Molecular rare event simulations

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

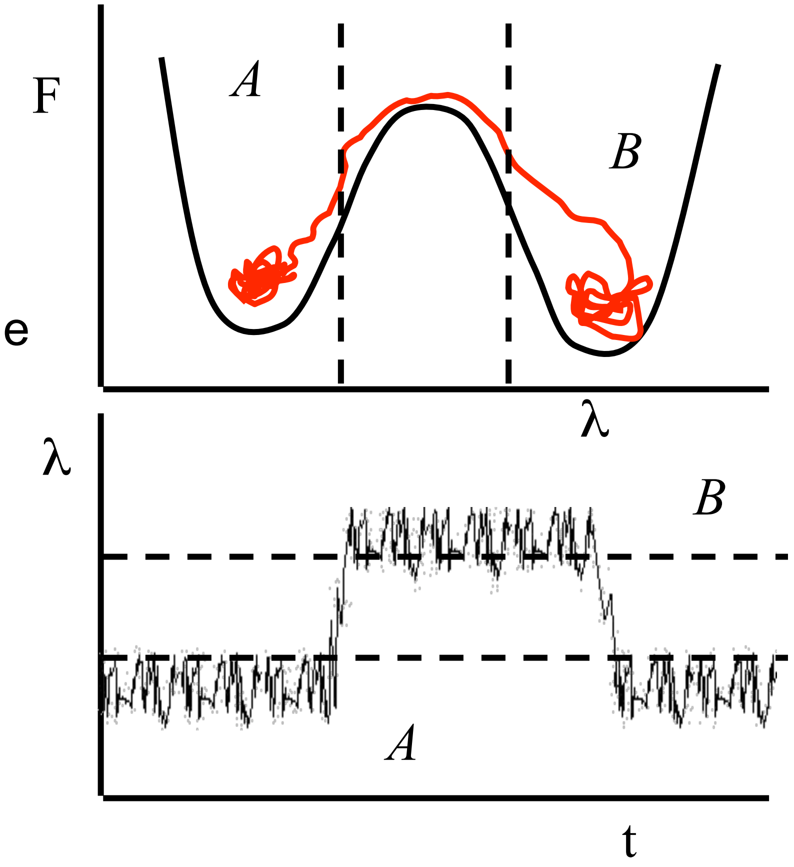
Rare events

Interesting transitions in complex systems

- solution chemistry
- protein folding
- enzymatic reactions
- complex surface reactions
- nucleation

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

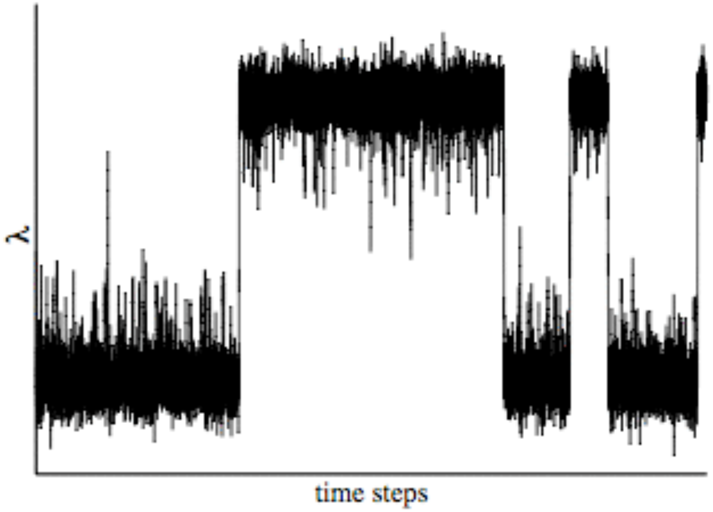
dominated by collective, rare events

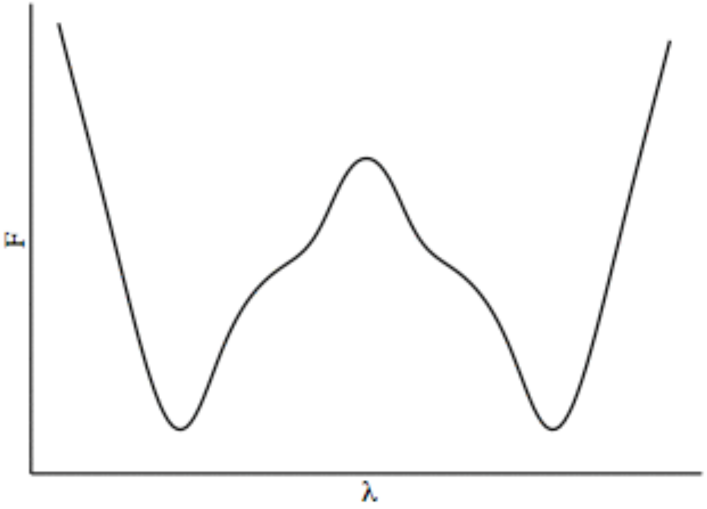


Rare events

The characteristic of a rare event is that the time to cross the barrier is much smaller than the time spend in the stable states

$$\tau_{mol} \lll \tau_{stable}$$

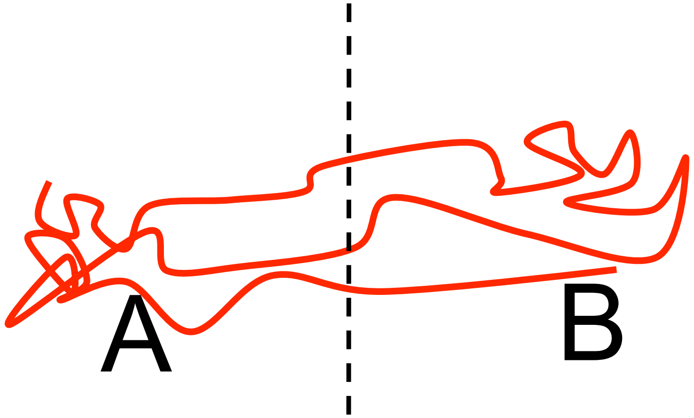




$$e^{-\beta F(\lambda^*)} \equiv \langle \delta(\lambda(r) - \lambda^*) \rangle = \int d\rho(r)\delta(\lambda(r) - \lambda^*)$$

Rate constants

- Definition of rate for A to B event
  - number of transitions per time unit

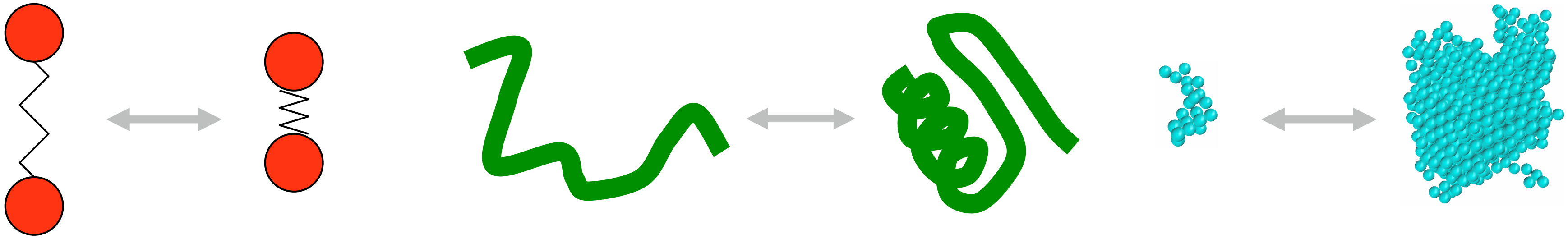


$$k_{AB} = \lim_{\mathcal{T}\to\infty} \frac{N_{A\to B}(\mathcal{T})}{t^{A}_{tot}(\mathcal{T})}$$

- Alternative: the inverse of the mean residence time

$$k^{-1}_{AB} = \lim_{\mathcal{T}\to\infty} \frac{1}{N_{A\to B}(\mathcal{T})} \sum_{i=1}^{N_{A\to B}} t^{A}_{i} \equiv t^{mr}_{A}$$

Phenomenological reaction kinetics



A ⇌ B

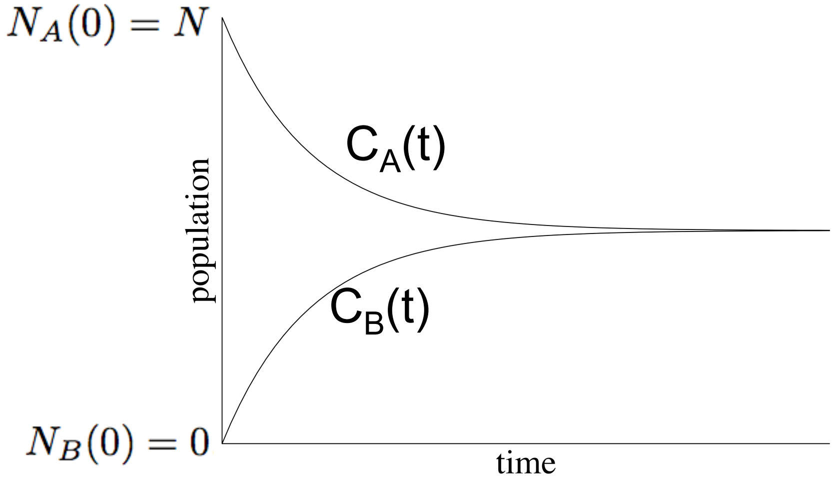
$$\dot{c}_A = -k_{AB}\, c_A + k_{BA}\, c_B$$

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

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

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

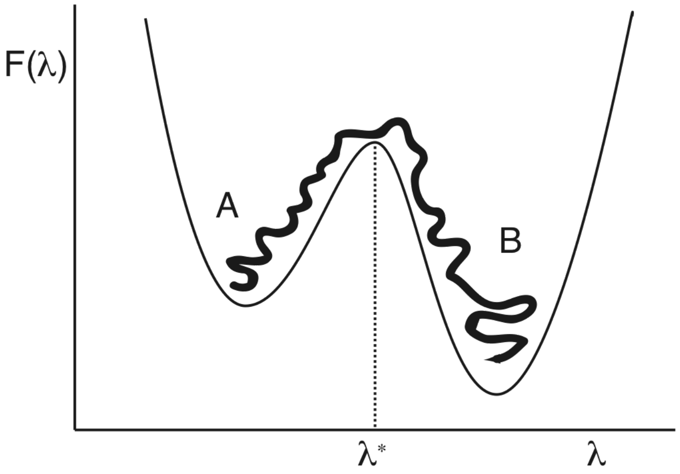
$$\tau_{\text{rxn}}^{-1} = k_{AB} + k_{BA}$$

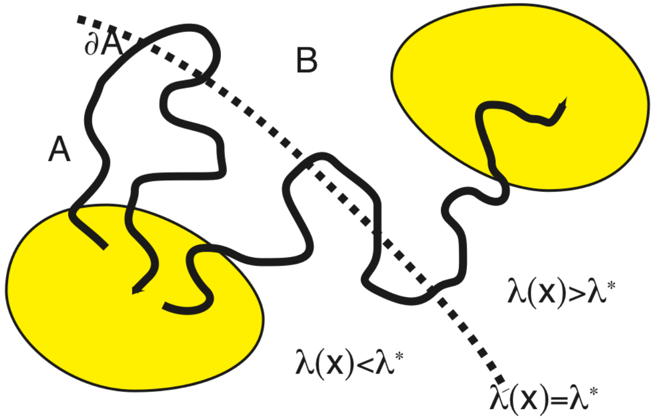


Transition state theory

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

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



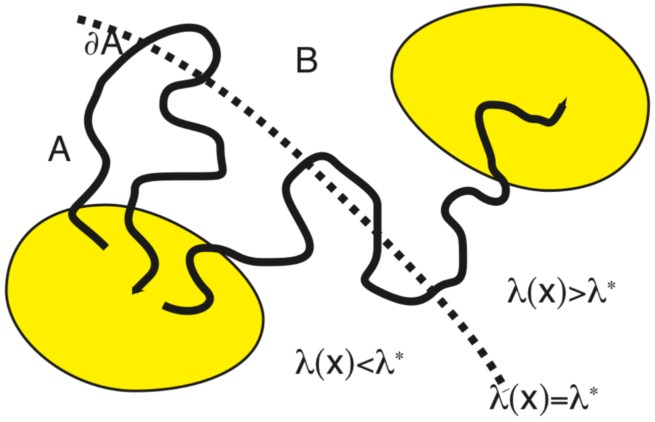


- How do we get to this expression?

Derivation of TST

Start with definition of stable region Ω=A,B

$$h_\Omega(r) = \begin{cases} 1 & \text{if } r \in \Omega \\ 0 & \text{otherwise} \end{cases}$$



$$\int_\Omega \rho(r)dr = \langle h_\Omega(r) \rangle$$

together they span the whole space

$$\langle h_A \rangle + \langle h_B \rangle = 1$$

The residence time in A is

$$t_A^{mr} = \lim_{\mathcal{T} \to \infty} \frac{2}{N(T)} \int_0^{\mathcal{T}} h_A(r(t))dt$$

where N(T) is number of crossings on an infinite trajectory

Rate constant in TST

The rate is the inverse residence time

$$k_{AB}^{TST} = (t_A^{mr})^{-1} = \frac{\nu^{TST}}{\langle h_A \rangle}$$

and equals the number of crossings per unit time

$$\nu^{TST} = \lim_{\mathcal{T} \to \infty} \frac{N(\mathcal{T})}{2\mathcal{T}}$$

divided by the equilibrium population in A

$$\langle h_A \rangle = \lim_{\mathcal{T} \to \infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} h_A(r(t))dt = \int_A \rho(r)dr$$

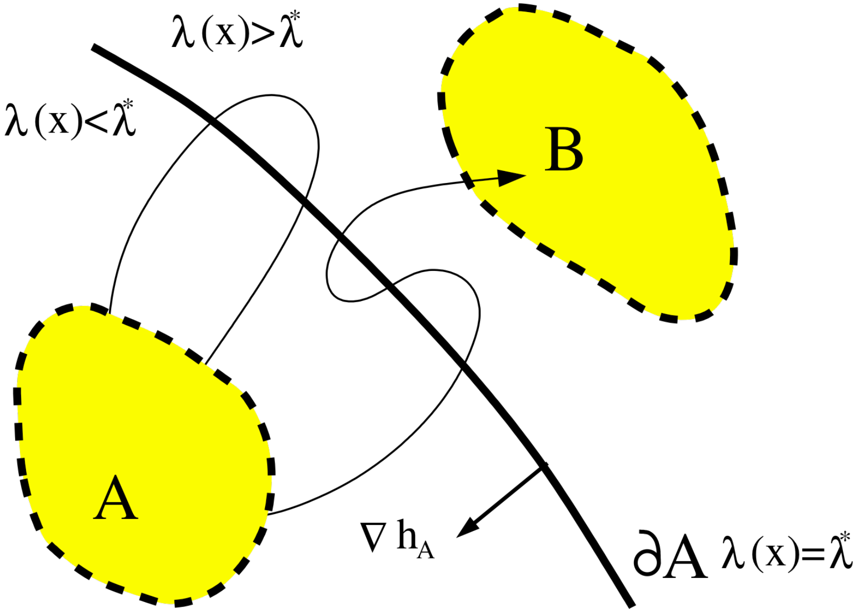
ergodicity

$$
\begin{aligned}
\nu^{TST} &= \lim_{\mathcal{T}\to\infty} \frac{1}{2\mathcal{T}} \int_0^{\mathcal{T}} |\dot{h}_A(r(t))| dt \\
&= \lim_{\mathcal{T}\to\infty} \frac{1}{2\mathcal{T}} \int_0^{\mathcal{T}} |\nabla h_A(r) \cdot v(t)| dt \\
&= \int dx \rho(x) |\nabla h_A(r) \cdot v|
\end{aligned}
$$

$$
\begin{aligned}
A &= \{r \in \mathbb{R}^n : \lambda(r) < \lambda^*\} \\
\partial A &= \{r \in \mathbb{R}^n : \lambda(r) = \lambda^*\}
\end{aligned}
$$

$$
h_A(r) = \theta(\lambda^* - \lambda(r))
$$

$$
h_B(r) = \theta(\lambda(r) - \lambda^*)
$$



….almost there…

$$\begin{aligned}\nu^{TST} &= \frac{1}{2}\int dx\rho(x)\delta(\lambda(r)-\lambda^*)|\nabla\lambda\cdot v| \\ &= \frac{1}{2}\int dx\rho(x)\delta(\lambda(r)-\lambda^*)|\dot\lambda| \\ &\equiv \frac{1}{2}\langle\delta(\lambda(r)-\lambda^*)|\dot\lambda|\rangle \\ &= \langle\delta(\lambda(r)-\lambda^*)\dot\lambda\theta(\dot\lambda)\rangle\end{aligned}$$

$$k_{AB}^{TST} = (t_A^{mr})^{-1} = \frac{\nu^{TST}}{\langle h_A\rangle}$$

$$h_A(r) = \theta(\lambda^* - \lambda(r))$$

$$k_{AB}^{TST} = \frac{\langle\delta(\lambda(r)-\lambda^*)\dot\lambda\theta(\dot\lambda)\rangle}{\langle\theta(\lambda^*-\lambda(r))\rangle}$$

Finally….

Multiply and divide k by $\delta(\lambda(r)-\lambda^*)$

$$
\begin{aligned}
k_{AB}^{TST} &= \frac{\langle \delta(\lambda(r)-\lambda^*)\dot{\lambda}\theta(\dot{\lambda})\rangle}{\langle \delta(\lambda(r)-\lambda^*)\rangle} \frac{\langle \delta(\lambda(r)-\lambda^*)\rangle}{\langle \theta(\lambda^*-\lambda(r))\rangle} \\
&= \langle \dot{\lambda}\theta(\dot{\lambda})\rangle_{\lambda=\lambda^*} \frac{e^{-\beta F(\lambda^*)}}{\int_{-\infty}^{\lambda^*} e^{-\beta F(\lambda')} d\lambda'}
\end{aligned}
$$

constraint to
dividing surface

$$e^{-\beta F(\lambda^*)} \equiv \langle \delta(\lambda(r)-\lambda^*)\rangle = \int d\rho(r)\delta(\lambda(r)-\lambda^*)$$

Rate as equilibrium quantity

Dynamical property has been changed into static equilibrium quantity!

$$k_{AB}^{TST} \propto e^{-\beta(F(\lambda^*)-F(\lambda_A))} \equiv e^{-\beta \Delta F}$$

indeed exponential dependence on free energy barrier

prefactor

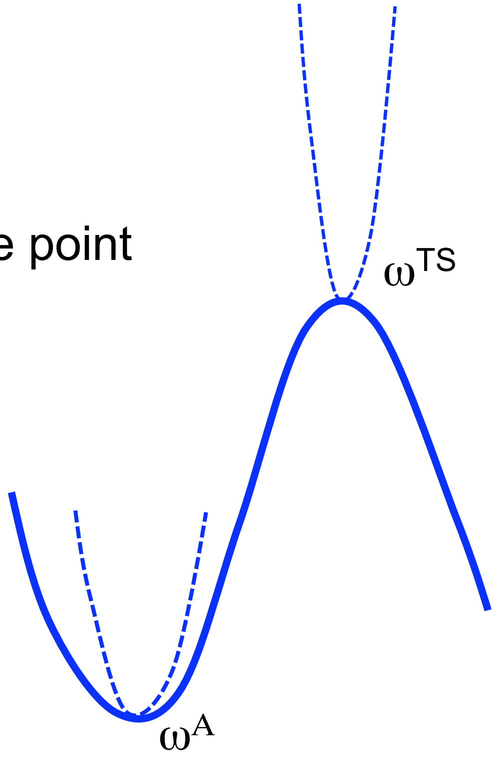
$$\langle \dot{\lambda}\theta(\dot{\lambda})\rangle_{\lambda=\lambda^*} = \frac{1}{2}\langle|\dot{\lambda}|\rangle_{\lambda=\lambda^*}$$

can be computed assuming Gaussian momenta

$$\frac{1}{2}\langle|\dot{\lambda}|\rangle_{\lambda=\lambda^*} = \sqrt{\frac{k_B T}{2\pi m}}$$

Harmonic TST

Assume quadratic form of the potential close to the minimum and saddle points (good approximation in gas phase)



Denoting ω_i eigenfrequencies along minima and saddle point

$$k_{\mathrm{TST}}^{ha} = \frac{1}{2\pi} \frac{\prod_{i=1}^{n} \omega_i^A}{\prod_{i=1}^{n-1} \omega_i^{TS}} \exp(-\beta \Delta V)$$

one more 'mode' in stable minima than in TS

$$k_{\mathrm{TST}}^{ha}(T) = \nu \exp(-\beta \Delta V)$$

v is entropic factor that is large if saddle point is wide and small if its is narrow

RRKM theory

Sometimes it is better to work in the microcanonical ensemble (e.g. in gas phase)

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

Then the rate constant becomes (Rice, Ramsperger, 1927, Kassel 1928, Marcus, 1951)

$$k_{\mathrm{RRKM}}(E) = \frac{1}{2\pi} \frac{\prod_{i=1}^{n} \omega_i^A}{\prod_{i=1}^{n-1} \omega_i^{TS}} \left[\frac{E - V_{TS}}{E - V_A} \right]^{n-1}$$

The T dependent rate follows then from integration over the energies.

$$k(T) = \frac{\int dE\, k(E) g(E) \exp(-\beta E)}{\int dE\, g(E) \exp(-\beta E)}$$

for large systems with many d.o.f.'s k(E) ≈ k(T)

Computing barriers

Clearly, barrier is most important for rare event

But how to obtain this barrier?

Straightforward MD or MC and then use $\beta F(\lambda') = -\ln\langle\delta(\lambda(r) - \lambda')\rangle$ is inefficient

There are several tricks to sample barriers

- **Thermodynamic integration**
- **Umbrella sampling**
- **Metadynamics**

Thermodynamic integration

The free energy follows from the derivative

$$\beta F(\lambda_1) - \beta F(\lambda_0) = \int_{\lambda_0}^{\lambda 1} \left(\frac{\partial F}{\partial \lambda} \right) d\lambda$$

The derivative of the free energy is known as the mean force

$$\left(\frac{\partial F}{\partial \lambda} \right) = \left\langle \frac{dU}{\partial \lambda} \right\rangle \equiv \langle f(\lambda) \rangle$$

we can compute the force at a particular lambda value by adding a constraint to the Lagrangian

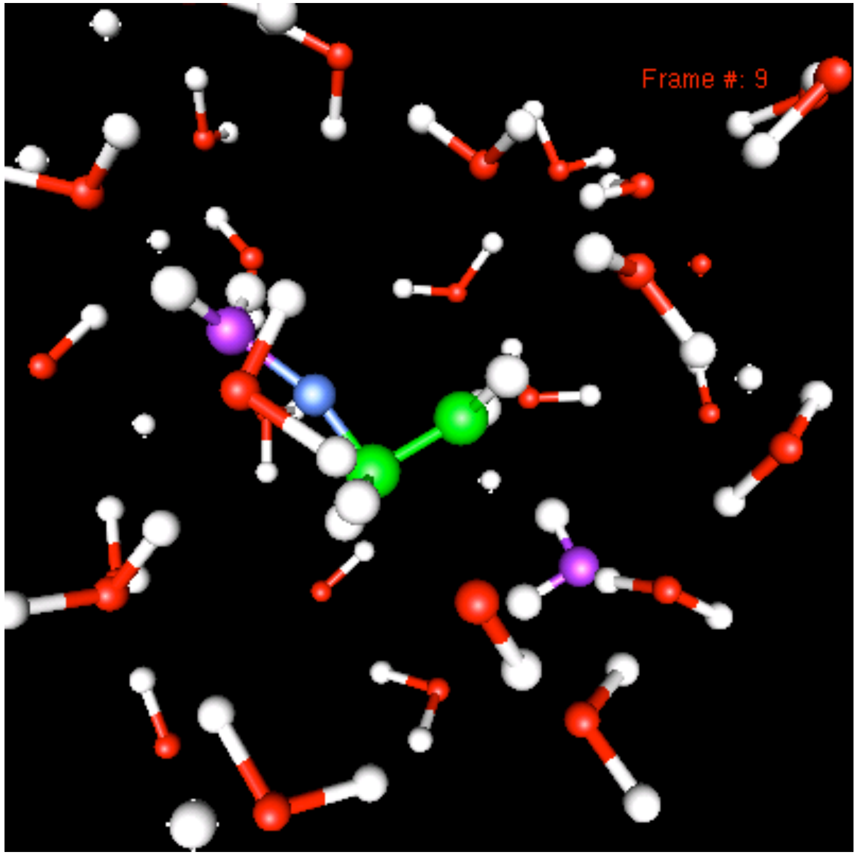
$$\mathcal{L}_c(\dot{r}^N, r^N) = \mathcal{L}(\dot{r}^N, r^N) - g(\sigma(r^N) - \lambda)$$

constraint force

the constraint force follows from the Lagrange multiplier

$$\langle f(\lambda) \rangle = \langle g \rangle$$

Example: Alkene hydration



$C_2H_4 + H_2O$ <--> CH_3CH_20H

SYSTEM

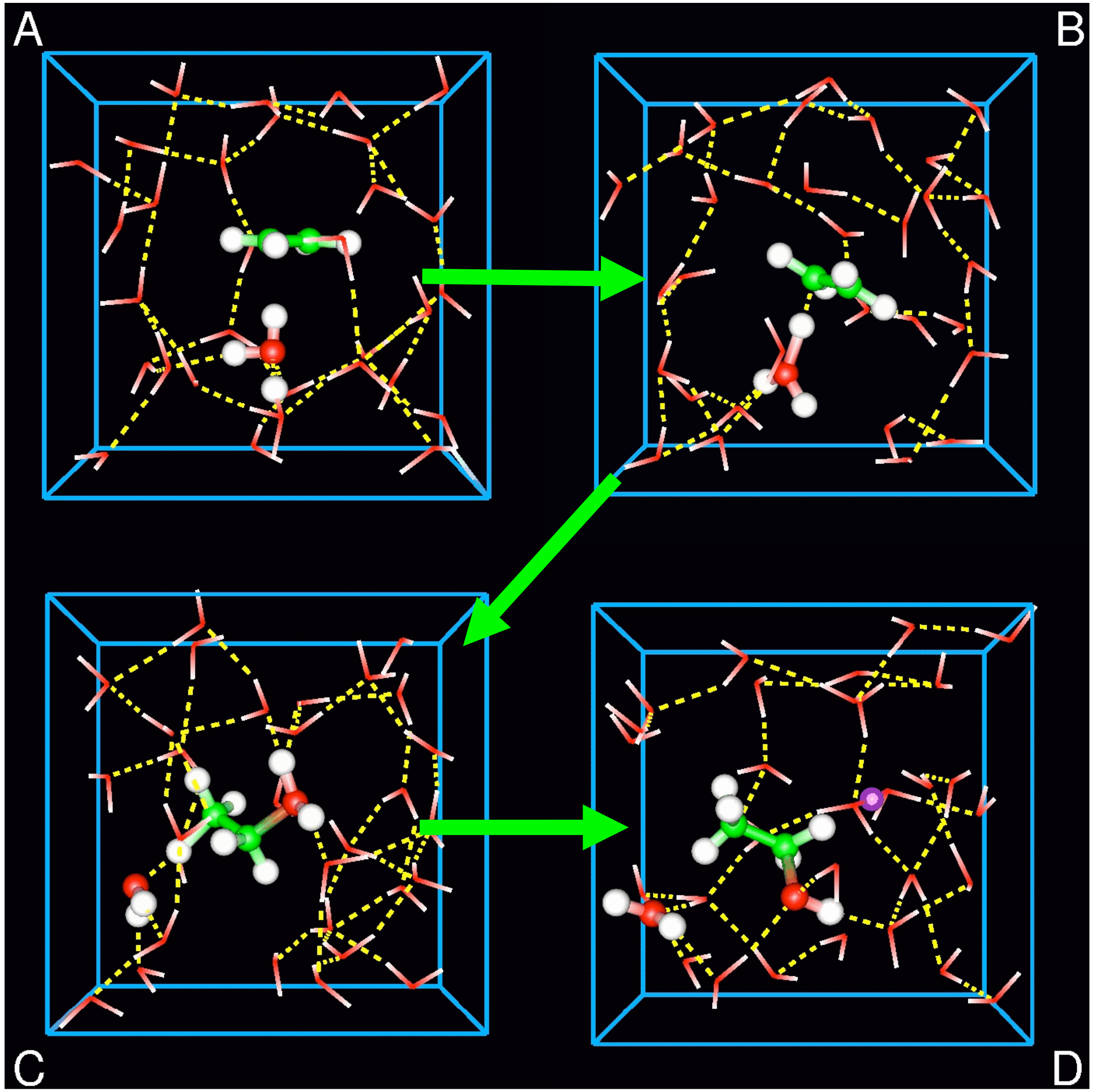
$32\ H_2O\ + H^+\ + C_2H_4$

T=300K

REACTION COORDINATE Q

$Q = R_{OH} - R_{HC}$

T. Van Erp, E-J Meijer
Angew. Chem, 43, 1660 (2004).



Q=-1.1 Å

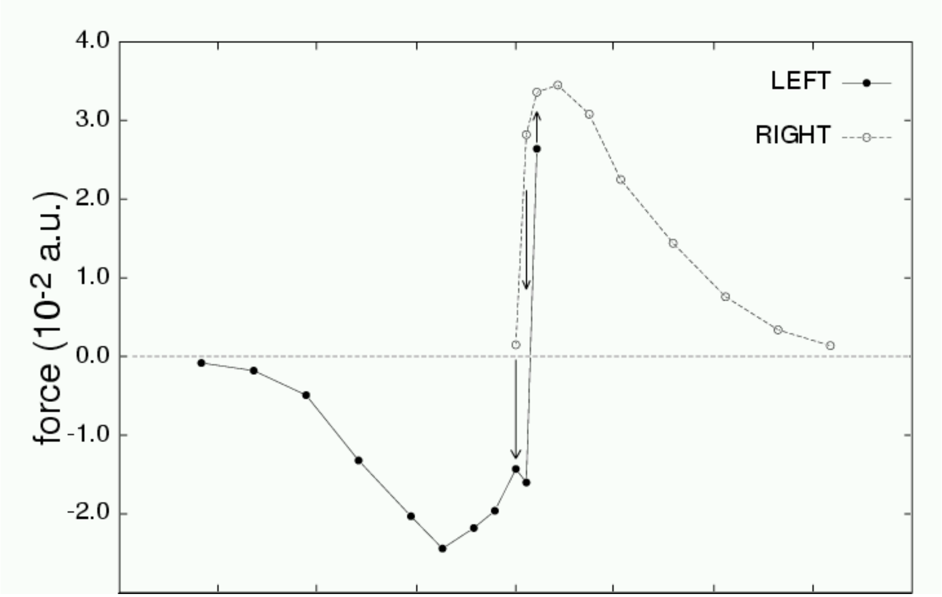
Q=0.0 Å

Q=0.1 Å

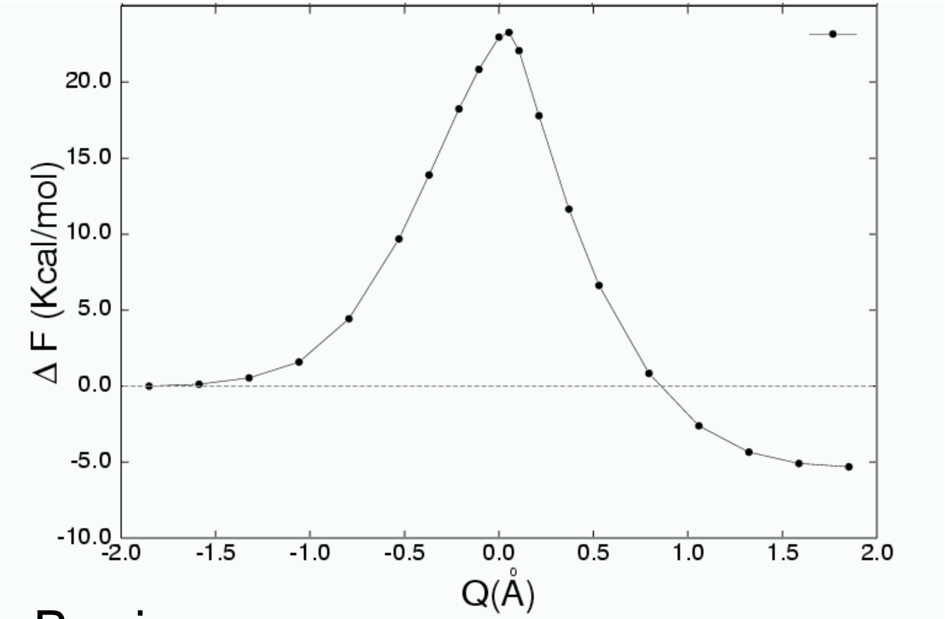
Q=1.1 Å

Example: Alkene hydration

CONSTRAINT FORCE



FREE ENERGY PROFILE



Reaction Barrier

CPMD-BLYP	23 kcal/mol
Exp: Gas Phase	50-100
MP2: Gas Phase	58
Exp: Low Density Acid Solution	33
BLYP: Gas Phase + Acid	24

Umbrella sampling

The regular distribution of an order parameter λ is

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

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

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

where V_{bs} is the bias potential

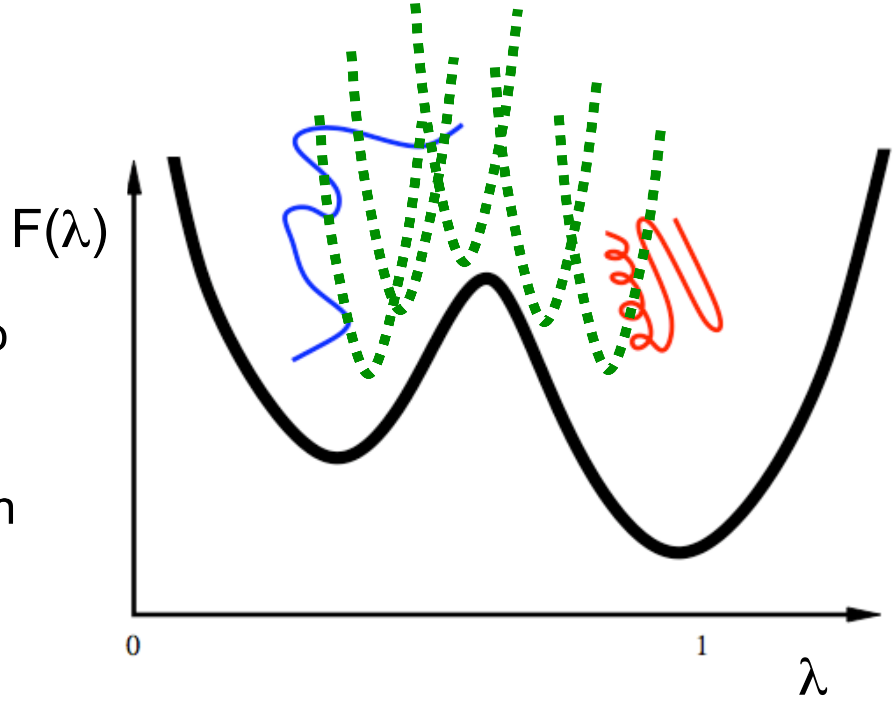
Free energy is given by

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

Umbrella sampling

Biasing potential can take any functional form
simplest one quadratic

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



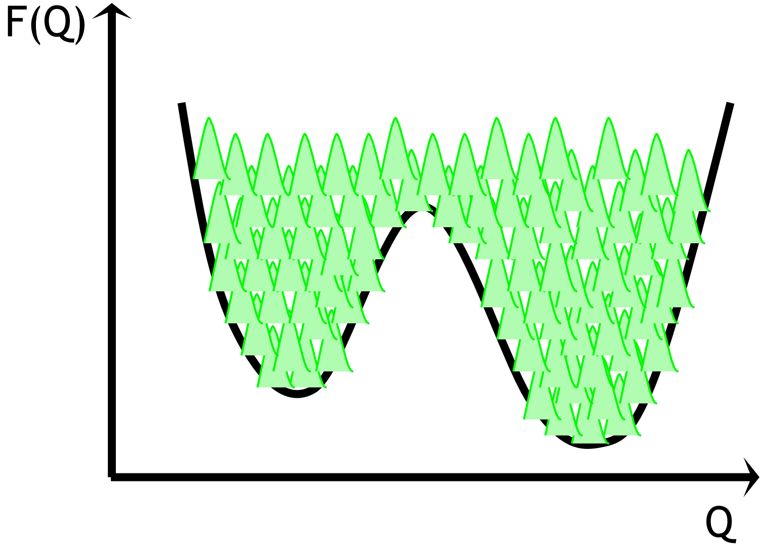
Biasing potential forces system in to unlikely region

Also used in flooding, local elevation hyperdynamics

Metadynamics

- Recent method to get free energy in a single simulation (Laio & Parrinello PNAS 2002)
- Adds history dependent biasing potential to forcefield 'on the fly'

$$V(s;t) = w \sum_{t'<t} e^{-\frac{(s-s(x(t'))^2}{2\sigma}}$$



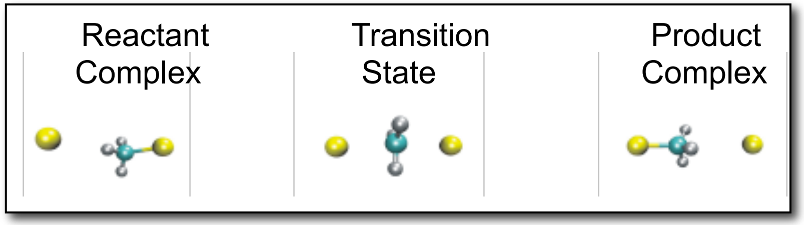
- s = predefined order parameters
- w = height of hills
- σ = width of gaussians

- w becomes smaller every cycle

$$F(s) = -\lim_{t\to\infty} V(s;t)$$

SN2 reaction beteen Cl- and CH3Cl

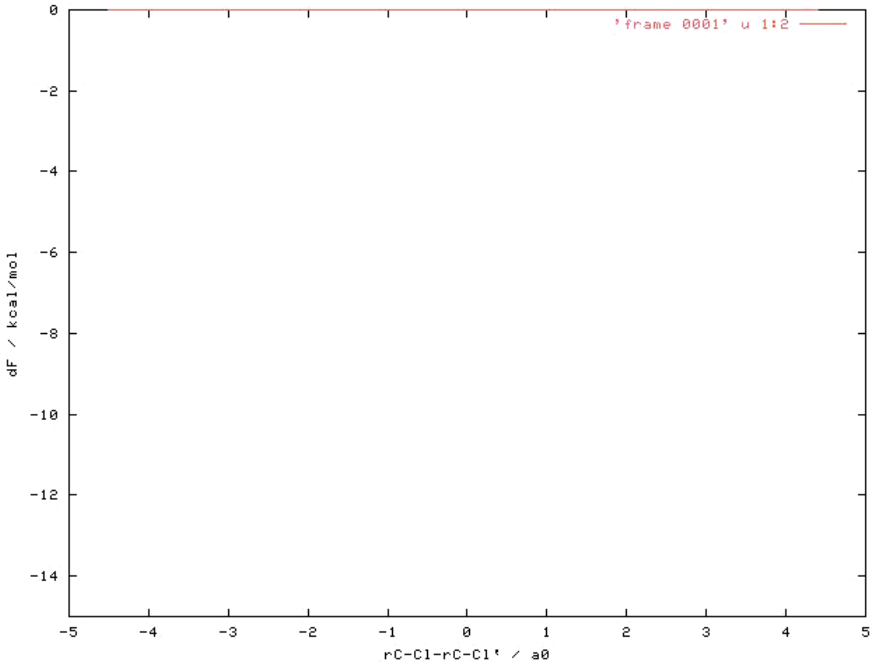
$S(R) = r_{C\text{-}Cl} - r_{C\text{-}Cl`}$

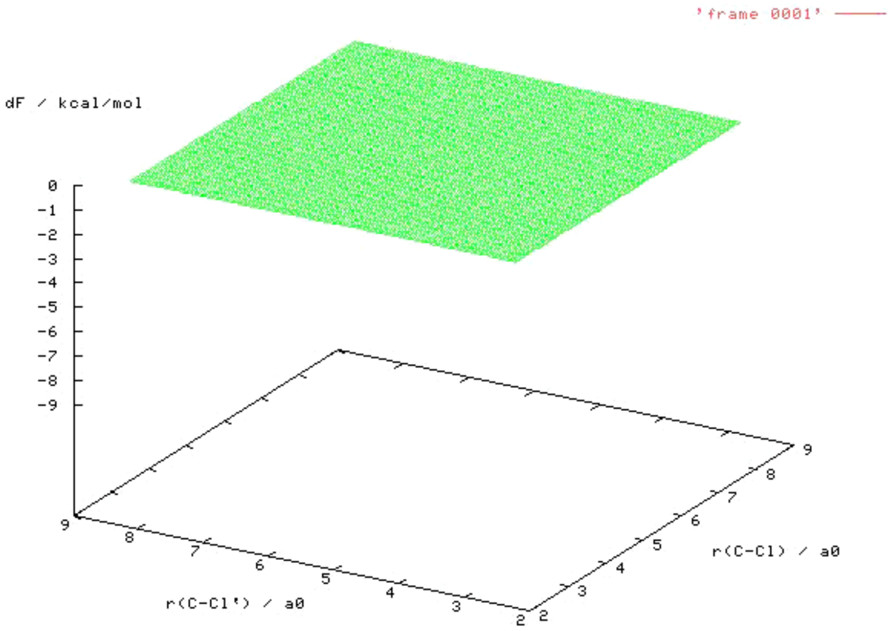

Reactant Complex

Transition State

Product Complex

$S_1(R) = r_{C\text{-}Cl}$
$S_2(R) = r_{C\text{-}Cl`}$



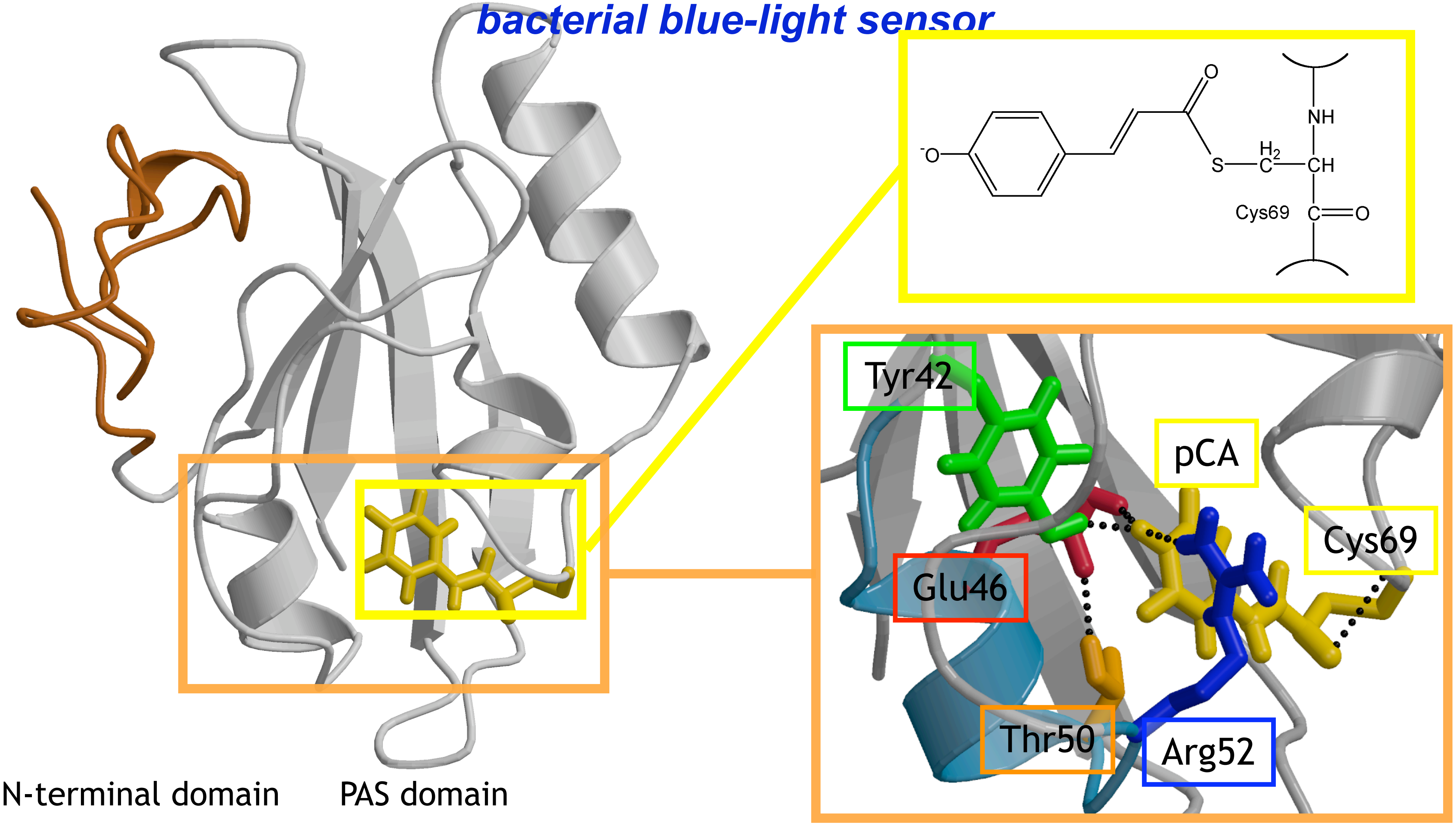


Meta-dynamics can relax the requirement of choosing a good reaction coordinate

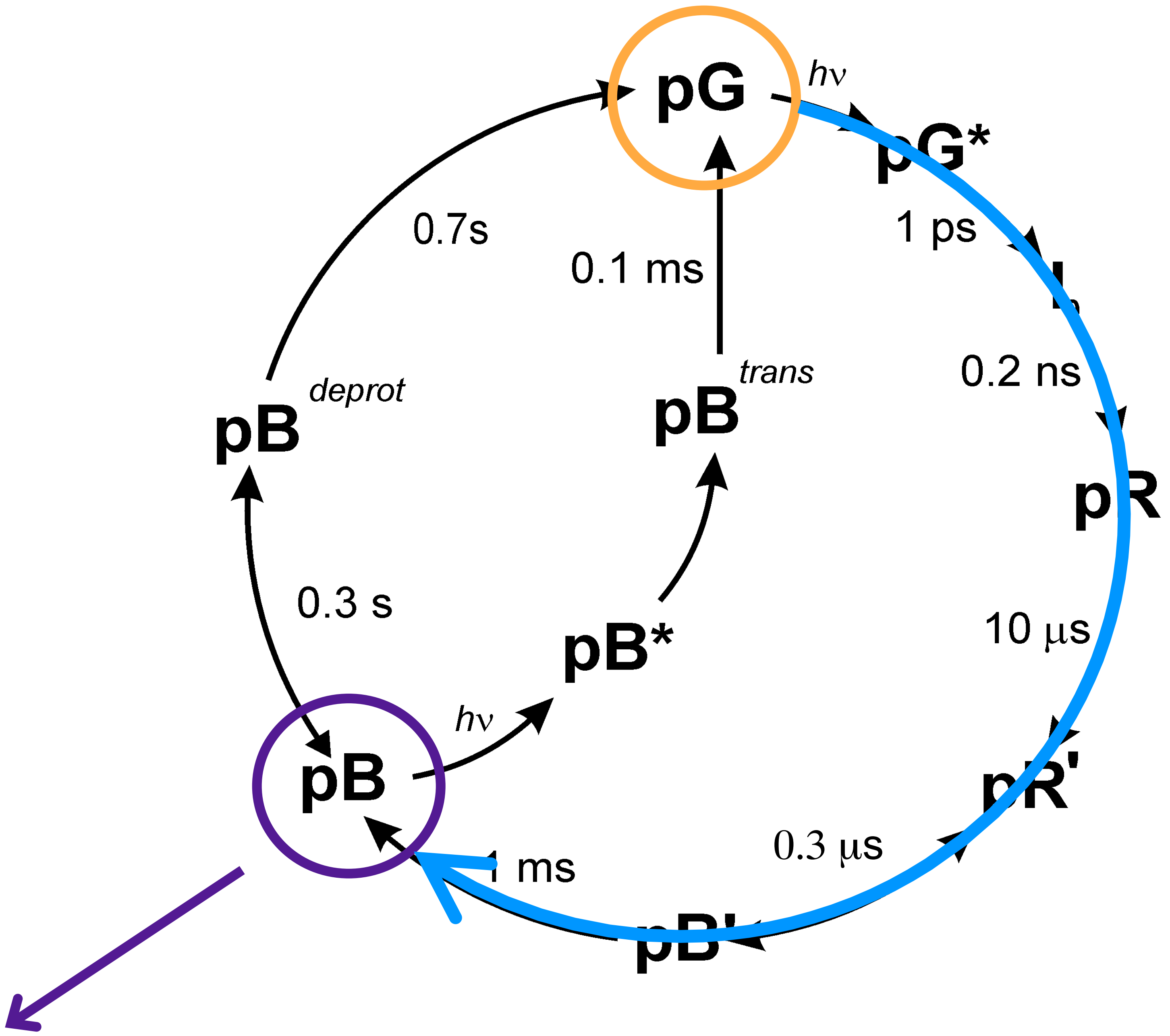
Bernd Ensing, Alessandro Laio, Michele Parrinello and Michael L. Klein
J. Phys. Chem. B **109** (2005), 6676-6687

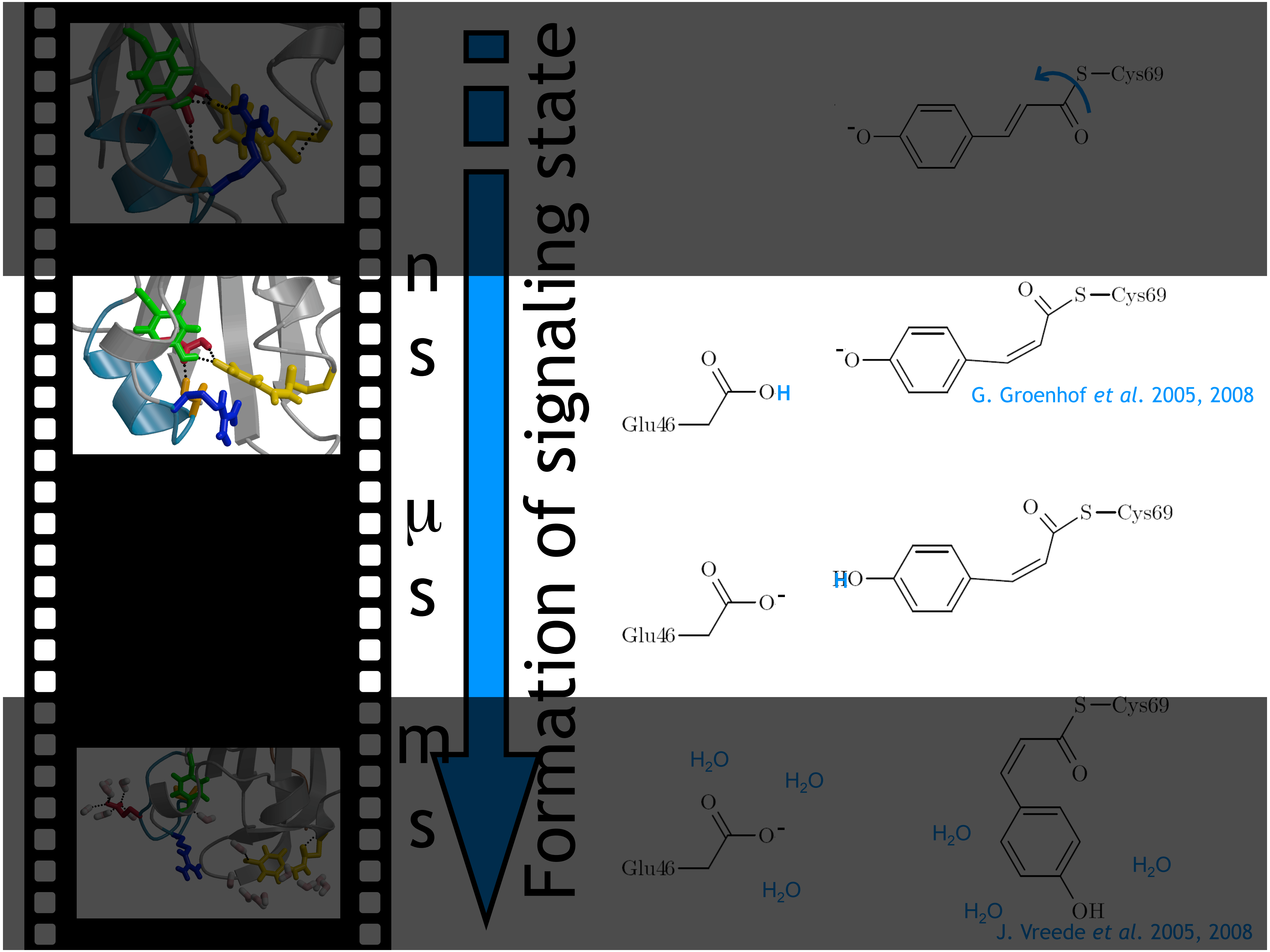
Photoactive Yellow Protein

bacterial blue-light sensor



Absorption of a blue-light photon triggers the photo cycle


pG
hν
pG*
1 ps
I
0.2 ns
pR
10 μs
pR'
0.3 μs
pB'
1 ms
pB
hν
pB*
pB
trans
0.1 ms
pB
deprot
0.3 s
0.7s



Formation of signaling state

ns

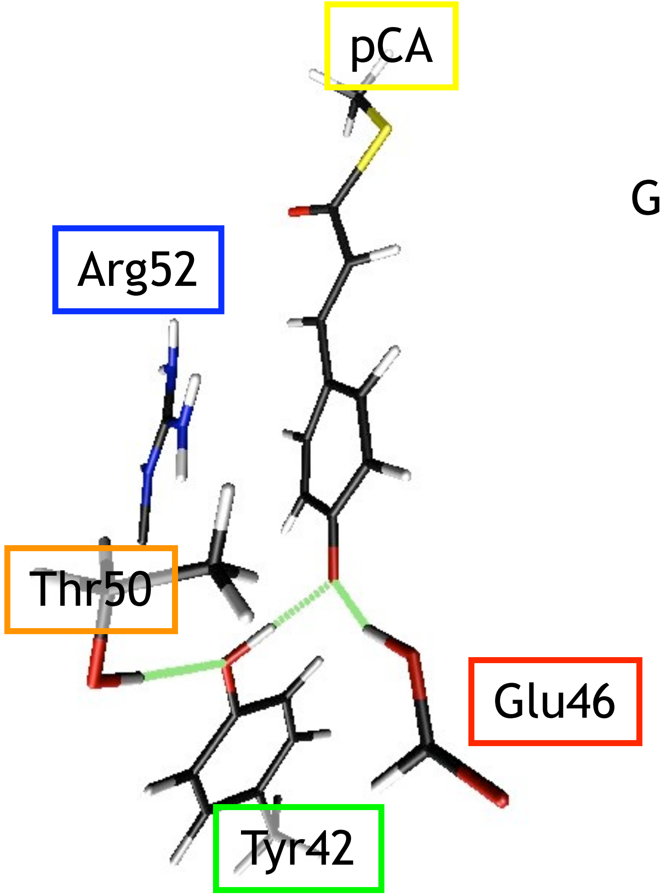
µs

ms

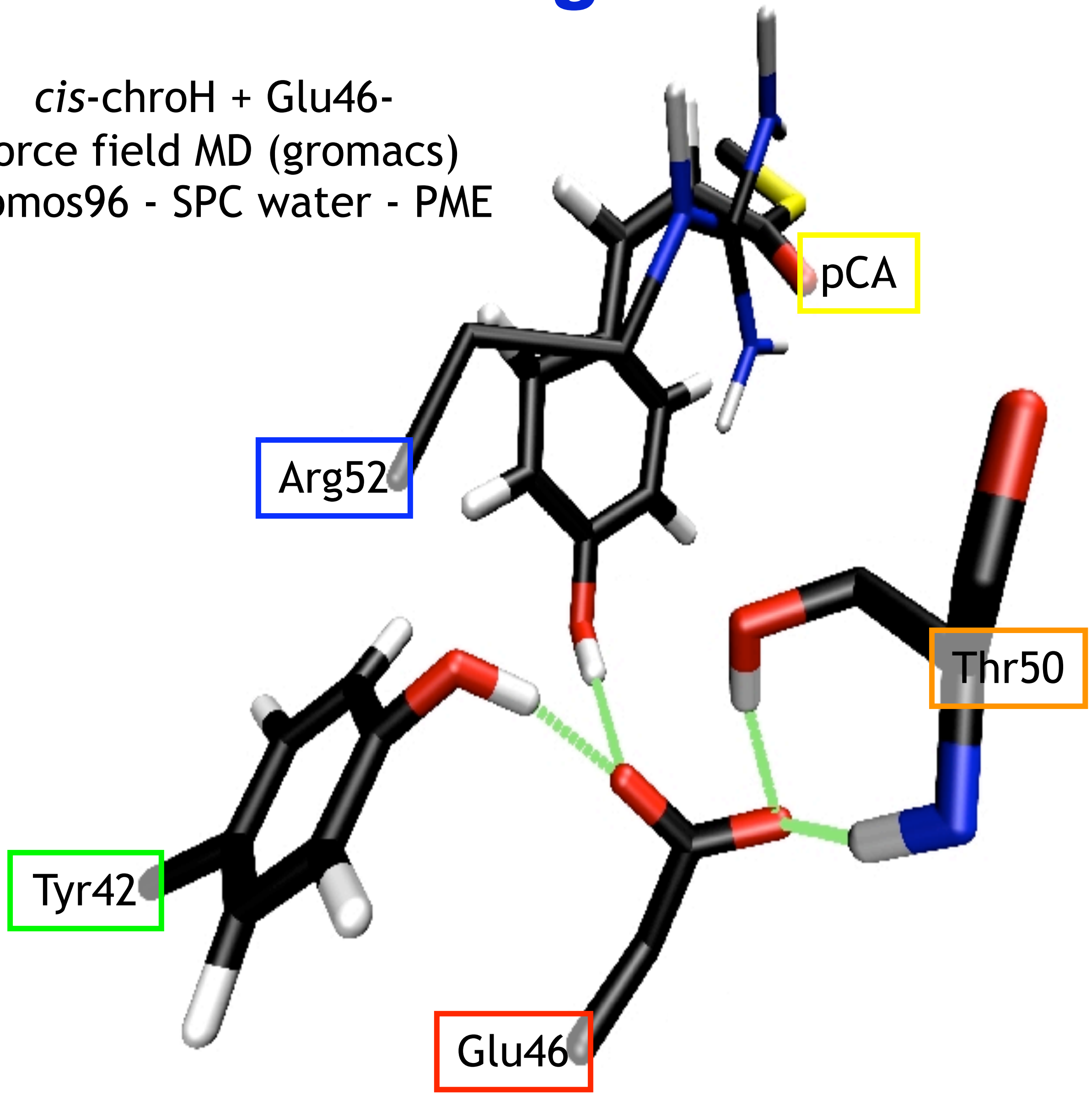
G. Groenhof *et al*. 2005, 2008

J. Vreede *et al*. 2005, 2008

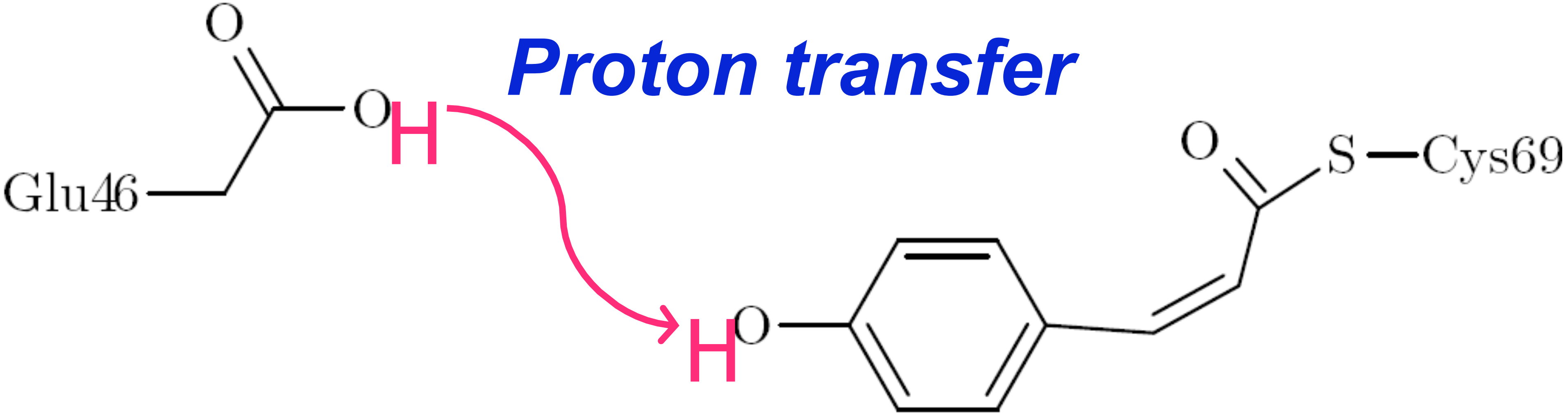
Hydrogen bond rearrangements



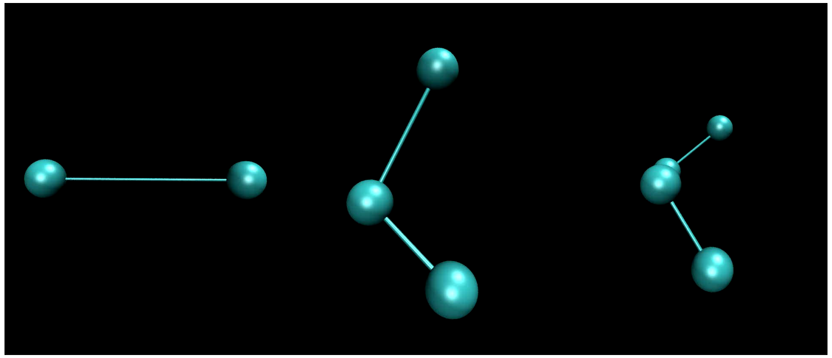
cis-chroH + Glu46-
Force field MD (gromacs)
Gromos96 - SPC water - PME



Proton transfer

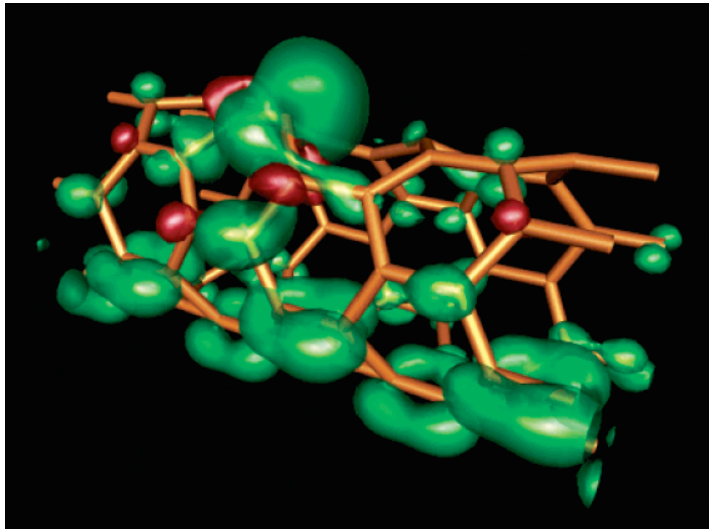


force field



bonded interactions are fixed

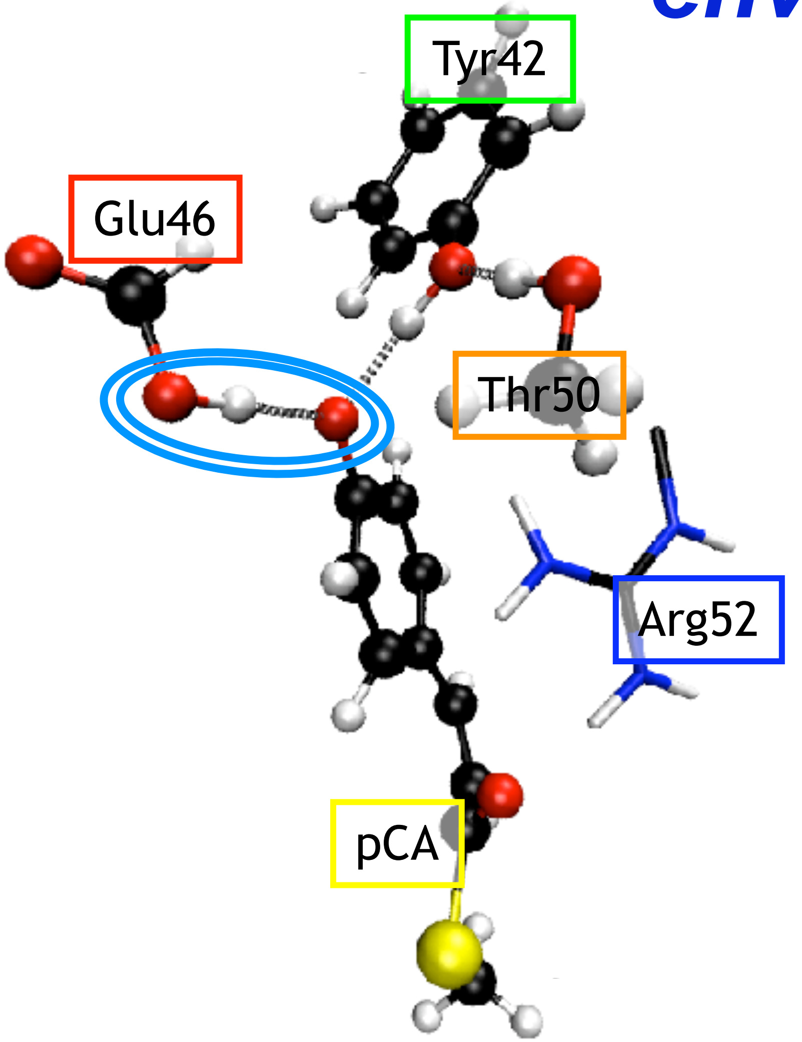
quantum (DFT)



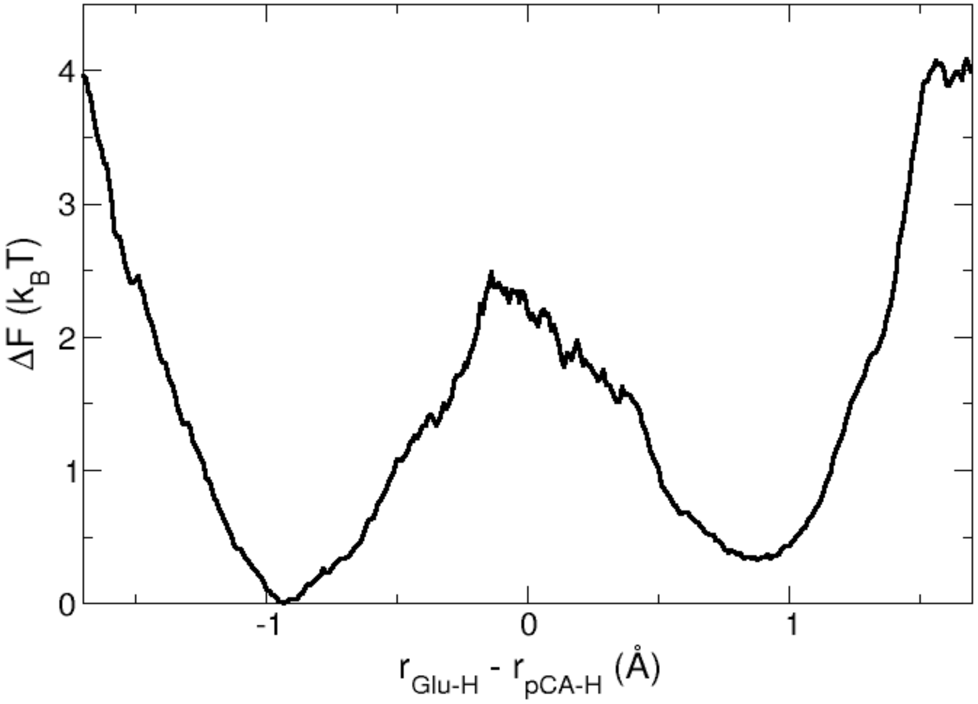
calculate electron density

bonds can change

Proton transfer in a protein-like environment



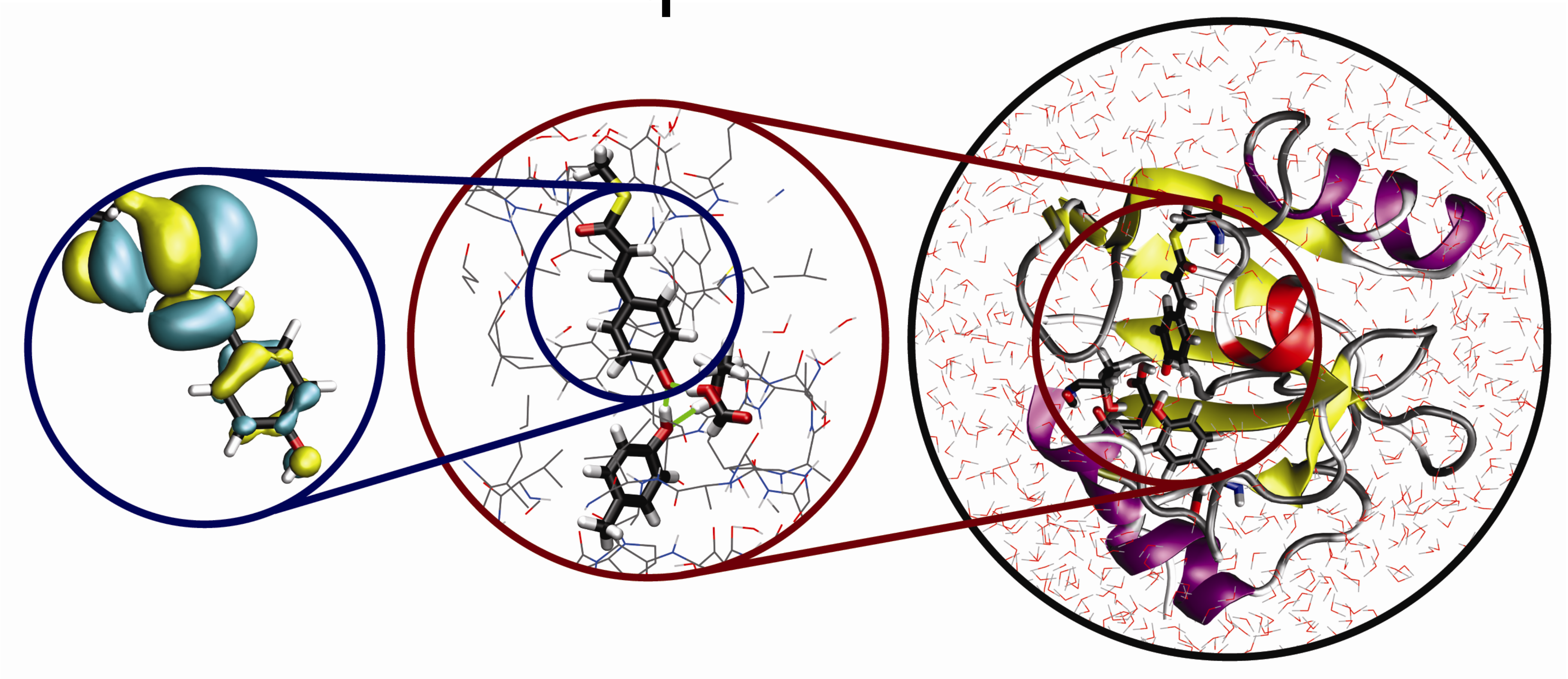
Car-Parrinello MD
DFT-based, NVT, 300K,
cg-BLYP, time step 0.1 fs



E.J.M. Leenders et al.*J. Phys. Chem. B* **2007**

Proton transfer in the protein

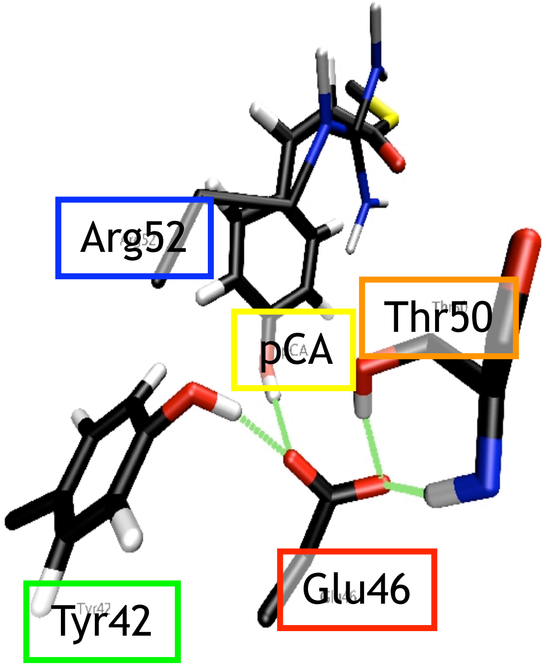
qm/mm



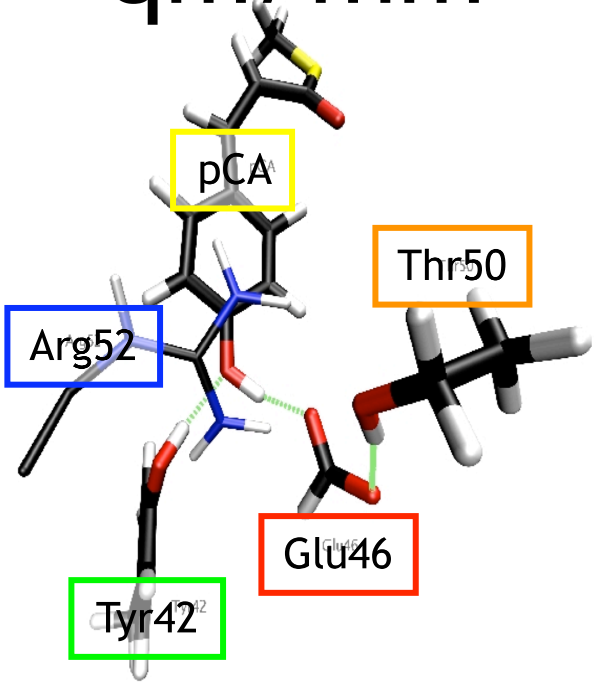
No proton transfer!

Hydrogen bond rearrangements

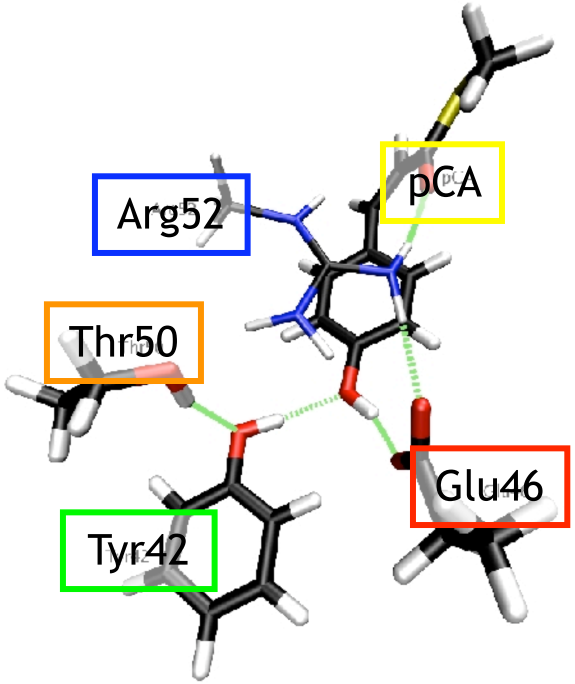
Force field



qm/mm

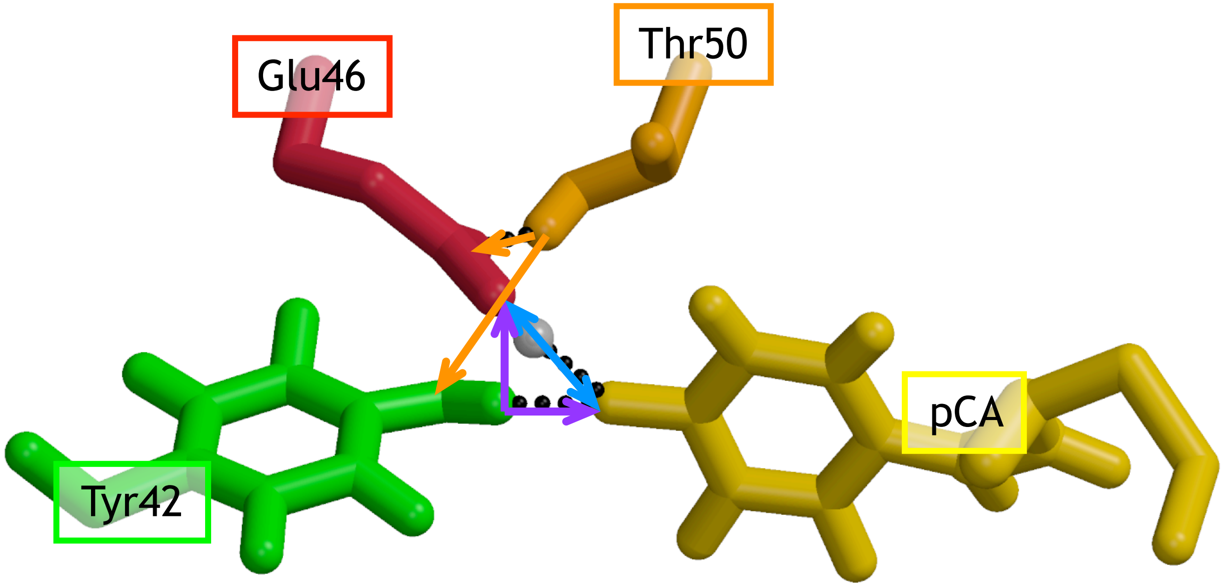


DFT

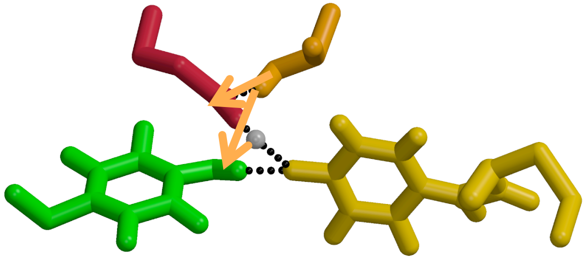


Metadynamics in QMMM

Shift Thr50

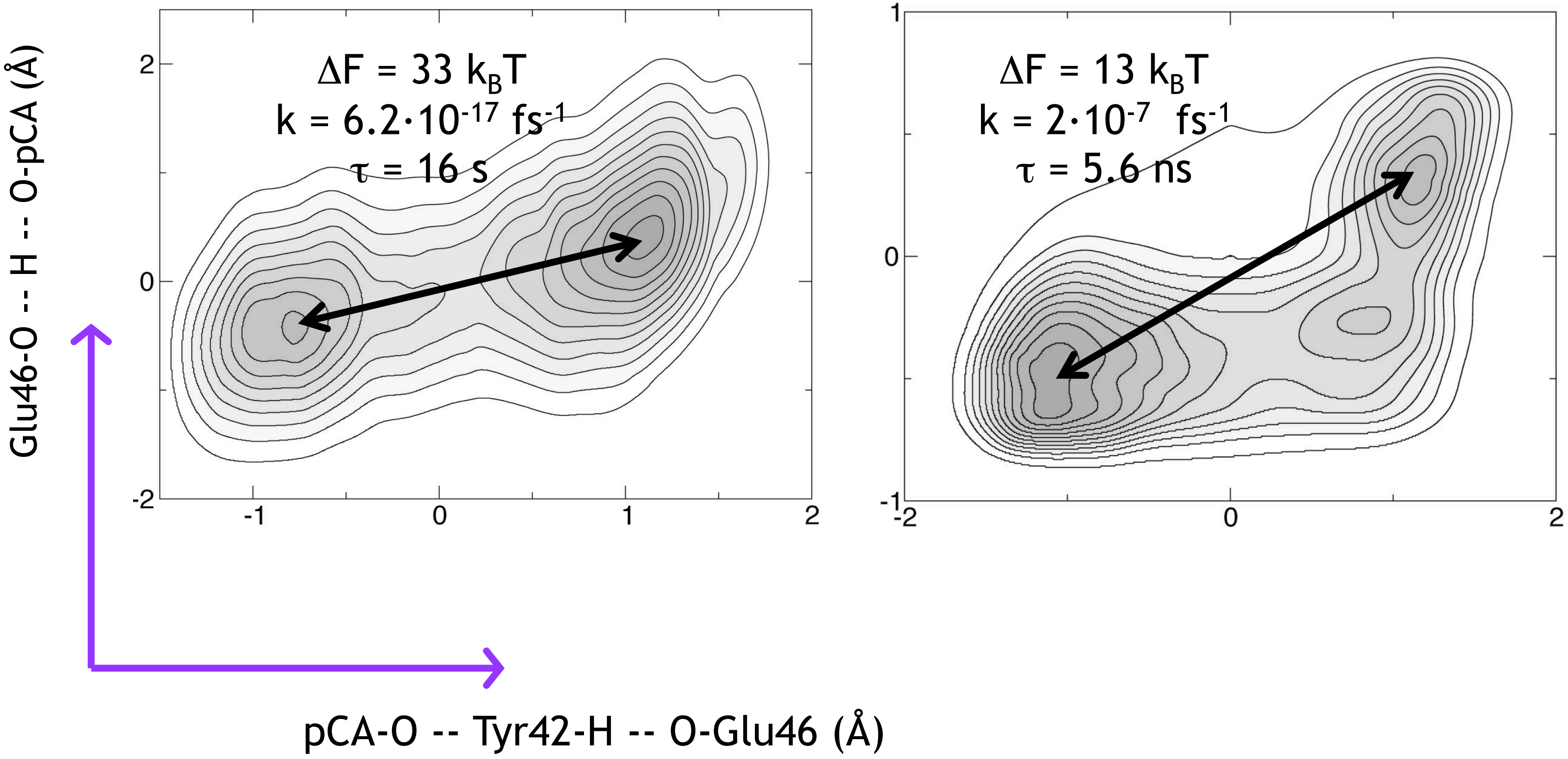


Free energy profiles



$k = \nu \exp(-\Delta F/RT)$

$\tau = 1/k$



$\Delta F = 33\ k_BT$

$k = 6.2 \cdot 10^{-17}\ fs^{-1}$

$\tau = 16\ s$

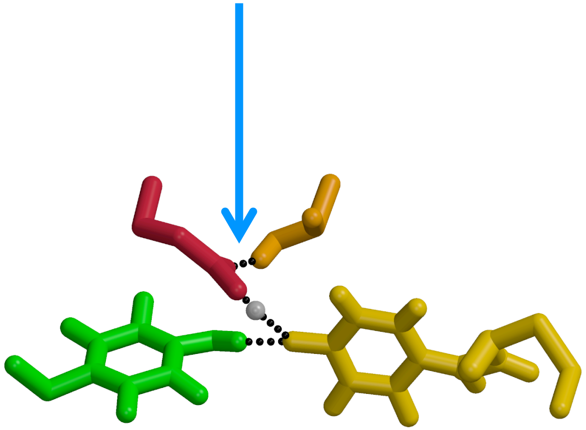
$\Delta F = 13\ k_BT$

$k = 2 \cdot 10^{-7}\ fs^{-1}$

$\tau = 5.6\ ns$

Alternative mechanism

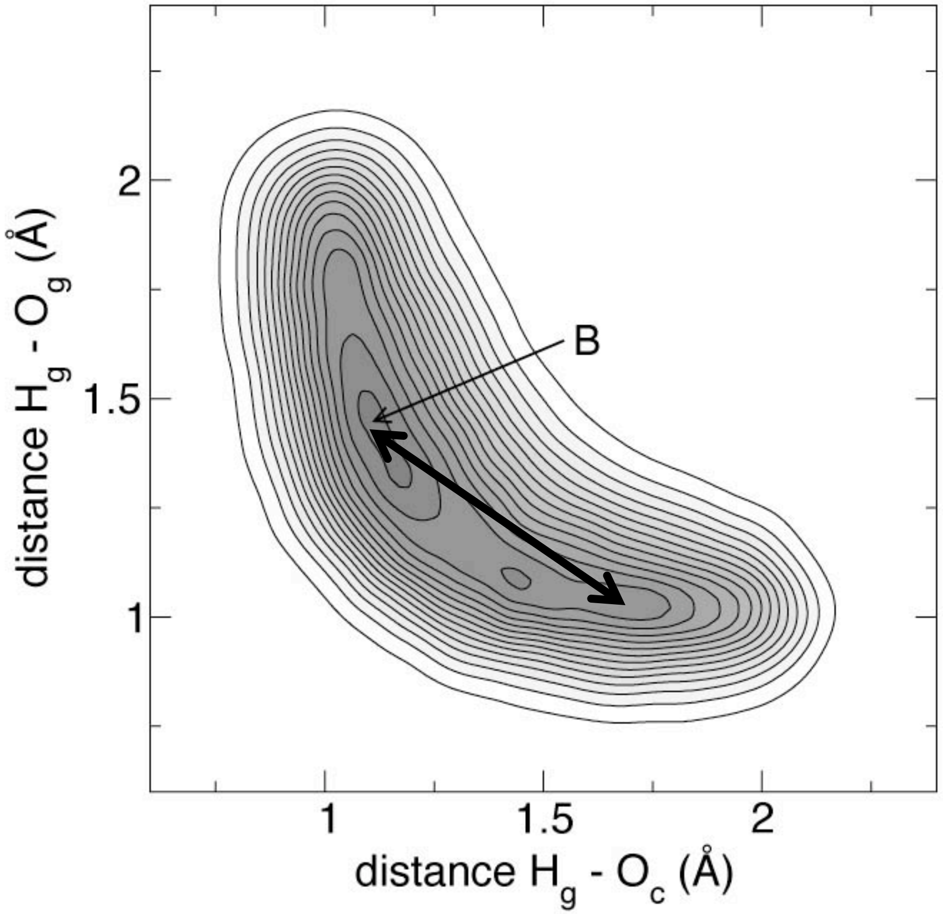
chain of water molecules



$\Delta F = 8\ k_BT$

$k = 1\ ns^{-1}$

Large error!



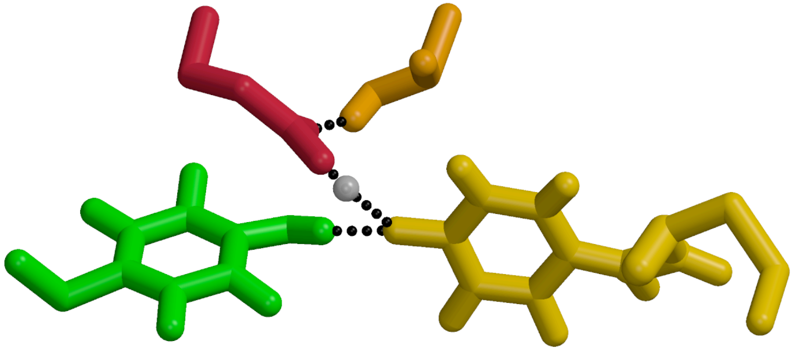
In summary

Long time scale of proton transfer is due to rearrangements of the hydrogen bond network

Charge compensation of Glu46 is very important for intramolecular proton transfer in PYP

Difference of 20 k_BT!

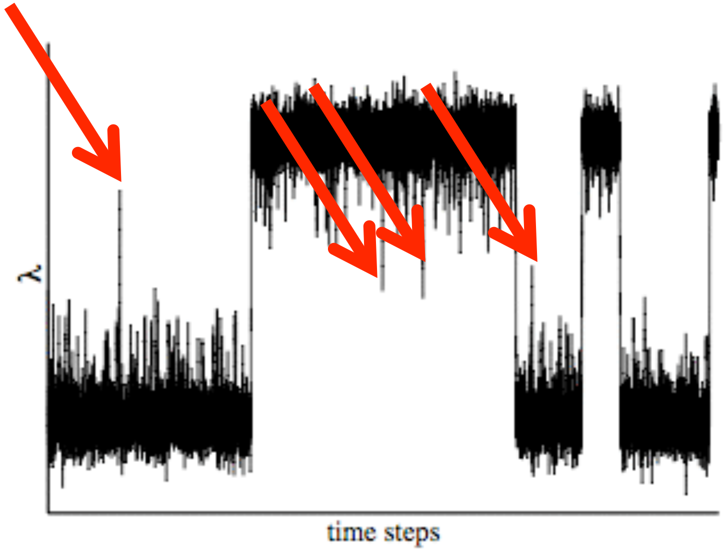
Alternative proton transfer via water

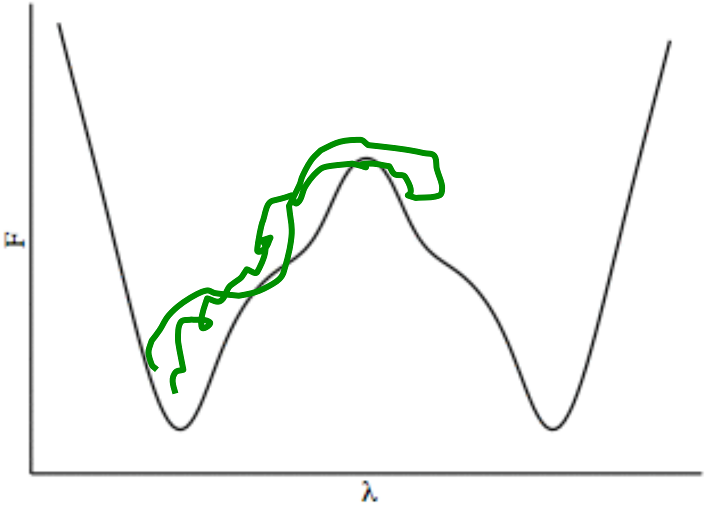


Problem with TST

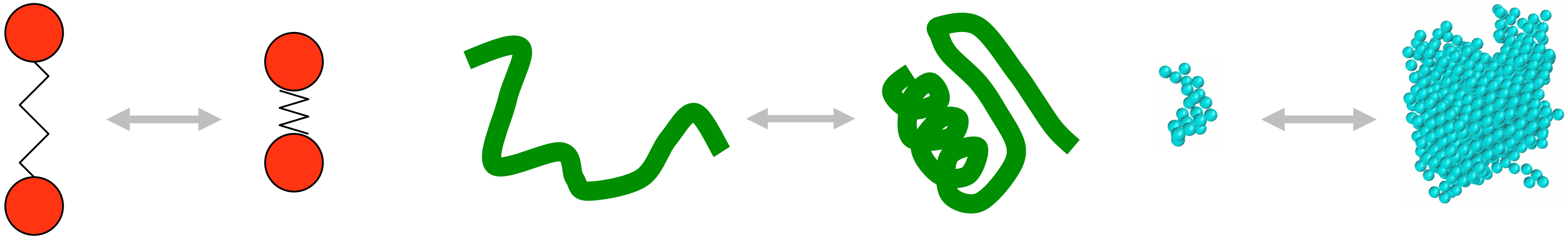
There are recrossings that cause overestimation of the rate constant

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





Back to the phenomology



A ⇌ B

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

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

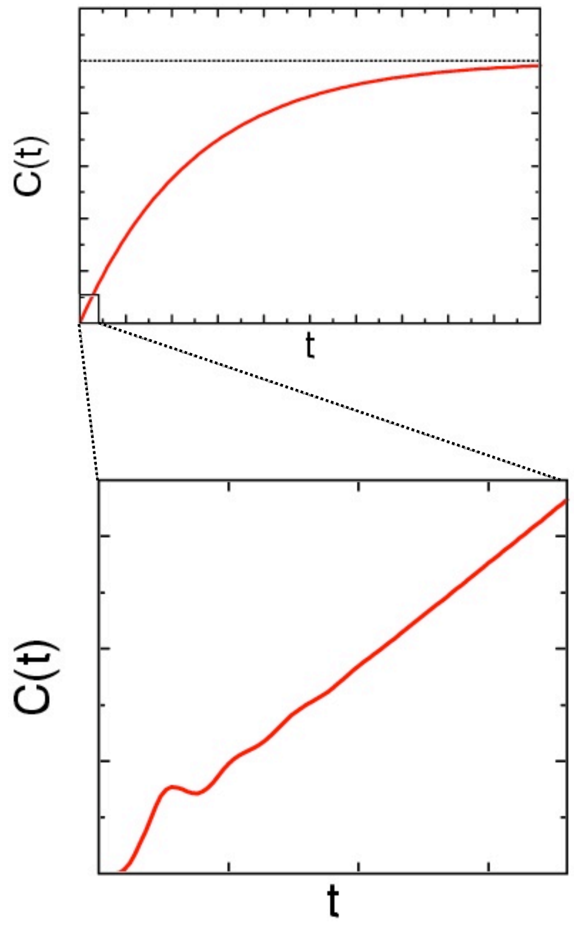
$$\boxed{\Delta c_A(t) = \Delta c_A(0) \exp(-t/\tau_{\mathrm{rxn}})}$$

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

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

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

Microscopic description



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

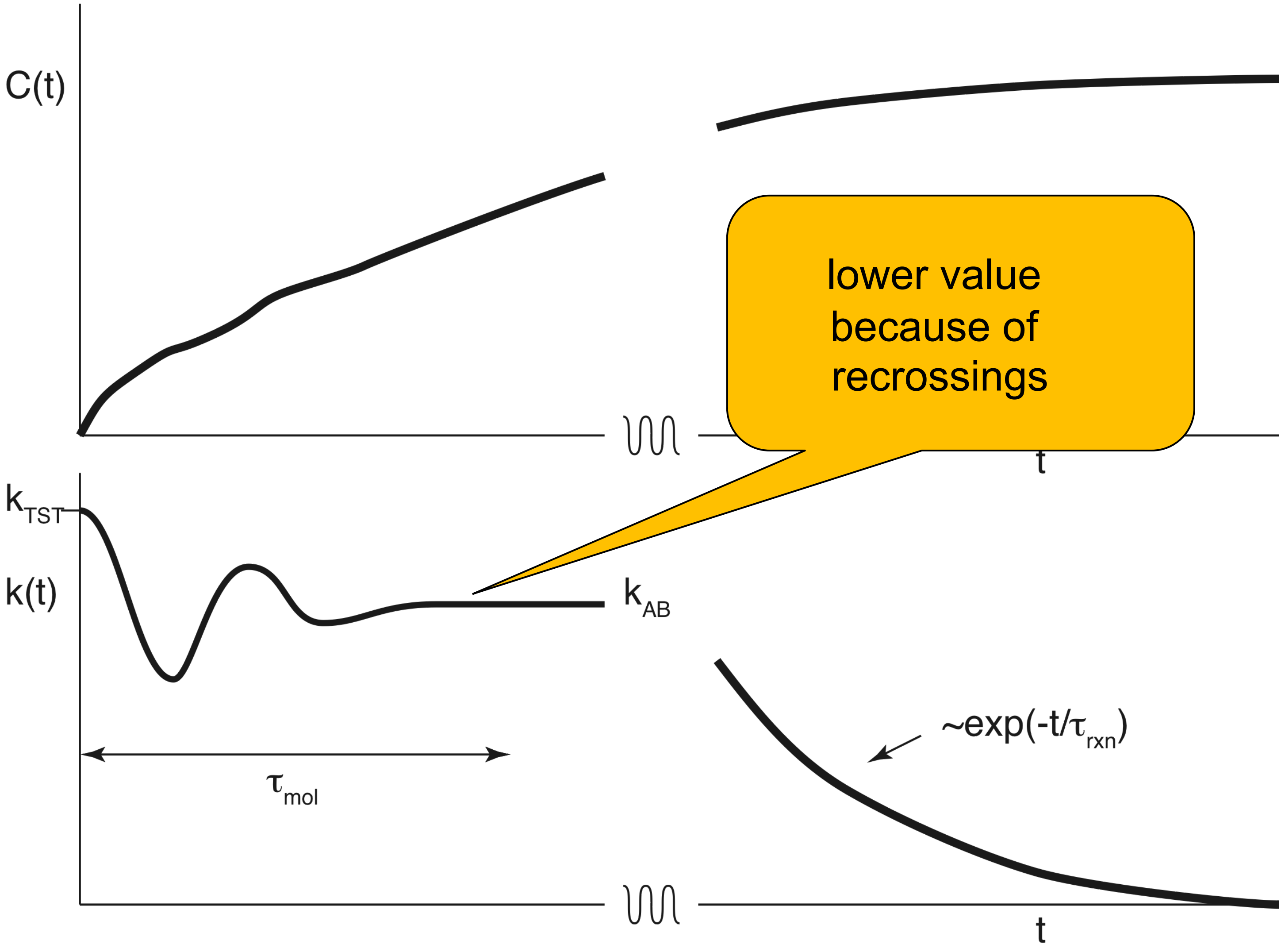
$$\boxed{\frac{\langle h_A(x_0)h_B(x_t)\rangle}{\langle h_A\rangle} = \langle h_B\rangle \left(1 - e^{-t/\tau_{\mathrm{rxn}}}\right)}$$

$$\text{for} \quad \tau_{\mathrm{mol}} < t \ll \tau_{\mathrm{rxn}}$$

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

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

Decay of correlation function



lower value because of recrossings

Bennet-Chandler approach

Start with time correlation function $C(t) \equiv \frac{\langle h_A(0) h_B(t) \rangle}{\langle h_A \rangle}$

use time invariance $\langle h_A(0) h_B(t) \rangle = \langle h_A(-t) h_B(0) \rangle$

take time derivative

$$\frac{d}{dt}\langle h_A(0) h_B(t) \rangle = \frac{d}{dt}\langle h_A(-t) h_B(0) \rangle =$$
$$-\langle \dot{h}_A(-t) h_B(0) \rangle = -\langle \dot{h}_A(0) h_B(t) \rangle$$

apply h function definition

$$h_A(r) = \theta(\lambda^* - \lambda(r))$$
$$h_B(r) = \theta(\lambda(r) - \lambda^*)$$

Bennet-Chandler approach

$$\dot{C}(t) = \frac{\langle \delta(\lambda^* - \lambda_0)\dot{\lambda}_0 \theta(\lambda_t - \lambda^*)\rangle}{\langle \theta(\lambda^* - \lambda_0)\rangle}$$

Multiply and divide by $\delta(\lambda^* - \lambda_0)$

$$\begin{aligned}
\dot{C}(t) &= \frac{\langle \delta(\lambda^* - \lambda_0)\dot{\lambda}_0 \theta(\lambda_t - \lambda^*)\rangle}{\langle \delta(\lambda^* - \lambda_0)\rangle} \frac{\langle \delta(\lambda^* - \lambda_0)\rangle}{\langle \theta(\lambda^* - \lambda_0)\rangle} \\
&= \langle \dot{\lambda}_0 \theta(\lambda_t - \lambda^*)\rangle_{\lambda_0 = \lambda^*} \frac{e^{-\beta F(\lambda^*)}}{\langle \theta(\lambda^* - \lambda_0)\rangle} \\
&\equiv R(t) \frac{e^{-\beta F(\lambda^*)}}{\langle \theta(\lambda^* - \lambda_0)\rangle}
\end{aligned}$$

Comparison with TST

$$\dot{C}(t) = \frac{\langle \delta(\lambda^* - \lambda_0)\dot{\lambda}_0\theta(\lambda_t - \lambda^*)\rangle}{\langle \delta(\lambda^* - \lambda_0)\rangle}\frac{\langle \delta(\lambda^* - \lambda_0)\rangle}{\langle \theta(\lambda^* - \lambda_0)\rangle}$$

Conditional average: $\dot{\lambda}_0\theta(\lambda_t - \lambda^*)$
given that we start on top of the barrier

$$k^{T}_{AB} \qquad \frac{\qquad}{\langle \delta(\lambda(r) - \lambda^*)\rangle} \frac{\delta(\lambda(r) - \lambda^*}{\langle \theta(\lambda^* -}$$

Probability to find λ on top of the barrier

$$\begin{aligned}\dot{C}(t) &= k^{TST}_{AB}\frac{\langle \dot{\lambda}_0\theta(\lambda_t - \lambda^*)\rangle_{\lambda_0=\lambda^*}}{\langle \dot{\lambda}_0\theta(\dot{\lambda}_0)\rangle_{\lambda_0=\lambda^*}} \\ &\equiv k^{TST}_{AB}\,\kappa(t)\end{aligned}$$

κ(t) is the transmission coefficient, accounting for recrossings

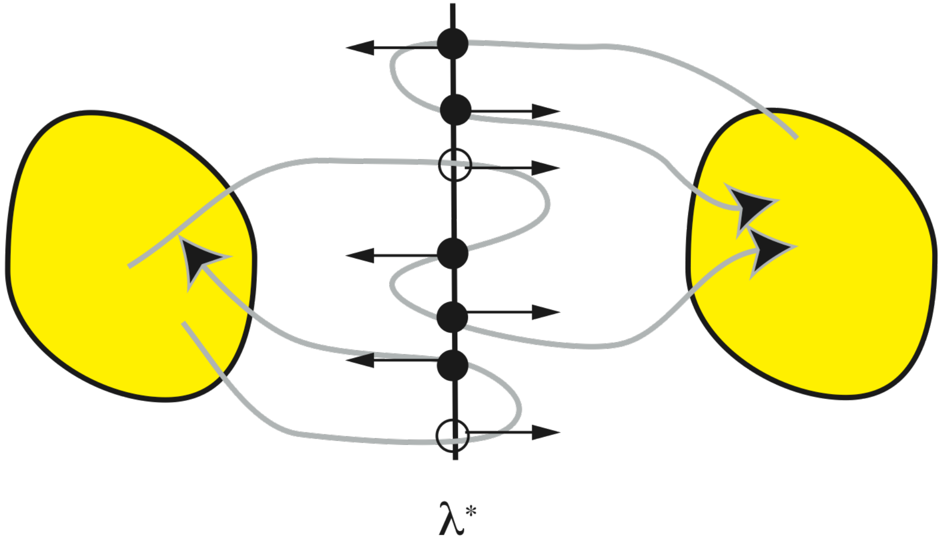
BC Computational scheme

Determine TST rate the probability from the free energy

$$k_{AB}^{TST} = \langle \dot{\lambda}\theta(\dot{\lambda})\rangle_{\lambda=\lambda^*} \frac{e^{-\beta F(\lambda^*)}}{\int_{-\infty}^{\lambda^*} e^{-\beta F(\lambda')} d\lambda'}$$

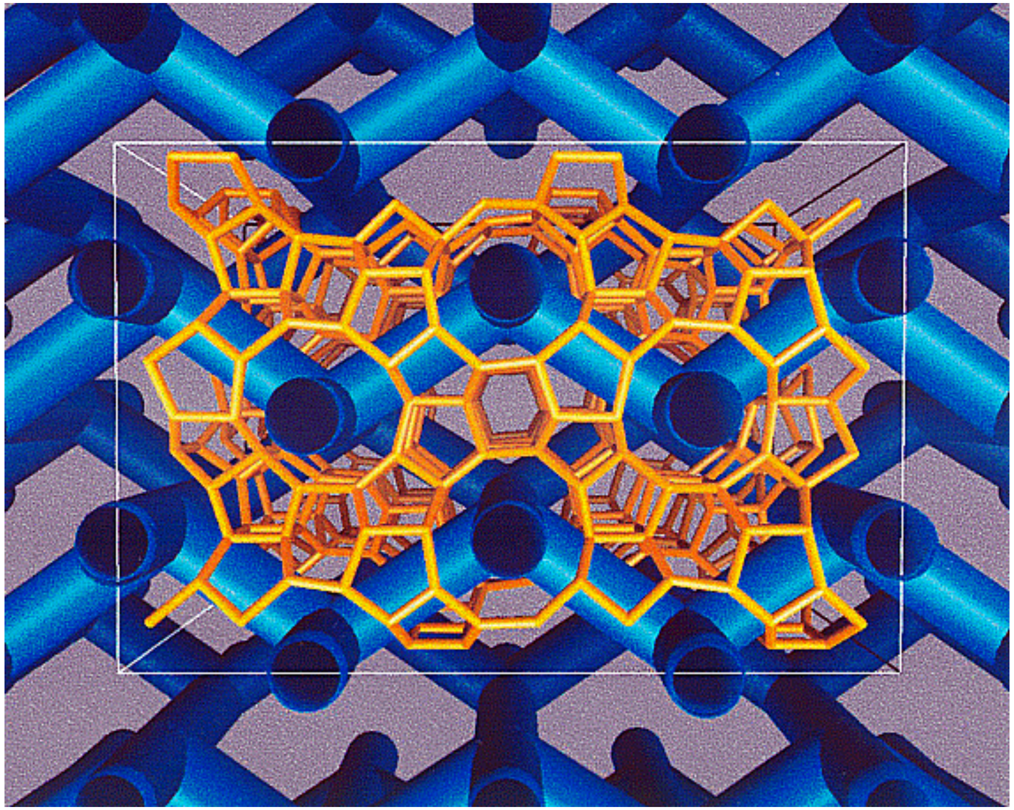
Compute the conditional average from a MD simulation

$$\kappa(t) = \frac{\langle \dot{\lambda}_0 \theta(\lambda_t - \lambda^*)\rangle_{\lambda_0=\lambda^*}}{\langle \dot{\lambda}_0 \theta(\dot{\lambda}_0)\rangle_{\lambda_0=\lambda^*}}$$



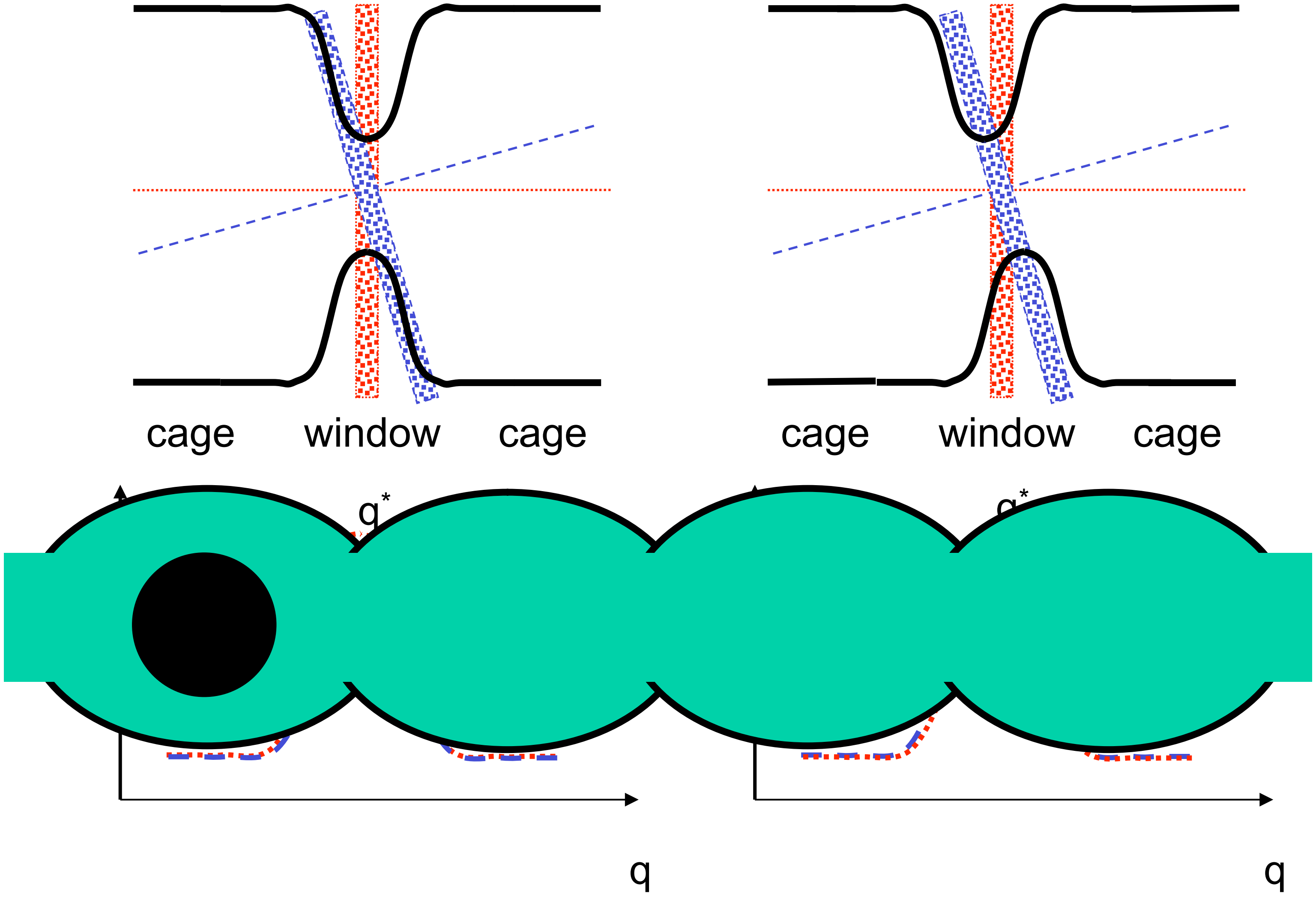
Example diffusion in zeolite

- Zeolites imporant class of materials
- Diffusion of alkanes in matrix is poorly described
- Approach
  - molecular simulation of alkanes in fixed zeolite frame
  - Unitied atom FF by Dubbeldam et al.

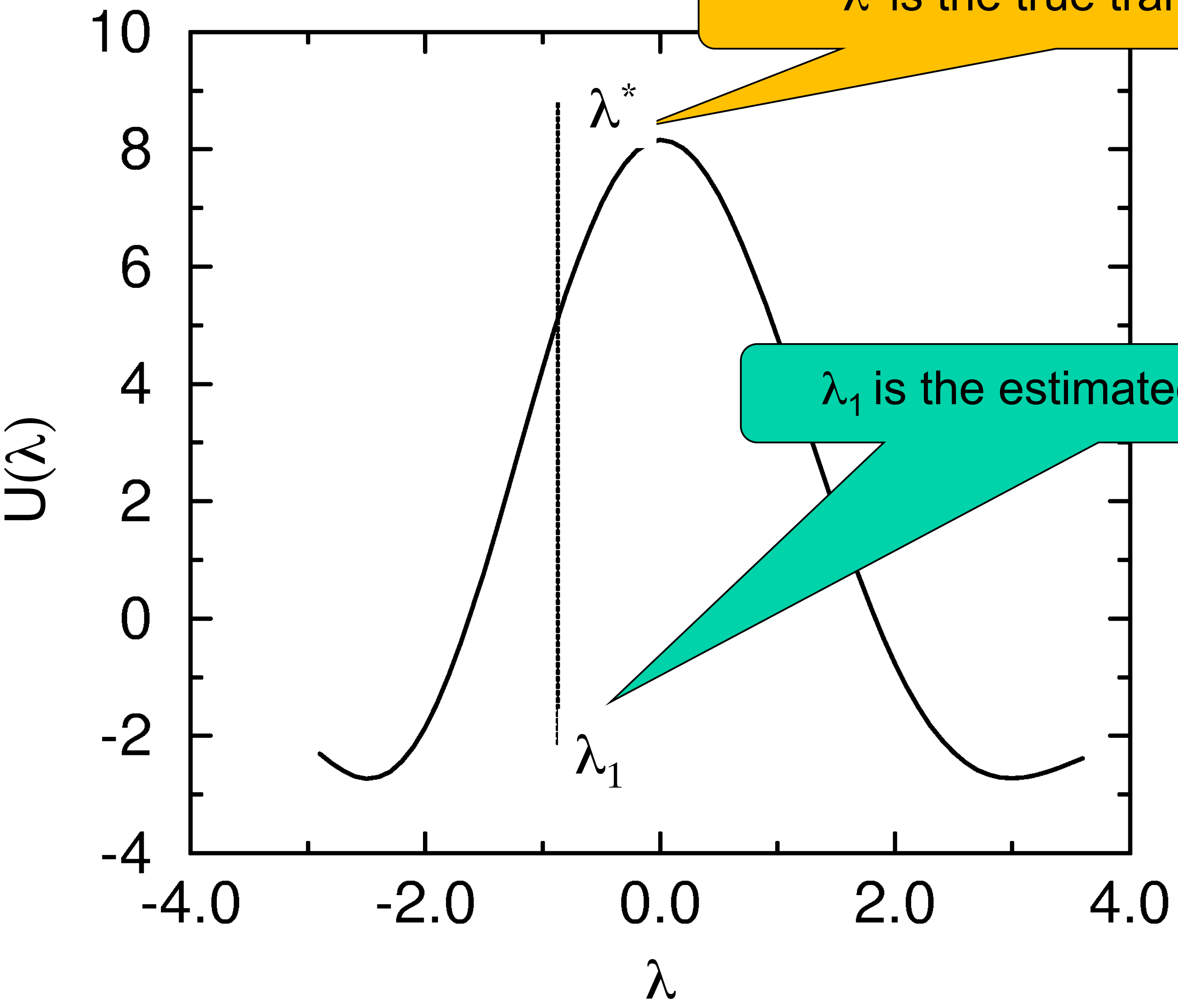


D. Dubbeldam, et al., J. Phys. Chem. B, 108, 12301, 2004

Reaction coordinate

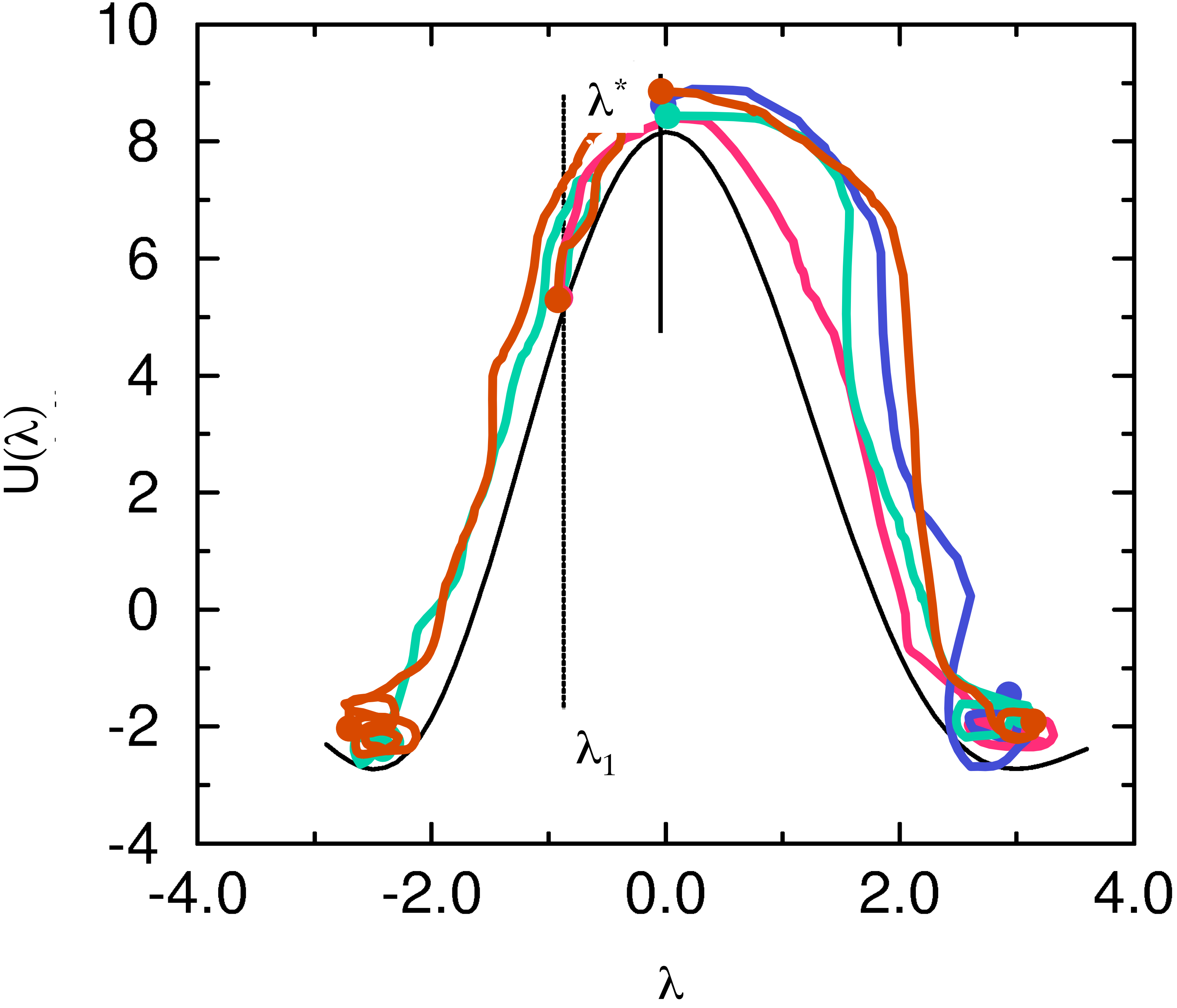


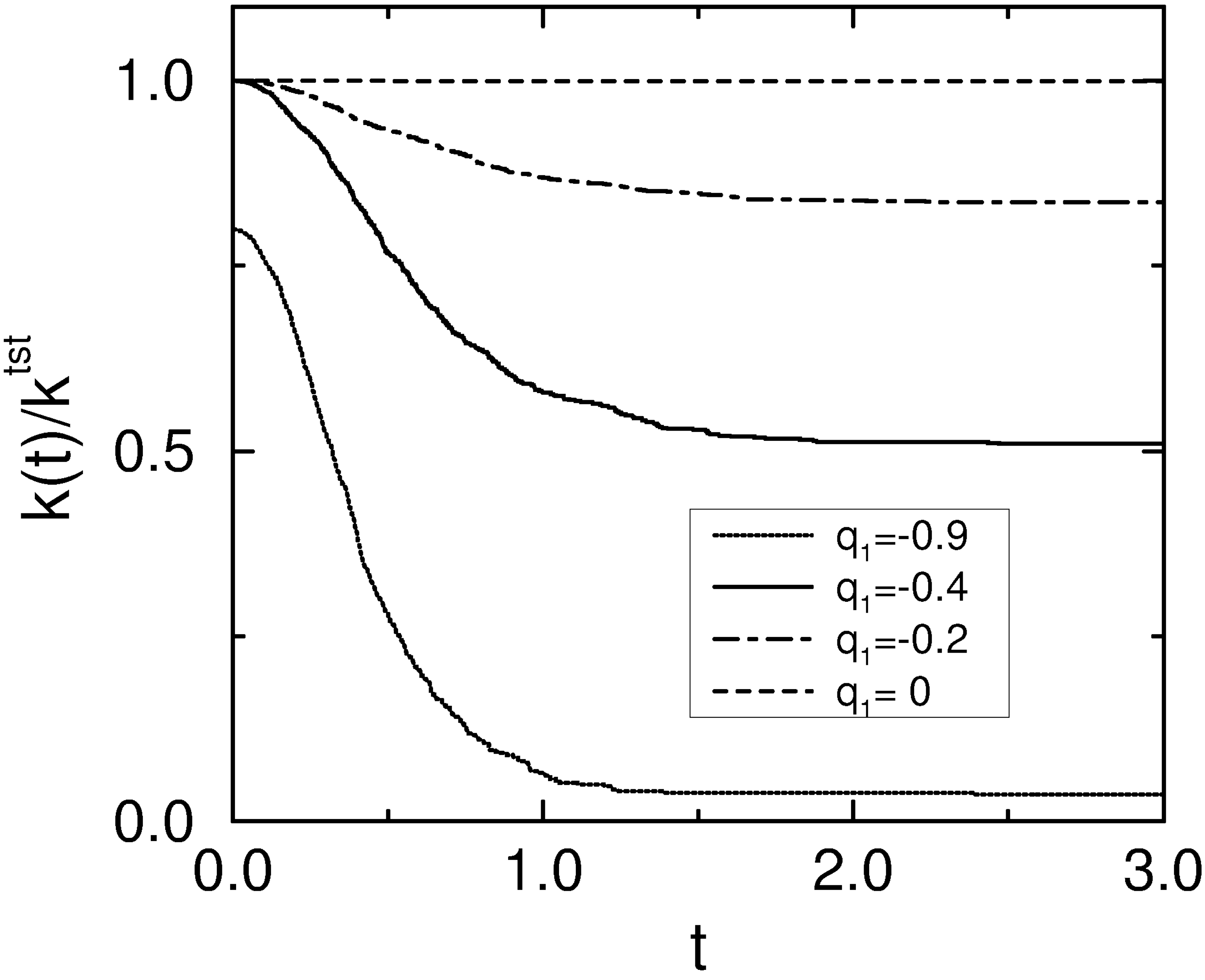
Ideal gas particle and a hill



λ^* is the true transition state

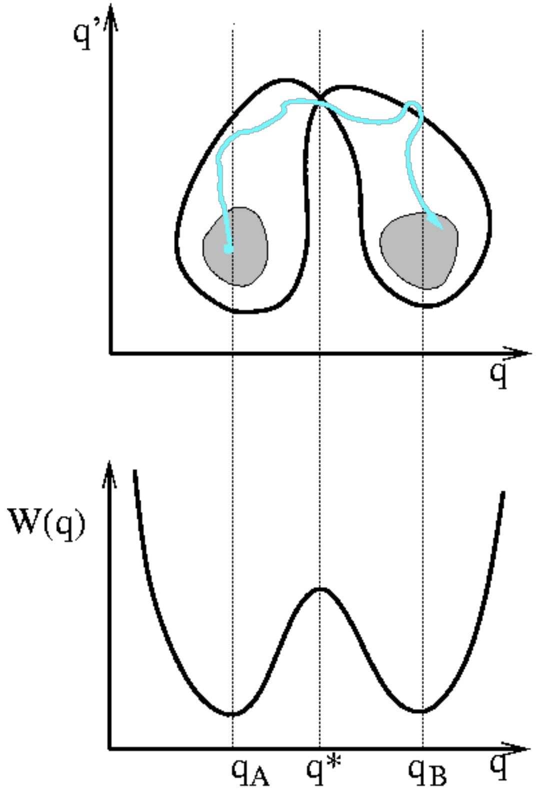
λ_1 is the estimated transition state


10
8
6
4
2
0
-2
-4
λ^*
λ_1
$U(\lambda)$
-4.0
-2.0
0.0
2.0
4.0
λ


k(t)/k^{tst}
t
0.0
0.5
1.0
0.0
1.0
2.0
3.0
q1=-0.9
q1=-0.4
q1=-0.2
q1= 0

Breakdown of FE approach

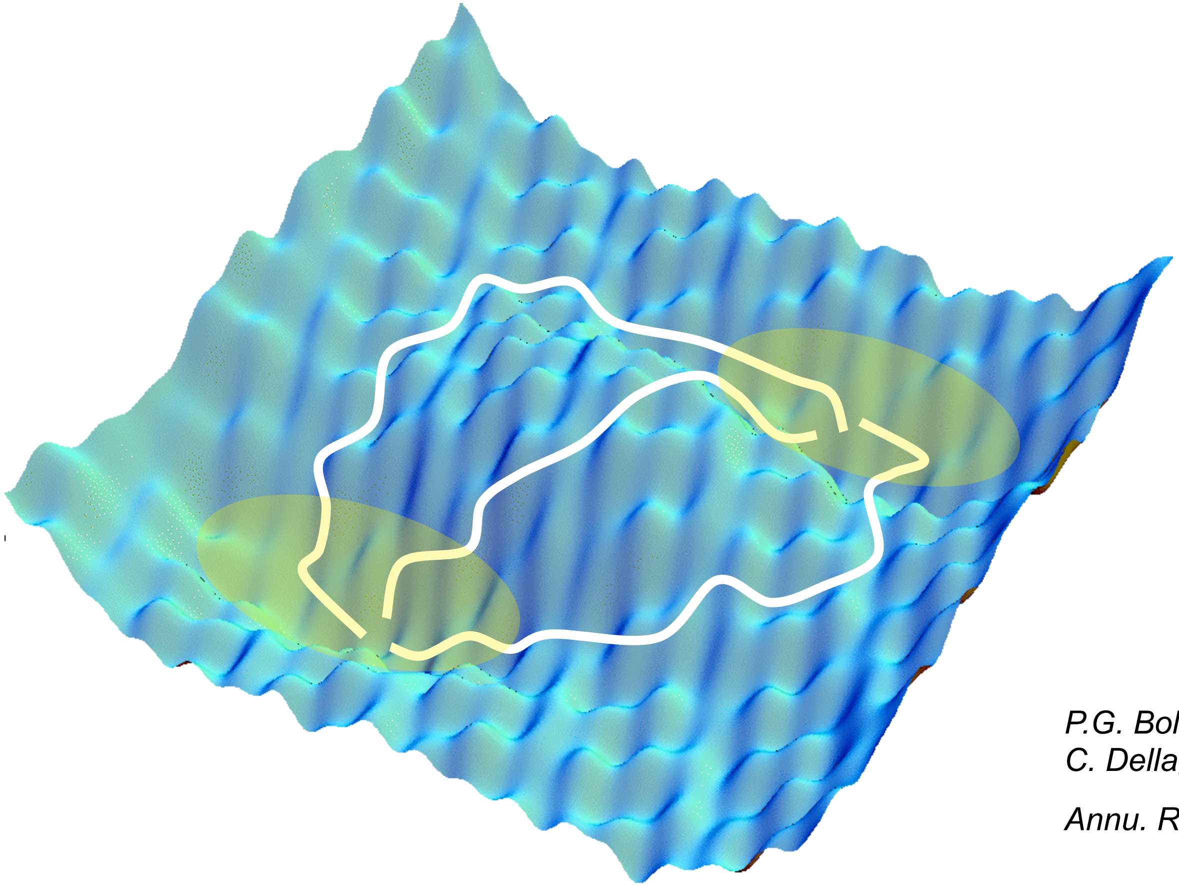
kappa can become immeasurable low if the reaction coordinate is at the wrong value the reaction coordinate is wrongly chosen



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

If the reaction coordinate is not known, the wrong order parameter can lead to wrong transition states, mechanism and rate constants

Calls for unbiased path sampling



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

Annu. Rev. Phys. Chem 2002

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