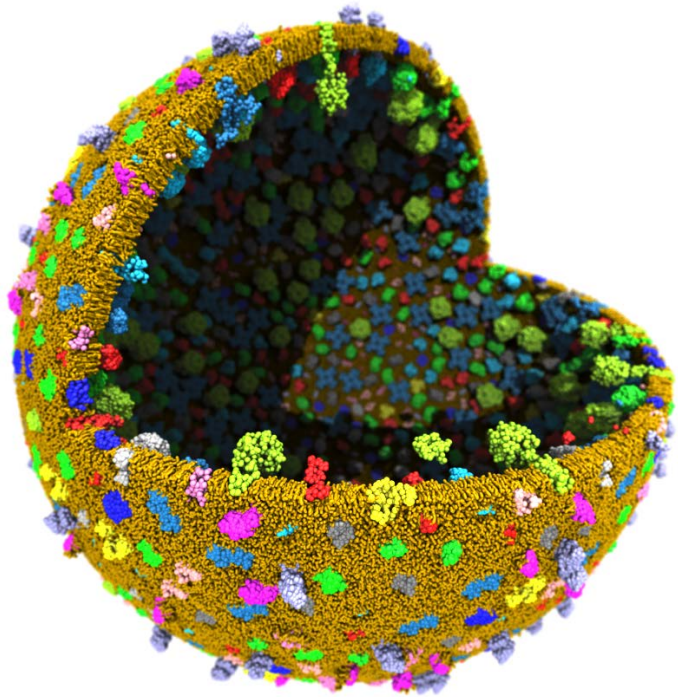
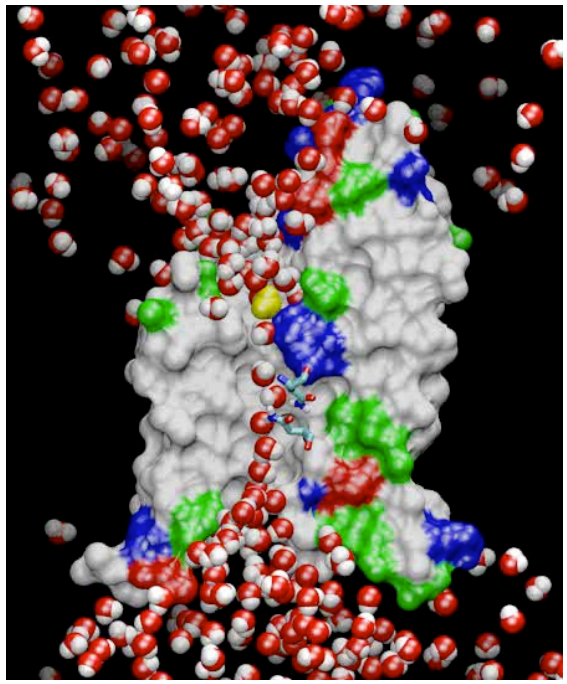
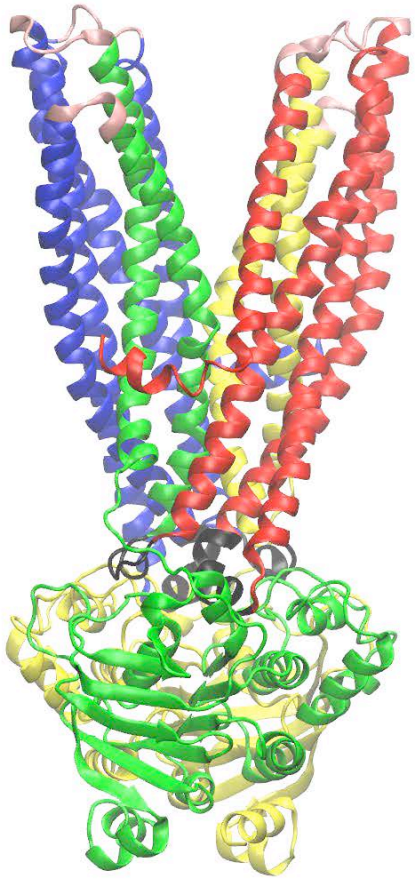
Enhanced Sampling and Free Energy Applications in Biomolecular Modeling

Emad Tajkhorshid

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



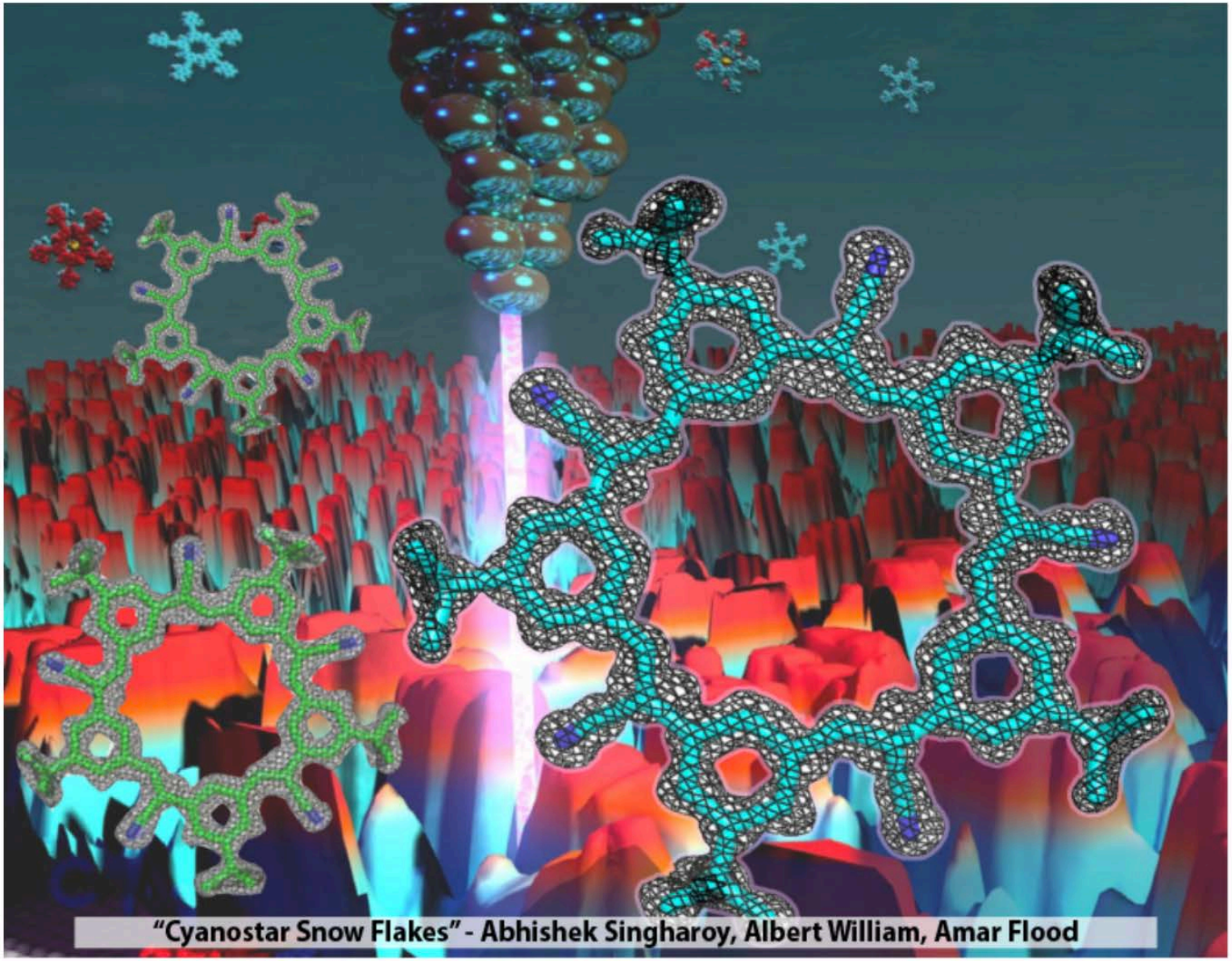




NIH-sponsored Workshop on
Theoretical and Computational Biophysics

2018

"Hands-on" Workshop on Enhanced Sampling and Free-Energy Calculation at Urbana, IL



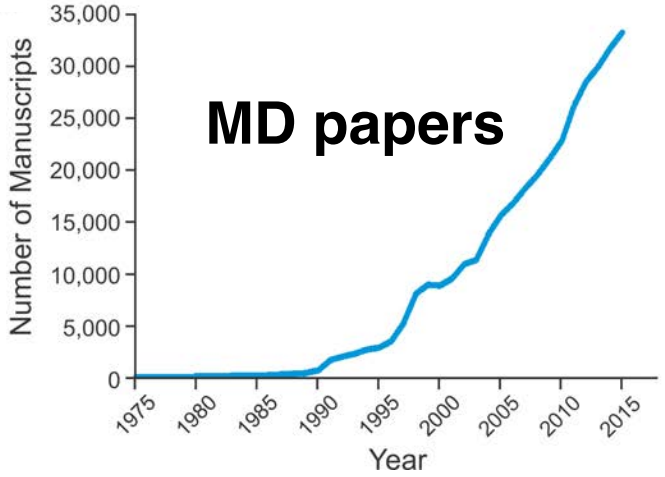
NIH P41 Biotechnology Center for Macromolecular Modeling and Bioinformatics

University of Illinois at Urbana-Champaign

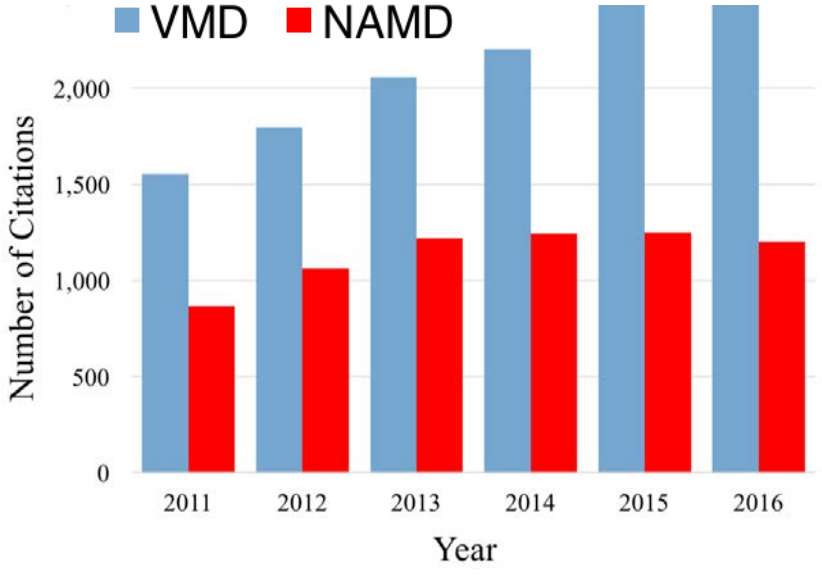
www.ks.uiuc.edu

NAMD
Scalable Molecular Dynamics

VMD
Visual Molecular Dynamics



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

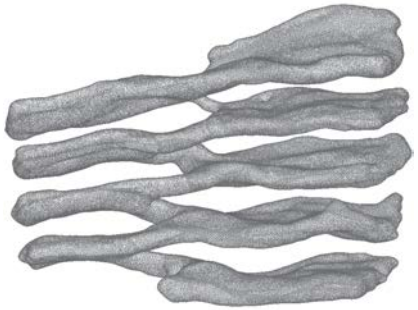


Serving a Large and Fast Growing Community

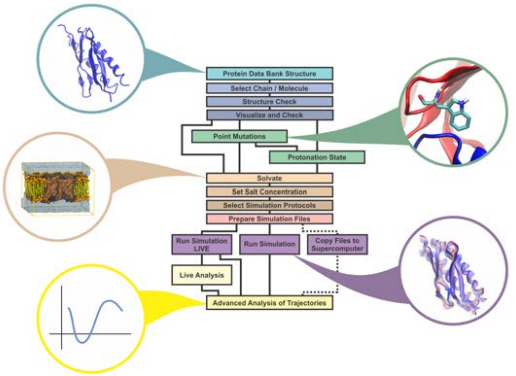
- Deploying Center's flagship programs NAMD and VMD on all major computational platforms from commodity computers to supercomputers

- 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

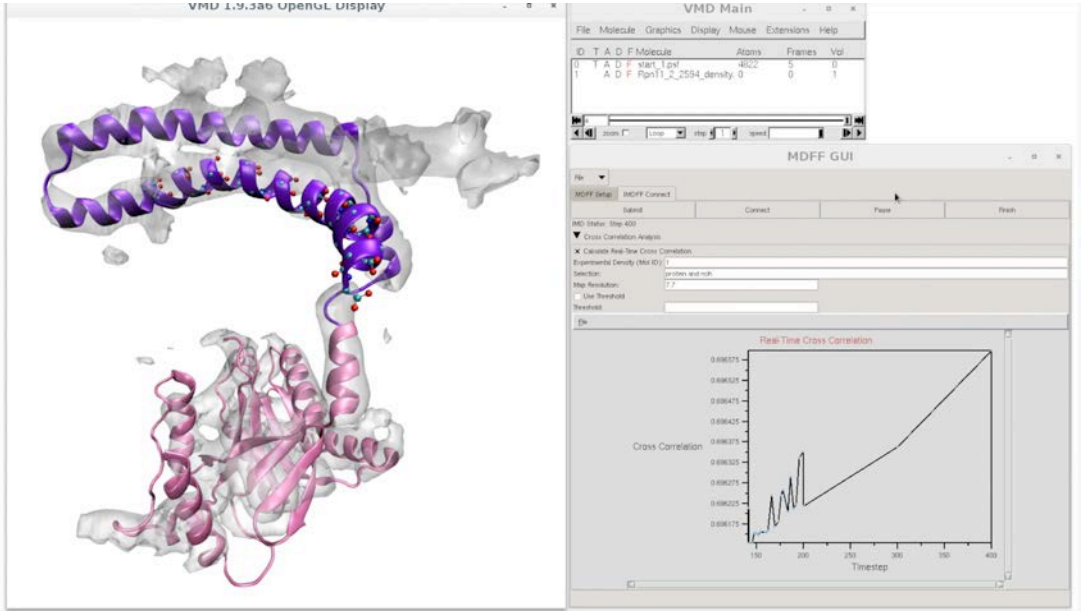


Exploiting State of the Art Hardware Technology

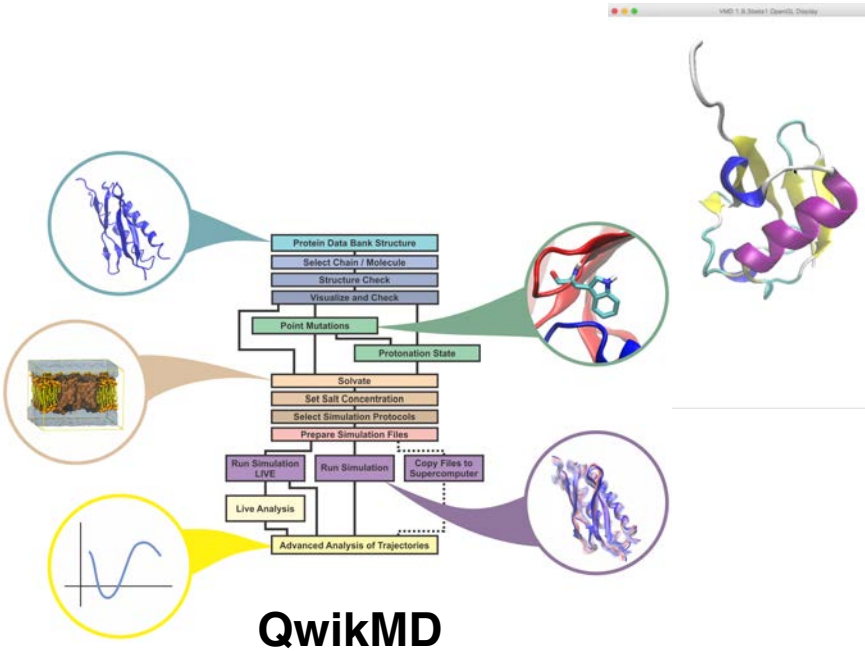
- Software available and optimized on all national supercomputing platforms (even before they come online)
- Decade-long, highly productive relationship with NVIDIA
- The first CUDA Center of Excellence funded by NVIDIA
- Consistently exploring opportunities for new hardware technology
  - Remote visualization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

Technology Made Highly Accessible to the Community

interactive MDFF

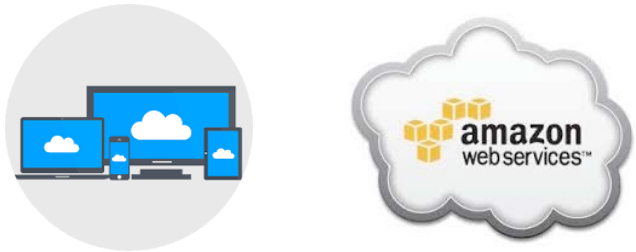


Developed primarily for experimental users



QwikMD

VMD Plugin for Setup and Analysis of **NAMD** Simulations



Vigorous Training Through Hands-On Workshops

53 Workshops on Computational Biophysics

- Online Workshops on Simulating Membrane Channels
- In-residence workshops for visiting researchers
- Local workshops on hardware and coding

1600+ Researchers Trained Since 2003

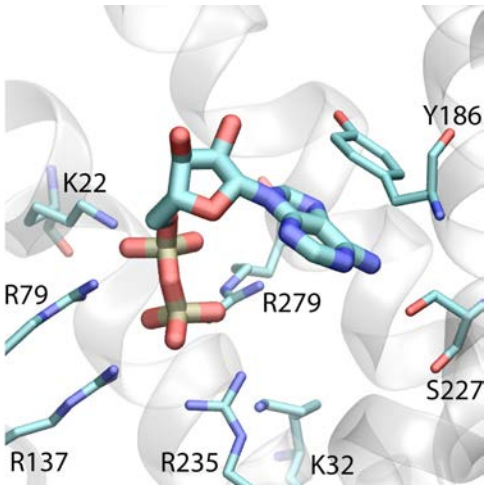
High school students to professional faculty
Computational to experimental backgrounds
National to international and minority communities

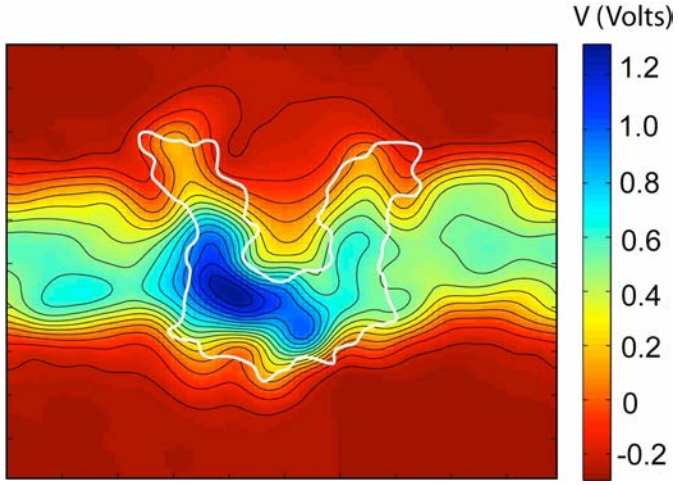
~2,000 Pages of Self-Study Tutorial Material

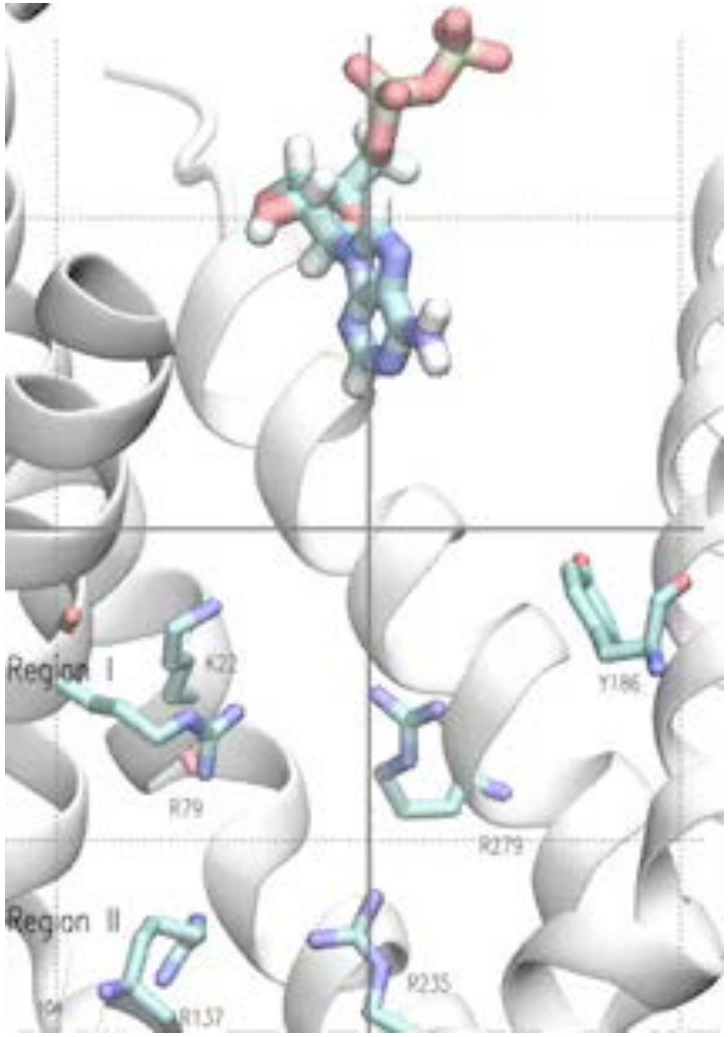
Slides, recorded lectures, and video tutorials also available

Microscopic View of Molecular Phenomena

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

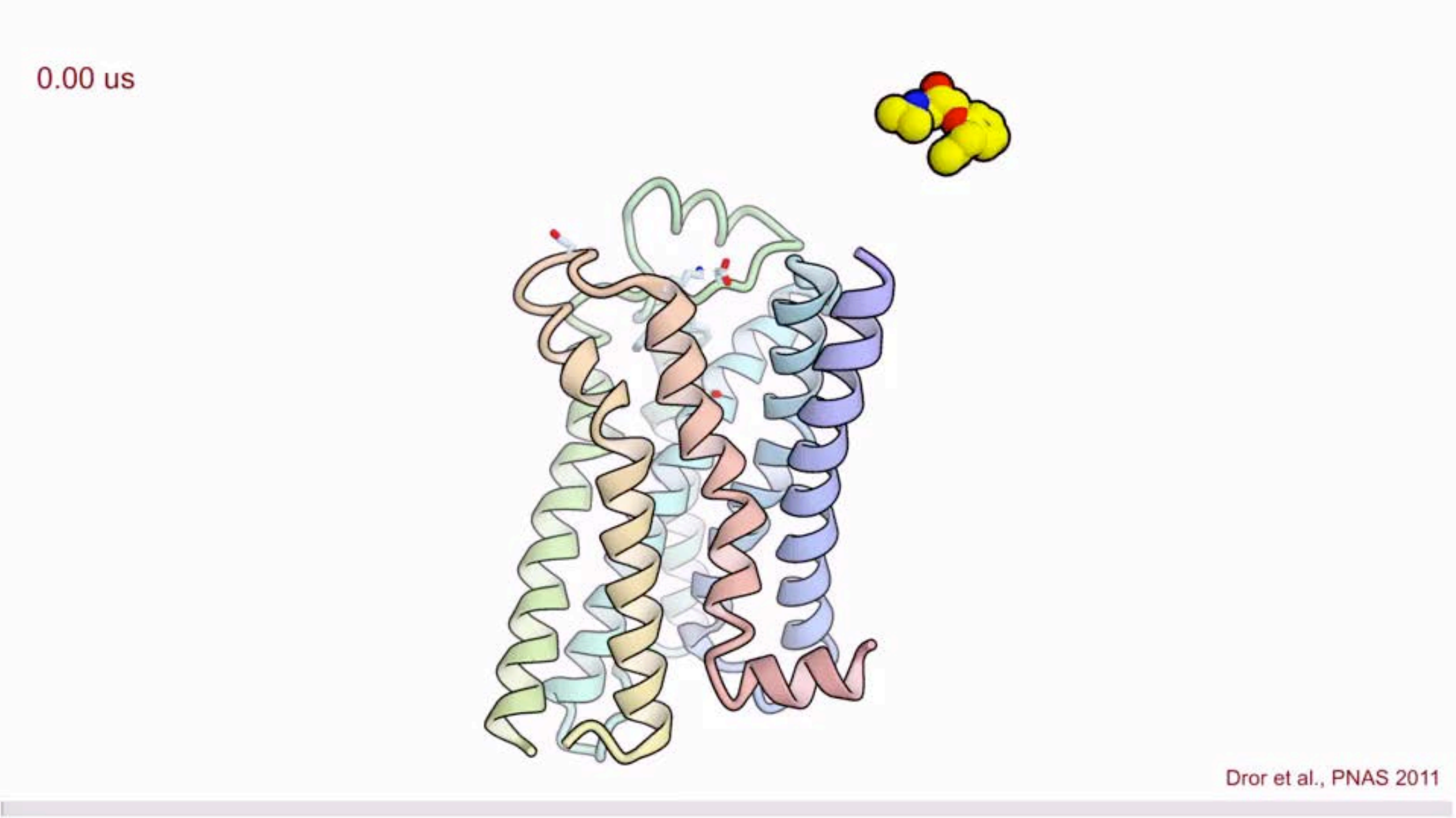






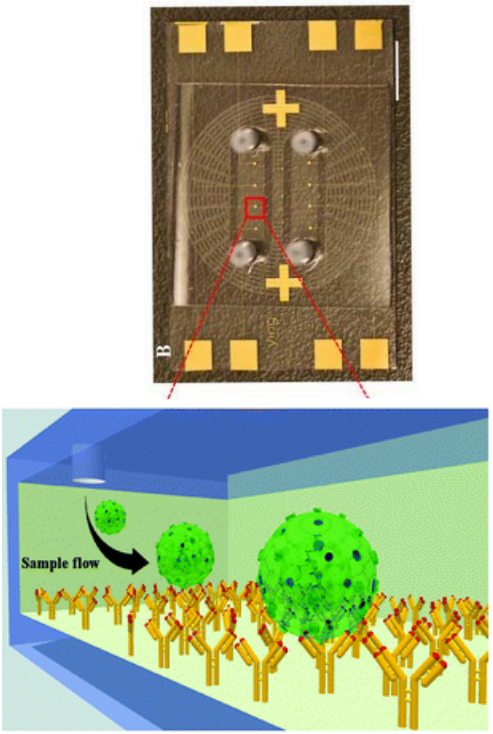
Binding of a small molecule to a binding site
Y. Wang & E.T. PNAS 2010

Microscopic View of Molecular Phenomena



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

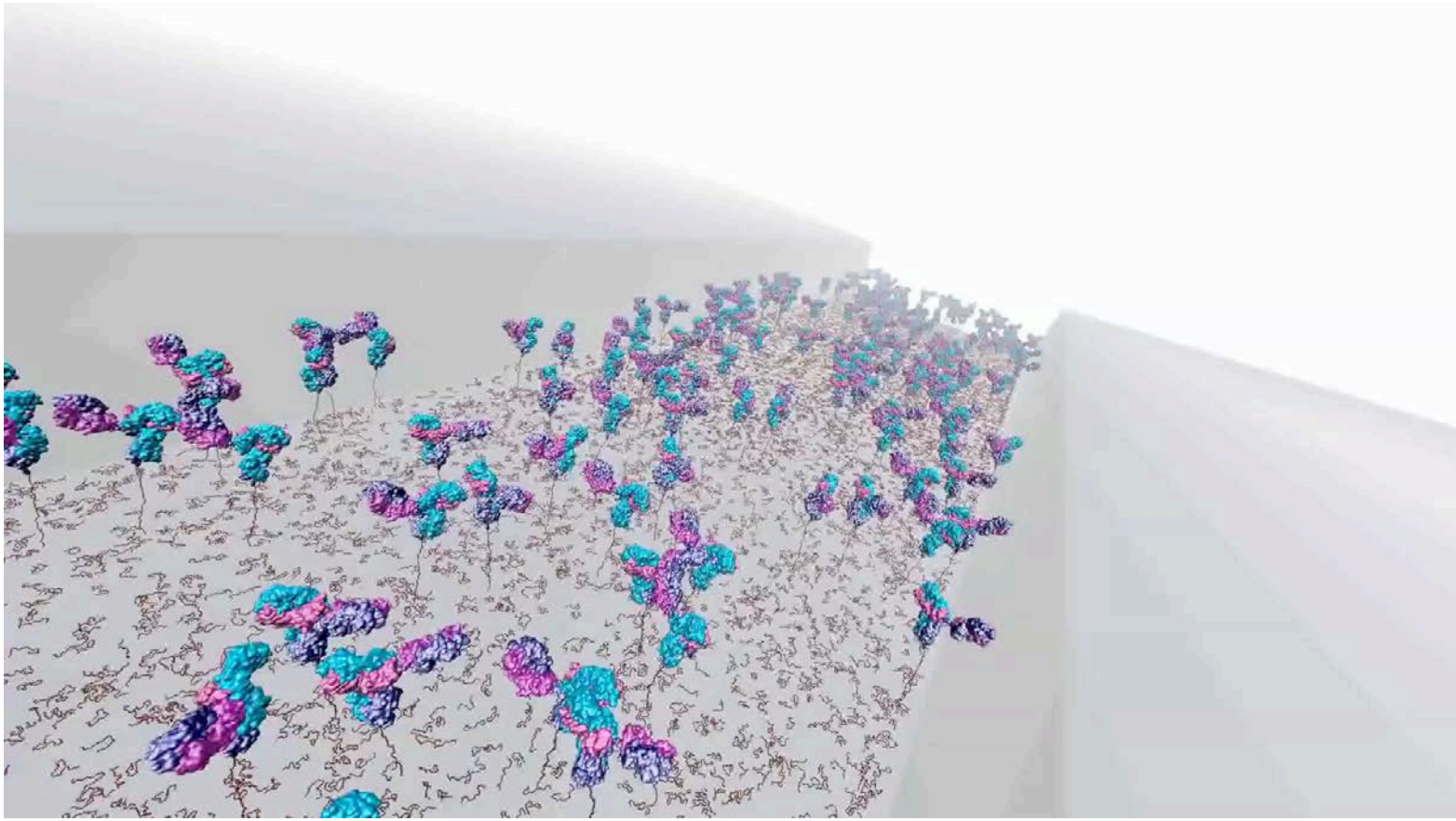
Microscopic View of Molecular Phenomena
Nano-biotechnology



HIV subtype identification

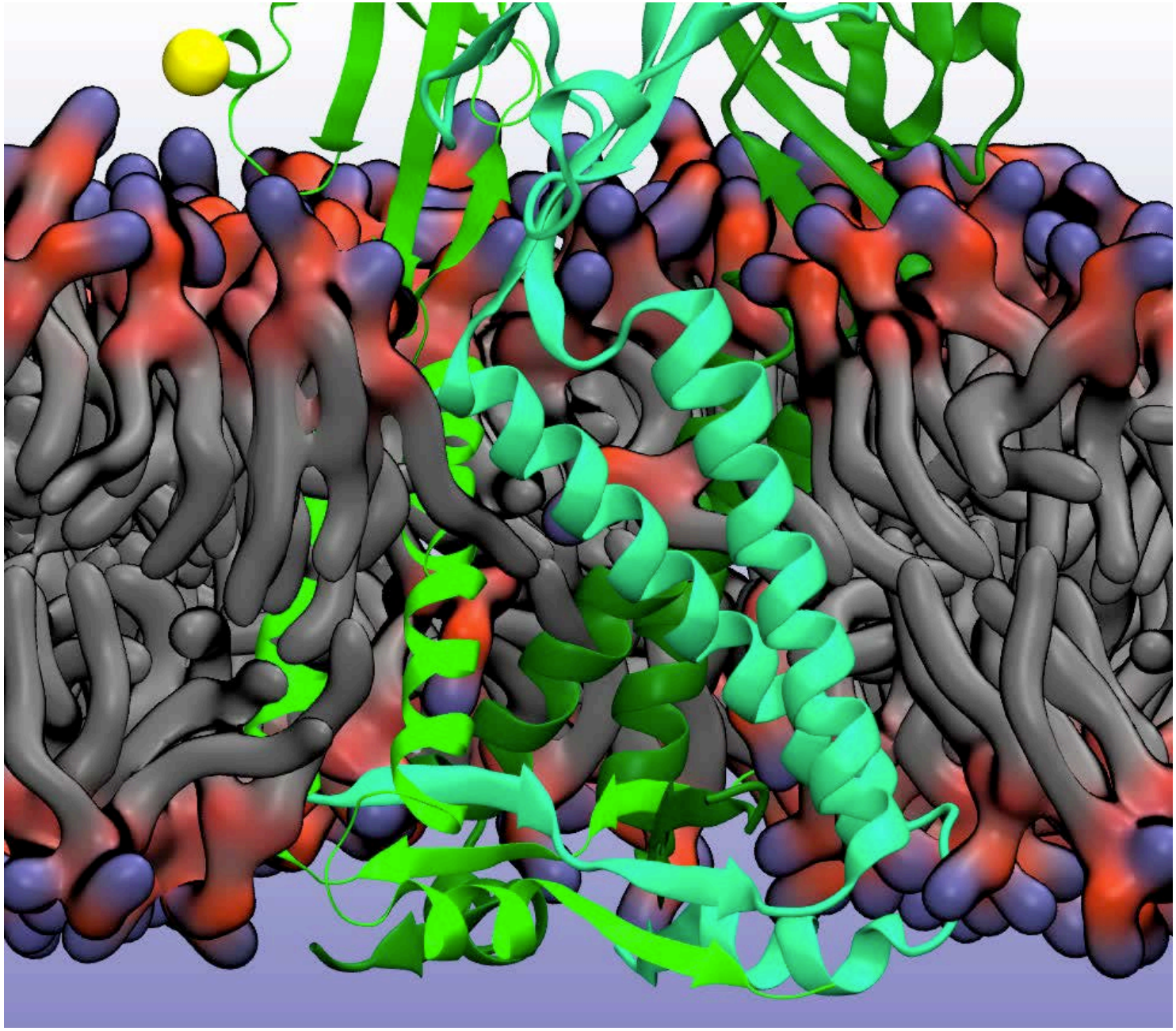
Lab Chip 2012

Functionalized nanosurface with antibodies



Created by **nanoBIO Node** tools

Most Detailed and Dynamic Microscopic View



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

Battling the Timescale

non-Equilibrium MD simulations

Free Energy Methods

Enhanced Sampling Techniques

Battling the Timescale - Case I

Steered Molecular Dynamics is a non-equilibrium method by nature

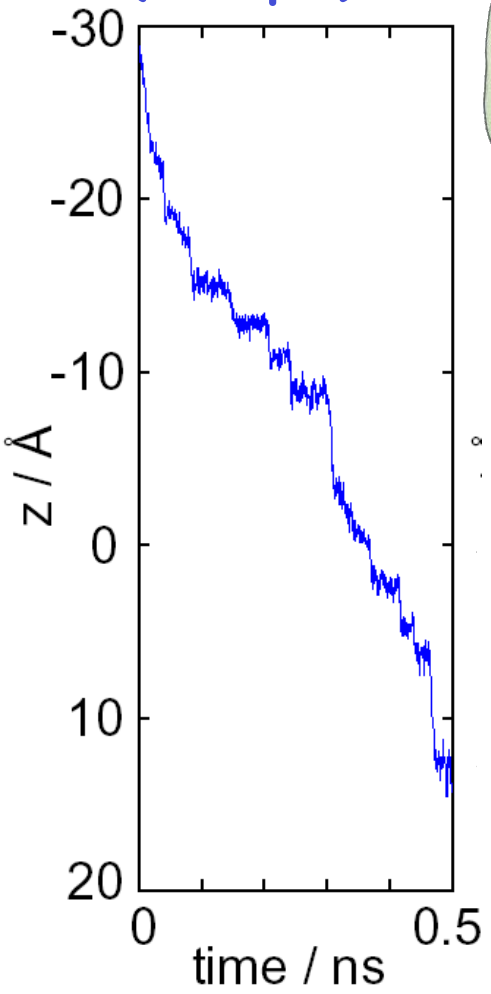
- A wide variety of events that are inaccessible to conventional molecular dynamics simulations can be probed.

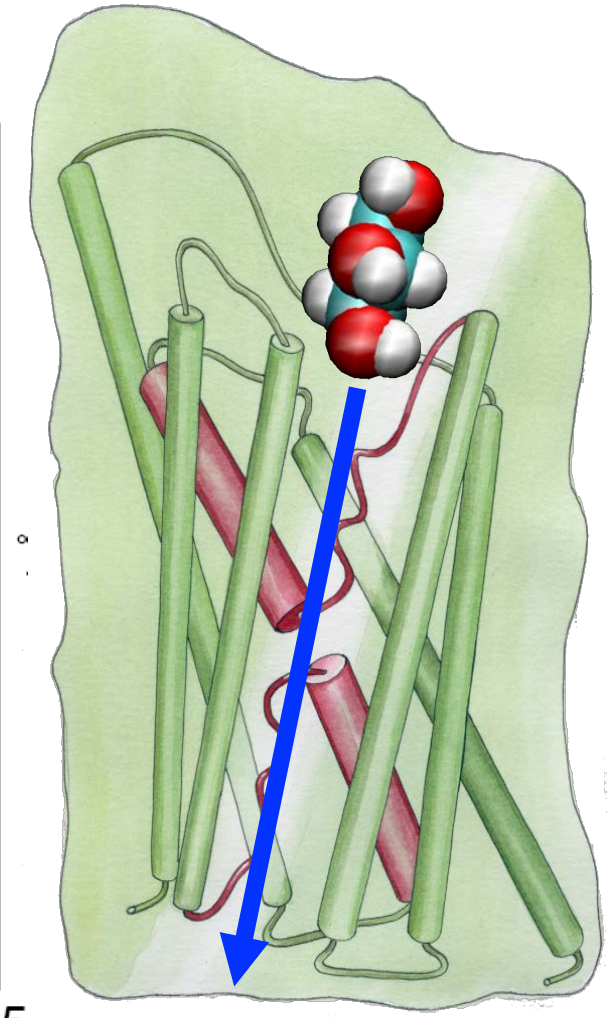
- The system will be driven, however, away from equilibrium, resulting in problems in describing the energy landscape associated with the event of interest.

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

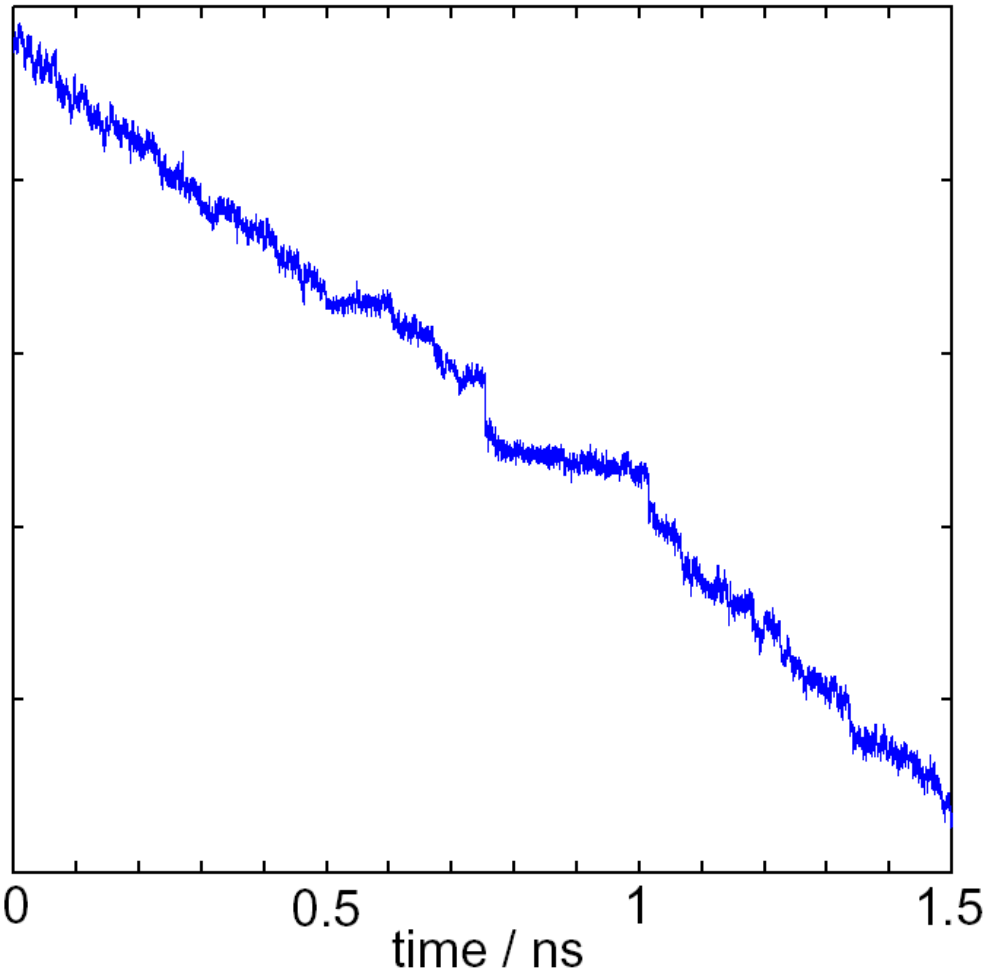
Steered Molecular Dynamics

constant force
(250 pN)



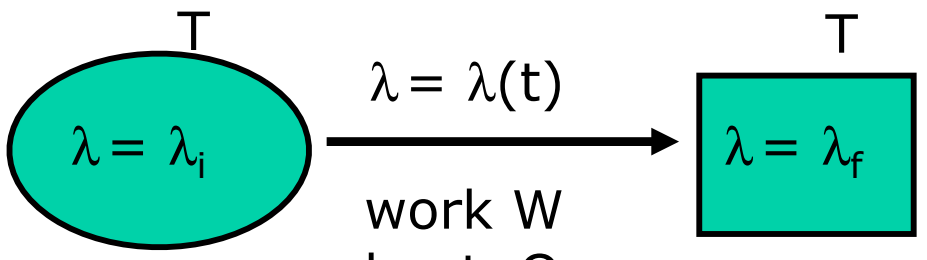


constant velocity
(30 Å/ns)

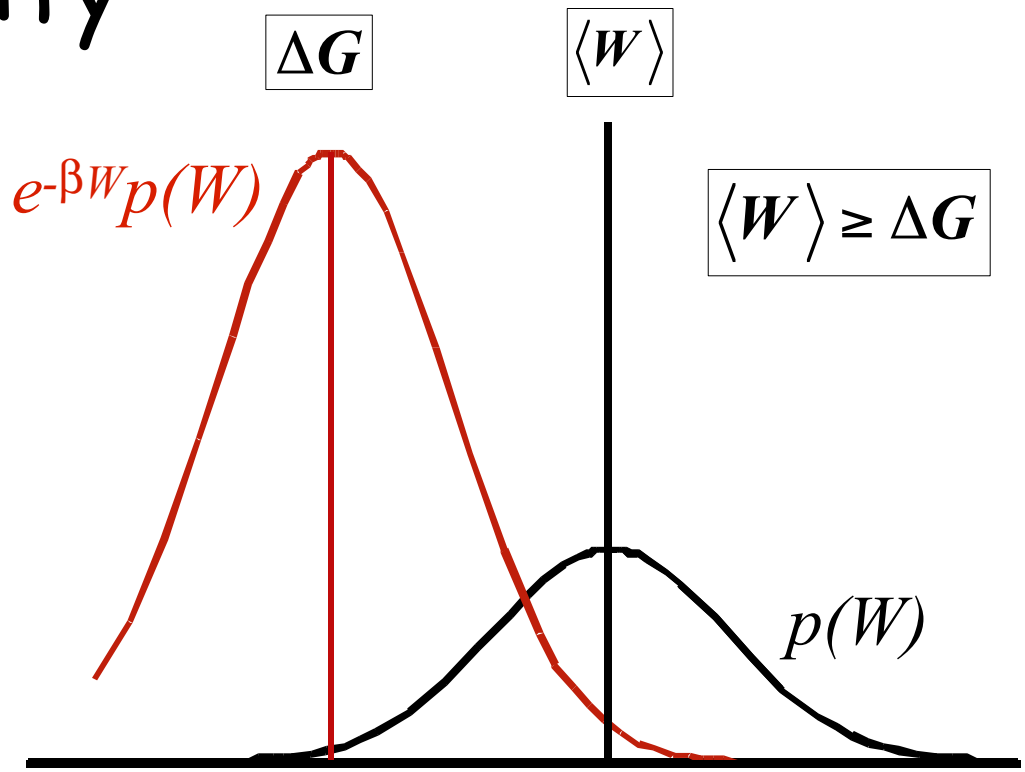


Jarzynski's Equality

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)

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

$$\beta = 1/k_B T$$

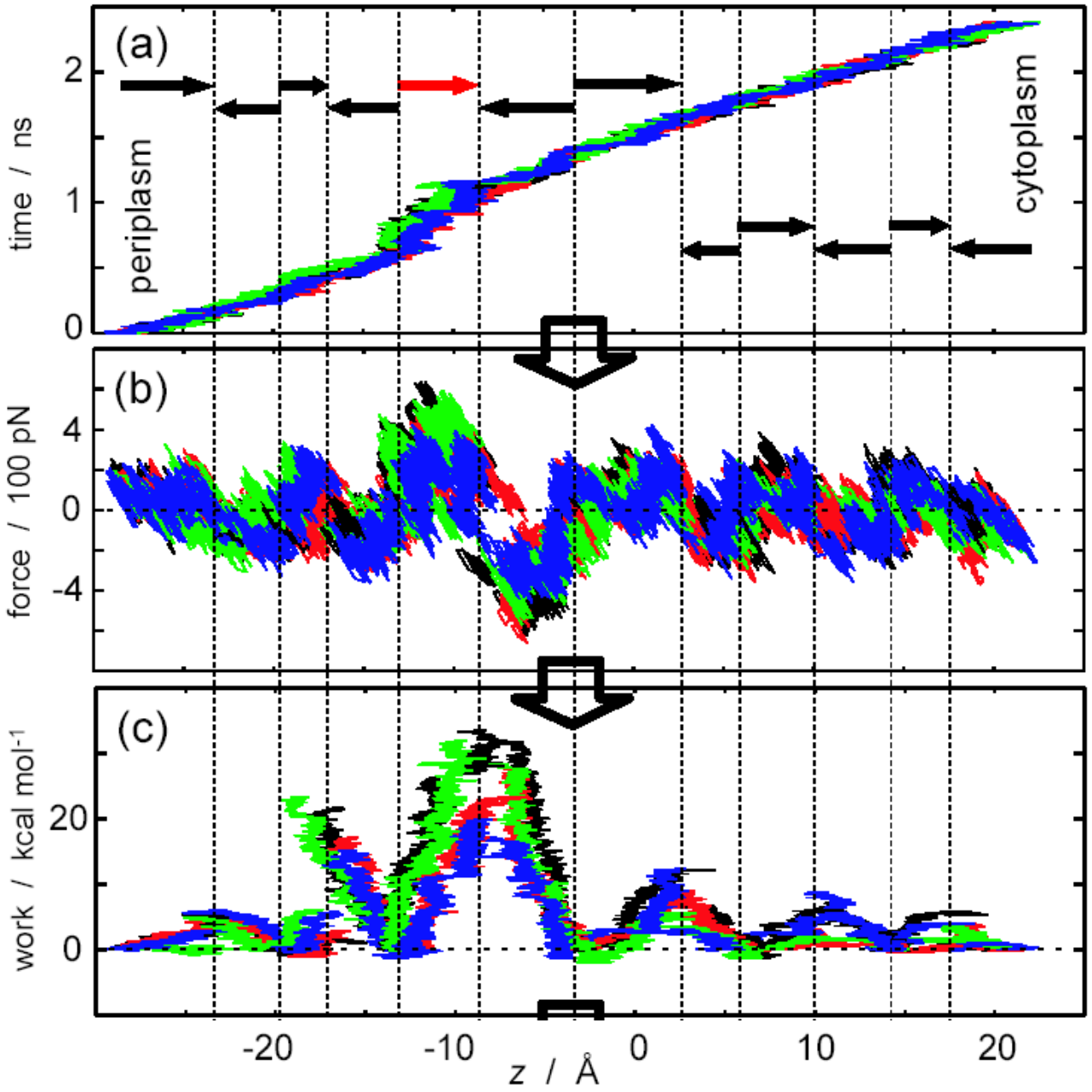
In principle, it is possible to obtain free energy surfaces from repeated non-equilibrium experiments.

Constructing the Potential of Mean Force

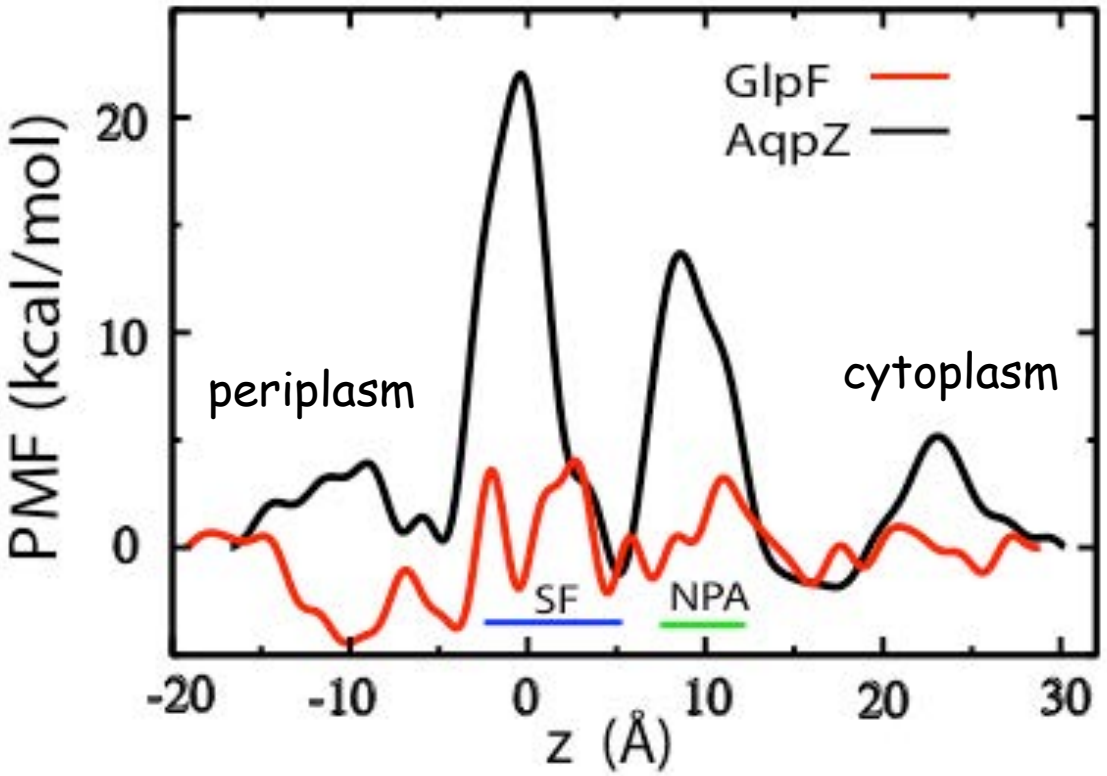
4 trajectories
v = 0.03, 0.015 Å/ps
k = 150 pN/Å

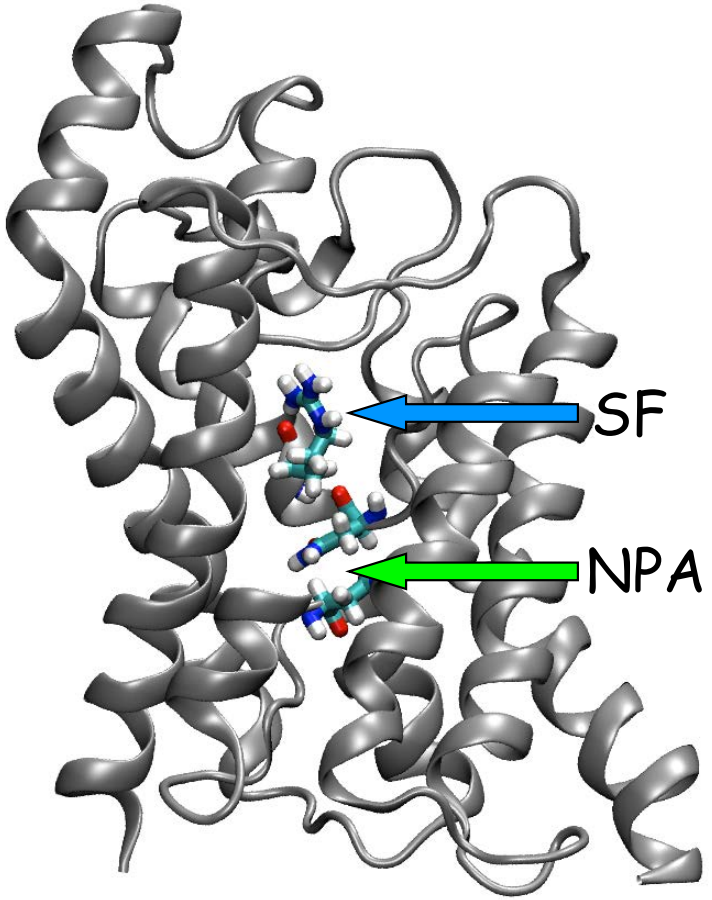
$$f(t) = -k[z(t) - z_0 - vt]$$

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



Three fold higher barriers



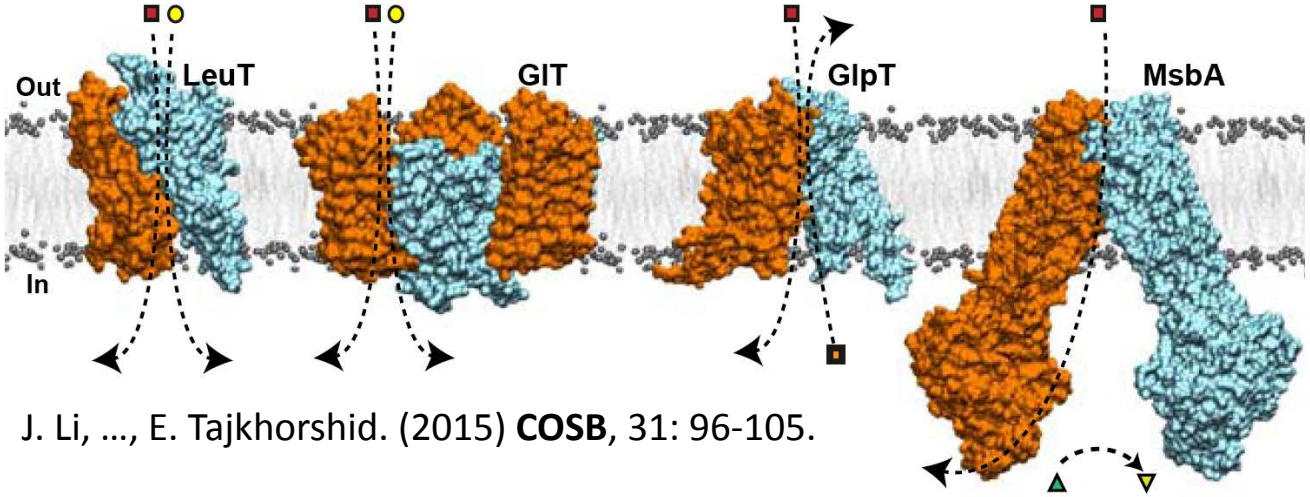


AqpZ 22.8 kcal/mol

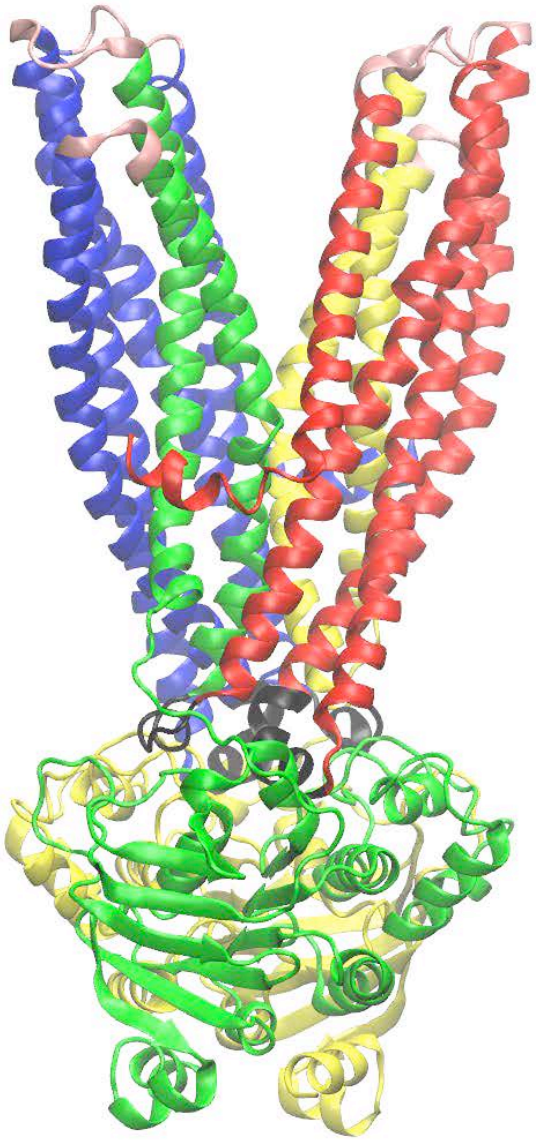
GlpF 7.3 kcal/mol

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

Battling the Timescale - Case II
Biased (nonequilibrium) simulations



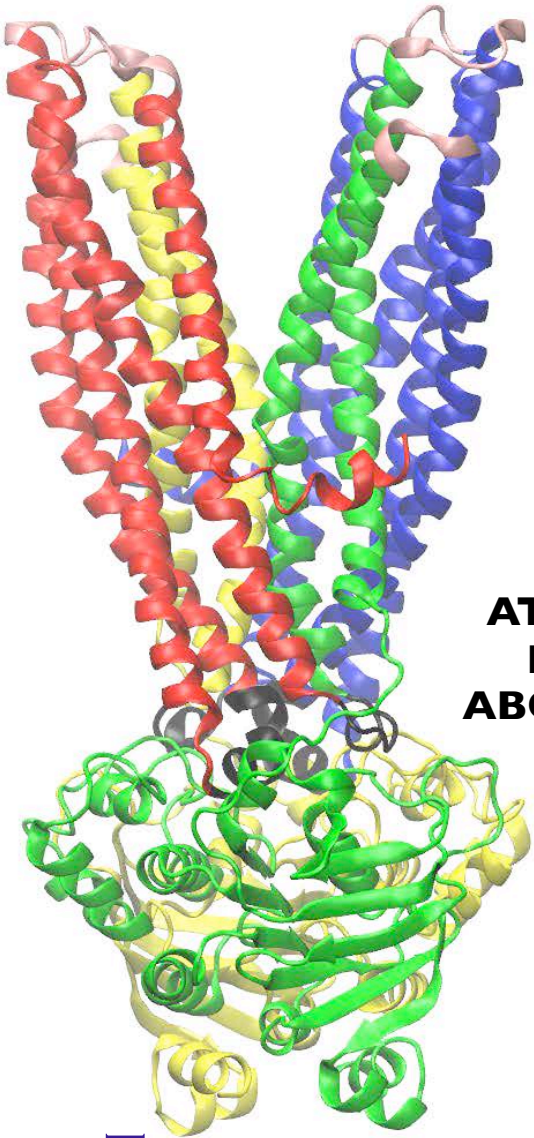
J. Li, …, E. Tajkhorshid. (2015) **COSB**, 31: 96-105.



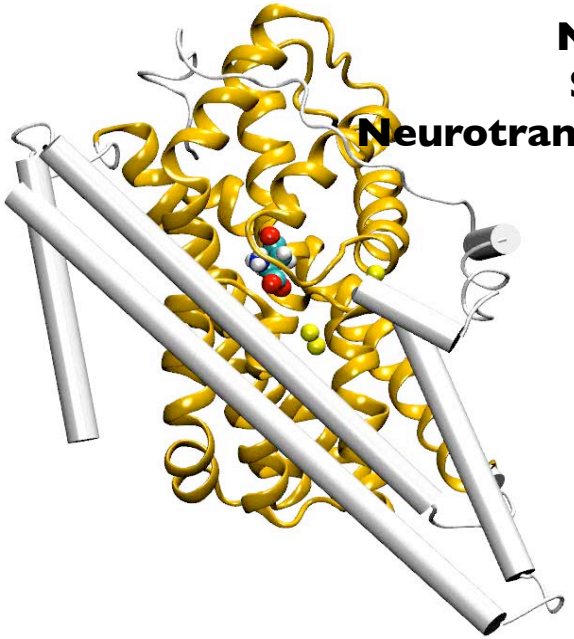
- **Neurotransmitter Uptake**
  - » Norepinephrine, serotonin, dopamine, glutamate,…
- **Gastrointestinal Tract**
  - » Active absorption of nutrients
  - » Secretion of ions
- **Kidneys**
  - » Reabsorption
  - » Secretion
- **Pharmacokinetics of all drugs**
  - » Absorption, distribution, elimination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

COMPLEX

Diverse Structural Transitions Involved

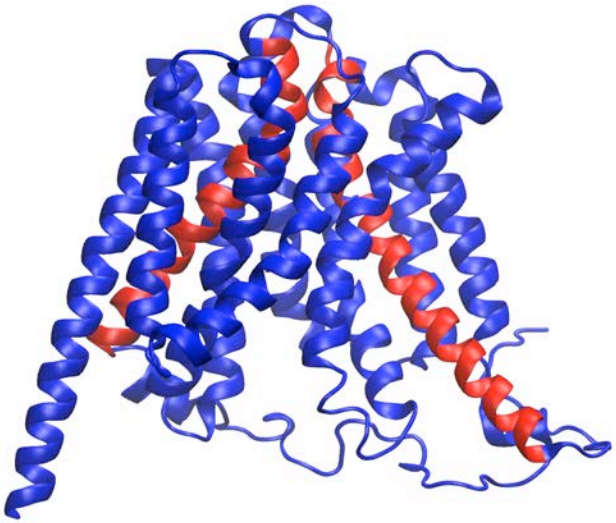


ATP-Driven Primary ABC Exporter



Na-coupled Secondary Neurotransmitter Transporter

Secondary Phosphate Antiporter



Biasing Techniques 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.

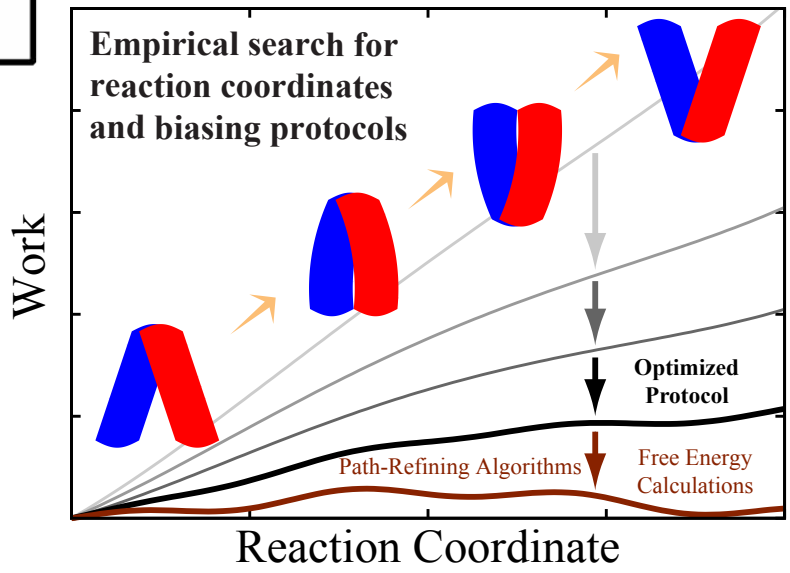
I.2 Optimizing the Biasing Protocols
Systematic search for a practical biasing protocol by using different combinations of collective variables.

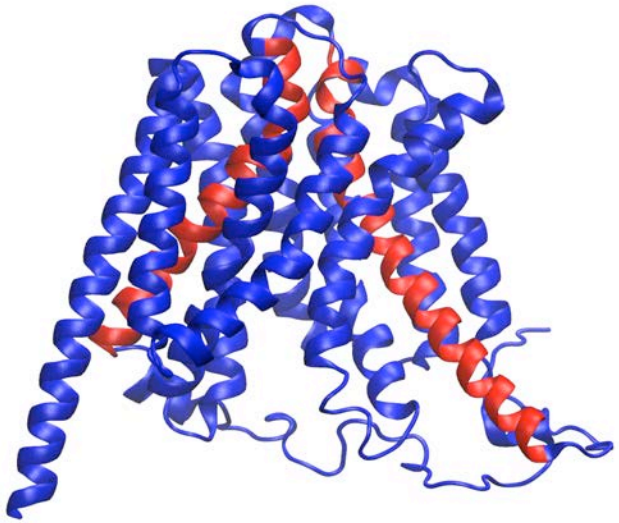
II. Optimizing the Transition Pathway
Use all of the conformations available to generate the most reliable transition pathway:
1. Bayesian approach for combining the data
2. Post-hoc string method (analysis tool)
3. String method with swarms of trajectories

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.

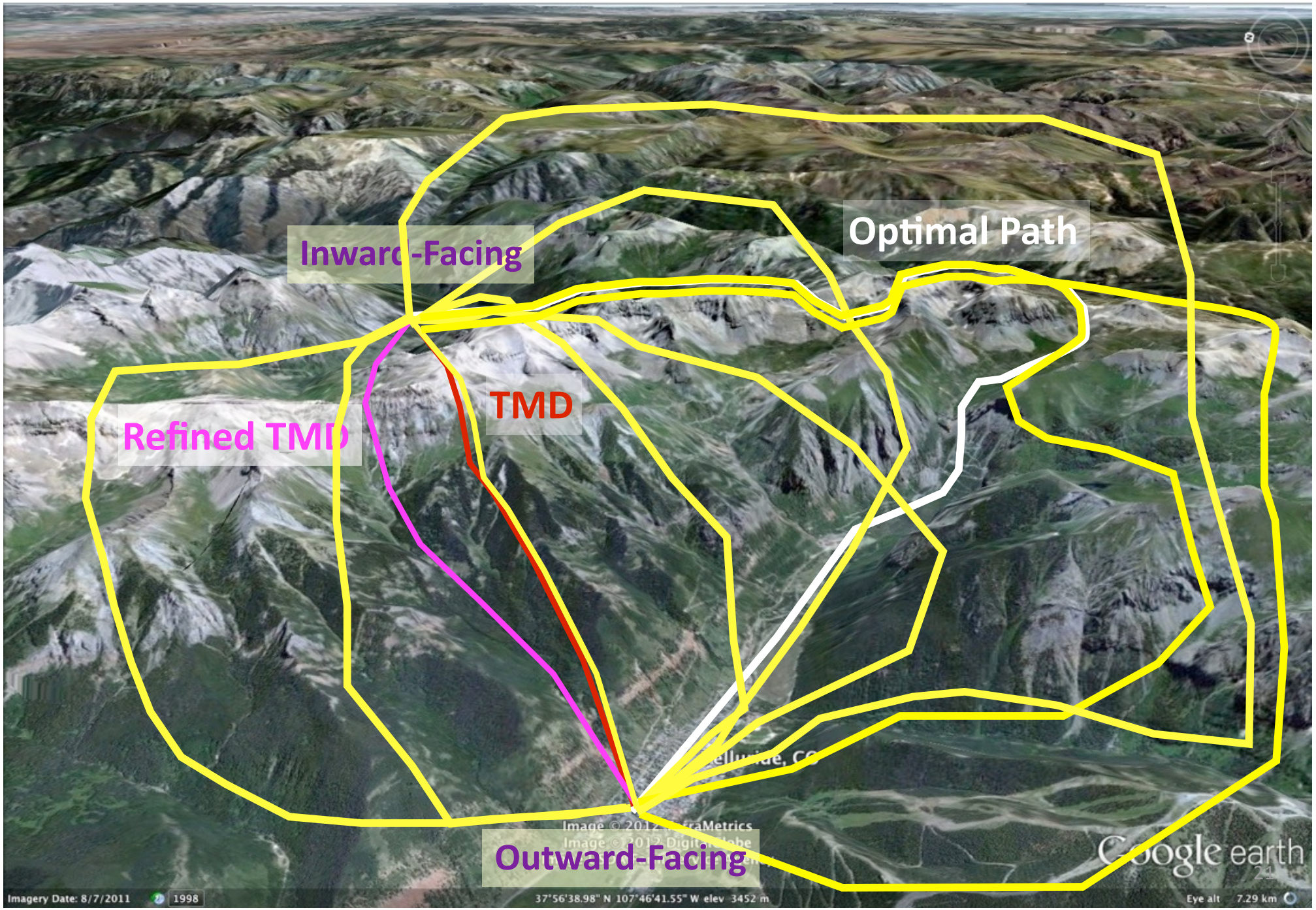
Mahmoud Moradi





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

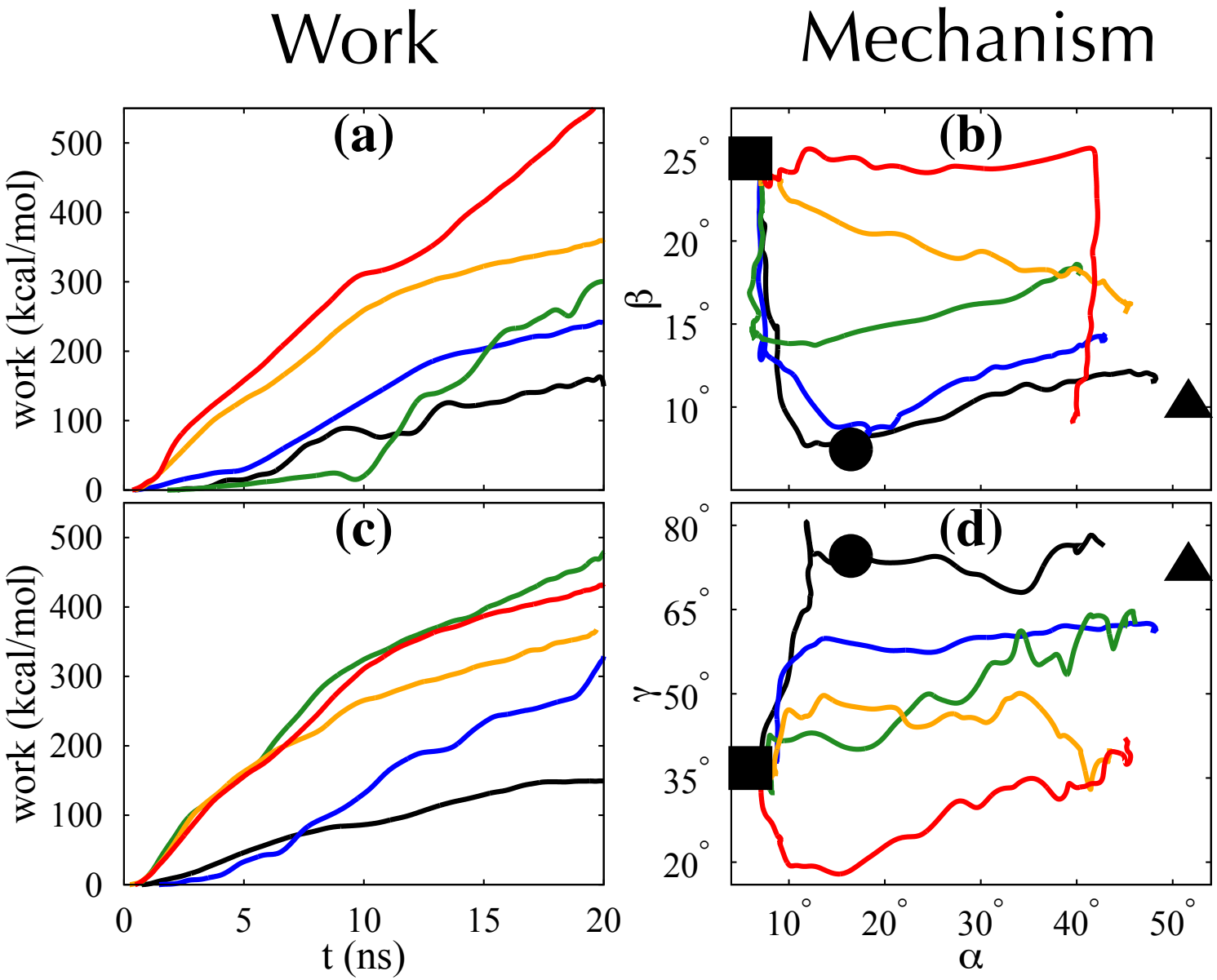
Collective variables:
RMSD, distance,
R_g, angle, …
orientation quaternion

Final state

Total simulation time

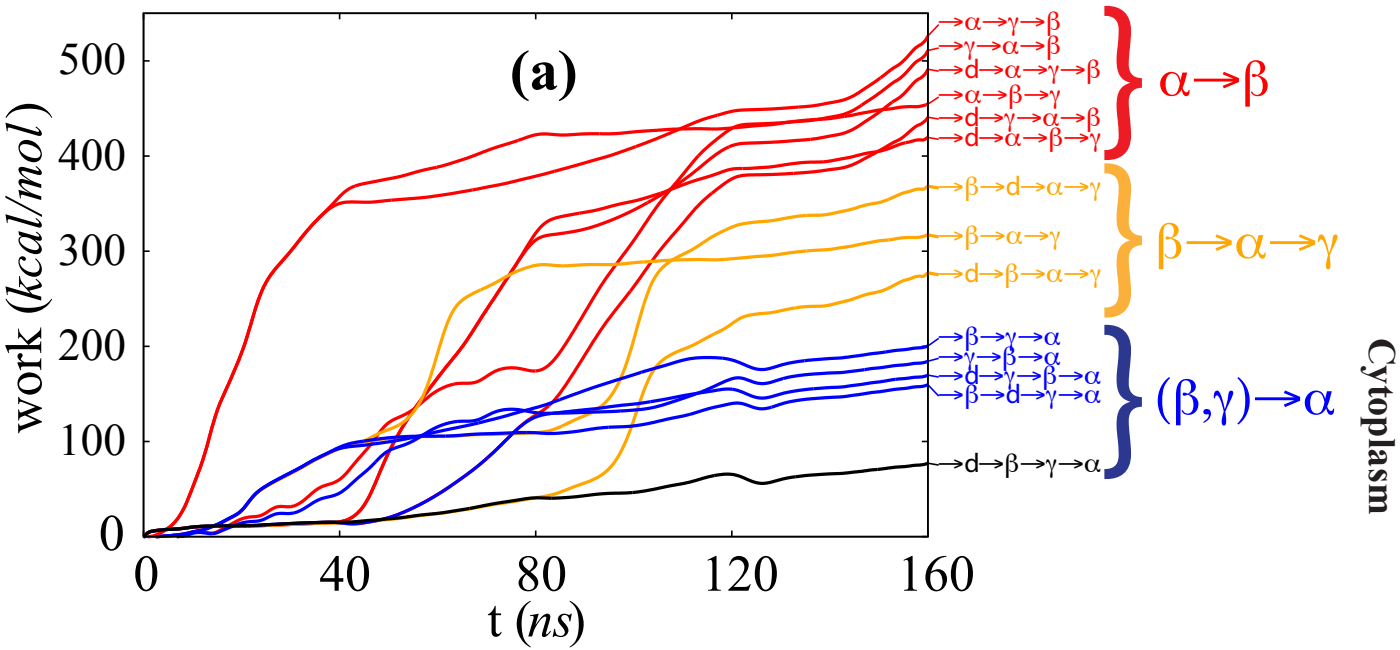
M. Moradi and ET (2013) **PNAS**, 110:18916–18921.
M. Moradi and ET (2014) **JCTC**, 10: 2866–2880.
M. Moradi, G. Enkavi, and ET (2015) **Nature Comm.**, 6:8393.

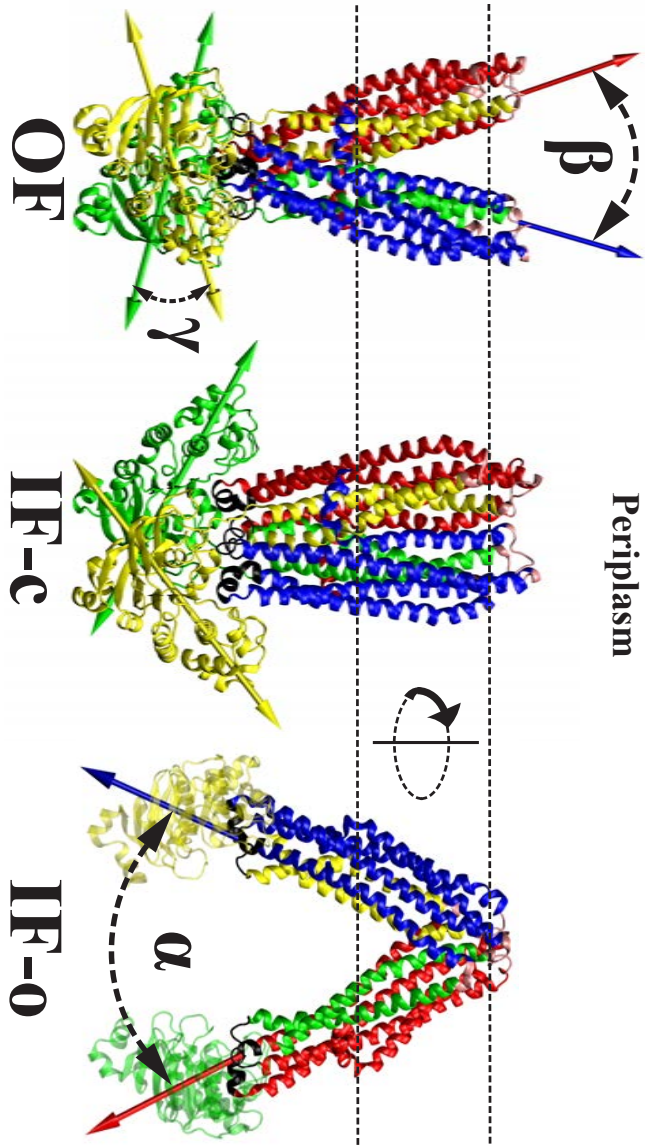
Progressively Optimizing the Biasing Protocol/Collective Variable using non-Equilibrium Work as a Measure of the Path Quality



Example set taken from a subset of 20 ns biased simulations

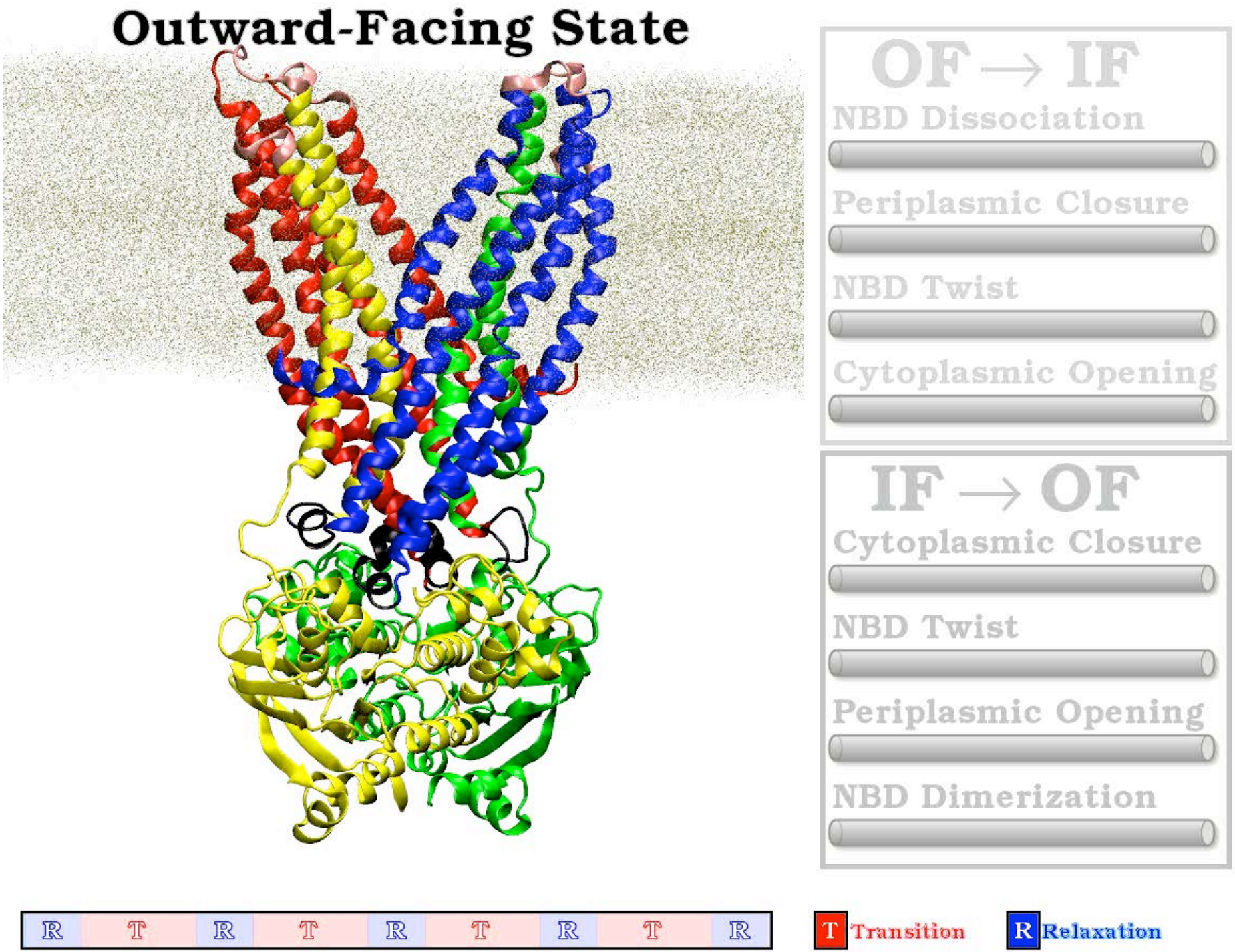
Mechanistic Insight From Transition Pathways in ABC exporters from Non-Equilibrium Simulations





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

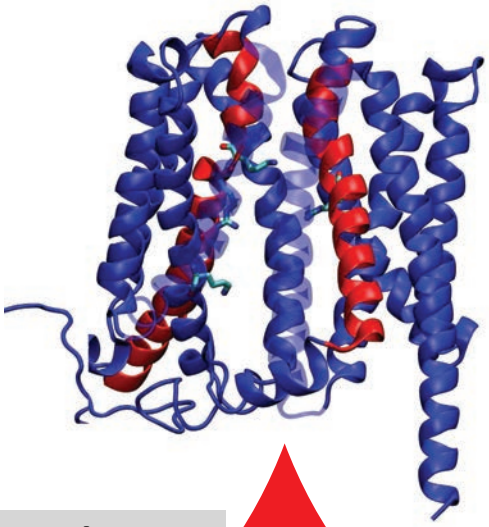
M. Moradi and ET (2014) **JCTC**, 10: 2866–2880.



NBD Doorknob Mechanism

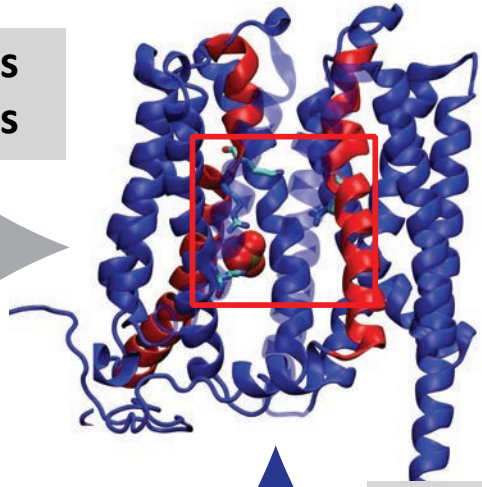
M. Moradi and ET (2013) **PNAS**, 110:18916–18921.

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

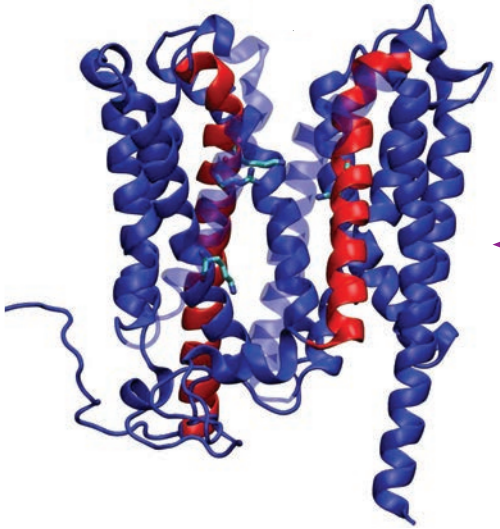


IF_{bound}

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

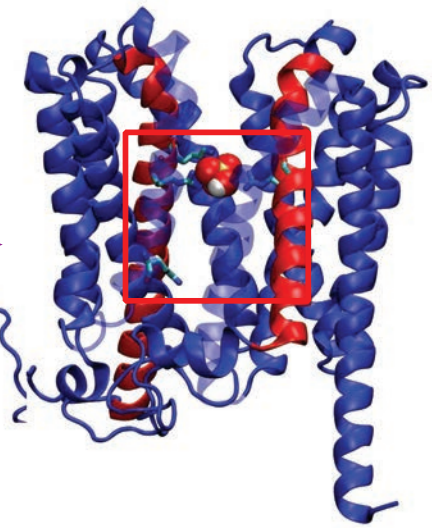
150 replicas

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}

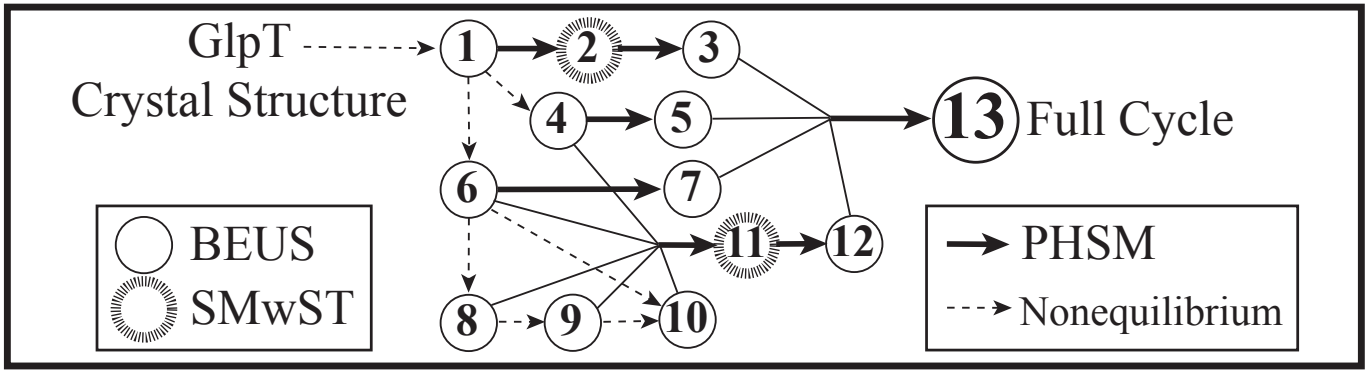
30 r x 20 ns
30 r x 20 ns
30 r x 20 ns



OF_{bound}

Simulation protocols

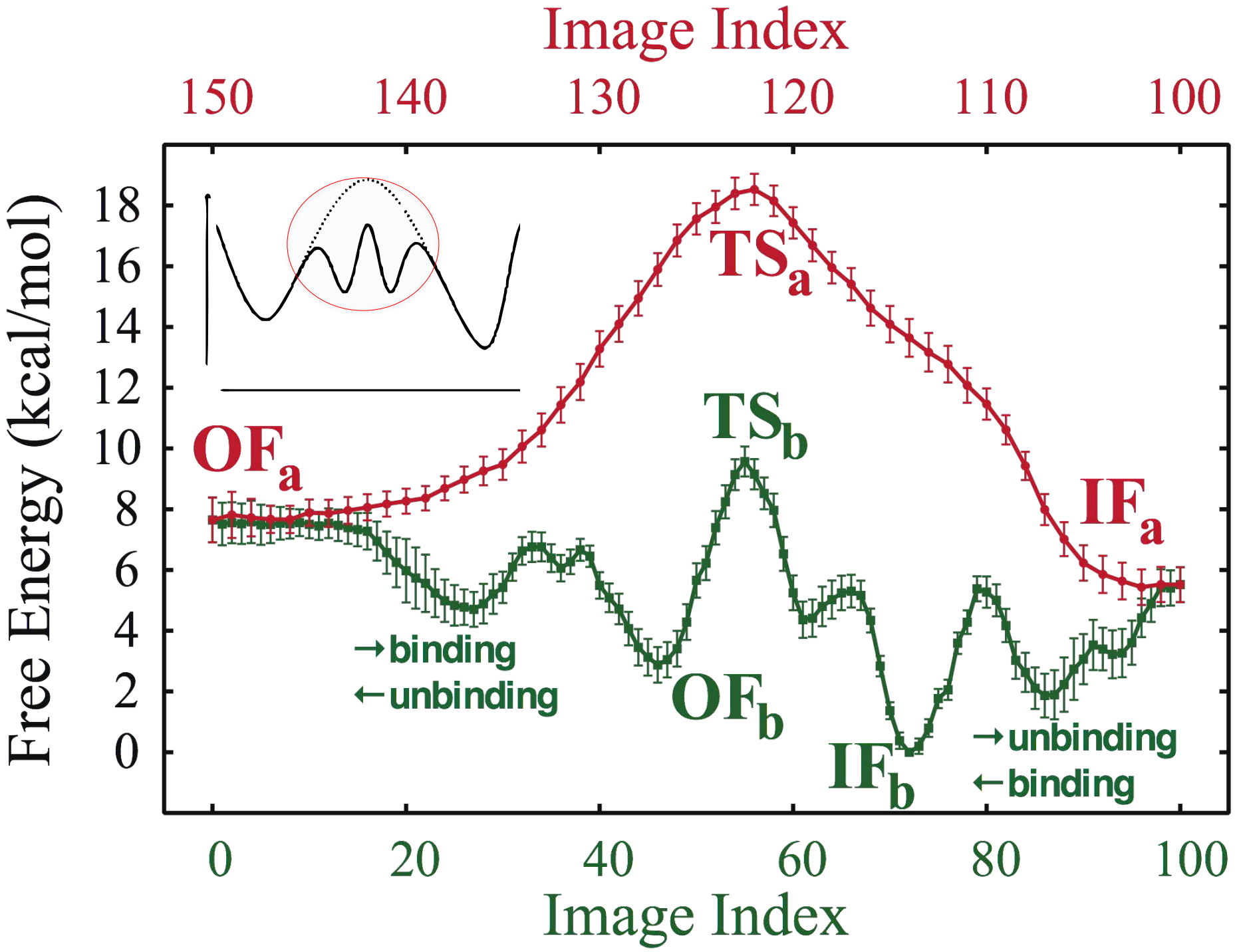
	Transition	Technique	Collective Variables	# of Replicas × Runtime		
1	$IF_a \leftrightarrow OF_a$	BEUS	(Q_1,Q_7)	12 × 40 ns	=	0.5 µs
2		SMwST	{Q}	1000 × 1 ns	=	1 µs
3		BEUS	{Q}	50 × 20 ns	=	1 µs
4	$IF_a \leftrightarrow IF_b$	BEUS	Z_{Pi}	30 × 40 ns	=	1.2 µs
5		BEUS	$(\{Q\}, Z_{Pi})$	30 × 40 ns	=	1.2 µs
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

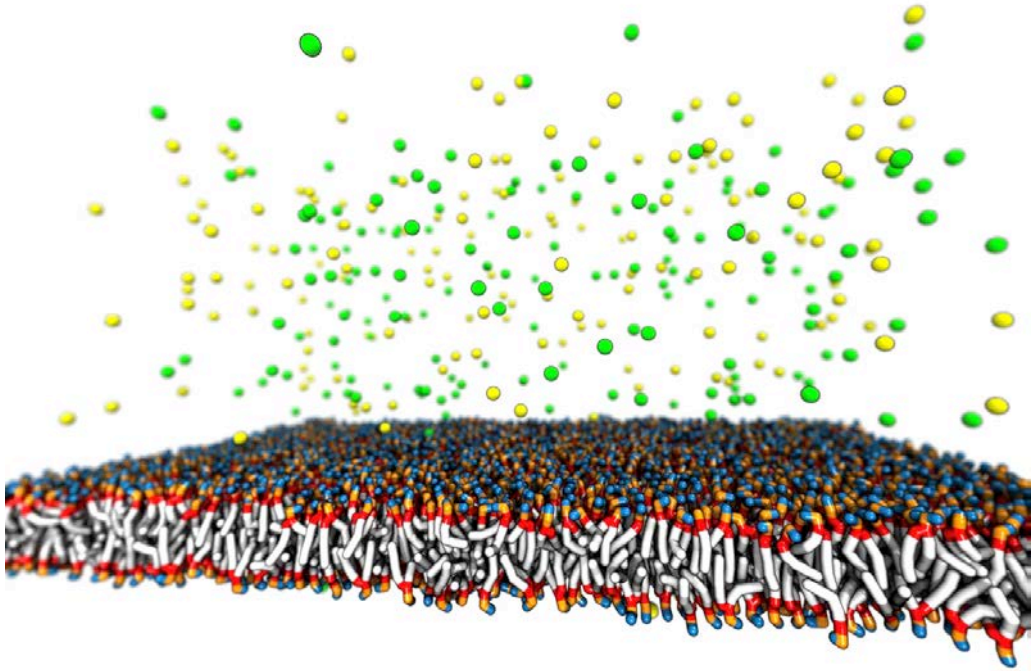
M. Moradi, G. Enkavi, and ET (2015) **Nature Communication**, 6: 8393.



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

Battling the Timescale - Case III
Multiscale Simulations

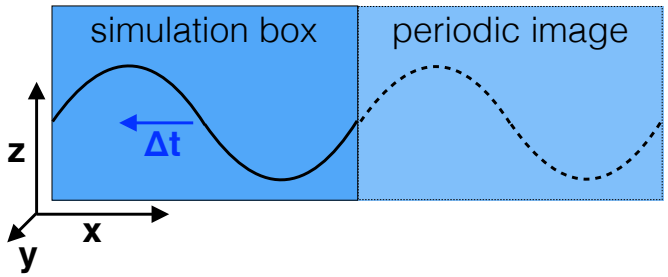
Membrane Budding/Fusion

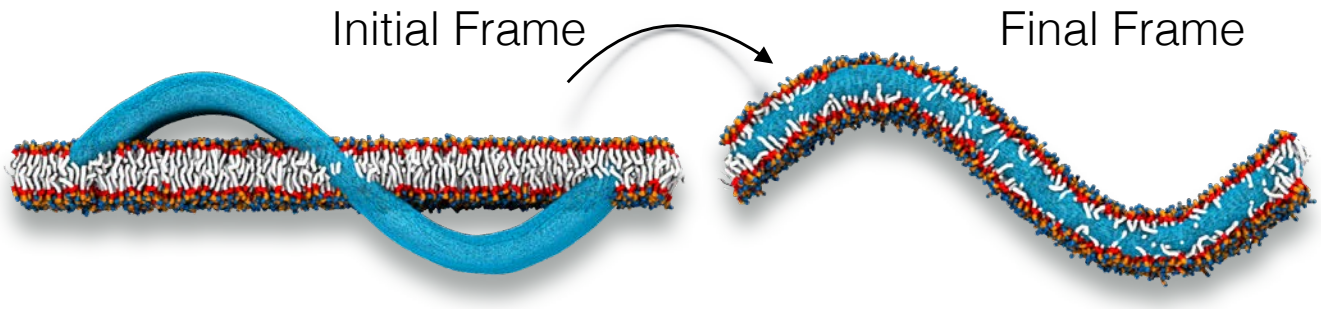


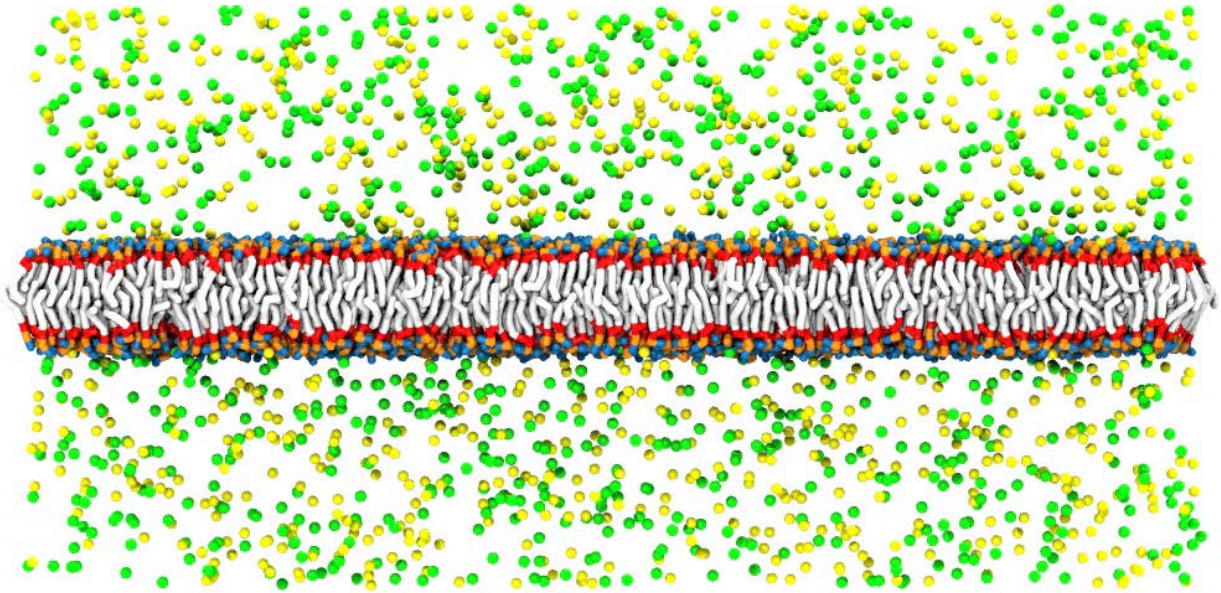
Combining multiple replica simulations and coarse-grained models to describe membrane fusion

Workflow for Multi-Scale Modeling

Parametrically Defined Sine Function

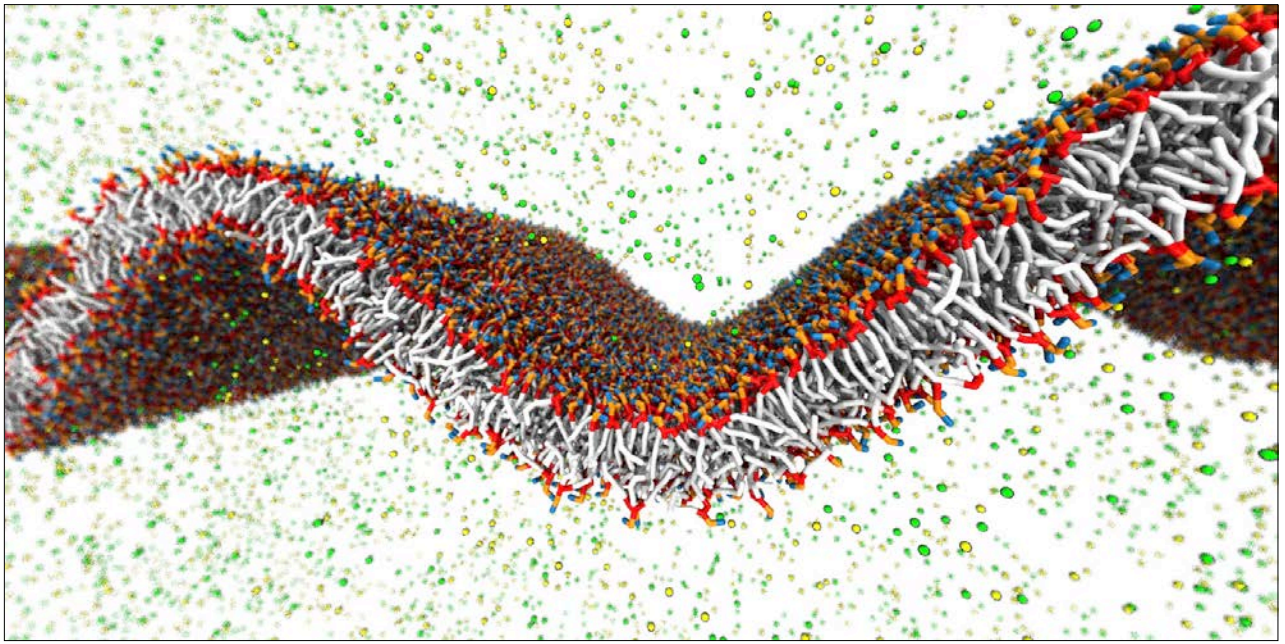






Christopher Mayne, Tajkhorshid Lab

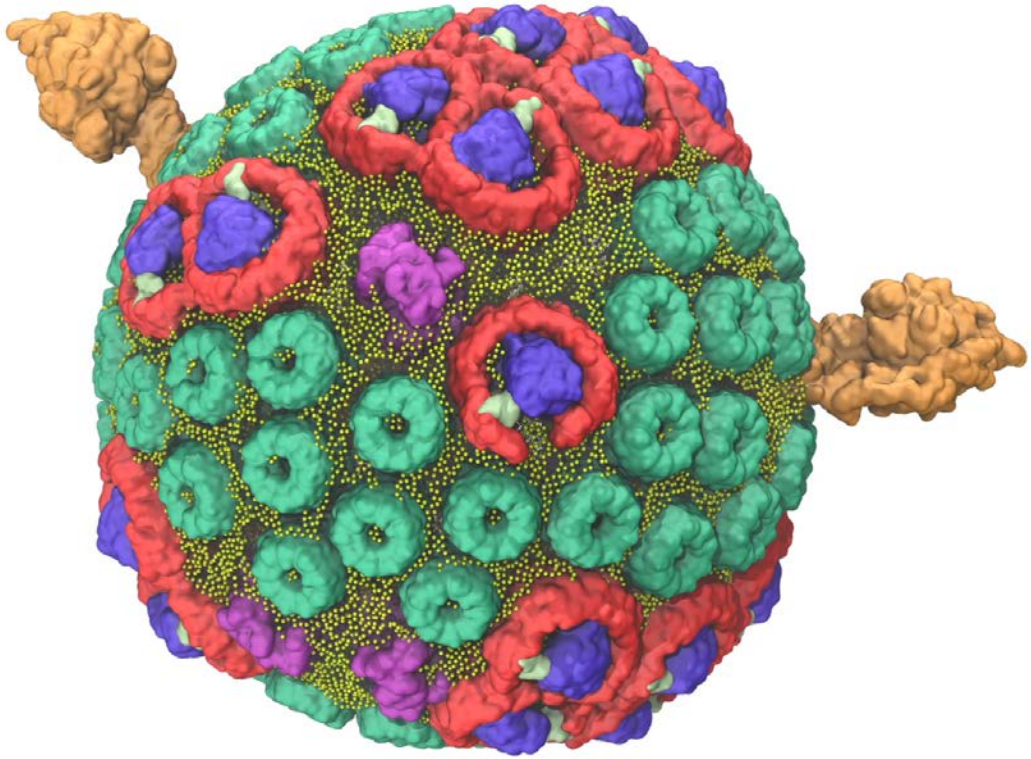
Workflow for Multi-Scale Modeling



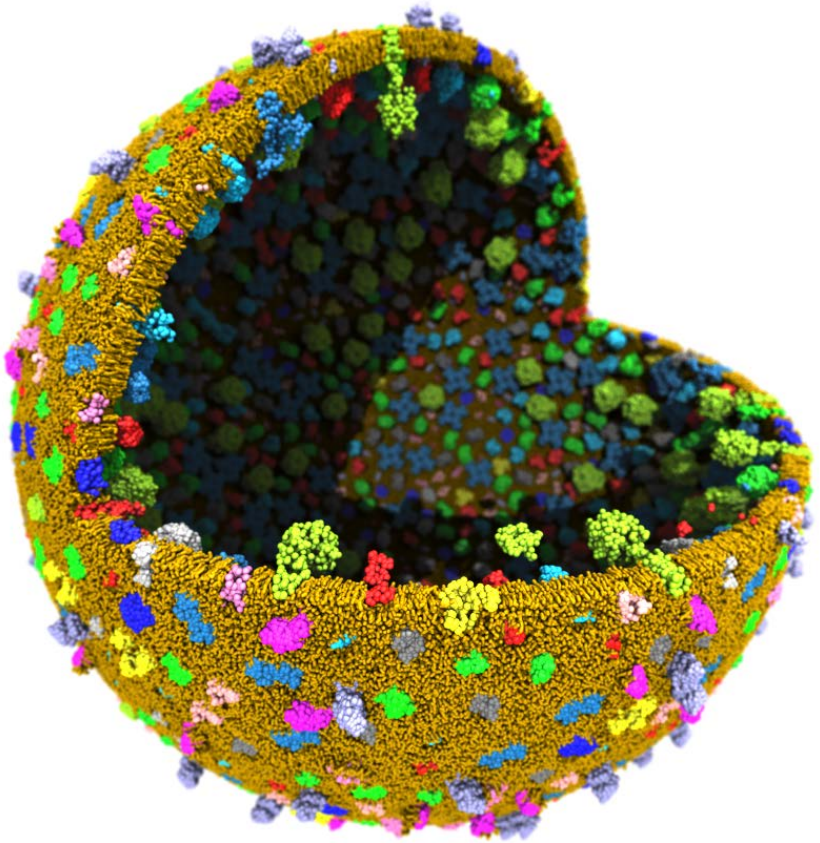
Christopher Mayne, Tajkhorshid Lab

Applications of Computational Methodologies to Cell-Scale Structural Biology

Using simulations as a “structure-building” tool



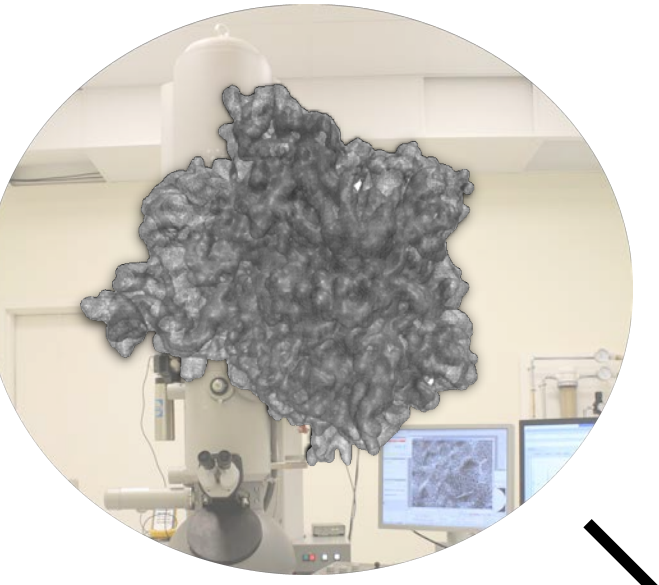
The most detailed model of a chromatophore



Computational model of a minimal cell envelope

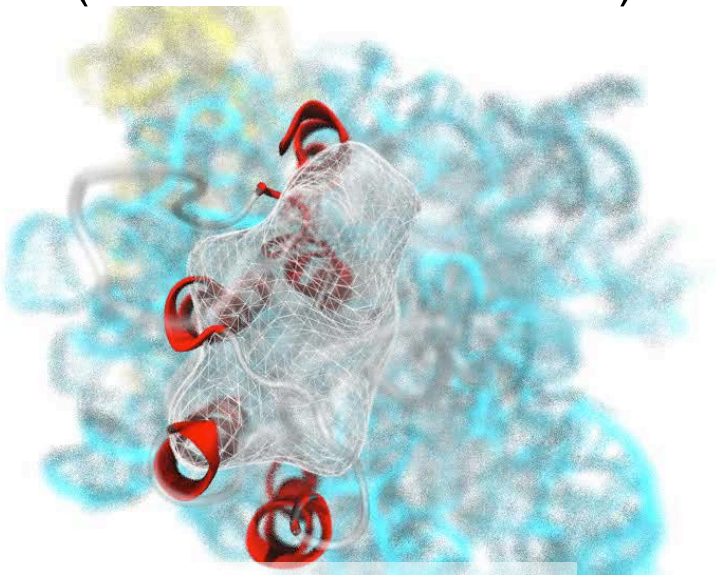
Molecular Dynamics Flexible Fitting (MDFF)

Electron Microscope



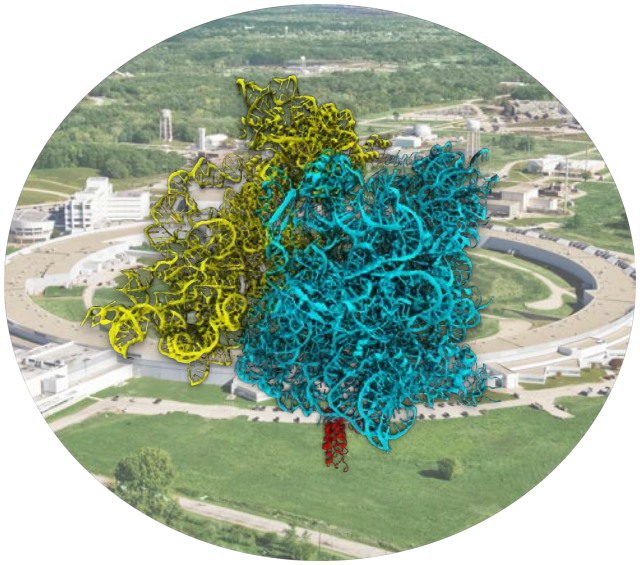
cryo-EM density map

(Ribosome-bound YidC)

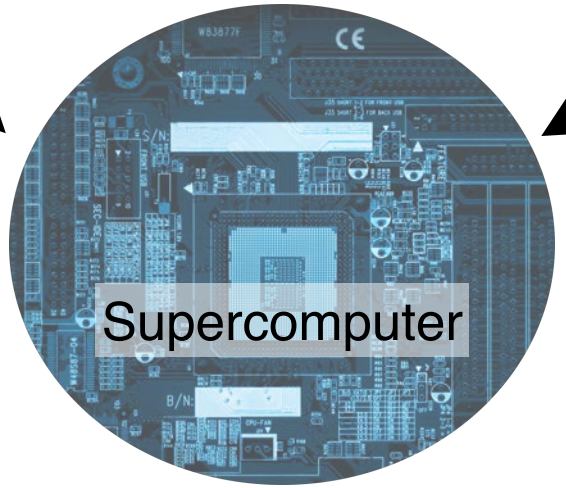


Match through MD

APS Synchrotron

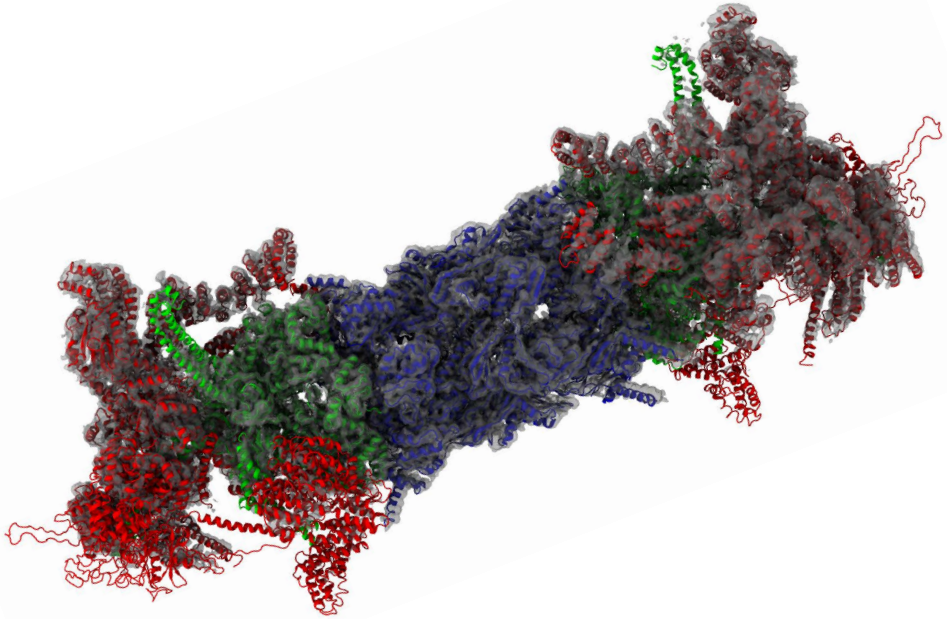


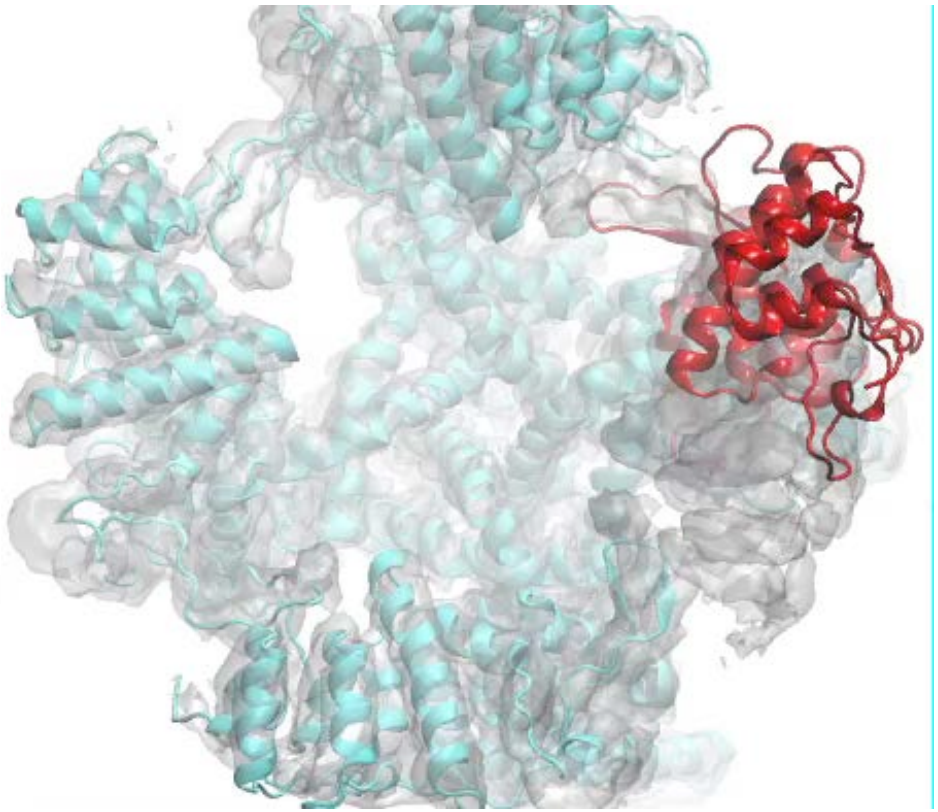
crystallographic structure



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

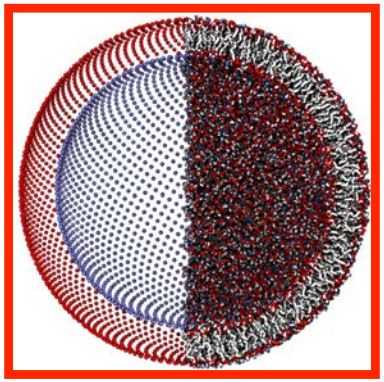
Molecular Dynamics Flexible Fitting (MDFF)



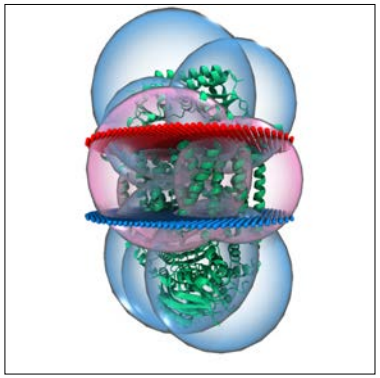


Automated Protein Embedding into Complex Membrane Structures

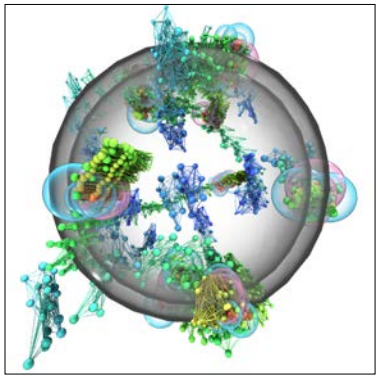
Vesicle Construction



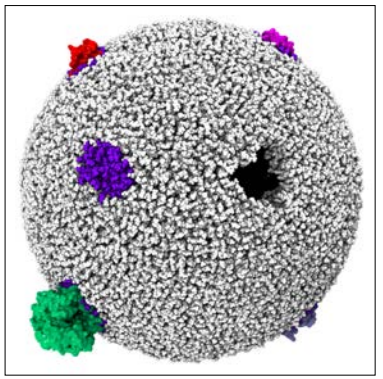
Coarse Grain Protein

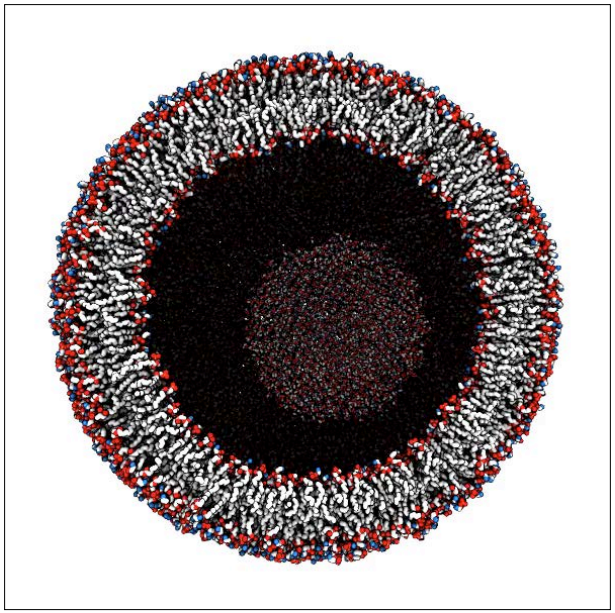


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

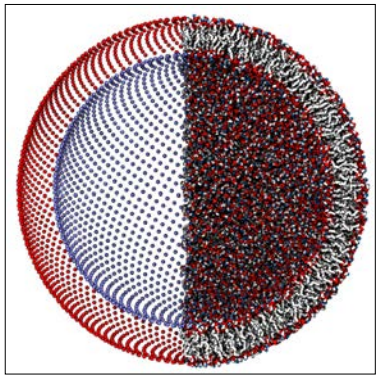
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

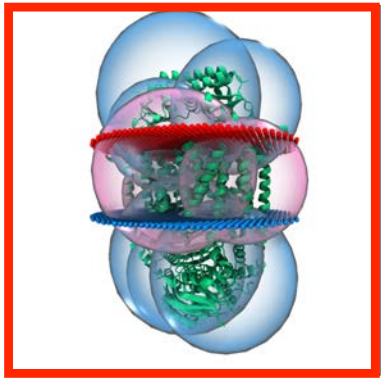
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

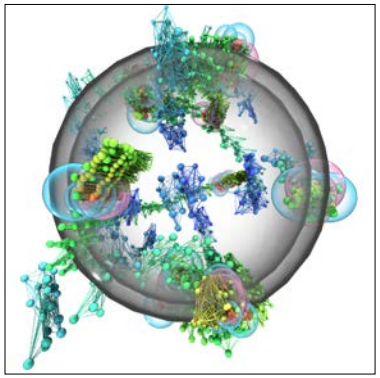
Vesicle Construction



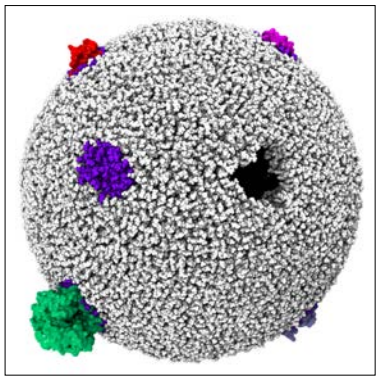
Coarse Grain Protein

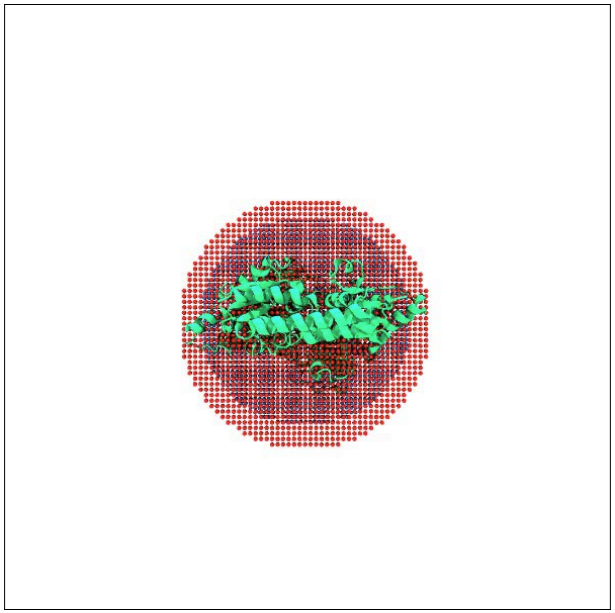


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

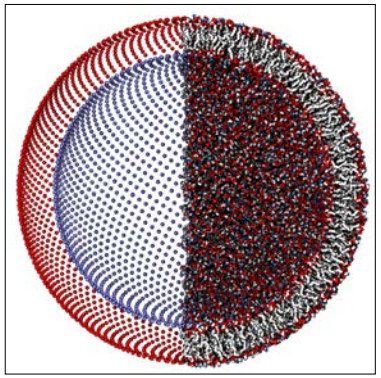
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

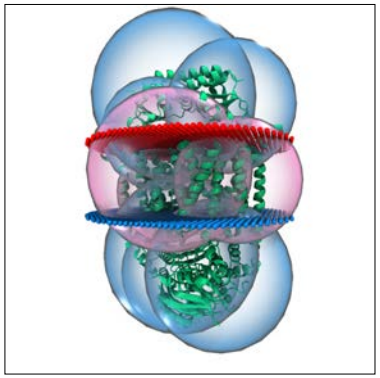
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

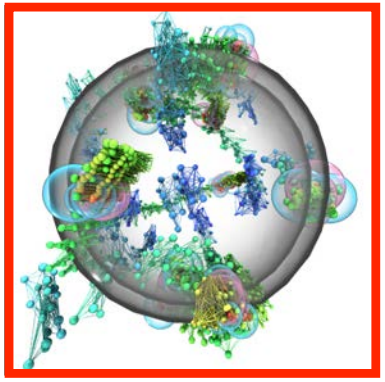
Vesicle Construction



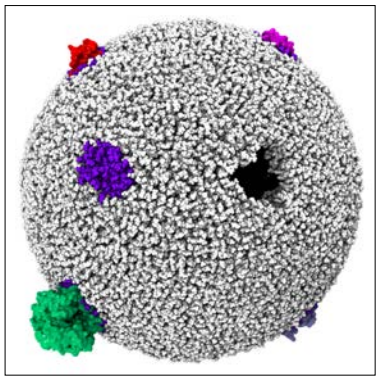
Coarse Grain Protein

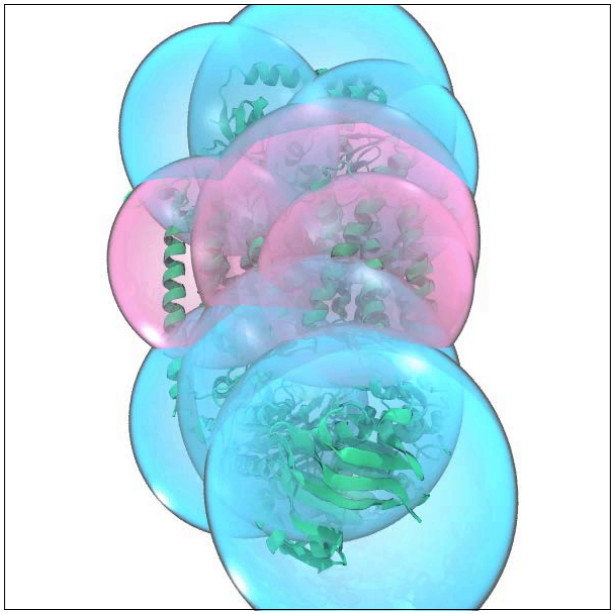


CG Protein Placement



Combine Lipid + Protein





Distribution of proteins across the membrane surface (dense environment)

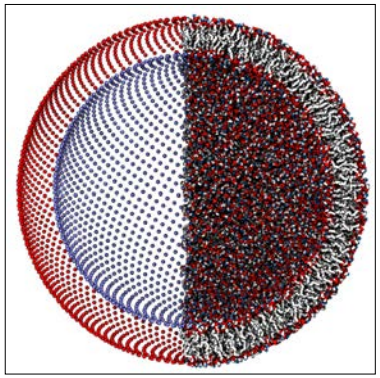
- Ability the handle a variety of protein geometries
- Proper orientation of proteins in relation to the membrane surface
- Generalizable and automated method for membranes of arbitrary shape

Embedding proteins into the membrane

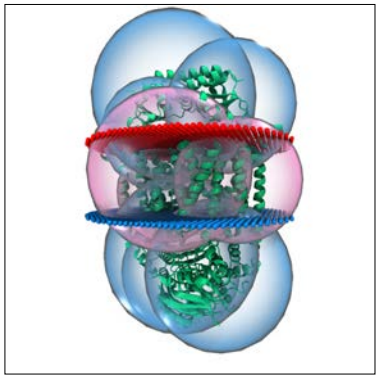
- Account for surface area occupied by proteins in inner and outer leaflets
- Proper lipid packing around embedded proteins

Automated Protein Embedding into Complex Membrane Structures

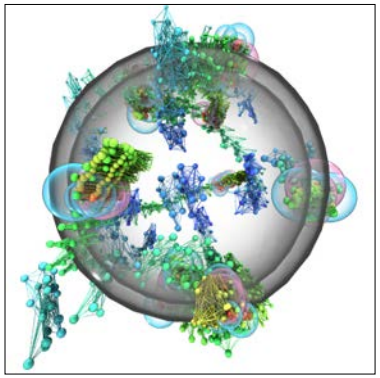
Vesicle Construction



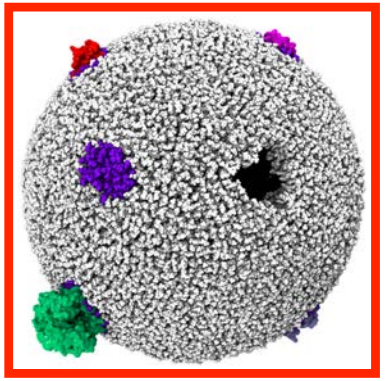
Coarse Grain Protein

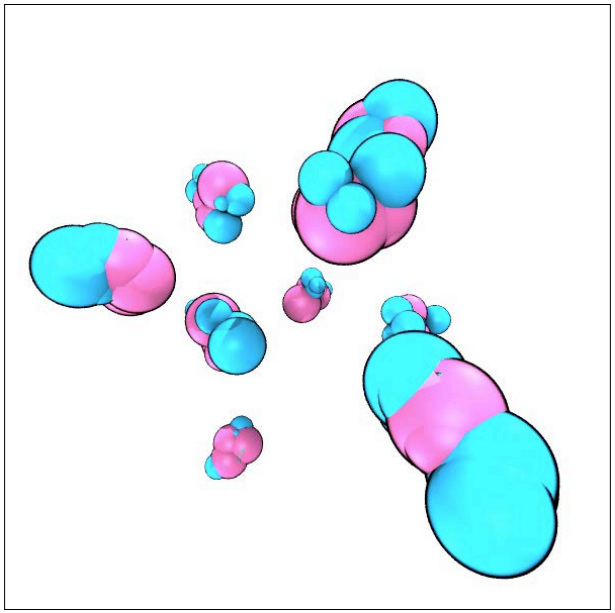


CG Protein Placement



Combine Lipid + Protein





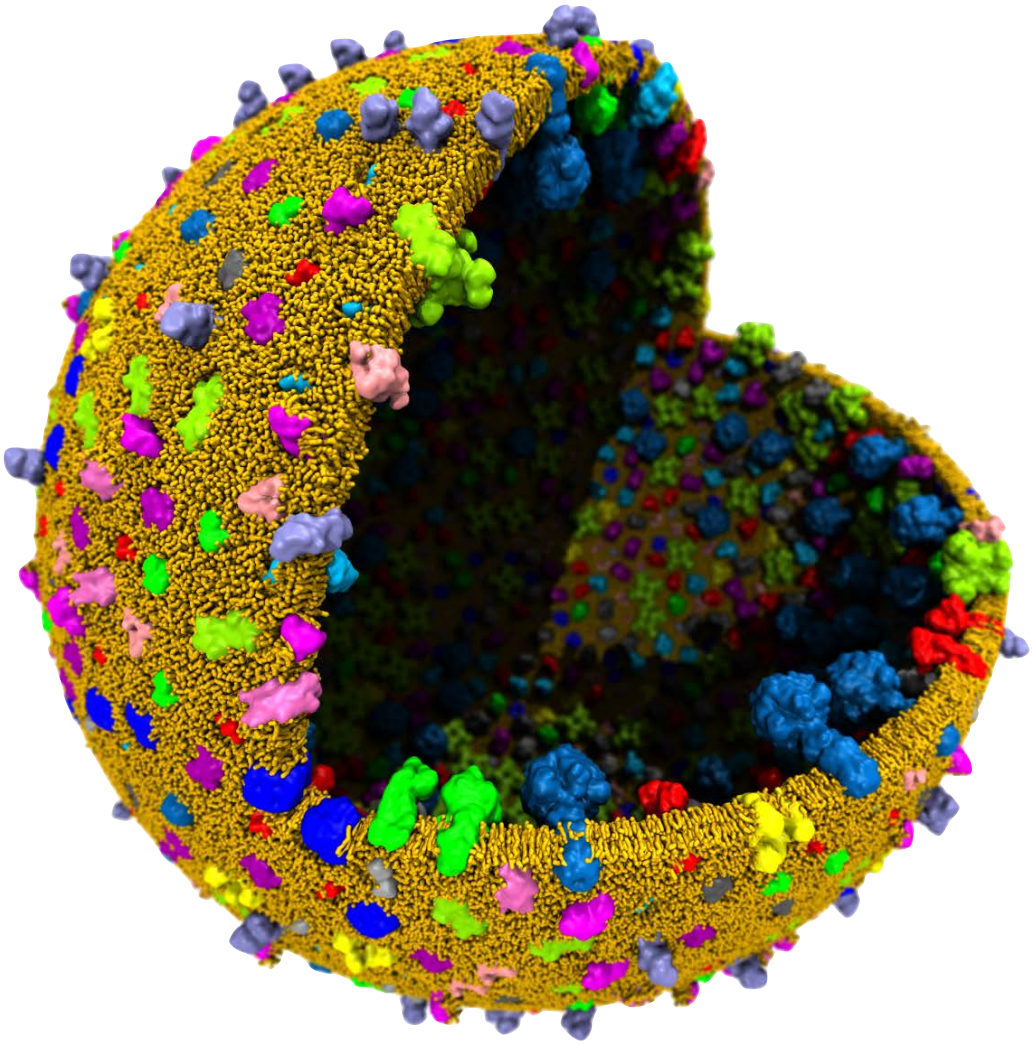
Distribution of proteins across the membrane surface (dense environment)

- 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

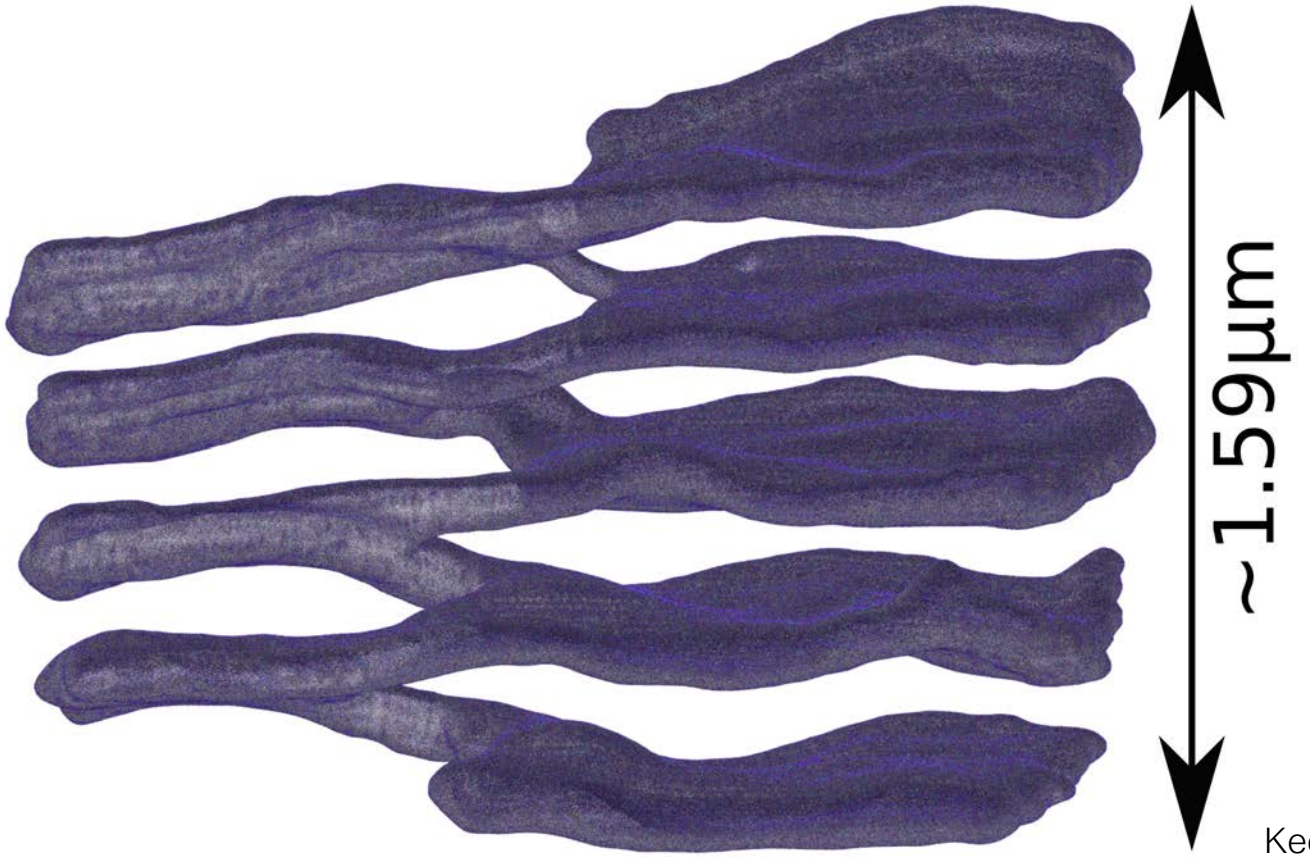
3.7 M lipids (DPPC), 2.4 M Na^+ & Cl^- ions,
104 M water particles (4 H_2O / particle)

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**.

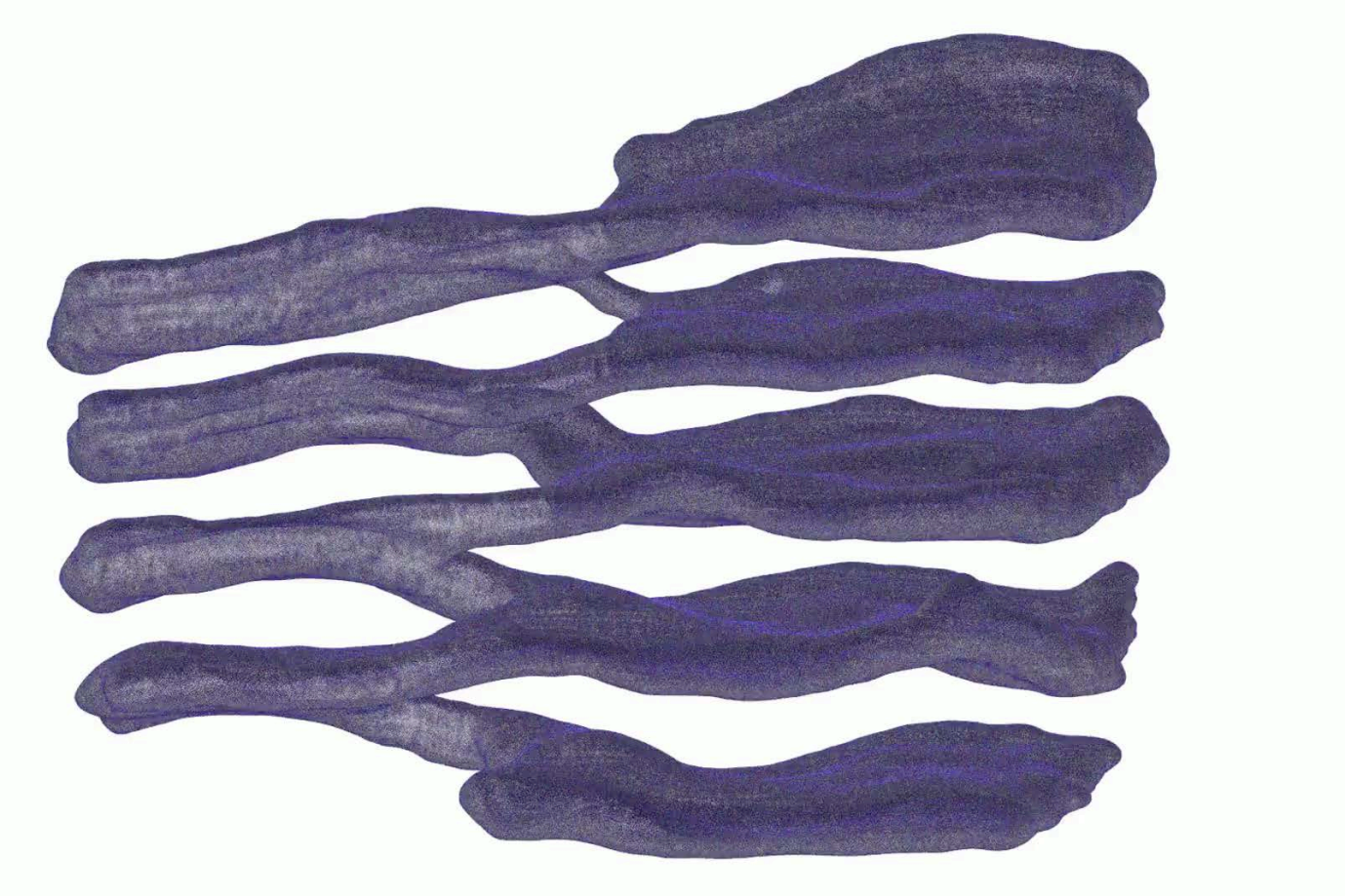
Outer Leaflet
Inner Leaflet
Cholesterol
POPC
POPE
POPI
POPS
Sphingomyelin
Cardiolipin

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

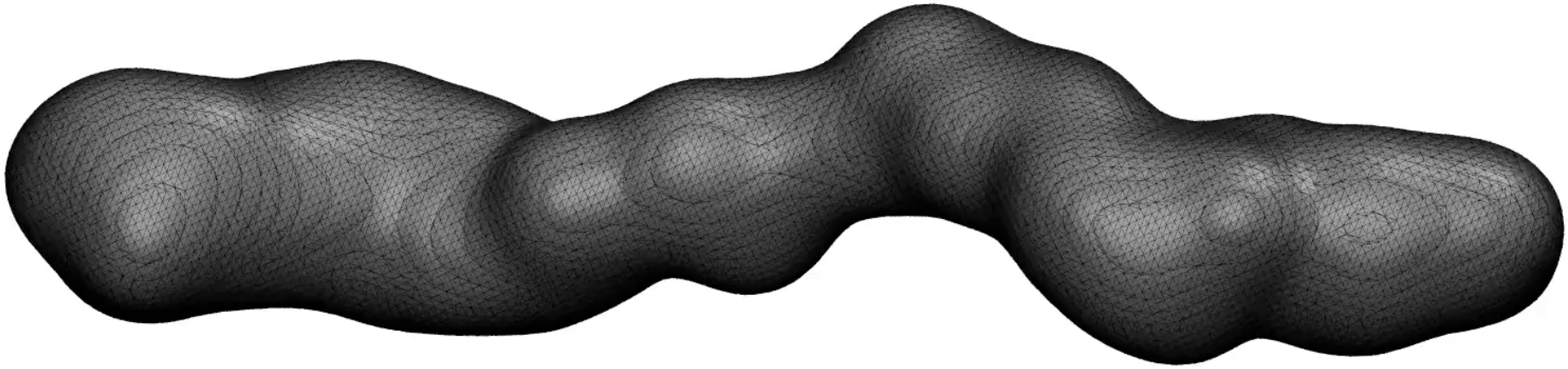
- Outer Leaflet
- Inner Leaflet
- Cholesterol
- POPC
- POPE
- POPI
- POPS
- Sphingomyelin
- Cardiolipin

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

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

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

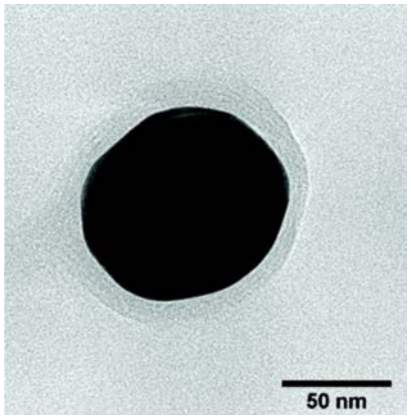
xMAS Builder



Obtain 3D mesh from an experimental technique

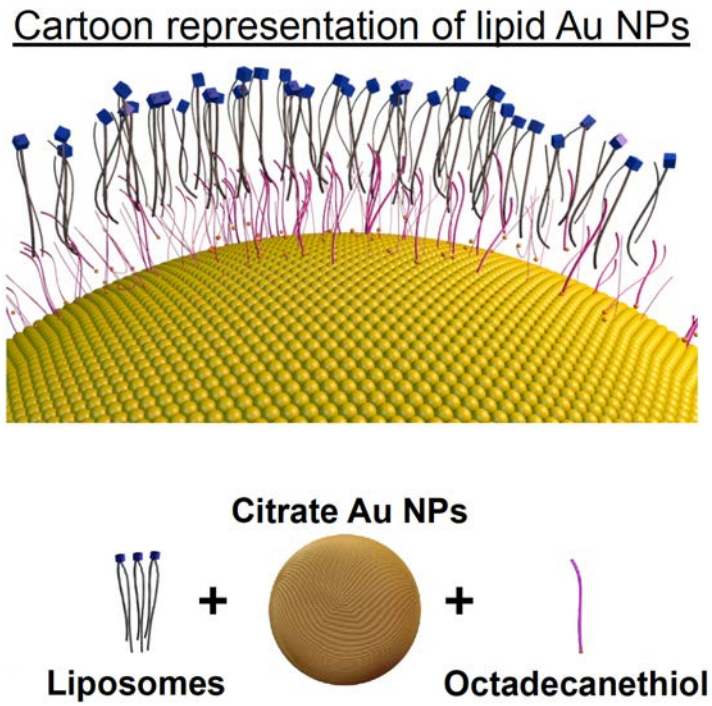
Nano-biotechnology
Gold Nanoparticles as Delivery Vehicles

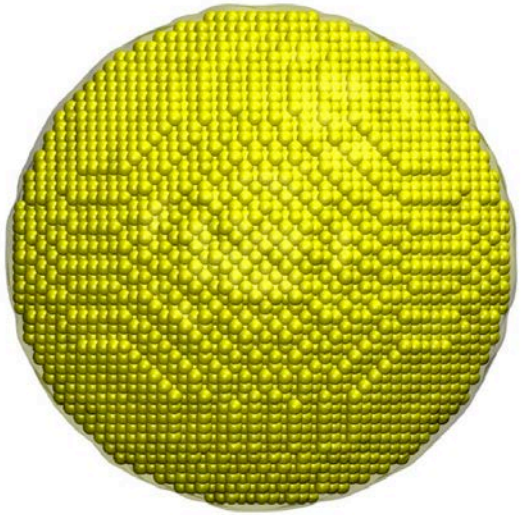
Transmission Electron Micrograph



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

Schematic model with no prediction power





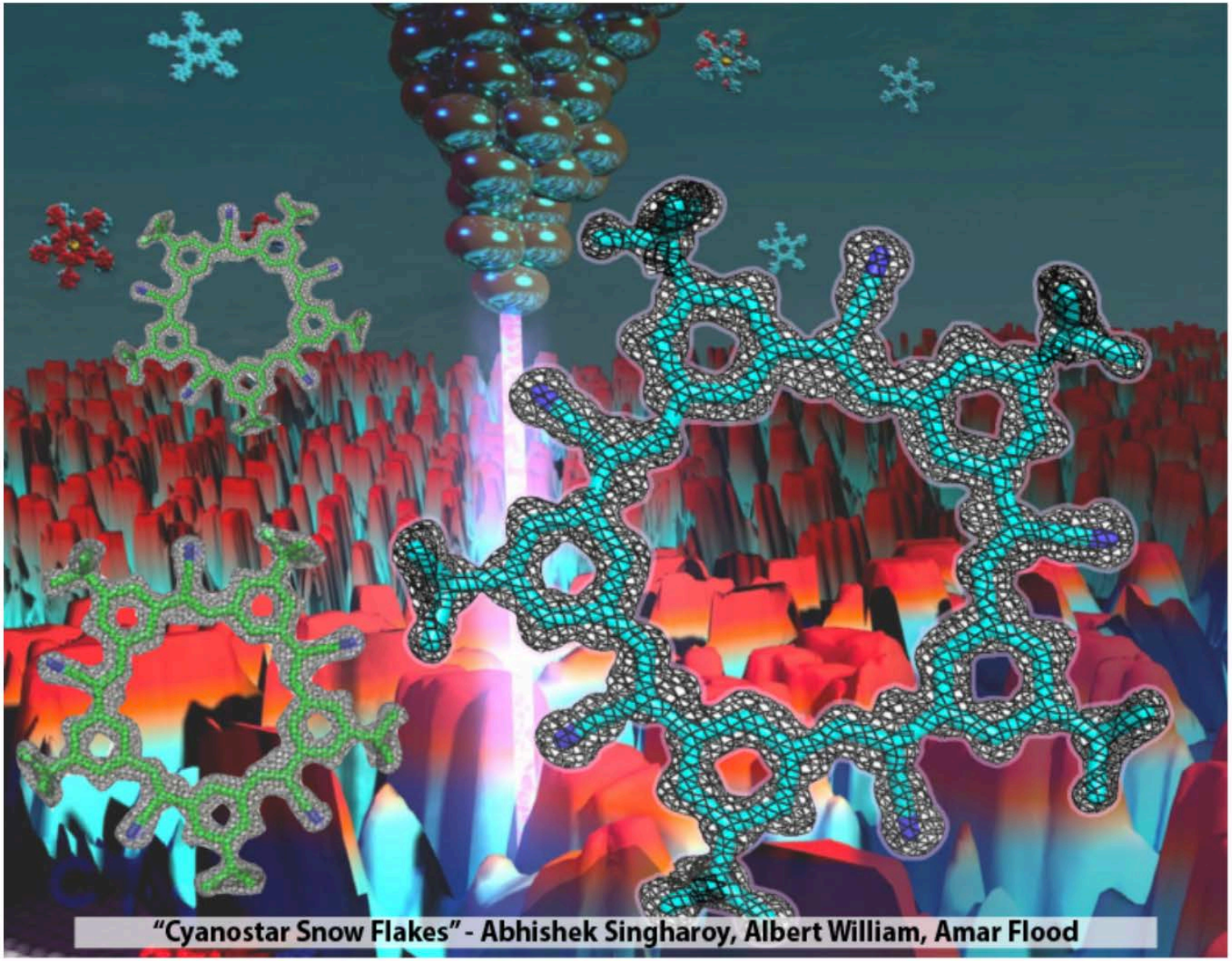
Experiment:
Murphy Lab

Modeling/Simulation:
Tajkhorshid Lab

NIH-sponsored Workshop on
Theoretical and Computational Biophysics

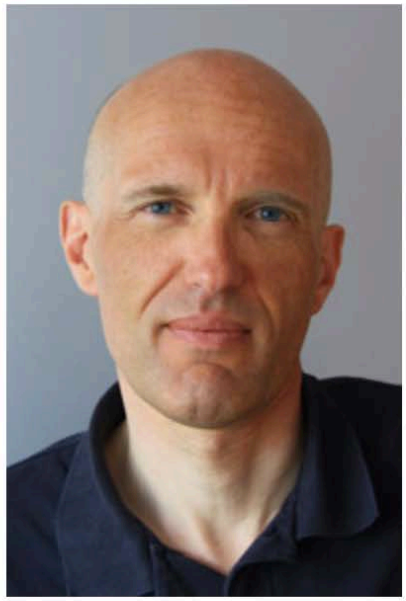
2018

"Hands-on" Workshop on Enhanced Sampling and Free-Energy Calculation at Urbana, IL



Workshop Instructors

Emad Tajkhorshid

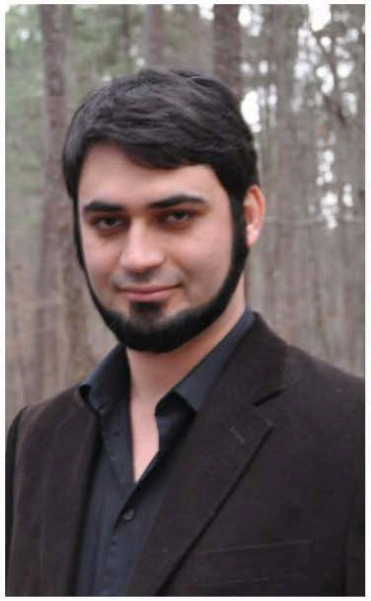


Chris Chipot

Brian K. Radak

Jérôme Hénin

Wei Jiang



Mahmoud Moradi

Rafael C. Bernardi

Program

(subject to changes)

Location: Beckman Institute, Room 5602

Monday, September 10: Alchemical and Geometrical Free-Energy Calculations

09:00-09:20	Welcome and Brief Overview , Emad Tajkhorshid
09:20-09:45	**Applications of Enhanced Sampling and Free-Energy Calculation Methods in Modern Biophysical Problems, Emad Tajkhorshid**
09:45-10:00	**Introduction to Alchemical and Geometrical Free-Energy Calculations, Chris Chipot**
10:00-10:20	Coffee Break
10:20-12:20	**Alchemical and Geometrical Free-Energy Calculations in NAMD, Chris Chipot**
12:20-12:40	Q & A
12:40-14:00	Lunch Break
14:00-15:20	Tutorials
15:20-15:40	Coffee Break
15:40-18:00	Tutorials

Tuesday, September 11: Transition Path Sampling Methods and Constant pH Simulations

09:00-10:20	**Transition Path Sampling Methods, Chris Chipot**
10:20-10:40	Coffee Break
10:40-12:00	**Hybrid Non-Equilibrium Molecular Dynamics/Metropolis Monte Carlo Calculations for Constant pH Simulations, Brian K. Radak**
12:00-12:20	Q & A
12:20-14:00	Lunch Break
14:00-15:20	Tutorials
15:20-15:40	Coffee Break
15:40-18:00	Tutorials

Wednesday, September 12: Geometrical Transformations and Collective Variables

09:00-10:20	**Geometrical Free-Energy Methods: Strengths and Limitations, Jérôme Hénin**
10:20-10:40	Coffee Break
10:40-11:40	**Designing, Implementing and Optimizing Collective Variables in VMD and NAMD, Jérôme Hénin**
11:40-12:00	Q & A
12:00-13:20	Lunch Break
13:20-15:30	Tutorials
15:30-16:30	Coffee Break + Meet the Developers
16:30-16:50	Group Picture and Social
16:50-18:00	Tutorials
19:00-21:00	Workshop Dinner (place to be determined; sign in during the workshop)

Thursday, September 13: Specialized Algorithms for Enhanced Ergodic Sampling

09:00-10:20	**Specialized Algorithms for Enhanced Ergodic Sampling, Chris Chipot**
10:20-10:40	Coffee Break
10:40-12:00	**Accelerating Convergence of Free-Energy Calculation with Replica Exchange Solute Tempering, Wei Jiang**
12:00-12:20	Q & A
12:20-14:00	Lunch Break
14:00-15:20	Tutorials
15:20-15:40	Coffee Break
15:40-18:00	Tutorials

Friday, September 14: Complex Reaction Pathways & QM/MM Simulations

09:00-10:20	**Exploring Complex Reaction Pathways, Mahmoud Moradi**
10:20-10:40	Coffee Break
10:40-12:00	**Free-Energy Calculations and Enhanced Sampling Methods in conjunction with QM/MM calculations, Rafael C. Bernardi**
12:00-12:20	Q & A
12:20-14:00	Lunch Break
14:00-15:20	Tutorials
15:20-15:40	Coffee Break
15:40-18:00	Tutorials

Tutorials

Below are planned tutorials listed by workshop day. More TCBG tutorials are available **here**.

Monday, September 10: Alchemical and Geometrical Free-Energy Calculations

- **A Tutorial on Alchemical Free Energy Perturbation Calculations in NAMD**
- **A Tutorial on Adaptive Biasing Force Calculations in NAMD**
- **Protein:ligand Standard Binding Free Energies: A Tutorial for Alchemical and Geometrical Transformations**

Tuesday, September 11: Transition Path Sampling Methods and Constant pH Simulations

- **String Method with Swarms of Trajectories: A Tutorial for Free-energy Calculations along a Minimum-action Path**
- **Constant pH tutorial**

Wednesday, September 12: Geometrical Transformations and Collective Variables

- **Colvars module (source code and supporting material)**
- **Performing Metadynamics Simulations Using NAMD**
- **Protein:ligand Standard Binding Free Energies: A Tutorial for Alchemical and Geometrical Transformations**
- **A Tutorial on Adaptive Biasing Force Calculations in NAMD**

Thursday, September 13: Specialized Algorithms for Enhanced Ergodic Sampling

- **Methods for calculating Potentials of Mean Force**
- **A Tutorial on One-dimensional Replica-exchange Umbrella Sampling**
- **Adaptive Multilevel Splitting Method: Isomerization of Alanine Dipeptide**

Friday, September 14: Complex Reaction Pathways & QM/MM Simulations

- **Exploring Complex Conformational Transition Pathways**
- **NAMD-QM/MM Tutorial**