Transition path sampling and analysis of complex activated (bio)molecular processes

ICTP-SISSA-CECAM Workshop on Molecular Dynamics and its Applications to Biological Systems

Sept 13-17 2021

Peter Bolhuis
van 't Hoff institute for Molecular Sciences
University of Amsterdam, The Netherlands

ACMM Amsterdam Center for Multiscale Modeling

Outline

- Introduction
- Rare events

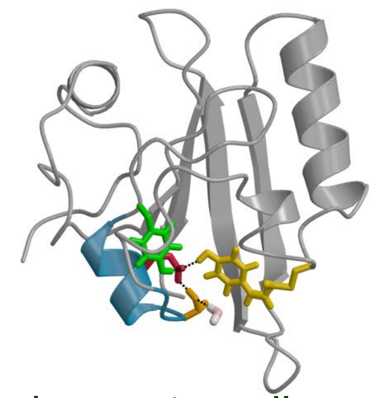
part 1:

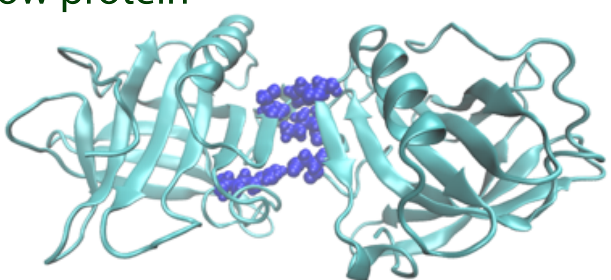
- Transition Path Sampling
- Committor & Reaction coordinate analysis
- Rate constants with transition interface sampling
- reaction networks with multiple state TPS/TIS
- advanced developments & machine learning
- OPS software

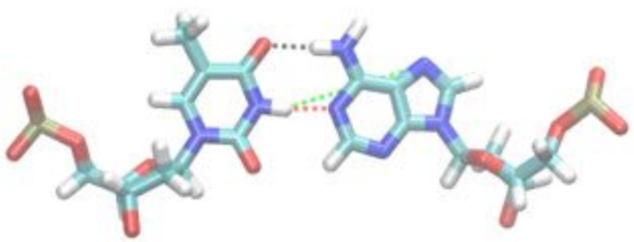
part 2:

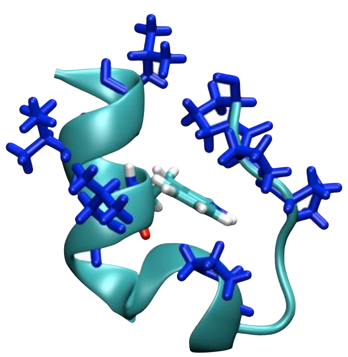
- imposing kinetic constraints
- path reweighting with Maximum Caliber

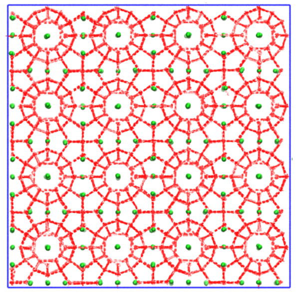
- conclusions


photoactive yellow protein


protein dissociation


DNA base pair rotation


Trp cage folding


gas hydrate formation

Molecular Dynamics

Aim: predicting complex molecular processes difficult to access in experiments

$$m\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$$

bonded interactions

$$V(\mathbf{r}) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} v_n (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i<j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

non-bonded interaction

© J. Vreede

Molecular Dynamics

Aim: predicting complex molecular processes difficult to access in experiments

$$m\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$$

bonded interactions

$$V(\mathbf{r}) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} v_n (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i<j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

non-bonded interaction

© J. Vreede

Molecular Dynamics

Aim: predicting complex molecular processes difficult to access in experiments

$$m\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$$

bonded interactions

$$V(\mathbf{r}) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} v_n (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i<j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

non-bonded interaction

© J. Vreede

Classical MD is able to yield at atomistic resolution

- **equilibrium statistics:** free energy landscapes, stable structures, transition states, …
- **kinetics:** rates, mechanisms, transport properties, …

Classical MD has two important sources of error:

- **the sampling problem (part 1)**
- the systematic force field error (part 2)

Molecular Dynamics

Aim: predicting complex molecular processes difficult to access in experiments

$$m\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$$

bonded interactions

$$V(\mathbf{r}) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} v_n (1 + \cos(n\phi - \phi_0))$$
$$+ \sum_{i<j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

non-bonded interaction

© J. Vreede

Classical MD is able to yield at atomistic resolution

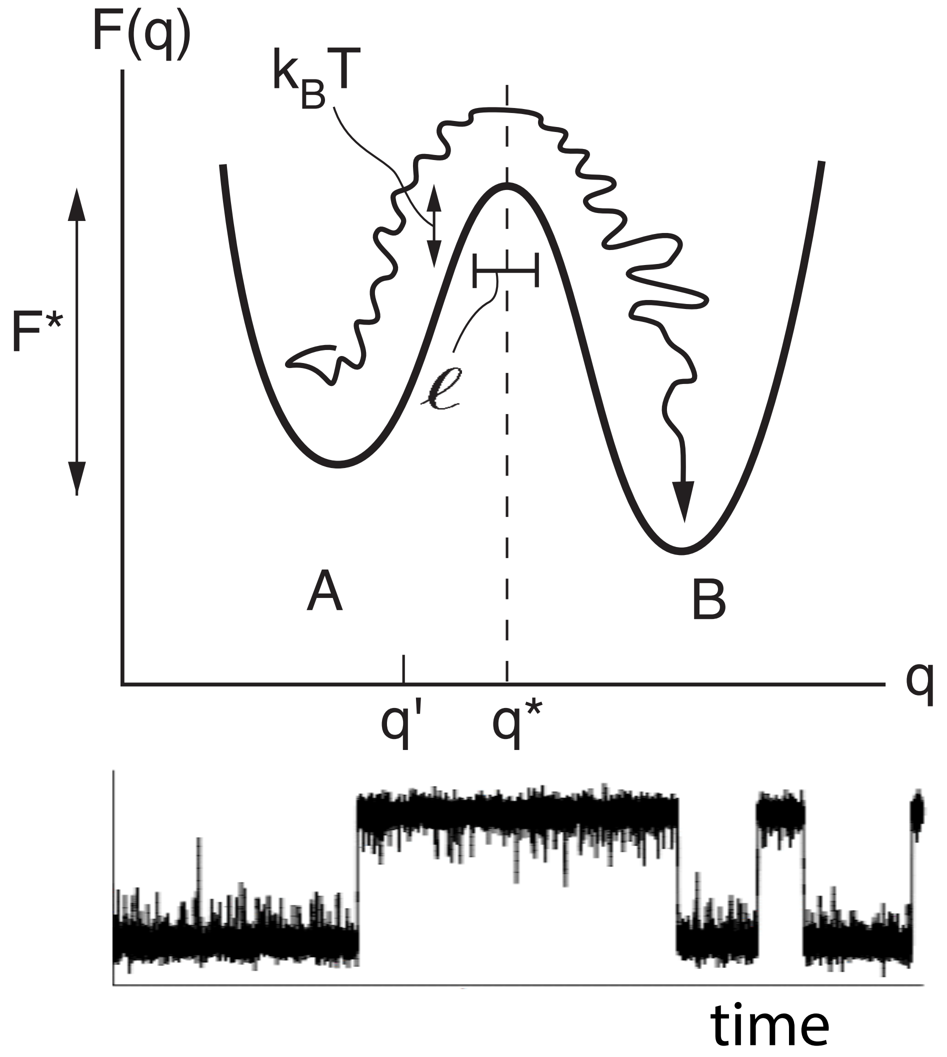
- **equilibrium statistics:** free energy landscapes, stable structures, transition states, …
- **kinetics:** rates, mechanisms, transport properties, …

Classical MD has two important sources of error:

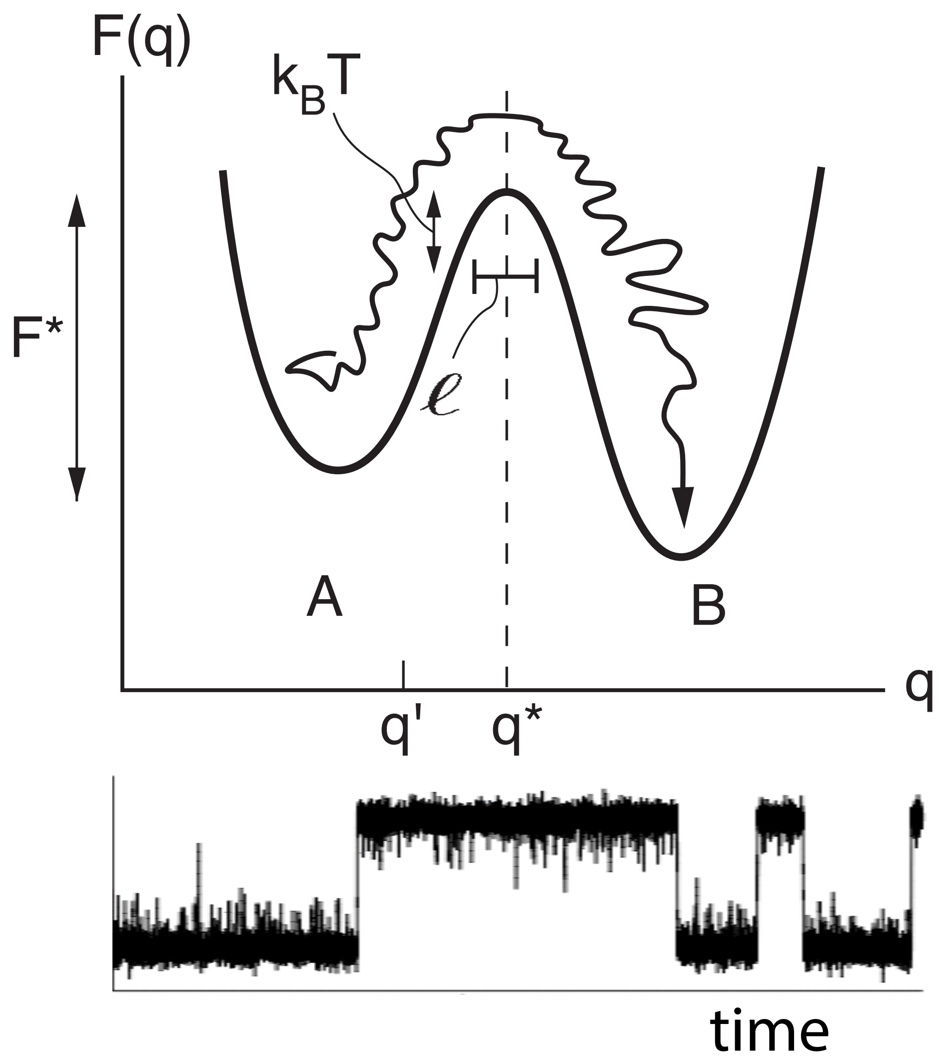
- **the sampling problem (part 1)**
- the systematic force field error (part 2)

current MD limited to sub-millisecond, most activated events much longer

Rare event sampling

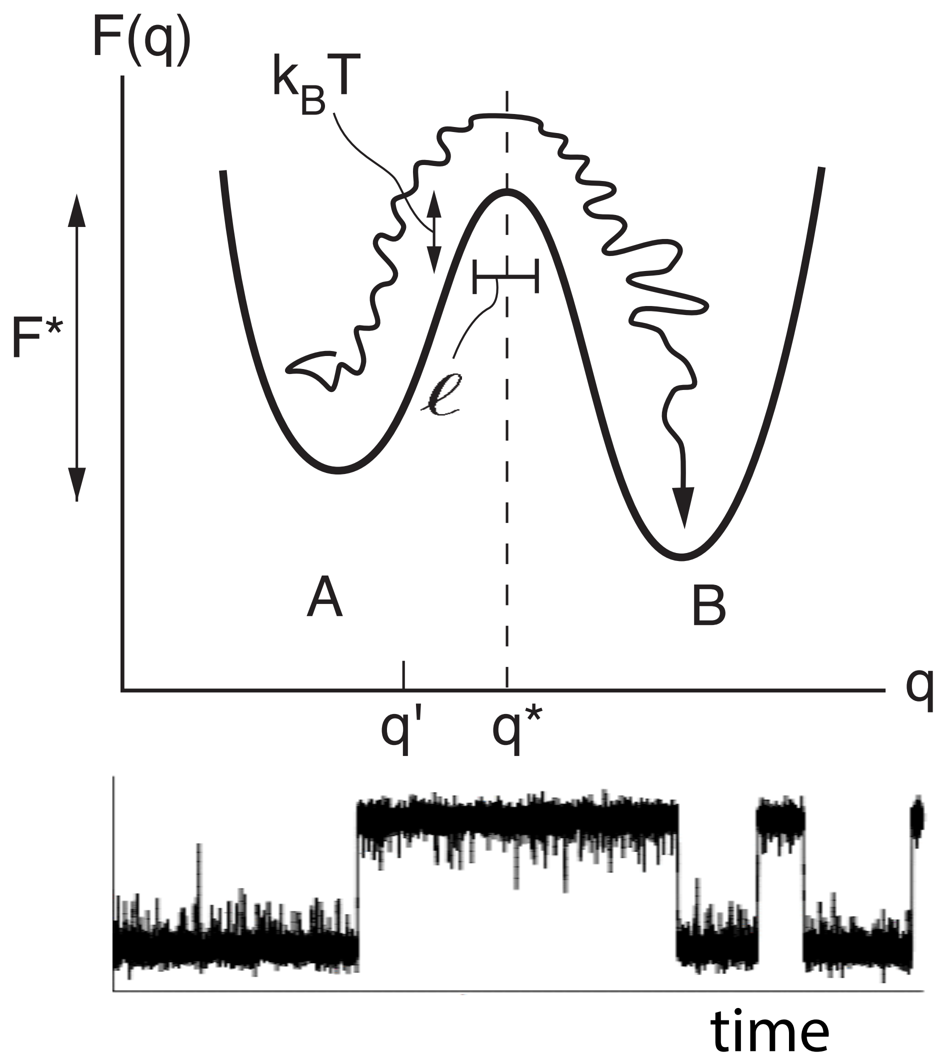


Rare event sampling

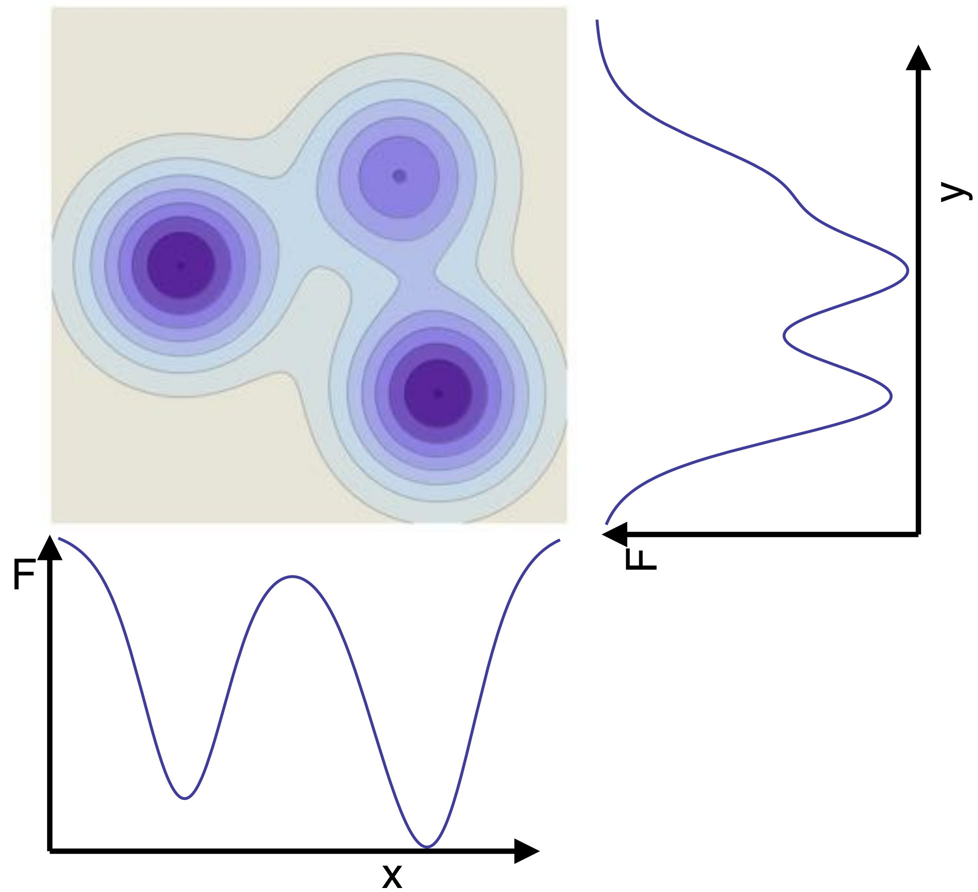


- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics

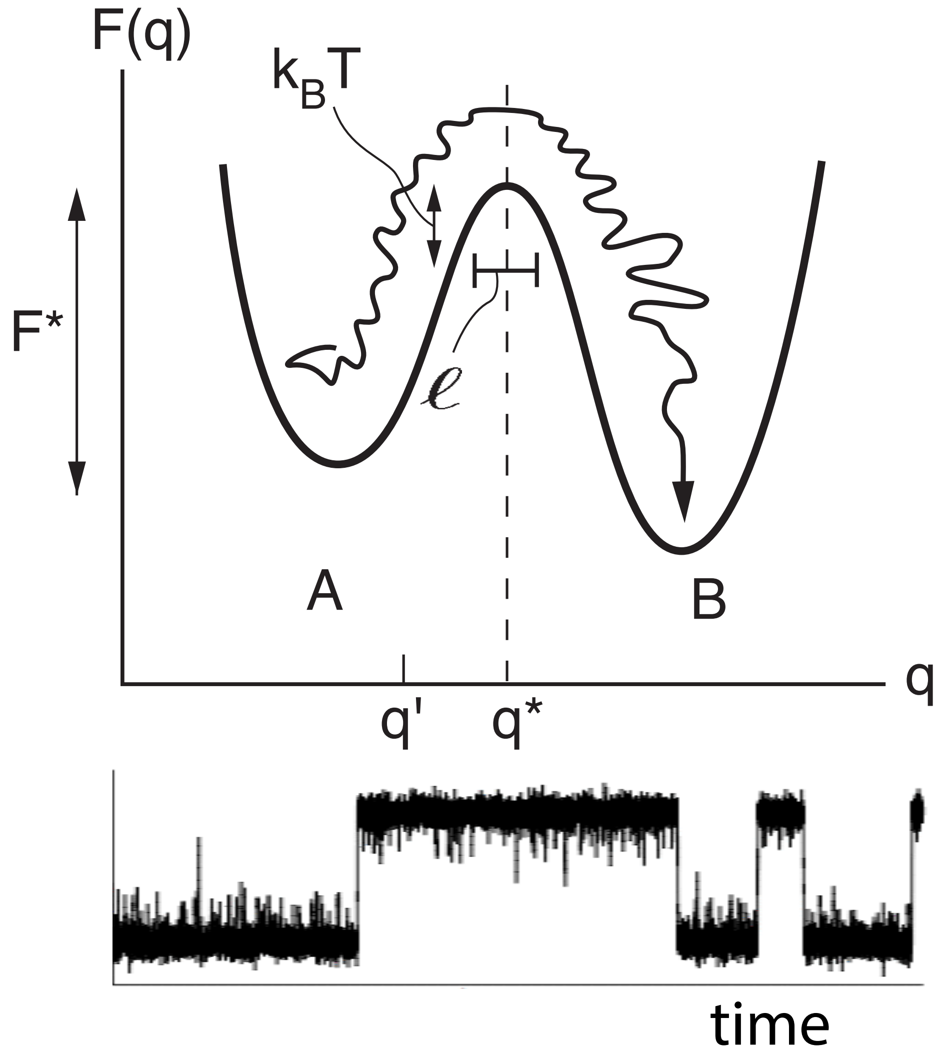
Rare event sampling



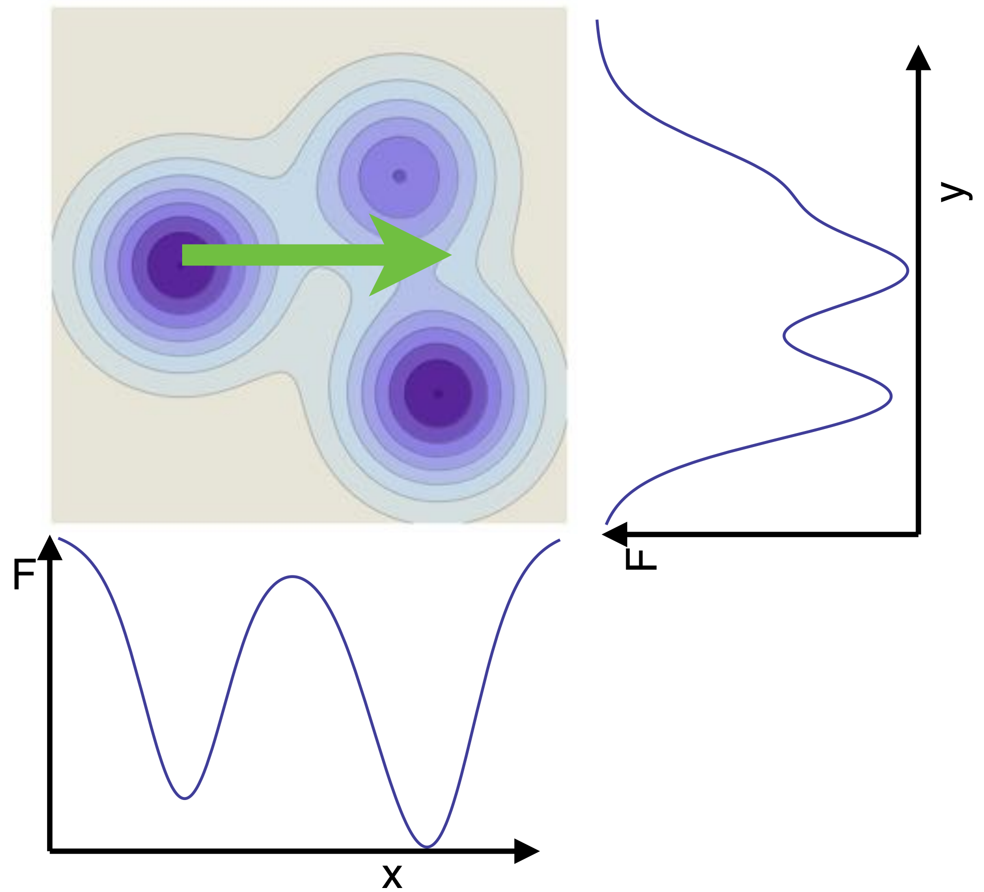
- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics



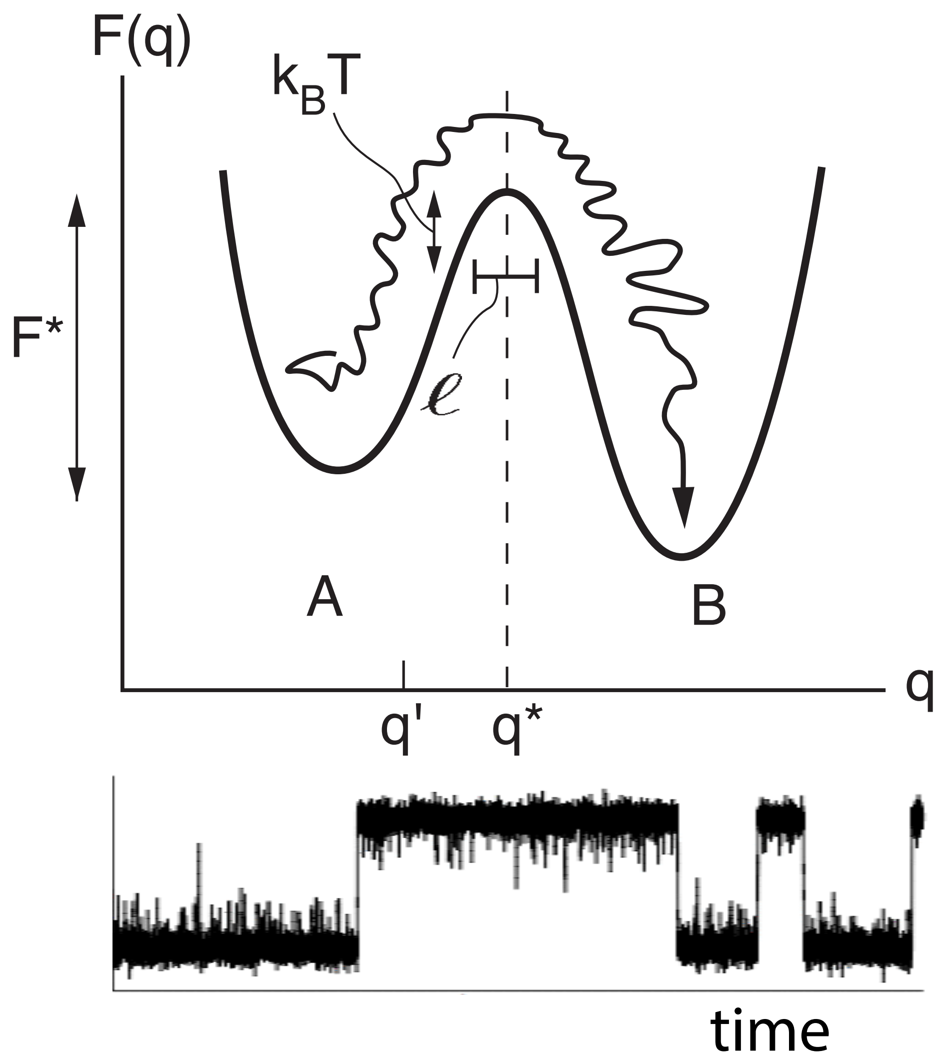
Rare event sampling



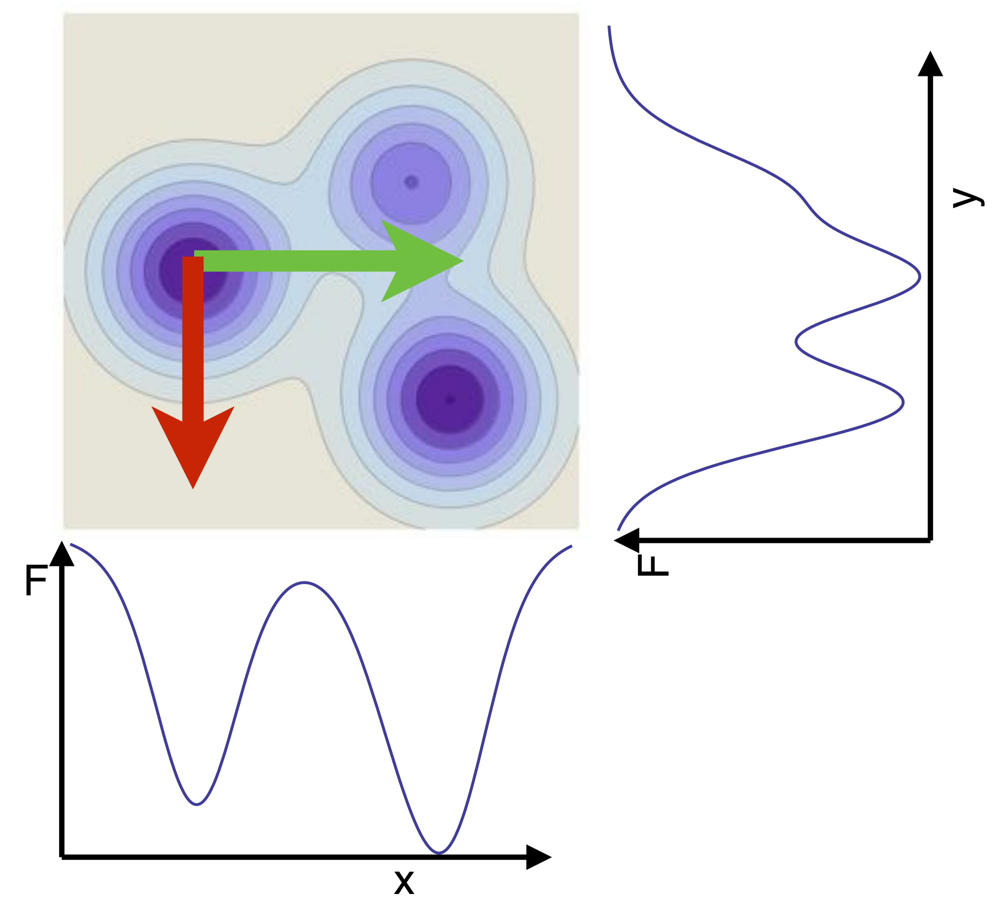
- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics



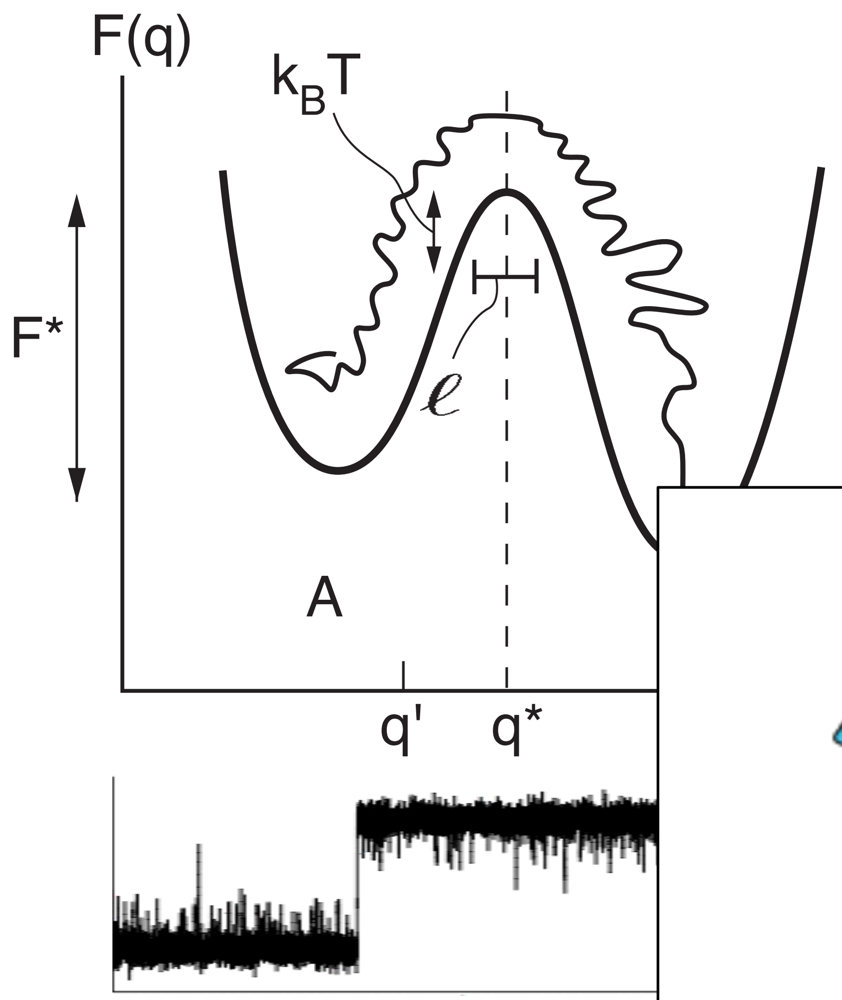
Rare event sampling

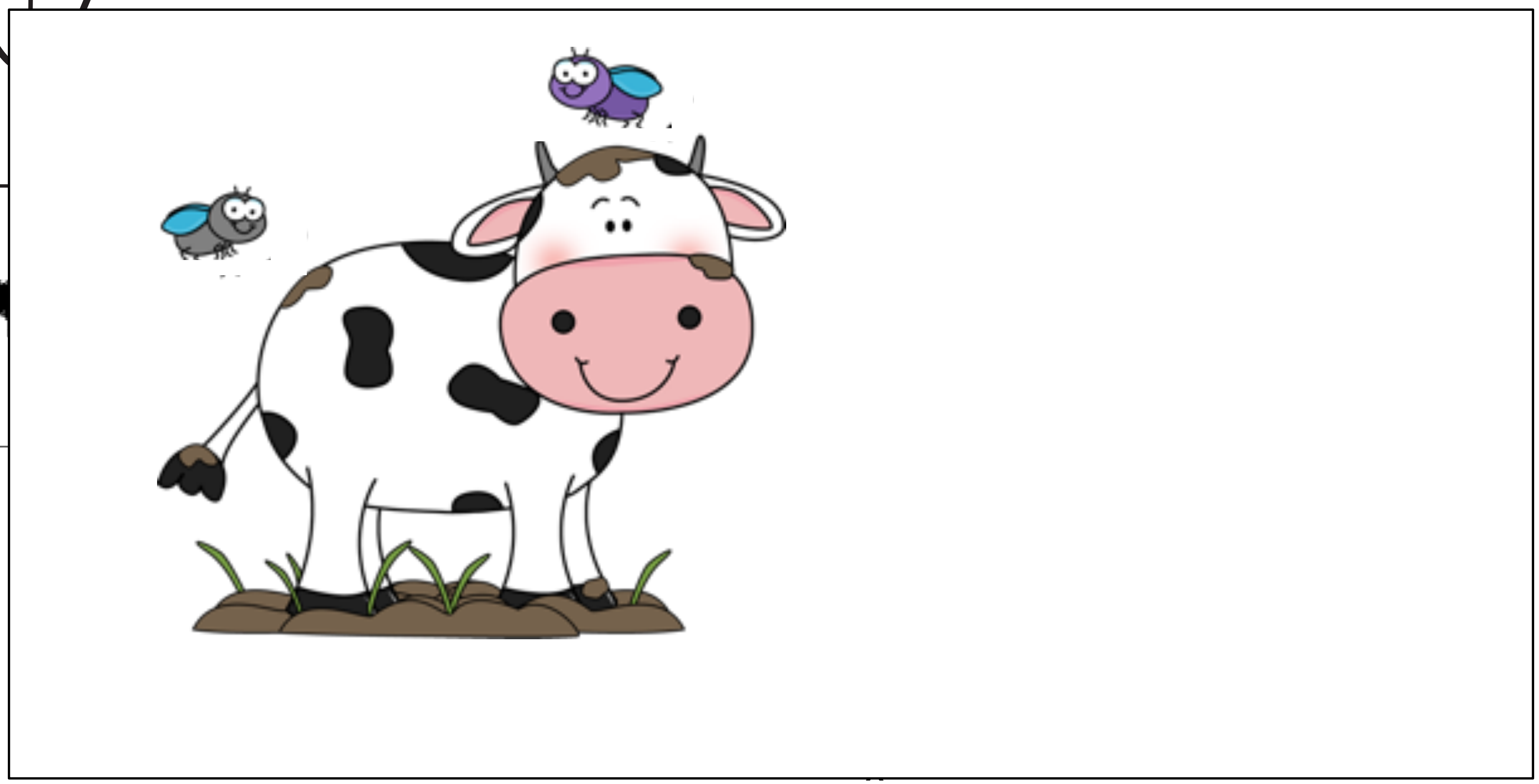


- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics



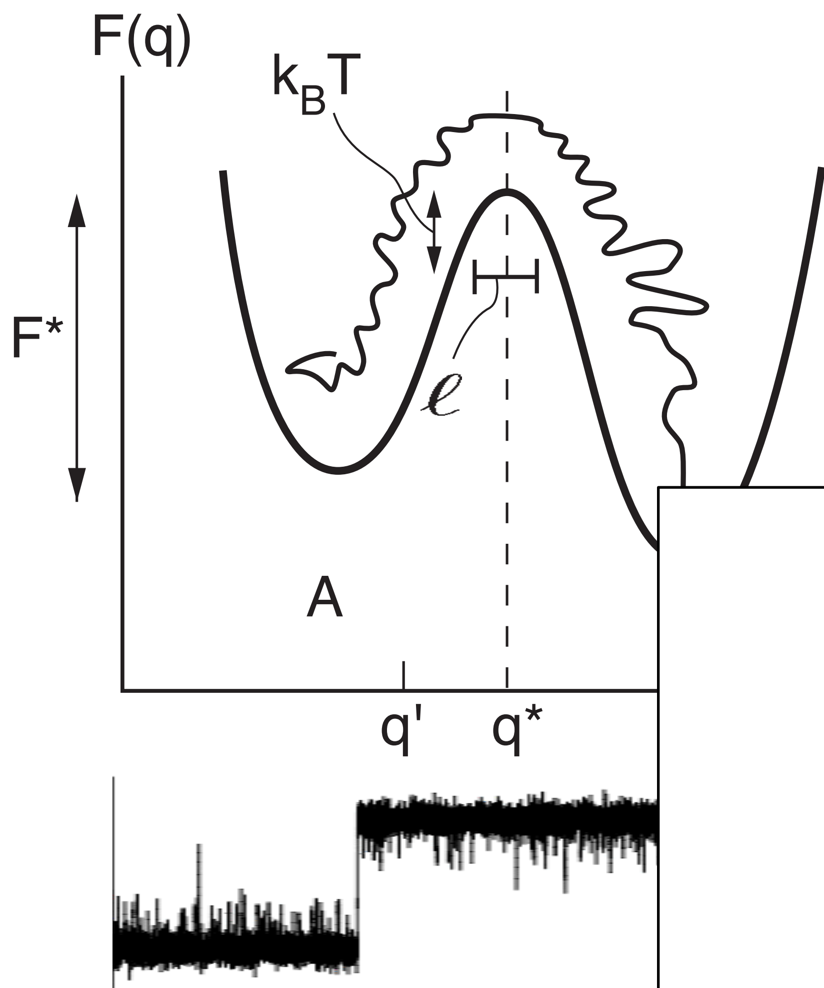
Rare event sampling



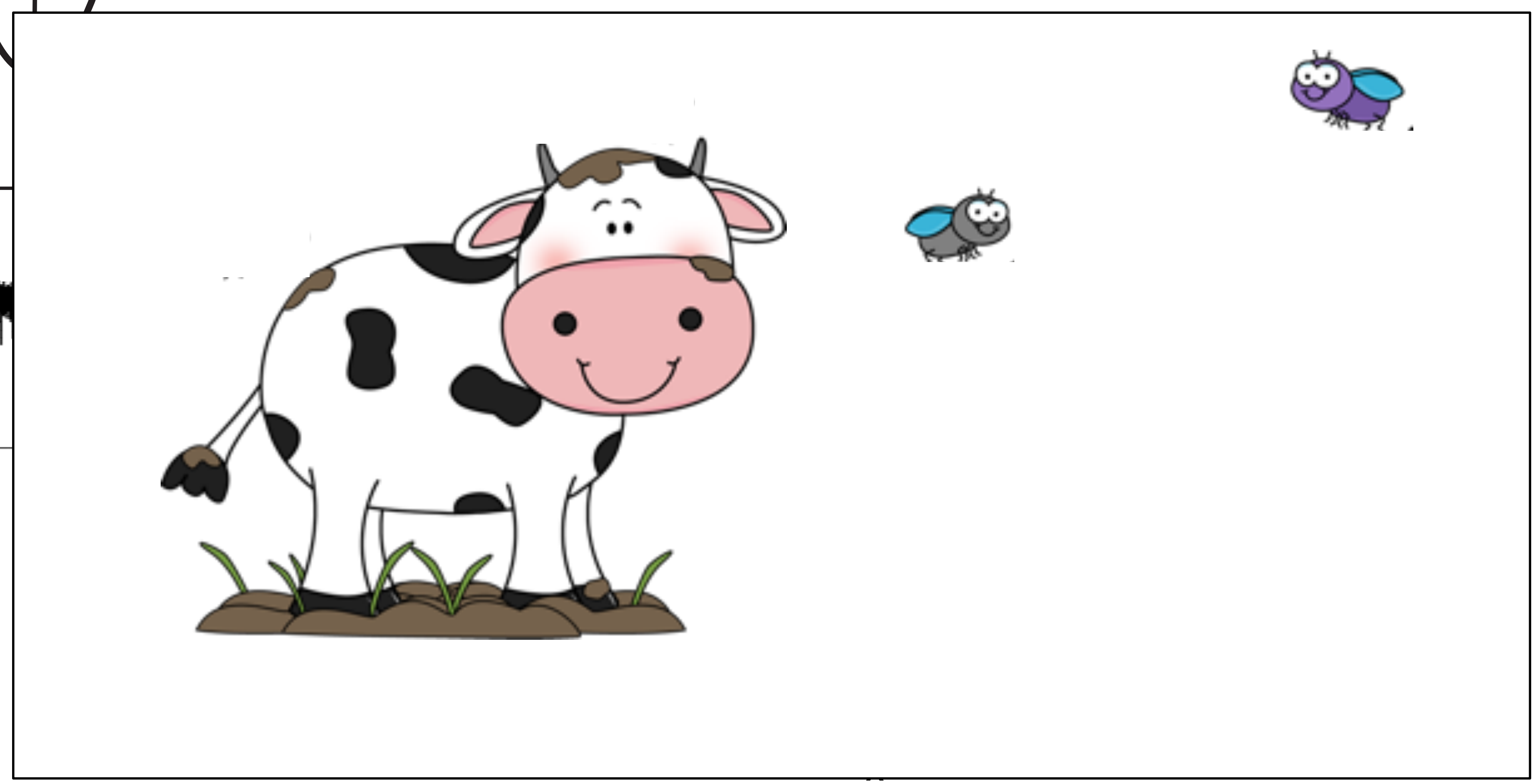


- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics

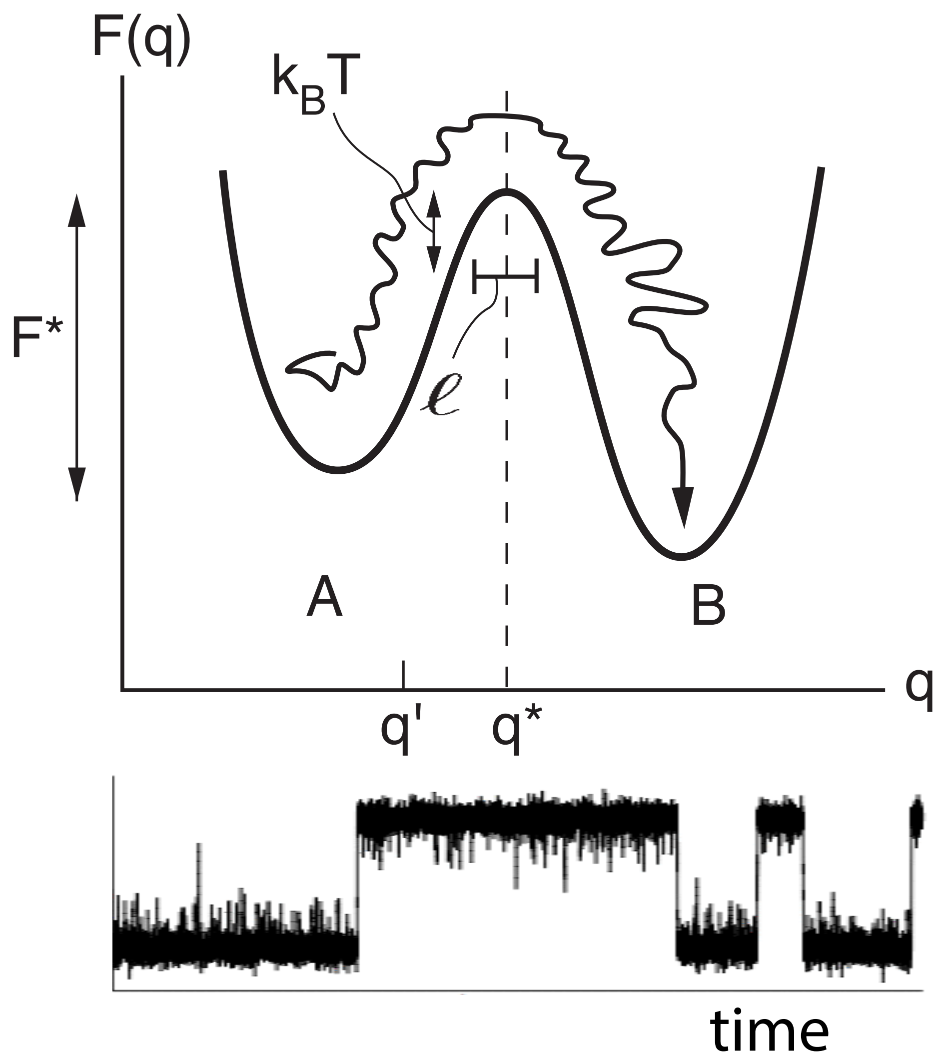
Rare event sampling



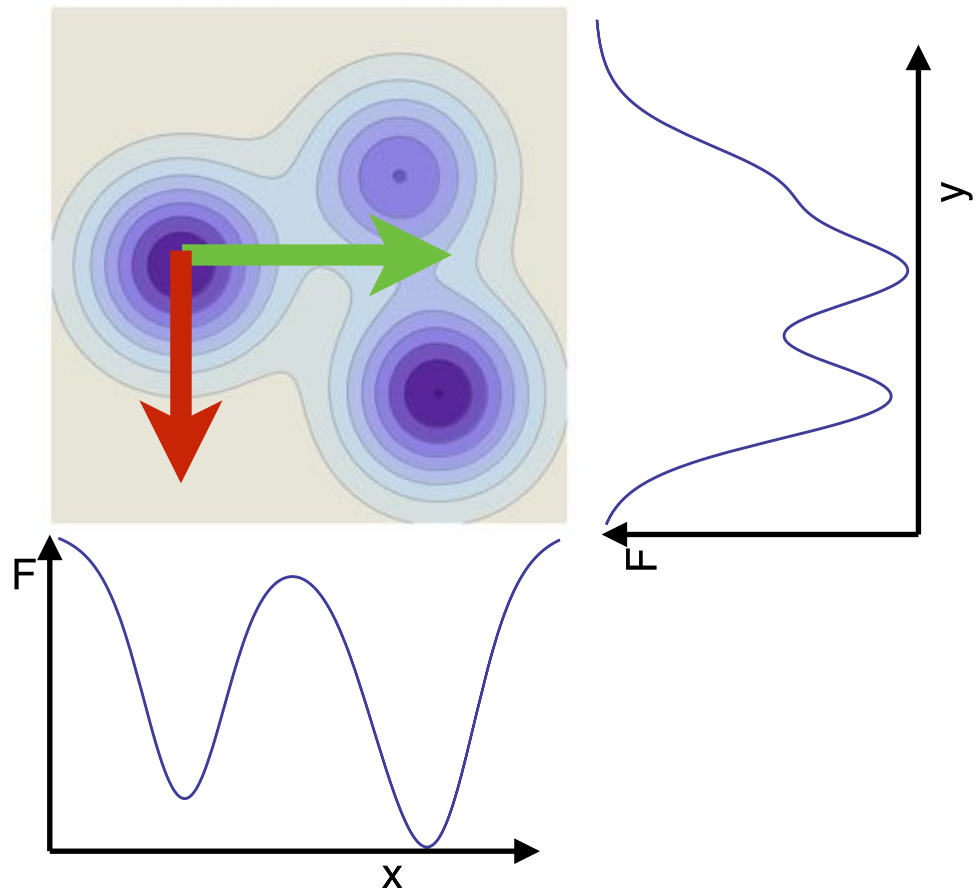
- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics



Rare event sampling



- transition state search futile in high dimensions
- enhanced sampling technique usually requires good reaction coordinate
  - umbrella sampling
  - metadynamics



The collective variable problem

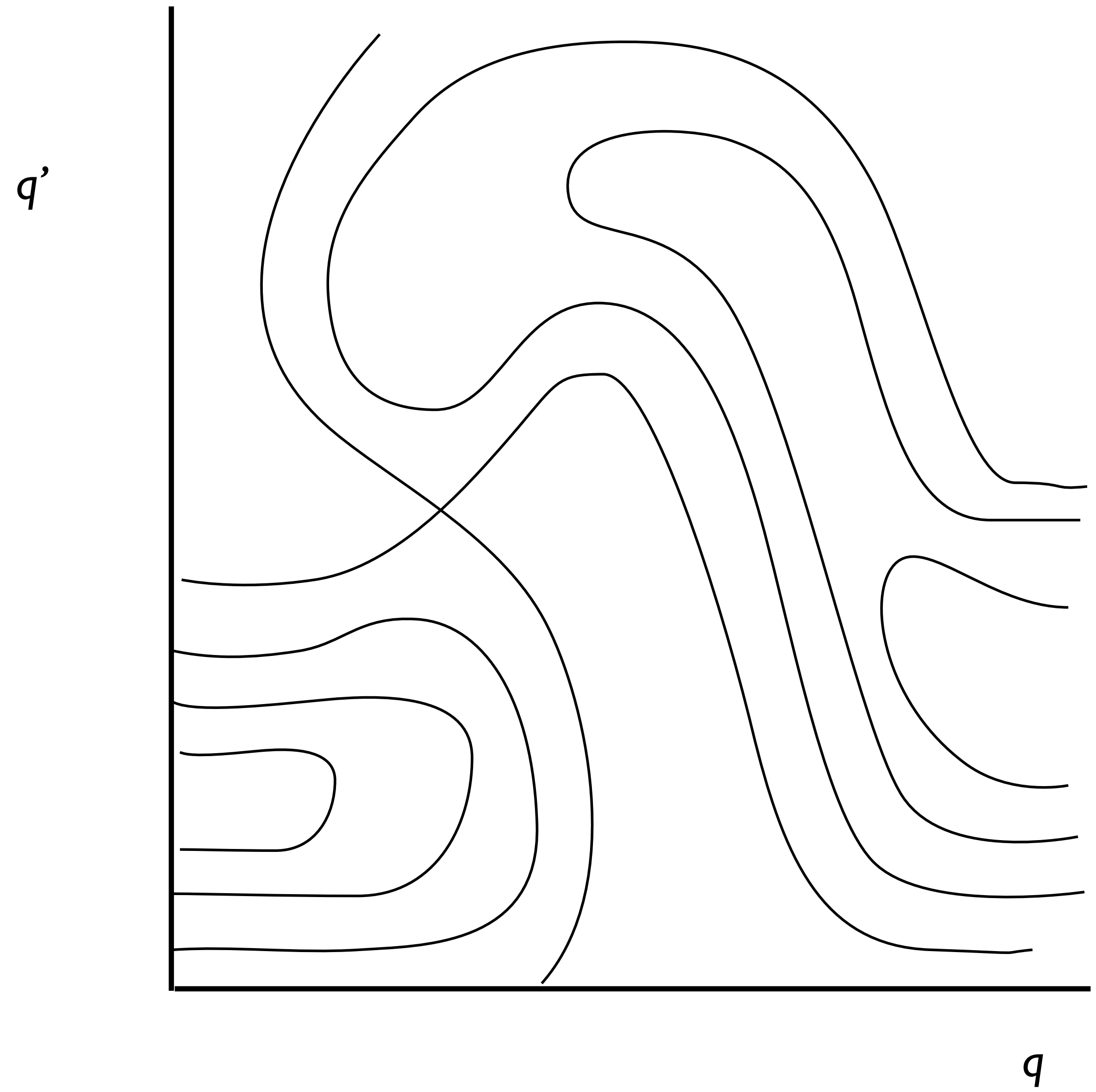
Objectives: free energy barrier, rates, transition states and mechanism.

But if reaction coordinate is not correctly represented by the collective variable, all these might be wrong!

The collective variable problem

Objectives: free energy barrier, rates, transition states and mechanism.

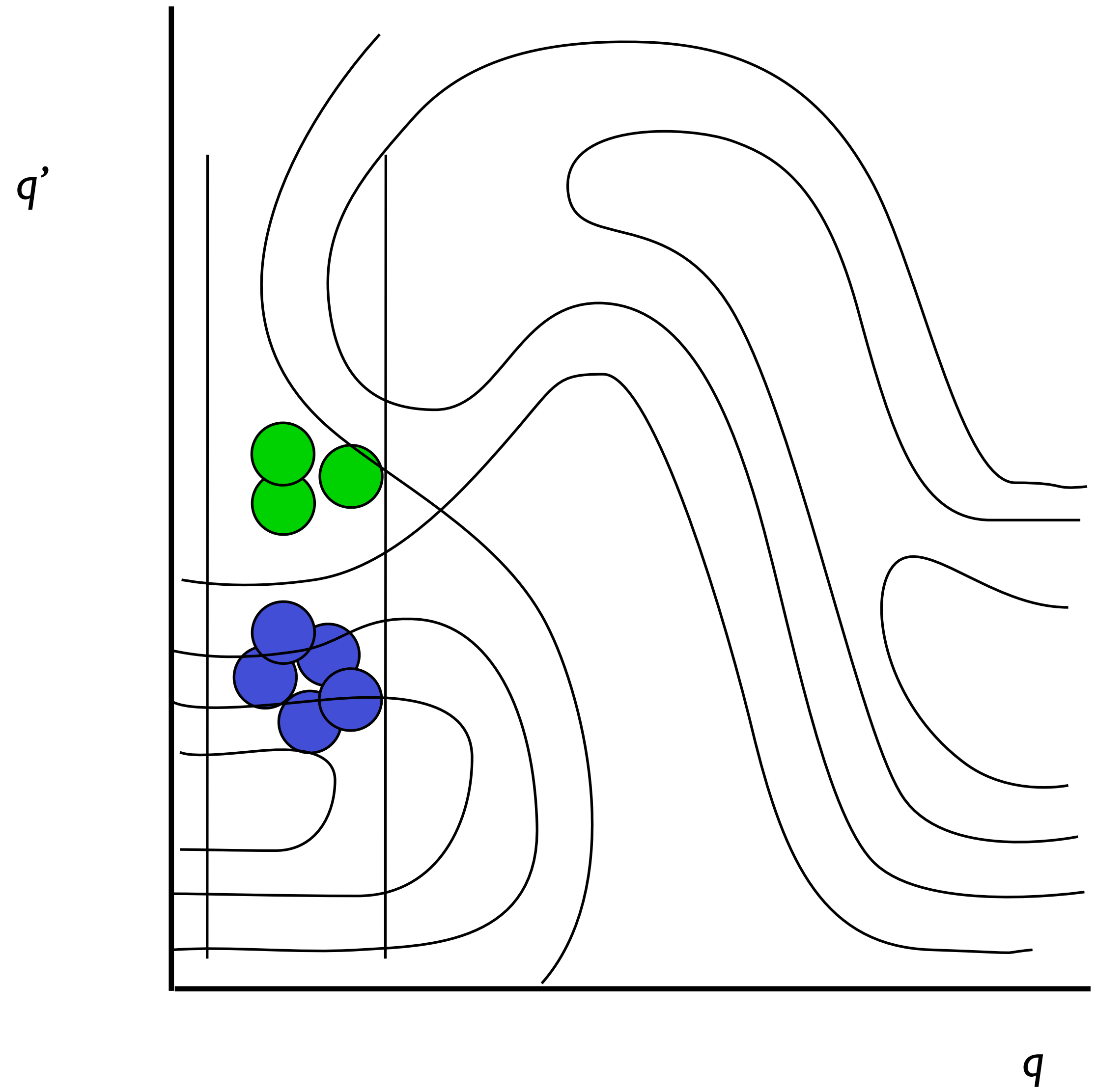
But if reaction coordinate is not correctly represented by the collective variable, all these might be wrong!



The collective variable problem

Objectives: free energy barrier, rates, transition states and mechanism.

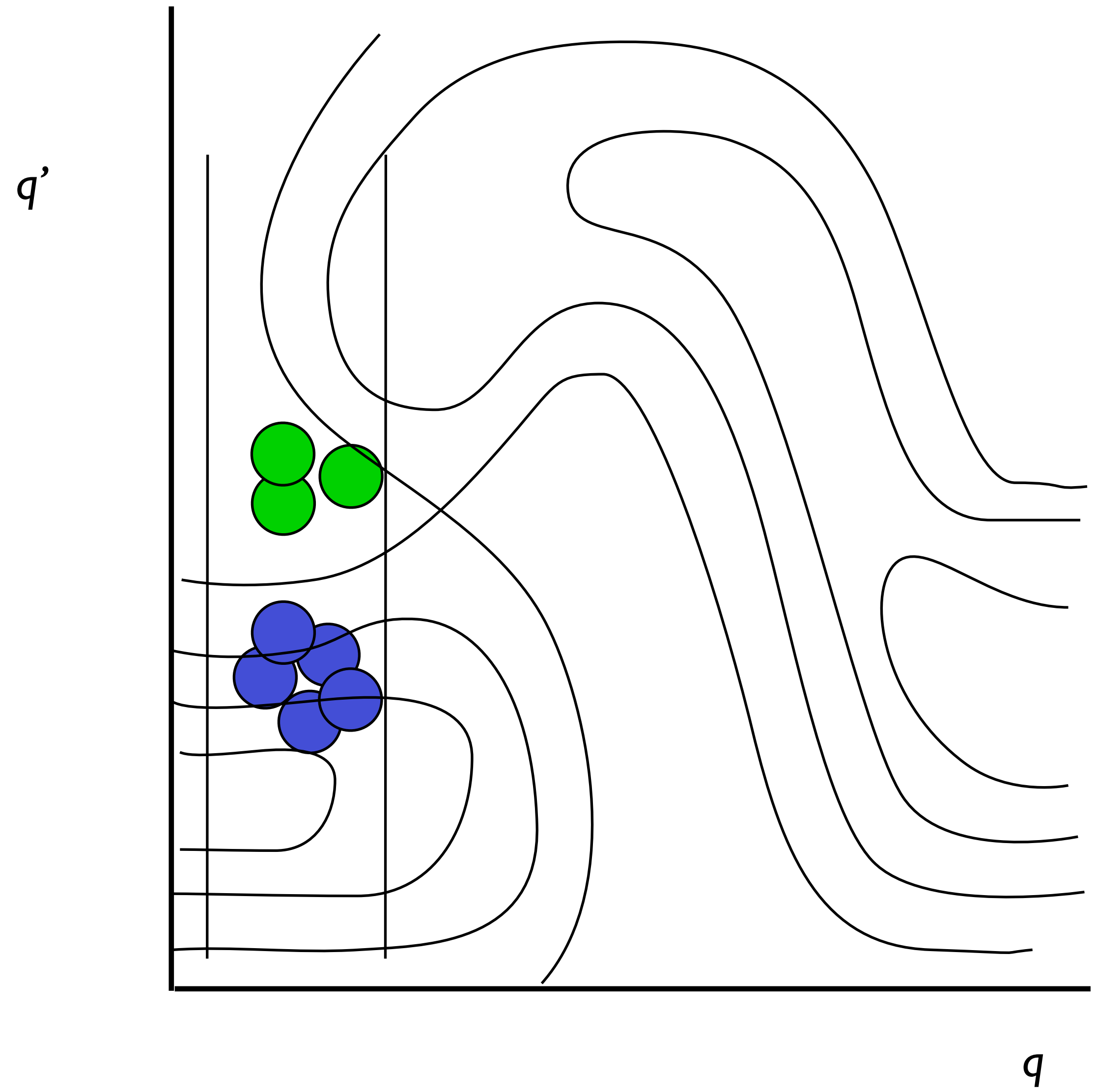
But if reaction coordinate is not correctly represented by the collective variable, all these might be wrong!



The collective variable problem

Objectives: free energy barrier, rates, transition states and mechanism.

But if reaction coordinate is not correctly represented by the collective variable, all these might be wrong!

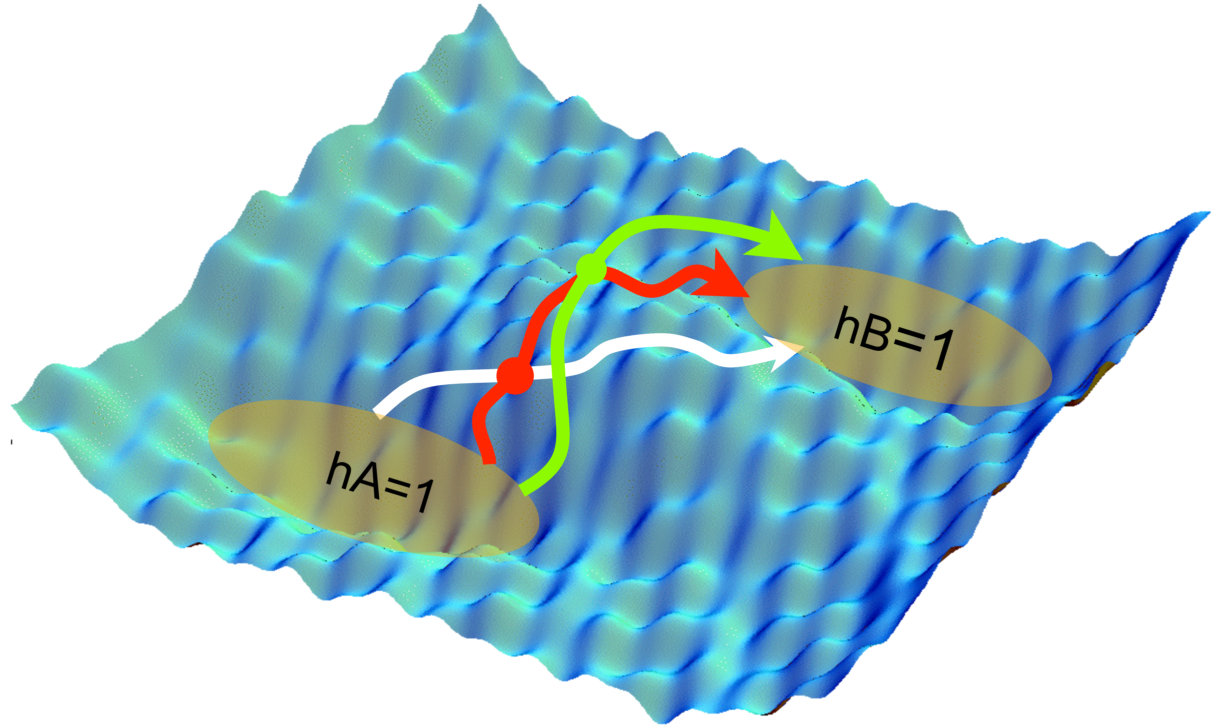


Need for methods that create pathways without prior knowledge of the RC:
Transition path sampling

Part 1: Transition path sampling

- Importance sampling of the rare event path ensemble
- yields paths, mechanisms, reaction coordinates, kinetics, and free energy

PGB, Chandler, Dellago, Geissler, Annu. Rev. Phys. Chem 2002; Dellago, PGB, Adv Polym Sci, 2009

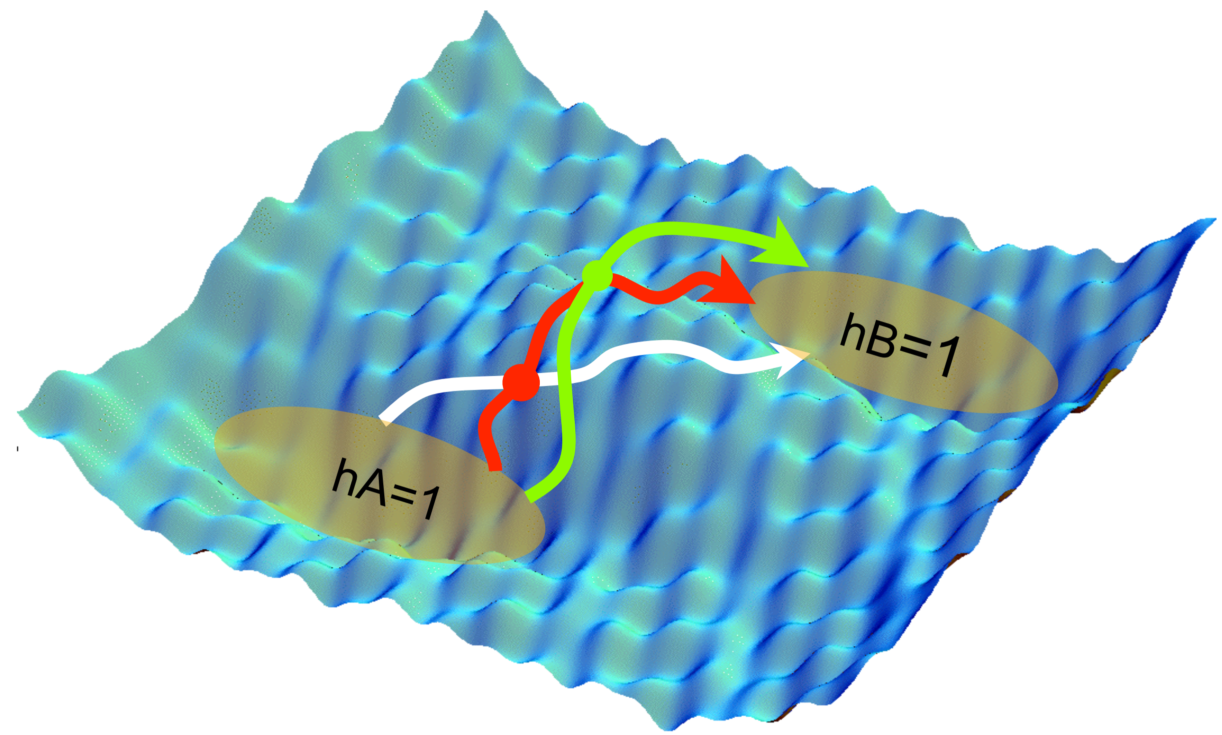


Part 1: Transition path sampling

- Importance sampling of the rare event path ensemble

- yields paths, mechanisms, reaction coordinates, kinetics, and free energy

  PGB, Chandler, Dellago, Geissler, Annu. Rev. Phys. Chem 2002; Dellago, PGB, Adv Polym Sci, 2009



- TPS philosophy: Path ensembles → mechanism → kinetics → Free Energy

- TPS gives exponential speed up w.r.t to rare event time scale
- advantages: unbiased dynamics, exact rates, independence of CVs

- Advanced Software Packages available

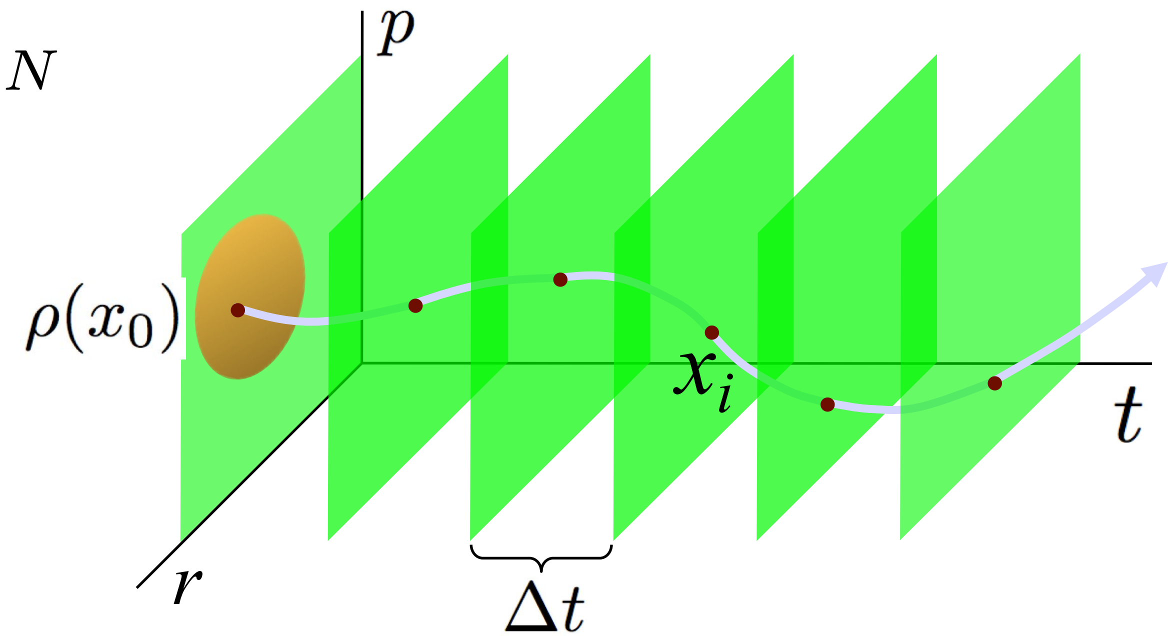
OpenPathSampling

PyRETIS
— rare events in Python

Transition path probability density

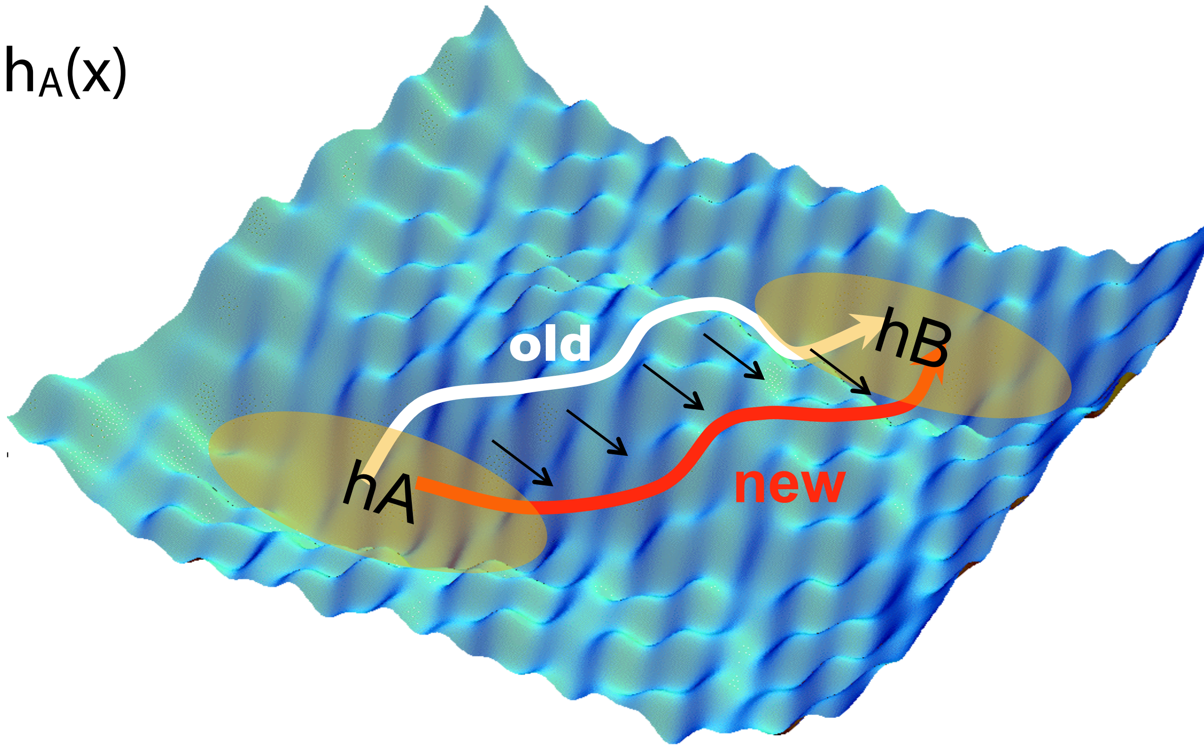
$$\mathbf{x}(L) = \{x_0, x_1, \ldots, x_L\} \qquad x \in \mathbb{R}^{6N}$$

$$\mathcal{P}\,[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} p(x_i \to x_{i+1}),$$



Define stables states A and B by indicator functions $h_A(x)$

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$



Path probability distribution

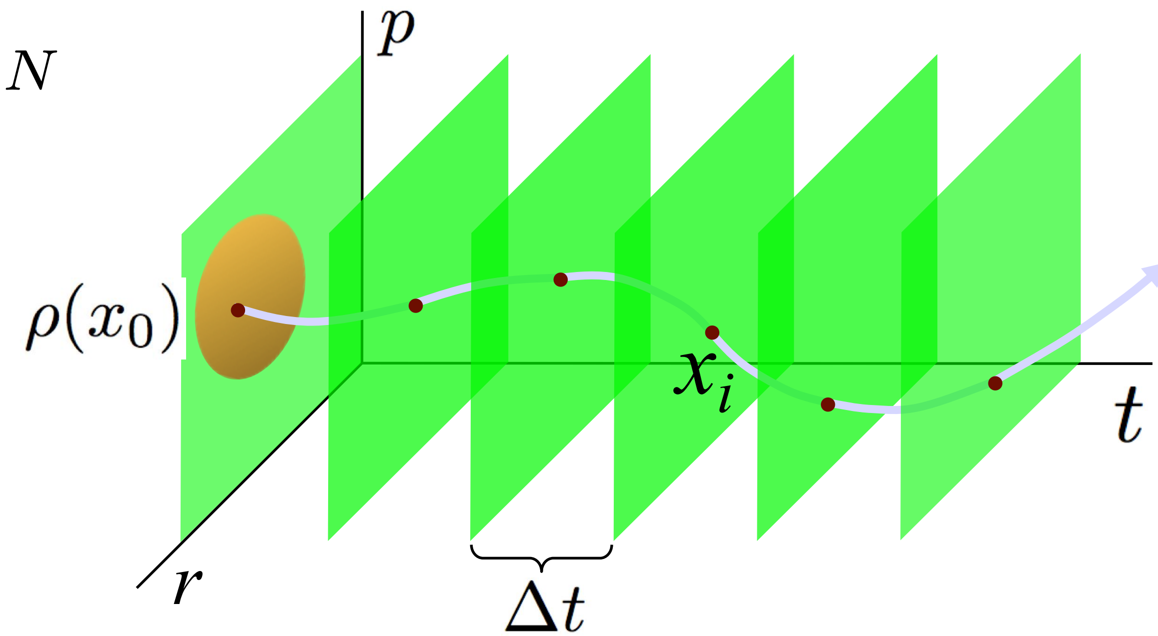
$$\mathcal{P}_{AB}[\mathbf{x}(L)] = h_A(x_0)\mathcal{P}[\mathbf{x}(L)]h_B(x_L)/Z_{AB}(L)$$

Importance sampling using Metropolis-Hastings :

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A[x_0^{(n)}]h_B[x_L^{(n)}]\min\left[1, \frac{\mathcal{P}[\mathbf{x}^{(n)}]\mathcal{P}_{gen}[\mathbf{x}^{(n)} \to \mathbf{x}^{(o)}]}{\mathcal{P}[\mathbf{x}^{(o)}]\mathcal{P}_{gen}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}]}\right].$$

Transition path probability density

$$\mathbf{x}(L) = \{x_0, x_1, \ldots, x_L\} \qquad x \in \mathbb{R}^{6N}$$

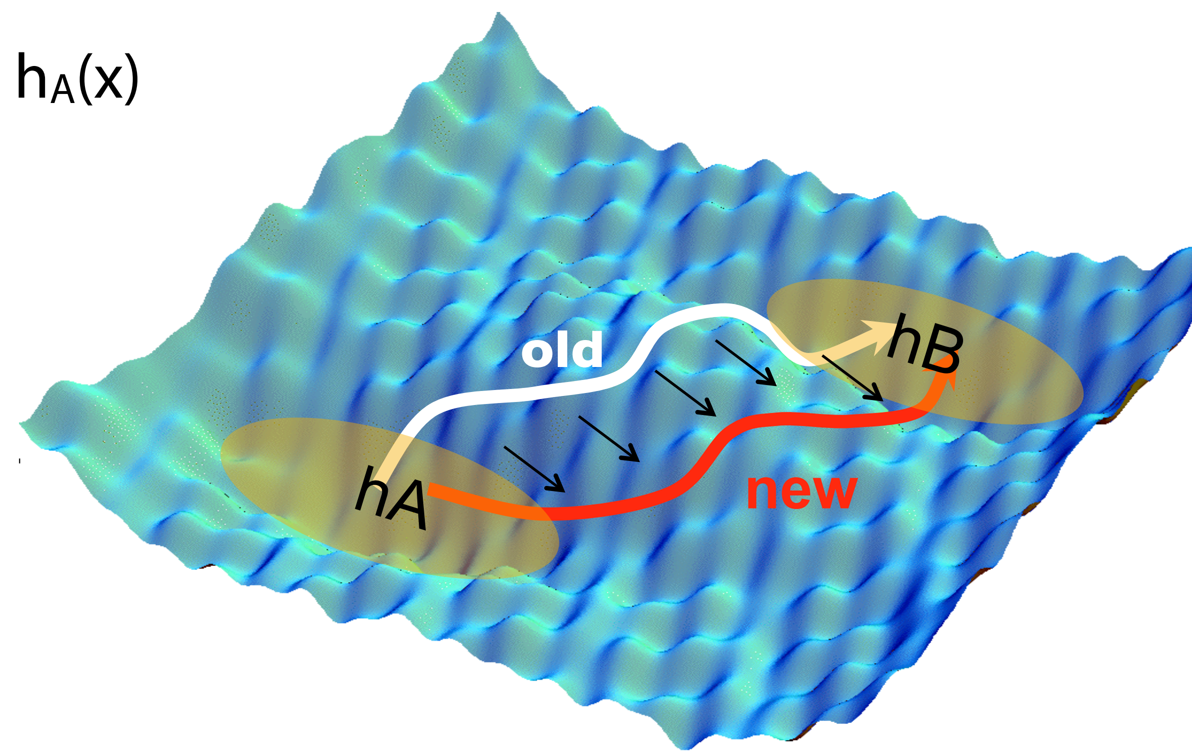


$$\mathcal{P}\ [\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} p(x_i \to x_{i+1}),$$

Initial (Boltzmann) distribution

Define stables states A and B by indicator functions $h_A(x)$

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$



Path probability distribution

$$\mathcal{P}_{AB}[\mathbf{x}(L)] = h_A(x_0)\mathcal{P}[\mathbf{x}(L)]h_B(x_L)/Z_{AB}(L)$$

Importance sampling using Metropolis-Hastings :

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A[x_0^{(n)}]h_B[x_L^{(n)}]\min\left[1, \frac{\mathcal{P}[\mathbf{x}^{(n)}]\mathcal{P}_{gen}[\mathbf{x}^{(n)} \to \mathbf{x}^{(o)}]}{\mathcal{P}[\mathbf{x}^{(o)}]\mathcal{P}_{gen}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}]}\right].$$

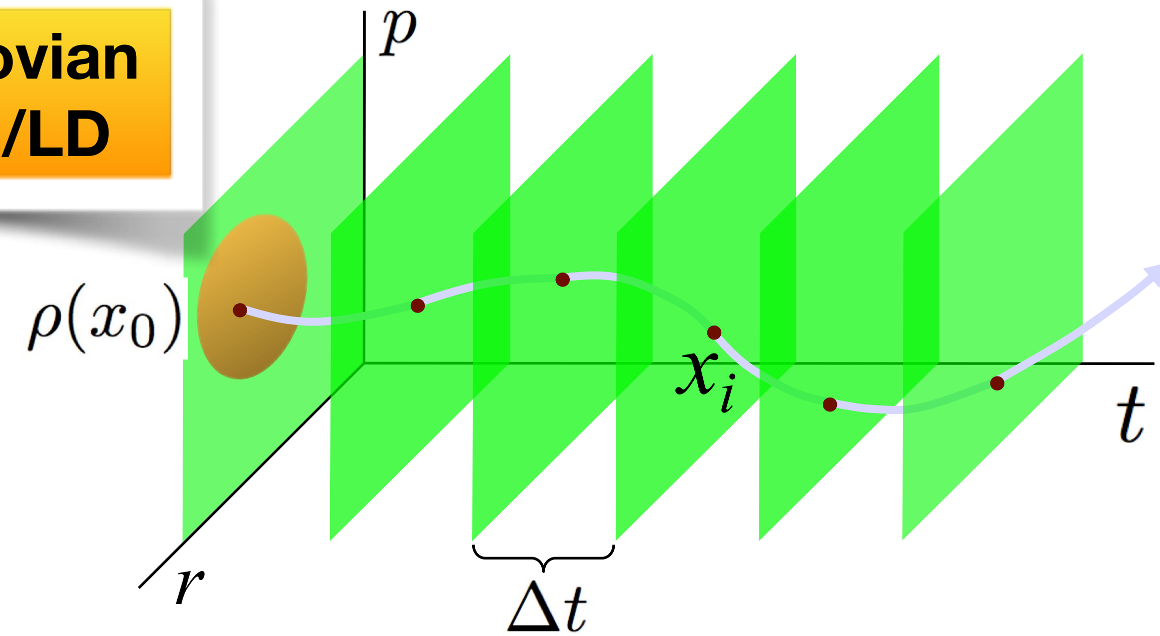
Transition path probability density

$$\mathbf{x}(L) = \{x_0, x_1, \ldots, x$$

short time Markovian propagator: MD/LD

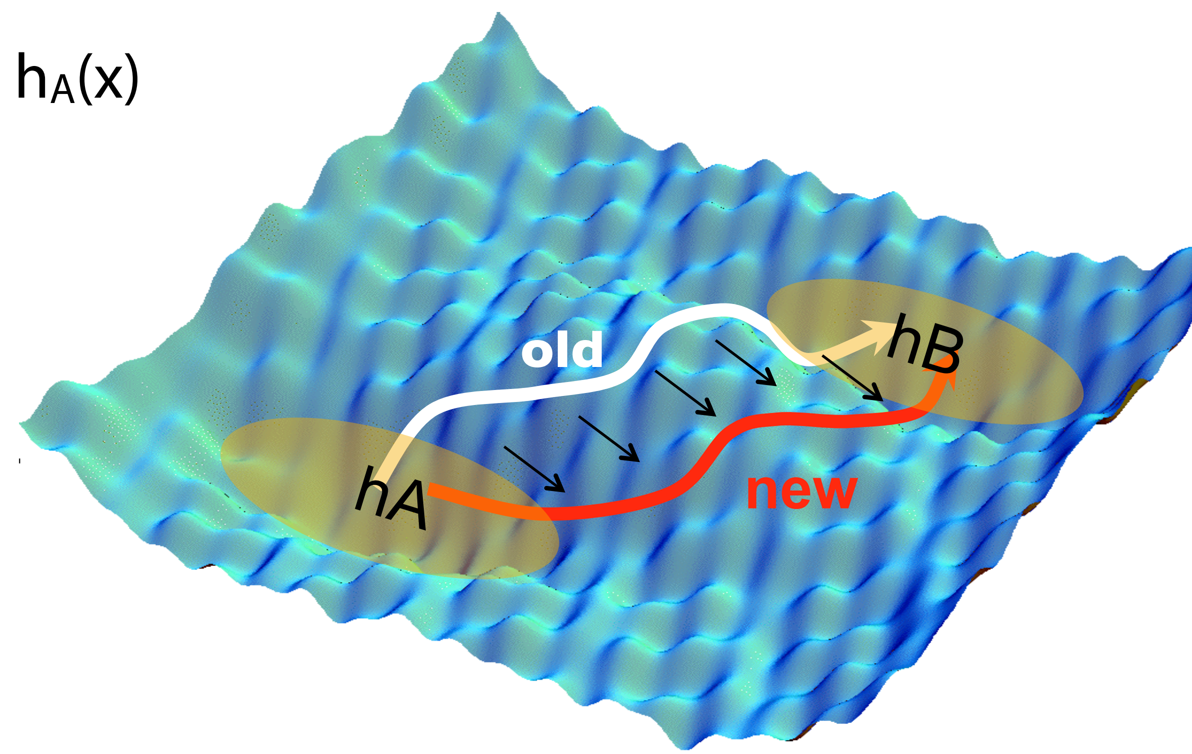
$$\mathcal{P}\ [\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} p(x_i \rightarrow x_{i+1}),$$

Initial (Boltzmann) distribution



Define stables states A and B by indicator functions $h_A(x)$

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$



Path probability distribution

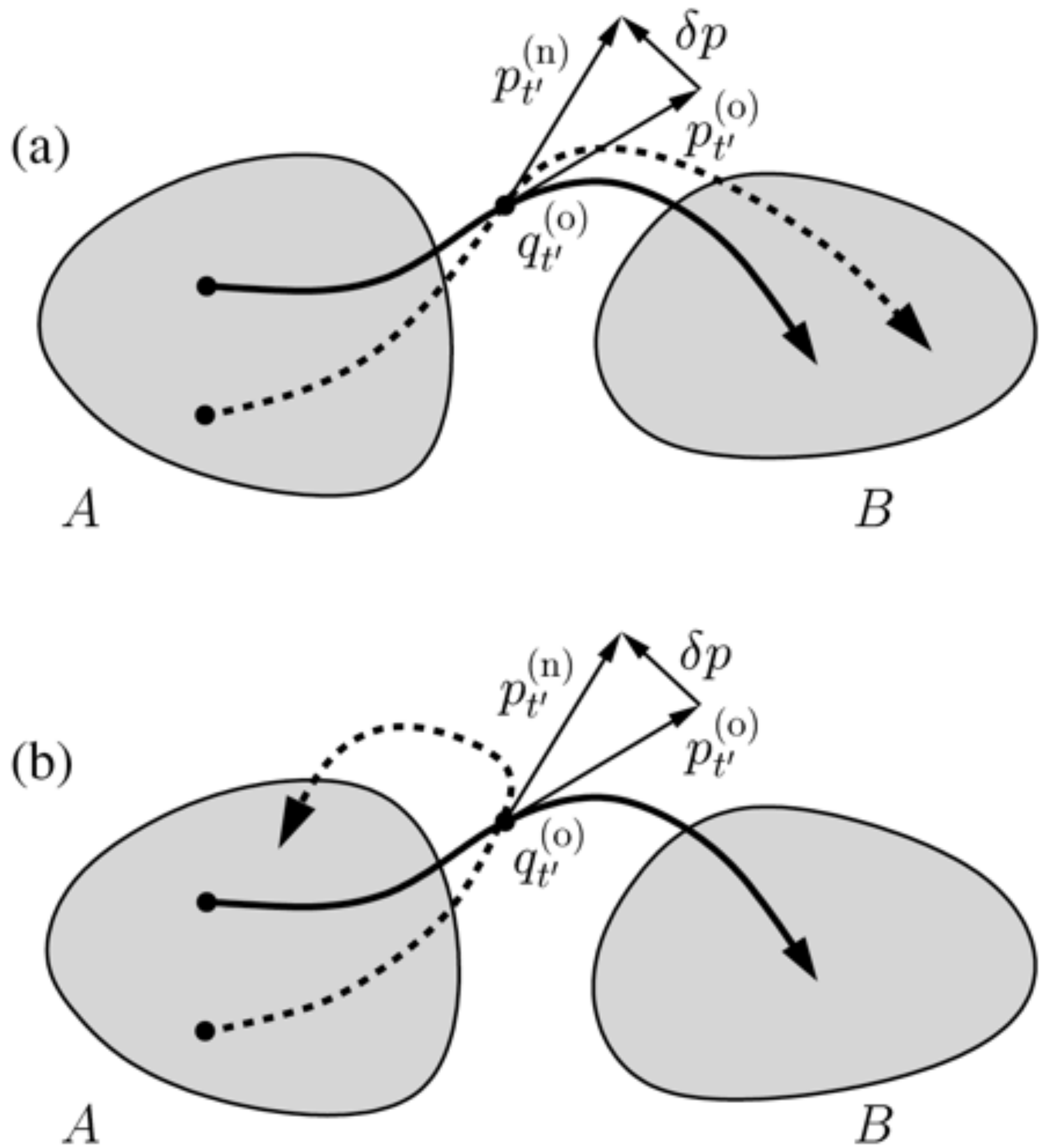
$$\mathcal{P}_{AB}[\mathbf{x}(L)] = h_A(x_0)\mathcal{P}[\mathbf{x}(L)]h_B(x_L)/Z_{AB}(L)$$

Importance sampling using Metropolis-Hastings :

$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A[x_0^{(n)}]h_B[x_L^{(n)}] \min\left[1, \frac{\mathcal{P}[\mathbf{x}^{(n)}]\mathcal{P}_{gen}[\mathbf{x}^{(n)} \rightarrow \mathbf{x}^{(o)}]}{\mathcal{P}[\mathbf{x}^{(o)}]\mathcal{P}_{gen}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}]}\right].$$

Shooting move

-



accept

reject

$$P_{\text{acc}}[x^{(\text{o})}(\mathcal{T}) \rightarrow x^{(\text{n})}(\mathcal{T})] = h_A[x_0^{(\text{n})}]h_B[x_{\mathcal{T}}^{(\text{n})}]$$

$$h_A(t) = \begin{cases} 1 & if \quad x_t \in A \\ 0 & if \quad x_t \notin A \end{cases}$$

PGB, C. Dellago and D. Chandler, Faraday Discuss., 1998, 110, 421

Acceptance rule

backward MD shot

$$P_{gen}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = p_{gen}(x_{\tau'}^{(o)} \to x_{\tau'}^{(o)}) \prod_{i=\tau'}^{L-1} p(x_i^{(n)} \to x_{i+1}^{(n)}) \prod_{i=1}^{\tau'} \bar{p}(x_i^{(n)} \to x_{i-1}^{(n)})$$

forward MD shot

$\bar{p}(x \to y) = p(\bar{x} \to \bar{y})$ backward in time by momenta reversal $\bar{x} = \{r, -p\}$ for $x = \{r, p\}$

assuming symmetric generation probability

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)}) \min \left[1, \frac{\rho(x_0^{(n)})}{\rho(x_0^{(o)})} \prod_{i=1}^{\tau'} \frac{p(x_i^{(n)} \to x_{i+1}^{(n)})}{\bar{p}(x_{i+1}^{(n)} \to x_i^{(n)})} \times \frac{\bar{p}(x_{i+1}^{(o)} \to x_i^{(o)})}{p(x_i^{(o)} \to x_{i+1}^{(o)})} \right]$$

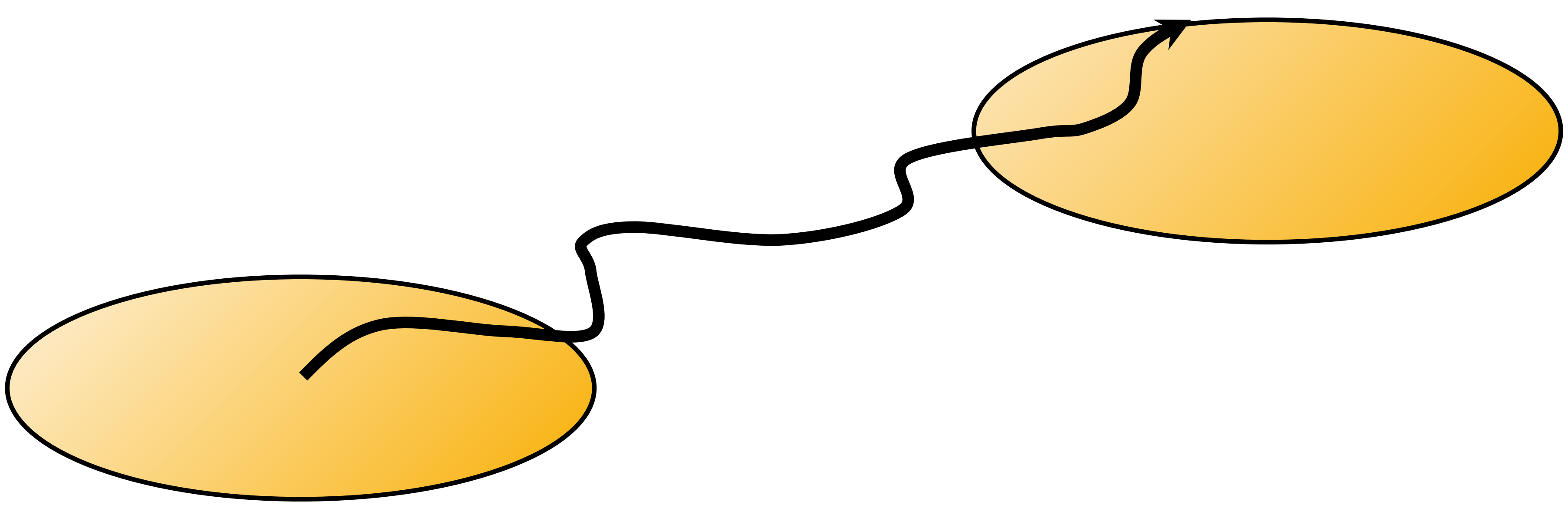
microscopic reversibility $$\frac{p(x \to y)}{\bar{p}(y \to x)} = \frac{\rho(y)}{\rho(x)}$$

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)}) \min \left[1, \frac{\rho(x_{\tau'}^{(n)})}{\rho(x_{\tau'}^{(o)})} \right]$$

$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)})$ for constant energy at shooting point

Standard TPS algorithm

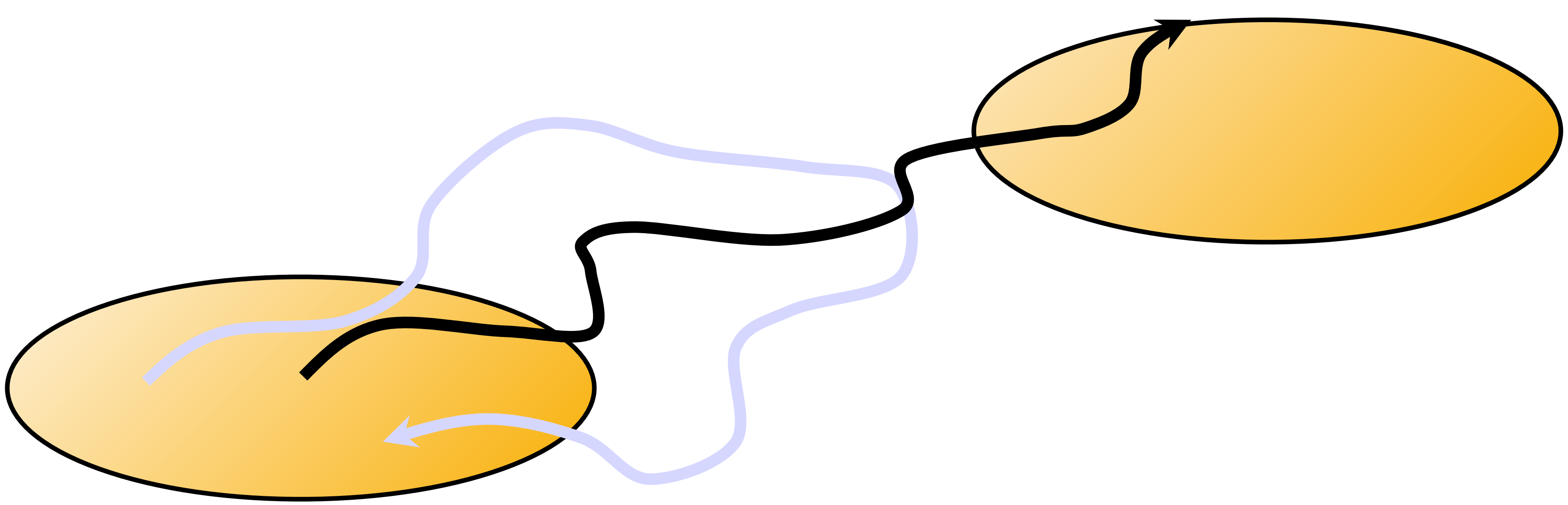
- take existing path
- choose random time slice *t*
- change momenta slightly at *t*
- integrate forward and backward in time to create new path of length L
- accept if A and B are connected, otherwise reject and retain old path
- calculate averages
- repeat



C.Dellago, PGB, P. L Geissler, , Adv. Chem Phys, 2002

Standard TPS algorithm

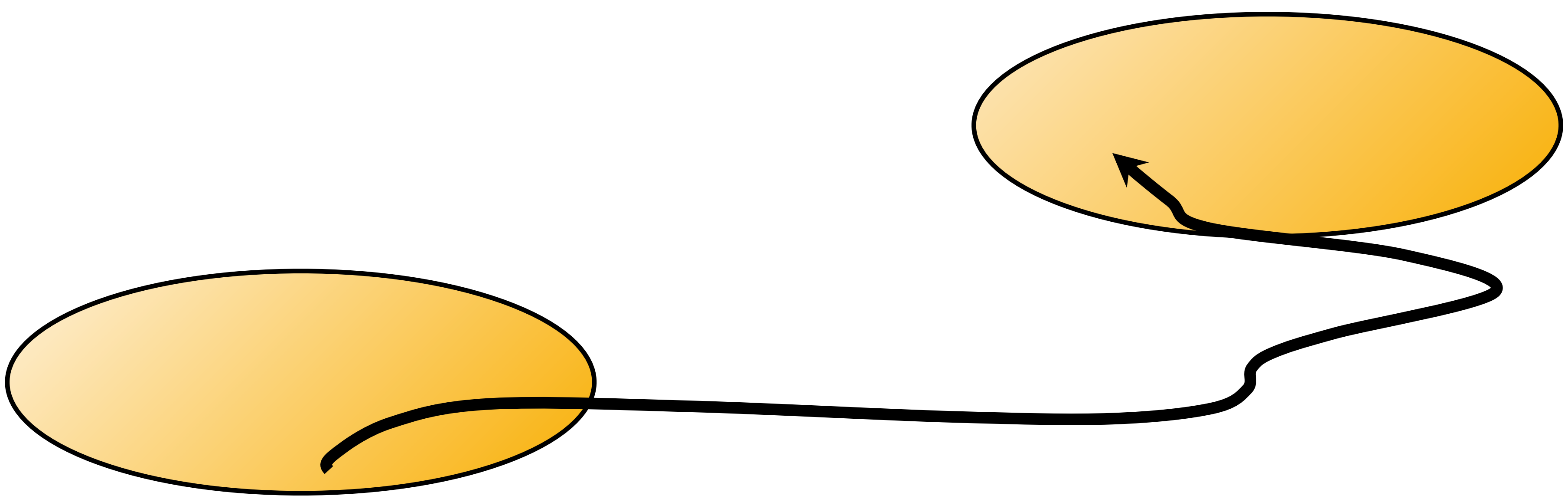
- take existing path
- choose random time slice *t*
- change momenta slightly at *t*
- integrate forward and backward in time to create new path of length L
- accept if A and B are connected, otherwise reject and retain old path
- calculate averages
- repeat



C.Dellago, PGB, P. L Geissler, , Adv. Chem Phys, 2002

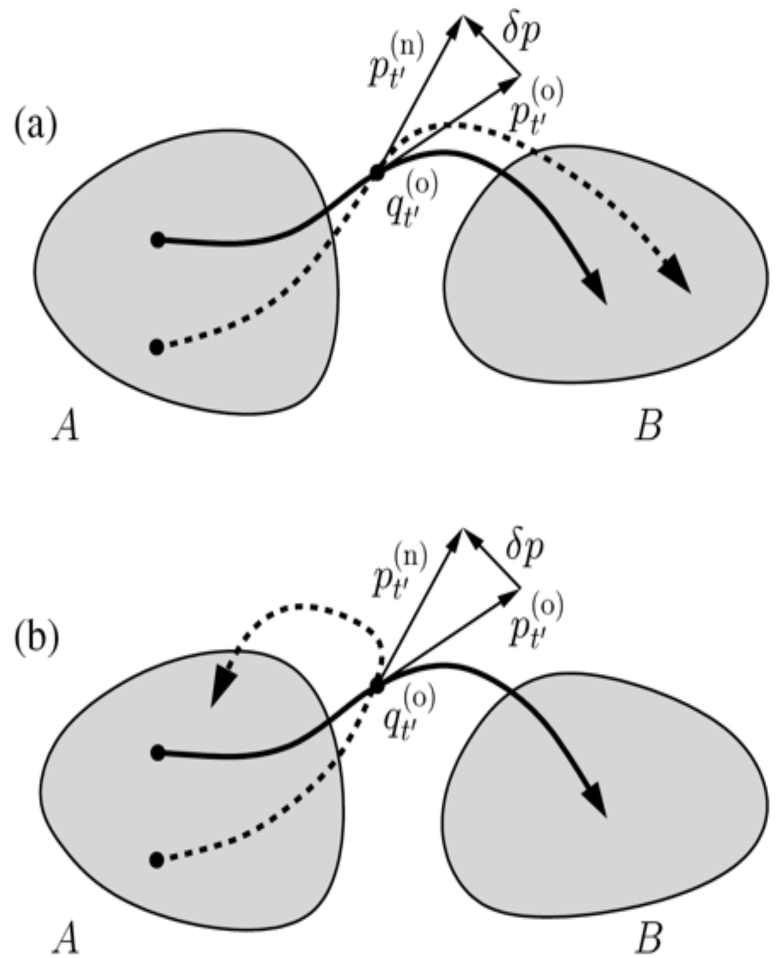
Standard TPS algorithm

- take existing path
- choose random time slice *t*
- change momenta slightly at *t*
- integrate forward and backward in time to create new path of length L
- accept if A and B are connected, otherwise reject and retain old path
- calculate averages
- repeat



C.Dellago, PGB, P. L Geissler, , Adv. Chem Phys, 2002

Shooting moves



accept

reject

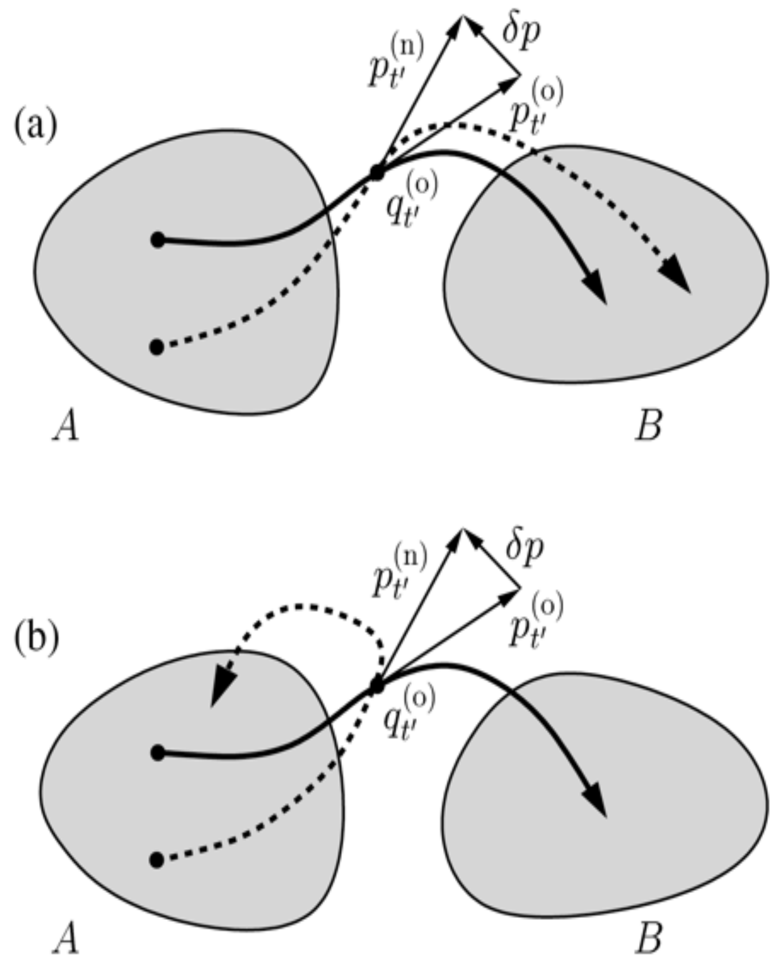
$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = h_A[x_0^{(\mathrm{n})}]h_B[x_{\mathcal{T}}^{(\mathrm{n})}]$$

arbitrary frame selection probability $p_{sel}(\tau, x)$

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)}) \min\left[1, \frac{p_{sel}(\tau', \mathbf{x}^{(n)})}{p_{sel}(\tau, \mathbf{x}^{(o)})}\right]$$

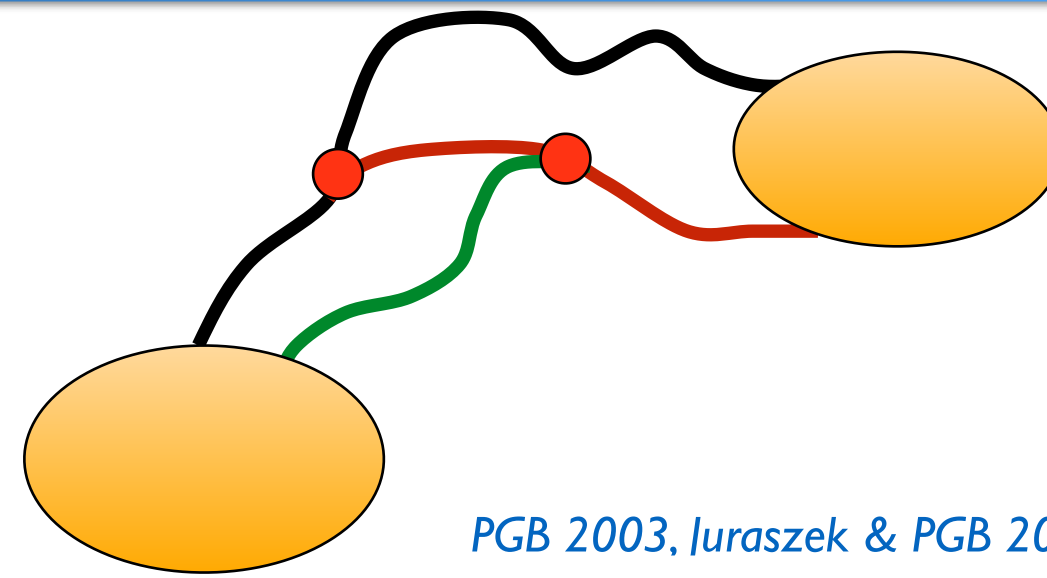
PGB, C. Dellago and D. Chandler, Faraday Discuss., 1998, 110, 421

Shooting moves



accept

reject



PGB 2003, Juraszek & PGB 2006)

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)})\min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$

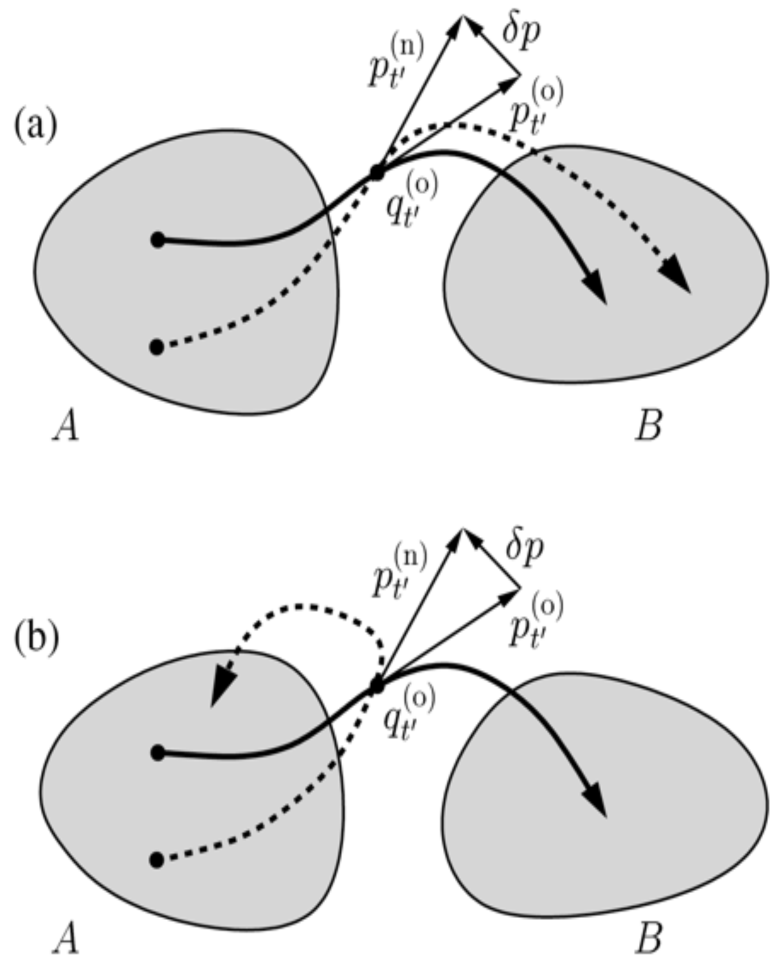
$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = h_A[x_0^{(\mathrm{n})}]h_B[x_{\mathcal{T}}^{(\mathrm{n})}]$$

arbitrary frame selection probability $p_{sel}(\tau,x)$

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)}) \min\left[1, \frac{p_{sel}(\tau', \mathbf{x}^{(n)})}{p_{sel}(\tau, \mathbf{x}^{(o)})}\right]$$

PGB, C. Dellago and D. Chandler, Faraday Discuss., 1998, 110, 421

Shooting moves



accept

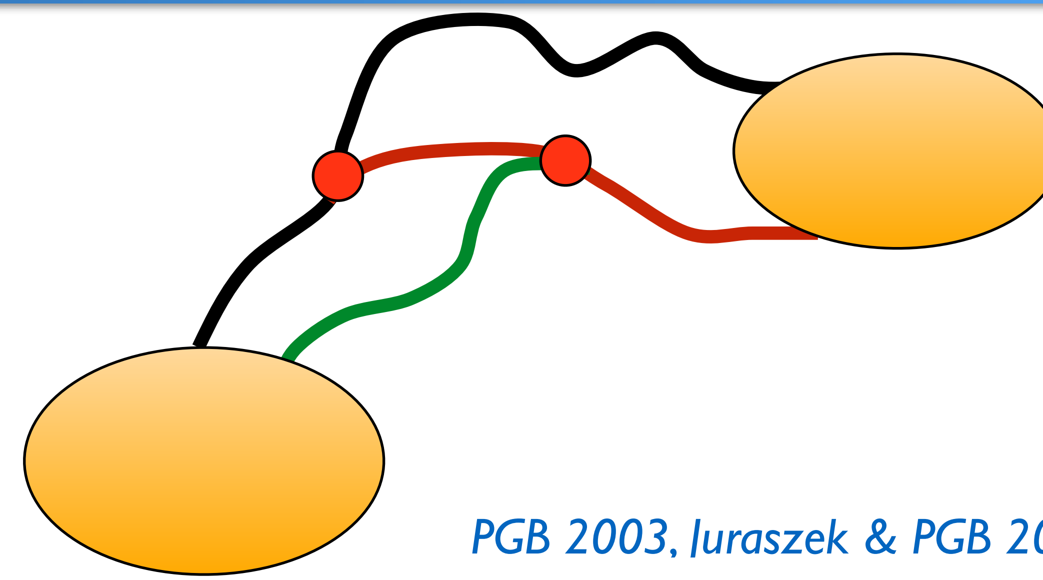
reject

$$P_{\text{acc}}[x^{(\text{o})}(\mathcal{T}) \to x^{(\text{n})}(\mathcal{T})] = h_A[x_0^{(\text{n})}]h_B[x_{\mathcal{T}}^{(\text{n})}]$$

arbitrary frame selection probability $p_{sel}(\tau, x)$

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)}) \min\left[1, \frac{p_{sel}(\tau', \mathbf{x}^{(n)})}{p_{sel}(\tau, \mathbf{x}^{(o)})}\right]$$

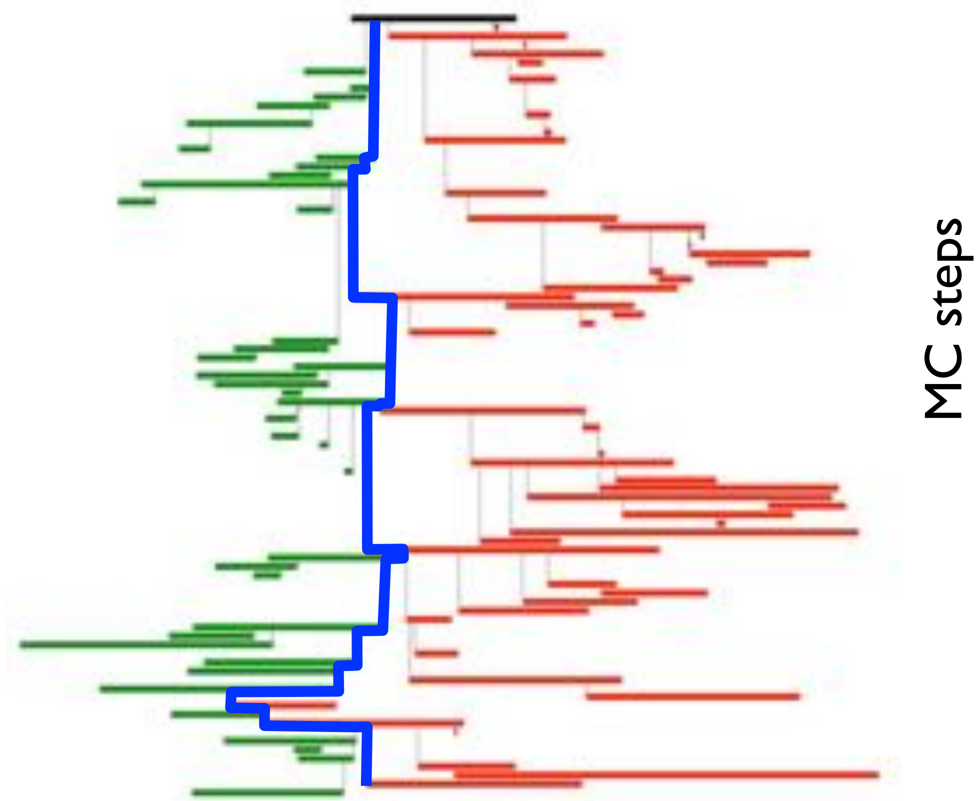
PGB, C. Dellago and D. Chandler, Faraday Discuss., 1998, 110, 421



PGB 2003, Juraszek & PGB 2006)

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)})\min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$

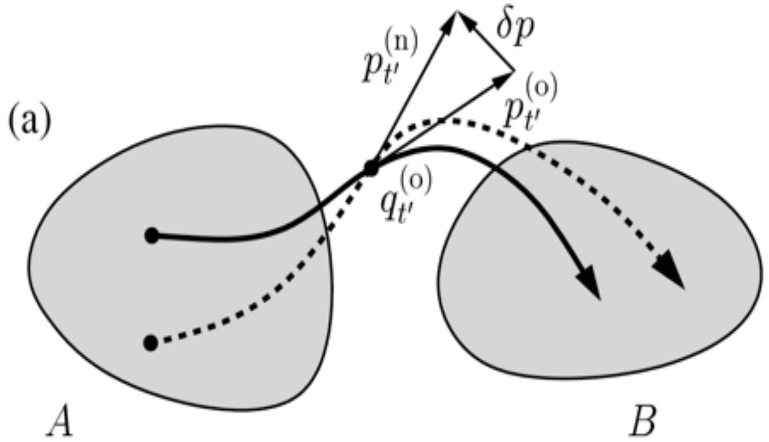
One way flexible shooting efficient but needs to be checked for decorrelation of paths



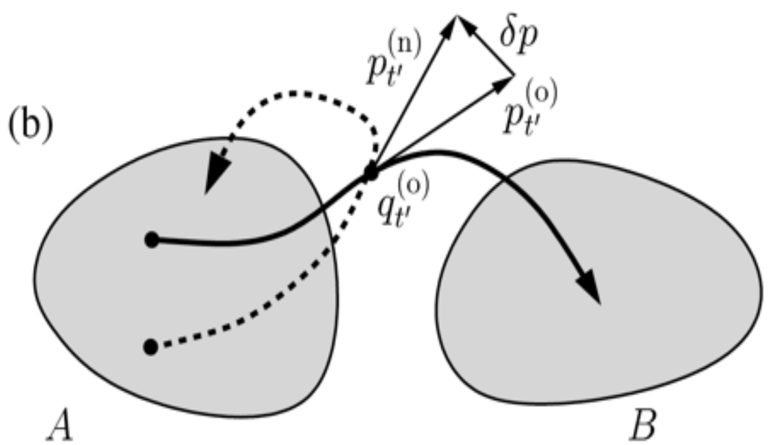
Many shooting variants

PGB and Swenson, Adv. Theor. Simul. 4, 2000237 (2021)

Shooting moves



accept



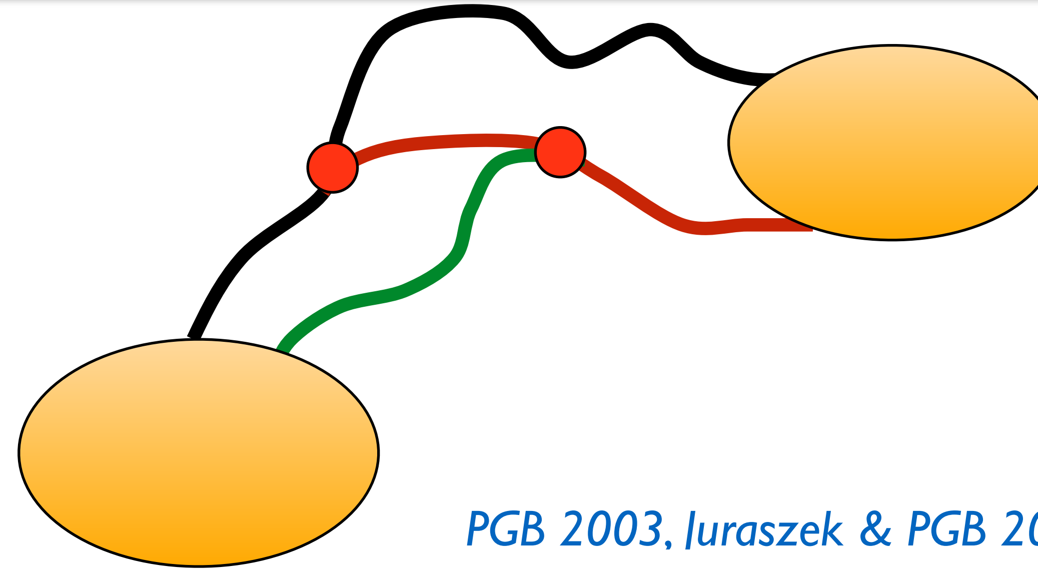
reject

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \rightarrow x^{(\mathrm{n})}(\mathcal{T})] = h_A[x_0^{(\mathrm{n})}]h_B[x_{\mathcal{T}}^{(\mathrm{n})}]$$

arbitrary frame selection probability $p_{sel}(\tau,x)$

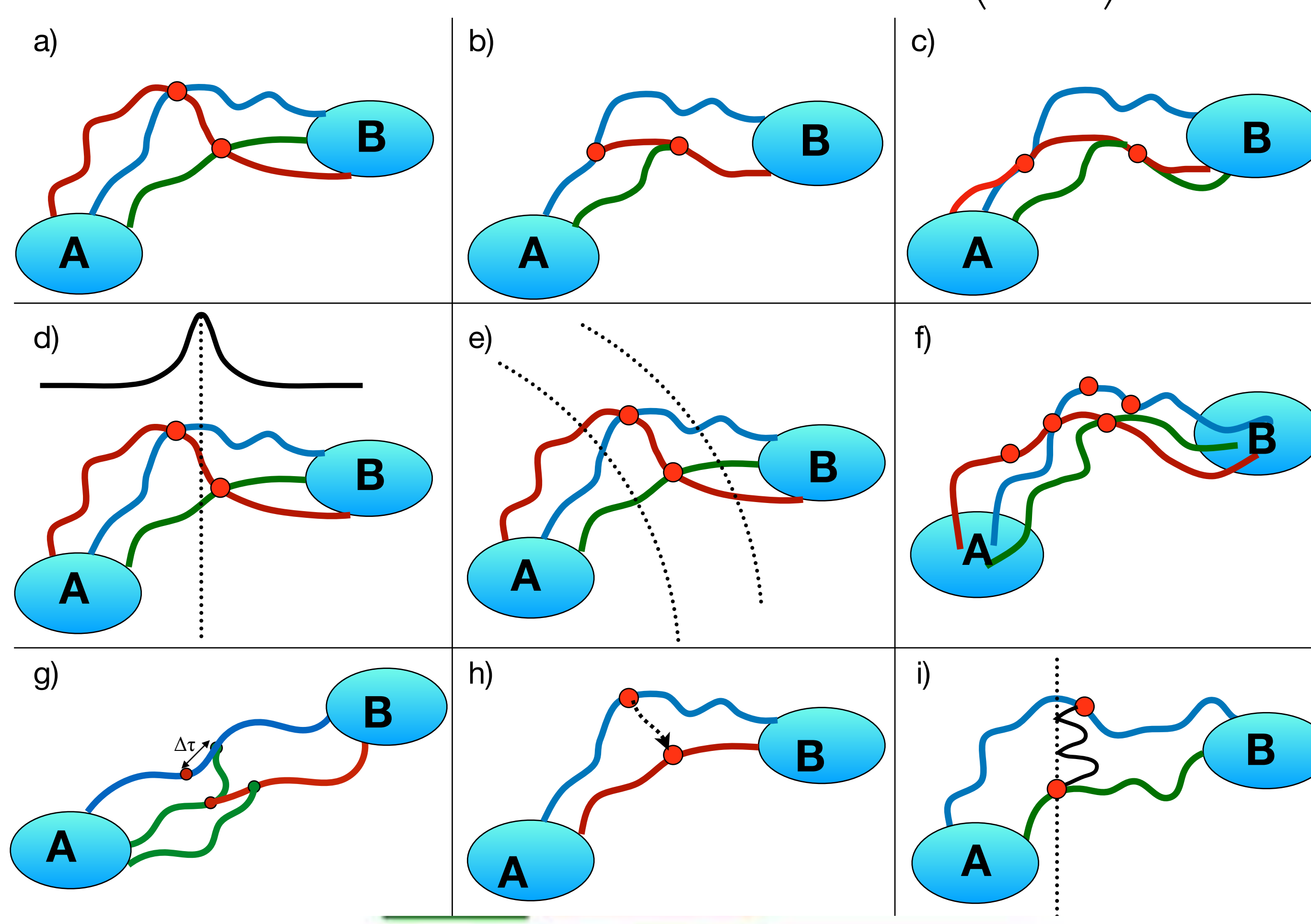
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)}) \min\left[1, \frac{p_{sel}(\tau', \mathbf{x}^{(n)})}{p_{sel}(\tau, \mathbf{x}^{(o)})}\right]$$

PGB, C. Dellago and D. Chandler, Faraday Discuss., 1998, 110, 421



PGB 2003, Juraszek & PGB 2006)

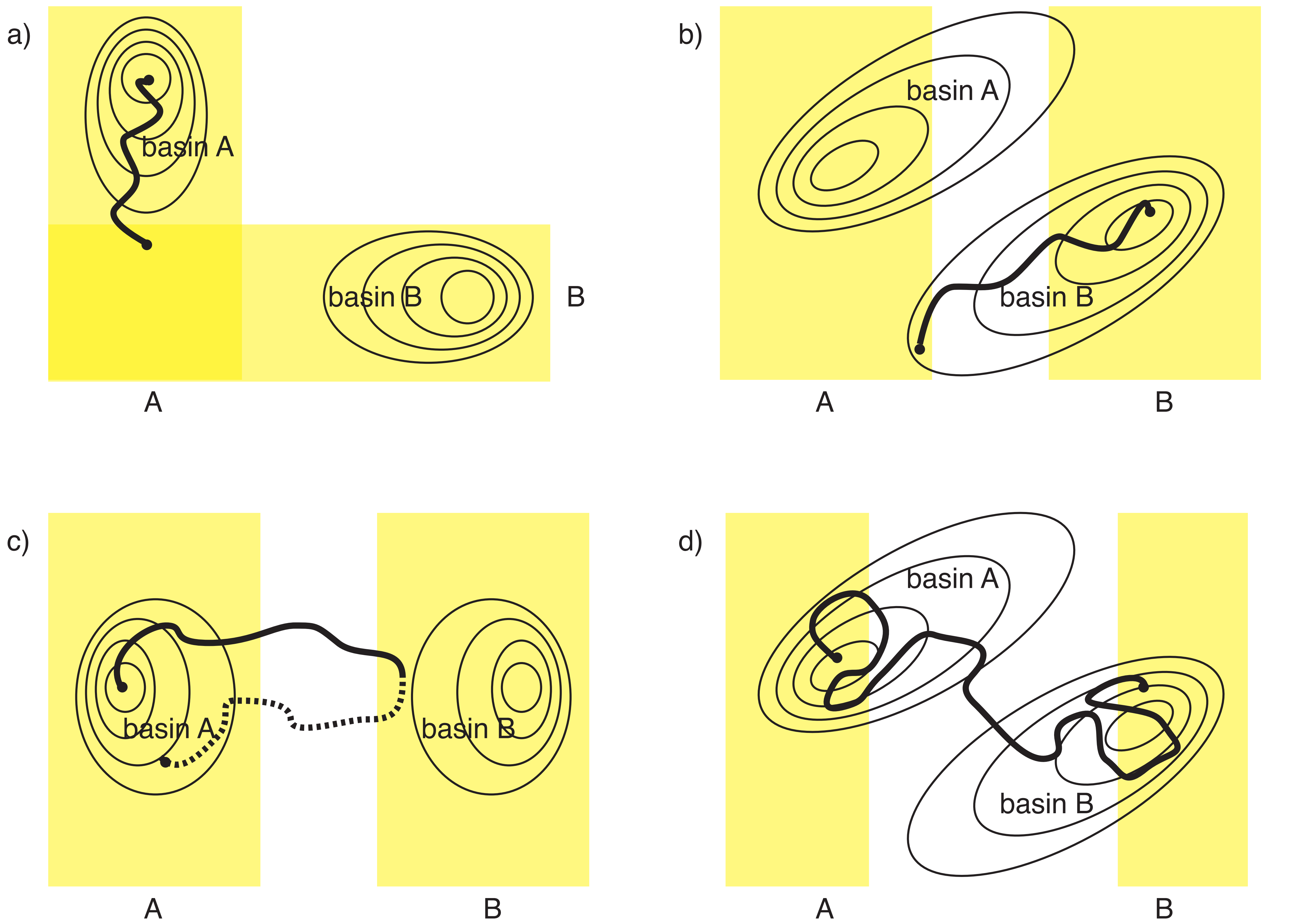
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)})\min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$



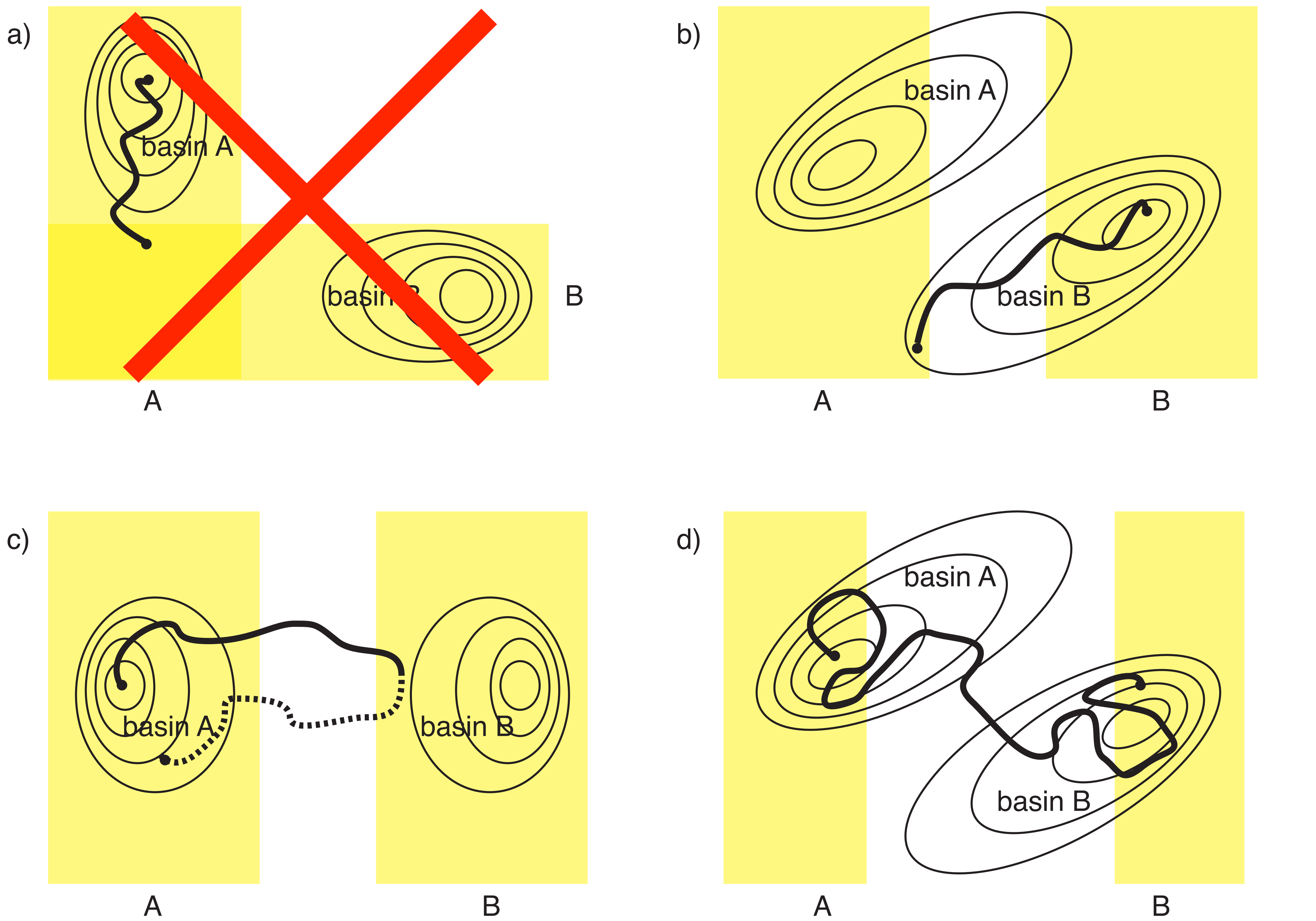
Many shooting variants

PGB and Swenson, Adv. Theor. Simul. 4, 2000237 (2021)

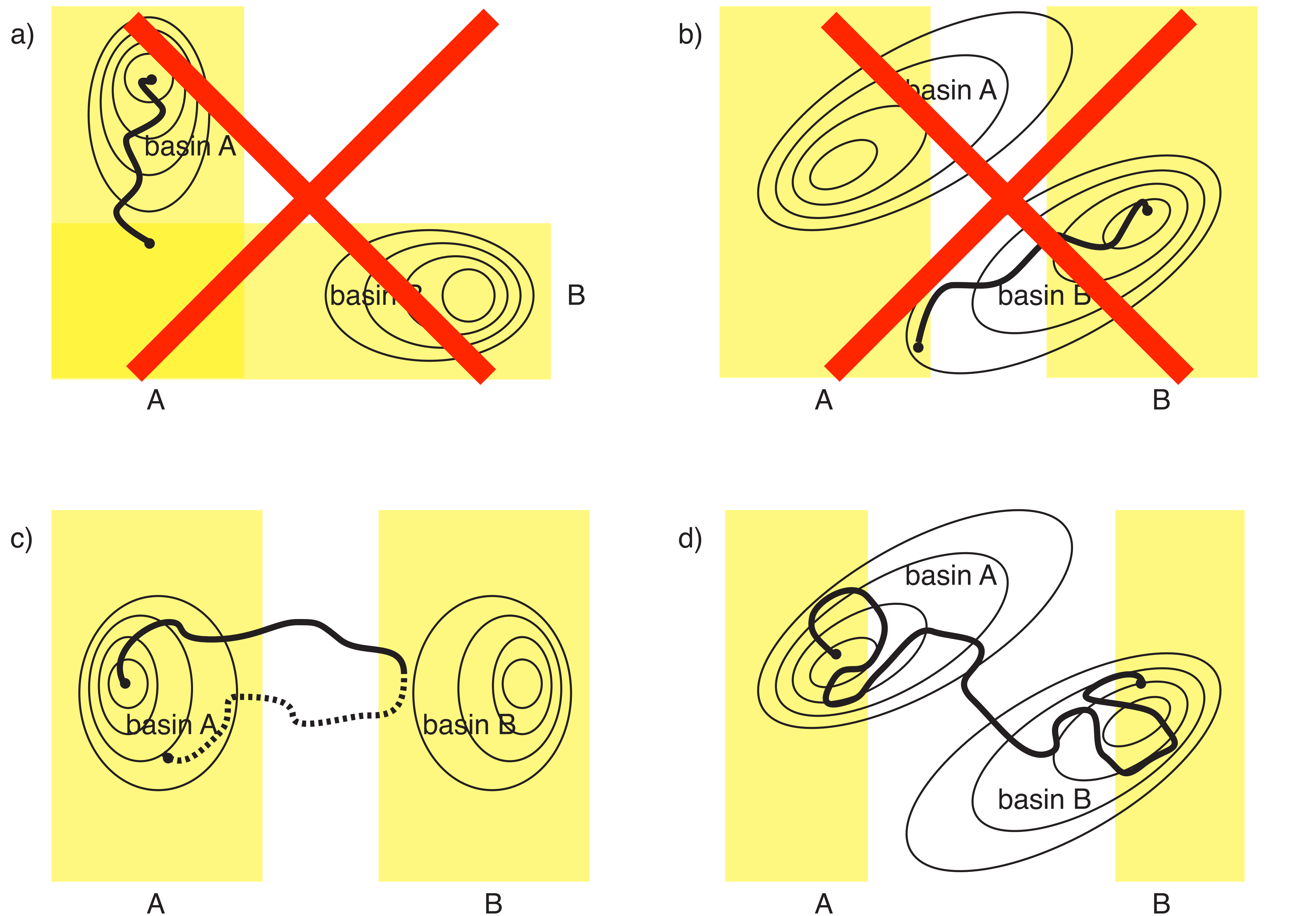
How do we define states?



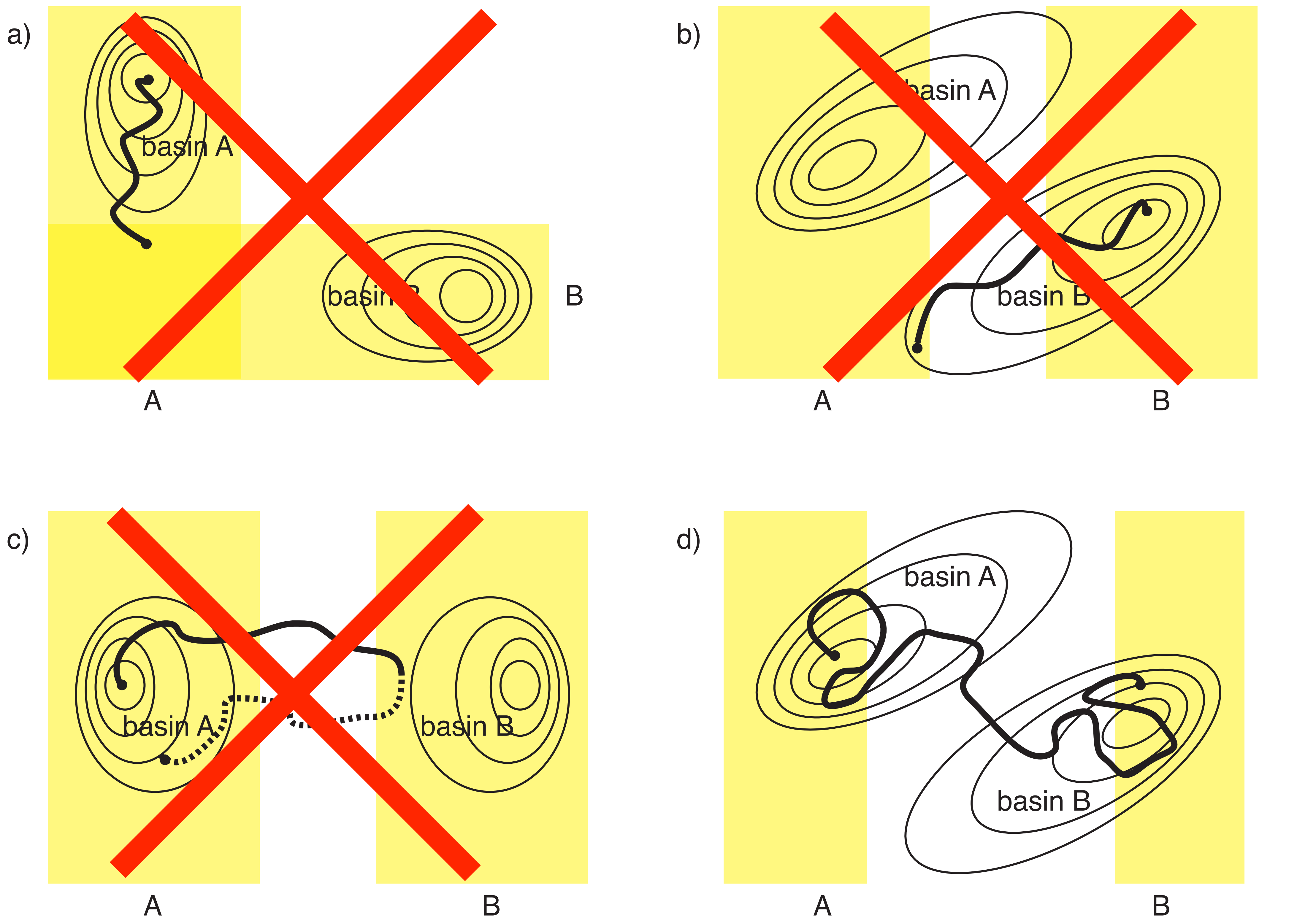
How do we define states?



How do we define states?

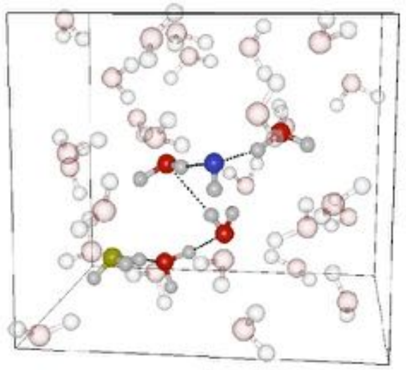


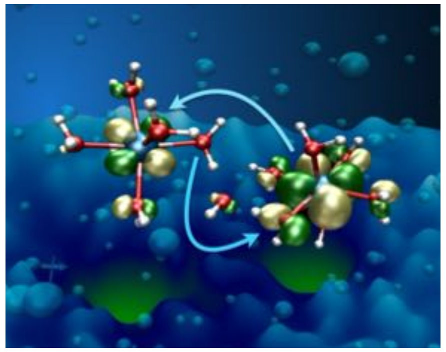
How do we define states?



Selected TPS applications

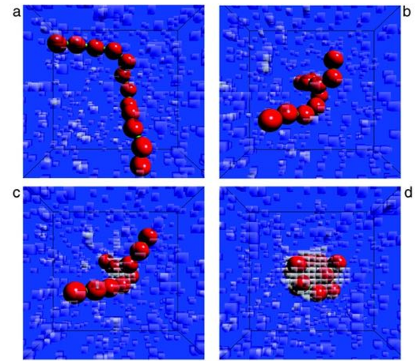
Chemical reactions in solution

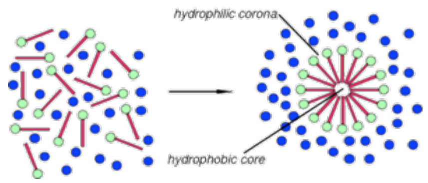




Geissler et al Science 2001; Tiwari and Ensing, Farad Disc 2016; Joswiak et al, PNAS 2017 ….

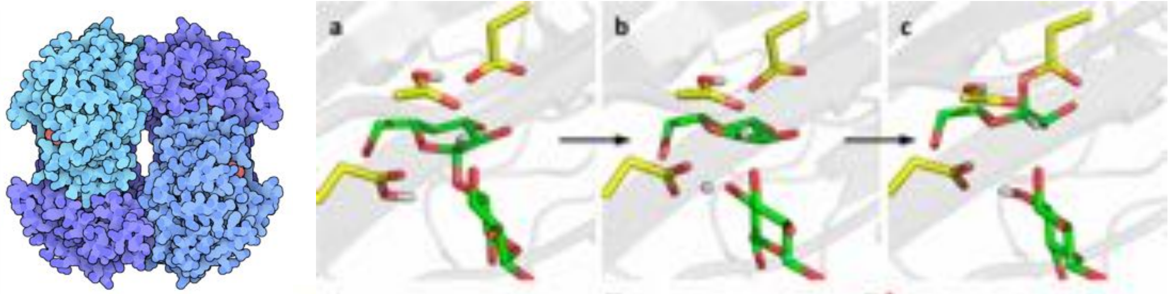
Microphases

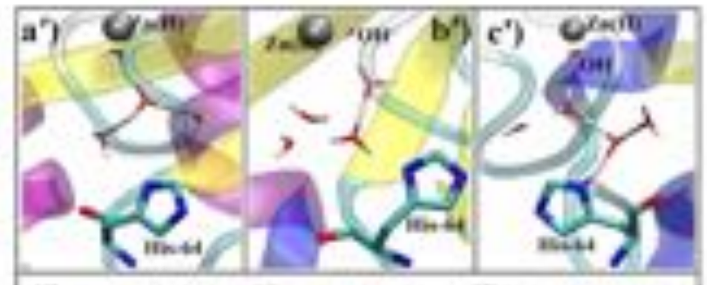


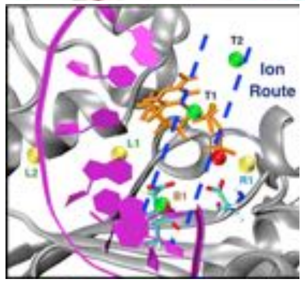


Ten Wolde et al PNAS 2002; Pool & PGB JCP 2007

Enzymatic reactions

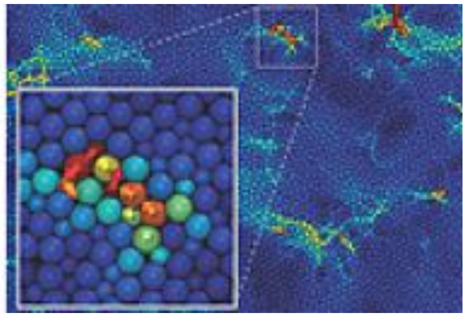


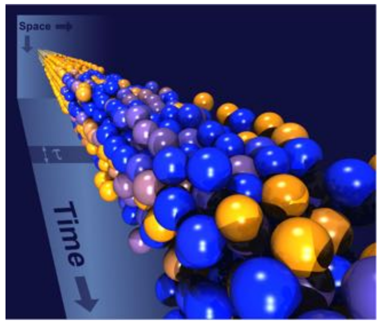




Basner et Schwarz, JACS 2005; Knott et al, JACS 2013;Li et al JACS 2014; Paul and Taraphder, ChemPhysChem 2020; Silveira et al, JPCB 2021;….

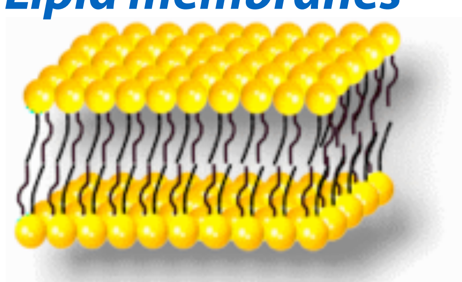
Glasses

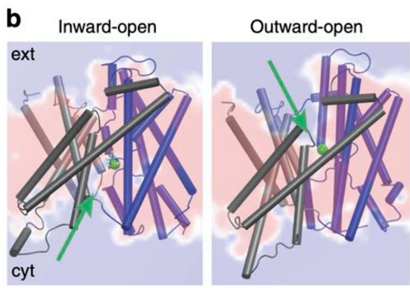


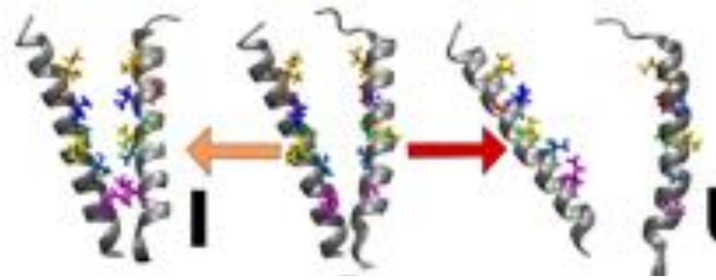


Hedges et al Science 2009; Jack etc al PRL 2011; Turci et al PRX 2017; ….

Lipid membranes

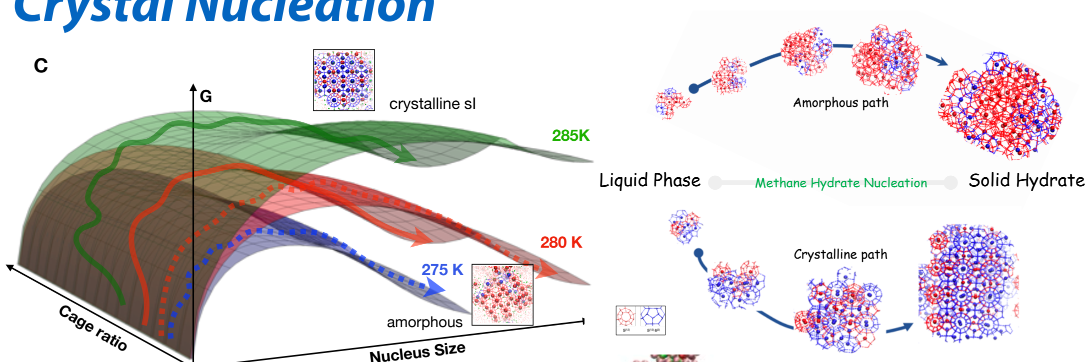


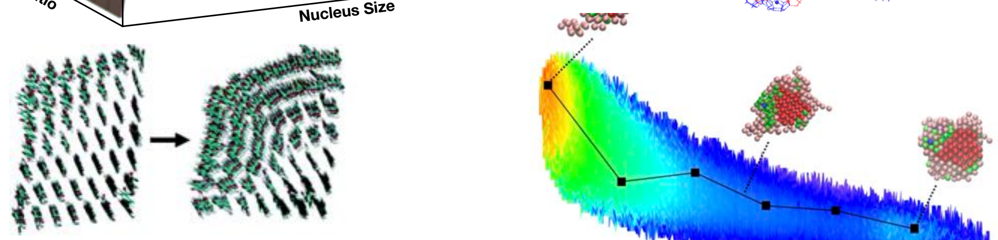




Marti & Csajka 2004; Okazaki et al Nat Comm. 2019 . Domanski, et al PLOS Comput. Biol. 2020,; …..

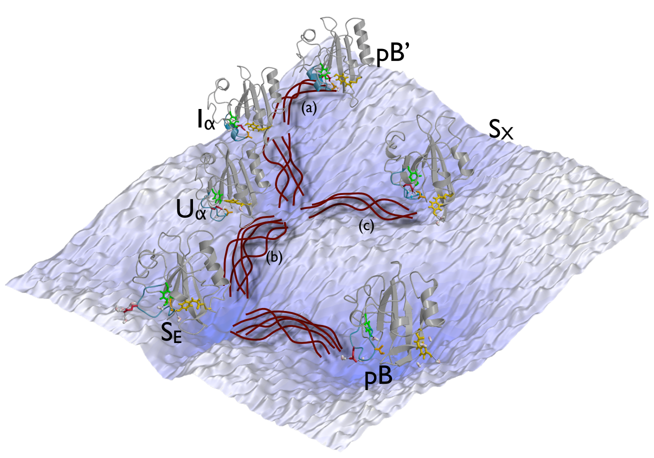
Crystal Nucleation

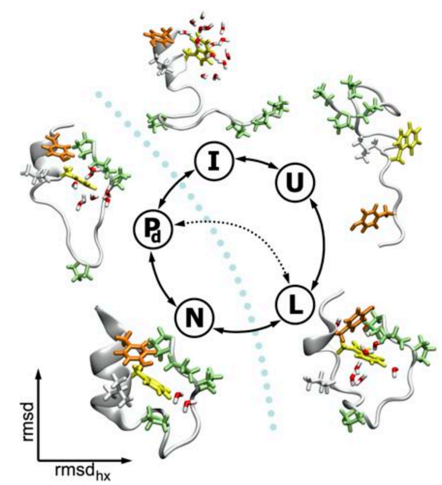


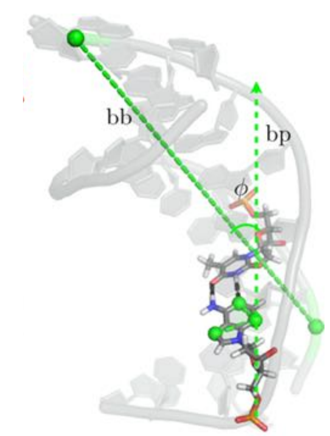


Moroni et al PRL 2005, Bekcham et al JACS 2007l Lechner et al. PRL 2011; Diaz Leines & Rogal JPCB 2018; Arjun et al PNAS 2019….

Biomolecular conformational change

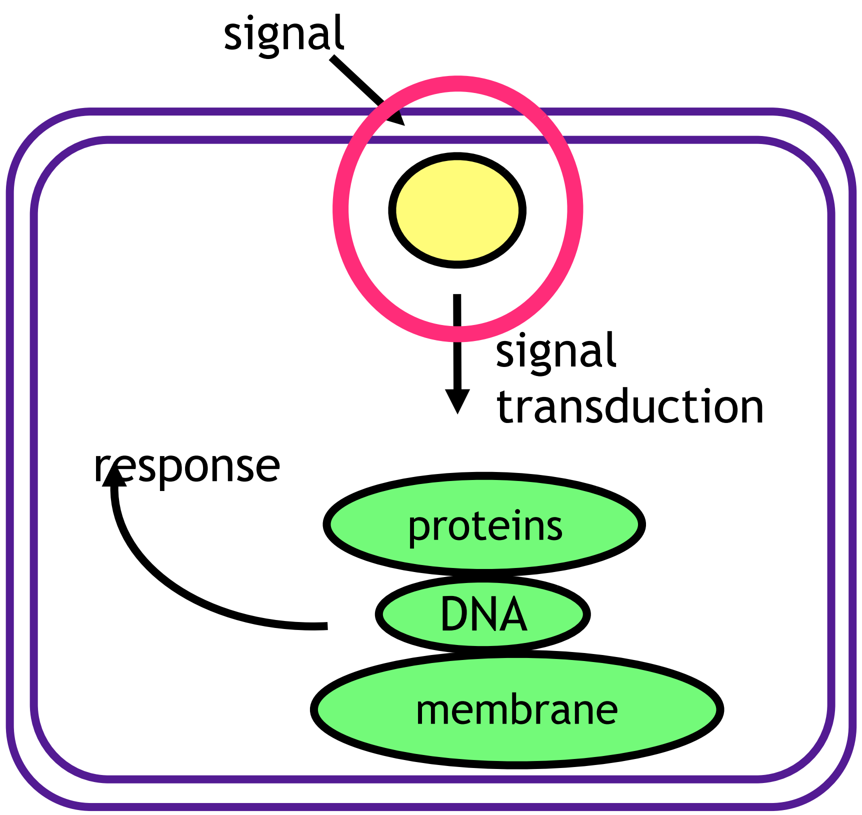


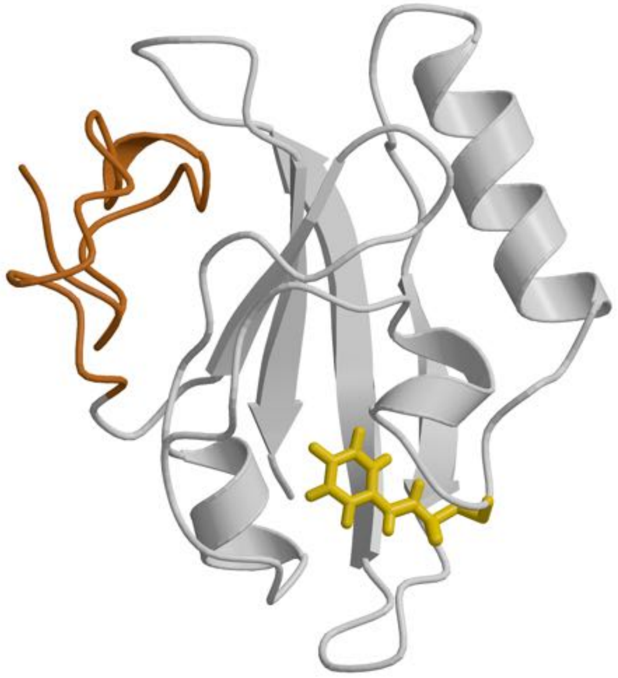




Bolhuis PNAS 2003; Juraszek & Bolhuis 2006; Vreede et al PNAS 2010; Best & Hummer PNAS 2016; Brotzakis & PGB, JPCB 2019, Vreede et al. NAR 2019…….

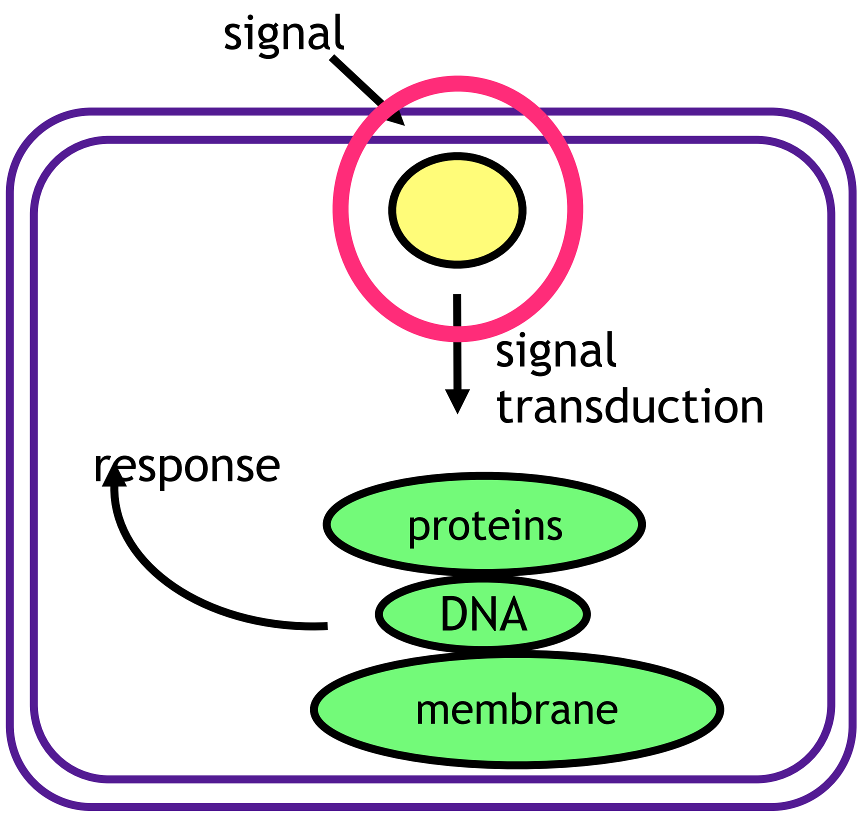
Photoactive yellow protein

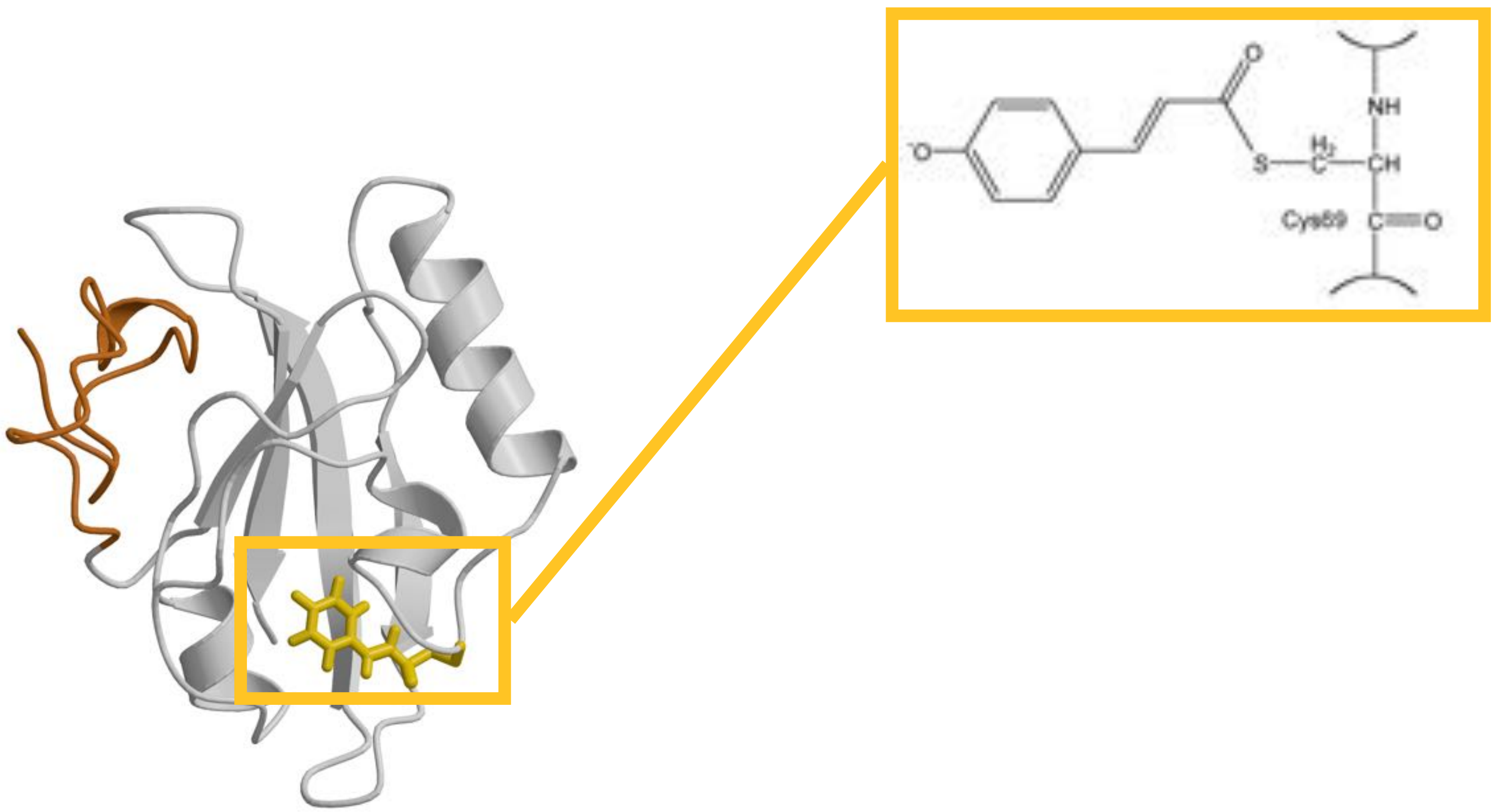




Photoactive yellow protein

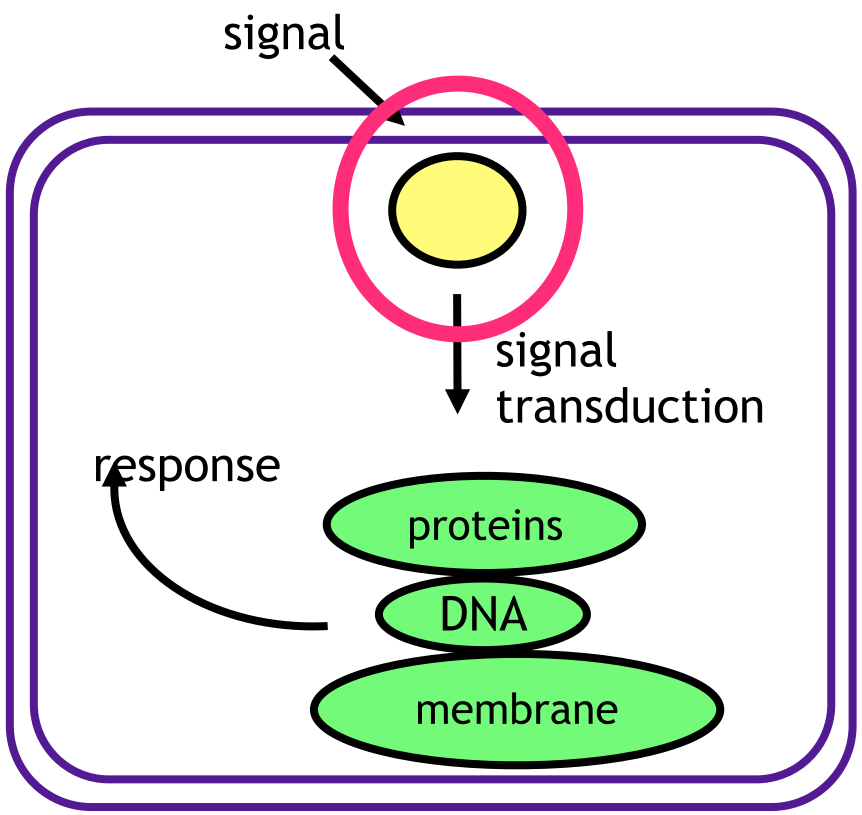
Photoactive yellow protein

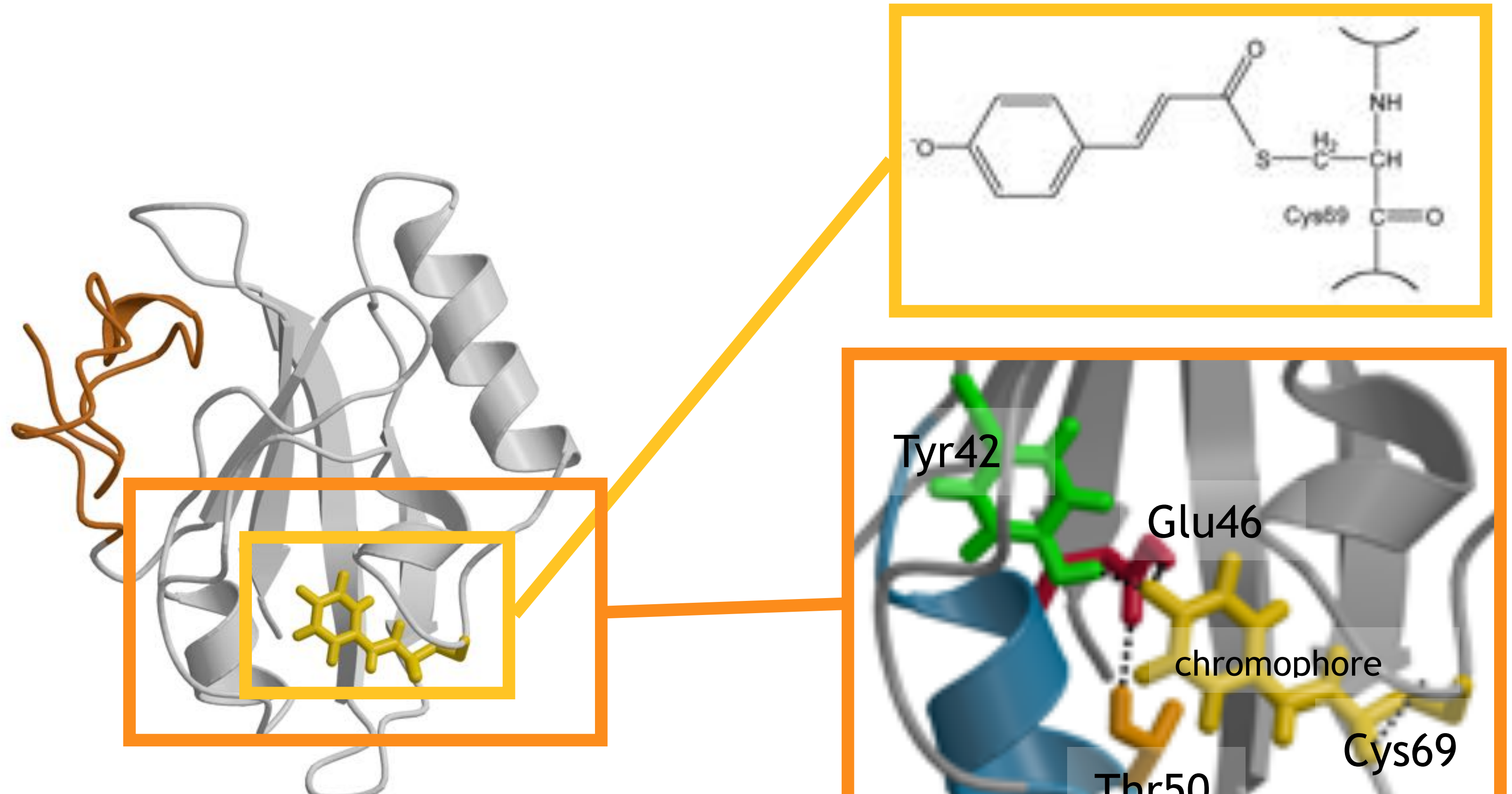




Photoactive yellow protein

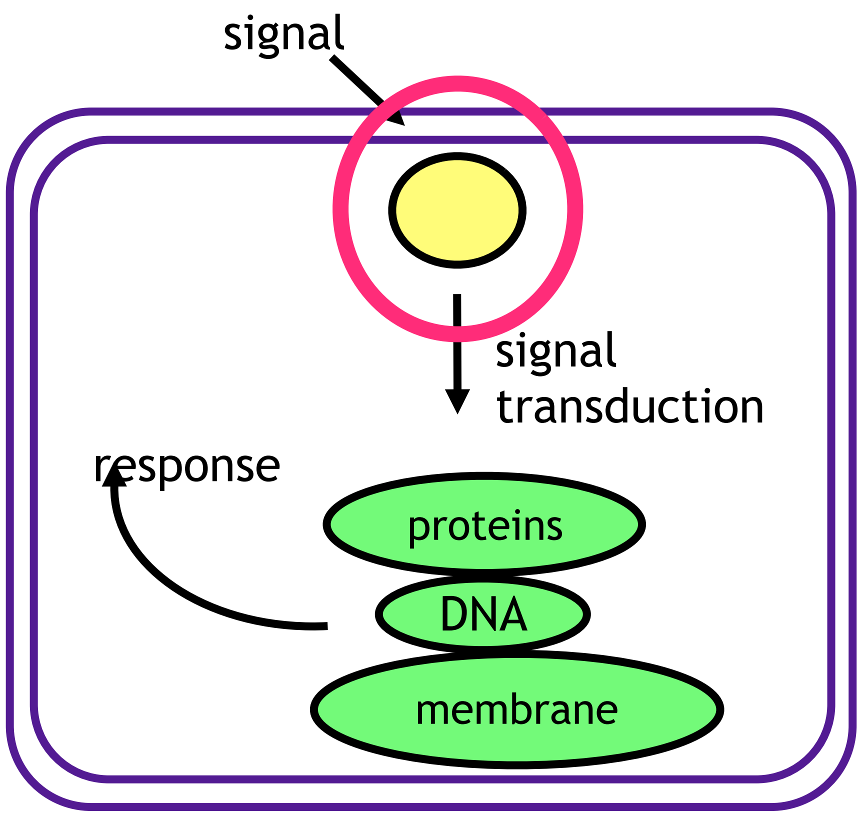
Photoactive yellow protein

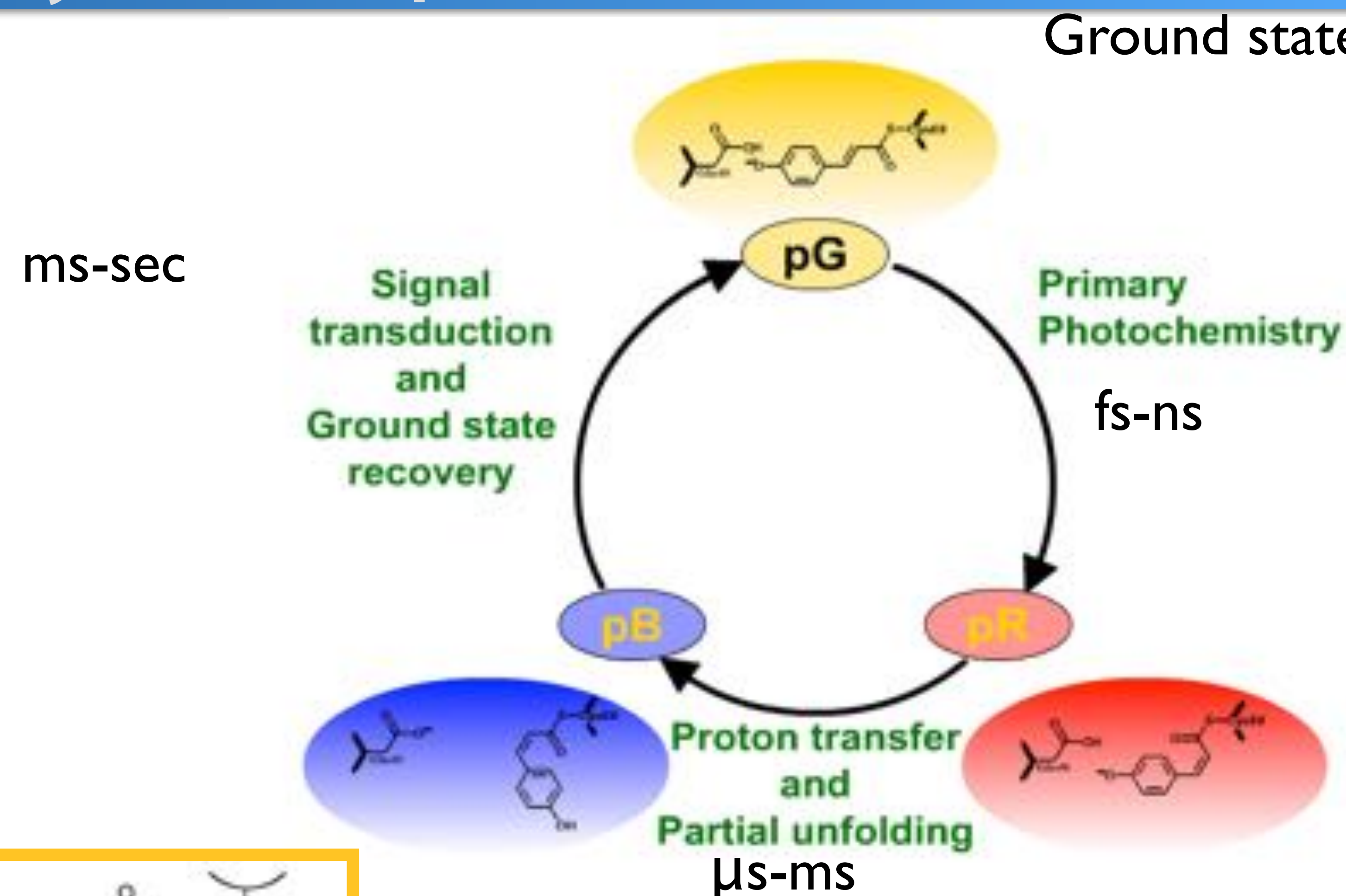


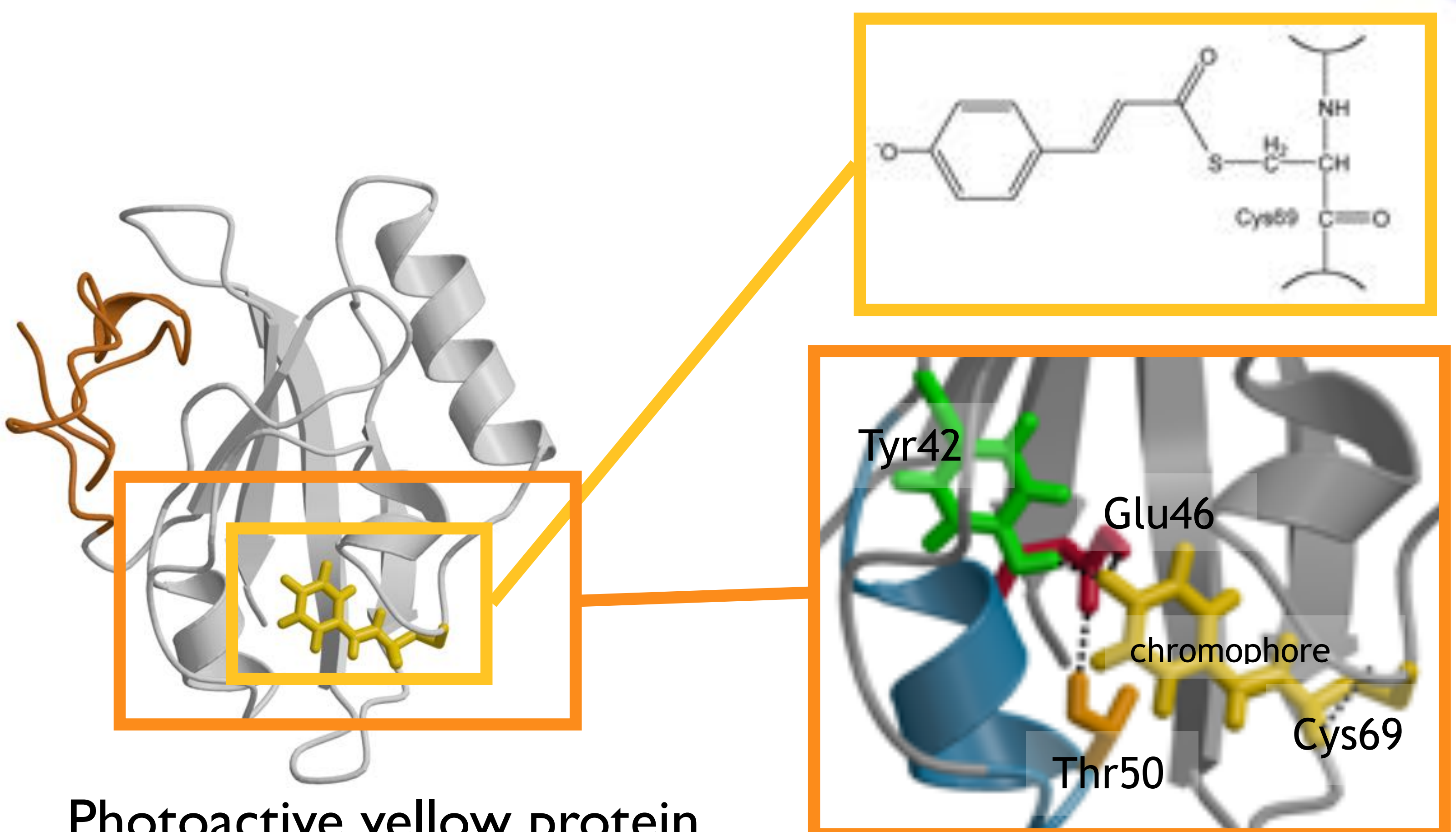


Photoactive yellow protein

Photoactive yellow protein

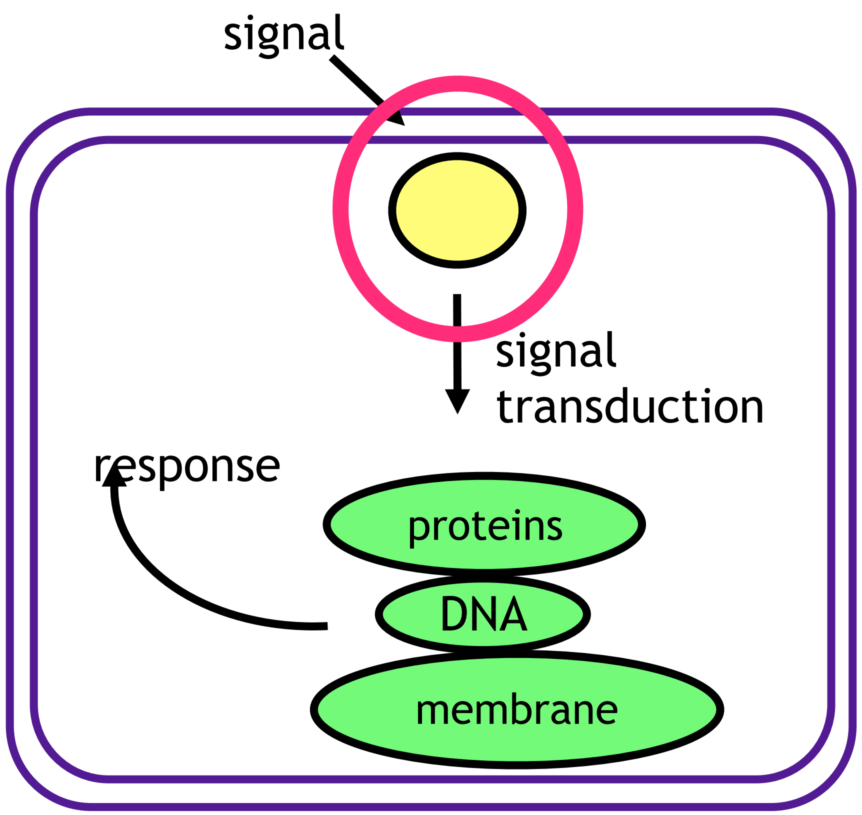


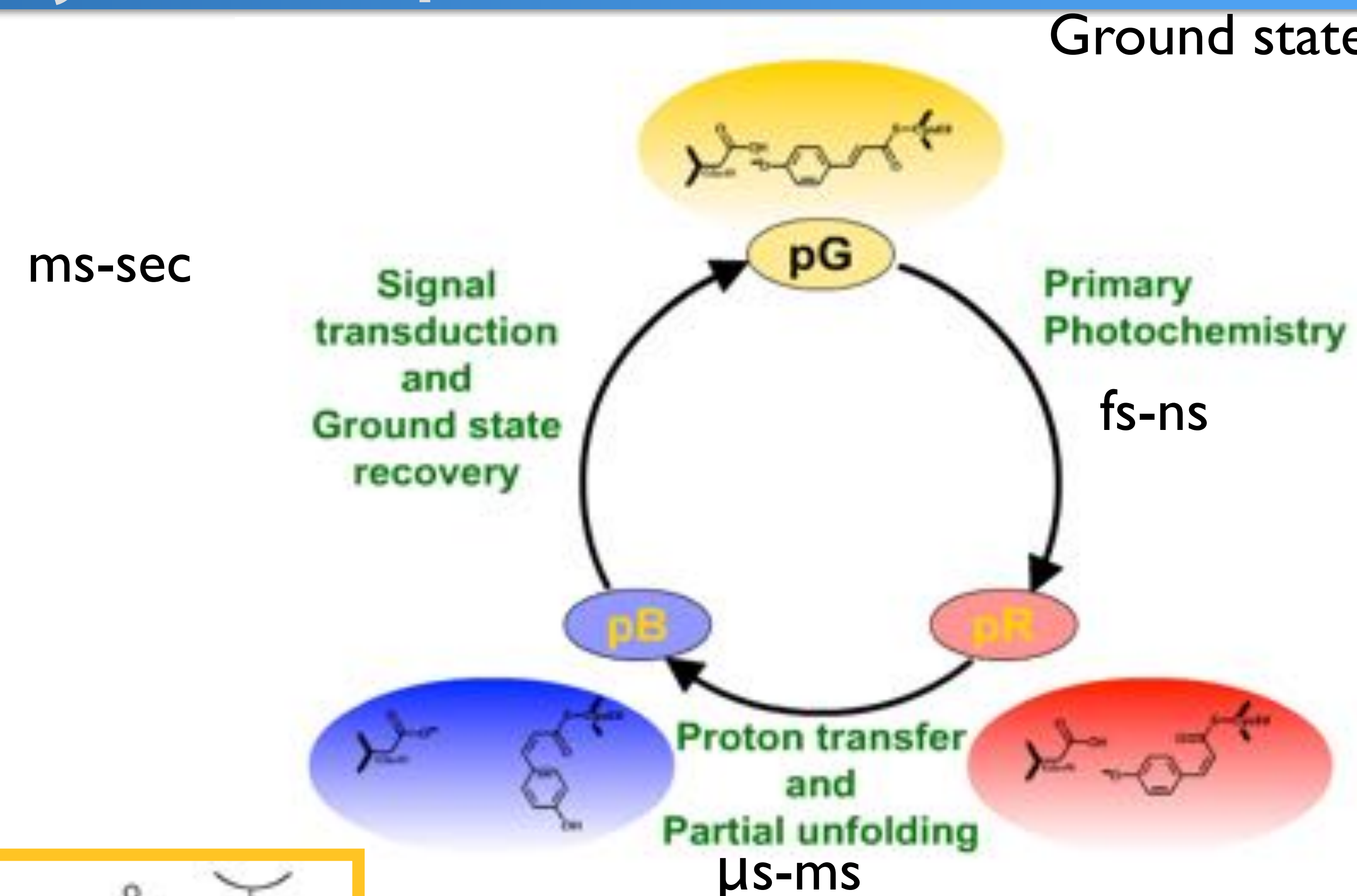


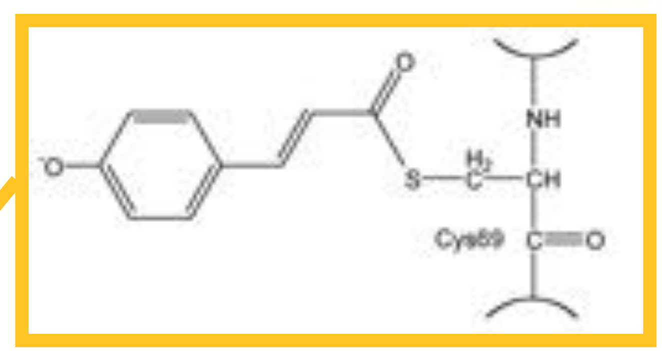


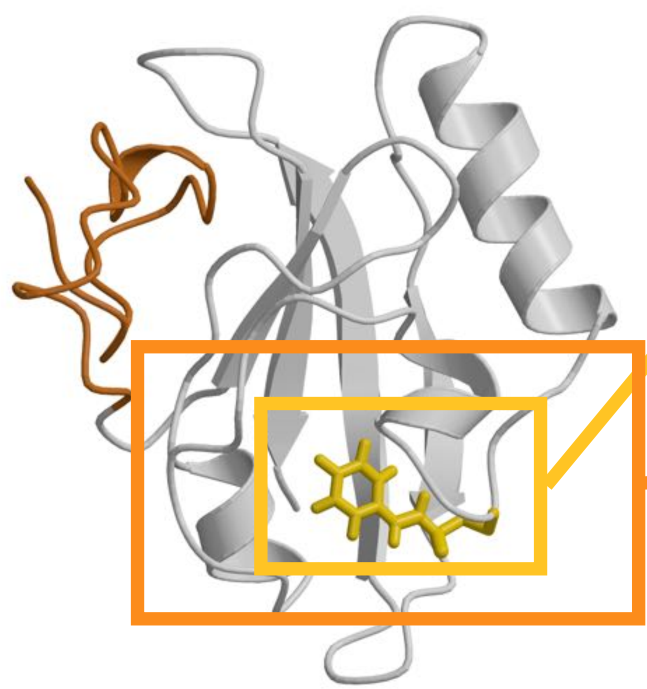
Photoactive yellow protein

Photoactive yellow protein

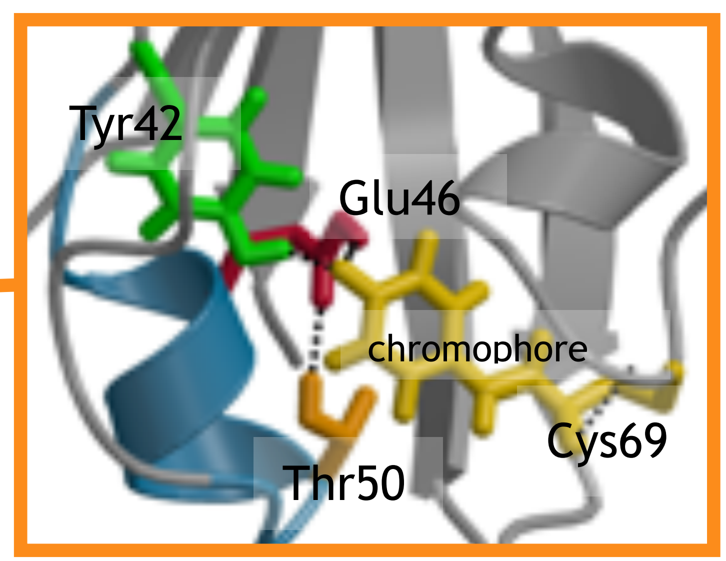








Photoactive yellow protein



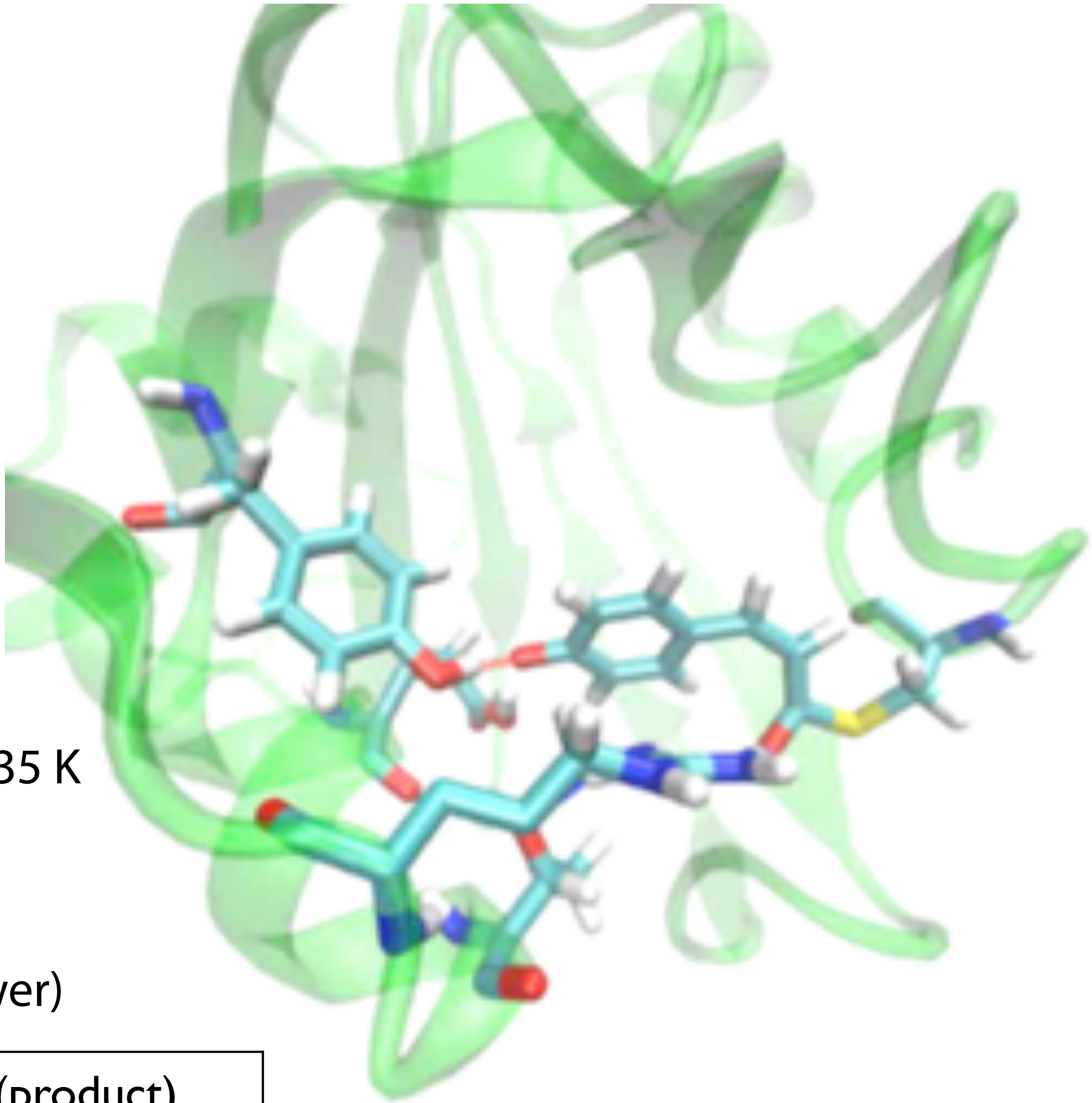
Question: What is the mechanism for amplifying signal?

We studied 2 steps:
1) proton transfer
2) partial unfolding

Path sampling of proton transfer

System

- 28244 atoms
- CPMD/QMMM
- BLYP functional
- Electronic mass 750 au
- QM region: pCA, Glu46,Tyr42, Thr50, Arg52
- Gromos96 force field



TPS settings

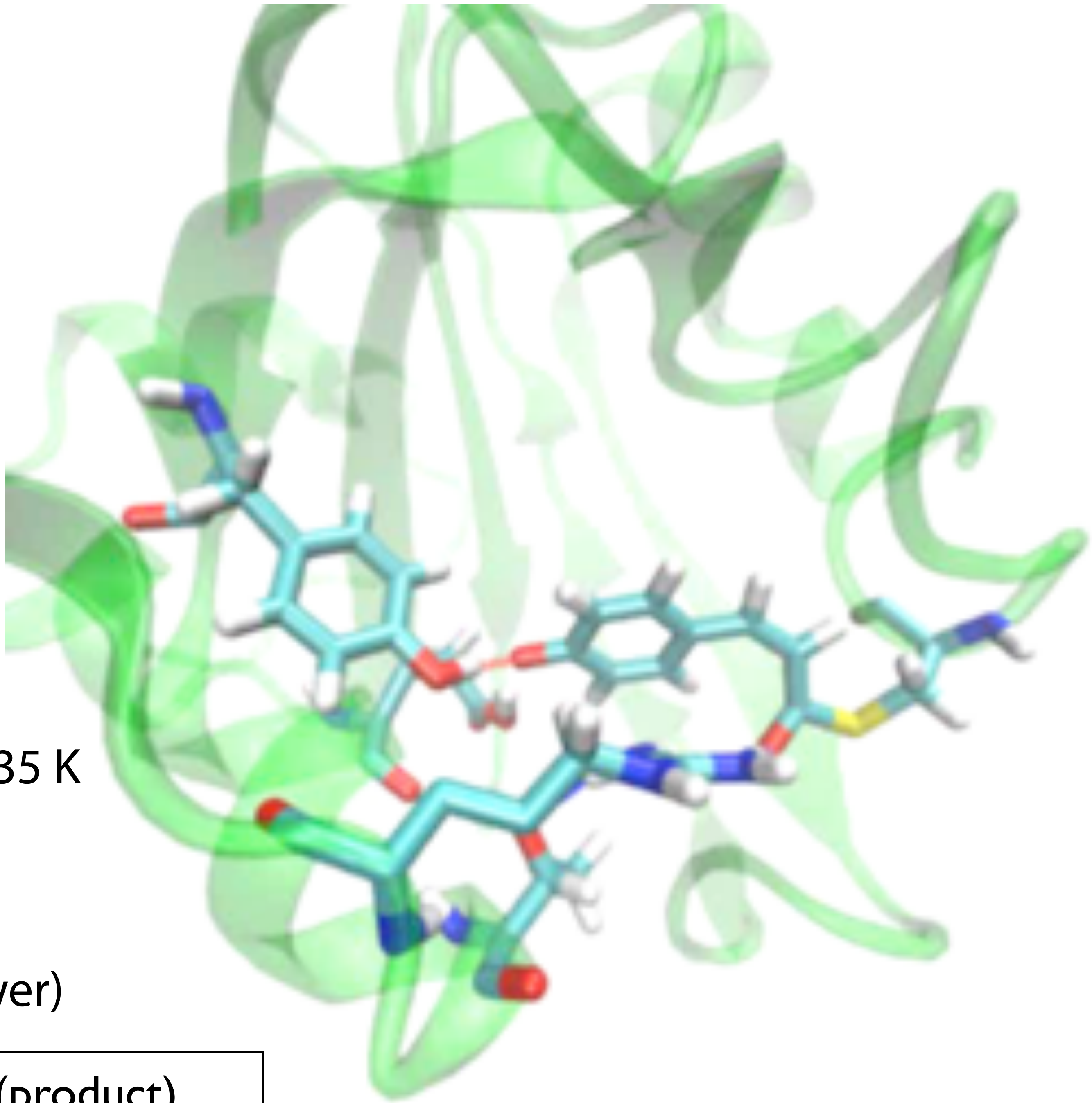
- two way shooting, perturbation temp 35 K
- 160 paths/ 50% acceptance
- average path length 0.5-1.5 ps
- reaction time microseconds (10^6 x slower)

stable states	pR (reaction)	pB' (product)
pCA-Glu46(H)	> 1.60 A	< 0.98 A
OX2-Tyr42	> 3.70 A	< 1.80 A
OX1-Tyr42	> 5.30 A	< 1.80 A

Path sampling of proton transfer

System

- 28244 atoms
- CPMD/QMMM
- BLYP functional
- Electronic mass 750 au
- QM region: pCA, Glu46,Tyr42, Thr50, Arg52
- Gromos96 force field



TPS settings

- two way shooting, perturbation temp 35 K
- 160 paths/ 50% acceptance
- average path length 0.5-1.5 ps
- reaction time microseconds (10^6 x slower)

stable states	pR (reaction)	pB' (product)
pCA-Glu46(H)	> 1.60 A	< 0.98 A
OX2-Tyr42	> 3.70 A	< 1.80 A
OX1-Tyr42	> 5.30 A	< 1.80 A

Transition path sampling of partial

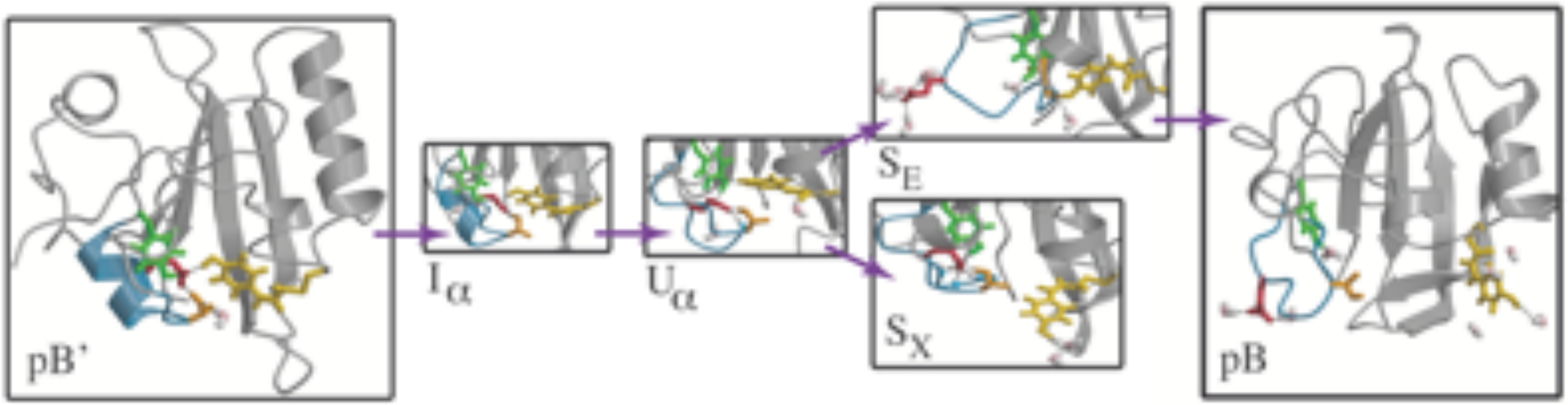


Table 1. Statistics of the TPS ensembles. The average path length is a weighted average over the whole ensemble. Decorrelated pathways have lost the memory of the previous decorrelated pathway. The aggregate time is the ensemble aggregate length

	$pB' - I_\alpha$	$U_\alpha - S_E$	$U_\alpha - S_X$	$S_E - pB$
acceptance	41%	25%	38%	44%
avg. path length	105 ps	1.8 ns	1.5 ns	1.7 ns
accepted paths	3847	305	584	311
decorr. paths	180	18	7	29
aggregate time (μs)	1.0	2.3	2.3	1.2

Vreede, Juraszek, PGB, PNAS ***107*** *2397 (2010)*

Transition path sampling of partial

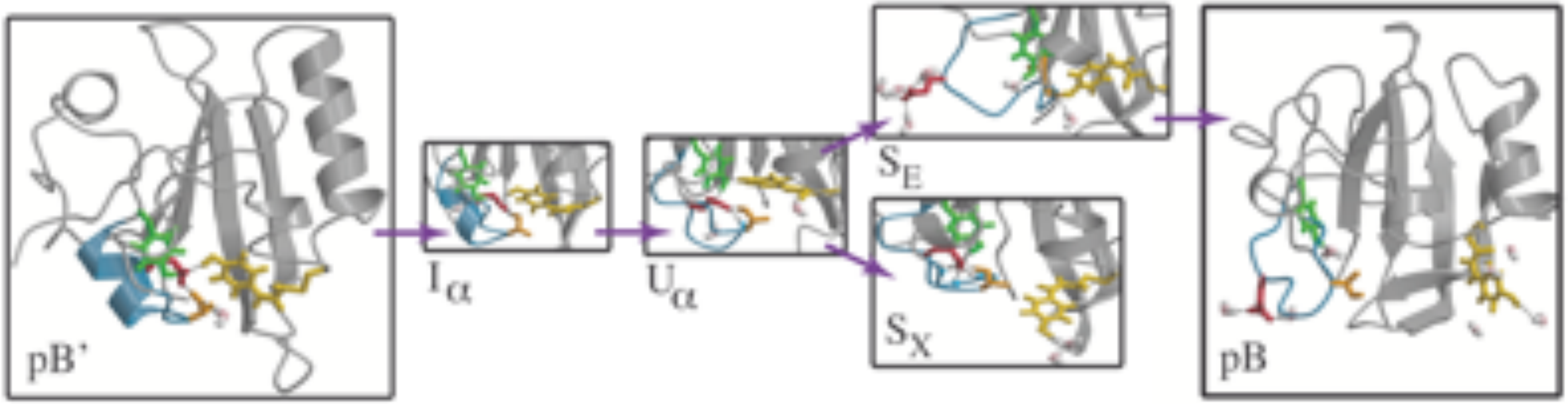
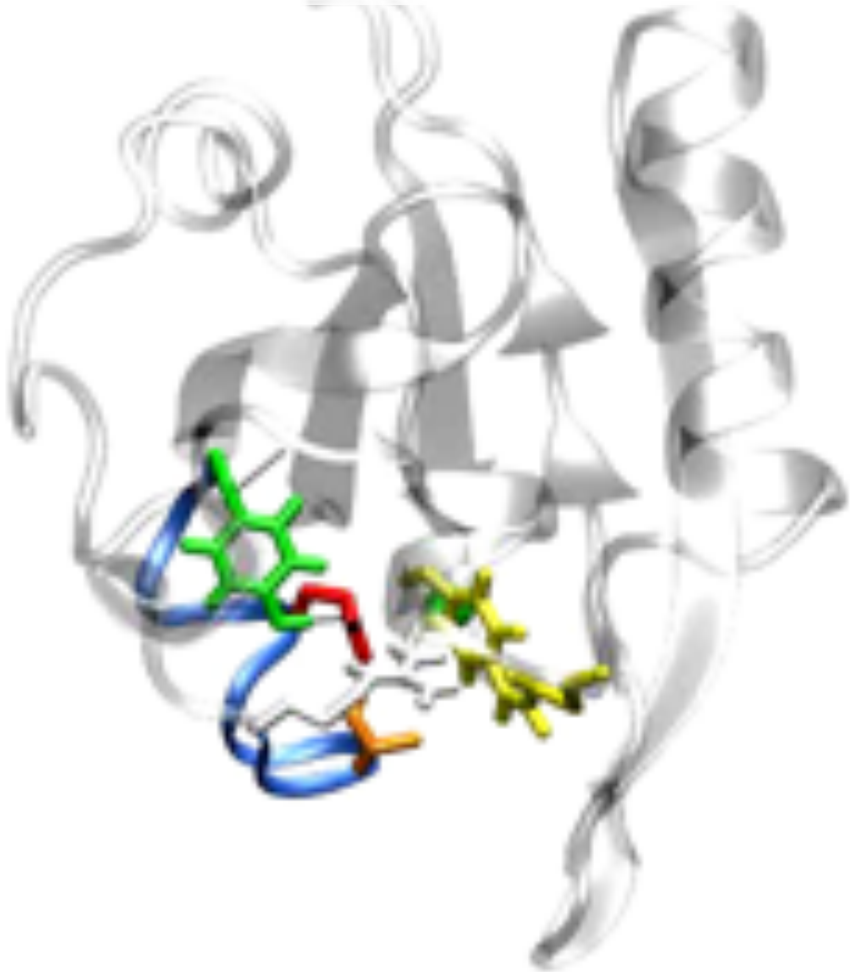


Table 1. Statistics of the TPS ensembles. The average path length is a weighted average over the whole ensemble. Decorrelated pathways have lost the memory of the previous decorrelated pathway. The aggregate time is the ensemble aggregate length

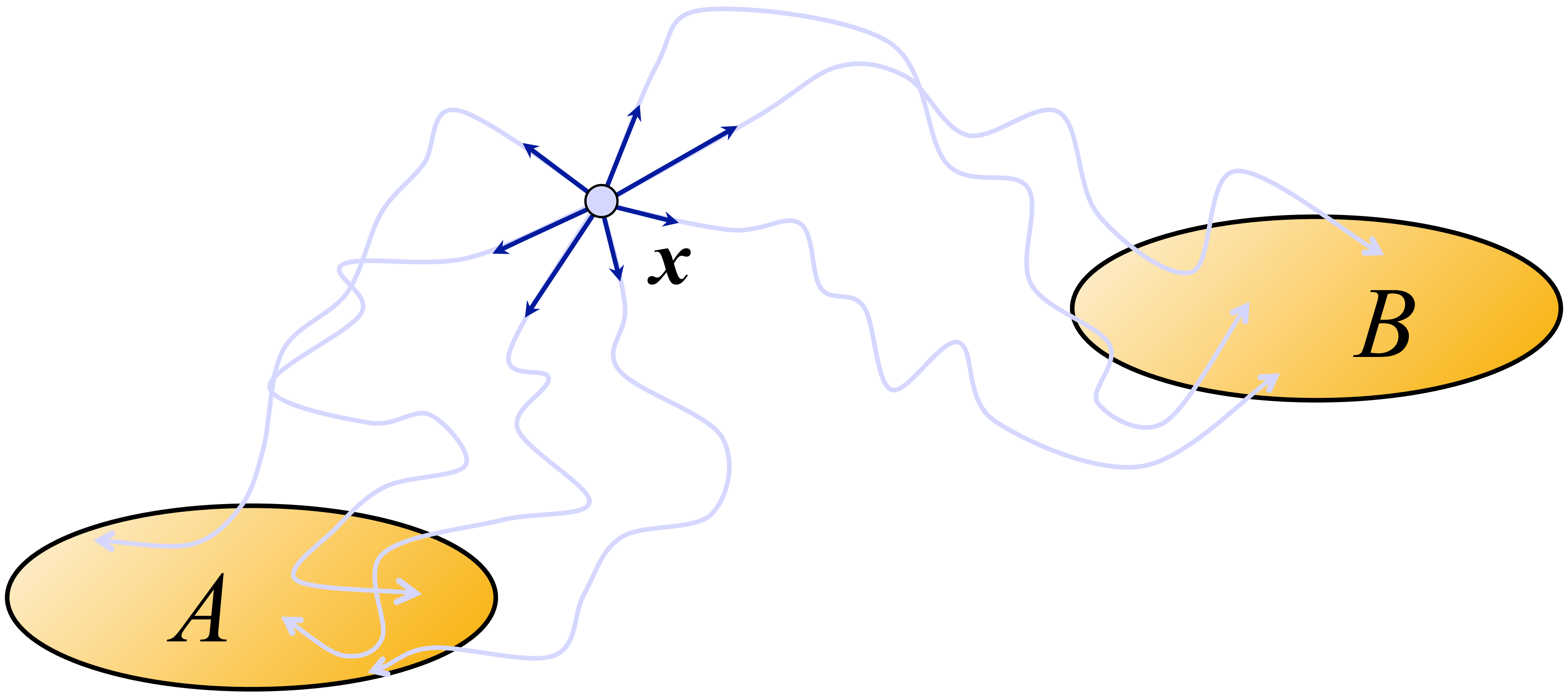
	$pB' - I_\alpha$	$U_\alpha - S_E$	$U_\alpha - S_X$	$S_E - pB$
acceptance	41%	25%	38%	44%
avg. path length	105 ps	1.8 ns	1.5 ns	1.7 ns
accepted paths	3847	305	584	311
decorr. paths	180	18	7	29
aggregate time (μs)	1.0	2.3	2.3	1.2



Vreede, Juraszek, PGB, PNAS ***107*** *2397 (2010)*

Transition states by committor

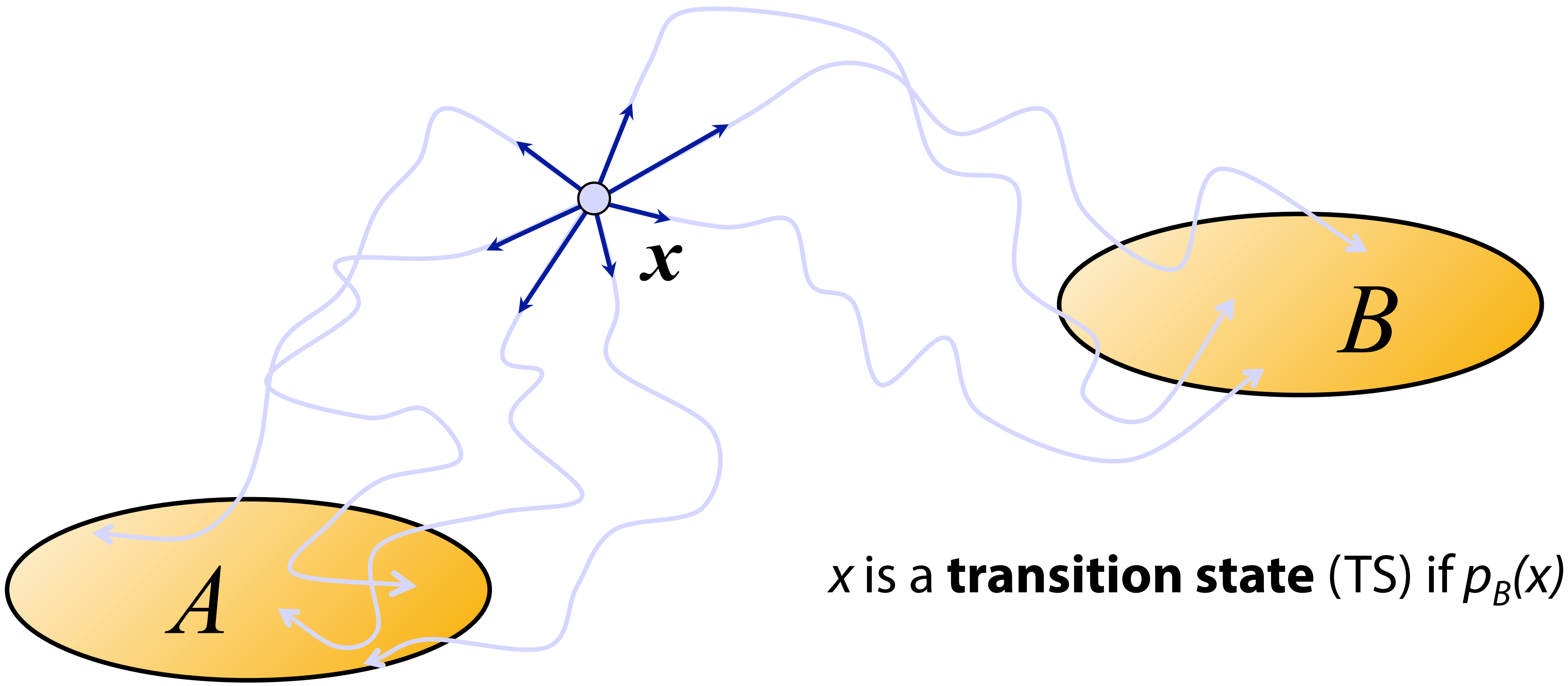
$p_B(x) =$ probability that a trajectory initiated at configuration *x* relaxes into *B*



L. Onsager, Phys. Rev. **54**, 554 (1938). M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991) V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Transition states by committor

$p_B(x) =$ probability that a trajectory initiated at configuration *x* relaxes into *B*

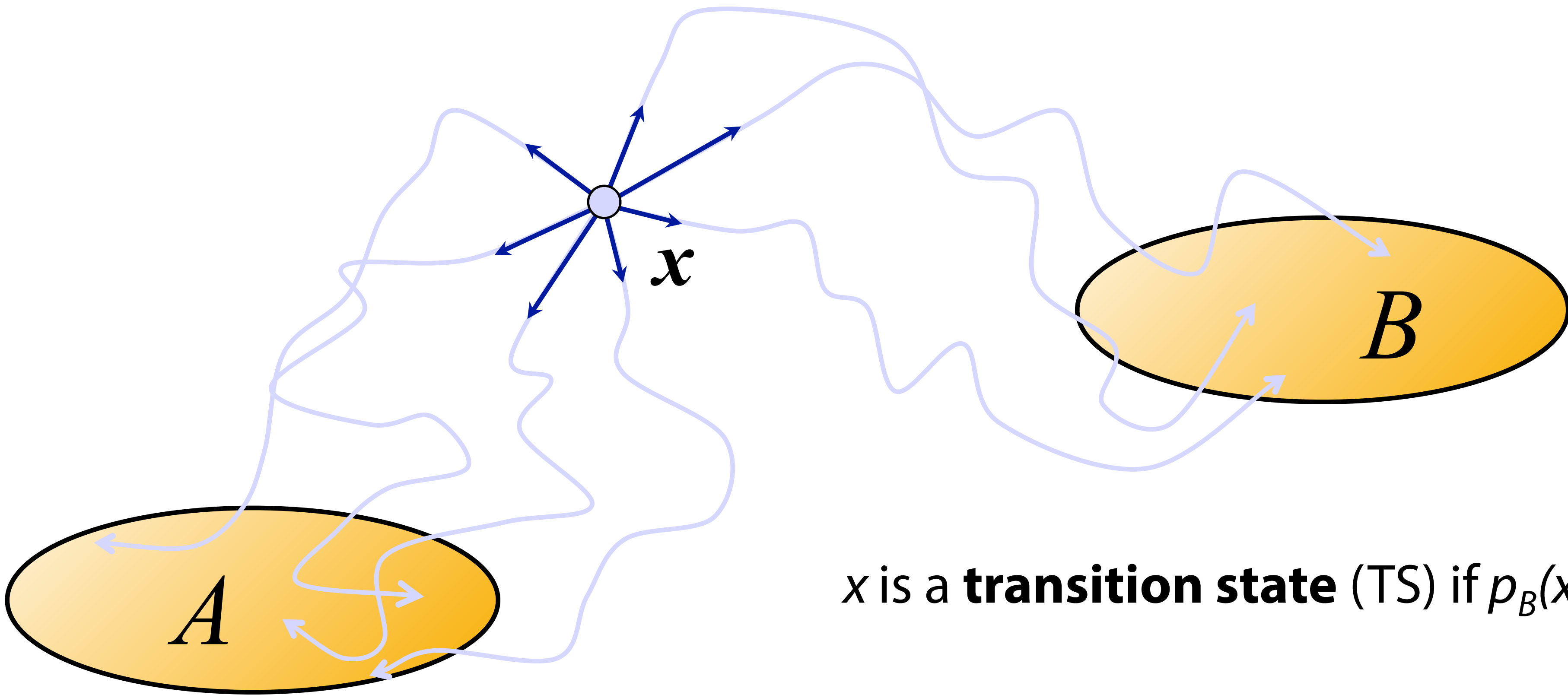


x is a **transition state** (TS) if $p_B(x) = p_A(x) = 0.5$

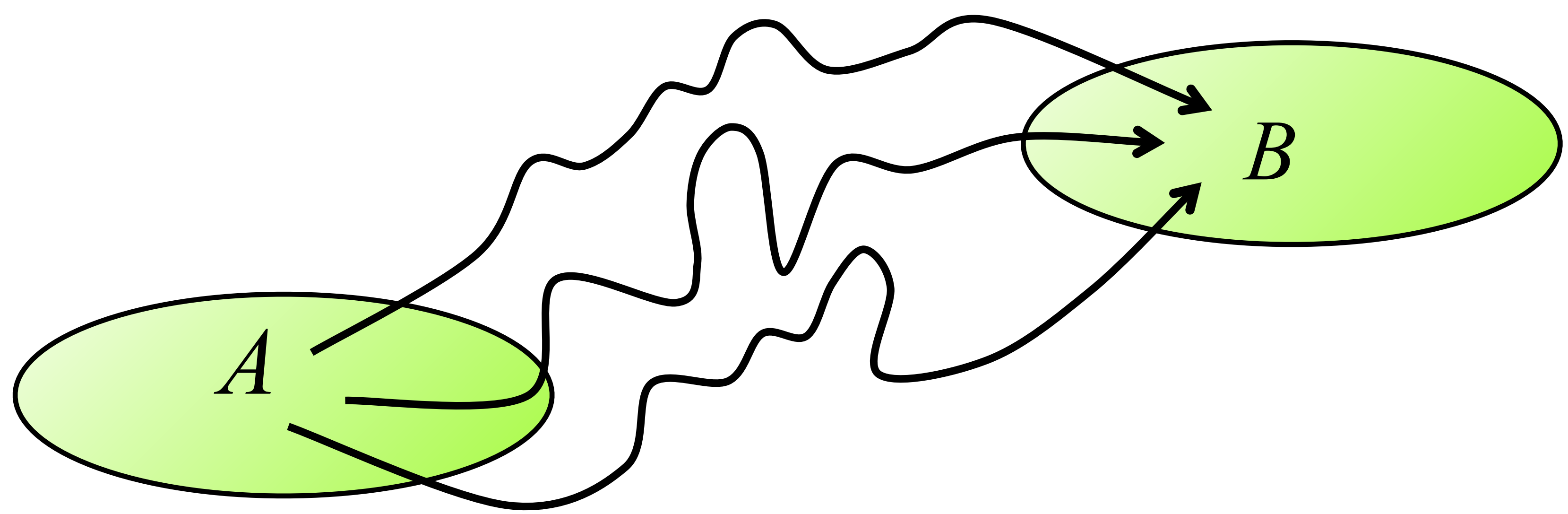
L. Onsager, Phys. Rev. **54**, 554 (1938). M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991) V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Transition states by committor

$p_B(x) =$ probability that a trajectory initiated at configuration *x* relaxes into *B*



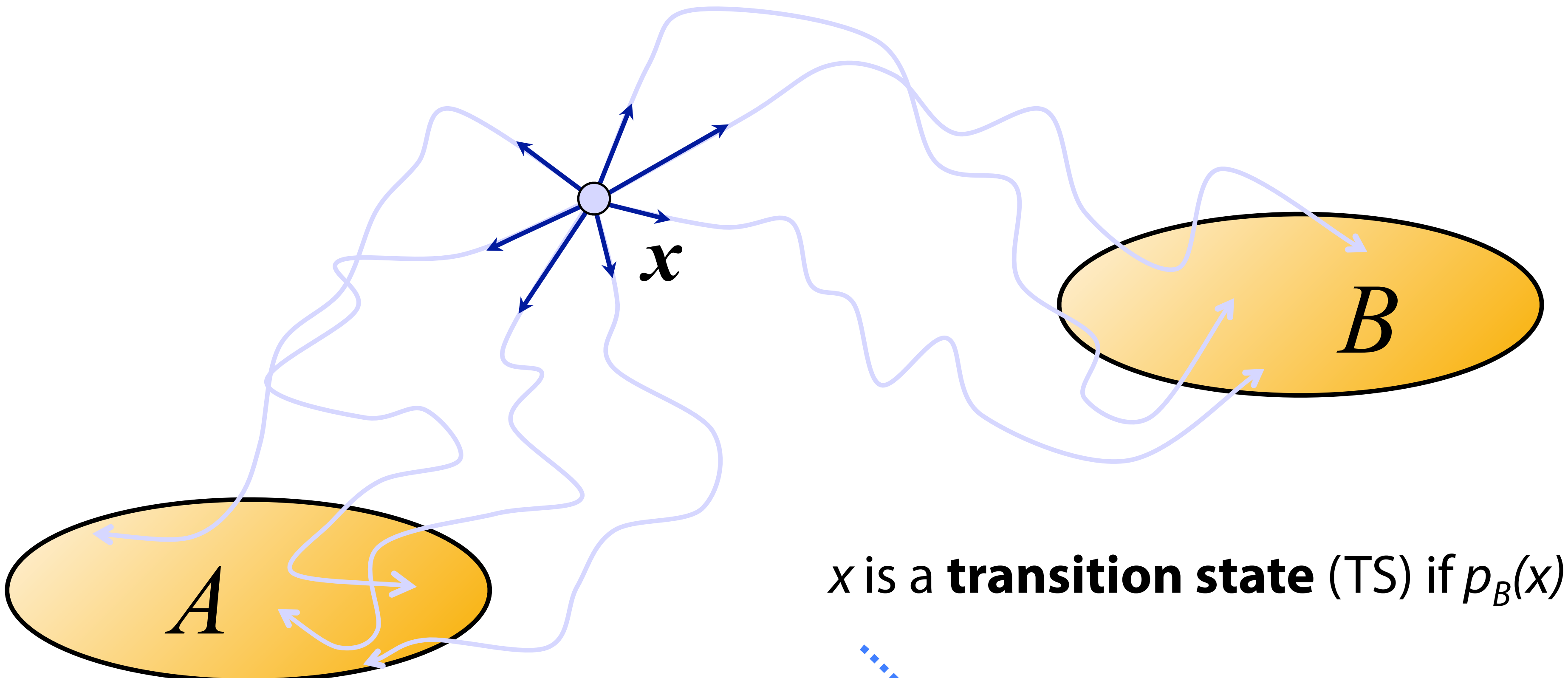
x is a **transition state** (TS) if $p_B(x) = p_A(x) = 0.5$



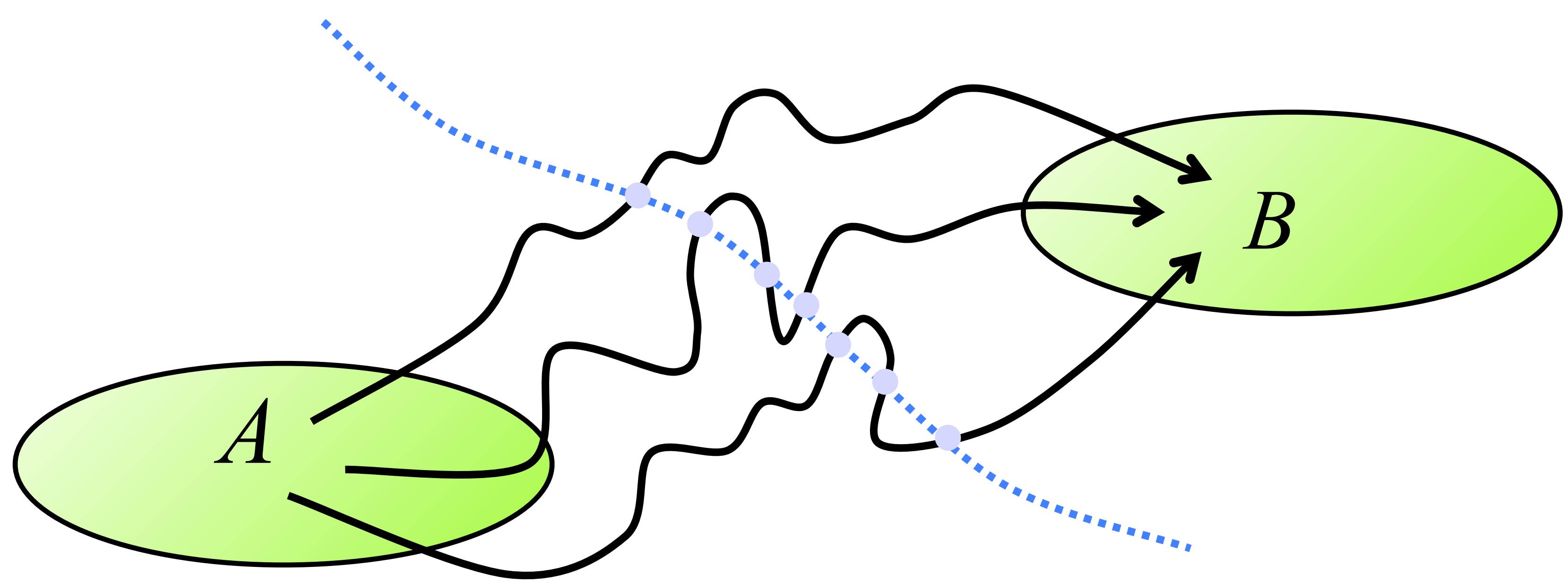
L. Onsager, Phys. Rev. **54**, 554 (1938). M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991) V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Transition states by committor

$p_B(x) =$ probability that a trajectory initiated at configuration *x* relaxes into *B*



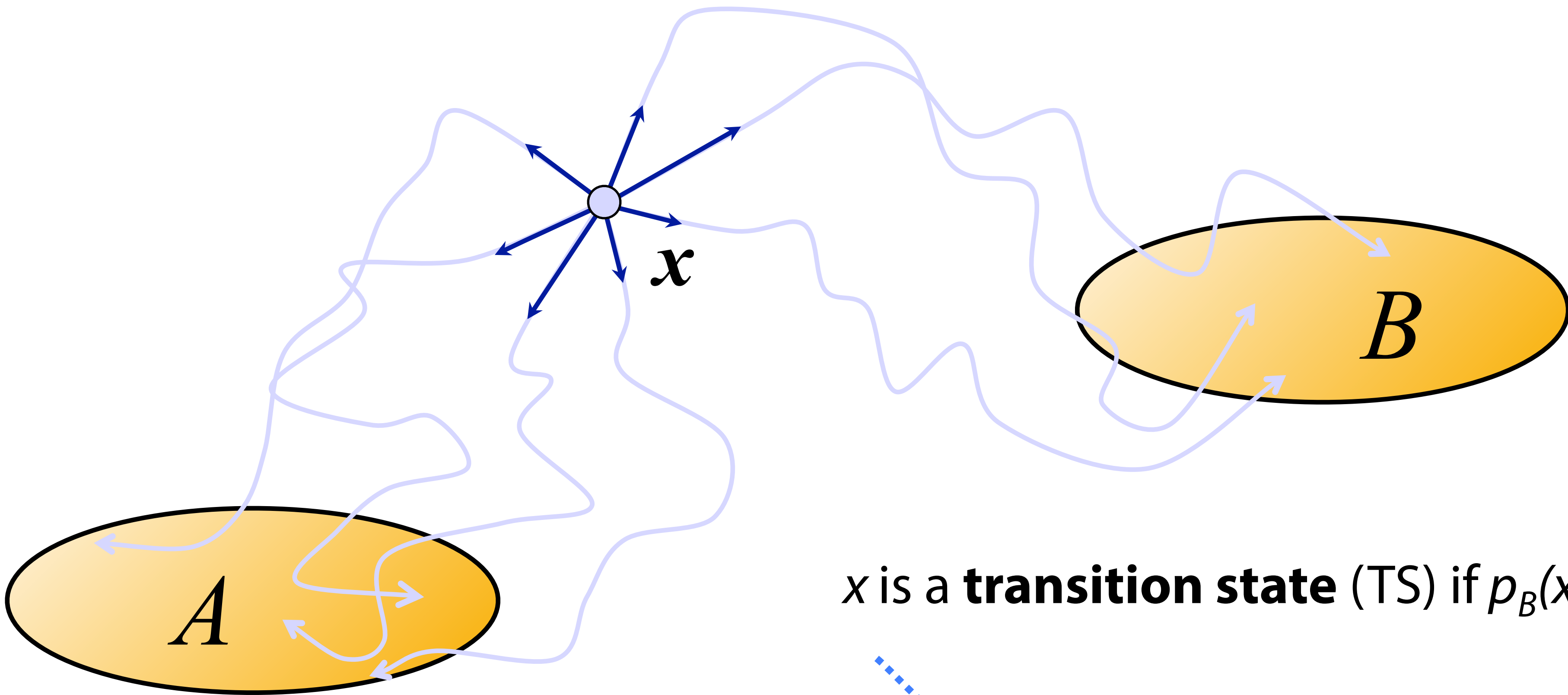
x is a **transition state** (TS) if $p_B(x) = p_A(x) = 0.5$



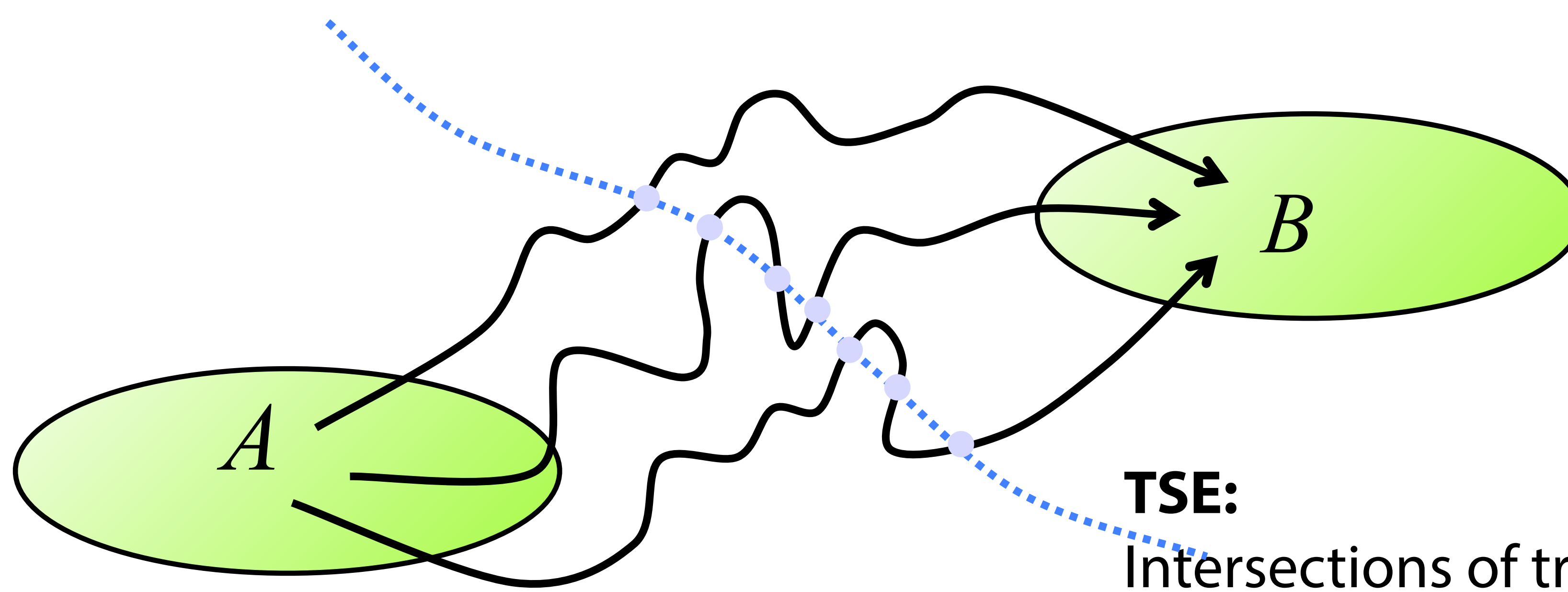
L. Onsager, Phys. Rev. **54**, 554 (1938). M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991) V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Transition states by committor

$p_B(x) =$ probability that a trajectory initiated at configuration *x* relaxes into *B*



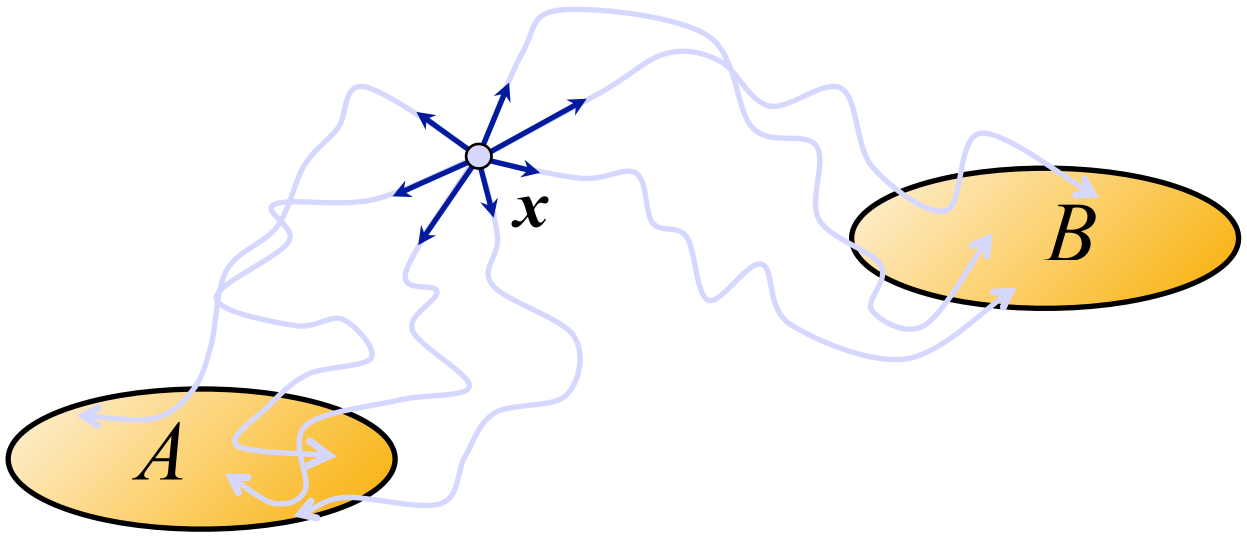
x is a **transition state** (TS) if $p_B(x) = p_A(x) = 0.5$



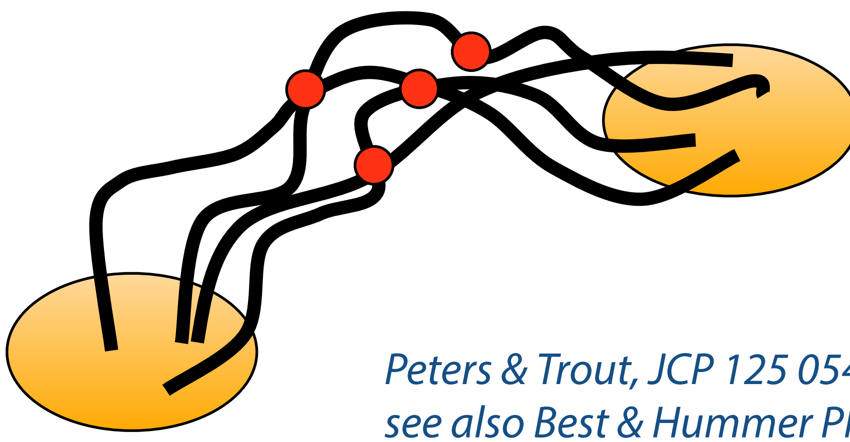
TSE:
Intersections of transition pathways with the p_B=1/2 surface

L. Onsager, Phys. Rev. **54**, 554 (1938). M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991) V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Reaction coordinate analysis



- Committor $p_B(x)$ is THE reaction coordinate
- Committor is high dimensional function; difficult to gain physical insight
- **dimensionality reduction:** find best low dimensional order parameter combination that best represents committor

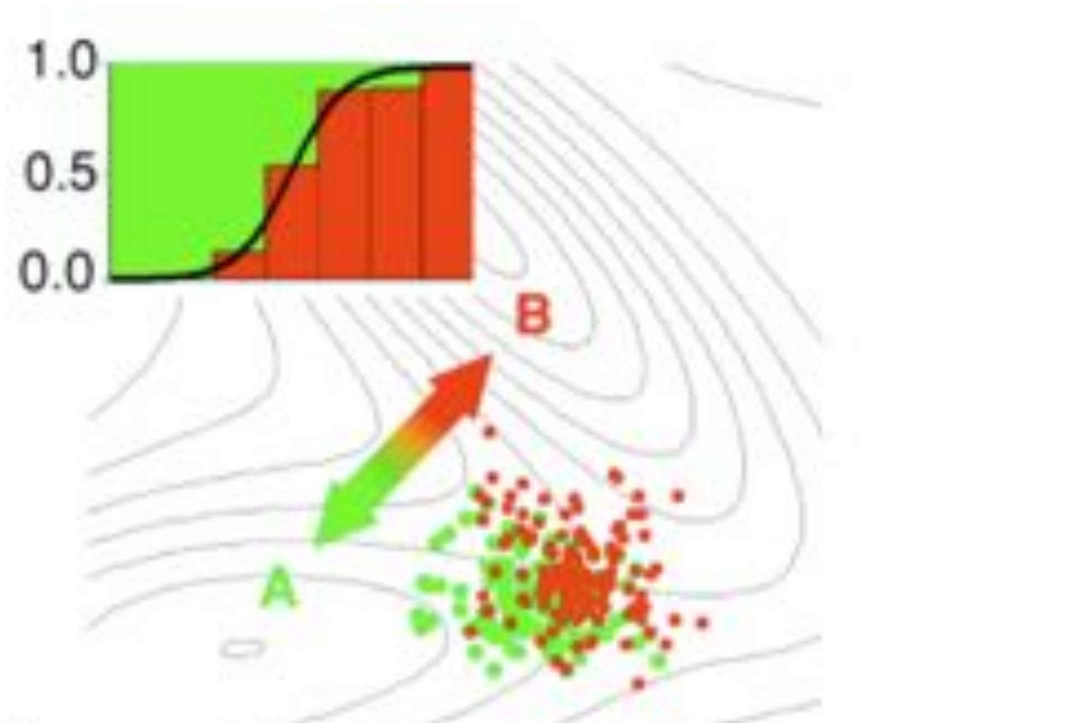


- Interpret each TPS shot as a committor attempt.
  Use info to optimise reaction coordinate model $r(q_1, q_2, ..)$

Peters & Trout, JCP 125 054108(2006)
see also Best & Hummer PNAS (2005)

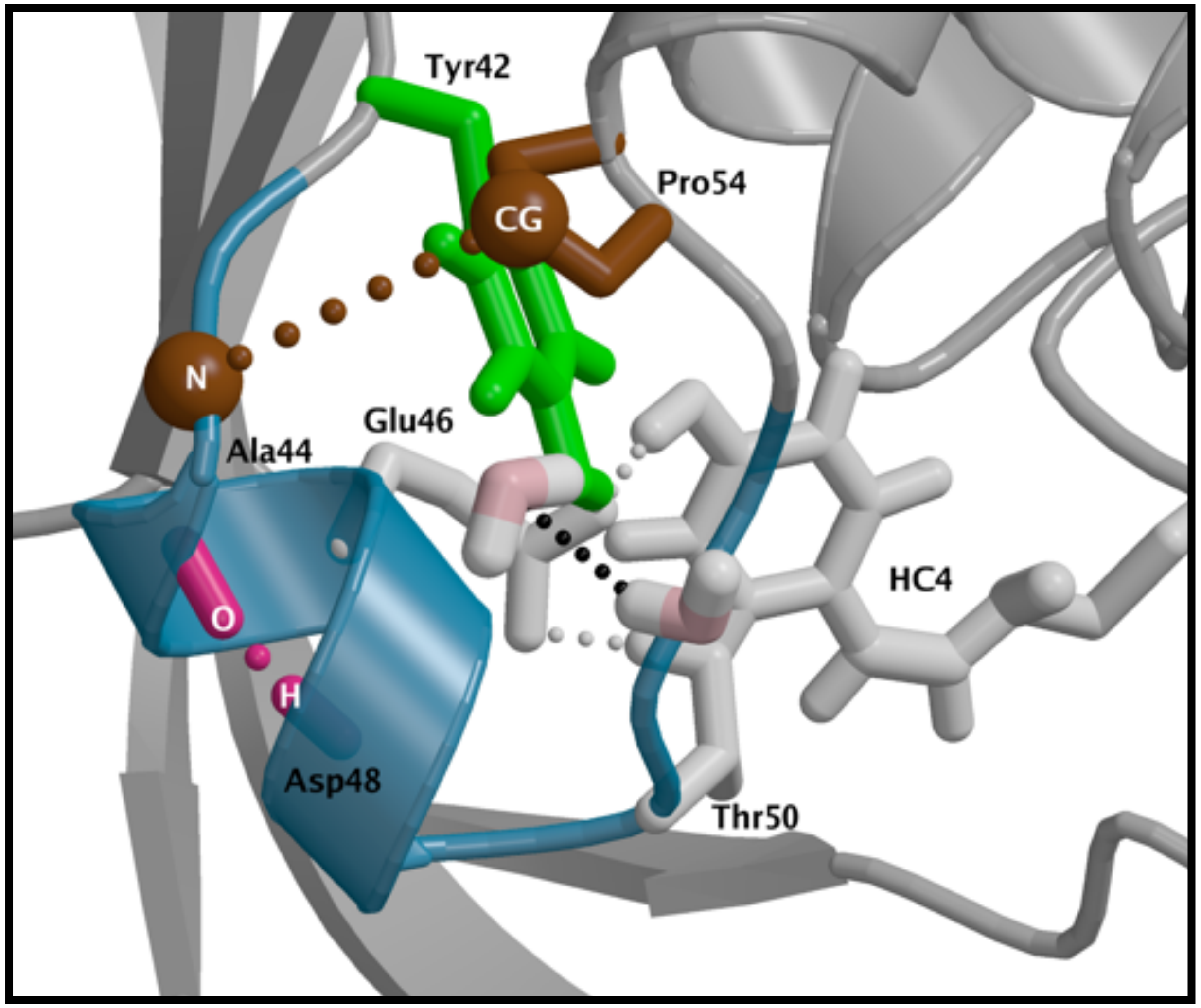
- Likelihood maximisation of predicted committor model

$$L(\alpha) = \prod_{i=1}^{N_B} p_B(r(q(\mathbf{x}_i^{(B)})) \prod_{i=1}^{N_A} (1 - p_B(r(q(\mathbf{x}_i^{(B)})))$$



- **result**: best model for the data given

Reaction coordinate of helix$_{\alpha 3}$ unfolding



Reaction coordinate by likelihood maximization *(Peters & Trout, JCP 2006)*

Order Parameters involved (out of 78):

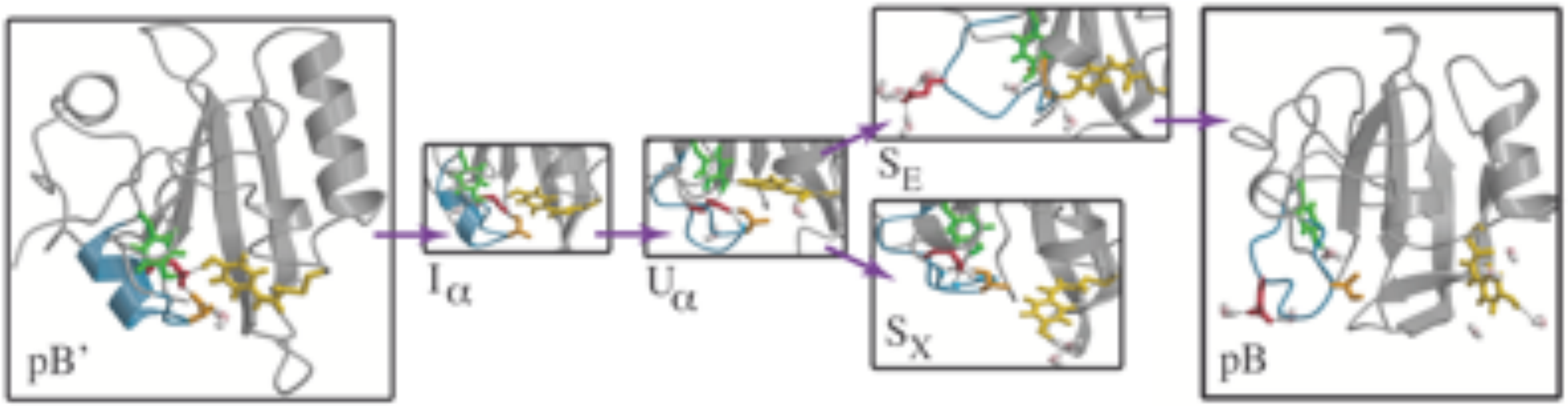
- RMSD$_{\alpha}$
- nwY42 : water molecules around Tyr42
- dPA : distance Ala44(N) - Pro54(Cγ)
- dhb2 : distance Ala44(O) - Asp48(H)

δLmin = 4.17

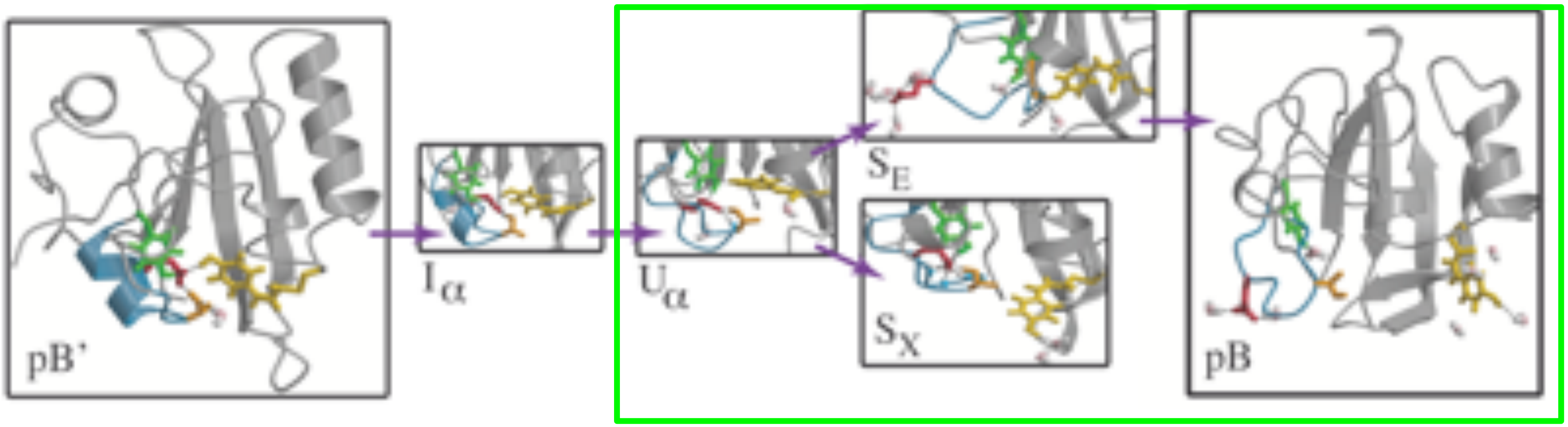
n	ln L	RC
1	-2117	3.89–29.10 × rmsdα
2	-2098	3.88–26.35 × rmsdα − 0.19 × nwY42
3	-2085	5.11–16.81 × rmsdα − 4.68 × dhb2 − 2.55 × dPA

*Vreede, Juraszek, PGB, PNAS **107** 2397 (2010)*

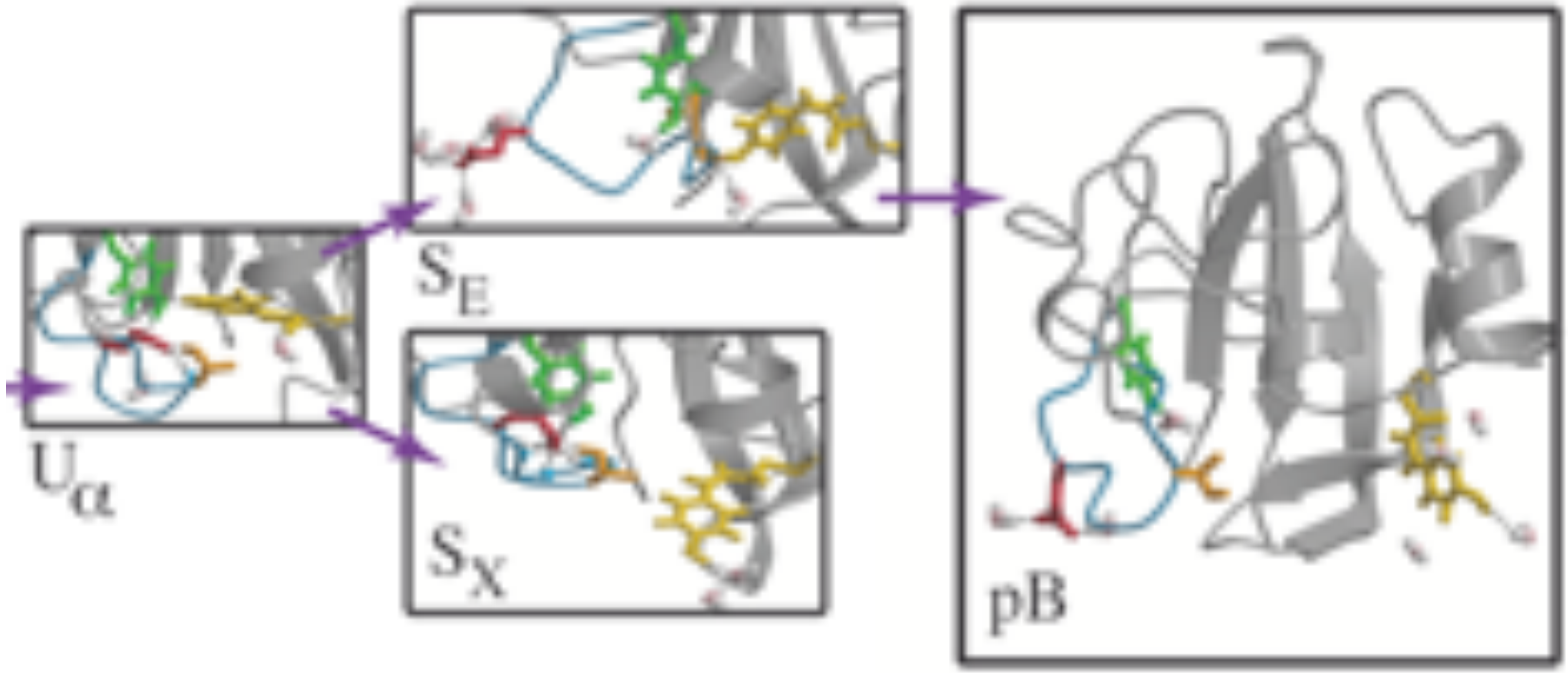
Solvent exposure transitions



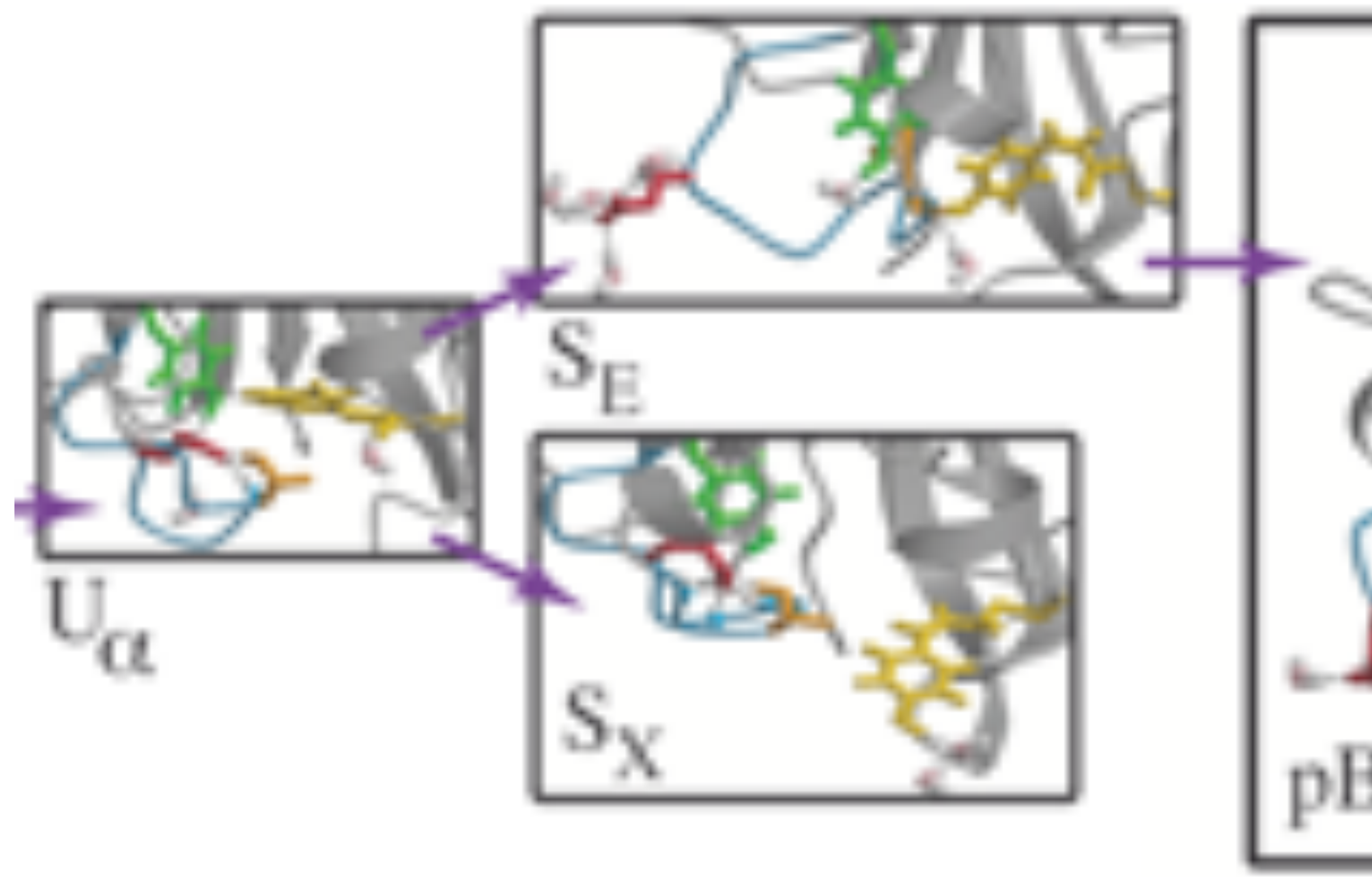
Solvent exposure transitions



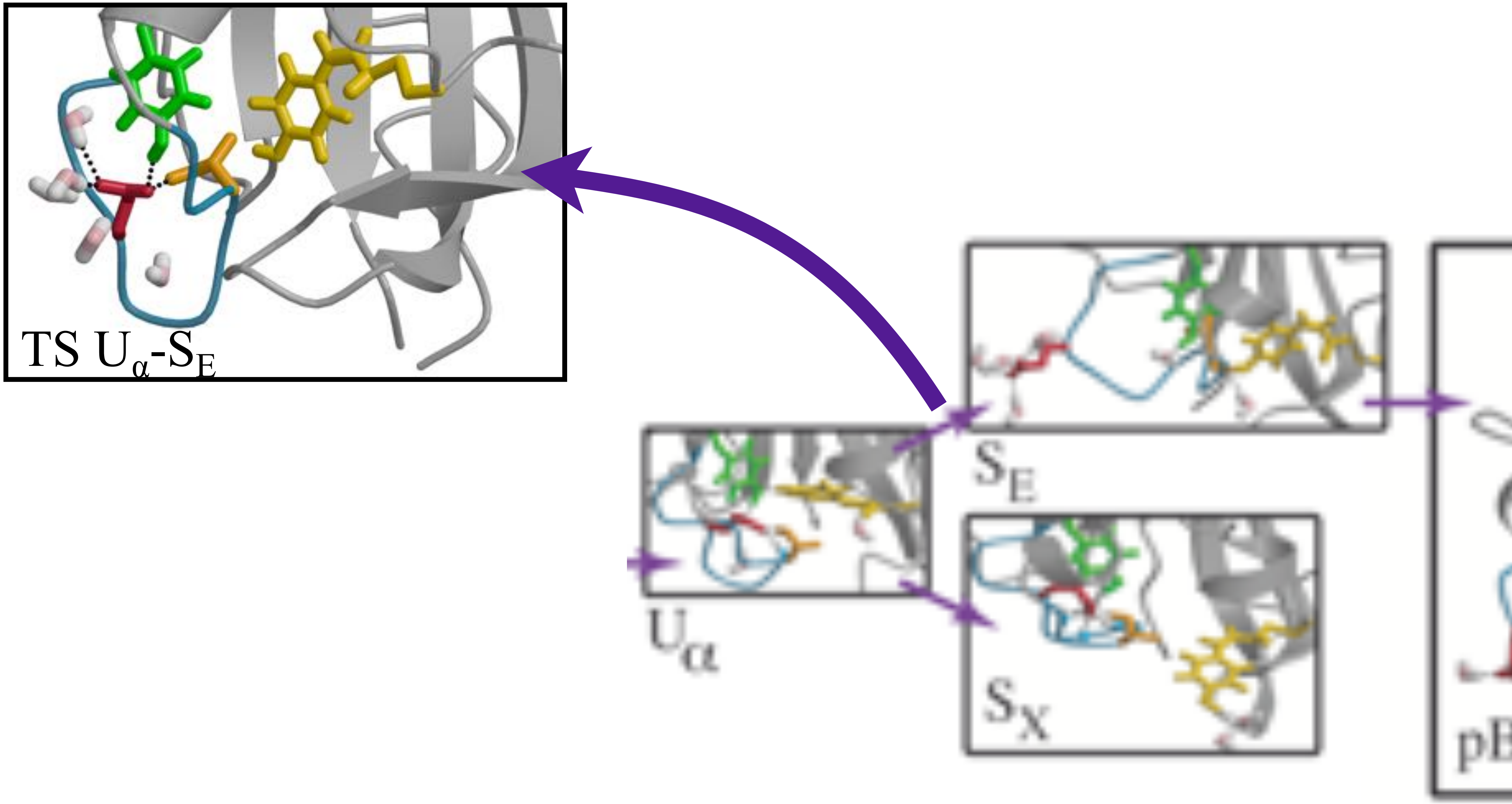
Solvent exposure transitions



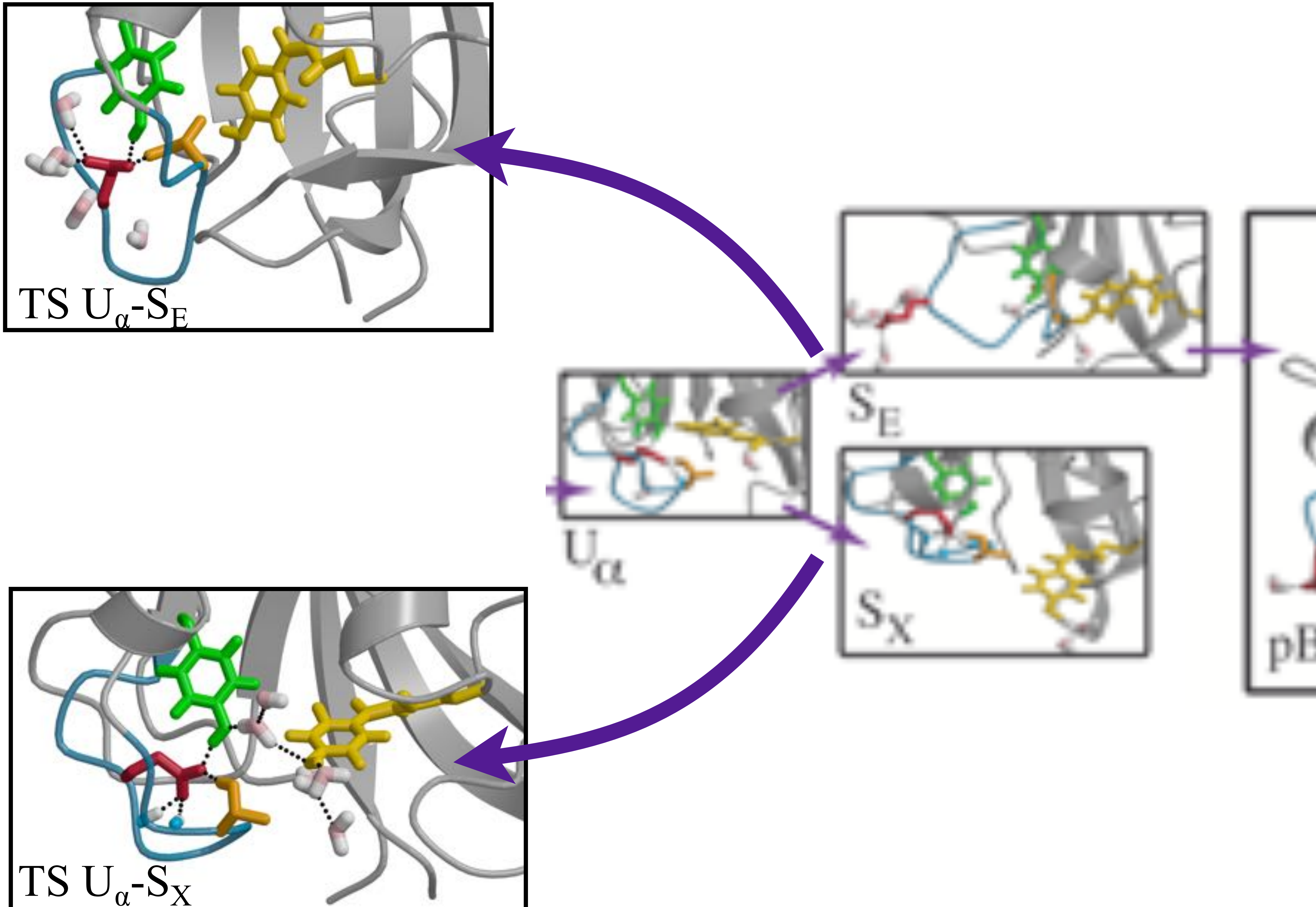
Solvent exposure transitions



Solvent exposure transitions

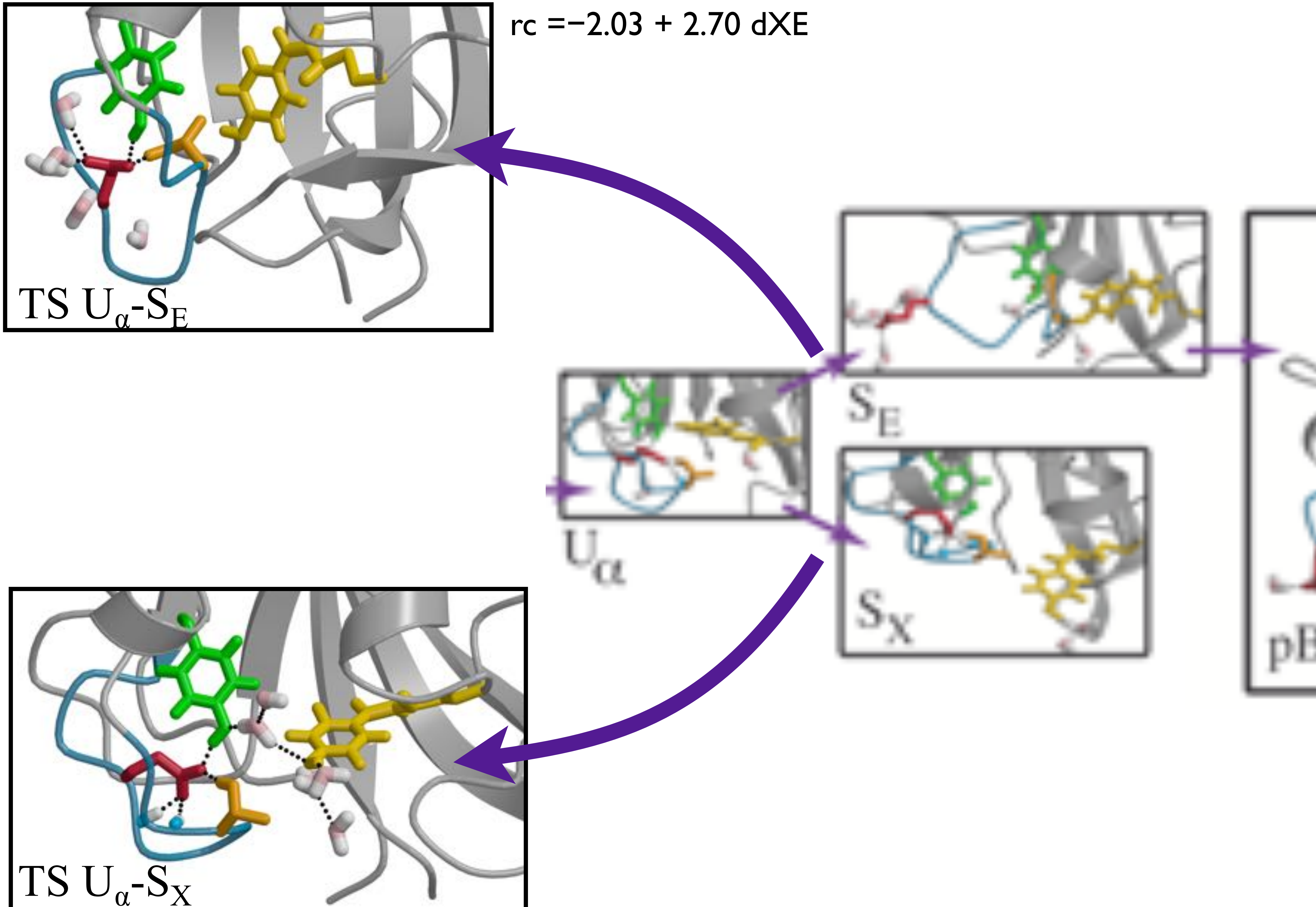


Solvent exposure transitions



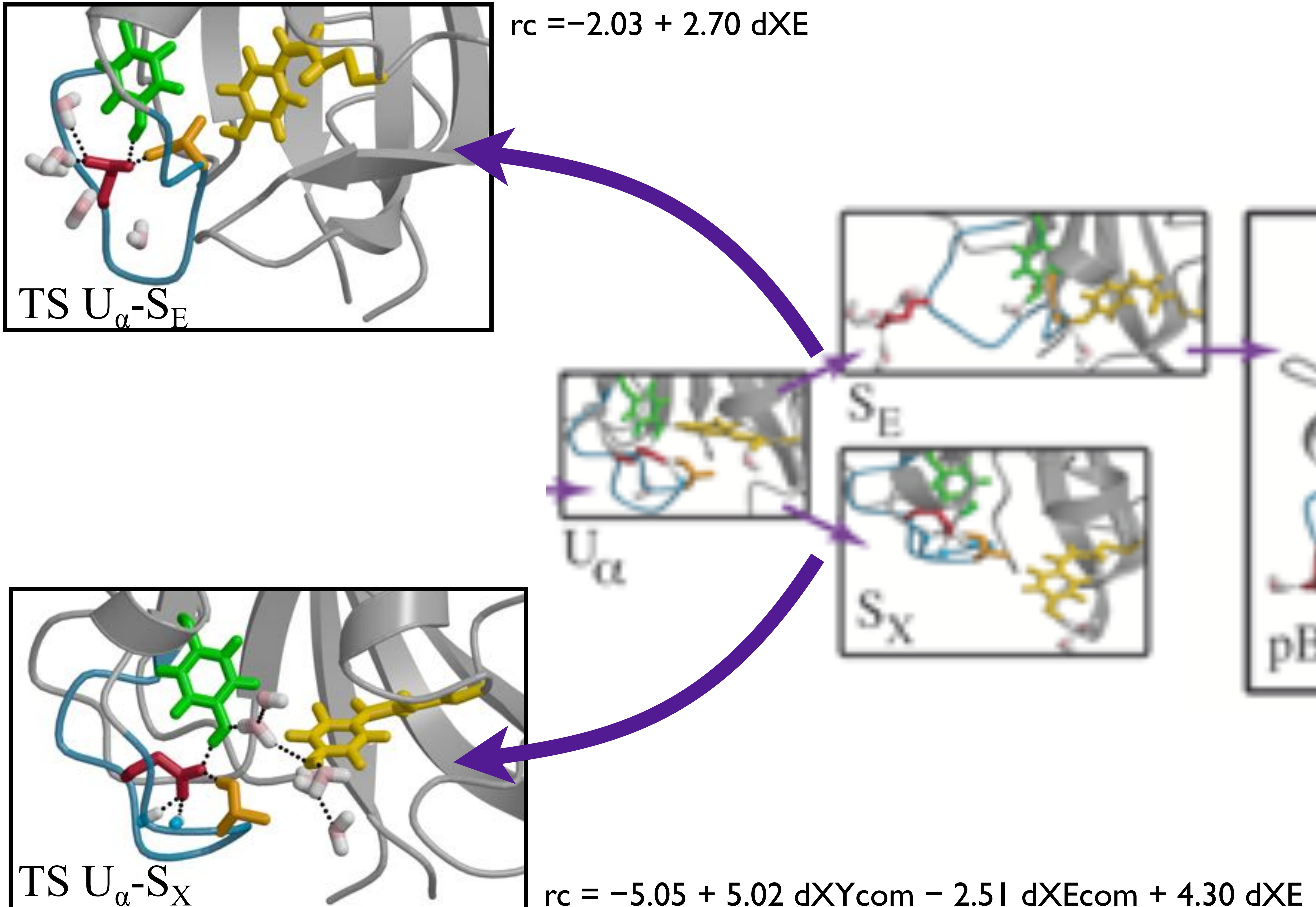
Solvent exposure transitions

rc =−2.03 + 2.70 dXE



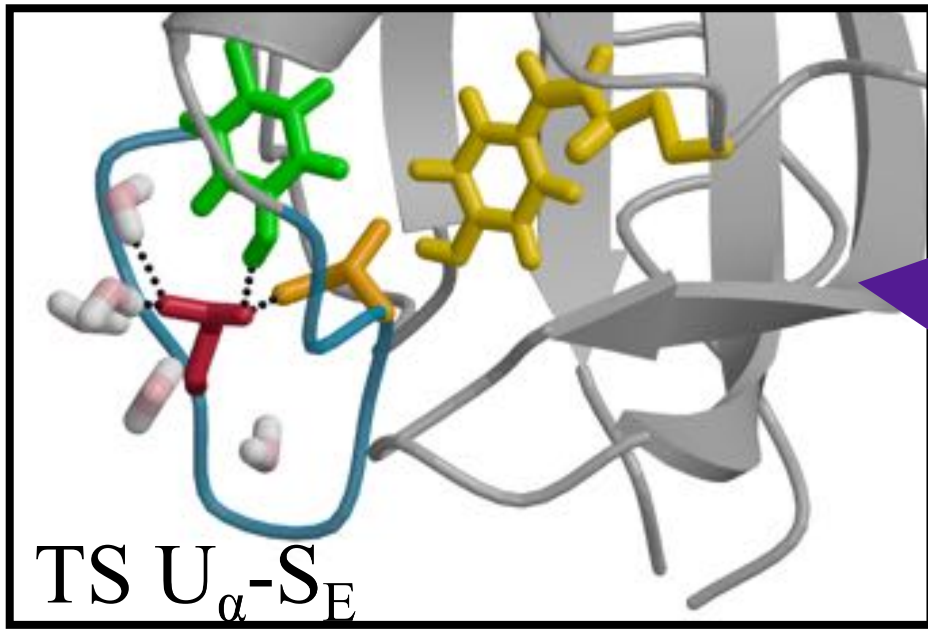
Solvent exposure transitions

rc $= -2.03 + 2.70$ dXE



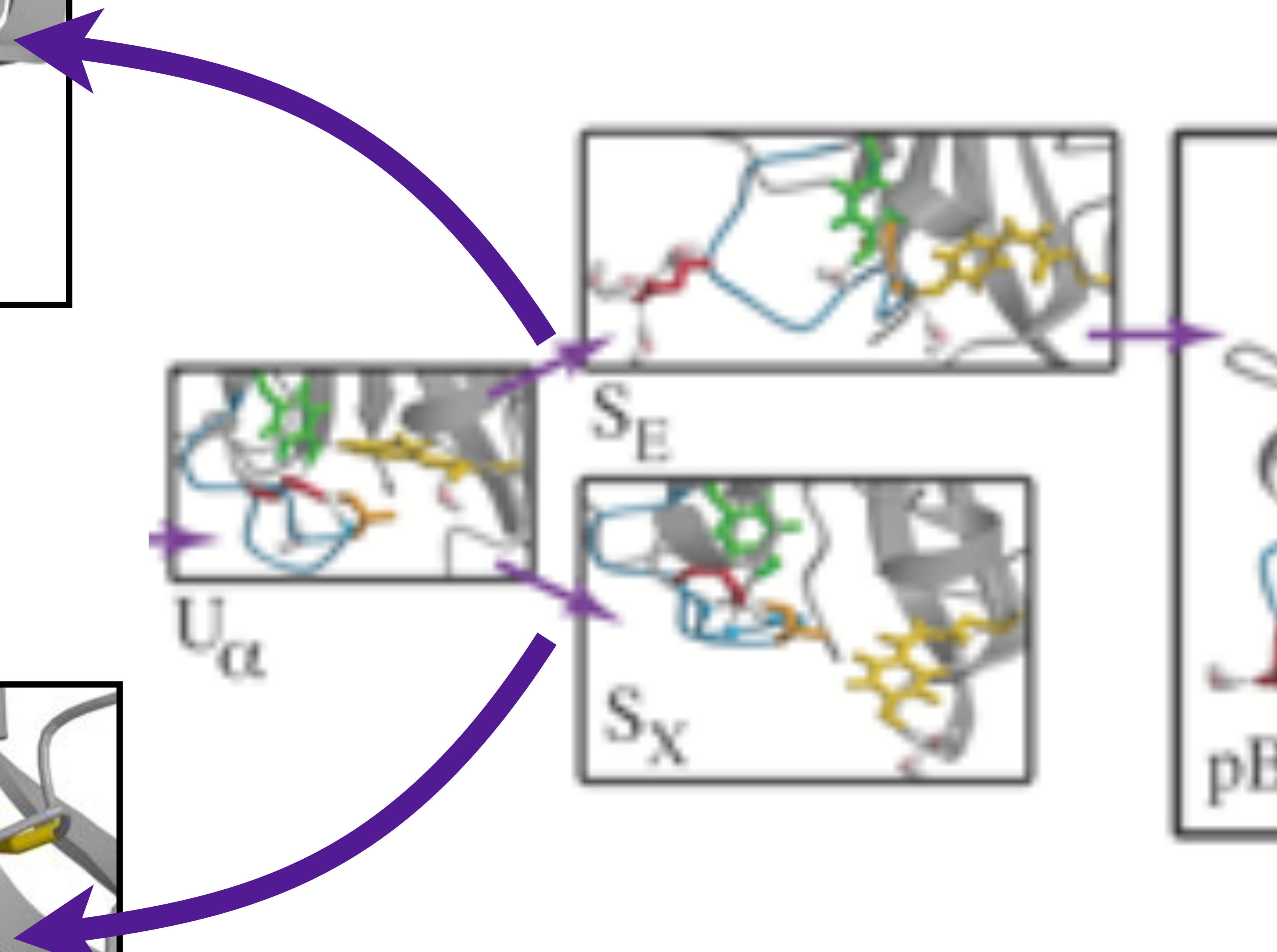
rc $= -5.05 + 5.02$ dXYcom $- 2.51$ dXEcom $+ 4.30$ dXE

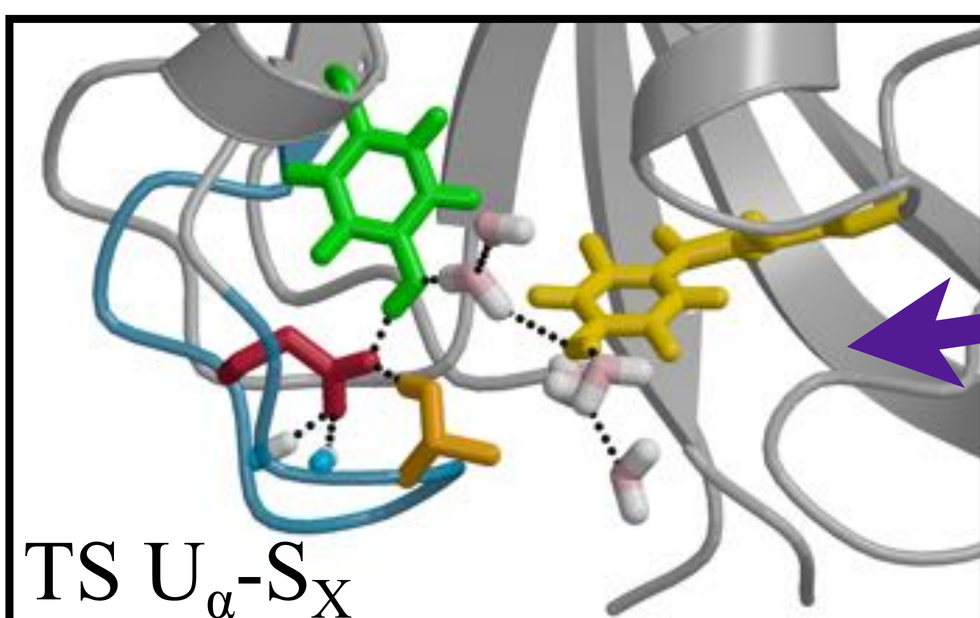
Solvent exposure transitions



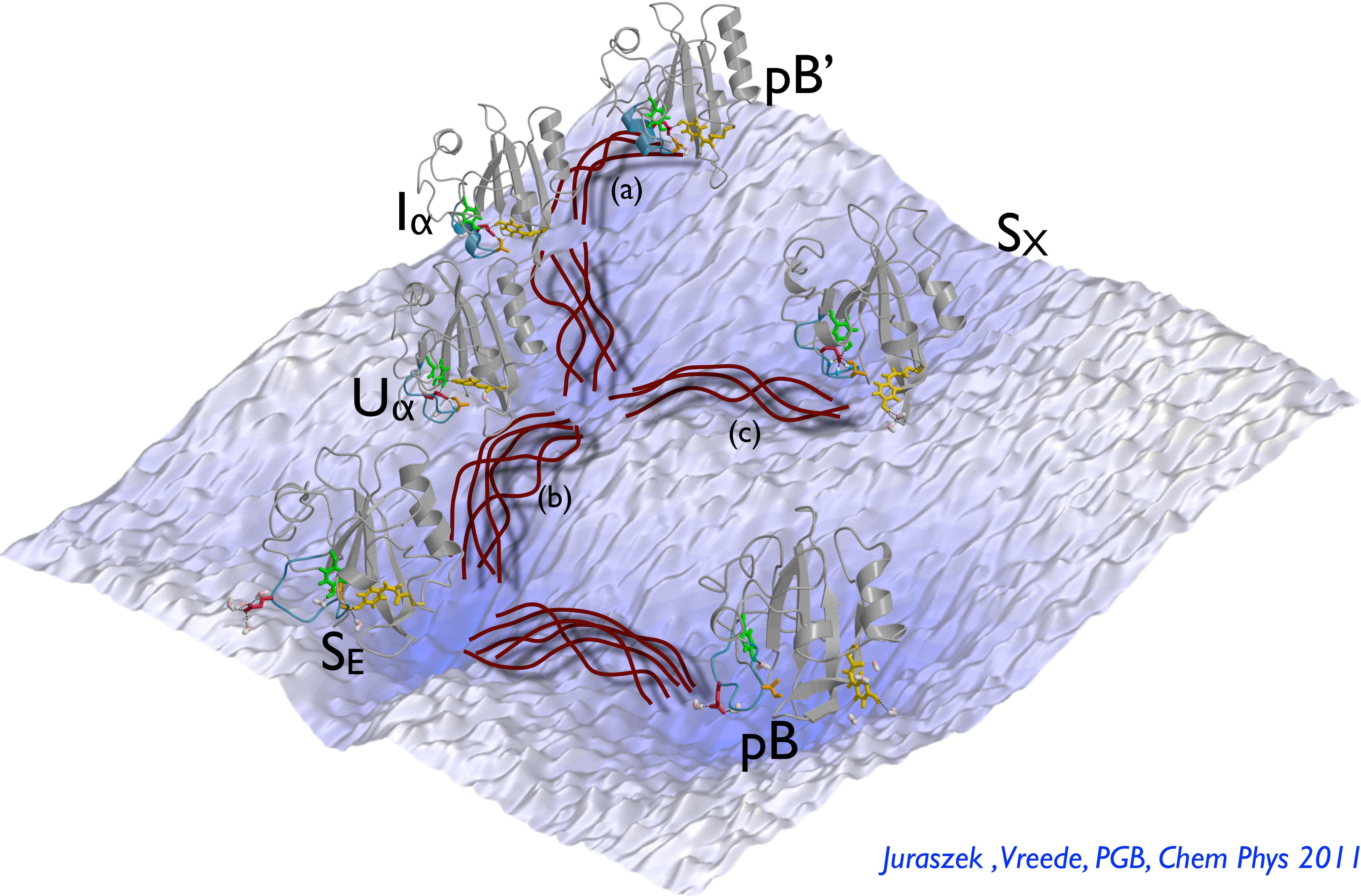
rc = −2.03 + 2.70 dXE

rate limiting step 16 k_BT: $k \approx 1\ ms^{-1}$





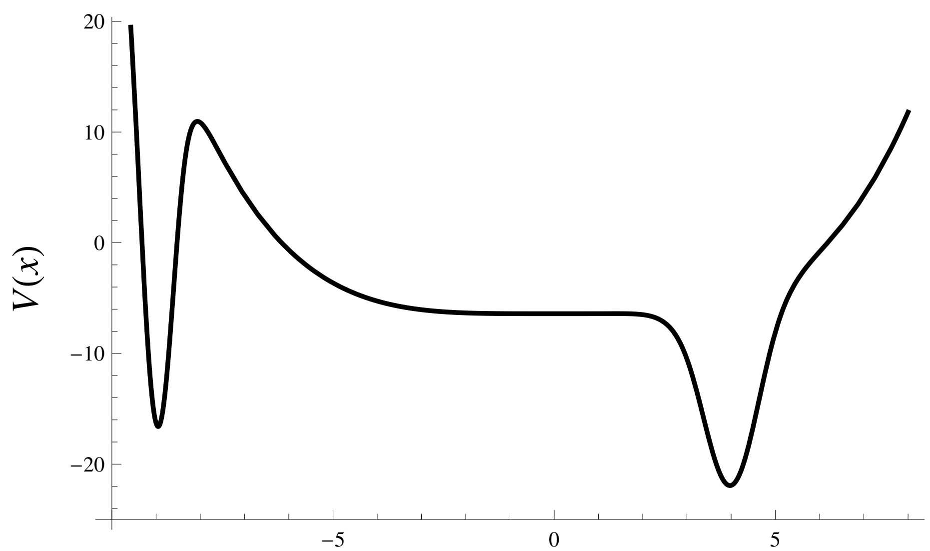
rc = −5.05 + 5.02 dXYcom − 2.51 dXEcom + 4.30 dXE

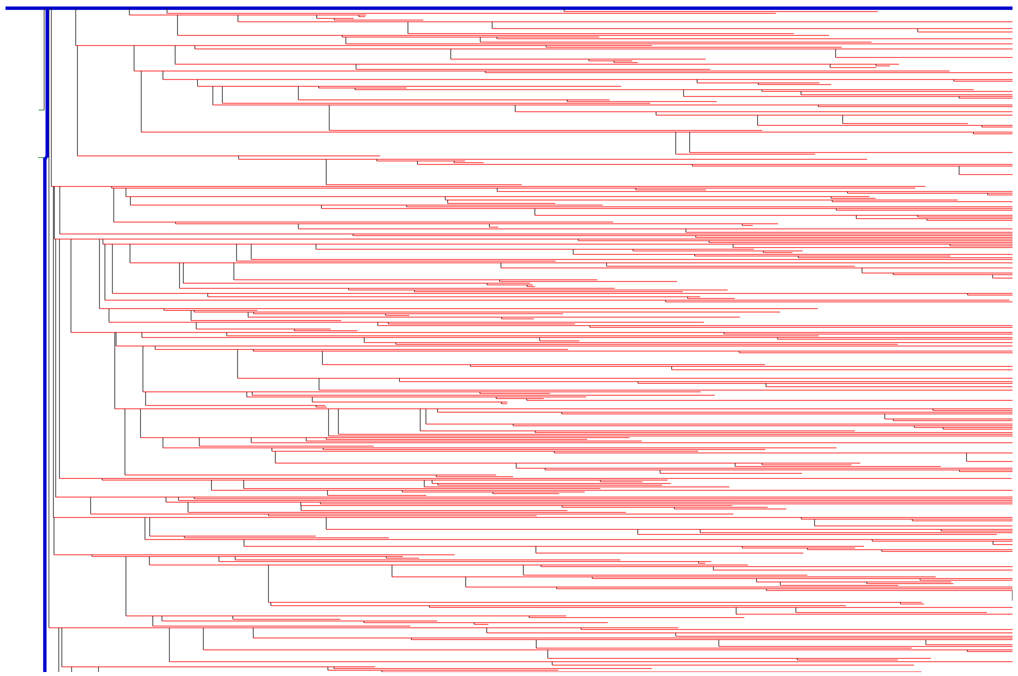


Juraszek ,Vreede, PGB, Chem Phys 2011

Spring shooting for asymmetric barriers

- uniform one way shoot has bad decorrelation





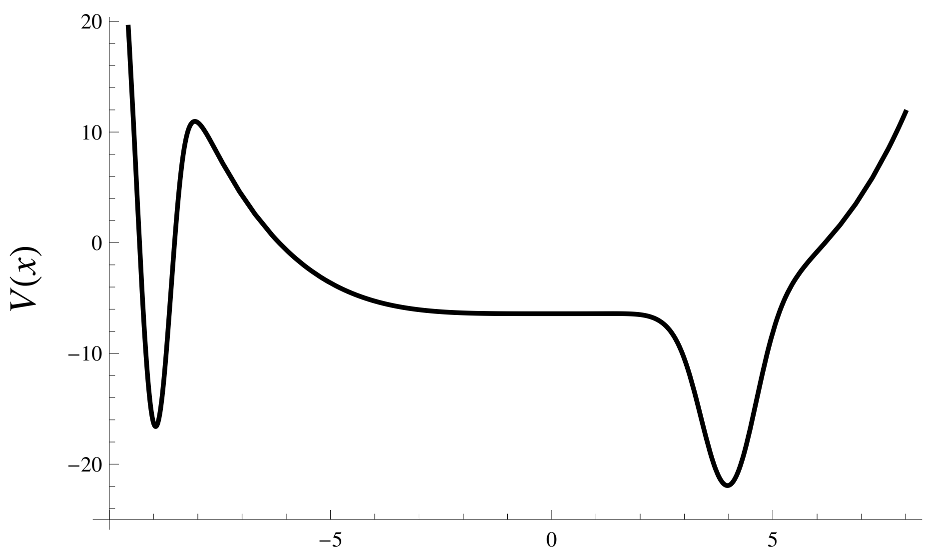
bad decorrelation

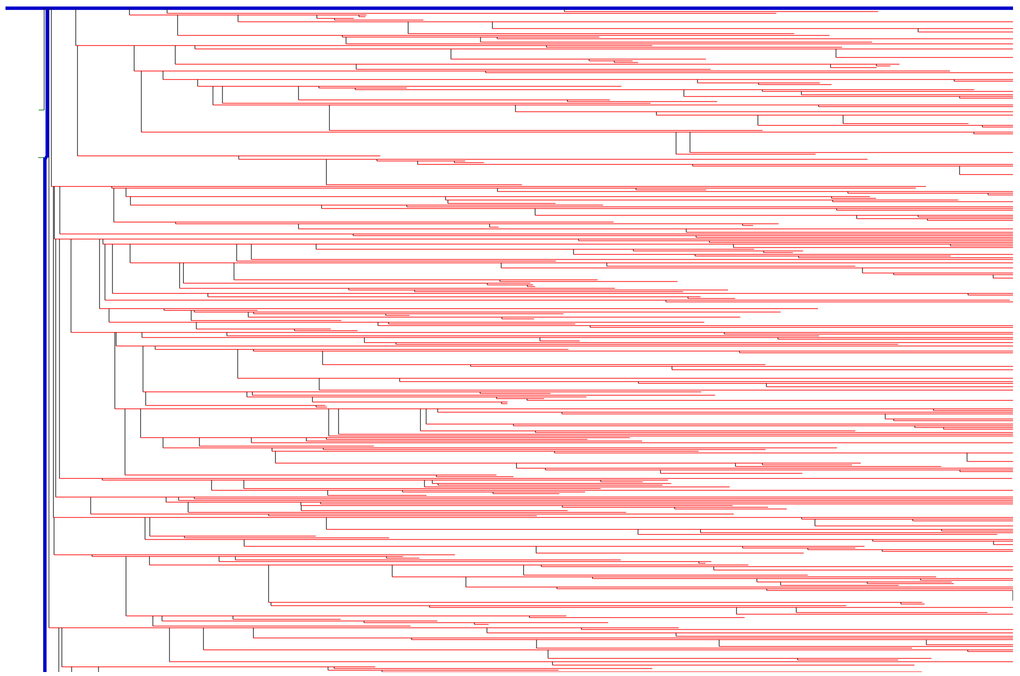
- spring shooting algorithm:

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

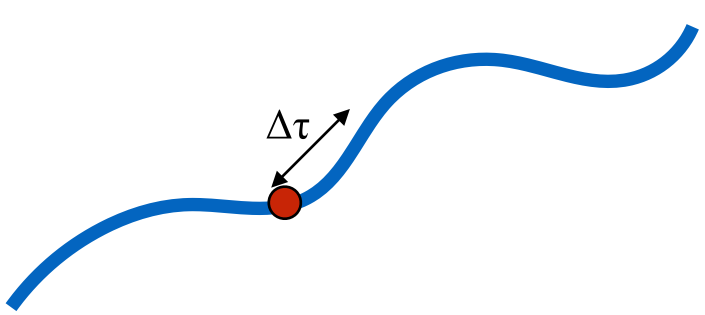
- uniform one way shoot has bad decorrelation





bad decorrelation

- spring shooting algorithm:

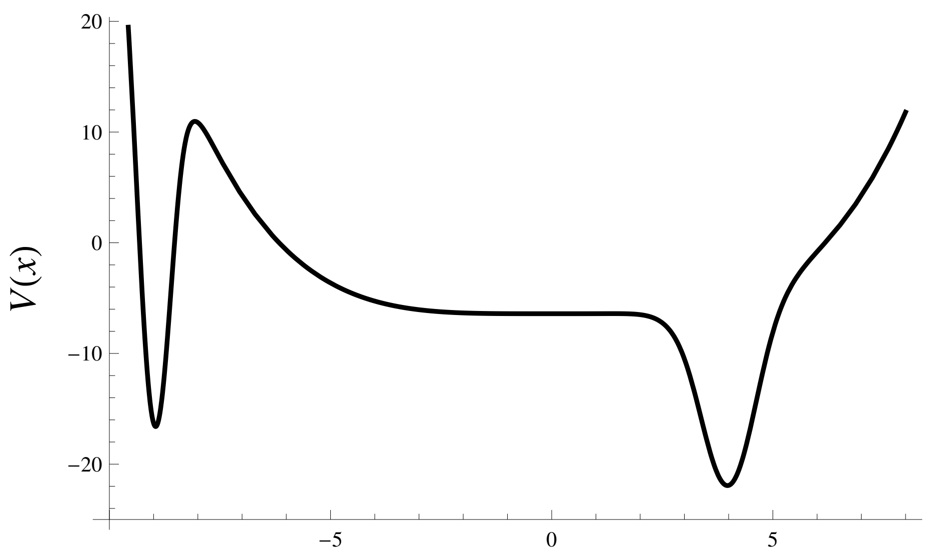


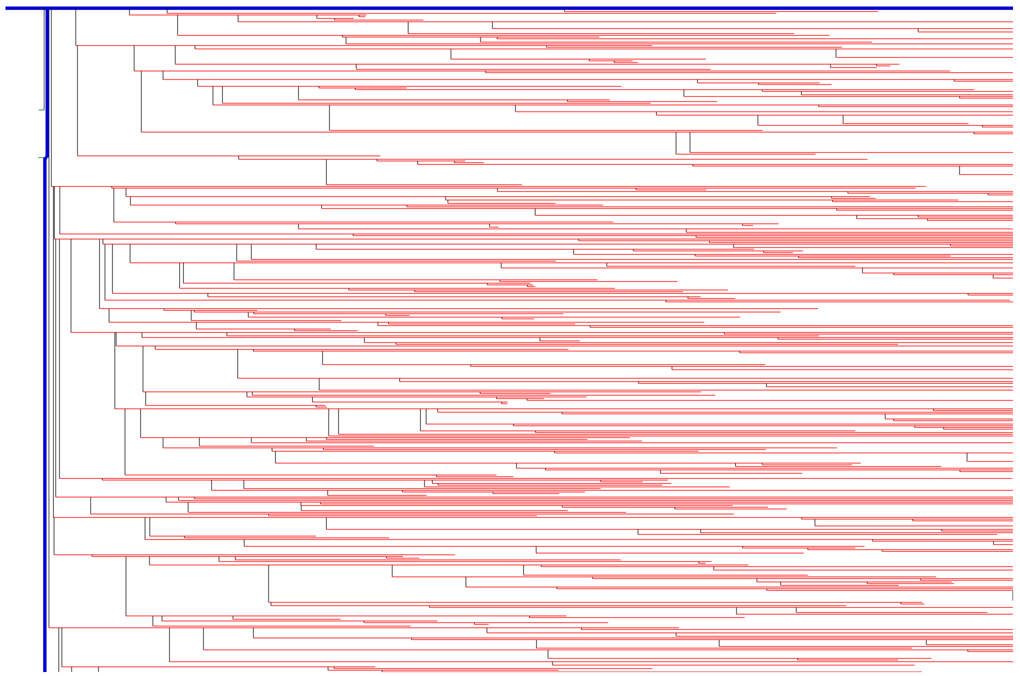
$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

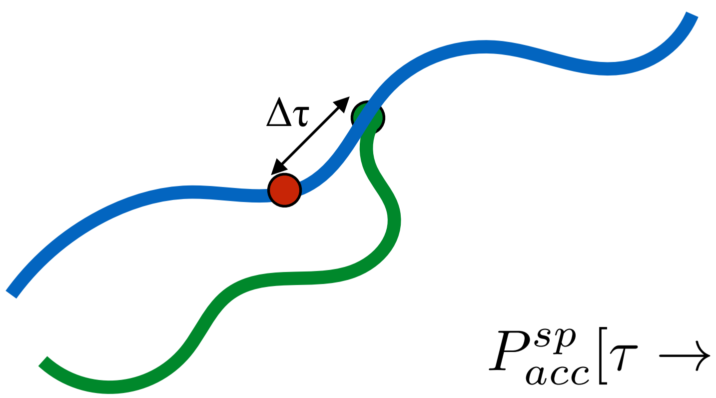
- uniform one way shoot has bad decorrelation





bad decorrelation

- spring shooting algorithm:

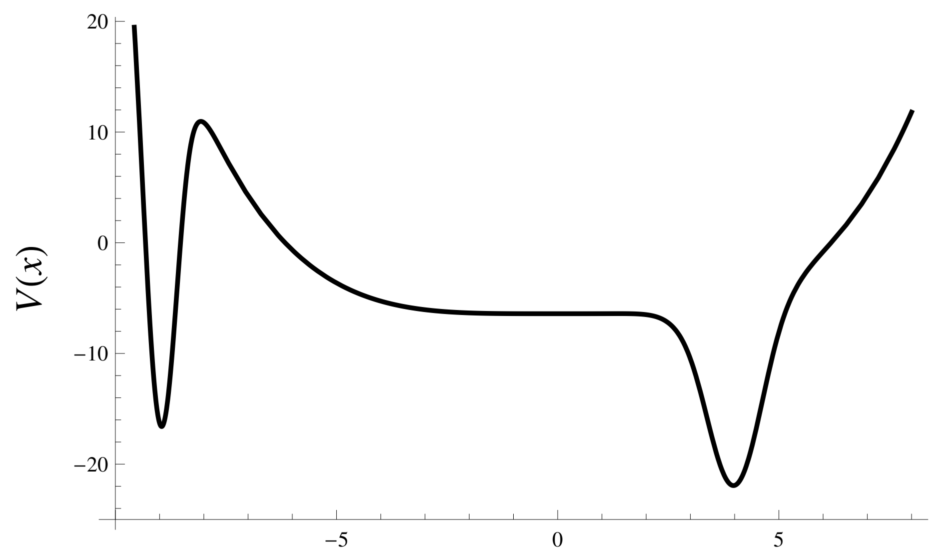


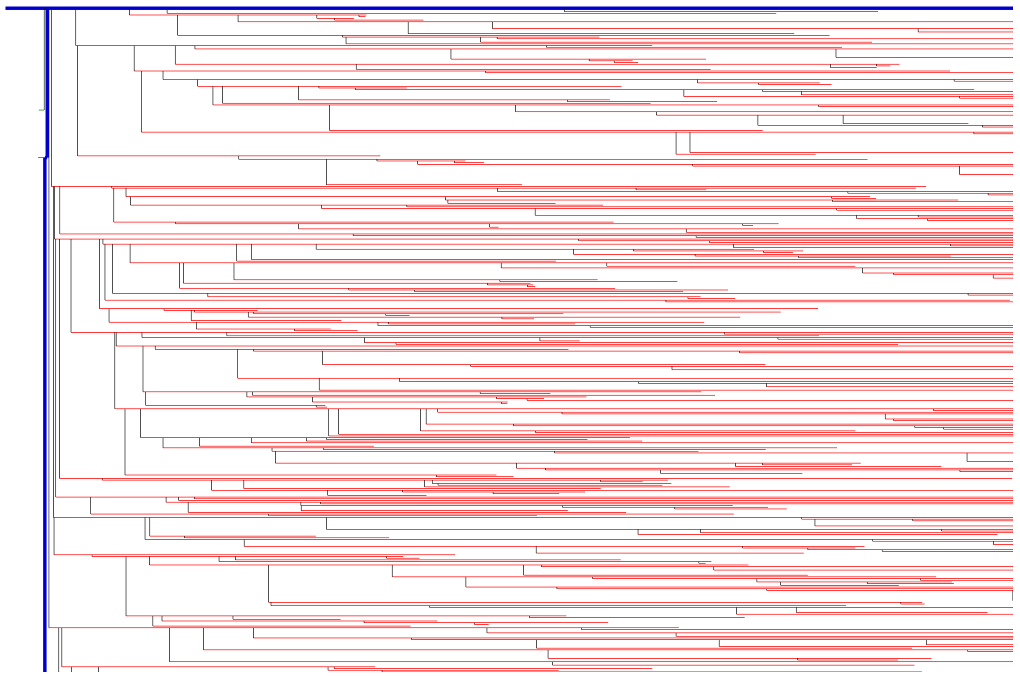
$$P^{sp}_{acc}[\tau \rightarrow \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

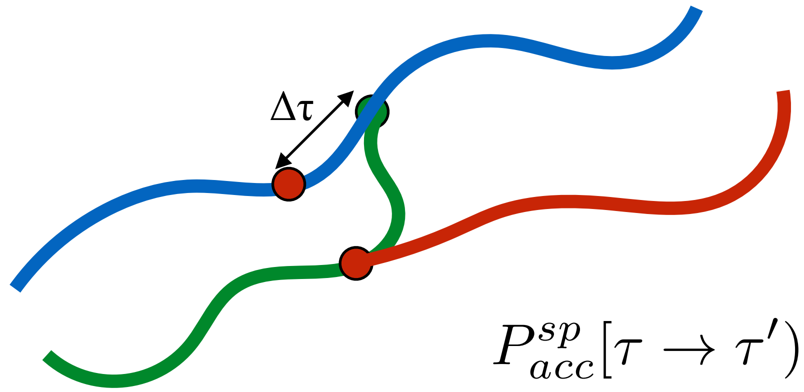
- uniform one way shoot has bad decorrelation





bad decorrelation

- spring shooting algorithm:

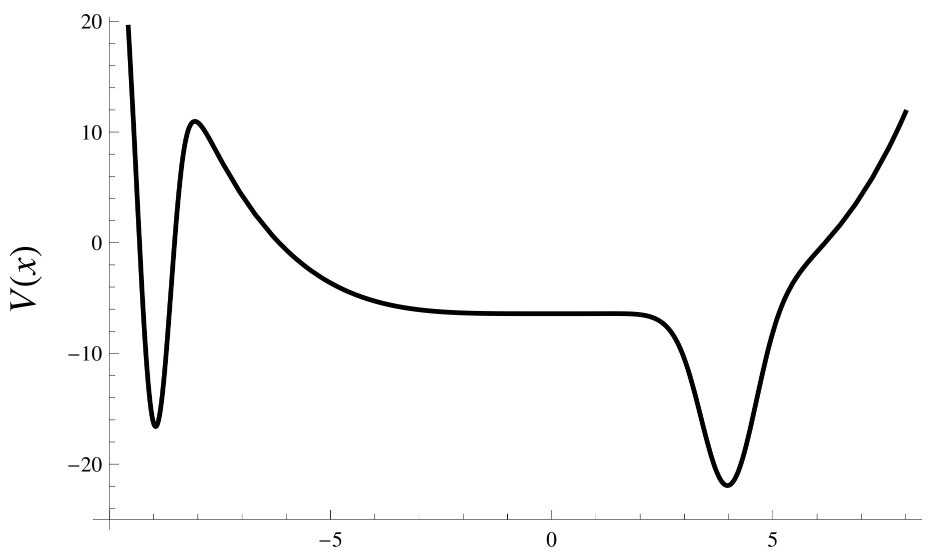


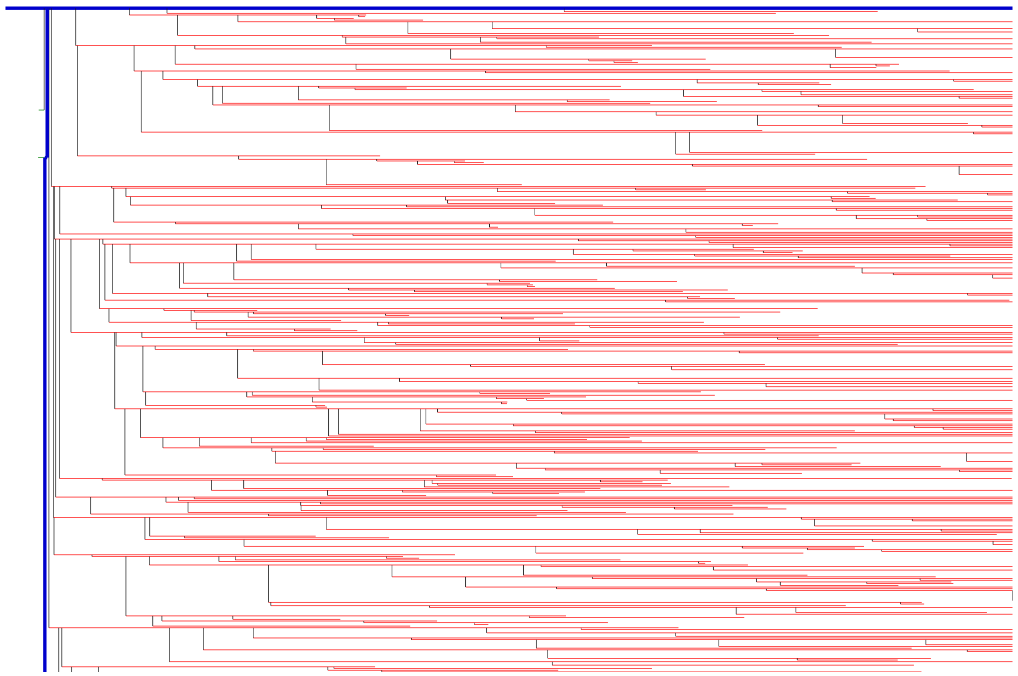
$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

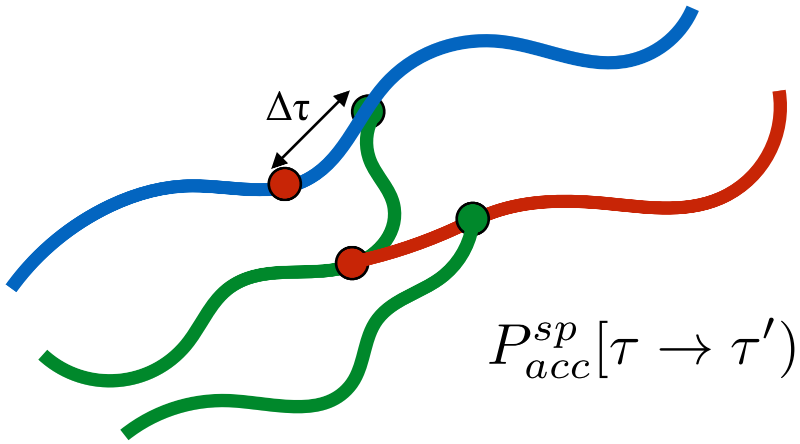
- uniform one way shoot has bad decorrelation





bad decorrelation

- spring shooting algorithm:

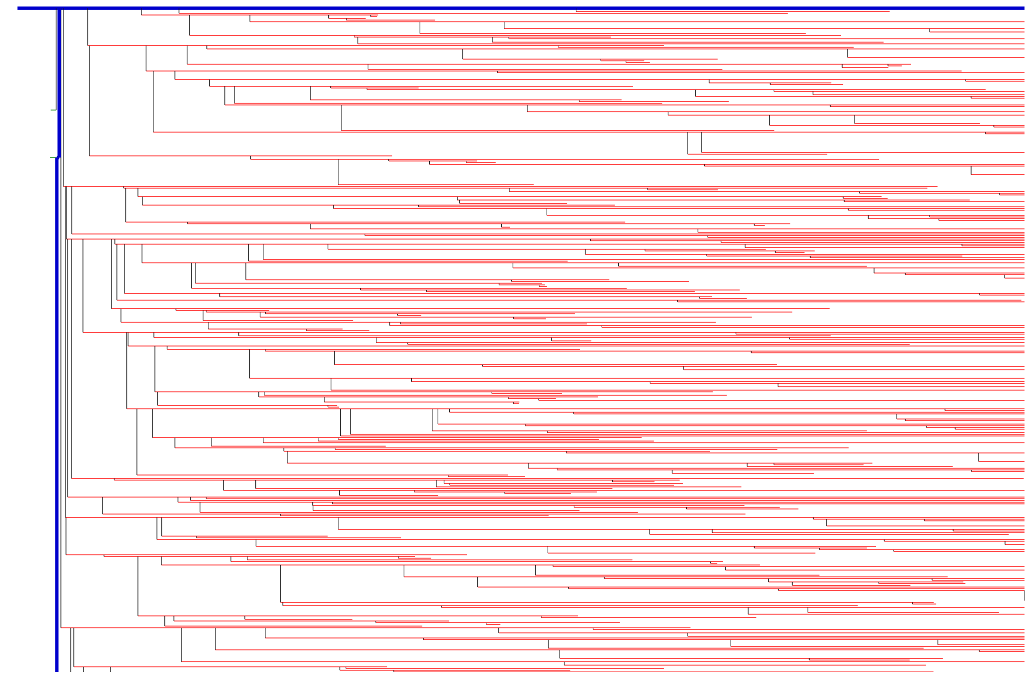


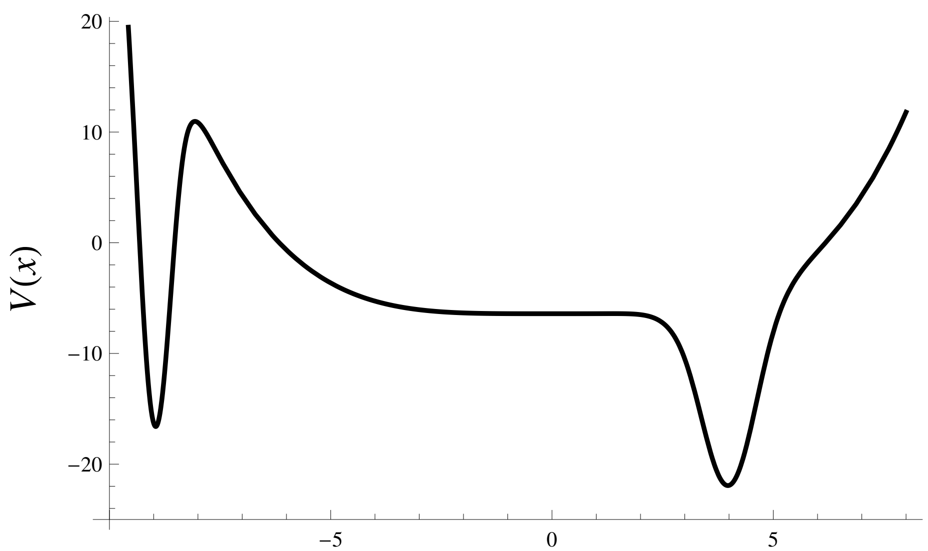
$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

- uniform one way shoot has bad decorrelation

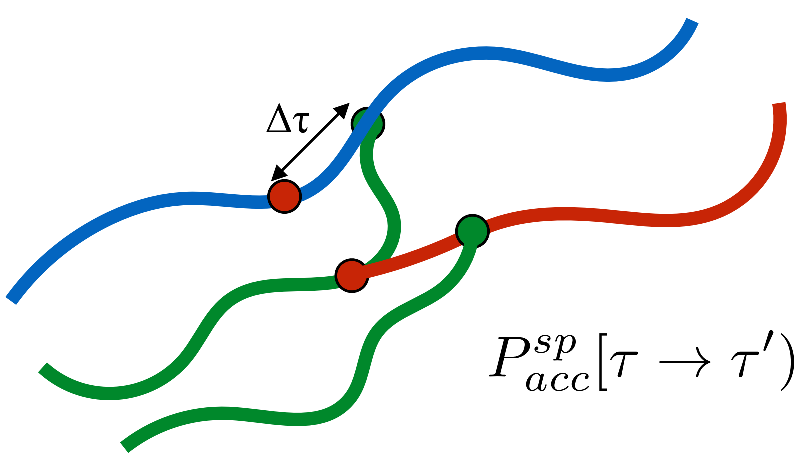




- spring shooting algorithm:

$$p_{sel}(\tau) = c\exp(sk\tau)$$

$$P^{sp}_{acc}[\tau \to \tau'] = \min\left[1, \frac{\exp(sk\tau')}{\exp(sk\tau)}\right] = \min[1, e^{sk\Delta\tau}]$$

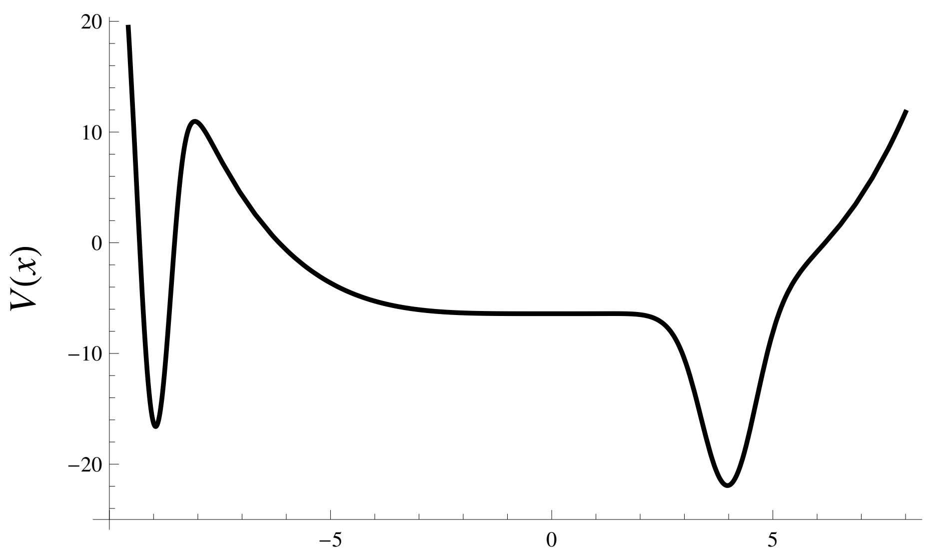


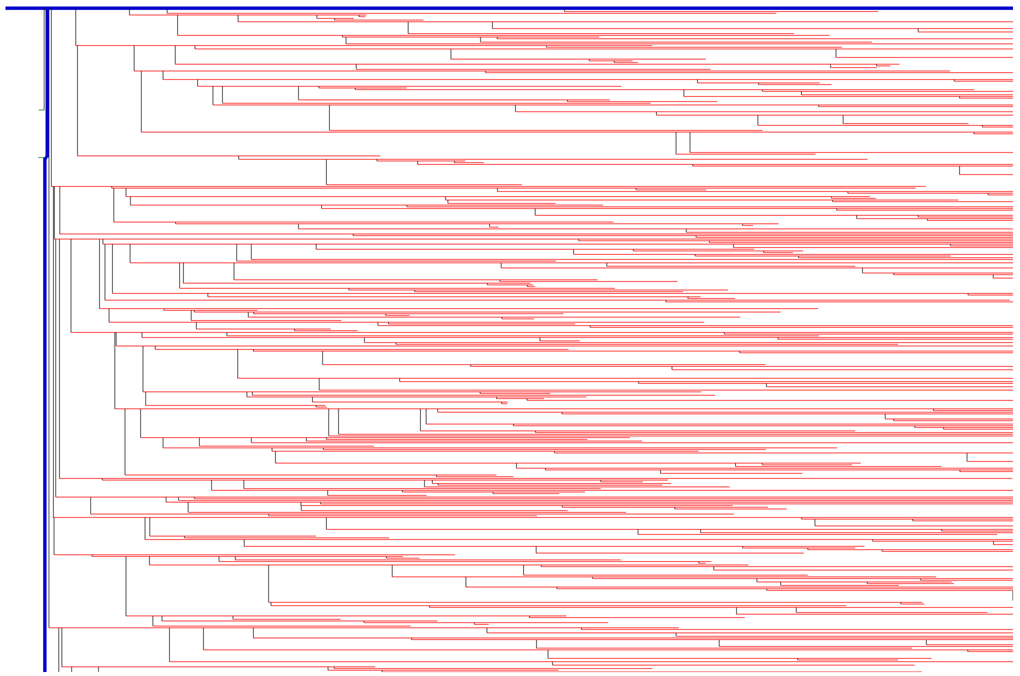
$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

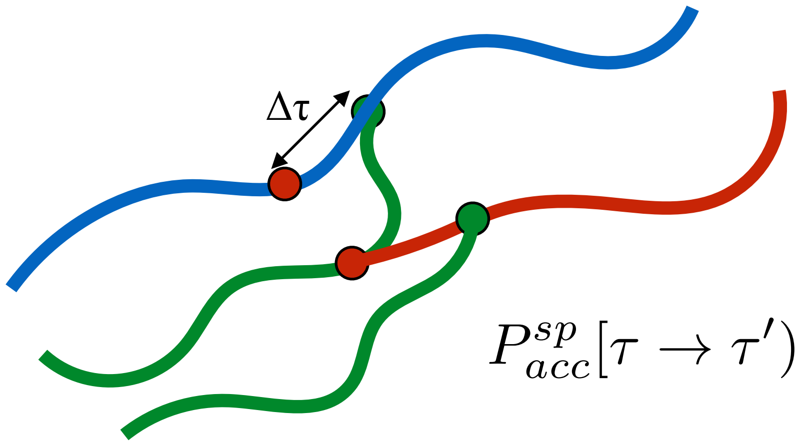
- uniform one way shoot has bad decorrelation





bad decorrelation

- spring shooting algorithm:

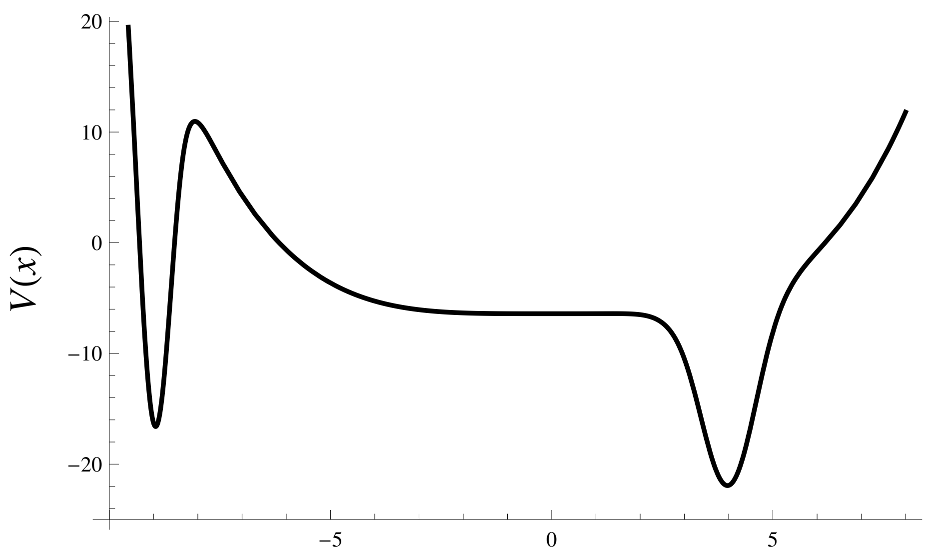


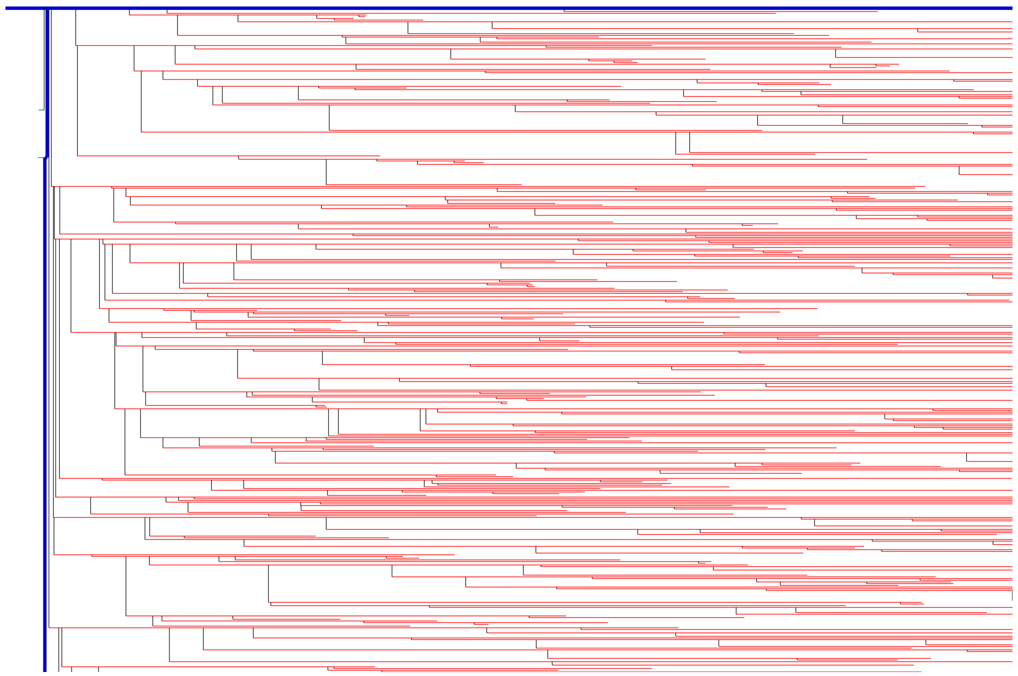
$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

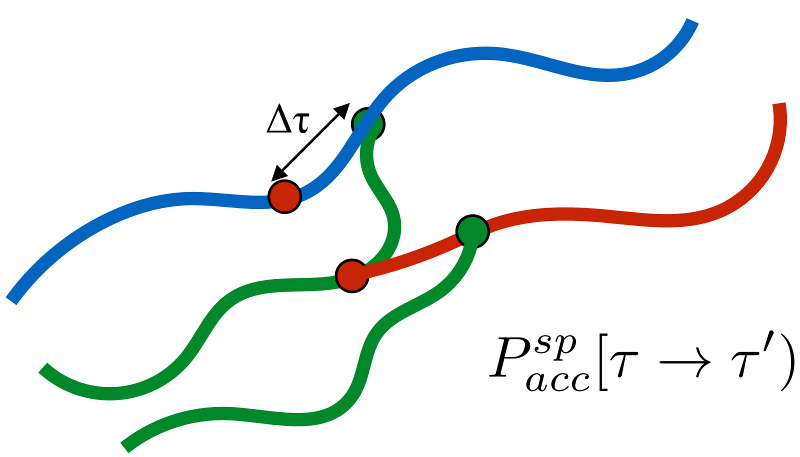
- uniform one way shoot has bad decorrelation



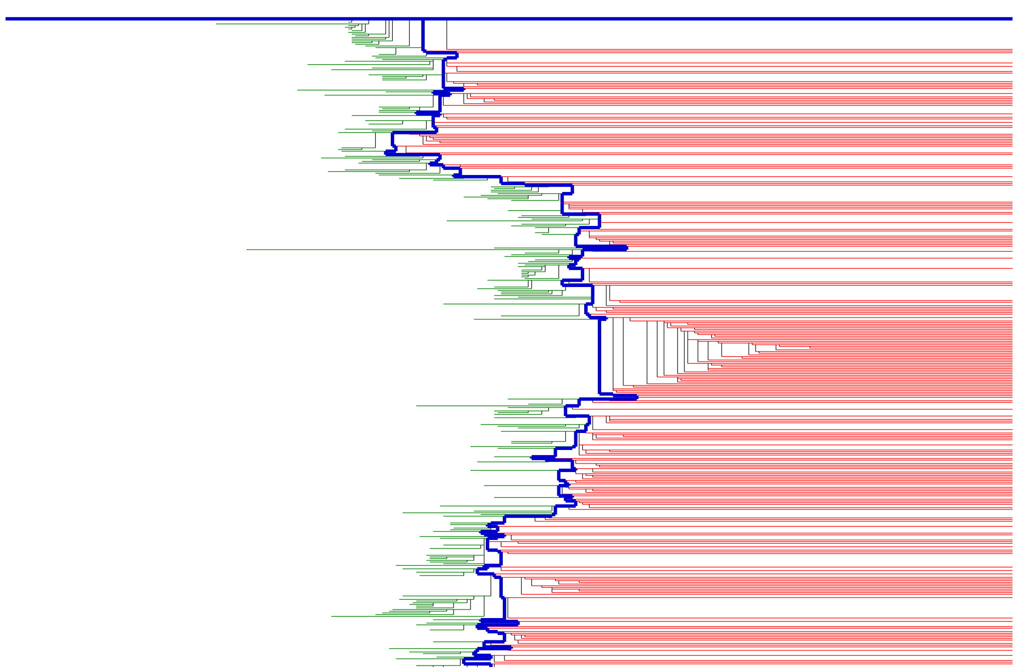


bad decorrelation

- spring shooting algorithm:



$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

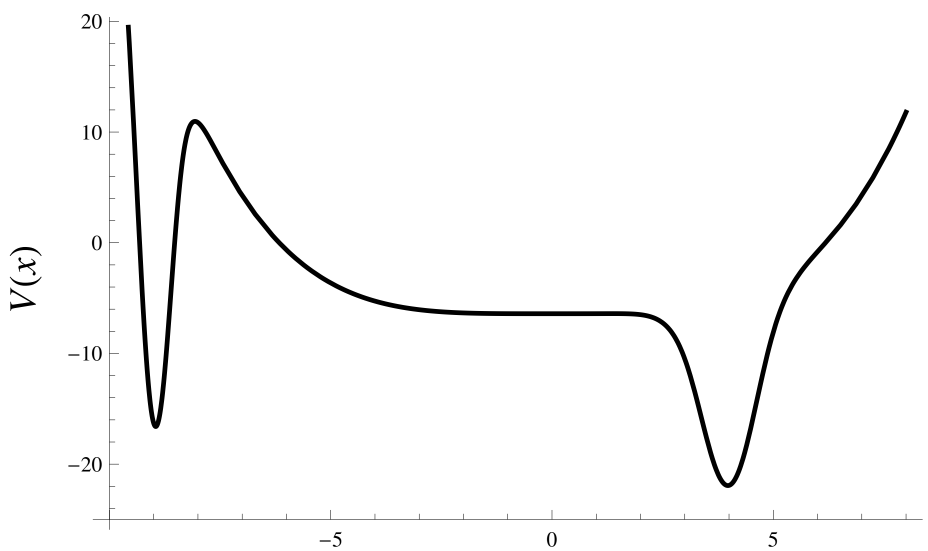


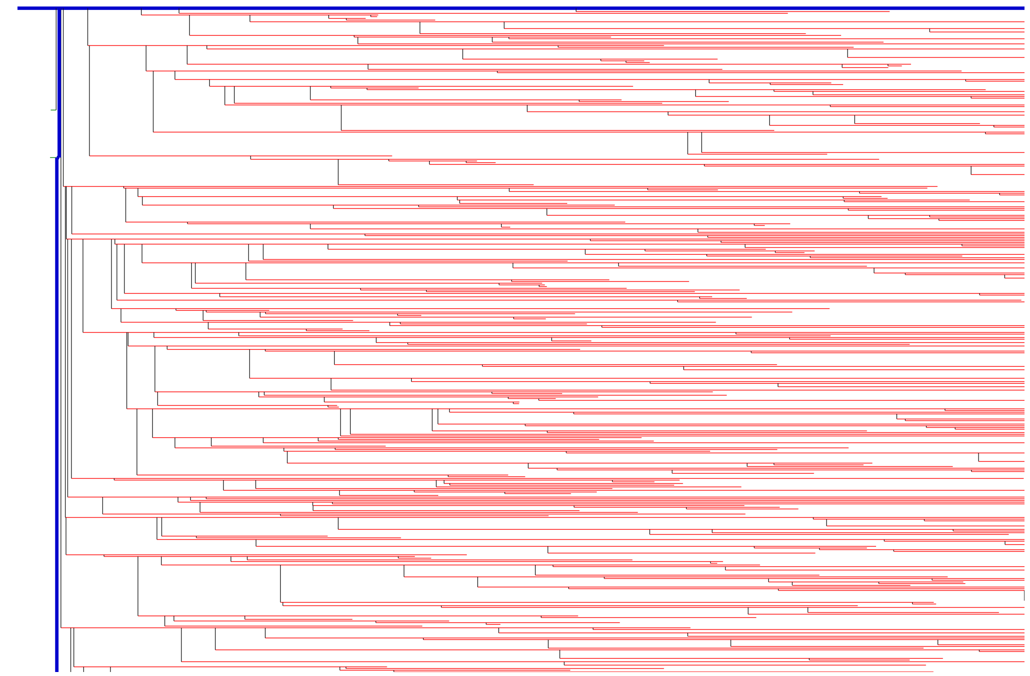
good decorrelation

Z.F. Brotzakis, PGB, JCP (2016)

Spring shooting for asymmetric barriers

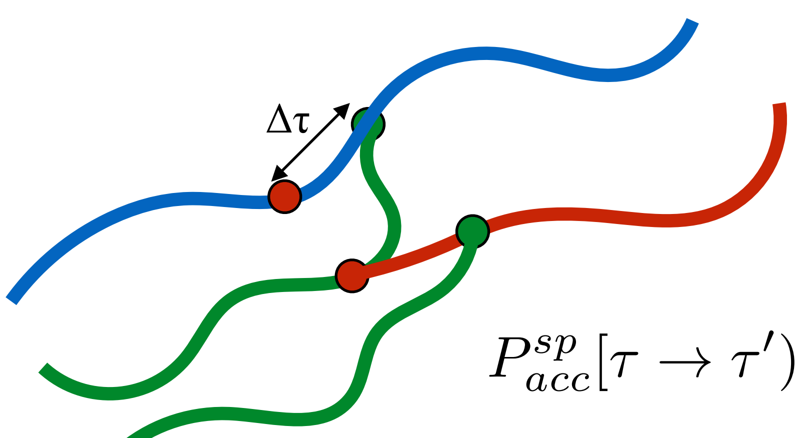
- uniform one way shoot has bad decorrelation





bad decorrelation

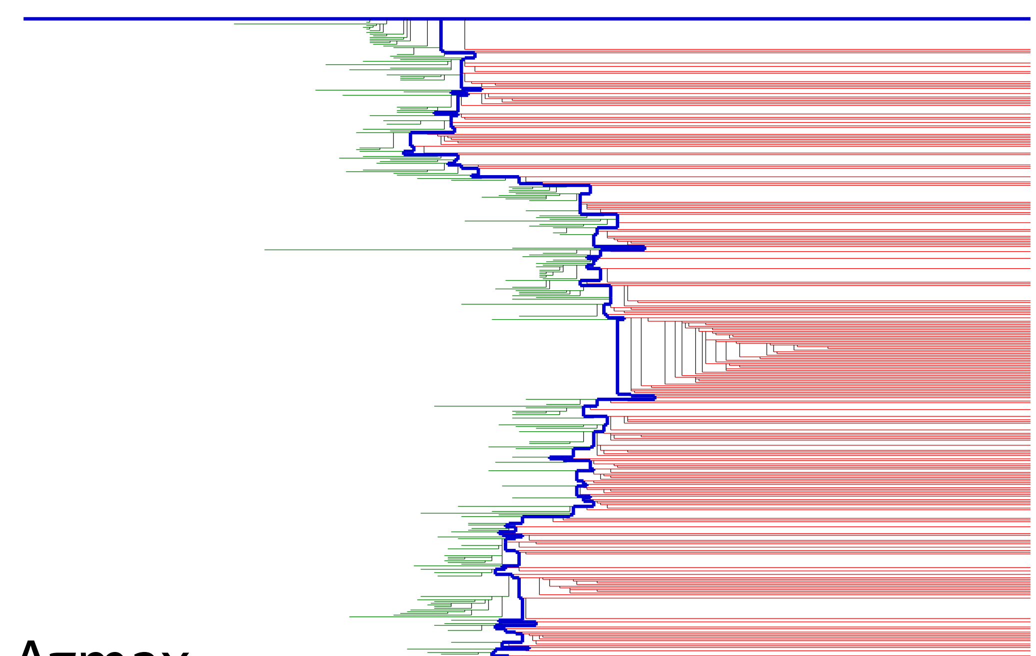
- spring shooting algorithm:



$$P^{sp}_{acc}[\tau \to \tau') = \min[1, e^{sk\Delta\tau}]$$

pro: much better decorrelation

con: need optimisation of k and Δτmax



good decorrelation

Z.F. Brotzakis, PGB, JCP (2016)

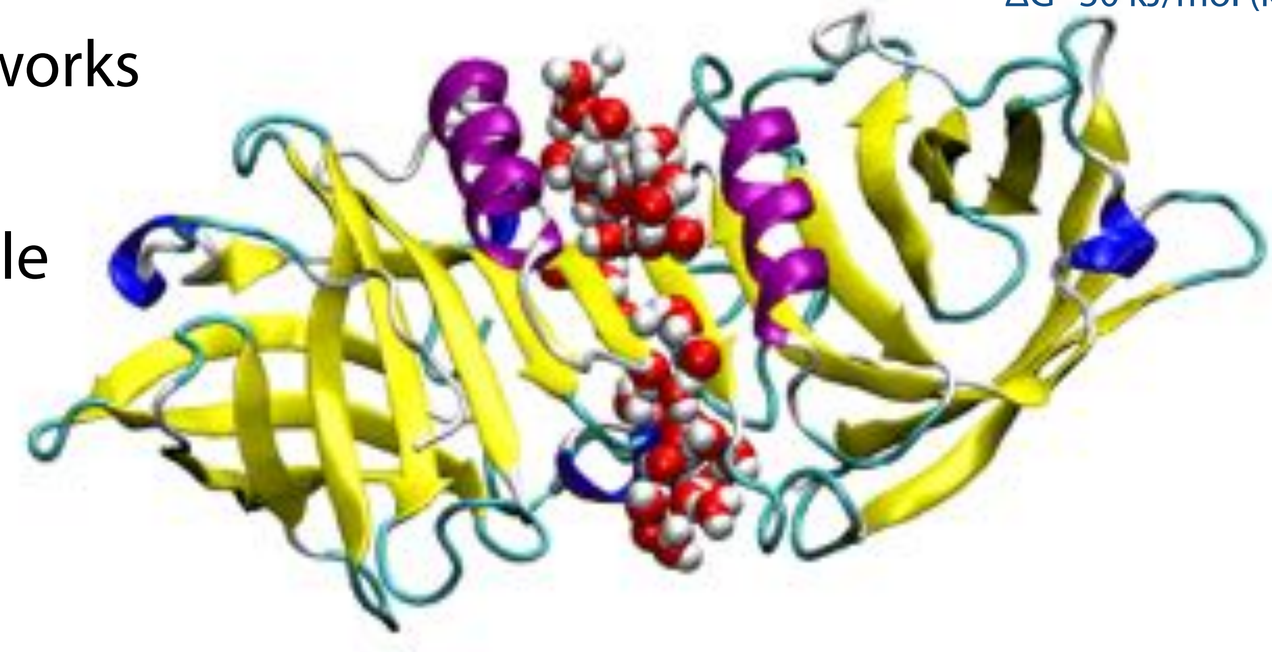
Protein dissociation

System: 65000 atoms,
AMBER99SB-ILDM
T=300K, P=1 atm,
ΔG~30 kJ/mol (k_{off}< 0.1 s^{-1})

important for signalling, regulation networks

TPS crucial to simulate on molecular scale

system: beta-lactoglobulin dimer
- important for food industry
- forms dimer in native state.



*Z.F. Brotzakis, PGB, JPCB **123**, 1883 (2019)*

Protein dissociation

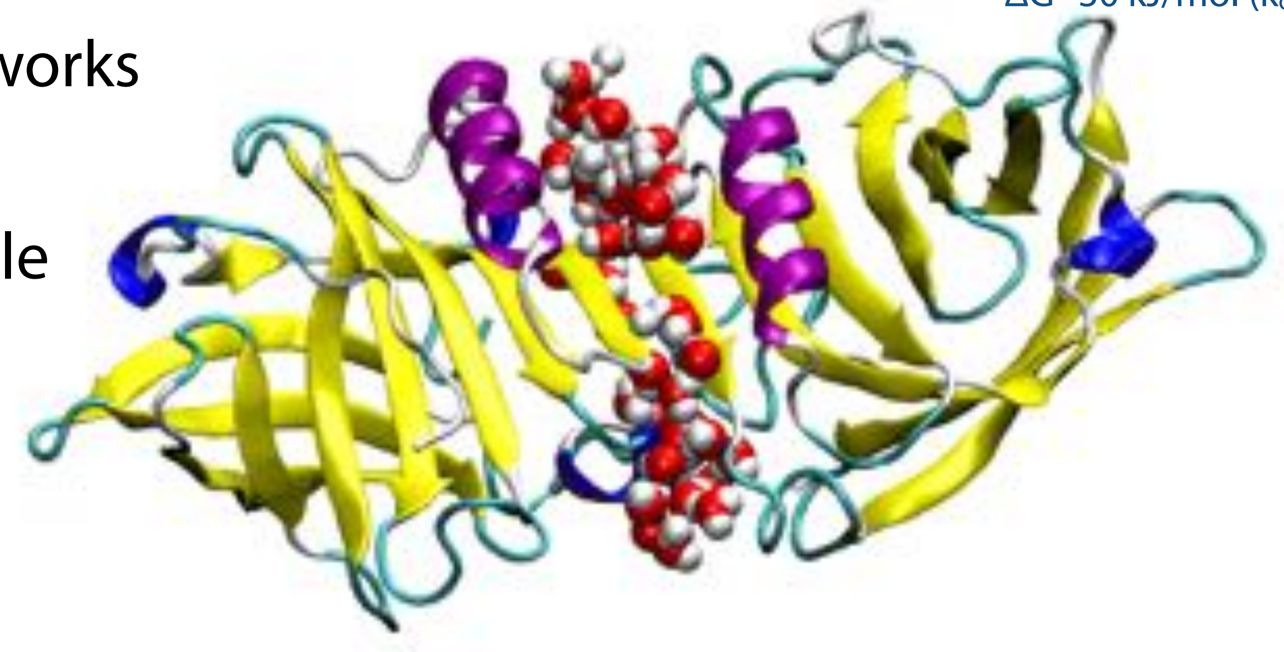
System: 65000 atoms,
AMBER99SB-ILDM
T=300K, P=1 atm,
ΔG~30 kJ/mol (k_{off}< 0.1 s^{-1})

important for signalling, regulation networks

TPS crucial to simulate on molecular scale

system: beta-lactoglobulin dimer

- important for food industry
- forms dimer in native state.



*Z.F. Brotzakis, PGB, JPCB **123**, 1883 (2019)*

Protein dissociation

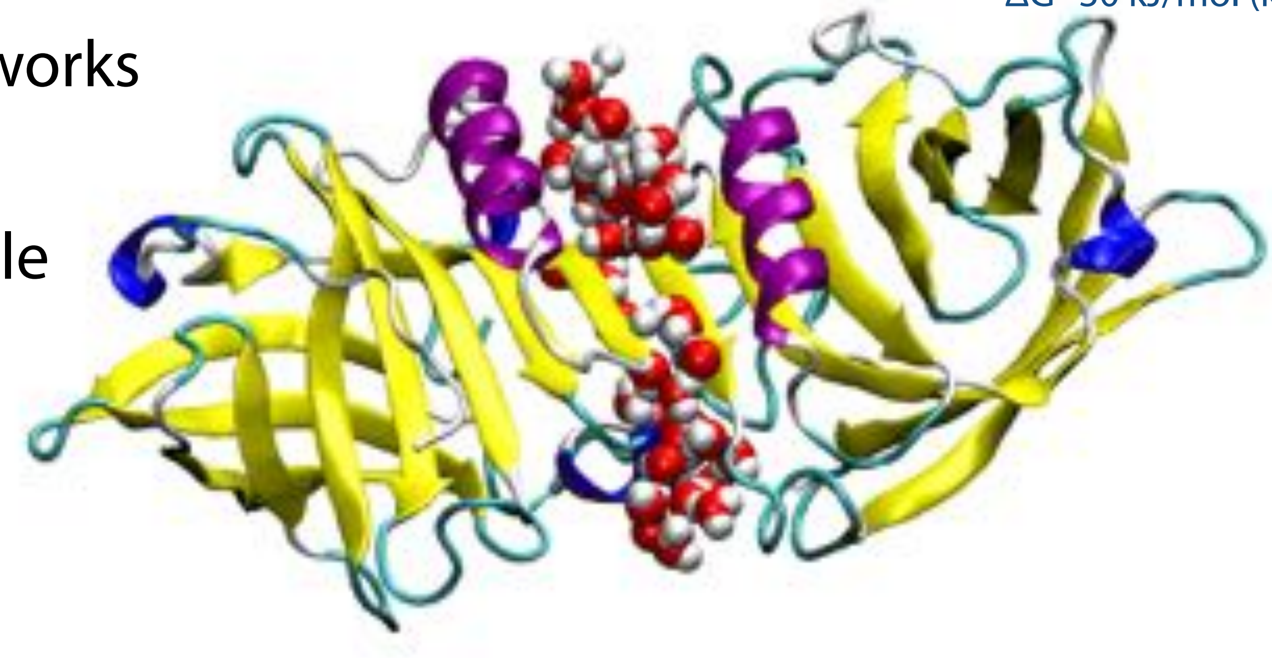
System: 65000 atoms,
AMBER99SB-ILDM
T=300K, P=1 atm,
ΔG~30 kJ/mol (k_{off}< 0.1 s^{-1})

important for signalling, regulation networks

TPS crucial to simulate on molecular scale

system: beta-lactoglobulin dimer

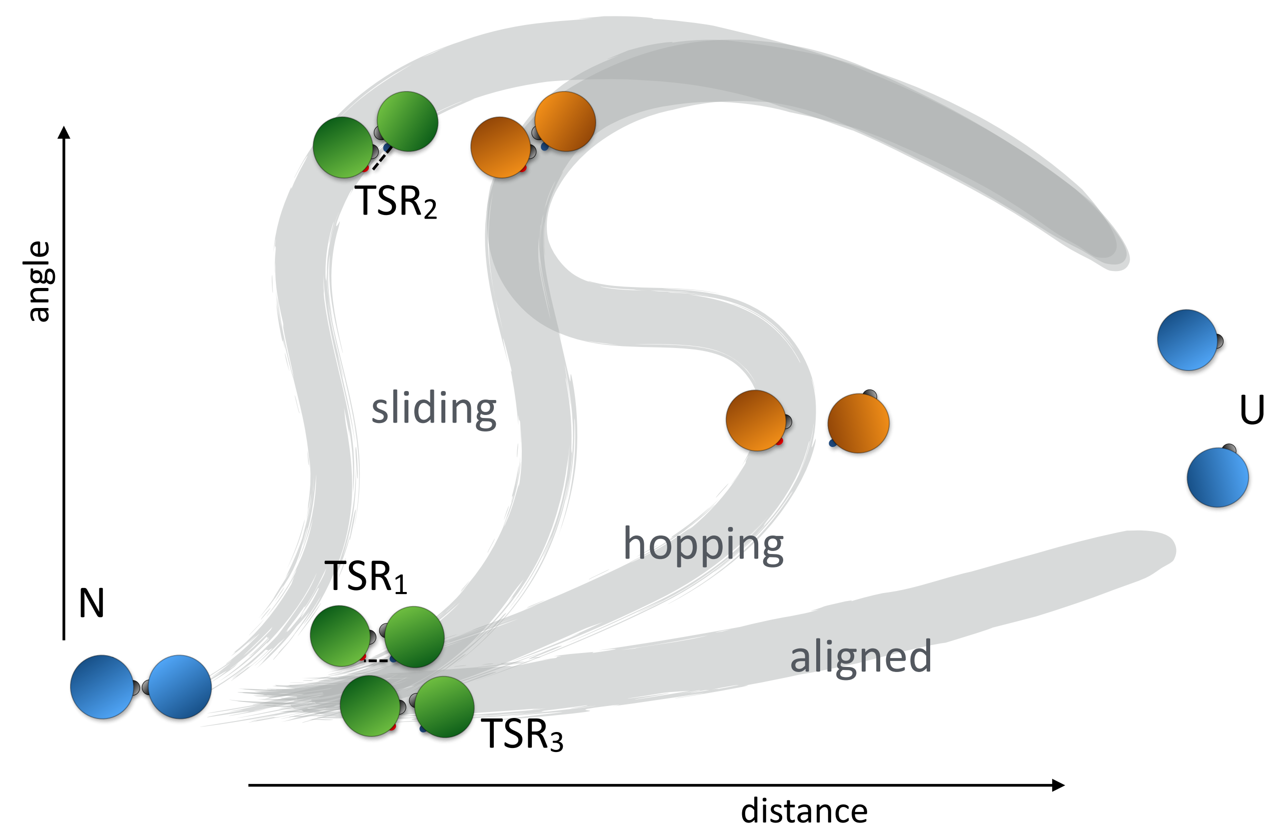
- important for food industry
- forms dimer in native state.



After sampling 100's of trajectories

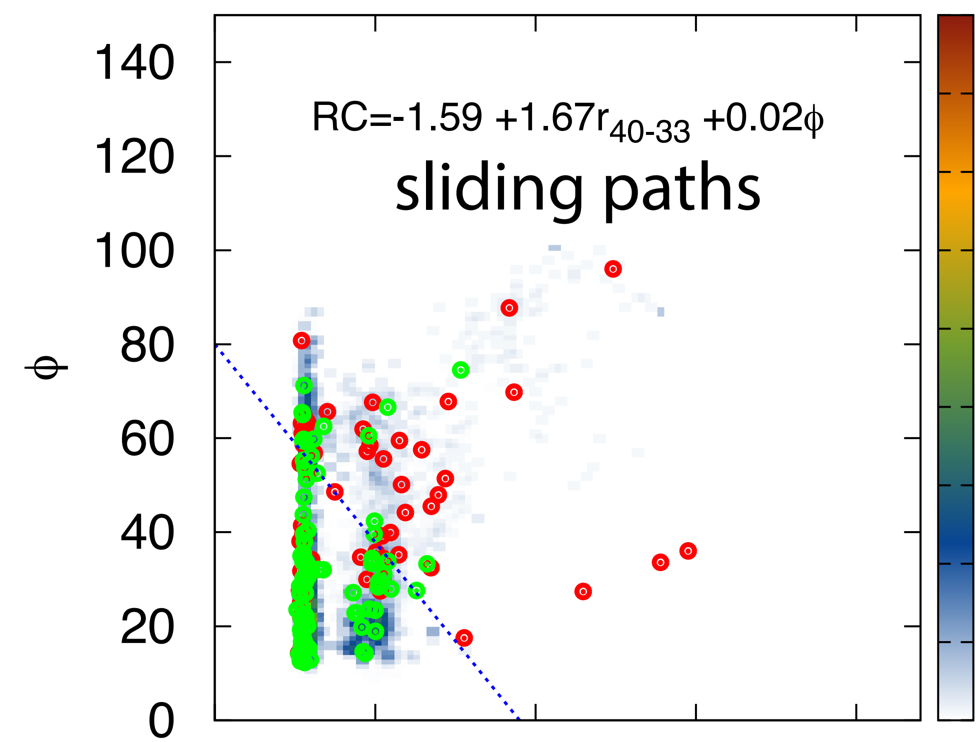
we find several mechanisms

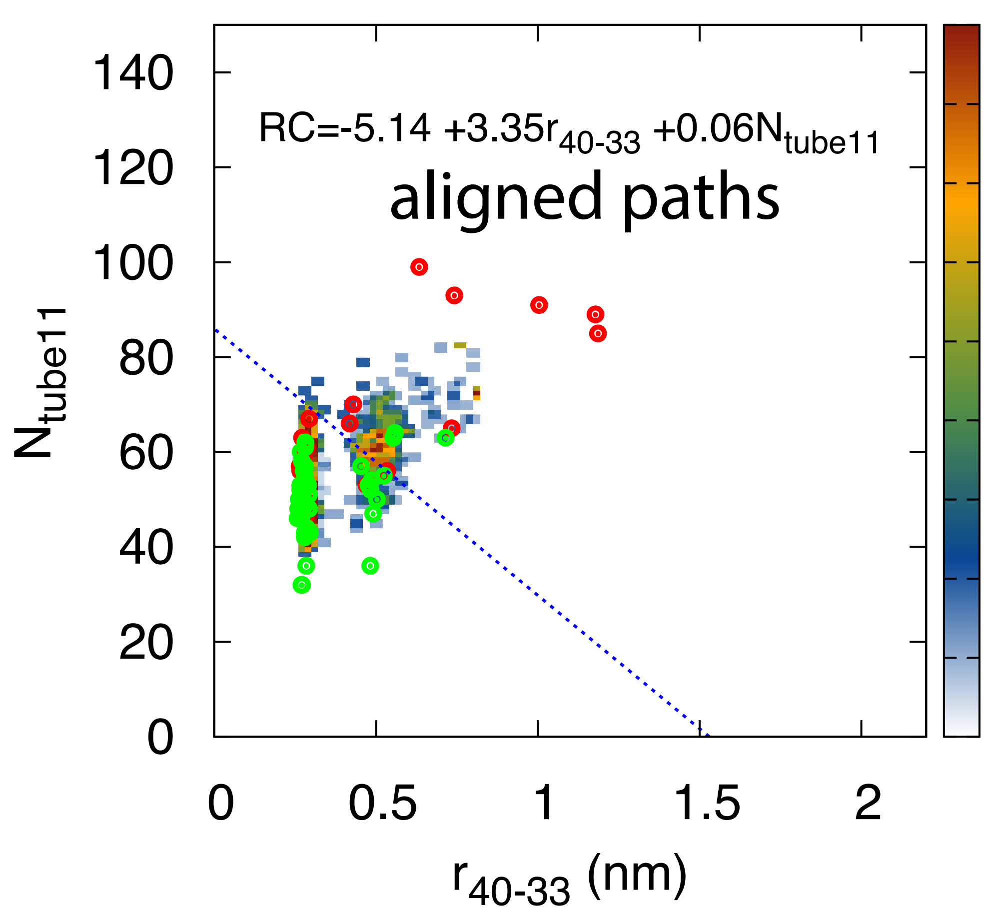
how do we understand molecular
nature of the mechanism
and identify transition states



*Z.F. Brotzakis, PGB, JPCB **123**, 1883 (2019)*

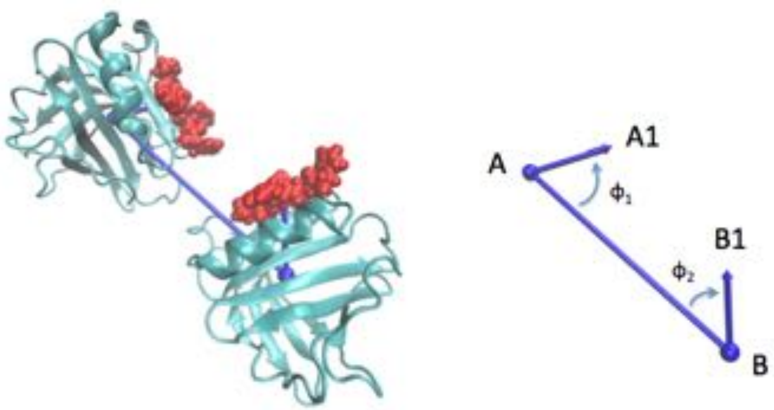
Reaction coordinate analysis





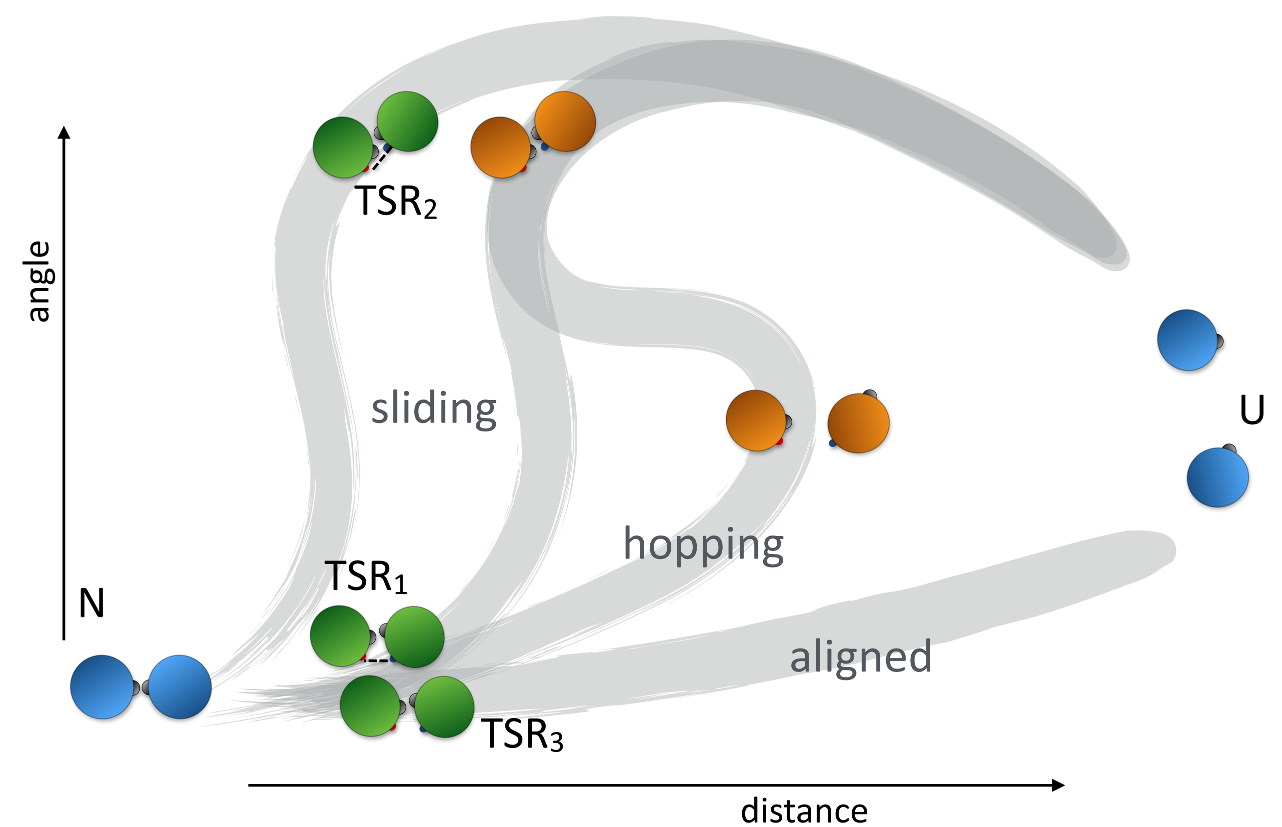
important ingredients sliding paths:

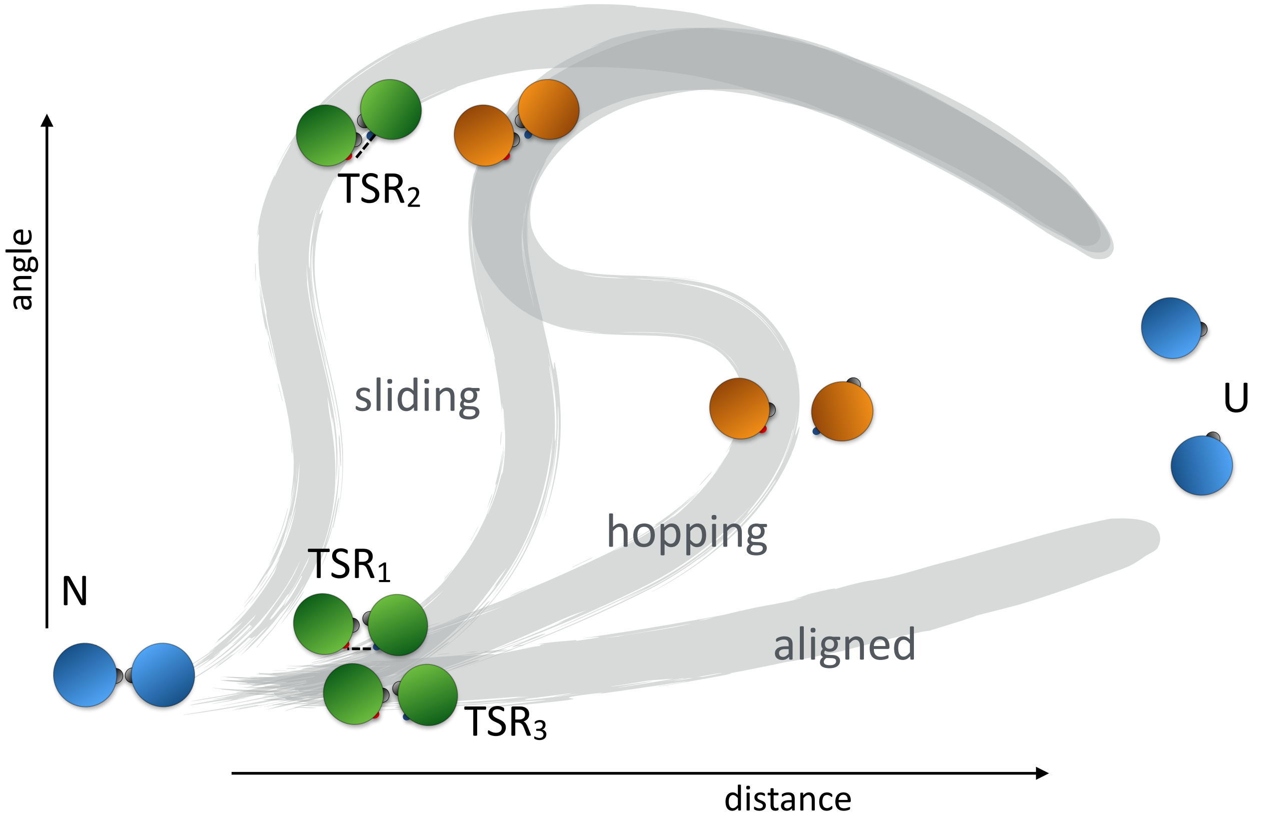
- salt bridge R40-D33
- angle phi



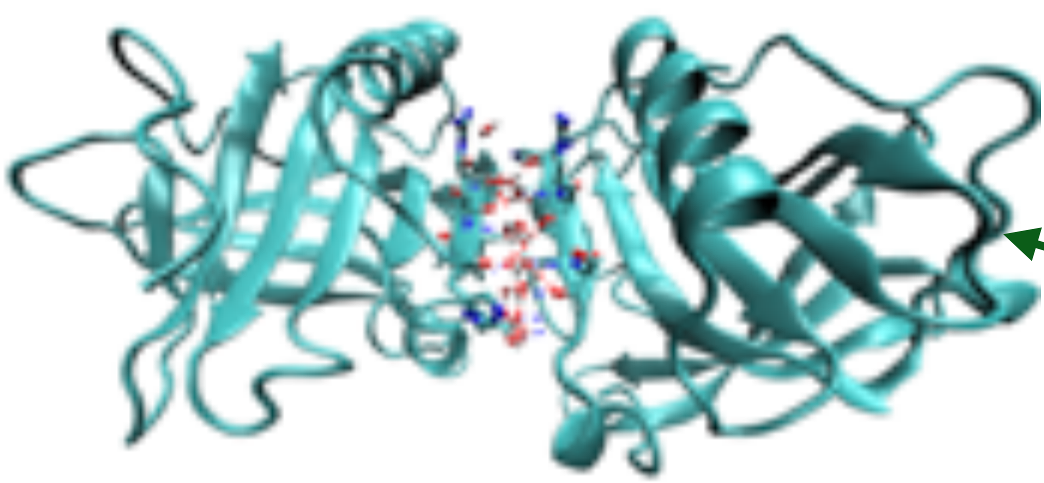
aligned paths:

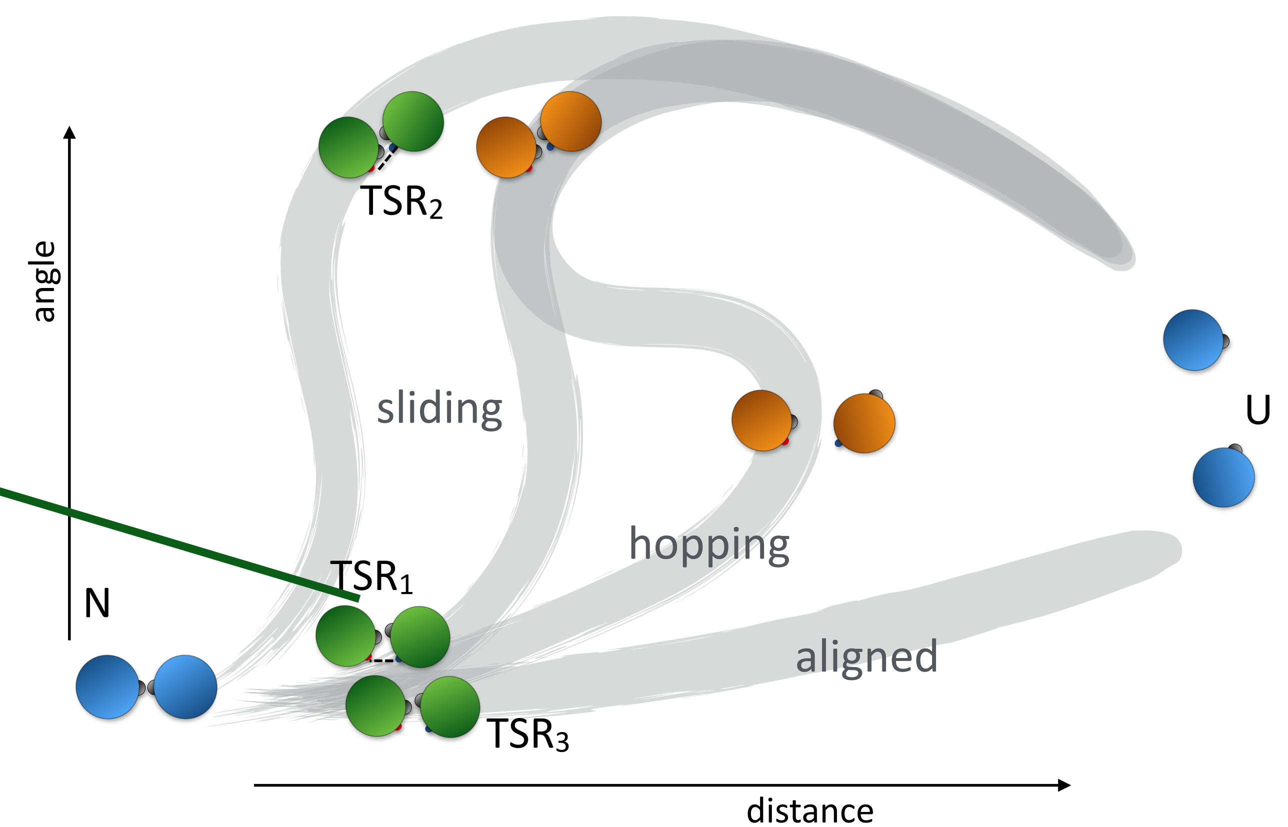
- salt bridge
- number of waters between proteins



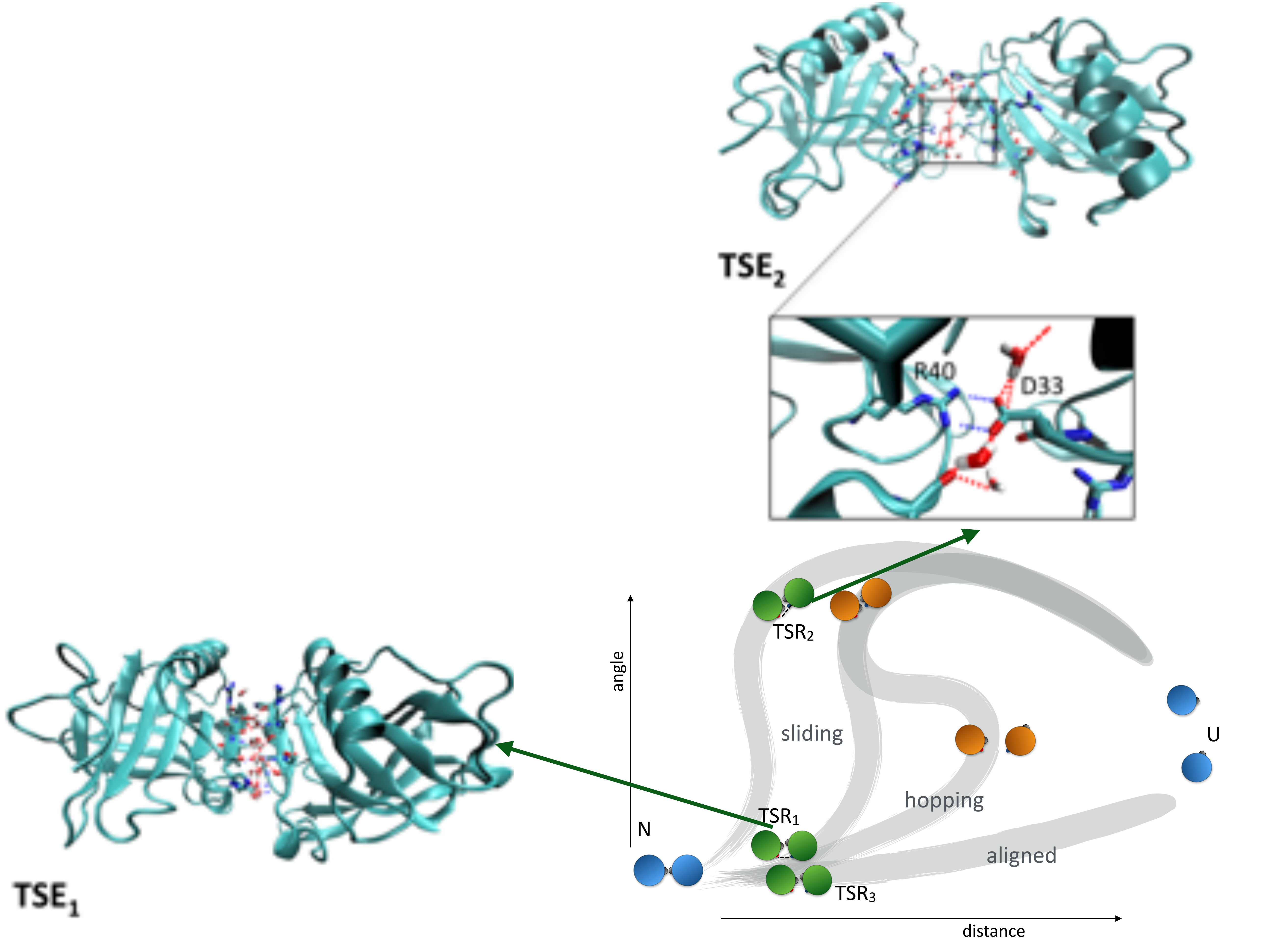


Z.F. Brotzakis, PGB, JPCB ***123****, 1883 (2019)*

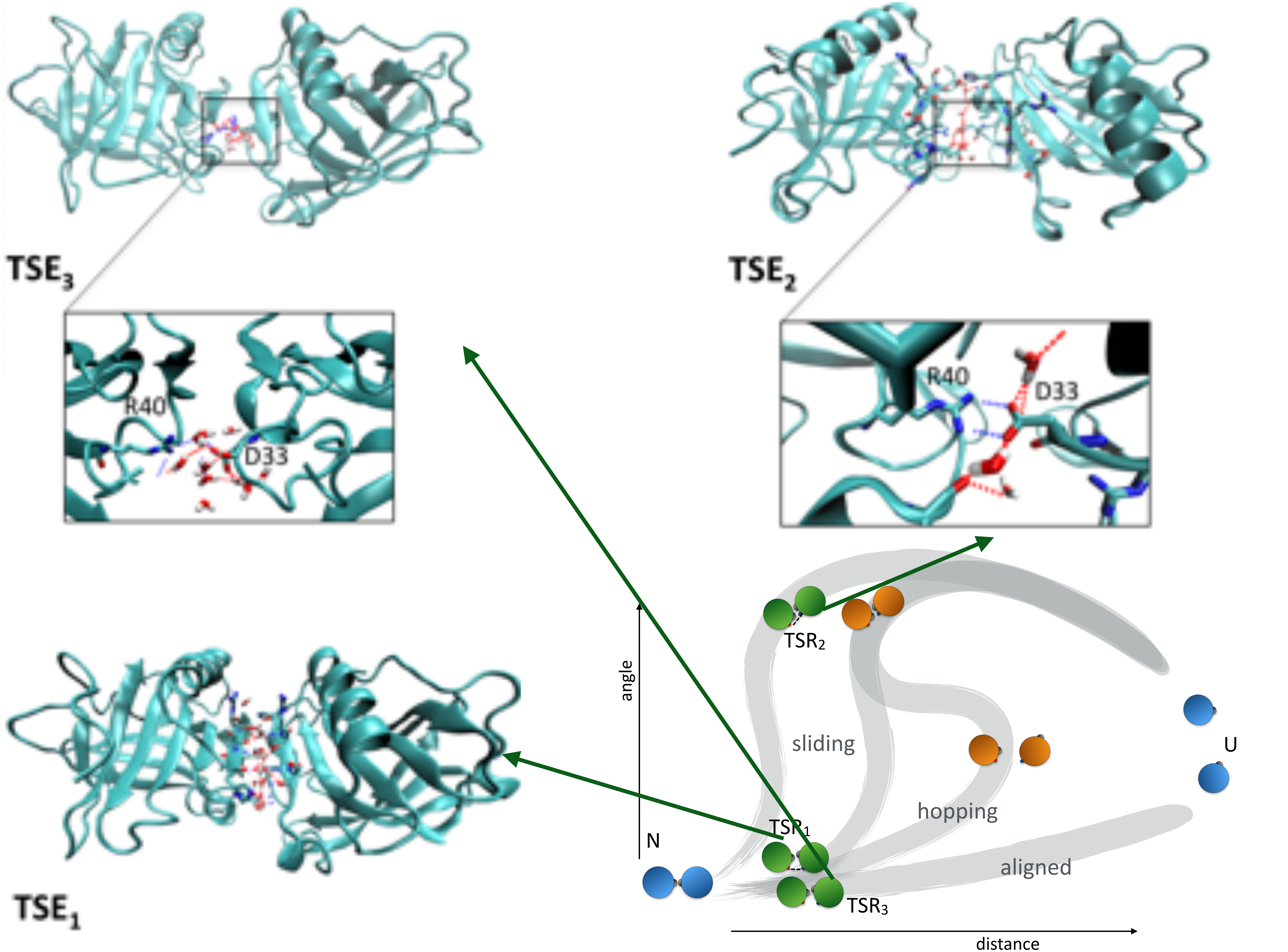

TSE$_1$



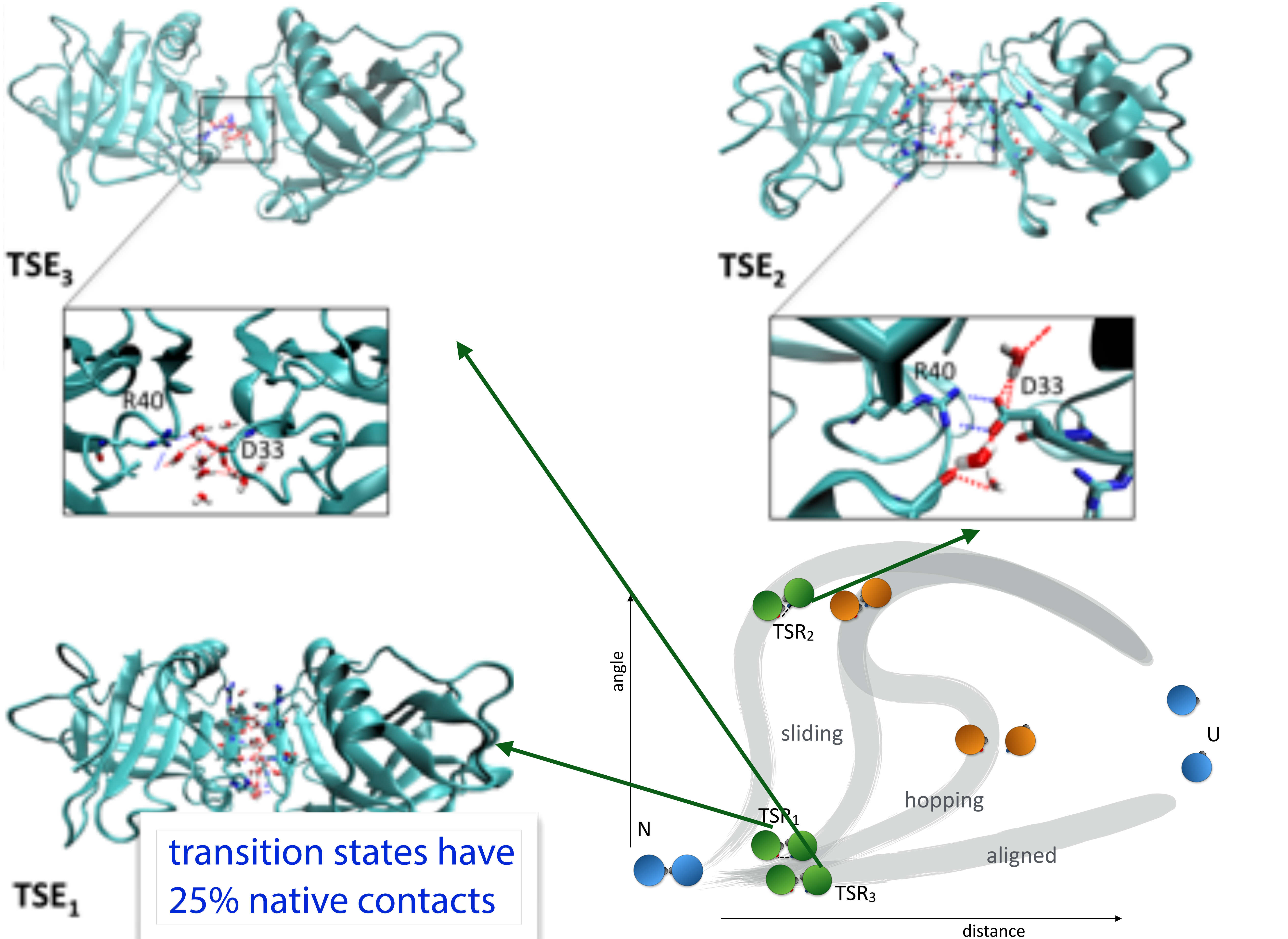
*Z.F. Brotzakis, PGB, JPCB **123**, 1883 (2019)*



Z.F. Brotzakis, PGB, JPCB ***123****, 1883 (2019)*

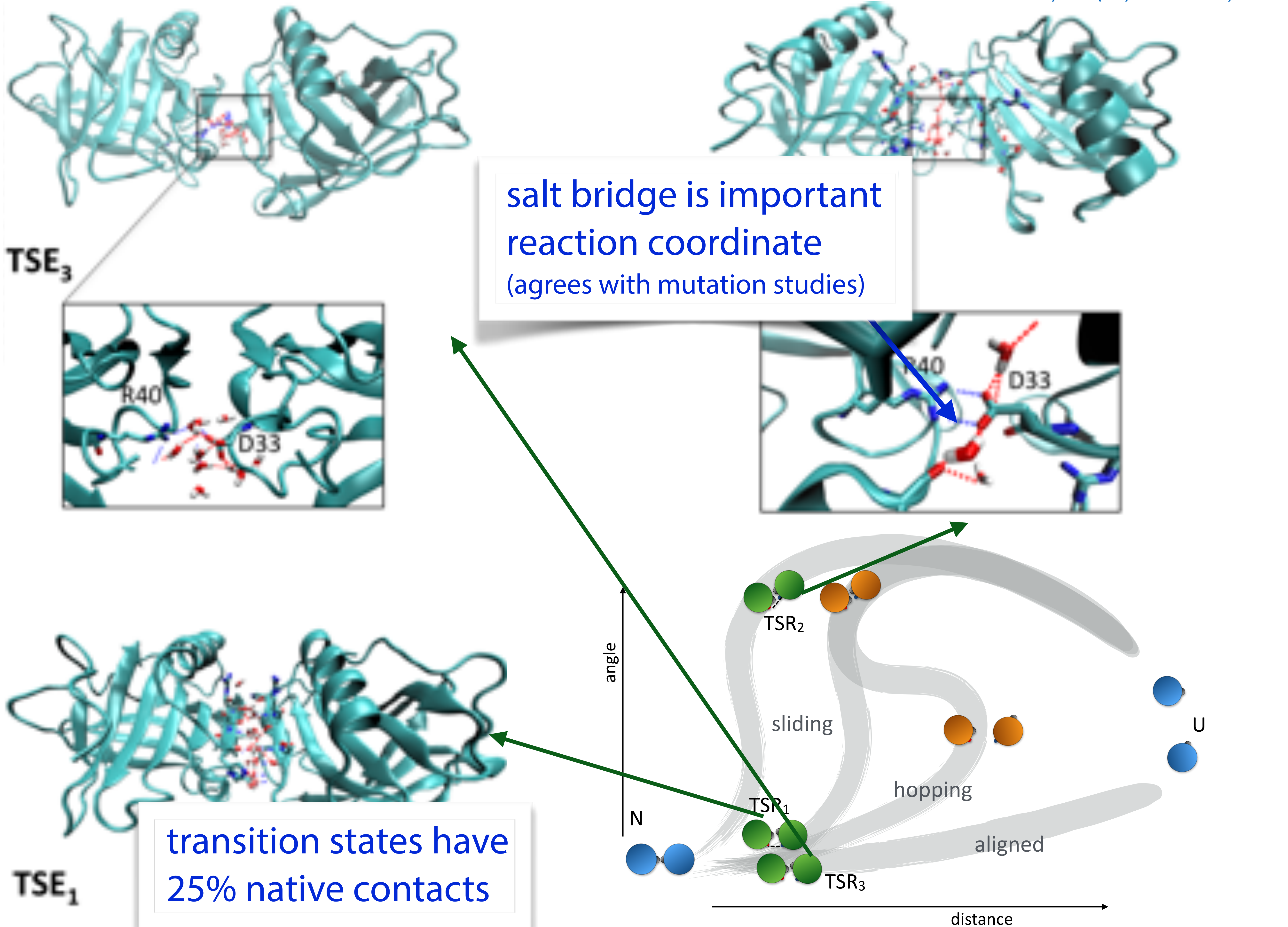


Z.F. Brotzakis, PGB, JPCB ***123****, 1883 (2019)*



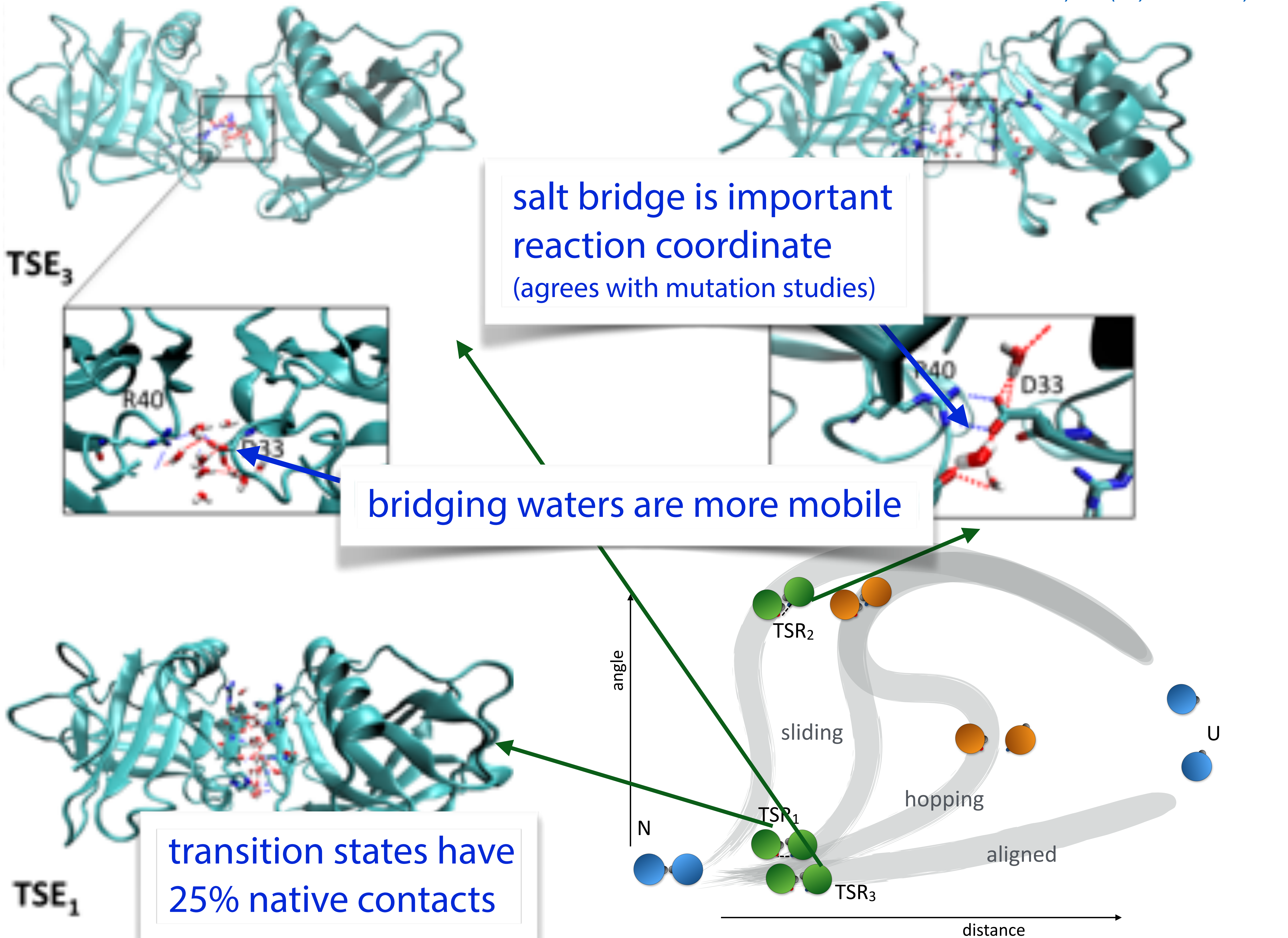
Z.F. Brotzakis, PGB, JPCB ***123***, *1883 (2019)*

Sakurai K and Goto Y. J. Biol. Chem., 277(28):25735–40, 2002.



*Z.F. Brotzakis, PGB, JPCB **123**, 1883 (2019)*

Sakurai K and Goto Y. J. Biol. Chem., 277(28):25735–40, 2002.



Z.F. Brotzakis, PGB, JPCB ***123****, 1883 (2019)*

Outline

- Introduction
- Rare events

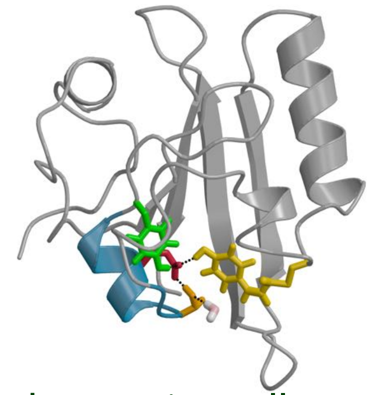
part 1:

- Transition Path Sampling
- Committor & Reaction coordinate analysis
- Rate constants with transition interface sampling
- reaction networks with multiple state TPS/TIS
- advanced developments & machine learning
- OPS software

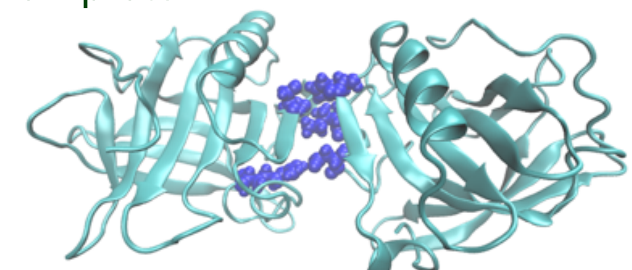
part 2:

- imposing kinetic constraints
- path reweighting with Maximum Caliber

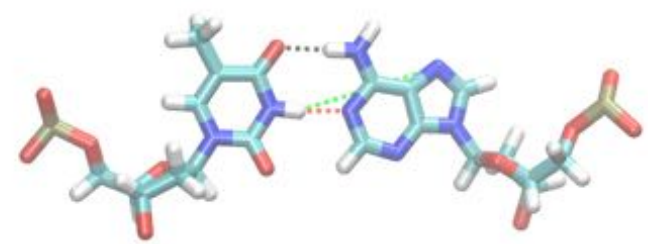
- conclusions



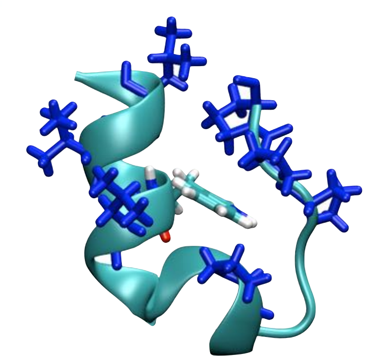
photoactive yellow protein



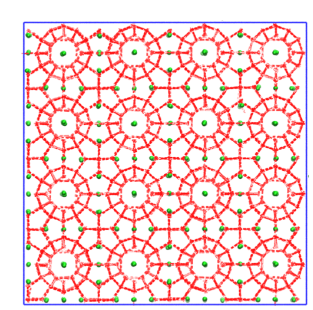
protein dissociation



DNA base pair rotation



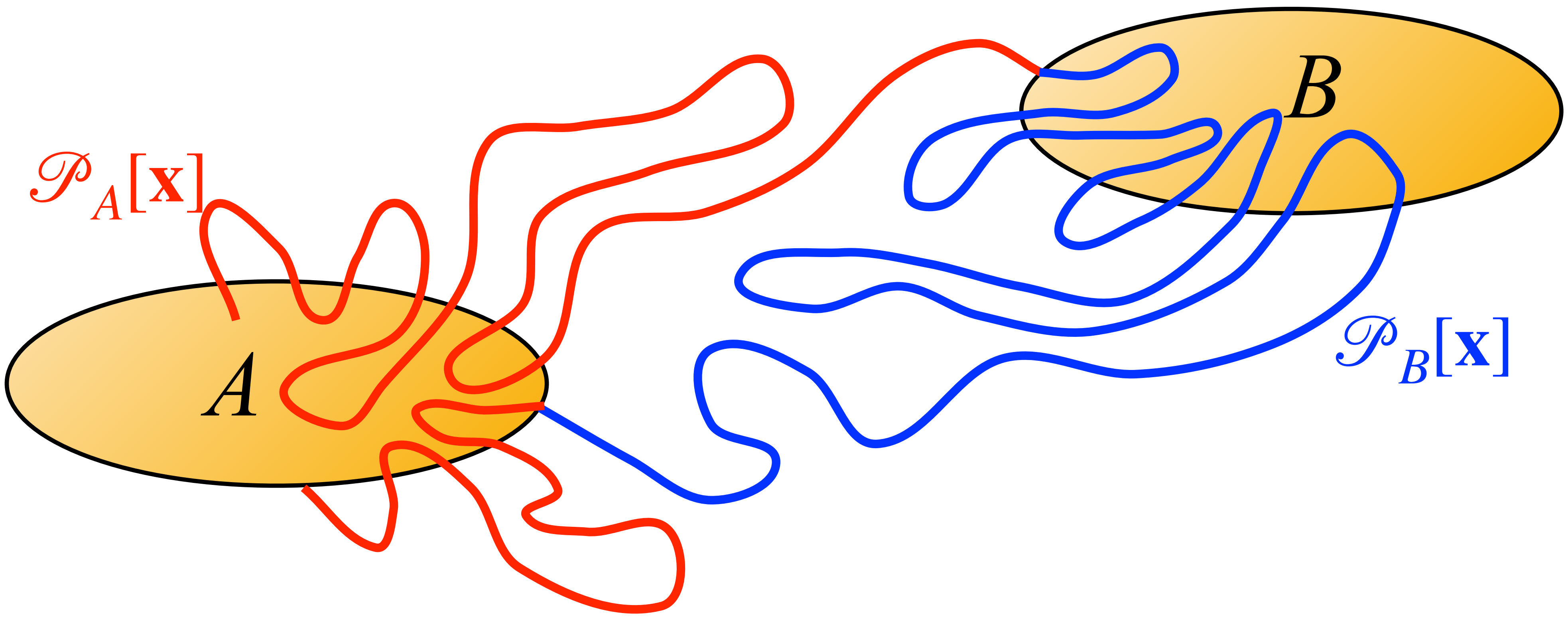
Trp cage folding



gas hydrate formation

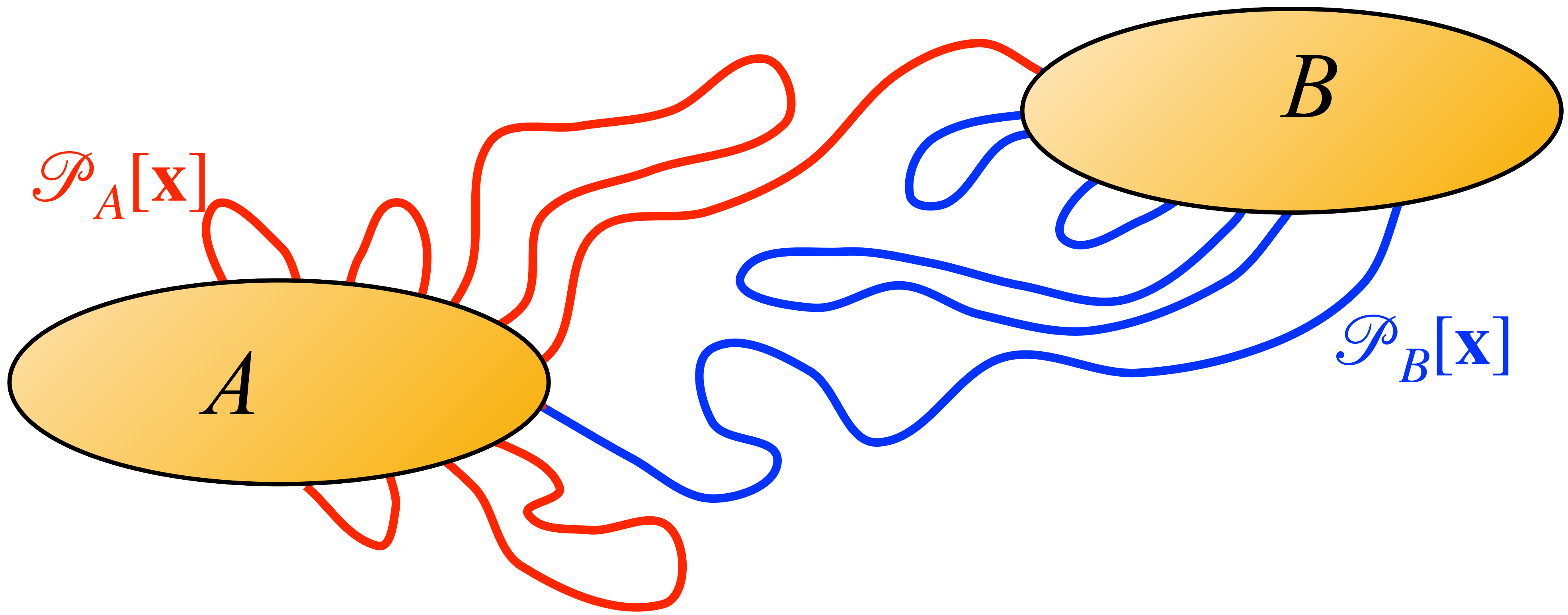
Kinetics: Transition interface sampling

van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



Kinetics: Transition interface sampling

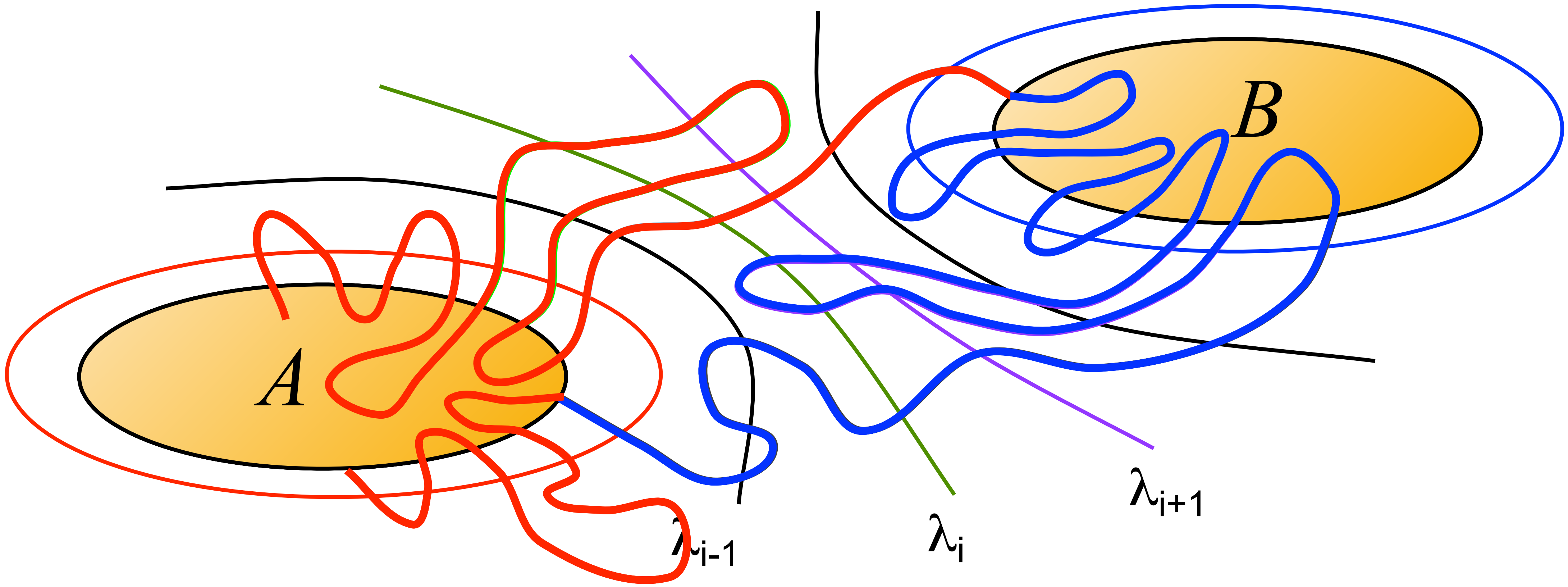
van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

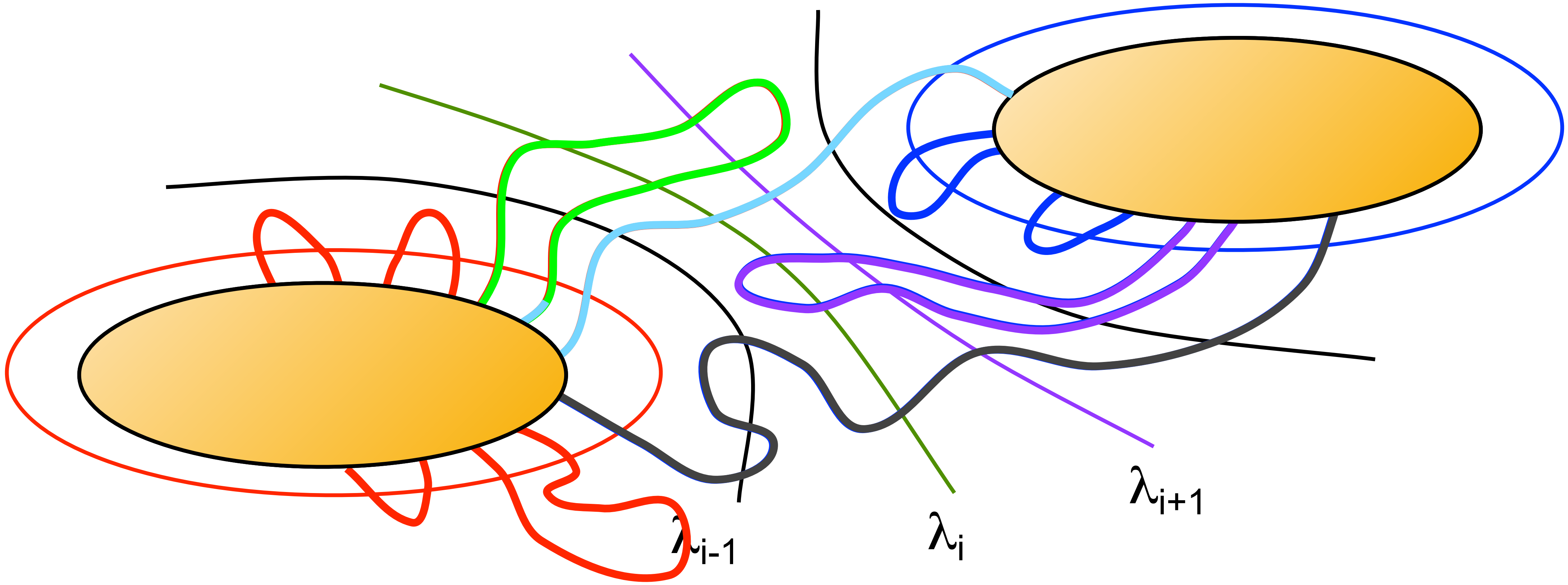
van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

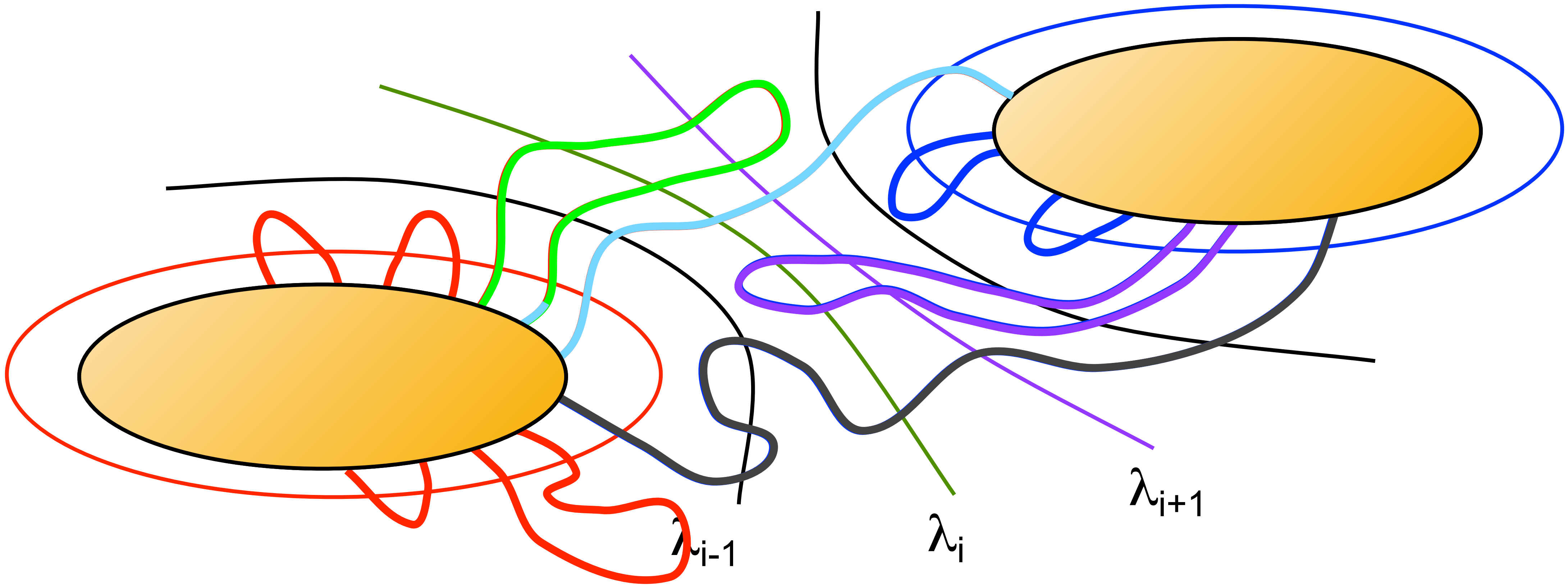
van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

compute $P_A(\lambda_{i+1} | \lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

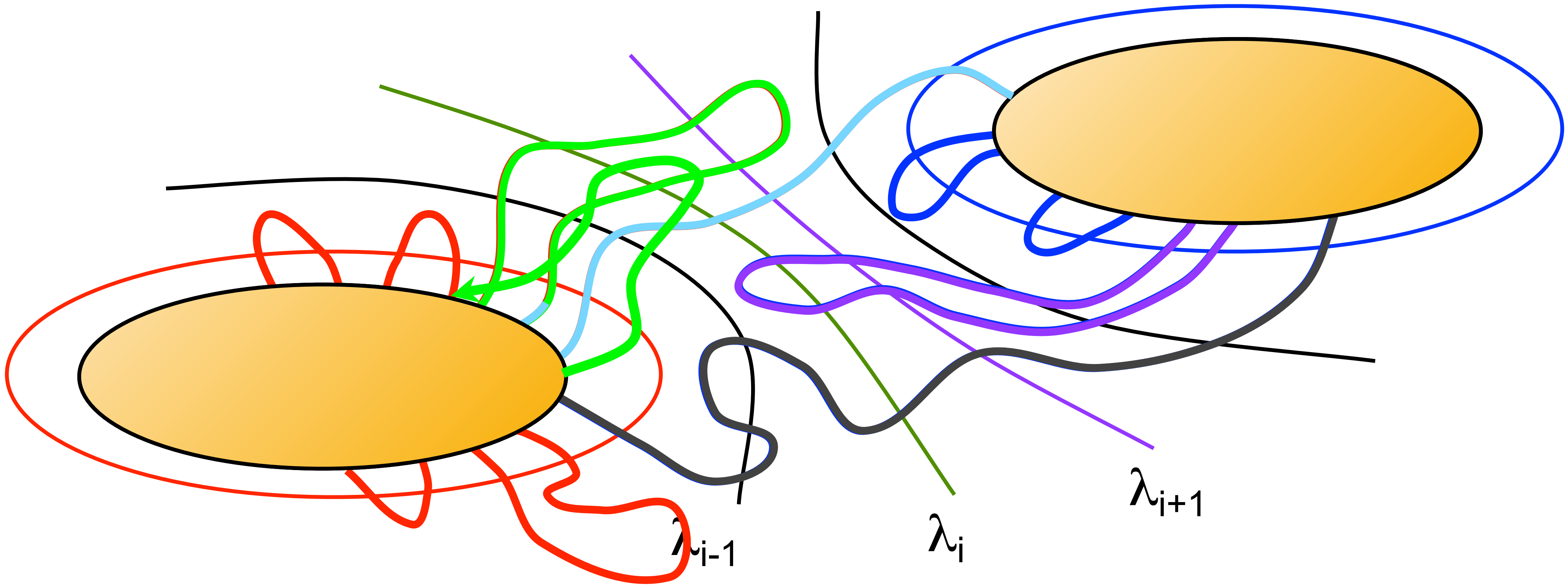
$$k_{AB} = \phi_0 P_A(\lambda_B | \lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

ϕ_0 is flux through initial state interface λ_0

Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118***, *7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

compute $P_A(\lambda_{i+1} | \lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

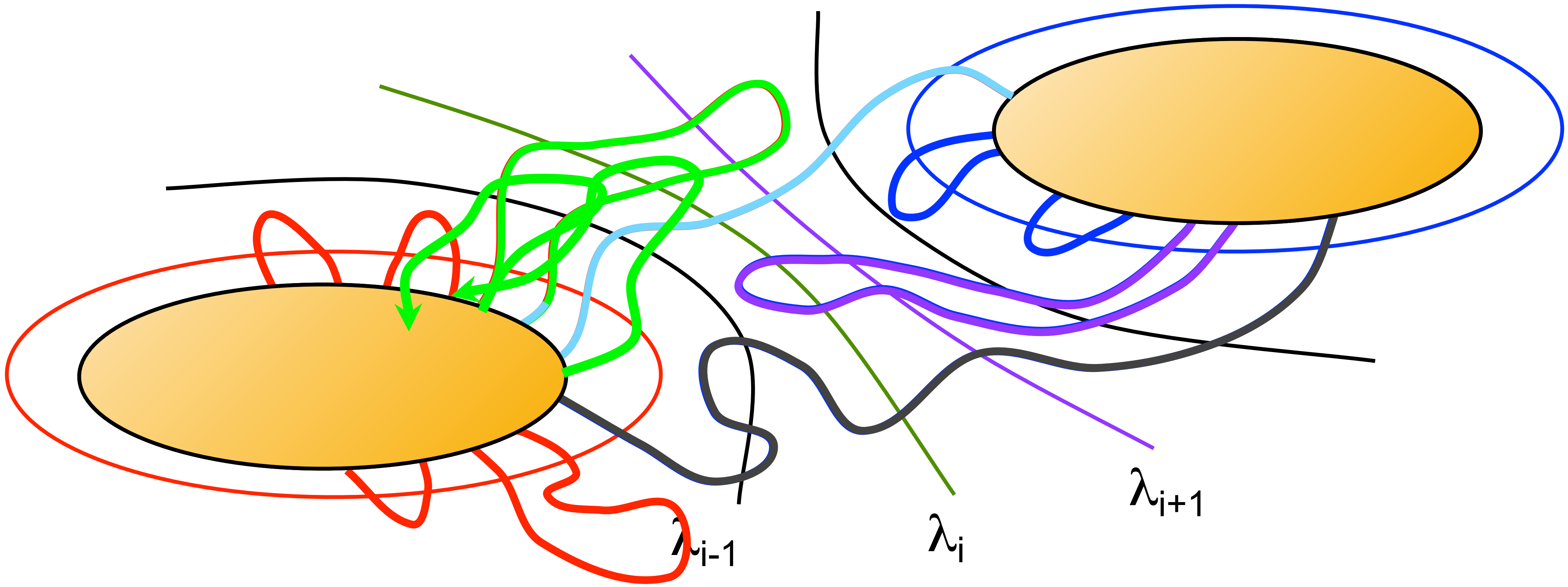
$$k_{AB} = \phi_0 P_A(\lambda_B | \lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

ϕ_0 is flux through initial state interface λ_0

Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

compute $P_A(\lambda_{i+1} | \lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

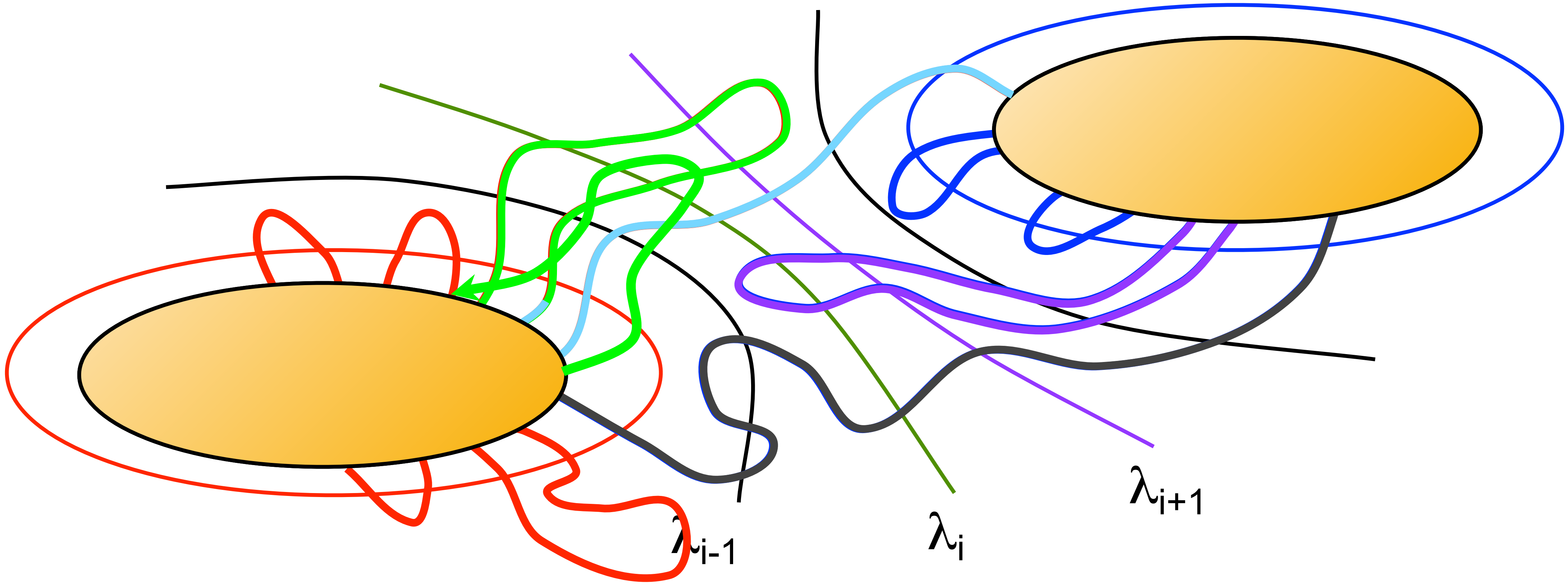
$$k_{AB} = \phi_0 P_A(\lambda_B | \lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

ϕ_0 is flux through initial state interface λ_0

Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118***, *7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

compute $P_A(\lambda_{i+1} | \lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

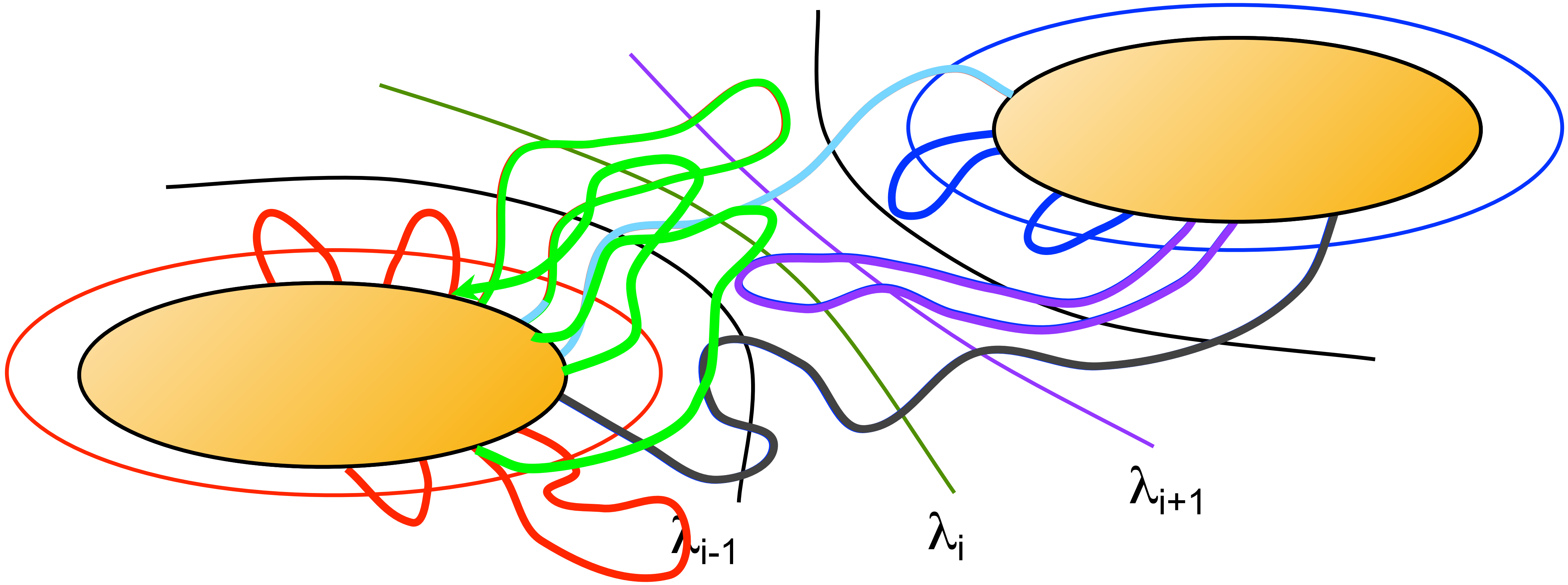
$$k_{AB} = \phi_0 P_A(\lambda_B | \lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

ϕ_0 is flux through initial state interface λ_0

Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118*** *, 7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

compute $P_A(\lambda_{i+1} | \lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

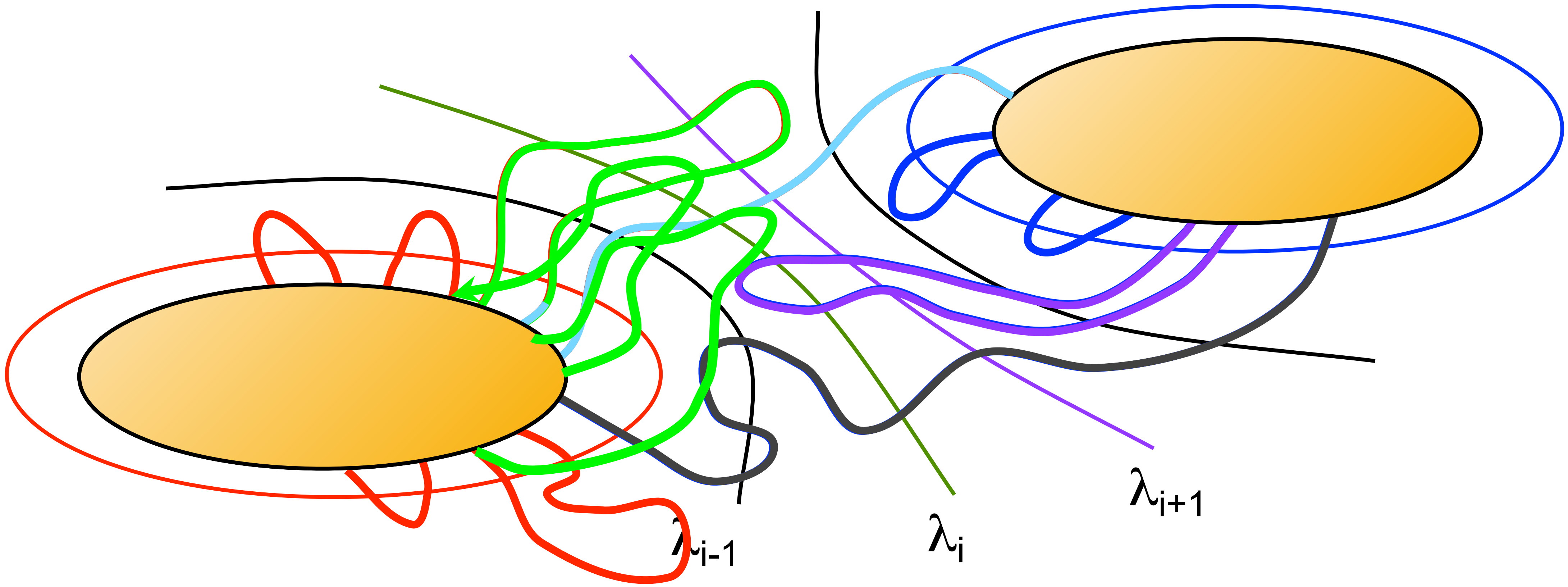
$$k_{AB} = \phi_0 P_A(\lambda_B | \lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

ϕ_0 is flux through initial state interface λ_0

Kinetics: Transition interface sampling

Introduce set of interfaces λ_i

van Erp, Moroni, P.GB, J. Chem. Phys. ***118***, *7762 (2003)*
Cabriolu, Skjelbred Refsnes, PGB, van Erp JCP ***147,*** *152722 (2019)*



for each interface i **sample** pathways that cross λ_i with flexible shooting move

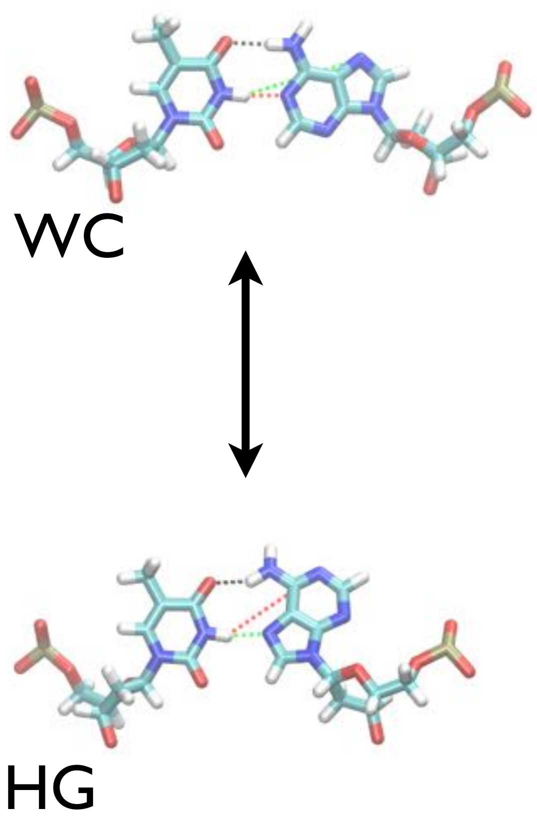
compute $P_A(\lambda_{i+1}|\lambda_i)$ = probability that path crossing λ_i for first time after leaving A reaches λ_{i+1}

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0) = \phi_0 \prod_{i=0}^{n-1} P_A(\lambda_{i+1}|\lambda_i)$$

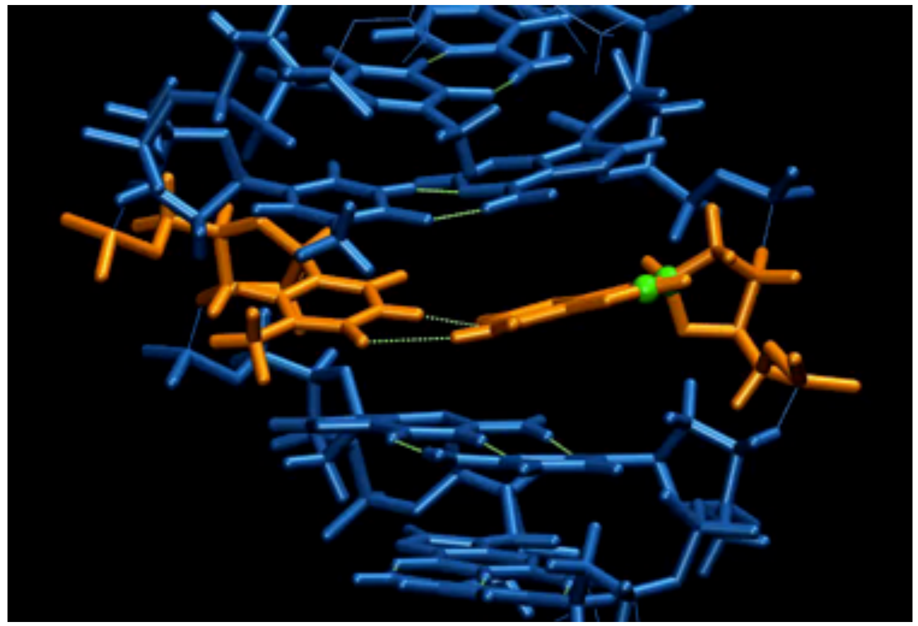
ϕ_0 is flux through initial state interface λ_0

Yields exact rates independent on λ
(note: also basis of FFS)

DNA Hoogsteen base pair formation

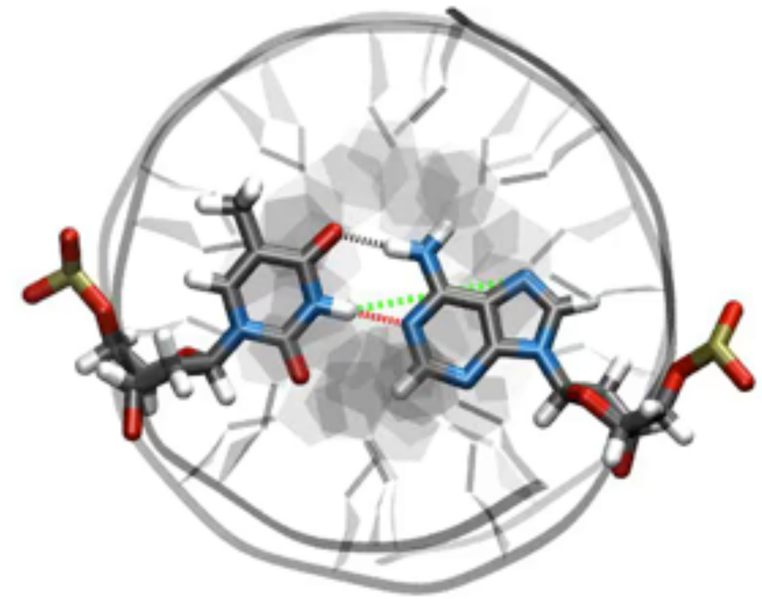


path by Conjugate peak refinement



Nikolova, Kim, Wise, O'Brien, Andricioaei and Al-Hashimi Nature 470, 498 (2011)

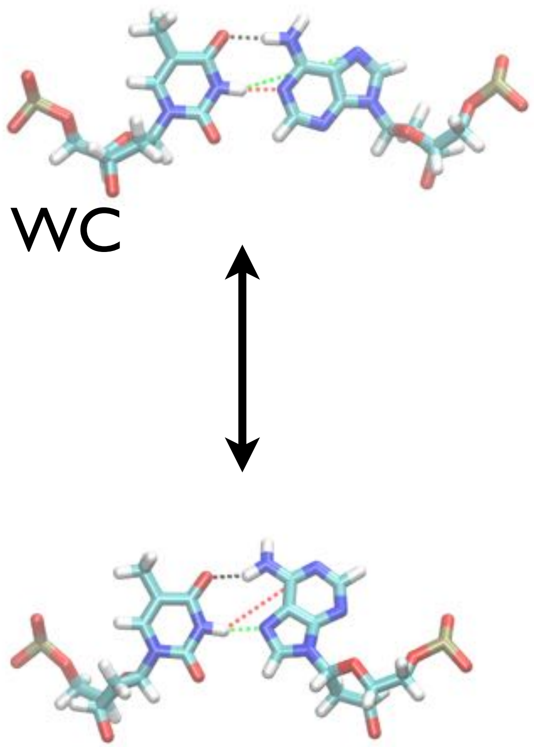
path by Transition path sampling



AMBER03 force field & parmbsc0/1

Vreede, Pérez de Alba Ortíz, PGB Swenson, NAR, 47, 11069 (2019)

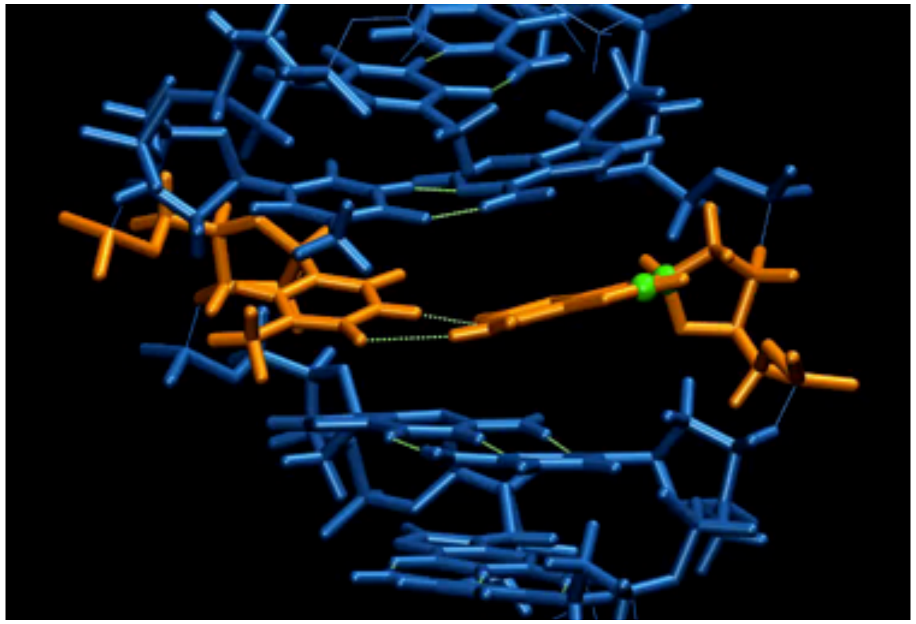
DNA Hoogsteen base pair formation



WC

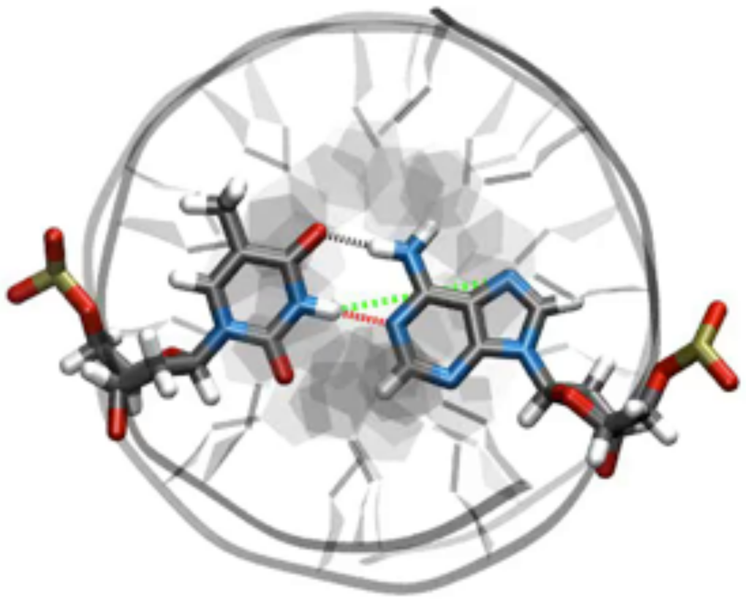
HG

path by Conjugate peak refinement



Nikolova, Kim, Wise, O'Brien, Andricioaei and Al-Hashimi Nature 470, 498 (2011)

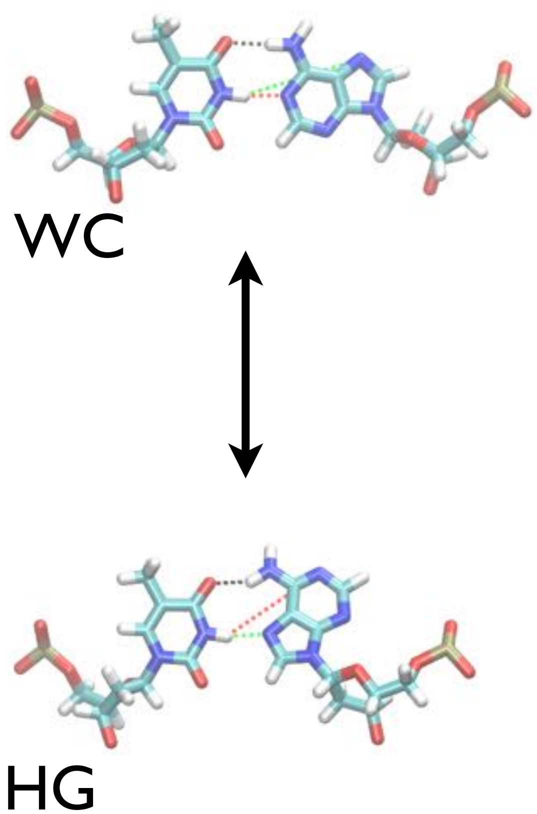
path by Transition path sampling



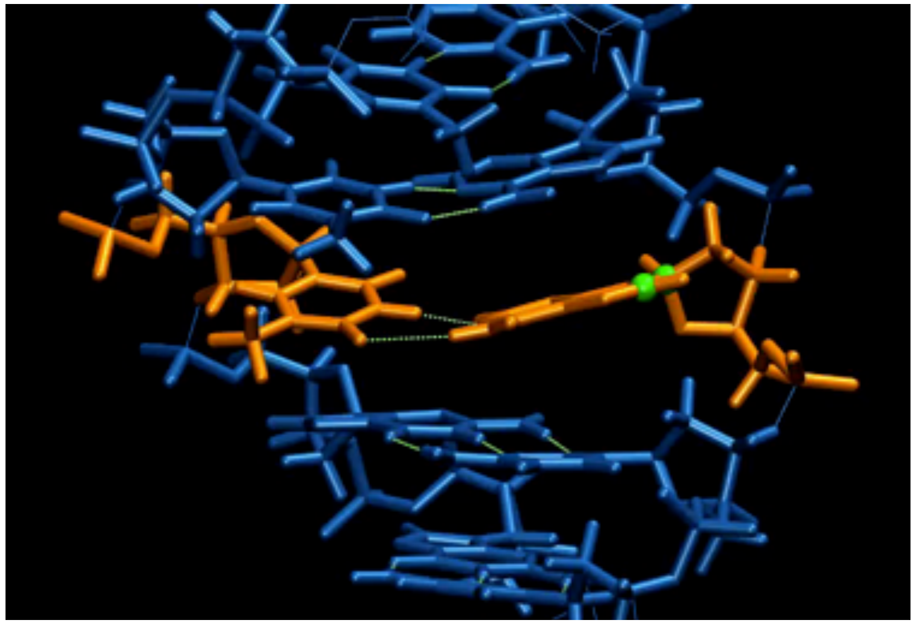
AMBER03 force field & parmbsc0/1

Vreede, Pérez de Alba Ortíz, PGB Swenson, NAR, 47, 11069 (2019)

DNA Hoogsteen base pair formation

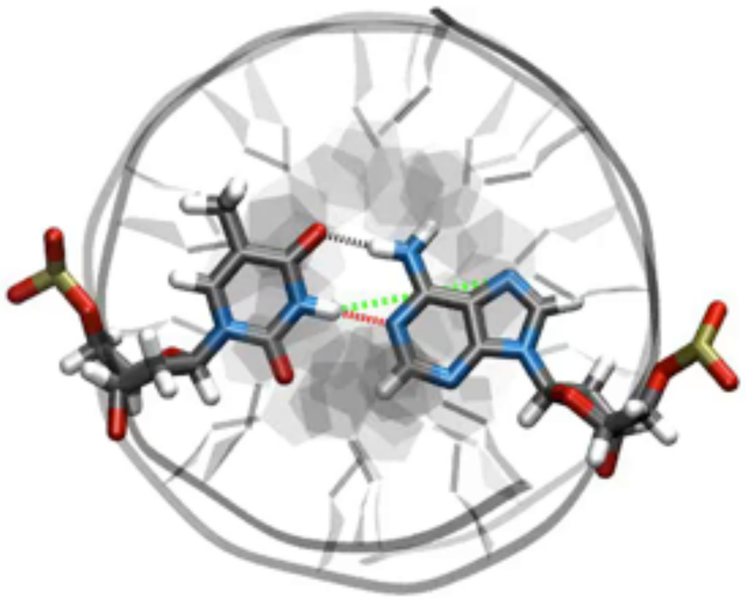


path by Conjugate peak refinement



Nikolova, Kim, Wise, O'Brien, Andricioaei and Al-Hashimi Nature 470, 498 (2011)

path by Transition path sampling



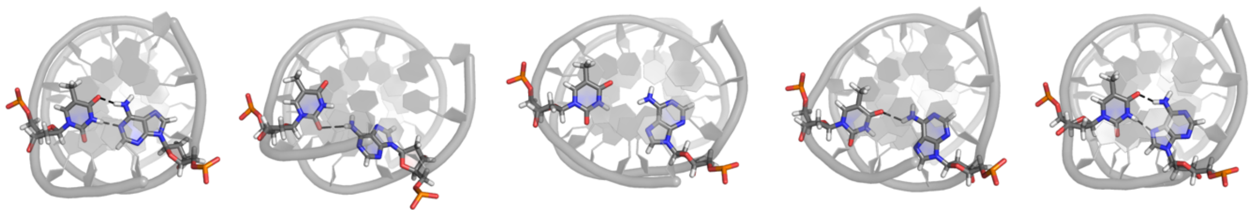
AMBER03 force field & parmbsc0/1

Vreede, Pérez de Alba Ortíz, PGB Swenson, NAR, 47, 11069 (2019)

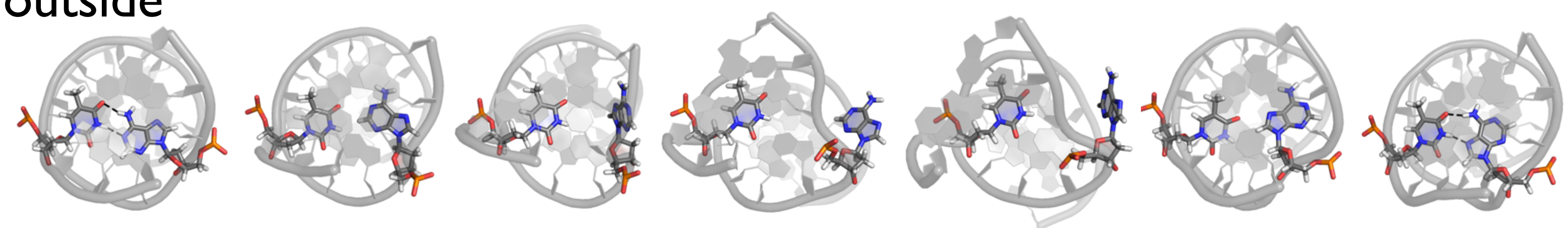
DNA Hoogsteen base pair formation

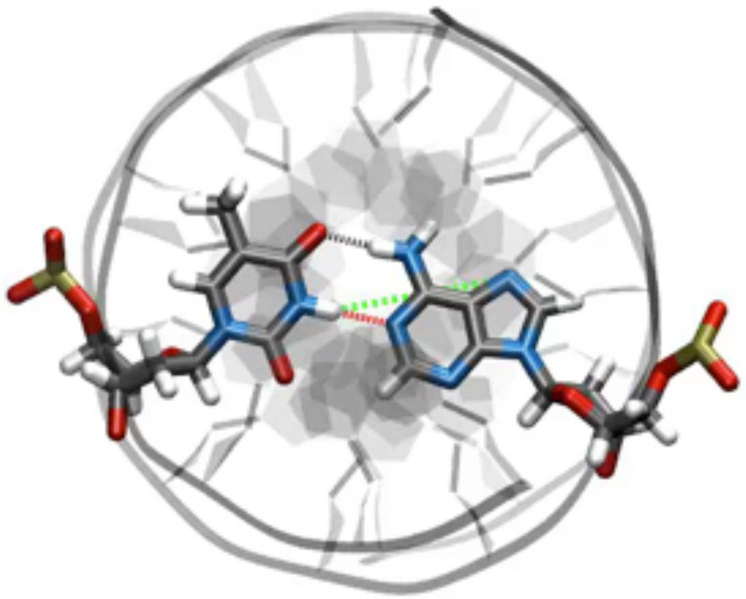
path by Transition path sampling

inside



outside





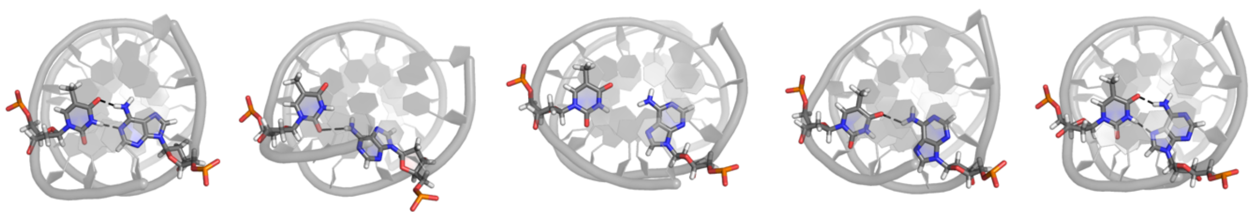
AMBER03 force field &
parmbsc0/1

Vreede,Pérez de Alba Ortíz,PGB Swenson, NAR, 47,11069 (2019)

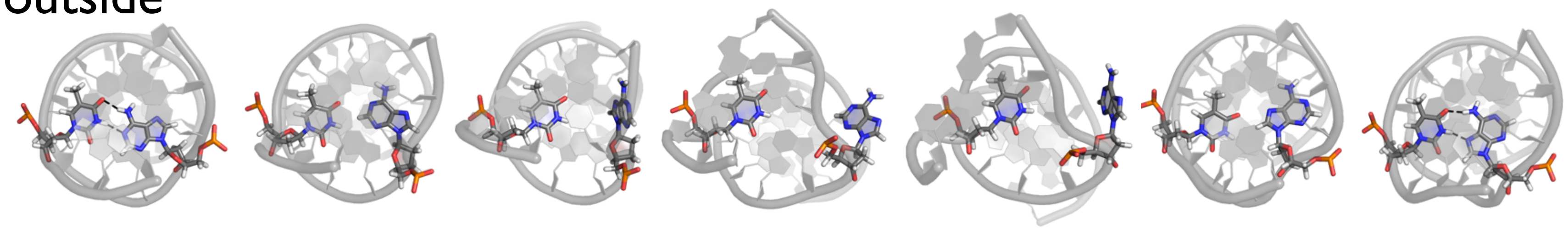
DNA Hoogsteen base pair formation

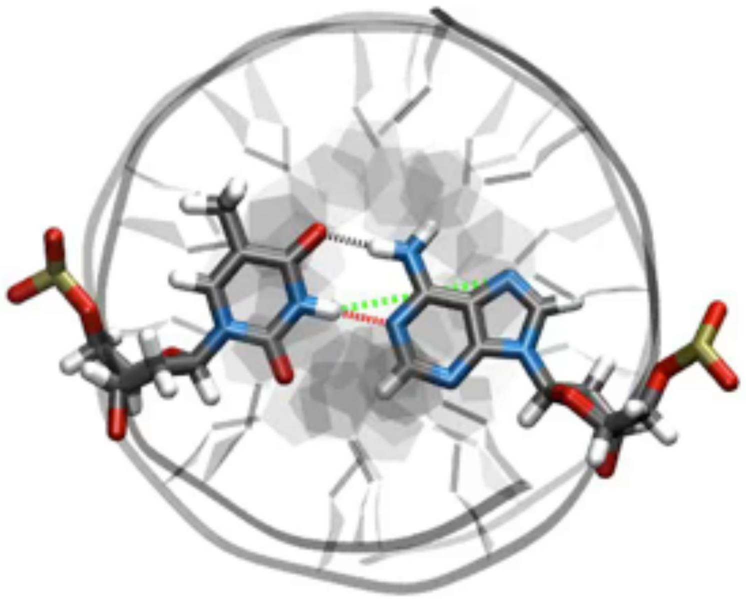
path by Transition path sampling

inside



outside





AMBER03 force field &
parmbsc0/1

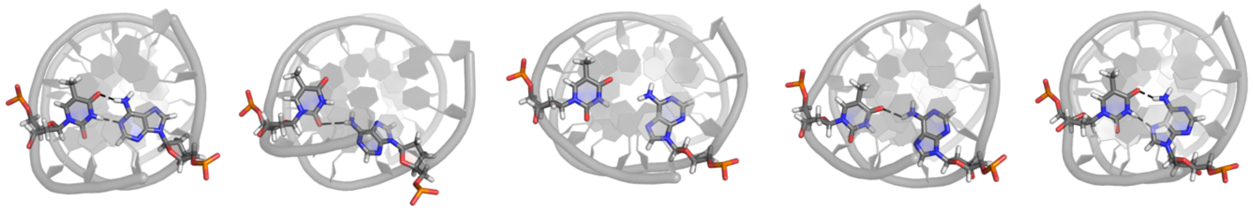
what is the rate constant of this process?

Vreede,Pérez de Alba Ortíz,PGB Swenson, NAR, 47,11069 (2019)

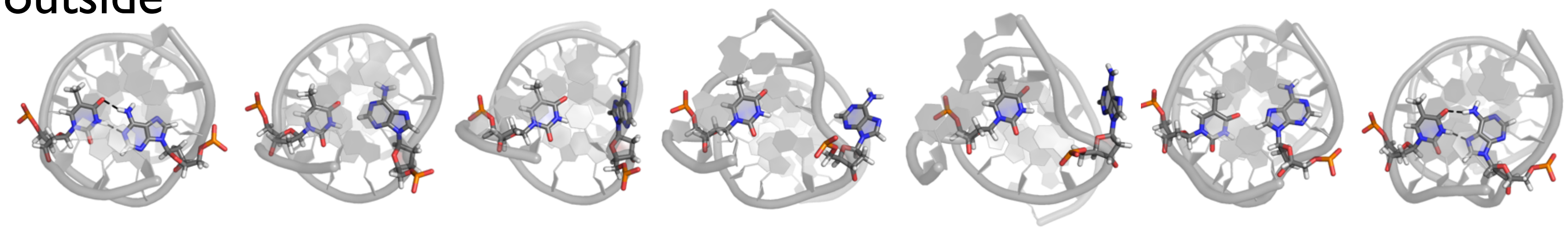
DNA Hoogsteen base pair formation

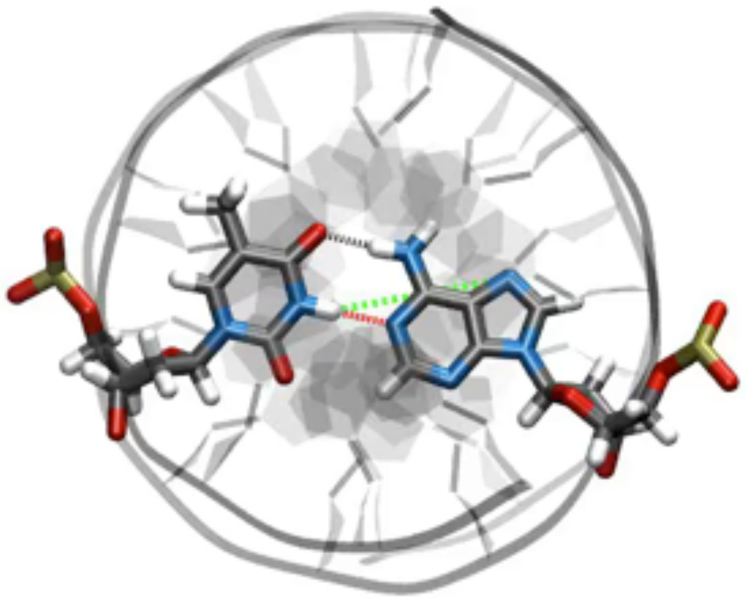
path by Transition path sampling

inside



outside

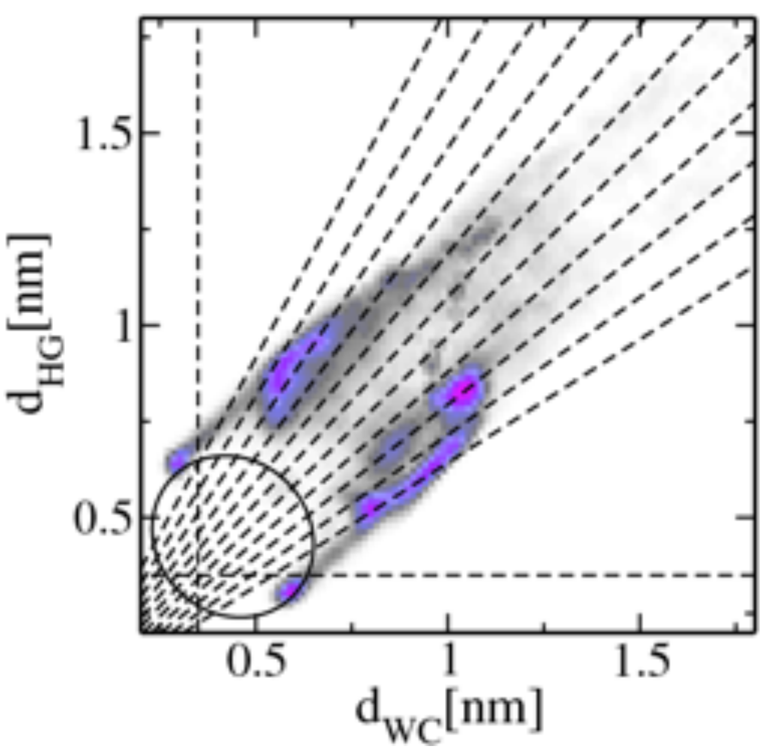


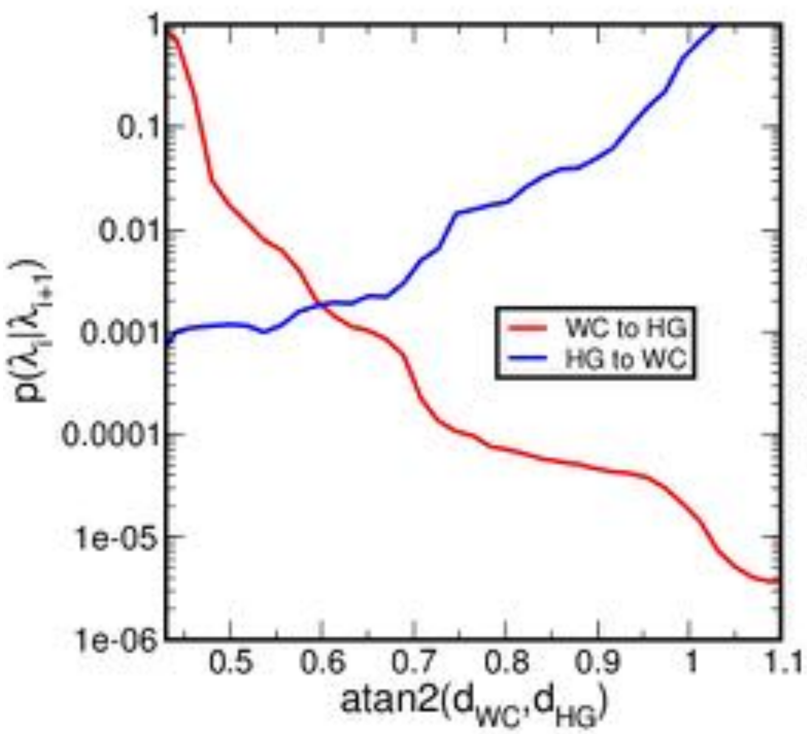


AMBER03 force field & parmbsc0/1

what is the rate constant of this process?

	experiment	TIS
$k_{WC \to HG}$ (s^{-1})	14.2 ± 1.03	742
$k_{HG \to WC}$ (s^{-1})	3670	$1.6 \cdot 10^5$
ΔG $(k_B T)$	5.5	5.4



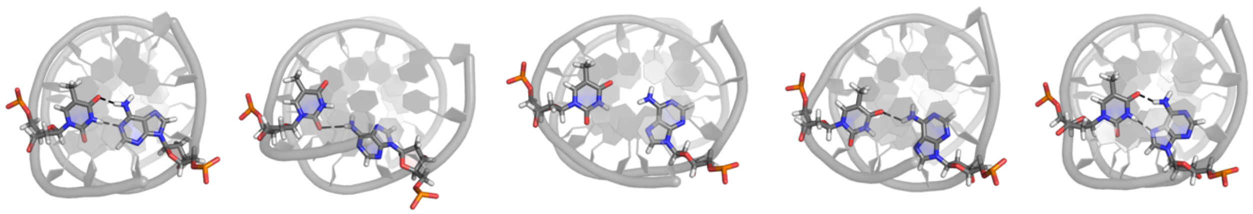


Vreede, Pérez de Alba Ortíz, PGB Swenson, NAR, 47, 11069 (2019)

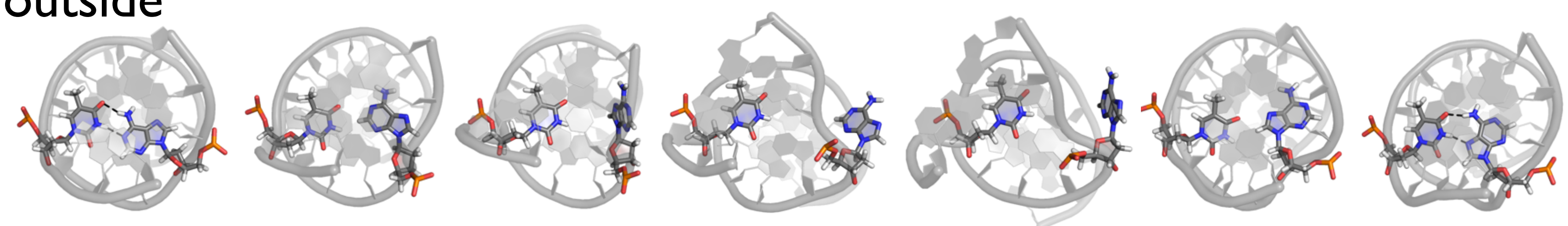
DNA Hoogsteen base pair formation

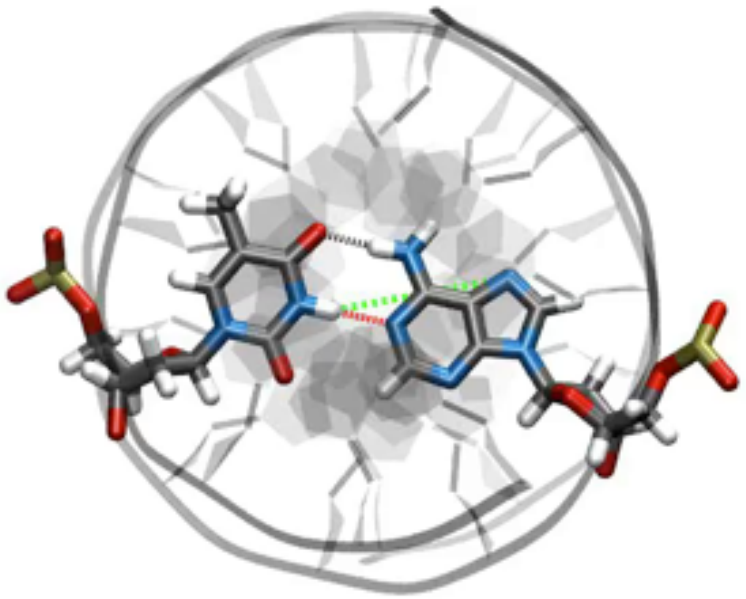
path by Transition path sampling

inside



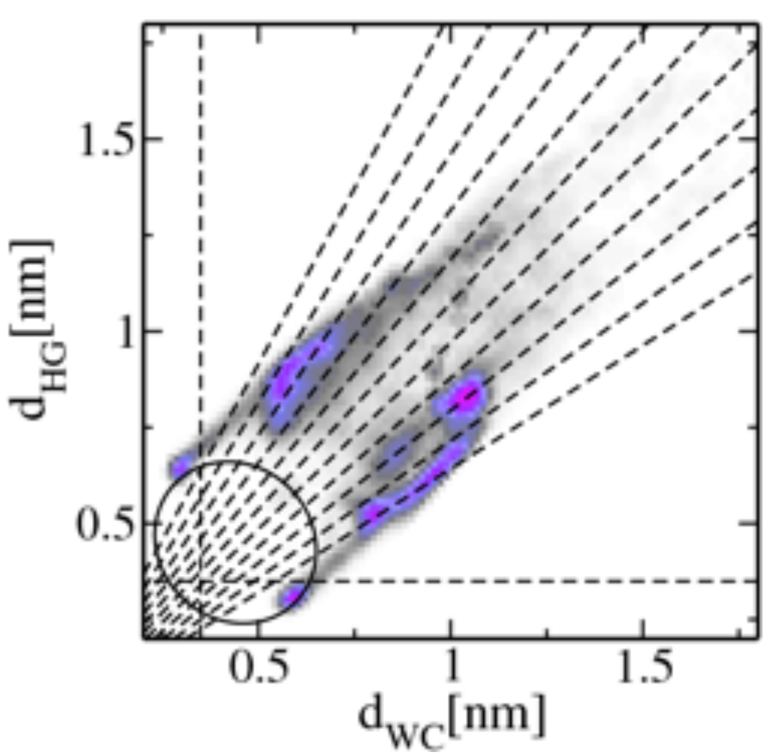
outside

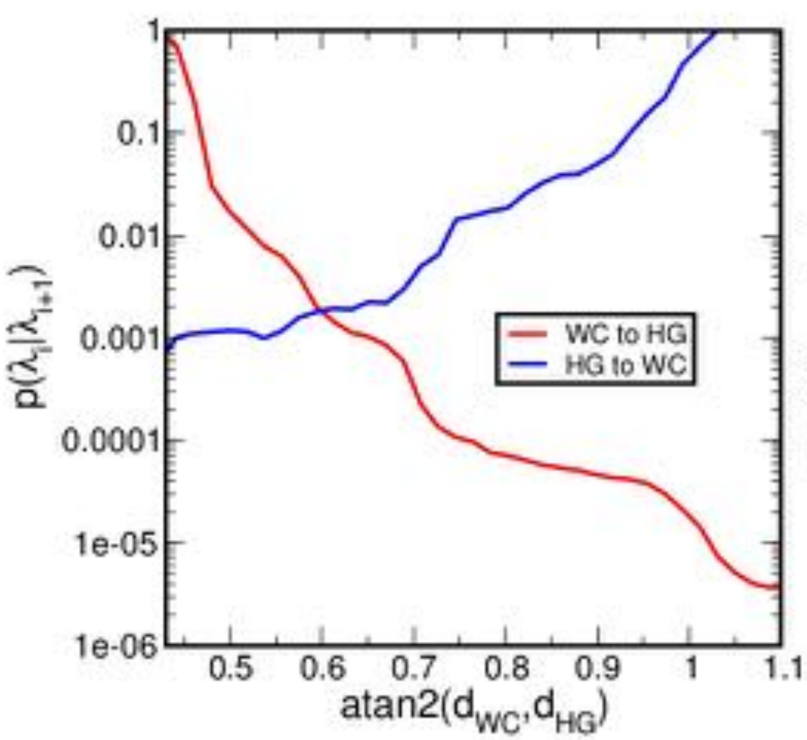




AMBER03 force field & parmbsc0/1

what is the rate constant of this process?



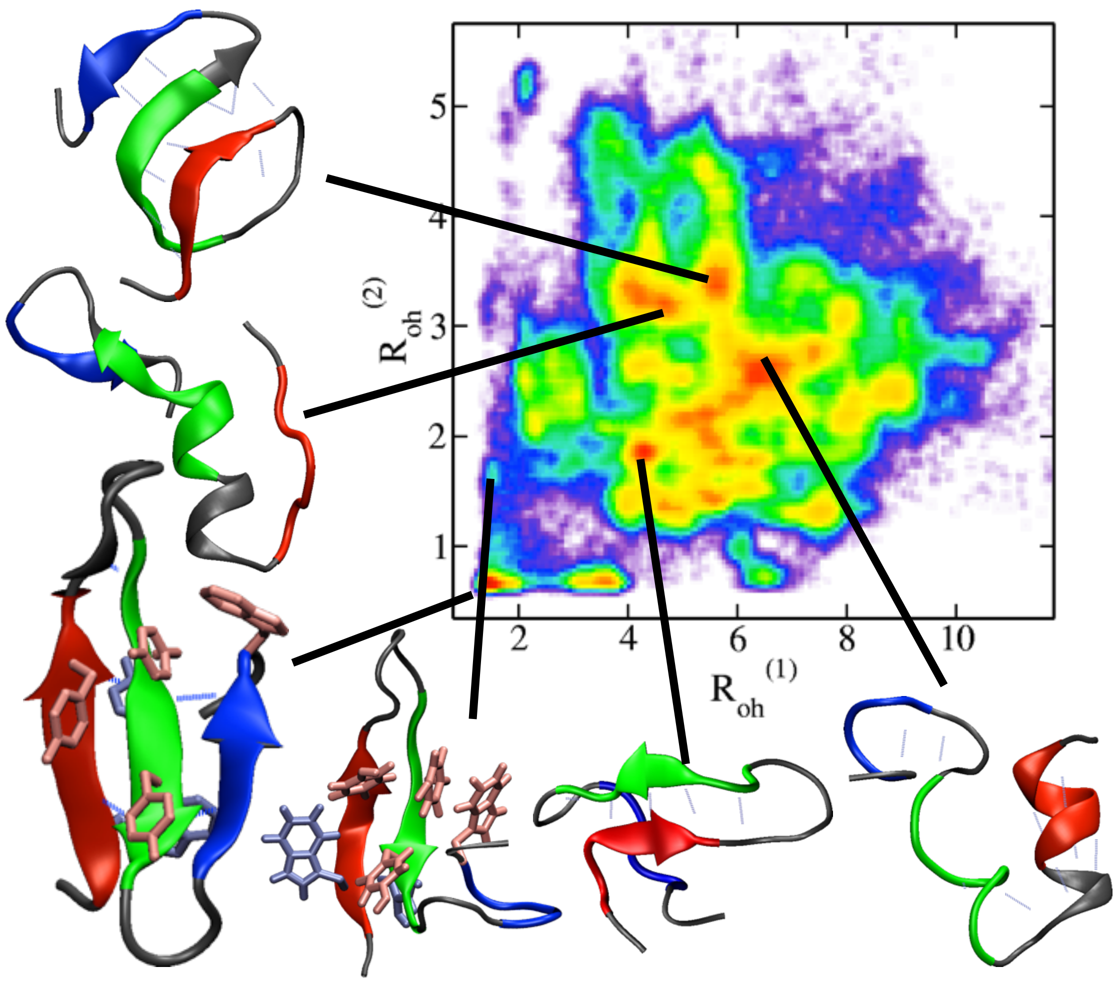


	experiment	TIS
$k_{WC\rightarrow HG}$ (s^{-1})	14.2 ± 1.03	742
$k_{HG\rightarrow WC}$ (s^{-1})	3670	$1.6 \cdot 10^5$
ΔG $(k_B T)$	5.5	5.4

mismatch with experiment:
force field or missing transition?

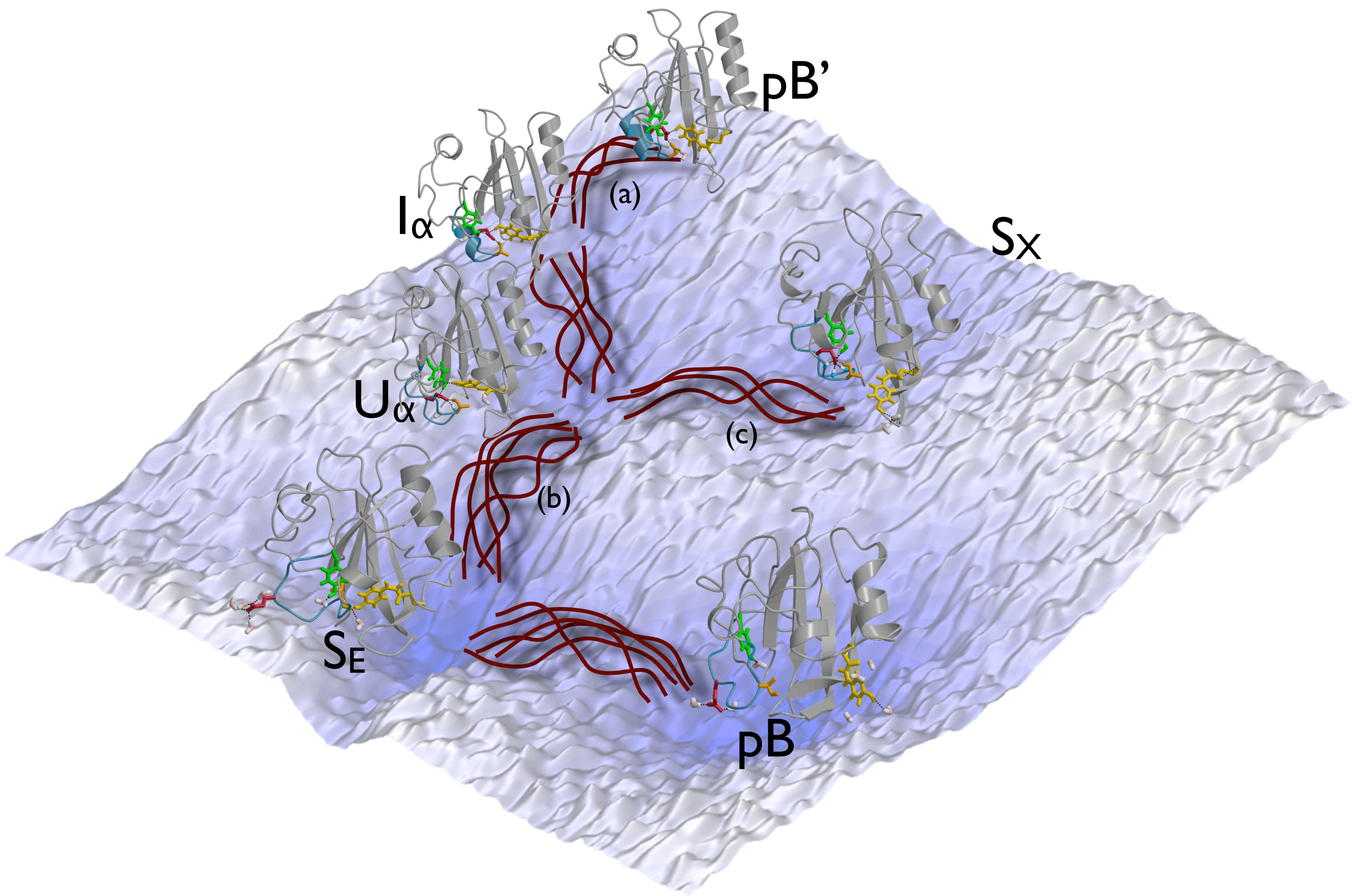
Vreede,Pérez de Alba Ortíz,PGB Swenson, NAR, 47,11069 (2019)

Sampling complex free energy landscapes



WW domain folding

J. Juraszek and PGB, Biophys. J.98, 646 (2010).

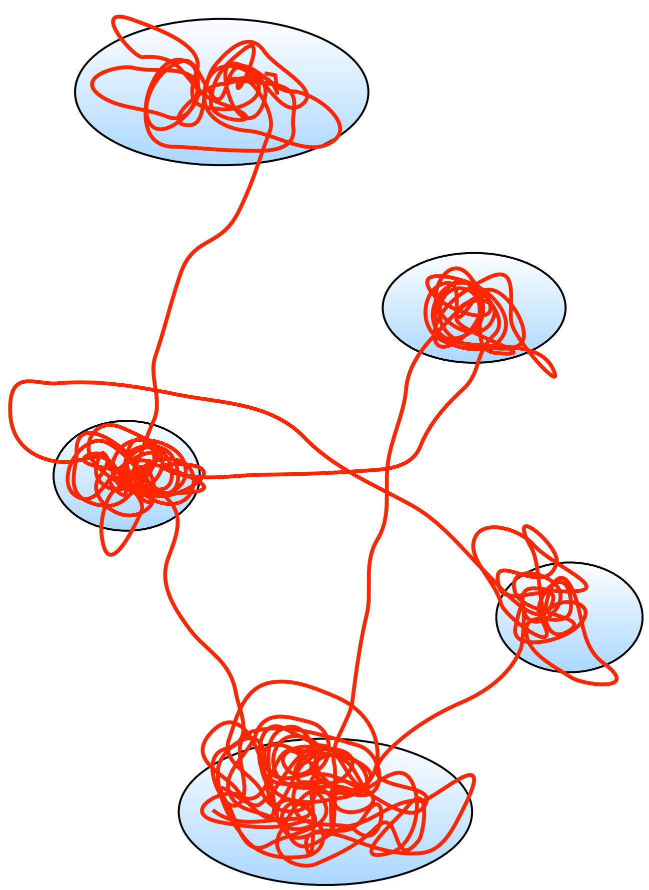


PYP signal transduction

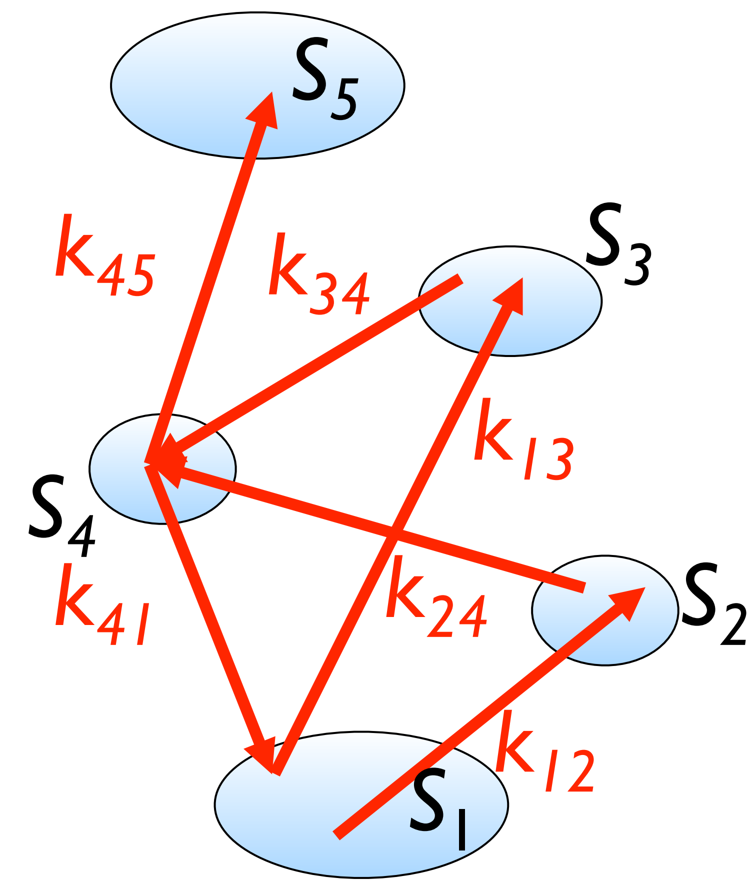
J.Vreede J, J. Juraszek and PGB PNAS 107 2397(2010)

Markov state model

molecular dynamics trajectory



coarse grained trajectory



integrate equations of motion

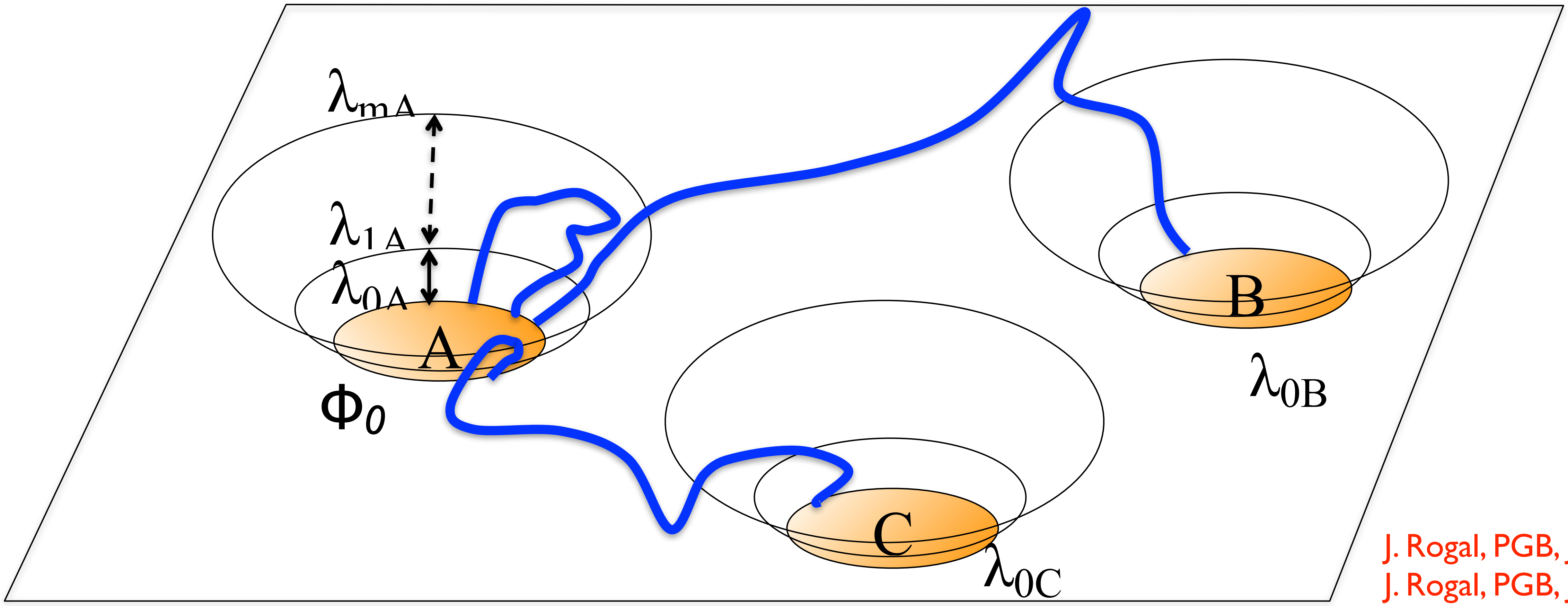
time step $\Delta t \approx$ fs

$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} k_{ji} p_j(t) - \sum_{j \neq i} k_{ij} p_i(t)$$

master equation,
solve analytically or by KMC
time step set by rates

See also work of Noe, Chodera, Pande, etc

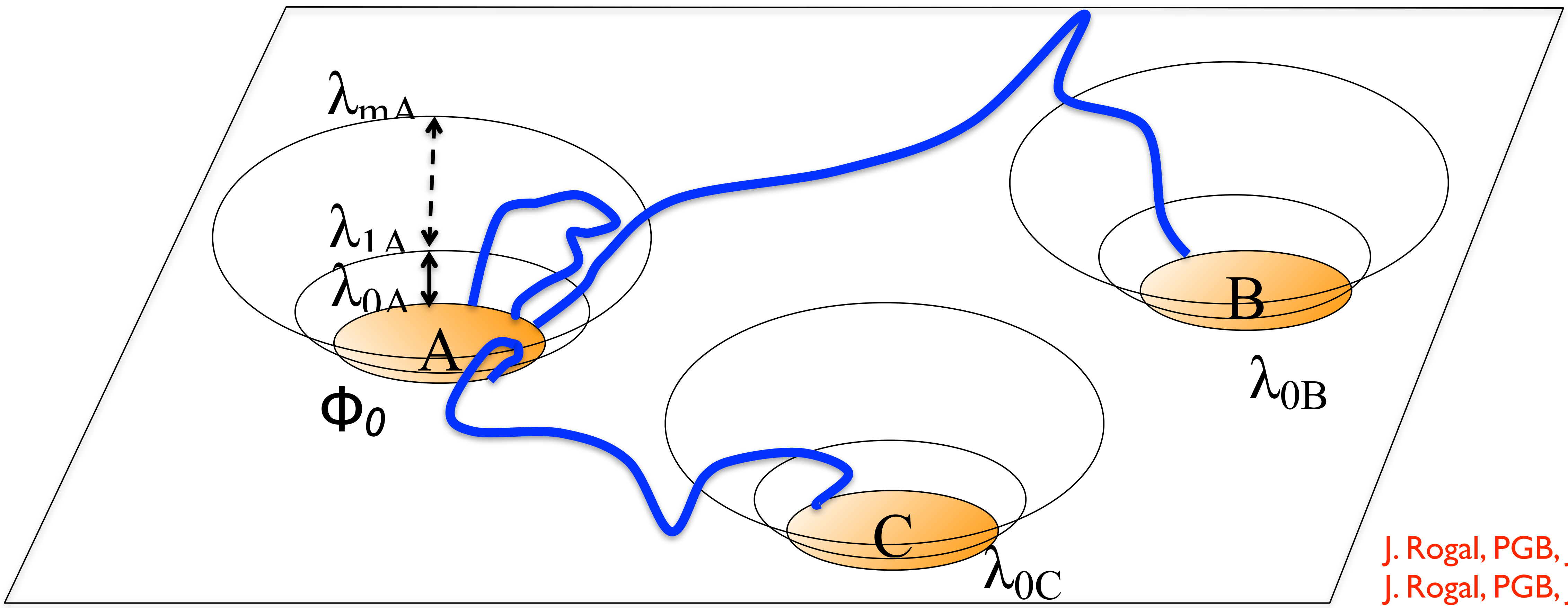
Multiple state transition interface sampling



J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

$P_A(\lambda_{(s+1)A}|\lambda_{(s+1)A})$ = probability path crossing s for first time after leaving A reaches s+1 before A

Multiple state transition interface sampling

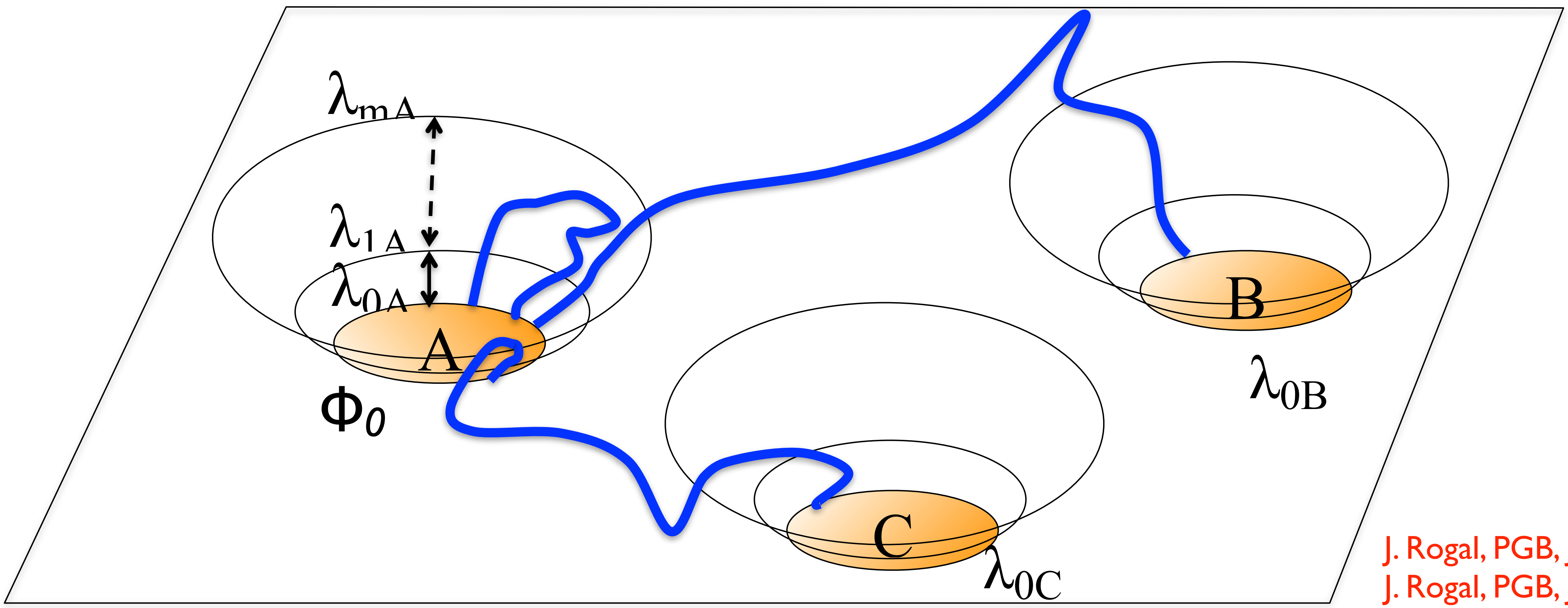


J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

$P_A(\lambda_{(s+1)A}|\lambda_{(s+1)A})$ = probability path crossing s for first time after leaving A reaches s+1 before A

$$k_{Ai} = \frac{\langle \phi_{\lambda_{m_A}} \rangle}{\langle h_A \rangle} \cdot P_A(\lambda_{0_i}|\lambda_{m_A})$$

Multiple state transition interface sampling



J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

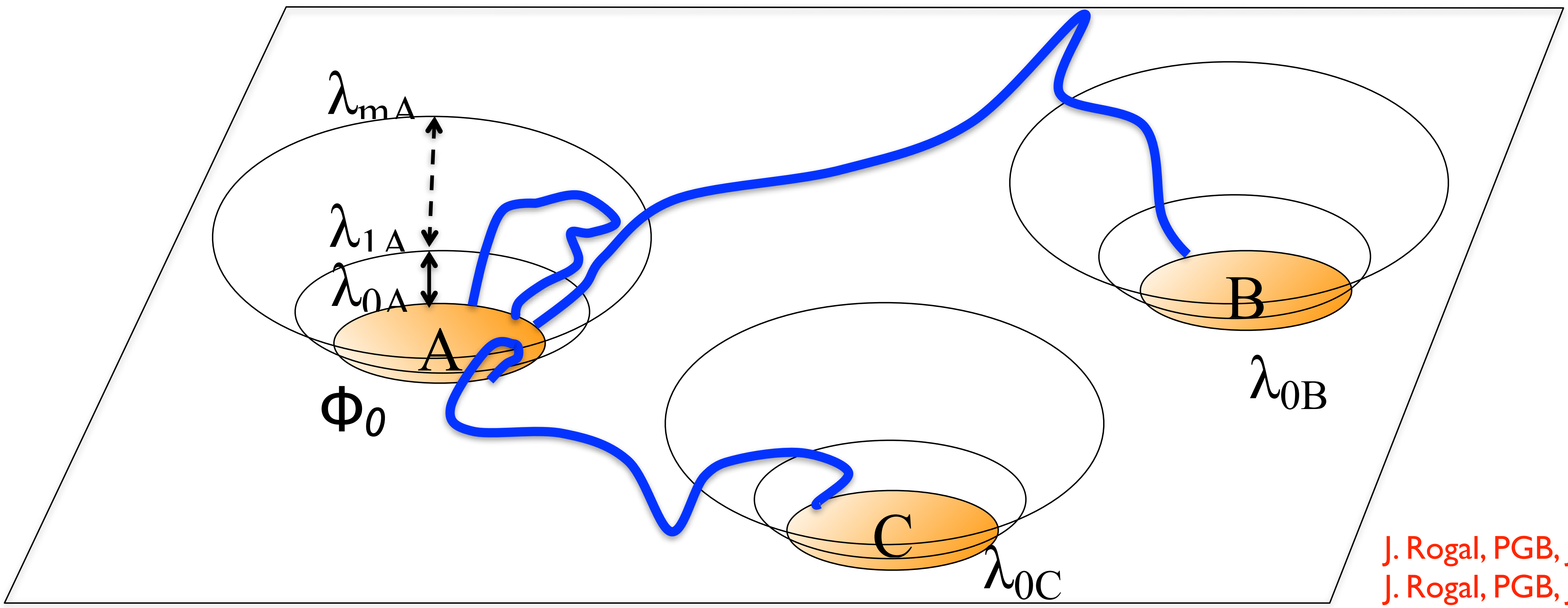
$P_A(\lambda_{(s+1)A}|\lambda_{(s+1)A})$ = probability path crossing s for first time after leaving A reaches s+1 before A

$$k_{Ai} = \underbrace{\frac{\langle \phi_{\lambda_{m_A}} \rangle}{\langle h_A \rangle}} \cdot P_A(\lambda_{0_i}|\lambda_{m_A})$$

TIS:

$$\frac{\langle \phi_A \rangle}{\langle h_A \rangle} \prod_{s=0}^{m-1} P_A(\lambda_{(s+1)_A}|\lambda_{s_A})$$

Multiple state transition interface sampling



J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

$P_A(\lambda_{(s+1)A}|\lambda_{(s+1)A})$ = probability path crossing s for first time after leaving A reaches s+1 before A

$$k_{Ai} = \underbrace{\frac{\langle \phi_{\lambda_{m_A}} \rangle}{\langle h_A \rangle}} \cdot \underbrace{P_A(\lambda_{0_i}|\lambda_{m_A})}$$

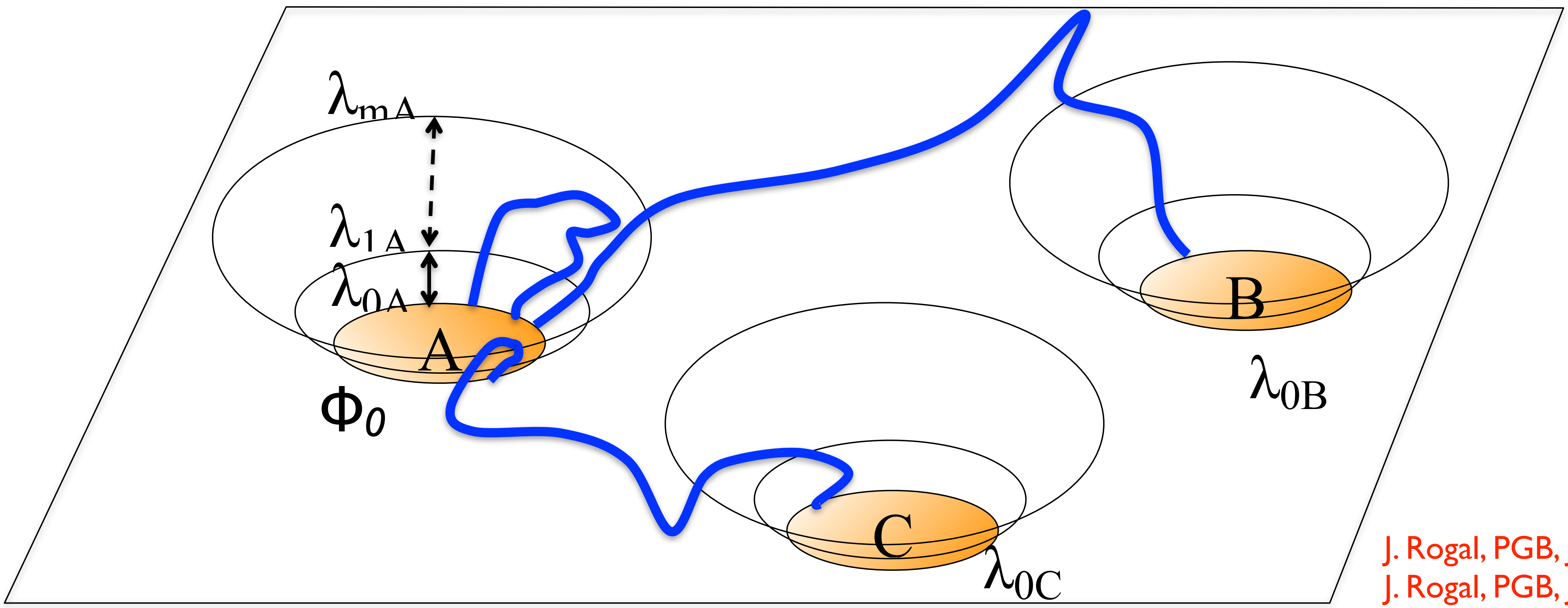
TIS:

$$\frac{\langle \phi_A \rangle}{\langle h_A \rangle} \prod_{s=0}^{m-1} P_A(\lambda_{(s+1)_A}|\lambda_{s_A})$$

MSTIS:

$$\frac{\text{no. of pathways coming from A, cross } \lambda_{mA}\text{, end i}}{\text{no. of pathways coming from A, cross } \lambda_{mA}}$$

Multiple state transition interface sampling



J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

$P_A(\lambda_{(s+1)A}|\lambda_{(s+1)A})$ = probability path crossing s for first time after leaving A reaches s+1 before A

$$k_{Ai} = \underbrace{\frac{\langle \phi_{\lambda_{m_A}} \rangle}{\langle h_A \rangle}}\cdot \underbrace{P_A(\lambda_{0_i}|\lambda_{m_A})}$$

TIS:

$$\frac{\langle \phi_A \rangle}{\langle h_A \rangle} \prod_{s=0}^{m-1} P_A(\lambda_{(s+1)_A}|\lambda_{s_A})$$

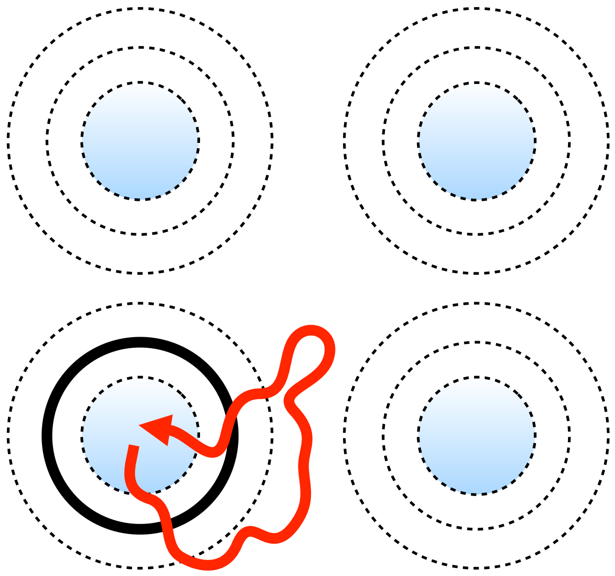
MSTIS:

$$\frac{\text{no. of pathways coming from A, cross } \lambda_{mA}\text{, end i}}{\text{no. of pathways coming from A, cross } \lambda_{mA}}$$

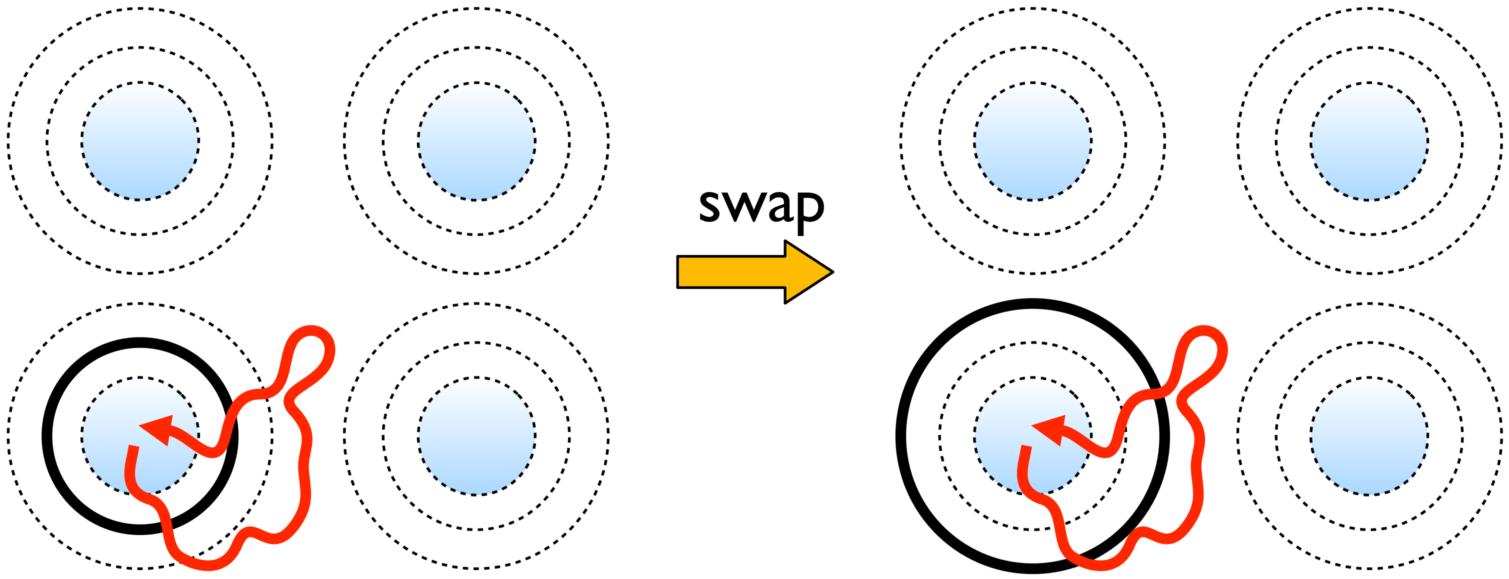
rates can be used in Markov state model $\frac{dp_i(t)}{dt} = \sum_{j\neq i} k_{ji}p_j(t) - \sum_{j\neq i} k_{ij}p_i(t)$

Single replica MSTIS

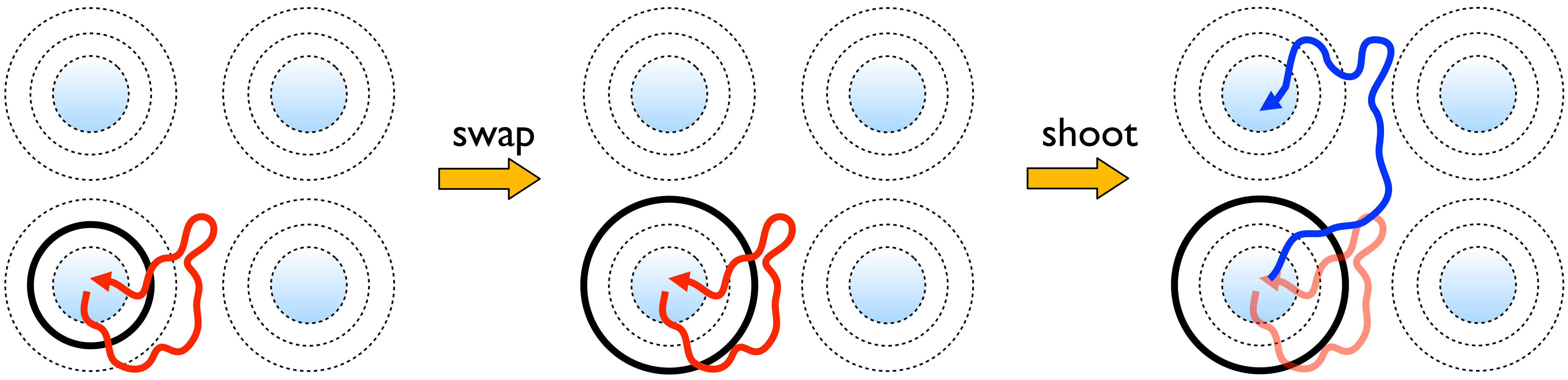
Single replica MSTIS



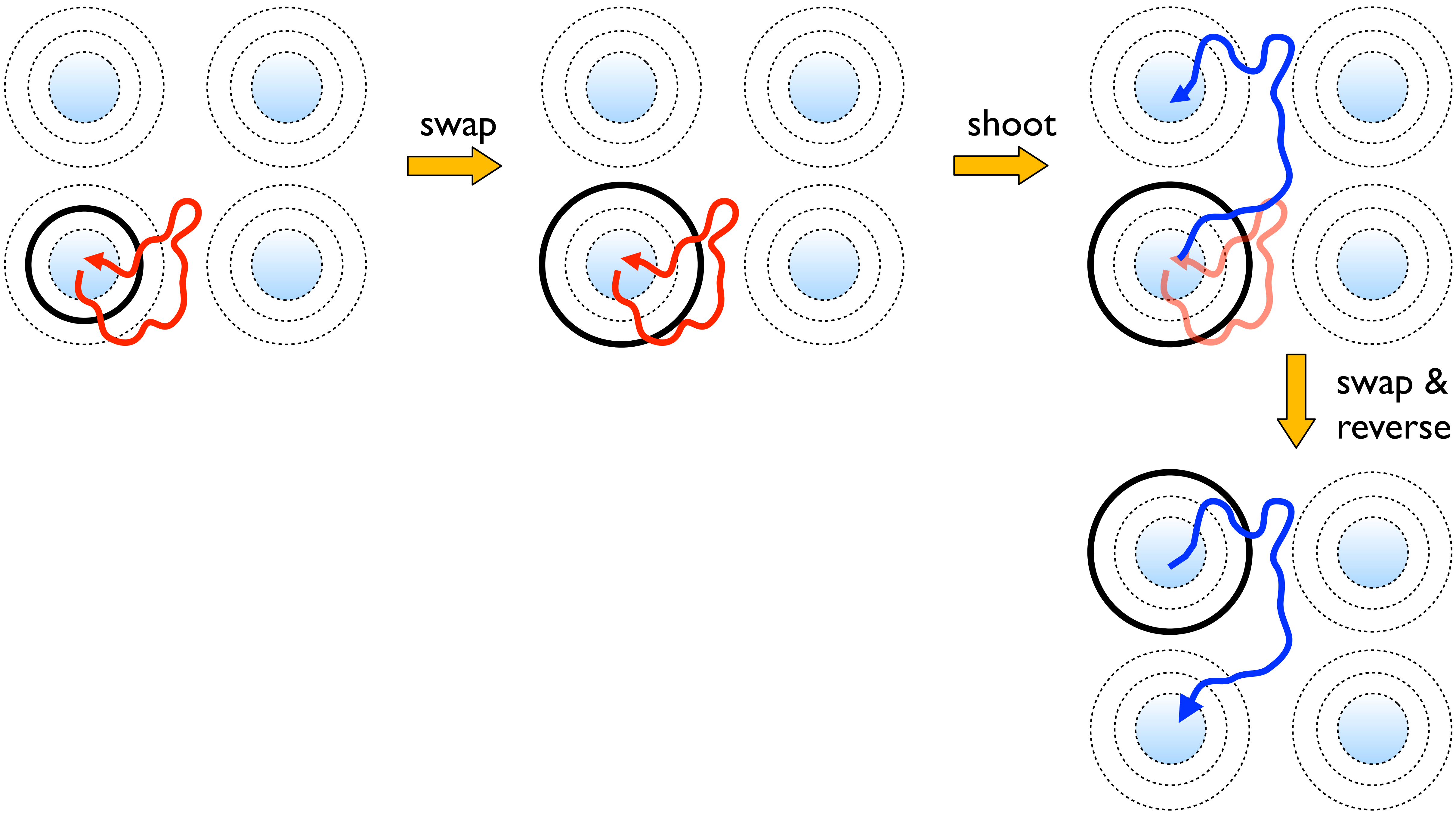
Single replica MSTIS



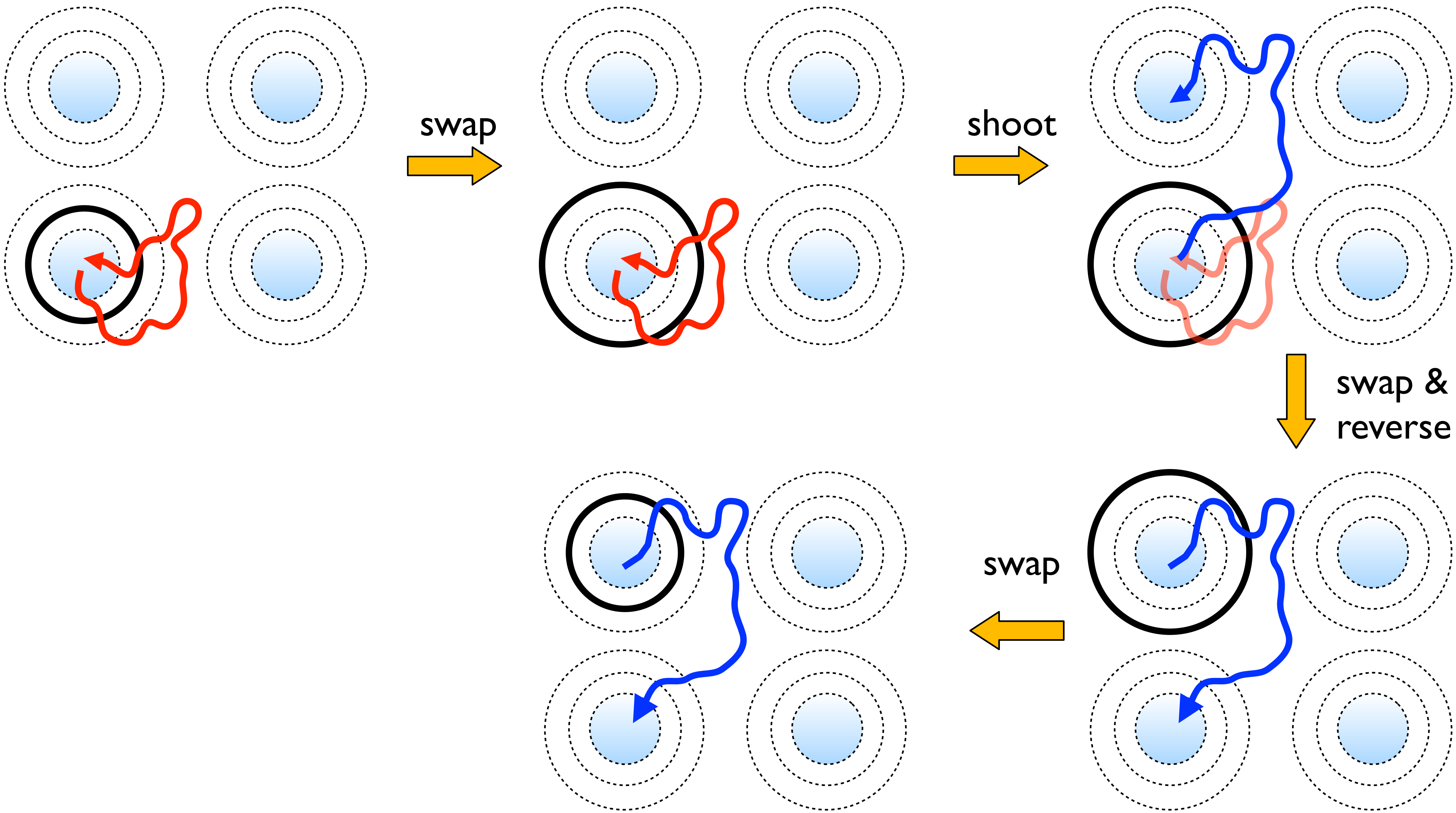
Single replica MSTIS



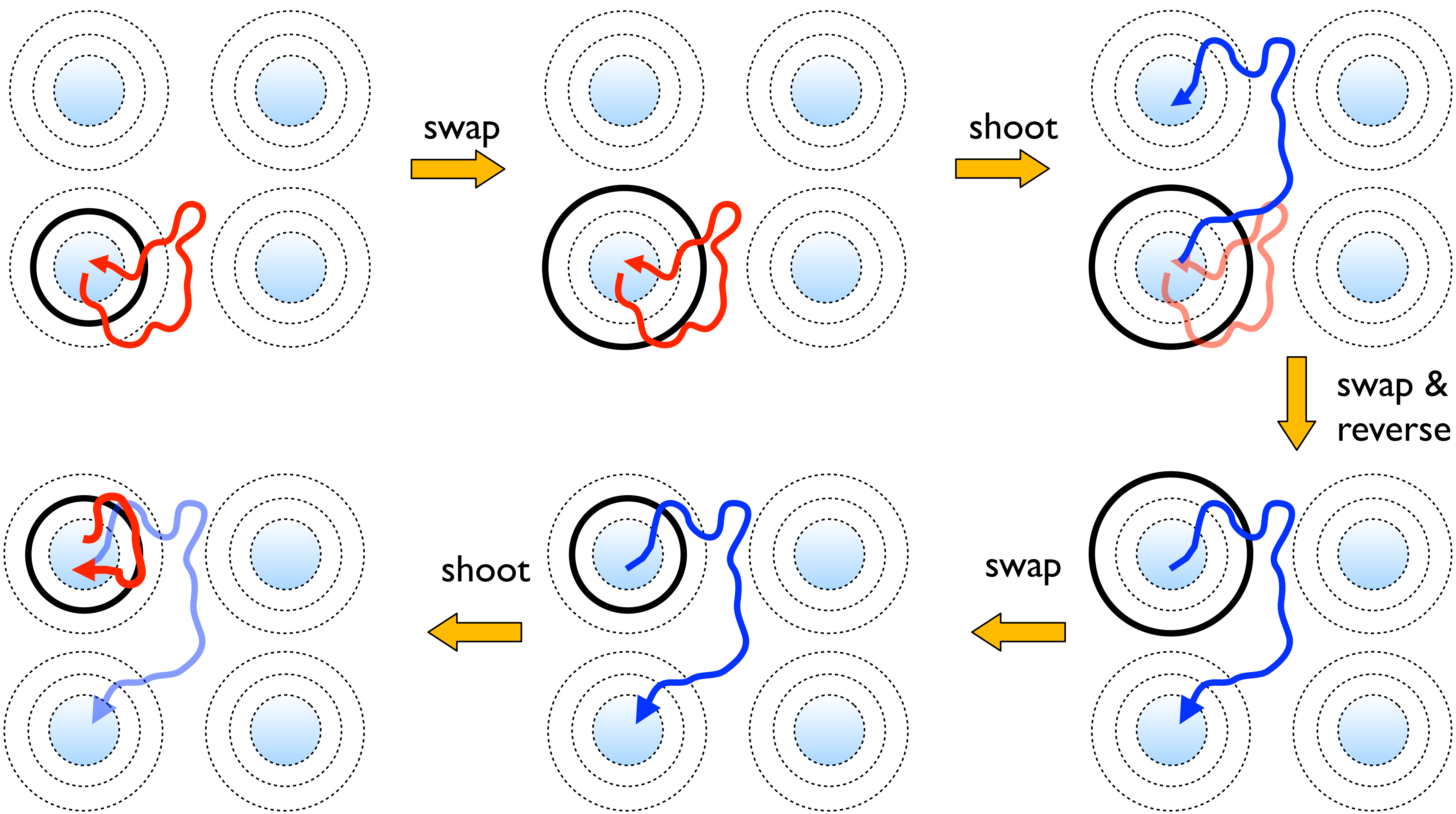
Single replica MSTIS



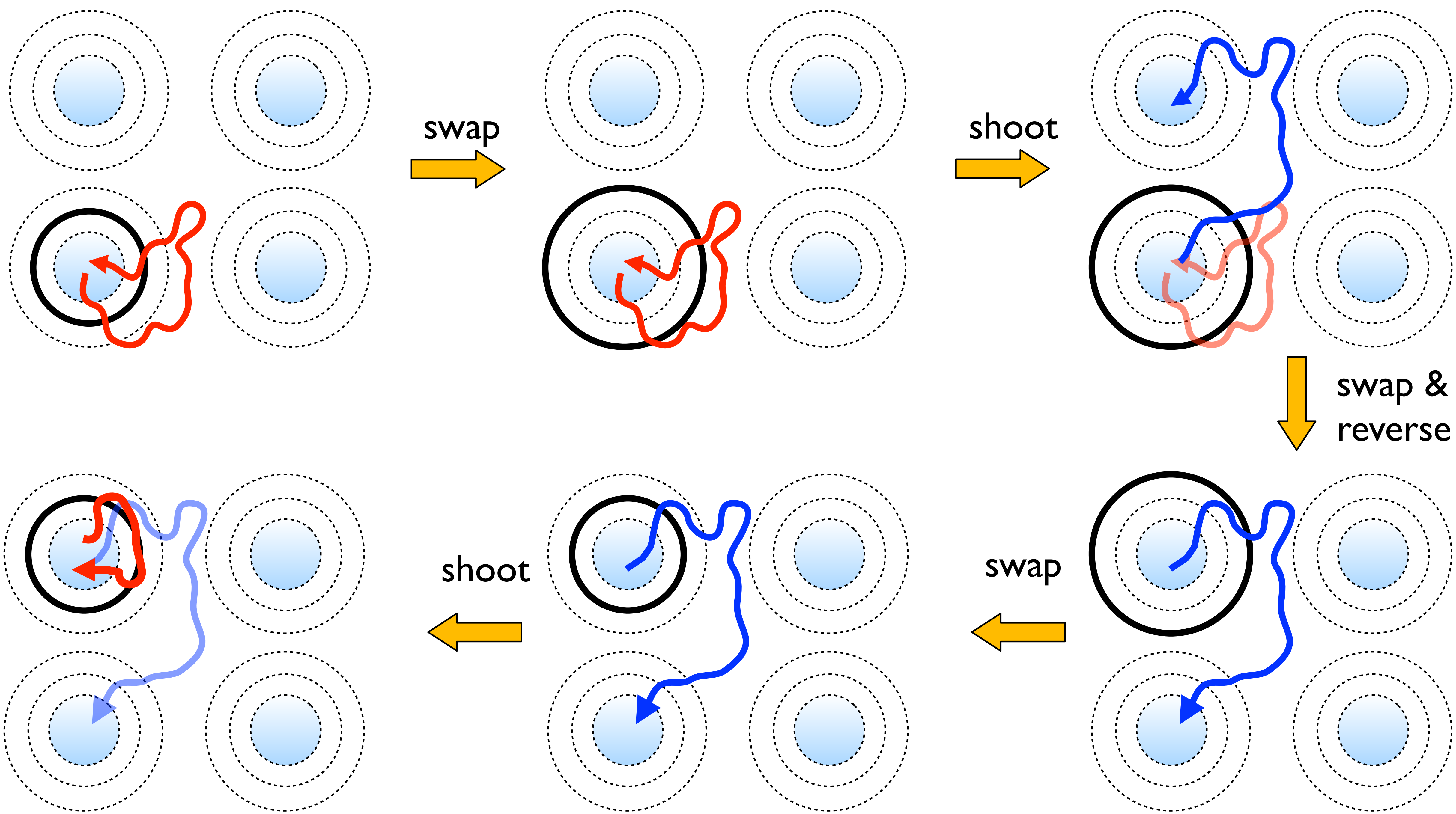
Single replica MSTIS



Single replica MSTIS



Single replica MSTIS



Problem: interfaces close to stable states will be favored
Solution: bias with e.g. Wang Landau scheme

Du, PGB, JCP 2013

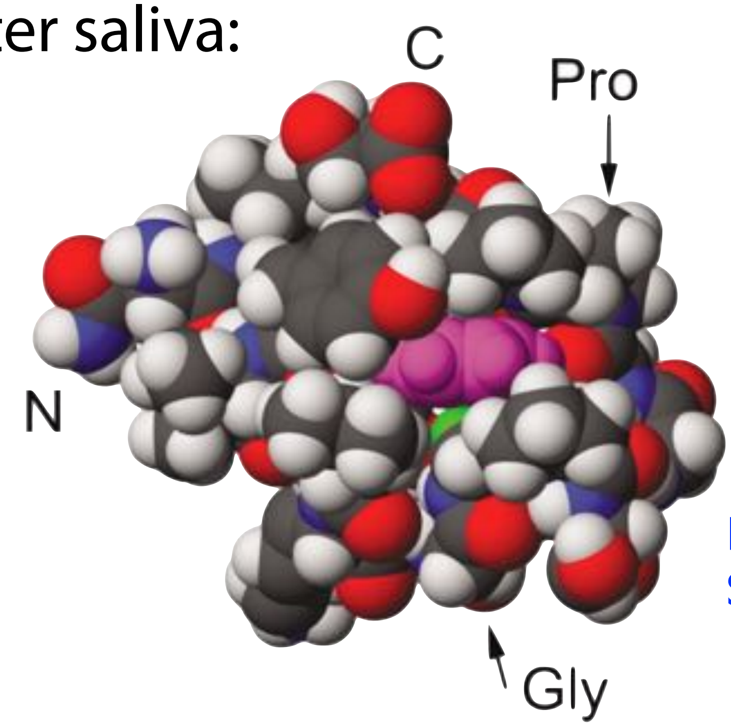
Trp-cage folding

- 20-residue fragment obtained from Gila monster saliva: **α-helix**, **3_{10}-helix**, **polyproline helix**

Salt-bridge

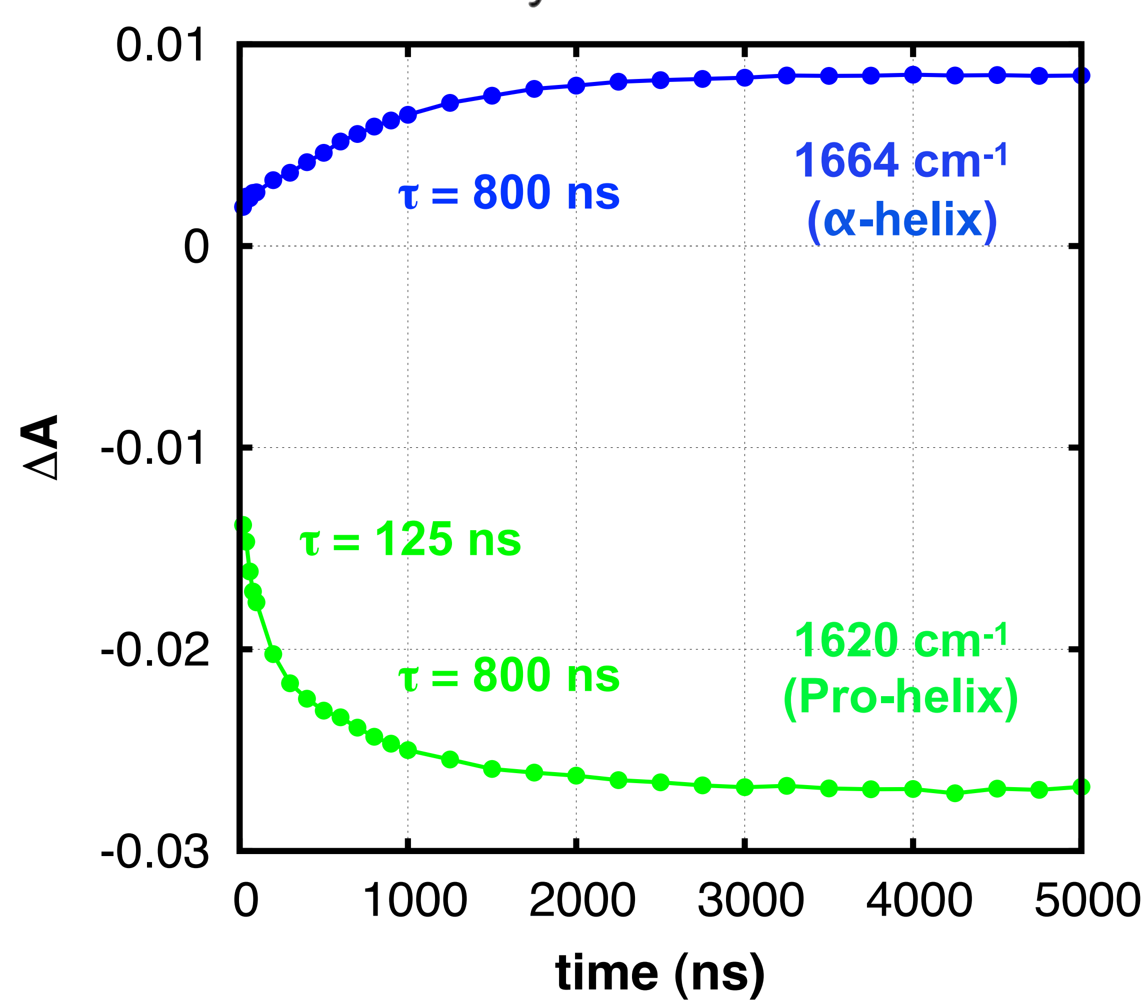
NAYAQ WLKDG GPSSG RPPPS

Tyrosine Tryptophan Glycine Proline



Neidigh & al., Nature Struct.Biol. **9**, 425 (2002)

- Folds on the microsecond timescale
- 2-state folder, experimental rate 4 μs
- T-jump vibrational spectroscopy (IR) shows bi-exponential relaxation kinetics
  ⇒ (un)folding involves an intermediate state
- timescales for different temperature, T=300 K τ_1=150 ns, t_2= 2.2 μs



H. Meuzelaar, K. A. Marino, A. Huerta-Viga, M. R. Panman, L.E. J. Smeenk, A.J. Kettelarij, J.H. van Maarseveen, P. Timmerman, PGB, and S. Woutersen
JPCB 2013.

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

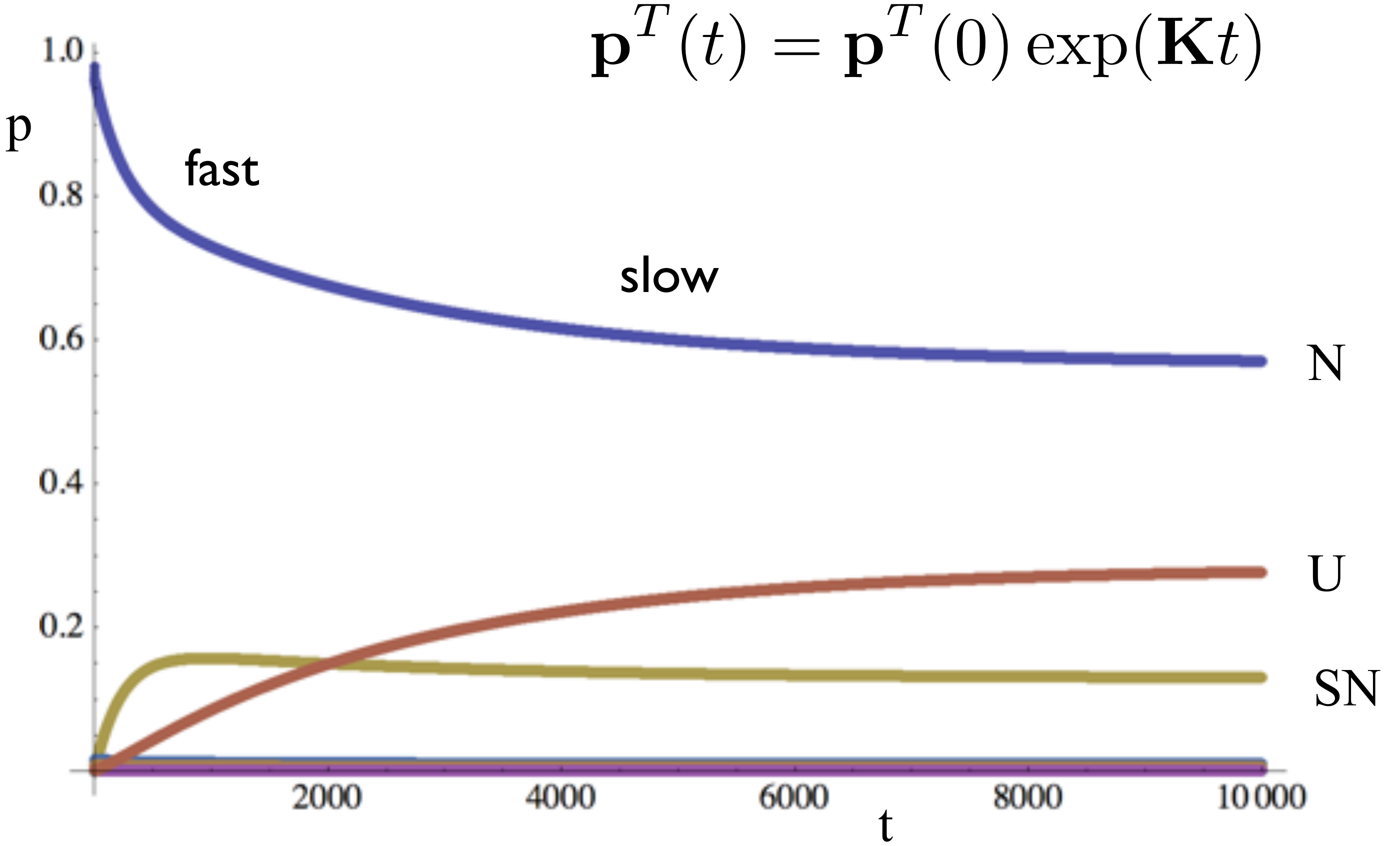
Du & PGB, JCP ***140****, 195102 (2014)).*

to end

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

$$\mathbf{p}^T(t) = \mathbf{p}^T(0)\exp(\mathbf{K}t)$$



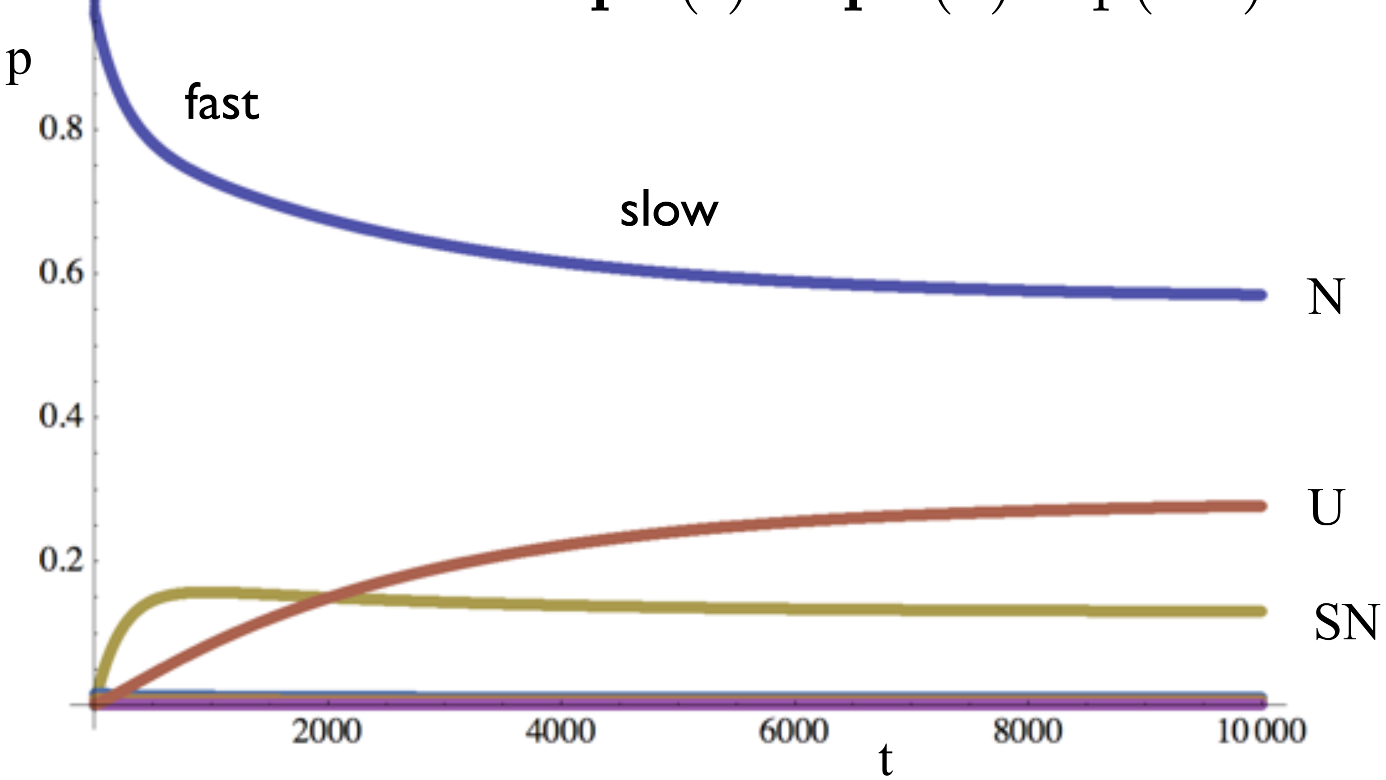
Du & PGB, JCP ***140***, *195102 (2014)).*

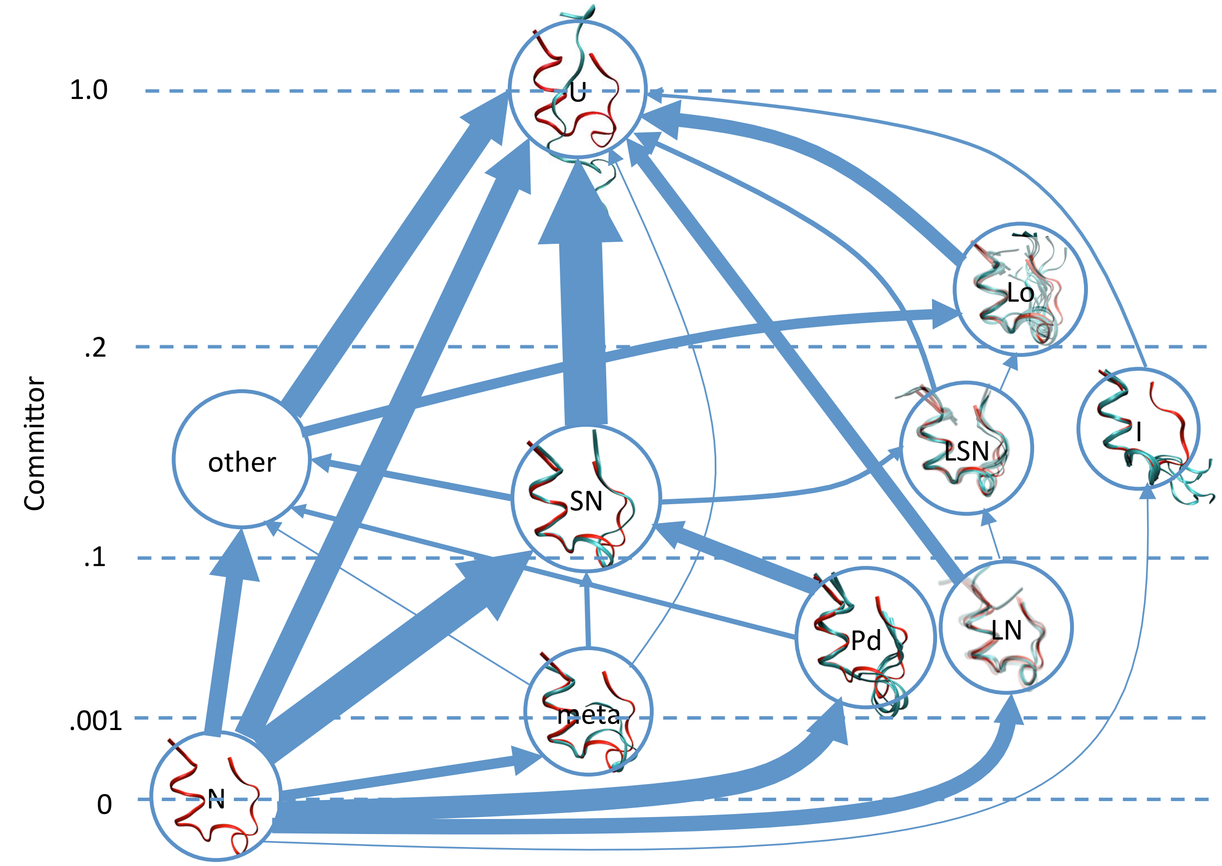
to end

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

$$\mathbf{p}^T(t) = \mathbf{p}^T(0)\exp(\mathbf{K}t)$$





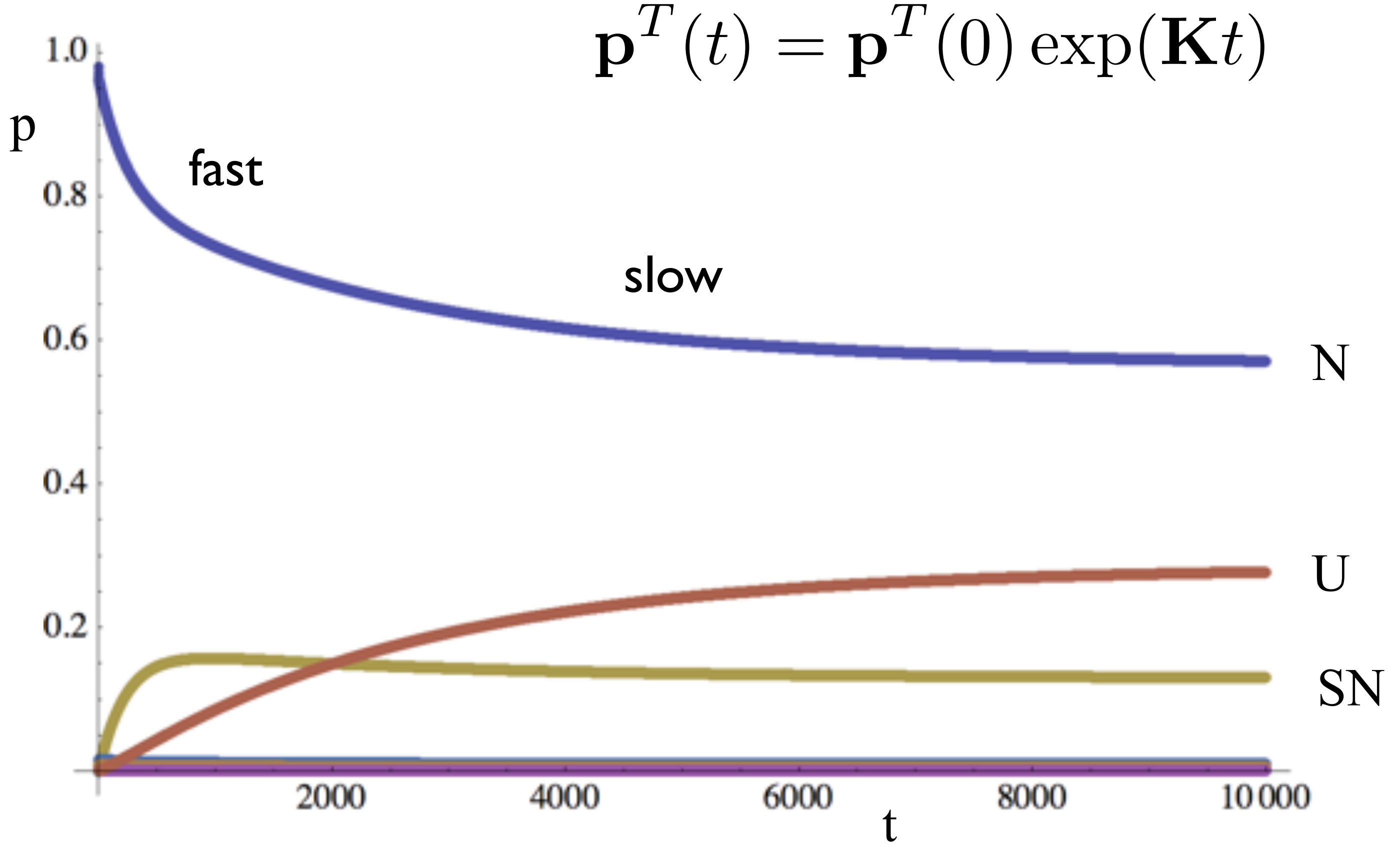
*Du & PGB, JCP **140**, 195102 (2014)).*

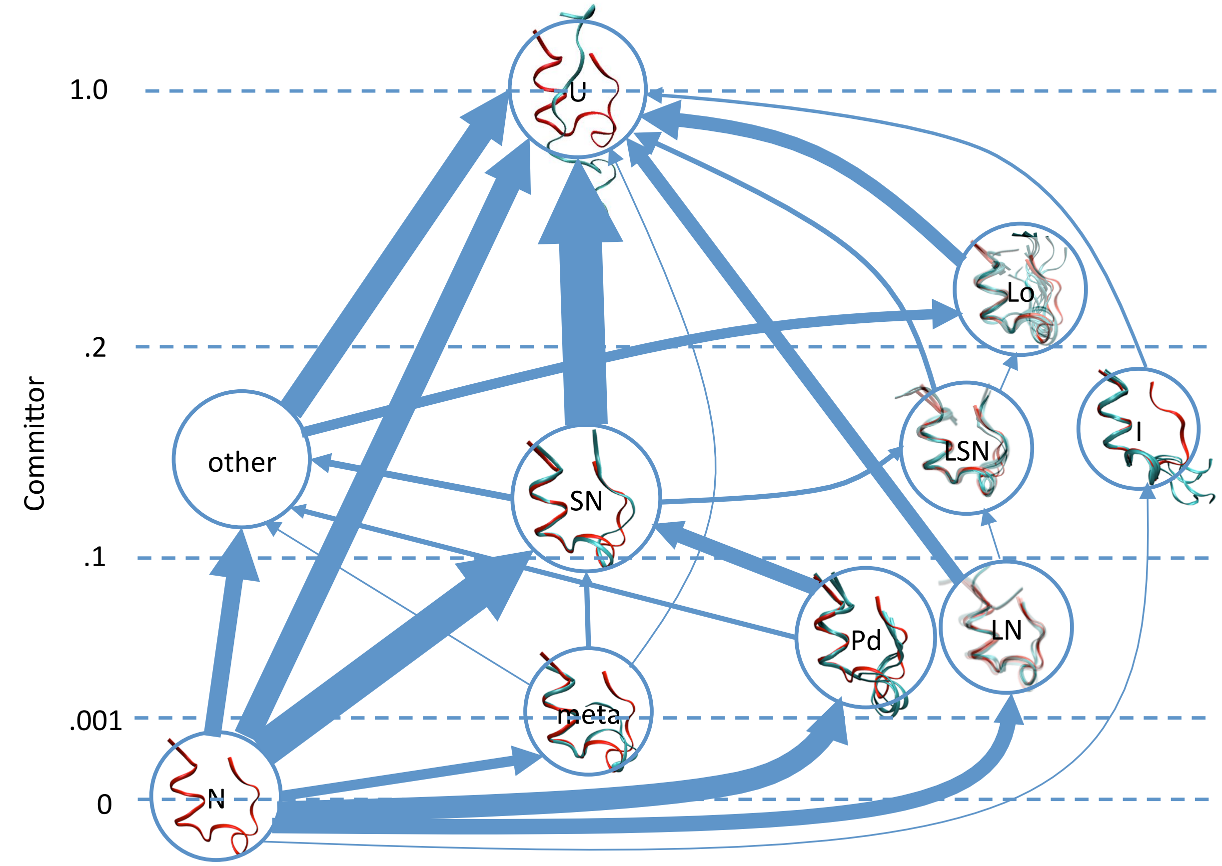
to end

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

$$\mathbf{p}^T(t) = \mathbf{p}^T(0)\exp(\mathbf{K}t)$$





fast time scale 200 ns

slow time scale 2 μs

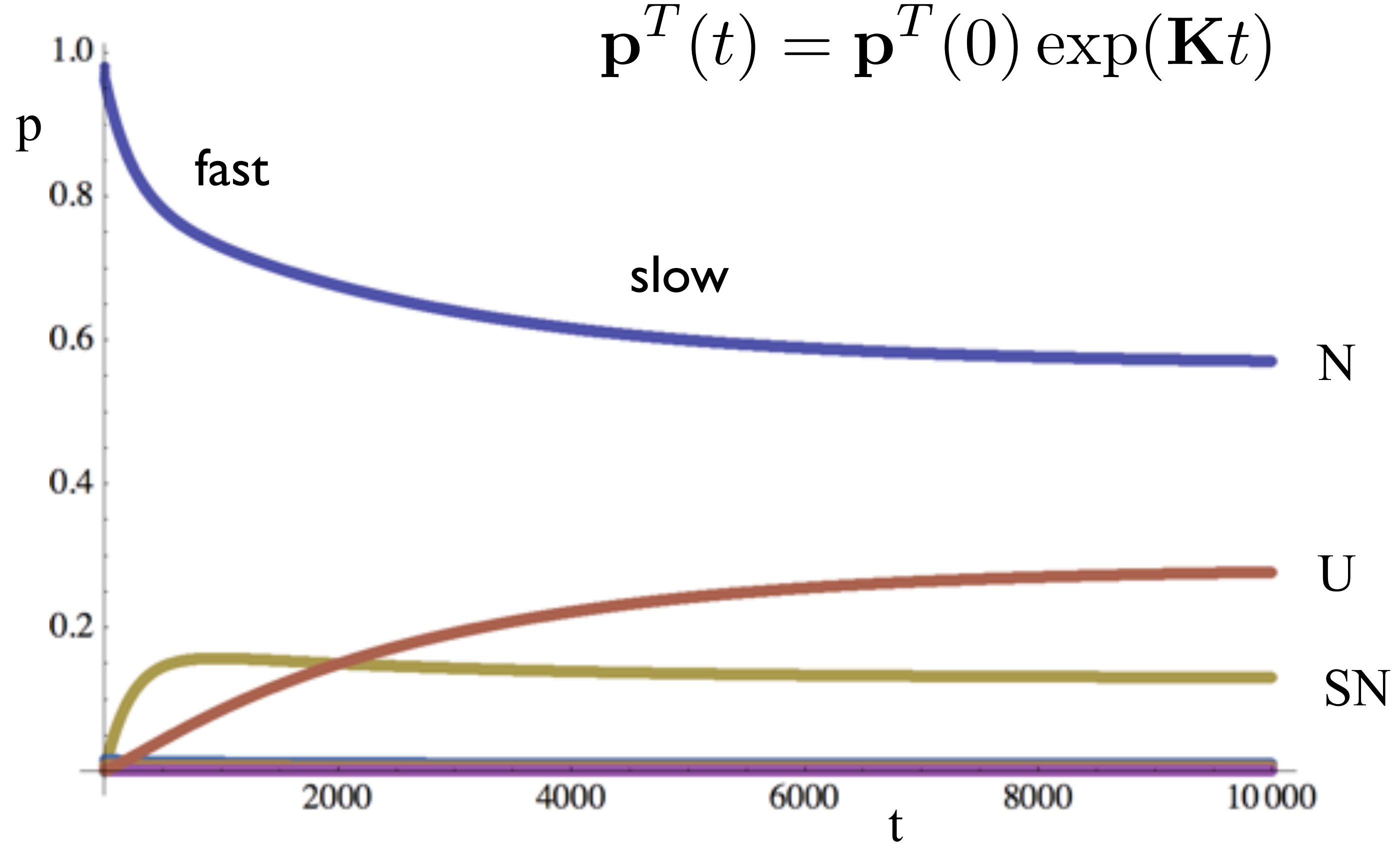
Du & PGB, JCP ***140****, 195102 (2014)).*

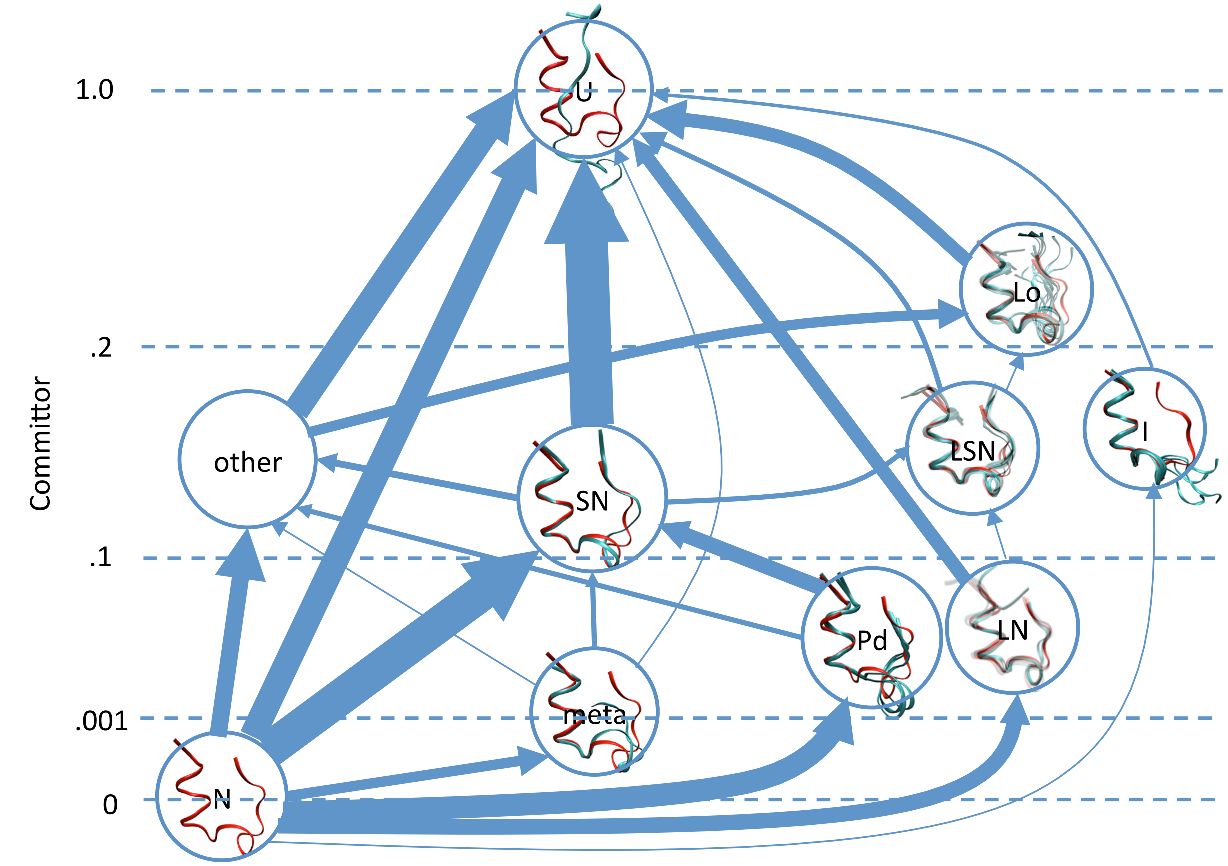
to end

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

$$\mathbf{p}^T(t) = \mathbf{p}^T(0)\exp(\mathbf{K}t)$$





fast time scale 200 ns

slow time scale 2 μs

Experimental t_1=150 ns, t_2= 2.2 μs

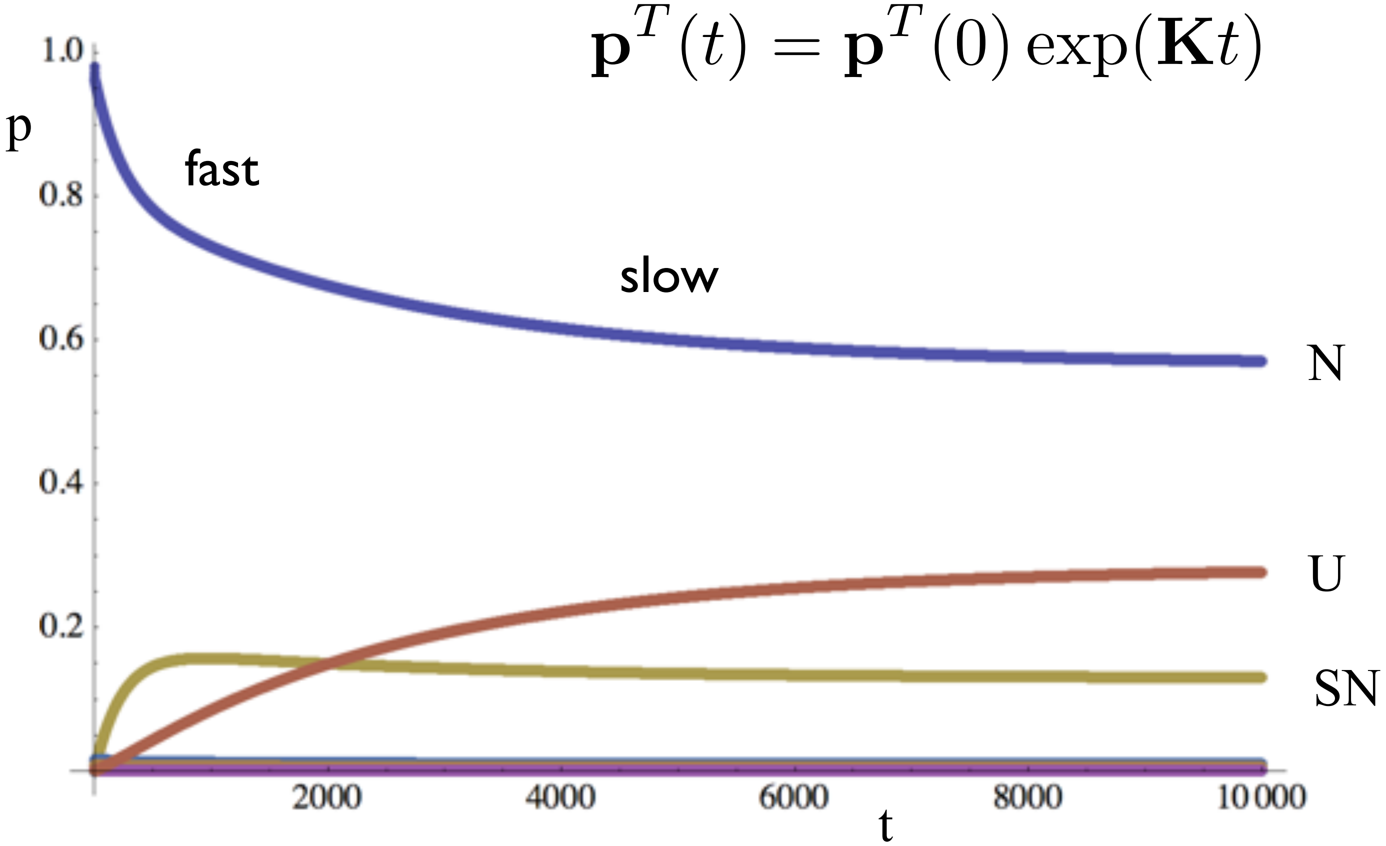
Du & PGB, JCP ***140****, 195102 (2014)).*

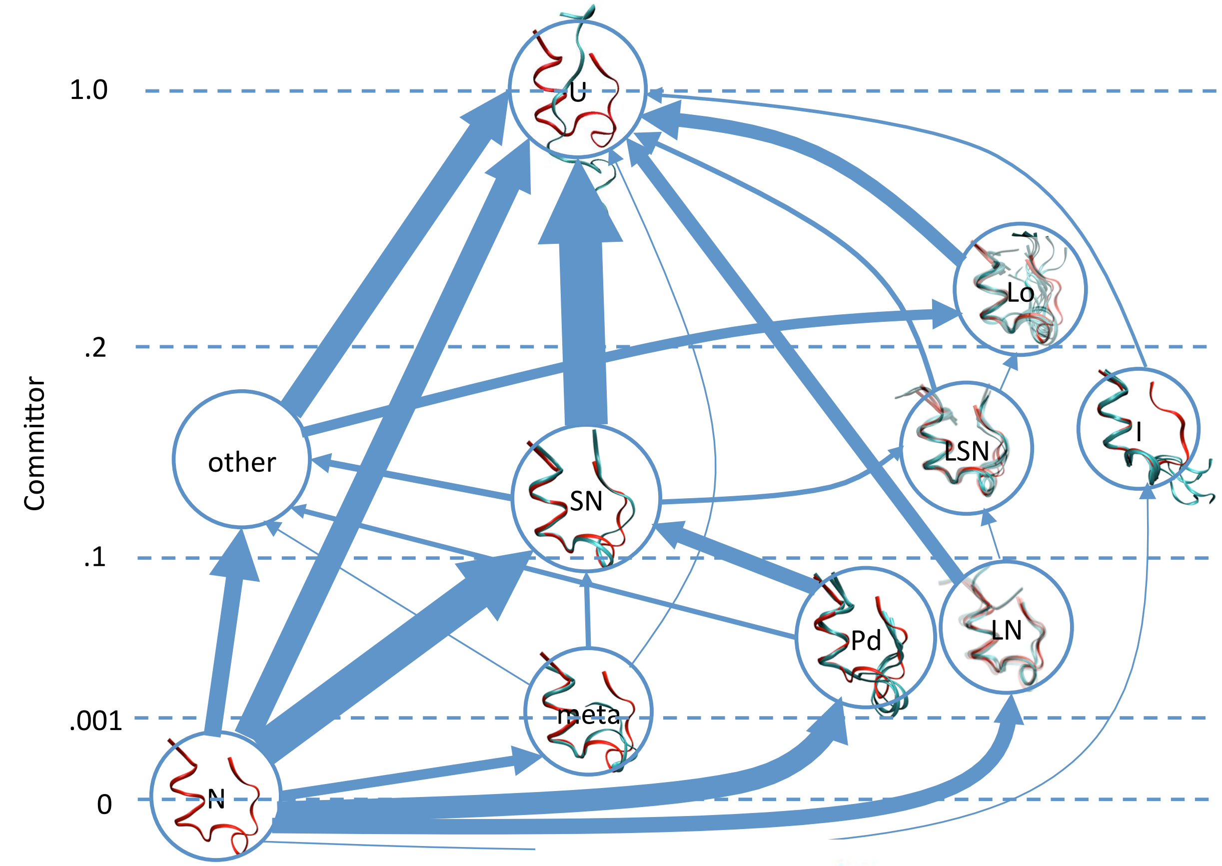
to end

Kinetics from MSTIS rate matrix

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	other state	U
N	—	3.75×10^{-3}	2.33×10^{-4}	4.67×10^{-4}	1.65×10^{-2}	5.35×10^{-3}	2.43×10^{-3}		1.04×10^{-4}		1.00×10^{-5}	2.12×10^{-7}	9.08×10^{-5}	2.35×10^{-5}
PN	6.68×10^{-1}	—	6.73×10^{-4}	3.66×10^{-4}	8.61×10^{-3}	3.48×10^{-3}	2.21×10^{-3}		7.16×10^{-5}		2.02×10^{-4}		1.70×10^{-3}	4.92×10^{-5}
SN	1.18×10^{-3}	1.91×10^{-5}	—	4.48×10^{-6}	2.88×10^{-4}	8.16×10^{-4}	2.85×10^{-5}	8.81×10^{-4}		2.55×10^{-5}	1.10×10^{-4}	2.58×10^{-8}	1.05×10^{-3}	2.26×10^{-4}
Mg	4.47×10^{-1}	1.97×10^{-3}	8.50×10^{-4}	—	3.45×10^{-1}		8.25×10^{-2}		3.57×10^{-5}			2.37×10^{-6}	1.49×10^{-3}	
meta	7.65×10^{-1}	2.24×10^{-3}	2.64×10^{-3}	1.67×10^{-2}	—	3.68×10^{-3}	7.85×10^{-3}	2.19×10^{-5}	3.42×10^{-4}		1.50×10^{-4}	8.59×10^{-7}	1.07×10^{-3}	9.01×10^{-5}
Pd	4.87×10^{-1}	1.78×10^{-3}	1.47×10^{-2}		7.22×10^{-3}	—	8.42×10^{-5}	1.01×10^{-4}		1.61×10^{-4}	1.46×10^{-4}	2.56×10^{-6}	4.79×10^{-3}	8.32×10^{-5}
LN	1.01×10^{-1}	5.16×10^{-4}	2.35×10^{-4}	3.59×10^{-3}	7.06×10^{-3}	3.85×10^{-5}	—	6.35×10^{-4}	2.16×10^{-3}		6.42×10^{-5}	7.31×10^{-6}		5.52×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	—		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	6.05×10^{-2}	2.34×10^{-4}		2.17×10^{-5}	4.29×10^{-3}		3.02×10^{-2}		—			2.71×10^{-6}		
Lo			2.27×10^{-3}			7.98×10^{-4}		8.95×10^{-3}		—	4.04×10^{-4}	1.74×10^{-6}	5.14×10^{-2}	8.69×10^{-3}
I	1.27×10^{-2}	1.44×10^{-3}	2.76×10^{-2}		4.10×10^{-3}	2.04×10^{-3}	1.95×10^{-3}	6.74×10^{-4}		1.13×10^{-3}	—	3.77×10^{-6}	1.25×10^{-2}	6.50×10^{-3}
W	1.00×10^{-2}		2.42×10^{-4}	1.17×10^{-4}	8.77×10^{-4}	1.33×10^{-3}	8.30×10^{-3}	1.01×10^{-4}	2.21×10^{-4}	1.83×10^{-4}	1.41×10^{-4}	—	1.05×10^{-5}	1.97×10^{-1}
other	9.16×10^{-3}	9.63×10^{-4}	2.10×10^{-2}	1.57×10^{-4}	2.34×10^{-3}	5.31×10^{-3}		7.65×10^{-4}		1.15×10^{-2}	1.00×10^{-3}	2.24×10^{-8}	—	2.94×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	—

$$\mathbf{p}^T(t) = \mathbf{p}^T(0)\exp(\mathbf{K}t)$$





fast time scale 200 ns

slow time scale 2 μs

Experimental t_1=150 ns, t_2= 2.2 μs

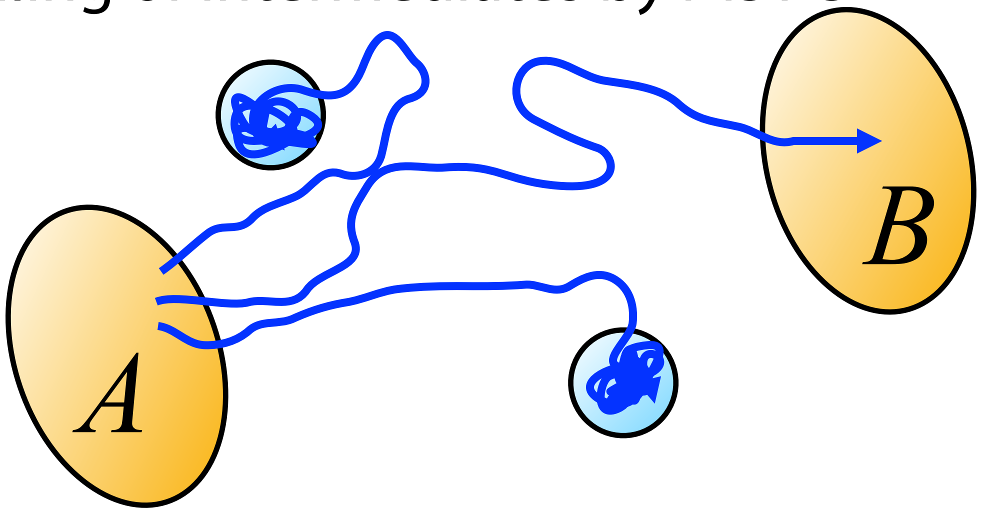
SN state

to end

*Du & PGB, JCP **140**, 195102 (2014)).*

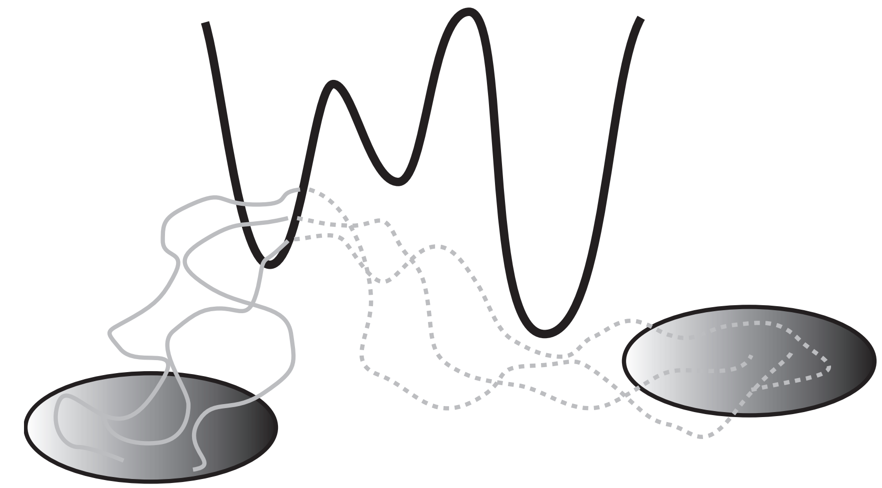
Evolution of path sampling algorithms

Handling of intermediates by MSTIS



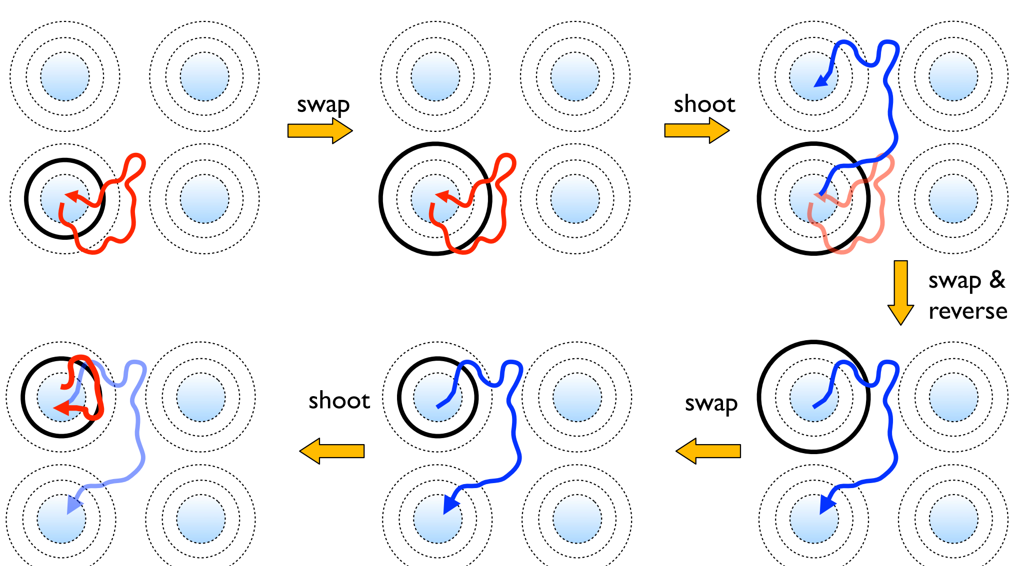
J. Rogal, PGB, JCP **129**, 224107 (2008).

Improved convergence by RETIS



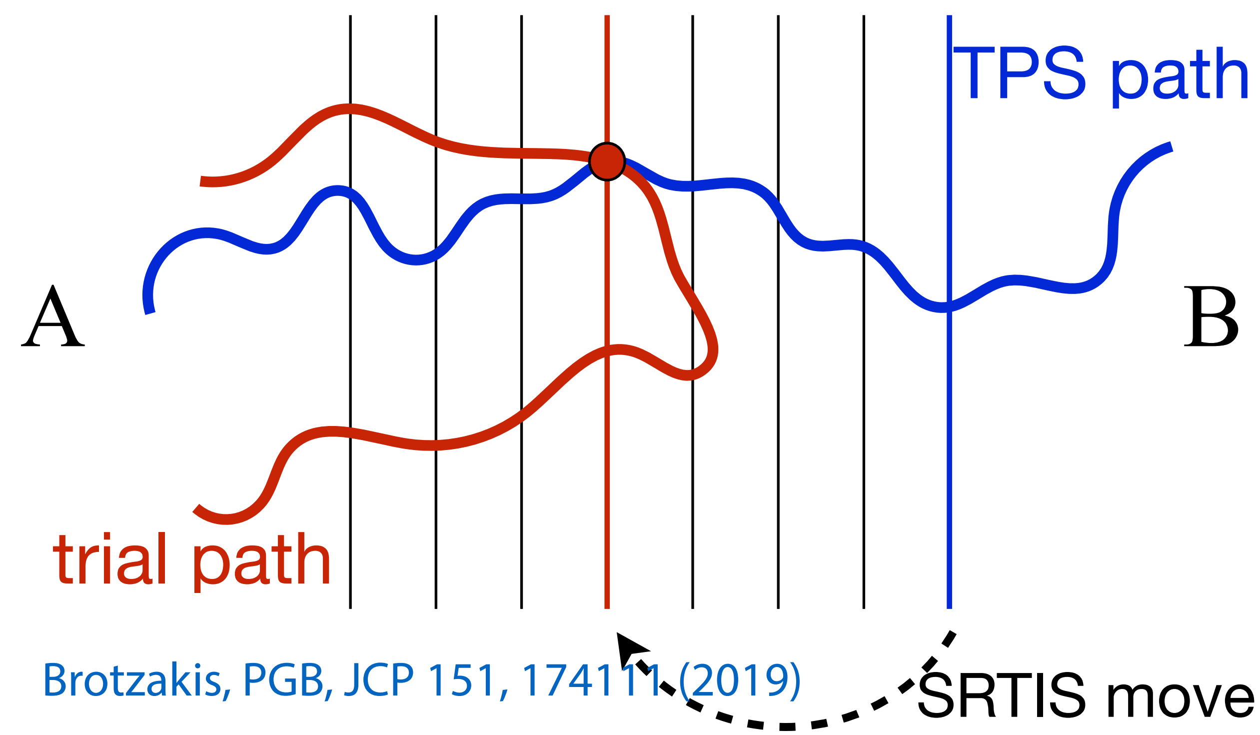
T.S. van Erp, PRL **98**, 268301 (2007)
PGB, JCP **129**,114108 (2008)

Avoiding large amount of replicas by SRTIS



Du & PGB, JCP **140**, 195102 (2014)).

RPE from TPS by Virtual Interface Exchange

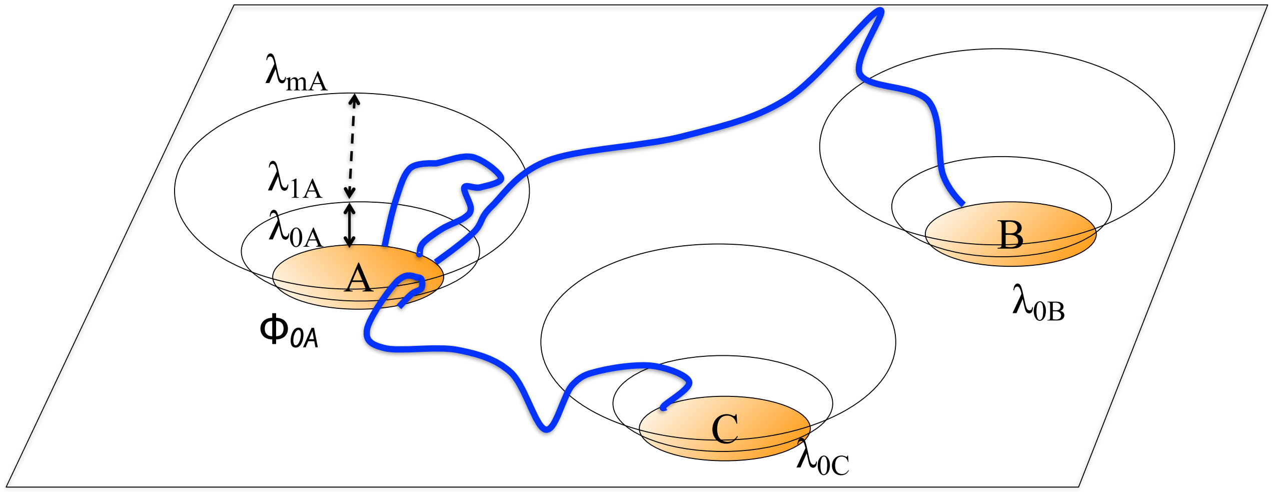


Brotzakis, PGB, JCP 151, 174111 (2019)

Reweighting schemes allow reconstruction of unbiased dynamical trajectory ensemble

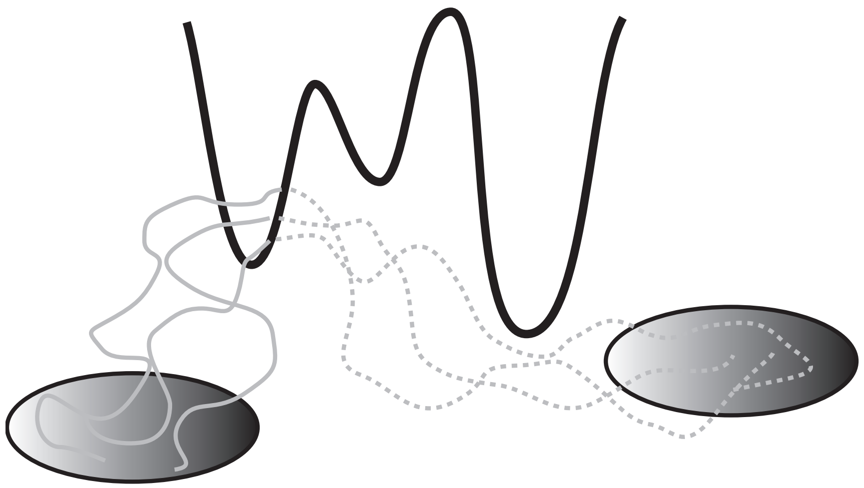
Evolution of path sampling algorithms

Handling of intermediates by MSTIS



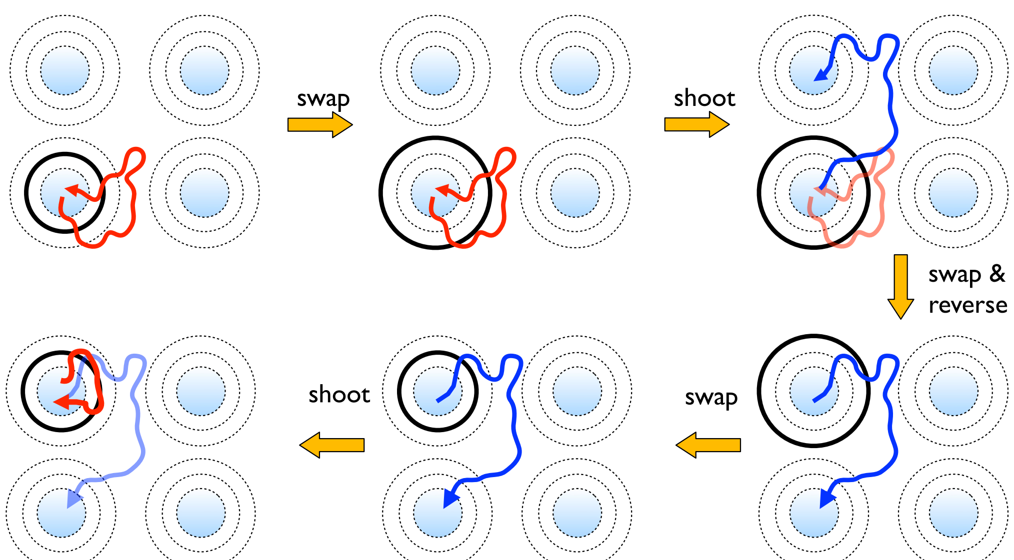
J. Rogal, PGB, JCP **129**, 224107 (2008).

Improved convergence by RETIS



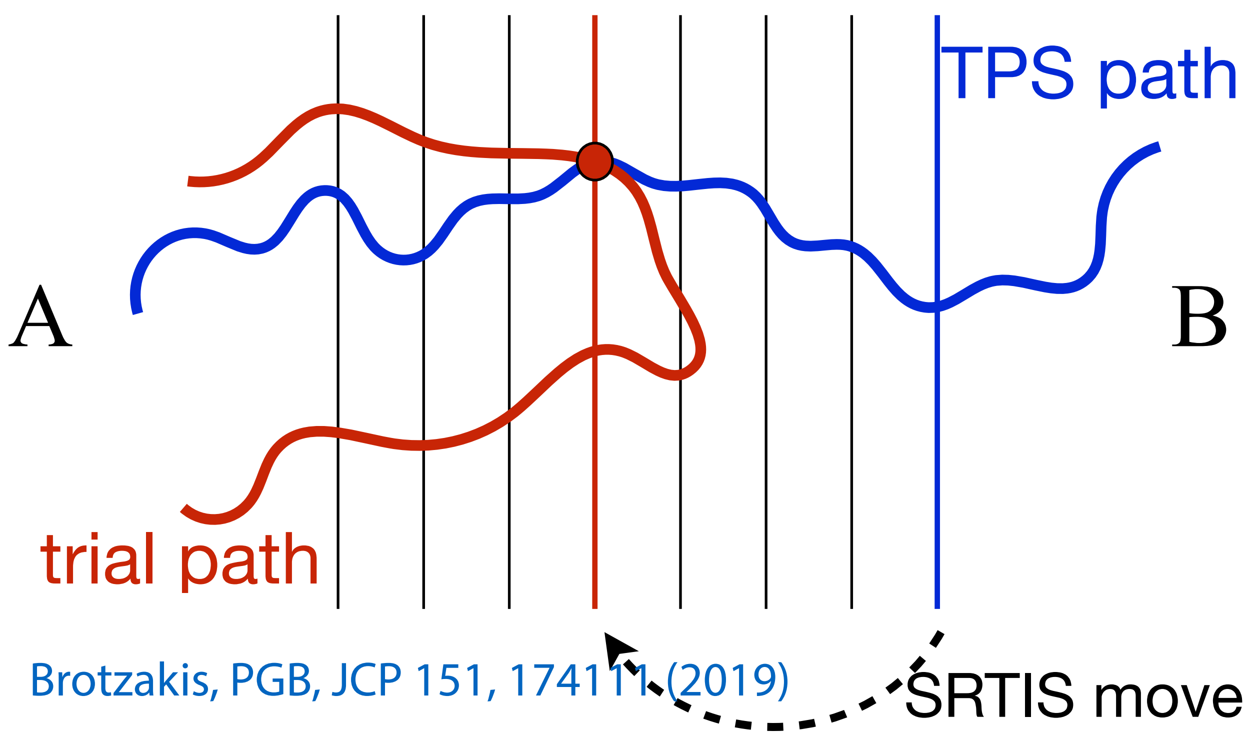
T.S. van Erp, PRL **98**, 268301 (2007)
PGB, JCP **129**,114108 (2008)

Avoiding large amount of replicas by SRTIS



Du & PGB, JCP **140**, 195102 (2014)).

RPE from TPS by Virtual Interface Exchange

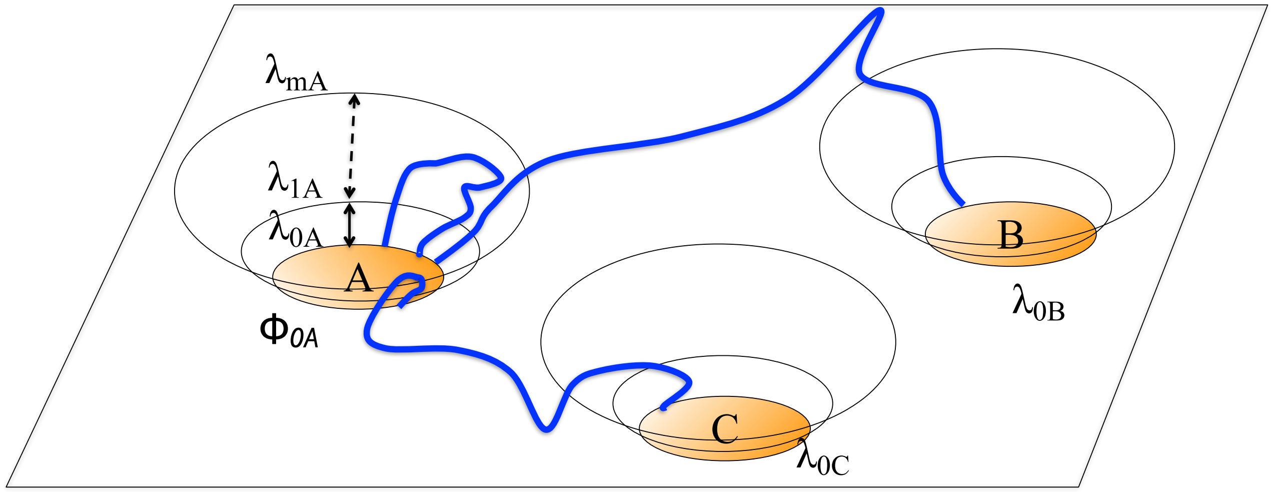


Brotzakis, PGB, JCP 151, 174111 (2019)

Reweighting schemes allow reconstruction of unbiased dynamical trajectory ensemble

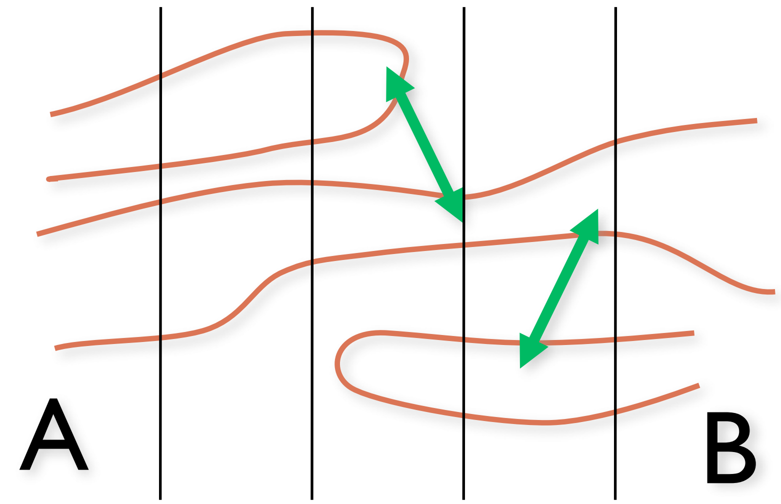
Evolution of path sampling algorithms

Handling of intermediates by MSTIS



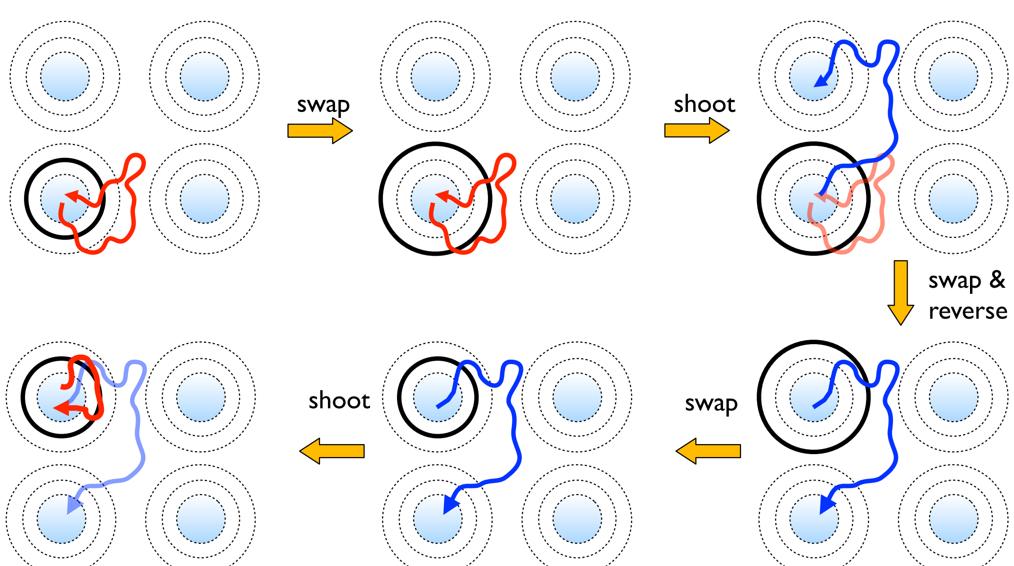
J. Rogal, PGB, JCP **129**, 224107 (2008).

Improved convergence by RETIS



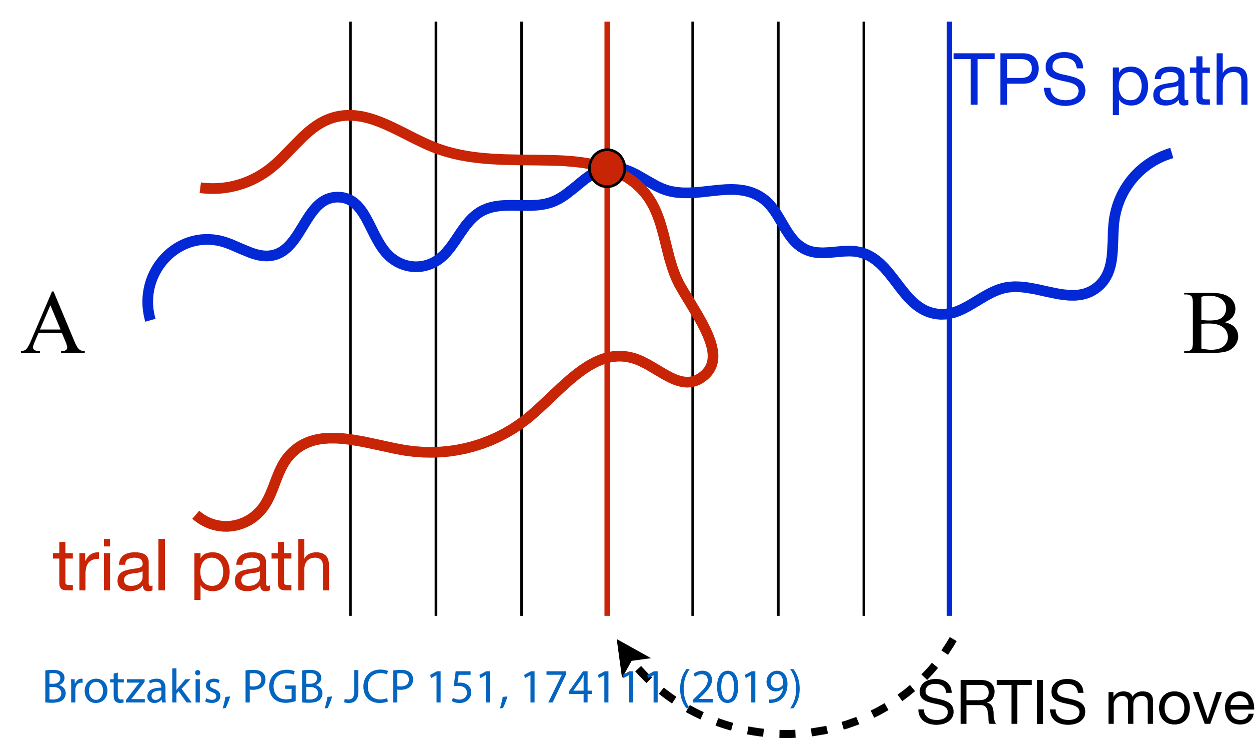
T.S. van Erp, PRL **98**, 268301 (2007)
PGB, JCP **129**,114108 (2008)

Avoiding large amount of replicas by SRTIS



Du & PGB, JCP **140**, 195102 (2014)).

RPE from TPS by Virtual Interface Exchange

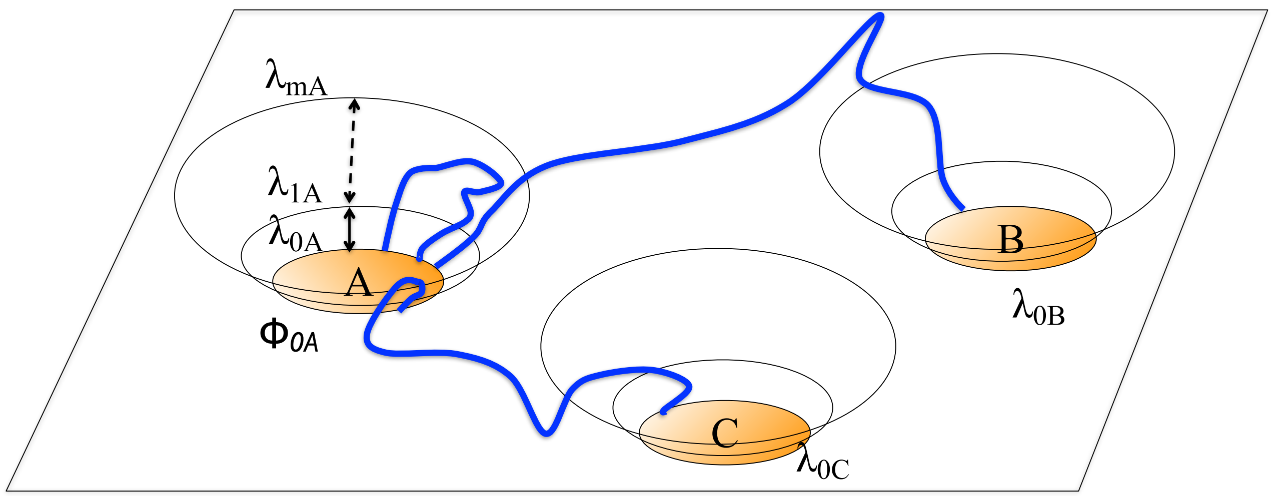


Brotzakis, PGB, JCP 151, 174111 (2019)

Reweighting schemes allow reconstruction of unbiased dynamical trajectory ensemble

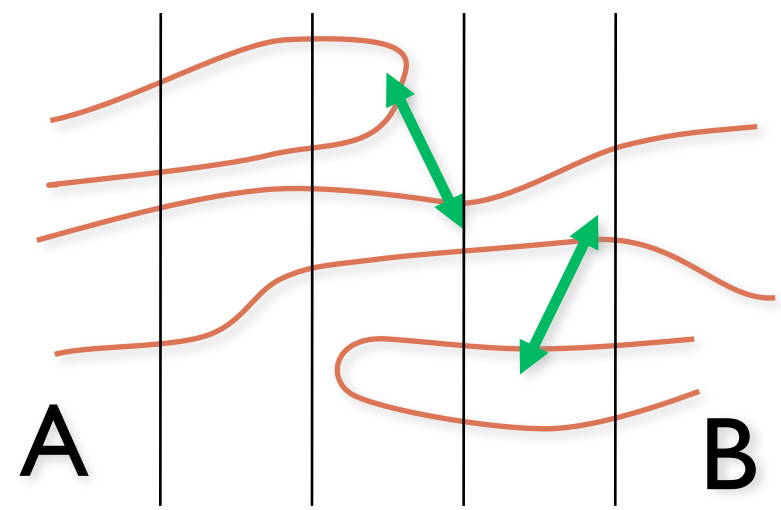
Evolution of path sampling algorithms

Handling of intermediates by MSTIS



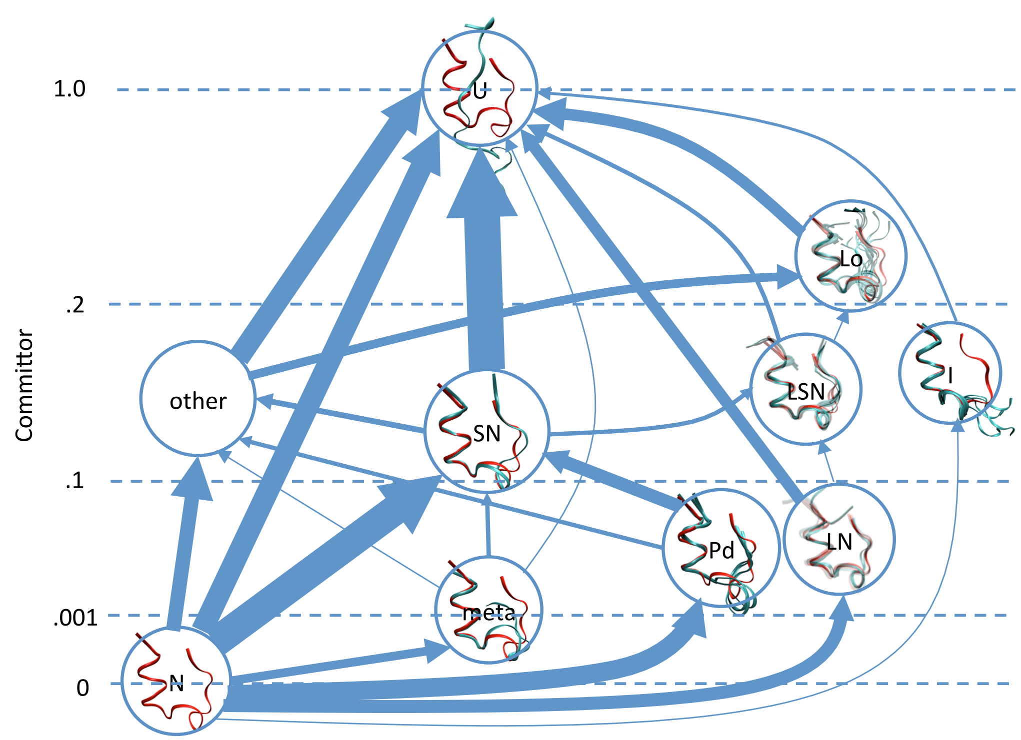
J. Rogal, PGB, JCP **129**, 224107 (2008).

Improved convergence by RETIS

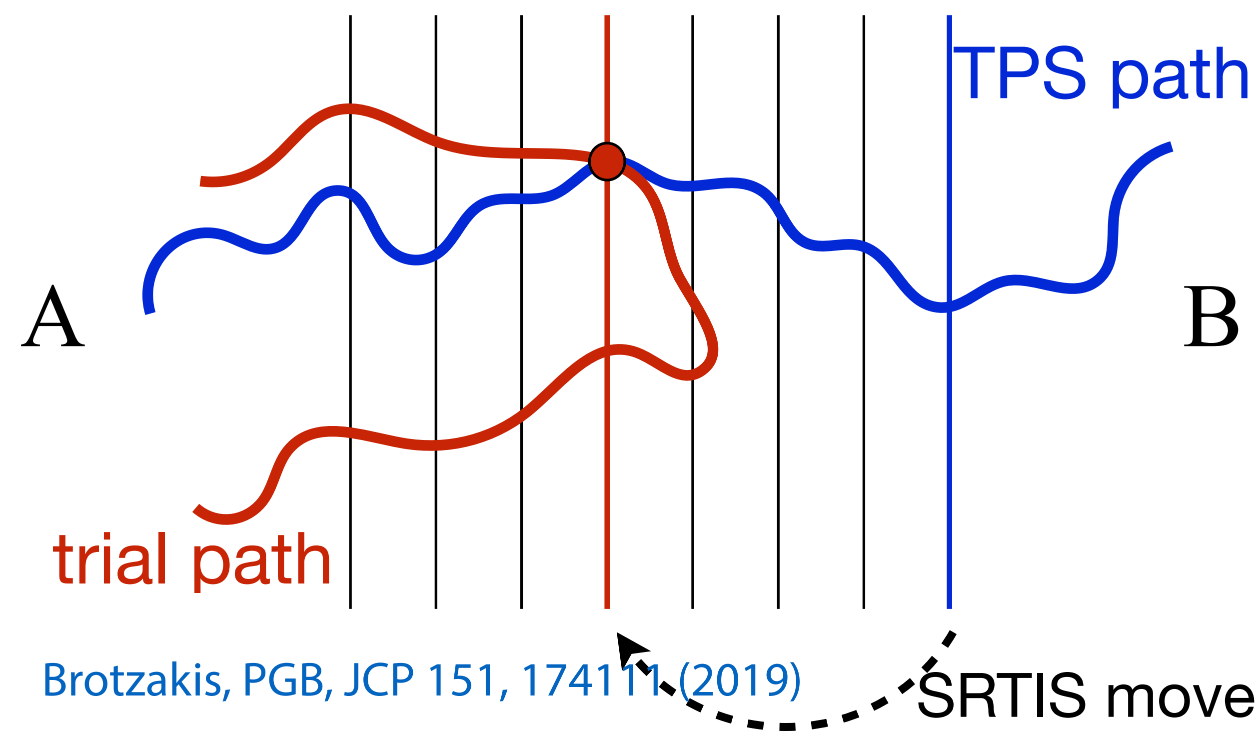


T.S. van Erp, PRL **98**, 268301 (2007)
PGB, JCP **129**,114108 (2008)

Avoiding large amount of replicas by SRTIS



RPE from TPS by Virtual Interface Exchange

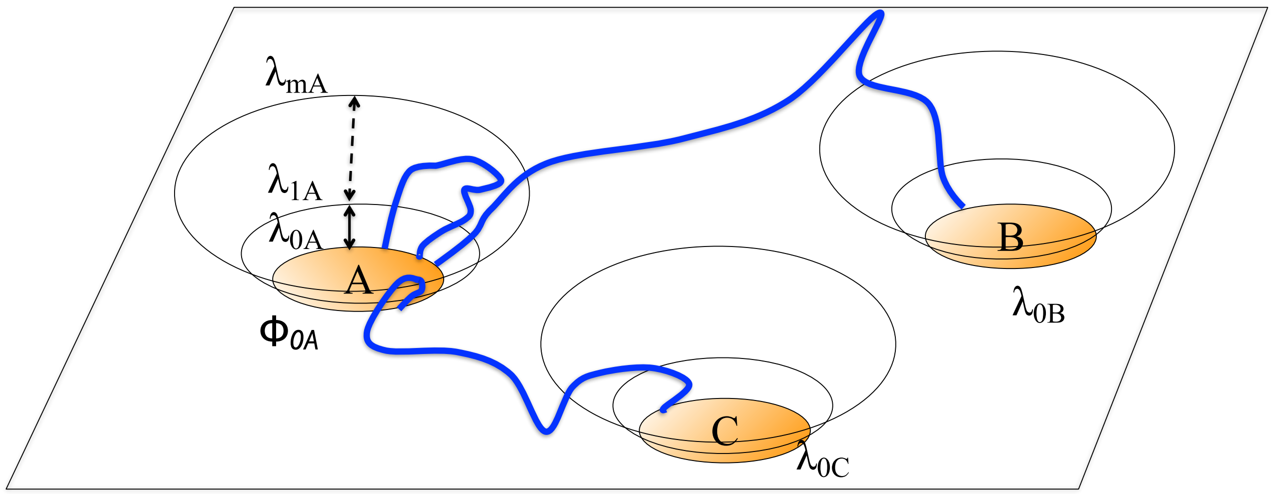


Brotzakis, PGB, JCP 151, 174111 (2019)

Reweighting schemes allow reconstruction of unbiased dynamical trajectory ensemble

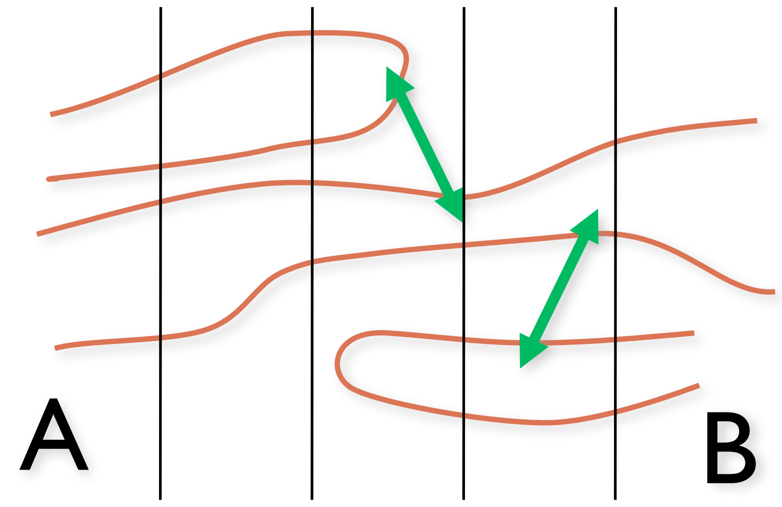
Evolution of path sampling algorithms

Handling of intermediates by MSTIS



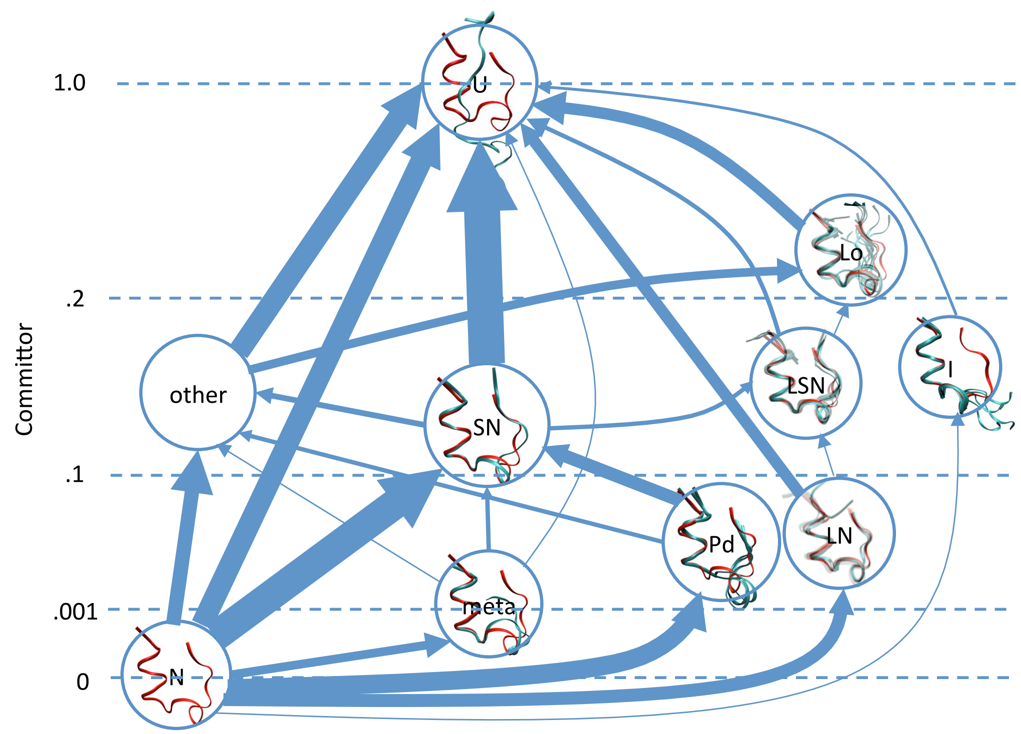
J. Rogal, PGB, JCP **129**, 224107 (2008).

Improved convergence by RETIS

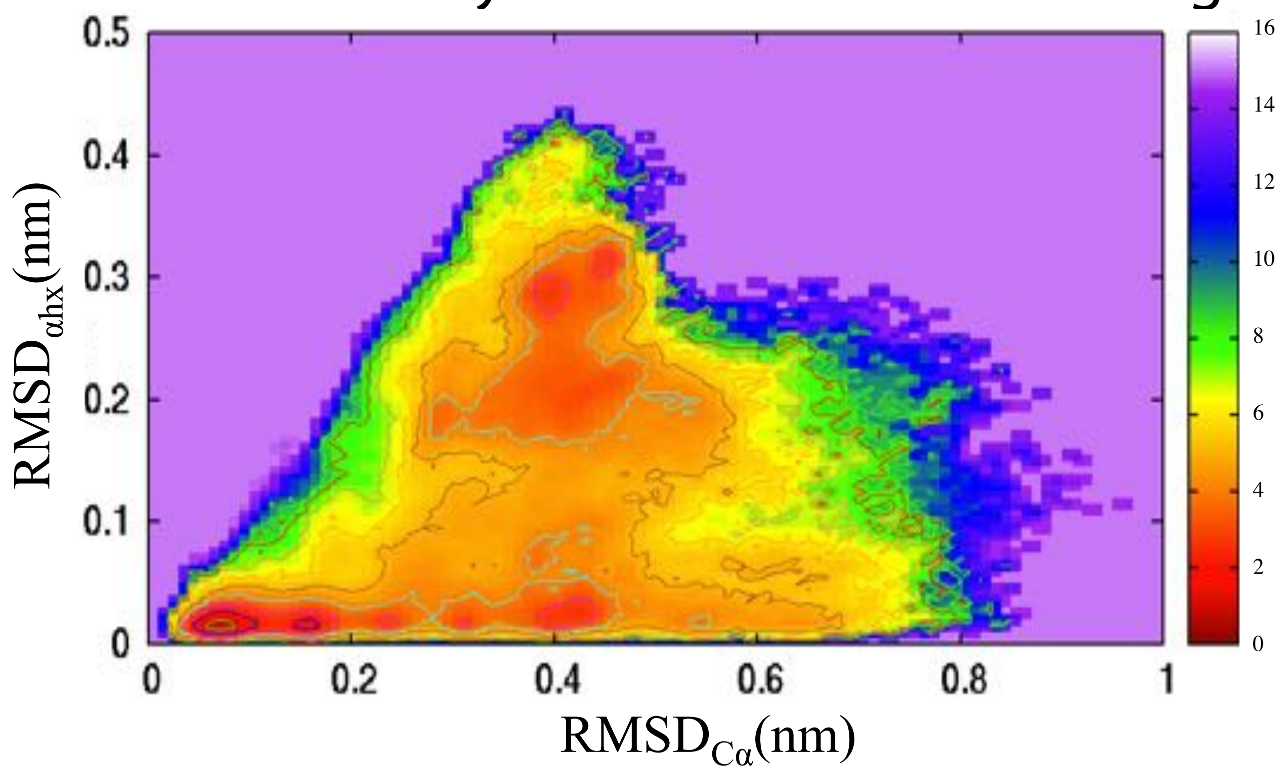


T.S. van Erp, PRL **98**, 268301 (2007)
PGB, JCP **129**,114108 (2008)

Avoiding large amount of replicas by SRTIS



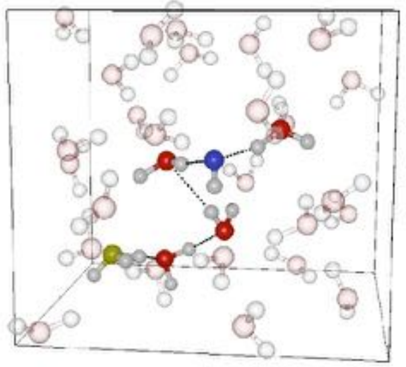
RPE from TPS by Virtual Interface Exchange

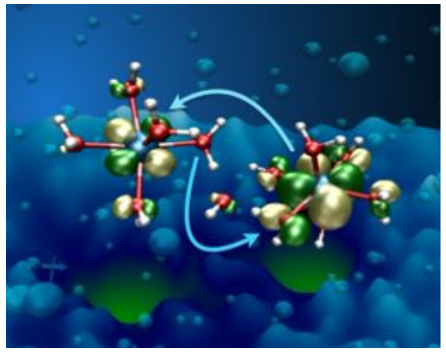


Reweighting schemes allow reconstruction of unbiased dynamical trajectory ensemble

Selected TPS applications

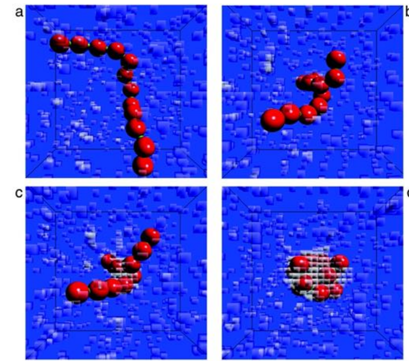
Chemical reactions in solution

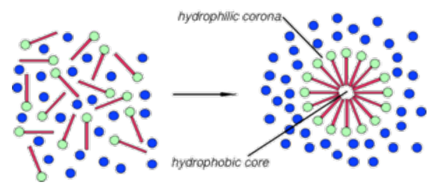




Geissler et al Science 2001; Tiwari and Ensing, Farad Disc 2016; Joswiak et al, PNAS 2017 ….

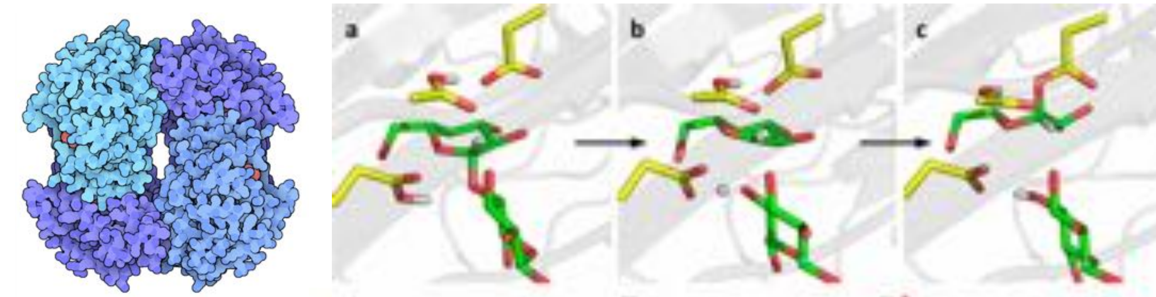
Microphases

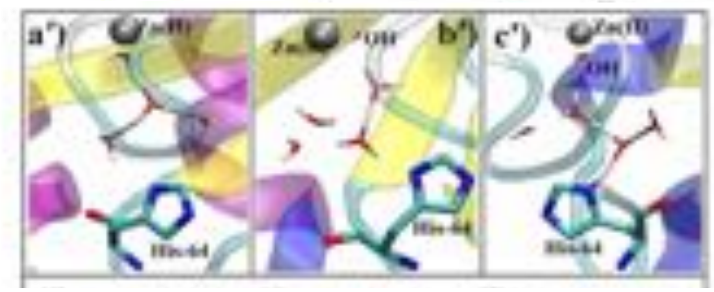


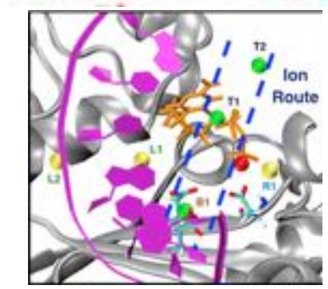


Ten Wolde et al PNAS 2002; Pool & PGB JCP 2007

Enzymatic reactions

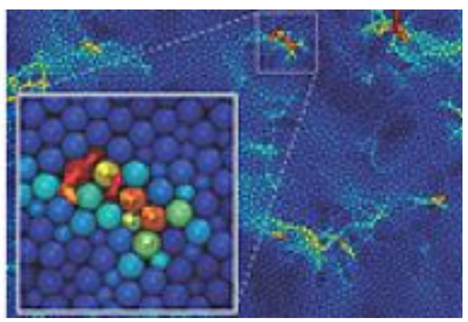


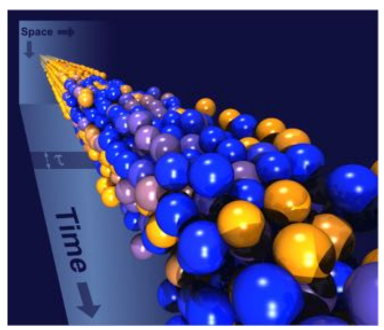




Basner et Schwarz, JACS 2005; Knott et al, JACS 2013;Li et al JACS 2014; Paul and Taraphder, ChemPhysChem 2020; Silveira et al, JPCB 2021;….

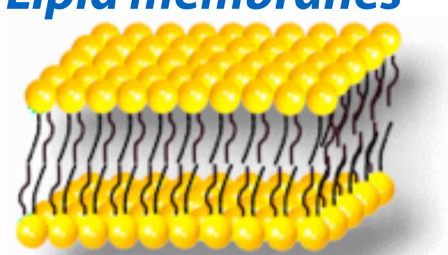
Glasses

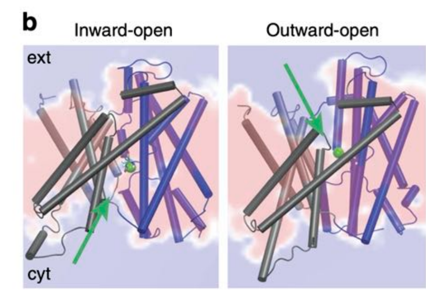


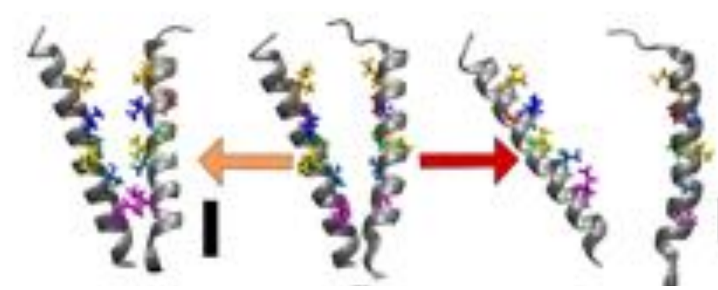


Hedges et al Science 2009; Jack etc al PRL 2011; Turci et al PRX 2017; ….

Lipid membranes

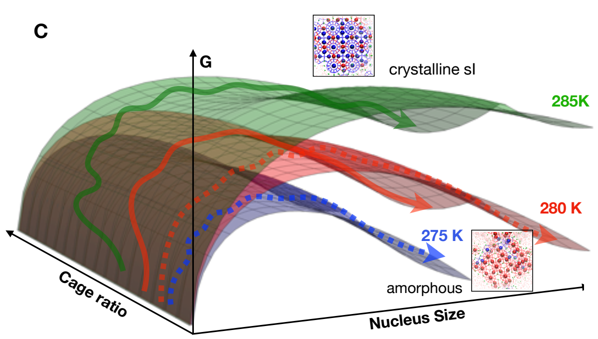


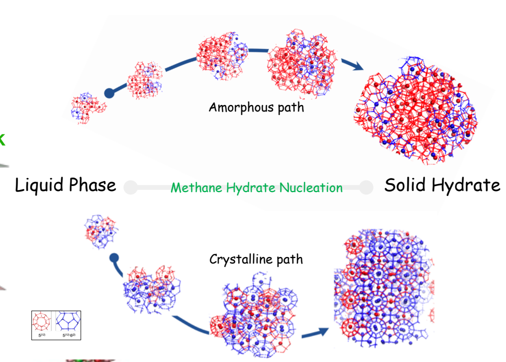


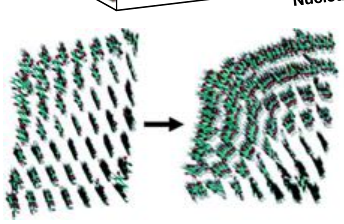


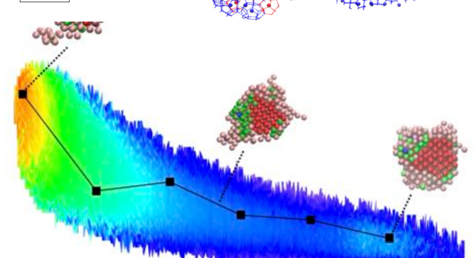
Marti & Csajka 2004; Okazaki et al Nat Comm. 2019 . Domanski, et al PLOS Comput. Biol. 2020,; …..

Crystal Nucleation



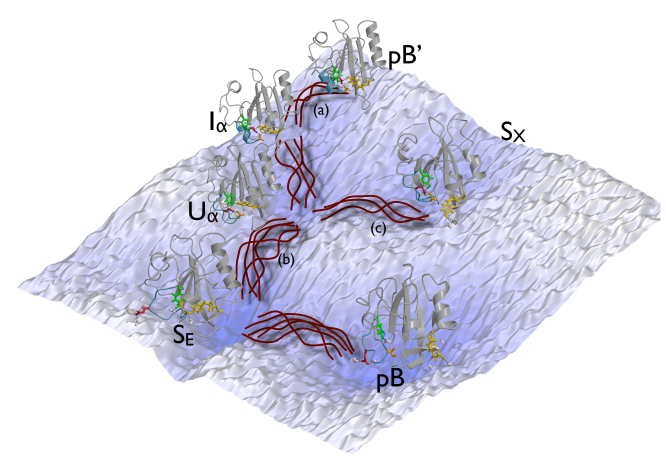


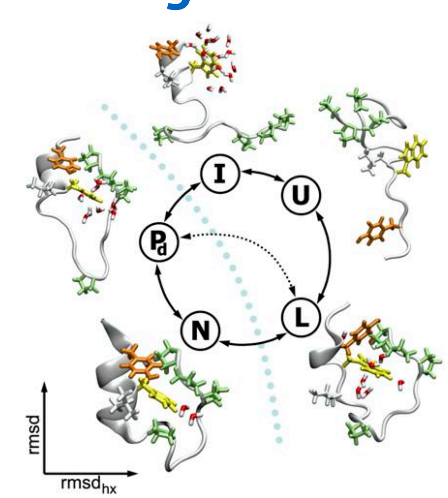


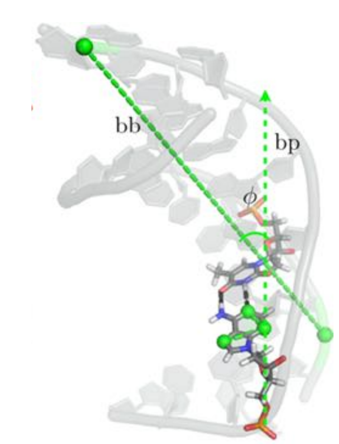


Moroni et al PRL 2005, Bekcham et al JACS 2007l Lechner et al. PRL 2011; Diaz Leines & Rogal JPCB 2018; Arjun et al PNAS 2019….

Biomolecular conformational change







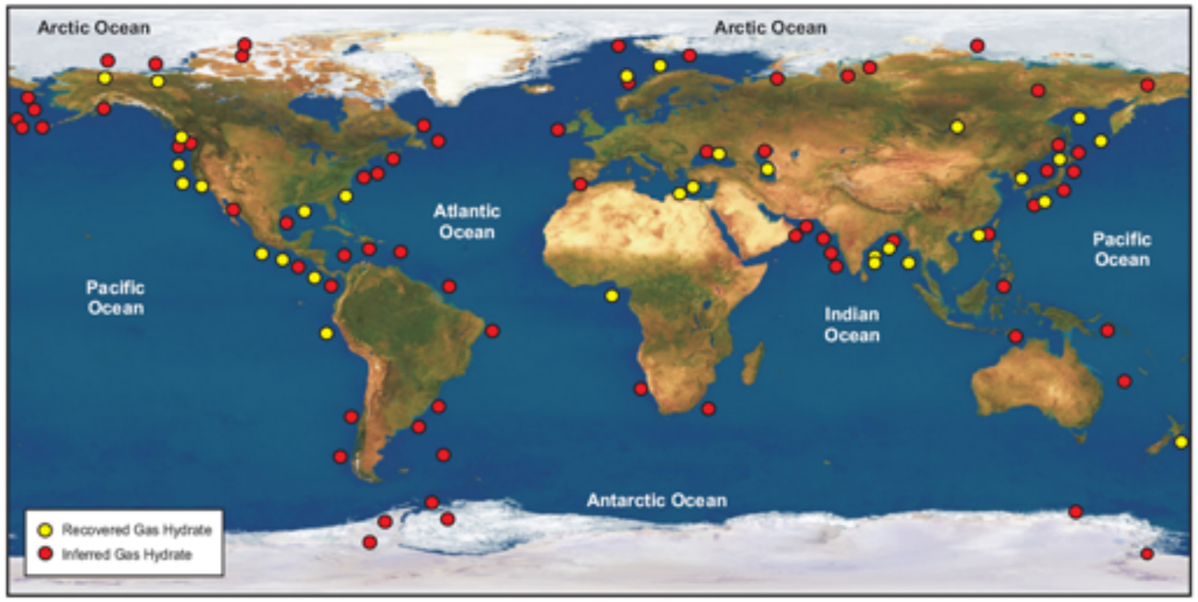
Bolhuis PNAS 2003; Juraszek & Bolhuis 2006; Vreede et al PNAS 2010; Best & Hummer PNAS 2016; Brotzakis & PGB, JPCB 2019, Vreede et al. NAR 2019…….

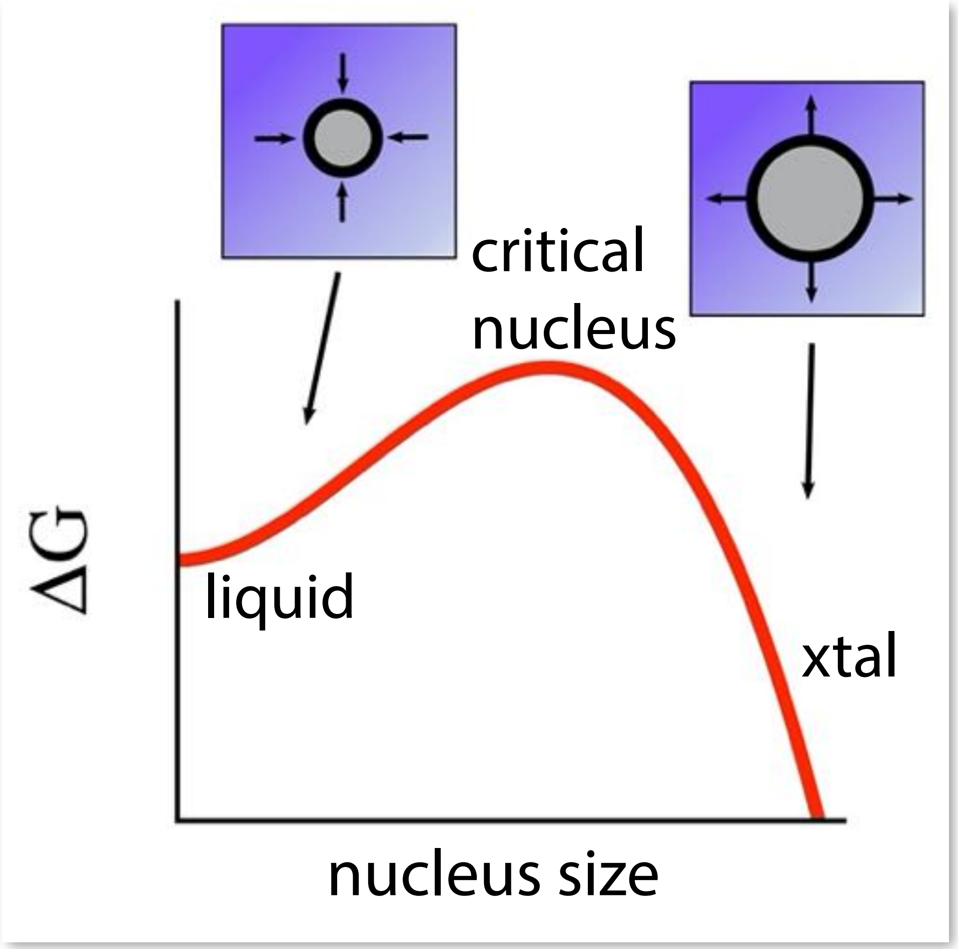
Gas hydrate nucleation

- Gas hydrates provide
  - large repository of natural gas
  - possibility for CO2 sequestration
  - problem in pipes
- How do hydrates nucleate?

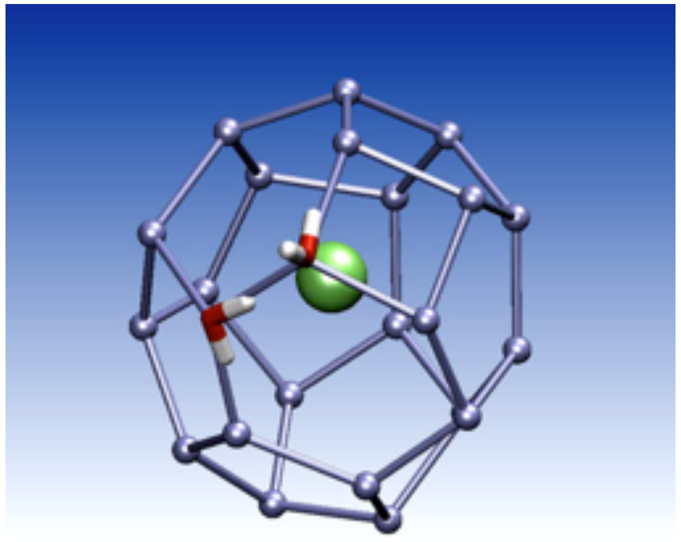
- Methane hydrate nucleation hypothesis
  - amorphous critical nucleus
  - transforms into crystalline form

- experimental testing difficult: simulations

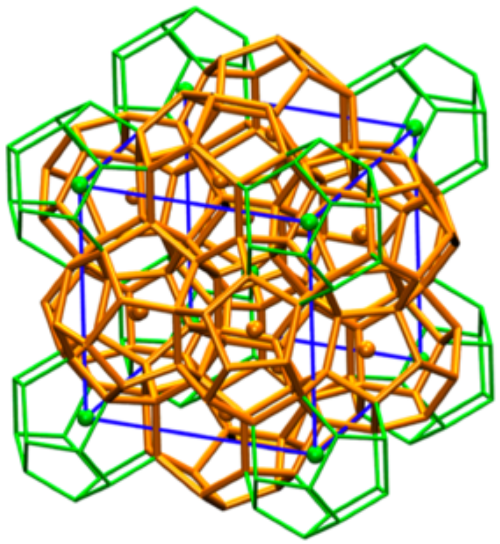




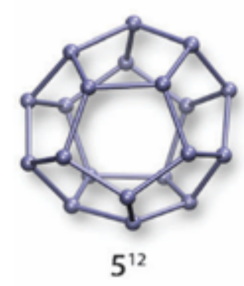
Stable phase and Order Parameters



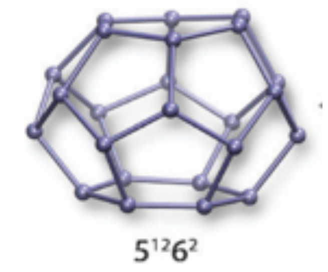
single methane cage



SI Unit Cell Crystal

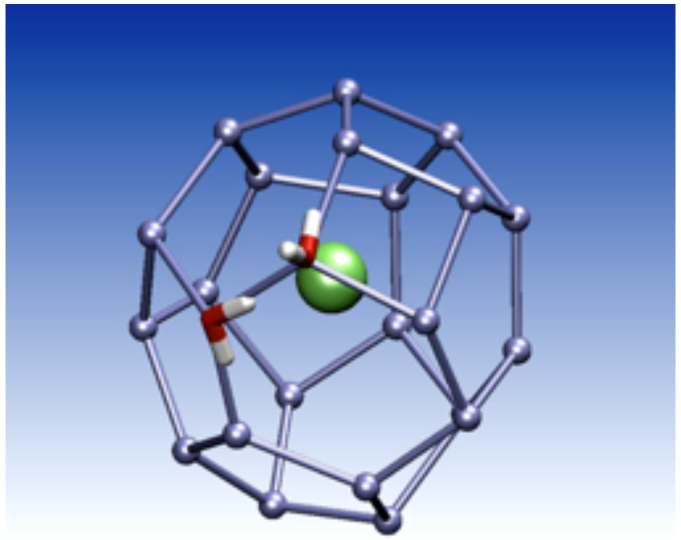


Small Cage

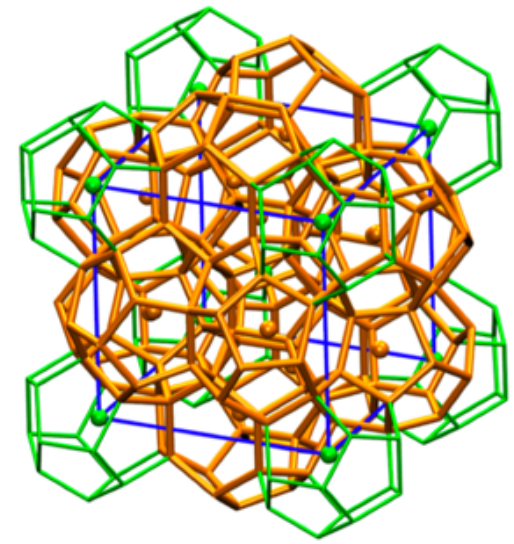


Big Cage

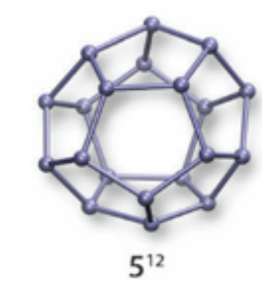
Stable phase and Order Parameters



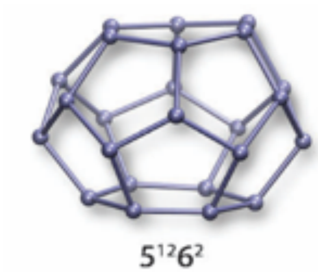
single methane cage



SI Unit Cell Crystal



Small Cage

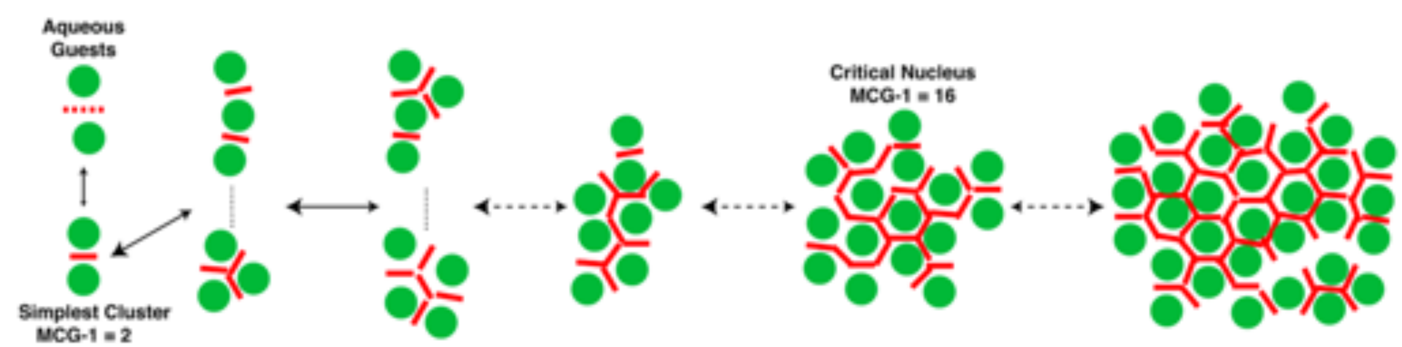


Big Cage

---

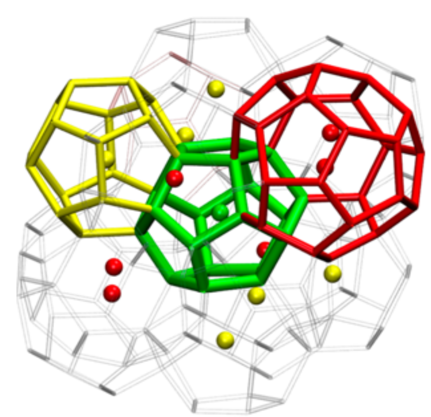
we focus on two families of order parameters

SIZE: MCG
(Mutually Coordinated Guests)



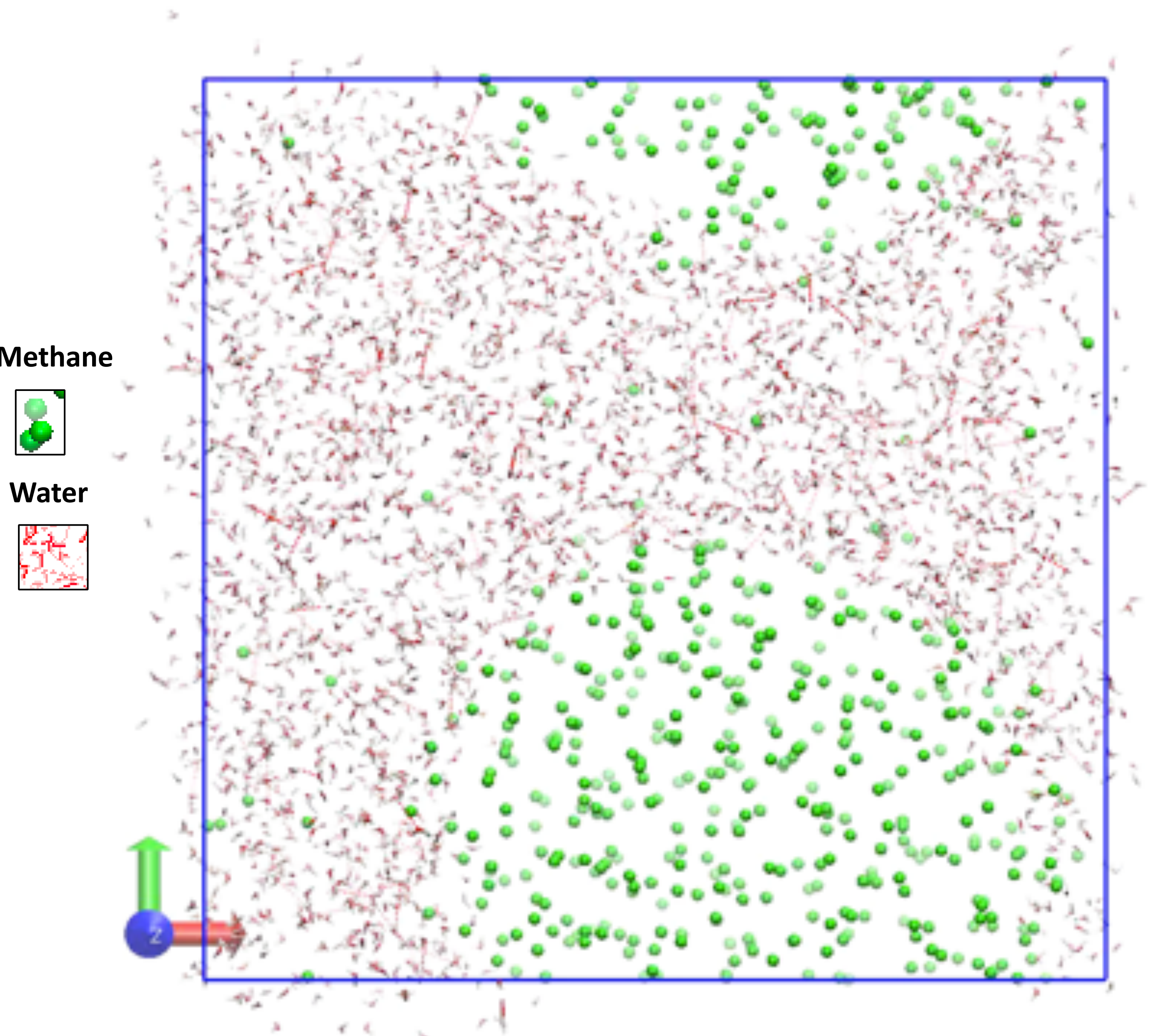
Barnes et. al (2014)

SHAPE: Cage type Identification
Cluster Analysis using MCG
Cage Ratio $5^{12}6^{2}/5^{12} = 3$ for pure SI crystal
smaller than 1 for amorphous solids

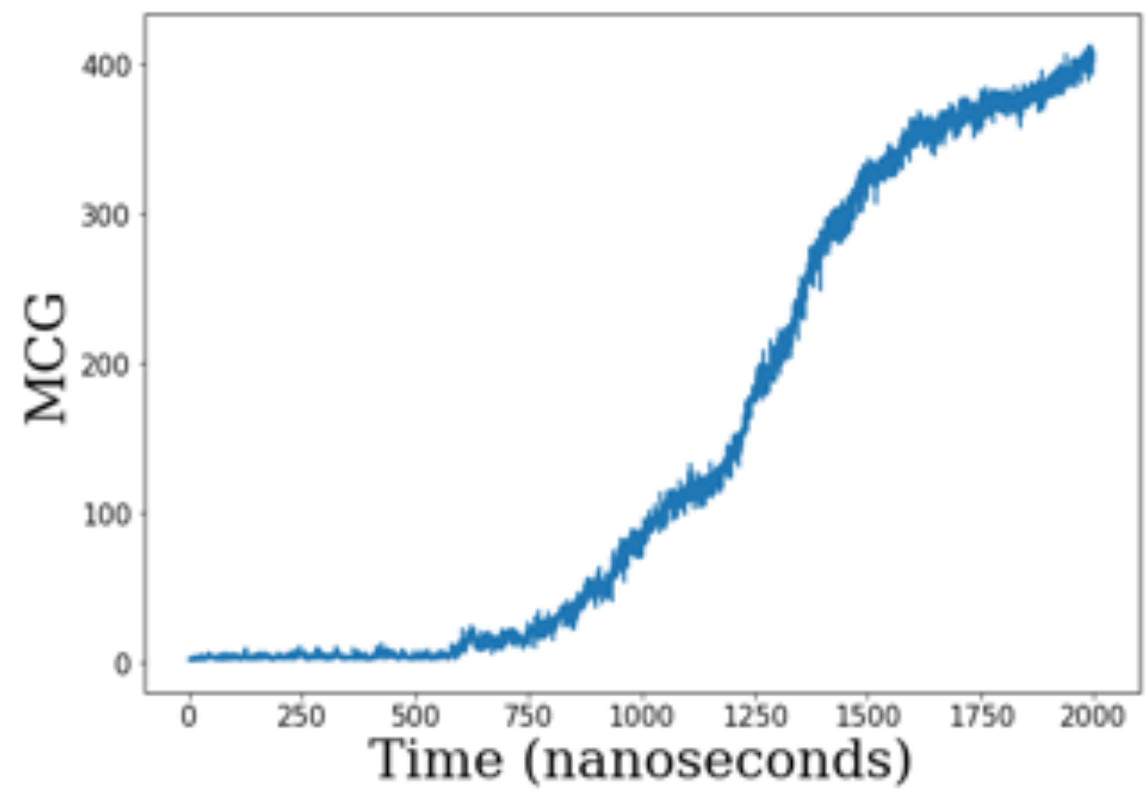


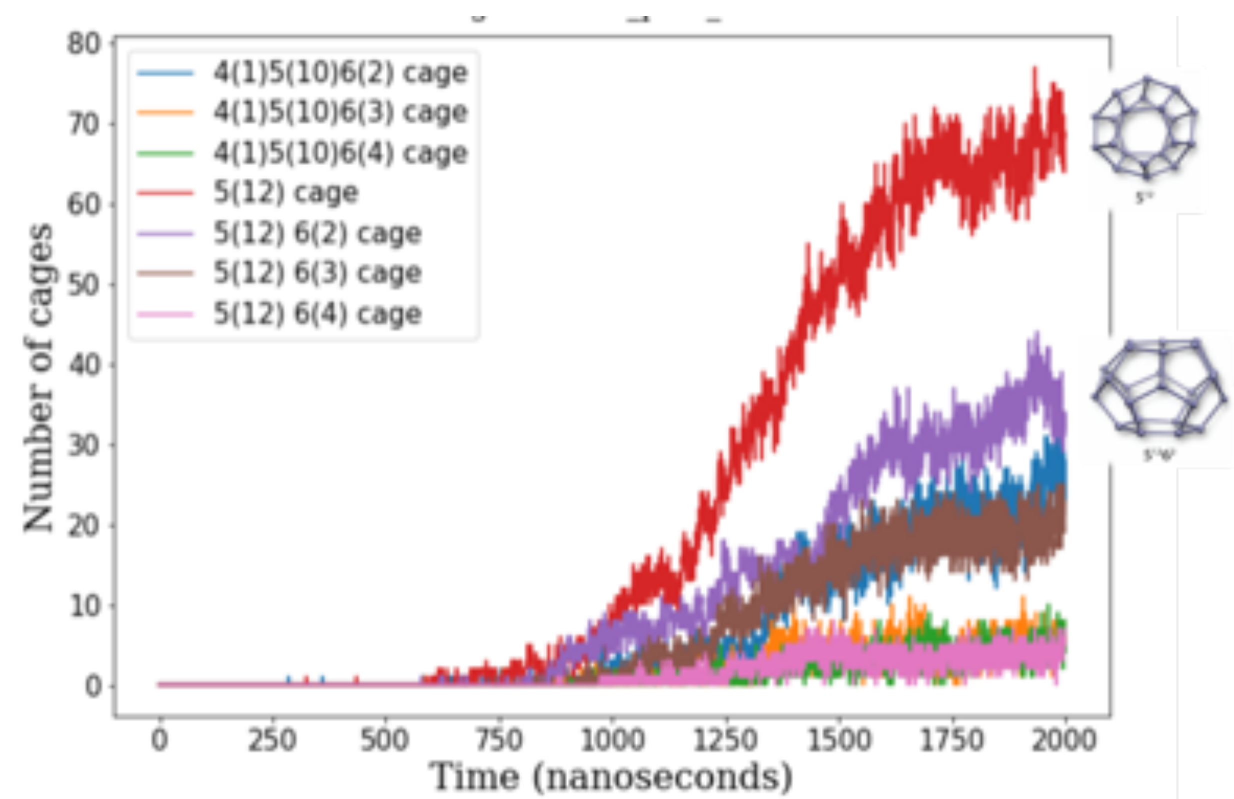
Arjun, Berendsen, PGB, PNAS, 2019

Brute force MD at 250 K forms amorphous solid



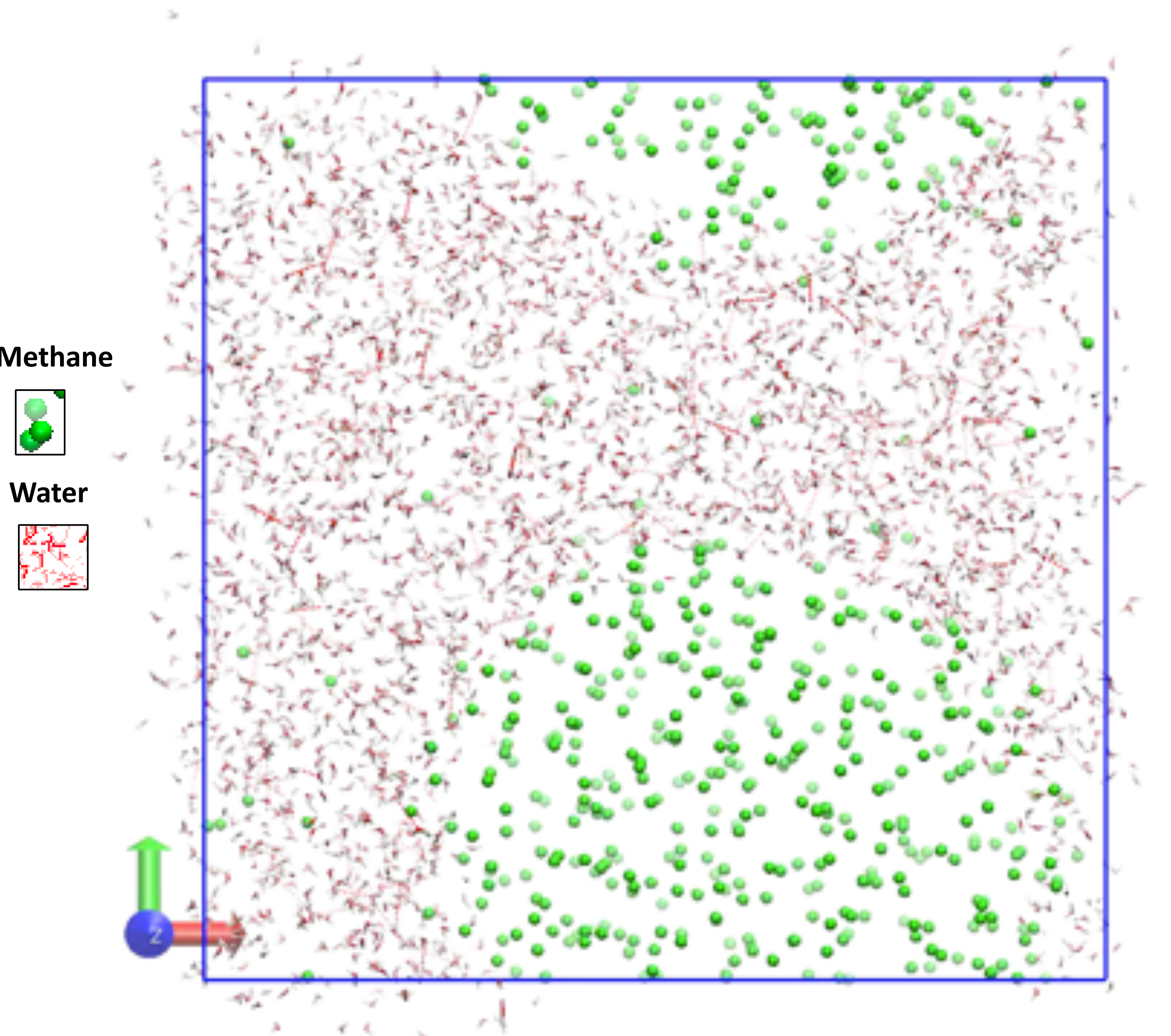
2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K, 2 µs



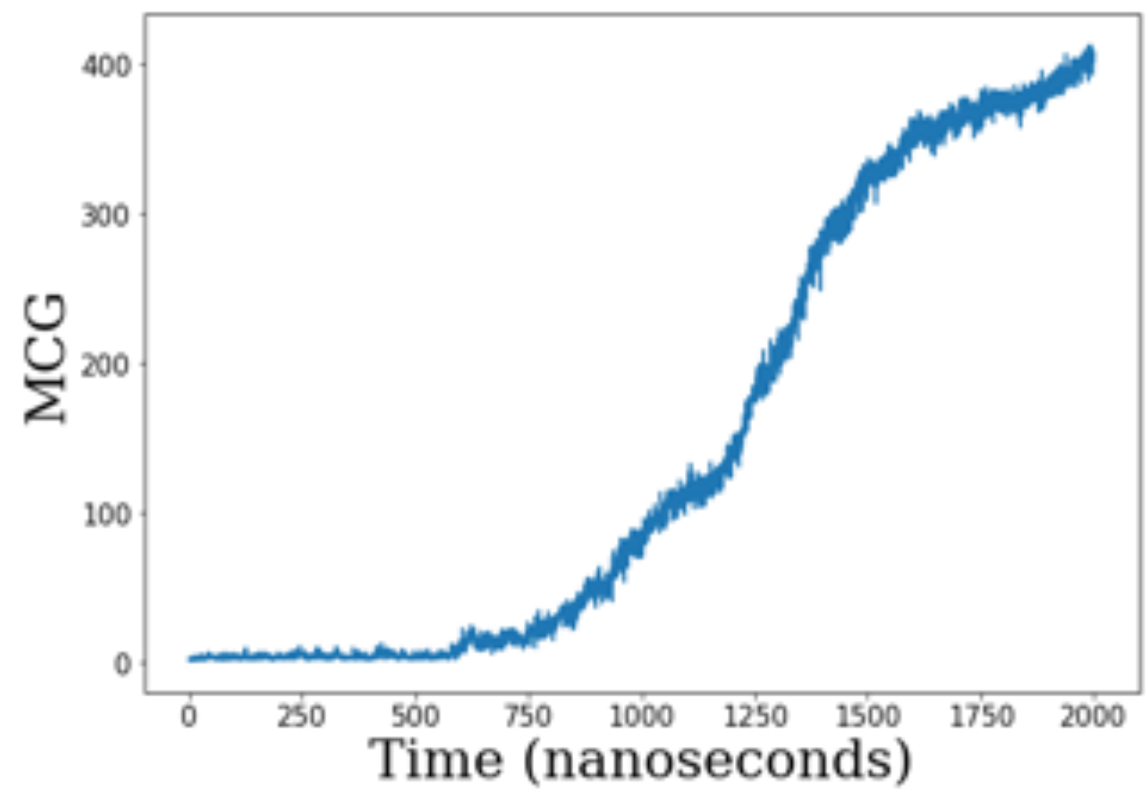


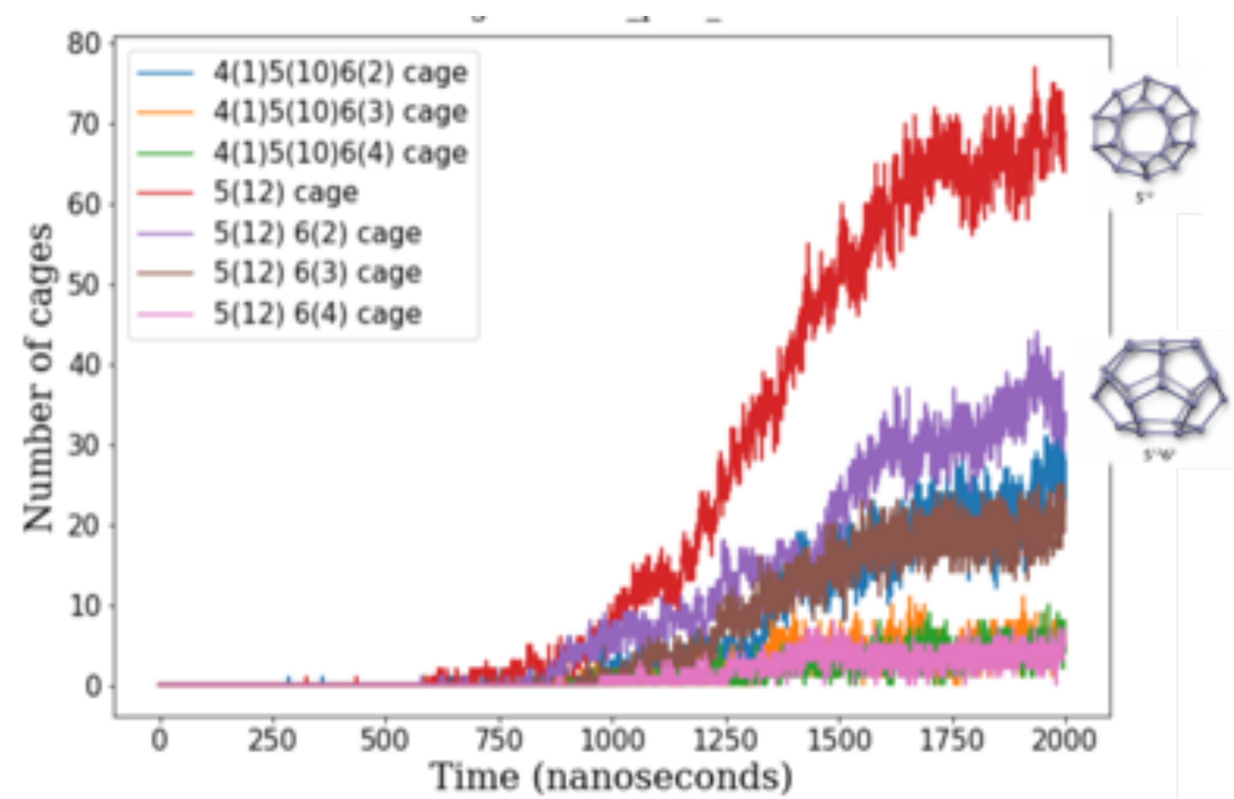
Walsh, M. R Science 2009, 326, 1095–1098

Brute force MD at 250 K forms amorphous solid



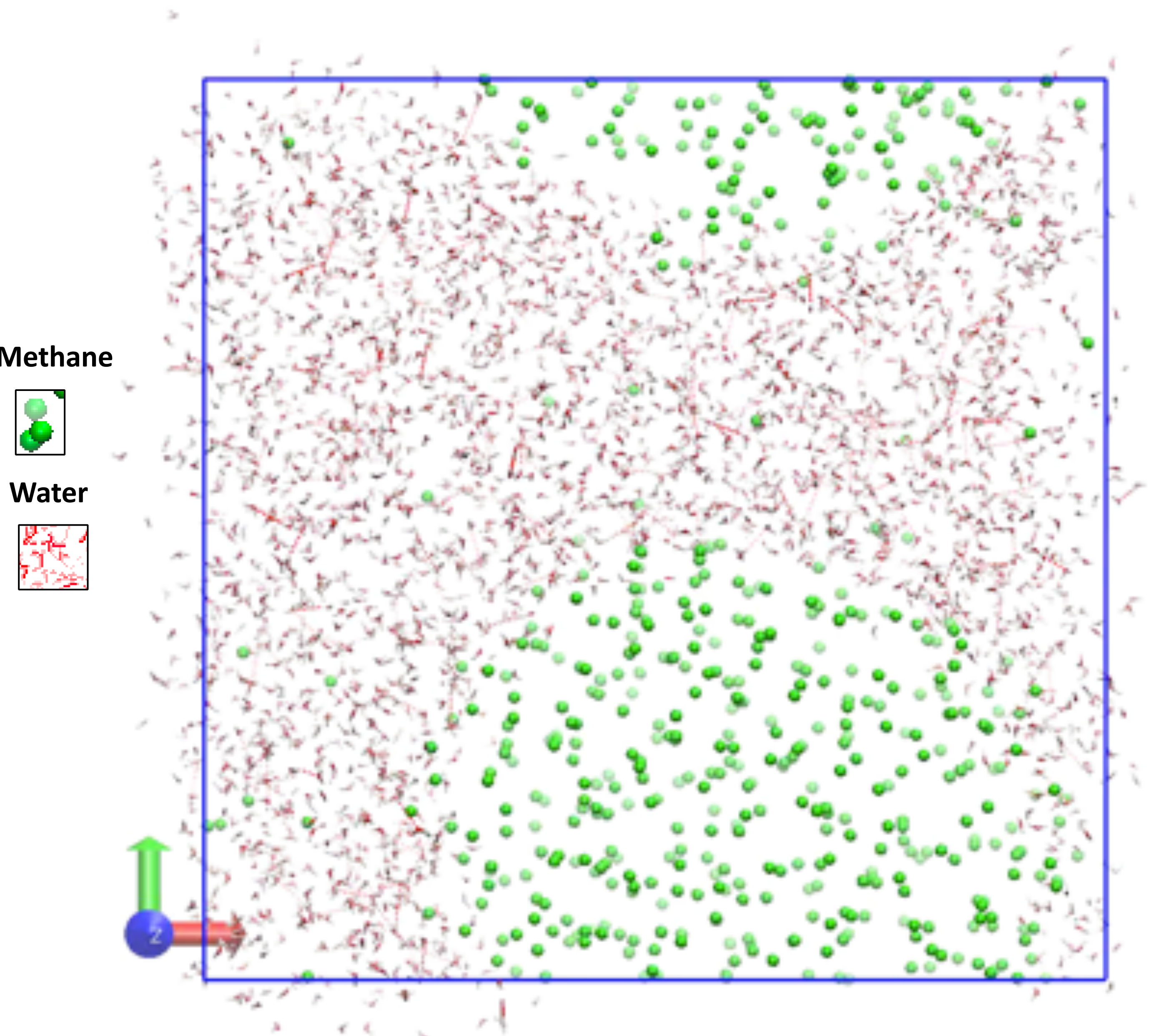
2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K, 2 μs



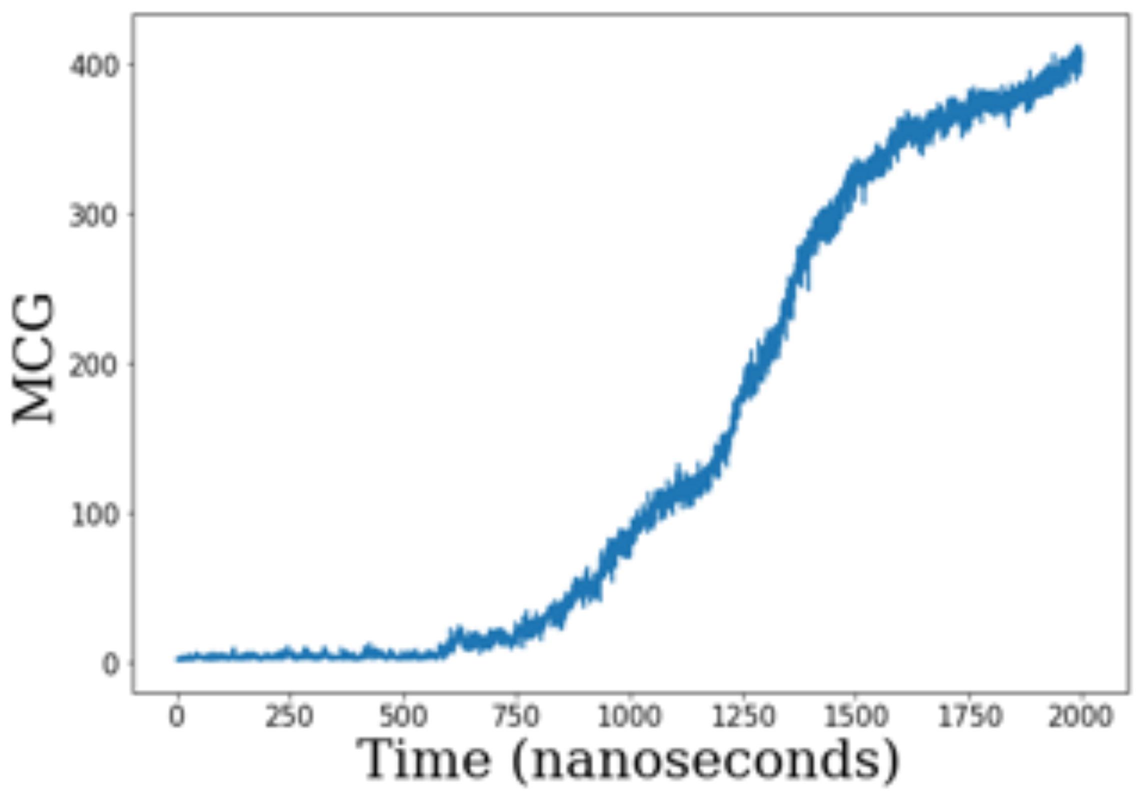


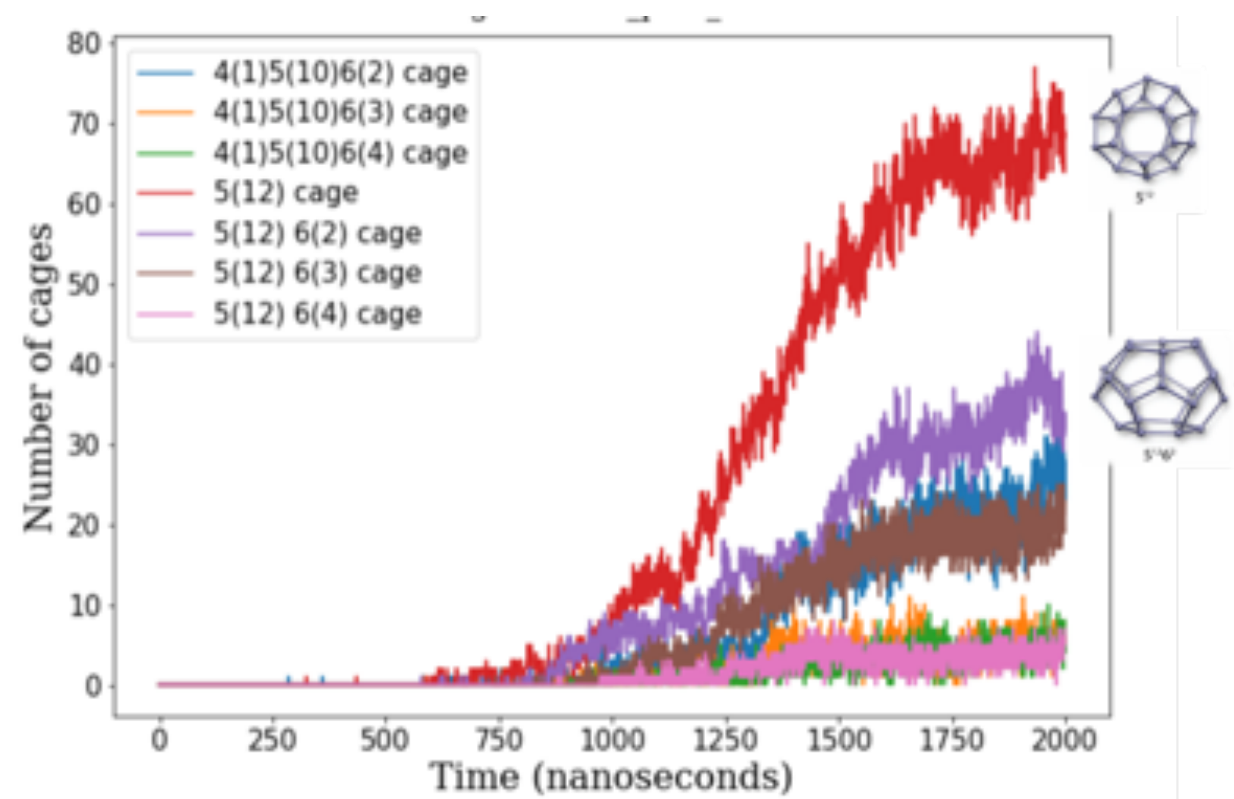
Walsh, M. R Science 2009, 326, 1095–1098

Brute force MD at 250 K forms amorphous solid



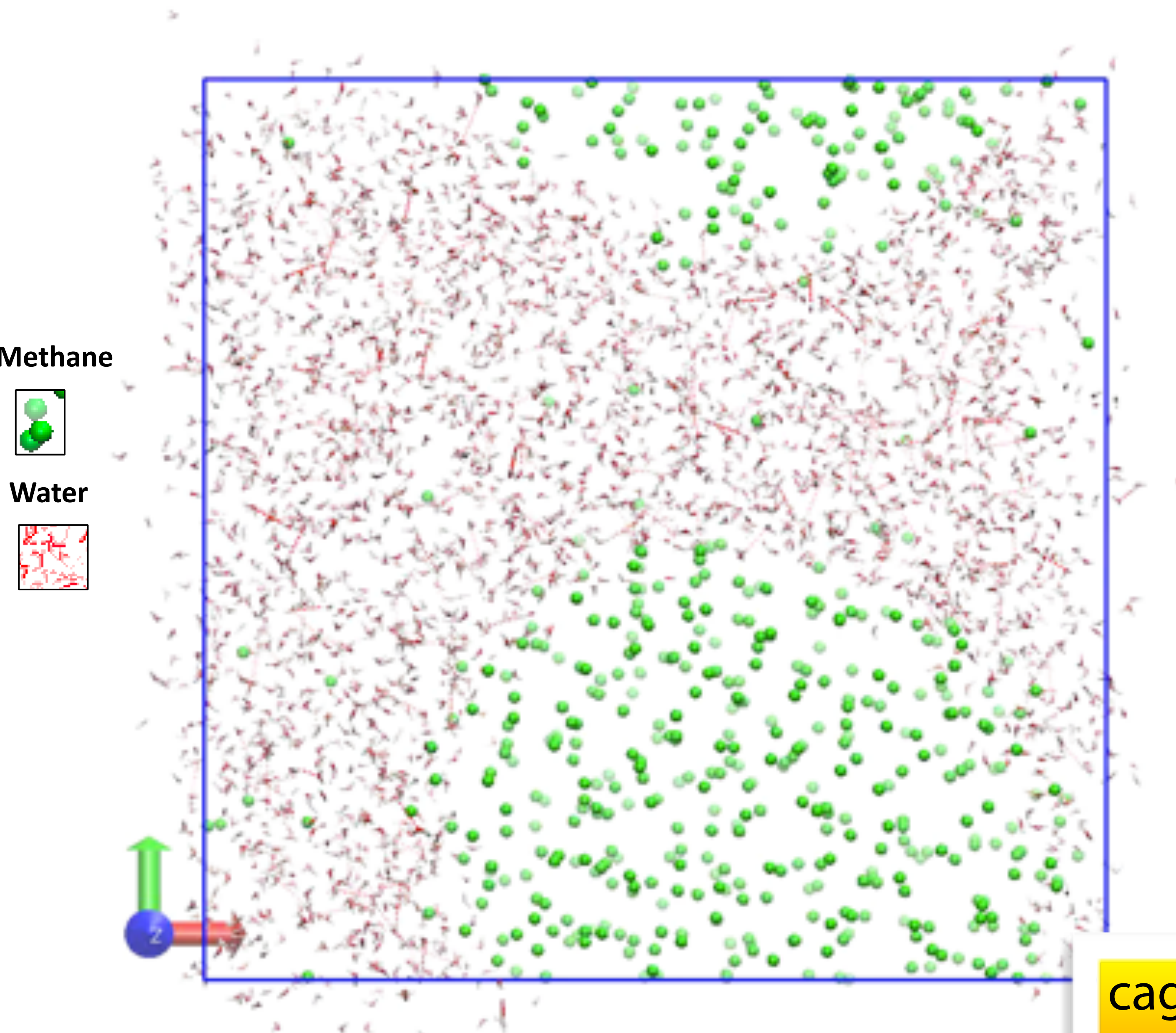
2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K, 2 µs

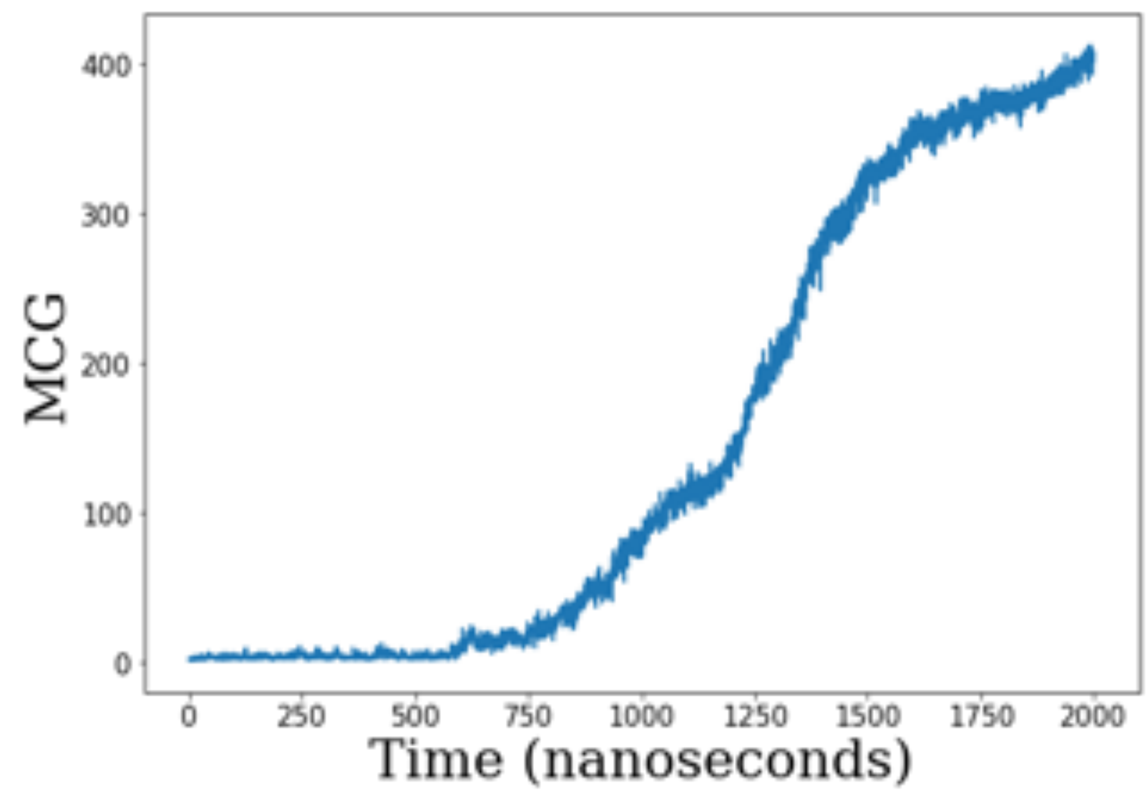


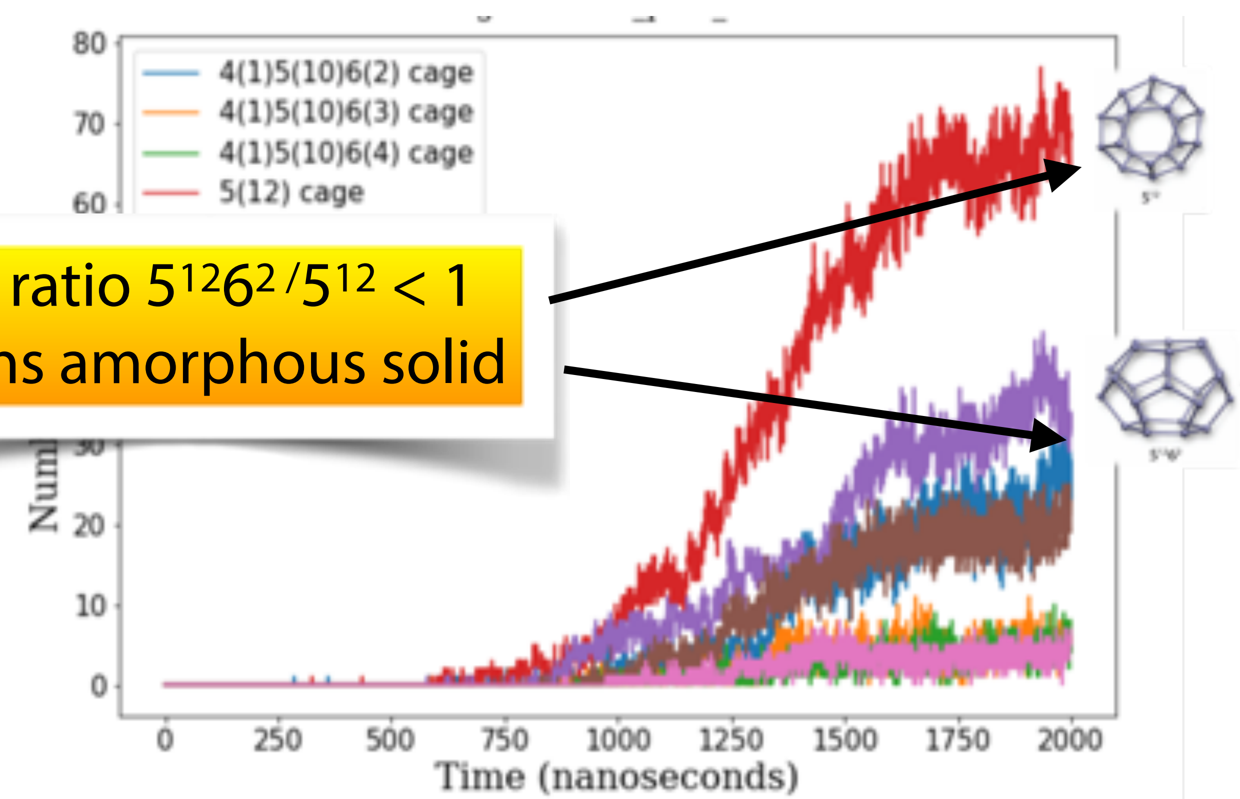


Walsh, M. R Science 2009, 326, 1095–1098

Brute force MD at 250 K forms amorphous solid





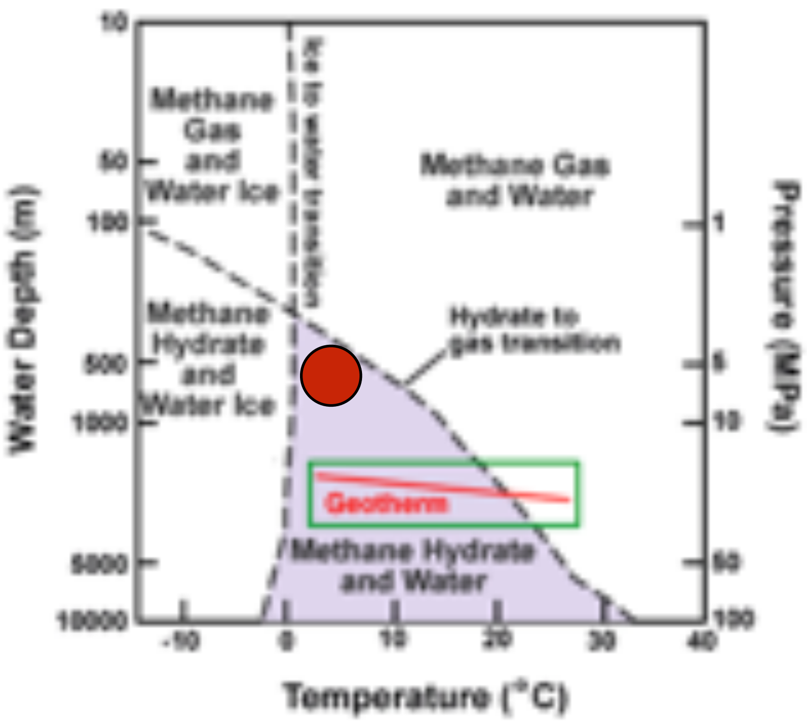


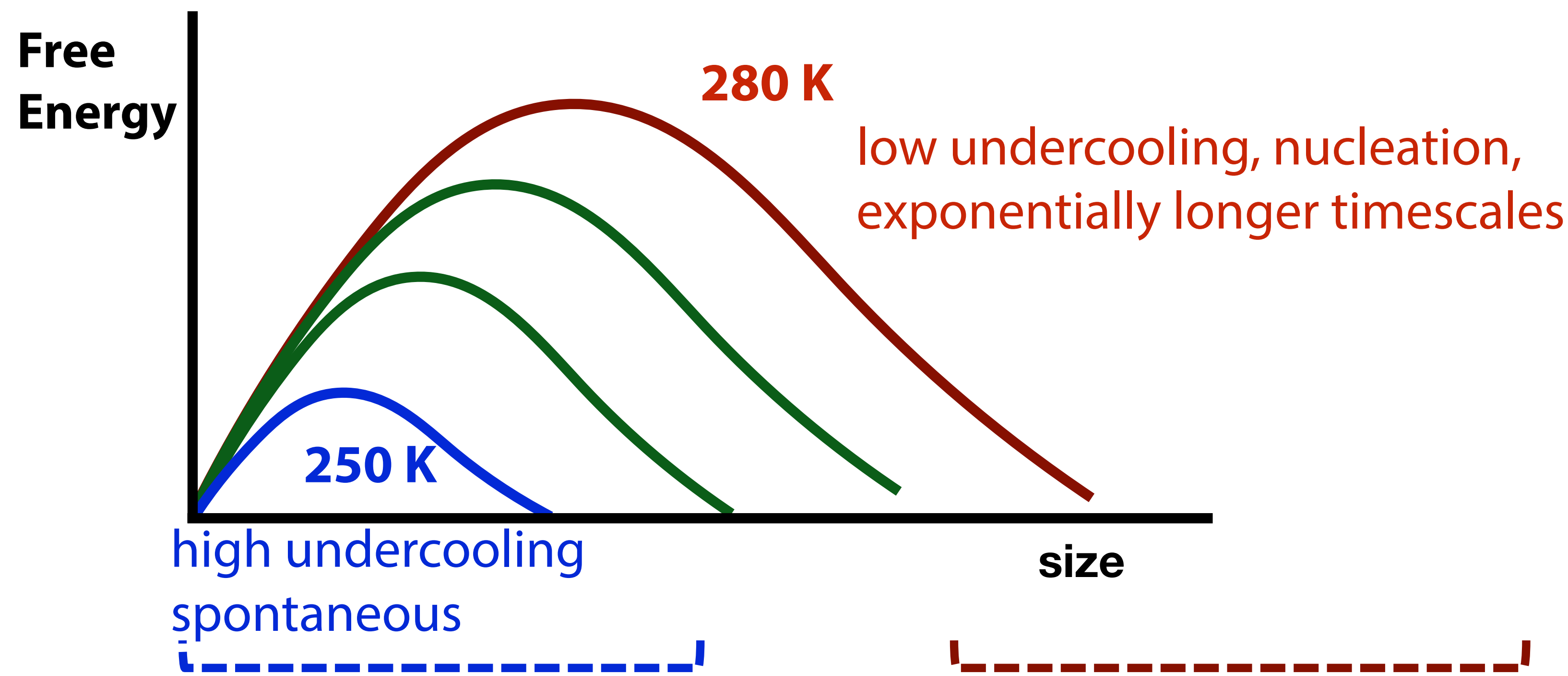
cage ratio $5^{12}6^{2}/5^{12} < 1$ means amorphous solid

2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K, 2 μs

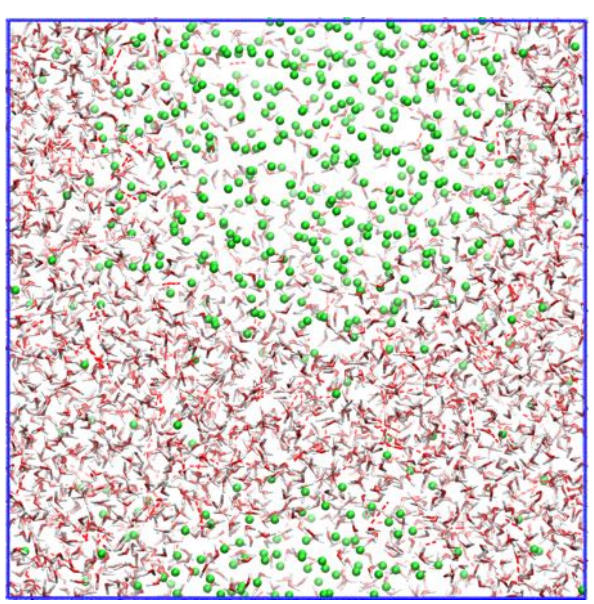
Walsh, M. R Science 2009, 326, 1095–1098

Solidification dependent on temperature

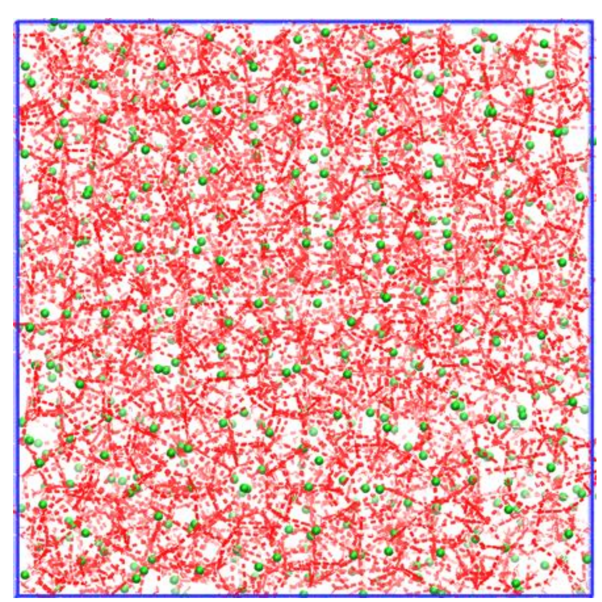




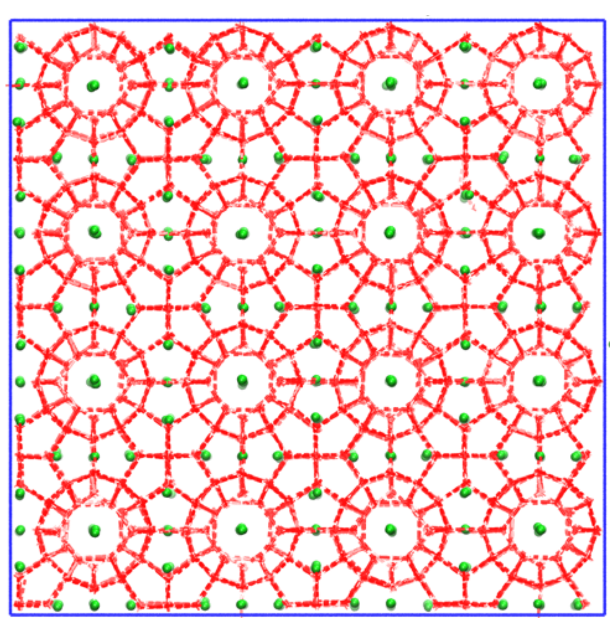
Liquid



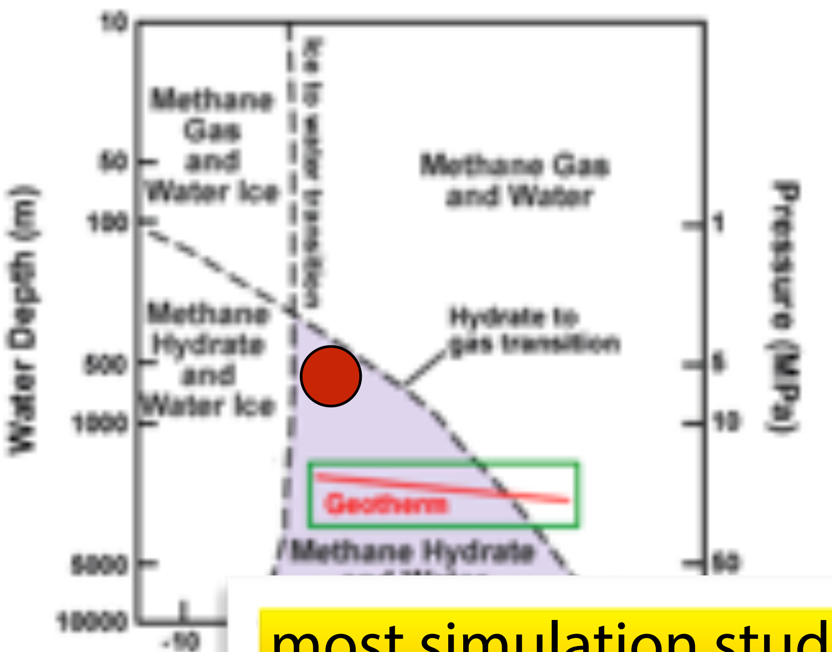
Amorphous Solid



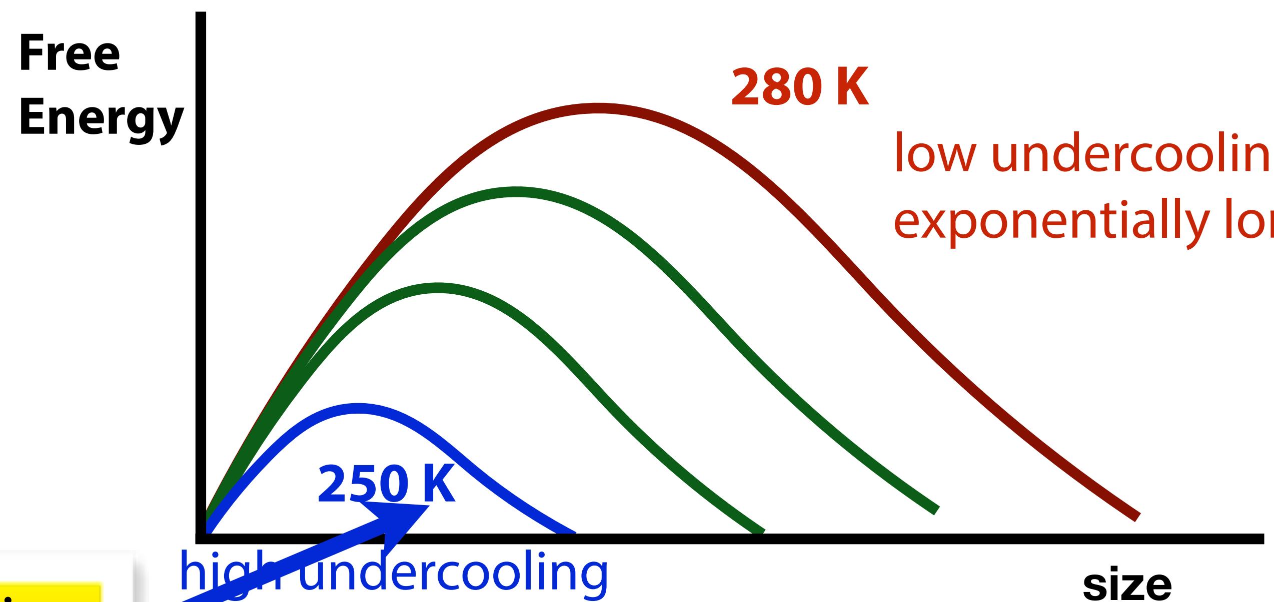
Crystalline Solid



Solidification dependent on temperature



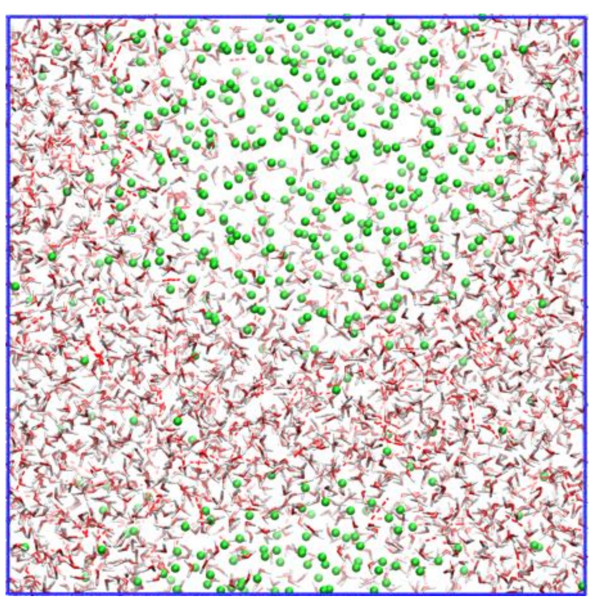
most simulation studies are over here



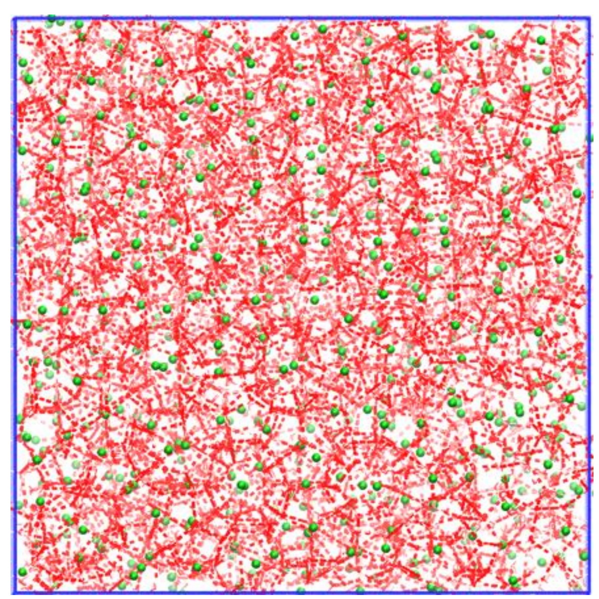
low undercooling, nucleation, exponentially longer timescales

high undercooling spontaneous

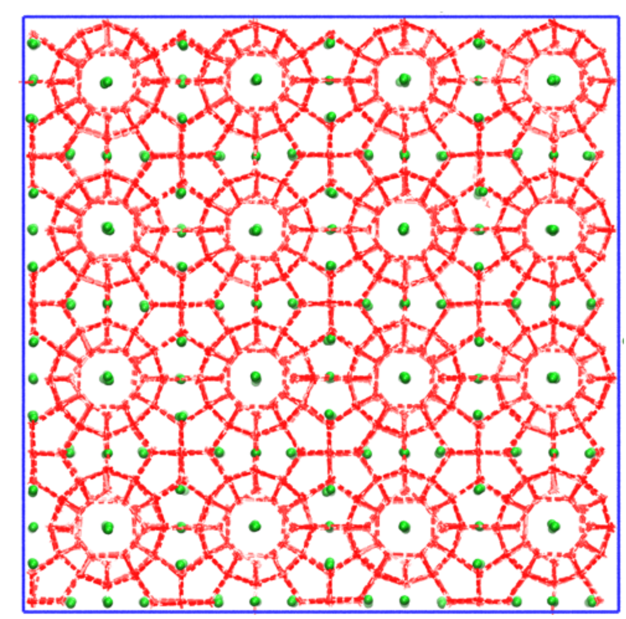
Liquid



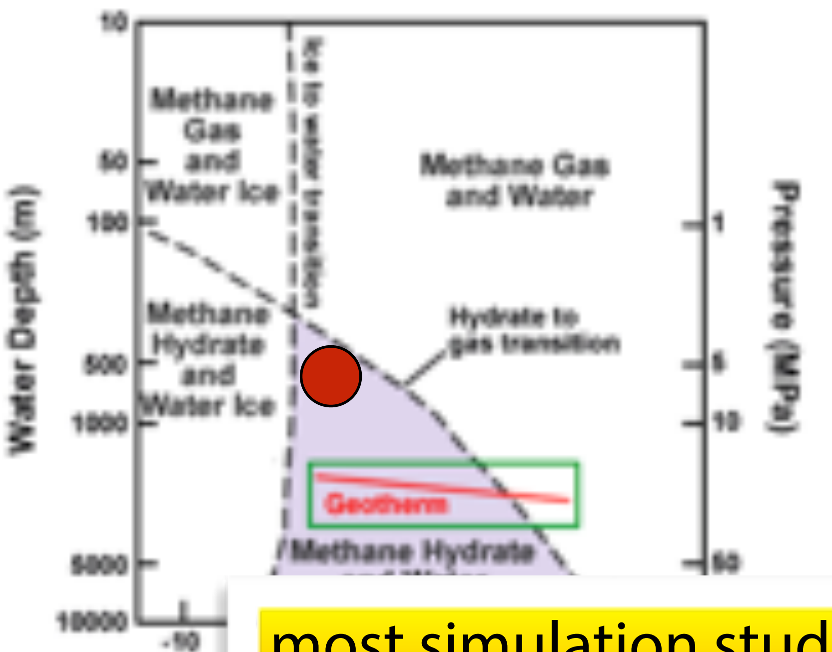
Amorphous Solid

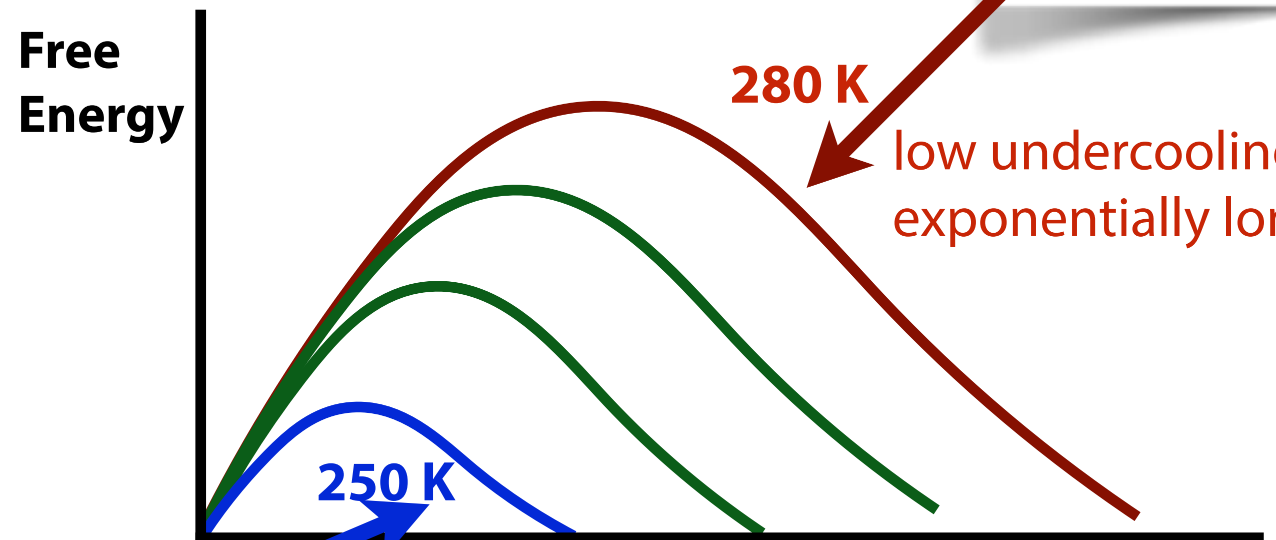


Crystalline Solid



Solidification dependent on temperature





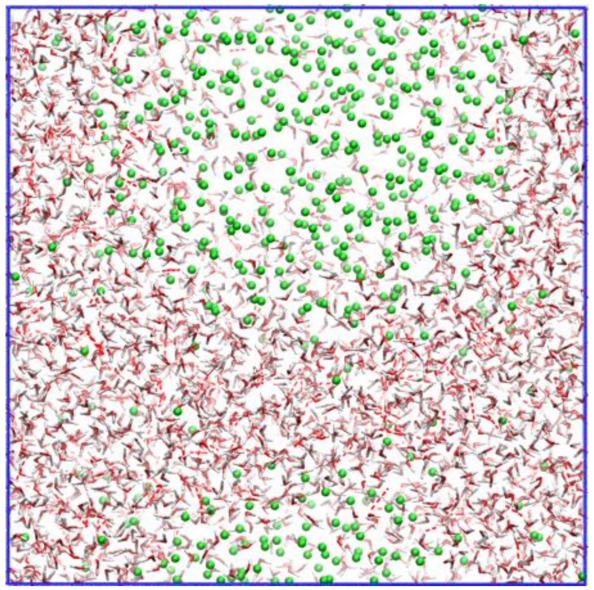
experimentally relevant
TPS required

low undercooling, nucleation,
exponentially longer timescales

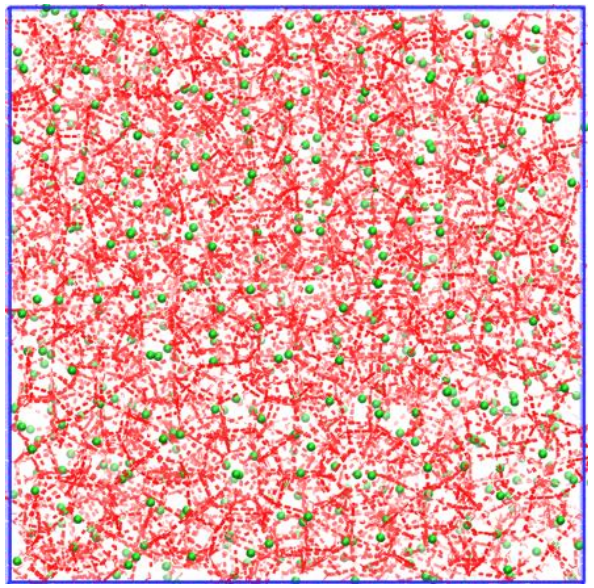
most simulation studies
are over here

high undercooling
spontaneous

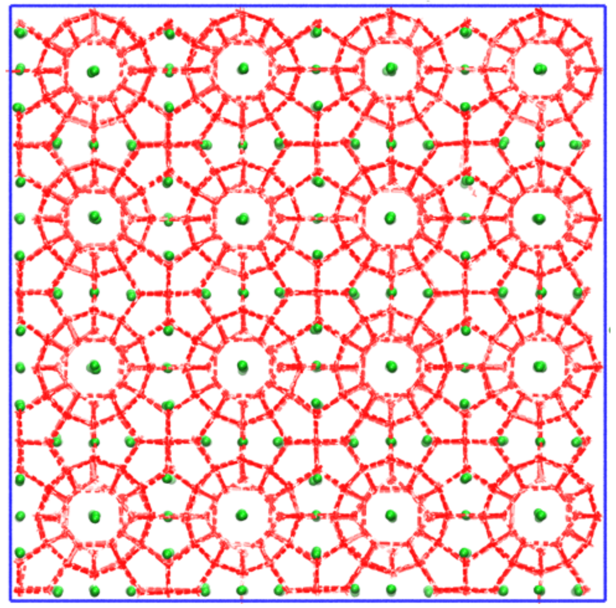
Liquid



Amorphous Solid

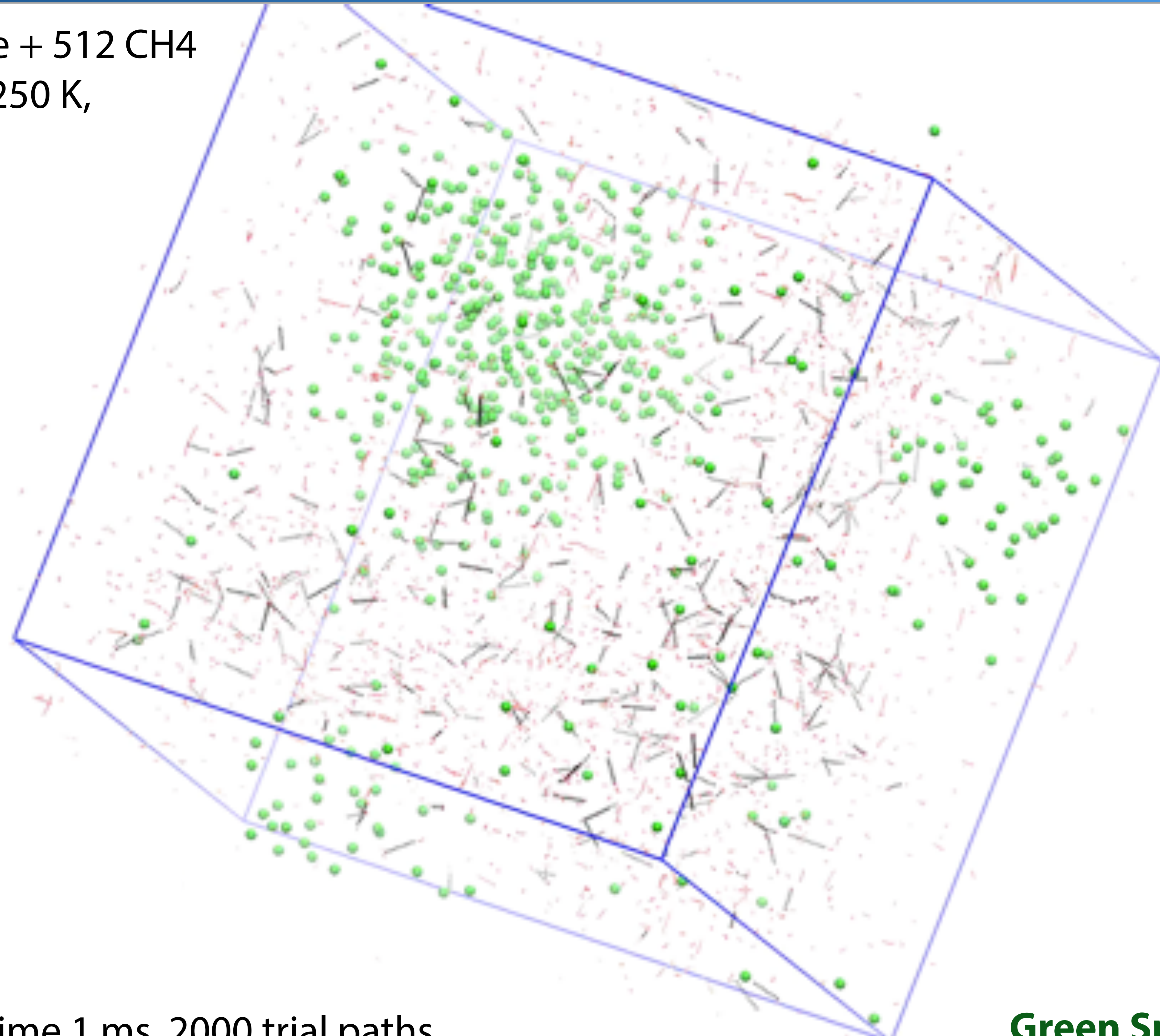


Crystalline Solid



Path sampling at 280 K

2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K,



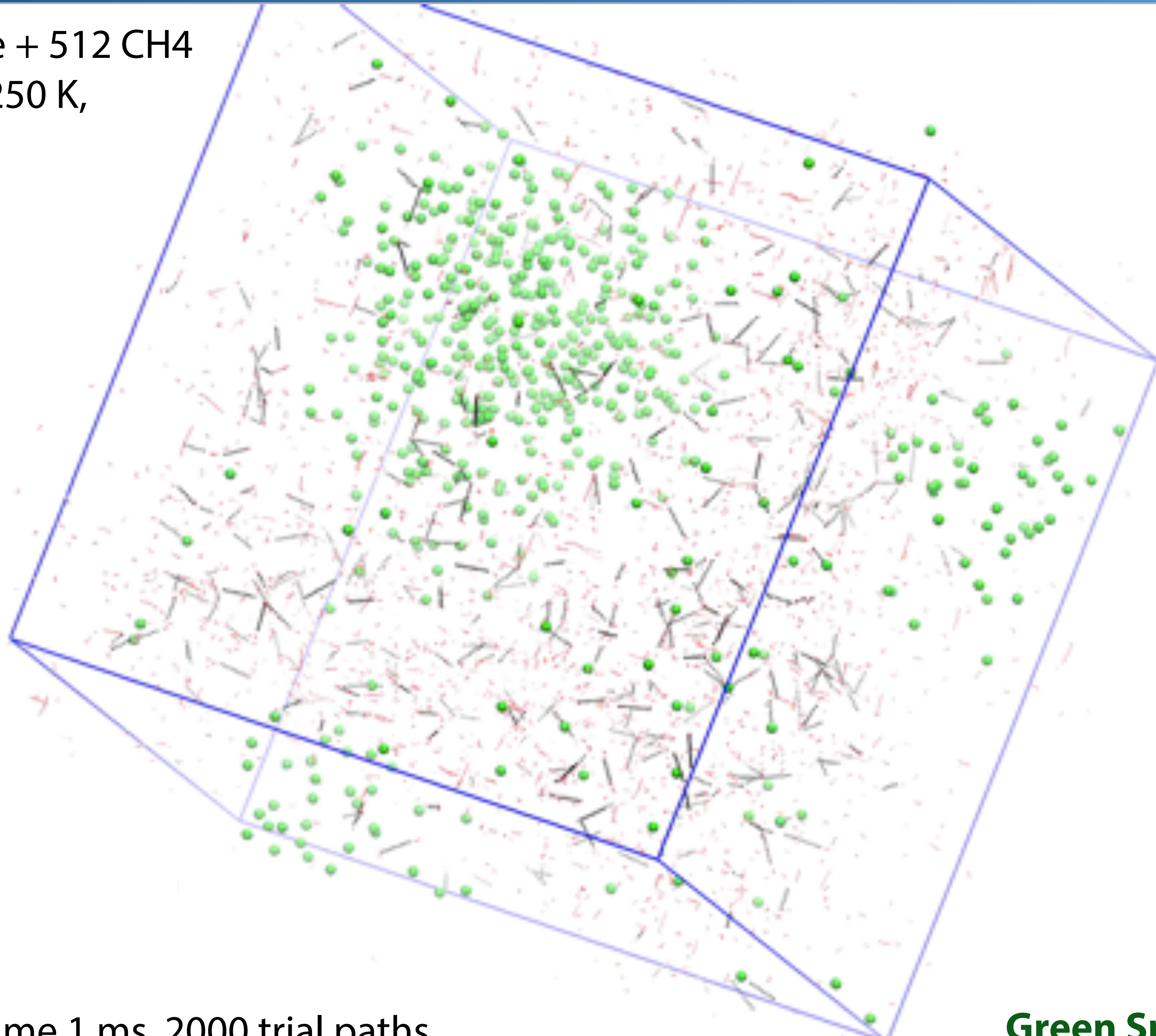
simulation time 1 ms, 2000 trial paths,
acceptance 33%, >200 decorrelated paths,
average path length 500 ns
induction time > 30 kyears

Green Spheres – Methane
Dotted Lines – Water
Hydrogen Bonds

Arjun, Berendsen, PGB, PNAS, 2019

Path sampling at 280 K

2944 TIP4P/ice + 512 CH4
NPT 500 atm 250 K,

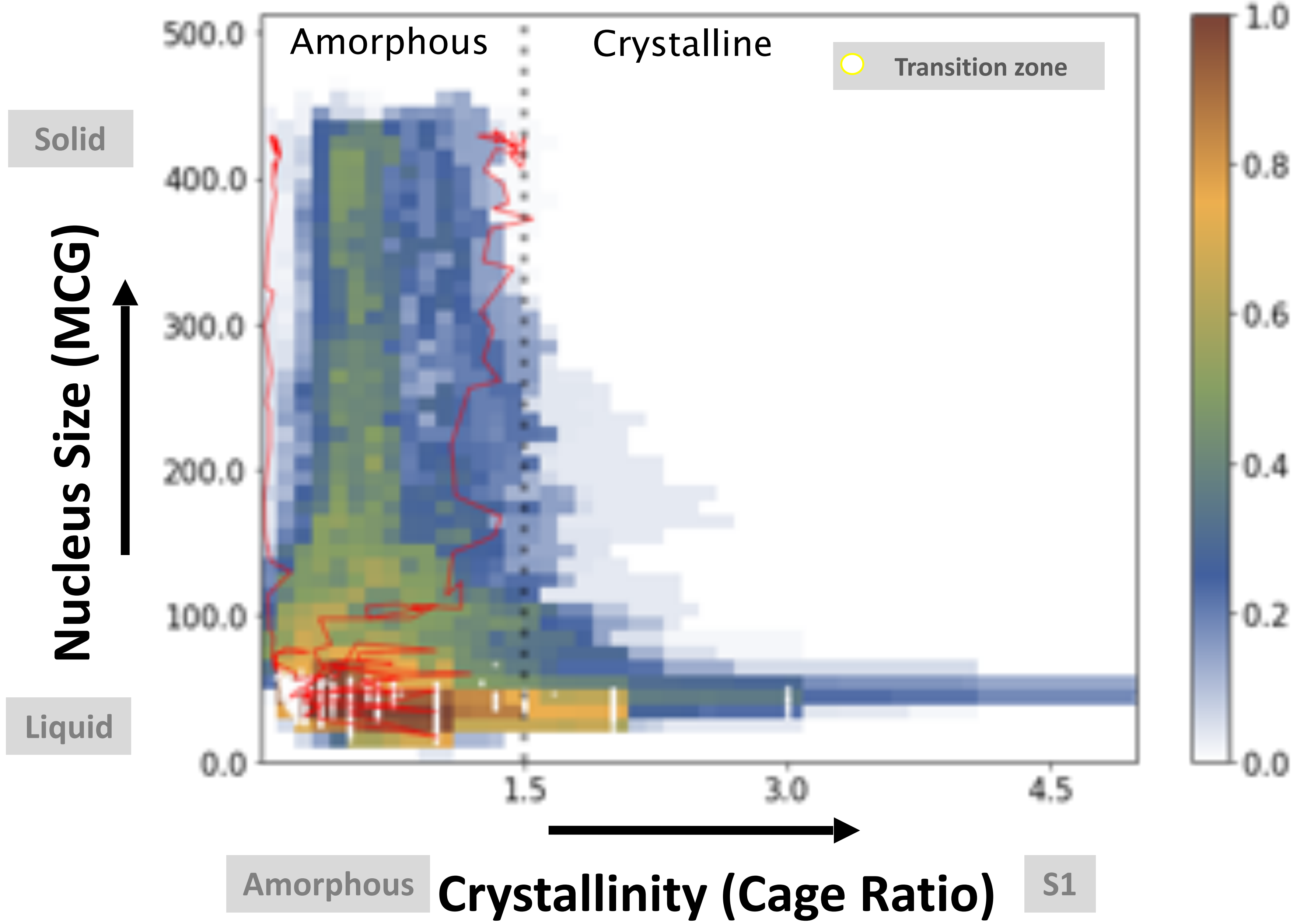


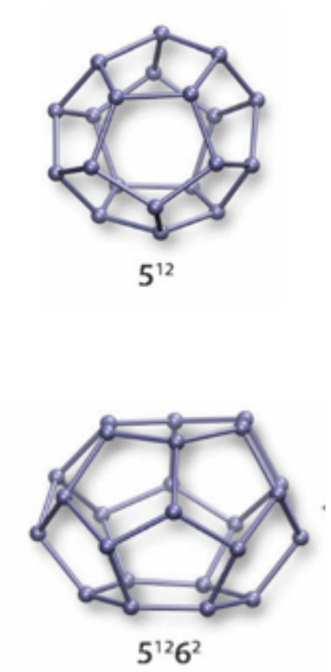
simulation time 1 ms, 2000 trial paths,
acceptance 33%, >200 decorrelated paths,
average path length 500 ns
induction time > 30 kyears

Green Spheres – Methane
Dotted Lines – Water
Hydrogen Bonds

Arjun, Berendsen, PGB, PNAS, 2019

path density T=270 K

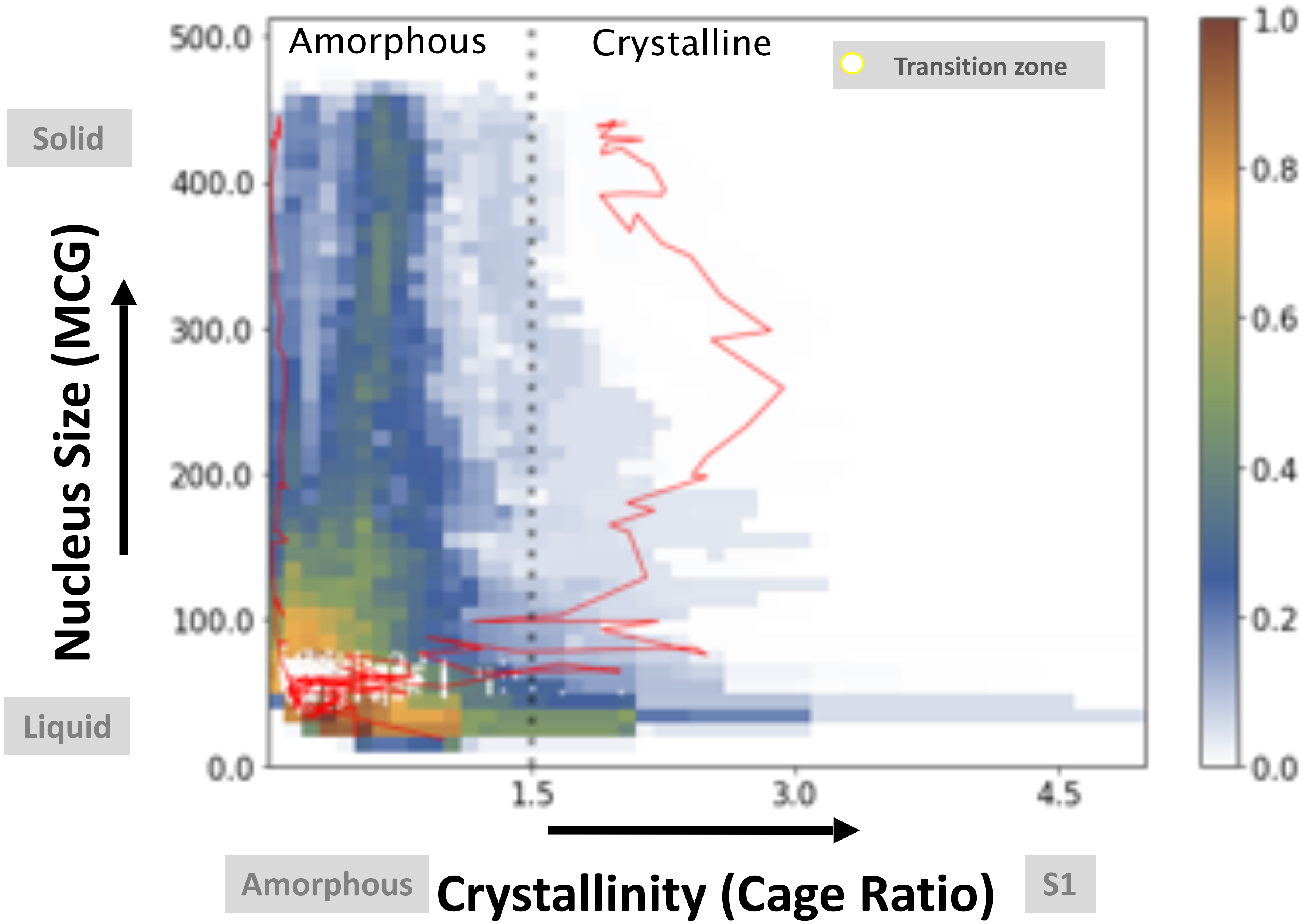


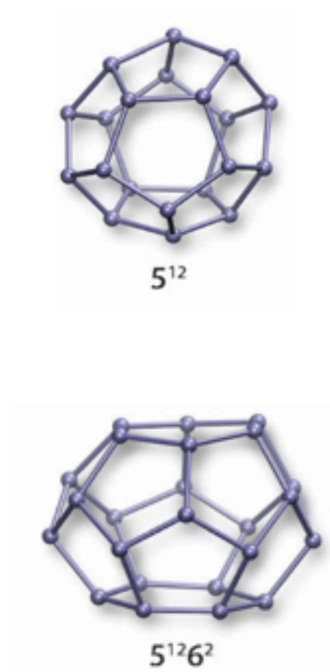


Pure S1 Ratio
5(12)6(2)/ 5(12) = 3

Arjun, Berendsen, PGB, PNAS, 2019

path density T=275 K

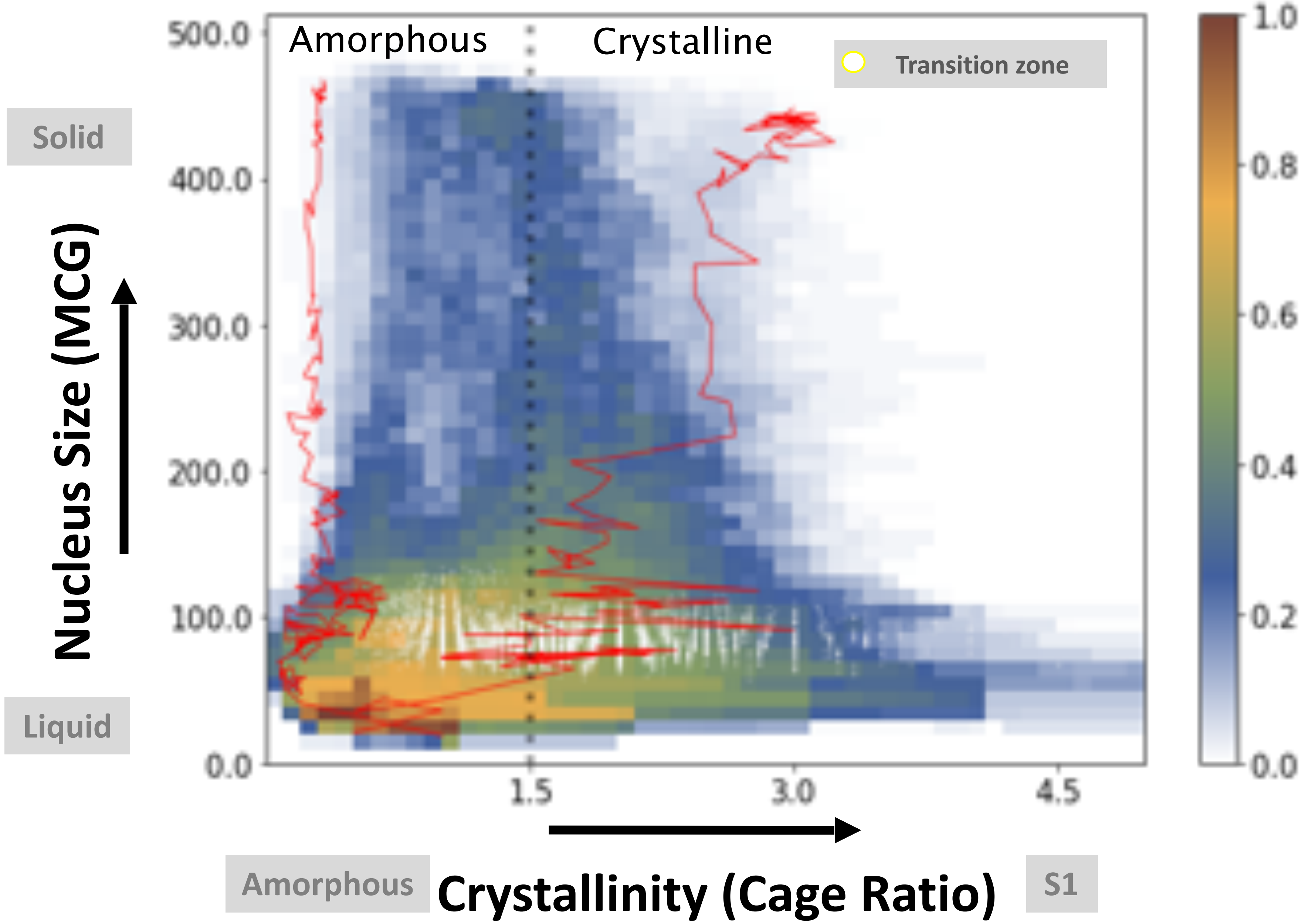


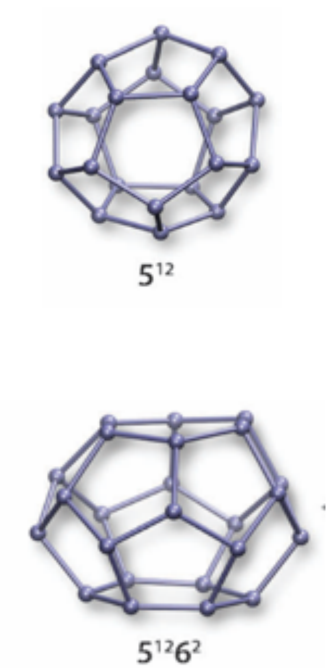


Pure S1 Ratio
5(12)6(2)/ 5(12) = 3

Arjun, Berendsen, PGB, PNAS, 2019

path density T=280 K

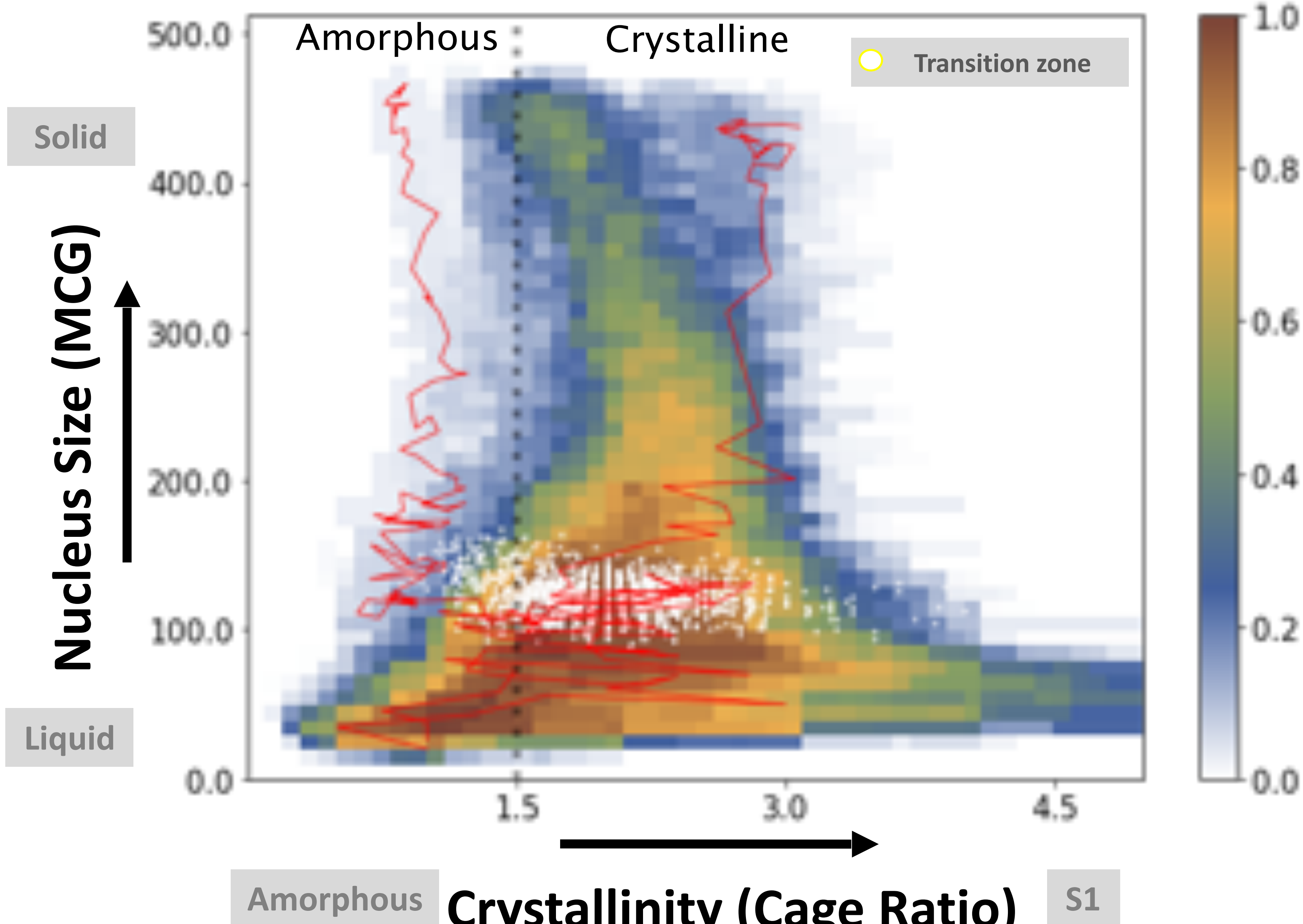


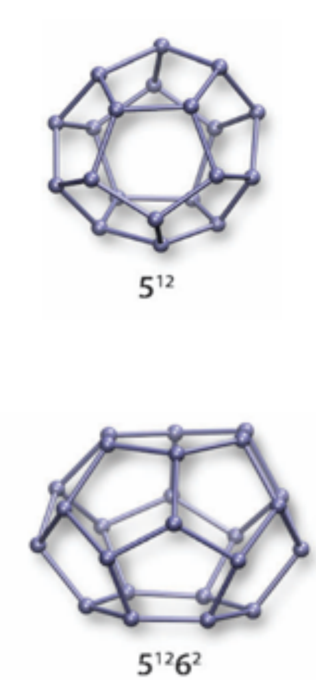


Pure S1 Ratio
5(12)6(2)/ 5(12) = 3

Arjun, Berendsen, PGB, PNAS, 2019

path density T=285 K



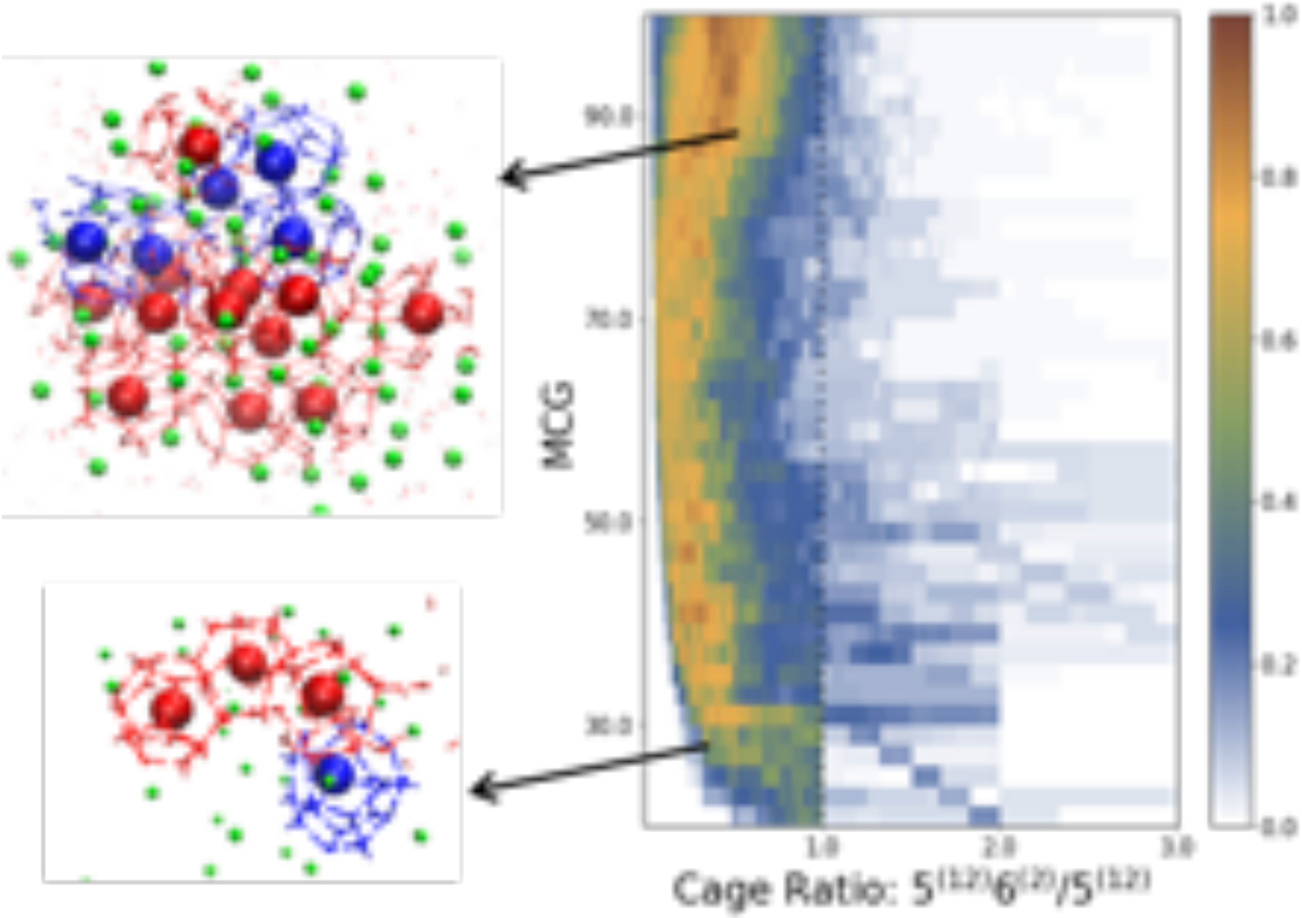


Pure S1 Ratio
5(12)6(2)/ 5(12) = 3

Arjun, Berendsen, PGB, PNAS, 2019

At 280 K two channels compete

Amorphous Pathways

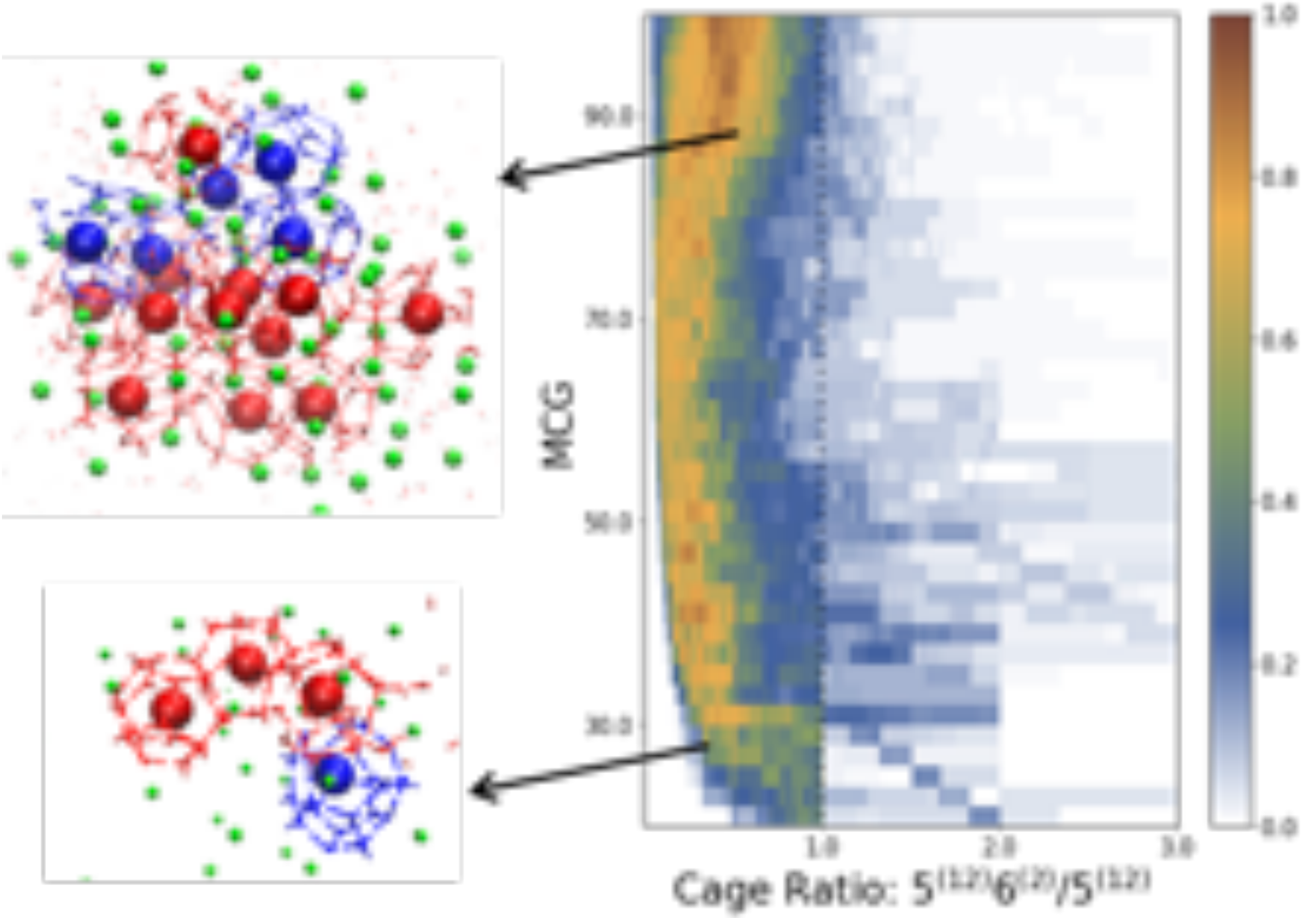


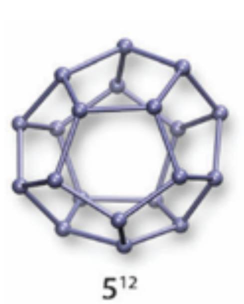
Red Cage

Arjun, Berendsen, PGB, PNAS, 2019

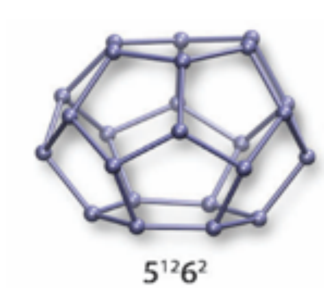
At 280 K two channels compete

Amorphous Pathways





Red Cage

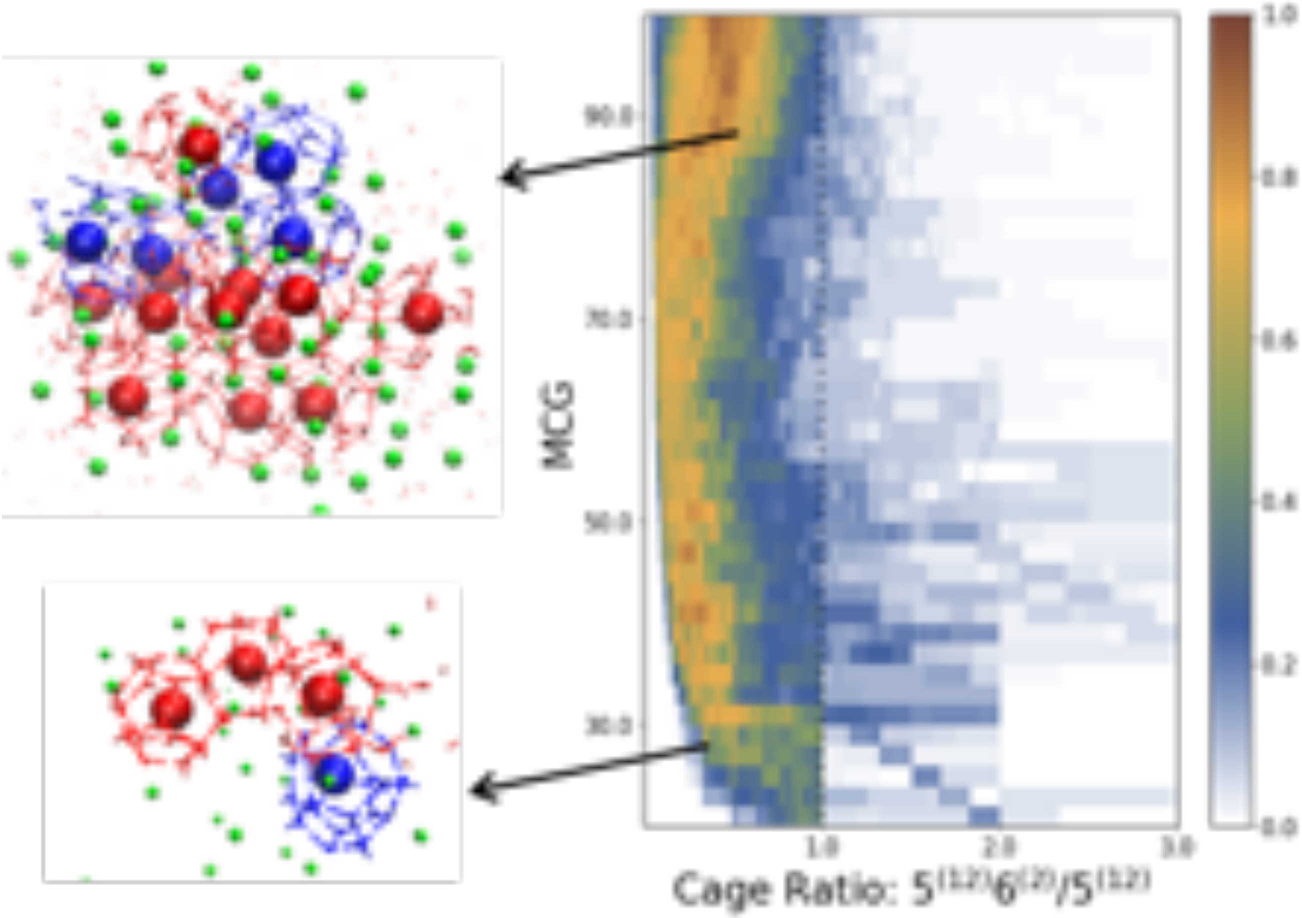


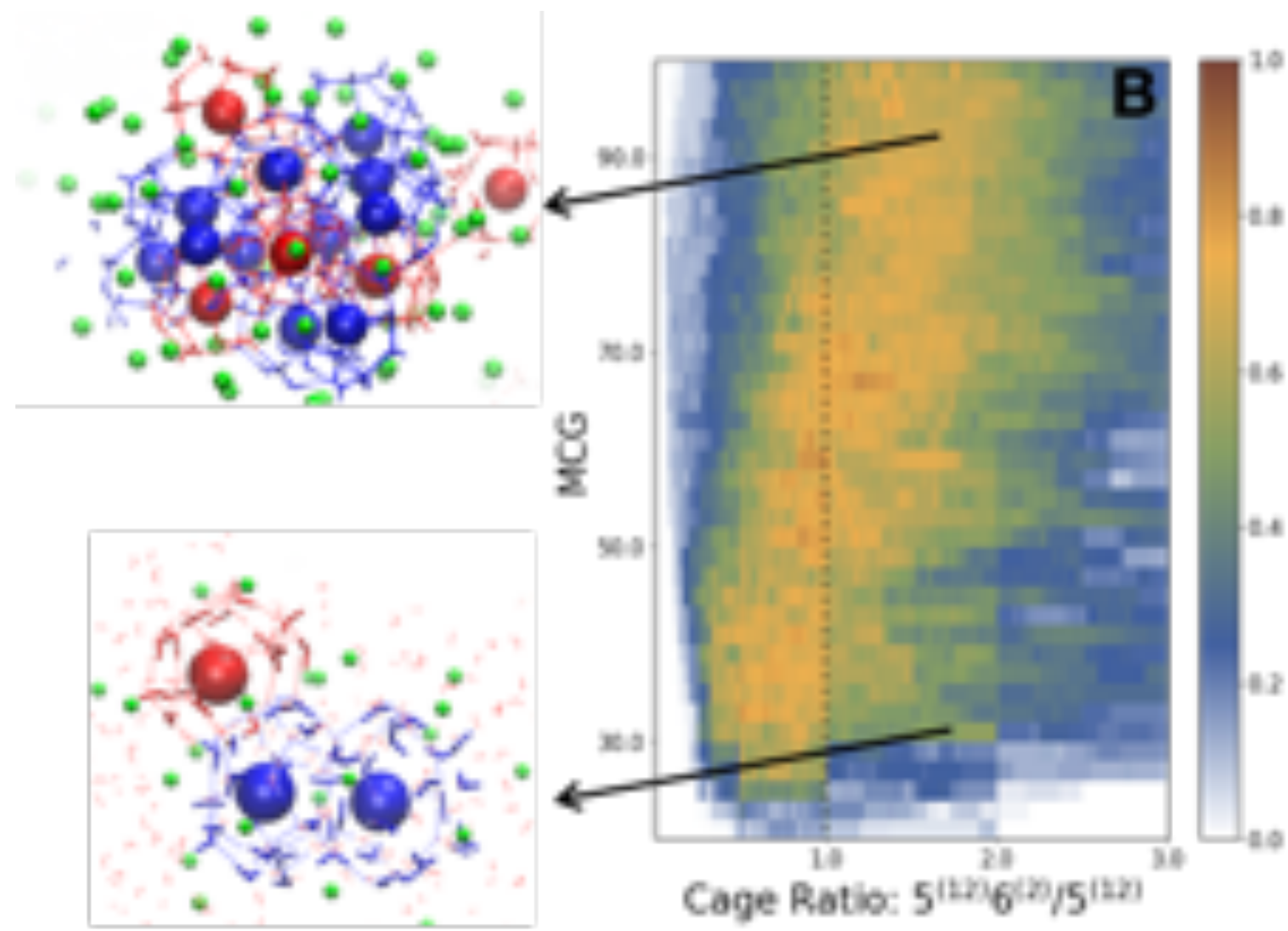
Arjun, Berendsen, PGB, PNAS, 2019

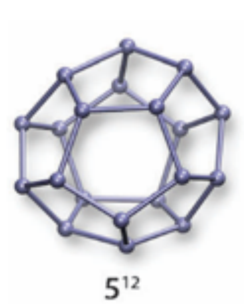
At 280 K two channels compete

Amorphous Pathways

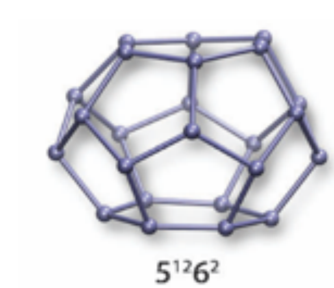
Crystalline Pathways







Red Cage

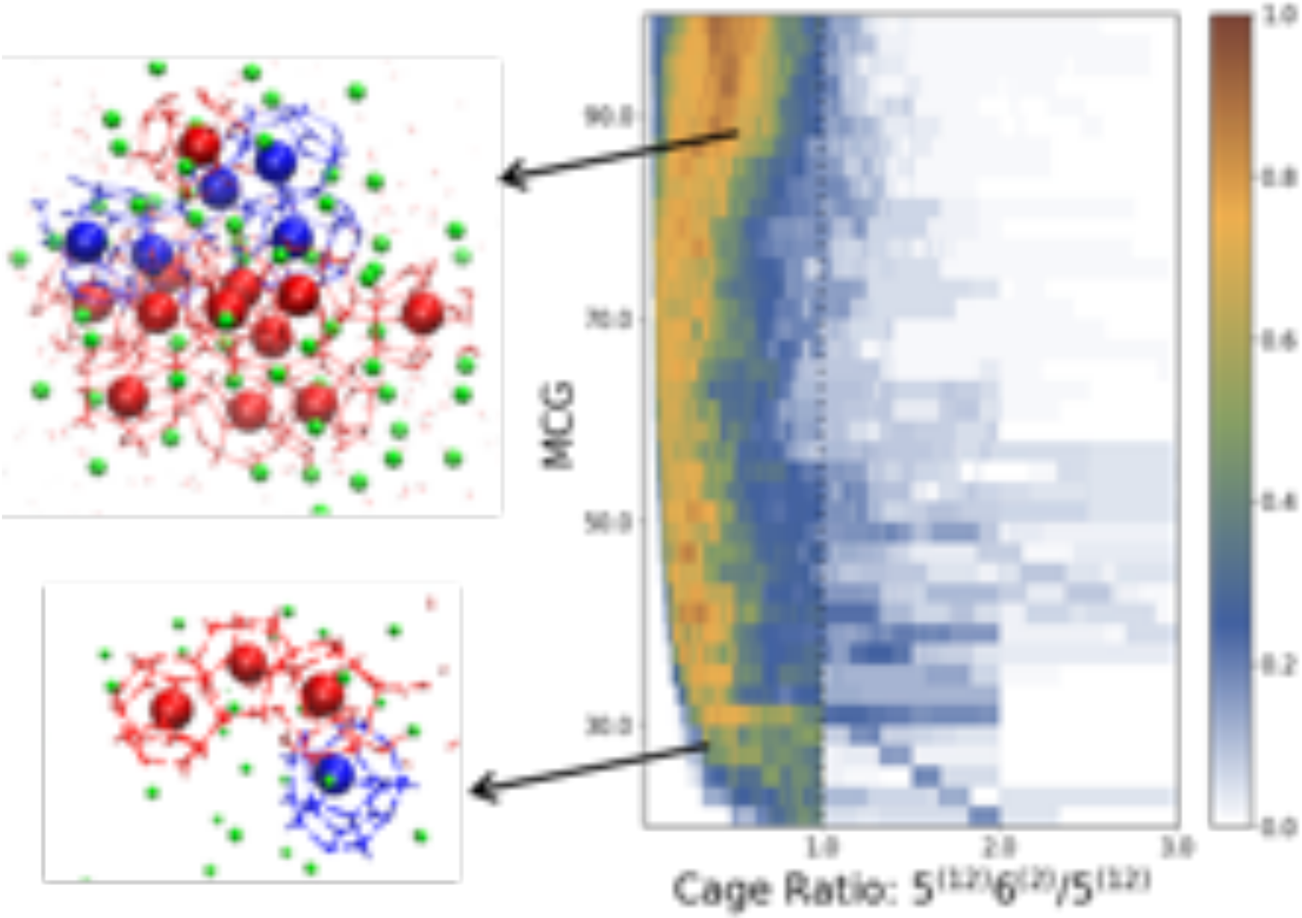


Blue Cage

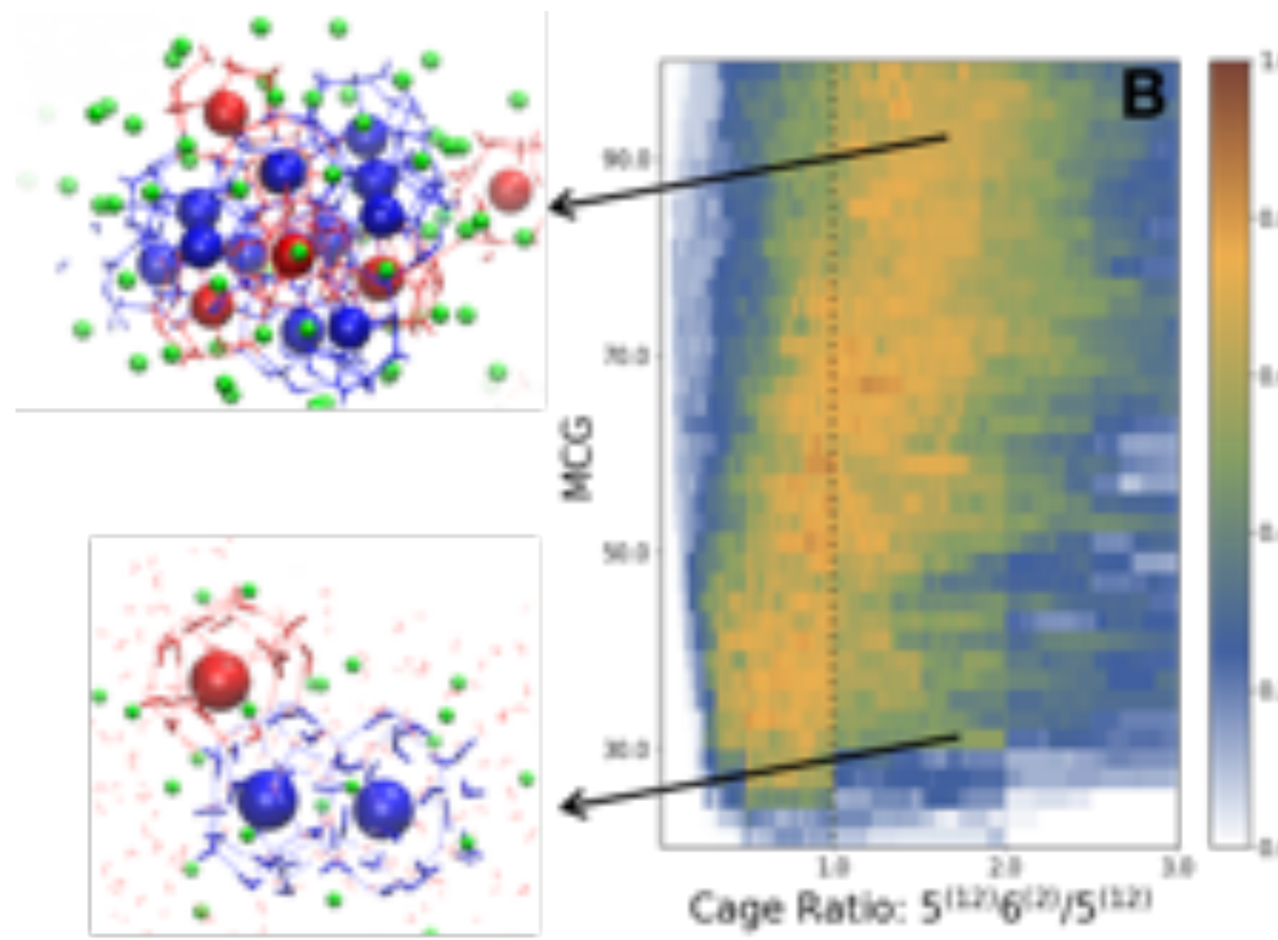
Arjun, Berendsen, PGB, PNAS, 2019

At 280 K two channels compete

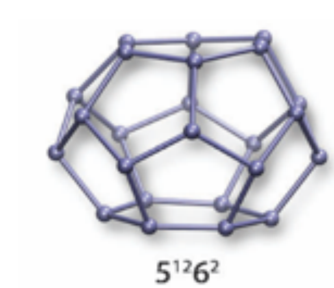
Amorphous Pathways



Crystalline Pathways



Red Cage



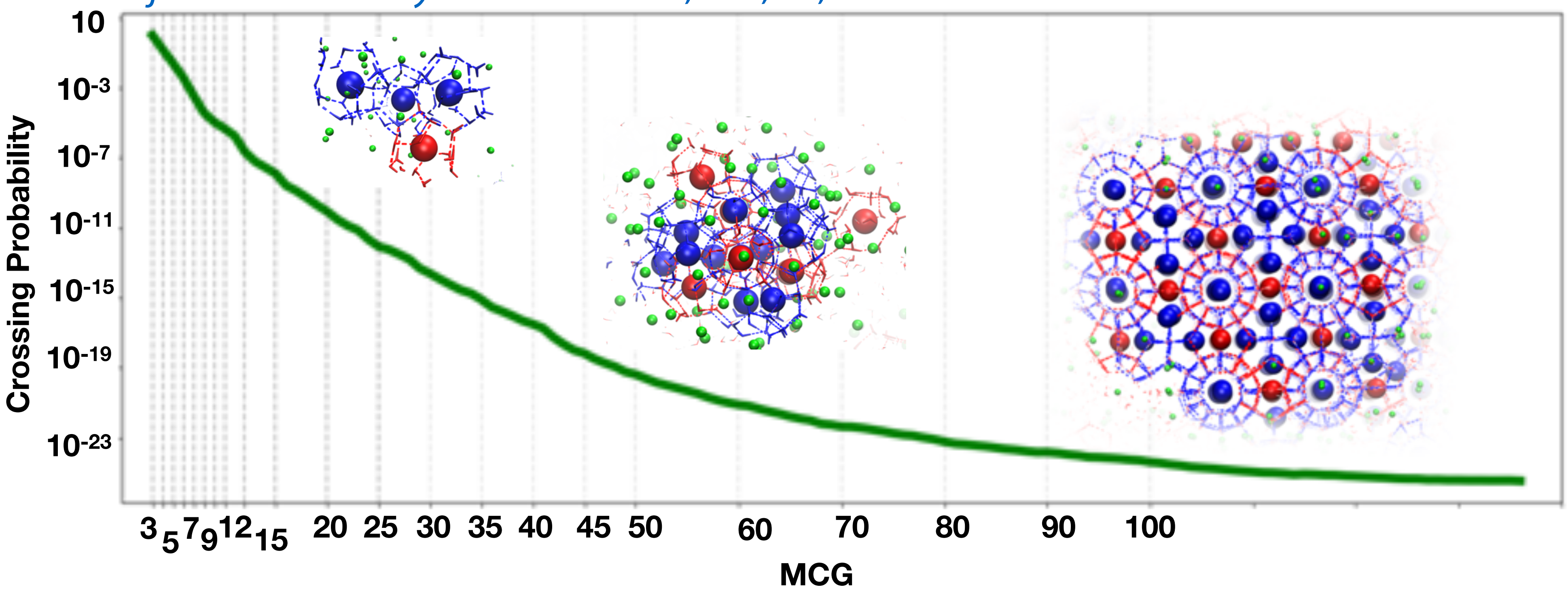
Blue Cage

does not follow Ostwald step rule: metastable phase avoided

Arjun, Berendsen, PGB, PNAS, 2019

Nucleation rate at 280K

Arjun and PGB . Phys. Chem. B 2020, 124, 37, 8099

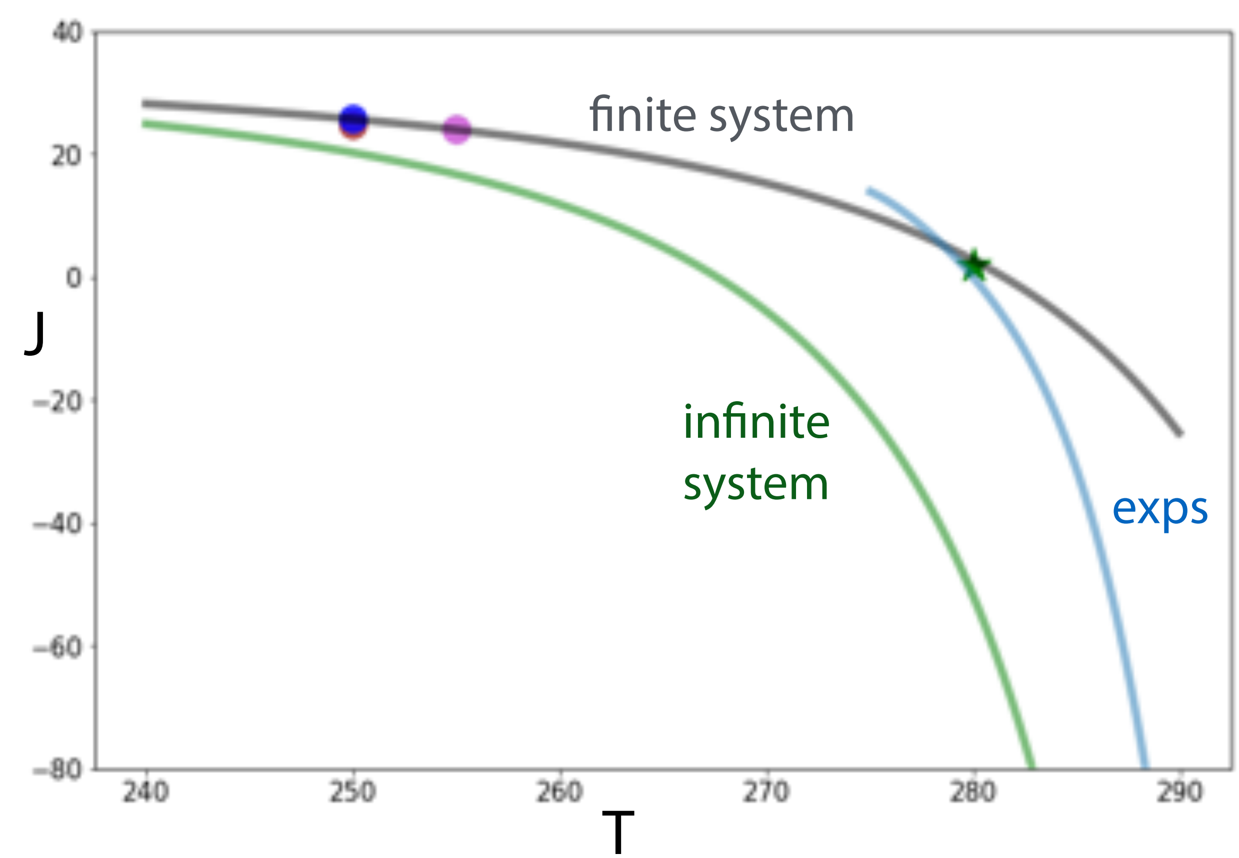


$J = \varphi\, P(\lambda)$

$\varphi = 2.1\ ns^{-1}$

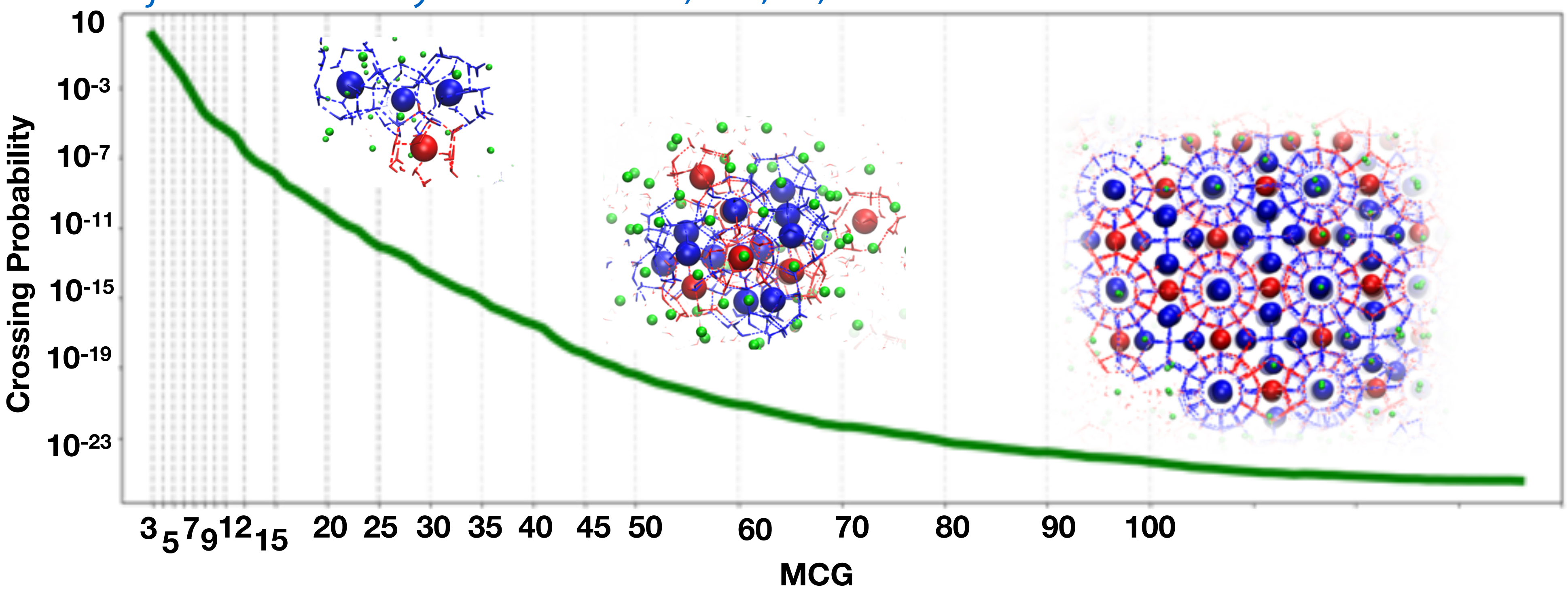
T = 280 K P = 500 atm

J= 500 nuclei $cm^{-3}\ s^{-1}$



Nucleation rate at 280K

Arjun and PGB . Phys. Chem. B 2020, 124, 37, 8099



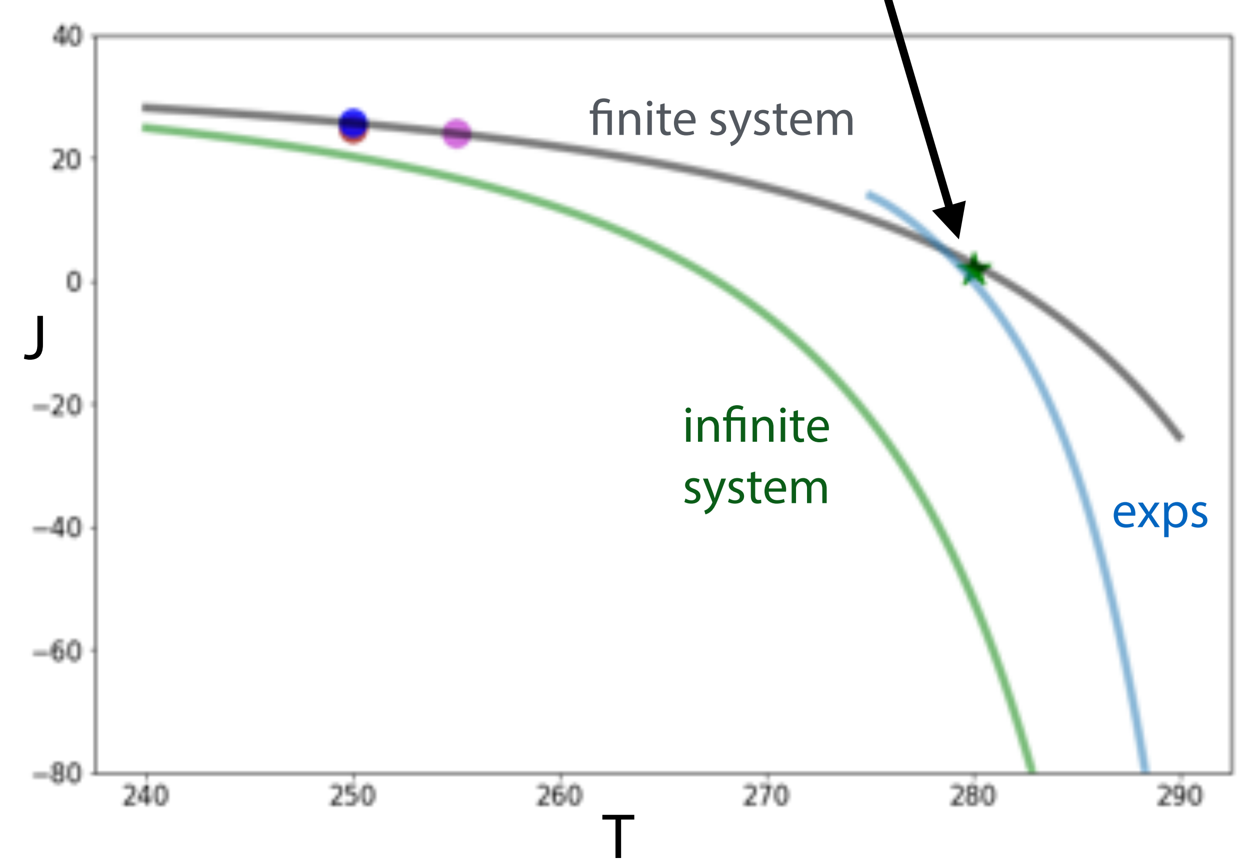
$J = \varphi\, P(\lambda)$

$\varphi = 2.1\ ns^{-1}$

T = 280 K P = 500 atm

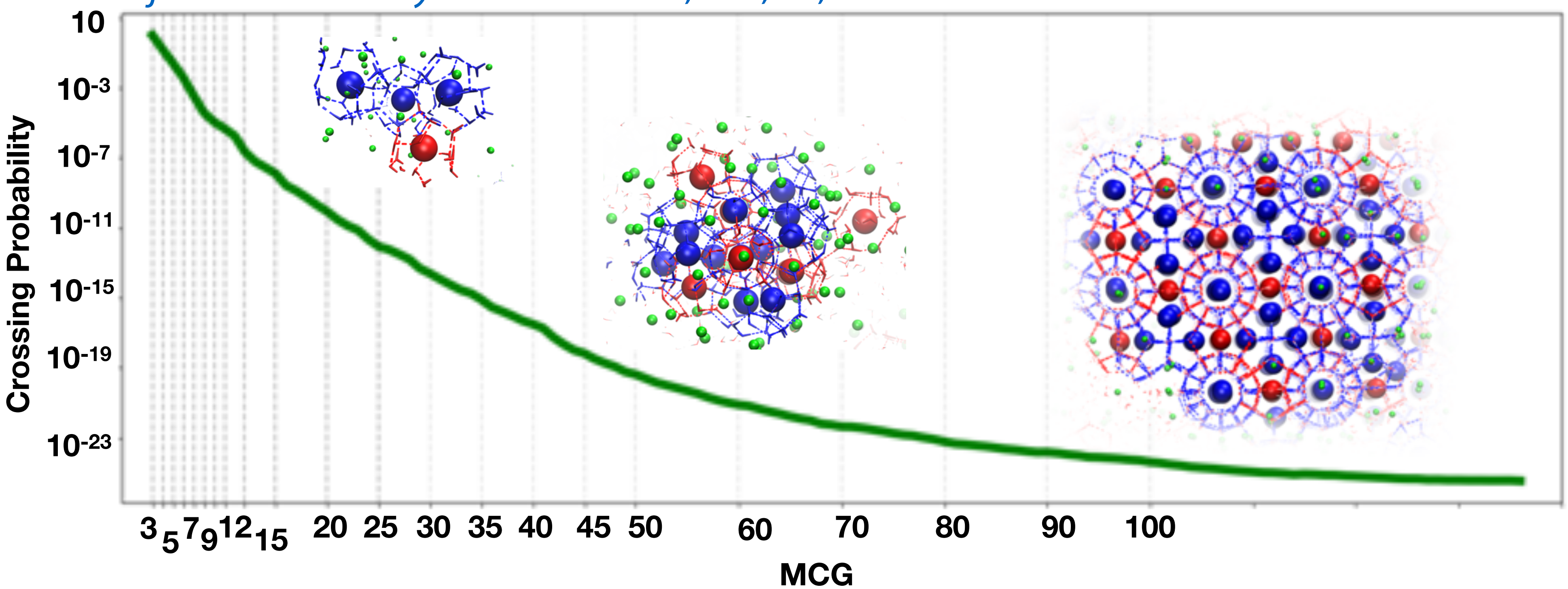
J= 500 nuclei cm^{-3} s^{-1}

rate in apparent agreement with experiments
(Toutham et al, Molecules 24, 1055 (2019))



Nucleation rate at 280K

Arjun and PGB . Phys. Chem. B 2020, 124, 37, 8099



$J = \varphi\, P(\lambda)$

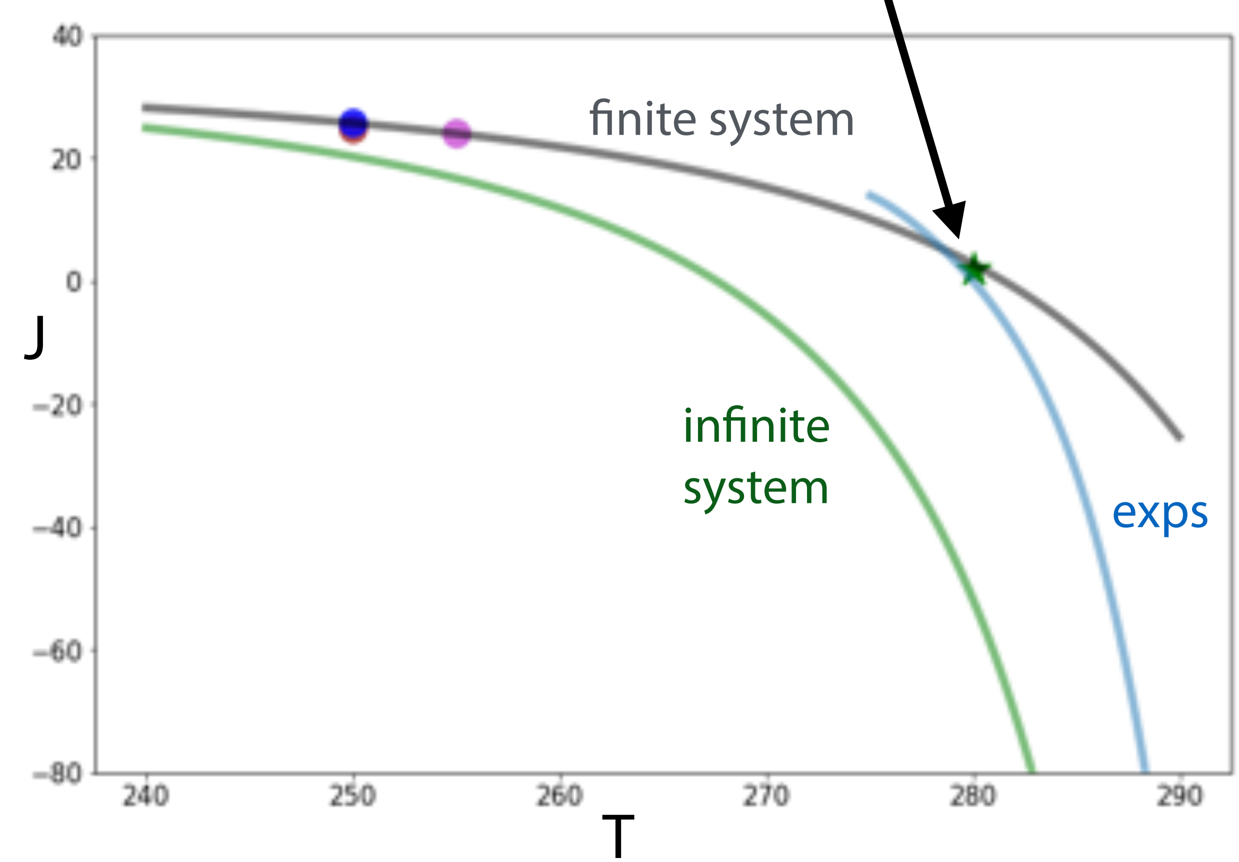
$\varphi = 2.1\ ns^{-1}$

T = 280 K P = 500 atm

J= 500 nuclei $cm^{-3}\ s^{-1}$

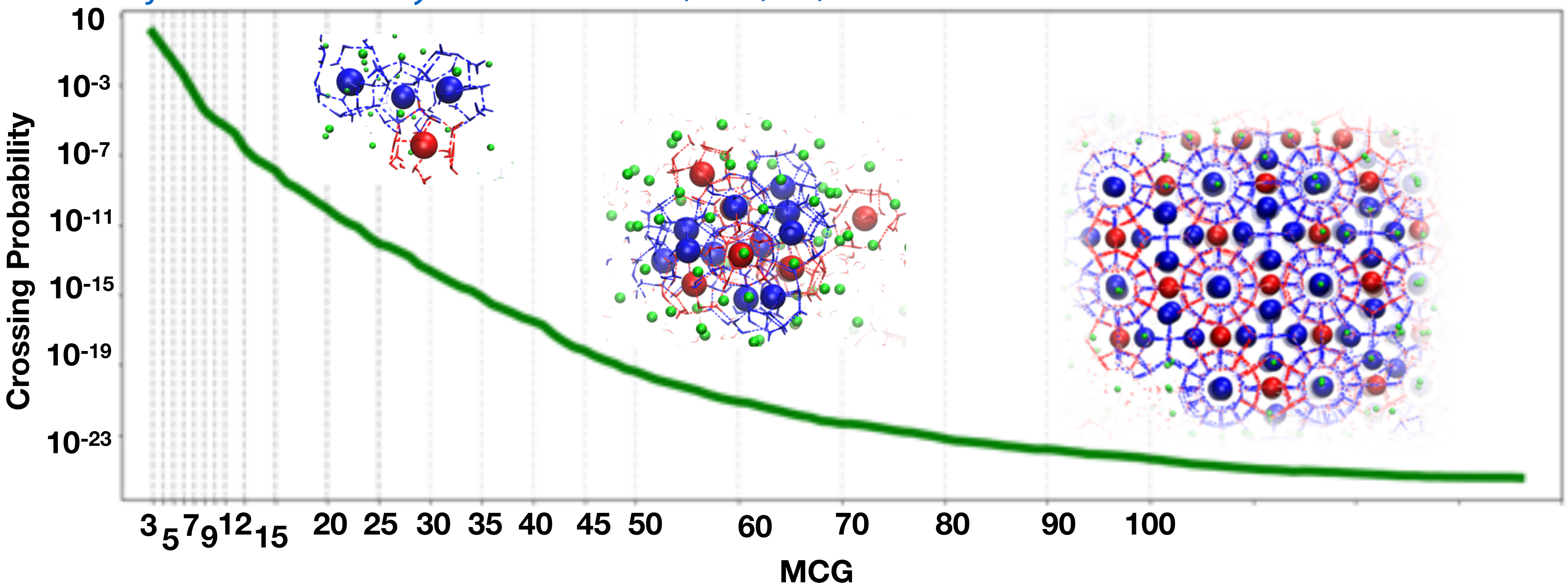
rate in apparent agreement with experiments *(Toutham et al, Molecules 24, 1055 (2019))*

fortuitous, rate after finite size correction much lower



Nucleation rate at 280K

Arjun and PGB . Phys. Chem. B 2020, 124, 37, 8099



$J = \varphi \, P(\lambda)$

$\varphi = 2.1 \text{ ns}^{-1}$

T = 280 K P = 500 atm

J= 500 nuclei cm^{-3} s^{-1}

rate in apparent agreement with experiments *(Toutham et al, Molecules 24, 1055 (2019))*

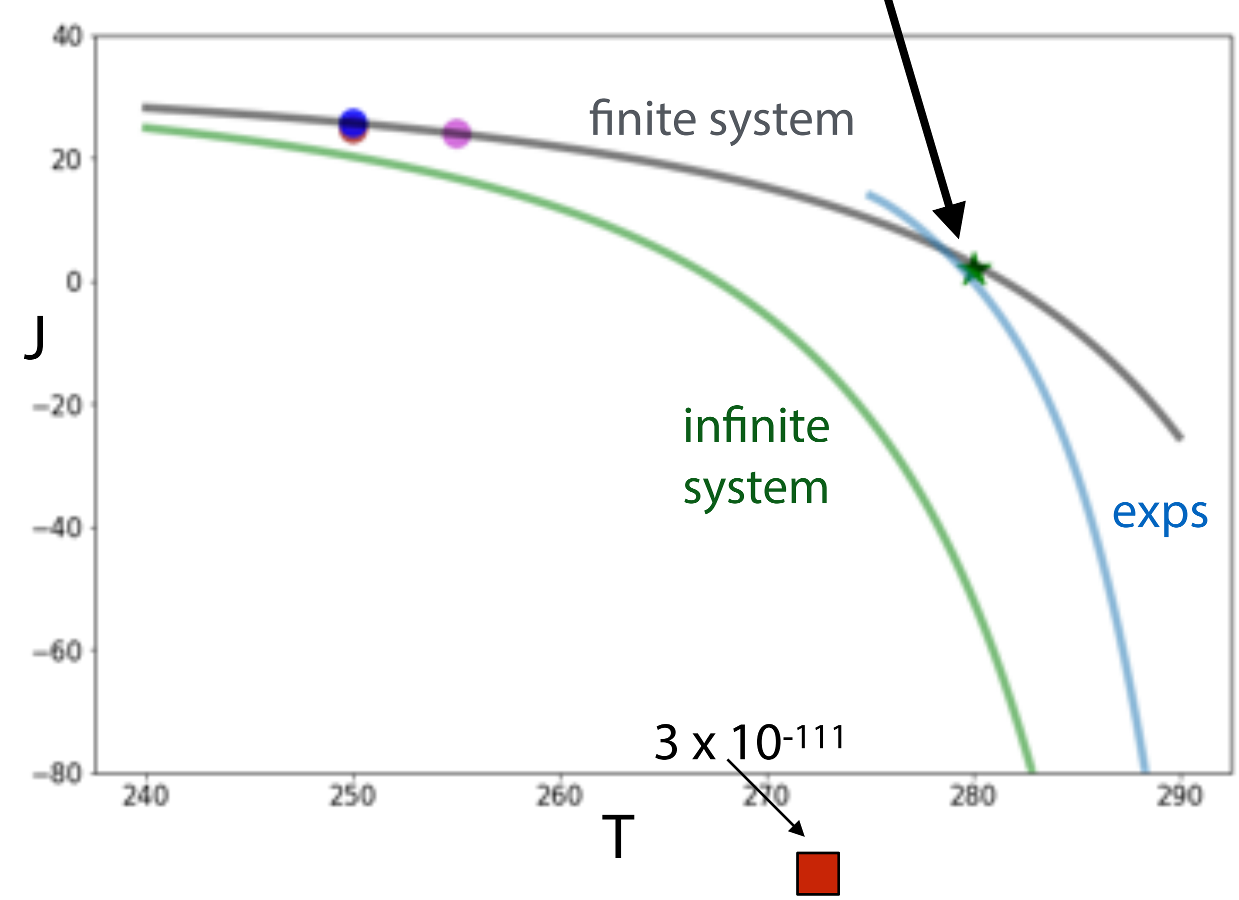
fortuitous, rate after finite size correction much lower

still rate much higher than previous predictions

J=3 x 10^{-111} nuclei cm^{-3} s^{-1}

T = 273 K P = 900 atm

Knott et al J. Am. Chem. Soc. 134, 19544 (2012)



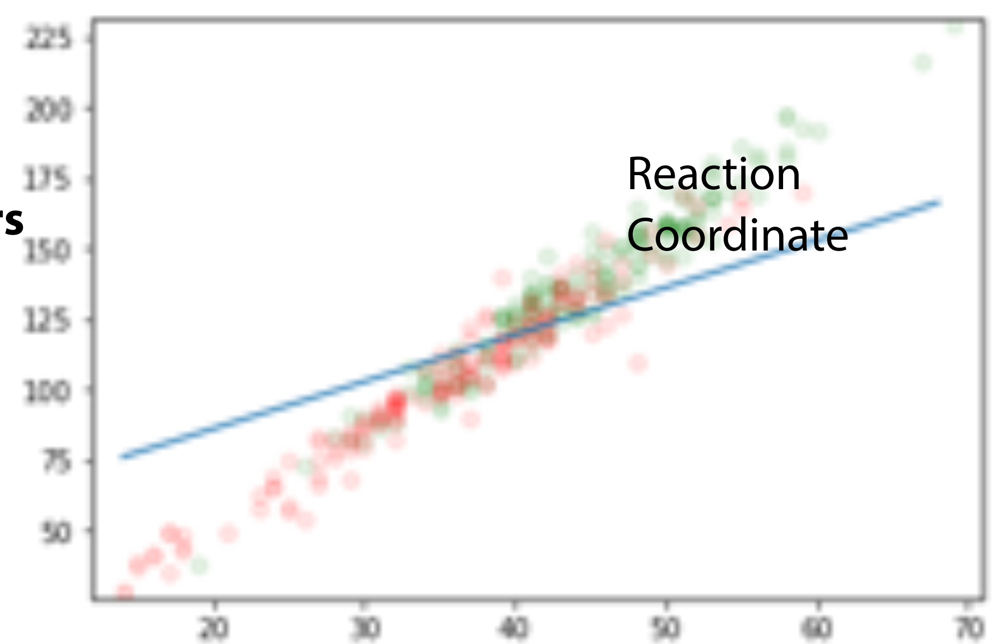
Reaction Coordinate Analysis

LME gives best model allowing two collective variables

High Undercooling

(270 and 275K)

No of Waters in Nucleus

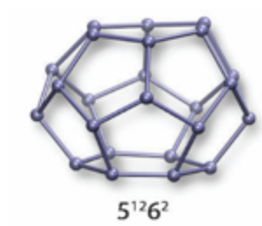


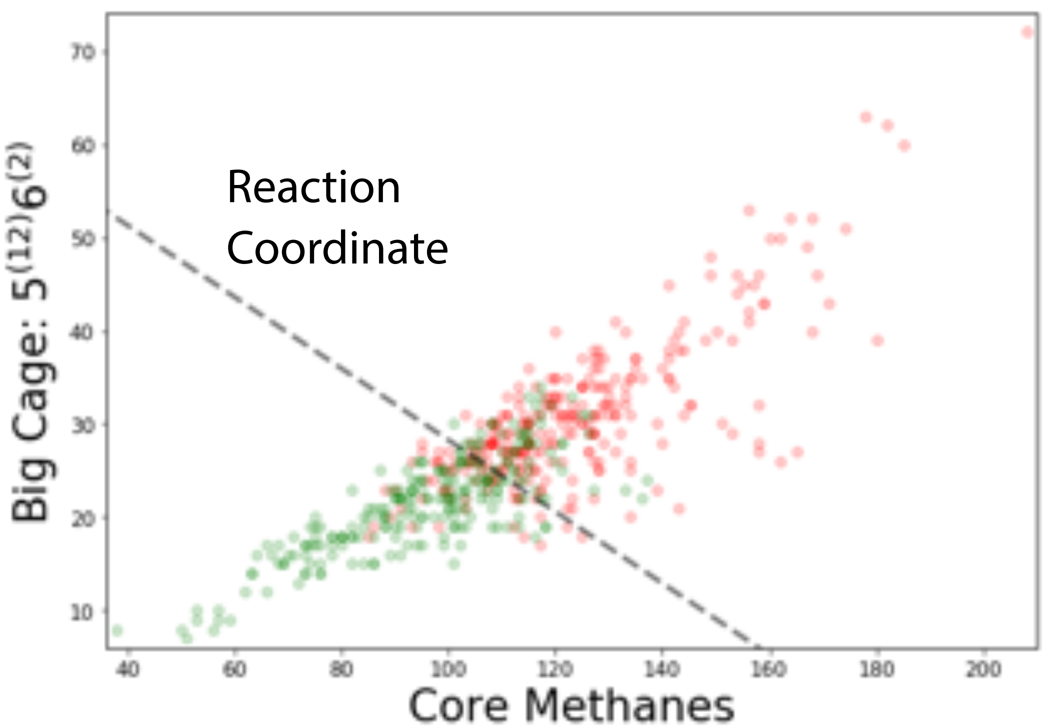
Methane (MCG)

Low Undercooling

(280 and 285K)

Big Cages





High undercooling

- Only **Size** matters

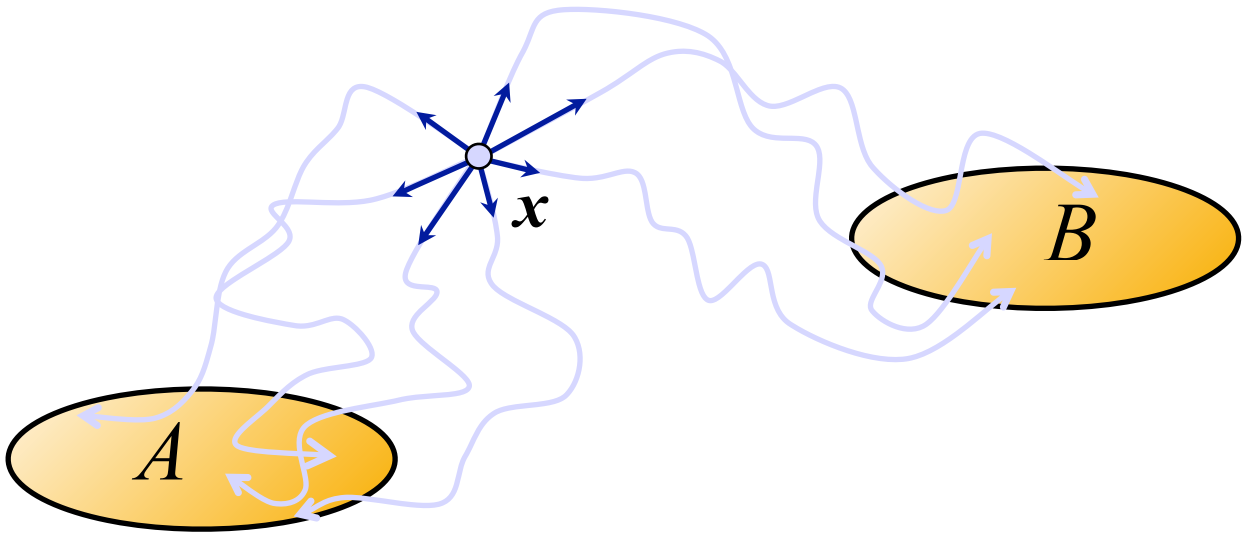
Low undercooling

- Both **Size** and **Structure** are important

important collective variables : nucleus size and big cage content

Arjun, Berendsen, PGB, PNAS, 2019

Machine learning of reaction coordinates

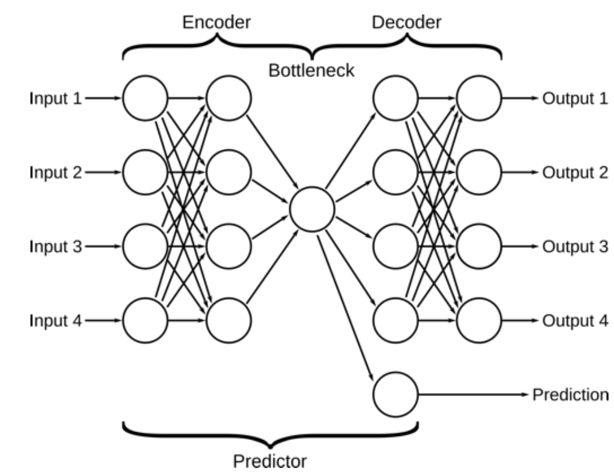


- Committor $p_B(x)$ is THE reaction coordinate
- Committor is high dimensional function; difficult to gain physical insight
- **dimensionality reduction:** find best low dimensional order parameter combination that best represents committor

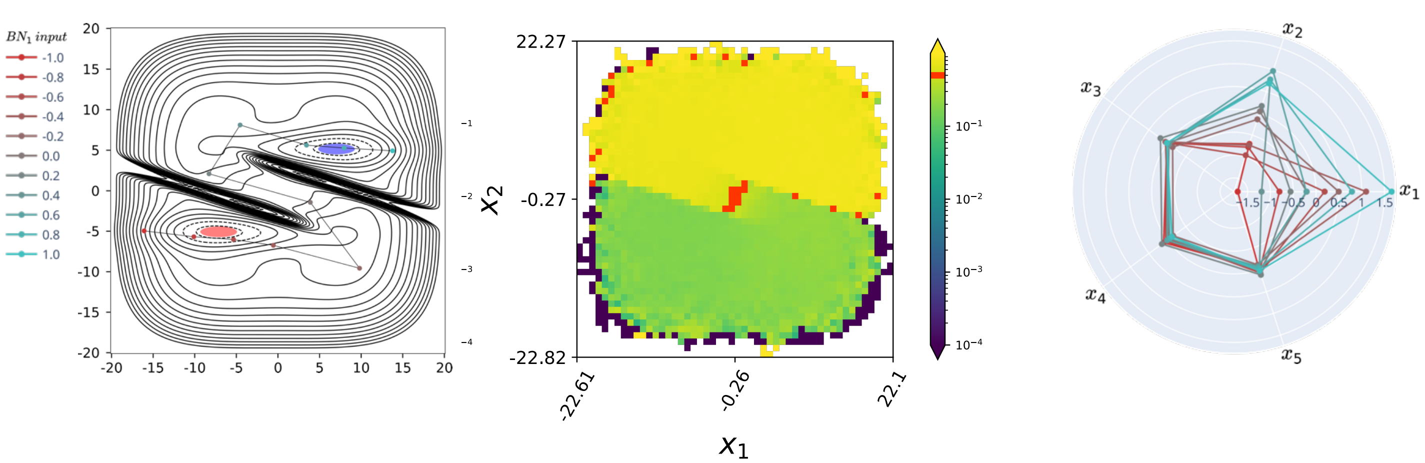
- Interpret each path of reweighted path ensemble as shot. Use info to optimise reaction coordinate model $r(q_1, q_2, ..)$

$$L(\alpha) = \prod_{i=1}^{N_B} p_B(r(q(\mathbf{x}_i^{(B)}))\prod_{i=1}^{N_A}(1 - p_B(r(q(\mathbf{x}_i^{(B)})))$$

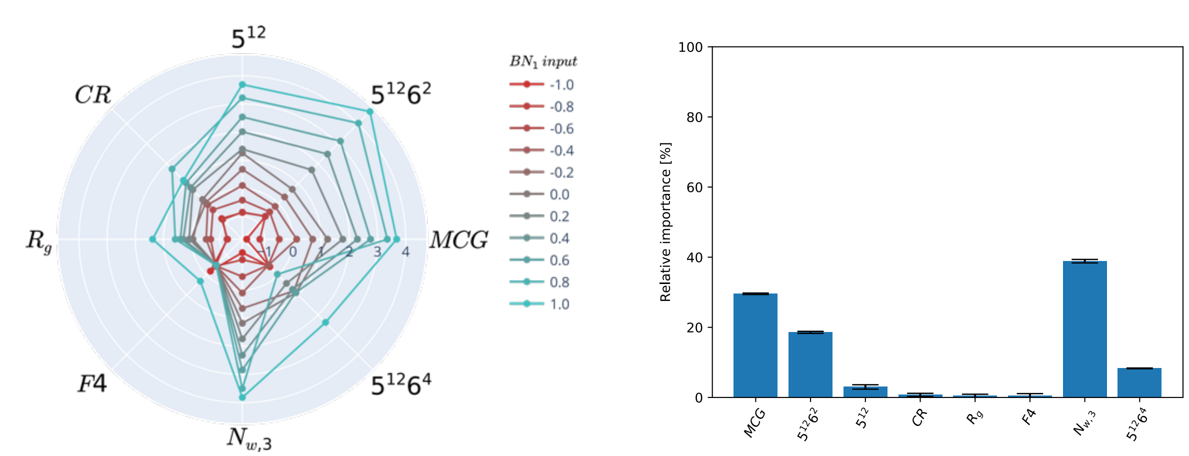
- Likelihood maximisation of predicted committor model
- use auto encoder to find optimal CV combination
-



extended autoencoder



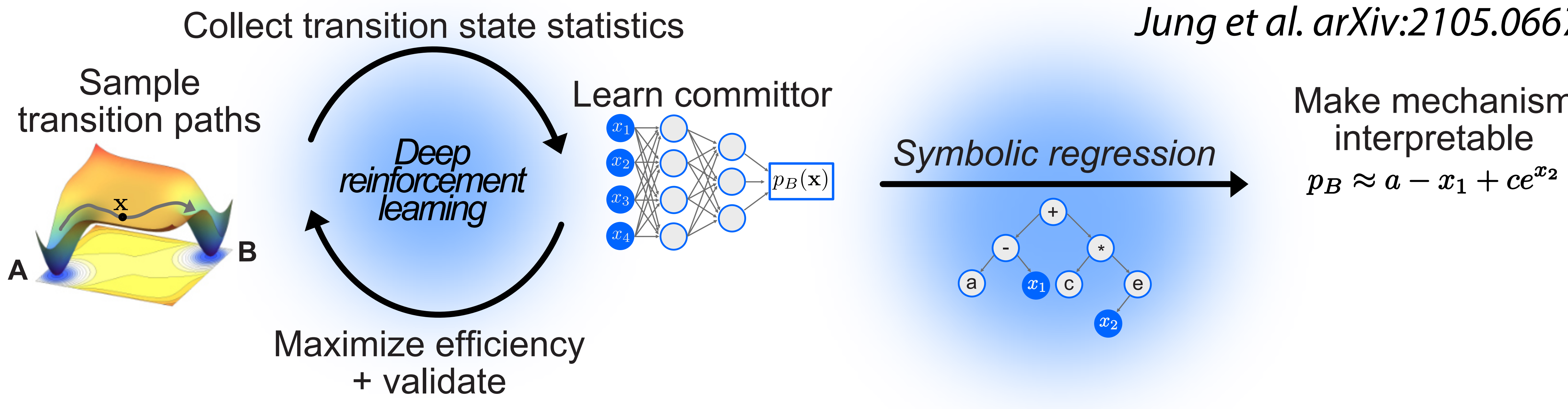
M. Frassek, Arjun, PGB, JCP 2021 doi 10.1063/5.0058639



hydrate latent space path

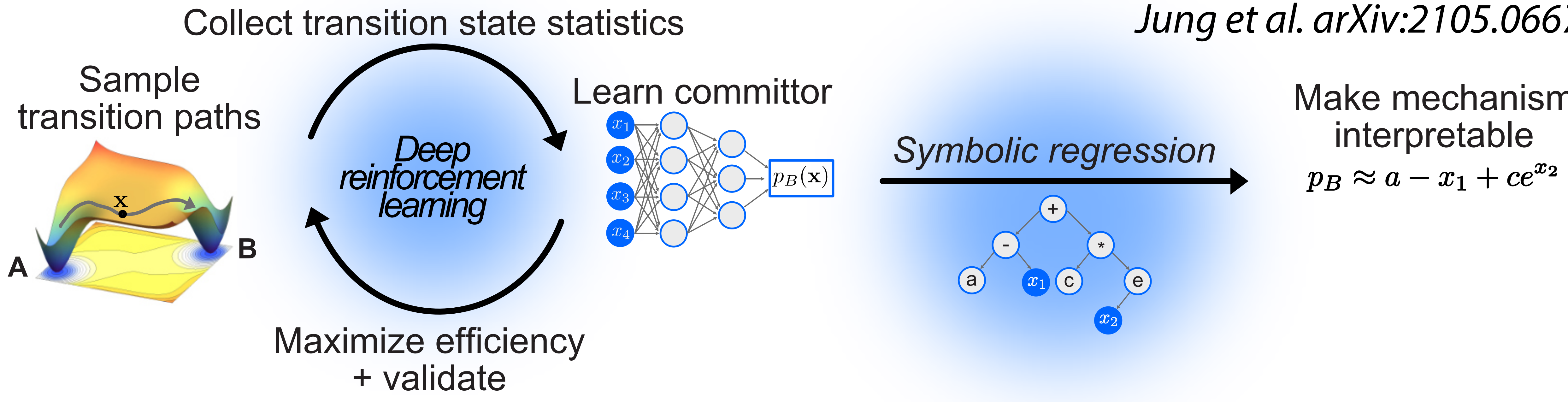
Learning the sampling & the RC together

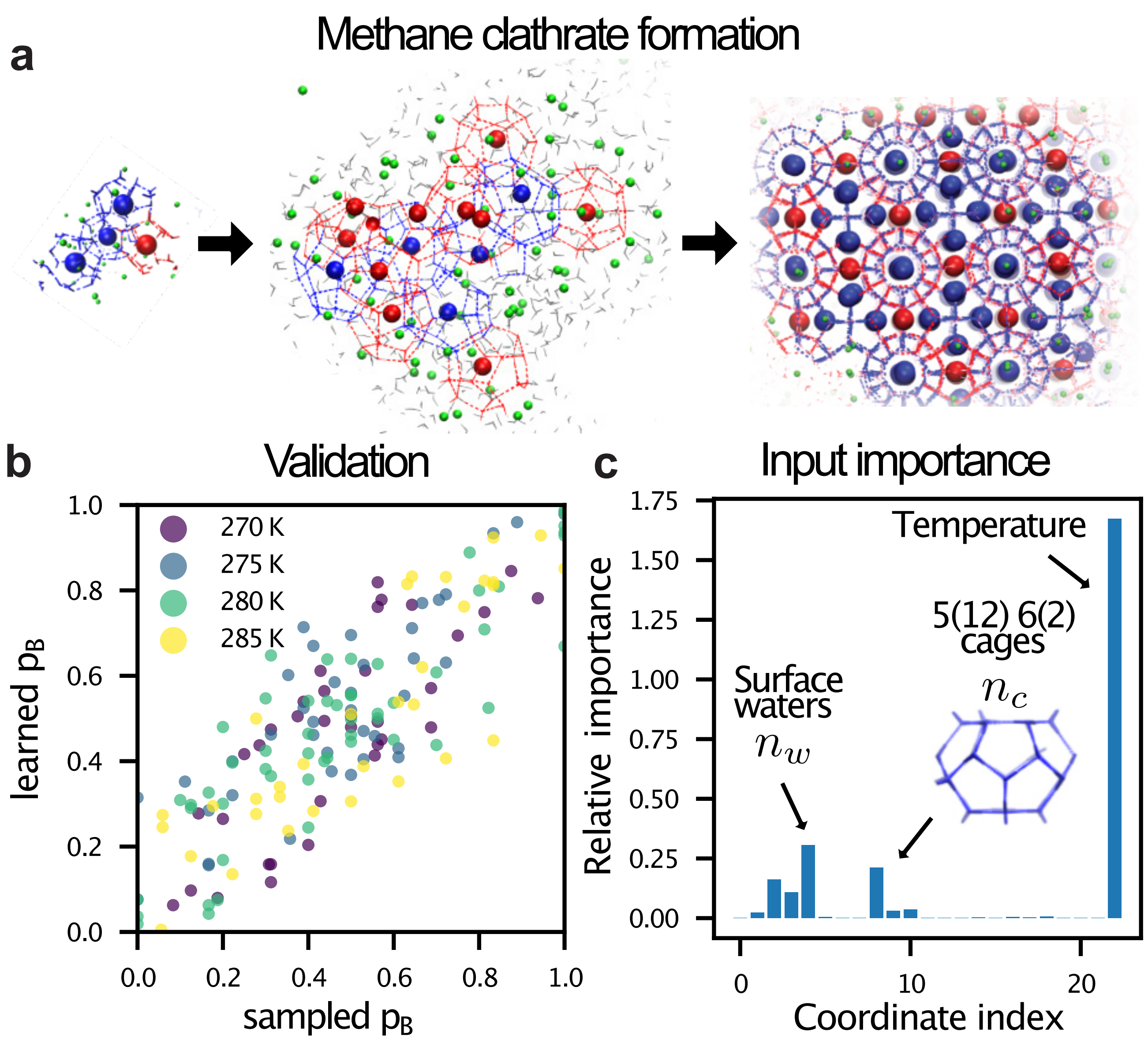
Jung et al. arXiv:2105.06673



Learning the sampling & the RC together

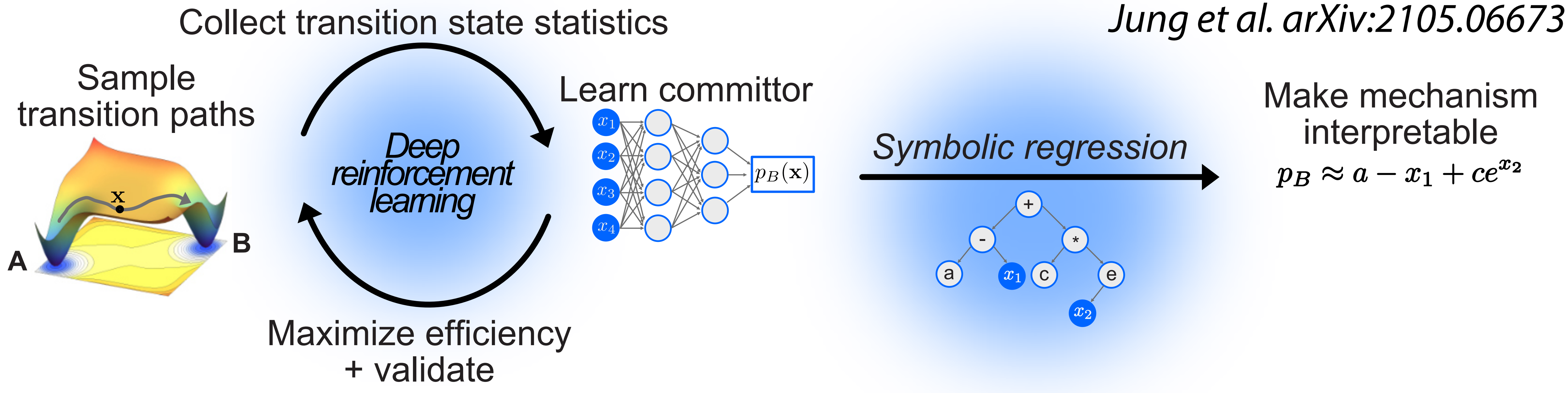
Jung et al. arXiv:2105.06673

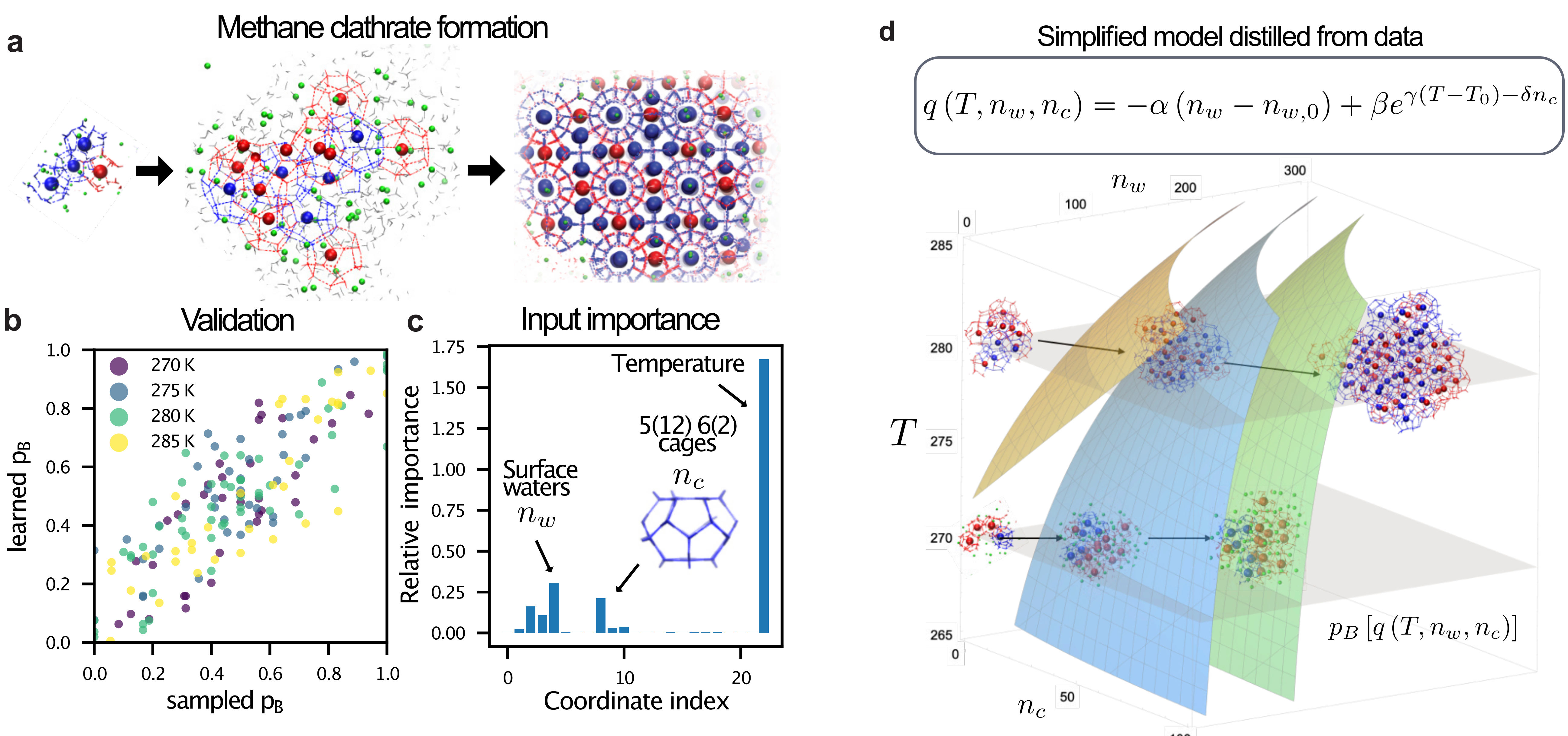




Learning the sampling & the RC together

Jung et al. arXiv:2105.06673





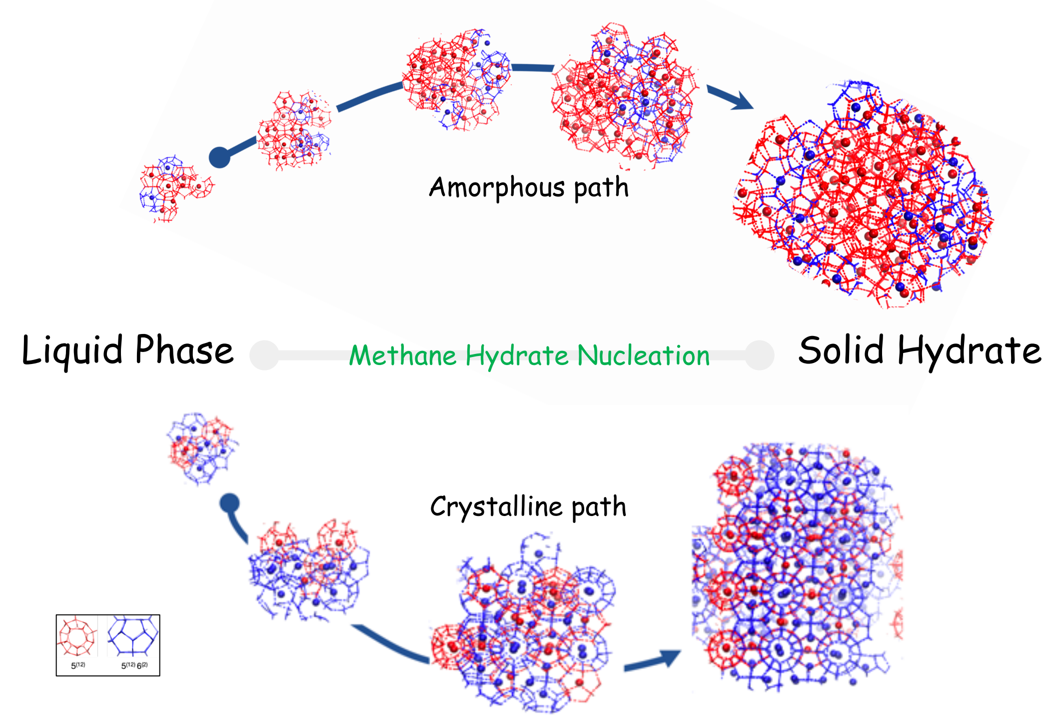
Summary hydrate formation

- TPS of hydrate formation under natural conditions using realistic models
  - path ensemble (over 1 ms) shows broad distribution of transitions
  - at high undercooling forms amorphous solid, at low undercooling forms crystal
  - not a gradual shift: at 280 K both routes can coexist
  - nucleation rate closer to experimentally predicted range

- reaction coordinate analysis
  - only size of the nucleus important at high undercooling
  - size & structure of nucleus important at low undercooling (anti correlated)
  - machine learning can identify non-linear function

- Polymorph selection:
  - 5(12)6(2) cages important at high T
  - occurs in precritical regime

- crystallisation at low undercooling does not follow Ostwald step rule: metastable phase avoided



Arjun, Berendsen, PGB, PNAS, 2019

Arjun and PGB . J Phys. Chem. B 2020, 124, 37, 8099

The OpenPathSampling package

- a python library for path sampling simulations
  - works with OpenMM and simple dynamics
  - Gromacs, Lammps support
  - uses MdTraj, OpenMM
- OPS allows flexible definition of
  - states, trajectory ensembles
  - sets of interfaces, networks of transitions
- OPS provides algorithms for sampling
  - TPS, (fixed or flexible length) MSTPS
  - TIS, MSTIS, RETIS (SRTIS)
  - committors, reactive flux
- OPS provides analysis tools
  - crossing probabilities
  - rates, free energies, path densities…..

OpenPathSampling

http://openpathsampling.org

Twitter: @pathsampling
Development at:
http://github.com/openpathsampling/

Swenson, Prinz, Noe, Chodera, PGB, JCTC, 2019

✓ **Easy to use**: Beginners can quickly learn to use it

✓ **Easy to extend**: Advanced users can use it to develop new methods

✓ **Independent of dynamics engine**: Useful in many fields and to the broadest audience

Conclusions

- transition path sampling yields unbiased ensemble of reactive trajectories
- committor based analysis yields reaction coordinate
- TIS yields kinetic rate constants predictions
- Multiple state versions allow sampling of kinetic reaction network
- Reweighted path ensemble allows evaluation of full reaction coordinate
- Simultaneous path sampling & RC analysis possible with Machine Learning
- Open Path Sampling makes all of this available to the community: www.openpathsampling.org

Conclusions

- transition path sampling yields unbiased ensemble of reactive trajectories
- committor based analysis yields reaction coordinate
- TIS yields kinetic rate constants predictions
- Multiple state versions allow sampling of kinetic reaction network
- Reweighted path ensemble allows evaluation of full reaction coordinate
- Simultaneous path sampling & RC analysis possible with Machine Learning
- Open Path Sampling makes all of this available to the community: www.openpathsampling.org

Transition path sampling allows exploration and understanding of kinetics of complex rare event protein and DNA dynamics

Acknowledgements

OPS


David Swenson


Jan-Hendrik Prinz


John Chodera

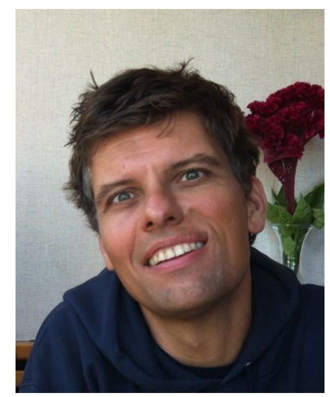

Frank Noe

UvA


Faidon Brotzakis


Arjun


Jocelyne Vreede


Bernd Ensing

Coll.


Titus van Erp


Christoph Dellago


Gerhard Hummer


Roberto Covino

OpenPathSampling

NWO

ecam

UvA AI4Science Laboratory

Outline

- Introduction
- Rare events

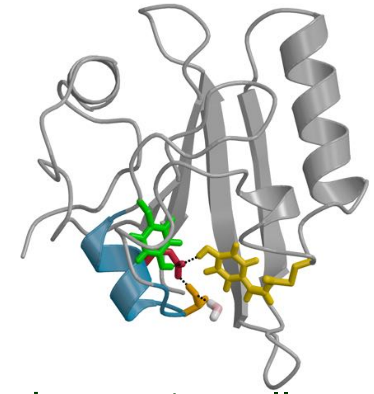
part 1:

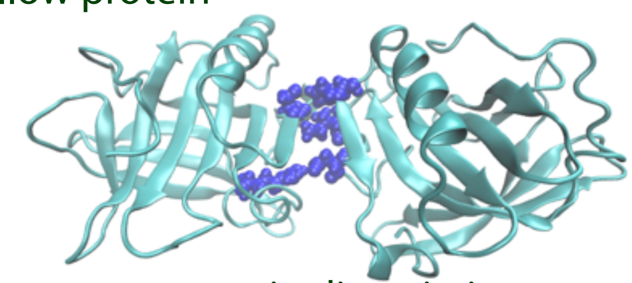
- Transition Path Sampling
- Committor & Reaction coordinate analysis
- Rate constants with transition interface sampling
- reaction networks with multiple state TPS/TIS
- advanced developments & machine learning
- OPS software

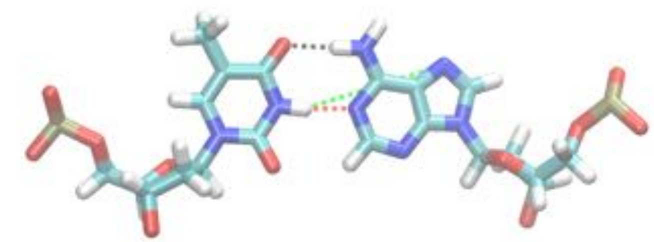
part 2:

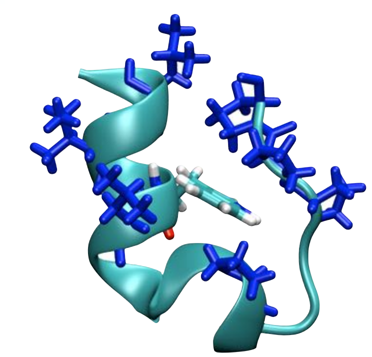
- **imposing kinetic constraints**
- path reweighting with Maximum Caliber

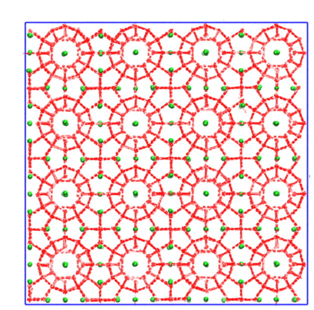
- conclusions


photoactive yellow protein


protein dissociation


DNA base pair rotation


Trp cage folding


gas hydrate formation

Part 2: Imposing experimental kinetics

- Classical MD can predict statics and dynamics but has two sources of error:
  - sampling problem
  - **systematic force field error**

- Combining MD with experiments can compensate FF errors, using the maximum entropy (MaxEnt) framework. *(Vendruscolo et al; Cesari, Reißer and Bussi, Computation 2018, 6, 15)*

- Can we do the same for kinetics?

*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Part 2: Imposing experimental kinetics

- Classical MD can predict statics and dynamics but has two sources of error:
  - sampling problem
  - **systematic force field error**

- Combining MD with experiments can compensate FF errors, using the maximum entropy (MaxEnt) framework. *(Vendruscolo et al; Cesari, Reißer and Bussi, Computation 2018, 6, 15)*

- Can we do the same for kinetics?

Obvious strategy:

Problem : (re)computing kinetics very expensive.

Better option: reuse existing trajectory data, and correct for error

Take cue from MaxEnt: put experimental constraint on path distribution, using the maximum path entropy approach

Brotzakis, Vendruscolo, PGB, PNAS ***118*** *, e2012423118 (2021).*

Maximum Caliber

the Maximum Caliber approach (Jaynes 1980) is a variational path based framework used in (non)-equilibrium statistical mechanics (mostly used in discrete systems, see e.g. Dill et al)

the path entropy (caliber) is

$$S[\mathcal{P}\|\mathcal{P}^0] = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]},$$

now optimise path distribution $\mathscr{P}[\mathbf{x}]$

$$\mathcal{P}^{MC}[\mathbf{x}] = \arg\max S[\mathcal{P}\|\mathcal{P}^0],$$

$$\text{subject to: } \begin{cases} \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] s_i[\mathbf{x}] = \langle s_i[\mathbf{x}] \rangle = s_i^{exp} \\ \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] = 1. \end{cases}$$

constraint could be kinetic rate constant

maximisation (with method of Lagrange multipliers) gives

$$\mathcal{L} = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]} - \nu \left(\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] - 1 \right) - \sum_i \mu_i \left(\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] s_i[\mathbf{x}] - s_i^{exp} \right),$$

leading to $\mathcal{P}^{MC}[\mathbf{x}] \propto e^{-\sum_i \mu_i s_i[\mathbf{x}]} \mathcal{P}^0[\mathbf{x}].$

with path probability $\mathcal{P}^0[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} p(x_i \to x_{i+1}),$

Maximum Caliber

the Maximum Caliber approach (Jaynes 1980) is a variational path based framework used in (non)-equilibrium statistical mechanics (mostly used in discrete systems, see e.g. Dill et al)

the path entropy (caliber) is

$$S[\mathcal{P}\|\mathcal{P}^0] = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]},$$

now optimise path distribution $\mathscr{P}[\mathbf{x}]$

$$\mathcal{P}^{MC}[\mathbf{x}] = \arg\max S[\mathcal{P}\|\mathcal{P}^0],$$

$$\text{subject to: } \begin{cases} \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}]s_i[\mathbf{x}] = \langle s_i[\mathbf{x}]\rangle = s_i^{exp} \\ \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] = 1. \end{cases}$$

constraint could be kinetic rate constant

maximisation (with method of Lagrange mu

$$\mathcal{L} = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]} - \nu\left(\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \cdots - s_i^{exp}\right),$$

similar to MaxEnt, but depends on path x

leading to $\mathcal{P}^{MC}[\mathbf{x}] \propto e^{-\sum_i \mu_i s_i[\mathbf{x}]}\mathcal{P}^0[\mathbf{x}].$

with path probability $\mathcal{P}^0[\mathbf{x}] = \rho(x_0)\prod_{i=0}^{L-1} p(x_i \to x_{i+1}),$

Maximum Caliber

the Maximum Caliber approach (Jaynes 1980) is a variational path based framework used in (non)-equilibrium statistical mechanics (mostly used in discrete systems, see e.g. Dill et al)

the path entropy (caliber) is

$$S[\mathcal{P}\|\mathcal{P}^0] = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]},$$

now optimise path distribution $\mathscr{P}[\mathbf{x}]$

$$\mathcal{P}^{MC}[\mathbf{x}] = \arg\max S[\mathcal{P}\|\mathcal{P}^0],$$

$$\text{subject to: } \begin{cases} \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}]\, s_i[\mathbf{x}] = \langle s_i[\mathbf{x}]\rangle = s_i^{exp} \\ \int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] = 1. \end{cases}$$

constraint could be kinetic rate constant

maximisation (with method of Lagrange mu

$$\mathcal{L} = -\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ln \frac{\mathcal{P}[\mathbf{x}]}{\mathcal{P}^0[\mathbf{x}]} - \nu\left(\int \mathcal{D}\mathbf{x}\mathcal{P}[\mathbf{x}] \ldots - s_i^{exp}\right),$$

similar to MaxEnt, but depends on path x

leading to $\mathcal{P}^{MC}[\mathbf{x}] \propto e^{-\sum_i \mu_i s_i[\mathbf{x}]}\mathcal{P}^0[\mathbf{x}].$

with path probability $\mathcal{P}^0[\mathbf{x}] = \rho(x_0)\prod_{i=0}^{L-1} p(x_i \to x_{i+1}),$

also similarities to tilting of path ensembles, e.g. s-ensemble *(see Hedges et al. Science 323, 1309 (2009)*

MaxCal for rate constant constraint

TIS gives expressions for rate constant based on path ensembles

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0), \qquad P_A(\lambda|\lambda_0) = \int \mathcal{D}\mathbf{x}\mathcal{P}_A[\mathbf{x}]\theta(\lambda_{max}[\mathbf{x}] - \lambda),$$

we impose correct rate at all interfaces λi; standard optimisation gives

$$\mathcal{P}_A^{MC}[\mathbf{x}] \propto e^{f_A(\lambda_{max}[\mathbf{x}])}\mathcal{P}_A^0[\mathbf{x}],$$

$$f_A(\lambda_{max}[\mathbf{x}]) \equiv -\sum_{i=1}^{n} \mu_i \theta(\lambda_{max}[\mathbf{x}] - \lambda_i) P_A(\lambda_n|\lambda_i),$$

Brotzakis, Vendruscolo, PGB, PNAS ***118*** *, e2012423118 (2021).*

MaxCal for rate constant constraint

TIS gives expressions for rate constant based on path ensembles

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0), \qquad P_A(\lambda|\lambda_0) = \int \mathcal{D}\mathbf{x}\mathcal{P}_A[\mathbf{x}]\theta(\lambda_{max}[\mathbf{x}] - \lambda),$$

we impose correct rate at all interfaces λi; standard optimisation gives

$$\mathcal{P}_A^{MC}[\mathbf{x}] \propto e^{f_A(\lambda_{max}[\mathbf{x}])}\mathcal{P}_A^0[\mathbf{x}],$$

$$f_A(\lambda_{max}[\mathbf{x}]) \equiv -\sum_{i=1}^{n} \mu_i \theta(\lambda_{max}[\mathbf{x}] - \lambda_i) P_A(\lambda_n|\lambda_i),$$

f(λ) is not fixed except at λ_B: determine *f(λ)* from projected **configurational density**:

$$\rho_A^{MC}(\lambda) \propto \int \mathcal{D}\mathbf{x}\mathcal{P}_A^0[\mathbf{x}]e^{f_A(\lambda_{max}[\mathbf{x}])}\sum_{k=0}^{L[\mathbf{x}]}\delta(\lambda(x_k) - \lambda).$$

This should be equal the density obtained from a MaxEnt approach

$$\rho^{ME}(x) \propto e^{-\mu g(x)}\rho^0(x), \longrightarrow \rho^{ME}(\lambda) \propto e^{-\mu g(\lambda)}\rho^0(\lambda),$$

But what is g(λ)?

Brotzakis, Vendruscolo, PGB, PNAS ***118*** *, e2012423118 (2021).*

MaxCal for rate constant constraint

TIS gives expressions for rate constant based on path ensembles

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0), \qquad P_A(\lambda|\lambda_0) = \int \mathcal{D}\mathbf{x}\mathcal{P}_A[\mathbf{x}]\theta(\lambda_{max}[\mathbf{x}] - \lambda),$$

we impose correct rate at all interfaces λi; standard optimisation gives

$$\mathcal{P}_A^{MC}[\mathbf{x}] \propto e^{f_A(\lambda_{max}[\mathbf{x}])}\mathcal{P}_A^0[\mathbf{x}],$$

$$f_A(\lambda_{max}[\mathbf{x}]) \equiv -\sum_{i=1}^{n} \mu_i \theta(\lambda_{max}[\mathbf{x}] - \lambda_i) P_A(\lambda_n|\lambda_i),$$

f(λ) is not fixed except at λ_{B}: determine *f(λ)* from projected **configurational density**:

$$\rho_A^{MC}(\lambda) \propto \int \mathcal{D}\mathbf{x}\mathcal{P}_A^0[\mathbf{x}]e^{f_A(\lambda_{max}[\mathbf{x}])} \sum_{k=0}^{L[\mathbf{x}]} \delta(\lambda(x_k) - \lambda).$$

This should be equal the density obtained from a MaxEnt approach

$$\rho^{ME}(x) \propto e^{-\mu g(x)} \rho^0(x), \longrightarrow \rho^{ME}(\lambda) \propto e^{-\mu g(\lambda)} \rho^0(\lambda),$$

But what is g(λ)? **determine from known exp. equilibrium constant (fraction)**

$$\frac{\int d\lambda g(\lambda)\rho(\lambda)}{\int d\lambda \rho(\lambda)} = K_{exp} = \frac{1}{1 + k_{BA}^{exp}/k_{AB}^{exp}}$$

Brotzakis, Vendruscolo, PGB, PNAS ***118***, *e2012423118 (2021).*

MaxCal for rate constant constraint

TIS gives expressions for rate constant based on path ensembles

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0), \qquad P_A(\lambda|\lambda_0) = \int \mathcal{D}\mathbf{x}\mathcal{P}_A[\mathbf{x}]\theta(\lambda_{max}[\mathbf{x}] - \lambda),$$

we impose correct rate at all interfaces λi; standard optimisation gives

$$\mathcal{P}_A^{MC}[\mathbf{x}] \propto e^{f_A(\lambda_{max}[\mathbf{x}])}\mathcal{P}_A^0[\mathbf{x}],$$

$$f_A(\lambda_{max}[\mathbf{x}]) \equiv -\sum_{i=1}^{n} \mu_i \theta(\lambda_{max}[\mathbf{x}] - \lambda_i) P_A(\lambda_n|\lambda_i),$$

f(λ) is not fixed except at λ_B: determine *f(λ)* from projected **configurational density**:

$$\rho_A^{MC}(\lambda) \propto \int \mathcal{D}\mathbf{x}\mathcal{P}_A^0[\mathbf{x}]e^{f_A(\lambda_{max}[\mathbf{x}])}\sum_{k=0}^{L[\mathbf{x}]}\delta(\lambda(x_k) - \lambda).$$

This should be equal the density obtained from a MaxEnt approach

$$\rho^{ME}(x) \propto e^{-\mu g(x)}\rho^0(x), \longrightarrow \rho^{ME}(\lambda) \propto e^{-\mu g(\lambda)}\rho^0(\lambda),$$

But what is g(λ)? **determine from known exp. equilibrium constant (fraction)**

$$\frac{\int d\lambda g(\lambda)\rho(\lambda)}{\int d\lambda \rho(\lambda)} = \frac{\int d\lambda p_B(\lambda)\rho(\lambda)}{\int d\lambda \rho(\lambda)} = \frac{\int d\lambda \rho_B(\lambda)}{\int d\lambda \rho(\lambda)} = K_{exp}$$

Brotzakis, Vendruscolo, PGB, PNAS ***118***, *e2012423118 (2021).*

MaxCal for rate constant constraint

TIS gives expressions for rate constant based on path ensembles

$$k_{AB} = \phi_0 P_A(\lambda_B|\lambda_0), \qquad P_A(\lambda|\lambda_0) = \int \mathcal{D}\mathbf{x}\mathcal{P}_A[\mathbf{x}]\theta(\lambda_{max}[\mathbf{x}] - \lambda),$$

we impose correct rate at all interfaces λi; standard optimisation gives

$$\mathcal{P}_A^{MC}[\mathbf{x}] \propto e^{f_A(\lambda_{max}[\mathbf{x}])}\mathcal{P}_A^0[\mathbf{x}],$$

$$f_A(\lambda_{max}[\mathbf{x}]) \equiv -\sum_{i=1}^{n} \mu_i \theta(\lambda_{max}[\mathbf{x}] - \lambda_i) P_A(\lambda_n|\lambda_i),$$

f(λ) is not fixed except at λ_B: determine *f(λ)* from projected **configurational density**:

$$\rho_A^{MC}(\lambda) \propto \int \mathcal{D}\mathbf{x}\mathcal{P}_A^0[\mathbf{x}]e^{f_A(\lambda_{max}[\mathbf{x}])}\sum_{k=0}^{L[\mathbf{x}]}\delta(\lambda(x_k) - \lambda).$$

This should be equal the density obtained from a MaxEnt approach

$$\rho^{ME}(x) \propto e^{-\mu g(x)}\rho^0(x), \longrightarrow \rho^{ME}(\lambda) \propto e^{-\mu g(\lambda)}\rho^0(\lambda),$$

g(λ) is committor function!

But what is g(λ)? **determine from known exp. equilibrium constant (fraction)**

$$\frac{\int d\lambda g(\lambda)\rho(\lambda)}{\int d\lambda \rho(\lambda)} = \frac{\int d\lambda p_B(\lambda)\rho(\lambda)}{\int d\lambda \rho(\lambda)} = \frac{\int d\lambda \rho_B(\lambda)}{\int d\lambda \rho(\lambda)} = K_{exp}$$

*Brotzakis, Vendruscolo, PGB, PNAS **118** , e2012423118 (2021).*

Posterior density distributions and committor

MaxEnt reweighting of density gives

$$\rho_A(\lambda) = \rho_A^0(\lambda)\, e^{\mu_A p_B(\lambda)}$$

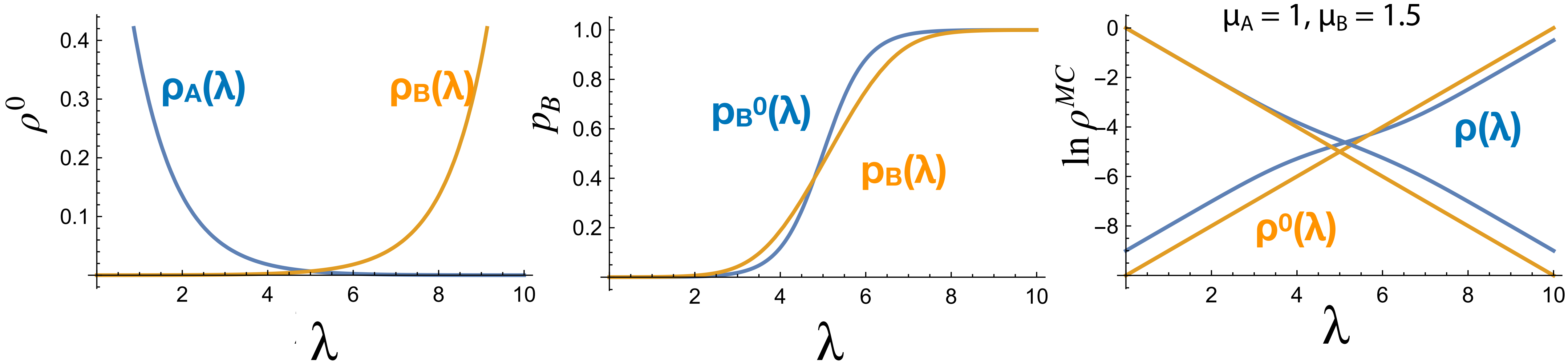
$$\rho_B(\lambda) = \rho_B^0(\lambda)\, e^{-\mu_B p_B(\lambda)}\, e^{\mu_A}$$

$$e^{\mu_A} = \frac{k_{AB}^{exp}}{k_{AB}^{0}}$$

$$e^{\mu_B} = \frac{k_{BA}^{exp}}{k_{BA}^{0}}$$

yielding corrected committor

$$p_B(\lambda) = \frac{\rho_B(\lambda)}{\rho_A(\lambda) + \rho_B(\lambda)}$$



*Brotzakis, Vendruscolo, PGB, PNAS **118** , e2012423118 (2021).*

Posterior density distributions and committor

MaxEnt reweighting of density gives

$$\rho_A(\lambda) = \rho_A^0(\lambda)\, e^{\mu_A p_B(\lambda)}$$

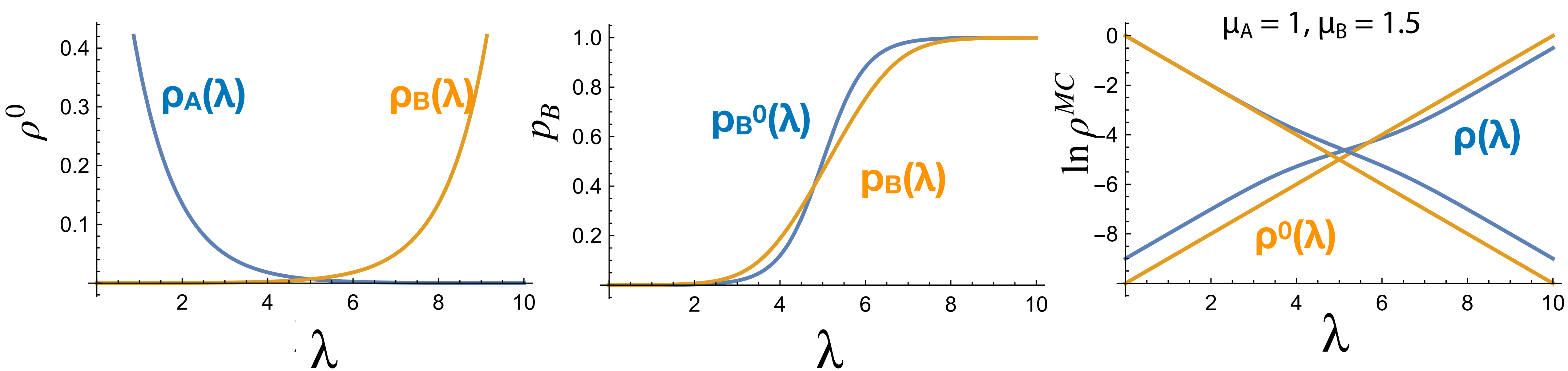
$$\rho_B(\lambda) = \rho_B^0(\lambda)\, e^{-\mu_B p_B(\lambda)}\, e^{\mu_A}$$

$$e^{\mu_A} = \frac{k_{AB}^{exp}}{k_{AB}^0}$$

$$e^{\mu_B} = \frac{k_{BA}^{exp}}{k_{BA}^0}$$

yielding corrected committor

$$p_B(\lambda) = \frac{\rho_B^0(\lambda)}{\rho_A^0(\lambda)\, e^{-\mu_A}\, e^{(\mu_A+\mu_B)p_B(\lambda)} + \rho_B^0(\lambda)}$$



*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Posterior density distributions and committor

MaxEnt reweighting of density gives

$$\rho_A(\lambda) = \rho_A^0(\lambda) e^{\mu_A p_B(\lambda)}$$

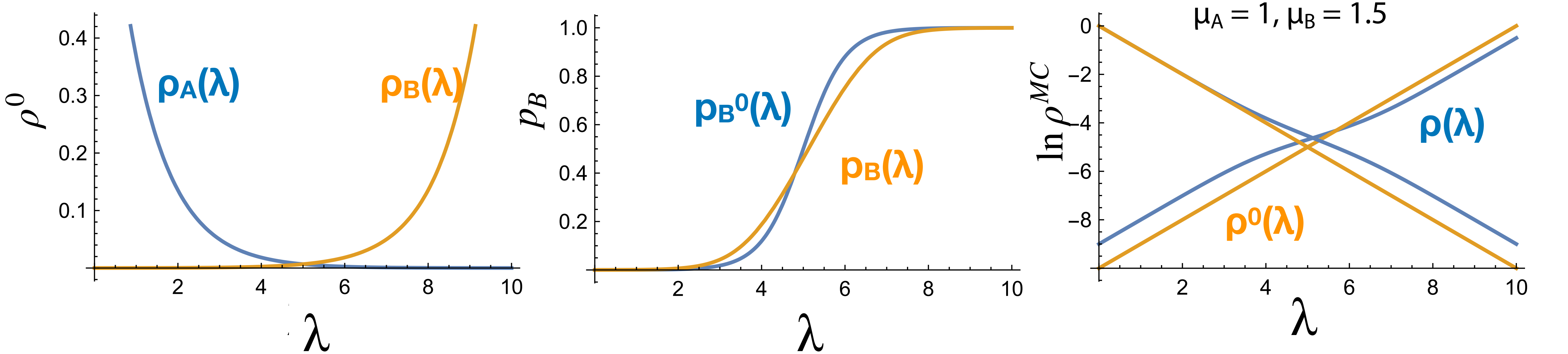
$$\rho_B(\lambda) = \rho_B^0(\lambda) e^{-\mu_B p_B(\lambda)} e^{\mu_A}$$

$$e^{\mu_A} = \frac{k_{AB}^{exp}}{k_{AB}^0}$$

$$e^{\mu_B} = \frac{k_{BA}^{exp}}{k_{BA}^0}$$

yielding corrected committor

$$p_B(\lambda) = \frac{\rho_B^0(\lambda)}{\rho_A^0(\lambda) e^{-\mu_A} e^{(\mu_A+\mu_B)p_B(\lambda)} + \rho_B^0(\lambda)}$$



The MaxEnt posterior density should be same as the MaxCal density posterior

$$\rho_A^{MC}(\lambda) = e^{-\mu g(\lambda)} \rho_A^0(\lambda), \qquad \text{with} \qquad g(\lambda) = p_B(\lambda)$$

*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Posterior density distributions and committor

MaxEnt reweighting of density gives

$$\rho_A(\lambda) = \rho_A^0(\lambda)\, e^{\mu_A p_B(\lambda)}$$

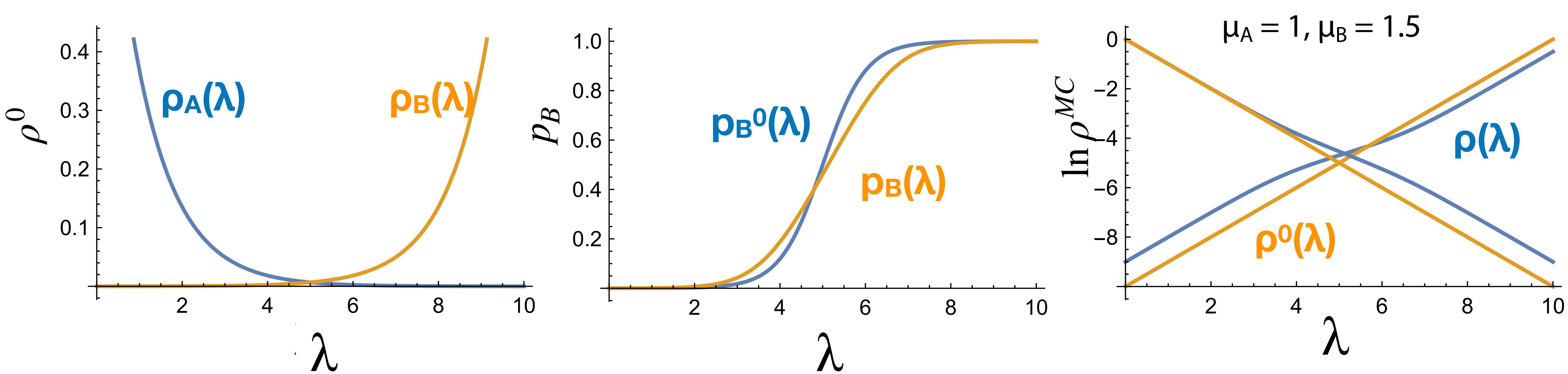
$$\rho_B(\lambda) = \rho_B^0(\lambda)\, e^{-\mu_B p_B(\lambda)}\, e^{\mu_A}$$

$$e^{\mu_A} = \frac{k_{AB}^{exp}}{k_{AB}^0}$$

$$e^{\mu_B} = \frac{k_{BA}^{exp}}{k_{BA}^0}$$

yielding corrected committor

$$p_B(\lambda) = \frac{\rho_B^0(\lambda)}{\rho_A^0(\lambda)\, e^{-\mu_A}\, e^{(\mu_A+\mu_B)p_B(\lambda)} + \rho_B^0(\lambda)}$$



The MaxEnt posterior density should be same as the MaxCal density posterior

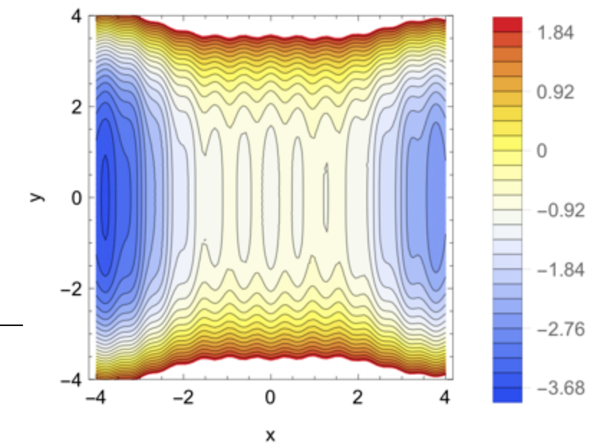
$$\rho_A^{MC}(\lambda) = e^{-\mu g(\lambda)}\rho_A^0(\lambda), \qquad \text{with} \qquad g(\lambda) = p_B(\lambda)$$

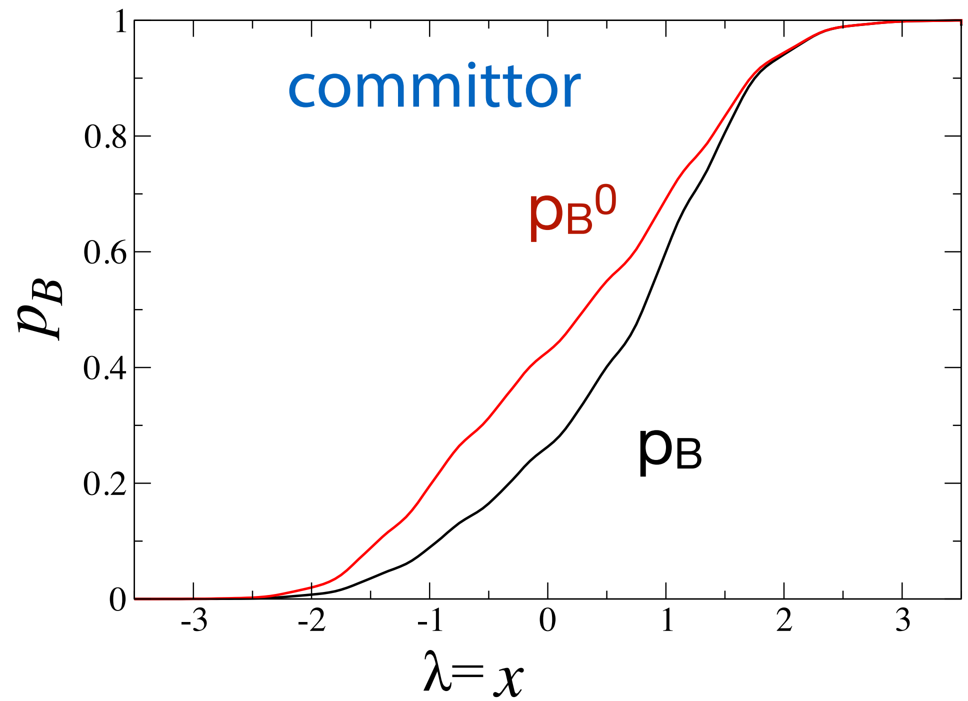
gives f(λ) !

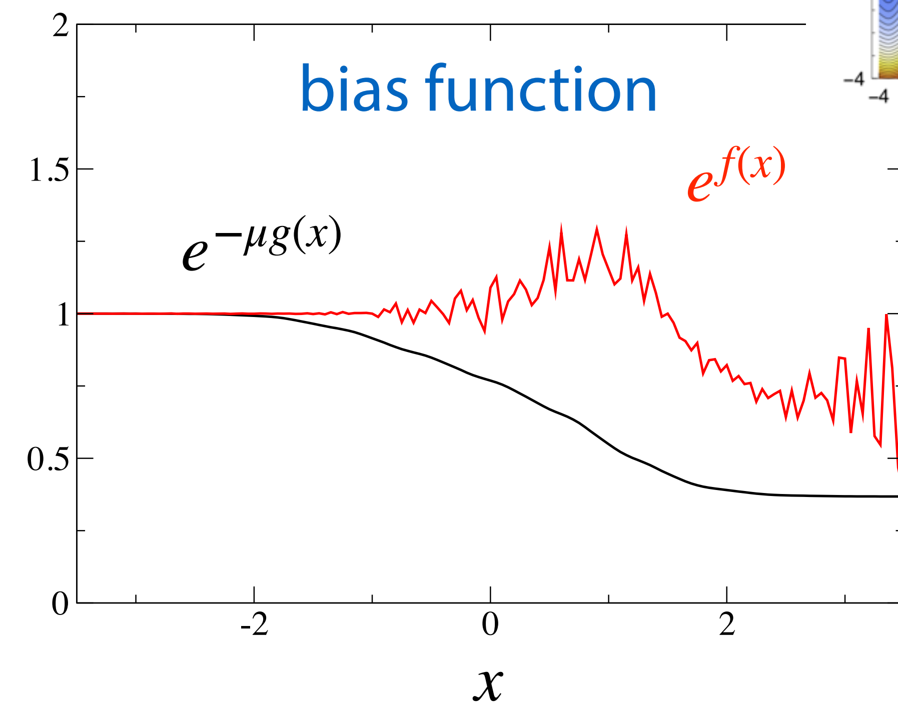
*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

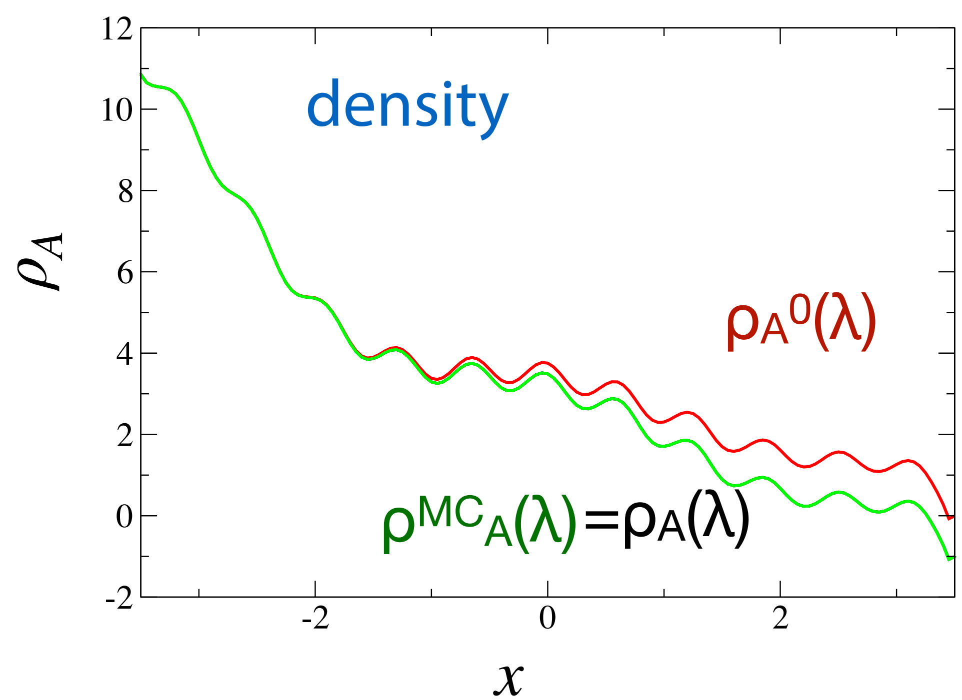
Application to 2D potential

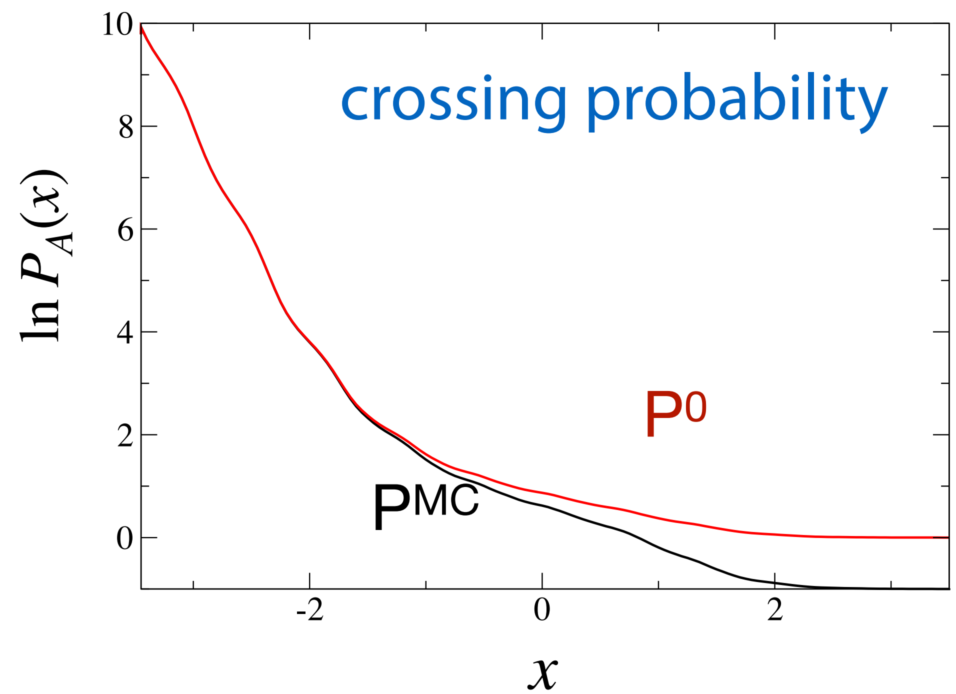
AB rate correction $\mu_A=-1$, BA rate correction $\mu_B=0$







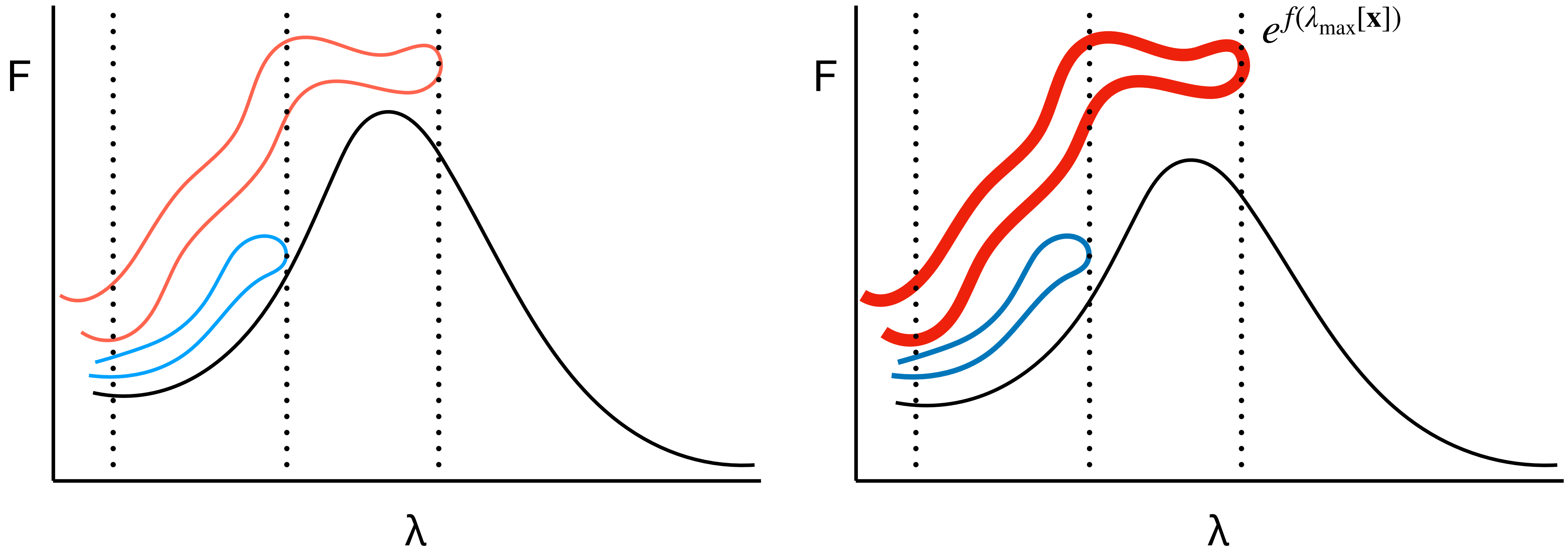




*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Interpretation of the MaxCal method

- MaxCal based method reweights existing trajectories or path ensembles (from MD, TIS, FFS, or other adaptive schemes).
- reweighting based on progress along λ: made more/less probable in the path ensembles
- rate constants are automatically constrained to the correct value (via $f_{A,B}(\lambda)$).
- fixing f(λ) requires a bias g(λ) based on the committor function
- method is enslaved to the original dynamics: so only distribution of initial conditions for paths is altered via the reweighting, the trajectories themselves do not change: analogous (but not identical) to microcanonical trajectory reweighting (e.g in Nested TPS)



*Brotzakis, Vendruscolo, PGB, PNAS **118** , e2012423118 (2021).*

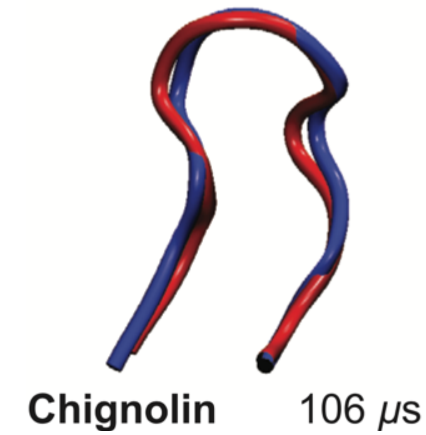
Application to chignolin folding

DE Shaw trajectory yield rate constants at exp. melting 341 K

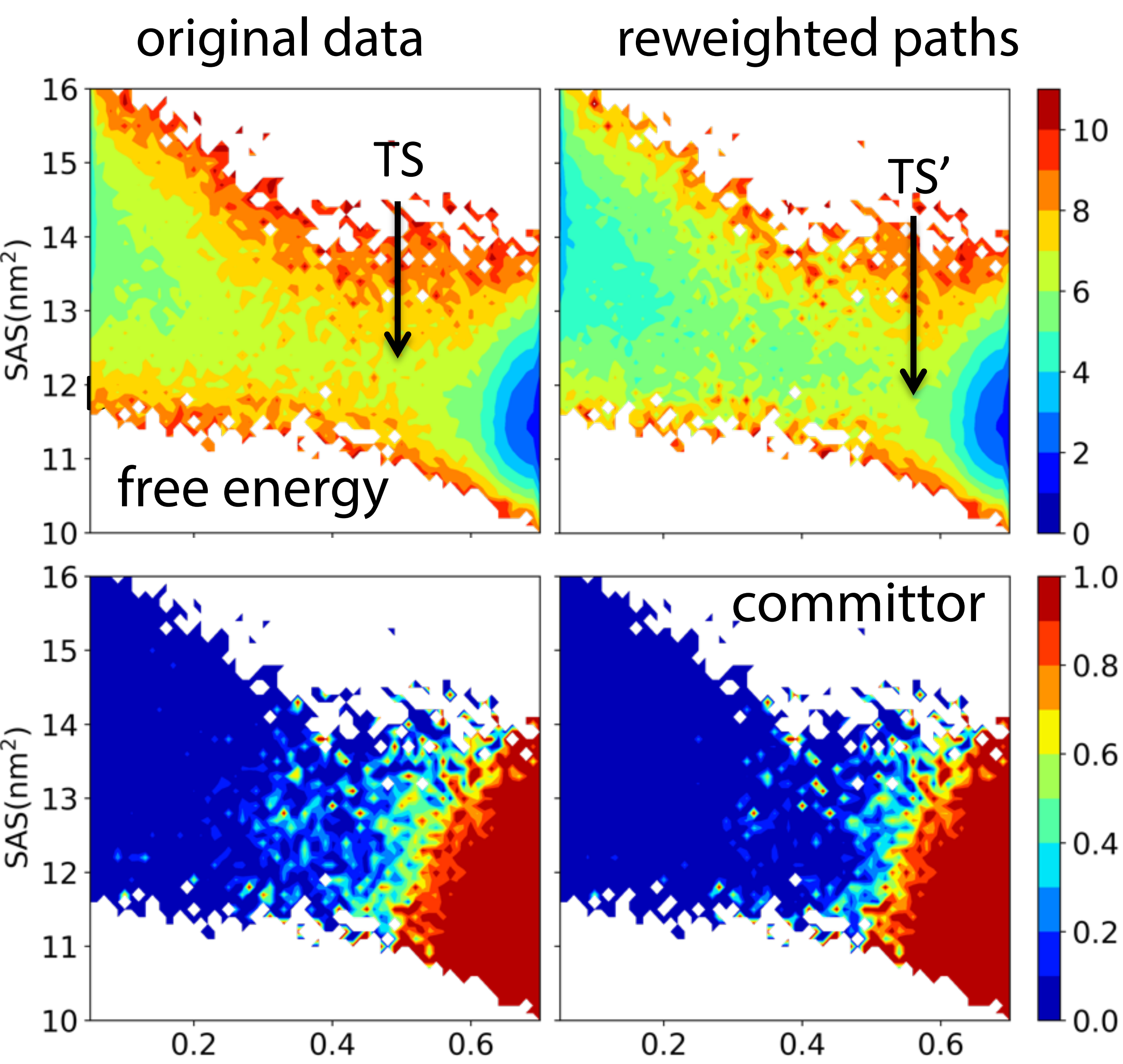
$k_f = 1.667\ \mu s^{-1}$

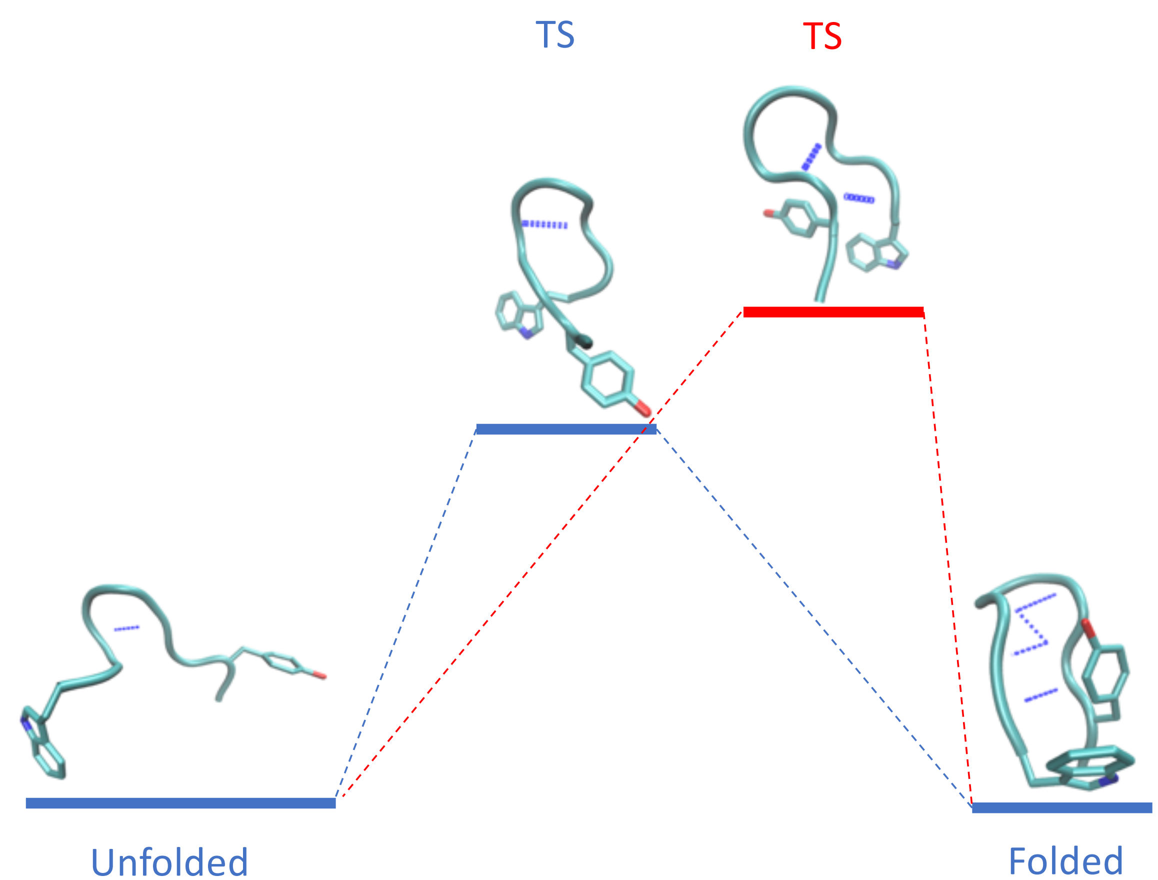
$k_u = 0.455\ \mu s^{-1}$

correct FF error by constraining folding rate also to $k_f = 0.455\ \mu s^{-1}$



Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2011). Science, 334(6055), 517–520





*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Application to chignolin folding

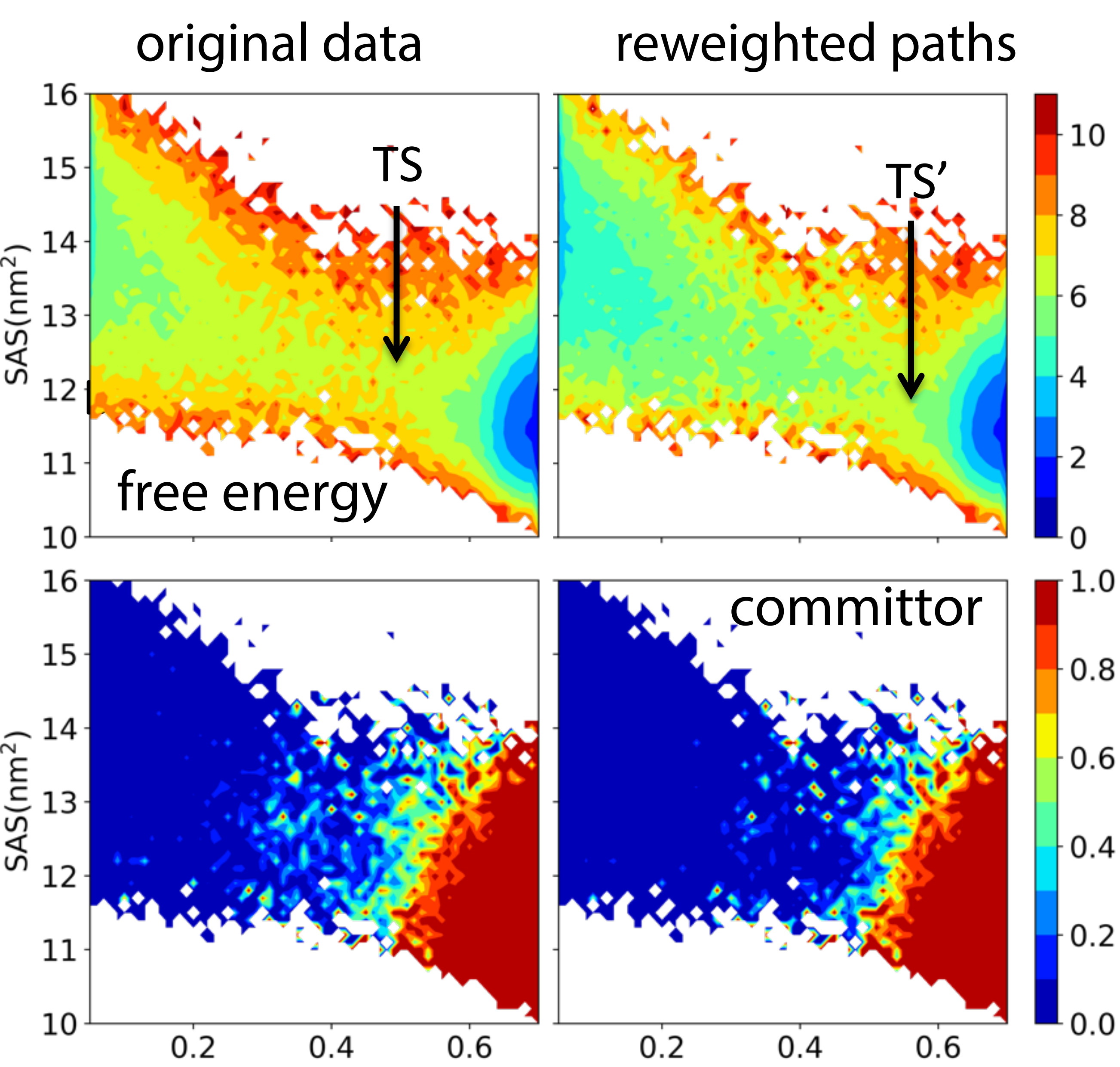
DE Shaw trajectory yield rate constants at exp. melting 341 K

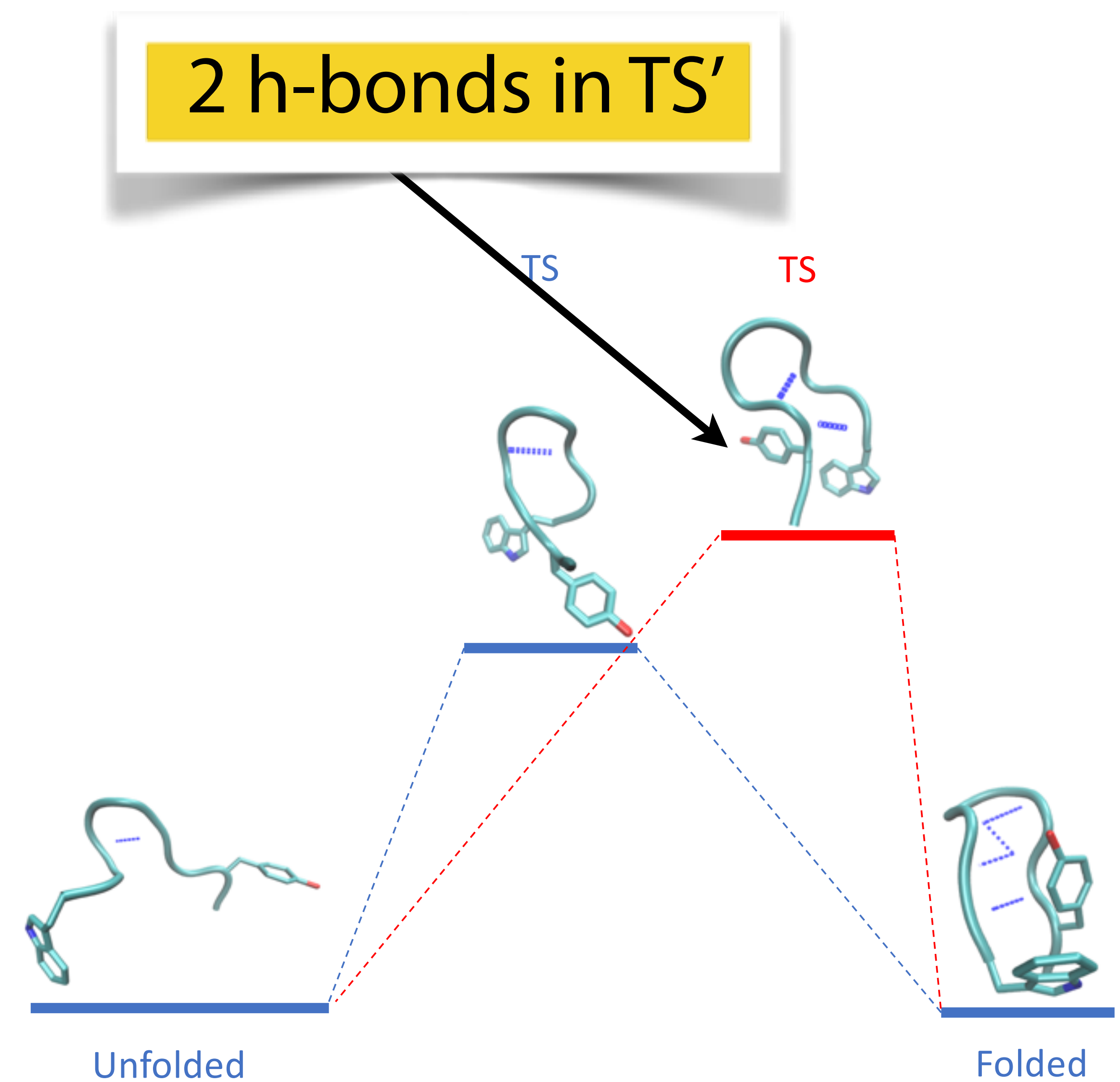
$k_f = 1.667\ \mu s^{-1}$

$k_u = 0.455\ \mu s^{-1}$

correct FF error by constraining folding rate also to $k_f = 0.455\ \mu s^{-1}$

Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2011). Science, 334(6055), 517–520

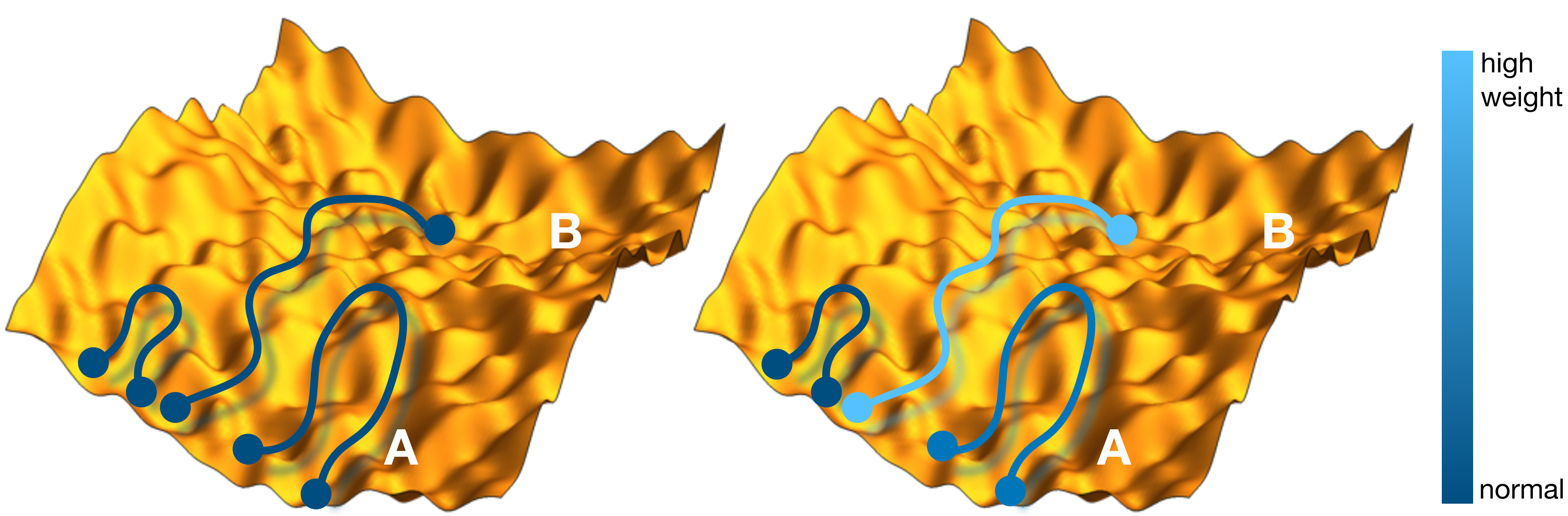




*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

Summary

- New general method that
  - imposes experimental dynamical constraints on path ensembles with MaxCal
  - can impose rate constant, and yields consistent free energy (configurational density) correctly, via committor.
  - is post-processing, no costly reevaluation of path ensembles needed.
  - reveals shifting of transition states and mechanisms.
  - is applicable to many processes in biology, physics, chemistry & material science.
- Outlook :
  - improve force fields, e.g. by computing the derivative of the path ensemble based rate constants w.r.t. to FF parameters.
  - make independence from CV definition



original paths **kinetically reweighted paths**

OpenPathSampling

*Brotzakis, Vendruscolo, PGB, PNAS **118**, e2012423118 (2021).*

FAQ

- Why does shooting work?
  - Stable states are attractors of the trajectories, while molecular chaos cause paths to diverge quickly
- How many trajectories?
  - About 10^3 trajectories per ensemble should suffice, but more is better.
- How long do the paths need to be?
  - Long enough to be able to relax to the stable states (about τ_{mol}). More quantitative measures based on correlation functions.
- How does it scale?
  - Just like MD, linear in time, linear in system size. Larger systems might need longer time.
- How to create initial path?
  - Many ways: pulling, high T, interpolation, metadynamics, committor analysis, TIS, etc
- What requirements do stable state definitions have?
  - Should distinguish states, but also representative. Path should quickly find state, but not with a false positive.
- What about other types of dynamics?
  - All eqs of motion, e.g. *ab initio* MD, Langevin, dynamic MC can be used.
- What about multiple channels and intermediates?
  - Should be included in sampling. Otherwise use e.g. RETIS, MSTIS or other advance path sampling methods.