IOWA STATE UNIVERSITY
REU Project – Department of Computer Science and Chemistry

Charlie Klawitter

SIMCODES REU Program
July 31, 2025

Machine Learning-Enhanced Computational Modeling of Metal-Protein Interactions

Abstract

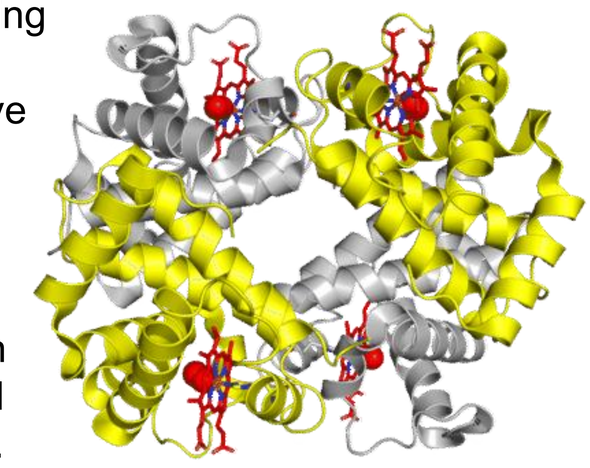
Metals are essential to biological processes, ranging from enzyme catalysis to DNA binding. Classical methods of modeling their interactions with proteins and solvents fail to capture the complexity of metal-ion binding. These approaches limit progress in drug discovery and bioengineering. This research aims to validate and refine a quantum mechanical computational approach that combines chemistry and machine learning. The goal is to use machine learning to enhance the modeling of metal-protein interactions with accuracy and efficiency.

Objectives

1. Validate a fragmentation-based computational strategy for modeling metal-protein interactions.
2. Integrate machine learning to enhance predictions of metal binding in biological systems.
3. Create a scalable workflow for future large-scale screening of metal-targeting therapeutics

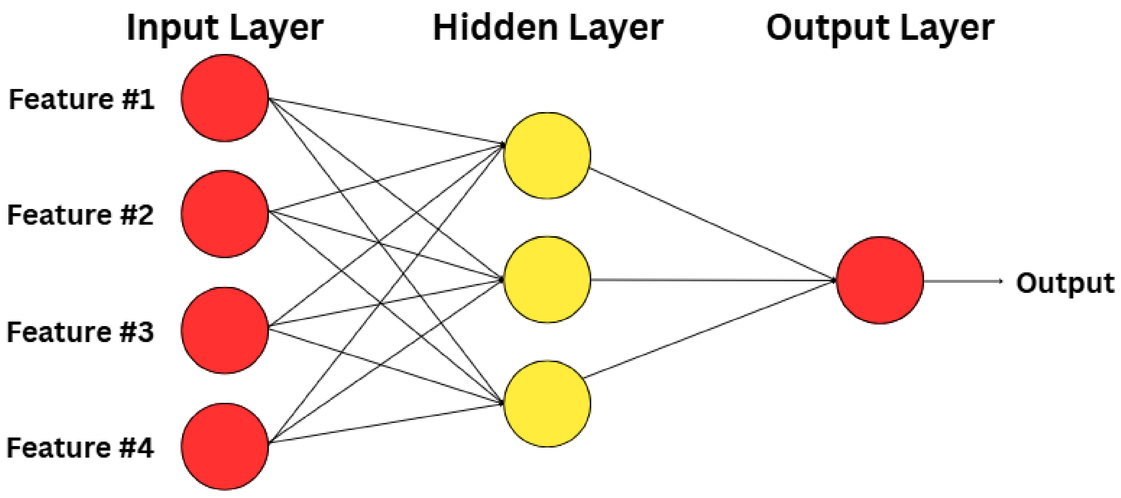
Data

The data developed for training the machine learning model consists of four key components of the Effective Fragment Potential (EFP) method. The Undamped Electrostatic Energy, Intermolecular Overlap Integral, Distance between Centroids, and the Ground Truth Electrostatic Energy.



Methods

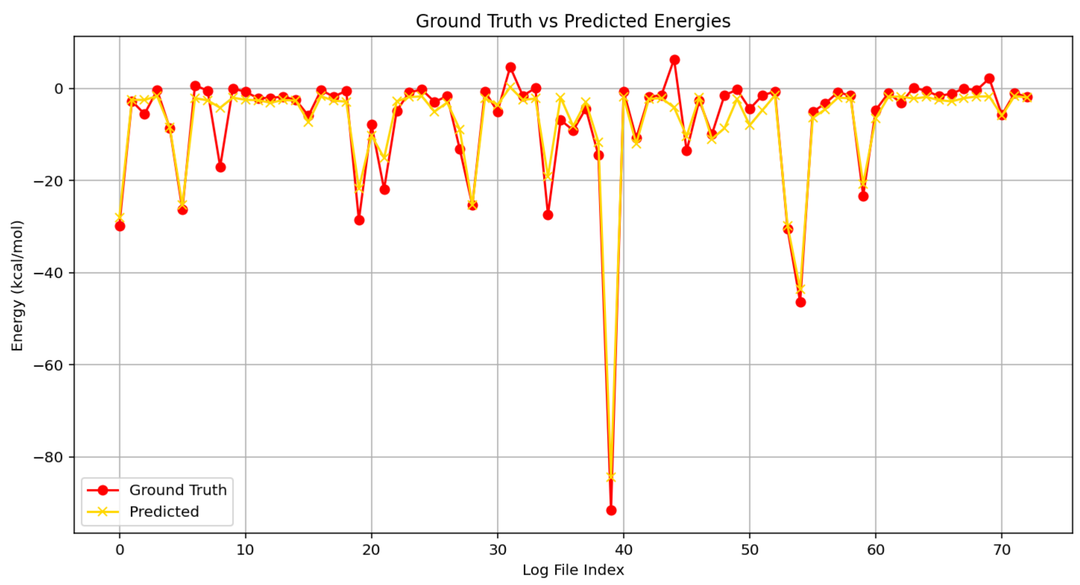
- Model: Neural Network implemented in PyTorch with hidden layers
- Activation Function: LeakyReLU
- Regularization: BatchNorm and Dropout
- Training: 287 samples, 80/20 train/test split, MSE loss, Adam optimizer, and 1000 epochs
- Evaluation: MAE, RMSE, and R^2

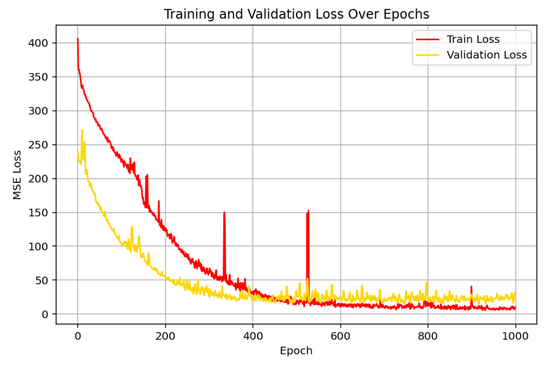


Results

Running the Neural Network yields high accuracy in a matter of seconds. Using K-fold cross-validation, the model returns results of:

- MAE: 1.9253
- RMSE: 3.075
- R^2: 0.969





Sample Neural Network Energy Prediction Table

Sample Log Index	Ground Truth Energy	Predicted Energy	Error
16	-12.657	-12.556	0.102
1	-0.592	-0.592	0.136
28	-2.001	-1.812	0.183
17	-0.531	-0.763	0.232
53	-0.004	-0.343	0.346
44	-0.581	-0.948	0.368
55	-23.372	-23.748	0.377
25	-1.741	-2.124	0.383
21	-10	-9.789	0.211
41	-1.891	-1.913	0.022

Conclusion

This research demonstrates that neural networks can accurately and efficiently predict key energy components relevant to the EFP method, establishing a strong foundation for future machine learning enhancements. Specifically, it lays the groundwork for predicting Polarization and Dispersion energies. The key is that machine learning can efficiently support computational modeling of metal-protein interactions. This project highlights the growing connection between quantum chemistry and machine learning, providing a promising future for interdisciplinary research.

References

1. Dürr, S. L., Levy, A., & Rothlisberger, U. (2023). Metal3D: A general deep learning framework for accurate metal ion location prediction in proteins. *Nature Communications*, *14*(1).
2. Gordon, M. S., Freitag, M. A., Bandyopadhyay, P., Jensen, J. H., Kairys, V., & Stevens, W. J. (2000). The effective fragment potential method: a QM-based MM approach to modeling environmental effects in Chemistry. *The Journal of Physical Chemistry A*, *105*(2), 293–307.
3. Kim, Y. L., Han, Y., Evans, J. W., & Gordon, M. S. (2021). Effective fragment potential-based molecular dynamics studies of diffusion in acetone and Hexane. *The Journal of Physical Chemistry A*, *125*(16), 3398–3405.

Scan to view full code & documentation on GitHub

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2348724.