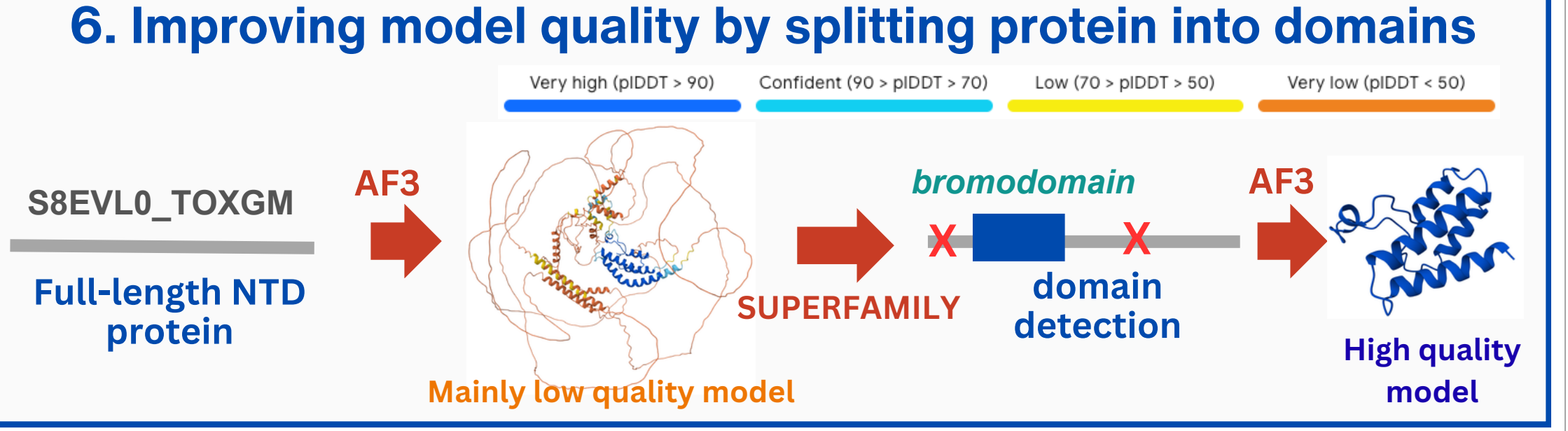
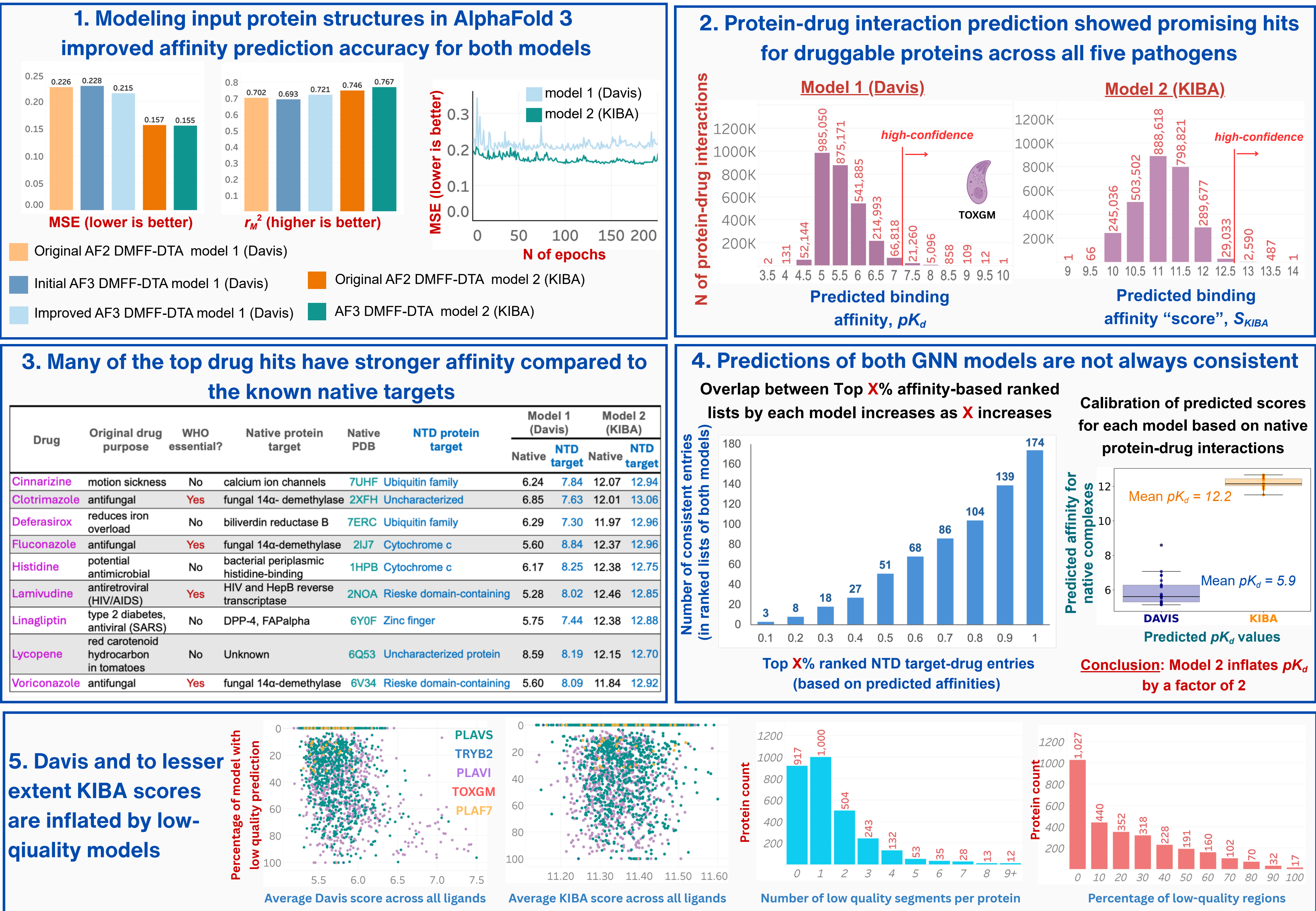
Hypothesis: Training on improved AF3-generated protein structures will increase the prediction accuracy of the GNN-based models.

Specific Goal 2: Using the improved deep learning models, develop a computational pipeline for virtual screening of drugs against protein targets. Apply this pipeline to generate a comprehensive list of FDA-approved drugs and their putative novel NTD protein targets.

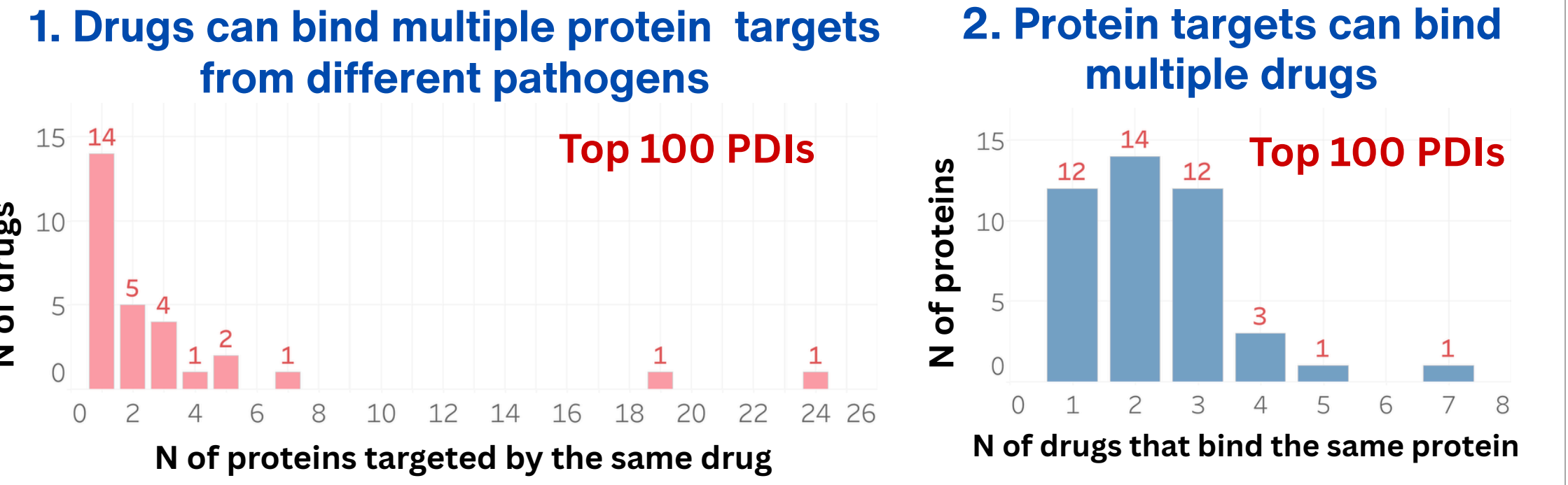
METHODS

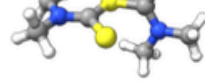
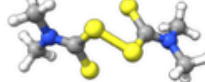
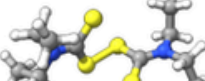
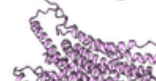
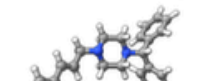
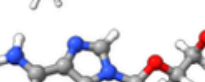
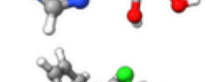
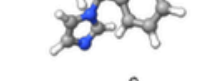


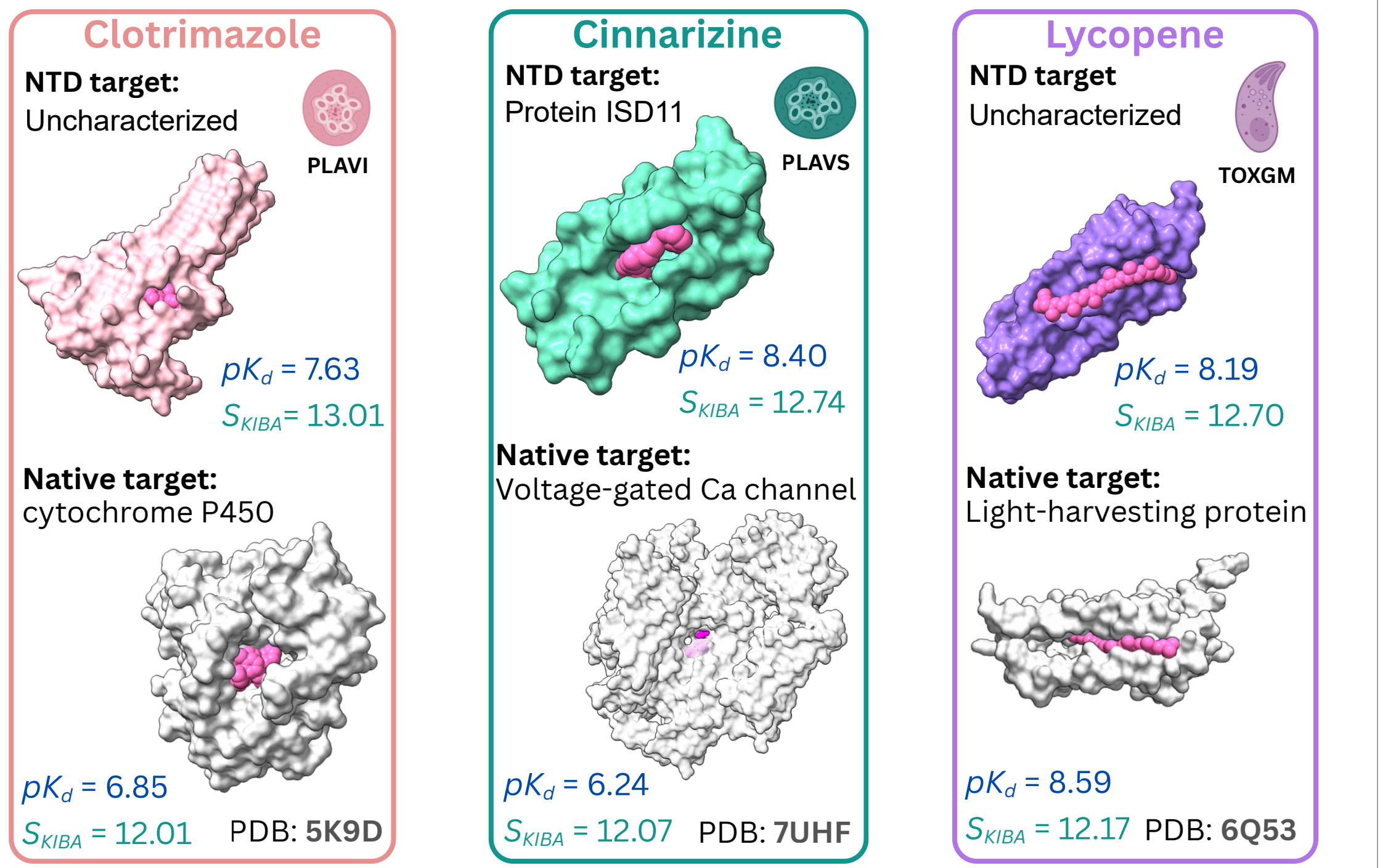
RESULTS / VALIDATION / ANALYSIS



APPLICATION



Most frequent drugs in Top 100 list				Most frequent NTD targets in Top 100 list					
Drug Name	Drug 3D structure	Targets in Top 100 list	Species in Top 100 list	Target Uniprot ID	Protein name	Protein 3D structure	Species	Drugs	
Tetramethylthiuram monosulfide		24	2	A0A125YL10_TOXGM	Rieske domain-containing protein		Toxoplasma gondii	7	
Thiram		19	2	S8GEB4_TOXGM	Tyrosine kinase-like protein		Toxoplasma gondii	5	
Disulfiram		7	2	A0A125YR49_TOXGM	H(+)-exporting diphosphatase		Toxoplasma gondii	4	
Cinnarizine		5	2	S8G8U4_TOXGM	Zinc finger motif-containing protein		Toxoplasma gondii	4	
Adenosine		5	1	S8G760_TOXGM	Cyclin2 related protein		Toxoplasma gondii	4	
Clotrimazole		4	2	ASKAE8_PLAVS	Cerine/threonine protein kinase		Plasmodium vivax (Sal-I)	3	
Fluconazole		3	1						



CONCLUSIONS

- Developed an automated computational framework for virtual screening of drugs against proteins based on the protein sequences
- Improved a recently published **GNN** model **DMFF-DTA** by training it on the new dataset of protein 3D structures generated with **AF3**
- Integrated results of **two DMFF-DTA** models trained on independent experimental sets using Borda ranking aggregation
- Applied this framework to a large-scale analysis of **5** pathogenic species causing deadly Neglected Tropical Diseases
- Framework automatically screened **9,663,644** potential protein-drug interactions between **3,386** pathogenic proteins annotated as 'druggable' by research community and **2,854** FDA-approved bioactive compounds
- Many compounds in the top 100 ranked pairs are **anti-fungal** drugs listed as WHO essential medicine
- Some of the identified drugs, such as **thiram** and **linagliptin**, have shown to exhibit antibacterial and antiviral properties
- An exciting discovery is **lycopen**, a natural compound found in tomato, strawberry, and other fruits and berries with recent evidence of its antimicrobial activity

IMPACT: This work offers a cost-effective and scalable alternative to traditional drug discovery, accelerating the identification of drug candidates for NTD that disproportionately affect vulnerable populations worldwide.

FUTURE RESEARCH

- Model Improvement:** The current DMFF-DTA can potentially be further improved through training on the larger available datasets of known ligand binding affinities
- Reducing false-positive rate:** Potential source of false-positive hits: incorrectly modeled protein, specifically for large proteins or proteins with unstructured regions
Solution: instead of modeling 3D structure of entire protein, model the structures of its individual **protein domains** (structurally compact and independent units of proteins)
- In Silico Confirmation:** How does one know that a ligand actually inhibits a function?
 - Find if this ligand targets a functional site of the protein
 - This site may be a protein-, DNA-, or RNA-binding site, etc.
- Experimental Validation:** Selected drugs should be experimentally tested, first in culture and then in animal models

REFERENCES

- [1] Branda F, et al. Assessing the burden of neglected tropical diseases in low-income communities: challenges and solutions. *Viruses*. 2024;17(1)
- [2] Annual research & development funding for infectious diseases, 2023. Our World in Data. 2024
- [3] Aguiere F, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nature Rev. Drug Disc.* 2008;7
- [4] He H, et al. Dual modality feature fused neural network integrating binding site information for drug target affinity prediction. *NPU Digital Medicine*. 2025;8(1)
- [5] Chai Discovery team, Boltraud J, et al. Chai-1: Decoding the molecular interactions of life. *BioRxiv*. 2024; 2024-10
- [6] Tang C, et al. Lycopen attenuates Staphylococcus aureus-induced inflammation via inhibiting α -hemolysin expression. *Microbes and Infection*. 2021;23(9-10):104853