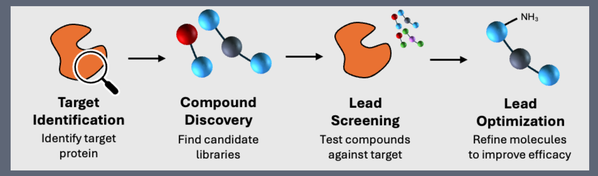
Unleashing Human-AI Creativity in Drug Discovery: Joint Embedding of Molecules, Proteins, and Language

Saathvik Kannan
MIT Computational Biology Group (PI: Manolis Kellis)

Acknowledgements: We thank Vinay Singamsetty for his work on integrating AffinityNet onto Mantis

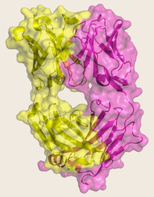
Why drug discovery needs a new map?



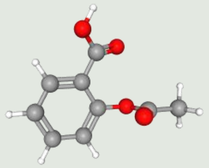
- **1 in 30,000**: Success rate
- **6 to 10 years** to develop
- **$985 million** spent/drug

Drug design today is complex and slow. We built an AI system that helps scientists explore, understand, and invent new drugs faster—by mapping molecules, proteins, and language into a shared landscape.

Our multimodal prediction model, **AffinityNet,** learns a shared embedding space for drugs, proteins, and patents. It enables accurate affinity prediction and powers other creative drug-discovery tasks.



Proteins



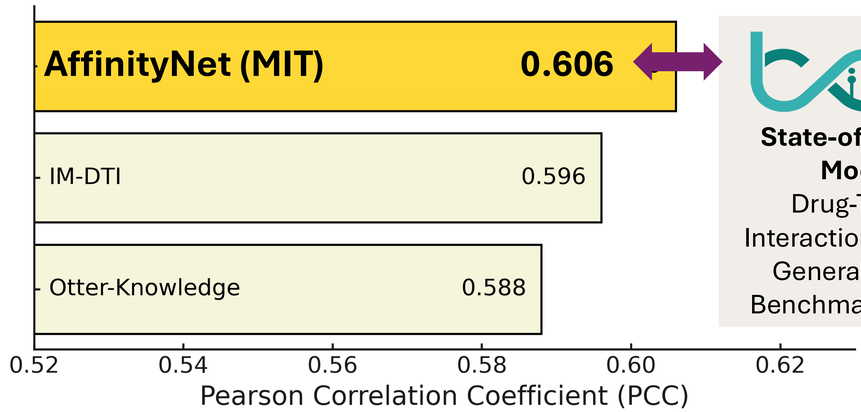
Molecules

MAAQDSAL Sequence

Domain Information

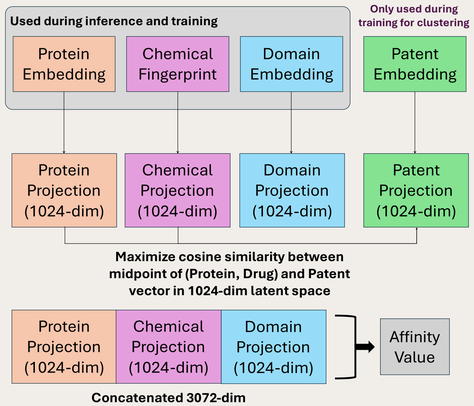
Patent Documents

Morgan Fingerprint

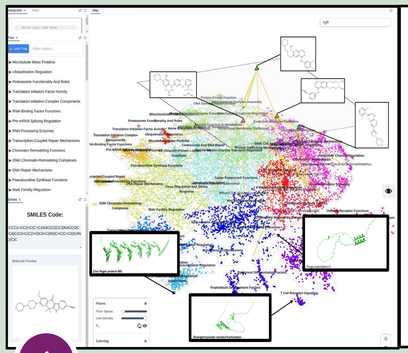


State-of-the-art Model
Drug-Target Interaction Domain Generalization Benchmark Group

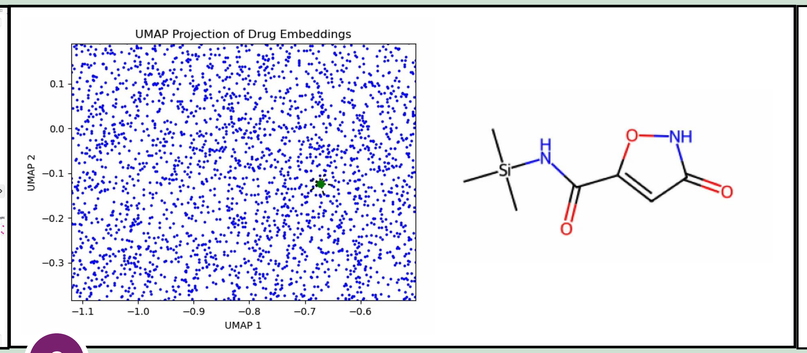
Joint embedding



Tools built on the map



1 **Navigate the chemical-protein space in real-time**



2 **AI fills the gaps humans haven't yet explored in chemical space**

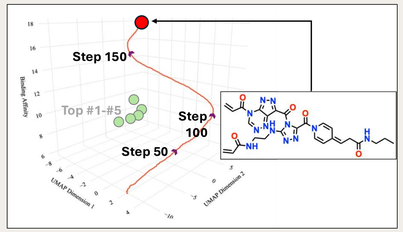
Under development

- Molecular Weight Optimization Agent
- Immunogenicity and PK/PD Optimization Agent
- Human-in-the-loop

3 **Iterative refinement through reinforcement learning and feedback**

Generating molecules for targets

1. Start with a protein target
   - Protein that causes a disease
2. Scan thousands of known drugs
   - AI checks which one is best. "finding correct key for the lock"
3. Use "gradient ascent"
   - AI tweaks the molecule. "Climbing uphill to find the best view"
4. Use "Outbeddings"
   - Turn the ideal shape back into real, druggable molecule.



Gradient ascent enables the generation of precise compounds for a specific protein target using the **AffinityNet** model.