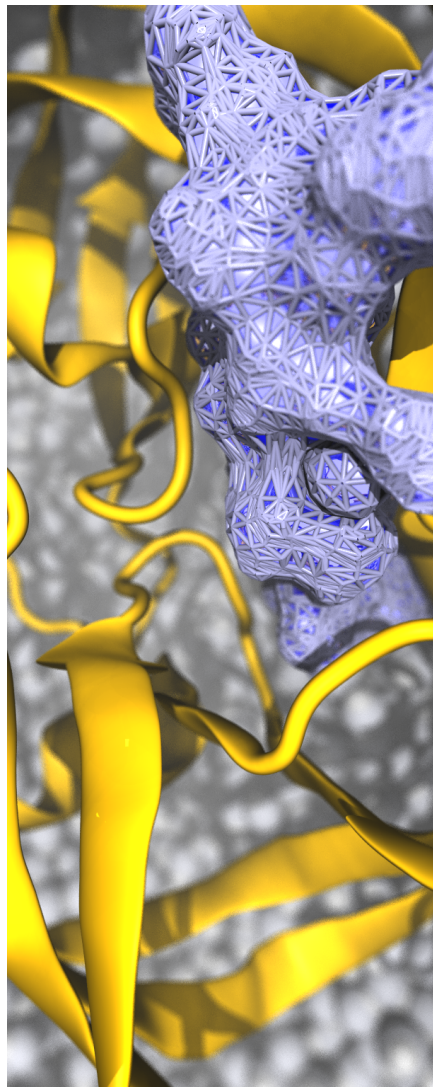
QwikMD: Making Molecular Dynamics Simulations of Biological Systems Easy

Rafael C. Bernardi

rcbernardi@ks.uiuc.edu
www.ks.uiuc.edu/~rcbernardi

Klaus Schulten Group - Theoretical and Computational Biophysics Group
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign
Urbana, IL



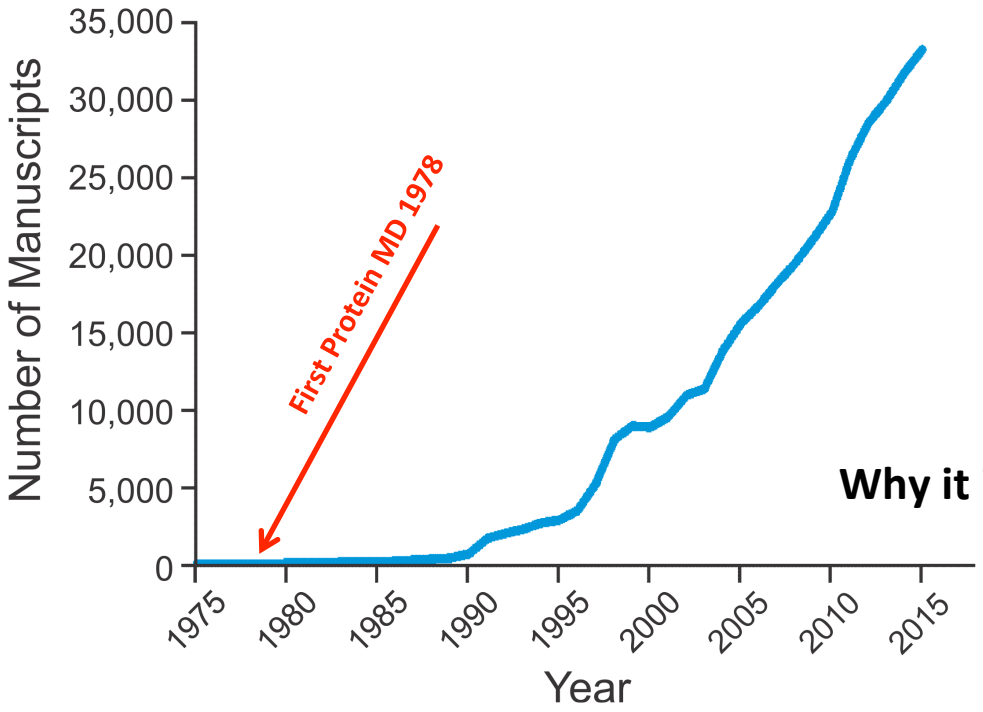
What is Molecular Dynamics?

"Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that ***all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms****."*

Richard Feynman

The Feynman Lectures on Physics: Mainly Mechanism, Radiation and Heat (1963)

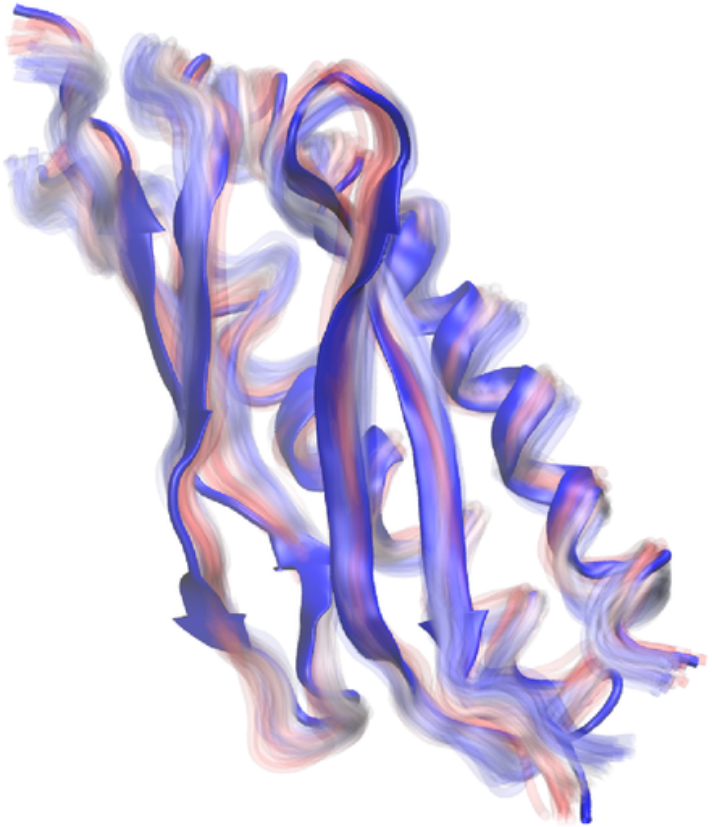
Development of Molecular Dynamics over the past decades:



Why it is so Important?

What is Molecular Dynamics?

Myoglobin Structure
Kendrew (1962 Chemistry Nobel Prize)

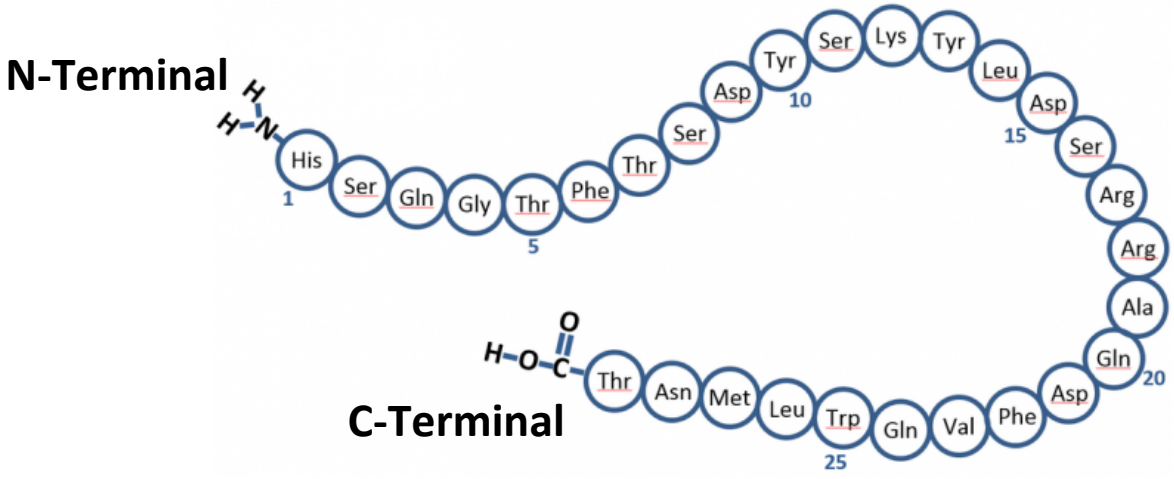


Dynamics plays an important role.

What are Proteins?

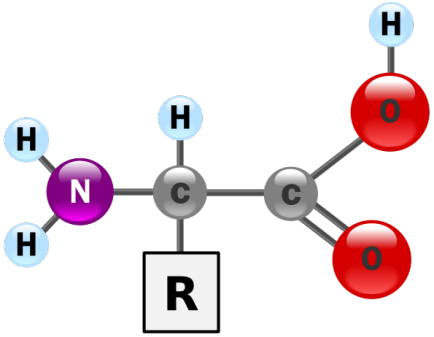
Proteins are large biomolecules, or macromolecules, consisting of one or more long chains of amino acid residues.

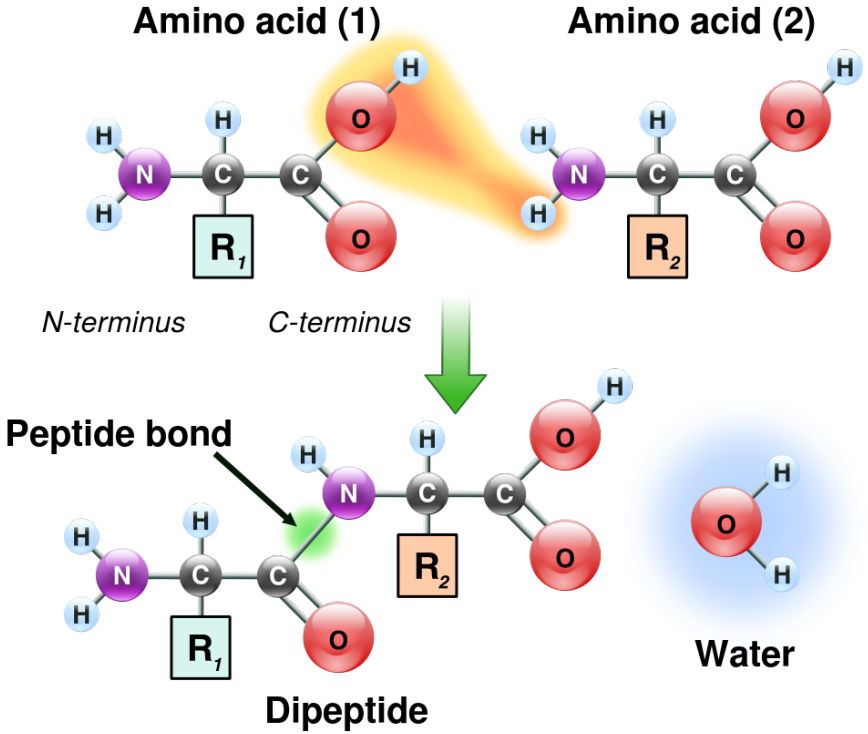
Wikipedia



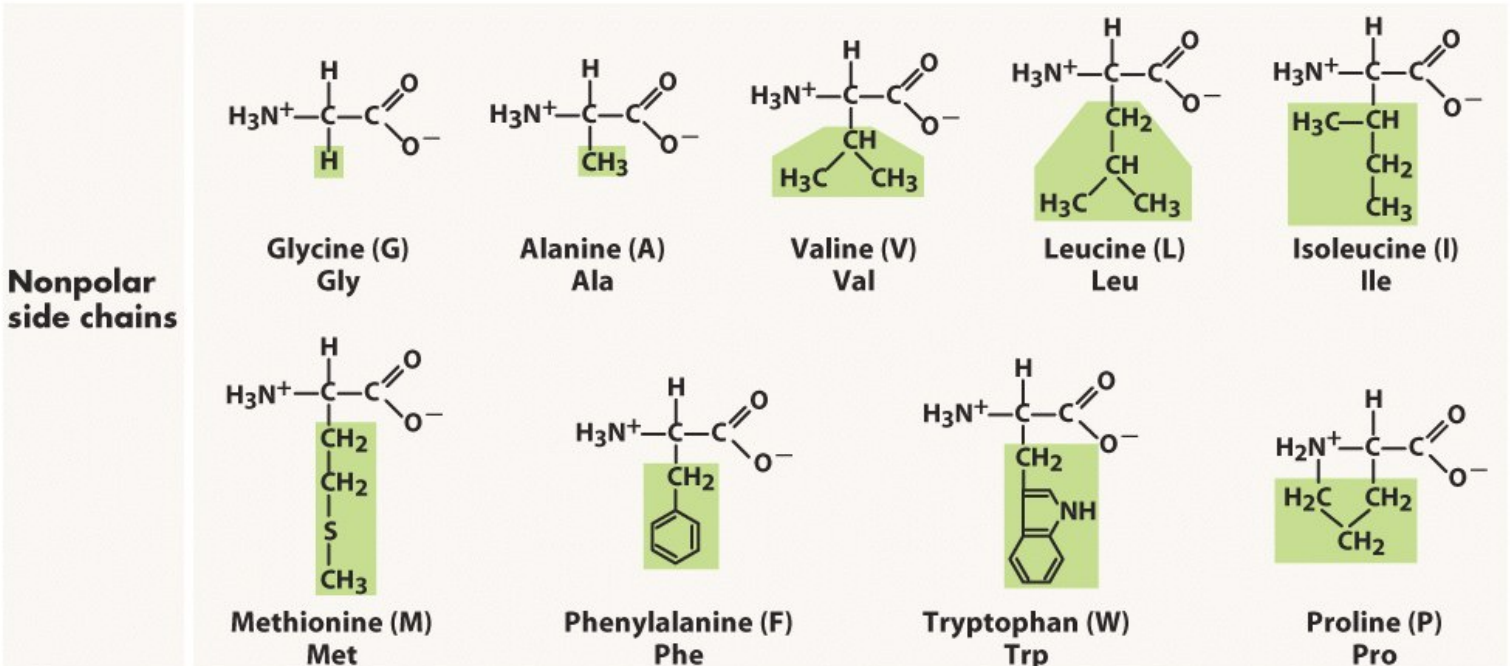
What are Proteins?

Amino Acids:

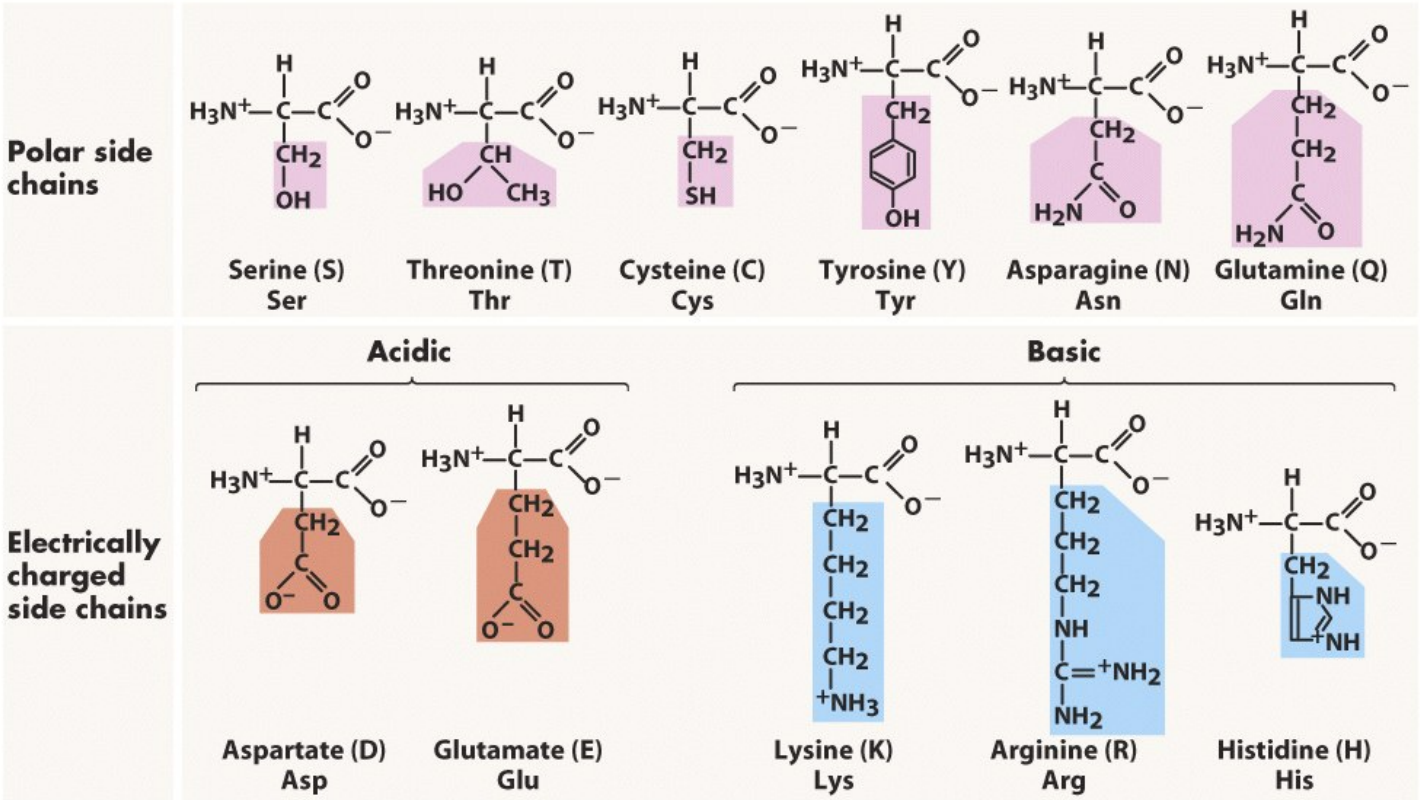




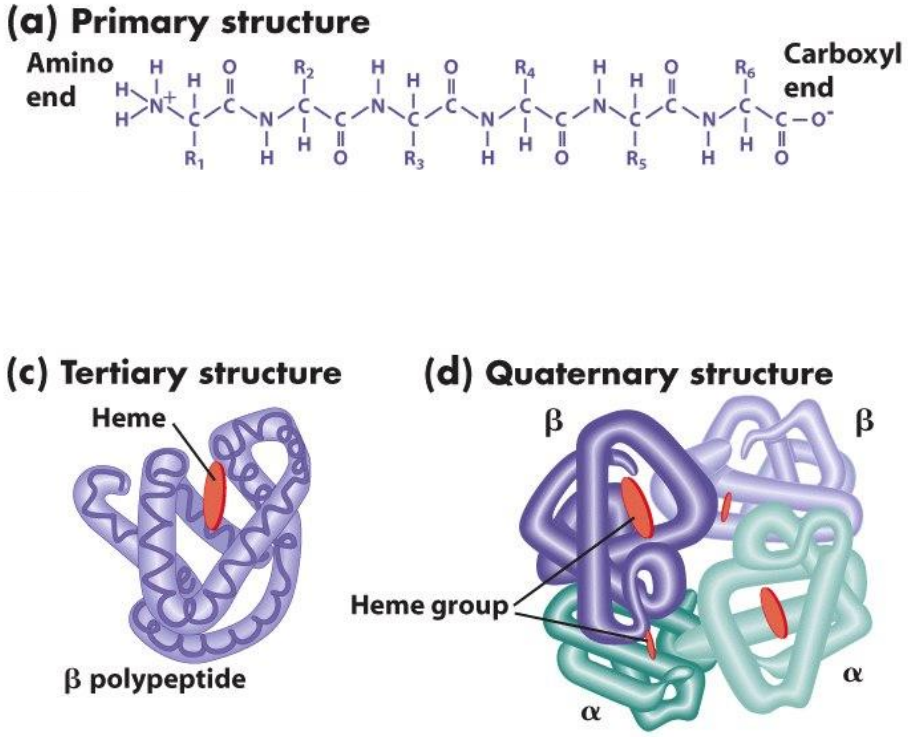
The Different Amino Acids



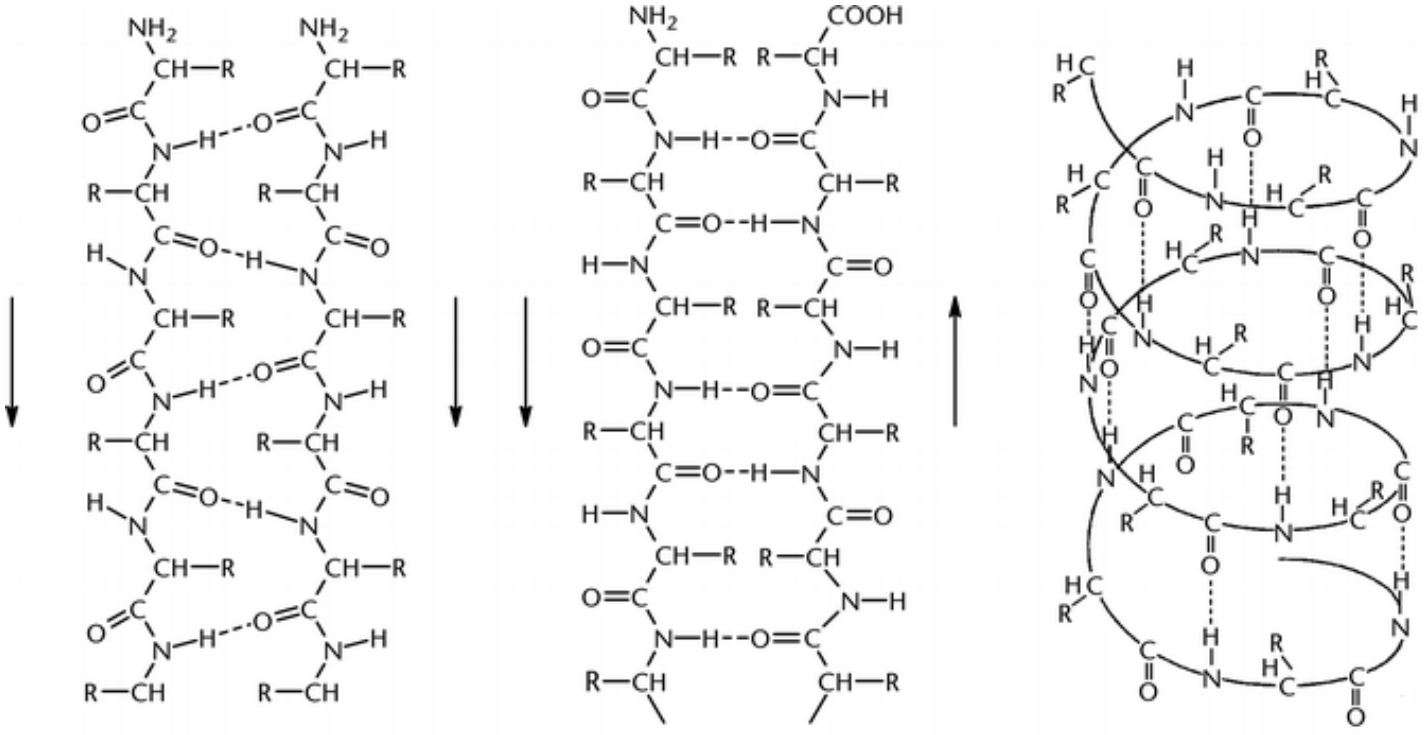
The Different Amino Acids



Protein Structure:



Protein Secondary Structure:



Parallel β pleated sheet

Antiparallel β pleated sheet

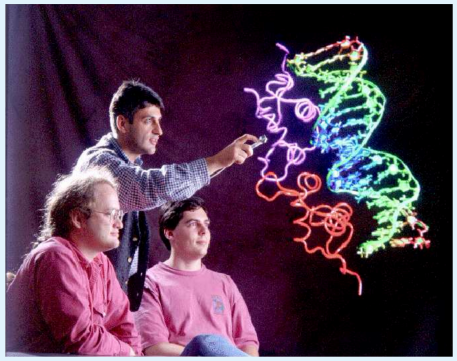
Right-handed α helix

Protein Structure Determination:

- X-Ray Crystallography
- Nuclear Magnetic Resonance (NMR)
- Cryo-Electron Microscopy
- Homology Modeling and *ab initio* Modeling

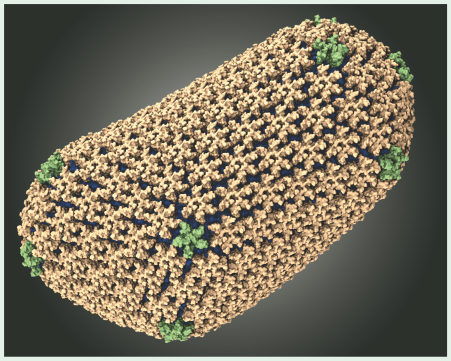
1958

Manual Construction of Structural Models



1996

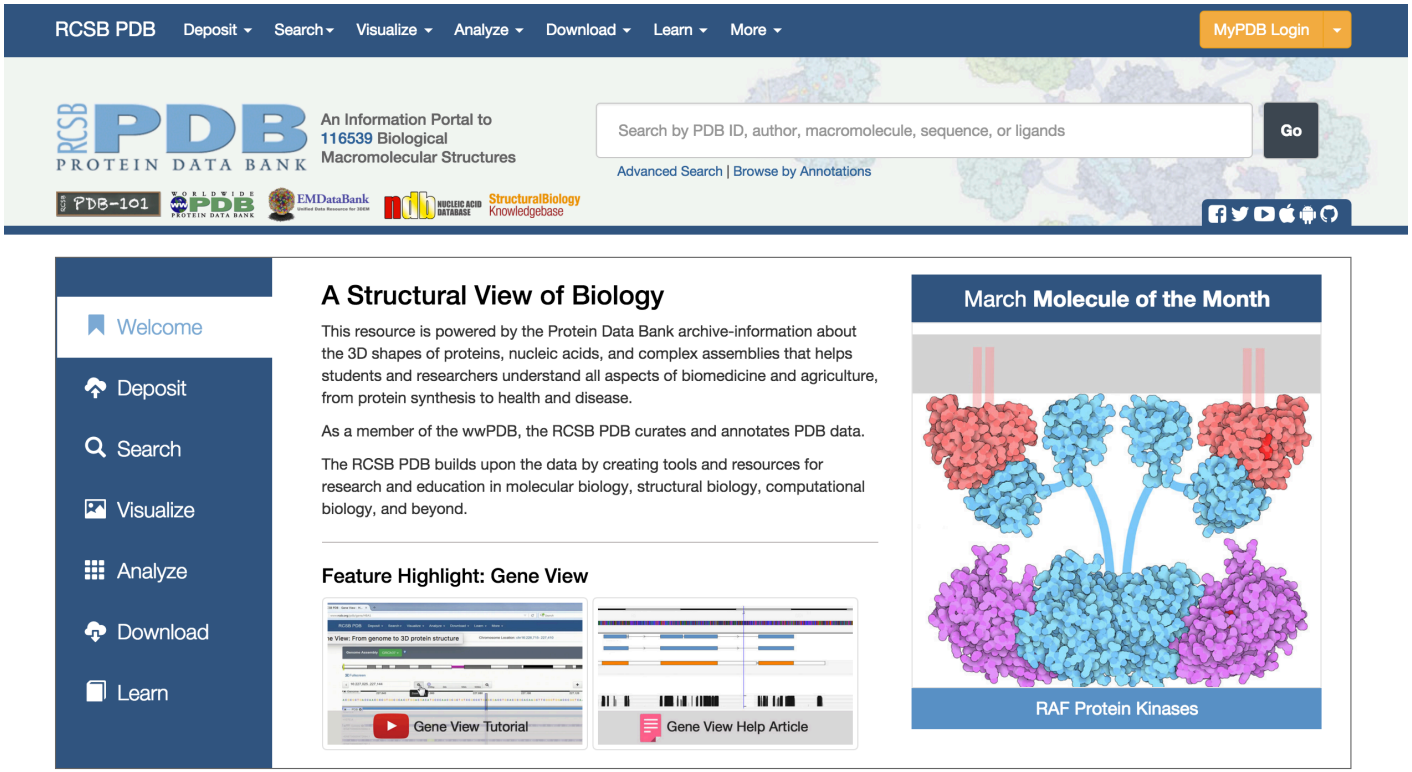
Computer-aided Construction of Structural Models with Manual Positioning of Atoms



2013

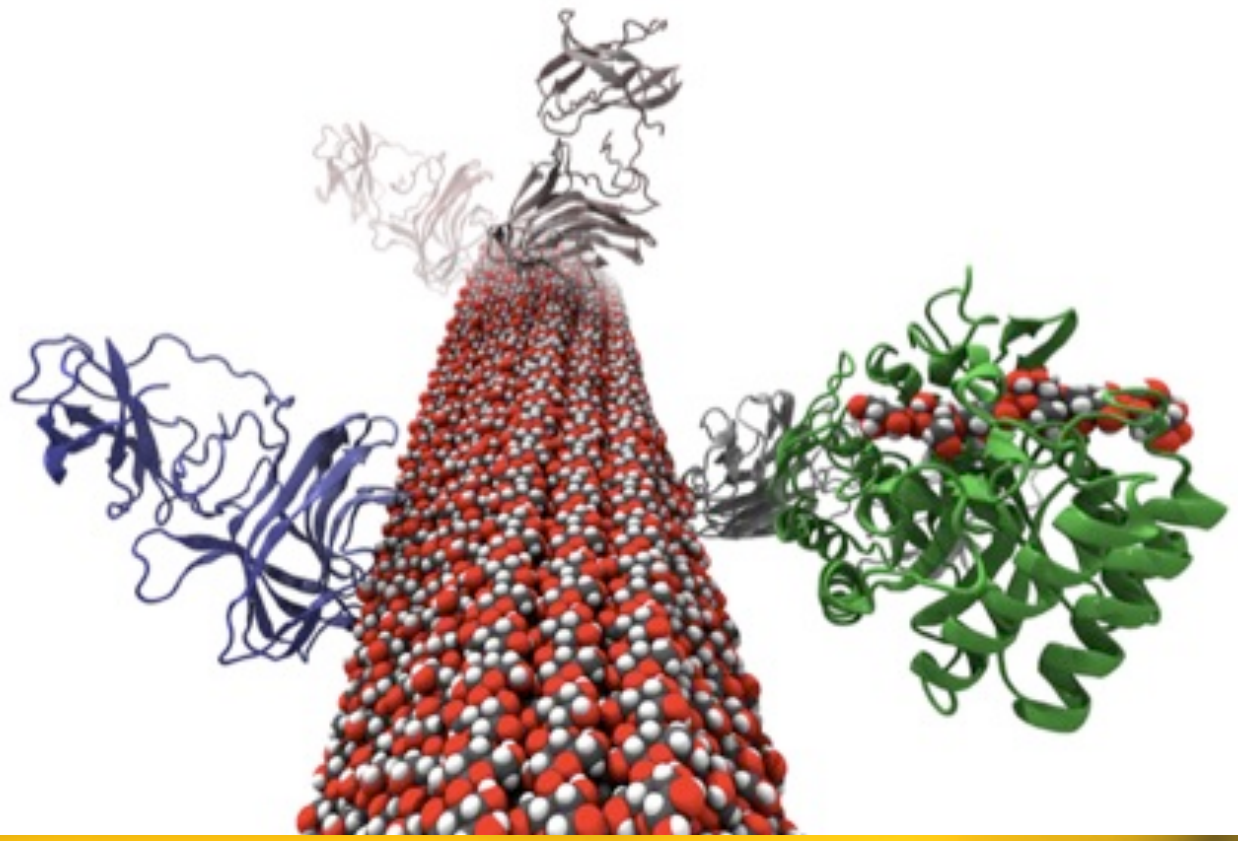
Computational Construction of Structural Models with Automated Positioning of Atoms

What is the Protein Data Bank?



Are there other Biomolecules?

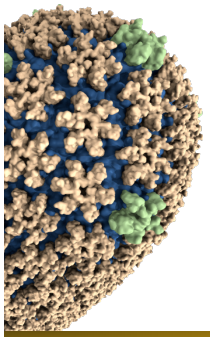
- Nucleic Acids
- Lipids
- Carbohydrates
- Fatty Acids
- Hormones
- Sterols
- …

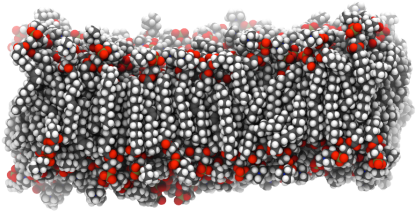


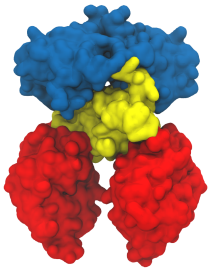
Questions?

The Computational Microscope

What our microscope is made of?

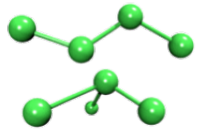






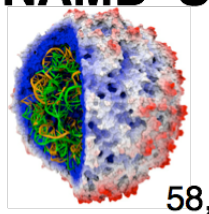
Molecular Dynamics Simulations - Theory

Chemistry

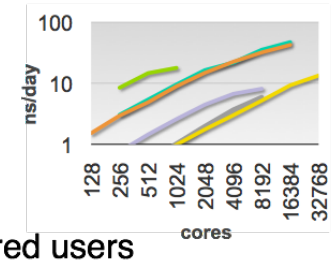


$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} + \underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} + \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

NAMD Software



Virus

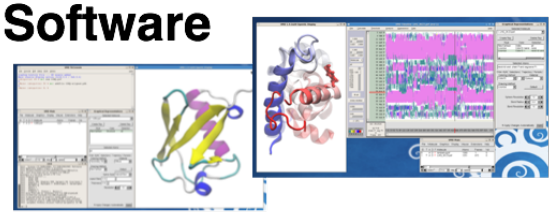


58,000 registered users

Physics

$$m_i \frac{d^2\vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla}U(\vec{R})$$

VMD Software



240,000 registered users

Math

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i}\vec{F}_i(t)$$

(repeat ***one billion times*** = microsecond)

Uses simple Physics concepts of Classical Mechanics.

$$V = V_{str} + V_{bend} + V_{oop} + V_{tors} + V_{vdW} + V_{es}$$

$$\vec{F}_i = -dV / d\vec{r}_i$$

$$\vec{a}_i = \vec{F}_i / m$$

$$\vec{s}_i = \vec{s}_{0i} + \vec{v}_i t + \tfrac{1}{2}\vec{a}_i t^2$$

$$V = V_{str} + V_{bend} + V_{oop} + V_{tors} + V_{vdW} + V_{es}$$

$$V_{es,ij} = Q_i Q_j / \varepsilon_r R_{ij}$$

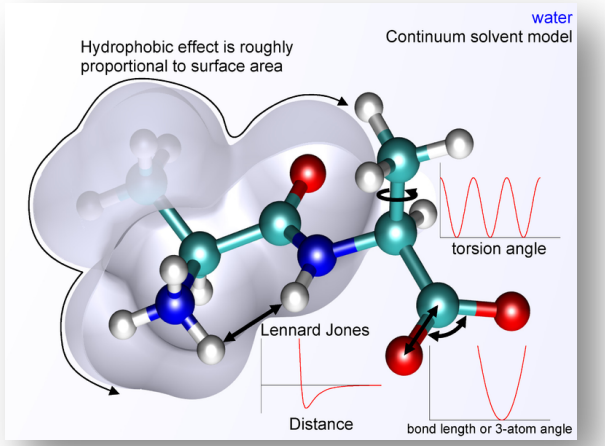
$$V_{vdW,ij} = \varepsilon_{IJ}\left[\left(\frac{R_{IJ}}{R_{ij}}\right)^{12} - 2\left(\frac{R_{IJ}}{R_{ij}}\right)^{6}\right]$$

$$V_{tors,ijkl} = \frac{1}{2}\left[V_1(1+\cos\varphi) + V_2(1-\cos 2\varphi) + V_3(1+\cos 3\varphi) + \ldots\right]$$

$$V_{oop} = \frac{1}{2} k_{oop} \omega_{oop}^2$$

$$V_{bend,ijk} = \frac{1}{2} k_{IJK}\left(\theta_{ij} - \theta^0{}_{IJK}\right)^2$$

$$V_{str,ij} = \frac{1}{2} k_{IJ}\left(l_{ij} - l^0{}_{IJ}\right)^2$$



$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{\text{total}}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \quad \alpha = 1, 2 \ldots N,$$

where m_α is the mass of atom α, $\vec{r}_\alpha$ is its position, and U_{total} is the total potential energy that depends on all atomic positions and, thereby, couples the motion of atoms.

$$M\dot{v} = F(r) - \gamma v + \sqrt{\frac{2\gamma k_B T}{M}} R(t),$$

where M is the mass, $v = \dot{r}$ is the velocity, F is the force, r is the position, γ is the friction coefficient, k_B is the Boltzmann constant, T is the temperature, and $R(t)$ is a univariate Gaussian random process. Coupling to the reservoir is modeled by adding the fluctuating (the last term) and dissipative ($-\gamma v$ term) forces to the Newtonian equations of motion. To integrate the Langevin equation, NAMD uses the Brünger–Brooks–Karplus (BBK) method, a natural extension of the Verlet method for the Langevin equation. The position recurrence relation of the BBK method is

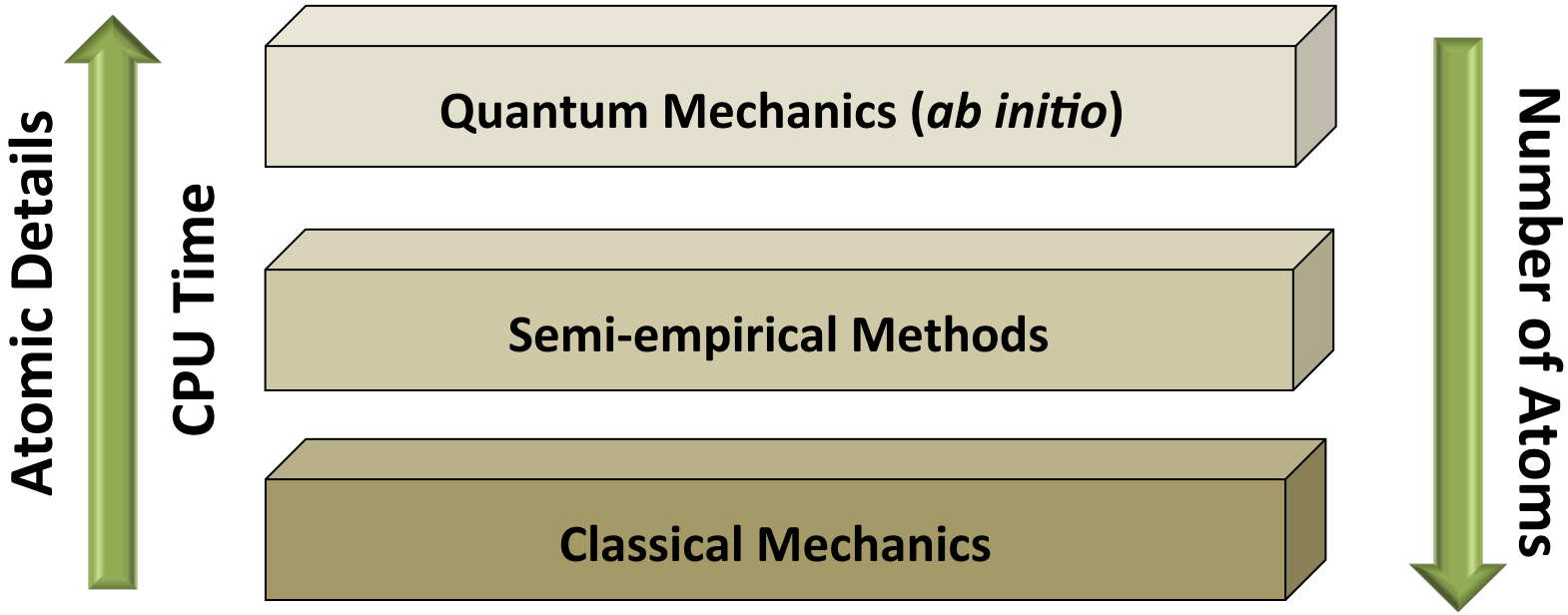
$$r_{n+1} = r_n + \frac{1 - \gamma\Delta t/2}{1 + \gamma\Delta t/2}(r_n - r_{n-1}) + \frac{1}{1 + \gamma\Delta t/2}\Delta t^2\left[M^{-1}F(r_n) + \sqrt{\frac{2\gamma k_B T}{\Delta M}} Z_n\right]$$

where Z_n is a set of Gaussian random variables of zero mean and variance 1. The BBK integrator requires only one random number for each degree of freedom. The steady-state distribution generated by the BBK method has an error proportional to Δt^2, although the error in the time correlation function can have an error proportional to Δt.

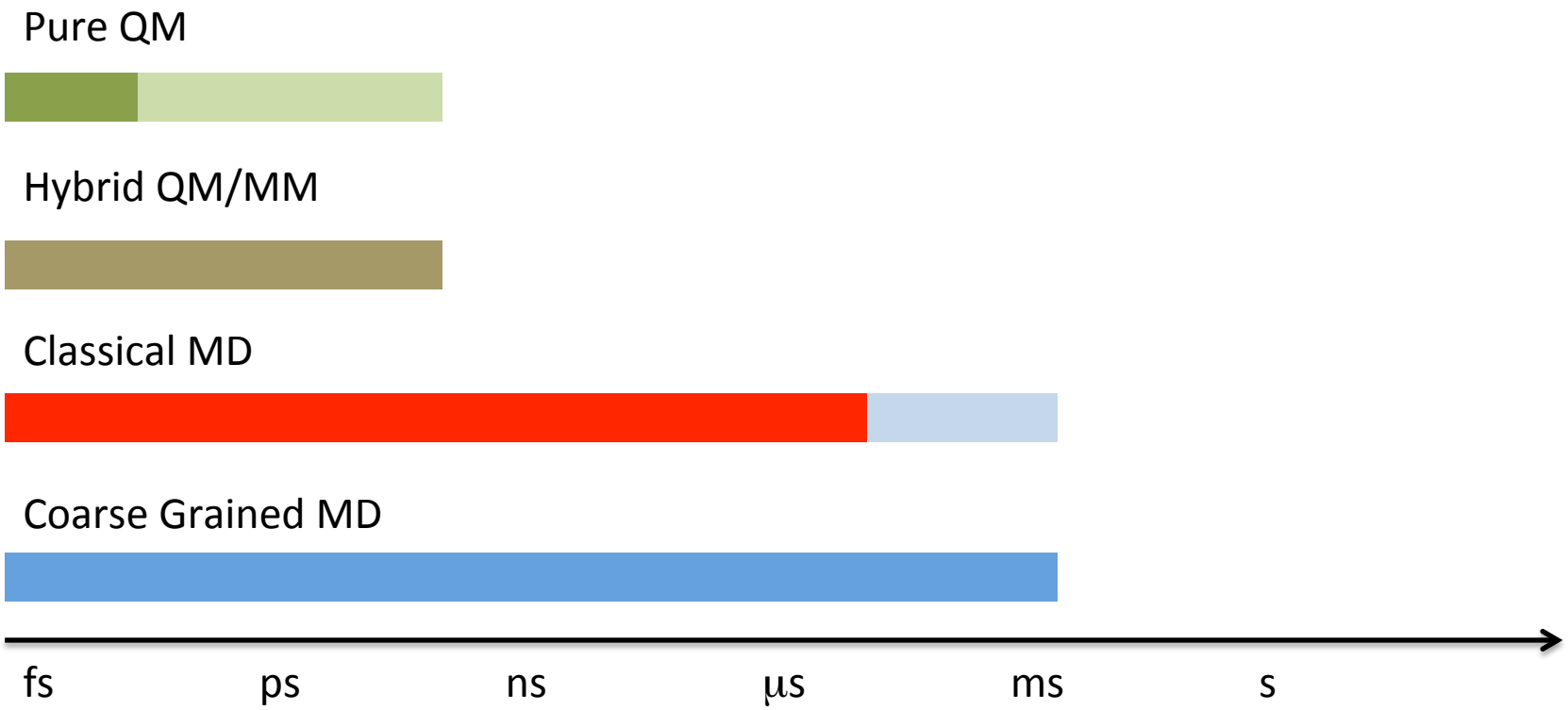
Why not to use Quantum Mechanics?

$$\frac{-\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2}+U(x)\Psi(x,t)=i\hbar\frac{\partial\Psi(x,t)}{\partial t}$$

Classical Mechanics vs. Quantum Mechanics

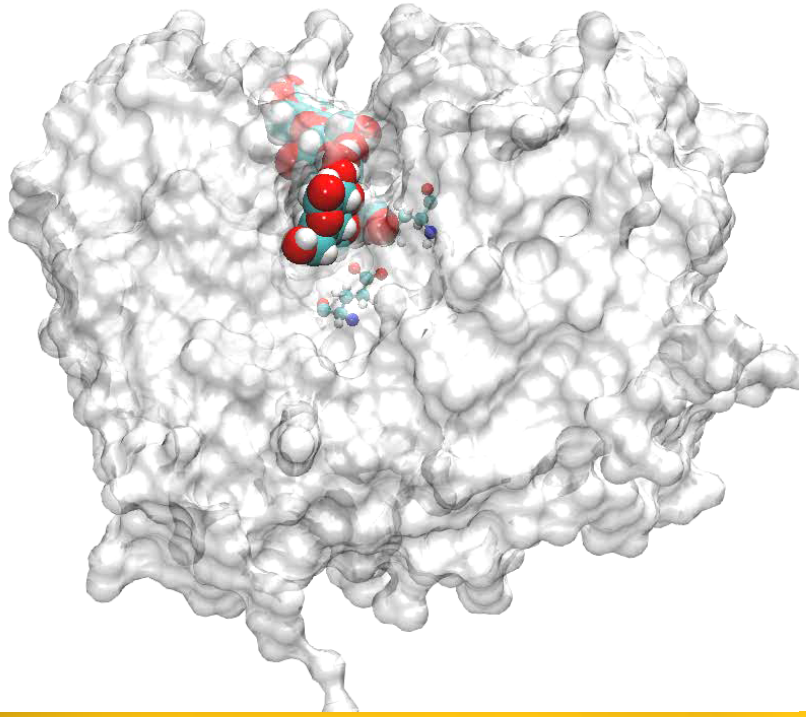


The Time Scale Problem



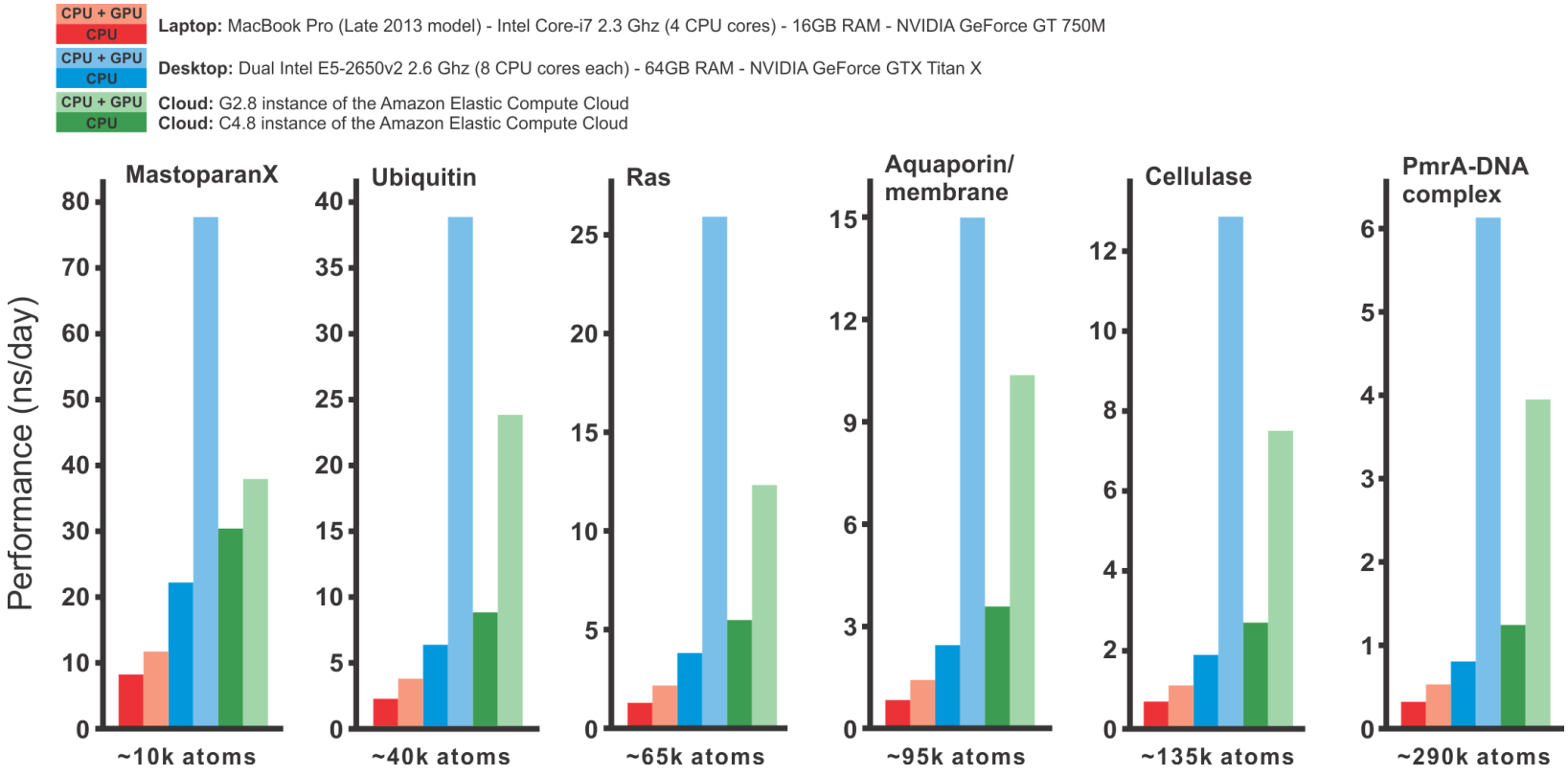
NAMD QM/MM will be released in late 2016

The Time Scale Problem



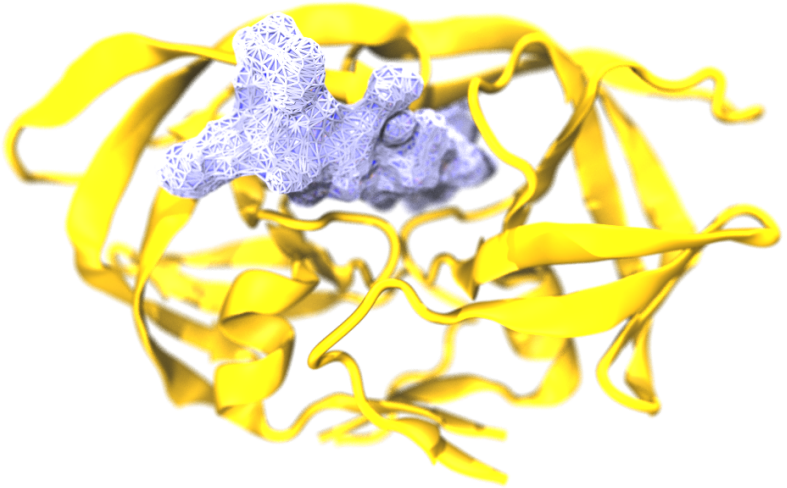
What timescale we can simulate?

Performance of NAMD simulations on different computer platforms



QwikMD

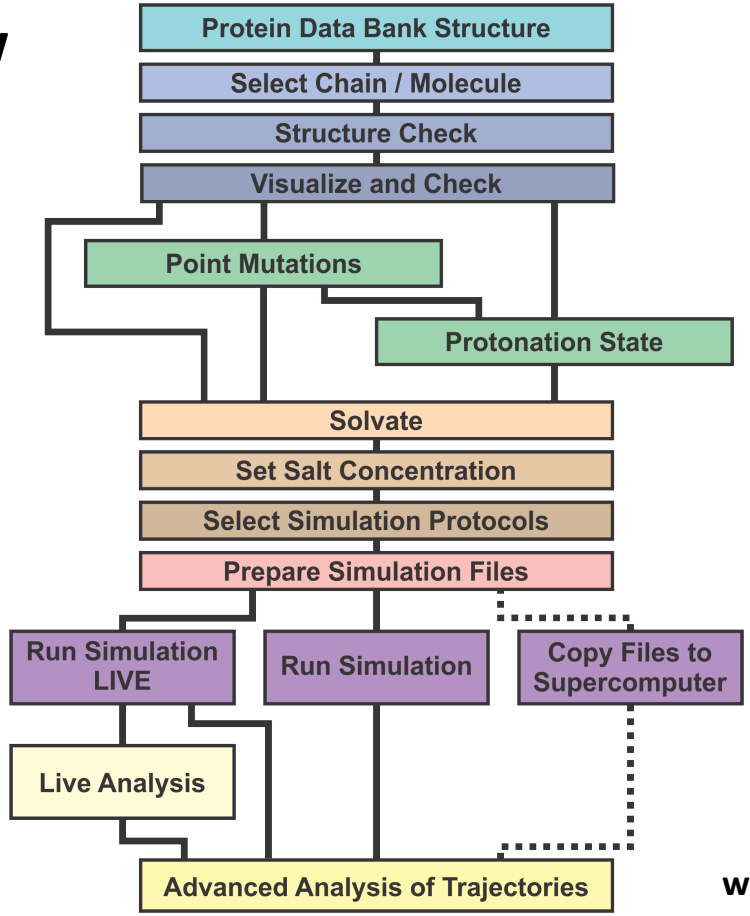
Employing QwikMD, a user is able to prepare an MD simulation in just a few minutes, allowing studies of point mutations, partial deletions and even atomic force microscopy experiments.



QwikMD assists a new user in performing 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.

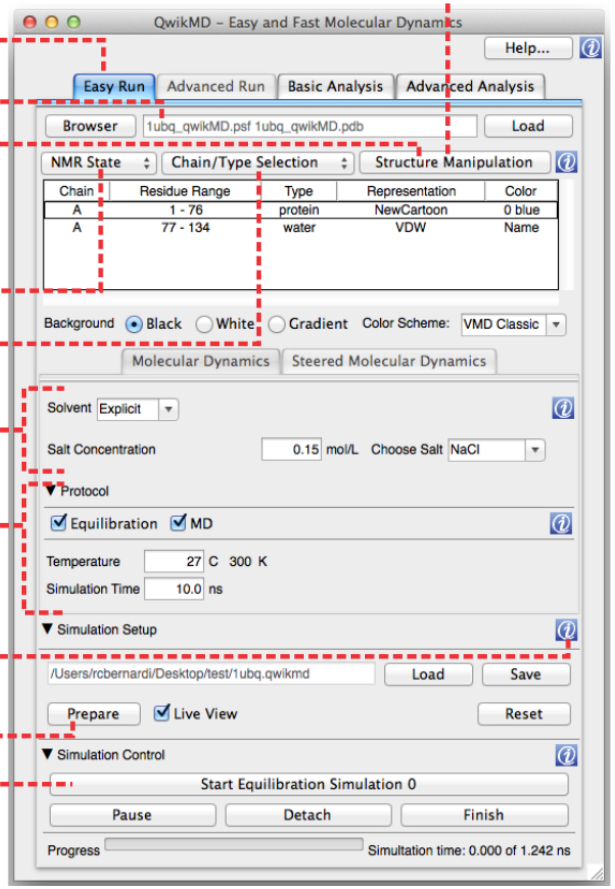
www.ks.uiuc.edu/Research/qwikmd

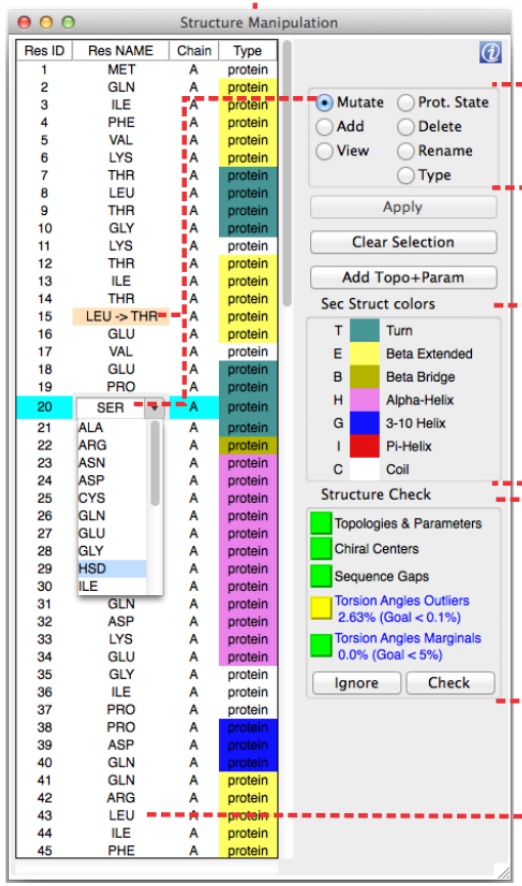
QwikMD Workflow



www.ks.uiuc.edu/Research/qwikmd

QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts





Select: Easy Run or Advanced Run Analysis Tools can also be selected

Browse for a PDB file or type PDB code

In the structure manipulation window the QwikMD user can delete molecules or parts of a protein sequence; perform point mutations; change protonation states; check the structure; insert membrane model (Advanced mode only)

For structures solved by NMR, select the conformational state

Select the chains and type of molecule to be included in the simulation

Select the environment: implicit or explicit solvent, and salt concentration

In the Protocol section the user selects temperature and duration of the simulation and also if an equilibration will be performed

"Info Buttons" provide the user with a variety of information, from protocols to a guidance on how to perform simulations and how to check results

When "Prepare" is clicked, QwikMD will invoke several scripts to perform all the operations set before, such as mutations, solvation, ...

Starts the simulation, either in "Live View" mode (if selected) or in "Background"

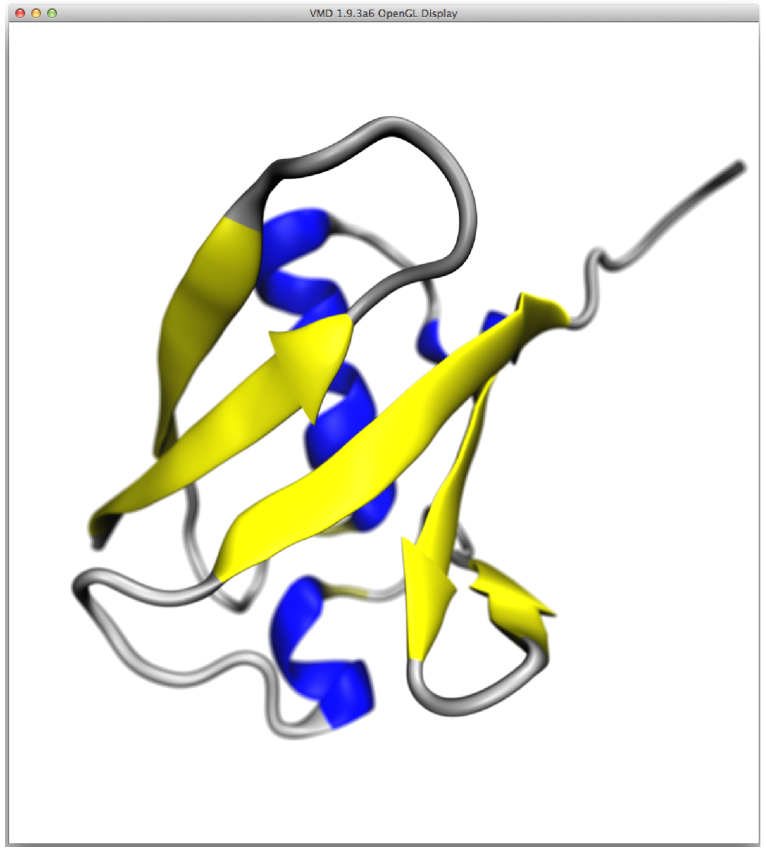
QwikMD user can easily perform point mutations. To do so, select "Mutate" and click on the amino acid from the sequence amino acid table. A list of all possible mutations will allow the user to select the desired mutation.

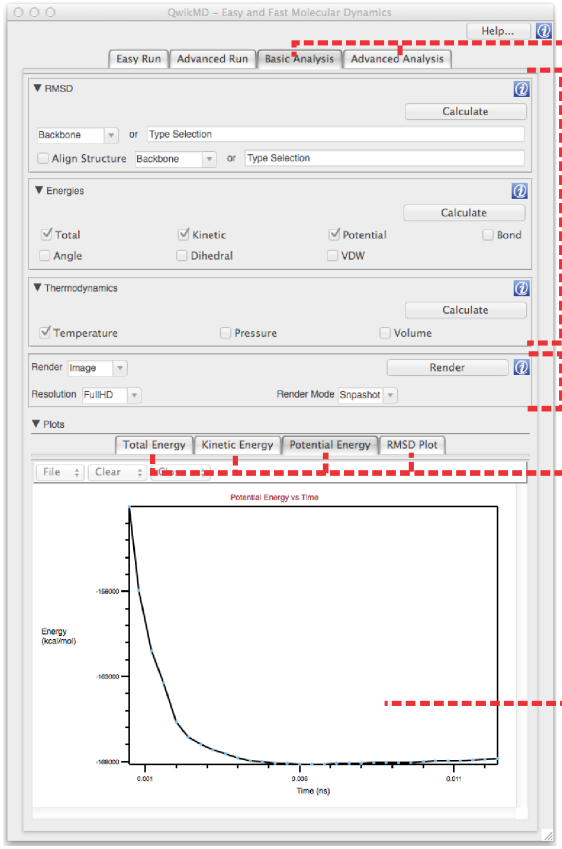
The list of amino acid in the protein sequence is colored according to the secondary structure.

An automated check of the structure is performed when a PDB is loaded by QwikMD. If problems arise they will be marked in the "Structure Check" Tab. The user will be guided on how to fix the problems found.

All molecules of the system are presented in a list an can be separated by *type* or *chain*. The QwikMD user can easily delete parts of structure, change protonation states, perform point mutations, among other actions.

QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts





QwikMD users can select between: (1) "Basic Analysis", which include most common analysis methods used to check how stable is the structure in the simulation; or (2) "Advanced Analysis", which includes several of the most used analysis tools in VMD, i.e., Hydrogen Bond count, and Solvent Accessible Surface Area (SASA).

Here the user can select the analysis to be performed when "Calculate" is clicked.

VMD is known for its structure image rendering capabilities. In QwikMD a quick-render tab allows for a fast high-quality rendering, employing the most used settings for shadows, colors, materials, ...

Multiple analysis can be performed at the same time. The resulting plots will be presented in different tabs.

In a simulation on *live view* mode the plot will be updated while the simulation is performed

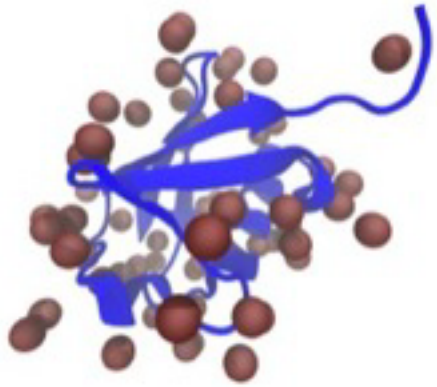
Molecular Dynamics can be used to calculate a diverse set of properties:

- Free-energy (transition between two structural states)
- Mechanical Properties
- Viscosity
- Thermodynamics Properties
- Effects of structural changes in the above properties
- …

QwikMD Hands-on

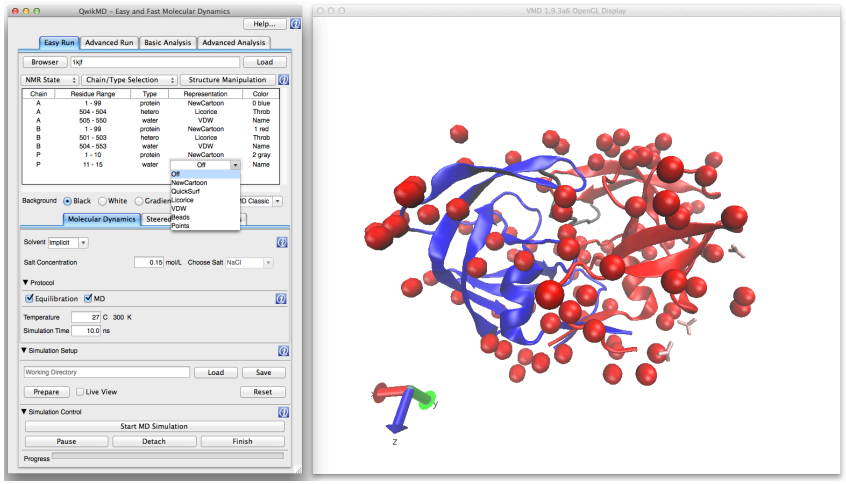
Case 1: Ubiquitin

Ubiquitin is a small protein that is found in almost all cellular tissues in humans and other eukaryotic organisms, which helps to regulate the processes of other proteins in the body.



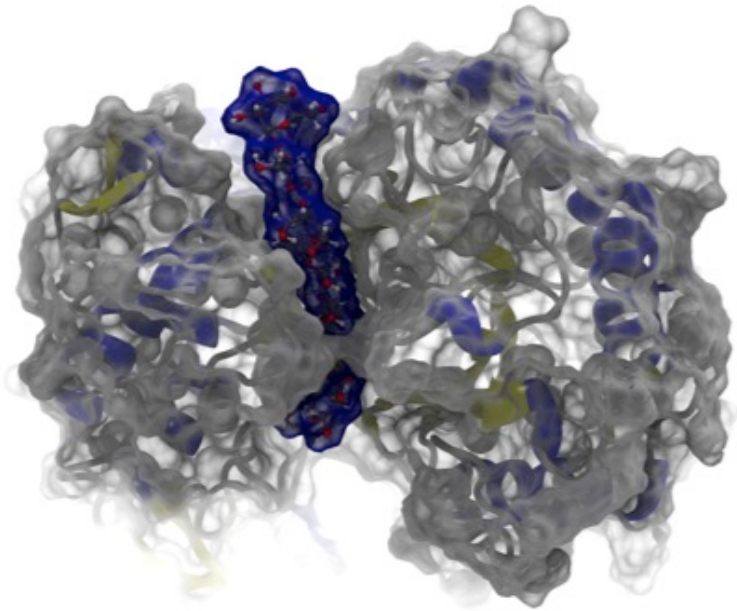
Case 2: HIV-Protease

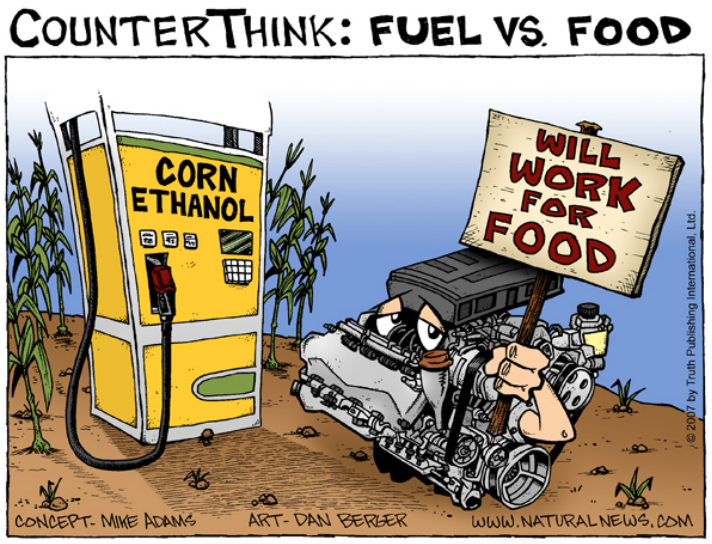
HIV-1 protease is a retroviral aspartyl protease (retropepsin) that is essential for the life-cycle of HIV, the retrovirus that causes AIDS. HIV protease cleaves newly synthesized polyproteins at the appropriate places to create the mature protein components of an infectious HIV virion. Without effective HIV protease, HIV virions remain uninfectious.



Case 3: Cellulase

Cellulase refers to a group of enzymes which, acting together, hydrolyze cellulose. Cellulose is a linear polysaccharide of glucose residues connected by β-1,4 linkages

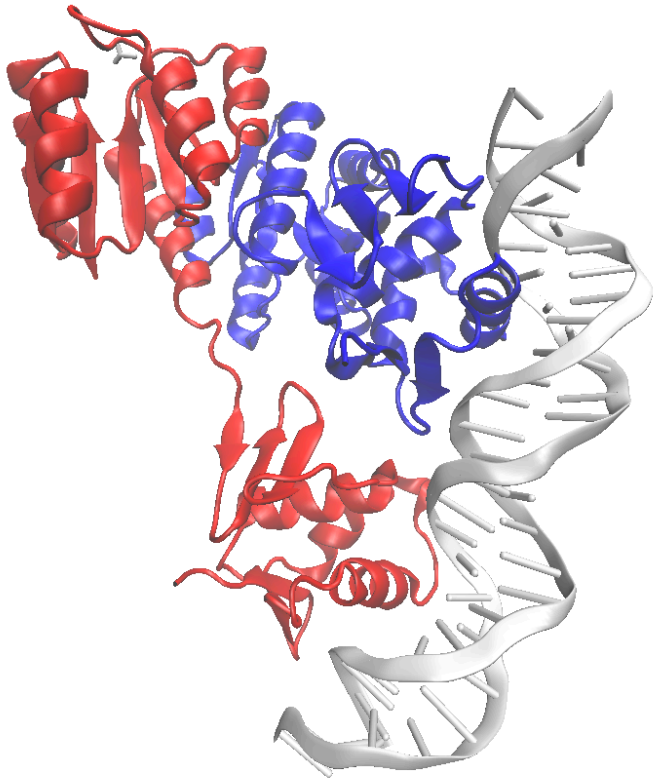




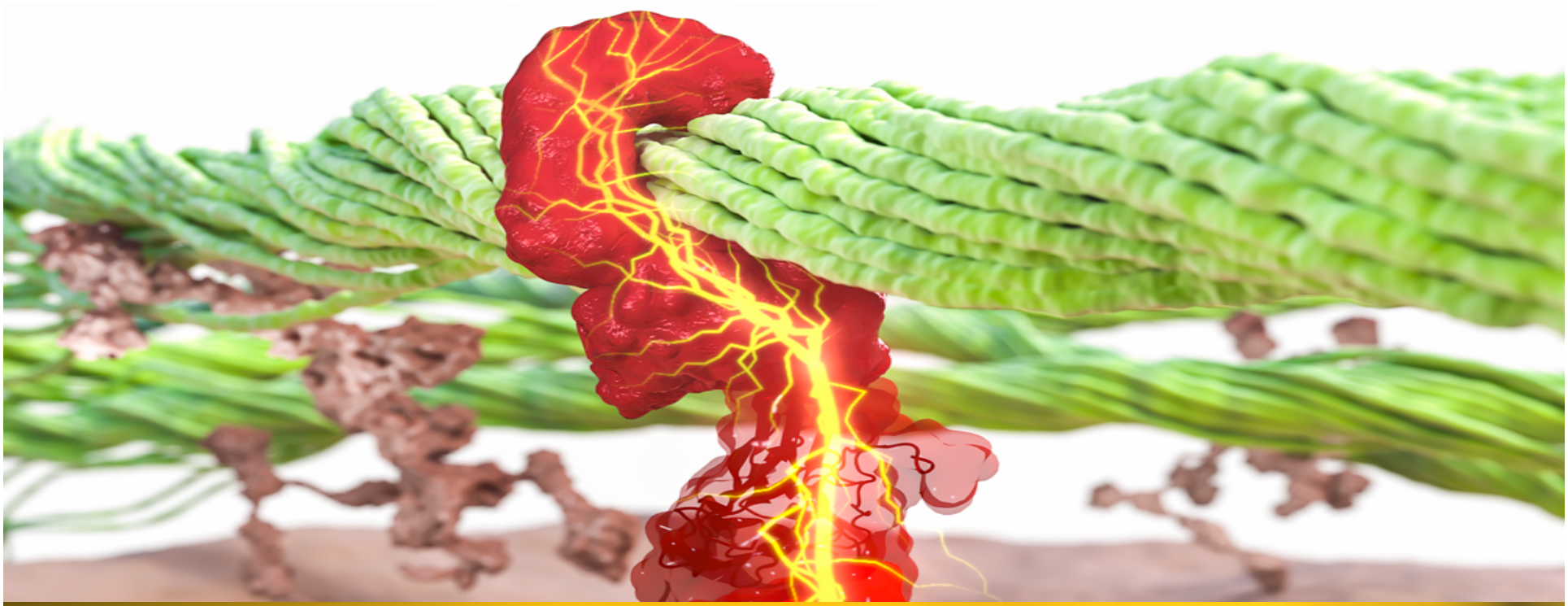
Case 4: DNA/Protein Complex

Deoxyribonucleic acid is a molecule that carries most of the genetic instructions used in the development, functioning and reproduction of all known living organisms and many viruses.

Protein–DNA interactions often regulate the biological function of DNA, usually the expression of a gene. Among the proteins that bind to DNA are transcription factors that activate or repress gene expression by binding to DNA motifs and histones that form part of the structure of DNA and bind to it less specifically.



Case 5: Cellulosomes - Pulling Experiment



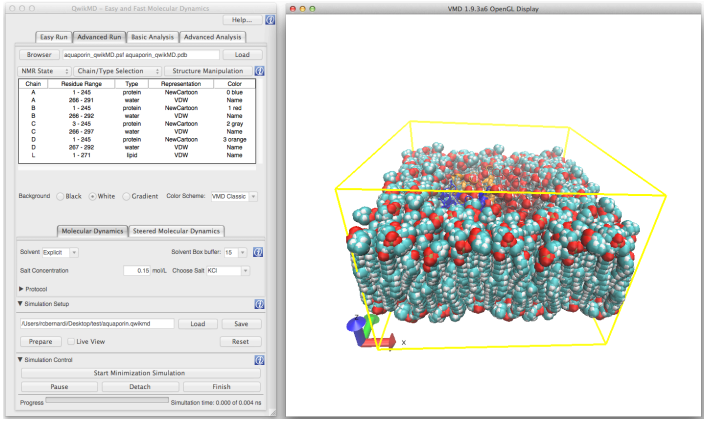
Ultrastable Protein complexes

Challenging environments have guided nature in the development of ultrastable protein complexes. Specialized bacteria produce discrete multi-component protein networks called cellulosomes to effectively digest lignocellulosic biomass.

Certain cellulosomal ligand–receptor interactions exhibit extreme resistance to applied force.

Case 6: Membrane Protein

Membrane proteins are proteins that interact with, or are part of, biological membranes. They are one of the common types of protein along with soluble globular proteins, fibrous proteins, and disordered proteins. They are targets of over 50% of all modern medicinal drugs. It is estimated that 20–30% of all genes in most genomes encode membrane proteins.



Newer Advanced Version of QwikMD required
Download QwikMD from: www.ks.uiuc.edu/~rcbernardi


Thank You