Bridging the timescale gap with transition path sampling

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

ACMM Amsterdam Center for Multiscale Modeling

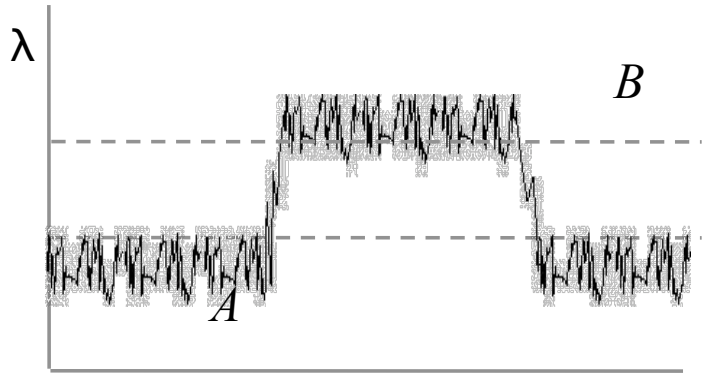
Outline

- Part 1
  - Rare events
  - The need for unbiased transition paths
  - Sampling the path ensemble
- Part 2
  - Analyzing the path ensemble
  - Calculation of rate constants
  - Transition state ensemble and reaction coordinate
- Part 3
  - Application to protein folding

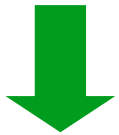
Rare events

Interesting transitions in complex fluids

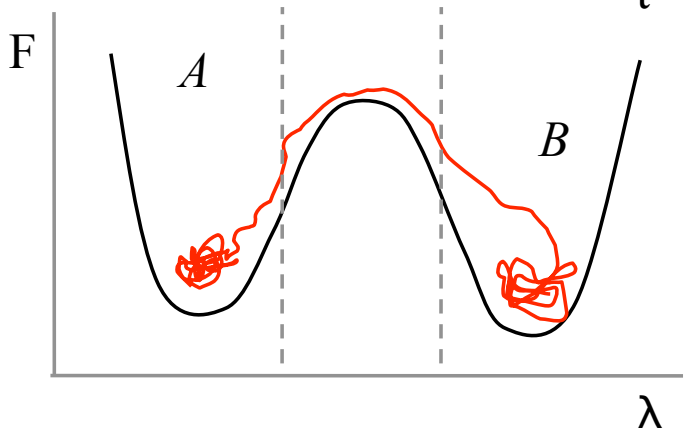
- solution chemistry
- phase transitions
- protein folding
- enzymatic reactions
- nucleation
- complex surface reaction
- membrane fusion

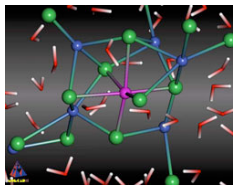


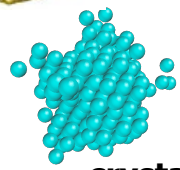
These reactions happen on a long time scale compared to the molecular timescale (eg solvent motion)



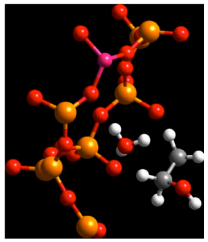
dominated by collective, rare events: straightforward MD is unpractical



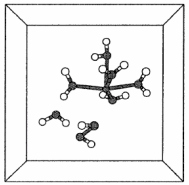


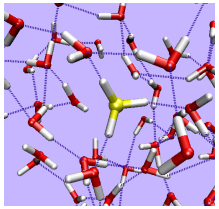


crystallisation

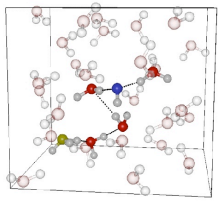


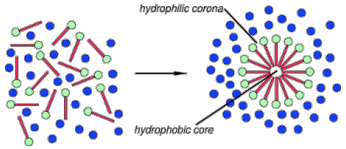
catalysis



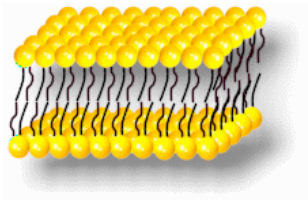


solution reactions



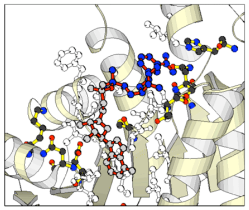


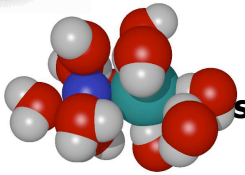
Rare Events



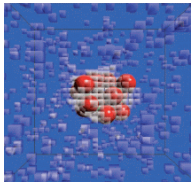
complex fluids

enzyme reactions

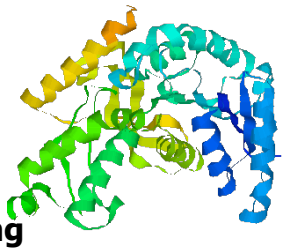


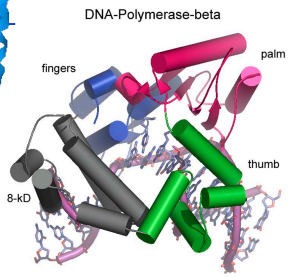


solvent effects

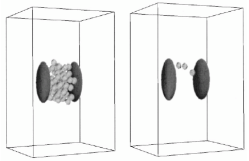


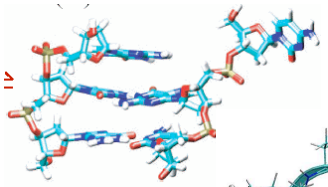
folding & binding

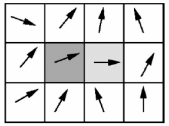


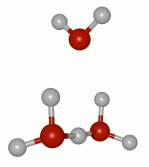


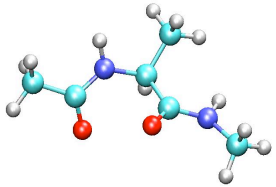
isomerization

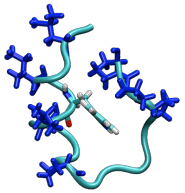


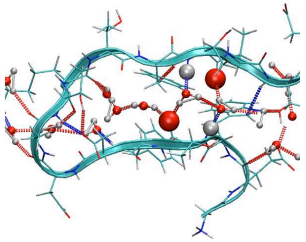


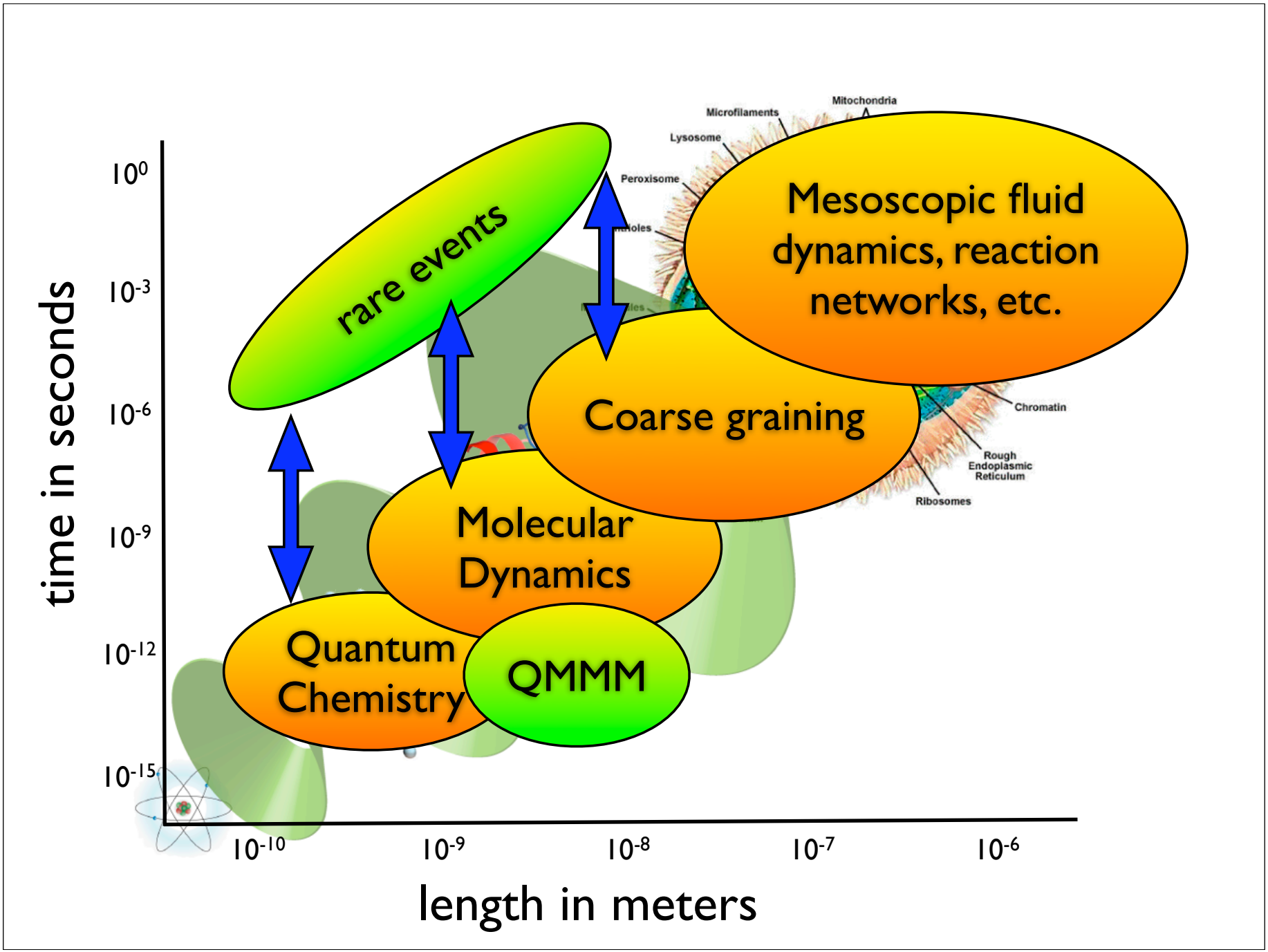




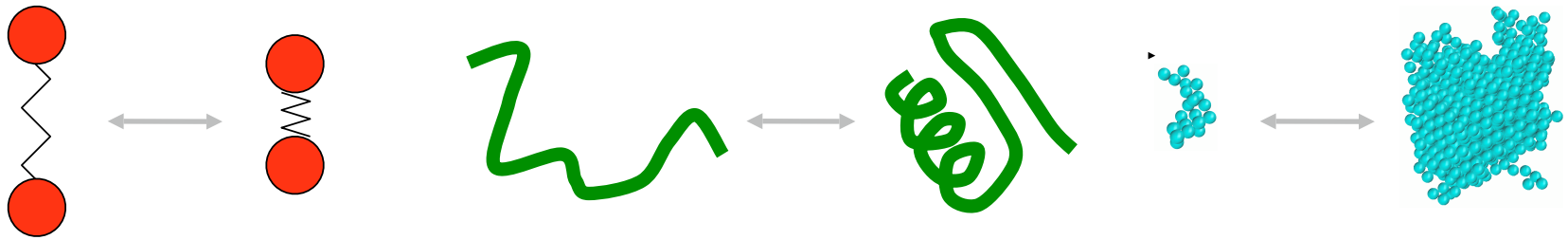







time in seconds
10^{0}
10^{-3}
10^{-6}
10^{-9}
10^{-12}
10^{-15}
10^{-10}
10^{-9}
10^{-8}
10^{-7}
10^{-6}
length in meters
rare events
Mesoscopic fluid dynamics, reaction networks, etc.
Coarse graining
Molecular Dynamics
Quantum Chemistry
QMMM
Mitochondria
Microfilaments
Lysosome
Peroxisome
Chromatin
Rough Endoplasmic Reticulum
Ribosomes

Two state reaction kinetics



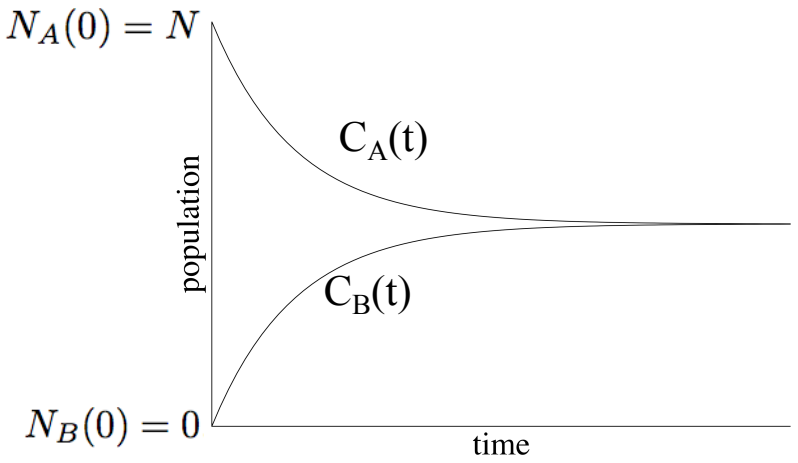
A ⇌ B

$$\dot{c}_A = -k_{AB}\, c_A + k_{BA}\, c_B$$
$$\dot{c}_B = \quad k_{AB}\, c_A - k_{BA}\, c_B$$

$$\Delta c_A(t) = c_A(t) - \langle c_A \rangle$$

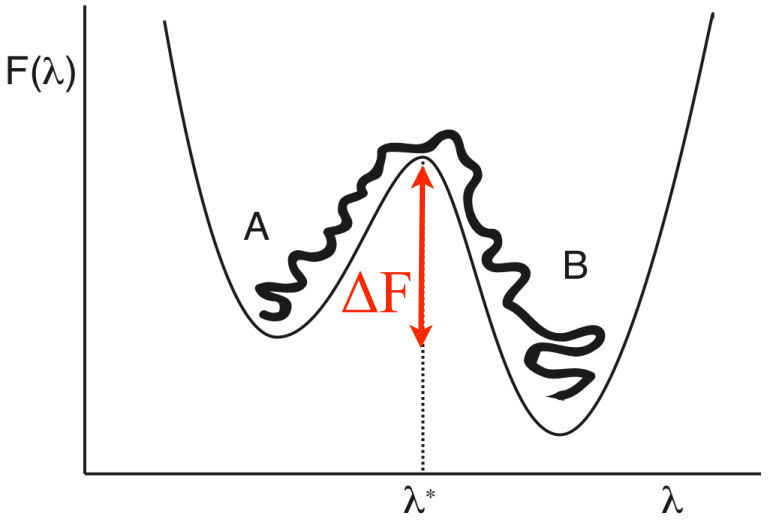
$$\Delta c_A(t) = \Delta c_A(0) \exp(-t/\tau_{\text{rxn}})$$

$$\tau_{\text{rxn}}^{-1} = k_{AB} + k_{BA}$$



Transition state theory

- Oldest attempt to describe rare events
  - Marcelin 1915
  - Eyring 1935
  - Wigner 1938

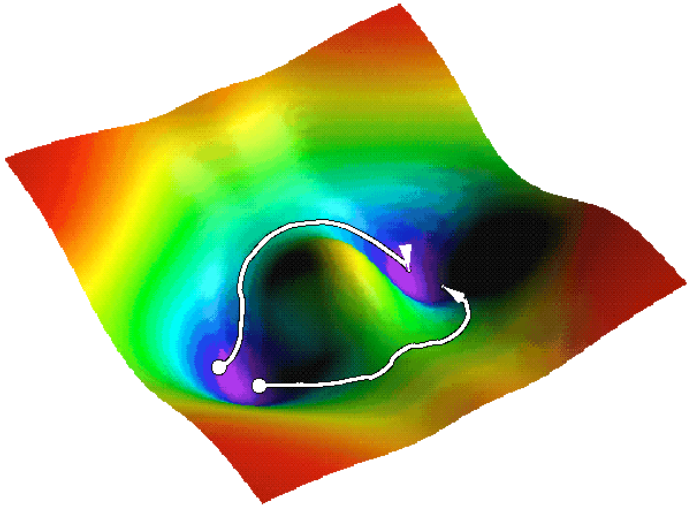


$$k_{AB}^{TST} \equiv \nu e^{-\beta \Delta F}$$

kinetic prefactor

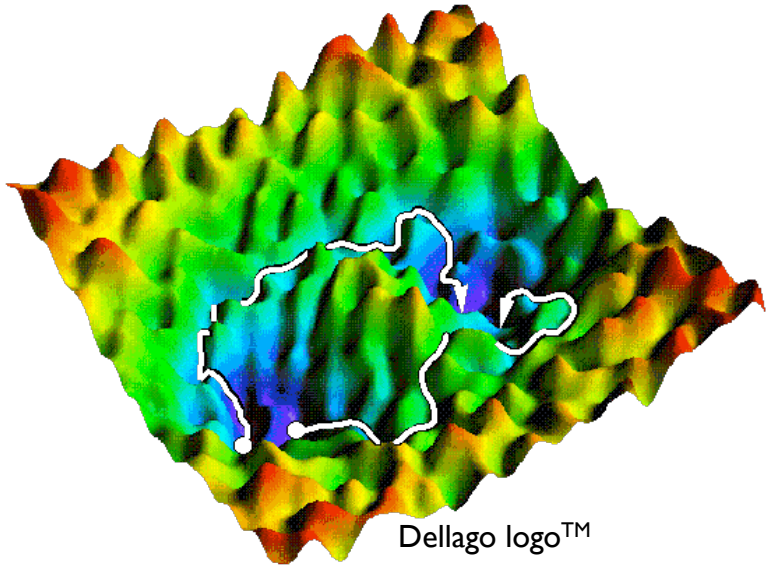
- Why can't we simply apply TST to all problems?
  - assumes that all trajectories that reach top with positive velocity will end in B
  - not always easy due to rough energy landscapes

Smooth vs rough energy



- # saddle points limited
- reaction coordinate known
- pathways can be enumerated
- TST (Arrhenius) applies

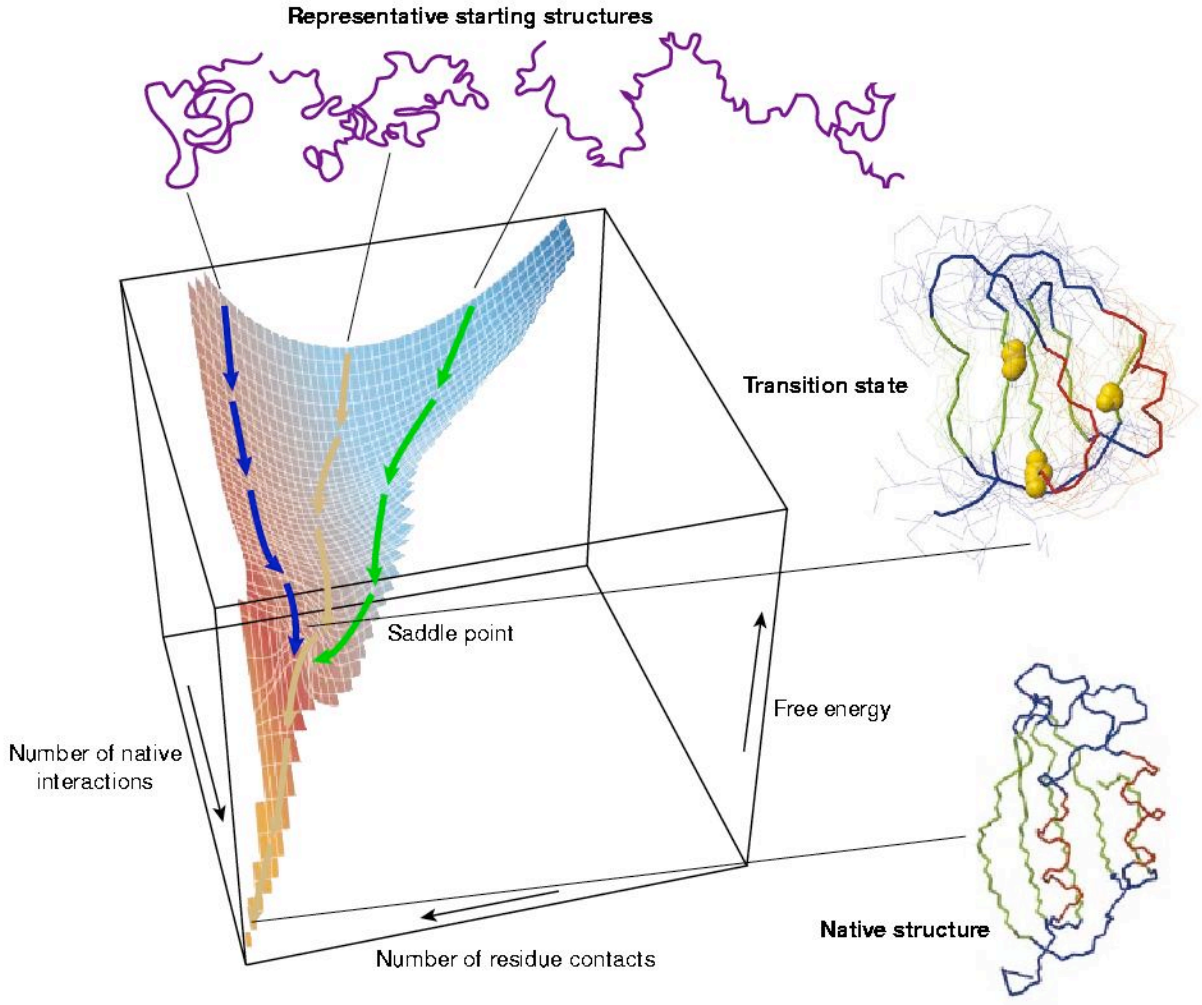
- saddle points uncountable
- reaction coordinates unknown
- entropy important, many pathways
- TST still applies?



Dellago logo™

How do we explore?
Usually by free energy surface

Free energy for protein folding



Taken from Dobson, Nature, 2003

Questions

- How do we get free energy? (thermodyamics)
- How do we compute rate constants? (kinetics)
- How do find transition states? (mechanism)
- What is the reaction coordinate of a reaction/process?
- How do we know we have the correct reaction coordinate?

How do we get free energy as a function of an order parameter λ ?

Free energy by umbrella sampling

The regular distribution of an order parameter λ is

$$P(\lambda) = \langle \delta[\lambda - \lambda(x)] \rangle = \frac{\int dx \rho(x) \delta[\lambda - \lambda(x)]}{\int dx \rho(x)}$$

$$\rho(x) = e^{-\beta \mathcal{H}(x)}/Z \qquad Z = \int e^{-\beta \mathcal{H}(x)} dx$$

multiplying both sides with exp(-βV_{bs}) gives

$$P_{bs}(\lambda) = \frac{\int dx \rho(x) \exp\left[-\beta V_{bs}(\lambda(x)) \delta(\lambda - \lambda(x))\right]}{\int dx \rho(x) \exp\left[-\beta V_{bs}(\lambda(x))\right)}$$

where V_{bs} is the bias potential

Free energy is given by

$$\beta F(\lambda) = -\ln P_{bs}(\lambda) - \beta V_{bs}(\lambda) + const$$

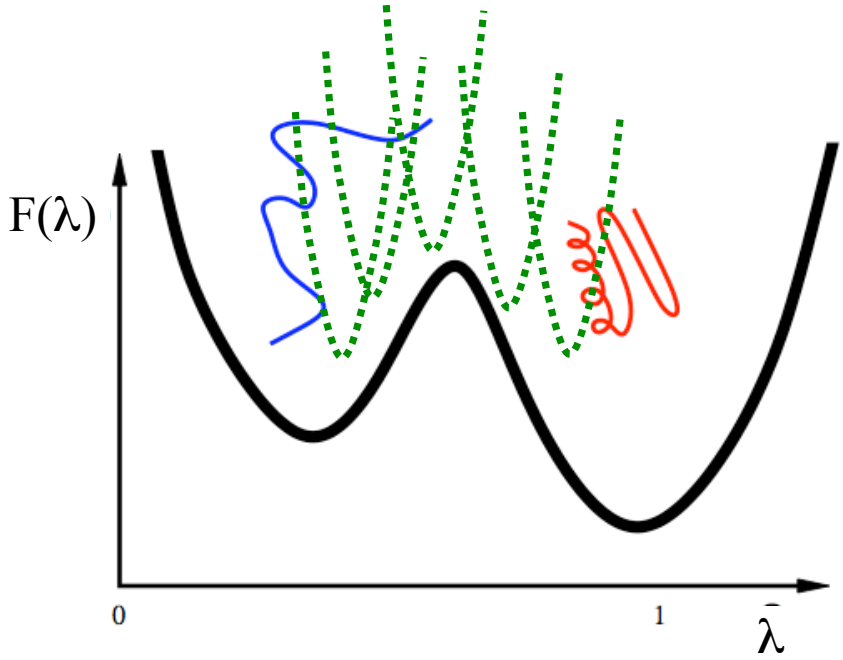
Umbrella sampling

Biasing potential can take any functional form

simplest one quadratic

$$V_{bs}(\lambda(x)) = c(\lambda(x) - \lambda_i)^2$$

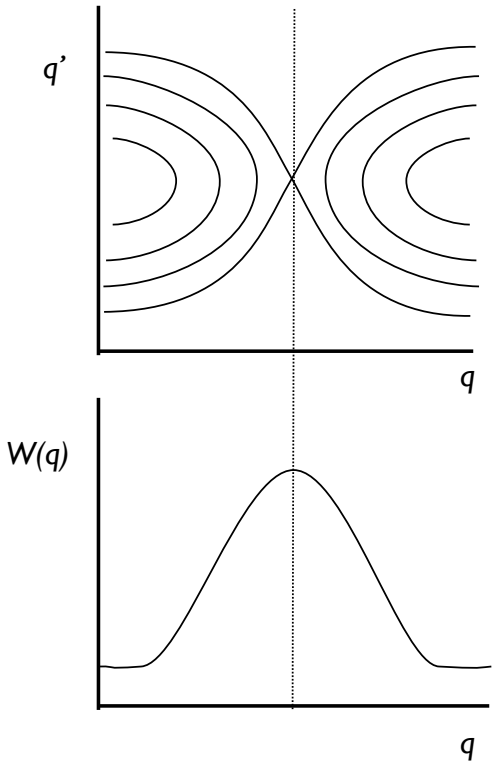
Biasing potential forces system in to unlikely region

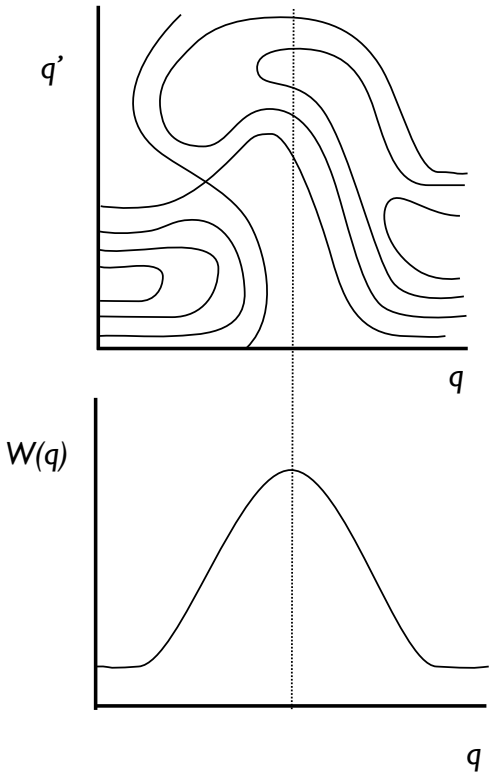


Breakdown of biased sampling

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

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



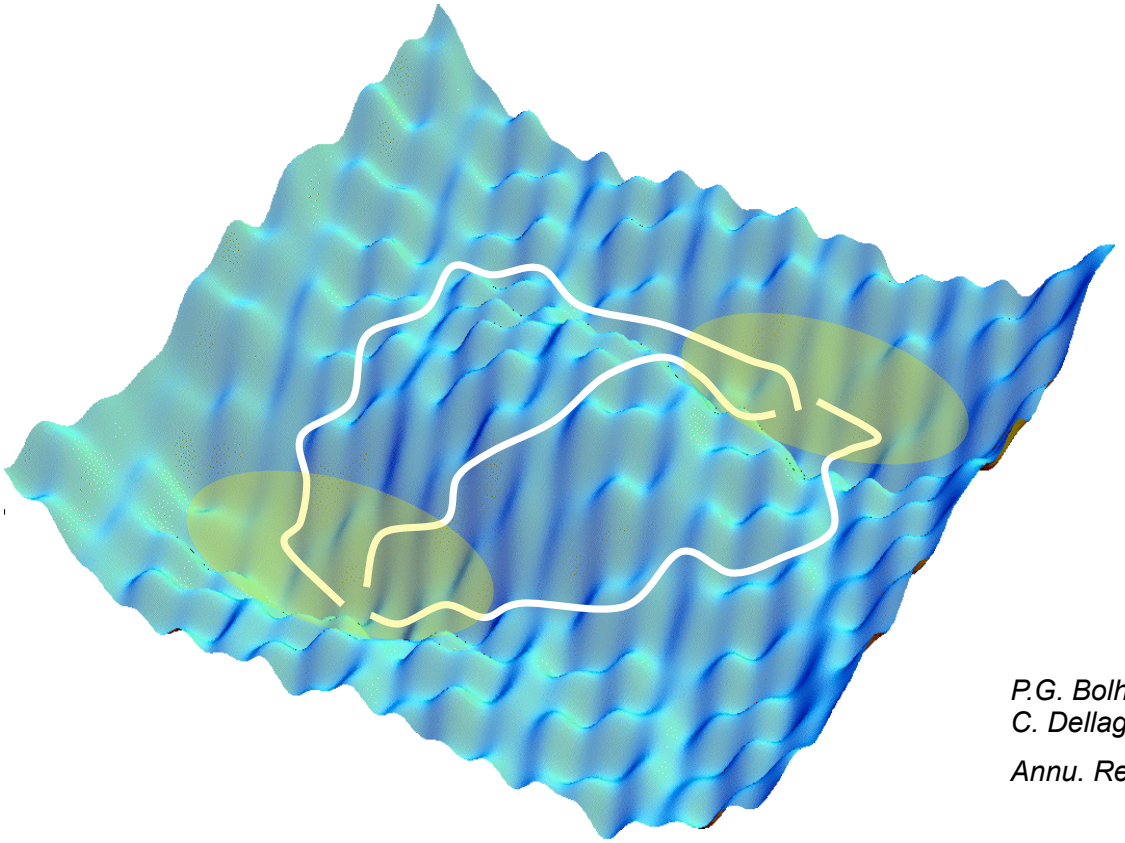


$$W(q) = -kT \ln \int dq' \exp\{-\beta E(q,q')\}$$

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

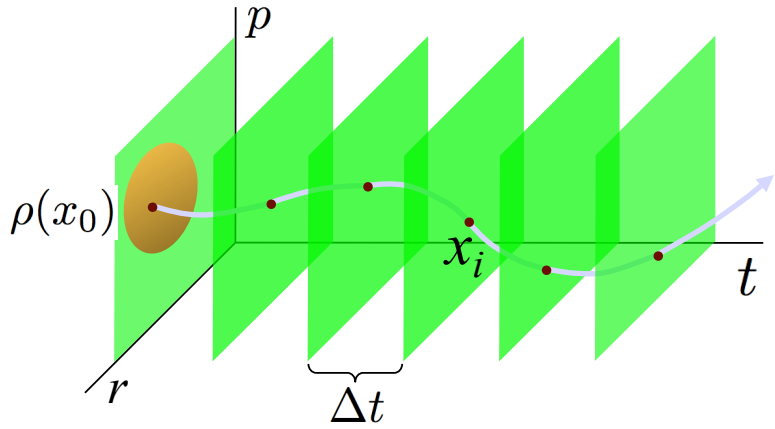
Annu. Rev. Phys. Chem 2002

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

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

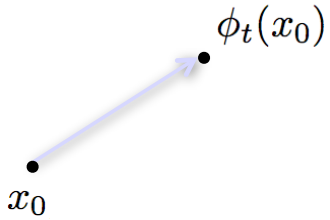
$$\mathcal{P}[\mathbf{x}(L)] = \rho(x_o) \prod_{i=0}^{L-1} p(x_i \to x_{i+1})$$

Can be defined for deterministic and stochastic dynamics

Newtonian (Hamiltonian) dynamics:

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

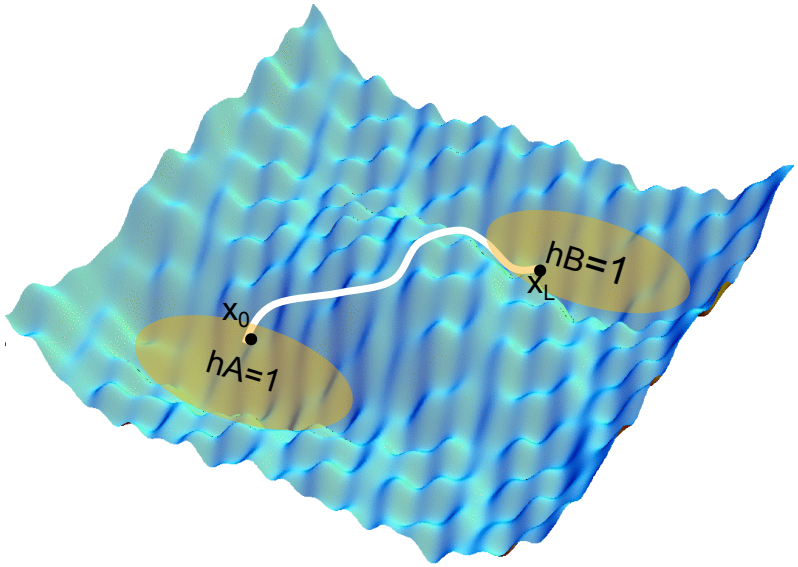
$$\dot{p} = -\frac{\partial \mathcal{H}(r,p)}{\partial r}$$



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

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 $h_A(x)$

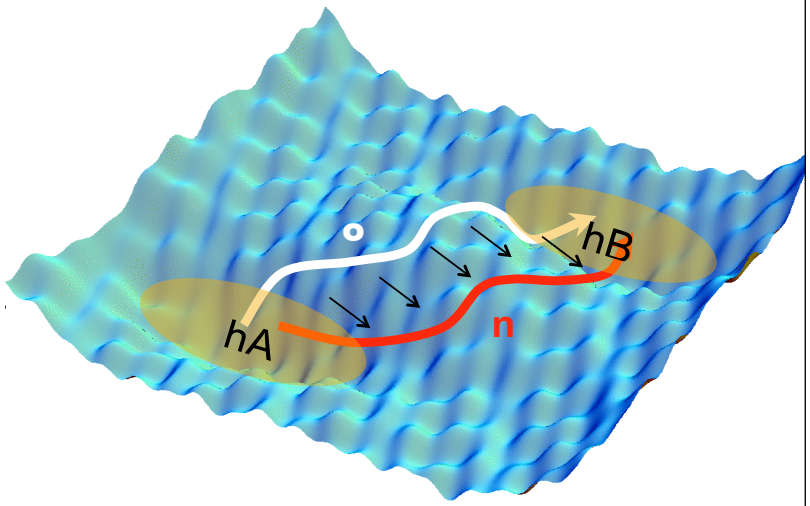
$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Path probability distribution

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

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

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

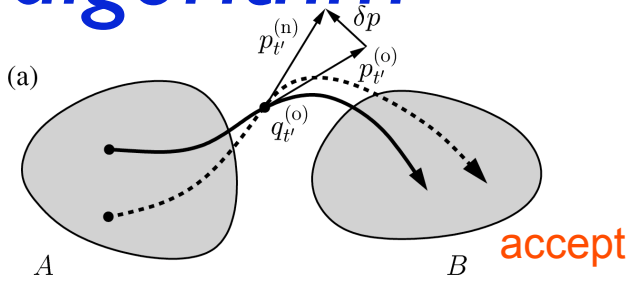


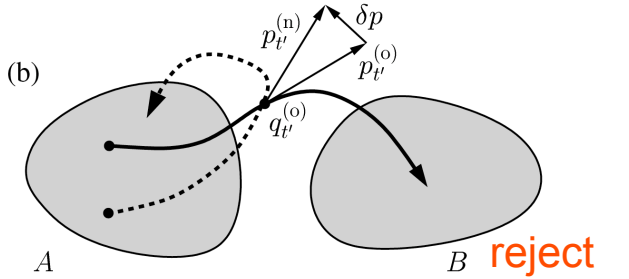
Importance sampling using Metropolis rule :

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

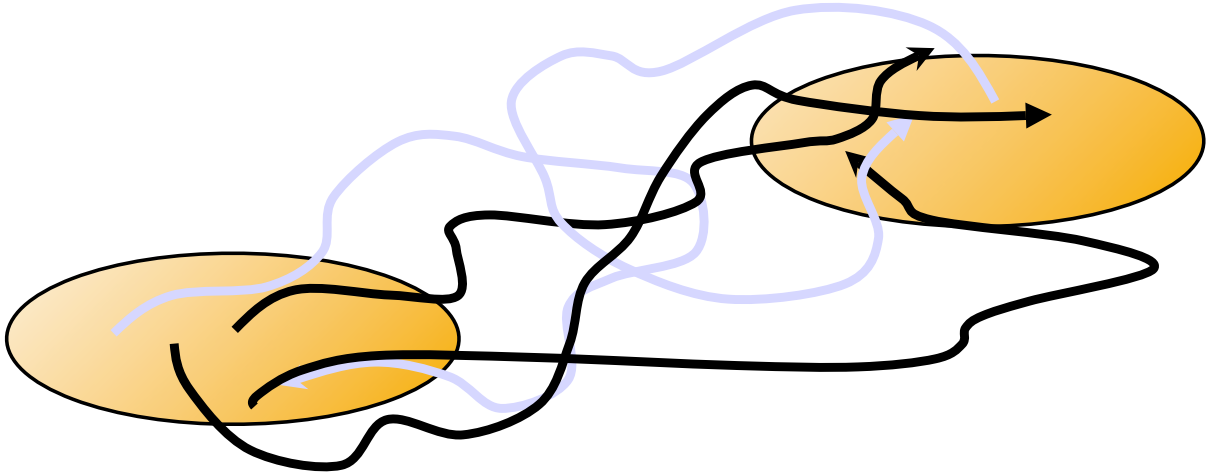
Standard shooting algorithm

- 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

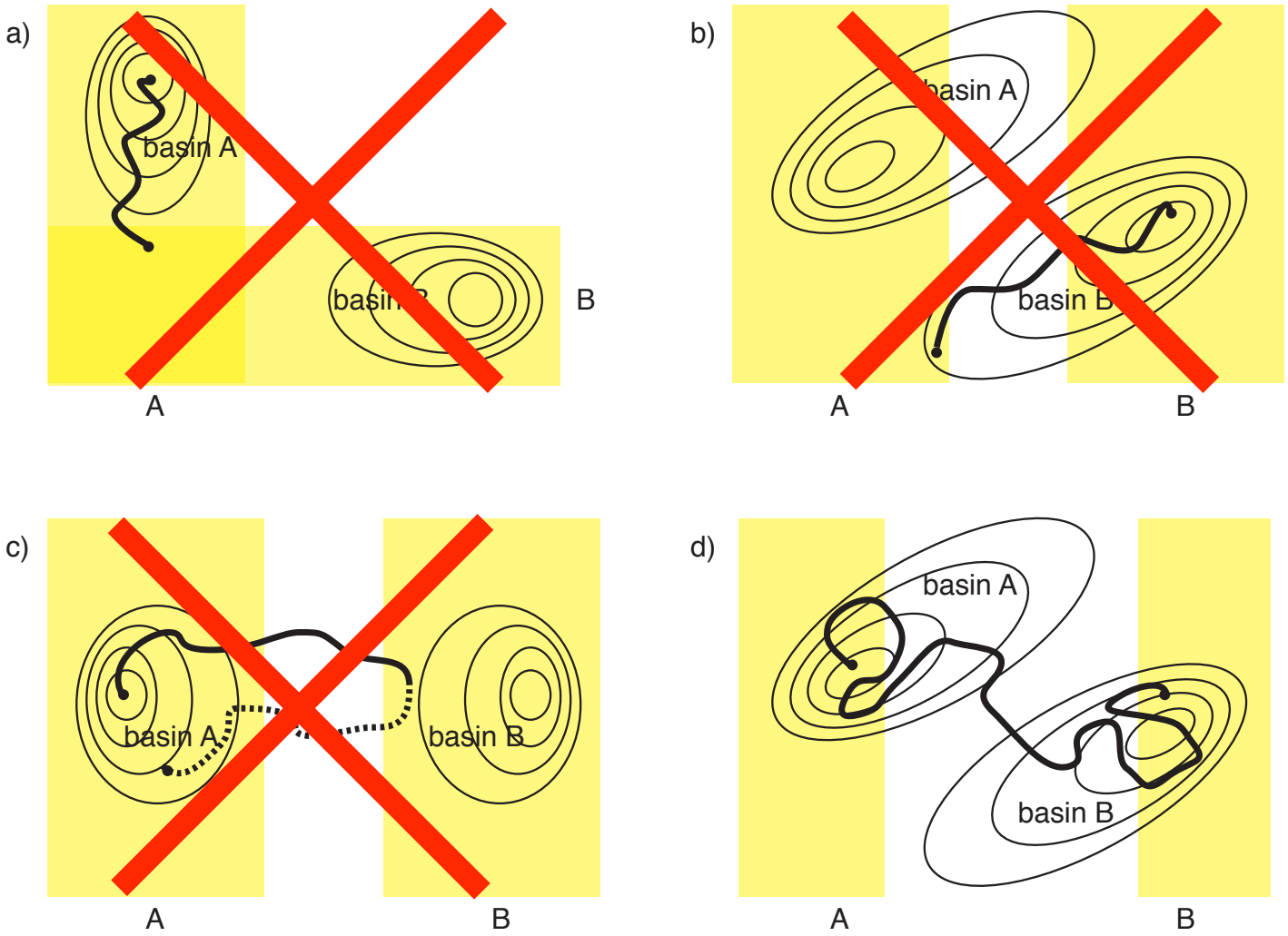


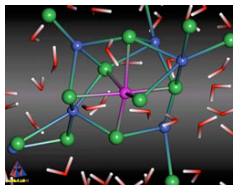


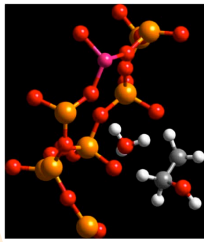
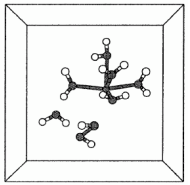
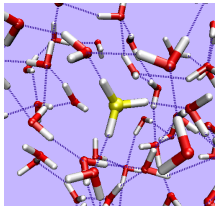
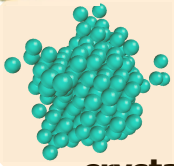
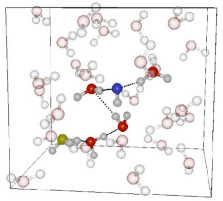
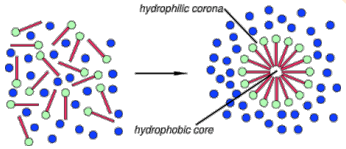
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_T^{(n)})$$

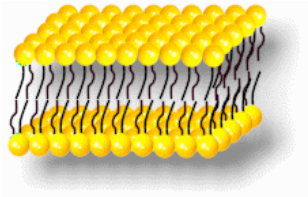
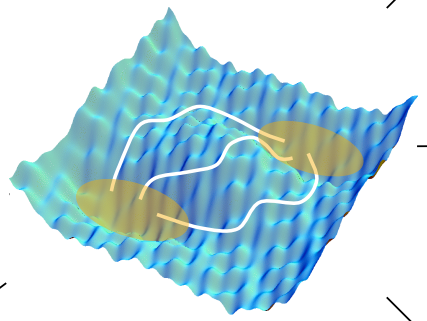
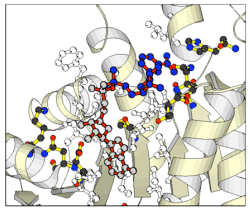
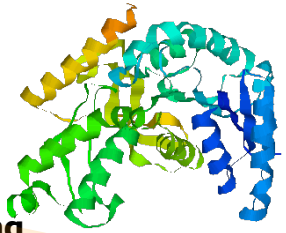
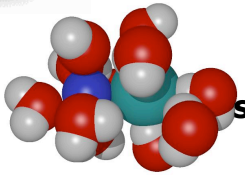
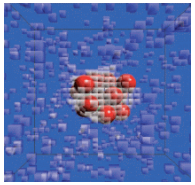
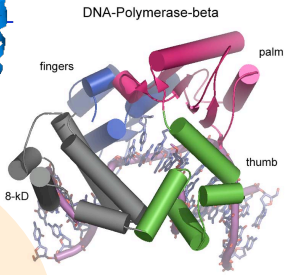


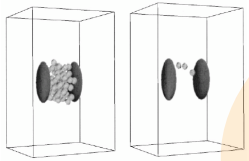
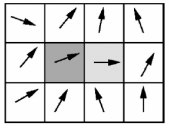
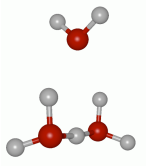
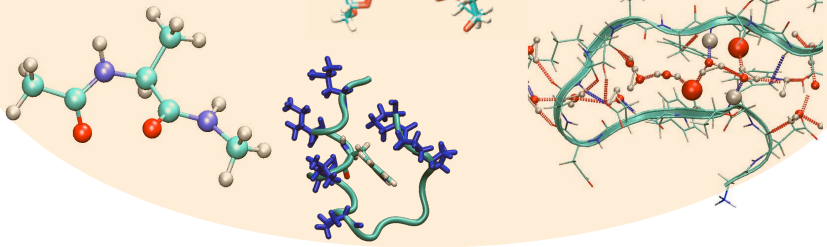
Definition of stable states



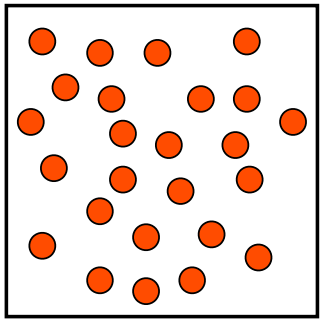







crystallisation
catalysis
reactions


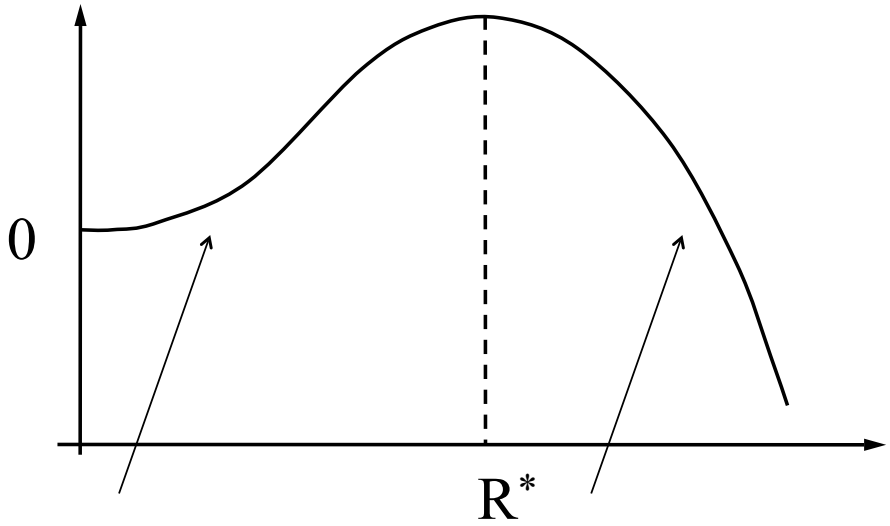

complex fluids

enzyme reactions



solvent effects

folding & binding
conformational change


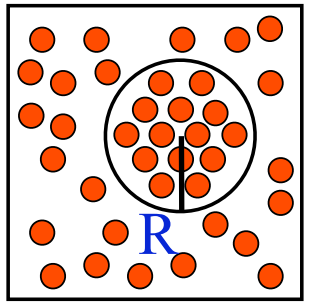





Classical nucleation



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

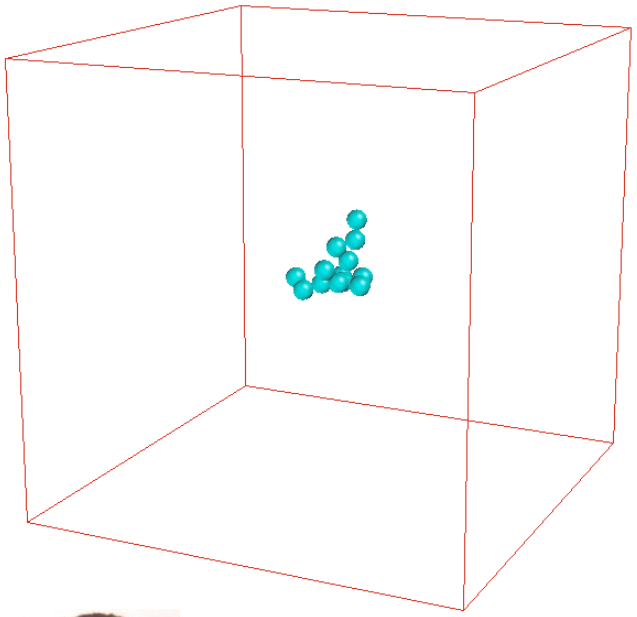
ρ: density

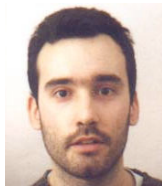
Path sampling of nucleation

TIS in NPH ensemble, as density and temperature change

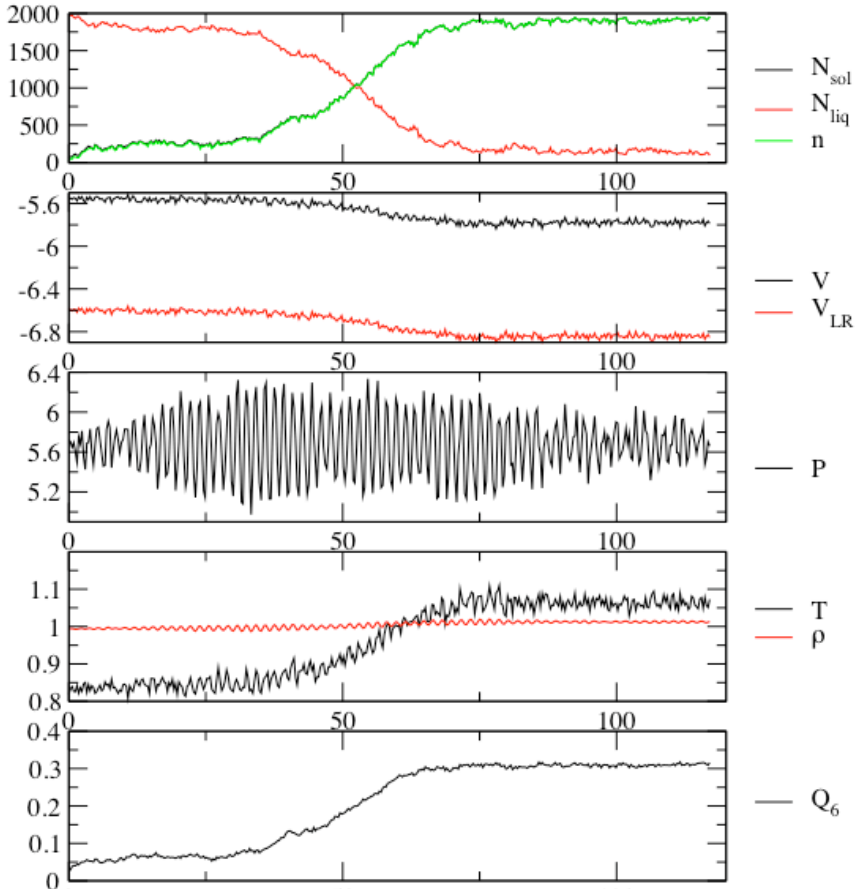
N=10000, P=5.68 H=1.41 (25 % undercooling)

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





Daniele Moroni



Outline

- Part 1
  - Rare events
  - The need for unbiased transition paths
  - Sampling the path ensemble
- Part 2
  - Analyzing the path ensemble
  - Calculation of rate constants
  - Transition state ensemble and reaction coordinate
- Part 3
  - Application to protein folding

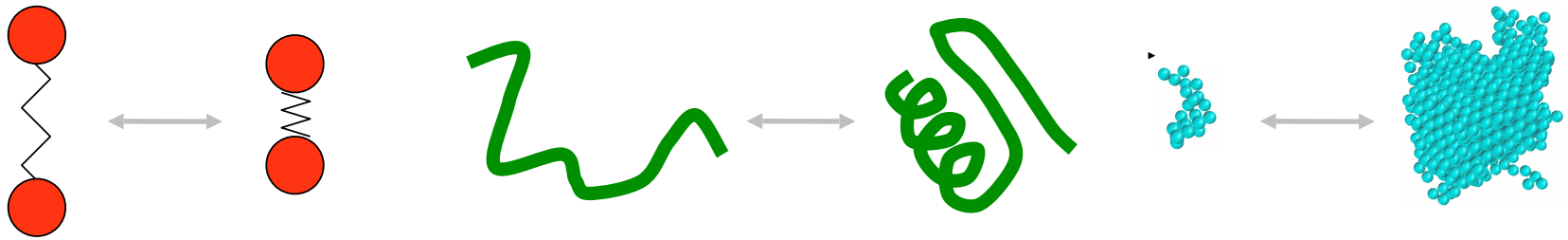
Sampling paths is only the beginning

- *Eugene Wigner:* ***"It is nice to know that the computer understands the problem. But I would like to understand it too."***

- Path ensemble needs to be further explored to obtain:
  - Rate constants
  - Free energy
  - Transition state ensembles
  - Mechanistic picture
  - Reaction coordinate

- Illustrative example: crystal nucleation

Back to the phenomology



$$\mathbf{A} \rightleftharpoons \mathbf{B}$$

$$\dot{c}_A = -k_{AB}\, c_A + k_{BA}\, c_B$$
$$\dot{c}_B = \quad k_{AB}\, c_A - k_{BA}\, c_B$$

$$\Delta c_A(t) = c_A(t) - \langle c_A \rangle$$

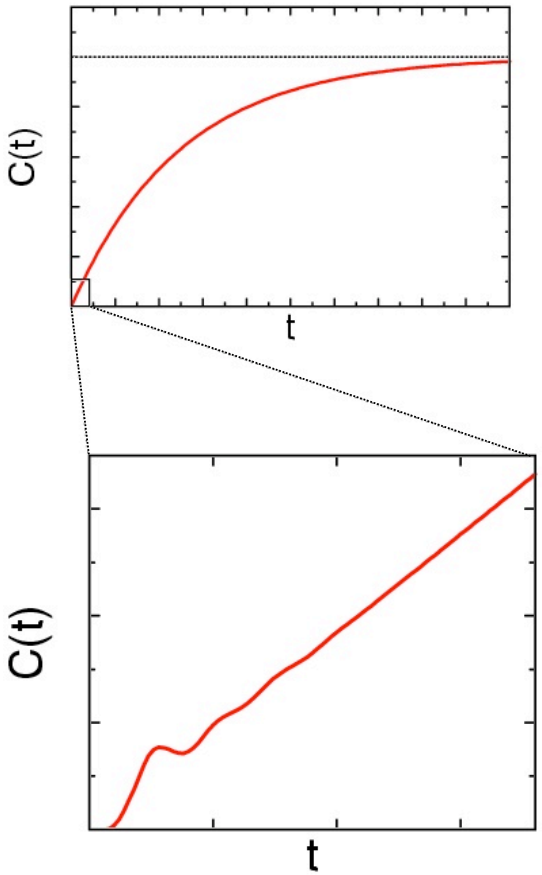
$$\Delta c_A(t) = \Delta c_A(0) \exp(-t/\tau_{\mathrm{rxn}})$$

$$\tau_{\mathrm{rxn}}^{-1} = k_{AB} + k_{BA}$$

$$N_A(0) = N$$
$$N_B(0) = 0$$

$$P(B,t|A,0) = \frac{\langle N_B \rangle}{N}\left(1 - e^{-t/\tau_{\mathrm{rxn}}}\right)$$

Microscopic description



$$P(B,t|A,0) = \frac{\langle h_A(x_0)h_B(x_t)\rangle}{\langle h_A\rangle}$$

$$\frac{\langle h_A(x_0)h_B(x_t)\rangle}{\langle h_A\rangle} = \langle h_B\rangle\left(1 - e^{-t/\tau_{\rm rxn}}\right)$$

for $\tau_{\rm mol} < t \ll \tau_{\rm rxn}$

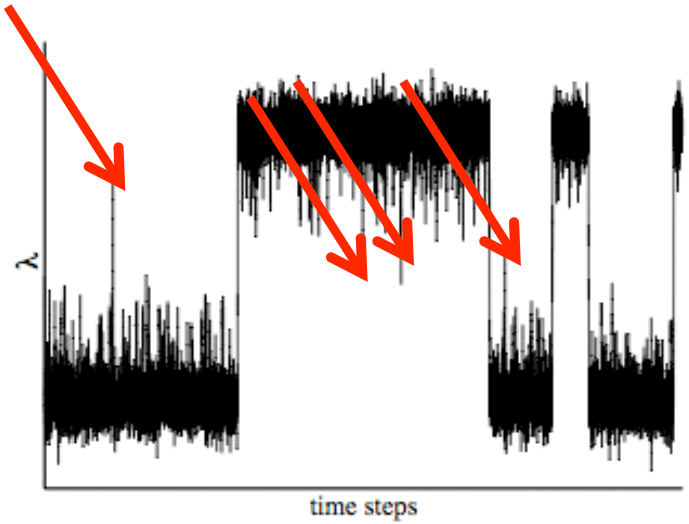
$$C(t) \equiv \frac{\langle h_A(x_0)h_B(x_t)\rangle}{\langle h_A\rangle} = k_{AB}t$$

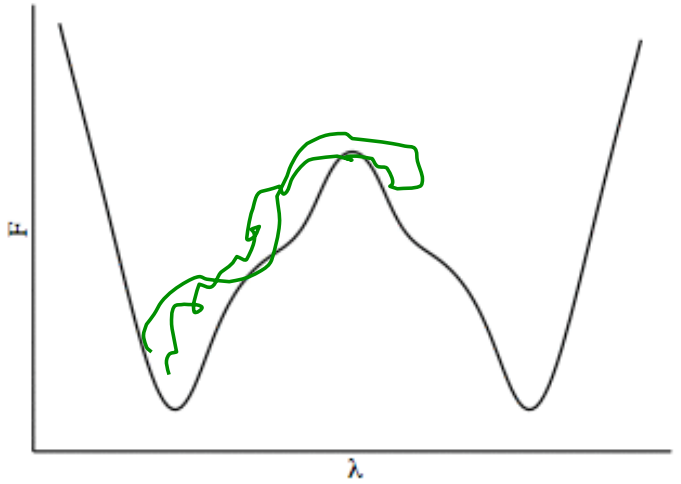
$$k(t) \equiv \dot{C}(t)$$

Recrossings

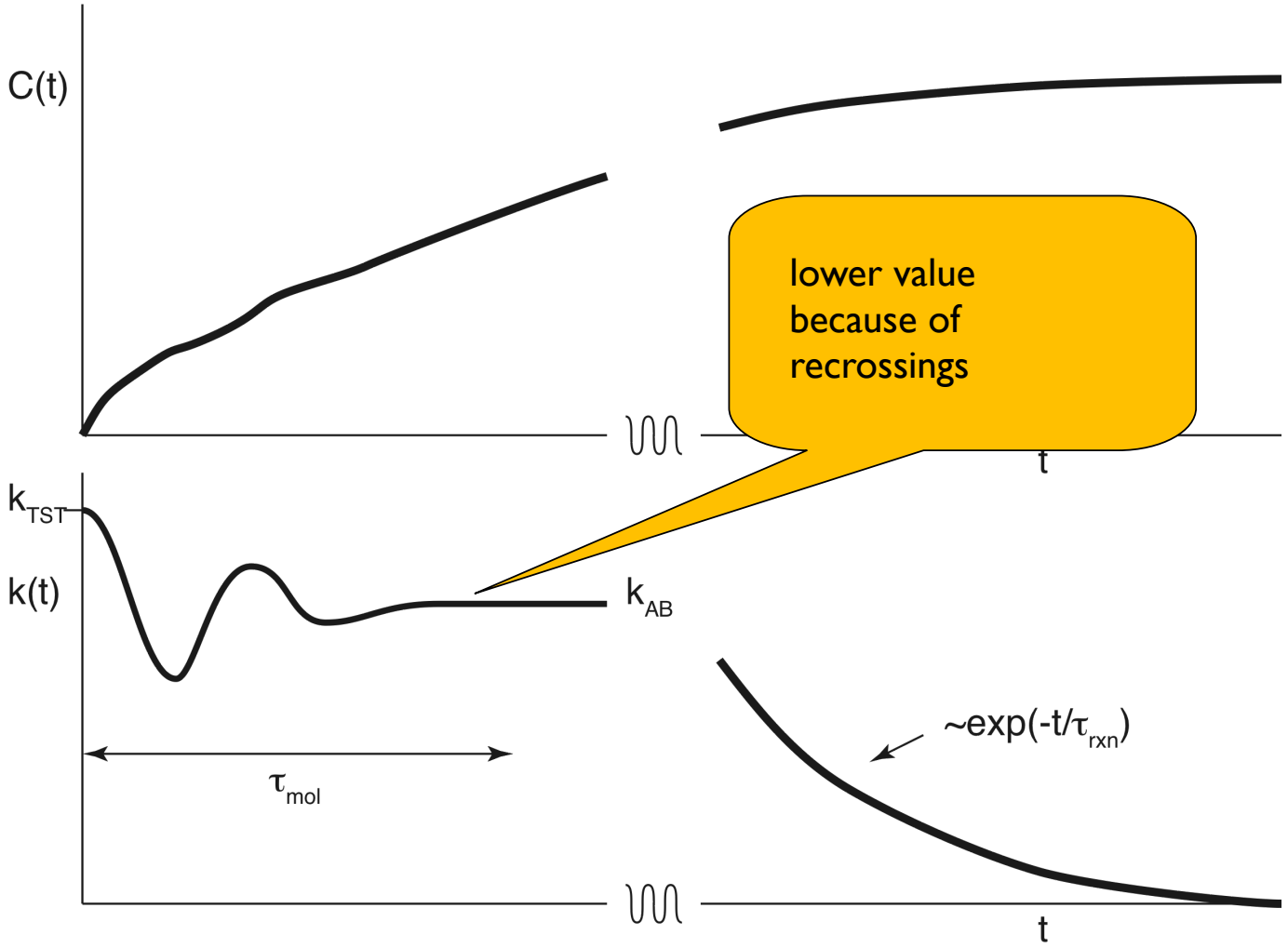
There are recrossings that cause overestimation of the rate constant

trajectories that seem to overcome the barrier but in fact bounce back

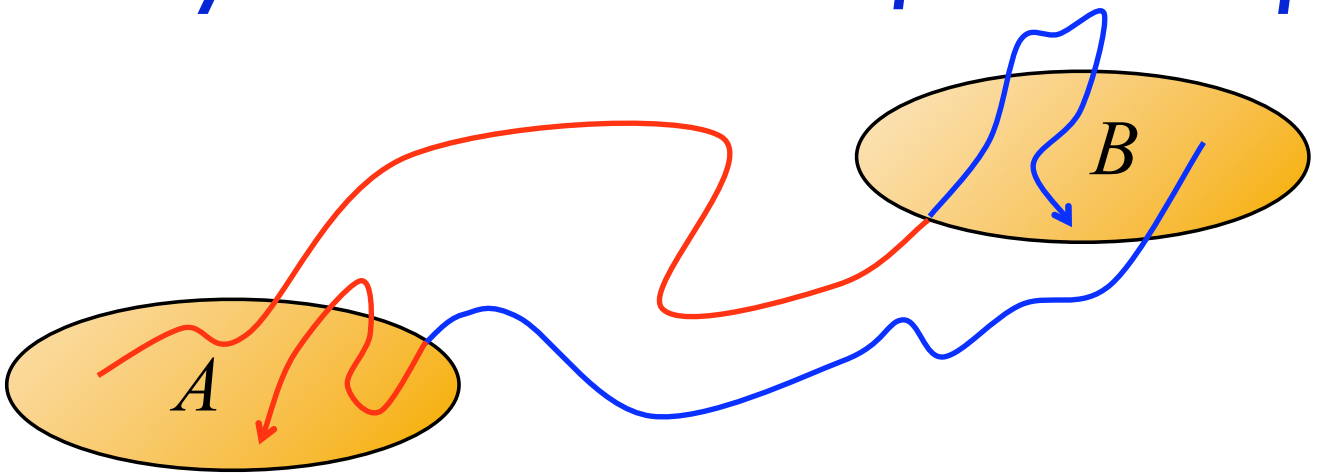




Decay of correlation function



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

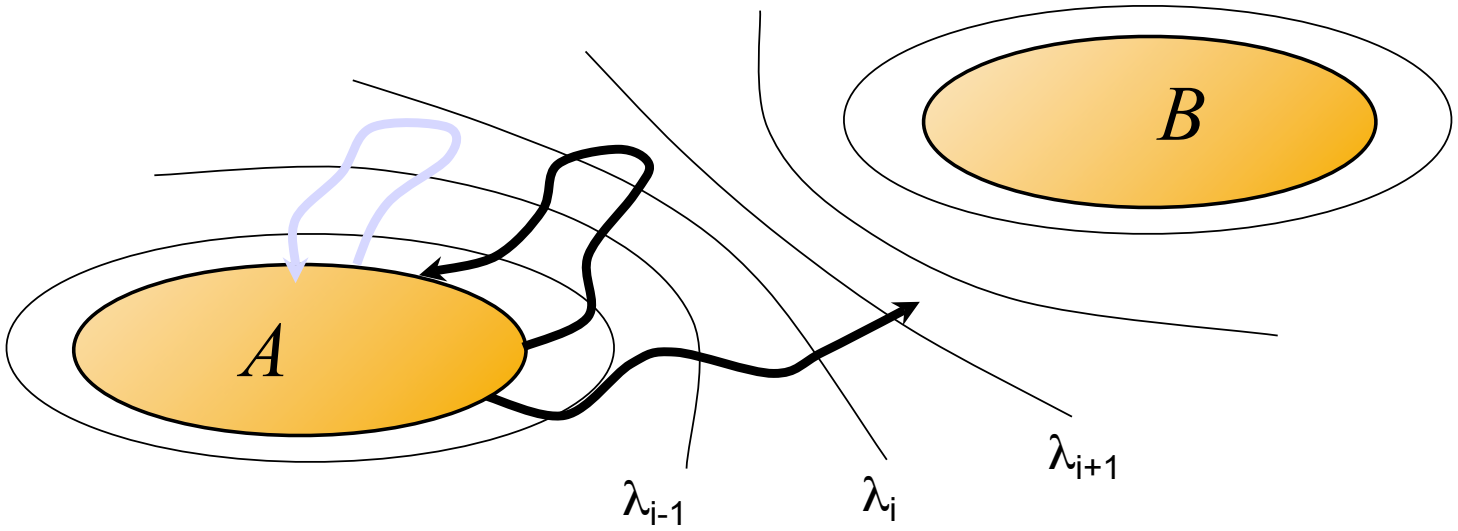
$$\mathcal{C}(t) \equiv \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t) \rangle}{\langle h_{\mathcal{A}} \rangle}$$

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

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

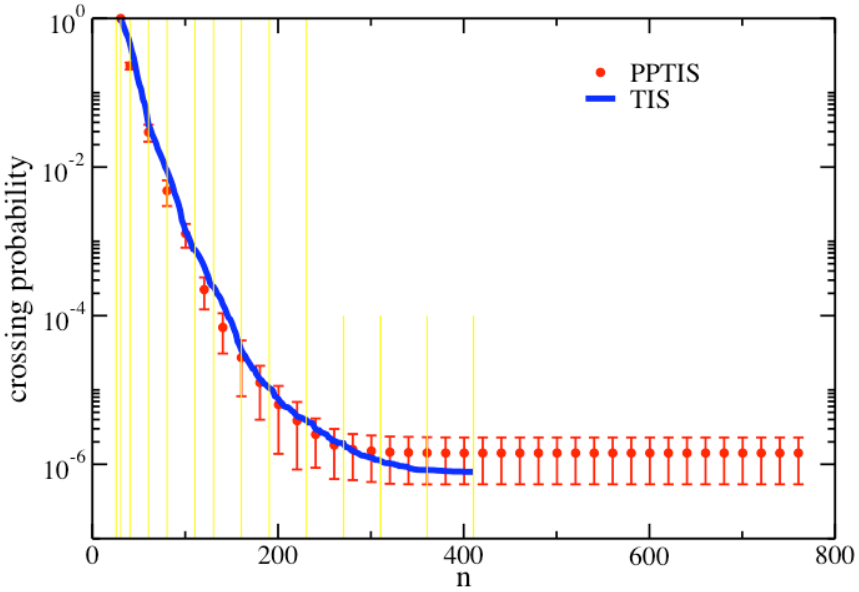


$P_A(\lambda_{i+1} \mid \lambda_i)$ = probability that path crossing i for first time after leaving A reaches i+1 before A

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

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

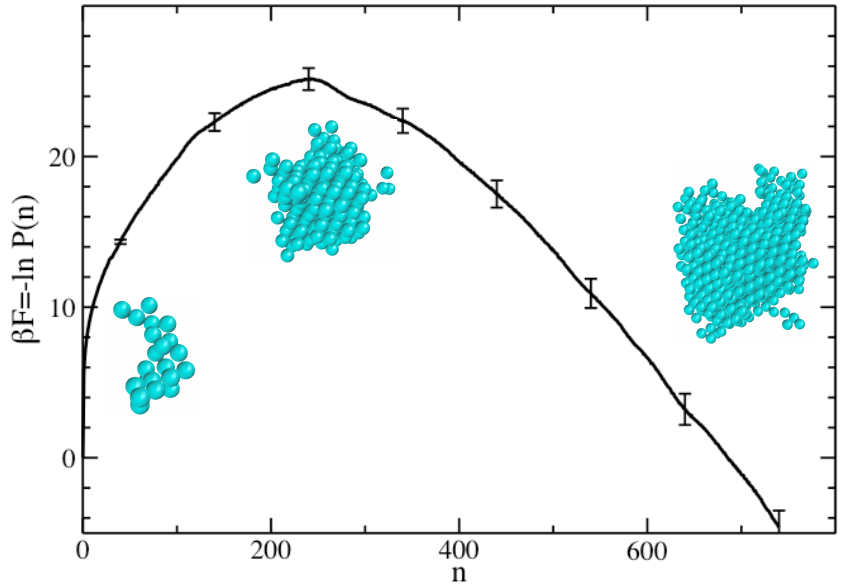


Nucleation rate

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



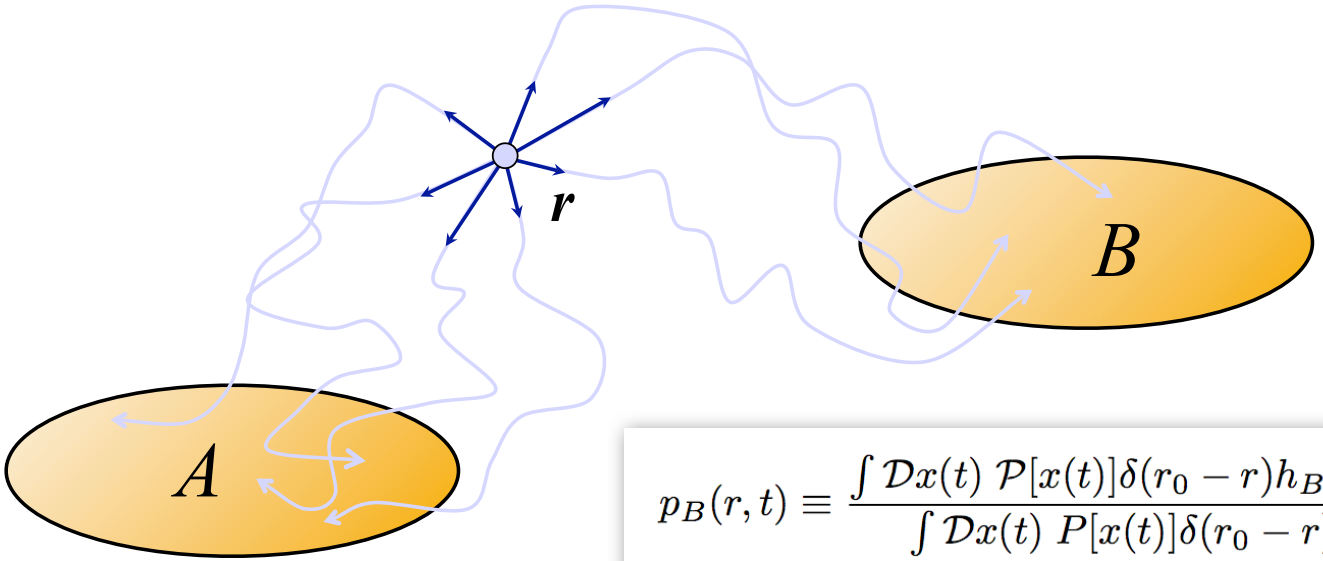
Free energy follows directly
Moroni, van Erp, Bolhuis, PRE, 2005

Structural analysis?

Committor

(aka p-fold, splitting probability)

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



$$p_B(r,t) \equiv \frac{\int \mathcal{D}x(t)\ \mathcal{P}[x(t)]\delta(r_0 - r)h_B(x_t)}{\int \mathcal{D}x(t)\ P[x(t)]\delta(r_0 - r)}$$

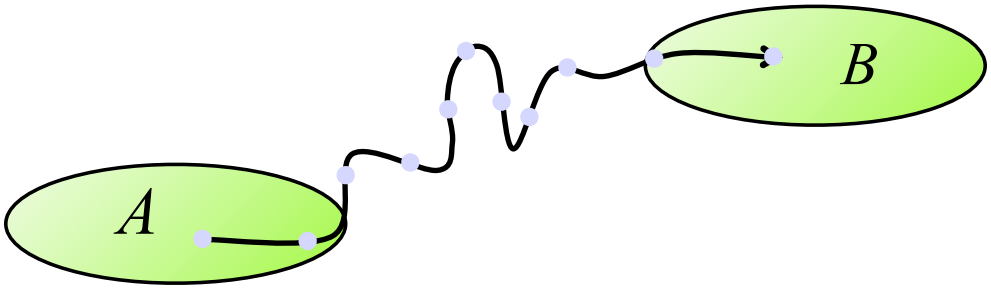
$$p_B(r,t) \approx \frac{1}{N}\sum_{i=1}^{N} h_B(x_t^{(i)})$$

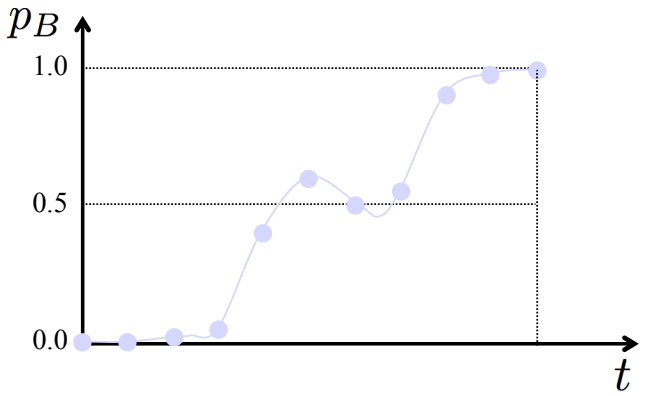
$$\sigma = \sqrt{\langle (p_B^{(N)} - p_B)^2 \rangle} = \sqrt{\frac{p_B(1-p_B)}{N}}$$

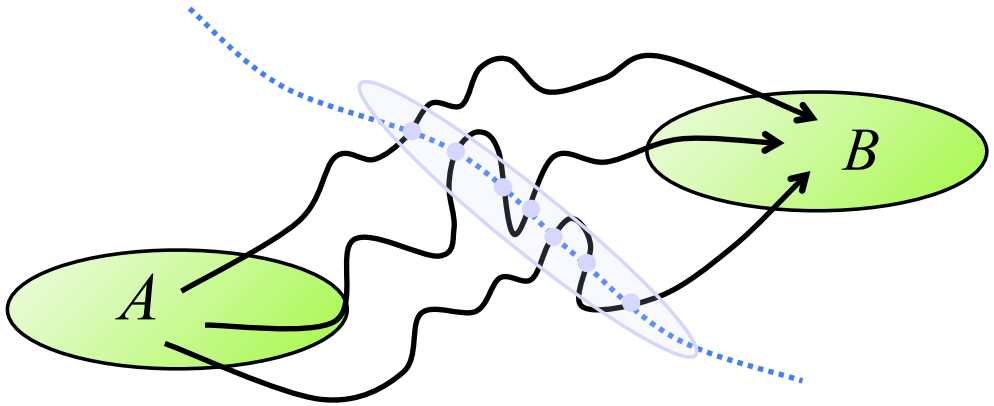
L. Onsager, Phys. Rev. **54**, 554 (1938).
M. M. Klosek, B. J. Matkowsky, Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991)
V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, J. Chem. Phys. **108**, 334 (1998)

Transition state ensemble

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





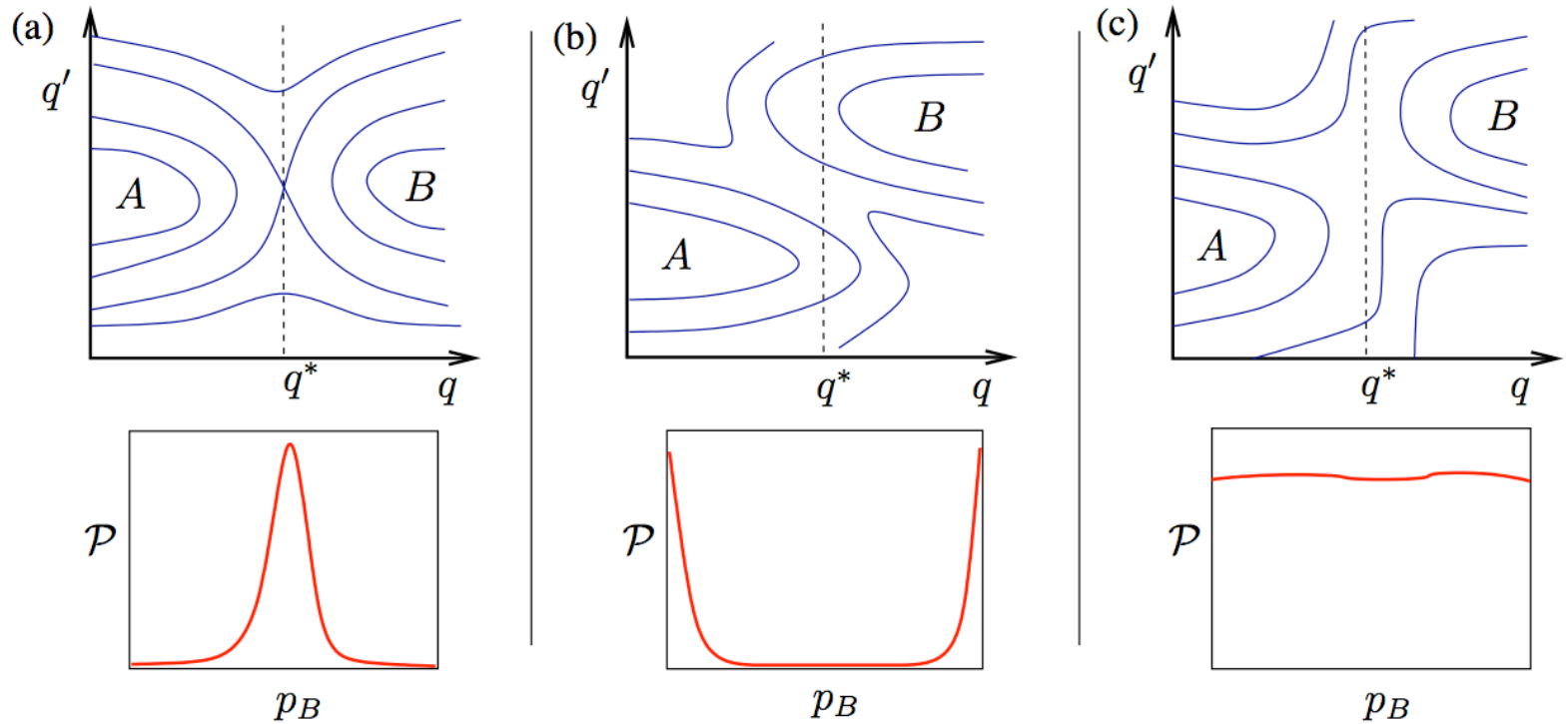


TSE:

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

Committor analysis

An attempt to find out the reaction coordinate

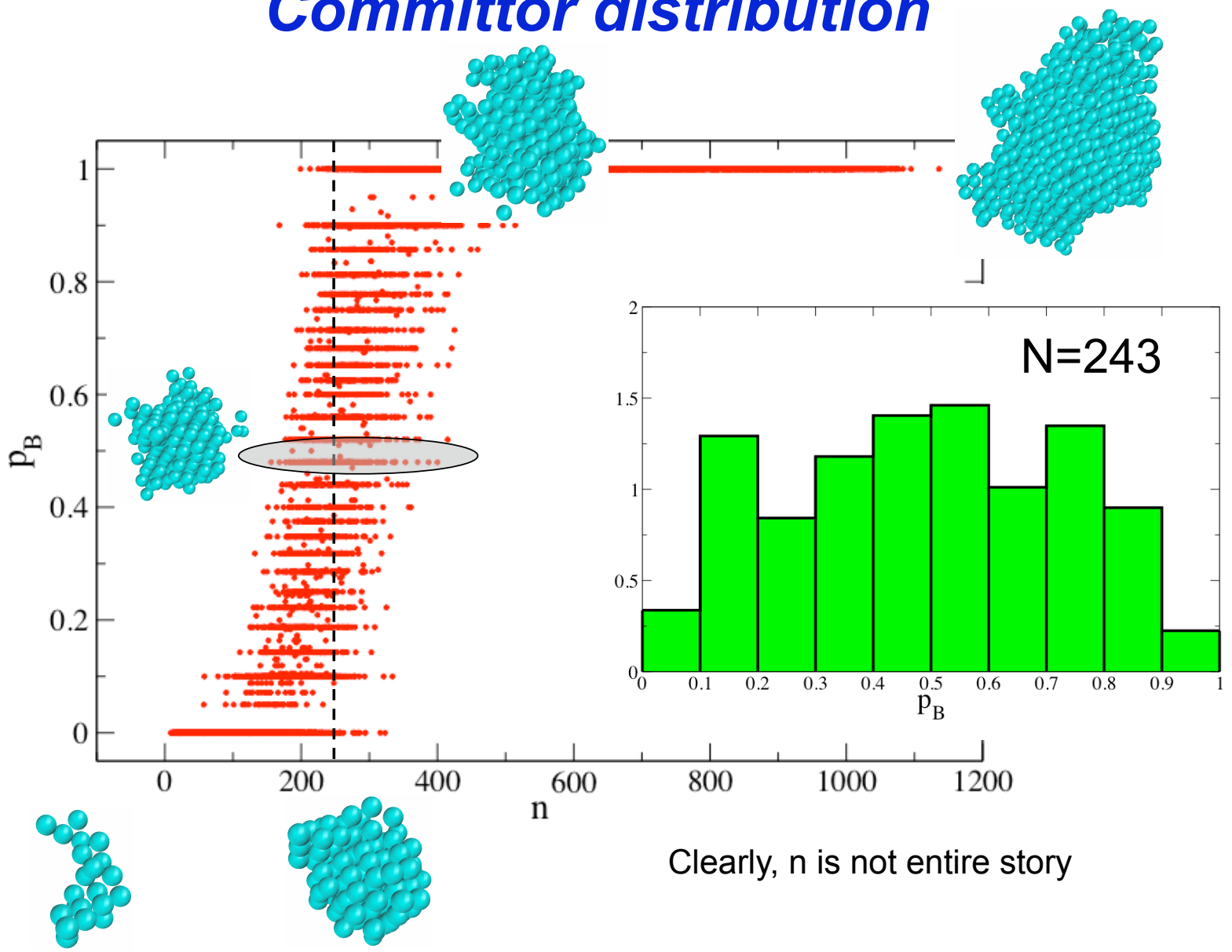


Good RC

Bad RC

analysis very expensive: requires pB histogram for every q

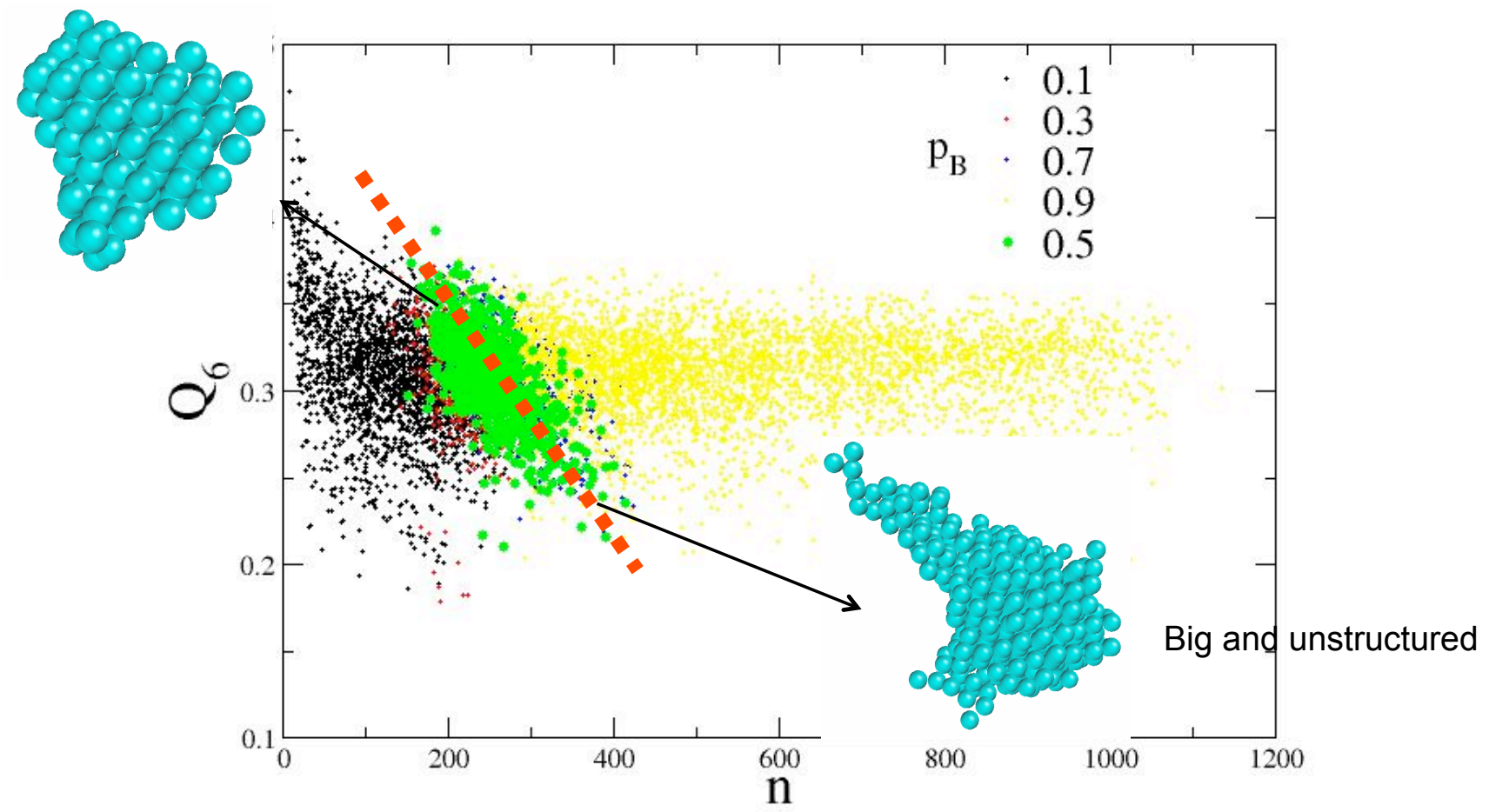
Committor distribution



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

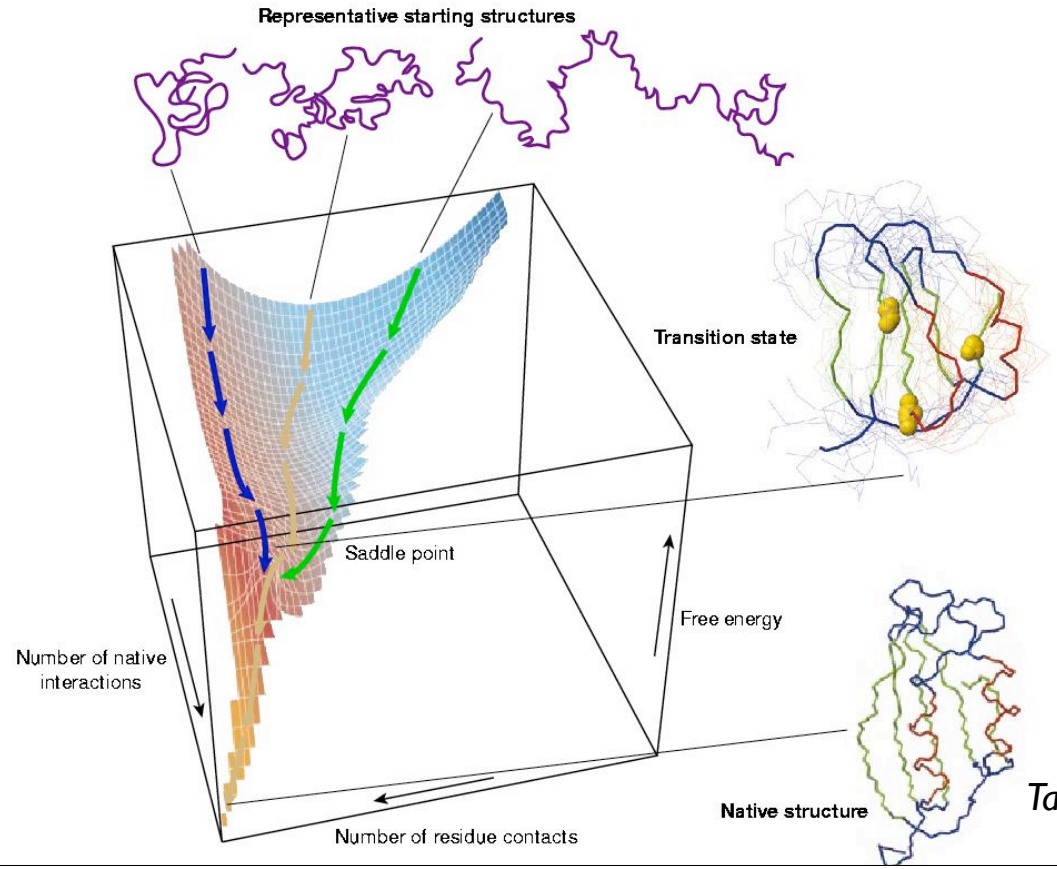
Moroni, ten Wolde, Bolhuis PRL 2005

Outline

- Part 1
  - Rare events
  - The need for unbiased transition paths
  - Sampling the path ensemble
- Part 2
  - Analyzing the path ensemble
  - Calculation of rate constants
  - Transition state ensemble and reaction coordinate
- Part 3
  - Application to protein folding

How do proteins find their native state?

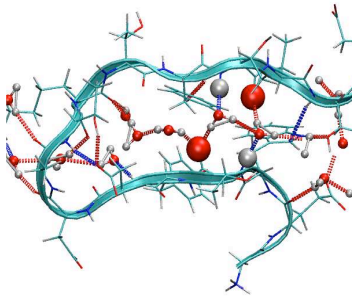
- Guided by free energy landscape
  - how is this related to folding kinetics?
  - mechanisms important to understand misfolding (Alzheimer,CJD, etc)



Taken from Dobson, Nature, 2003

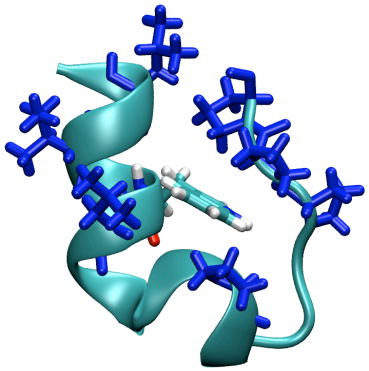
Applications of TPS to folding

GB1 hairpin



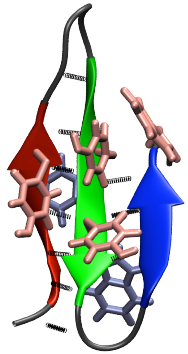
PGB PNAS 2003
Biophys J. 2005

Trp cage



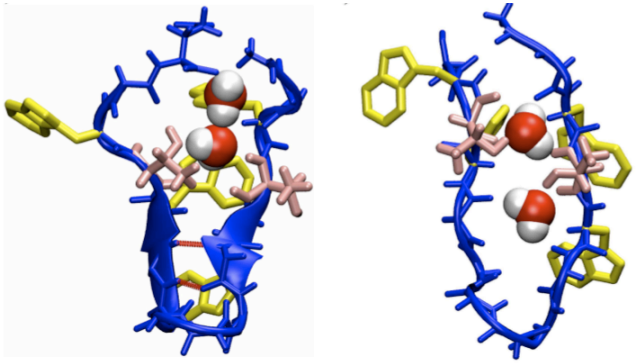
J. Juraszek, PGB PNAS 2006
Biophys. J. 95 4246 (2008)

WW domain



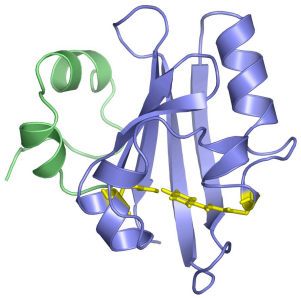
J. Juraszek, PGB Biophys J. in press

Trpzip4



J. Juraszek, PGB JPCB inpress

Photoactive yellow protein



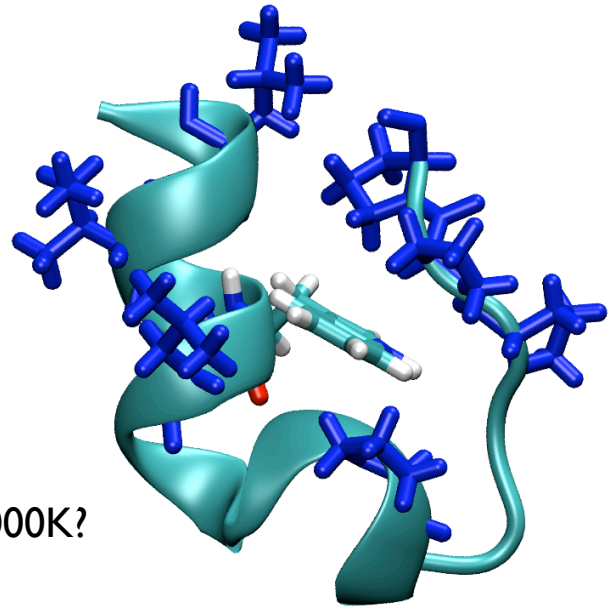
J.Vreede J. Juraszek, PGB, PNAS, in press

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*)

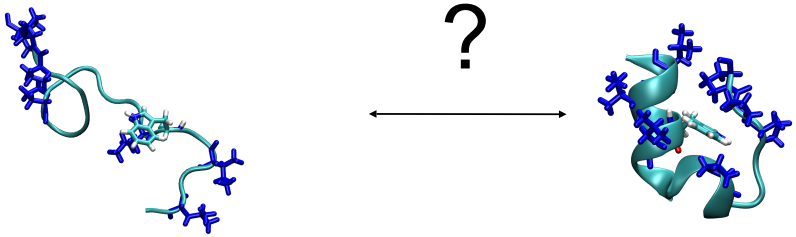


System:

1L2Y in 2800 waters

OPLSAA, PME, Nose-Hoover, GROMACS

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



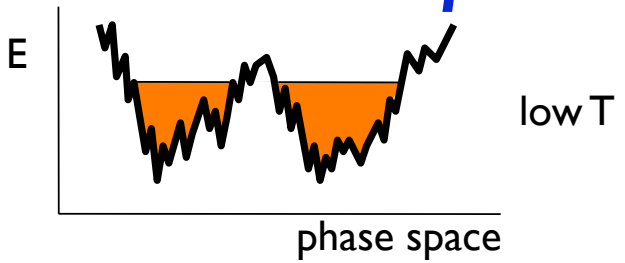
Strategy:

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

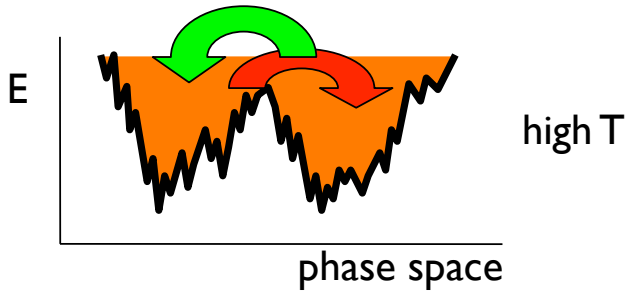
Jarek Juraszek

Replica Exchange/Parallel tempering

High barriers in energy landscape:
difficult to sample

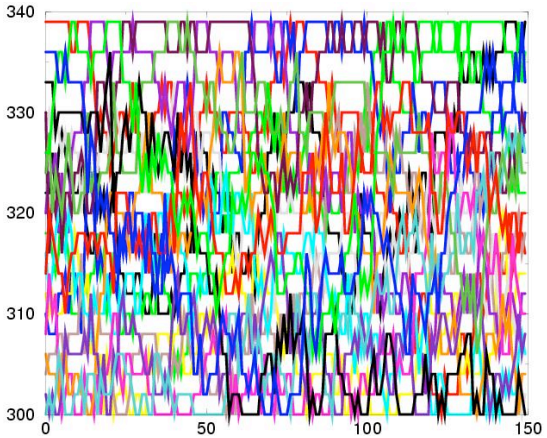


Barriers effectively low: easy to sample



REMD: m replica's running MD in the NVT ensemble at a different temperature $\beta_1, \beta_2 .. \beta_m$.

Allow swapping between replicas using Metropolis Monte Carlo



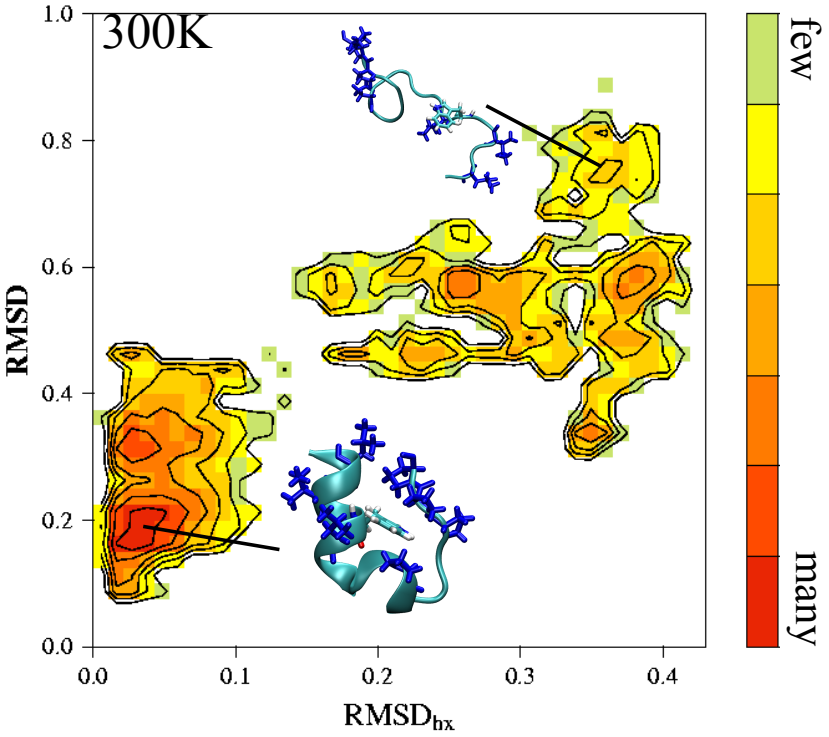
$$P_{acc}(i \leftrightarrow j) = \min\left[1, e^{(\beta_i - \beta_j)(U_i - U_j)}\right]$$

A swap between two systems of different temperatures (β_i,β_j) is accepted if their energies are close.

Advantage: no order parameters needed

Replica exchange of Trp-cage

initially folded

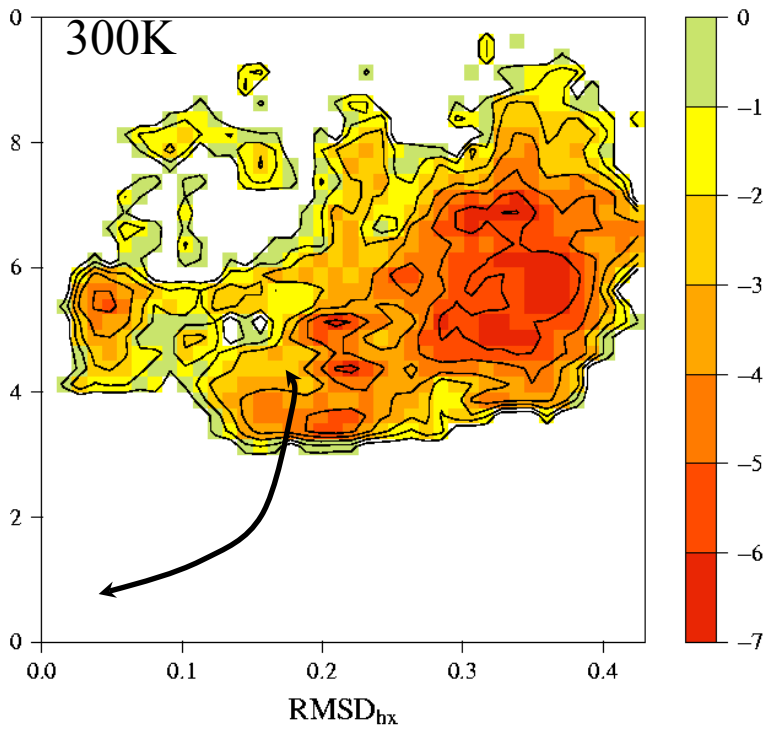


64 replicas

Temp range 272-555K

30 ns per replica

initially unfolded

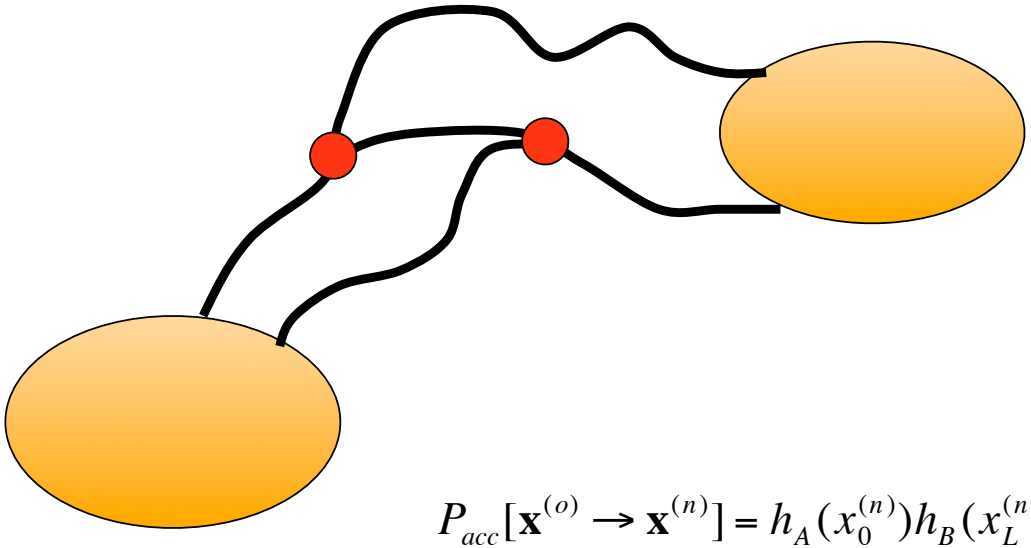


barrier towards native state not crossed:

Path sampling

Flexible one way shooting

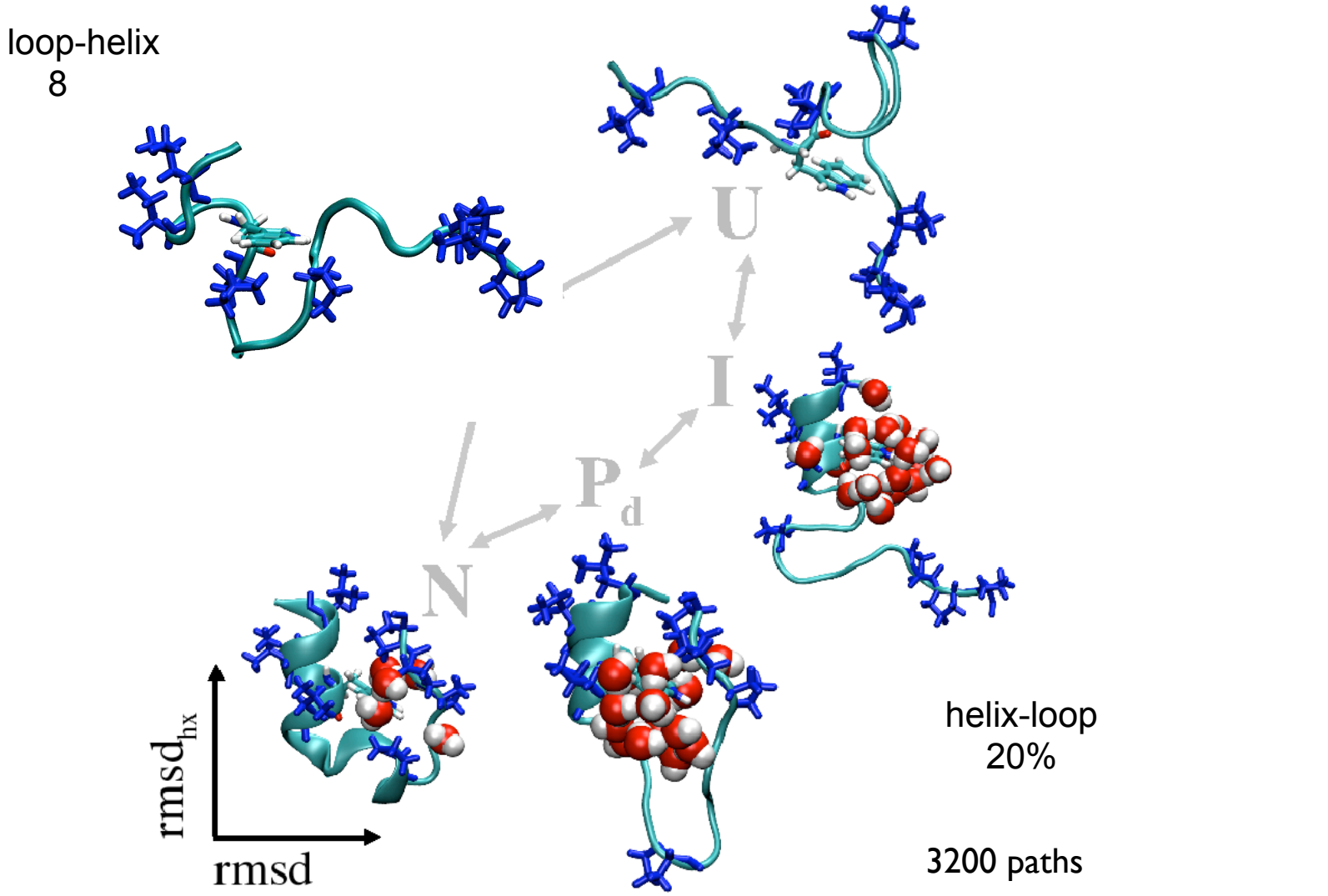
- Variable length shooting (PGB 2003, Juraszek & PGB 2006)
  - Choose new shooting point randomly from old path
  - Integrate in one direction until one stable states is reached



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

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

Parallel folding pathways



3200 paths

J. Juraszek, PGB PNAS 2006

N-L rates for Trp-cage

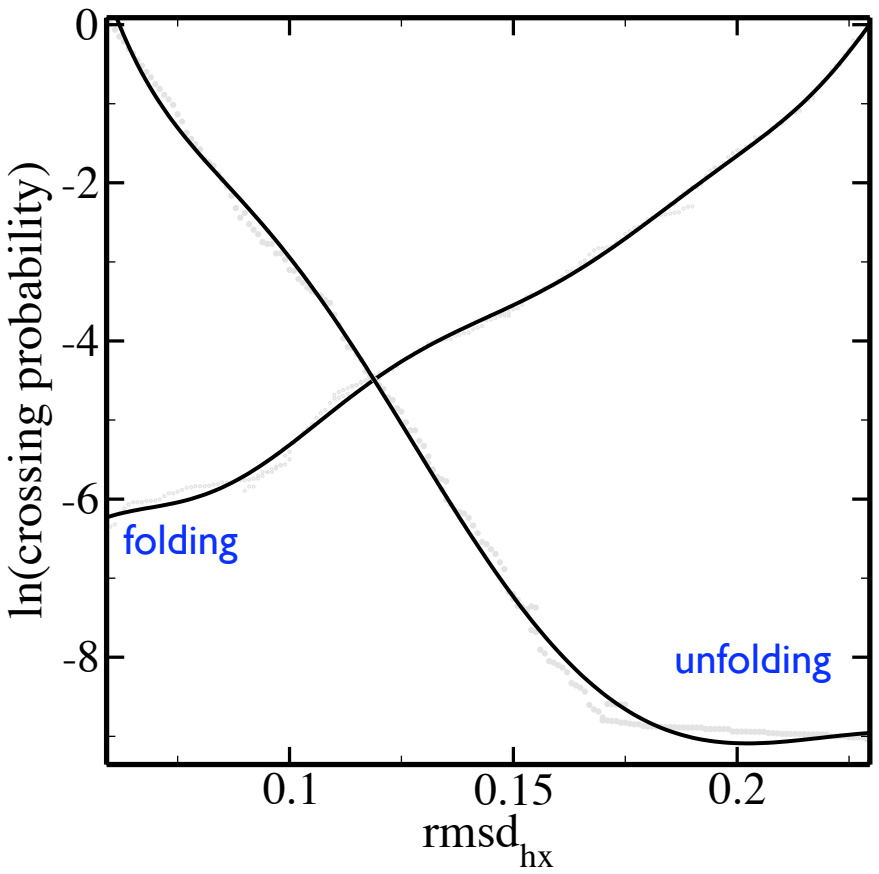
TIS, 6 interfaces

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

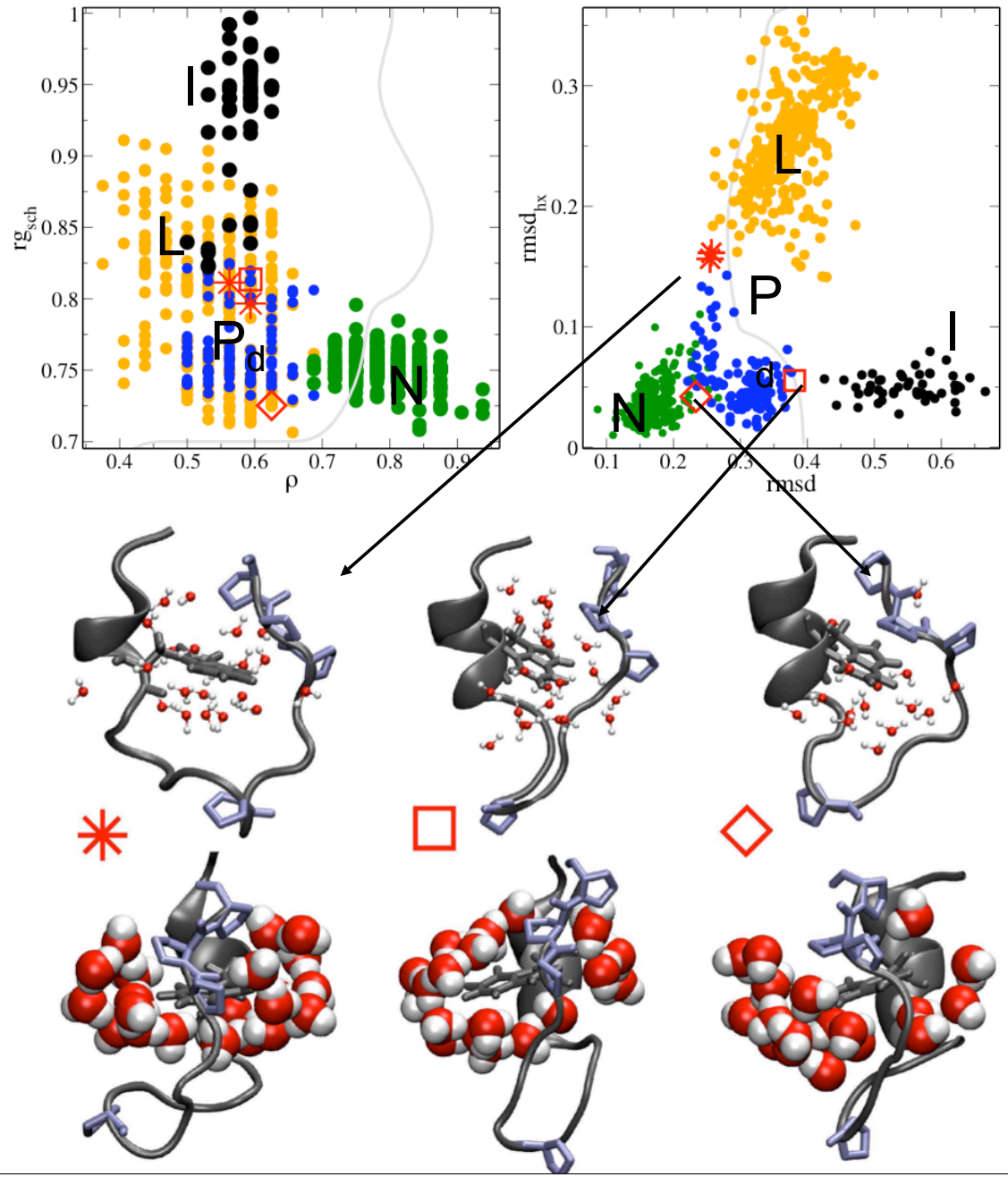
Exp: k_{unf} = 0.08 μs^{-1}

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

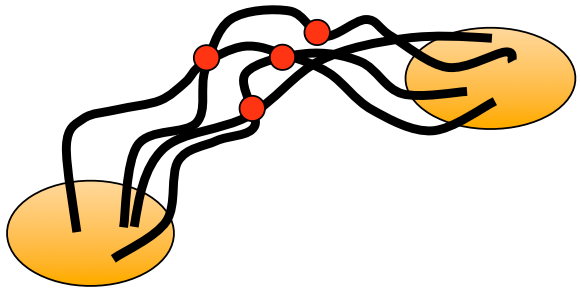
$$p_B(x) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(q(x)\right]$$

- parametrize r as linear combination of q

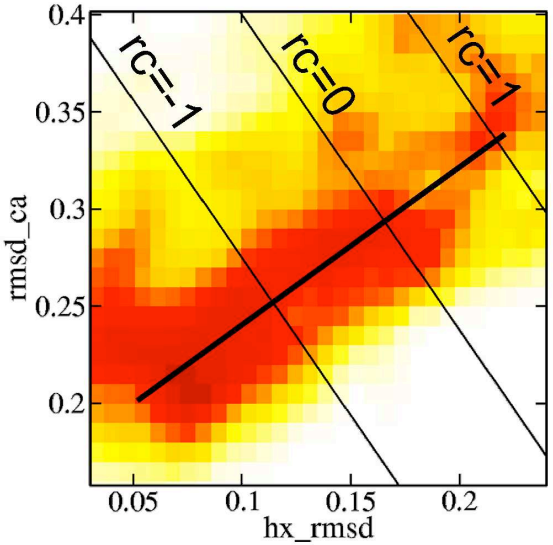
$$r(\mathbf{x}) = \sum_i \alpha_i q(\mathbf{x}) + \alpha_0$$

- 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

Conclusion

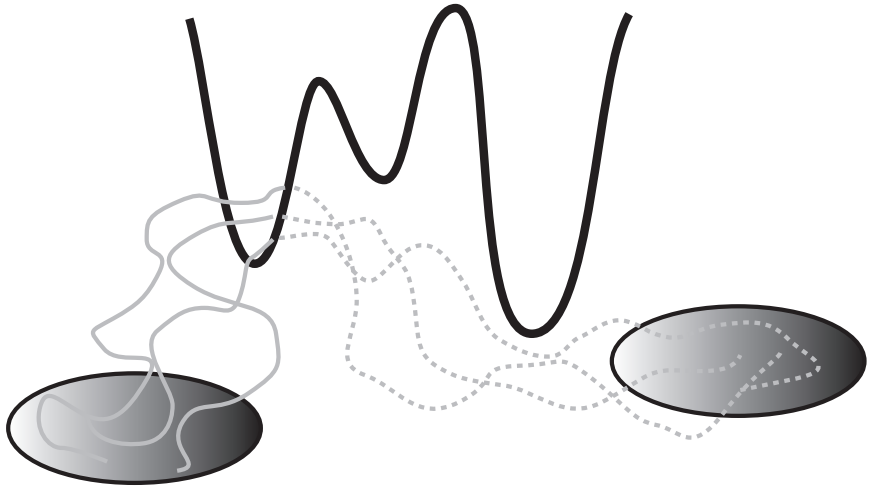
- TPS, TIS
  - can be used for wide range of rare event processes
  - has no need for reaction coordinate, just stable state definitions
  - gives true, unbiased molecular dynamical reaction pathways
  - do not assume reaction tube
  - yields correct rate constant, no suffering from low transmission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

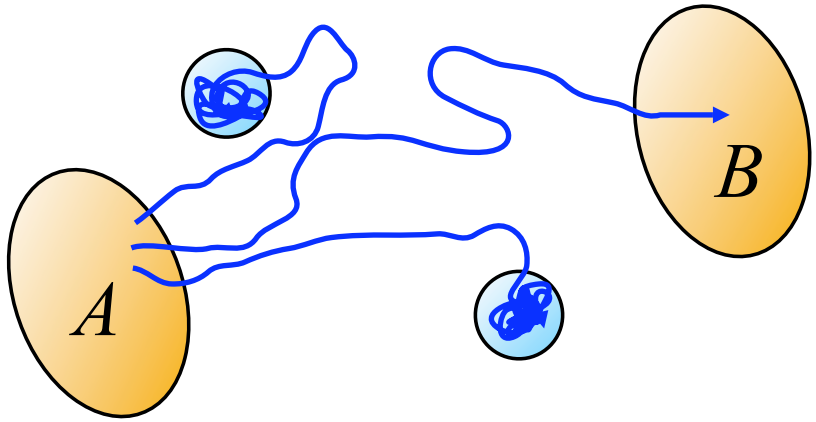
Challenges for path sampling

- 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).

Acknowledgments

- Group members
  - Jocelyne Vreede
  - Marieke Schor
  - Lizhe Zhu
  - Weina Du
  - Kristen Marino
  - Rosanne Zeiler

- Ex-group members
  - Jarek Juraszek
  - Elske Leenders
  - Jutta Rogal

- Funding: FOM/NWO/EU

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

several PhD positions open