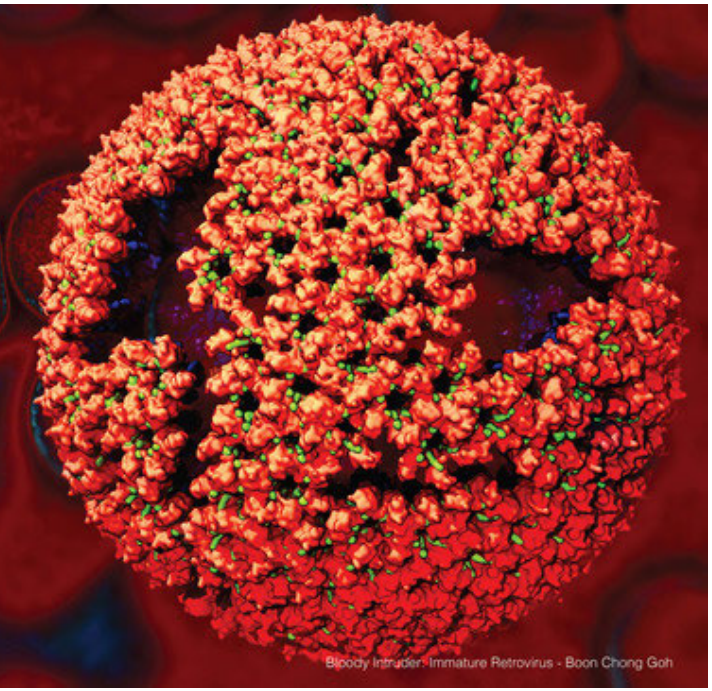
Hands-on Workshop on Computational Biophysics

NIH Center for Macromolecular Modeling and Bioinformatics



April 2017

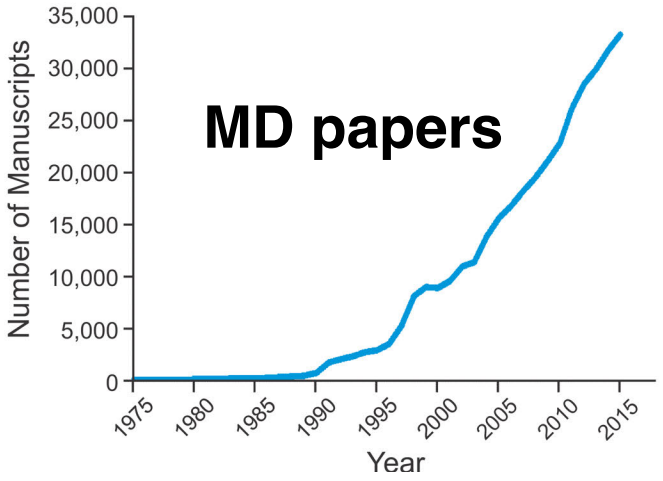
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign
Urbana, IL

Emad Tajkhorshid
NIH Center for Macromolecular Modeling and Bioinformatics

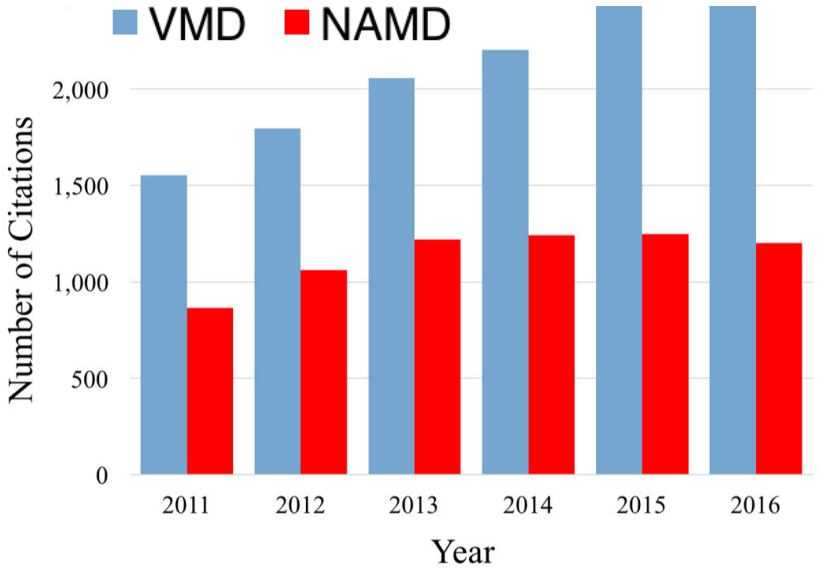
Computational Structural Biology and Molecular Biophysics
Department of Biochemistry
Center for Biophysics and Quantitative Biology

NIH Center for Macromolecular Modeling and Bioinformatics

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

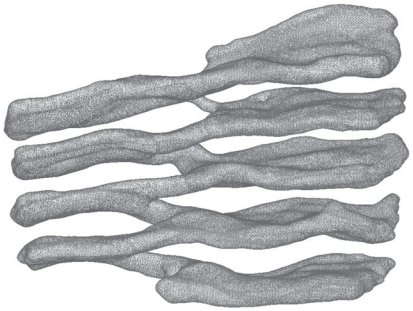


Serving a Large and Fast Growing Community

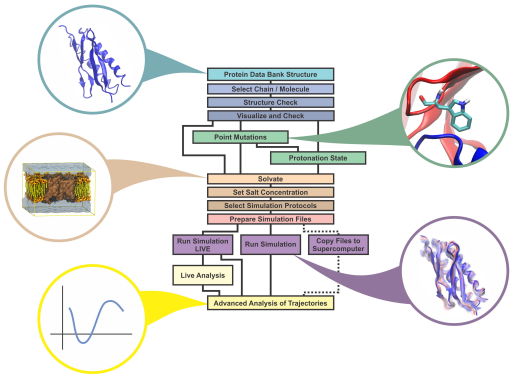
- Deploying Center's flagship programs NAMD and VMD on all major computational platforms from commodity computers to supercomputers

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

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

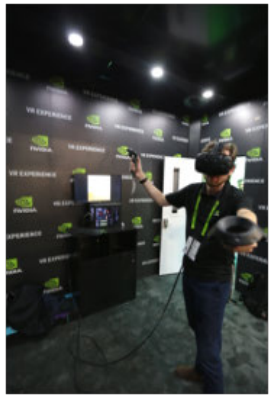


- Enhanced software accessibility
    - QwikMD, interactive MDFF, ffTk, simulation in the Cloud, remote visualization

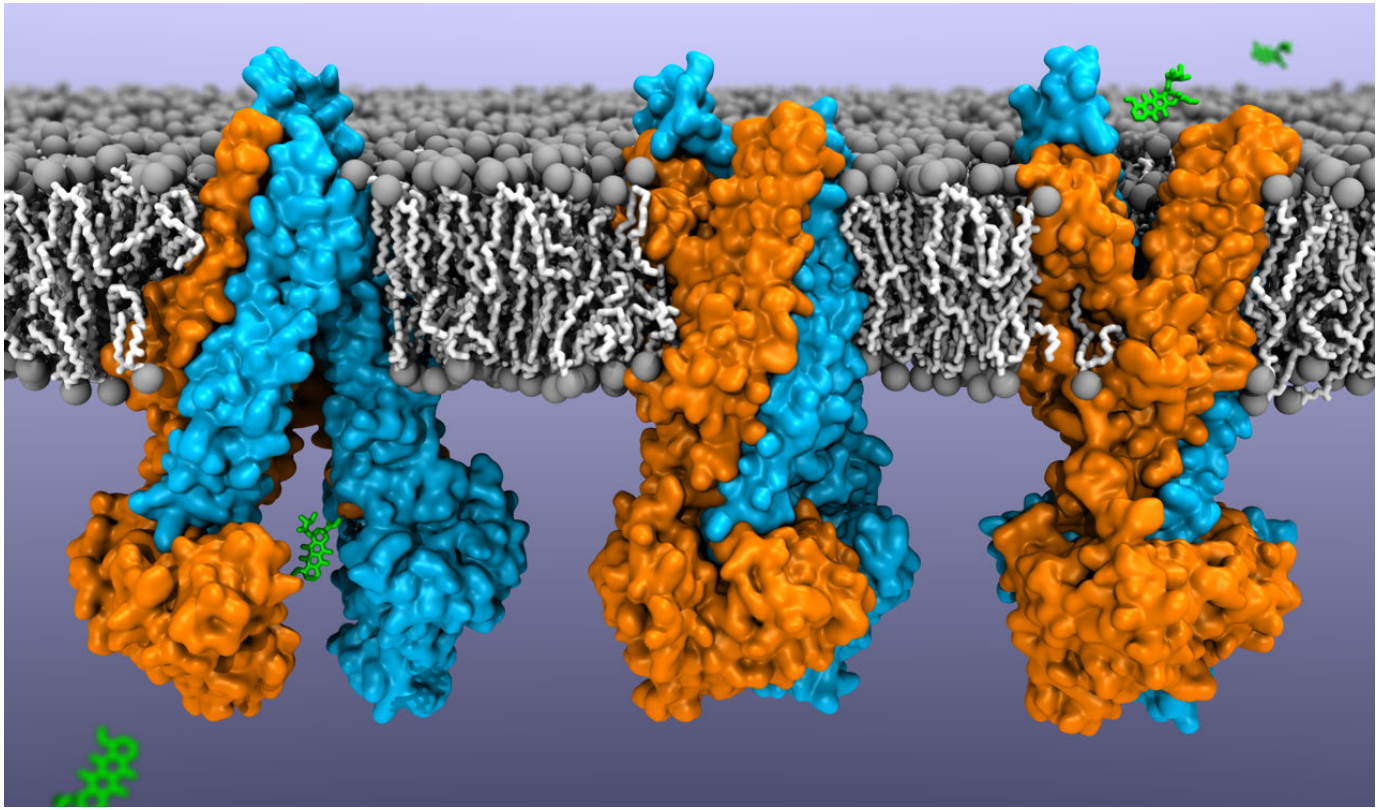


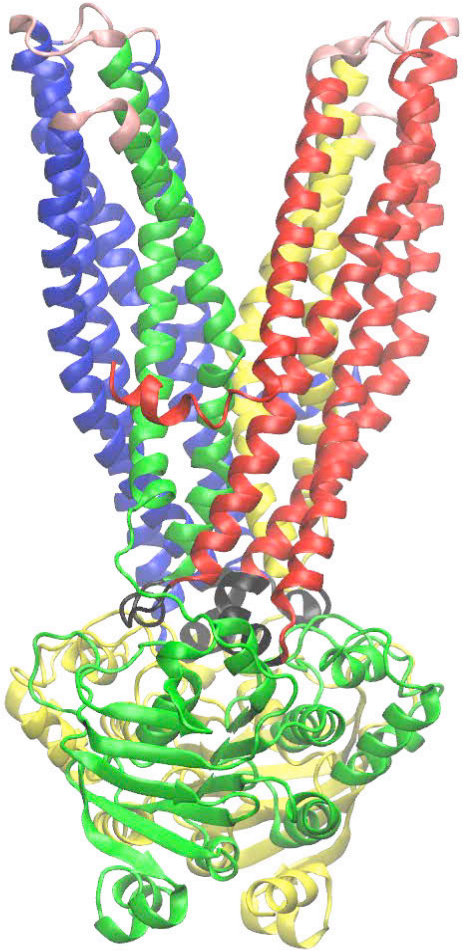
Exploiting State of the Art Hardware Technology

- 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





Emad Tajkhorshid

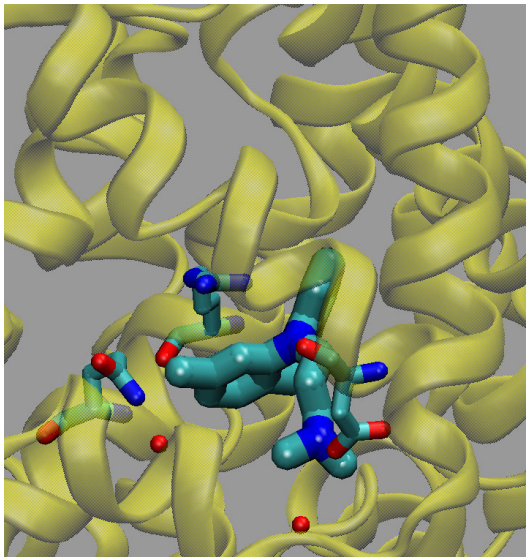
NIH Center for Macromolecular Modeling and Bioinformatics

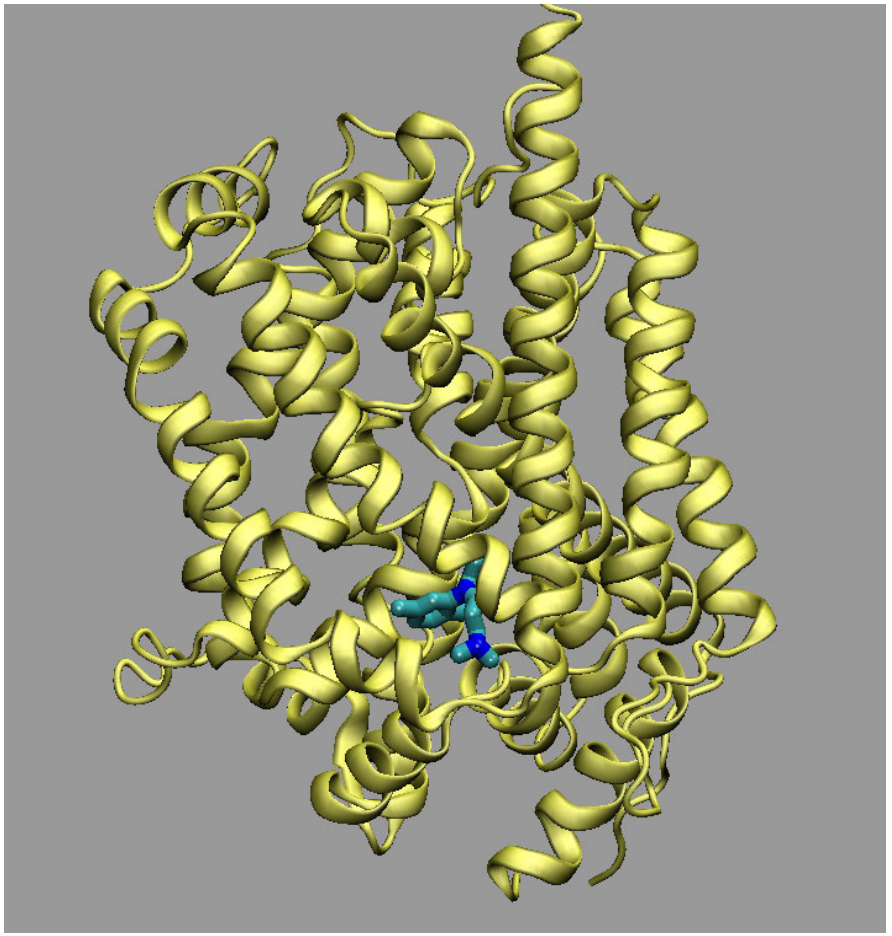
Computational Structural Biology and Molecular Biophysics
Department of Biochemistry
Center for Biophysics and Quantitative Biology
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

Structure / Dynamics @ nanoscale

Why Structural Biology at Nanoscale?

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

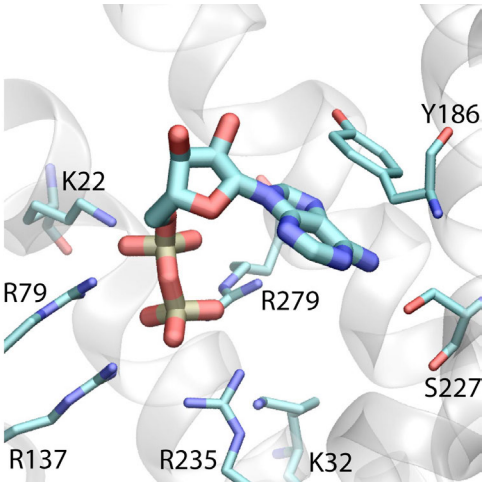


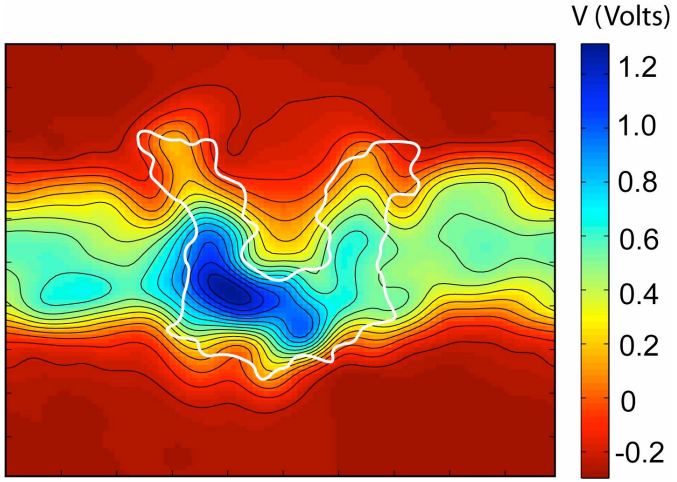


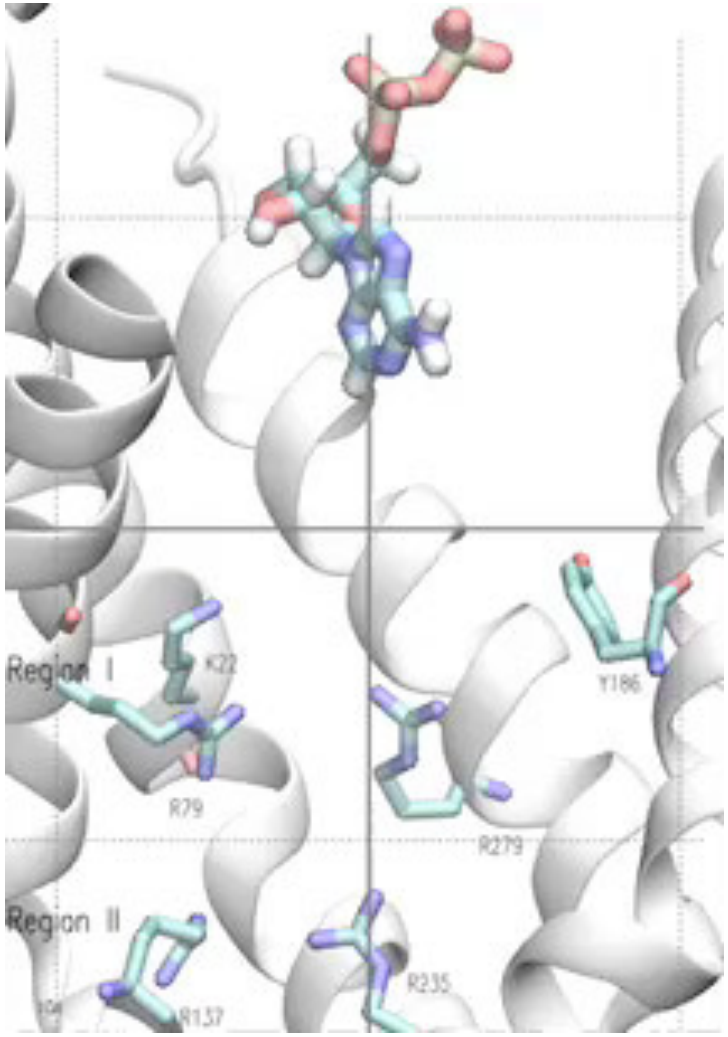
Antidepressant binding site in a neurotransmitter transporter.
Nature 448: 952-956 (2007)

Why Structural Biology at Nanoscale?

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



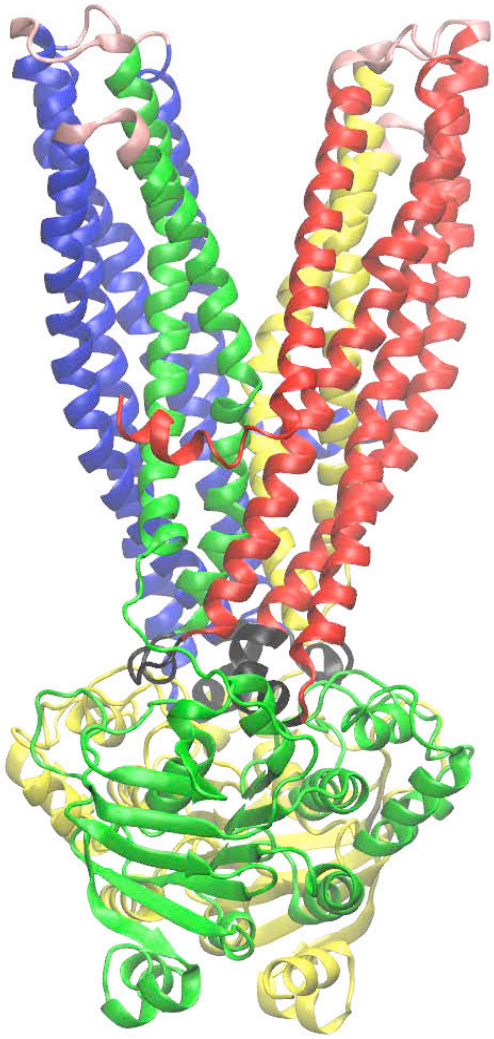




Binding of a small molecule to a binding site
Y. Wang & E.T. PNAS 2010

Why Structural Biology at Nanoscale?

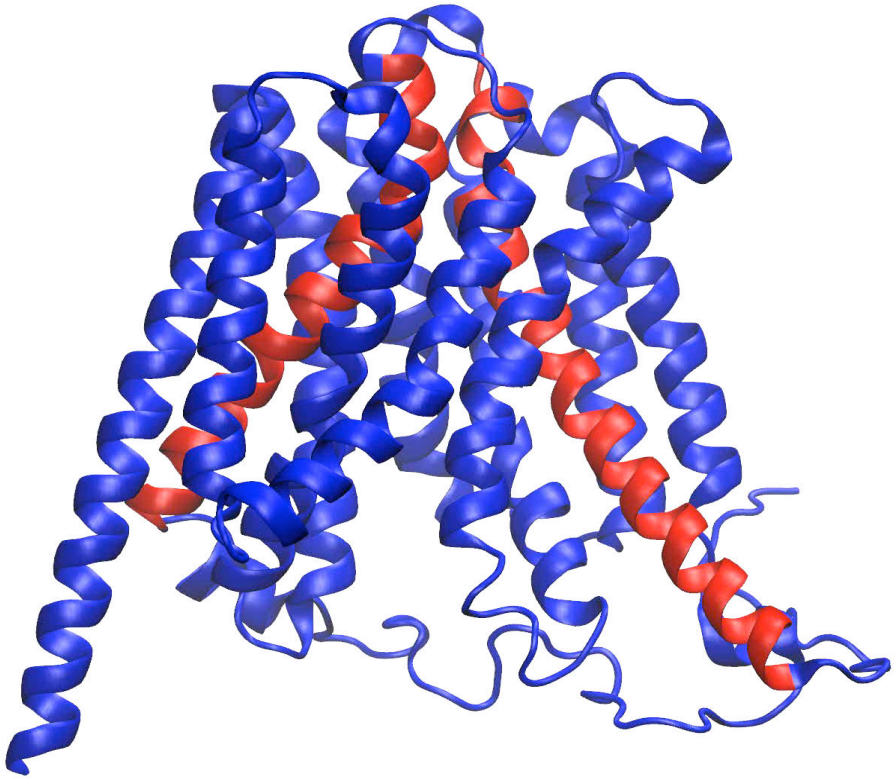
- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology

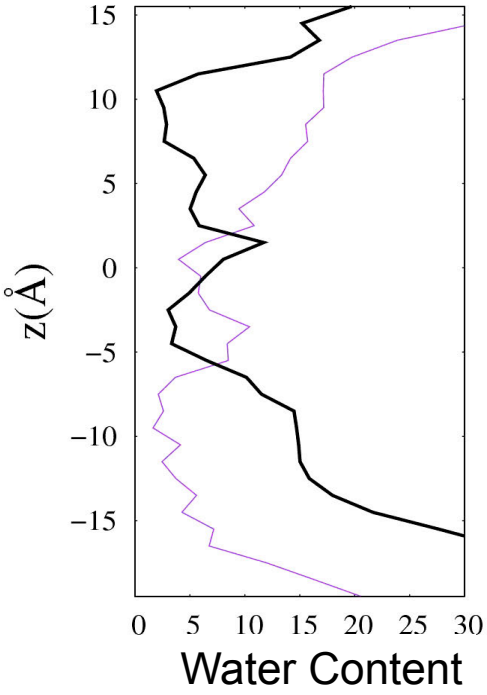


Structural changes underlying function
M. Moradi & E. T. PNAS 2013

Why Structural Biology at Nanoscale?

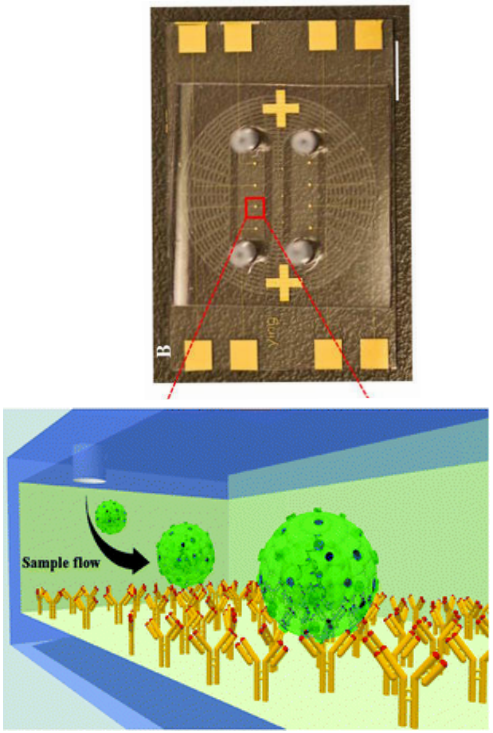
- ✦ Mechanisms in Molecular Biology
- ✦ Molecular Basis of Disease
- ✦ Drug Design
- ✦ Nano-biotechnology





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

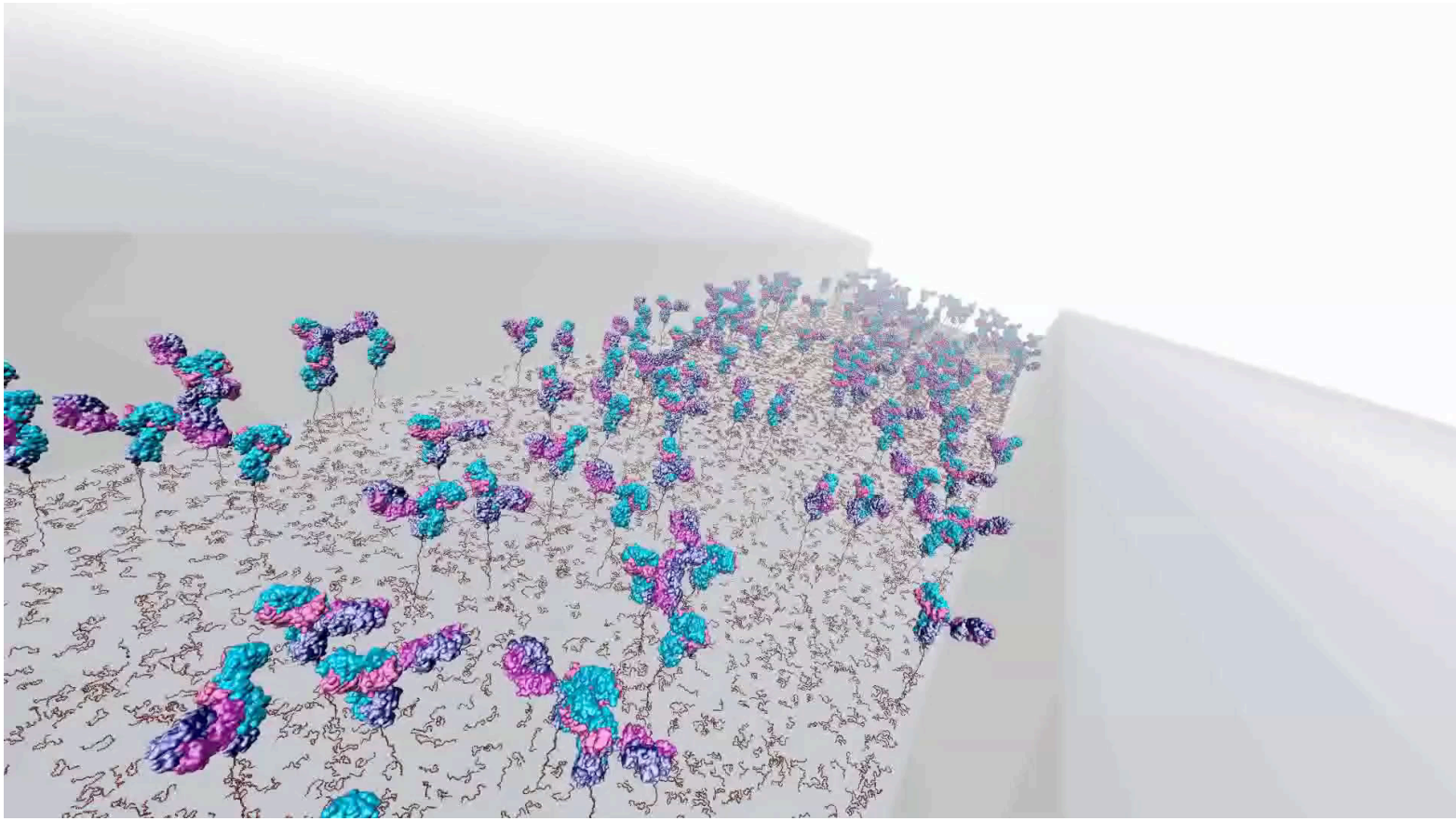
Nano-biotechnology
Microfluidic Sensing Devices



HIV subtype identification

Lab Chip 2012

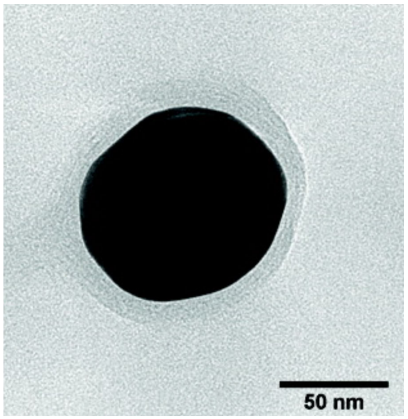
Functionalized nanosurface with antibodies



Created by **nanoBIO Node** tools

Nano-biotechnology
Gold Nanoparticles as Delivery Vehicles

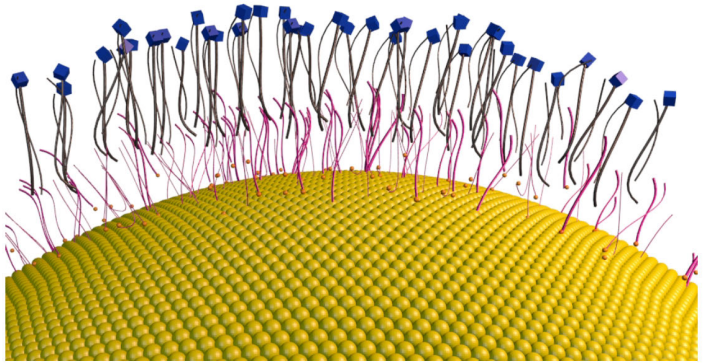
Transmission Electron Micrograph

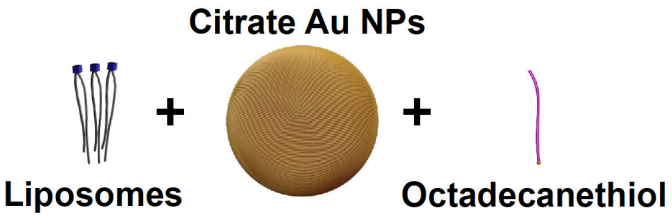


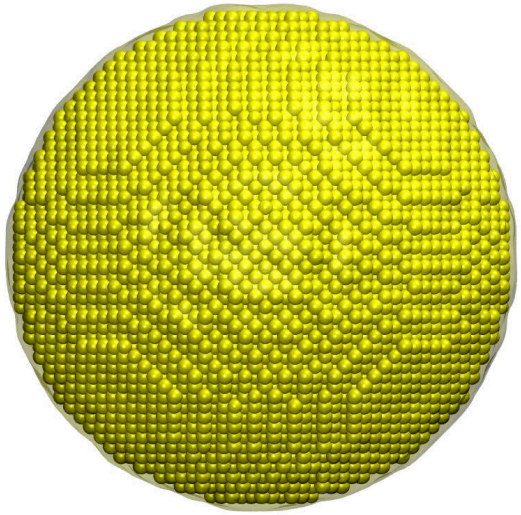
Yang, J. A.; Murphy, C. J. Langmuir 2012, 28, 5404–5416

Schematic model with no prediction power

Cartoon representation of lipid Au NPs







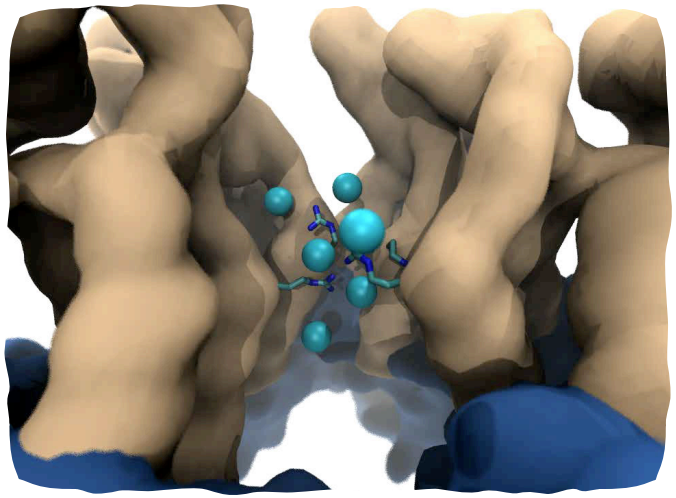
Experiment:
Murphy Lab

Modeling/Simulation:
Tajkhorshid Lab

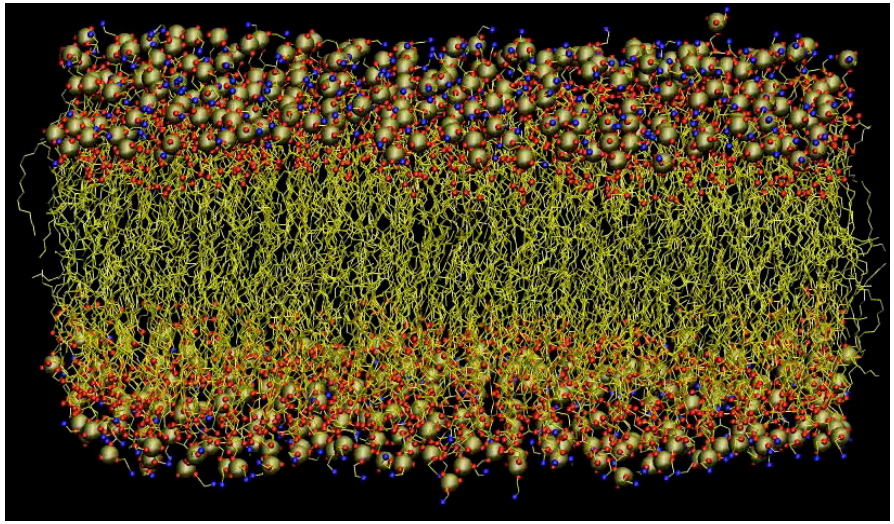
Applications of Computational Methodologies to Structural Biology

Simulation of the dynamics of the molecular system (MD)

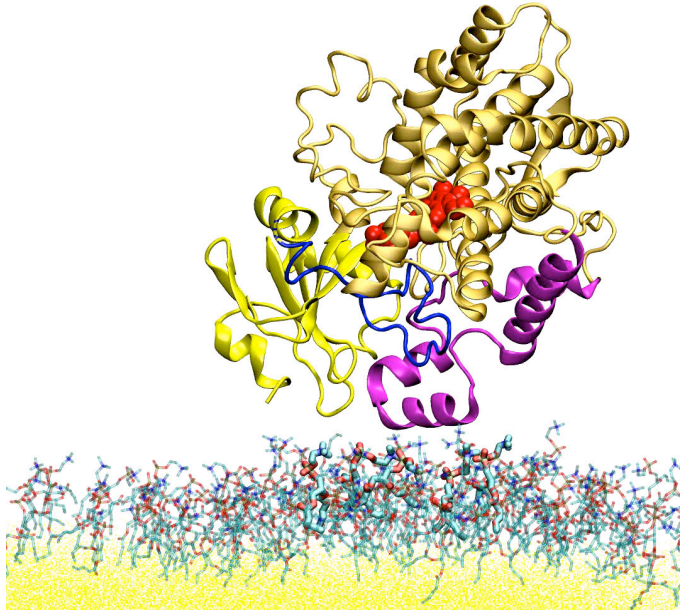
- Calculating ensemble-averaged properties of microscopic systems to compare to macroscopic measurements
- Providing a molecular basis for function
- Describing the molecular/structural changes underlying function
- …



Hydration at the interface of viral shell proteins

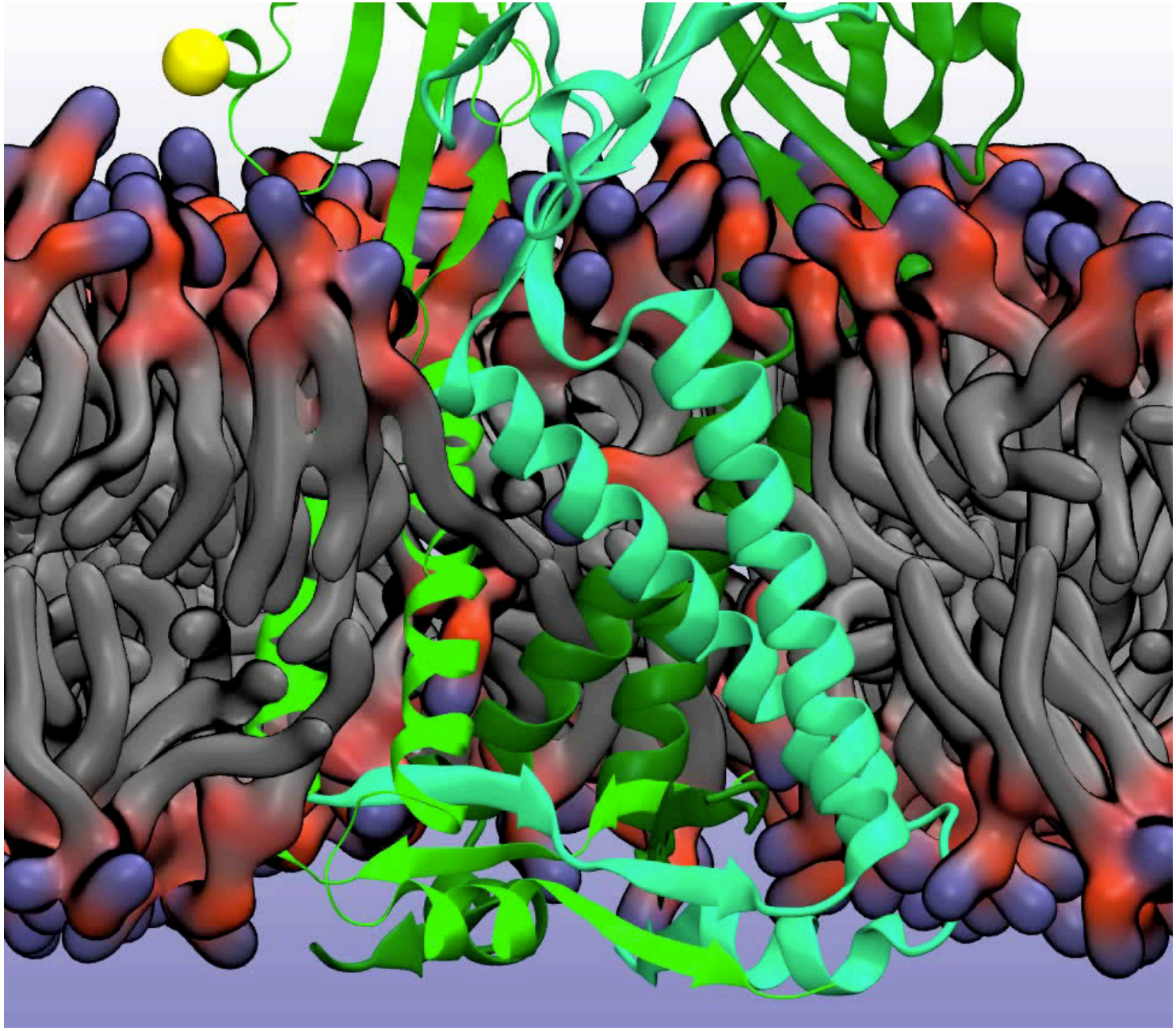


Thermal fluctuations of a phospholipid bilayer



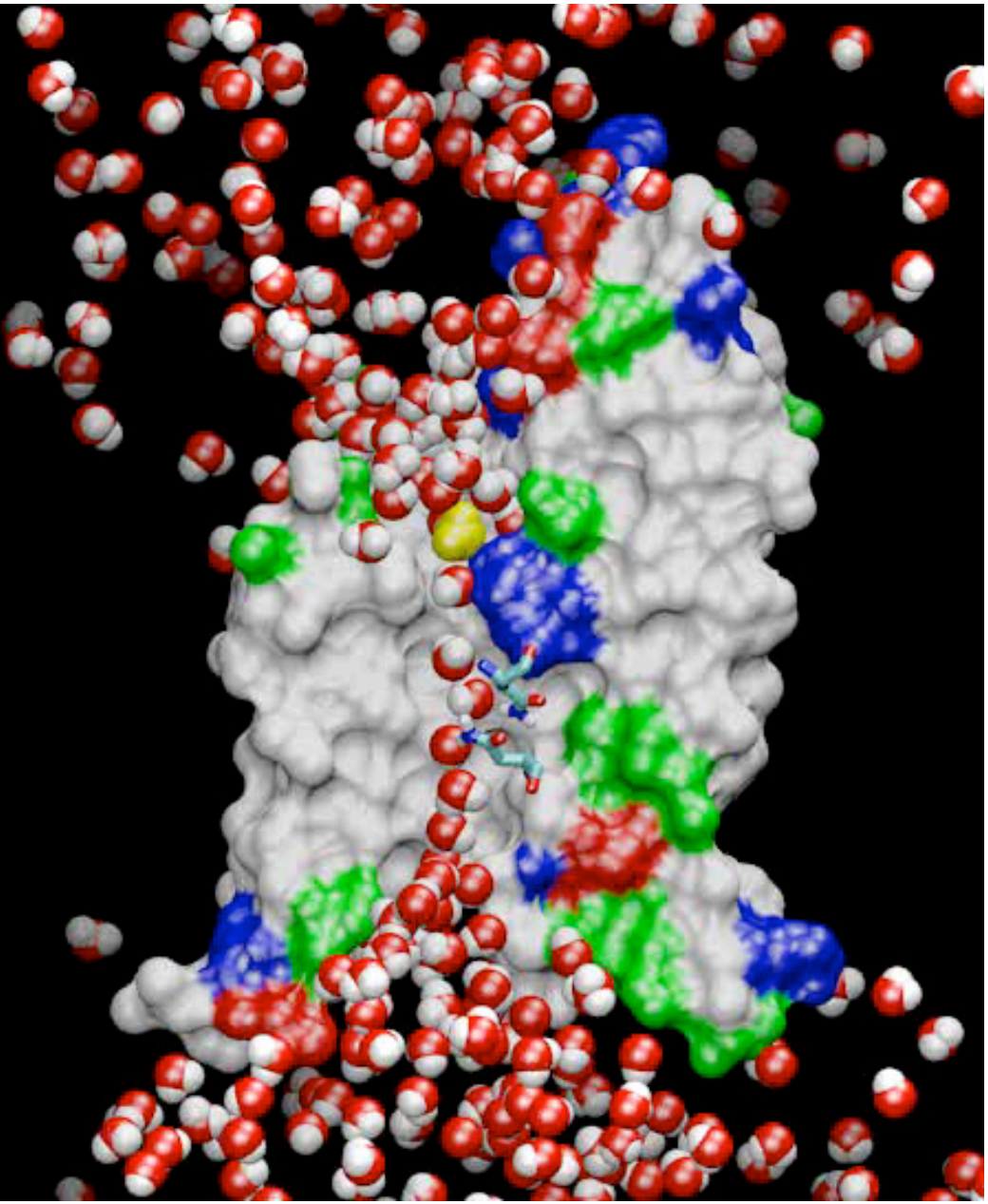
Membrane binding of a coagulation protein

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

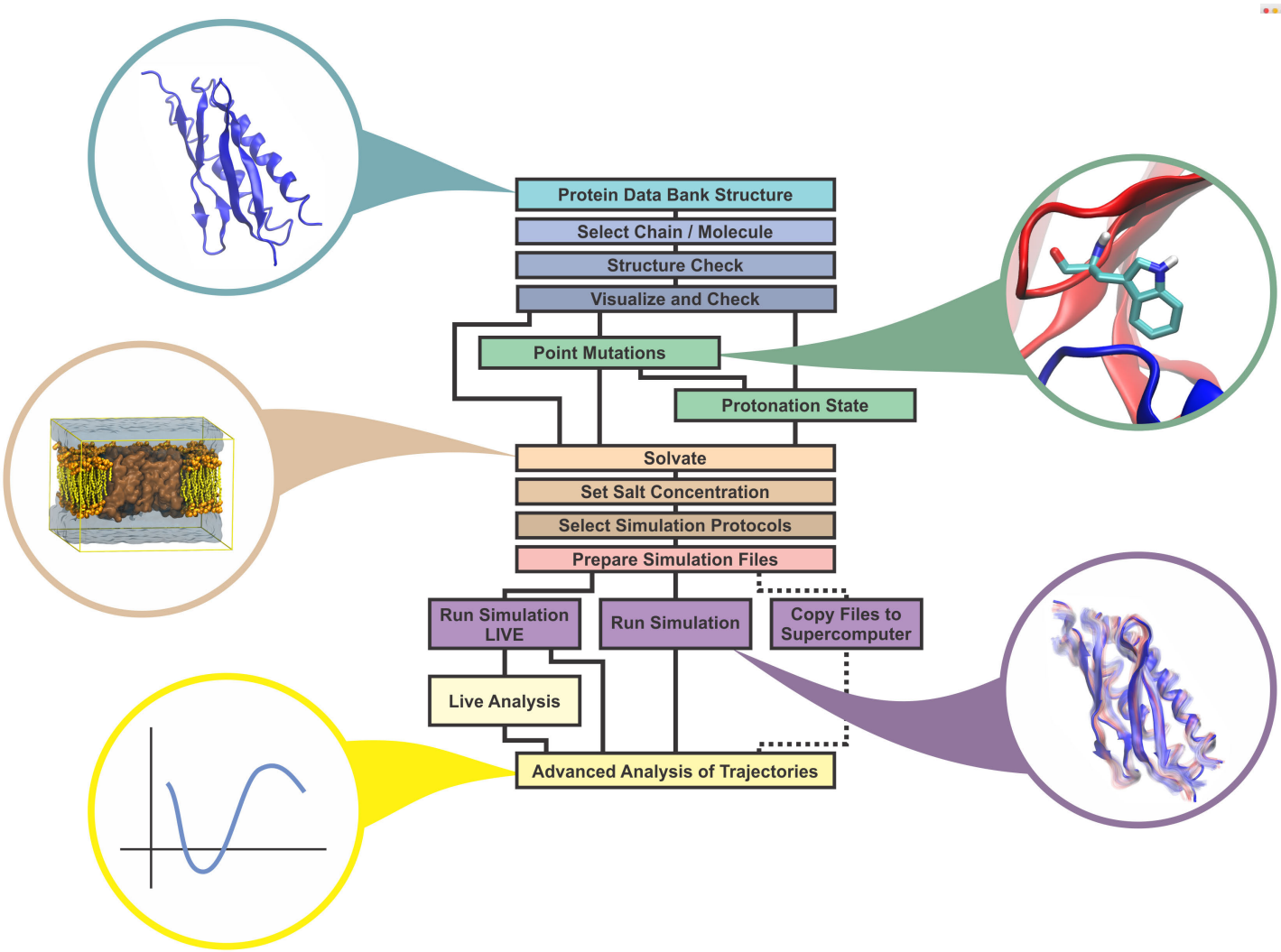
Major advantage:

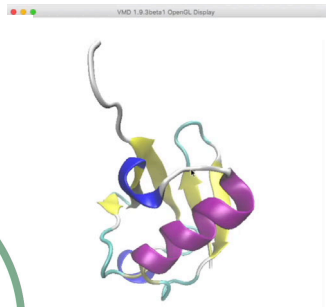
- Unparalleled spatial and temporal resolutions, simultaneously

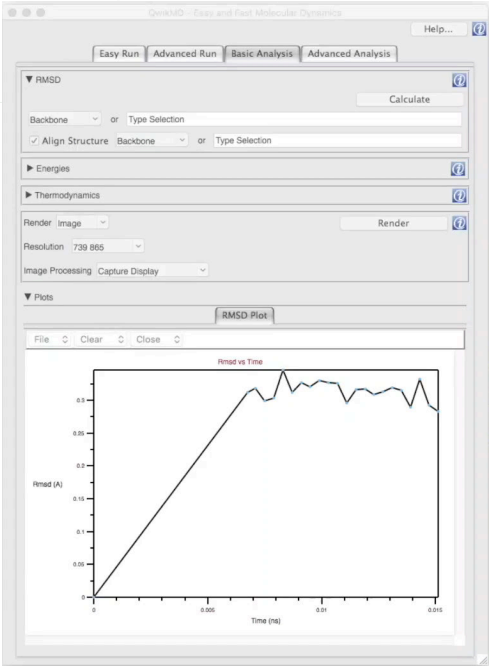
Steps in a Typical MD Simulation

- 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







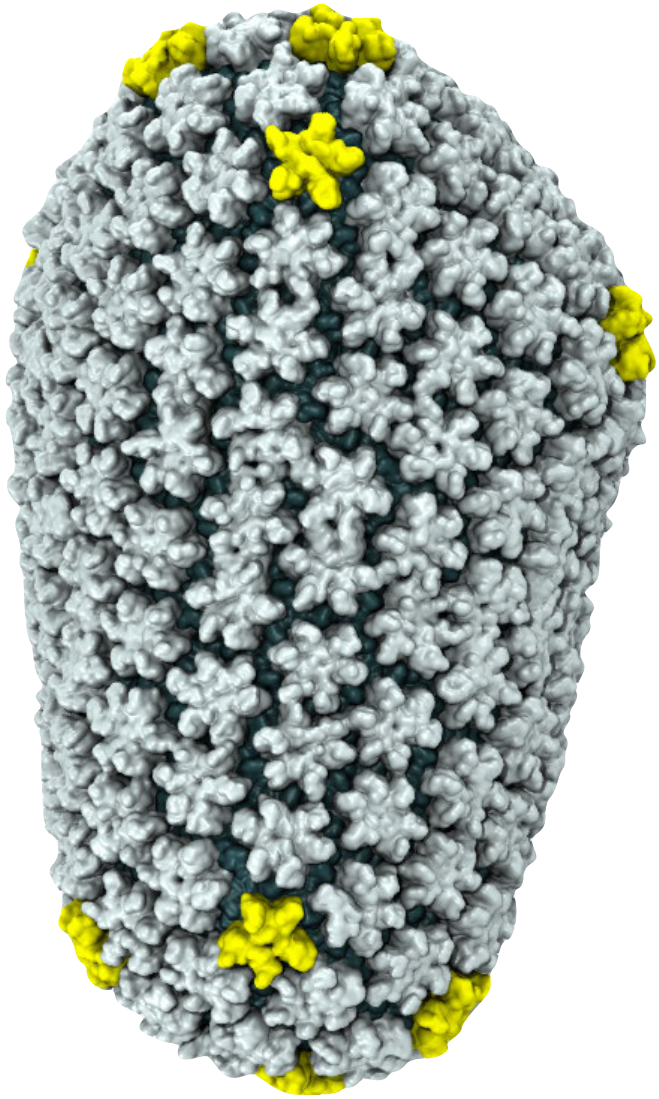
Ribeiro, J. V., …, Schulten, K.. QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. ***Sci. Rep***. 6, 26536; doi: 10.1038/srep26536 (**2016**)

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

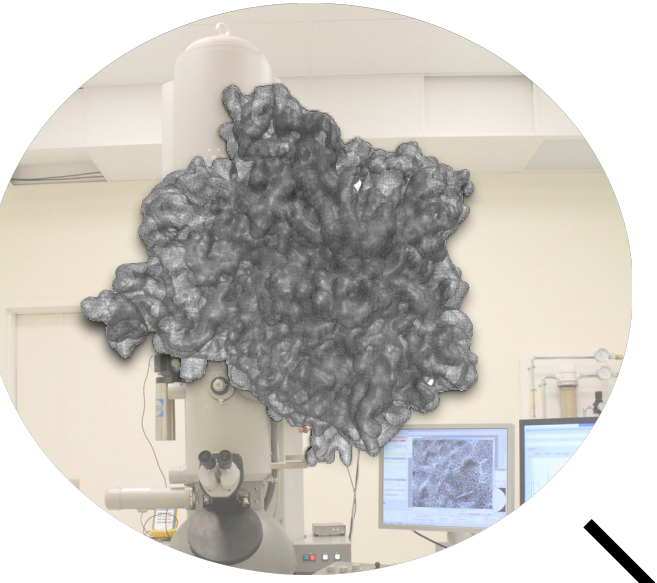
- NMR
- X-ray
- Electron Microscopy
- …



Structural model of HIV virus

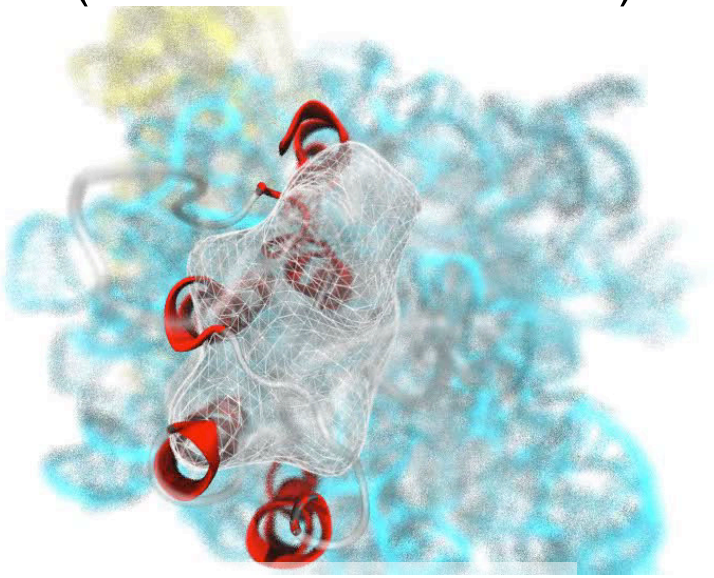
Molecular Dynamics Flexible Fitting (MDFF)

Electron Microscope



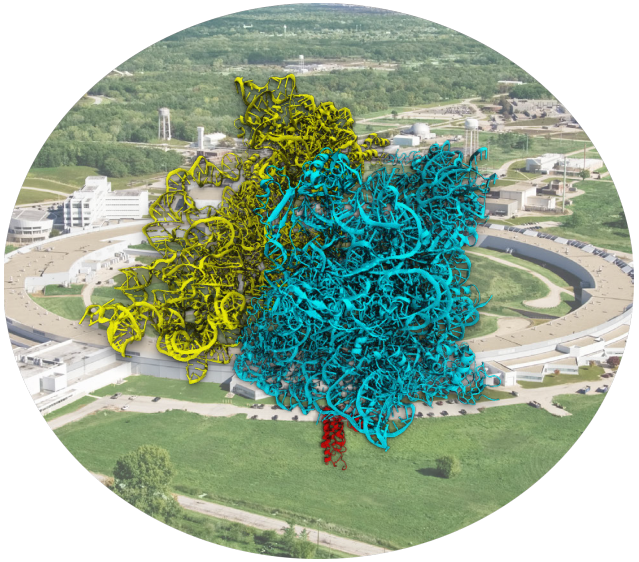
cryo-EM density map

(Ribosome-bound YidC)

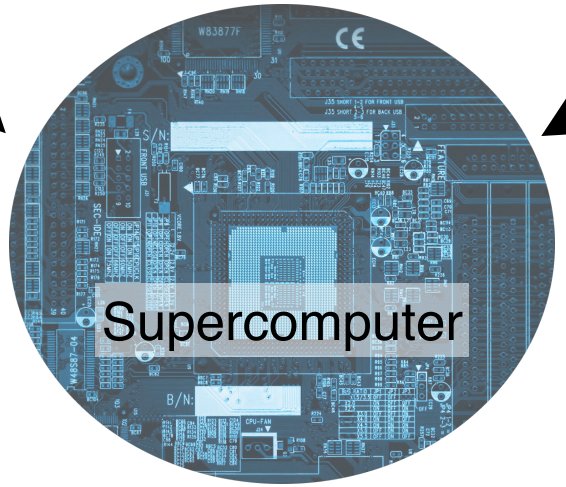


Match through MD

APS Synchrotron



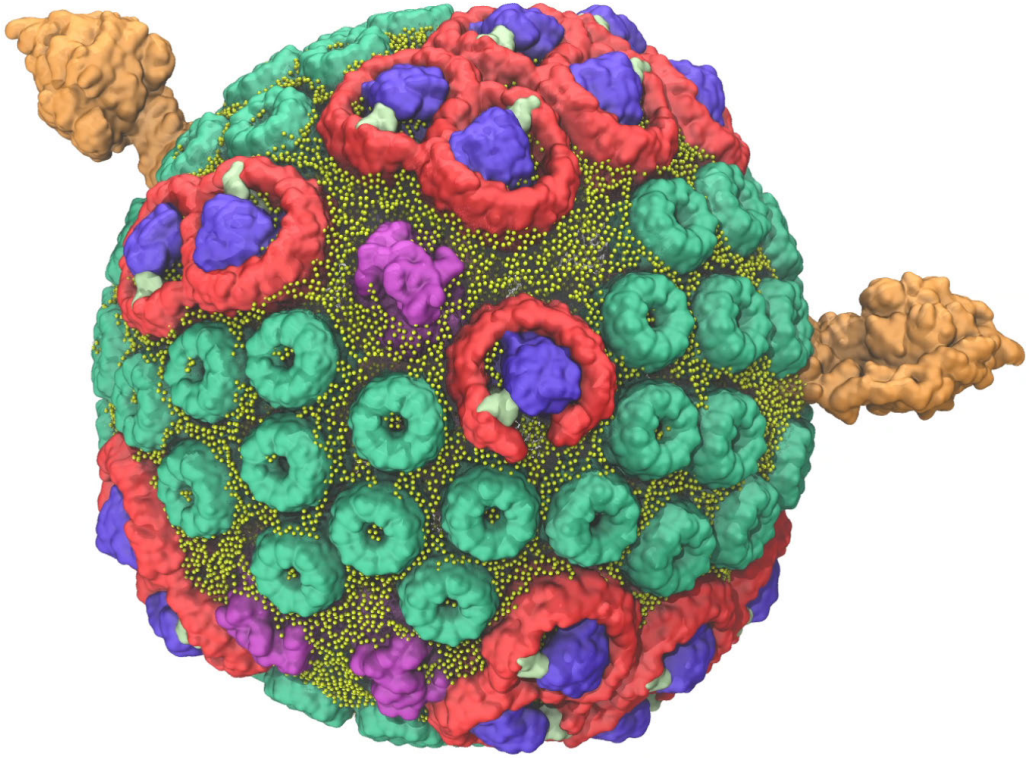
crystallographic structure



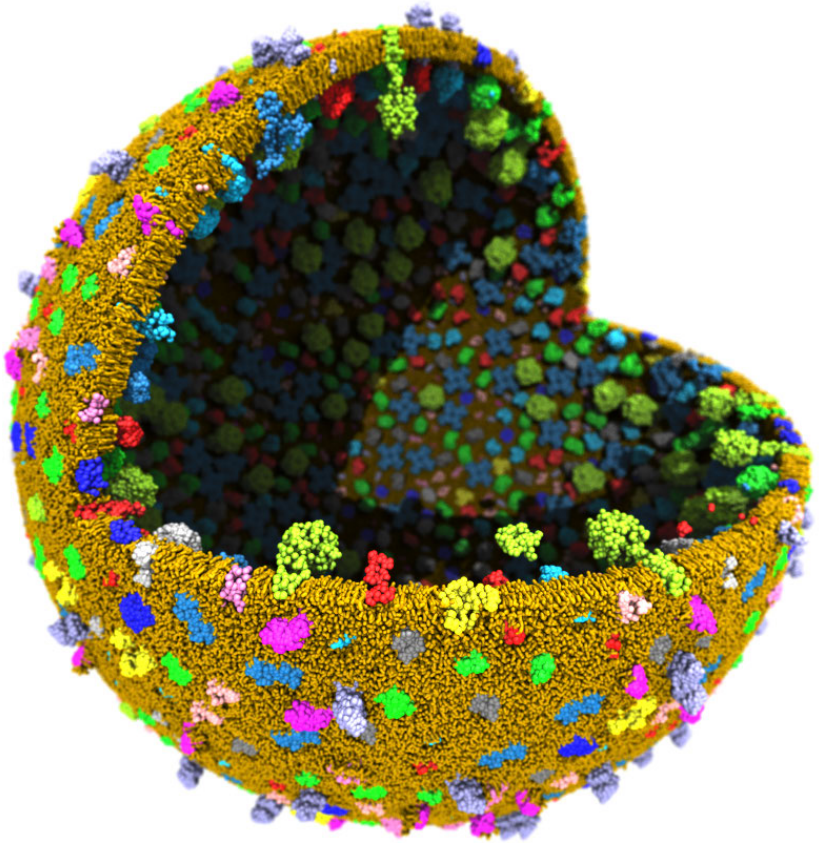
[1] Trabuco et al. *Structure* (2008) 16:673-683.
[2] Trabuco et al. *Methods* (2009) 49:174-180.

Applications of Computational Methodologies to Cell-Scale Structural Biology

Using simulations as a “structure-building” tool



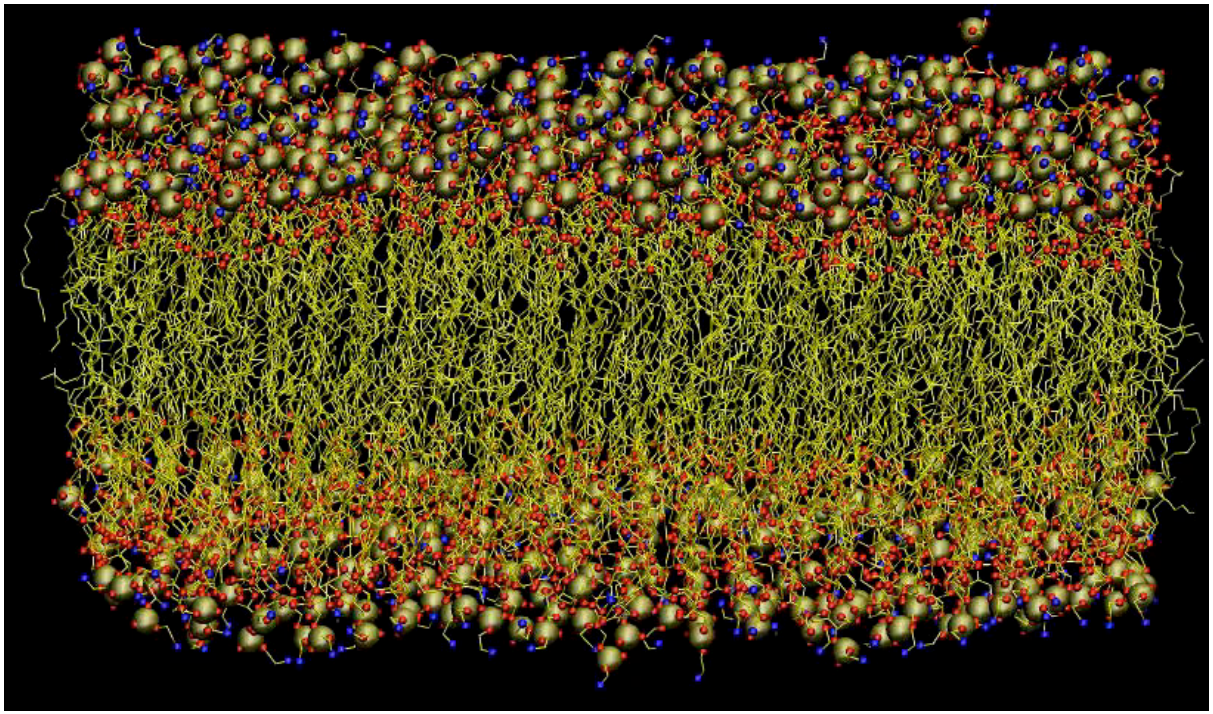
The most detailed model of a chromatophore

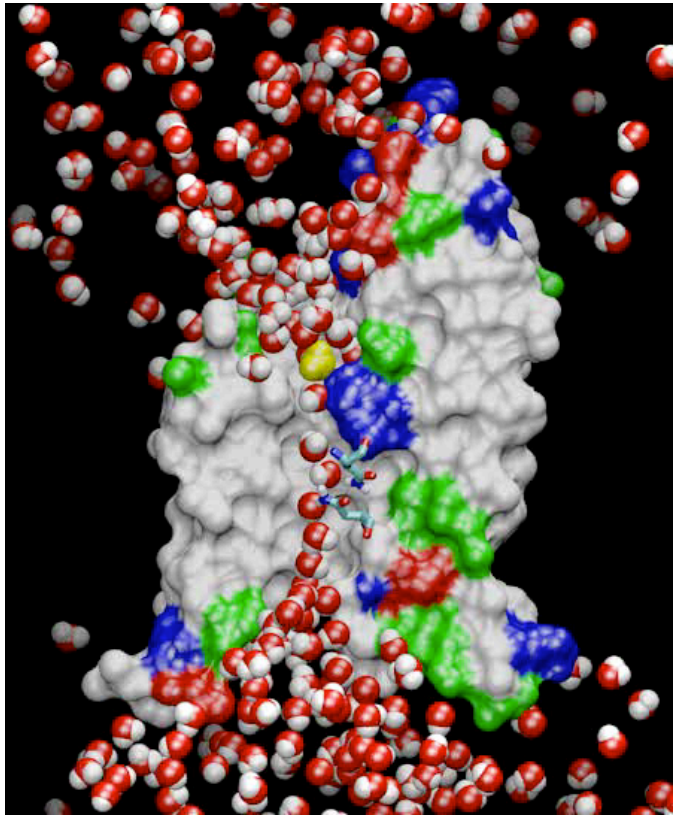


Computational model of a minimal cell envelope

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

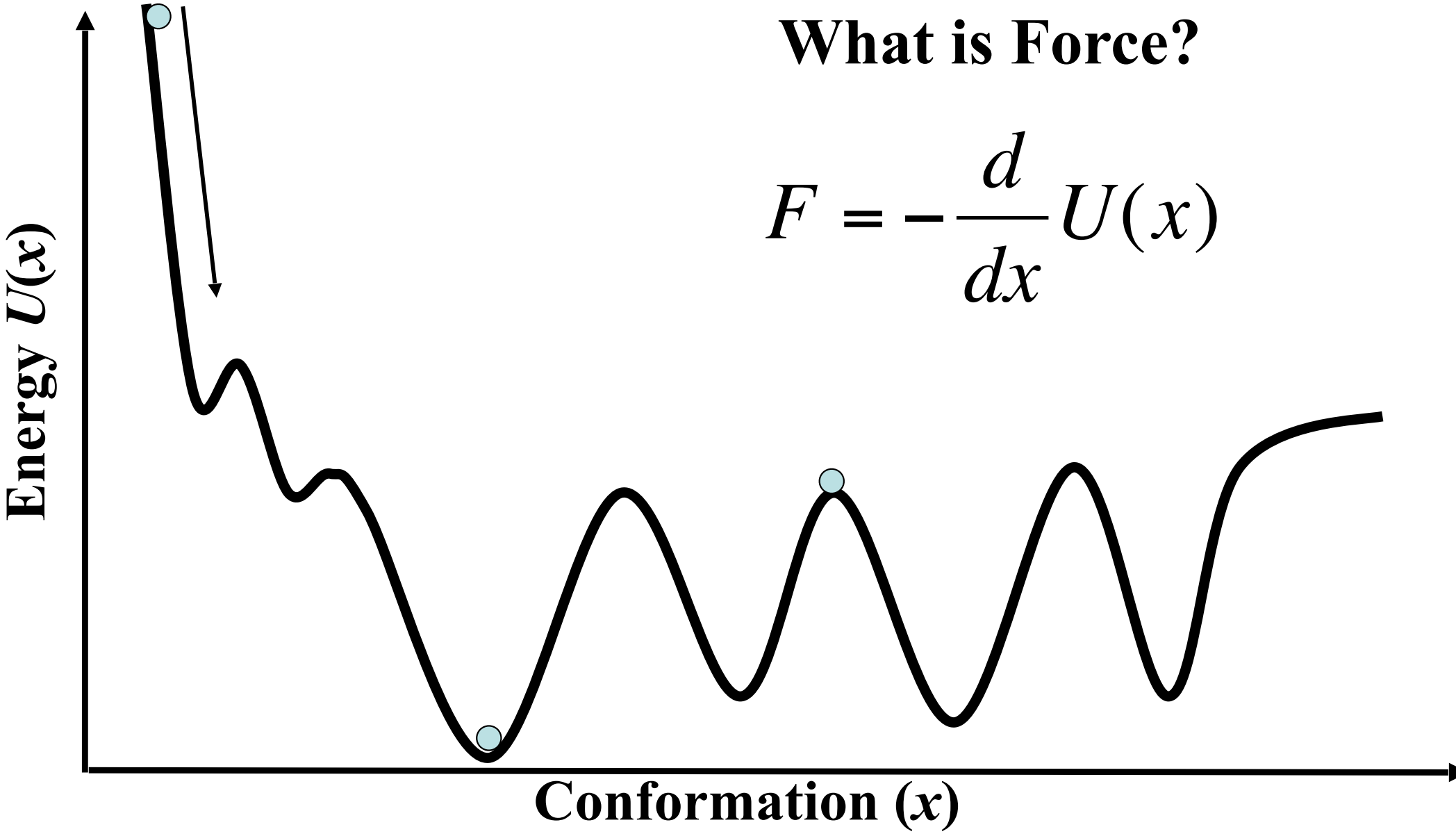
$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t$$

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

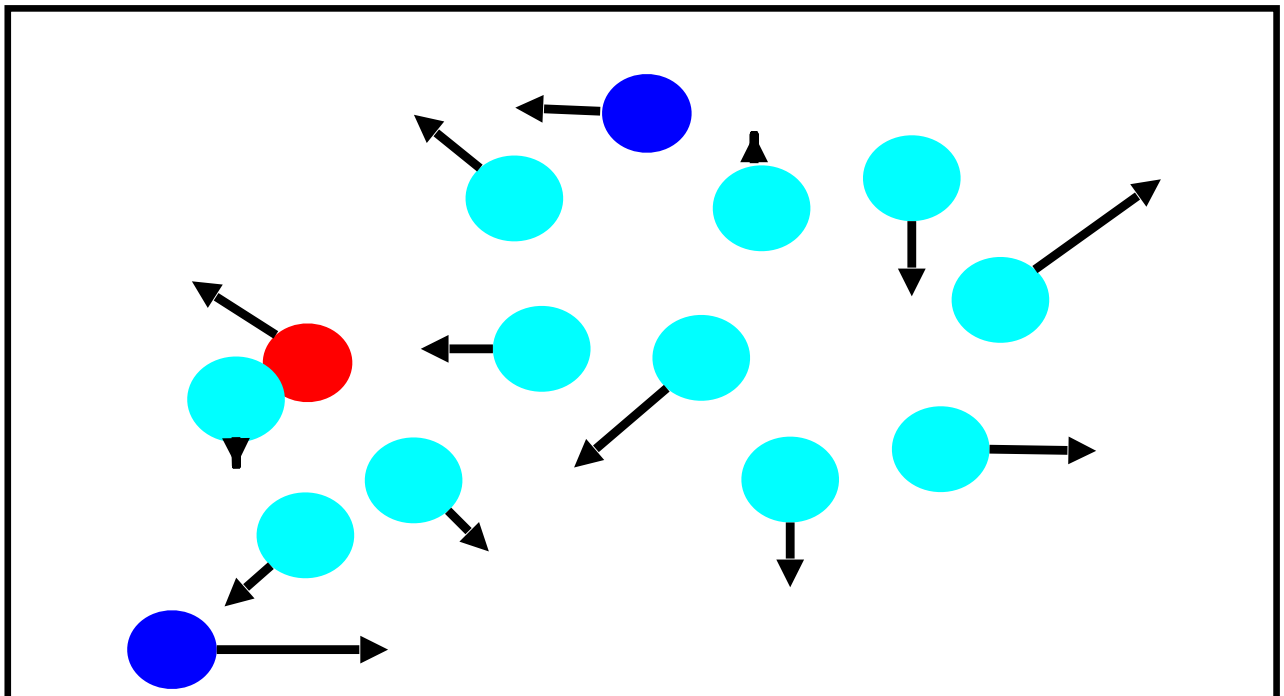
$$\boldsymbol{a}(t) = \boldsymbol{F}(t)/m$$

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

Potential Energy (hyper)Surface



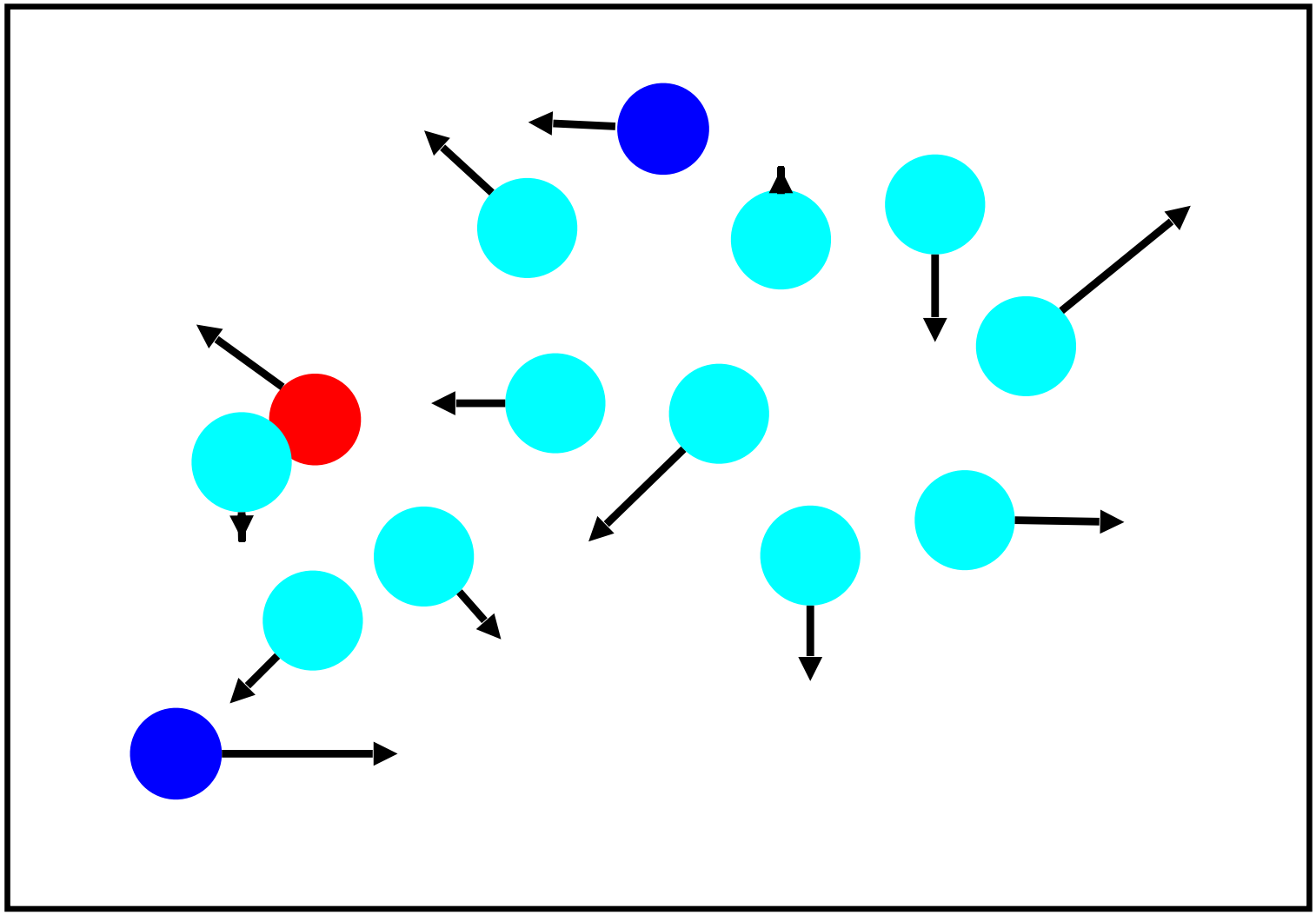
Classical Molecular Dynamics



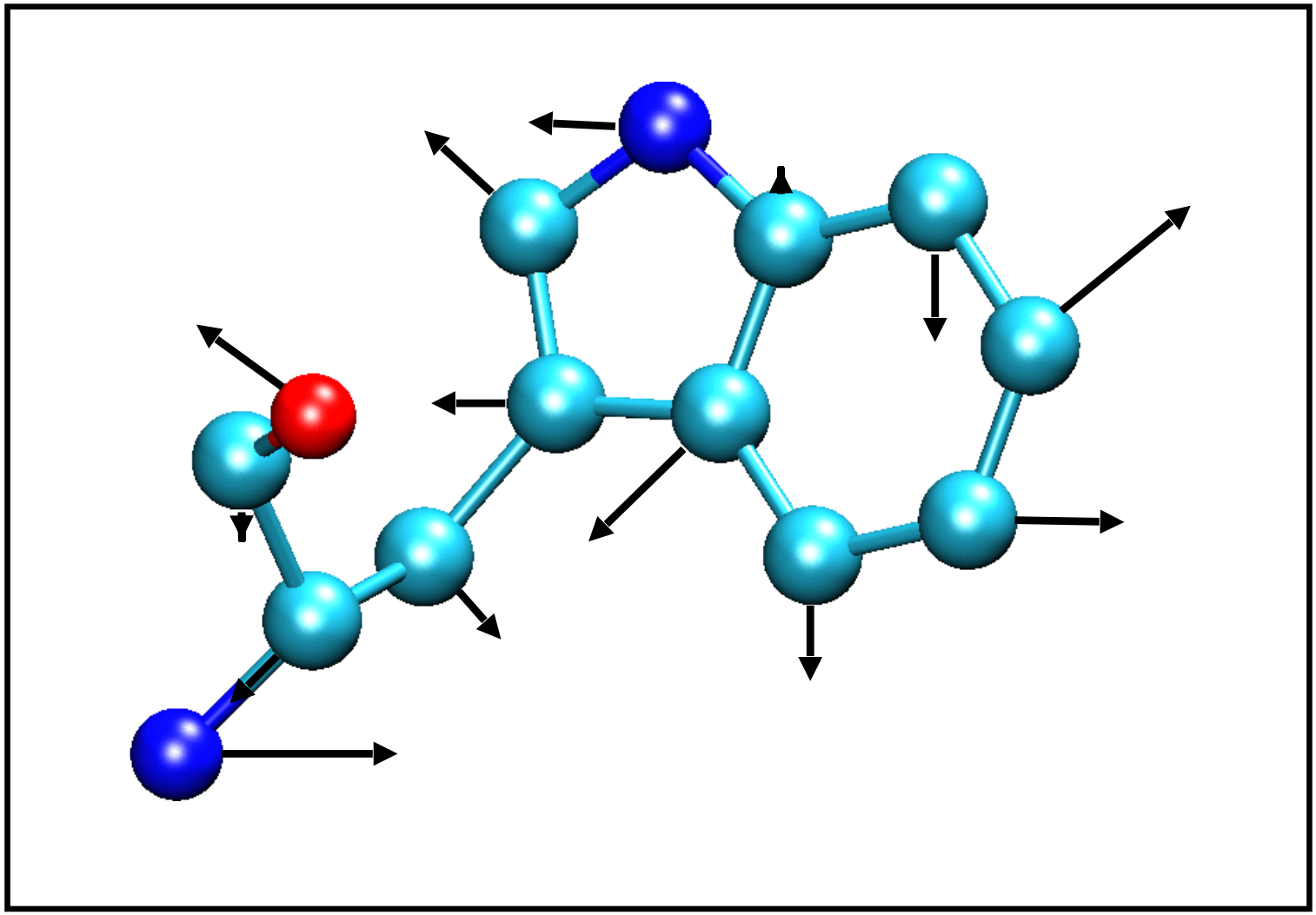
$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

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

Classical Molecular Dynamics



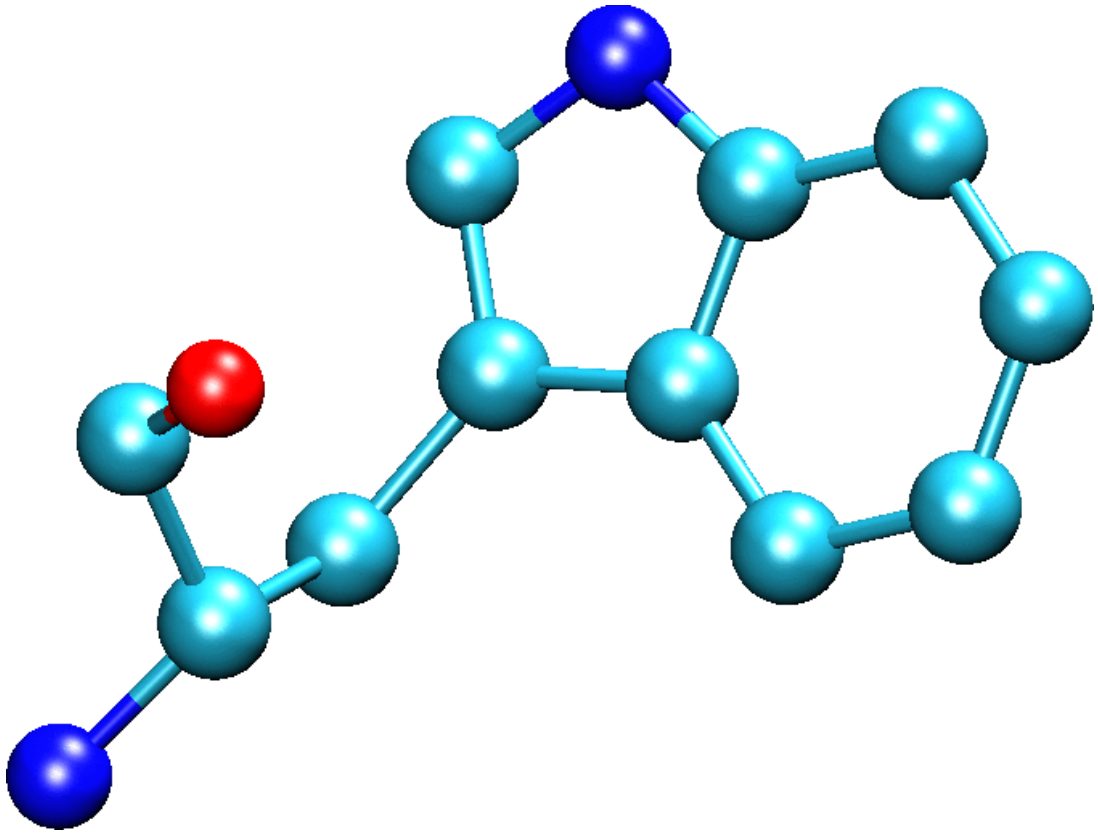
Classical Molecular Dynamics



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

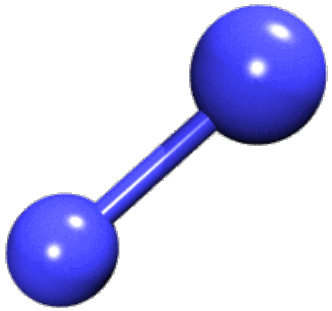
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle

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

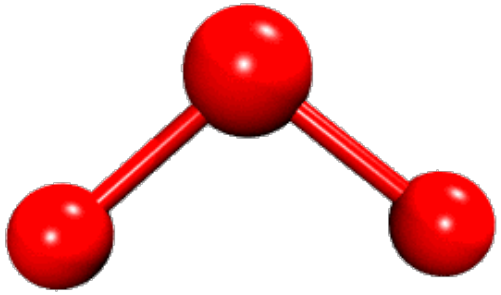
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

Energy Terms Described in the CHARMm Force Field

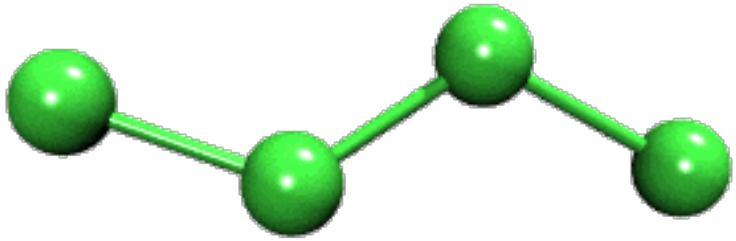
Bond



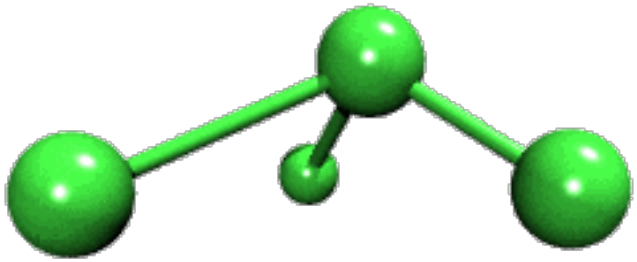
Angle



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:

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

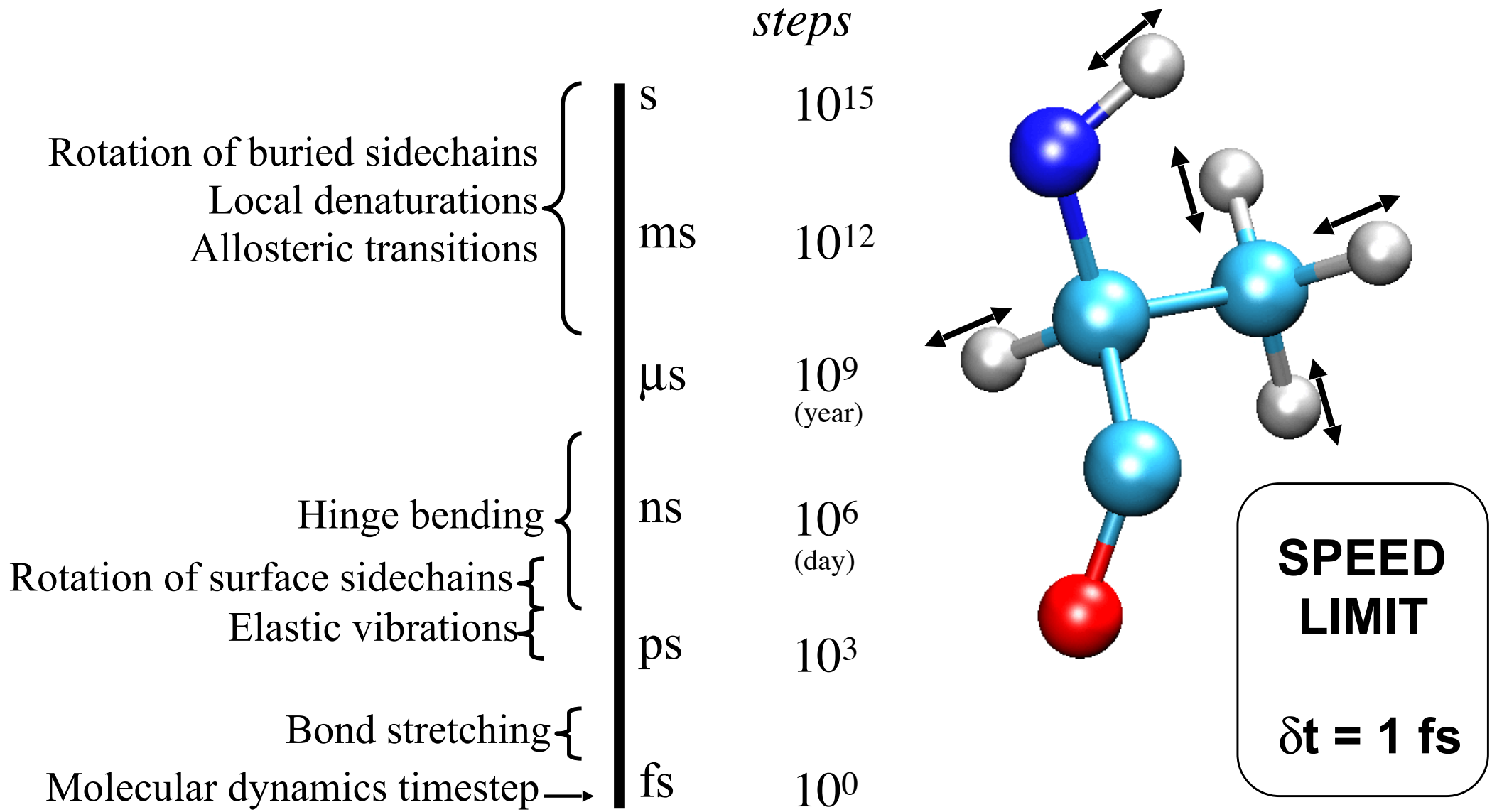
yields a set of 3N coupled 2nd-order differential equations that can be propagated forward (or backward) in time.

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

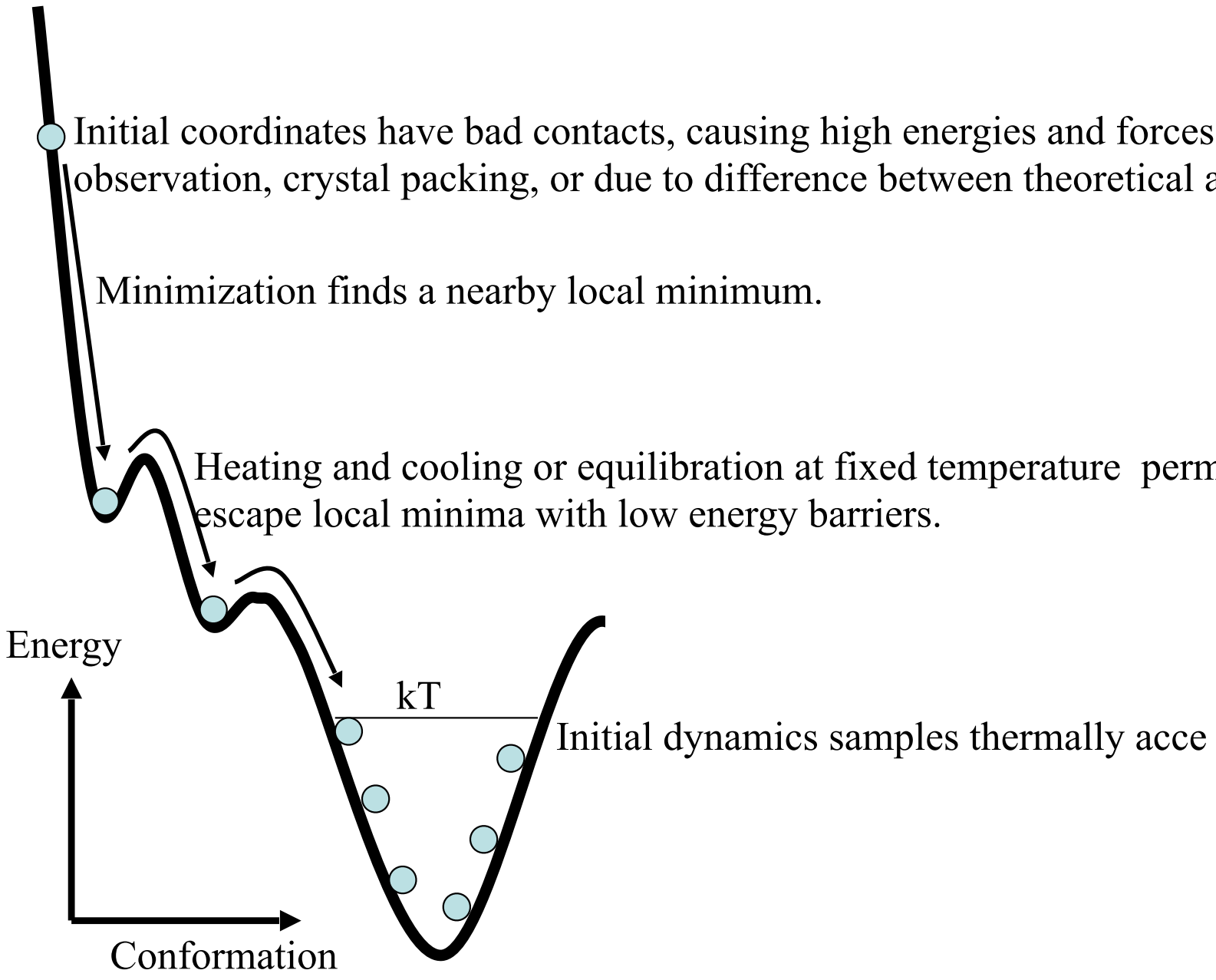
Langevin dynamics deals with each atom separately, balancing a small friction term with Gaussian noise to control temperature:

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

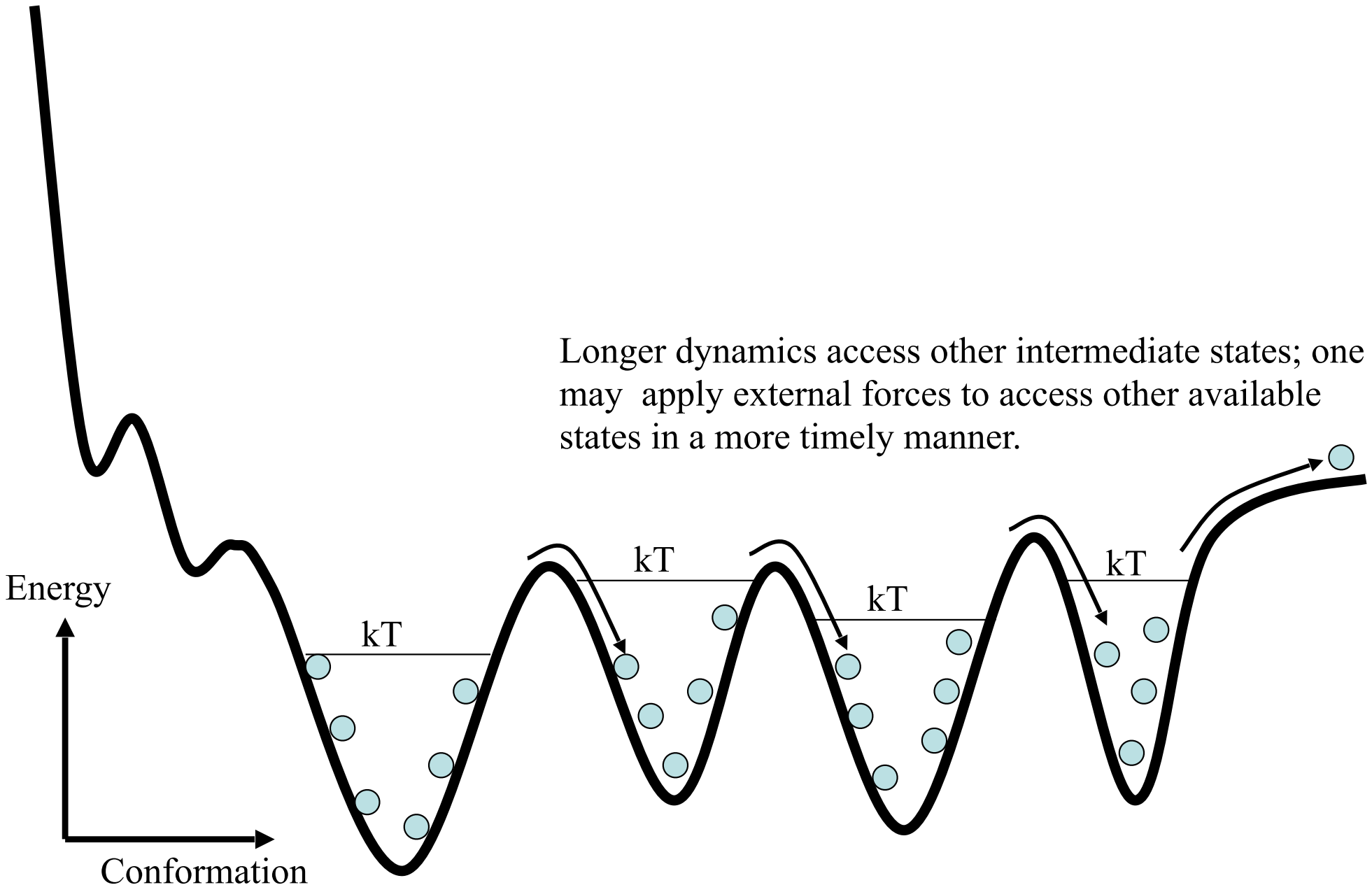
The most serious bottleneck



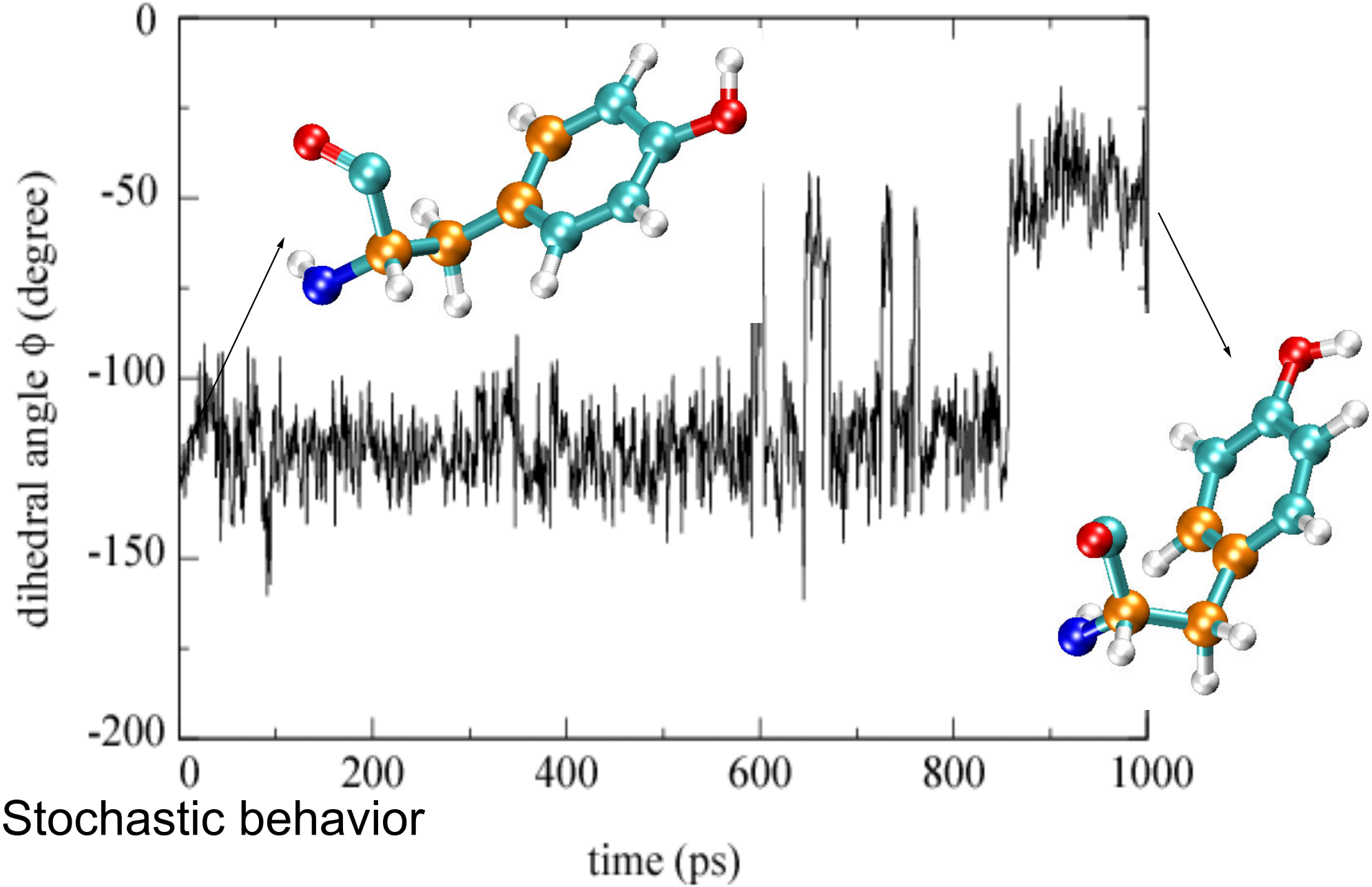
Molecular Dynamics to Sample Energy Landscape



Molecular Dynamics to Sample Energy Landscape



Patience is required to observe Molecular Events



Stochastic behavior

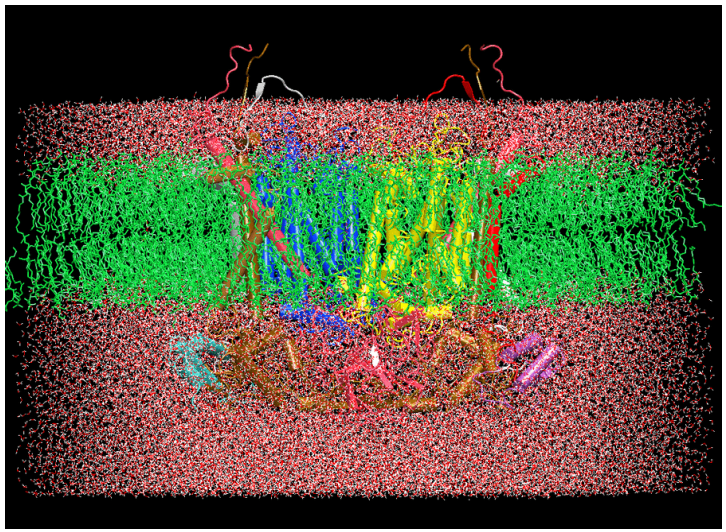
Steps in a Typical MD Simulation

- 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

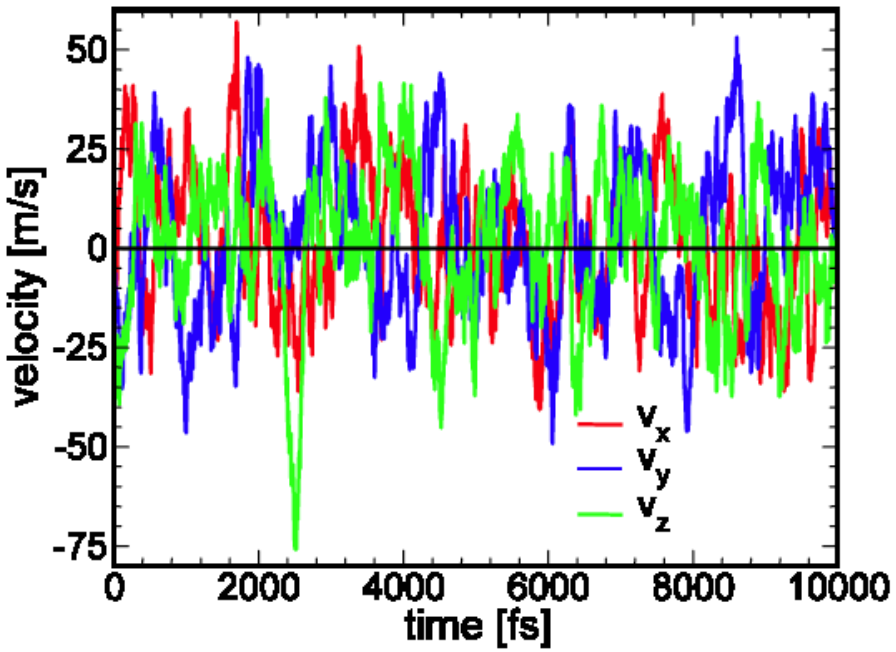
$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t$$

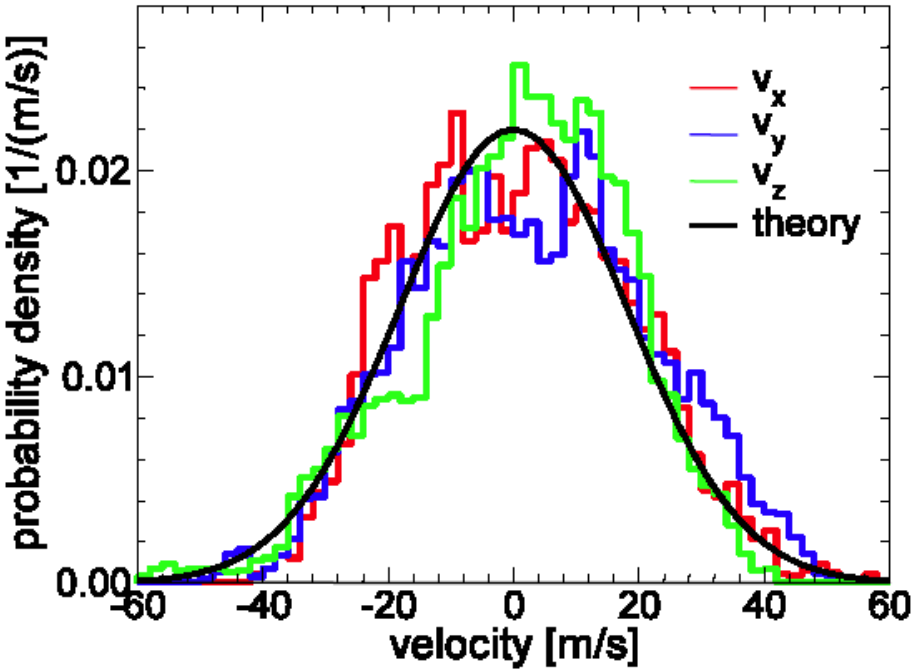
$$\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \boldsymbol{a}(t)\delta t$$

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

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

Maxwell Distribution of Atomic Velocities





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

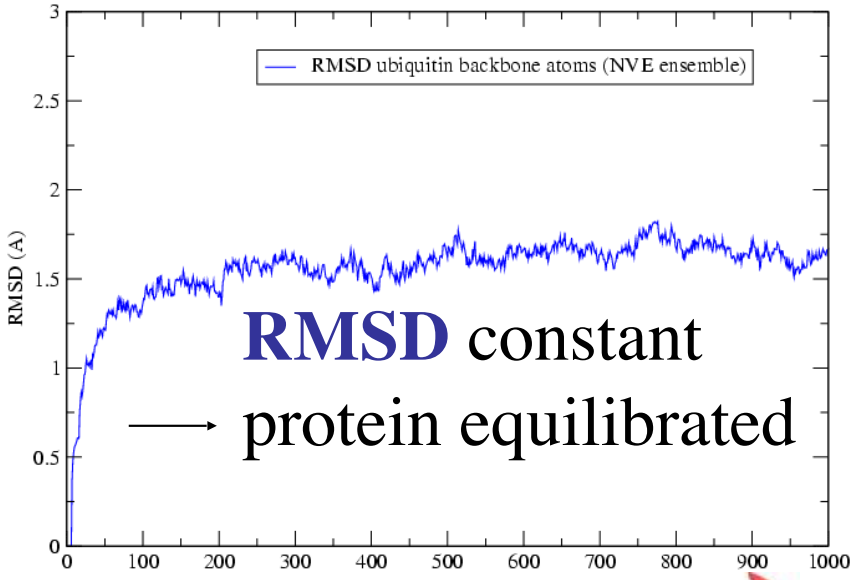
$$\sigma = x,\ y,\ z$$

Equilibrium Properties of Proteins

Ubiquitin

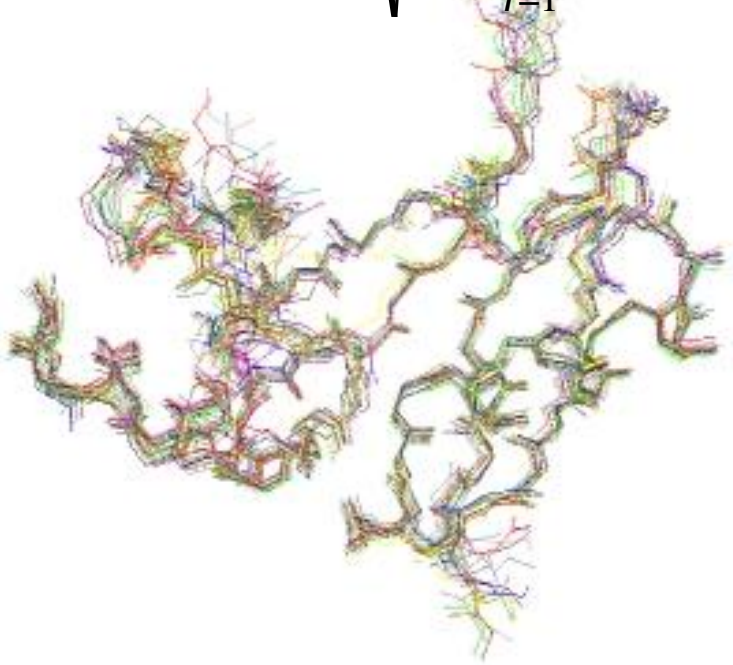
Root Mean Squared Deviation: measure for equilibration and protein flexibility

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



RMSD constant

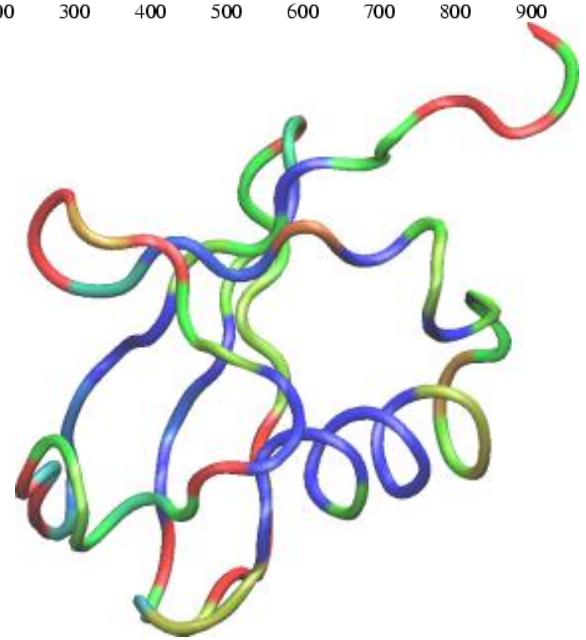
⟶ protein equilibrated



NMR structures

aligned together to see flexibility

Protein sequence exhibits characteristic permanent flexibility!

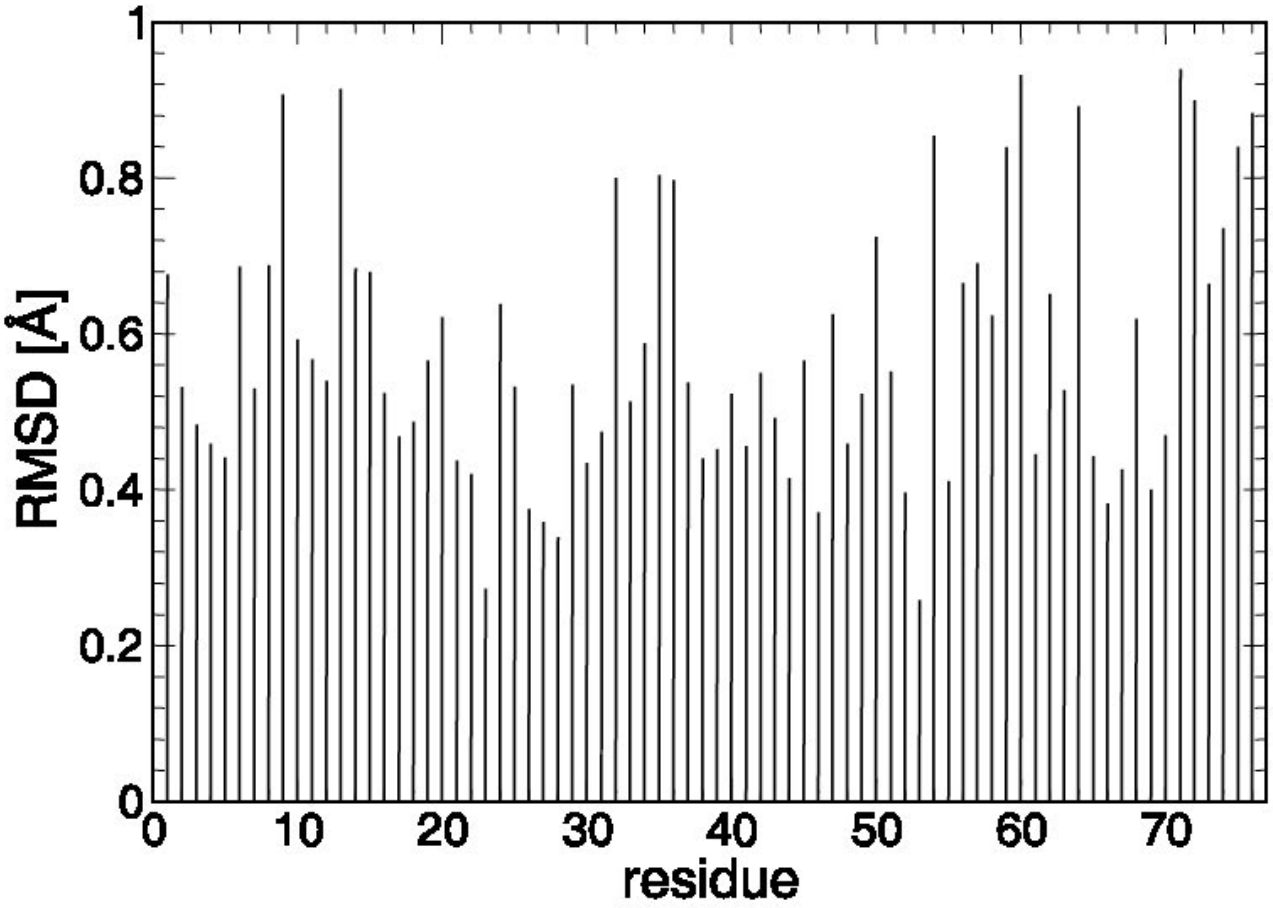


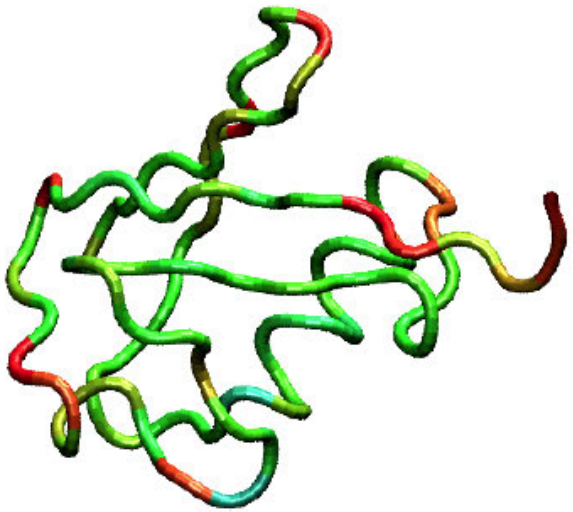
MD simulation

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

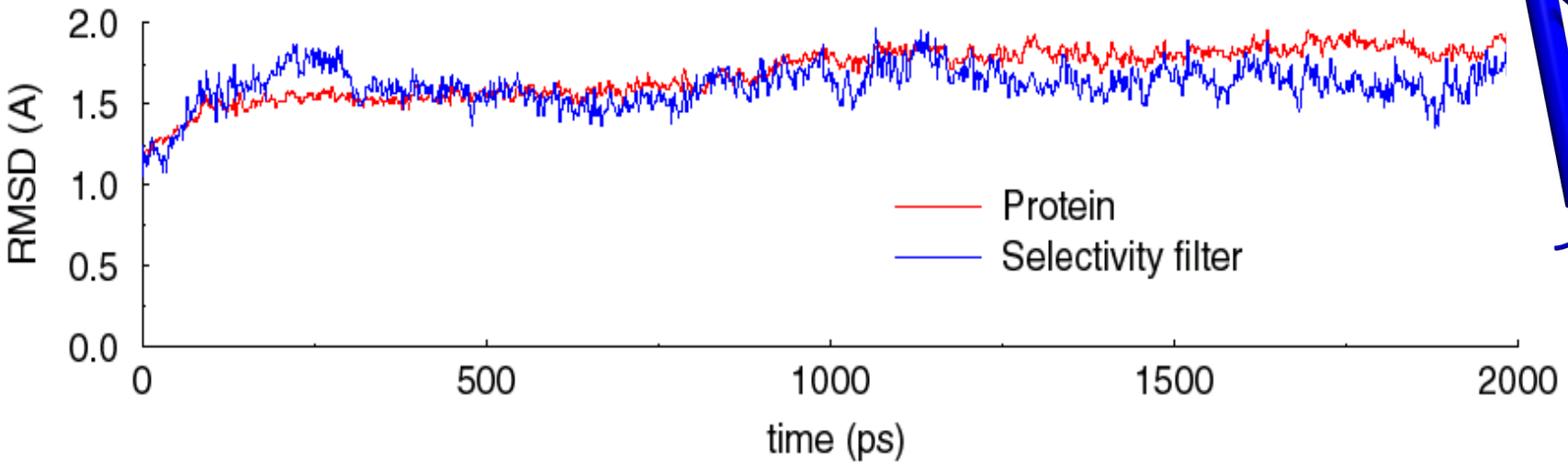
Thermal Motion of Ubiquitin from MD

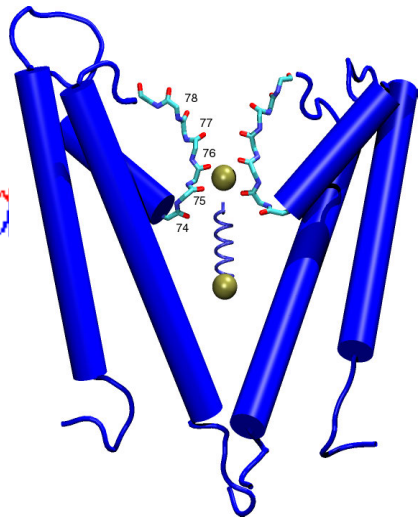
RMSD values per residue



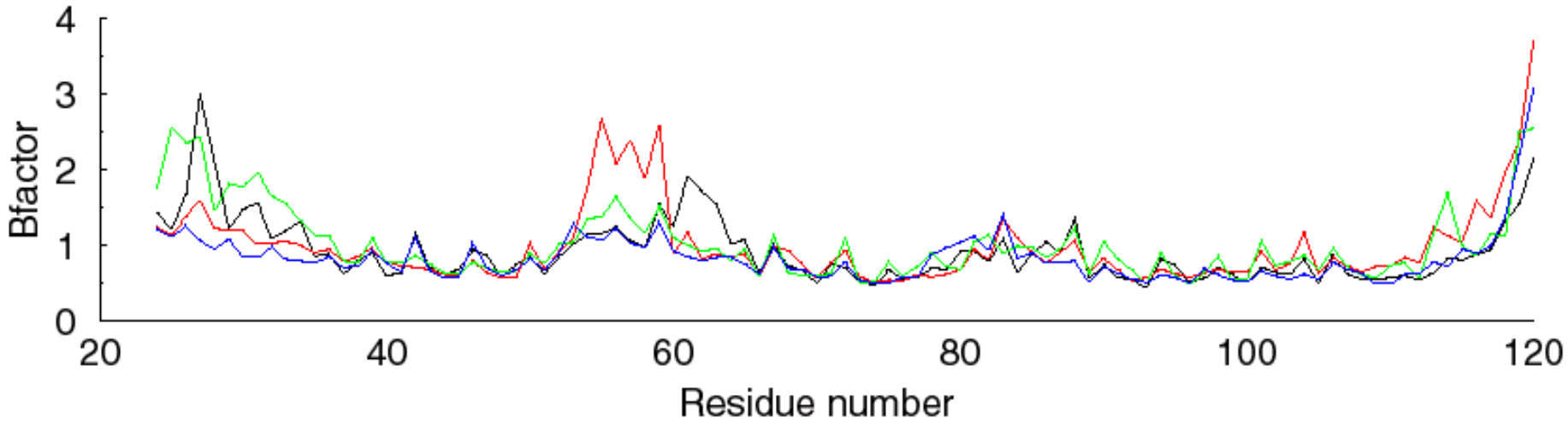


MD Results





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.

Battling the Timescale

non-Equilibrium MD simulations

Reduced Representations

Battling the Timescale - Case I

Steered Molecular Dynamics is a non-equilibrium method by nature

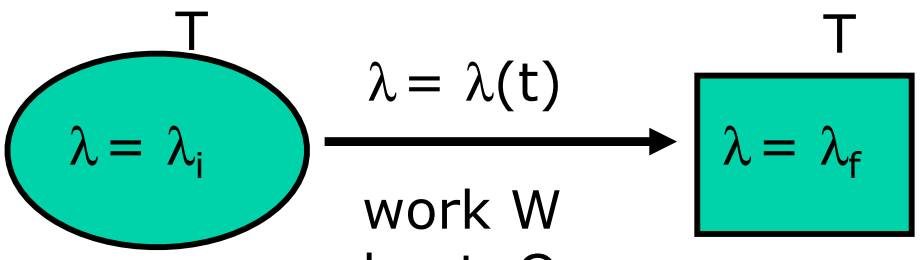
- A wide variety of events that are inaccessible to conventional molecular dynamics simulations can be probed.

- The system will be driven, however, away from equilibrium, resulting in problems in describing the energy landscape associated with the event of interest.

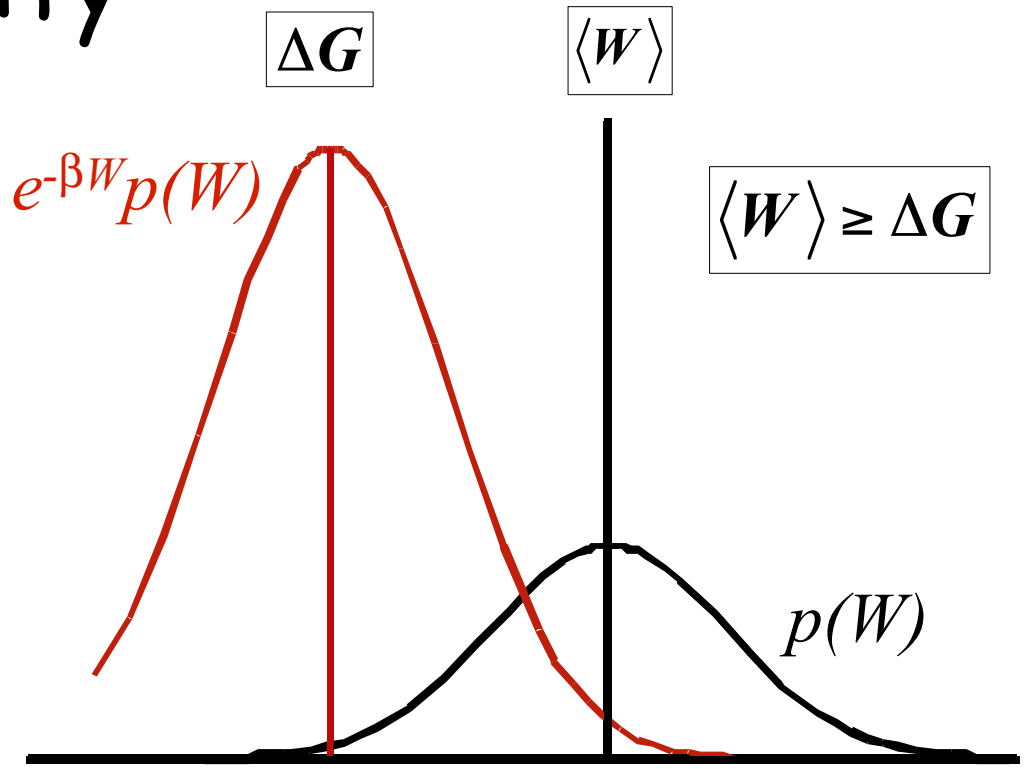
Second law of thermodynamics $\longrightarrow W \geq \Delta G$

Jarzynski's Equality

Transition between two equilibrium states



$$\Delta G = G_f - G_i$$



C. Jarzynski, *Phys. Rev. Lett.*, **78**, 2690 (1997)
C. Jarzynski, *Phys. Rev. E,* **56**, 5018 (1997)

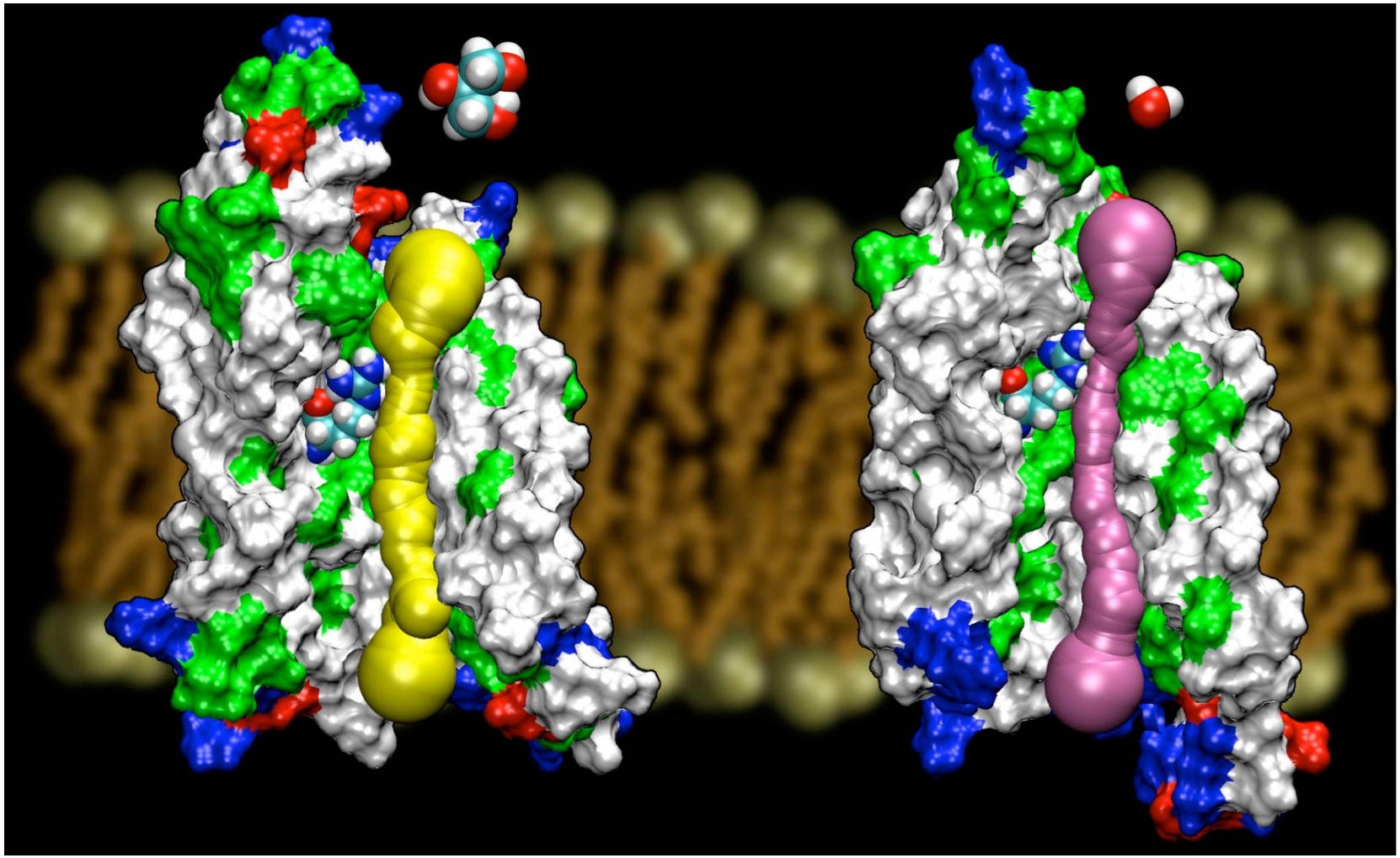
$$\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta G}$$

$$\beta = 1/k_B T$$

In principle, it is possible to obtain free energy surfaces from repeated non-equilibrium experiments.

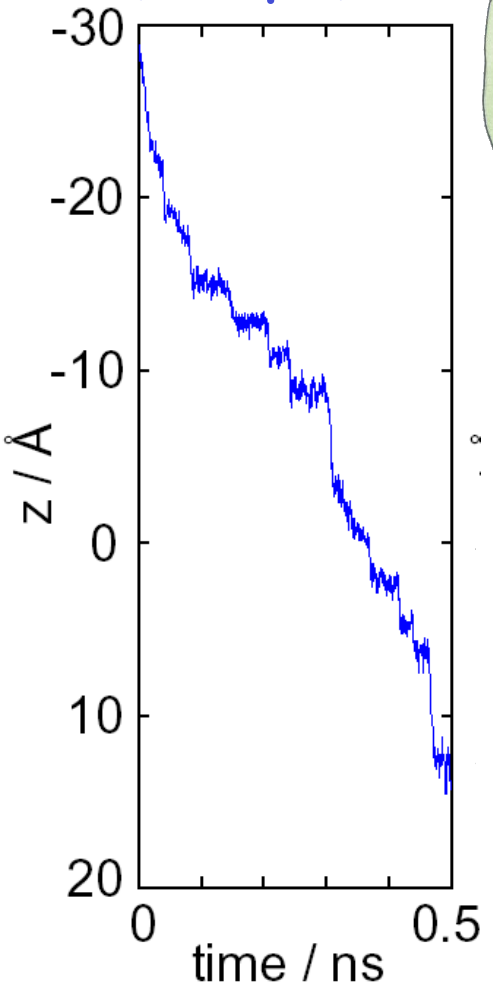
AqpZ *vs*. GlpF

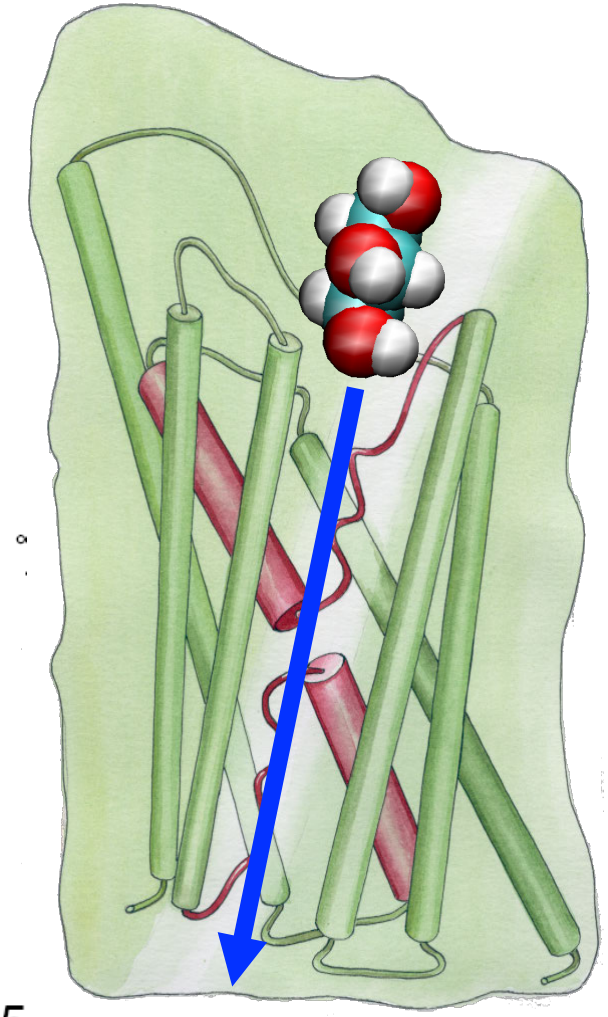
- Both from *E. coli*
- AqpZ is a pure water channel
- GlpF is a glycerol channel
- We have high resolution structures for both channels



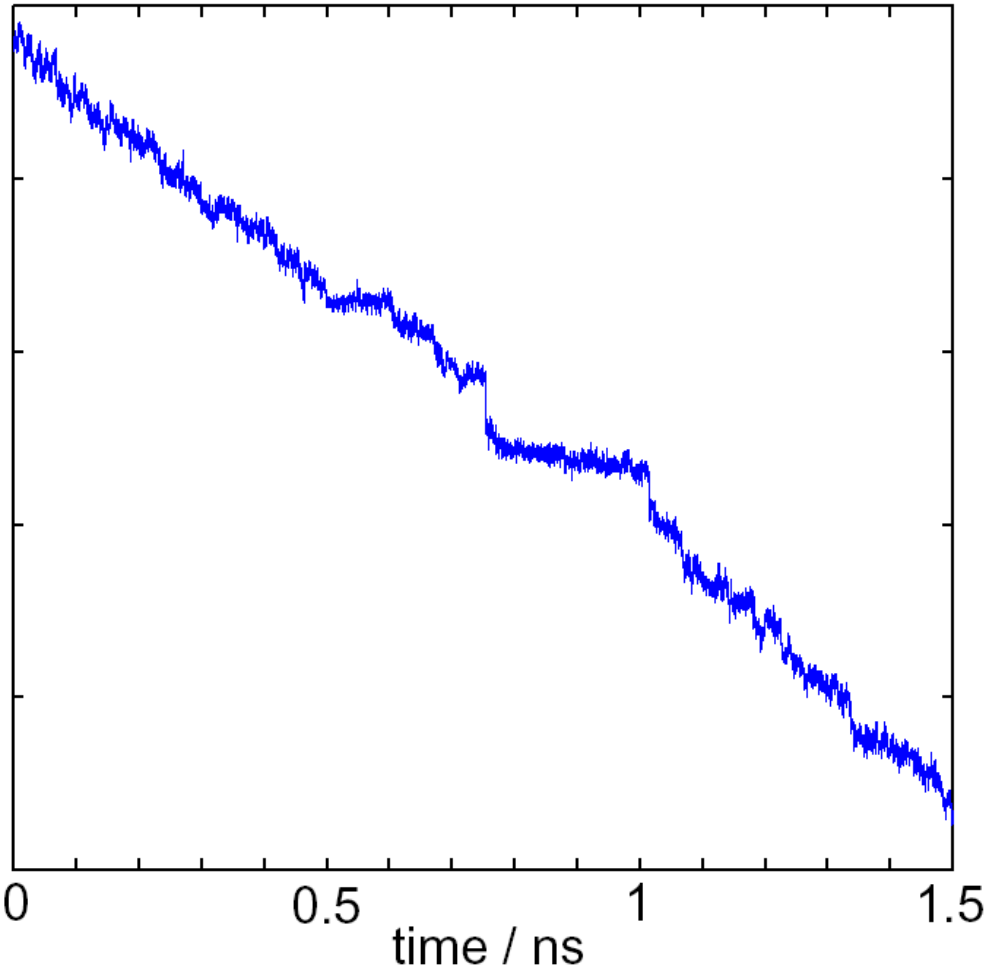
Steered Molecular Dynamics

constant force
(250 pN)

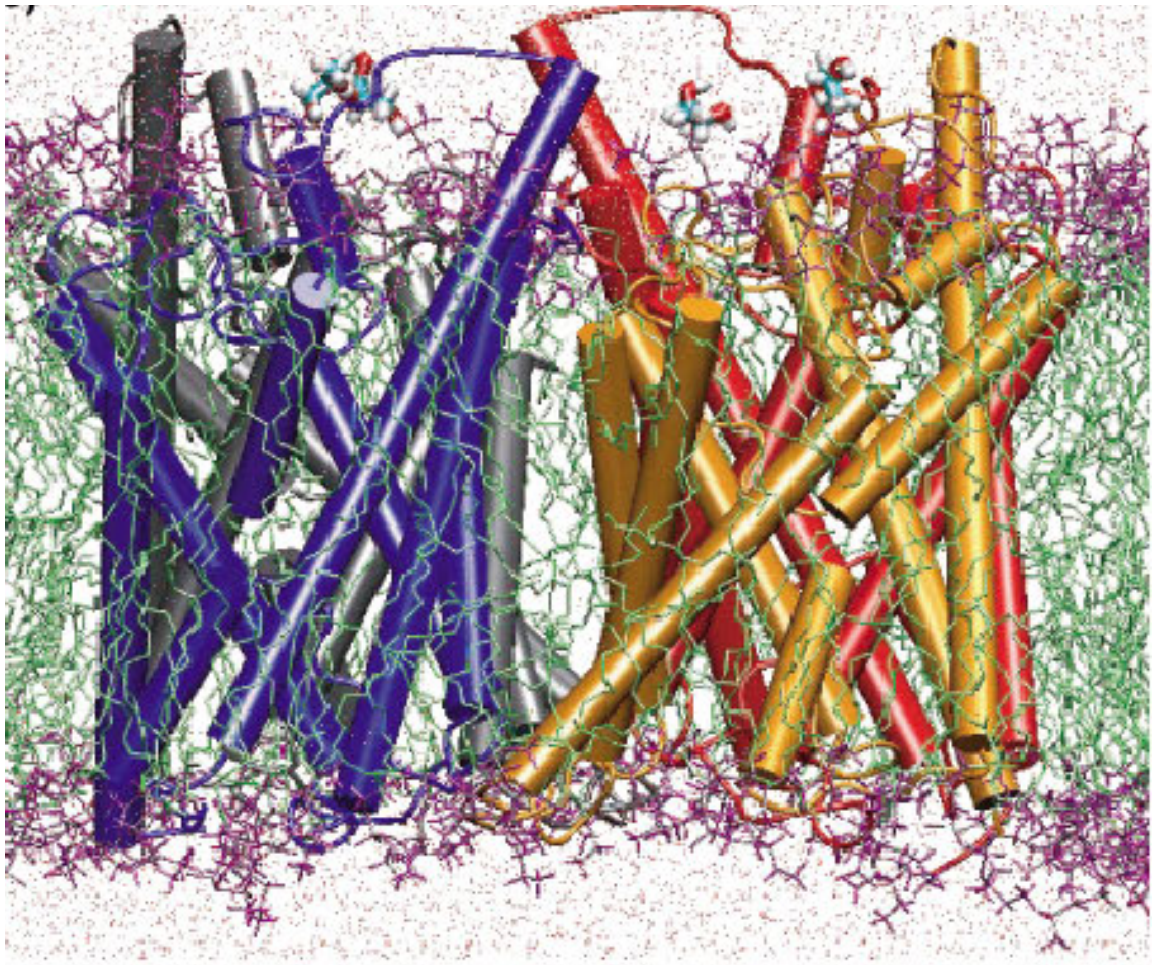


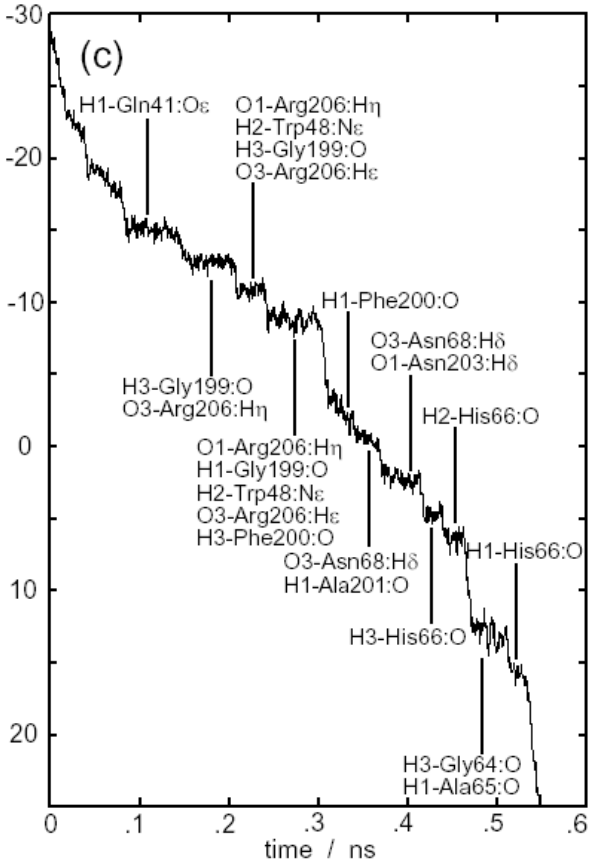


constant velocity
(30 Å/ns)



SMD Simulation of Glycerol Passage





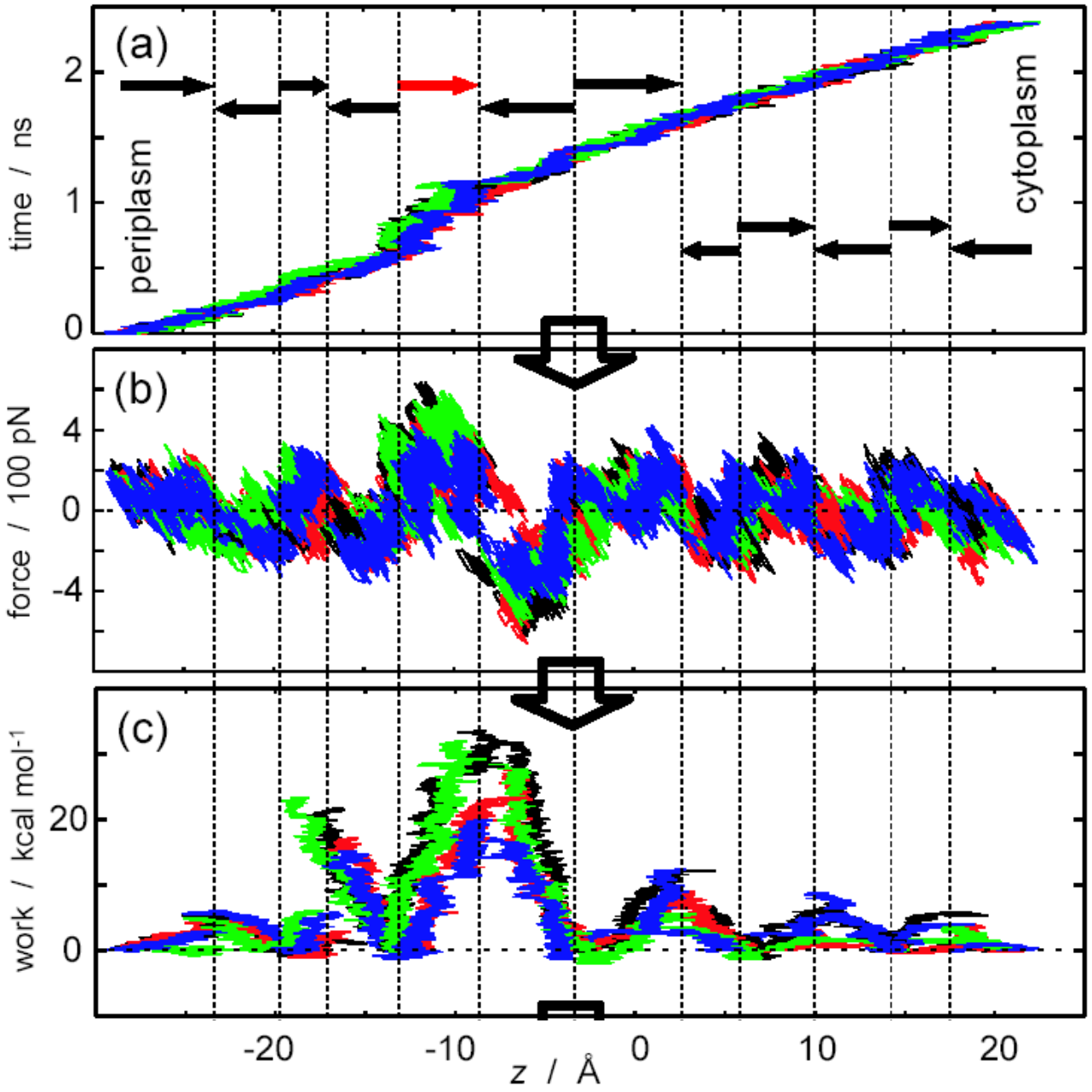
Trajectory of glycerol pulled by **constant force**

Constructing the Potential of Mean Force

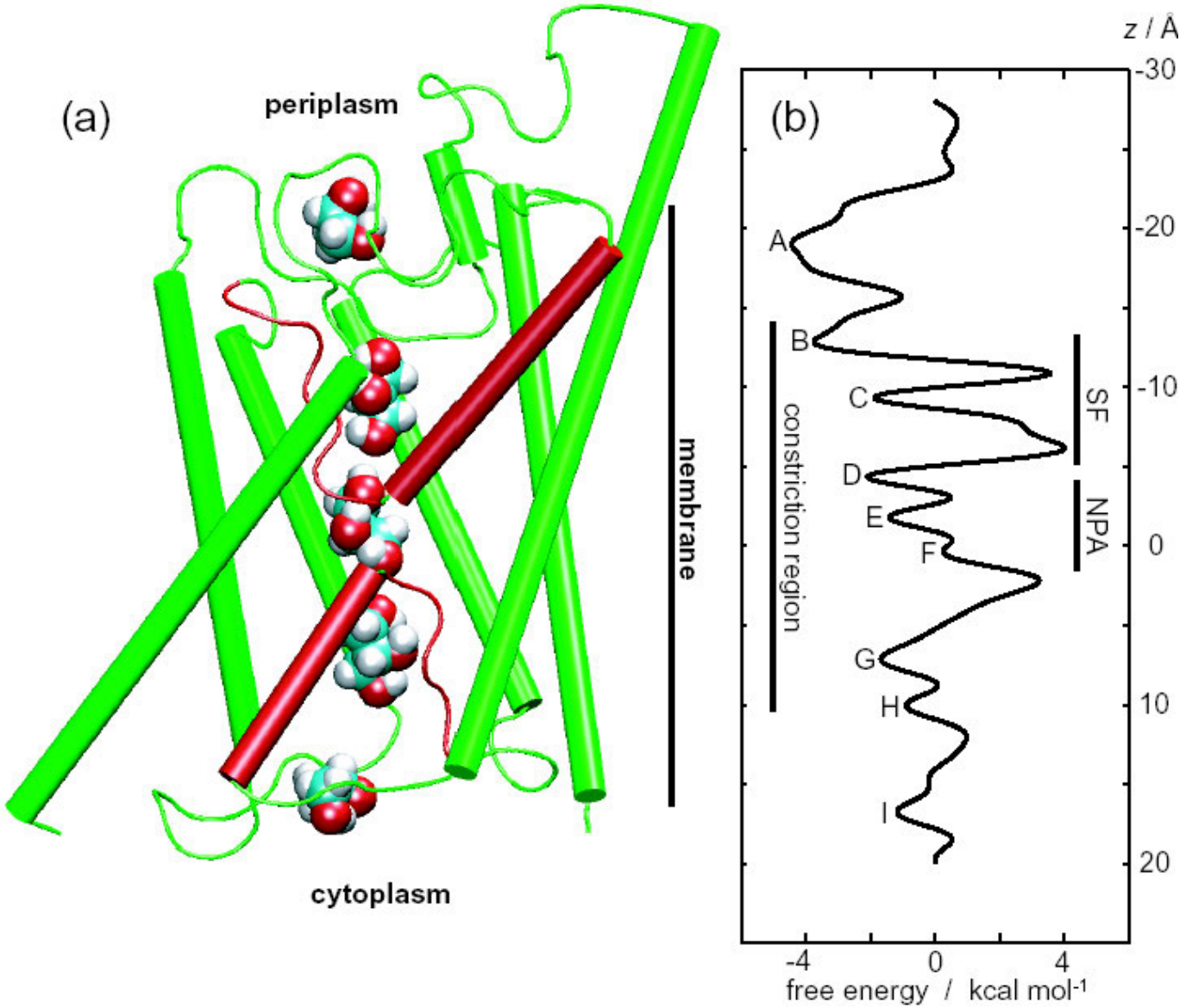
4 trajectories
v = 0.03, 0.015 Å/ps
k = 150 pN/Å

$$f(t) = -k[z(t) - z_0 - vt]$$

$$W(t) = \int_0^t dt'\, vf(t')$$



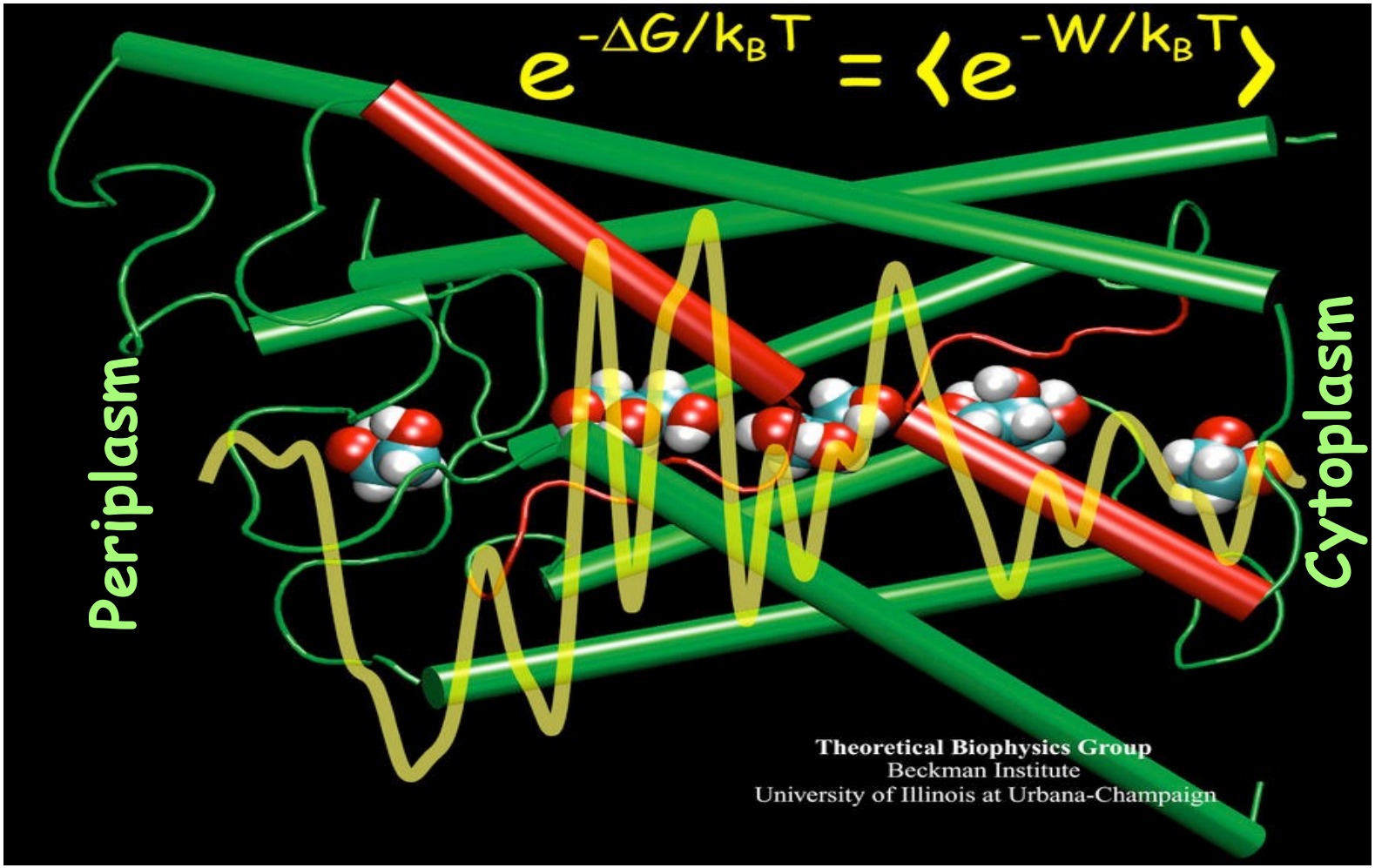
Features of the Potential of Mean Force



- Captures major features of the channel
- The largest barrier ≈**7.3 kcal/mol; exp.: 9.6±1.5 kcal/mol**

Jensen et al., *PNAS*, 99:6731-6736, 2002.

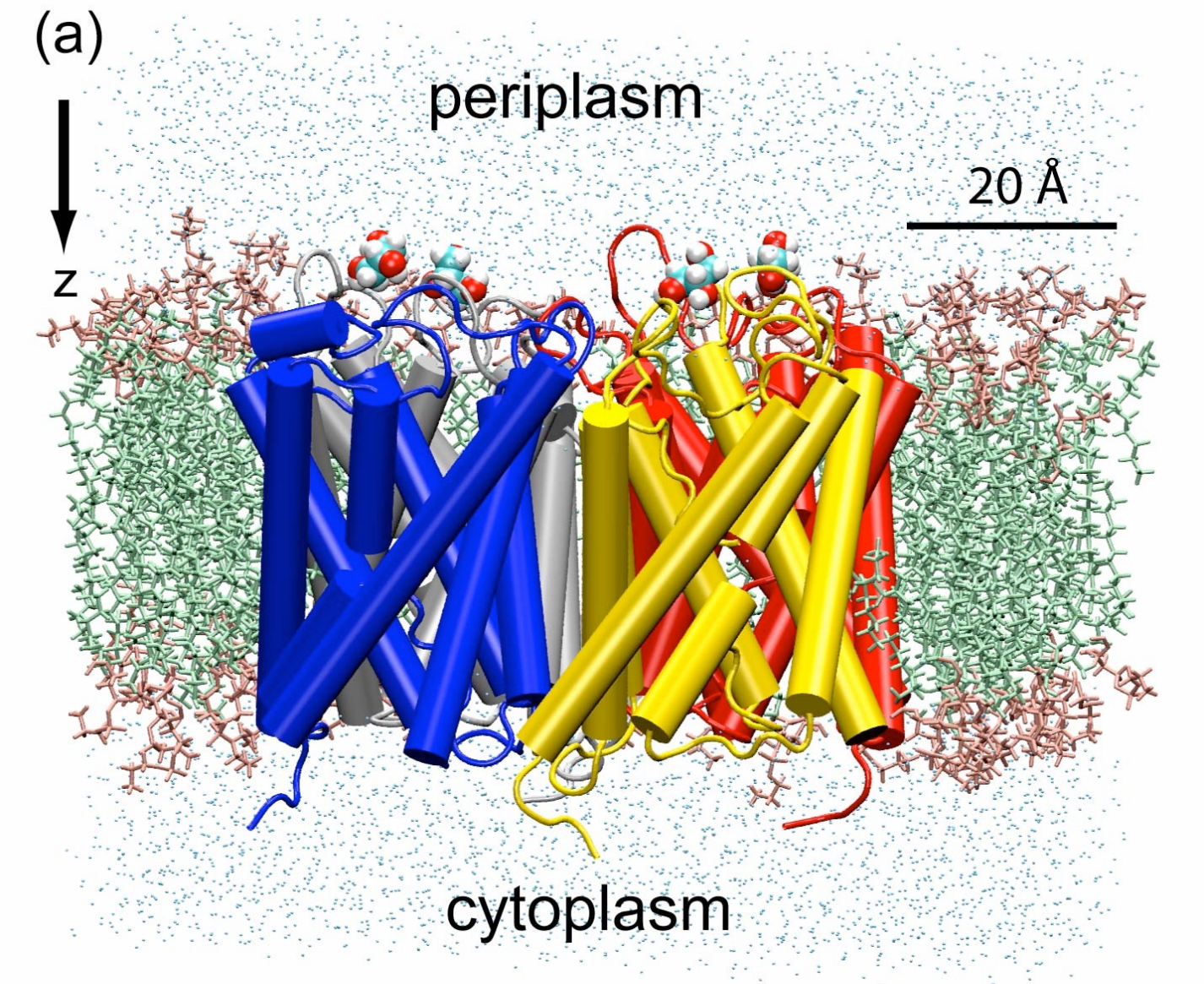
Features of the Potential of Mean Force



Asymmetric Profile in the Vestibules

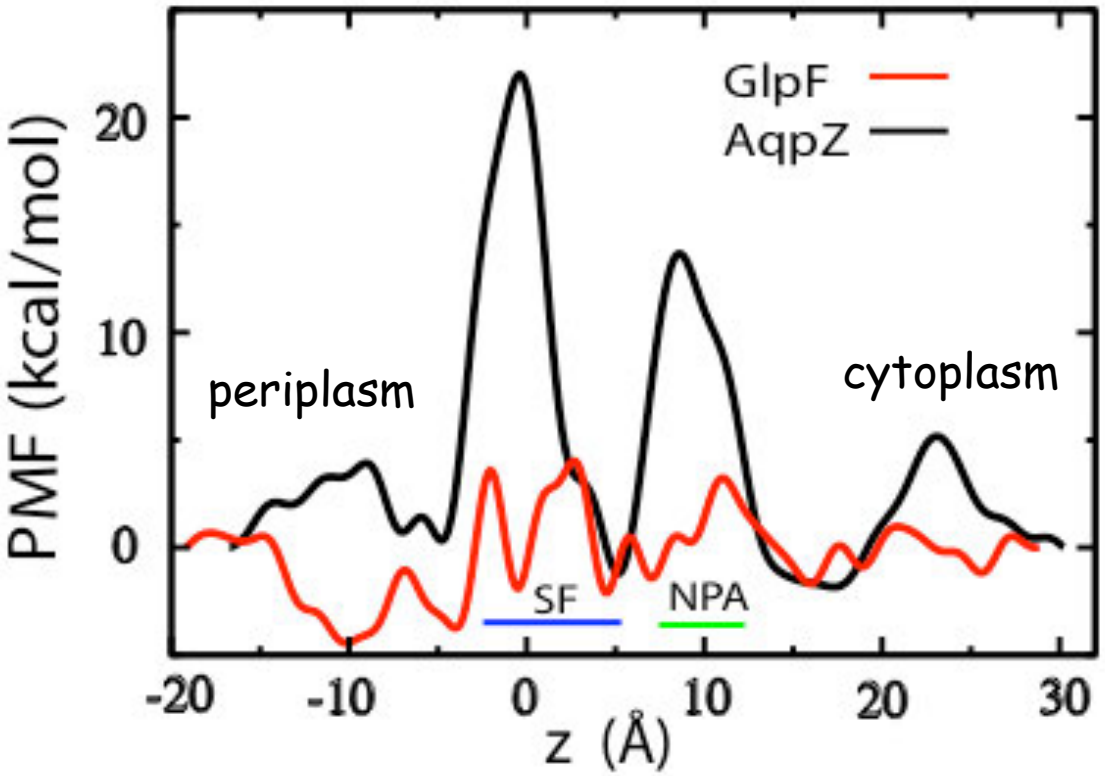
Jensen et al., *PNAS*, 99:6731-6736, 2002.

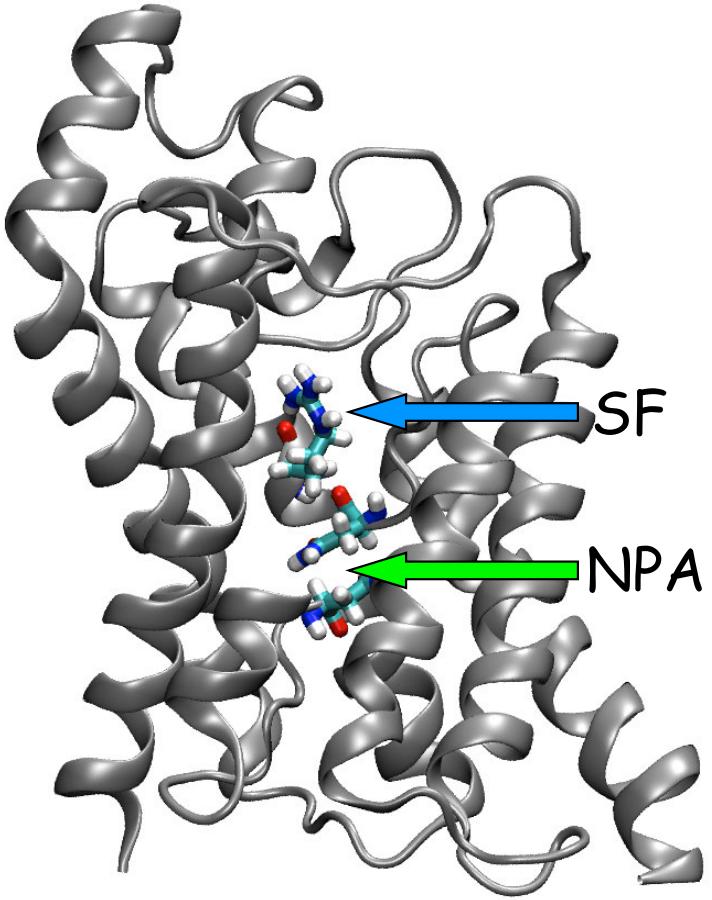
Artificial induction of glycerol conduction through AqpZ



Y. Wang, K. Schulten, and E. Tajkhorshid ***Structure*** 13, 1107 ***(***2005)

Three fold higher barriers

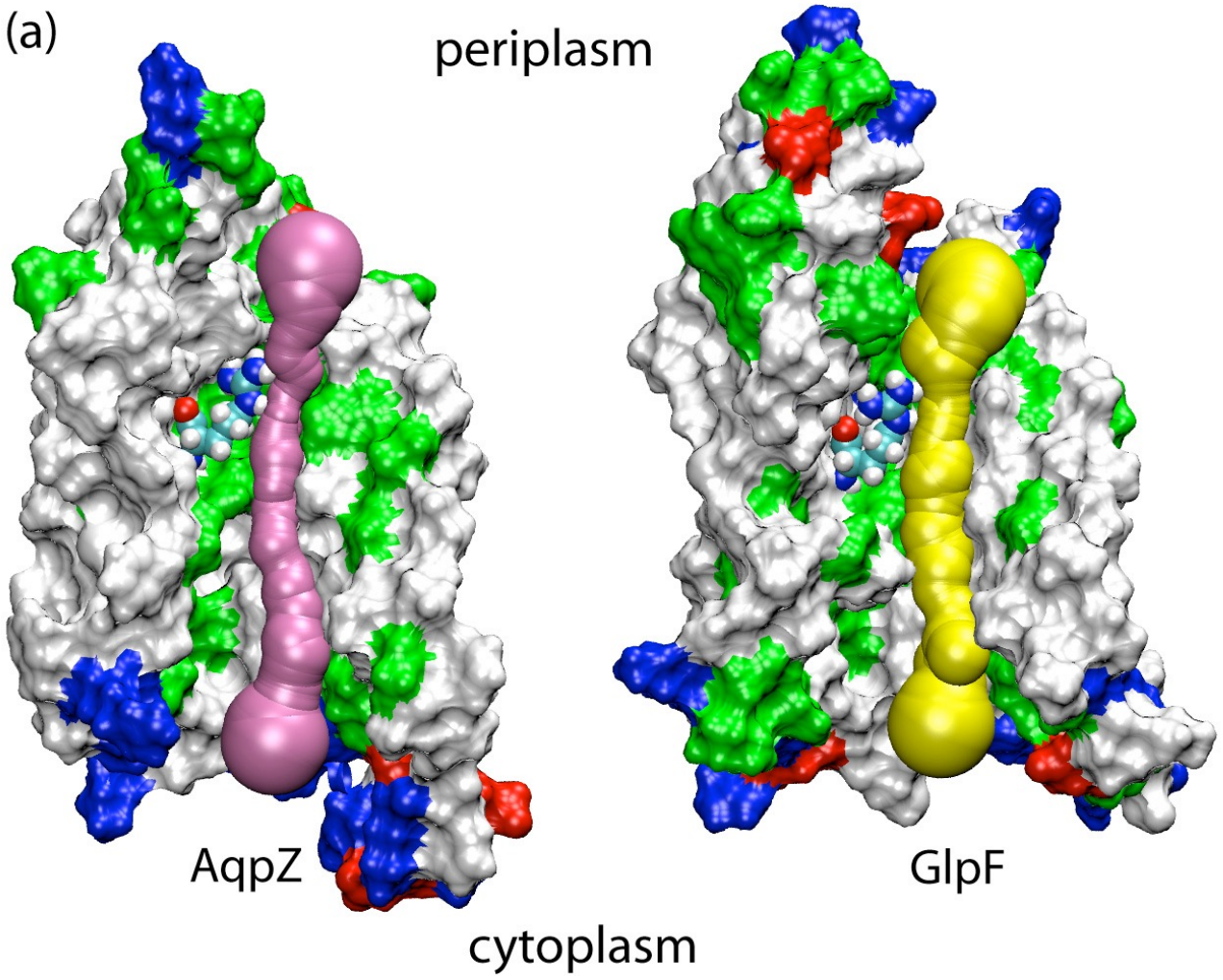


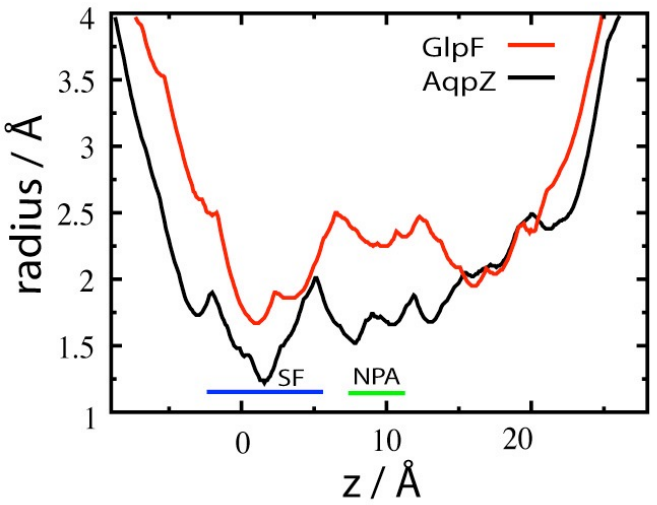


AqpZ 22.8 kcal/mol
GlpF 7.3 kcal/mol

Y. Wang, K. Schulten, and E. Tajkhorshid ***Structure*** 13, 1107 ***(***2005)

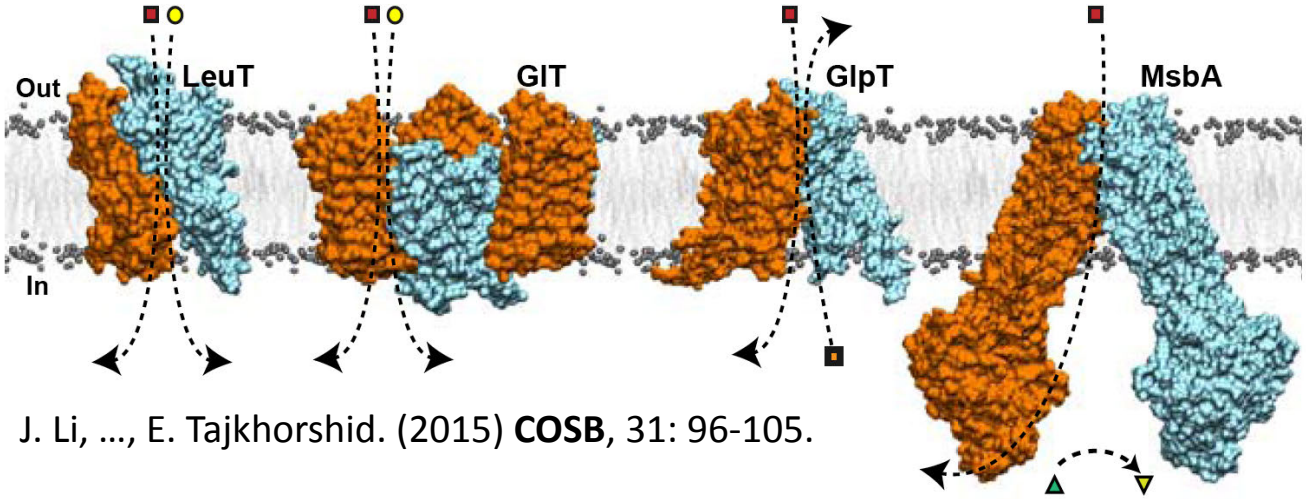
Could it be simply the size?



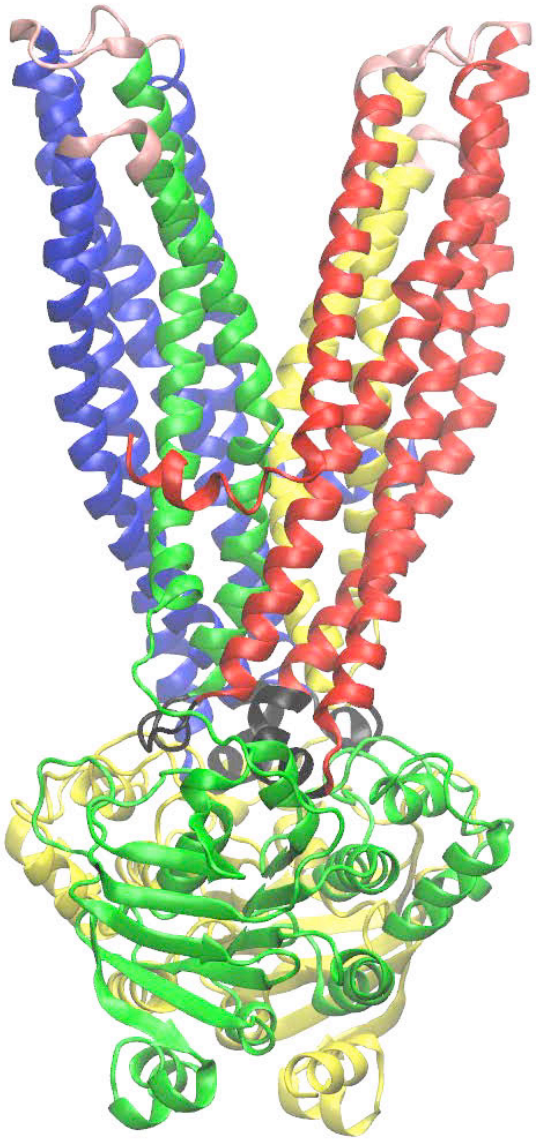


Y. Wang, K. Schulten, and E. Tajkhorshid ***Structure*** 13, 1107 ***(***2005)

Battling the Timescale - Case II
Biased (nonequilibrium) simulations

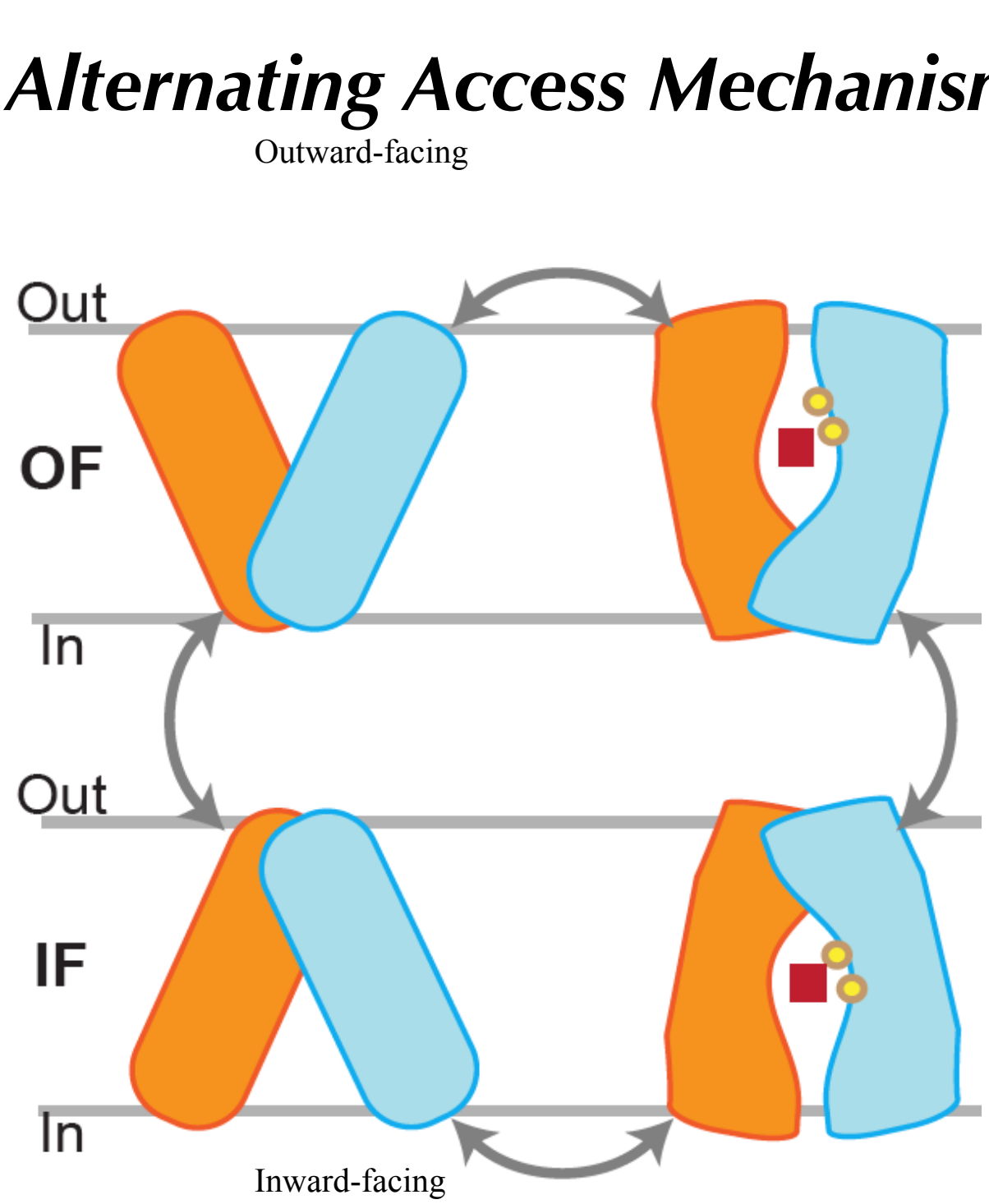


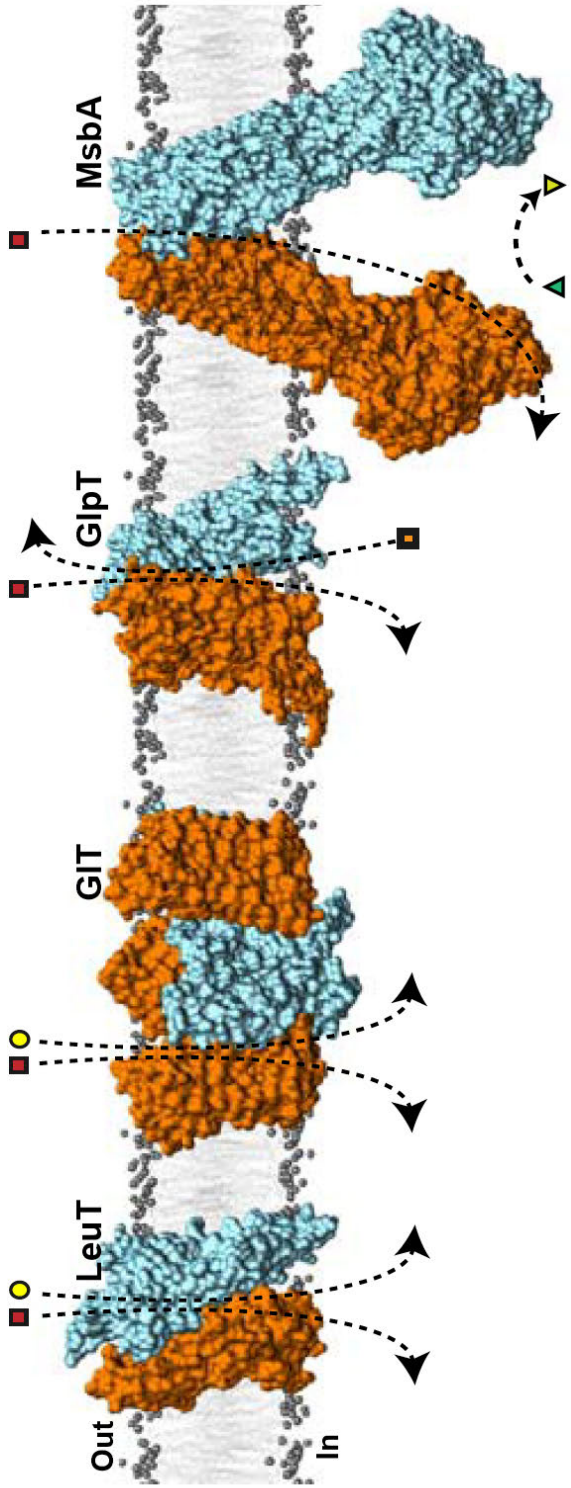
J. Li, …, E. Tajkhorshid. (2015) **COSB**, 31: 96-105.



- **Neurotransmitter Uptake**
  - » Norepinephrine, serotonin, dopamine, glutamate,…
- **Gastrointestinal Tract**
  - » Active absorption of nutrients
  - » Secretion of ions
- **Kidneys**
  - » Reabsorption
  - » Secretion
- **Pharmacokinetics of all drugs**
  - » Absorption, distribution, elimination
  - » Multi-drug resistance in cancer cells

Alternating Access Mechanism



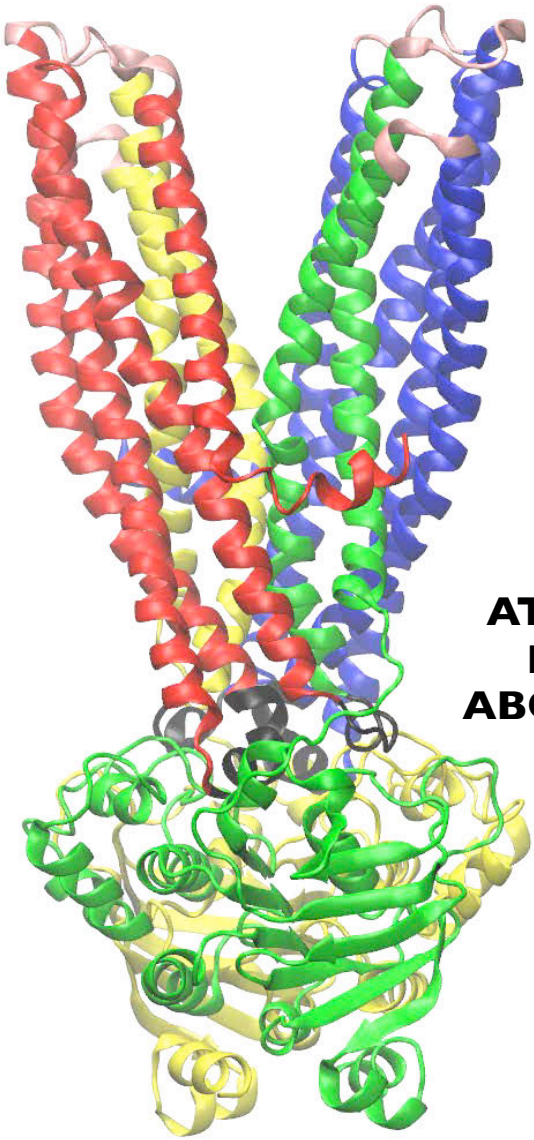


Jardetzky O. *Nature* **211**: 969–970 (1966)

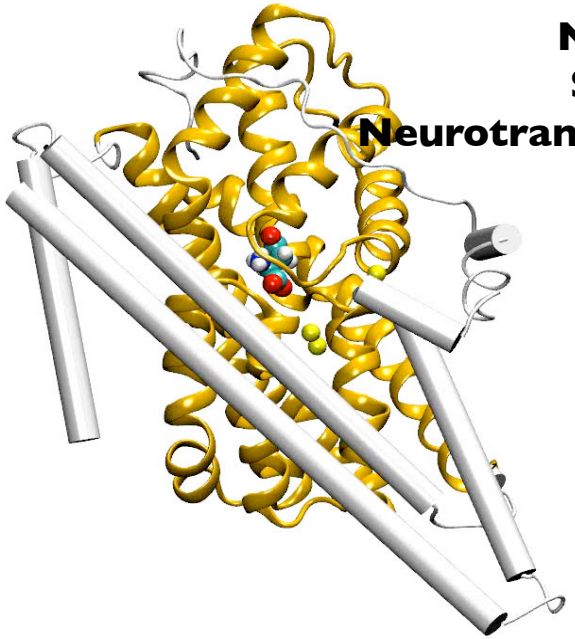
J. Li, …, E. Tajkhorshid. (2015) **COSB**, 31: 96-105.

COMPLEX

Diverse Structural Transitions Involved

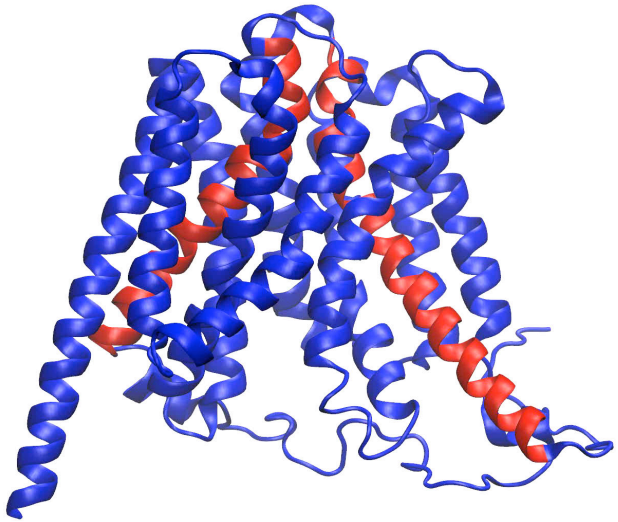


ATP-Driven Primary ABC Exporter



Na-coupled Secondary Neurotransmitter Transporter

Secondary Phosphate Antiporter



NON-EQUILIBRIUM METHODS ARE REQUIRED.

Complex Processes Require Complex Treatments

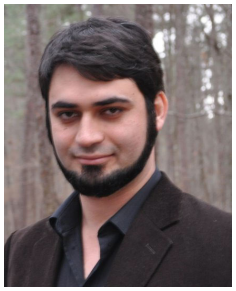
I.1 Defining Practical Collective Variables
Empirical search for practical collective variables for inducing the conformational changes involved in the transition.

I.2 Optimizing the Biasing Protocols
Systematic search for a practical biasing protocol by using different combinations of collective variables.

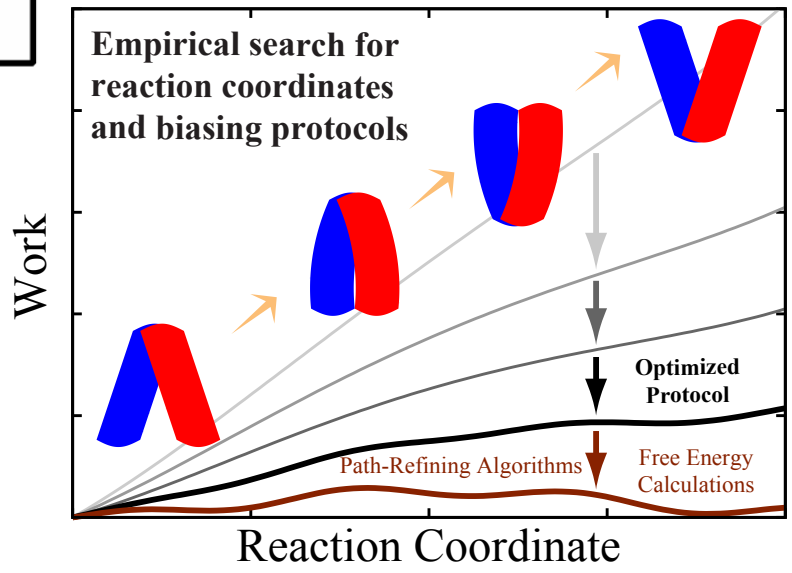
II. Optimizing the Transition Pathway
Use all of the conformations available to generate the most reliable transition pathway:
1. Bayesian approach for combining the data
2. Post-hoc string method (analysis tool)
3. String method with swarms of trajectories

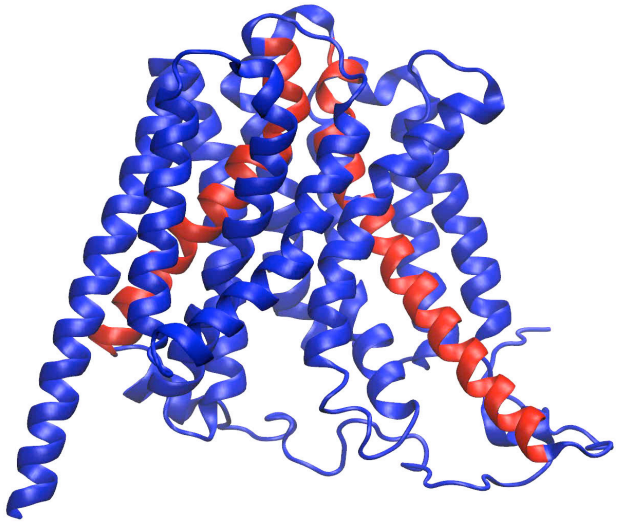
III.1 Free Energy Calculations
Using the most relevant collective variables (from I.1), biasing protocol (from I.2), and initial conformations (from I.2).

III.2 Assessing the Sampling Efficiency
Detecting the poorly sampled, but potentially important regions, e.g., by using PCA.



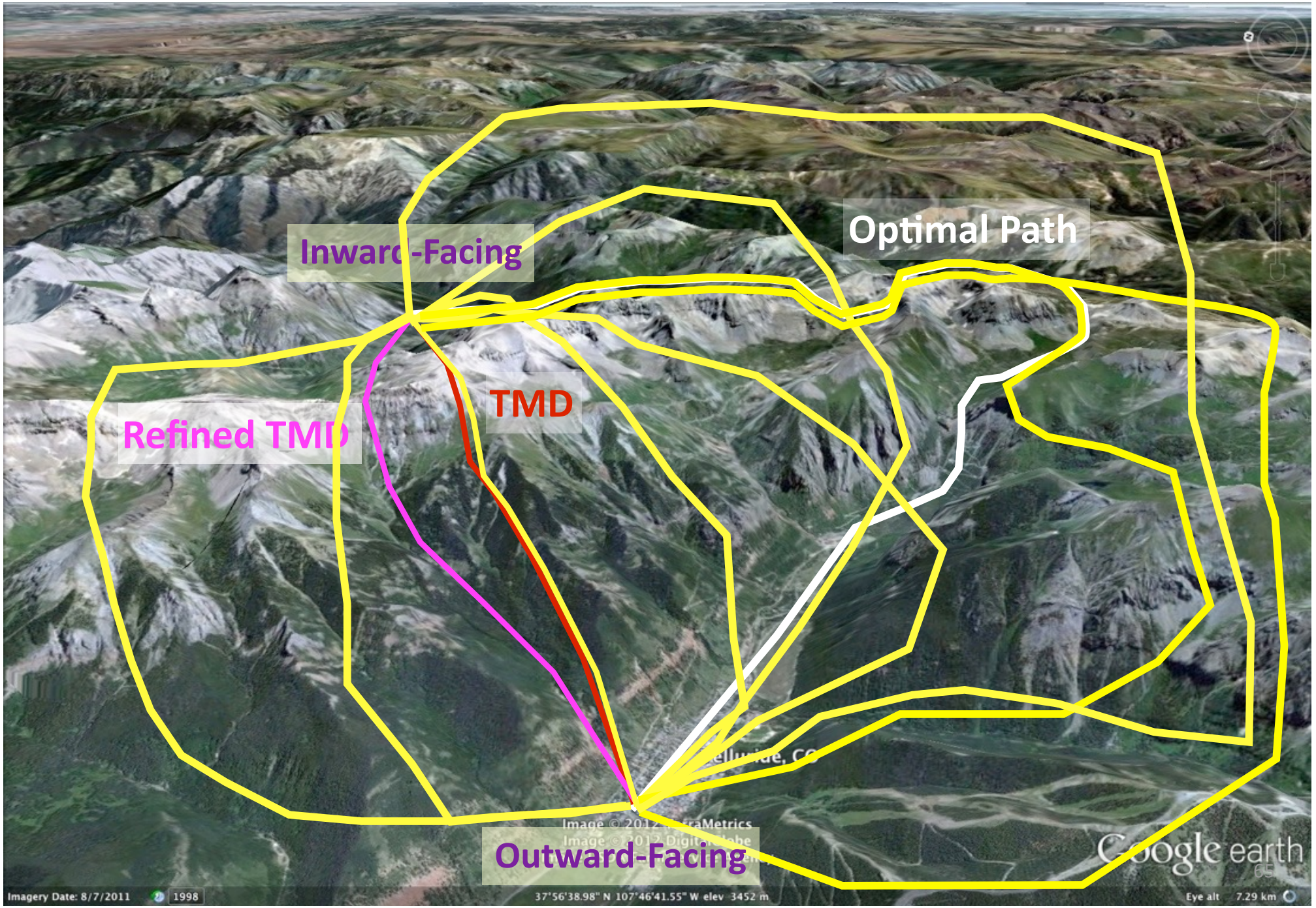
Mahmoud Moradi





M. Moradi and ET (2013) **PNAS**, 110:18916–18921.
M. Moradi and ET (2014) **JCTC**, 10: 2866–2880.
M. Moradi, G. Enkavi, and ET (2015) **Nature Comm.**, 6:8393.

Aggressive Search of the Space



Non-equilibrium Driven Molecular Dynamics:

Applying a time-dependent external force to induce the transition

Along various pathways/mechanisms (collective variables)

Harmonic constant

Initial state

$$U_{dr}(\boldsymbol{x}, t) = \frac{1}{2}k\left(\xi(\boldsymbol{x}) - \xi_A + (\xi_B - \xi_A)\frac{t}{\mathrm{T}}\right)^2$$

Biasing potential

Collective variables:
RMSD, distance,
R_g, angle, …
orientation quaternion

Final state

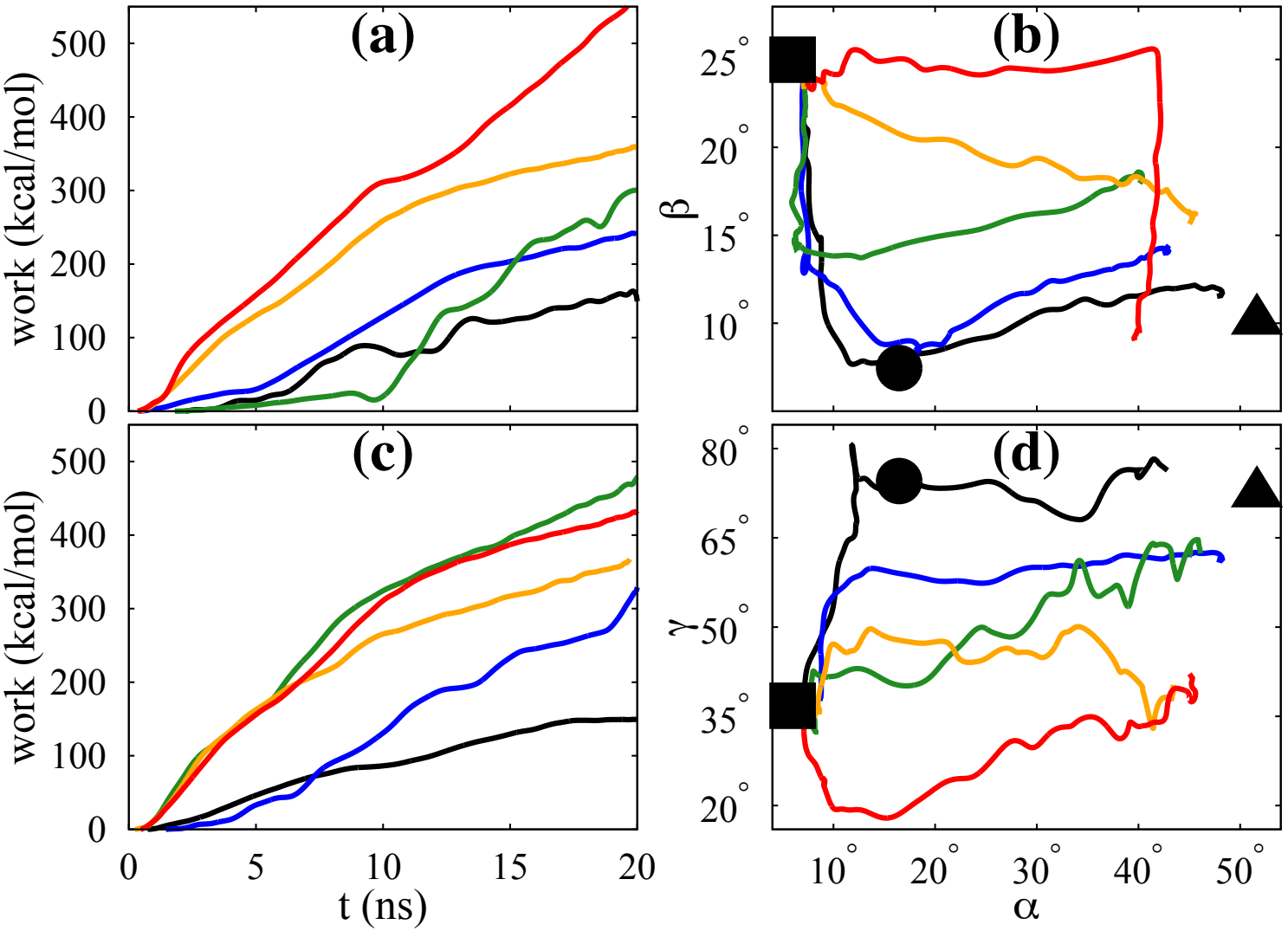
Total simulation time

M. Moradi and ET (2013) **PNAS**, 110:18916–18921.
M. Moradi and ET (2014) **JCTC**, 10: 2866–2880.
M. Moradi, G. Enkavi, and ET (2015) **Nature Comm.**, 6:8393.

Progressively Optimizing the Biasing Protocol/Collective Variable using non-Equilibrium Work as a Measure of the Path Quality

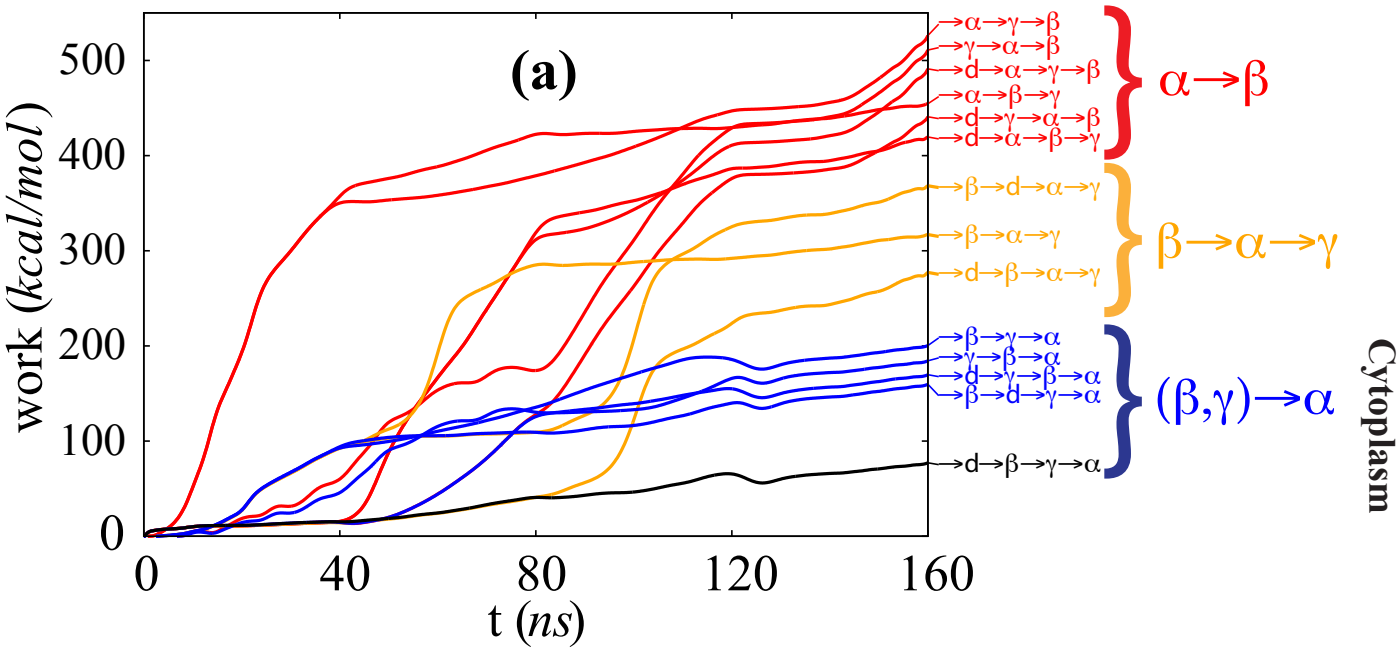
Work

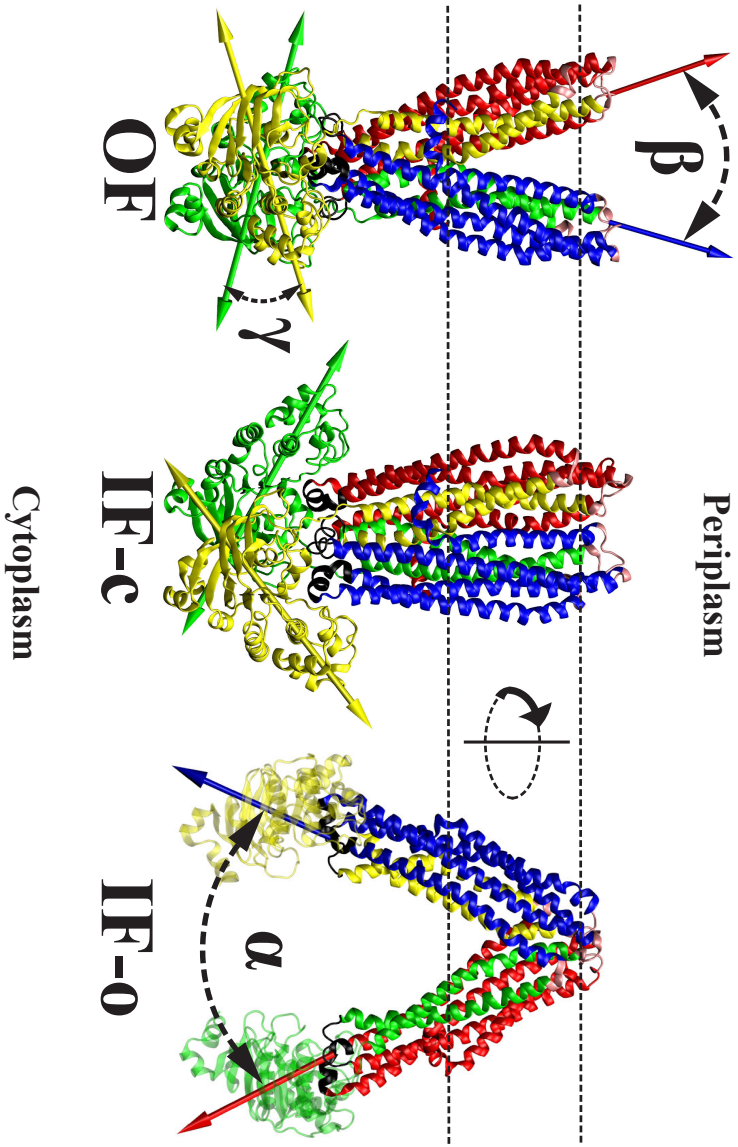
Mechanism



Example set taken from a subset of 20 ns biased simulations

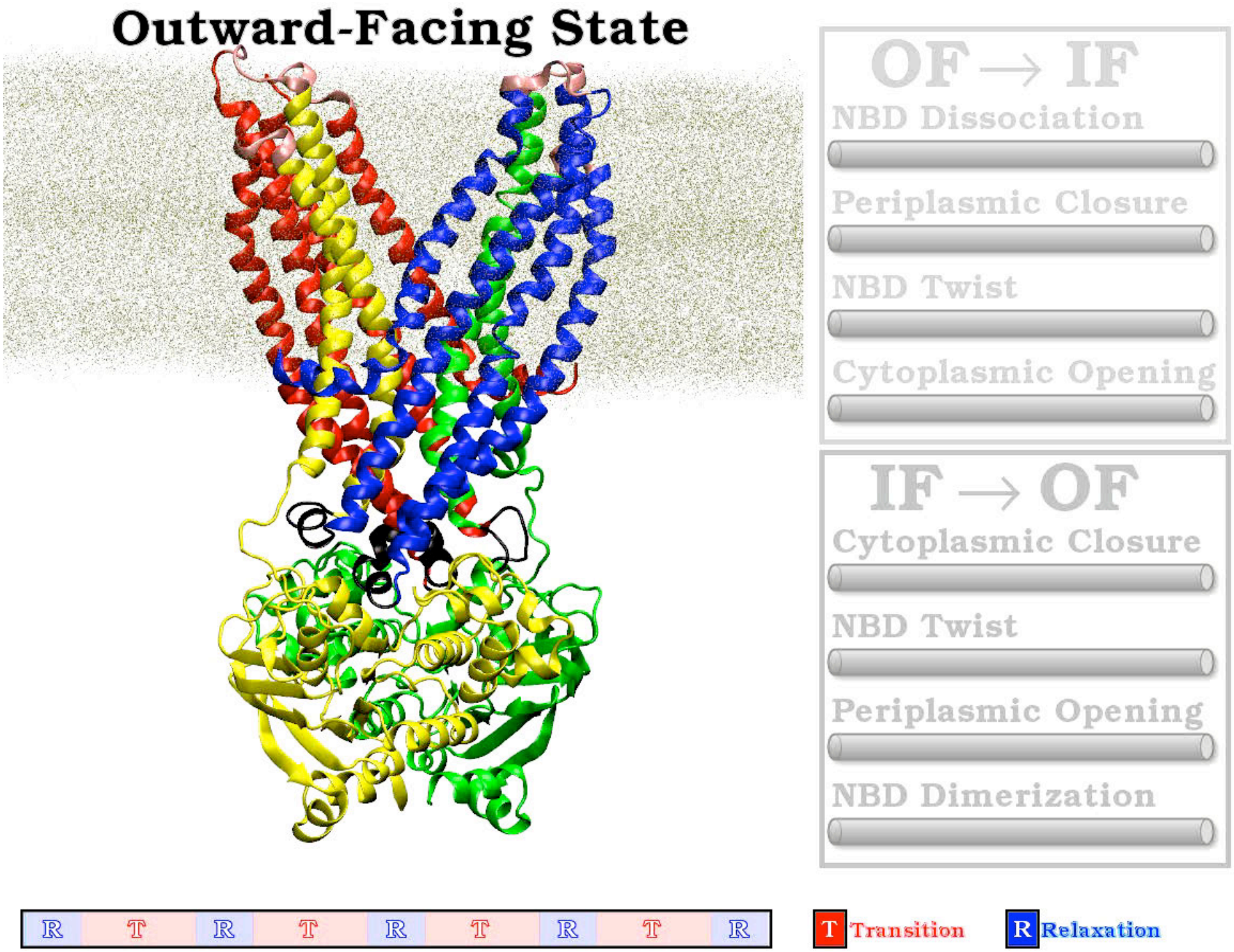
Mechanistic Insight From Transition Pathways in ABC exporters from Non-Equilibrium Simulations





M. Moradi and ET (2013) **PNAS**, 110:18916–18921.

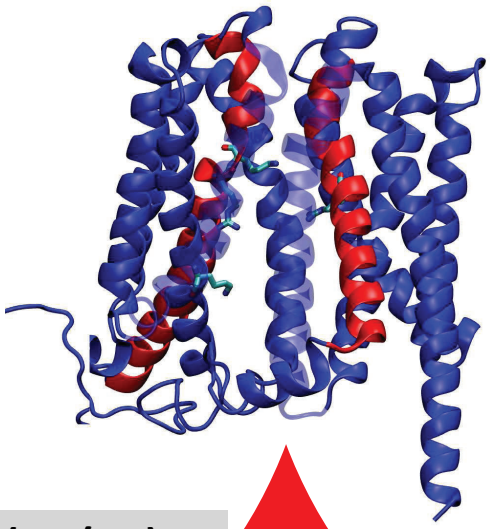
M. Moradi and ET (2014) **JCTC**, 10: 2866–2880.



NBD Doorknob Mechanism

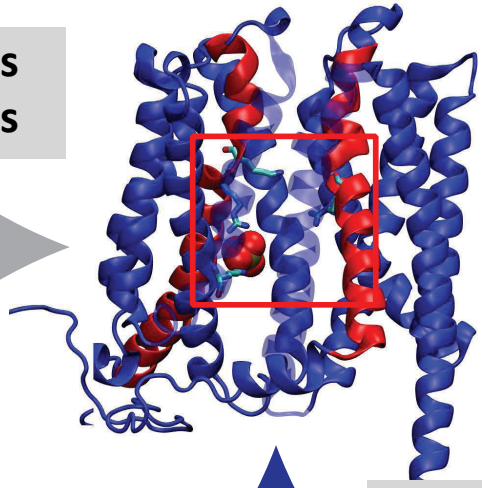
M. Moradi and ET (2013) **PNAS**, 110:18916–18921.

Describing a Complete Cycle (Adding Substrate) Requiring a Combination of Multiple Collective Variables



IF_{apo}

30 r x 20 ns
30 r x 20 ns

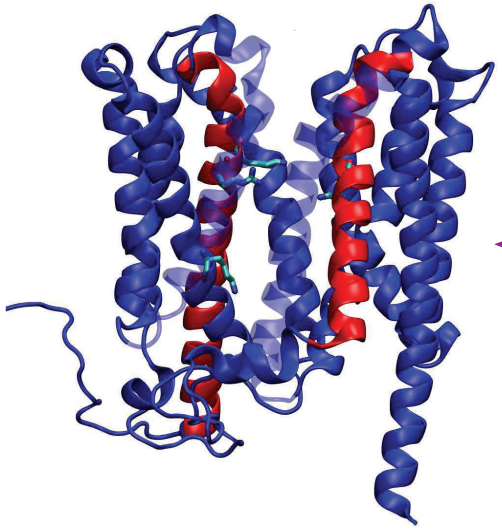


IF_{bound}

12 replicas x 40 ns (H1/H7)
50 replicas x 20 ns (10 Hs)

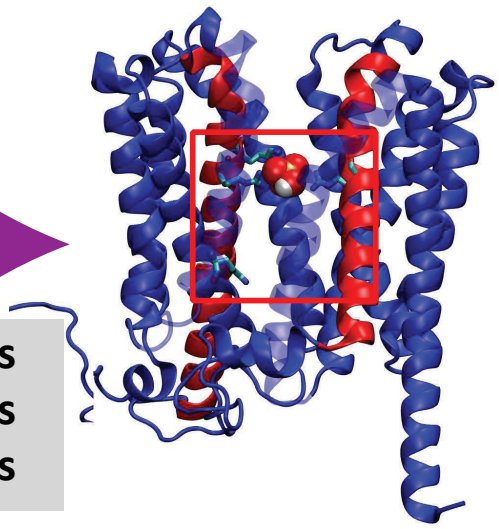
150 replicas

12 replicas x 40 ns (H1/H7)
24 replicas x 20 ns (H1/H7)
200 replicas (2D) x 5 ns
50 replicas x 20 ns



OF_{apo}

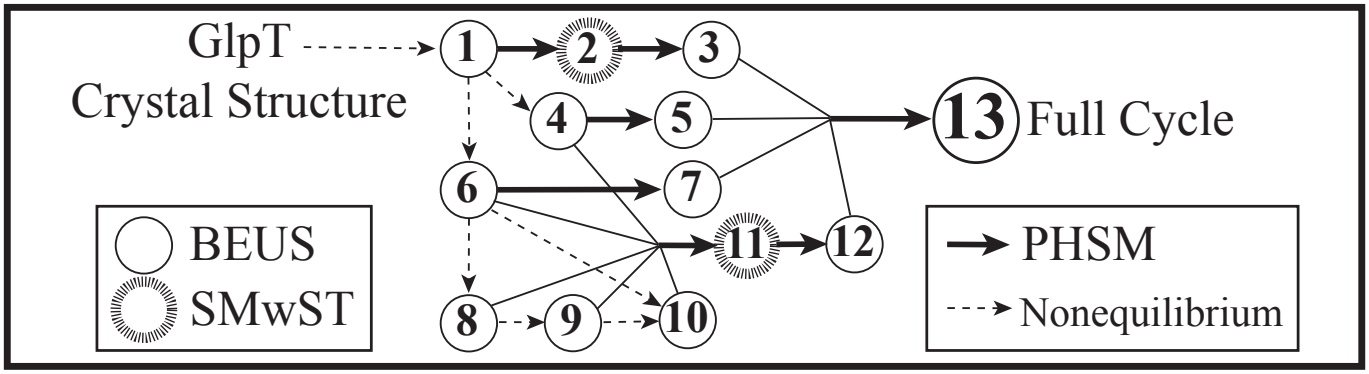
30 r x 20 ns
30 r x 20 ns
30 r x 20 ns



OF_{bound}

Simulation protocols

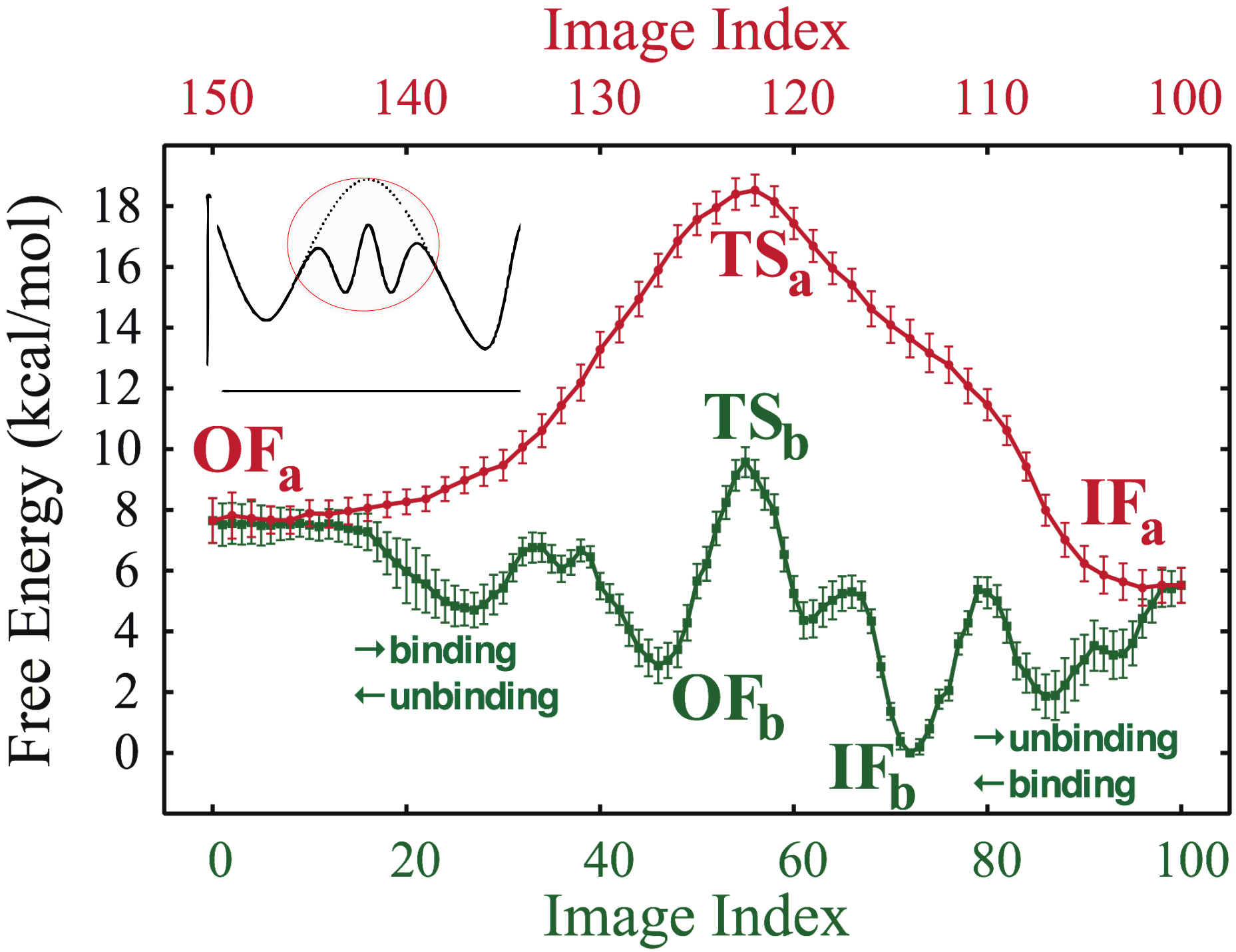
	Transition	Technique	Collective Variables	# of Replicas × Runtime		
1	$IF_a \leftrightarrow OF_a$	BEUS	(Q_1,Q_7)	12 × 40 ns	=	0.5 µs
2		SMwST	{Q}	1000 × 1 ns	=	1 µs
3		BEUS	{Q}	50 × 20 ns	=	1 µs
4	$IF_a \leftrightarrow IF_b$	BEUS	Z_{Pi}	30 × 40 ns	=	1.2 µs
5		BEUS	$(\{Q\}, Z_{Pi})$	30 × 40 ns	=	1.2 µs
6	$OF_a \leftrightarrow OF_b$	BEUS	Z_{Pi}	30 × 40 ns	=	1.2 µs
7		BEUS	$(\{Q\}, Z_{Pi})$	30 × 40 ns	=	1.2 µs
8	$IF_b \leftrightarrow OF_b$	BEUS	(Q_1,Q_7)	24 × 20 ns	=	0.5 µs
9		BEUS	Z_{Pi}	15 × 30 ns	=	0.5 µs
10		2D BEUS	$(\Delta RMSD, Z_{Pi})$	200 × 5 ns	=	1 µs
11		SMwST	$(\{Q\}, Z_{Pi})$	1000 × 1 ns	=	1 µs
12		BEUS	$(\{Q\}, Z_{Pi})$	50 × 20 ns	=	1 µs
13	**Full Cycle**	**BEUS**	**$(\{Q\}, Z_{Pi})$**	**150 × 50 ns**	=	**7.5 µs**
	Total Simulation Time					**18.7 µs**



U.S. DEPARTMENT OF ENERGY
INCITE
LEADERSHIP COMPUTING

BLUE WATER NCSA

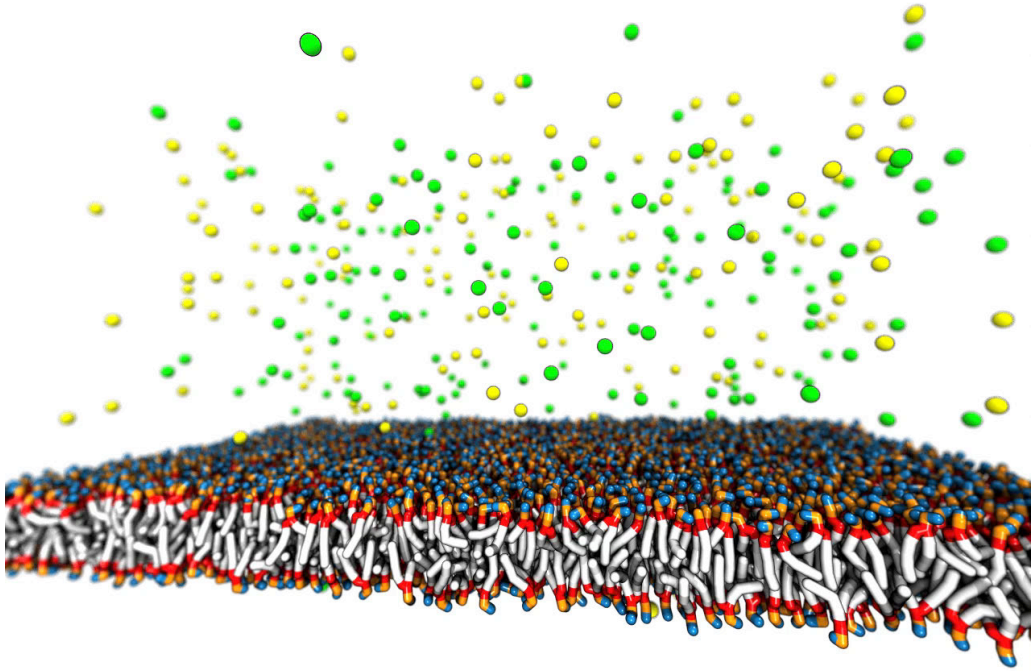
M. Moradi, G. Enkavi, and ET (2015) **Nature Communication**, 6: 8393.



M. Moradi, G. Enkavi, and ET (2015) **Nature Communication**, 6: 8393.

Battling the Timescale - Case III
Multiscale Simulations

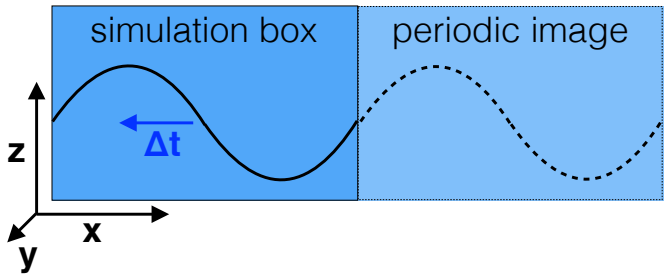
Membrane Budding/Fusion

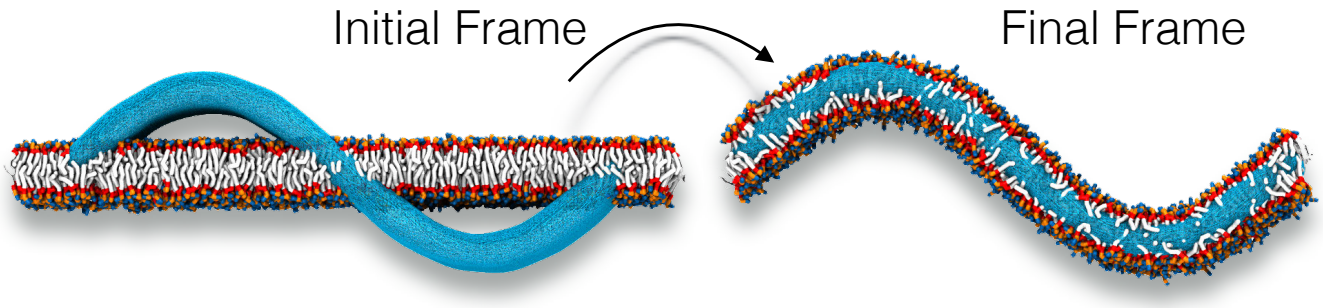


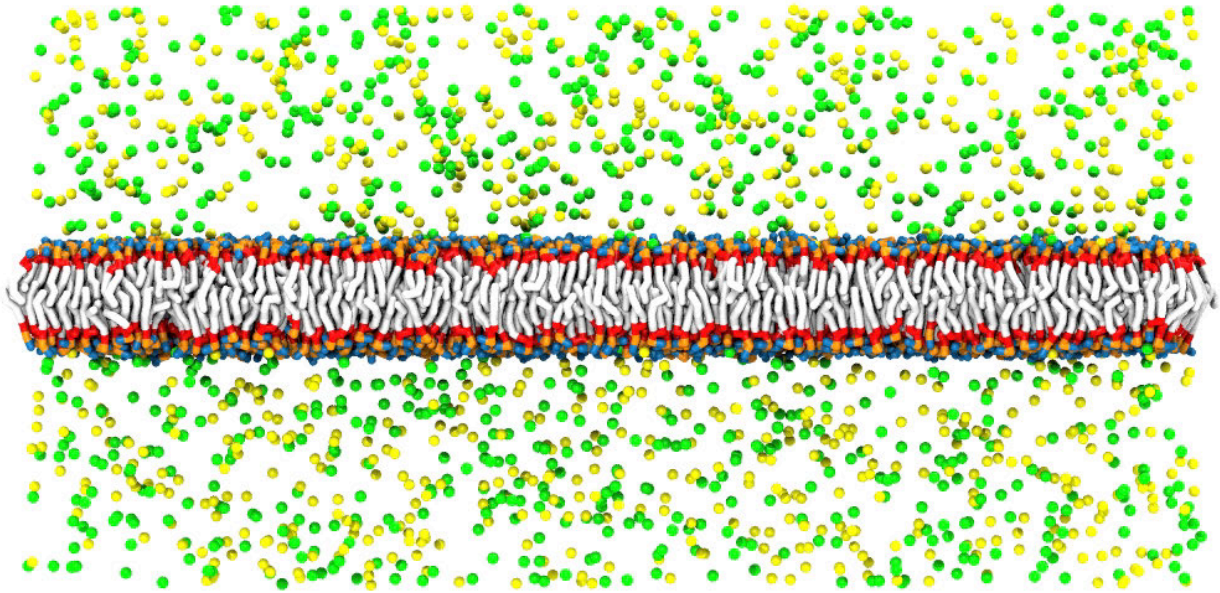
Combining multiple replica simulations and coarse-grained models to describe membrane fusion

Workflow for Multi-Scale Modeling

Parametrically Defined Sine Function

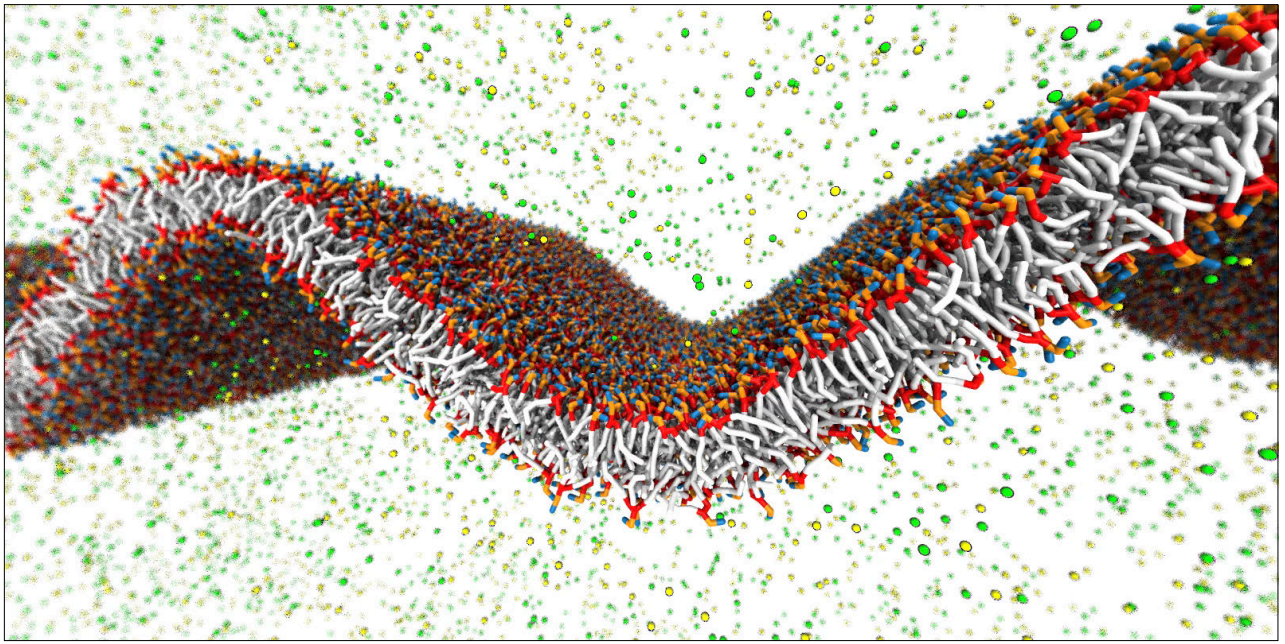






Christopher Mayne, Tajkhorshid Lab

Workflow for Multi-Scale Modeling

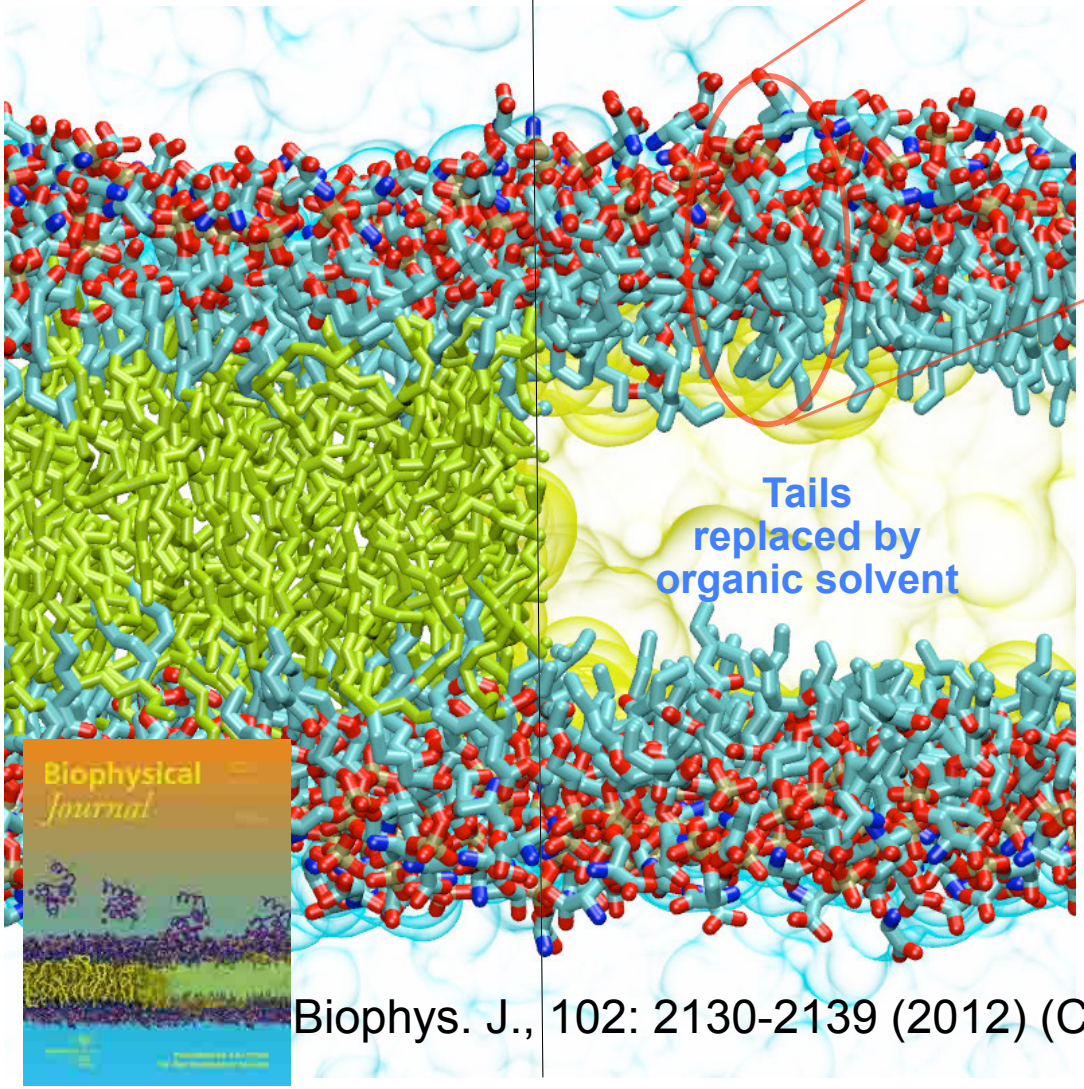


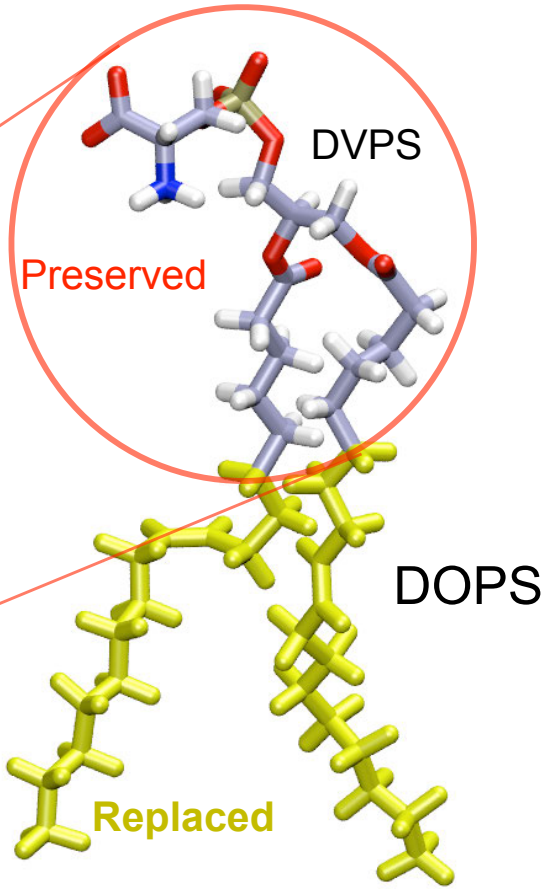
Christopher Mayne, Tajkhorshid Lab

Battling the Timescale - Case IV
Reduced Representations

Highly Mobile Membrane Mimetic model

Full model | **HMMM model**





Advantages

Increased mobility of lipids

Retain explicit headgroups allowing for atomic details

Biophys. J., 102: 2130-2139 (2012) (Cover Article)

Zenmei Ohkubo

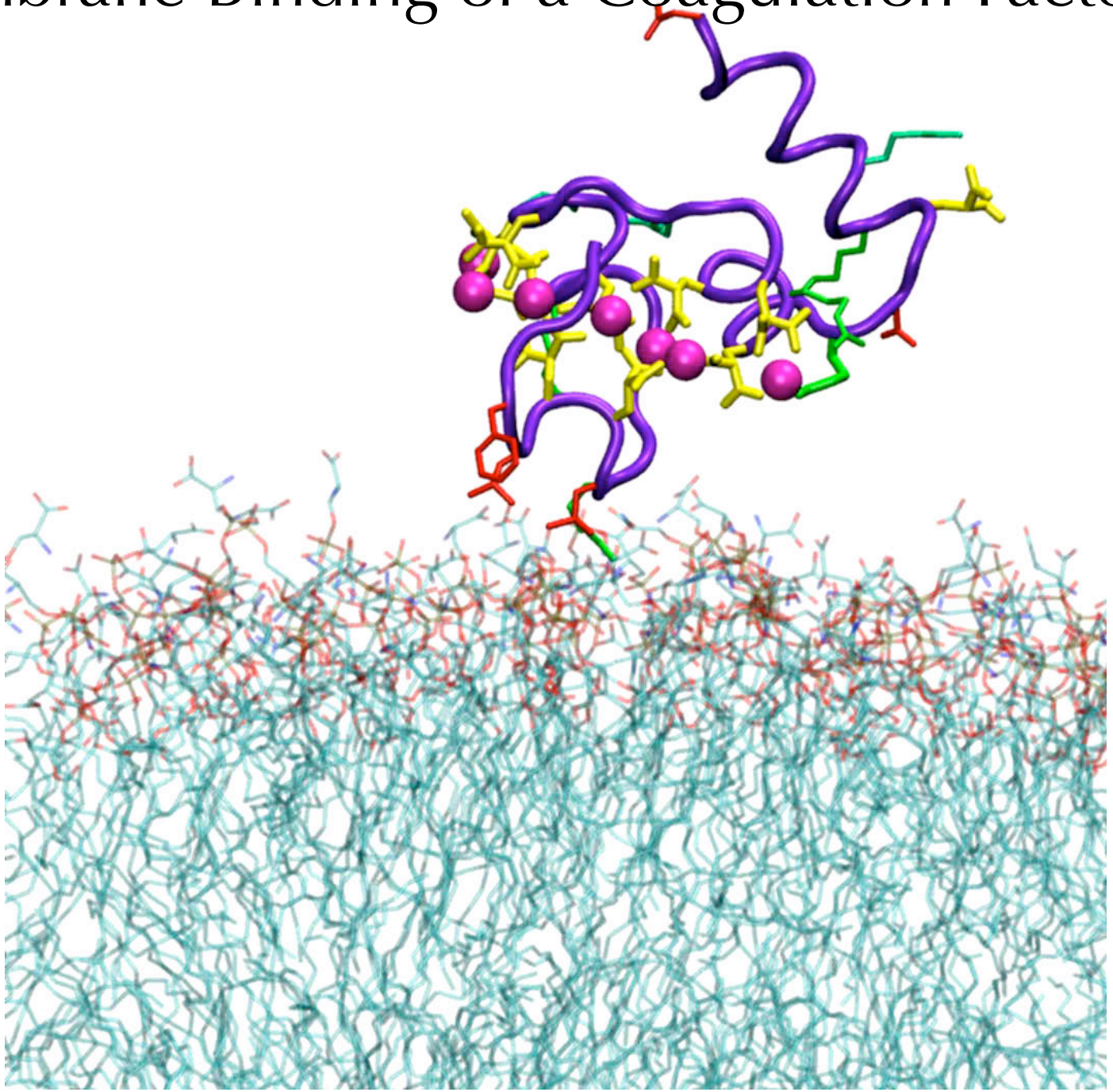
Mark Arcario

Taras Pogorelov

Josh Vermaas

Javier Baylon

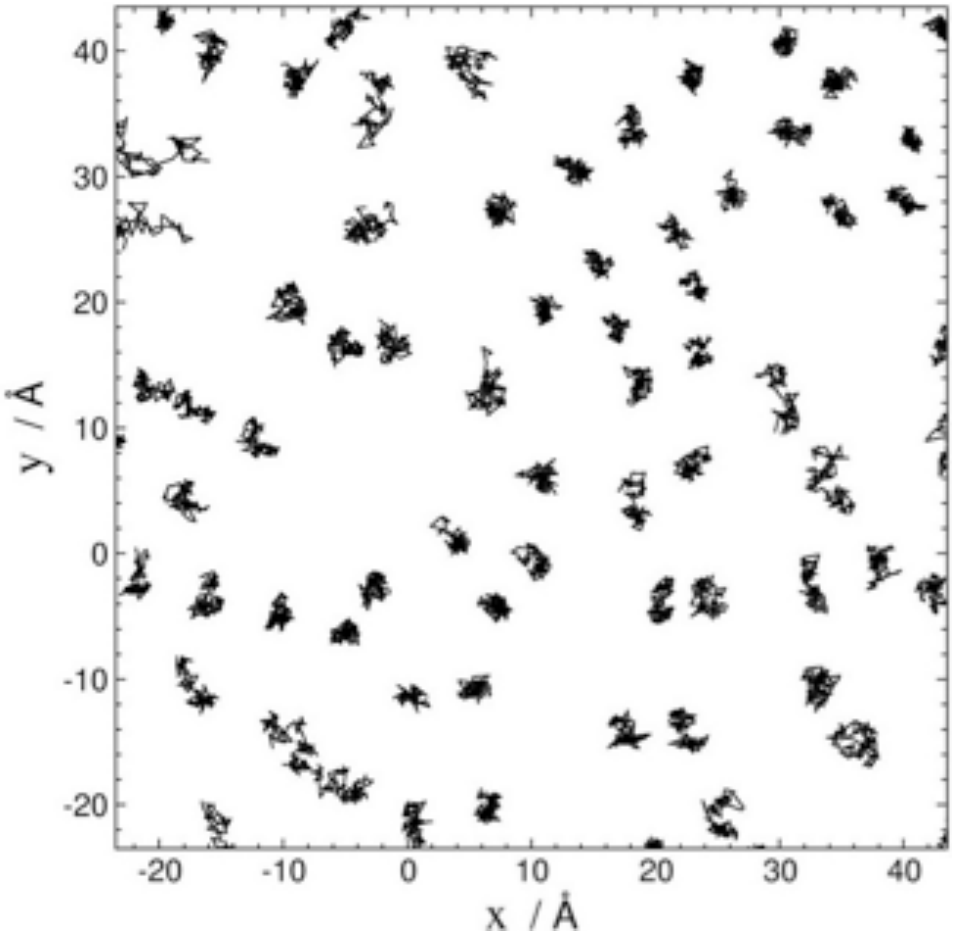
Membrane Binding of a Coagulation Factor



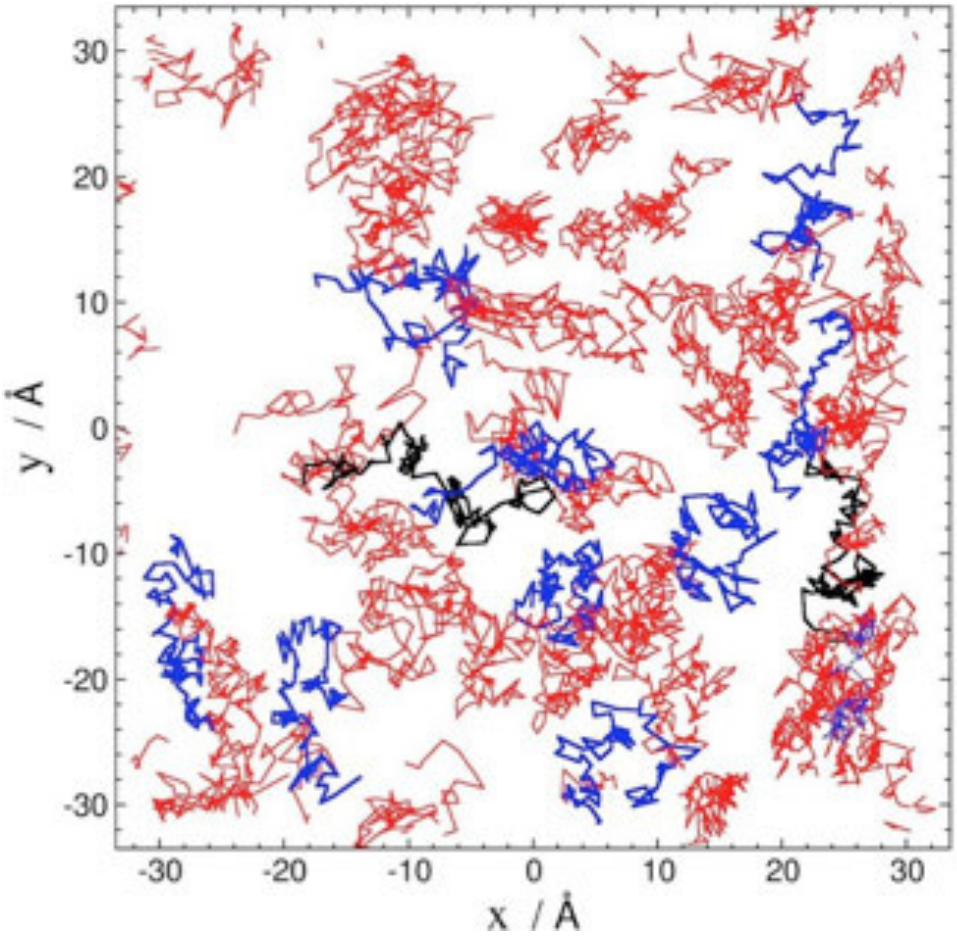
Z. Ohkubo and E. Tajkhorshid, Structure **2008**

Enhanced Lipid Lateral Diffusion

Without Compromising Atomic Details of the Headgroups

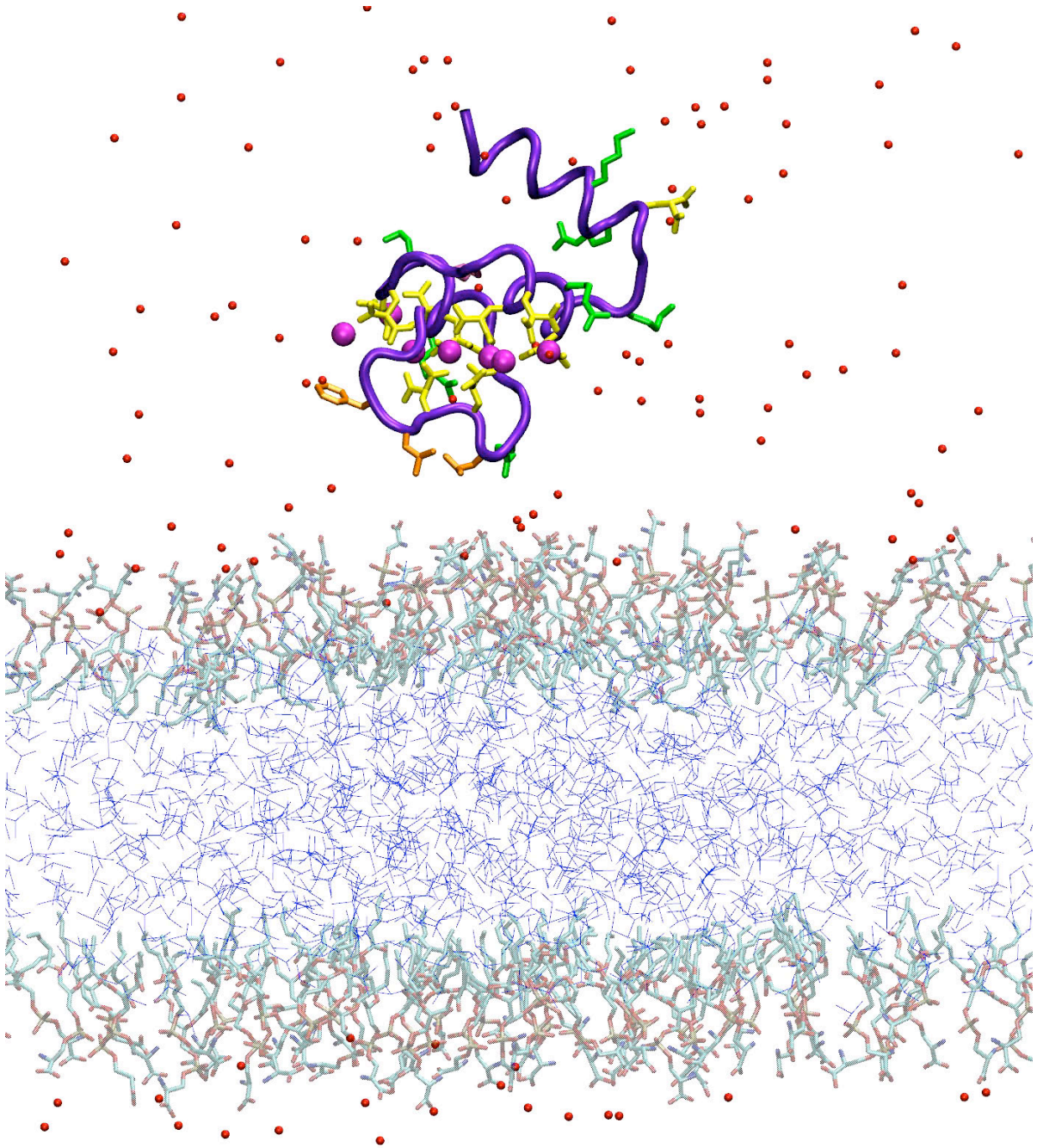


Conventional membrane (**10 ns**)



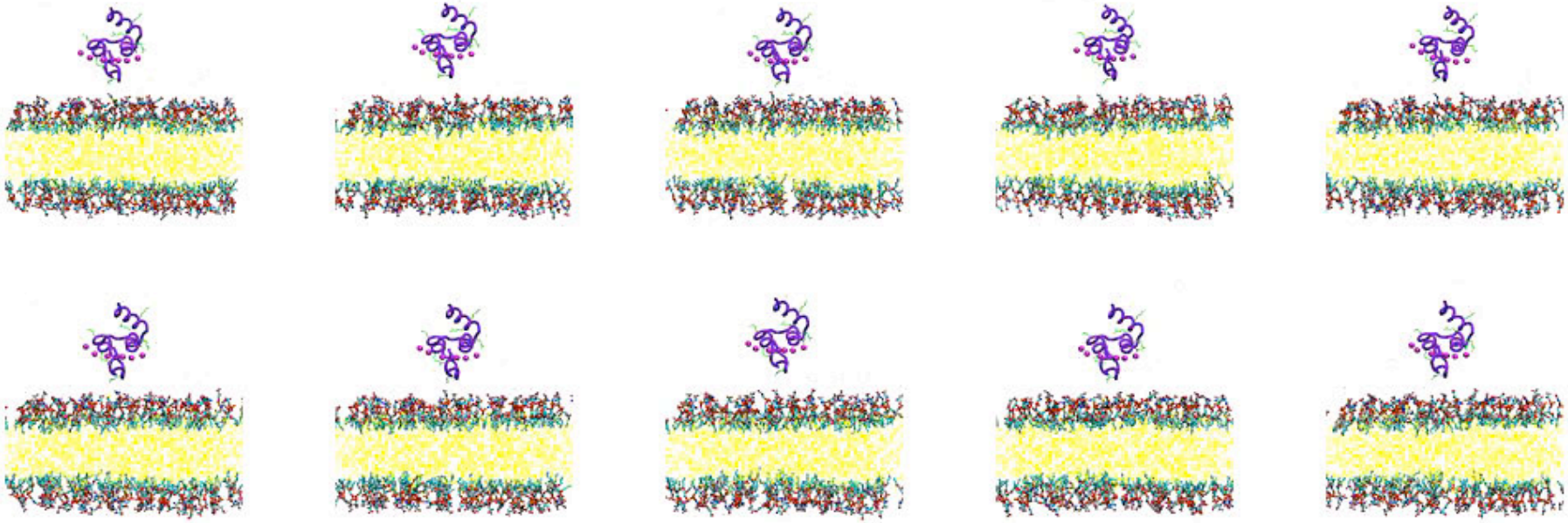
HMMM membrane (**1 ns**)

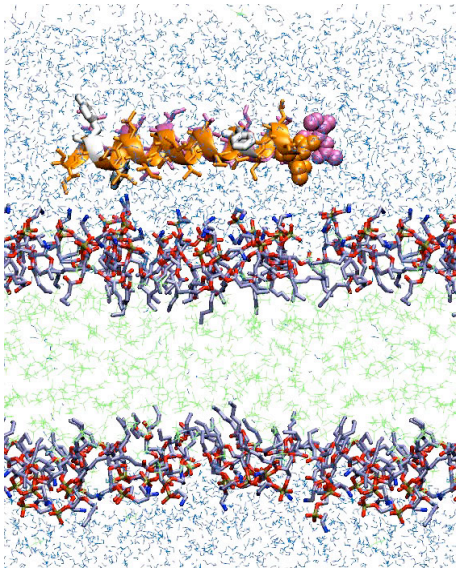
PS-Dependent Spontaneous Insertion of FVII-GLA

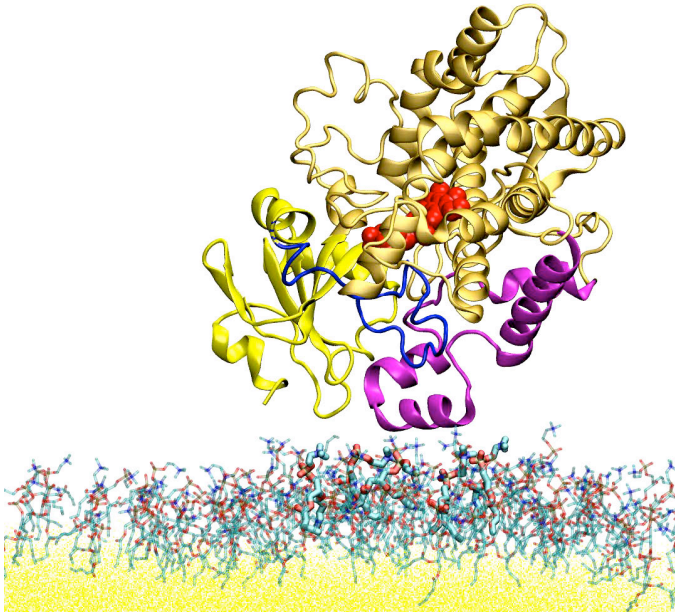


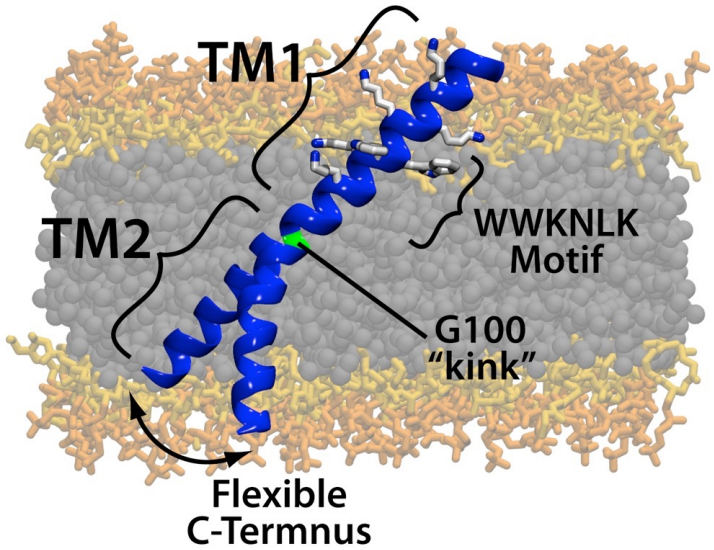
Zenmei Ohkubo

HMMM - More Efficient Computational Model for Membrane Proteins





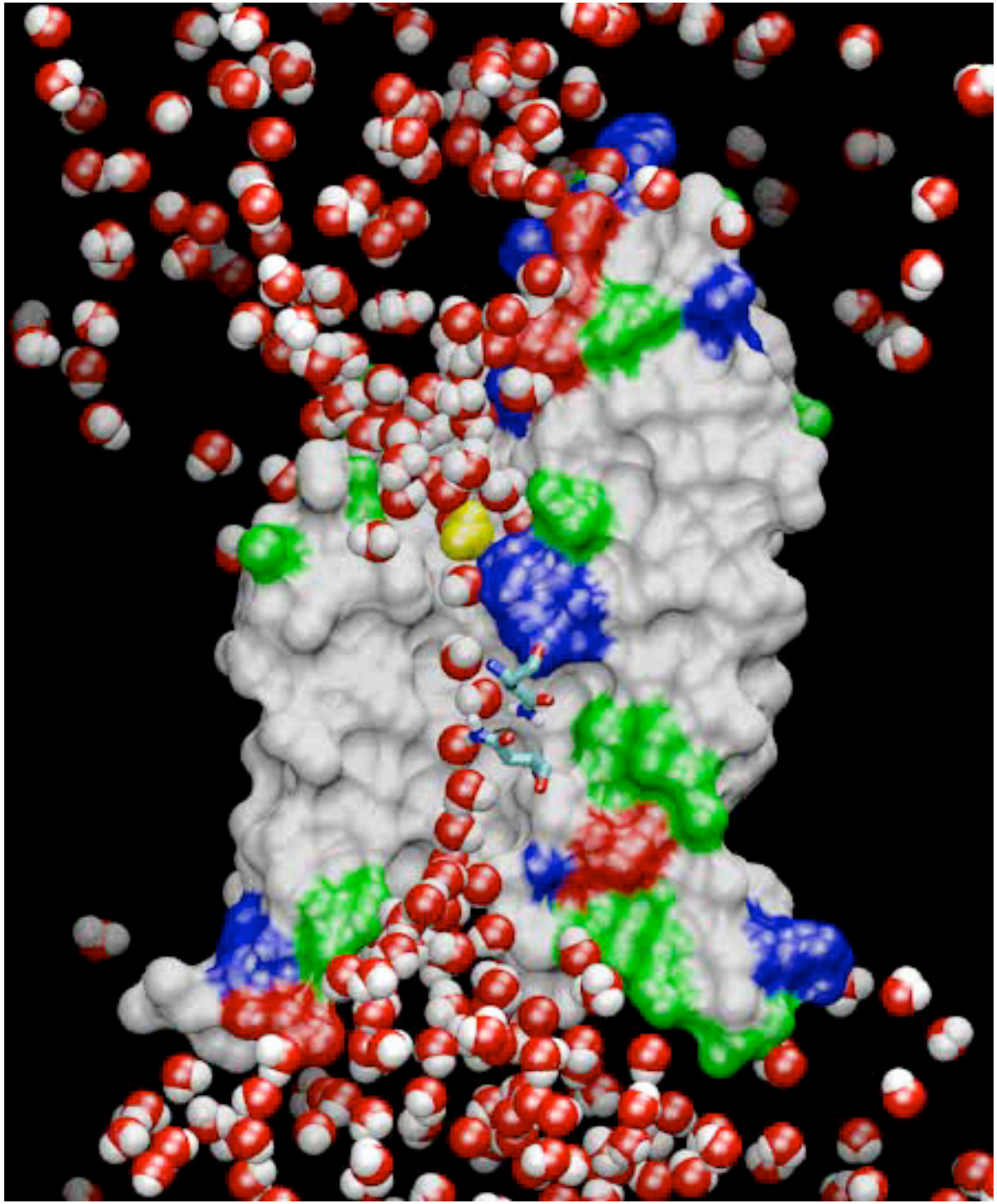




R E M E M B E R:

One of the most useful advantages of simulations over experiments is that you can modify the system as you wish: You can do modifications that are not even possible at all in reality!

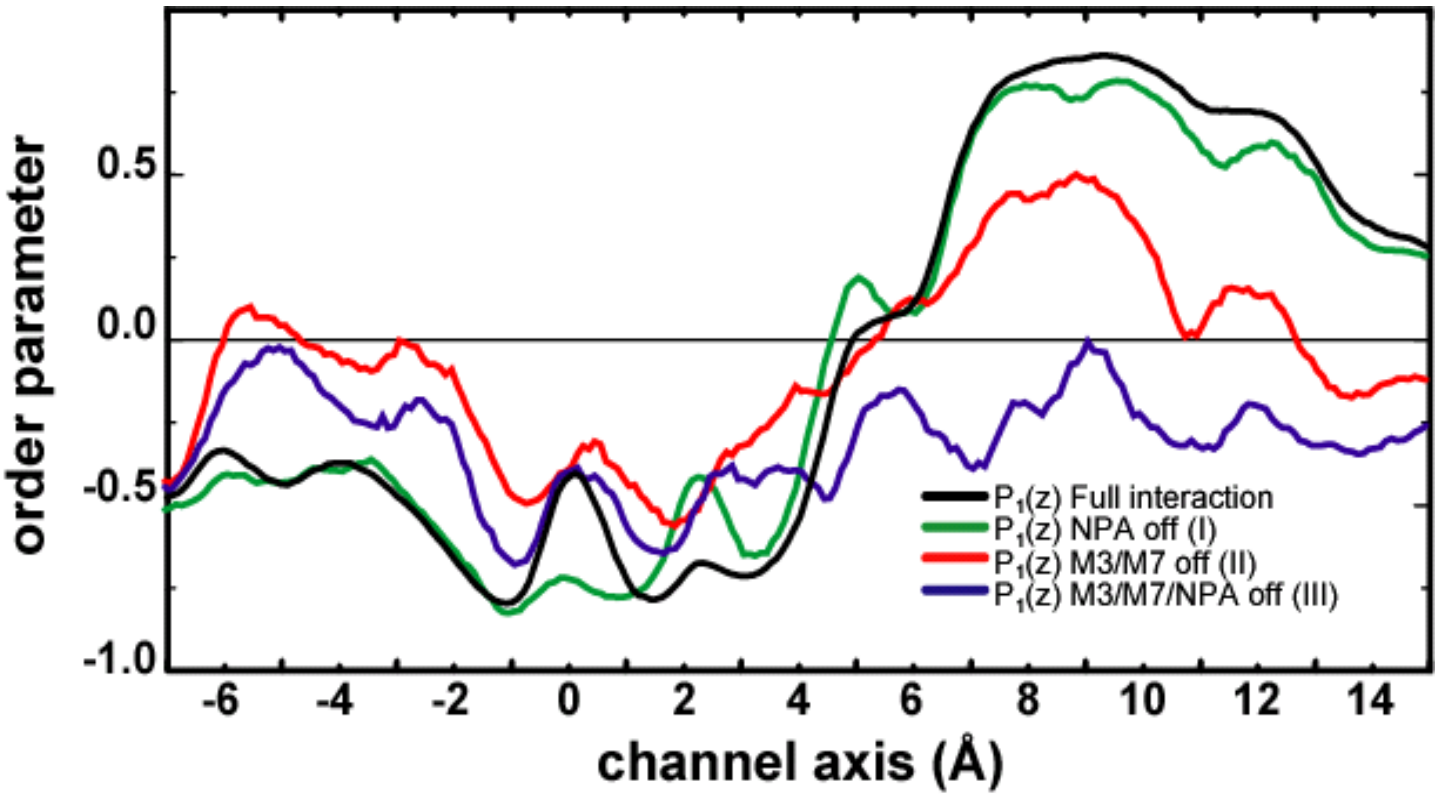
This is a powerful technique to test hypotheses developed during your simulations. Use it!

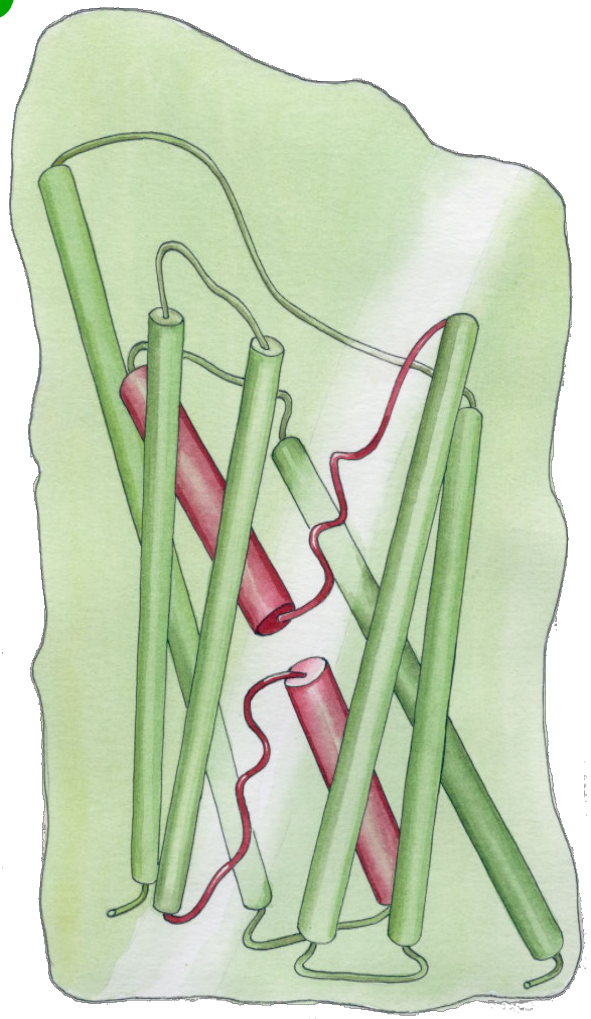


Animation available at the Nobel web site

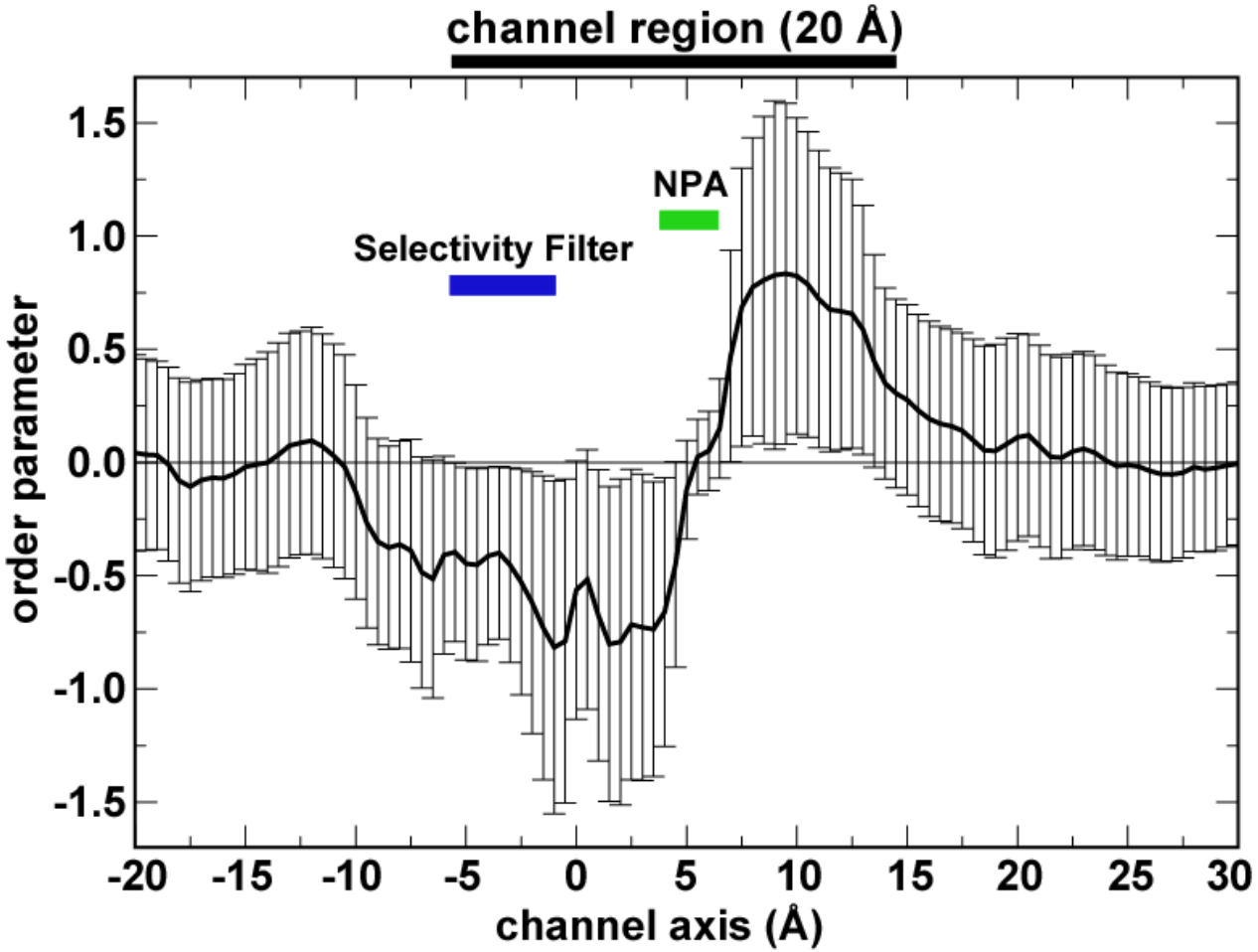
E. T., et al., ***Science*** 2002.

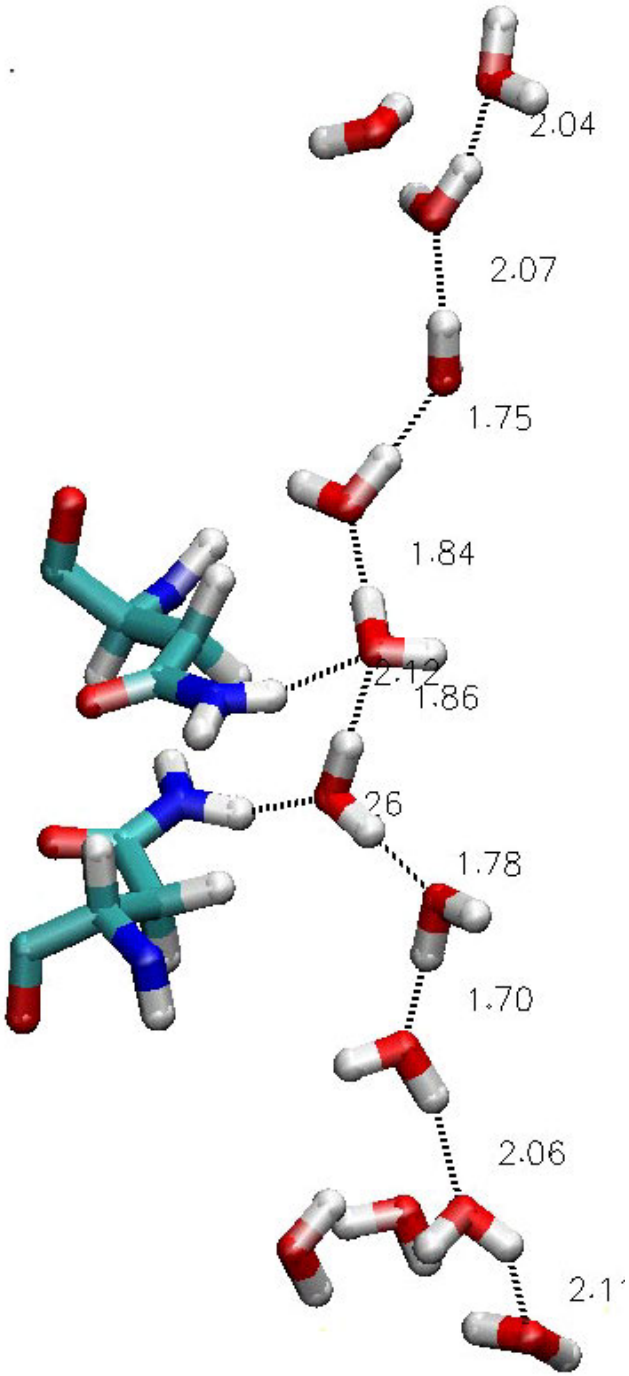
Electrostatic Stabilization of Water Bipolar Arrangement





E. T., et al., ***Science*** 2002.


channel region (20 Å)
NPA
Selectivity Filter
order parameter
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-20
-15
-10
-5
0
5
10
15
20
25
30
channel axis (Å)


2.04
2.07
1.75
1.84
2.12
1.86
26
1.78
1.70
2.06
2.1