Introduction: SASSIE Interface and CHD1

SASSIE allows its users to generate time-efficient simulations and models for molecules while also transforming original data into experimental data. When compared to real MD simulations, the SASSIE interface proves to be much faster and therefore a more practical tool for simulations, calculations, and analysis.

The CHD1 chromatin remodeler is capable, under certain mutations, of increasing ones susceptibility to various types of cancers and other diseases. The name of this protein is derived from three distinguishing elements: its two chromodomains, its single helicase-like ATPase motor, and its DNA-binding region (Hauk et al., 2010).



Figure 1. Domain map illustrates all of the regions that were altered in SASSIE.

Although progress has been made in gaining relevant information about this remodeler, there is still much more to be known about how specific regions of the protein function and how they are regulated. It has been observed that this "ATP-driven" protein can put together, slide, and detach nucleosomes from DNA; nevertheless, how these actions are carried out have yet to be understood (Hauk et al., 2010). SASSIE allows for a fast an efficient analysis of the structure functions of the CHD1 protein; however, this interface can also be used for many other bio-molecular structures.

Regions

In order to further understand how the protein reacts and regulates, moving various regions were key to opening up the molecule. We started out with two samples: a wild type and a mutant. The wild type was the original shape of the protein with no alterations. The mutant on the other hand had three vital amino acid residues that were altered (265,266, and 268); when this mutation took place, the protein increased in size.

Next it was decided that it would be beneficial to alter the regions listed below and shown in Figure 1.

- Tail 1 (residues 142 181)
- Region X (residues 261 266)
- Side Hinge (residues 346 362)
- Region Y (residues 844 862)
- Tail 2 (residues 925 942)

SASSIE Modules

During this process there were seven vital modules used in the CHD1 analysis.

Tools

1. Data Interpolation: Used to interpolate the original data. Produces an I(Q) vs. Q graph once complete.

000	Data Interpolation		
project name :	run_1		
input filename :	kakM.dat	output filename :	kakM.dat
I(0) :	9.17	I(0) error :	1.09
new delta q :	0.01	number of new points :	21
version 0.3 : 03/12/09			Wed Jul 30 12:05:07 2014

Figure 2. The data interpolation Module.

- I(0) and I(0) error should be entered correctly or graphs in the chi-square filter runs will not display the best fits.
- New Delta q determines the spacing between the number of points.
- Number of New Points determines how many points
- Coordinate Tools: This module creates new pdb and dcd files from specific data. This was used to create a better fit for one set of runs. To do this the bestworst.txt

file located in the filter run directory; in this file find the structure number that corresponded to the best chi-square (X^2) value. Use the two structure number values after the best value for the creation of the new dcd.

project name :	mutR4	input file path :	2/run_3/MMC_3/generate/
input pdb file :	mutR2_best.pdb	input (pdb or dcd) file :	mutR2_best.dcd
option (single_frame,range,text_file,weight_file):	range	value (frame #, frame range, or filename):	4264-4266
output filename (pdb or dcd):	mutR4_best.dcd		
version 0.1 : 08/29/12			Fri Jul 11 15:59:34 2014

Figure 3. The coordinate tools module.

- Input PDB and DCD files use the inputs from previous run.
- Option input "single_frame" for creating a new pdb and "range" for creating a new dcd.
- Value enter a single value digit for creating a new pdb and a range of numbers (Ex: 1-3) if creating a new dcd.
- 3. Merge Utilities: This module can be used for merging multiple runs together. This was another method used in order to get better fitting data.

project name :	mutant		
reference pdb	mmutant.pdb	merge option (0 == dcd & sas, 1 == sas only) :	0
project name paths :	run_1,run_2,run_3,run_4	dcd filenames :	cd,mutant3.dcd,mutant4.dcc
enter SAS type (1=Xtal2sas: 2=Cryson: 3=Crysol)	3	sas path :	crysol
version 0.1 : 10/24/12 Sun Jul 27 09:38:06 2014			Sun Jul 27 09:38:06 2014

Figure 4. The merge utilities module.

- Reference PDB original pdb used
- Project Name Paths only input the name of the folder not the entire path.
- DCD File Names similar to project path names accept your simply inputting the name of the dcd file.
- SAS Path input the name of the SAS path (in this case "crysol" was used.

Simulate

4. Monomer Monte Carlo: This module generates new structures and also runs simulation using implicit solvent force fields. The latter of these two functions provides user with an idea of how the molecule would move and react with forces present in water. Produces a *Structure vs. Rg* graph once complete.

User Input Section			
project name :	mutant4	output filename (dcd) :	mutant4.dcd
input file path :	/tomlab/Research/mmutant	input filename (pdb) :	mmutant.pdb
number of trial attempts :	20000	return to previously accepted structure :	10
temperature (K) :	300.0	molecule type (protein or rna) :	protein
Molecule Specific Input			
number of flexible regions to vary :	4	enter MAX dtheta for each region :	30.0,30.0,30.0,30.0
first residue per region :	1,119,204,783	number contiguous residues per region :	39,5,16,17
structure alignment: low residue :	284	structure alignment: high residue :	350
Non-Standard / Specialized Input			
overlap basis :	heavy	overlap cutoff :	0.9
low Rg cutoff :	0.0	high Rg cutoff :	400.0
Z coordinate filter (0==no, 1==yes) :	0	Z cutoff (angstroms) :	0.0
atomic constraints (0==no, 1==yes) :	0	constraint filename :	constraints.txt
non-bonding energies (0==no, 1==yes) :	0	non-bonding scaling factor :	1.0
psf file path :	/tomlab/Research/mmutant	psf file name :	mmutant.psf
parameter file path :	/usr/local/bin/sassie/simulat	parameter file name :	par_all27_prot_na.inp
		directed Monte Carlo (0 = = no or Ra value) :	45

Figure 5a. The monomer monte carlo module.

- Number of Trial Attempts How many attempts should be ran. The larger the number of attempts, the longer the run will take.
- Number of Flexible Regions how many regions do you want to vary (must be entered from least to greatest).
- Number of Continues Residues per Region the region boundaries (Ex: Tail 1, residues 1-39).
- Overlap Basis In this section the user inputs a keyword; this keyword allows one to select a certain category of atoms in the molecule.
- Overlap Cutoff As this value becomes lower, more bending angles are rejected; when set at a high value fewer bending angles are rejected. Permits a certain degree of flexibility if the sample data it stuck at an overlapping point in the structure.
- Directed Monte Carlo This input allows the user to increase the number of low chi-squared values.

Calculate

5. Crysol: This module calculates the scattering of the structures generated in he Monomer Monte Carlo run and creates the SANS data files. Important to match the run with the data interpolation run. If not matched properly the chi-filter step would not run smoothly.

User Input Section			
project name :	mutant4		
trajectory input file path (dcd or pdb) :	C/run_4/mutant4/generate	trajectory file name (dcd or pdb) :	mutant4.dcd
pdb reference input file path :	C/run_4/mutant4/generate	pdb reference file name :	mmutant.pdb
crysol executable name :	/usr/local/bin/crysol.exe	delete alm/flm/sav files (1=Y/0=N):	1
Crysol Input Section			
option	0	max harmonic order :	15
Fibonacci grid order :	17	maximum s value :	0.2
number of points :	21	contrast of solvent shell :	0.03
electron density of solvent :	0.334	explicit hydrogens (Y/N) :	Ν
version 0.3 : 3/24/09			Tue Jul 15 10:11:18 2014

Figure 6. The crysol module.

- Number of Points Determines the number of points. Make sure the number of points in the crysol run is the same of that in the data interpolation run
- Maximum S-value It is essential that you adjust the maximum s-value in the crysol run so that it will have the correct spacing. (Ex: If interpolation run is set at 21 points, it only made sense to match the crysol run to 21 points as well. In this case a maximum s-value of 0.20 would provide the correct spacing needed to match the data in order to proceed to the next step successfully).

<u>Analyze</u>

6. Chi-Square Filter: Produces the *Chi-Square vs. Rg* and *SAS Spectra* plots. With these plots we are able to see areas of improvement that could be made in future runs, which would create better fits for data. This module also produces the .txt files used in the Density Plot. For creating an unequal weights file for your data two filter runs should ran.

🖲 🔿 🔿 Chi-Square Filter				
project name :	mutant	path to SAS files	Enter path to SAS files	
SAS exp. data file (goal) :	./nsfit.data	I(0) :	9.17	
X2 high cutoff :	40.0	X2 high cutoff file (output) :	x2highweights.txt	
X2 low cutoff :	15.89	X2 low cutoff file (output) :	x2lowweights.txt	
Rg high cutoff :	60.0	Rg high cutoff file (output) :	rghighweights.txt	
Rg low cutoff :	40.0	Rg low cutoff file (output) :	rglowweights.txt	
enter SAS type (1=Xtal2sas : 2=Cryson : 3=Crysol) :	3	use reduced X2 (0==no, 1==yes) :	1	
version 0.5 : 07/14/08 Tue Jul 29 15:00:42 2014				
Run Chi-Square Filter Program				
0 %				

Figure 7a. The chi-square filter module.

- X2 Cutoffs Cutoff inputs for the chi-squared values. This input is important when you plan on creating an un-equal weights file for the Density Plot Module; in the second filter run ten points should be added to the best chi-square value (x2 low cutoff input), and then the x2 high cutoff input value should be increased so that it is at least 20.0 points higher than the x2 low.
- Rg Cutoffs The cutoff inputs for the Rg values.
- 7. Density Plot: In this step we use the filter data to create cube files. With these cube files, we can view the molecules in VMD. This allows us to see the motion of the various segments.

00	Density Pl	ot	
project name :	mutant	input filename :	erge/generate/mutant.dcd
reference pdb :	rge/generate/mmutant.pdb	output filename :	mutant
xlength :	600	grid spacing :	6.0
ylength :	600	save occupancy data (Y/N)?	Ν
zlength :	600	number of segments :	1
Click to enter segment information			
equalweights (1=yes,0=no) :	1	weights filename :	Enter Weights path + file nar
Run Density Program			
	0 %		



\varTheta 🔿 🔿 Segment Region Input			
Variables for segment number 1			
number of ranges :	4		
enter low regions :	1,119,204,783		
enter high regions :	39,124,220,800		
segment basis:	CA		
segment name:	CHD1		
Continue			

Figure 8b. The segment information.

- X, Y, and Z length increases the box size of the molecule.
- Equal Weights used to create equal and unequal weight

- Weights Filename leave blank for equal weights, enter "lowweights.txt" when you want to create an unequal cube file.
- Segment Region Input
 - a) Number of Ranges How many regions
 - b) Enter Low Regions Low residue numbers for each region
 - c) Enter high Regions High residue numbers for each region
 - d) Segment Basis Names the region or category of atoms.
 - e) Segment Name Name of DNA or protein.
 - f)

SASSIE Flow Chart



Figure 9. **SASSIE flow chart.** *This step is optional. In the CHD1 runs the merge utilities and coordinate tool modules were used to get better fitting data.

Results: Wild Type and Mutant CHD1

After several complete runs with the SASSIE interface, it was noted that the data could obtain better fits by merging the runs using the merge utilities tool, and proceeding with the chi-square filter and density plot modules. When this was complete the following results were obtain (see figure 8 and figure 9 below).



Figure 8a. The X^2 vs. Rg plot for the wild type.



Figure 8b. The SAS Spectra plot for the wild type.



Figure 9a. The X^2 vs. Rg plot for the mutant.



Figure 9a. The SAS Spectra plot for the mutant.

When compared to the plots from each individual run these graphs proved to have a better fit; however, these results did not fit as well as expected. From the unsatisfactory plots it was inferred that something must be wrong with the data; after this was agreed

upon, new data for both the wild type and the mutant were brought back, and tried through the same process.

Upon receiving the new data, it was decided that the coordinate tools module should be used to create a best pdb and dcd for the wild type and kak samples. After the interpolation of the new data, and coordinate tool runs were complete, a monomer monte carlo of 40,000 steps for the next two sets of runs. Once finished the runs used with this new data was merged and filtered and the new plots were created; when zooming in on the Chi-Square filter graphs displayed very unique steep concentrated regions (see figure 10, figure 11 below).





Figure 10a. The X² vs. Rg plot for the wild type



Figure 10b. The X² vs. Rg plot zoomed in for domain 0<y<100 for the wild type



Figure 10c. The SAS Spectra plot for the wild type.



Figure 11a. The X^2 vs. Rg plot for the mutant.



Figure 10b. The X^2 vs. Rg plot zoomed in for domain 0 < y < 100 for the wild type



Figure 11c. The SAS Spectra plot for the mutant.

Using the equal and unequal cube files created in the density plot module, VMD renders of the wilt type and the mutant were created (see figure 12 and figure 13 below). The silver mass represents the equal weight, and the blue mass represents the unequal weight.



Figure 12. VMD Render of wild type showing equal and unequal weights.



Figure 13. VMD render of mutant showing equal and unequal weights.

Conclusion and Suggested Future Direction:

The plots from the final merged data show an excellent fits, thus demonstrating SASSIE's ability to indicate flaws in ones data. With these results the assumption that something was wrong with the old data is clearly validated. In the future it would be beneficial to look more into what the steepness of the plots in figures 11c. and 12c. could indicate. If given provided with samples other than the CHD1 chromatin remodeler one could utilize the SASSIE interface to analyze and investigate the structure functions of various DNA samples that are just as vital as the CHD1; this could possibly lead to the solution and understanding of other health issues similar to those involved with the CHD1 protein.

References

- 1. Curtis, J.E., et al., *SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints.* Comp. Phys. Comm., 2012. **183**(2): p. 382-389.
- 4. 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. Kline, S.R., *Reduction and analysis of SANS and USANS data using IGOR Pro.* J. Appl. Crystallog., 2006. **39**: p. 895-900.
- Sharma, A, et al., *Crystal Structure of the Chromodomain Helicase DNA-binding Protein 1 (Chd1) DNA-binding Domain in Complex with DNA.* The Journal of Biological Chemistry., 2011. 49: p. 42099-42103.