Stephanie Mason · Dr. Filip Jagodzinski · Dr. Brian Chen · Tim Woods

EXPLORING PROTEIN CAVITIES

through rigidity analysis

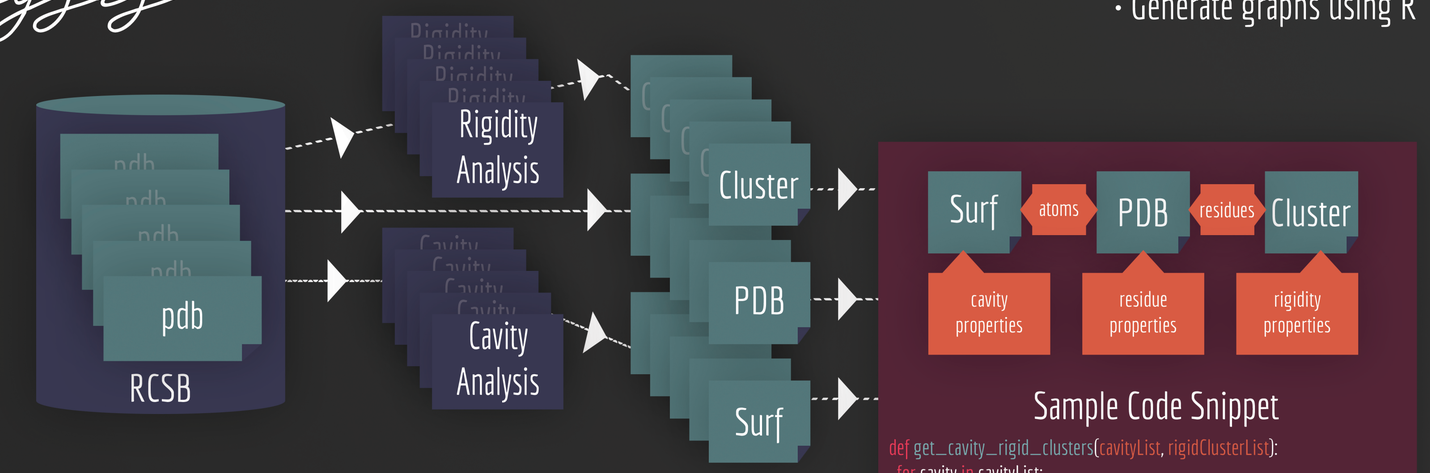
BACKGROUND
and motivation

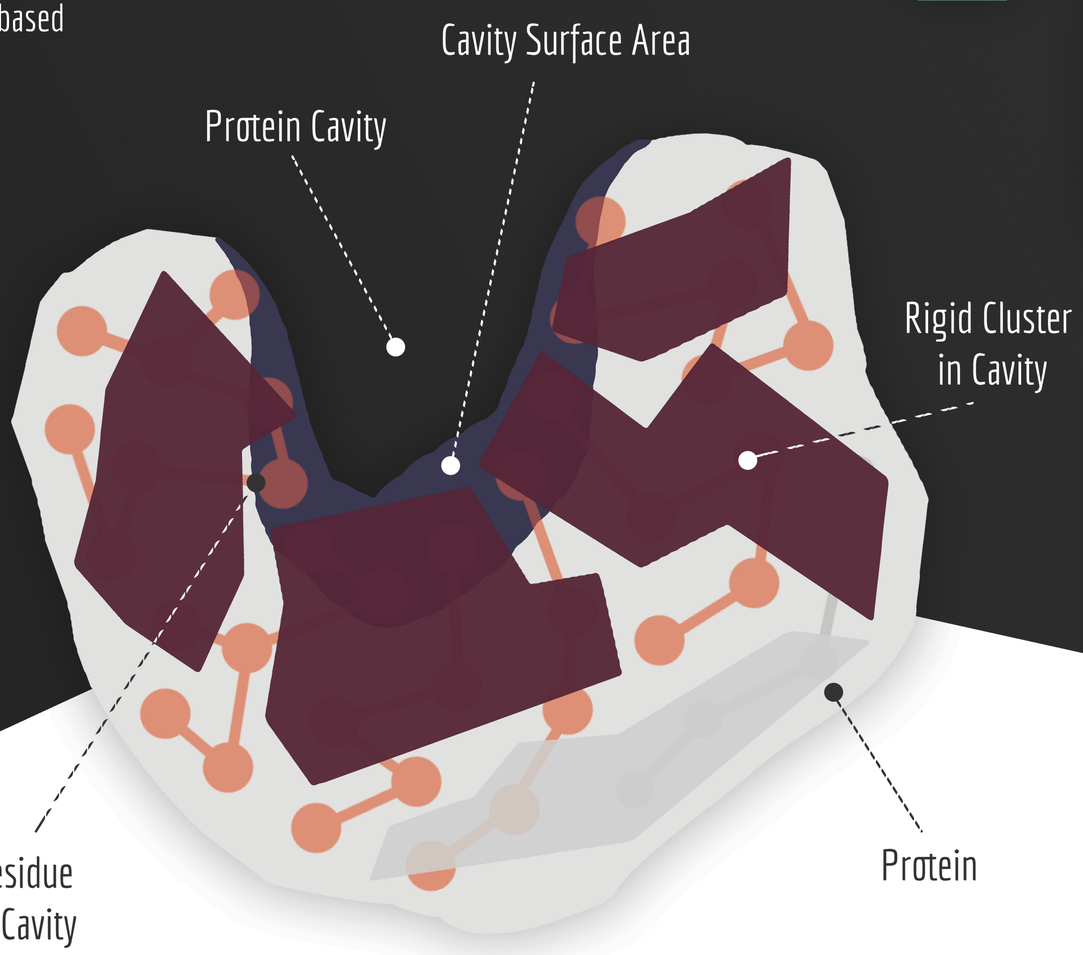
Cavities in proteins facilitate a variety of biochemical processes. The shapes and sizes of cavities are factors that contribute to specificity of ligand binding and docking with other biomolecules. [1] A deep understanding of cavity properties may enable new insights into protein-protein interactions, ligand binding, and structure-based drug design studies. In this work, we:

- Hypothesize that rigidity properties of protein cavities are dependent on cavity surface area
- Explore how biological properties such as size and residue membership of protein cavities correlate with the flexibility of the cavity
- Utilize an existing, efficient graph theoretic rigidity algorithm [2]
- Enumerate a set of cavity rigidity metrics, and demonstrate their use in characterizing over 4,000 protein chains comprising tens of thousands of cavities
- Show that cavity size indeed correlates with some-but not all-cavity rigidity metrics.

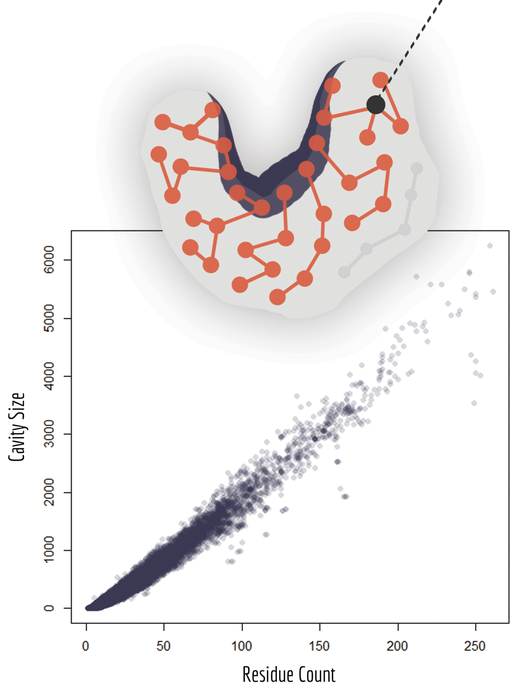
METHODS
and metrics

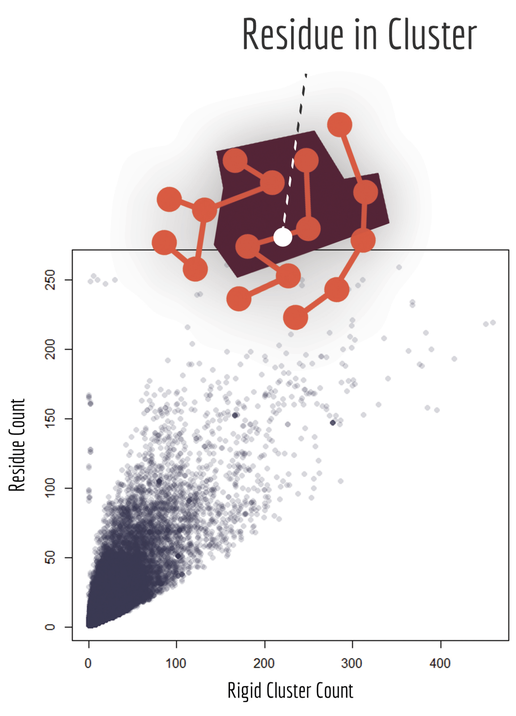
- Obtain protein structure data files from the Protein Data Bank
- Use KINARI (graph theory analysis) [2] to assess rigidity properties of proteins
- Use VASP-E [3] to assess cavity surface area and amino acid identities
- Aggregate high volumes of data output from these programs with Python
- Generate graphs using R

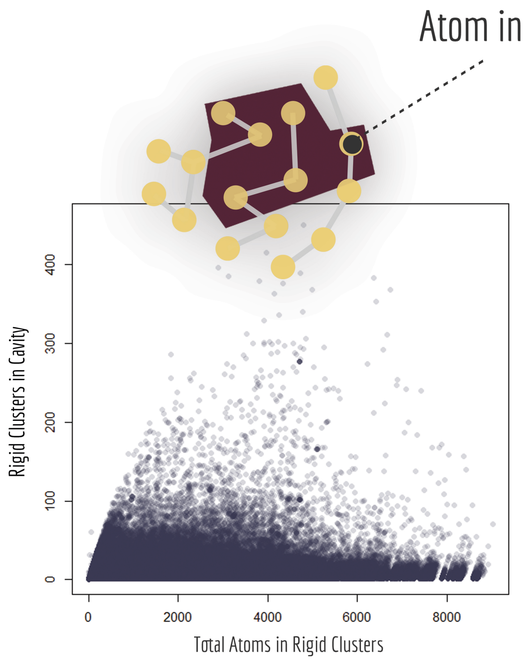




RESULTS
and visualization







Several interesting relationships were observed in our aggregate data. In addition to the relationships shown to the left, we also found that:

- Number of rigid clusters has a positive correlation with cavity size
- Small cavities have no correlation with the size of rigid clusters participating in them
- Larger cavities are composed predominantly of small rigid clusters

CONCLUSION
and looking forward

With an ultimate goal of enabling a large-scale assessment of the rigidity properties of protein cavities we have enumerated a series of cavity rigidity metrics. We have demonstrated their use by exploring 44,342 cavities from among 3,818 proteins chains randomly selected from the PDB. We have found several interesting relationships amongst the metrics of highest interest. This work is still in its early stages.

In the future, we will explore further the correlations that exist among cavity size and other rigidity and biological properties. For example, we will explore how counts of the different types of amino acids that are participating in rigid clusters of a cavity correlate with cavity size. There are many statistical analyses that need to be performed on our aggregate data to better quantify the existing relationships. There are also some interesting regions on the graphs that warrant closer investigation, which will require looking closely at individual proteins and their properties. In the long term, we plan to analyze the majority of the 130,000+ protein structures available from the protein data bank.

[1] Liang, J., Woodward, C., & Edelsbrunner, H. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein science, 7(9), 1884-1897. [2] Fox, N., Jagodzinski, F., & Streinu, I. (2012). Kinari-lib: a C++ library for pebble game rigidity analysis of mechanical models. Minisymposium on Publicly Available Geometric/Topological Software. [3] Chen, B. Y. (2014). Vasp-e: Specificity annotation with a volumetric analysis of electrostatic isopotentials. PLoS Comput Biol, 10(8), e1003792.

WESTERN WASHINGTON UNIVERSITY