NOTE: This PDF file is for reference purposes only. This lab should be accessed directly from the web at https://sassie-web.chem.utk.edu/sassie2/docs/sample_work_flows/ssrna_example/ssrna_example.html.

Return to Main Documents Page

ssRNA Example SASSIE Workflow

In this workflow example we will model the conformation of a portion of an intron from the HIV-1 viral single-stranded RNA. Retroviral RNAs are often exported from the cell nucleus with some or all of their introns. These RNAs eventually serve as genomes (encoding for Gag and other viral proteins) that are encapsulated in virions that leave the cell to infect other cells.

Residues	Color	Flexible?
1-23	cyan	No
24-30	red	Yes
31-46	blue	No
47-55	green	Yes
56-80	magenta	No

The RNA portion to be examined consists of 80 nucleotides and is called trunc2a. A starting structure was built using the RNAComposer tool. The energy-minimized structure, which is divided into 5 different regions, is shown in the figure above. The regions are listed by residues in the table alongside the structure. The 2 regions that were identified as flexible (colored red and green) are indicated in the table. It is not known whether the base pairings in the other 3 regions remain intact in solution. Therefore, SAXS data were obtained on the trunc2a sample at concentrations of 1.0 and 1.42 mg/ml. The data were consistent at these two concentrations, so the higher concentration will be used here for modeling.

The purpose of this exercise is to test whether good fits to the SAXS data can be obtained by allowing only the green and red regions in the figure to be flexible. If structures with large enough Rg values can't be obtained, it will be necessary to allow some of the residues in the blue region to be flexible also in order to "open" up the structure and allow structures with larger Rg values to be sampled.

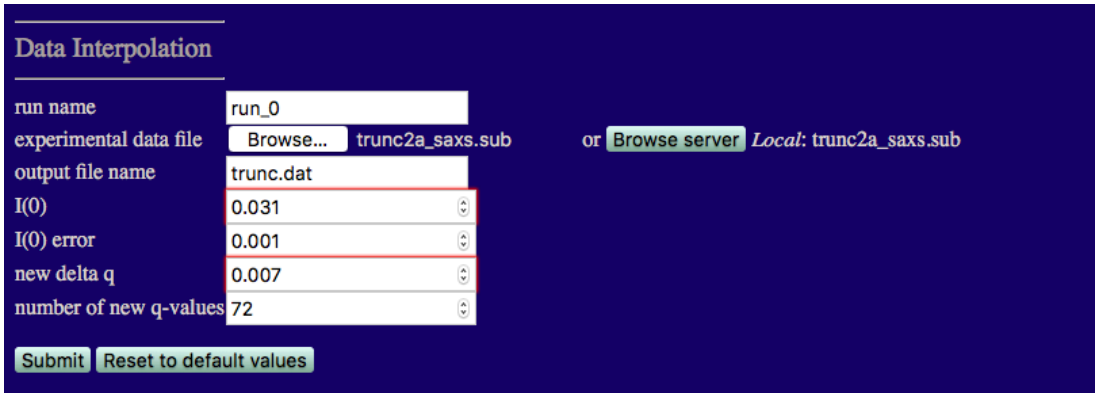
To begin, log into your SASSIE-web account and choose an existing project or create a new project for your work. In this example, the project name is **test**.

Download the trunc2a SAXS data and the minimized starting trunc2a structure.

Data Interpolation

Here we interpolate the SAXS data from trunc2a ssRNA at a concentration of 1.42 mg/ml. More information on the Data Interpolation module is available in the Data Interpolation documentation.

Select the 'Data Interpolation' button from this 'Tools' menu. You should now see a page like the one below. This page is used to enter all of the information needed to do the data interpolation.



The figure shows the values for each field as required for data interpolation.

Edit the values on your screen to match the screenshot. An explanation of the field and how to edit it can be found below.

run name: user defined name of folder that will contain the results.

experimental data file: Name of input file with experimental data with at least three columns: q, I(q), and error in I(q). Here we use the *trunc2a_saxs.sub* file.

output file name: Name of file that will contain the interpolated data. Here we choose the name *trunc.dat*.

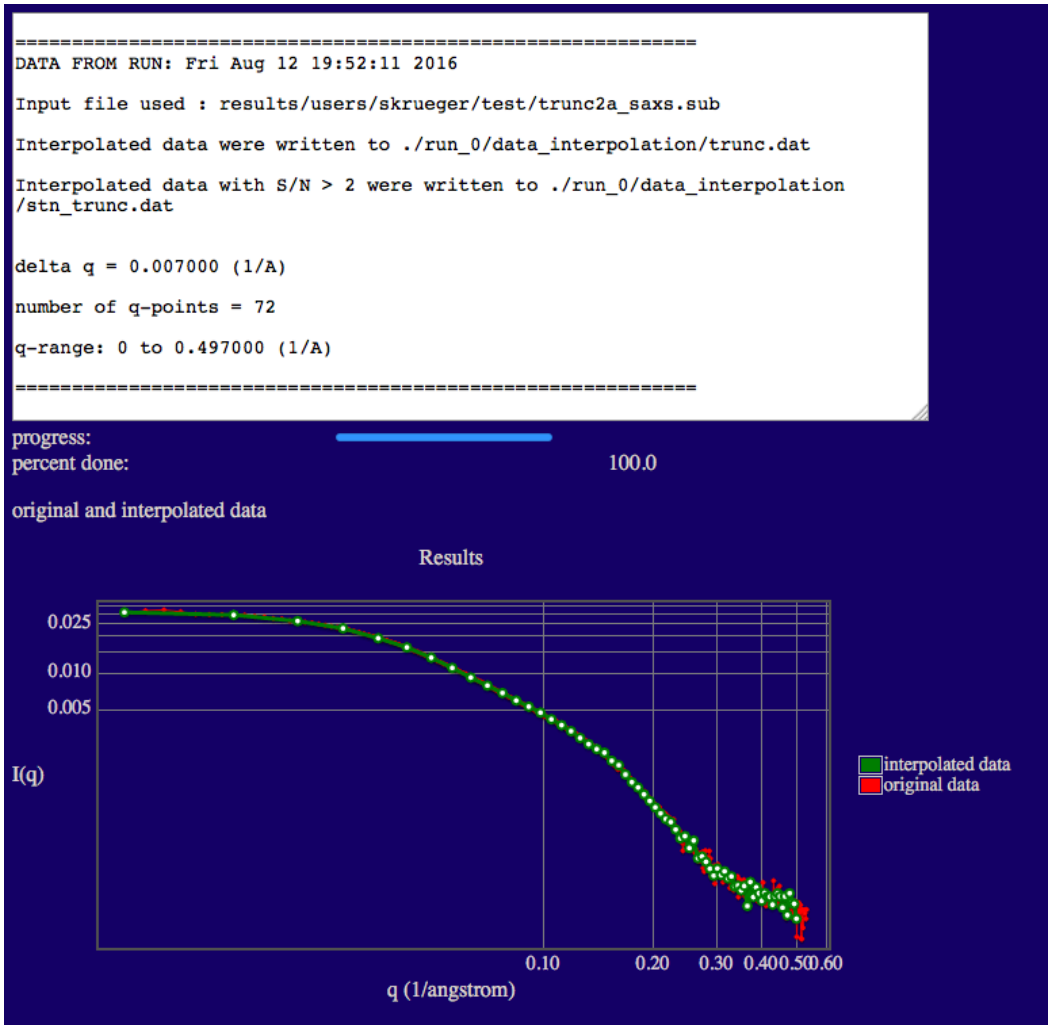
I(0): Experimentally determined value of scattering intensity at q = 0. Here we used the value of 0.031 that was derived from a Guinier fit to the data.

I(0) error: Experimentally determined value of the error of the scattering intensity at q = 0. Here we use the value of 0.001 that was obtained from the Guinier fit to the data.

new delta q: Desired spacing of q-values (1/Angstrom). This should be chosen so that your first interpolated data point falls within the q-range of the experimental data. For this tutorial, the value has been set to 0.007 since the first data point occurs at a value of ~0.007.

number of new q-values: Integer number of desired q-values. For this tutorial, the value has been set to 72 to that the maximum q value is ~0.5.

Click on the 'Submit' button to start simulation. Once complete the output should look similar to the figure below.



The output will show a plot of the original and interpolated data, the name of the input file and the name of the interpolated data file as well as the directory in which it is located.

Note that roll-over help will indicate options to resize, zoom and reset the view of the plot.

What have we generated:

test/run_0/data_interpolation

- *trunc.dat*: text file containing the interpolated SANS data

- *stn_trunc.dat*: text file containing a the interpolated SANS data truncated at the q-value where the signal-to-noise drops below a value of 2

PDB Scan and Structure Minimization

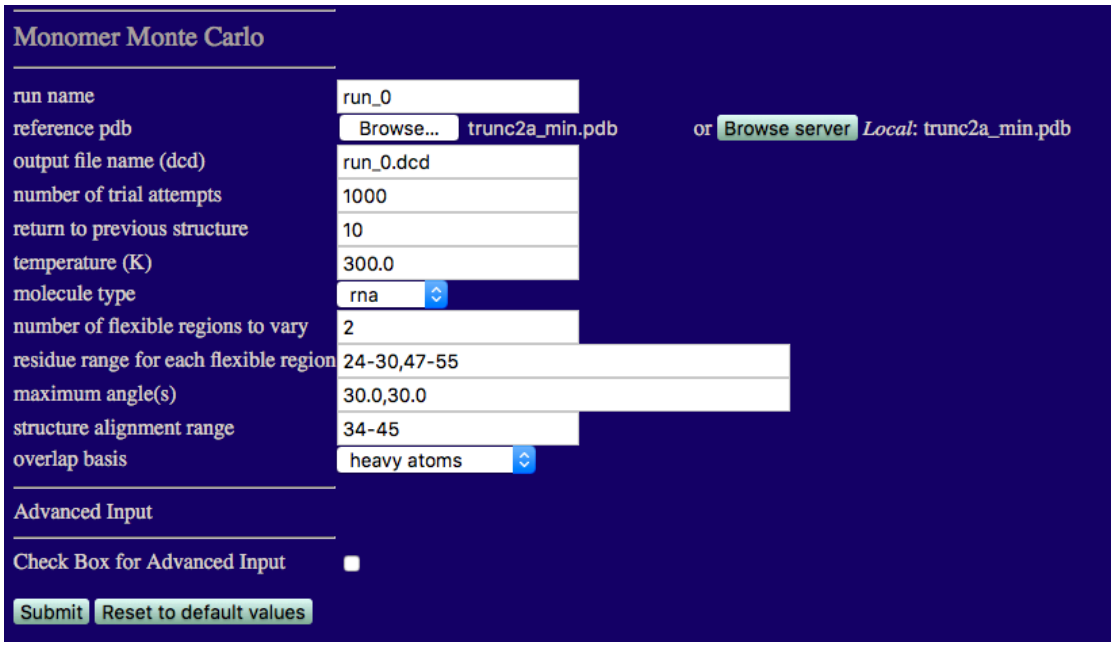
PDB Scan only supports PDB files for proteins at this time. Therefore, the ssRNA PDB file had to be checked and corrected to conform to the CHARMM naming convention by hand. As mentioned above, the starting structure was minimized prior to the simulation. This can be done in SASSIE using the Energy Minimization module. More information can be found in the Energy Minimization documentation.

Structure Variation - Monte Carlo (Monomer)

Here we use the Monomer Monte Carlo module to vary the ssRNA structure. More information can be found in the Monomer Monte Carlo documentation.

Select the 'Simulate' button from the Main Menu of SASSIE-web and then 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. An explanation of some of the fields can be found below.

reference pdb: The starting structure for the simulation. Here we use the trunc2a_min.pdb 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.

return to previous structure: Number of discarded structures in a row that are considered before returning to a randomly-selected structure that was previously accepted

number of flexible regions to vary: single number

residue range for each flexible region: comma-separated list of the range of residues to vary for each flexible region

maximum angle(s): comma-separated list of the maximum angle sampled in a single Monte Carlo step for each flexible region

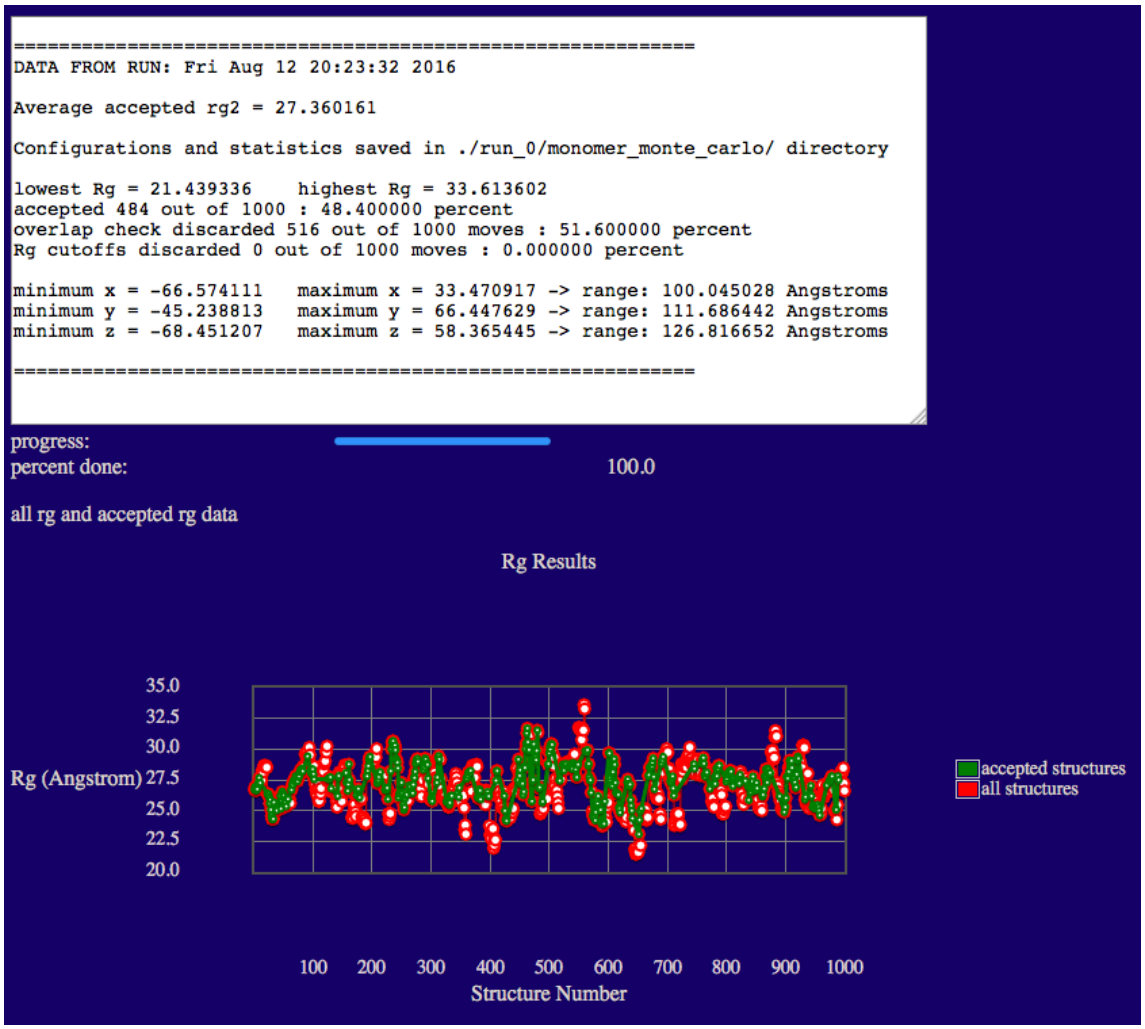
structure alignment region: a single range of residues for structural alignment of all the flexible segments. This makes it easy to make visual comparisons of each frame in the output trajectory.

overlap basis: Select either heavy atoms, all, backbone or enter atom name. The atom name option will spawn futher inputs:

- **overlap basis:** Enter an atom name to check for overlap.
- **overlap cutoff (angsgtroms):** Overlap basis atoms closer than this distance defines an overlap condition.

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. When the run is finished, some statistics are output in the white output box. Notice that **484 out of 1000** structures were accepted in this case. 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:

test/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

Visualization

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

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

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

- Click on the triangle next to the 'test' 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 (run_0/monomer_monte_carlo/trunc2a_min.pdb) 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. Note that both of these files can be found in the run_0 directory you just downloaded. 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.

Initial SAS Curve Calculation - Single Structure

We will use SasCalc to calculate the SAXS curves from the generated structures. More information can be found in the SasCalc documentation. The starting structure **must** be a complete structure without missing residues or atoms (including hydrogen atoms) in order to obtain accurate scattering profiles. Atom and residue naming must be compatiable with those defined in the CHARMM force field.

The SasCalc module is first run using the "converged number of golden vectors" option on just one structure. The SASSIE workflow operates by calculating the scattering intensities at evenly spaced Q values and matching these against interpolated experimental values.

The file trunc.dat contains our previously interpolated experimental data. In order to create the correct data points in our theoretical curves we need three pieces of information:

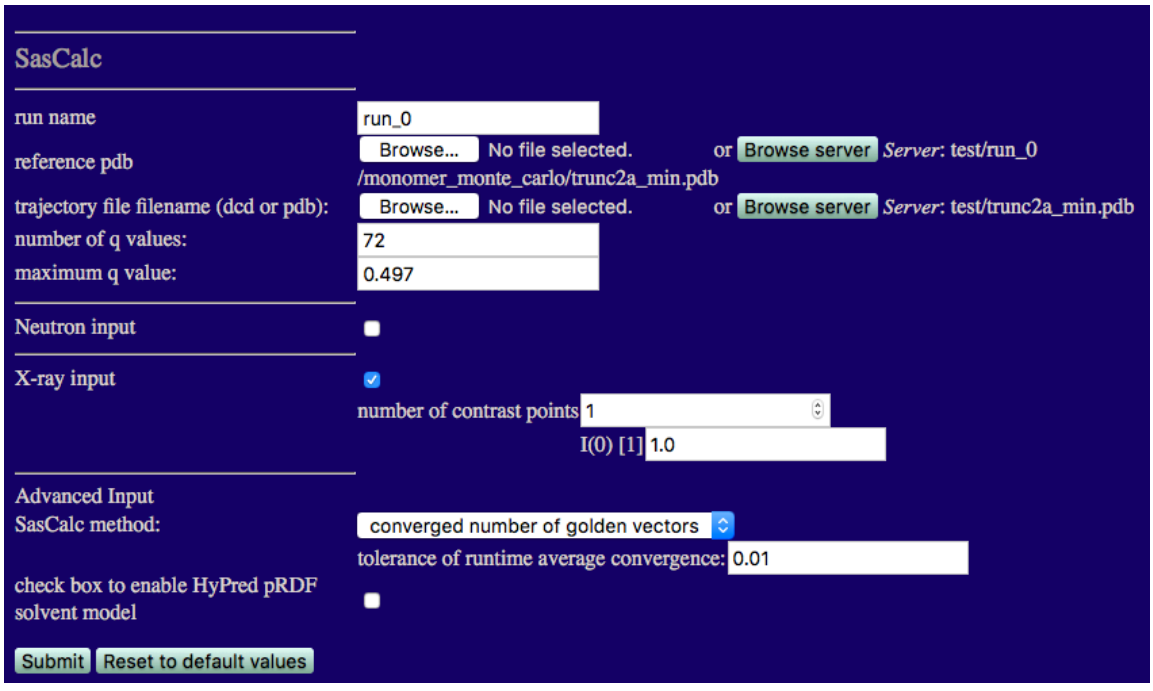
- Intensity and Q=0, I(0): 0.031
- Maximum value of Q: 0.497 (units are inverse Angstroms)
- Number of points in the curve: 72

To begin:

- Select 'Beta' from the Main Menu.
- Click the 'SasCalc' button.

Now you need to enter the information to run the scattering calculator. SasCalc can be used to calculate the scattering for SAXS and SANS and/or for several SANS contrasts at the same time.

Other than the values listed above you can keep the default values for this tutorial, except for the Advanced Input section. Since this is the first time we are running SasCalc on this structure, we are using the **"converged number of golden vectors"** option. Choose this option from the **SasCalc method** menu in the Advanced Input section of the page (see figure below).

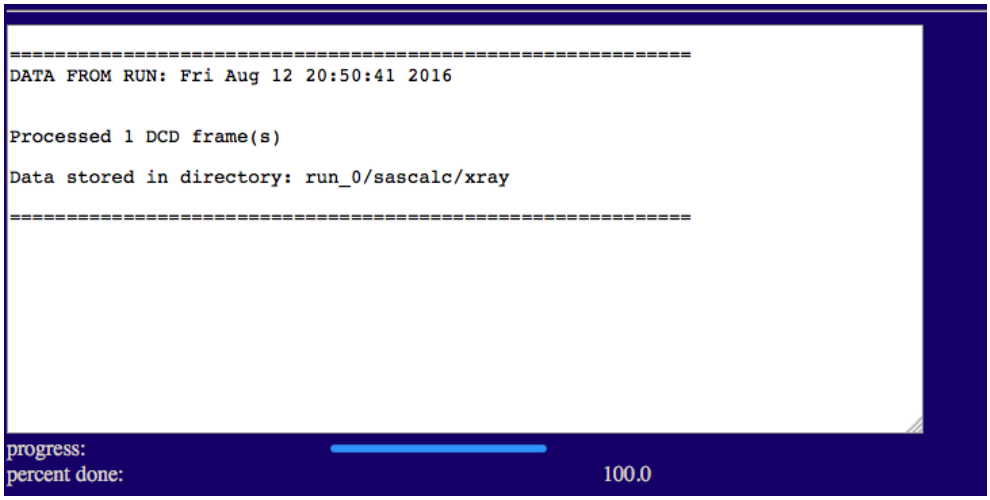


The single structure that we used to start the simulation is used as both the **reference pdb** and the **trajectory file filename** (PDB in this case) so it is already uploaded to the SASSIE-web server. Thus, you can either upload it again from your local computer or locate it on the server and read it from there.

To read the file from the server:

- Click on the 'Browse server' button next to the appropriate field.
- Navigate to the file test/run_0/monomer_monte_carlo/trunc2a_min.pdb
- Click 'OK'

Click on the 'Submit' button to start the calculation. A scattering curve will be calculated for the starting structure (the progress bar should reach 100% and a message stating the run finished appear in the window beneath when the job has completed). Note that the files are written to a the directory sascalc/xray.



What have we generated:

test/run_0/sascalc/xray

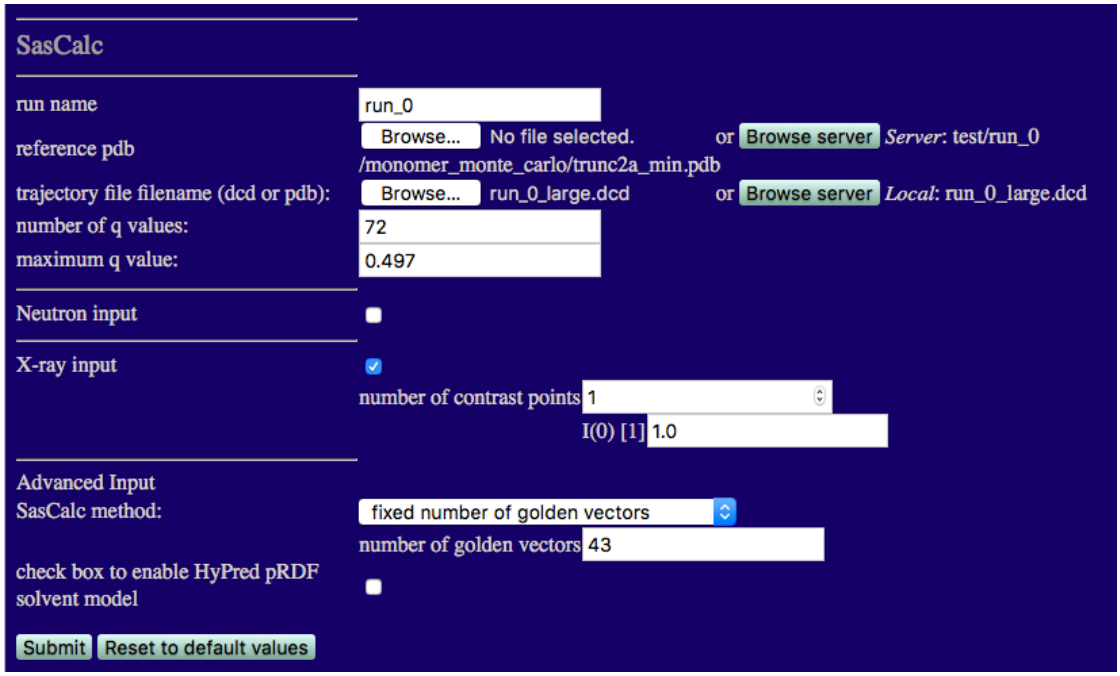
- *run_0_00001.iq*: files containing theoretical scattering data
- *run_0_00001.log*: log files containing information about the structure and calculation inputs

Second SAS Curve Calculation - Entire Ensemble

Next we calculate a theoretical scattering curve for each of the trial structures we have generated. The run_0_00001.log file from the inital SAS Curve calculation indicates that 43 golden vectors were required for convergence to the desired tolerance (0.01 in this case).

```
#Structural Information:
1. PDB input file = /share/apps/genapp/sassie2/results/users/skrueger/test/run_0/monomer_monte_carlo
/trunc2a_min.pdb
2. Number of atoms = 2603; Mw = 26072.08
3. Dimensions = x: -43.08, 19.62, y: -36.90, 54.89, z: -17.31, 14.78
4. Maximum radial dimension Dmax = 115.70 A (contrast-independent)
5. Molecular center of mass = x: -7.66, y: 19.78, z: -3.33 (contrast-independent)
6. Molecular Rg = 26.315412 A (contrast-independent)
#Scattering Intensity:
7. Source = x-ray
8. I(q) .vs. q file = run_0/sascalc/xray/run_0_00001.iq
9: Convergence tolerance used = 0.010000
   Converged number of golden vectors (for complete scattering profile) = 43
10. Io =      1.000
11. Center of Mass = x: -7.65, y: 19.79, z: -3.32 (contrast-dependent)
12. Rg = 26.476143 (contrast-dependent)
```

We now use this information to calculate the scattering curves for all of the generated structures using the **"fixed number of golden vectors"** option from the **SasCalc method** menu as shown below. Since 484 structures are not sufficient to perform a complete analysis of conformation space, we will use a previously generated DCD file run_0_large.dcd that has almost 10,000 structures for further analysis. Download this file and use it as the trajectory filename (see figure below).

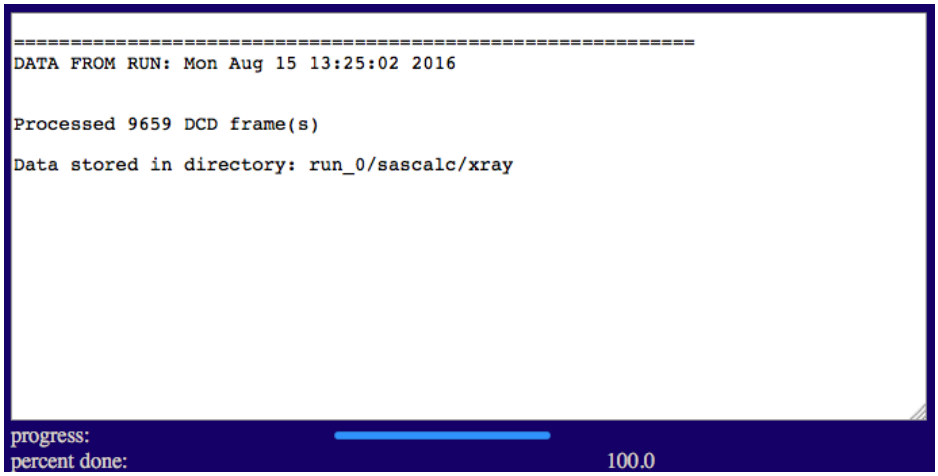


reference pdb: the starting structure that is already on the SASSIE-web server

trajectory file filename: the run_0_large.dcd file that you just downloaded

When all input fields are complete:

- Click 'Submit'.
- **NOTE**: This job takes ~16 min using GPUs, including setup time. Total time can be checked in the Jobs Manager after the job is finished.



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). Note that the files written during the initial SAS curve calculation will be **overwritten** since we chose the same run name (run_0) in both cases. If you wish to save the files from the initial calculation, use a different run name.

What have we generated:

test/run_0/sascalc/xray

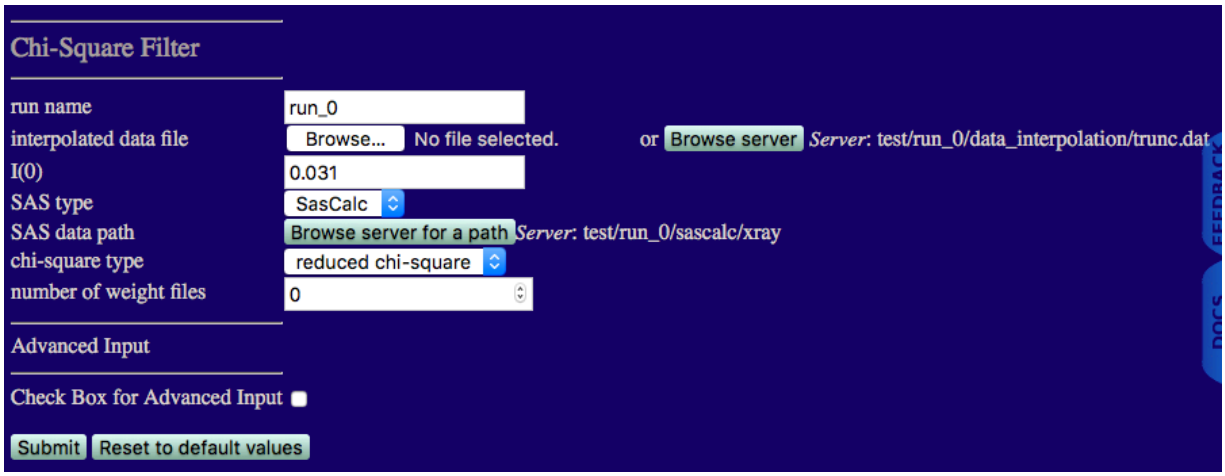
- *.*iq*: files containing theoretical scattering data for all frames in the DCD file (9659 files in this case).
- *.*log*: log files containing information about each structure and calculation inputs (9659 files in this case).

Initial 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 using Chi-Square Filter. More information can be found in the Chi-Square Filter documention.

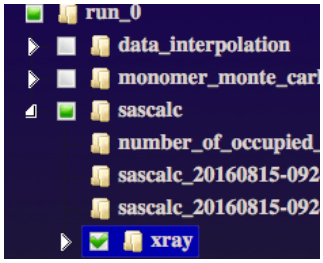
- Select 'Analyze' from the Main Menu.
- Click the 'Chi-Square Filter' button.

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 the test project folder and select the run_0/sascalc/xray folder (see figure below)



- Click 'OK'

interpolated data file:

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

I(0):

- Enter the value 0.031

We eventually may want to create 'weight files' that record which frames meet criteria that make them successful models of our data. This means those with low chi square values. However, we don't know the range of chi square values we have at this stage. So, we set the 'number of weight files' to 0 at this time.

Sas type:

- Choose SasCalc from the menu

number of weight files:

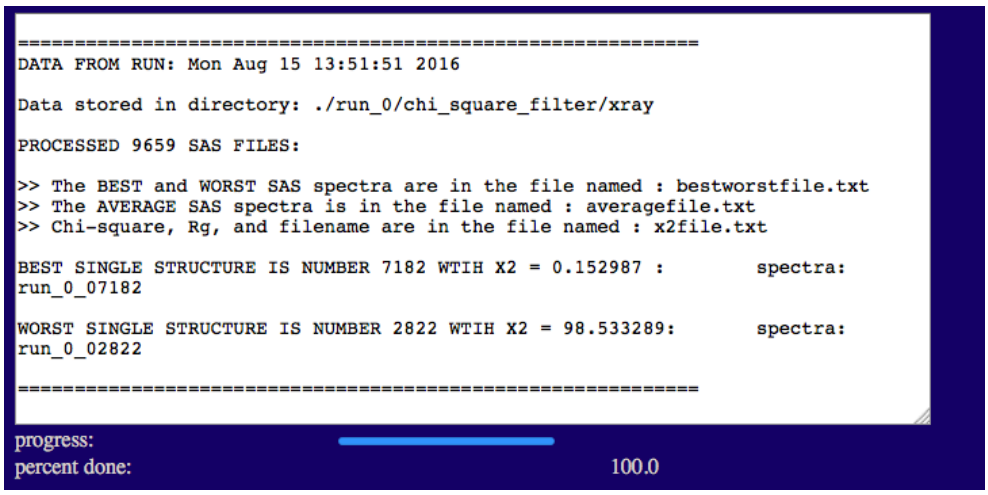
- Set this to 0 using the down arrow associated with the input box. If you type '0' in the box, press the TAB button to make sure this value is

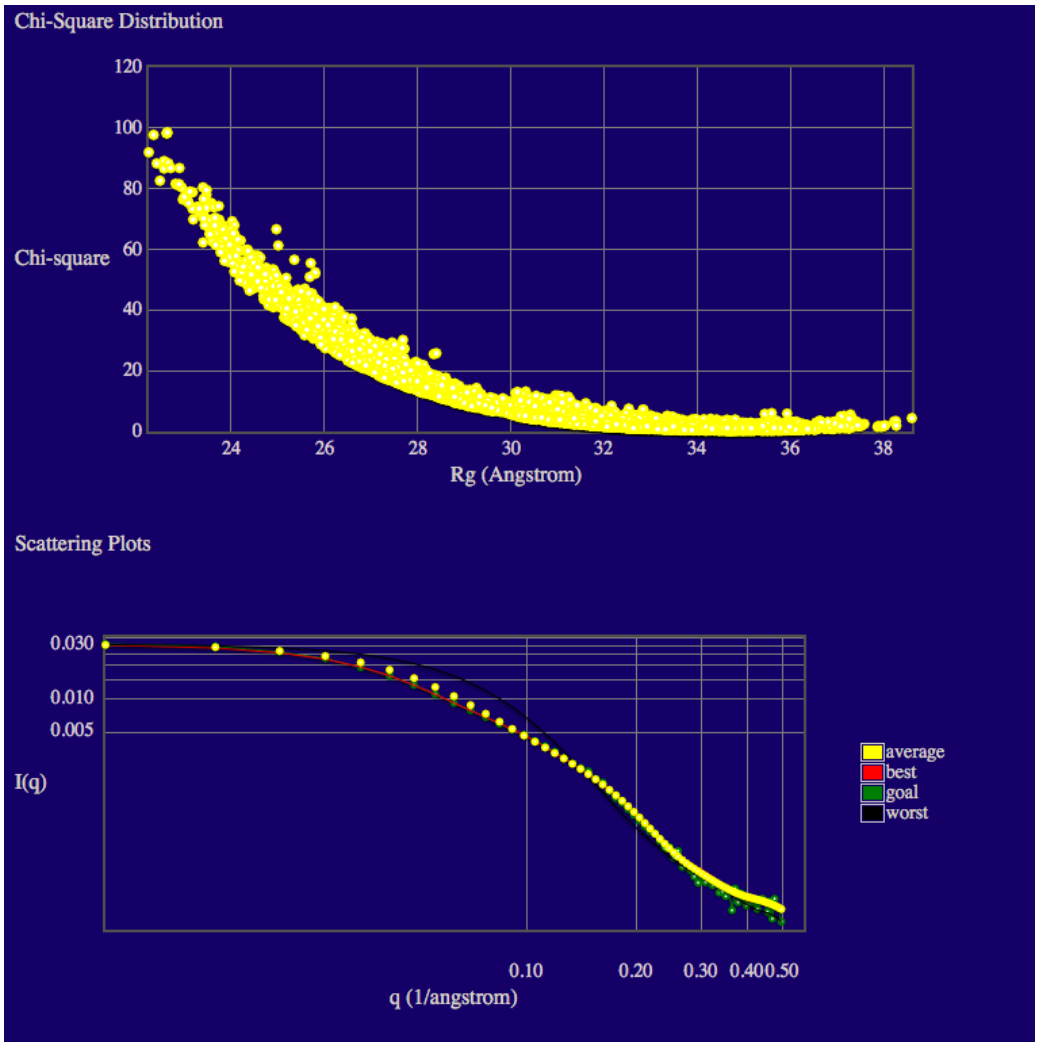
accepted.

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 'SasCalc' and 'reduced chi-square'.

Click 'Submit'.

Once complete you should see outputs similar to those below.





In the text output you will see the minimum chi square (X2) values is given.

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).

Examine the x2 vs Rg plot using the mouse commands: drag to pan, double-click to zoom, to reset zoom and pan: click on title, axis labels or live coordinates box.

What have we generated:

test/run_0/chi_square_filter/xray

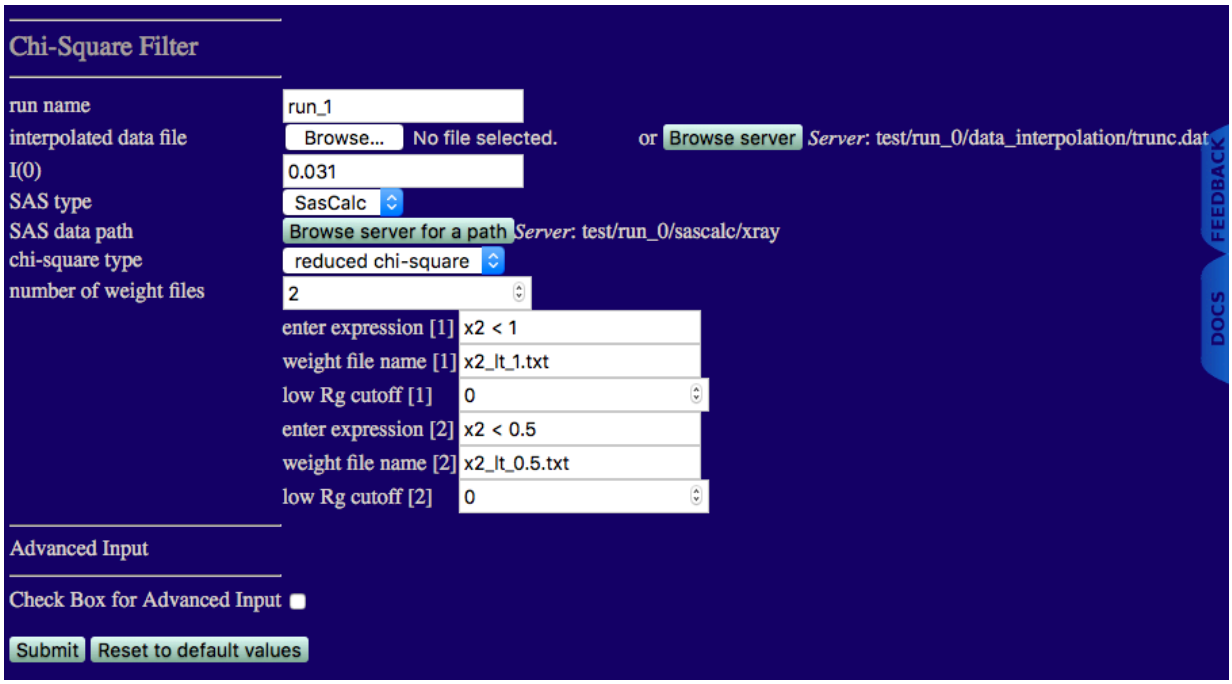
- *averagefile.txt*: average scattering curve for all structures
- *bestworstfile.txt*: best and worst scattering curves selected from all structures
- *sas_spectra_plot.txt*: goal, best, worst and average scattering curves
- *x2_vs_rg_plot.txt*: chi squared against radius of gyration for all structures
- *x2file.txt*: chi squared for all structures

/spectra

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

Second SAS Curve Comparison

Now that we know the range of chi square values that we have, we can compare the theoretical curves to the data a second time and create two weight files that flag all structures with chi square values below a certain number. Now, we set the 'number of weight files' to 2.



run name:

- Since we already have a chi_square_filter folder in the run_0 directory, set the run name to run_1.

number of weight files:

- Set this to 2 using the down arrow associated with the input box.

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

enter expression[1]:

- Enter the following expression:

```
x2 < 1
```

This selects all frames with a chi square less than 1.

weight file name[1]:

- Enter x2_lt_1.txt

low Rg cutoff[1]:

- Enter a value if you wish to also restrict the Rg range to be above this value. The default value is 0 so that all Rg values are acceptable.

enter expression[2]:

- Now enter the following expression for the second weight file:

```
x2 < 0.5
```

This selects all frames with a chi square less than 3.0. Adjust this value if necessary to suit the results from your simulation.

weight file name[2]:

- Enter x2_lt_0.5.txt

low Rg cutoff[2]:

- Enter a value if you wish to also restrict the Rg range to be above this value. The default value is 0 so that all Rg values are acceptable.
- Click 'Submit'.

Once complete the outputs should be exactly the same as for the previous run. The only difference is that the two weight files can now be found in the chi_square_filter/xray folder.

What have we generated:

test/run_1/chi_square_filter/xray

- *averagefile.txt*: average scattering curve for all structures
- *bestworstfile.txt*: best and worst scattering curves selected from all structures
- *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
- *x2_lt_1.txt*: weights file selecting only frames with chi squared < 1.
- *x2_lt_0.5.txt*: weights file selecting only frames with chi squared < 0.5

/spectra

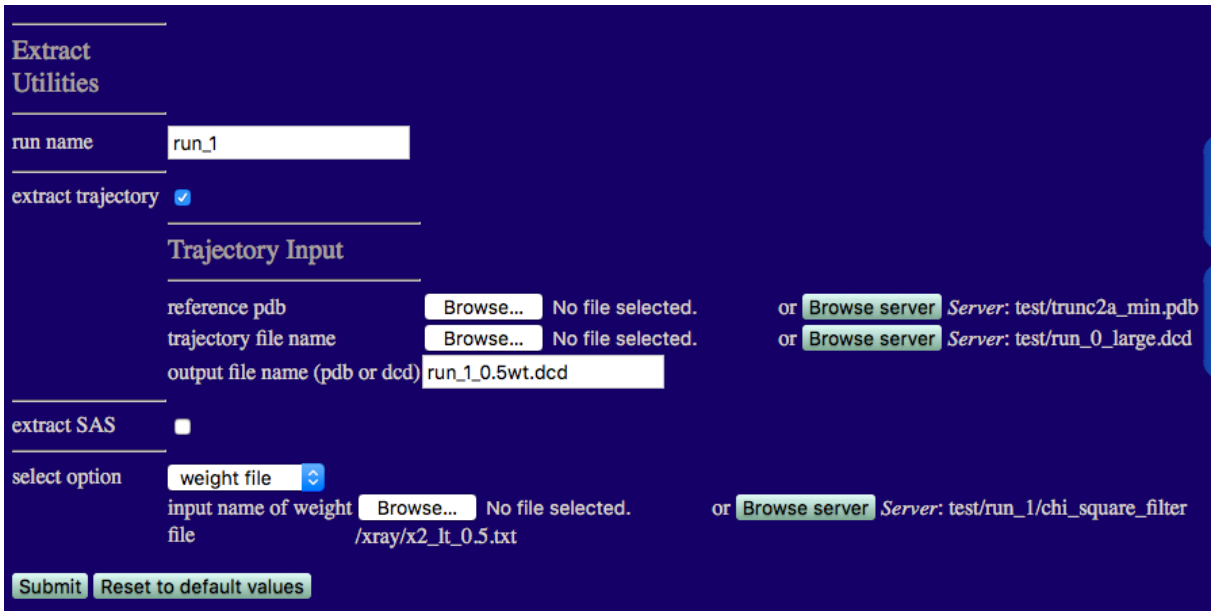
- *spec_*.ciq*: scattering curves scaled to correct I(0) for each structure

Trajectory Filtering

Now we can filter out the best fit structures and vizualize them using the Extract Utilities. More information can be found in the Extract Utilities documentation.

We will extract the frames from run_0_large.dcd that fit the data with x2 values less than 0.5.

- Select 'Tools' from the Main Menu.
- Click the 'Extract Utilities' button.



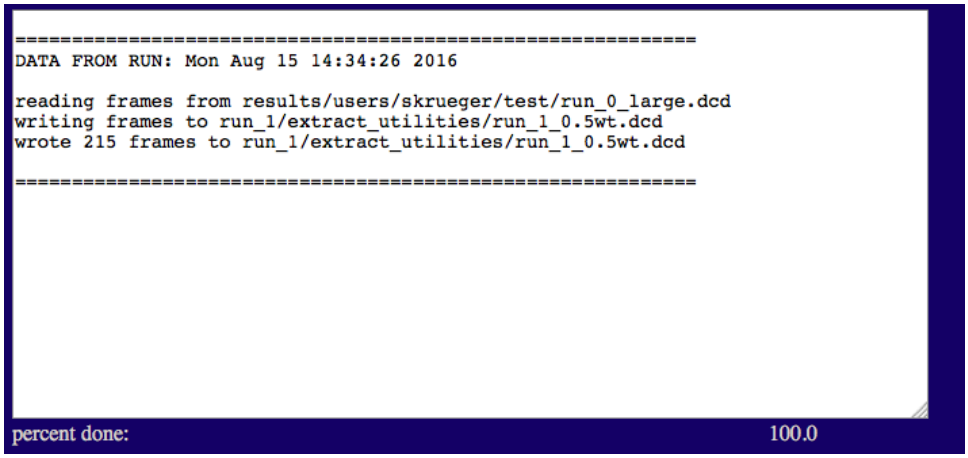
In this module we can select structures from run_0_large.dcd using the weight files generated in the Chi-Square Filter module.

In this case, we chose to select the weight file from the server.

- Check the tick box labelled 'extract trajectory' (this will reveal the options shown in the screenshot)
- Select the trunc2a_min.pdb 'reference pdb' and the run_0_large.dcd file from the server (or from your local computer).
- Input 'run_1_0.5wt.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_0.5.txt' file generated in the last step that selects only the frames which had a chi squared value of less 0.5.
- Click 'Submit'

When the process is finished your output should look like the one below.



What have we generated:

test/run_1/extract_utilities

- *run_1_0.5wt.dcd*: DCD containing only the frames for which the theoretical scattering curve is a good match to experiment.

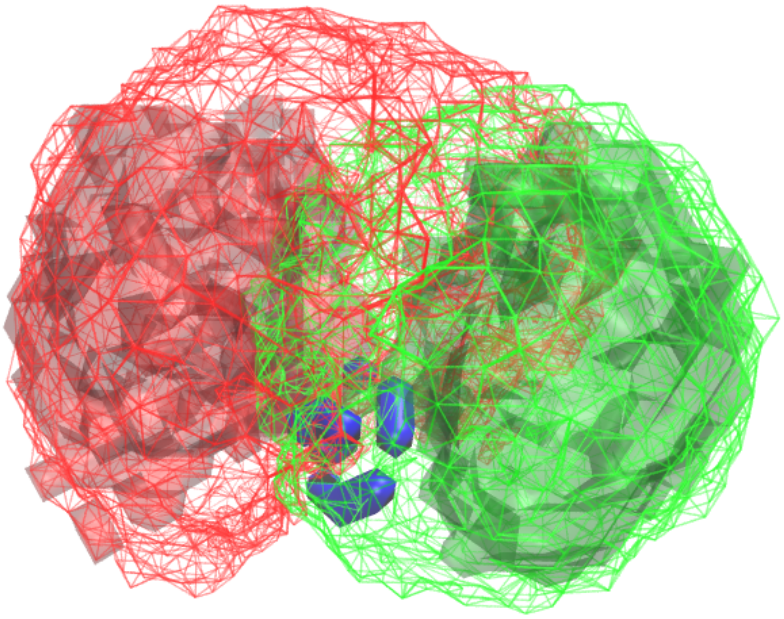
Visualization

Download the 'run_1_0.5wt.dcd' file as you did the unfiltered DCD and then vizualize both the 'run_0_large.dcd' and 'run_1_0.5wt.dcd' trajectories in VMD. (You will need to load a suitable PDB in both cases as before). Can you tell if the filtered structures are more compact than many of those in the unfiltered DCD? This is often difficult to do by examining the trajectories directly.

Density Plot

Another way to visualize the structures sampled in the 'run_0_large.dcd' and 'run_1_0.5wt.dcd' files for comparison is to use the Density Plot module. The purpose of the Density Plot module is to generate files with volumetric data to visualize results. Often, this is used to visualize sub-ensembles of structures that are in agreement with experimental data. More information can be found in the Density Plot documentation.

The density plot below shows the envelope sampled by all of the accepted structures as well as that sampled by only the best fit structures with x2 < 0.5. The blue region is the envelope represented by residues 31-46, which is essentially the fixed alignment region that we defined in the Monomer Monte Carlo module. The red regions represent the envelope sampled by residues 1-30 for all accepted structures (mesh) and for the best fit structures (solid). The green regions represent the envelope sampled by residues 47-80 for all accepted structures (mesh) and the best fit structures (solid). This representation makes it easier to see that the envelope represented by the best fit structures is significantly smaller than that represented by all of the accepted structures.

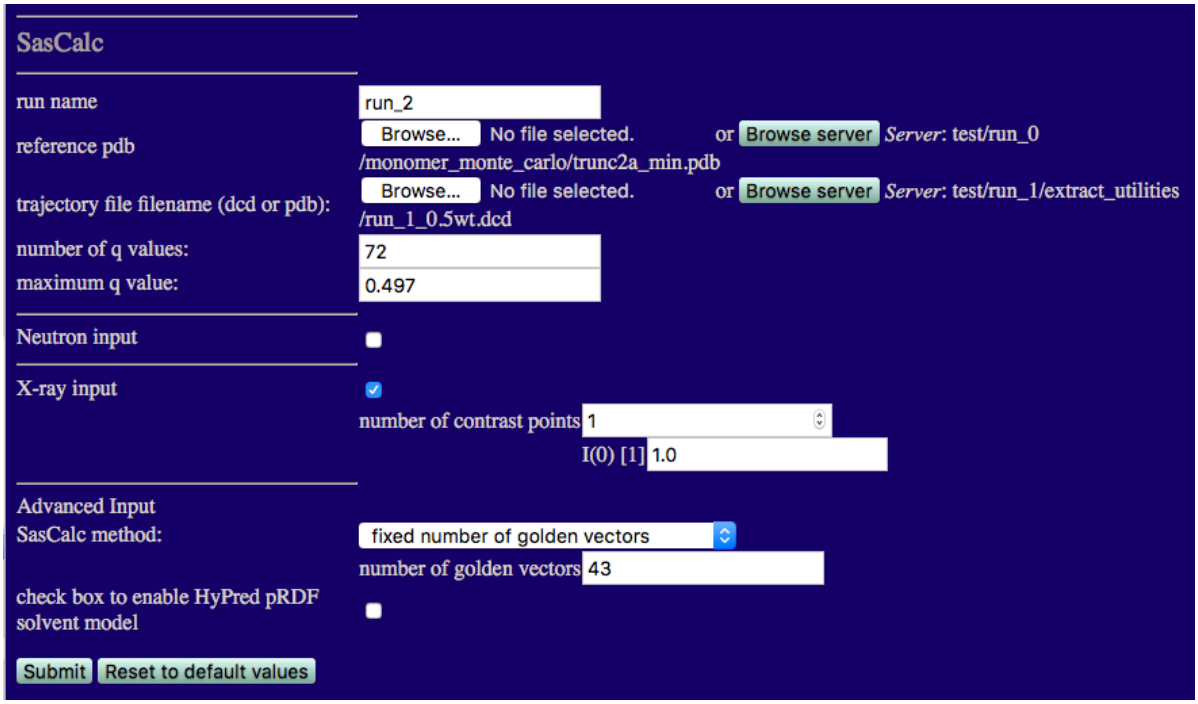


Instructions on how to make the above density plot can be found here.

Final SAS Curve Comparison

Now, we can compare the structures with x2 < 0.5 to the SAXS data by calculating their theoretical SAXS curves and comparing them to the SAXS data again.

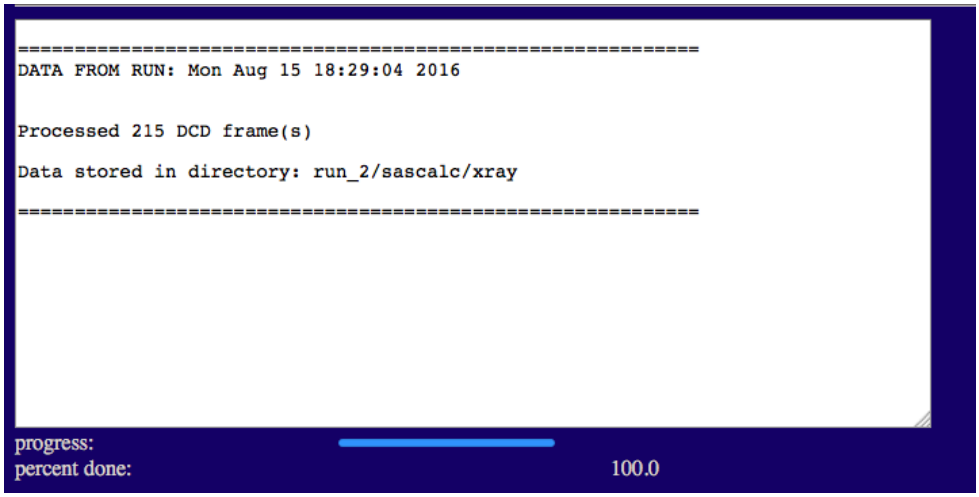
First, calculate the theoretical SAXS curves using SasCalc.



run name:

Set the run name to run_2. Once the inputs have been entered, click 'Submit'.

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

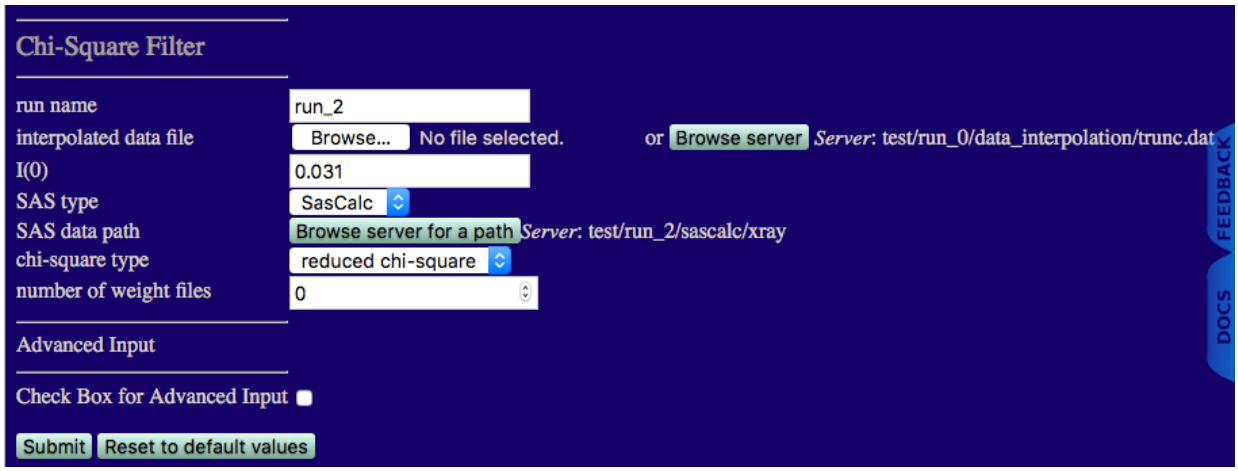


What have we generated:

test/run_2/sascalc/xray

- *.*iq*: files containing theoretical scattering data for all frames in the DCD file (692 files in this case).
- *.*log*: log files containing information about each structure and calculation inputs (692 files in this case).

Then, compare the theoretical SAXS curves to the SAXS data using Chi-Square Filter.



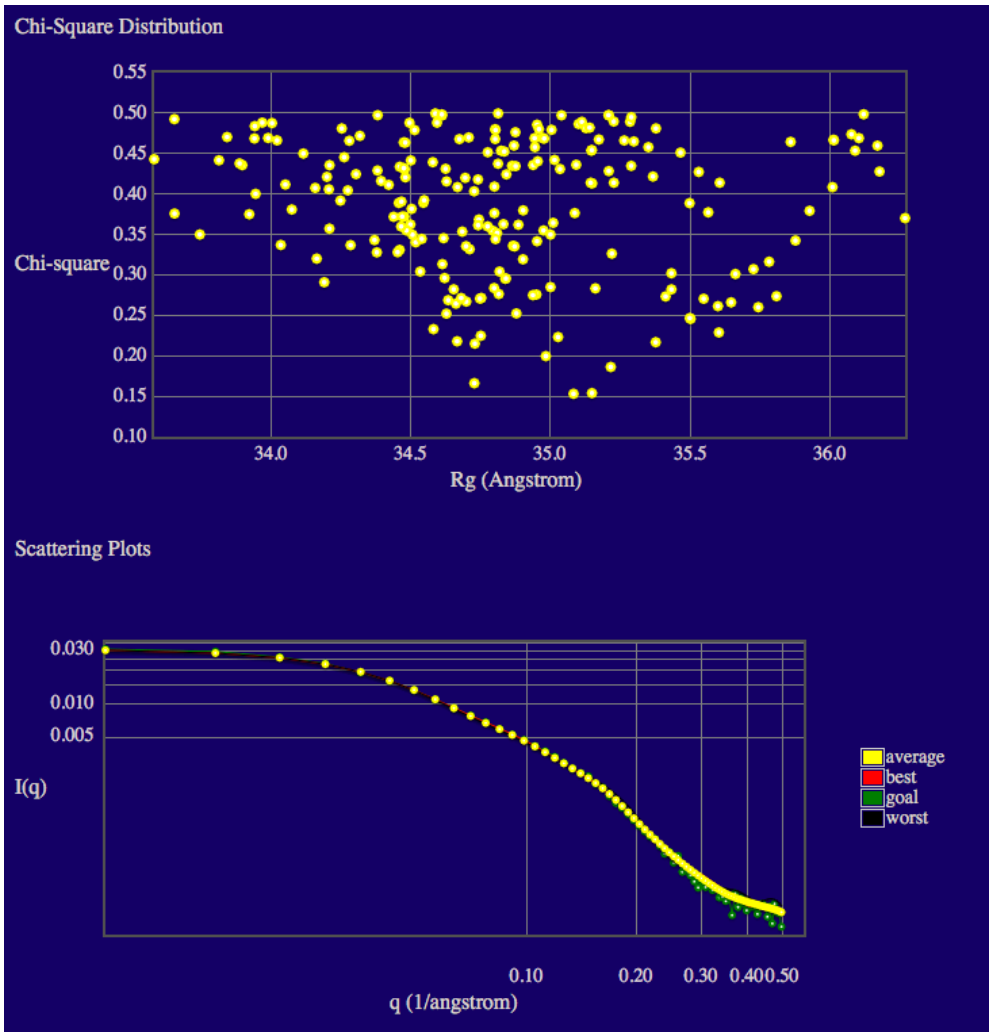
run name:

- Since we already have a chi_square_filter folder in the run_1 directory, set the run name to run_2.
- **number of weight files**
- Set the 'number of weight files' to 0 since we are already dealing with the best fit structures.
- Click 'Submit'.

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

```
============================================================
DATA FROM RUN: Mon Aug 15 18:31:12 2016

Data stored in directory: ./run_2/chi_square_filter/xray

PROCESSED 215 SAS FILES:

>> The BEST and WORST SAS spectra are in the file named : bestworstfile.txt
>> The AVERAGE SAS spectra is in the file named : averagefile.txt
>> Chi-square, Rg, and filename are in the file named : x2file.txt

BEST SINGLE STRUCTURE IS NUMBER 149 WTIH X2 = 0.152987 :         spectra:
run_2_00149

WORST SINGLE STRUCTURE IS NUMBER 26 WTIH X2 = 0.499600:          spectra:
run_2_00026

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

progress:
percent done: 100.0



Notice that the best, worst and ensemble average curves now all fit the data well. This means that any single structure in the ensemble of structures with x2 < 0.5 and any combination of structures in the ensemble, including the entire ensemble itself, represents a good fit to the SAXS data.

What have we generated:

test/run_2/chi_square_filter

- *averagefile.txt*: average scattering curve for all structures
- *bestworstfile.txt*: best and worst scattering curves selected from all structures
- *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

/spectra

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

Visualization

Download the contents of the chi_square_filter directory and reproduce the output plots. This can be accomplished by plotting x2 vs Rg from the x2_vs_rg_plot.txt file and by plotting the goal, best, worst and average SAXS curves vs q from the sas_spectra_plot.txt file using your favorite plotting program.

Minimization of Structures

The structures in trajectories such as run_0_large.dcd or run_1_0.5wt.dcd should be minimized for at least 1000 steps if the trajectories will undergo further processing. More information can be found in the Energy Minimization documentation.

References

1. 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