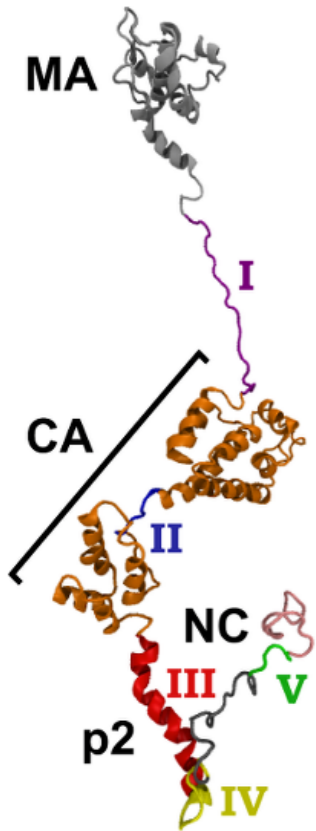


Monomer Monte Carlo single chain IDP or ID-NA



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

Monomer Monte Carlo

run name:

reference pdb: hiv1_gag.pdb OR Local: hiv1_gag.pdb

output file name (dcd):

number of trial attempts:

return to previous structure:

temperature (K):

molecule type:

number of flexible regions to vary:

maximum angle sampled for each region:

residue range for each flexible region:

structure alignment: low residue:

structure alignment: high residue:

overlap basis:

enter single chain information

Complex Monte Carlo

Complex Monte Carlo

run name

reference pdb rna_protel...mplex.pdb OR Local: rna_protein_complex.pdb

output file name (dcd)

number of trial attempts

return to previous structure

temperature (K)

Complex Specific Input

Enter TOTAL number of segments:

number of flexible segments:

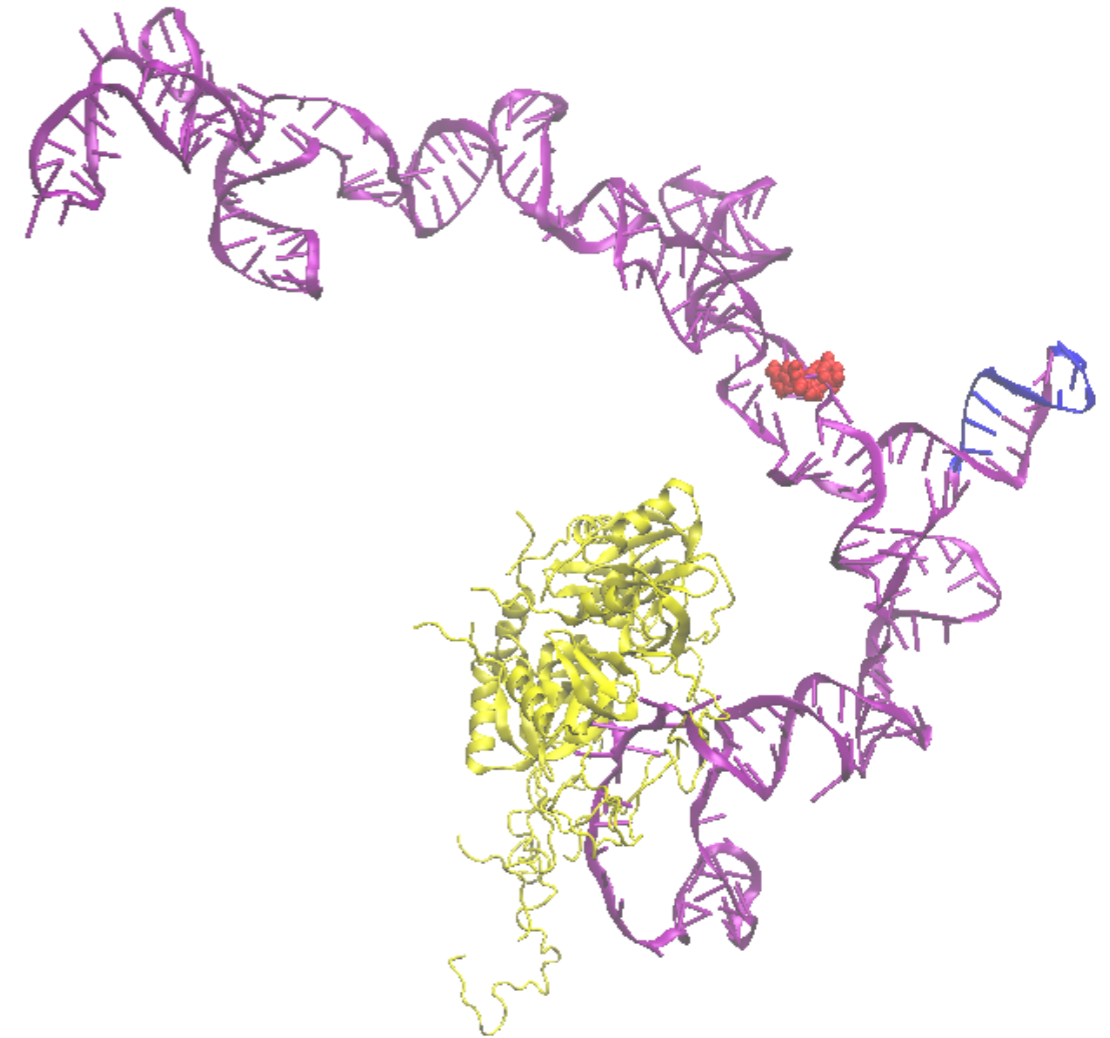
molecule type [1] flexible segment name [1] number of flexible regions [1]

flexible residue range(s) [1] maximum angle(s) [1] structure alignment range [1]

overlap basis

Advanced Input

Check Box for Advanced Input



enter total number of segments
then
enter number of flexible segments
then information on the regions that are flexible

Complex Monte Carlo

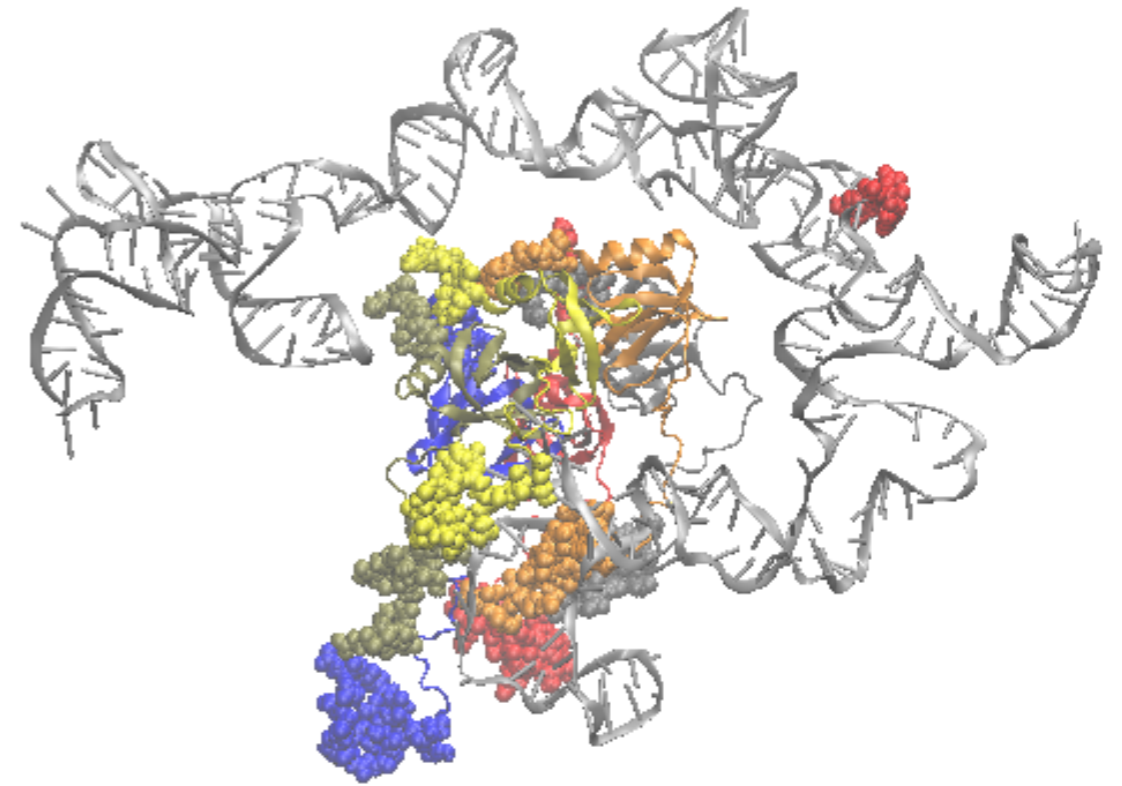
7 segments total

Protein segments:

**HFQ1, HFQ2, HFQ3, HFQ4,
HFQ5, HFQ6**

RNA segment:

RNA1



each protein segment
has two flexible regions

RNA segment has
a single flexible region

Complex Monte Carlo

| | | | | | | | | | | | |
|-------------------|---------|---------------------------|------|--------------------------------|---|-------------------------------|------------|----------------------|-----------|-------------------------------|-------|
| 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 |

Complex Monte Carlo

HIV Integrase

12 segments total

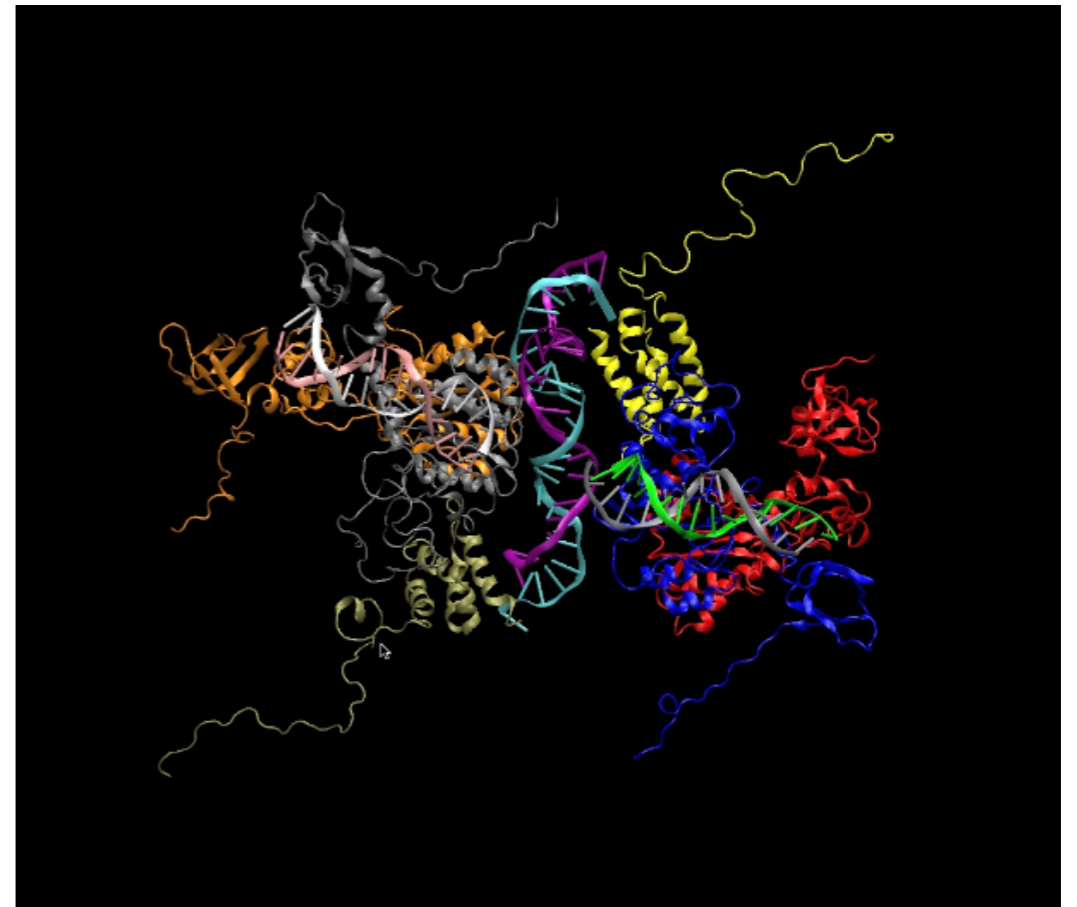
Protein segments:

**INT1, INT2, INT3, INT4,
LED1, LED2**

DNA segments:

**DNA1, DNA2, DNA3, DNA4,
DNA5, DNA6**

ds-DNA is NOT a single chain



Each protein segment
has two flexible regions

NO flexible regions in
DNA

Complex Monte Carlo

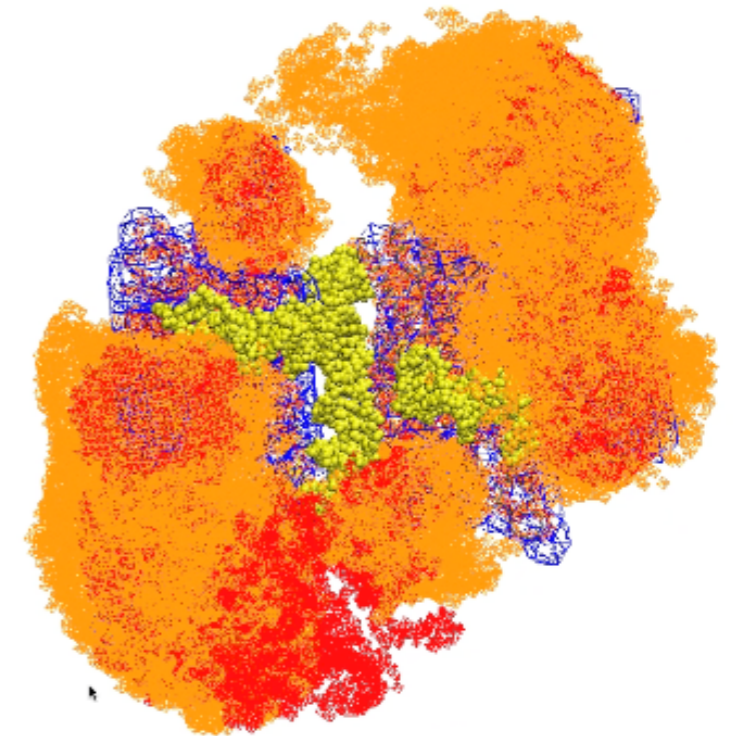
Complex Specific Input

Enter TOTAL number of segments:

number of flexible segments:

| | | | | |
|-------------------|--------------------------------------|---------------------------|-----------------------------------|--------------------------------|
| molecule type [1] | <input type="text" value="protein"/> | flexible segment name [1] | <input type="text" value="INT1"/> | number of flexible regions [1] |
| molecule type [2] | <input type="text" value="protein"/> | flexible segment name [2] | <input type="text" value="INT2"/> | number of flexible regions [2] |
| molecule type [3] | <input type="text" value="protein"/> | flexible segment name [3] | <input type="text" value="INT3"/> | number of flexible regions [3] |
| molecule type [4] | <input type="text" value="protein"/> | flexible segment name [4] | <input type="text" value="INT4"/> | number of flexible regions [4] |
| molecule type [5] | <input type="text" value="protein"/> | flexible segment name [5] | <input type="text" value="LED1"/> | number of flexible regions [5] |
| molecule type [6] | <input type="text" value="protein"/> | flexible segment name [6] | <input type="text" value="LED2"/> | number of flexible regions [6] |

overlap basis



No input for DNA (or anything that you are not sampling ... like first RNA protein example)

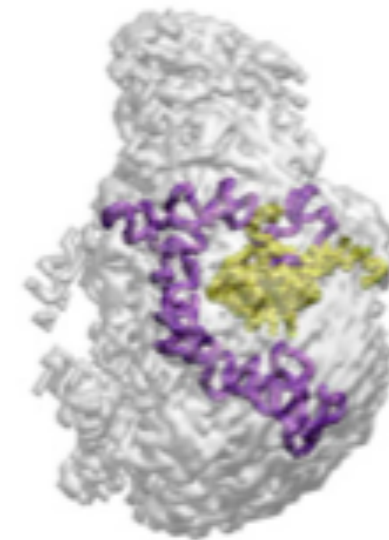
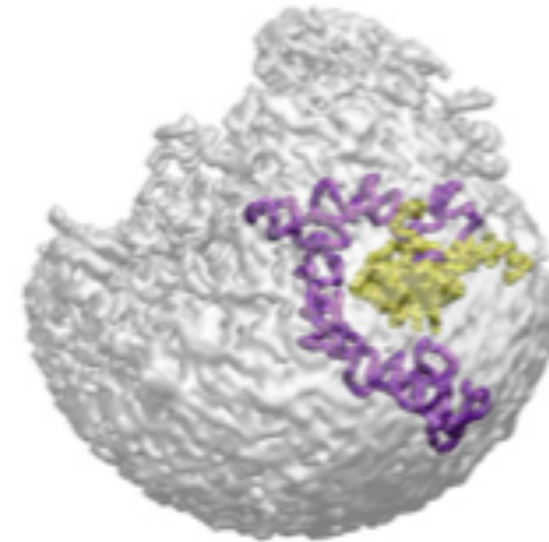
Complex Monte Carlo

Go to DOCS page for
Complex Monte Carlo

and work through the
examples ...

get comfortable with inputs

start a job or two



[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

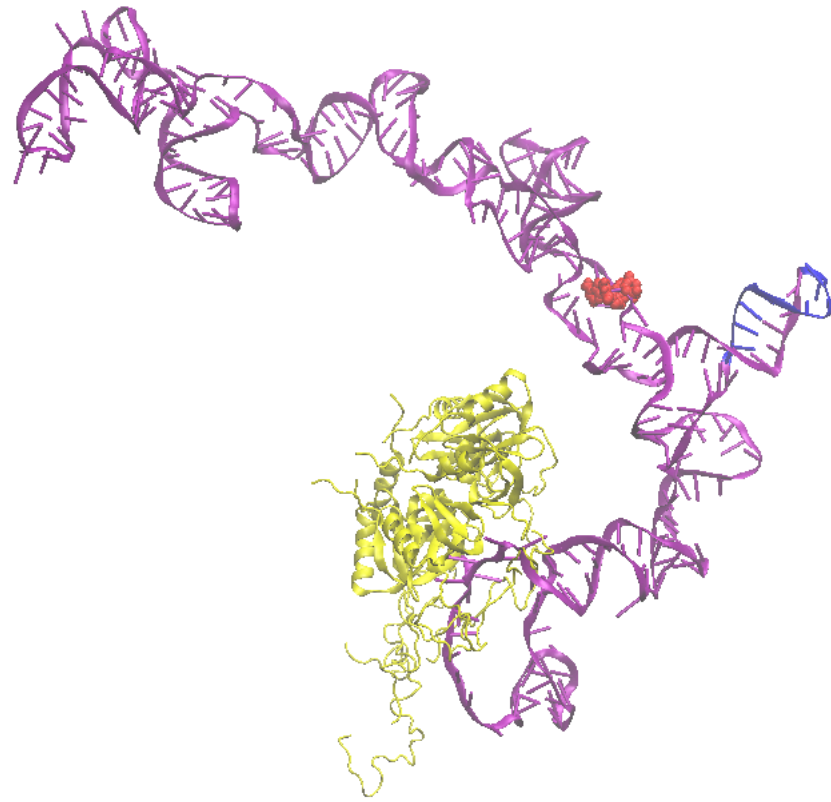
The purpose of the module is to perform a molecular simulation of a multiple chain protein and/or single stranded nucleic acid by sampling backbone torsion angles.

Notes

- The starting structure must be a complete structure without missing residues. Atom and residue naming must be compatible 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 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 an 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 finally 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 Utilities](#) 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 necessary 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 Utilities](#) using the periodic sampling option.
-

Screen Shots and Description of Input Fields

This example generates a series of structures to sample configurations of an 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

reference pdb ma_protel...mplex.pdb OR Local: ma_protein_complex.pdb

output file name (dcd)

number of trial attempts

return to previous structure

temperature (K)

Complex Specific Input

Enter TOTAL number of segments:

number of flexible segments:

molecule type [1] flexible segment name [1] number of flexible regions [1]

flexible residue range(s) [1] maximum angle(s) [1] structure alignment range [1]

overlap basis

Advanced Input

Check Box for Advanced Input

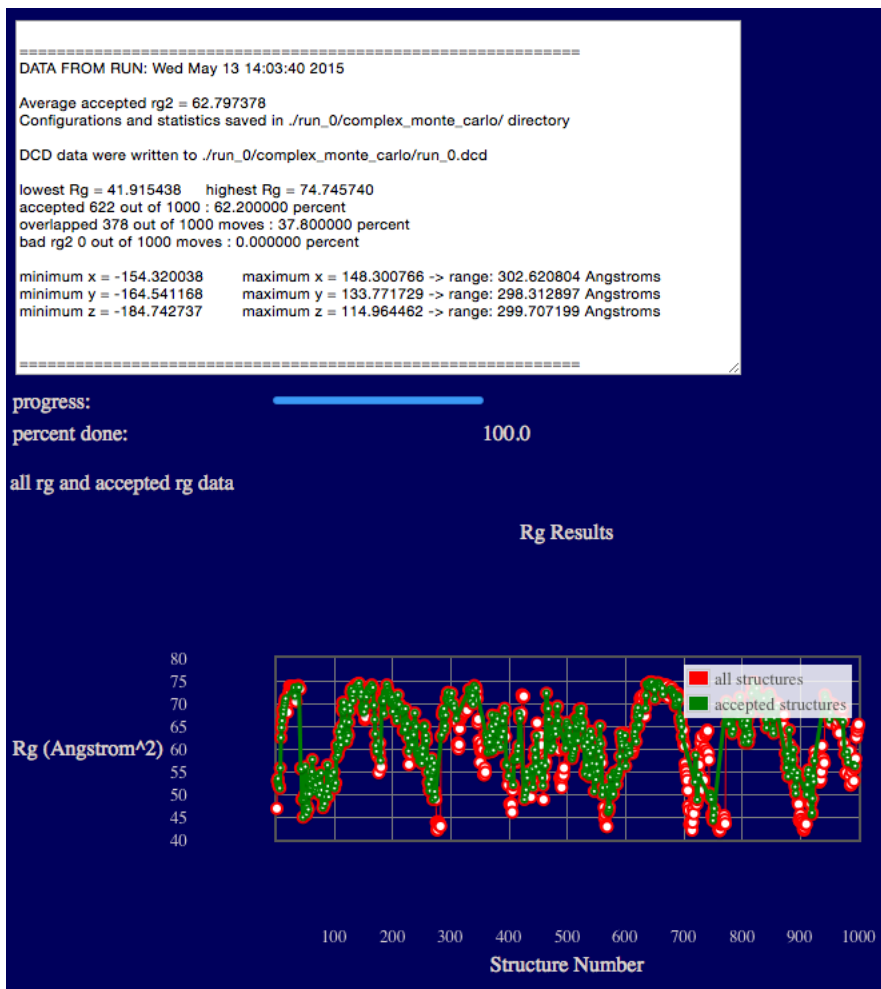
- **run name** user defined name of folder that will contain the results.
- **reference pdb** PDB file with naming information and coordinates of the starting structure.
- **output file name (dcd)** Name of output DCD file containing accepted structures from the simulation.
- **number of trial attempts** 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 hyphens 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 further inputs:

- **overlap basis** Enter an atom name to check for overlap.
- **overlap cutoff (angstroms)** 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.

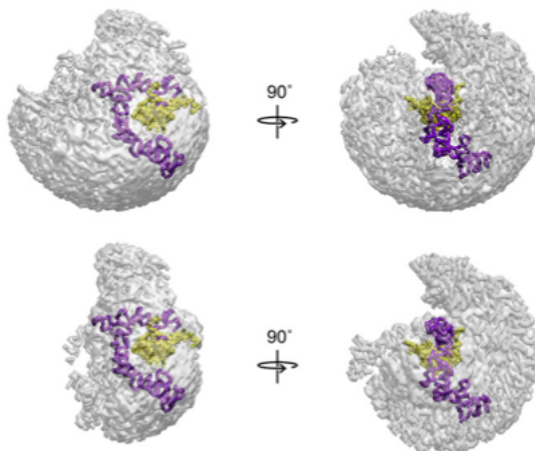
```

./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_rq_results_data.txt](#)
[run_0.dcd.accepted_rq_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 whether 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 | <input type="text" value="run_0"/> |
| reference pdb | <input type="button" value="Choose File"/> <input type="text" value="rna_protein_complex.pdb"/> or <input type="button" value="Browse server"/> <i>Local: rna_protein_complex.pdb</i> |
| output file name (dcd) | <input type="text" value="run_0.dcd"/> |
| number of trial attempts | <input type="text" value="100000"/> |
| return to previous structure | <input type="text" value="20"/> |
| temperature (K) | <input type="text" value="300.0"/> |

Complex Specific Input

| | |
|---------------------------------|---|
| Enter TOTAL number of segments: | <input type="text" value="7"/> |
| number of flexible segments: | <input type="text" value="7"/> <input type="button" value="⊕"/> |

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.

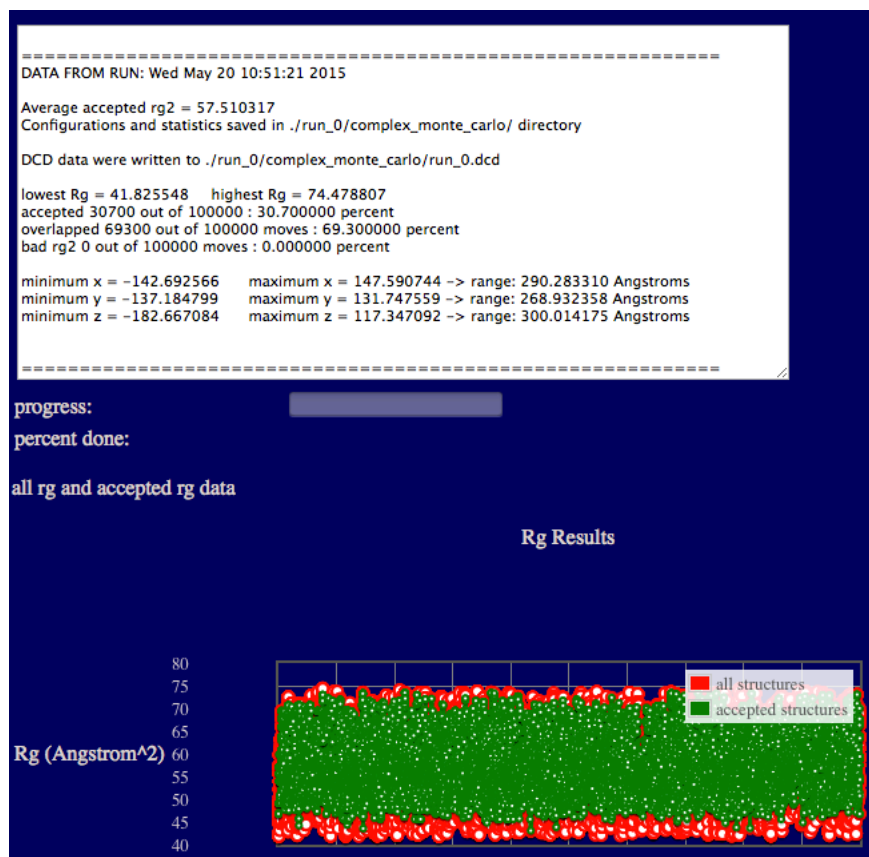
| | | | | | | | | | |
|-------------------|---------|---------------------------|------|--------------------------------|---|-------------------------------|------------|----------------------|-----------|
| molecule type [1] | protein | flexible segment name [1] | HFQ1 | number of flexible regions [1] | 2 | flexible residue range(s) [1] | 2-5,67-100 | maximum angle(s) [1] | 30.0,30.0 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |

overlap basis

Advanced Input

Check Box for Advanced Input

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