Contents:

1. .cfg files

2. .gpf files

3. Finding MESes from GPFs

1. .cfg files / Configuration Files

1.1 : Why?

A .cfg file lets you control the hydrophobic center distances for each hydrophobic residue, as well as the "allowance constant," or AC. The AC is the fraction of points that will be allowed to fall outside of the sphere during Minimal Enclosing Sphere (MES) generation, represented in decimal. That is to say, if you want to make sure that 90% of points fall inside the MES, you would set the AC to .1, so 10% of points will be allowed to fall outside of the MES. If you want 80% inside, you would set the AC to .2.

A .cfg file also allows you to control the assumed hydrophobic cender distances along the alpha-beta carbon vector. The MES generation script takes in a .gpf file as input, which only has data about the locations of the alpha-carbon backbone, as well as the beta carbons of hydrophobic residues and the specific identity of each hydrophobic residue. The set of points for the MES script is created by assuming the "hydrophobic center" of each hydrophobic residue falls a certain distance along the alpha-beta carbon vector, originating from the alpha carbon, and that this distance is determined by the specific identity of the residue.

1.2 What and how?

A .cfg file is very simple to make. It is a series of lines of text. Each line has two components: A tag and a number. The tag of each line indicates what it's for. The tag of each line is either the single-letter symbol for the hydrophobic residue, if it indicates a hydrophobic center distance, or the phrase "AC" to indicate the allowance constant. An example is below:

A 2.14

F 2.71

AC .2

This indicates that Alanine residues are being set 2.14 along the Alpha-Beta carbon vector, Tyrosine residues are being set 2.71 along the A-B carbon vector, and the allowance constant is set to .2. To see a full example, check the default.cfg file. Note that you do not have to indicate values for every single residue and also the AC. If no value is indicated for any parameter, they will be set to default values which you can see inside the GPFConfigParser.py script.

2. .gpf files / Geometric Protein Files

1. What and why?

A GPF consists of two components. The first is many lines of coordinates (x, y, z) of the alpha-carbon backbone of a single-chain protein. The second is sets of paired coordinates for the alpha and beta carbons of each hydrophobic residue, tagged with the identity of the hydrophobic residue. A GPF captures the important information necessary for the MES generation script. A GPF can be generated from the name of a single-chain protein in the Protein Data Bank (PDB). After being generated, they are placed in the gpf\_files