

Improving Gromacs trajectory analysis

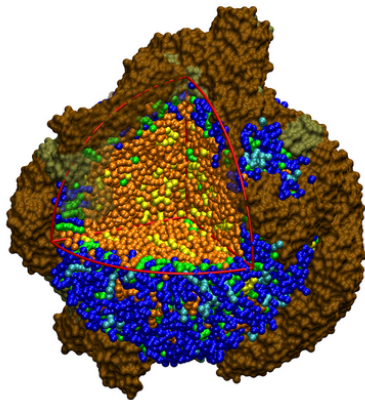
Dynamic selections and library functions

Teemu Murtola

January 30, 2009

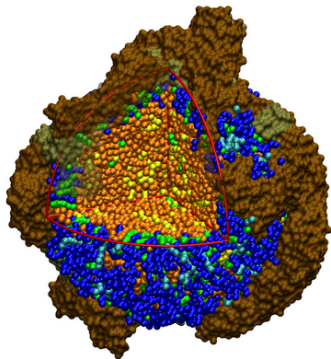
Model System

- Main components
 - Lipid droplet
 - Protein on the surface
 - Water (not shown)
- Simulated with MARTINI
- Five types of lipids:
 - POPC, PPC (surface)
 - CHES, TRIOL (core)
 - CHOL (both)
- Protein (surface)



Motivation 1/2

- Global properties easy to calculate using Gromacs tools
 - E.g., $\rho(r)$ using `g_rdf`
- How about local properties?
 - E.g., near to/far from the protein?
- VMD has flexible selections, but heavy calculations not easy to implement



Motivation 2/2

- Solution idea: general-purpose *analysis library*
 - No need to implement, e.g., dynamic selections for every tool separately
 - All tools can take advantage of library improvements
 - Framework for implementing trajectory analysis tools in general
- Main goals:
 - Flexible dynamic selections similar to VMD (no more index files)
 - Only minor modifications to existing programs
 - Transparency for the tool implementer
 - Easy access to center-of-mass data
 - Documentation using Doxygen

Example: Radial Density

Standard radial density calculation:

```
>make_ndx  
[First group: all lipids]  
[Each molecule type in separate group]  
>g_rdf -com -ng 6  
[Select groups]
```

Example: Radial Density

Standard radial density calculation:

```
>make_ndx  
[First group: all lipids]  
[Each molecule type in separate group]  
>g_rdf -com -ng 6  
[Select groups]
```

New radial density calculation (for POPC only for simplicity):

```
>g_rdf -com -select "resname POPC PPC CHOL CHES TRIOL; resname POPC"
```

Example: Radial Density

Standard radial density calculation:

```
>make_ndx  
[First group: all lipids]  
[Each molecule type in separate group]  
>g_rdf -com -ng 6  
[Select groups]
```

New radial density calculation (for POPC only for simplicity):

```
>g_rdf -com -select "resname POPC PPC CHOL CHES TRIOL; resname POPC"
```

New radial density calculation (full):

```
>cat selection.dat  
"Lipids"  resname POPC PPC CHOL CHES TRIOL;  
"POPC"   resname POPC;  
"PPC"    resname PPC;  
"CHOL"   resname CHOL;  
"CHES"   resname CHES;  
"TRIOL"  resname TRIOL;  
"Protein" group "Protein";  
>g_rdf -com -sf selection.dat
```

Example: More Radial Densities

For heads and tails separately:

```
>g_rdf -com \  
-select 'Lipids"      resname POPC PPC CHOL CHES TRIOL; \  
      "POPC head" resname POPC and not name C*; \  
      "POPC tail" resname POPC and name C*'
```

Example: More Radial Densities

For heads and tails separately:

```
>g_rdf -com \  
-select "Lipids"   resname POPC PPC CHOL CHES TRIOL; \  
          "POPC head" resname POPC and not name C*; \  
          "POPC tail" resname POPC and name C*
```

Under the protein:

```
>cat selection.dat  
lipids      = resname POPC PPC CHOL CHES TRIOL;  
underprot   = insolidangle center com of lipids span group "Protein" cutoff 5;  
  
"Lipids"    lipids;  
"POPC"      resname POPC and underprot;  
"PPC"       resname PPC and underprot;  
"CHOL"      resname CHOL and underprot;  
"CHES"      resname CHES and underprot;  
"TRIOL"     resname TRIOL and underprot;  
>g_rdf -com -sf selection.dat
```

More Complex Example

- Partition CHES molecules into groups based on their center-of-mass
 - Closer than 4 nm to the center
 - Between 4 nm and 6 nm
 - Near the surface (>6 nm)
 - Near protein
 - Far from protein
- Could be used, e.g., for diffusion analysis

More Complex Example

- Partition CHES molecules into groups based on their center-of-mass
 - Closer than 4 nm to the center
 - Between 4 nm and 6 nm
 - Near the surface (>6 nm)
 - Near protein
 - Far from protein
- Could be used, e.g., for diffusion analysis

```
>cat selection.dat
lipids      = rename POPC PPC CHOL CHES TRIOL;
underprot  = insolidangle center com of lipids span group "Protein" cutoff 5;
rdist      = dist from com of lipids;
"CHES (r < 4)"          rename CHES and rdist < 4;
"CHES (4 < r < 6)"      rename CHES and rdist >= 4 and rdist < 6;
"CHES (6 < r, near protein)"  rename CHES and rdist >= 6 and underprot;
"CHES (6 < r, not near prot.)" rename CHES and rdist >= 6 and not underprot;
>g_jumpdist ... -type res_com -selrpos res_com -sf selection.dat
```

(Some) Implementation Details

- Index files not completely gone:
 - Old command-line syntax (`-n`, `-ng`) still works
 - Possible to use within selections (`group "Atoms1", group 2`)
- Large dynamic selections may be slow, but this is usually not the bottleneck

Programming considerations:

- Well-documented template
- Only minor modification to existing codes needed
- Should be easy to use with no additional overhead (my `g_rdf` implementation some 15% shorter)
- Selection language easy to extend with new keywords
- Bison/Lex-based parser for the selections
- \approx 12 000 lines of C code (including docs)

Summary and Future

- Analysis library implementation with flexible text-based dynamic selections
- Current status:
 - Library is working, but more extensive testing is needed
 - Only a limited number of selection keywords implemented
 - Only a limited number of tools using the library
- Will be made publicly available once time allows