In Silico Screening of Natural Compounds as Novel Drug Targets for the Treatment of Multiple Myeloma

Rousselene Larson

Advisor: Dr. Rakesh Kaundal

Utah State University

Multiple Myeloma

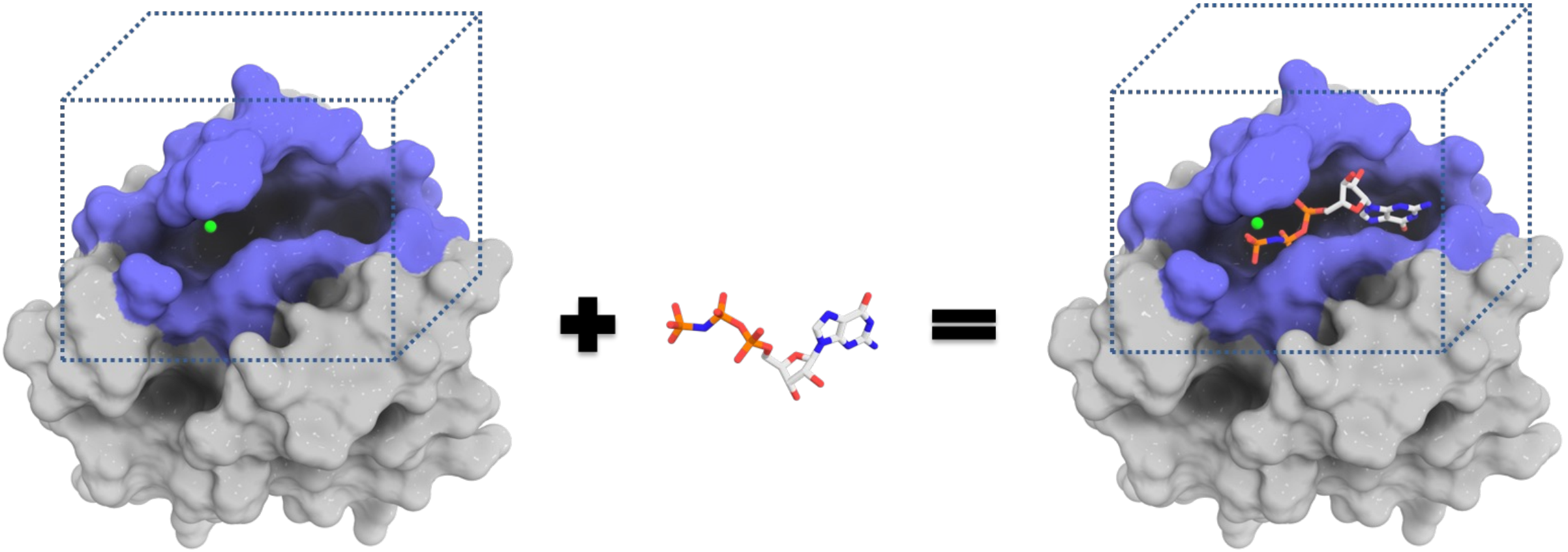
- Second most prevalent type of cancer in the US
- 34,920 new cases in 2021
- 12,410 deaths

Hypothesis

- Toxicity and resistance are correlated with existing therapies for MM, emphasizing the need for novel, effective therapeutics.
- Due to their biological activity, small molecules found in plant-based natural compounds provide therapeutic benefits that affect human protein expression.
- Docking methods seem crucial in order to develop a comprehensive solution.

Objective

- Mine literatures to identify M-proteins and generate their 3D structures.
- Generate natural compounds dataset from various databases (ZINC, PubChem, NPCARE).
- Perform docking and post-docking analysis.
- Identify the binding sites where the protein interact with the natural compound (ligand).
- Perform simulation to study the molecular forces acting on the complex.
- Generate a list of potential drug candidates for MM.
- Web server implementation: https://bioinfo.usu.edu/myDockDB



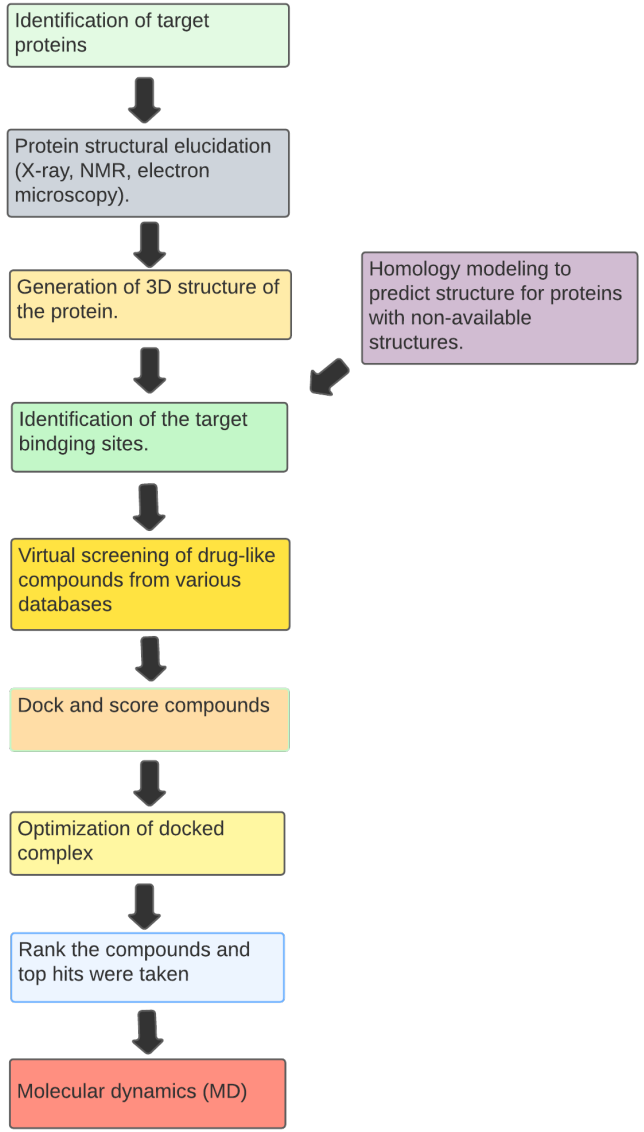
Molecular Docking

Rigid Flexible Semi-flexible

M-Proteins Collection

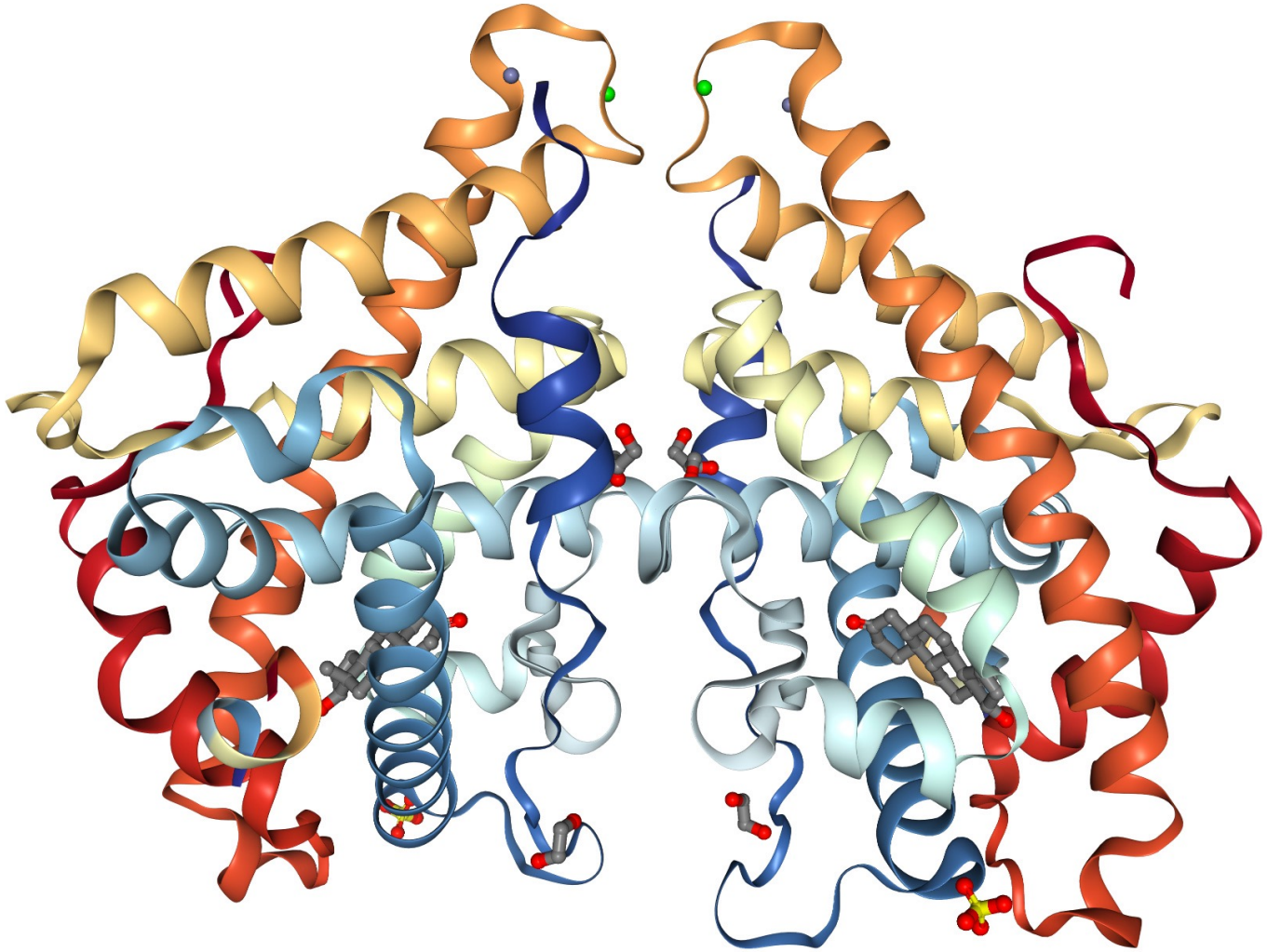
GENE TYPES	# OF GENES COLLECTED
SPEP (serum)	16
UPEP (urine)	322
LCP1 (T-cells)	41
Others	15

Methods


Identification of target proteins
Protein structural elucidation (X-ray, NMR, electron microscopy).
Generation of 3D structure of the protein.
Homology modeling to predict structure for proteins with non-available structures.
Identification of the target bindging sites.
Virtual screening of drug-like compounds from various databases
Dock and score compounds
Optimization of docked complex
Rank the compounds and top hits were taken
Molecular dynamics (MD)

Generate 3D Structure for the Receptor (Protein)

- PDB structure
- Homology Modeling



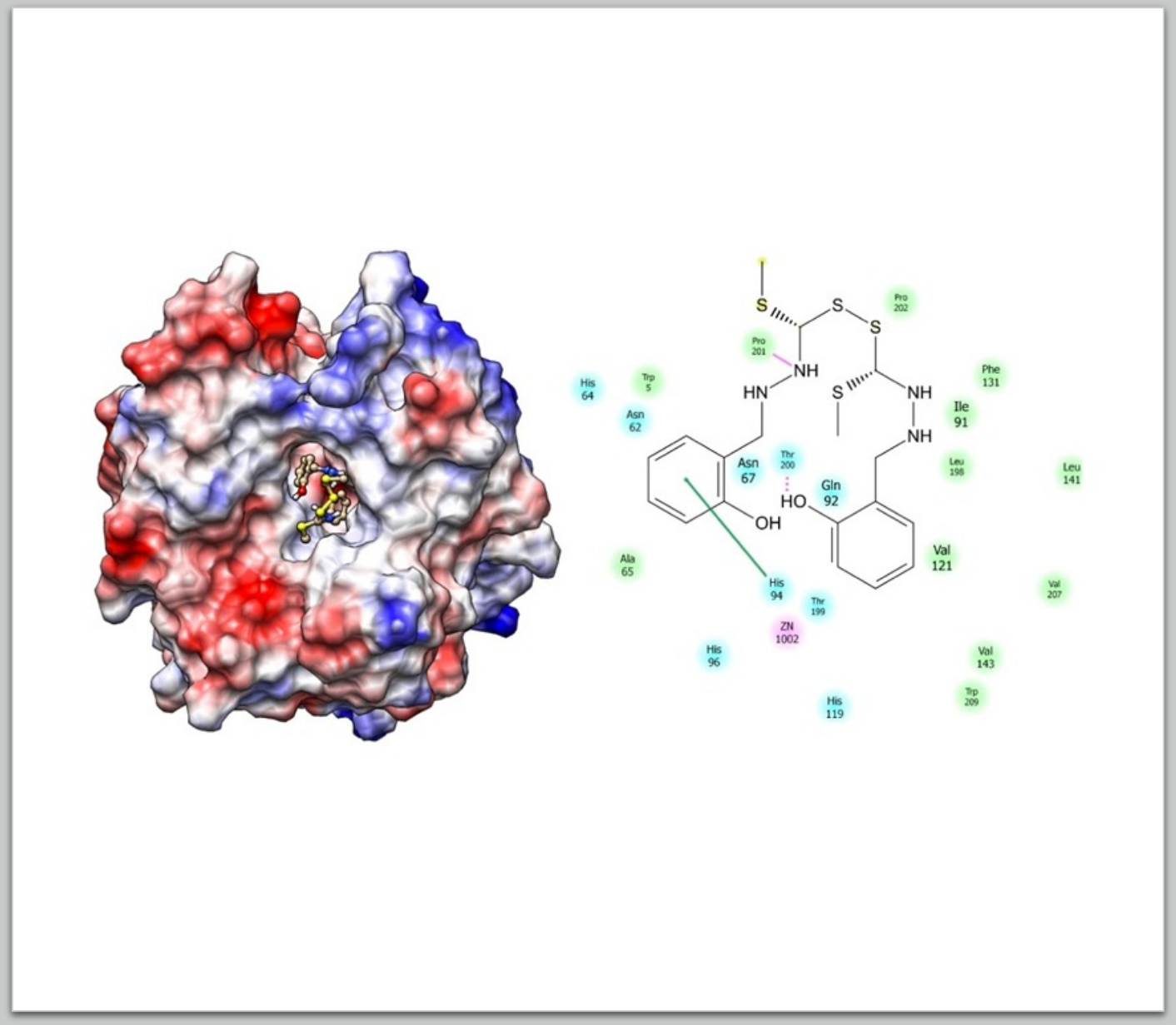
Preparation of the Receptor Molecule

- Removal of the water molecules
- Stabilizing charges
- Filling in the missing residues

Identification of the Active Sites

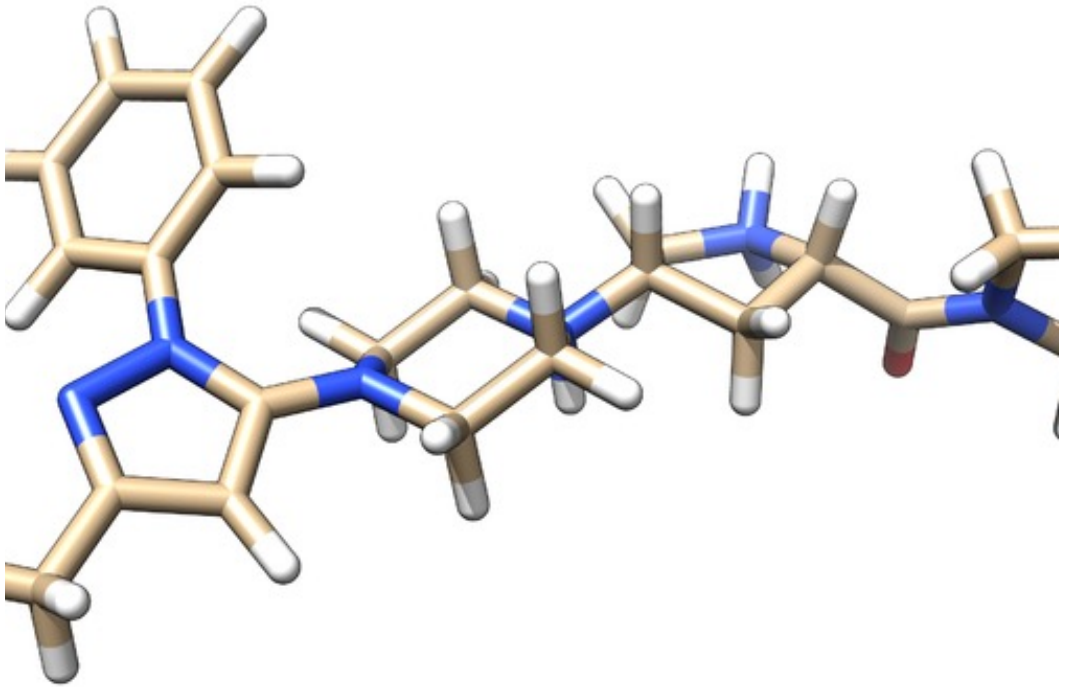
- When these protein cavities come into contacts with any foreign molecules, they become active and are referred to as active sites in the pocket.

- The receptor may have many active sites but the one of the interest is selected.



Natural Compound(ligand) Preparation

- Ligands is obtained from various databases like ZINC, and PubChem.
- While selecting the ligand, the LIPINSKY'S RULE OF 5 is applied.



Ligands Collection

- *Initial Compounds
- **Filtered Compounds

Filter Used	Natural Compound Databases					
	PubChem	Input	Output	Zinc	Input	Output
REOS		15,970*	5,387		224,205*	123,223
PAINS 1		5,387	5,369		123,223	123,074
PAINS 2		5,369	5,365		123,074	122,938
PAINS 3		5,365	5,286**		122,938	116,714**

Also duplicates found between ZINC and PubChem compounds were removed. **3,715** duplicates were found.

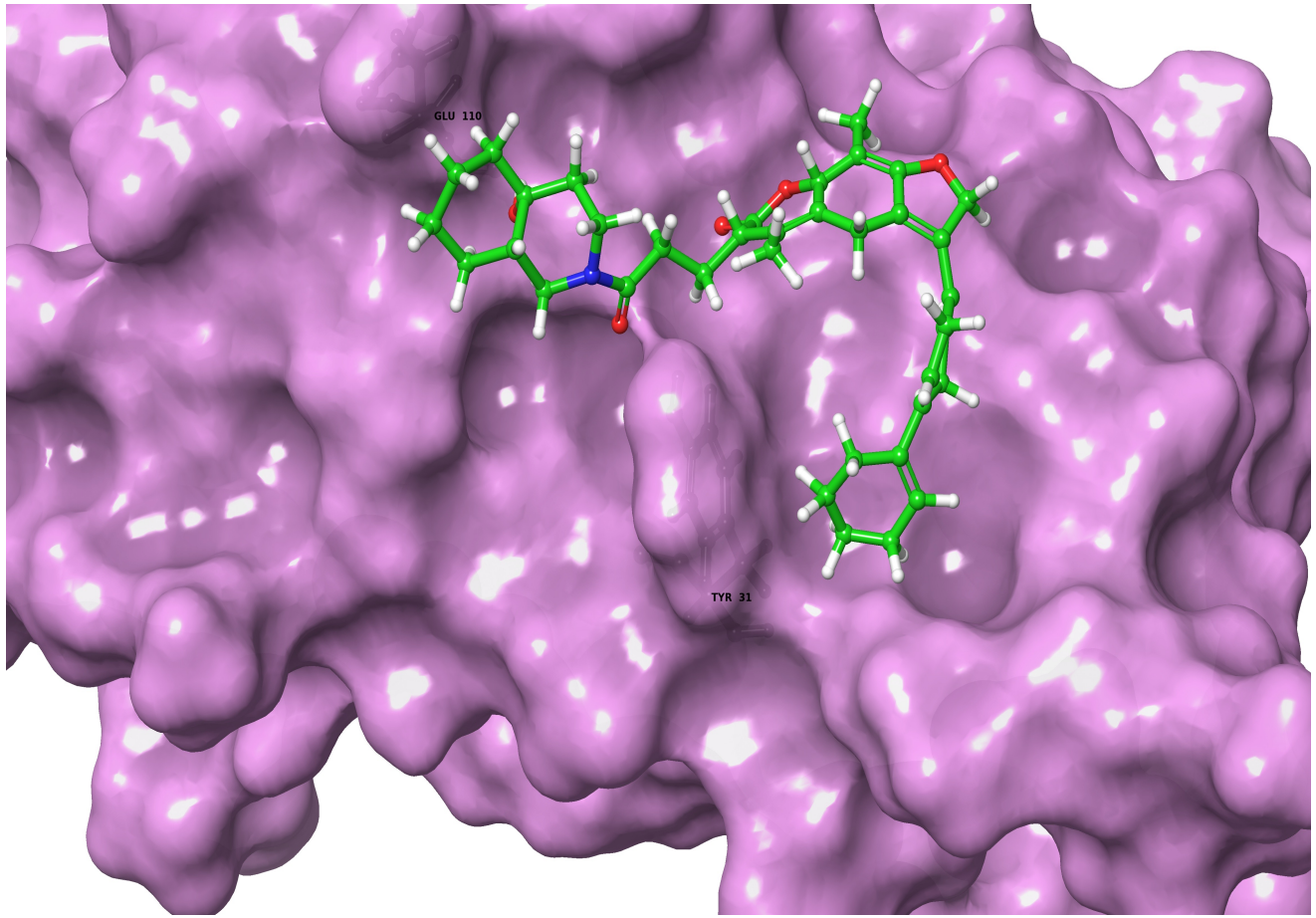
Result and Discussion

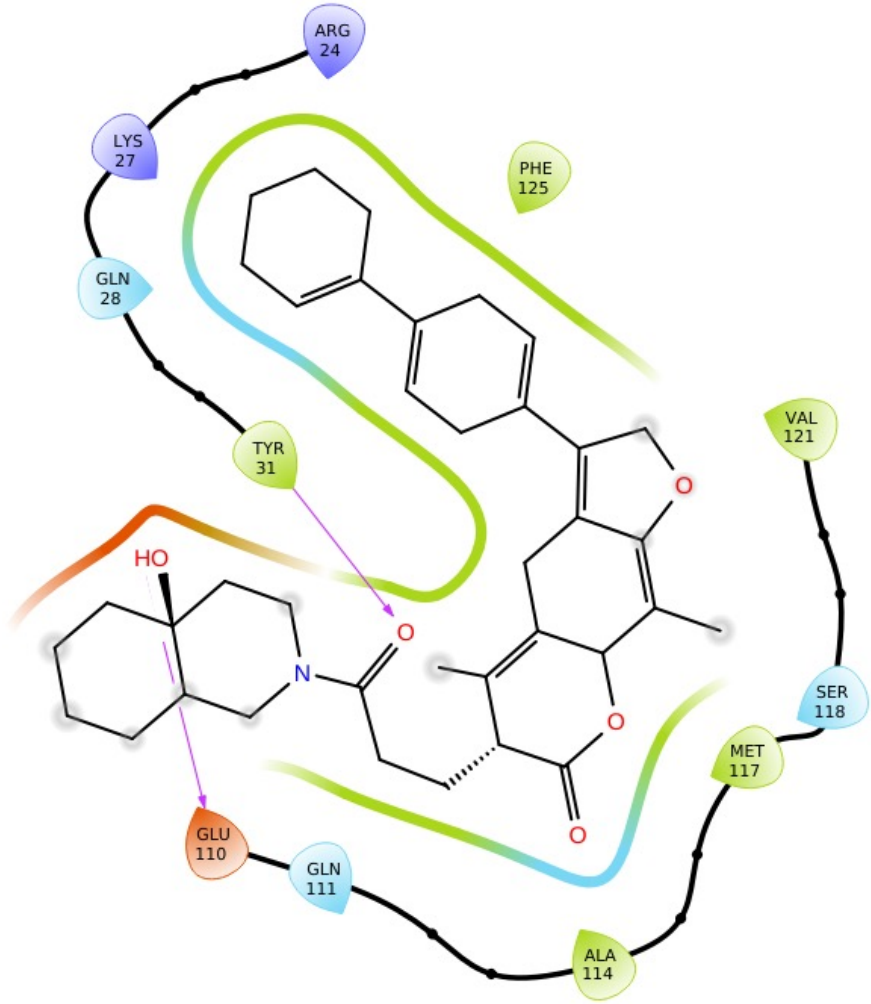
- **IL6** (PDB: ALU) produced by marrow stomal cells contribute to the overactivity of osteoclasts causing bone lesion in MM patients.

- **TP53** (PDB: 2BIP) is the most mutated protein in human cancers.

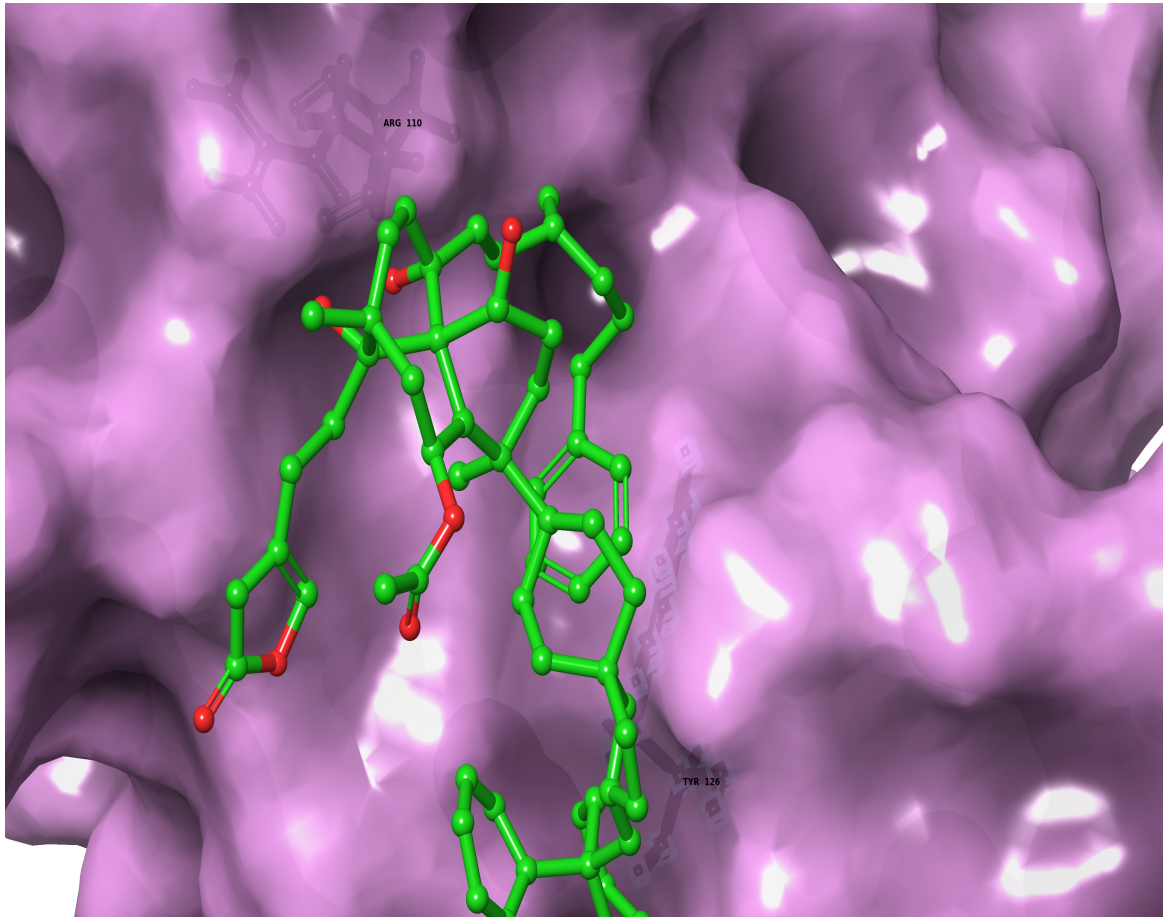
- **NRAS** (PDB: 6ZIO) RAS mutations cause cells grow uncontrollably. They also make cells resistant to some available cancer therapies.

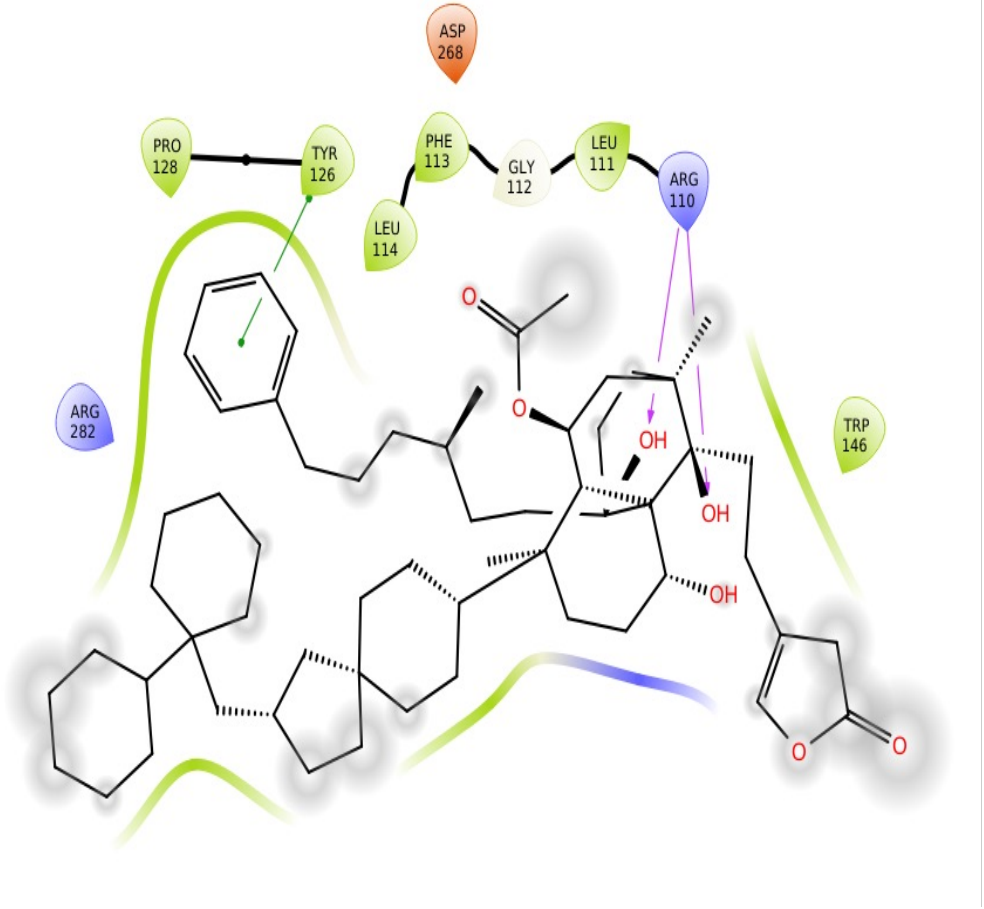
ALU_ ZINC000001900625 Complex Structure



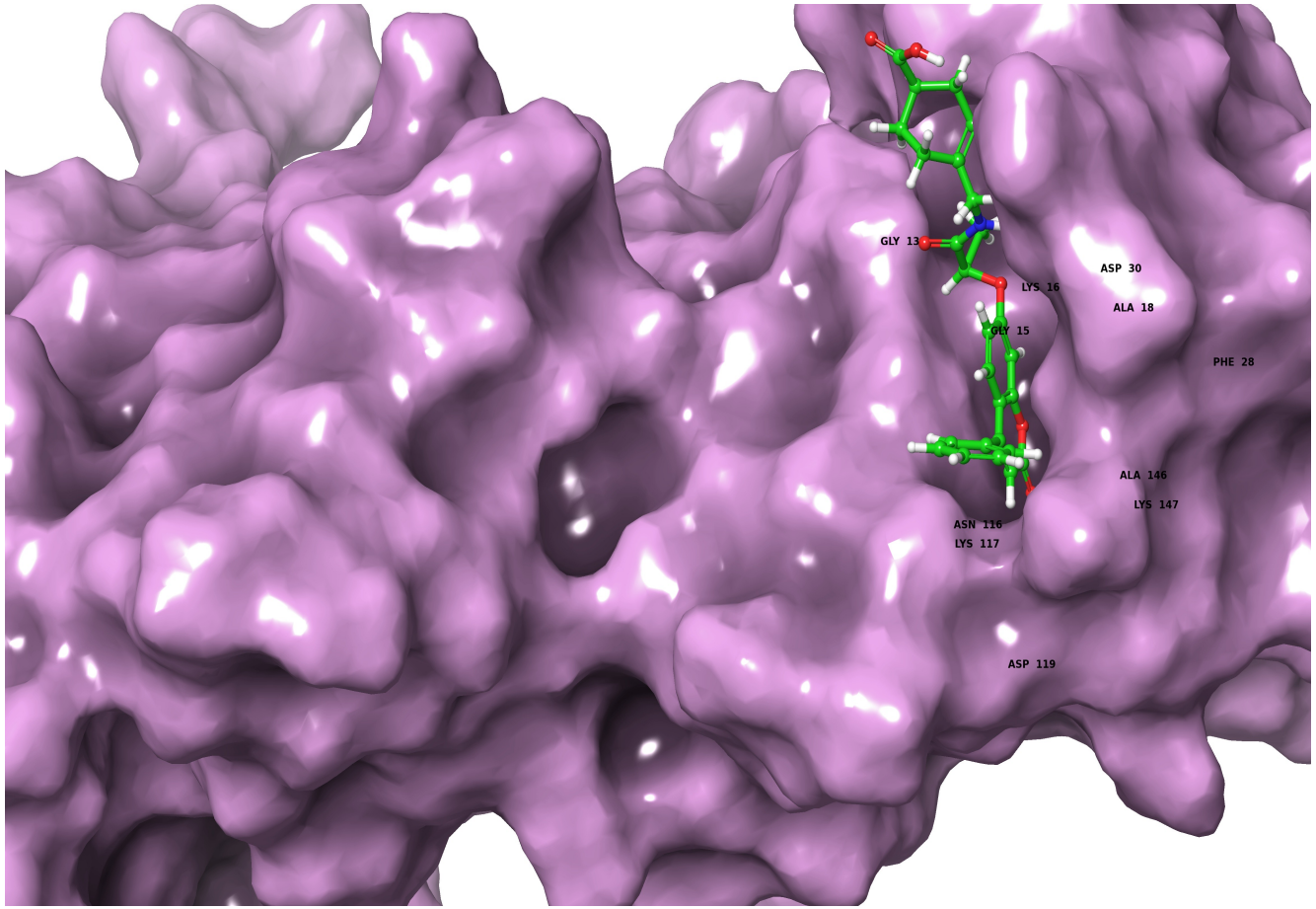


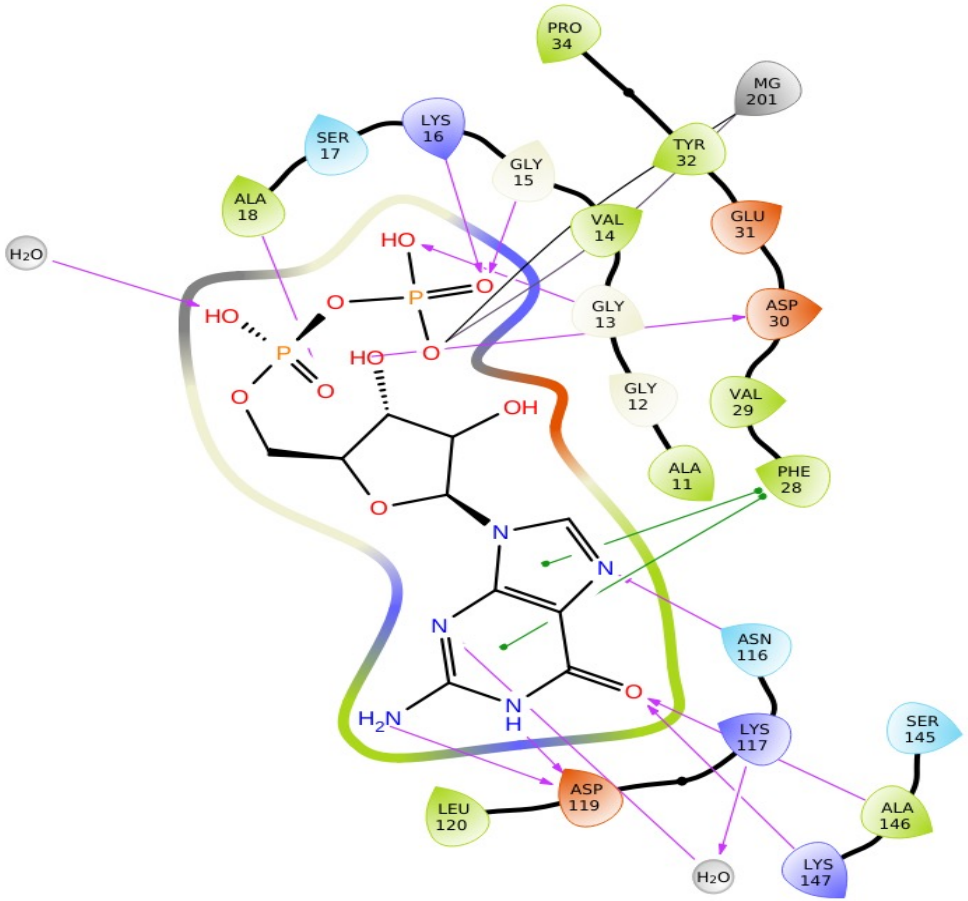
2BIP_ ZINC000150359959 Complex Structure





6ZIO_ ZINC000006142058 Complex Structure



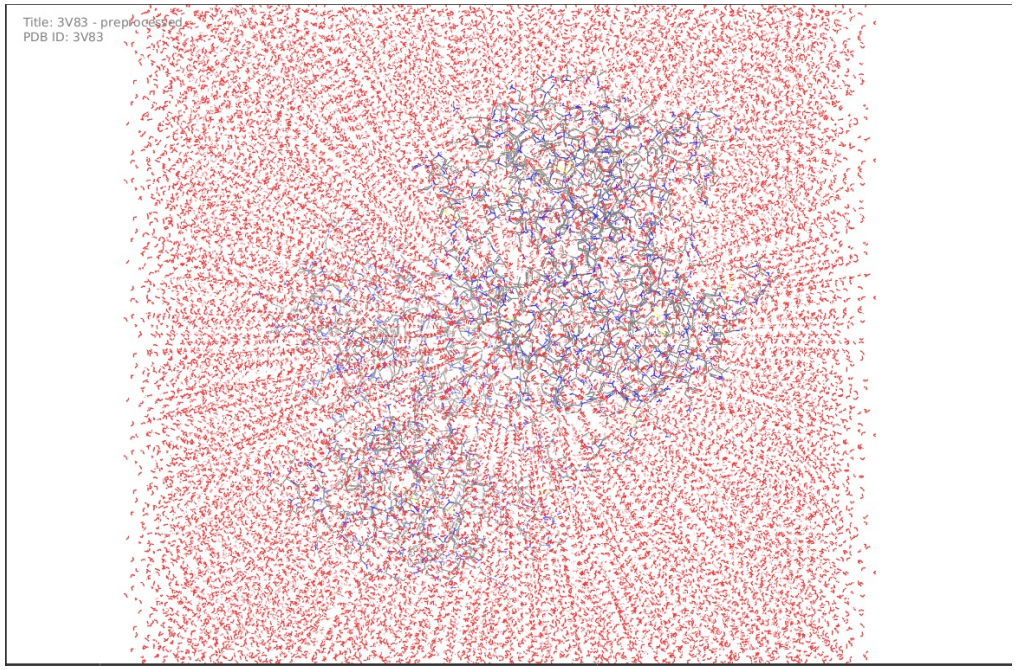


Top 10 Lead Compounds and their Docked Inhibitors

Gene	PDB_id	Compounds	Binding affinity (kcal/mol)	Interacting Amino Acid	Types of Bonds	Bond Distance (A)
TP53	2BIP	ZINC000150359959 ZINC000150359969 ZINC000150359978 ZINC000150359935 ZINC000150359973 ZINC000085542224 ZINC000085541700 ZINC000150359931 ZINC000085594244 ZINC000095909362 **2bip_CID_3213**	-8.0 -7.8 -7.4 -7.3 -7.2 -7.2 -6.8 -6.3 -5.8 -5.6 **-6.1**	Arg110, Arg110, Tyr126 Tyr126 Asp268 Tyr220 Asp268 Tyr126, Leu111, Leu111 H2o, Gly262 Asp268 Glu258, H2o, H2o Tyr126 **Thr211**	2 H-Bond, pi-pi stacking 1 H-Bond 1 H-Bond 1 H-Bond 1 H-Bond 3 H-Bond 2 H-Bond 1 H-Bond 3 H-Bond 1 H-Bond **1 H-Bond**	1.92, 2.00, 4.64 1.96 2.00 2.14 2.08 2.05,2.64, 2.66 1.54, 2.25 2.32 2.01, 2.06, 2.41 2.15 **2.36**
IL6	1ALU	ZINC000001900625 ZINC000150359978 ZINC000276935046 ZINC000150359973 ZINC000150359959 ZINC000150359935 ZINC000150359969 ZINC000085541700 ZINC000085541935 ZINC000085594244 **1alu_F_TLA**	-7.8 -7.7 -7.5 -6.8 -6.8 -6.6 -5.9 -5.7 -5.7 -5.6 **-4.7**	Glu 110, Tyr 31 Lys 66, H2o, H2o H2o Lys70, Phe74 Gln28 H2o Tyr31 Asn63 H2o Leu92, H2o, H2o **Arg104, Thr43**	2 H-Bond 3 H-bond 1 H-bond 1 H-Bond, 1 pi-pi stacking 1 H-Bond 1 H-Bond 1 H-Bond 1 H-Bond 1 H-Bond 3 H-Bond **2 H-Bond**	2.38, 2.18 2.19, 1.25, 2.08 1.54 2.26, 3.75 2.21 2.57 2.63 2.37 2.42 2.72, 1.58, 2.78 **1.92, 2.34**
KRAS	6ZIO	ZINC000006142058 ZINC000011867065 ZINC000253499958 ZINC000150359959 ZINC000150359935 ZINC000150359978 ZINC000150359931	-11.6 -10.2 -9.3 -8.4 -7.8 -7.7 -7.5	Asp30, H2o, Ala18, Lys16, Gly13, Gly15, Asn116… Lys147, Lys117, Asn116, H2o, Phe28, Phe28 Arg68, Gln99 Tyr32, H2o, Phe28 Arg68, Gln99 Gly13, Asp30, Lys117 Arg102, H2o	13 H-Bond, 2 pi-pi stacking 3 H-Bond, 2 pi-pi stacking 2 H-Bond 2 H-Bond, 1 pi-pi stacking 2 H-Bond 3 H-Bond 2 H-Bond	1.75, 2.61, 1.95, 2.03, 1.87, 2.11, 2.35… 2.07, 2.60, 2.29, 4.74, 5.36 2.02, 2.14 2.15, 1.26, 4.70 2.02, 2.14 2.18, 1.76, 2.15 2.67, 2.72

Molecular Dynamic Simulations

- MD is performed to study the conformational stability of protein-ligand complexes and their changes.
- Our system is simulated up to 100ns.
- RMSD and RMSF values of the c-alpha atoms were calculated to reveal thermodynamic stability.



RMSD Calculation

- The root mean square deviation (RMSD) indicates how stable the ligand is with respect to the protein and its binding pocket

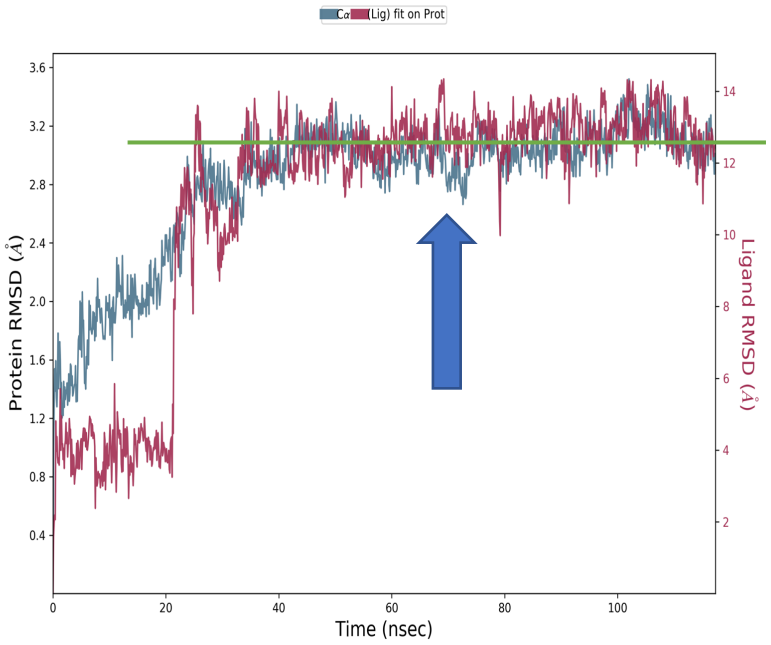
- Equilibrium is reached

- Well within the range of acceptable RMDS ~2-3 A

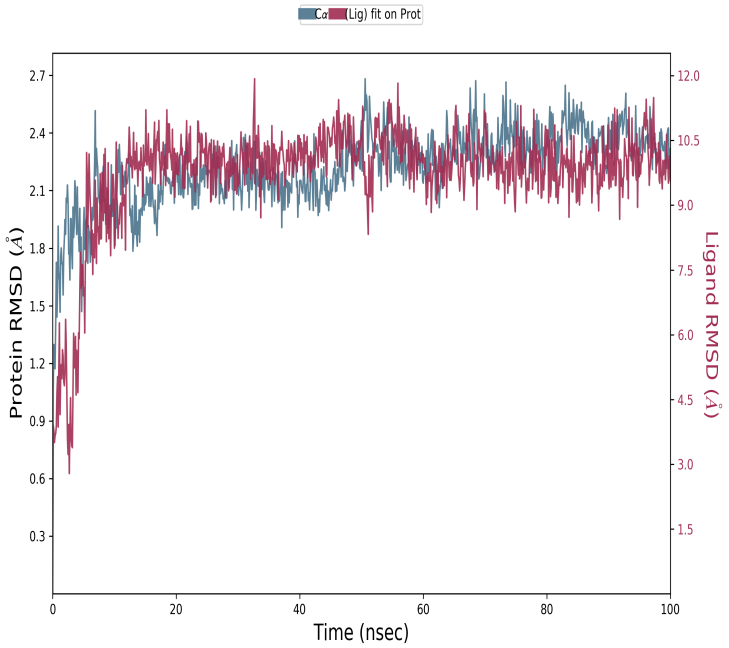
- Indicate the compounds are bound tightly within the cavity of the proteins

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i'(t_x)) - r_i(t_{ref}))^2}$$

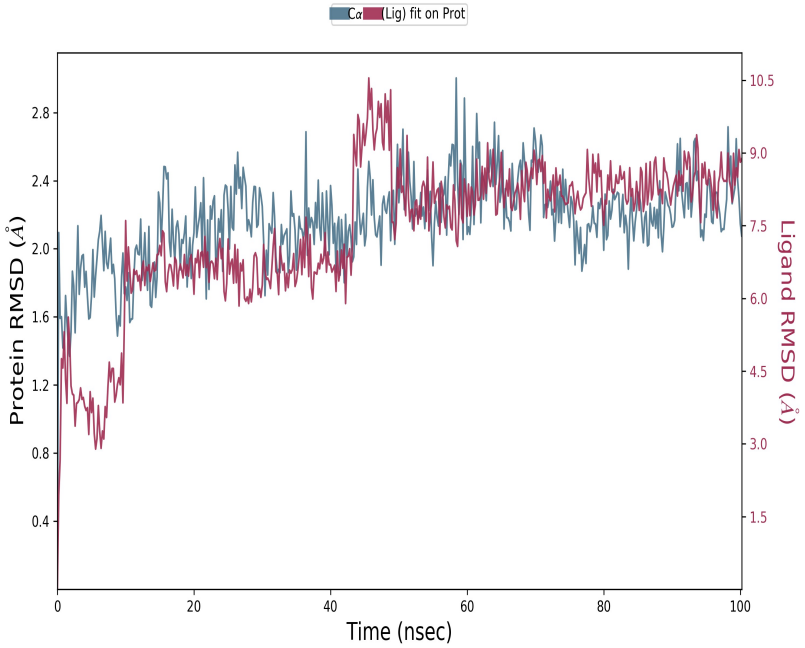
RMSD



1ALU



2BIP



6ZIO

RMSF Calculation

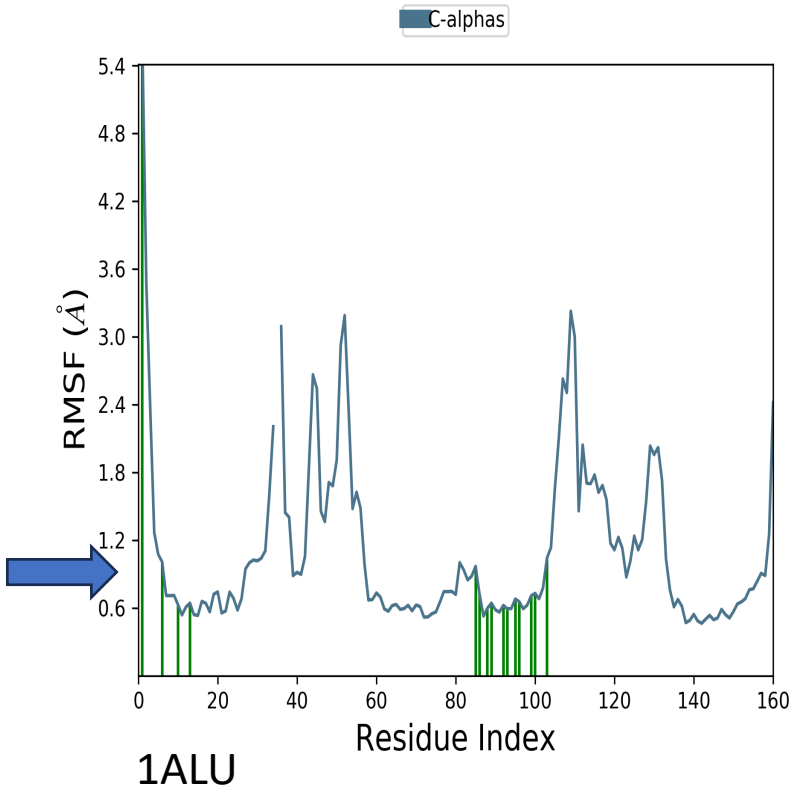
- The root mean square fluctuation(RMSF) is used to examine the binding affinity efficiency of the lead compound with each gene

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_i'(t)) - r_i(t_{ref}))^2}$$

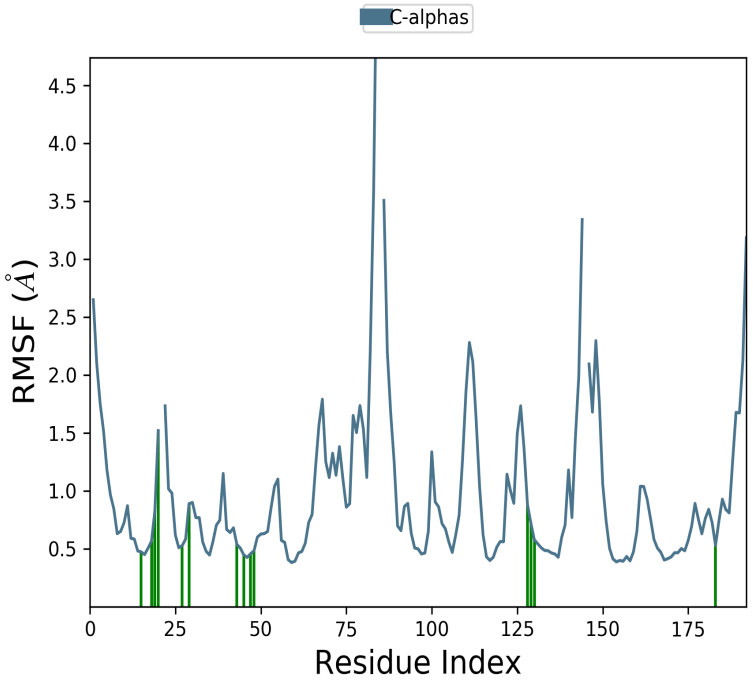
- RMSF ~ 2 A

- The binding pockets of each protein is quite stable during 100 ns period of MD simulation

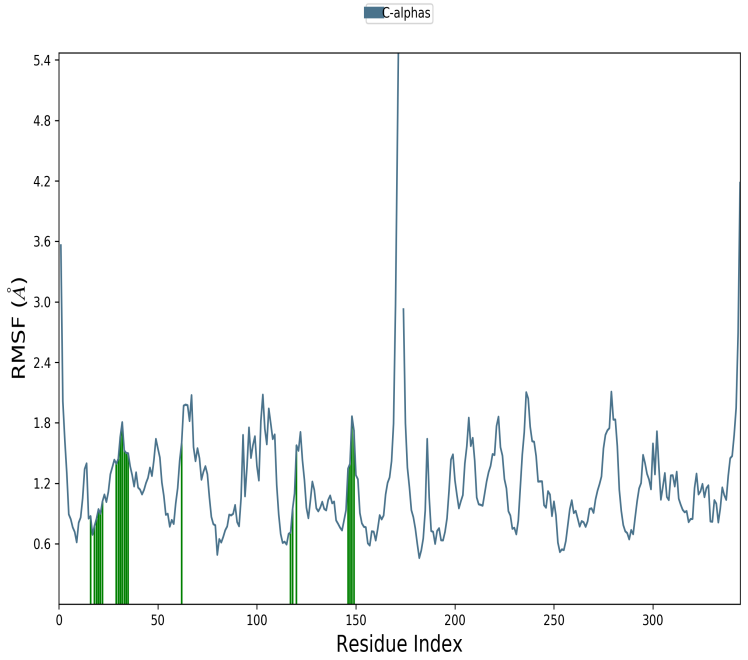
RMSF



1ALU



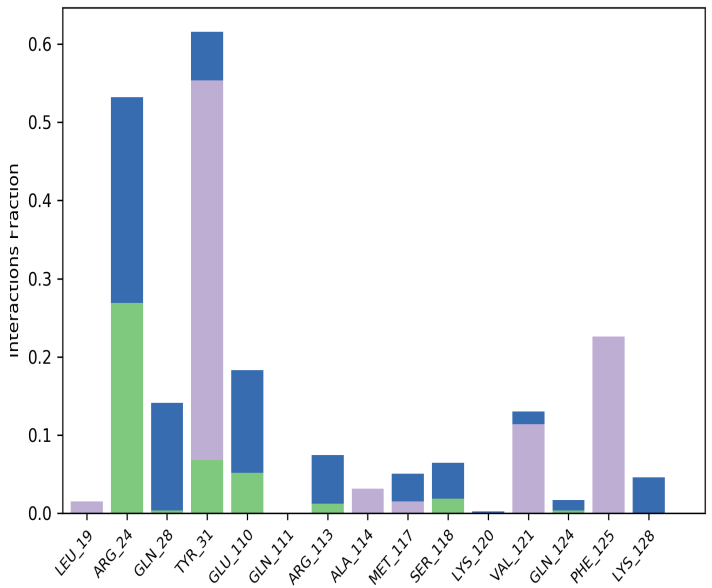
2BIP



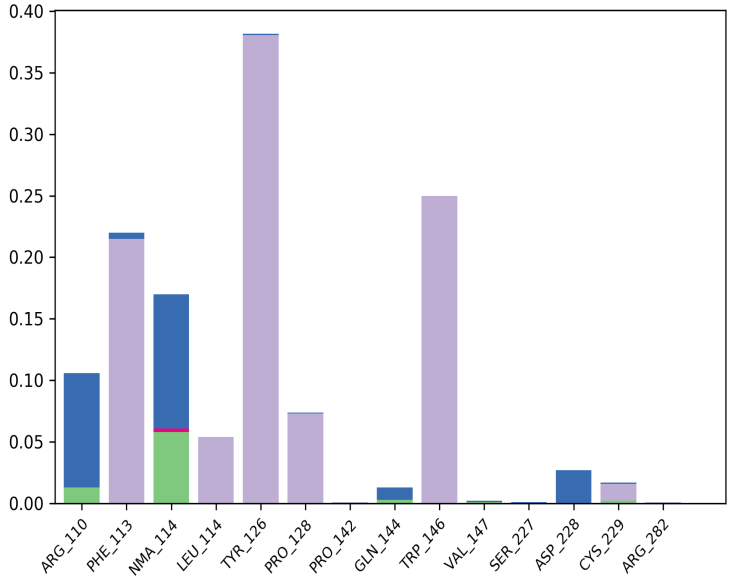
6ZIO

****Peaks indicate areas of the protein that fluctuate the most during the simulation**

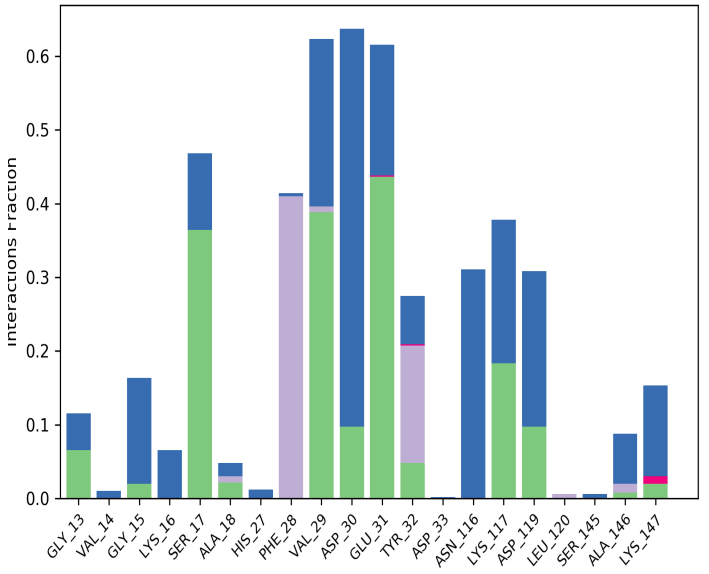
Binding Types



1ALU



2BIP



6ZIO

What we needed

- 342 proteins
- 118,285 ligands
- 40,453,470 total docking runs
- Node 6,612 core
- 2.7 million core hours
- 1.3 TB of results
- Database implementation

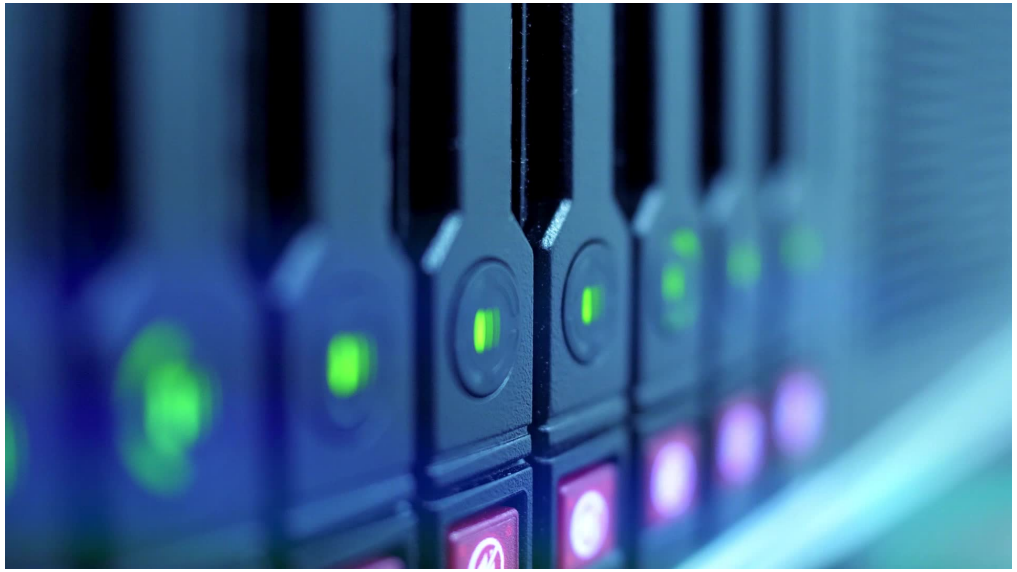
Computational Challenges

We simply did not have the time

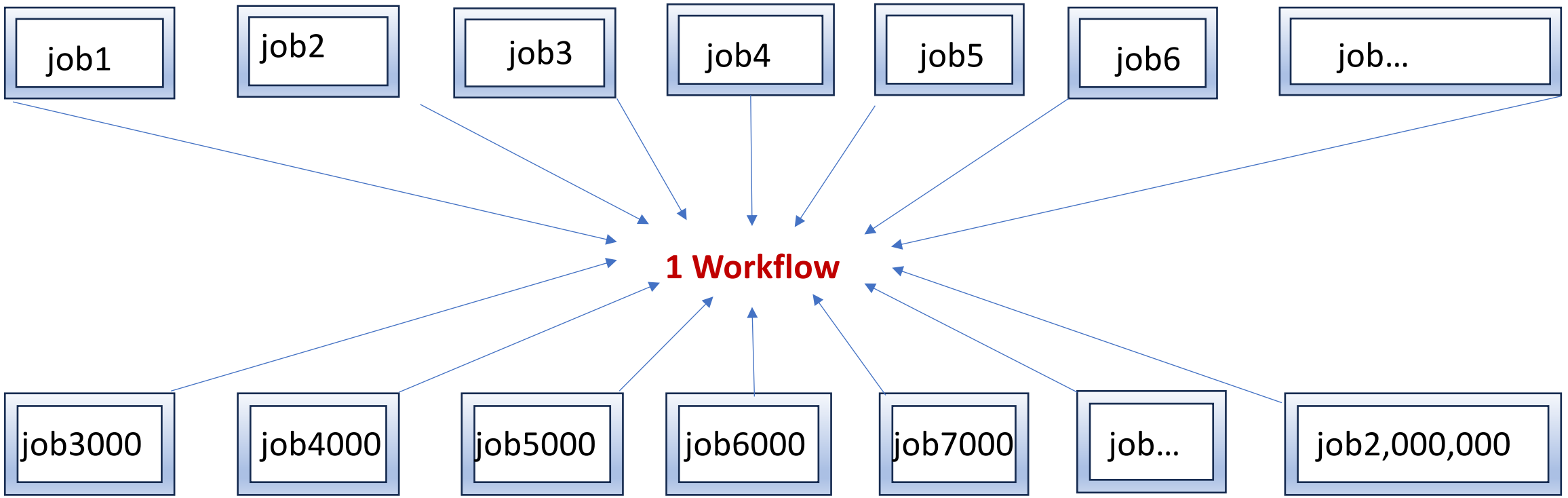
We simply did not have the computing resources

HTCondor System

- Excellent scheduling and computing power
- Tasks distribution more powerful than what a GPU system can handle
- Greatly increase computer capacity
- Greatly reduces time Consumption
- Perform 40,453,470 total docking runs in just a month.



my_HTC Docking Job



Web-Server Implementation

Home | Datasets | Help | Search Menu

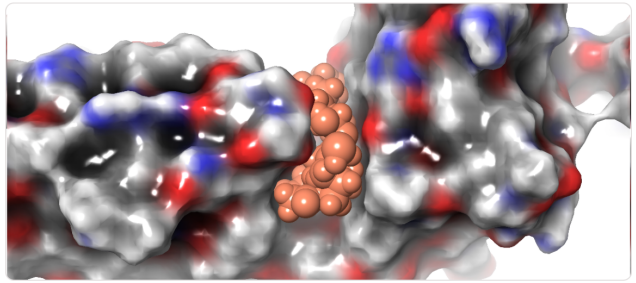
myDockDB: Multiple Myeloma Proteins and Natural Compounds Docking WEB-resource

Overview

Molecular Docking: the molecular docking approach may be used to represent the atomic level interaction between a small molecule and a protein, allowing us to define small molecule behavior in target protein binding sites as well as elucidate key biochemical processes. The docking procedure consists of two main steps: predicting the ligand structure as well as its position and orientation within these sites (known as pose) and determining the binding affinity.

Multiple Myeloma (also known as cancer of the plasma cells) is a disease where malignant plasma cells, also known as myeloma cells, create clones of themselves and accumulate in the bone marrow. The **M-protein**, or M- spike, or paraprotein, or myeloma protein, is an antibody or immunoglobulin secreted by malignant plasma cells. Most myeloma patients have it found in their blood and/or urine. Many research groups have been working into find a solution to this devasting condition. However, toxicity and resistance are correlated with existing therapies for MM, emphasizing the need for novel, effective therapeutics.

Due to their biological activity, small molecules found in natural compounds provide therapeutic benefits that affect human protein expression. Here we present M3BioDock, a database implemented to provide to general audience a way to check the binding affinity of a multiple myeloma protein to a natural compound.



Get Protein Record

PDB ID

3V83

View Information

Get Protein Record

Gene Name

TF

View Information

Please cite below if you use the database:

Rousselene, R., myDockDB: Molecular Docking of Multiple Myeloma Proteins, (2023). DOI: 10.1186/s13007-022-00897-9

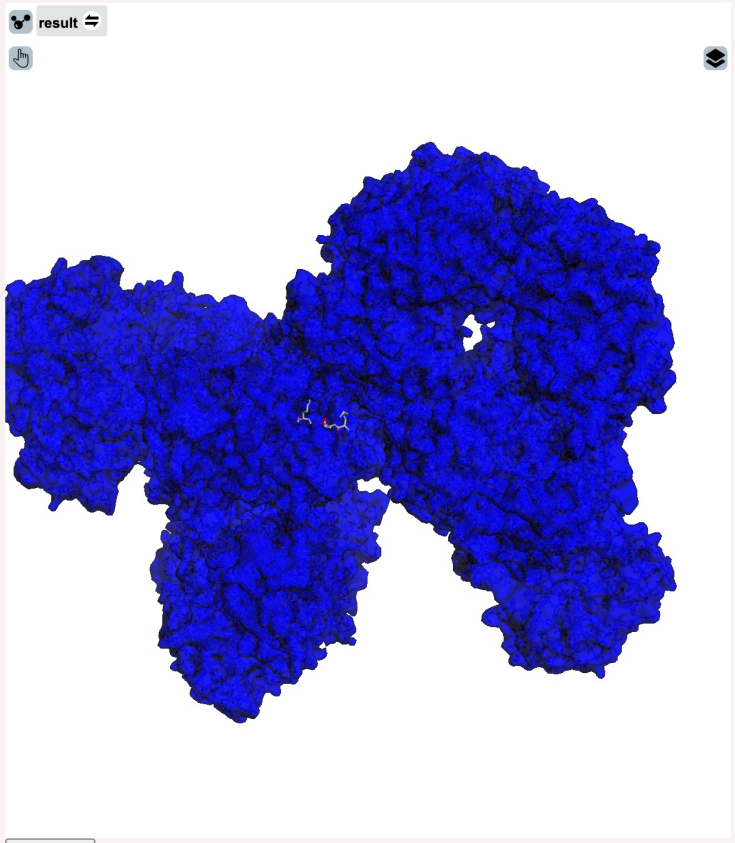
URL: https://bioinfo.usu.edu/myDockDB

Search result

Top Compounds that Docked to your Protein

Protein	Ligand	Complex	Affinity	Download	Visualization
3V83	139585659	3V83_139585659	-9.3	3V83_139585659	
3V83	139585574	3V83_139585574	-8.3	3V83_139585574	
3V83	139585730	3V83_139585730	-7.9	3V83_139585730	
3V83	139585579	3V83_139585579	-7.9	3V83_139585579	
3V83	139585460	3V83_139585460	-7.7	3V83_139585460	
3V83	139585675	3V83_139585675	-7.6	3V83_139585675	
3V83	139585741	3V83_139585741	-7.6	3V83_139585741	
3V83	139585431	3V83_139585431	-7.3	3V83_139585431	
3V83	139585598	3V83_139585598	-7.3	3V83_139585598	
3V83	139585713	3V83_139585713	-7.2	3V83_139585713	

User Submitted



Download

Select Different Conformation (1 to 10):
1

Visualize Ligand Only:

Select Another Ligand or Complex

Download Results

Protein	Ligand	Complex	Affinity	Download	Visualization
3V83	139585659	3V83_139585659	-9.3	3V83_139585659	
3V83	139585574	3V83_139585574	-8.3	3V83_139585574	
3V83	139585730	3V83_139585730	-7.9	3V83_139585730	
3V83	139585579	3V83_139585579	-7.9	3V83_139585579	
3V83	139585460	3V83_139585460	-7.7	3V83_139585460	
3V83	139585675	3V83_139585675	-7.6	3V83_139585675	
3V83	139585741	3V83_139585741	-7.6	3V83_139585741	
3V83	139585431	3V83_139585431	-7.3	3V83_139585431	
3V83	139585598	3V83_139585598	-7.3	3V83_139585598	
3V83	139585713	3V83_139585713	-7.2	3V83_139585713	

Conclusion

- At the present time, there is no known cure for Multiple Myeloma. Efforts to control the disease are therefore more focused on maintenance and rehabilitation.
- In this research we use the strategy of drug repurposing to show how plant-based natural compounds can be very effective against Multiple Myeloma.
- The RMSD and RMSF results we obtain show that the c-alpha atoms of these proteins reveal a strong thermodynamic conformational stability with these natural compounds.
- HTCondor has made it possible to make this data available for our community.

UtahStateUniversity
CENTER FOR INTEGRATED BIOSYSTEMS

UtahState University
PLANTS, SOILS & CLIMATE

Acknowledgment

Advisor: Rakesh Kaundal

Graduate students:

Naveen Duhan

Colin Bearnhart