The Packing Fraction of Protein-Protein Interfaces

Integrated Workshop
O'Hern and Regan Labs
August 8, 2016

1 Background

1.1 SpyTag and SpyCatcher

Many applications in biotechnology and synthetic biology require bringing different proteins together in a defined fashion. Researchers typically use non-covalent interactions to assemble these multi-protein complexes. For example, metabolic engineers use peptide tags and scaffolds to co-localize enzymes in a pathway, facilitating the inter-enzyme transfer of intermediates and increasing the efficiency of the pathway. Non-covalent interactions can be highly specific allowing researchers to select multiple protein-peptide pairs that do not cross-react. However, when assembling multi-protein complexes using noncovalent protein-peptide pairs, the yield of complex formation depends on the dissociation constant and intracellular concentrations of the molecules involved. In contrast, the SpyTag-SpyCatcher system enables in vivo assembly through an irreversible, covalent bond. SpyTag (ST) and SpyCatcher (SC) are an engineered split protein pair based on CnaB2, the immunoglobulin-like collagen adhesion domain of *Streptococcus pyogenes* [3]. When ST binds to SC, a covalent amide bond rapidly forms between Asp7 of ST and Lys31 of SC . Formation of this bond is efficient over a wide range of conditions, making the ST-SC system a powerful alternative to non-covalent protein-peptide pairs [3].

Currently, only one ST-SC binding pair exists, limiting potential applications. Designing additional ST-SC-like pairs that do not cross-react with the current ST and SC will allow for more sophisticated and controlled applications, without the dissociation constant and concentration dependences that limit assembly by non-covalent interactions. We are developing methods to predict which mutations to the ST-SC system will produce stable binding pairs. We hope to use two properties of a binding pair to predict their stability: 1) the number and severity of atomic clashes predicted using the repulsive Lennard-Jones interatomic potential and 2) the packing efficiency. The computational side of this project will be focusing on the packing fraction calculations of various mutations. In the wet lab, we will be focusing on measuring the reactivity of ST-SC and the cross-reactivity of the ST-SC system with a new system: SnoopCatcher-SnoopTag. SnoopCatcher-SnoopTag is a second engineered binding pair based on an adhesion domain from *Streptococcus pneumoniae*, that forms a covalent bond between a Lys and a Asn residue upon binding [6]. In the wet lab experiments, we will confirm that the Snoop and Spy systems do not interact with each other.

1.2 Packing Fraction

Packing fraction is a property that measures the extent to which an object fills all of the available space in a system. For example, if one carefully packs oranges in a box, the maximum packing fraction you can obtain is ≈ 0.74, where the oranges fill 74% of the

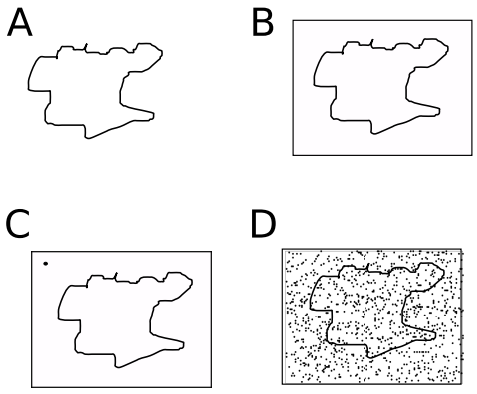

Figure 1: Schematic of Monte Carlo sampling method to find the area of an object by (A) taking an object, (B) drawing a box around it, (C) generating a random point and determining if it is inside the object and (D) repeating with a large number of points until Eq. 1 can be used to calculate the area.

box. Packing fraction is an important metric for understanding the tightness of protein cores or protein-protein binding. By understanding the packing fraction typically found in proteins, one can begin to understand the structural and mechanical properties of proteins. For example, random loose packing would indicate flexibility in protein cores while a high crystalline packing fraction would indicate extensive rigidity. In a recent paper [2], we have shown that the packing fraction of protein cores is ≈ 0.56, significantly below crystal close packing 0.74. In this project, we will be finding the packing fraction of residues at the interface of a protein-peptide binding system. We will then mutate the residues and see how that affects the packing fraction of the system. The extent to which the mutation creates or removes voids in the protein will give us insight into the stability of the mutated structure.

1.3 Monte Carlo Sampling

Monte Carlo methods use random sampling to calculate high dimensional integrals, for example, the volume of high-dimensional geometrical structures and partition functions. They are used for a wide range of applications varying from molecular dynamics simulations to particle physics to astrophysics. We will be using Monte Carlo integration methods to study the volume of a complex three dimensional structure, i.e. a protein. Monte Carlo sampling can be used calculate area or volume in the following way (Fig. 1):

1. Place the object of interest in a box. (Fig. 1B)
2. Generate a random position in the box. (Fig. 1C)
3. Determine if the coordinate is inside the object. Keep track of the number of coordinates in the object.
4. Repeat steps 2 and 3 many times (Fig. 1D)

If enough points are generated then the following is true:

$$\frac{\text{Area of Object}}{\text{Area of Box}} = \frac{\text{\# of Points in Object}}{\text{Total \# of Points}} \tag{1}$$

This type of sampling can be used to calculate the area or volume of any shape where one can determine if a point is inside or outside of the object.

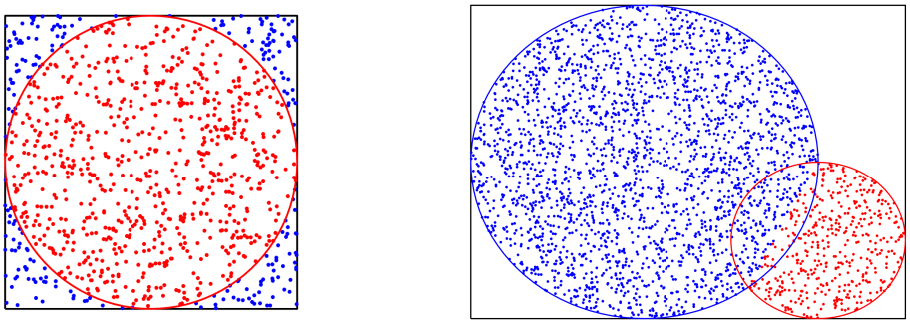

Figure 2: Examples of using Monte Carlo sampling to find the area of an object. A) Finding the area of a circle. B) Finding the area of two overlapping circles

2 Project Details

2.1 Monte Carlo Calculations of a Sphere

The first method you will develop is calculating the volume of a sphere using Monte carlo sampling. Use the function *find_circle_area* as an example of how this works in 2D (Fig. 2, left). Write a function that takes the sphere radii and center coordinates as input and returns the volume of the sphere. Calculate the accuracy as a function of the number of points in the simulation where:

$$Accuracy = \frac{(V_m - V_t)}{V_t} \tag{2}$$

where V_m is the measured volume and V_t is true volume of the sphere calculated using $V = \frac{4}{3}\pi r^3$. Create a plot that shows the accuracy of your model as a function of the number of points in the simulation and determine the number of points needed to obtain an accuracy of 99%.

2.2 Monte Carlo Calculations of Overlapping Spheres

A protein is made up of a set of overlapping spheres. Each point in the protein needs to be correctly assigned to the sphere it is part of, placing dividing planes between the spheres. Write a function to calculate the volume of each of 2 overlapping spheres. Once again, a 2d example is available in *area_overlapping_circles* (Fig. 2, right). The input should be the center and radii of each sphere and the function should return the volume of each sphere where each point in the overlapping region is assigned to only one sphere. We will discuss the best methods to determine which side of the dividing plane between the spheres a given

point is on. To check your accuracy, compare the total volume of the two spheres to the actual volume using the following equation [1] and plot the accuracy using Eq 2 :

$$V = \frac{4}{3}\pi r^3 + \frac{4}{3}\pi R^3 - \frac{\pi(R + r - d)^2(d^2 + 2dr - 3r^2 + 2dR + 6rR - 3R^2)}{12d} \tag{3}$$

Where r and R are the two radii and d is the distance between the centers of the spheres. The 3rd term in this equation is the volume of the overlapping region.

2.3 Volume of Protein Atoms

We now want to calculate the volume of each atom in a protein, allowing us to find the packing fraction of the protein. Edit the function *calc_protein_volume* to find the volume of each atom in a given protein. As in Section 2.2, each point in an overlapping region of two atoms should be assigned to only one atom based on the dividing plane between the two atoms. *calc_protein_volume* takes a protein structure and returns a matrix containing the volume of each atom.

2.4 Packing Fraction of SpyTag/SpyCatcher interface

As discussed above, the ST-SC system is a unique protein-peptide binding pair that forms a covalent bond. We will be mutating the interface and measuring the change in packing fraction upon mutation. The first step is finding the packing fraction of the interface. The interface is defined as any amino acid residue in the protein that is solvent exposed in the individual proteins but buried after binding occurs. STSC_wt.mat contains a subset of the residues of the crystal structure of STSC that are part of the interface. Calculating the packing fraction also involves knowing the actual and voronoi volumes of each atom. The voronoi volumes of STSC_wt are provided in the file STSC_wt_vor.txt.

Write a function that loads the wt structure, calls *calc_protein_volume* and combines the actual and voronoi volumes to get the packing fraction of the system. We will discuss various methods of choosing which atoms to include in the packing fraction calculations to get the most accurate results. Plot the packing fraction of the wild type structure on the distribution of interface packing fractions provided.

2.5 Study Mutations

Now that we know the packing fraction of the wild type system, we will mutate the structure and calculate the packing fraction of each mutated structure. Modify the functions produced above to run on any system. More detailed instructions on generating batch jobs and submitting them to the computer cluster will be provided when this step is reached. After finding the volumes of the mutated structures, calculate the packing fraction of each and compare to the distribution of packing fractions found at protein interfaces. Identify which mutations are expected to be stable based on their packing fraction.

3 Schedule

Week 1

Day	Time		Location
Monday	1-1:30pm	Project Overview	Mason 200
	1:30pm +	Read background and references. Begin Section 2.1	Mason 313
Tuesday	1pm +	Complete Sections 2.1 and 2.2	Mason 313
Wednesday	1-1:30 pm	Meet with Corey via Skype	Bass 326
	1:30-3:30pm	Transform plasmids and plate cells	Bass 317
	3:30pm+	Begin Section 2.3	Bass 326
Thursday	1-1:30pm	Pick cells and replate	Bass 317
	2pm +	Continue Section 2.3	Mason 313
Friday	1-1:30pm	Collect cells and prepare for purification	Bass 317
	2pm +	Finish Section 2.3 and start Section 2.4	Mason

Week 2

Day	Time		Location
Monday	1pm +	Purify Proteins. Work on Section 2.4	Bass 317
Tuesday	1pm+	Purify Proteins. Work on Section 2.5	Bass 317
Wednesday	1-1:30 pm	Meet with Corey via Skype	Bass 326
	1:30pm +	Mix proteins and run gels. Submit jobs to cluster for Section 2.5	Bass 317
Thursday	1-3pm	Calculate packing fraction and analyze results.	Mason 313
	3pm-5pm	Write up experimental and computational results	

References

[1] E.W. Weisstein. "Sphere-Sphere Intersection." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Sphere-SphereIntersection.html.

[2] J.C. Gaines, W.W. Smith, L. Regan and C.S. O'Hern (2016) Random close packing in protein cores. *PRL E* 93:32415.

[3] B.Zakeri, J.O. Fierer, E. Celik, E.C. Chittock, U. Schwarz-Linek, V.T. Moy and M. Howarth (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *PNAS* 109:E690.

[4] L. Li, J.O. Fierer, T.A. Rapoport and M. Howarth (2014) Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* 426:309.

[5] G. Veggiani, B. Zakeri and M. Howarth (2014) Superglue from bacteria: unbreakable bridges for protein nanotechnology. *Trends in Biotechnology* 10:506.

[6] G.Veggiani, T. Nakamura, M. D. Brenner, R.V. Gayet, J. Yan, C. V. Robinson, and M. Howarth (2016) Programmable polyproteams built using twin peptide superglues. *PNAS* 113:1202.