Tutorial 2. SIRAH force field in GROMACS

Hybrid solvation: Plugging SIRAH solvent to your atomistic system

By Matias Machado
Mail any comment or suggestion to *spantano@pasteur.edu.uy*

This tutorial shows how to apply the hybrid solvation approach of SIRAH force field to speed up the simulation of an atomistic solute. The example system contains a DNA fiber surrounded by a shell of atomistic waters, which are embedded in coarse-grained (CG) molecules, called WT4, to represent bulk water. The general procedure is extensible to any other solute. The hybrid solvation methodology is well tested for SPC, SPC/E, TIP3P atomistic water models. The main references for this tutorial are: Darré et al. *WAT4?* [JCTC, **2010**, 6:3793], Darré et al. *All-atoms/CG solvation* [JCTC, **2012**, 8:3880], Gonzalez et al. *Transferable All-atoms/CG solvation* [J Phys Chem B, **2013**, 117:14438], Machado et al. *SIRAH Tools* [Bioinformatics, **2017**, 32:1568]. We strongly advise you to read these articles before starting the tutorial.

Required Software
GROMACS 4.5.5 or later version properly installed in your computer. The molecular visualization program VMD (freely available at *www.ks.uiuc.edu/Research/vmd*).

Prior knowledge
How to perform a standard atomistic molecular dynamic simulation with GROMACS.

Hands on
0) Download the file *sirah_[version].gmx.tgz* from *www.sirahff.com* and uncompress it into your working directory. Notice: *[version]* should be replaced with the actual package version e.g.: x2_18-09

```
tar -xzvf sirah_[version].gmx.tgz
```

You will get a folder *sirah_[version].ff/* containing the force field definition, the SIRAH Tools in *sirah_[version].ff/tools/*, molecular structures to build up systems in *sirah_[version].ff/PDB/*, frequently asked questions in *sirah_[version].ff/tutorial/SIRAH_FAQs.pdf* and the required material to perform the tutorial in *sirah_[version].ff/tutorial/2/* .

Make a new folder for this tutorial:

```
mkdir tutorial2; cd tutorial2
```

Create the following symbolic links in the folder *tutorial2*:

```
ln -s ../sirah_[version].ff sirah.ff
```

1) Convert the PDB file of the solute *dna.pdb* into GROMACS format:

```
pdb2gmx -f sirah.ff/tutorial/2/dna.pdb -o dna.gro
```

When prompted, choose AMBER99SB and then TIP3P as water model.

Notice: GROMACS' commands in 5.x and later versions start with "*gmx*".

2) For GROMACS to recognize SIRAH, edit your topology file *topol.top* adding the following lines **after** the force field definition:

Topology before editing	Topology after editing

`; Include forcefield parameters`
`#include "amber99sb.ff/forcefield.itp`
	`; Include forcefield parameters`
	`#include "amber99sb.ff/forcefield.itp"`
	
	`#include "./sirah.ff/hybsol_comb2.itp"`
	`#include "./sirah.ff/solv.itp"`

Notice: *hybsol_comb2.itp* is a parameter file, while *solv.itp* links to the topologies of WT4 and SIRAH ions. The choice of the right parameter file depends on the chosen atomistic force field (see Table 1). The plug-in works smoothly with the implemented force fields of the GROMACS distribution. When using a customized force field (e.g. for lipids) choose the parameter file according to the combination rule and check that the atom type you are using for the atomistic water match the DEFAULT in the SIRAH parameter file, which is OW. If they don't match then rename OW accordingly to your definition.

3) Solvate the system

Define the simulation box of the system:

```
editconf -f dna.gro -o dna_box.gro -bt octahedron -d 2 -c
```

Then add an atomistic water shell:

```
genbox -cp dna_box.gro -cs spc216.gro -o dna_shell.gro -shell 1
```

Notice: in GROMACS 5.x and later versions the command genbox was renamed to "*gmx solvate*"

Finally add the CG solvent:

```
genbox -cp dna_shell.gro -cs ./sirah.ff/wt4tip3p.gro -o dna_sol.gro
```

Edit the *[molecules]* section from *topol.top* to include the number of SOL and WT4 molecules:

Hint! You can use the following command line to get these numbers

```
grep -c OW dna_sol.gro; grep -c WP1 dna_sol.gro
```

Topology before editing	Topology after editing

`[ molecules ]`
`; Compound  #mols`
`DNA_chain_A  1`
`DNA_chain_B  1`
	`[ molecules ]`
	`; Compound  #mols`
	`DNA_chain_A  1`
	`DNA_chain_B  1`
	`SOL  3580`
	`WT4  2659`

Notice: The number of added SOL and WT4 molecules (3580, 2659) may change according to the software version.

Add CG counterions:

```
grompp\
 -f ./sirah.ff/tutorial/2/CPU/em_HYBSOL.mdp\
 -p topol.top\
 -c dna_sol.gro\
 -o dna_sol.tpr

genion -s dna_sol.tpr -o dna_ion.gro -np 38 -pname NaW
```

When prompted, choose to substitute WT4 molecules by NaW ions.

Notice: The available ionic species in SIRAH force field are: Na+ (NaW), K+ (KW) and Cl- (ClW). One ion pair (e.g. NaW-ClW) each 34 WT4 molecules renders a salt concentration of ~0.15M (see Appendix 1). Be aware that SIRAH ions remain within the CG phase. So, if the presence of atomistic electrolytes in close contact with the solute is important to describe the physics of the system you will have to add them.

Edit again the *[molecules]* section in *topol.top* to include the 38 NaW ions and the correct number of WT4 molecules.

Before running the simulation it may be a good idea to visualize your molecular system. CG molecules are not recognized by molecular visualizers and will not display correctly. To fix this problem you may generate a PSF file of the system using the script *g_top2psf.pl*:

```
./sirah.ff/tools/g_top2psf.pl -i topol.top -o dna_ion.psf
```

Notice: This is the basic usage of the script *g_top2psf.pl*, you can learn other capabilities from its help:

```
./sirah.ff/tools/g_top2psf.pl -h
```

Use VMD to check how the hybrid system looks like:

```
vmd dna_ion.psf dna_ion.gro -e ./sirah.ff/tools/sirah_vmdtk.tcl
```

Notice: VMD assigns default radius to unknown atom types, the script *sirah_vmdtk.tcl* sets the right ones. It also provides a kit of useful selection macros and coloring methods. Use the command *sirah_help* in the Tcl/Tk console of VMD to access the manual pages.

4) Run the simulation

The folder *sirah.ff/tutorial/2/CPU/* contains typical input files for energy minimization (*em_HYBSOL.mdp*), equilibration (*eq_HYBSOL.mdp*) and production (*md_HYBSOL.mdp*) runs. Please check carefully the input flags therein.

Create an index files adding a group for WT4 and NaW:

```
echo -e "r WT4 | r NaW\nq\n" | make_ndx -f dna_ion.gro -o dna_ion.ndx
```

Make a new folder for the run:

```
mkdir run; cd run
```

Energy Minimization:

```
grompp\
 -f  ../sirah.ff/tutorial/2/CPU/em_HYBSOL.mdp\
 -p  ../topol.top\
 -po    em.mdp\
 -n  ../dna_ion.ndx\
 -c  ../dna_ion.gro\
 -o     dna_em.tpr

mdrun -deffnm dna_em &> EM.log &
```

Equilibration:

```
grompp\
 -f  ../sirah.ff/tutorial/2/CPU/eq_HYBSOL.mdp\
 -p  ../topol.top\
 -po    eq.mdp\
 -n  ../dna_ion.ndx\
 -c     dna_em.gro\
 -o     dna_eq.tpr

mdrun -deffnm dna_eq &> EQ.log &
```

Production (10ns):

```
grompp\
 -f  ../sirah.ff/tutorial/2/CPU/md_HYBSOL.mdp\
 -p  ../topol.top\
 -po    md.mdp\
 -n  ../dna_ion.ndx\
 -c     dna_eq.gro\
 -o     dna_md.tpr

mdrun -deffnm dna_md &> MD.log &
```

Notice: You can find example input files for the GPU-CPU version of *mdrun* at folder *GPU/* within *sirah.ff/tutorial/2/* . GPU flags were set for GROMACS 4.6.7, different versions may complain about some specifications.

That's it! Now you can analyze the trajectory.

Process the output trajectory at folder *run/* to account for the Periodic Boundary Conditions (PBC):

```
trjconv\
 -s   dna_em.tpr\
 -f   dna_md.xtc\
 -o   dna_md_pbc.xtc\
 -n   ../dna_ion.ndx\
 -ur  compact -center\
 -pbc mol
```

When prompted, choose "*DNA*" for centering and "*System*" for output.

Now you can check the simulation using VMD:

```
vmd ../dna_ion.psf ../dna_ion.gro dna_md_pbc.xtc\
 -e ../sirah.ff/tools/sirah_vmdtk.tcl
```

Table 1. SIRAH parameter file to include in the topology according to the chosen atomistic force field.

		Atomistic force field				
Parameter file	Combination rule	GMX	GROMOS	AMBER	CHARMM	OPLS
hybsol_comb1.itp	1	X	X			
hybsol_comb2.itp	2			X	X	
hybsol_comb3.itp	3					X

Appendix 1: Calculating ionic concentrations

$\rho_{WT4} = \rho_{H2O}$ = 1000 g/L

MW_{H2O} = 18 g/mol

1 WT4 ~ 11 H_2O

$$M = \frac{mol}{V} \quad ; \quad n = mol\, N_A \quad ; \quad \rho = \frac{m}{V} \quad ; \quad m = mol\, MW$$

$$V = \frac{m}{\rho} = \frac{mol\, MW_{H_2O}}{\rho} = \frac{n_{H_2O} MW_{H_2O}}{N_A \rho} \quad ; \quad M = \frac{mol}{V} = \frac{n_{ion}}{N_A V} = \frac{n_{ion}}{N_A} \frac{N_A \rho}{n_{H_2O} MW_{H_2O}} = \frac{n_{ion} 1000}{n_{WT4}(11)(18)} \sim 5 \frac{n_{ion}}{n_{WT4}}$$

Number of WT4 molecules per ion at 0.15M: $n_{WT4} = 5 \frac{n_{ion}}{M} = \frac{5(1)}{0.15} \sim 34$