Advanced path sampling of rare events in biomolecular systems

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

Outline

- Simulation of biomolecular systems

- Basic TPS
  - shooting algorithms
  - stable state definitions
  - example Photoactive Yellow Protein
  - reaction coordinate analysis

- Advanced path sampling
  - rates by Transition Interface Sampling (TIS)
  - replica exchange TIS and multiple state TIS
  - single replica multiple state TIS
  - example Trp-cage folding network

- Conclusion

Protein self-assembly

understanding the cellular processes

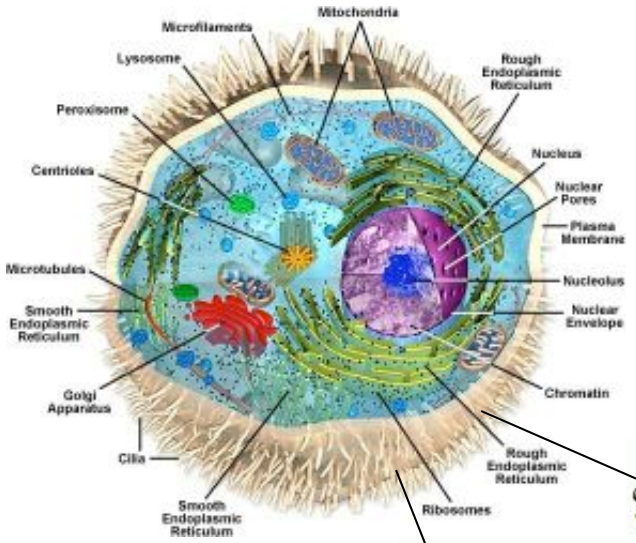
- folding
- structure formation (cytoskeleton)
- complex formation (regulation)
- neurodegenerative/genetic diseases
- …..

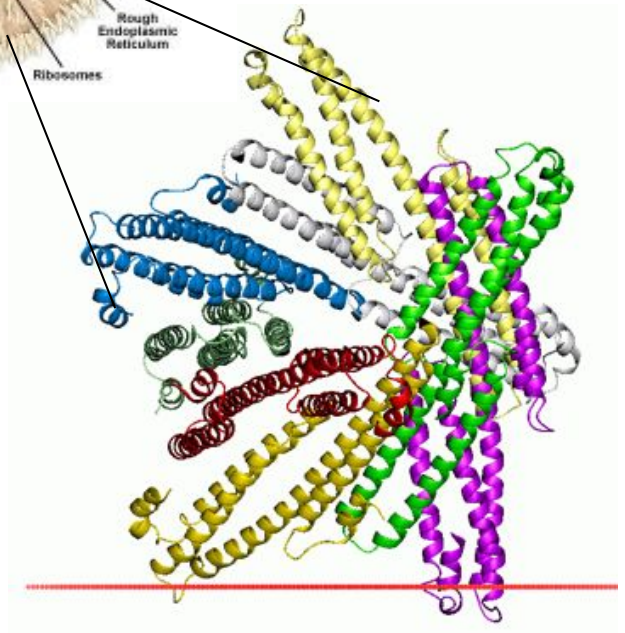
novel self assembling biomaterials

- artificial tissue
- smart packaging
- self healing coatings
- sensors

Challenge

understanding and predicting protein assembly with advanced molecular simulation



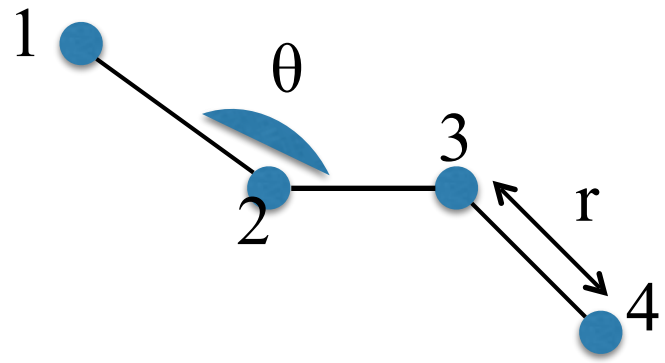


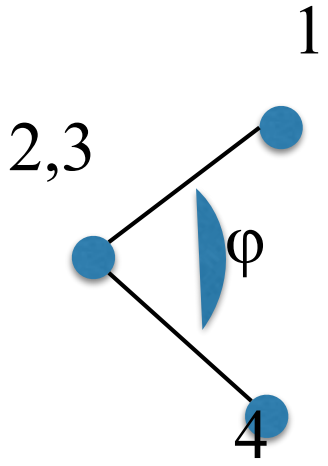
All-atom force fields for biomolecules

- Potential energy for protein

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

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





vdW interactions only between non-bonded |i-j|>4

Currently available empirical force fields

- CHARMm (MacKerrel et 96)
- AMBER (Cornell et al. 95)
- GROMOS (Berendsen et al 87)
- OPLS-AA (Jorgensen et al 95)
- ENCAD (Levitt et al 83)
- …..

- Subtle differences in improper torsions, scale factors 1-4 bonds, united atom rep.
- Partial charges based on empirical fits to small molecular systems
- Amber & Charmm also include ab-initio calculations
- Not clear which FF is best : top 4 mostly used

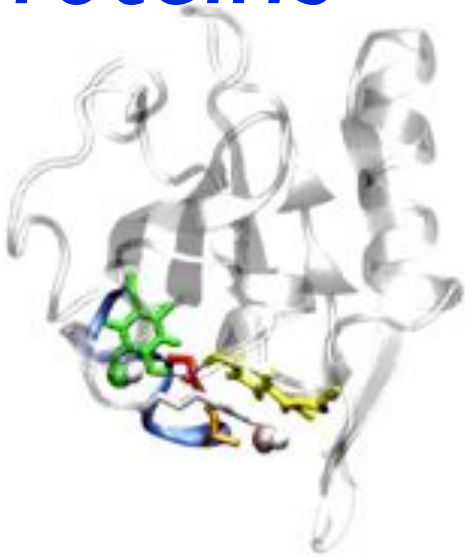
- Water models also included in description
  - TIP3P, TIP4P
  - SPC/E
- Current limit: 10^6 atoms, microseconds (with Anton ms)

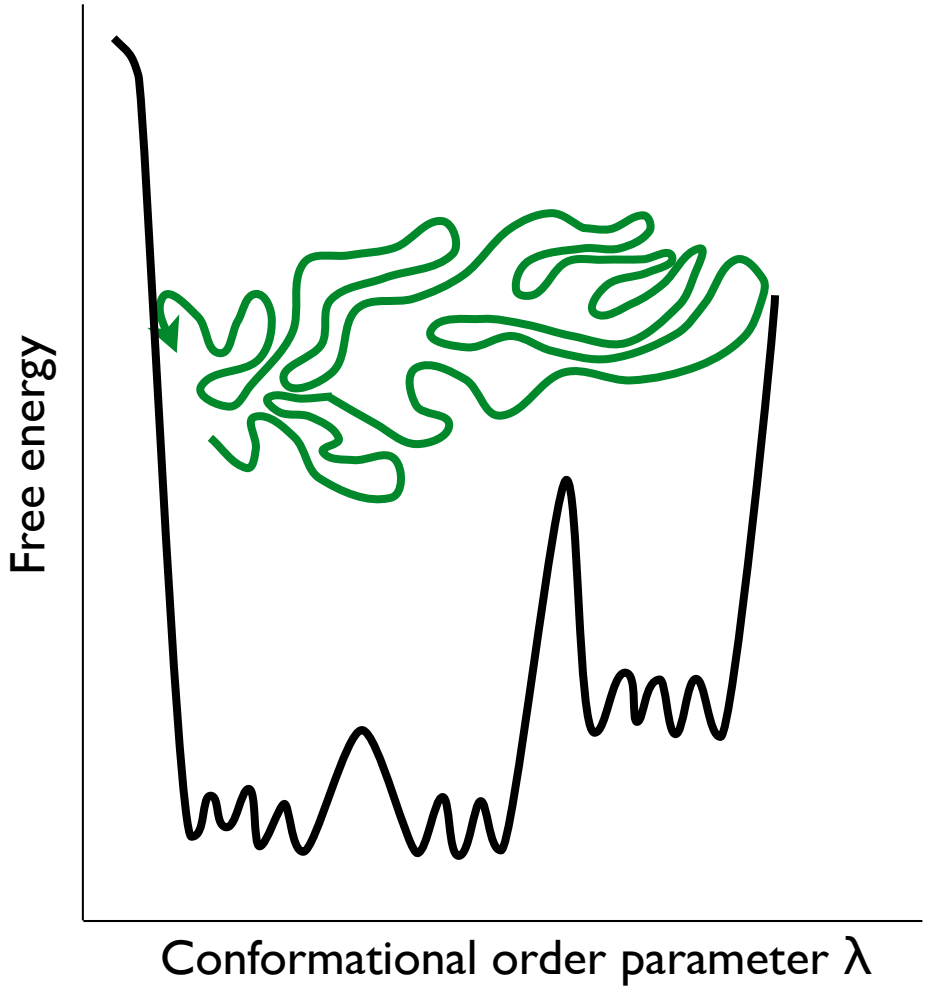
Molecular Dynamics of proteins

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

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

$$\mathbf{F} = -\frac{\partial \mathbf{V}}{\partial \mathbf{r}}$$





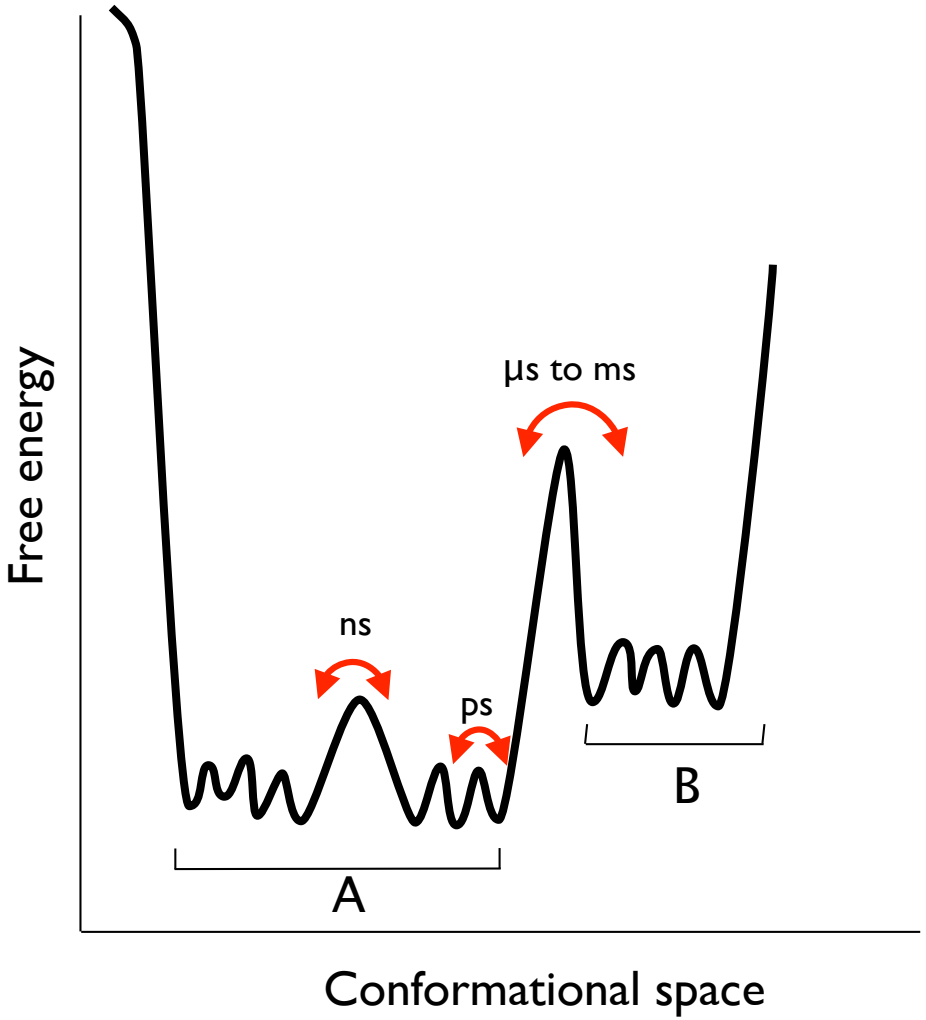
MD yields

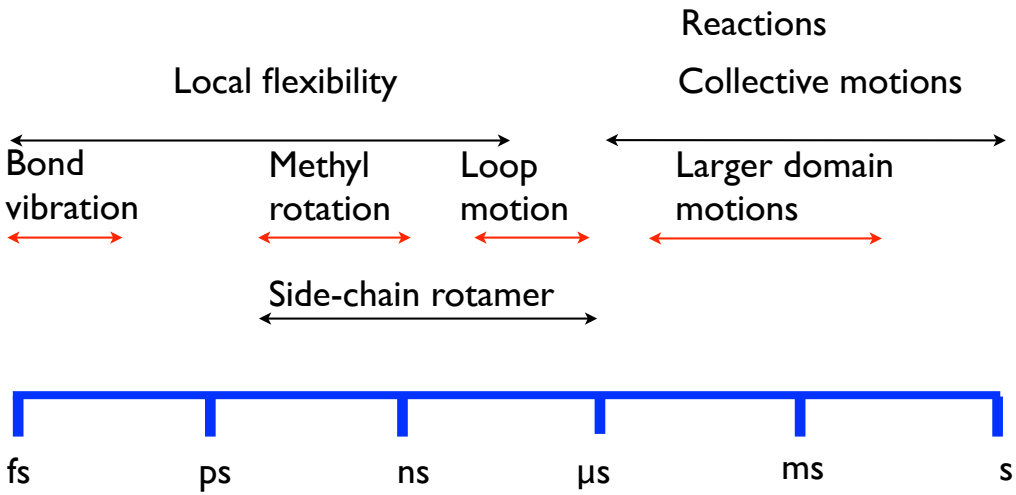
- **equilibrium statistics:** free energy landscapes, stable structures, transition states, …
- **kinetics:** rates, mechanisms, transport properties, …

FE landscape: high dimensional trajectory can be projected onto collective variable λ

$$e^{-\beta F(\lambda)} = \langle \delta(\lambda(\mathbf{r}) - \lambda) \rangle$$

Timescales in proteins

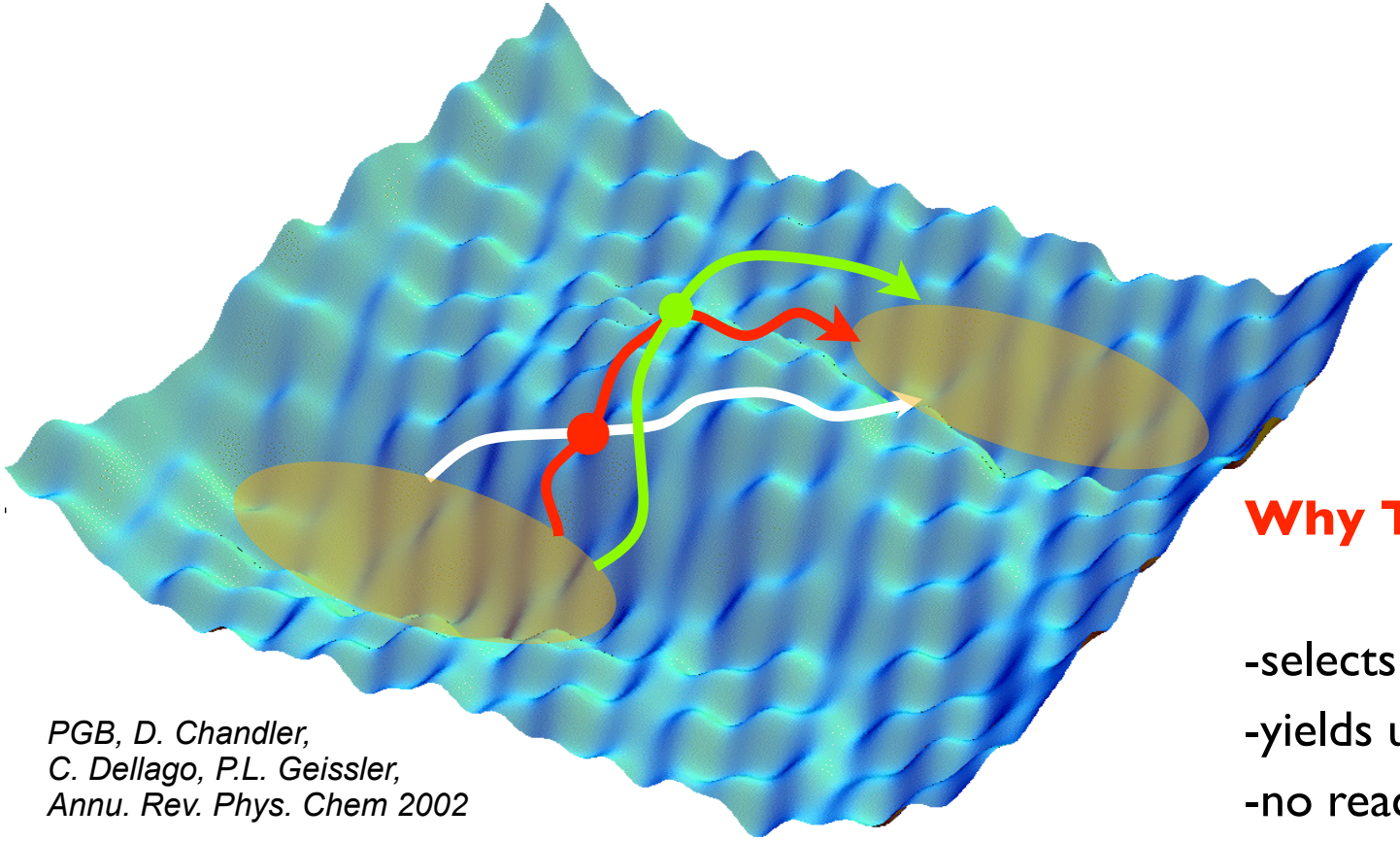




- Straightforward MD inefficient
- enhanced sampling: thermodynamic integration, umbrella sampling, hyper dynamics, adaptive biasing force, metadynamics, ….
  makes exponential barrier problem linear
  requires good reaction coordinate

Transition path sampling

Importance sampling of the rare event path ensemble:
all dynamical trajectories that lead over (high) barrier and connect stable states.



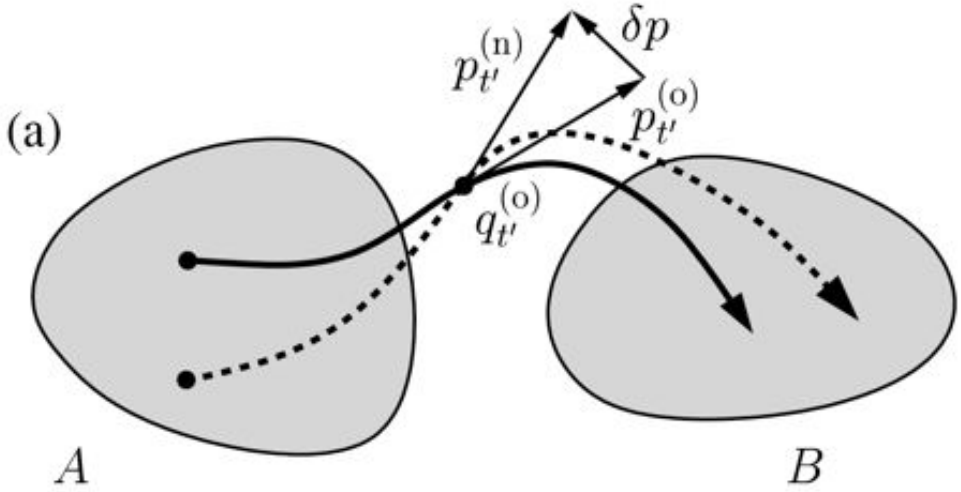
PGB, D. Chandler,
C. Dellago, P.L. Geissler,
Annu. Rev. Phys. Chem 2002

C. Dellago, PGB, Adv Polym Sci, 2009

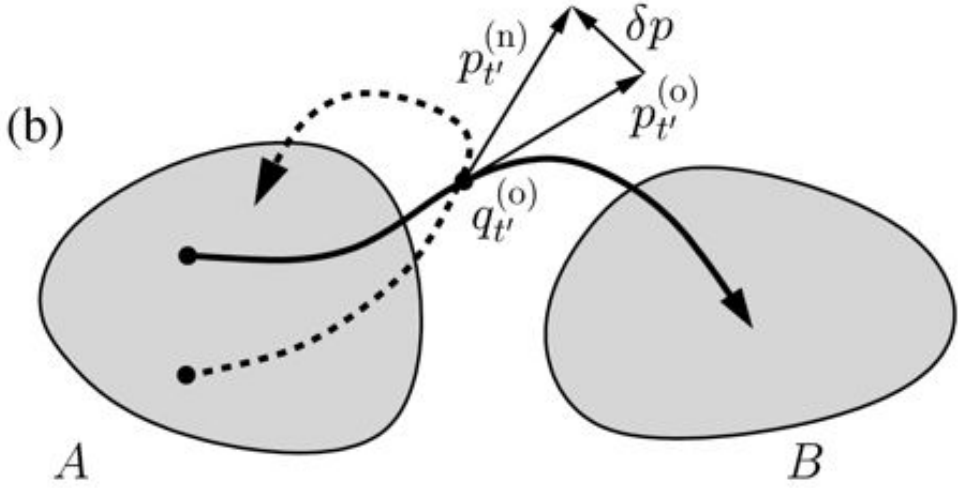
Why TPS?

-selects important rare paths
-yields unbiased molecular dynamics
-no reaction coordinate needed
-reaction coordinate from committor
-rate constants

Shooting moves



accept

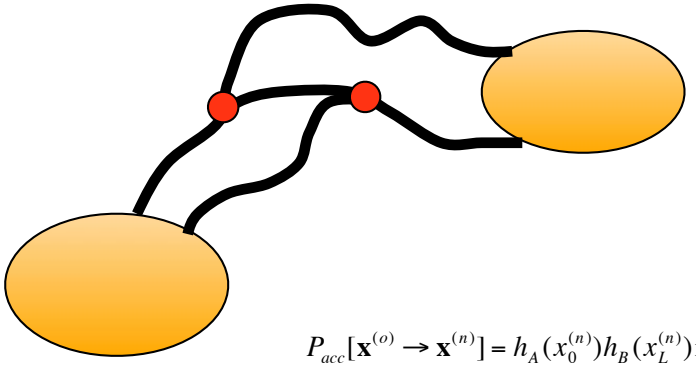


reject

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

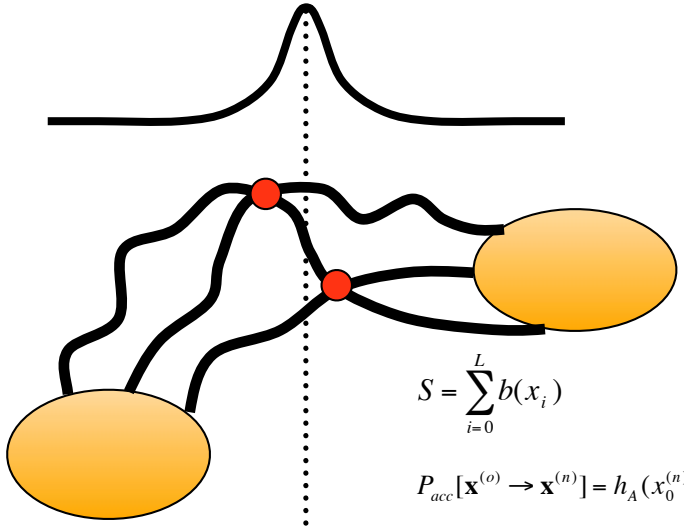
$$h_A(t) = \begin{cases} 1 & if \quad x_t \in A \\ 0 & if \quad x_t \notin A \end{cases}$$

Flexible one way shooting (TIS)



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

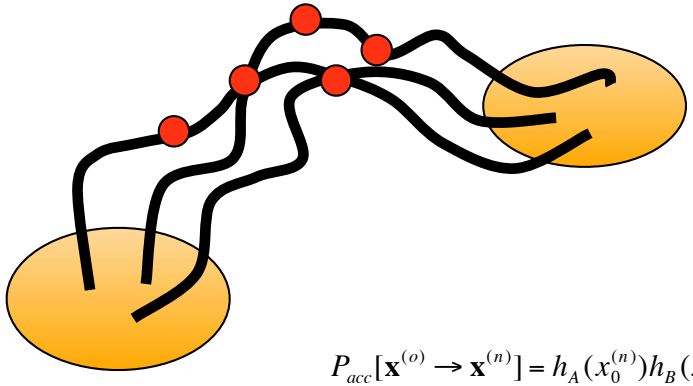
Biased shooting



$$S = \sum_{i=0}^{L} b(x_i)$$

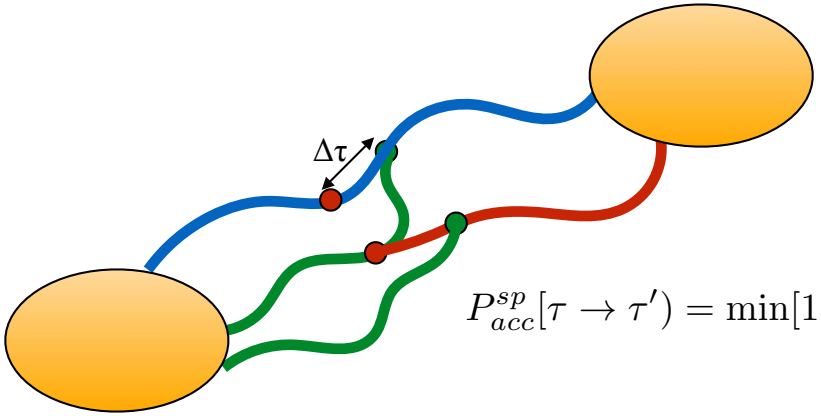
$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)}) \min\left(1, \frac{S^{(o)}}{S^{(n)}}\right)$$

Aimless shooting



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

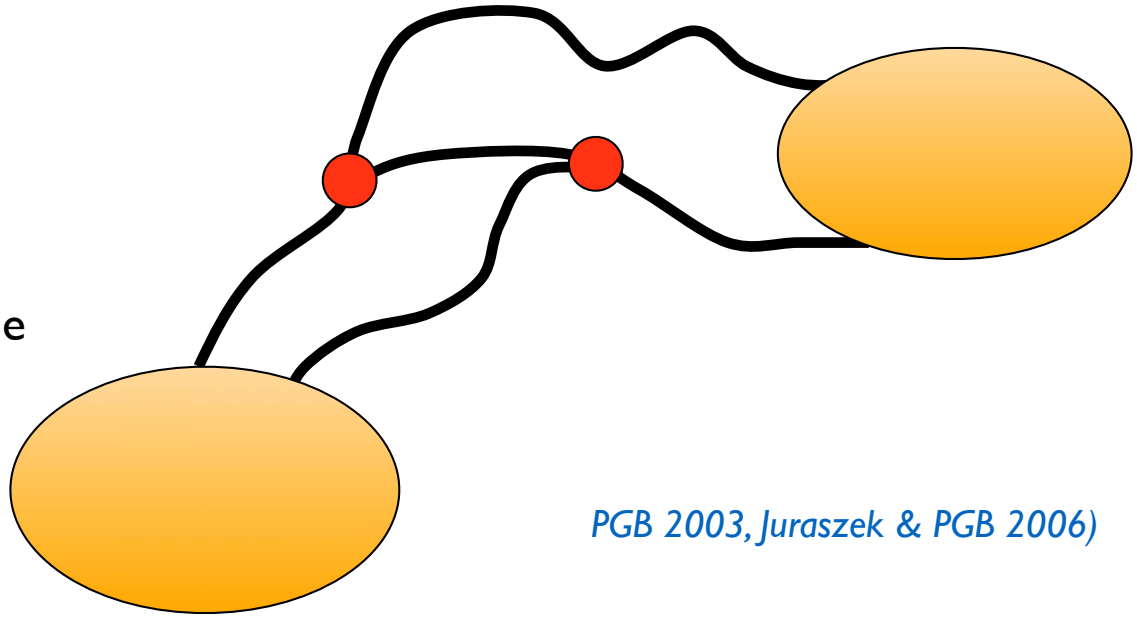
Spring shooting



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

Flexible one way shooting

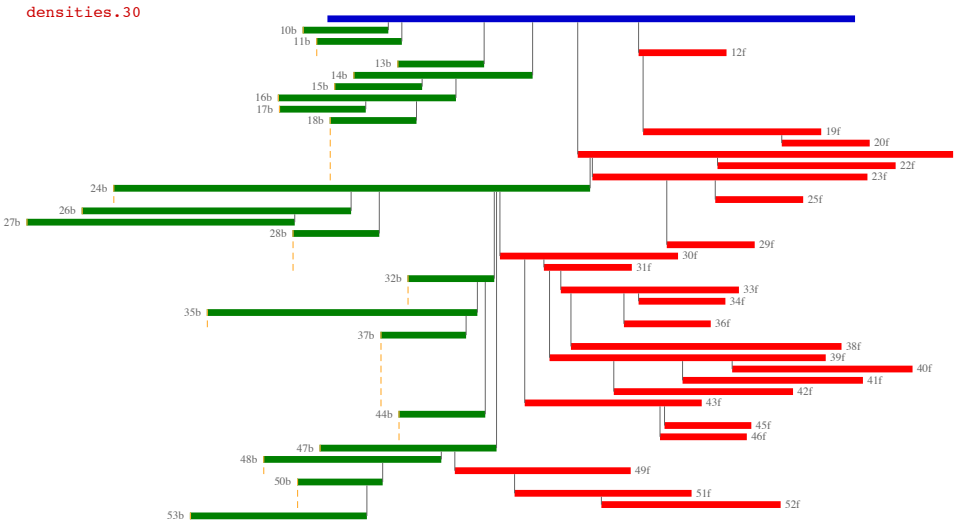
- Algorithm
  - Choose new shooting point randomly from old path $p_{sel} = 1/L$
  - Do **not** alter momenta
  - Integrate in one direction until one stable states is reached
  - keep old partial path, accept new partial with probability



PGB 2003, Juraszek & PGB 2006)

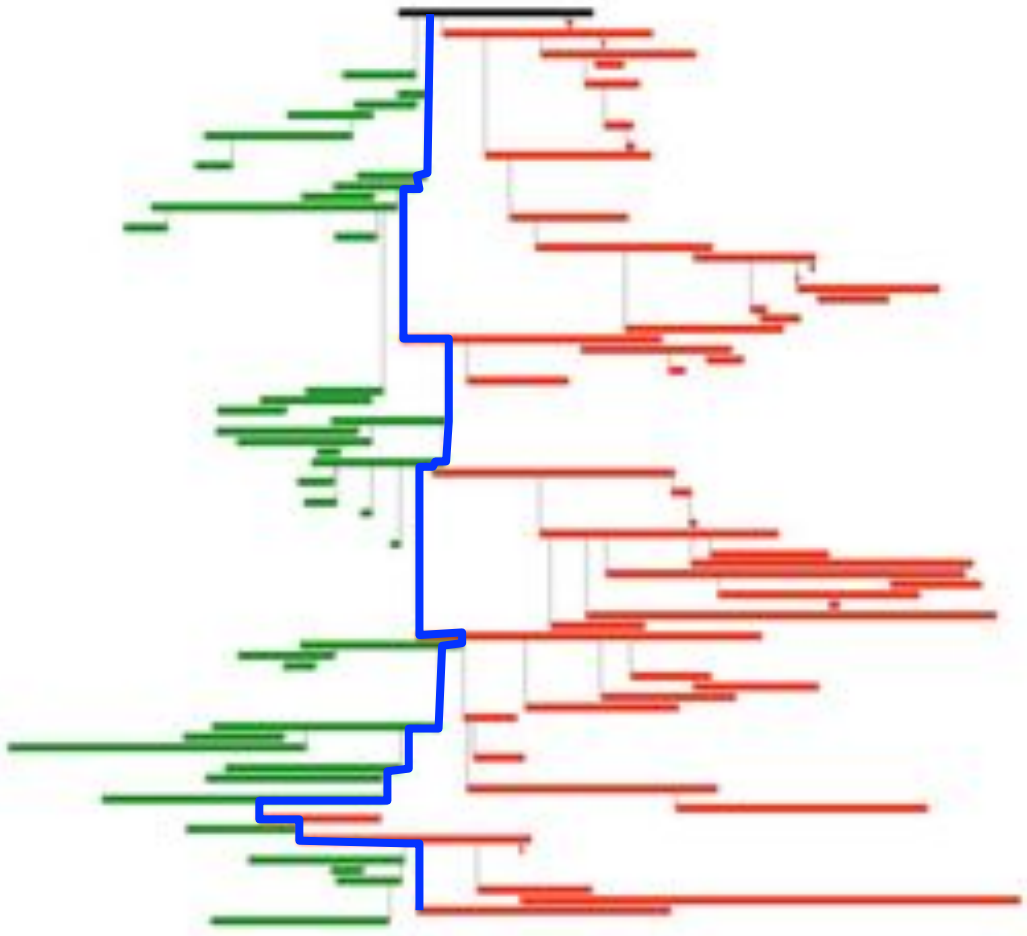
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)}) \min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$

- higher acceptance, better convergence for diffusive transitions and long pathways
- requires some stochastic dynamics, e.g. thermostat
- needs check for decorrelation of paths
- useful for diffusive (bio)systems

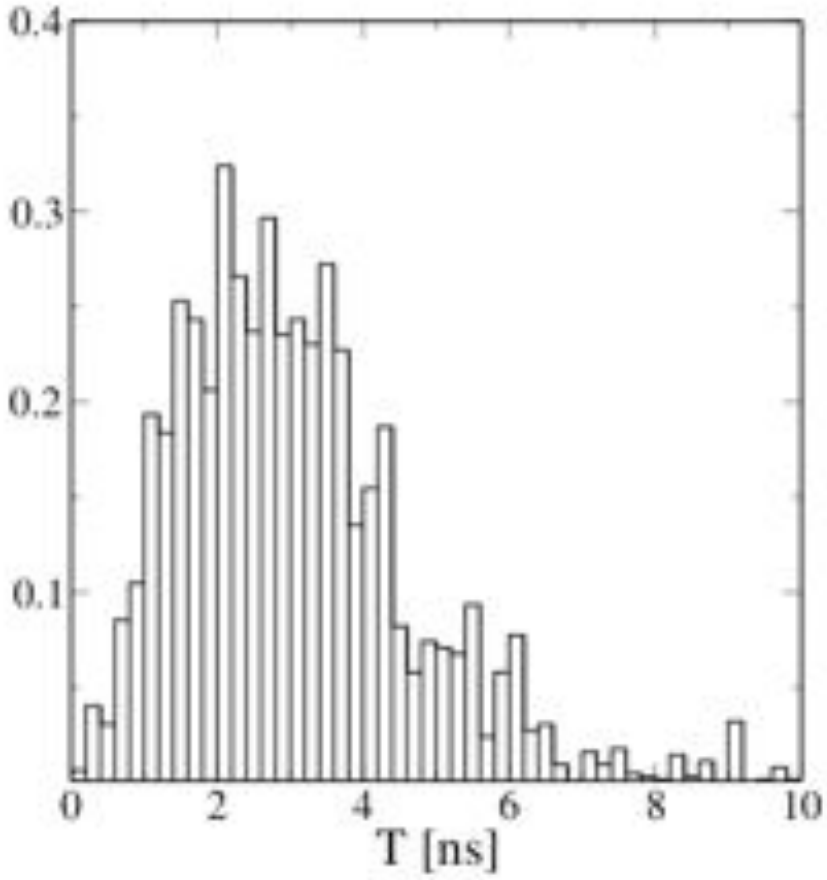


Path sampling indicators

Least changed path

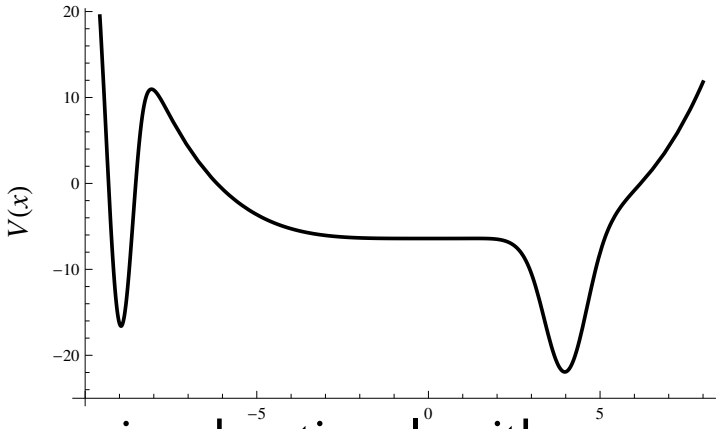


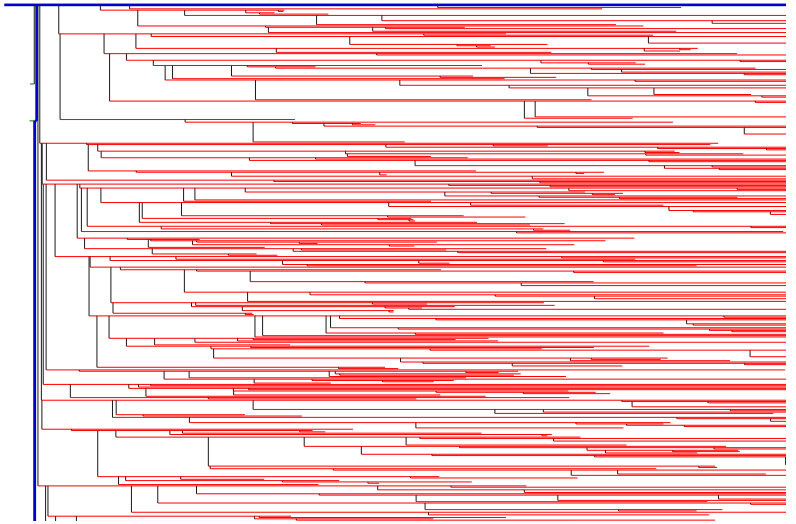
Path length distribution



Spring shooting for asymmetric barriers

- uniform one way shoot has bad decorrelation



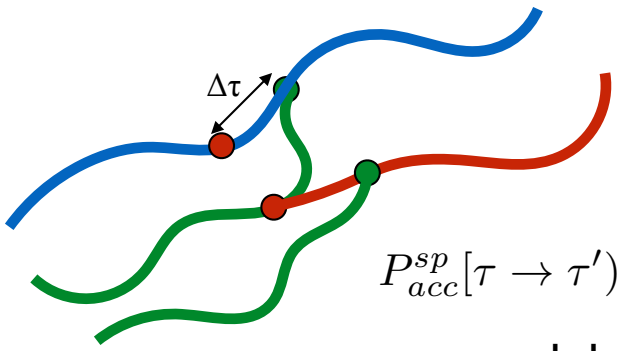


bad decorrelation

- spring shooting algorithm:

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

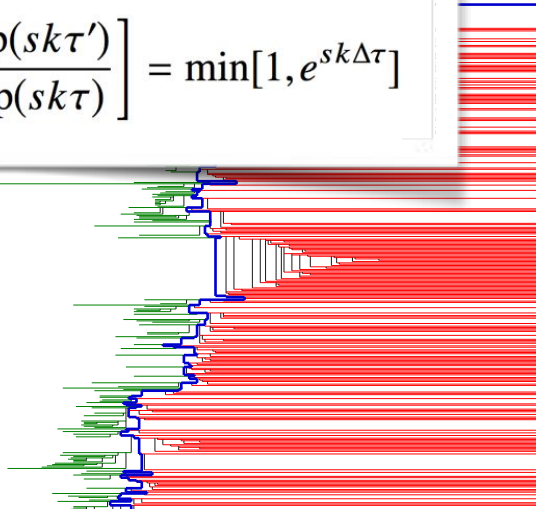
$$P_{acc}^{sp}[\tau \to \tau'] = \min\left[1, \frac{\exp(sk\tau')}{\exp(sk\tau)}\right] = \min[1, e^{sk\Delta\tau}]$$



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

pro: much better decorrelation

con: need optimisation of k and Δτmax



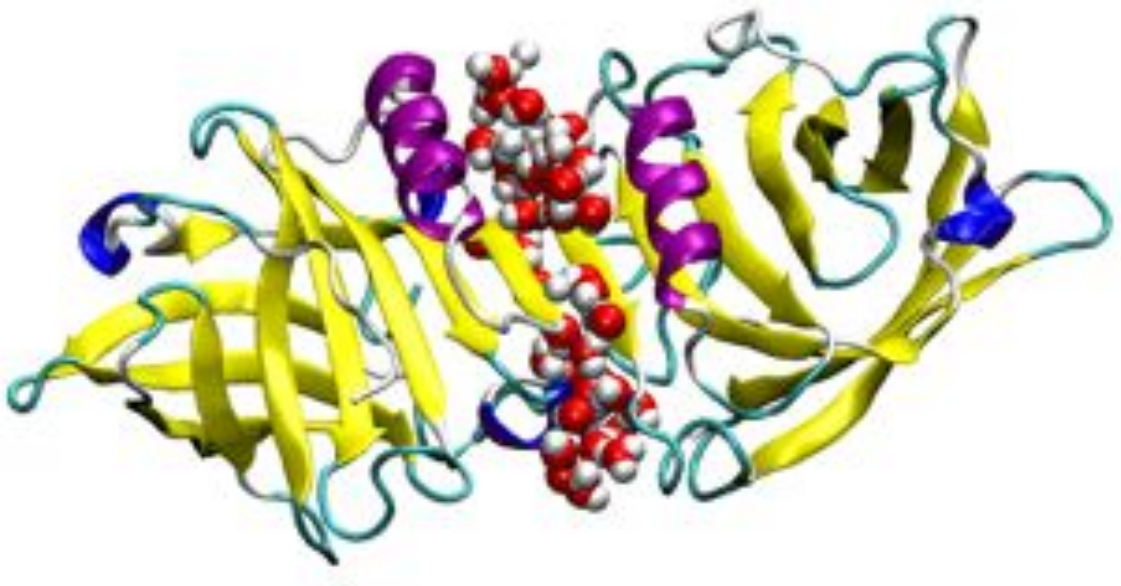
good decorrelation

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

Protein association/dissociation

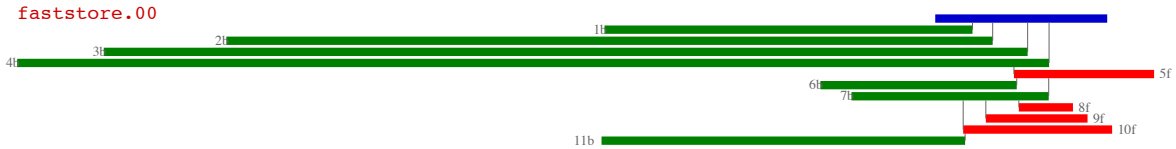
System

- beta-lac dimer (2AKQ)
- ~65000 atoms
- AMBER99SB
- T=300K, P=1 atm
- dissociation ΔG=30 kJ/mol



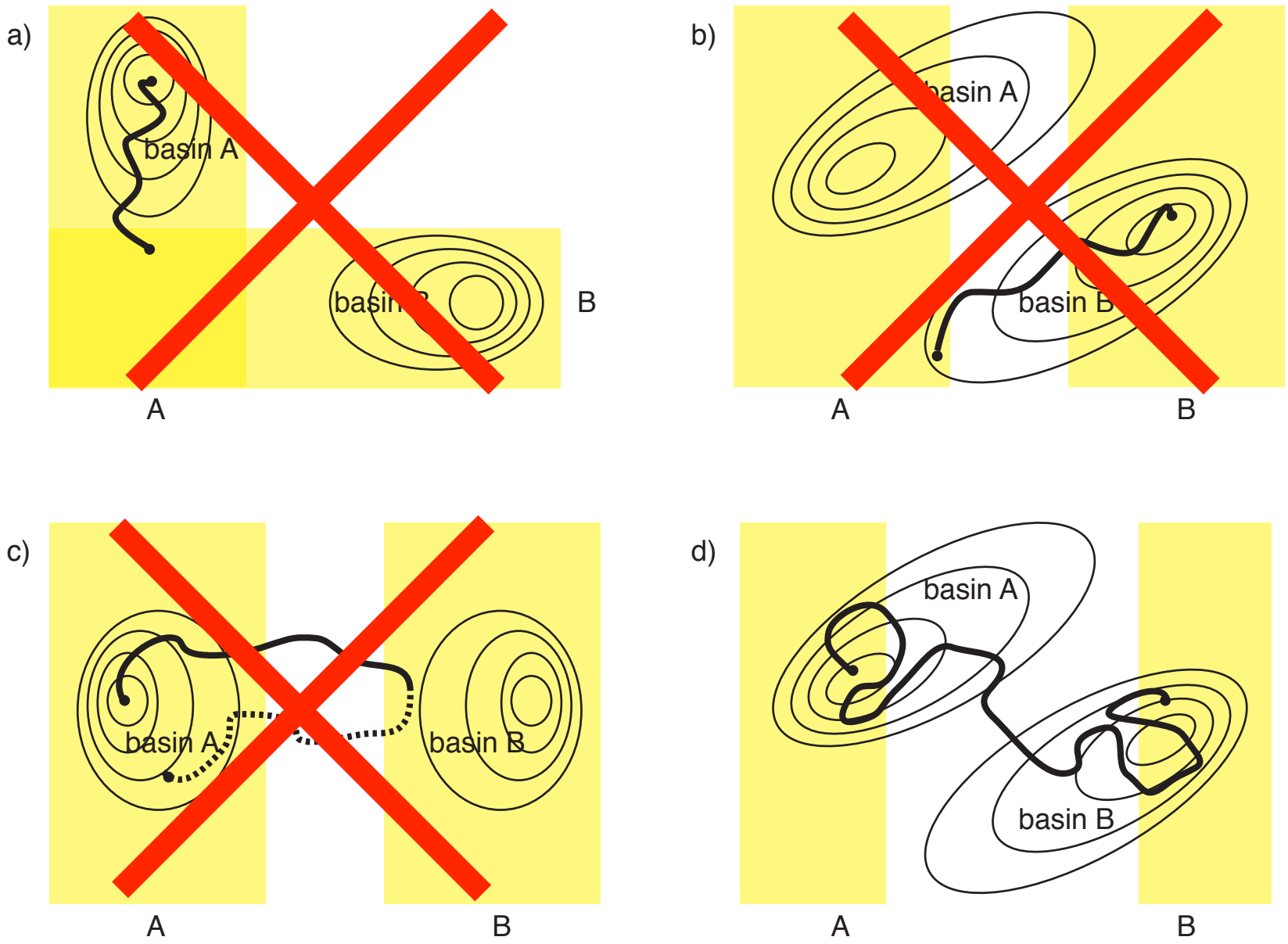
TPS of extremely asymmetric barrier

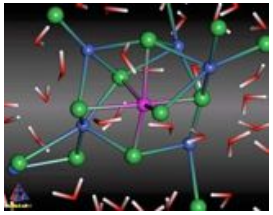
- spring shooting algorithm
- spring const 5 kT, dmax = 200
- 50 ps between frames
- max path length 50 ns
- acceptance 30%

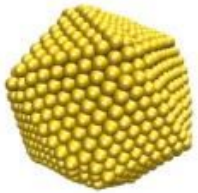


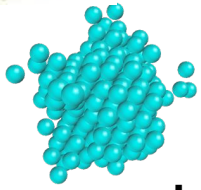
Z.F. Brotzakis, PGB, JCP, (2016)

Definition of stable states

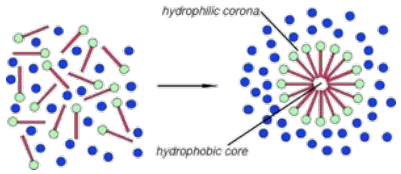


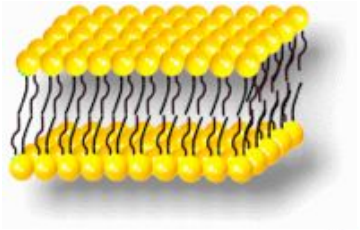




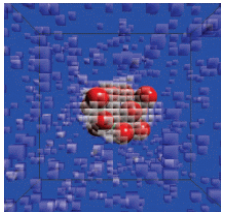


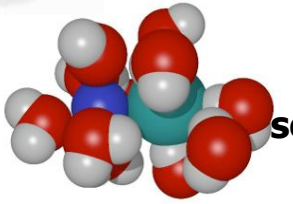
crystallisation



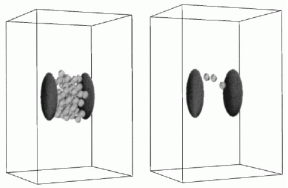


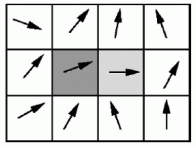
complex fluids

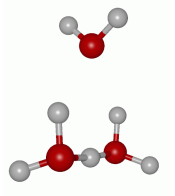




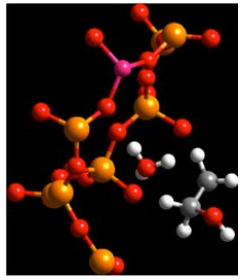
solvent effects

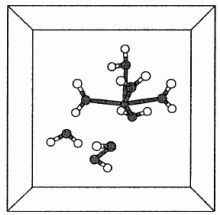




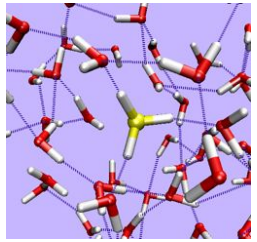


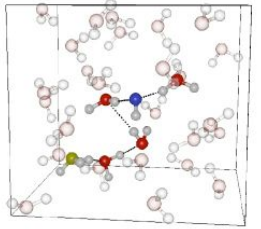
catalysis

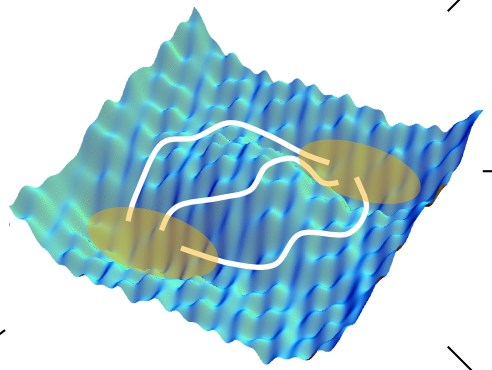


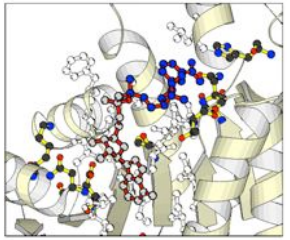


reactions

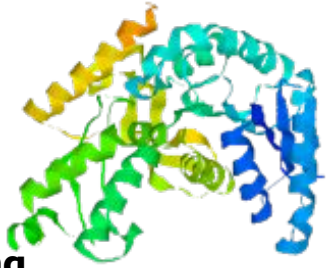




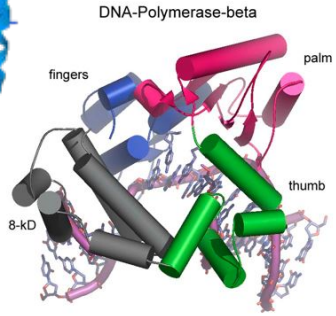




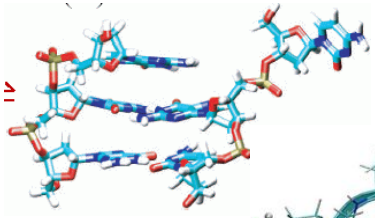
enzyme reactions

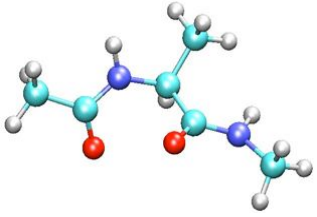


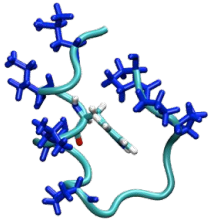
folding & binding

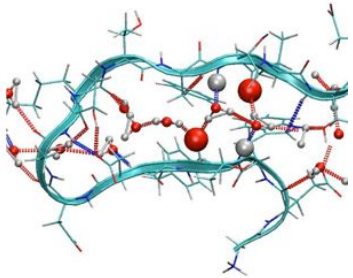


isomerization

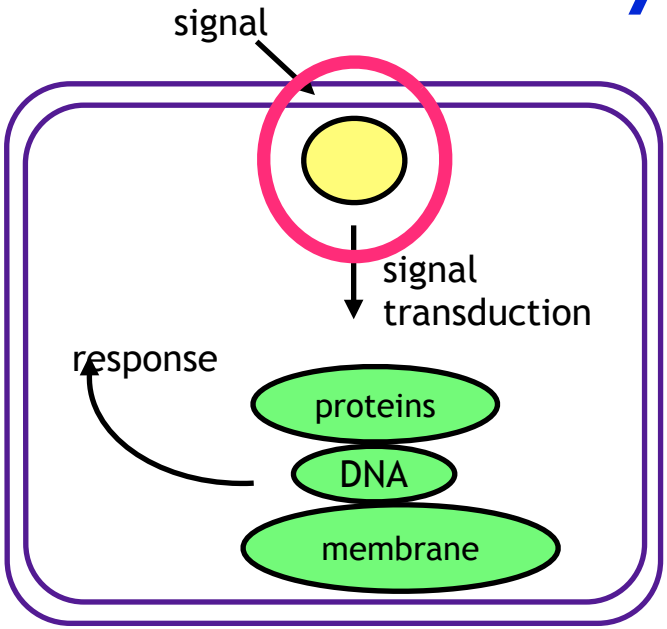


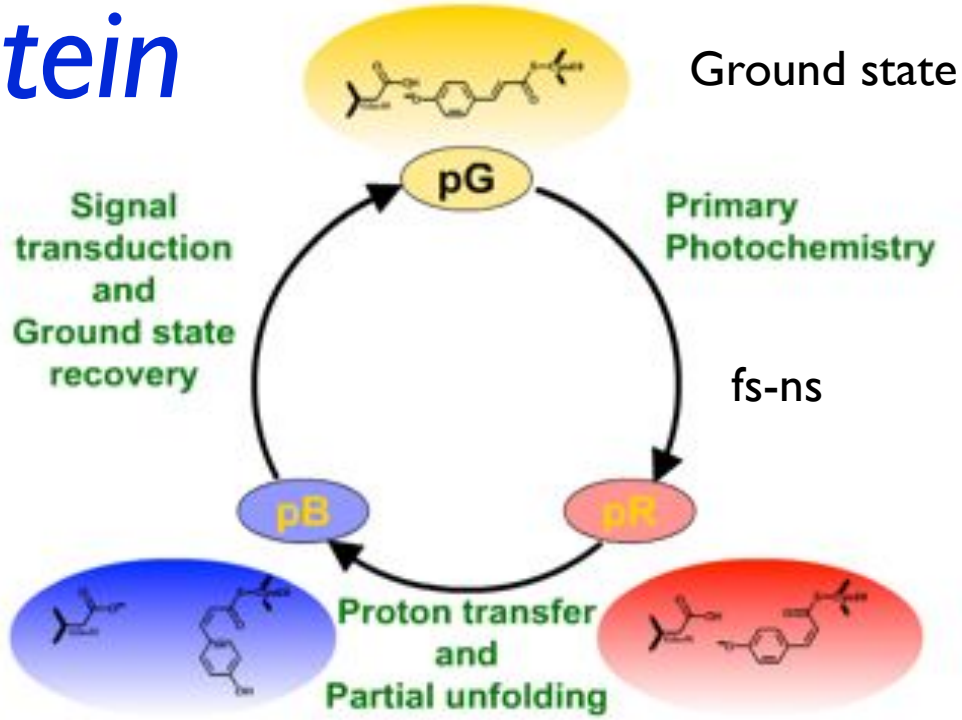


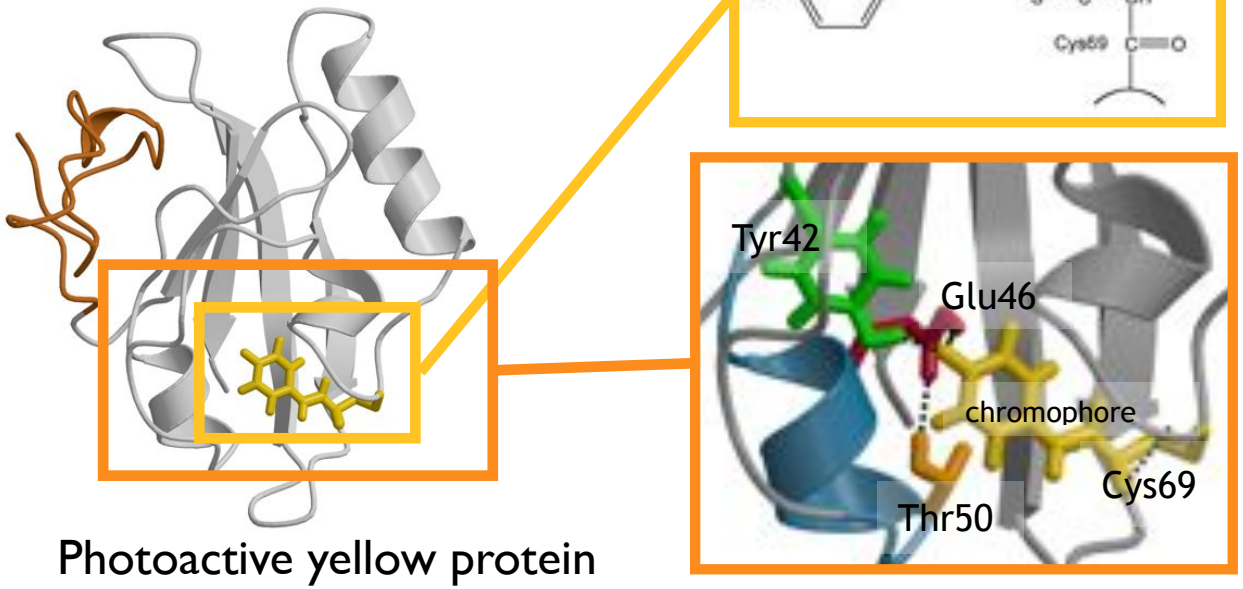




Photoactive yellow protein







Photoactive yellow protein

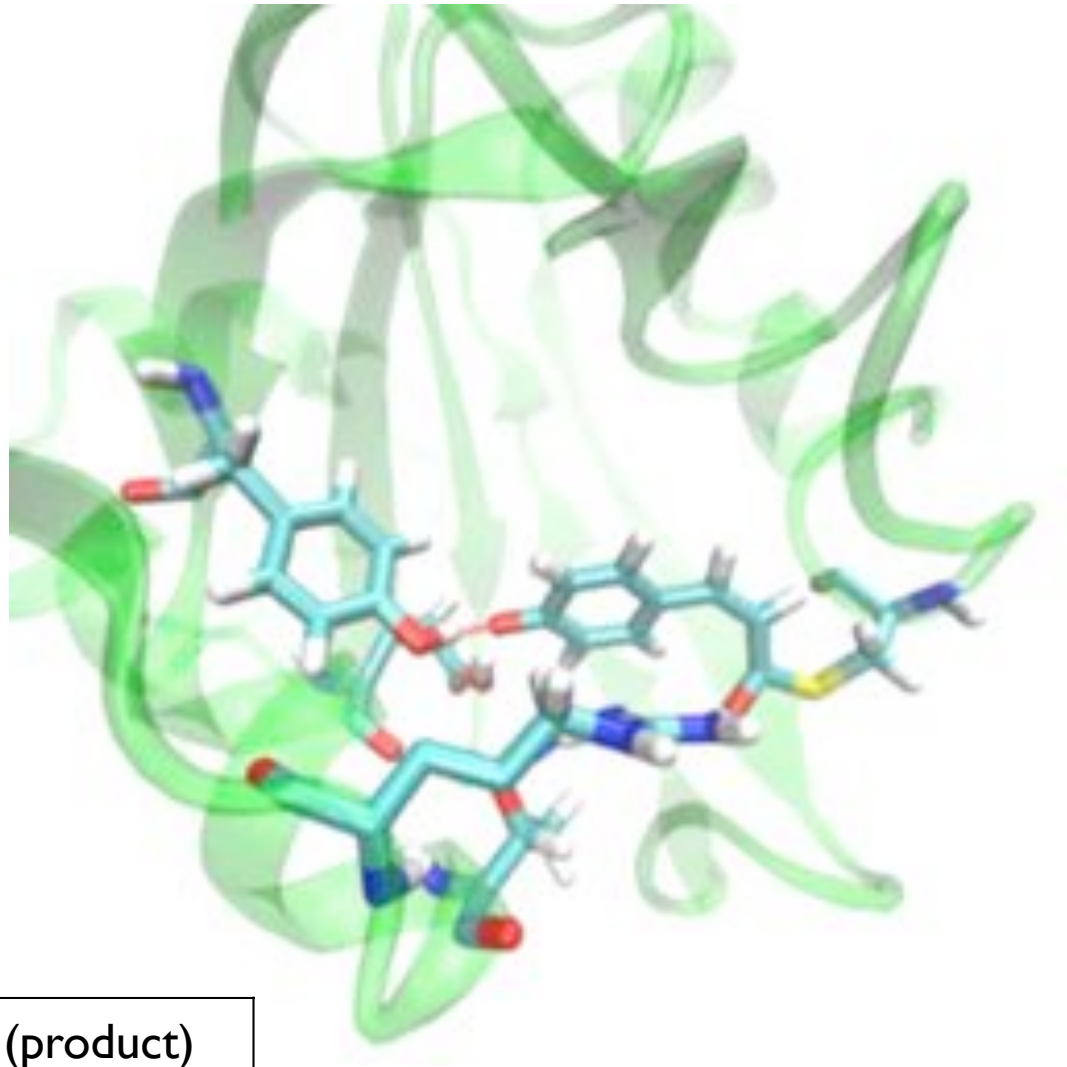
Question: What is the mechanism for amplifying signal?

We studied 2 steps:

1) proton transfer
2) partial unfolding

TPS of proton transfer

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



- TPS: two way shooting, perturbation temp 35 K
- 160 paths/ 50% acceptance
- average path length 0.5-1.5 ps
- reaction time microseconds

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

Transition path sampling of partial unfolding

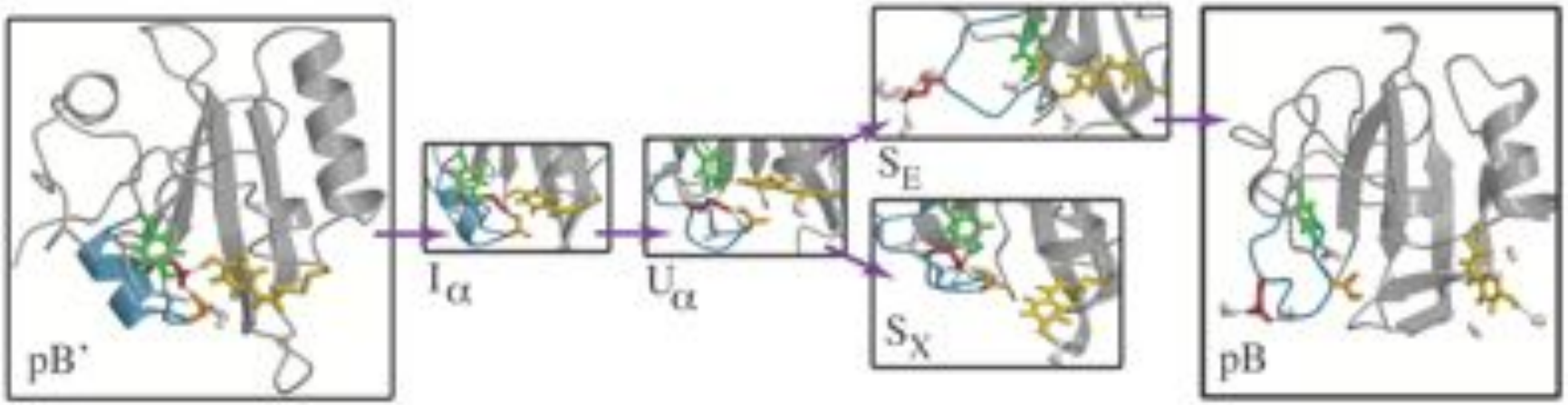
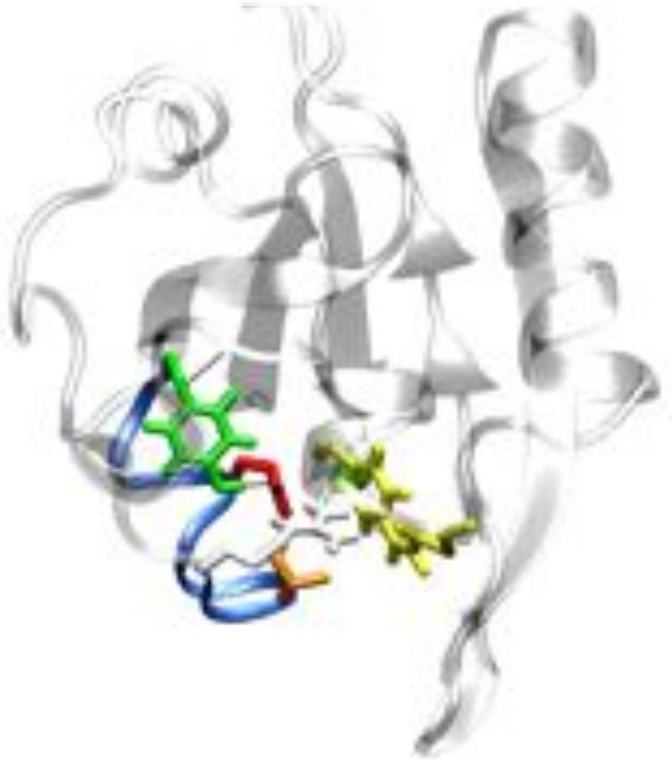


Table 1. Statistics of the TPS ensembles. The average path length is a weighted average over the whole ensemble. Decorrelated pathways have lost the memory of the previous decorrelated pathway. The aggregate time is the ensemble aggregate length

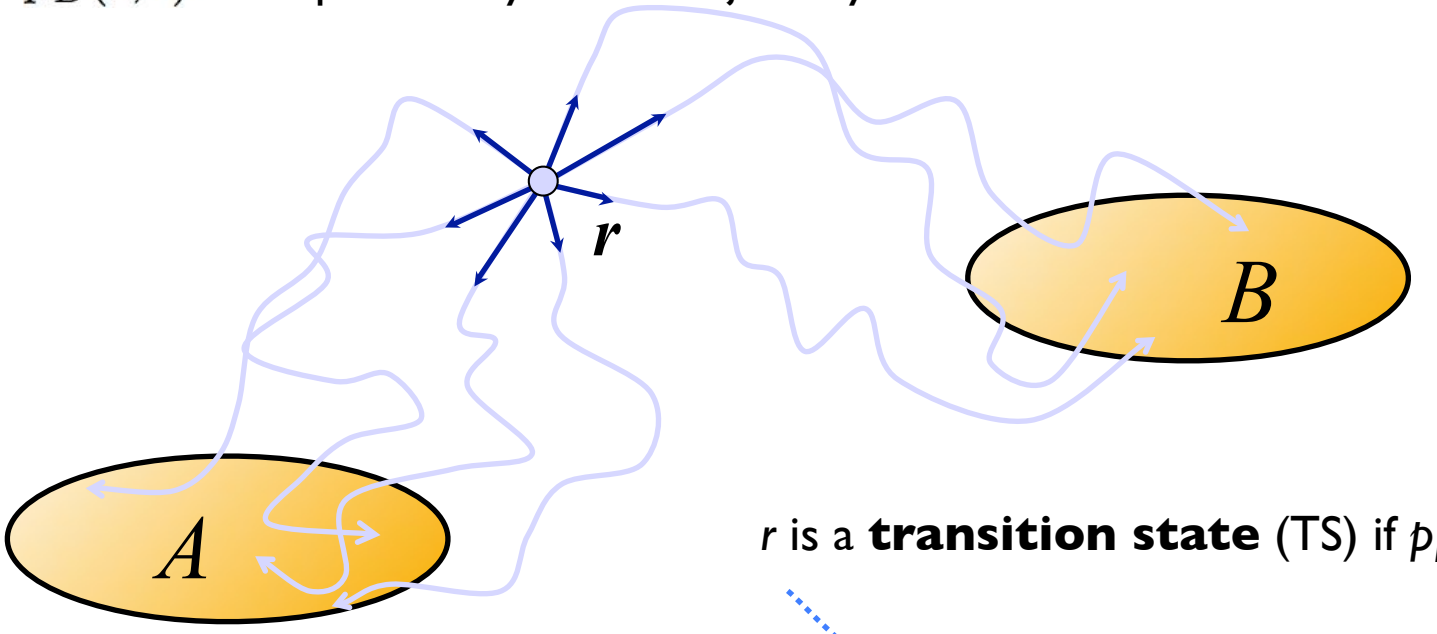
	$pB' - I_\alpha$	$U_\alpha - S_E$	$U_\alpha - S_X$	$S_E - pB$
acceptance	41%	25%	38%	44%
avg. path length	105 ps	1.8 ns	1.5 ns	1.7 ns
accepted paths	3847	305	584	311
decorr. paths	180	18	7	29
aggregate time (μs)	1.0	2.3	2.3	1.2



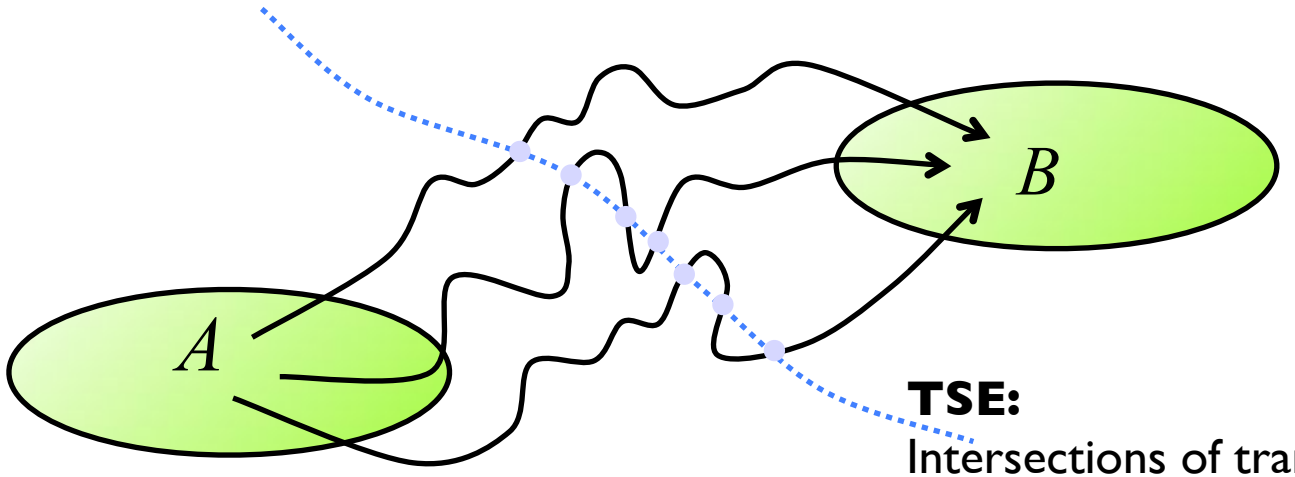
Vreede, Juraszek, PGB, PNAS 2010

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)

Reaction coordinate analysis

- Each TPS shot is a committor attempt. Use this information to optimise model of reaction coordinate *r*

- The probability that structure *x* with rc *r* is on a transition path (for diffusive dynamics)

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

- Assume committor function to be

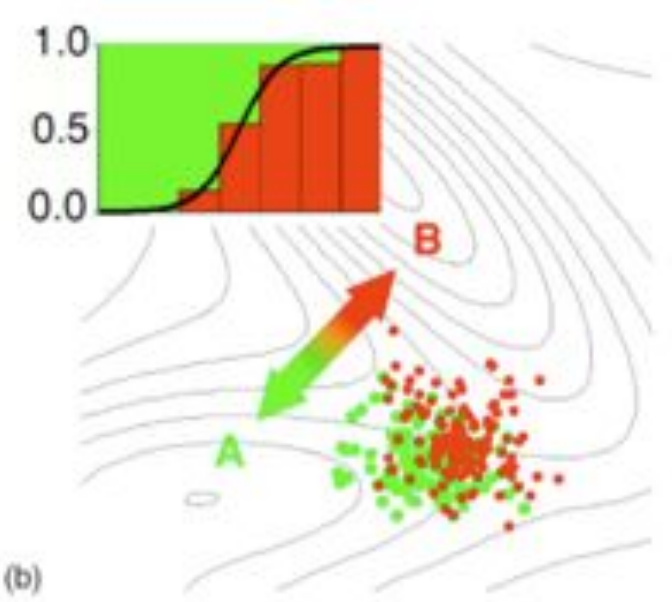
$$p_B(x) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(q(x)\right]$$

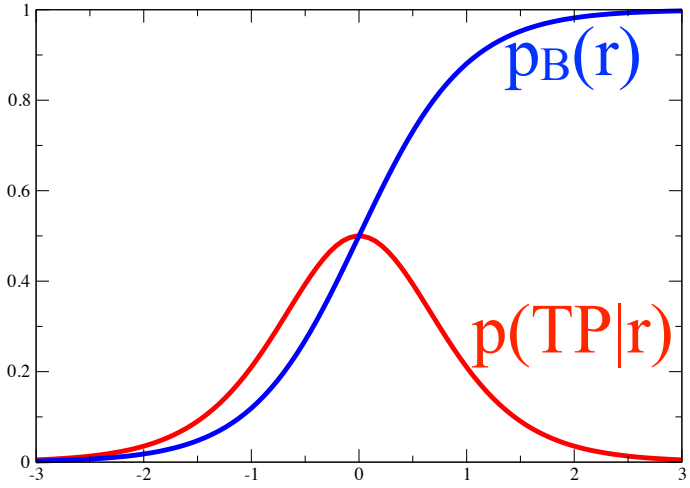
- parametrize *r* as linear combination of q

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

- best *r* is maximizing likelihood

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



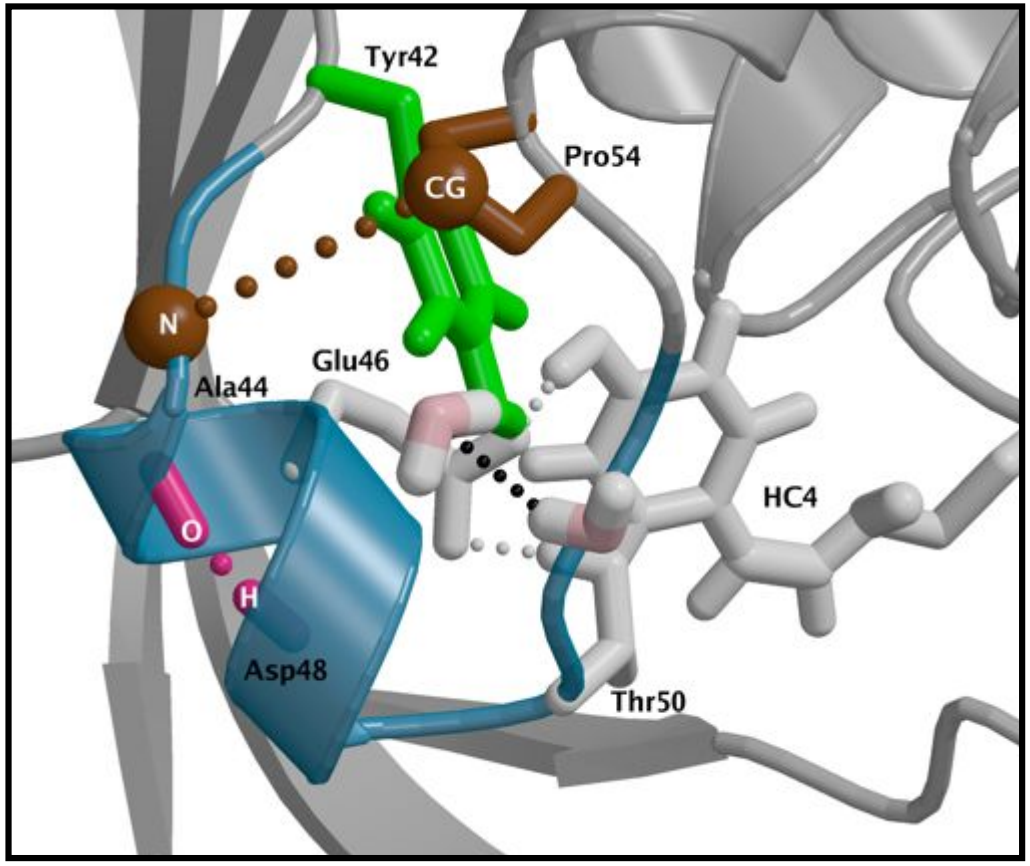


Included order parameters

Number of waters around	
pCA	nw_X
Tyr42	nw_Y
Glu46	nw_Z
Distance between atoms	
pCA-O4' – Tyr42-OH	*dXY*
pCA-O4' – Glu46-CD	*dXE*
pCA-O1 – Cys69-N	*dOaC*
Glu46-CD – Thr50-OG1	*dET*
Glu46-CD – Tyr42-OH	*dYE*
Arg52-CZ – Asp97-CG	*dRD*
Lys64-NZ – Thr70-OG1	*dKT*
pCA-O1 – Asp97-N	*dOaN*
pCA-O4' – Ile49-N	*dXI*
pCA-O4' – Thr50-N	*dXT*
pCA-O4' – Arg52-N	*dXR*
pCA-O4' – Asp97-N	*dXN1*
pCA-O4' – Asp97-CG	*dXN2*
Ala44-N – Pro54-CG	*dPA*
Gly47-CA – Arg52-O	*dGR*
Glu46-CD – Asn43-ND2	*dEN*
Glu46-CD – Gly51-N	*dEG*
Asn43-O – Gly47-H	*dhb1*
Ala44-O – Asp48-H	*dhb2*
Ala45-O – Ile49-H	*dhb3*
Glu46-O – Thr50-H	*dhb4*
Gly47-O – Gly51-H	*dhb5*
Asp20-CG – Lys55-NZ	*dDK*
Asp24-CG – Lys55-NZ	*dDK2*
Glu9-CD – Lys110-NZ	*dEK*
Glu12-CD – Lys110-NZ	*dEK2*
K111-NZ – Glu116-CD	*dKE*

Distance between center of mass of side chains	
pCA – Tyr42	dXY^{com}
pCA – Glu46	dXE^{com}
pCA – Phe62	$dXF1^{com}$
pCA – Phe96	$dXF2^{com}$
pCA – Ile49	dXI^{com}
Lys64 – Thr70	dKT^{com}
Distance between center of mass of groups of residues	
13–17 – 114–116	dN – loop
35–37 – 98–101	dloop1
35–37 – 114–116	dloop2
rmsd	
11–15	$rmsd_{\alpha1}$
19–23	$rmsd_{\alpha2}$
43–51	$rmsd_{\alpha}$
62–68	$rmsd_{C1}$
75–86	$rmsd_{C2}$
111–116	$rmsd_{loop}$
Dihedral angles in pCA	
N-CA-CB-SG	dih_{CACB}
CA-CB-SG-C1	dih_{CBSG}
Other	
Number of hydrogen bonds in α3	*nhb*
Cosines of dihedral angles ϕ in α3	$\phi_{42} - \phi_{53}$
Cosines of dihedral angles ψ in α3	$\psi_{42} - \psi_{53}$

Reaction coordinate of $helix_{\alpha 3}$ unfolding



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

Order Parameters involved (out of 78):

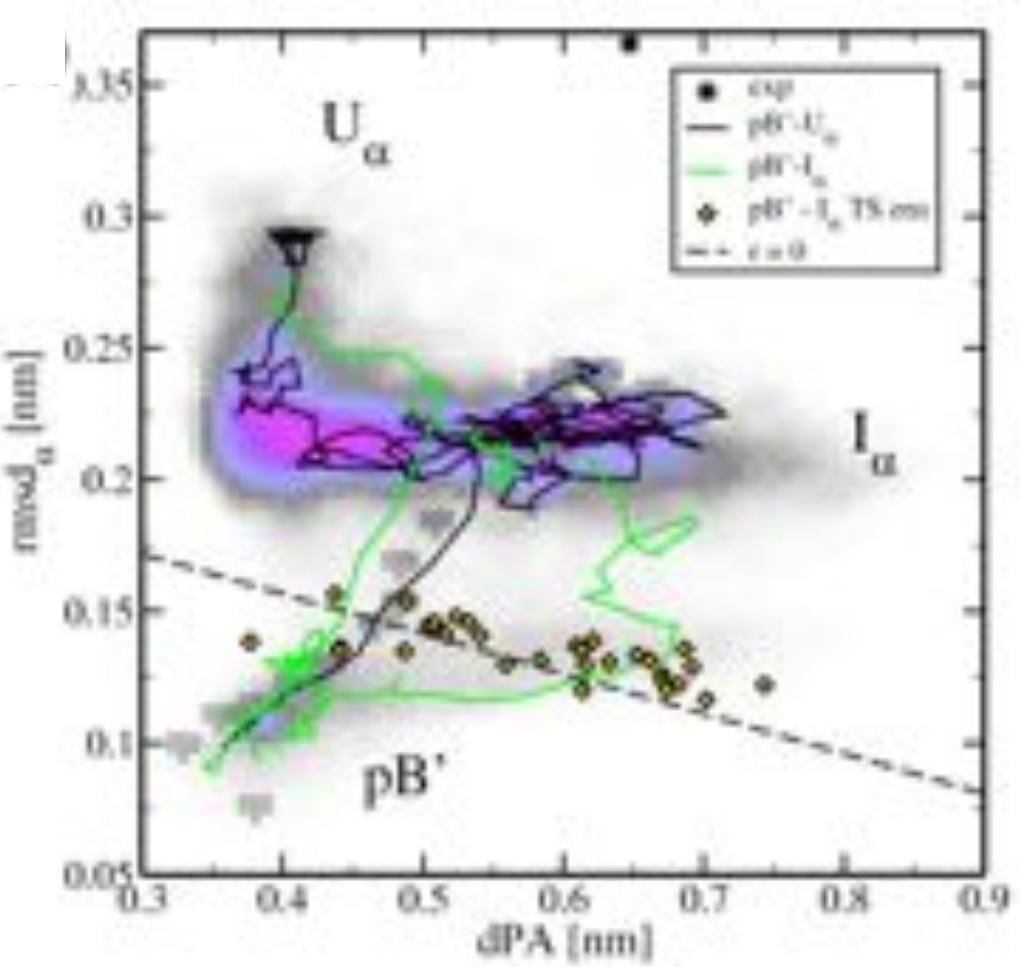
- $RMSD_{\alpha}$
- nwY42 : water molecules around Tyr42
- dPA : distance Ala44(N) - Pro54(Cγ)
- dhb2 : distance Ala44(O) - Asp48(H)

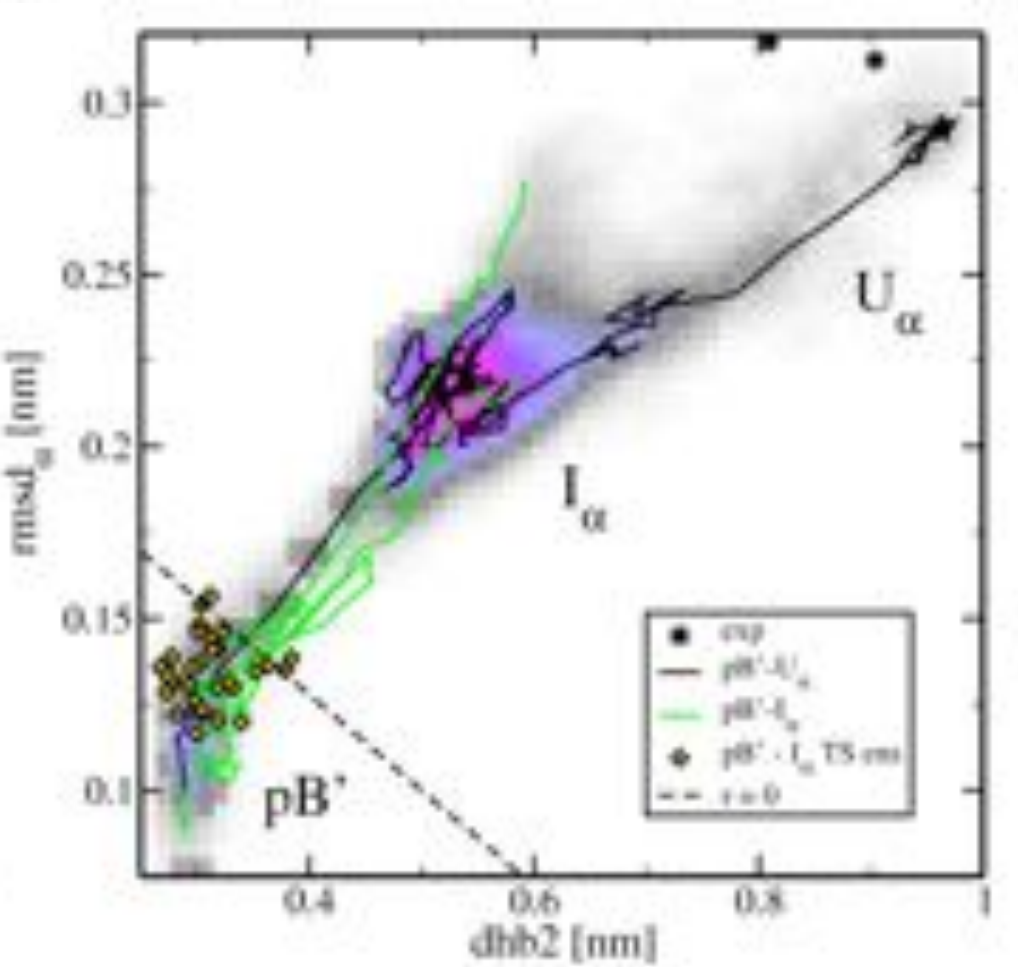
$\delta L_{min} = 4.17$

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

Reaction coordinate pB'→ I_α

r =5.11 -16.28 $rmsd_{\alpha 3}$ -4.68 dhb2- 2.55 d_{PA}

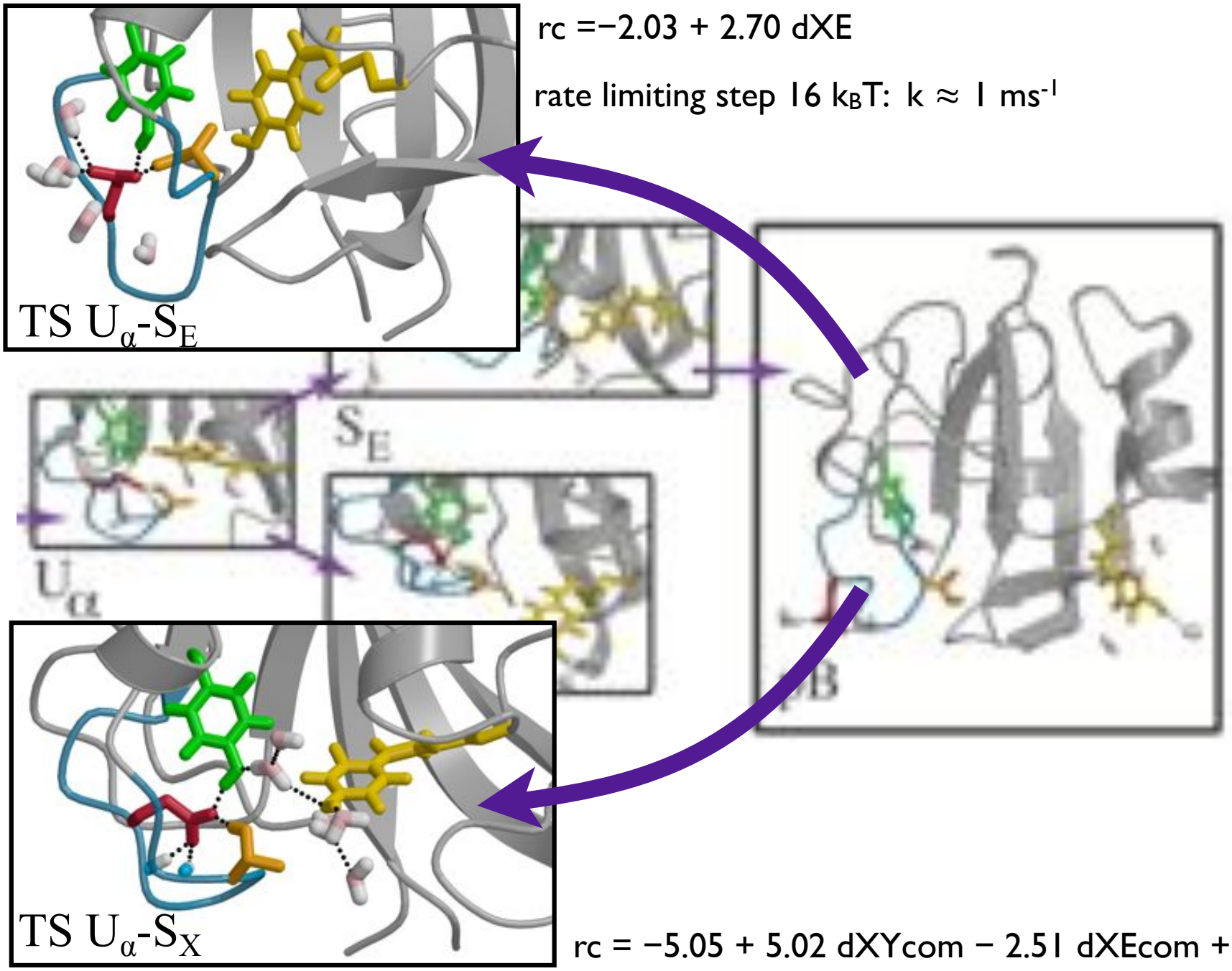




Committor check:
30% of predicted TSE is a true transition state.

Vreede, Juraszek, PGB, PNAS 2010

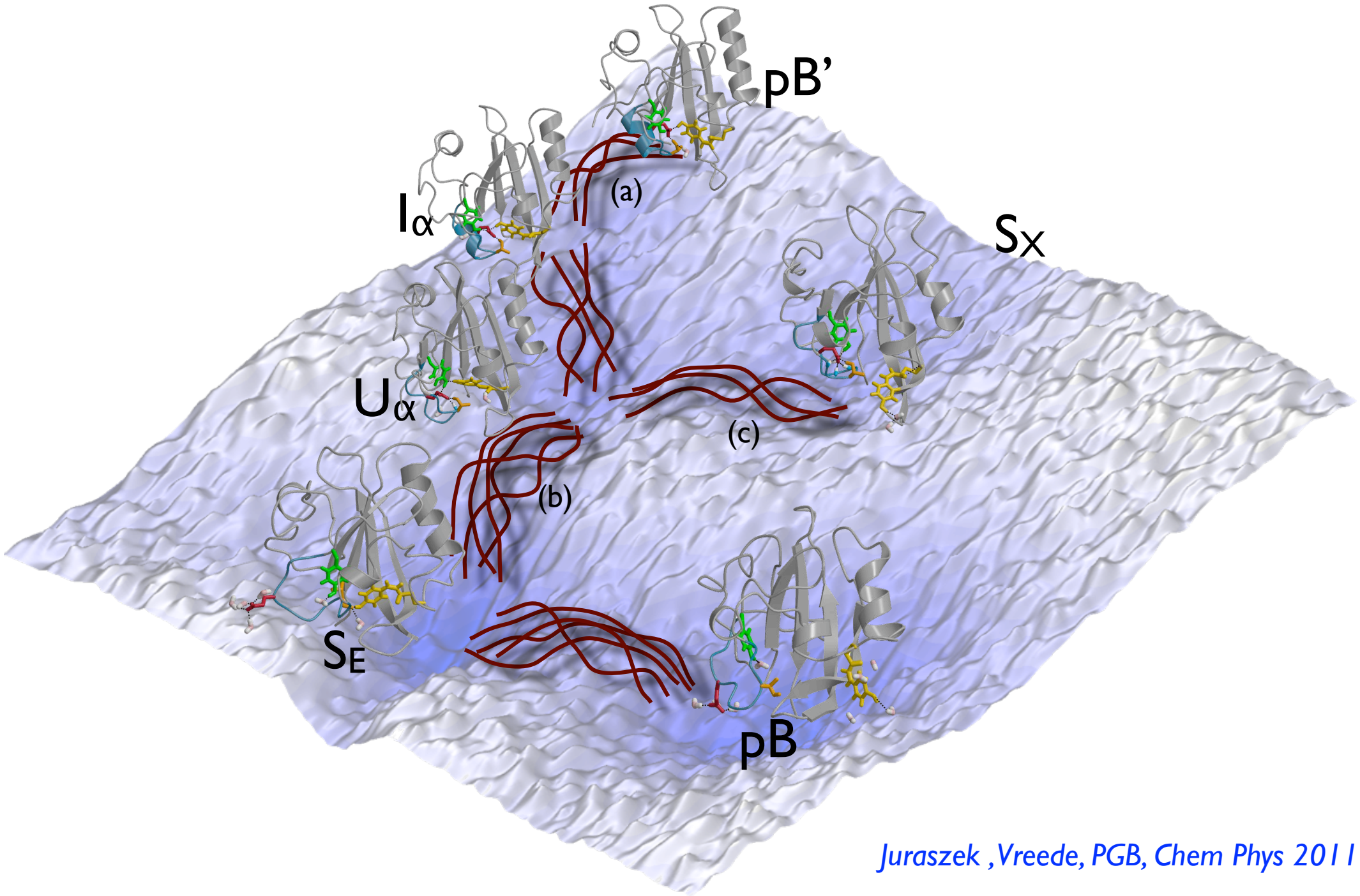
Solvent exposure transitions



rc =−2.03 + 2.70 dXE

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

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



Juraszek ,Vreede, PGB, Chem Phys 2011

Outline

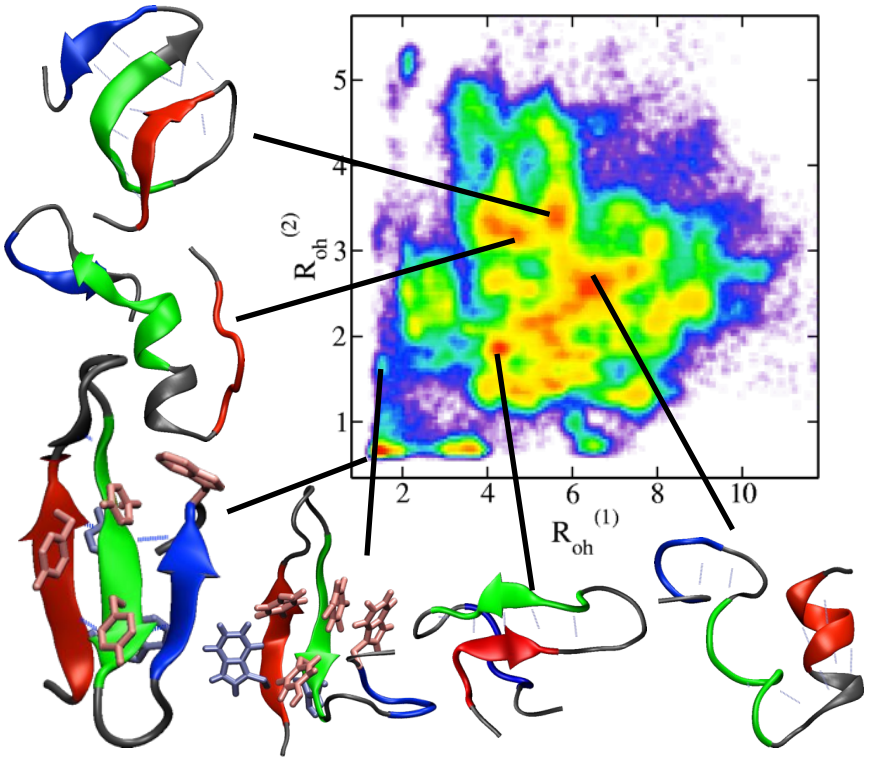
- Simulation of biomolecular systems

- Basic TPS
  - shooting algorithms
  - stable state definitions
  - example Photoactive Yellow Protein
  - reaction coordinate analysis

- Advanced path sampling
  - rates by Transition Interface Sampling (TIS)
  - replica exchange TIS and multiple state TIS
  - single replica multiple state TIS
  - example Trp-cage folding network

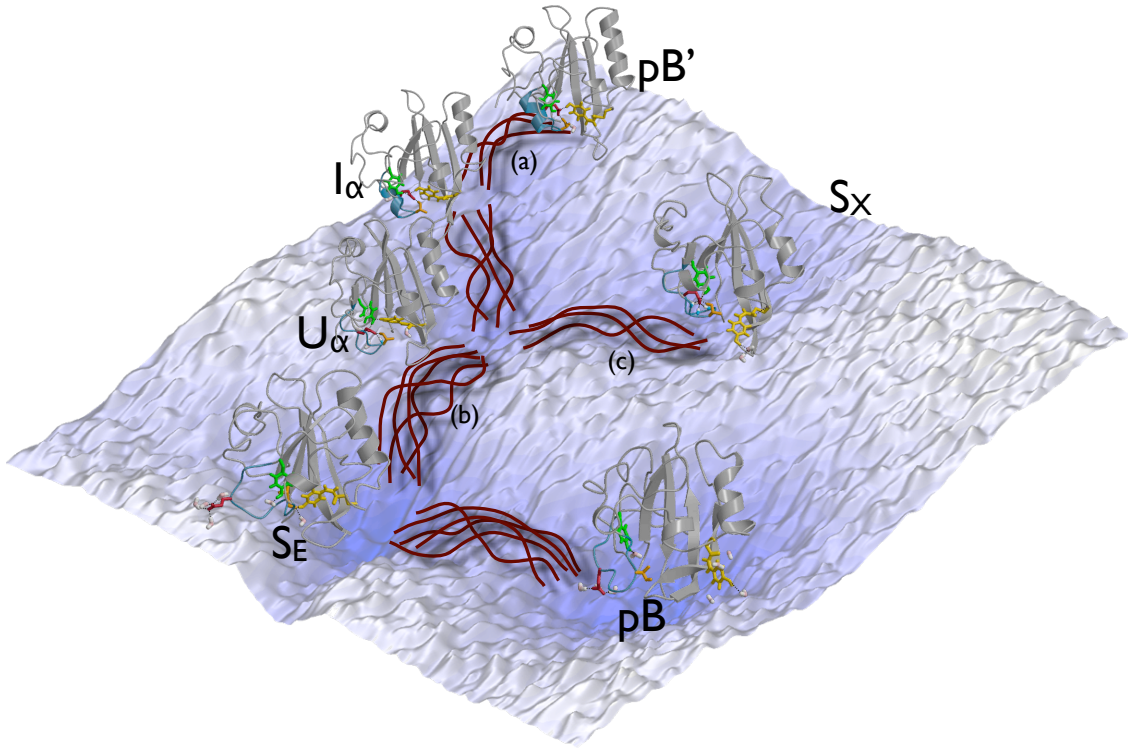
- Conclusion

Rough free energy landscapes



WW domain folding

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

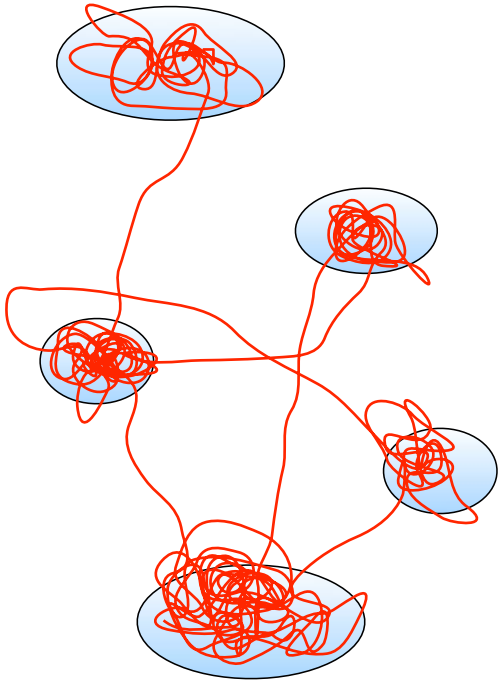


PYP signal transduction

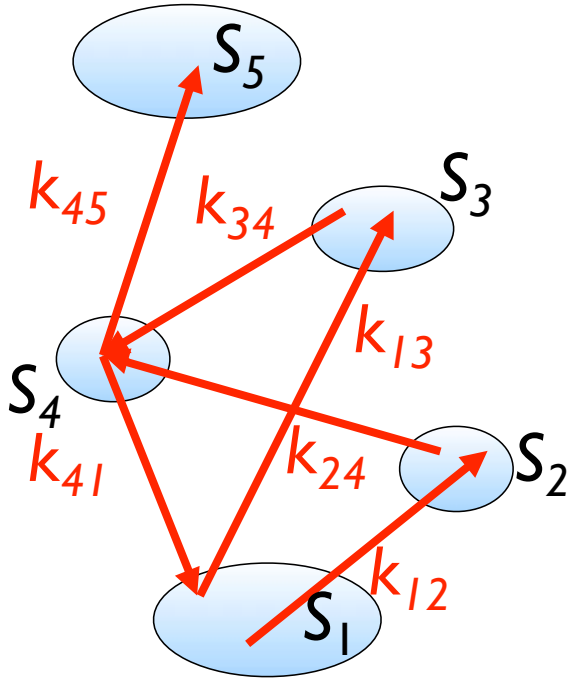
J.Vreede J, J. Juraszek and PGB PNAS 107 2397(2010)

Markov state model

molecular dynamics trajectory



coarse grained trajectory



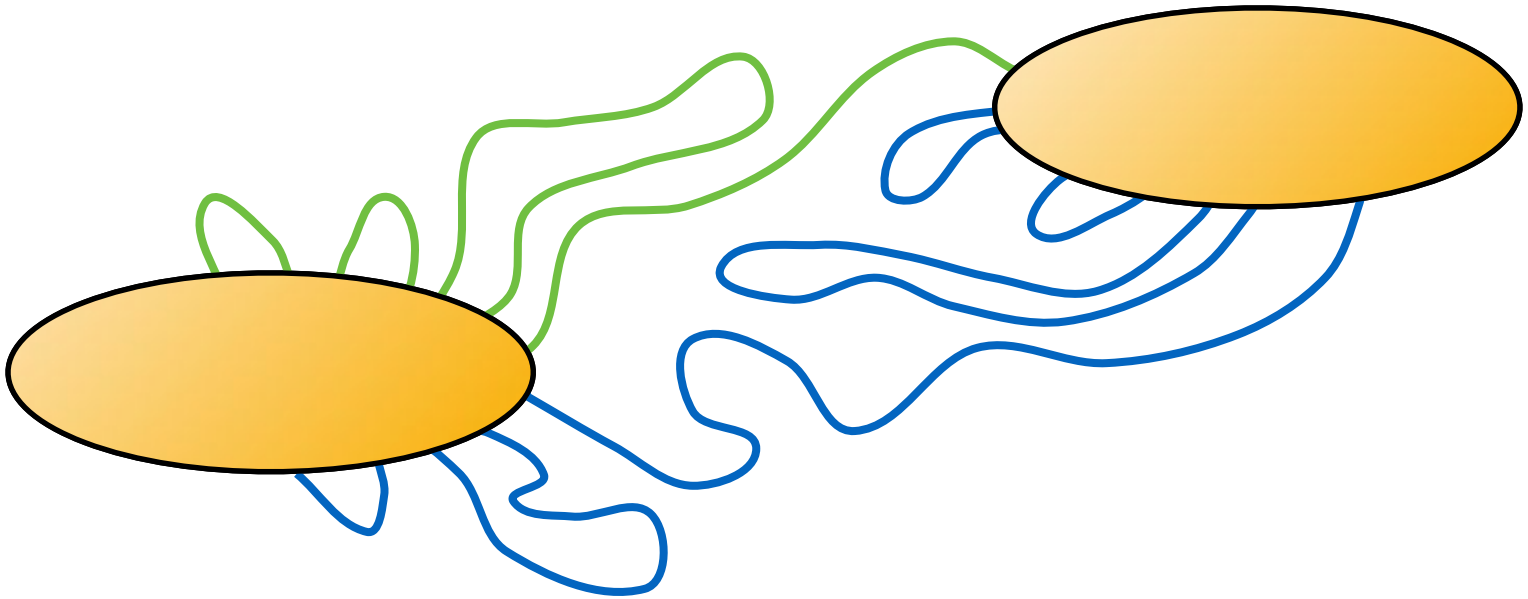
integrate equations of motion

time step Δt ≈ fs

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

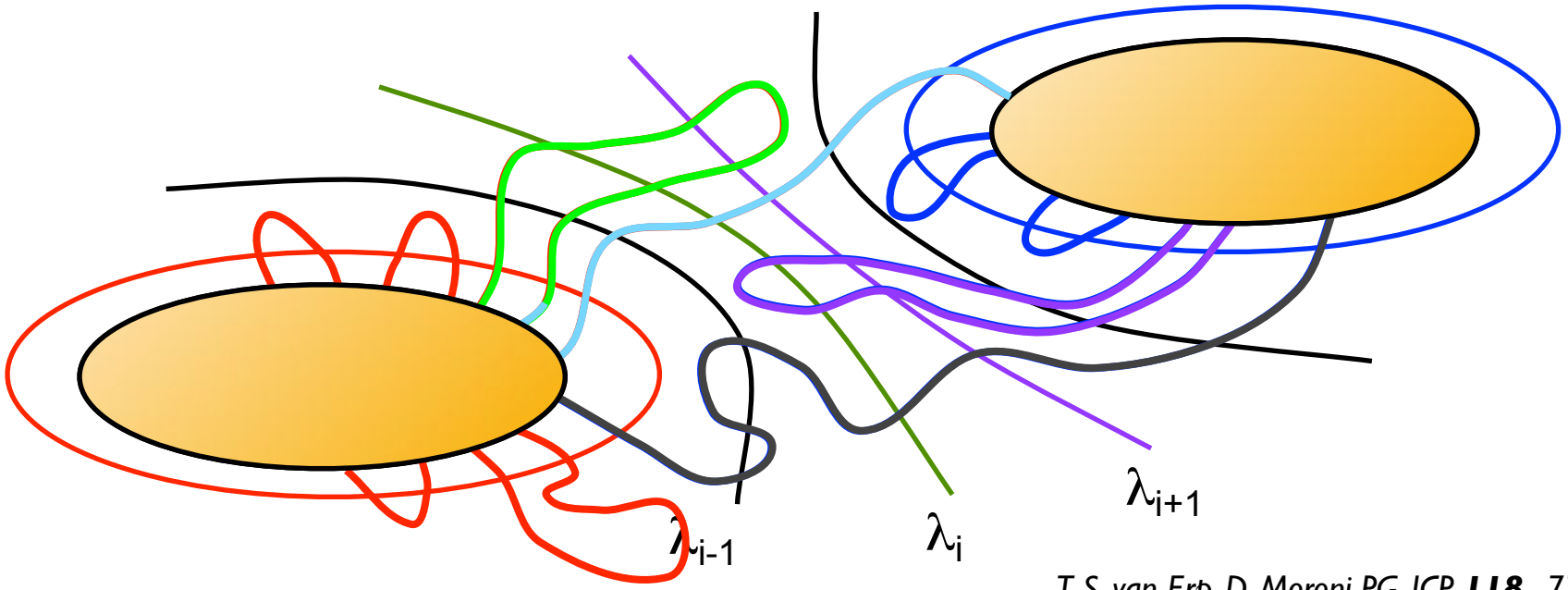
master equation,
solve analytically or by KMC
time step set by rates

Transition interface sampling



Transition interface sampling

Introduce set of interfaces λi



*T. S. van Erp, D. Moroni PG, JCP **118** , 7762 (2003)*

TIS : for each interface i sample pathways that cross λi

$P_A(\lambda|\lambda_i)$ = probability that path crossing λi for first time after leaving A reaches λ before A

Sample with shooting, replica exchange, reversal, first/last interface move

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

Pros and cons

- Advantages TPS/TIS
  - correct rate (recrossings are counted)
  - no reaction coordinate needed
  - access to the entire path space by reweighting
  - mechanistic insight through committor analysis

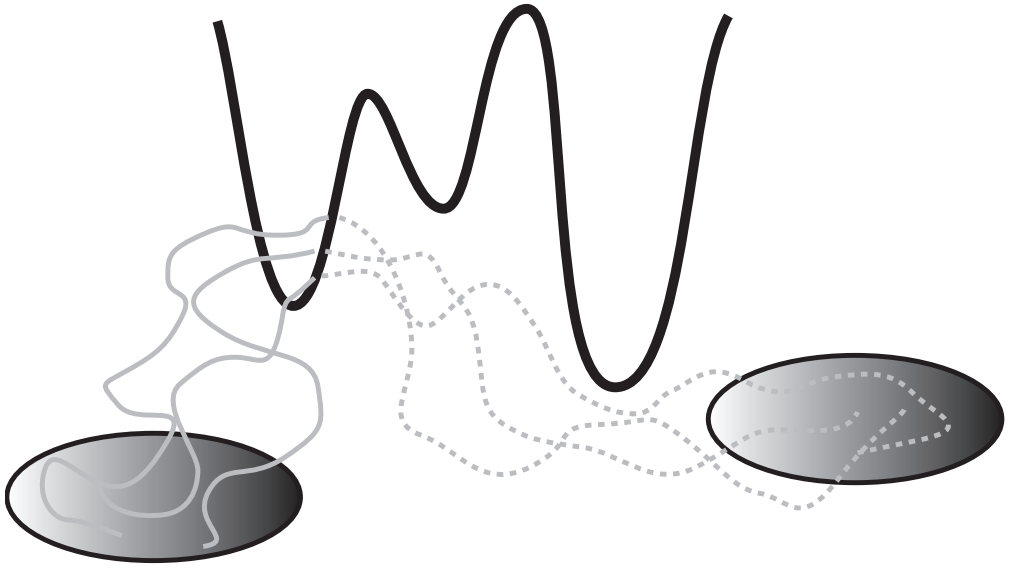
- disadvantage of TPS/TIS
  - stable states need to be carefully defined: core sets
  - not easy to implement
  - computationally expensive

- challenges and convergence issues of TPS/TIS
  - can get trapped in intermediate metastable states
  - multiple channels not easily sampled (addressed by RETIS)

Challenges for path sampling

Multiple channels

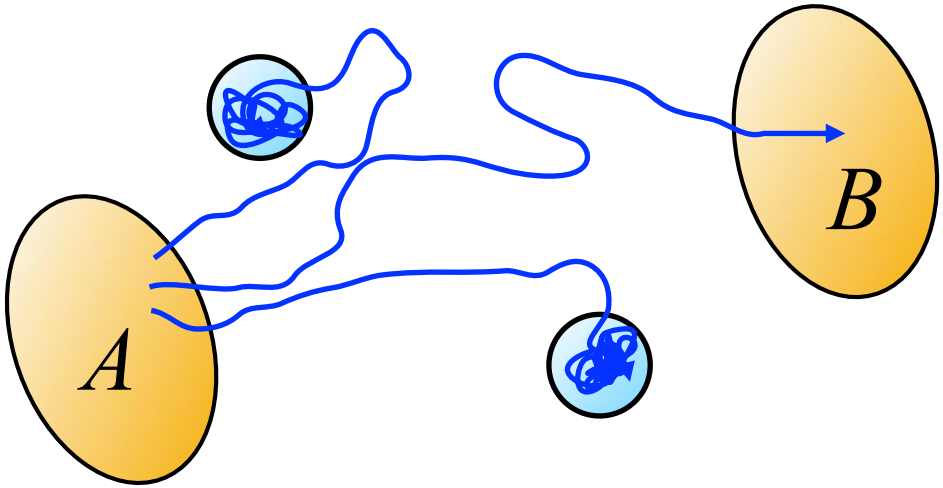
- multiple channels are not sampled properly with shooting
- Replica exchange TIS



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

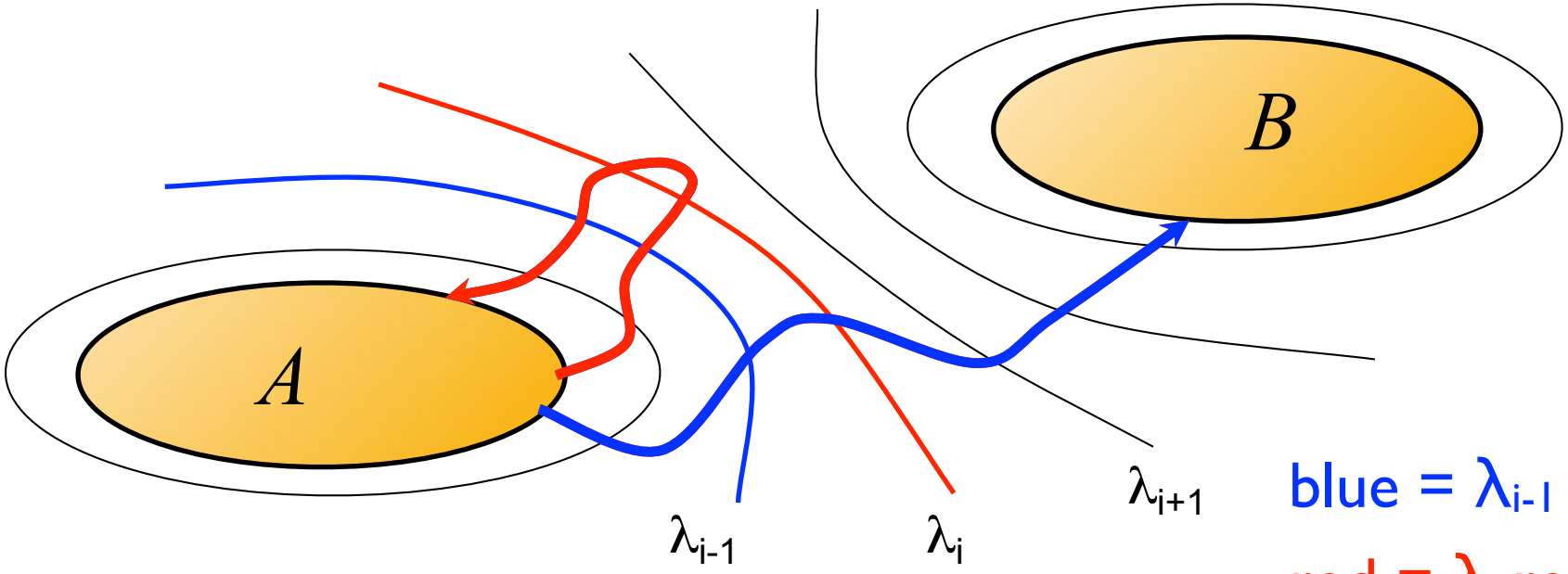
Presence of intermediates

- paths become very long because of intermediates
- Multiple state TIS



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

Path replica exchange



blue = λ_{i-1} replica

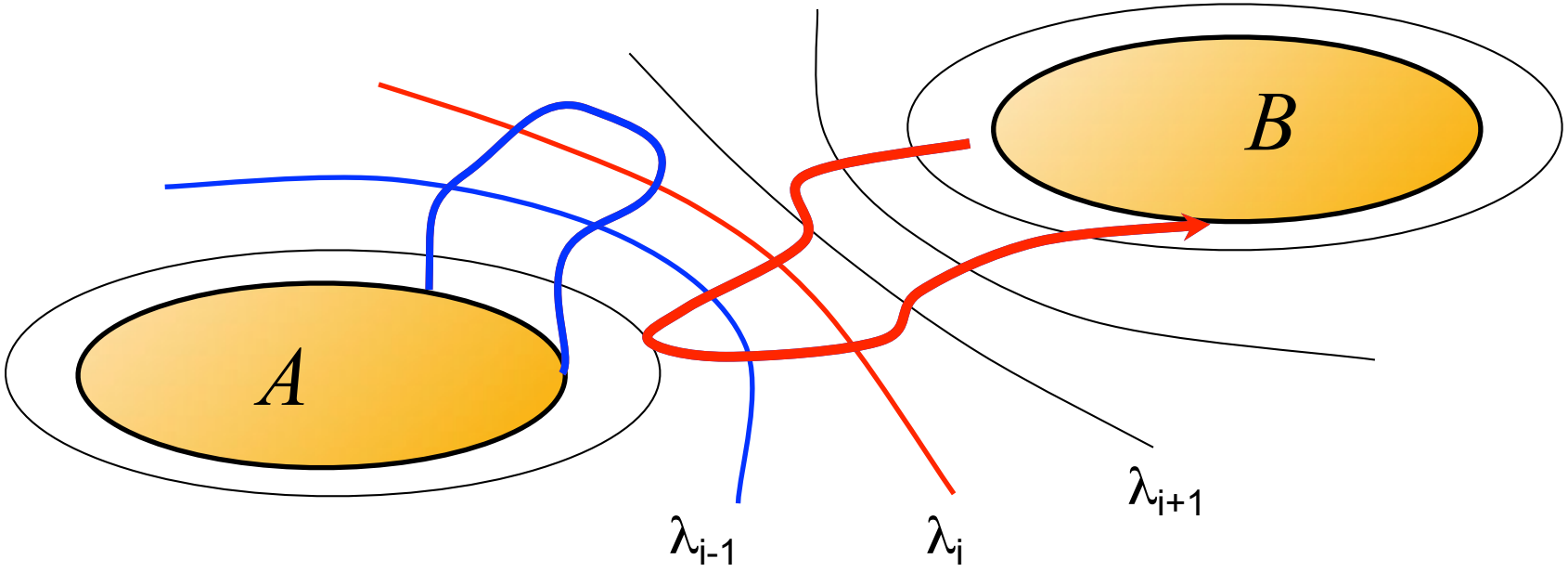
red = λ_i replica

$$P_{acc}(i \leftrightarrow j) = \min\left(1, \frac{g_{\lambda i}[\mathbf{x}^{(j)}(L^{(j)})]g_{\lambda j}[\mathbf{x}^{(i)}(L^{(i)})]}{g_{\lambda i}[\mathbf{x}^{(i)}(L^{(i)})]g_{\lambda j}[\mathbf{x}^{(j)}(L^{(j)})]}\right)$$

$$g_\lambda[\mathbf{x}(L)] = \begin{cases} 1 & \text{if path crosses } \lambda \\ 0 & \text{otherwise} \end{cases}$$

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

Include paths starting in B

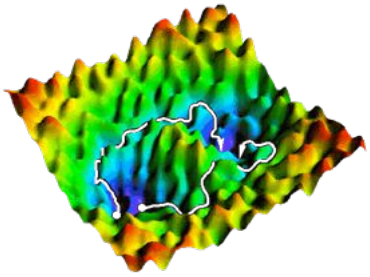


$$P_{acc}(i \leftrightarrow j) = \min\left(1, \frac{g_{\lambda i}[\mathbf{x}^{(j)}(L^{(j)})]g_{\lambda j}[\mathbf{x}^{(i)}(L^{(i)})]}{g_{\lambda i}[\mathbf{x}^{(i)}(L^{(i)})]g_{\lambda j}[\mathbf{x}^{(j)}(L^{(j)})]}\right)$$

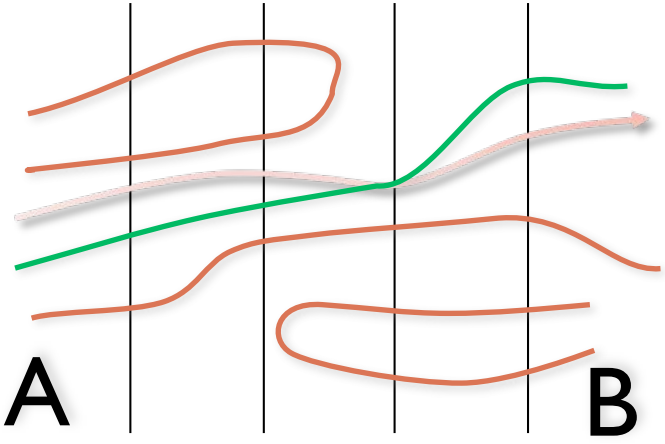
$$g_\lambda[\mathbf{x}(L)] = \begin{cases} 1 & \text{if path crosses } \lambda \\ 0 & \text{otherwise} \end{cases}$$

Samples AA, AB, BA and BB paths
- -shooting move
- -time reversal move
- -exchanges

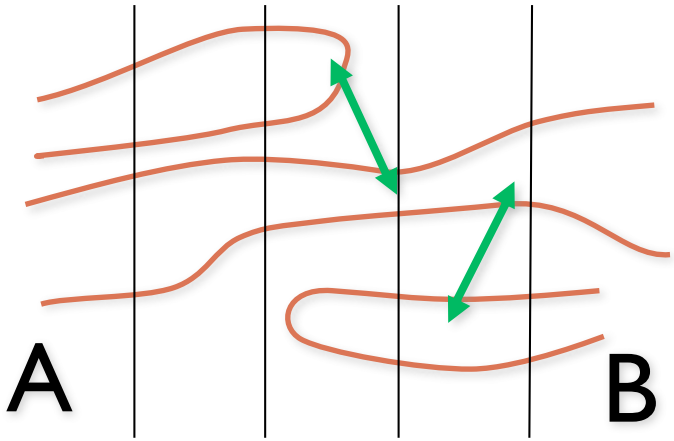
Replica Exchange Transition Interface Sampling



Shooting Move

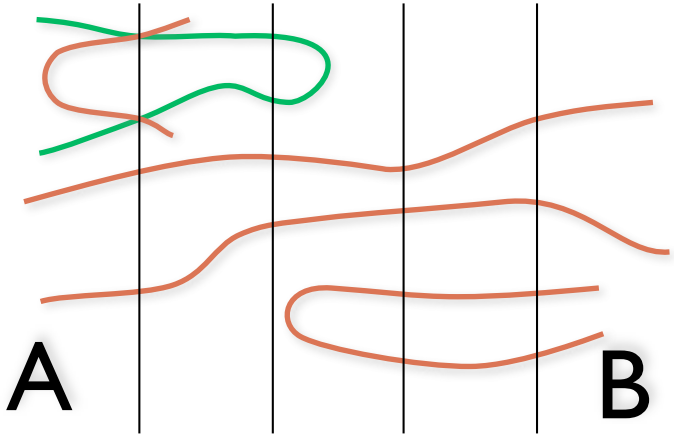


Exchange Move

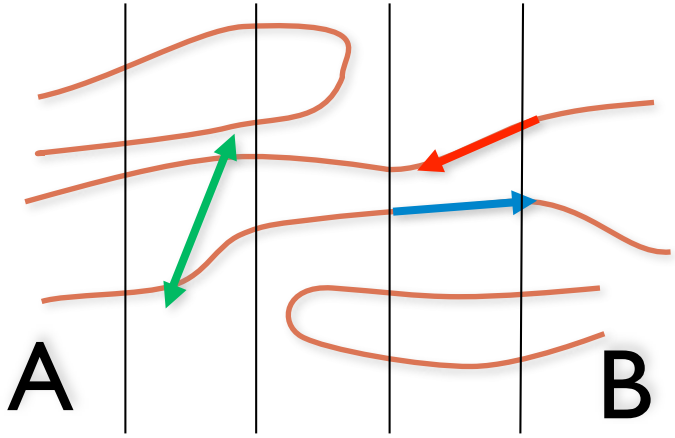


T.S. van Erp, PRL **98**, 268301 (2007)
P.G. Bolhuis, JCP **129**,114108 (2008)

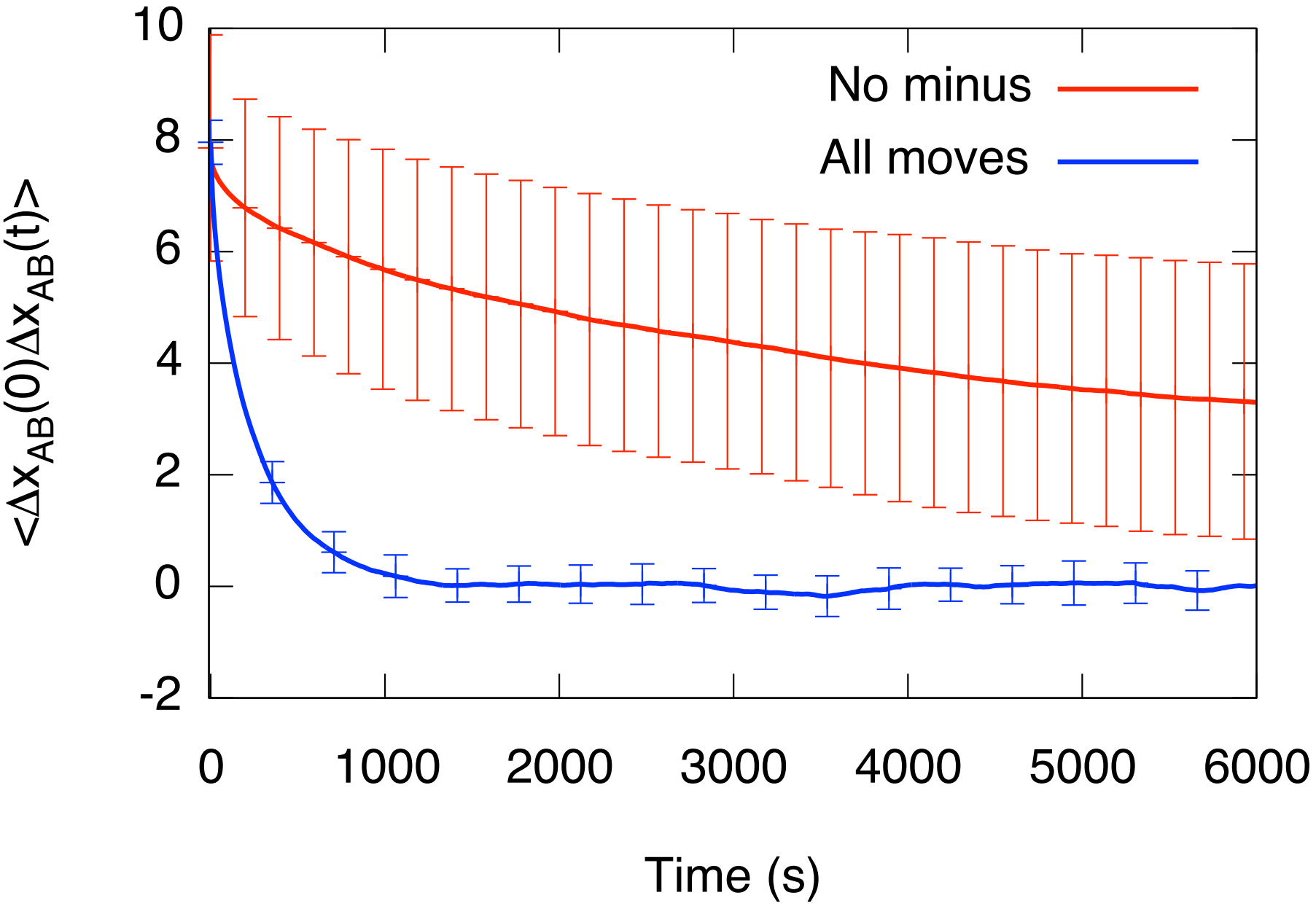
First Interface Move



Reversal Move

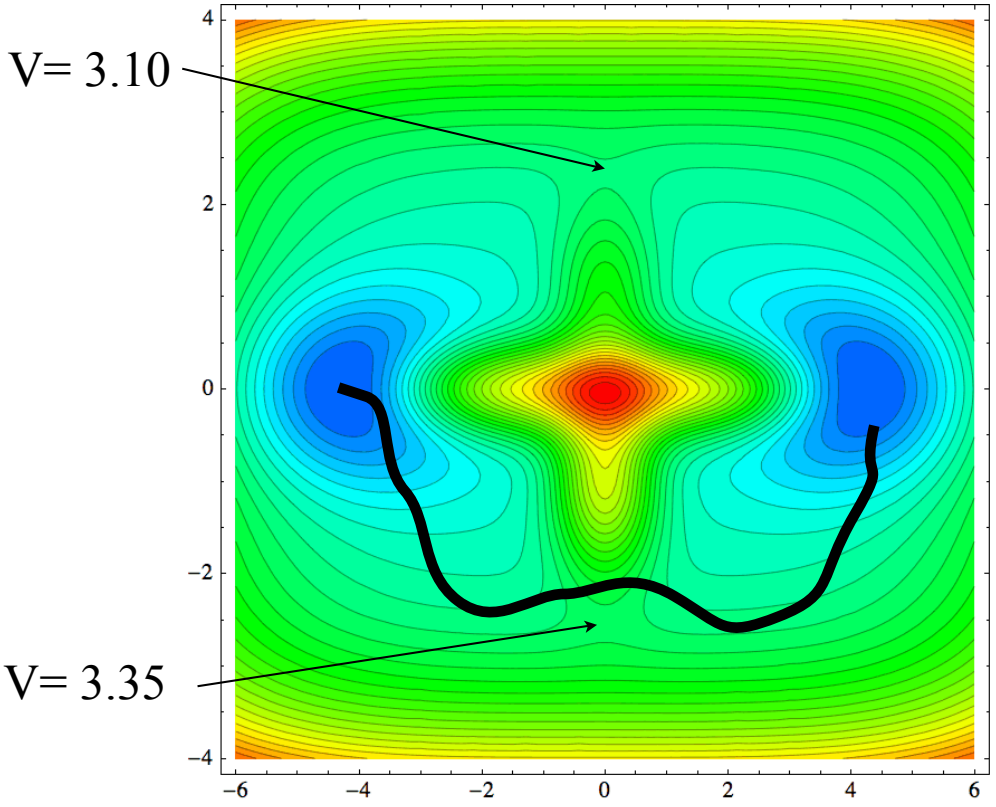


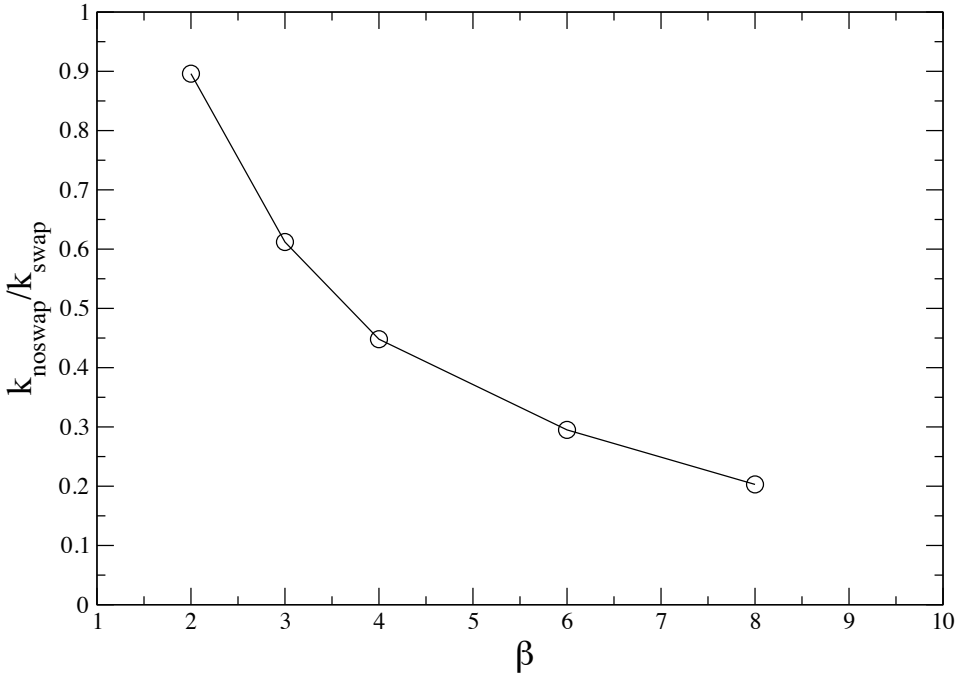
Center of Mass Autocorrelation



Necessity of exchange

- illustrated on two channels with different barrier height.

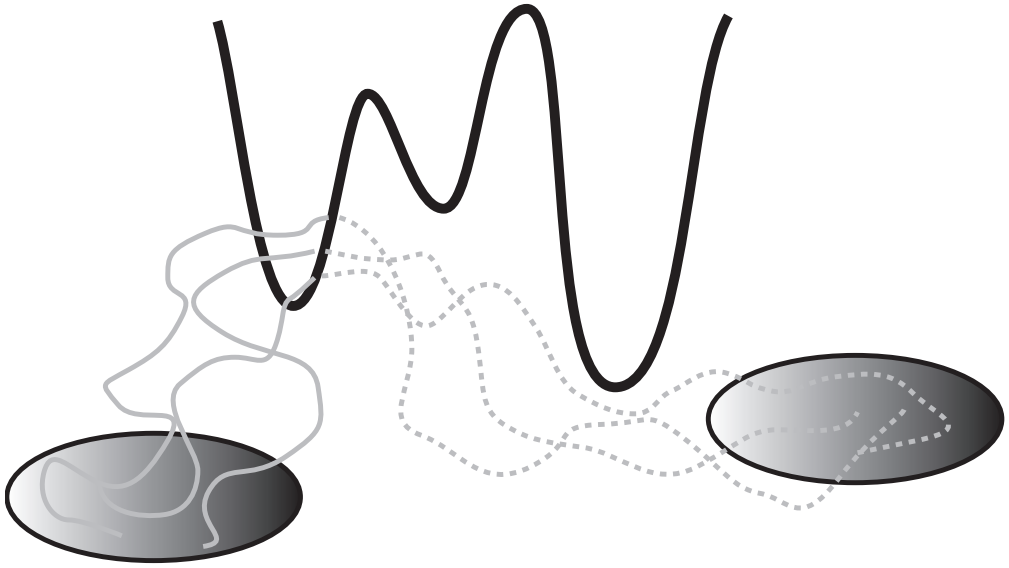




Challenges for path sampling

Multiple channels

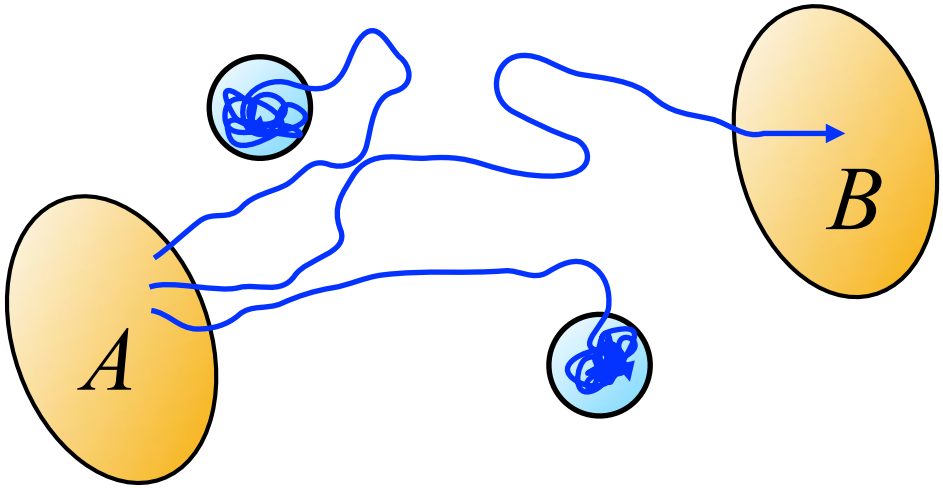
- multiple channels are not sampled properly with shooting
- Replica exchange TIS



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

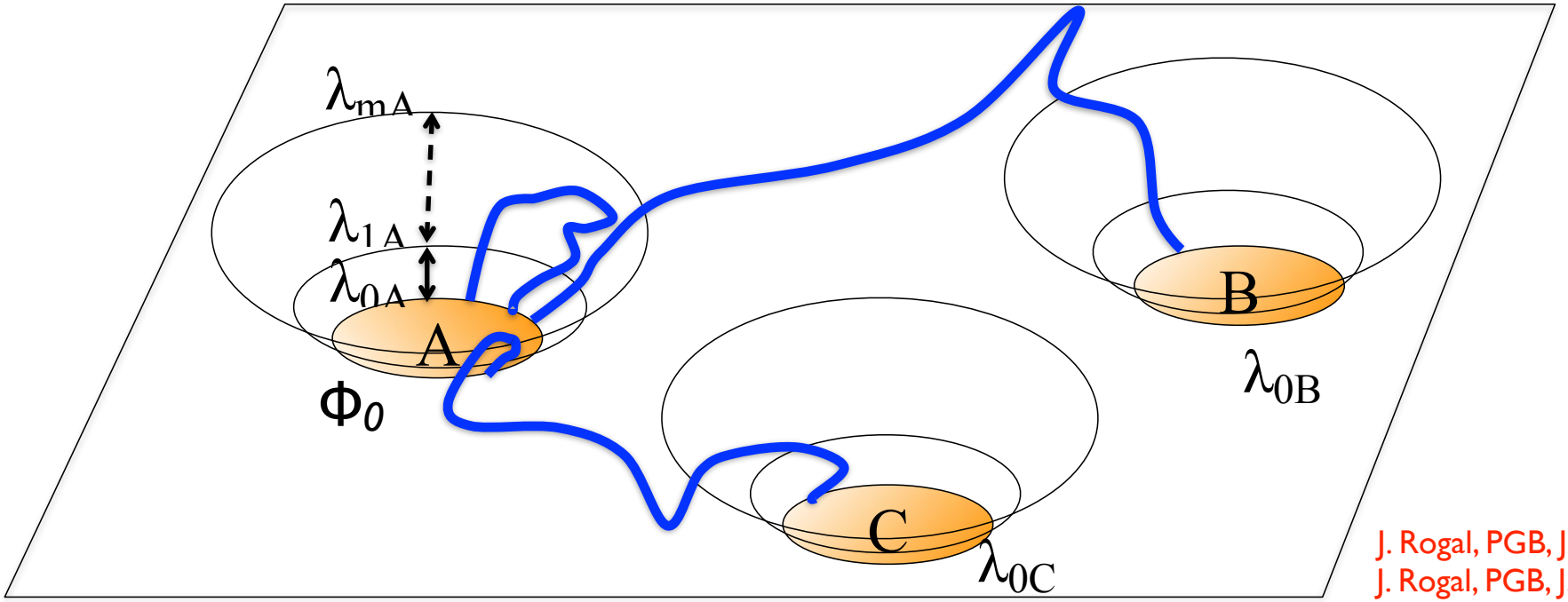
Presence of intermediates

- paths become very long because of intermediates
- Multiple state TIS



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

Multiple state transition interface sampling



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

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

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

TIS:

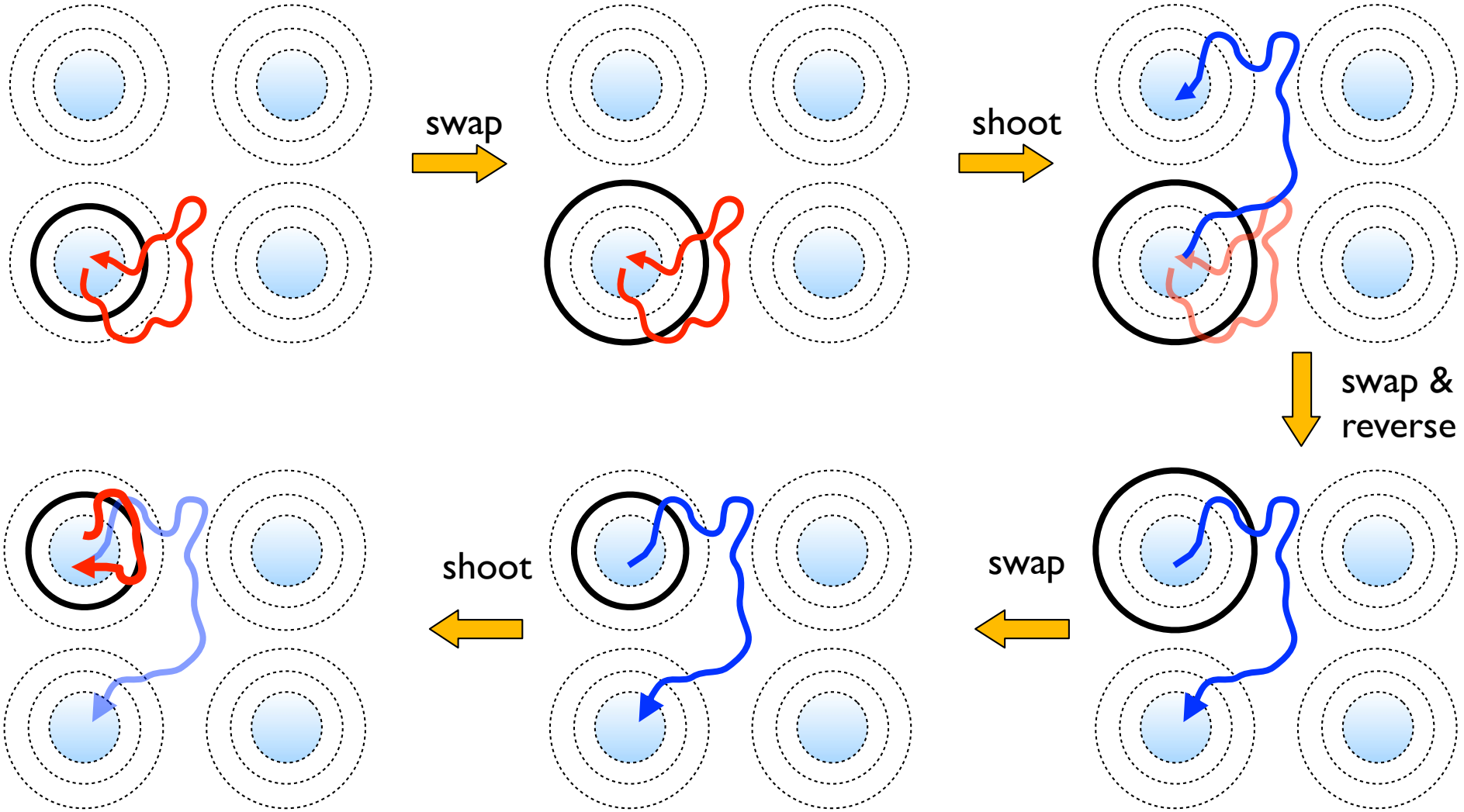
$$\frac{\langle \phi_A \rangle}{\langle h_A \rangle} \prod_{s=0}^{m-1} P_A(\lambda_{(s+1)_A}|\lambda_{s_A})$$

MSTIS:

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

rates can be used in Markov state model

Single replica MSTIS



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

Single replica Wang-Landau path sampling

- a single replica walks along set of interfaces
- each interface of i for state **J** has a *density of paths* g_i and histogram h_i
- each time a interface is sampled $g_i = g_i * f$ and $h_i = h_i + 1$
- a swap between interfaces is accepted with:

$$P_{acc}^{wl}[\lambda_i \leftrightarrow \lambda_{i+1}] = \min\left[1, \frac{g_i}{g_{i+1}}\right]$$

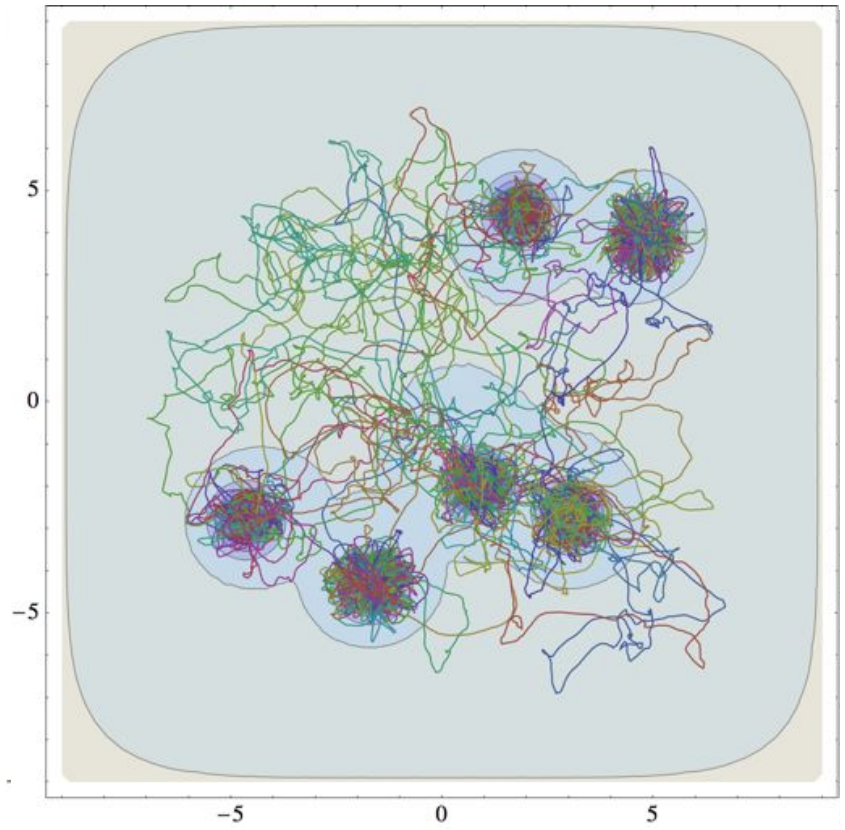
- when path switches from **J→J** to **J→K**, allow switch to new set of interfaces for **K**
- if histogram is "flat" then
  - reset histogram $h_i = 0$
  - reduce factor $f = \sqrt{f}$, continue
- weights g_i reflect ratio of pathways on each interface = crossing probability

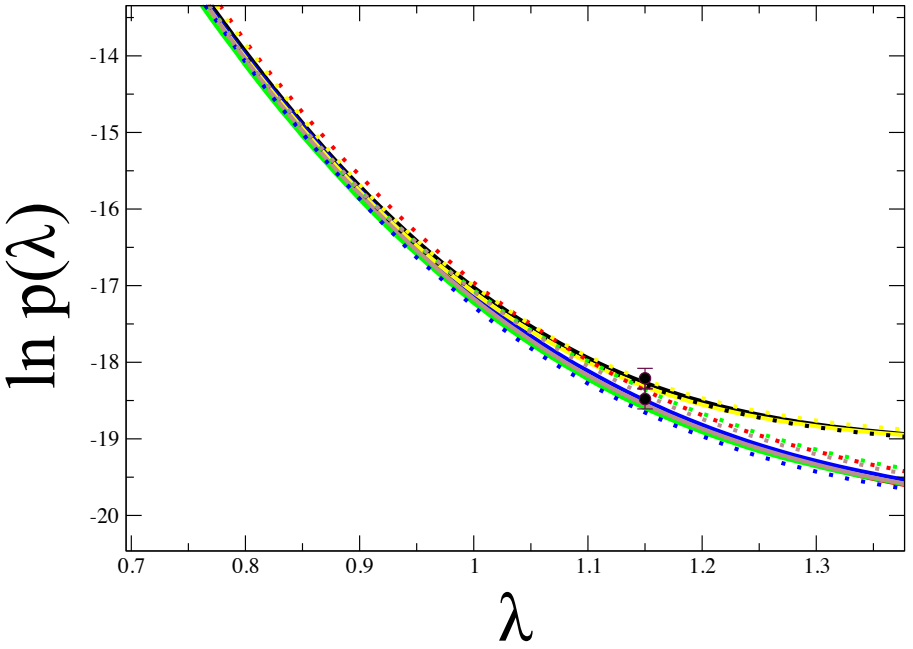
advantage:
only one replica needed

disadvantage:
need to wait until histogram is flat
only correct in limit of f→0

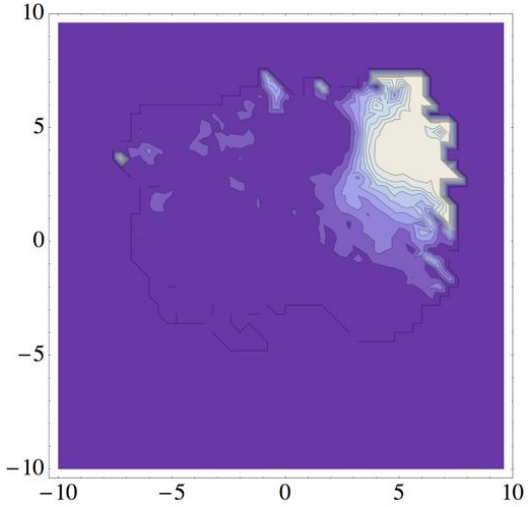
F. Wang and D.P. Landau, PRL 86, 2050 (2001)
W. Du and PGB, JPC, 139, 044105, (2013)

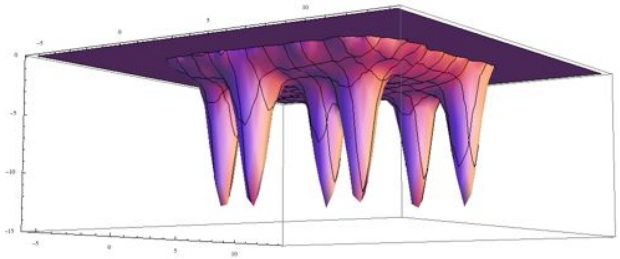
Langevin dynamics on 2D potential





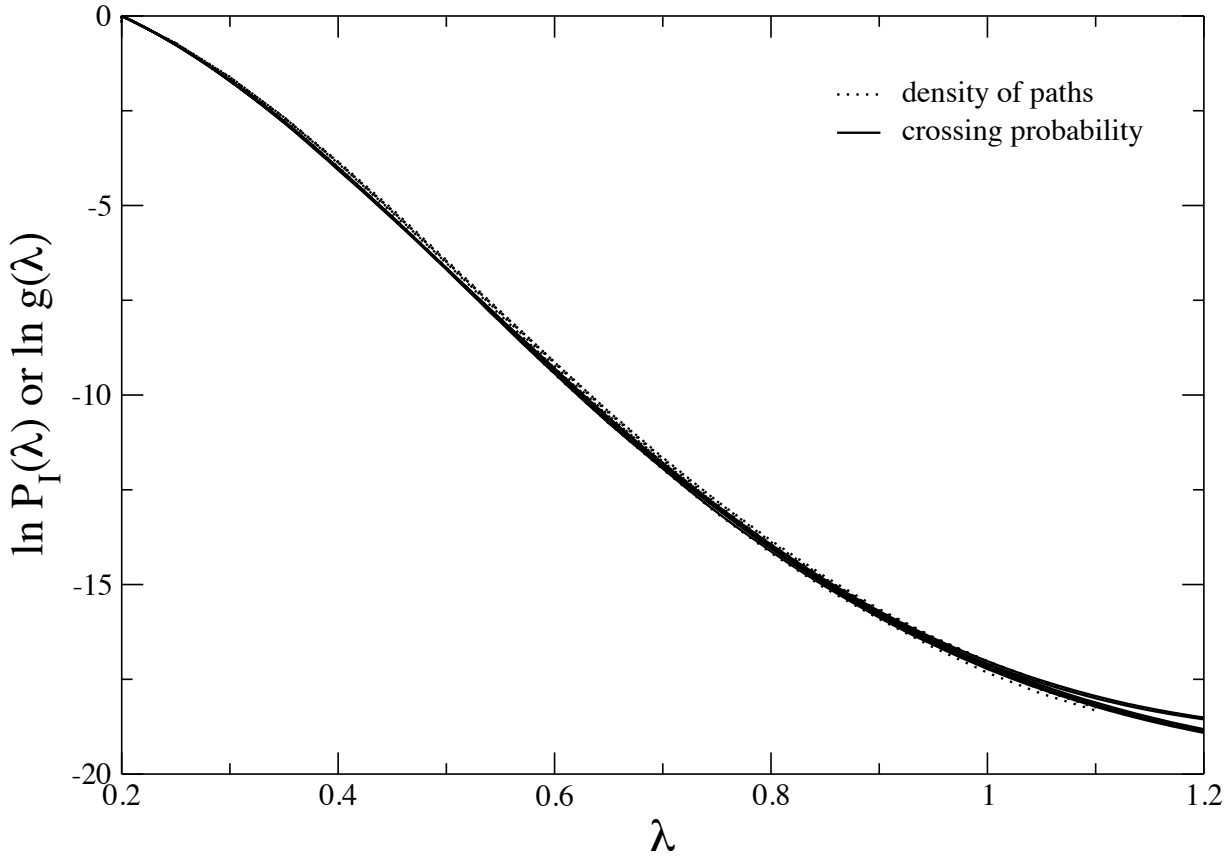
- β=8, γ=2.5 , Δt =0.05
- number of interfaces 20
- metric: distance to center of state
- shooting, time reversals and swaps
- reweighting gives FE and committor





Improve on Wang-Landau

- WL converges slowly
- improve by imposing fixed bias
- DOP turns out to converge to crossing probability $P_l(\lambda)$
  - because $P_l(\lambda_i)/P_l(\lambda_j)$ is probability to reach λ_j from λ_i

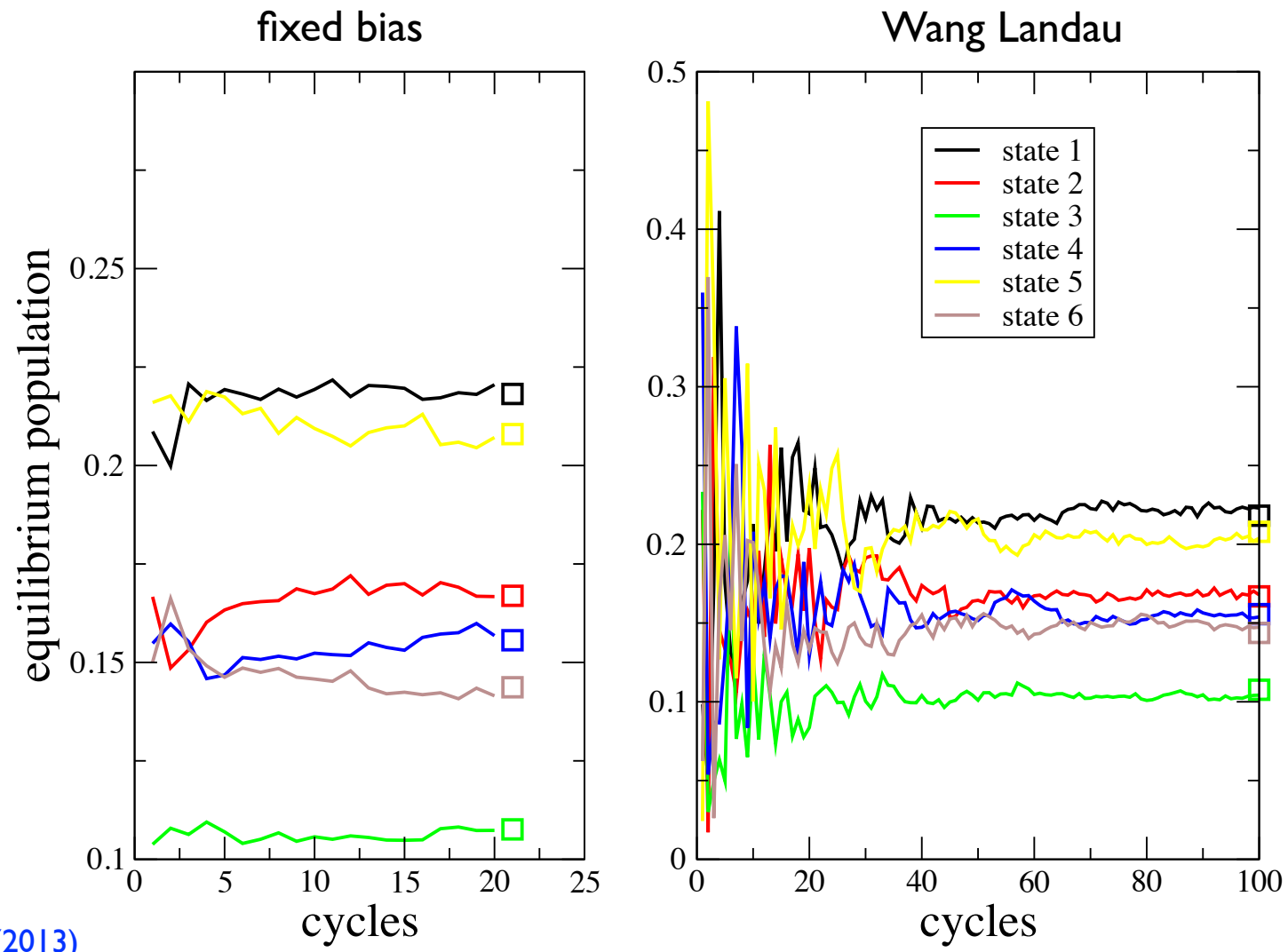


- perfect bias for flat histogram is crossing probability $P_l(\lambda)$

Convergence fixed bias vs WL

- use equilibrium population to assess convergence

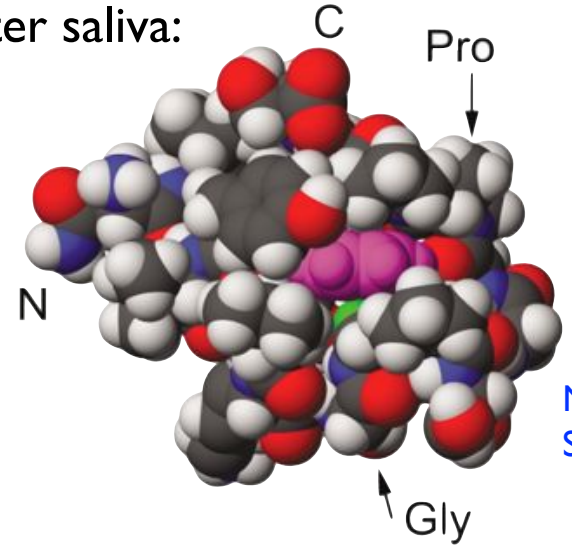
$$\mathbf{p}_{eq} = \lim_{t\to\infty} \mathbf{p}^T(t) = \mathbf{p}^T(0) \lim_{t\to\infty} \exp(\mathbf{K}t)$$



W. Du and PGB,
JPC, 139, 044105, (2013)

Trp-cage folding

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



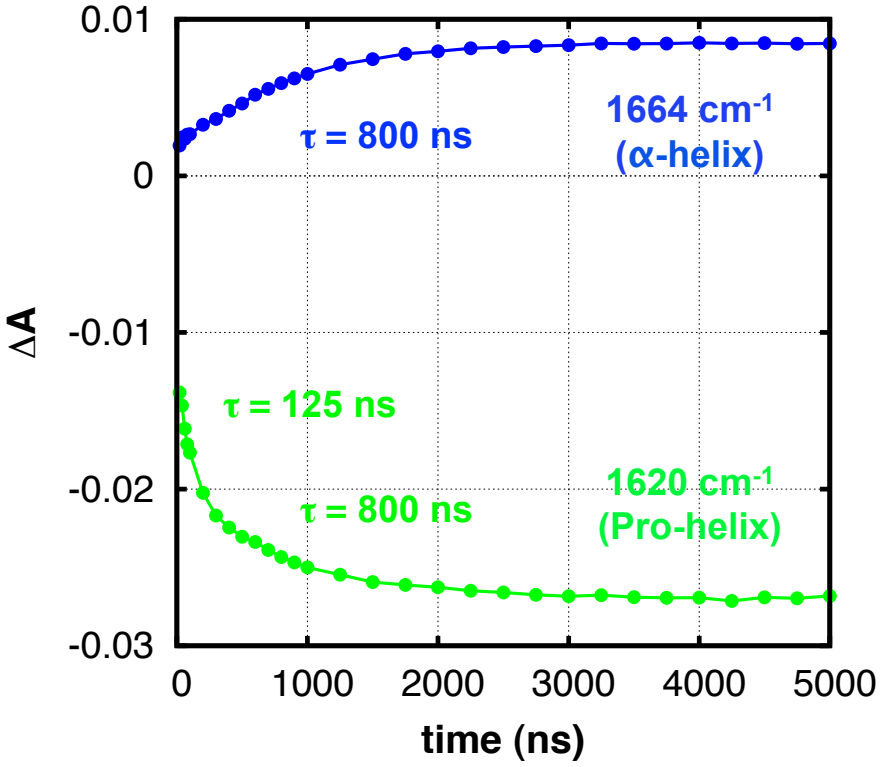
Neidigh & al., Nature Struct.Biol. **9**, 425 (2002)

Salt-bridge

NAYAQ WLKDG GPSSG RPPPS

Tyrosine Tryptophan Glycine Proline

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

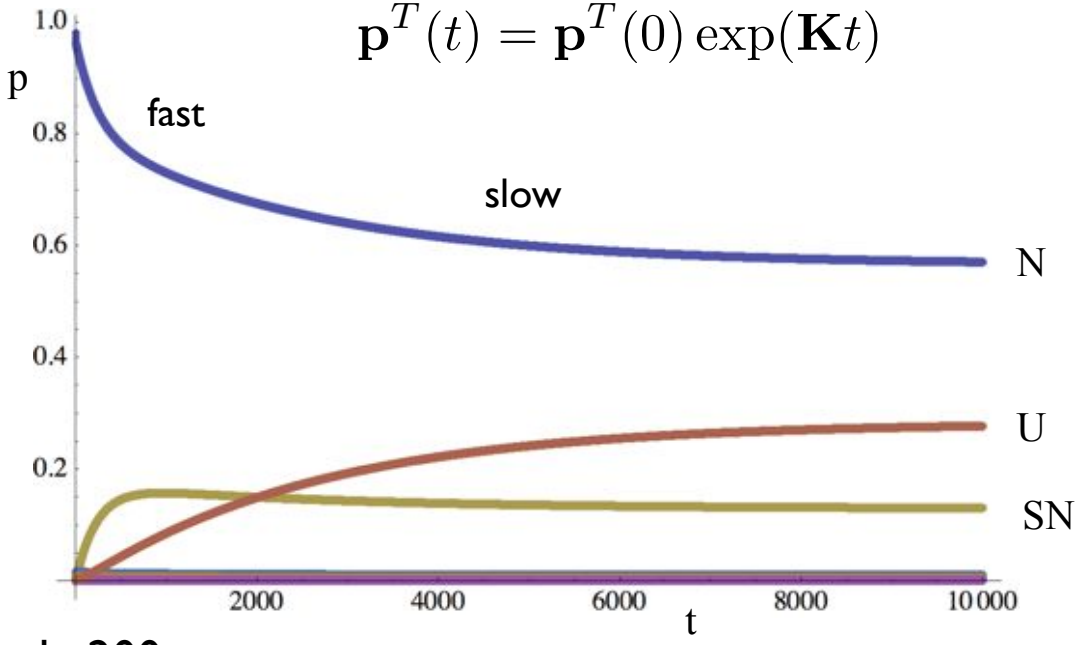


H. Meuzelaar, K. A. Marino, A. Huerta-Viga, M. R. Panman, L.E. J. Smeenk, A.J. Kettelarij, J.H. van Maarseveen, P. Timmerman, PGB, and S. Woutersen
JPCB 2013.

Kinetics from rate matrix

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

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

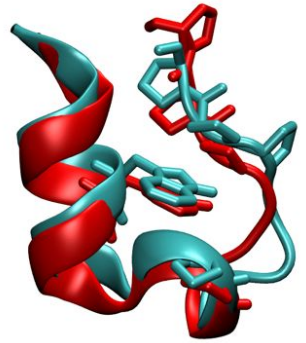


fast time scale 200 ns

slow time scale 2 μs

Experimental t_1=150 ns, t_2= 2.2 μs

SN state



Transition path theory analysis

Analysis of large kinetic matrix for more insight

E, Vanden-Eijnden, J. Stat. Phys 2006
Noe et al., PNAS 2009

compute forward committor q_i for each state i

$$q_i^+ = \sum_{k \in \mathbf{I}} T_{ik} q_k^+ + T_{iU} \qquad \mathbf{T} = \exp(\mathbf{K}\tau)$$

Compute flux from i to j $\quad f_{ij} = \pi_i q_i^- T_{ij} q_j^+$

compute effective flux $\quad f_{ij}^+ = \max\{0, f_{ij} - f_{ji}\}$

overall rate constant $\quad k_{NU} = \dfrac{\sum_{i \neq N} \pi_N T_{Ni} q_i^+}{\tau \sum_{i \in \mathbf{M}} \pi_i q_i^-}$

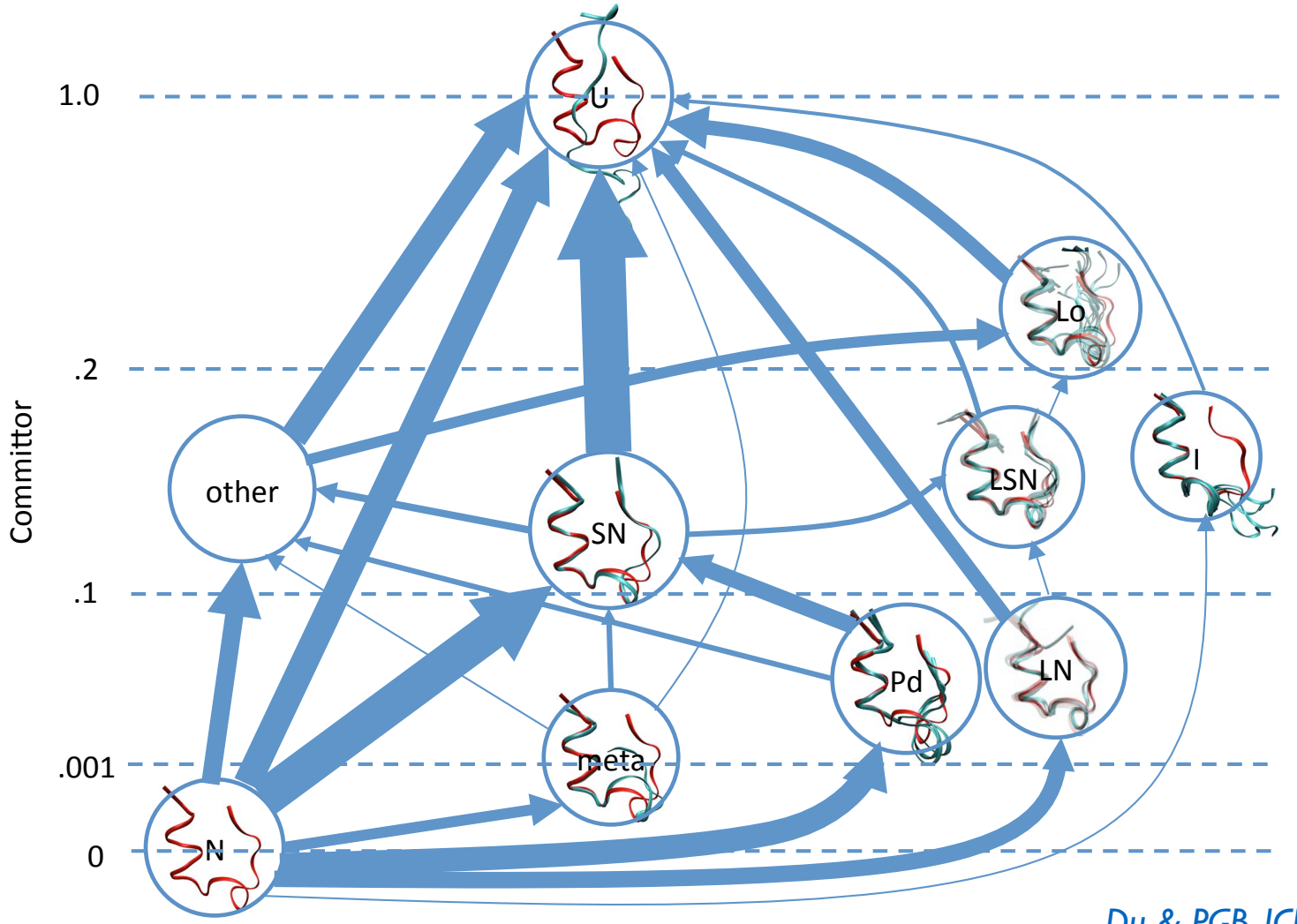
k_{NU} is $1.01 \times 10^{-4} ns^{-1} \approx (9.9 \mu s)^{-1}$,

$k_{UN} = 4.17 \times 10^{-4} ns^{-1} \approx (2.4 \mu s)^{-1}$

$k_{NU}exp = (12 \mu s)^{-1}$

$k_{UN}exp = \approx (4.1 \mu s)^{-1}$

TPT flux analysis

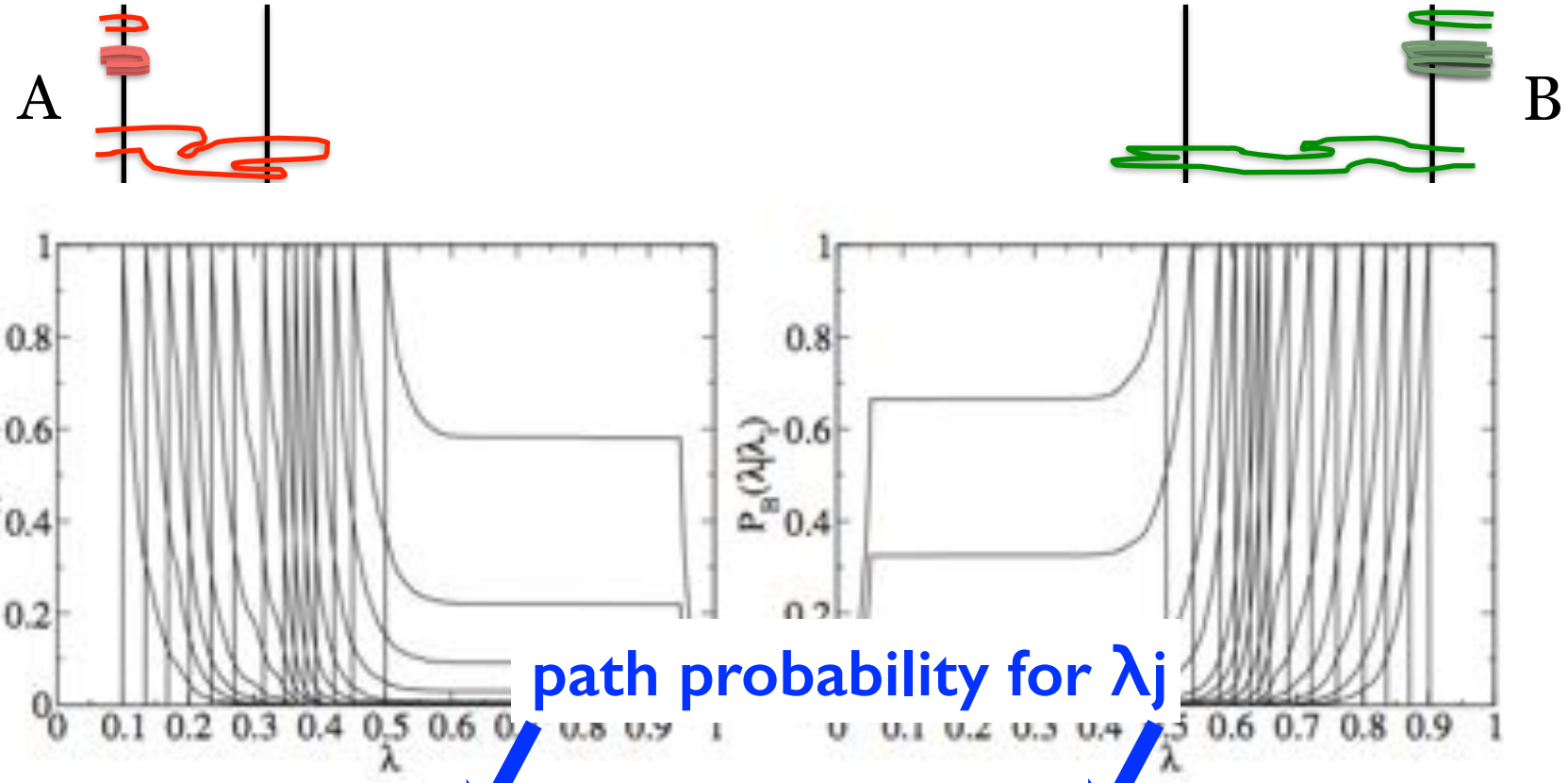


*Du & PGB, JCP **140**, 195102 (2014)).*

158 μs MD, around 70000 trajectories, represents around 15 ms of time

rate matrix and flux are non sparse, many pathways possible

The Reweighted Path Ensemble



path probability for λj

$$\mathcal{P}[\mathbf{x}^L] = c_A \sum_{j=1}^{n-1} \mathcal{P}_{A\Lambda_j}[\mathbf{x}^L] W^A[\mathbf{x}^L] + c_B \sum_{j=1}^{n-1} \mathcal{P}_{B\Lambda_j}[\mathbf{x}^L] W^B[\mathbf{x}^L]$$

$$W^A[\mathbf{x}^L] = \sum_{i=1} \bar{w}_i^A \theta(\lambda_{max}[\mathbf{x}^L] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{max}[\mathbf{x}^L])$$

WHAM weights

Rogal et al , J. Chem. Phys. 133, 174109 (2010)
Lechner et al .J. Chem. Phys. 133 174110 (2010).

Projection of Reweighted Path Ensemble

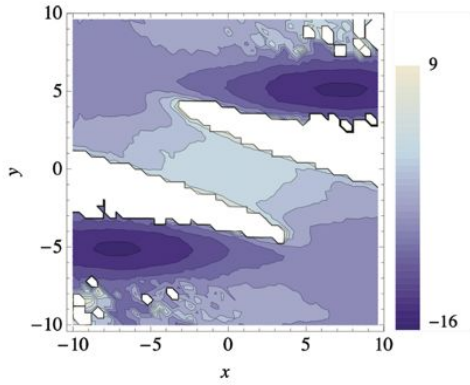
RPE can be used to project the conditional path dependent population density

$$\rho_{ij}(\mathbf{q}) = \langle h_i(\mathbf{x}_0) h_j(\mathbf{x}_L) \delta(\mathbf{q}(\mathbf{x_k}) - \mathbf{q}) \rangle_{RPE}$$

$$\rho(\mathbf{q}) = \rho_{AA}(\mathbf{q}) + \rho_{AB}(\mathbf{q}) + \rho_{BA}(\mathbf{q}) + \rho_{BB}(\mathbf{q}) = \langle \delta(\mathbf{q}(\mathbf{x_k}) - \mathbf{q}) \rangle_{RPE}$$

and thus the free energy landscape

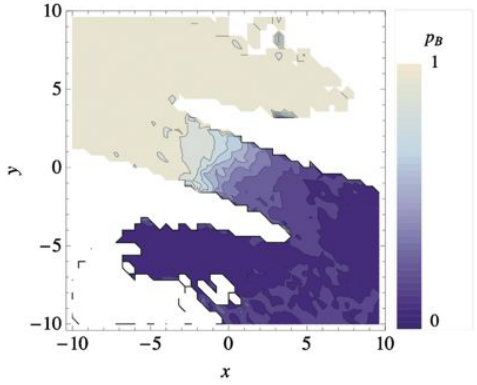
$$F(\mathbf{q}) = -k_B T \ln(\rho(\mathbf{q})) + const,$$



and the (averaged) committor

$$p_A(\mathbf{q}) = \frac{\rho_{AA}(\mathbf{q}) + \rho_{BA}(\mathbf{q})}{\rho(\mathbf{q})}$$

$$p_B(\mathbf{q}) = \frac{\rho_{AB}(\mathbf{q}) + \rho_{BB}(\mathbf{q})}{\rho(\mathbf{q})}.$$

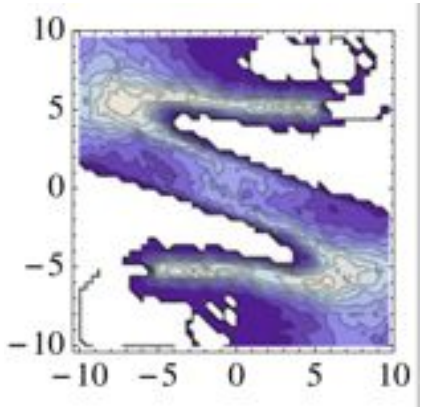


W Lechner, PGB,
J. Stat. Phys. 145 841 (2011).

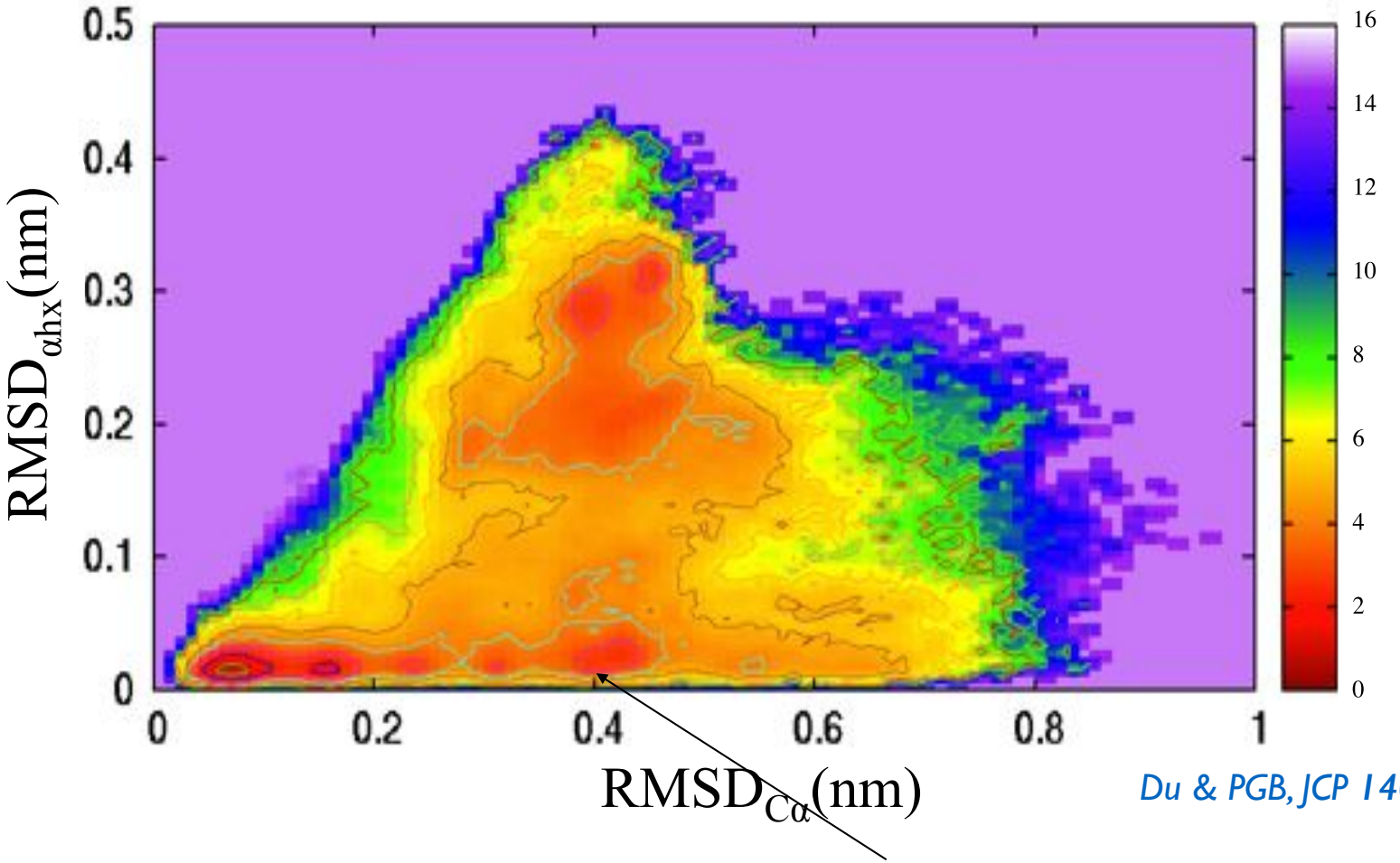
and the path density

$$n_{ij}(\mathbf{q}) = \langle h_i(\mathbf{x}_0) h_j(\mathbf{x}_L) h_{\mathbf{q}}(\mathbf{x}^L) \rangle_{RPE}$$

$$h_{\mathbf{q}}(\mathbf{x}^L) = \begin{cases} 1 & \text{if path visits } \mathbf{q} \\ 0 & \text{otherwise} \end{cases}$$

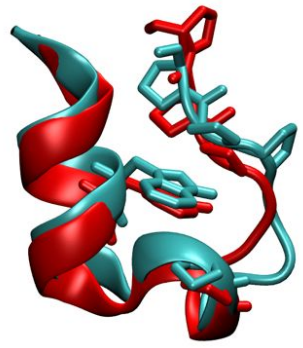


RPE Free energy for Trp-cage



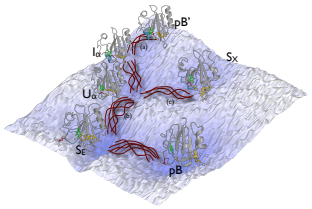
*Du & PGB, JCP **140**, 195102 (2014)).*

I state

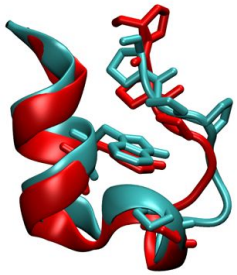


- The I state looks like a state, but it is not stable
- Free energy projection can be misleading

Summary



- Transition path sampling gives unbiased paths of protein conformational changes
  - millisecond light-induced unfolding mechanism in PYP
  - association/dissociation of protein dimers
  - reaction coordinate requires advanced data analysis of simulations



- Single replica MSTIS samples equilibrium network of Trp Cage
  - asynchronous sampling scheme (in contrast to parallel replica exchange)
  - corroborates experimental evidence for near native intermediate
  - structure and correct time scales predicted by single replica MSTIS

- **Path sampling gives quantitative insight in rough free energy landscapes, kinetics and mechanism simultaneously**