**A Tutorial of the CAPT Package**

Here we explain how to use the CAPT package to analyze Molecular Dynamics simulation (MD) trajectory data by taking the Alanine dipeptide data (Ala) as an example. The related files are in the folder “code/example”.

The procedure contains 4 steps, which can be divided into two stages: data preparation and data analysis.

**Stage I: data preparation**

**Step 1**. Create a project folder to store your data. Take Ala as an example, we create a folder called "Ala".

**Step 2.** Prepare the torsion angle files and trajectory index file. Suppose your molecule contains *n* pairs of phi-psi torsion angles and your MD data has *K* trajectories of this molecule, each trajectory containing *m* conformations.

Put the *k*-th pairs of phi and psi torsion angles for all conformations from all trajectories in the two angle files “Ala/Phi/Phi*k*.Rdata” and “Ala/Phi/Psi*k*.Rdata”, respectively, where *k=1, 2, ..., n*. Each angle file contains a numeric vector called “angV” of m\*K angle values in the range of (-π, π). Angles from neighbouring conformations shall also be put next to each other in the corresponding vector. The angle file shall be in R data file format (i.e., .Rdata file).

b) Put the trajectory index (a number within (*1, 2, ..., K)*) the file "traj.Rdata", which is an R data file containing a numeric vector called “traj” of the same length as all above “angV”. A number within “traj” tells which trajectory the corresponding element of “angV” is from. An obvious choice is to set “traj” using R code line “traj=c(rep(1, m), rep(2, m), ...., rep(K,m))”, and arrange angle files accordingly.

Note: R package “bio3d” can be used to compute all torsion angles from 3D coordinates of conformations.

**Stage II: data analysis using main.R**

**Step 3**. Set your parameters for CAPT by revising the following variables in the header of main.R;

**ProjectName**: the name of the folder where you store the data, e.g., “Ala”

**ProjectPath**: the path of the folder where you store the data, e.g., “./example/Ala/”

**TrajectoryFile**: the path and name of the trajectory file prepared in Step 2 (b), e.g., “./example/Ala/traj.Rdata”

m: the number of pairs of torsion angles in the molecule

**S0**: minimal size of a cluster to be generated

**Sc**: minimal size of a cluster to be further partitioned

**Pc**: cutoff of self-transition probability

**P0:** minimal self-transition probability (Pc >=P0)

**KernelType**: the kernel type used for density estimation; KernelType can be one of {"von", "gaussian", "epanechnikov"}; to use von-mises kernel, you need to install R packages circular and NPCirc.

**Step 4**. **main.R** will then initialize the R working environment, then call the CAPT function to get the partition tree. Users usually do not need to change anything beyond the header part. You can just copy&paste the code in main.R to R console, or run “Rscript main.R” in your console. The result file will be put within ProjectPath.