

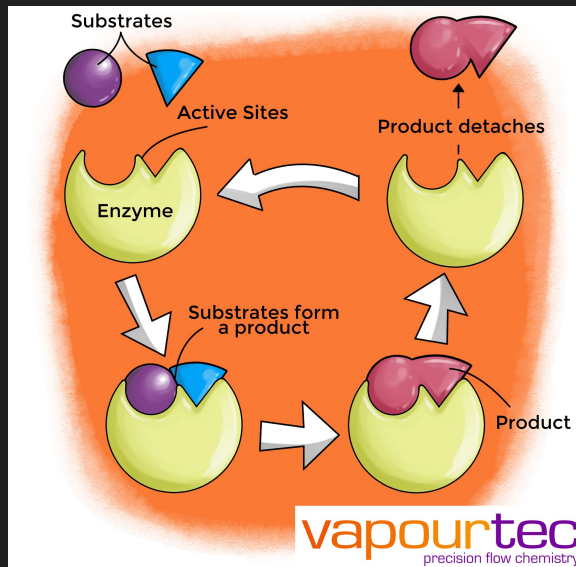
# Automating the Fragmentation of Proteins

SIMCODES 2025

Mentors: Theresa L. Windus,  
Ryan M. Richard  
Daniel Woodard

# Goal

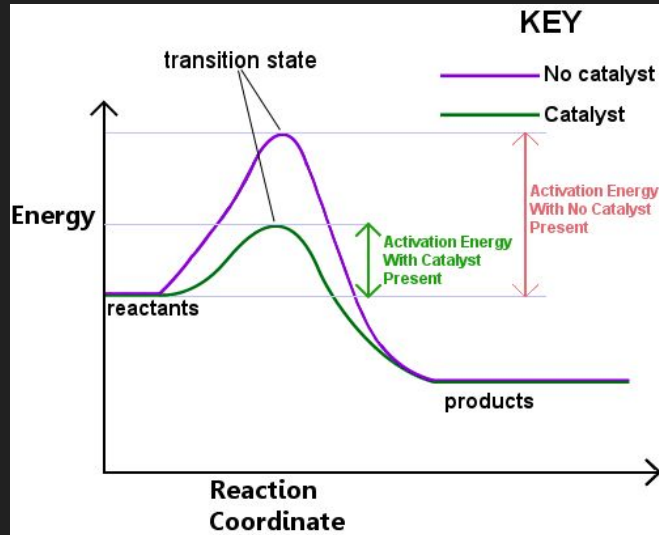
To predict the energy of enzymatic reaction using AI/ML for better predictions.



# Background

Enzymes are biological catalyst that helps with the functions inside of the body.

Enzymes are made up of proteins and amino acids.



## Structure of Amino Acid

MN

Glycine	Alanine	Serine	Threonine	Cysteine
Valine	Leucine	Isoleucine	Methionine	Proline
Phenylalanine	Tyrosine	Tryptophan	Aspartate	Glutamine
Asparagine	Glutamine	Histidine	Lysine	Arginine

# Challenges and solutions

It is difficult to simulate enzymes and proteins.

As the size of the molecule increase, the cost and time proportionally increase also.

Method of fragmentation using Hartree Fock method and Self Consistent Field approximation.

# Progress

Python script for PDB files.

Python script for xyz to nw files, energy reader, and difference.

Run job using the nova system using NWChem

```
3
Water
H    0.7493682    0.0000000    0.4424329
O    0.0000000    0.0000000   -0.1653507
H   -0.7493682    0.0000000    0.4424329|
```

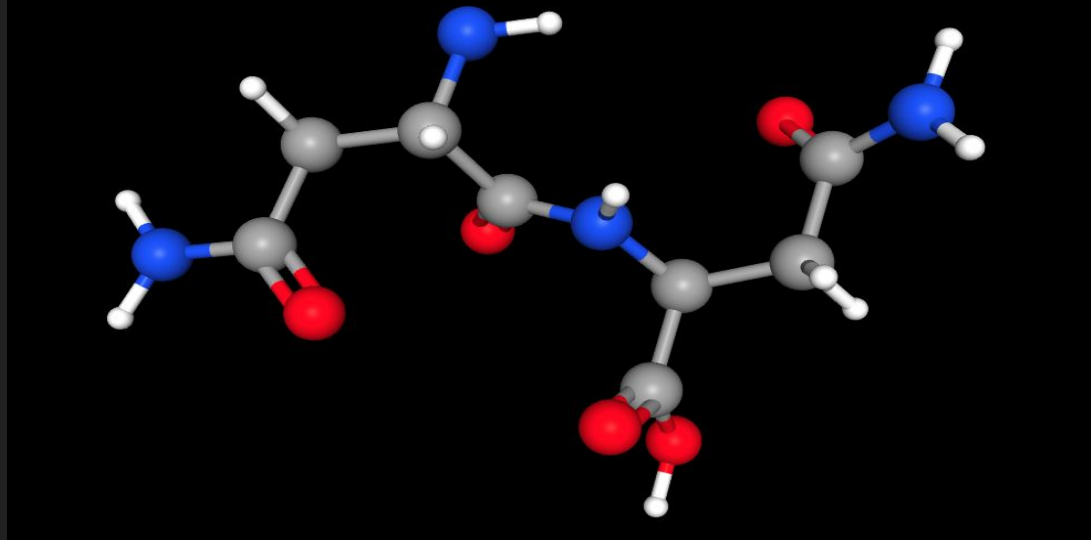
```
start h-test-sto
geometry
H    0.7493682    0.0000000    0.4424329
O    0.0000000    0.0000000   -0.1653507
H   -0.7493682    0.0000000    0.4424329
end
```

```
basis
| * library sto-3g
end
```

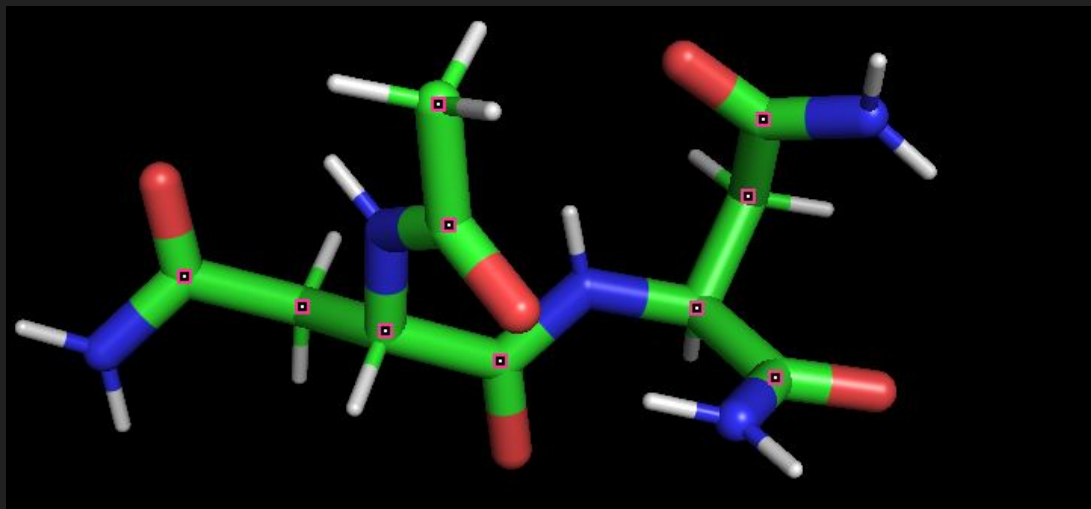
```
task scf optimize
task scf freq
```

# Example ASN ASN

Top is from [PubChem](#)



The bottom is from PyMOI of xyz file.



# What I have learned / most challenging

Python.

How to use and navigate Unix.

Interact and navigate a server.

Molecular modeling software.

Basics Quantum and computational chemistry.

# What I hope to accomplish

To finish running the rest of the amino acids.

Fully simulate one protein from the amino acids.



# Acknowledgements

NSF

SIMCODES program ISU.

# Citation

<https://www.medicalnewstoday.com/articles/319704#what-do-enzymes-do>

[https://simcodes-isu.github.io/projects/fragmenting\\_protein/](https://simcodes-isu.github.io/projects/fragmenting_protein/)

National Center for Biotechnology Information (2025). PubChem Compound Summary for CID 16122514, Asn-Asn. Retrieved July 10, 2025 from <https://pubchem.ncbi.nlm.nih.gov/compound/Asn-Asn>.

<https://github.com/nutjunkie/IQmol/blob/master/share/fragments/Molecules/Water.xyz>