Return to Main Documents Page

SASSIE-web: Quick Start

Introduction

SASSIE-web is an online simulation and analysis tool for the modelling of biomolecular structures using small angle scattering data. It is based on the original, standalone, program SASSIE (Curtis *et al.* 2012) and retains all of its core features. This guide is designed to get you familiar with the basic features of the program as quickly as possible. The features covered will be:

- Monomer Monte Carlo simulation to create an ensemble of varied trial protein structures
- Calculation of theoretical scattering from protein models
- Comparison of theoretical and experimental scattering data

Important Note: Before you start this tutorial you will need to register for an account for and login to SASSIE-web. Instructions on how to register can be found here.

The only data needed to work through this tutorial are:

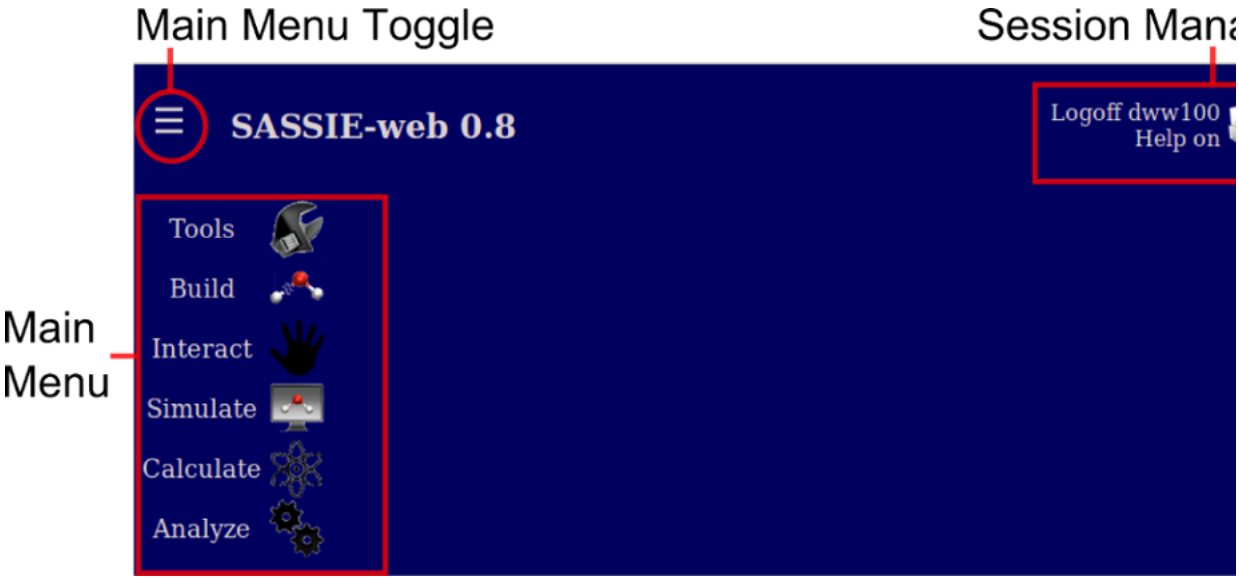
- A starting structure in PDB format: gag_start.pdb
- An experimental scattering curve: gag_scattering.dat

You should download these files to your computer now. A good idea is to create a directory called *SASSIE-web-tutorial* and save all downloads from the tutorial in this location.

You will also need to familiarize yourself with a molecular viewing program that can display PDB and DCD files. We recommend VMD and provide a quick tutorial here.

SASSIE-web Interface

Once logged in to SASSIE-web the page should look something like the figure below (minus the red boxes highlighting the various sections of the interface).



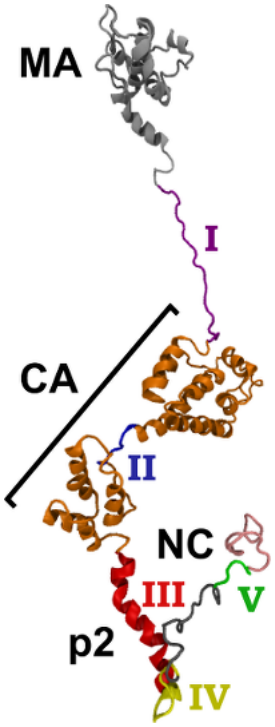
- **Session Management**: Section containing the following icons:
  - Filing Cabinet: File browser
  - Cogs: Job management
  - Head: User configuration
- **Main Menu Toggle**: Hides or shows the main menu
- **Main Menu**: Provides links to the different categories of modules available in SASSIE

During this tutorial, when instructed to select something from the **Main Menu** but no menu is visible on the left hand side of the page you must click on the **Main Menu toggle** to reveal it.

Starting Structure

In this tutorial we will model the conformation of the HIV-1 Gag protein following the study of Datta *et al.* 2007. HIV Gag is a long polyprotein which is cleaved to form the functional proteins required by the virus. The viral proteins which form the domains in Gag are the matrix (MA), capsid (CA), p2 and

nucleocapsid (NC). A structure stitched together from evidence based on crystal structures and models of the individual domains is shown below. This structure will be used as the starting structure for our simulations.



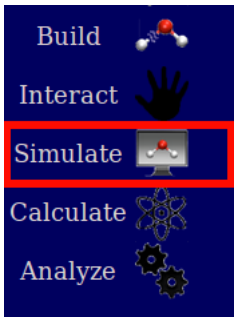
Domain	Flexible Region	Residues
MA		1 - 122
linker	I	123 - 144
CA		145 - 276
	II	277 - 282
		283 - 353
p2	III	354 - 377
linker	IV	378 - 389
NC		390 - 407
	V	408 - 412
		413 - 432

Datta *et al.* 2007 identified 5 flexible regions (labelled I-V), which are highlighted in the figure above. The table alongside the picture shows the residues which make up each region (we are going to need this information to select regions to be varied when we run Monte Carlo simulations).

Structure Variation - Monte Carlo (Monomer)

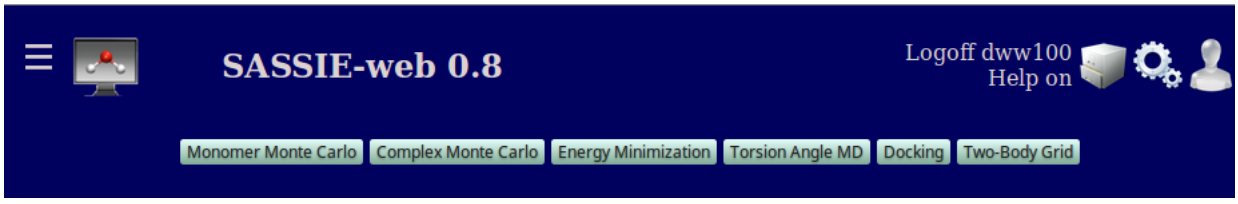
The primary way to vary structures in SASSIE is via Monte Carlo simulations which rotate the backbone dihedral angles of flexible regions within proteins to sample a wide range of structures. Here we setup and run such a simulation, before visualizing the range of stuctures produced.

Select the 'Simulate' button from the main menu of SASSIE-web (left hand side of the main page).

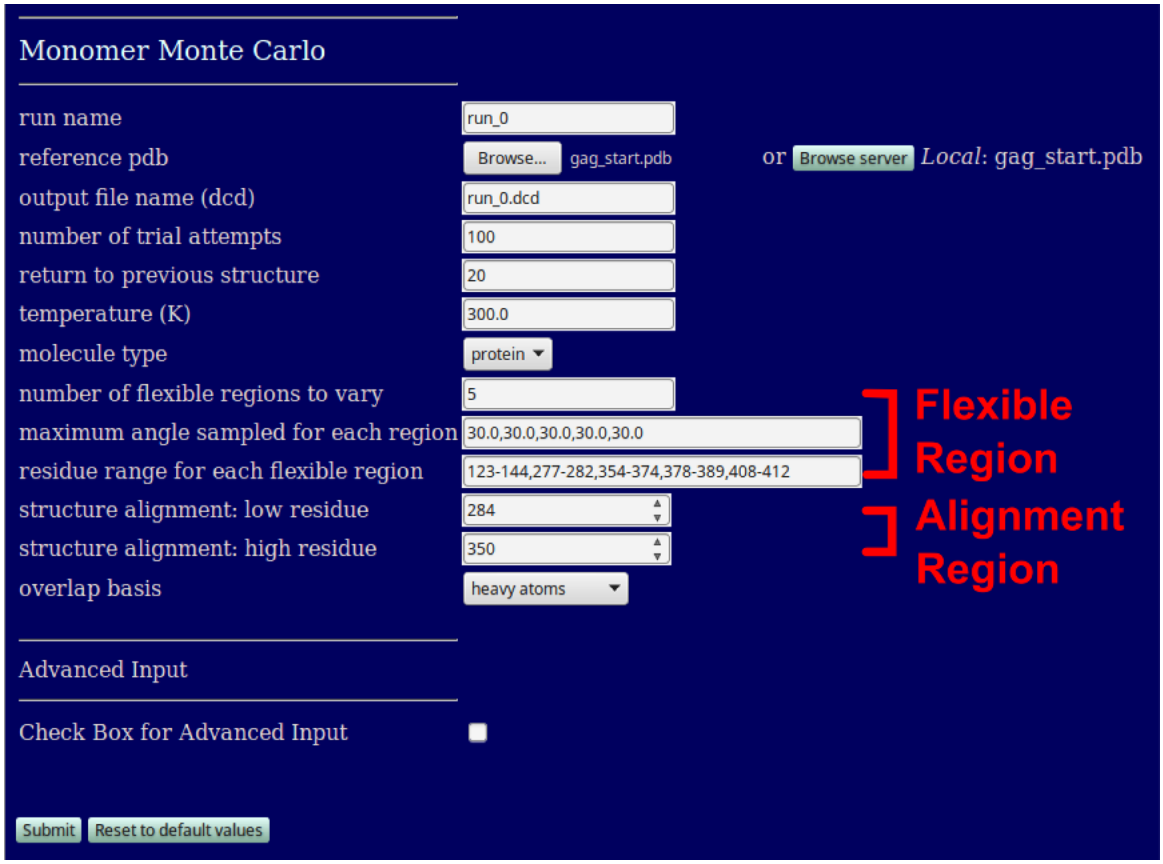


This will reveal a list of buttons for each simulation module running across the top of the page (just below the top bar with the **Session Management** and **Main Menu Toggle** icons).

Click on the 'Monomer Monte Carlo' button.



You should now see a page like the one below. This page is used to enter all of the information needed to run a Monte Carlo simulation.



The figure shows the values for each field as required for our simulation.

Edit the values on your screen to match the screenshot and check that the values match the table provided above for the starting structure. An explanation of the field and how to edit it can be found below.

reference pdb The starting structure for the simulation. Here we use the *gag_start.pdb* file.

- Click on the 'Choose File' button then browse to location in which you saved the file).

number of trial attempts Number of times the simulation will try to vary the structure (some structures will be discarded by the Monte Carlo algorithm) For this tutorial set the value to 1000. For real studies tens of thousands of structures are needed.

The two most system specific sections of the input are highlighted in red. The values entered in these sections detemine the regions to be varied by the simulation and those used to align the structures for comparison.

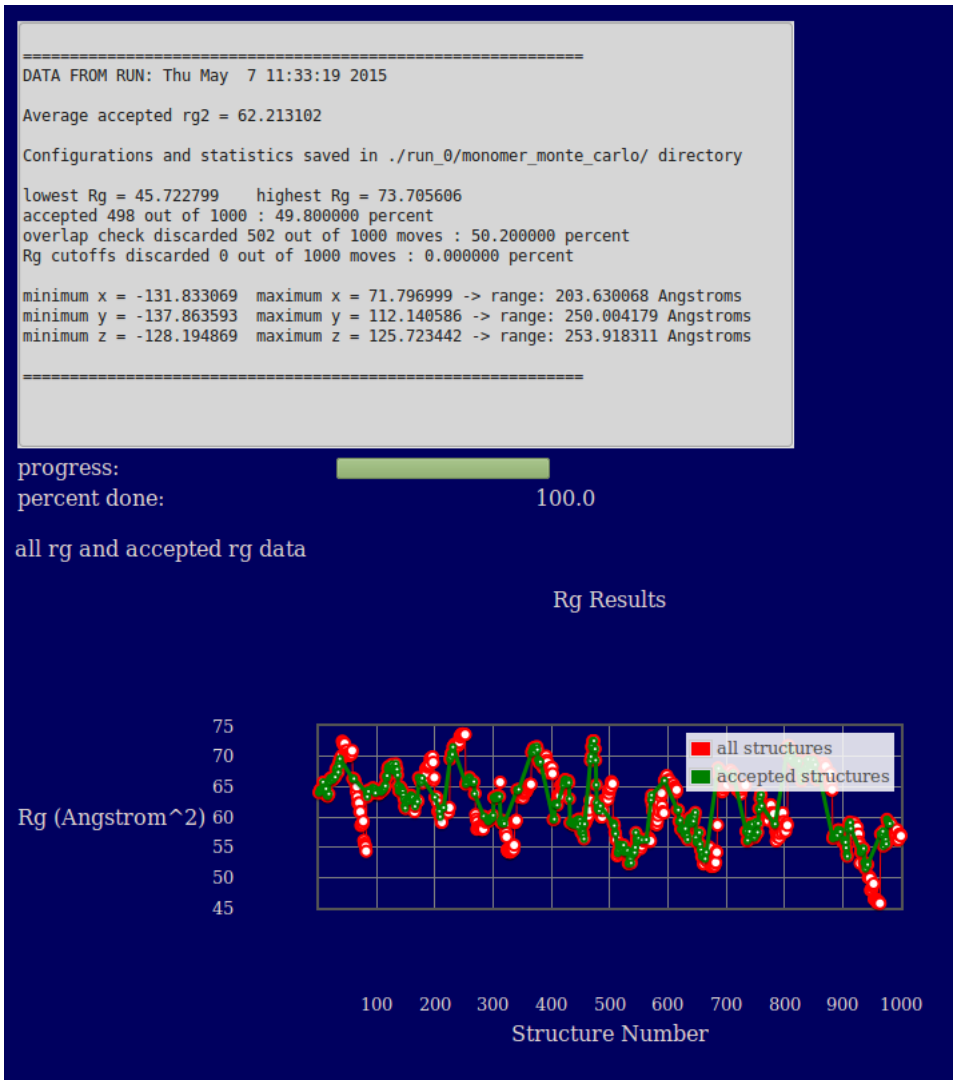
Flexible Region Selected regions of the protein will be varied using the Monte Carlo alorithm. The entries are expected in the formats listed below and shown in the figure.

- number of flexible regions to vary: single number
- enter maximum angle sampled in a single Monte Carlo step for each region: comma separated list
- first residue per region: comma separated list
- number contiguous residues per region: comma separated list

Alignment Region Each frame will be aligned using the residues between the low and high residues selected. This makes it easy to make visual comparisons of each frame in the output trajectory.

Once you have understood the input fields and made sure that your values agree with the figure click on the 'Submit' button to start simulation.

As the run continues the progress bar beneath the submit button should update. A graph beneath this should will show the variation of the radius of gyration over the steps of the Monte Carlo simulation. Once complete the output should look similar to the figure below.



What have we generated:

no_project_specified/run_0/monomer_monte_carlo

- *run_0.dcd*: DCD file containing all of the structures accepted by out Monte Carlo simulation
- *run_0.dcd.accepted_rg_results_data.txt*: text file containing radius of gyration for all structures that made it into the DCD file
- *run_0.dcd.all_rg_results_data.txt*: text file containing radius of gyration for all structures generated
- *run_0.dcd.stats*: text file containing statistics for our Monte Carlo run

Vizualization

You should now download the output trajectory using the file browser.

- Click on the filing cabinet icon in the Session Management area.

We did not select a project when starting this tutorial, so our files will be saved in the 'no_project_specified' directory

Note: the 'Configurations and statistics saved in ' line in the output gives a relative path).

- Click on the trangle next to the 'no_project_specified' directory name to reveal the 'run_0' directory created by our simulation.
- Check the box next to 'run_0' to select it for download.
- Beneath the file tree is an option labelled 'Compression type'. Select an option suitable for your operating system from the list (for Windows select 'zipped' for Linux 'bzip2 tarball').

- Click the 'Download' button.

A progress bar will appear monitoring the upload of your files to the server. Once complete a link will appear beneath the download button.

- Click on this to download your data.

Once the download is complete uncompress the file in a location of your choice.

You should now load the PDB gag_start.pdb (you will find this in the run_0 directory you just downloaded) and DCD (run_0/monomer_monte_carlo/run_0.dcd) into VMD to observe the variation produced even in our very short Monte Carlo simulation. Remember the DCD file contains coordinates alone, you need to load the PDB first so that the visualization software knows about the atoms they represent and how they are connected.

SAS Curve Calculation

Next we calculate a theoretical scattering curve for each of the trial structures we have generated. The SASSIE workflow operates by calculating the scattering intensities at evenly spaced Q values and matching these against interpolated experimental values.

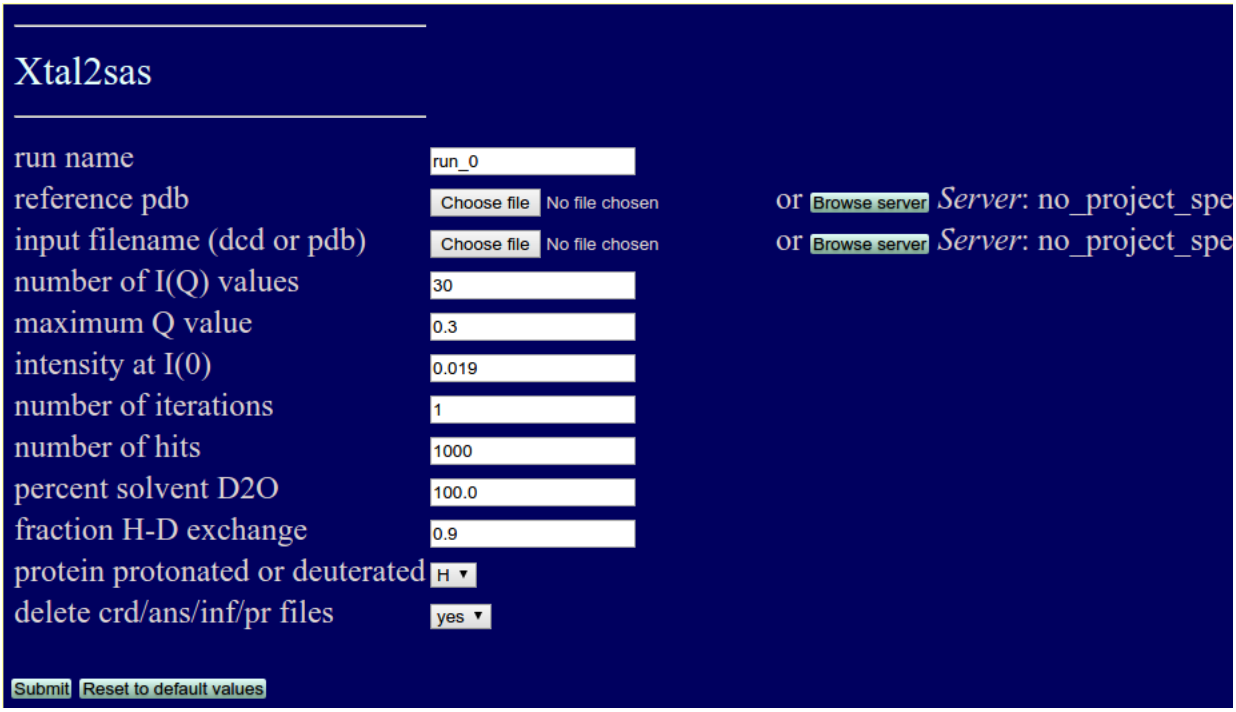
The file gag_scattering.dat contains previously interpolated experimental data (for information on how to do this on your own data see the documentation for the SASSIE Data Interpolation module). In order to create the correct data points in our theoretical curves we need three pieces of information:

- Intensity and Q=0, I(0): 0.019
- Maximum value of Q: 0.3 (units are inverse Angstroms)
- Number of points in the curve: 30

A number of scattering calculators are available in SASSIE, Here we use Xtal2sas.

- Select Calculate from the Main Menu.
- Select Xtal2sas from the Module Menu.

Now you need to enter the information to run the scattering calculator. Other than the values listed above you can keep the default values for this tutorial (see figure below).



The **reference pdb** is the same one we used to start the simulation and the **input filename** (DCD) comes from the previous step, so both are already uploaded to the SASSIE-web server. To select both of them follow you the same procedure:

- Click on the 'Browse server' button next to the appropriate field.
- Navigate to the file requires
  - PDB: no_project_specified/gag_start.pdb
  - DCD: no_project_specified/run_0/generate/run_0.dcd
- Click 'OK'

Now all input fields should be complete.

- Click 'Submit'

A scattering curve will be calculated for all of the structures generated by the Monte Carlo simulation (the progress bar should reach 100% and a message stating the run finished appear in the window beneath when the job has completed).

What have we generated:

no_project_specified/run_0/xtal2sas

- *.*iq*: files for each structure we generated containing theoretical scattering data
- *.*log*: log files for each run of xtal2sas
- *.*ans*: input file used for each xtal2sas run

SAS Curve Comparison

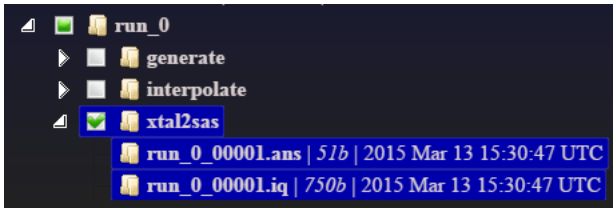
Now we compare our theoretical curves to the experimental data to see which of our structures are plausible models of the real protein.

- Select Analyze from the Main Menu.
- Select Chi-Square Filter from the Module Menu.

We now need to select the path containing the theoretical scattering curves and the file containing the experimental data. In addition we need to input the value of I(0) to enable comparison of the two curves (see the picture below).

To set the path to the scattering curves generated in the previus step:

- Click on the 'Browse server for a path' button
- Navigate to and select the run_0/xtal2sas folder (see picture below)



- Click 'OK'

interpolated data file

- Click on the 'Choose File' button
- Navigate to and select the gag_scattering.dat file on your local computer
- Click 'OK'

I(0)

- Enter the value 0.019

We now need to create files which record which frames meet criteria which make them successful models of our data. This means those with low chi square values.

number of weight files

- Set this as 1.

Weight files contain information on which frames in our simulation meet specific criteria provided in the expression box.

enter expression

- Enter the following expression:

> $x2 < 5$

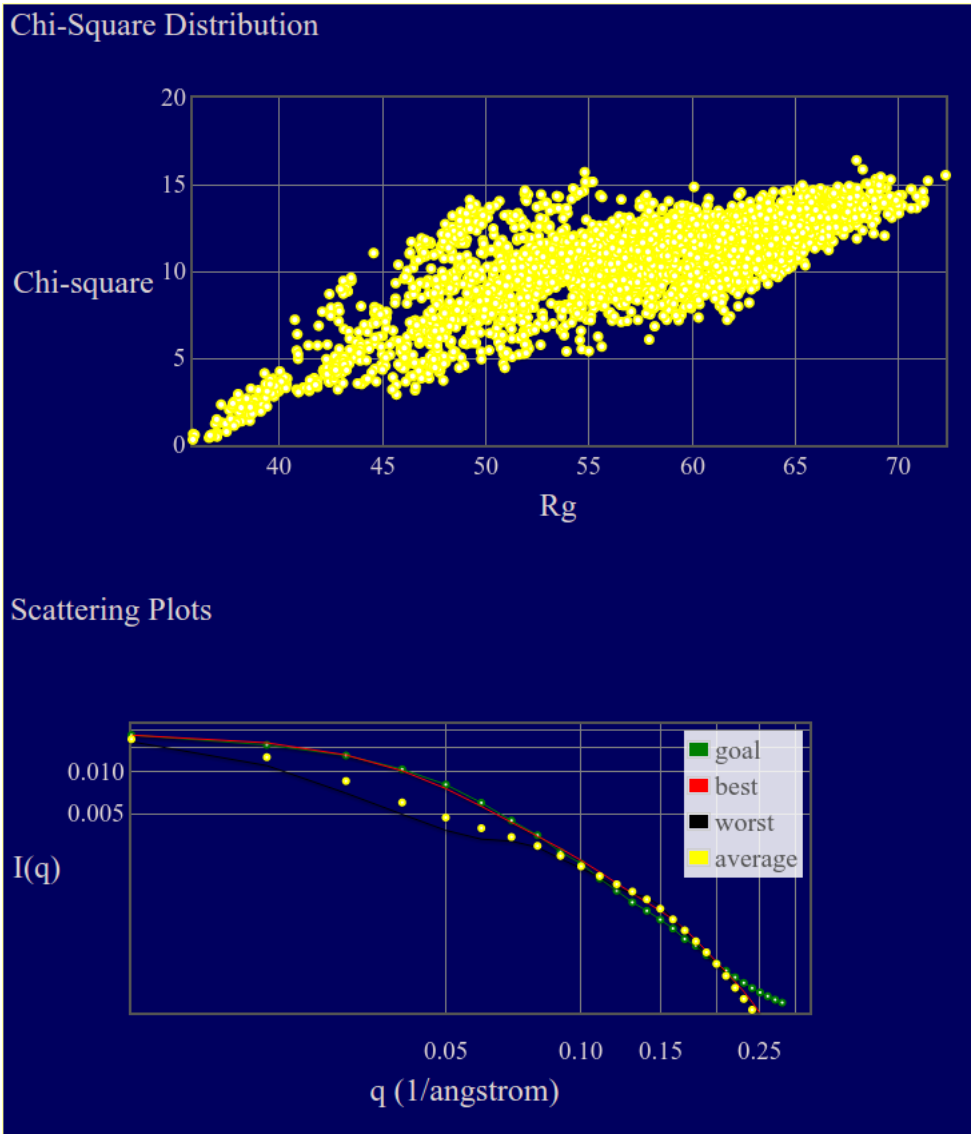
This selects all frames with a chi square less than 5.

weight file name

- Enter x2_lt_5.txt
- Click 'Submit'.

Note: There are list boxes that allow the selection of the format of the input theoretical curves and the metric used to compare the curves. Here we wish to use the defaults of 'xtal2sas' and 'reduced chi-square'.

Once complete you the run you should see outputs like those below.



The top plot shows the variation of chi squared (y-axis) with the radius of gyration (x-axis). Chi squared is a measure of the quality of fit of the theoretical curve to the experimental one. It is a percentage and the lower the value the better.

The bottom plot shows a direct comparison of the best, worst and average theoretical curves with experiment (goal).

In the text output you will see the minimum che square (X2) values is given. If this is higher than 2 then you can alter the X2 low cutoff to a value just above the minimum value to ensure some structures are selected in the next step.

What have we generated:

no_project_specified/run_0/filter

- *averagefile.txt*: average scattering curve for all structures
- *bestworstfile.txt*: best and worst scattering curves selected from all structures
- *rglowweights.txt*: weights file selecting only frames with Rg < 40
- *rghighweights.txt*: weights file selecting only frames with Rg < 60
- *sas_spectra_plot.txt*: goal, best, worst and average scattering curves
- *x2_vs_rg_plot.txt*: chi squared against radius of gyration for all structures
- *x2file.txt*: chi squared for all structures
- *x2lowweights.txt*: weights file selecting only frames with chi squared < 2
- *x2highweights.txt* weights file selecting only frames with chi squared < 10

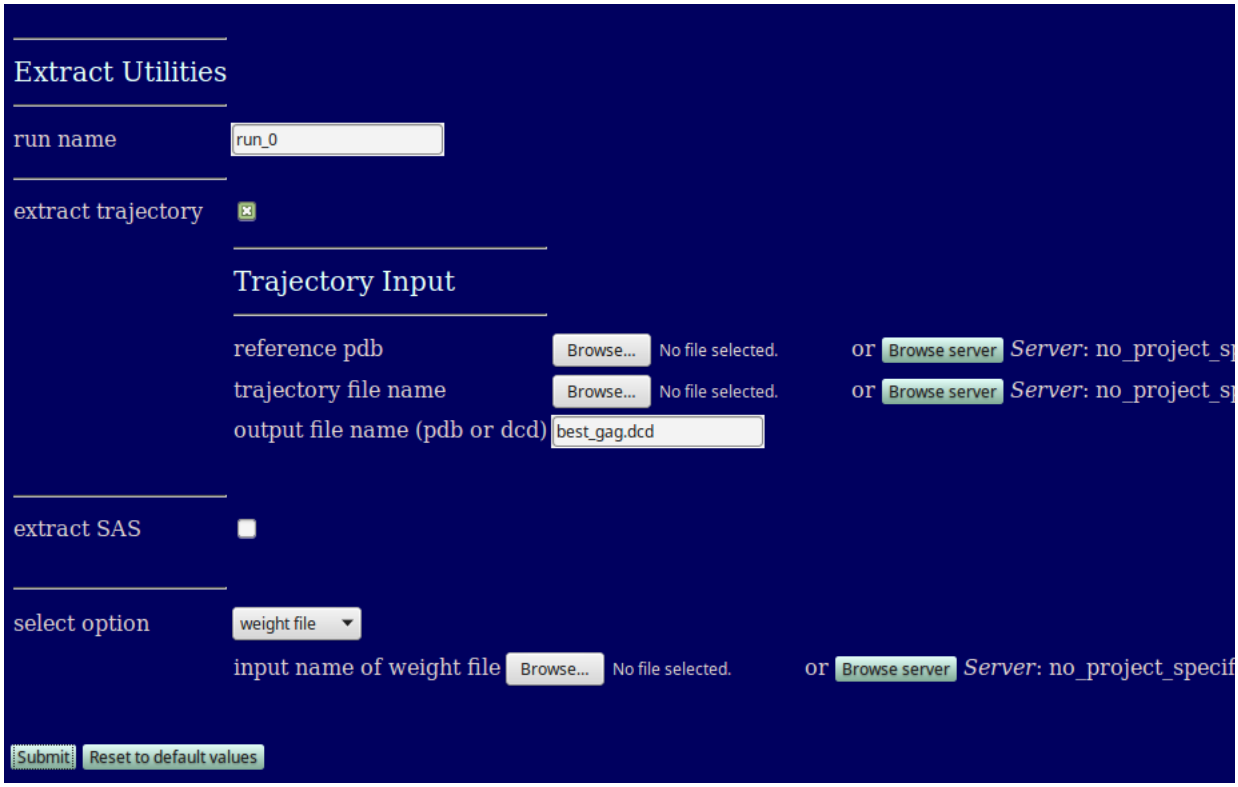
spectra

- *.*ciq*: scattering curves scalled to correct I(0) for each structure

Trajectory Filtering

The final stage of this tutorial is to filter out the best fit structures and vizualize them.

- Select Tools from the Main Menu
- Select the Extract Utilities module



In this module we can select structures from the DCD we created from the Monte Carlo simulation using the weight files generated in the Chi-Square Filter module.

By now you should be familiar with how to select files from the server, so no detailed instructions are provided here.

- Check the tick box labelled 'extract trajectory' (this will reveal the options shown in the screenshot)
- Select the usual 'reference pdb' and the DCD output from the Monte Carlo simulation
- Input 'best_gag.dcd' as the 'output filename'
- Choose 'weight file' from the 'select option' listbox.
    - Where it says 'input name of weight file' select the 'x2_lt_5.txt' file generated in the last step

The 'x2lowweights.txt' file selects only the frame which had a chi squared value of less 5.0. If you noticed that you did not get any structures this good in your simulation then go back the the chi square filter and use a more generous cut off.

- Click 'Submit'

When the process is finished you should get an input saying how many frames were selected like that below:

```
============================================================
DATA FROM RUN: Sun May  3 04:55:43 2015

reading frames from /share/apps/genapp/sassie2/results/users/d/no_project_specified/run_0/monomer_monte_carlo/run_0.dcd
writing frames to run_0/extract_utilities/best_gag.dcd
wrote 33 frames to run_0/extract_utilities/best_gag.dcd

============================================================
```

In the event that none of your frames pass the filter then you can download these preprepared files and try the filtering process:

- DCD: all_gag.dcd
- weights file: x2lowweights.txt

What have we generated:

no_project_specified/run_0/extract_utilities

- *best_gag.dcd*: DCD containg only the frames for which the theoretical scattering curve is a good match to experiment.

Vizualization

Download the 'best_gag.dcd' file as you did the unfiltered DCD and then vizualize the structure again in VMD (you will need to load the PDB first as before). You should see that the filtered structures are all noticeably more compact that the starting structure and the majority of those in the unfiltered DCD.

References

1. Conformation of the HIV-1 Gag Protein in Solution S. A. K. Datta, J. E. Curtis, W. Ratcliff, P. K. Clark, R. M. Crist, J. Lebowitz, S. Krueger, A. Rein, J. Mol. Biol. 365, 812-824 (2007). BIBTex, Endnote, Plain Text
2. SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints J. E. Curtis, S. Raghunandan, H. Nanda, S. Krueger, Comp. Phys. Comm. 183, 382-389 (2012). BIBTeX, EndNote, Plain Text

Return to Main Documents Page

Go to top

Supported via CCP-SAS a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant