

MMC & SASSIE Overview

CCP-SAS 2015

Atomistic Modeling for Small Angle Scattering:
A Short Course

May 26-28, 2015

Institut Laue-Langevin, France

Barriers: BUILD; EQUILIBRATE; PROPAGATE; ANALYZE

What software package(s) and force-fields do I use?

Starting structure?

How do I clean up the structure?

How do I set up a trajectory (time or **space**)?

How do I calculate scattering observables correctly?

1 ns

VMD Main

File Molecule Graphics Display Mouse Extensions Help

ID	T	A	D	F	Molecule	Atoms	Frames	Vol
0	T	A	D	F	solvated_complex0_min	397028	500	0

0

zoom Once step 10 speed

Graphical Representations

Selected Molecule

0: solvated_complex0_min0.psf

Create Rep Delete Rep

Style	Color	Selection
NewCartoon	SegName	not water
Points	SegName	water

Selected Atoms

water

Draw style Selections Trajectory Periodic

Coloring Method

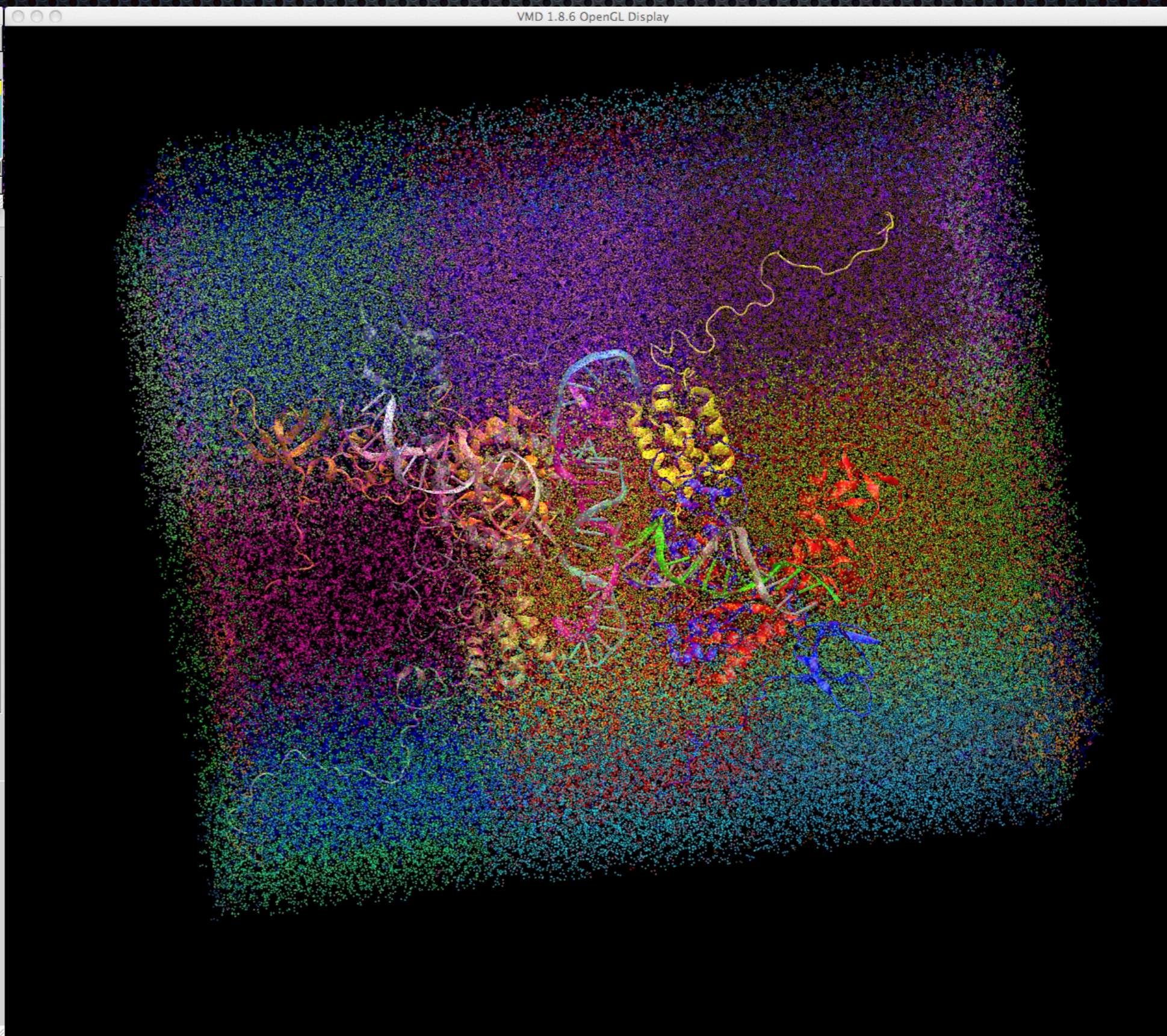
SegName Opaque

Drawing Method

Points Default

Size 1

Apply Changes Automatically Apply



1 ns

VMD Main

File Molecule Graphics Display Mouse Extensions Help

ID	T	A	D	F	Molecule	Atoms	Frames	Vol
0	T	A	D	F	solvated_complex0_min	397028	500	0

0

zoom Once step 10 speed

Graphical Representations

Selected Molecule

0: solvated_complex0_min0.psf

Create Rep Delete Rep

Style	Color	Selection
NewCartoon	SegName	not water
Points	SegName	water

Selected Atoms

water

Draw style Selections Trajectory Periodic

Coloring Method

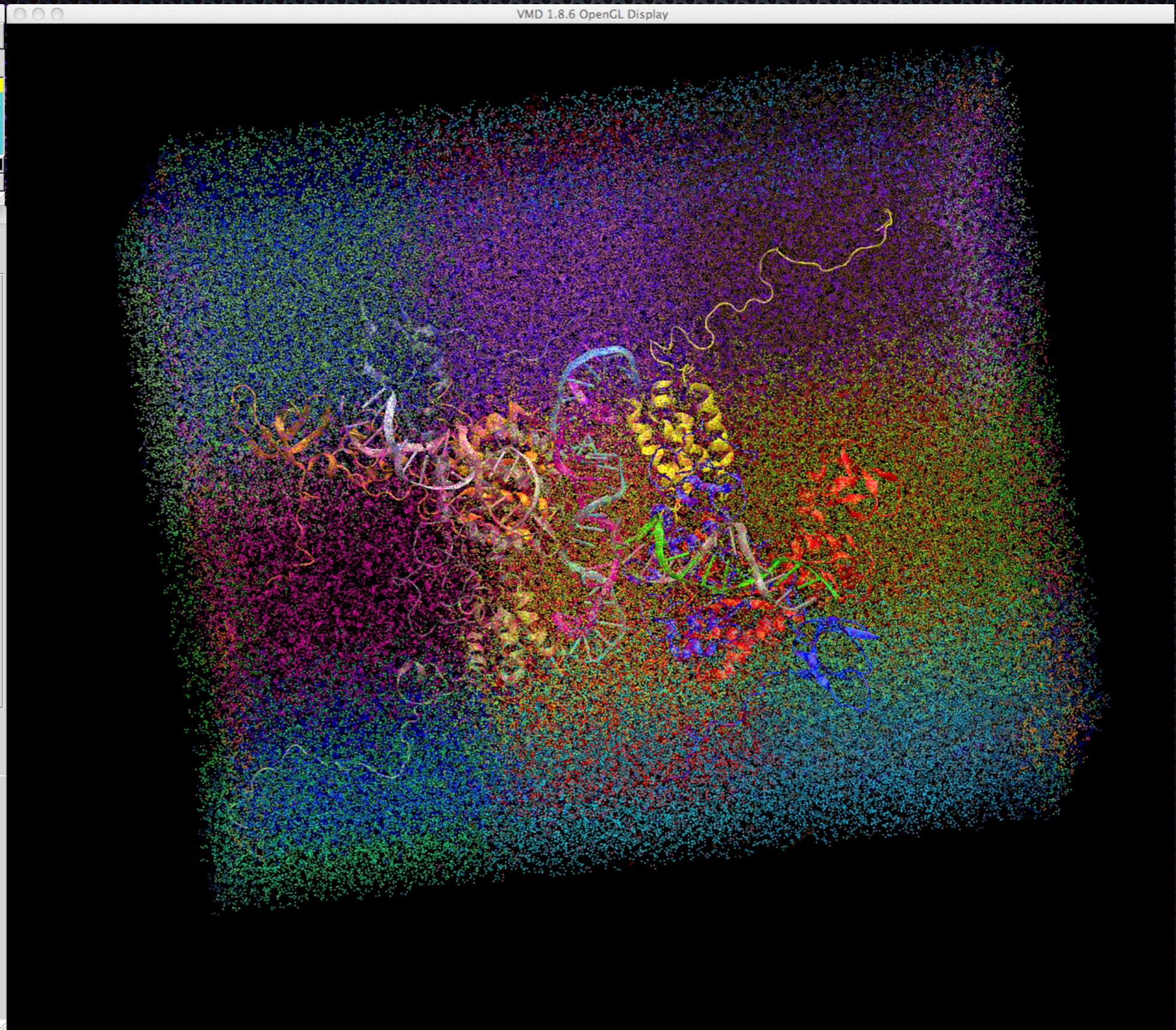
SegName Opaque

Drawing Method

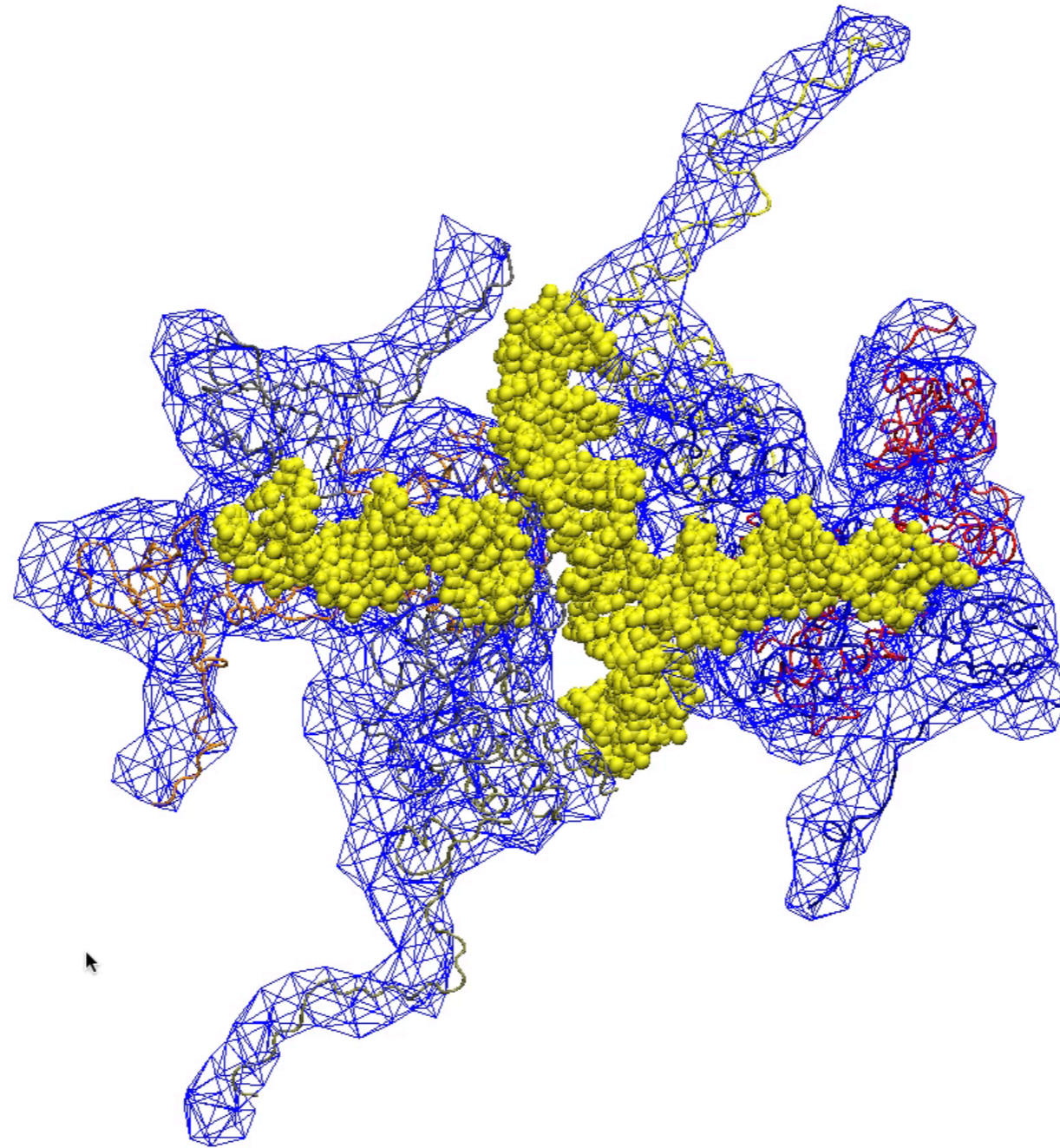
Points Default

Size 1

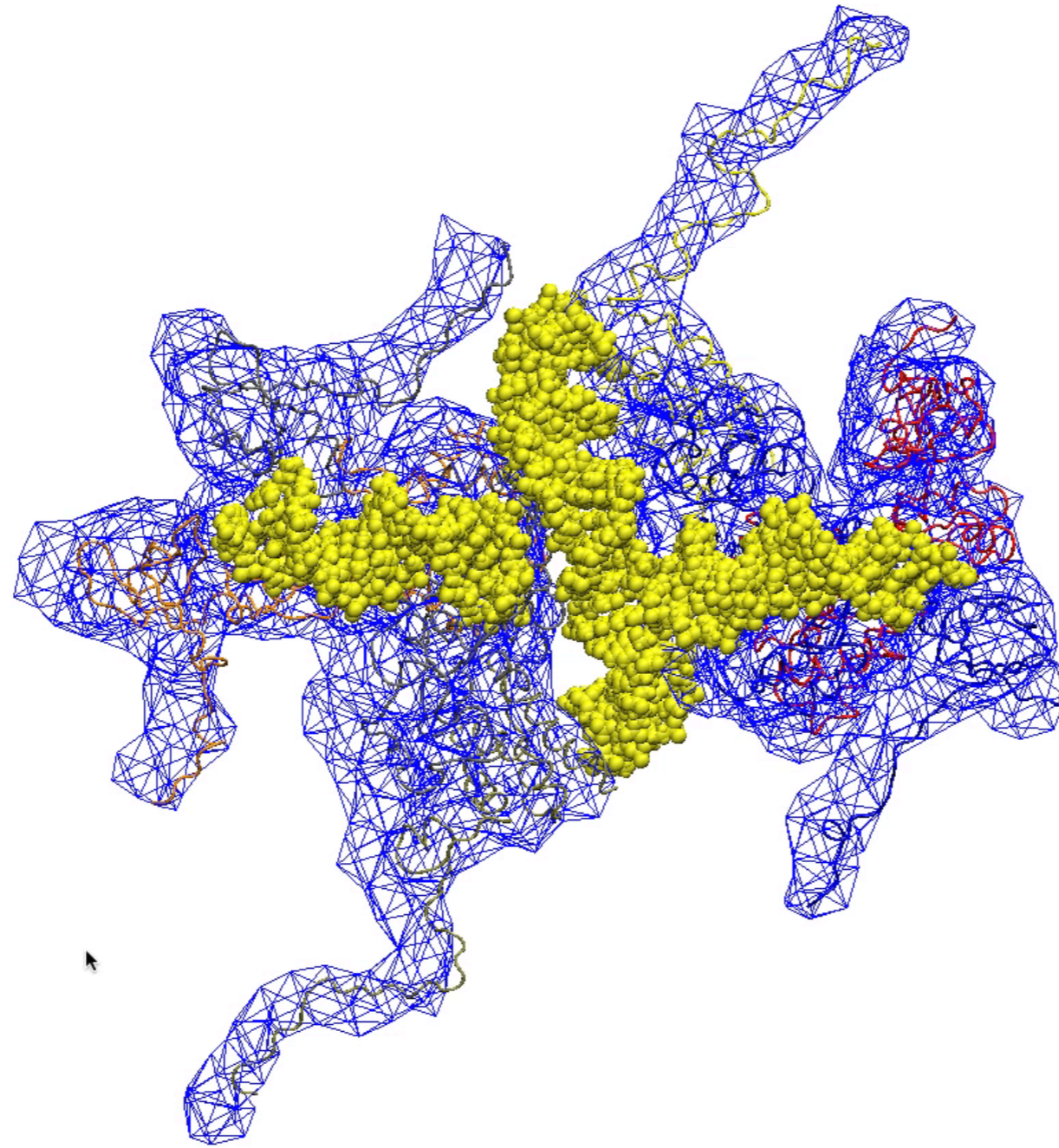
Apply Changes Automatically Apply



1 ns: 8 CPU 13 days



1 ns: 8 CPU 13 days



Options (today)

We don't have ANTON . . . (and even if we did, maybe you don't) . . . and even then 25 us may not be long enough, etc. etc.

For many, yet not all, IDP/ID problems a fully-atomistic simulation is impractical for an average user.

For many, IDP/ID problems, even with massive computational resources, to obtain an answer for a particular starting model is going to take time.

A number of hybrid MD/MC & MC routines exist to sample configuration space yet they are cumbersome to use and not ready for prime time.

Rigid body modeling packages exist (docking, etc.), but they don't handle flexible bits or scattering calculations.

In 2004 there were fewer options

MMC

Monte Carlo samples configurations by generating moves (translational, rotational, etc.), within the molecular potential of your system. These structures taken together make up an ensemble of configurations covering “phase space”

There is an equivalence where by calculated observables, let's say for example the density of a system $\rho = \rho(r)$

$$\langle \rho_i(r) \rangle_{time} = \langle \rho_i(r) \rangle_{phasespace}$$

This means that the “time-average” of a quantity calculated by MD is equivalent to the “ensemble-average” of the same quantity as calculated by Monte Carlo means. *For ergodic systems.*

See: Frenkel & Smit for a nice description from QM to CM.

MMC

Monte Carlo samples configurations by generating moves (translational, rotational, etc.), within the molecular potential of your system. These structures taken together make up an ensemble of configurations covering “phase space”

There is an equilibrium
example the de

let's say for

146

Chapter 6. Molecular Dynamics in Various Ensembles

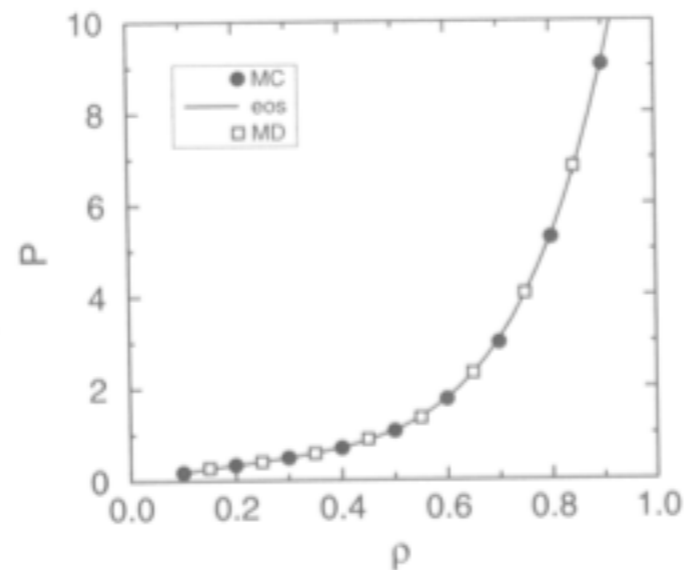


Figure 6.2: Equation of state of the Lennard-Jones fluid ($T = 2.0$ and $N = 108$); comparison of the Molecular Dynamics results using the Andersen thermostat (open symbols) with the results of Monte Carlo simulations (closed symbols) and the equation of state of Johnson *et al.* [62].

This means that the time-average of a quantity calculated by MD is equivalent to the “ensemble-average” of the same quantity as calculated by Monte Carlo means. *For ergodic systems.*

See: Frenkel & Smit for a nice description from QM to CM.

MC

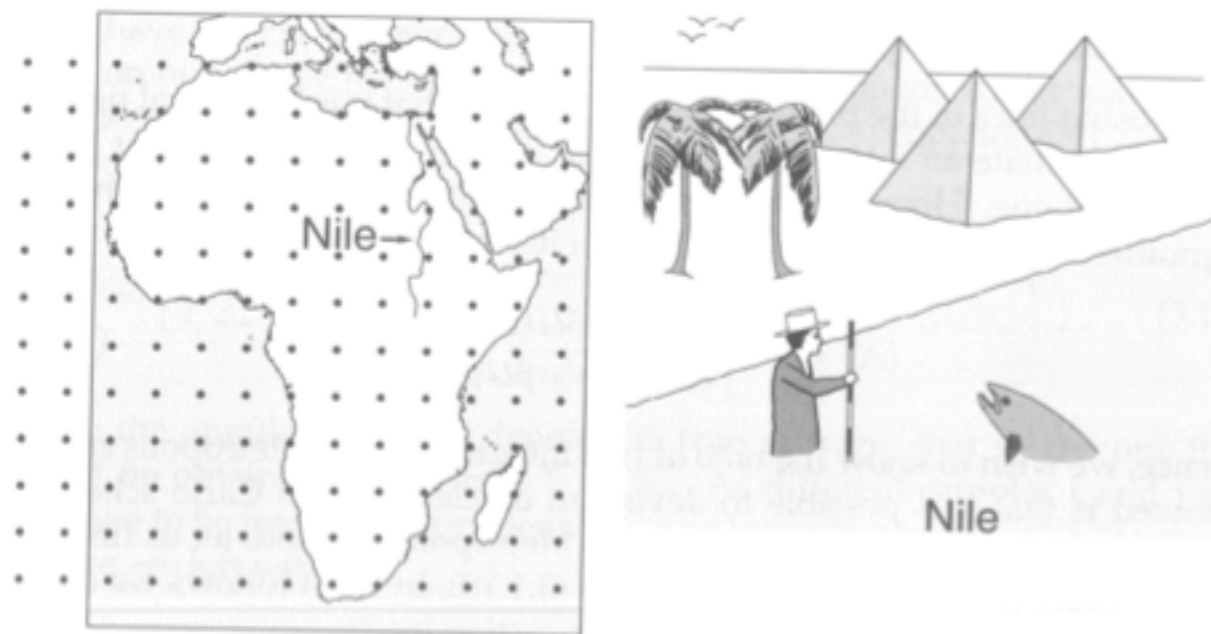


Figure 3.1: Measuring the depth of the Nile: a comparison of conventional quadrature (left), with the Metropolis scheme (right).

Two ways to sample.

(1) Quadrature

You measure the depth of the “Nile” on a equidistant grid. --> get depth

(2) “Metropolis”

a) trial move -->
rejected if you are not in
the water

(b) trial move is
accepted if you are in
the water.

For molecular systems the important part of space (r , θ , ϕ , etc.) is related to the molecular potential.

After every move you measure the depth. The average of all measurements is the depth of the Nile

See: Frenkel & Smit for a nice description of MC (& MD).

MC

For molecules, we generate points in configuration space with a relative probability proportional to the Boltzmann factor. So for a move from “old == o” to “new == n”, the probability that a move is accepted: $\text{acc}(o \rightarrow n)$

$$\frac{\text{acc}(o \rightarrow n)}{\text{acc}(n \rightarrow o)} = \exp(-\beta [U(n) - U(o)])$$

If $U(n) < U(o)$ then the trial move is accepted.

If $U(n) > U(o)$ then the trial move is accepted if $\exp(-\beta [U(n) - U(o)]) < 1$

```
for i in xrange(number_of_steps):
```

```
    x(o) = pick_a_random_atom(random(0,natoms))
```

```
    Eold = energy ( x(o) )
```

```
    Xnew = x(o) + random(0,1)*delta_x
```

```
    Enew = energy( Xnew )
```

```
    If Enew < Eold:
```

```
        x(o) = Xnew
```

```
    elif ( random(0,1) < exp(-beta*(Enew-Eold) )):
```

```
        x(o) = Xnew
```

MC

Size of moves in a Markovian chain:

“A MC sampling scheme is optimal if it yields the lowest statistical error of $\langle A \rangle$ for a given length of CPU time”

In practice, you merely want to cover configuration space.

For an observable A , the mean square error in A should be inversely proportional to the number of uncorrelated configurations is given amount of CPU time.

The point is that there is no golden rule.

Ad-hoc definitions mean, to me, that I should use common sense.

Each problem type needs to be optimized.

MMC (with rotation)

So, to have a fully-atomistic MMC (vary all bonds and angles) our averages and hence our energy calculations become numerically much more complicated. **For large flexible systems, fully-atomistic MMC is slower than (or as slow as) MD!**

$$\langle A \rangle = \frac{\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\beta H(\mathbf{p}^N, \mathbf{r}^N)] A(\mathbf{p}^N, \mathbf{q}^N)}{\int d\mathbf{p}^N d\mathbf{r}^N \exp[-\beta H(\mathbf{p}^N, \mathbf{r}^N)]}$$

$$\text{where } H(\mathbf{p}^N, \mathbf{r}^N) = \sum_j p_j^2 / (2m_j) + U(\mathbf{r}^n)$$

but for rotational moves (now in generalized coordinates);

$$H(\mathbf{p}^N, \mathbf{r}^N) = \frac{1}{2} \mathbf{p} \cdot G^{-1} \cdot \mathbf{p} + U(\mathbf{r}^n)$$

when placed into our expression for $\langle A \rangle$ leads to $|G|^{\frac{1}{2}}$ in the numerator.

MMC with frozen dof

An alternative strategy is to only sample the configurational landscape accessible by rotating the dihedral bonds of macromolecules.

Use the CHARMM force-field to sample only the dihedral potential using the Metropolis criterion, then if you have interesting result, you minimize the entire ensemble of structures.

Sampling on a grid (40 degree spacing for HIV-1 Gag) --> 10^{67} configurations are required.

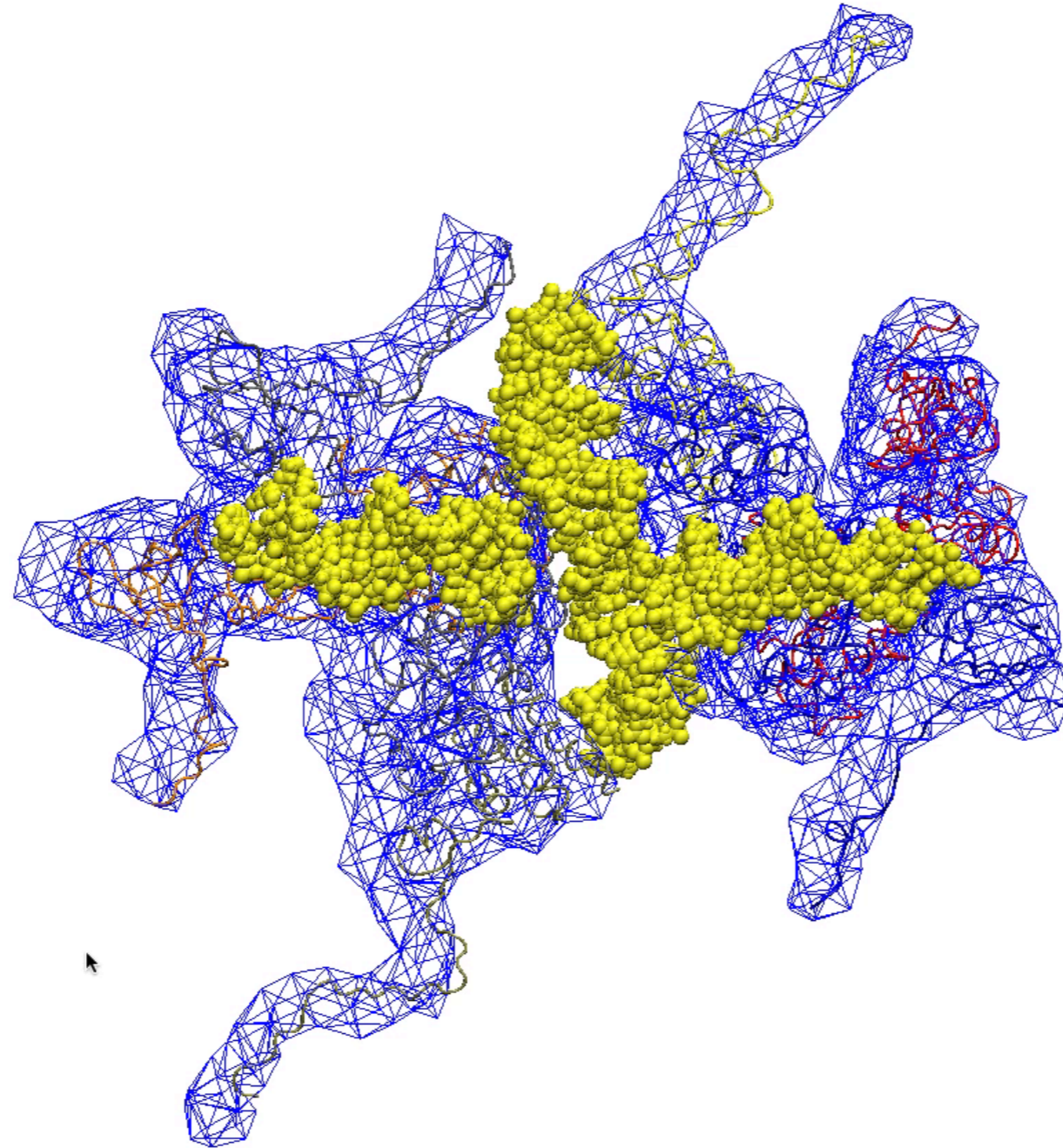
Sampling from Ramachandran plot is fundamentally limiting.

Advantages:

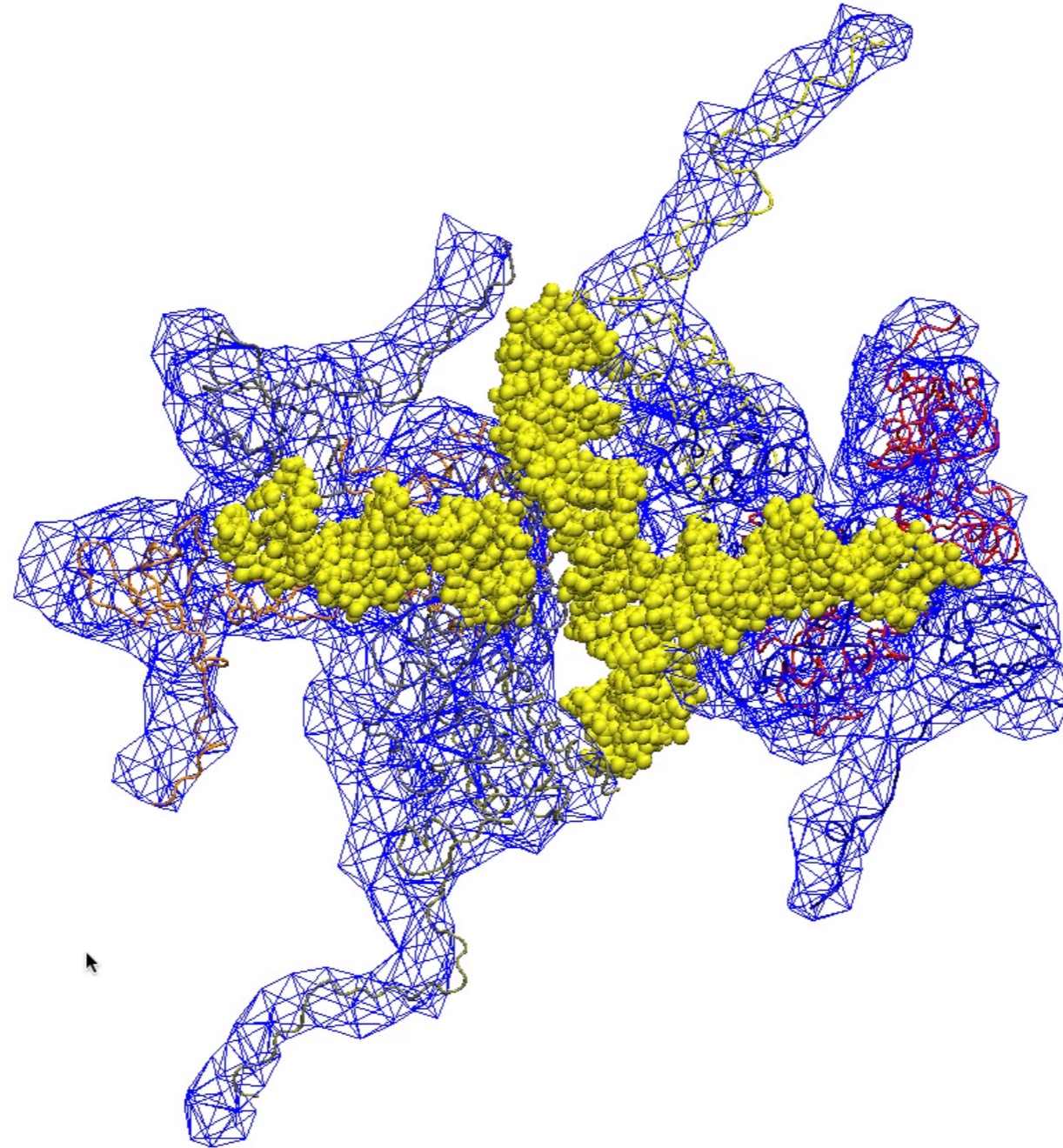
Generate configurations quickly to test hypotheses.

Structures are minimized and ready for further analyses IF YOU WISH.

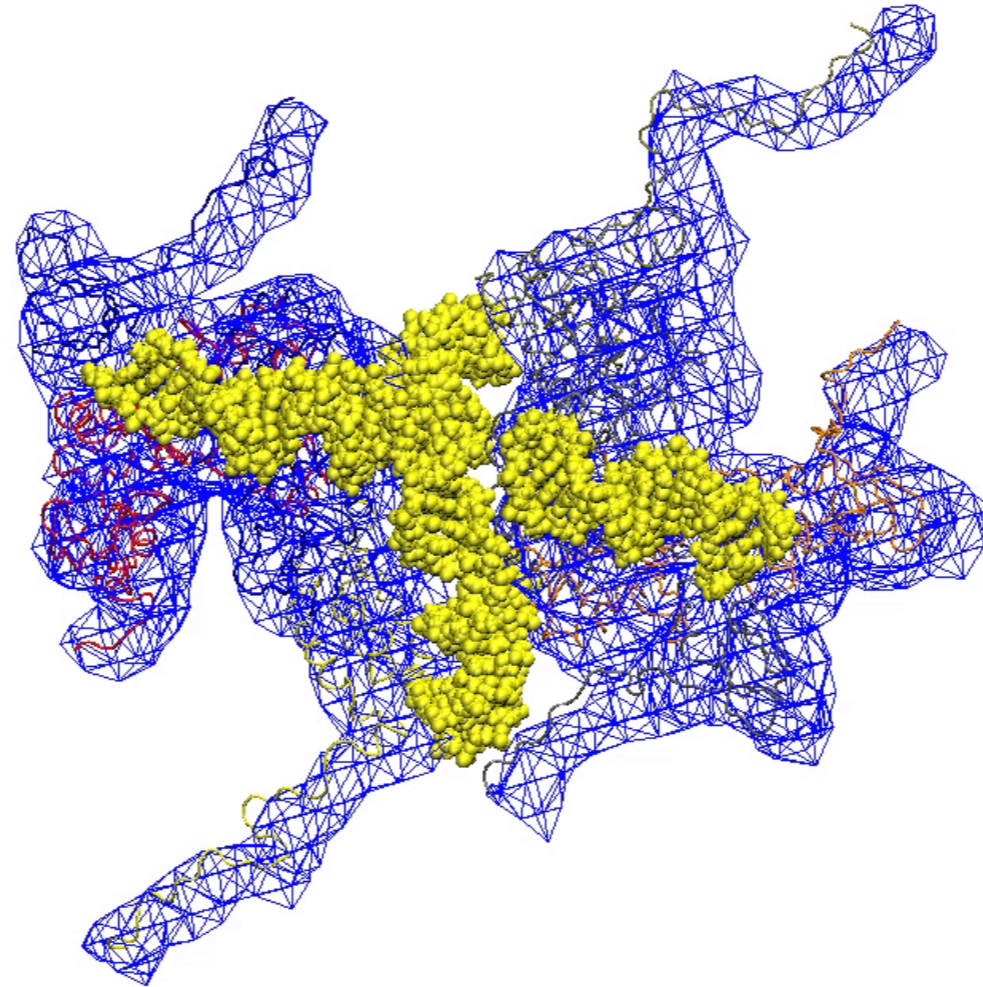
1 ns: 8 CPU 13 days



1 ns: 8 CPU 13 days



15 min. then 45 min. : 1 CPU



backbone torsion sampling

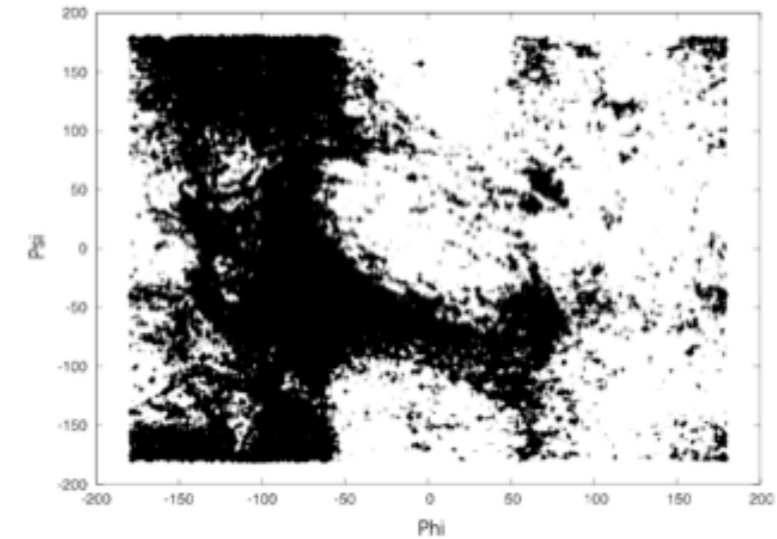
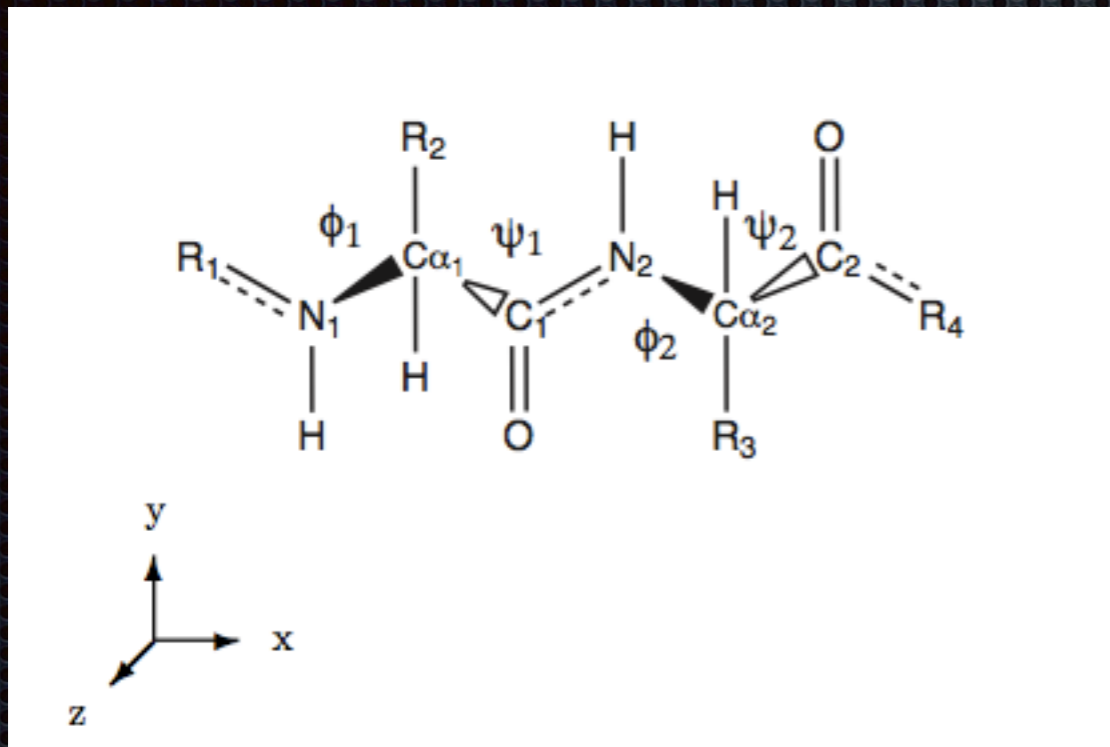


Fig. 5. The values of accepted angular values of ϕ and ψ for the 65 amino-acids located in the five flexible regions of the 10^5 energy minimized full-length HIV-1 Gag protein structures. Angles are shown in degrees.

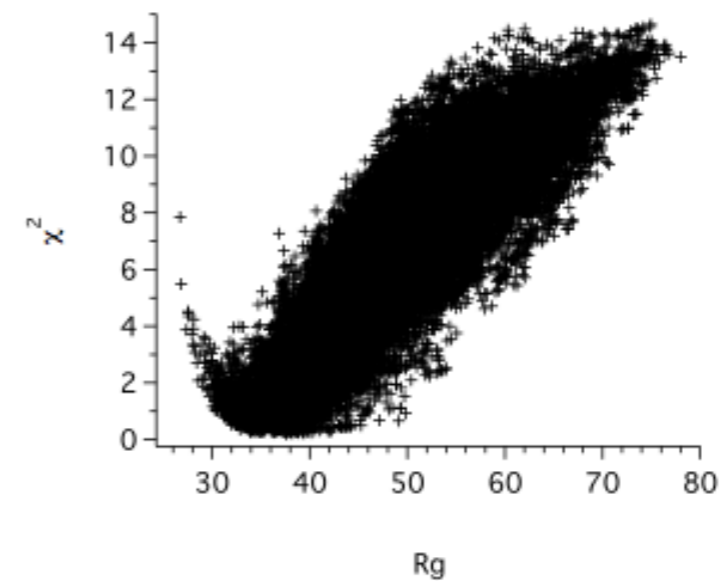
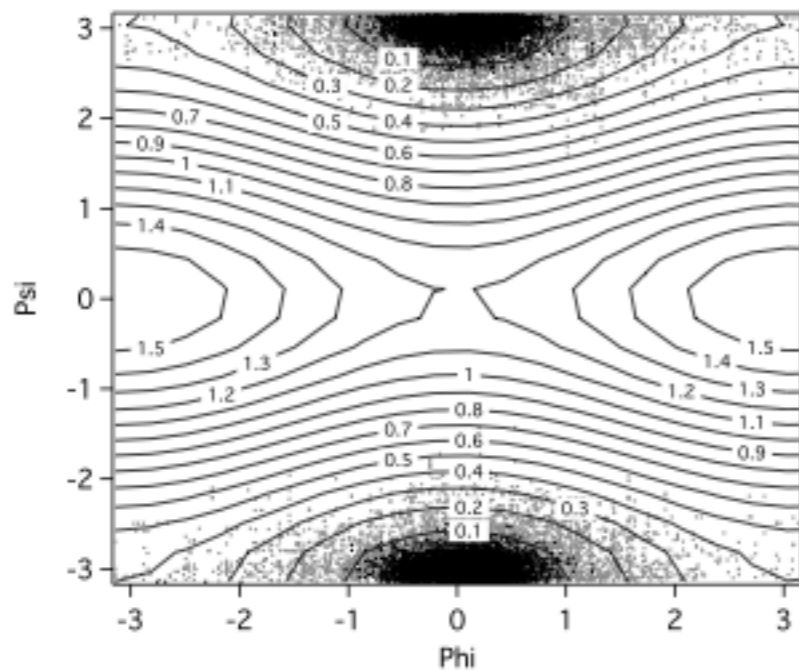
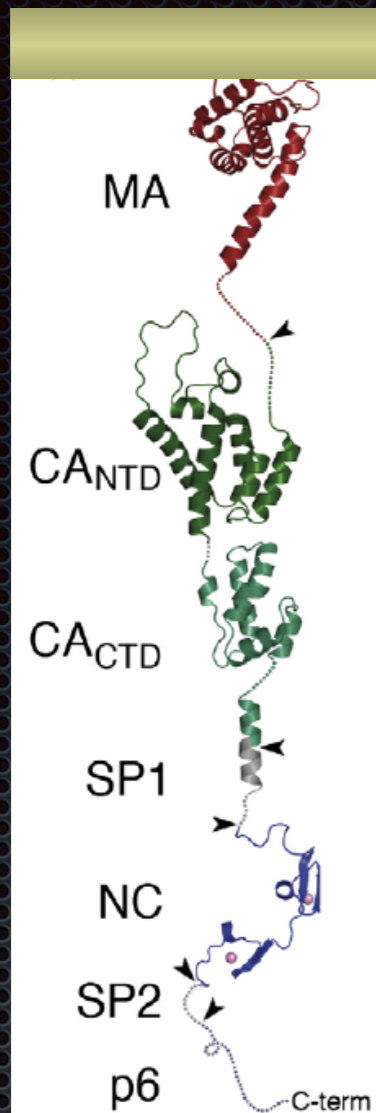


Fig. 6. A plot of χ^2 versus R_g obtained from 10^5 non-overlapping full-length HIV-1 Gag protein structures.

representative envelope/st.



123-144 (21)

278-283 (5)

354-378 (24)

378-389 (11)

408-412 (4)

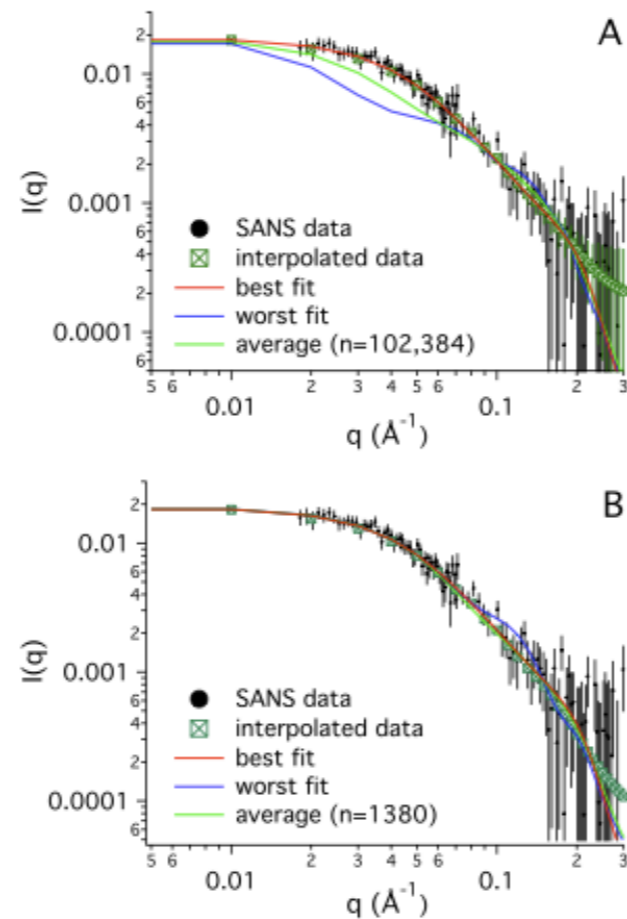


Fig. 7. The plot of experimental and theoretical SANS profiles obtained from 10^5 non-overlapping full-length HIV-1 Gag protein structures.

shown in Figure 6. Here, χ^2 is calculated using

$$\chi^2 = \frac{1}{(N-1)} \frac{\sum_Q (I_{exp}(Q) - I_{calculated}(Q))^2}{\sigma_{exp}(Q)^2} \quad (7)$$

where $I_{exp}(Q)$ is the experimentally determined SANS profile, $I_{calculated}(Q)$ is the value obtained using SASSIE, $\sigma_{exp}(Q)$ is the experimentally determined Q -dependent variance, and the sum was taken over $N = 19$ grid points of momentum transfer, Q . A plot of the best, worst, average, and experimental SANS profiles is shown in Figure 7.

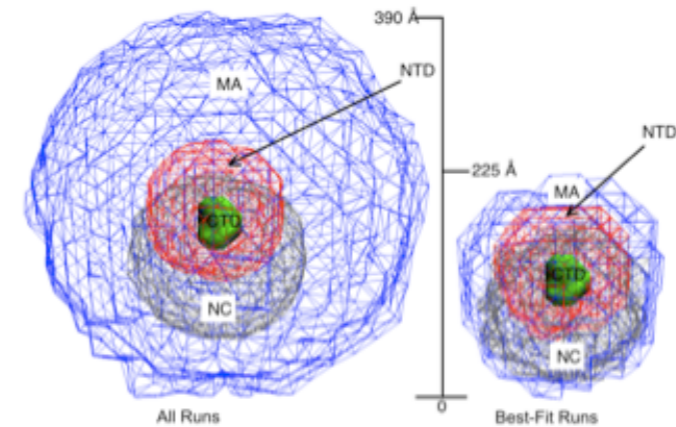


Fig. 8. Iso-density plots of all non-overlapping HIV-1 Gag structures (left: "All Runs") and structures with $\chi^2 \approx 1.0$ (right: "Best-Fit Runs"). The different regions of Gag are colored for clarity (MA:blue, NTD:red, CTD:green, and NC:grey). Structures used to generate the plots were aligned on the alpha-carbon atoms of the CTD region.

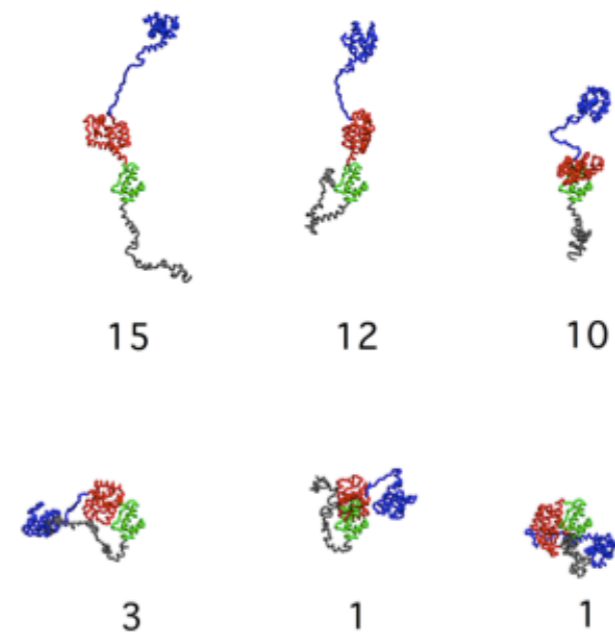


Fig. 9. Tube representations of example HIV-1 Gag structures with χ^2 values shown below each structure. The different regions of Gag are colored for clarity (MA:blue, NTD:red, CTD:green, and NC:grey). Structures are aligned on the alpha-carbon atoms of the CTD region.

sassie-web

<https://sassie-web.chem.utk.edu/sassie2>

SASSIE-web : Beta

Login Help on 

- Tools 
- Build 
- Interact 
- Simulate 
- Calculate 
- Analyze 

FEEDBACK

DOCS

sassie-web -- **Build**

Software to clean-up structure files

and

generate force field topology files

Build – evaluate and clean up structure files

PDB Scan Generates a report that characterizes the user supplied PDB file (alpha).

PDB Rx Attempts to correct mistakes in user supplied PDB file (alpha).

CG Builder Tools to assist the generation of coarse-grain structures (alpha).

sassie-web -- Simulate

Monomer Monte Carlo: single chain protein or ss-NA (think IDP / ID-NA)

Complex Monte Carlo: multiple single chain protein or ss-NA (think IDP / ID-NA complexes)

Energy Minimization: anything FF supports (NAMD)

Torsion Angle MD: protein/NA/Carbs. FF supports (CHARMM)

Docking: anything FF supports

Two-Body Grid: anything FF supports

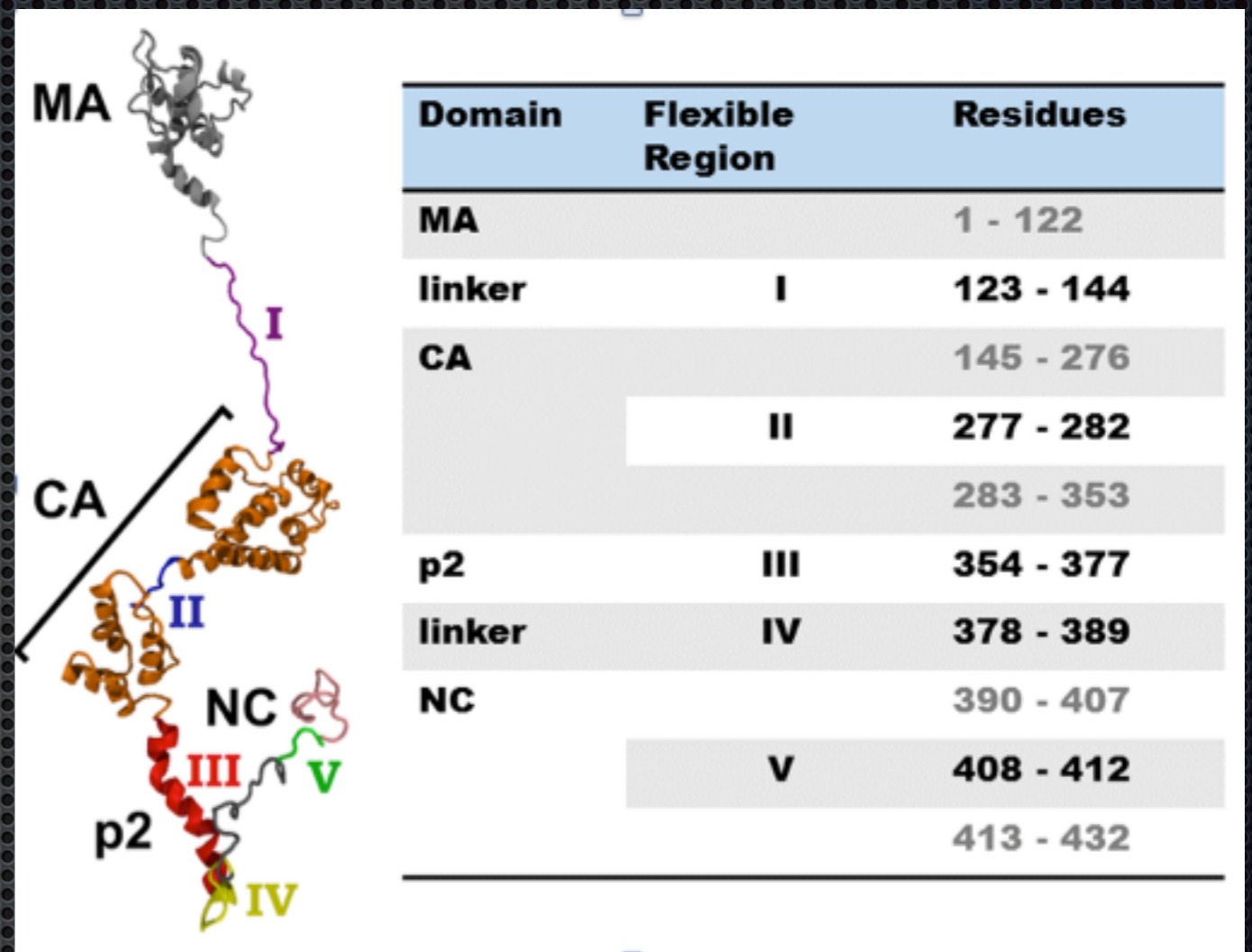
Rigid body

Monomer Monte Carlo

Single chain IDP / ID-NA
simple N → C or 5' – 3'
backbone torsion sampling

“What if” questions and
ensembles for further analysis

Configurations are correlated
Not usable with internal loops
Not usable for ds NA moves



Generate
Ensemble

Energy Minimize
&
Calculate I(q)

Compare to Exp.:
Chi-square Filter
Density Plot

Monomer Monte Carlo II

Monomer Monte Carlo

run name

reference pdb 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

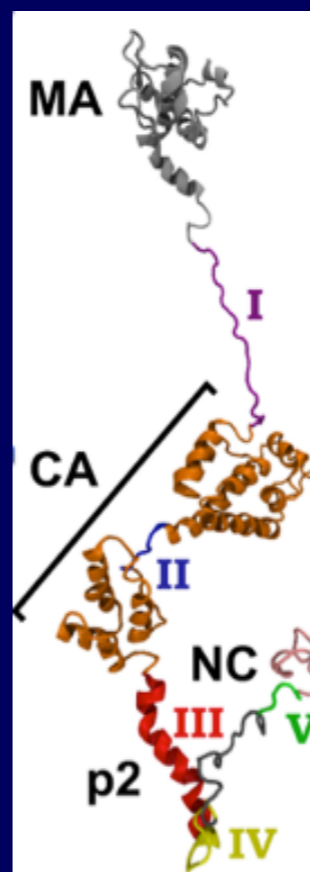
structure alignment: high residue

overlap basis

OR Local: hiv1_gag.pdb

Advanced Input

Check Box for Advanced Input



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 III

Rg filter(s)

(eg. $55 < Rg < 60$)



low Rg cutoff

55

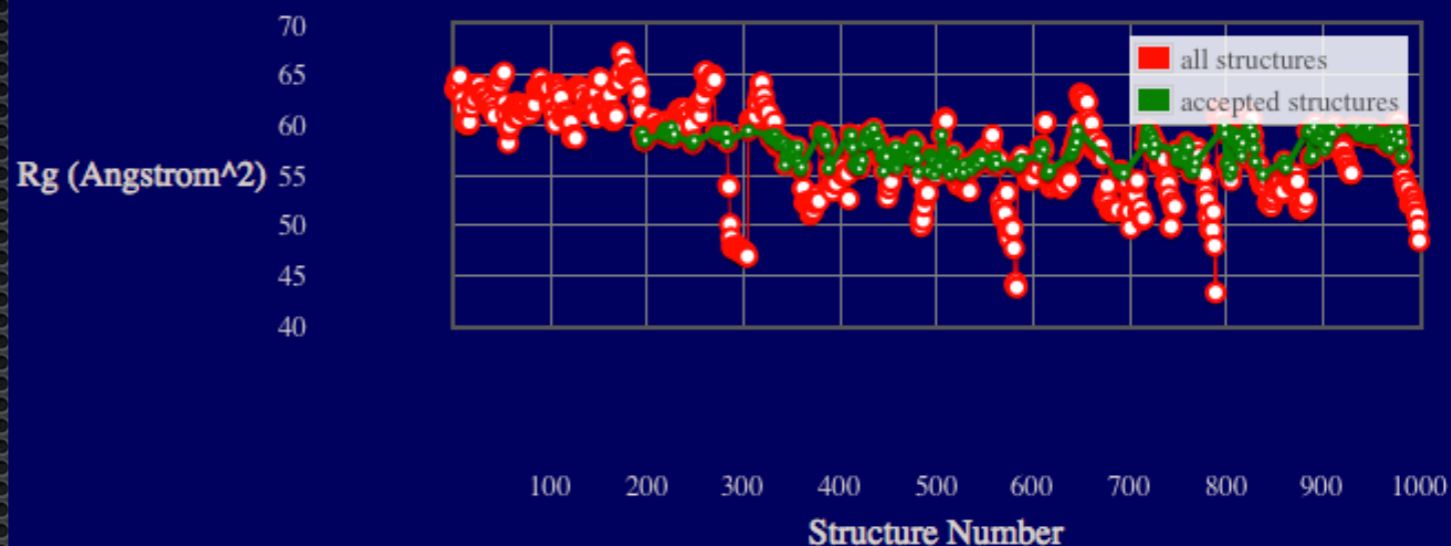
high Rg cutoff

60

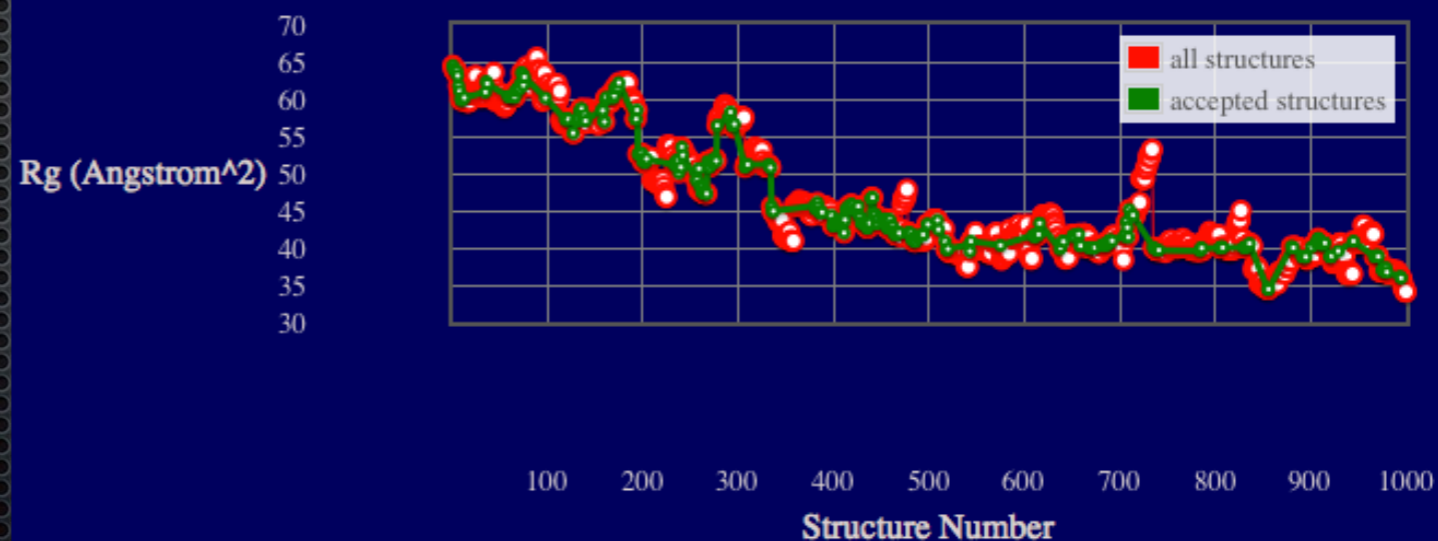
Directed Monte Carlo
(bias to an Rg value)

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

Rg Results



Rg Results



Monomer Monte Carlo IV

check box to use Z coordinate filter



Z filter

Z cutoff (angstroms) 0.0

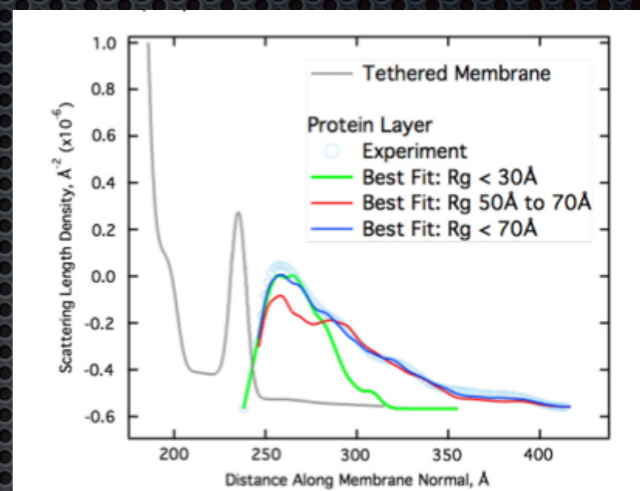
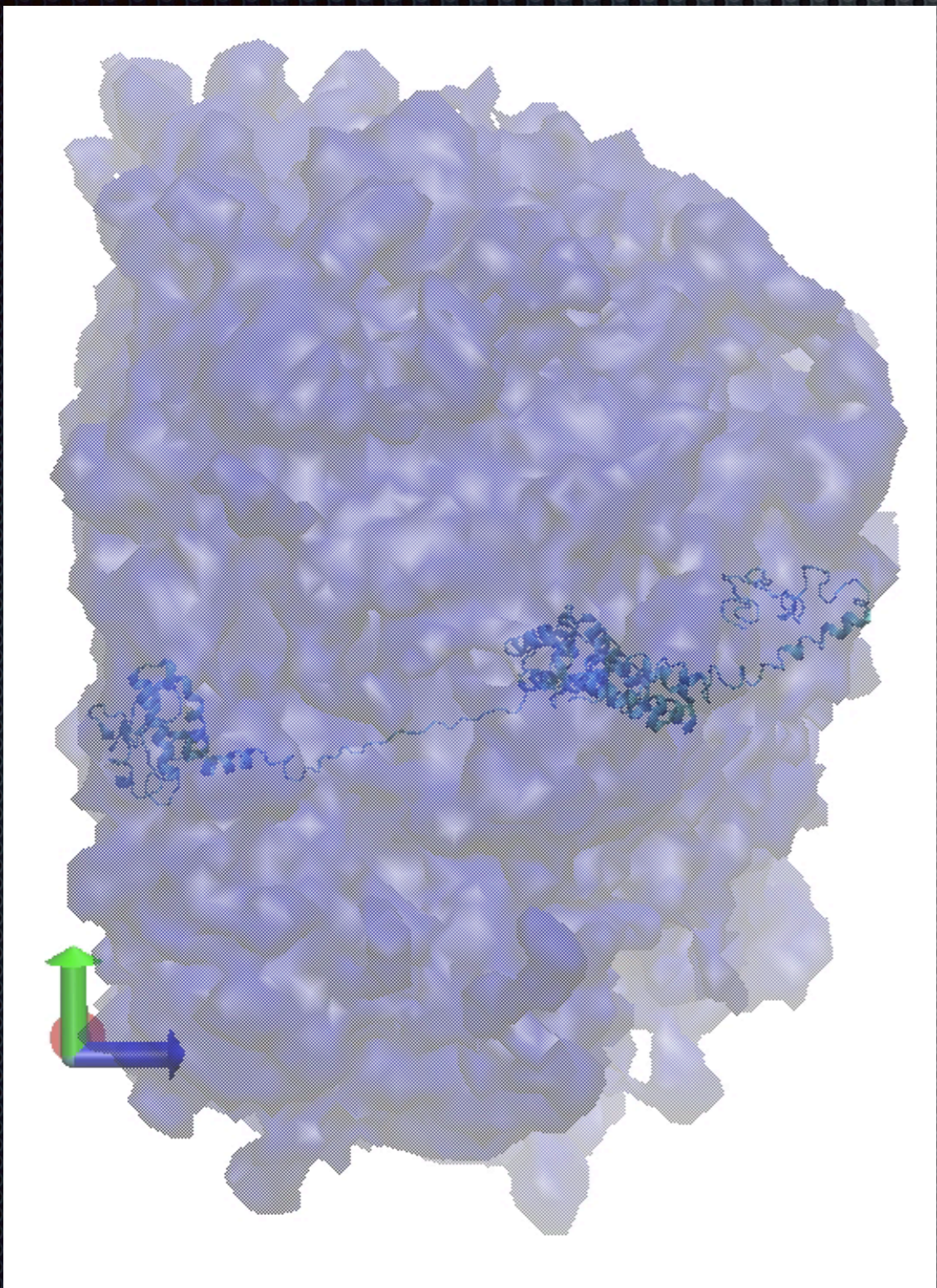


Fig. 2. Molecular modeling of Gag protein structure on the membrane using SLDMOL to fit experimentally derived SLD profiles. The experimentally derived SLD profiles for Gag, light blue circles, are shown in the compact conformational state, panel A, and the extended state, panel B. Using the SASSIE software package 100,000 independent conformations of the Gag protein were generated and subsequently filtered based on R_g values. An average SLD profile using a Monte-Carlo optimization to weight the structures provided a best fit for each R_g range and can be seen as solid lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Monomer Monte Carlo IV

check box to use Z coordinate filter



Z filter

Z cutoff (angstroms) 0.0

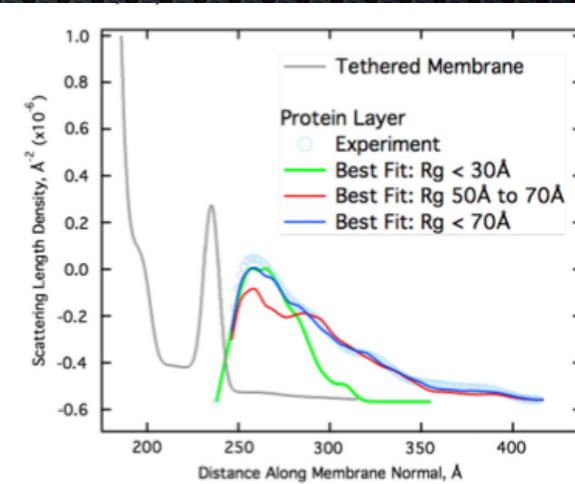
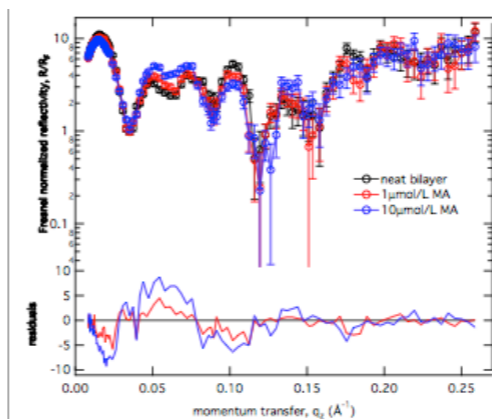
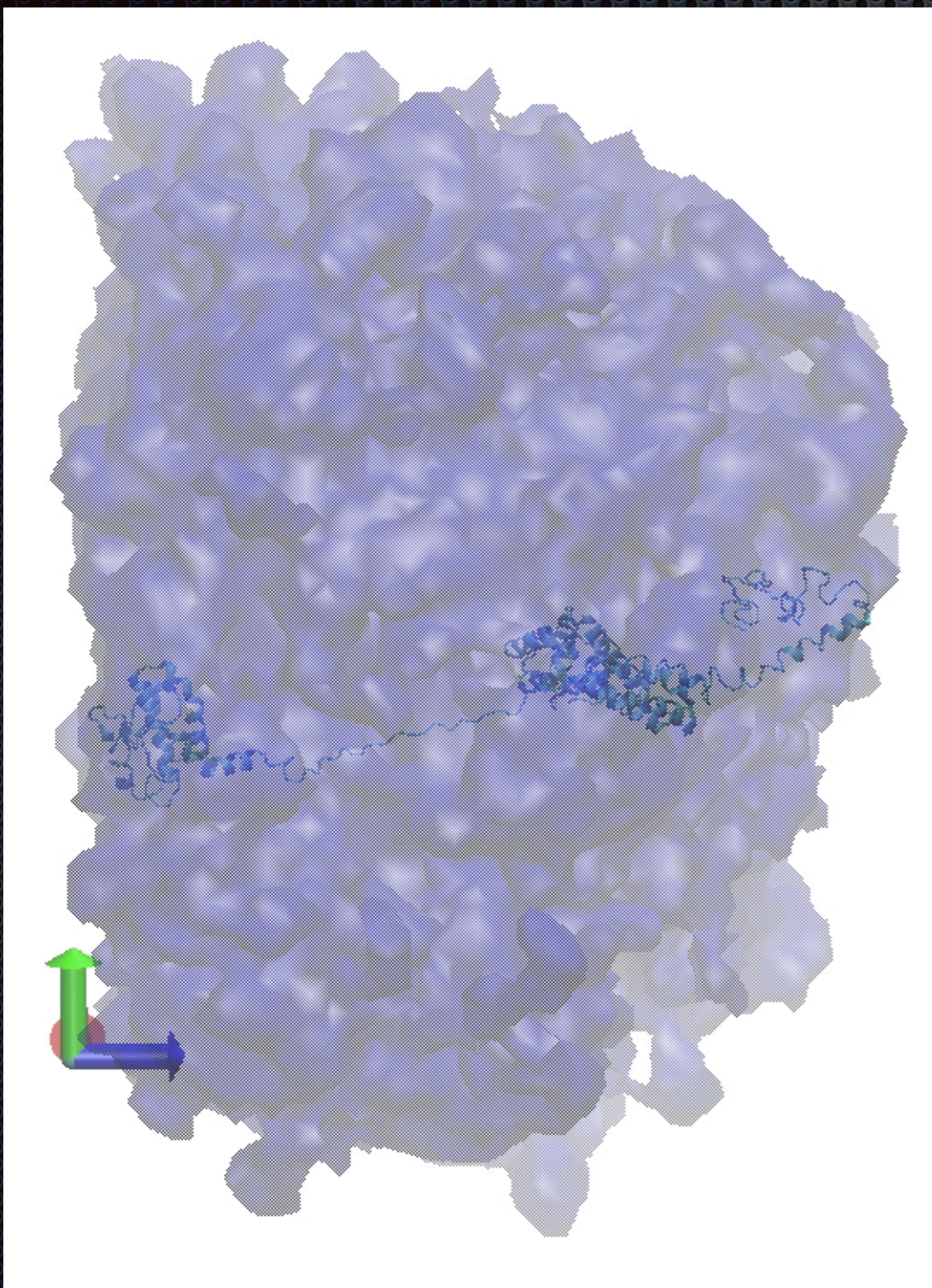
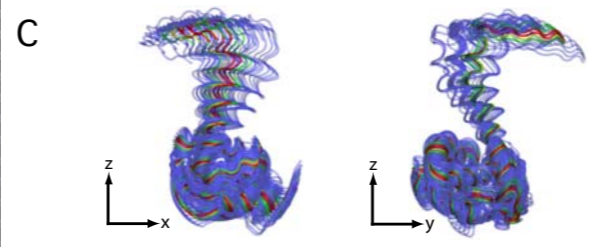
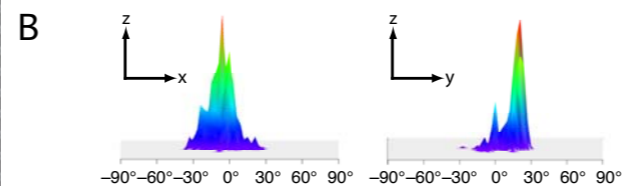
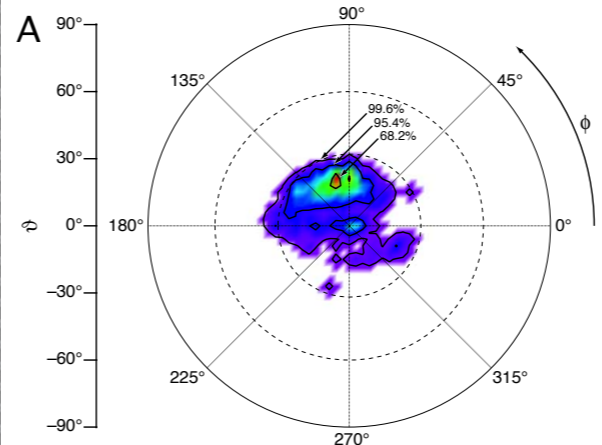


Fig. 2. Molecular modeling of Gag protein structure on the membrane using SLDMOL to fit experimentally derived SLD profiles. The experimentally derived SLD profiles for Gag, light blue circles, are shown in the compact conformational state, panel A, and the extended state, panel B. Using the SASSIE software package 100,000 independent conformations of the Gag protein were generated and subsequently filtered based on R_g values. An average SLD profile using a Monte-Carlo optimization to weight the structures provided a best fit for each R_g range and can be seen as solid lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Monomer Monte Carlo IV

check box to use Z coordinate filter

Z filter

Z cutoff (angstroms)

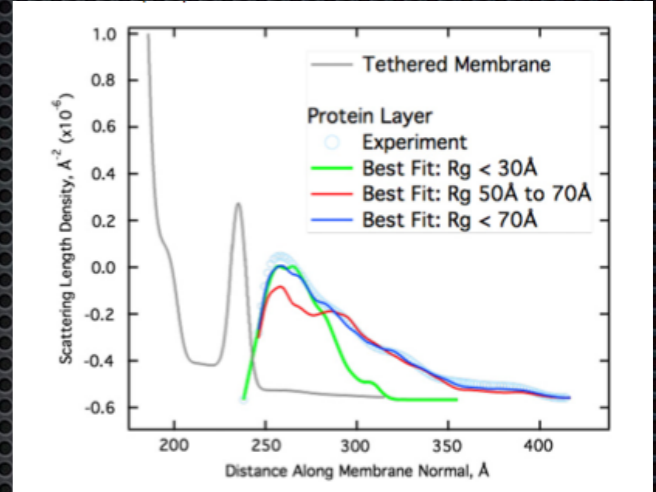
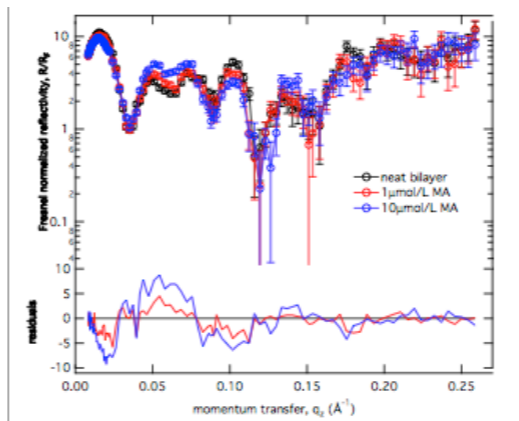
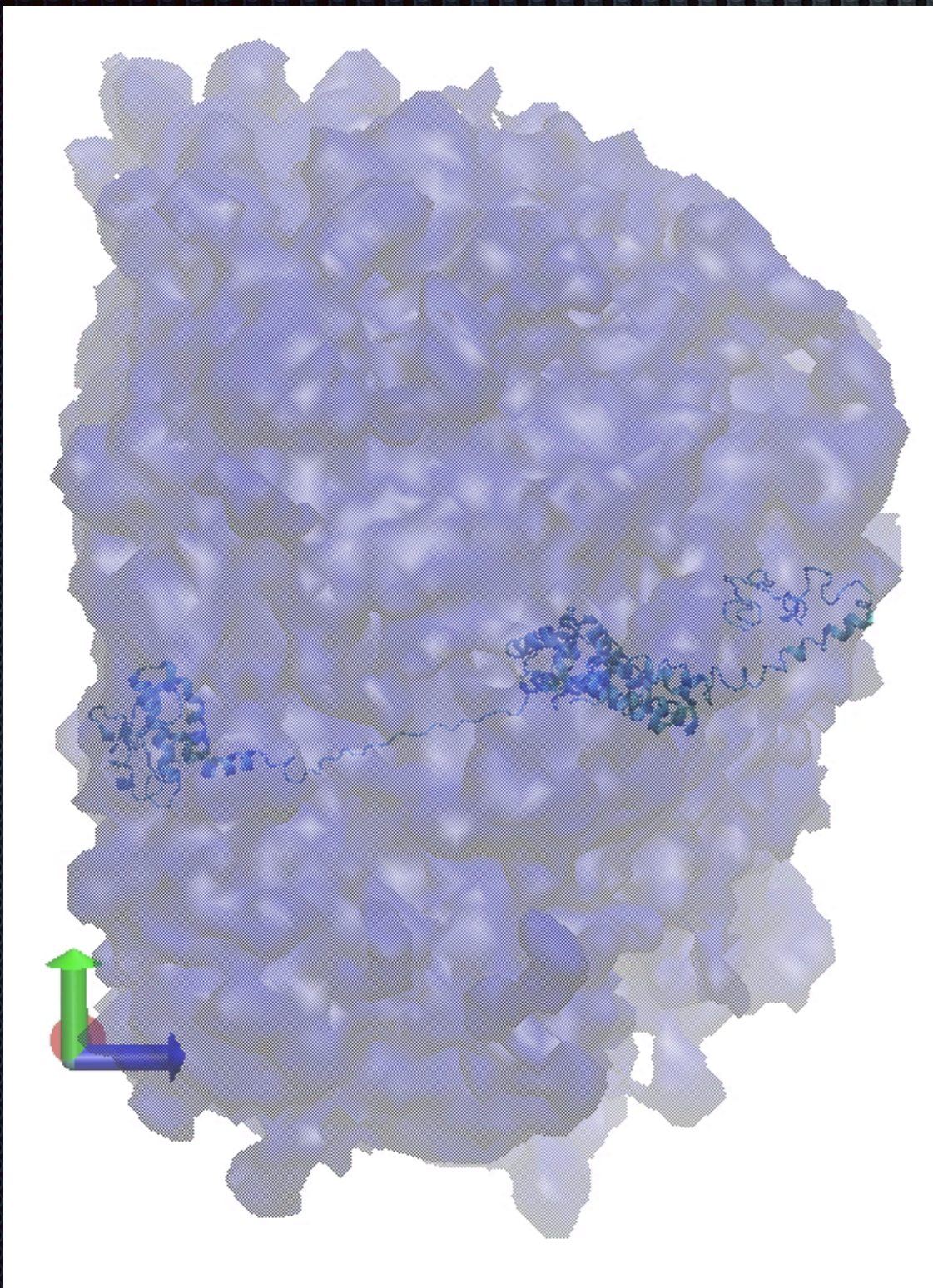
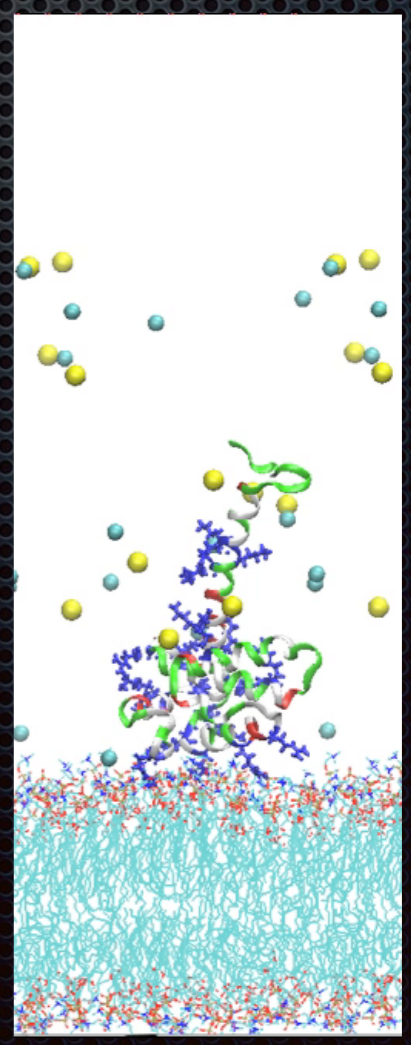
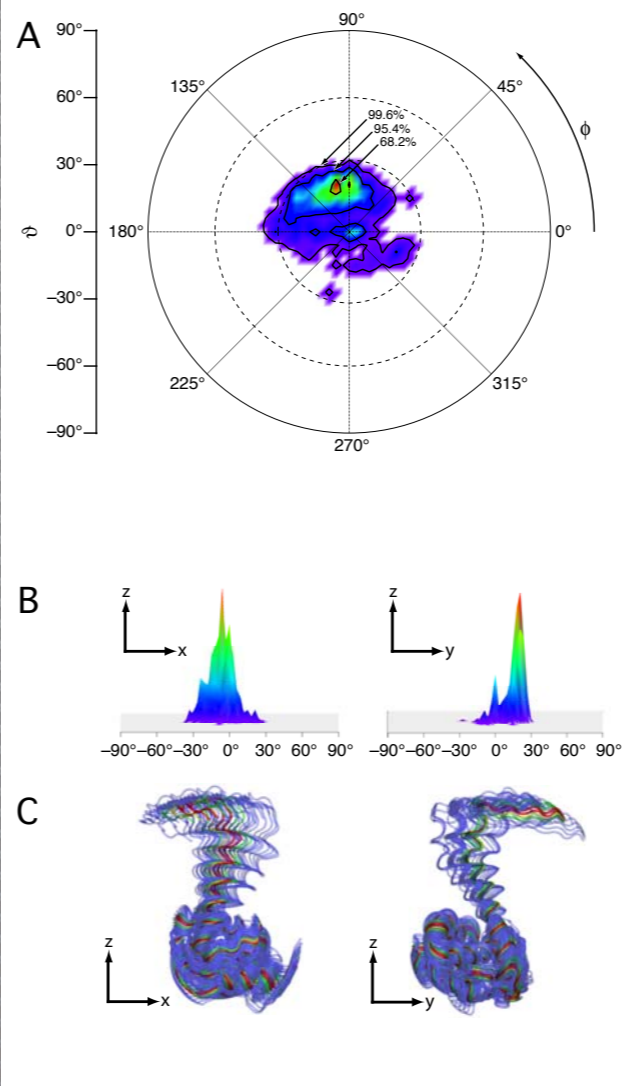


Fig. 2. Molecular modeling of Gag protein structure on the membrane using SLDMOL to fit experimentally derived SLD profiles. The experimentally derived SLD profiles for Gag, light blue circles, are shown in the compact conformational state, panel A, and the extended state, panel B. Using the SASSIE software package 100,000 independent conformations of the Gag protein were generated and subsequently filtered based on Rg values. An average SLD profile using a Monte-Carlo optimization to weight the structures provided a best fit for each Rg range and can be seen as solid lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Monomer Monte Carlo IV

check box to use Z coordinate filter

Z filter

Z cutoff (angstroms)

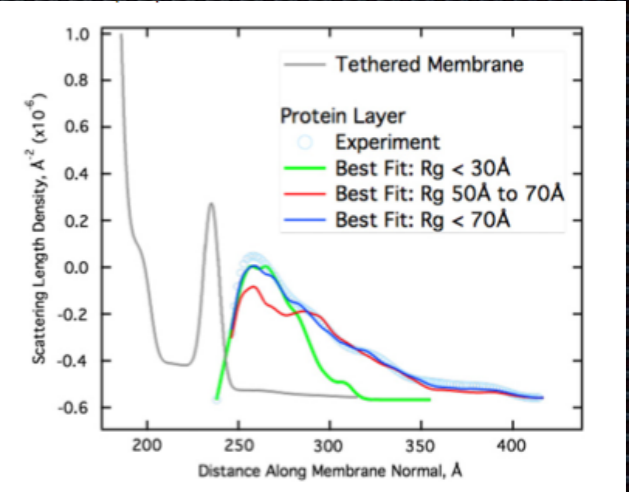
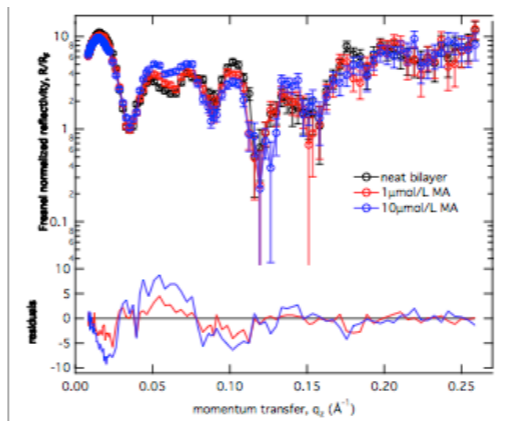
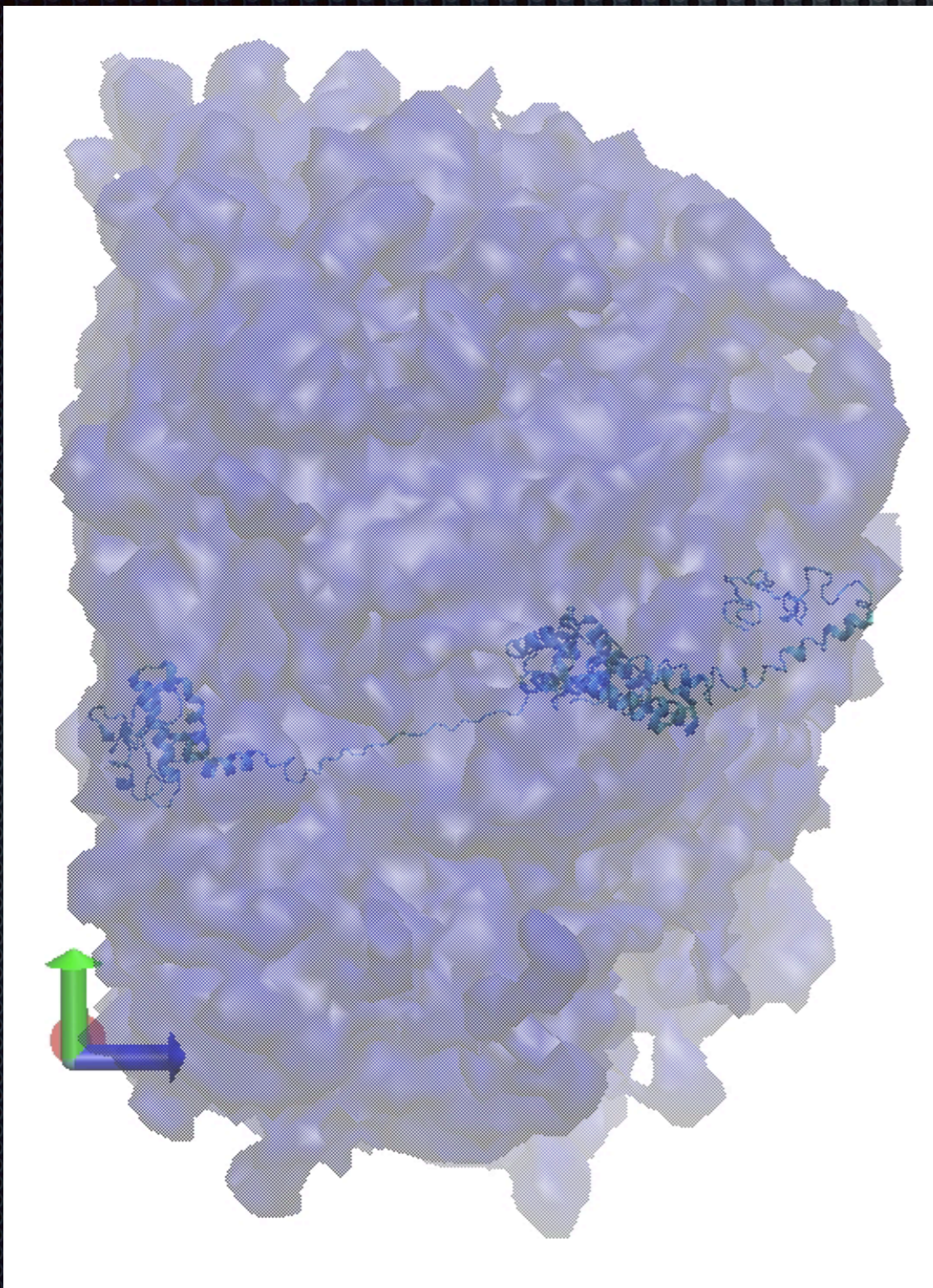
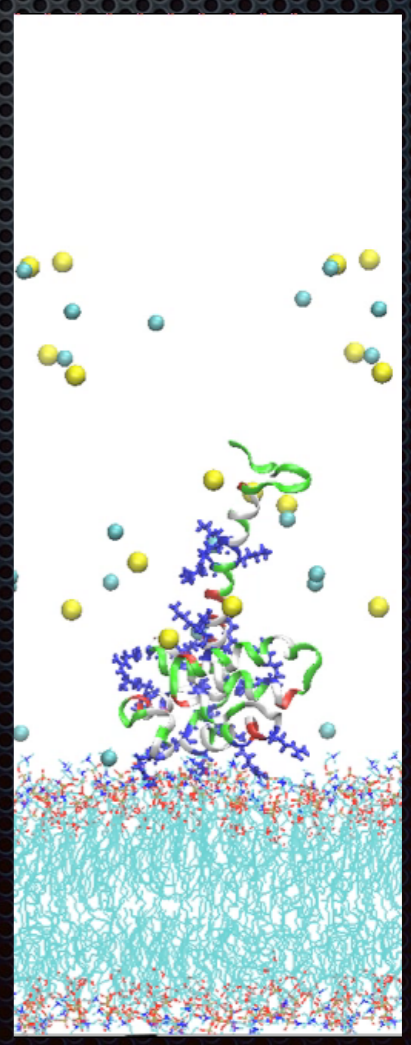
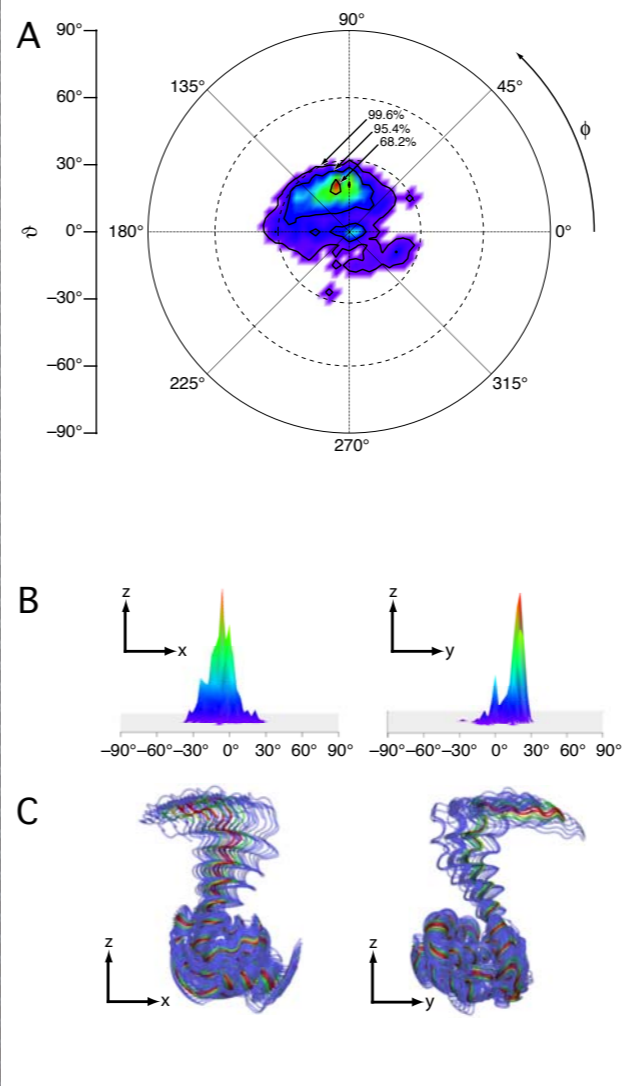
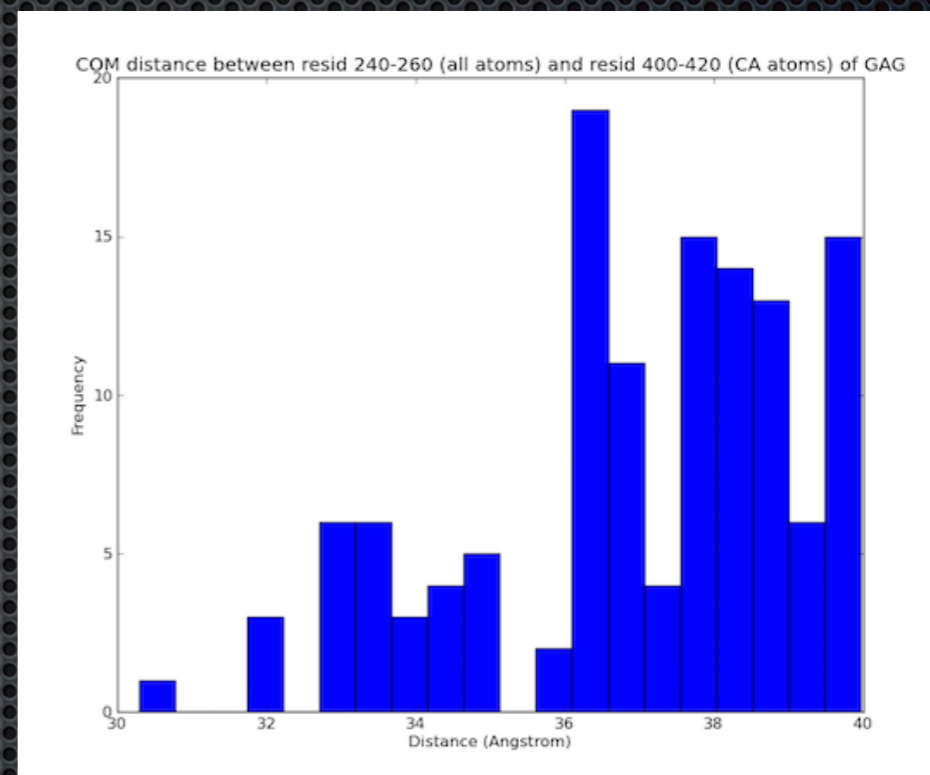
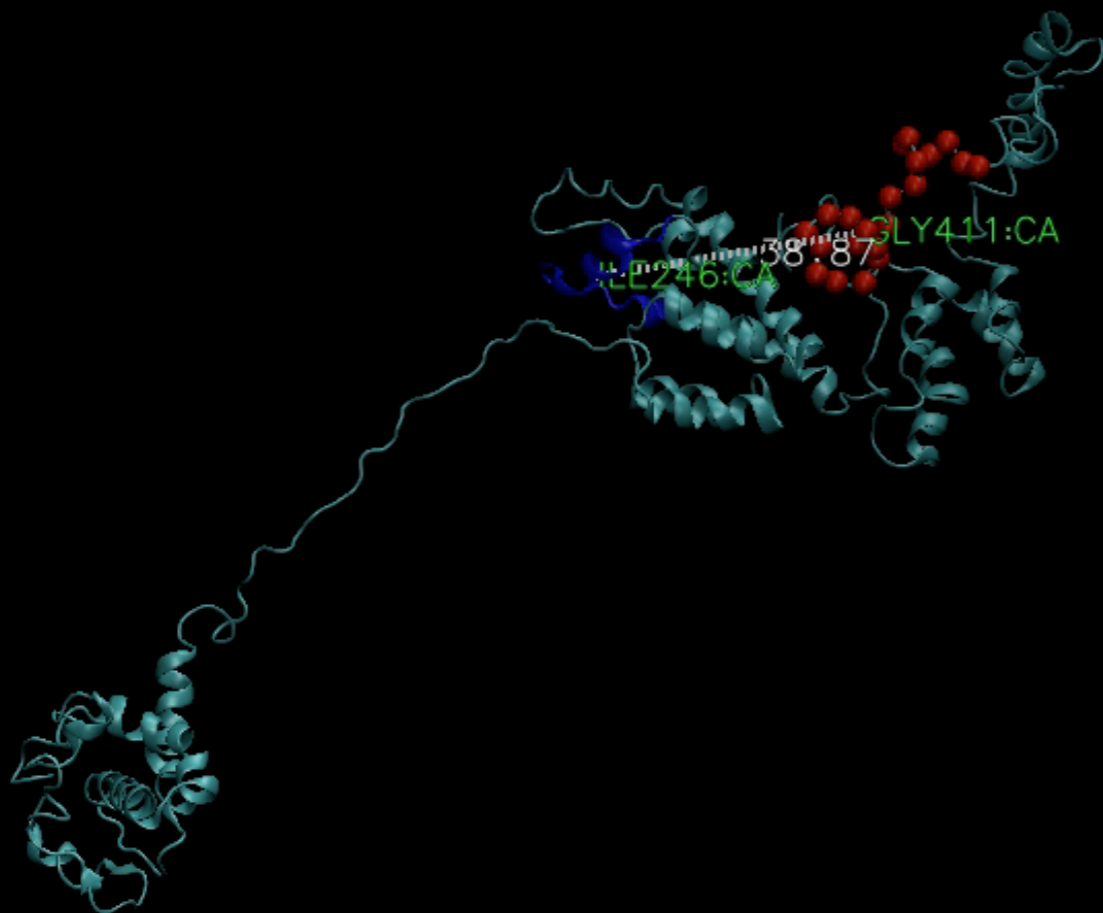


Fig. 2. Molecular modeling of Gag protein structure on the membrane using SLDMOL to fit experimentally derived SLD profiles. The experimentally derived SLD profiles for Gag, light blue circles, are shown in the compact conformational state, panel A, and the extended state, panel B. Using the SASSIE software package 100,000 independent conformations of the Gag protein were generated and subsequently filtered based on Rg values. An average SLD profile using a Monte-Carlo optimization to weight the structures provided a best fit for each Rg range and can be seen as solid lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



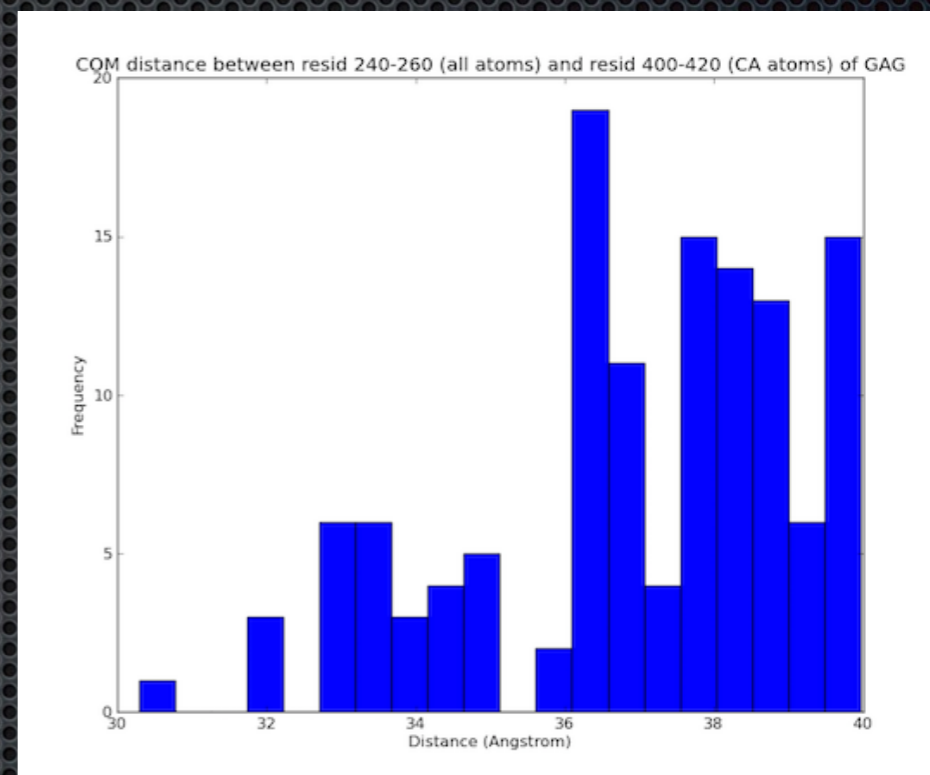
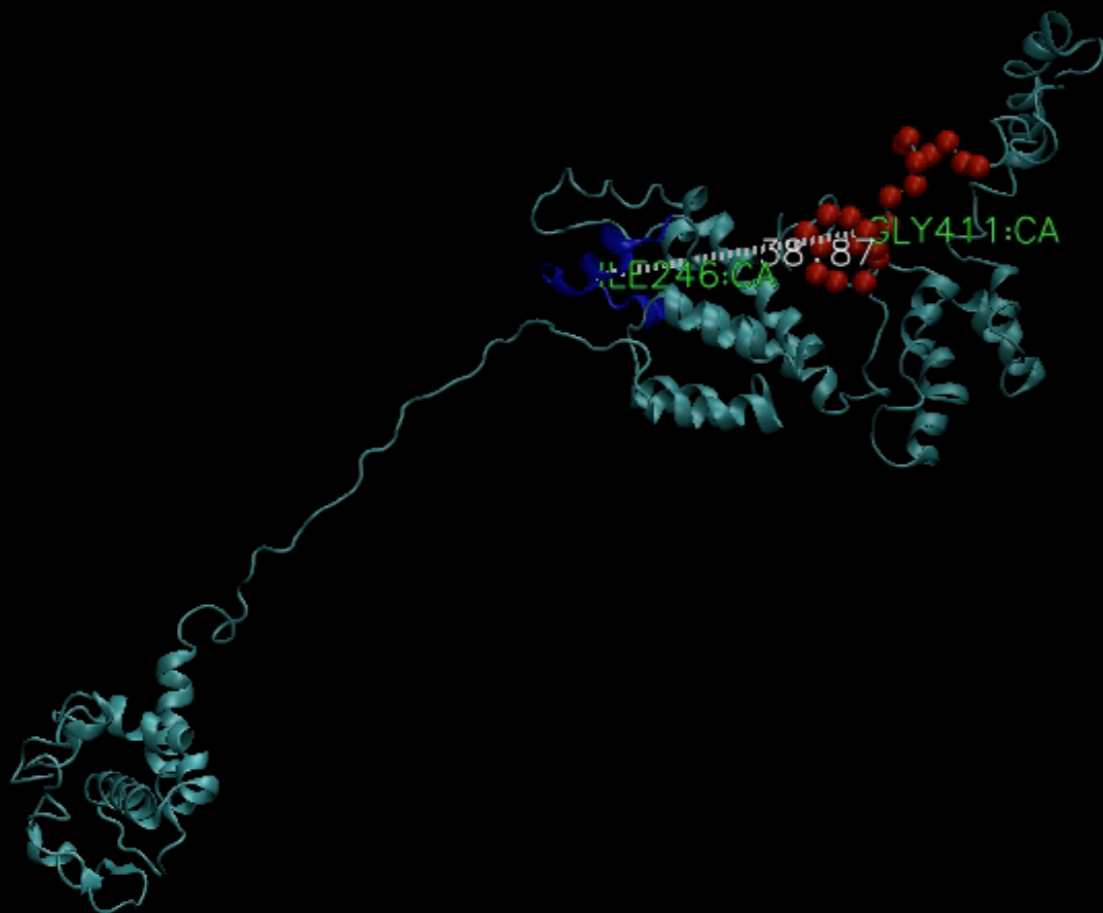
Monomer Monte Carlo V



constraint file name constraints.txt OR Local: constraints.txt

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

Monomer Monte Carlo V



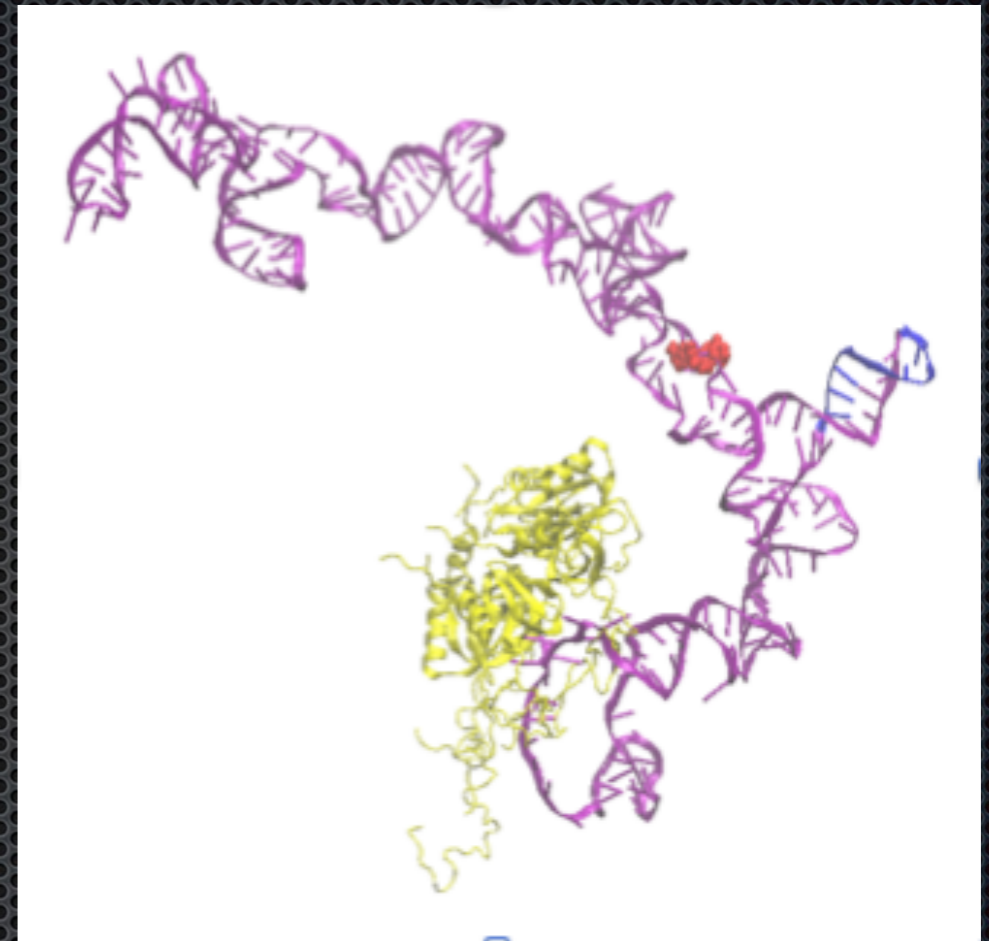
constraint file name constraints.txt OR Local: constraints.txt

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

Complex Monte Carlo

Multiple, single chain IDP / ID-NA
simple N \rightarrow C or 5' – 3'
backbone torsion sampling

“What if” questions and
ensembles for further analysis
Not all chains need to be flexible
(ds DNA etc. can be present)
Configurations are correlated
Not usable with internal loops
Not usable for ds NA moves



Generate
Ensemble

Energy Minimize
&
Calculate $I(q)$

Compare to Exp.:
Chi-square Filter
Density Plot

sassie-web -- Complex MC II

Rg filter(s)

Monte Carlo

Z-filter

Constraints etc.

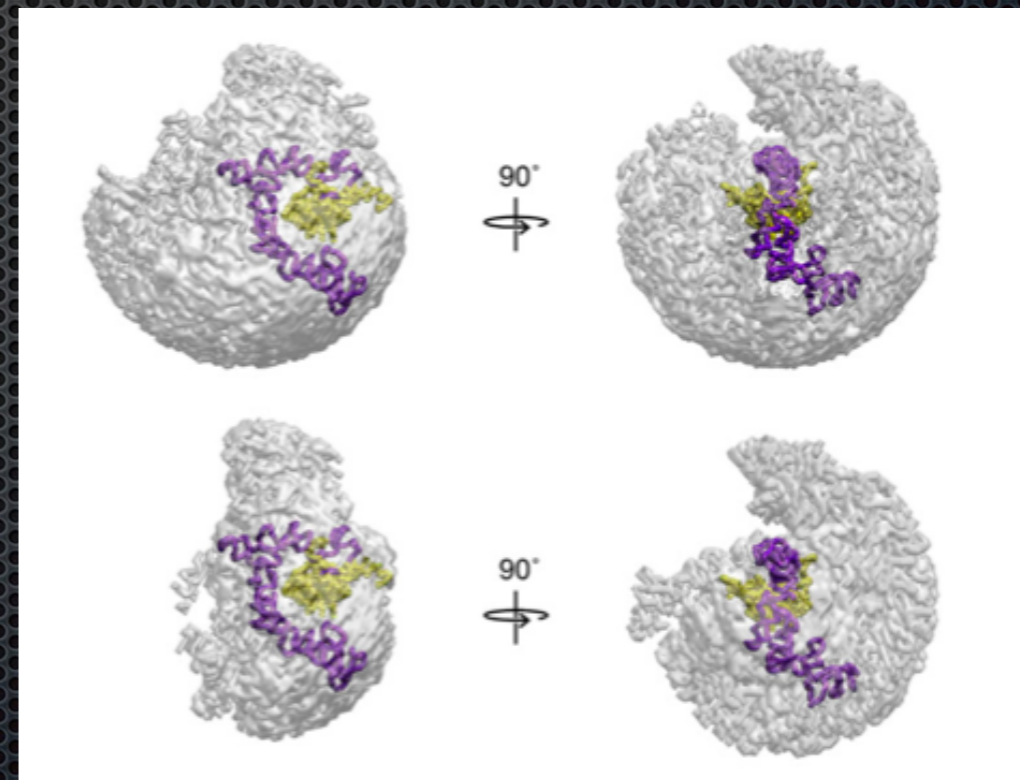
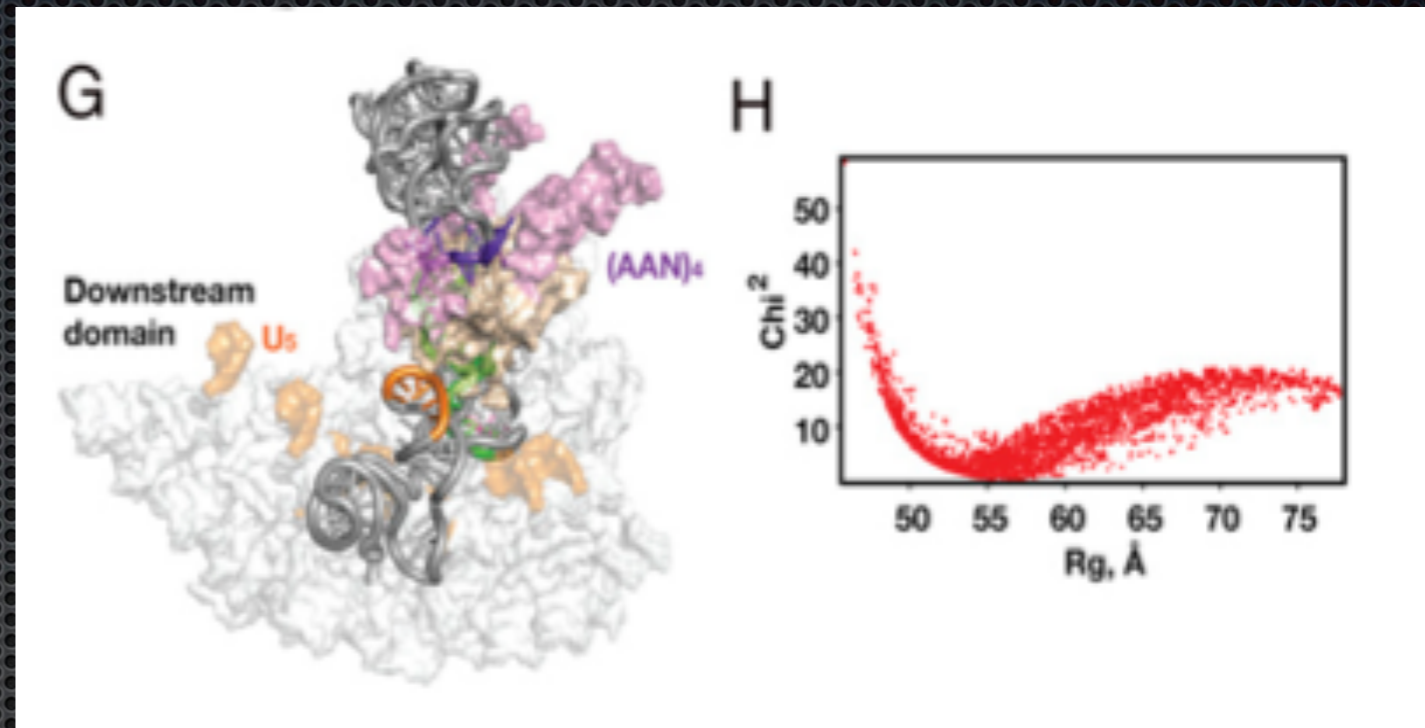
Example: RNA – protein complex

→ many experimental constraints (SHAPE)

→ 12 flexible regions in 6 protein segments (tails)

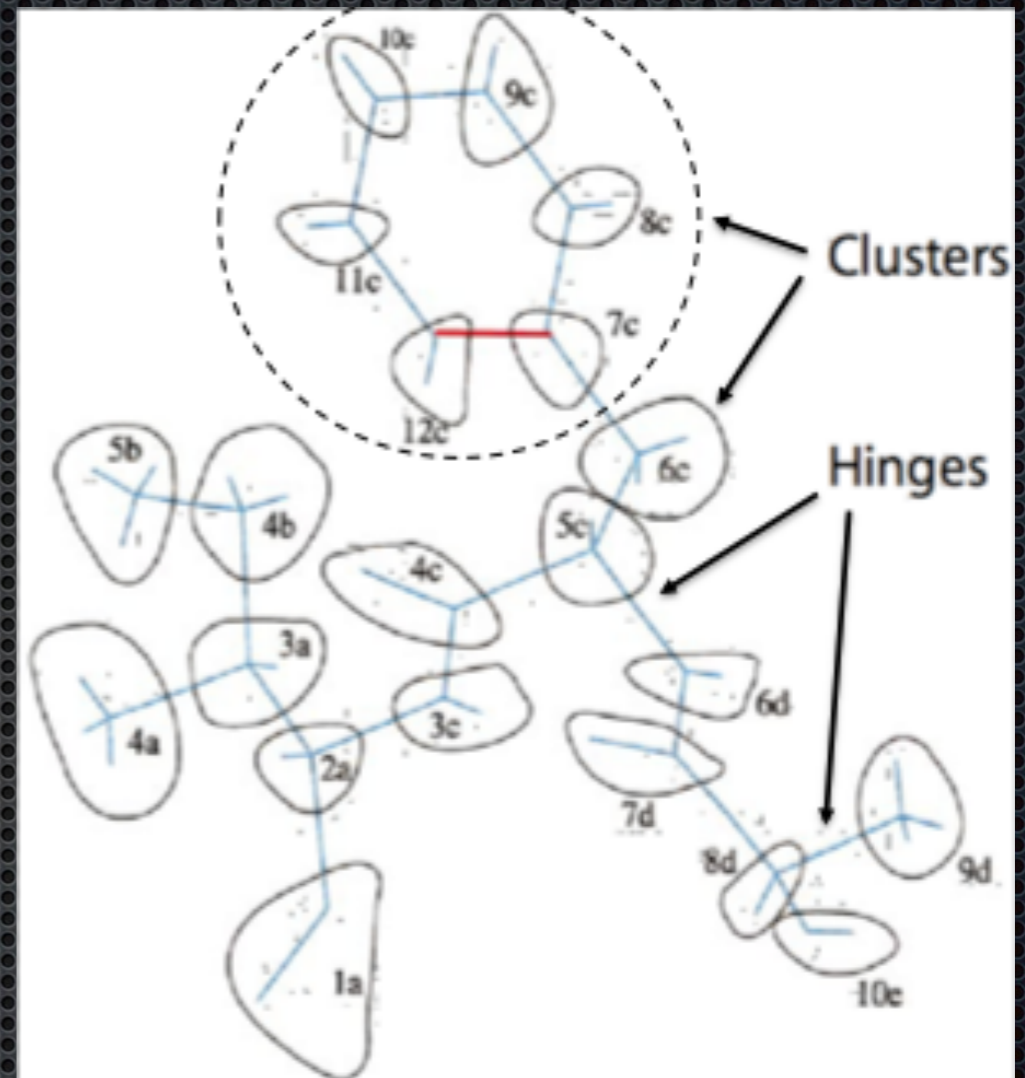
→ several RNA torsions sampled (systematically)

→ best fitting structures indicated a restricted range of motion:
opening and closing



Torsion Angle MD

- Much more robust than Cartesian MD (no high-frequency bond/angle vibrations)
- Larger MD time steps
- Sample the most relevant degrees of freedom (dihedrals)
- Convenient classification of rigid and flexible regions
- Can handle internal loops
- Requires extensive energy minimization to start



Run MC then TAMMD: multi-scale sampling use SAS as a guide

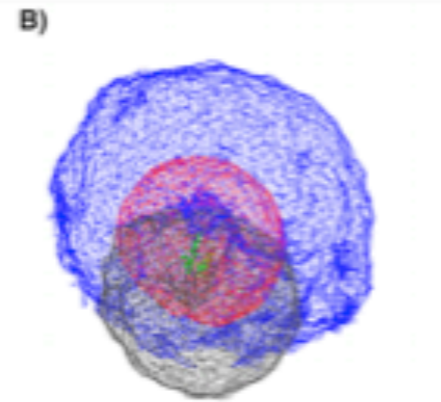
J. Chen, W. Im, C. Brookes, JCC 26, 1565 (2005), Zhang et al. (in preparation)

sassie-web -- Torsion Angle MD

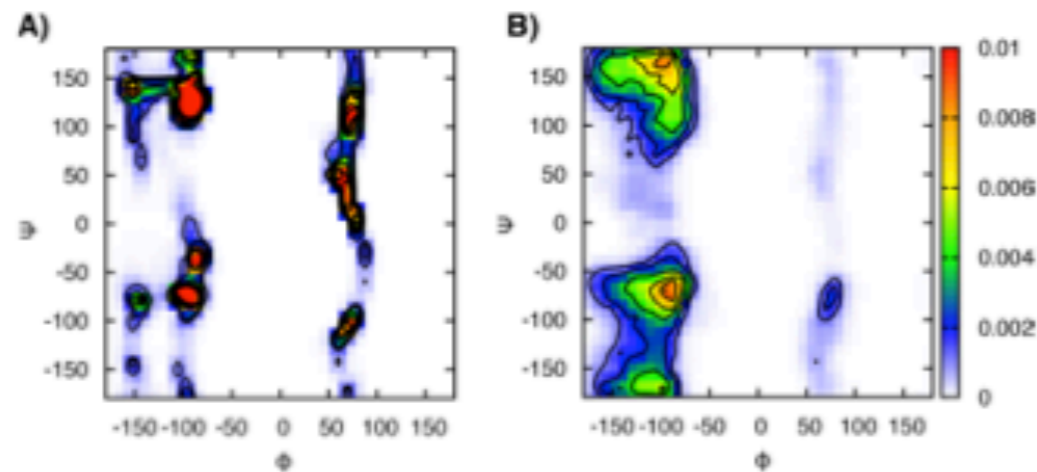
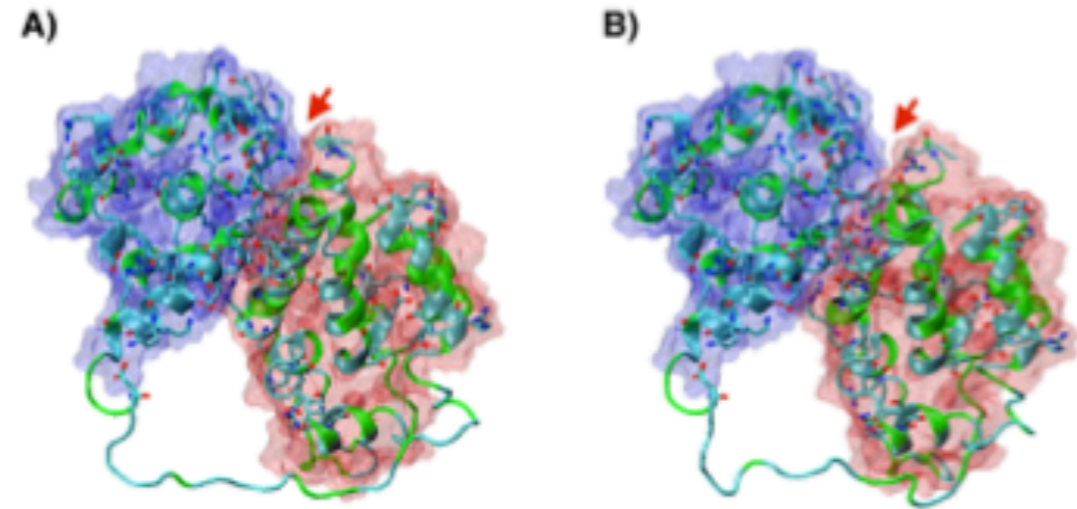
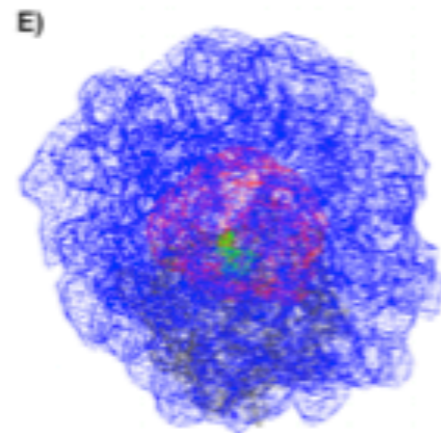
TAMD : combine MC w/ fast MD

Protein, DNA, RNA,
carbohydrates . . . and complexes

MC (50000)

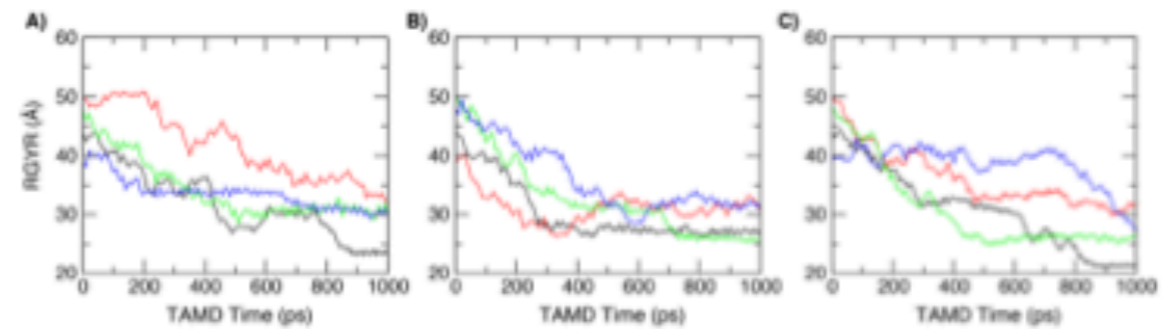


Spatial Clusters
(~100)



Zhang et al. (in preparation)

Implicit solvent effects

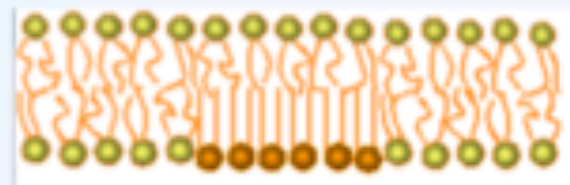


EFF1, SASA, GB, ACE

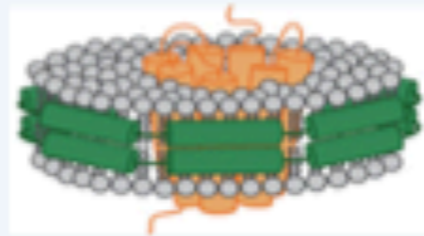
CG Sim. For Chem.

- Martini / LAMPPS etc. accessible *now* via SASCALC (contrast coming soon)
- But some systems are even bigger and may benefit from specialized CG Methods

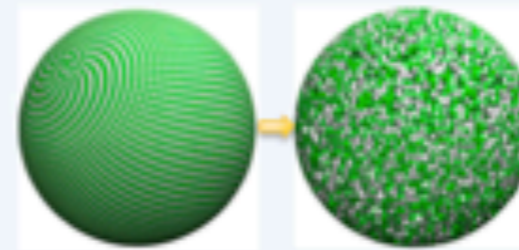
Coarse-grain model systems



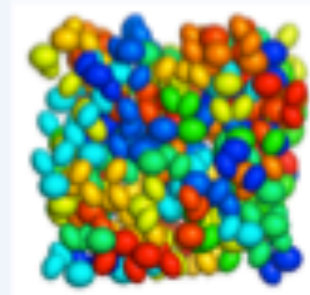
Lipid rafts



Nanodiscs

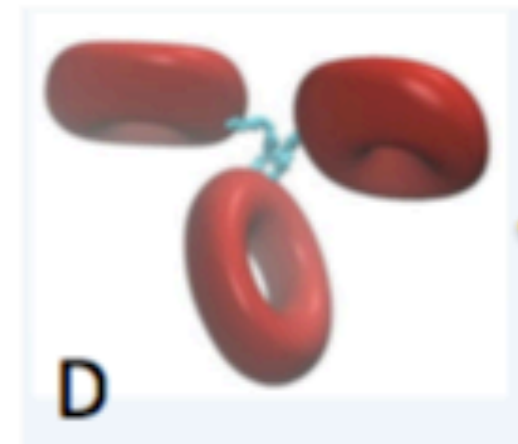
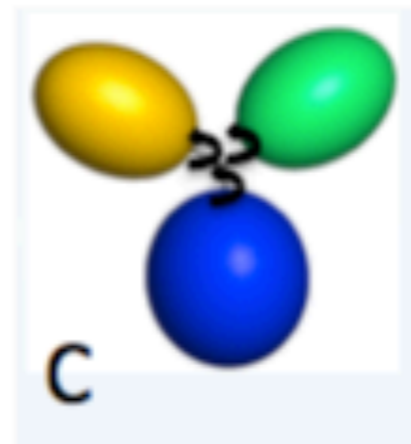
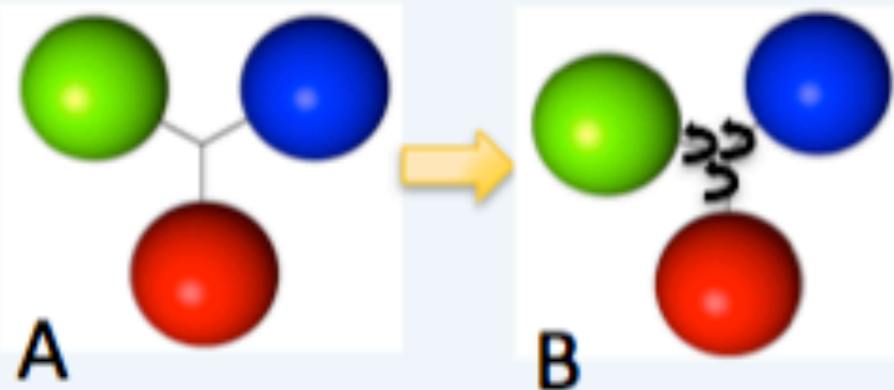


Lipid rafts on
sphere surfaces



High-density
protein systems

Apply the level of detail required by experiment



Zhang & Curtis (in preparation)

$$S(\mathbf{q}) = 1 + \rho \int_V d\mathbf{r} e^{-i\mathbf{q}\cdot\mathbf{r}} g(\mathbf{r})$$

sassie-web -- Calculate

Input a TRAJECTORY and get an ensemble of scattering profiles

Calculate -- generate theoretical scattering data from structures

SasCalc Calculates neutron and X-ray scattering profiles from input structures. Will replace xtal2sas once completed.

SasCalc-MD Calculates neutron and X-ray scattering profiles from MD trajectories with explicit water.

Xtal2sas Calculates neutron scattering profiles from input structures.

Cryson Calculates neutron scattering profiles from input structures.

Crysol Calculates X-ray scattering profiles from input structures.

SCT Calculate Calculates neutron and X-ray scattering profiles from input structures.

SLD MOL Calculates neutron and X-ray reflectivity scattering length density from user supplied structures. Utilities for experimental planning and isotopic labeling and optimization of ensemble populations are supplied.

EM to SANS Calculates neutron scattering profile from user supplied electron density map.

sassie-web -- Analyze

Analyze -- compare theoretical data to experimental data and advanced analysis methods

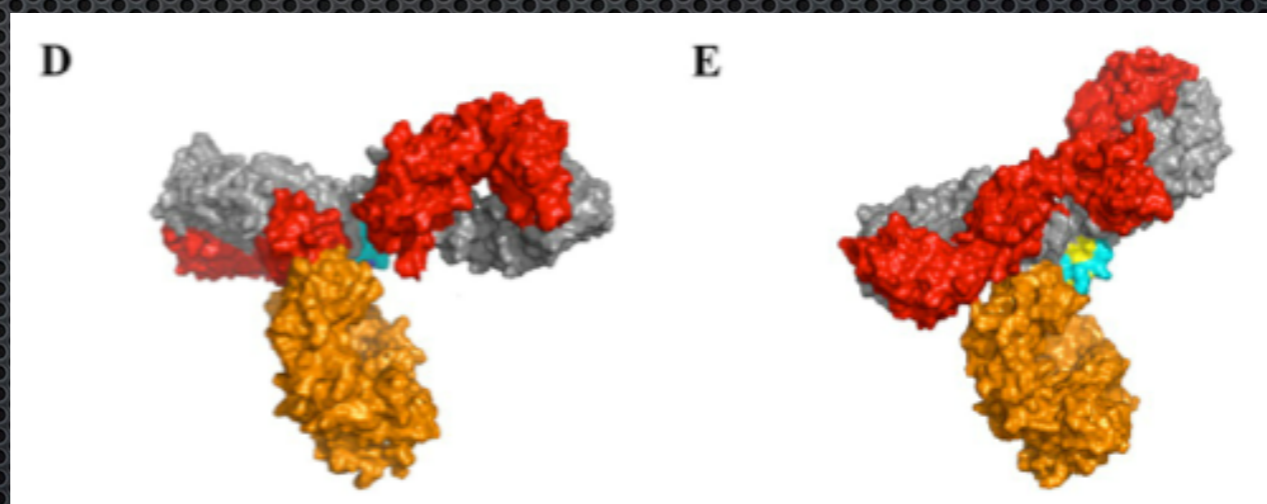
Chi-Square Filter Compares calculated scattering profiles to user supplied interpolated scattering profiles. No fitting.

SCT Analyze Compares calculated scattering profiles to user supplied interpolated scattering profiles. No fitting.

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

APBS Calculates non-polar solvation energy using adaptive Poisson-Boltzman solver from user supplied structures.

Both $\chi^2 \sim 1$



Free Energy differs by > 400 kCal/mol

Clark et al. J. Phys. Chem. B 2013, 117, 14029–14038, 2013

sassie-web -- Tools

Tools -- general utilities

Contrast Calculator Experimental planning to predict scattering length density and $I(0)$ from input PDB file, Fasta sequence file, or chemical formula.

Align Moves structures from input structure file so that they overlap with a reference input structure.

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

Extract Utilities Allows user to extract individual structures and/or SAS profiles from a larger input structure file and/or folder containing SAS profiles.

Merge Utilities Allows user to combine user supplied structures and calculated scattering profiles into a single set of new files.

Small-Angle Scattering Contrast Calculator for Protein and Nucleic Acid Complexes in Solution, J. Appl. Cryst. 46, 1889-1893 (2013).

sassie-web



20+ modules ported using GenApp

sassie-web.chem.utk.edu/sassie2

User persistent data structure

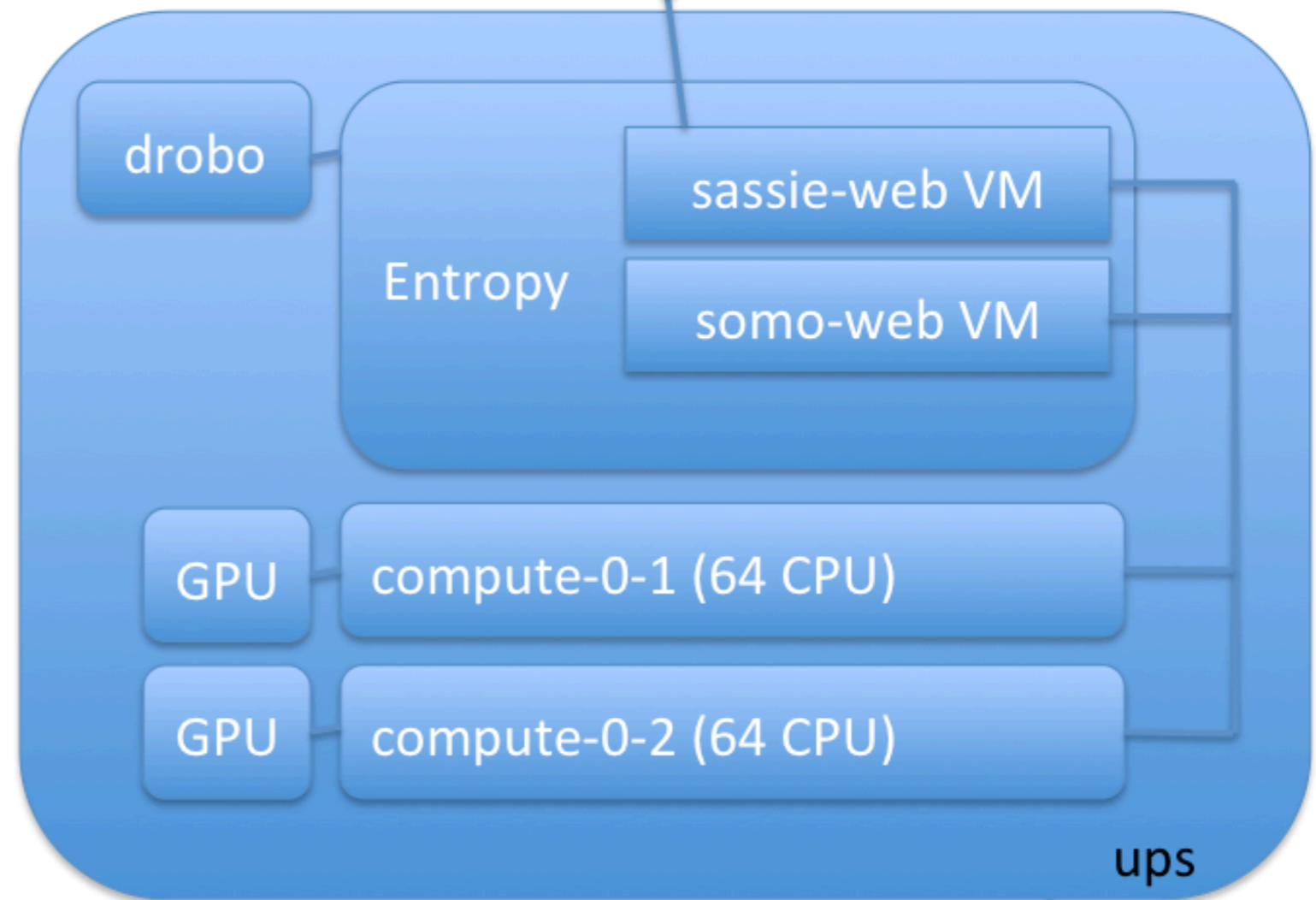
128 CPU / ~20,000 GPU cluster / 30 TB

beta testers Google Group:

CCP-SAS SASSIE

<https://groups.google.com/forum/#!forum/madscatt>

Allows fewer stand-alone roll-outs && potential GenApp stand-alone GUI version(s) && iOS & Android Apps.



Remote web-monitoring
& control

Please contribute feedback to Group!

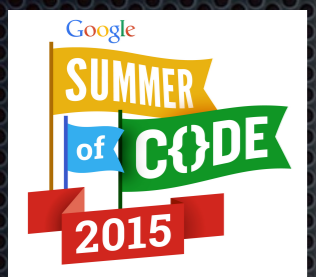


HPC is relative

Basic server: mobile-entropy



Prototype server: Entropy → sassie-web



Enable advanced MD & APBS



Langan & Smith



HPC is relative

☰

SASSIE-web : Beta

Login Help on

Tools

Build

Interact

Simulate

Calculate

Analyze

DOCS FEEDBACK -web



2015

Enable advanced MD & APBS

Langan & Smith



sassie-web -- work in progress

<https://sassie-web.chem.utk.edu/sassie2>

The screenshot displays the SASSIE-web Beta interface. At the top, there is a navigation menu with a hamburger icon on the left, the title "SASSIE-web : Beta" in the center, and "Login Help on" with a user icon on the right. A vertical sidebar on the left contains six categories: "Tools" (wrench icon), "Build" (molecular model icon), "Interact" (hand icon), "Simulate" (computer monitor icon), "Calculate" (atom icon), and "Analyze" (gears icon). The main content area features a feedback form titled "Please give your feedback" with a close button (X) in the top right corner. The form includes the following fields and options: "Server date" (2015 May 05 01:50:08 UTC), "Email address" (gibblygop@gmail.com), "Repeat email address" (gibblygop@gmail.com), "Subject" (empty text box), "Suggestion" (radio button), "Important" (radio button), "Critical" (radio button), "Provide a reference job(s) if available" (dropdown menu showing "Select a job or jobs"), and "Your comments" (large text area). A "Submit" button is located at the bottom left of the form. On the right side of the interface, there are two vertical buttons labeled "FEEDBACK" and "DOCS".

Ease of use, then HPC and advanced simulation tools