Tutorial for Hi-CO simulated-annealing molecular dynamics simulation

Masae Ohno, Tadashi Ando, David G. Priest and Yuichi Taniguchi

This tutorial will explain how to use the simulated annealing-molecular dynamics (SA-MD) simulation to build 3D chromosome structures at nucleosome level based on potentials derived from Hi-CO experiments. Here we will build a structure of chromosome I of yeast as an example. Each step will contain a brief explanation of input and output. Commands you need to type are shown with a gray background throughout the tutorial. Details of simulation methods are described in the paper "Ohno et al., *Nature Protocols*. 20XX, **XXX**(X):XXX-XXX" (the information will be updated after the publication). The authors request that all published works which utilize this software include a citation of the paper.

Step 0: Computer and simulation programs.
Programs are pre-compiled for a x86_64 CPU architecture on Linux. The SA-MD simulation program is parallelized with OpenMP. For efficient computations, we recommend use of computers with multi-core CPUs.

After downloading `HiCO_SAMD_Programs.zip` on a Linux computer, you need to unzip the file.

1-1) `unzip HiCO_SAMD_Programs.zip`

1-2) Change directory to `HiCO_SAMD_Programs`:
`cd HiCO_SAMD_Programs`

1-3) Type:
`chmod u+x 0_Programs/bin/*.exe */*.sh`

`HiCO_SAMD_Programs` has following directories and files:

```
HiCO_SAMD_Programs/
  ├ 0_Programs/
  │      └ bin/
  │          ├ LD.exe
  │          ├ Setup.exe
  │          └ Param.exe
  ├ 1_Setup/
  │      ├ Inward.tsv
  │      ├ Outward.tsv
  │      ├ Tandem-.tsv
  │      ├ Tandem+.tsv
  │      ├ model_linker_length_ncp_133bp.out
  │      ├ Setup.sh
  │      └ ChrI_Results
  │          ├ ChrI.dat
  │          └ ChrI.hic
  │
  ├ 2_Minimization_1/
  │      ├ 3Dcoor.tsv
  │      ├ formula.tsv
  │      ├ Param.sh
  │      ├ LD.sh
  │      └ ChrI_Results
  │          ├ out.par
  │          ├ out.psf
  │          ├ out.pqr
  │          ├ template.pqr
  │          ├ LD.out
  │          ├ TRAJ.dcd
  │          └ RESTART.rst
  │
```

```
├ 3_Minimization_2/
│      ├ 3Dcoor.tsv
│      ├ formula.tsv
│      ├ Param.sh
│      ├ LD.sh
│      └ ChrI_Results
│           ├ out.par
│           ├ out.psf
│           ├ out.pqr
│           ├ template.pqr
│           ├ LD.out
│           ├ TRAJ.dcd
│           └ RESTART.rst
│
├ 4_SAMD_1/
│      ├ formula.tsv
│      ├ LD.sh
│      └ ChrI_Results
│           ├ LD.out
│           ├ TRAJ.dcd
│           └ RESTART.rst
│
└ 5_SAMD_2/
       ├ formula.tsv
       ├ LD.sh
       └ ChrI_Results
            ├ LD.out
            ├ TRAJ.dcd
            └ RESTART.rst
```

In `0_Programs/bin` directory, there are three programs, `LD.exe`, `Setup.exe`, and `Param.exe`, which are pre-compiled execution files. `LD.exe` is a program for performing Langevin dynamics simulations and minimizations. `Setup.exe` is a program for extracting Hi-CO experimental data of the target region of chromosomes to be modeled in the simulations. `Param.exe` is used for generating parameter files required for simulations (`LD.exe`) from Hi-CO data extracted by `Setup.exe` program. Five directories, `1_Setup`, `2_Minimization_1`, `3_Minimization_2`, `4_SAMD_1`, and `5_SAMD_2`, correspond to steps of the simulation. In each of the directories, there are sub-directories `ChrI_Results`, in which there are output files we obtained as references. Let's assume that you are in the directory `HiCO_MD_Programs`.

Step 1: Setup

In this step, you extract Hi-CO experimental data of the target region, Chromosome I in this tutorial, from raw Hi-CO data files, `Inward.tsv`, `Outward.tsv`, `Tandem-.tsv`, and `Tandem+.tsv`.

1-1) Change directory to `1_Setup`:
`cd 1_Setup`

1-2) Cut out the lines regarding chromosome I from `model_linker_length_ncp_133bp.out` file and save it as `ChrI.dat` using a text editor. Here, `model_linker_length_ncp_133bp.out` is a table that lists the index of the nucleosome locus and its corresponding genomic position (bp).

1-3) Edit the third line of `Setup.sh` using a text editor as follows:
`NAME=ChrI`.

1-4) Run Setup program.
`./Setup.sh`

1-5) Now, you can see an output file, `ChrI.hic`. To check files, type
`ls`

Step 2: Minimization_1

In the second step, you generate a parameter file required for MD simulations and build an initial structure of chromosome I using `Param.sh`. You also perform energy minimization of the initial structure using `LD.sh`. The generated initial conformation has unphysical overlaps or clashes between particles that cause numerical instability in simulations. To remove such clashes, an energy minimization process is necessary. In this tutorial, energy minimizations are performed two times with different parameters that control the strength of excluded volume effects. The overall structure is superimposed into a low-resolution 3D model obtained in a previous Hi-C study (Lesne A et al, Nat Methods, 11, 1141-1143, 2014). This superimposition is performed by restraining positions of histone particles at every 5,000 bp to the coordinate of the low-resolution model stored in `3Dcoor.tsv` by a harmonic potential.

2-1) Change directory to `2_Minimization`:
`cd ../2_Minimization_1`

2-2) Edit `Param.sh` as follows:
`NAME=ChrI`
`ELJ=0.1` (set a small value to avoid numerical instability at the first stage of minimization.)
etc.

2-3) Run Param program.
`./Param.sh`

2-4) You now find four output files.
`ls`
`out.pqr`: Initial coordinate of histone and DNA particles, which can be visualized in 3D with a program such as VMD (see below).
`template.pqr`: Coordinate of histone particles in the low-resolution model for the superimposition.
`out.par`: Parameter file required for following simulations.
`out.psf`: Structure information file required for visualizing simulation trajectories (see below).

Next, you run energy minimization.

2-5) Edit `LD.sh`.
In the file, default values for simulation parameters are already written. Meanings of the parameters are also explained in the file. For parallel execution, change the value "`OMP_NUM_THREADS`" at line 3, which depends on the computer you use. If you do not know the number of CPU cores of your machine, please type "`lscpu`" to display information about the CPU architecture.

2-6) Run minimization.
`./LD.sh`
(formula.tsv file is necessary to run LD.exe. Please do not remove the file.)

2-7) You find three output files.
`ls`
`LD.out`: Output of the simulation.
`TRAJ.dcd`: Trajectory of coordinates of particles during the simulation.
`RESTART.rst`: Restart file for next simulation.

Step 3: Minimization_2

In the third step, you perform the 2nd stage minimization, where the parameter controlling the strength of excluded volume is set to 10.0.

3-1) Change directory to `3_Minimization_2`:
`cd ../3_Minimization_2`

3-2) Edit `Param.sh` to set `ELJ=10.0`.

3-3) Run Param program.
`./Param.sh`
3-4) You find four output files, `out.pqr`, `template,pqr`, `out.par`, and `out.psf`.
`ls`

Next, you run energy minimization.

3-5) Edit `LD.sh`.
3-6) Run minimization.
`./LD.sh`
3-7) You find three output files, `LD.out`, `TRAJ.dcd`, and `RESTART.rst`.
`ls`

Step 4: SAMD_1
In the fourth step, you perform a Langevin dynamics simulation from temperature 600 K to 300 K for 125,000 steps. The temperature linearly decreases with steps. To randomize the positions of particles of the minimized structure, this high temperature at the initial step is necessary. Parameter file generated in Step 3 is used.

4-1) Change directory to `4_SAMD_1:`
`cd ../4_SAMD_1`
4-2) Edit `LD.sh.` Then run Langevin dynamics simulation:
`./LD.sh`
4-3) You find three output files, `LD.out`, `TRAJ.dcd`, and `RESTART.rst`.
`ls`

Step 5: SAMD_2
In the fifth step, you perform Langevin dynamics simulation from temperature 300 K to 0.1 K for 125,000 steps.

5-1) Change directory to `5_SAMD_2:`
`cd ../5_SAMD_2`
5-2) Edit `LD.sh.` Then run Langevin dynamics simulation:
`./LD.sh`
5-3) You find three output files, `LD.out`, `TRAJ.dcd`, and `RESTART.rst`.
`ls`

Visualizing SA-MD simulation results using VMD
You can use the VMD software to visualize the structure and its trajectory as the time steps proceed. For downloading and installation, please visit on the VMD original home page at https://www.ks.uiuc.edu/Research/vmd/.

To visualize the trajectory, launch VMD -> File -> New Molecule -> Click Browse button and select `3_Minimization_2/out.psf`. Click Load button. -> Click Browse button and select `4_SAMD_1/TRAJ.dcd`. Click Load button. -> Click Browse button and select `5_SAMD_2/TRAJ.dcd`. Click Load button. -> Click right arrow button in the main console. Now, you can see a trajectory movie.