A Temporal Analysis of the Effect of Protein Flexibility on Binding Cavity Shape

Steven Stinson, Ziyi Guo, Trevor Kuhlengel, Brian Chen
Department of Computer Science and Engineering

Background

Protein binding preferences control how living systems operate. Studying binding preferences has many applications, including:

- Designing effective drugs
- Understanding molecular systems
- Reengineering proteins
- Studying disease

Problem

Software designed to compare protein structure has been shown to be effective in determining binding specificity, but they struggle comparing proteins in different poses, or conformations.

Approach

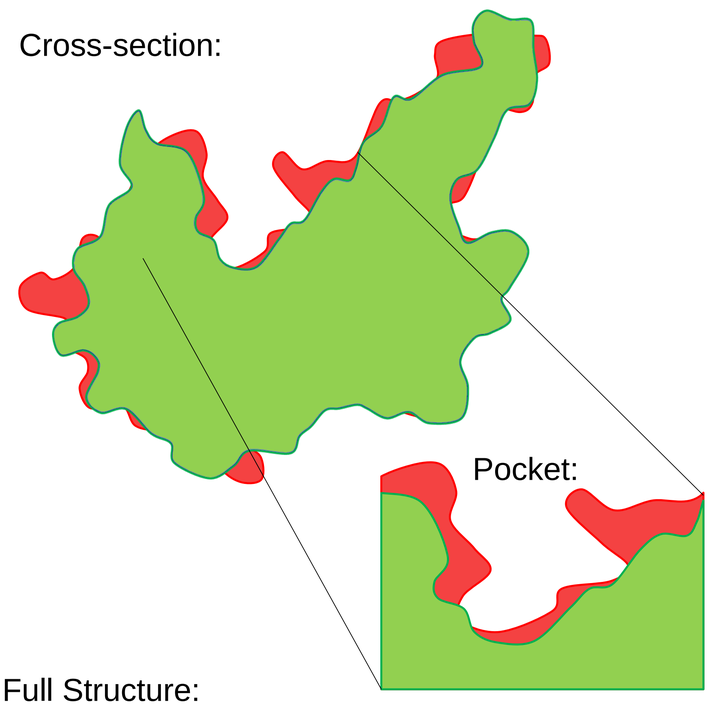
To make such comparisons possible, we sample structural variations from molecular dynamics simulations and build statistical models of the kinds of variations that can occur.

Hypothesis

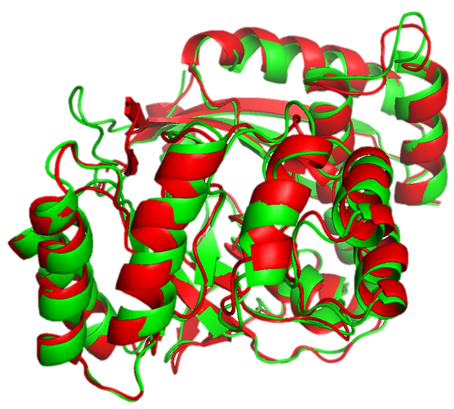
We hypothesize that we can create an accurate model for variation in shape of binding pockets and determine the effect molecular dynamics has on the binding preferences of proteins.

An enolase, PDB id: 1MDR
Red: 1 nanosecond Green: 100 nanoseconds

Cross-section:



Full Structure:



Distributed Molecular Dynamics Simulation

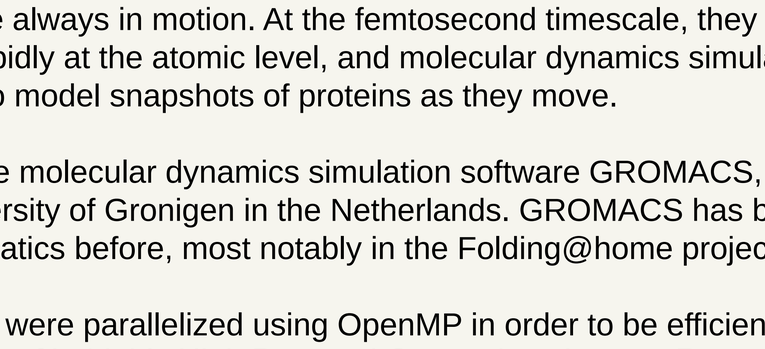
Proteins are always in motion. At the femtosecond timescale, they are vibrating rapidly at the atomic level, and molecular dynamics simulations enable us to model snapshots of proteins as they move.

We used the molecular dynamics simulation software GROMACS, developed at the University of Gronigen in the Netherlands. GROMACS has been used in Bioinformatics before, most notably in the Folding@home project.

Simulations were parallelized using OpenMP in order to be efficiently run on Corona, one of Lehigh's High Powered Computing clusters. Runs occur on four separate nodes, each with 16 cores, meaning a total of 64 processor cores are utilized.

Boolean Set Operations

VASP is a software package designed to allow us to compute set operations (union, intersection, difference) on two aligned proteins as if they were solid objects.



Regions conserved between two proteins may bind similar molecular fragments, but differences in cavity shape will cause different specificity. Variations in protein motion may throw off this analysis, however, so we analyze these motions to determine a bound on this effect.

Experimental Methods

Data Set

We are analyzing subfamilies of the canonical serine proteases; trypsins, chymotrypsins, and elastases; and subfamilies of the enolase superfamily, enolases, mandelate racemases, and muconate laconizing enzymes. Each subfamily exhibits binding preferences that differ from other subfamilies, providing a rich set to test our method.

Training Stage

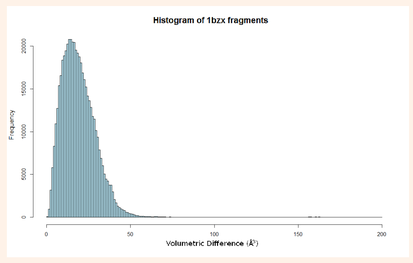
Step 1: Simulate the proteins for 100 million steps, generating snapshots of the protein in motion.
Step 2: Align the snapshots atomically to the original protein, allowing for structural comparison.
Step 3: Generate cavities of each snapshot by analyzing the molecular surface.

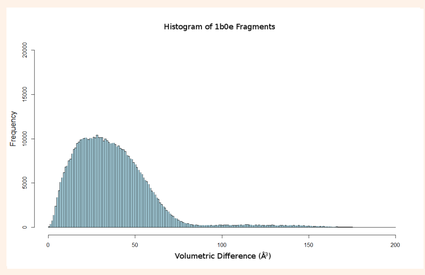
Testing Stage

Step 1: Select two proteins from different subfamilies and simulate.
Step 2: Measure the largest volumetric differences between all pairs of snapshots in each simulation.
Step 3: Statistically analyze the largest volumetric differences.

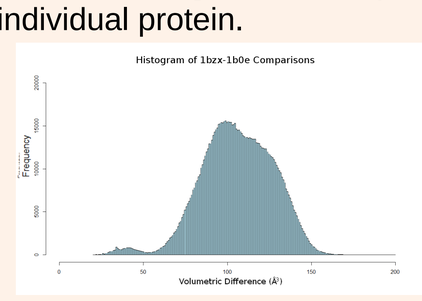
Statistical Analysis

Pairwise comparisons within one protein provide a baseline amount of variation within that protein. The cavity of each snapshot of each protein is compared with every other cavity of every other snapshot of the same protein to measure variation in volume.





We then do comparisons between the two proteins, showing that there is a distinct level of variation that is significant compared to the baseline variation within each individual protein.



Summary

- Cavity shape and therefore binding preferences are significantly affected by protein flexibility
- Molecular dynamics simulation with Gromacs allows us to view this motion
- Volumetric Analysis of all of the "snapshots" of the protein will give insight to the magnitude of these mutations of shape

References

Berendsen, et al. (1995) GROMACS: A message-passing parallel moleculardynamics implementation. *Comp. Phys. Comm.* 91: 43-56.
Chen BY, Honig B (2010) VASP: A Volumetric Analysis of Surface Properties Yields Insights into Protein-Ligand Binding Specificity. PLoS Comput Biol 6(8): e1000881. doi:10.1371/journal.pcbi.1000881
Chen BY, Bandyopadhyay S (2011) VASP-S: A Volumetric Analysis and Statistical Model for Predicting Steric Influences on Protein-Ligand Binding Specificity," bibm, pp.22-29, 2011 IEEE International Conference on Bioinformatics and Biomedicine, 2011

LEHIGH UNIVERSITY