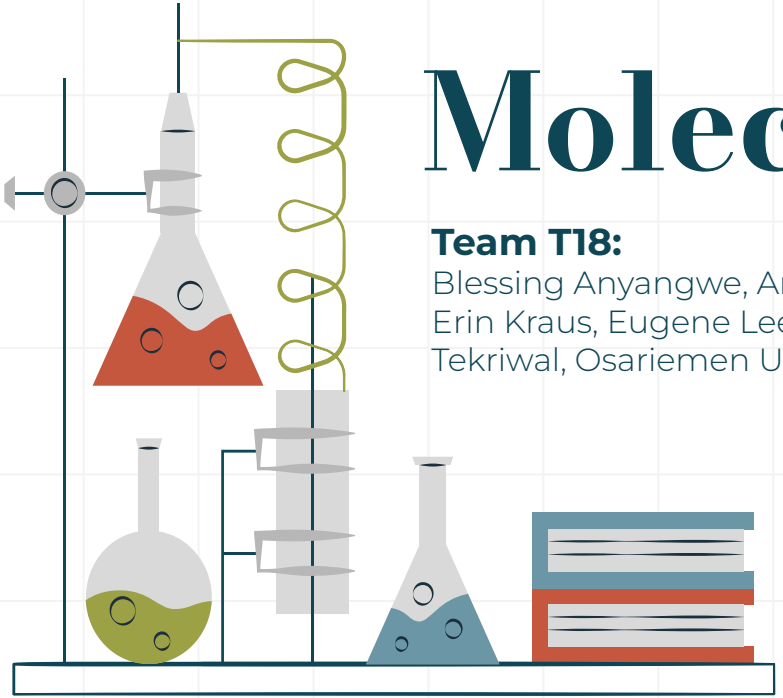
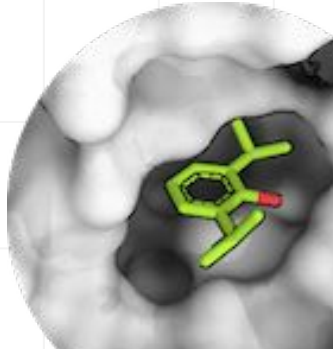


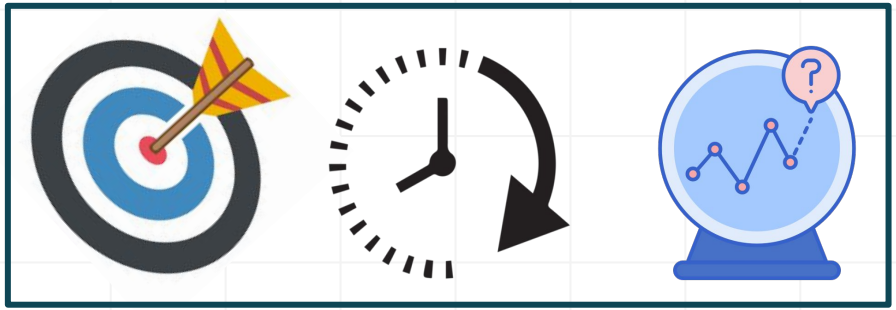
Molecular Docking

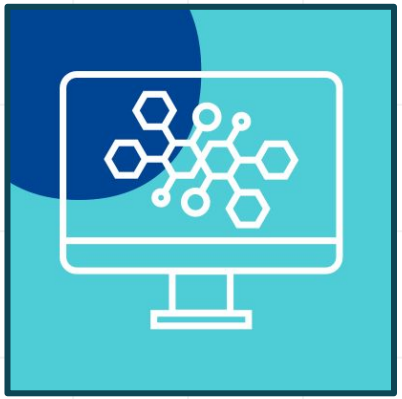
Team T18:
Blessing Anyangwe, Arushi Desai, Elizabeth Fishman, Kevin Jin, Kai Kim, Erin Kraus, Eugene Lee, Angelina Li, Bridget Liu, Nicholas Sardy, Aarna Tekriwal, Osariemen Unuigbe, Alexander Zatuchney, Eric Zhu



Our Purpose

Assess and compare the accuracy of molecular docking methods

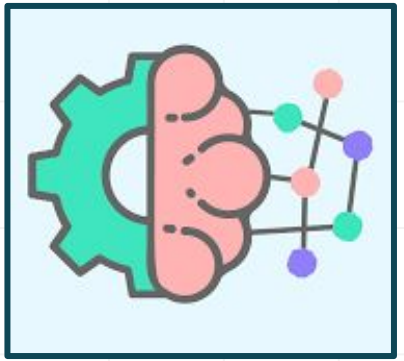




Traditional docking

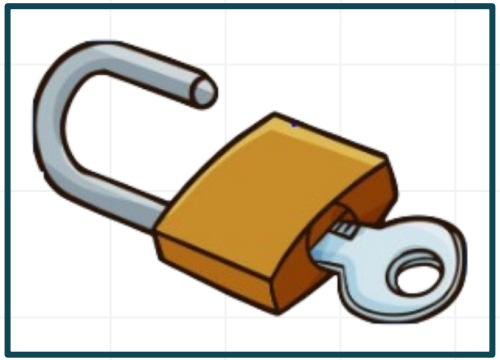
VS.

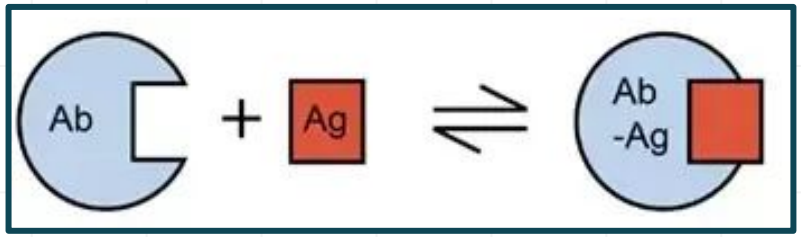
Deep Machine learning docking



Protein-Ligand Binding

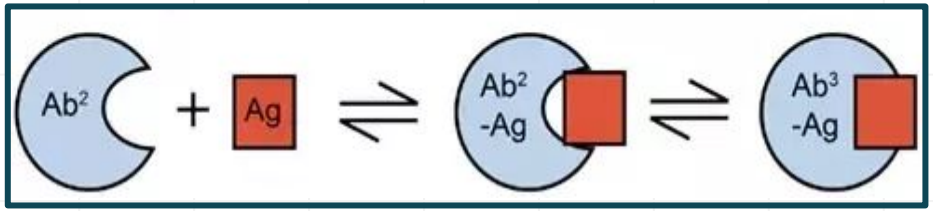
LOCK-AND-KEY





or

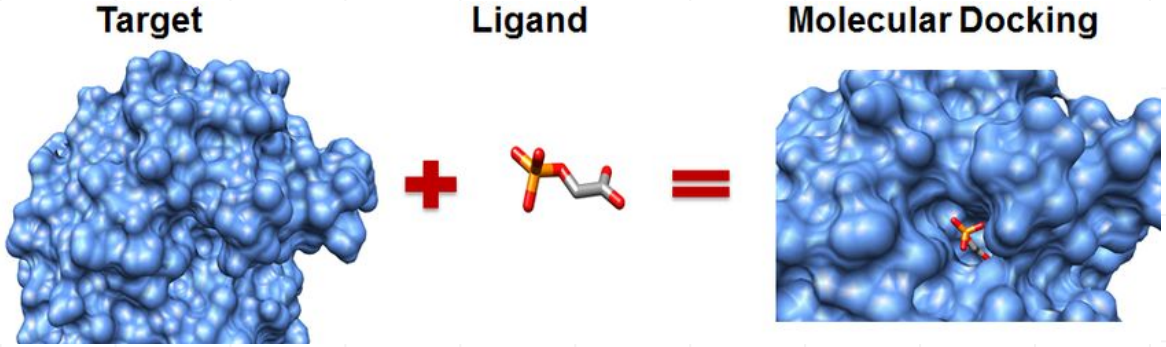
INDUCED FIT



Molecular Docking

Predicts **binding conformation** between **ligand** and **target protein**

- *In-silico* method
- Widespread applications in pharmaceutical fields
- Traditional programs vs. AI-based programs



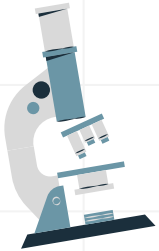
Traditional Molecular Docking Programs

2 main parts:

Search Algorithm

Generates protein-ligand binding poses

- Conformational search
- Fragmentation
- Monte Carlo
- Genetic Algorithm
- Tabu Search



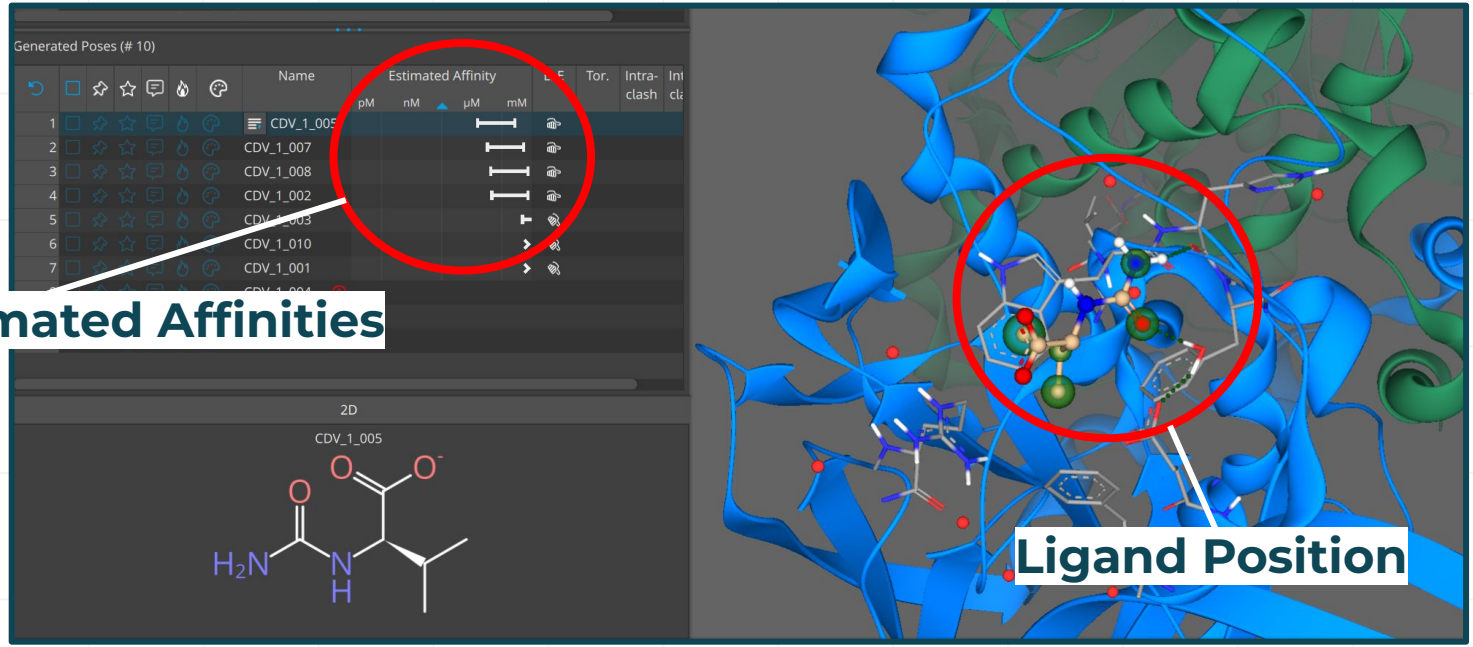
Score Algorithm

Calculates binding affinity

- Different metrics (kcal/mol vs. M)
- Determines optimal binding conformation

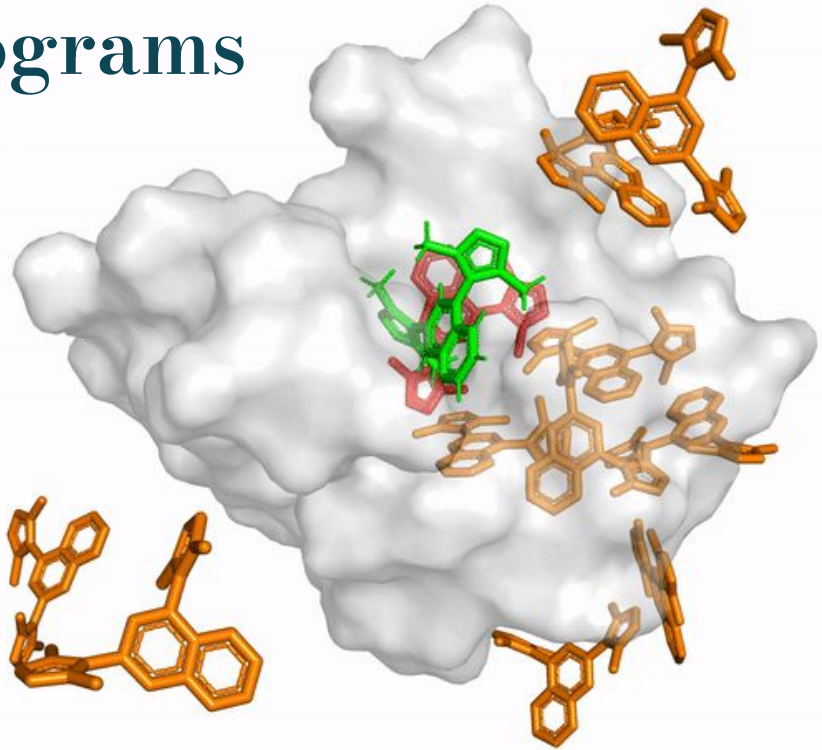
SeeSAR

A traditional docking method



Machine Learning-Based Molecular Docking Programs

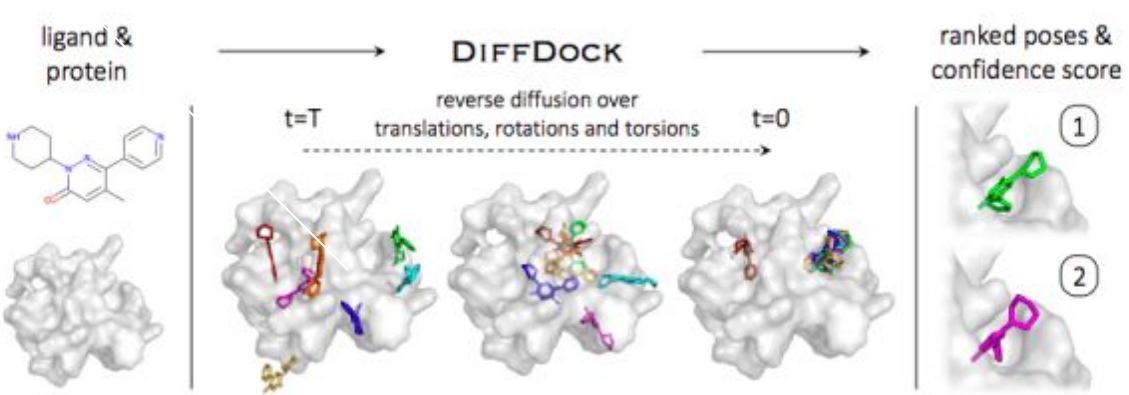
- Optimization and accuracy through reverse diffusion
- “Learn” and predict data using scoring function

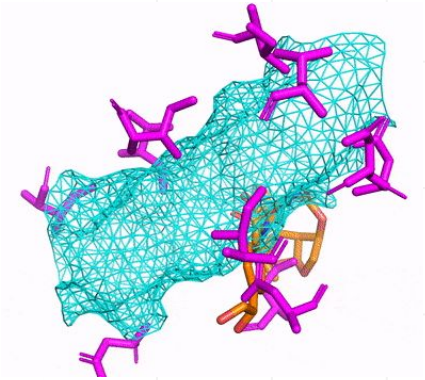


Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, Tommi Jaakkola, DiffDock: “Diffusion Steps, Twists,and Turns for Molecular Docking” 2022

DiffDock-Pocket

A deep learning-based pocket-docking method



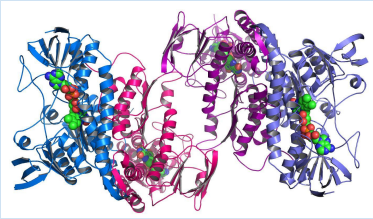


Process Overview

Step 1

Protein identification

- Establish protein dataset



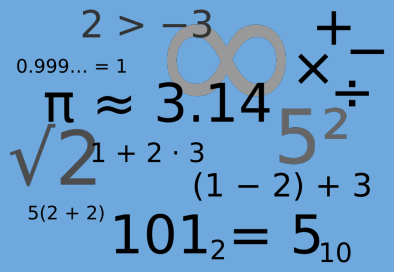
Step 2

Molecular docking

- SeeSAR
- DiffDock-Pocket

Step 3

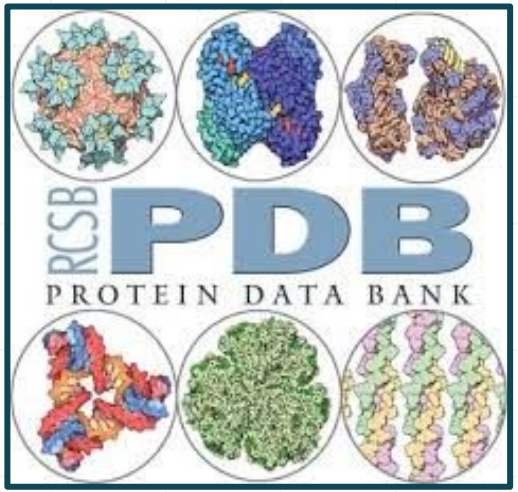
Analysis/Evaluation: compare generated RMSDs



Protein Dataset

- DockGen and why it's important

189 -> 135

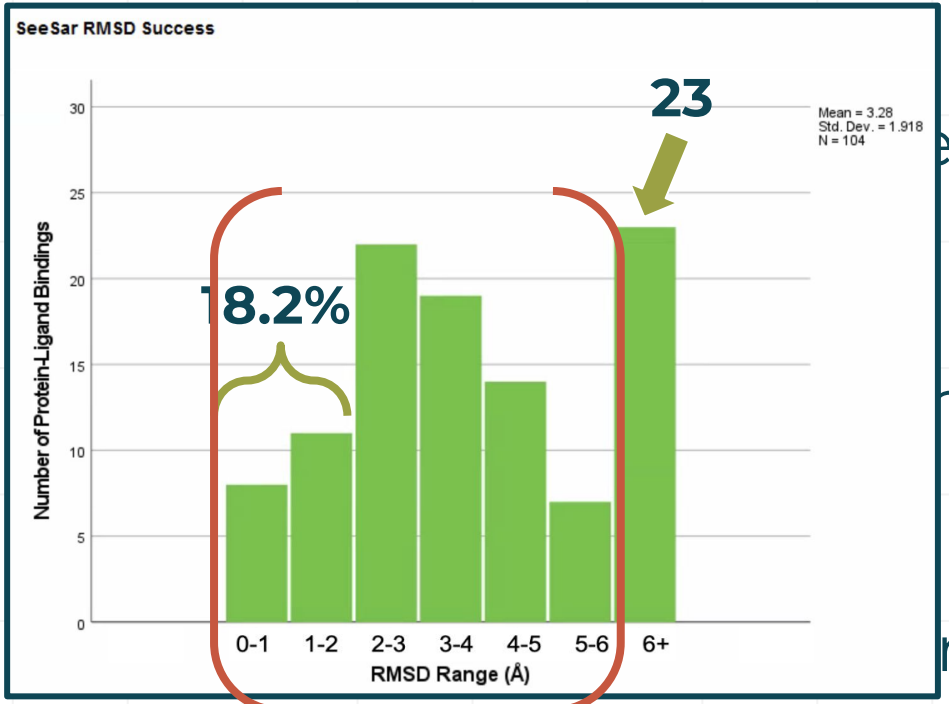


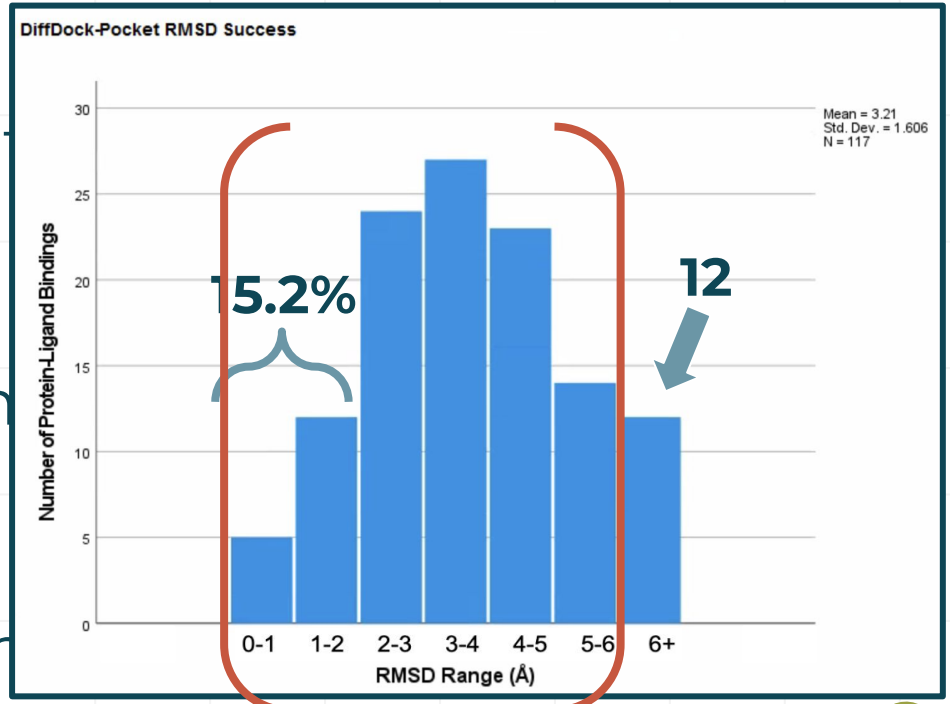
Calculating RMSD Values

- Python script
- Loaded molecules
- Matched atom order and aligned molecules
- Calculated RMSD

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$

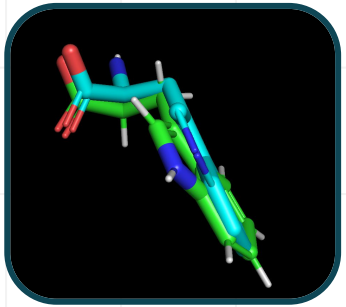
Results

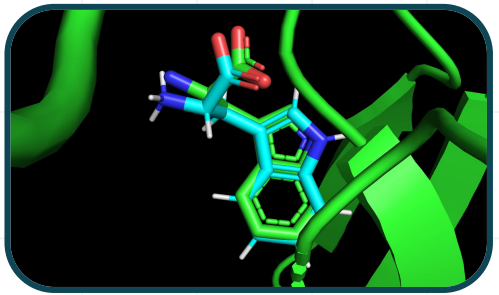

SeeSar RMSD Success
Mean = 3.28
Std. Dev. = 1.918
N = 104
Number of Protein-Ligand Bindings
30
25
20
15
10
5
0
18.2%
23
0-1
1-2
2-3
3-4
4-5
5-6
6+
RMSD Range (Å)


DiffDock-Pocket RMSD Success
Mean = 3.21
Std. Dev. = 1.606
N = 117
Number of Protein-Ligand Bindings
30
25
20
15
10
5
0
15.2%
12
0-1
1-2
2-3
3-4
4-5
5-6
6+
RMSD Range (Å)

Limitations of the Study

- Solely using RMSD/Ångström
  - Small vs. Large structures
- Interface Limitations
- Operating System Optimizations



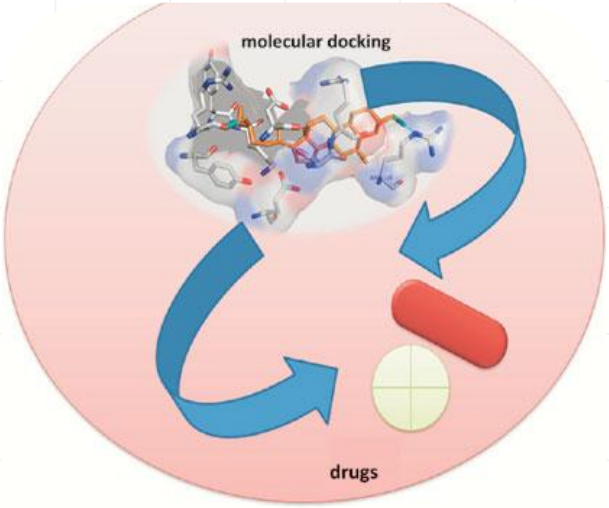


Future Research

- RMSD Calculators
- Alter the Dataset

Future Research

Why do this research?



Dockingin drug designing

Acknowledgements

- Mr. Hannes Stärk and Mr. Michael Plainer
- GSNJS Faculty
- Dr. David Cincotta
- Katerina Pouathas and Joel Moses
- GSNJS Alumni and Parents of Alumni

Thank you!

Any questions?