

Investigating Rigidity Properties of Protein Cavities

Stephanie Mason
Western Washington University
516 High Street
Bellingham, WA 98225
stephanie.mason@wwu.edu

Brian Chen
Lehigh University
19 Memorial Drive West
Bethlehem, PA 18015
chen@cse.lehigh.edu

Tim Woods
Western Washington University
516 High Street
Bellingham, WA 98225
woodst@wwu.edu

Filip Jagodzinski
Western Washington University
516 High Street
Bellingham, WA 98225
filip.jagodzinski@wwu.edu

ABSTRACT

Cavities in proteins facilitate a variety of biochemical processes. The shapes and sizes of cavities are factors that contribute to specificity in ligand binding, and docking with other biomolecules. 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 explore how biological properties such as size and residue membership of protein cavities correlate with the flexibility of the cavity as computed using an efficient graph theoretic rigidity algorithm. **We hypothesize that various rigidity properties of protein cavities are dependent on cavity surface area.** In this work we enumerate a set of cavity rigidity metrics, and demonstrate their use in characterizing over 120,000 cavities from approximately 2,500 chains. We show that cavity size indeed does correlate with some – but not all – cavity rigidity metrics.

CCS CONCEPTS

•Computing methodologies → ; •Applied computing → *Molecular structural biology*; *Bioinformatics*;

KEYWORDS

Protein; Cavity; Rigidity

1 INTRODUCTION

Many biological functions performed by proteins are known to be dependent on the size, structure, and placement of a protein's cavities. For example, ligand binding sites are strongly correlated with the largest and deepest cavity in a protein, and the specific geometry of a cavity has relevance for ligand design studies[18]. Figure 1 shows a few cavities for typical chains. The size and

geometry of cavities has proven useful in making predictions about protein-protein interactions and protein druggability [10, 16]. Such inferences depend on a large body of training data relating to cavity properties. Where the space of such properties can be expanded to reflect new biophysical characteristics like molecular rigidity, the capability of all inference building algorithms can be enhanced.

Our specific goal is to describe a qualitative and quantitative survey of the rigidity properties of a large set of protein cavities. We leverage the speed of rigidity analysis [12], which builds a mechanical model of a protein from which an associated graph containing nodes, hinges, and bars is constructed. This enables us to calculate structural information about cavities for a large set of proteins in a reasonable amount of time, which we then analyze to infer whether or not over-arching trends are discernible. The rigidity analysis approach that we have chosen is different from other approaches, which either require time and labor intensive work on physical proteins, or else require computationally expensive energetics calculations.

We define several metrics of interest: cavity surface area, residues in cavities, rigid clusters in cavities, and the total number of atoms which are members of rigid clusters that are part of a cavity. New tools for exploring these metrics may provide insights on how cavities are affected by mutations, and how cavities facilitate drug-ligand binding, and protein-protein interactions. In the long term, understanding the relationship between protein cavities and rigidity

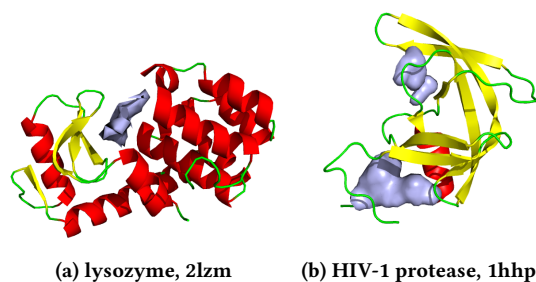


Figure 1: A cavity can be in a deep cleft of a protein (a), or can be near a biomolecule's surface (b). Identified cavities are shown purple. Secondary structures are red and yellow (α helix and β sheets). Rendered in PyMol.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM-BCB'17, August 20-23, 2017, Boston, MA, USA.

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. 978-x-xxxx-xxxx-x/YY/MM... \$15.00

DOI: <http://dx.doi.org/10.1145/3107411.3107502>

analysis can lead to new approaches for *in silico* experiments useful for making predictions about protein function and how it is affected by such variables as residue mutations, drug/ligand binding, or environmental factors. By performing this large scale assessment we begin to understand generalized relationships and create a basis for more complex, detailed computations and biological insights.

2 RELATED WORK

In vitro studies investigating the roles of protein cavities have been conducted on a variety of biomolecules. The scope and breadth of such studies highlights the importance of that work. Experiments elucidating the roles of cavities have been conducted *in vitro* on a variety of physical proteins.

For example, Musah *et al.* have studied the binding thermodynamics of cytochrome *c* [19]. They report on the binding of a small molecule to mutant structures, and reveal that cavities exhibit strong specificity for heterocyclic cations. Among 18 X-ray resolved structures with bound molecules, they showed that cavities induced by the mutations were able to exclusively bind specific molecules. This shows that changes to a cavity’s configuration can drastically affect the binding properties of molecules.

In other work, Bade *et al.*, have studied cytotoxic T lymphocytes and peptide-human leukocyte antigen complexes [1]. They explored the role of a single amino acid substitution, and identified six pockets which play a specificity role in restricting antigen binding. They showed that antigen polymorphism affects which cavities the antigen is compatible with.

In other research highlighting the important role of cavities, synthetic C peptides were engineered to study drug targets of the HIV-1 gp4 envelope protein [4]. Crystal structures of the core of gp41 were resolved, and showed that 3 C helices pack against a central coiled coil when the protein is in its function-activation conformation. The helices make strong contacts with hydrophobic cavities on the surface of the coiled coil.

Binda, *et al.*, studied monoamine oxidase B (MAO B), which due to its role in a variety in neurological disorders, is a common target for antidepressants and neuroprotective drugs [3]. Several beta helices and coils at various positions of the oxidase were identified to play key roles in the protein’s binding affinity with substrates. The active site was identified as a 420 \AA^2 hydrophobic cavity interconnected with a second cavity of almost the same size. This has yielded understanding of the catalytic mechanism of the oxidase, and drug design studies consider the induced fit of ligand-enzyme interactions [2].

In yet other work, Elliott *et al.* studied the serpin family of serine proteinase inhibitors to improve understanding of their roles in the inflammatory, coagulation, and fibrinolytic cascades [7]. They identified five cavities as potential drug targets which were hypothesized to be associated with conformational changes of the proteinase inhibitor family of molecules. The ability of the serine proteinase inhibitors to undergo structural rearrangements enables them to perform their functions, which causes a host of adverse medical conditions such as cirrhosis, thrombosis, and angioedema. Knowledge of the cavities involved has enabled the rational design of drug agents to prevent conformational transitions and pharmacologically control the debilitating diseases they cause.

Computational approaches have also been developed, which aim to identify protein cavities to enable predicting drug binding sites. Such tools often leverage insights from machine learning. SCREEN utilizes a Random Forest approach to identify druggable cavities, analyzing surface cavities of nonredundant proteins crystallized with drugs [20]. MetaPocket 2.0 is a popular web services that predicts drug binding sites with approximately 75% accuracy [22]. Fpocket detects and identifies ligand binding pockets through use of alpha spheres and Voronoi tessellation [9].

These studies show how knowledge of protein cavities has far reaching implications across a variety of research efforts. Due to the labor intensive work needed to conduct studies in physical proteins, research efforts most often focus on a single protein or a class of proteins. Thus, knowledge gained about the characteristics and roles of specific cavities applies only to the biomolecules being investigated. A few large scale studies have been conducted, in which the properties of protein cavities have been categorized by their function in relation to their size [16, 18]. There is a close relationship between the size of cavities and the active site of proteins, with a majority of proteins binding ligands at their largest cavity [16]. Additionally, the number of cavities correlates with the size of the protein. However, even though the largest cavity is the binding site in most cases, certain methods of measuring cavities have revealed there is no correlation between the size of the protein and the size of the active site.

3 METHODS

Our computational pipeline includes several steps (Figure 2). For each protein in our data set of PDB structures, we identify the residues in the cavities on the molecule’s surface and calculate the rigidity properties of each protein. We generate a data set which aggregates structural information from a PDB file, the information output by the cavity detection program, and the rigidity analysis.

We analyzed 2,497 non-redundant chains from a randomly selected group of PDBs, all of which were resolved using X-ray crystallography. From these structures we used a cavity detection radius of 1 solvent Radii, and eliminated all cavities smaller than 1 \AA^2 . We were left with a list of 123,872 cavities.

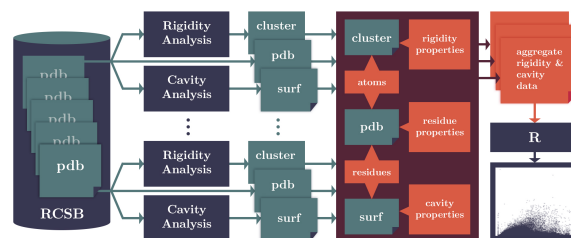


Figure 2: Computation pipeline. We identify cavities among protein chains in PDBs. We use an efficient rigidity analysis approach to identify rigid clusters of atoms in a protein. The structure (pdb), rigidity (cluster), and cavity (surf) files are aggregated to generate details of the rigidity properties of the cavities. From the aggregate metrics we plot various cavity-rigidity-atom properties.

3.0.1 Identifying Cavities. The software *VASP-E* [5] was used to identify cavities and to calculate their surface areas. The surface area is the sum of the areas of all triangles defined to be on the surface based on the molecular surface area [6]. Molecular surfaces are defined as C1-smooth surfaces that surround the atoms of a protein. It is generated conceptually by rolling a ball around the surface, and defined by the region of space that ball cannot occupy. [17] The residues that participate in each cavity were also identified as the set of amino acids closest to the cavity surface. Specifically, the surface is defined as a mesh of triangles and every triangle has an atom of the structure closest to it. That atom is part of some amino acid, which is added to a list. Consider the following output of the cavity calculation program:

```
Pocket 0 (area: 23.60): (2, 4, 5, 7, 12, 45, 46, 47)
Pocket 1 (area: 12.34): (11, 13, 15, 16, 17)
Pocket 2 (area: 6.04): (18, 19, 20)
```

Three pockets are identified, with IDs 0, 1 and 2, with cavity surface areas of 23.60, 12.34, and 6.04 \AA^2 . The largest pocket (ID 0) has participating residues 2, 4, 5, 7, 12, 45, 46, and 47, the second largest pocket has participating residues 11, 13, 15, 16, and 17, and so forth. From among the proteins that we studied, the largest cavity was 6,257 \AA^2 . We did not include any cavities with surface area less than 1 \AA^2 because this represents the geometry of an amino acid.

3.0.2 Rigidity Analysis. Rigidity analysis [13, 14, 21] is a fast graph-based method that provides information about the flexibility of proteins, which are known to contain regions of varying degrees of rigidity [15]. In rigidity analysis, atoms and their chemical interactions are used to construct a mechanical model of a molecule. A

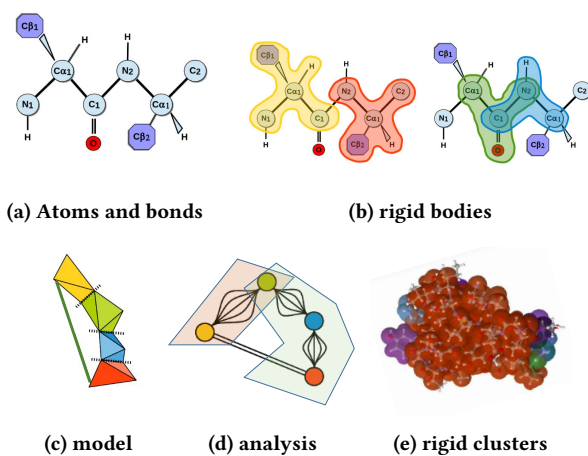


Figure 3: Rigidity Analysis: atoms from a PDB file and bonds among them (a) are analyzed to determine rigid bodies (b), the smallest rigidity components in a molecule. A mechanical model (c) representing the degrees of freedom among the rigid bodies is converted to an associated graph (d) on which rigidity analysis is performed. Its output is used to identify rigid clusters of atoms in the protein (e).

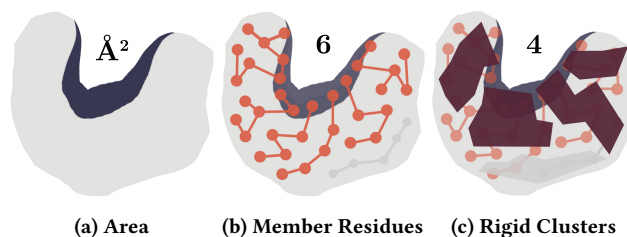


Figure 4: Structure Based Metrics. The properties of protein cavities that we are exploring are the surface area (a), the number of residues participating in the cavity (b, which has six residues) and the number of rigid clusters participating in the cavity (c, which has four rigid clusters).

graph is constructed from the model in which each body is associated to a node, a hinge between two bodies to five edges, and a bar is associated to an edge. Pebble game algorithms [11] are used to analyze the rigidity of the graph. The results are used to infer the rigid and flexible regions of the protein. See Figure 3 for an overview of rigidity analysis.

We use the freely available KINARI rigidity software [8] to calculate the rigidity properties of each protein in our data set. The rigidity analysis output is an XML file containing the identified rigid clusters of atoms. The following is a sample output of the rigidity analysis of a protein; it identifies atoms 2 and 3 (the IDs from the PDB file) as being members of the rigid body with ID 0.

```
<body id="0">
  <pointSet>
    <point id="2"/>
    <point id="3"/>
  </pointSet>
</body>
```

3.0.3 Aggregation of Data. Rigidity data files contain information about atom IDs and clusters; PDB files contain information about residue IDs and atom IDs; and cavity data files contain information about residue IDs and rigid clusters. None of these files individually contain comprehensive information about rigidity properties and residue properties of the protein. In order to associate the rigid clusters with the cavities, the residue identities from the cavity surfaces are cross referenced with a PDB file to identify the alpha carbon associated with that residue. The atom IDs of a rigid cluster are scanned, and if the alpha carbon exists in that rigid cluster it is associated with the cavity.

The output of our aggregation step is a tabular summative record of all cavities at least 1 \AA^2 in size. A sample output below, lists from left-to-right the protein ID, chain ID, cavity ID, cavity surface area, count of residues, count of rigid clusters, and total atoms in those rigid clusters.

```
1e9w A 0 23.6024493 16 6 63
1e9w A 1 17.0409083 17 3 45
```

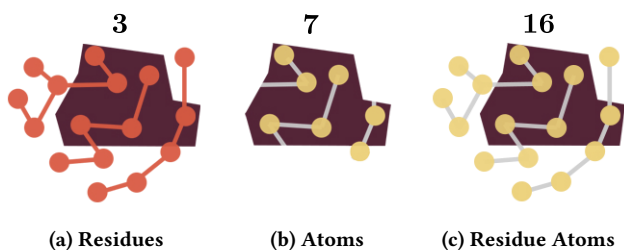


Figure 5: Rigidity Based Metrics. The rigid cluster metrics we tally include the count of residues in a rigid cluster (a, count of 3), the count of atoms in a rigid cluster (b, count of 7), and the count of atoms among the residues that have membership in a rigid cluster, irrespective if the atoms are themselves part of the rigid cluster or not (c, count of 16).

4 CAVITY / RIGIDITY METRICS

The aggregate data, along with the rigidity, cavity, and PDB data, permit us to define a number of metrics for exploring the rigidity properties of the proteins in our data set. The cavity-rigidity metrics that we explore are the following:

- $C_{resCount}$: number of residues in a cavity
- $C_{rigidClustCount}$: number of rigid clusters in a cavity
- $C_{clusterAtomCount}$: count of atoms in the rigid clusters that have membership in a cavity

We are particularly interested in exploring how the surface area of a cavity correlates with other metrics, including the cavity-rigidity metrics, because the size of a cavity is known to determine structural and functional properties of a protein [10, 16]. Because rigidity analysis is quick, identifying relationships between cavities and their rigidity properties has the potential to complement existing studies at a fraction of the computation time, and without the need for laborious wet lab work. Ultimately we aim to quantify the cavity-rigidity relationships to improve cavity analysis software and research. We identify two broad types of metrics: those that are structure based, and those that are rigidity based. We describe both in the following two subsections.

4.0.1 Structure Based Metrics. Several properties of cavities we are exploring are from a biological context, and are taken directly from existing PDB data. These include surface area of protein cavities having size greater than 1 \AA^2 (Figure 4a), and the number of residues that participate in a cavity (Figure 4b).

4.0.2 Rigidity Based Metrics. We identify three rigidity metrics of cavities. We tally the count of residues in a rigid cluster (Figure 5a), the count of atoms in a rigid cluster (Figure 5b), and the count of atoms among all of the residues participating in a rigid cluster, irrespective if the atoms themselves are members of the rigid cluster or not (Figure 5c). The implications of these particular properties of cavities are still largely unexplored, and these metrics are the foundation that will permit us in future work to quantify how properties of atoms in rigid clusters relate to cavity size.

5 RESULTS

Following through with our motivation to explore the rigidity properties of protein cavities, we enumerated the relationship among several of the metrics in Section 4. For each pair of metric correlations we explored, we generated a plot. We discuss these here.

5.0.1 Residue count versus cavity surface area. As a validation step, we plotted the surface area of cavities versus the number of residues in those cavities (Figure 6). Although the 20 naturally occurring amino acids are composed of different numbers of atoms and are therefore different sizes, on average a cavity composed of more residues is expected to be larger than one composed of fewer residues. A linear relationship (correlation coefficient: 0.975) for cavity size versus the number of residues in a cavity follows this logic. Subsequent graphs containing these metrics have very similar overall shapes attributable to this linear relationship.

5.0.2 Rigid cluster count versus cavity surface area and residue count. The correlation coefficient between the rigid cluster count of a cavity and the size of the cavity (Figure 7) was 0.761. **The plot suggests that a small number of rigid clusters will not dominate a large cavity.** There are only a few isolated cases where a large cavity surface area contains a low count of rigid clusters, and these cases are isolated from the behavior of the majority of the cavities plotted.

We observed that **for a particular cavity surface area, the counts of rigid clusters that have membership in that cavity fall predominantly within a lower and upper limit.** For example, in the upper plot in Figure 7, cavity surface areas of approximately $1,000 \text{ \AA}^2$ are composed of between 30 and 150 rigid clusters, but most of the rigid clusters are between 30 and 100 atoms. Similarly, for cavities greater than $1,000 \text{ \AA}^2$, there are no cases where 1 or 2 rigid clusters only are members of those cavities. There are a handful of outlier cases where a large cavity is made up of a small count of rigid clusters (the line of outliers along the left side of the top and bottom of Figure 7). A few examples of these proteins are

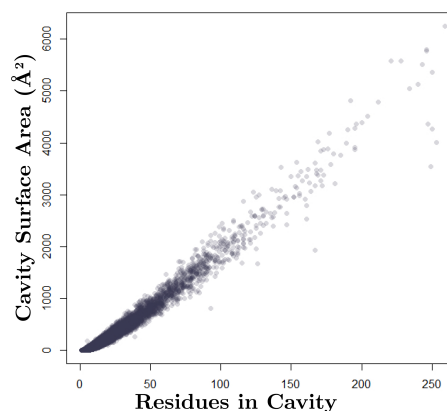


Figure 6: Validation of our Calculations, Residue Count vs Cavity Surface Area. The surface area of cavities has a linear relationship with the number of residues participating in the cavity, with a correlation coefficient of 0.975.

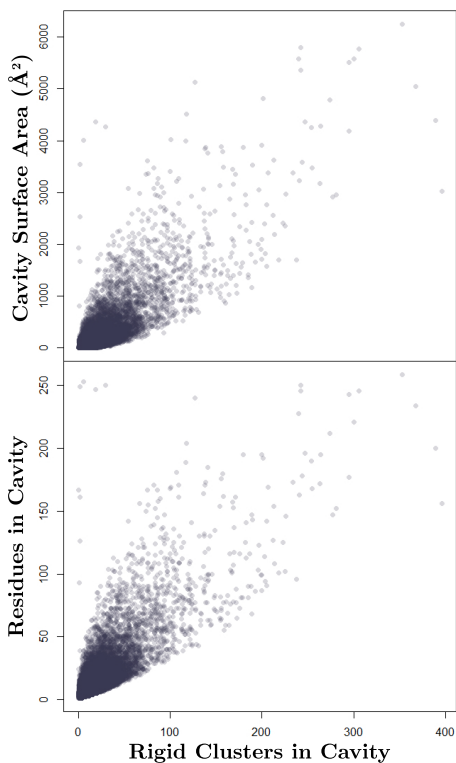


Figure 7: Rigid Cluster Count vs Cavity Surface Area and Residue Count. With the exception of a select number of proteins containing unusual characteristics (the points along the left hand side of the graphs, specifying large cavities dominated by single rigid clusters), the ranges of the count of rigid clusters in a cavity falls within a distinct range, which grows wider as cavities increase in size. The correlation coefficients for cavity surface area and residue count versus rigid cluster count are 0.761 and 0.769 respectively.

PDB structures 1rea, 1tgl, 1tia, and 1tic. Our preliminary investigation into these particular proteins reveals that they all have a large, shallow concave surface, and that there is a single rigid cluster that consumes the entire cavity.

5.0.3 Total Atoms in Rigid Clusters versus cavity surface area, residue count, and rigid cluster count. For smaller cavities (less than $1,000\text{\AA}^2$, or approximately 50 residues), there does not exist a discernible correlation between the total atoms in the rigid clusters that are members of a cavity and the number of residues in the cavity. This can be seen in the dense placing of points at the bottom of the graphs for total atoms in rigid clusters versus cavity surface area/residues/rigid clusters in the cavity (Figure 8), and in the low correlation coefficients of 0.128 for residue count and 0.109 for cavity surface area. However, **in cavities with sizes greater than $1,000\text{\AA}^2$, it is more likely that those cavities contain smaller rigid clusters.** This can be seen in each of the subplots in Figure 8, where there are far more points to the left of 4,000 for each Cavity size than there are to the right of 4,000. The relationship between

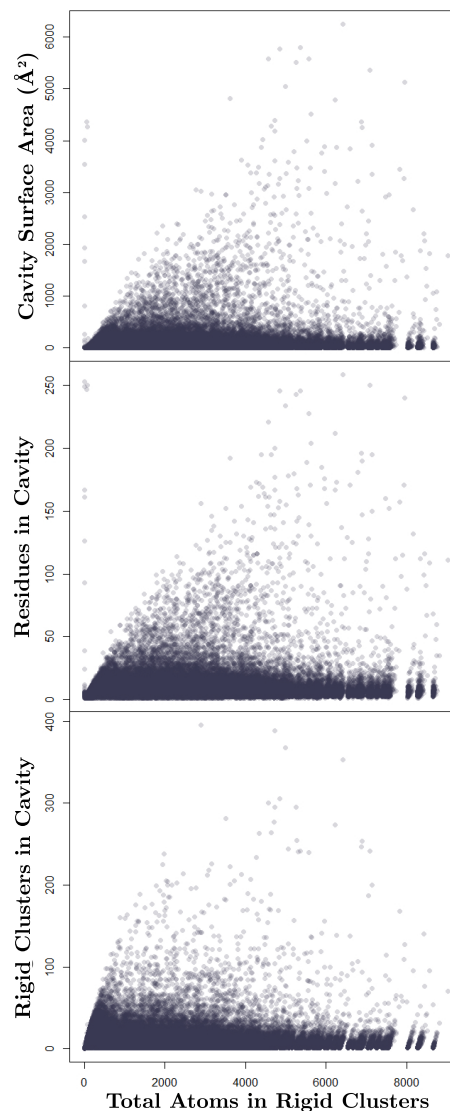


Figure 8: Total Atoms in Rigid Clusters vs Rigid Clusters in Cavity, Residues in Cavity, and Cavity Surface Area. While there is no correlation between these metrics (having correlation coefficients of 0.109 for cavity surface area, 0.128 for residues in cavity, and -0.103 for rigid clusters versus total atoms in rigid clusters), there is a heavier distribution of smaller rigid clusters (fewer total atoms) for cavities larger than 1000\AA^2 or which have more than 50 residues. Under this threshold there is an even spread of rigid cluster sizes.

rigid clusters and their atom counts likely affects some of the outcomes of how rigid clusters are correlating with the size of the cavity/residues in the cavity.

5.0.4 Rigid cluster count versus cavity surface area versus total rigid atoms. A three dimensional graph (Figure 9) of our metrics best shows how a relationship exists between the size of cavities, the number of rigid clusters participating in those cavities, and the

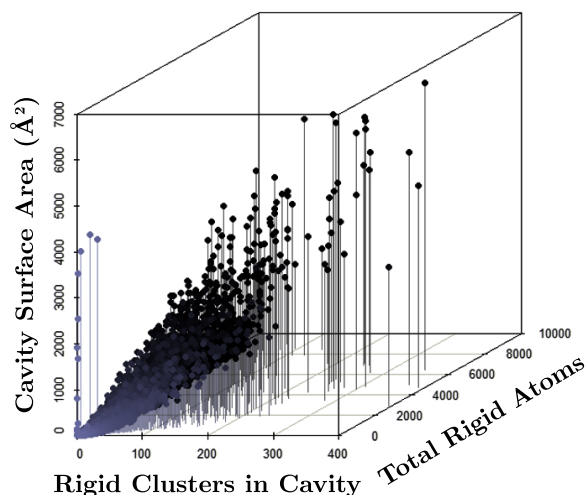


Figure 9: Rigid Cluster Count vs Cavity Surface Area vs Total Atoms in Rigid Clusters. Larger surface areas coincide with higher rigid cluster counts/more overall rigid atoms.

number of atoms contained in those rigid clusters. There are more rigid clusters associated with bigger cavities, and these rigid clusters grow larger as the cavity surface area increases. This suggests that larger rigid clusters are correlated with larger cavities. There are strong limits on the range of how many rigid clusters exist in a cavity of a given size, as well as a limit on the range of how large those rigid clusters can be. There are some exceptions to these ranges where there is low rigid cluster count to cavity surface area ratio.

6 CONCLUSION

A large scale assessment of the rigidity properties of protein cavities is our ultimate goal. We have enumerated a series of cavity rigidity metrics and demonstrated their use in exploring a small subset of available PDB data. Of the correlations among the rigidity and cavity metrics that we explored, we were able to identify several interesting relationships. As expected, the number of rigid clusters has a positive correlation with cavity size (Figure 7). Additionally we observed that for small cavities, no correlation exists between the size of rigid clusters and the size of the cavity; cavities smaller than approximately $1,000\text{\AA}^2$ were composed of both small and large rigid clusters from among the proteins that we studied. However, for larger cavities, we observed that they are composed predominantly of rigid clusters composed of residues with less than 4,000 total atoms (Figure 8). Our initial hypothesis that rigidity properties of cavities vary with cavity surface area is true.

For future work, we will explore the correlations that exist among cavity size and other rigidity and biological properties. For example, how counts of the different types of amino acids that are participating in rigid clusters of a cavity correlate with surface area. Additionally we will investigate how secondary and tertiary structures of protein cavities interact with rigidity metrics. Many of our outliers, on cursory inspection, had similar features such as a large, shallow cavities. Creating more selective data sets that contain

similar structures (for example, proteins with largest cavities over a certain threshold, or classes of proteins such as transmembrane proteins with beta-barrels or alpha helix domains) and looking in close detail at their rigidity analysis/cavity data may lead to useful insights that are not apparent in an overview study. There are also statistical analyses that need to be performed on our existing data to better quantify the relationships. Finally, we plan to analyze the majority of the more than 120,000 protein structures in the PDB.

REFERENCES

- [1] C. Bade-Doeding, T. Huyton, B. Eiz-Vesper, and R. Blasczyk. 2011. The composition of the F pocket in HLA-A* 74 generates C-terminal promiscuity among its bound peptides. *Tissue antigens* 78, 5 (2011), 378–381.
- [2] C. Binda, M. Li, F. Hubálek, N. Restelli, D. Edmondson, and A. Mattevi. 2003. Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proceedings of the National Academy of Sciences* 100, 17 (2003), 9750–9755.
- [3] C. Binda, P. Newton-Vinson, F. Hubálek, D.E. Edmondson, and A. Mattevi. 2002. Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nature structural & molecular biology* 9, 1 (2002), 22–26.
- [4] D. Chan, C.T. Chutkowski, and P.S. Kim. 1998. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proceedings of the National Academy of Sciences* 95, 26 (1998), 15613–15617.
- [5] B.Y. Chen. 2014. Vasp-e: Specificity annotation with a volumetric analysis of electrostatic isopotentials. *PLoS Comput Biol* 10, 8 (2014), e1003792.
- [6] B.Y. Chen and B. Honig. 2010. VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS Comput Biol* 6, 8 (2010), e1000881.
- [7] P. R. Elliott, X.Y. Pei, T.D. Dafforn, and D.A. Lomas. 2000. Topography of a 2.0\AA structure of $\alpha 1$ -antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Science* 9, 7 (2000), 1274–1281.
- [8] N. Fox, F. Jagodzinski, and I. Streinu. 2012. KINARI-lib: a C++ library for pebble game rigidity analysis of mechanical models. *Minisymposium on Publicly Available Geometric/Topological Software* (2012).
- [9] V. Le Guilloux, P. Schmidtke, and P. Tuffery. 2009. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics* 10, 1 (2009), 168.
- [10] N. Huang and M.P. Jacobson. 2010. Binding-site assessment by virtual fragment screening. *PLoS One* 5, 4 (2010), e10109.
- [11] D.J. Jacobs and B. Hendrickson. 1997. An algorithm for two-dimensional rigidity percolation: the pebble game. *J. Comput. Phys.* 137, 2 (1997), 346–365.
- [12] D.J. Jacobs, A.J. Rader, L.A. Kuhn, and M.F. Thorpe. 2002. Protein flexibility predictions using graph theory. *Journal of Molecular Graphics and Modelling* 2 (2002), 195–207.
- [13] D.J. Jacobs, A.J. Rader, M.F. Thorpe, and L.A. Kuhn. 2001. Protein Flexibility Predictions using Graph Theory. *Proteins* 44 (2001), 150–165.
- [14] D.J. Jacobs and M.F. Thorpe. 1995. Generic rigidity percolation: the pebble game. *Physical review letters* 75, 22 (1995), 4051.
- [15] A.G. Ladurner and A.R. Fersht. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *Journal of Molecular Biology* 273, 1 (1997), 330 – 337.
- [16] R.A. Laskowski, N.M. Luscombe, M.B. Swindells, and J.M. Thornton. 1996. Protein clefts in molecular recognition and function. *Protein science: a publication of the Protein Society* 5, 12 (1996), 2438.
- [17] B. Lee and F.M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology* 55, 3 (1971), 379IN3–400IN4.
- [18] J. Liang, C. Woodward, and H. Edelsbrunner. 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein science* 7, 9 (1998), 1884–1897.
- [19] R. Musah, G.M. Jensen, S.W. Bunte, R.J. Rosenfeld, and D.B. Goodin. 2002. Artificial protein cavities as specific ligand-binding templates: characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *Journal of molecular biology* 315, 4 (2002), 845–857.
- [20] M. Nayal and B. Honig. 2006. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins: Structure, Function, and Bioinformatics* 63, 4 (2006), 892–906.
- [21] I.R. Tsang and I.J. Tsang. 1999. Cluster size diversity, percolation, and complex systems. *Physical Review E* 60, 3 (1999), 2684.
- [22] Z. Zhang, Y. Li, B. Lin, M. Schroeder, and B. Huang. 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* 27, 15 (2011), 2083–2088.