PROBLEM STATEMENT

Formal: Utilizing EquiBind for In-Silico Selection of Anti-Cancer Drug Candidates

Human: Using a deep learning model to rank 250,000 chemicals based on what sticks best to a human protein linked to cancer growth

MEET RAS

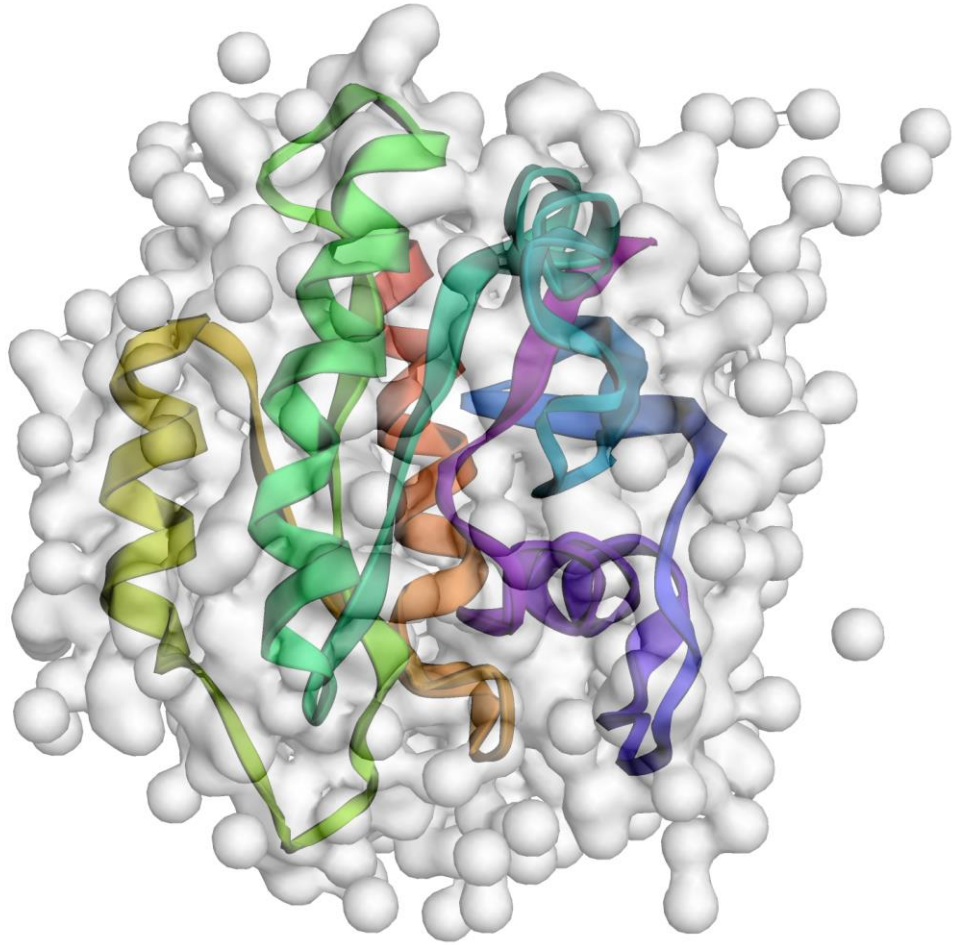
RAS protein: Crucial in cellular signaling pathways

Regulates cell growth, differentiation, and survival

Mutations in RAS genes: Disrupt cellular balance

RAS dysfunction: Implicated in various human cancers

Targeting RAS: Promising avenue for cancer treatment



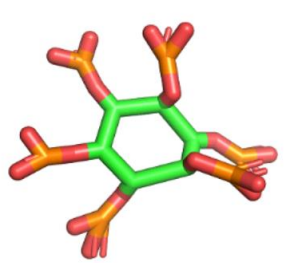
REQUIRED BACKGROUND INFO

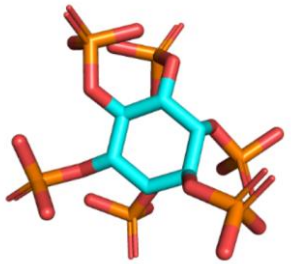
Ligand: A molecule that can bind to a specific site on a larger molecule, often a protein, and produce a biological effect.

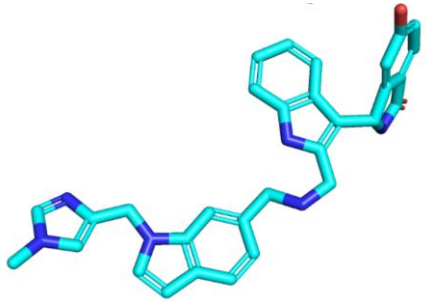
Torsion Angles: The angle of rotation around a covalent bond between atoms in a molecule.

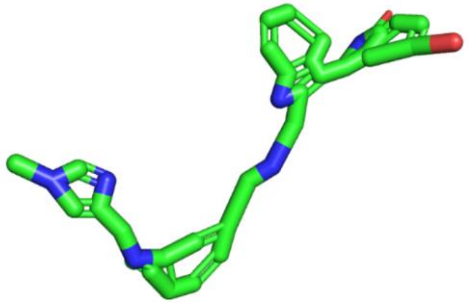
Conformation: The specific spatial arrangement ("shape") of atoms in a molecule that can be attained by rotation around its single bonds.

Complex: A group of two or more molecules or macromolecules that bind together to form a larger functional unit.

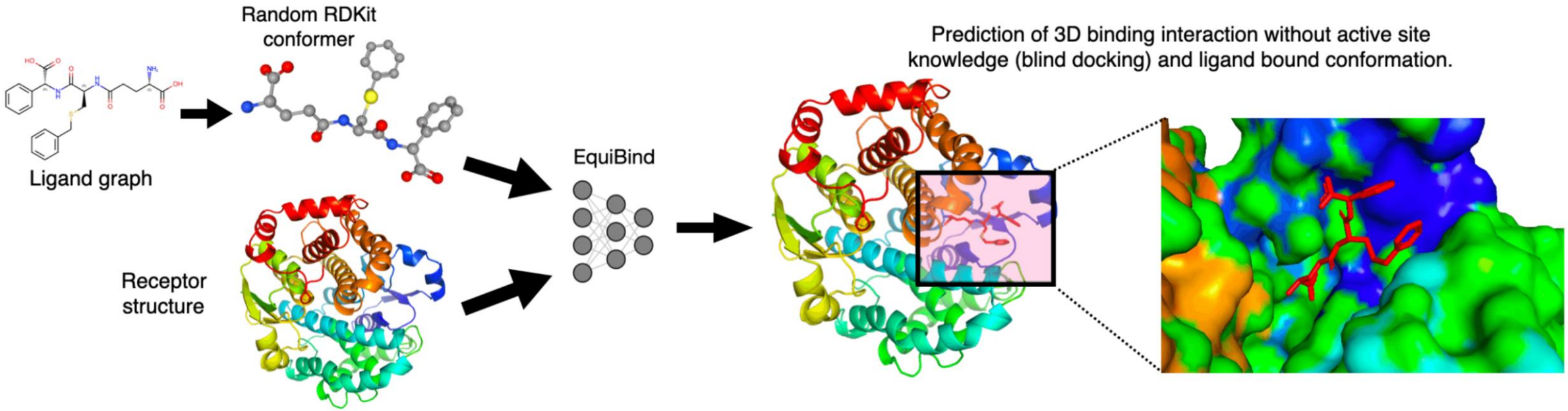




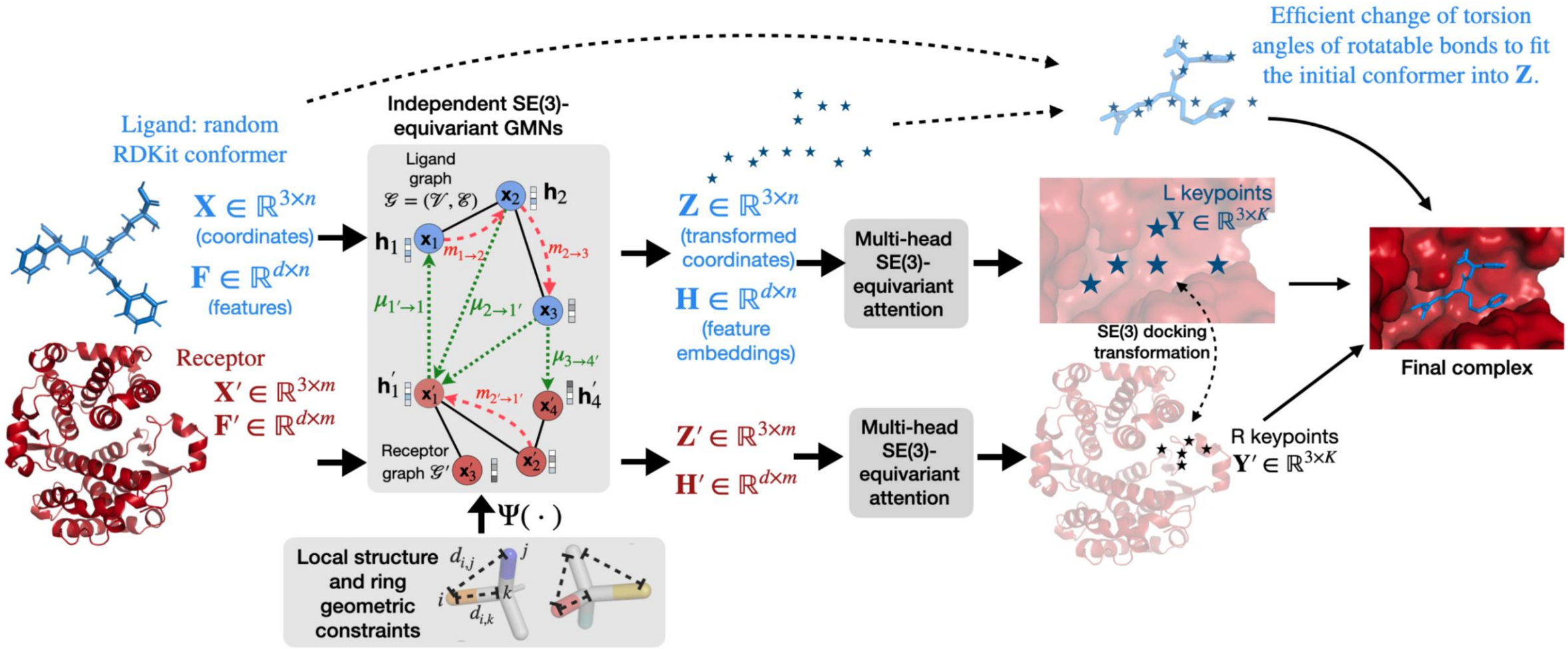




THE GOAL OF EQUIBIND



THE EQUIBIND ARCHITECTURE

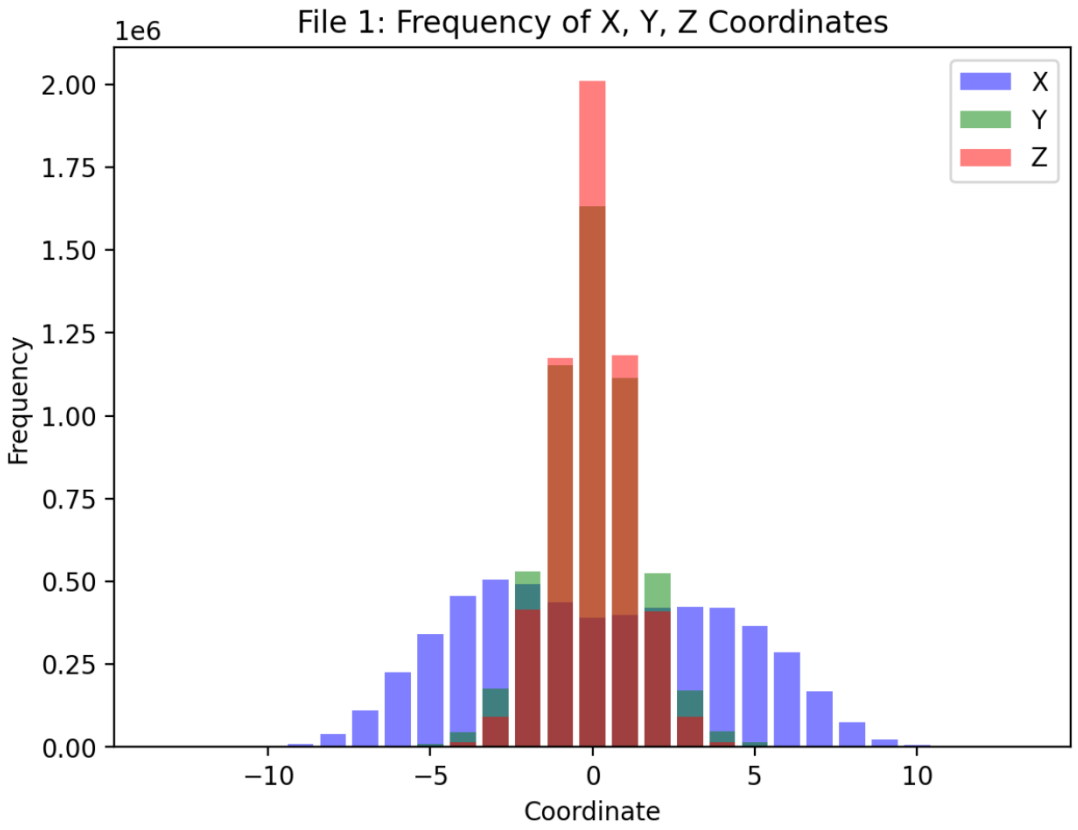


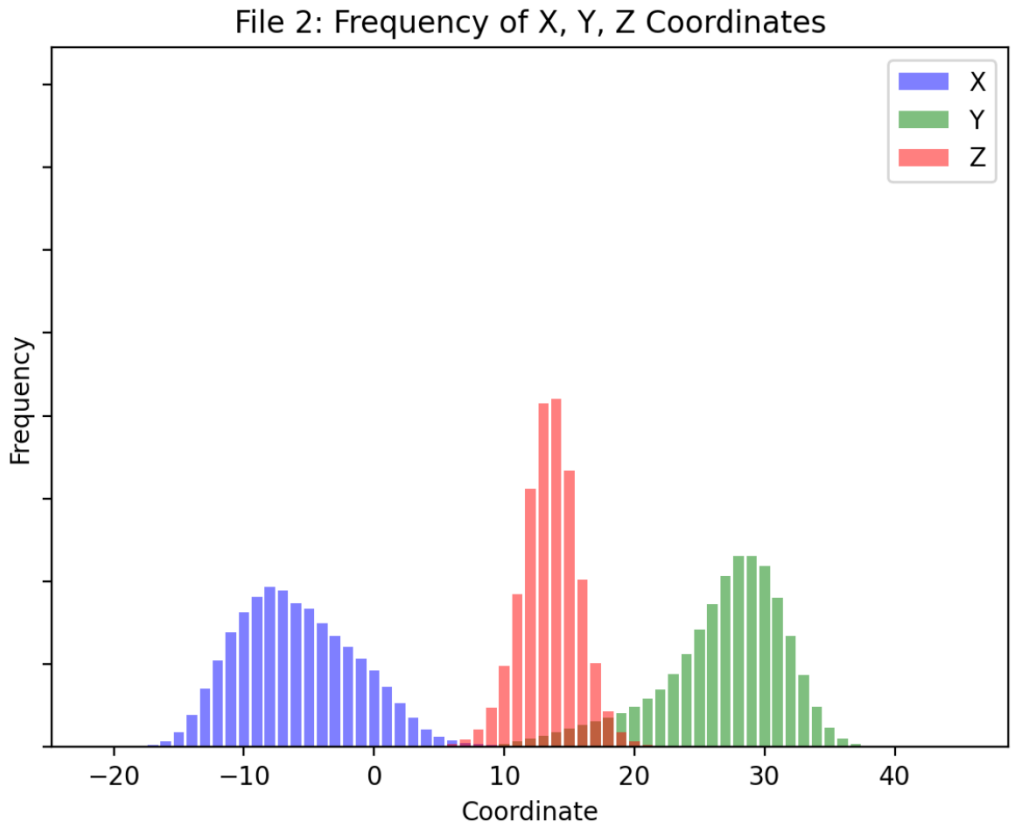
DATA: 5P21 AND THE ZINC250K DATASET

5P21: A well-known structure of the RAS protein, a family of proteins that play a critical role in cellular signaling pathways. The Protein Data Bank (PDB) file of 5P21 contains detailed information about the 3D structure of the RAS protein, including the spatial coordinates of all its atoms. Researchers can use this PDB file to study the RAS protein's structure and interactions with other molecules.

ZINC250k: A collection of over 250,000 unique, commercially available chemical compounds often used in drug discovery research. These compounds are represented using SMILES strings, which are a simple way to describe their chemical structures. Researchers can use the ZINC250k dataset to screen for potential drug candidates, focusing on molecules that may interact with specific biological targets to treat various diseases.

EQUIBIND'S OUTPUT



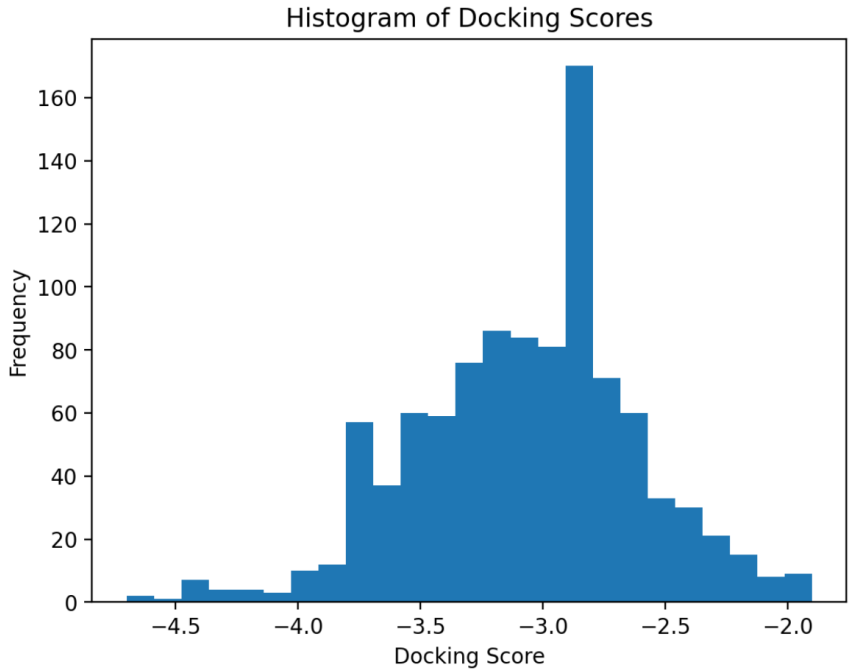


RANKING WITH SMINA

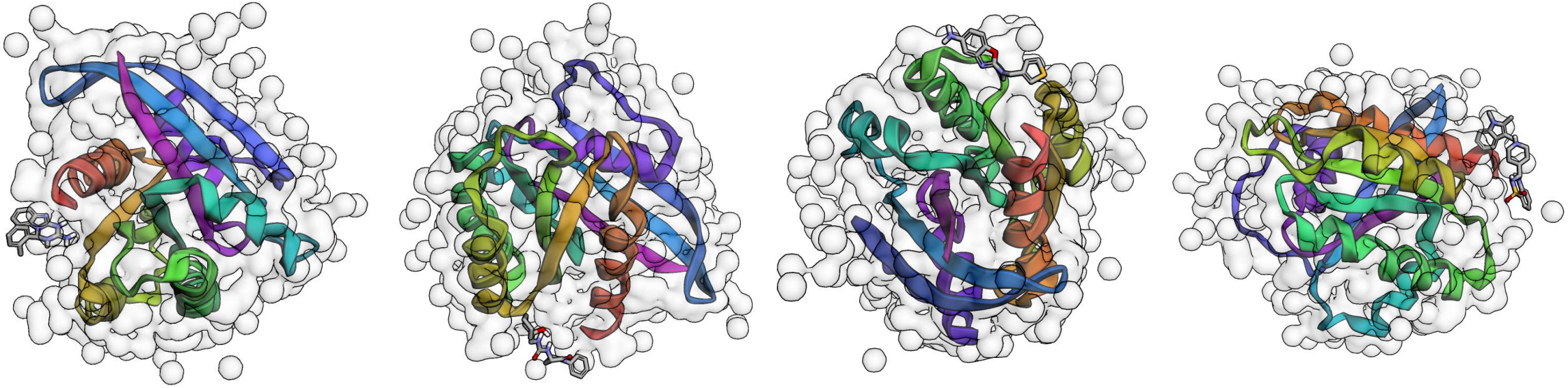
EquiBind is ultimately a prediction model. It will always take in a ligand and find a docking location and orientation on a protein. This means that it will even create unrealistic complexes. The model itself has no ability to score one ligand over another. Here is where SMINA enters.

SMINA is a powerful open-source molecular docking tool that simulates and predicts small molecule interactions with target proteins. It refines, scores, and ranks binding poses, with the lowest score indicating the highest affinity for the target protein.

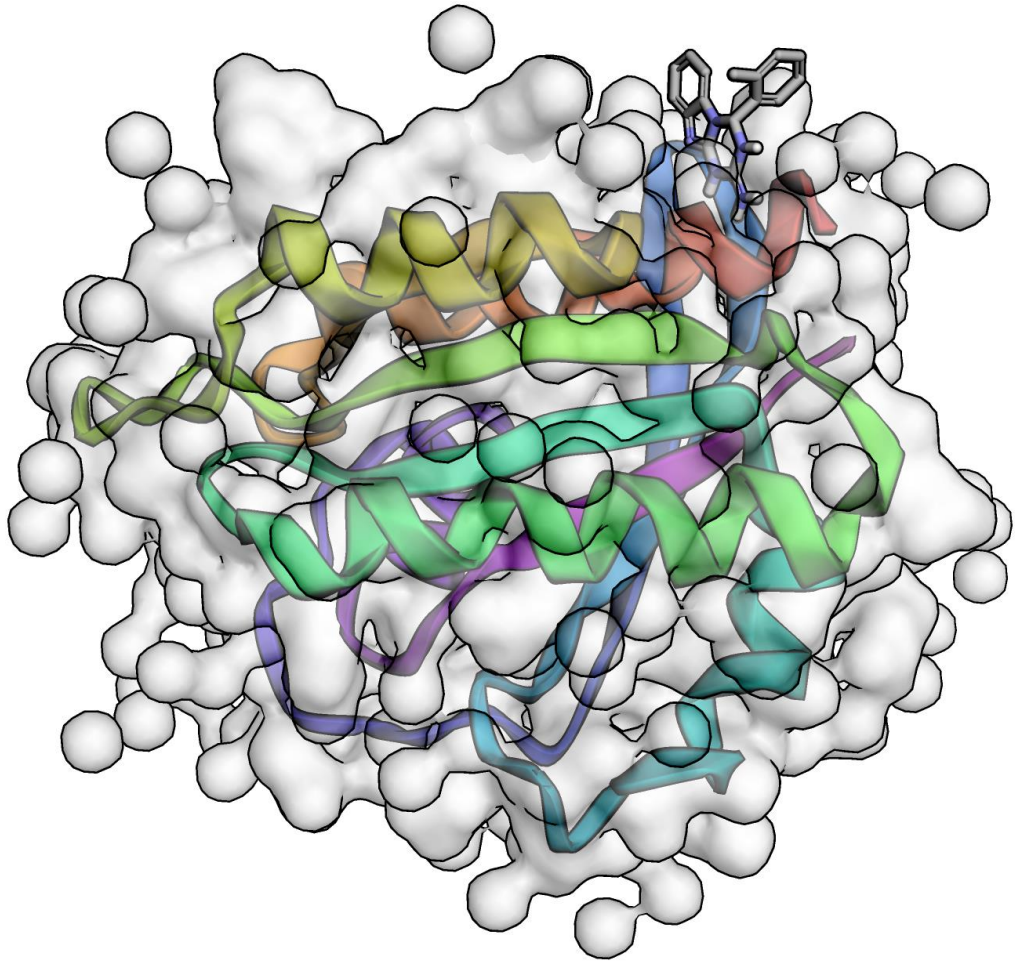
	Molecule index	Score
0	282	-4.7
1	844	-4.6
2	671	-4.5
3	27	-4.4
4	253	-4.4
5	425	-4.4
6	446	-4.4
7	505	-4.4
8	519	-4.4
9	745	-4.4
10	214	-4.3



EQUIBIND'S OUTPUT



RANKING WITH SMINA



THE PROS AND CONS

GOOD

Improved accuracy: Directly predicts binding locations and poses using SE(3)-equivariant deep learning.

Increased efficiency: Bypasses traditional candidate sampling, scoring, ranking, and fine-tuning steps.

Versatility: Applicable to a wide range of protein targets and ligand datasets.

BAD

EquiBind's predictions tend to be less accurate when the distances between ligand and protein positions are under 5 angstroms. To achieve the highest accuracy in these cases, it's recommended to fine-tune the results using specialized commercial software, such as Glide.

This additional step may add complexity and cost to the overall process.