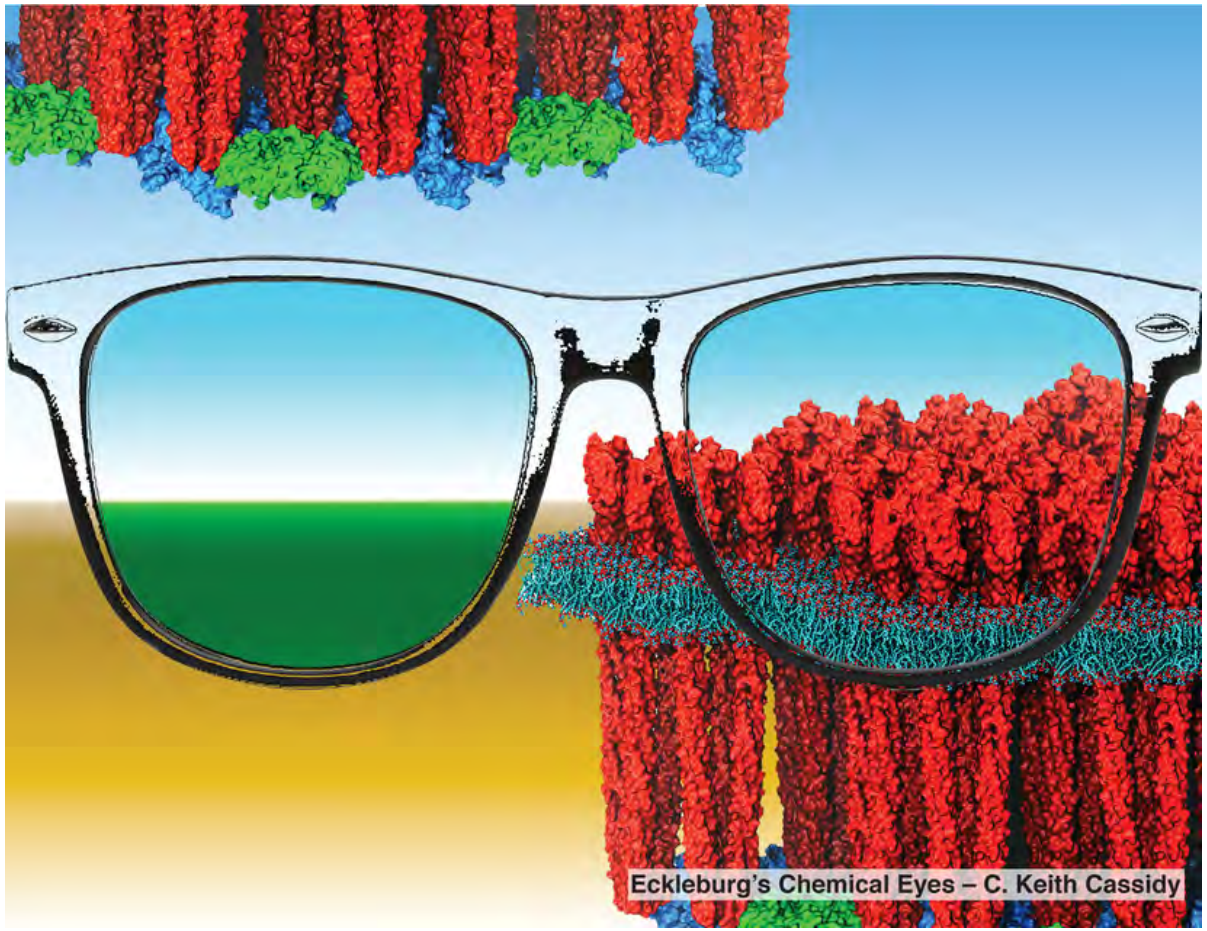
Hands-on Workshop on Computational Biophysics

May 21 -25, 2018

Pittsburgh Supercomputing Center



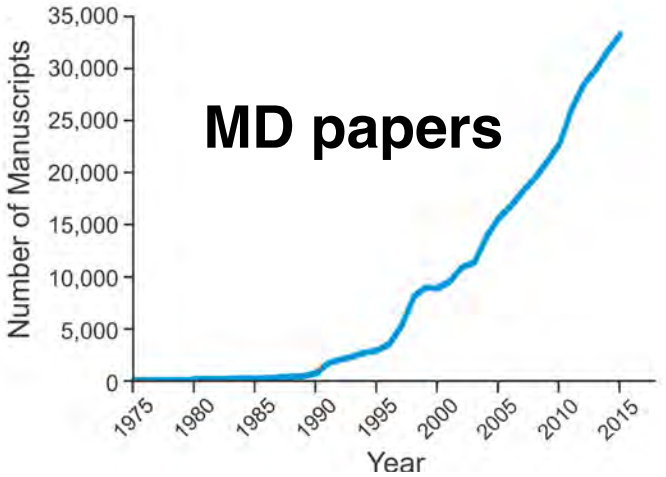
Emad Tajkhorshid

NIH Center for Macromolecular Modeling and Bioinformatics
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

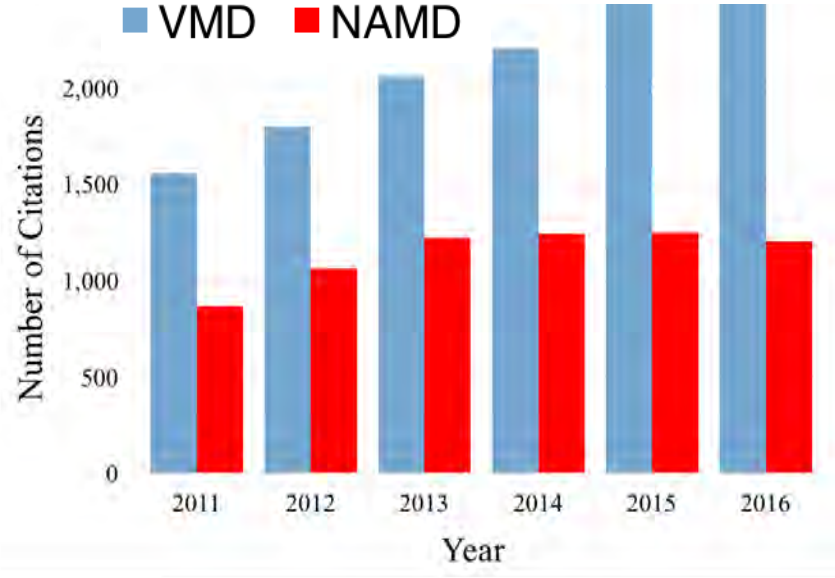
NIH P41 Center for Macromolecular Modeling and Bioinformatics

University of Illinois at Urbana-Champaign

Serving the large and fast growing community of biomedical researchers employing molecular modeling and simulation technologies



103,000 VMD users
19,000 NAMD users
17,000 NIH funded
1.4 million web visitors
228,000 tutorial views

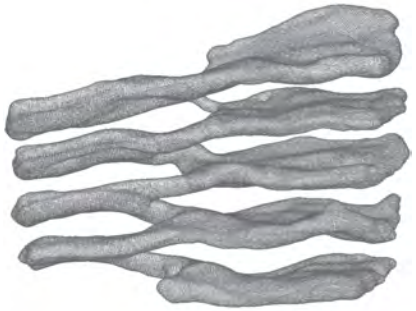


Serving a Large and Fast Growing Community

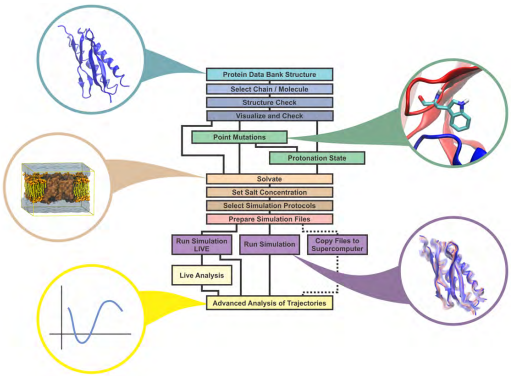
- Deploying Center's flagship programs NAMD and VMD on all major computational platforms from commodity computers to supercomputers

- Consistently adding user-requested features
    - simulation, visualization, and analysis

- Covering broad range of scales (orbitals to cells) and data types



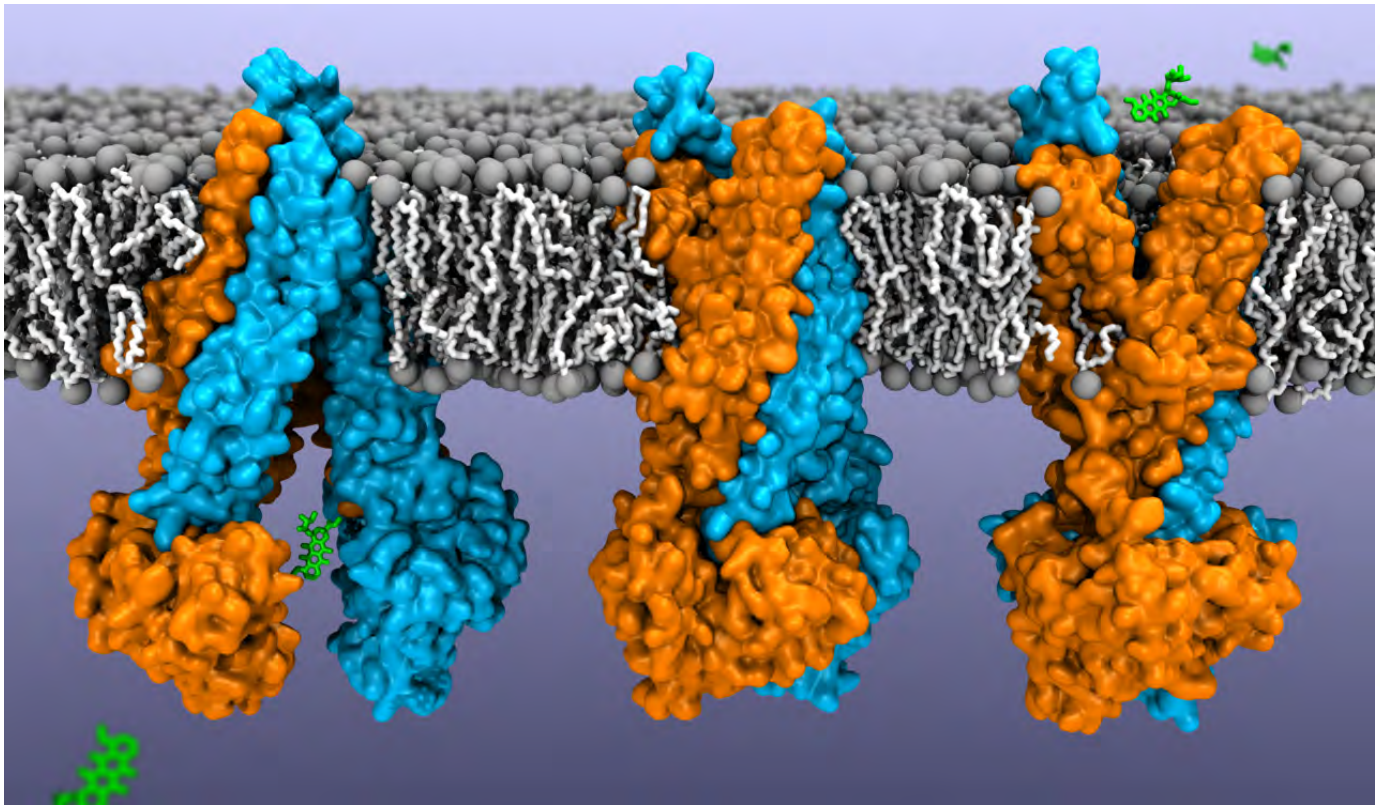
- Enhanced software accessibility
    - QwikMD, interactive MDFF, ffTk, simulation in the Cloud, remote visualization

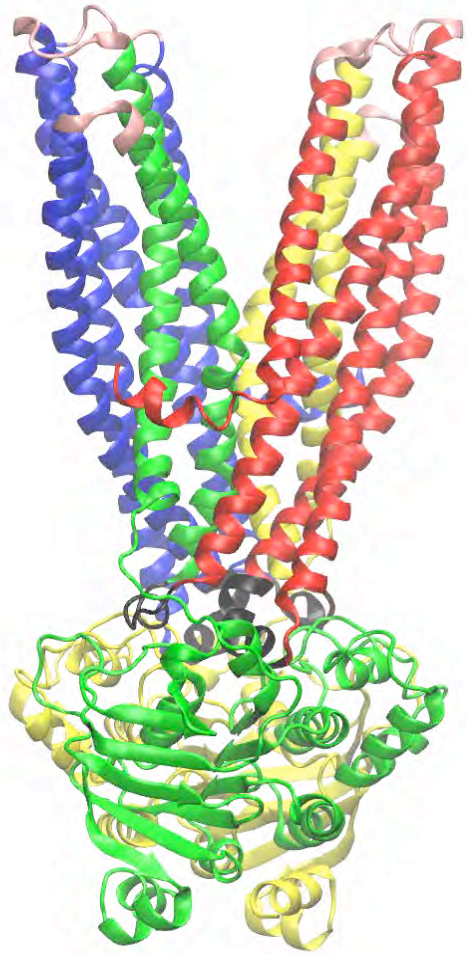


Exploiting State of the Art Hardware Technology

- Software available and optimized on all national supercomputing platforms (even before they come online)
- Decade-long, highly productive relationship with NVIDIA
- The first CUDA Center of Excellence funded by NVIDIA
- Consistently exploring opportunities for new hardware technology
  - Remote visualization
  - Virtual Reality
  - Handheld devices

Computational Structural Biology Describing Biomolecules at Nanoscale

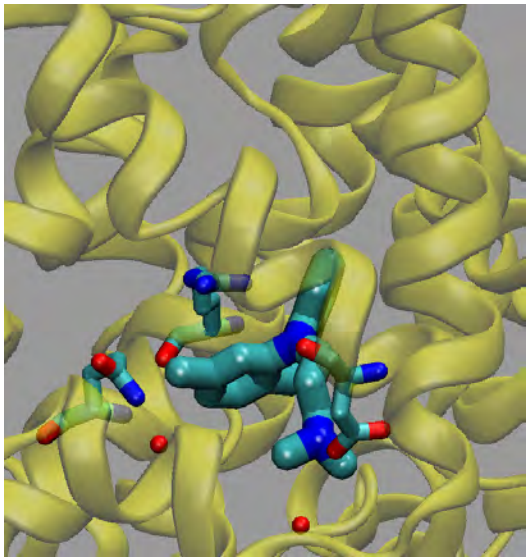


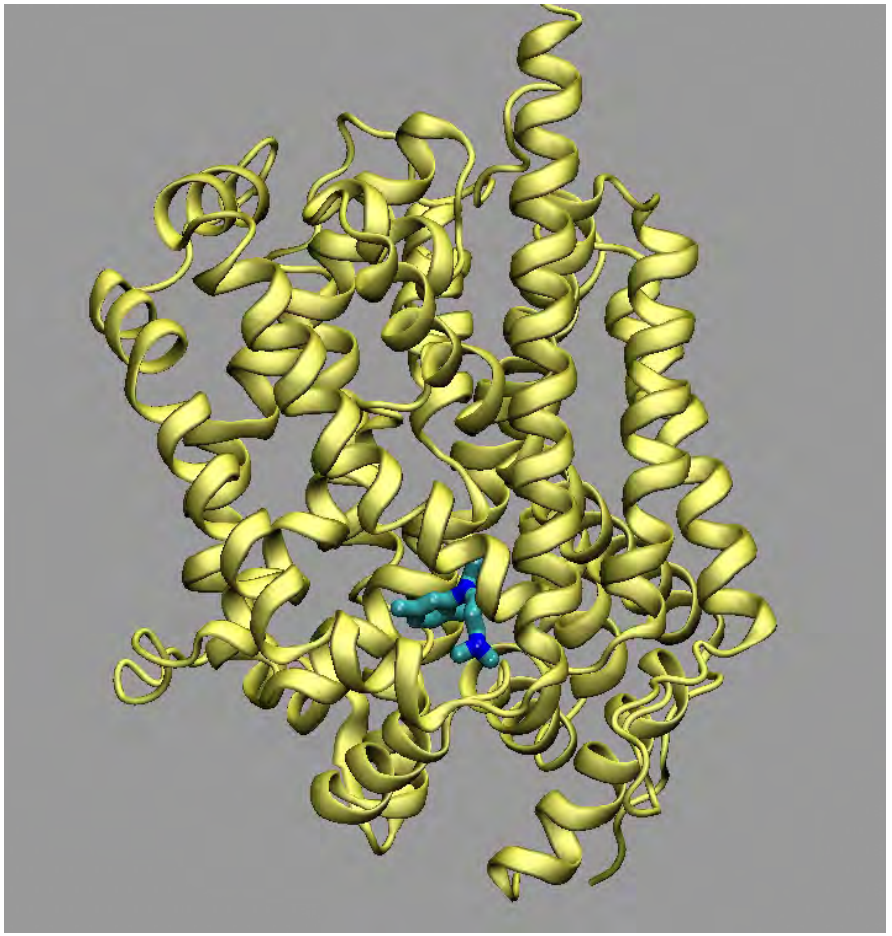


Structure / Dynamics
@ nanoscale

Why Structural Biology at Nanoscale?

- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology

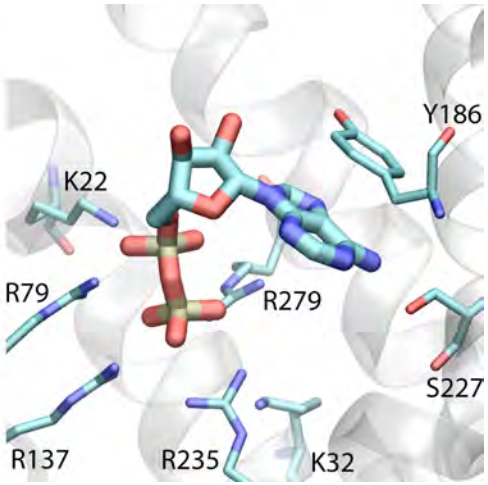


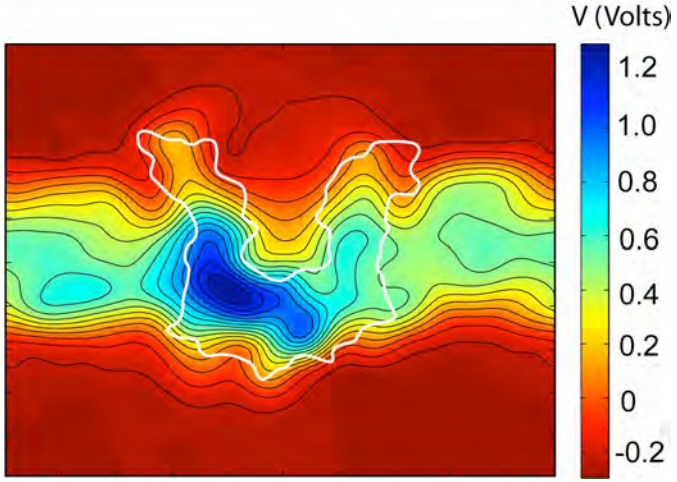


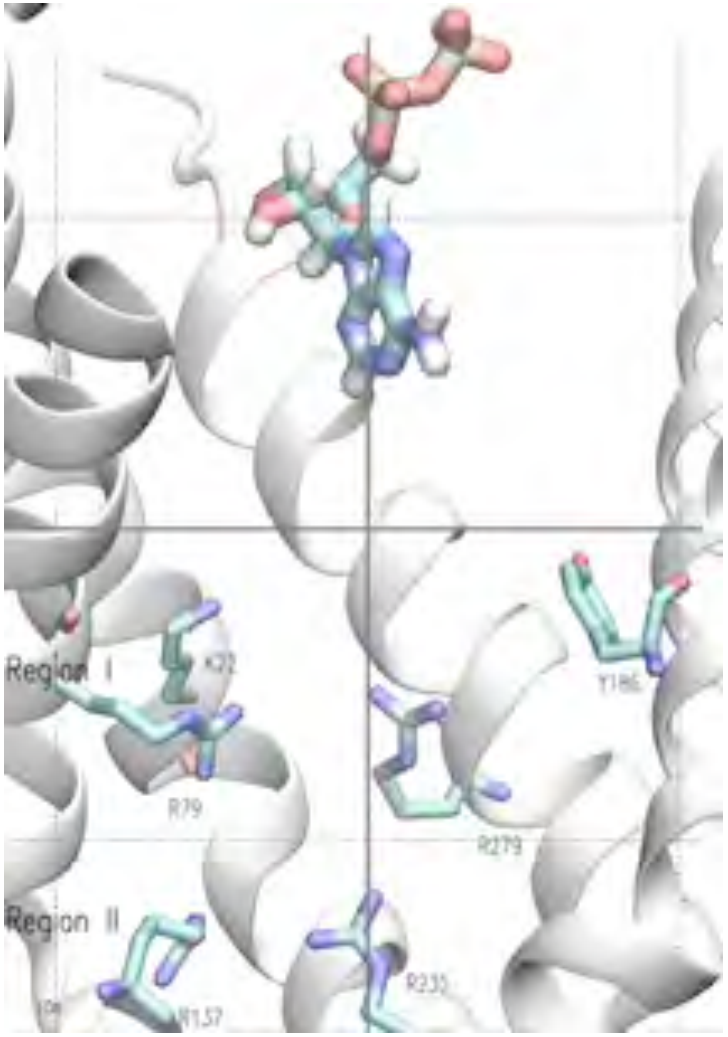
Antidepressant binding site in a neurotransmitter transporter.
Nature 448: 952-956 (2007)

Why Structural Biology at Nanoscale?

- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology

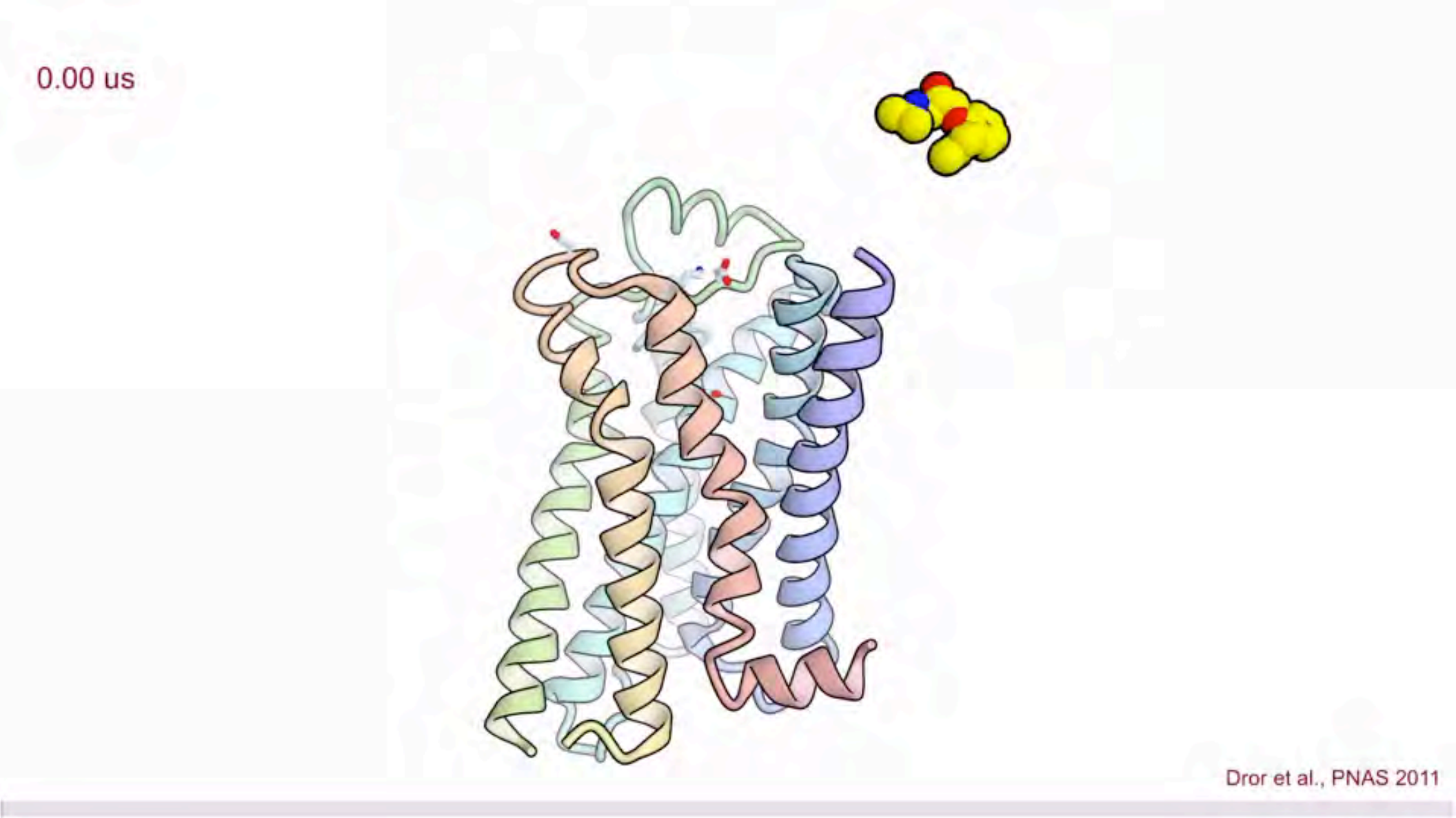






Binding of a small molecule to a binding site
Y. Wang & E.T. PNAS 2010

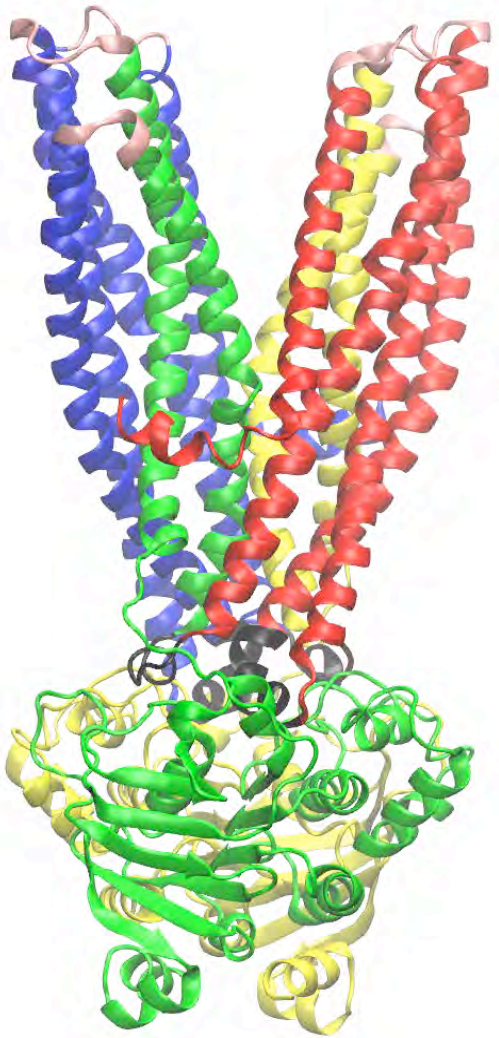
Why Structural Biology at Nanoscale?



Drug binding to a GPCR
Dror, …, Shaw, PNAS, 108:13118–13123 (2011)

Why Structural Biology at Nanoscale?

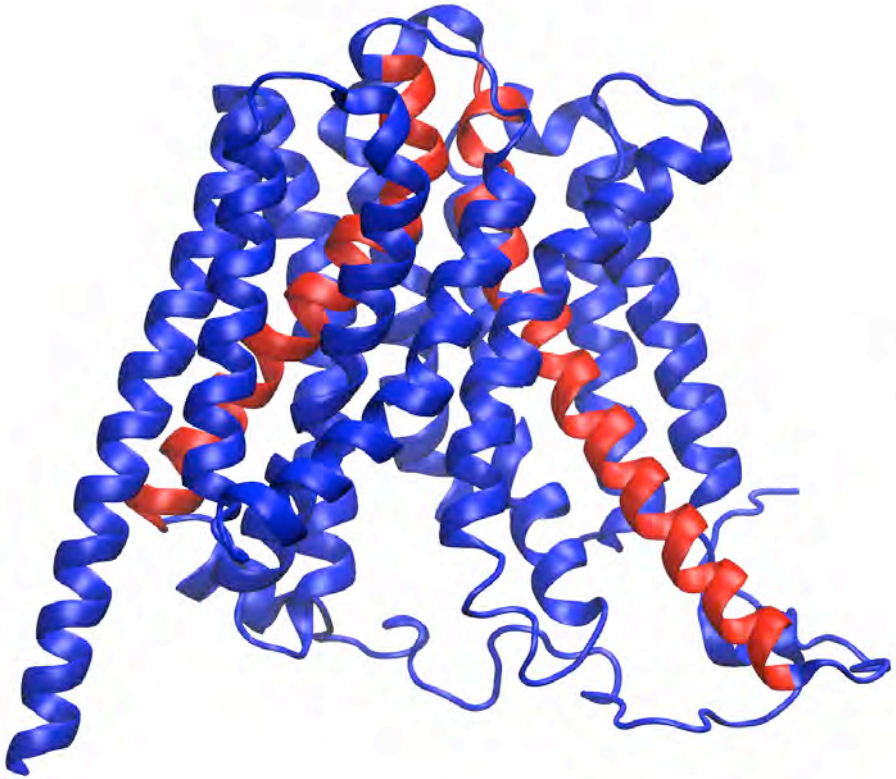
- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology

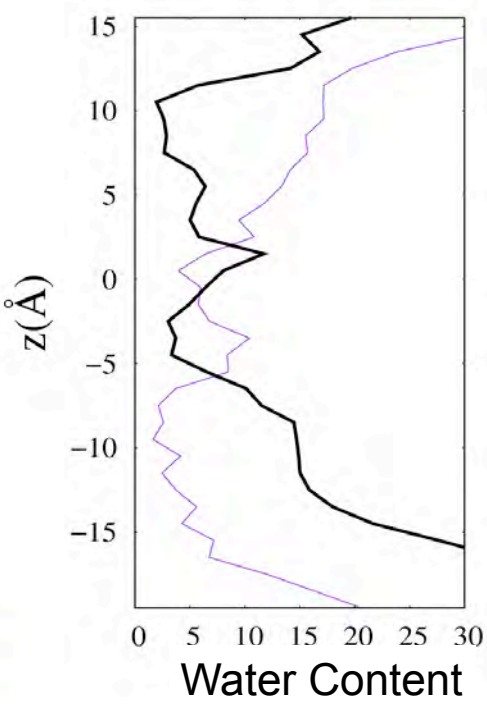


Structural changes underlying function
M. Moradi & E. T. PNAS 2013

Why Structural Biology at Nanoscale?

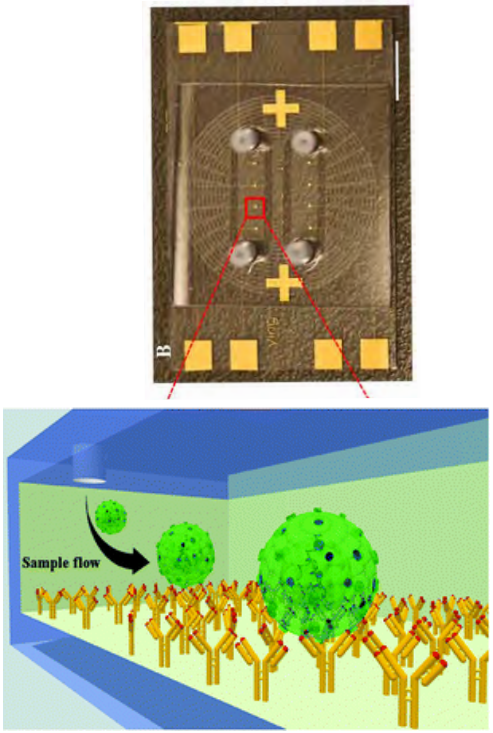
- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology





Structural changes underlying function
M. Moradi, G. Enkavi, & E. T. Nature Comm. 2015

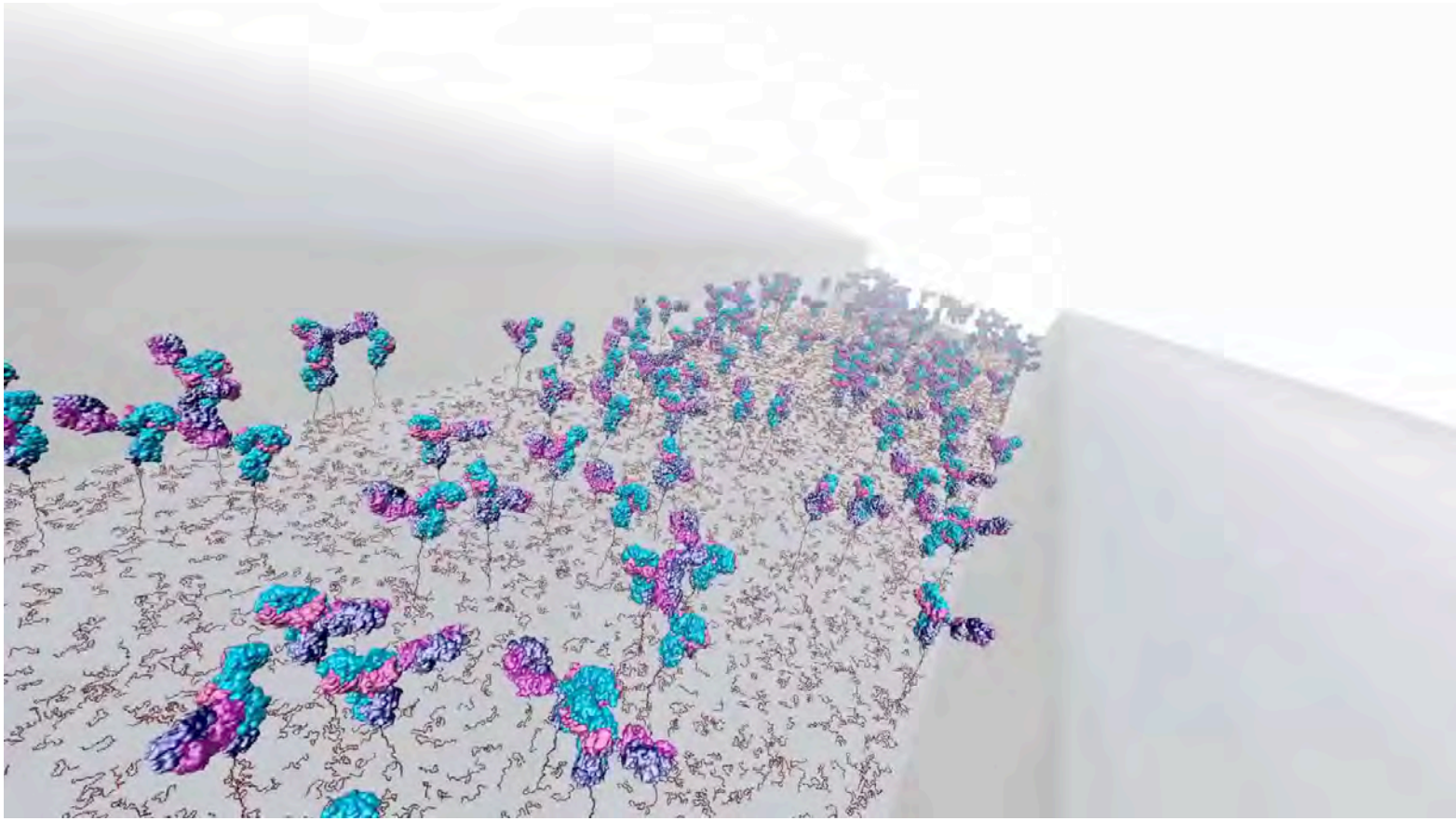
Nano-biotechnology
Microfluidic Sensing Devices



HIV subtype identification

Lab Chip 2012

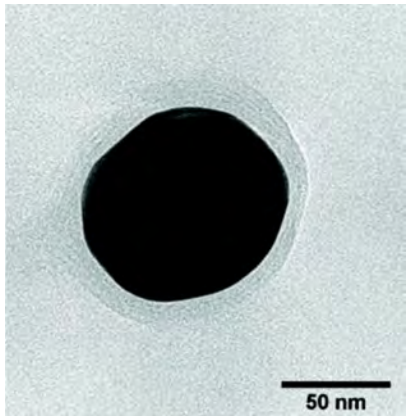
Functionalized nanosurface with antibodies



Created by **nanoBIO Node** tools

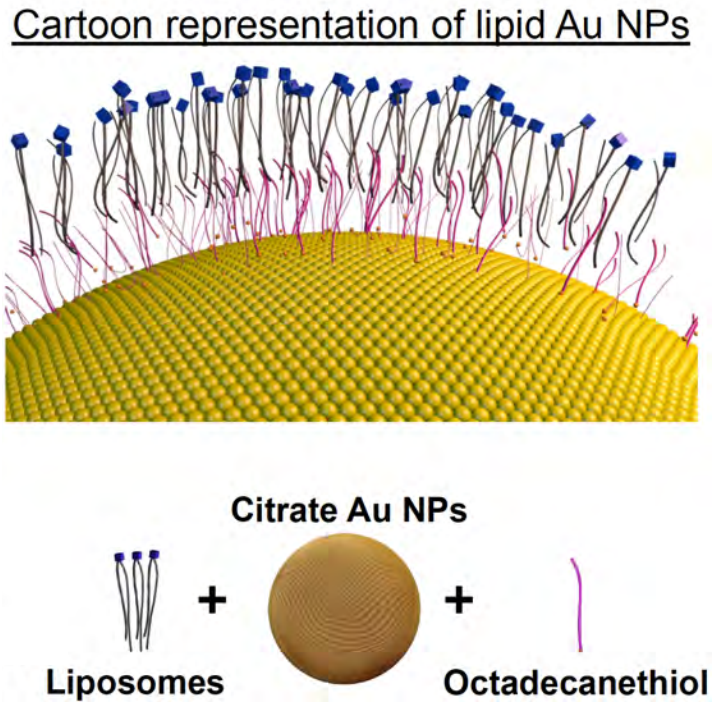
Nano-biotechnology
Gold Nanoparticles as Delivery Vehicles

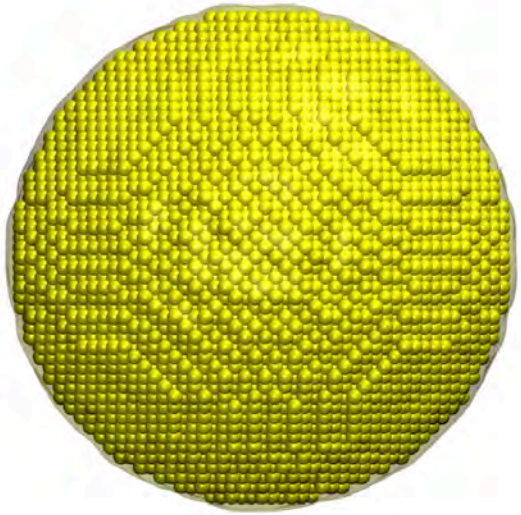
Transmission Electron Micrograph



Yang, J. A.; Murphy, C. J. Langmuir 2012, 28, 5404–5416

Schematic model with no prediction power





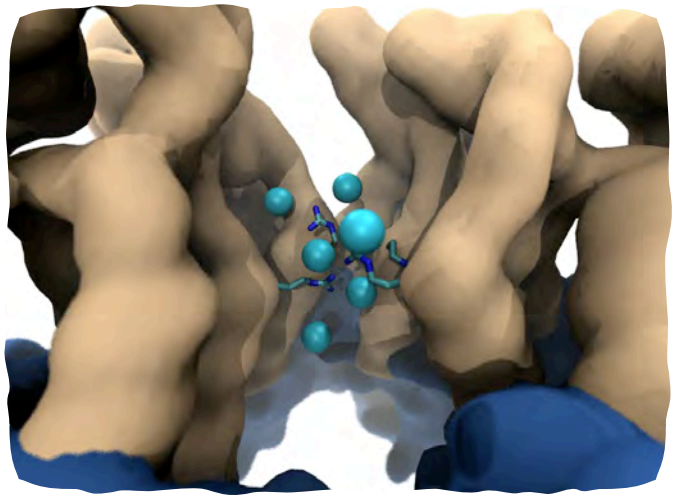
Experiment:
Murphy Lab

Modeling/Simulation:
Tajkhorshid Lab

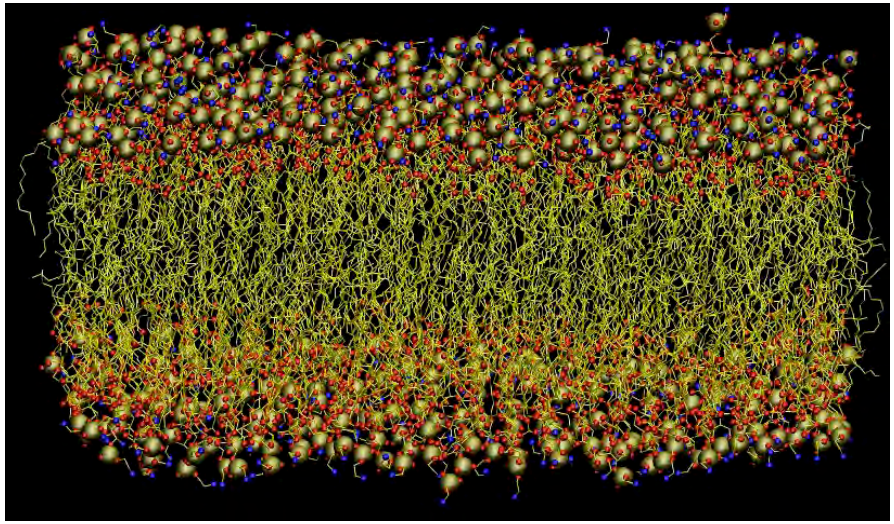
Applications of Computational Methodologies to Structural Biology

Simulation of the dynamics of the molecular system (MD)

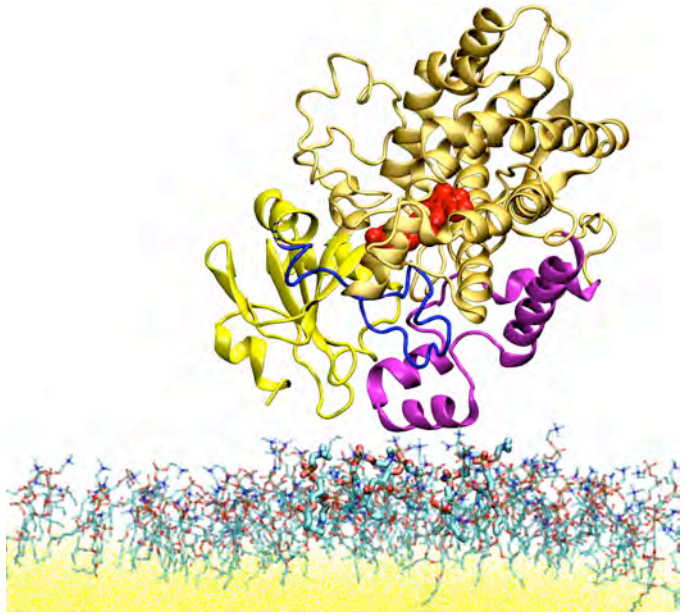
- Calculating ensemble-averaged properties of microscopic systems to compare to macroscopic measurements
- Providing a molecular basis for function
- Describing the molecular/structural changes underlying function
- …



Hydration at the interface of viral shell proteins

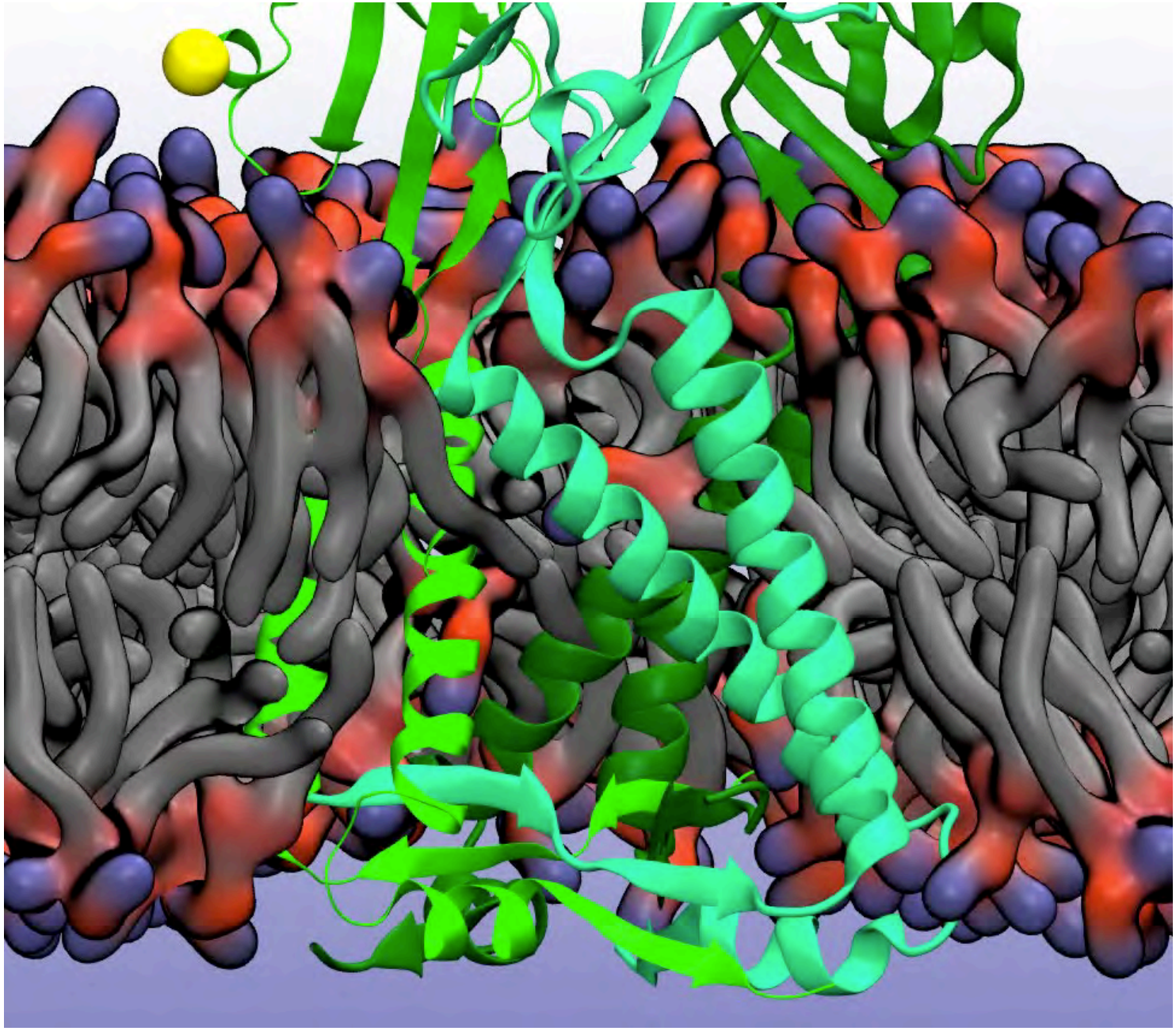


Thermal fluctuations of a phospholipid bilayer



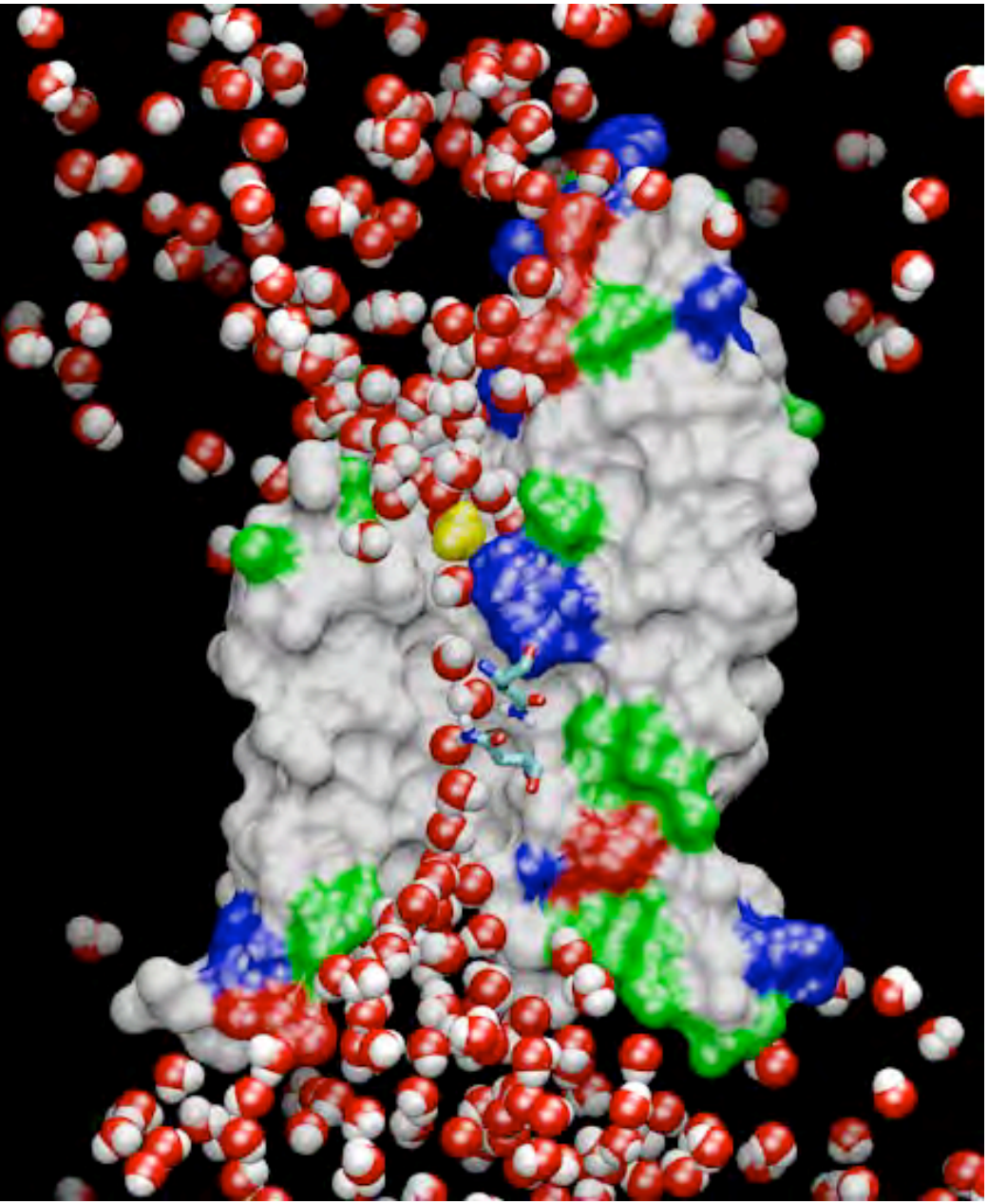
Membrane binding of a coagulation protein

Lipid Protein Interaction



S. Mansoor, …, E. Tajkhorshid, E. Gouaux, Nature, 2016.

Molecular Dynamics Simulations



Solving the Newtonian equations of motion for all particles at every time step

Major limitations:

- Time scale / sampling
- Force field approximations

SPEED LIMIT

1 fs

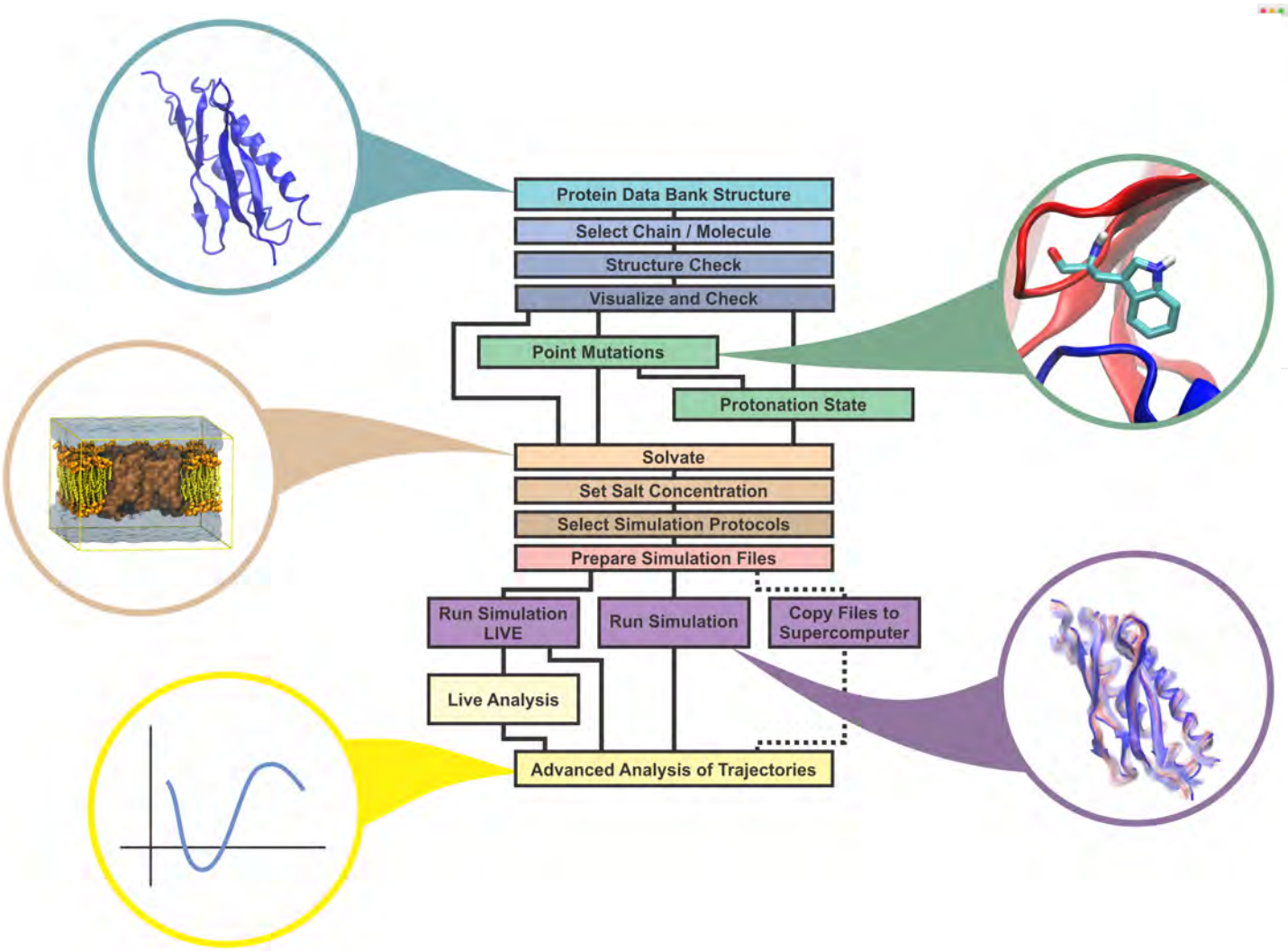
Major advantage:

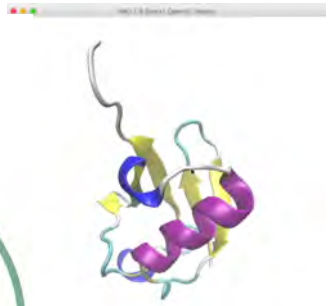
- Unparalleled spatial and temporal resolutions, simultaneously

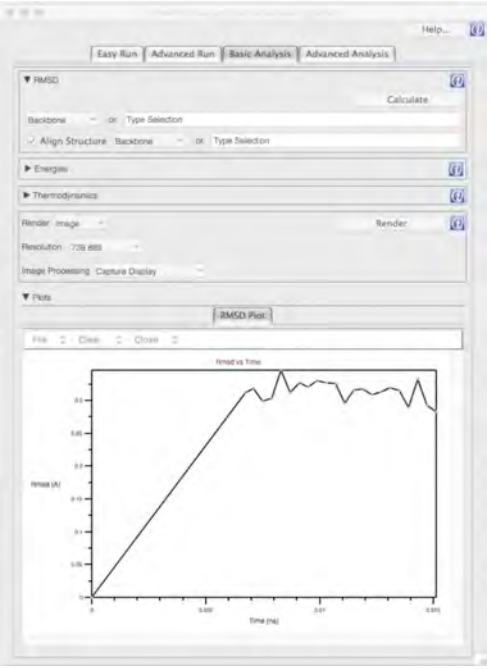
Steps in a Typical MD Simulation

- 1. Prepare molecule
  - Read in pdb and psf file
- 2. Minimization
  - Reconcile observed structure with force field used (T = 0)
- 3. Heating
  - Raise temperature of the system
- 4. Equilibration
  - Ensure system is stable
- 5. Dynamics
  - Simulate under desired conditions (NVE, NpT, etc)
  - Collect your data
- 6. Analysis
  - Evaluate observables (macroscopic level properties)
  - Or relate to single molecule experiments

QwikMD- Gateway to Easy Simulation







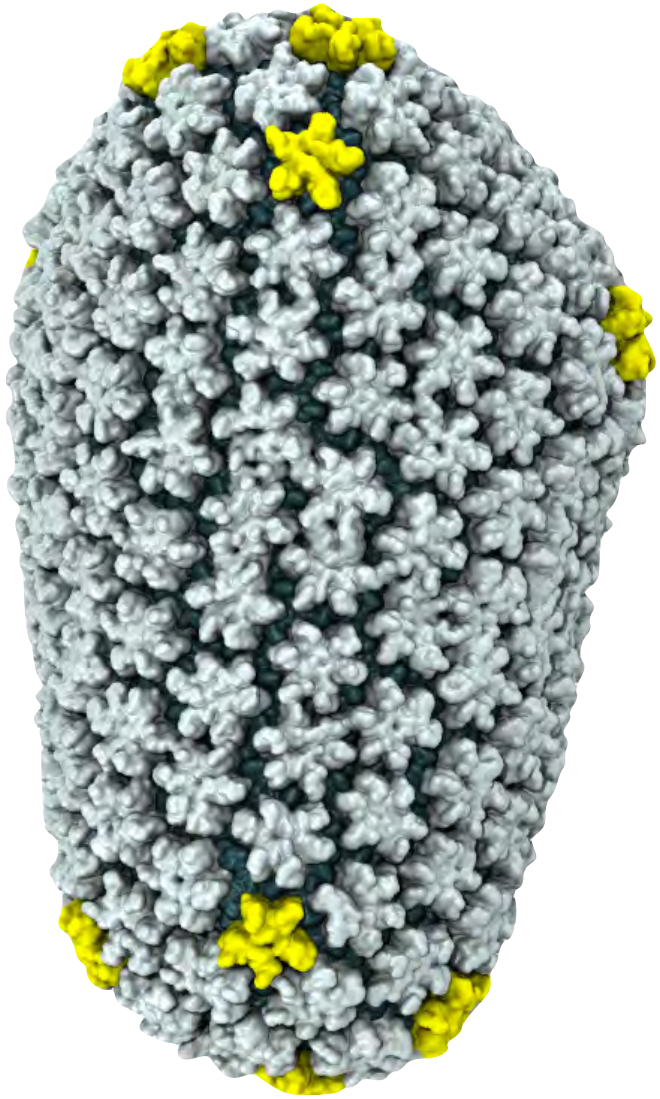
Ribeiro, J. V., …, Schulten, K.. QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. ***Sci. Rep***. 6, 26536; doi: 10.1038/srep26536 (**2016**)

Applications of Computational Methodologies to Cell-Scale Structural Biology

Using computational methods as "structure-building" tools

All experimental Structural biological approaches heavily rely on computational methods to analyze their data

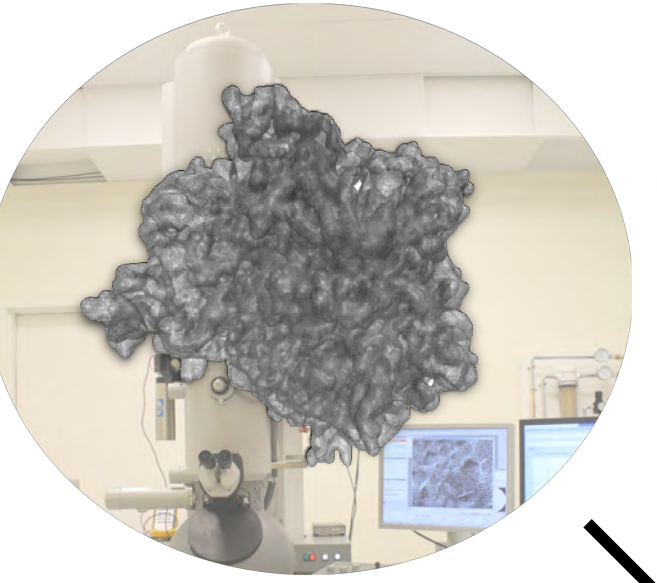
- NMR
- X-ray
- Electron Microscopy
- …



Structural model of HIV virus

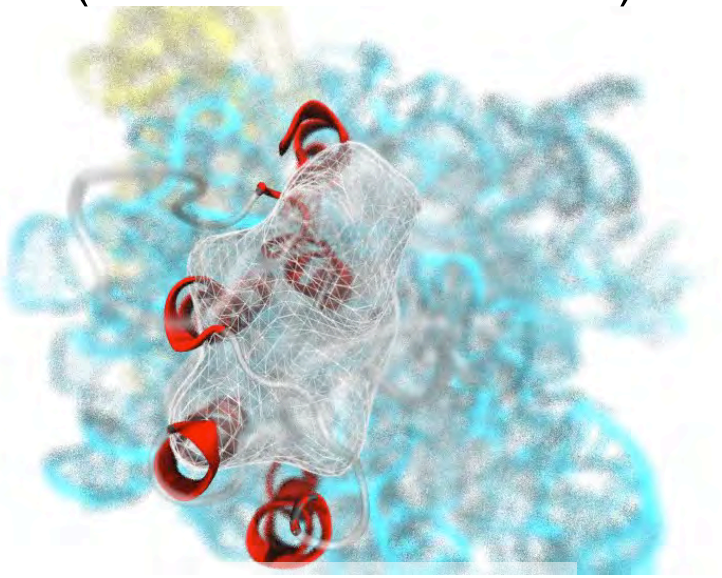
Molecular Dynamics Flexible Fitting (MDFF)

Electron Microscope



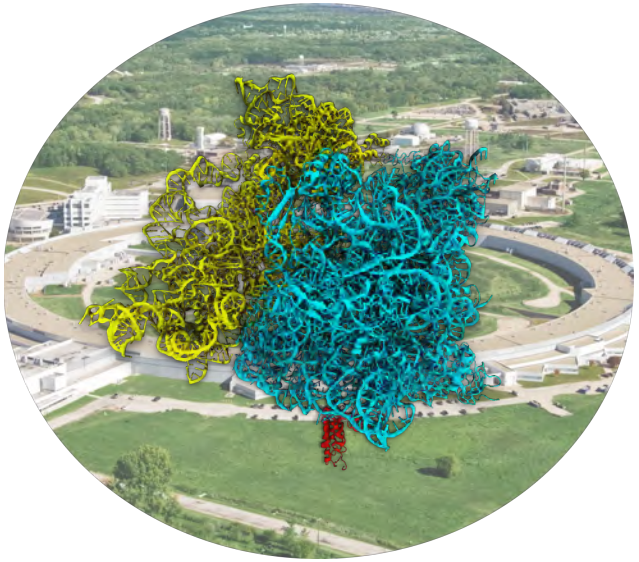
cryo-EM density map

(Ribosome-bound YidC)

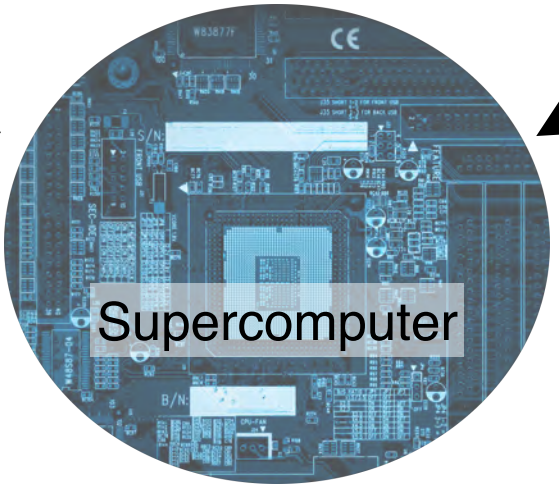


Match through MD

APS Synchrotron



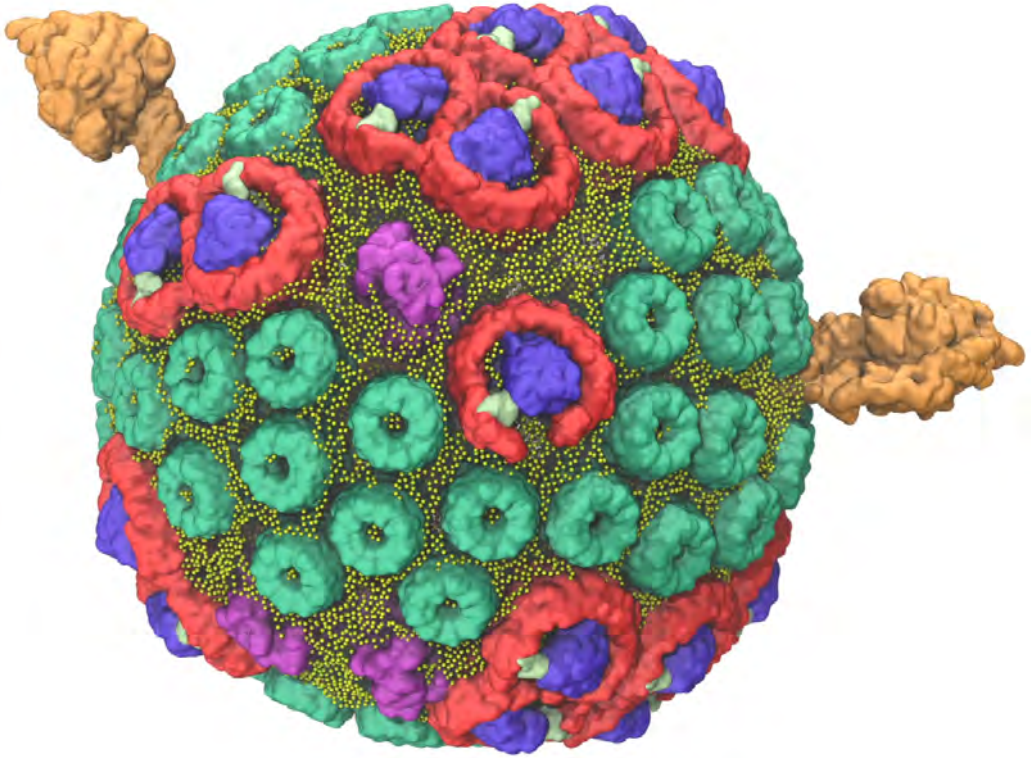
crystallographic structure



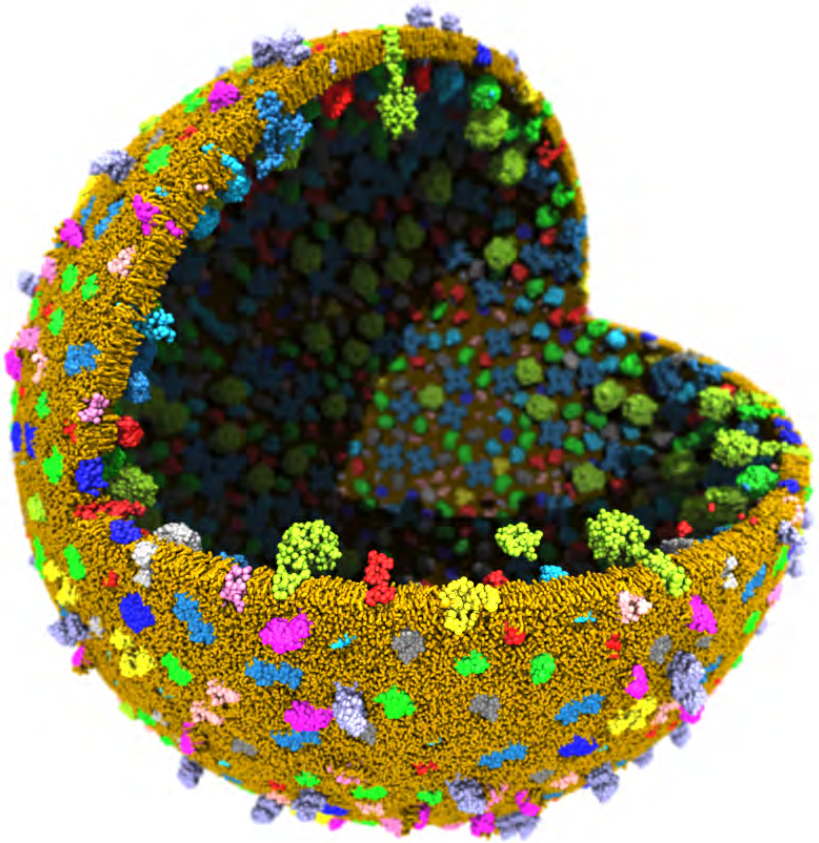
[1] Trabuco et al. *Structure* (2008) 16:673-683.
[2] Trabuco et al. *Methods* (2009) 49:174-180.

Applications of Computational Methodologies to Cell-Scale Structural Biology

Using simulations as a “structure-building” tool



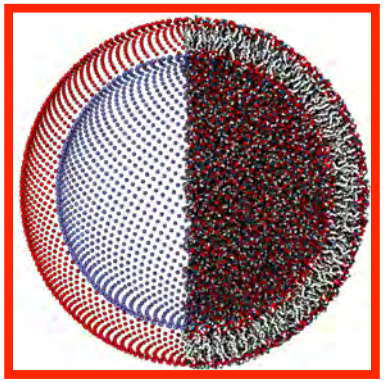
The most detailed model of a chromatophore



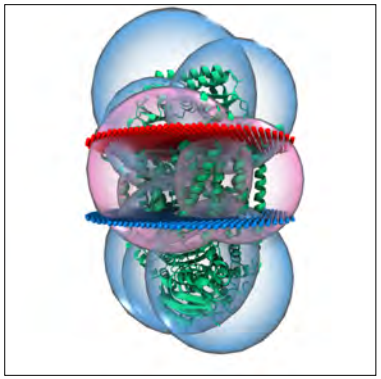
Computational model of a minimal cell envelope

Automated Protein Embedding into Complex Membrane Structures

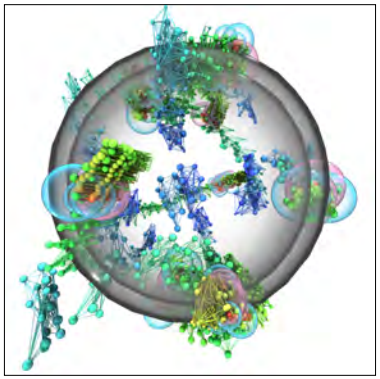
Vesicle Construction



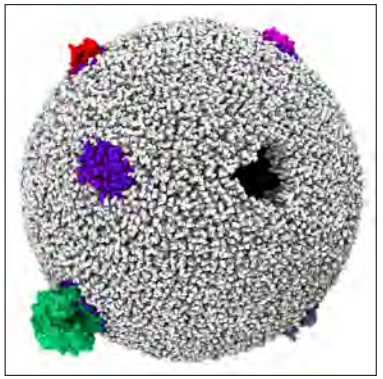
Coarse Grain Protein

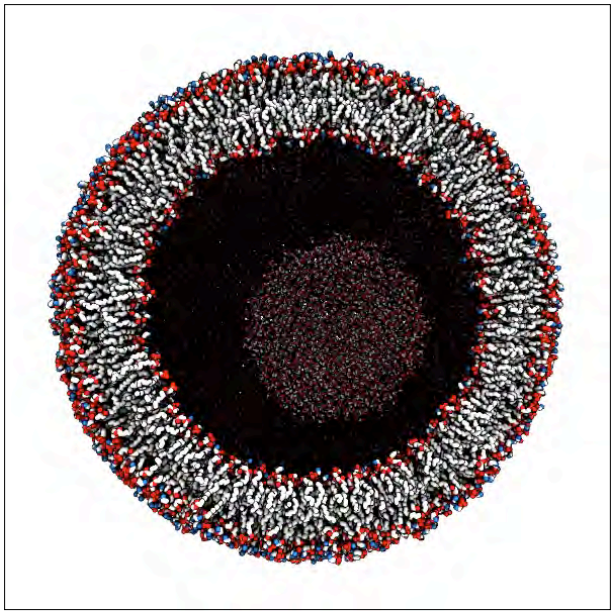


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

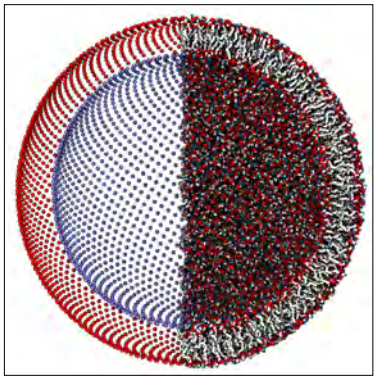
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

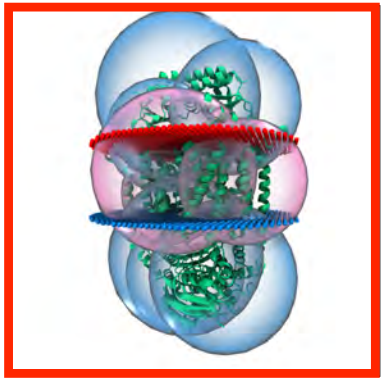
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

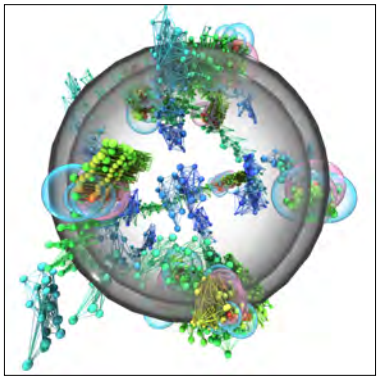
Vesicle Construction



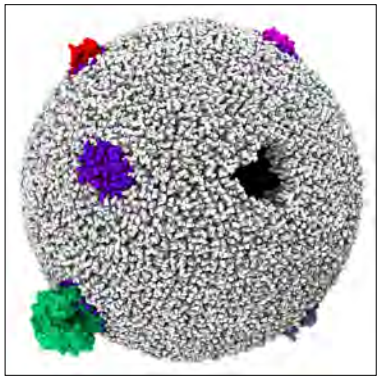
Coarse Grain Protein

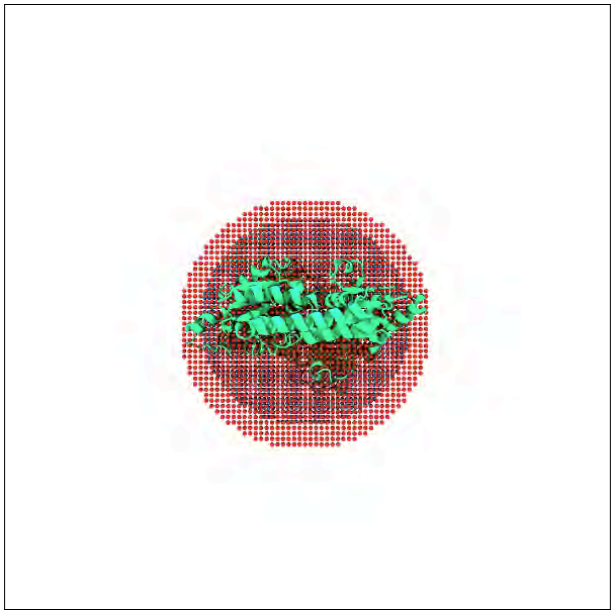


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

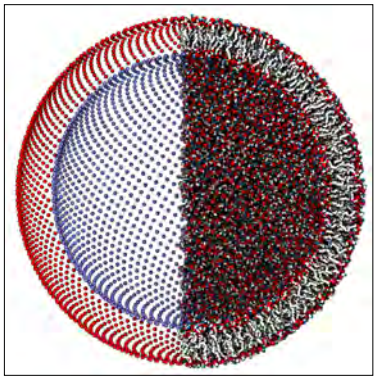
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

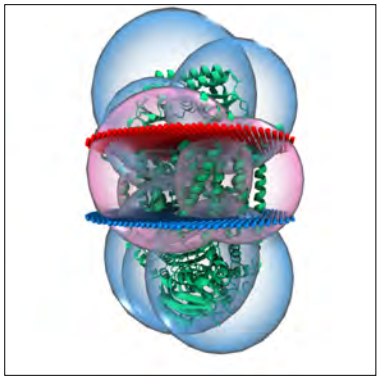
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

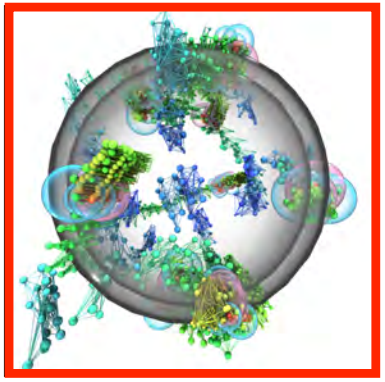
Vesicle Construction



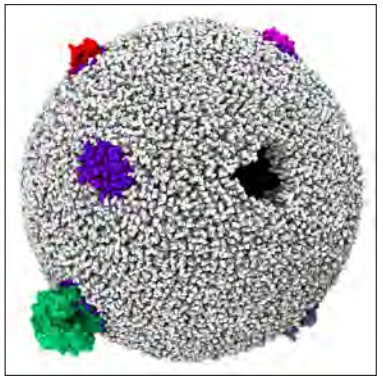
Coarse Grain Protein

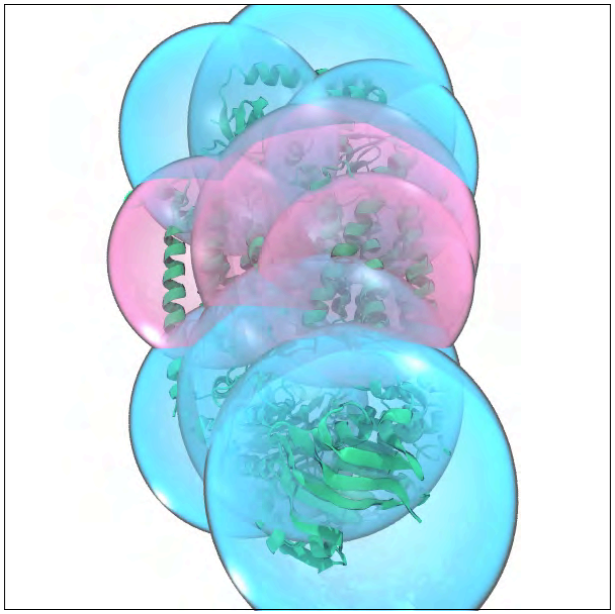


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

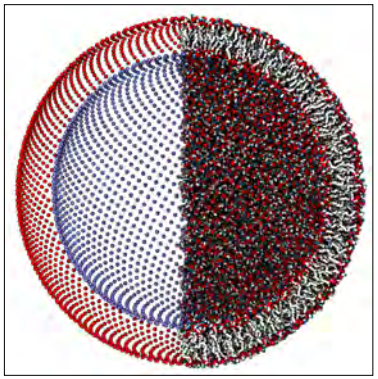
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

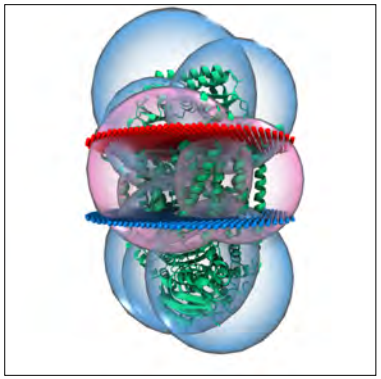
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

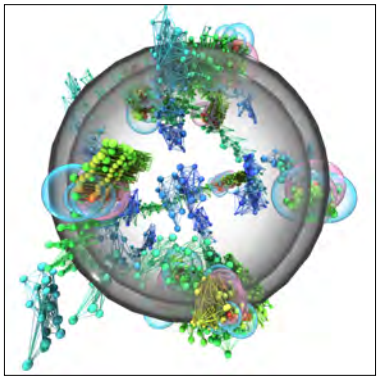
Vesicle Construction



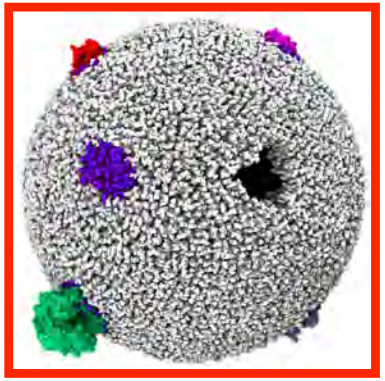
Coarse Grain Protein

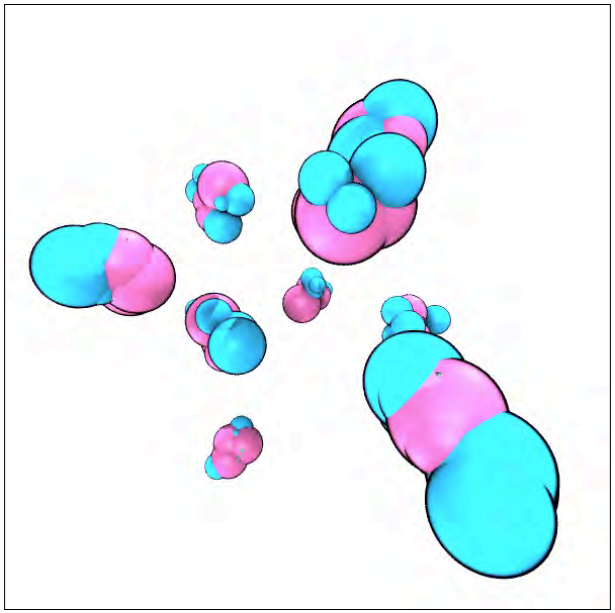


CG Protein Placement



Combine Lipid + Protein





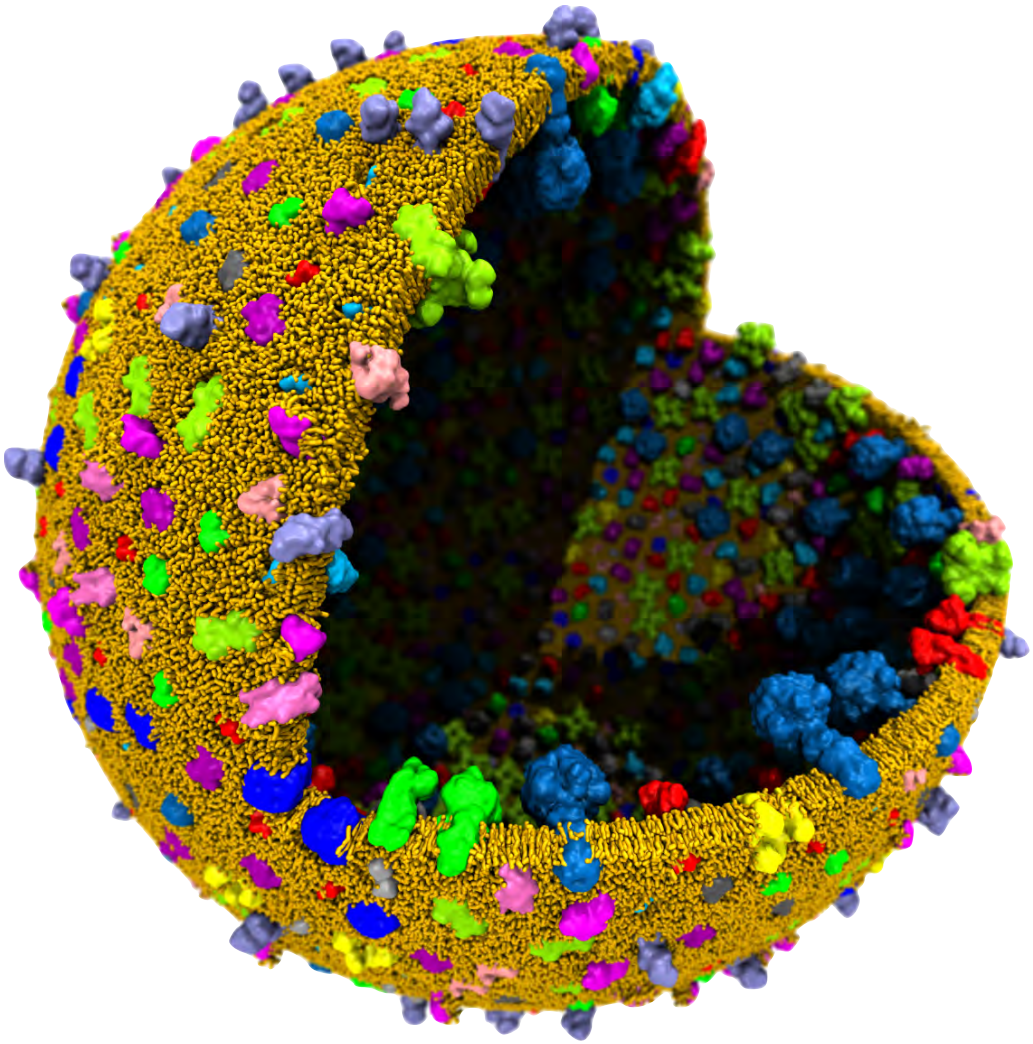
Distribution of proteins across the membrane surface (dense environment)

- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

113 million Martini particles representing **1 billion** atoms



0.4 μm

Protein Components	Copy #
Aquaporin Z	97
Copper Transporter (CopA)	166
F1 ATPase	63
Lipid Flipase (MsbA)	29
Molybdenum transporter (ModBC)	130
Translocon (SecY)	103
Methionine transporter (MetNI)	136
Membrane chaperon (YidC)	126
Energy coupling factor (ECF)	117
Potassium transporter (KtrAB)	148
Glutamate transporter (Glt_{Tk})	41
Cytidine-Diphosphate diacylglycerol (Cds)	50
Membrane-bound protease (PCAT)	57
Folate transporter (FolT)	134
	1,397

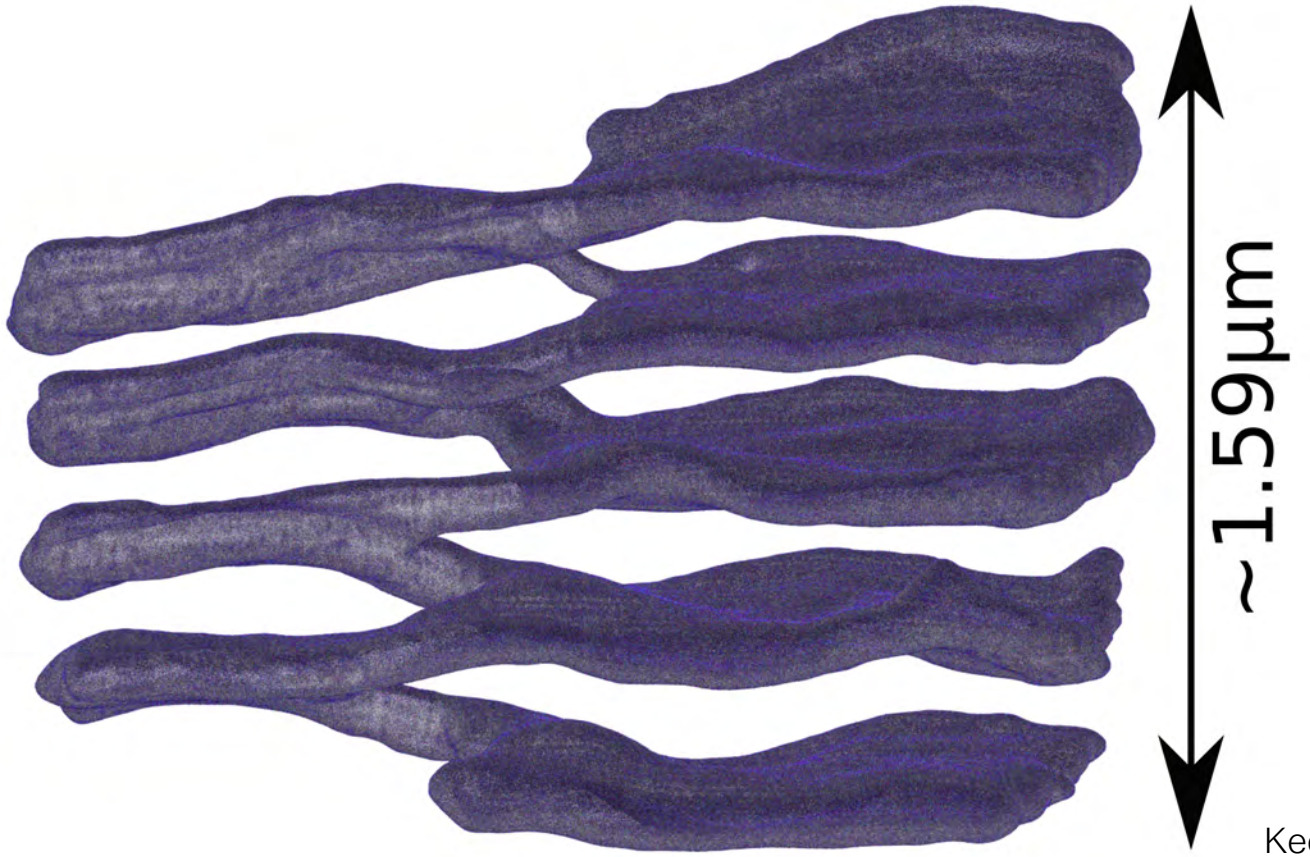
3.7 M lipids (DPPC), 2.4 M Na^+ & Cl^- ions,
104 M water particles (4 H_2O / particle)

Applications of Computational Methodologies to Cell-Scale Structural Biology

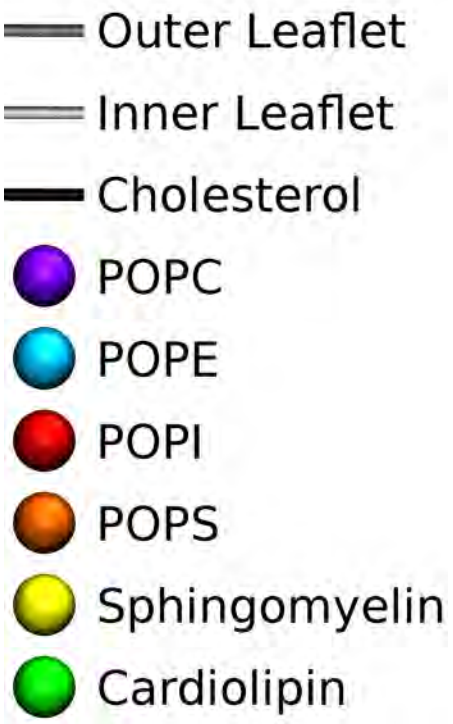
Guided Construction of Membranes from Experimental Data

E**x**perimentally-Derived **M**embrane of **A**rbitrary **S**hape Builder

Terasaki Ramp
~4 Billion Atoms



Terasaki et al., *Cell*, **2013**.

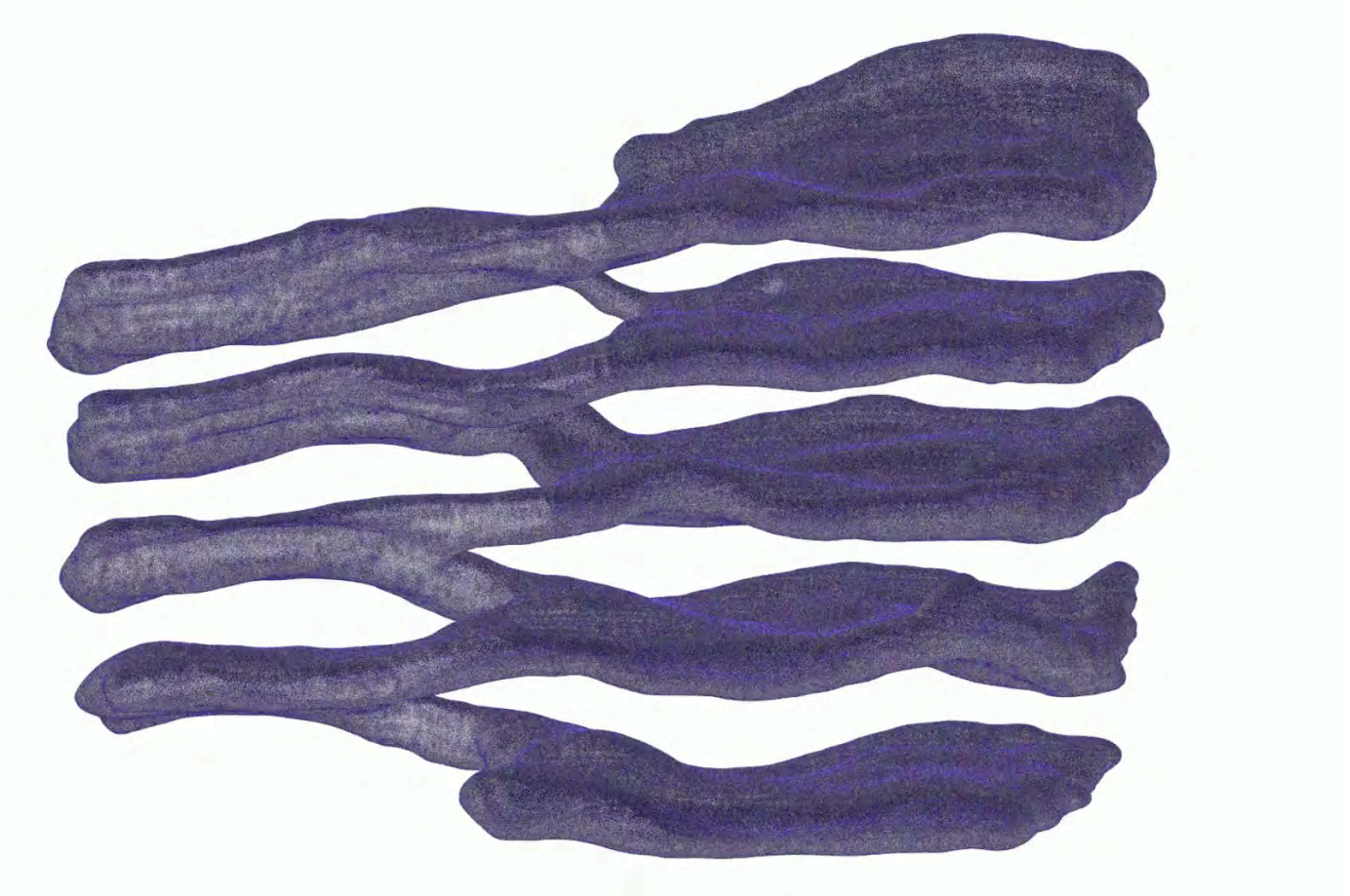


Keenan and Huang, *J. Dairy Sci.*, **1972**.

Applications of Computational Methodologies to Cell-Scale Structural Biology

Guided Construction of Membranes from Experimental Data

E**x**perimentallv-Derived **M**embrane of **A**rbitrarv **S**hape Builder



Terasaki Ramp
~4 Billion Atoms

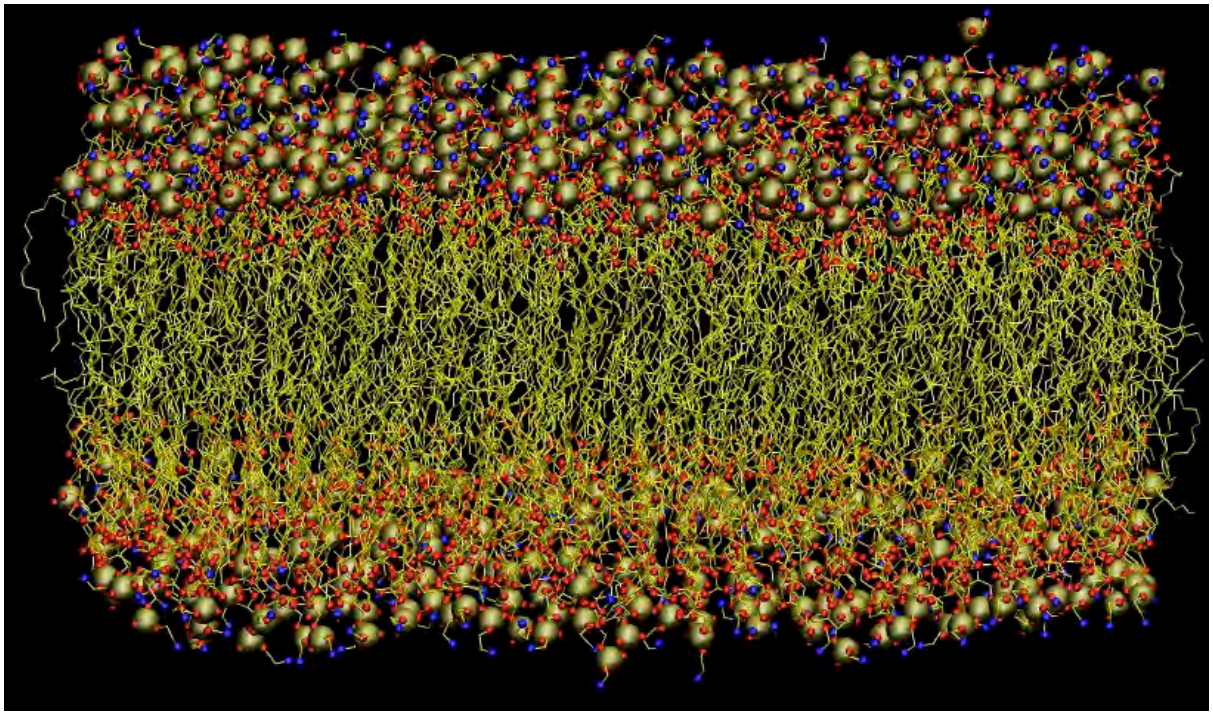
Outer Leaflet
Inner Leaflet
Cholesterol
POPC
POPE
POPI
POPS
Sphingomyelin
Cardiolipin

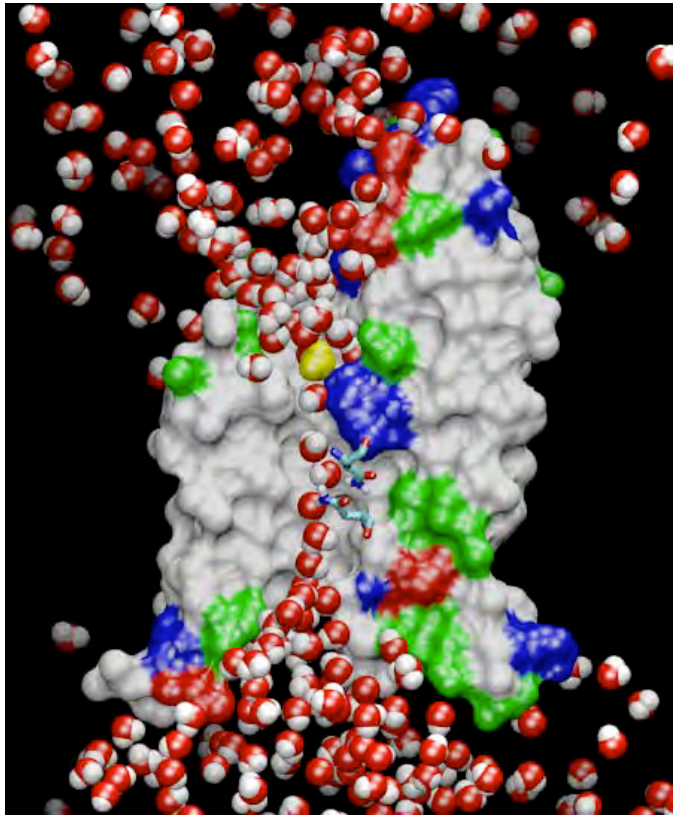
Terasaki et al., *Cell*, **2013**.

Keenan and Huang, *J. Dairy Sci.*, **1972**.

Molecular Dynamics Simulation

- Generating a thermodynamic ensemble (Sampling / Statistic)
- Taking into account fluctuations/dynamics in interpretation of experimental observables
- Describing molecular processes + free energy
- Help with molecular modeling





Classical Molecular Dynamics

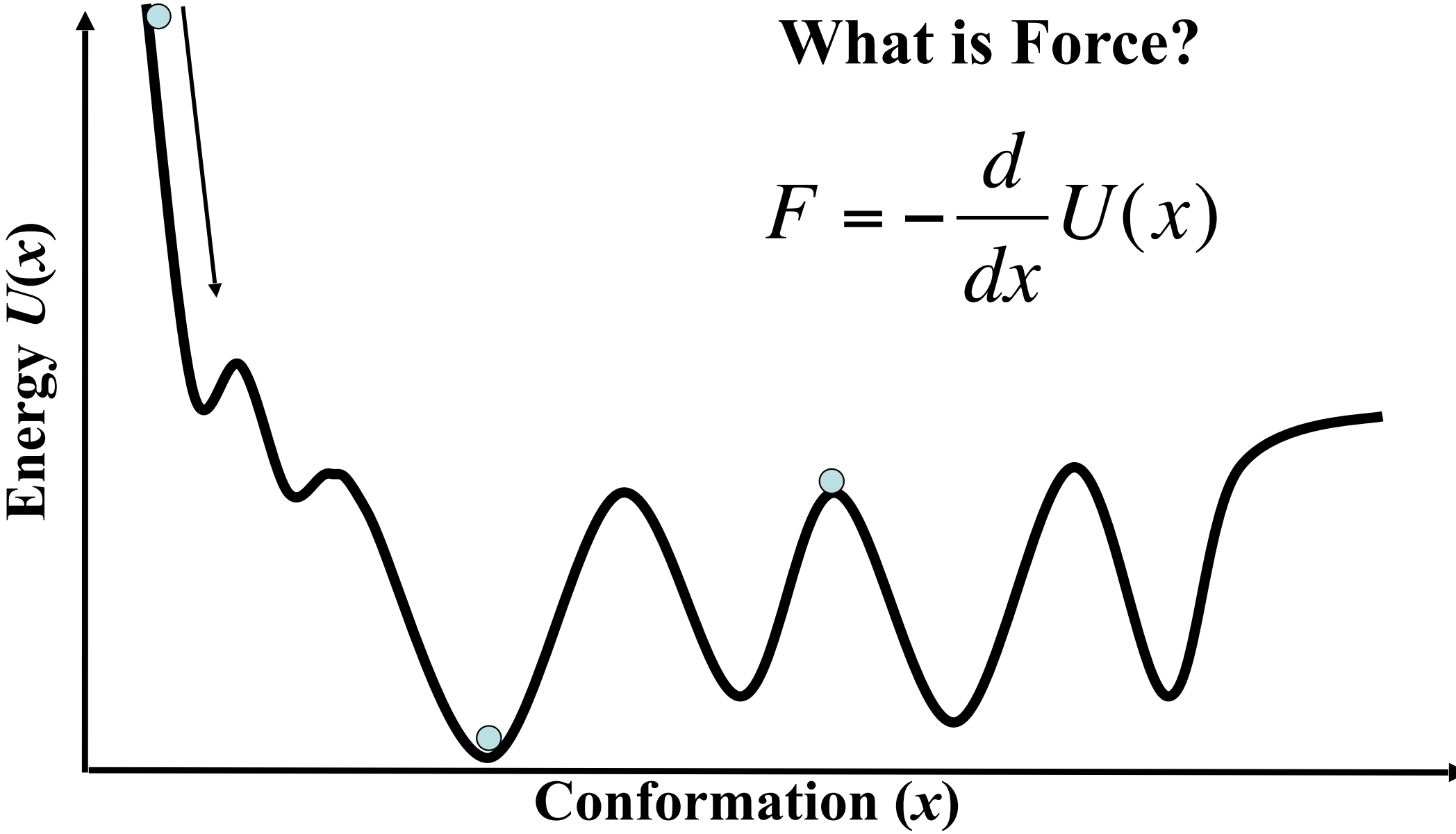
$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t$$

$$\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \boldsymbol{a}(t)\delta t$$

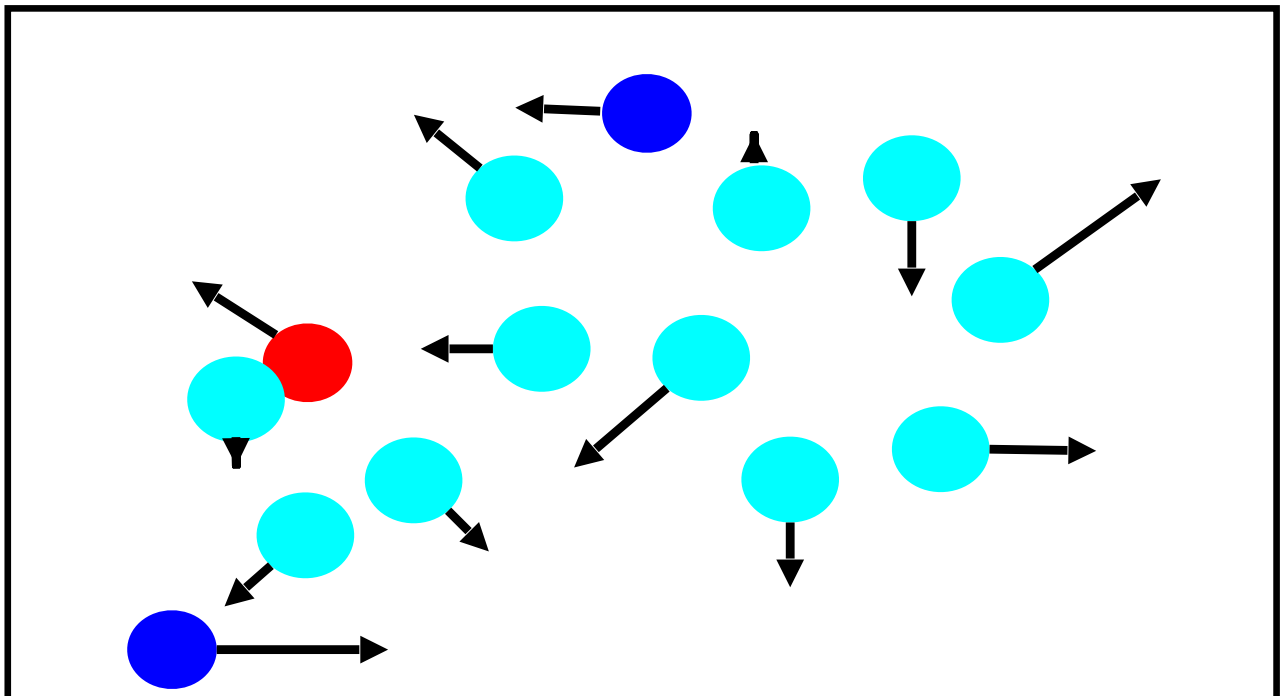
$$\boldsymbol{a}(t) = \boldsymbol{F}(t)/m$$

$$\boldsymbol{F} = -\frac{d}{d\boldsymbol{r}}U(\boldsymbol{r})$$

Potential Energy (hyper)Surface



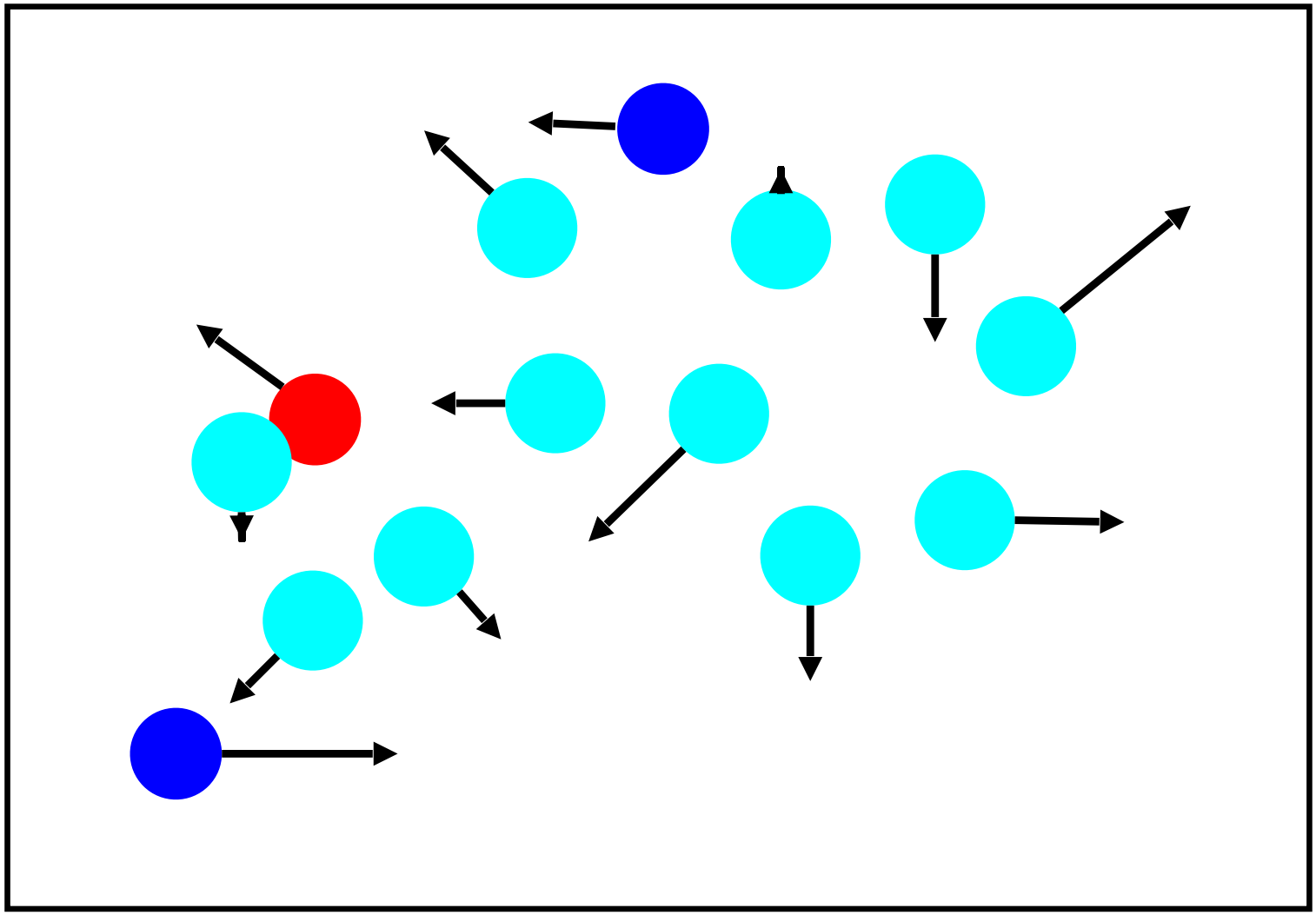
Classical Molecular Dynamics



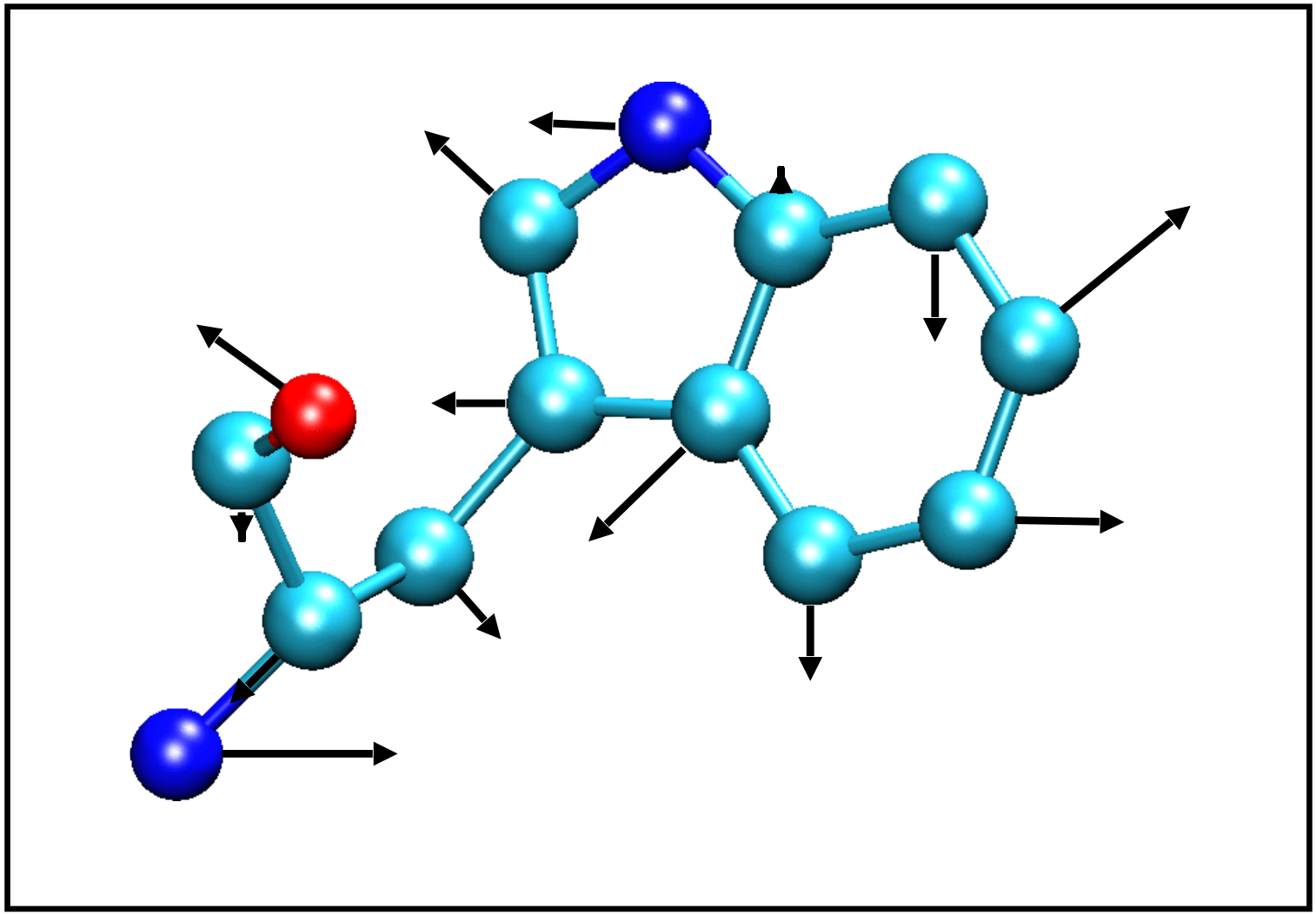
$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

$$\mathbf{F}(\mathbf{r}) = \left(-\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}^{\,2}} - 12\frac{\varepsilon_{ij}}{|r_{ij}|}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]\right)\hat{\mathbf{r}}_{ij}$$

Classical Molecular Dynamics



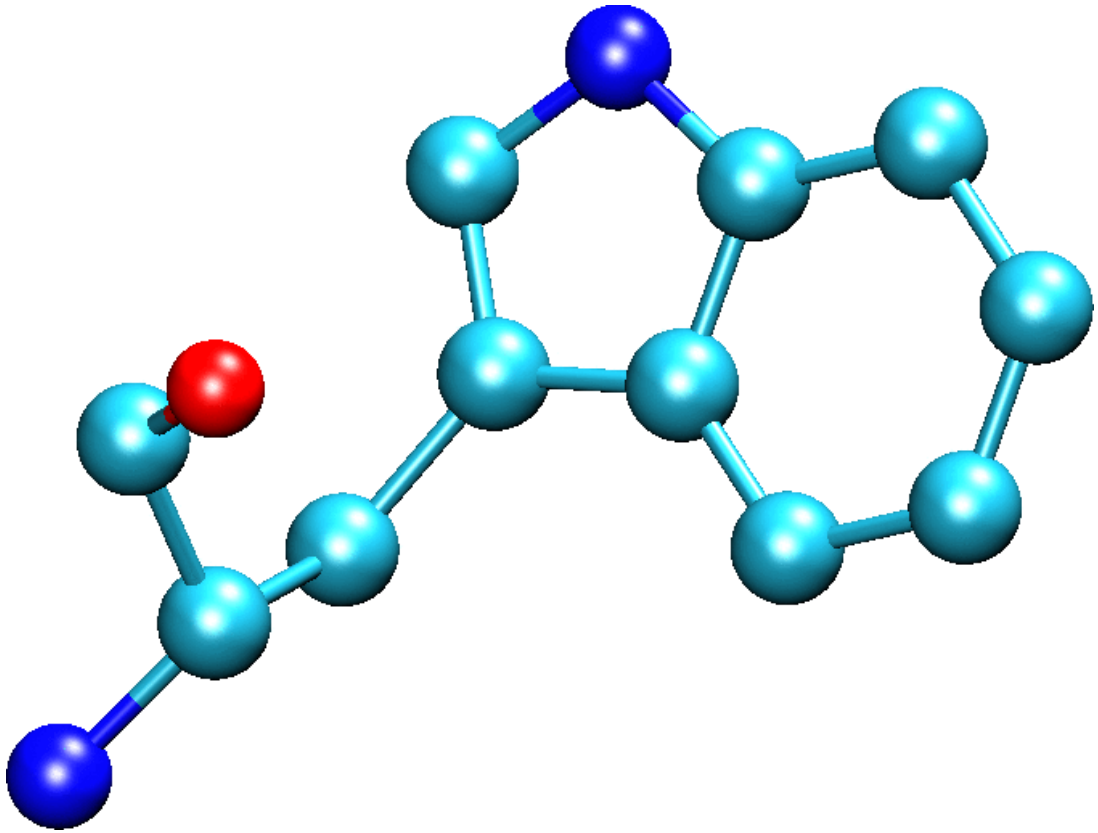
Classical Molecular Dynamics



Bond definitions, atom types, atom names, parameters, ….

What is a Force Field?

In molecular dynamics a molecule is described as a series of charged points (atoms) linked by springs (bonds).



To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms, one uses a force field.
The force field is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

Energy Functions

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

U_{bond} = oscillations about the equilibrium bond length

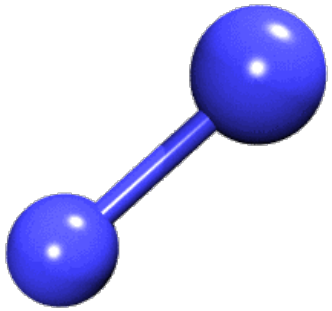
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle

$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond

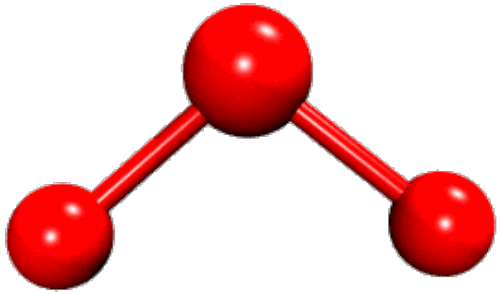
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

Energy Terms Described in the CHARMm Force Field

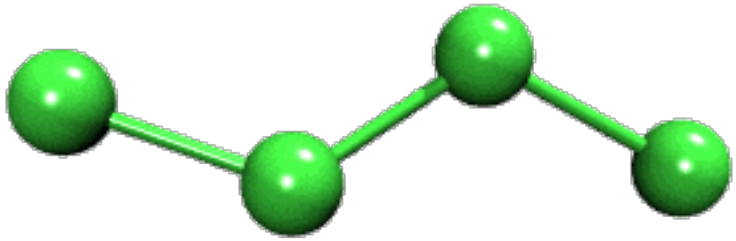
Bond



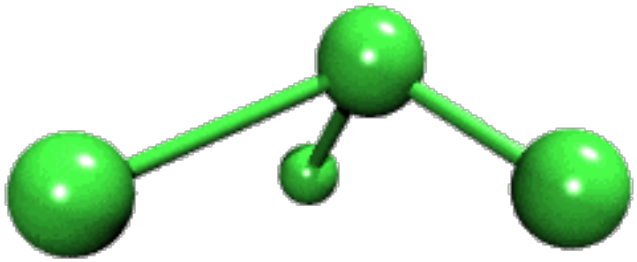
Angle



Dihedral



Improper



Classical Dynamics

F=ma at 300K

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

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

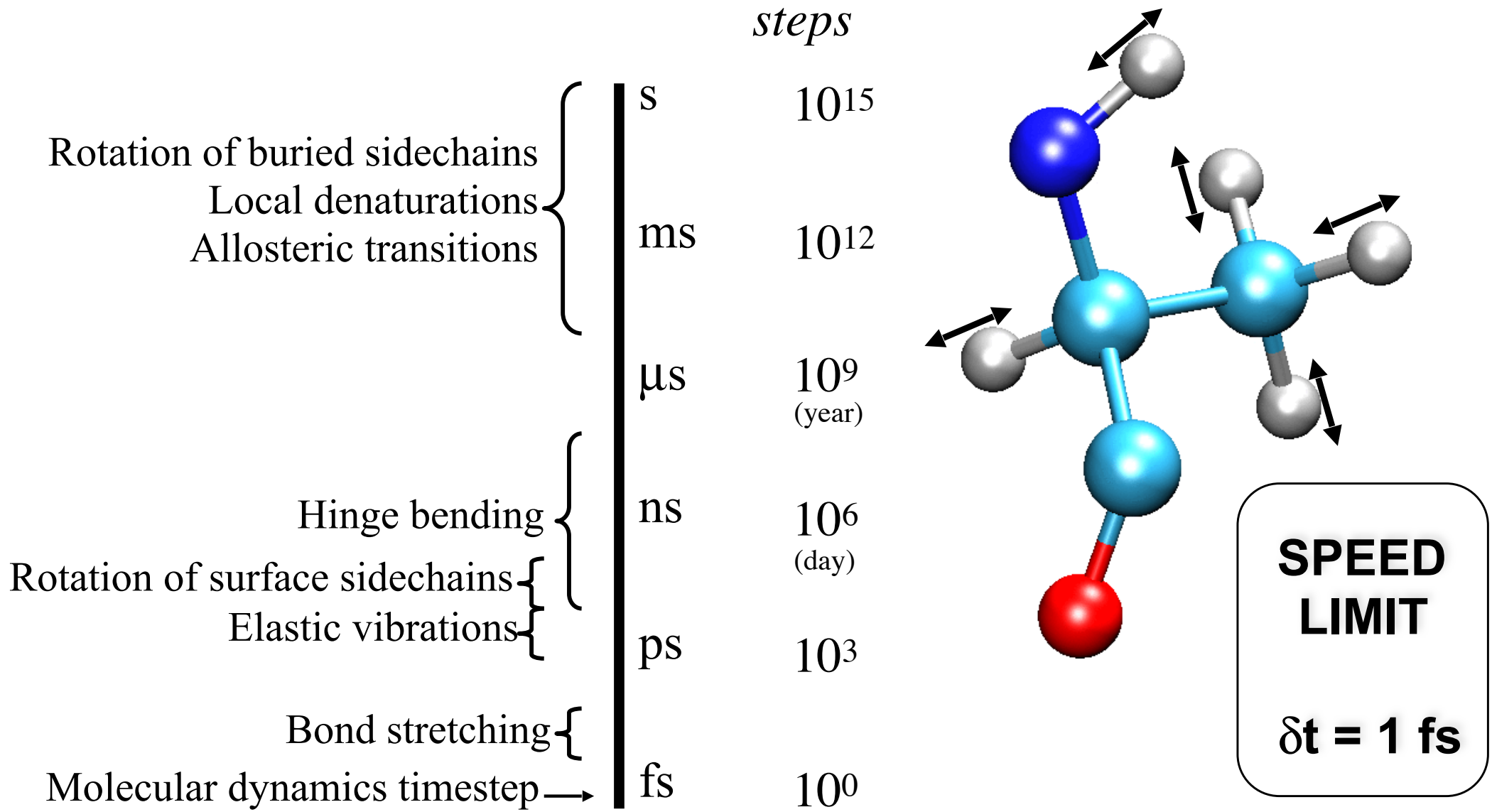
yields a set of 3N coupled 2nd-order differential equations that can be propagated forward (or backward) in time.

Initial coordinates obtained from crystal structure, velocities taken at random from Boltzmann distribution.

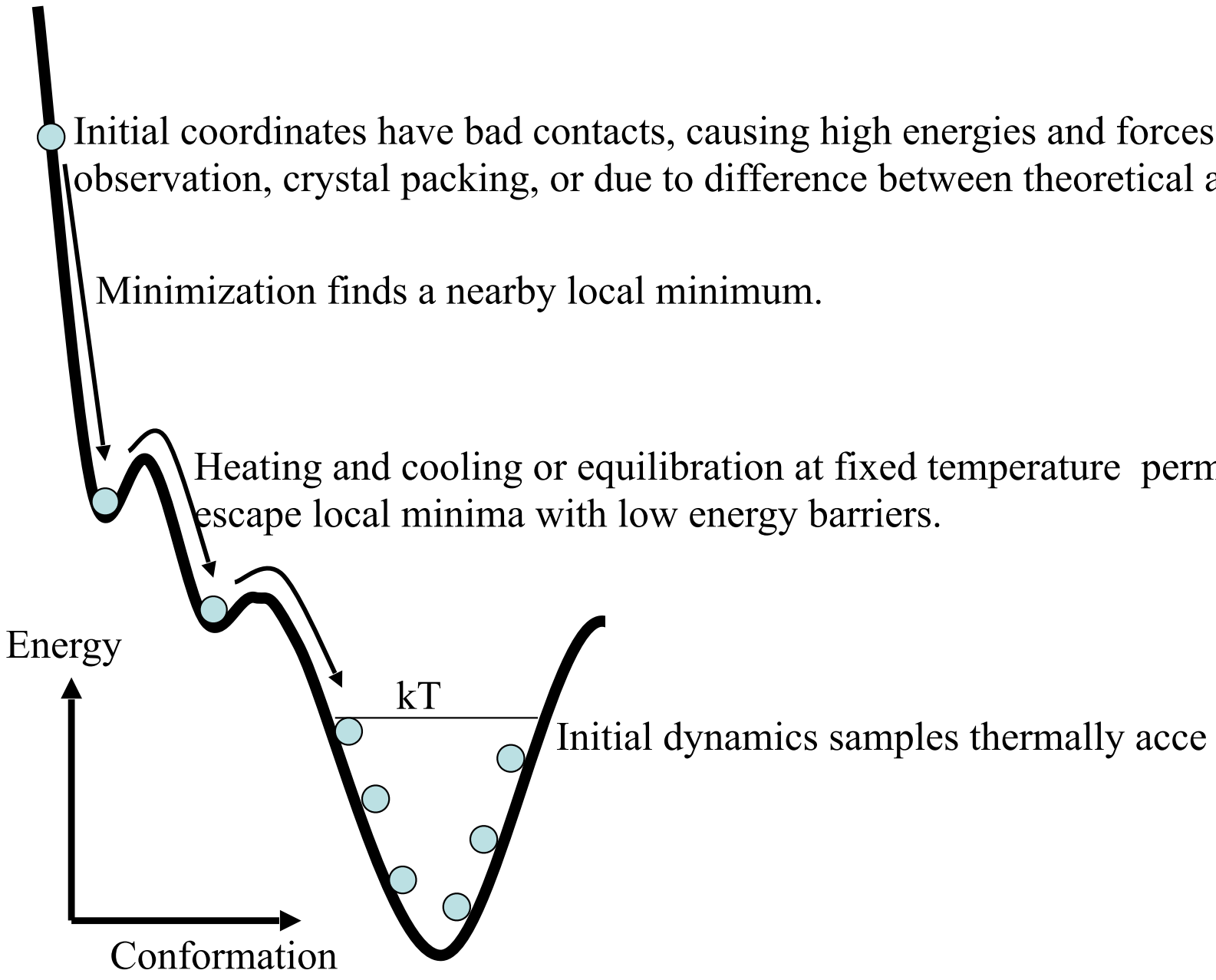
Langevin dynamics deals with each atom separately, balancing a small friction term with Gaussian noise to control temperature:

$$m\,\ddot{\vec{r}} \;=\; \vec{F}(\vec{r}) \;-\; \gamma\, m\,\dot{\vec{r}} \;+\; \vec{R}(t)$$

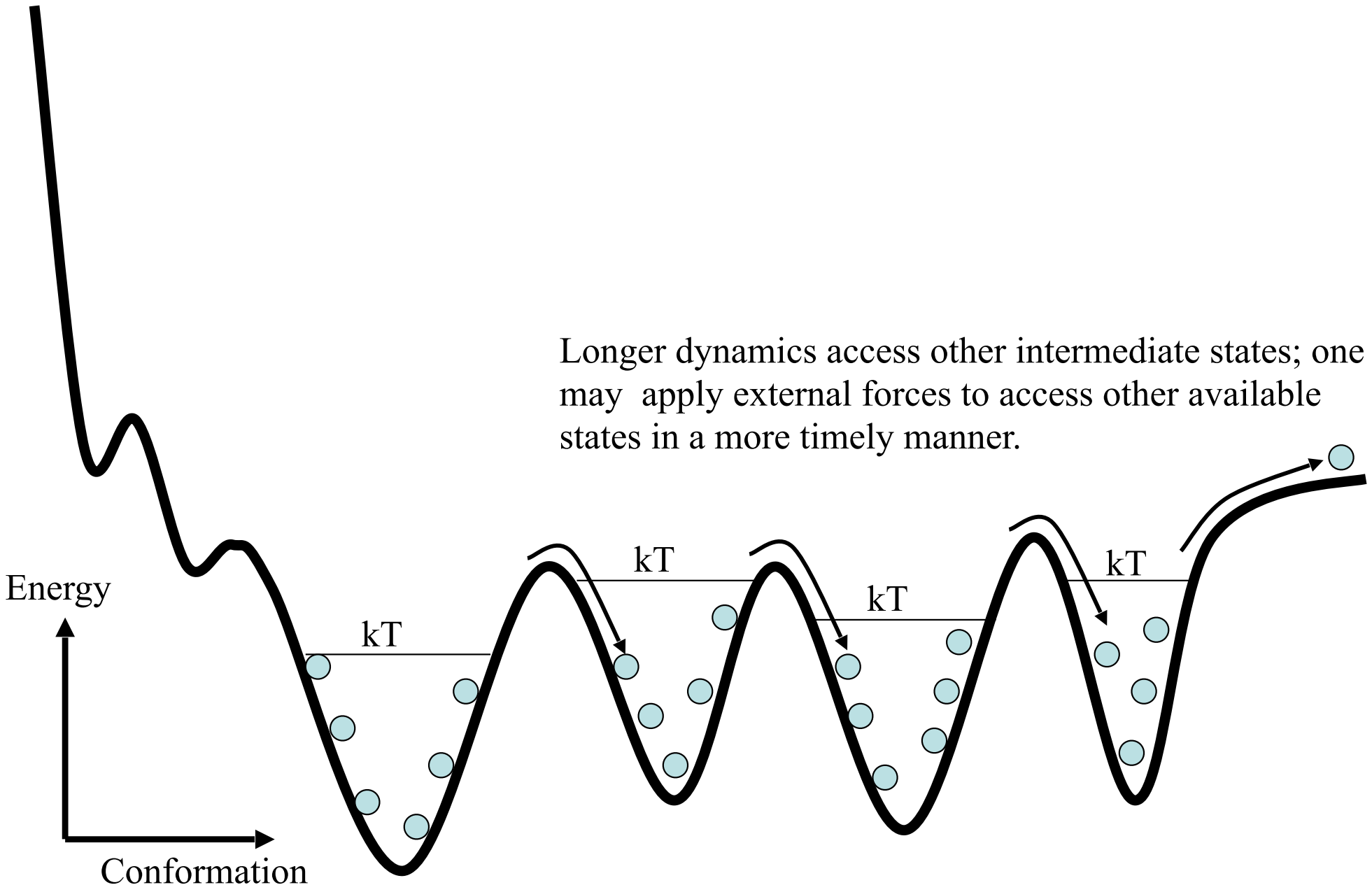
The most serious bottleneck



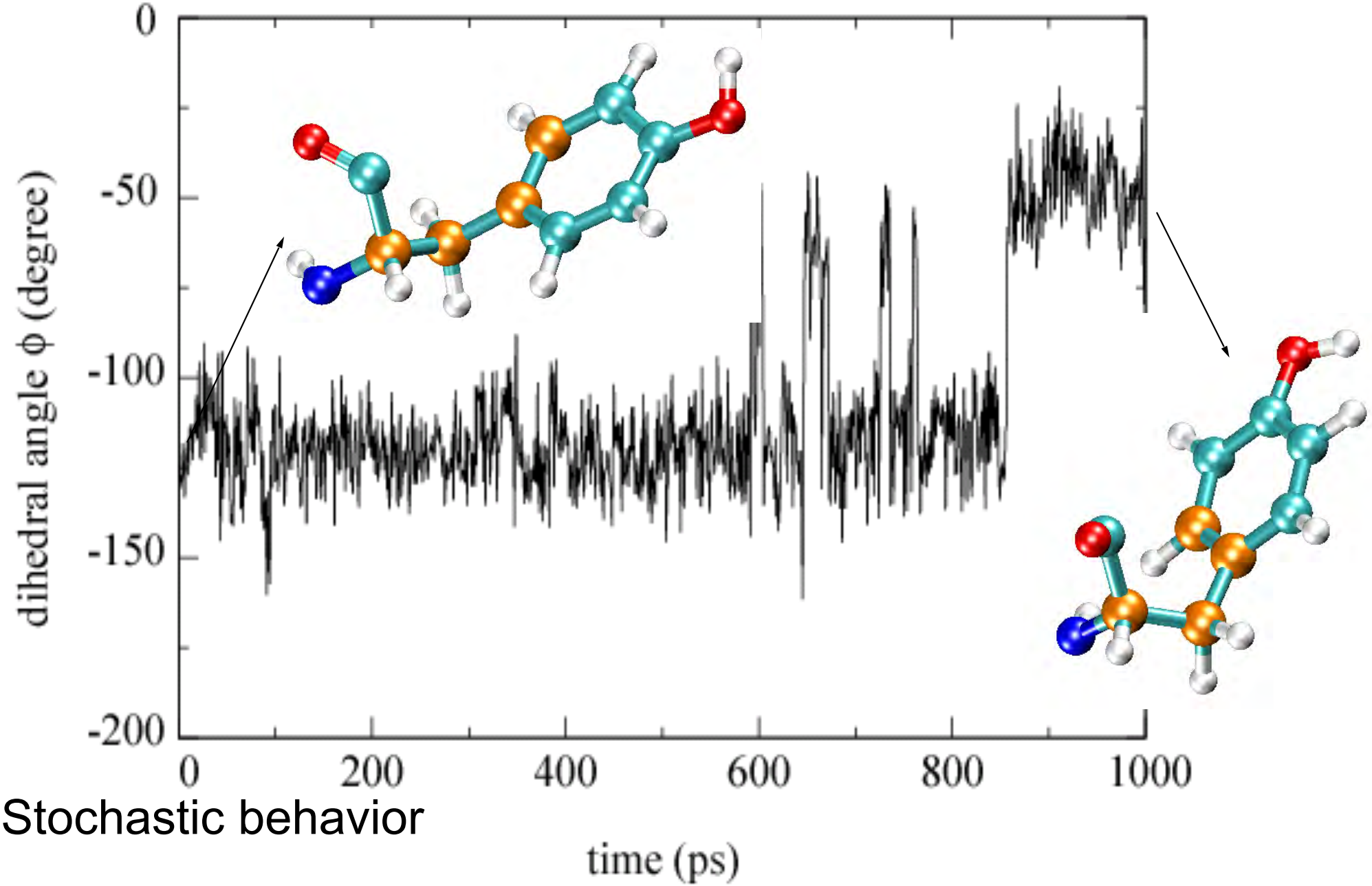
Molecular Dynamics to Sample Energy Landscape



Molecular Dynamics to Sample Energy Landscape



Patience is required to observe Molecular Events



Stochastic behavior

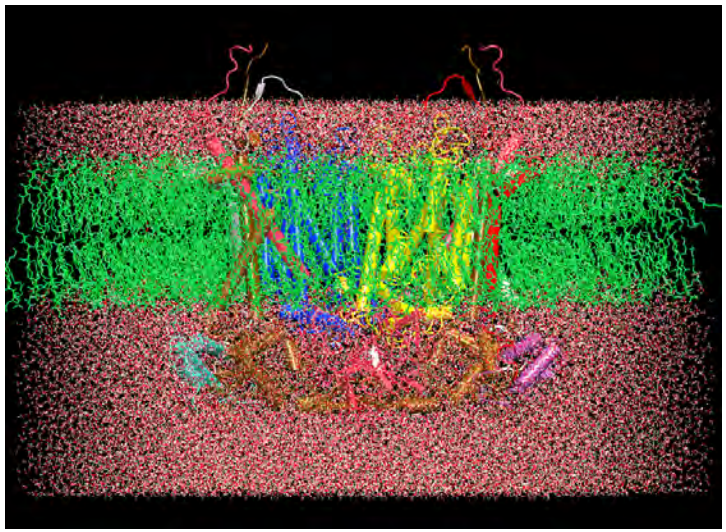
Steps in a Typical MD Simulation

- 1. Prepare molecule
  - Read in pdb and psf file
- 2. Minimization
  - Reconcile observed structure with force field used (T = 0)
- 3. Heating
  - Raise temperature of the system
- 4. Equilibration
  - Ensure system is stable
- 5. Dynamics
  - Simulate under desired conditions (NVE, NpT, etc)
  - Collect your data
- 6. Analysis
  - Evaluate observables (macroscopic level properties)
  - Or relate to single molecule experiments

Preparing Your System for MD
Solvation

Biological activity is the result of interactions between molecules and occurs at the interfaces between molecules (protein-protein, protein-DNA, protein-solvent, DNA-solvent, etc).

Why model solvation?
• many biological processes occur in aqueous solution
• solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc



How to model solvation?
• explicit treatment: solvent molecules are added to the molecular system
• implicit treatment: solvent is modeled as a continuum dielectric

Classical Molecular Dynamics

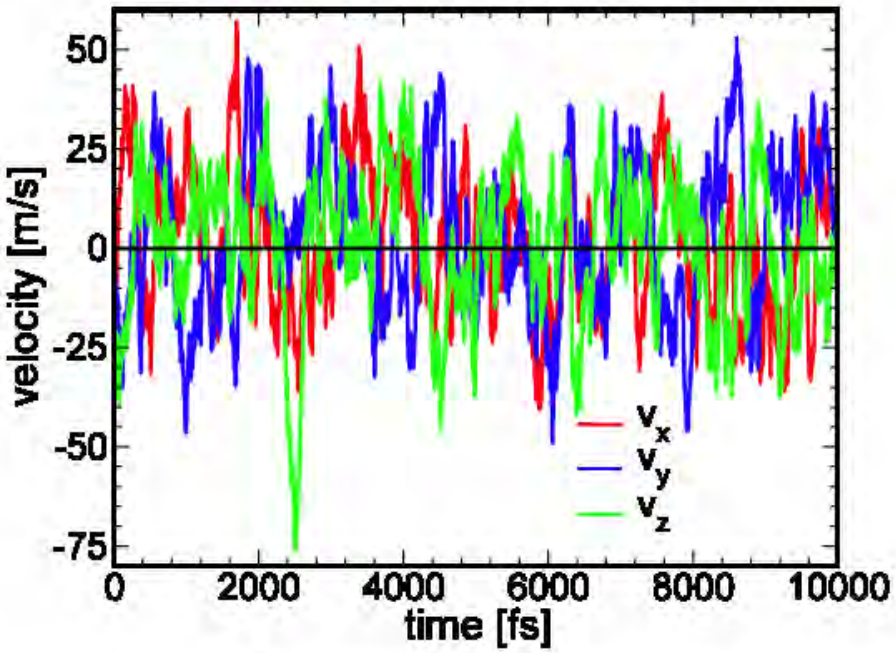
$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t$$

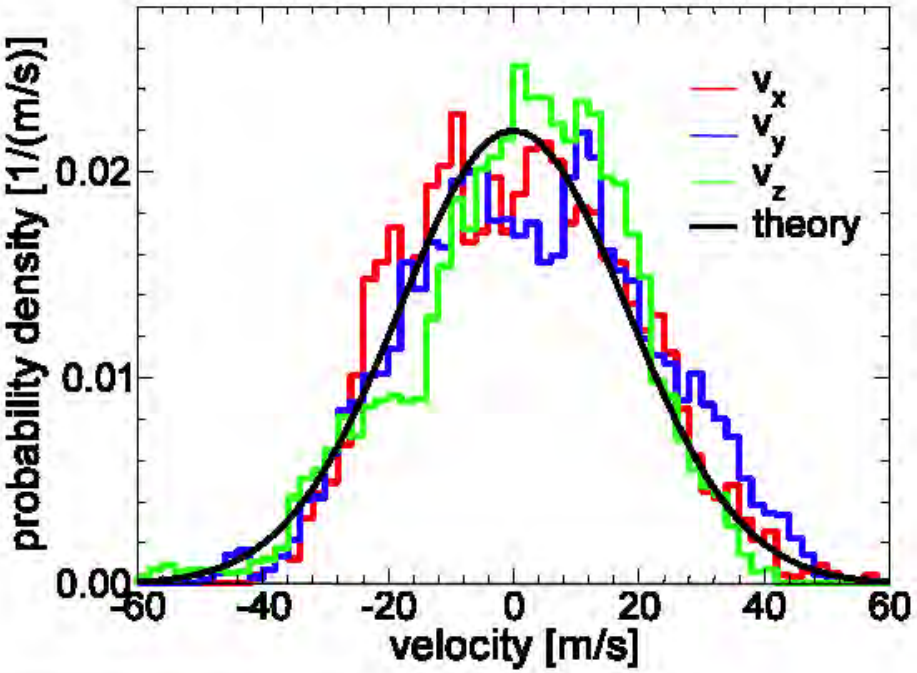
$$\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \boldsymbol{a}(t)\delta t$$

$$\boldsymbol{a}(t) = \boldsymbol{F}(t)/m$$

$$\boldsymbol{F} = -\frac{d}{d\boldsymbol{r}}U(\boldsymbol{r})$$

Maxwell Distribution of Atomic Velocities





$$p(v_\sigma) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left[-\frac{m v_\sigma^2}{2 k_B T}\right]$$

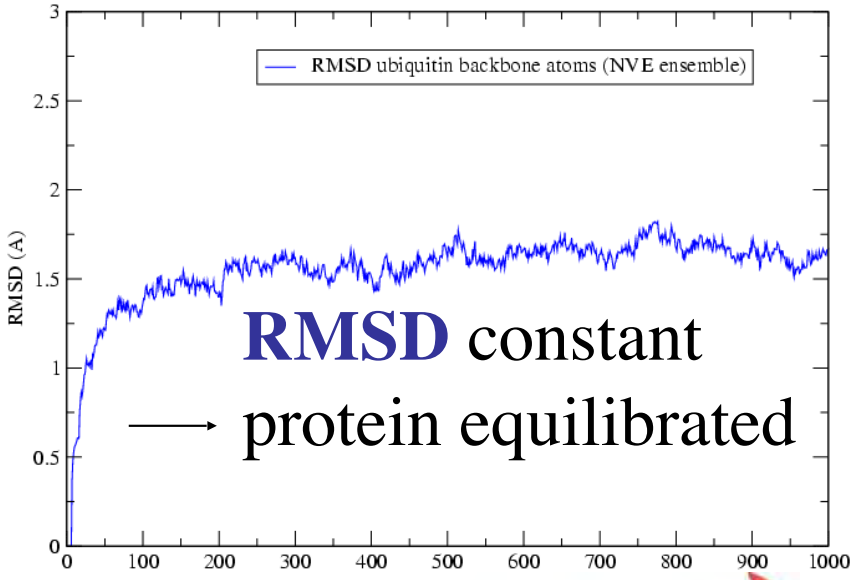
$$\sigma = x,\ y,\ z$$

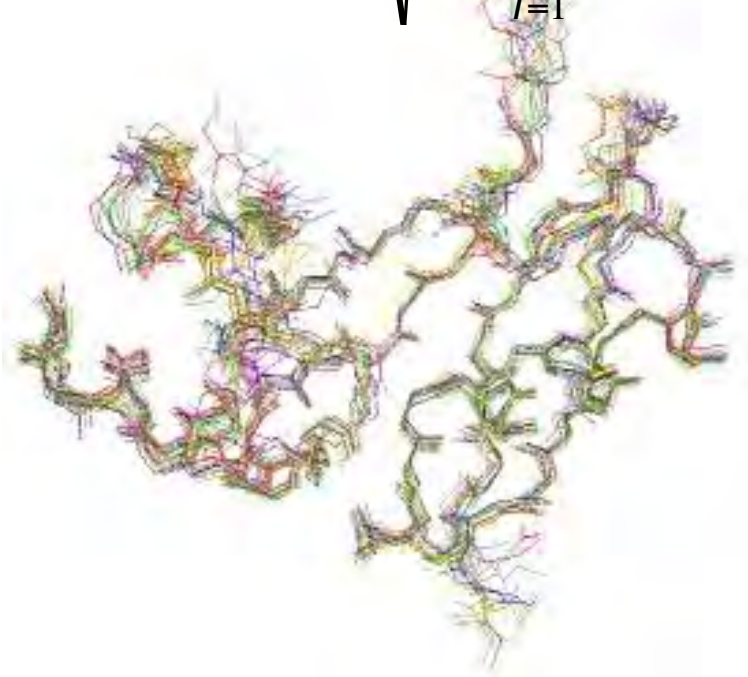
Equilibrium Properties of Proteins

Ubiquitin

Root Mean Squared Deviation: measure for equilibration and protein flexibility

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(R_i(t) - R_i(0)\right)^2}$$

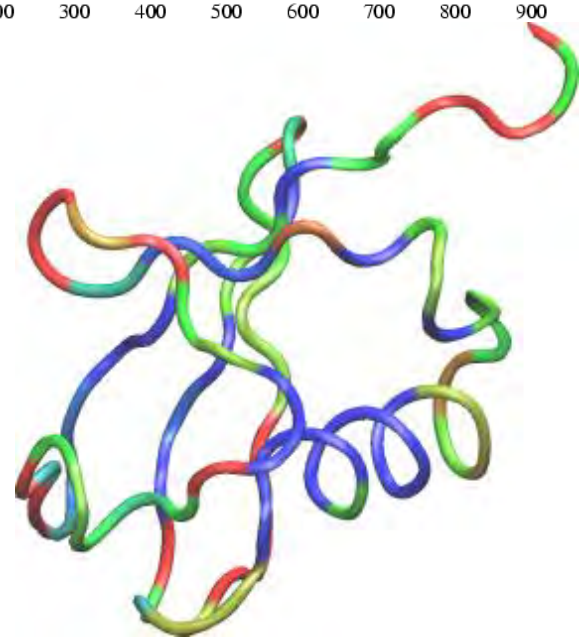




NMR structures

aligned together to see flexibility

Protein sequence exhibits characteristic permanent flexibility!

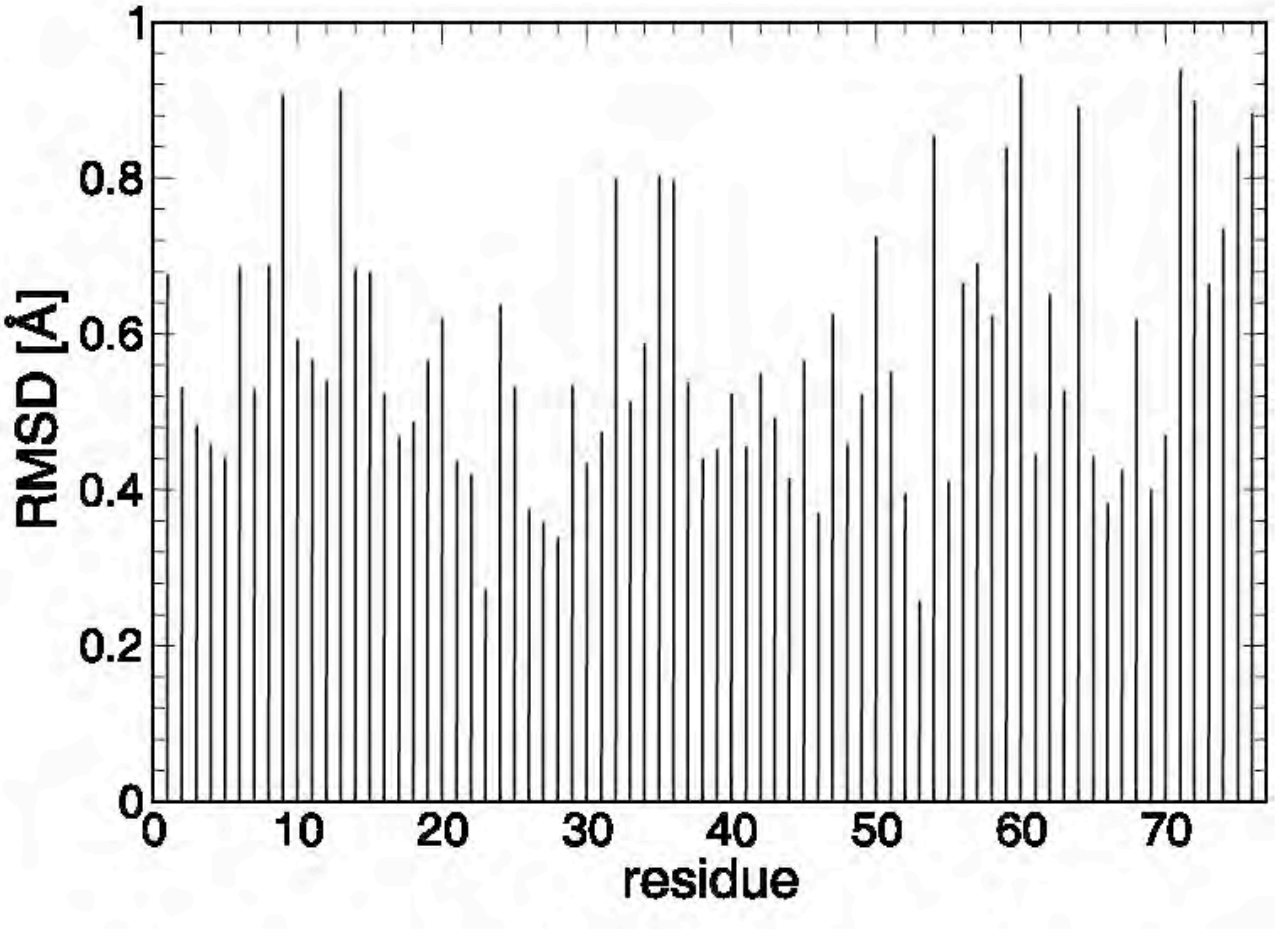


MD simulation

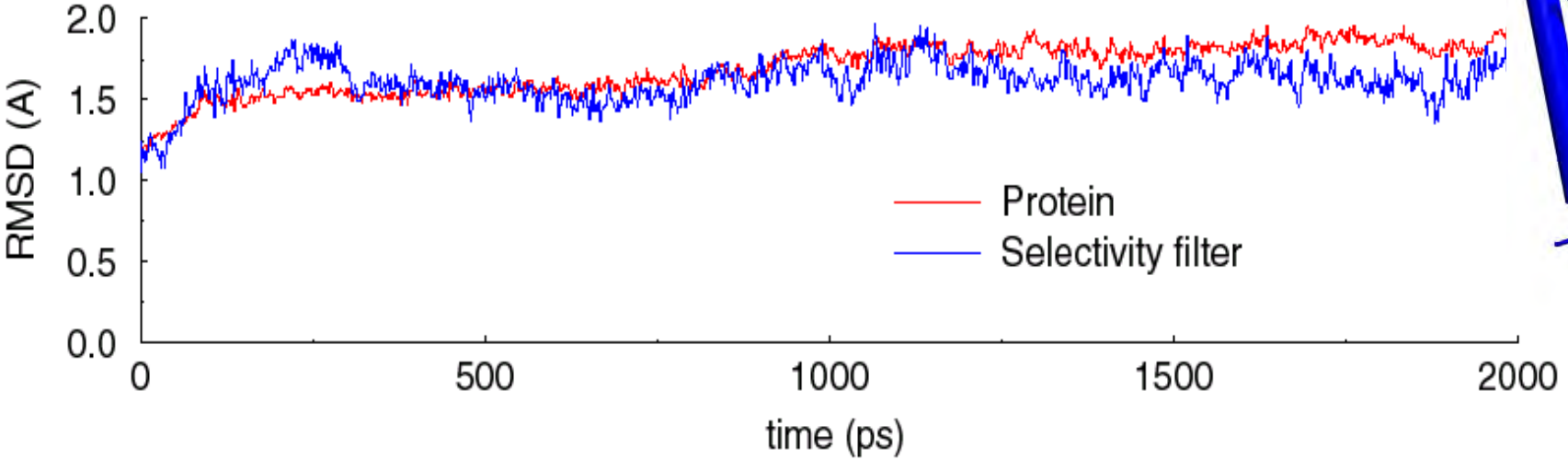
The color represents mobility of the protein through simulation (red = more flexible)

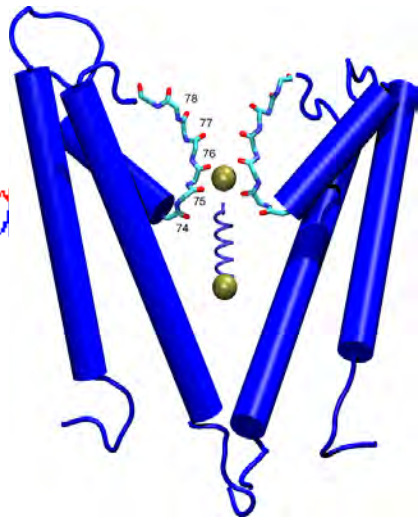
Thermal Motion of Ubiquitin from MD

RMSD values per residue

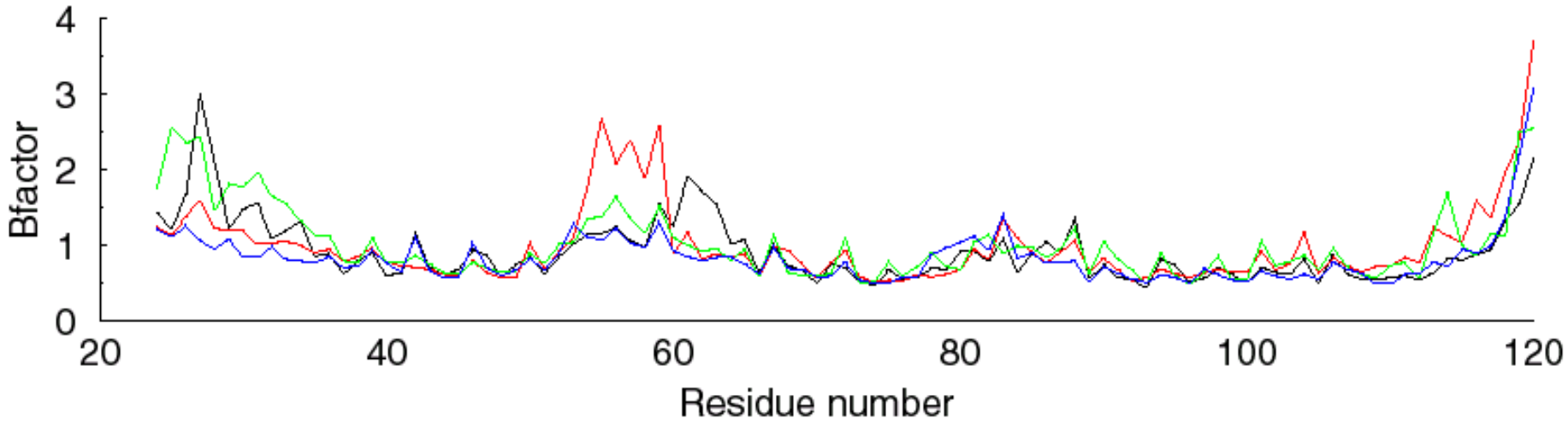


MD Results





RMS deviations for the KcsA protein and its selectivity filer indicate that the protein is stable during the simulation with the selectivity filter the most stable part of the system.



Temperature factors for individual residues in the four monomers of the KcsA channel protein indicate that the most flexible parts of the protein are the N and C terminal ends, residues 52-60 and residues 84-90. Residues 74-80 in the selectivity filter have low temperature factors and are very stable during the simulation.