Return to Main Documents Page

Data Interpolation

Creates a new data file that is spaced in a uniform grid from input experimental data file.

Accessibility

The Data Interpolation module is accessible from the Tools section of the main menu.

Basic Usage

The purpose of the module is create new files containing data generated by performing cubic-splines interpolation of the input data.

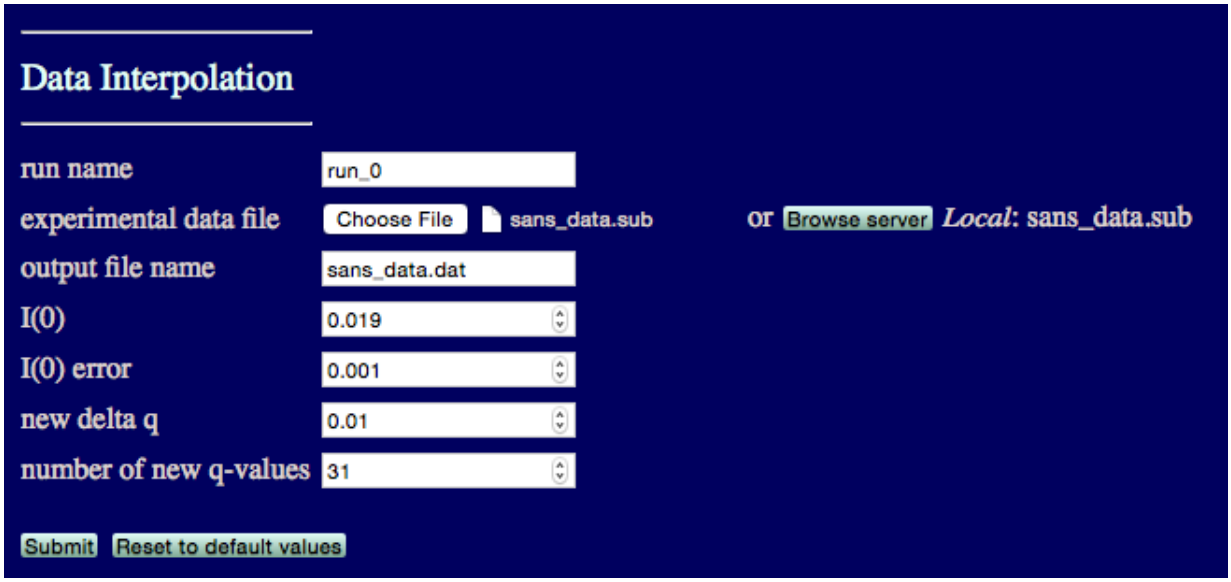
Users supply an input data file containing three columns: q, I(q), and error of I(q) at each q-value. In addition, a value of I(0) and the error of I(0) usually obtained by Gunier analysis. The user also specifies the spacing of number of q-values to be generated by the interpolation algorithm.

Notes

- Any data set with smoothly varying data can be interpolated. SANS, SAXS, X-ray or neutron reflectivity, etc. The output labels are q and I(q) by default and can not be changed in the browser view.
- Do not attempt to use the algorithm to extrapolate data (i.e. estimate data outside the range of the input data set).
- All background correction and data reduction pertaining to the input data set have to be completed before using this module.
- Two files are generated. The first is the interpolated data set given your desired number of q-values and q-range. The other is truncated at the point that the signal-to-noise ratio drops below 2.
- More data points may not be mathematically justified. For SAS there is limited information content related to the size of the molecule measured in the experimental scattering data. 15 to 31 points are generally used. See BIOISIS for a theoretical and pratical reasoning regarding the number of points one should use.
- The number of q-points, range of q, and the spacing of the q-points used to create the interpolated data files MUST match the input settings that you use in subsequent modules to calculate SAS profiles (Calculate) that are subsequently used to compare theoretical and experimental data (Chi-Square Filter).
- Users do not have to generate interpolated data using this module to input experimental data in subsequent modules. Any third-party method can be used but the number of q-points, range of q, and the spacing of the q-points MUST match what is used in subsequent modules as mentioned above.
- The final range of q will be from 0 to a maximum q-value of (number of q-values - 1) * delta q, where delta q is the spacing between adjacent q-values entered by the user.

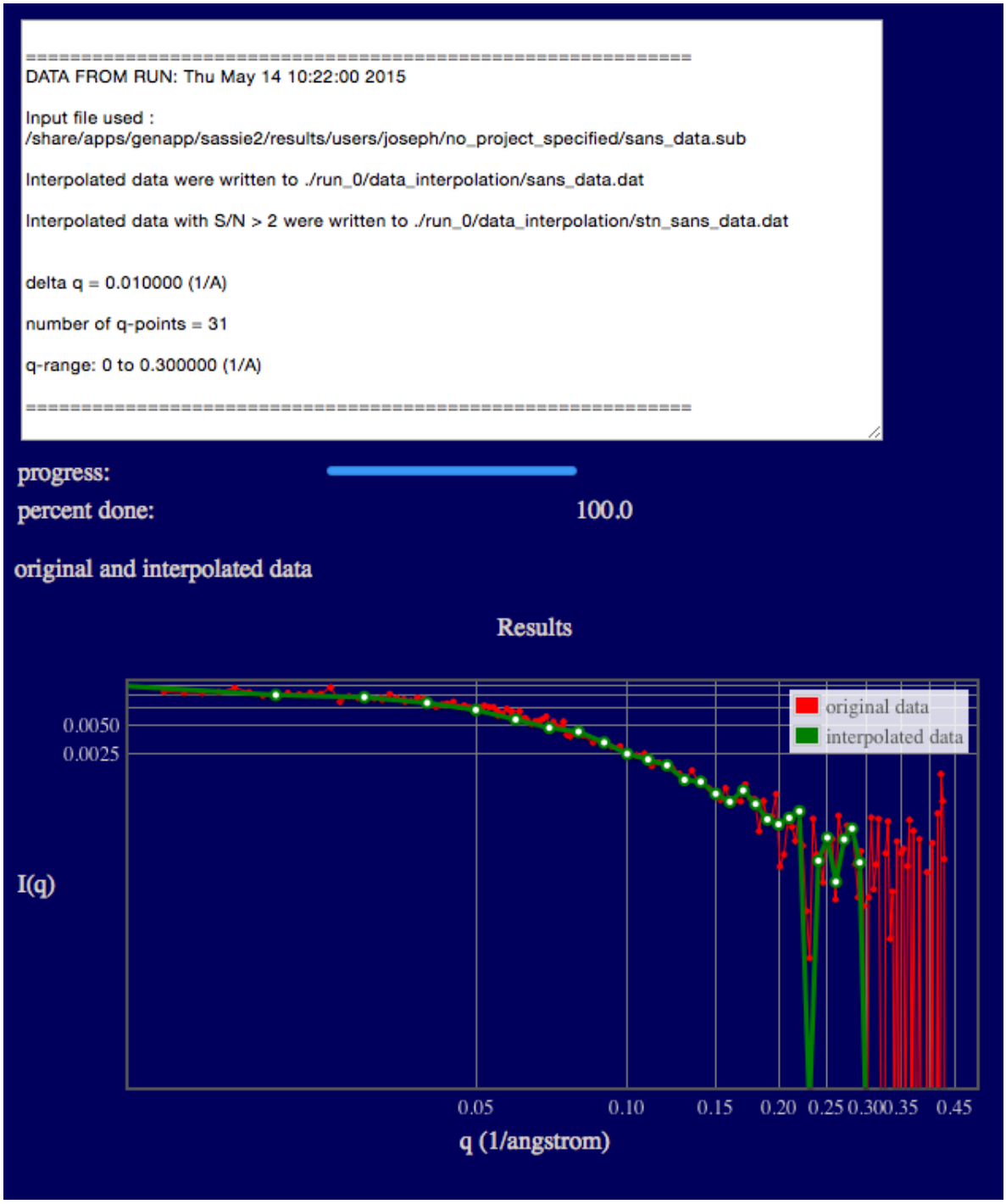
Screen Shots and Description of Input Fields

This example generates interpolated data files for an example input data file containing 171 points from q = 0.0199 to 0.4305.



- **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)
- **output file name:** Name of file that will contain the interpolated data.
- **I(0):** Experimentally determined value of scattering intensity at q = 0.
- **I(0) error:** Experimentally determined value of the error of the scattering intensity at q = 0.
- **new delta q:** Desired spacing of q-values (1/Angstrom).
- **number of new q-values:** Integer number of desired q-values.

Example Output



The output will show a plot of the original and interpolated data as well as the name of the input file used as well as the path and names of the interpolated data file.

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

Results will be written to a new directory within the given "run name". For example, in the figure it is noted that

the structures and dimensions were saved files within the current project directory within the chosen "run name" directory. In addition, the q-values are reported to be used in subsequent SAS calculator module(s): see Calculate.

```
./run_0/data_interpolation/sans_data.dat

./run_0/data_interpolation/stn_sans_data.dat
```

Visualization

None

Files Used and Created in Example

- input file

  sans_data.sub

- output files

  sans_data.dat
  stn_sans_data.dat

Limitations

The cubic-splines algorithm may give sub-standard results for certain types of data sets. Make sure to compare the interpolated results with the original data.

Reference(s) and Citations

1. Numerical Recipes: The Art of Scientific Computing (3rd ed.) W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, New York: Cambridge University Press. ISBN 978-0-521-88068-8 (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 Tools

Return to Main Documents Page

Go to top

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

Return to Main Documents Page

Chi-Square Filter

Compares calculated scattering profiles to user supplied interpolated experimental scattering profiles.

Accessibility

The Chi-Square Filter module is accessible from the Analyze section of the main menu.

Basic Usage

The purpose of the module is create compare theoretical scattering profiles to interpolated experimental data.

Users supply an input data file containing three columns: q, I(q), and error of I(q) at each q-value. In addition, a value of I(0 usually obtained by Gunier analysis.

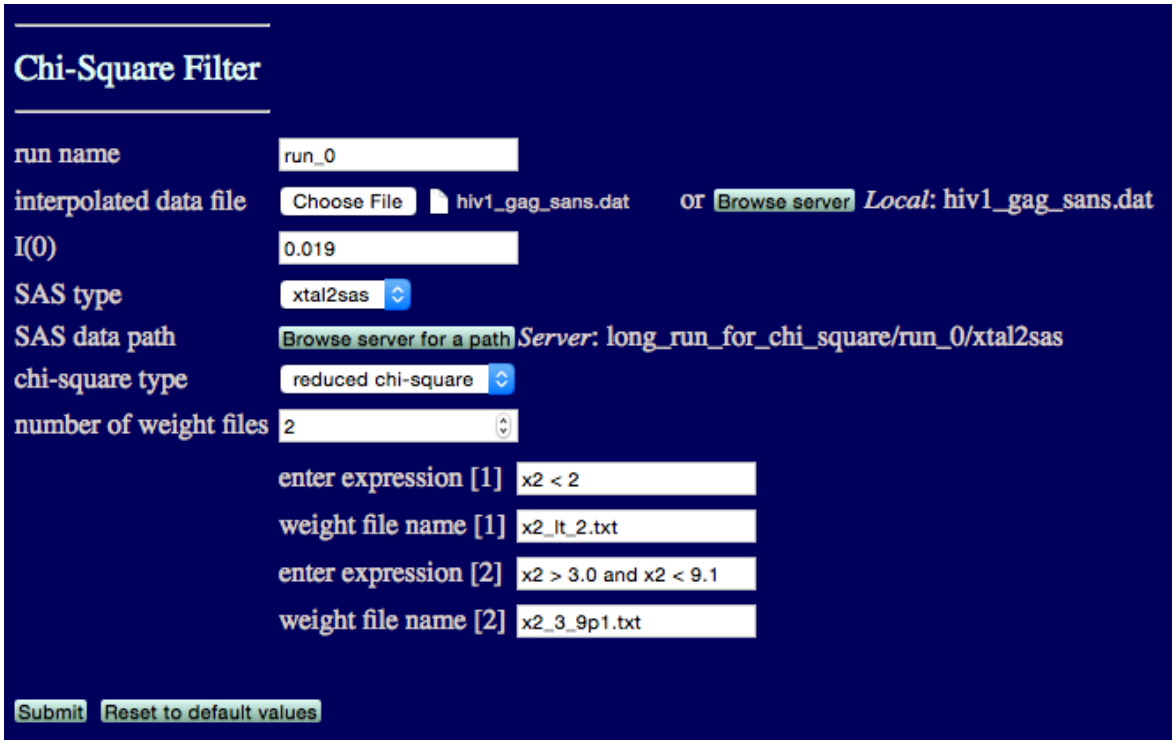
Notes

- The number of q-points, range of q, and the spacing of the q-points used to create the interpolated data files MUST match the input settings that you used in subsequent modules to calculate SAS profiles (Calculate) that are used in this module to compare theoretical and experimental data.
- Any third-party method can be used to create the input interpolated data file but the number of q-points, range of q, and the spacing of the q-points MUST match what is used in the module used to calculate the SAS profiles.
- Three mathematical options are provided to evaluate the quality of the comparison. 1) reduced chi-square, 2) chi-square, and 3) Pearson chi-square.

OPEN put PNGs of equations here

Screen Shots and Description of Input Fields

This example compares a large set of scattering profiles to interpolated experimental data for the hiv1_gag protein.



- **run name:** user defined name of folder that will contain the results.
- **interpolated data file:** Name of input file with interpolated data with at least three columns: q, I(q), and error in I(q)
- **I(0):** Experimentally determined value of scattering intensity at q = 0.
- **SAS type** Select the name of the SAS calculator that was used to generate the SAS profiles to compare to experimental data.
- **SAS data path** Select the directory that contains the SAS profiles.
- **chi-square type:** Desired spacing of q-values (1/Angstrom).
- **number of weight files** Enter an integer number to indicate a number of weight files you wish to generate. Weight files are useful to extract (or merge) portions of trajectories and SAS files.
    - **enter expression** Simple expression indicating the filter to indicate conditions you wish to satsify. Only keywords "rg" and "x2" are allowed.
    - **weight file name** Name of text file containing weights.

The expression syntax is simple, for example,

```
x2 < 3.3
```

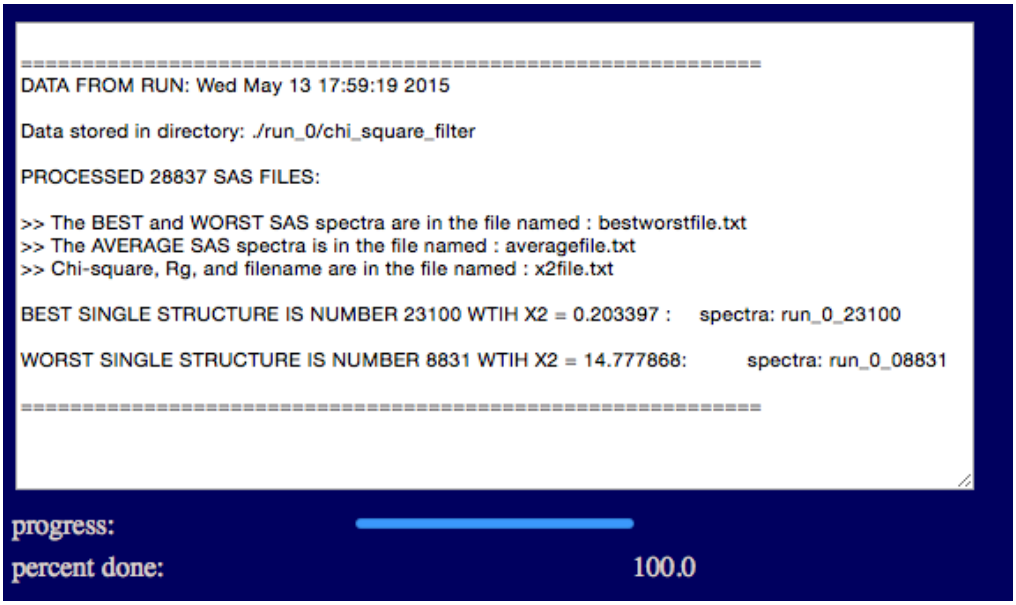
will set all frames that have a chi-square value less than 3.3 to 1.0 in the weight file, all other frames will be set to zero. You can link expressions together, and it is best to do so using parentheses to prevent errors,

```
        (rg > 20.2) and (x2 < 30)
```

will set frames with Rg values greater than 20.2 and less than 30 equal to 1.0, otherwise they are set to zero.

One can not simultaneously select for both "rg" and "x2" in the same expression.

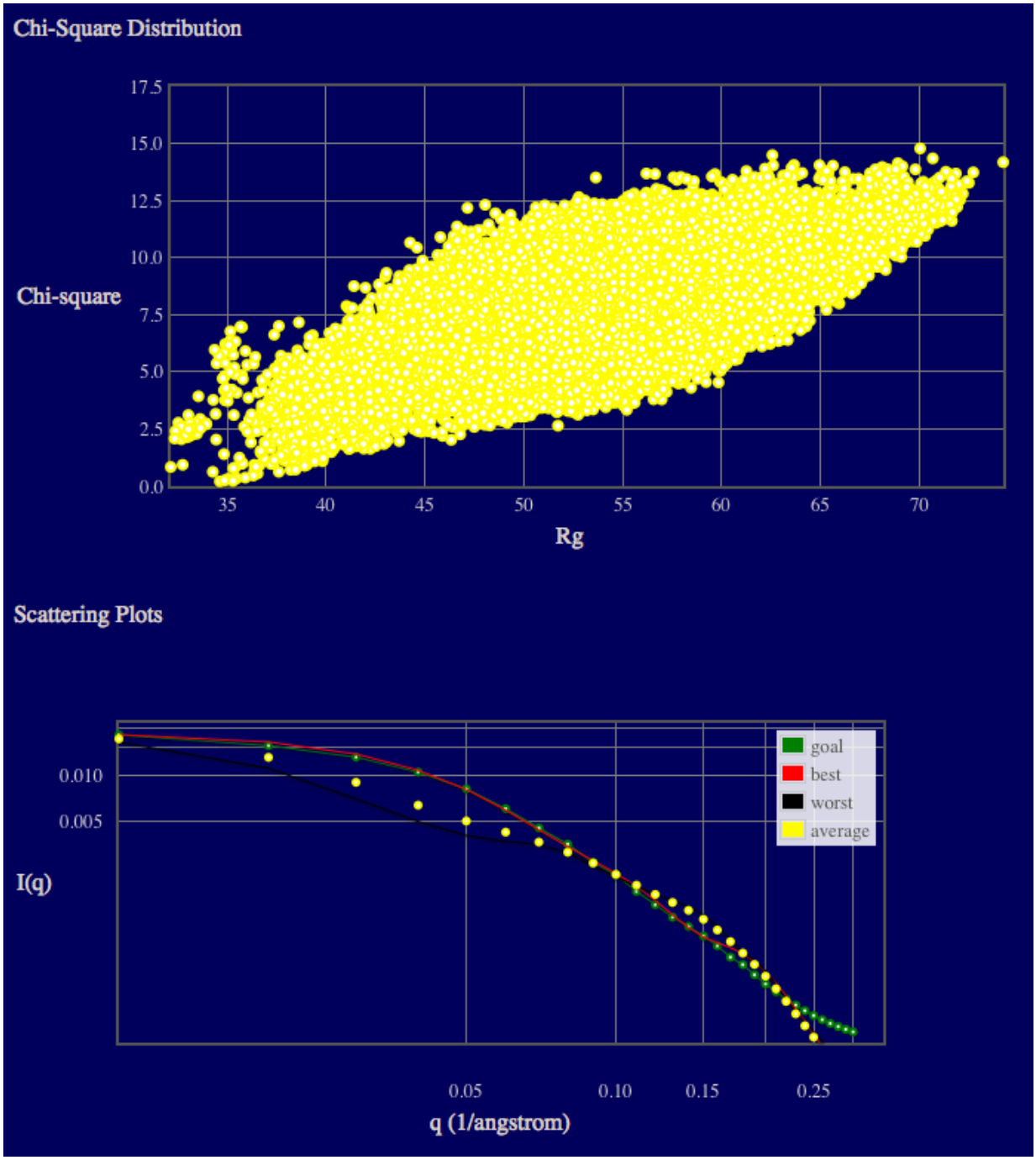
Example Output



The output will show a plot of the interpolated data and scattering profiles from the best, worst, and average structures in the ensemble of scattering profiles.

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

Results will be written to a new directory within the given "run name". For example, in the figure it is noted that the structures and dimensions were saved files within the current project directory within the chosen "run name" directory.




```
./run_0/chi_square_filter/bestworstfile.txt
./run_0/chi_square_filter/averagefile.txt
./run_0/chi_square_filter/x2file.txt
```

Files Used and Created in Example

- input file

  hiv1_gag.dat

- output files

  sas_spectra_plot.txt
  x2_vs_rg_plot.txt x2file.txt

  averagefile.txt
  bestworstfile.txt

  x2_lt_2.txt
  x2_3_9p1.txt

Visualization

None

Limitations

Simple comparison is done between theoretical and experimental data.

Reference(s) and Citations

1. Numerical Recipes: The Art of Scientific Computing (3rd ed.) W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, New York: Cambridge University Press. ISBN 978-0-521-88068-8 (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 Analyze

Return to Main Documents Page

Go to top

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

Return to Main Documents Page

Density Plot

Calculates volumetric gaussian cube files from user supplied structures. These files can be subsequently rendered in VMD and Pymol.

Accessibility

The Density Plot module is accessible from the Analyze section of the main menu.

Basic Usage

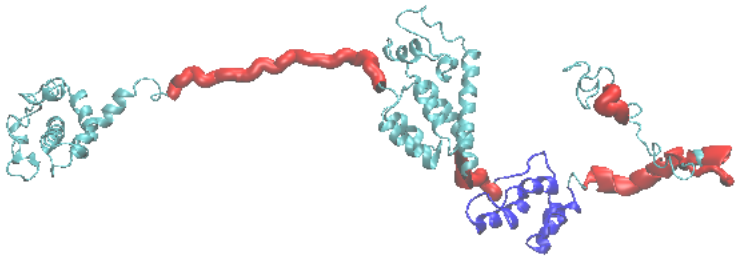
The purpose of the module is generate files with volumetric data to visualize results. Often, this is used to visualize sub-ensembles of structures that are in agreement with experimental data.

Notes

- Typically, between 10,000 to 50,000 structures are required to enable a clear visualization consistent with adequate configuration space for most problems.
- For a molecule with a single segment, the files with "complete" in the file name are identical. The files named with "region" in the name are for the regions identified by the user. For molecules with multiple segments, each will have a "complete" file with all the information for each region in that segment. Then the single, overall, "complete" file (without the a segment name) will have data from all segments with all regions.
- The dimensions of the cube are automatically re-sized, if a structure doesn't fit within the input dimensions. This is done so that each structure in your input trajectory will be inside the final cube.
- All cube files are normalized to account for the number of atoms. There is an option to save the raw occupancy data if users wish to process the data in a different manner.
- Output file naming is discussed in the first example output section below.

Screen Shots and Description of Input Fields

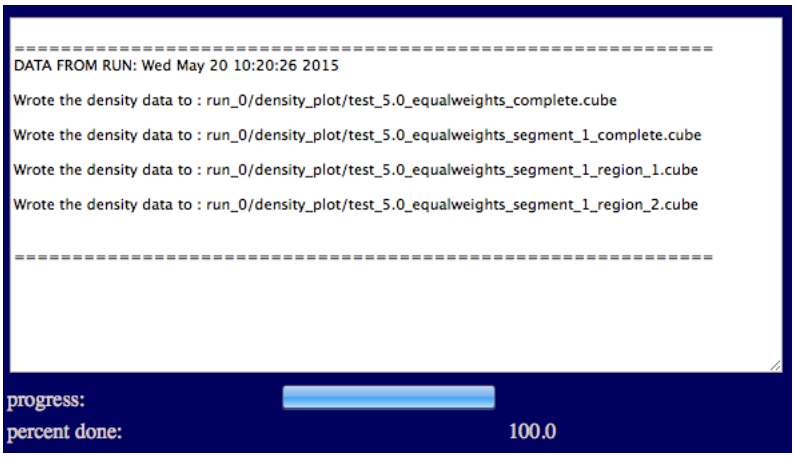
This example creates gaussian cube files sample configurations of the HIV-1 Gag protein generated in the Monomer Monte Carlo "constraints" example. The cartoon of the starting structure highlights the flexible regions (red) and structure alignment region (blue) used in the simulation.



Note that the constraints used in the example allowed the acceptance of trial configurations so that only structures with the center of mass of atoms in residues 240 to 260 were within 40.0 angstroms of the center of mass of CA atoms in residues 400 to 420 were valid.

- **run name** user defined name of folder that will contain the results.
- **reference pdb** PDB file with naming information and coordinates of the starting structure.
- **input filename (dcd or pdb):** DCD or PDB file with coordinates that will be used to create gaussian cube files.
- **output file name prefix** Unique text to insert in output file names.
- **x,y,z-length(s) (angstroms)** Estimate of the size of overall cube to contain your molecule. Note, box will re-size to fit all structures in your trajectory if the input values are too small.
- **grid spacing (angstroms)** Cubic length of each voxel.
- **save occupancy information** Choose to keep non-normalize occupancy values for each voxel.
- **number of segments** An integer value indicating the number of segments in molecule.
    - **segment name** Name of particular segment.
    - **number of ranges** An integer value indicating the number of regions to create independent gaussian cube files.
    - **residue region(s)** Residue numbers defining each region in segment. The number of pairs should match the number of regions for the given segment. Pairs of integers separated by hypens with each pair separated by commas.
    - **segment basis** Name of the atom to use to determine if a voxel is occupied.
- **check box to use a weight file**
    - **weights file name** Name of file with weights to apply to structures to include or exclude from the gaussian cube file.

Example Output



The output will indicate the list of files that were written to disk.

Results are written to a new directory within the given "run name" as noted in the output.

Several files are generated and saved to the "run name" density_plot directory.

```
run_0/density_plot/test_5.0_equalweights_complete.cube

run_0/density_plot/test_5.0_equalweights_segment_1_complete.cube

run_0/density_plot/test_5.0_equalweights_segment_1_region_1.cube

run_0/density_plot/test_5.0_equalweights_segment_1_region_2.cube
```

Output file naming

The general recipe is:

"output file name prefix_"grid spacing"_"weight file flag"_segment_"segment number"_region_"region number".cube

In this example:

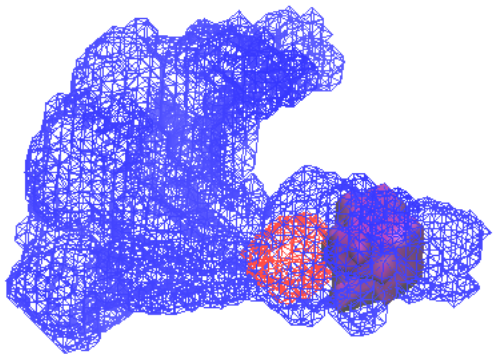
output file name prefix : test
grid spacing: 5.0
equal weight file not used:equal_weights
1 segment in molecule: 1
2 regions in segment: region_1 region_2

Region_1 : 240-260
Region_2 : 400-420

This allows one to experiment with different conditions and keep similar files organized.

Visualization

In the figure below the envelope sampled by all accepted structures is shown along with smaller envelopes described below. The envelopes were visualized using VMD. Three envelopes are displayed in the figure. The blue mesh represents all atoms for each frame of the trajectory, the red mesh represents only CA atoms in region 1 (resid 240-260), and the solid purple mesh represents only CA atoms in regions 2 (resid 400-420).



A step by step guide to create this figure using VMD is found in the VMD visualization of gaussian cube files: Example 1 page.

Files Used and Created in Example

- input files

  hiv1_gag.pdb
  hiv1_gag_monte_carlo_constraints.dcd

- output files

  test_5.0_equalweights_complete.cube
  test_5.0_equalweights_segment_1_complete.cube
  test_5.0_equalweights_segment_1_region_1.cube
  test_5.0_equalweights_segment_1_region_2.cube

Limitations

Module creates output files that are visualized using other program(s) as there are no visualization options built into sassie-web.

Reference(s) and Citations

1. VMD - Visual Molecular Dynamics W. Humphrey, A. Dalke, K. J. Schulten, Molec. Graphics, 14, 33-38 (1996). 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 Analyze

Return to Main Documents Page

Go to top

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

Return to Main Documents Page

Return to Density Plot

VMD visualization of gaussian cube files: Example 1

This page covers the steps used to visualize the gaussian cube files created in example 1 in Density Plot. This example uses VMD 1.9.2 but should be nearly the same for older versions.

It is useful to carry out this example to first familiarize yourself with VMD.

Download example files:

The following links allow you to download the files uesed and created in the example. Note that these files can be downloaded in-line with the descriptions below.

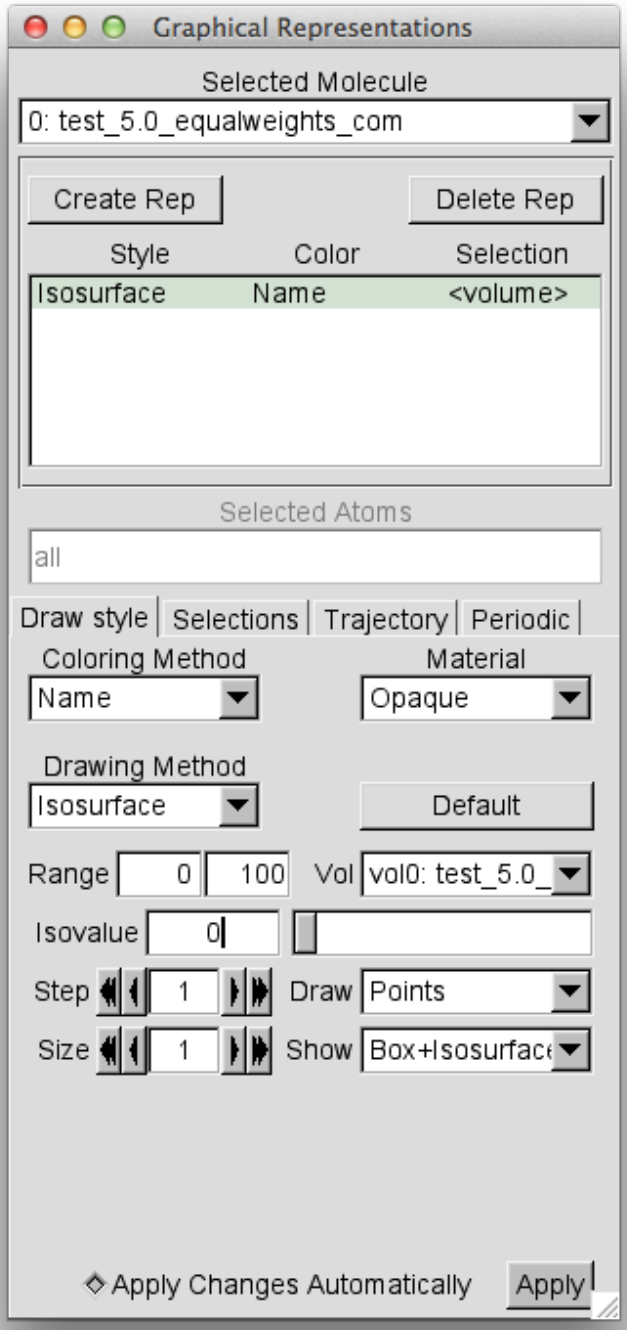
hiv1_gag.pdb

test_5.0_equalweights_complete.cube
test_5.0_equalweights_segment_1_region_1.cube
test_5.0_equalweights_segment_1_region_2.cube

Open VMD program and load "complete" cube file

From **"File"** chose **"New Molecule"** and load the test_5.0_equalweights_complete.cube file.

After loading, open the **"Graphical Reprentations"** widget from **"Reprentations"** selection in the **"Graphics"** pull-down tab. Then, for the **"Drawing Method"** option choose **"Isosurface"**. This should give you the following window.



Note that the default settings are to **Draw "Points"** and **Show "Box+Isosurface"**. Change these settings

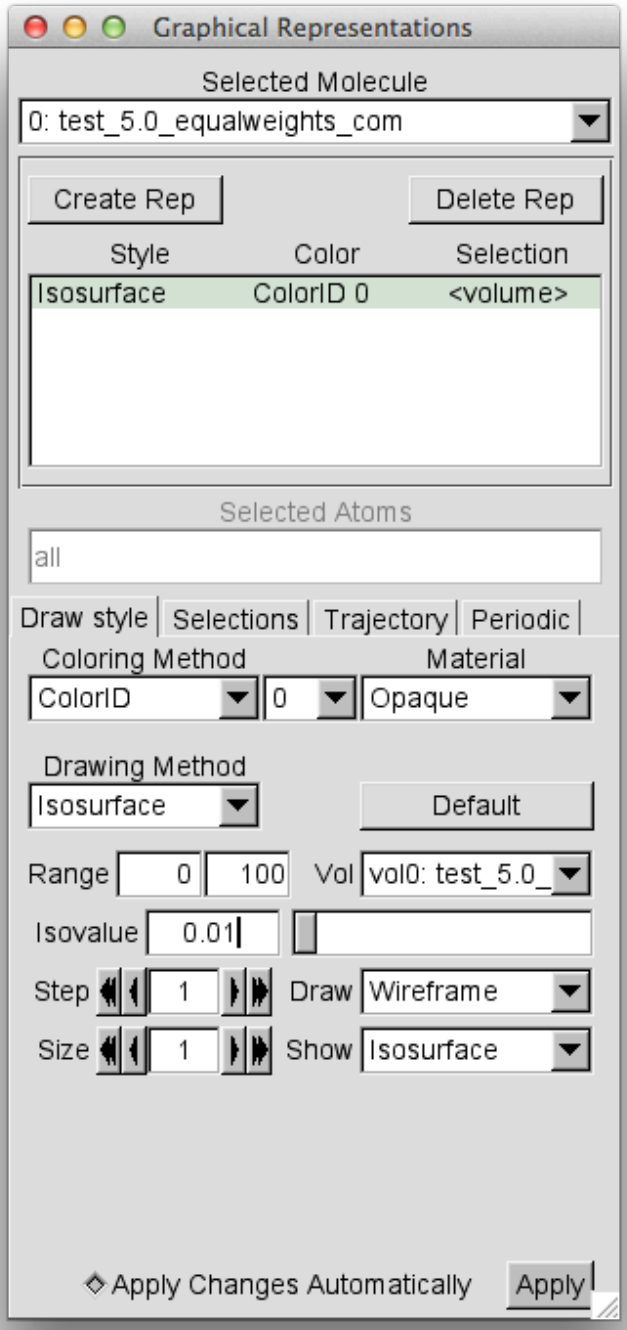
to **"Wireframe"** and **"Isosurface"** respectively.

Then change the **"Isovalue"** from the default **0** to **0.01** and press **return**. This is a reasonalbly low value to capture the complete volumetric data set.

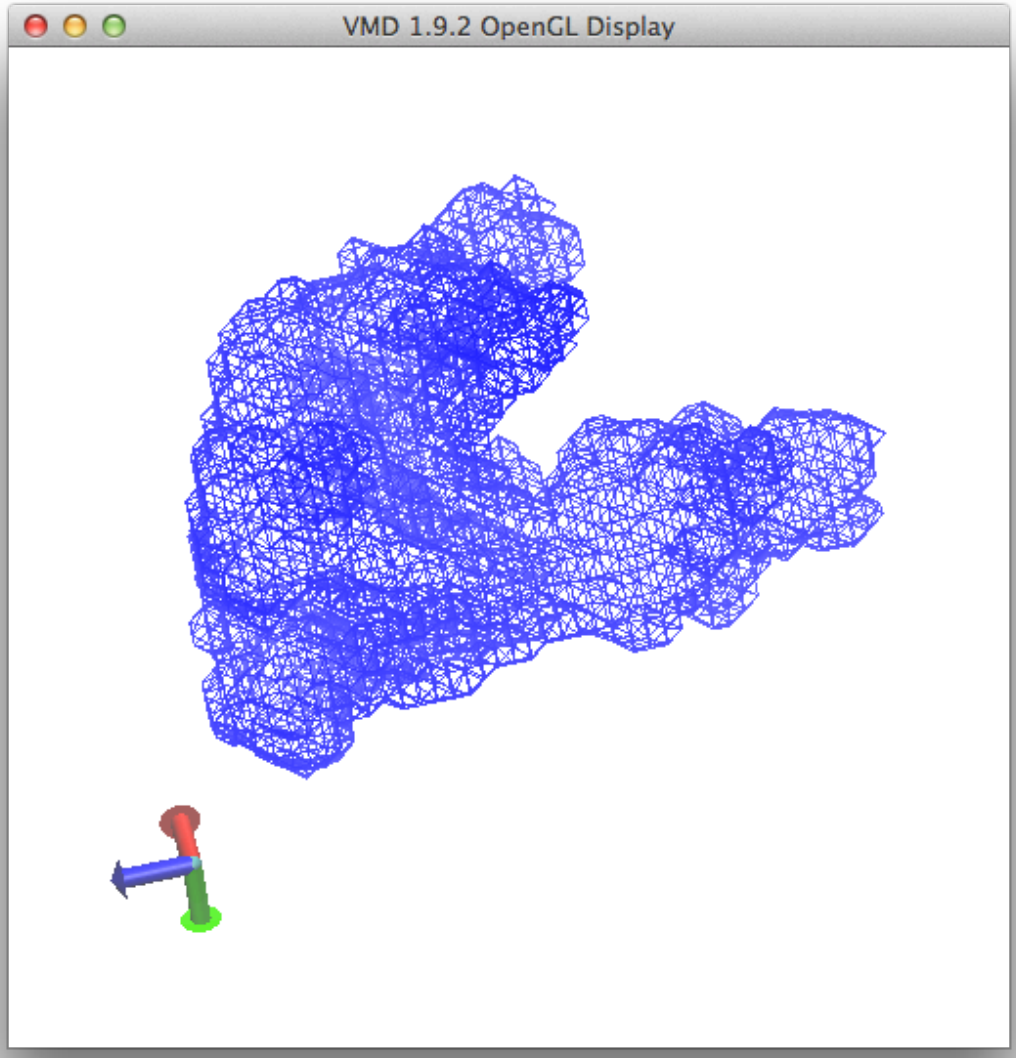
Note: When loading multiple gaussian cube files, as shown below, make sure that the **"Isovalue"** setting is the same for all files, otherwise the visulaization may be misleading.

Then change the **"Coloring Method"** to **"Color ID"** and select **0** as the option.

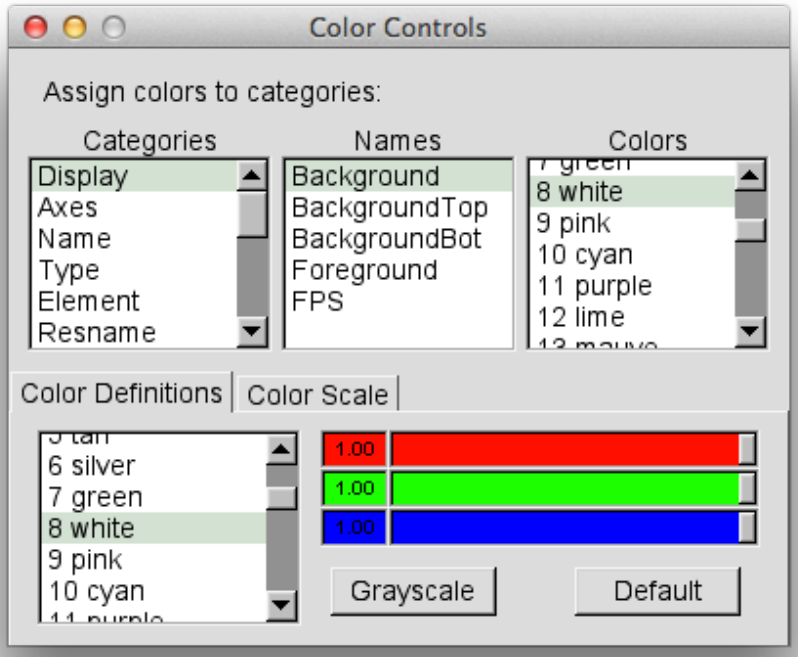
Your **"Graphical Representations"** window should look similar to this.



And the plot, once re-oriented for visual purposes, is shown below.



Your version will probably have a black background. To switch the color of the background to white, choose the **"Colors"** option from the **"Graphics"** pull-down tab. Select **"Display"** in the **"Catagories"** section, **"Background"** in the **"Names"** section, and **"8 white"** in the **"Colors"** section, as shown in the picture below.



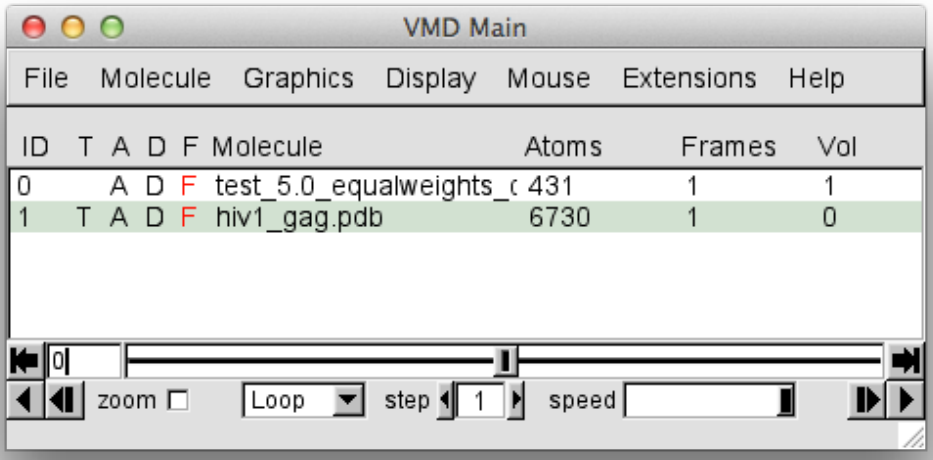
At this point you are displaying the region of three-dimensional space for all structures that were accepted from the constrained Monomer Monte Carlo example.

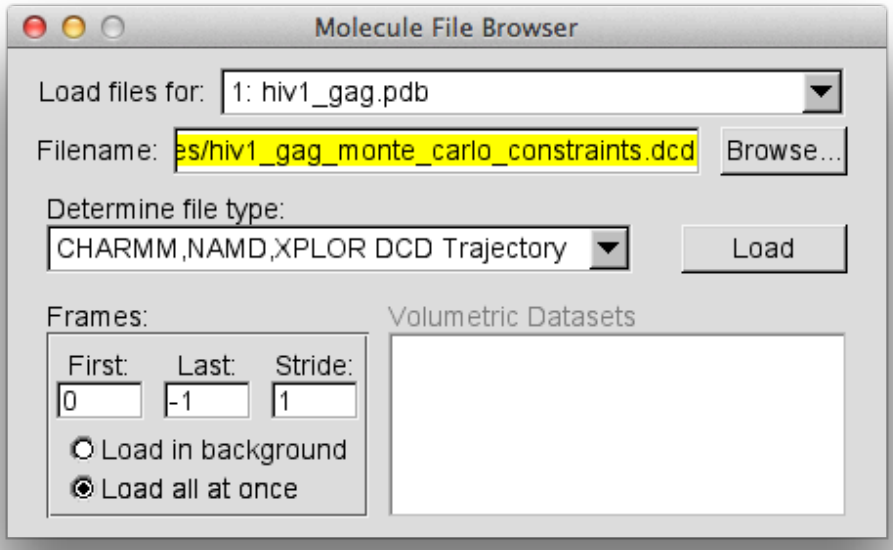
Load original PDB and DCD trajectory

In this section the original structure along with the accepted structures from the simulation will be loaded into a new molecule in the same VMD session described above.

From **"File"** chose **"New Molecule"** and load the hiv1_gag.pdb file. Then, in the same **"Molecule File Browser"** load the trajectory from hiv1_gag_monte_carlo_constraints.dcd. Your **"VMD Main Menu"** and **"Molecule File Browser"** should look like the screen shots below.

NOTE It is faster to select **"Load all at once"** option in the **"Molecule File Browser"** when loading a large trajectory. This sidesteps the animation of each frame as the trajectory loads.

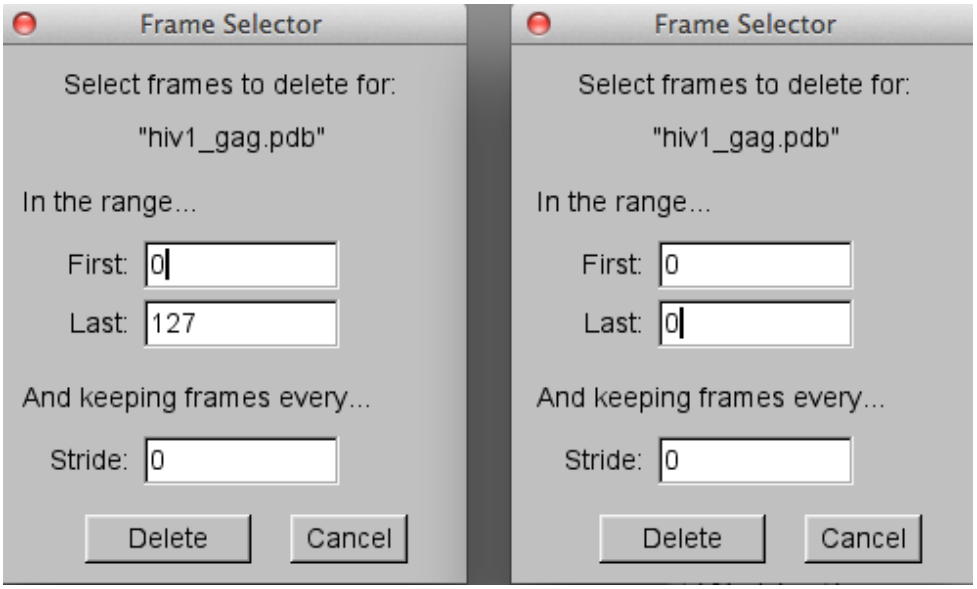




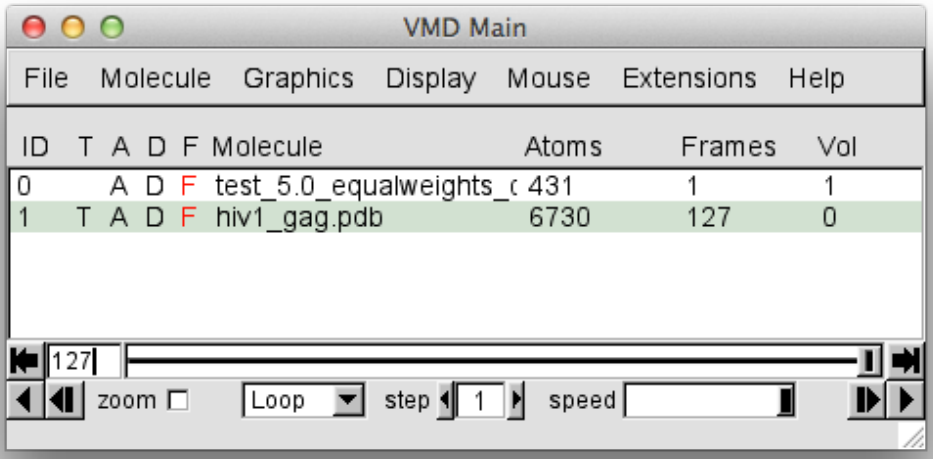
The trajectory contains 127 frames. The **"hiv1_gag.pdb"** entry in the **"Molecule File Browser"** will indicate that there are 128 frames loaded. This is because the PDB of the starting structure was loaded before the trajectory. That original structure may or may not have been accepted. So for completeness, this single frame should be removed.

NOTE VMD counts frames starting at 0, not 1.

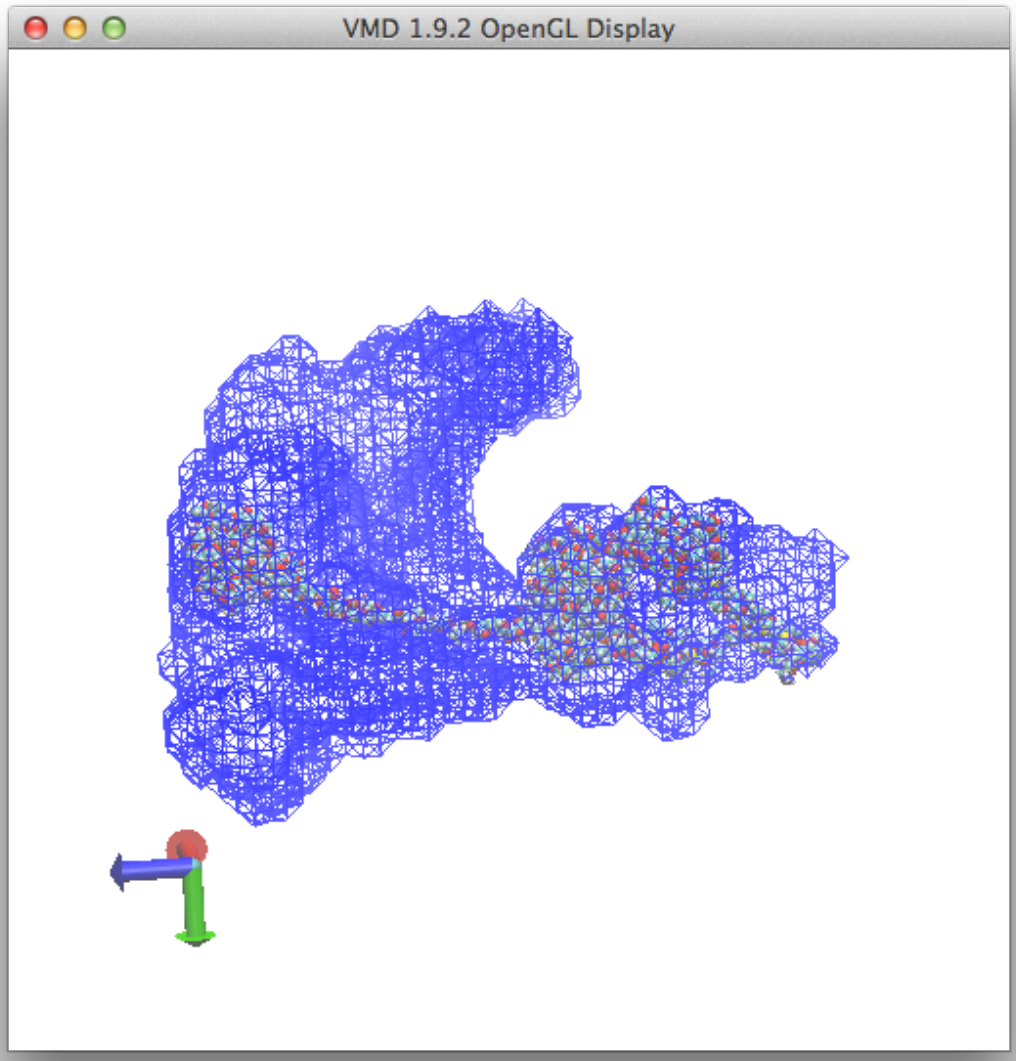
To remove frame 0, the coordinates from the starting PDB file, select the **"Delete Frames..." option from the** "Molecule" **pull-down tab. Then in the** "Frame Selector" **pop-up windown change the value of the** "Last" **frame from** 127 **to** 0** as shown on the right side of the image below.



Then hit the **"Delete"** button to complete the task. The **"VMD Main Menu" should now indicate that only 127 frames are in the** "hiv1_gag.pdb"** molecule.



In the Open GL window, it may be difficult to see the protein structure. In the **"Graphical Reprentations"** window, select **"VDW"** in the **"Drawing Method"** option box. This will give you something similar to the view below.



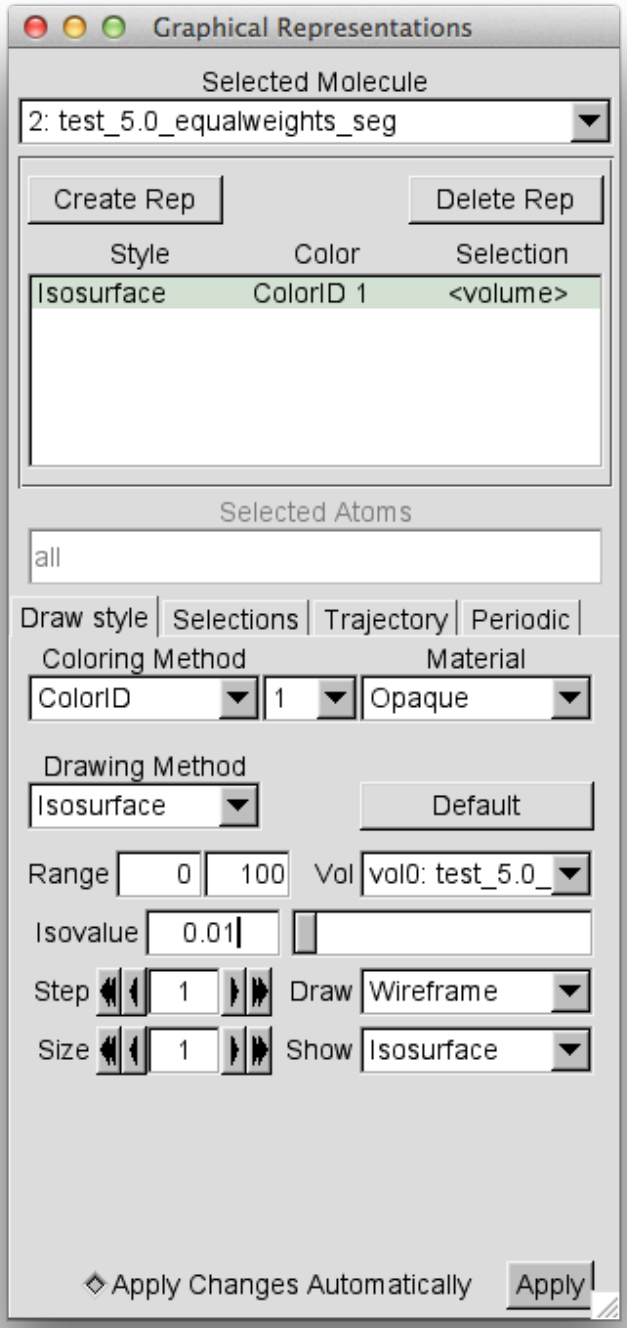
Load gaussian cube file for range 1

From **"File"** chose **"New Molecule"** and load the test_5.0_equalweights_segment_1_region_1.cube file. As described above, then select the **"Graphical Reprentations"** widget from **"Reprentations"** selection in the **"Graphics"** pull-down tab. Then, for the **"Drawing Method"** option choose **"Isosurface"**. Then

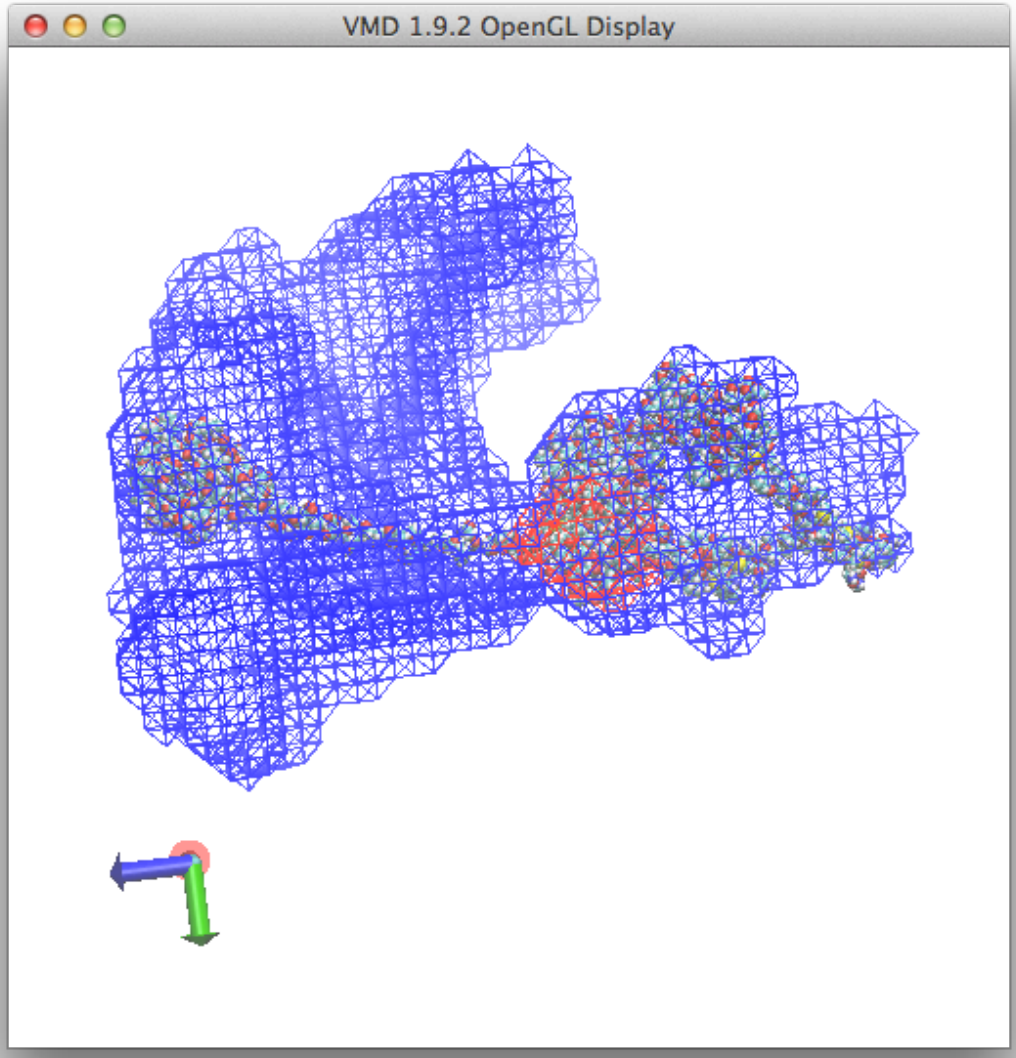
change the **"Isovalue"** from the default **0** to **0.01** and press **return**.

Then change the **"Coloring Method"** to **"Color ID"** and select **1** as the option.

This should give you the following window.



and a view in the OpenGL display similar to this

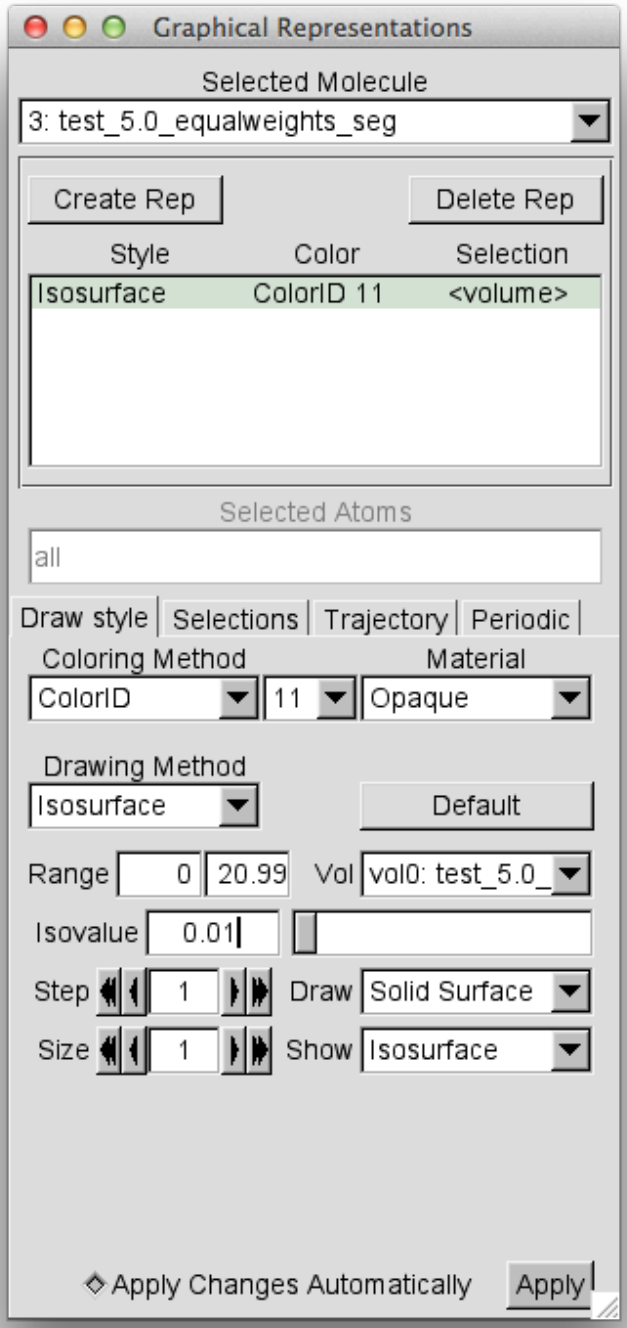


Load gaussian cube file for range 2

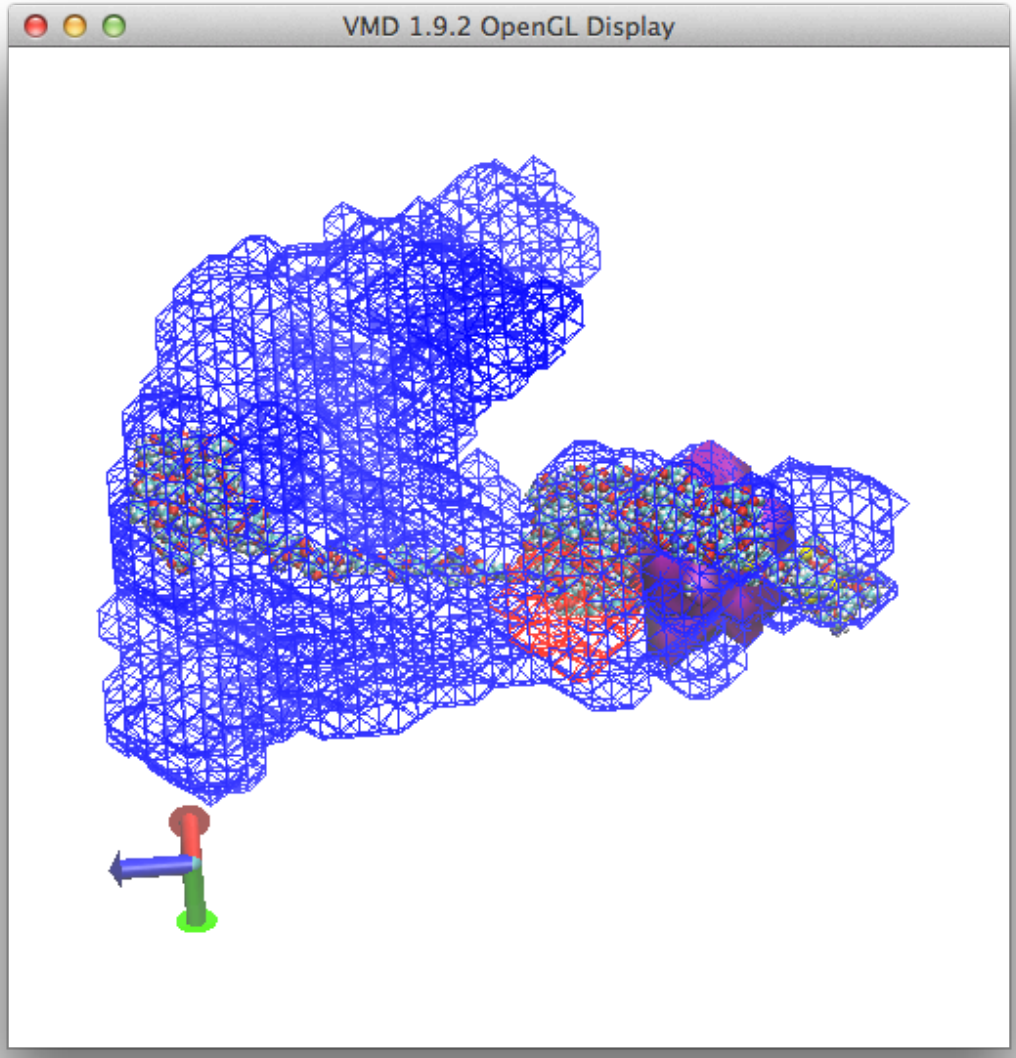
From **"File"** chose **"New Molecule"** and load the test_5.0_equalweights_segment_1_region_2.cube file. As described above, then select the **"Graphical Reprentations"** widget from **"Reprentations"** selection in the **"Graphics"** pull-down tab. Then, for the **"Drawing Method"** option choose **"Isosurface"**. Then change the **"Isovalue"** from the default **0** to **0.01** and press **return**.

Then change the **"Coloring Method"** to **"Color ID"** and select **11** as the option.

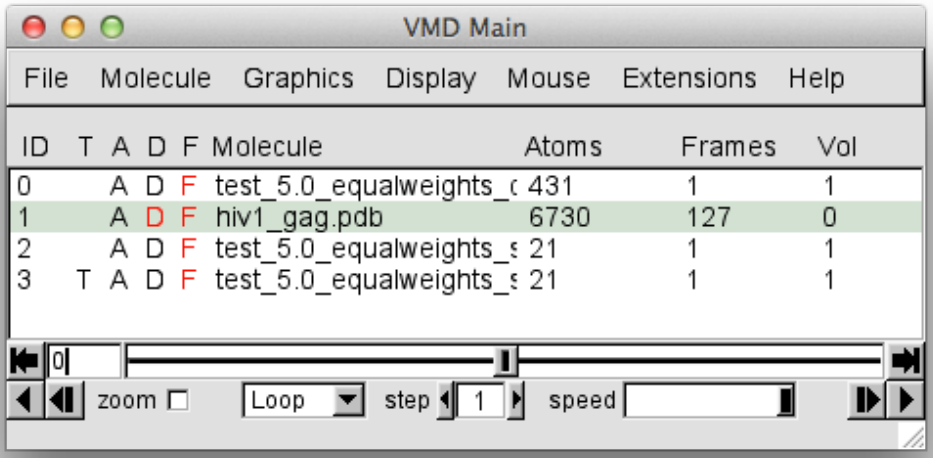
This should give you the following window.



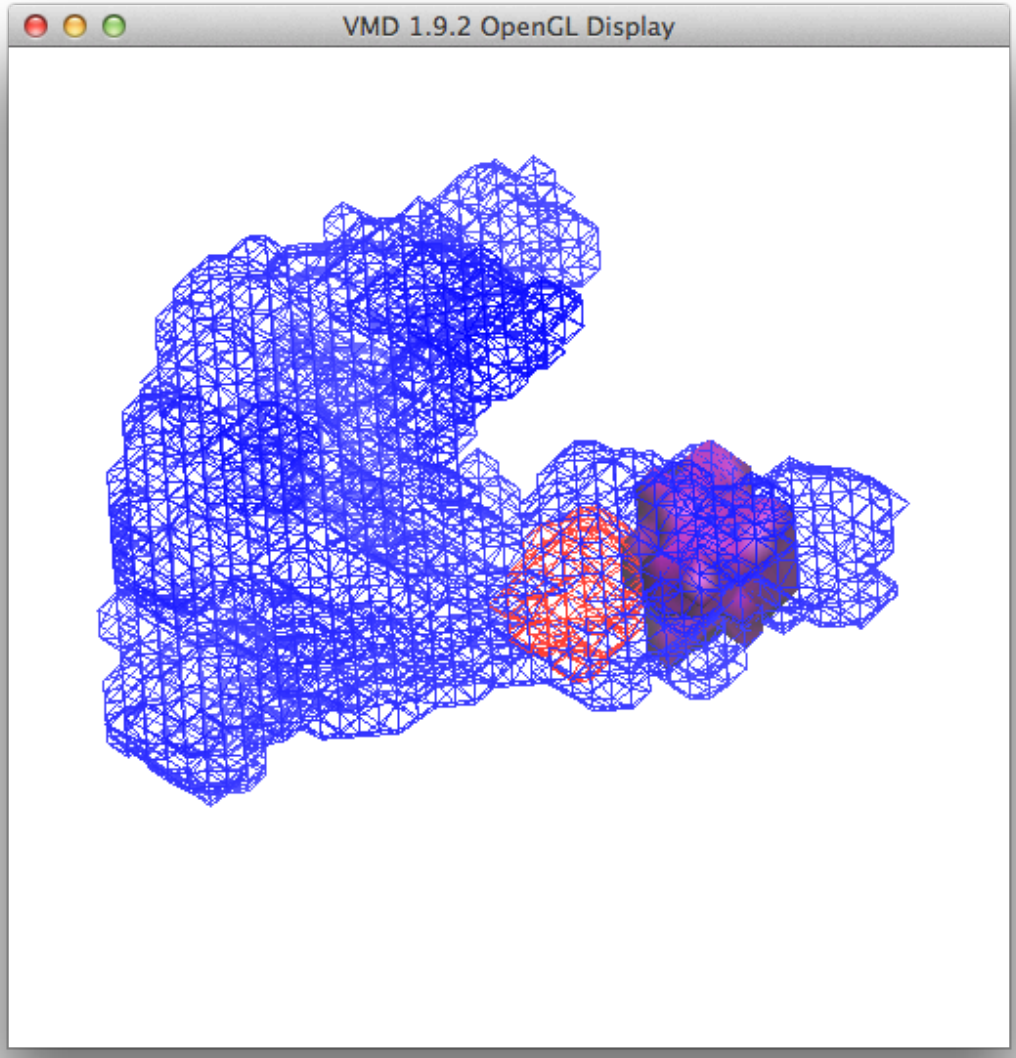
and a view in the OpenGL display similar to this



To obtain a final image similar to that shown in Density Plot example 1, merely toggle the **"D"** or **display** option for the **"hiv1_gag.pdb"** molecule as shown below.



To remove the axes label select **"Axes" "off"** from the **"Display"** pull-down tab. This will result in the final image.



Notes

- Typically, between 10,000 to 50,000 structures are required to enable a clear visualization consistent with adequate configuration space for most problems.

Reference(s) and Citations

1. VMD - Visual Molecular Dynamics W. Humphrey, A. Dalke, K. J. Schulten, Molec. Graphics, 14, 33-38 (1996). BIBTeX, EndNote, Plain Text

---

Return to Density Plot

Return to Analyze

Return to Main Documents Page

Go to top

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