Return to Main Documents Page

Monomer Monte Carlo

Performs molecular Monte Carlo simulation of single chain protein or single chain nucleic acids.

Accessibility

The Monomer Monte Carlo module is accessible from the Simulate section of the main menu.

Basic Usage

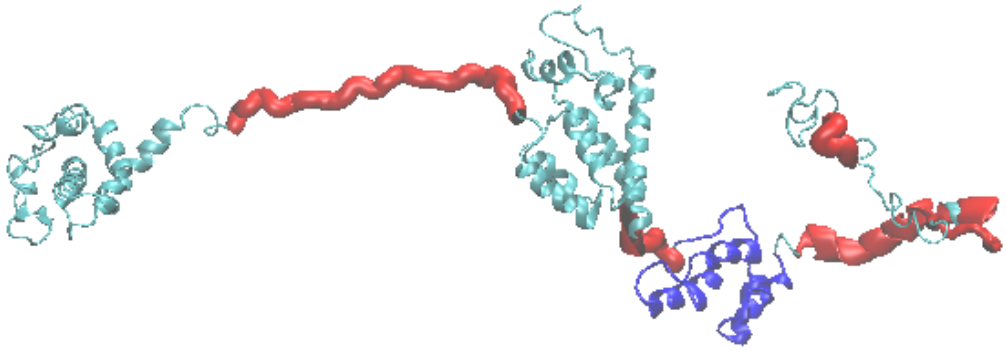
The purpose of the module is perform a molecular simulation of an input single chain protein or single chain nucleic acid by sampling backbone torsion angles.

Notes

- The starting structure must be a complete structure without missing residues. Atom and residue naming must be compatiable with those defined in the CHARMM force field See Notes on Starting Structures and Force Fields and PDB Scan for further details.
- Only single chain protein or single chain nucleic acids are supported. Systems with multiple chains can be modeled using Complex Monte Carlo.
- The output file format is DCD since in most cases many structures are generated. There is no option to save the output files in PDB format. One can use Extract Utilities to convert DCD files to multi-frame PDB files.
- Structures are generated by Markov Monte Carlo sampling of backbone torsion angles. Energetics of torsion angles are determined using CHARMM force field parameters.
- Typically, between 10,000 to 50,000 structures are required to sample adequate configuration space for most problems.
- Parameters are supplied to help guide the Monte Carlo sampling such as temperature, control of single move angle sampling per region, and directed Monte Carlo options to guide the radius of gyration (Rg) to a user supplied value.
- A utility is provided to overlap accepted structures onto a single reference frame. This is useful to visualize relative configuration coverage in an ensemble.
- Several options are offered to check for atomic overlap: heavy atoms, all, backbone, and atom name. If one chooses the atom name option, then the user will be prompted to supply an atom name that should exist in all residues and a overlap distance cutoff value. Other options set the cutoff distance automatically.
- In Advanced Input options are provided to reject structures based on Rg value, position of atoms in the Z-direction and via atomic constraints provided as a list in a text file as described in Constraints. These options are not mutually exclusive and can be used in the same run as needed.
- Typical workflows involve generating an ensemble of structures using this module, then energy minimizing the ensemble using Energy Minimization, then calculating scattering from the ensemble using modules in Tools, and finaly comparing results to experimental data using modules in Analyze.
- In many situations, multiple runs need to be carried out to find structures that cover configuration space and have scattering profiles that are in agreement with experimental data. One can use Merge Utilties to combine both the structures (DCD files) and SAS profiles into a single new DCD and single directory will correctly numbered SAS profiles.
- To simulate long random coil regions, usually at the ends of globular proteins, it is often neccessary to sub-sample accepted structures as adjacent structures can be correlated. To obtain adequate power-law scaling, one can sub-sample a trajectory using Extract Utitilies using the periodic sampling option.

Screen Shots and Description of Input Fields

This example generates a series of structures to sample configurations of the HIV-1 Gag protein. The cartoon of the starting structure highlights the flexible regions (red) and structure alignment region (blue).



Monomer Monte Carlo

run name	run_0
reference pdb	Choose File hiv1_gag.pdb or Browse server *Local*: hiv1_gag.pdb
output file name (dcd)	hiv1_gag_monte_carlo.d
number of trial attempts	10000
return to previous structure	20
temperature (K)	300.0
molecule type	protein
number of flexible regions to vary	5
maximum angle sampled for each region	30.0,30.0,30.0,30.0,30.0
residue range for each flexible region	123-144,277-282,354-374,378-389,408-412
structure alignment: low residue	284
structure alignment: high residue	350
overlap basis	heavy atoms

Advanced Input

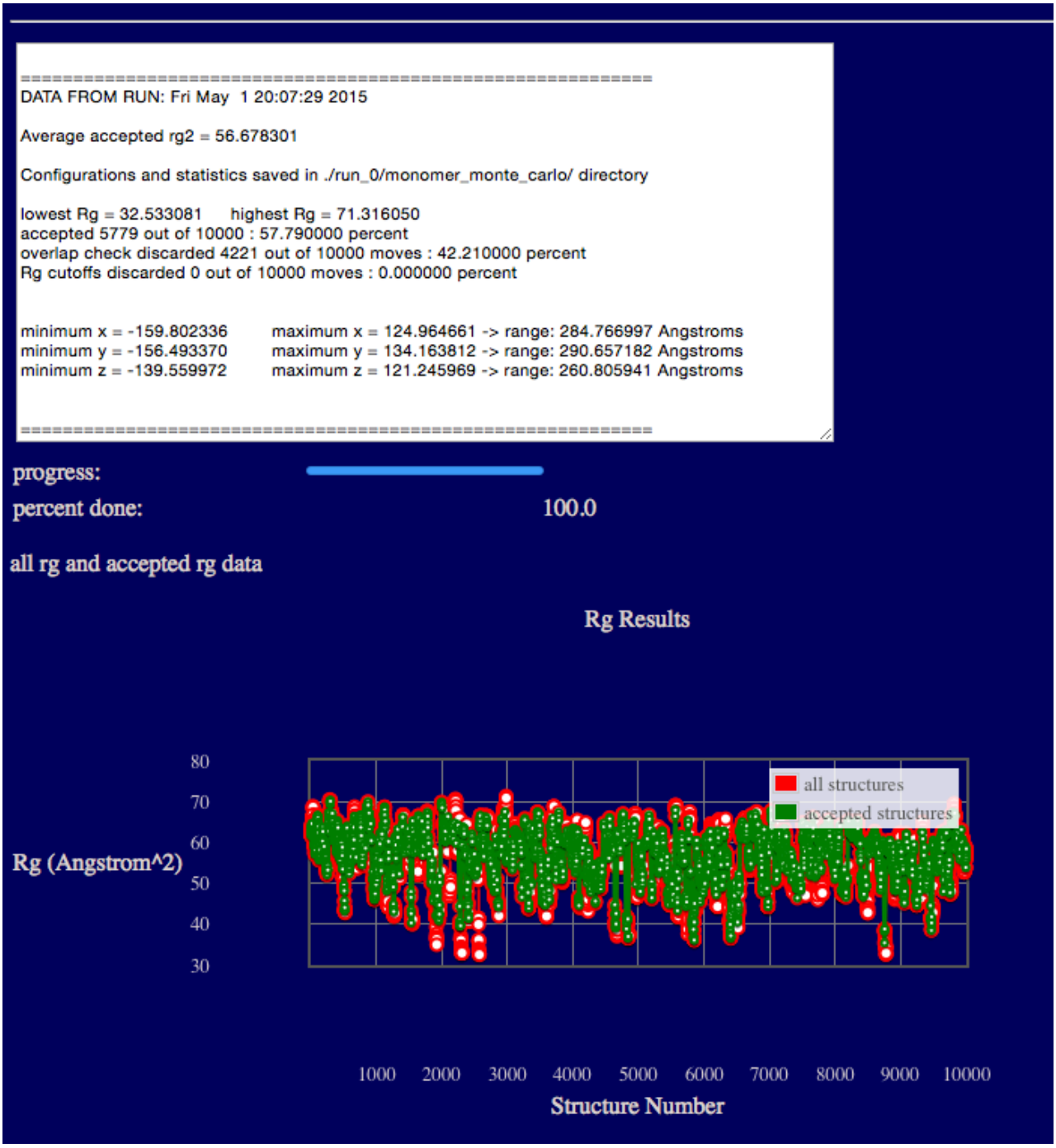
Check Box for Advanced Input

Submit Reset to default values

- **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.
- **output file name** Name of ouput DCD file containing accepted structures from the simulation.
- **number of trial attempts** Number of Monte Carlo moves to attempt.
- **return to previous structure** After this number of Monte Carlo moves fails to find an accepted configuration, re-load a previously accepted structure.
- **temperature (K)** Simulation temperature.
- **molecule type** Select either protein or RNA.
- **number of flexible regions to vary** An integer value indicating the number of regions to sample backbone torsions.

- **maximum angle sampled for each region** Angle, in degrees, that can be sampled in a single move for each region.
- **residue range for each flexible region** Residue numbers defining each flexible region.
- **structure alignment: low residue** Residue to define the beginning of region used to align all structures.
- **structure alignment: high residue** Residue to define the end of region used to align all structures.
- **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.

Example Output



The output will indicate various Rg values from the ensemble, acceptance and overlap statistics, and dimensions of the accepted structures in the final ensemble.

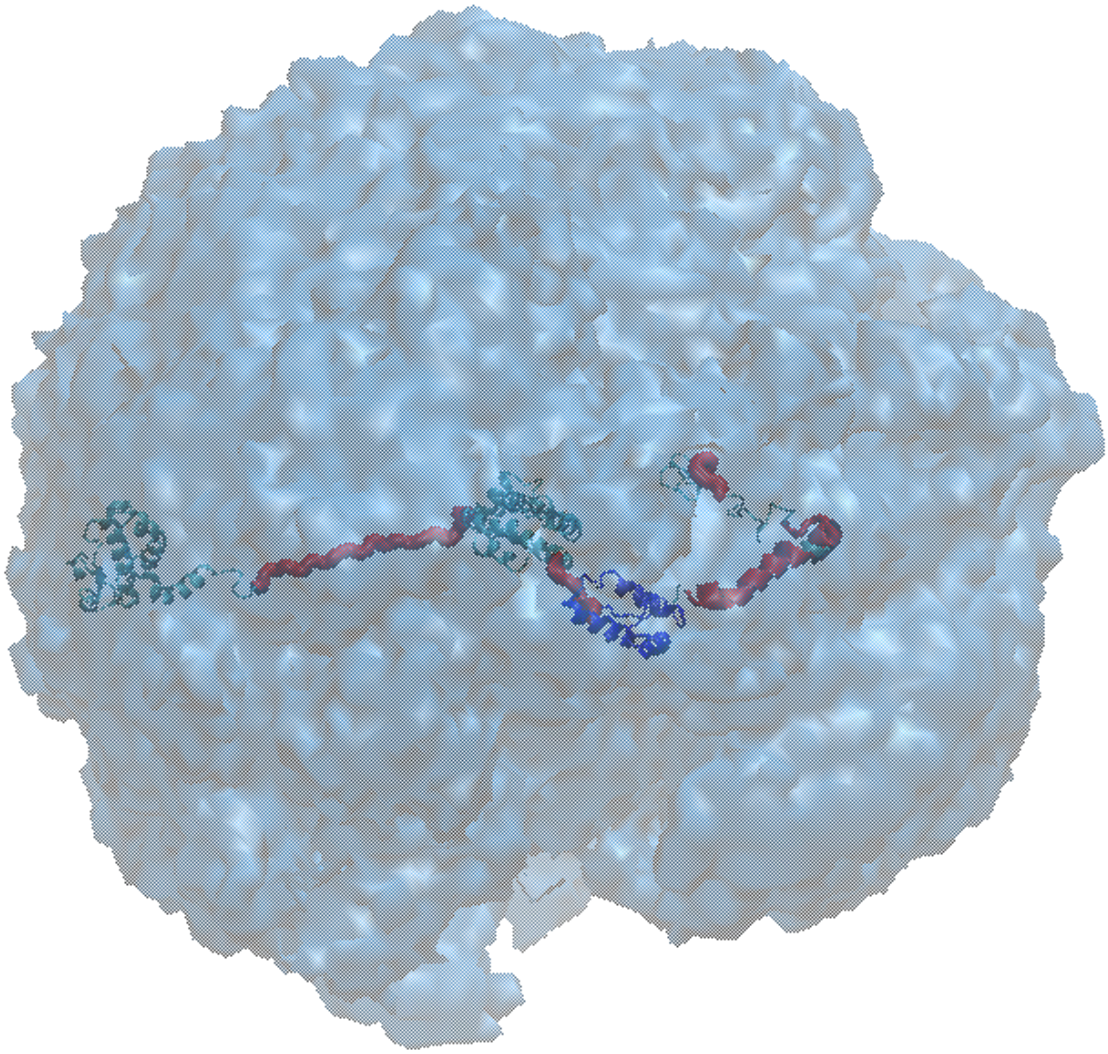
Results are written to a new directory within the given "run name" as noted in the output. In addition, a plot of Rg versus structure number is shown.

Several files are generated and saved to the "run name" monomer_monte_carlo directory. A copy of the original input PDB file, the output DCD file containing accepted structures, files with Rg values as shown in the plot on the web-page, and run statistics.

```
./run_0/monomer_monte_carlo/hiv1_gag.pdb

./run_0/monomer_monte_carlo/hiv1_gag_monte_carlo.dcd

./run_0/monomer_monte_carlo/hiv1_gag_monte_carlo.dcd.all_rg_results_data.txt

./run_0/monomer_monte_carlo/hiv1_gag_monte_carlo.dcd.accepted_rg_results_data.txt

./run_0/monomer_monte_carlo/hiv1_gag_monte_carlo.dcd.stats
```

Visualization

In the figure below, the original input structure of hiv1_gag inside the envelope sampled by all accepted structures. The envelope was created using the Density Plot module.



Files Used and Created in Example

- input files

  hiv1_gag.pdb

- output files

  caution: DCD file is > 450 MB

  hiv1_gag_monte_carlo.dcd

hiv1_gag_monte_carlo.dcd.all_rg_results_data.txt
hiv1_gag_monte_carlo.dcd.accepted_rg_results_data.txt
hiv1_gag_monte_carlo.dcd.stats

Advanced Input Options

The input variables are listed below.

- **low Rg cutoff** Structures with Rg values less than this value are discarded.
- **high Rg cutoff** Structures with Rg values greater than this value are discarded.
- **check box to use Z coordinate filter** Check box to implement the ability to discard structures with any Z coordinates with a value less than the user supplied Z cutoff value.
    - **Z cutoff** Value in angstroms to determine wheter a structure should be discarded.
- **directed Monte Carlo (0==no or Rg value)** Enter a non-zero value to use an extra energy term in the Monte Carlo sampling to favor Rg values towards the supplied value. The default value is zero which indicates that no bias is implemented.
- **check box to use atomic constraints** Check box to implement the ability to discard structures that do not satisfy the atomic / geometric constraints provided in the user defined constraint file.
    - **constraint file name** Choose a text file with constraint definitions. See Constraints for guidance as to how to create such a file with desired constraints.

In the following sections examples will be shown for the various options in the Advanced Input section.

Advanced Example 1: Rg cutoffs

This example uses the low Rg and high Rg cutoff inputs to restrict accepted structures to be between 55 and 60. Note that the example reports the number and percent of Rg values that do not satisfy the input cutoffs.

Advanced Input

Check Box for Advanced Input

low Rg cutoff 55

high Rg cutoff 60

check box to use Z coordinate filter

directed Monte Carlo (0==no or Rg value) 0

check box to use atomic constraints

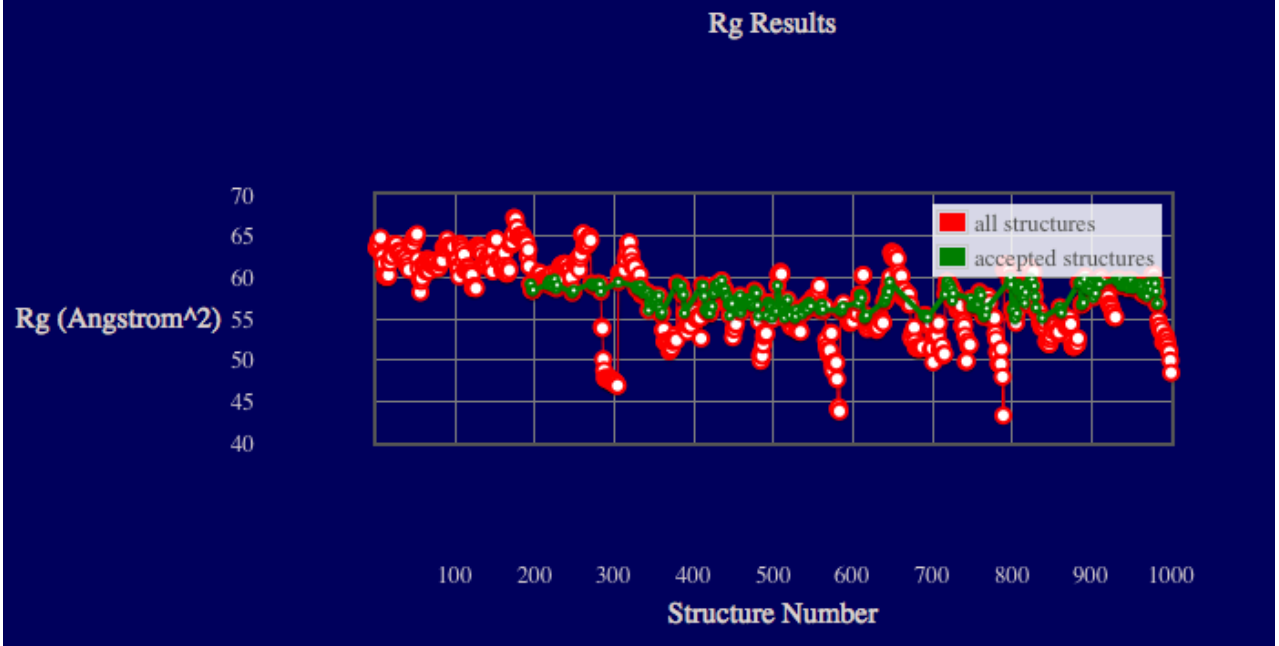
Submit Reset to default values

```
============================================================
DATA FROM RUN: Fri May  1 20:01:39 2015

Average accepted rg2 = 57.697269

Configurations and statistics saved in ./run_0/monomer_monte_carlo/ directory

lowest Rg = 43.271851      highest Rg = 67.231124
accepted 350 out of 1000 : 35.000000 percent
overlap check discarded 330 out of 1000 moves : 33.000000 percent
Rg cutoffs discarded 320 out of 1000 moves : 32.000000 percent


minimum x = -111.877888        maximum x = 69.571427 -> range: 181.449315 Angstroms
minimum y = -139.721074        maximum y = 58.253960 -> range: 197.975033 Angstroms
minimum z = -96.732521   maximum z = 102.004141 -> range: 198.736663 Angstroms


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

progress:

percent done: 100.0

all rg and accepted rg data



- input files

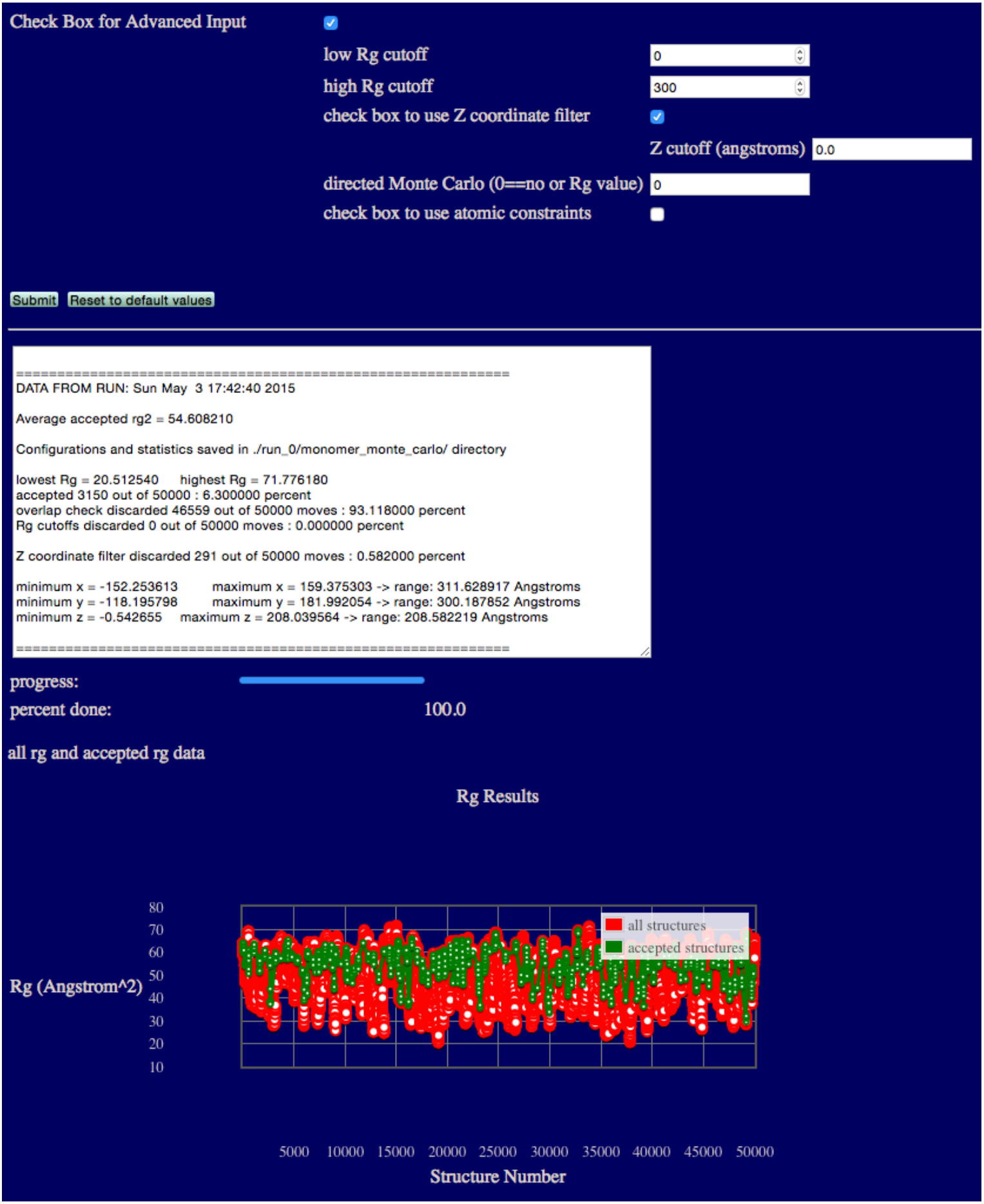
  hiv1_gag.pdb

- output files

  hiv1_gag_monte_carlo_rg_55_to_60.dcd
  hiv1_gag_monte_carlo_rg_55_to_60.dcd.all_rg_results_data.txt
  hiv1_gag_monte_carlo_rg_55_to_60.dcd.accepted_rg_results_data.txt
  hiv1_gag_monte_carlo_rg_55_to_60.dcd.stats

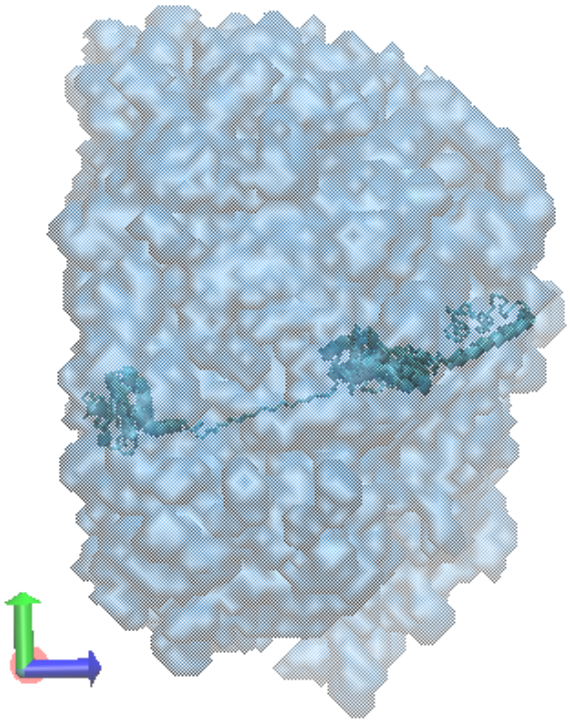
Advanced Example 2: Z coordinate filter

This example restrict accepted structures to be those with all Z coordinates to be greater than 0.



Visualization : Z coordinate filter

In the figure below, the original input structure of hiv1_gag inside the envelope sampled by all accepted structures. The envelope was created using the Density Plot module.



- input files

  hiv1_gag_on_membrane.pdb

- output files

  caution: DCD file is > 240 MB

  hiv1_gag_on_membrane.dcd
  hiv1_gag_on_membrane.dcd.all_rg_results_data.txt
  hiv1_gag_on_membrane.dcd.accepted_rg_results_data.txt
  hiv1_gag_on_membrane.dcd.stats

Advanced Example 3: Directed Monte Carlo

This example biases the Monte Carlo sampling to accept Rg values closer to 30.

Advanced Input

Check Box for Advanced Input

low Rg cutoff 0

high Rg cutoff 300

check box to use Z coordinate filter

directed Monte Carlo (0==no or Rg value) 30

check box to use atomic constraints

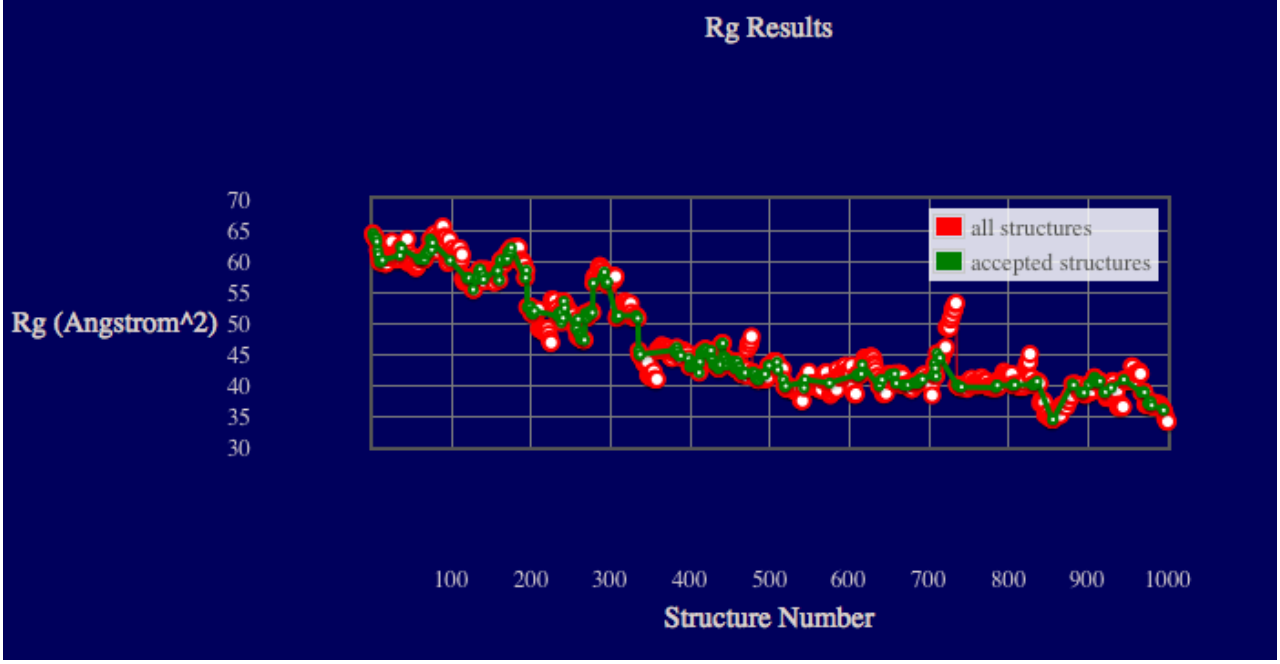
Submit Reset to default values

```
============================================================
DATA FROM RUN: Fri May  1 20:31:26 2015

Average accepted rg2 = 47.282767

Configurations and statistics saved in ./run_0/monomer_monte_carlo/ directory

lowest Rg = 34.073593     highest Rg = 65.576689
accepted 328 out of 1000 : 32.800000 percent
overlap check discarded 672 out of 1000 moves : 67.200000 percent
Rg cutoffs discarded 0 out of 1000 moves : 0.000000 percent


minimum x = -126.411141       maximum x = 51.737952 -> range: 178.149093 Angstroms
minimum y = -118.585359       maximum y = 87.510928 -> range: 206.096287 Angstroms
minimum z = -107.891278       maximum z = 113.100985 -> range: 220.992263 Angstroms


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

progress:

percent done:

all rg and accepted rg data



- input files

  hiv1_gag.pdb

- output files

  hiv1_gag_monte_carlo_directed_rg_30.dcd
  hiv1_gag_monte_carlo_directed_rg_30.dcd.all_rg_results_data.txt
  hiv1_gag_monte_carlo_directed_rg_30.dcd.accepted_rg_results_data.txt
  hiv1_gag_monte_carlo_directed_rg_30.dcd.stats

Advanced Example 4: Atomic Constraints

This example only accepts structures that satisfy the user defined atomic constraints. The segment name of the protein in the hiv1_gag.pdb is "GAG". The following single line, supplied in the user supplied file "constraints.txt" will filter the structures so that onlyt structures with the center of mass of atoms in residues 240 to 260 is within 40.0 angstroms of the center of mass of CA atoms in residues 400 to 420.

Note that the constraint syntax is robust and allows for sophisticated selections, see Constraints for further details.

```
GAG 240-260 : GAG 400-420 CA : 40.0 : COM : COM
```

Check Box for Advanced Input ☑

low Rg cutoff 0

high Rg cutoff 300

check box to use Z coordinate filter ☐

directed Monte Carlo (0==no or Rg value) 0

check box to use atomic constraints ☑

constraint file name [Choose File] constraints.txt or [Browse server] *Local*: constraints.txt

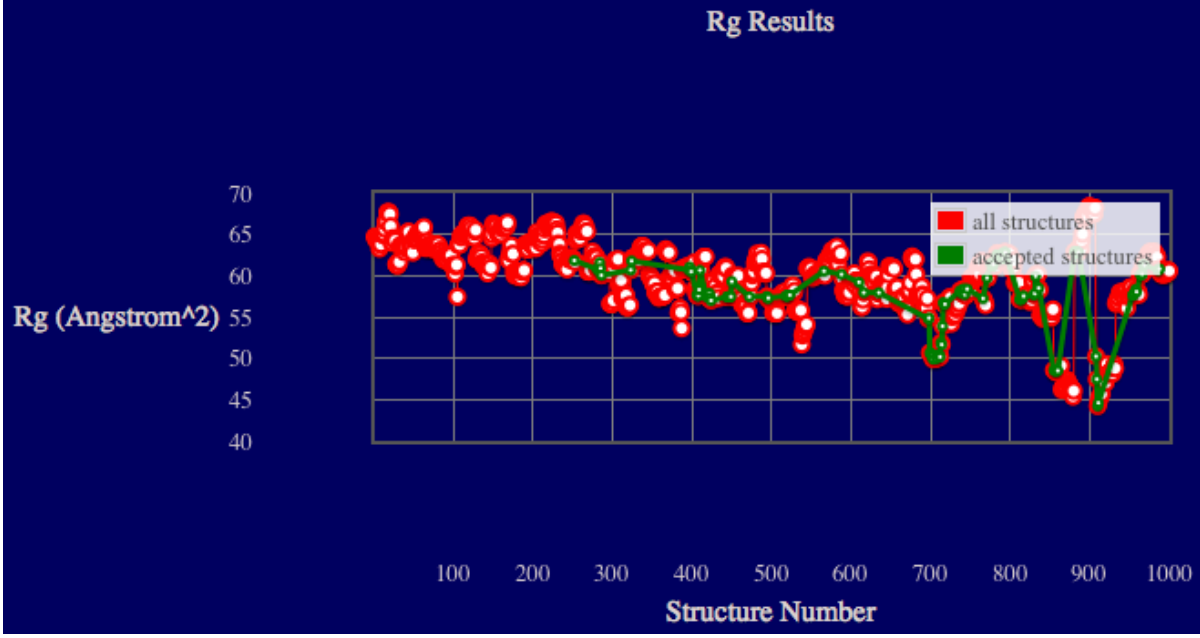
[Submit] [Reset to default values]

```
================================================================
DATA FROM RUN: Sat May  2 14:31:03 2015

Average accepted rg2 = 57.541894

Configurations and statistics saved in ./run_0/monomer_monte_carlo/ directory

lowest Rg = 44.224827     highest Rg = 68.399032
accepted 127 out of 1000 : 12.700000 percent
overlap check discarded 491 out of 1000 moves : 49.100000 percent
Rg cutoffs discarded 0 out of 1000 moves : 0.000000 percent

constraint filter(s) discarded 382 out of 1000 moves : 38.200000 percent

minimum x = -41.745841   maximum x = 88.742463 -> range: 130.488304 Angstroms
minimum y = -87.353106   maximum y = 62.849013 -> range: 150.202119 Angstroms
minimum z = -87.398178   maximum z = 119.414998 -> range: 206.813176 Angstroms

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

progress:

percent done: 100.0

all rg and accepted rg data



- input files

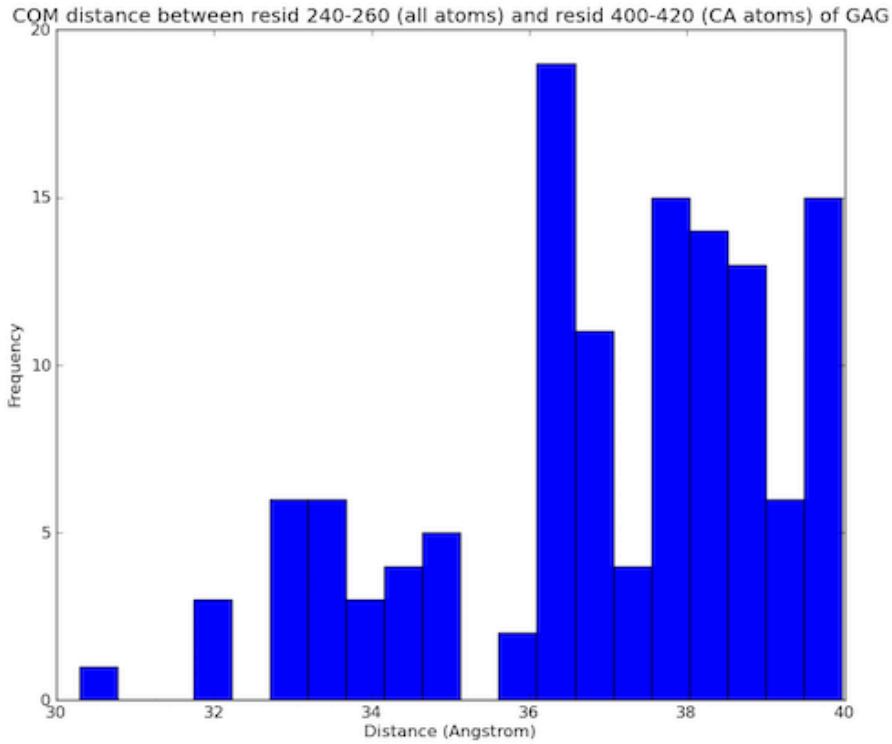
  hiv1_gag.pdb
  constraints.txt

- output files

hiv1_gag_monte_carlo_constraints.dcd
hiv1_gag_monte_carlo_constraints.dcd.all_rg_results_data.txt
hiv1_gag_monte_carlo_constraints.dcd.accepted_rg_results_data.txt
hiv1_gag_monte_carlo_constraints.dcd.stats

Visualization (Advanced Usage)

In the figure below a plot of distances between the center of mass of residues 240 to 260 and the center of mass of CA atoms for residues 400-420 are shown for accepted structures from the simualtion utilizing constraints.



Limitations

The program is written so that linear polymers of proteins and single-stranded nucleic acids are simulated over a specific selection of residues in a single direction.

Reference(s) and Citations

1. A solution for the best rotation to relate two sets of vectors W. Kabsch, Acta Crystallog. sect. A 32 922-923 (1976). BIBTeX, EndNote, Plain Text
2. A discussion of the solution for the best rotation to relate two sets of vectors W. Kabsch, Acta Crystallog. sect. A 34 827-828 (1978). BIBTeX, EndNote, Plain Text
3. CHARMM: The energy function and its parameterization with an overview of the program A. D. MacKerel Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, The Encyclopedia of Computational Chemistry, John Wiley & Sons: Chichester, 271-277 (1998). BIBTex, Endnote, Plain Text
4. 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
5. 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 Simulate

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

Complex Monte Carlo

Performs molecular Monte Carlo simulation of multiple chain protein and single chain nucleic acids.

Accessibility

The Complex Monte Carlo module is accessible from the Simulate section of the main menu.

Basic Usage

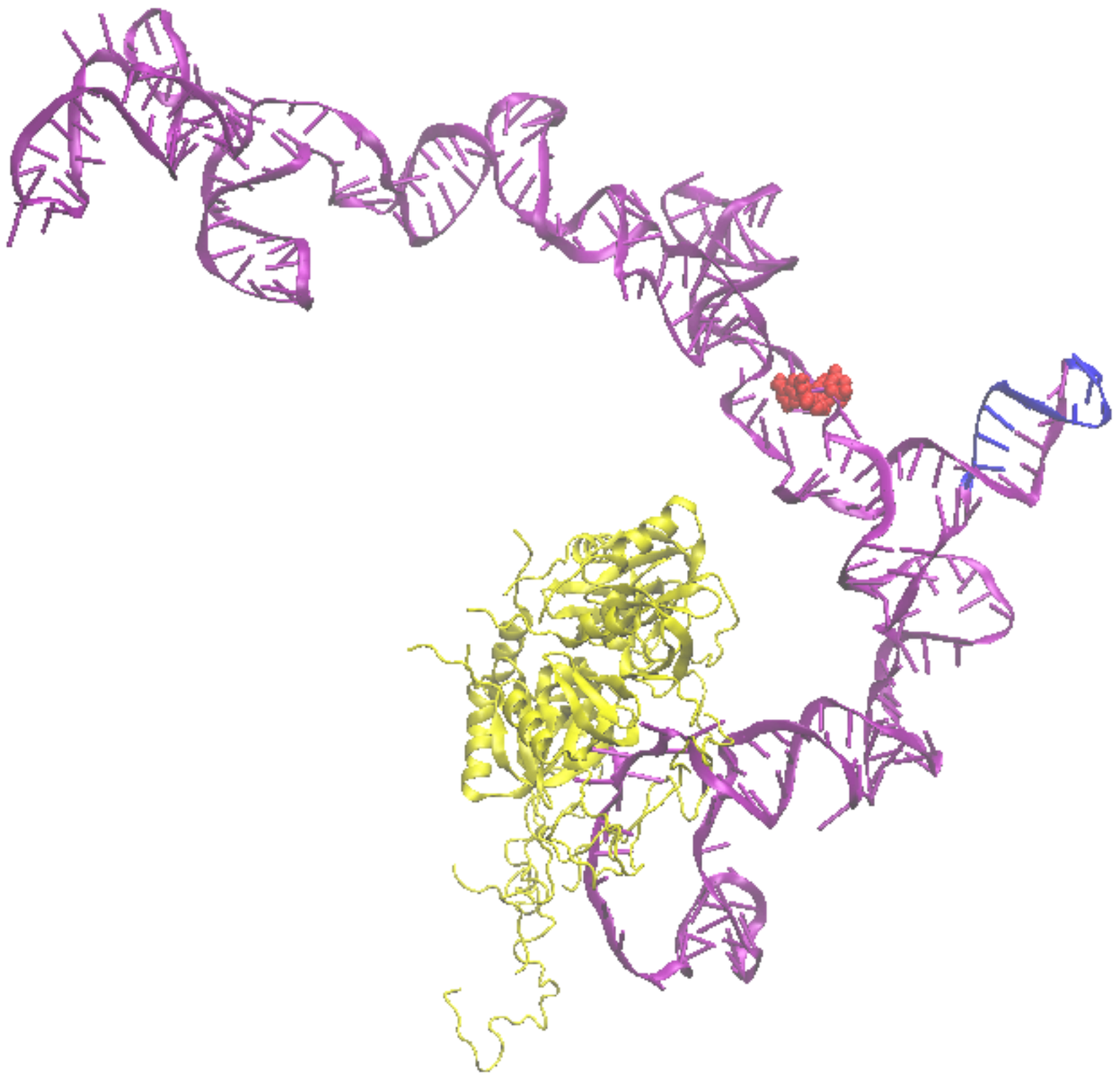
The purpose of the module is perform a molecular simulation of an multiple chain protein and/or single stranded nucleic acid by sampling backbone torsion angles.

Notes

- The starting structure must be a complete structure without missing residues. Atom and residue naming must be compatiable with those defined in the CHARMM force field See Notes on Starting Structures and Force Fields and PDB Scan for further details.
- The output file format is DCD since in most cases many structures are generated. There is no option to save the output files in PDB format. One can use Extract Utilities to convert DCD files to multi-frame PDB files.
- Structures are generated by Markov Monte Carlo sampling of backbone torsion angles. Energetics of torsion angles are determined using CHARMM force field parameters.
- While only single chain backbone torsion angles can be sampled, the complete system can have regions that are static and overlap is considered. For example, one can have a composite system of a multi-domain protein, single-stranded DNA ends, along with double stranded DNA (ds-DNA) chains. The ds-DNA chains will not be sampled, but will be can be present and are treated appropriately. Therefore, not all segments need to have move-sets defined, nor do all segments need to be sampled to use this module.
- Typically, between 10,000 to 50,000 structures are required to sample adequate configuration space for most problems.
- Parameters are supplied to help guide the Monte Carlo sampling such as temperature, control of single move angle sampling per region, and directed Monte Carlo options to guide the radius of gyration (Rg) to a user supplied value.
- A utility is provided to overlap accepted structures of each segment onto a single reference frame. This is required to carry out the simulation and is also useful to visualize relative configuration coverage in an ensemble.
- Several options are offered to check for atomic overlap: heavy atoms, all, backbone, and atom name. If one chooses the atom name option, then the user will be prompted to supply an atom name that should exist in all residues and a overlap distance cutoff value. Other options set the cutoff distance automatically.
- In Advanced Input options are provided to reject structures based on Rg value, position of atoms in the Z-direction and via atomic constraints provided as a list in a text file as described in Constraints. These options are not mutually exclusive and can be used in the same run as needed.
- Typical workflows involve generating an ensemble of structures using this module, then energy minimizing the ensemble using Energy Minimization, then calculating scattering from the ensemble using modules in Tools, and finaly comparing results to experimental data using modules in Analyze.
- In many situations, multiple runs need to be carried out to find structures that cover configuration space and have scattering profiles that are in agreement with experimental data. One can use Merge Utilties to combine both the structures (DCD files) and SAS profiles into a single new DCD and single directory will correctly numbered SAS profiles.
- To simulate long random coil regions, usually at the ends of globular proteins, it is often neccessary to sub-sample accepted structures as adjacent structures can be correlated. To obtain adequate power-law scaling, one can sub-sample a trajectory using Extract Utitilies using the periodic sampling option.

Screen Shots and Description of Input Fields

This example generates a series of structures to sample configurations of a rna protein complex. The cartoon of the starting structure highlights the flexible regions (red balls) and structure alignment region (blue). RNA is shown in purple and protein is yellow.



Complex Monte Carlo

run name: run_0

reference pdb: Choose File | rna_protei...mplex.pdb | or Browse server | *Local*: rna_protein_complex.pdb

output file name (dcd): run_0.dcd

number of trial attempts: 1000

return to previous structure: 20

temperature (K): 300.0

Complex Specific Input

Enter TOTAL number of segments: 7

number of flexible segments: 1

molecule type [1]: rna | flexible segment name [1]: RNA1 | number of flexible regions [1]

flexible residue range(s) [1]: 128-129 | maximum angle(s) [1]: 30 | structure alignment range [1]: 20-30

overlap basis: heavy atoms

Advanced Input

Check Box for Advanced Input

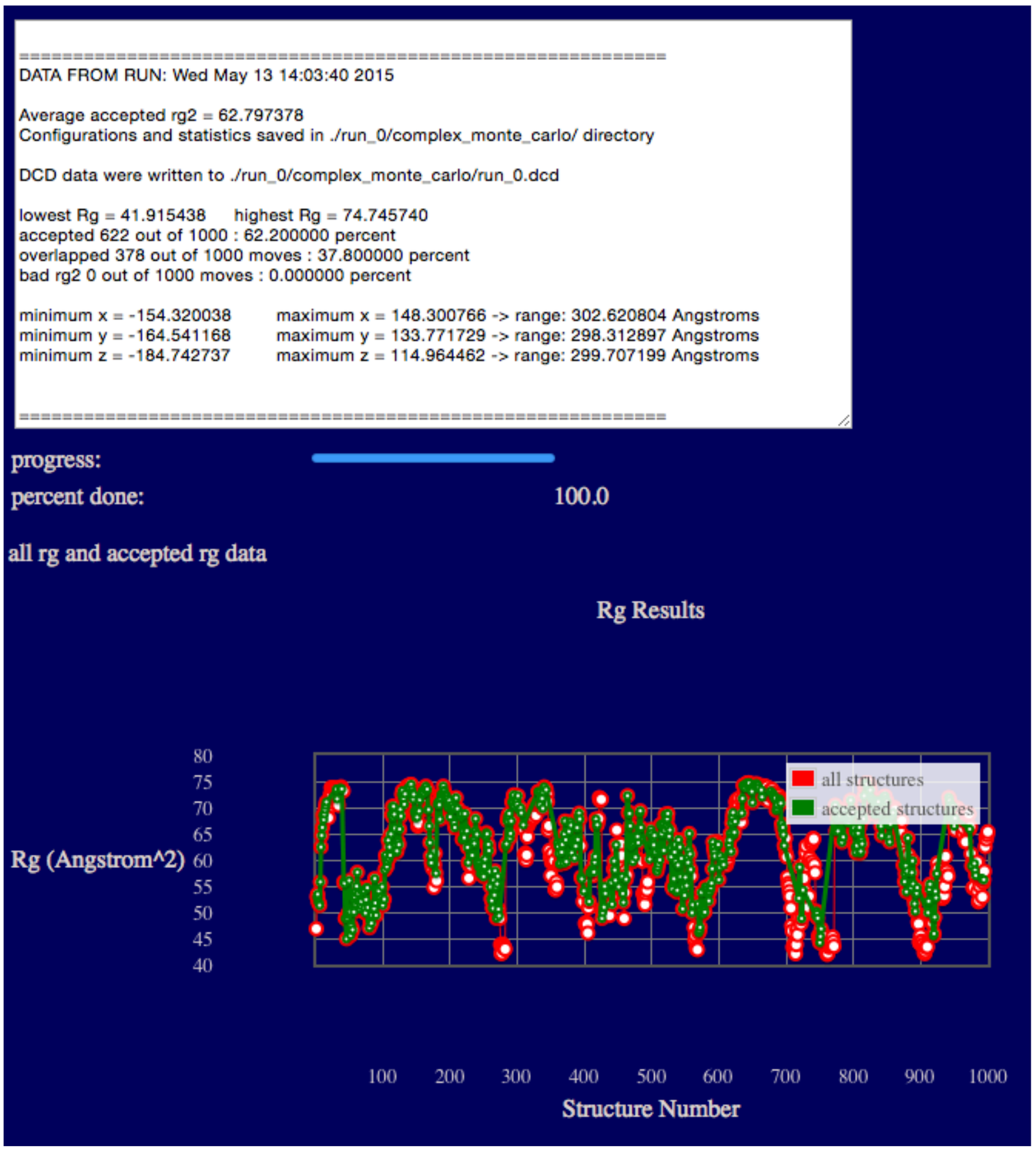
- **run name** user defined name of folder that will contain the results.
- **reference pdb** PDB file with naming information and coordinates of the starting structure.

- **output file name (dcd)** Name of ouput DCD file containing accepted structures from the simulation.
- **number of trial attempts** Number of Monte Carlo moves to attempt.
- **return to previous structure** After this number of Monte Carlo moves fails to find an accepted configuration, re-load a previously accepted structure.
- **temperature (K)** Simulation temperature.

Complex Specific Input

- **Enter TOTAL number of segments** An integer value indicating the number of rigid and flexible segments in the input PDB file.
- **number of flexible segments** An integer value indicating the number of regions to sample backbone torsions.
    - **molecule type** Select either protein or RNA.
    - **flexible segment name** Name of particular flexible segment .
    - **number of flexible regions** An integer value indicating the number of regions to sample backbone torsions.
    - **flexible residue range(s)** Residue numbers defining each flexible region in segment. The number of pairs should match the number of flexible regions for the given segment. Pairs of integers separated by hypens with each pair separated by commas.
    - **maximum angle(s)** Angle, in degrees, that can be sampled in a single move for each region.
    - **structure alignment range** Residue to define the beginning of region used to align structures for the given segment.
- **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.

Example Output



The output will indicate various Rg values from the ensemble, acceptance and overlap statistics, and dimensions of the accepted structures in the final ensemble.

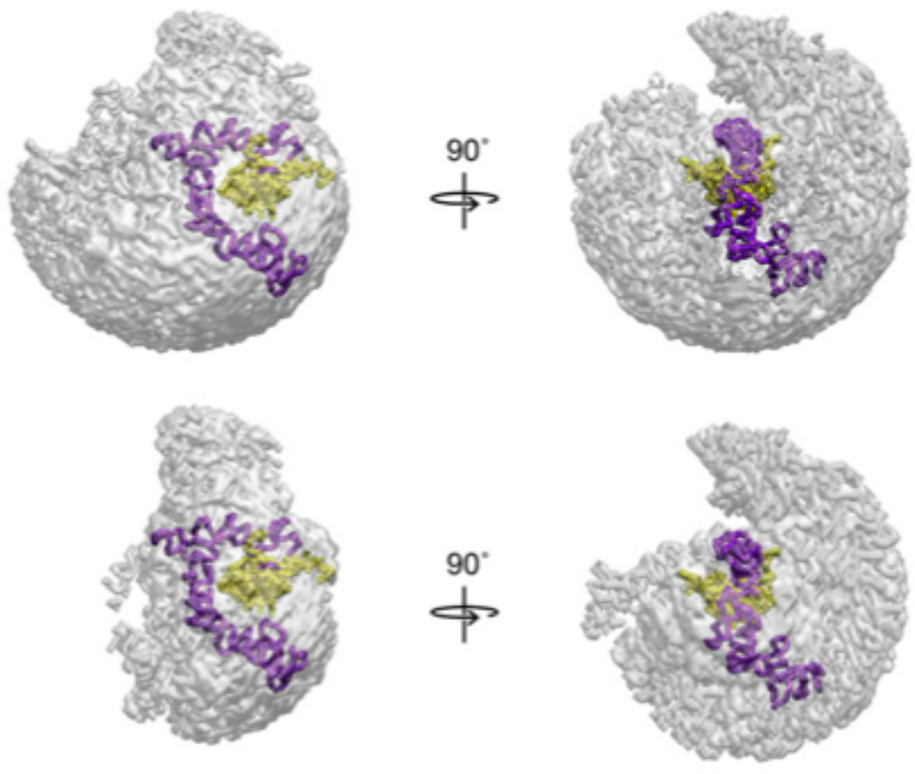
Results are written to a new directory within the given "run name" as noted in the output. In addition, a plot of Rg versus structure number is shown.

Several files are generated and saved to the "run name" monomer_monte_carlo directory. A copy of the original input PDB file, the output DCD file containing accepted structures, files with Rg values as shown in the plot on the web-page, and run statistics.

```
./run_0/monomer_monte_carlo/rna_protein_complex.pdb

./run_0/complex_monte_carlo/run_0.dcd

./run_0/complex_monte_carlo/run_0.dcd.all_rg_results_data.txt

./run_0/complex_monte_carlo/run_0.dcd.accepted_rg_results_data.txt

./run_0/complex_monte_carlo/run_0.dcd.stats
```

Visualization

In the figure below, the original input structure of rna protein complex inside the envelope sampled by accepted structures for a longer complex monte carlo run. The top two density plots represent all accepted structures while the bottom two density plots represent the region of space for structures with reduced chi-square values less than 1.5 (see reference below for more information). The envelope was created using the Density Plot module while filtering against experimental data was carried out using the Chi-Square Filter module. From this diagram one can see that only a subset of structures in a confined set of space are consistent with the experimental SAS data.



Files Used and Created in Example

- input files

  rna_protein_complex.pdb

- output files

  run_0.dcd
  run_0.dcd.all_rg_results_data.txt
  run_0.dcd.accepted_rg_results_data.txt
  run_0.dcd.stats

Advanced Input Options

The input variables are listed below.

- **low Rg cutoff** Structures with Rg values less than this value are discarded.
- **high Rg cutoff** Structures with Rg values greater than this value are discarded.
- **check box to use Z coordinate filter** Check box to implement the ability to discard structures with any Z coordinates with a value less than the user supplied Z cutoff value.
  - **Z cutoff** Value in angstroms to determine wheter a structure should be discarded.
- **directed Monte Carlo (0==no or Rg value)** Enter a non-zero value to use an extra energy term in the Monte Carlo sampling to favor Rg values towards the supplied value. The default value is zero which indicates that no bias is implemented.
- **check box to use atomic constraints** Check box to implement the ability to discard structures that do not satisfy the atomic / geometric constraints provided in the user defined constraint file.
  - **constraint file name** Choose a text file with constraint definitions. See Constraints for guidance as to how to create such a file with desired constraints.

The Advanced Input options are used in same way as described in Monomer Monte Carlo.

Multi-chain Complex Monte Carlo Simulation Example

This example uses the same system used above with the additional caveat that you will allow 13 regions be flexible.

The inputs for the run are shown below.

Complex Monte Carlo

run name	run_0
reference pdb	Choose File rna_protein_complex.pdb or Browse server *Local*: rna_protein_complex.pdb
output file name (dcd)	run_0.dcd
number of trial attempts	100000
return to previous structure	20
temperature (K)	300.0

Complex Specific Input

Enter TOTAL number of segments:	7
number of flexible segments:	7

There are six protein segments (HFQ1, HFQ2, HFQ3, HFQ4, HFQ5, HFQ6) and one rna segment (RNA1). Each of the protein segments has both N- and C-terminal flexible regions and the RNA segment has a single flexible regions.

molecule type [1]	protein	flexible segment name [1]	HFQ1	number of flexible regions [1]	2	flexible residue range(s) [1]	2-5,67-100	maximum angle(s) [1]	30.0,30.0	structure alignment range [1]	20-30
molecule type [2]	protein	flexible segment name [2]	HFQ2	number of flexible regions [2]	2	flexible residue range(s) [2]	2-5,67-100	maximum angle(s) [2]	30.0,30.0	structure alignment range [2]	20-30
molecule type [3]	protein	flexible segment name [3]	HFQ3	number of flexible regions [3]	2	flexible residue range(s) [3]	2-5,67-100	maximum angle(s) [3]	30.0,30.0	structure alignment range [3]	20-30
molecule type [4]	protein	flexible segment name [4]	HFQ4	number of flexible regions [4]	2	flexible residue range(s) [4]	2-5,67-100	maximum angle(s) [4]	30.0,30.0	structure alignment range [4]	20-30
molecule type [5]	protein	flexible segment name [5]	HFQ5	number of flexible regions [5]	2	flexible residue range(s) [5]	2-5,67-100	maximum angle(s) [5]	30.0,30.0	structure alignment range [5]	20-30
molecule type [6]	protein	flexible segment name [6]	HFQ6	number of flexible regions [6]	2	flexible residue range(s) [6]	2-5,67-100	maximum angle(s) [6]	30.0,30.0	structure alignment range [6]	20-30
molecule type [7]	rna	flexible segment name [7]	RNA1	number of flexible regions [7]	1	flexible residue range(s) [7]	128-129	maximum angle(s) [7]	30.0	structure alignment range [7]	20-30

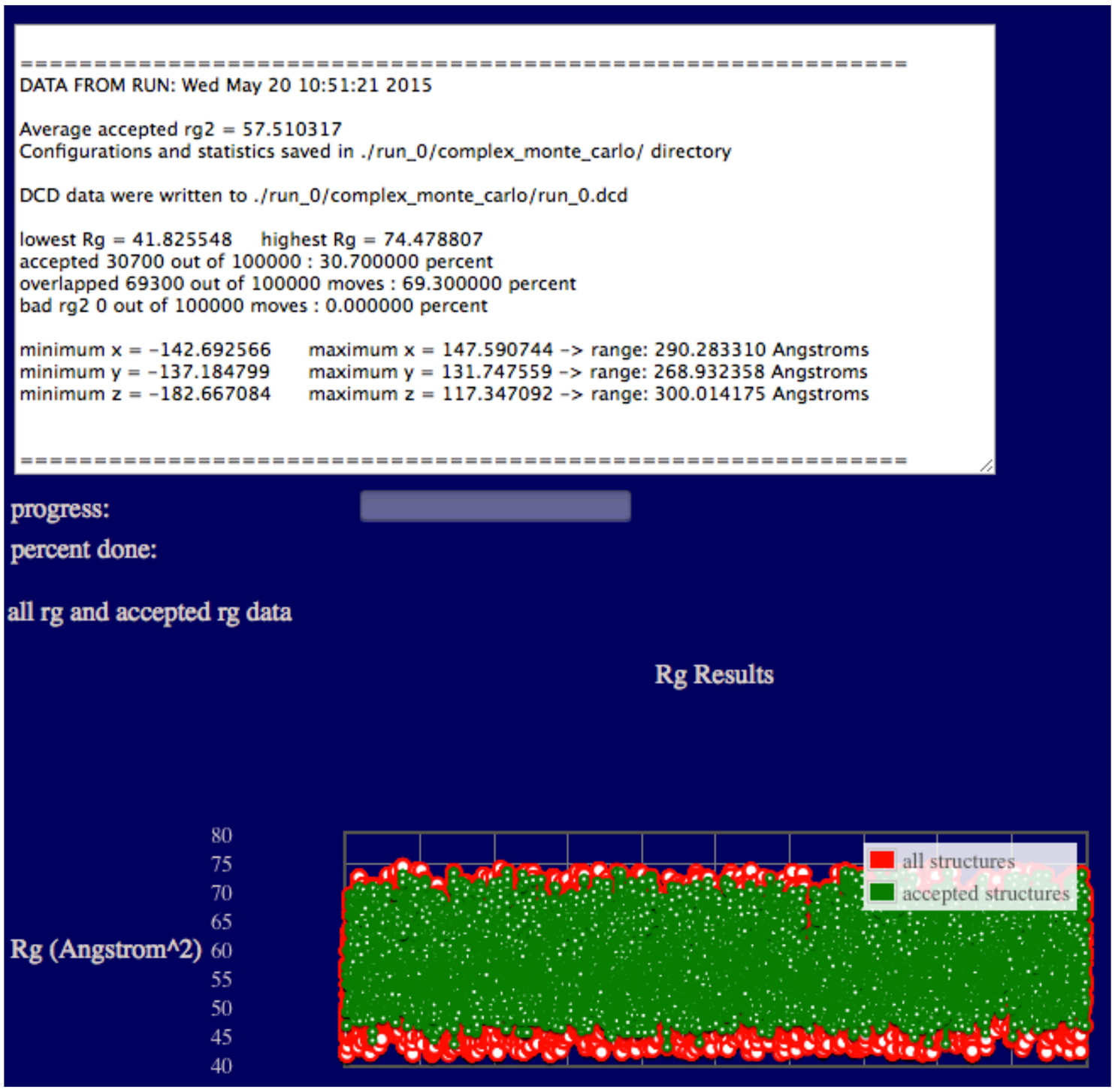
overlap basis heavy atoms

Advanced Input

Check Box for Advanced Input

Submit Reset to default values

Multi-chain Output



The output will indicate various Rg values from the ensemble, acceptance and overlap statistics, and dimensions of the accepted structures in the final ensemble.

Results are written to a new directory within the given "run name" as noted in the output. In addition, a plot of Rg versus structure number is shown.

Several files are generated and saved to the "run name" monomer_monte_carlo directory. A copy of the original input PDB file, the output DCD file containing accepted structures, files with Rg values as shown in the plot on the web-page, and run statistics.

```
./run_0/monomer_monte_carlo/rna_protein_complex.pdb

./run_0/complex_monte_carlo/run_0.dcd

./run_0/complex_monte_carlo/run_0.dcd.all_rg_results_data.txt

./run_0/complex_monte_carlo/run_0.dcd.accepted_rg_results_data.txt

./run_0/complex_monte_carlo/run_0.dcd.stats
```

Limitations

The program is written so that linear polymers of proteins and single-stranded nucleic acids are simulated over a specific selection of residues in a single direction.

Reference(s) and Citations

1. A solution for the best rotation to relate two sets of vectors W. Kabsch, Acta Crystallog. sect. A 32 922-923 (1976). BIBTeX, EndNote, Plain Text
2. A discussion of the solution for the best rotation to relate two sets of vectors W. Kabsch, Acta Crystallog. sect. A 34 827-828 (1978). BIBTeX, EndNote, Plain Text
3. CHARMM: The energy function and its parameterization with an overview of the program A. D. MacKerel Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, The Encyclopedia

of Computational Chemistry, John Wiley & Sons: Chichester, 271-277 (1998). BIBTex, Endnote, Plain Text

4. Atomistic ensemble modeling and SANS of intrinsically disordered protein complexes: applied to MCM helicase S. Krueger, J. E. Curtis, S. Raghunandan, Z. Kelman, Biophys. J. 101, 2999-3007 (2011). BIBTex, Endnote, Plain Text
5. 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
6. Structural model of an mRNA in complex with the bacterial chaperone Hfq Y. Peng, J. E. Curtis, X. Fang, S. A. Woodson, Proc. Natl. Acad. Sci. USA 111, 17134-17139 (2014). BIBTex, Endnote, Plain Text

---

Return to Simulate

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

Energy Minimization

Performs energy minimization and molecular dynamics simulation of input structures.

Accessibility

The Energy Minimization module is accessible from the Simulate section of the main menu.

Basic Usage

The purpose of the module is to use energy minimization and molecular dynamics to remove bad contacts from biomolecular models parameterized in the CHARMM forcefield.

Users will input a reference PDB file name, along with matching starting structure (in either PDB or DCD format) and CHARMM topology (PSF) files. Four modes of operation can be chosen:

1. minimization alone
2. minimization followed by molecular dynamics
3. minimization followed by molecular dynamics leading to a second round of minimization
4. molecular simulation (energy minimization and/or molecular dynamics) with a user supplied input file.

These four options are explored in three cases below.

Notes

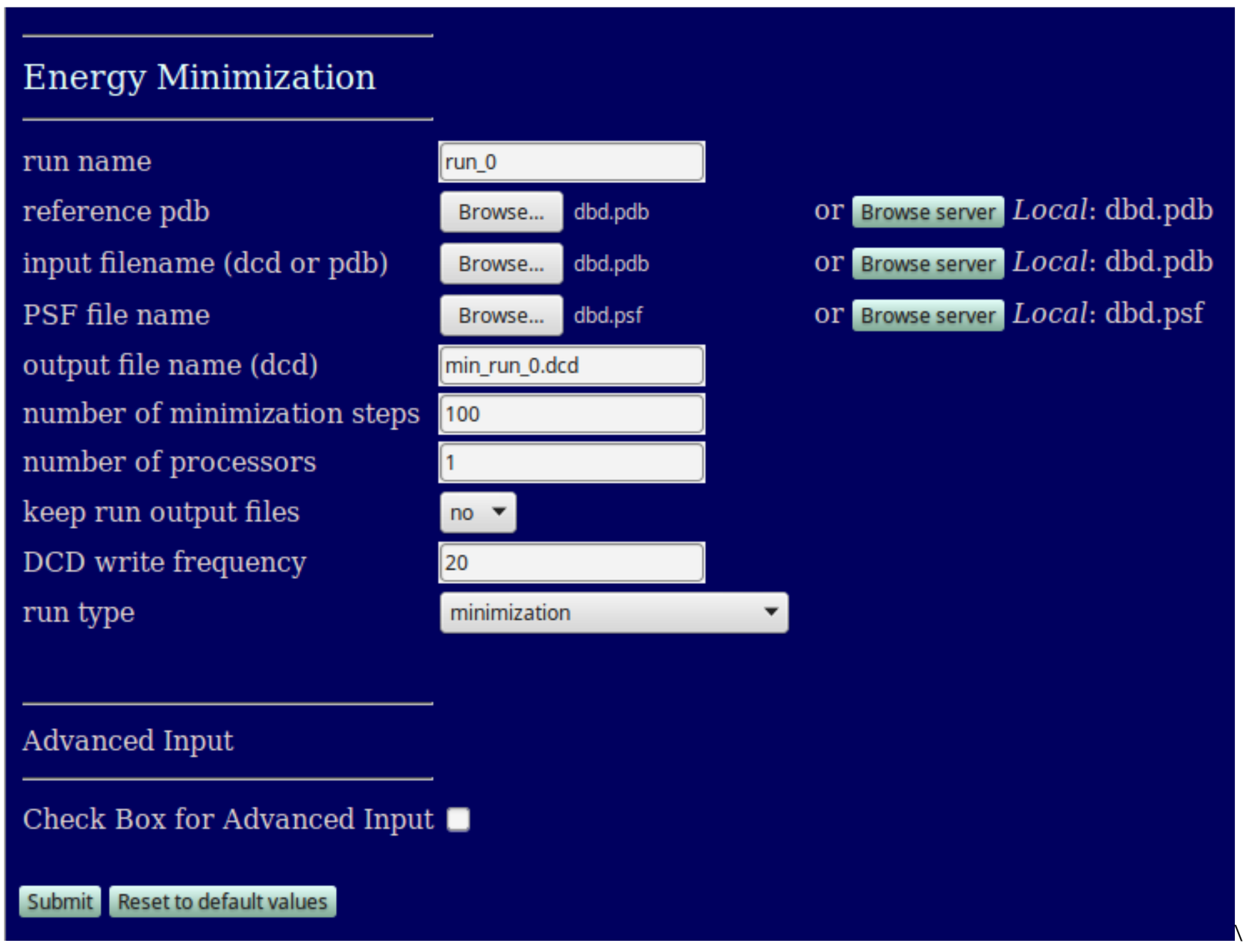
- NAMD (version 2.9) is used as the simulation engine
- Both minimization and molecular dynamics are performed using the Generalized Born implicit solvent model
- If a DCD file is selected as the input file then simulations are run on each frame

Screen Shots and Description of Input Fields

These examples show minimization of a single structure and a minimization and molecular dynamics run using advanced inputs from a DCD containing several structures.

Case 1: Single Structure Minimization

This example minimizes a single structure input as a PDB using the default CHARMM 27 forcefield.



- **run name:** user defined name of folder that will contain the results.
- **reference pdb:** PDB file with naming information for coordinates that will be extracted.
- **input filename (dcd or pdb):** file containing starting conformation(s) for simulation. The number of atoms must match that in the reference pdb. For files with multiple frames each one will be simulated.

- **PSF file name (dcd or pdb):** PSF file with topology information, must match the reference pdb and input dcd/pdb
- **output file name (dcd):** filename for the output DCD contaiing the final frames resulting from simulation
- **number of minimization steps:** number of steps of the conjugate gradient minimization to apply to each structure
- **number of processors:** number of processors used to run the simulation
- **keep run output files:** choice of whether to retain NAMD log files and other output from each simulation
- **DCD write frequency:** user determined number of steps between the output of structures from each simulation
- **run type:** select which of the four combinations of minimization and molecular dynamics to run

---

Example Output


```
./run_0/minimization/dbd.pdb
./run_0/minimization/dbd.psf
./run_0/minimization/min_run_0.dcd
./run_0/minimization/min_run_0.dcd.pdb
```

Files Created in Example

- Input files

  dbd.pdb

  dbd.psf

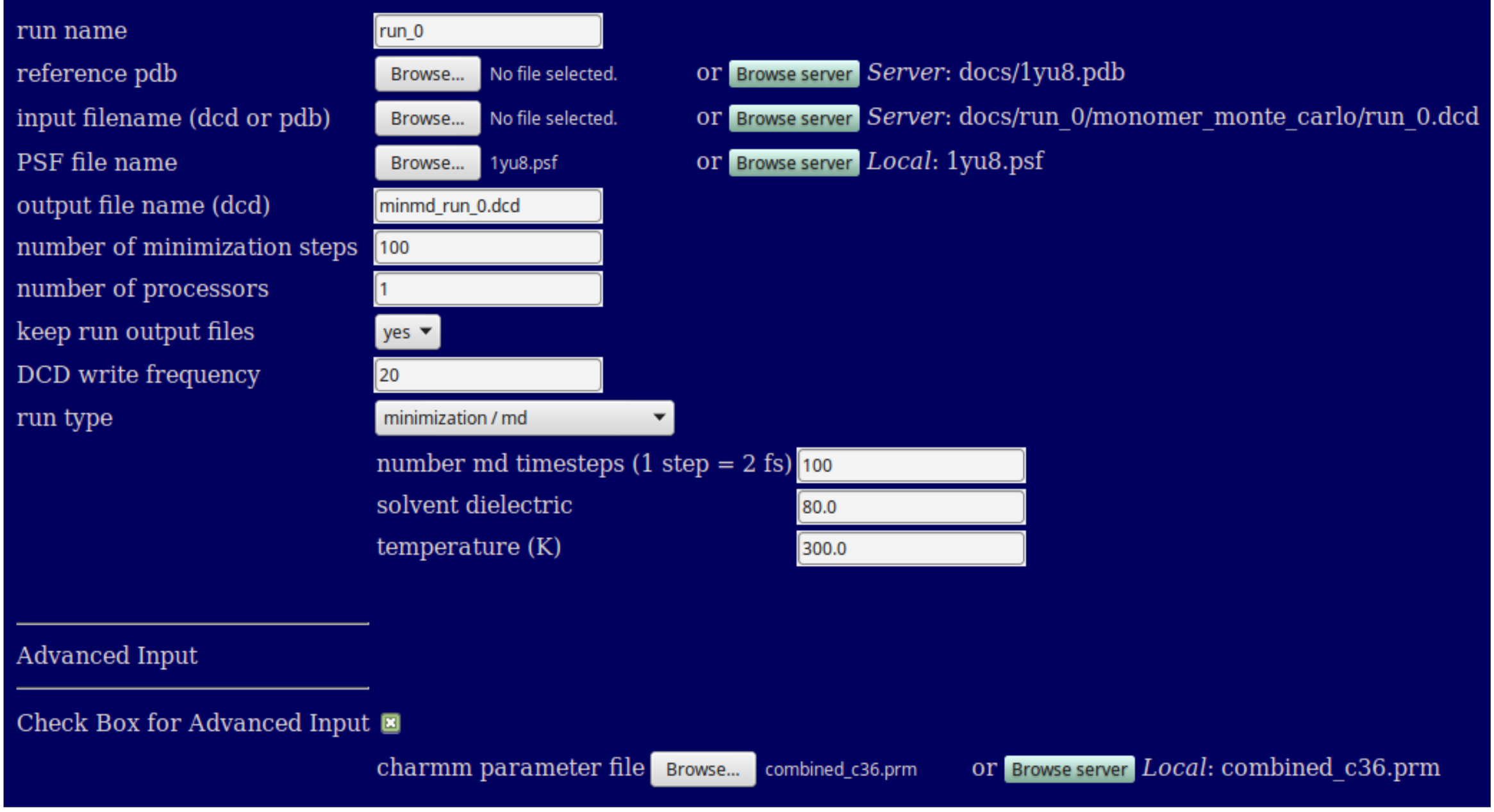
- Output files

  min_run_0.dcd

  min_run_0.dcd.pdb

Case 2: Minimization and Molecular Dynamics of Multiple Structures

This example minimizes and then runs moleclar dynamics on every structure in the input DCD using the CHARMM 36 forcefield.



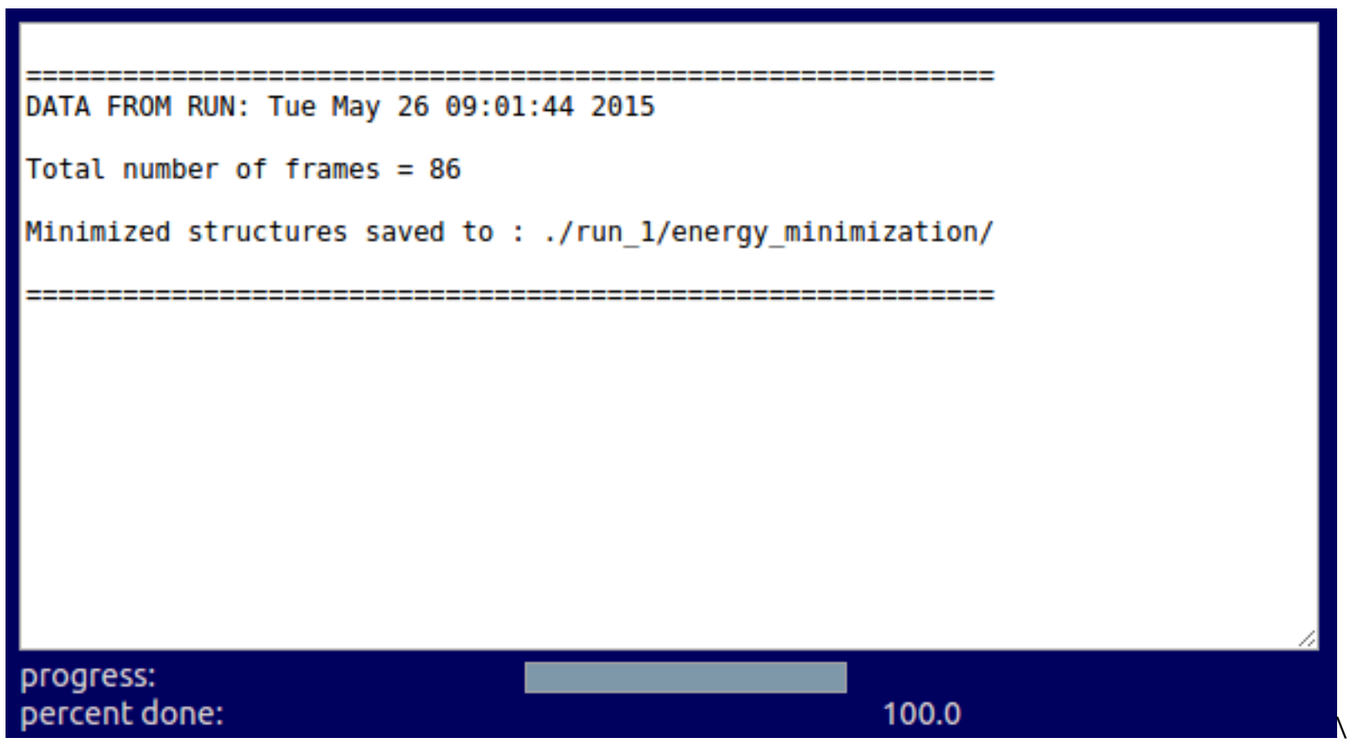
When a **run type** of minimization/md or minimization/md/minimization are selected additional options related to the molecular dynamics simulation are revealed.

- **number md timesteps (1 step = 2 fs):** number of molecular dynamics steps to be run if selected as part of the run type (ignored for minimization alone runs)
- **solvent dielectric:** Defines the dielectric of the implicit solvent used in teh simulation, usually 78.5 or 80 (the latter of which is the default).
- **temperature (K):** temperature in Kelvin at which simulatons will be performed

The advanced input section is accessed by ticking the "Check Box for Advanced Input".

- **charmm parameter file** users can specify a NAMD formatted PRM file describing a particular version of the CHARMM forcefield (this must agree with the input PSF). The default is for no file to be supplied and the CHARMM 27 forcefield is used for simulation.

Example Output




```
./run_0/minimization/1yu8.pdb
./run_0/minimization/1yu8.psf
./run_0/minimization/minmd_run_0.dcd
./run_0/minimization/minmd_run_0.dcd.pdb
min_00001.out
...
min_000086.out
temp.inp
```

Files Used and created in Example

- Input files

  1yu8.pdb

  1yu8.psf

- Output files

  minmd_run_0.dcd

  minmd_run_0.dcd.pdb

  min_00001.out

Case 3: Advanced Usage: User Supplied Input Files

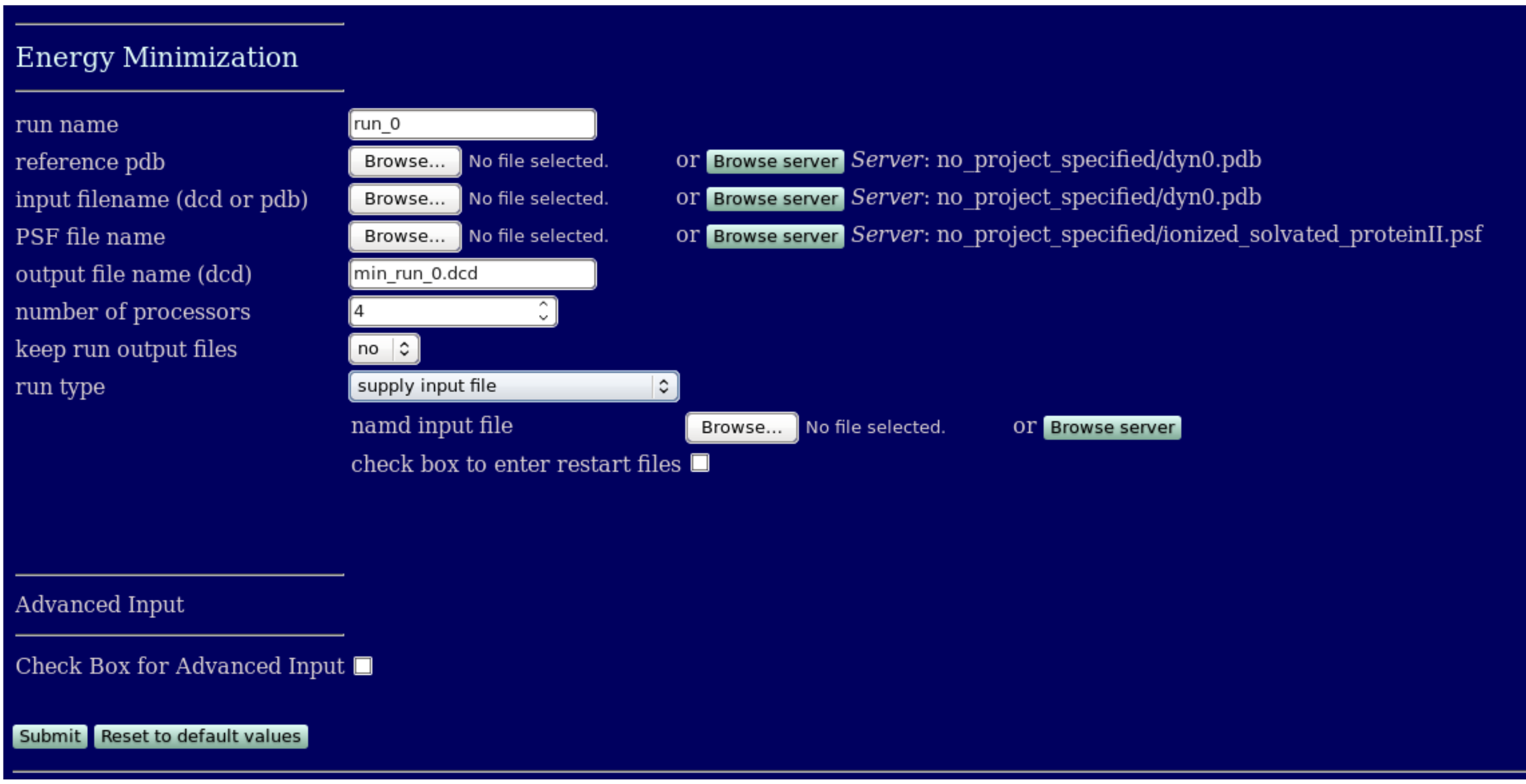
It is possible to provide an input file that allows for the simulation of more complicated systems.

This aspect is for experienced users.

When a **run type** of **supply input file** is chosen then one can select the name of the **namd** input file name.

The PDB, (DCD/PDB), PSF, and output file name read in from SASSIE-web take precendence for the files that may be listed in the user supplied input files. Specifically, the following input file keywords are overridden by the values provided by the userinput in SASSIE-web

```
coordinates
structure
paratypecharmm
parameters
outputname
DCDfile
```

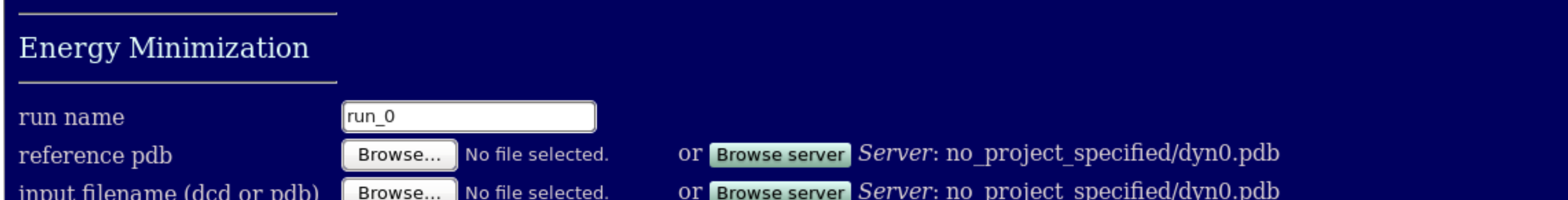


It is also possible to upload velocity and extended system files which is often required for continuuing / restarting a previous trajectory.

When a **run type** of **supply input file** is chosen and one selects the **check box to enter restart files** then the option to enter the filename(s) of **velocity restart file** and/or **extended system restart file** is provided.

If you check the box to enter restart files then the following keywords are overridden by the values provided by the user in SASSIE-web

```
velocities
extendedSystem
```



NOTES

- If you upload a velocity restart file you must not specify a temperature in your uploaded input file. In addition, if you upload an extended system file then you should not specify a set of cellBasisVector1, cellBasisVector2, cellBasisVector3 definitions.
- Relative paths for other input and output files are not accommodated. In other words, lines such as

```
restartname             output/dyn0.rest.pdb
```

should be written as

```
restartname             dyn0.rest.pdb
```

- Currently, user supplied input files are NOT filtered to assure the capability to run a successful simulation.

Files Used and Created in Case 3 Example

- Input files

  dyn0.pdb

  ionized_solvated_proteinII.psf

  dyn1

  dyn0.vel

  dyn0.xsc

- Output files

  []

Reference(s) and Citations

1. Scalable molecular dynamics with NAMD James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. J. Comput. Chem. 26, 1781-1802 (2005). BIBTeX EndNote Plain Text
2. NAMD User's Guide
3. 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 Simulate

Return to Main Documents Page

Go to top

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