SASSIE-Beta Testing using Small Angle X-ray Scattering Patterns from Mono-Ubiquitin

Oyindayo Hassan

Xavier University of Louisiana, *New Orleans, LA*

Abstract

SASSIE, a component of CCP-SAS (collaborative computational project for small angle scattering) is a computational tool used to generate atomistic models of molecular systems, calculate the SANS, SAXS, and neutron reflectivity profiles from these atomistic structures, and compare the resulting scattering data directly to experimental data. [1] I spent this summer testing SASSIE's user friendliness to non-experts, using Mono-Ubiquitin as my model mechanism. This paper will highlight the complications I encountered during my usage of SASSIE- Beta.

1. Introduction

The goal of CCP-SAS is to create new and enable existing computational tools to model scattering data in real space to dramatically improve accessibility by non-experts.[2] To test the accessibility of CCP-SAS's SASSIE- Beta, I used experimental SAXS (small angle x-ray scattering) data from Mono-Ubiquitin.

Ubiquitin is a protein that aids in various cellular functions, such as, protosomal degradation, DNA damage repair, cell division, and apoptosis. It is also detected that approximately 5% of the genome regulates Ubiquitin signals.[3]

As an open browser, there is no specific way in which one must uses SASSIE. Figure 1 gives one example in which SASSIE can be used.

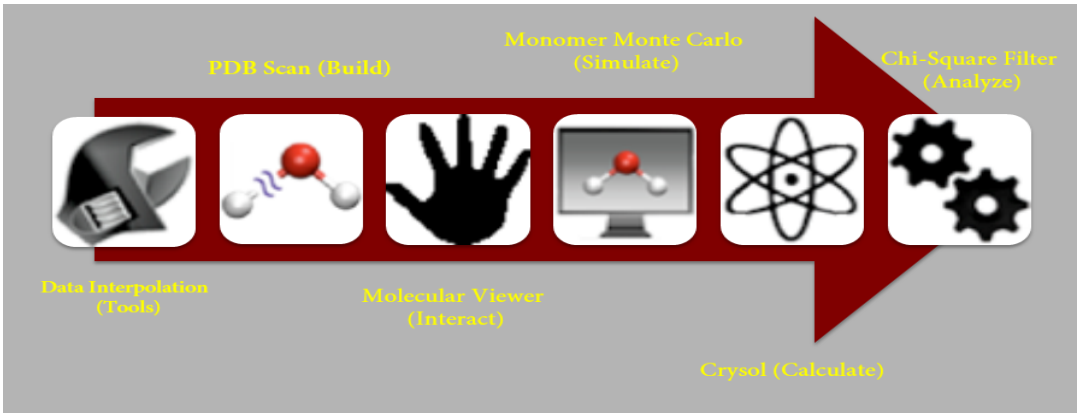


Figure 1: SASSIE user flowchart

2. Materials and Methods

2.1. Molecular Dynamics Simulation

Rather that utilizing SASSIE's ability to generate atomistic models of molecular systems, I used AMBER (assisted model building and energy refinement) and VMD (visual molecular dynamics) to simulate the molecular dynamics and produce my mono-ubiquitin trajectory. In AMBER I used xLEap to obtain my parameter, topology and coordinate files that describe my proteins molecular interactions. Next I used pmemd to create input files that define the program setting for the molecular dynamics of mono-ubiquitin. These input files included energy minimization, slow heating of the system at a constant volume and temperature, and the molecular dynamics production of the system at a constant pressure and temperature. Using the output files from my last molecular dynamics production run, I then visualized my resulting mono-ubiquitin trajectory in VMD and created the pdb and dcd files that I would later use in SASSIE.

3. Complications

3.1. Q-value spacing

The first complication I faced during my use of SASSIE occurred as a result of my "new delta q" and "number of q- values" input during data-interpolation, and caused the failure of my chi-squared analysis as seen in figure 2.

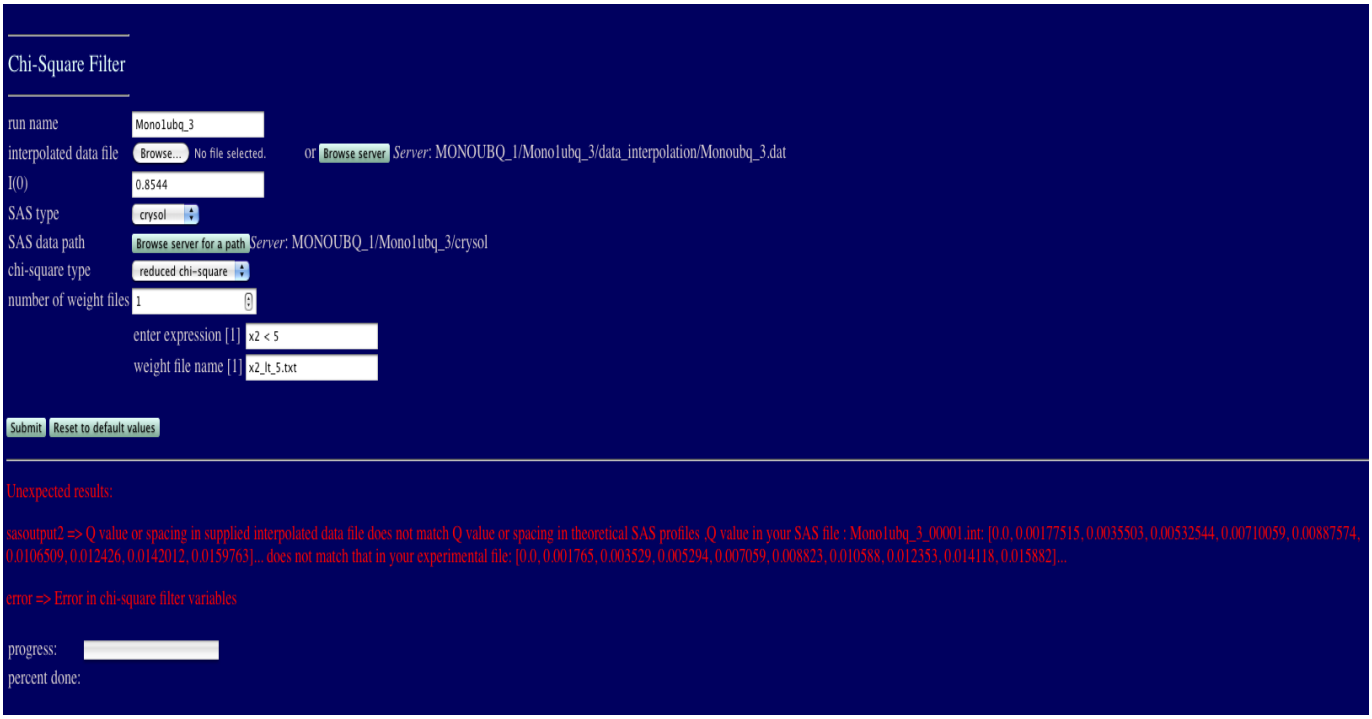


Figure 2: q-value spacing error

The unexpected failure of my chi-squared analysis resulted because of the q- value spacing in the theoretical SAS data created by crysol, did not match the spacing in the interpolated experimental data. Since this difference between the experimental and theoretical SAS data is one of 0.00001, it is evident that SASSIE is extremely sensitive to this q-value spacing.

3.2. Large Chi-Squared values

The second complication I encountered was the problem of large chi-squared values calculated by SASSIE. Using the Guinier Equation,

$$\ln(I(q)) = \ln\left(I_0\right) - \frac{q^2 R_g^{\ 2}}{3}$$

when qRg <<1.5

I initially calculated an I(0) of 0.8544. However, I later discovered that this value was not a proper fit to my data because the q-values selected when calculating my equation were not linear. Using this value of 0.8544 with one dcd frame of my protein, I obtained a chi-square value of 1261! After it was called to my attention that I should try using an I(0) of 0.7, I attained a chi-squared value of 60.93. Although this was an improvement from my previous chi-squared value, it is not one that would be readily acceptable by the scientific community.

To insure that the error was not originating from my experimental data or simulated molecule, I calculated the chi-squared value of frame 48 using Crysol on the command line. With the same parameters, and experimental data I used in SASSIE, I obtained a chi-squared value of 1.653. This result begged the question of what is causing SASSIE to give such large chi-squared values.

Using the following chi-squared equation, the interpolated experimental SAX data, and theoretical SAS data outputted from SASSIE's crysol, I manually calculated a chi-squared value of 13.58.

$$x^2 = \sum_{i=1}^{N_p} \left[\frac{I_{exp} - cI_{theo}}{\sigma}\right]^2 = 13.582$$

$$c = \left[\sum_{i=1}^{N_p} \frac{I_{exp} - I_{theo}}{\sigma^2}\right]\left[\sum_{i=1}^{N_p} \frac{I_{theo}^{\ 2}}{\sigma^2}\right]^{-1} = 7.003^{-7}$$

The variable "c" is the scaling factor that is required in the equation because the theoretical intensity for each q value differs from its correlating experimental intensities by a magnitude of five.

Why is it that using the same data interpolated and calculated by SASSIE to manually calculate chi-square gave me a value far better than that calculated by SASSIE? I believe that this had to do with different scaled theoretical intensities. I decided to calculate my I(0) using the scaling factor, and obtained a value of 0.676.

$$c * I_{theo} = I(0) = 0.676$$

Using this I(0) value, I recalculated my chi-squared for frame 48 and achieved a value of 13.69. Seeing how much changing my I(0) value caused my chi-squared value in SASSIE to fluctuate, I recalculated the I(0) value with the Guinier equation, selecting more linier q-values, and obtained a new I(0) of 0.695. Using this new I(0) value and changing my "contrast of solvation shell", in SASSIE's crysol tool, from 0.03 to 0.00, I was finally able to obtain an acceptable chi-squared value of 1.754, using 50 frames of mono-ubiquitin.

3.3. SASSIE vs Crysol command line

While my best structure was frame 14 with a chi-squared value of 1.754, my worst structure was frame 48 with a chi-squared value of 77.54.

This shows that an I(0) of 0.695 was compatible with some structures such as frame 14, but not as compatible with others, such as my frame 48. Using this chi-squared data along with the data I obtained with different I(0) values in SASSIE and the data collected from Crysol on the command line, I created a comparison table. The purpose of this comparison, as seen in figure 3, was to determine if the best and worst chi-squared values correlated with the same frames from mono-ubiquitin, between the different computational tools, and varying I(0) values.

It is clear that the structures with the best and worst structures did not correlate amongst SASSIE and Crysol, but rather, contradicted in some cases, as seen when comparing frame 14 in both tools to frame 8 (when one value increases, the other decreases). I then decided to visualize and compare these frames using VMD.

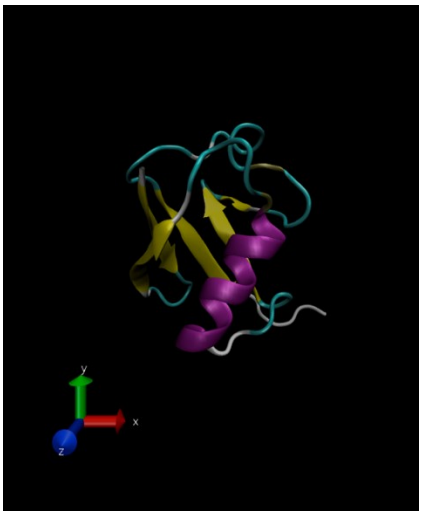
Frame#	SASSIE-B X^2			Crysol X^2
	I(0):0.6952	*I(0):0.7*	*I(0):0.676*	
1	3.76	8.47	21.24	0.72
2	6.80	13.25	17.52	1.08
3	31.77	26.80	86.58	2.77
4	11.71	20.71	12.58	0.89
5	22.52	34.59	11.53	1.31
6	8.09	11.62	30.12	1.26
7	21.13	32.70	12.08	0.76
8	8.57	15.97	15.63	0.68
9	16.14	26.57	11.49	0.77
10	15.36	25.38	12.28	0.72
11	14.57	24.88	10.38	0.93
12	8.42	16.23	13.89	0.79
13	10.58	19.46	11.94	0.89
...	...	...	...	...
14	1.75	5.12	24.38	0.77
40	3.08	3.31	37.80	0.83
41	15.62	25.33	13.75	1.96
42	26.14	38.75	13.04	3.31
43	6.64	13.00	17.69	0.86
44	5.52	12.29	15.01	0.88
45	9.53	17.59	14.02	0.71
46	26.38	39.71	10.53	1.12
47	5.18	10.55	20.12	0.85
48	77.55	99.38	28.81	1.65
49	43.90	60.94	13.70	2.54
50	21.16	33.23	10.18	1.84

Best X^2

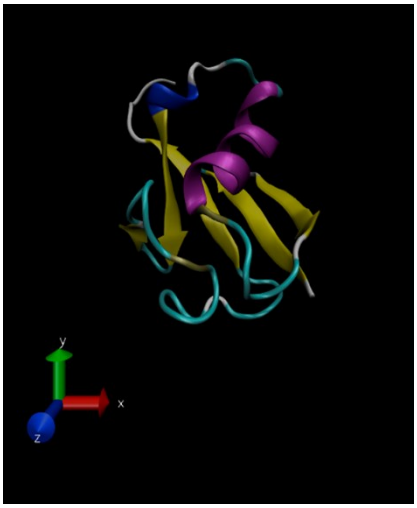
Worst X^2

Figure 3: chi-squared comparison table

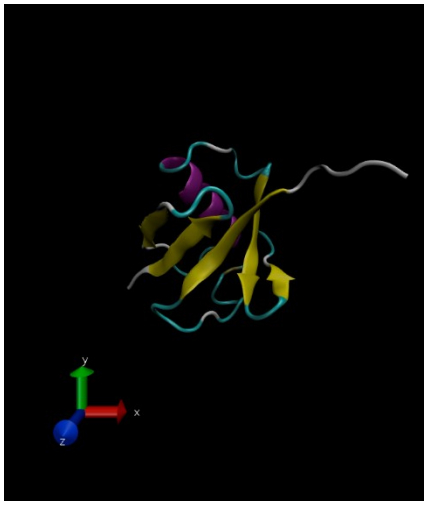
Figure 4: VMD image of Frames



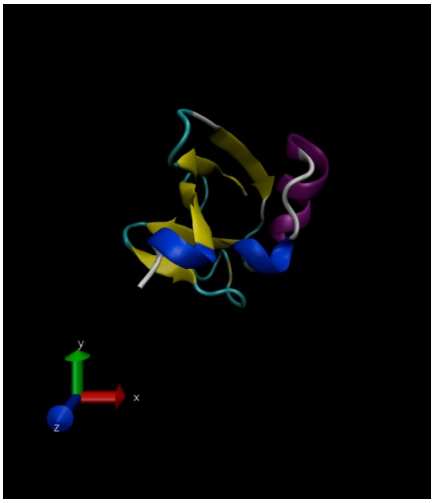
Frame 8 (best x^2 Crysol)



Frame 14 (best x^2 SASSIE)



Frame 42 (worst x^2 Crysol)



Frame 48 (worst x^2 SASSIE)

Although it is evident that these structures look very different, because I did not obtain a RMSD (root mean square deviation) of the different frames, this data in not significant.

5. Conclusion

SASSIE brings together a number of computational tools to one web browser, making it far easier for the non- expert to obtain data without having to fully understand how to manipulate several different tools. However, the variations in my chi-squared data show that the non-expert must be somewhat of an expert to obtain usable data. Finally, the comparison between SASSIE and Crysol command line indicate that results vary dependent on the tool used.

As a result of the problems I encountered during my use of SASSIE-Beta, my suggestions for future improvements to this revolutionary computational tool include, an adjustment to its extreme sensibility to the number of q-values and q-value spacing. The addition of auto-scaling I(0) values when analyzing the chi-squared distribution. Along with the inclusion of the "run name" in the job manager to assist in the location of saved files.

6. Acknowledgments

First and foremost, I would like to thank God. My advisor, Dr. Anderson Sunda-Meya from Xavier University of Louisiana's Physics Department. Dr. Tom Irving, and Dr. Jeff Wereszczynski, of the Illionis Institute of Engineering, and Joseph Curtis for this great opportunity. And everyone in the lab for all of their support and answered questions.
Thank you!

7. References

[1] "CCP-SAS Project." CCP-SAS. N.p., n.d. Web. 05 Aug. 2015. <http://ccpsas.org/about.html>.

[2] Curtis, Joseph E., and Krueger, Susan. "SASSIE." SASSIE. NIST/NCNR, n.d. Web. 05 Aug. 2015. <http://www.smallangles.net/sassie/SASSIE/SASSIE_HOME.html>.

[3] Wereszczynski, Jeff "Ubiquitin Structures". PowerPoint. Chicago, Illinois. 05 Aug. 2015.