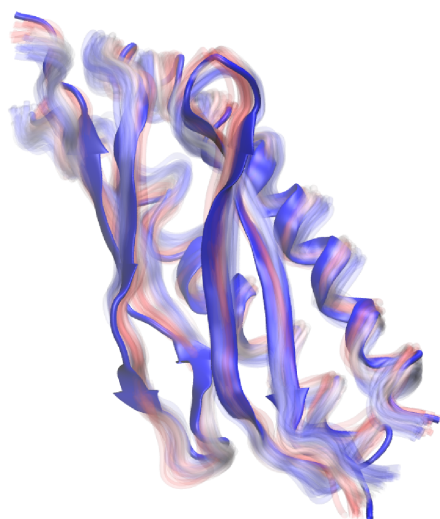


University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group
Computational Biophysics Workshop

QwikMD - Easy Molecular Dynamics with NAMD and VMD



Tutorial by Rafael C. Bernardi, Till Rudack, Joao V. Ribeiro,
Angela Barragan, Muyun Lihan, Rezvan Shahoei and Yi Zhang

June 30 2017

QwikMD Developers:
Joao V. Ribeiro
Rafael C. Bernardi
Till Rudack

A current version of this tutorial is available at
<http://www.ks.uiuc.edu/Training/Tutorials/>

Contents

1	Introduction	3
1.1	NAMD	3
1.2	QwikMD	4
2	Required programs	4
2.1	For Linux/Mac Users:	5
2.2	For Windows Users:	5
3	Installing the Required Programs	5
3.1	VMD	5
3.2	QwikMD	6
3.3	NAMD	6
4	Running my First Molecular Dynamics Simulation	7
4.1	Ubiquitin in implicit solvent	7
4.1.1	Preparing structures and starting simulations	9
4.1.2	Analyzing during a Live Simulation	10
4.2	Ubiquitin in a Water Box	12
4.2.1	Starting a New Simulation	12
4.2.2	Creating a Salt Solution	13
4.3	Running your Simulation outside of QwikMD	14
5	Tackling common scientific problems	15
5.1	Cancer mutation in Ras	16
5.2	HIV protease	19
5.3	Proton transport through a membrane by bacteriorhodopsin	25
6	Steered Molecular Dynamics	31
6.1	Biomolecular interactions during protein unfolding	31
6.1.1	Preparing a SMD system	32
6.1.2	Analyzing during a Live Simulation	33
6.2	Setting-up steered molecular dynamics to study protein complex interaction	34
7	Structure refinement with MDFF	37
8	Closing Remarks	40

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. To assist both experimentalists and any novice to MD to overcome the initial learning curve barrier of MD simulation software, we developed a VMD-NAMD plugin, QwikMD, that connects the widely employed and user-friendly molecular graphics program VMD to the widely used MD program NAMD. By employing QwikMD, one 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 serves 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 in an automatic tool! QwikMD is an automatic tool that aims to prepare as many different biological system as possible, however covering all possible systems that are studied employing MD simulations is a hard quest. Please inspect the system prepared by QwikMD carefully before starting the MD simulations. We are working hard to develop new features and extend QwikMD applicability and if you have suggestions 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 files:

- a Protein Data Bank (pdb) file which stores atomic coordinates and/or velocities for the system. PDB files are available for many proteins at <http://www.pdb.org>, or can be generated using various structure modeling programs.

- 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. For more details check NAMD Tutorial.



Warning! The goal of this tutorial is to introduce QwikMD and NAMD by performing short molecular dynamics simulations. Therefore, the examples provided are optimized so simulations can be done in a reasonable period of time in 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 the user to start and analyze MD simulations. The plugin helps, in particular 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 loading the PDB to generating the NAMD configuration files - are performed automatically but giving the user all the information necessary to reproduce the system created and the simulations performed. The **Live View** 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 them 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:** Included in VMD since the version 1.9.3

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

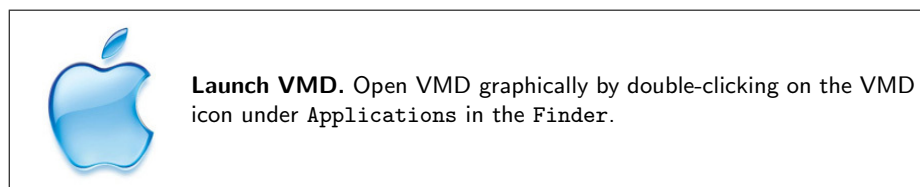
The VMD source code and binary distributions can be obtained after registering at the VMD web page. Download the appropriate distribution file with your web

browser. Windows binary distributions are self extracting, so once the distribution file is downloaded, proceed to the installation directions. For source distributions and Linux binary distributions, uncompress and untar the file. This will produce a subdirectory named `vmd-1.9.3`. Unless otherwise specified, all references to VMD code will be from this subdirector. To install the pre-compiled VMD binaries, `cd` to the extracted folder, open the `configure` file and set the installation directory location (`$install_bin_dir` and `$install_library_dir`) to the desired location. After the edition and save of the `configure` file, `cd src` and excute the command `make install`. One can add the installation location (`$install_bin_dir`) to the `PATH` environment variable for convenience. For more information regarding VMD installation and source code compilation please visit <http://www.ks.uiuc.edu/Research/vmd/current/ig/ig.html>.

3.2 QwikMD

QwikMD: QwikMD is included in VMD in version 1.9.3 and later.

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



- 2 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.

If you start your first MD simulation with QwikMD you will be asked to specify the path to your NAMD installation. The automatic path assignment might fail on some operating systems due to individual permission settings. In case the automatic path assignment fails please set the path manually using the following instructions.

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 operating 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:$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:$PATH
```

If you are new to Linux, visit our guide at:

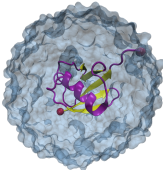
<http://www.ks.uiuc.edu/Training/Tutorials/Reference/unixprimer.html>

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” or “This PC” 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. Also make sure that you (and not only root) have permission to execute `namd2`. 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”).

4.1 Ubiquitin in implicit solvent

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)

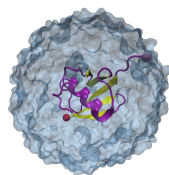
of eukaryotic organisms. Here we are going to simulate Ubiquitin with implicit solvent, where water molecules are represented by a dielectric constant.

- 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 representations. 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” and the color to “Name” 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 has 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

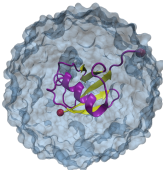
An implicit solvent model is a simulation technique that eliminates the need for explicit water atoms by including many of the effects of solvent in the interatomic force calculation. For example, polar solvent acts as a dielectric and screens (lessens) electrostatic interactions. The elimination of 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 (GBIS) 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 atom's 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 won't 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 Tanner, D.E. *et al.*, J. Chem. Theory Comput., 7:3635-3642, 2011 <http://www.ks.uiuc.edu/Publications/Papers/paper.cgi?tbcode=TANN2011A>.



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.

4.1.1 Preparing structures and starting simulations

- 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 **Chain/Type Selection** 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 View** box under the **Simulation Setup** option and click **Prepare**. You will have to select a folder where the input file (.qwikmd) will be saved. During this step the autopsf 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, a series of operations is performed to prepare the system and the simulation configuration files with the settings selected by user. Two folders will be created in the working directory: files created during the preparation step are stored in the “**setup**” folder, while files needed to run the MD simulations are stored in the “**run**” folder.



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

QwikMD helps the user to prepare NAMD configuration files to run in a range of computers, from the largest supercomputers with high-end parallel platforms to the smallest laptop computers. QwikMD also allows, 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 Equilibration Simulation** Button and select the number of CPU cores to be used to perform the simulation. Interactive Molecular Dynamics (IMD) used in the Live Simulation may take a few seconds to start. In the console window is possible that the message message “**ERROR) Error connecting to localhost on port 3000**” shows up until the connection between VMD and NAMD is established.



NOTE:! The connection between VMD and NAMD during the initiation process of an IMD simulation may take from a few seconds to minutes, particularly in explicit solvent simulations. If the connection is never established, you may want to make sure the network port 3000 is not blocked by the firewall definitions or if any errors were introduced during the preparation phase by checking the simulation log file in the “**run**” folder.



Warning!! Always click **Finish**, **Reset** or close QwikMD window before leaving VMD. NAMD activities started from Live Simulations will continue to run in background for all the steps requested unless you make sure the simulation was aborted before leaving VMD.

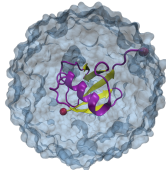
4.1.2 Analyzing during a Live Simulation

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 commonly employed analysis tools, also allowing the analysis while performing live NAMD sections.

During the simulation, QwikMD allows the user to perform common analysis in two analysis tabs: “Basic Analysis” and “Advanced Analysis”. Click on the “Basic Analysis” tab on the top and there you can perform a few analyses during the Live Simulation.

- 7 First, let’s analyze how the Kinetic and Potential energies evolve during the initial phase of the simulation. Expand the “Energies” section by clicking on the ► symbol, check the “Kinetic” and “Potential” boxes and then press “Calculate”. The plot is displayed in the **Plots** section of the tab. Each plot can be exported and saved using one of the options available under the “File” menu of each individual plot.
- 8 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.
- 9 Follow the evolution of the Temperature while the system undergoes a temperature ramp. The option to plot temperature is under the thermodynamics tab
- 10 Calculate the Root-Mean-Square Deviation (RMSD). In the RMSD section you can select the region to be evaluated and the superimposition options. Select “Backbone” as RMSD evaluation region, check the “Align Structure” box and “Backbone” as the alignment region to eliminate system rotational and translational motions from the calculation and then click on **Calculate**.



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 $C\alpha$ 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 Mössbauer 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 .

4.2 Ubiquitin in a Water Box

4.2.1 Starting 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 and the edge size is defined by:

$$box_{edge} = (\sqrt{x^2 + y^2 + z^2}) + 15$$

where x , y and z are the dimensions of structure in the three axis. Box dimensions in Å. The big water box is 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 consists in 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.

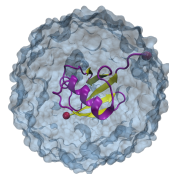
Different water molecule models can be used in an MD simulations. NAMD currently supports the 3-site TIP3P water model, the 4-site TIP4P water model, and the 5-site SWM4-NDP water model (from the Drude force field). As the standard water model for CHARMM force field, TIP3P is the model employed in the simulations prepared with QwikMD.

- 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.



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.

- 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



Structure Manipulation. This button will open a new window titled Structure Manipulation/Check where the user can perform point 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. Structure Manipulation/Check 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, in the example described further in the Cohesin-Dockerin complex in the section 6.2, the Ca^{2+} ions have the residue name CA, whereas this residues is defined in the CHRAMM force field with the residue name CAL. Structure Manipulation/Check allows the user to rename CA ions to proper Calcium parameters that will be compatible with the CHARMM force field.

- 4 To perform simulations in explicit solvent we need to change part of QwikMD standard protocol. For that, first select the Explicit solvent model.

4.2.2 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.

- 5 Perform a simulation in explicit solvent is not an easy task for your computer. Change the Simulation Time, under the protocol drop arrow, to 0.1 ns, to be able to perform this tutorial simulation in a more realistic time. Depending on the processing power of your machine, this simulation may still take some time. Please notice that “Simulation time” refers to the production “MD” protocol step.
- 6 To run a simulation in background, make sure the **Live View** box is de-select 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.
- 7 To run the simulation click the **Start Equilibration Simulation** Button. VMD windows will be unavailable until the simulation is finished.

4.3 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 stores the files created during the preparation step in the “**setup**” folder, while files needed to run the MD simulations are stored in the “**run**” folder. These two folders are created inside the folder with the same name as the QwikMD input file (`*.qwikmd`) defined by the user when preparing the system. This input file (with `.qwikmd` extension) allows the user to load the simulations performed with QwikMD and also previously structure manipulation operation and simulation options, like amino acid residues mutations or salt concentration.

To run a simulation prepared with QwikMD in a computer cluster or super-computer, one needs to copy only the “**run**” folder.

- 1 Running simulation in a command line

In a terminal window, navigate to the “run” folder of the simulation prepared by QwikMD and type the following command:

```
namd2 +p <Number of CPUs to parallelize to>
qwikmd_equilibration_0.conf > qwikmd_equilibration_0.log &
```

Analyzing Trajectories

To analyze a simulation executed outside of QwikMD, one can load the file with `.qwikmd` extension. If any trajectory was generated, meaning if the files with the extension `.dcd` exists in the “run” folder, QwikMD will load the trajectories into VMD.

- 2 Load the file “Ubiquitin.qwikmd” from the tutorial files using the Load button in the **Simulation Setup**.
- 3 Select **Initial Structure**, `qwikmd_equilibration_0` and `qwikmd_production_1` trajectories to be loaded.

Once the frames are loaded into VMD, one can performed the analysis present in both “Basic” and “Advanced” tabs.

Image Rendering

QwikMD offers a very simple way to render your VMD scene and save it as a picture, however the options available are limited. For more advanced rendering options, please follow the “VMD Images and Movies Tutorial” in our website: <http://www.ks.uiuc.edu/Training/Tutorials/>.

- 4 To render the current VMD scene displayed in the VMD display, select the “Basic Analysis” tab and beneath the “Thermodynamics” you can select the option for the image render.
- 5 Select the image resolution and the image processing method (render) to be use to produce the image.
- 6 Press Render and select the location and filename to save the image.

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.

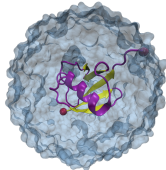
5.1 Cancer mutation in Ras

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 GDP and Pi, Ras is switched from the active GTP-bound on-conformation 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. Indeed, about 20% of all cancer tumors are associated to malfunctions of the Ras protein through the so called oncogenic mutations, which most of 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 the oncogenic mutation, you need first 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 using the PDB code 1QRA and click Load.

A warning will appear that one or more residues could not be identified. To prepare the MD simulation of the Ras protein we need to introduce the correct parameters for the GTP molecule into QwikMD. The Structures in the PDB often times consist of not only standard amino acids but also DNA, ligands or other small molecules. If the structure loaded contains non-standard molecules for which the CHARMM force field parameters are not included in QwikMD, you will not be able to prepare the MD simulation without further modification. If no parameters exist in the literature, you will need to parameterize the molecule. Alternatively you can delete the non-identified part if it is not relevant for structural properties or function. But usually this is not the case. The non-identified parts are colored with throbbing colors in your structure and highlighted in red in the **Structure Manipulation/Check** window. In case of Ras the ligand GTP could not be 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 **Structure Manipulation**. You can see all non-identified parts highlighted in red (Fig. 1). In this case, to add parameters for non-identified GTP, click on **Add Topo+Param** and then hit + to select `toppar_all36_na_nad_ppi_gdp_gtp.str` from the tutorial files. Change the type of GTP from nucleic to hetero and press **Apply** (Fig. 2). Hit **Reset** in the QwikMD main window and the force field parameter for GTP is now successfully added to the database.

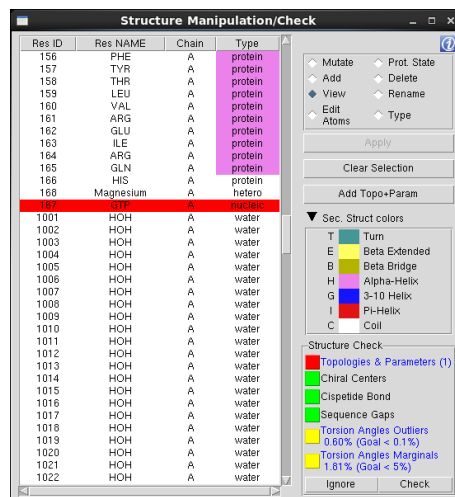


Figure 1: Unrecognized residues are mark in red in the **Structure Manipulation/Check** window.

- 3 Reload the Ras structure and now all molecules of the structure should be recognized correctly by QwikMD. Beside the hetero atoms of GTP

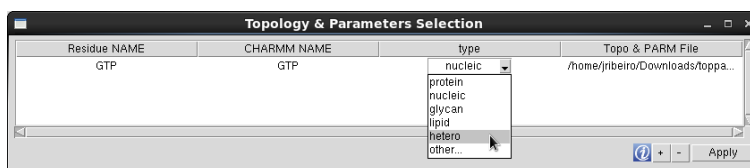


Figure 2: In the **Topology & Parameters Selection** window one can add missing topology and parameters information for non-standard residues.

(residue ID 167) and the magnesium ion (residue ID 168), represented as a small sphere, the Ras structure has some waters molecules resolved in the crystal, which are revealed as spheres in a VDW representation. These waters can play a crucial functional role. However, for simplicity we here delete them by clicking on **Chain/Type Selection** 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.

Analyzing Trajectories

- 5 Once 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 **Calculate** to open a window for the kinetic and potential energy plot. In order to generate a plot to follow the temperature press the **Plot** button in the **Conditions** panel.
- 6 Press **Finish** to stop the equilibration. Navigate in your file browser to the run folder in your work directory. The file `qwikmd.equilibration_0.restart.coor` is the last state structure shown on the screen as you pressed finish. Load the psf file “`mol new <psf file name>.psf`”, followed by the `qwikmd.equilibration_0.restart.coor` “`mol addfile qwikmd.equilibration_0.restart.coor`” into the VMD session and save the coordinates of all atoms to a file named `Ras_WT_equil.pdb` File → Save Coordinates... → Selected Atoms: “all” → Save.

Ras G12V mutation simulation

- 7 Redo steps one and two for the Ras wild type simulation and then go to the **Structure Manipulation/Check** window. Select **Mutate** in the table options. Click on **GLY** with the residue ID 12 and choose **VAL** 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.
- 8 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 perform the same steps described before to generate the pdb file to **Ras_G12V_equil.pdb**.

Comparison of the RasWT and Ras G12V mutation simulations

- 9 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 **rename 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**.

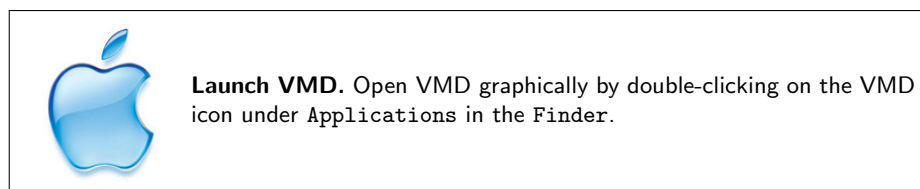
You can observe that there are almost no overall structural differences between the inactive oncogenic G12V Ras mutation (red) and the active Ras wild type structure (green). In addition the site chain atoms added by introducing valine instead of glycine do not have an effect on the structure or positioning 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.

5.2 HIV protease

In this section, we are going to use QwikMD to prepare HIV-1 protease's molecular dynamics simulations. HIV-1 protease is a retroviral aspartyl protease and it is essential for the HIV life-cycle, the well known retrovirus causing the Acquired Immune Deficiency Syndrome (AIDS). This protease has been extensively studied, either computationally and experimentally, and several structural properties can be found in the literature. In this section we will use MD simulations

to analyze the structure in the bound and unbound state. The HIV-1 protease “flaps” are a two very distinctive loops that enclose the substrate in the binding site and are very stable in the presence of the substrate (bound state). However, in the absence of the substrate (unbound state), the “flaps” become very flexible, where the backbone can move as large as 7 Å (Miller *et al.* Science, 246, pp: 1149-1152, 1989.).

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



- 2 Open QwikMD by clicking Extensions → Simulation → QwikMD menu item in the VMD Main window.
- 3 To perform a MD simulation of the HIV protease structure, go to **Advanced Run** tab. Load the PDB by typing in PDB ID: 1KJF and press the Load button.
- 4 A warning will appear that one or more residues could not be identified. Always make a visual check by rotating and looking at details of the molecular system. In the Display window, molecules with throbbing colors were not identified by QwikMD. Fig. 3
- 5 To facilitate visualizing the complex structure, one can turn off the representation of water molecules present in the crystallographic structure. Click in “VDW” representation of water and select “OFF”. Fig. 4
- 6 A visual inspection of the loaded structure shows a few small molecules and ions with their colors throbbing. QwikMD did not identify these molecules and ions as having force field parameters available, at least not with the name they are presented in the PDB file.
- 7 Go to **Structure Manipulation/Check** window to check what problems are present. Here, certain topologies and parameters are missing, and that some torsion angles are considered outliers and marginals in the Ramachandran plot.
- 8 Check the molecules and ions that have missing topologies and parameters, which are marked in red. The name ACT for Acetate is not the same as the name in the CHARMM36 force field (Fig. 5)

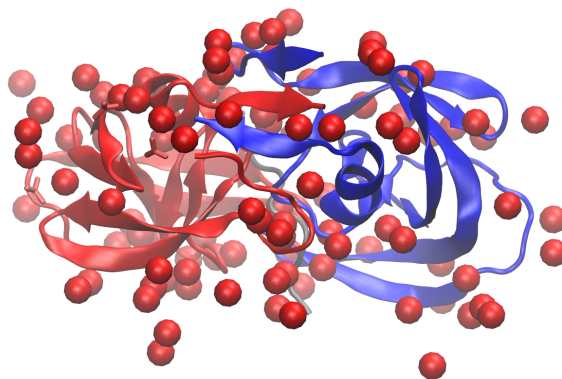


Figure 3: HIV-1 protease (PDB code 1kjf) loaded into VMD by QwikMD. During the loading process, QwikMD checks the structure for possible problems and assign representation schemes for easy structure identification. The protein is represented in NewCartoon and colored in red, blue and gray, the water oxygen atoms as VDW and colored in red and the acetate ion molecules is represented as Licorice and with throbbing colors to mark the unrecognized residues.

- 9 To rename the Acetate, select “Rename” and click on the name ACT. A list of possible substitutes will become available, select Acetate. QwikMD automatically detects that other molecules have the same name (rename) and asks if they should all be renamed. Click “Yes”.
- 10 Now that QwikMD recognizes the residue, another check routine compares if the atoms name in the structure correspond with the atom names defined in the force field for the selected residue name. In this case, none of the atoms have the correct name, so QwikMD ask if the user wants to assign the correct name. Click “Yes” and rename the atoms as shown in the Fig. 6.
- 11 After renaming ACT to Acetate and change the atoms name, check if the problems were fixed by clicking the “Check” button. Here, the Ramachandran Outliers can be ignored. Click on ‘Ignore’. To learn how to fix a problem found by QwikMD, click on the hyperlink (in blue) when a problem is found.
- 12 Close the **Structure Manipulation/Check** window and select the **Advanced Run** tab. Here, select “Explicit Solvent” as solvent model, check the box **Minimal Box** and **Buffer** of 20 Å. Then, select 0.15 mol/L as concentration of salt NaCl. Set the “Temperature” to 27 °C. In this section

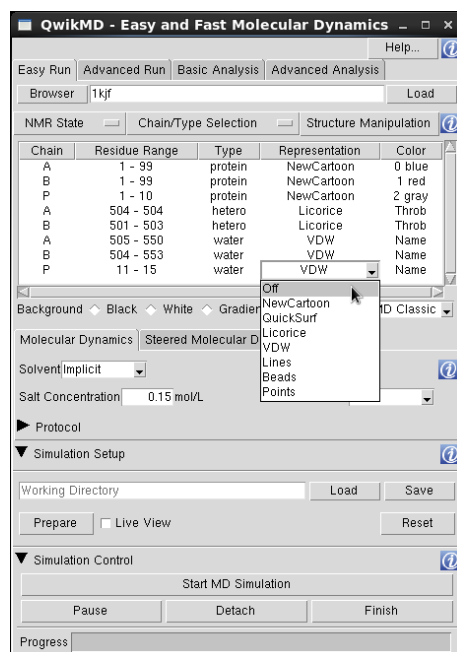


Figure 4: QwikMD main window after loading the HIV-1 protease (PDB code 1kjf). The main table presents the information regarding the different chains of the structure and the correspondent representation and coloring mode. One can use this table to change the chain representation by clicking in the value intended to change and selecting one of the options from the drop-down menu.

we will execute the default protocol steps in the MD tab (Minimization, Annealing and Equilibration), and changing the “n Stpes” of the MD step from 500,000 to 10,000,000, meaning from 1 ns to 20 ns (the default integration step is 2 fs).

Note: To reduce the size of the files generated by NAMD, you can select one of the protocol steps and click on the **Edit** button in the protocol table to edit the configuration files manually and change the saving frequency values. Be careful when changing the saving frequencies of DCD (trajectory file), energies or SMD forces, since the amount of data generated can be insufficient for a proper statistical analysis of the property to be computed.

- 13 Select the MD protocol step and then click on the + button to extend the simulation for another 20 ns. Depending on where you are running your simulations, some time is beneficial to split one big simulation in several smaller steps to overcome, for instance, overcrowded supercomputer queues and clusters walltime limitations.

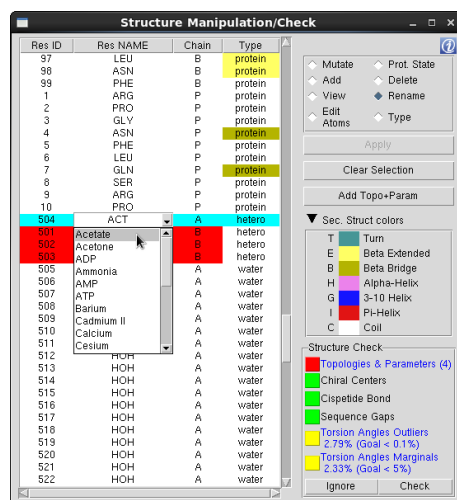


Figure 5: The user must use **Structure Manipulation/Check** window to change the name of the unrecognized residues (marked in red) to use the correct parameters in the CHRAMM force field. To rename the residue, select “Rename”, click on the residue in the table and select the correct name from the drop-down menu.

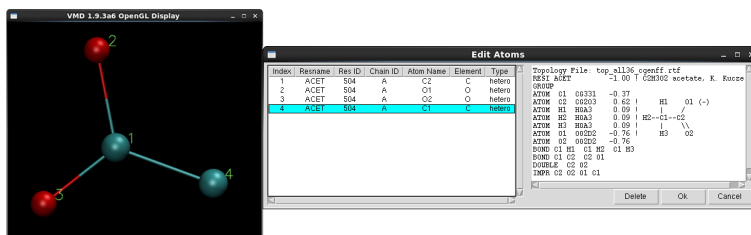



Figure 6: In the **Edit Atom** window, one can change the residue identifier number (Res ID) and the atoms name to match the topology file of the CHARMM force field. This window also display the topological information found based on the residue name in the **Structure Manipulation/Check** window and isolate the molecule to be edited in VMD OpenGL window with the matching atoms indexes for reference.

The duration of the MD simulations must be according to the system in study and its stability must be evaluated before collecting frames for analysis. Simulation’s stability can be evaluated by analyzing, for instance, the RMSD, the different energies components, temperature and pressure. For tutorial purposes we will simulate 40 ns, whereas in a scientific application research project, one should extend this time for timescales encompassing the structural motions

in study.

	<p>Warning!! The <code>Minimal Box</code> option reduces significantly the size of the water box. This option rotates the system to reduce water box volume and add water molecules to complete a padding of at least the value selected in <code>Buffer</code> on the surface of the structure. This option must be used carefully to ensure the protein does not interact with its image in the next cell if periodic boundary conditions are used.</p>
---	--

- 14 To prepare all the files necessary for the simulation, click “Prepare”. This step can take a few minutes.
- 15 A visual inspection is always important. Make sure everything looks as expected. QwikMD automatically makes a very simple water box representation, helping users to better observe the protein and the substrates inside the water box.

You prepared the bound state of the HIV-1 Protease. To prepare the unbound state, Reset QwikMD and follow steps from **2** to **13**. In this next step we want to delete the substrate, which can be simply done by:

- 16 On the `Select Chain/Type` menu of QwikMD main window, deselect the “chain P and protein”. In this way, the substrate will not be considered during the preparation phase.
- 17 Prepare the system with a different name for future comparison between bound and unbound state.
- 18 To start the simulation in the local computer, click on the “Start Minimization Simulation” button.

Analyzing Trajectories

The final results of the simulations prepared previously are located in the supporting files of this tutorial under HIVProtease folder with the names of “HIVprotease_with_substrate.qwikmd” and “HIVprotease_without_substrate.qwikmd”.

Note: The resulting trajectories were stripped of waters and solvent ions mole for file size reduction. You can remove the water molecules and solvent ions molecules from your simulations by selecting the options “Don’t load water molecules” and “Don’t load solvent ion molecules” in the **Loading Trajectories** window. This process can take a few minutes, however is executed once. Options only available for the Unix (Linux & Mac) VMD versions. In this window, you can also chose to not load all the frames saved in the trajectory files by increasing the “Stride” number present in the bottom of the window (available for all platforms).

- 19 Load the “HIVprotease_with_substrate.qwikmd” file using the Load button in the **Simulation Setup** section.
- 20 Select the **Initial Structure**, MD and MD.1.
- 21 Calculate the RMSD, with the “Align Structure” option on and the selection “(chain A B) and protein” for both RMSD calculation and alignment entries, and export the plot using the option “Export to ASCII matrix” under the **File** menu in the plot window.
- 22 Perform the same analysis for the unbound state by loading the “HIVprotease_without_substrate.qwikmd”. You can perform this analysis in the same VMD session or opening another VMD so more analysis can be performed to both systems simultaneously.
- 23 Plot both calculations for better comparison using your favorite plotting program.

The RMSD analysis returns larger RMSD values and large amplitude of the RMSD values for the unbound state when comparing to the bound state. You can isolate the RMSD calculation only to the “flaps” region (residues numbers 35 to 57 of the chain A and B) by using the selection “(resid 35 to 57) and protein” in the RMSD calculation atom selection. You can also calculate the RMSF (Root Mean Square Fluctuation) to have an insight of which residues are more flexible.

5.3 Proton transport through a membrane by bacteriorhodopsin

Bacteriorhodopsin is a membrane protein in archaea in charge of pumping protons across a membrane. Powered by the green light, bacteriorhodopsin produces an electrochemical proton gradient across the membrane which in turn is used by the ATP synthase to produce ATP. Deep inside bacteriorhodopsin, a molecule called retinal is buried. Retinal, which contains a string of carbons, changes conformation under light absorption from a straight form to a bent form. This conformational change makes the proton transport across the membrane possible.

In the next section, QwikMD is used to prepare a membrane embedded bacteriorhodopsin model for MD simulations.

Bacteriorhodopsin simulation

- 1 Restart VMD.
- 2 Open QwikMD by clicking **Extensions** → **Simulation** → **QwikMD** menu item in the VMD Main window.
- 3 Open **Structure Manipulation** window.

- 4 Press **Add Topo+Param**. We need to add topology information and force field parameters for retinal.
- 5 Press **+** and choose **RetinalTop.str** in the browser.
- 6 In the **Topology & Parameters Selection** window for **type** choose **protein** from the drop-down menu.
- 7 Press **Apply** and continue with **Yes**.
- 8 Reset **QwikMD**.
- 9 Load the structure of bacteriorhodopsin by typing the PDB ID **1QHJ** in the blank space and then clicking **Load**.
- 10 Proceed with **OK**.
- 11 Select **Advanced Run**.
- 12 Click on the **Structure Manipulation** window.
- 13 In bacteriorhodopsin, residue Lysine (LYS) 216 of the Bacteriorhodopsin is linked to a protonated Schiff's base, which by its turn is bound to a retinal molecule. In the PDB file, the Schiff's base and the retinal are considered one residue with the resid 300 and rename RET. However, in the topology file you loaded earlier, the complex Lysine, Schiff's base and Retinal is defined as one single residue with the rename LYR. In order to treat LYS + Schiff base + Retinal as one residue, they have to be defined as the same residue. In the following few steps we will change the Res NAME for both LYS 216 to LYR and change the Schiff base + Retinal residue from 300 to 216.
- 14 Rename the residue LYS 216 to LYR by checking the **Rename** option (on the top right corner of the window) and clicking on LYS next to Res ID 216 and choose LYR from the drop-down menu.
- 15 Press **No** when prompt to rename all residues named LYS.
- 16 Press **Yes** when prompt to change the atoms' names to match CHARMM topologies.
- 17 Rename the nitrogen atom with **index** 9 from **NZ** to **N16** to be consistent with the topology and parameters provided for LYR.
- 18 Press **Ok**.
- 19 Change the table option to **Type** to change the residue type of Res ID 300 from hetero to protein.

- 20 Rename the Res ID 300 from RET to LYR.
- 21 Check option next to Edit Atoms.
- 22 Select Res ID 300 and press Edit.
- 23 Change the Res ID from 300 to 216 for all the atoms by clicking on 300 and typing in 216.
- 24 Press Ok. You can find the compound residue (resname LYR) in Fig. 7 shown in Licorice; the carbon chain is colored in cyan.

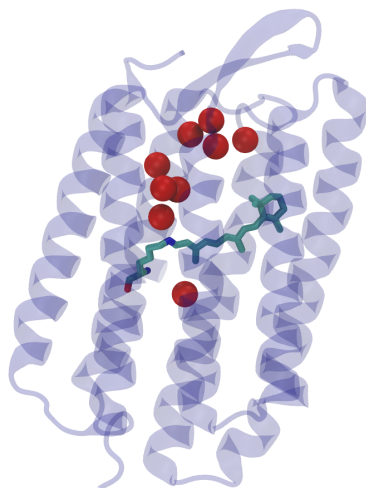


Figure 7: Retinal (in licorice representation) embedded deep inside bacteriorhodopsin.

- 25 To delete the lipids, check the circle next to Delete and select the lipids (Res NAME PH1) and press Apply.
- 26 Also delete the water molecules that are not internally bound to the protein, which are RES ID 405 and RES IDs 410 to 444.
- 27 In the same window, open the Membrane menu on the right panel and for the Box dimensions choose 100x100. Click on Box to display the borders of the membrane and make sure the protein is placed reasonably in the membrane (See Fig. 9). You can use Translate and/or Rotate and the --, -, +, ++ buttons to adjust the box to the protein transmembrane region. Click on Generate once you are satisfied with the box's position.

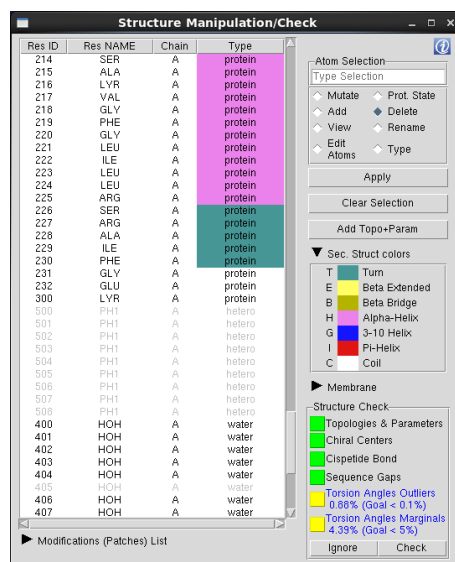


Figure 8: The Structure Manipulation window once all the topologies and parameters exist.

- 28** In order to prepare a 10 ns simulation for bacteriorhodopsin in a membrane surrounded environment with explicit solvent, in the QwikMD main menu choose **Explicit** from the **Solvent** drop-down and **15 Å** for **buffer**.

Note: When the protein is inserted in the membrane, QwikMD selects the solvent model “Explicit Solvent” and disables the “Minimal Box” option, since the solvent box for membrane complexes is created in a slightly different way. To create the effect of a “infinite” membrane, the periodic boundary conditions are defined to be the limits of the membrane (in the membrane axis) and any water molecules is added to the membrane axis. The solvent molecules are only placed above and beneath the membrane, with a buffer of the value selected by the user.

- 29** Go to the **Protocol** tab and click on **n Steps** of the MD protocol and change the number of steps from 500,000 (1 ns) to 5,000,000 (10 ns).
- 30** Press the **Prepare** button under the **Simulation Setup** section of QwikMD main window.
- 31** You can either run the simulation on your machine by pressing **Start MD Simulation** in the **Simulations Control**. QwikMD automatically recognizes the maximum number of processors and uses them as default but you can use fewer if you want. Press **OK** to start the simulation. A note will pop up stating that the simulation is run in the background so

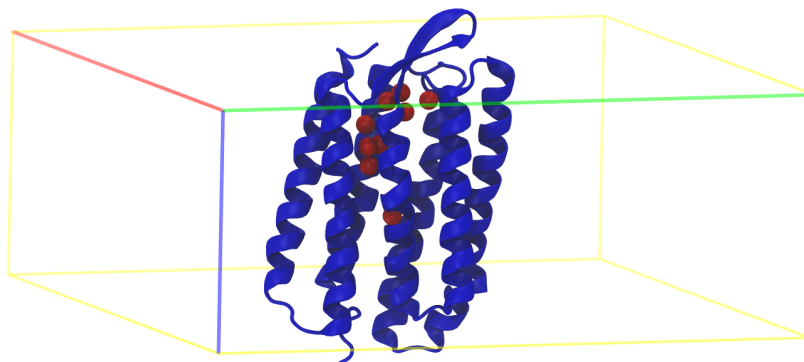


Figure 9: Bacteriorhodopsin placed in a box, prior to membrane being generated.

the VMD window will freeze. Proceed with **Yes**. Alternatively you can start each of the steps, for example Minimization, Annealing, Equilibration, ... in the terminal by going to the run folder and execute NAMD with the command: `namd2 +p <Number of CPUs to parallelize to> Filename.conf > Filename.log & .` Make sure that you execute the protocol steps in the same order as they appear in the protocol table, since every step is dependent on the files of the previous step. Note that whether you start through VMD or the terminal you will have to restart each of the steps manually. It does not happen automatically. Running 10 ns for all the systems could take a couple of days, so if you do not want to wait until the simulations are done you can continue right away with the analysis using the supporting files of this tutorial, under the “Bacteriorhodopsin”. Note that the dcd writing frequency was set to 100000 steps (every 200 ps) just for practical reasons of the tutorial to avoid producing large dcd files. For a real simulation the dcd file should be written more frequently. One can always remove steps of the trajectory or parts of the simulation system like the solvent (or hydrogen atoms) during the trajectories loading process in QwikMD as explained in the section 5.2.

Analysis of the trajectory

- 32** Go to the **Simulation Setup** tab and press load and load the .qwikMD of the system you want to analyze.
- 33** You can choose the parts of the simulation you want to load in the **Loading Trajectory** window. As the trajectory files could become very big you can reduce the size of data loaded in by the stride option or by filtering out the parts of the system like solvent molecules or hydrogen atoms.

- 34** To visualize the changes in secondary structure you have to download `sscachel.tcl` from the tutorial files into your working directory and open the TK console in VMD through **Extension** → **TK console** and source the file by typing `source sscachel.tcl` and `start_sscachel` in the TK console.
- 35** You can go through the trajectory by using the slider in the VMD main window.
- 36** Go to the **Basic Analysis** tab in QwikMD window and open the **RMSD** menu. Check the **Align Structure** box and press **Calculate**. Then open the **Plot** menu and you can analyze the RMSD there. You can export the plot to an postscript file, an xmgrace file or to an ASCII matrix or vector file for further analysis in any plot program.

Vacuum and explicit solvent simulations

In this section, we will set up two other independent simulations of 10 ns in different environments: one in vacuum and one in explicit solvent (without the membrane environment).

- 37** In order to prepare the systems in both cases follow steps **1** to **25**. Follow step **26** for the vacuum simulation but not for the explicit solvent simulation.
- 38** To set up the simulations: For the vacuum simulation, choose **Vacuum** in the drop-down menu for **Solvent** (Note that for vacuum simulation NVT ensemble is considered.). For the explicit solvent simulation, choose **Explicit** in the drop-down menu for **Solvent**, check **Minimal Box** and select 20 Å for **Buffer**. And then follow steps **29** to **31** for both the simulations.



Warning!! The **Minimal Box** option reduces significantly the size of the water box. This option rotates the system to reduce water box volume and add water molecules to complete a padding of at least the value selected in **Buffer** on the surface of the structure. This option must be used carefully to ensure the protein does not interact with its image in the next cell if periodic boundary conditions are used.

- 39** Follow steps **32** to **36** for each of these simulations to load the corresponding trajectory and perform basic analysis, e.g. calculate RMSD.

Comparing Vacuum, explicit solvent and membrane + explicit solvent simulation results for the Bacteriorhodopsin

- 40** Plot RMSD for all the three simulation systems. You will observe that for any simulation RMSD increases along the trajectory before it plateaus. Also, the smallest RMSD is obtained for the membrane-embedded Bacteriorhodopsin simulation. The reason is the following. RMSD measures

the deviation from the initial structure which in this case was the crystal structure 1QHJ. In order to grasp the most natural conformation for Bacteriorhodopsin, which is a membrane protein, the protein was expressed in a lipid environment for crystallography. Therefore, the simulation of membrane-embedded Bacteriorhodopsin results in a more similar structure to the crystal structure (i.e. smaller RMSD values) than the structures obtained from simulations in vacuum or water environment (i.e. larger RMSD values).

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.

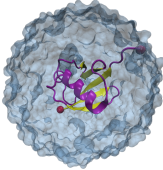
6.1 Biomolecular interactions during protein unfolding

- 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.

6.1.1 Preparing a SMD system

- 3 Load the Deca-alanine structure located in the supporting files for this tutorial by clicking on the **Browser** button, select the file "smd.ini.pdb" and then click on **Load**.
- 4 A warning will appear showing possible problems found in the structure. Click in **Structure Manipulation** to open the Residue Selection window and check the identified problems in the **Structure Check** section. The possible problem identified in this structure is related with the Ramachandran plot outliers, which is not problematic for our case study.
- 5 To perform SMD simulations, click on the **Steered Molecular Dynamics** tab within the **Easy Run** tab.
- 6 Click in **Protocol 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 Deca-alanine, set **Pulling Distance** to 20 Å and **Pulling Speed** to 1000 Å/ns. That will prepare a simulation of 0.02 ns.
- 9 To perform a SMD simulation we need to select an amino acid residue to be pulled. Click in **Pulling 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 **Anchoring residues** and a window will open. Select the last residue of Deca-alanine (Res ID 10) and click **Apply**.

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

- 11 Since we are running the simulation in implicit solvent and the system in study is relative simple, we will skip the “Equilibration” and “MD” and execute the “SMD” in the first place. Deselect the steps “Equilibration” and “MD” under the protocol section and leave only the “SMD” selected.

Note: The SMD forces profile is strictly connected with the intra-molecular interactions of the different residues. Make sure that you start a SMD simulation from a well equilibrated structure, particularly if explicit solvent model is used.

- 12 To run a **Live Simulation**, make sure the **Live View** box is selected and click **Prepare**. You will have to select a folder where the files will be saved. During this step the autopsf function of VMD is called to create the necessary files to run the MD simulation. QwikMD will also create the necessary NAMD configuration files.
- 13 To run the simulation click in the **Start Simulation** Button. Interactive Molecular Dynamics (IMD) used in the Live Simulation may take a few seconds to start. In the console window is possible that the message message “(ERROR) Error connecting to localhost on port 3000” shows up until the connection between VMD and NAMD is established.



NOTE:! The connection between VMD and NAMD during the initiation process of an IMD simulation may take from a few seconds to minutes, particularly in explicit solvent simulations. If the connection is never established, you may want to make sure the network port 3000 is not blocked by the firewall definitions or if any errors were introduced during the preparation phase by checking the simulation log file in the “run” folder.

6.1.2 Analyzing 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 **Basic Analysis** or **Advanced Analysis** tab on the top you can perform a few analysis during the Live Simulation.

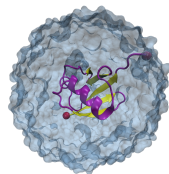
- 14 To check how much force is necessary to unfold this peptide you can, in the **Advanced Analysis** tab, click on **Calculate** the SMD Force.
- 15 At the same time, check the number of Hydrogen bonds you are breaking by clicking on, in the **Advanced Analysis** tab, **Calculate** the Hydrogen Bonds.

- 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?

1000 times slower: Prepare the same simulation, but this time select a pulling speed 1000 times slower, meaning 1 Å/ns and execute the simulation in the background (Live View option off). Analyze the forces profile and compare with the simulations performed previously. You can find the final results for this simulation in the supporting files for this tutorial, under the “SMD” folder, “Deac-alanine.qwikmd”. You can read more about Deca-alanine SMD experiments in the paper also located in the supporting files for this tutorial.

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 cellulosomes. 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. In the Display window, a few ions and small molecules with throbbing colors were not identified by QwikMD.
- 5 On QwikMD window click in the `Chain/Type Selection` and deselect water molecules. As you can see in the Display window, sulfate, the molecule that was 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, go to `Structure Manipulation` and click in `Delete` in the right-hand side and click in the residue you want to delete and click `Apply`.
- 6 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 `Apply`.
- 7 The `Structure Check` at the right bottom shows possible structural problems including missing topologies and parameters. To learn how to fix a problem found by QwikMD, click on the hyperlink (in blue). To ignore outliers, click in `Ignore`. Before closing the `Structure Manipulation/Check` window, click in `Check` in the `Structure Check` to confirm that no other problems are present. After check the structure, press “Ignore” so mark that you acknowledge the cispeptide bond present in the structure
- 8 To perform SMD simulations, click on the `Steered Molecular Dynamics` tab.
- 9 Click in `Protocol options` to select SMD parameters.
- 10 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.
- 11 To break Coh:Doc apart we set “Puling Distance” to 250 Å and “Pulling Speed” to 0.25 Å/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.

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

13 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 **Anchoring 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.

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

Note: In very long simulations, like the one prepared in this example, one can split the SMD simulations in several small steps to overcome, for instance, overcrowded supercomputer queues and clusters walltime limitations. For that, use the **Advanced Run** tab and create extensions of the SMD simulations of 10 or 100 ns.

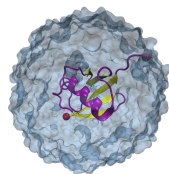
Running your Simulation outside of QwikMD. To run long steered molecular dynamics simulations with NAMD at least four files are required: a Protein Data Bank (pdb) file a Protein Structure File (psf), a force field parameter file and a configuration file. During the preparation steps, where the system 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.

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

- `namd2 qwikmd_equilibration_0.conf > qwikmd_equilibration_0.log &`
- `namd2 qwikmd_production_1.conf > qwikmd_production_1.log &`
- `namd2 qwikmd_production_smd_2.conf > qwikmd_production_smd_2.log &`

- 16 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 Structure refinement with MDFF

MDFF for cryo-EM The molecular dynamics flexible fitting (MDFF) method can be used to flexibly fit atomic structures into density maps. The method consists of adding external forces proportional to the gradient of the density map into a molecular dynamics (MD) simulation of the atomic structure. In this section, we are going to apply MDFF to Adenylate Kinase. The files necessary for this section are located in the supporting files of this tutorial under the QwikMDFF folder.



Windows Installation! Currently there is an issue with using MDFF related tools on Windows computers. In order to use MDFF (and the volutil plugin) on Windows, you will need to work in the following directory: C:\Users\Username\AppData\Local where Username is your Windows username. You can create a new folder to work in, but it must be under that directory. Also, this may not be under the C: drive on your installation, but should be under whatever drive your 'Users' folder is located. To see the AppData folder, you may need to turn on "Hidden items" under the "View" menu in File Explorer. You should "cd" to the directory on the TkConsole or set the directory as your working directory in the MDFF GUI.

- 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 To prepare a MDFF simulation, remember to go to **Advanced Run** tab. Load the PDB by using **Browser** to load the file `1ake_docked-noh.pdb`. Fig. 10

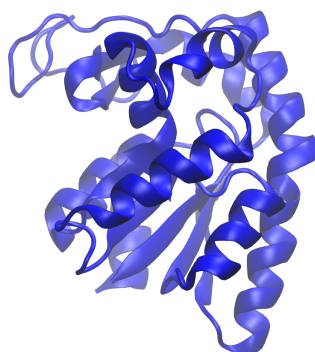


Figure 10: Initial model for the Adenylate Kinase

- 4 A warning will appear that one or more potential problems were detected. Always make a visual check by rotating and looking at details of the molecular system. In the Display window, molecules with throbbing colors were not identified by QwikMD. Fig. 11
- 5 Go to **Structure Manipulation** window, click in **Check** in the Structure Check to confirm that no problems are present. Here, the Ramachandran Outliers can be ignored. Fig. 12
- 6 To perform MDFF, click on the **MDFF** tab. We are using the default values under the **Protocol Option**, which means that none of the protein atoms are fixed and restraints are applied to maintain the secondary structure, cis peptide bonds, and the chirality during MDFF.
- 7 Check the **Live View** and click **Prepare** to set up a working directory where the input files for MDFF are created.
- 8 After that, QwikMD closes and directs you to the MDFF GUI.
- 9 In the MDFF GUI, click on the **MDFF File** menu. Add the target density for the MDFF run. Set up a Grid Scaling Factor of 0.3. After the density is loaded, click on **Generate NAMD files**. Fig. 13

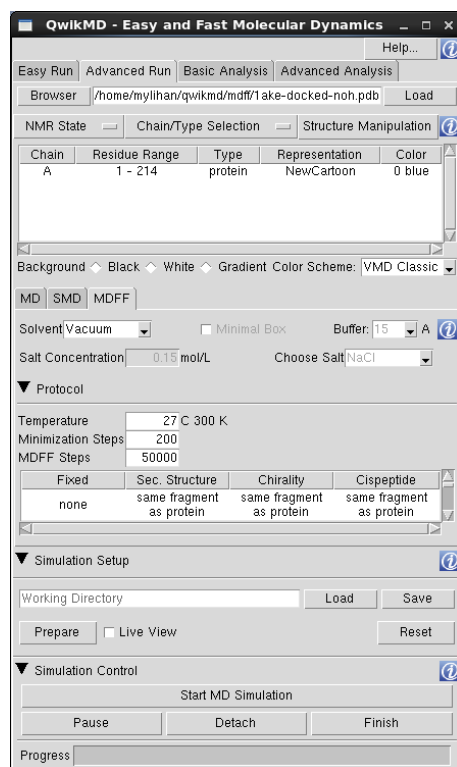


Figure 11: QwikMD main window after loading 1ake_docked-noh.pdb.

- 10 You may want to change the representation of the molecule. Open the Graphical Representation Window of VMD. Change the representation to the density to an Isosurface drawn as Solid Surface with a Transparent Material. Fig. 14
- 11 In the MDFF GUI, go to the MDFF Connect tab.
- 12 To check the quality of the fit to the density on the fly, click on the **Cross Correlation Analysis** menu and specify the mol ID of the density file and its resolution (5 Å).
- 13 To start the iMDFF run, press first Submit and then Connect. Depending on your hardware, it might take between minutes up to half an hour until the initial structure is successfully fitted to the density.
- 14 Compare your fitted result with the experimentally observed model for the target conformation by adding the experimental structure to the VMD through File → New Molecule, Browse and Load 4ake.target.pdb, from the folder you have downloaded in step 1.

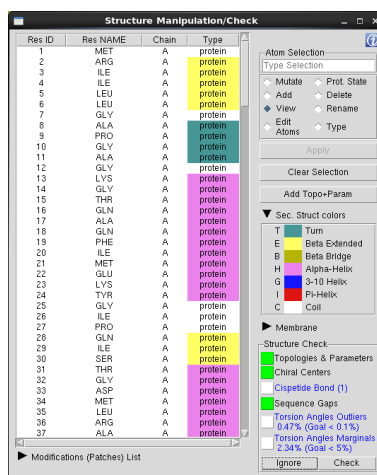


Figure 12: The Structure Manipulation window. Check if any problem exists.

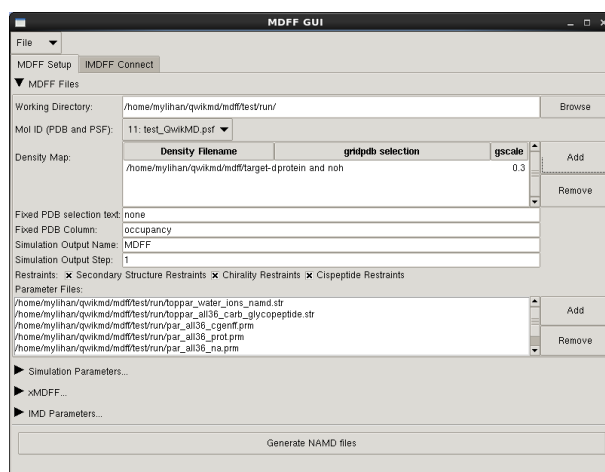


Figure 13: In the MDFF GUI, load the density and prepare MDFF simulations.

- For better comparison, change the Drawing Method to New Cartoon, the Coloring Method to Color ID and the Color to green (7). Fig. 15

8 Closing Remarks

Computer simulations of biomolecular systems have grown rapidly over the past few decades, passing from simulating very small proteins in vacuum to simu-

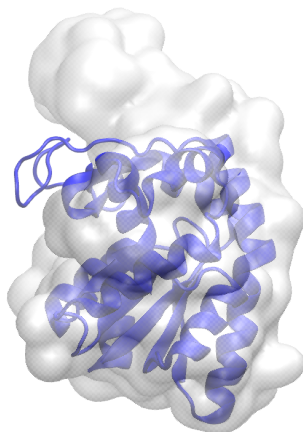


Figure 14: Visualization of the target density (in Isosurface representation) for the fitting.

lating 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,

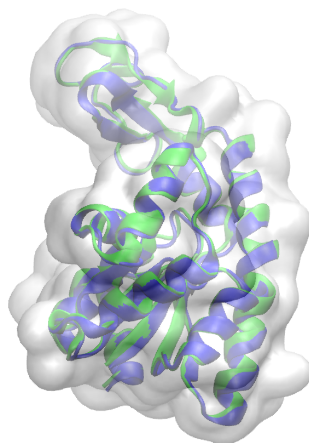


Figure 15: Comparison of the MDFF result (blue) with the experimentally observed model (green).

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/>