GENESIS

Generalized-ensemble simulation system

理化学研究所 計算科学研究機構
粒子系生物物理研究チーム
Computational Biophysics Research Team
RIKEN Advanced Institute for Computational Science

GENESIS tutorial 1:

Usage of machines, Compile of GENESIS, and Basic MD

Workshop “Frontiers in Computational Biophysics and Biochemistry”

2017/02/28

C. Kobayashi

Contents

- 13:30 – 15:00 GENESIS tutorial 1
  - Overview of GENESIS & tutorials
  - Usage of machine for hands-on tutorial
  - Compile of GENESIS
  - Tutorial 1: Basic MD

GENESIS

Generalized-ensemble simulation system

理化学研究所 計算科学研究機構
粒子系生物物理研究チーム
Computational Biophysics Research Team
RIKEN Advanced Institute for Computational Science

A software package for Molecular Dynamics (MD) simulation of biological system

Designed to extend limitations in system size and time scale of MD simulation.

- Highly parallelized schemes for supercomputers like K computer
- Development of enhanced sampling algorithms using multiple replicas
- Use of all atom force fields (FFs) and Coarse-Grained (CG) models

GENESIS

Generalized-ensemble simulation system

理化学研究所 計算科学研究機構
粒子系生物物理研究チーム
Computational Biophysics Research Team
RIKEN Advanced Institute for Computational Science

- Development :
  Computational Biophysics Research Team, RIKEN AICS
  - Project leader : Y. Sugita
  - Main developers : J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Ando, M. Kamiya, T. Yoda, M. Feig
- Current version : 1.1.3
- License : GPLv2
- Website : http://www.aics.riken.jp/labs/cbrt/

  You can download source code and tests
- Publication
  - [Version 1.0] J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, and Y. Sugita, *WIREs Comput. Mol. Sci.*, **5**, 310-323 (2015).

GENESIS

Generalized-ensemble simulation system

理化学研究所 計算科学研究機構
粒子系生物物理研究チーム
Computational Biophysics Research Team
RIKEN Advanced Institute for Computational Science

- Two MD programs
  - SPDYN (**SP**atial decomposition **DYN**amics simulation)
    - High performance and scalability
    - All atom FFs
  - ATDYN (**AT**omic decomposition **DYN**amics simulation)
    - Usage of All atom FFs and CG models
    - Readable code for easy implementation by users

Feature	ATDYN	SPDYN
Decomposition scheme	Atomic decomposition	Domain decomposition
REMD/string method	○	
All atom FFs	CHARMM, AMBER	
CG models	MARTINI, SMOG, Clementi, Karanicolas-Brooks	MARTINI
r-RESPA	×	○
GPU computation	×	○
Mixed precision	×	○

- Available functions in both ATDYN & SPDYN
  - Minimization
    - Steepest Decent
  - Integrator
    - Leapfrog
    - Velocity Verlet
  - Ensemble
    - NVE
    - NVT
      - Langevin
      - Bussi
      - Berendsen
    - NPT
      - Langevin-piston
      - Bussi
      - Isotropy of Simulation box
        Isotropic (default), Semi-iso, An-iso, XY-fixed
  - Constraints
    - SHAKE (Leapfrog)
    - RATTLE (Velocity Verlet)
    - SETTLE
  - FFT
  - Restraint functions
    - Position
    - Bond
    - Angle
    - Dihedral angles
    - RMSD
  - Steered MD
  - Targeted MD

- Other analysis tools
  - Convert trajectories（trjcnv）
    - ✧ crd_convert
    - ✧ pcrd_convert（parallel I/O）
    - ✧ remd_convert（REMD）
  - Analyze trajectories（trjana）
    - ✧ mbar_analysis
    - ✧ pathcv_analysis
    - ✧ pmf_analysis
    - ✧ gval_analysis
    - ✧ rmsd_analysis（RMSD、RMSF）
    - ✧ trj_analysis（distance, angle, etc）
    - ✧ wham_analysis
  - Convert restart files（rstcnv）
    - ✧ rst_convert
    - ✧ rst_upgrade
  - Principal Component Analysis (PCA)（pcaana）
    - ✧ avecrd_analysis
    - ✧ eigmat_analysis
    - ✧ flccrd_analysis
    - ✧ prjcrd_analysis
  - Input generator of String method
    - ✧ rpath_generator

- **Tutorials**:

  2/28

  – Tutorial 1 (13:30-15:00) **Basic MD** (C. Kobayashi)

  – Tutorial 2 (15:30-17:00) **Advanced MD** (J. Jung)

  3/1

  – Tutorial 3 (13:40-15:10) **Replica exchange MD** (T. Mori)

  – Tutorial 4 (15:30-17:00) **String method** (Y. Matsunaga)

- **Lectures**

  3/1

  – Lecture 1 (9:30-10:00) **High performance MD** (J. Jung)

  – Lecture 2 (10:00-10:30) **Replica exchange MD** (T. Mori)

  – Lecture 3 (10:30-11:00) **String method** (Y. Matsunaga)

  – Lecture 4 (11:20-11:50) **QM/MM** (Y. Yagi)

  – Lecture 5 (11:50-12:20) **CG and Brownian dynamics** (C. Kobayashi)

Contents

- 13:30 – 15:00 GENESIS tutorial 1
  - Overview of GENESIS & tutorials
  - Usage of machine for hands-on tutorial
  - Compile of GENESIS
  - Tutorial 1: Basic MD

Access cloud machine

In this tutorial, Microsoft azure cloud machines are used.

0. Please make sure if your machine connects to internet.

1. Please login a server if you have your account & password.

```
% ssh userXX@XXXXXXXXXXXXXXXXXXXXXXXXXXXX
        login node
% ssh userXX@XX
        calculation node
```

userXX: your account
Please check back-side of your name card

2. Please check your home directory if tutorial files are prepared.

```
% pwd
/home2/userXX
% ls -l
drwxr-xr-x. 4 userXX user  46 Feb 25 09:35 Compile
drwxr-xr-x. 2 userXX user 123 Feb 25 10:21 env
drwxr-xr-x. 8 userXX user 121 Feb 21 19:16 Tutorial_1
drwxr-xr-x. 3 userXX user  19 Feb 22 16:53 Tutorial_2
drwxr-xr-x. 4 userXX user  32 Feb 22 19:57 Tutorial_3
…
```

2’. If you use your machine, you can download all stuff from a link from GENESIS/Workshop websites:

Files for this tutorial

Compile/: Compile of GENESIS

- genesis-1.1.3
  Recent GENESIS code
- tests-1.1.3
  Compile test

Tutorial_1/: Basic MD

- 1_setup
- 2_minimization
- 3_heating
- 4_equilibration
- 5_production
- 6_analysis

env/: Set of environment (Do not remove & change!)

- setpath*

Contents

- 13:30 – 15:00 GENESIS tutorial 1
  - Overview of GENESIS & tutorials
  - Usage of machine for hands-on tutorial
  - Compile of GENESIS
  - Tutorial 1: Basic MD

Compile of GENESIS (1)

Go to src in GENESIS

```
% cd Compile/genesis-1.1.3/src
```

Source tree of GENESIS

```
.:
COPYING
README
src/          source codes
bin/          applications
              (after compile)

./src:
GENESIS_VERSION
Makefile.am
Makefile.in
aclocal.m4
analysis/     Analysis tools
atdyn/        Code for atdyn
bootstrap
cleanup
config.h.in
configure
configure.ac
depcomp
install-sh
lib/          Common modules
Missing
spdyn/        Code for spdyn

./src/analysis:
Makefile.am
Makefile.in
libana/       Common module of
              analysis
pcaana/       code for PCA
rpath_generator/
              rpath_generator
rstcnv/       rst_convert
trjana/       trj_analysis
trjcnv/       crd/pcrd/remd_convert
```

Compile of GENESIS (2)

Recommend compilers

Intel , Fujitsu compiler

gfortran newer than 4.4.7

lapack, blas libraries are required to compile genesis

1. Compile option is set by configure (autoconf), and package is generated by make command

```
% ./bootstrap
% ./configure
% make
% make install
```

←CPU only usage with double precision

2. All packages are in genesis-1.1.3/bin

By compile of GENESIS MD, ATDYN, SPDYN, and the other analysis tools are generated.

Compile of GENESIS (3)

How to compile depending on computational environments

0. You can see help;

```
% ./configure --help
```

1. Compile without any options (all programs with double precision)

```
% ./configure
% make install
```

2. Compile of spdyn with mixed precision **(SPDYN only!)**

```
% ./configure --enable-single
% make install
```

3. Compile of spdyn with GPU computation with mixed precision **(SPDYN only!)**

```
% ./configure --enable-gpu --enable-single --with-cuda=/usr/local/cuda-8.0
% make install
```

4. Compile with 'debug' mode

```
% ./configure --enable-debug=3
% make install
```

Compile test of GENESIS (1)

To confirm compile of GENESIS, we strongly recommend to execute ‘regression test’ prepared by the developers team.
You can download a tarball for the regression test from ‘GENESIS’ website.

```
% cd Compile/tests-1.1.3/regression_test
```

Tree of regression tests of GENESIS-1.1.3

./regression_test:		test_remd.py	scripts for REMD
build/	Inputs for regression tests	test_rpath.py	scripts for string method
charmm.py		test_rpath_atdyn/	tests for rpath(atdyn)
genesis.py		test_rpath_spdyn/	tests for rpath(spdyn)
param/	FF parameters	test_spdyn/	tests for functions only in spdyn
test.py	regression test’s script		
test_atdyn/	tests for functions only in atdyn		
test_common/	tests for common functions in spdyn & atdyn		
test_nonstrict.py			
test_parallel_IO/	tests for parallel I/O		
test_remd/	tests for REMD		
test_remd.csh			

Compile test of GENESIS (2)

How to execute ‘regression tests’

```
% cd ~/Compile/tests-1.1.3/regression_tests
% export PATH_GENESIS=/home2/data/genesis/bin.CPU.dp
% export OMP_NUM_THREADS=3
% ./test.py “mpirun -np 8 ${PATH_GENESIS}/atdyn”
% ./test.py “mpirun -np 8 ${PATH_GENESIS}/spdyn”
```

In this tests, please set # of processes (MPI) = 8

# of threads (KMP/OMP) should be less than # of processes due to speed.

How to execute test for GPU computations

```
% export PATH_GENESIS=/home2/data/genesis/bin.GPU.sp
% export OMP_NUM_THREADS=3
% ./test.py “mpirun -np 8 ${PATH_GENESIS}/spdyn” gpu
```

How to execut test for REMD, R-PATH

```
% export PATH_GENESIS=/home2/data/genesis/bin.CPU.dp
% export OMP_NUM_THREADS=3
% ./test_remd.py “mpirun -np 8 ${PATH_GENESIS}/spdyn”
% ./test_remd.py “mpirun -np 8 ${PATH_GENESIS}/atdyn”
% ./test_rpath.py “mpirun -np 8 ${PATH_GENESIS}/spdyn”
% ./test_rpath.py “mpirun -np 8 ${PATH_GENESIS}/atdyn”
```

Note: double- or mixed- precision will be detected by automatically.

Basic Usage of GENESIS

How to execute spdyn/atdyn

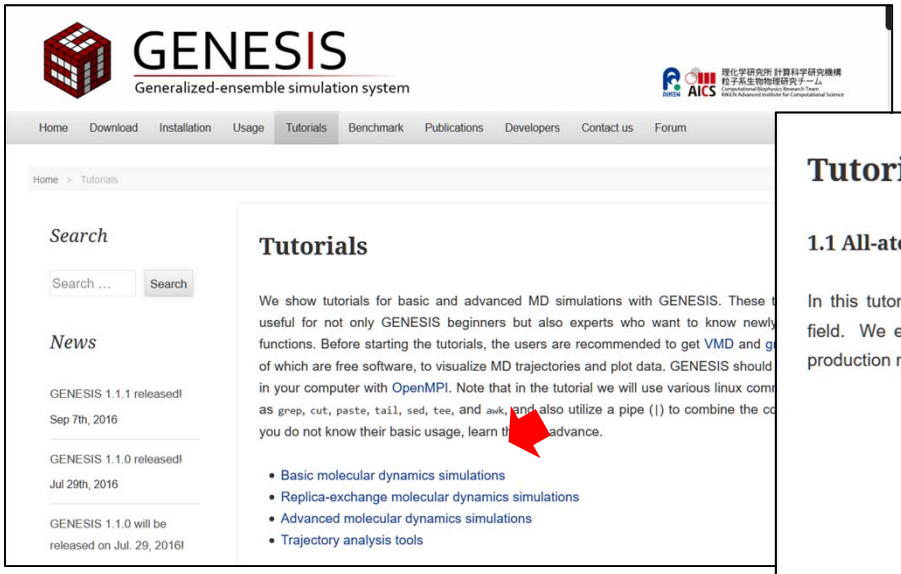
```
% export PATH_GENESIS=/home2/data/genesis/bin.CPU.dp
% export OMP_NUM_THREADS=3
% mpirun -np 8 ${PATH_GENESIS}/atdyn INP
% mpirun -np 8 ${PATH_GENESIS}/spdyn INP
```

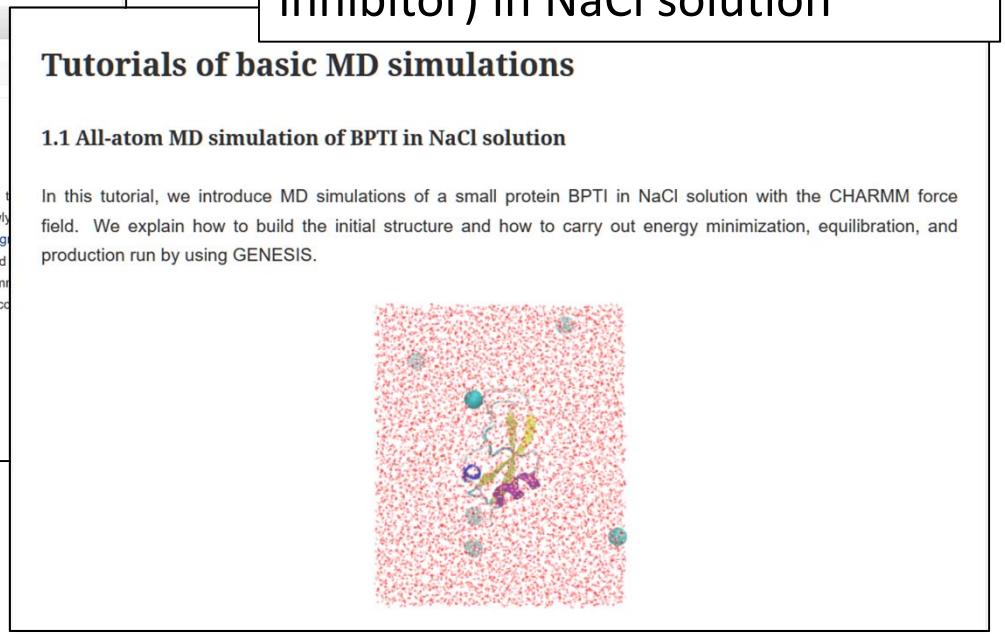
Since the machine is quit slow to access files, the regression test are skipped today.

Tutorial 1: MD of soluble protein

http://www.aics.riken.jp/labs/cbrt/tutorial/basic_md_tutorials/tutorial-1-1/

All atom MD simulation of BPTI (Bovine pancreatic trypsin inhibitor) in NaCl solution





1. Generation Simulation system (1_setup/)

   GENESIS does not have a builder for simulation sytem and read molecule files generated by CHARMM/VMD/AMBER/Gromacs.

   Today, we use molecule files generated by VMD.
   You can find an instruction for building in *"1.1.1 building a simulation system"*

Tutorial 1: Minimization(2_minimization/)

The initial structure often contains non-physical steric clashes and/or unstable geometry.

At first, we need to remove such clashes by minimizing potential energy of the simulation system before production runs.

Steepest Decent method is available in GENESIS 1.1

Tutorial_1/2_minimization/
- run.inp : control files of GENESIS
- run.sh : job scripts
- output/ : results

Control file for Minimization

```
[INPUT]
topfile = ../1_setup/top_all27_prot_lipid.rtf # topology file
parfile = ../1_setup/par_all27_prot_lipid.prm # parameter file
psffile = ../1_setup/ionize.psf            # protein structure file
pdbfile = ../1_setup/ionize.pdb             # PDB file
reffile = ../1_setup/ionize.pdb             # reference for restraints

[OUTPUT]
dcdfile = run.dcd  # DCD trajectory file
rstfile = run.rst  # restart file

[ENERGY]
electrostatic    = PME      # [CUT
switchdist       = 10.0     # switch distance
cutoffdist       = 12.0     # cutoff distance
pairlistdist     = 13.5     # pair-list distance
contact_check    = YES
pme_ngrid_x      = 72       # grid size_x in [PME]
pme_ngrid_y      = 80       # grid size_y in [PME]
pme_ngrid_z      = 72       # grid size_z in [PME]

[MINIMIZE]
nsteps       = 1000     # number of steps
eneout_period = 100     # energy output period
crdout_period = 100     # coordinates output period
rstout_period = 1000    # restart output period

[BOUNDARY]
type         = PBC      # [PBC,NOBC]
box_size_x    = 70.8250  # box size (x) in [PBC]
box_size_y    = 83.2579  # box size (y) in [PBC]
box_size_z    = 69.0930  # box size (z) in [PBC]

[SELECTION]
group1       = an:CA | an:C | an:O | an:N  # index of restraint group 1

[RESTRAINTS]
nfunctions    = 1     # number of functions
function1     = POSI  # restraint function t
constant1     = 10.0  # force constant
select_index1 = 1     # restrained group
```

Use Particle Mesh Ewald (PME) for long-range interaction

Positional Restraints
Target atoms are selected in [SELECTION]

Report too long/short distances or steric crash.
In addition, large forces due to steric crash are tempered during simulations

Numbers of PME grids: the number should be multiples of 2,3, and 5 due to restriction of FFT library.
The numbers are selected automatically from # of domains, pme_max_spacing

Number of steps for minimization

Executing Minimization

Execute minimization

```
% cd  ~/Tutorial_1/2_minimization
% ./run.sh
```

When your job is done, these outputs are shown.

run.dcd : Trajectory with dcd format (binary)
run.out : Output (energy/property) of GENESIS (ascii)
run.rst : Restart file for next simulation (binary)

Output(run.out)

```
*************************************************
*                                               *
*            GENESIS SPDYN                      *
*                                               *
*     A Molecular Dynamics Simulator with       *
*       Spatial Decomposition Scheme            *
*                                               *
*          Developed by RIKEN AICS              *
*                                               *
*************************************************
```

Output file of GENESIS

There are seven paragraphs in GENESIS ([STEP 0]-[STEP 6])

[STEP 0] : Information for compile & computational environment

[STEP 1] : Parameters that you set

[STEP 2] : parallelization (numbers of process and threads)

[STEP 3] : Information of molecule and energy functions

[STEP 4] : Single point energy by initial coordinates

[STEP 5] : Show energies and properties during simulation

```
INFO:     STEP   POTENTIAL_ENE         RMSG         BOND        ANGLE    UREY-BRADLEY      DIHEDRAL      IMPROPER          CMAP       VDWAALS         ELECT RESTRAINT_TOTAL
 --------------- --------------- --------------- --------------- ---------------
INFO:        0   -101597.1585       30.2479     11247.1385     2939.9532       74.9561      260.8373       62.6065      -72.0093    11798.1692   -127908.8099        0.0000

INFO:      100   -114456.2091        5.4597      3977.1511     2353.8183       51.5709      256.4578       22.6346      -74.3462     9730.9841   -130775.5114        1.0317
```

[STEP 6] : ELAPS

```
Output_Time> Averaged timer profile (Min, Max)
 total time     =    137.159
  setup         =      5.843
  dynamics      =    131.316
   energy       =    122.896
   integrator   =      0.838
   pairlist     =      5.774 (      5.629,      5.963)
 energy
  bond          =      0.657 (      0.587,      0.708)
  angle         =      0.717 (      0.696,      0.751)
  dihedral      =      1.275 (      1.236,      1.295)
  nonbond       =    109.986 (    109.658,    110.141)
   pme real     =     83.219 (     81.500,     84.919)
   pme recip    =     26.077 (     24.516,     27.847)
  restraint     =      0.221 (      0.178,      0.251)
 integrator
  constraint    =      0.000 (      0.000,      0.000)
  update        =      0.000 (      0.000,      0.000)
  comm_coord    =      0.000 (      0.000,      0.000)
  comm_force    =      0.000 (      0.000,      0.000)
  comm_migrate  =      0.000 (      0.000,      0.000)
```

Tutorial 1:Heating (annealing) (3_heating/)

After minimization, temperature are gradually increased to target temperature with ‘heating (annealing)’ simulation.

In GENESIS, heating simulation is executed by ‘annealing=yes’

Tutorial_1/3_heating/
- run.inp : control files of GENESIS
- run.sh : job scripts
- output/ : results

Extracted input file

```
[INPUT]
topfile = ../1_setup/top_all27_prot_lipid.rtf # topology file
(skip)
rstfile = ../2_minimization/run.rst          # restart file
```

```
[DYNAMICS]
integrator     = LEAP  # [LEAP,VVER]
nsteps         = 5000      # number of MD steps
timestep       = 0.002     # timestep (ps)
(skip)
annealing      = YES       # simulated annealing
anneal_period  = 50        # annealing period
dtemperature   = 3         # temperature change at annealing (K)
```

LEAP: Leapfrog
VVER: Velocity Verlet

Increasing 3K every 50 steps

```
[CONSTRAINTS]
rigid_bond     = YES       # constraints all bonds
                           # involving hydrogen

[ENSEMBLE]
ensemble       = NVT       # [NVE,NVT,NPT]
tpcontrol      = LANGEVIN  # thermostat
temperature    = 0.1       # initial temperature (K)
```

NOTE:
‘nsteps’ is short due to time restriction. Please increase steps when you will run actual simulations.

Executing Heating

Execute jobs

```
% cd  ~/Tutorial_1/3_heating
% ./run.sh
```

When your job is done, these outputs are shown.

run.dcd : Trajectory with dcd format (binary)
run.out : Output (energy/property) of GENESIS (ascii)
run.rst : Restart file for next simulation (binary)

Extract of output file

[STEP 5]

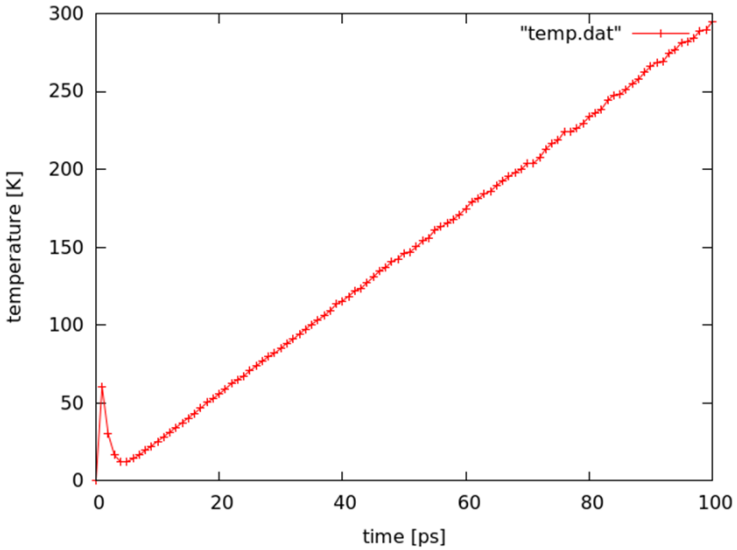
```
INFO:      STEP          TIME      TOTAL_ENE   POTENTIAL_ENE     KINETIC_ENE          RMSG          BOND         ANGLE    UREY-
BRADLEY       DIHEDRAL       IMPROPER          CMAP       VDWAALS         ELECT RESTRAINT_TOTAL     TEMPERATURE
VOLUME
 --------------- --------------- --------------- --------------- ---------------
(..skip..)
Simulated_Annealing_Leapfrog> Anneal temperature from   213.100 to    216.100
INFO:      3650        7.3000   -121015.8486   -134668.2442     13652.3956        14.8398       129.1861       316.5600
34.9546       301.9510       20.3742       -76.0000     17444.0571   -152886.2023       46.8752       184.8766    407423.5098

Simulated_Annealing_Leapfrog> Anneal temperature from   216.100 to    219.100
INFO:      3700        7.4000   -120524.6696   -134365.7978     13841.1281        14.8065       135.7281       319.1936
42.6740       291.7295       14.9113       -82.0820     17460.5756   -152587.8837       39.3559       187.4323    407423.5098
```

Checking output file

- Check properties from output file
  - You can extract temperature from output fie and see its change by plotting graphs using a graph tool (ex. gnuplot)

```
% sh temp.sh
% gnuplot
gnuplot> plot "temp.dat" w lp
```

Change of temperature




```
$ cat temp.sh
#!/bin/bash

grep "^INFO" output/run.out | tail -n
+2 | awk '{print $3, $17}' >temp.dat
```

Tutorial 1: Equilibration (4_equilibration/)

After heating system, simulations with conditions in production are required to equilibrate system. (This step is quite important for production runs)

Tutorial_1/4_equilibration/
- run.inp : control files of GENESIS
- run.sh : job scripts
- output/ : results

Extracted input file

Output trajectory file. The file can be input of analysis tools

```
[OUTPUT]
dcdfile = run.dcd  # DCD trajectory file
```

The following parameters should be identical to those in prodction runs.

[ENERGY]
[CONSTRAINTS]
[RESTRAINS]
[ENSEMBLE]

```
[ENSEMBLE]
ensemble     = NPT      # [NVE,NVT,NPT]
tpcontrol    = LANGEVIN  # thermostat and barostat
temperature  = 300.0    # initial temperature (K)
pressure     = 1.0      # target pressure (atm)
```

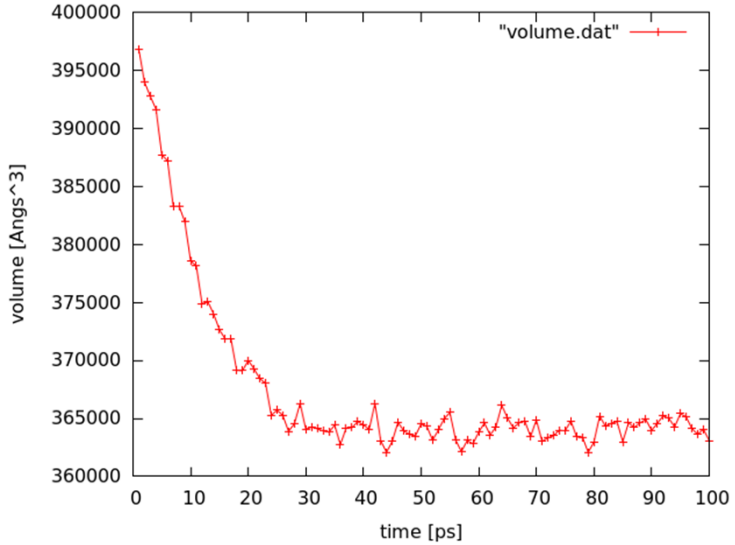
In this tutorial, 'nsteps' is limited due to time restriction. However, in actual simulations, 'nsteps' is much longer. (User needs to check not only stabilities of temperature and energy, but also properties she/he needs to check.)

Checking output file

- Check properties from output file
  - You can extract volume from output fie and see its change by plotting graphs using a graph tool (ex. gnuplot)

```
$ sh vol.sh
$ gnuplot
gnuplot> plot "vol.dat" w lp
```

Change of volume



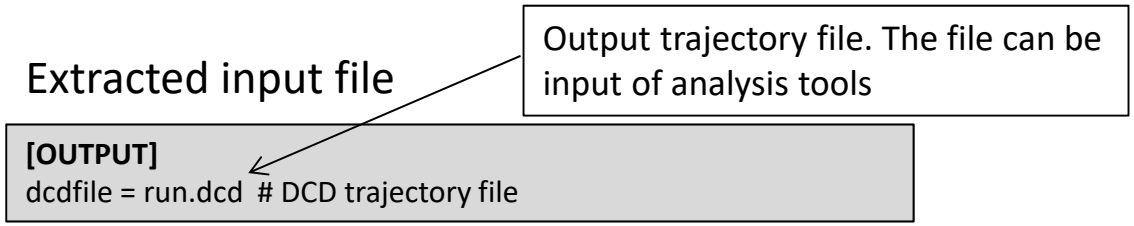

```
$ cat vol.sh
#!/bin/bash

grep "^INFO" output/run.out | tail -n
+2 | awk '{print $3, $17}' > vol.dat
```

Tutorial 1:Production(5_production/)

‘Production’ to investigate dynamics, structure, and so on

Tutorial_1/5_production/
- run.inp : control files of GENESIS
- run.sh : job scripts
- output/ : results

Extracted input file

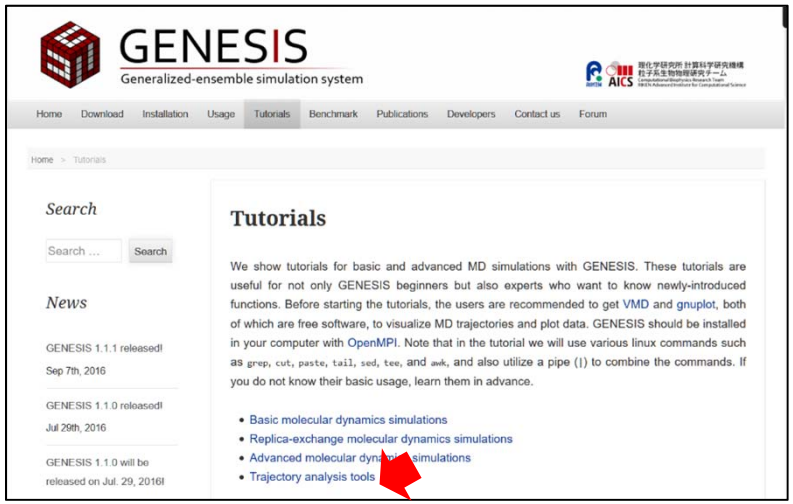


In this tutorial, ‘nsteps’ is limited due to time restriction. However, in actual simulations, ‘nsteps’ is much longer. (User needs to check not only stabilities of temperature and energy, but also properties she/he needs to check.)

Tutorial 1:Analysis(6_analysis/)

- Analysis of GENESIS from
  - GENESIS provides a set of analysis tools which processes simulation trajectories.

```
(Example)
./6_analysis:
1_RMSD              # RMSD
2_DIST              # Distance
3_RMSF              # RMSF
4_PCA               # PCA
```



- In this tutorial, we calculate RMSD (Root mean square distance: shows the deviation of structure from a reference coordinate) of Ca atoms in the system by using “trjana/rmsd_analysis” tool.

```
./1_RMSD:
run.inp                 # GENESIS input
run.sh                  # Batch script
```

Input file for Analysis

- Trajectory、selection、fitting

```
[INPUT]
reffile = ../../1_setup/ionize.pdb  # PDB file

[OUTPUT]
rmsfile       = run.rms      # RMSD file

[TRAJECTORY]
trjfile1      = ../../5_production/output/run.dcd # trajectory
file
md_step1      = 500000       # number of MD steps
mdout_period1  = 500         # MD output period
ana_period1    = 500         # analysis period
repeat1       = 1
trj_format     = DCD          # (PDB/DCD)
trj_type       = COOR         # (COOR/COOR+BOX)
trj_natom      = 0            # (0:uses reference PDB atom count)

[SELECTION]
group1        = an:CA # selection group 1

[FITTING]
fitting_method = TR+ROT # method
fitting_atom   = 1            # atom group
mass_weight    = NO           # mass-weight is not allowed!

[OPTION]
check_only     = NO           # (YES/NO)
analysis_atom  = 1            # atom group
```

Only Cα atoms

Fit to a reference structure with translation and rotation.

The number of repetition of trajectory information (for multiple trjfile)

YES : No actual calculation is done.

- run.sh is a shell script to achieve the analysis tools.

```
rmsd_analysis run.inp > run.out
```

Distance, angle, and dihedral angles can also be calculated using 'trj_analysis'.

GENESIS Trouble shooting

- GENESIS abnormally terminates.
  1. Run short MD (nstep=20) using “contact_check” option. It checks the structure before entering MD steps. It gives message as “too short distance” in log file if the structure is not adequate (e.g. containing many bad-contacts). In that case, preliminary minimization followed by short equilibration is recommended.

  2. Recompile the program using “./configure --enable-debug=3”

     It checks the number of atoms inside cells while running simulation. The calculation stops when access violation occurs. In that case, please change the related variables, e.g. maximum number of cell.

  3. Please inform the error in GENESIS Forum