

## Tutorial how to use the NAMD interface on GenePattern

### A. Methods

We use NAMD for our simulations. This software was designed in order to perform simulations in structural biology and is available without charge for academic purposes.

The guide for this software is available online:

<http://www.ks.uiuc.edu/Research/namd/2.6b1/ug.pdf>

NAMD is using the CHARMM force-fields. The CMARMM force field is a reasonable model for atomistic interactions within proteins and between different proteins (including covalent bonds, H-bonds, electrostatic interactions and vdW interactions).

We will use steered molecular dynamics (SMD) to apply forces to the molecule and mimic with this method a tensile test. This method is widely used in the biomechanical community. To apply load, C $\alpha$  atoms at one end are fixed and the force is applied on the C $\alpha$  atom at the other end. The SMD technique is equivalent to attaching one end of a virtual harmonic spring to the end of the system and pulling at the virtual atom on the other end of the spring. The SMD method thus mimics an experiment where one end of the molecule is fixed (e.g. on a gold surface), while the other end is pulled at with the AFM cantilever tip. Using this technique, different loading conditions (e.g. tensile and bending) can be realized.

The force experienced by the virtual atom are given by:

$$F = k(v \cdot t - x) \quad (1)$$

Here,  $x$  is the displacement of the pulled atom,  $v$  the pulling velocity,  $t$  the time step, and  $k$  is the spring constant.

Like for the other problems, the NAMD software is made accessible via the GenePattern website.

### B. Getting Started

In order to run NAMD **you need four files**:

1. **The coordinate file**, also called .pdb file. This file describes the structure of your protein and gives the precise position of each single atom inside the protein. In addition to the protein atoms, it also contains water molecules, which surround the protein like a skin. (PDB= protein data bank, <http://www.pdb.org/pdb/home/home.do>). The PDB ID of our protein is 1gk7. You can also look for other proteins and visualize them with VMD.) However, structures directly from the PDB are typically not “ready” for direct simulation with NAMD and require significant pre-processing. We have done this already, so we encourage using the structures provided to you on Stellar.

2. **The structure file**, also called .psf file. This file describes the structure of the protein and the kind of bonds and interactions between the individual atoms (e.g. single bond, double bond, etc. )
3. **The fixed atoms file** (fix.pdb). This file has exactly the same structure as the pdb file. In **the second column after the z coordinate** we indicate with “1”, which atoms are fixed and thus cannot be moved, “0” indicates that they are free to move. Find an example below.
4. **The SMD file** (smd.pdb). This file has the same structure as the .pdb file and the fix file. In the **first column after the z coordinate** we indicate with “1”, which atoms are pulled, “0” means atoms are not pulled.

Example for a fixed atom file; the pdb file and the smd file have exactly the same structure:

ATOM	ID	Type	Residue		x	y	z	Tag for SMD	Tag for FIX			
ATOM	2	HT1	GLY	P	100	-28.735	-0.515	-2.217	0	0	P1	H
ATOM	3	HT2	GLY	P	100	-27.508	0.694	-2.355	1	0	P1	H
ATOM	4	HT3	GLY	P	100	-29.112	1.078	-2.616	0	0	P1	H
ATOM	5	CA	GLY	P	100	-28.671	0.852	-0.58	0	1	P1	C

This table shows the general structure of a .pdb file; the fix.pdb and smd.pdb have the same file format but different columns signify specific boundary conditions. In the example, Atom 3 is tagged for SMD (being pulled at), and atom 5 is held fixed. Note that we use two different files for SMD and FIX to avoid any confusion.

### C. Set up the simulation:

For our simulations we use the GenePattern platform developed by Ivica Ceraj.

1. You find the platform following this link: <http://acishpc.mit.edu:8083/gp/index.jsp>  
Choose “task” →NAMD→run
2. You will be connected to the page (MIT certificate necessary) where the simulation parameters can be set. As this page equals to the parameters file, you only need the other four files mentioned above (pdb, psf, fix, smd) to run a simulation. These files can be uploaded here.
  - a. Upload four files: Initial structure (.pdb file), structure .psf file, as well as fix.pdb and smd.pdb for the BCs.
  - b. The following parameters exist and can be modified:
    - **dcdfreq**. Indicates after how many iteration steps the atomistic configuration is stored. These are snapshots, which are later assembled to the vmd-movie. Please choose the frequency by considering the number of iteration steps (“run”, see below). The overall number of snapshots should not exceed 1,000 (e.g. when you choose 500,000 steps →dcdfreq=500). Otherwise the out.dcd file will become too large (100 MB or more!).

- **Fixatoms**: always on, except you would like to run an energy minimization or a similar simulations, where no boundary conditions are necessary (not applicable here).
- **SMD**: always on, except you would like to run an energy minimization or a similar simulations, where no boundary conditions are necessary.
- **SMD\_vel**: describes how far (in Å) the smd-atoms are pulled per iteration (1 iteration =1fs (1E-15s)). A reasonable number is 0.0001 or smaller. The smaller the “better” as it is closer to typical experimental conditions, but you have to consider the number of iteration steps to get reasonable displacement within reasonable simulation time.
- **smd\_DIR**: gives the direction of the pulling force. We are pulling along the z-axis. However, you can also run bending simulations or compression (then, you may choose z = - 1). Then you have to change the direction of the applied force.
- **smdK**: The spring constant between the virtual, with smd\_vel moved atom and the protein. 10 is a reasonable number (unit: kcal/mol/Angstrom<sup>2</sup>).
- **set\_temperature**: Defines the temperature of the environment. 300 K is a reasonable number and is used in most simulations on proteins. But you could in principle analyse the structure at any temperature you want to find out if there is any difference in behaviour.
- **Run**: Here you define the number of iterations: consider that the computer performs approximately 1300 steps/minute (for this particular structure). You can thus estimate how long the simulation will approximately take. Additionally, the displacement for the complete unfolding of the protein is of about 100Å. Thus you can calculate how many steps you need at a given pulling velocity (e.g. SMD\_vel=0.0001 →1,000,000 steps →a bout 13h). If you have e.g. a very small pulling velocity, that you might be interested in focusing only on the first 20 Angstrom of the displacement.

### 3. Run the simulation:

Logout is possible, just as for all other GenePattern applications.

Before making longer runs (e.g. 1,000,000 or more steps) it is highly recommended to check out a short run to find out if everything is working fine, and take a look at the results. (e.g. perform a study of 10,000-100,000 steps)

## D. Post processing: Analysing the results

### 1. The following resulting files are of particular interest:

- **energy**: Energy over the iteration steps. The file contains various energy contributions, at each step, in the following format:

```
ETITLE:    TS  BOND    ANGLE  DIHED  IMPRP  ELECT  VDW  BOUNDARY
MISC      KINETIC  TOTAL  TEMP   TOTAL2  TOTAL3  TEMPAVG
```

For example, the 12<sup>th</sup> column provides the total energy (see energy.png as plot)

- **energy.png**: plots the total energy (kinetic plus potential) over iteration steps.
- **out.coor**: give the coordinates of the atoms at the end of the simulation
- **force\_displacement.png**: Plots the force over the displacement

- **Namd.log**: contains all the important information, which are collected during the simulation. The energy is extracted from this file, as well as the F\_x\_plot.
- **out.dcd**: This file includes the trajectories of the simulation. Used in VMD the simulation results can be visualized as a movie. To do this, first load a PDB file (e.g. the initial geometry, and then select: “File→Load data into molecule”).
- **top\_all22\_prot\_inp**: Topology files. Provides the information about the amino acids, the bonds etc. (not needed here, just for your information)
- **out.namd**: File includes all the parameters of the simulation, as it was passed on to NAMD.

## E. Visualizing the results

VMD (available here: <http://www.ks.uiuc.edu/Research/vmd/>) can be used for visualizing the results of your simulation. It can also be used for visualizing other proteins and structures (e.g. after download from the protein data bank).

The necessary files are (i) a valid PDB file and (2) a dcd file. The following steps are necessary for visualizing the results (windows-version):

- a. Open VMD
- b. →file→new molecule → a molecular file browser will appear
- c. browse the file you want to load (e.g. the initial geometry) and choose as file type “PDB” →load
- d. Go to the main menu again.
- e. Click on the name once with the left mouse button, and select the previous molecule in the “VMD Main” window →the name will be coloured in yellow → click with the right mouse button now and choose “load data into molecule”
- f. Browse the out.dcd file and choose as file type: “CHARMM, NAMD, XPLOR, DCD Trajectories” (this normally happens automatically) →load the file
- g. The .dcd file is loading. This can take a couple of seconds.
- h. Now you can watch the movie, change the speed of the movie etc, by controlling it in the VMD Main window

More details about the visualization will be shown in the lecture/additional tutorial. There is also a VMD-tutorial and a VMD-manual online:

<http://www.ks.uiuc.edu/Research/vmd/>

## F. References

1. Alberts B., *Molecular Biology of the Cell*, 2002: Taylor & Francis
2. Wang N., Stamenovic D., *Mechanics of Vimentin intermediate filaments*. J. Muscle Research and Cell Motility, 2003. **23**: p. 535-540
3. Janmey P. A., U. Euteneuer, P. Traub, M. Schliwa, *Viscoelastic Properties of Vimentin Compared with Other Filamentous Biopolymer Networks*. J. of Cell Biology, 1991. **113**: p-155-160
4. Strelkov S. V., H. Herrmann, U. Aebi, *Molecular architecture of intermediate filaments*. BioEssays, 2003. **25**: p. 243-251

5. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal Of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
6. Brooks B.R., Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M, *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations*. J. comput. chem., 1983, **4**(2): p. 187-217
7. Lu H, Israilewitz B, Krammer A, Vogel V, Schulten K, *Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation*. Biophysical Journal, 1998 **75**: p. 662-671
8. Humphrey W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics, 1996. **14** (1): p. 33