Application of the Markov State Model to Molecular Dynamics of Biological Molecules

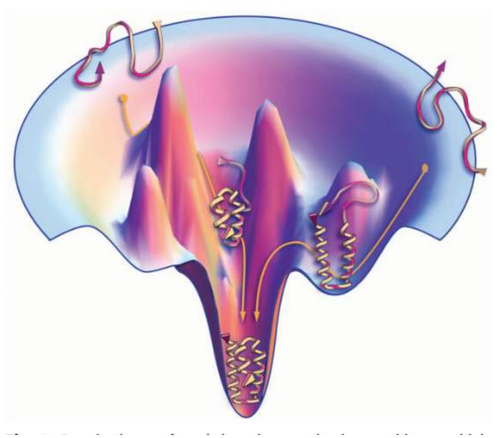
Speaker: Xun Sang-Ni

Supervisor: Prof. Wu

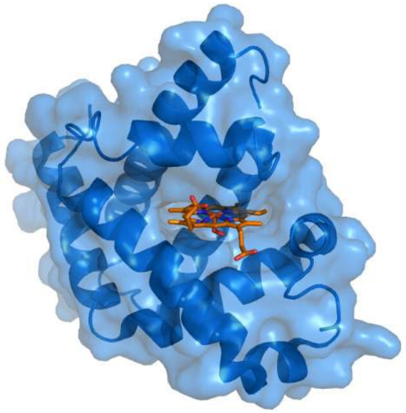
Dr. Jiang

Introduction

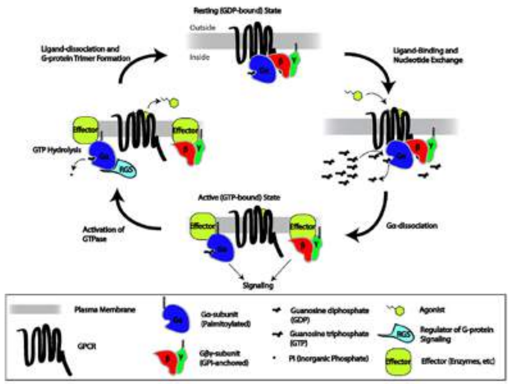
Conformational changes of proteins : essential part of many biological processes



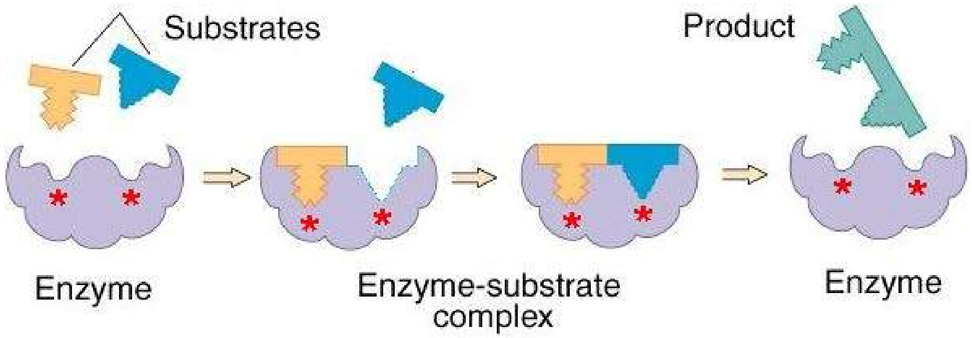
Protein folding



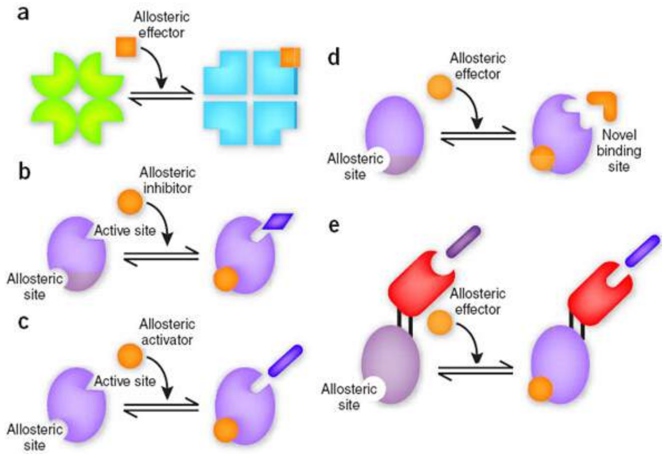
Ligand binding



Signal transduction



Enzymatic catalysis

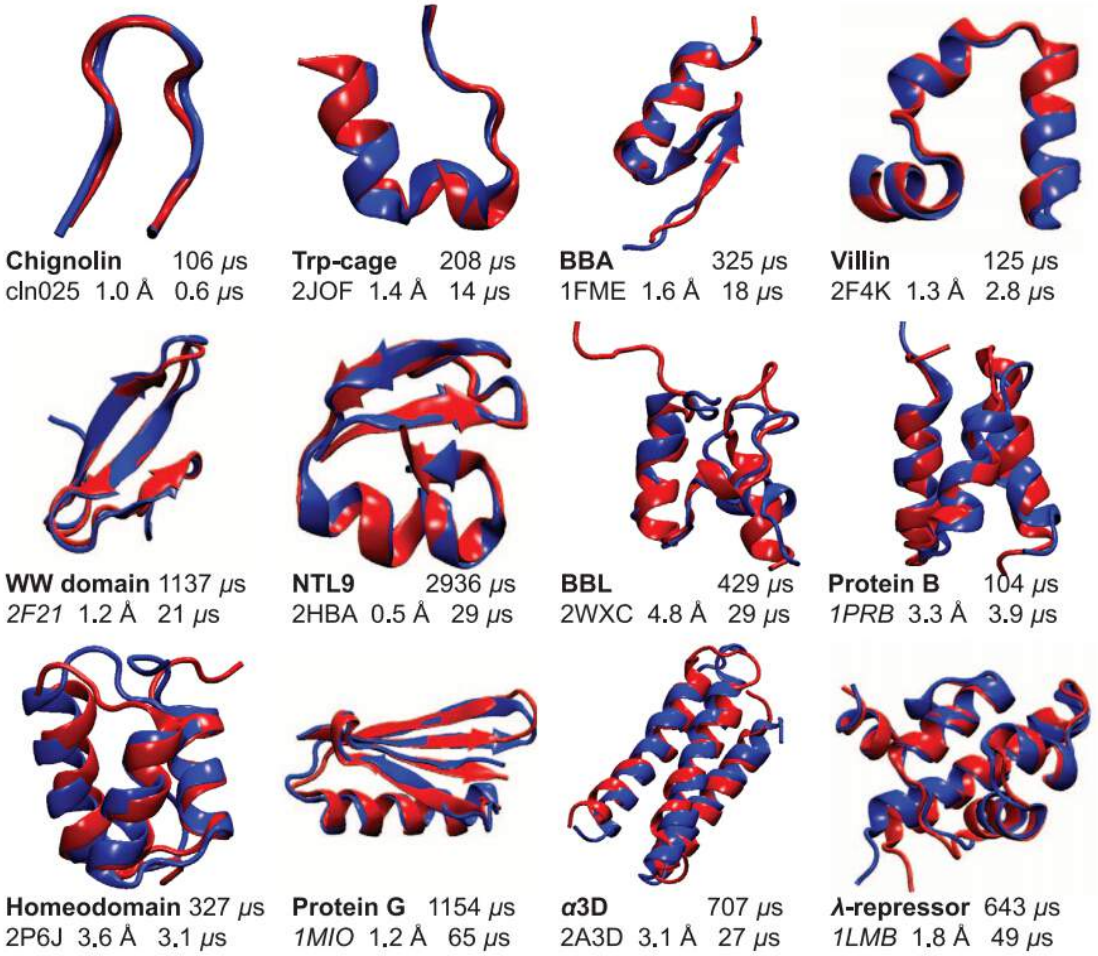


Allosteric regulation

http://www.wikipedia.org/

Introduction

Molecular dynamics simulations can describe the dynamics of molecules at atomic level

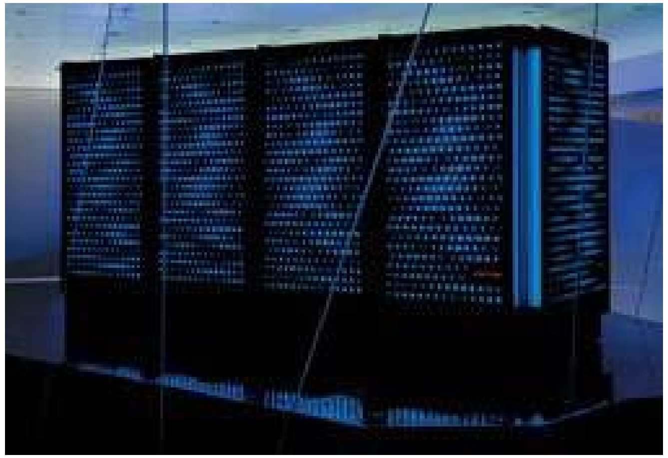


How Fast-Folding Proteins Fold

Kresten Lindorff-Larsen,[1*†] Stefano Piana,[1*†] Ron O. Dror,[1] David E. Shaw[1,2†]

Science AAAS

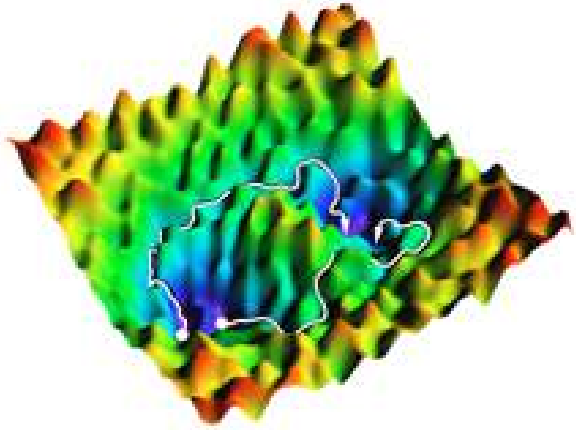
How Fast-Folding Proteins Fold
Kresten Lindorff-Larsen *et al.*
Science **334**, 517 (2011);
DOI: 10.1126/science.1208351



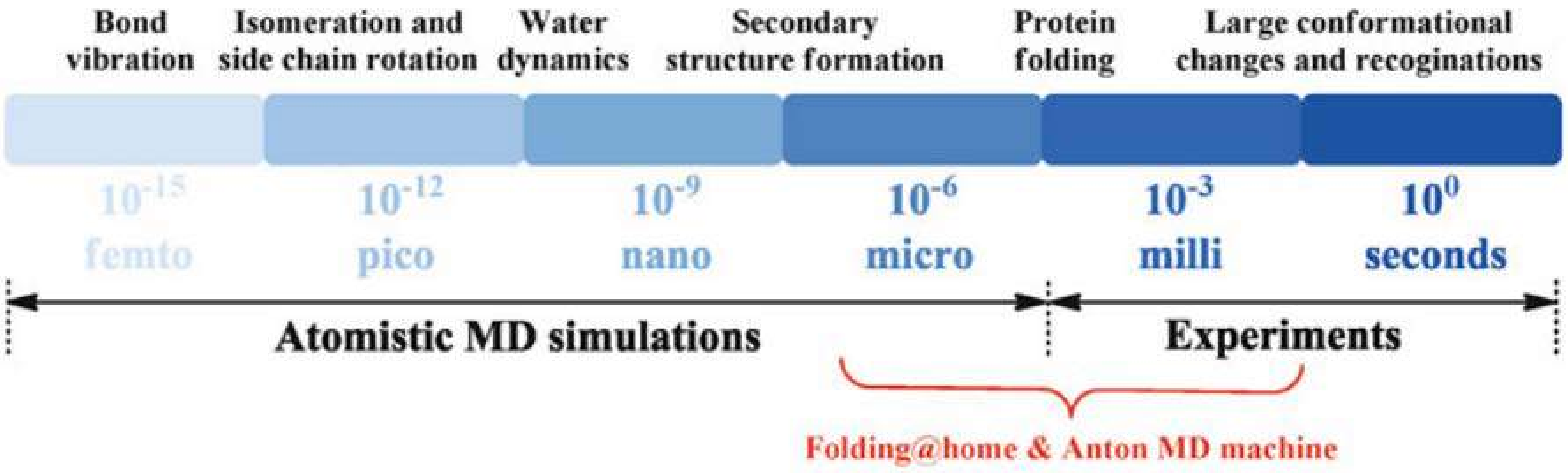
Anton MD mechine

Introduction

Problem1: Timescale (conventional MD methods) are limited



Rugged energy surface
in high - dimensional system



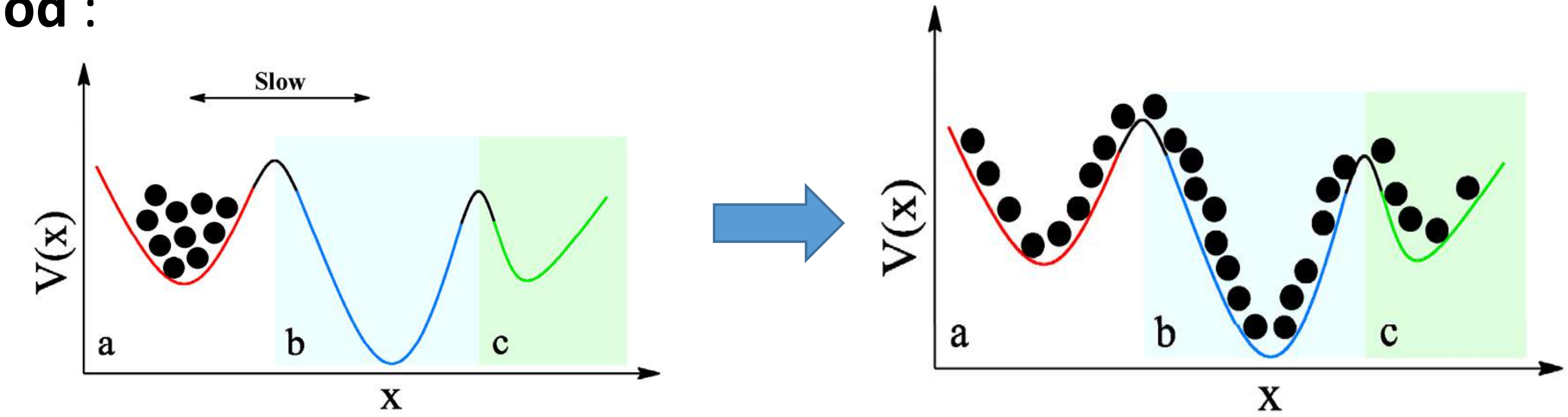
http://gold.cchem.berkeley.edu/The_Chandler_Group/Research.html

Introduction

Available tools to solve Problem1:

Enhanced sampling method :

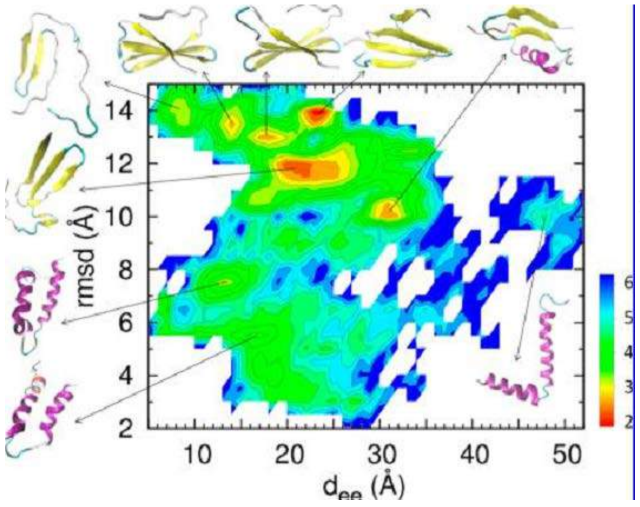
Replica exchange MD
Targeted MD
Steered MD
Metadynamics MD

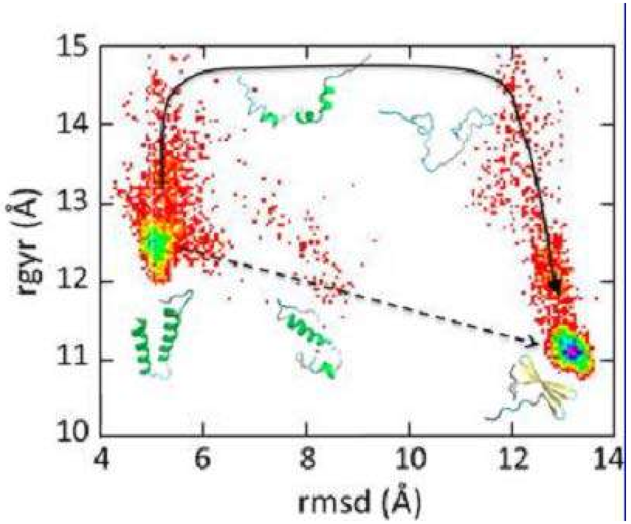


Reaction coordinates (RMSD, Q, Rg) ?

Markov State Model

Free energy surface





Robert DM et al,
Da LT, Sheong FK, Silva DA, Huang X, Adv Exp Med Biol, 2014, 805: 29 - 66.

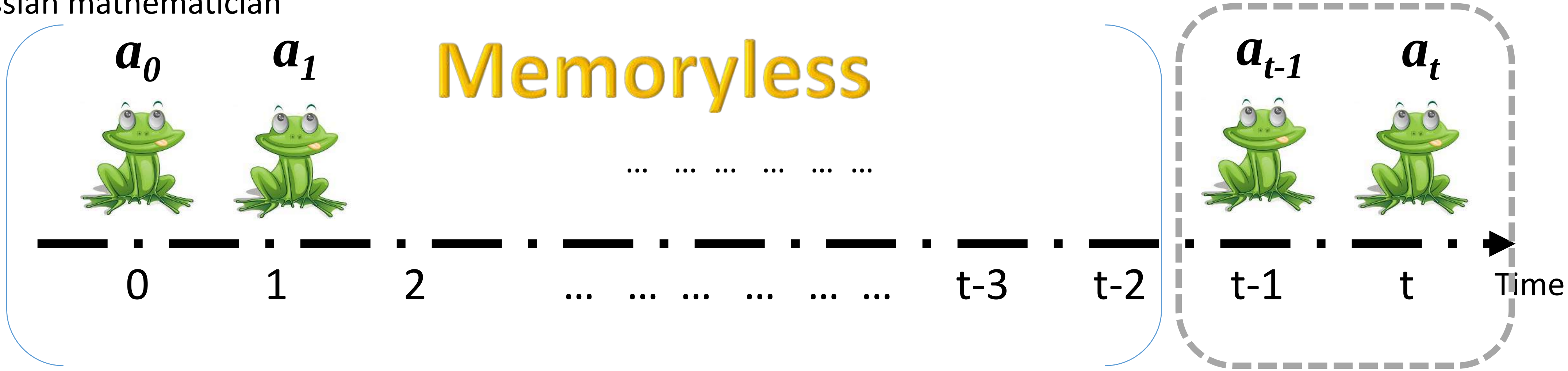
What is Markov State Model

Andrey Markov
Russian mathematician

Markov Chain

Definition: *A discrete time stochastic process* $X_0, X_1, X_2, \ldots$ *is a Markov chain if:*

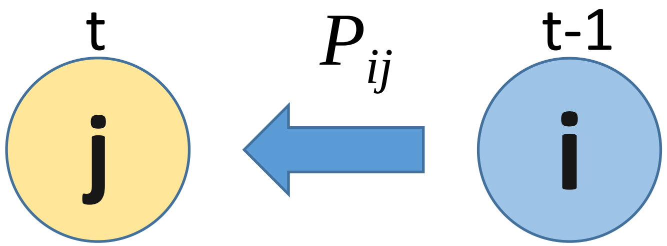
$$\mathbf{Pr}\,(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_0 = a_0) = \mathbf{Pr}(X_t = a_t \mid X_{t-1} = a_{t-1})$$



$\mathbf{Pr}(X_t = a_t \mid X_{t-1} = a_{t-1})$

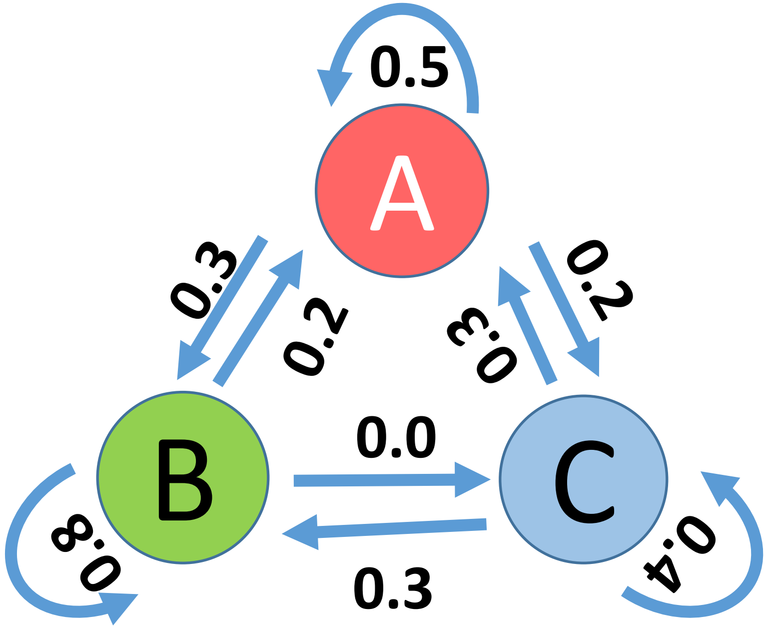
Michael Mitzenmacher, Probability and Computing, 2005, 153 – 156.

What is Markov State Model



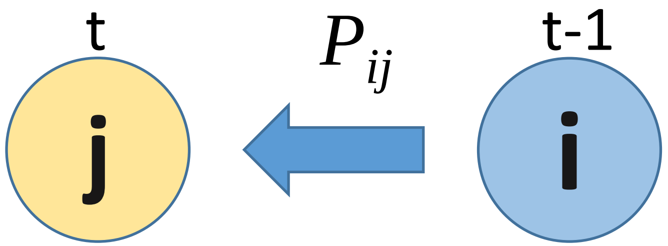
$$P_{ij} = Pr(X_t = j \mid X_{t-1} = i)$$

A simple example: 3 state model

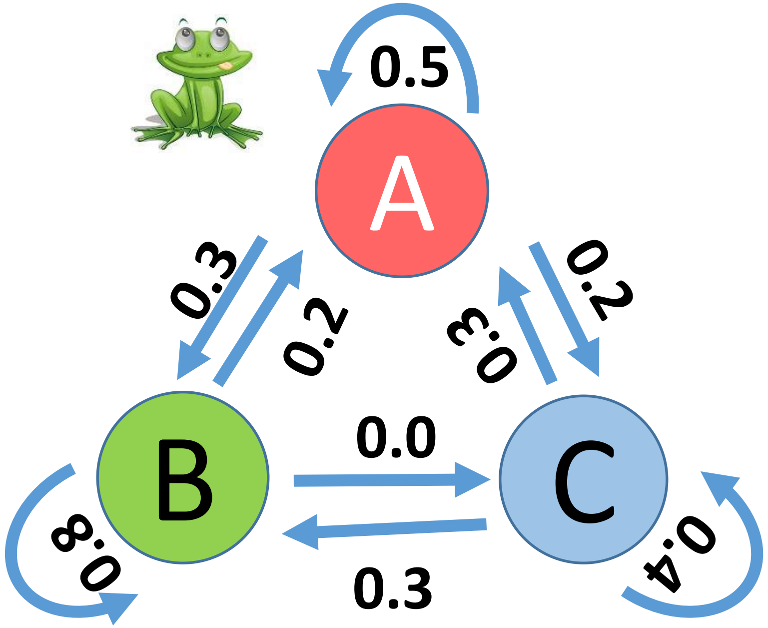


Time	P_A	P_B	P_C
0	1	0	0

What is Markov State Model



$$P_{ij} = Pr(X_t = j \mid X_{t-1} = i)$$



A simple example: 3 state model

Time	P_A	P_B	P_C
0	1	0	0
τ	0.5	0.3	0.2

P_A' p_{AA} P_B' p_{BA} P_C' p_{CA}

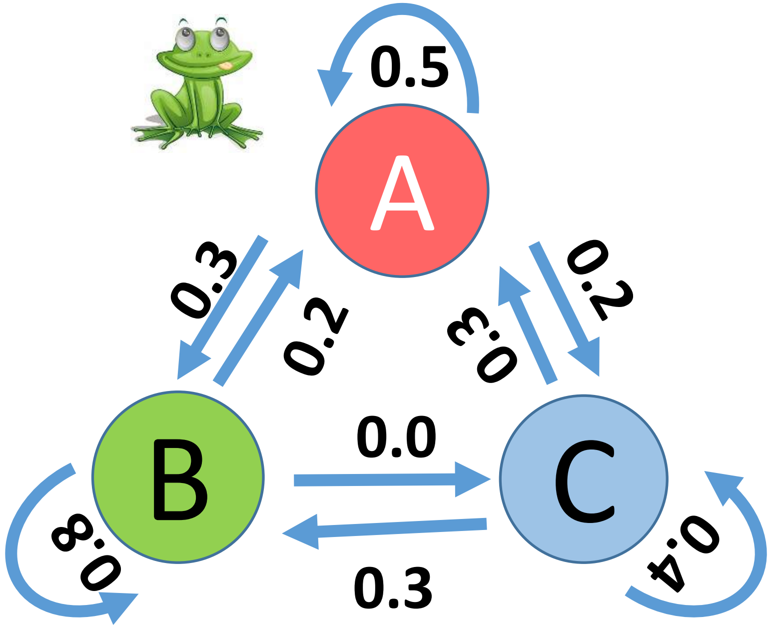
$P_A = 1 \times 0.5 + 0 \times 0.2 + 0 \times 0.3 = 0.5$

$P_B = 1 \times 0.3 + 0 \times 0.8 + 0 \times 0.3 = 0.3$

$P_C = 1 \times 0.2 + 0 \times 0.0 + 0 \times 0.4 = 0.2$

What is Markov State Model

$$P_{ij} = Pr(X_t = j \mid X_{t-1} = i)$$



A simple example: 3 state model

Time	P_A	P_B	P_C
0	1	0	0
τ	0.5	0.3	0.2
2τ	0.37	0.45	0.18

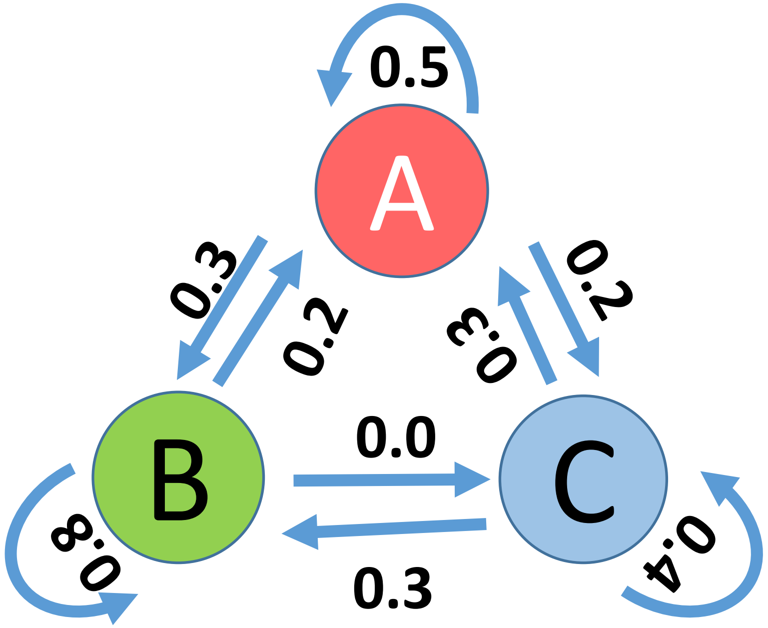
$$P_A = \underbrace{0.5}_{P_A'} \times \underbrace{0.5}_{p_{AA}} + \underbrace{0.3}_{P_B'} \times \underbrace{0.2}_{p_{BA}} + \underbrace{0.2}_{P_C'} \times \underbrace{0.3}_{p_{CA}} = 0.37$$

$$P_B = 0.5 \times 0.3 + 0.3 \times 0.8 + 0.2 \times 0.3 = 0.45$$

$$P_C = 0.5 \times 0.2 + 0.3 \times 0.0 + 0.2 \times 0.4 = 0.18$$

What is Markov State Model

A simple example



From:

$$T(\tau) = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.4 \end{bmatrix}$$

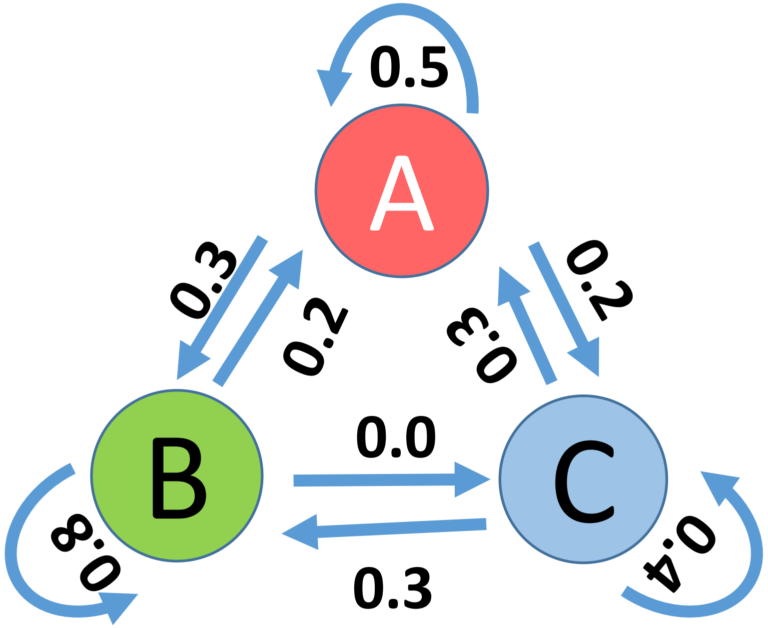
State: columns *A*, *B*, *C*; To: rows *A*, *B*, *C*

$$P(0) = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

$$P(1\tau) = T(\tau)P(0) = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.4 \end{bmatrix} \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}$$

What is Markov State Model

A simple example



From:

$$T(\tau) = \begin{array}{c} \textit{State:} \end{array} \begin{array}{ccc} A & B & C \end{array} \quad \textbf{To:}$$

$$T(\tau) = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.4 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix} \qquad P(0) = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

$$P(2\tau) = T(\tau)P(1\tau) = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.4 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.37 \\ 0.45 \\ 0.18 \end{bmatrix}$$

What is Markov State Model

A simple example

$$P(13\tau) = \begin{bmatrix} 0.30002 \\ 0.59993 \\ 0.10005 \end{bmatrix} \quad P(14\tau) = \begin{bmatrix} 0.30001 \\ 0.59996 \\ 0.10002 \end{bmatrix} \quad P(15\tau) = \begin{bmatrix} 0.30001 \\ 0.59998 \\ 0.10001 \end{bmatrix}$$

$$P(n\tau) = T(\tau)P((n\text{-}1)\tau)$$

$$P(n\tau) = [T(\tau)]^n P(0)$$

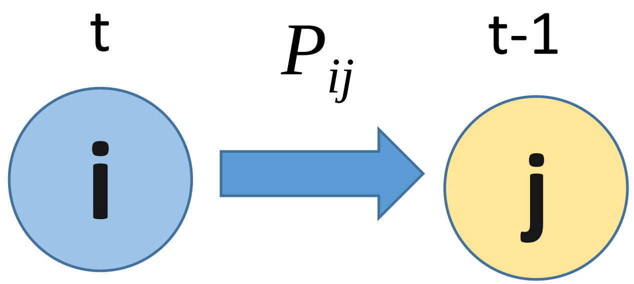
What is Markov State Model

Markov Chain

Definition: *A discrete time stochastic process*

$X_0, X_1, X_2, \ldots$ *is a Markov chain if:*

$$\mathbf{Pr}\,(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_0 = a_0) = \mathbf{Pr}(X_t = a_t \mid X_{t-1} = a_{t-1})$$



$$P_{ij} = Pr(X_t = j \mid X_{t-1} = i)$$

One-step transition matrix

$$T(\tau) = \begin{pmatrix} P_{0,0} & P_{0,2} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & & \vdots & \\ P_{i,0} & P_{t,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & & \vdots & \end{pmatrix}$$

Michael Mitzenmacher, Probability and Computing, 2005, 153 – 156.

Application of MSM to MD simulation

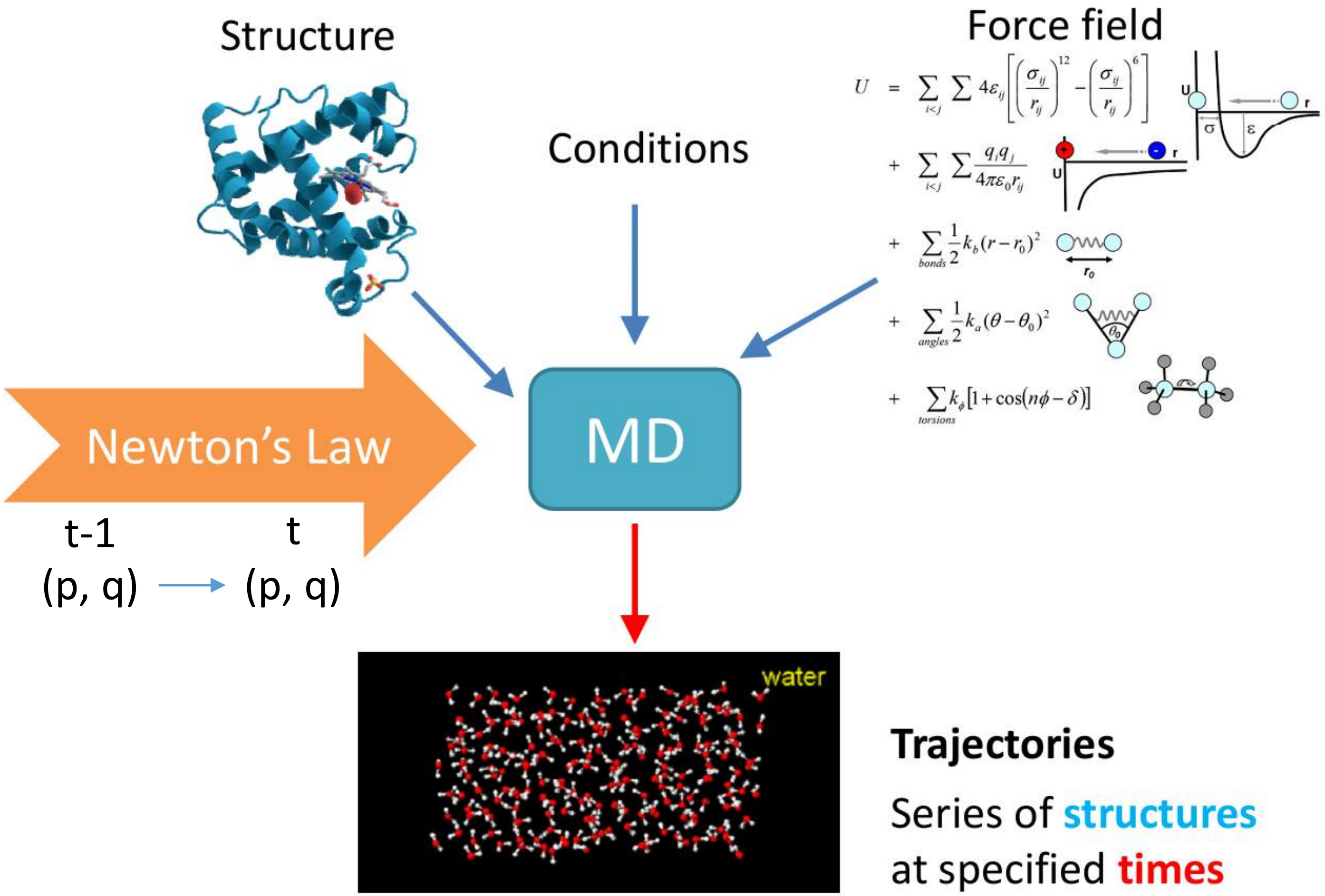
Markov State Model

*Transition modes between **different states***

*A **discrete** time process*

***Independent** of the history before the last step*

MD simulation



Trajectories

Series of structures at specified times

Application of MSM to MD

How can MSM be applied to MD?

Prof. Vijay S pande

Folding@home

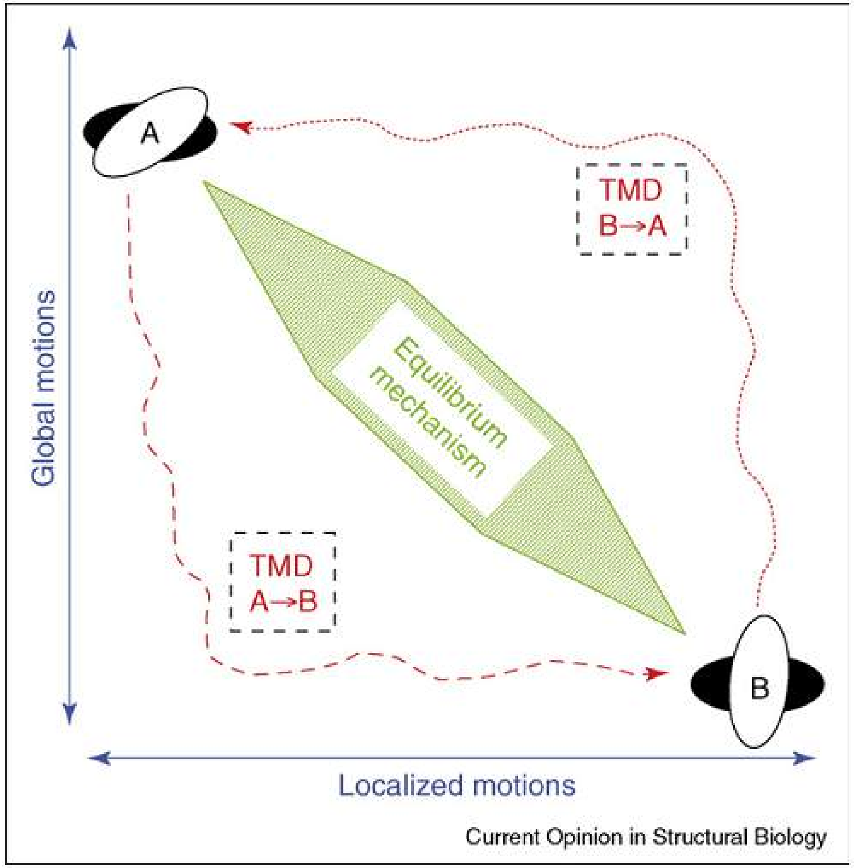
MSM builder

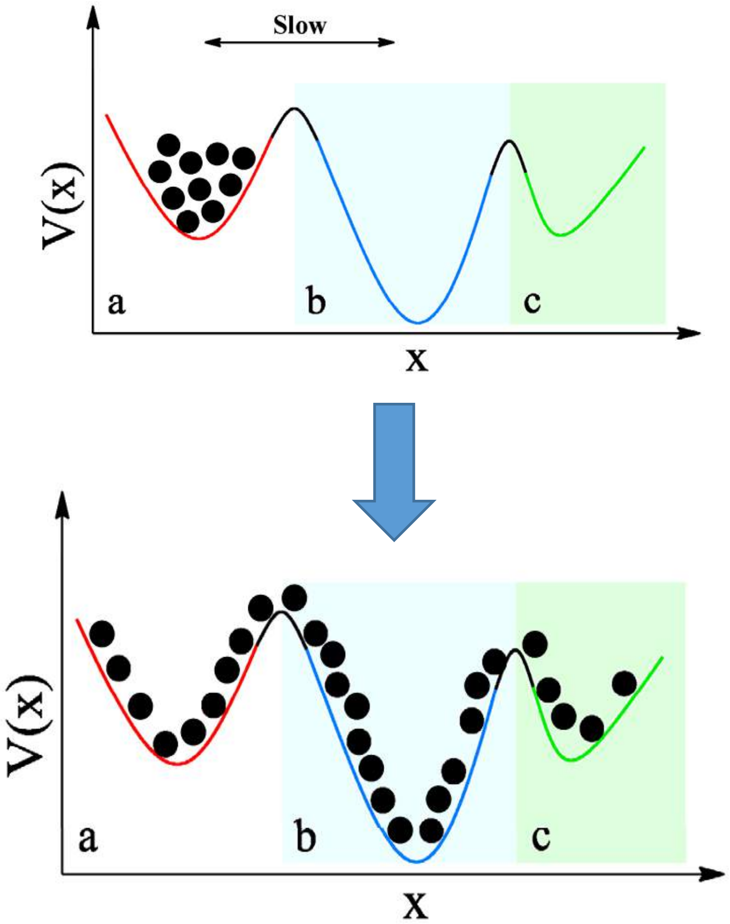
http://pande.stanford.edu/

Application of MSM to MD

How can MSM be applied to MD?

1. Enhanced sampling





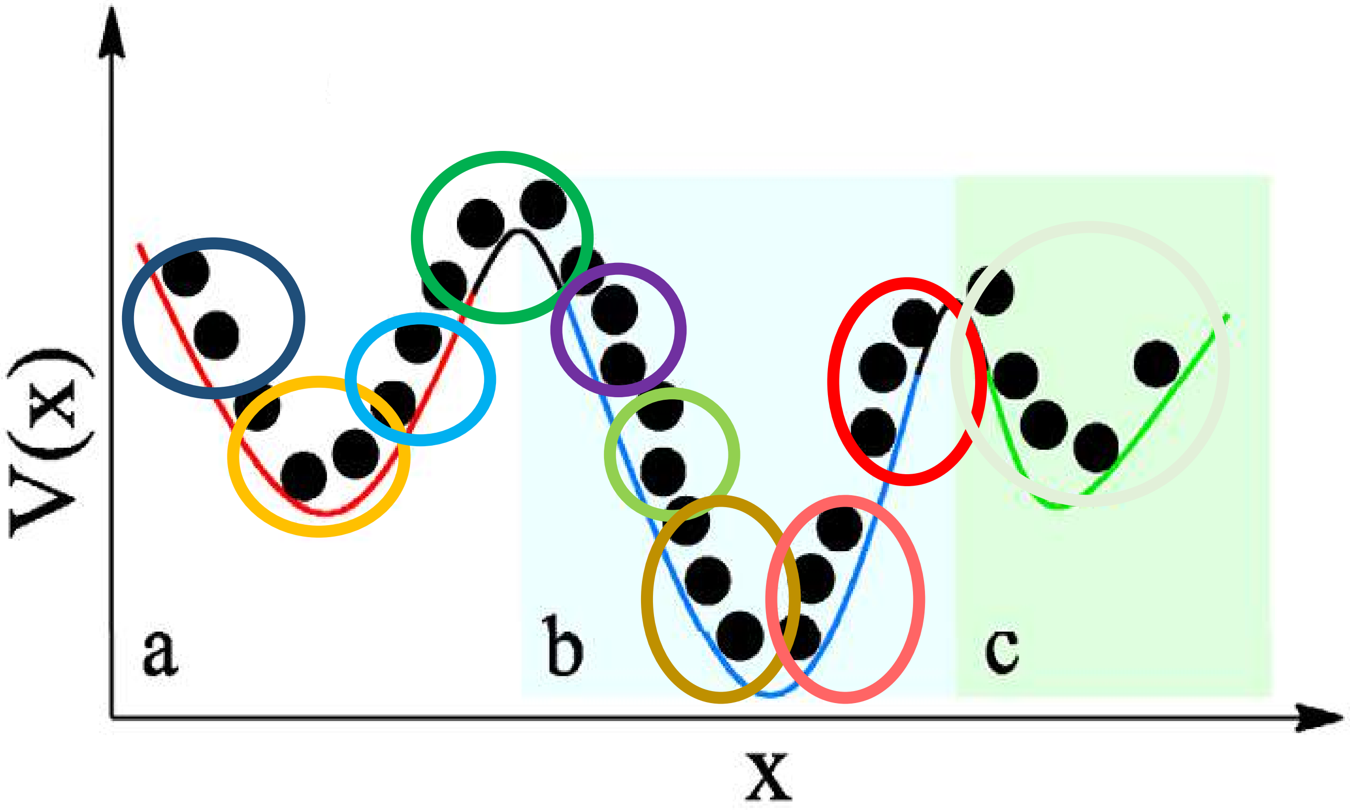
Eg. Targeted MD

Noé F, Fischer S, Curr Opin Struct Biol, 2008, 18(2):154-162.

Application of MSM to MD

How can MSM be applied to MD?

2. Geometric clustering for seeding

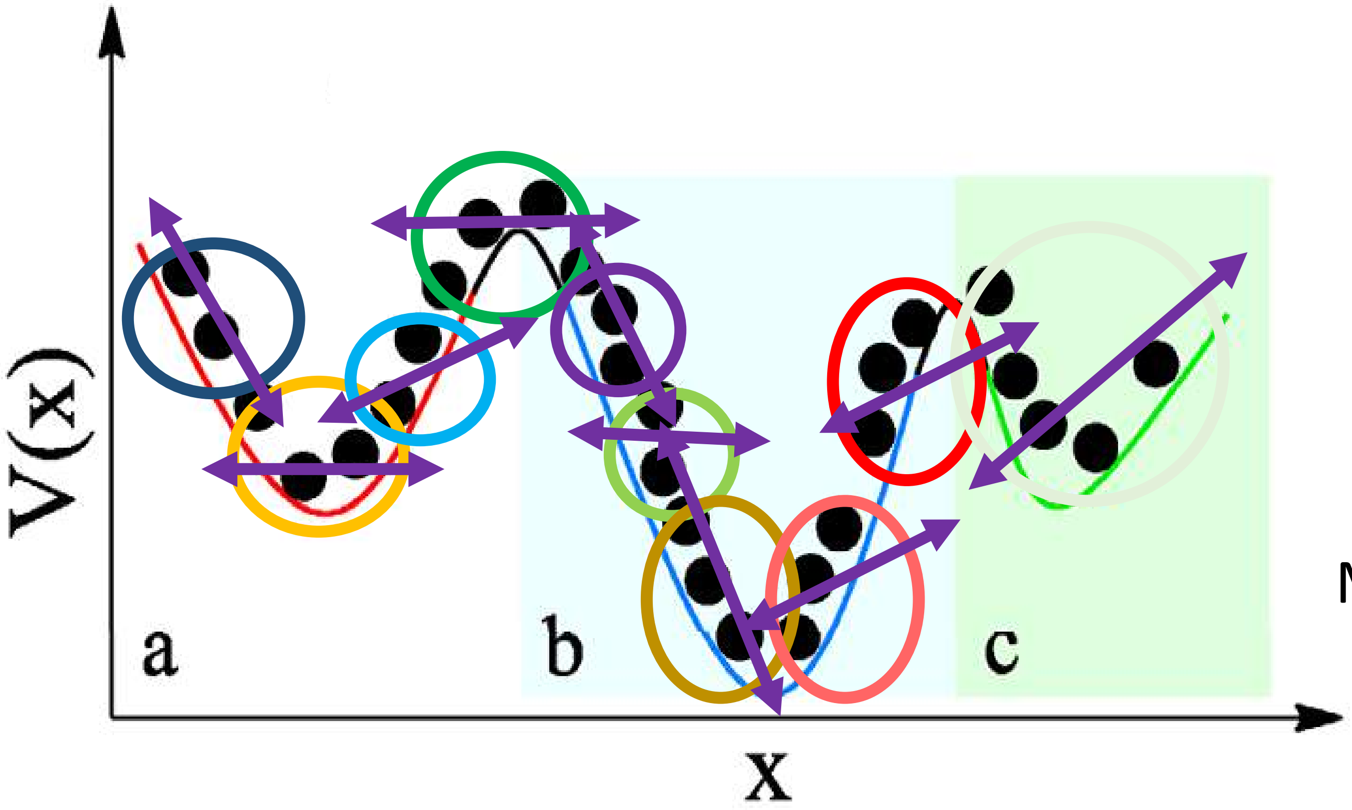


K-center, K-means

Application of MSM to MD

How can MSM be applied to MD?

3. Unbiased short MD simulations initiated from the seeds clusterd



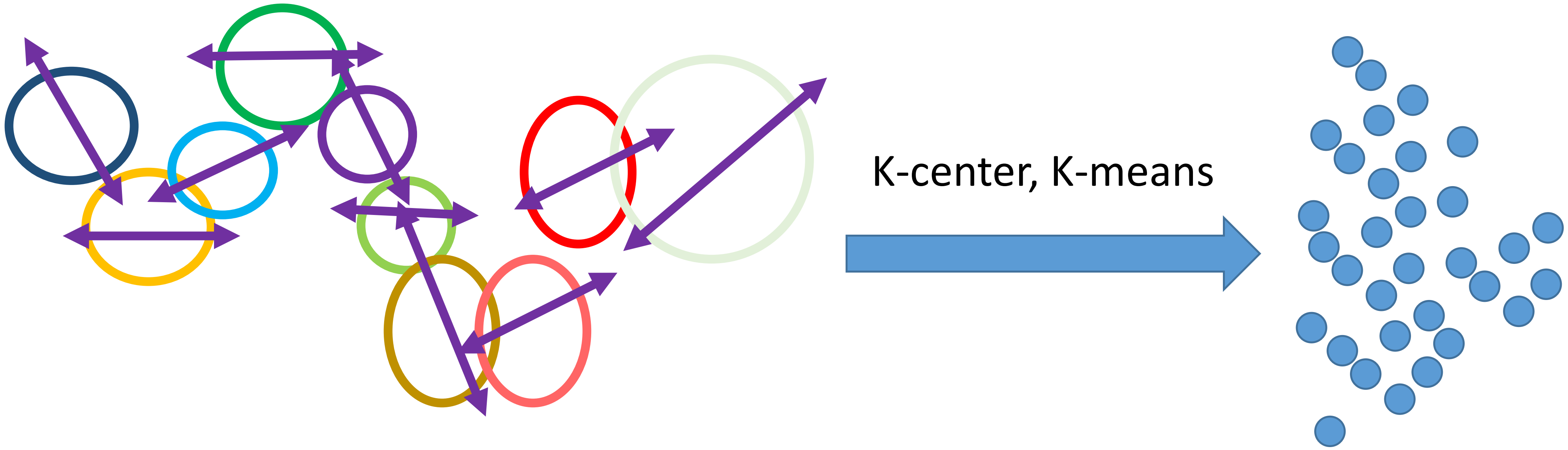
M clusters

M trajectories

Application of MSM to MD

How can MSM be applied to MD?

4. Geometric clustering for microstates



N trajectories **N microstates**

Application of MSM to MD

How can MSM be applied to MD?

5. Building the transition count matrix (Transition probability matrix)

1 1 2 2 2 1 2 1 2 1 2

$N_{11}=1$ $N_{12}=4$ $N_{21}=3$ $N_{22}=2$

From/to	State 1	State 2
State 1	1	4
State 2	3	2

Fulfilling Detailed Balance

$$N^{symm} = \frac{N + N^T}{2}$$

From/to	State 1	State 2
State 1	1	3.5
State 2	3.5	2

$$P_{ij} = \frac{N_{ij}^{symm}}{\sum(N_{ij}^{symm})}$$

From/to	State 1	State 2
State 1	0.222	0.778
State 2	0.636	0.364

Da LT, Sheong FK, Silva DA, Huang X, Adv Exp Med Biol, 2014, 805: 29 - 66.

Application of MSM to MD

How can MSM be applied to MD?

5. Building the transition probability matrix

$$T_{ij}(\tau) = \frac{N_{ij}(\tau)}{\sum_j N_{ij}(\tau)} \longrightarrow \mathbf{T}(\tau) = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & & \\ \vdots & & \ddots & \\ P_{n1} & & & P_{nn} \end{bmatrix}$$

Transition Probability Matrix

P(nτ) is a vector of microstate populations at time nτ

$$P(n\tau) = [T(\tau)]^n P(0)$$

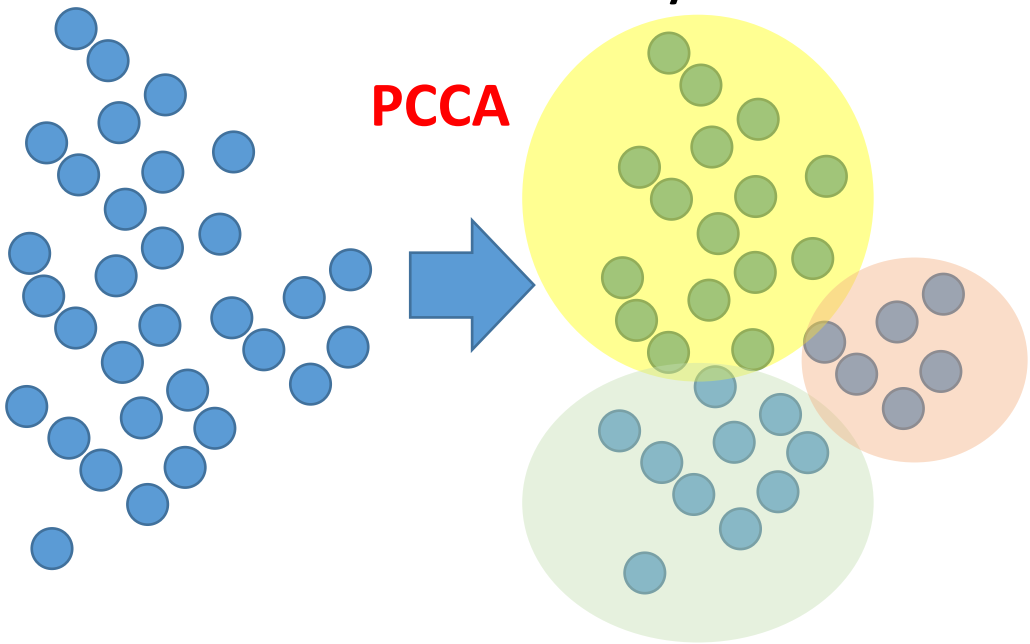
Da LT, Sheong FK, Silva DA, Huang X, Adv Exp Med Biol, 2014, 805: 29 - 66.

Application of MSM to MD

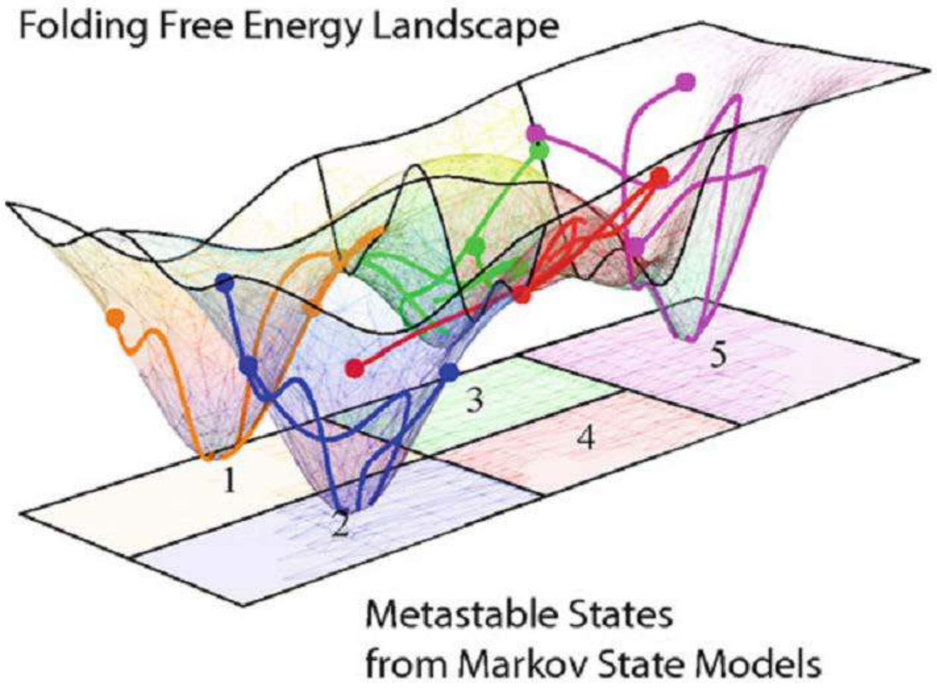
How can MSM be applied to MD?

6. Lumping microstates into microstates.

Perron cluster cluster analysis



microstates **macrostates**

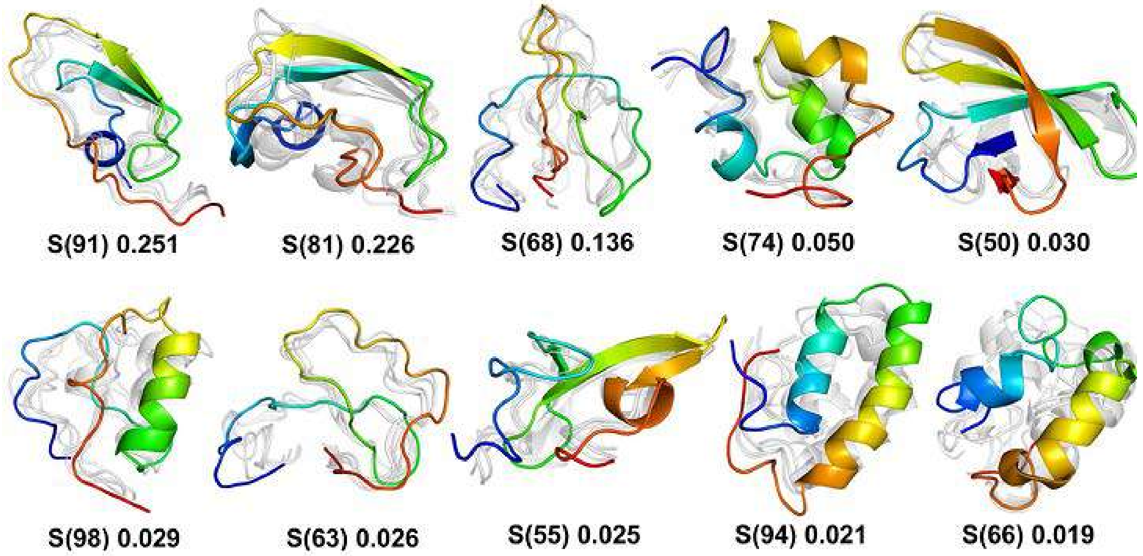


Da LT, Sheong FK, Silva DA, Huang X, Adv Exp Med Biol, 2014, 805: 29 - 66.

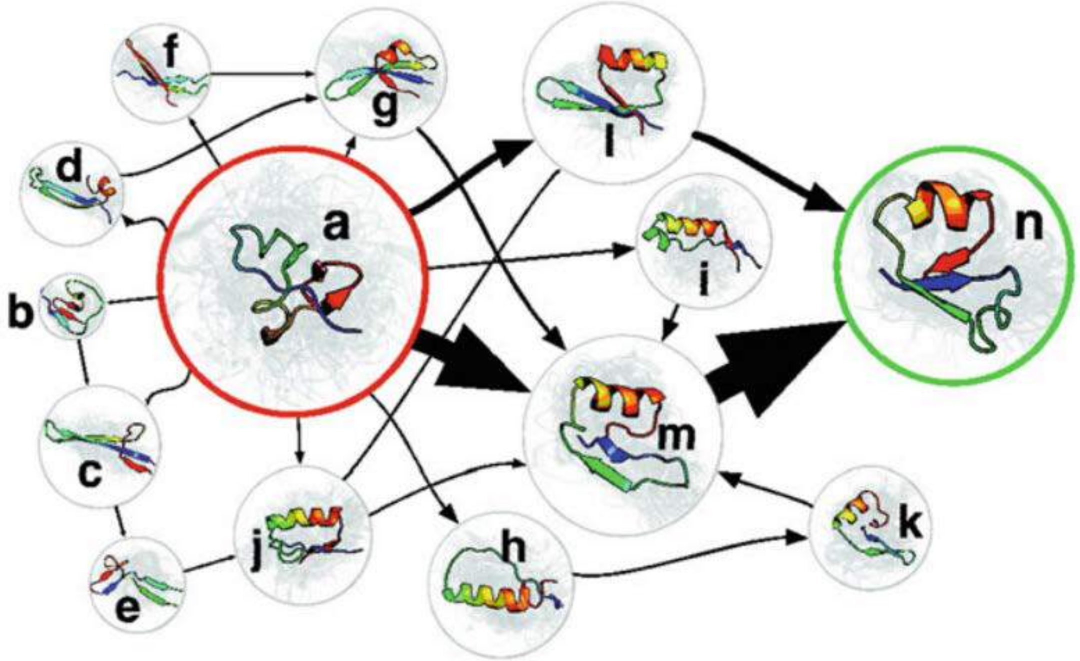
Application of MSM to MD

How can MSM be applied to MD?

7. Analysis of markov state model



The 10 most populated macrostates from the coarse-grained MSM with their equilibrium populations



The top 15 folding fluxes

Transition pathway theory

Da LT, Sheong FK, Silva DA, Huang X, Adv Exp Med Biol, 2014, 805: 29 - 66.
Li SS et al, J. Chem. Theory Comput, 2014, 10(6): 2255–2264

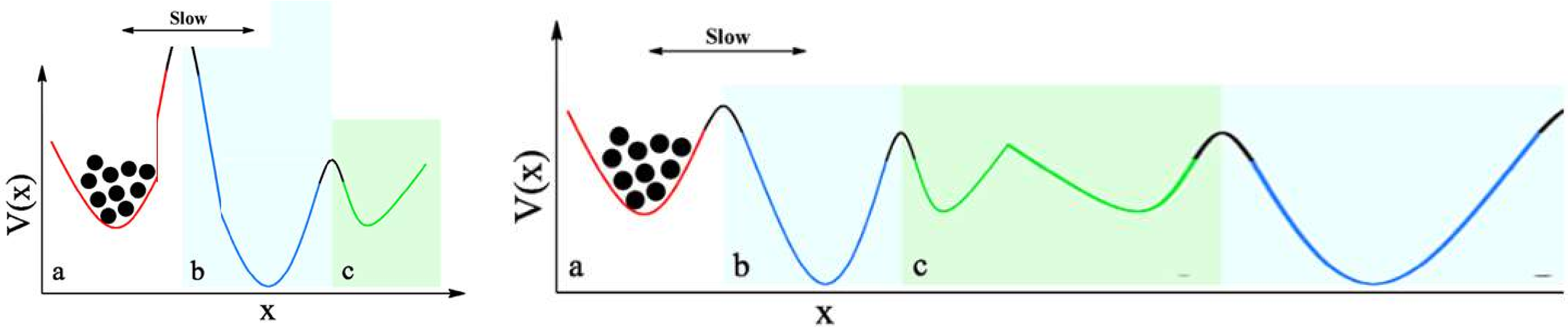
Advantages of MSM

- Enhanced sampling +
  multi-parallel short conventional MD simulations

- No specified reaction coordinates (clustering)
- Thermodynamic information
  Equilibrium populations
- Kinetic information
  Dominant pathways

Cautions of MSM

- Sufficiency sampling is still challenging



- Accurate force fields

- Connections to experiments

Pande VS et al, Methods, 2010, 52(1):99-105.

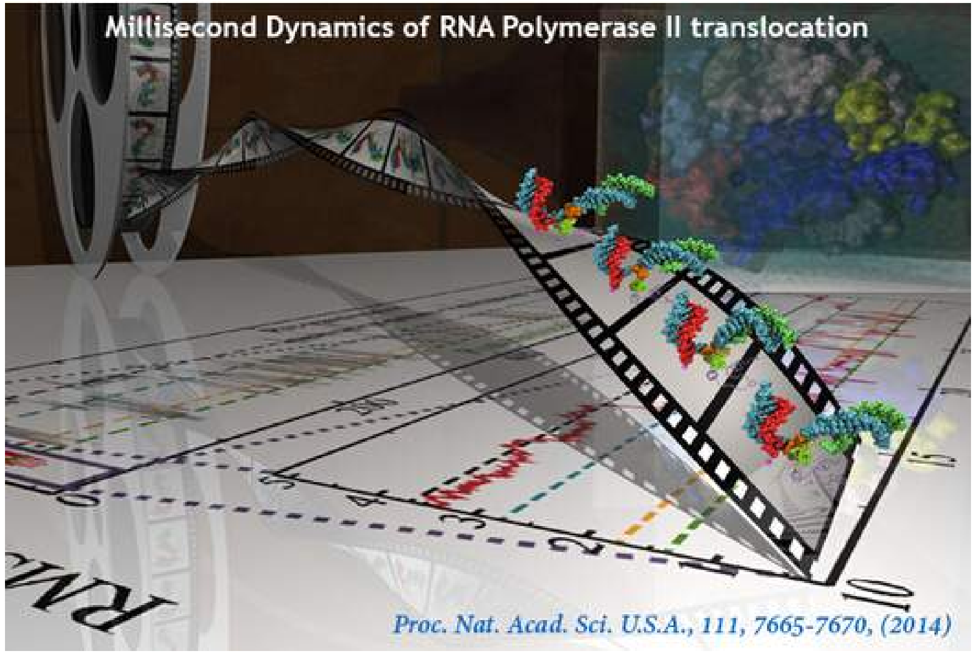
Popularity of MSM

Key word = ‘markov state model’ + ‘molecular dynamics simulation’

Popularity of MSM

Recent notable work

Prof. Huang Xuhui



Dr. Da Lintai

- **July 2014**
  Congs on Lu on our collaborative project with Prof. Dong Wang's lab at UCSD on Asymmetric Recognition of Phosphodiester Linkage by RNA Polymerase II, which has just been accepted by PNAS! See ***PNAS*, 111, E3269, (2014)**
- **June 2014**
  Congs on Shuo on his paper on developing a new method to qualify the protein-ligand recognition mechanisms which has just been accepted by PLoS Comp. Bio.! See ***PLoS. Comp. Bio.*, 10(8):e1003767, (2014)**
- **June 2014**
  Shuo's work on virtual screening of protein inhibitors, a project led by Prof. Nancy Ip's group, has just been published on ***PNAS*, 111, 9959, (2014)**! Congs! **See Press release here**
- **June 2014**
  Congs on Lu on her paper on photosynthesis which is just published on Nature Communications! ***Nature Communications*, 5:4170, (2014)**. See **Media report here**

Summary

Markov state model is a mathematical model applied to molecular dynamic simulations

- Large timescale equilibrium simulations
- Clustering (Higher dimensional data)
- Thermodynamic information (equilibrium populations)
- Kinetic information (dominant pathways)
- Increasing popularity

Thank you !