Automated Generation of Transition Metal containing Active Sites

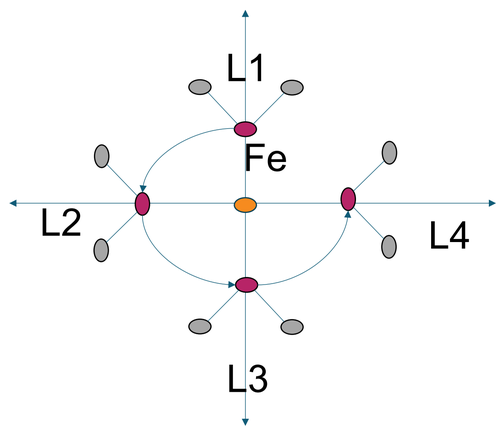
Robert Stahl, Computer and Information Sciences, University of Delaware, Newark, DE 19711
Mentor: Atiya Banerjee, Department of Chemistry, Brookhaven National Laboratory, Upton, NY 11973

Abstract

The geometrical makeup of metal active sites in proteins present an important opportunity for structure generation and machine learning in modern computational theory. Current approaches for automatic generation of these sites in the chemical language makeup suffers from a lack of information related to the unique electronic aspects of transition metals. In order to gain better control over the construction of these sites, the present work consists of using principles of graph theory chemistry to automate the generation of these active sites efficiently. Our methodology uses connectivity theory in addition to empirical information obtained from quantum mechanical considerations to arrive at a representative structure. This approach has been successful in generating models of diiron active sites connected with multiple different bridge structures as well as monoiron active sites. These models have been capable of successfully extracting electronic properties from chemical computations.

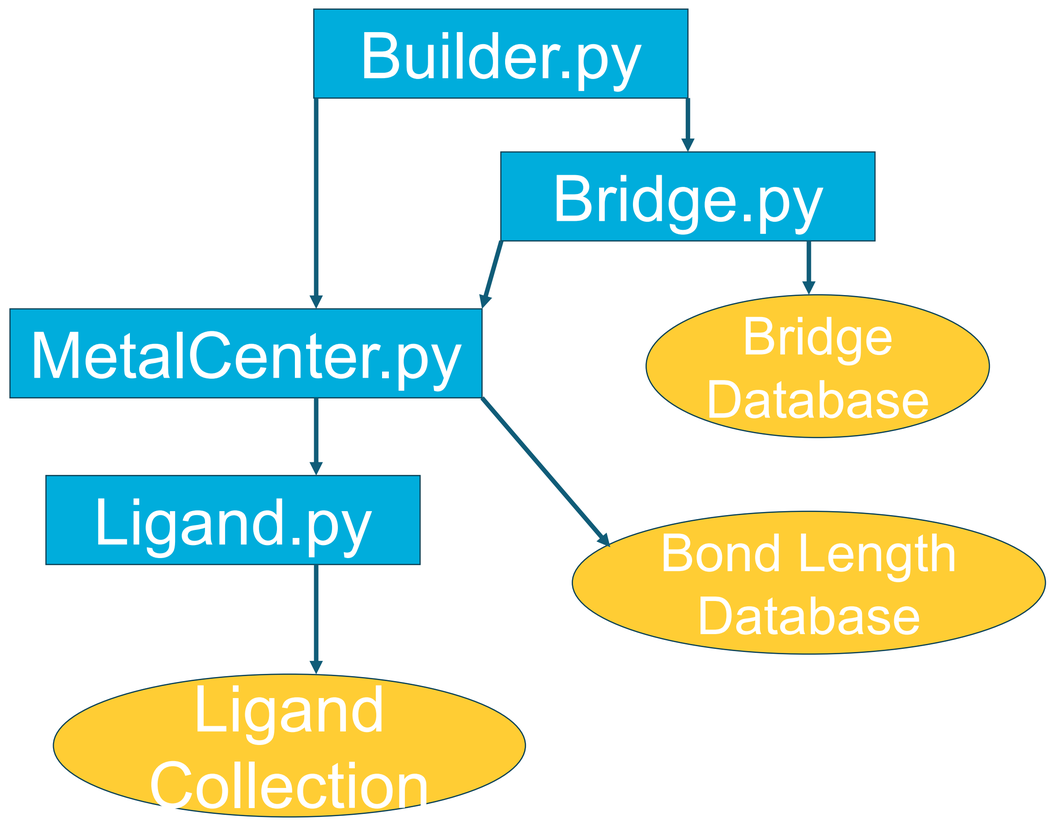
Introduction

- For the formation of active sites, principles of graph theory can be applied in order to generate any active site. Any chemical structure like tetrahedral or octahedral can be created by performing translations on planar ligands in order to properly place them relative to a metal center. The visual on the right shows how rotation can create the ligand placements on 2 planes for octahedral structures.
- From this, additional complexity can be added using any other translations required to form a proper structure. Irregularities can be accounted for using algorithmic transformations, and additional structure information like bridge connections can be loaded in to properly form the active site.
- Ideally this software will be a tool to further other research by providing reliable model generation methods for these active sites.



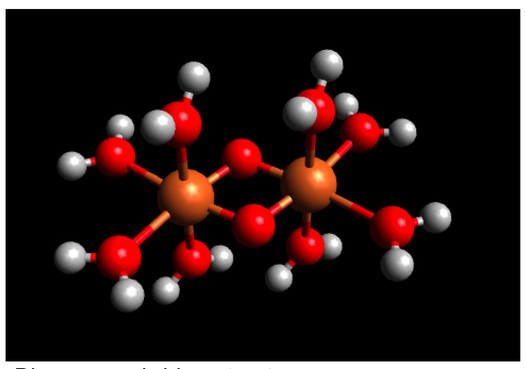
Methods

- Construction of three databases: a bridge database containing chemical properties of oxygen bridges to connect the metals, a bond length database to hold the bond distances between the metal atoms and the ligands attached to them, and a collection of planar ligands to load and attach to the metals.
- Creation of a software package for holding the chemical information and constructing the active site based on inputs from the user.

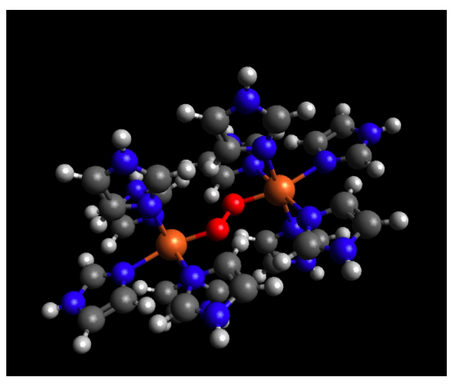


- Builder controls the overall building process. This is the script the user runs and adjusts the active site as they like. From there, if it is a diiron system it creates a Bridge object in order to generate a bridge structure and calculate the two metal coordinates. If it is a monoiron system, Bridge directly calls a MetalCenter object at the origin and constructs ligands off of this center.
- From the Bridge object, two MetalCenter objects are created and stored. These MetalCenters create and attach the Ligand objects that are attached to it. This is accomplished by reading in ligand Protein Data Base (.pdb) files from the ligand collection and performing transformations on the Ligand object in order to align it properly before attaching it.
- Once all of these objects have been created and adjusted, Builder reads in the information concerning all the atoms and connections and creates a Protein Data Base file containing the final active site for use with chemical computations.

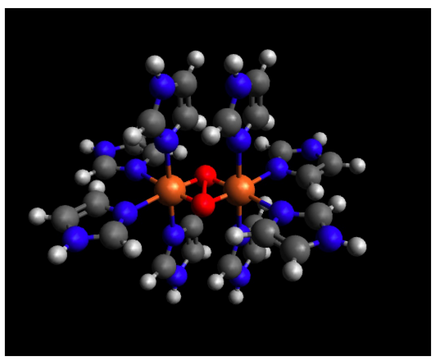
Results



Bis-mu-oxo bridge structure



Trans-1-2-peroxo bridge structure



N2-N2-peroxo bridge structure

- The current package can generate active sites with multiple bridge coordinations and octahedral, bipyramidal or tetrahedral coordinations on the metals.
- The models generated have been successful in being used for computational calculations to gather Mossbauer parameters.
- The resulting software can be combined in the future with machine learning in order to efficiently form structures for further research. Machine learning on chemical properties could recommend a structure that can be generated using this software package. From there these models can be run into chemical computations in order to further optimize the complex.

Conclusion

- The current software package is capable of generating active sites containing multiple mental centers with connecting ligand coordinations. This is a useful tool for computational chemists seeking to perform research on these active sites that lacked generation methods through chemical languages like SMILES.
- Further expansion upon the software could include expanding current functionality within the software itself as well as integrating the model generator with machine learning techniques. Expanding the implementation would take the form of added functionality to integrate bidentate and tridentate ligand placements as well as further algorithmic complexity to account for irregularities in chemical structures. In addition, this generation software could be combined with machine learning to formulate software capable of analyzing the chemical properties of an active site and generating potential structures for further analysis.

Acknowledgements

I would like to give a special thanks to my mentor Dr. Atiya Banerjee for his guidance and support throughout this project. I would also like to thank Dr. Mehmed Ertem for his insights that aided me to create the project. I would also like to extend my thanks to fellow interns Bryana Lopez and Ambika Natarajan for their consistent help and positive attitude. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

Reference

- Hohenberger, J., Ray, K. & Meyer, K. The biology and chemistry of high-valent iron–oxo and iron–nitrido complexes. *Nat Commun* **3,** 720 (2012). https://doi.org/10.1038/ncomms1718
- Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, Philip E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, Volume 28, Issue 1, 1 January 2000, Pages 235–242, https://doi.org/10.1093/nar/28.1.235
- F. Riaz and K. M. Ali, "Applications of Graph Theory in Computer Science," 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, 2011, pp. 142-145, doi: 10.1109/CICSyN.2011.40.

U.S. DEPARTMENT OF ENERGY
UNIVERSITY OF DELAWARE
www.bnl.gov
Brookhaven National Laboratory