Efficient sampling of molecular dynamics trajectories connecting arbitrary metastable states

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

ACMM Amsterdam Center for Multiscale Modeling

Transition path sampling

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

ACMM Amsterdam Center for Multiscale Modeling

Rare events

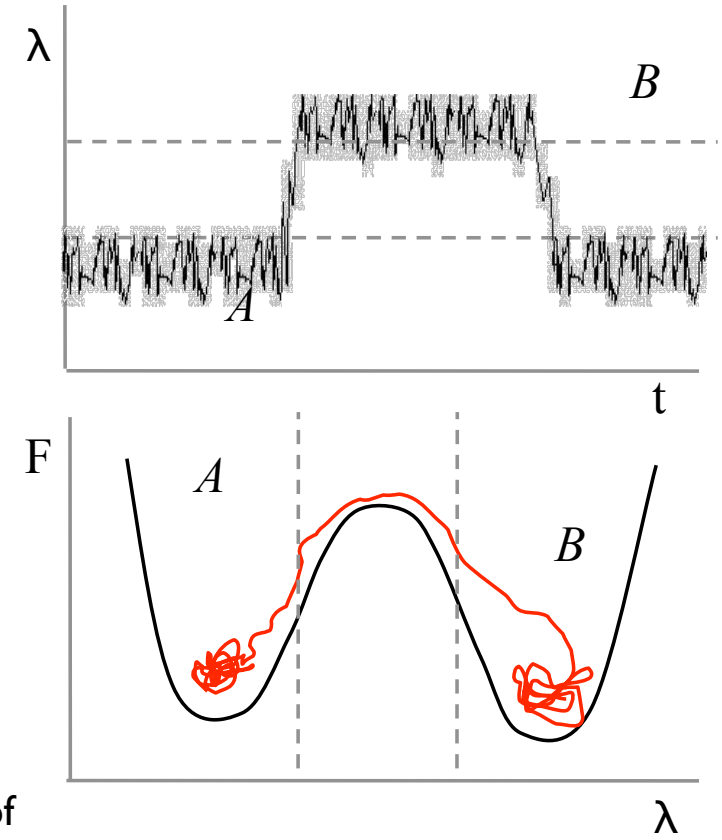
Interesting transitions in complex fluids

- solution chemistry
- phase transitions
- protein folding
- enzymatic reactions
- nucleation
- complex surface reaction
- membrane fusion

These reactions happen on a long time scale compared to the molecular timescale (eg solvent motion)

dominated by collective, rare events: straightforward MD is unpractical

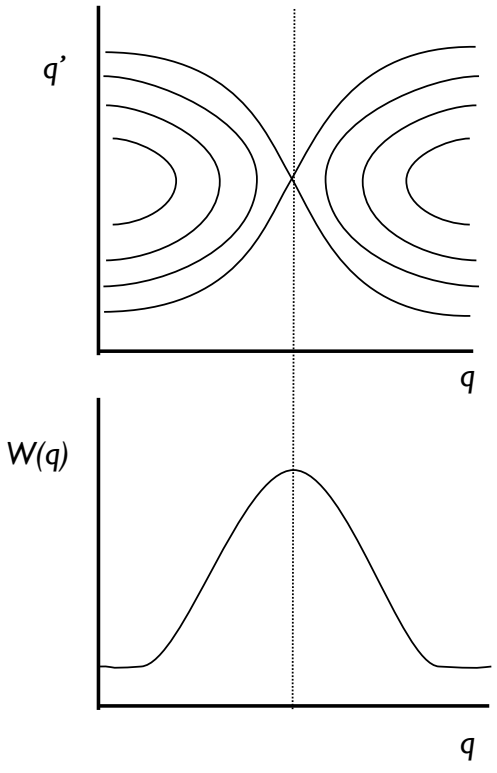
Usual tactics: compute free energy as a function of order parameter λ

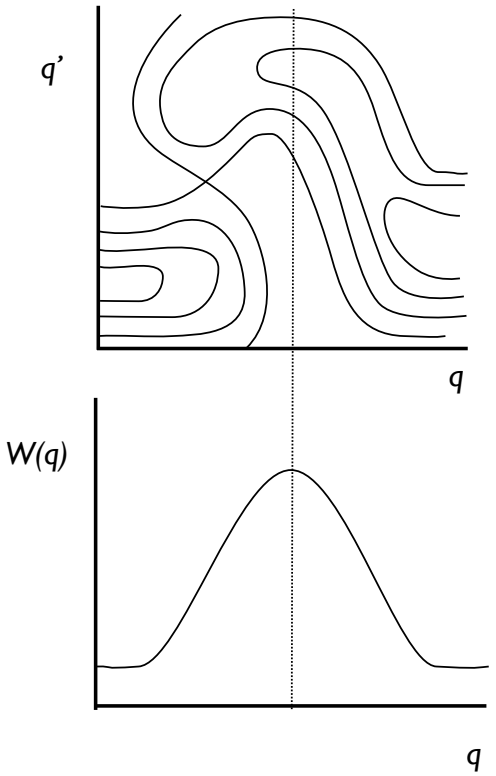


Breakdown of biased sampling

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

But if RC is not correct, all these might be wrong!



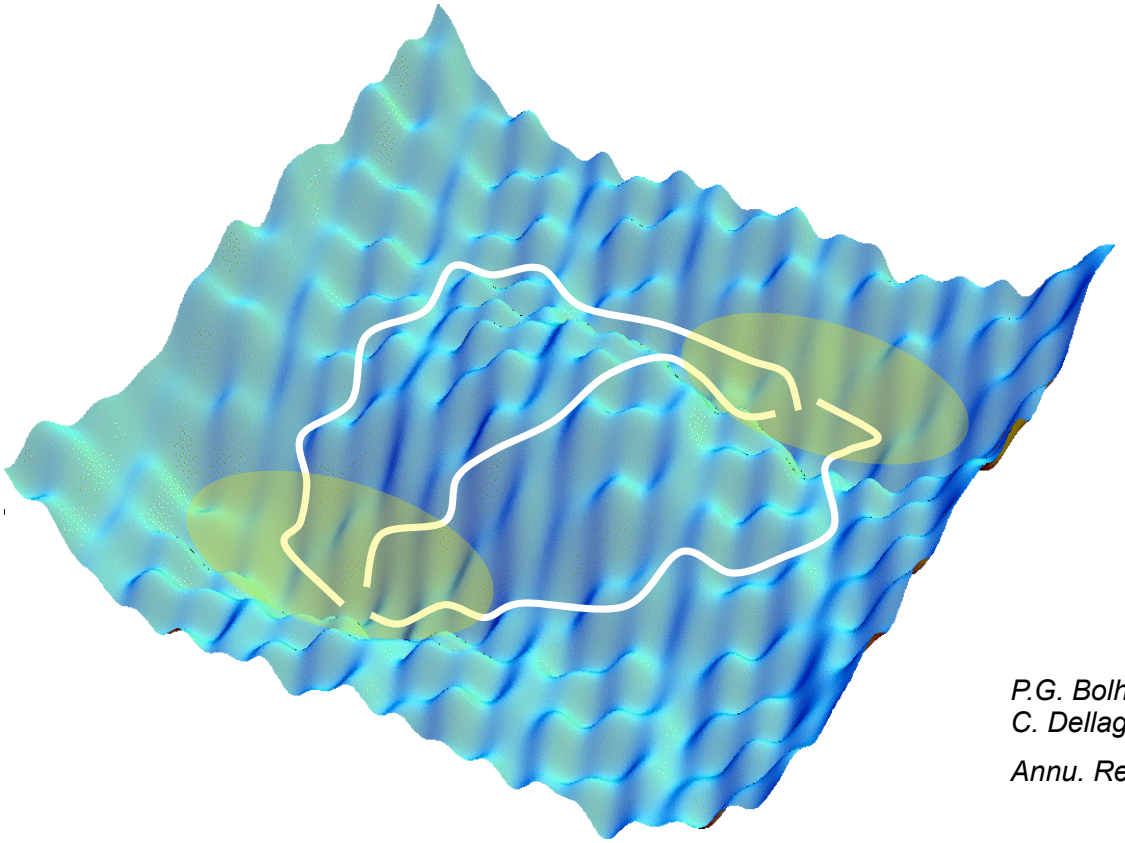


$$W(q) = -kT \ln \int dq' \exp\{-\beta E(q,q')\}$$

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

Transition path sampling

Transition path sampling



P.G. Bolhuis, D. Chandler,
C. Dellago, P.L. Geissler

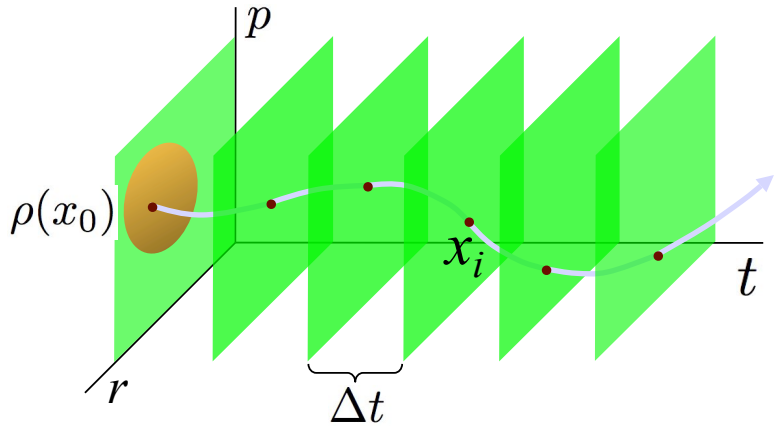
Annu. Rev. Phys. Chem 2002

Importance sampling of the path ensemble:
all trajectories that lead over barrier and connect stable states.

Path probability density

System consisting of N particles in 3D

$$x = \{r_1, r_2 \ldots r_N; p_1, p_2 \ldots p_N\} \in \mathbb{R}^{6N}$$



Discrete representation

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

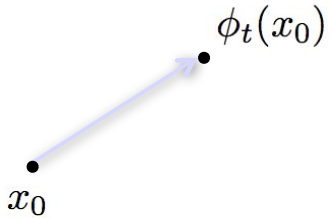
$$\mathcal{P}[\mathbf{x}(L)] = \rho(x_o) \prod_{i=0}^{L-1} p(x_i \to x_{i+1})$$

Can be defined for deterministic and stochastic dynamics

Newtonian (Hamiltonian) dynamics:

$$\dot{r} = \frac{\partial \mathcal{H}(r,p)}{\partial p}$$

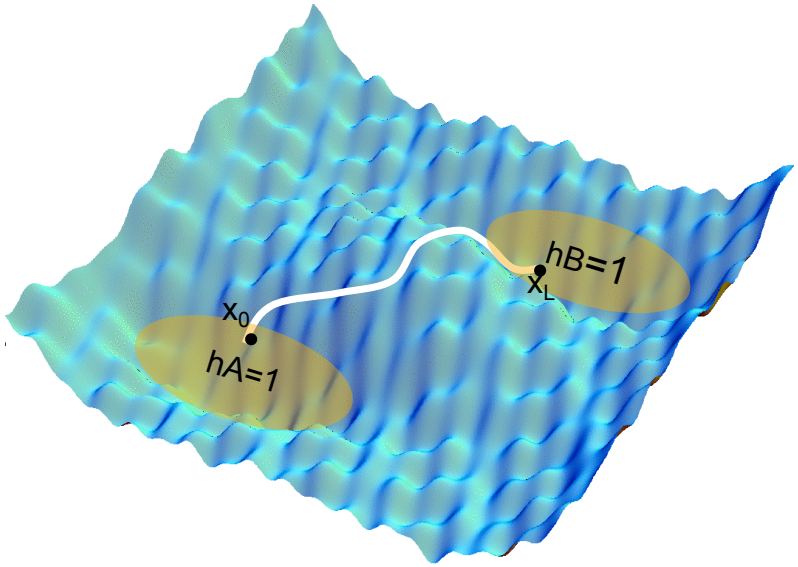
$$\dot{p} = -\frac{\partial \mathcal{H}(r,p)}{\partial r}$$



$$p(x_i \to x_{i+1}) = \delta\left[x_{i+1} - \phi_{\Delta t}(x_i)\right]$$

Canonical initial conditions $\rho(x) = \exp\{-\beta \mathcal{H}(x)\}/Q$ $Q(\beta) = \int dx \exp\{-\beta \mathcal{H}(x)\}$

Transition path probability density



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

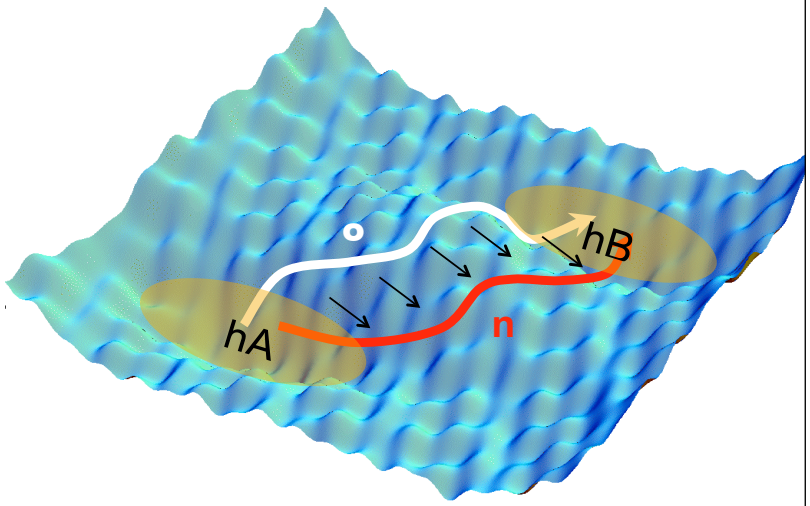
$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Path probability distribution

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

$$Z_{AB}(L) \equiv \int \mathcal{D}\mathbf{x}(L)h_A(x_0)\mathcal{P}[\mathbf{x};L]h_B(x_L)$$

$$\int \mathcal{D}\mathbf{x}(L) = \int \ldots \int dx_0 dx_1 \ldots dx_L$$

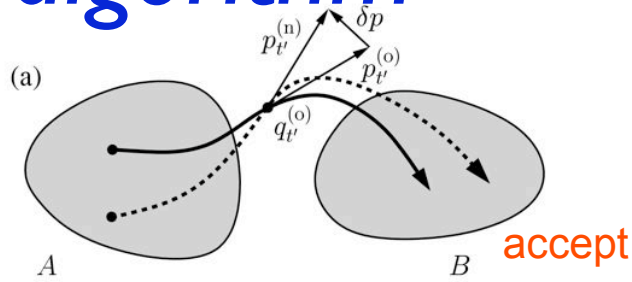


Importance sampling using Metropolis rule :

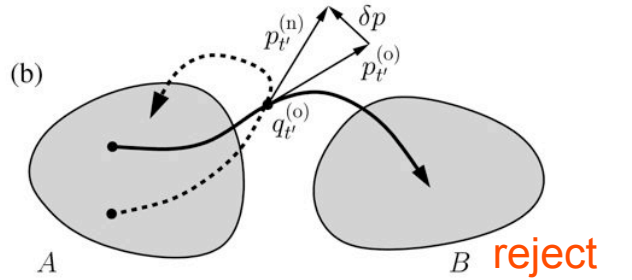
$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A[x_0^{(n)}]h_B[x_L^{(n)}] \min\left[1, \frac{\mathcal{P}[\mathbf{x}^{(n)}]\mathcal{P}_{gen}[\mathbf{x}^{(n)} \to \mathbf{x}^{(o)}]}{\mathcal{P}[\mathbf{x}^{(o)}]\mathcal{P}_{gen}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}]}\right].$$

Standard shooting algorithm

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

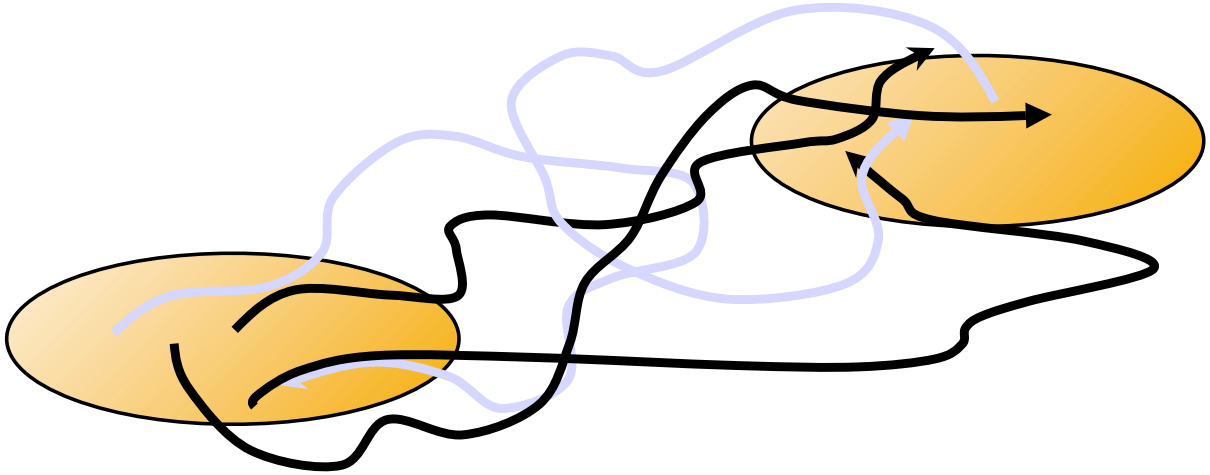


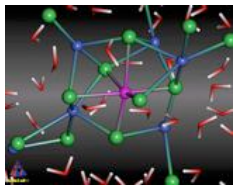
accept

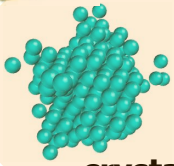


reject

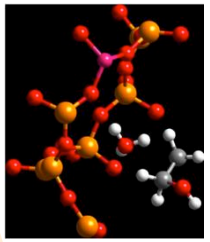
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_T^{(n)})$$



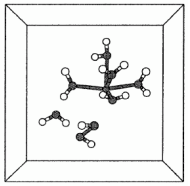




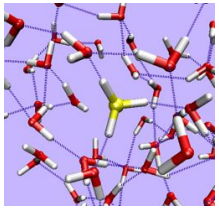
crystallisation

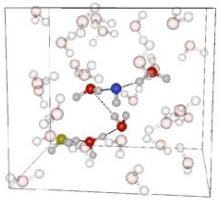


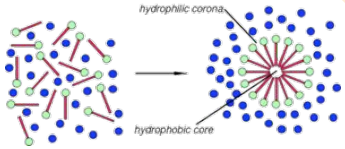
catalysis

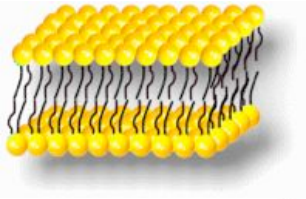


reactions

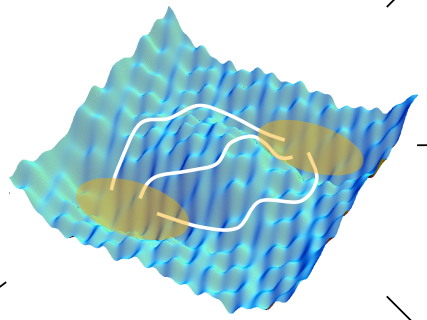


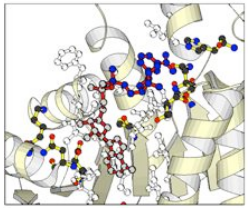




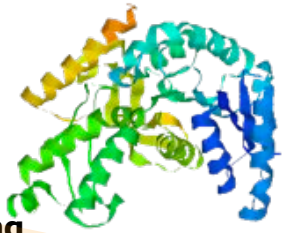


complex fluids

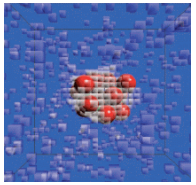




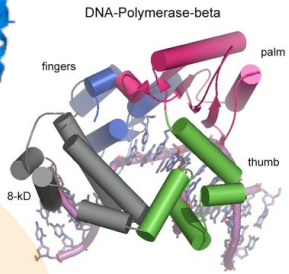
enzyme reactions



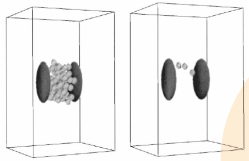
solvent effects

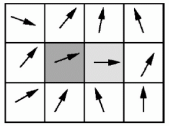


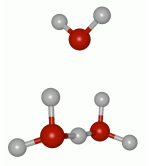
folding & binding

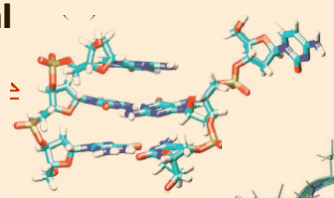


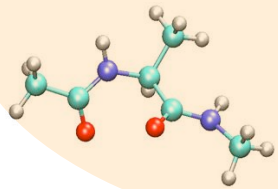
conformational change

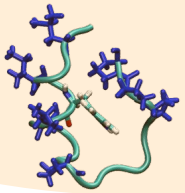


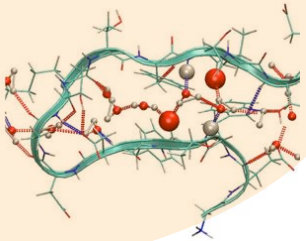






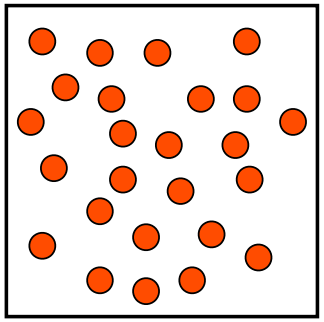




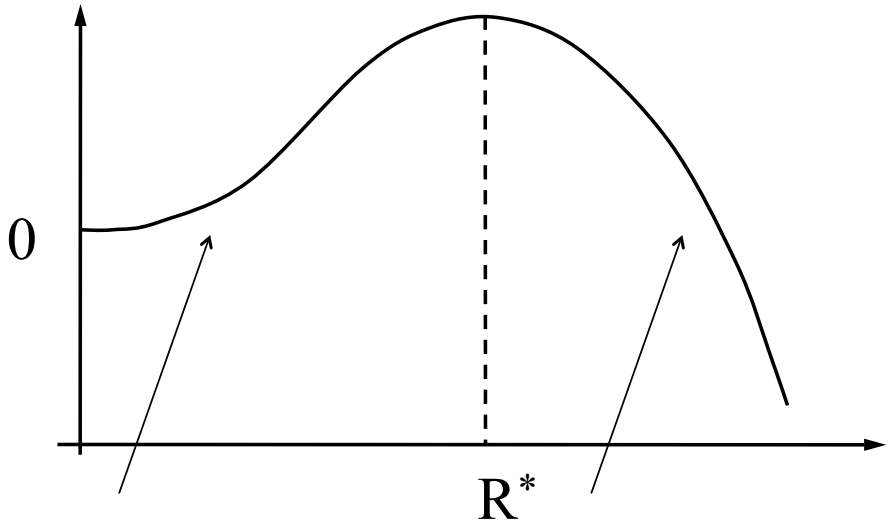


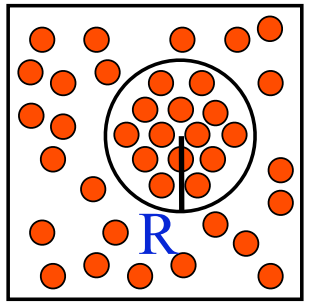
Warwick 2009

Classical nucleation



Liquid





Crystal nucleus

surface bulk

$$\Delta G = 4\pi R^2\gamma - \frac{4}{3}\pi R^3 \rho \Delta\mu_{ls}$$

–How does the crystal form?

–What is the structure of the critical nucleus

–Is classical nucleation theory correct?

- What is the barrier?
- Rate constant

γ : surface tension

$\Delta\mu$: chem. pot difference

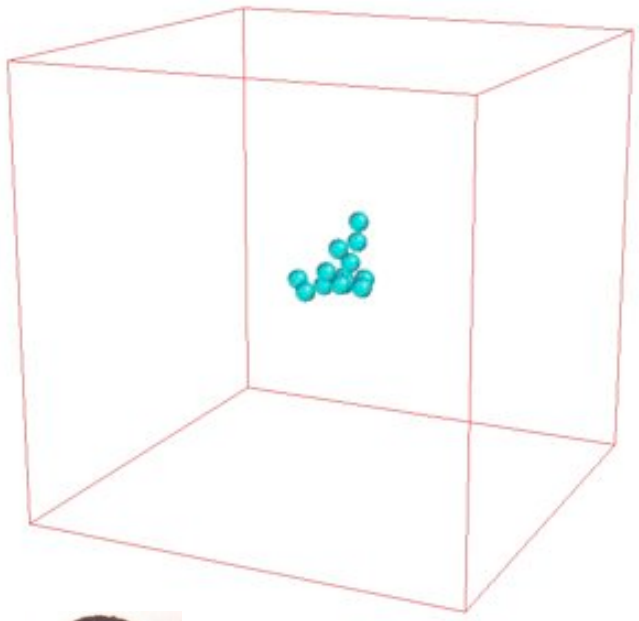
ρ: density

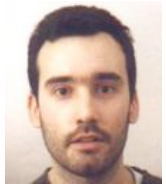
Path sampling of nucleation

TIS in NPH ensemble, as density and temperature change

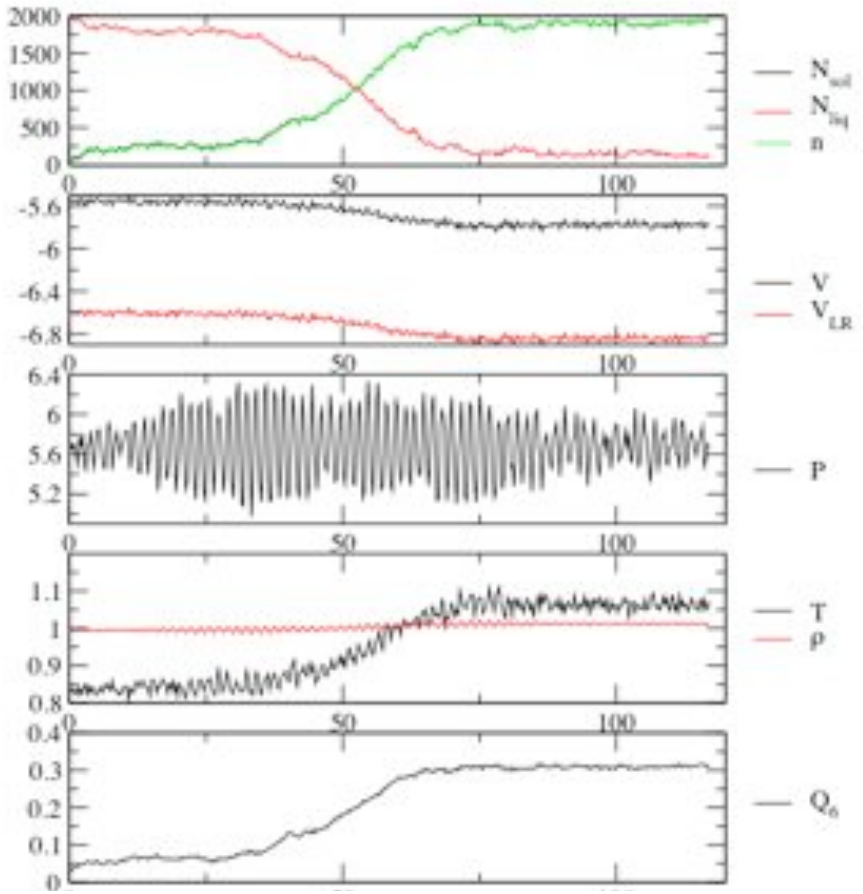
N=10000, P=5.68 H=1.41 (25 % undercooling)

order parameter is number of particles in solid cluster n (based on bond order q6)

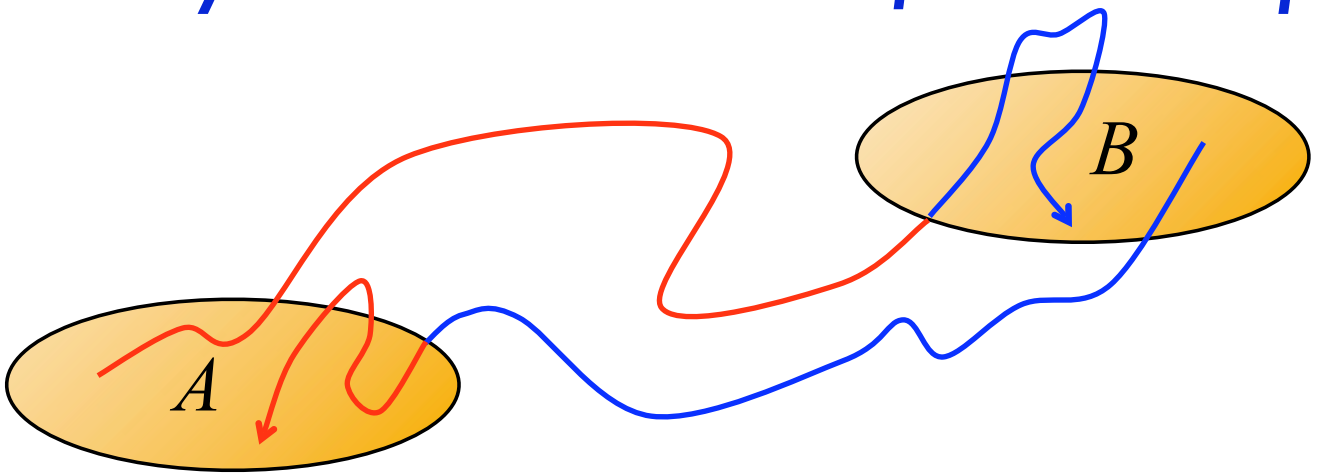




Daniele Moroni



Rates by transition interface sampling



Overall states in phase space:

$\mathcal{A}$ going *back in time* **A** reached first

$\mathcal{B}$ going *back in time* **B** reached first

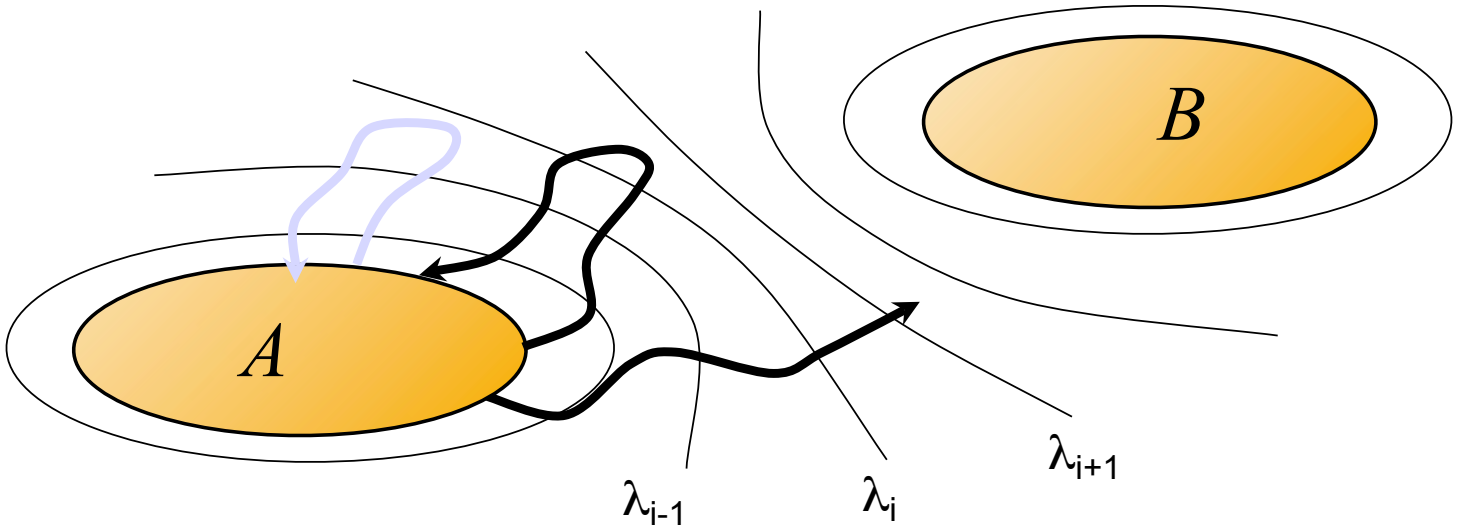
$$\mathcal{C}(t) \equiv \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t) \rangle}{\langle h_{\mathcal{A}} \rangle}$$

$$k_{AB} = \frac{\langle h_{\mathcal{A}}(x_0) \dot{h}_{\mathcal{B}}(x_0) \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle}$$

T. S. van Erp, D. Moroni and P. G. Bolhuis, J. Chem. Phys. **118** , 7762 (2003)

T. S. van Erp and P. G. Bolhuis, J. Comp. Phys. **205**, 157 (2005)

Rates by Transition interface sampling

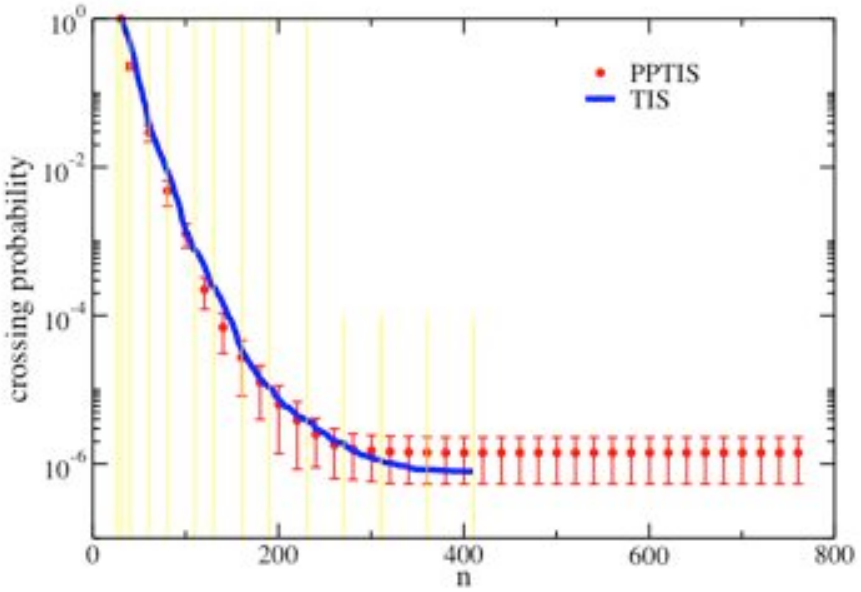


$P_A(\lambda_{i+1} \mid \lambda_i)$ = probability that path crossing i for first time after leaving A reaches i+1 before A

$$k_{AB} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} P_A(\lambda_B | \lambda_A) = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

flux $\frac{\langle \phi_A \rangle}{\langle h_A \rangle} = \frac{1}{\Delta t} \frac{N_c^+}{N_{\mathrm{MD}}}$

Sample paths with
- -shooting
- -time reversal moves for AA paths

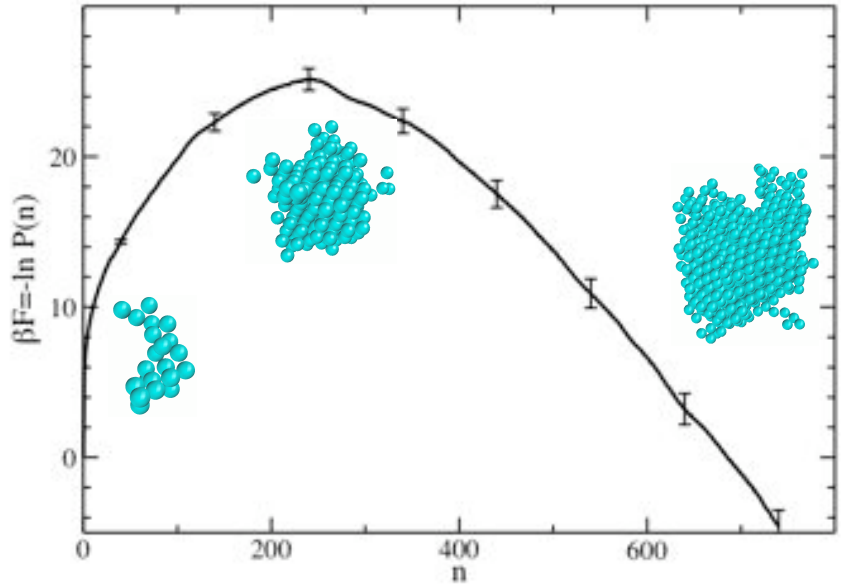


Nucleation rate

Order parameter n = number of solid-like particles in crystal nucleus

$$k_{AB} = (1.0 \pm 0.8) \times 10^{-6}$$

Moroni, ten Wolde, Bolhuis, PRL, 2005

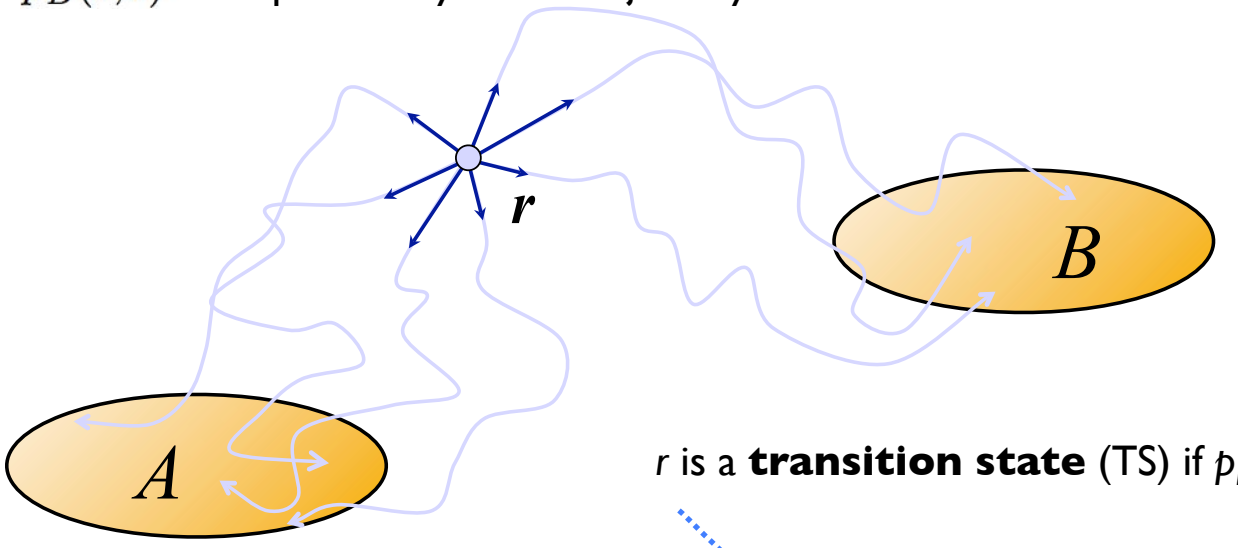


Free energy follows directly
Moroni, van Erp, Bolhuis, PRE, 2005

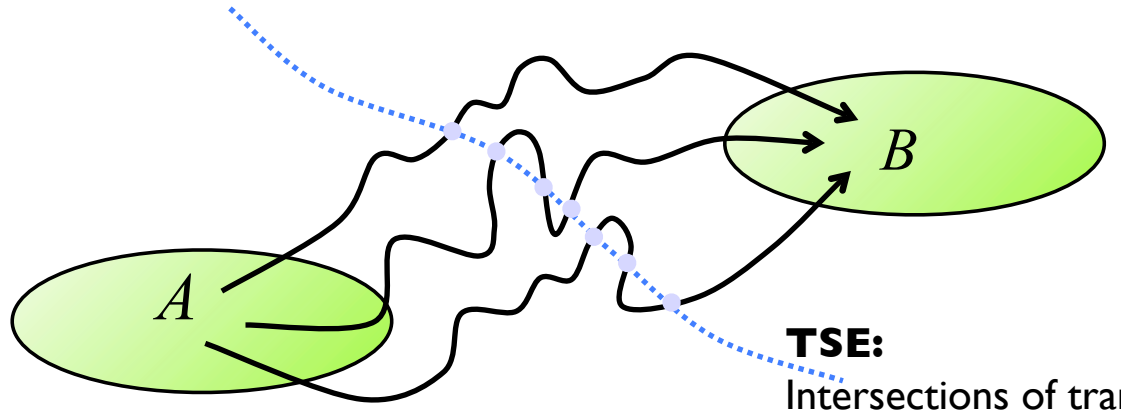
Structural analysis?

Transition states by committor

$p_B(r,t) =$ probability that a trajectory initiated at *r* relaxes into *B*



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

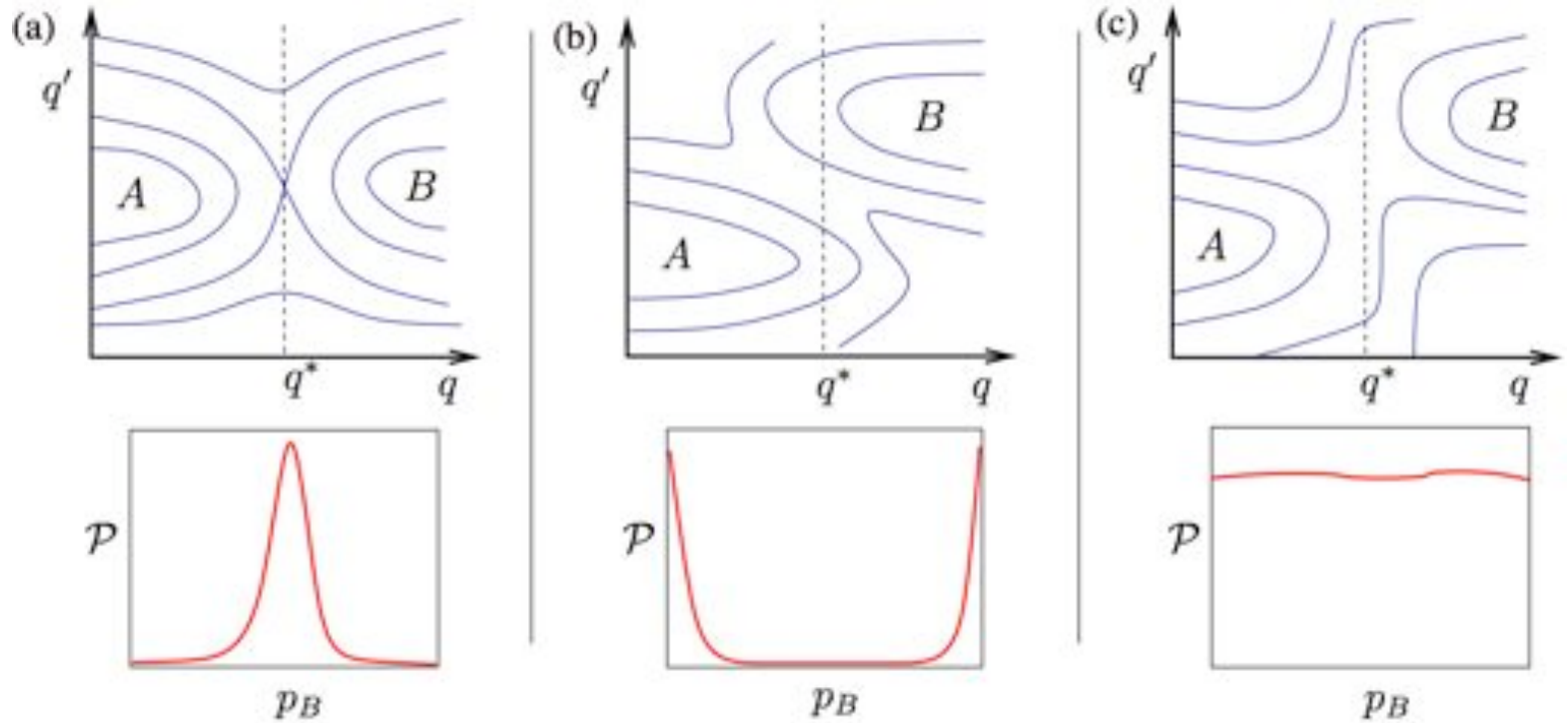


TSE:
Intersections of transition pathways with the p_B=1/2 surface

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)

Committor analysis

An attempt to find out the reaction coordinate

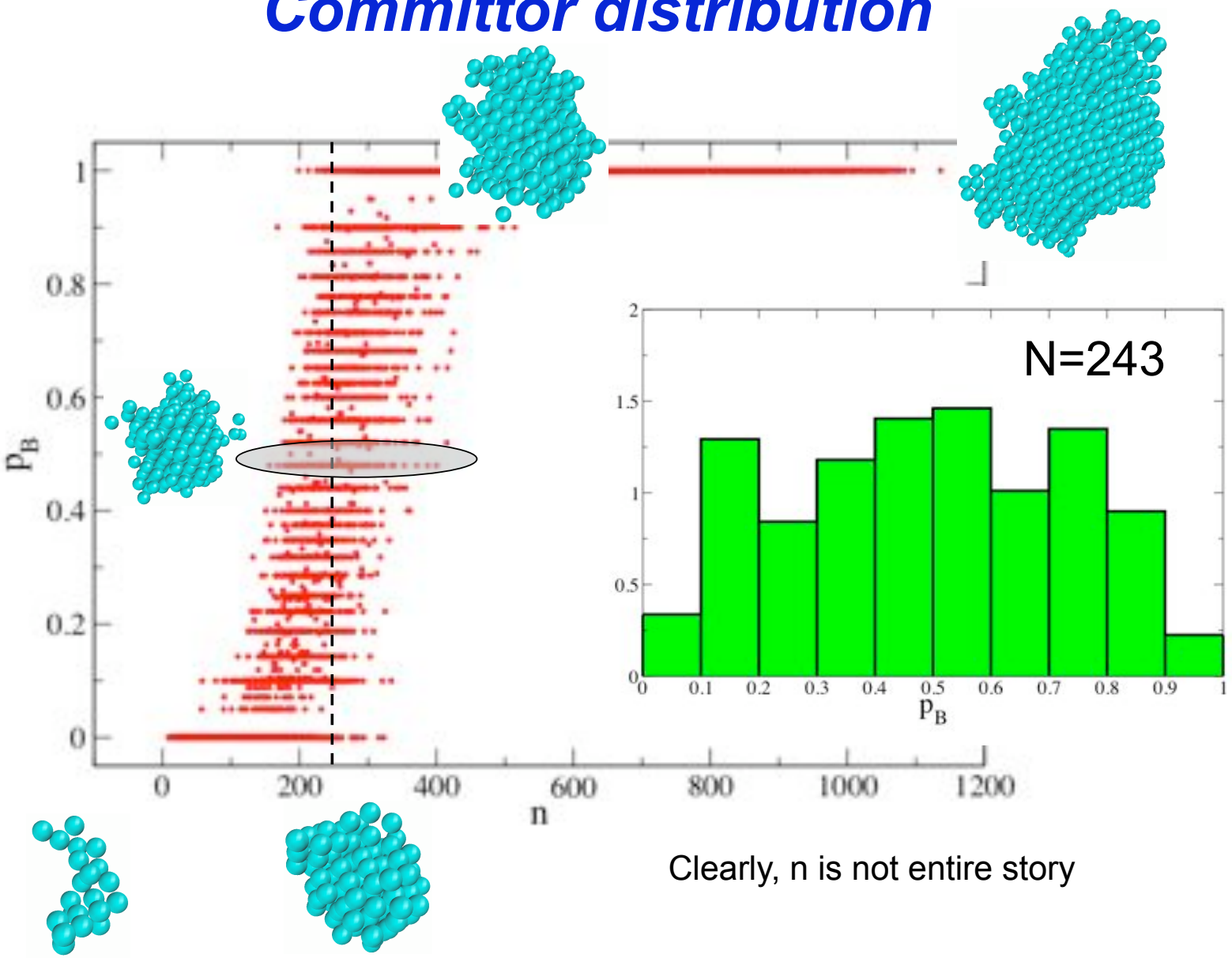


Good RC

Bad RC

analysis very expensive: requires pB histogram for every q

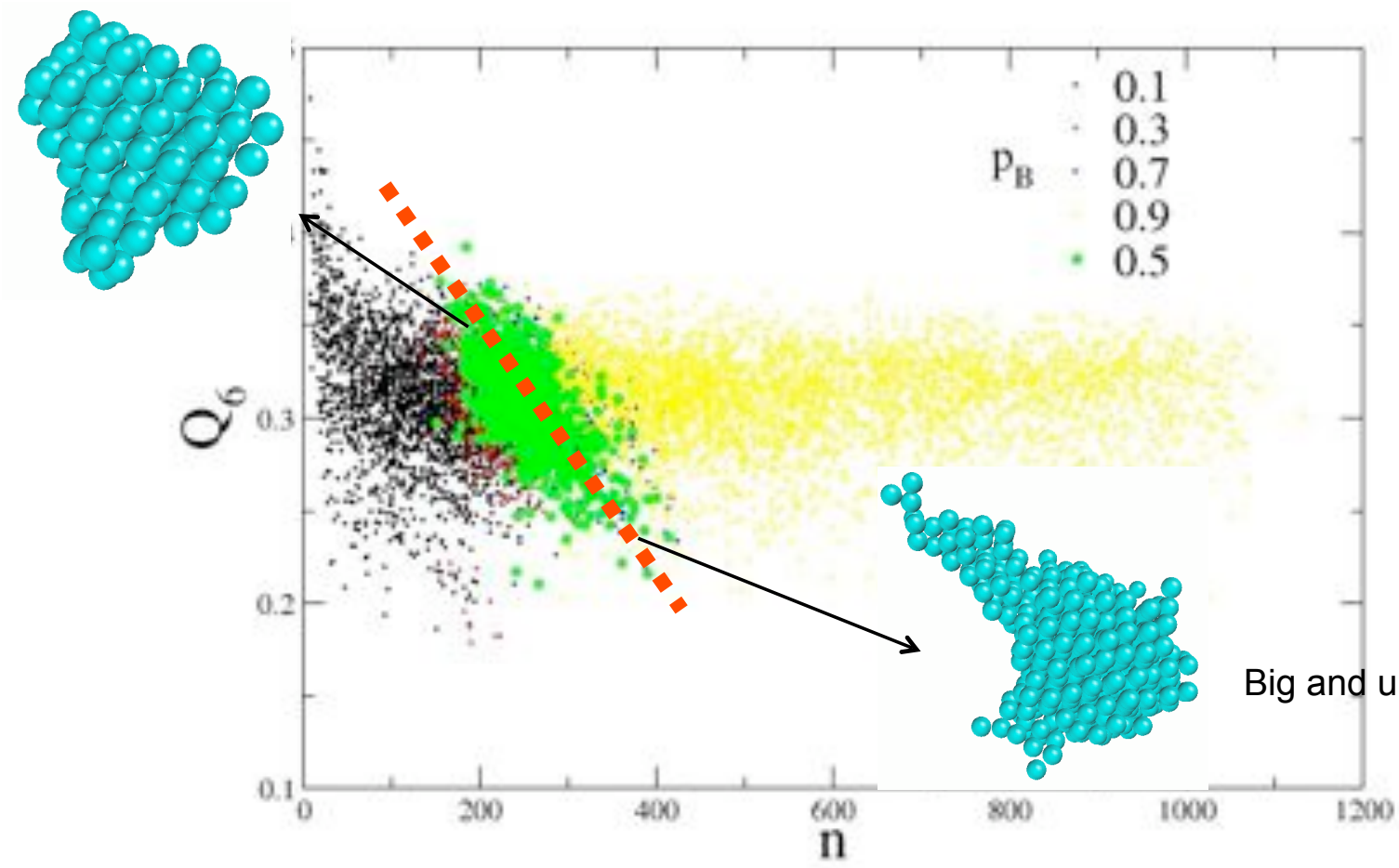
Committor distribution



Clearly, n is not entire story

Structure in the TSE

Small and structured



Big and unstructured

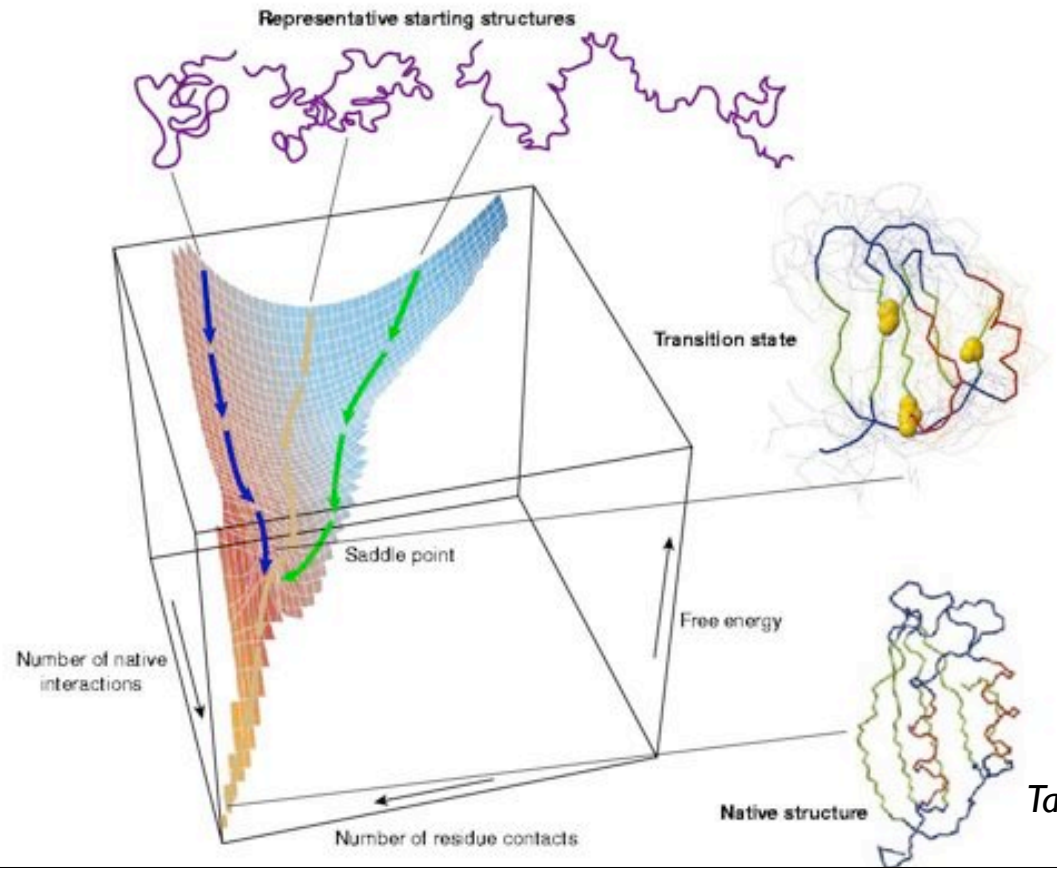
Conclusion crystallization

- Crystal nucleation very diffusive.
- Interplay between size and structure in critical nucleus.
  - combination of n and Q_6 better reaction coordinate
- Many crystal nucleation pathways
  - If critical cluster is small, it is more FCC structured
  - If critical cluster is larger, it is less FCC structured.
- Large BCC content: Ostwalds step rule.
- However, exact reaction coordinate still not completely known

Moroni, ten Wolde, Bolhuis PRL 2005

How do proteins find their native state?

- Guided by free energy landscape
  - how is this related to folding kinetics?
  - mechanisms important to understand misfolding (Alzheimer,CJD, etc)



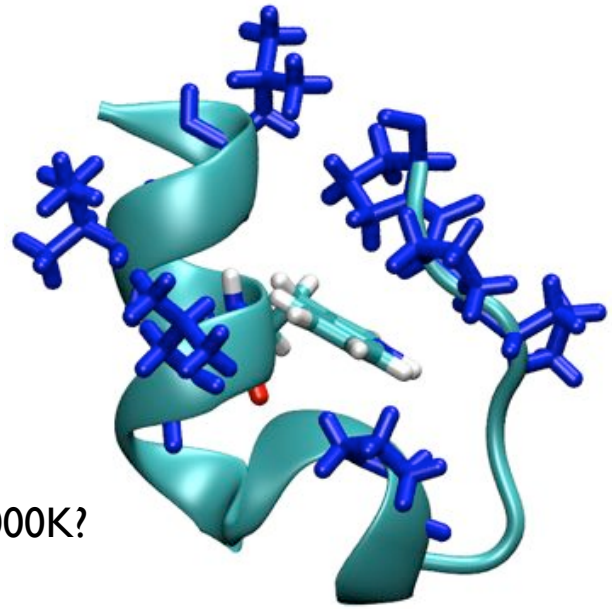
Taken from Dobson, Nature, 2003

Folding of Trp-cage

20-residue protein NLYIQ WLKDG GPSSG RPPPS

2-state folder, experimental rate 4 μs

(*Andersen et al, Nature 2002, Zhou et al. PNAS 2004, others*)

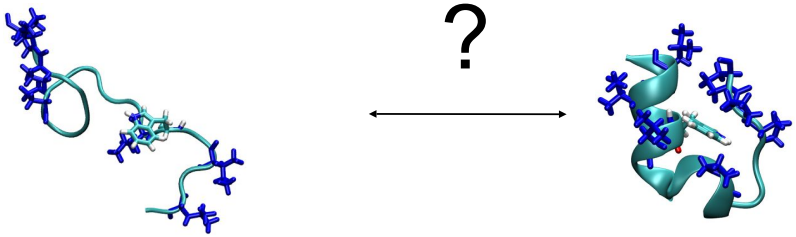


System:

1L2Y in 2800 SPC waters

OPLSAA, PME, Nose-Hoover, GROMACS

What is folding mechanism and kinetics in explicit water at 3000K?



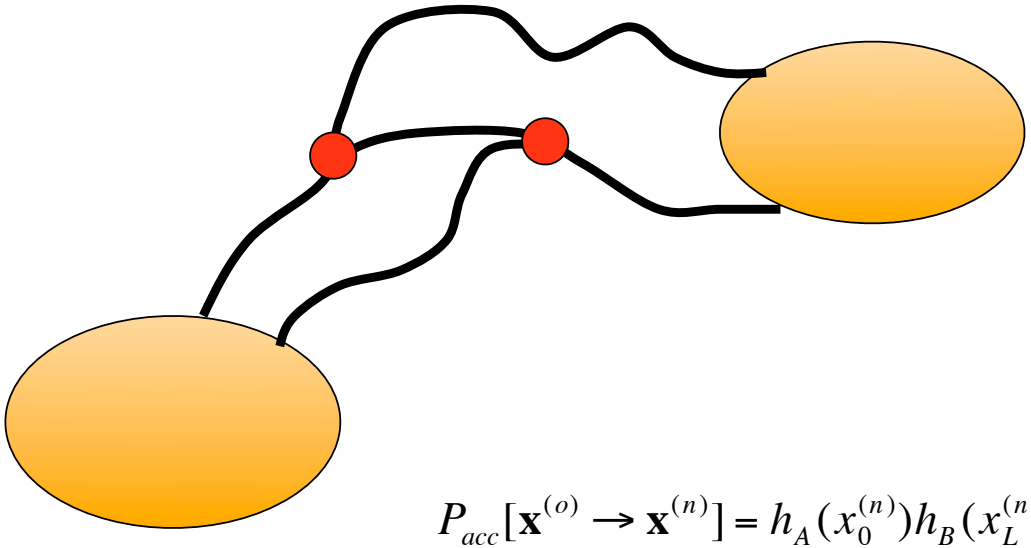
Strategy:

- Stable states by PT/REM
- Mechanism by path sampling
- rate by TIS

Jarek Juraszek

Flexible one way shooting

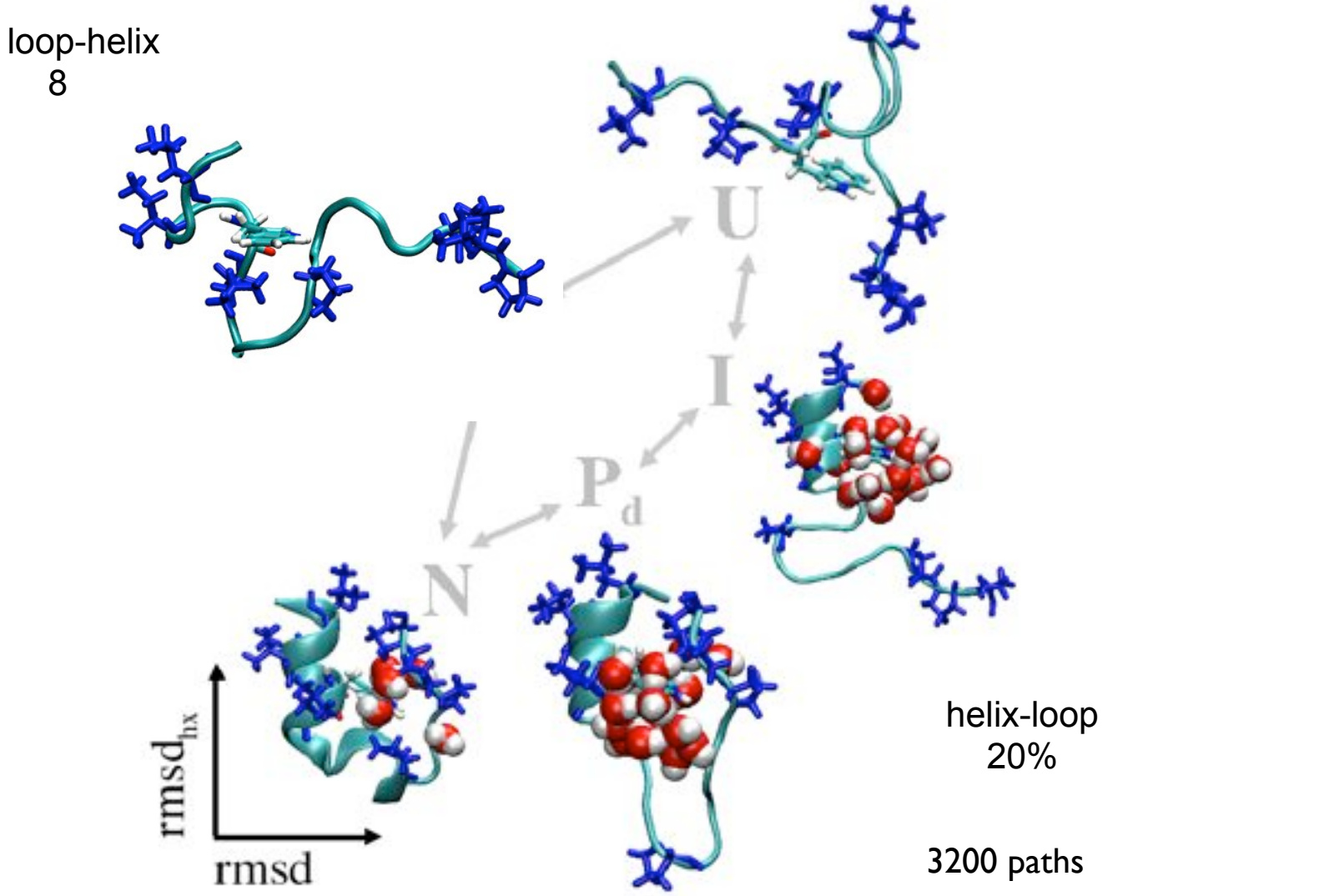
- Variable length shooting (PGB 2003, Juraszek & PGB 2006)
  - Choose new shooting point randomly from old path
  - Integrate in one direction until one stable states is reached



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

- higher acceptance, better convergence for diffusive transitions and long pathways
- requires some stochastic dynamics

Parallel folding pathways



N-L rates for Trp-cage

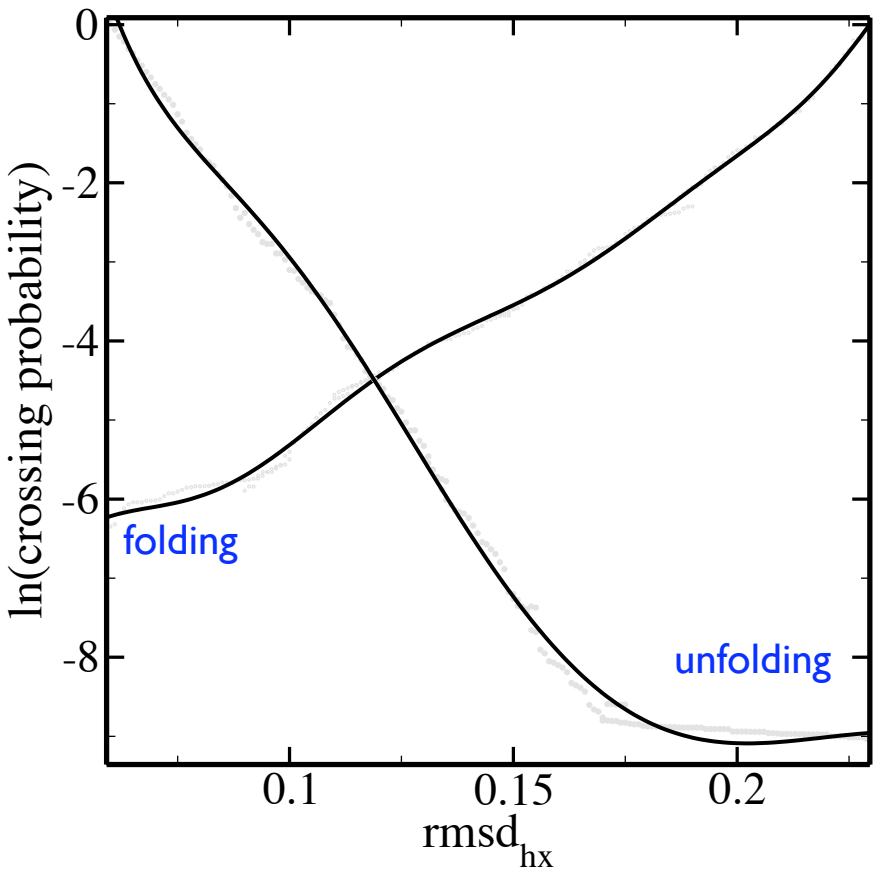
TIS, 6 interfaces

ln P_{unf} = -9
Φ_A (λ =0.06)= 6.6 ns^{-1}
k_{unf} = 0.8 μs^{-1}

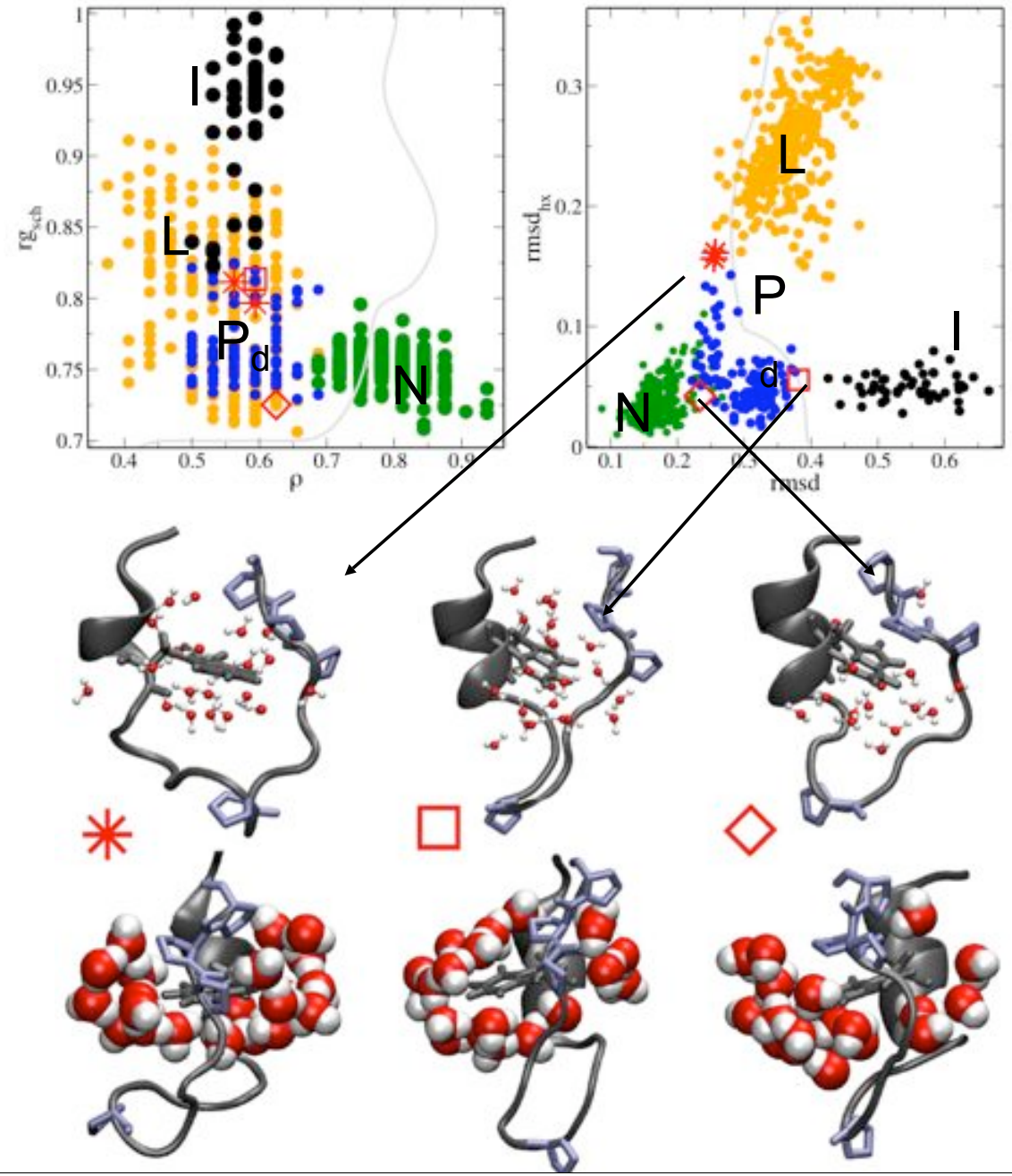
Exp: k_{unf} = 0.08 μs^{-1}

ln P_{fol} = -6.3
Φ_A (λ=0.23) = 1 ns^{-1}
k_{fol} = 2.5 μs^{-1}
corrected k_{fol} = 0.2 μs^{-1}

Exp. k_{fol} = 0.24 μs^{-1}



TS can fall inside stable state in FE landscape



Likelihood maximization

- Each TPS shot can be seen as a committor shot. Based on this look for best model of reaction coordinate *r*
- The probability p(TP|r) to be on a transition path provided we are at a structure x with rc r is (for diffusive dynamics)

$$p(TP|r) = 2p_B(r)(1 - p_B(r))$$

- Assume committor function to be

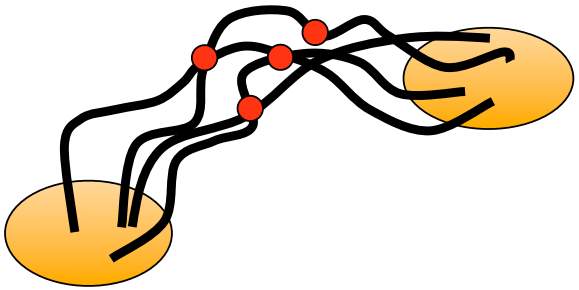
$$p_B(x) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(q(x)\right]$$

- parametrize r as linear combination of q

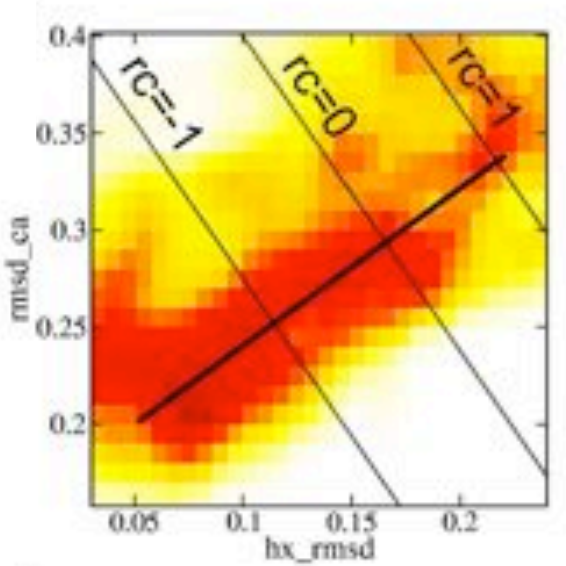
$$r(\mathbf{x}) = \sum_i \alpha_i q(\mathbf{x}) + \alpha_0$$

- best r is maximizing likelihood

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



Peters & Trout, JCP 125 054108(2006)



rc = -4.5 + 13 $rmsd_{hx}$ + 8 $rmsd_{ca}$

Summary Trp-cage

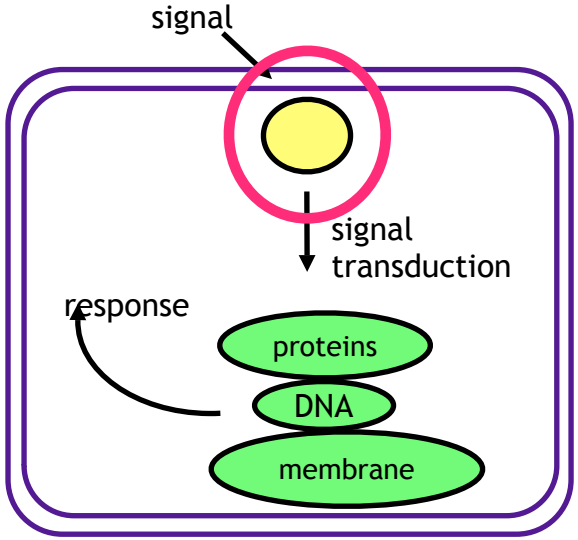
- TPS can sample all-atom folding pathways even for events with μs time scales
- Shows switching between mechanisms

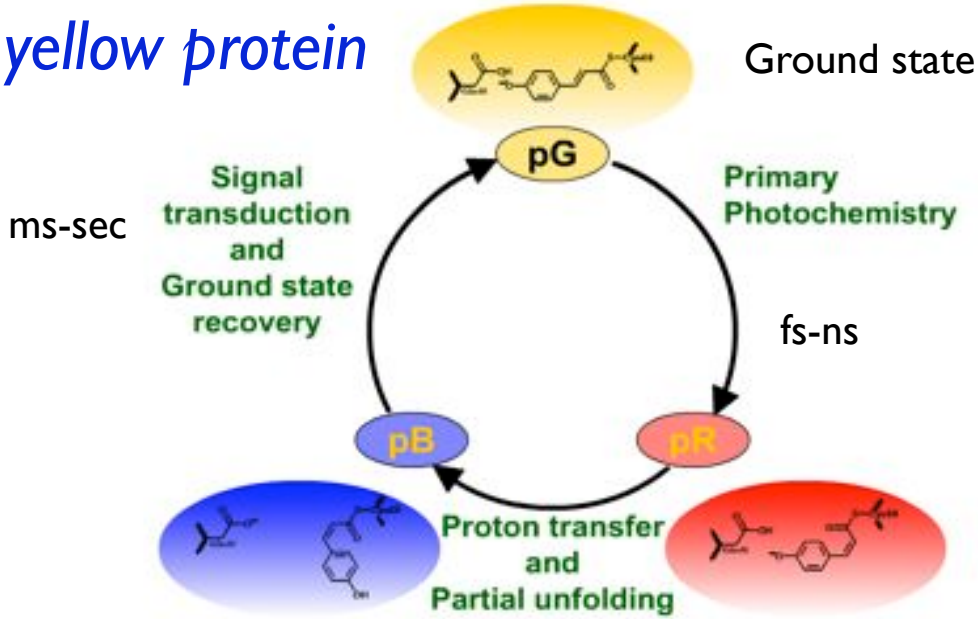
- Folding rate of Trp-cage compares to experiment, unfolding not

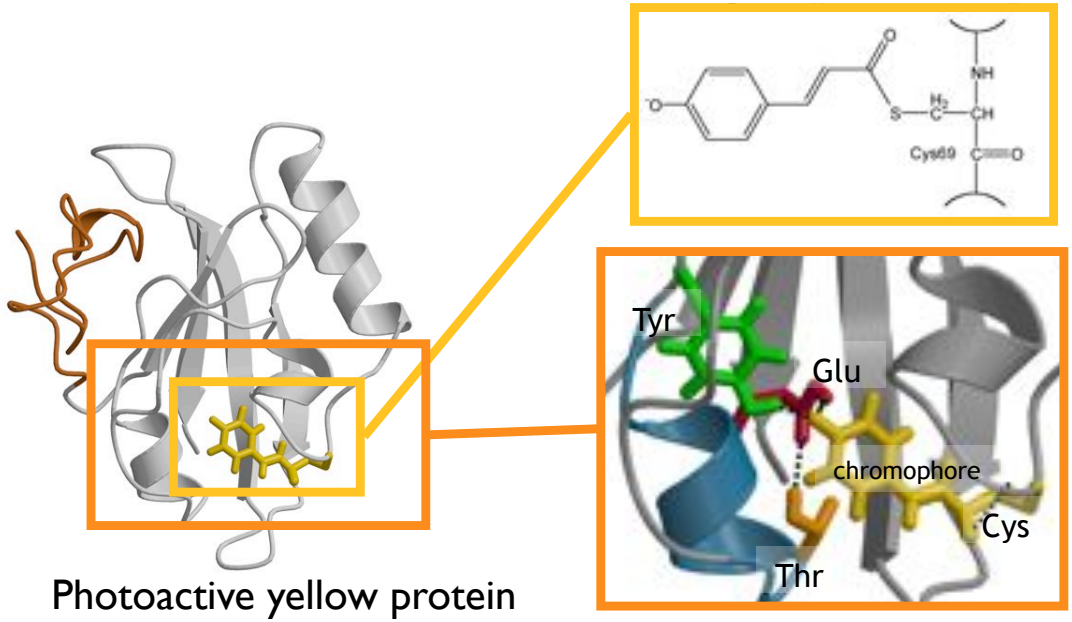
- Transition state ensemble (TSE) :
  - characterized by solvation
  - water expulsion is last step upon folding.
  - water dynamics probably no part of RC at TSE , water structure is.
  - does not always correspond with a FE landscape saddle

- Reaction coordinate involves secondary structure rmsd as well as global rmsd

Signalling proteins: Photoactive yellow protein







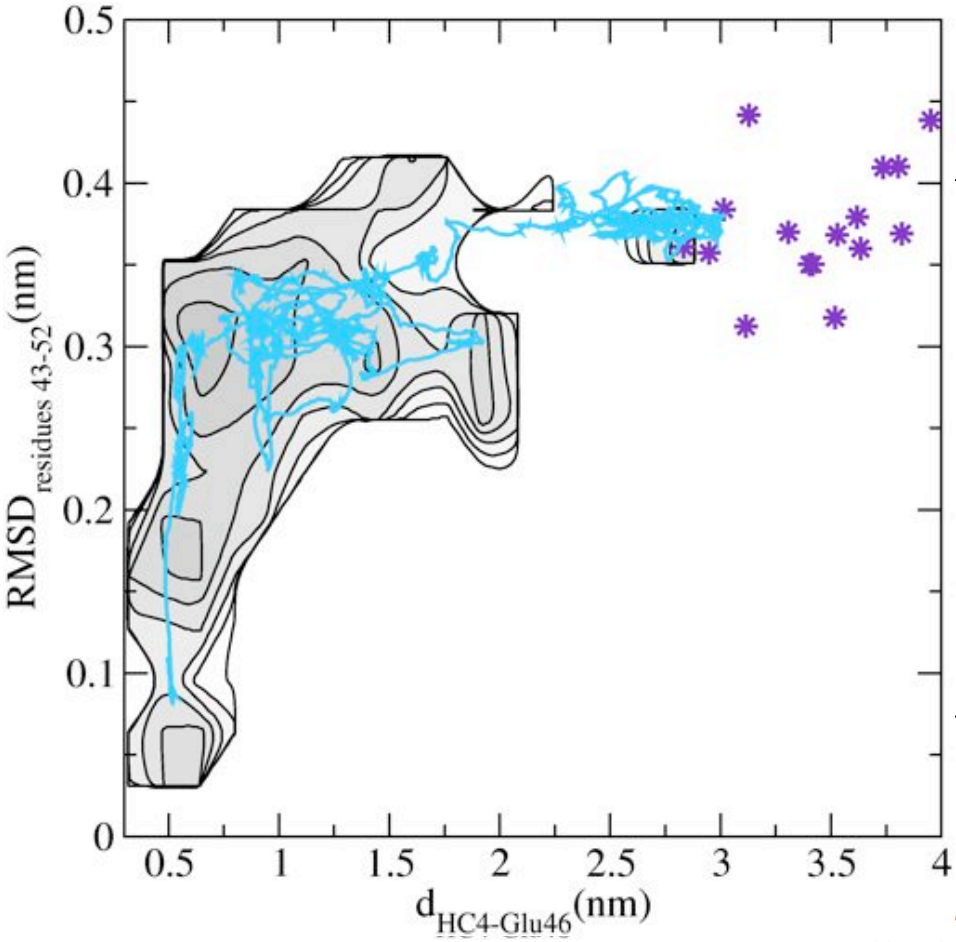
Photoactive yellow protein

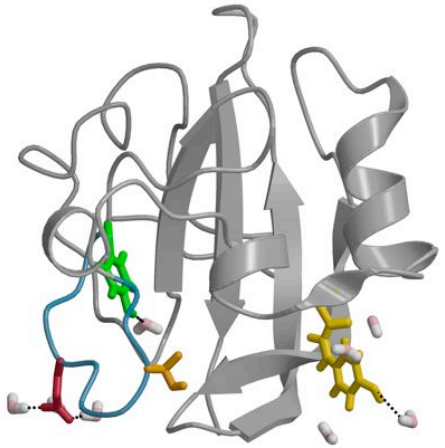
Question: What is the mechanism for amplifying signal?

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

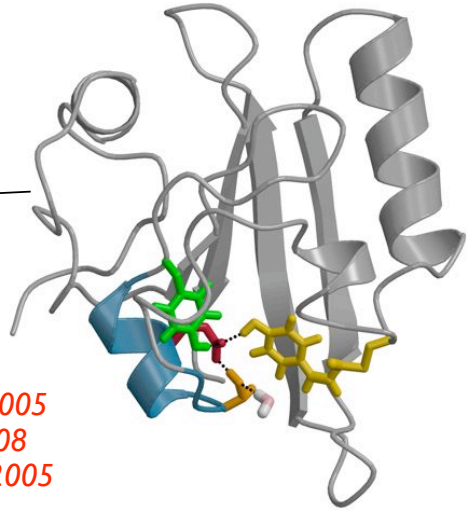
Partial unfolding: REMD free energy

Unfolding happens on sub-millisecond timescale
(1000 times longer is possible by direct MD)





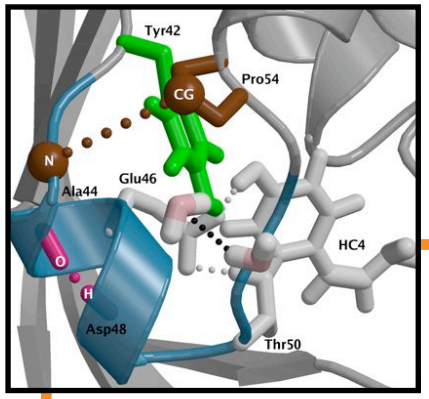
Helix 43-51 unfolded
Chromophore and Glu46 are solvent exposed
Agreement with experiment



et al. Biophys. J. 2005
Vreede et al. Proteins 2008
Bernard et al. Structure 2005

TPS of unfolding reaction

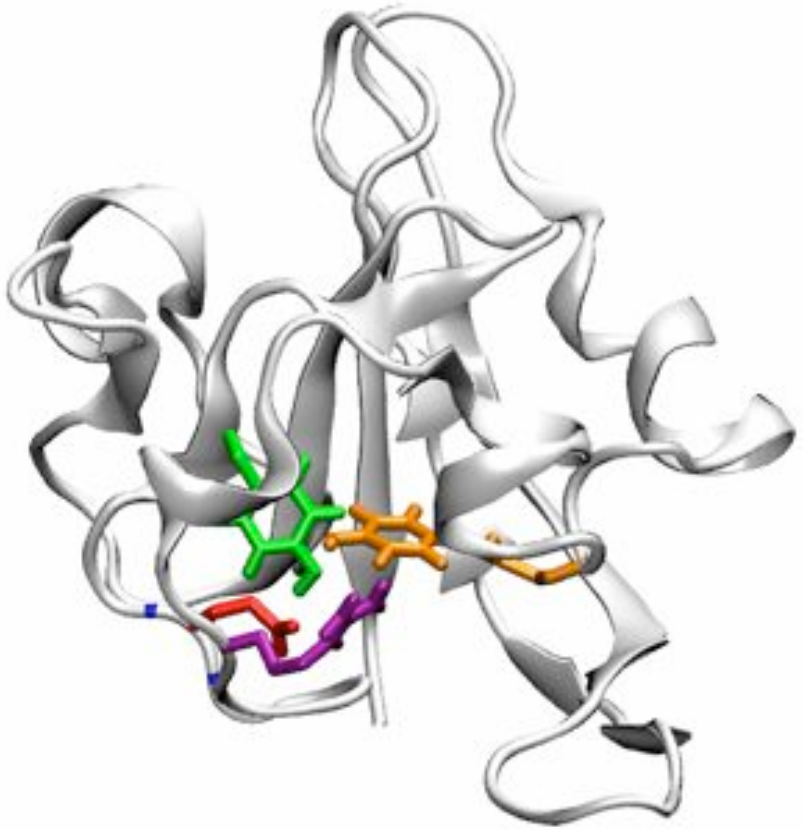
Solvent exposure Glu46 and chromophore

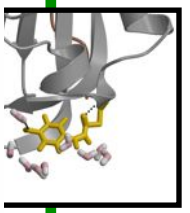


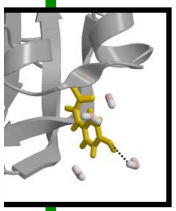
$I_{\alpha 3}$

folded

Unfol





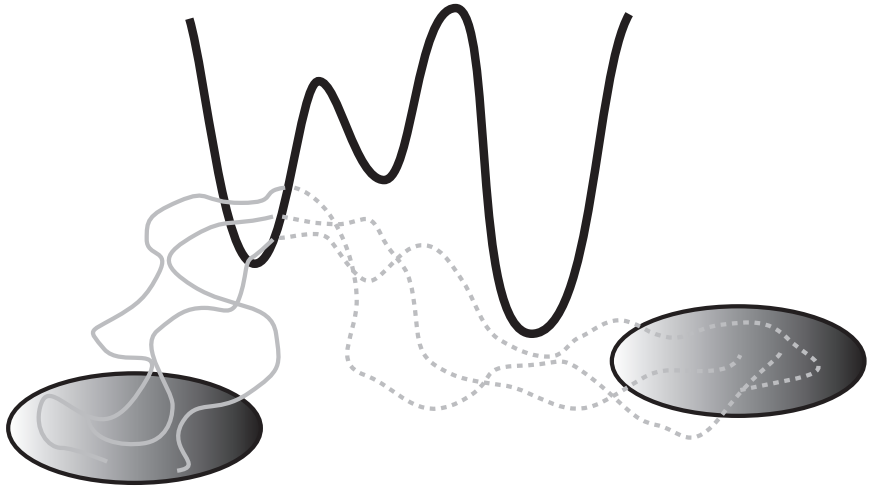


preparation

Solvent exposure of Glu46 and chromophore is governed by fluctuations in internal hydrogen bonds

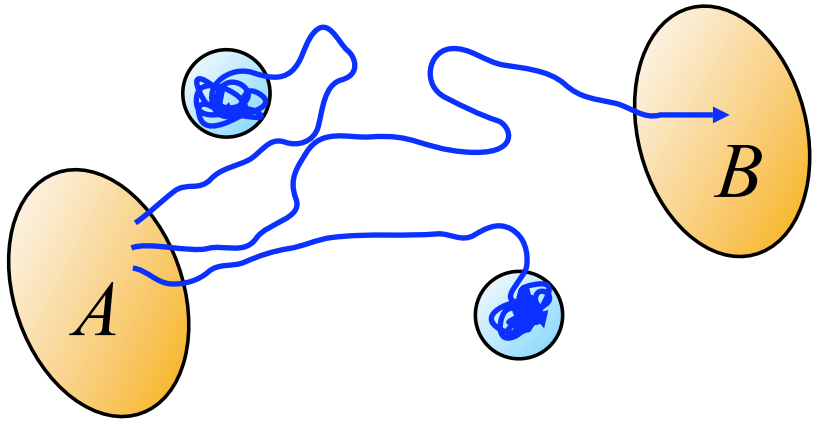
Challenges for path sampling

- Multiple channels
  - multiple channels are not sampled properly with shooting



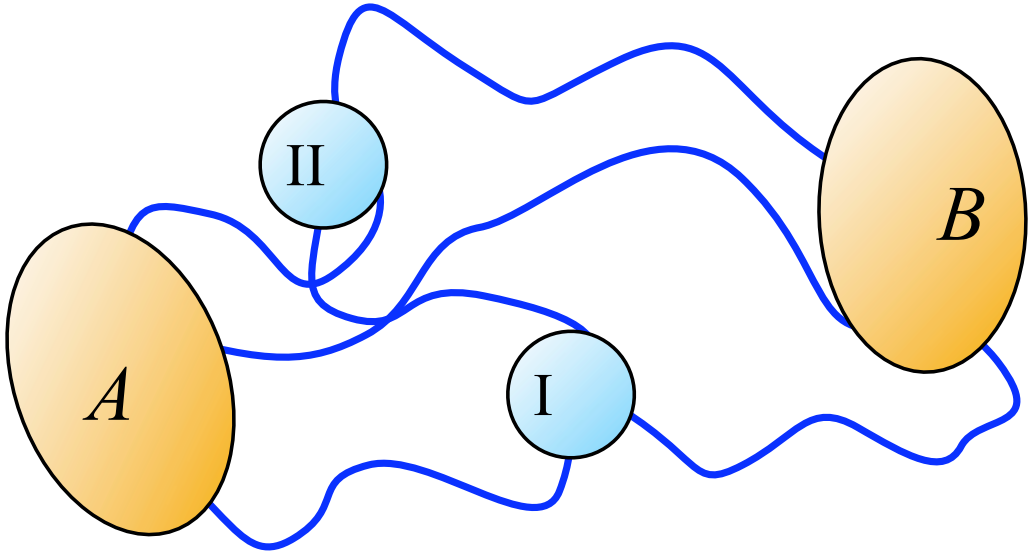
T.S. van Erp, PRL 98, 268301 (2007)
PGB, JCP 129,114108 (2008)

- Presence of intermediates
  - paths become very long because of intermediates



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

Multiple state path ensemble



- path ensemble: $\mathcal{P}_{\mathrm{MSTPS}} = \sum_{i,j\neq i} \mathcal{P}_{ij}\left[\mathbf{x}(L)\right]$

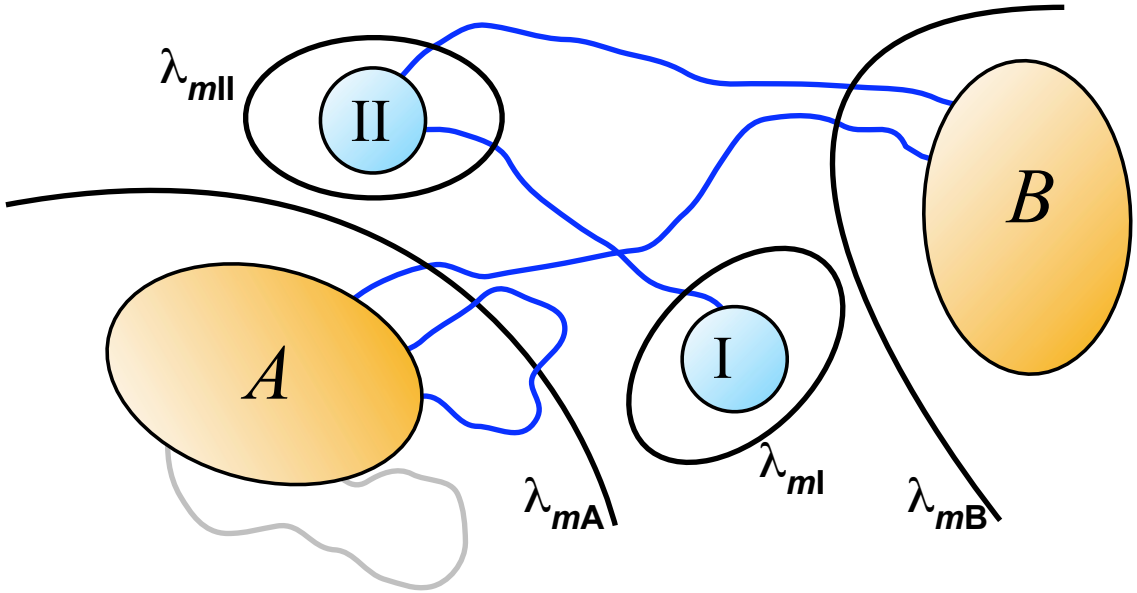
$$\mathcal{P}_{ij}\left[\mathbf{x}(L)\right] \equiv Z^{-1} \prod_k \bar{h}_k\left[\mathbf{x}(L)\right] h_i(x_0) \mathcal{P}\left[\mathbf{x}(L)\right] h_j(x_L)$$

- normalization:

$$Z \equiv \int \mathcal{D}\mathbf{x}(L) \mathcal{P}\left[\mathbf{x}(L)\right] \prod_k \bar{h}_k\left[\mathbf{x}(L)\right] \sum_{i,j\neq i} h_i(x_0) h_j(x_L)$$

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

Multiple state TIS



path ensemble: $\mathcal{P}_{\mathrm{MSTIS}} = \sum_{i,j} \mathcal{P}_{ij}\left[\mathbf{x}(L)\right]$

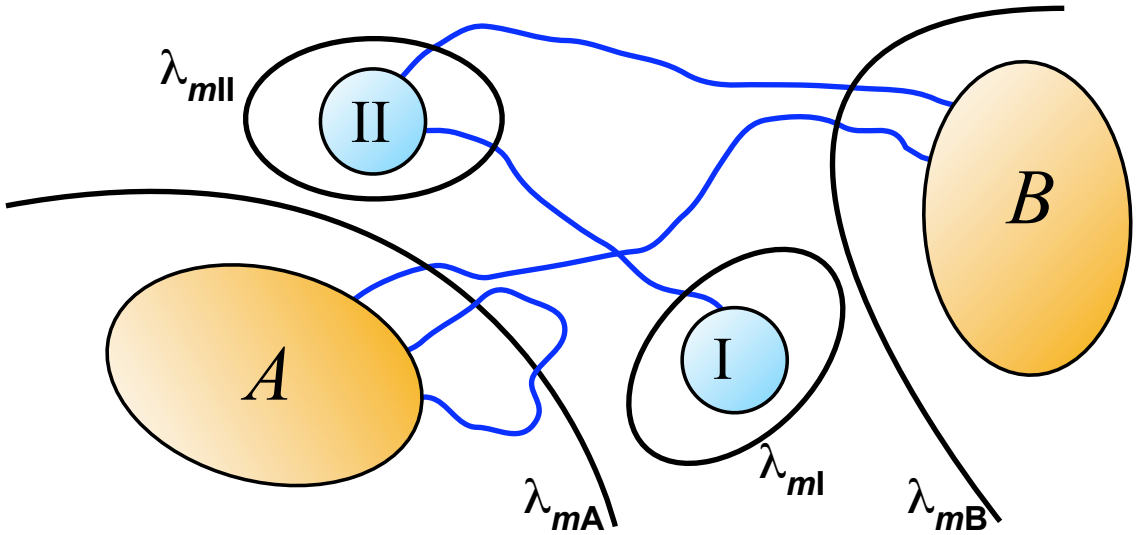
$$\mathcal{P}_{ij}\left[\mathbf{x}(L)\right] \equiv Z^{-1} \prod_{k} \bar{h}_k\left[\mathbf{x}(L)\right] h_i(x_0) \mathcal{P}\left[\mathbf{x}(L)\right] h_j(x_L) \hat{h}_i^m\left[\mathbf{x}(L)\right]$$

$$\hat{h}_{ij}^m[\mathbf{x}(L)] = \begin{cases} 1 & \text{if} \quad \exists\{t|0 \le t \le L\}: \quad x_t \in \Lambda_{mi}^- \\ & \quad \wedge \quad \exists\{t|0 \le t \le L\}: \quad x_t \in \Lambda_{mi}^+ \\ 0 & \text{otherwise} \end{cases}$$

$$\Lambda_{mi}^- = \{x|\lambda(x) < \lambda_{mi}\}$$

$$\Lambda_{mi}^+ = \{x|\lambda(x) > \lambda_{mi}\}$$

Multiple state rates



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

TIS:

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

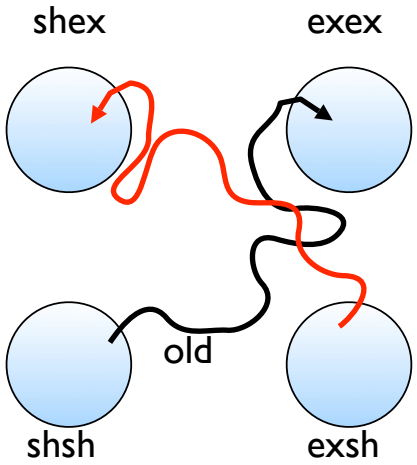
TPS:

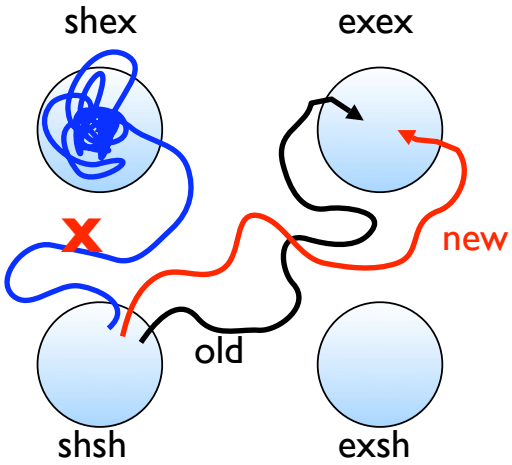
$$\frac{\text{no. of pathways coming from A, cross } \lambda_{mA}, \text{ end i}}{\text{no. of pathways coming from A, cross } \lambda_{mA}}$$

rates can be used in markovian state model

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

Multiple state or two state sampling?





- one simulation
- high acceptance
- fast decorrelation

- one simulation for each transition
- lower acceptance
- slower decorrelation

in simple model gain factor of 10

Further improvements

-biasing pathways to enhance sampling of rare paths

-combination with replica exchange

Conclusion

- TPS, TIS
  - can be used for wide range of rare event processes
  - has no need for reaction coordinate, just stable state definitions
  - gives true, unbiased molecular dynamical reaction pathways
  - do not assume reaction tube
  - yields correct rate constant, no suffering from low transmission coefficient
  - RC from LM methods

- Disadvantages
  - final state has to be known
  - multiple channels can be difficult (RETIS alleviates this)
  - long lived metastable states have to be treated separately or by MSTPS

- When is path sampling worthwhile?
  - rare event in complex system (when straightforward MD is inefficient)
  - complex unknown RC
  - other methods fail to do proper sampling

Acknowledgments

- Group members
  - Jarek Juraszek
  - Jocelyne Vreede
  - Maarten Wolf
  - Elske Leenders
  - Marieke Schor
  - Lizhe Zhu
  - Jutta Rogal

- Collaborators
  - Evert Jan Meijer
  - Bernd Ensing
  - Klaas Hellingwerf (Amsterdam)
  - Christoph Dellago (Vienna)

- Funding
  - FOM
  - NWO
  - EU

several PhD positions open