Molecular rare event simulations

Peter Bolhuis
van ‘t Hoff institute for Molecular Sciences
University of Amsterdam, The Netherlands

Outline

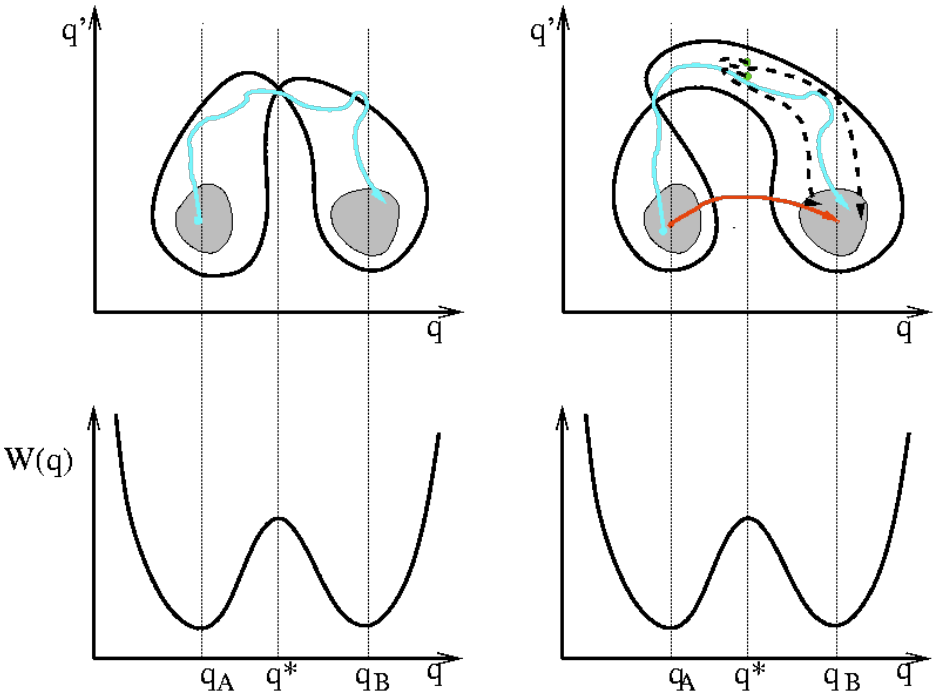
- Part 1
  - Rare events
  - The need for unbiased transition paths
  - Sampling the path ensemble

- Part 2
  - Analyzing the path ensemble
  - Calculation of rate constants
  - Transition state ensemble and reaction coordinate

- Part 3
  - Application to protein folding

The reaction coordinate problem

If RC is not correct, transition states, rates and mechanism 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, e.g.

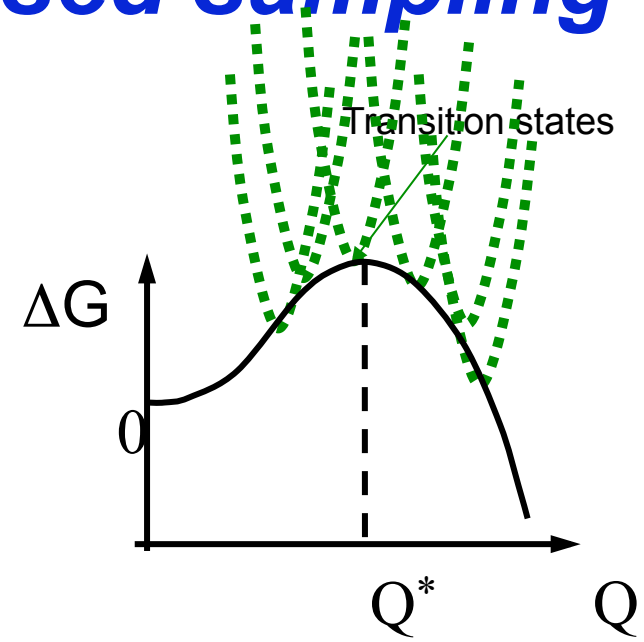
- –Parallel replica, hyperdynamics, metadynamics, etc
- –Nudged Elastic Band, string methods, etc
- –Transition path sampling

Breakdown of biased sampling

Free energy landscape has barrier

Calculate FE by umbrella sampling

- –by applying biasing potential
- –as a function of order parameter Q

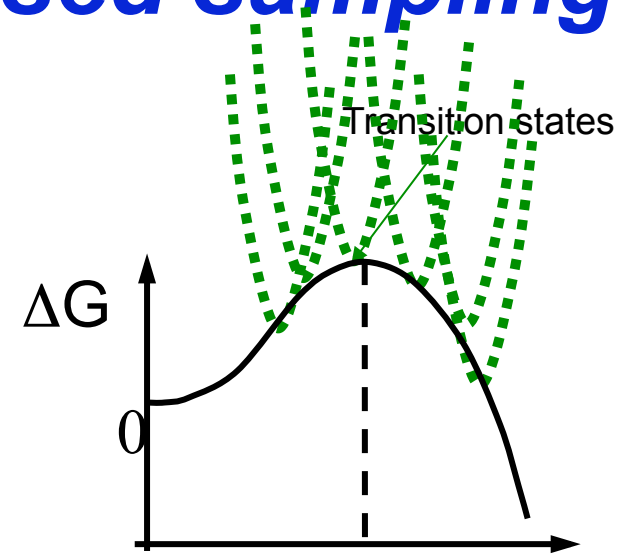


Breakdown of biased sampling

Free energy landscape has barrier

Calculate FE by umbrella sampling

- –by applying biasing potential
- –as a function of order parameter Q

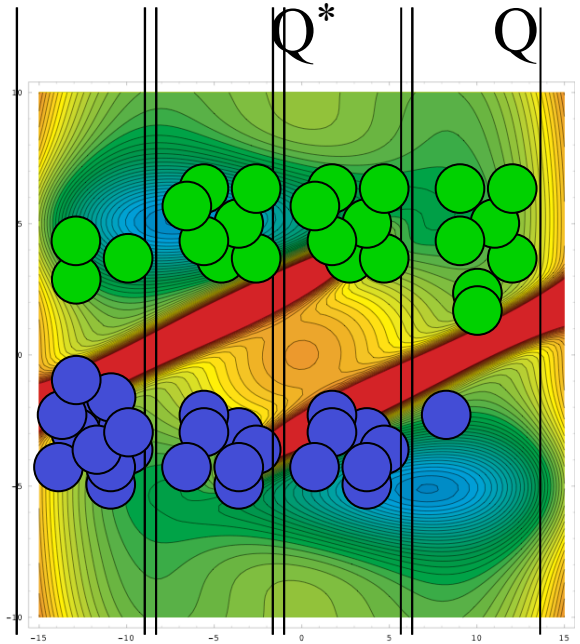


Finding the correct reaction coordinate is difficult!

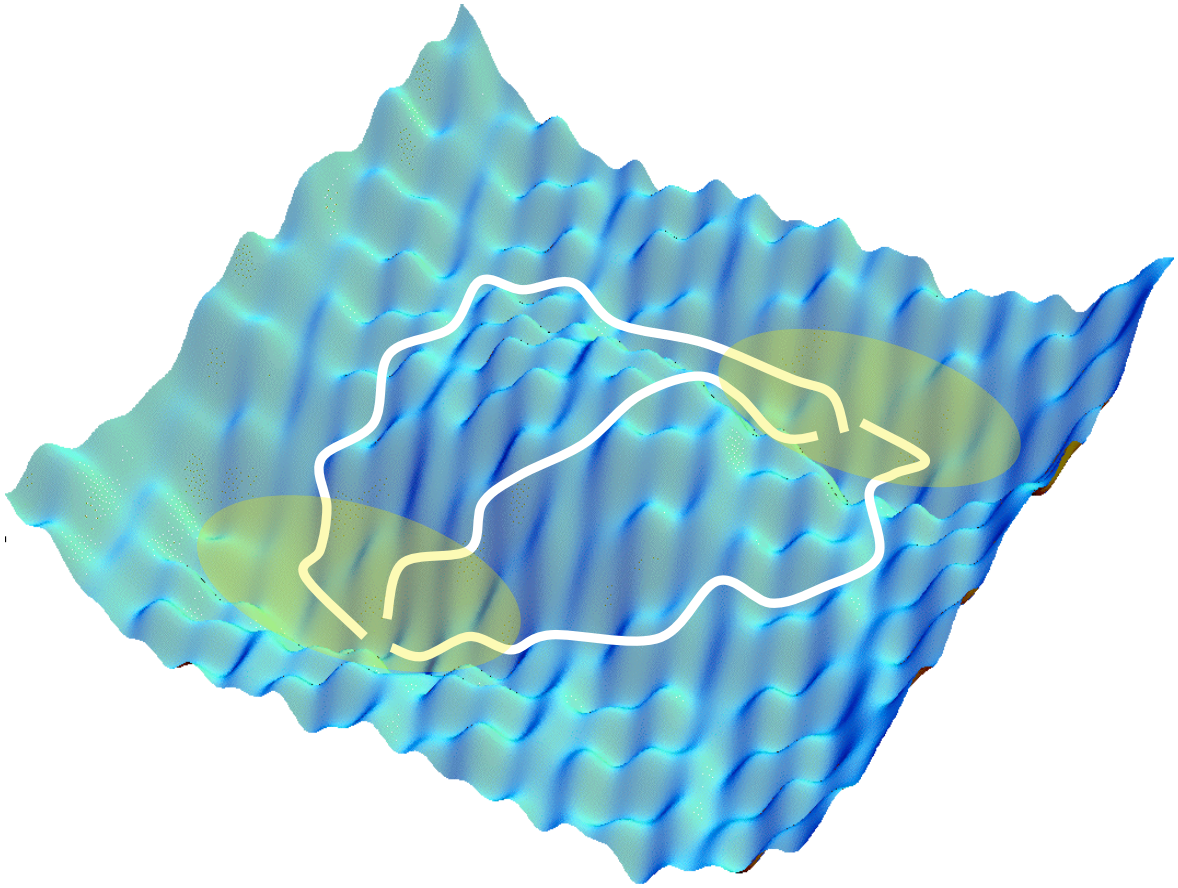
Rare events by biasing along reaction coordinate

Hysteresis because reaction coordinate not known in advance

Mechanism and kinetics are not correct



Two ended methods



Methods that take the entire path and fix the begin and end point

Nudged Elastic Band

Suitable for finding saddle points on complex PES

chain of states $\{r^{(0)}, r^{(1)}, \ldots, r^{(M)}\}$

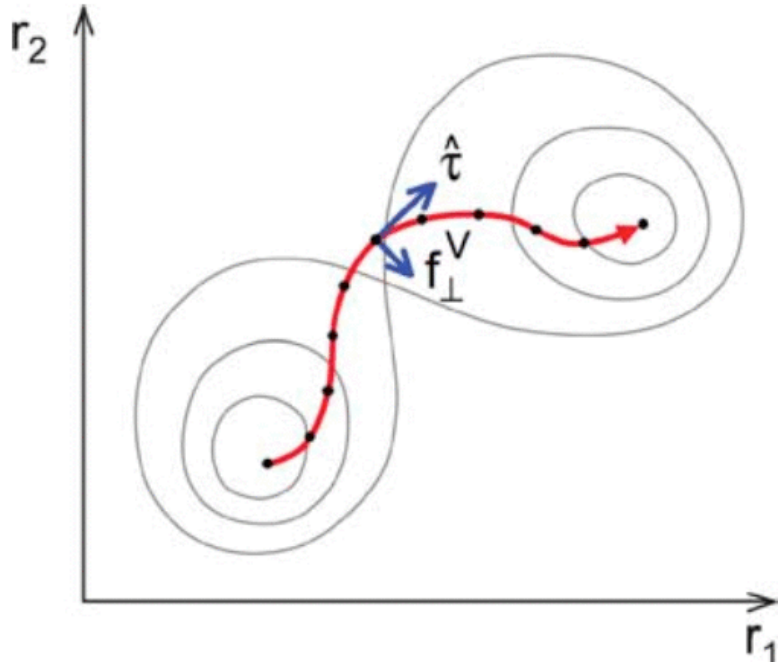
define object function

$$S[r^{(0)}, r^{(1)}, \ldots, r^{(M)}] = \sum_{i=0}^{M} V[r^{(i)}] + \frac{k}{2} \sum_{i=1}^{M} [r^{(i)} - r^{(i-1)}]^2$$

that should be minimized.

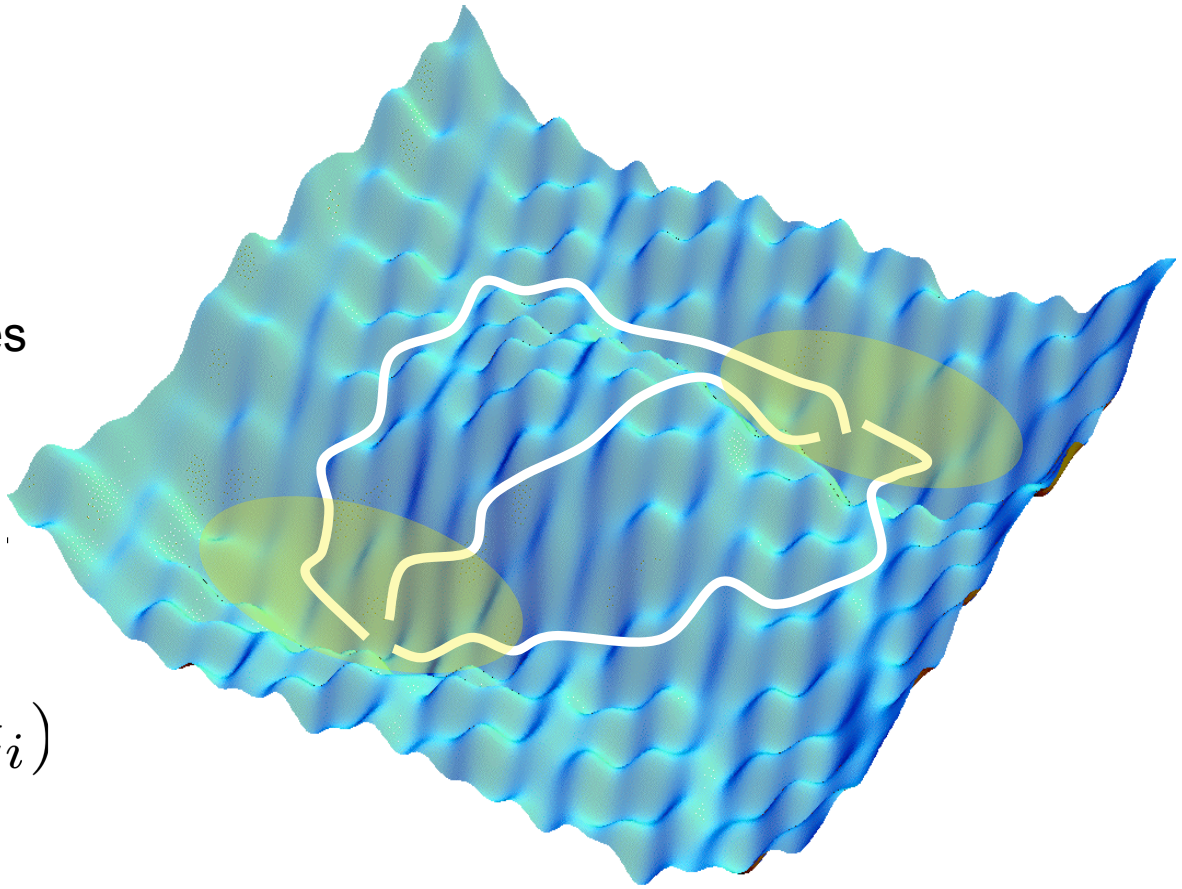
$$f^{\mathrm{NEB}}[r^{(i)}] = f_{\perp}^{V}[r^{(i)}] + f_{\parallel}^{s}[r^{(i)}]$$

$$\hat{\tau} = \frac{r^{(i+1)} - r^{(i)}}{|r^{(i+1)} - r^{(i)}|}$$



H. Jonsson et al.
WorldScientific,Singapore, p.385(2000)

Minimize the action



To get dynamical trajectories minimize action

$$S_{cont} = \int_{t_0}^{t_1} dt L(t)$$

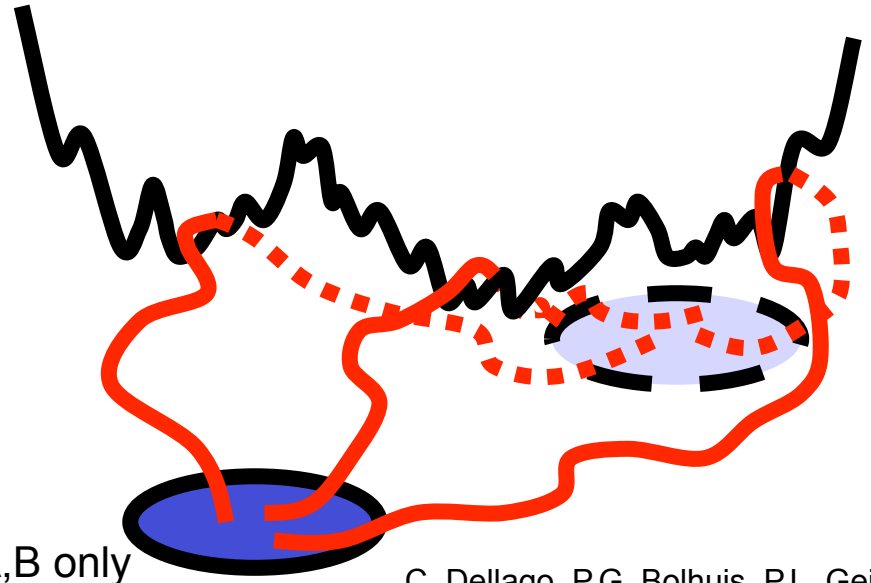
$$S_{disc} = \Delta t \sum_{i=0}^{i_{max}} \mathcal{L}(t_i)$$

$$\mathcal{L}(r(t)) = \sum_{i=1}^{N} m_i \dot{r}_i^2 - U(r_1, r_2, \ldots r_N)$$

problem : only one trajectory is obtained

Transition path sampling

Samples the path ensemble:
all trajectories that lead over barrier



C. Dellago, P.G. Bolhuis, P.L. Geissler
Adv. Chem. Phys. **123**, 1 2002

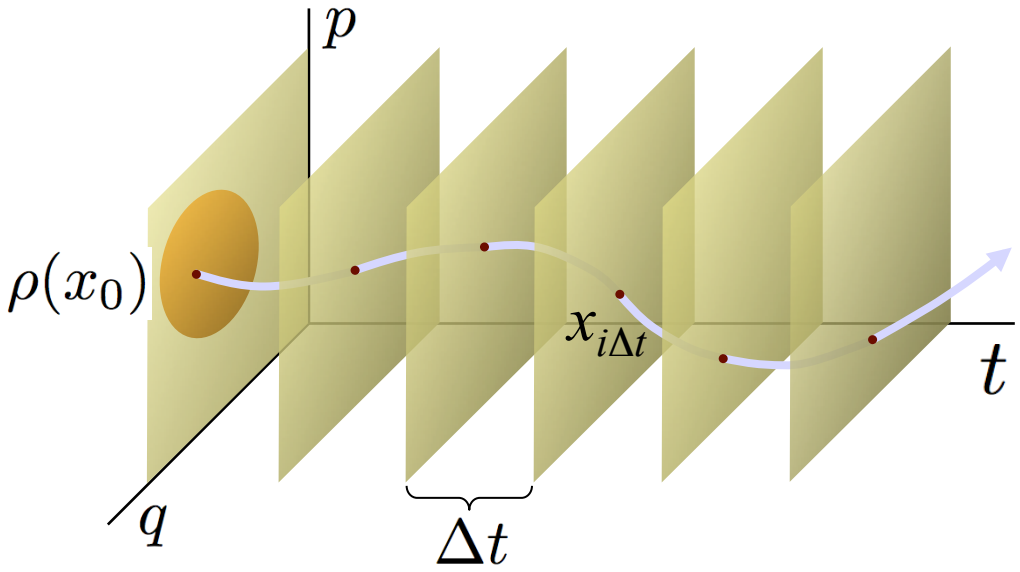
- Sampling by Monte Carlo
- Requires definition of stable states A,B only
- Results in ensemble of pathways
- Reaction coordinate is a result of simulation not an input
- Allows for calculation of rate constants

Apply when process of interest
  - is a rare event
  - is complex and reaction coordinate is not known

Examples: autodissociation of water, organic reactions in solution, protein folding

Path probability density

$$x(\mathcal{T}) \equiv \{x_0, x_{\Delta t}, x_{2\Delta t}, \ldots, x_{\mathcal{T}}\}$$ Path = Sequence of states



$$\mathcal{P}[x(\mathcal{T})] = \rho(x_0) \prod_{i=0}^{\mathcal{T}/\Delta t - 1} p(x_{i\Delta t} \to x_{(i+1)\Delta t})$$

Transition probabilities

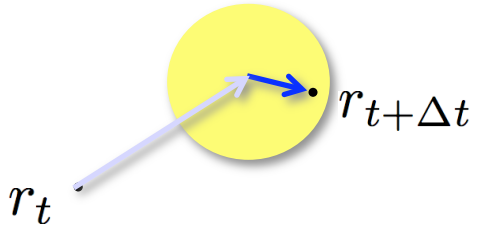
Initial conditions

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

Microcanonical: $\rho(x) = \delta[E - \mathcal{H}(x)]/g(E)$ $\quad g(E) = \int dx \delta[E - \mathcal{H}(x)]$

Brownian dynamics

$$m\gamma\dot{r} = -\frac{\partial V(r)}{\partial r} + \mathcal{F}$$



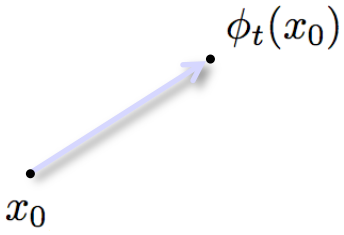
$$p(r_t \to r_{t+\Delta t}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(r_{t+\Delta t} - r_t + \frac{\Delta t}{\gamma m}\frac{\partial V}{\partial r})^2}{2\sigma^2}\right\}$$

$$\sigma^2 = \frac{2k_B T}{m\gamma}\Delta t$$

Newtonian dynamics

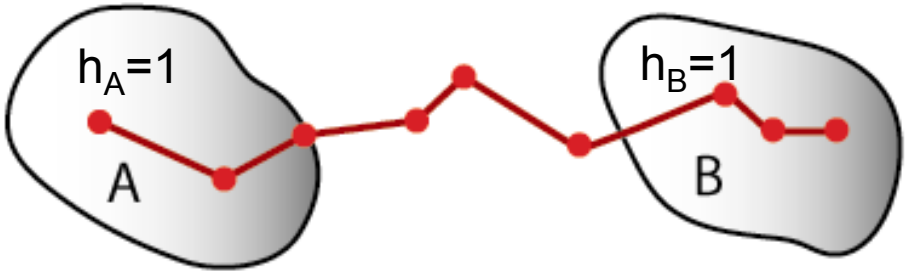
$$\dot{r} = \frac{\partial\mathcal{H}(r,p)}{\partial p}$$

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



$$p(x_t \to x_{t+\Delta t}) = \delta[x_{t+\Delta t} - \phi_{\Delta t}(x_t)]$$

Transition path ensemble

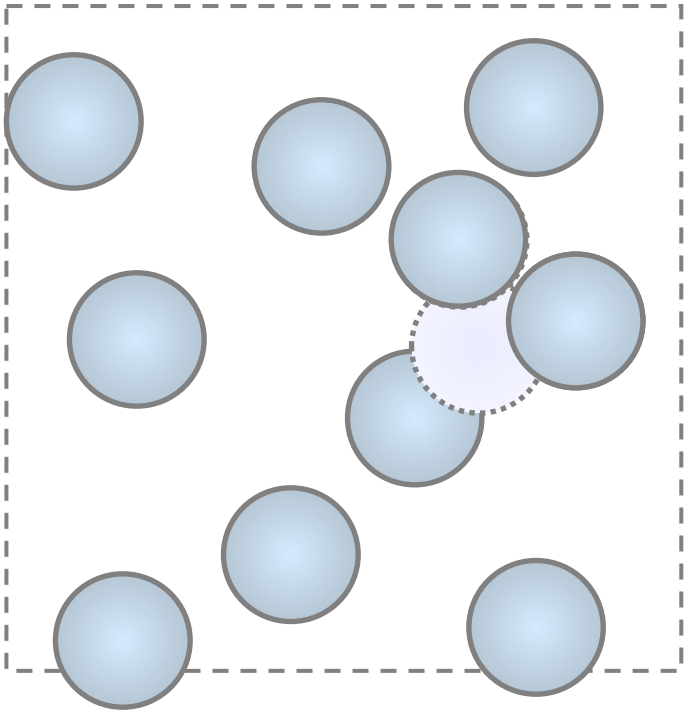


$$\mathcal{P}_{AB}[x(\mathcal{T})] \equiv h_A(x_0)\mathcal{P}[x(\mathcal{T})]h_B(x_\mathcal{T})/Z_{AB}(\mathcal{T})$$

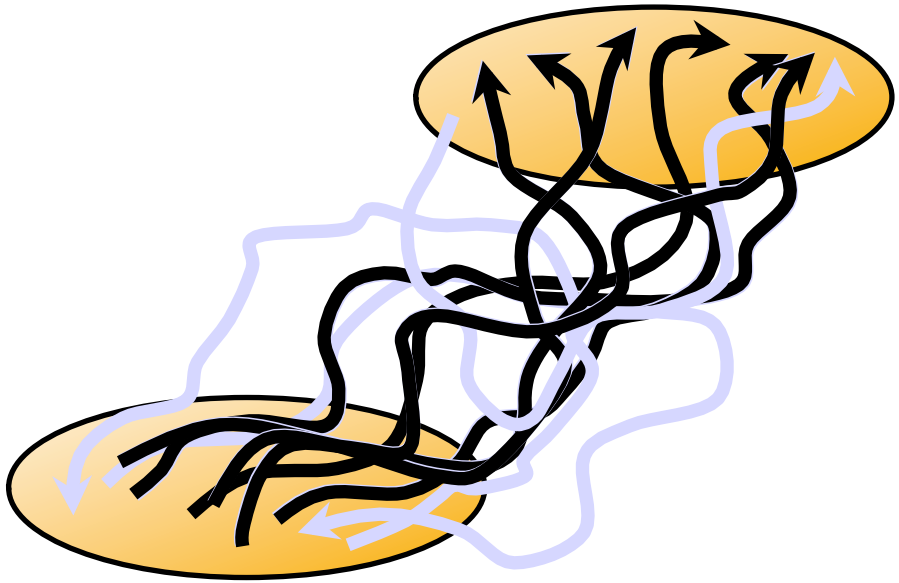
$$Z_{AB}(\mathcal{T}) \equiv \int \mathcal{D}x(\mathcal{T})\, h_A(x_0)\mathcal{P}[x(\mathcal{T})]h_B(x_\mathcal{T})$$

$$\int \mathcal{D}x(\mathcal{T}) \equiv \int \cdots \int dx_0 dx_{\Delta t} dx_{2\Delta t} \cdots dx_\mathcal{T}$$

Sampling the path ensemble



Monte Carlo simulation

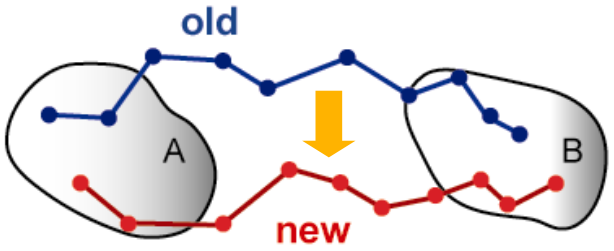


transition path sampling

Metropolis MC of pathways

1. Generate **new** path from **old** one

$$x^{(o)}(\mathcal{T}) \longrightarrow x^{(n)}(\mathcal{T})$$



2. Accept **new** path according to **detailed balance**:

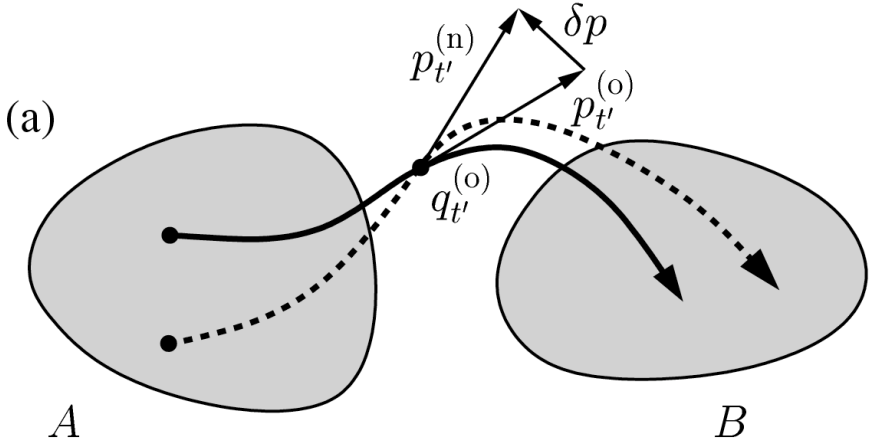
$$\mathcal{P}_{AB}[x^{(\mathrm{o})}(\mathcal{T})]\pi[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = \mathcal{P}_{AB}[x^{(\mathrm{n})}(\mathcal{T})]\pi[x^{(\mathrm{n})}(\mathcal{T}) \to x^{(\mathrm{o})}(\mathcal{T})]$$

$$\pi[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = P_{\mathrm{gen}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] \times P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})]$$

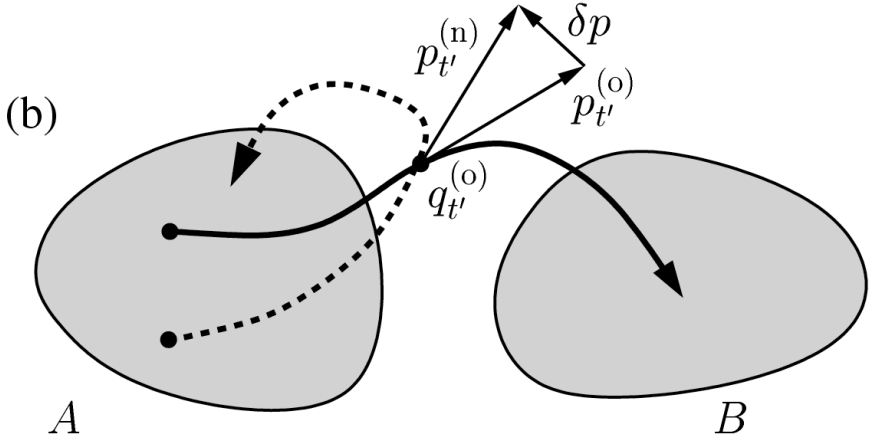
3. Satisfy detailed balance with the **Metropolis rule**:

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

Shooting moves



accept



reject

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

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

Shooting algorithm

$$P^{\mathrm{f}}_{\mathrm{gen}}[x^{\mathrm{o}}(\mathcal{T}) \to x^{\mathrm{n}}(\mathcal{T})] = \prod_{i=t'/\Delta t}^{\mathcal{T}/\Delta t-1} p\left(x^{(\mathrm{n})}_{i\Delta t} \to x^{(\mathrm{n})}_{(i+1)\Delta t}\right)$$

$$P^{\mathrm{b}}_{\mathrm{gen}}[x^{\mathrm{o}}(\mathcal{T}) \to x^{\mathrm{n}}(\mathcal{T})] = \prod_{i=1}^{t'/\Delta t} \bar{p}\left(x^{(\mathrm{n})}_{i\Delta t} \to x^{(\mathrm{n})}_{(i-1)\Delta t}\right)$$

$$P_{\mathrm{gen}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = p_{\mathrm{gen}}[x^{(\mathrm{o})}_{t'} \to x^{(\mathrm{n})}_{t'}] \prod_{i=t'/\Delta t}^{\mathcal{T}/\Delta t-1} p\left(x^{(\mathrm{n})}_{i\Delta t} \to x^{(\mathrm{n})}_{(i+1)\Delta t}\right) \times \prod_{i=1}^{t'/\Delta t} \bar{p}\left(x^{(\mathrm{n})}_{i\Delta t} \to x^{(\mathrm{n})}_{(i-1)\Delta t}\right)$$

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = h_A[x^{(\mathrm{n})}_0] h_B[x^{(\mathrm{n})}_{\mathcal{T}}] \min\left[1, \frac{\rho\left(x^{(\mathrm{n})}_0\right)}{\rho\left(x^{(\mathrm{o})}_0\right)} \prod_{i=0}^{t'/\Delta t-1} \frac{p\left(x^{(\mathrm{n})}_{i\Delta t} \to x^{(\mathrm{n})}_{(i+1)\Delta t}\right)}{\bar{p}\left(x^{(\mathrm{n})}_{(i+1)\Delta t} \to x^{(\mathrm{n})}_{i\Delta t}\right)} \times \frac{\bar{p}\left(x^{(\mathrm{o})}_{(i+1)\Delta t} \to x^{(\mathrm{o})}_{i\Delta t}\right)}{p\left(x^{(\mathrm{o})}_{i\Delta t} \to x^{(\mathrm{o})}_{(i+1)\Delta t}\right)}\right]$$

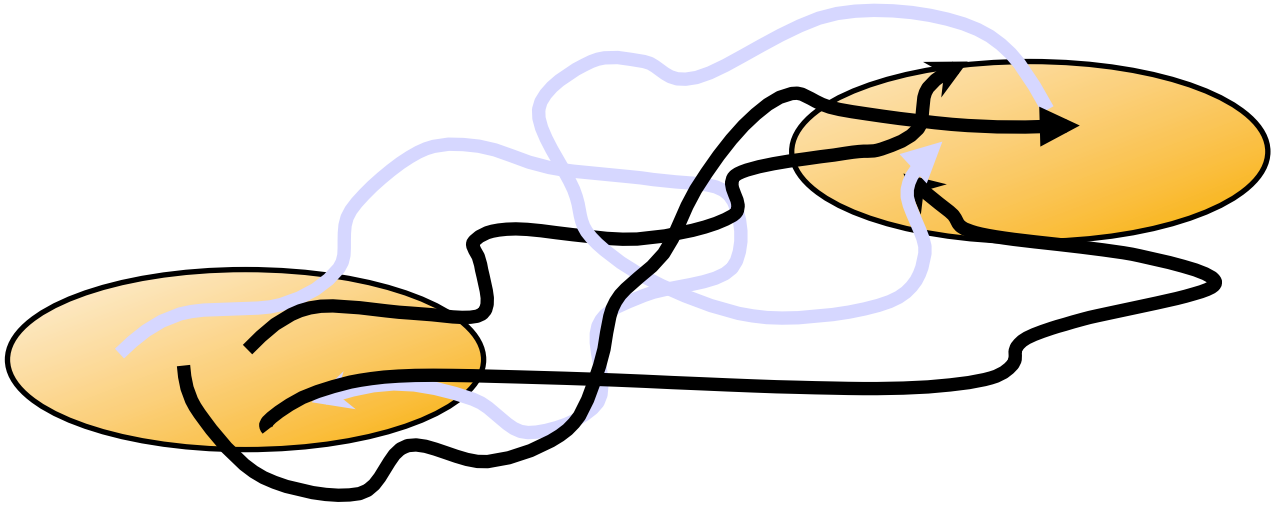
$$\frac{p(x \to y)}{\bar{p}(y \to x)} = \frac{\rho_0(y)}{\rho_0(x)}$$

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = h_A[x^{(\mathrm{n})}_0] h_B[x^{(\mathrm{n})}_{\mathcal{T}}] \min\left[1, \frac{\rho(x^{(\mathrm{n})}_{t'})}{\rho(x^{(\mathrm{o})}_{t'})}\right]$$

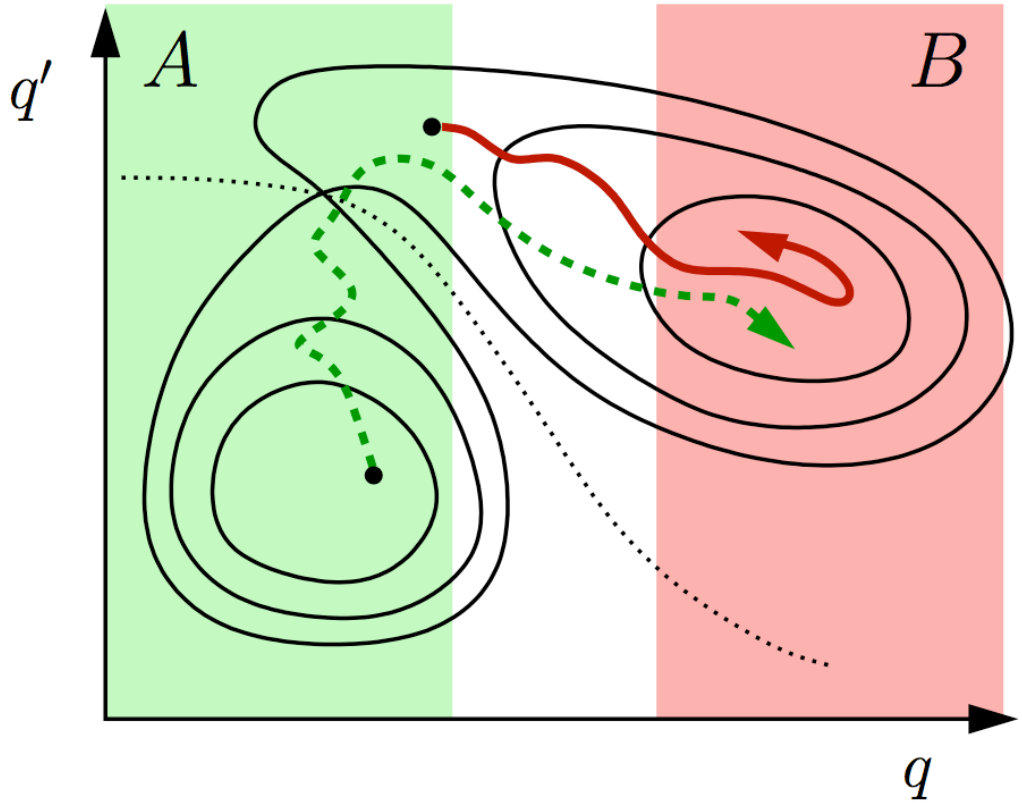
$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(\mathcal{T}) \to x^{(\mathrm{n})}(\mathcal{T})] = h_A[x^{(\mathrm{n})}_0] h_B[x^{(\mathrm{n})}_{\mathcal{T}}]$$

Standard TPS algorithm

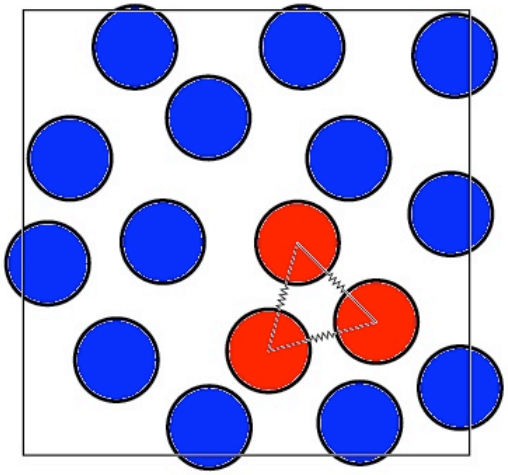
- take existing path
- choose random time slice *t*
- change momenta slightly at *t*
- integrate forward and backward in time to create new path of length L
- accept if A and B are connected, otherwise reject and retain old path
- calculate averages
- repeat

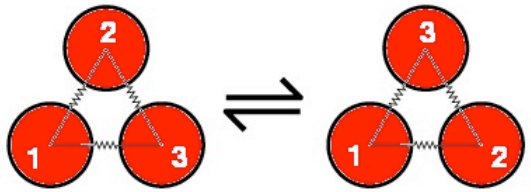


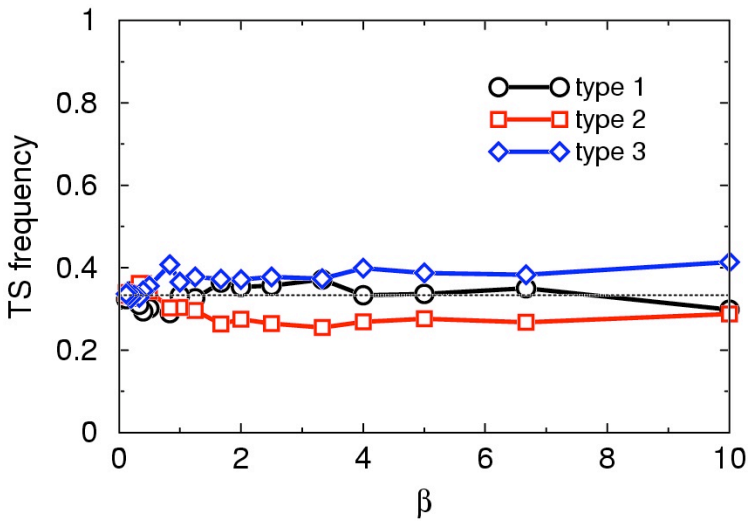
Definition of the stable states



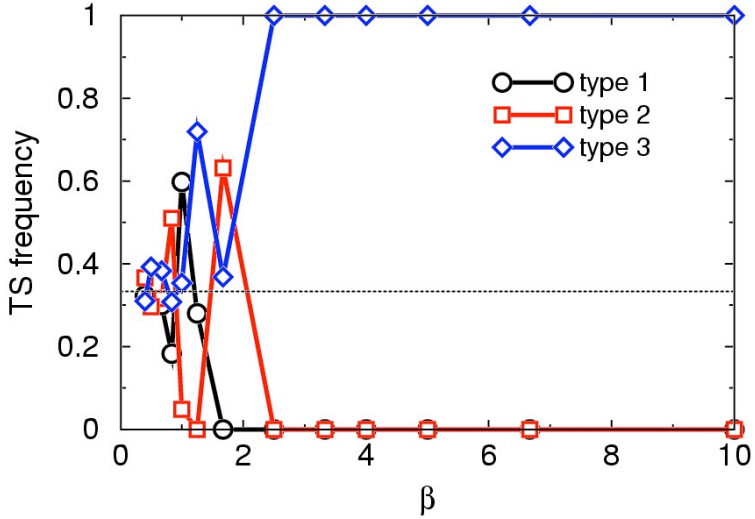
Parallel tempering





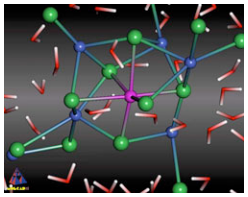


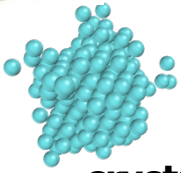
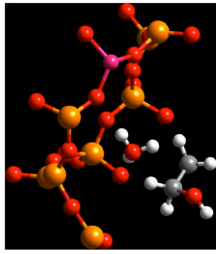
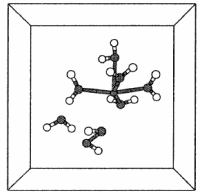
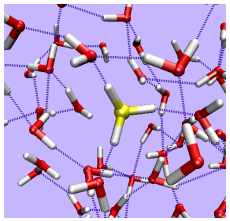
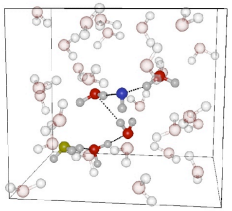
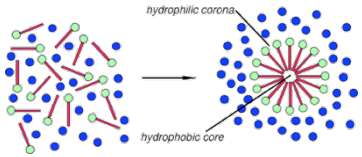
with parallel tempering

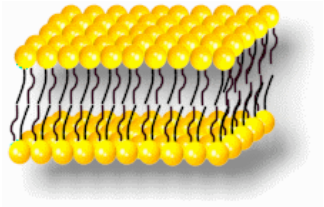
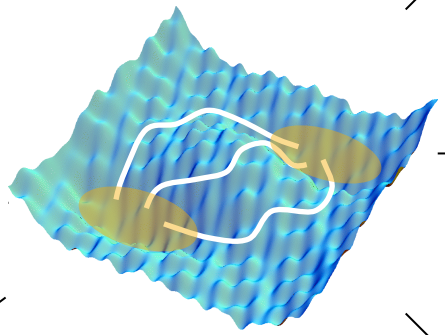
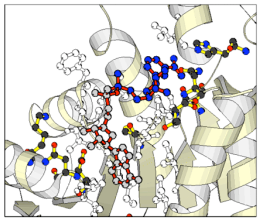
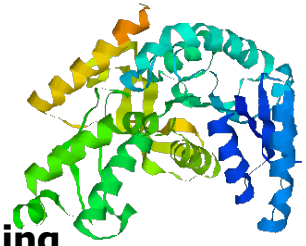
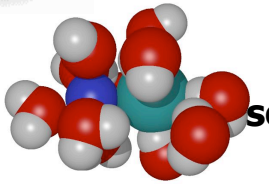
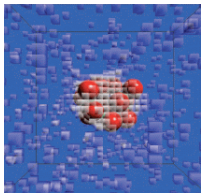
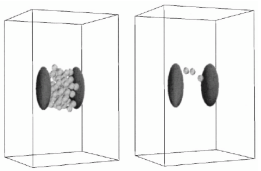
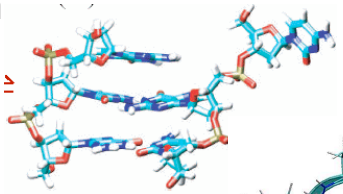
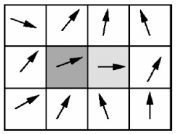
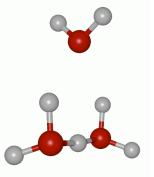
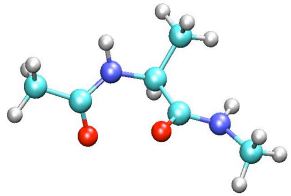
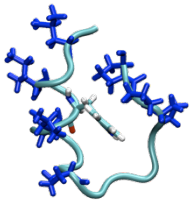
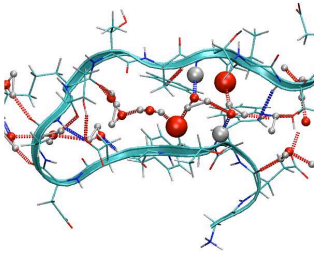
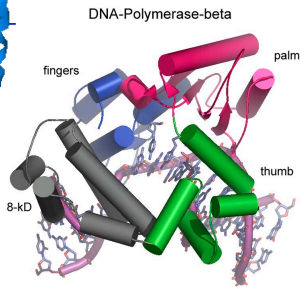


without parallel tempering

T. J. H. Vlugt and B. Smit, Phys. Chem. Comm. 2, 1 (2001).



crystallisation

catalysis

reactions

complex fluids

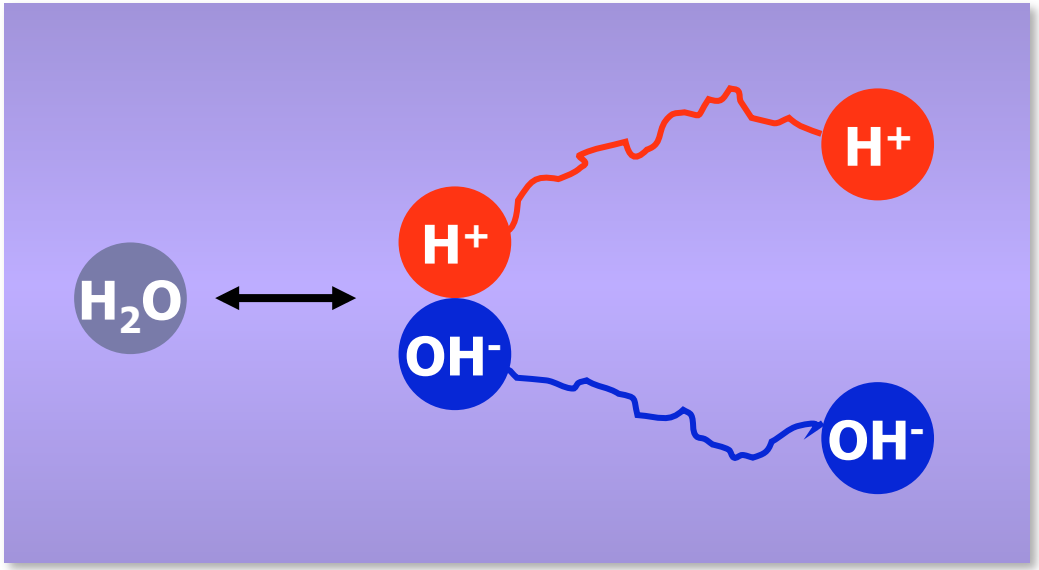
enzyme reactions

solvent effects

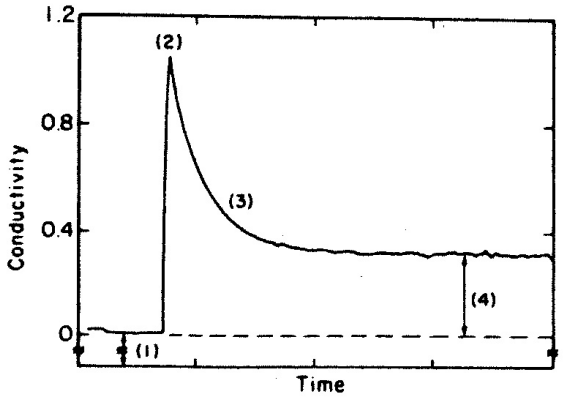
folding & binding

isomerization

Autoionization in liquid water



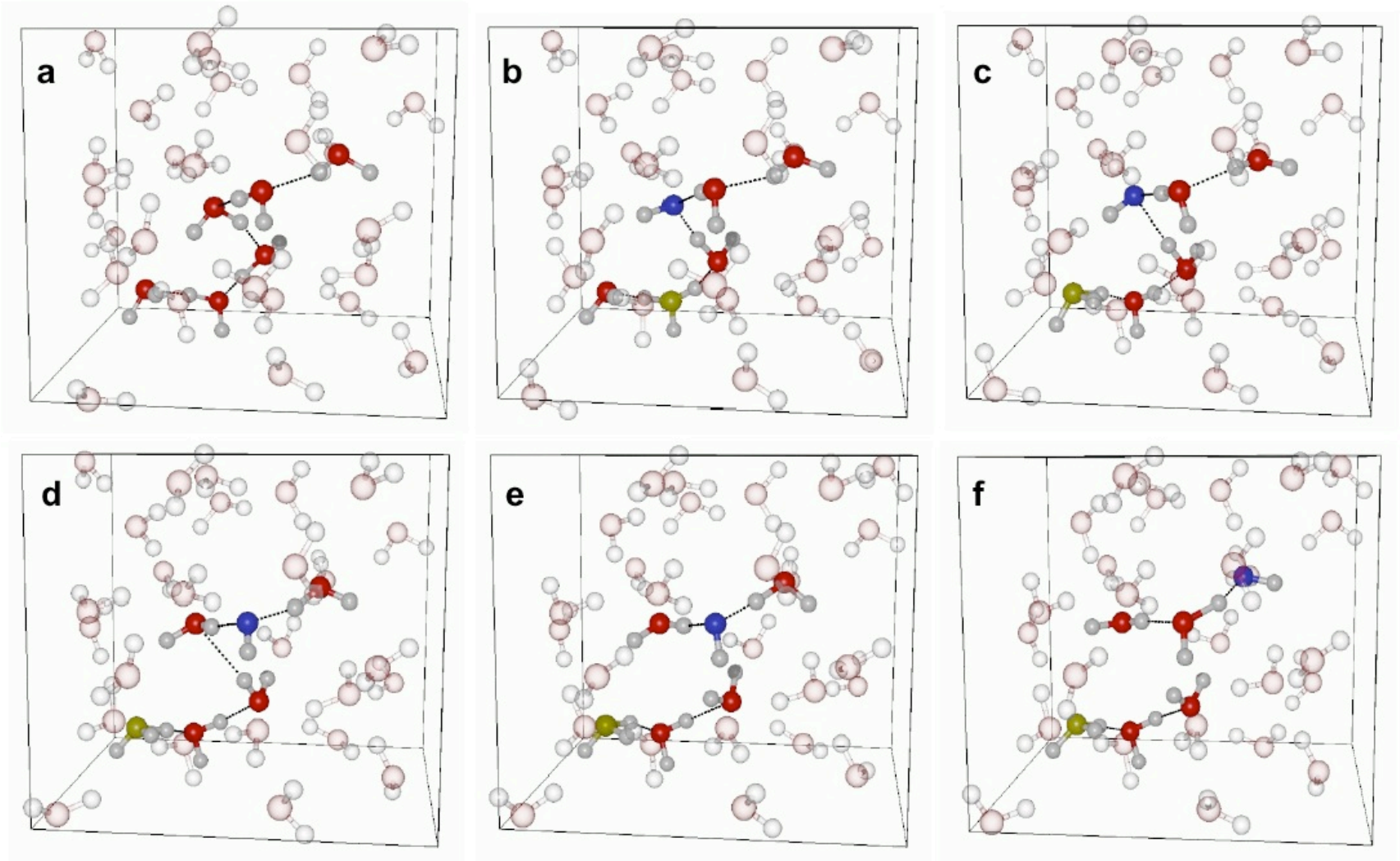
$$H_2O \rightleftharpoons OH^- + H^+$$



Average life time > 10 h

M. Eigen and L. De Maeyer, *Z. Elektrochemie* **59**, 987 (1955)

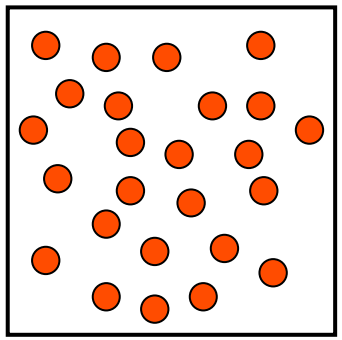
Autoionization in liquid water



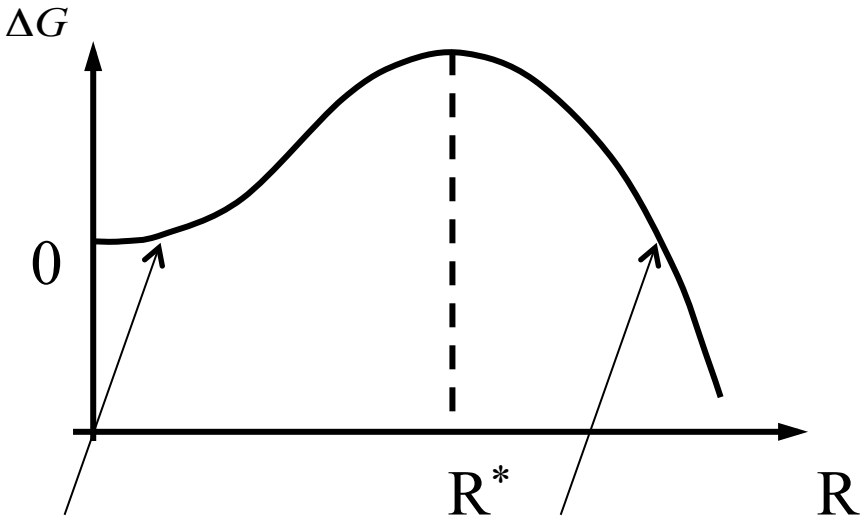
32 waters, BLYP functional CPMD code
10 uncorrelated trajectories of 200 fs,

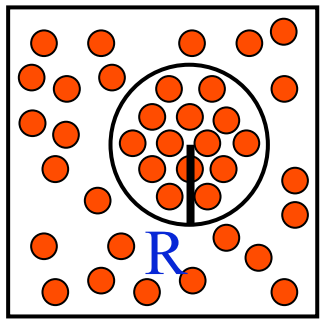
P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, M. Parrinello, Science ***291****, 2121 (2001).*

Classical nucleation (1926)



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

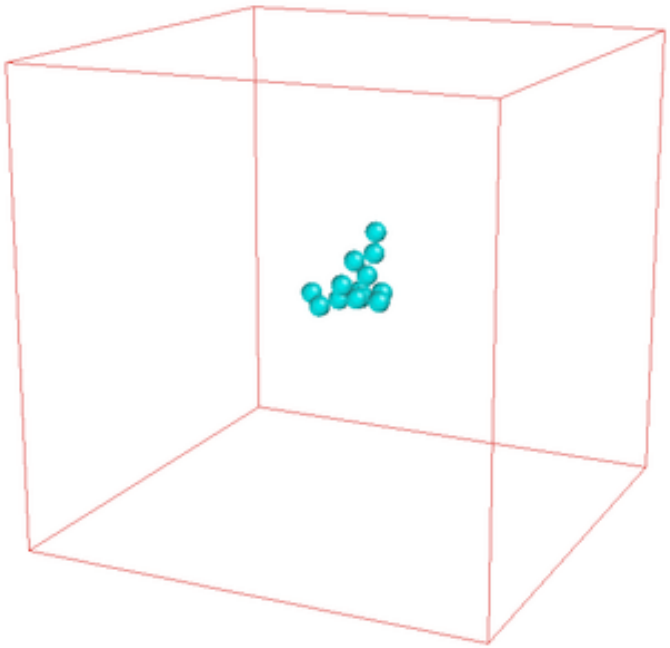
$\Delta\mu$: chem. pot difference

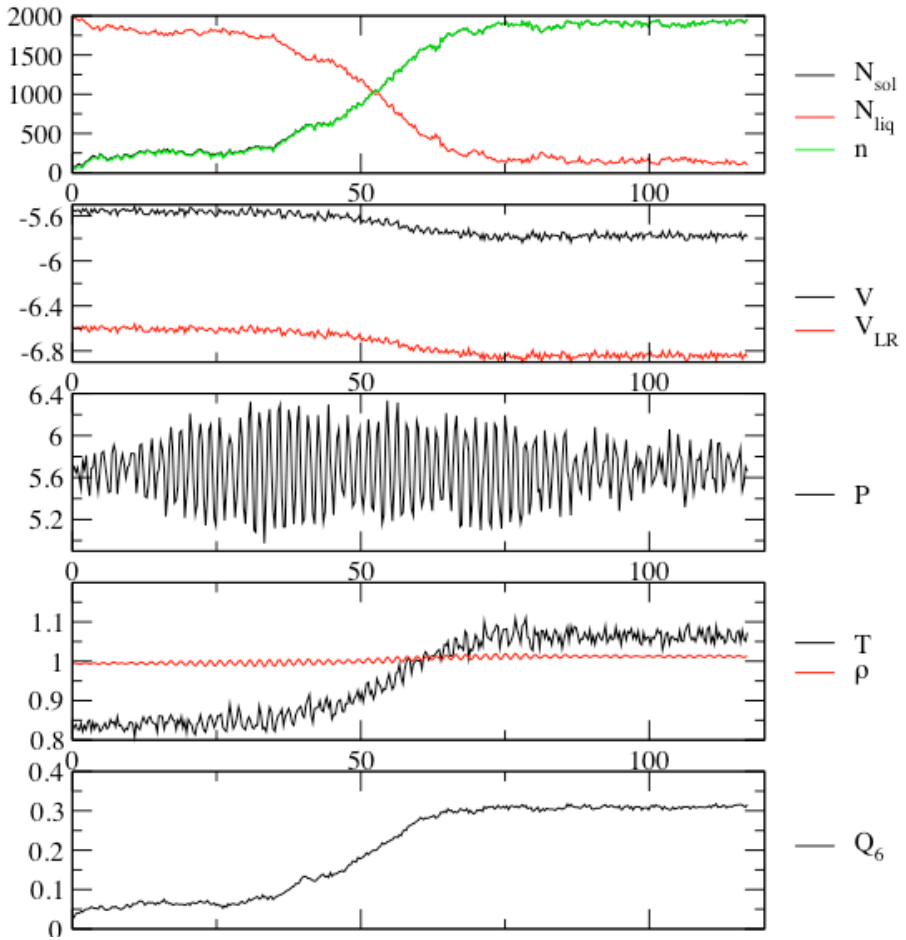
ρ: density

Path sampling of nucleation

TIS in NPH ensemble, as density and temperature change ***(Daniele Moroni)***

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





D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, *Phys. Rev. Lett.* **94**, 235703 (2005)

Sampling paths is only the beginning

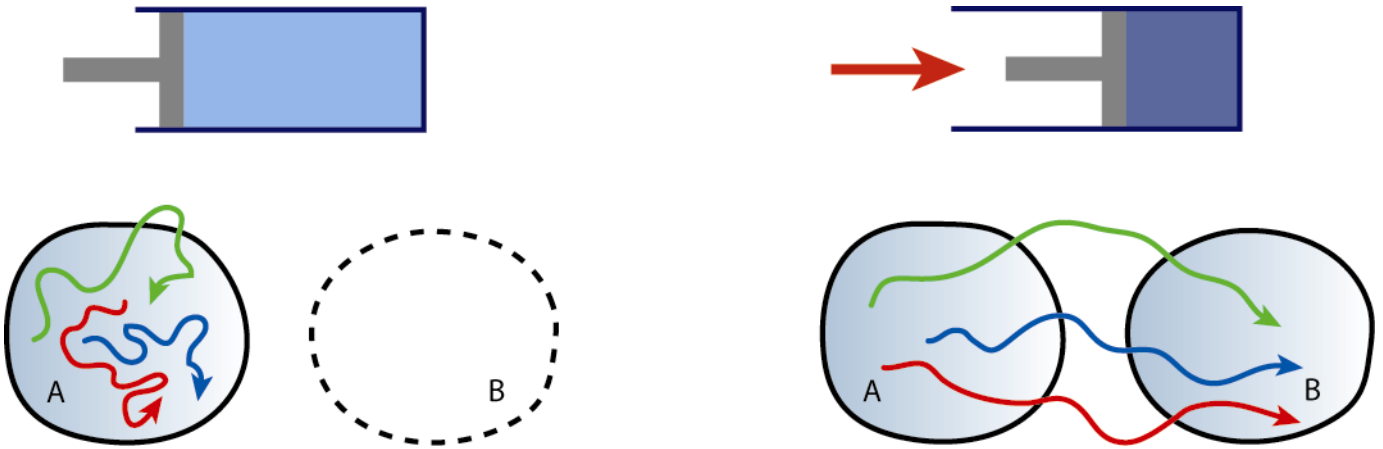
- *Eugene Wigner:* ***"It is nice to know that the computer understands the problem. But I would like to understand it too."***

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

- Illustrative example: crystal nucleation

Free energy of paths

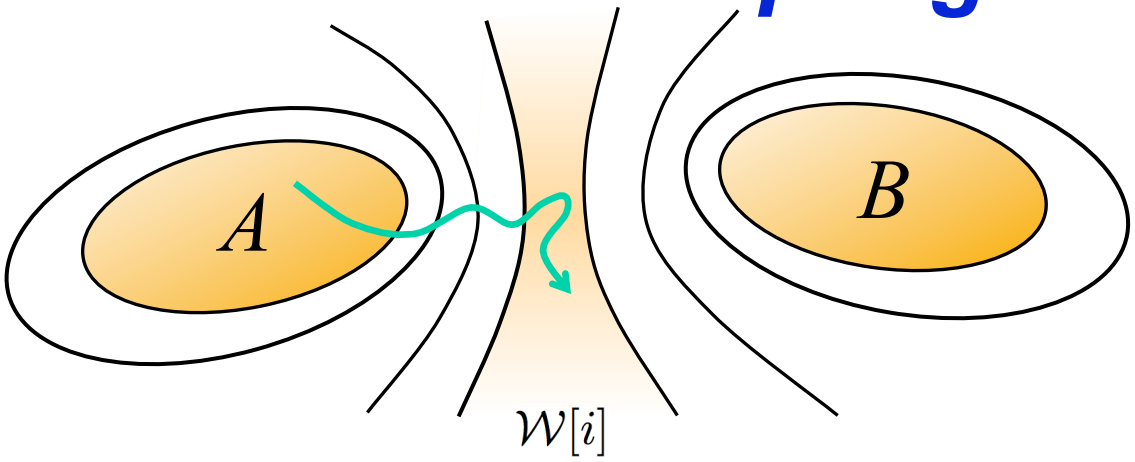
$$C(t) = \frac{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]h_B(x_t)}{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]} = \frac{Z_{AB}(t)}{Z_A}$$



$$W_{AB}(t) \equiv -\ln \frac{Z_{AB}(t)}{Z_A}$$

P.G. Bolhuis, D. Chandler, C. Dellago, and P.L. Geissler, *Annu. Rev. Phys. Chem.* **53,** 291 (2002)
C. Dellago and P. L. Geissler, *AIP Conf. Proc.* **690**, (2003)
P. L. Geissler and C. Dellago, *JPC B* **108**, 6667 (2004)

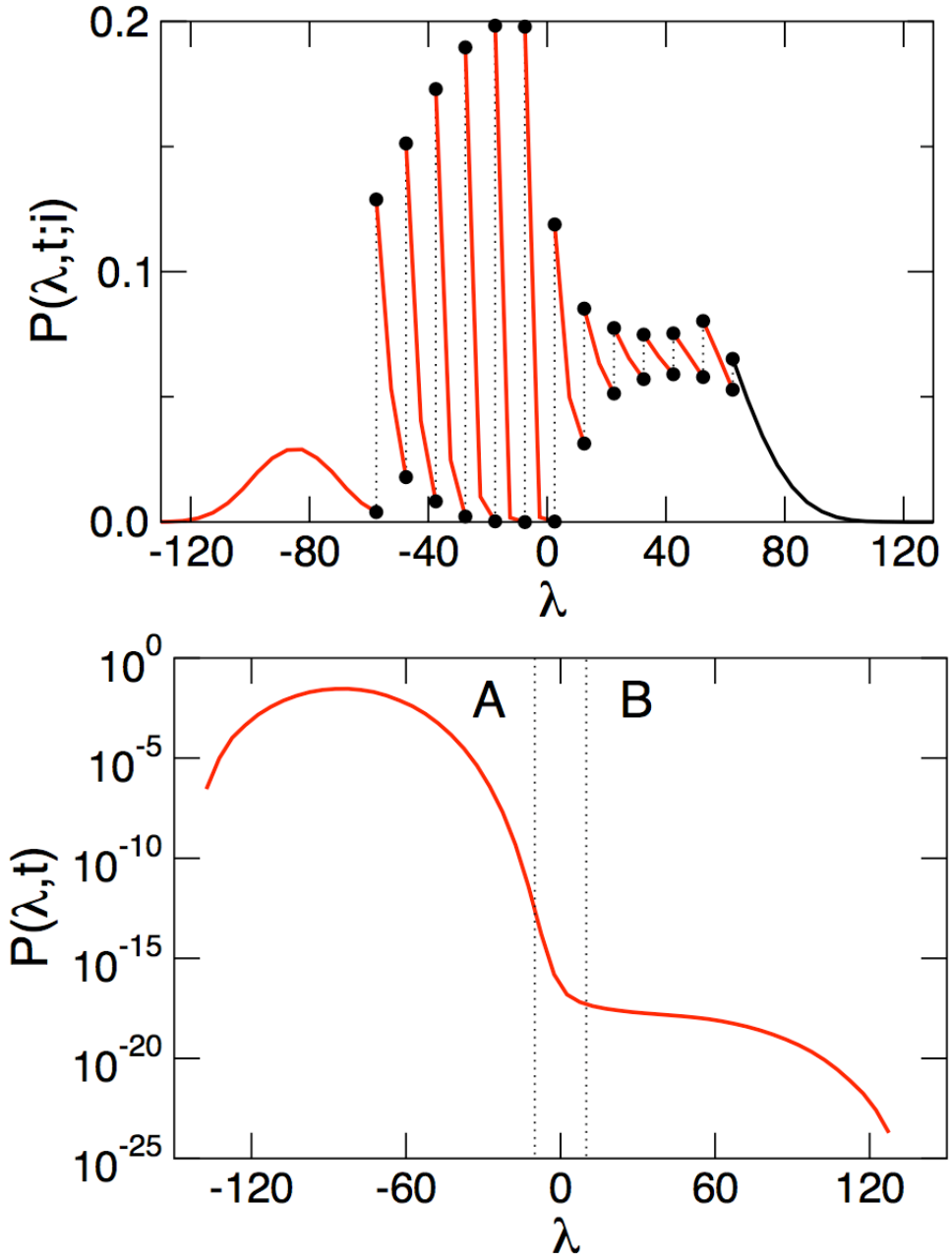
Umbrella sampling



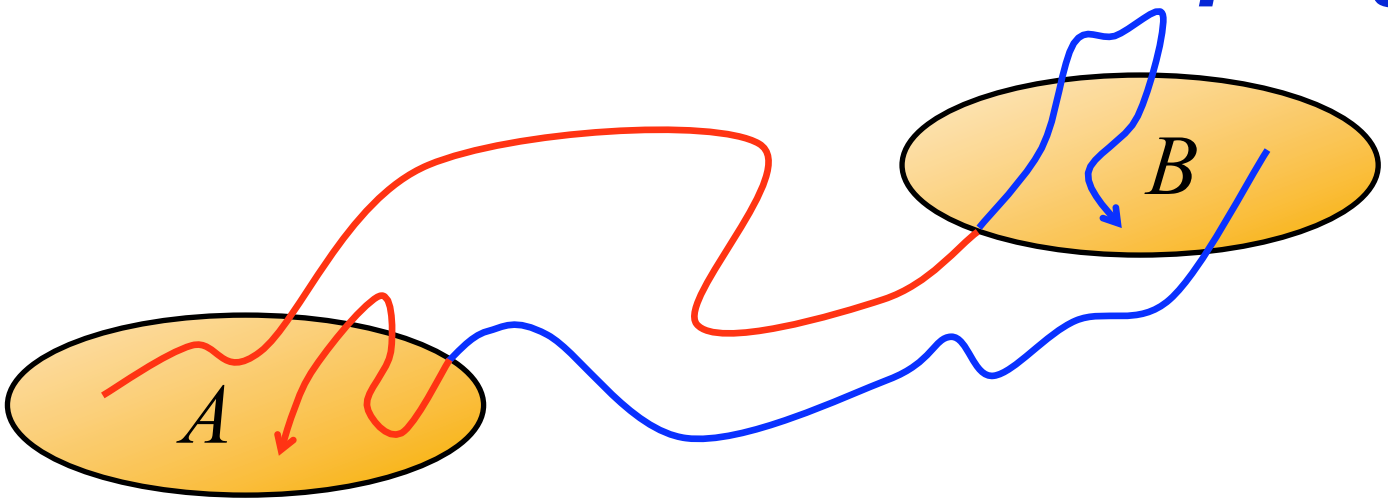
$$P_A(\tilde{\lambda},t) = \frac{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]\delta[\tilde{\lambda}-\lambda(x_t)]}{Z_A} = \langle \delta[\tilde{\lambda}-\lambda(x_t)]\rangle_A$$

$$C(t) = \exp[-W_{AB}(t)] = \int_{\lambda^B_{\min}}^{\lambda^B_{\max}} d\lambda\, P_A(\lambda,t)$$

$$P_{A\mathcal{W}[i]}(\tilde{\lambda},t) = \frac{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]h_{\mathcal{W}[i]}(x_t)\delta[\tilde{\lambda}-\lambda(x_t)]}{\int \mathcal{D}x(t)\, h_A(x_0)h_{\mathcal{W}[i]}(x_t)}$$
$$= \langle \delta[\tilde{\lambda}-\lambda(x_t)]\rangle_{A\mathcal{W}[i]}.$$


0.2
0.1
0.0
P(λ,t;i)
-120
-80
-40
0
40
80
120
λ
10^0
10^-5
10^-10
10^-15
10^-20
10^-25
P(λ,t)
A
B
-120
-60
0
60
120
λ

Transition interface sampling



Overall states in phase space:

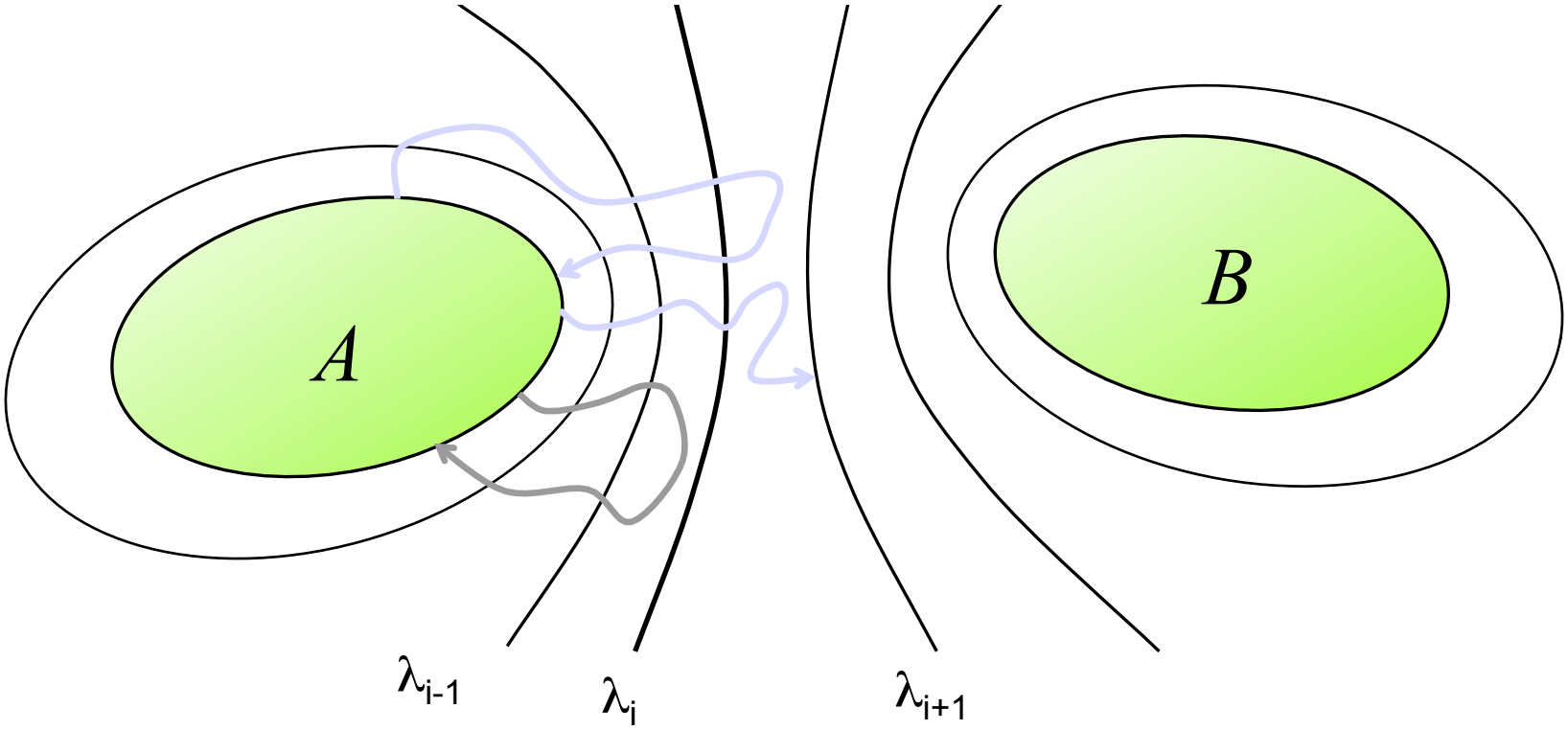
$\mathcal{A}$ going *back in time* **A** reached first

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

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

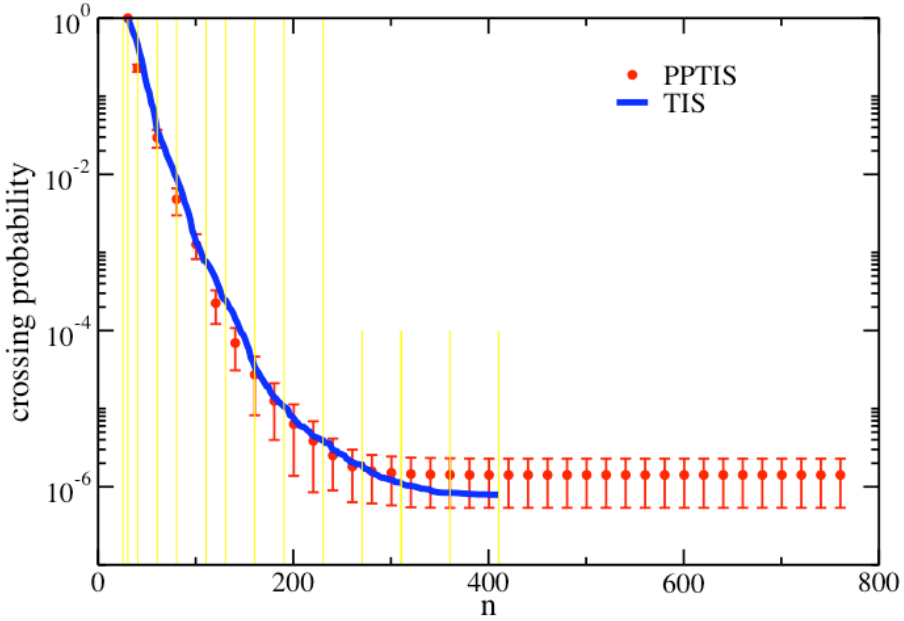
$$k_{AB} = \frac{\langle h_{\mathcal{A}}(x_0) \dot{h}_{\mathcal{B}}(x_0) \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle}$$

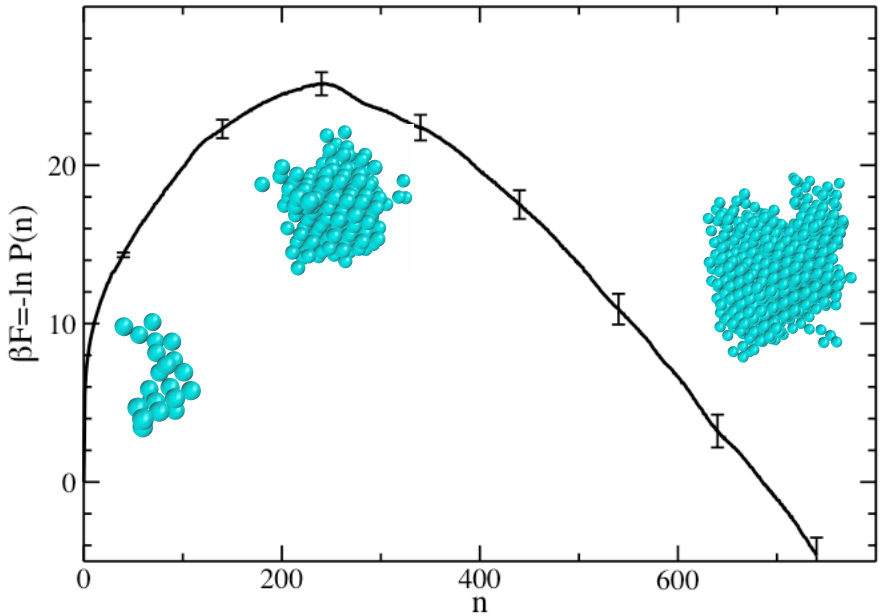
T. S. van Erp, D. Moroni and P. G. Bolhuis, *J. Chem. Phys*. **118** , 7762 (2003)
T. S. van Erp and P. G. Bolhuis, *J. Comp. Phys.* **205**, 157 (2005)



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

$$k_{AB}^{TIS} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^{n-1} P_A(\lambda_{i+1} \mid \lambda_i) = \Phi_A \prod_{i=1}^{n-1} P_A(\lambda_{i+1} \mid \lambda_i)$$





TIS results for nucleation

$$\mathcal{P}_A(B|1) = \prod_{i=1}^{n-1} \mathcal{P}_A(i+1|i) = 8 \times 10^{-7}$$

$$\frac{\langle \phi_{A1} \rangle}{\langle h_A \rangle} = 1.29$$

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

Memory loss reasonable

Free energy follows directly

Moroni, van Erp, Bolhuis, PRE, 2005

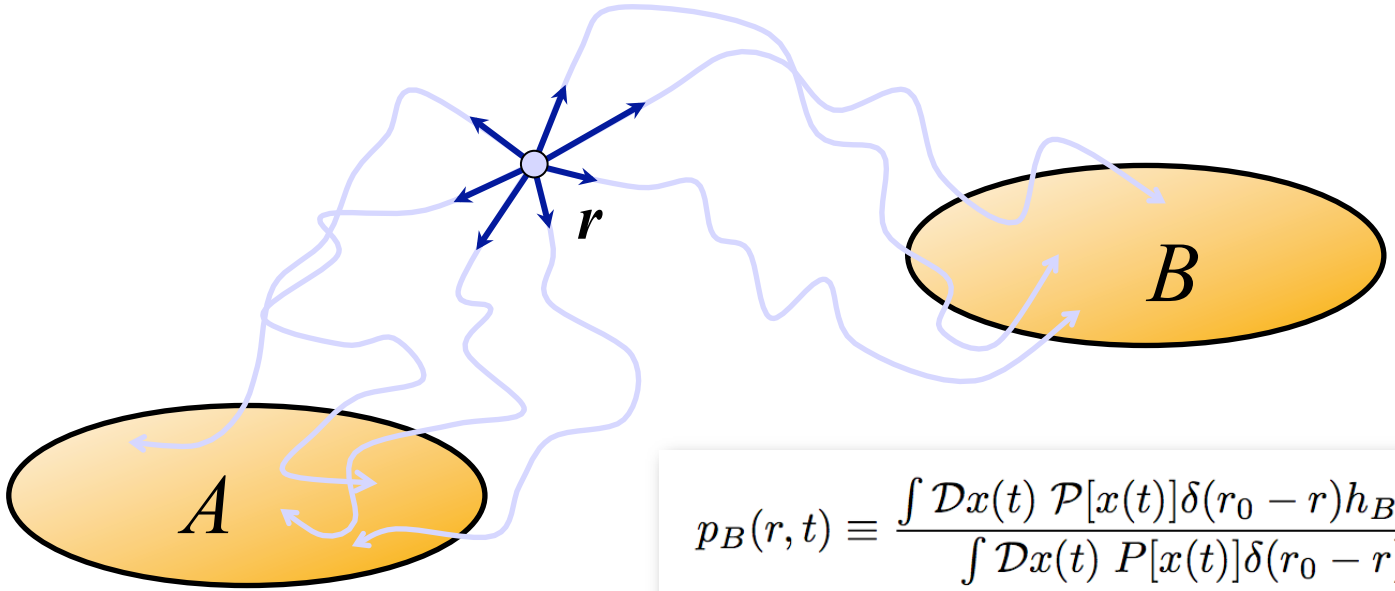
Agrees with ten Wolde (1995)

Structural analysis?

Committor

(aka p-fold, splitting probability)

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



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

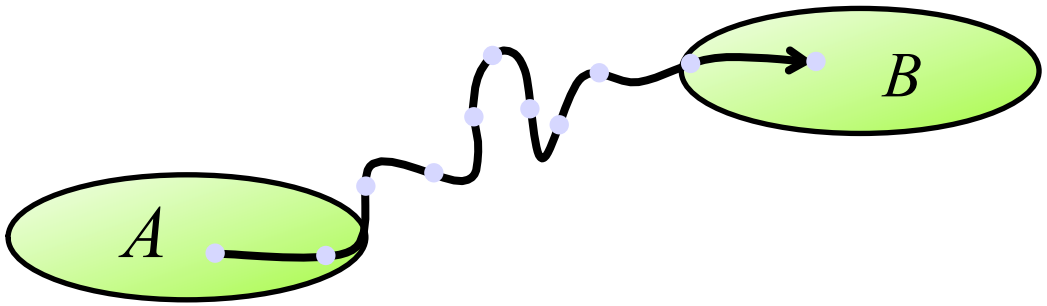
$$p_B(r,t) \approx \frac{1}{N}\sum_{i=1}^{N} h_B(x_t^{(i)})$$

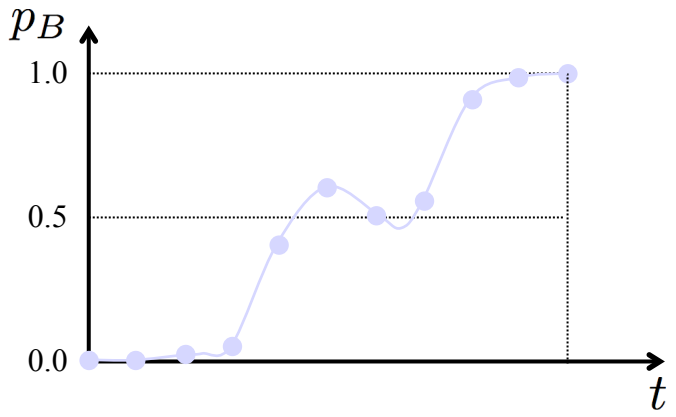
$$\sigma = \sqrt{\langle (p_B^{(N)} - p_B)^2 \rangle} = \sqrt{\frac{p_B(1-p_B)}{N}}$$

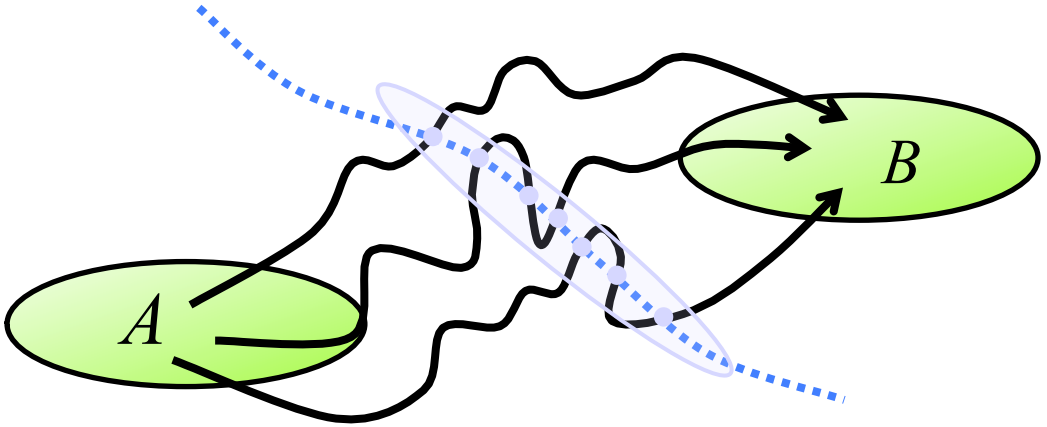
L. Onsager, *Phys. Rev.* **54**, 554 (1938).
M. M. Klosek, B. J. Matkowsky, Z. Schuss, *Ber. Bunsenges. Phys. Chem.* **95**, 331 (1991)
V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, *J. Chem. Phys.* **108**, 334 (1998)

Transition state ensemble

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



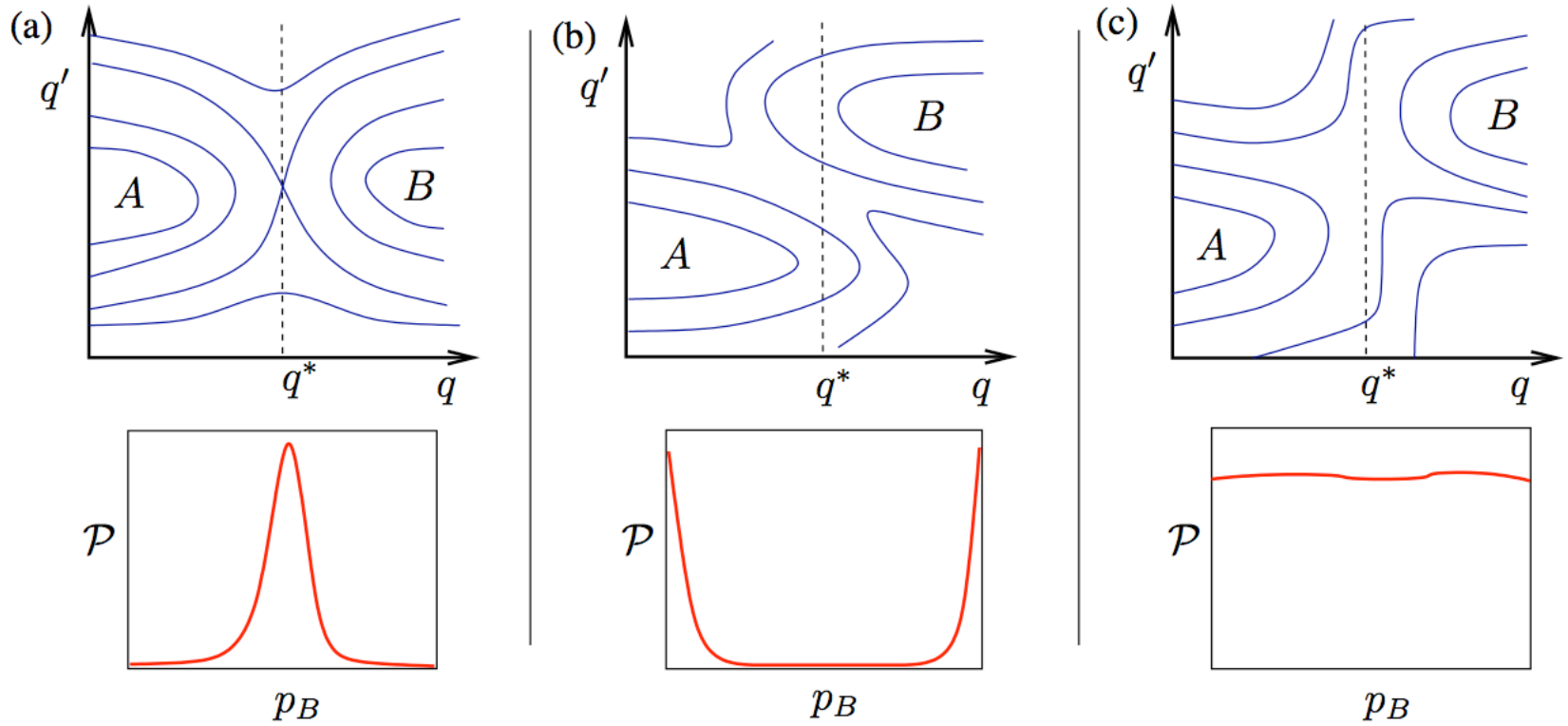




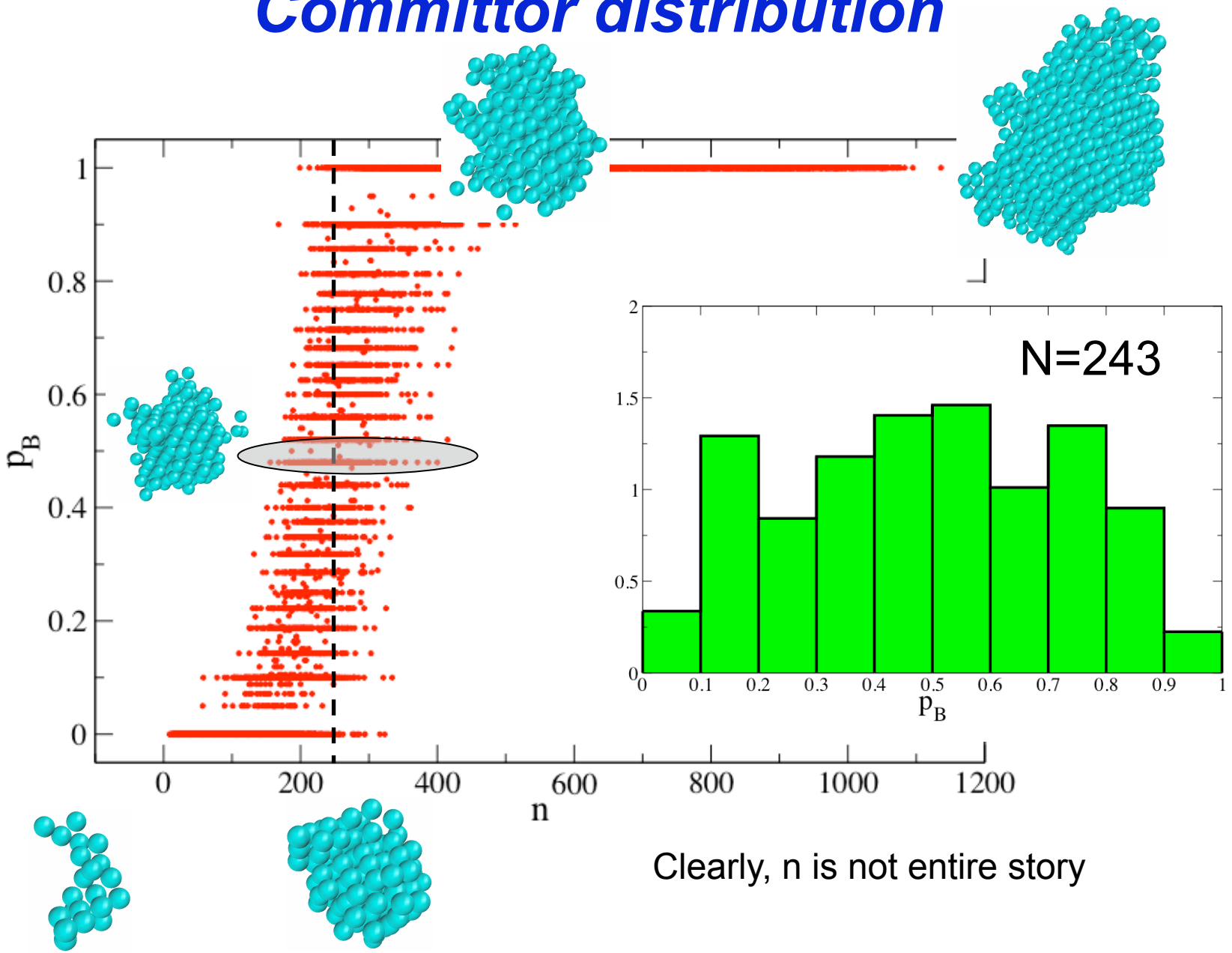
TSE:

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

Committor distributions



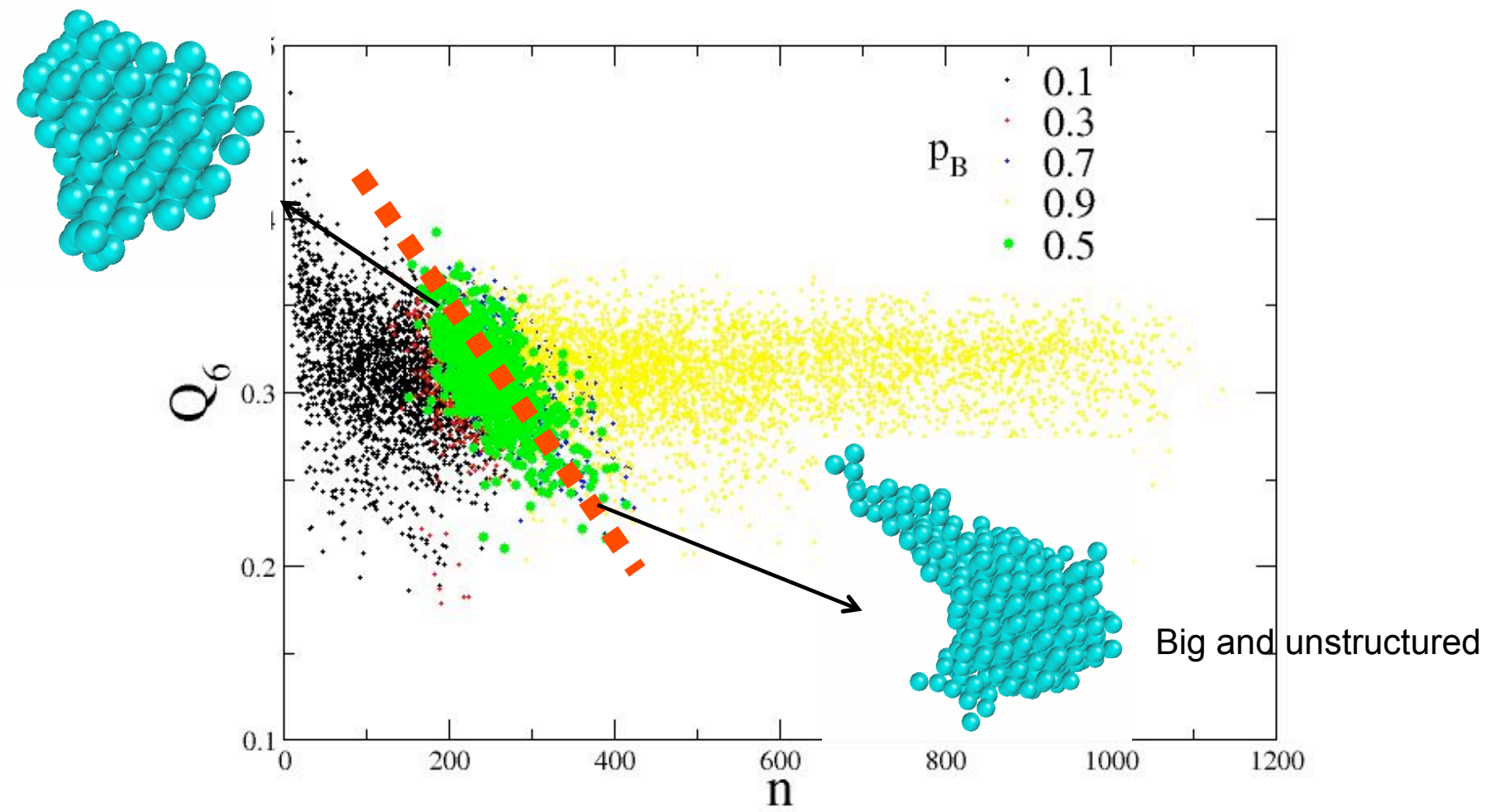
Committor distribution



Clearly, n is not entire story

Structure

Small and structured



Big and unstructured

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

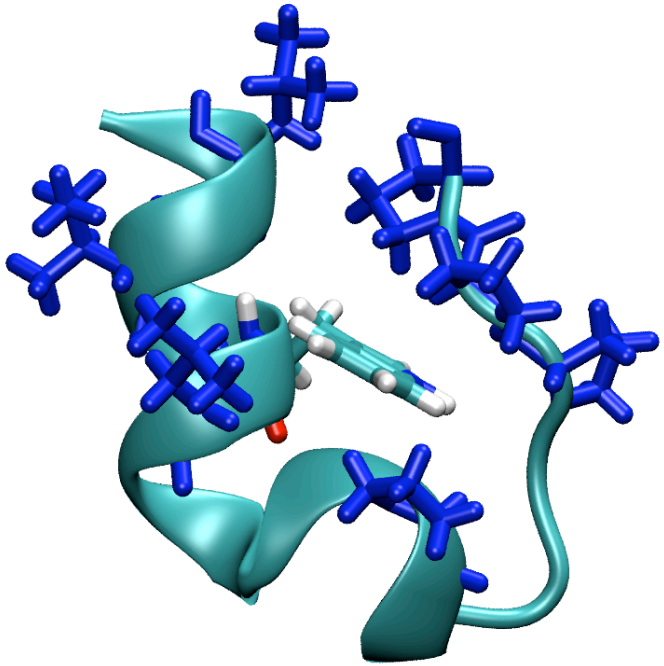
- **Part 3**
  - **Application to protein folding**

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*)

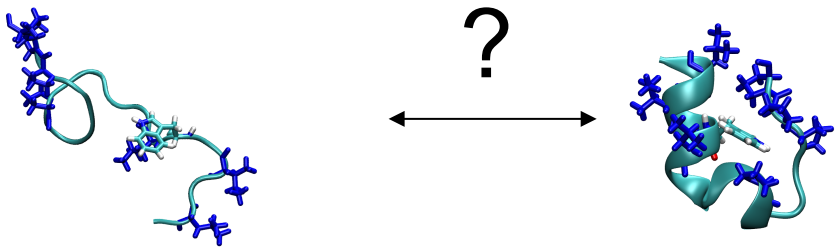


System:

1L2Y in 2800 SPC waters

OPLSAA, PME, Nose-Hoover, GROMACS

What is folding mechanism in explicit water at 300K?



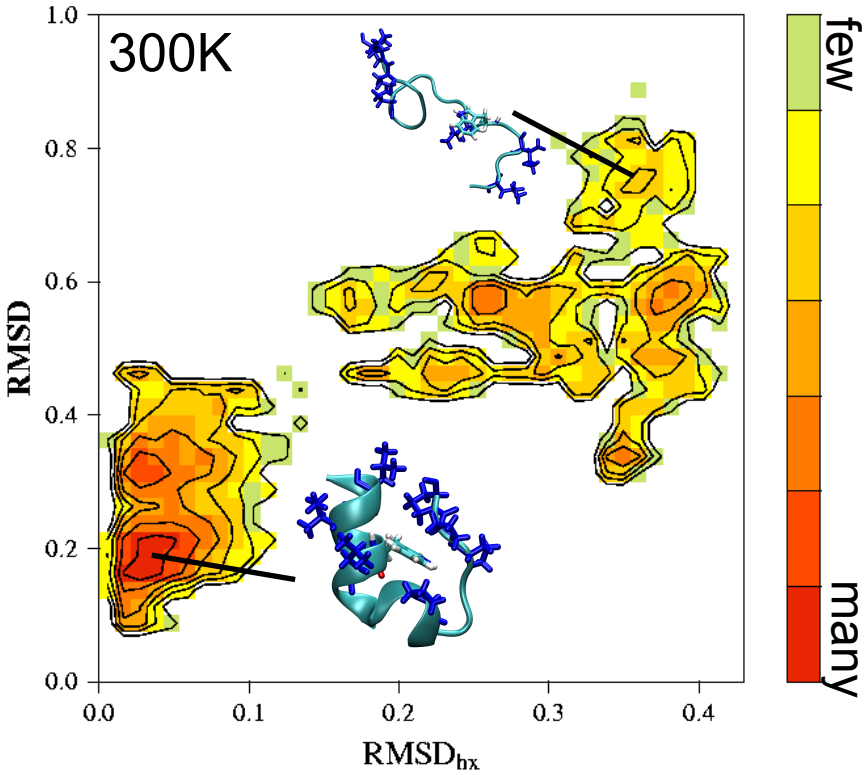
Strategy:

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

Jarek Juraszek and P.B., PNAS 2006

Replica exchange of Trp-cage

initially folded

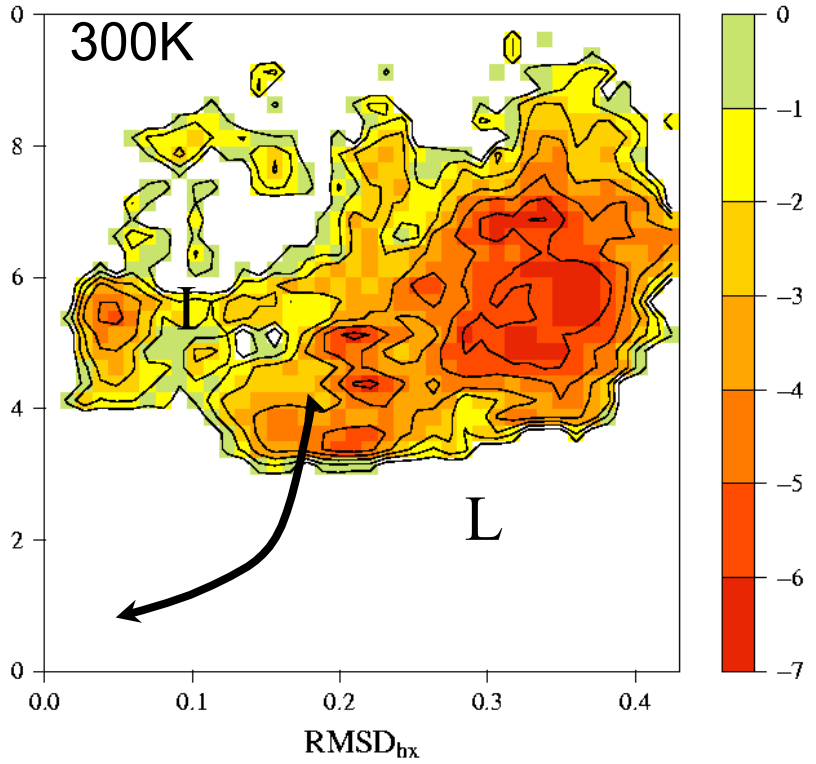


64 replicas

Temp range 272-555K

30 ns per replica

initially unfolded

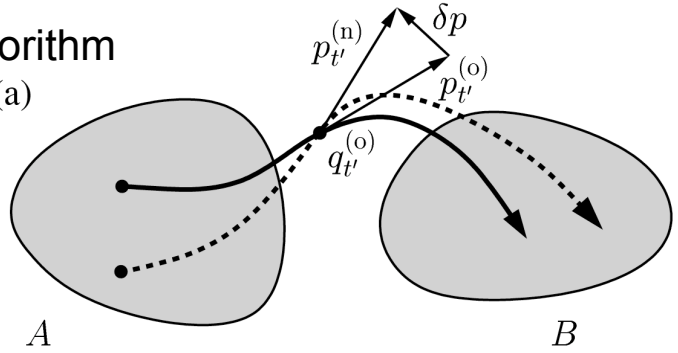


barrier towards native state not crossed:

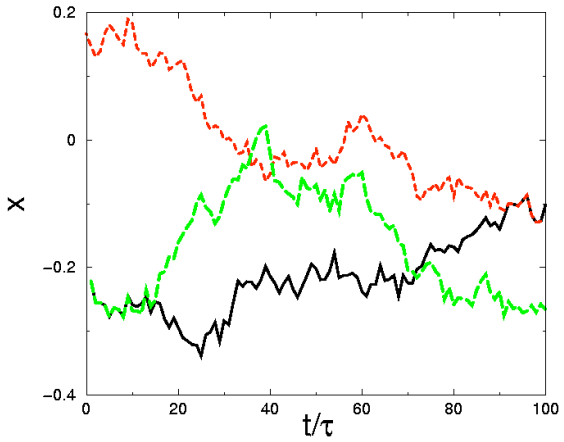
Path sampling

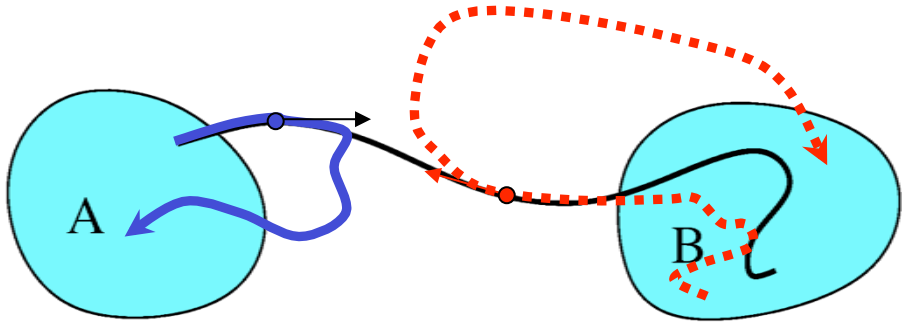
Acceptance problem

Original shooting algorithm



has problems with long paths (>1ps) because of molecular chaos (even with dp→0)

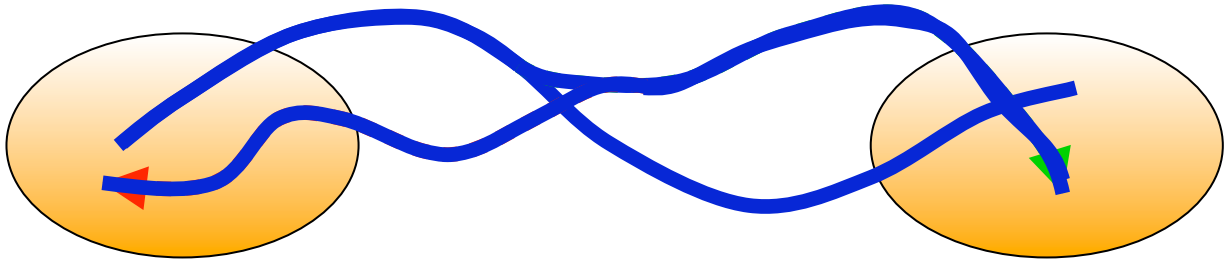




Result: very low acceptance

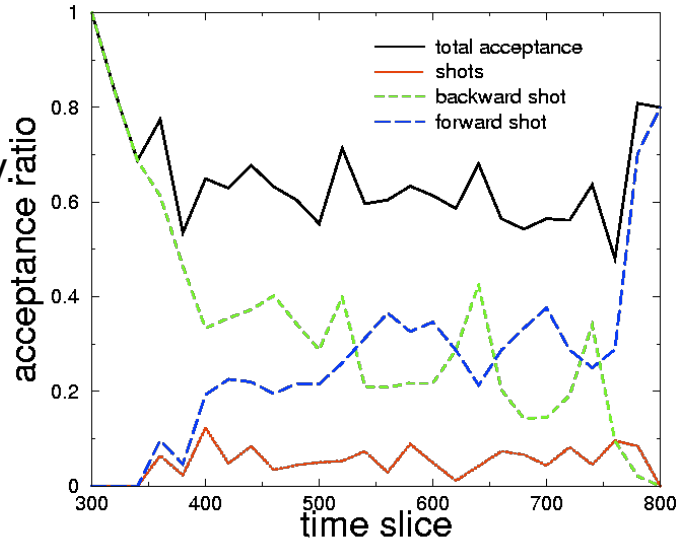
Modified path sampling

Long paths require stochastic sampling.



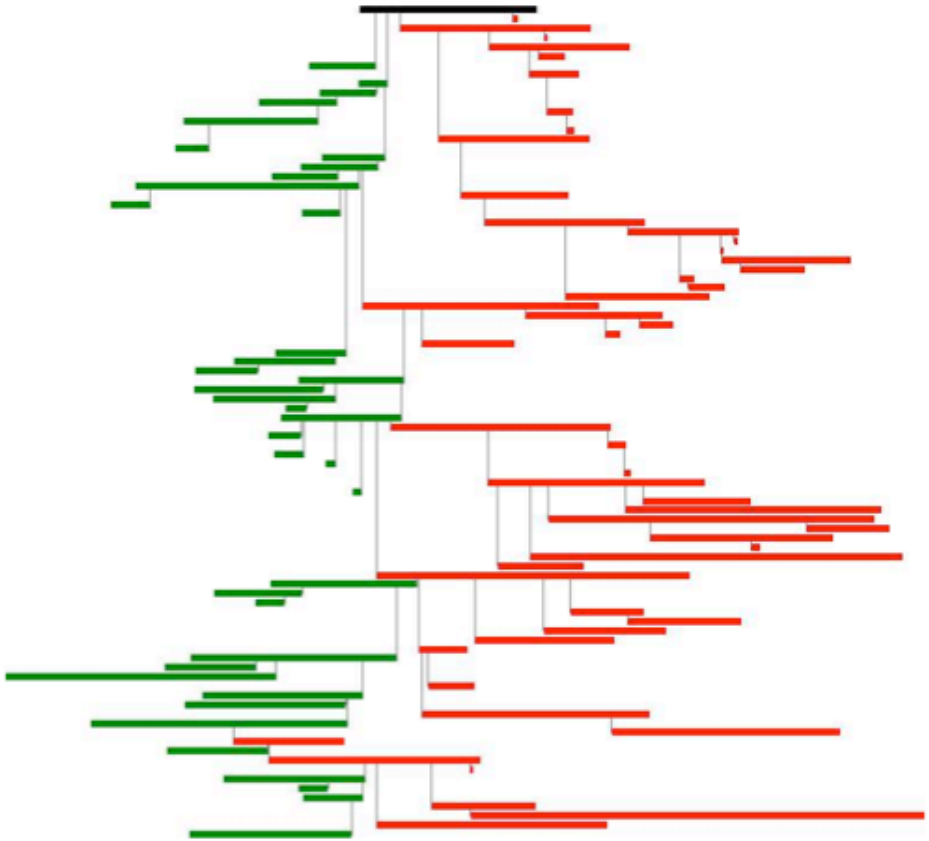
Shoot a trajectory using stochastic thermostat in just one arbitrary direction.

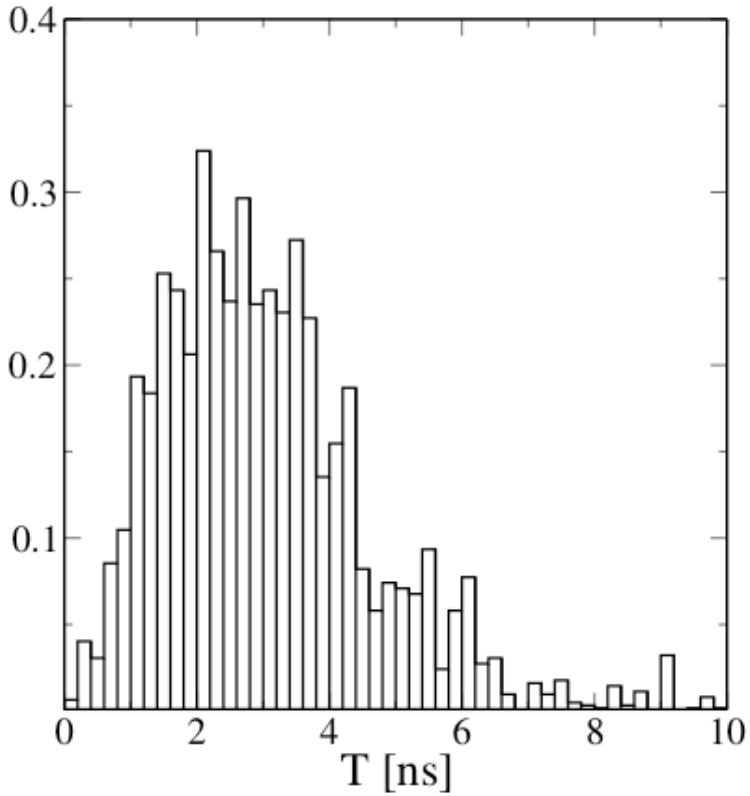
Path length fluctuates, no shifting moves necessary.



- Higher acceptance (>50% instead of <10%)
- Allows shorter trajectories
- No true deterministic dynamics

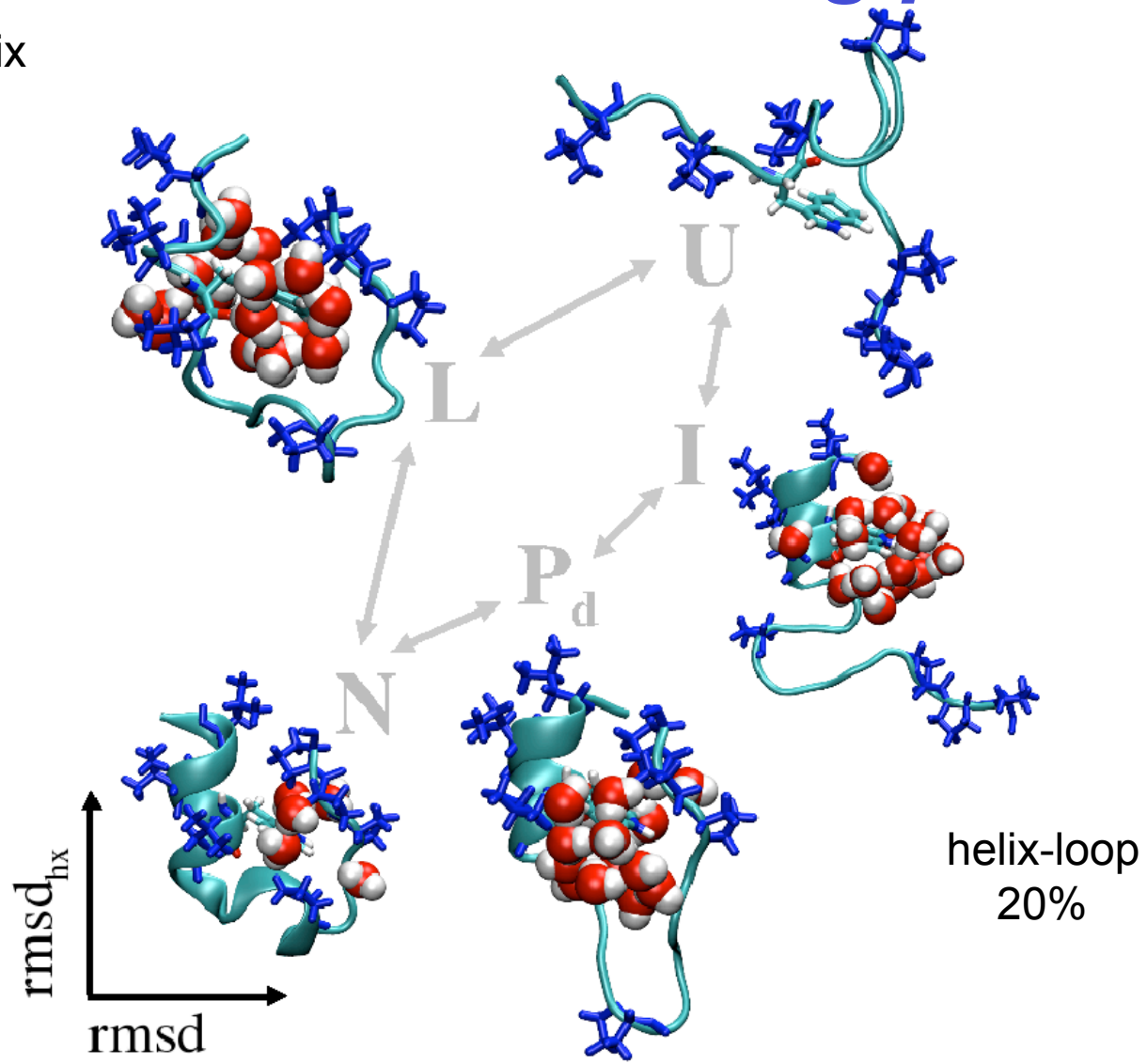
Path sampling indicators





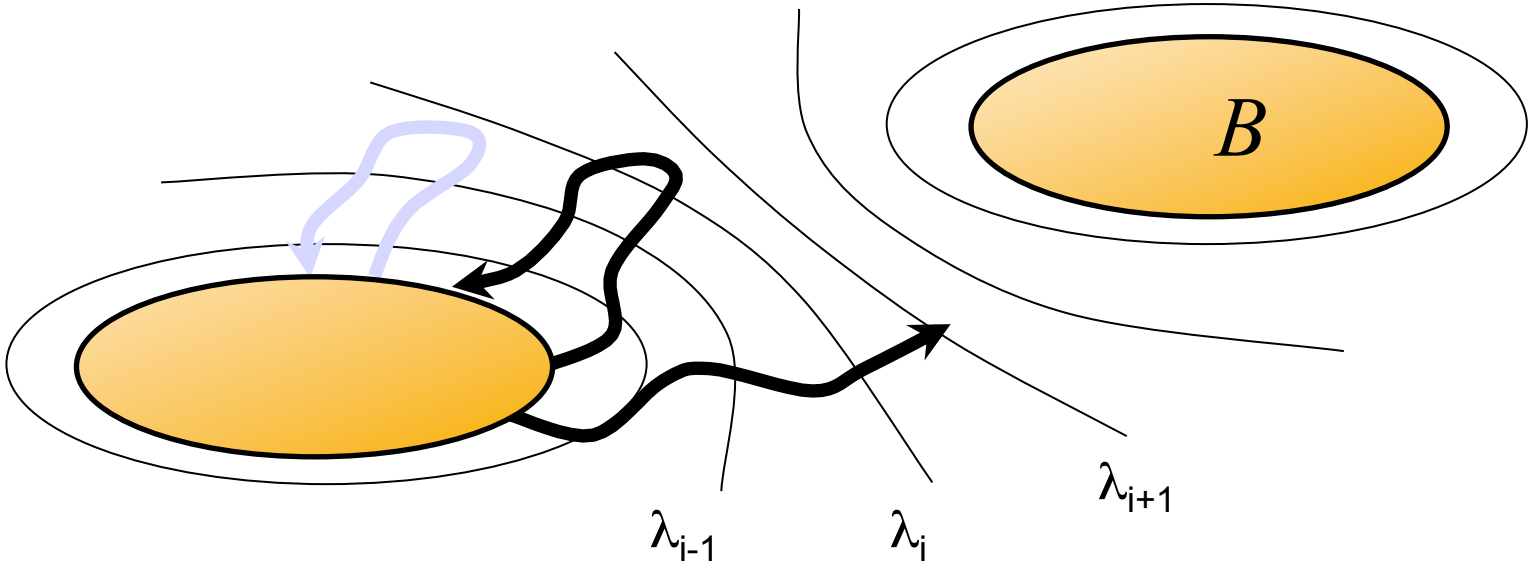
TPS reveals two folding pathways

loop-helix
80%



helix-loop
20%

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

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

N-L folding and unfolding rate

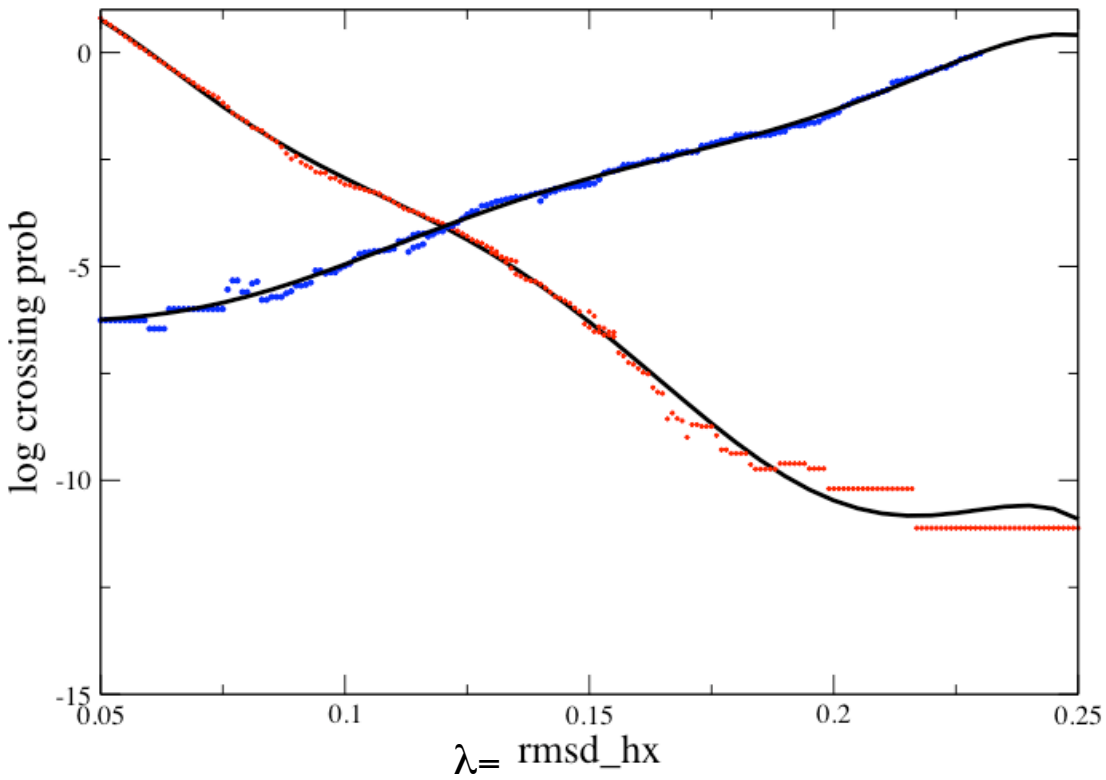
TIS, 6 interfaces

ln P_{unf} = -11
Φ_A (λ =0.06)= 6.6 ns^{-1}
k_{unf} = 0.11 μs^{-1}

Exp: k_{unf} = 0.08 μs^{-1}

ln P_{fol} = -6.3
Φ_A (λ=0.23) = 1 ns^{-1}
k_{fol} = 1.83 μs^{-1}

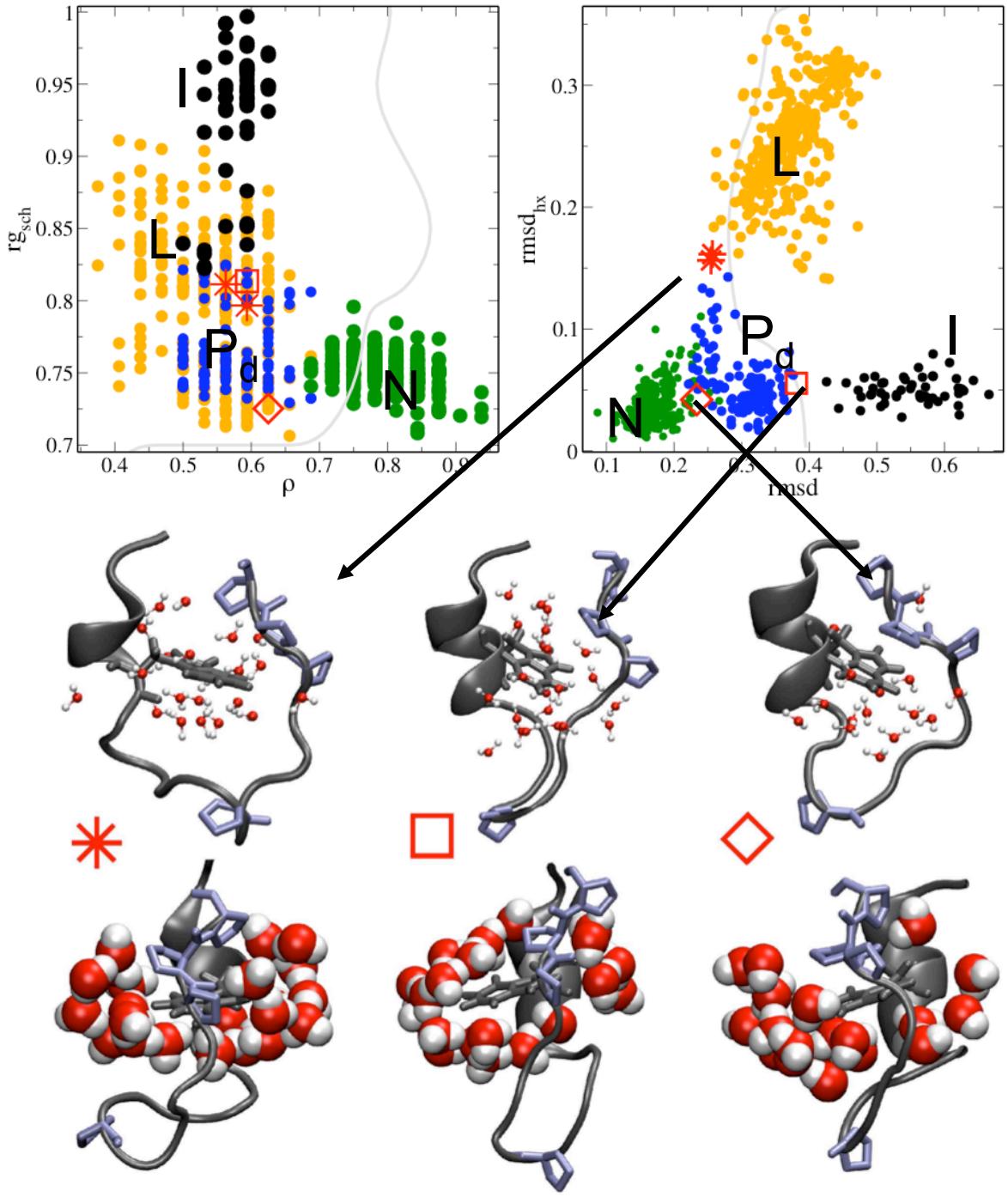
Exp. k_{fol} = 0.24 μs^{-1}



Difference in folding rate might be caused by L state being different from U state.

TSE

TS can fall inside stable state in FE landscape

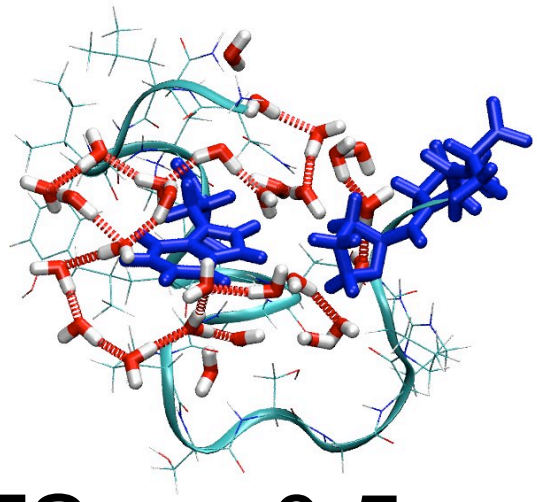


Water role in kinetics

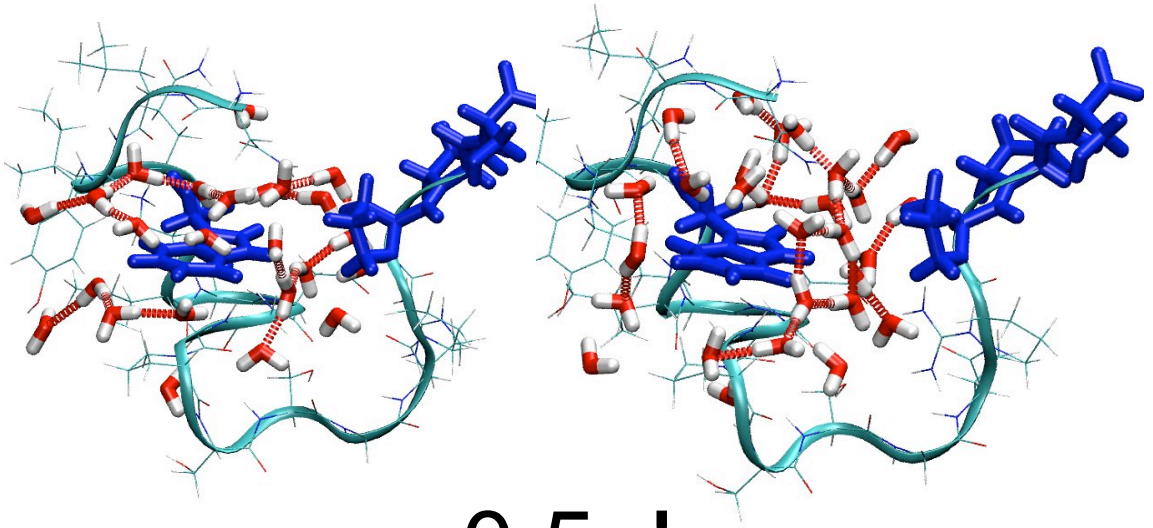
We can do a committor analysis:

- Start with TS and randomize waters, calculate committor

I-N path



TS:p_B=0.5



p_B=0.5 !

Role of water

- Water expulsion is last step upon folding.

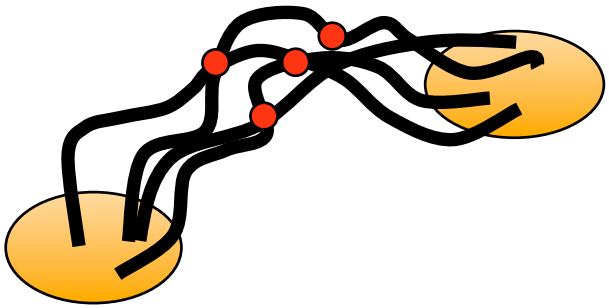
- P_{fold} calculations show instantaneous configuration of water is not important in reaction coordinate (for both L-N and I-N pathways).
  - Average path length is 3 ns: more than enough time to equilibrate

- Water plays a role, as implicit solvent predict different pathways and TS structure. Possible solution: some water molecules can play structural role:
  - Average residence time < 50 ps
  - Special very strongly bound bridging water molecules > 1 ns.

- In improved implicit solvent models these waters should be treated differently.

Maximum likelihood estimation

- Suppose we have TPS information, how do we get reaction coordinate *r* ?



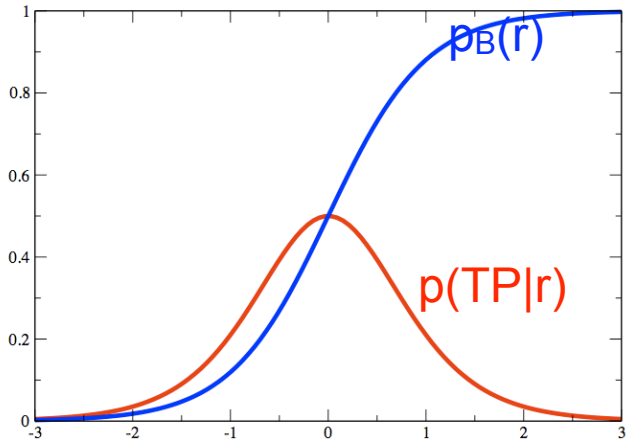
Peters & Trout, JCP 125 054108(2006)

- p(TP|x) is the probability to be on a Transition Path provided we are at x
- A model for this function could be

$$p(TP|r) \propto (1 - \tanh^2(r))$$

- For diffusive dynamics (or dynamics that quickly loses memory) this changes into a function of the committor pB

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



- The question: what is r(x) ?

Maximum likelihood estimation

- we are searching for a reaction coordinate r(x) that describes the reaction best.
- r(x) should be a combination of the collective variable q(x), e.g.

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

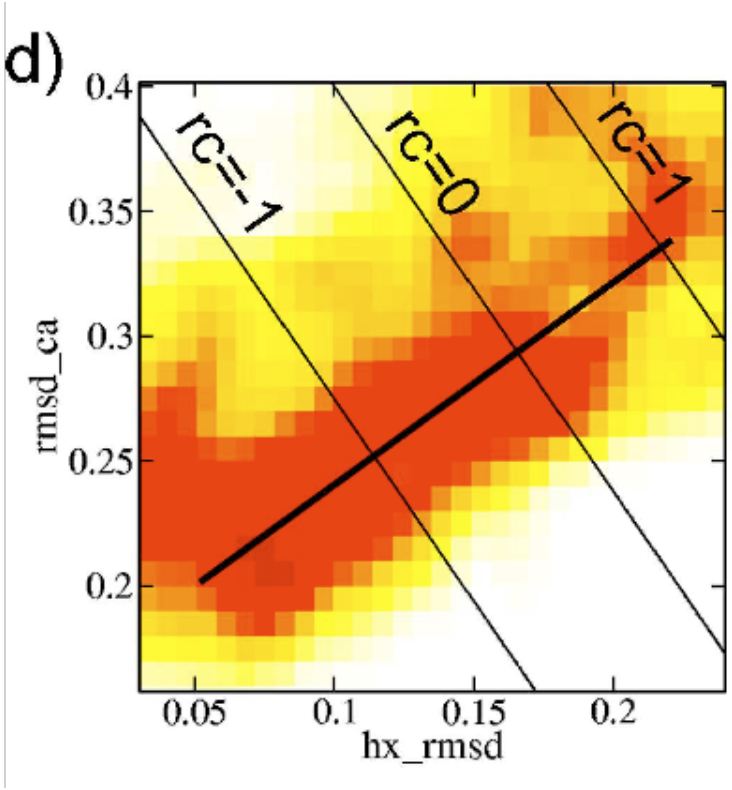
- suppose we have from aimless TPS shooting points x and the outcomes of the trajectories: $\{x^{A}\}$ and $\{x^{B}\}$

- Then the likelihood of a model predicting this outcome would be

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

- maximize this likelihood to obtain the optimal parameters α

Reaction coordinate analysis



TPS

Conclusions

- Separate folding pathways can be sampled with TPS

- Transition state ensemble (TSE) :
  - characterized by solvation
  - water dynamics probably no part of RC at TSE , water structure is.
  - does not always correspond with a FE landscape saddle
  - In general FE should not be trusted for kinetics (solution: flux through graphs, MSMs)

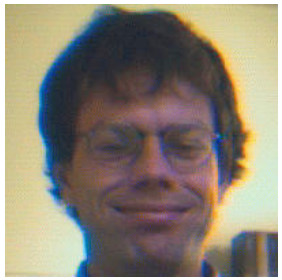
- Calculated TIS rate constant compares good to experiment for unfolding and less for folding: possibly a force field issue.

Bridging time and length scales in biophysical problems requires a combination of techniques

Summary

- Rare events difficult to simulate
- need for good potentials
- need for good MD simulation
- need for free energy calculations
- need for order parameter free methods:
  - NEB, TPS , string method
- need for reaction coordinate analysis methods

Acknowledgements



Evert Jan Meijer

Elske Leenders

Jocelyne Vreede

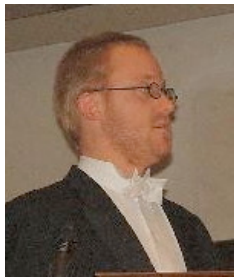
Jarek Juraszek

Christoph Dellago
Univ Vienna

David Chandler
UC Berkeley

Phill Geissler
UC Berkeley

Daniele Moroni
Cambridge



Titus van Erp
Univ Leuven