Tutorial 1. SIRAH force field in GROMACS

Simulation of a coarse grained DNA molecule in explicit solvent

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

This tutorial shows how to use the SIRAH force field to perform a coarse grained (CG) simulation of a double stranded DNA in explicit solvent (called WatFour, WT4) in four simple steps: 1 download; 2 map; 3 solvate and; 4 run. The main references for this tutorial are: Dans et al. *SIRAH DNA* [JCTC, **2010**, 6:1711], Darré et al. *WAT4?* [JCTC, **2010**, 6:3793], 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/1/* .

Make a new folder for this tutorial:

```
mkdir tutorial1; cd tutorial1
```

Create the following symbolic links in the folder *tutorial1*:

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

Notice: Files *residuetypes.dat* and *specbond.dat* are essential for the correct definition of molecular groups and auto-detection of disulfide bonds and cyclic DNA polymers.

1) Map the atomistic structure of a 20-mer DNA to its CG representation:

```
./sirah.ff/tools/CGCONV/cgconv.pl\
 -i sirah.ff/tutorial/1/dna.pdb\
 -o dna_cg.pdb
```

The input file *dna.pdb* contains all the heavy atoms composing the DNA molecule, while the output *dna_cg.pdb* preserves a few of them. Please check both PDB structures using VMD:

```
vmd -m sirah.ff/tutorial/1/dna.pdb dna_cg.pdb
```

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

```
./sirah.ff/tools/CGCONV/cgconv.pl -h
```

From now on it is just normal GROMACS stuff!

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

2) Use pdb2gmx to convert your PDB file into GROMACS format:

```
pdb2gmx -f dna_cg.pdb -o dna_cg.gro -merge all
```

When prompted, choose "*SIRAH force field"* and then "*SIRAH solvent models"*.

Notice: Getting warning messages of long bonds is fine and expected due to the CG nature of the residue topologies. However missing atom messages are errors which probably trace back to the mapping step. In that case, check your atomistic and mapped structures and do not carry on the simulation until the problem is solved.

Notice: Merging both DNA chains is convenient when planning to apply restraints between them.

During long simulations of DNA, capping residues may eventually separate. If you want to avoid this effect, which is called helix fraying, add Watson-Crick (WC) restraints at terminal base pairs. To add these restraints edit *topol.top* to include the file *WC_RST.itp* at the end of the *[moleculetype]* section:

Topology without WC restraints	Topology with WC restraints

`; Include Position restraint file`
`#ifdef POSRES`
`#include "posre.itp"`
`#endif`
	`; Include Position restraint file`
	`#ifdef POSRES`
	`#include "posre.itp"`
	`#endif`
	
	`; Watson-Crick restraints`
	`#include "./sirah.ff/tutorial/1/WC_RST.itp"`

3) Solvate the system

Define the simulation box of the system:

```
editconf -f dna_cg.gro -o dna_cg_box.gro -bt octahedron -d 2 -c
```

Add WT4 molecules:

```
genbox -cp dna_cg_box.gro -cs sirah.ff/wt416.gro -o dna_cg_sol.gro
```

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

Edit the *[molecules]* section in *topol.top* to include the number of added WT4 molecules:

Hint! If you forget to read the number of added WT4 molecules from the output of genbox, then use the following command line to get it

```
grep -c WP1 dna_cg_sol.gro
```

Topology before editing	Topology after editing

`[ molecules ]`
`; Compound  #mols`
`DNA_chain_A  1`
	`[ molecules ]`
	`; Compound  #mols`
	`DNA_chain_A  1`
	`WT4  3179`

Notice: The number of added WT4 molecules (3179) may change according to the software version.

Add CG counterions and 0.15M NaCl:

```
grompp\
 -f sirah.ff/tutorial/1/CPU/em_CGDNA.mdp\
 -p topol.top\
 -c dna_cg_sol.gro\
 -o dna_cg_sol.tpr
genion -s dna_cg_sol.tpr -o dna_cg_ion.gro -np 113 -pname NaW -nn 75 -nname ClW
```

When prompted, choose to substitute WT4 molecules by 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). Counterions were added according to Machado et al. *SPLIT* [JCTC, **2020**].

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

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_cg_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 CG system looks like:

```
vmd dna_cg_ion.psf dna_cg_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, coloring methods and a backmapping utility. 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/1/CPU/* contains typical input files for energy minimization (*em_CGDNA.mdp*), equilibration (*eq_CGDNA.mdp*) and production (*md_CGDNA.mdp*) runs. Please check carefully the input flags therein.

Create an index file:

```
echo "q" | make_ndx -f dna_cg_ion.gro -o dna_cg_ion.ndx
```

Notice: WT4 and CG ions (NaW and ClW) are automatically set to the group "*SIRAH-Solvent*".

Make a new folder for the run:

```
mkdir run; cd run
```

Energy Minimization:

```
grompp\
 -f  ../sirah.ff/tutorial/1/CPU/em_CGDNA.mdp\
 -p  ../topol.top\
 -po    em.mdp\
 -n  ../dna_cg_ion.ndx\
 -c  ../dna_cg_ion.gro\
 -o     dna_cg_em.tpr

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

Equilibration:

```
grompp\
 -f  ../sirah.ff/tutorial/1/CPU/eq_CGDNA.mdp\
 -p  ../topol.top\
 -po    eq.mdp\
 -n  ../dna_cg_ion.ndx\
 -c     dna_cg_em.gro\
 -o     dna_cg_eq.tpr

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

Production (100ns):

```
grompp\
 -f  ../sirah.ff/tutorial/1/CPU/md_CGDNA.mdp\
 -p  ../topol.top\
 -po    md.mdp\
 -n  ../dna_cg_ion.ndx\
 -c     dna_cg_eq.gro\
 -o     dna_cg_md.tpr

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

Notice: You can find example input files for the GPU-CPU version of *mdrun* at folder *GPU/* within *sirah.ff/tutorial/1/* . 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_cg_em.tpr\
 -f   dna_cg_md.xtc\
 -o   dna_cg_md_pbc.xtc\
 -n   ../dna_cg_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_cg_ion.psf ../dna_cg_ion.gro dna_cg_md_pbc.xtc\
 -e ../sirah.ff/tools/sirah_vmdtk.tcl
```

Appendix 1: Calculating ionic concentrations

$\rho_{WT4} = \rho_{H2O} = 1000$ g/L
$MW_{H2O} = 18$ g/mol
1 WT4 ~ 11 H_2O

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

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

Number of WT4 molecules per ion at 0.15M: $n_{WT\,4}=5\frac{n_{ion}}{M}=\frac{5(1)}{0.15}\sim 34$