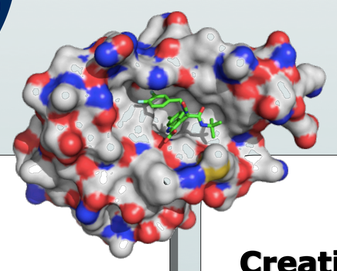
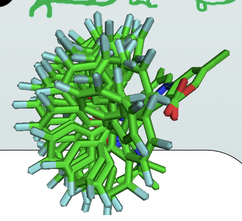
Knowledge-Based Ligand Conformer Generation for Virtual Drug Screening

Rafael Bergerman, Dr. David Koes
Drug Discovery, Systems and Computational Biology (DiSCoBio) Academy
University of Pittsburgh, Pittsburgh, PA 15260





Abstract

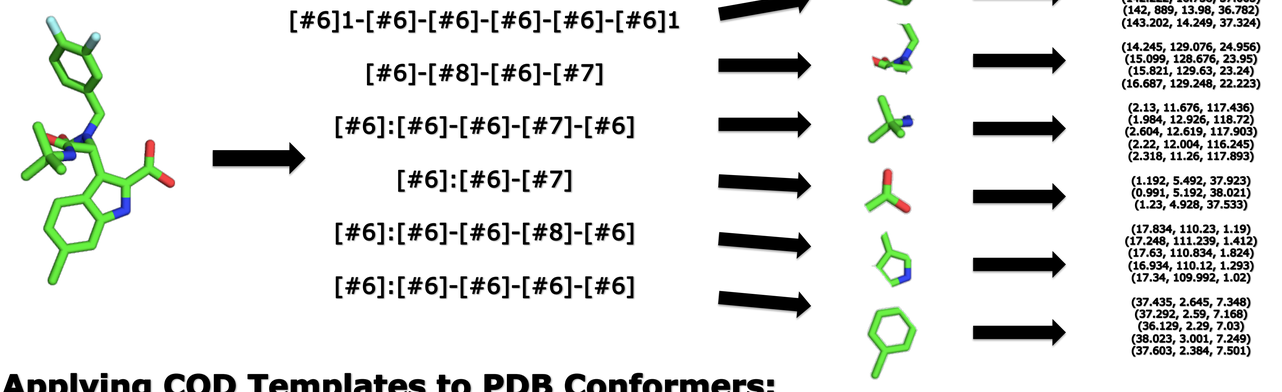
Conformer generation is the computational prediction of three-dimensional molecular conformations. Conformer generation is a critical component of virtual screening methods since these methods of computationally identifying potential drug-protein interactions require realistic three-dimensional structures.

Current methods of conformer generation are energy-based, where conformers are generated to minimize an empirically determined energy function. An alternative approach is enabled by the recent explosion of structural data, as exemplified by the Protein Data Bank (102,158 structures) and the Crystallography Open Database (293,056 structures). In this knowledge-based approach, conformers are generated by stitching together pieces of experimentally determined structures.

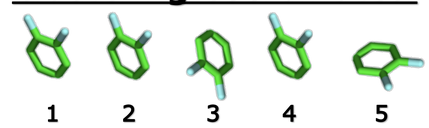
In this project we analyze the PDB and COD and show that the majority of bioactive conformers in the PDB can be reconstituted using experimental data from the COD.

Evaluating Knowledge-Based Conformers

Creating Fragment Template Library from COD:



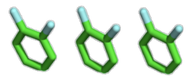
Clustering Coordinates:



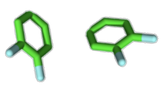
1 → 2, RMSD = 0.00450568
1 → 3, RMSD = 0.744103
1 → 4, RMSD = 0.0120437
1 → 5, RMSD = 0.747406

3 → 4, RMSD = 0.4738029

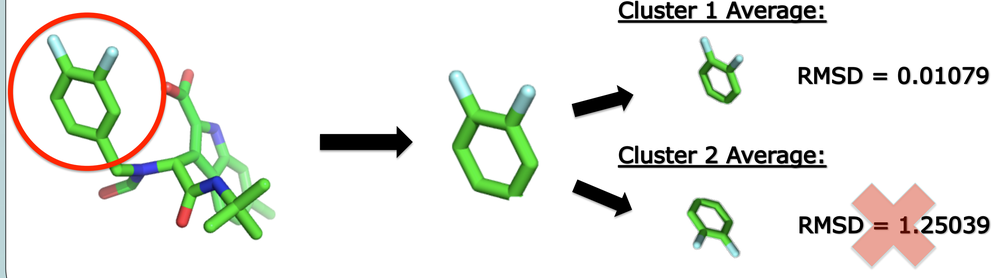
Cluster 1



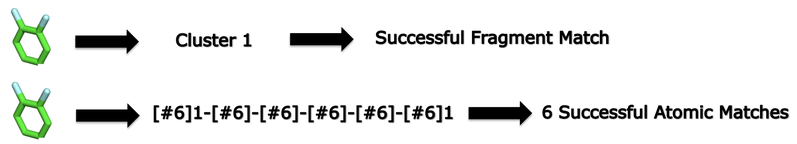
Cluster 2



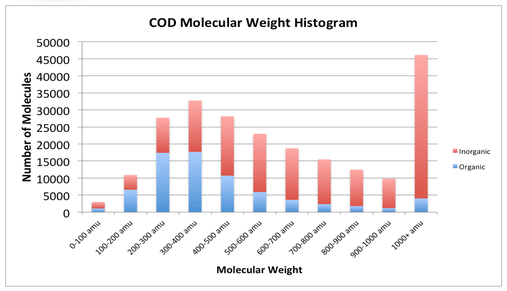
Applying COD Templates to PDB Conformers:

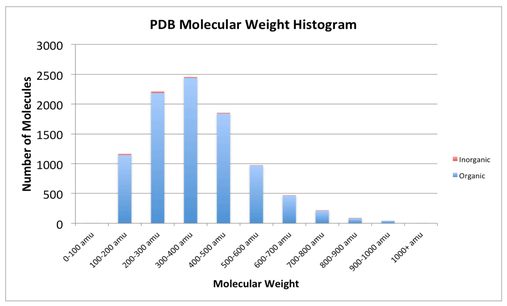


Atomic vs Fragment Coverage:

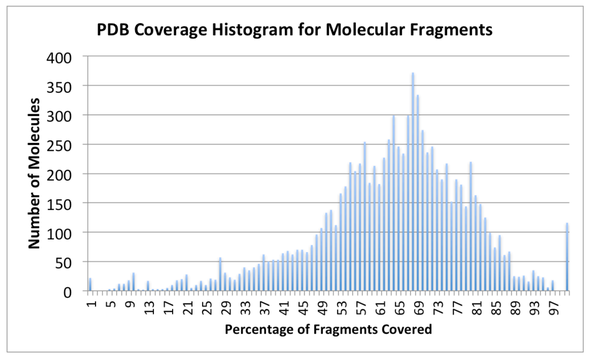


Preliminary Data

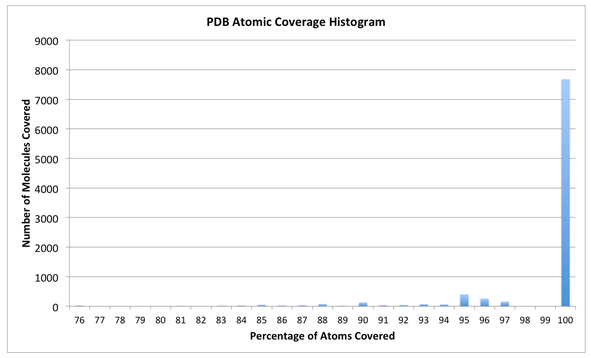




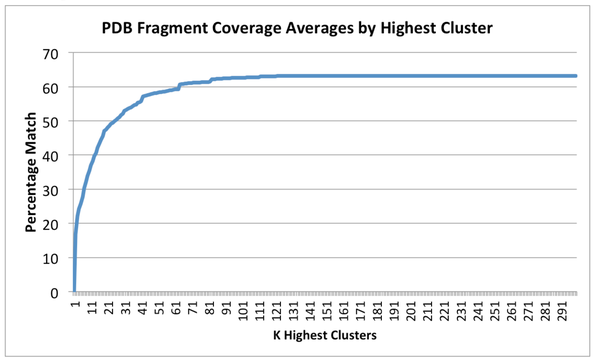
PDB Analysis from the COD Database



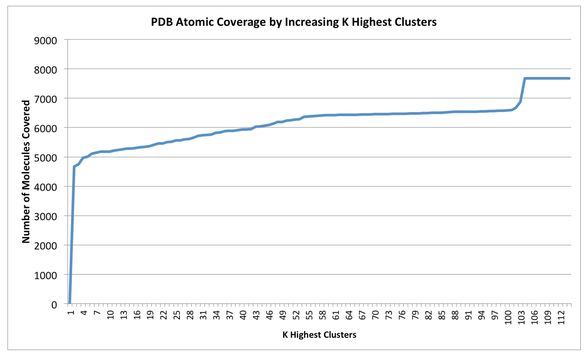
This figure shows the coverage histogram for the PDB database fragments. All the compounds were broken down into fragments and analyzed to see what percentage of the fragments were actually present in the COD within an RMSD of 0.5 Angstroms.



This figure shows the coverage histogram for the PDB database atoms. All PDB molecules were analyzed to see what percentage of atoms were covered by the fragments in a single molecule.



This figure shows the percentage match of the average molecule in the PDB Database based on the most popular clusters found from the COD. The X-Axis follows the increase in k, the highest cluster analyzed, while the Y-Axis follows the percentage match of a molecule to the K highest clusters.



This figure shows the atomic coverage in the PDB Database as the K highest cluster is increased. While the graph above analyzes the entire database with a maximum K value into a histogram, this graph follows the progression of the atomic coverage as K is increased.

Conclusions

These results suggest that knowledge-based conformer generation methods can be successful in reconstituting PDB ligands from the experimental data in the COD.

Future Directions

Conformers now need to be generated and tested to measure the efficacy of these results.

Acknowledgements

- David Koes
- Joseph Ayoob, Tim Lezon, Leah Russell, Mark Schurdak, and the entire DiSCoBio Program
- Megan Houlihan, Michael Lotze, and the entire UPCI Academy staff
- Sean Geiger and my family

References

"Biological Macromolecular Resource." *RCSB Protein Data Bank*. RCSB. Web. 05 Aug. 2014.
"Crystallography Open Database." *Crystallography Open Database*. Vilnius Development Group, 05 Aug. 2014. Web. 05 Aug. 2014.
"Daylight Fingerprinting." *Daylight*. Daylight Chemical Information Systems Inc. Web. 05 Aug. 2014.
"Python." *https://www.python.org* 05 Aug. 2014.