Data parsing for optimized molecular geometry calculations

Luke Rens, Dr. Christopher Ivancic, Dr. Brian Barngrover

Stephen F. Austin State University, Department of Computer Science, Nacogdoches, TX 75962

Abstract

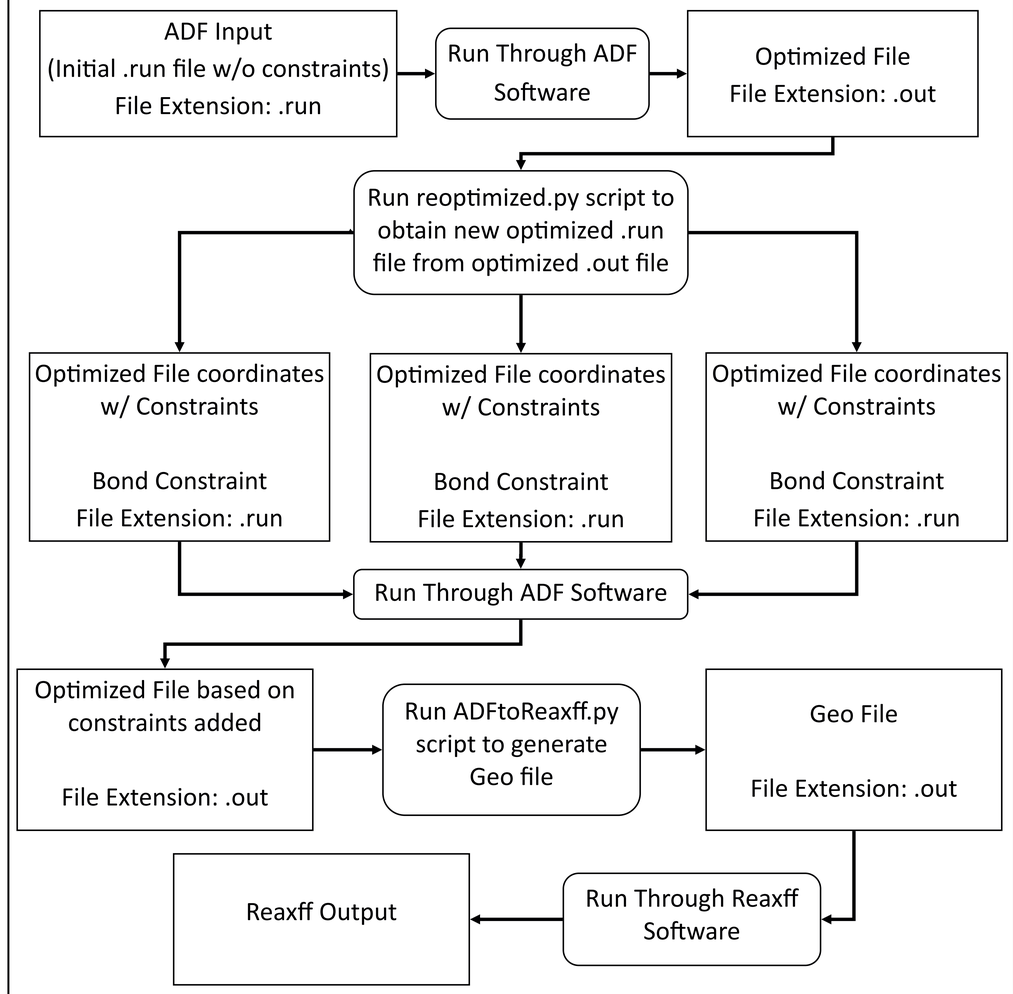
Amsterdam Density Functional (ADF) is software that understands and predicts structures, reactivity and spectra of molecules. When used in conjunction with Reactive force fields (ReaxFF) software chemical reactions involving those molecules based on the reactive force field approach. The text parsing scripts developed in Python were created to bridge the gap between ADF and ReaxFF.

Purpose

The purpose of this project is to optimize and streamline to process of using ADF and ReaxFF. There is no efficient way to effectively add constraints to a compound and run it through ADF, take the ADF output and create a file that can be run through Reaxff, then take that Reaxff output and come to conclusions on it. To streamline this process, scripts were developed using Python to parse information out of data generated by ADF.

$$E_{system} = E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + E_{coa} + E_{C2} + E_{triple} + E_{tors} + E_{conj} + E_{H\text{-}bond} + E_{vdWaals} + E_{Coulomb}$$

Optimization Process Design



Language

For this project a critical part is choosing which programming language to use in achieving success. Since the project calls for reading in large amounts of non-uniform data and parsing through said data, Python was chosen as its efficiency outweighs other languages such as C++ of Java. With Python the ability is gained to write scripts that are 100-200 lines that would take other languages 800-900 lines to complete the same task. This leads to more efficient and readable code that also makes maintainability more achievable.

Conclusion

Scripts were successfully able to bridge the gap between ADF and ReaxFF using Python. For the next phase of research, new text parsing scripts to be developed to replace the Monte Carlo Force Field Optimizer for finding best fit force fields for a training sets more efficiently.

Acknowledgments

- Stephen F. Austin Department of Chemistry and Biochemistry
- Stephen F. Austin Department of Computer Science
- Office of Research and Sponsored Programs