Biased sampling of unbiased dynamical trajectories

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

ACMM Amsterdam Center for Multiscale Modeling

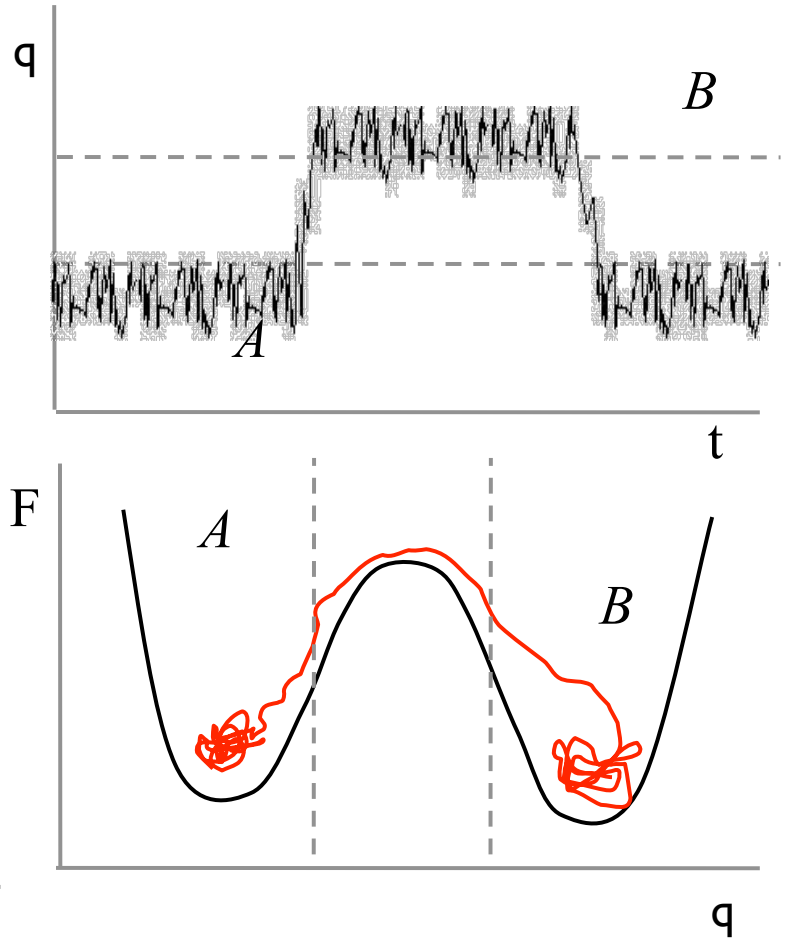
Outline

- introduction on transition path sampling
- example on signaling protein mechanism
- reaction coordinate analysis
- rate constants via transition interface sampling
- new algorithm: Wang Landau path sampling
- conclusions

Rare events

Interesting transitions in complex systems

- protein conformational change
- solution chemistry
- nucleation
- complex surface reaction



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

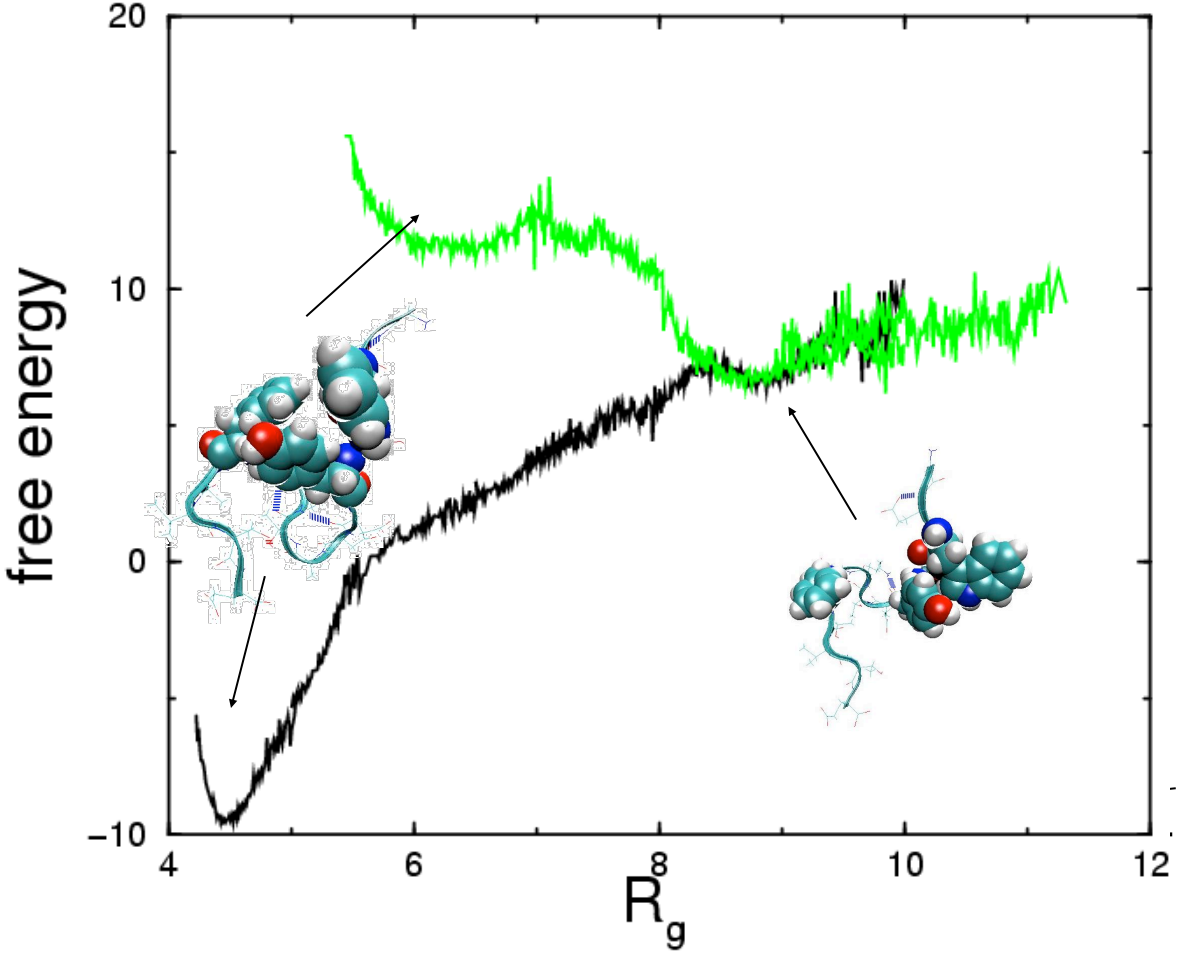
dominated by collective, rare events: straightforward MD is unpractical

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

Biased sampling of phase space

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

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



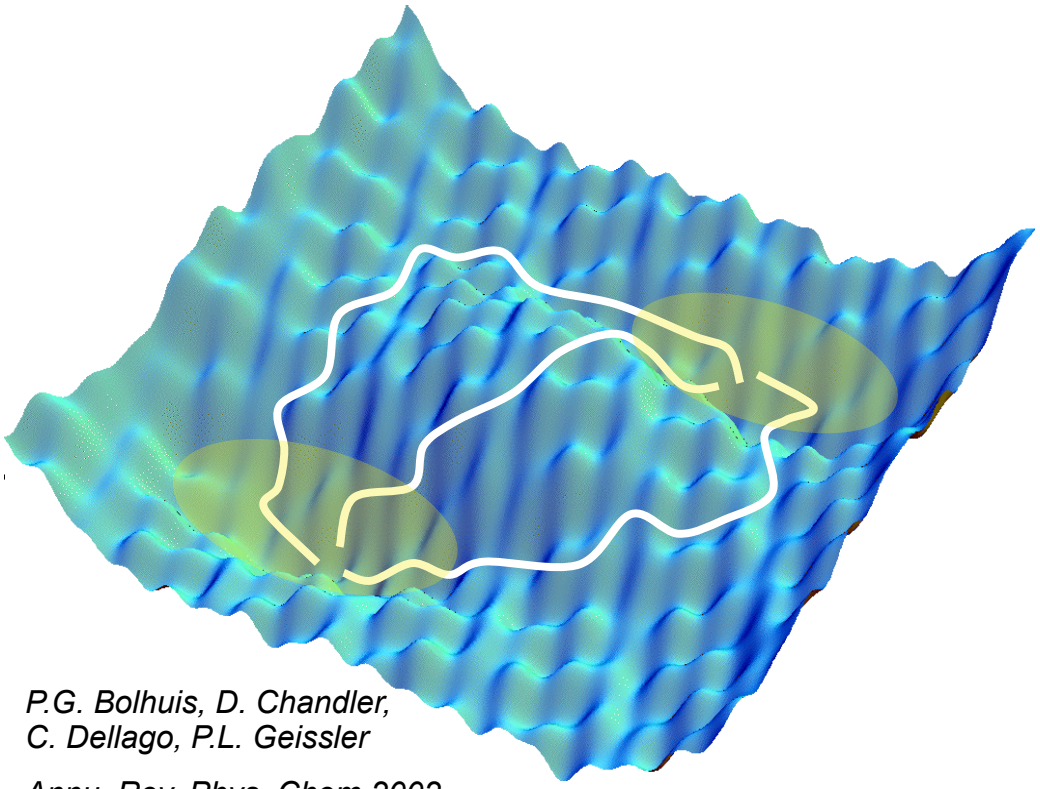
$$\beta W(q) = -\ln \int dq' e^{-\beta F(q,q')}$$

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

Transition path sampling

Transition path sampling

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



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

Annu. Rev. Phys. Chem 2002

Why TPS?

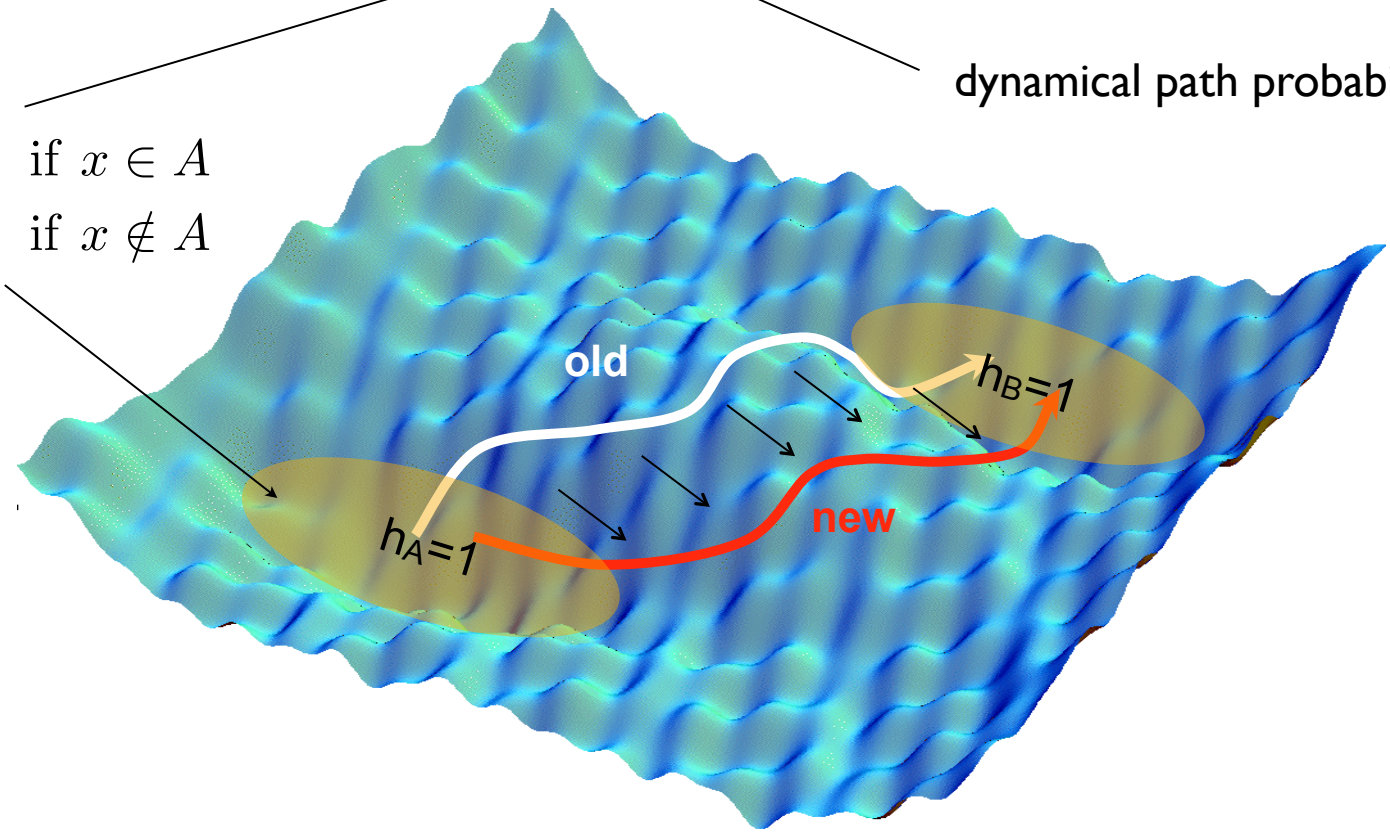
-selects unbiased rare paths
-no reaction coordinate needed
-RC follows from committor analysis
-rate constants

Importance sampling of paths

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

dynamical path probability

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

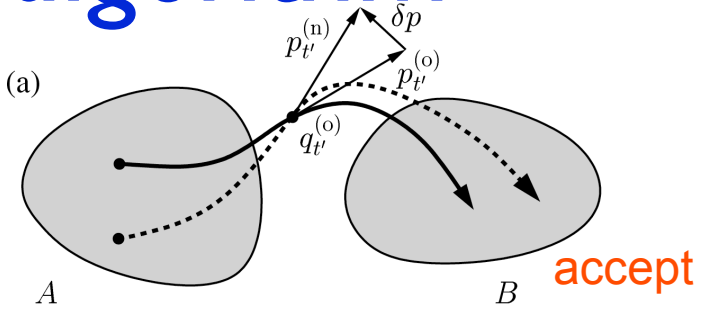


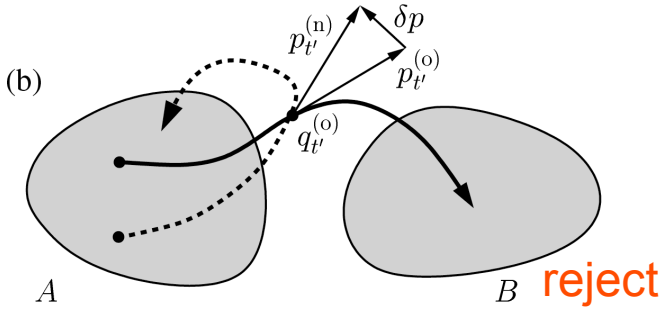
Metropolis rule:

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

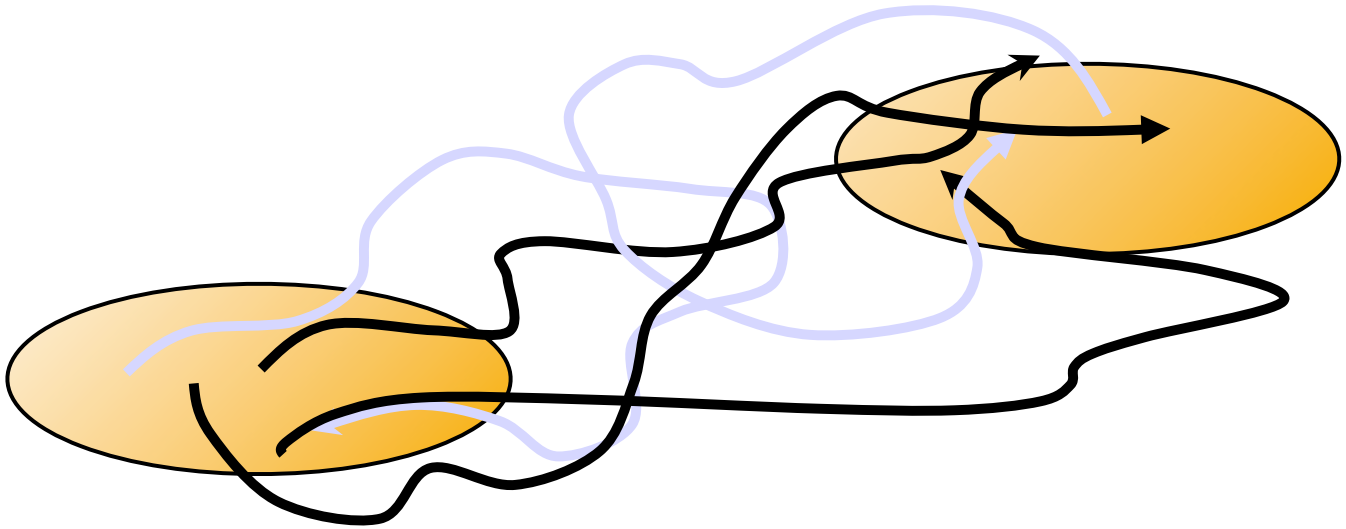
Standard shooting algorithm

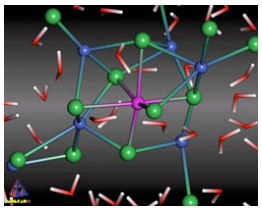
- take existing path
- choose random time slice *t*
- change momenta at *t*
- integrate forward and backward in time to create new path of length L (by MD)
- accept if A and B are connected, otherwise reject and retain old path
- calculate averages
- repeat

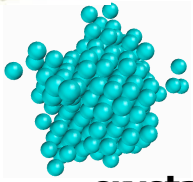




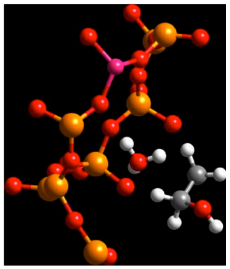
$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_A(x_0^{(n)}) h_B(x_T^{(n)})$$



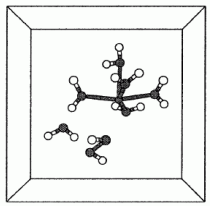


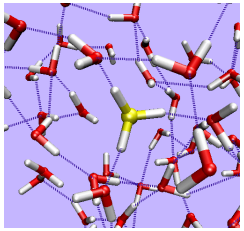


crystallisation

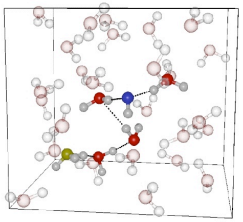


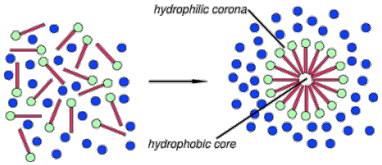
catalysis

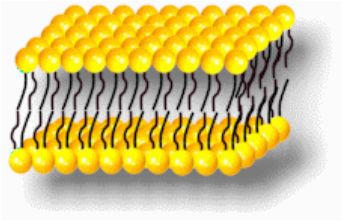




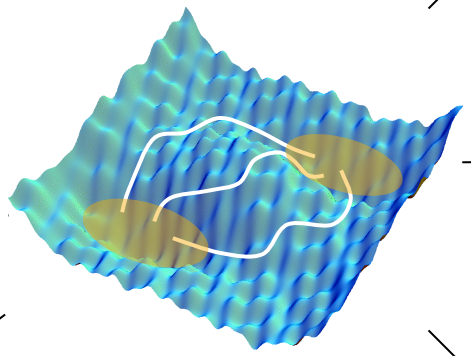
reactions

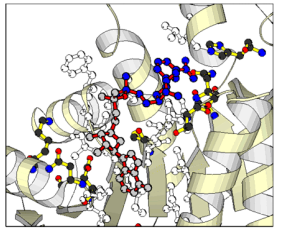




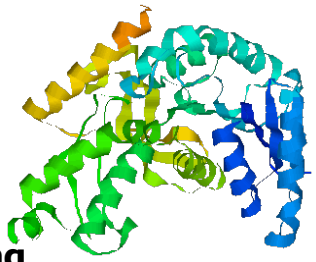


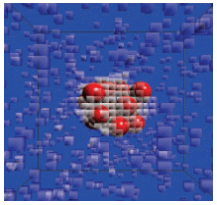
complex fluids

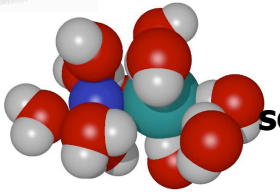




enzyme reactions

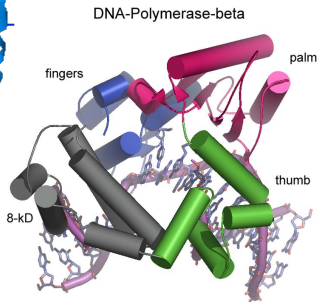


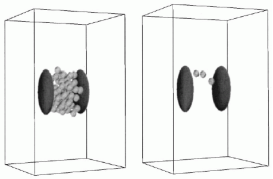




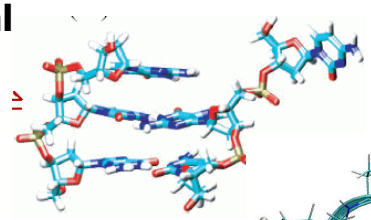
solvent effects

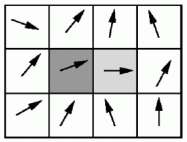
folding & binding

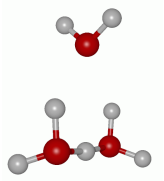


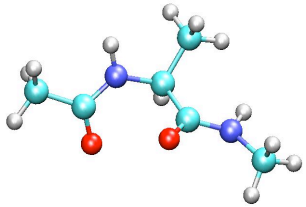


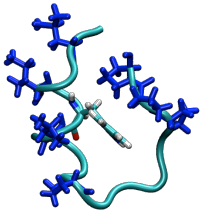
conformational change

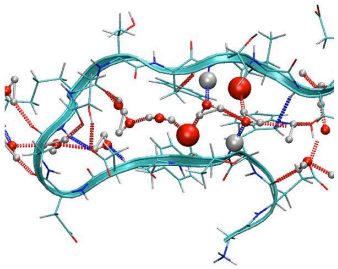




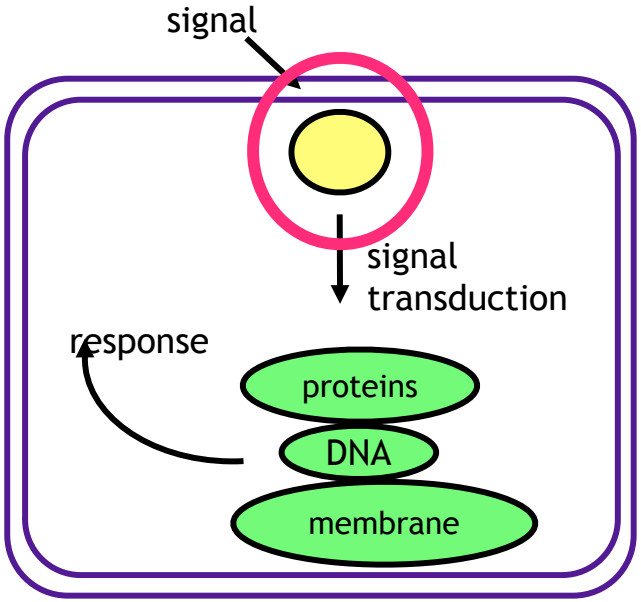


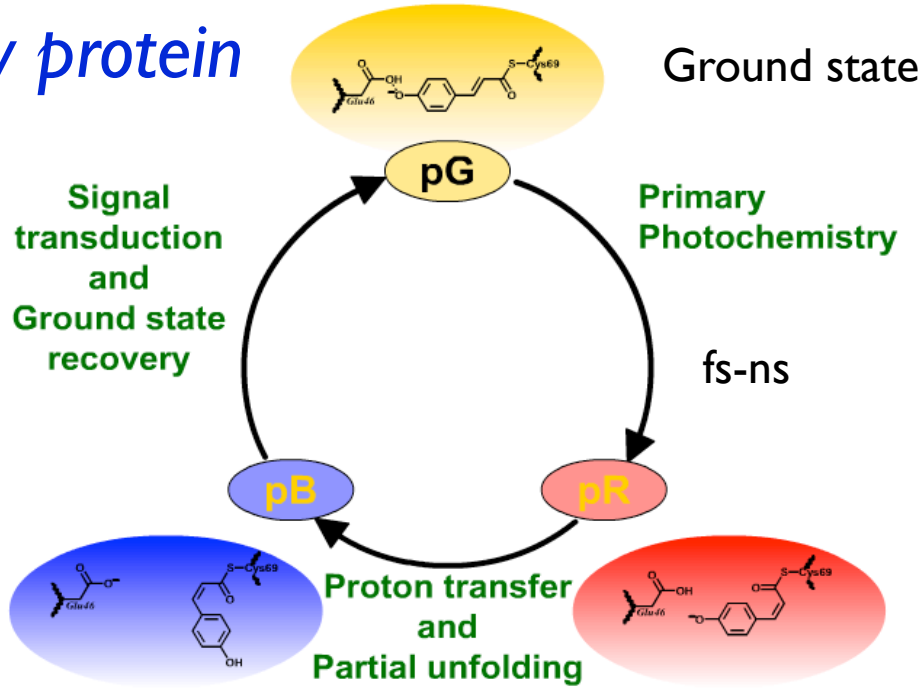


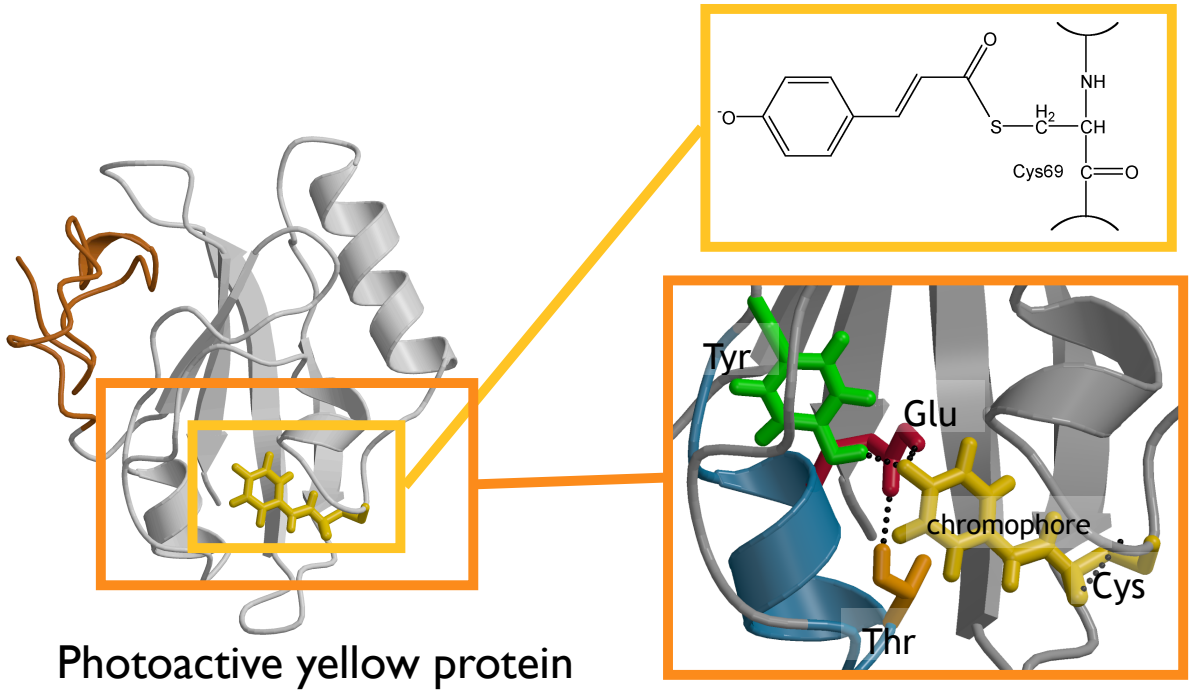




Signalling proteins: Photoactive yellow protein







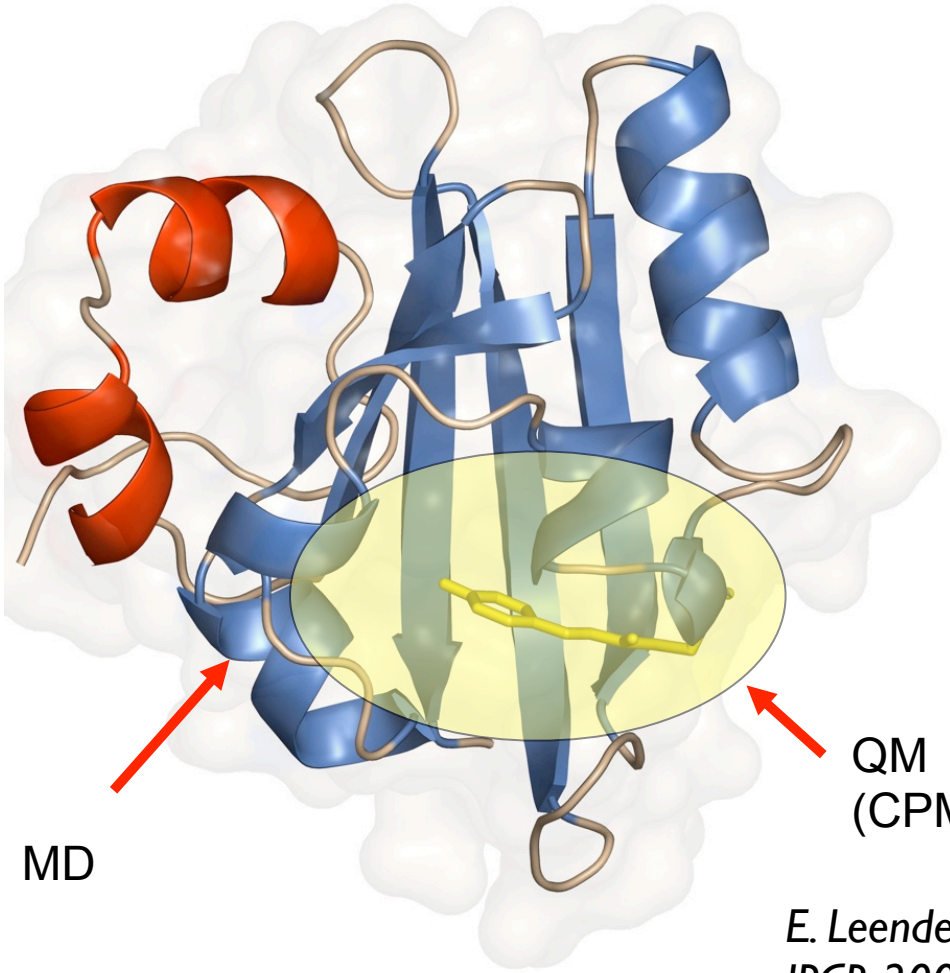
Photoactive yellow protein

Question: What is the mechanism for amplifying signal?

We studied 2 steps:

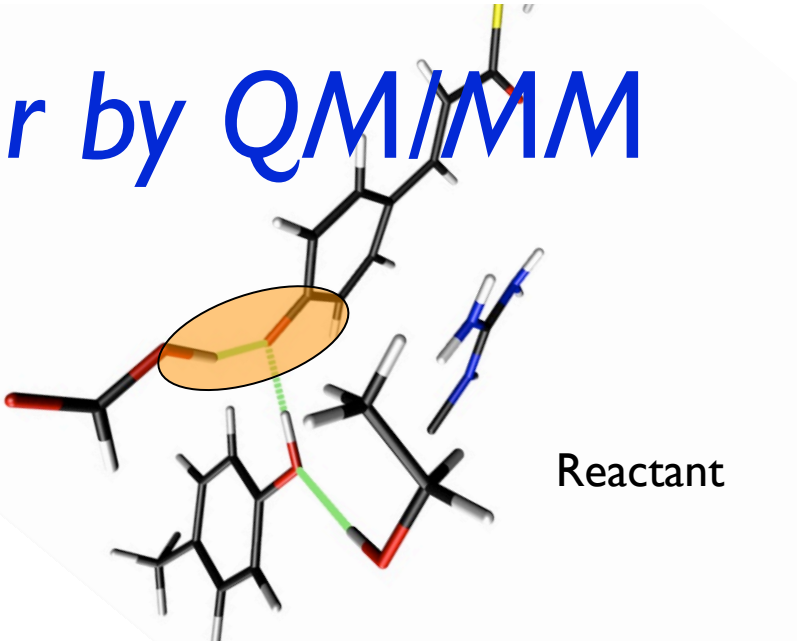
1) proton transfer
2) partial unfolding

Proton transfer by QM/MM



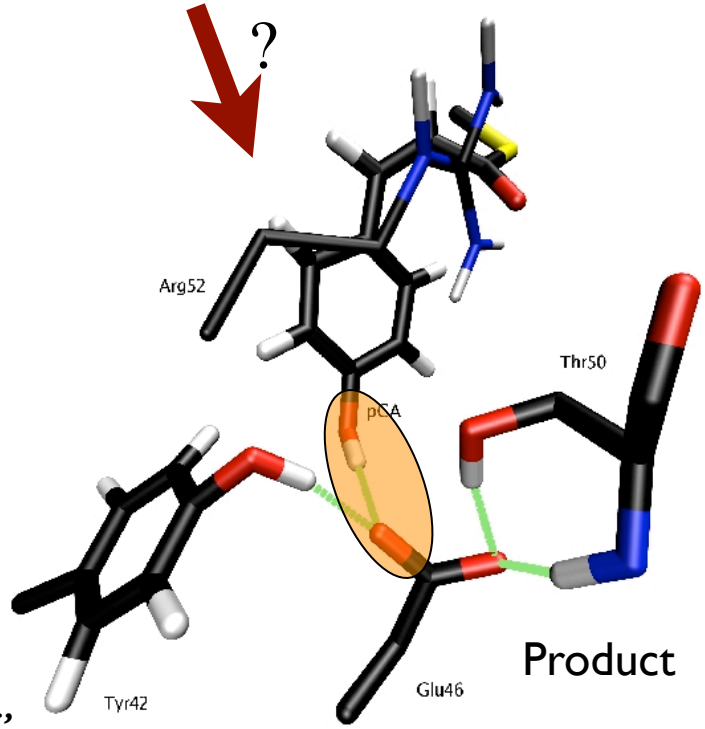
MD

QM
(CPMD)



Reactant

?

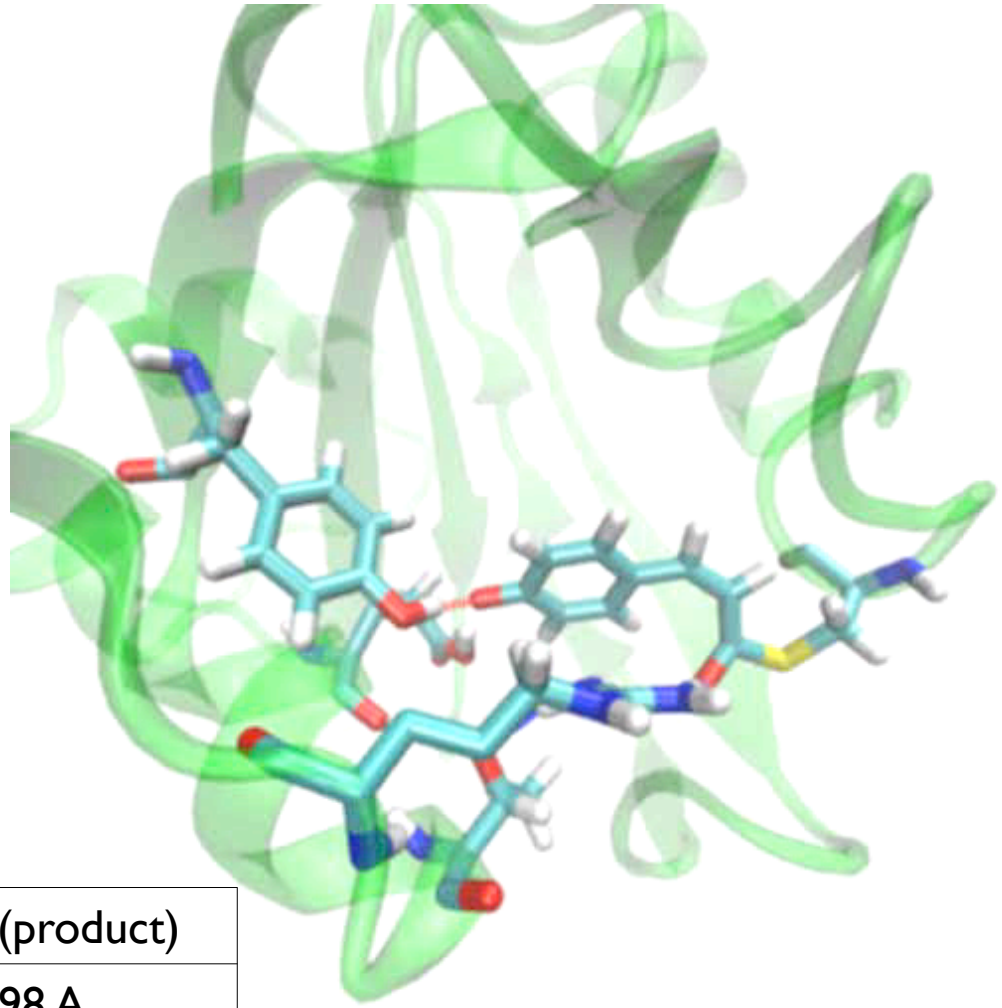


Product

E. Leenders et al., JPCB 2008

TPS of proton transfer

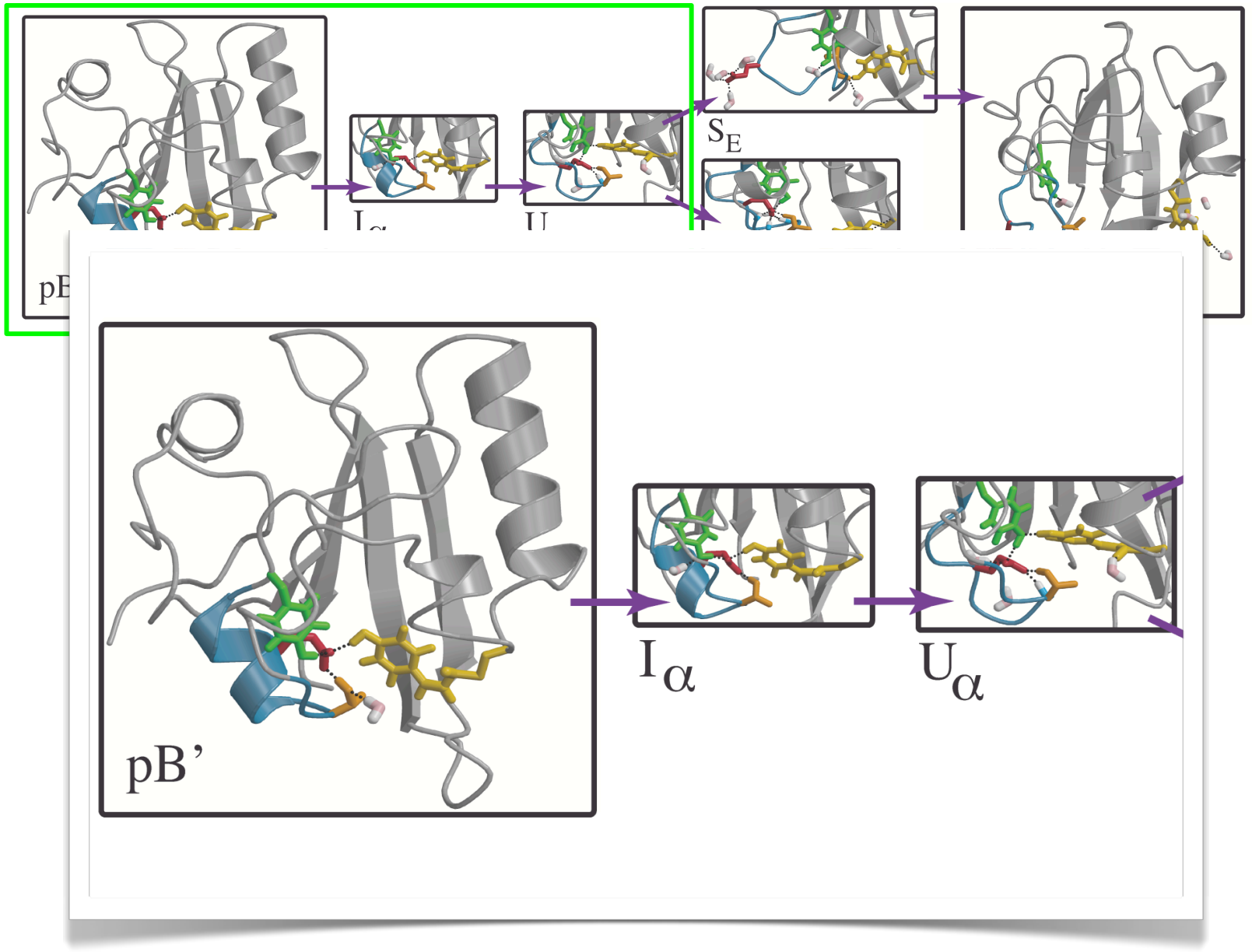
- 28244 atoms
- CPMD/QMMM
- BLYP functional
- Electronic mass 750 au
- QM region: pCA, Glu46, Tyr42, Thr50, Arg52
- Gromos96 force field



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

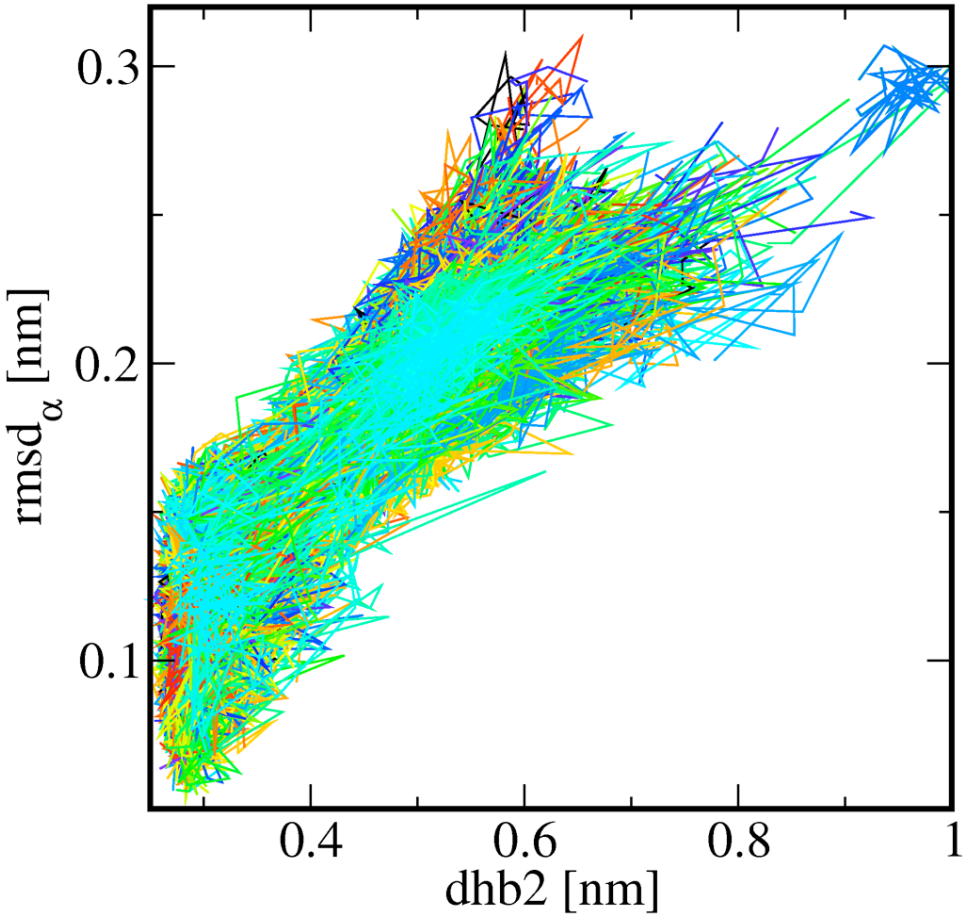
	pR (reaction)	pB' (product)
pCA-Glu46(H)	> 1.60 A	< 0.98 A
OX2-Tyr42	> 3.70 A	< 1.80 A
OX1-Tyr42	> 5.30 A	< 1.80 A

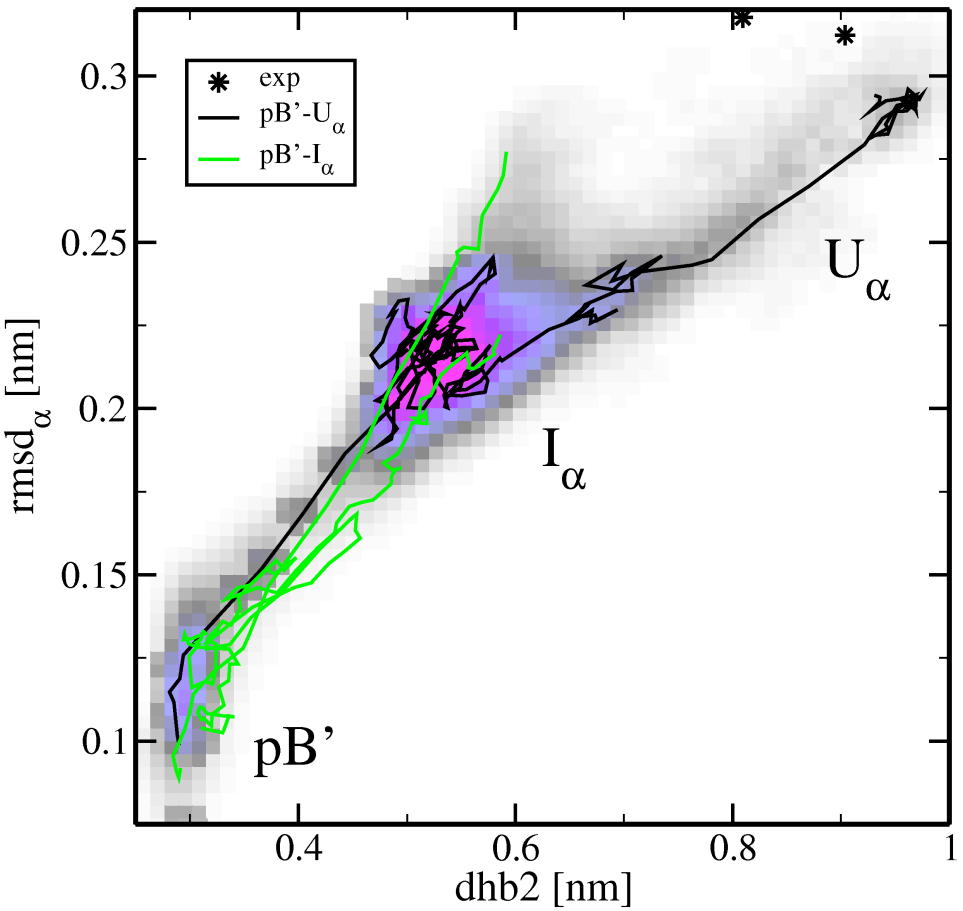
Transitions in the partial unfolding



Path ensemble

3847 accepted pathways
180 decorrelated pathways
~1 μs aggregate simulation time

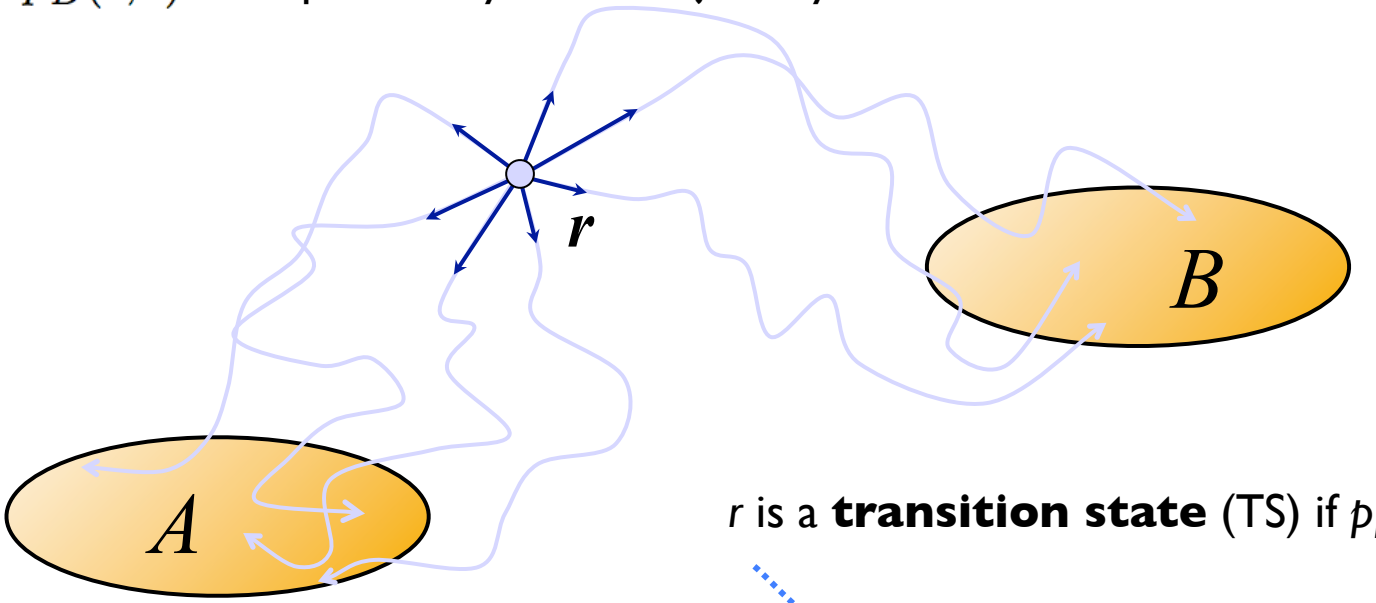




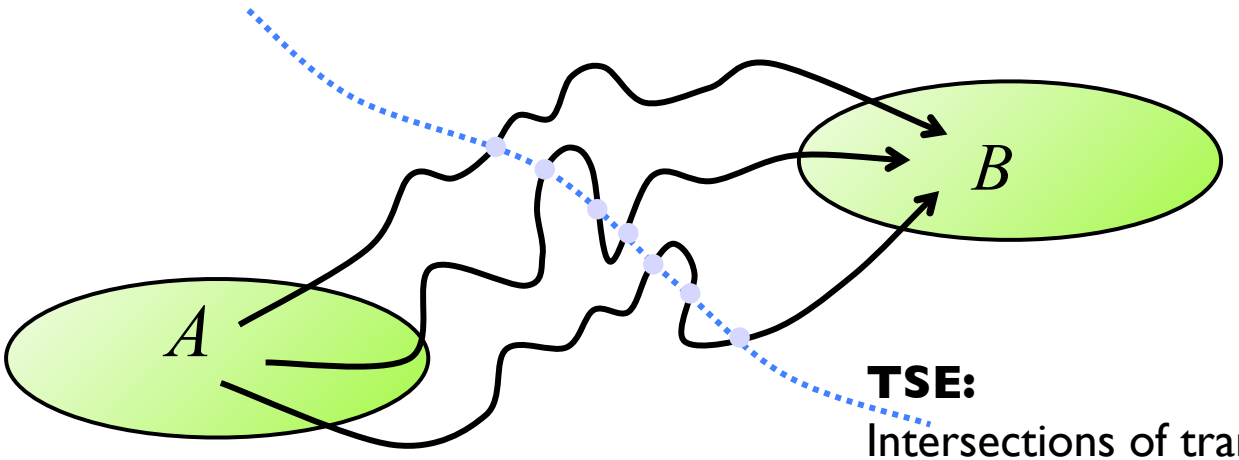
Which order parameters are relevant for the reaction coordinate?
What is the transition state?

Transition states by committor

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



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

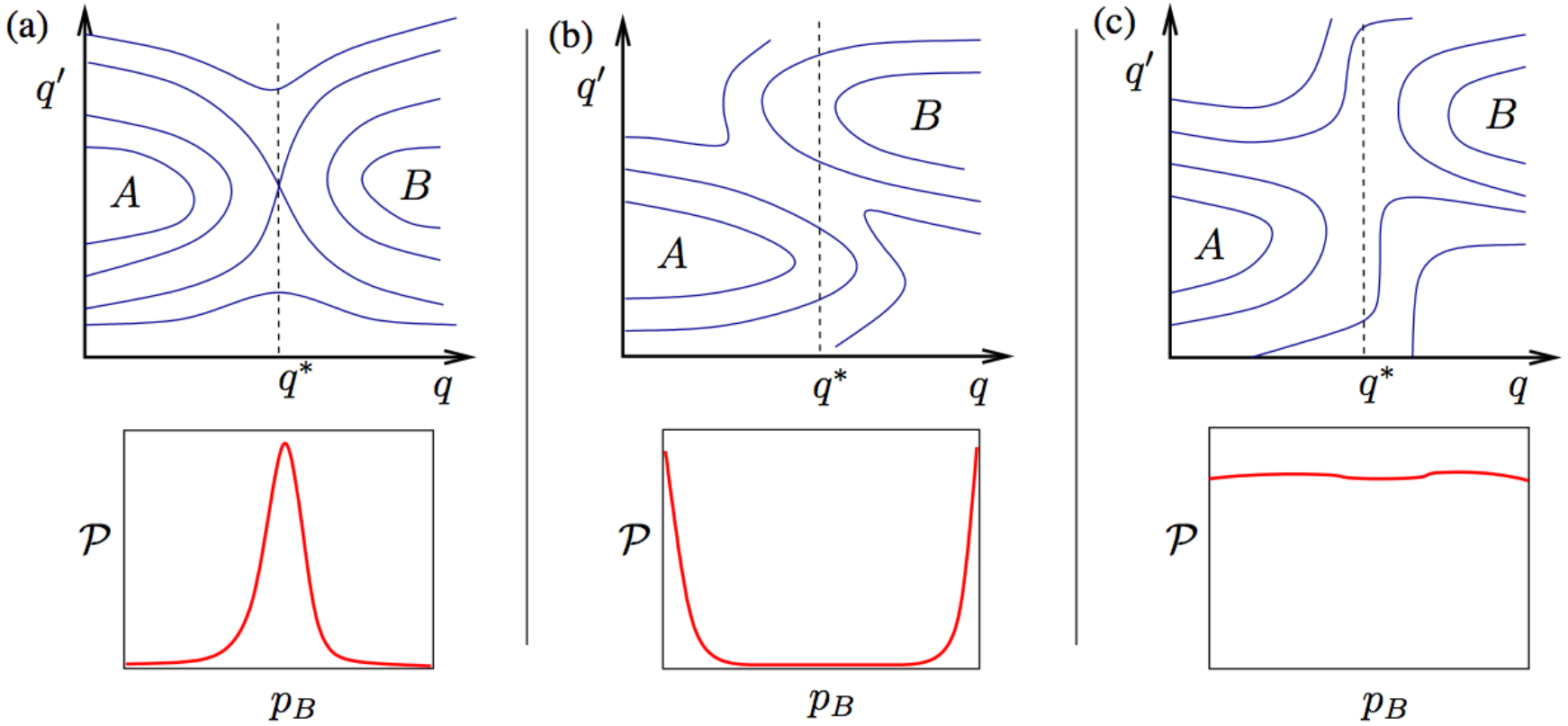


TSE:
Intersections of transition pathways with the $p_B = 1/2$ surface

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

Committor analysis

An attempt to find out the reaction coordinate



Good RC

Bad RC

analysis very expensive: requires pB histogram for every q

cheaper approaches:

GNN approach. Ma and Dinner, JPC **109** 6769 (2005)

Bayesian path distribution Best and Hummer, PNAS **102** 6732 (2005)

Likelihood Maximization. Peters and Trout, JCP **125**, 054108 (2006)

Likelihood maximization

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

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

- Assume committor function to be

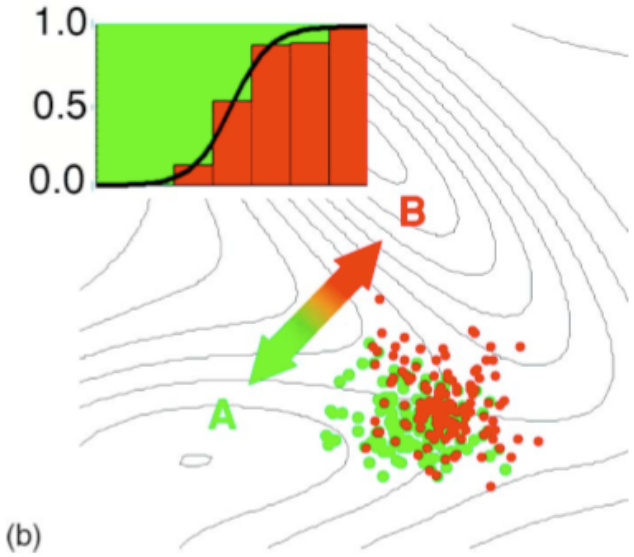
$$p_B(x) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(q(x)\right]$$

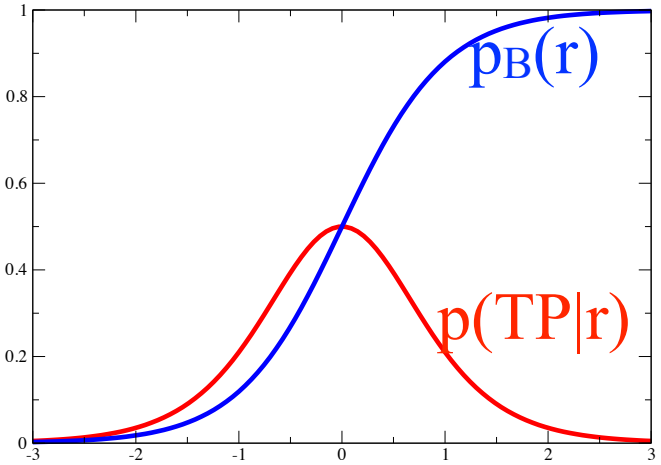
- parametrize r as linear combination of q

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

- best r is maximizing likelihood

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



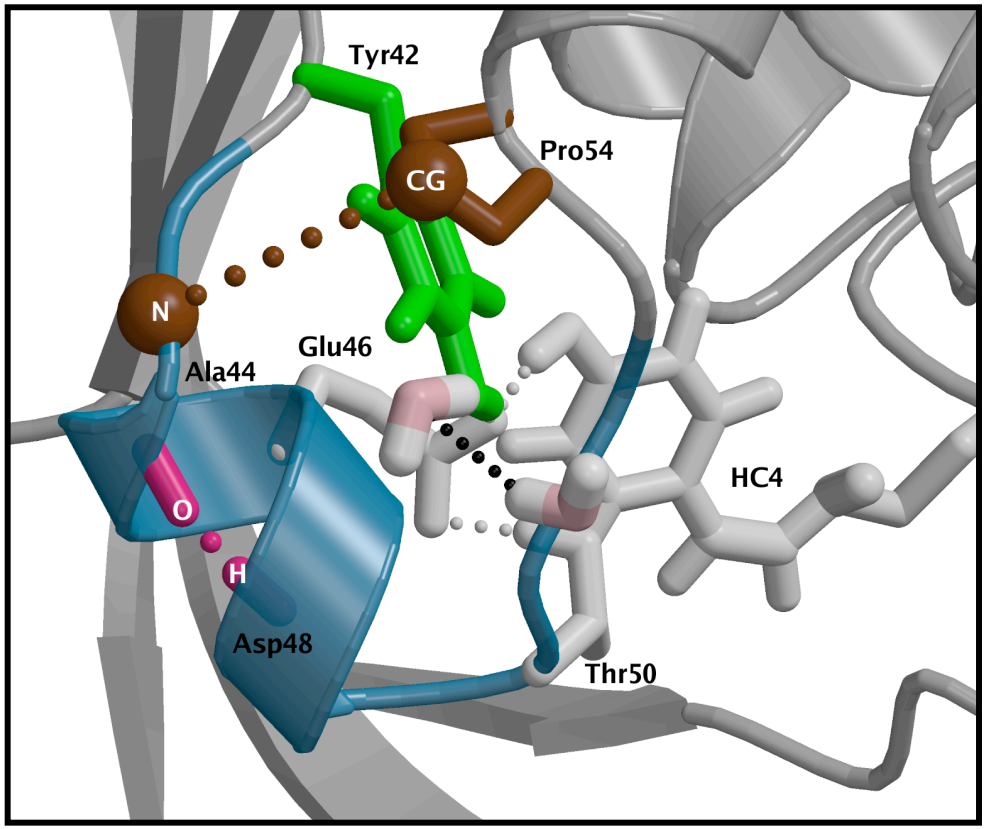


Included order parameters

Number of waters around	
pCA	nw_X
Tyr42	nw_Y
Glu46	nw_E
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	*dXN*1
pCA-O4′ – Asp97-CG	*dXN*2
Ala44-N – Pro54-CG	*dPA*
Gly47-CA – Arg52-O	*dGR*
Glu46-CD – Asn43-ND2	*dEN*
Glu46-CD – Gly51-N	*dEG*
Asn43-O – Gly47-H	*dhb*1
Ala44-O – Asp48-H	*dhb*2
Ala45-O – Ile49-H	*dhb*3
Glu46-O – Thr50-H	*dhb*4
Gly47-O – Gly51-H	*dhb*5
Asp20-CG – Lys55-NZ	*dDK*
Asp24-CG – Lys55-NZ	*dDK*2
Glu9-CD – Lys110-NZ	*dEK*
Glu12-CD – Lys110-NZ	*dEK*2
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	*d*loops1
35–37 – 114–116	*d*loops2
rmsd	
11–15	$rmsd_{N1}$
19–23	$rmsd_{N2}$
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 coordinates of helix$_{\alpha 3}$



Order Parameters involved:

RMSD$_{\alpha}$

nwY42 : water molecules around Tyr42

dPA : distance Ala44(N) - Pro54(Cγ)

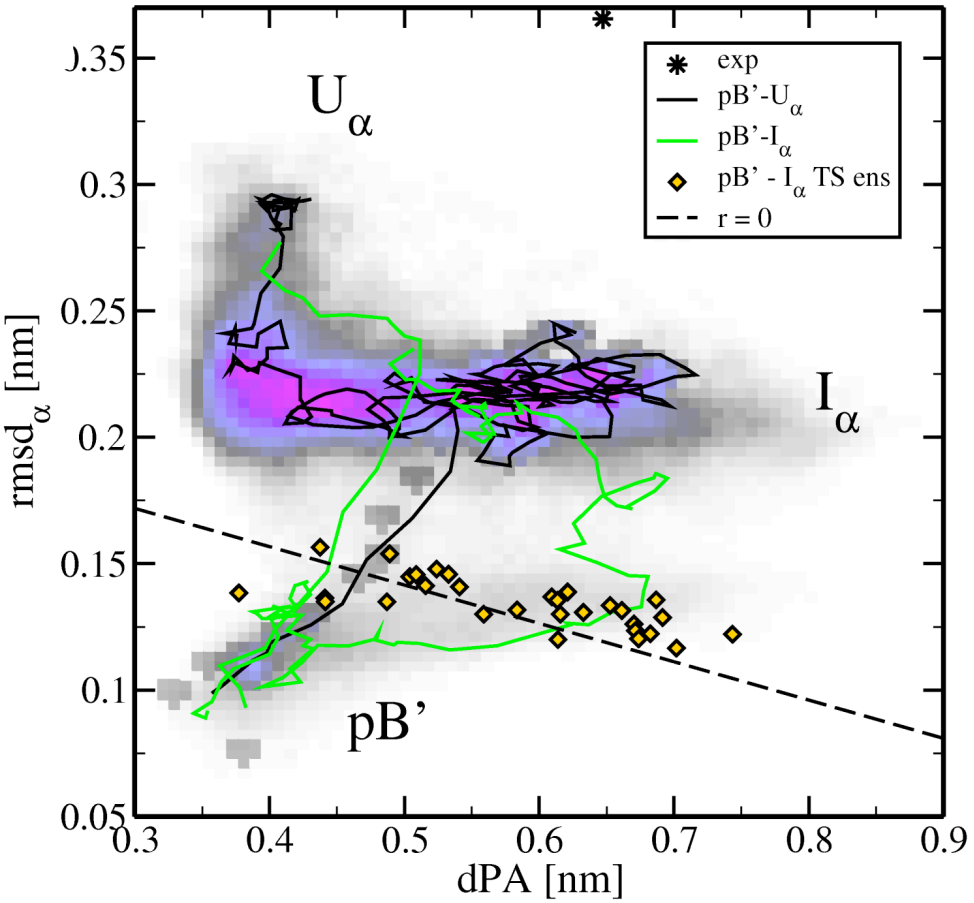
dhb2 : distance Ala44(O) - Asp48(H)

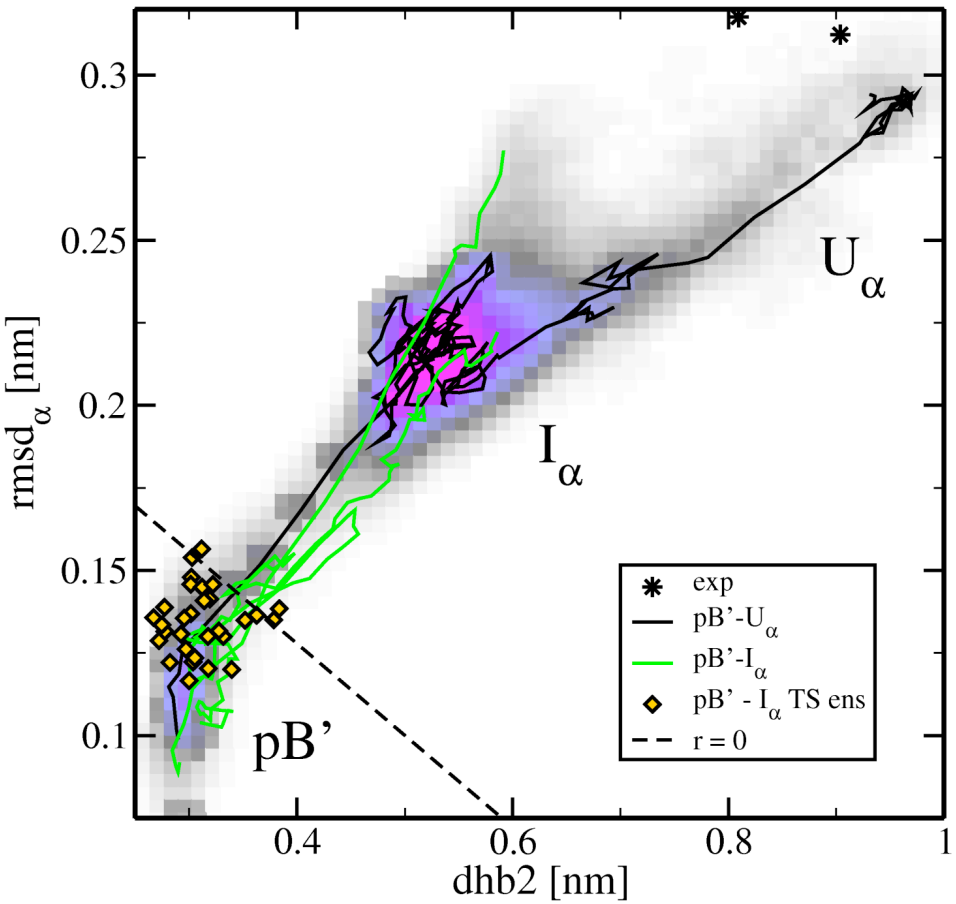
δLmin = 4.17

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

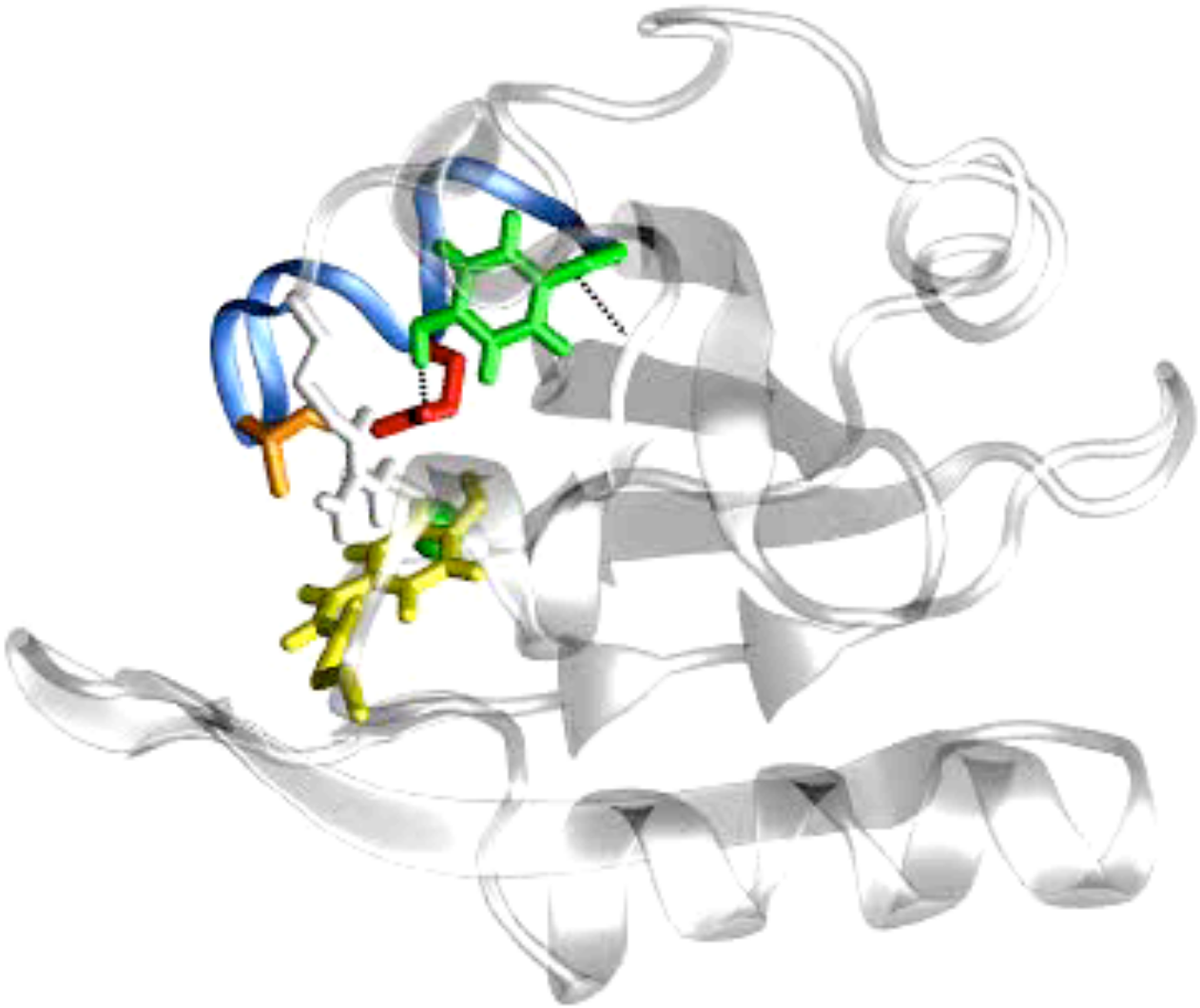
Reaction coordinates pB'→ I_α

rc =-3.49 + 15.28 $rmsd_{\alpha 3}$+ 5.65 dhb2+ 2.52 d_{PA}





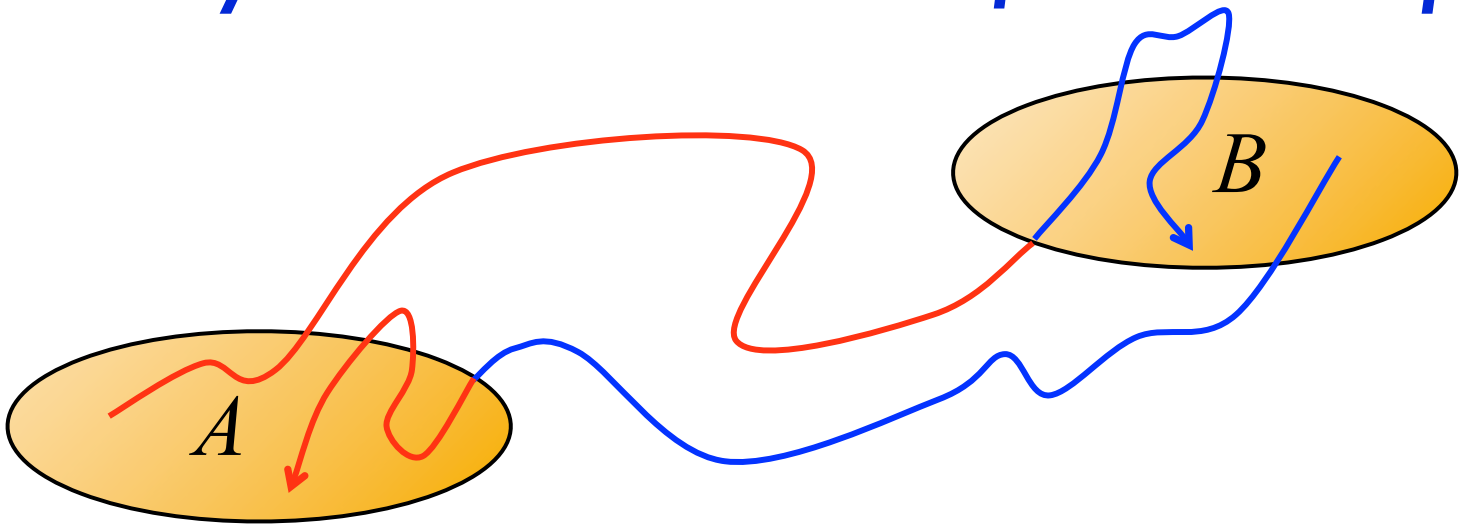
Vreede, Juraszek, PGB, PNAS 2010



Outline

- introduction on path sampling
- example on signaling protein mechanism
- reaction coordinate analysis
- **rate constants via transition interface sampling**
- new algorithm: Wang Landau path sampling
- conclusions

Rates by transition interface sampling



Overall states in phase space:

$\mathcal{A}$ points directly coming from **A**

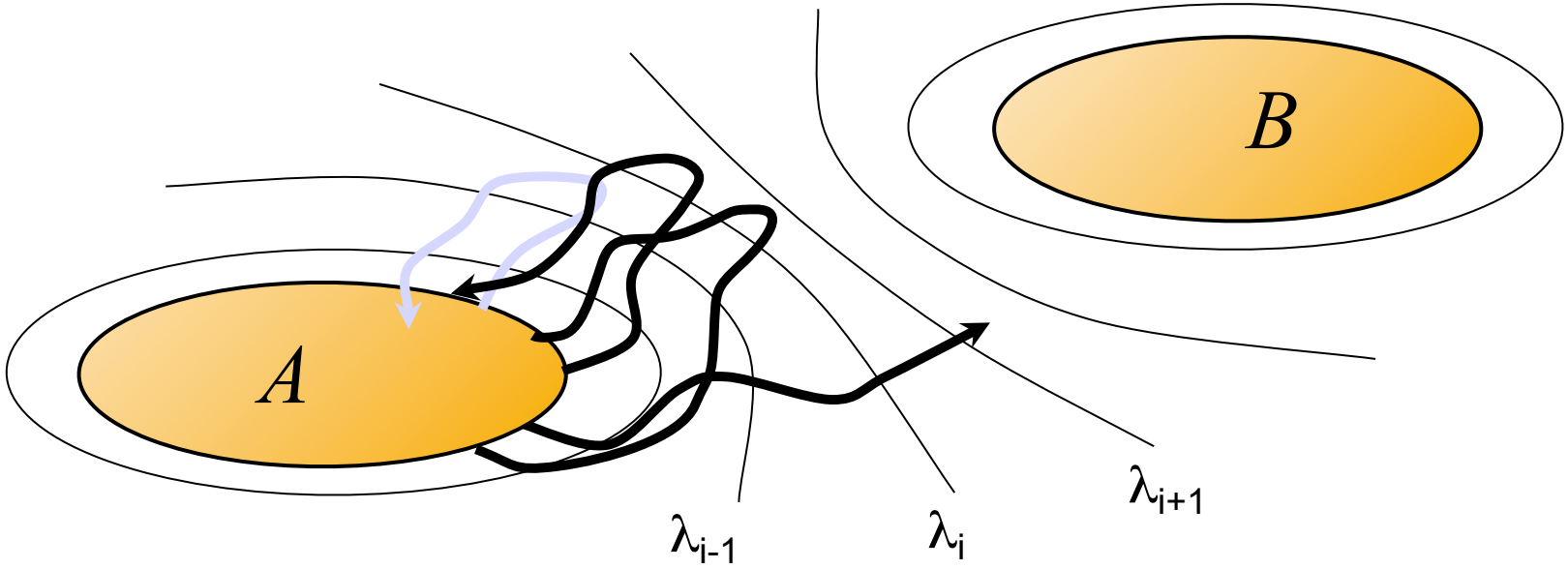
$\mathcal{B}$ points directly coming from **B**

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

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

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

Rates by transition interface sampling



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

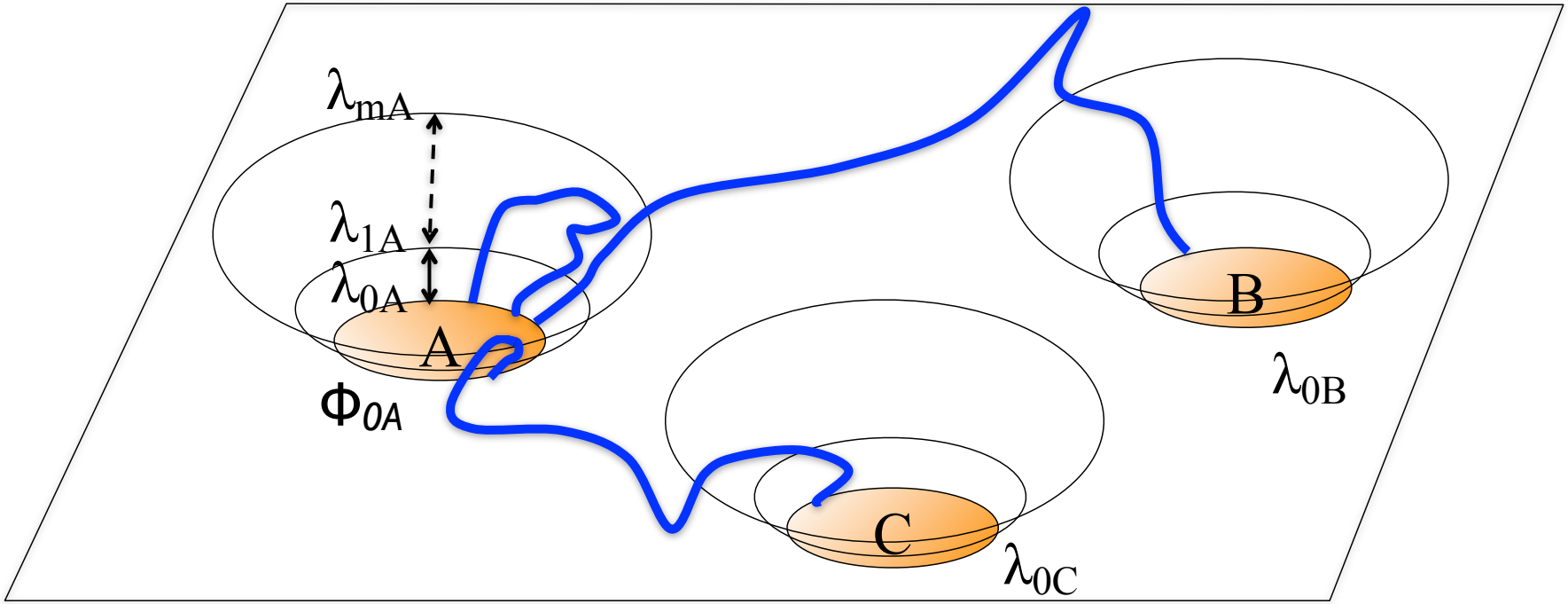
$$k_{AB} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} P_A(\lambda_B|\lambda_A) = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^{n-1} P_A(\lambda_{i+1}|\lambda_i)$$

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

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

Also the basis of FFS (ten Wolde and coworkers)

Multiple state TIS



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

TIS:

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

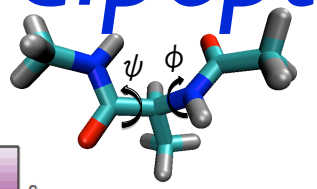
MSTIS:

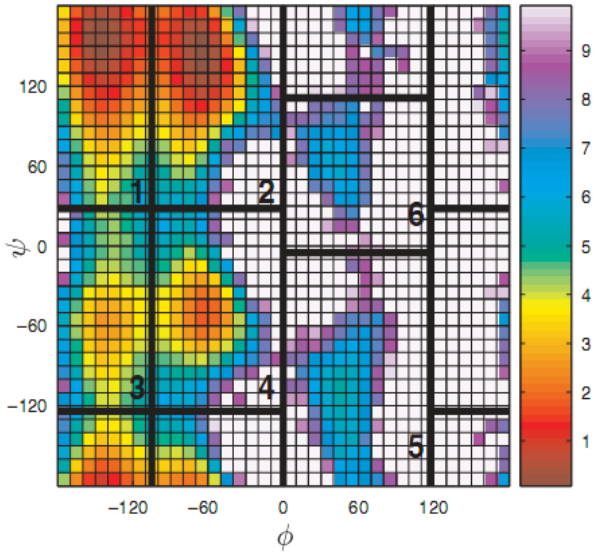
$$\frac{\text{no. of pathways coming from A, cross } \lambda_{mA}, \text{ end i}}{\text{no. of pathways coming from A, cross } \lambda_{mA}}$$

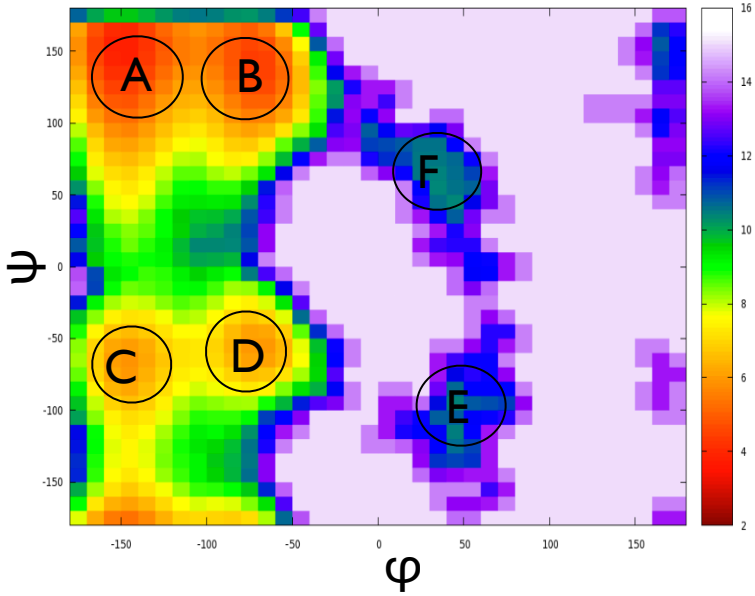
rates can be used in Markov state model

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

Alanine dipeptide in water

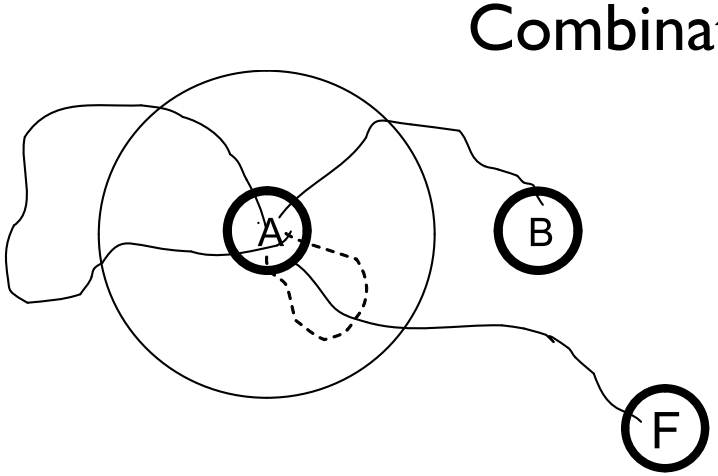


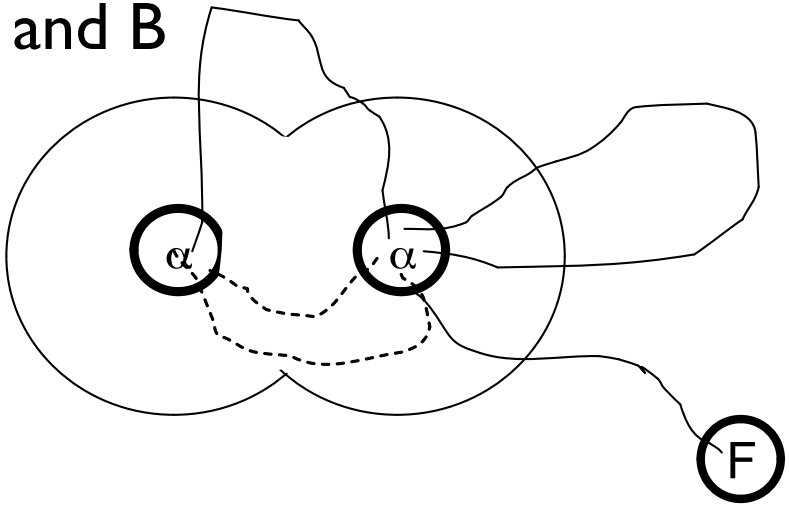




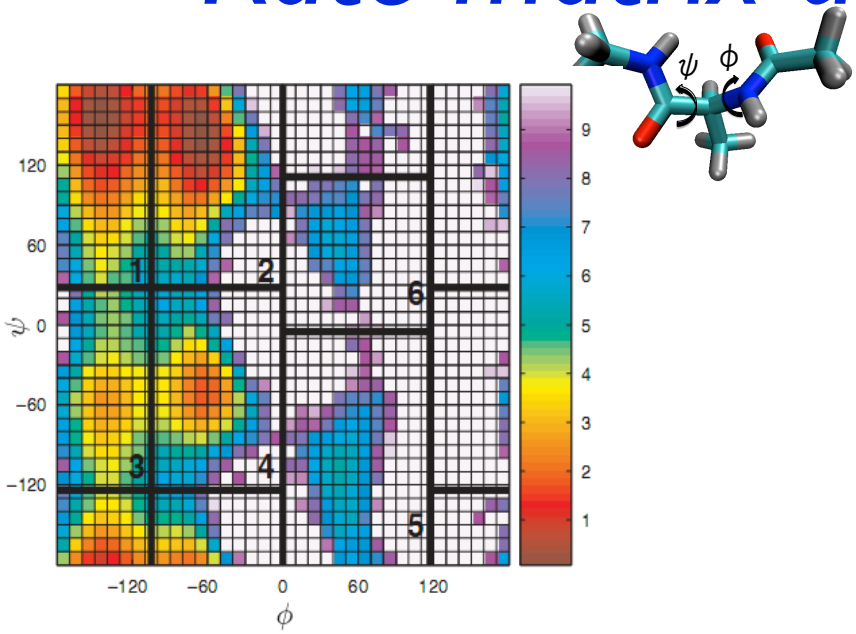
Chodera, Multiscal Model. Simul. 5, 1214-1226 (2009).

Combination of A and B

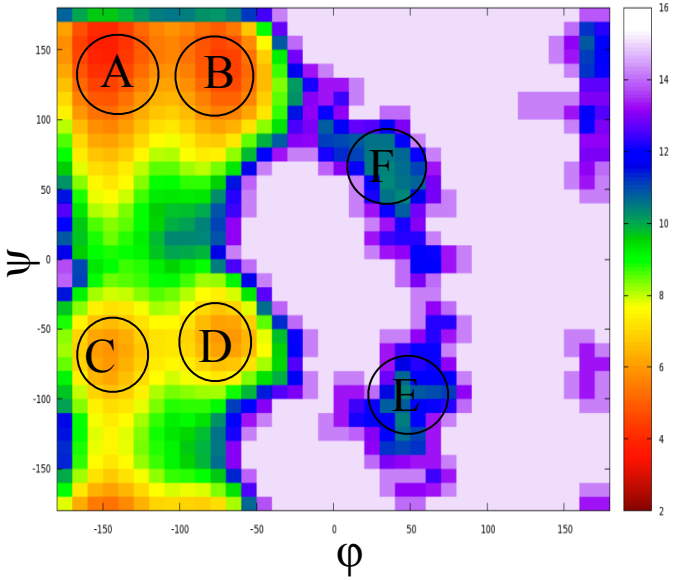




Rate matrix using MSTIS



Chodera, Multiscal Model. Simul. 5, 1214-1226 (2009).



Du, Marino & PGB, JCP 135, 145102 (2011).

From

To

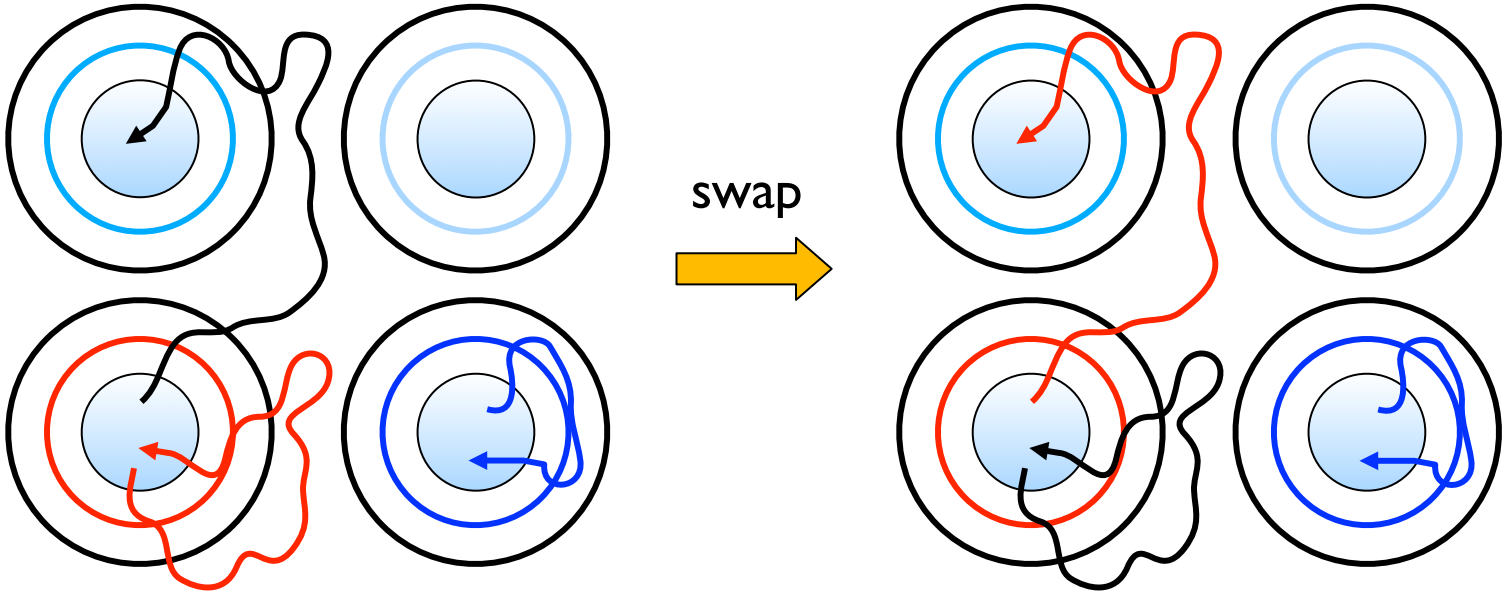
	1 and 2	3 and 4	5	6
1 and 2		0.0335	0.0011	0.073
3 and 4	0.0046		0.018	0
5	0	0.0001		0.023
6	0.0001	0	0.011	

From

	A & B	C & D	E	F
A & B		0.035	0.0011	0.038
C & D	0.0037		0.017	0.0002
E	0.000006	0.0001		0.01
F	0.0002	0.000004	0.008	

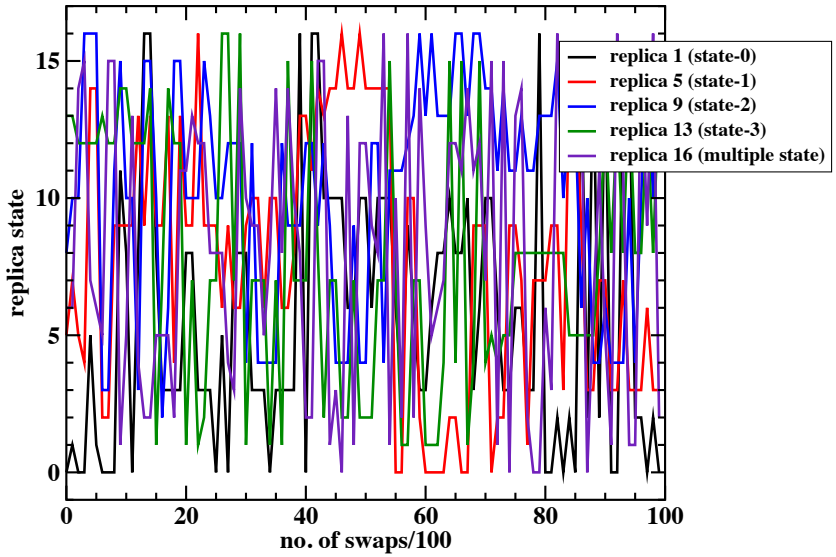
Multiple state path replica exchange

- one ensemble for each interface



advantage:
faster decorrelation
gain factor of 5-10

disadvantage:
need many replicas



Wang-Landau path sampling

- a single replica walks along set of interfaces
- each interface of i for state j has a density of states 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:

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

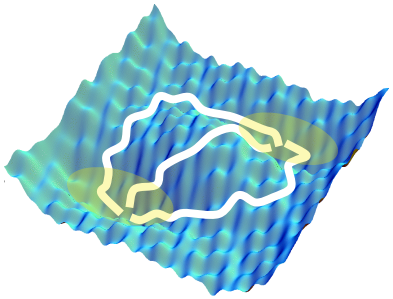
- when path switches from j→ j to j →k, allow switch to new set of interfaces for k
- if histogram is "flat" then
  - reset histogram $h_i = 0$
  - reduce factor $f = \sqrt{f}$, continue
- weights g_i reflect ratio of pathways on each interface = crossing probability

advantage:
only one replica needed

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

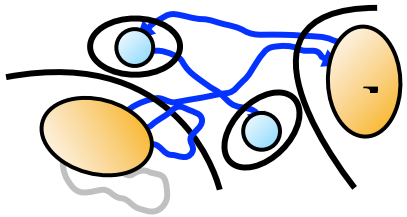
F. Wang and D.P. Landau, PRL 86, 2050 (2001)

Conclusions



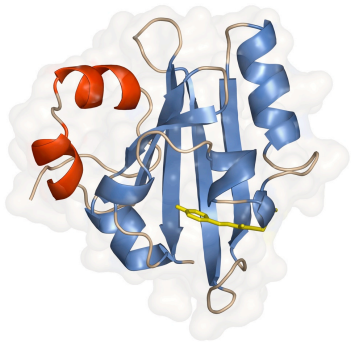
Brute force dynamics is inefficient for high barriers
Rare event techniques are exponentially more efficient than straightforward dynamics

Biasing simulation along predefined collective variables suffers from the reaction coordinate problem



In the TPS framework one can sample the path ensemble,
obtain rates via transition interface sampling,
optimize reaction coordinates using likelihood maximization

Replica exchange and Wang Landau TIS can sample the complete path ensemble, reweighting yields free energy, committor and rates directly



Multiple state TPS can sample systems with many intermediates

Biasing of unbiased trajectories solves the long time scale problem and allows us to understand complex processes such as protein conformational changes

Acknowledgments

- BRG members
  - Lizhe Zhu
  - Weina Du
  - Rosanne Zeiler
  - Kristen Marino
  - Kush Singhal
  - Ethayaraja Mani
  - Aatish Kumar
  - Juriaan Luiken
  - Ana Vila Verde
  - Marcin Nowosielsky
  - Aurora Cruz Cabeza
  - Siri van Keulen
- Ex members
  - Marieke Schor
  - Wolfgang Lechner
  - Jutta Rogal
  - Jarek Juraszek
  - Elske Leenders

- Funding
  - NWO
  - FOM
  - EU

- Collaborators
  - Evert Jan Meijer
  - Bernd Ensing
  - Jocelyne Vreede
  - Christopher Lowe
  - M. Cohen Stuart (Wageningen)
  - Willem Kegel (Utrecht)
  - Chistroph Dellago (Vienna)
  - Klaas Hellingwerf (UvA)
  - Huib Bakker (UvA)
  - Peter Schall (UvA)
  - Sander Woutersen (UvA)
  - Sander Tans (Amolf)