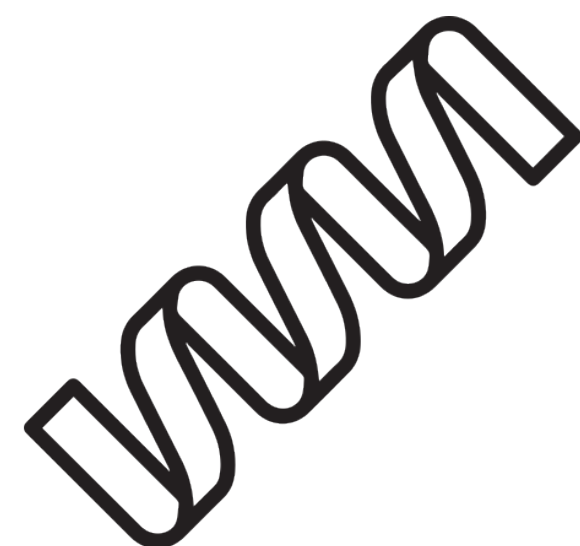
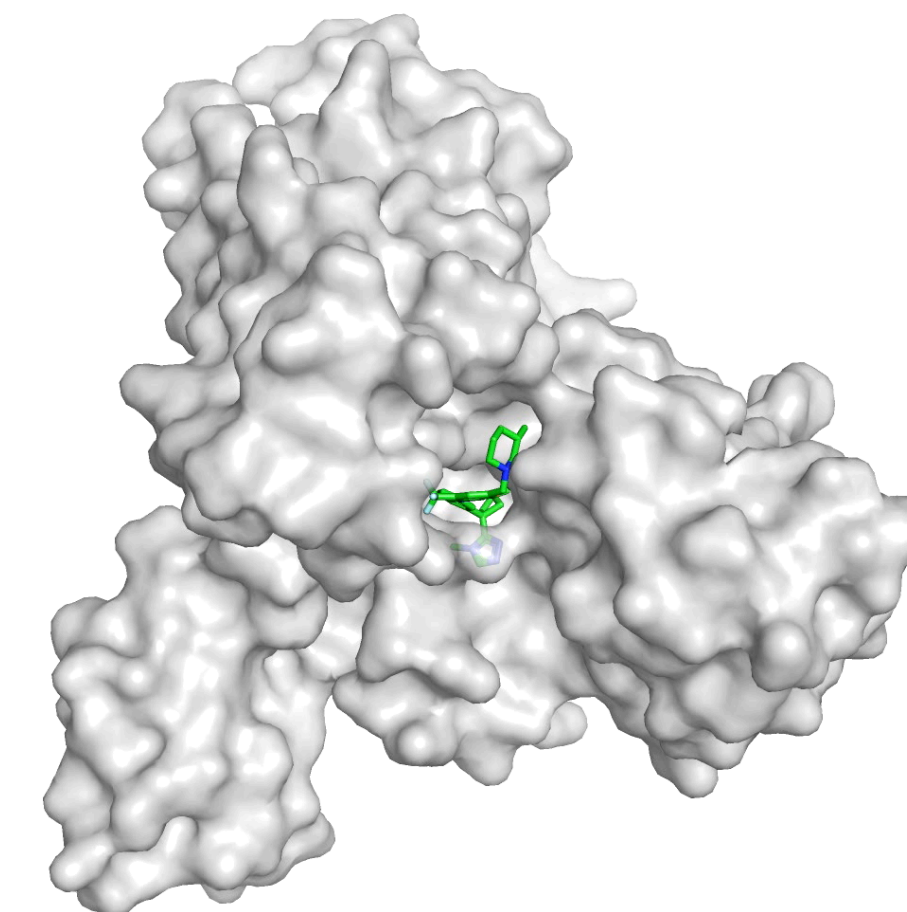
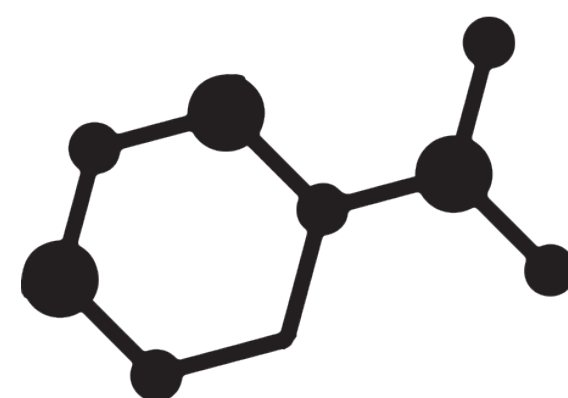


CACHE Challenge 4

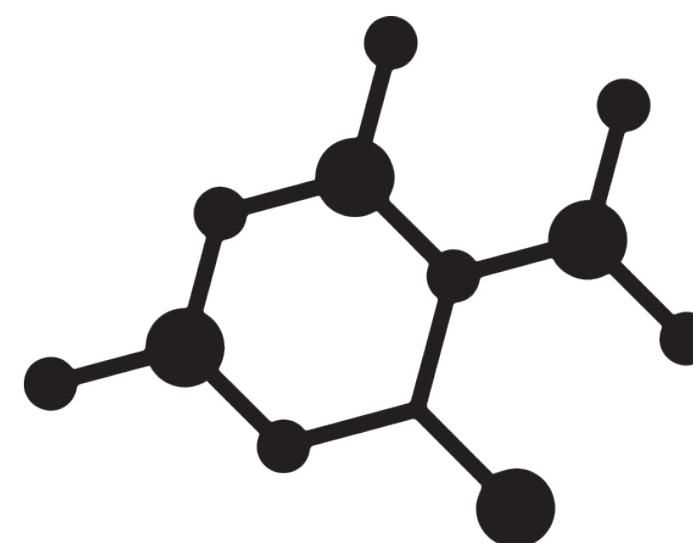
Discover new drugs for cancer immunotherapy



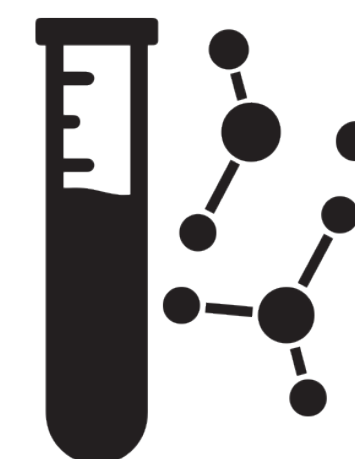
1. Identify a protein target



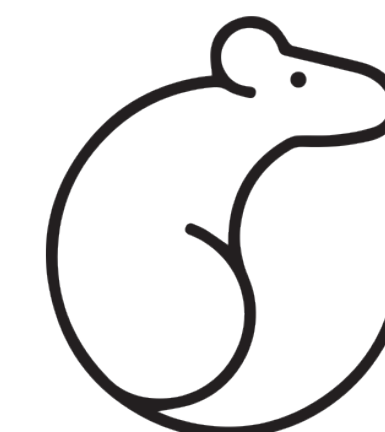
2. Find molecules that bind to target



3. Optimize molecules



4. Lab tests



5. Animal trials



6. Human trials

Akshay Sethia, Alice Cheng, Ferdie Krammer, Gbetondji Dovonon, Jude Wells

our approach

THE VIRTUAL SCREENING FUNNEL

Open source database of drug-like molecules (up to 30bn)

Initial screen based on mol properties
(weight, solubility, toxicity)

Fast ML scoring from molecule alone
(100 molecules per second)

Docking and scoring
(10s per molecule)

MMGBSA scoring
(60s per molecule)

Molecular dynamics
(5hr per molecule)

Experimental
validation

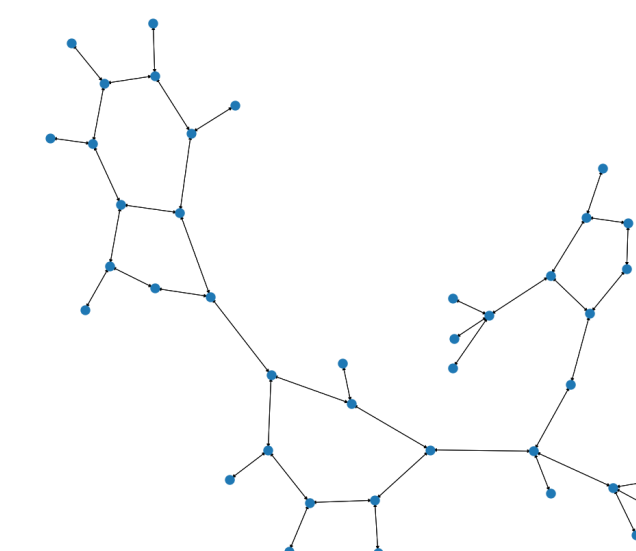
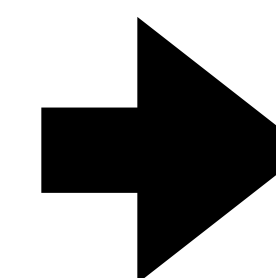
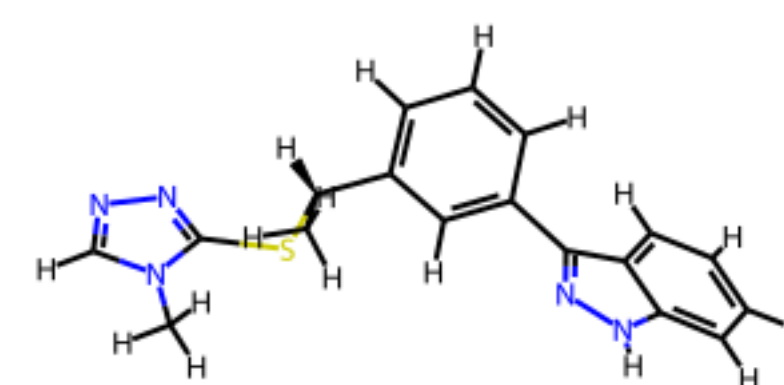
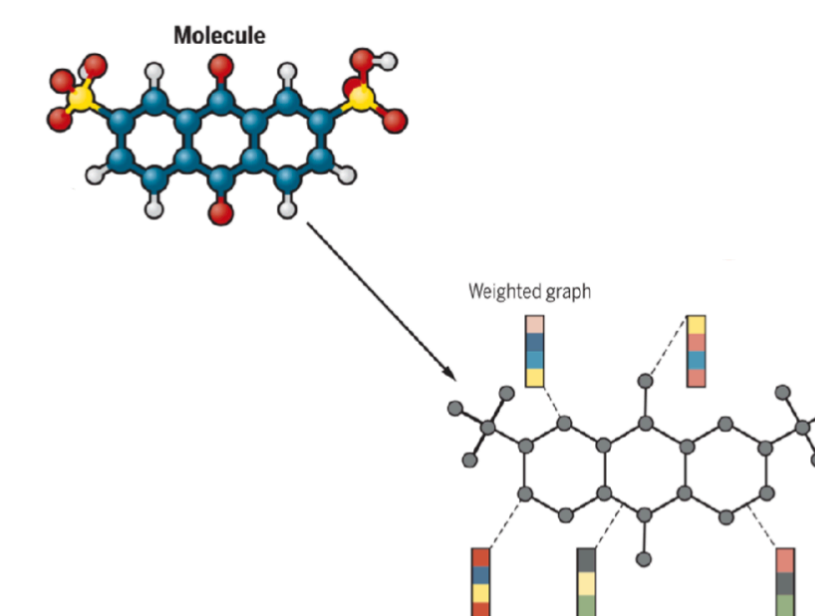
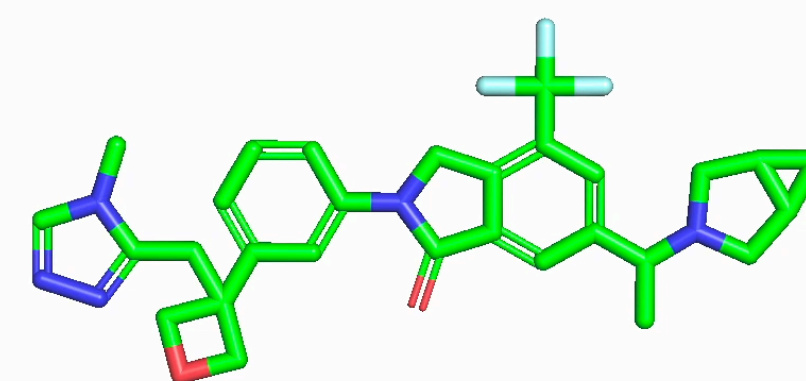
Optimize
hits

**Binary classification
experimentally observed
hits vs. decoy molecules**



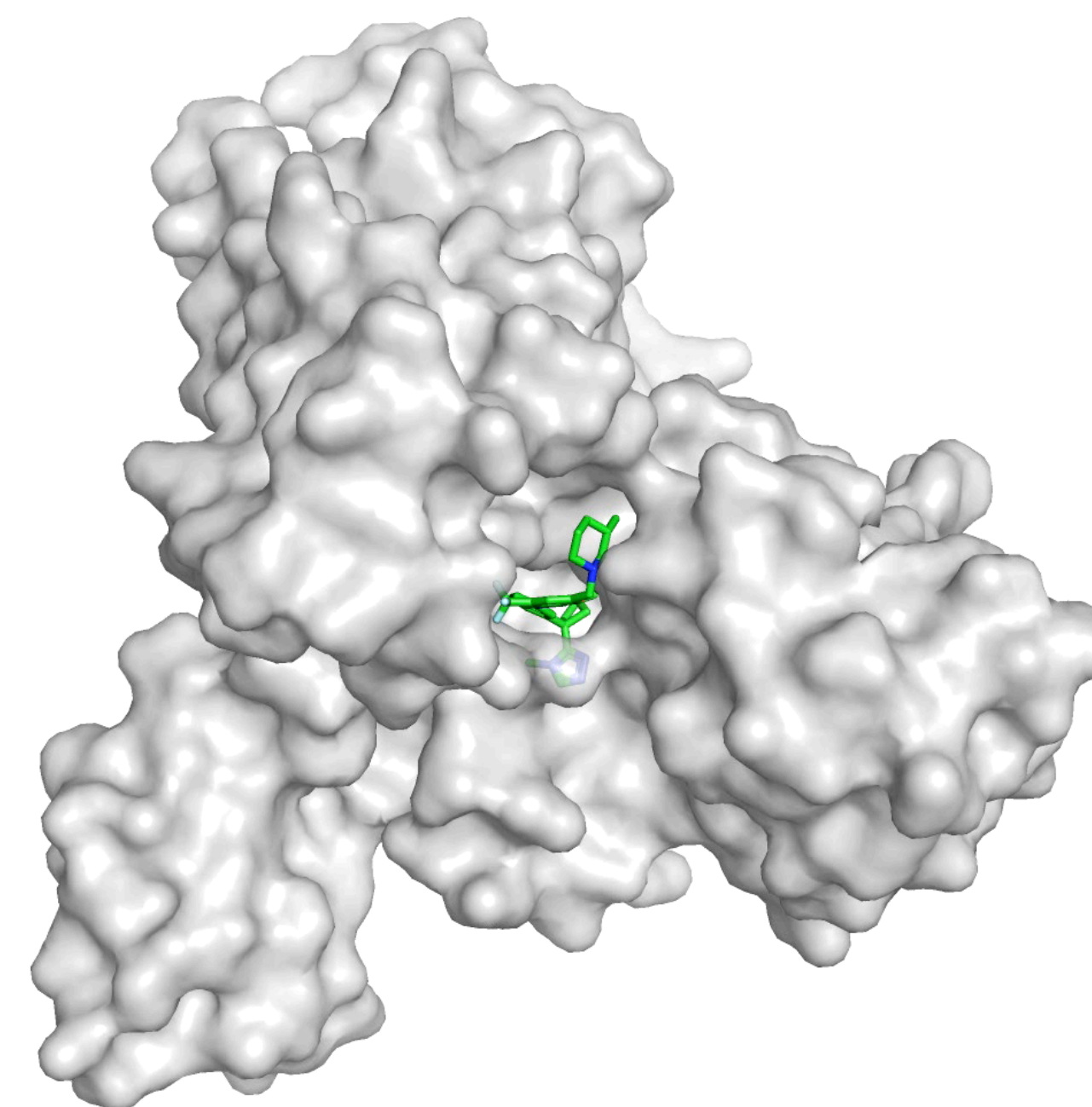
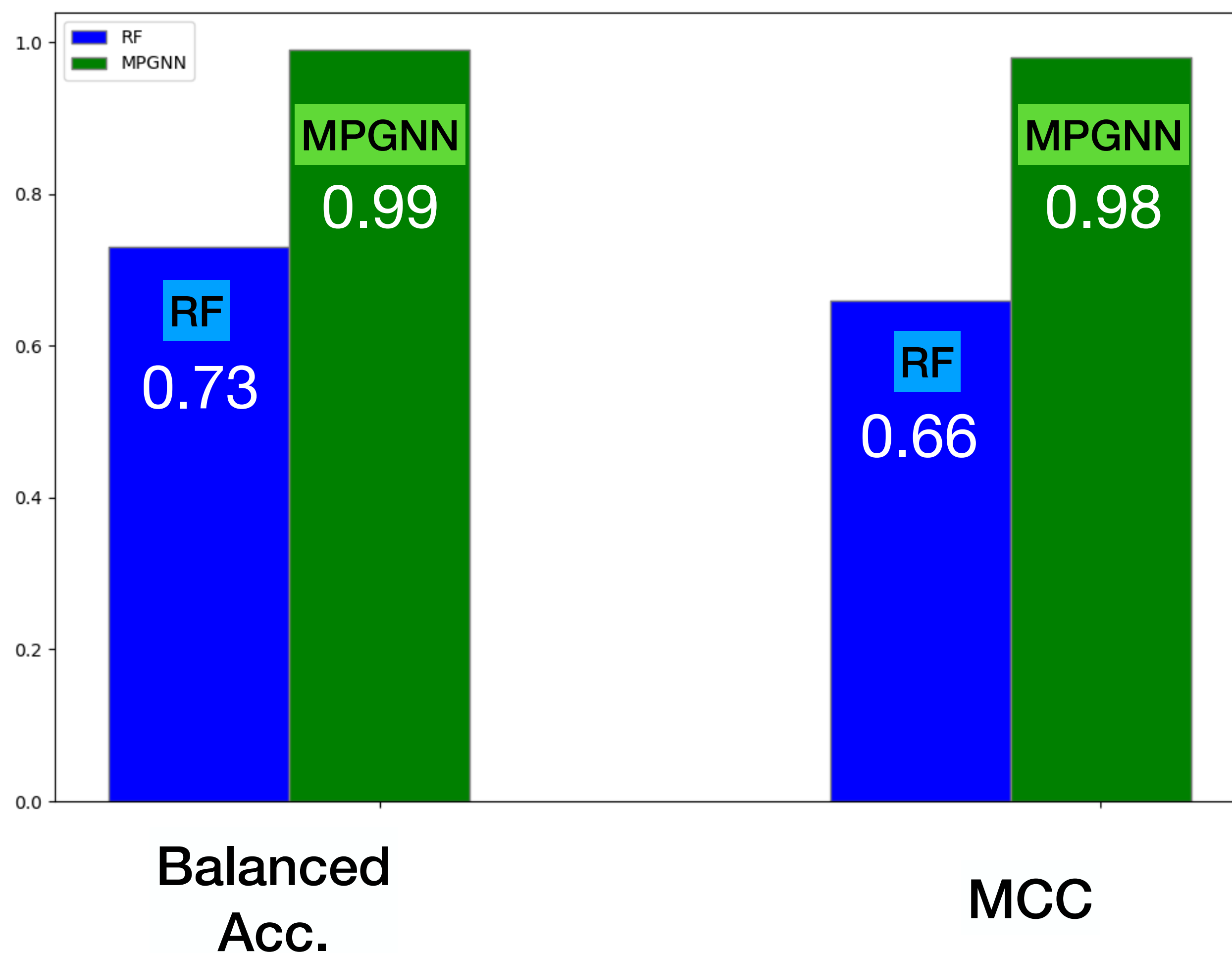
Molecules as graphs + GNNs

- Message passing graph neural networks
- Graph classification: one graph per molecule
- Atoms are nodes
- Chemical bonds are edges



results: message passing GNN vs RF

GNN doing suspiciously well



Future directions

- Develop custom graph embeddings for molecules
- Bayesian optimisation loop with ML + physics based scoring
- Transfer learning from other protein-ligand pairs