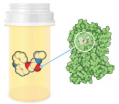
Improving Molecular Modeling through Advances in Force Field Development

Daniel Cabrera, Jessica Maat, Dr. David Mobley

Department of Chemistry, University of California Irvine - Software Engineering Major

Abstract

- Force fields are the underlying technology in molecular simulations that allow for predictive measures of different physical properties
- These molecular simulations have applications in computer aided drug design through accurate predictions of protein and ligand behavior
- Force fields consist of potential energy functions that describe bonded and non-bonded molecular interactions
- In order to improve, benchmark, and train force fields, chemically diverse datasets are important
- DANCE works to generate diverse datasets of molecules using fragment based fingerprints
- This allows for molecules to be filtered for specific molecular substructures and properties for force field development

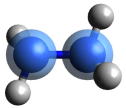


$$W_{AB} = \sum_{\mu \in A} \sum_{\nu \in B} P^2 \mu\nu$$

Wiberg bond order equation
SOURCE: docs.openeye.com

Introduction

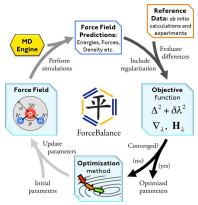
- Chemists make use of tools, such as ForceBalance, in order to parameterize force fields
- DANCE aids in this process by providing chemists with diverse datasets for parameterization
- This ultimately allows for improved molecular modeling and increasingly accurate predictions about molecular behavior



Single nitrogen-nitrogen bond
SOURCE: avogadro.cc

[#7:1]-[#7:2]

Single nitrogen-nitrogen bond
SMIRKS pattern



Wang et al. *J. Phys. Chem. Lett.*2014511885-1891

- DANCE currently implements fragment based fingerprinting for impropers and select torsion parameters
- This project extends this method to bonds, specifically single nitrogen-nitrogen bonds

Methods

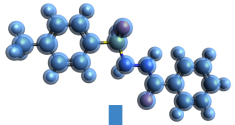
Develop a customizable fingerprint method to filter molecules based on their Wiberg bond order and neighboring atoms

↓

Curate a diverse dataset of molecules with single nitrogen-nitrogen bonds using the customizable DANCE fingerprint

↓

Perform chemical diversity analysis on the resulting datasets using MACCS key fingerprint and bond length



Single Nitrogen-Nitrogen bond: 7-7 (blue)
Neighbor 1: 6 Carbon (dark grey)
Neighbor 2: 16 Sulfur (yellow)
Neighbor 3: 1 Hydrogen (light grey)
Neighbor 4: 1 Hydrogen (light grey)
Wiberg bond order: 0.9447705833913734

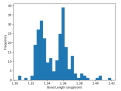
Fingerprinted Cc1ccc(cc1)S(=O)(=O)NNC(=O)c2ccccc2 molecule
SOURCE: avogadro.cc

Results: Bond Length

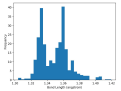
- We have established a customizable fingerprint for molecular bonds
- Used this fingerprint to generate a diverse database of molecules with single nitrogen-nitrogen bonds based on a fingerprint
- Analyzed the diversity of the dataset by comparing against a dataset of randomly selected molecules
- Bond length values appeared relatively similar between the two datasets
- MACCS key fingerprint values varied heavily, with more diverse values in the fingerprinted dataset

Bond Length Comparisons

- Fingerprinted Dataset



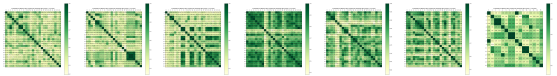
- Random Dataset



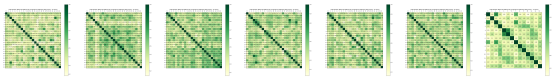
Results: MACCS Key Fingerprint

MACCS Key Fingerprint Comparisons

- Fingerprinted Dataset



- Random Dataset



Outlook

- DANCE proved to be effective at filtering datasets for specific molecular traits and electronic properties
- DANCE also demonstrated the ability to effectively generate diverse datasets using customizable fingerprinting
- Moving forward, we are looking to develop new fingerprinting methods and expand the molecular properties that DANCE is able to fingerprint for

Diagram of the DANCE pipeline steps.
SOURCE: dance.readthedocs.io

Acknowledgements

I'd like to thank Dr. David Mobley and Jessica Maat from the Mobley Group for welcoming me with open arms to the Mobley Lab and for their perpetual support and assistance on this project. I'd also like to thank Bryon Tjanaka for his help acclimating me into the DANCE program and the Mobley Lab. Finally, I would like to thank UROP for recognizing my project and providing me with the opportunity to present my research.