HydroDock

Tutorial

ver. 25-06-2021

Balázs Zoltán Zsidó[1], Rita Börzsei[1,2], Viktor Szél[1], Csaba Hetényi[1]

[1]*Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary.*

[2]*Department of Pharmacology, Faculty of Pharmacy, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary.*

HydroDock is a protocol for calculation of hydrated target-ligand complexes. The protocol produces the complexes from scratch, using the structures of unbound target, ligand and water molecules as input. HydroDock is based on open source and freely available software packages.

This Tutorial provides a practical help with the details of the usage of HydroDock following the steps described in the original publication (Zsidó et al., Determination of ligand binding modes in hydrated viral ion channels to foster drug design and repositioning, J Chem Inf Model, 2021, x, y-z, under review). Input and output files **marked with bold** are available in the GitHub repository of HydroDock except large trajectory files deposited on Google Drive.

Table of Contents

Step 1 – Dry docking

Software used: AutoDock 4.2.6 (http://autodock.scripps.edu/)
AutoDockTools (http://autodock.scripps.edu/resources/adt)
Available files: **step1.tgz**

For the dry docking, the dry target and the ligand structures were needed. Their respective *.pdbqt formats were created in AutoDock Tools using the following menu commands.

File>Read Molecule > **infl.pdb**
Add> H atoms
Save>As pdb: **infl_h.pdb**
Grid>Macromolecule>open> **infl_h.pdb**
Save as **infl.pdbqt**

Ligand>Input > **amantadine.pdb**
Torsion tree > Choose torsions
Output>Save as pdbqt> **amantadine.pdbqt**

The grid maps were created with the following command line, after setting the coordinates of the grid box in AutoDock Tools:

```
autogrid4 -p amanta_flu.gpf -l amanta_flu.glg
```

After setting the docking parameters in AutoDock Tools, the following command line was used for the docking:

```
autodock4 -p amanta_flu.dpf -l amanta_flu.dlg
```

The docked ligand file: **O_amanta_flu_rank_1.pdb**
The complex file: **dry_complex.pdb**

Step 2 – Building the water structure of the inner surface of the target channels

Software used: GROMACS 5.1.4. (https://www.gromacs.org/)
MobyWat 1.1 (http://mobywat.com/)
Available files: **step2.tgz, system.xtc** (a separate link is provided for the xtc file below)

The preparation of the box, and the other MD steps were carried out with Gromacs 5.1.4. As a first step, hydrogens were added to the target molecule (target.pdb) from the holo structure (target.pdb) and it was placed in a box and the box is filled up with water molecules. Coordinates of the box are stored in a file named b4em.gro. TIP3P water model, Amber99sb-ILDN force field and a dodecahedral box with 10 Å (=1 nm) spacing specified from the solute were used.

```
gmx pdb2gmx -water tip3p -ff amber99sb-ildn -ignh -f target.pdb
gmx editconf -o -d 1 -bt cubic -f conf.gro
gmx solvate -cp out -cs -o b4em -p topol
```

If the target has non-zero net charge Neutrality of the system have to be set by adding X copies of positive (Na+) or negative (Cl-) ions to the box, as PME (Particle Mesh-Ewald) summation was used for long range electrostatics.

```
gmx grompp -v -f st -c b4em -o em -p topol
gmx genion -s em.tpr -o ion_b4em -p topol -pname NA -np X
```
or
```
gmx genion -s em.tpr -o ion_b4em -p topol -nname CL -nn X
```

Before launching productive MD calculations energy minimization must be performed on the content of the box. Here, commands of a two step minimization are shown including steepest descent and a conjugated gradient runs. Minimizations are performed by the gmx mdrun program and the binary inputs are produced by grompp. Note, that –c ion_b4em can be specified instead of –c b4em if you have added neutralizing ions to your box. It is important to put position restraints on all solute heavy atoms with force constant of 10^3 $kJmol^{-1}nm^{-2}$ in both minimization steps. Position restraints must be defined in both mdp file ("define = -DPOSRES"), and the position restraints topology file ("posre*.itp" generated at the first step) must be present in the working directory. PME (Particle Mesh-Ewald) summation was used for long range electrostatics. Van der Waals and Coulomb interactions had a cut-off at 11 Å. Convergence threshold of the first step was set to 10^3 kJ mol^{-1} nm^{-1}, in the second step it was set to 10 kJ mol^{-1} nm^{-1} .

```
gmx grompp -v -f st -c ion_b4em -o st -p topol.top
gmx mdrun -v -s st -o st -c after_st -g st
gmx grompp -v -f cg -c after_st -o cg -p topol.top
gmx mdrun -v -s cg -o cg -c after_cg -g cg
```

Final inputs are produced from the energy minimized system after_cg and MD calculation can be launched using mdrun. Trajectory of the system is stored in an md.trr file specified at switch –o. It is important to put position restraints on all solute heavy atoms with force constant of 10^3 $kJmol^{-1}nm^{-2}$ during MD simulation. Position restraints must be defined in the mdp file ("define = -DPOSRES"), and the position restraints topology file ("posre*.itp" generated at the first step) must be present in the working directory. PME (Particle Mesh-Ewald) summation was used for long range electrostatics. Van der Waals and Coulomb interactions had a cut-off at 11 Å. For temperature-coupling the velocity rescale algorithm was used. Solute and solvent were coupled

separately with a reference temperature of 300 K and a coupling time constant of 0.1 ps. Pressure was coupled the Parrinello-Rahman algorithm with a coupling time constant of 0.5 ps, compressibility of 4.5×10^{-5} bar^{-1} and reference pressure of 1 bar.

```
gmx grompp -f md -o md -c after_cg -r after_cg -p topol.top -maxwarn 1
gmx mdrun -v -s md -e md -o md -c after_md -g md.log
```

Once you have your trajectory in an md.trr file fast conversions are recommended using trjconv into the portable xtc format.
Such conversions handle periodic boundary effects, center the system in the box and fit target molecules in subsequent frames on the top of the first frame. First, you will need to fit your trajectory onto the initial structure with waters using Gromacs tools. The following command specifies multi-frame fit using CA atoms of the protein backbone.

```
gmx confrms -one -f1 target.pdb -f2 md.tpr -o fit.pdb < EOF
3
3
EOF
gmx editconf -label A -f fit.pdb -o fit.pdb
gmx trjconv -f md.trr -s md.tpr -o pbc_1.xtc -pbc whole < EOF
0
EOF
gmx trjconv -f pbc_1.xtc -s md.tpr -o pbc_2.xtc -pbc cluster < EOF
1
0
EOF
gmx trjconv -f pbc_2.xtc -s md.tpr -o pbc_3.xtc -center -pbc mol -ur compact
< EOF
1
0
EOF
```

In the last command line, instead of –o system.xtc, a switch –o system_mdl.pdb or –sep –o system_.pdb can be specified for an NMR-type PDB file or separate PDB files, respectively. A topology file system_tpy.pdb can be easily produced for MobyWat by trjconv.

```
gmx trjconv -f pbc_3.xtc -s fit.pdb -o system.xtc -fit progressive < EOF
3
0
EOF
gmx trjconv -f pbc_3.xtc -s fit.pdb -o system_tpy.pdb -b 0 -e 0 -fit
progressive < EOF
3
0
EOF
```

The system.xtc (or the corresponding pdb files) and system_tpy.pdb can be used as input for MobyWat.
The system.xtc file is available at the following link:
https://drive.google.com/drive/folders/179LWLgZF2fdFiRH2rl9aBTbB6S8qydNv?usp=sharing

From the MD frames (system.xtc and system_tpy.pdb) MobyWat predicts water positions for the target surface based on clustering. If system_ref.pdb file exists with reference water positions after prediction the program automatically calculates matches between reference and predicted positions based on match tolerance (mtol). In our case target is denoted as chain A,

all 1000 frames are used with clustering tolerance (ctol) 1Å, prediction tolerance 2.5Å, maximum of 5Å distance (dmax) from heavy atoms and 1.5Å prediction tolerance. Output files are **O_system_prIDA.pdb** with all predicted water positions and **O_system_mtIDA.lst** with matched water IDs.

```
mobywat -t [A] -w Auto -n 0-1000 -f system.xtc -tpy system_tpy.pdb -r
system.ref -ctol 1 -ptol 2.5 -dmax 5 -mtol 1.5 -m Prediction -cls IDa
```

Step 3 – Merging and refinement

Software used: PyMol (https://www.schrodinger.com/products/pymol)
MobyWat 1.1 (http://mobywat.com/)
GROMACS 2020 (https://www.gromacs.org/)
Available files: **step3.tgz**

The dry docked complex and the hydrated target were opened in PyMol. Action: alignment was selected, and the hydrated target was aligned by Cα atoms to the dry docked complex (**dry_complex.pdb**), then with Save Molecule>hydrated target (current form), the hydrated target within the same coordinate system as the dry docked complex was saved.
In a text editor the water oxygen coordinates from the hydrated target were copied into the pdb file of the dry complex structure (**complex_all_wat.pdb**).
Then the Editing mode of MobyWat was used to eliminate overlapping water oxygens:

```
mobywat -f complex_all_wat.pdb -l [F] -t [A,B,C,D] -w Auto -m Editing -dmax
80 -dmin 1.75
```

After the –t the chains of the target are given in the bracket, and after –l the chain that contains the ligand is given.

This gives a file, that contains all the water oxygen positions without the overlapping ones (**O_0106.dtd**), practically this can replace all the water oxygen positions in the input file, and in a text editor in can be saved as a separate file, ready for the minimization (**complex_edited.pdb in the „minim" folder**).

Robust Refinement was performed as follows (from the input of the previous paragraph):

The first three commands are similar to as described in **Step 2**, for the explanation of each command please see the previous section.

Hydrogen atoms are added, TIP3P water model and AMBER99sb-ILDN force field are specified by pdb2gmx. Ligand topology files necessary for the pdb2gmx command are included, with a readme.txt about their usage in step3.tgz /topol folder.

```
gmx pdb2gmx -water tip3p -ff amber99sb-ildn -ignh -f complex_edited.pdb
```

The system is placed in a dodecahedral (to lower computational costs) shaped box:

```
gmx editconf -o -d 1 -bt dodecahedron -f conf.gro
```

The box is filled up by water molecules (apart from the predicted ones, which are kept):

```
gmx solvate -cp out -cs -o b4em -p topol
```

The system is neutralized by counter-ions (if necessary), write the necessary number of Cl or Na ions instead of X, and only use one of the lines:

```
gmx grompp -v -f steep -c b4em -o em -p topol
gmx genion -s em.tpr -o ion_b4em -p topol -pname NA -np X
gmx genion -s em.tpr -o ion_b4em -p topol -nname CL -nn X
```

Energy minimization (these can be performed on a PC as well in a short period of time):
sd1+cg1:

```
gmx grompp -v -f st1 -c ion_b4em -r ion_b4em -o st1 -p topol.top
gmx_d mdrun -v -s st1 -o st1 -c after_st1 -g st1
```

```
gmx grompp -v -f cg1 -c after_st1 -r after_st1 -o cg1 -p topol.top
gmx_d mdrun -v -s cg1 -o cg1 -c after_cg1 -g cg1
```
short md:
```
gmx grompp -f md -o md -c after_cg1 -r after_cg1 -p topol.top -maxwarn 2
gmx mdrun -v -s md -e md -o md -c after_md -g md.log
```
sd2+cg2:
```
gmx grompp -v -f st2 -c after_md -r after_md -o st2 -p topol.top
gmx_d mdrun -v -s st2 -o st2 -c after_st2 -g st2
gmx grompp -v -f cg2 -c after_st2 -r after_st2 -o cg2 -p topol.top
gmx_d mdrun -v -s cg2 -o cg2 -c after_cg2 -g cg2
```

After the 3rd step, the **after_cg2.gro** is created.

Step 4 – Generating MD snapshots of the target-ligand complex

Software used: GROMACS 2020 (https://www.gromacs.org/)
Available files: **step4.tgz, system.xtc** (a separate link is provided for the xtc file below)

From the **after_cg2.gro** file, an **ama.tpr** file is created (ligand topology files are available in the step3.tgz /topol folder):

```
gmx grompp -f md1 -o ama -c after_cg2 -r after_cg2 -p topol.top -maxwarn 2
```

And then the MD simulation if performed on a super computer preferably:

This line has to be embedded into a running script of the preferances of the given super computer, can be run on multiple cores and nodes:

```
gmx mdrun -v -s ama -e md_ama -o md_ama -c md_ama -g md_ama.log
```

After the MD simulation is finished these GROMACS command lines were used, similarly as in Step 2:

```
gmx confrms -one -f1 complex_edited.pdb -f2 ama.tpr -o fit.pdb << EOF
3
3
EOF
gmx trjconv -f md_ama.trr -s ama.tpr -o pbc_1.xtc -pbc whole << EOF
0
EOF
gmx trjconv -f pbc_1.xtc -s ama.tpr -o pbc_2.xtc -pbc cluster <<EOF
1
0
EOF
gmx trjconv -f pbc_2.xtc -s ama.tpr -o pbc_3.xtc -center -pbc mol -ur compact
<<EOF
1
0
EOF
gmx trjconv -f pbc_3.xtc -s fit.pdb -o system.xtc -fit progressive <<EOF
3
0
EOF
```

This way, the frames are placed back to the middle of the simulation box, broken periodic boundaries are fixed and all frames are aligned back to the original, to have matching coordinate systems and easier handling of the files.
The system.xtc file is available at the following link:
https://drive.google.com/drive/folders/1f0aEKbGF8MARlpyPdBDtYF7GCGueSaRD?usp=sharing

From the system.xtc file, separate ligand, or protein, or water files can also be extracted:

```
gmx trjconv -sep -f system.xtc -s fit.pdb -o ligand.pdb
```

In our case 13 was the number of the ligand cluster, this might vary of course.

Step 5 – The selection of the representative ligand binding modes from the MD trajectory file

Software and scripts used: GROMACS 2020 (https://www.gromacs.org/)
Shell scripts (*.sh) provided in step5.tgz file
Program rmsd provided in step5.tgz (exec. bin for Linux systems)
Available files: **step5.tgz**

Instructions to calculate an average coordinate file, its closest structure and an RMSD for the reference crystallographic structure.

1. From the starting position of the MD simulation (**start.pdb**), to create a **fit.pdb** for least squares fitting.

```
gmx confrms -one -f1 start.pdb -f2 ama.tpr -o fit.pdb <<EOF
3
3
EOF
```

2. A system.xtc file is created, from which the ligand structures will be extracted. All of the frames are fitted to fit.pdb by the following steps:

```
gmx trjconv -f md_ama.trr -s ama.tpr -o pbc_1.xtc -pbc whole <<EOF
0
EOF
gmx trjconv -f pbc_1.xtc -s ama.tpr -o pbc_2.xtc -pbc cluster <<EOF
1
0
EOF
gmx trjconv -f pbc_2.xtc -s ama.tpr -o pbc_3.xtc -center -pbc mol -ur compact
<<EOF
1
0
EOF
gmx trjconv -f pbc_3.xtc -s fit.pdb -o system.xtc -fit progressive <<EOF
3
0
EOF
```

3. Now use the following scripts calling the “echo”, “grep”, “awk”, “cat” and “rm” programs, that are available within Linux operating system, and the gmx trjconv program:

```
./average_1.sh
./average_2.sh
./average_3.sh
```

4. The previous steps will create average.txt, but these are only coordinates, so **average.pdb** (create a pdb file from the coordinates) is necessary to continue the following steps:

```
./calculation_1.sh
./calculation_2.sh
./calculation_3.sh
```

5. This results in result_1 _2 _3.log, these three files are copied into one: **result_final.log**, which is imported into an excel sheet: **excel results.xlsx**

In the excel file the RMSDs are numbered and the smallest is selected by the individual order function of the program.

Optionally, the structure with the smallest RMSD can be selected and compared to the reference ligand:

```
rmsd start -a sep20.pdb -r reference.pdb
```

sep20.pdb is the ligand from the 20th frame.
This last step gives the results shown in Table 4 of the original publication