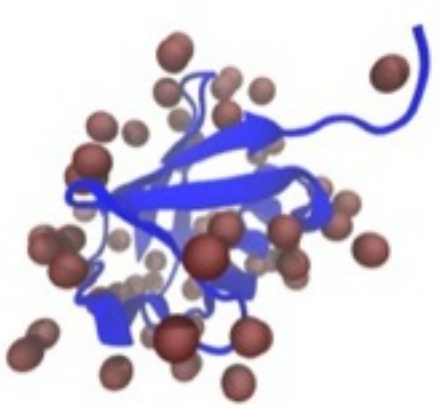
University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group
Computational Biophysics Workshop

qwikMD - Easy Molecular Dynamics with NAMD and VMD



Tutorial by Rafael C. Bernardi, Till Rudack and Joao Ribeiro

July 2015

A current version of this tutorial is available at
http://www.ks.uiuc.edu/Training/Tutorials/
Join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Contents

1 Introduction

The relationship between structure and function of proteins is directly connected to atomistic aspects of protein dynamics. As a major methodology in structural biology, molecular dynamics (MD) simulations permit the exploration of the physical mechanism underlying the function of proteins by examining their dynamical behavior. Although the advances in MD simulations are leading to a new level of knowledge of macromolecular complexes, reaching the million-to-billion atom regime, the vast majority of MD users are interested in straightforward MD simulations of relatively simple proteins. Particularly in the last few years, structural biology experimentalists interest in performing MD simulations to improve their knowledge of a protein structure/function relationship has increased greatly. To assist these experimentalists and any novice to MD to overcome the initial learning curve barrier of MD simulation software, we developed a user interface (plugin), qwikMD, that connects the widely employed and user-friendly molecular graphics program VMD to the widely adopted MD program NAMD. Employing the qwikMD , a user is able to setup an MD simulation in just a few minutes, allowing quick studies of point mutations, partial deletions and even atomic force microscopy experiments. The plugin makes it easy for a new user to perform MD simulations, while it also servers as a learning tool. Many info buttons provide the theoretical background underlying the MD procedures carried out in modern MD simulations. The *info button* windows also provide links to more complete explanations in our website or in some of our publications. Don't forget to use these buttons to learn more about qwikMD and MD simulations.

qwikMD still in test! qwikMD is a new tool that is not yet ready for all the possible biological systems that are studied employing MD simulations. We are working hard to make qwikMD fully operational soon and if you have suggestion to improve qwikMD please contact us.

1.1 NAMD

NAMD, recipient of a 2002 Gordon Bell Award and a 2012 Sidney Fernbach Award, is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. Based on Charm++ parallel objects, NAMD scales to hundreds of cores for typical simulations and beyond 500,000 cores for the largest simulations. NAMD uses the popular molecular graphics program VMD for simulation setup and trajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PLOR. NAMD is distributed free of charge with source code.

In order to run any MD simulation, NAMD requires at least four things:

- a Protein Data Bank (pdb) file which stores atomic coordinates and/or velocities for the system. PDB files may be generated by hand, but they are

also available via the Internet for many proteins at `http://www.pdb.org`.

- a Protein Structure File (psf) which stores structural information of the protein, such as various types of bonding interactions.
- a force field parameter file. A force field is a mathematical expression of the potential which atoms in the system experience. CHARMM, X-PLOR, AMBER, and GROMOS are four types of force fields, and NAMD is able to use all of them. The parameter file defines bond strengths, equilibrium lengths, etc.
- a configuration file, in which the user specifies all the options that NAMD should adopt in running a simulation. The configuration file tells NAMD how the simulation is to be run. Dor more details check NAMD Tutorial.

Warning! The goal of this tutorial is to introduce qwikMD and NAMD by performing some short molecular dynamics simulations. Therefore, the examples provided are optimized so simulations can be done in a reasonable period of time on a common computing facility. This means that some parameters and conditions under which simulations are done in this tutorial are not suitable for scientific studies. Whenever this happens it will be pointed out and alternatives or more appropriate parameters/conditions will be provided in case you want to improve the simulations and/or you have more computer power available.

1.2 qwikMD

qwikMD is a VMD plugin that aims on helping its users to start and analyze MD simulations. The plugin helps, particularly scientists that are starting to perform MD simulations to prepare the necessary files to run these simulations from desktop machines all the way to large supercomputers. All the necessary steps, from the PDB to the configuration file is created with simple procedures but giving the user all the information and scripts that were performed in background, so that one can use the plugin to learn how to prepare MD simulations. The `Live Simulation` option allows for the visualization and analysis of the simulation on the fly, helping new users to learn more about MD simulations and expert users to test their simulations before submitting it to run in a supercomputer. qwikMD integrates VMD and NAMD, two widely used software developed by the Theoretical and Computational Biophysics Group at University of Illinois at Urbana-Champaign.

2 Required programs

The following programs are required for this tutorial:

- **NAMD:** Available at `http://www.ks.uiuc.edu/Research/namd/` (for all platforms). Note that versions prior to NAMD 2.6 do not support the CMAP correction of the CHARMM forcefield.

- **VMD:** Available at `http://www.ks.uiuc.edu/Research/vmd/` (for all platforms)
- **qwikMD:** Available at `http://www.ks.uiuc.edu/qwikMD` (for all platforms)

2.1 For Linux/Mac Users:

- **Text Editor:** Nedit is a text editor which will be used throughout this tutorial to view and edit some of the files associated with the simulations. There are others such as pico, emacs, jot, and vi. Feel free to use whichever text editor you are most comfortable with.
- **Plotting Program:** We will use the free program xmgrace, available at `http://plasma-gate.weizmann.ac.il/Grace/`, to view and analyze output data from NAMD simulations. VMD also has an internal plotting program which may be used to examine output directly from NAMD log files. Other useful graphing programs are Mathematica, `http://www.wolfram.com/` (Purchase required), Matlab, `http://www.mathworks.com/` (Purchase required), and gnuplot, `http://www.gnuplot.info/` (Free download).

2.2 For Windows Users:

- **Text Editor:** We will use WordPad to view and edit some of the text files associated with the NAMD simulations. You may prefer to use Notepad or another text editor of your choosing. Microsoft Word is a word processing program and should not be used as a text editor.
- **Plotting Program:** We will use Microsoft Excel, available at `http://office.microsoft.com/en-us/FX010858001033.aspx`, to view and analyze output data from NAMD simulations. VMD also has an internal plotting program which we will use to examine output directly from NAMD log files. Other graphing programs which you may find useful are Mathematica, `http://www.wolfram.com/`, Matlab, `http://www.mathworks.com/`, and scilab, available at `http://www.scilab.org/`. Note that each of these programs, with the exception of scilab, requires you to purchase software. Scilab is a free program with capabilities nearly identical to Matlab. If Excel or any of the proprietary programs are unavailable to you, we encourage you to use scilab.

3 Installing the required Programs

3.1 VMD

The VMD source code and binary distributions can be obtained after registering at the VMD web page. Download the appropriate distribution file with your web browser. Windows binary distributions are self extracting, so once the

distribution file is downloaded, proceed to the installation directions below. For source distributions and Unix binary distributions, uncompress and untar the file. This will produce a subdirectory named `vmd-1.9.2`. Unless otherwise specified, all references to VMD code will be from this subdirectory, so cd there.

http://www.ks.uiuc.edu/Research/vmd/current/ig/ig.html

3.2 qwikMD

QwikMD is available in the newest version of VMD. If it is not available in your version you can download the plugin from `http://www.ks.uiuc.edu/qwikMD`

Download and uncompress the file in a known location.

1 Open VMD by typing `vmd` in the Terminal window.

Launch VMD. Open VMD graphically by double-clicking on the VMD icon under `Applications` in the `Finder`.

2 Open VMD console by clicking Extensions → Tk Console menu item in the VMD Main window.

3 Go to the downloaded qwikMD folder and type `play install.tcl`

4 Follow the instructions. Do not Install qwikMD in the same folder where you downloaded the plugin.

5 Close vmd and launch it again. Test if you can open qwikMD by clicking Extensions → Simulation → qwikMD menu item in the VMD Main window.

3.3 NAMD

You can build NAMD yourself or download binaries for a wide variety of platforms. To run qwikMD you need NAMD installed in your machine and available in your path.

For LINUX/MAC users:

Setting the Path: To start to use qwikMD you will need to add the namd2 directory to your path in order for the operational system to locate it. To perform that, add to your .bashrc (Linux) or .Profile (Mac) in your home folder the following line:

export PATH=complete.path.for.namd:complete.path.for.namd:$PATH

Where (complete.path.for.namd) is the complete path to the actual folder where the namd2 executable is available.

Example:

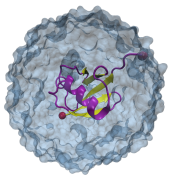
export PATH=/usr/local/NAMD_2.10:/usr/local/NAMD_2.10:$PATH

If you are new to Linux, visit our guide at:
http://www.ks.uiuc.edu/Training/Tutorials/Reference/unixprimer.html
For Windows users:

Setting the Path: To start to use qwikMD you will need to add the namd2 directory to your path in order for Windows to locate it. This can be accomplished by right-clicking Computer on the Desktop and selecting Properties Advanced system settings Advanced Environment Variables (the precise procedure may vary depending on your version of Windows). Under System variables, scroll down and select Path and then Edit. At the end of the long line in Variable Value, add a semi-colon ; then the full path to the directory containing namd2 (but do NOT add the executable namd2 at the end). Click OK. Now open a new command prompt. Regardless of the directory you are in, you should be able to type namd2 and run it.

4 Running my First Molecular Dynamics Simulation

Molecular dynamics (MD) is a computer simulation of physical movements of atoms and molecules, widely used to study biological systems. The atoms and molecules are allowed to interact, giving a view of the motion of the atoms and consequently of protein domains in the case of protein simulations. MD has emerged as an important research methodology covering systems to the level of millions of atoms.



Histidine Residues. Of the 20 amino acids, histidine is the only one which ionizes within the physiological pH range (∼7.4). This effect is characterized by the pK_a of the amino acid side chain. For histidine, the value is 6.04. This leads to the possibility of different protonation states for histidine residues in a protein, and makes the consideration of the proper state important in MD simulations. The viable states are one in which the δ nitrogen of histidine is protonated (listed with residue name "HSD" in the topology file), one in which the ϵ nitrogen of histidine is protonated ("HSE"), and one in which both nitrogens are protonated ("HSP").

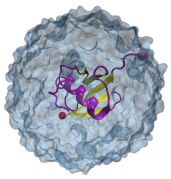
To start setting up a MD simulation with qwikMD and NAMD you will first load a PDB file. If this is your first time working with MD simulations of proteins you can use, as an example, the PDB code 1UBQ. Ubiquitin (PDB code 1UBQ) is a small regulatory protein that is present in almost all tissues (ubiquitously) of eukaryotic organisms. Here we are going to simulate Ubiquitin with implicit solvent, where water molecules are represented by a dieletric constant.

1 Open VMD by typing `vmd` in the Terminal window.

Launch VMD. Open VMD graphically by double-clicking on the VMD icon under `Applications` in the `Finder`.

2 Open qwikMD by clicking Extensions → Simulation → qwikMD menu item in the VMD Main window.

3 Load the PDB by typing the PDB code in the blank space and clicking `Load`. For this tutorial we are going to work with Ubiquitin, PDB code: 1ubq



Representations. To help the user to recognize the patterns of the system, qwikMD loads the structures using the more common represention. You can change the representation in the selection window of the qwikMD by clicking on the current representation and selecting a different option. Changing the representation of the protein to licorice you will note that the X-ray structure from the Protein Data Bank does not contain the hydrogen atoms of ubiquitin. This is because X-ray crystallography usually cannot resolve hydrogen atoms. The pdb file you will generate with psfgen along with the psf will contain guessed coordinates for hydrogen atoms of the structure. Later, energy minimization of the protein will ensure their positions are reasonable.

To perform MD simulations one have to mimic the environment of the protein, or any other molecule of interest. The most common solvent is water and there are two main ways to mimic the solvent effect. Either simulating all the atoms of the solvent explicit solvent model or by adding dielectric constant to the electrostatic calculation implicit solvent model. Next you will find a description of these models as well as a description on how to add salt to the water solution in order to make a more realistic solvent model.

Implicit Solvent

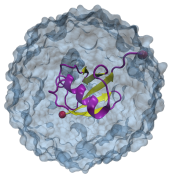
An implicit solvent model is a simulation technique that eliminates the need for explicit water atoms by including many of the effects of solvent in the interatomic force calculation. For example, polar solvent acts as a dielectric and screens (lessens) electrostatic interactions. The elimination of explicit water accelerates conformational explorations and increases simulation speed at the cost of not modeling the solvent as accurately as explicit models.

But be careful, because implicit solvent models eliminate explicit water molecules and represent water in an averaged manner, implicit solvent models are considered less accurate than explicit solvent models. Always use caution when employing implicit solvent for molecular dynamics research.

Generalized Born Implicit Solvent

Generalized Born implicit solvent models are one particular class of implicit solvent models. There are two parts to a GBIS calculation. First, the Born radius of each atom is calculated. An atoms Born radius represents the degree of exposure of an atom to solvent. Atoms on the surface of a protein are highly exposed to solvent, their electrostatic interactions will be highly screened and their Born radii will be small. Atoms buried in the center of a protein will not be very exposed to solvent, their electrostatics wont be screened much and their Born radii will be large. Second, inter-atomic electrostatic forces are calculated based on atom separation as well as the geometric mean of the two atoms Born radii.

qwikMD uses Generalized Born Implicit Solvent when the Implicit Solvent option is selected. You can learn more about Generalized Born method in the manuscript linked at the bottom of this window.



X-ray Crystallography. X-ray crystallography methods utilize the optical rule that electromagnetic radiation will interact most strongly with matter the dimensions of which are close to the wavelength of that radiation. X-rays are diffracted by the electron clouds in molecules, since both the wavelength of the X-rays and the diameter of the cloud are on the order of Angstroms. The diffraction patterns formed when a group of molecules is arranged in a regular, crystalline array, may be used to reconstruct a 3-D image of the molecule. Hydrogen atoms, however, are not typically detected by X-ray crystallography since their sizes are too small to interact with the radiation and since they contain only a single electron. The best X-ray crystallography resolutions currently available are around 0.9Å.

4 To perform simulations in implicit solvent we need to remove the oxygen atoms from water molecules that were present in the crystal structure. To do that, above the `Selection` window, click on `Select chain/type` and deselect the water molecules `A and water`. This will make the oxygen atoms (in red) disappear.

5 Let's first run a `Live Simulation` with standard protocols. Mark the `Live Simulation` box and click `Prepare`. You will have to select a folder where the files will be saved. During this step the autopsfgen function of VMD is called to create the necessary files to run the MD simulation. qwikMD will also create the necessary NAMD configuration files.

When the Prepare button is pressed, scripts to perform all the steps required by the settings selected by user. Two folders will be created in the Working directory: files created in the preparation step in a SETUP folder, while files needed to run the MD simulations are in a RUN folder.

NOTE:! If you want to run `Live Simulation`, make sure you have the corresponding box checked before you click `Prepare`.

NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. NAMD scales to hundreds of processors on high-end parallel platforms, as well as tens of processors on low-cost commodity clusters, and also runs on individual desktop and laptop computers.

qwikMD helps the user to prepare NAMD input files to run in a range of computers, from the largest supercomputers with high-end parallel platforms to the smallest laptop computers. qwikMD also allow, through the Interactive Molecular Dynamics interface of NAMD, to run live simulations where the user can look and analyze the trajectories while they are being created.

6 To run the simulation click in the Start Button. Interactive Molecular Dynamics used in the Live Simulation might take a few seconds to start. In the console window is possible that an Error message will show up until the connection between VMD and NAMD is established.

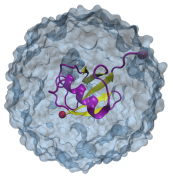
Warning!! Always click to FINISH your simulation before leaving VMD or closing qwikMD window. NAMD activities started from VMD will continue to run in background for all the steps requested unless you make sure the simulation was aborted before leaving qwikMD.

4.1 Analysing during a Live Simulation

VMD is a powerful tool for analysis of structures and trajectories and should be used as a tool to think. Numerous tools for analysis are available under the VMD Main menu item Extensions - Analysis. In addition to these built-in tools, VMD users often use custom-written scripts to analyze desired properties of the simulated systems. VMD Tcl scripting capabilities are very extensive, and provide boundless opportunities for analysis. qwikMD provides the user with some of the most employed analysis tools, allowing also the analysis while performing live NAMD sections.

During the simulation, clicking on the `Analysis` tab on the top you can perform a few analysis during the Live Simulation.

1 On RMSD, click on Calculate to check the RMSD.



Root-Mean-Square Deviation. The root-mean-square deviation (RMSD) is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins. In the study of globular protein conformations, one customarily measures the similarity in three-dimensional structure by the RMSD of the C atomic coordinates after optimal rigid body superposition. When a dynamical system fluctuates about some well-defined average position, the RMSD from the average over time can be referred to as the RMSF or root mean square fluctuation. The size of this fluctuation can be measured, for example using Mssbauer spectroscopy or nuclear magnetic resonance, and can provide important physical information. qwikMD allows the user to perform RMSD analysis during live NAMD simulations. More advanced options for RMSD analysis can be done with VMD plugins available in VMD Main menu item Extensions → Analysis → RMSD Trajectory Tool .

4.1.1 Ubiquitin in a Water Box

To start a new simulation don't forget to click FINISH and RESET qwikMD.

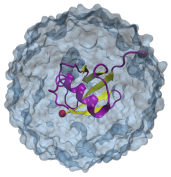
More realistic MD simulations are performed with explicit representation of every atom of the solvent, usually a solution of water and salt. The water box created by qwikMD is somewhat big for most studies. The big water box was adopted as a safety measure. It is common to see large conformational changes in proteins. These changes can make the water box too small, which is hard to be observed by someone new in the field. Ideally, one should work with a box, which is large enough that the protein does not interact with its image in the next cell if periodic boundary conditions are used. The use of periodic boundary conditions involves surrounding the system under study with identical virtual unit cells. The atoms in the surrounding virtual systems interact with atoms in the real system. These modeling conditions are effective in eliminating surface interaction of the water molecules and creating a more faithful representation of the in vivo environment than a water sphere surrounded by vacuum provides.

Many water molecules are available for MD simulations. NAMD currently supports the 3-site TIP3P water model, the 4-site TIP4P water model, and the 5-site SWM4-NDP water model (from the Drude force field). As the standard water model for CHARMM, TIP3P is the model employed in the simulations prepared with qwikMD.

2 Open VMD by typing `vmd` in the Terminal window.

Launch VMD. Open VMD graphically by double-clicking on the VMD icon under `Applications` in the `Finder`.

3 Open qwikMD by clicking Extensions → Simulation → qwikMD menu item in the VMD Main window.



NMR Structures. Nuclear magnetic resonance spectroscopy of proteins, usually abbreviated protein NMR, is a field of structural biology in which NMR spectroscopy is used to obtain information of the structure and dynamics of proteins, nucleic acids, and their complexes. It is not the case of the 1UBQ structure, solved by X-ray crystallography, but NMR structure in the PDB usually have multiple steps. In order to start a MD simulation one have to select one of the steps as the initial coordinates. It is usual, when running more than one simulation of the same system, to select different initial steps in each of the simulations to improve sampling of conformational structure.

4 Load the PDB by typing the PDB code in the blank space and clicking `Load`. For this tutorial we are going to work with Ubiquitin, PDB code: 1ubq

Select Resid. This button will open a new window where the user can do mutations, rename molecules that have wrong names (read more below), change protonation states, delete parts of the molecules and also inspect the structure with a interactive residue/molecule list. Select Resid is especially important in cases where one of the molecules/ions have wrong names, or names that are different from the name used in the CHARMM force field. For instance, it is common in the PDB that Ca2+ ions have the name CA. CHARMM recognize the name CA as alpha-Carbon of protein structures, and CA resname (residue name) is not recognized by CHARMM. Select Resid allow for the user to rename CA ions to proper Calcium parameters that will be compatible with the CHARMM force field.

5 To perform simulations in explicit solvent we need to change part of qwikMD standard protocol. For that, select the `Show Options` box.

Creating a Salt Solution

Ions should be placed in the water to represent a more typical biological environment. They are especially necessary if the protein being studied carries an excess charge. In that case, the number of ions should be chosen to make the system neutral. The ions present will shield the regions of the protein, which carry the charge, and make the entire system more stable. They should be placed in regions of potential minima, since they will be forced to those regions during the simulation anyway. The psf file contains the charge of each atom and may be used to determine the charge of the total system or parts of it.

One must set the desired salt concentration when preparing the simulation with qwikMD. The default Salt Concentration is 0.15 mol/L. Even if the Salt Concentration is set to ZERO, qwikMD will add ions to neutralize the total charge of the system. Remember, in a MD simulation with periodic boundary condition the total charge of the system should be ZERO.

6 Perform a simulation in explicit solvent is not an easy task for your computer. Change the Max Time of the simulation to 0.1, to be able to perform this tutorial simulation in a few minutes.

7 To run a simulation in background, make sure the `Live Simulation` box is deselect and click `Prepare`. You will have to select a folder where the files will be saved. During this step the autopsfgen function of VMD is called to create the necessary files to run the MD simulation. qwikMD will also create the necessary NAMD configuration files.

8 To run the simulation click in the Start Button. VMD will freeze during the simulation.

4.1.2 Running your Simulation outside of qwikMD

To run molecular dynamics simulations with NAMD at least four files are required: a Protein Data Bank (pdb) file a Protein Structure File (psf), a force field parameter file and a configuration file. During the preparation steps, where the system might be solvated, ionized, among other procedures, several files are created. qwikMD separates the files created in the preparation step in a SETUP folder, while files needed to run the MD simulations are in a RUN folder. These two folders are created inside the folder defined by the user in working directory window. With the same name as the folder created by the user, a file with .qwikMD extension allows the user to load simulations performed with qwikMD and also previously created preparation steps, like amino acid residues mutations or salt concentration.

To run a simulation prepared with qwikMD in a computer cluster or supercomputer, one needs to copy only the RUN folder.

9 Running

Run your simulation by typing in a Terminal window:

```
namd2 qwikMD_equilibration_0.conf > qwikMD_equilibration_0.log &
```

5 Tackling common scientific problems

After getting familiar with qwikMD by applying it to the toy example of ubiquitin you will now apply qwikMD to a real scientific scenario. One common problem in structural biology is that experiments reveal mechanistically relevant mutations for cellular processes but the detailed atomic structural changes evoked by the mutation often remain elusive. Here MD simulations can help. Instead of time consuming and expensive mutational studies through X-ray or NMR spectroscopic experiments, MD simulations are often a much faster and cheaper way to gain atomic insights in structural and dynamical changes through mutations. Prerequisite to solve this problem by MD simulation is only one experimentally resolved atomic structural model of the wild type structure as initial model.

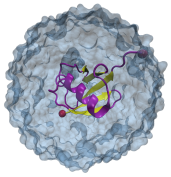
5.1 Cancer mutation in Ras

Here, we will use the example of the small GTPase Ras, a crucial switch regulating cellular signal transduction for cell growth. By GTP hydrolysis to diphosphate and Pi, Ras is switched from the active GTP-bound on- to the inactive GDP bound off-conformation and signal transduction is terminated. If the downregulation of Ras by GTP hydrolysis is disturbed, the nucleus receives an enduring signal for proliferation, resulting in uncontrolled cell growth that may lead to cancer. In deed, about 20 % of all cancer tumors are associate to malfunctions of the Ras protein through so called oncogenic mutations, which most off them effect the hydrolysis rate. One famous oncogenic mutation, which inactivates the switching mechanism of the Ras protein, is the G12V mutation. In order to investigate the atomistic structural differences between wild type and oncogenic mutation you first need to prepare and simulate the wild type structure in a similar manner as done previously for ubiquitin.

Ras wild type simulation

1 Load the active GTP bound structure of Ras by typing the PDB code 1QRA in the blank space and clicking `Load`.

A warning will appear that one or more residues could not be identified. In order to run MD simulations, force field parameters are need. Structures in the PDB often not only consist of standard amino acids but also DNA, ligands or other small molecules. If your loaded molecule contains non standard parts for which no force field parameters could be identified . This means you cannot run a MD simulation without further modification. If no parameters exist you need to parameterize it by yourself. Alternatively you can delete the non-identified part if it is not relevant for structure or function but usually this it not the case. The non-identified parts are flashing in your structure and are highlighted in red in the `Select Residue` window. In case of Ras the ligand GTP could not been identified.



Force field parameters. To run an MD simulation force field parameters are needed. The force field contains mathematical expression of the potentials which atoms in the system experience and which leads to the molecular motion. Every force field needs parameters to describe the potentials between the atoms. There are force field parameters available for all standard amino acids, DNA, several small molecules and much more. However, for some small molecules no parameters are available yet. Details about parameterization are described in the tutorial about the force field Tool Kit (ffTK), which is a VMD plugin that assists in the process of parameterization. However, force field parameterization is something for advanced users and usually a very challenging and time consuming process. As the outcome of the simulation strongly depends on the force field parameters, parameterization should be carried out very carefully.

2 Press `OK` in the warning window and click on `Select Residue`. You can see all non-identified parts highlighted in red. Click on `Type` to order the residues by type. To modify the second line, which contains GTP, select `Type` in the `Table Mode` box and click on nucleic. Choose hetero in the scroll down menu as the GTP is a hetero atom and not a nucleic acid.

3 Now the whole loaded structure is recognized correctly by qwikMD. Beside the hetero atoms of GTP (residue ID 166) and the magnesium ion (residue ID 167), represented by sticks, the Ras structure has some protein bound waters resolved, which are revealed as spheres in a van-der-Waals representation. These waters can play a crucial functional role. However, for simplicity we here delete them by clicking on `Select chain/type` and unchecking the third line `A and water`.

4 Now the structure is ready to set up the simulation. Following the same steps as in section 4.1.1 for ubiquitin you can prepare a live simulation of GTP bound Ras. For time reasons just choose implicit solvent. Note, if you want to run a scientific relevant simulation you should consider to use explicit solvent as solvation effect are in general mechanistically crucial, in particular to explain mutational effects.

5 Before you start the equilibration run, navigate to the analysis tab in the top. Here, you can choose to interactively calculate and plot some important properties during the MD simulation, like RMSD, energies, and temperature. Open an analysis window that will plot the RMSD of the backbone atoms by checking the `Backbone` option in the RMSD panel and clicking on `Calculate`. Uncheck `Total` and check `Kinetic` and `Potential` in the Energies panel and then press `Plot` to open a window for the kinetic and potential energy plot. In order to generte a plot to follow the temperature press the `Plot` button in the `Conditions` panel.

6 Change the representation of the protein to `licorice` and the color to `name` and then start the simulation by pressing `Start Equilibration Simulation`. First the system is minimized. The atoms are rearranged in order to optimize the energy. You can observe this in an increase of the RMSD and a decrease of the potential energy. After the minimization the system is slowly heated up. This means the system is coupled to an external temperature bath, which can be observed by an increase in the temperature plot. By this the system is getting an external energy, which leads to an increase in the kinetic energy.

7 Press `Finish` to stop the equilibration. Navigate in your file browser to the run folder in your work directory. The file `qwikmd_equilibration_0_constraints.pdb` is the last state structure shown on the screen as you pressed finish. Rename the file to `Ras_WT_equil.pdb`

Ras G12V mutation simulation

8 Redo steps one and two for the Ras wild type simulation and then go to the `Select Residue` window. Select `Mutate` in the `Table Mode`. Click on Gly with the residue ID 12 and choose valin from the scroll down menu. In order to introduce the G12V mutation you need to prepare the system by using the `Prepare` button. All simulation parameters should be the same as for the wild type simulation.

9 Run a simulation for the Ras G12V mutant in the same manner as for the wild type but choose a different work directory. After finishing the simulation press `Reset` and close qwikMD. Then navigate to the run folder in the G12V work directory and rename the file `qwikmd_equilibration_0_constraints.pdb` to `Ras_G12V_equil.pdb`.

Comparison of the RasWT and Ras G12V mutation simulations

10 Open VMD and load via `File New Molecule` the files `Ras_WT_equil.pdb` and `Ras_G12V_equil.pdb`. Open the Representation window in `Graphics Representation`. Select in the top menu `Ras_WT_equil.pdb` and click in the top window on the selection all. Then change the `Drawing Method` to `New Cartoon` and the `Coloring Method` to `Color ID 7`. Create a new representation (`Create Rep`) and select only the GTP by typing `resname GTP` in the `Selected Atom` line. Change the `Drawing Methode` to `Licorice`. Create another representation for the mutated residue by typing resid 12 in the `Selected Atoms` line. Change the `Drawing Methode` to `Licorice` and the `Coloring Methode` to `Color ID 7`. Now select `Ras_G12V_equil.pdb` in the top menu and create the same representations as for the wild type but choose `Color ID 1`.

You can observe that there are almost no overall structural differences between the inactive oncogenic G12V Ras mutation (red) and the active Ras wild type structure (green). In addition the site chain atoms added by introducing valine instead of glycine do not have an effect on the structure or positioning of the GTP. This reveals that the G12V mutant has no influence on the GTP bound state and implies that it has to have an effect on the intermediate state. So far there is no experimental method available to resolve the structure of the intermediate state.

6 Steered Molecular Dynamics

Among MD methods, steered molecular dynamics (SMD) simulations in which external forces are used to explore the response and function of proteins have become a powerful tool especially when combined with single molecule force spectroscopy SMFS employing atomic force microscopy. SMD has been successfully employed in a wide range of biological systems, from the investigation

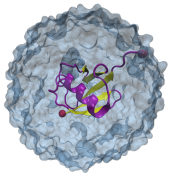
of protein mechanotransduction, to permeability of membrane channels, and the characterization of proteinreceptor interactions. SMD simulations have also been used to study force propagation through proteins.

6.1 Biomolecular interactions during protein unfolding

1 Open VMD by typing `vmd` in the Terminal window.

Launch VMD. Open VMD graphically by double-clicking on the VMD icon under `Applications` in the `Finder`.

2 Open qwikMD by clicking Extensions → Simulation → qwikMD menu item in the VMD Main window.



SMFS. Single-molecule force spectroscopy has emerged as a powerful tool to investigate the forces and motions associated with biological molecules and enzymatic activity. The most common force spectroscopy techniques are optical tweezers, magnetic tweezers and atomic force microscopy (AFM). In AFM, a microscopic tip that is located on the end of a cantilever picks up molecules adsorbed on a surface. Usually the cantilever tips are covalently functionalized with the molecules of interest. A piezoelectric controller then pulls up the cantilever. If some force is acting on the elastic cantilever (for example because some molecule is being stretched between the surface and the tip), this will deflect upward (repulsive force) or downward (attractive force). According to Hooke's law, this deflection will be proportional to the force acting on the cantilever. Deflection is measured by the position of a laser beam reflected by the cantilever.

3 Load the PDB by typing the PDB code in the blank space and clicking `Load`. For this tutorial we are going to work with Mastoparan X, a widely studied toxin, PDB code: 2czp

4 A warning will appear that one or more residues could not be identified. Click in `Select Residue` to open the Residue Selection window. The residue with ID 15 will be marked in RED as it was not recognized. This residue, in fact a NH2 terminal, can be deleted as terminals will be added to the molecule by during the preparation step by autopsf. For that, click in Delete in the right-hand side and click in the residue you want to delete (NH2 Res ID 15) and click in Apply. You may now close the Residue Selection window.

5 To perform SMD simulations, click on the `Steered Molecular Dynamics` tab.

6 Click in `Show options` to select SMD parameters.

7 First we will run a simulation with `Implicit Solvent`, at 0.15 mol/L of Salt and 27C.

8 To unfold Mastoparan X, set Max Length to 50 A and Pulling Speed to 100 A/ns. That will prepare a simulation of 0.5ns.

9 To perform a SMD simulation we need to select an amino acid residue to be pulled. Click in `Pull residues` and a window will open. Select the first residue (Res ID 1) and click Apply.

10 To perform a SMD simulation we also need to select an amino acid residue to be anchored, otherwise our molecule would just be translate during the simulation. Click in `Anchor residues` and a window will open. Select the last residue of Mastoparan X (Res ID 14) and click Apply.

SMD atoms: The VMD display now shows the amino acid residues that are your pulling atoms and your anchoring atoms.

11 To run a `Live Simulation`, make sure the `Live Simulation` box is selected and click `Prepare`. You will have to select a folder where the files will be saved. During this step the autopsfgen function of VMD is called to create the necessary files to run the MD simulation. qwikMD will also create the necessary NAMD configuration files.

12 To run the simulation click in the `Start` Button. Interactive Molecular Dynamics used in the Live Simulation might take a few seconds to start. In the console window is possible that an Error message will show up until the connection between VMD and NAMD is established.

6.1.1 Analysing during a Live Simulation

VMD is a powerful tool for analysis of structures and trajectories and should be used as a tool to think. Numerous tools for analysis are available under the VMD Main menu item Extensions - Analysis. In addition to these built-in tools, VMD users often use custom-written scripts to analyze desired properties of the simulated systems. VMD Tcl scripting capabilities are very extensive, and provide boundless opportunities for analysis. qwikMD provides the user with some of the most employed analysis tools, allowing also the analysis while performing live NAMD sections.

During the simulation, clicking on the `Analysis` tab on the top you can perform a few analysis during the Live Simulation.

13 To check how the equilibration is performed, plot the temperature while running your live simulation.

14 To go faster when training with this tutorial, you can go back to the `Run` tab and click on `Finish` after a few minutes observing the molecule moving. That will allow you to jump to the next step, in this case `eq MD`. You can also do the same and click on `Finish` after a few minutes observing the molecule moving.

Warning!! Always click to FINISH your simulation before leaving VMD or closing qwikMD window. NAMD activities started from VMD will continue to run in background for all the steps requested unless you make sure the simulation was aborted before leaving qwikMD.

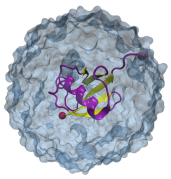
15 To check how much force is necessary to unfold this peptide you can, in the `Analysis` tab, click on `Calculate` the SMD Force.

16 At the same time, check the number of Hydrogen bonds you are breaking by clicking on `Calculate` the `Hydrogen Bonds`.

- **Do you see any correlation between the number of Hydrogen Bonds and the force necessary to unfold the complex?**
- **What if you pull a little slower? How different the Force profile will be?**
- **How is the force behaving at the end of the simulation? Why?**

6.2 Setting-up steered molecular dynamics to study protein complex interaction

To elucidate the molecular mechanisms at play that enable extreme mechanostability in some protein complexes, one can carry out all-atom steered molecular dynamics simulations of such complexes. One of the strongest protein-protein interaction ever found, namely cellulosomes cohesin-dockerin interaction, was characterized using this simulation approach in addition to single molecule force spectroscopy. SMD results showed the force increased with distance until the complex rupture. Forces observed were in the range of hundreds to thousands of pico-Newtons in both simulation and experiment.



Ultrastable protein interactions in cellulsomes. Cellulosomes are protein networks designed by nature to degrade lignocellulosic biomass. These networks comprise intricate assemblies of conserved subunits including catalytic domains, scaffold proteins, carbohydrate binding modules (CBMs), cohesins (Cohs), dockerins (Docs) and X-modules (XMods) of unknown function. Coh:Doc pairs form complexes with high affinity and specificity, and provide connectivity to a myriad of cellulosomal networks with varying Coh:Doc network topology. To know more check: Schoeler, C., et al. "Ultrastable cellulosome-adhesion complex tightens under load." Nature communications 5 (2014).

1 Open VMD by typing `vmd` in the Terminal window.

Launch VMD. Open VMD graphically by double-clicking on the VMD icon under `Applications` in the `Finder`.

2 Open qwikMD by clicking Extensions → Simulation → qwikMD menu item in the VMD Main window.

3 Load the PDB by typing the PDB code in the blank space and clicking `Load`. The ultrastable cellulosomal Cohesin:Dockerin complex that we are studying here can be loaded from PDB code: 4iu3

4 A warning will appear that one or more residues could not be identified. Click in `Select Residue` to open the Residue Selection window. The residues/molecules marked in RED were not recognized.

5 Calcium ions are presented as CA, which is the CHARMM name for alpha-carbons. In the right-hand side, click in Rename and then click in the CA and change the `Res NAME` to `Calcium`. qwikMD will offer to change the other molecules with same name that are also marked in RED. If you are confident about it click yes.

6 On `qwikMD` window go to the `Selection` window and change the representation of the both water groups to `Off`. As you can see, sulfate, the other molecule that was also not recognized, is present over the surface of our protein complex. The sulfate in these positions is only present due to crystallization effect. To remove the sulfate, click in Delete in the right-hand side and click in the residue you want to delete and click `Apply`.

7 To perform SMD simulations, click on the `Steered Molecular Dynamics` tab.

8 Click in `Show options` to select SMD parameters.

9 We will run the simulation with similar parameters that were used in the Nature Communications manuscript described above. For that, select Explicit Solvent, at 0.075 mol/L of Salt and 27C.

10 To break Coh:Doc apart we, set Max Length to 250 A and Pulling Speed to 0.25 A/ns. That will prepare a simulation of 1000 ns. Note that, if no unfolding is observed, the protein complex will break apart much earlier. In this kind of simulation, always check the results periodically to avoid bad usage of computer time.

11 To perform a SMD simulation we need to select an amino acid residue to be pulled. Click in `Pull residues` and a window will open. Select the last residue of chain A - Cohesin (Res ID 210) and click Apply.

12 To perform a SMD simulation we also need to select an amino acid residue to be anchored, otherwise our molecule would just be translate during the simulation. Click in `Anchor residues` and a window will open. Select the first residue of chain B - Dockerin (Res ID 5) and click Apply.

SMD atoms: The VMD display now shows the amino acid residues that are your pulling atoms and your anchoring atoms.

13 To run a simulation like this you will need a computer cluster or a supercomputer, so make sure the `Live Simulation` box is deselected and click `Prepare`. You will have to select a folder where the files will be saved. During this step the autopsfgen function of VMD is called to create the necessary files to run the MD simulation. qwikMD will also create the necessary NAMD configuration files.

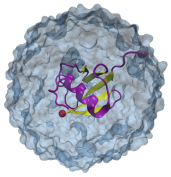
Running your Simulation outside of qwikMD. To run long steered molecular dynamics simulations with NAMD at least four files are required: a Protein Data Bank (pdb) file a Protein Structure File (psf), a force field parameter file and a configuration file. During the preparation steps, where the system might be solvated, ionized, among other procedures, several files are created. qwikMD separates the files created in the preparation step in a SETUP folder, while files needed to run the MD simulations are in a RUN folder. These two folders are created inside the folder defined by the user in working directory window. With the same name as the folder created by the user, a file with .qwikMD extension allows the user to load simulations performed with qwikMD and also previously created preparation steps, like amino acid residues mutations or salt concentration.

To run a simulation prepared with qwikMD in a computer cluster or supercomputer, one needs to copy only the RUN folder.

14 Running: Run your simulation by typing in a Terminal window:

- `namd2 qwikMD_equilibration_0.conf > equilibration_0.log &`
- `namd2 qwikMD_eqMD_1.conf > equMD_1.log &`
- `namd2 qwikMD_smd_2.conf > smd_2.log &`

15 To perform the analysis after running in another computer you use the saved qwikMD file to load your simulation in qwikMD and perform your analysis with the same tools as well as any other VMD tools.



Force Propagation. A combination of network-based correlation analysis and thermodynamic fluctuation theory allows for the visualization of stiff paths through protein complexes along which force is transmitted. For the presented here Coh:Doc complex from cellulosome, analysis of the force path suggests specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces. To learn how to perform network analysis check out the Dynamical Network Analysis tutorial. To know more check: Schoeler, C., et al. "Mapping Mechanical Force Propagation through Biomolecular Complexes" Nano Letters (2015).

7 Conclusions

Computer simulations of biomolecular systems have grown rapidly over the past few decades, passing from simulating very small proteins in vacuum to simulating large protein complexes in a solvated environment. All-atom MD simulations, employing classical mechanics, allowed the study of a broad range of biological systems, from small molecules such as anesthetics or small peptides, to very large protein complexes such as the ribosome or virus capsids. Hybrid classical/quantum MD simulations allowed the study of enzymatic activity or polarizable molecules in biological membranes. However, despite its success, MD simulations are still limited in two regards, inaccuracy of force fields and high computational cost. Such limitations can lead to inadequate sampling of conformational states, which in turn limits the ability to analyze and reveal functional properties of the systems being examined. All relevant states of a system must be reached in simulations in order for its dynamics and function to be meaningfully characterized. Molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes of biomolecules. However, biological molecules are known to have rough energy landscapes, with many local minima frequently separated by high-energy barriers, making it easy to fall into a non-functional state that is hard to jump out of in most conventional simulations.

As discussed, when running MD simulations it is very important to run more than one replica of the same system. Long trajectories usually also helps one to sample different conformations. Therefore a long simulation is important if big conformational changes are expected. If you want to learn more about sampling and molecular dynamics check our group publications at: http://www.ks.uiuc.edu/Publications/Papers/