



# Modeling of Serine Protease Substrate Catalysis in Synthetic Micelle Environment

SIMCODES 2025

NSF Funded

Mentors: Davit Potoyan and Xueyu Song

Alisa Steele



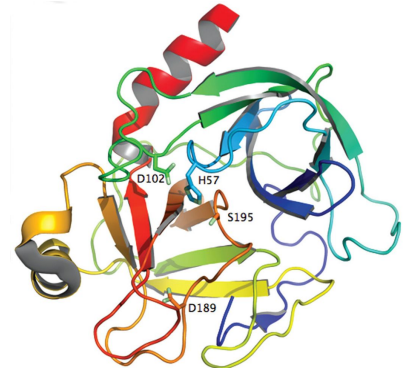
## Goal of Project

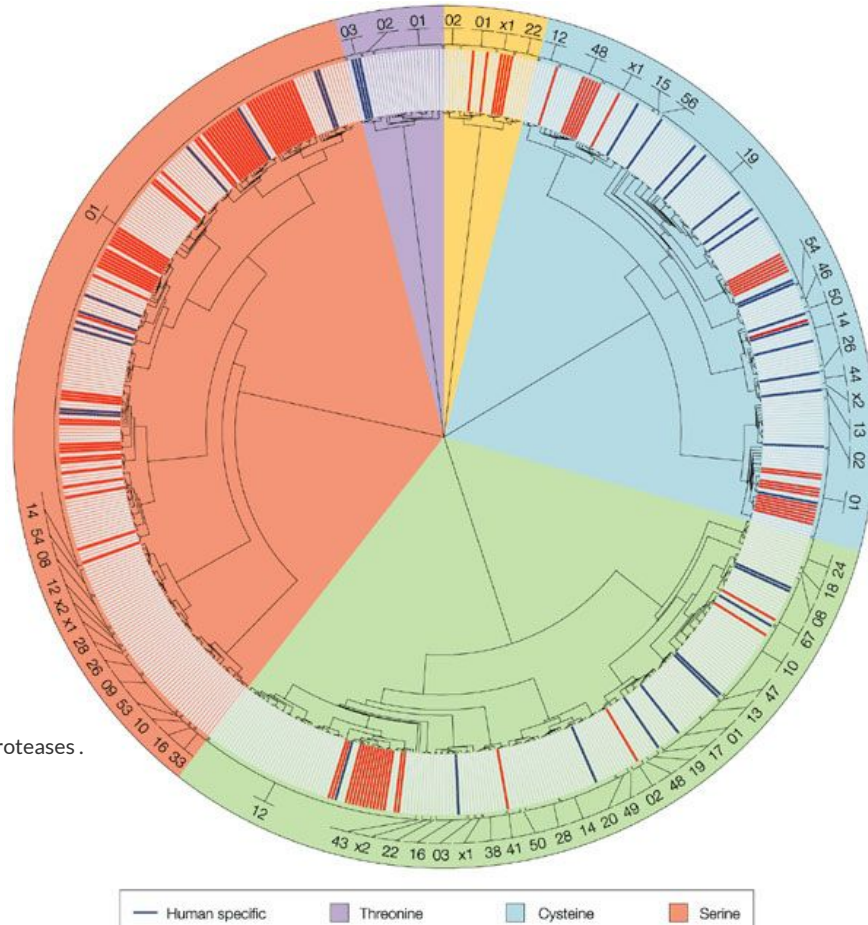
- Develop a machine learning-based simulation framework to predict catalytic efficiency within micelles

# Background Information - Serine Proteases

- Large family of enzymes which cleave peptide bonds in proteins by using a serine residue at the active site
- Found in almost all organisms
- Important for biological processes like programmed cell death, digestion, and immune responses
- Ex: thrombin, plasmin, trypsin

Di Cera E. Serine proteases. IUBMB Life. 2009 May;61(5):510-5. doi: 10.1002/iub.186. PMID: 19180666; PMCID: PMC2675663.





Neitzel, J. J. (2010) Enzyme Catalysis: The Serine Proteases .  
*Nature Education* 3(9):21



# Role in Project

- Data collection and preprocessing
  - Collect a dataset of serine protease structures from Protein Data Bank
  - Represent substrate molecules through SMILES notation
  - Perform molecular docking within enzyme's active site
  - Graph data from collected datasets to search for correlation and trends
- Computational Simulations
  - Perform molecular dynamics simulations of enzyme-ligand interactions inside enzymes \*
  - Investigate how enzyme environment influences binding affinity and transition state stability by modifying interactions within pocket \*
  - Compare computational predictions with experimental data \*



# What I've Learned

- Computational research experience
- Coding
- Interpretation of data
- Poster-making and presenting



## Current Progress

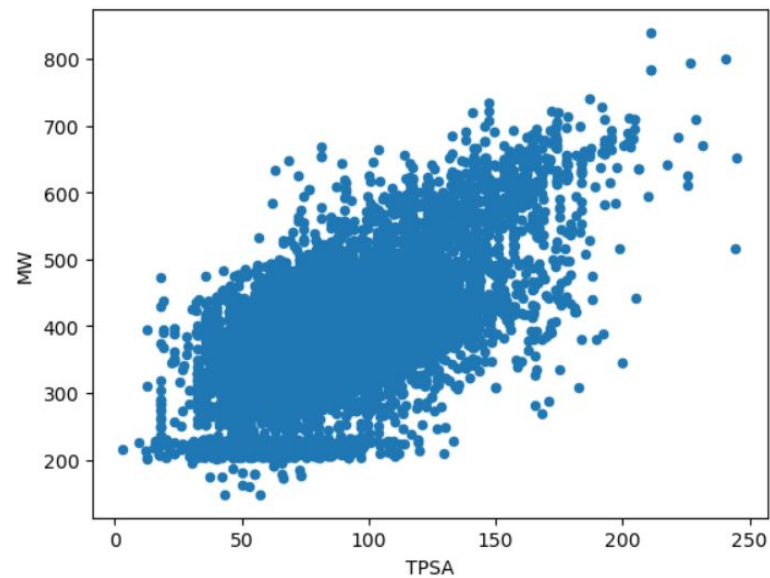
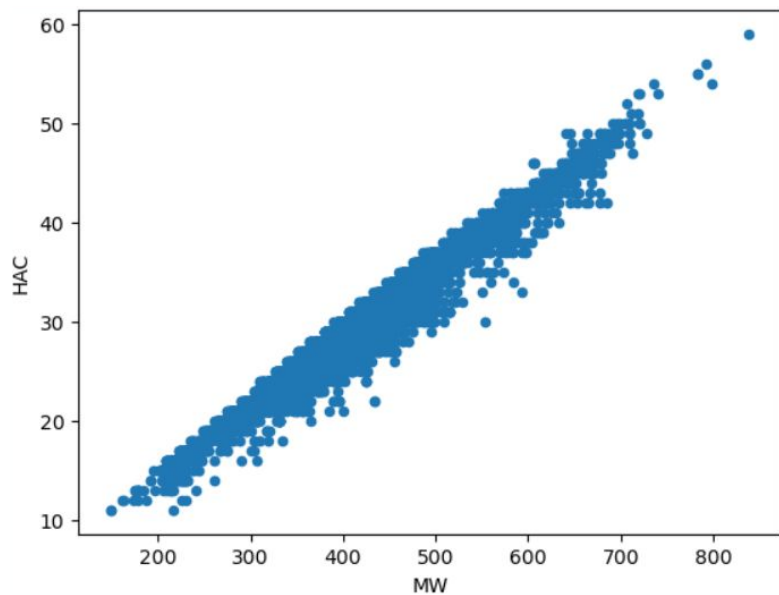
- $K_i$ ,  $K_{cat}$ ,  $K_m$ , and small molecule database
- All SMILES notations
- Comparative graphs
- Beginning docking

	IDNUMBER	Chemical_Name	CAS	PriceCoeff	MW	Description	FSP3	clogP	TPSA	H-acceptors	...	Activity_value	smiles
0	F2147-9803	1-(1-benzylpiperidin-4-yl)-3-methylimidazolidi...	72338-84-2	1.5	273.37	solid	0.56	0.00	26.79	1	...	3100.0	CN1CCN(C2CCN(CC3=CC=CC=C3)CC2)C1=O
1	F6541-1489	1-cyclopropyl-3-(1,2,3,4-tetrahydroisoquinolin...	2097865-50-2	1.5	256.34	oil	0.56	2.34	23.55	1	...	550.0	O=C1C(N2CCC3=C(C=CC=C3)C2)CCN1C1CC1
2	F1739-0011	4,6-dimethyl-2-(pyrrolidin-1-yl)pyridine-3-car...	693254-24-9	1.5	201.27	solid	0.50	0.00	39.92	2	...	3548.1	CC1=CC(C)=NC(N2CCCC2)=C1C#N
3	F3162-0106	6,8-dimethyl-[1,2,4]triazolo[4,3-b]pyridazine	18591-75-8	1.5	148.17	solid	0.29	0.00	43.08	3	...	7943.3	CC1=CC(C)=NN2C=NN=C12
4	F3260-0021	7-ethyl-3-methyl-8-(4-methylpiperidin-1-yl)-2,...	941873-43-4	1.5	291.35	solid	0.64	0.00	70.47	3	...	200.0	N1C(N2CCC(C)CC2)=NC2=C1C(=O)NC(=O)N2C



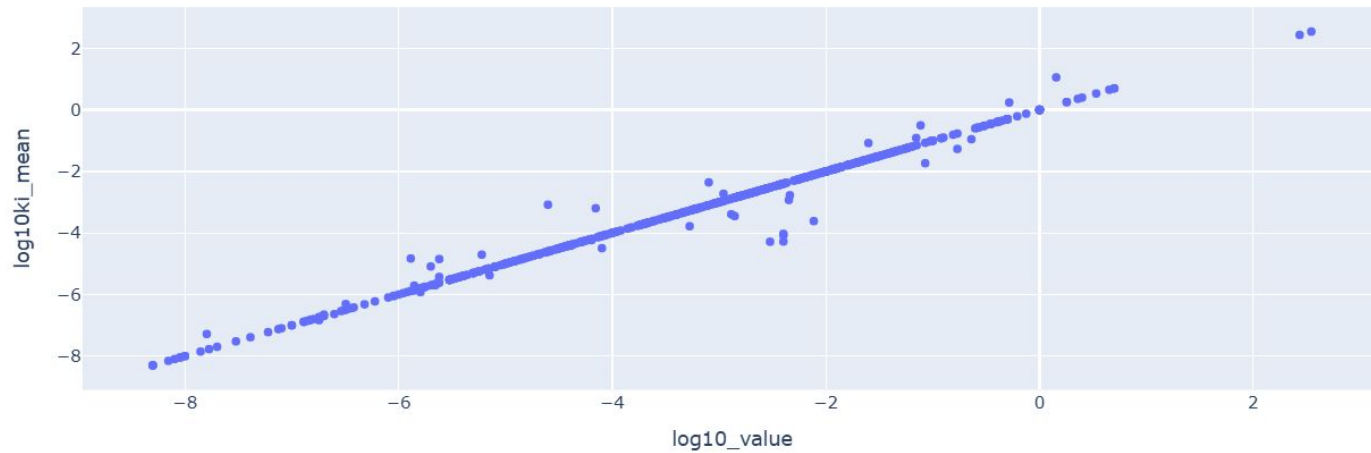


# Graphs



## Graphs (continued)

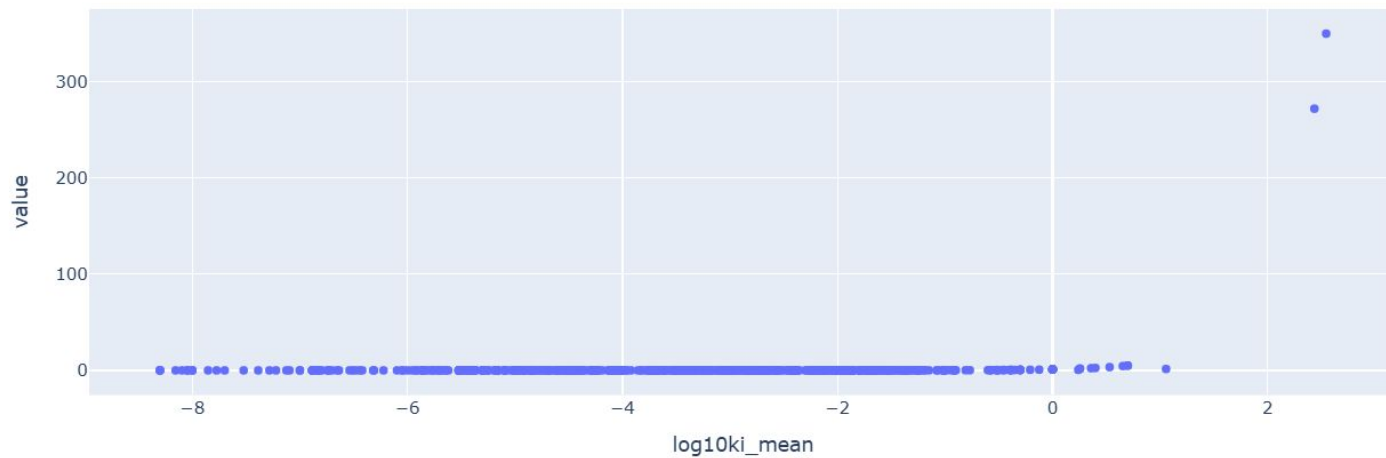
Scatter Plot: log10ki\_mean vs log10\_value





## Graphs (continued)

Scatter Plot: value vs log10ki\_mean





# Challenges

- Learning how to use Python
- Learning pace of research
- Docking



# Future Goals

- Continued project goals
  - Continue docking and docking analysis
  - Begin MD simulations and trajectory analysis
  - ML model building and evaluation
- Personal goals
  - Continue learning more about coding
  - Better understand docking
  - Learn to perform MD simulations and trajectory analysis



# Citations

- Di Cera E. Serine proteases. IUBMB Life. 2009 May;61(5):510-5. doi: 10.1002/iub.186. PMID: 19180666; PMCID: PMC2675663.
- Neitzel, J. J. (2010) Enzyme Catalysis: The Serine Proteases . *Nature Education* 3(9):21
- Prasanna S, Doerksen RJ. Topological polar surface area: a useful descriptor in 2D-QSAR. Curr Med Chem. 2009;16(1):21-41. doi: 10.2174/092986709787002817. PMID: 19149561; PMCID: PMC7549127.



# Acknowledgments

Funding for the SIMCODES REU comes from NSF