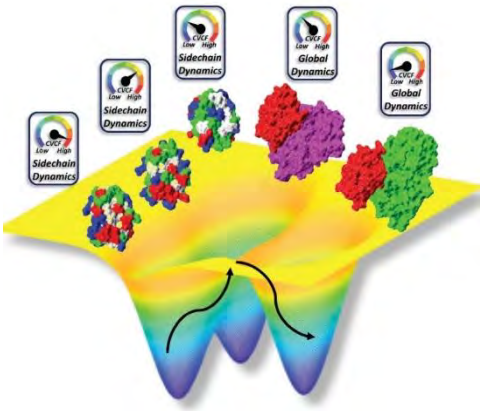
Quantitative Views of Protein Dynamics and the Energy Landscape through Molecular Dynamics Simulations

Sanjoy Paul, Mitradip Das, Krishna Kant Vishwakarma, Ravi Venkatramani

Department of Chemical Sciences
Tata Institute of Fundamental Research (TIFR)



SP

A bit about TIFR-Mumbai: Location

A bit about TIFR: Institute

- A premier national institute (Est. 1945, Under Dept Atomic Energy)
- Graduate programs in Physics, Chemistry, Biology, Mathematics, and Computer Science

- Major campuses all over India.

tifr — *Mumbai*

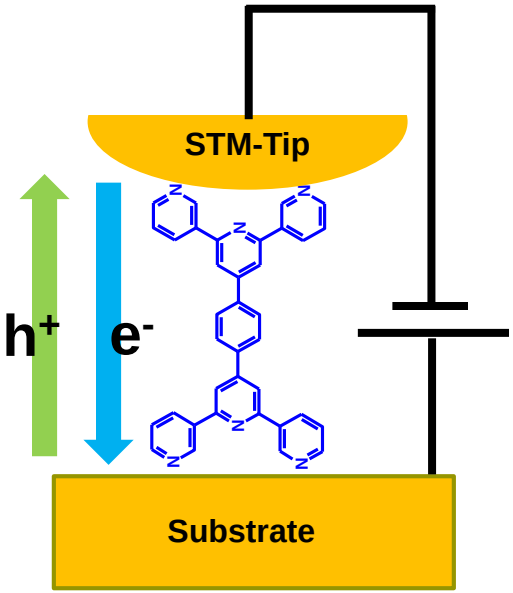
NCRA • TIFR — *Pune*

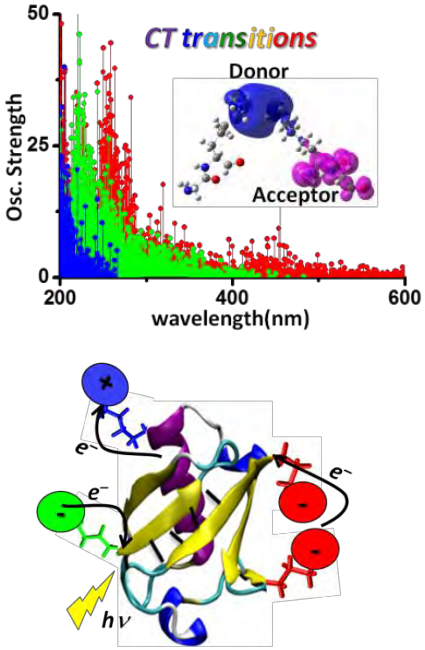
ncbs national centre for biological sciences — *Bangalore*

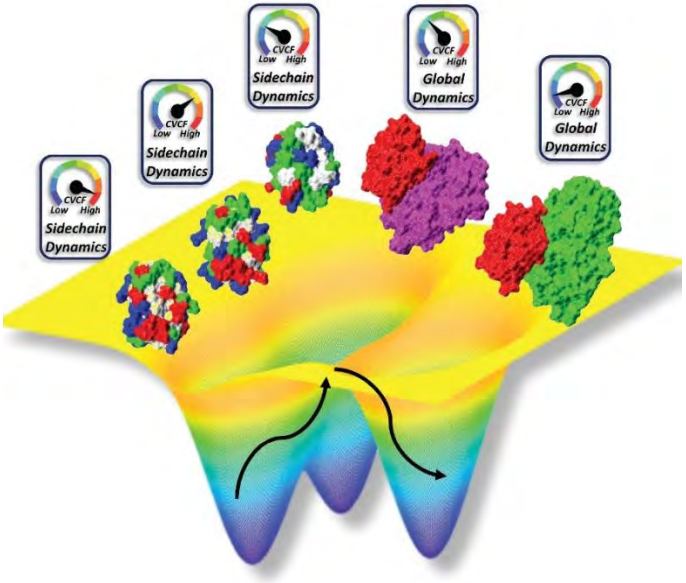
ICTS International Centre for Theoretical Sciences, Tata Institute of Fundamental Research — *Bangalore*

TCIS — *Hyderabad*

Computational Chemistry @ TIFR Mumbai







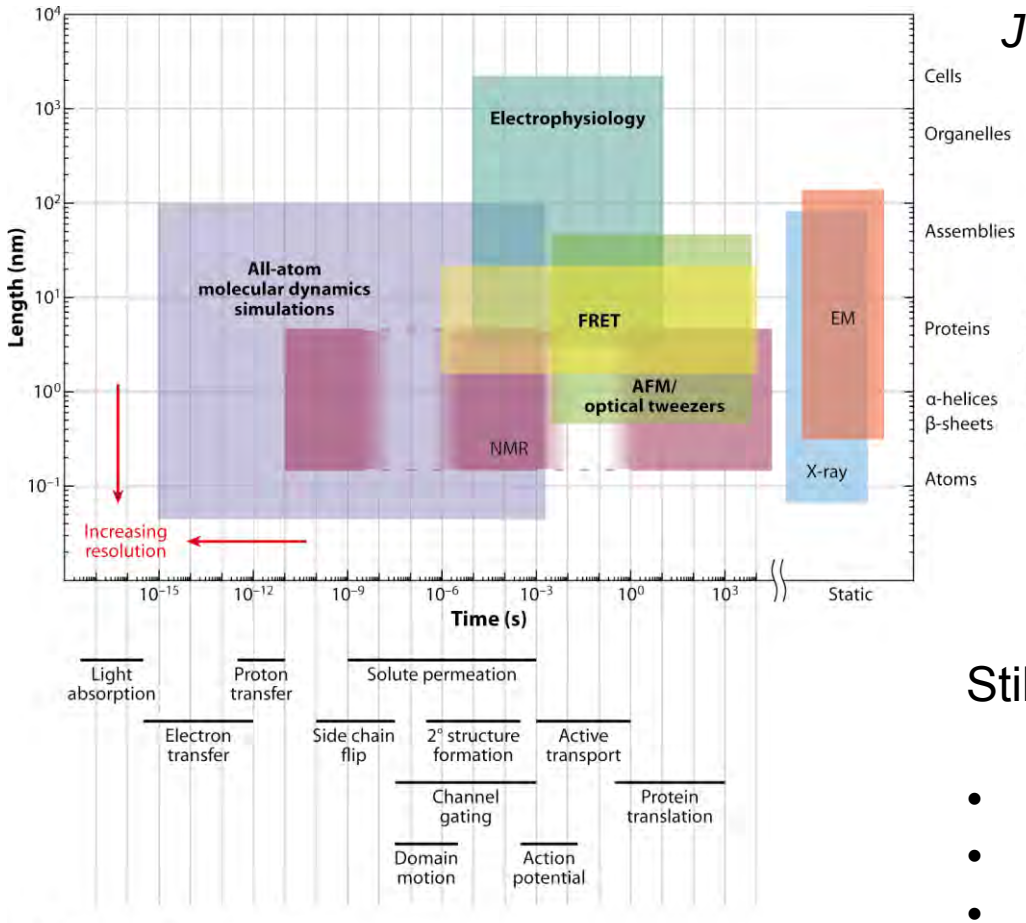
Pushing Electrons through Molecules

Dynamical Metrics to Describe Proteins their Interactions, and functions

Objectives

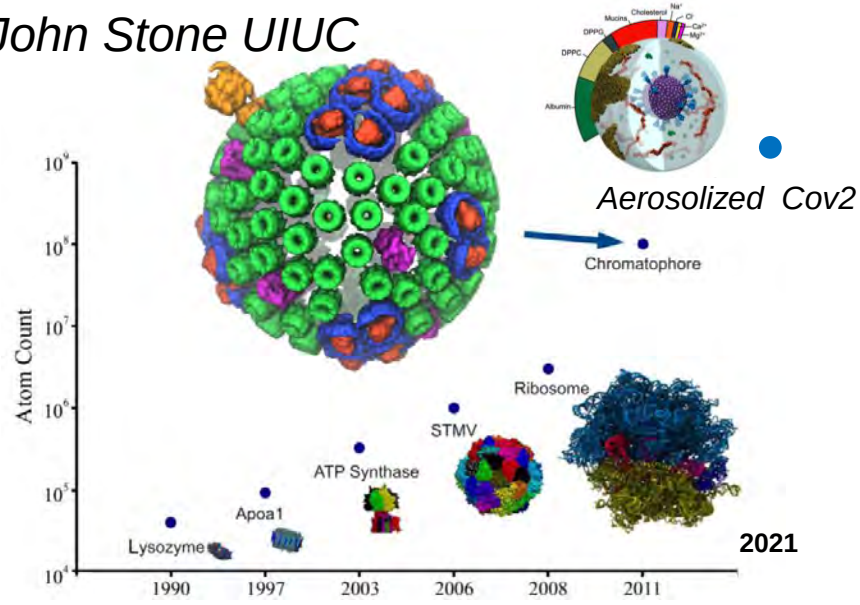
To develop fundamental principles which can enable technological breakthroughs

MD Simulations: A Computational Microscope



Dror RO, et al. 2012.
Annu. Rev. Biophys. 41:429–52

John Stone UIUC



Still a qualitative tool!

- Force Fields
- Implementation of thermostat/barostats
- Polarizability
- Simulation timescales
- Estimations of protein dynamics
- Quantitative comparisions

Is MD a qualitative/quantitative tool?

The Amber Project

Chemistry through a Computational Lens

AmberTools21 | Amber20 | Manuals | Tutorials | Force Fields | Contacts | History

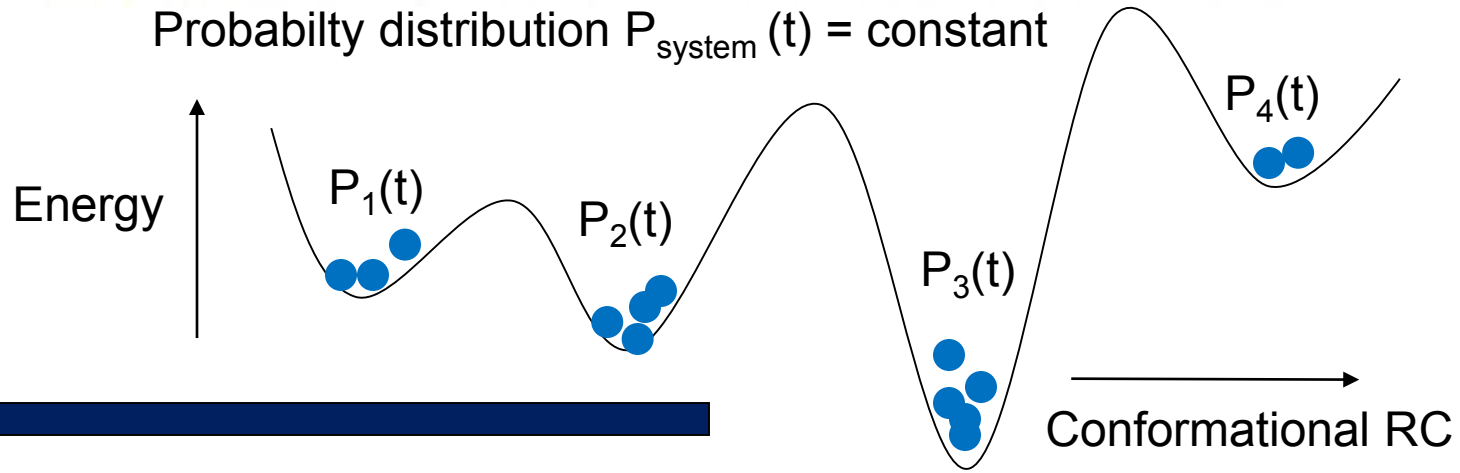
5 Introductory Case Studies

Useful links:
Amber Home
Download Amber
Installation
Amber Citations
GPU Support
Updates
Mailing Lists
For Educators
File Formats

3. Equilibration

In MD simulations, atoms of the macromolecules and of the surrounding solvent undergo a relaxation that usually lasts for tens or hundreds of picoseconds before the system reaches a stationary state. The initial nonstationary segment of the simulated trajectory are typically discarded in the calculation of equilibrium properties. This stage of the MD simulation is called equilibration stage.

Equilibration protocols are still largely a matter of personal preference. Some protocols call for very elaborate procedures involving gradually increasing temperature in a step-wise fashion while other more aggressive approach simply use a linear temperature gradient and heat the system up to the desired temperature.



Questions/objectives

- Is it possible to estimate equilibration along a biomolecular simulation trajectory?
  - Local equilibration
  - Single/Multiple trajectories
  - Comparing protein dynamics, protein stability/flexibility and more.

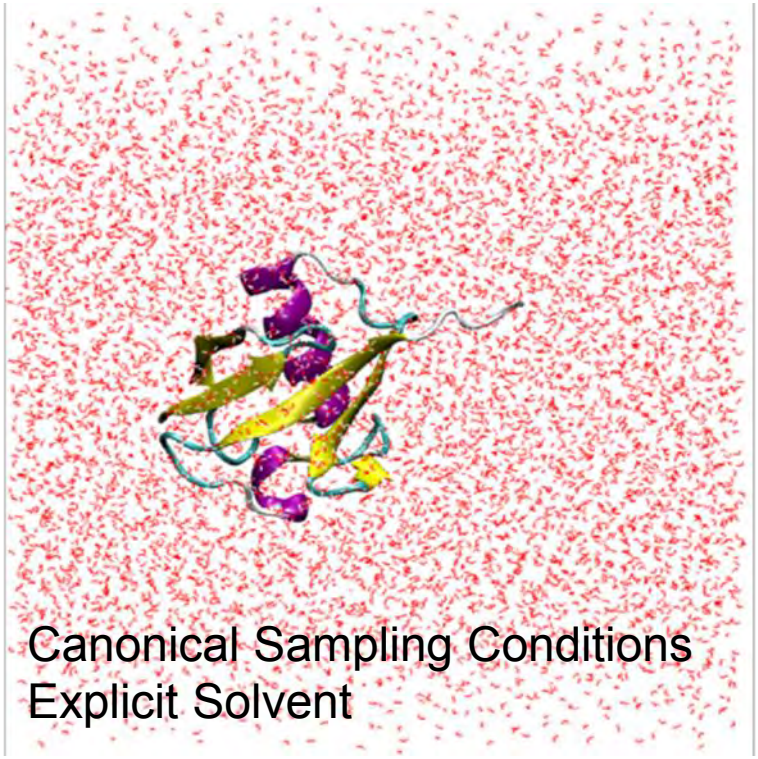
- Is it possible to extract converged reaction coordinates for biomolecular conformational transitions **before** equilibration?
  - Locally converged reaction coordinates
  - Single/Multiple Trajectories
  - Better biased sampling schemes, thermodynamics and kinetics, and more.

Simulation Methods: Monte-Carlo for Model Potentials,
MD for solvated proteins (10-20 x 0.25-1μs)

Protein Dynamics

Atomistic Molecular Dynamics (MD) + Principal Component Analysis (PCA)

Variance-Covariance matrix of atomic fluctuations from simulations



Canonical Sampling Conditions
Explicit Solvent

$$C_{ab} = \langle (x_a - \langle x_a \rangle)(x_b - \langle x_b \rangle) \rangle$$

x= Cartesian coordinates of N atoms; a,b=1……3N;

Solve $(\boldsymbol{C} - \boldsymbol{\lambda I})\boldsymbol{\xi} = \boldsymbol{0}$ to get:

Eigenvectors (Principal components) $\xi=(\xi_1,\xi_2,\ldots,\xi_{3N})$

Eigenvalues (variances) $\boldsymbol{\lambda}=(\lambda_1, \lambda_2,\ldots,\lambda_{3N})$

$$\sigma^2_{CVCF}(t) = Tr(\boldsymbol{C}(t)) = \sum_{i=1}^{3N-6} \lambda_i(t)$$

Cumulative Variance of Coordinate Fluctuations

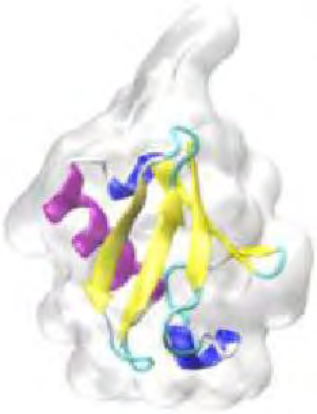
-Can be applied to any subset of atoms
-Can extract directional spring constants

Equipartition
+
Thermal Equilibrium

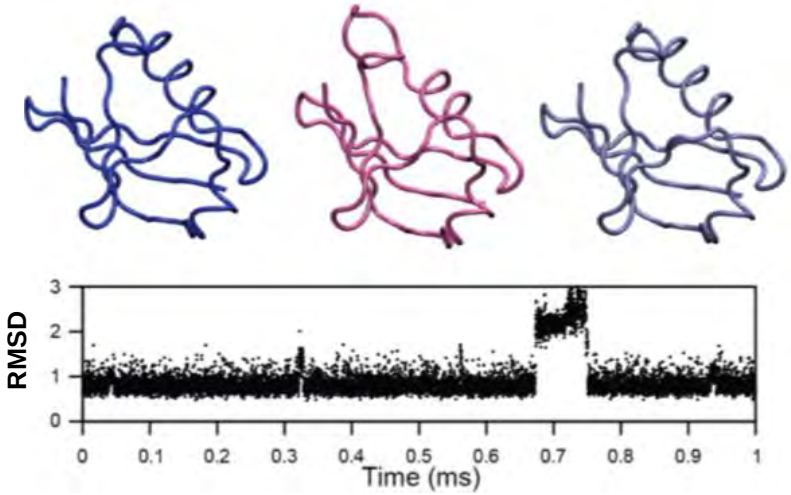
$$k_{CVCF} = \frac{k_B T}{(\sigma_{CVCF})^2}$$

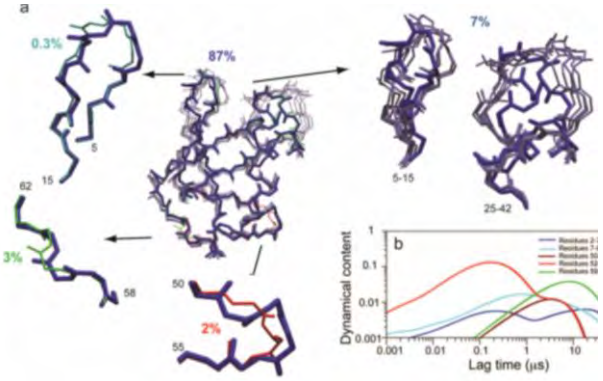
Assessing Equilibration in MD

D.E.Shaw et. al. JPC-B (2016)



Ubiquitin (Ub)



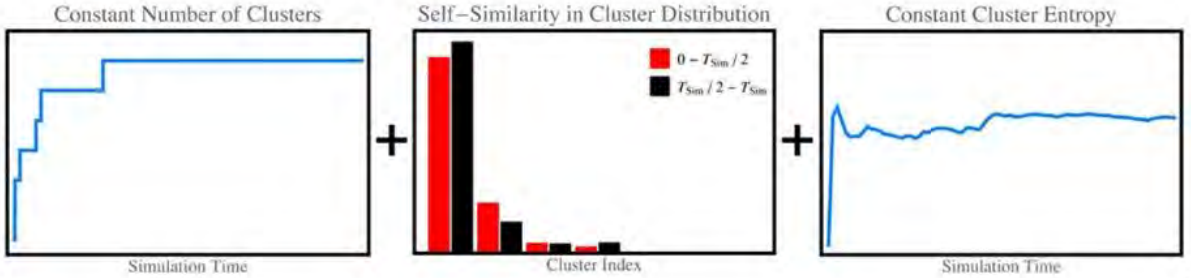


Bovine Pancreatic Trypsin Inhibitor BPTI

N_C

$$S_c = \sum_{i=1}^{N_C} P_i log P_i$$

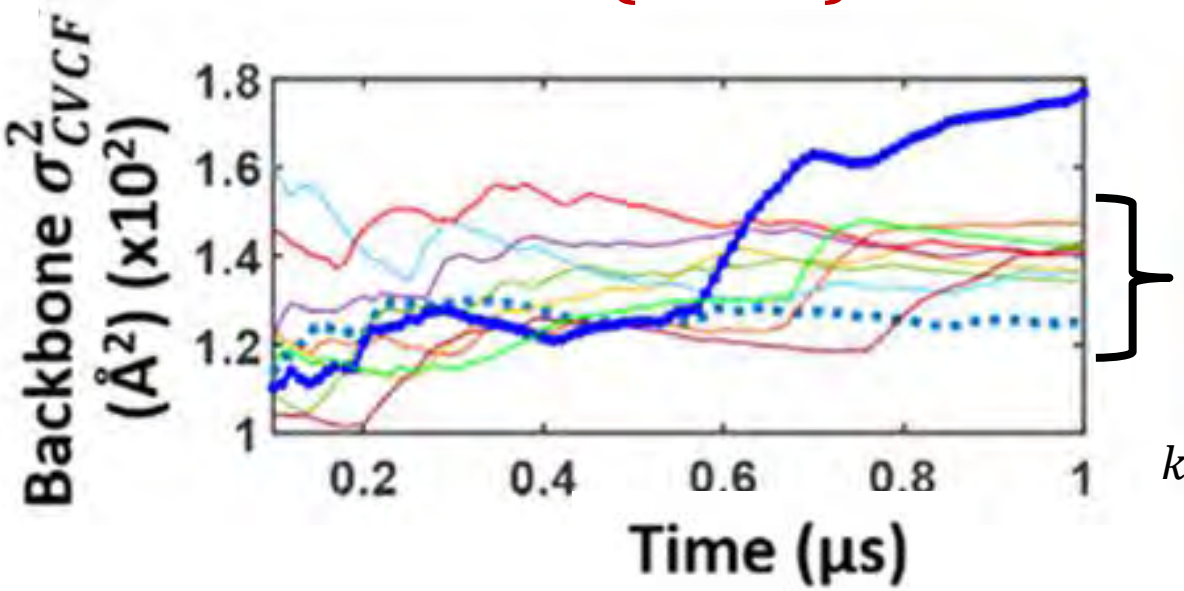
Multi-μs long native-state simulation



Simulation is self-consistent, but may not be equivalent to *True* convergence

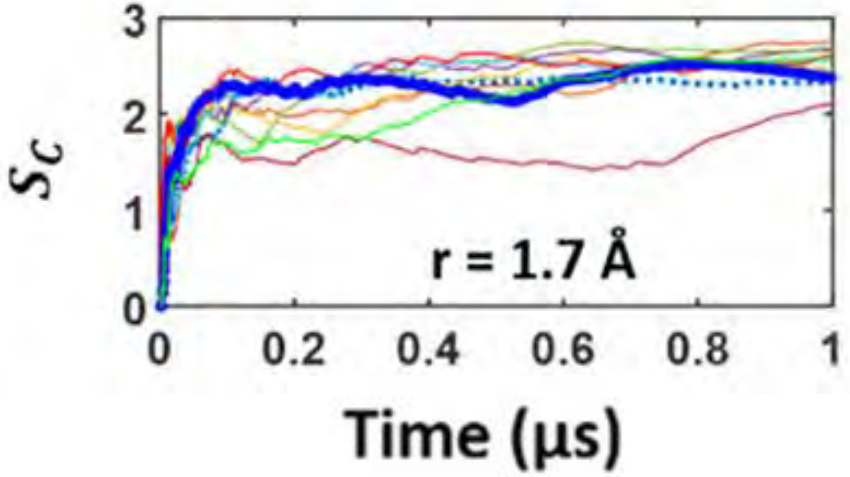
Sawle & Ghosh JCTC (2016)

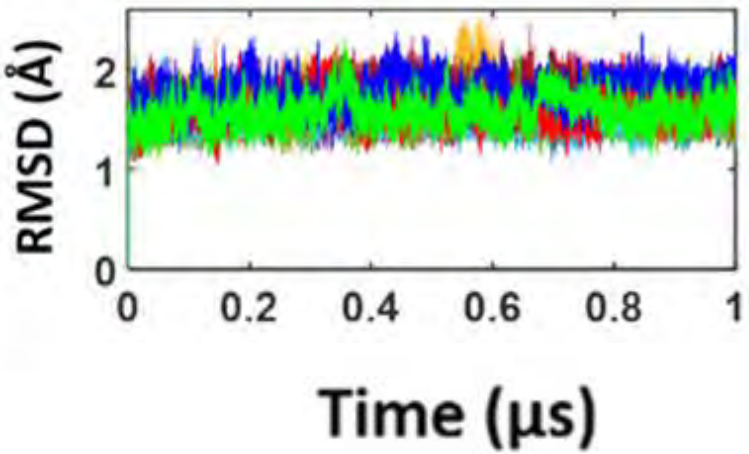
Cumulative Variance of Coordinate Fluctuations (CVCF)



Local Equilibria
Boltzmann Statistics!

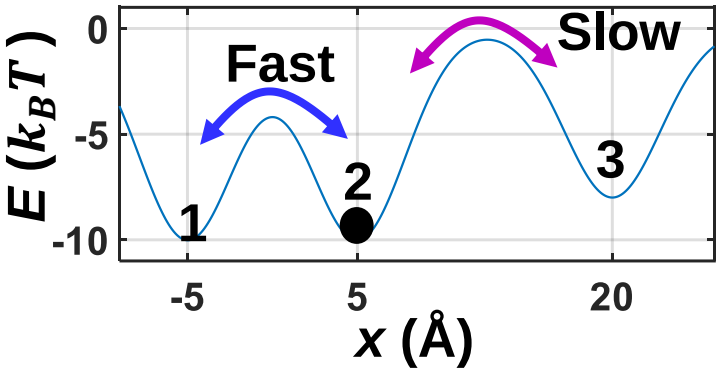
$$k_{avg}(t) = \frac{k_B T}{\langle \sigma^2_{CVCF}(t) \rangle}$$

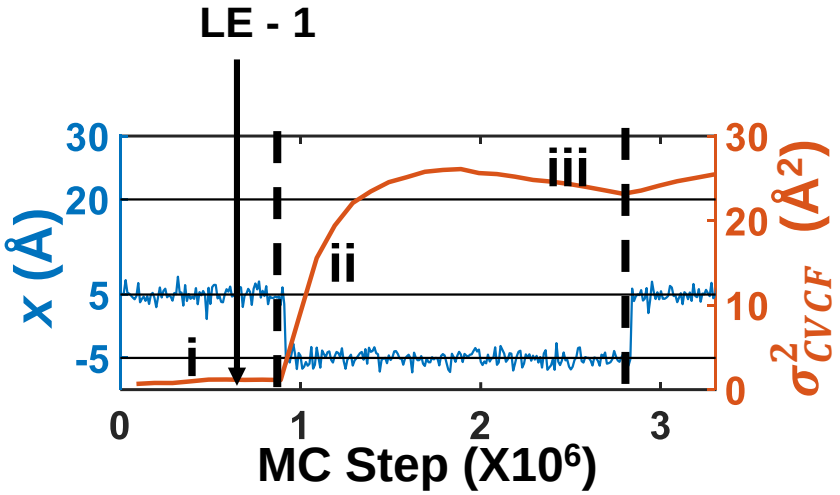


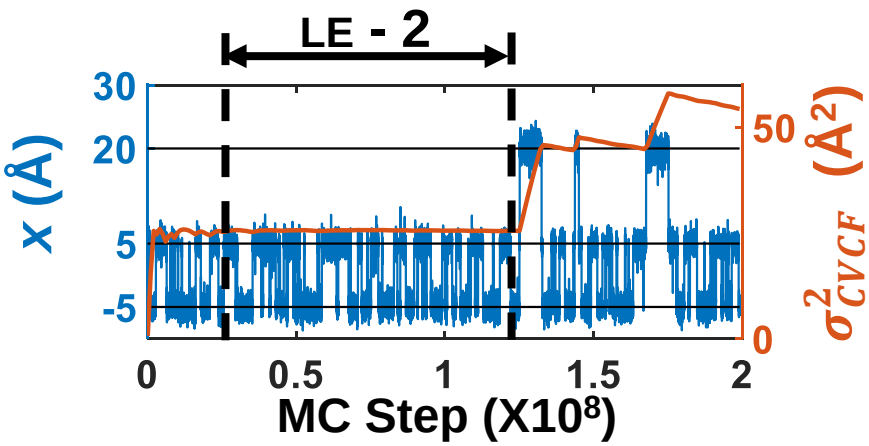


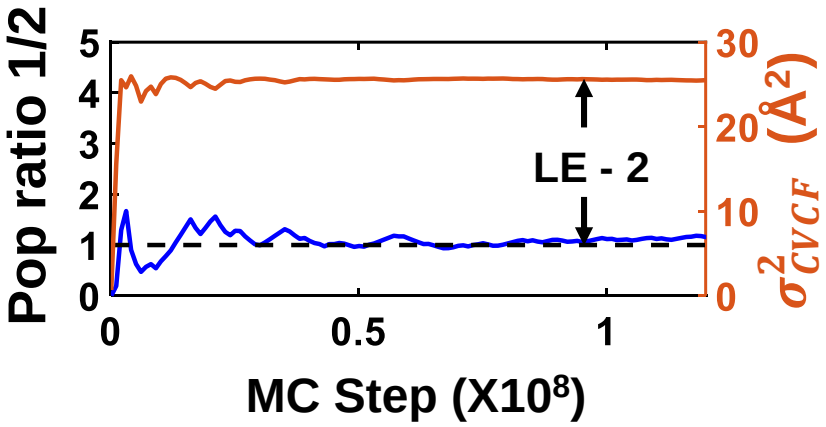
Paul, Koti, and RV JPC-B(2020)

CVCF: A metric to evaluate local equilibration



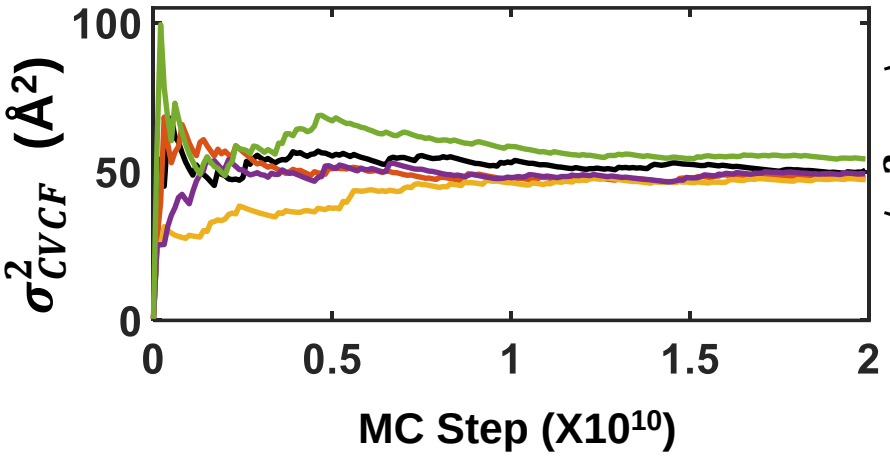


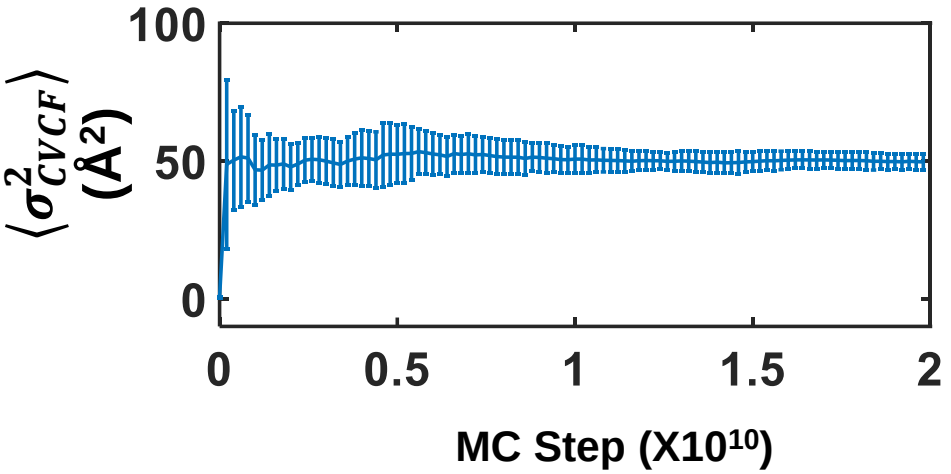


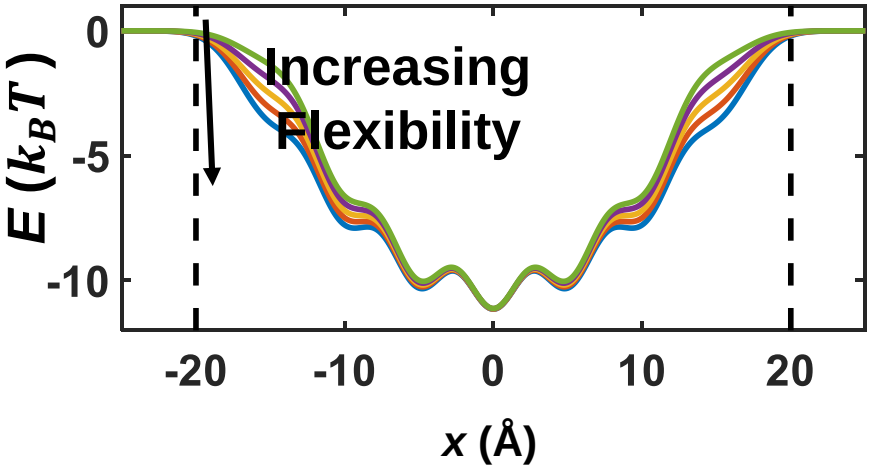


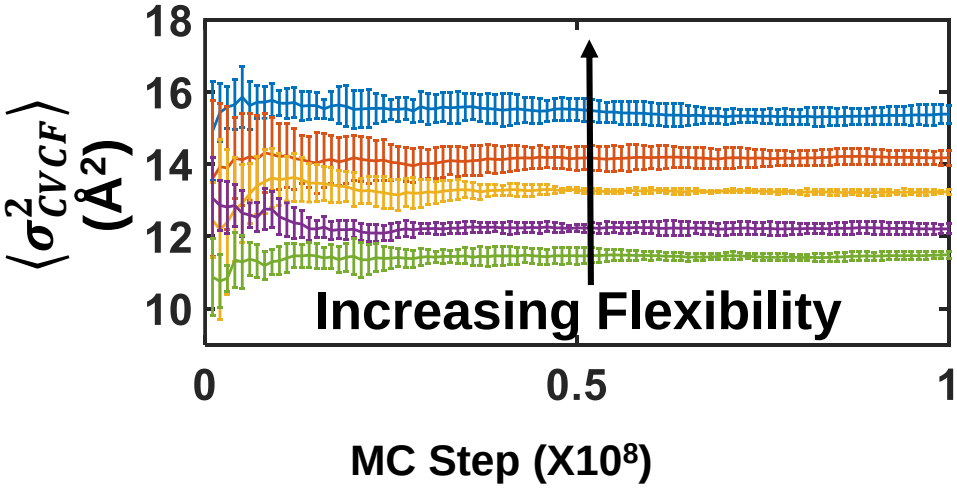
Paul, Koti, and RV JPC-B(2020)

CVCF and the energy landscape



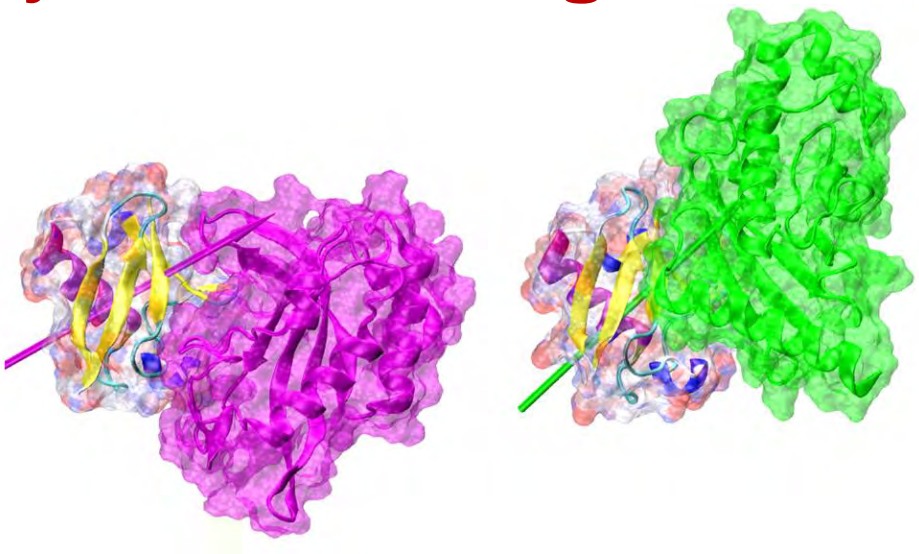


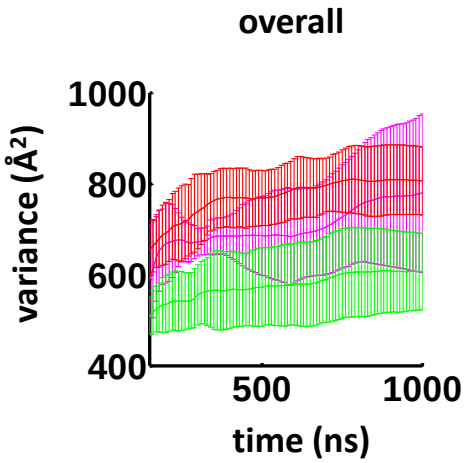


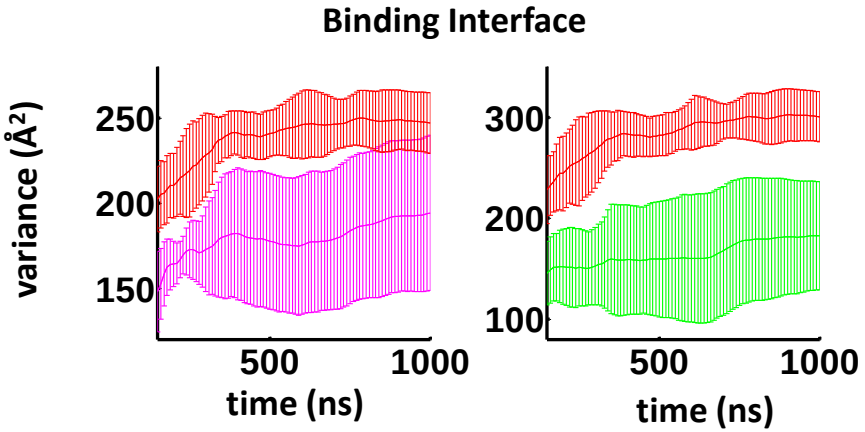


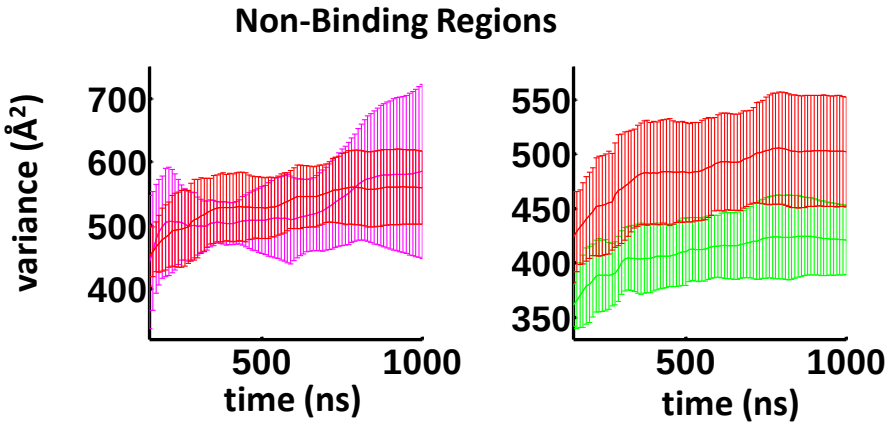
Paul, Koti, and RV JPC-B(2020)

Dynamical Changes in Ub with Complexation







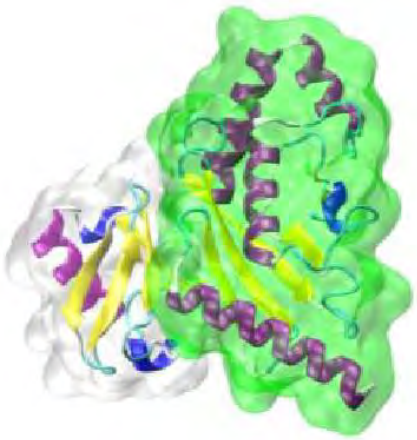


UBQ UBQ:UEV UBQ:UBCH5A

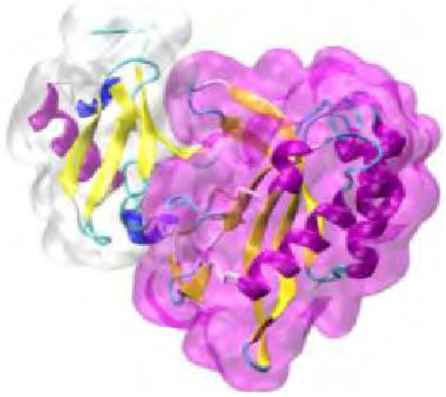
Paul, Koti, and RV JPC-B(2020)

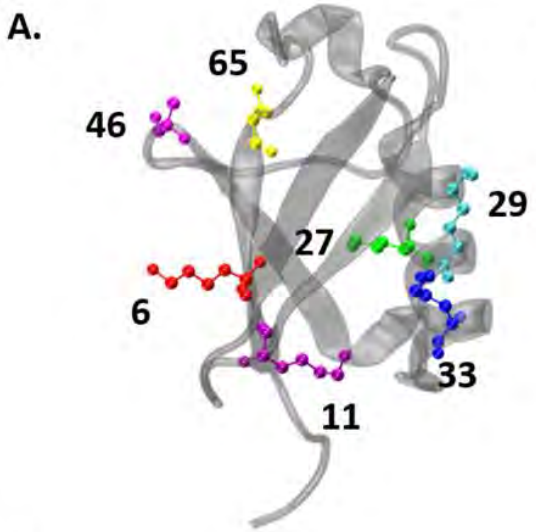
Some SMFS Predictions for Ub

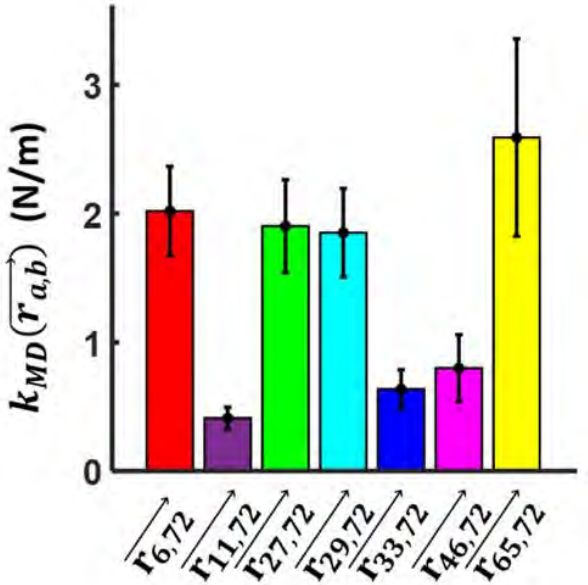
Ub:UBCH5A

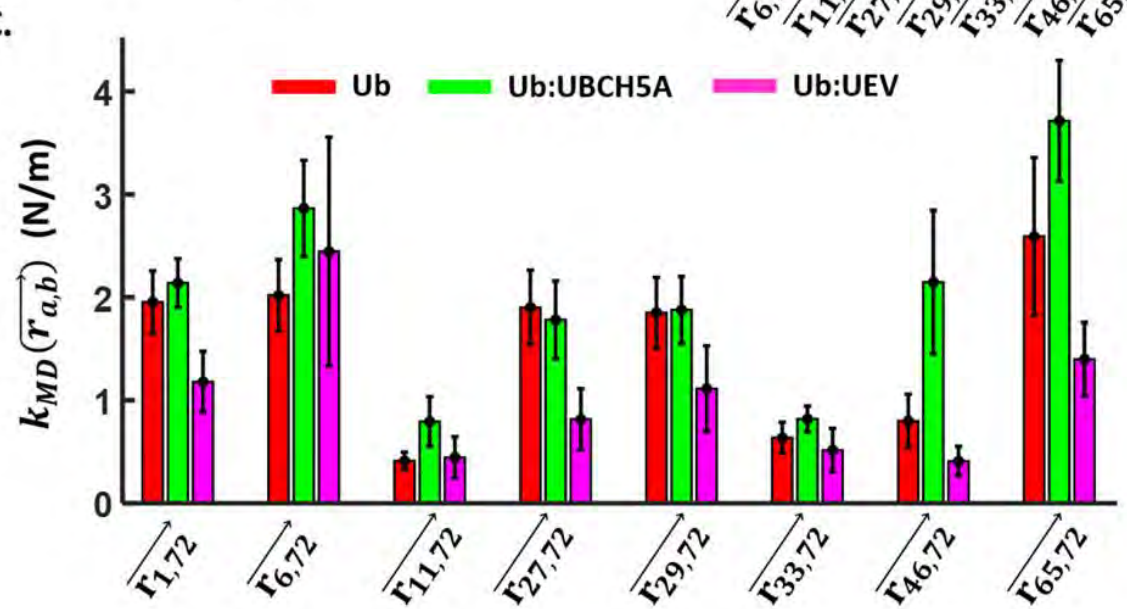


Ub:UEV









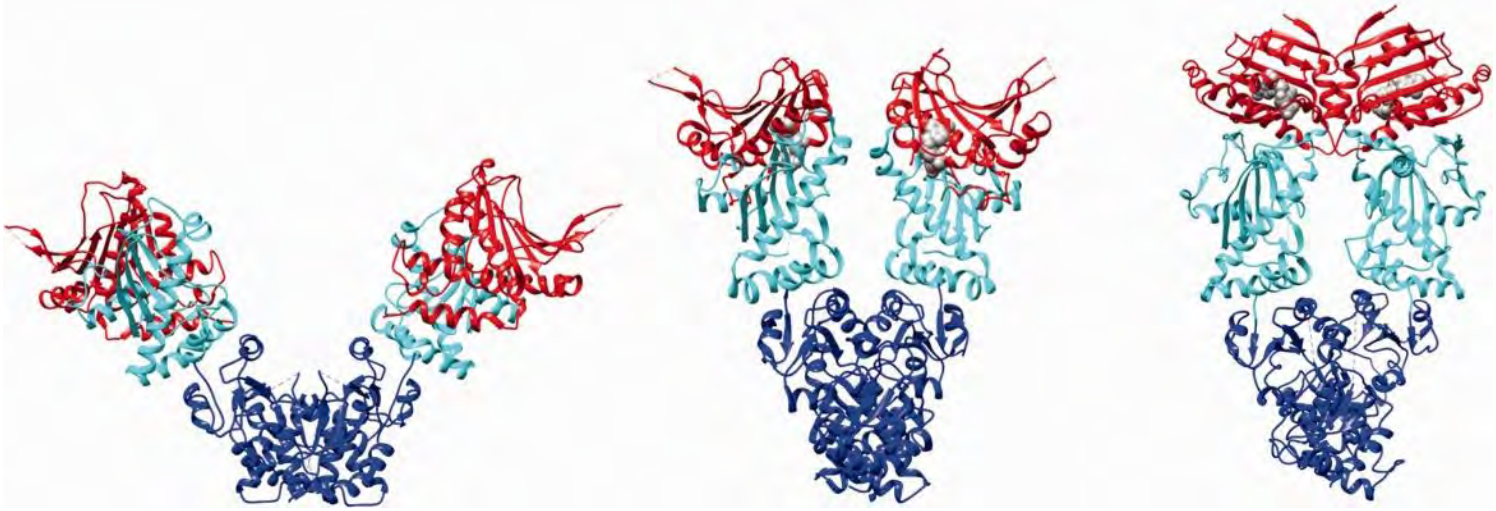
Paul & RV JCTC (2021)

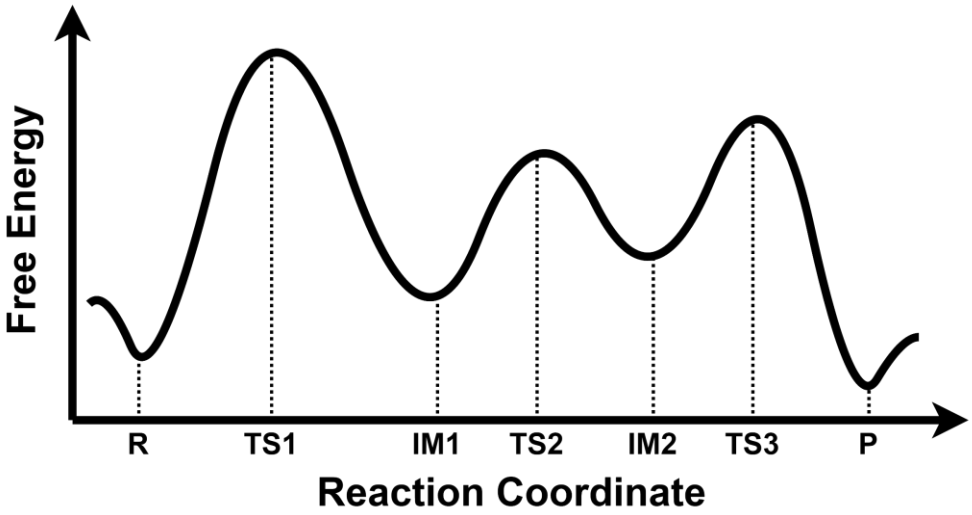
Questions/objectives

- Is it possible to estimate equilibration along a biomolecular simulation trajectory?
  - Local equilibration
  - Single/Multiple trajectories
  - Comparing protein dynamics, protein stability/flexibility and more.

- Is it possible to extract converged reaction coordinates for biomolecular conformational transitions **before** equilibration?
  - Locally converged reaction coordinates
  - Single/Multiple Trajectories
  - Better biased sampling schemes, thermodynamics and kinetics, and more.

Biomolecular Conformational Transitions





Accurate descriptions to RCs are essential for **quantitative description of free energy surface**, which provides information about **thermodynamics** and **kinetics** associated with conformational changes.

Biased Sampling methods can help

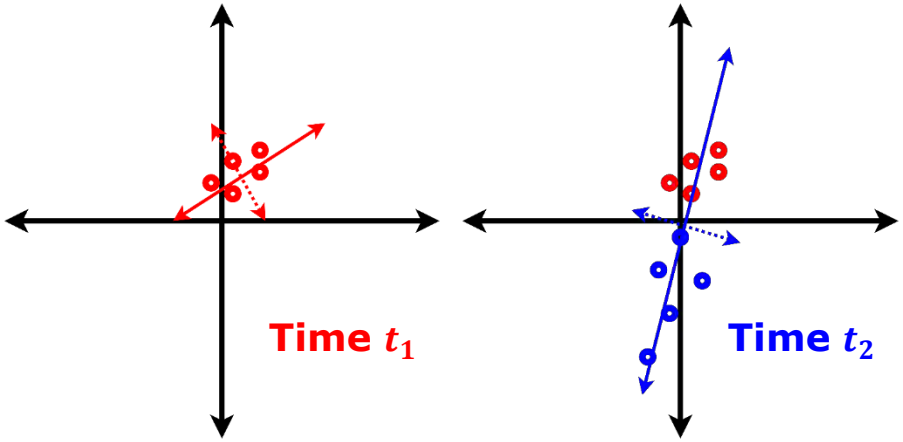
…. but with good reaction coordinates

Goal: *Extract reaction coordinates from dimensionality reduction techniques: e.g PCA*

The Mode Evolution Metric

Das & RV In-prep

Example: Time evolution of PCA in 2D system

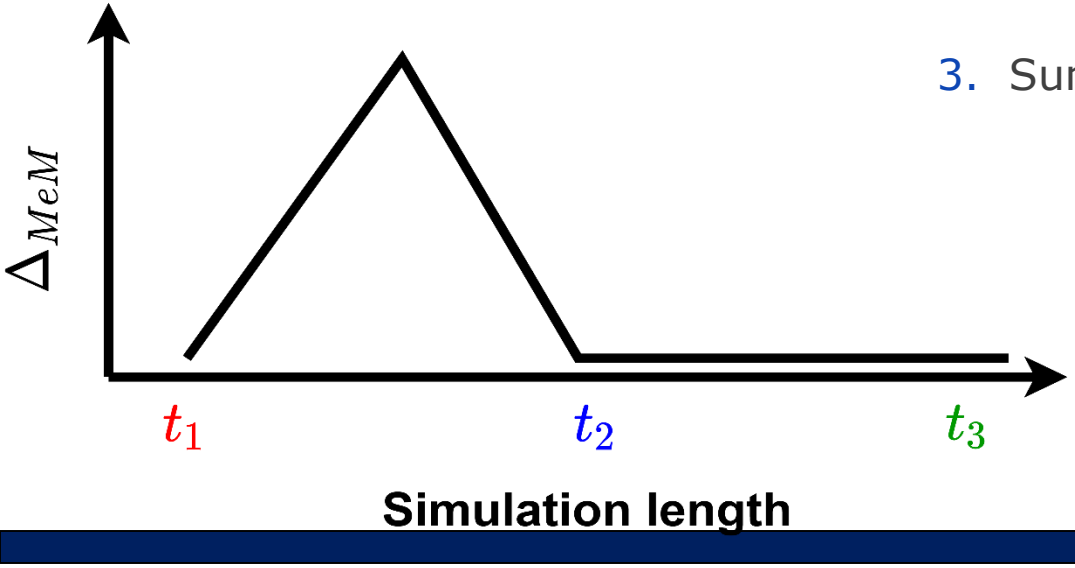


$$\Delta_{\mathrm{MeM}}(t, dt) = \sum_{i=1}^{3N-6} \lambda_i(t) \left[1 - \left|\hat{\xi}_i(t - dt/2) \cdot \hat{\xi}_i(t + dt/2)\right|\right]$$

3 2 1

1. Detects the change in directions. In case there is no change, the overall value is zero.
2. Scales by corresponding variance.
3. Summation over all modes.

Expected character of the metric:



Features:

- Zero at equilibrium or no PC change
- Proportional to variance along direction(s) of change.
- Value ranges between zero and total variance.

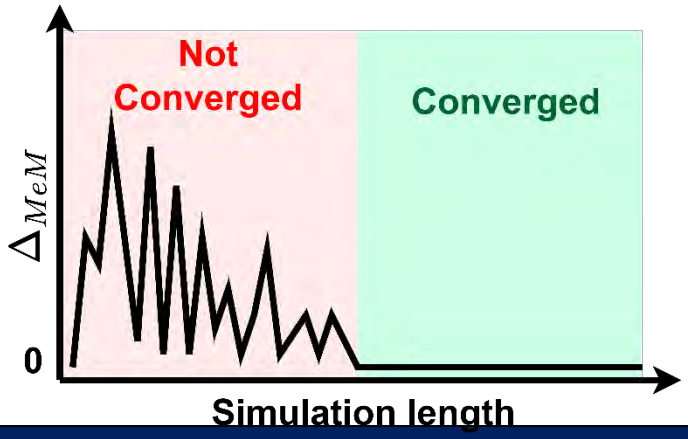
A Hypothesis:

Reaction coordinates of a system converges before the system reaches equilibrium.

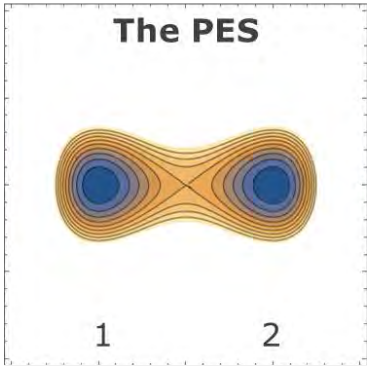
Das & RV In-prep

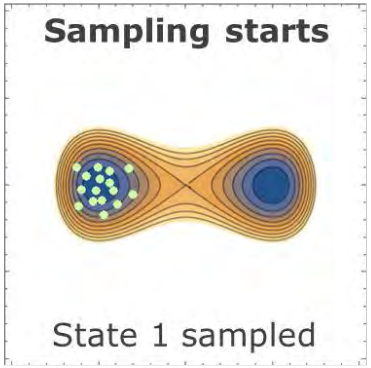
Basis of hypothesis:

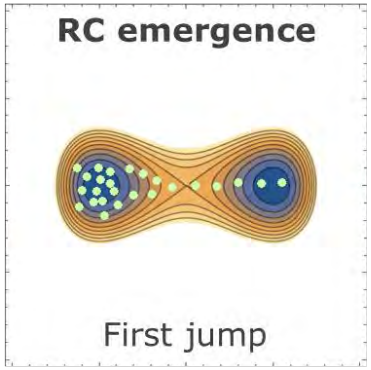
- Often in simulations, the system has rare transitions to states separated by high energy barrier.
- **RC Emergence**: RCs are encoded into the system which may appear during transitions.
- **Equilibrium**: Requires multiple transitions to build up Boltzmann statistics in both states.

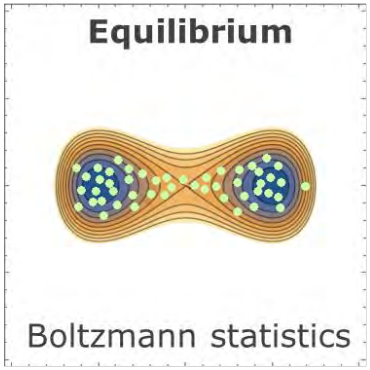


Example: Two state system







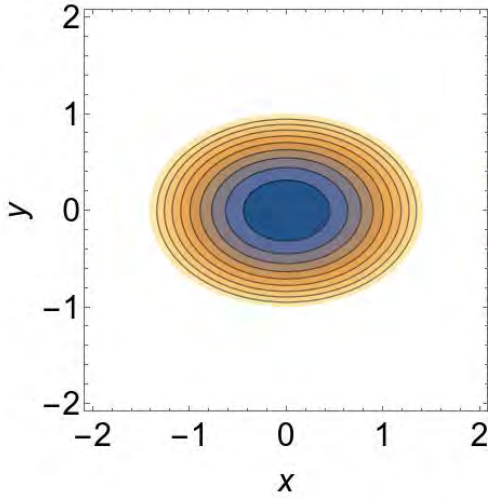


MC on Simple Model Potentials

Das & RV In-prep

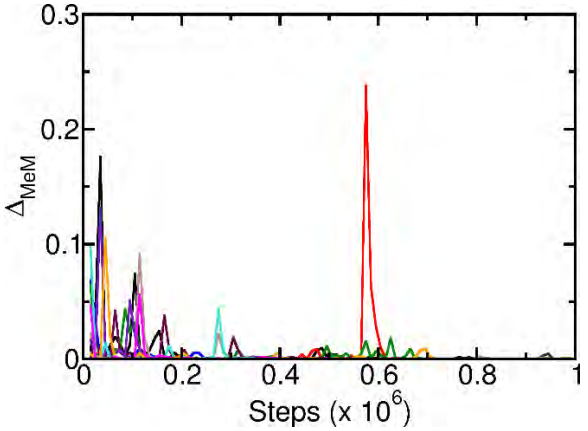
$$V(x,y) = \frac{1}{2}k_x x^2 + \frac{1}{2}k_y y^2$$

$$k_x/k_y = 2$$

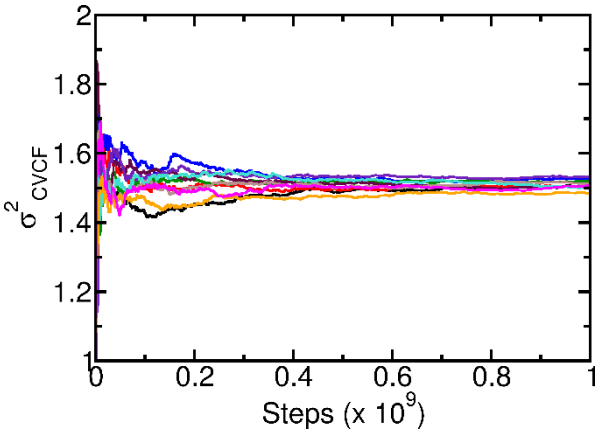


- RCs converge well before system equilibrates.

MeM



CVCF

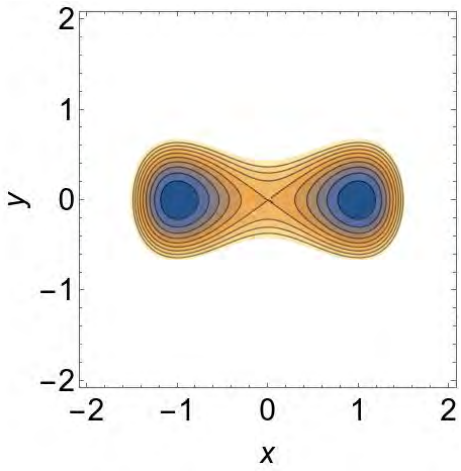


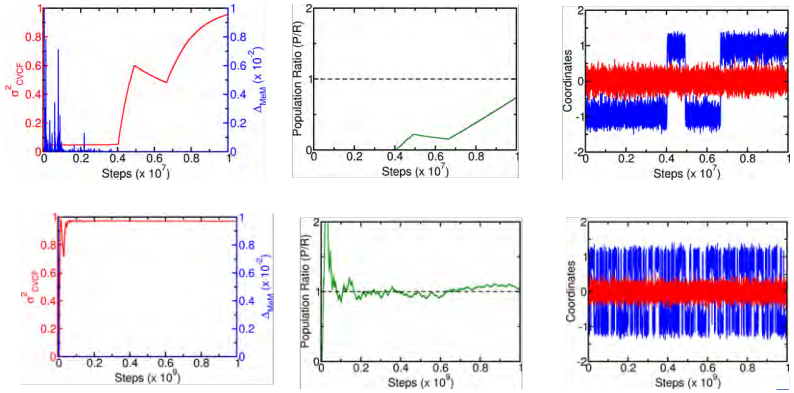
RC convergence steps	7×10^5
Equilibration steps (range)	$(2.3 - 6.1) \times 10^8$
Equilibration steps (mean)	3.8×10^8

MC on Simple Model Potentials

Das & RV In-prep

$$V(x,y) = 6 \times \left((x-1)^2 \times (x+1)^2 + 3.7 \times y^2\right)$$



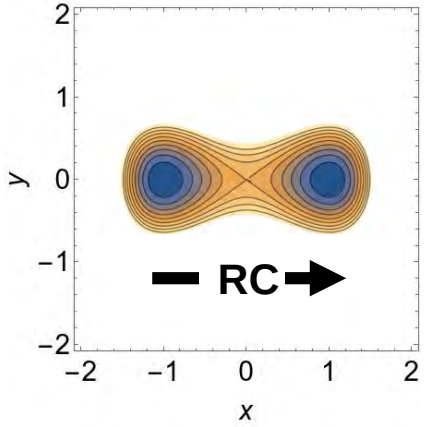


Traj. No.	Convergence (Steps)	First visit to P (Steps)	Equilibration (Steps)
1	1.6×10^6	11.5×10^6	3.7×10^8
2	5.2×10^6	5.2×10^6	6.9×10^8
3	0.5×10^6	0.6×10^6	6.2×10^8
4	4.5×10^6	4.5×10^6	3.0×10^8
5	2.1×10^6	4.7×10^6	4.6×10^8
6	0.9×10^6	1.0×10^6	3.0×10^8
7	4.1×10^6	4.1×10^6	1.7×10^8
8	2.3×10^6	3.1×10^6	5.3×10^8
9	0.9×10^6	0.9×10^6	6.7×10^8
10	1.9×10^6	1.9×10^6	5.1×10^8
Avg.	2.4×10^6	3.8×10^6	4.6×10^8

RCs converge well before system equilibrates.

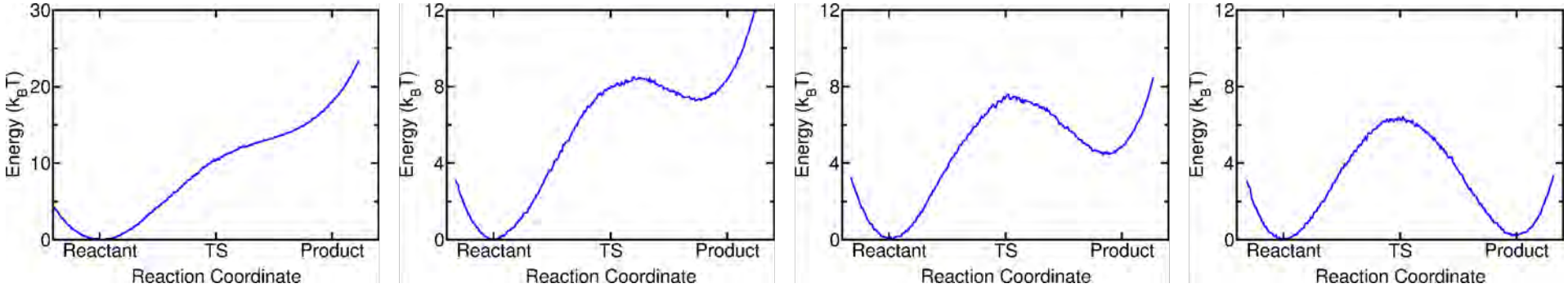
- RCs converge during or before the system first visits product state!

Converged RCs for Metadynamics



Das & RV In-prep

Slices along first PC

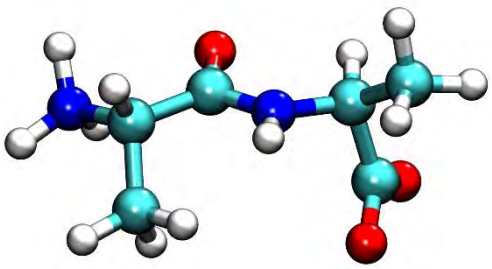


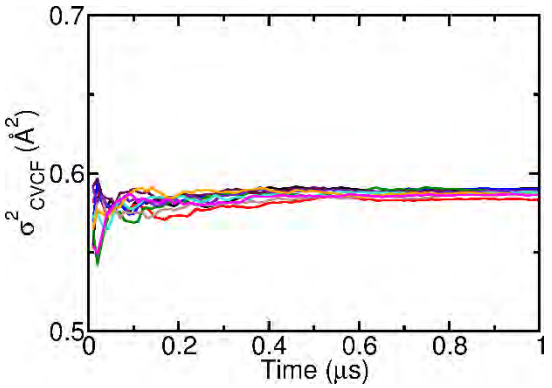
Steps	∞	6×10^5	9×10^5	1×10^6	2×10^6
Direction	$(1.0, 0.0)$	$(0.81, 0.59)$	$(0.90, 0.43)$	$(0.94, 0.43)$	$(1.0, 0.0)$
E(P)-E(R)	0	$??$	$7.3k_BT$	$4.5k_BT$	$0.2k_BT$
E(TS)-E(R)	$6k_BT$	$??$	$8.5k_BT$	$7.5k_BT$	$6.2k_BT$

Modest speedup of 30x (not optimized!)

Application to Alanine Dipeptide

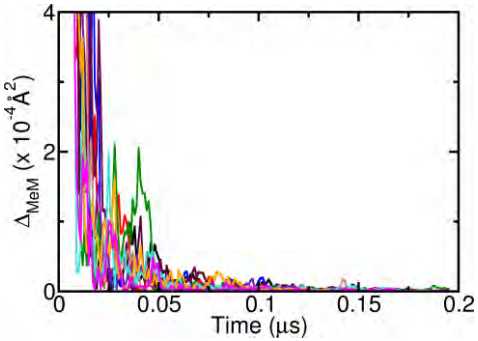
Das & RV In-prep

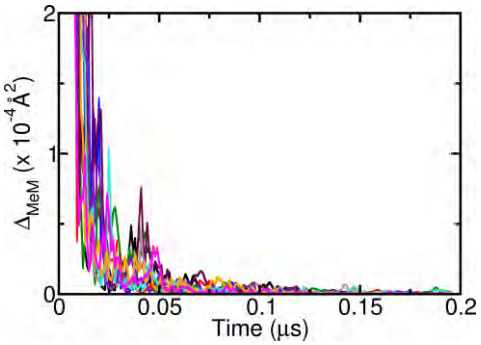


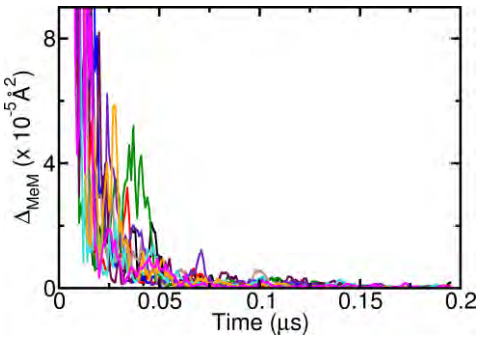


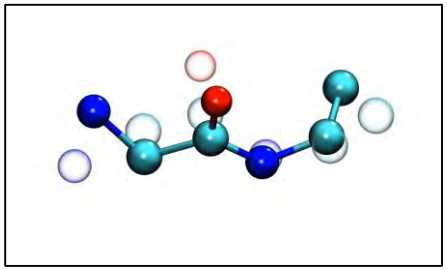
The system equilibrates ~0.36 μs (average).

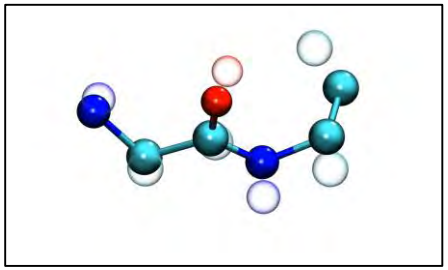
All modes converge within $0.05 - 0.2\ \mu s$.

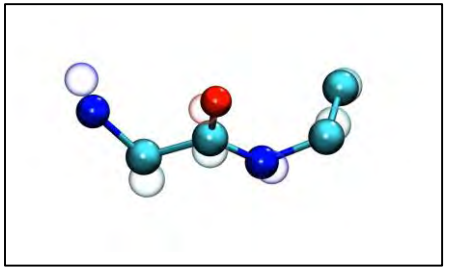










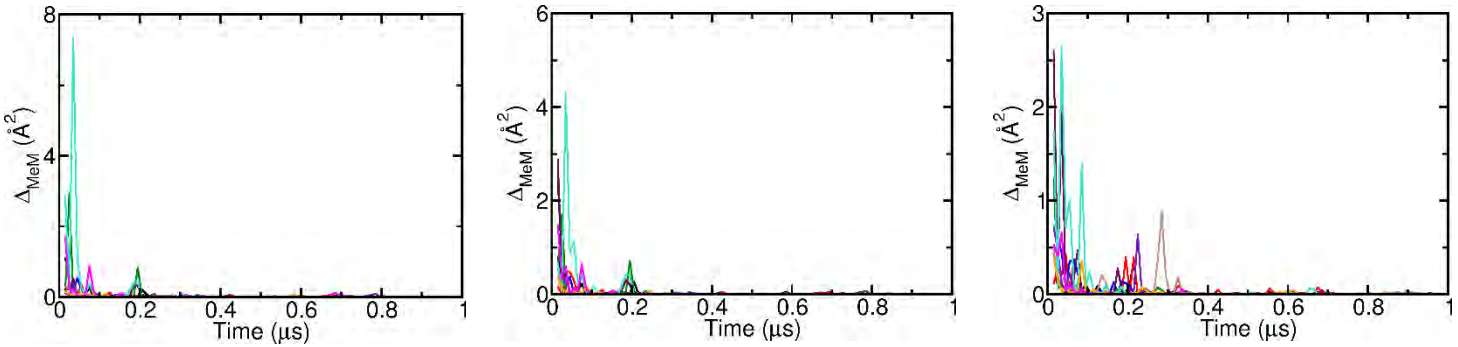


Application to ubiquitin

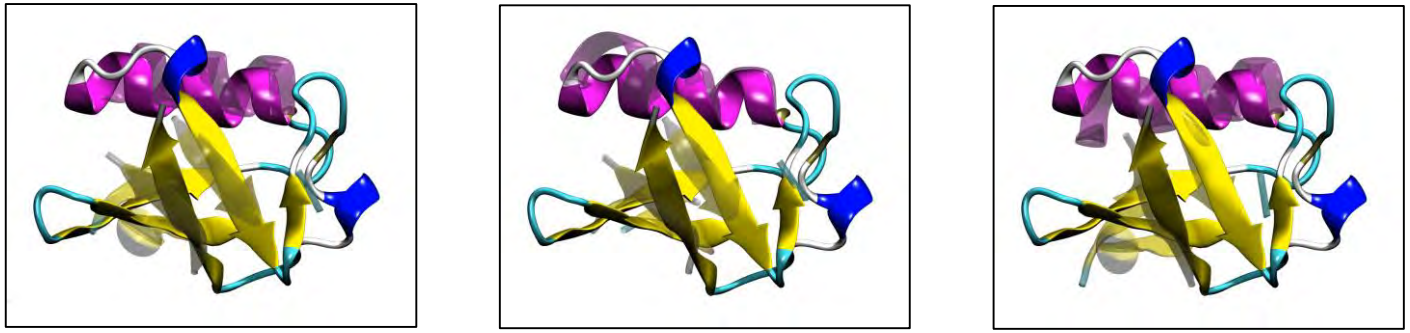
Das & RV In-prep

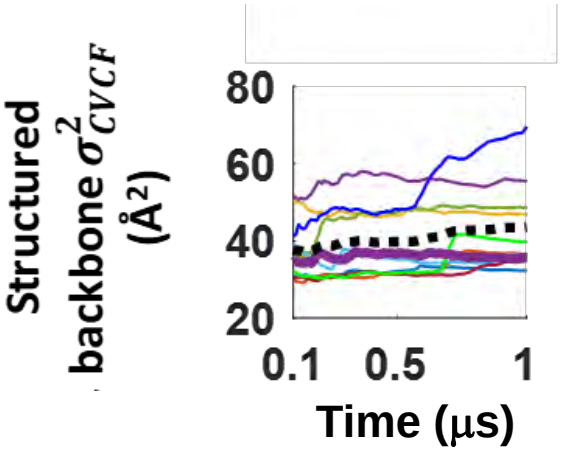
PC 1 **PC 2** **PC 3**

MeM



PC Mode





Questions/Objectives

- Is it possible to estimate equilibration along a biomolecular simulation trajectory?
  - Local equilibration
  - Single/Multiple trajectories
  - Comparing protein dynamics, protein stability/flexibility and more.

- Is it possible to extract converged reaction coordinates for biomolecular conformational transitions **before** equilibration?
  - Locally converged reaction coordinates
  - Single/Multiple Trajectories
  - Better biased sampling schemes, thermodynamics and kinetics, and more.

Acknowledgements

Funding: Department of Atomic Energy, Govt of India
Project no. 12-R&D-TFR-5.10-0100

Collaborators: Prof. A. S. R. Koti (Chemical Sciences)
Prof. Ullas Kolthur (Biological Sciences)