MDexciteR tutorial

This tutorial provides an introduction to the basic usage of MDexciteR in conjucntion with NAMD. All input/output manipulations, normal modes and linear algebra calculations are carried out with the R software using its internal functions and the routines included in the bio3d package[1],[2].

install R

R can be easily installed through the package manager. Alternatively, it may also be downloaded from the R-project web site.

```
apt-get install r-base-core
```

install the Bio3D package

To obtain the Bio3D package, start R and then type the following line:

```
install.packages("bio3d")
```

Additional instructions for obtaining and installing the Bio3D package can be found in the Installing Bio3D Vignette.

install NAMD

Download the NAMD executable.

To invoke NAMD from R, we advise users to edit the bash_profile file entering the place where the namd2 executable is found:

```
export PATH="/path-to-namd-executable/:$PATH"
```

Preparation steps

Before run MDeNM simulations, users must carry out an equilibration procedure. The current MDexciteR version requires both atomic coordinates and velocities in PDB format (NAMD automatically writes them as .coor and .vel, respectively). In addition, an extended system file (.xsc) is also required. Here, the CHARMM-GUI webserver was used to prepare inputs for MDeNM simulations.

The **solution builder** module was employed to build a Protein/Solution system. The lysozyme structure (PDB code 2lzt) was selected. All input generation steps were carried out using default parameters. In the last step we selected the generation of inputs for equilibration in NAMD.

After download and extraction of the generated files we moved to the "namd" folder and edited the **step4_equilibration.inp** file by adding the **(binaryoutput = no)** line. Then the equilibration run was carried out with NAMD. The coordinates and velocities obtained at the end of the procedure were taken as inputs for MDexciteR.

Running MDexciteR

The current **MDexciteR** version works as a R script. Here is an example of the files that should be included in the folder where calculations will be carried out:

```
step4_equilibration.coor              # coordinates after equilibration
step4_equilibration.vel               # velocities after equilibration
step4_equilibration.xsc               # NAMD extended system file
step3_input.psf                       # topology
toppar/                               # folder with forcefield parameters
run_mdenm_namd_2022.sh                # execution script
inputs.R                              # input file to be edited by users
mdexicter_namd_nm.R                   # mdexciter script (DO NOT EDIT)
config.namd                           # NAMD configuration file
catdcd.sh                             # script to concatenate outputs
```

Users may edit the **inputs.R** file. This file contains all parameters required to set up MDeNM simulations. The corresponding file given in this tutorial is already prepared to run 10 independnt MDeNM simulations in which combinations of the three lowest frequency coarse grained NMs are excited.

```
# MDexciteR Arguments
# coor.user
Coordinates in PDB format from the last frame of the equilibration run.
# vel.user
Velocities in PDB format from the last frame of the equilibration run.
# ext.system
NAMD Extended system file from the equilibration run
# topol
topology in psf format
# seg.user
protein segments to be considered for  NM calculations
# sel.res
subset of that undergo excitation effects. Type 'all' to include all protein atoms
# modes
subset of modes to build linear combinations
# enm.ff
ENM forcefield. Types included: 'aaenm', 'hca' and 'anm'.
See documentation at the bio3D website for more details.
# temp.user
maximun temperature of the normal modes space
# temp.md
parameter to control temperature and pressure in NAMD
# steps_md
MD steps per excitation
```

```
# maxcycles
total number of cycles of excitation/relaxation
# namd.input
the command line for executing MD should be inserted in quotes (see example file)
```

The **config.namd** file may be edited by experienced users to adapt the parameters for their specific needs (i.e. force field paramenters, pressur amd temperature controls, etc). The version provided as example enables the simulation of lysozyme in explicit solvent in the NPT ensemble using the langevin thermostat and barostat.

After setting the user defined parameters in the **inputs.R** file, MDeNM simulations can be launched with the following command (to be typed in the command line):

```
sh run_mdenm_namd_2022.sh
```

This script will ask for the number of desired replicas. Then, it will automatically creates numbered folders that will contain the output files from each independent replicate. Each folder will contain output coordinate files (PDB format) named **cycle_X.coor**, where X is a number starting from one up to the the last excitation. In the example, five excitations were selected, so the last coordinate output will be **cycle_5.coor**.

These coordinates from each independnt replicates can be easily concatenated by a large variety of tools. We provide a script based on the catdcd tool, allowing the concatenation of protein coordinates into a single DCD file per folder.

References

(1) Grant, B.; Rodrigues, A.; Elsawy, K.; Mccammon, A.; Caves, L. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695–2696. https://doi.org/10.1093/bioinformatics/btl461.

(2) Skjaerven, L.; Yao, X.; Scarabelli, G.; Grant, B. Integrating Protein Structural Dynamics and Evolutionary Analysis with Bio3d. *BMC Bioinformatics* **2015**, *15*, 399. https://doi.org/10.1186/s12859-014-0399-6.