

Molecular dynamics simulation of hevein domain

Go to the PDB web site and download PDB file *ITOW*. Make two copies of the PDB file; one for protein and one for ligand. Edit the protein file, choose a model (for example the first) and remove all other models and all lines except those starting with **ATOM**. Then remove hydrogen atoms. To do this you can use **egrep** command. The final PDB file should contain only non-hydrogen protein atoms of a single model. To create Gromacs topology and coordinates, type:

```
pdb2gmx -f protein_file.pdb -o protein -p protein
```

When you are asked to chose force field, chose Amber 99SB for protein and TIP3P for water. You can check the resulting files **protein.gro** and **protein.top**.

Now we have to create the box confining the protein. This can be done by typing:

```
editconf -f protein -o box -c -d 1.2
```

This will create the file **box.gro** with a minimal distance between the side of box and an atom 1.2 nm. Then type:

```
genbox -cs -cp box -p protein -o solvated
```

Luckily, the protein contains exactly the same number of positively and negatively charged groups, so its net charge is neutral. Otherwise, it would be necessary to exchange some water molecules for sodium cations or chloride anions to make the box neutral. Program **genion** can be used for this purpose.

Now we have to do energy minimization with following parameters in the file **em.mdp**:

```
constraints      = none
integrator       = steep
nsteps          = 1000
emtol           = 0
emstep          = 0.1
nstcomm         = 1
ns_type         = grid
coulombtype     = PME
rlist           = 1.0
rcoulomb        = 1.0
rvdw           = 1.0
```

Energy minimization can be done by typing:

```
grompp -f em -c solvated -p protein -o em1
mdrun -s em1 -o em1 -g em1 -c after_em1 -e em1
```

After several minutes you can check output files.

Now we can continue with a molecular dynamics simulation. Most protein simulation protocols start with geometry minimization. This is followed by a sequence of energy minimizations, low-temperature simulations and/or simulations with artificial forces restraining motions of protein atoms. These steps are necessary to “equilibrate” the system, for example to allow water molecules to find suitable binding poses on the protein surface. The equilibration phase is followed by a long “production” run at desired conditions. Here we will run only a single short simulation to demonstrate the concept.

The simulation uses following parameters stored in the file **md.mdp**:

```
constraints      = none
integrator       = sd
dt              = 0.001 ; ps
nsteps          = 1000 ; total 1 ps.
nstxout         = 10
nstvout         = 10
nstfout         = 0
nstlog          = 10
nstenergy       = 1000
nstlist         = 10
```

```

ns_type           = grid
coulombtype       = PME
rlist             = 1.0
rvdw             = 1.0
rcoulomb         = 1.0
Tcoupl           = v-rescale
Tc-grps          = System
tau_t            = 0.1
ref_t            = 10
energygrps       = System
gen_vel          = no

```

The simulation can be done by typing:

```

grompp -f md -c after_em1.gro -p protein -o md1
mdrun -s md1 -o md1 -g md1 -c after_md1 -e md1

```

After several minutes you can check the result in VMD.

Molecular dynamics simulation of hevein domain – GlcNAc₃ complex

Simulation of the unliganded hevein domain was relatively easy because it is (except the C-terminal amide) a pure protein. Gromacs can naturally handle proteins composed of standard amino acid residues (as well as a C-terminal amide). However, to simulate the hevein domain – GlcNAc₃ complex we have to generate a topology for the ligand. Gromacs alone cannot do it because it is not designed for carbohydrate simulations. Another package Amber contains parametrization of sugars called Glycam. Visit the Glycam web site, namely the link Carbohydrate builder. In the web form, build the structure by choosing D/L-isomer, residue type, configuration and linkage, Repeat this for three times to get GlcNAc₃. At the end, chose OH as an “aglycon”. Then submit the structure and wait until topology and coordinate files are built (they are supplied as *zip* or *tar.gz* archive).

Glycam web site produces files that are ready to be used with the program Amber, not with Gromacs. You can use a script `amb2gmx.p1` to convert them to the Gromacs format. However, there is one important pitfall when using Glycam force field with `amb2gmx.p1`!!! Energies of rotatable bonds can be approximated by a combination of goniometric functions of the torsion angle. Each goniometric function is defined by its amplitude, frequency and shift on the horizontal axis. For example, the function $-\sin(x)$ can be written either as function $\sin(x)$ multiplied by the amplitude -1 , or as function $\sin(x)$ shifted by π , namely $\sin(x - \pi)$. Authors of `amb2gmx.p1` script assumed that all people would use only the second option. However, the authors of Glycam used the first option. To avoid deletion of important torsion terms, you have to edit the file `amb2gmx.p1` and modify lines:

```

for(my $j=$i;$j<=$lines;$j++){
  my $period = abs($pn{$j});
  if($pk{$j}>0) {
    $V[$period] = 2*$pk{$j}$scal/$idivf{$j};
  }
}

```

to:

```

for(my $j=$i;$j<=$lines;$j++){
  my $period = abs($pn{$j});
  #if($pk{$j}>0) {
    $V[$period] = 2*$pk{$j}$scal/$idivf{$j};
  }
#}

```

where # means comment, so the program ignores the rest of the line.

The (modified) program can be used by typing:

```

./amb2gmx.p1 --prmtop amber-prmtop-filename --crd amber-crd-filename --outname outname

```

where you can chose the outname. It will create two files: `outname.gro` and `outname.top`. The program requires installation of AmberTools package.

The coordinates of the file generated by `amb2gmx.p1` are different from those found in the PDB file. There are two ways how to generate coordinates for GlcNAc₃ to be compatible with the PDB structure.

First, dirty way, you can convert the `outname.gro` to `outname.pdb` by program `editconf`. You can open the resulting `pdb` file together with the original PDB file of the protein-ligand complex in UCSF Chimera. Then you can manually fit and twist the sugar to the binding site. Finally, you can save it and convert back to `gro` format. However, you will never fit it exactly the same as in the experimental structure.

Second possibility, more laborious, is following. Copy lines of the trisaccharide from the original NMR structure to a separate PDB file and convert it to `gro` format, e.g.

`trisaccharide.gro`. Draw a large structure of the trisaccharide with all hydrogens on a sheet of paper. Then open the trisaccharide PDB in some molecular viewer and label all atoms. Write atom indexes to the structure on the paper. Concatenate `outname.gro` and `trisaccharide.gro` files to one file, e.g. `goodtrisaccharide.gro`. Using the drawn structure, place atoms from `trisaccharide.gro` to the same order as in `outname.gro` and then remove atoms of `outname.gro`.

Once you created trisaccharide conformation compatible with the receptor, you can combine them together. First, make a copy of the trisaccharide's topology file with the suffix `.itp`. Next, edit the new file and comment lines of the item `defaults` by typing `;` at the beginning. Copy the whole item `atomtypes` to the beginning of the file `protein.top` (just after inclusion of force field). Change atom types from `C`, `Ng` etc. to `XC`, `XNg` etc. This is necessary to avoid confusions between protein and ligand atom types. Comment the item `atomtypes` in the `itp` file and change atom types in `atoms` to `XC`, `XNg` etc.

Rename the molecule from `solute` to something more fancy, e.g. `glcnac3`. Then comment also lines of items `system` and `molecules` at the end of file. Then, edit the topology file of the protein, add line:

```
#include "glcnac3.itp"
```

between position restrain and water item at the end of file. Finally, in the `molecules` item add one `glcnac3` molecule.

Finally, concatenate files `protein.gro` and `goodtrisaccharide.gro`. Update the number of atoms in the second line, remove extra lines from the `goodtrisaccharide.gro` and place the box periodicity line from `protein.gro` to the end of file. Now you can run energy minimization and molecular dynamics simulation as described above.