Introduction

SASSIE is a program suite that enables users to perform tasks that fall into two broad categories: creating comprehensive models of molecular systems, and contrasting the data gained from its simulations to user-supplied experimental data. A primary benefit of utilizing SASSIE's interface is the wide variety of calculators that are available to analyze different types of scattering data (crysol, Xtal2sas, etc). SASSIE is equipped with a variety of molecular simulation methods that can be used to generate large amounts of structures, obtain SAS profiles from said structures, compare these profiles to experimental data, and output files that can be used to visualize results.
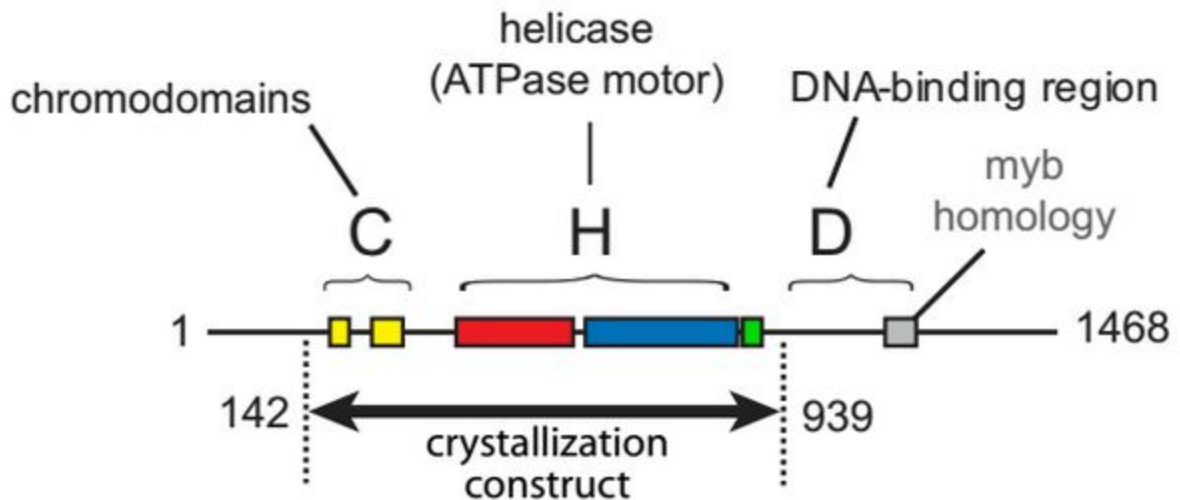
Figure 1: Depiction of Residue Ranges of CHD1's domains (Hauk et al.,2010)

SASSIE can be utilized for a variety of purposes, but in the interest of this inquiry it has been used to examine CHD1, a chromatin remodeler that regulates various parts of transcription. (SGD). Specifically, the mechanism by which the ATPase motor is regulated is not very well understood. By utilizing SASSIE, it was possible for comparisons to take place between two samples of CHD1: the unaltered wild type, and the mutated kak. The principal difference between the mutant kak and unaltered wild type is the mutation of three amino acid residues, being residue numbers 265, 266 and 268 from E D E (Glutamine, Aspartate, and Glutamine) to

K A K (Lysine, Alanine, and Lysine). By completing the process of generating structures from a PDB file, creating SAS profiles, comparing results to experimental SAXS data and then visualizing the results, insight into CHD1's structural behavior can be observed.

Methods

In order to compare the two CHD1 conformations, a variety of tools in SASSIE's interface were utilized. They have been presented in the order used for this inquiry.

Data Interpolation:

Data Interpolation is used to format original experimental data to be readily used with SASSIE, and is usually the first step in structure analysis. This method allows for data to be interpolated once at the beginning of the process instead of re-interpolating it for every SAS curve generated from structures later in the process. This does not reveal any structural information about the PDB file; it only prepares experimental data to be used in later steps. Note that the values entered here *must* match those entered in the calculation step (in this case, Crysol) to obtain results. Generates a DAT file for the *Crysol* step.

- I(0): Experimentally derived I(0), can be predicted via Guinier analysis. Scattering intensity at q = 0.
- new delta q: Determines spacing between data points.
- number of new q-values: Determines number of q-values for interpolated data set.

Monomer Monte Carlo:

This simulation uses a PDB file to generate a number of structures based on residue ranges and the number of trial attempts. Generates a DCD file for the *Crysol* and *Density Plot* steps.

- number of trial attempts: Determines the number of Monte Carlo moves to be attempted; most experiments utilize over 10,000 structures for adequate sampling.
- number of flexible regions to vary: Asks how many flexible regions are included in the submitted PDB file.
- residue range for each flexible region: Asks for the ranges of residues in the PDB file that should be considered "flexible" for the purposes of the simulation.
- structure alignment range: Asks for the ranges of residues in the PDB file that should be considered "rigid" for the purposes of the simulation.
- low / high Rg cutoff: Allows the user to restrict data collection to a predetermined radius of gyration.

Crysol:

Utilizing both the original PDB file and the DCD file generated in the *Monomer Monte Carlo* step, this tool calculates SAS (scattering) profiles. Note that if the following two parameters do not exactly match what is shown in the DAT file generated from the *Data Interpolation* step, the *Chi-Squared Filter* step will fail. Generates a SAS data path for the *Chi-Squared Filter* step.

- maximum q value: The highest q value present in the set of interpolated data.
- number of q points: The number of data points present in the set of interpolated data.

Chi-Squared Filter:

This tool compares the theoretical SAS curves generated in the *Crysol* step to the interpolated file generated in the *Data Interpolation* step. Allows users to determine the best, worst, and average structures generated across the entirety of SAS curves. Generates a TXT weight file for the *Density Plot* and *Extract Utilities* steps.

- I(0): Experimentally derived I(0), can be predicted via Guinier analysis. Scattering intensity at $q = 0$. Must match the value entered at the *Data Interpolation* step.

- enter expression: By entering different chi squared values into the given expression, it is possible to utilize the *Extract Trajectories* tool to visualize the structures that best fit your interpolated data set.

Extract Utilities:

This tool allows users to extract a particular set of frames from the *Chi-Squared Filter* step. This is primarily used to narrow down the number of frames rendered in the *Density Plot* step to include only the best-fit structures as determined by the value entered into the expression from the *Chi-Squared Filter* step. It utilizes the original PDB file, the DCD file from the *Monomer Monte Carlo* step, and the TXT weight file from the *Chi-Squared Filter* step. Generates a DCD file that can be rendered in VMD as an animation.

Density Plot:

This analysis tool creates CUBE files that can be rendered as structures in such programs as VMD and Pymol. It utilizes the original PDB file, the DCD file generated from the *Monte Carlo Monomer* step, and the TXT file from the *Chi-Squared Filter* step. Generates a CUBE file that can be rendered in VMD.

- number of segments: Integer number of different residue regions present in the PDB file.
- segment name: In this inquiry, CHD1 was entered here to match the PDB file.
- residue region(s): Asks for the residue ranges that define each segment. In this inquiry, the regions used were residues 1-35, 36-206, 207-229, 230-784, and 785-801.
- weight file: optional weight file from *Chi-Squared Filter* step. If included in procedure, a set of unequal weights CUBE files are generated. If it is not, a set of equal weights CUBE files are generated, that can both be rendered in VMD simultaneously.

Results

After obtaining data over several iterations, it was determined that more favorable chi-square results could be obtained by narrowing the value of Rg. These results were obtained by restricting the Rg from 35 to 40.
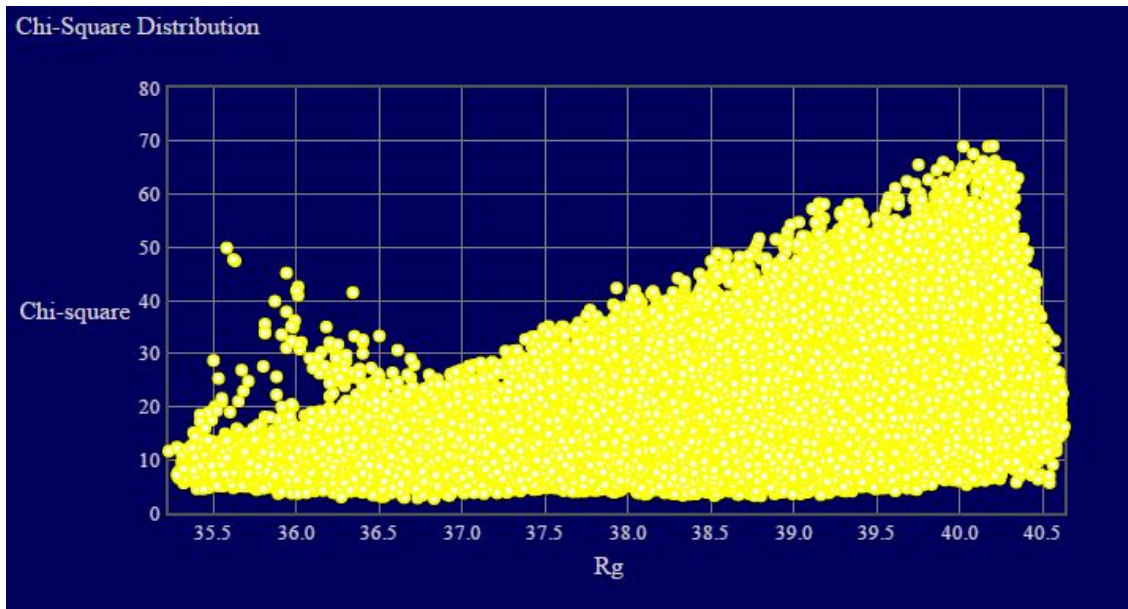

Figure 2: x^2 vs Rg plot for wild type ; lowest $x^2 = 2.66$

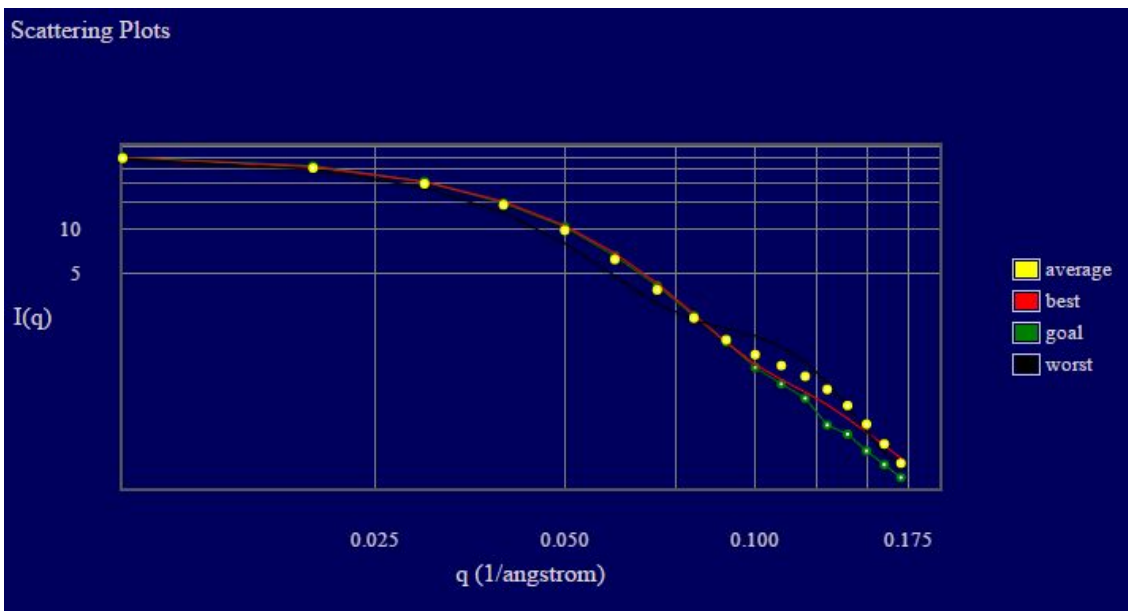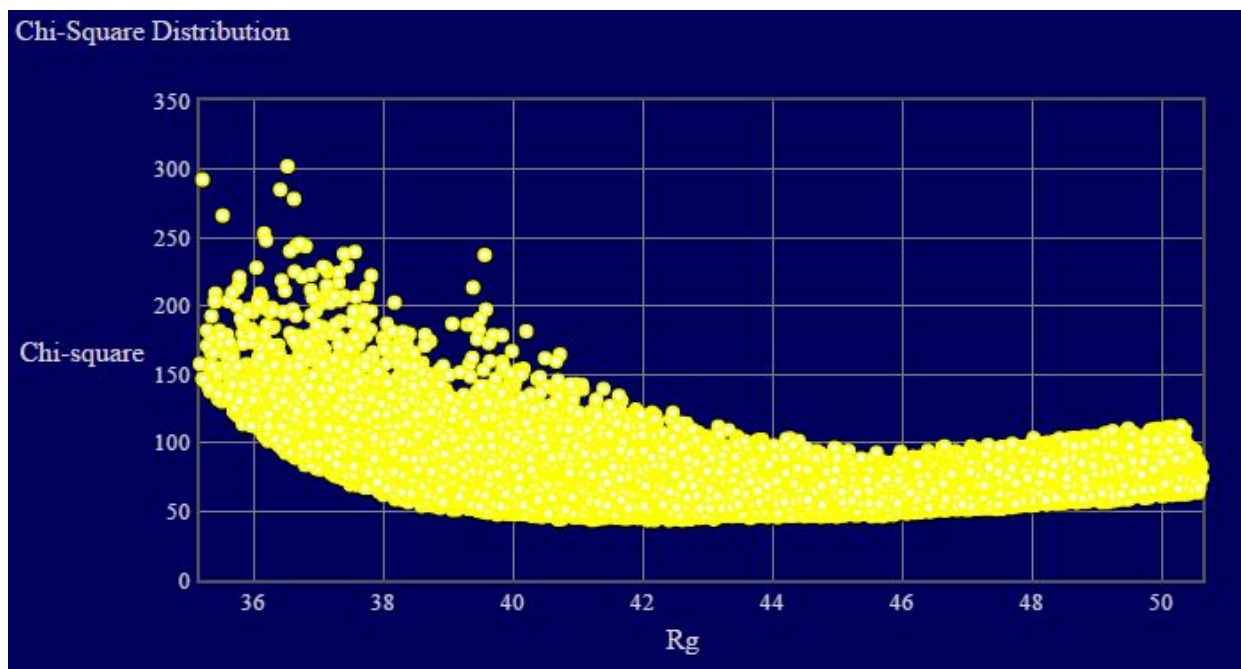**Figure 3: SAS Spectra plot for wild type**

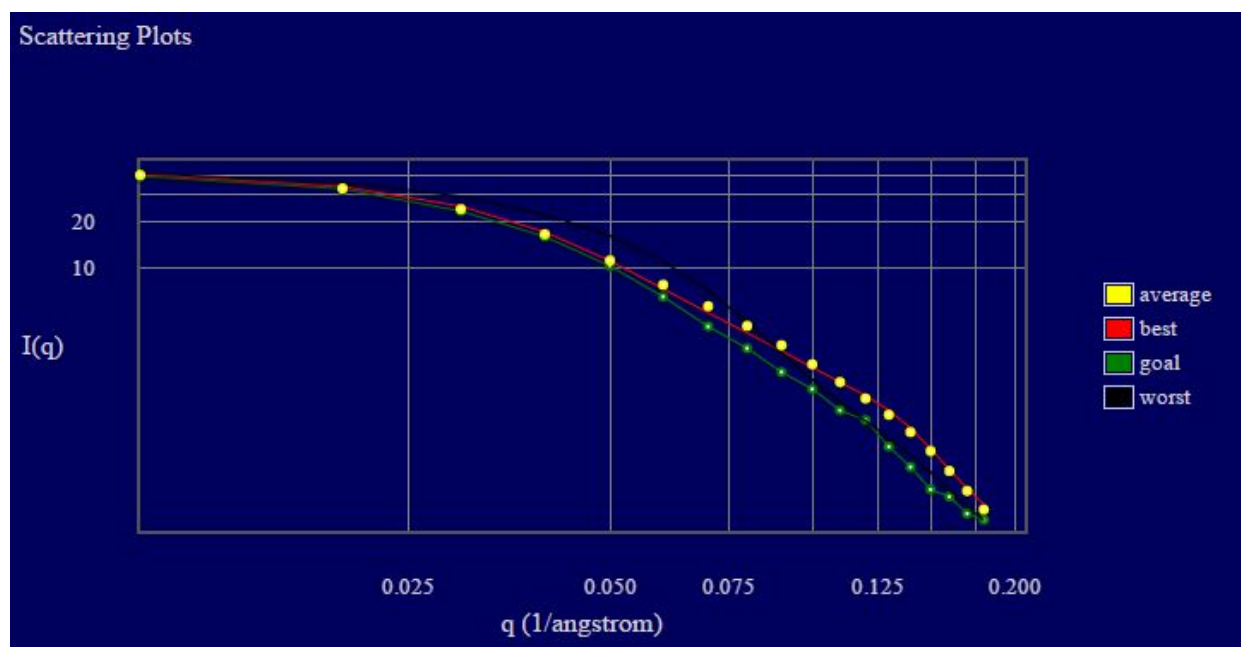**Figure 4: x^2 vs Rg plot for kak ; lowest $x^2 = 43.33$**

Figure 5: SAS Spectra plot for kak

Conclusions

One uncertainty worth investigating is the reason behind the high chi-squared value obtained from the kak in its x^2 vs Rg plot. Homogeneity of the original sample was cited as a potential reason, but further investigation is warranted. As expected, the kak sample demonstrated a more flexible hinge as compared to the wild type due to the repulsive nature about its hinge. Future studies should take advantage of the Merge Utilities function to further appropriate separate runs on the same protein.

While restricting the Rg values on the Monte Carlo helped lower chi-square values, it also caused subsequent failure of a density plot rendering. Thus, it is important to note that when rendering density plots in VMD, it is imperative to use data that is not Rg restricted in order to achieve the best results. Future studies should look into the possibility of merging an Rg restricted run with one that has not been Rg restricted to determine if the chi-square values produced are more favorable, SASSIE shows great promise as an atomistic modeler, and could easily tackle other problems that relate to the structure of complex proteins; this assumes, of course, that the problems are on a large enough scale to be observed with SAXS methods in the first place.

References

1. Hauk, G., et al., The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. Cell Press., 2010. **39**: p. 711-723.

2. "SGD." *CHD1*. Web.http://www.yeastgenome.org/locus/S000000966/overview 06 Aug. 2015.