

Molecular simulation of biological processes: predicting mechanisms and free energies

Jocelyne Vreede

Computational Chemistry
Van 't Hoff Institute for Molecular Sciences
University of Amsterdam
The Netherlands

Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

Part 2: Biased sampling

Part 3: Path sampling

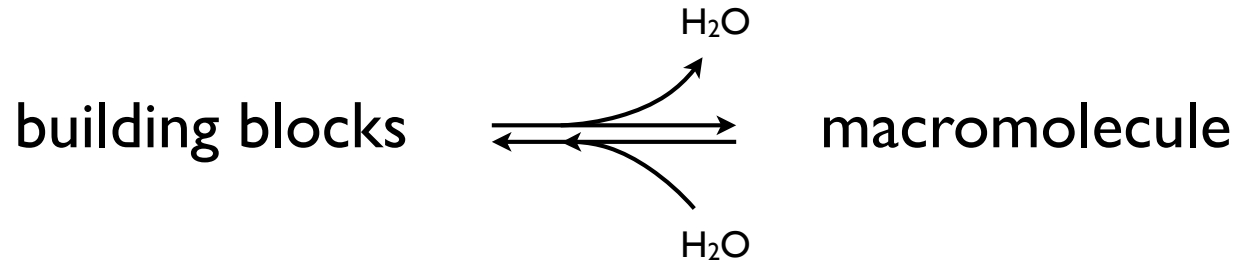
Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

Part 2: Biased sampling

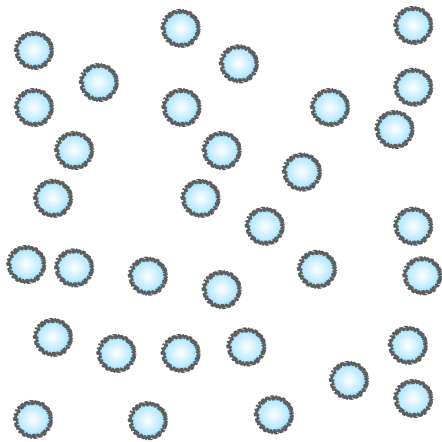
Part 3: Rare events

Biomolecules

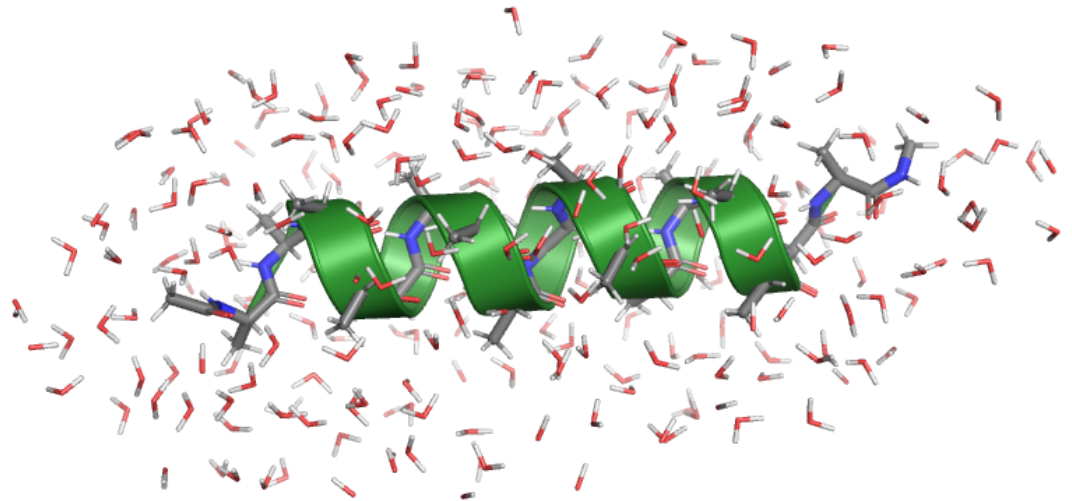


building block	macromolecule
amino acid	protein
nucleotide	nucleic acids
monosaccharide	cellulose, starch
lipid	membrane

Interactions in a biomolecular system



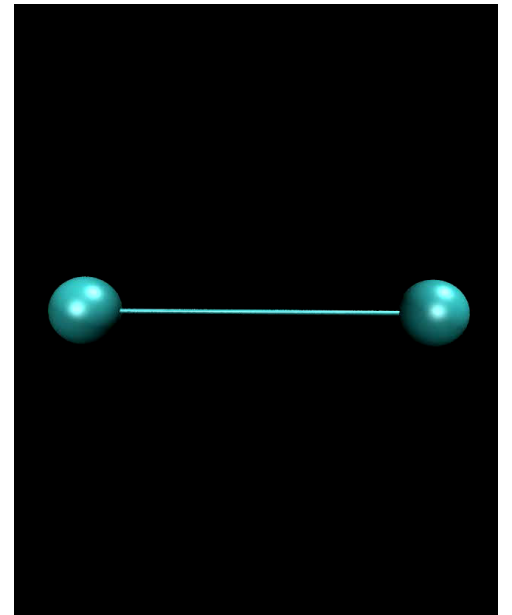
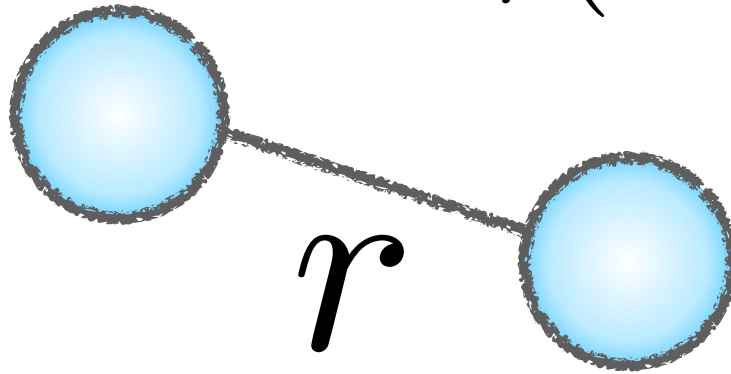
Lennard-Jones particles



α -helix in water

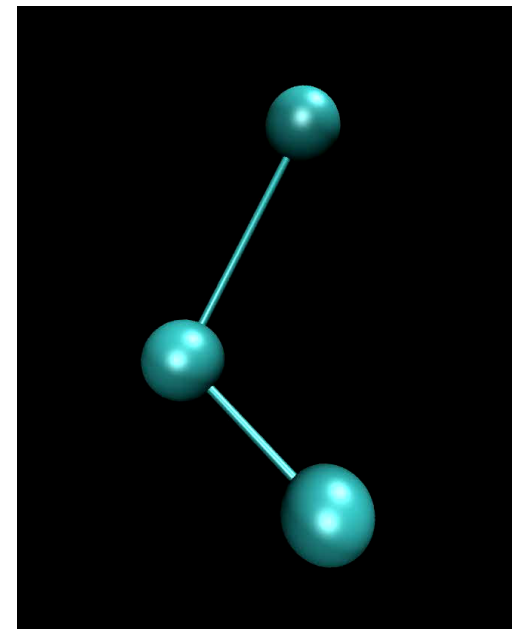
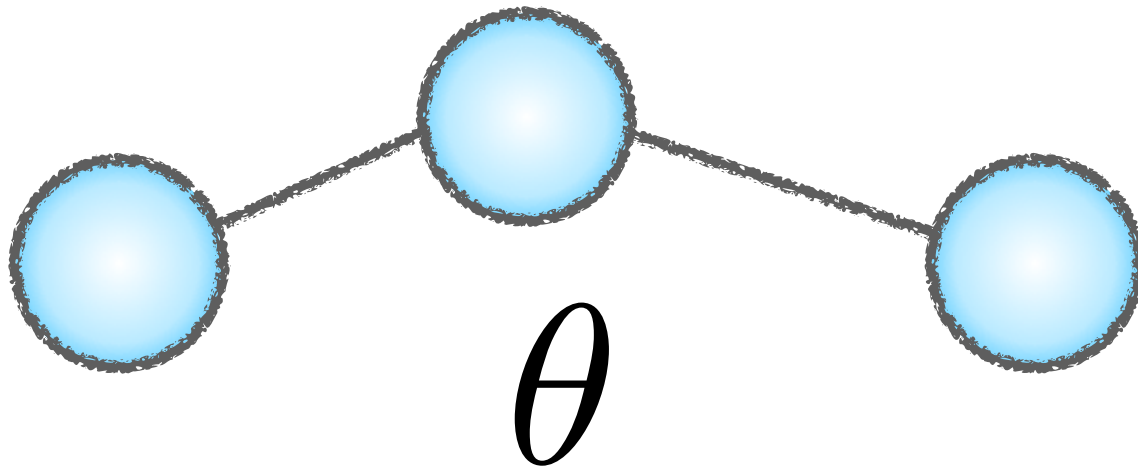
Bonded interactions

$$k_r (r - r_{eq})^2$$



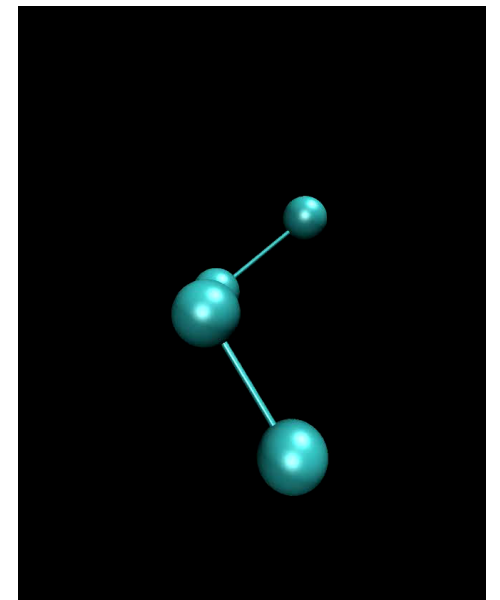
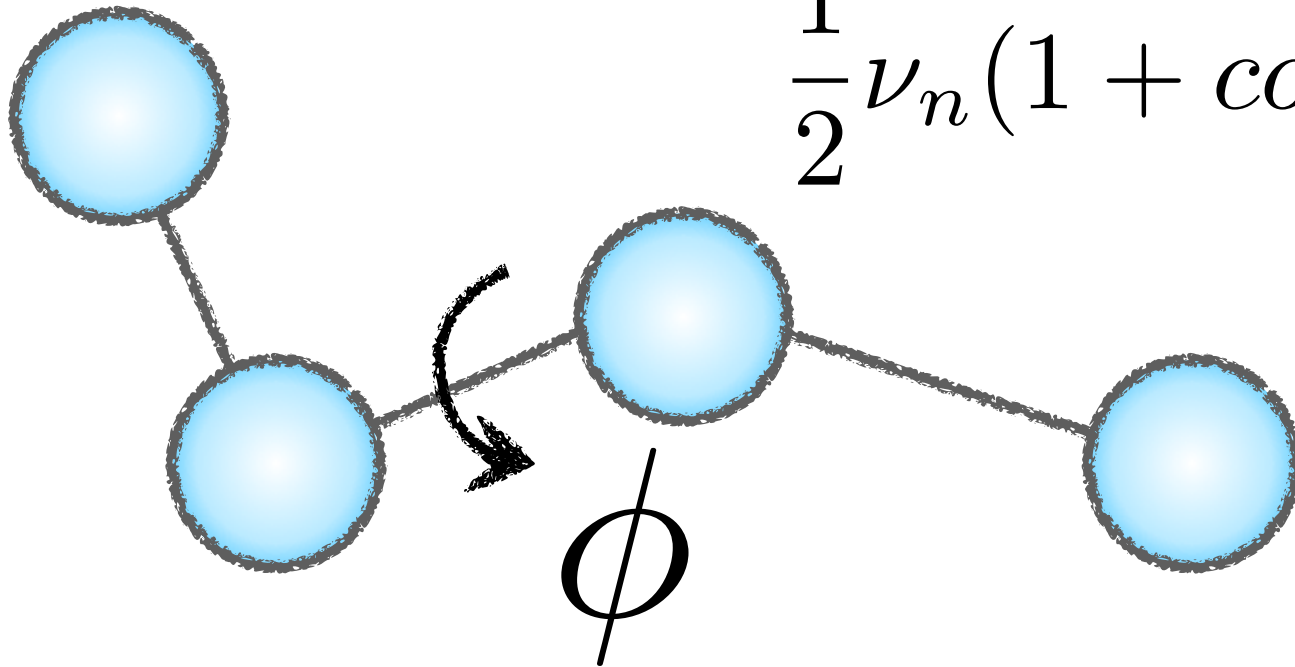
Bonded interactions

$$k_{\theta}(\theta - \theta_{eq})^2$$

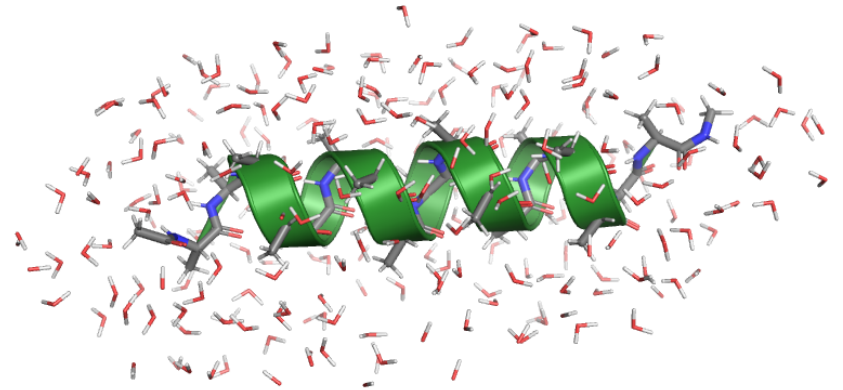


Bonded interactions

$$\frac{1}{2}\nu_n(1 + \cos(n\phi - \phi_0))$$



Bonded interactions



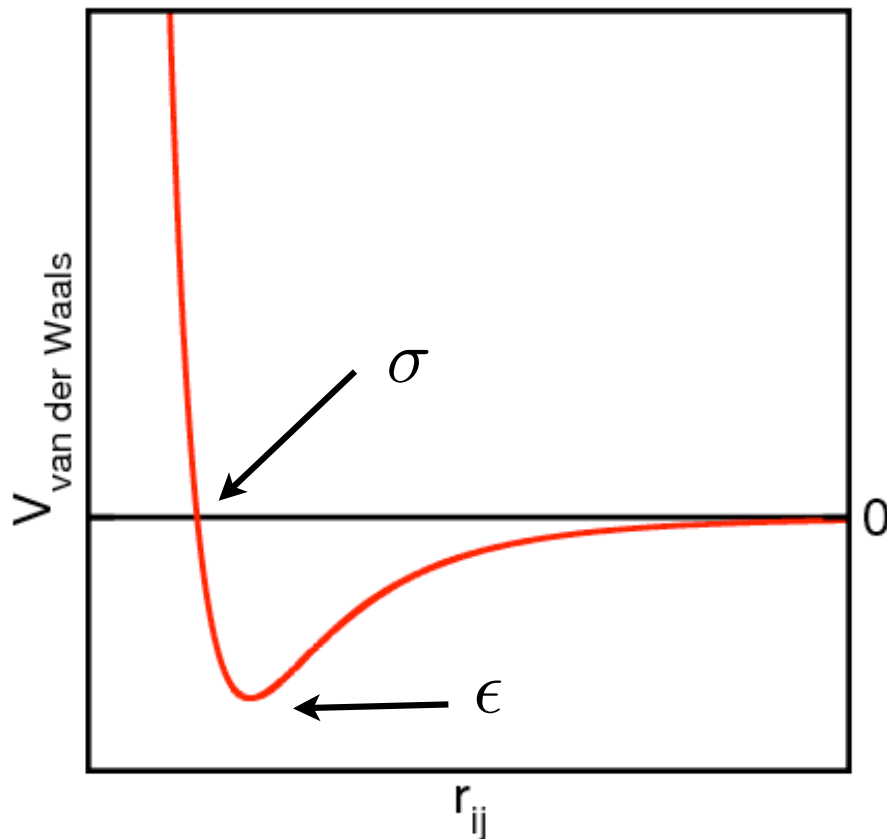
$$V(r) = \sum_{bonds} k_r (r - r_{eq})^2 +$$

$$\sum_{angles} k_\theta (\theta - \theta_{eq})^2 +$$

$$\sum_{dihedrals} \frac{1}{2} \nu_n (1 + \cos(n\phi - \phi_0))$$

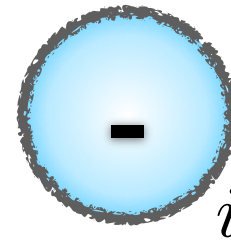
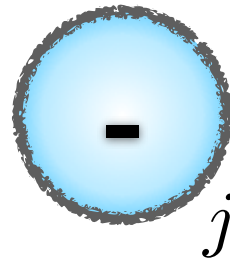
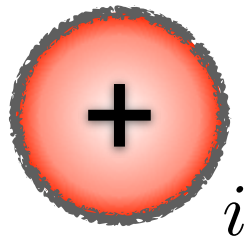
Non-bonded interactions

$$V(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$$



Lennard-Jones potential

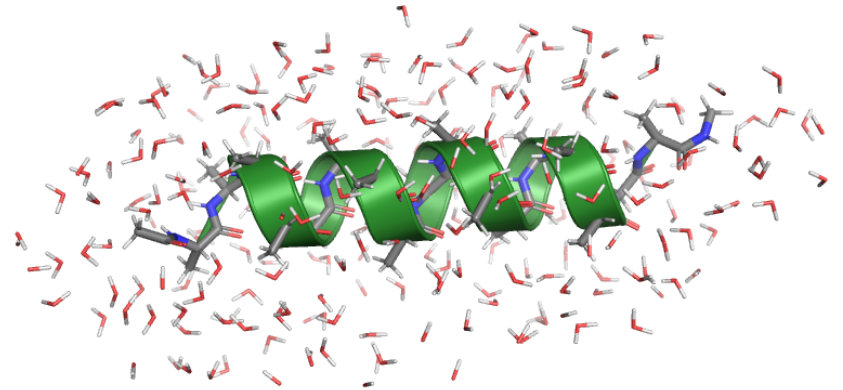
Electrostatic interactions



$$\frac{q_i q_j}{\epsilon_d r_{ij}}$$

Coulomb's law

Force field



$$V(r) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 +$$
$$\sum_{dihedrals} \frac{1}{2} \nu_n (1 + \cos(n\phi - \phi_0)) +$$
$$\sum_{i < j} \left(4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{\epsilon_d r_{ij}} \right)$$

MOLECULAR POTENTIAL ENERGY

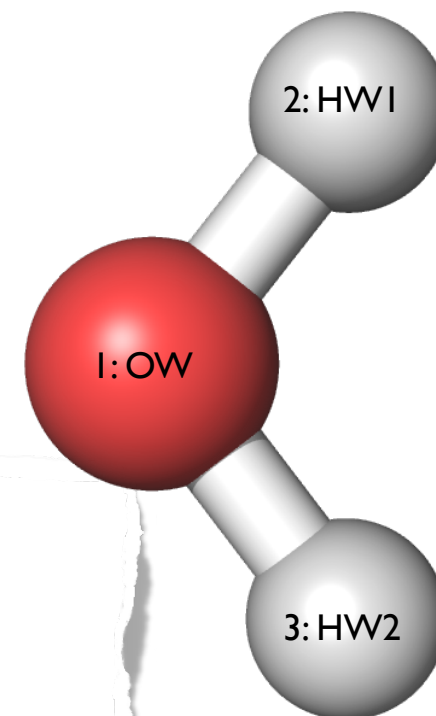
$$\begin{aligned}
 U = & \sum_{\text{All Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{All Angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{All Torsion Angles}} K_\phi [1 - \cos(n\phi + \delta)] \\
 & + \sum_{\text{All Nonbonded pairs}} \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right] \\
 & + \sum_{\text{All partial charges}} \frac{332 q_i q_j}{r}
 \end{aligned}$$

Hooke 1635
 Fourier 1768
 Van der Waals 1837
 Coulomb 1736

Simple sum over many terms

Topology

- List of atoms
- Connectivity



```
[ moleculetype ]  
; molname      nrexcl  
SOL            2
```

```
[ atoms ]  
; id  at type      res nr  res name  at name  cg nr  charge  mass  
1    OW           1      SOL      OW       1      -0.834  16.00000  
2    HW           1      SOL      HW1      1      0.417   1.00800  
3    HW           1      SOL      HW2      1      0.417   1.00800
```

```
[ bonds ]  
; i    j          funct  length  force_constant  
1      2          1      0.09572  502416.0    0.09572      502416.0  
1      3          1      0.09572  502416.0    0.09572      502416.0
```

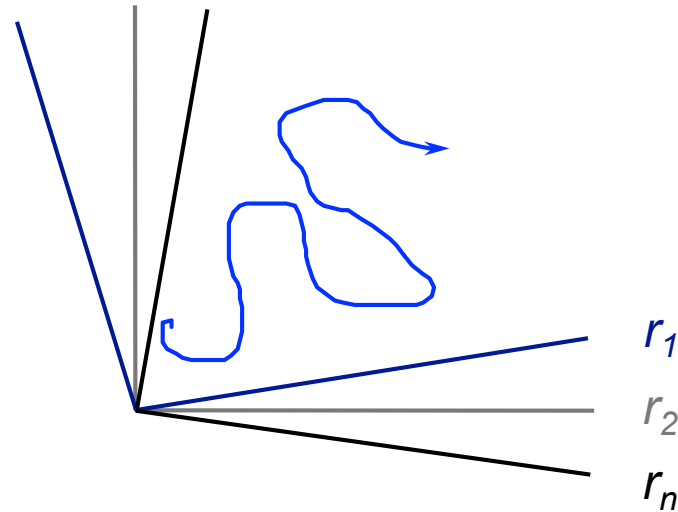
```
[ angles ]  
; i    j          k      funct  angle  force_constant  
2      1          3      1      104.52  628.02      104.52  628.02
```

Force field

$$\begin{aligned} V(r) = & \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \\ & \sum_{dihedrals} \frac{1}{2} \nu_n (1 + \cos(n\phi - \phi_0)) + \\ & \sum_{i < j} \left(4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{\epsilon_d r_{ij}} \right) \end{aligned}$$

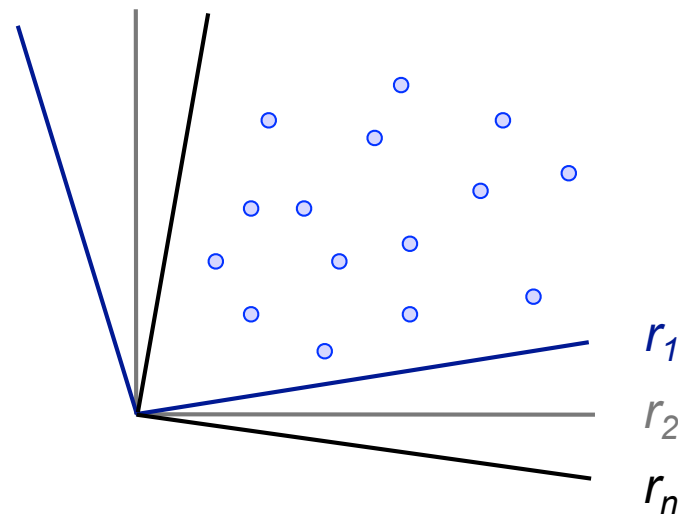
Sampling

Molecular dynamics:
solve equations of motion



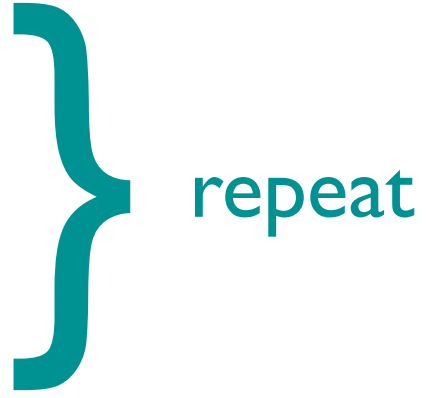
MD

Monte Carlo:
importance sampling

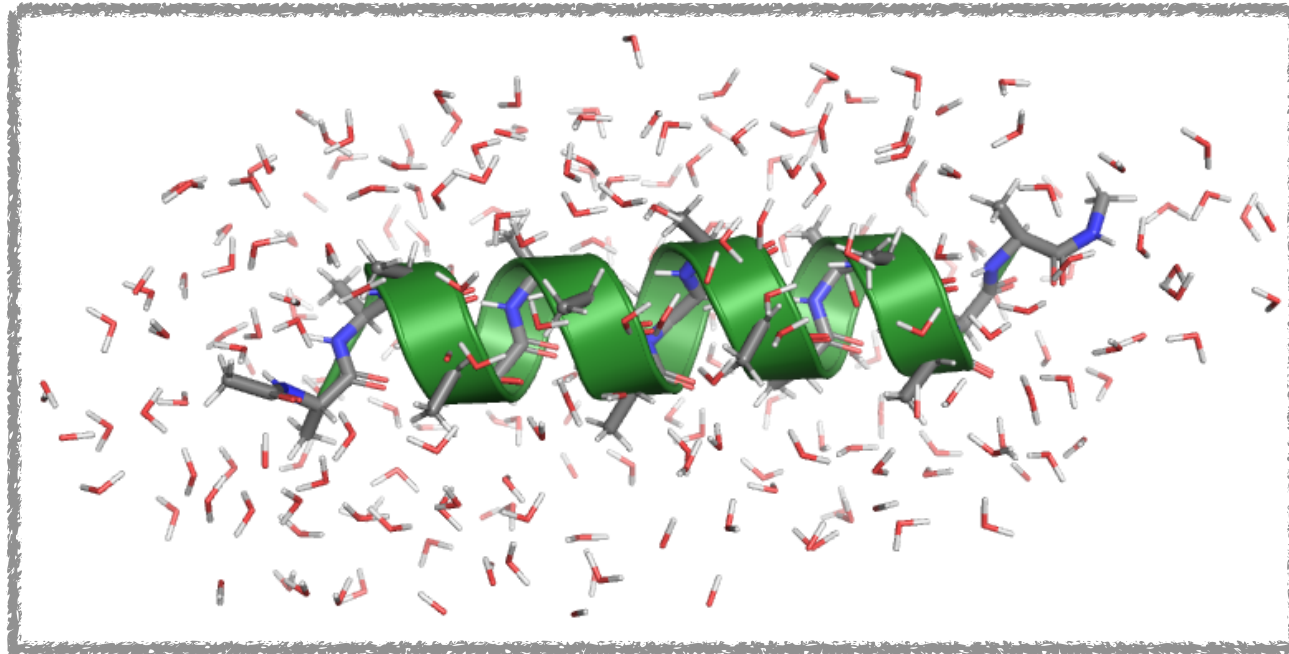


MC

Molecular dynamics

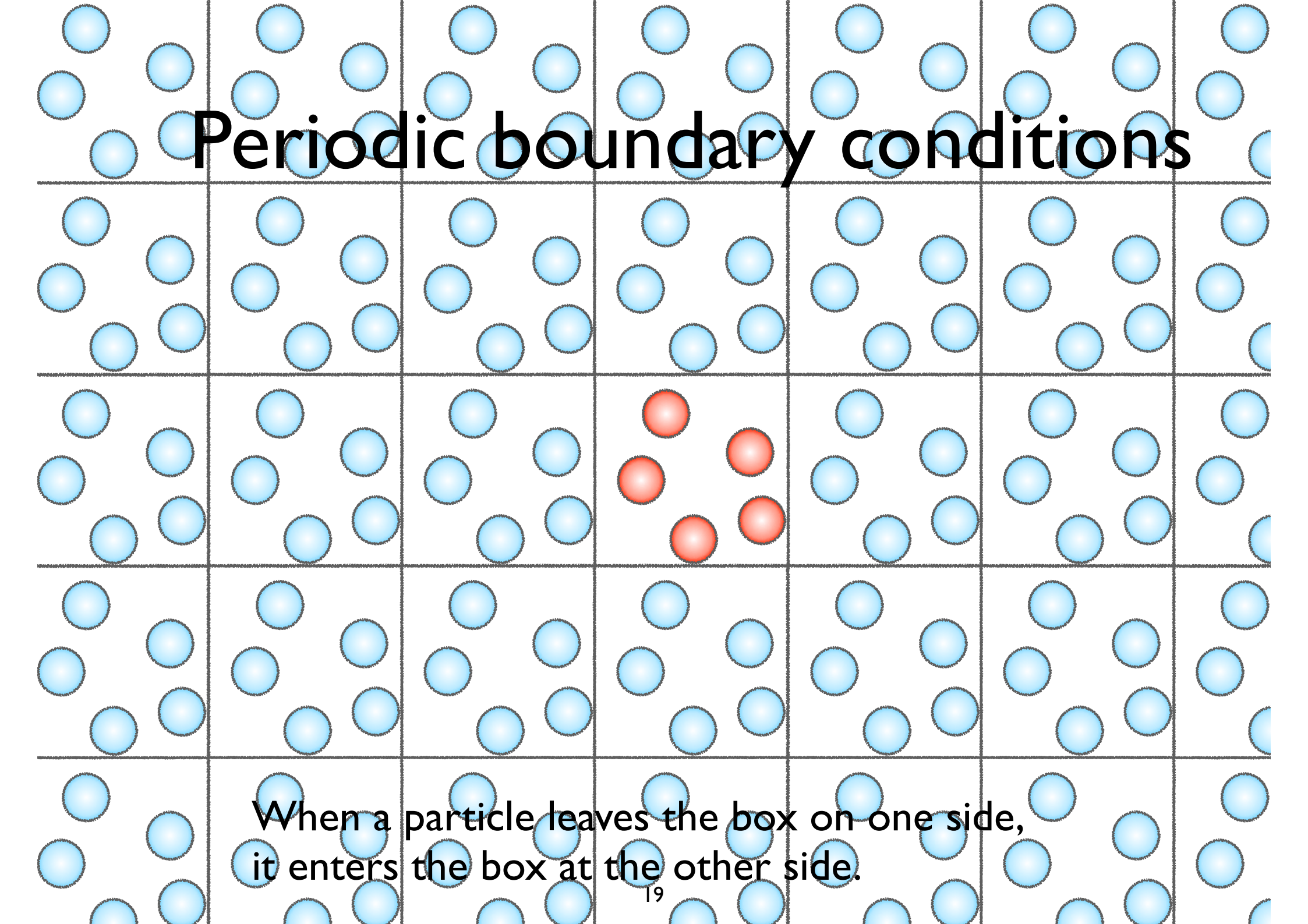
- assign positions and velocities to particles
 - compute forces on all particles
 - integrate equations of motion
 - measure properties
 - stop
- 
- repeat

Simulating an alpha-helix



constant N, p, T

Periodic boundary conditions

A 5x7 grid of particles is shown. Most particles are light blue circles with a white highlight. In the fourth column from the left, there are four red circles instead of blue ones. The grid is divided by thin black lines. The text 'Periodic boundary conditions' is centered at the top, and 'When a particle leaves the box on one side, it enters the box at the other side.' is at the bottom. A small '19' is at the bottom center.

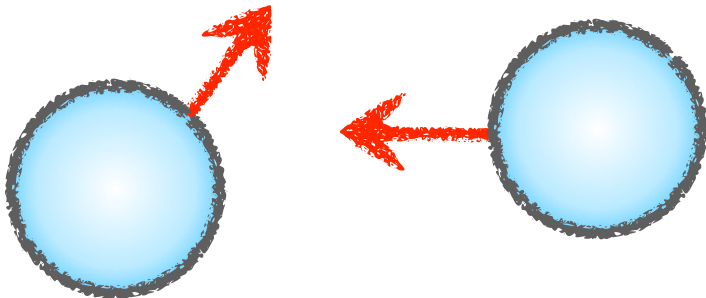
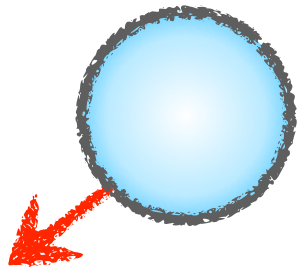
When a particle leaves the box on one side,
it enters the box at the other side.

Ambient conditions

- temperature $\left\langle \frac{1}{2}mv^2 \right\rangle = \frac{3}{2}k_B T$
- pressure $P = \rho k_B T + \frac{1}{V} \left\langle \sum_{i < j} f(r_{ij}) \cdot r_{ij} \right\rangle$

Thermostat

$$v^2 \sim T$$



change velocities by

- scaling

Berendsen and Bussi thermostats

- stochastic forces

Andersen thermostat

- adding additional variable that modulates the kinetic energy

Nosé-Hoover thermostat

Barostat

$$\frac{1}{V} \sim P$$



change volume by

- scaling

Berendsen barostat

- adding additional variable that modulates changes in volume

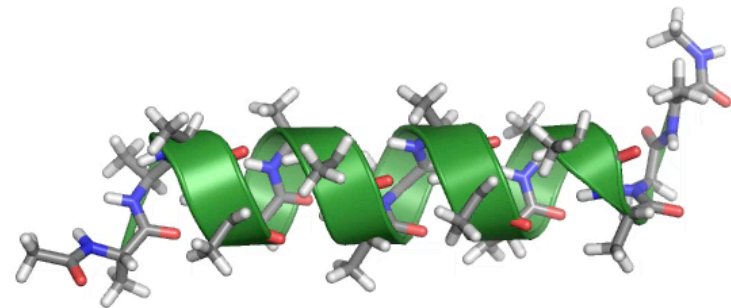
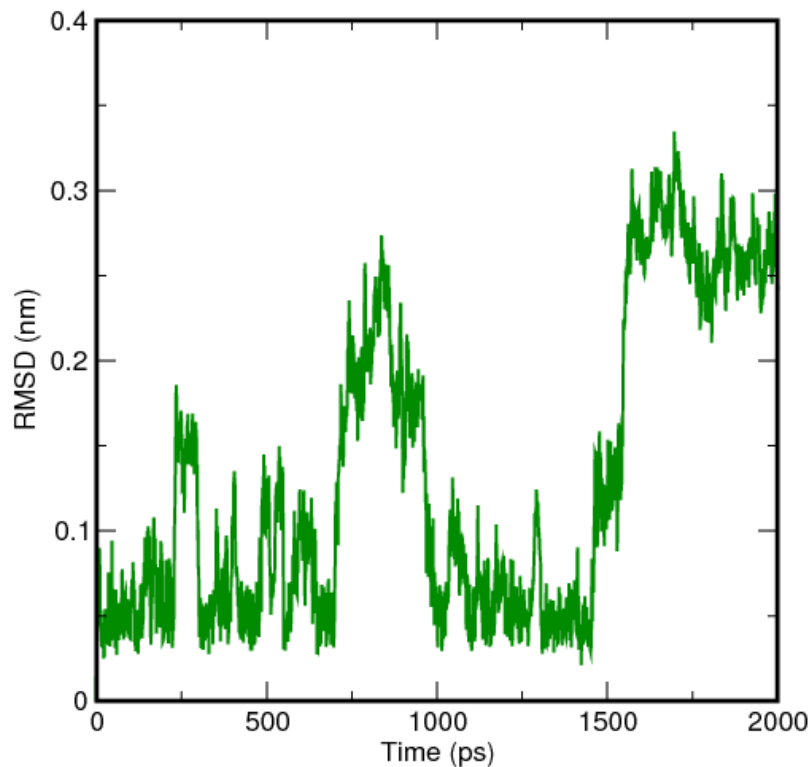
Andersen barostat

2 ns MD simulation of an alpha-helix

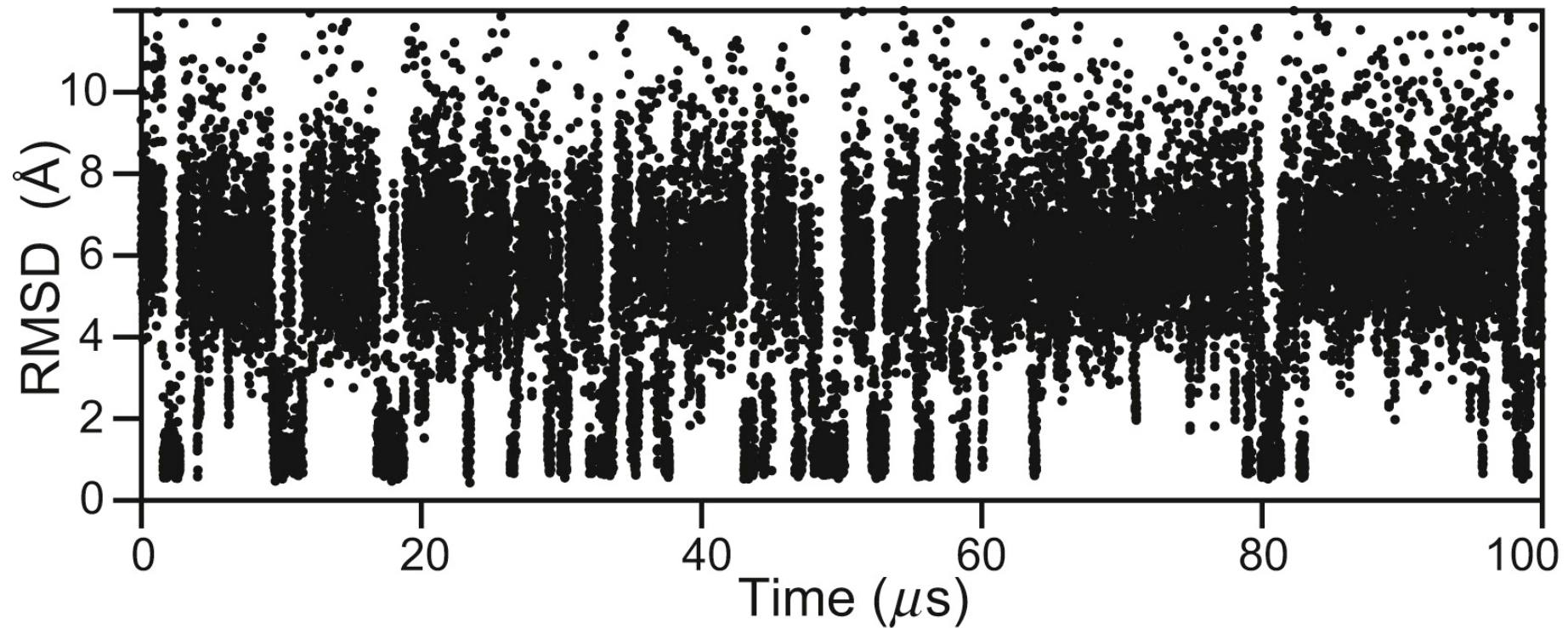
Root Mean Square Deviation
RMSD

$$\sqrt{\frac{\sum_i^N (r_i - \hat{r}_i)^2}{N}}$$

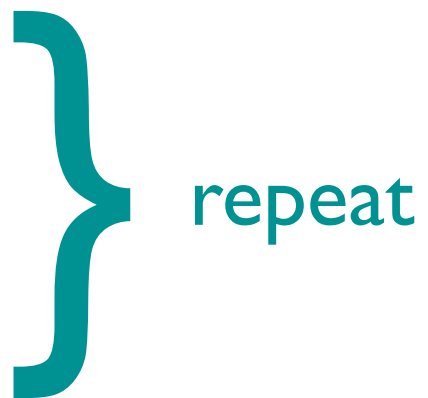
N - number of atoms
 i - atom index
 r - position
 $\hat{r}$ - reference position



Long MD simulation



Molecular dynamics

- assign positions and velocities to particles
 - compute forces on all particles
 - integrate equations of motion
 - measure properties
 - stop
- 
- repeat

When?

Ensemble averages

For properties that only depend on the configurational part, the probability $P(\mathbf{\Gamma})$ to find a configuration $\mathbf{\Gamma} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ is

$$P(\mathbf{\Gamma}) = \frac{1}{Q} \exp[-\beta U(\mathbf{\Gamma})]$$

with Q the partition function, U the internal energy and $\beta = 1/k_B T$.

The ensemble average of a property A is

$$\begin{aligned} \langle A \rangle &= \int P(\mathbf{\Gamma}) A(\mathbf{\Gamma}) d\mathbf{\Gamma} \\ \langle A \rangle &= \frac{1}{Q} \int \exp[-\beta U(\mathbf{\Gamma})] A(\mathbf{\Gamma}) d\mathbf{\Gamma} \end{aligned}$$

Ergodicity theorem

Suppose we have an ensemble average of a system defined by $U(\Gamma)$ obtained by MC.

$$\langle A \rangle = \frac{1}{Q} \exp[-\beta U(\Gamma)] A(\Gamma) d\Gamma$$

Now suppose we have an NVT molecular dynamics trajectory for the same system.
A time average over the trajectory is simply:

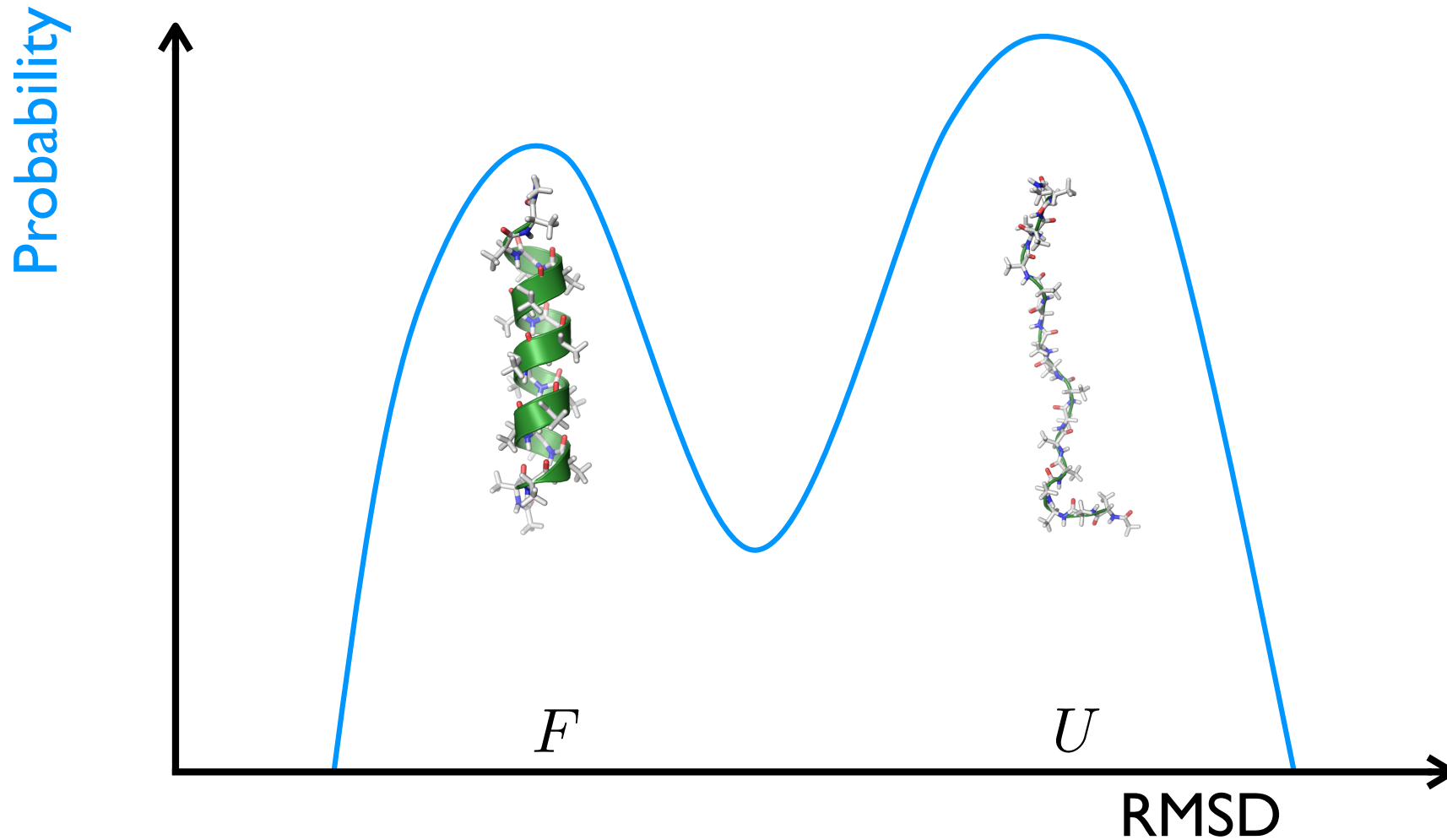
$$\bar{A} = \frac{1}{T} \int_0^T A(t) dt$$

The ergodicity theorem states that for an 'ergodic system',
the time average is equal to the ensemble average.

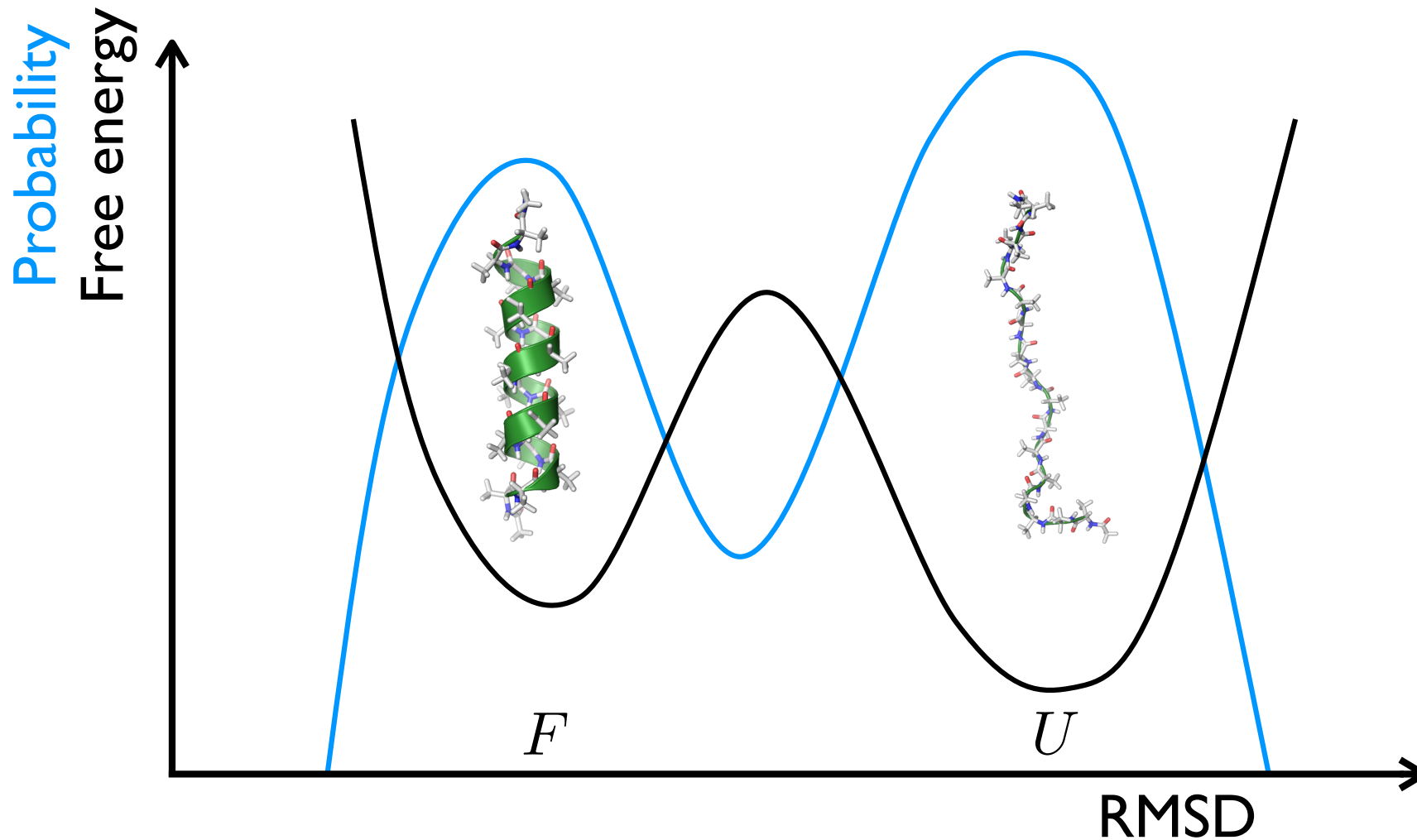
$$\bar{A} = \langle A \rangle \quad \text{MC and MD give the same averages!}$$

An MD simulation is done when the probability distribution no longer changes.

Probability histogram

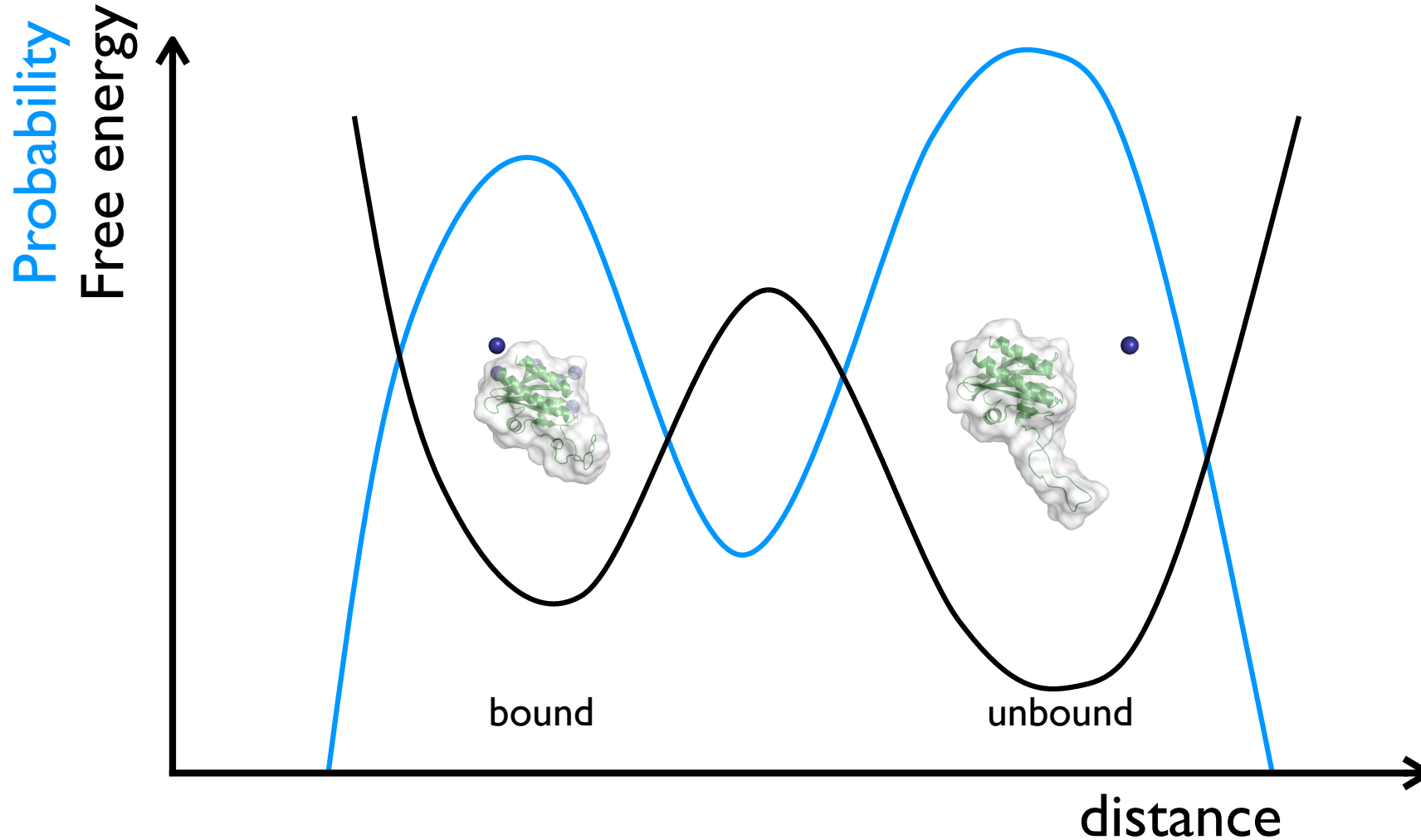


Free energy profile



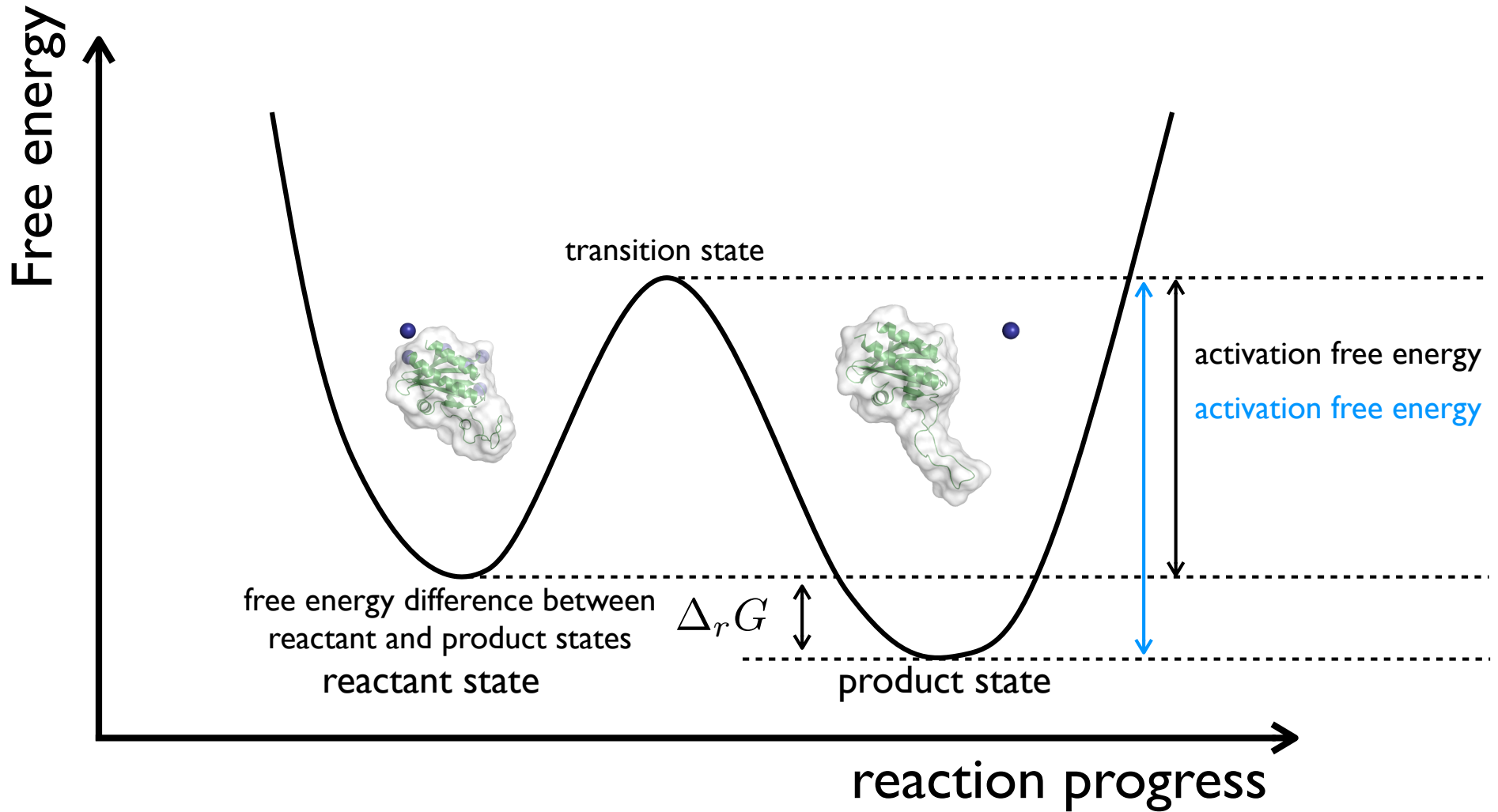
Free energy = $-\ln(\text{Probability})$ in units of $k_B T$

Free energy profile

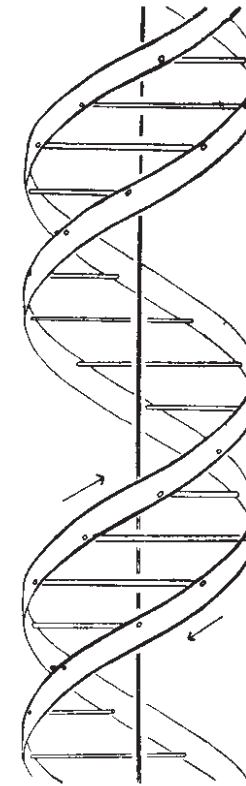


Free energy = $-\ln(\text{Probability})$ in units of $k_B T$

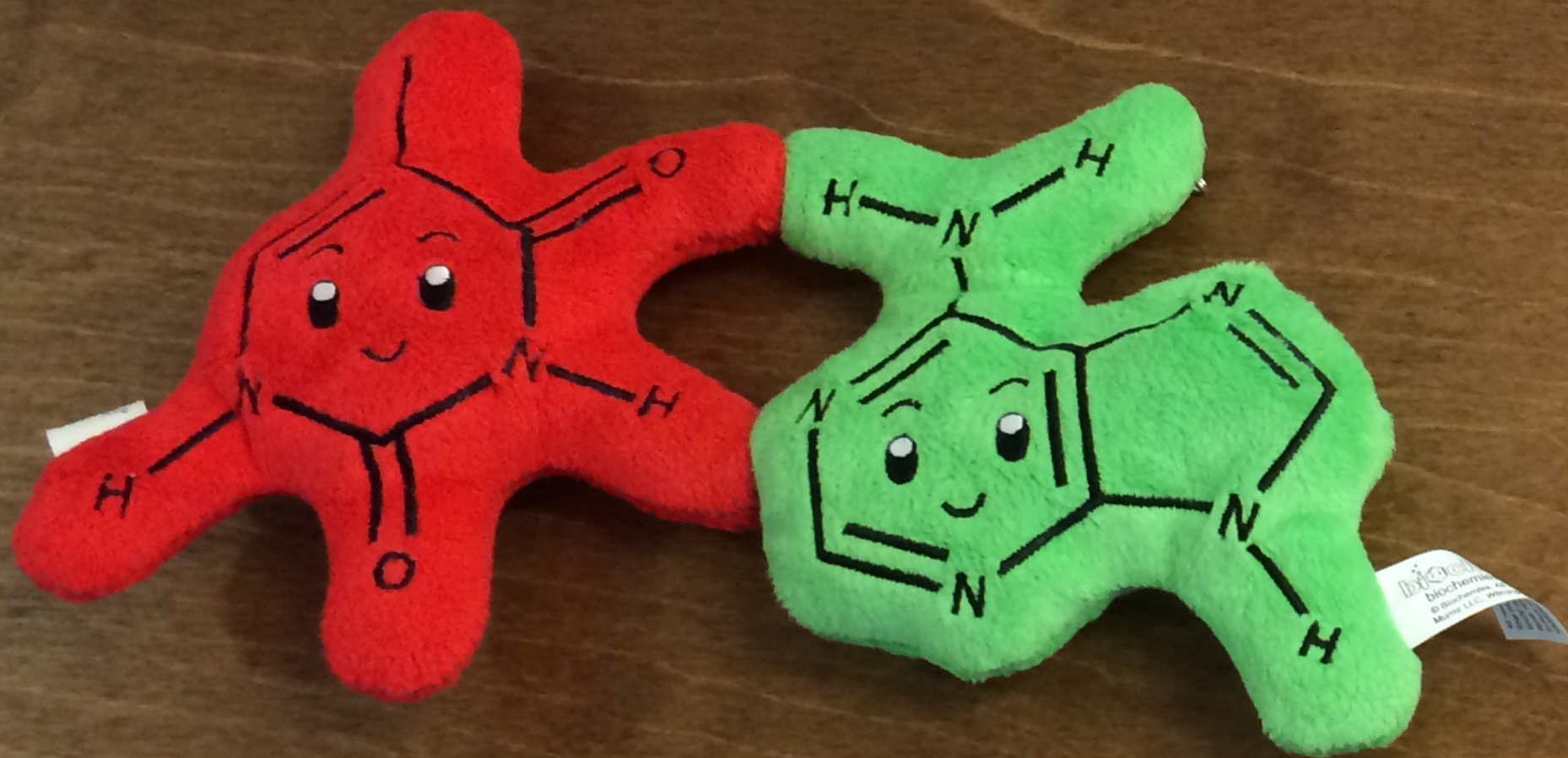
Free energy profile



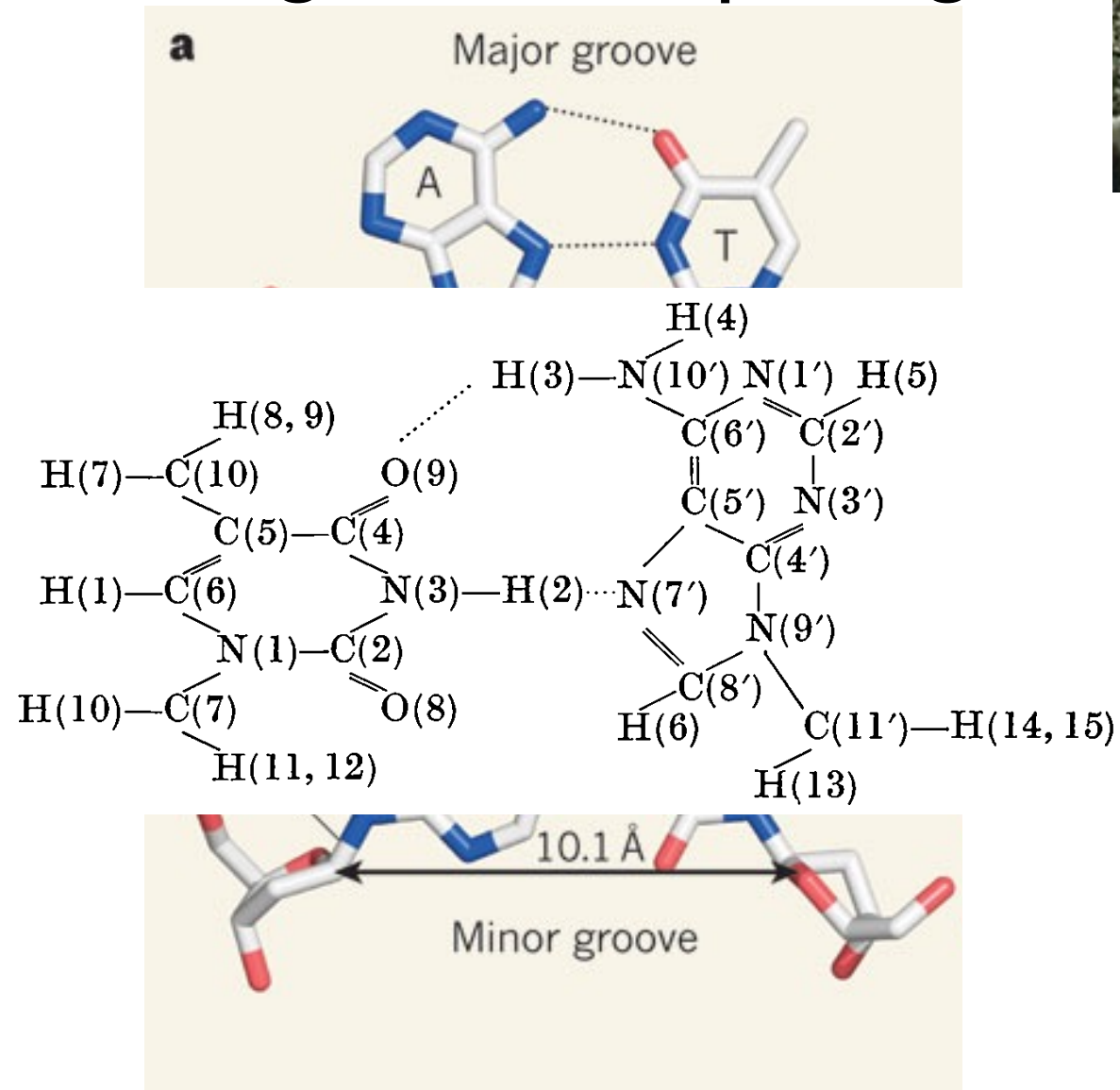
Structure of DNA

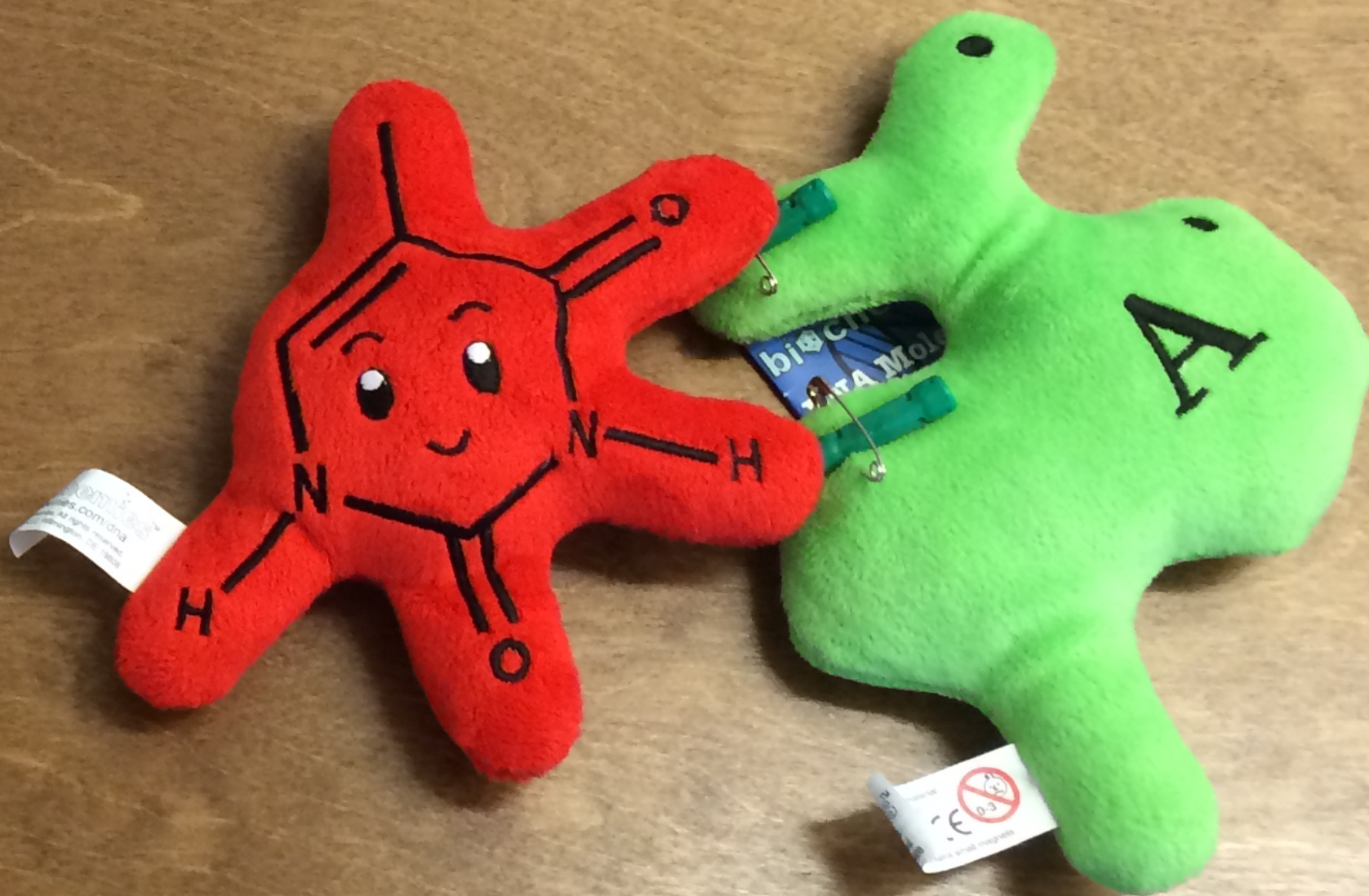


This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

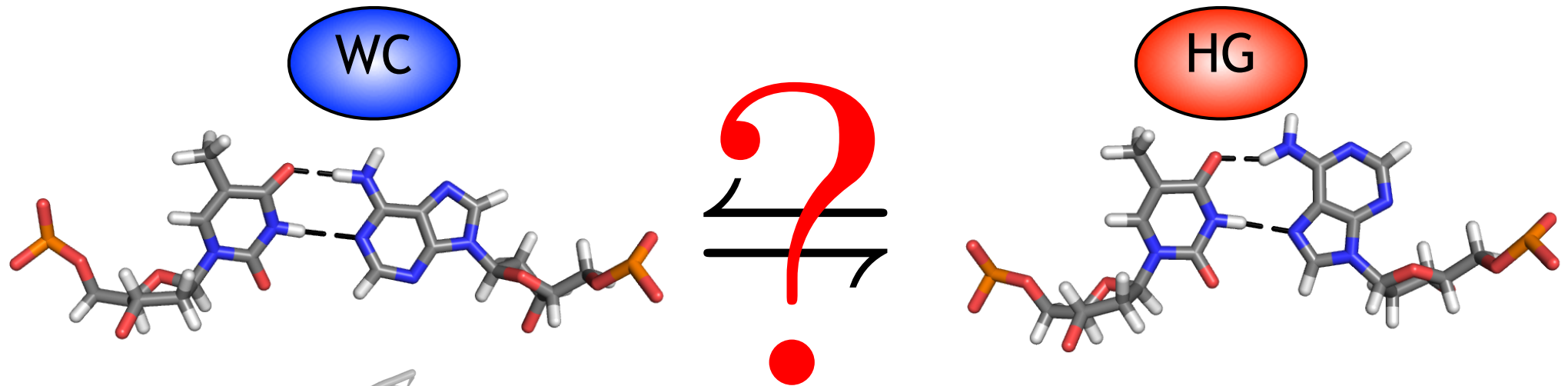


Hoogsteen basepairing

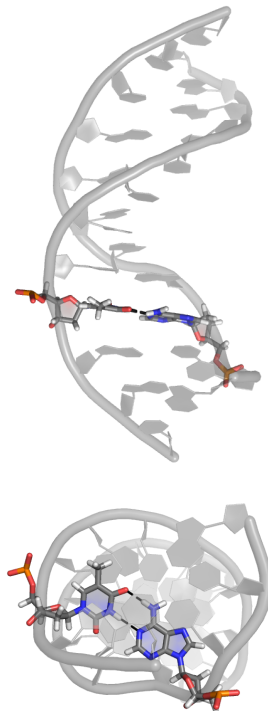




What is the mechanism?



a
A₆-DNA
5'
1 C—G
G—C
A—T
T—A
5 T—A 20
T—A
T—A
T—A
T—A
10 T—A 15
G—C
G—C
C—G
5'



Molecular Dynamics:

5' -GGATTTTTTGGC-3'

3' -CCTAAAAAACCG-5'

AMBER03

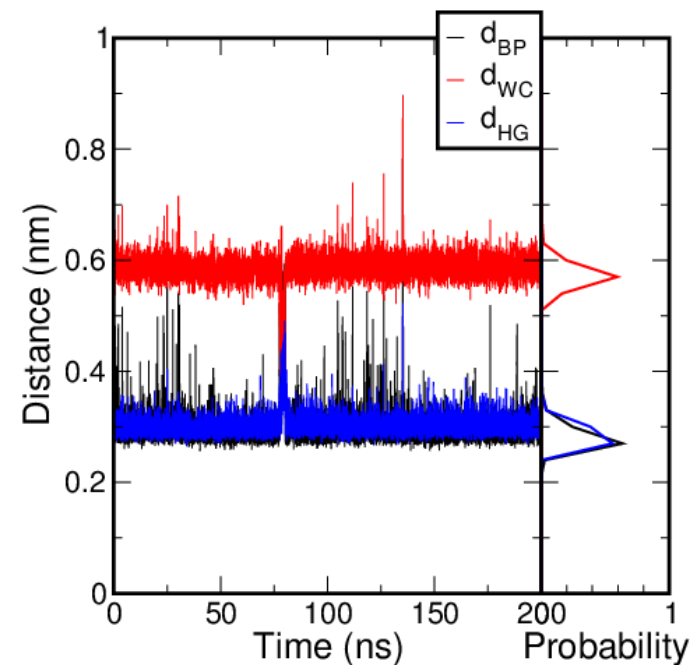
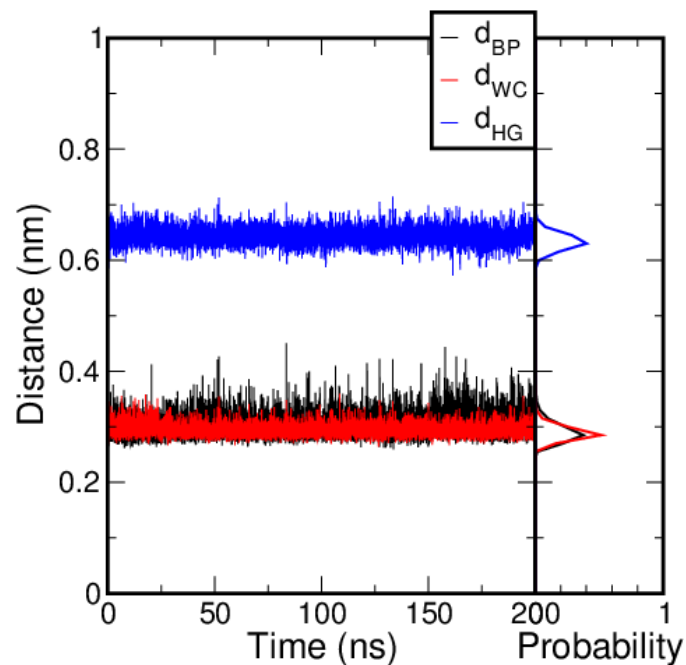
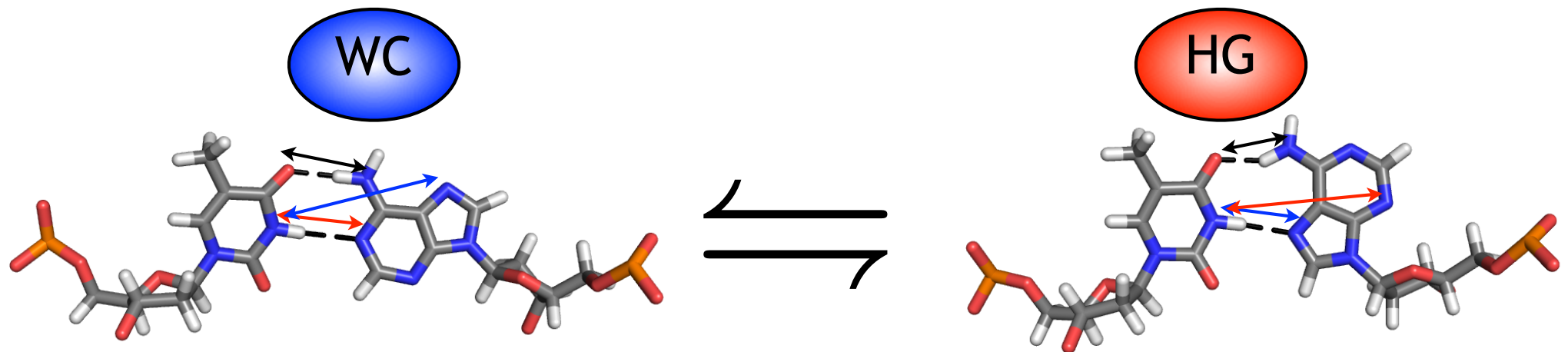
TIP3P water + 25 mM NaCl

NpT ensemble, T = 300 K

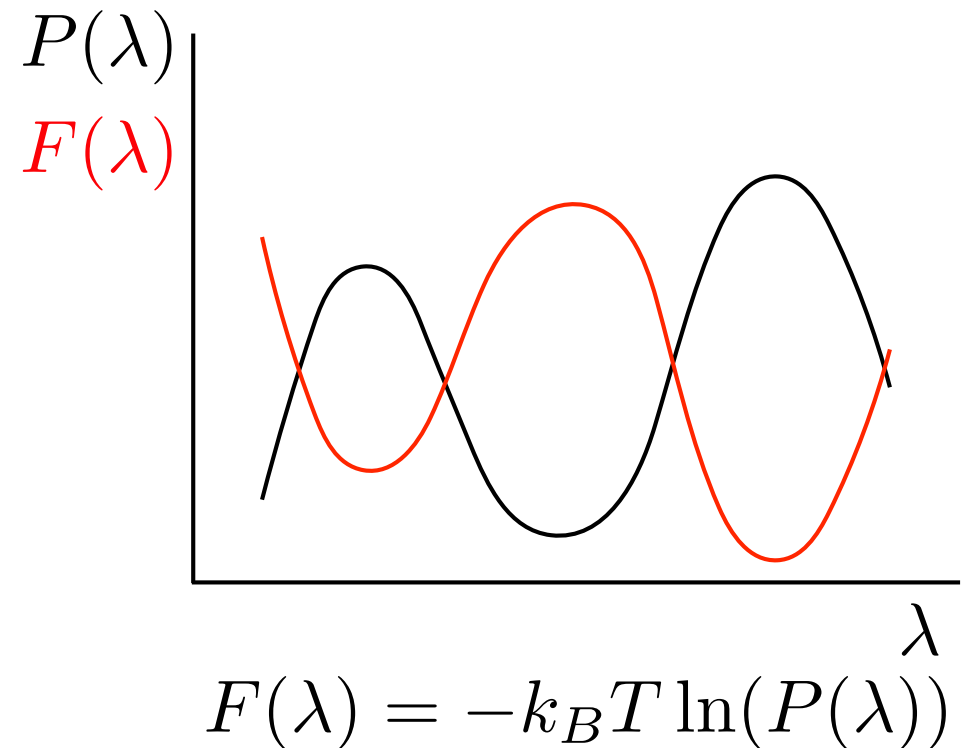
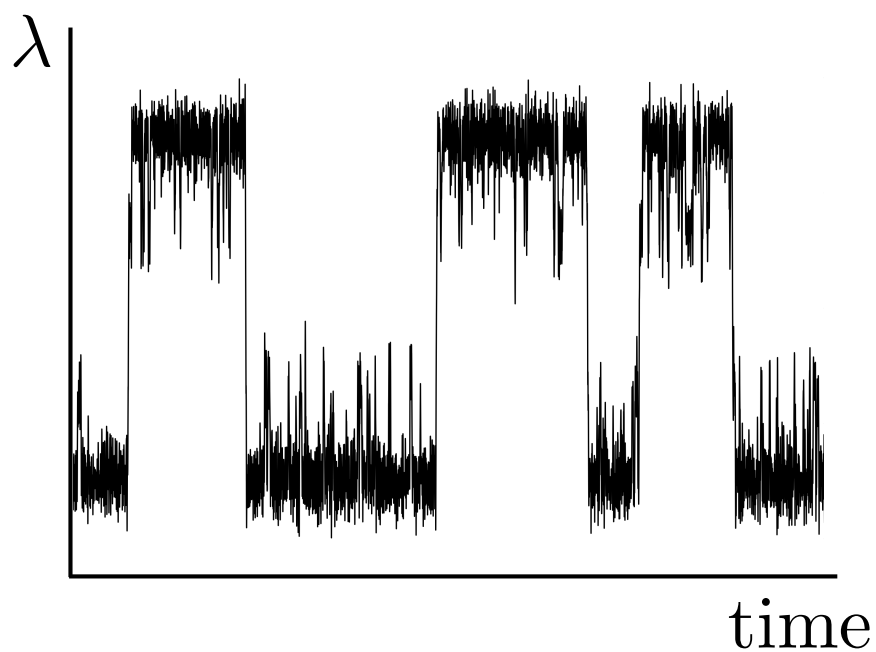
20868 atoms

gromacs v4.6.1

The transition between WC and HG does not occur within 200 ns.

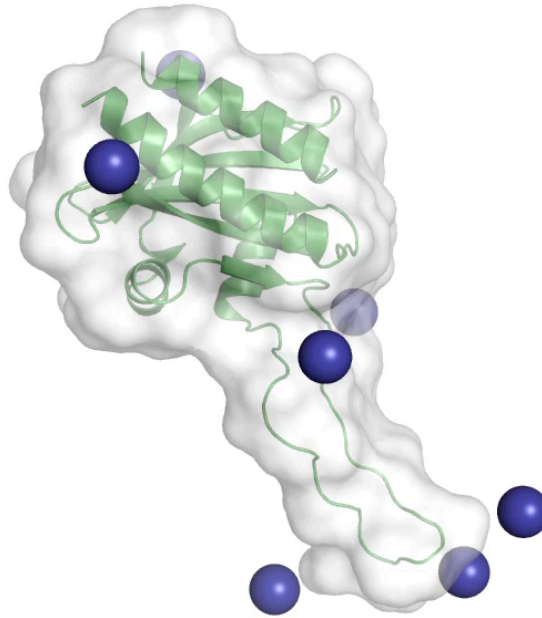


Computing a free energy profile from a molecular dynamics simulation



Problem: little sampling except in minima of $F(\lambda)$

All-atom force field molecular dynamics of biomolecular systems



$$F = m \frac{d^2 r}{dt^2} = ma$$

Given the potential, one can **numerically** integrate the trajectory of the whole system as a function of time.

$$F = -\frac{dV(r)}{dr}$$

The force F is given by the gradient of the potential $V(r)$.

$$V(r) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{1}{2} \nu_n (1 + \cos(n\phi - \phi_0)) + \sum_{i < j} \left(4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{\epsilon_d r_{ij}} \right)$$

bonds

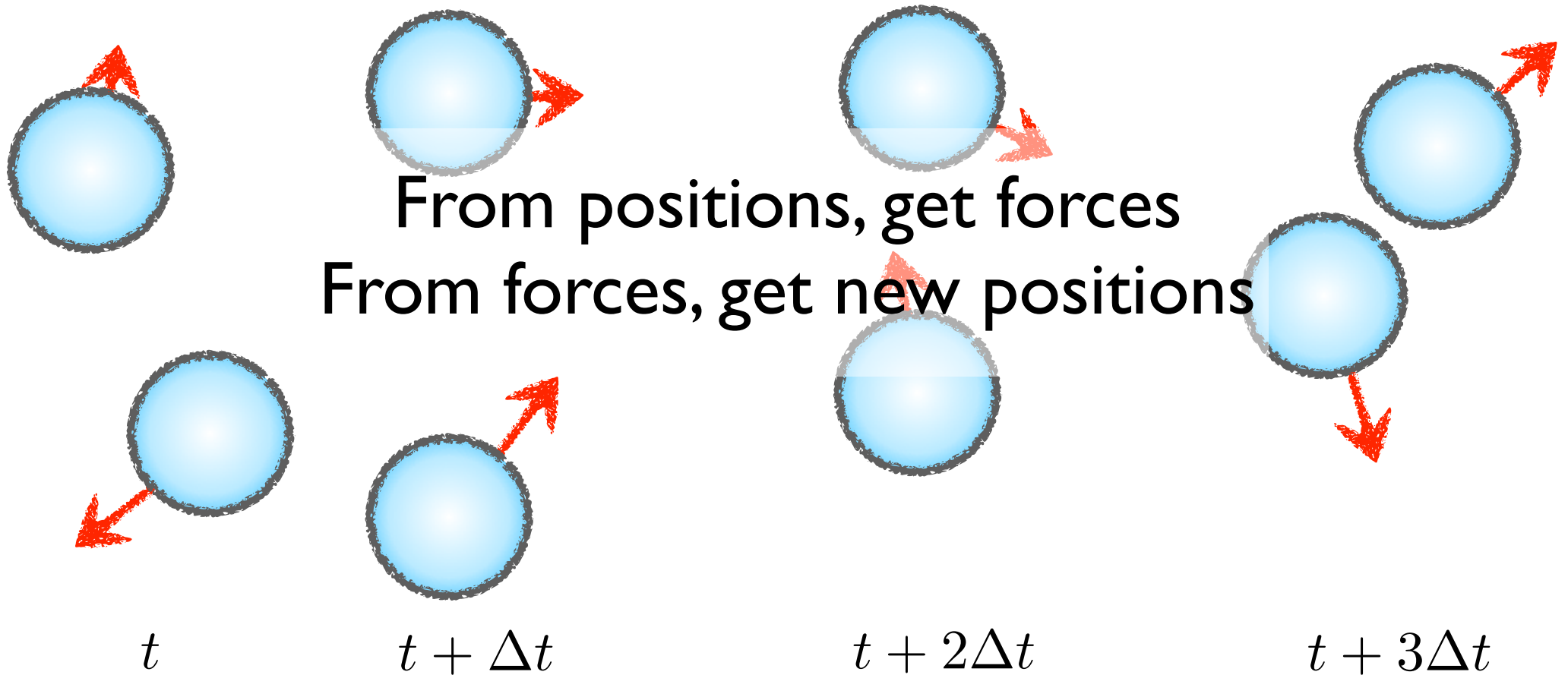
bends

torsions

non-bonded

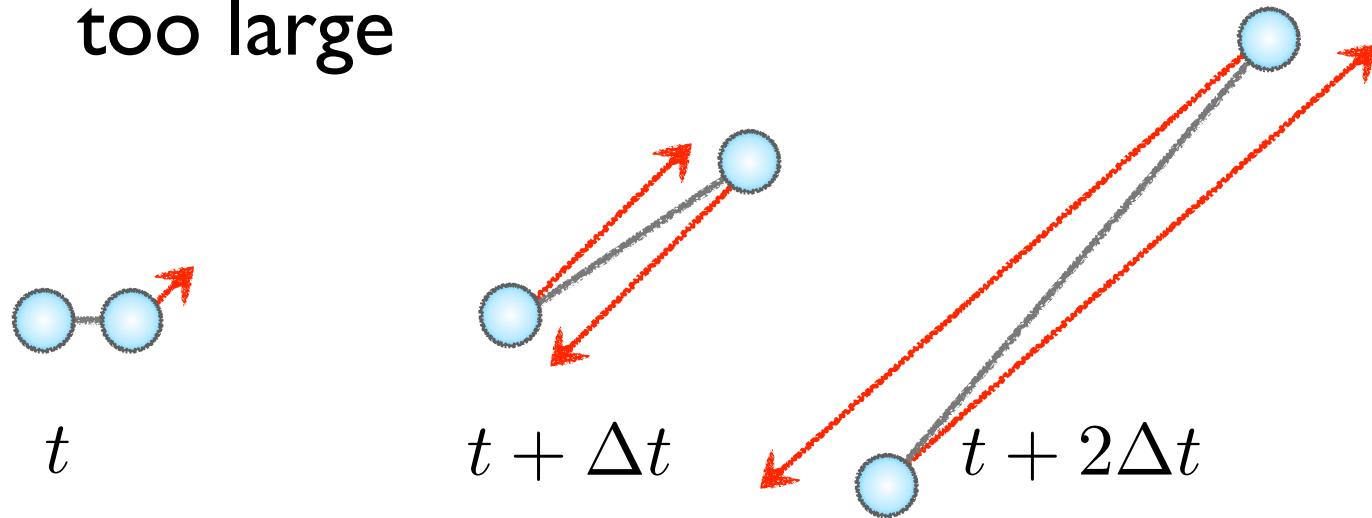
Numerical approach

Snapshots

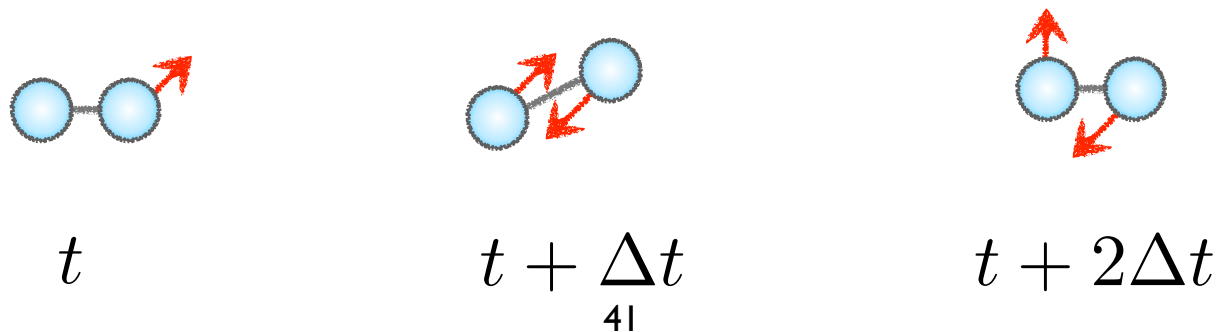


Choosing the time step

Δt too large

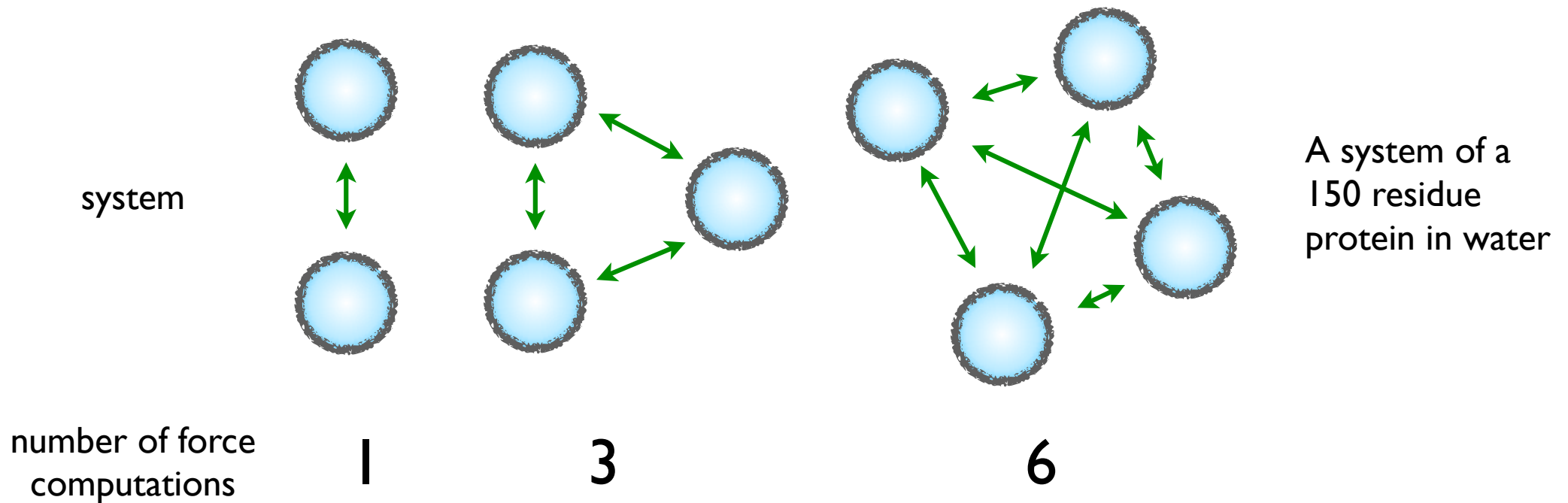


Δt just fine (2 fs for all-atom force field MD)



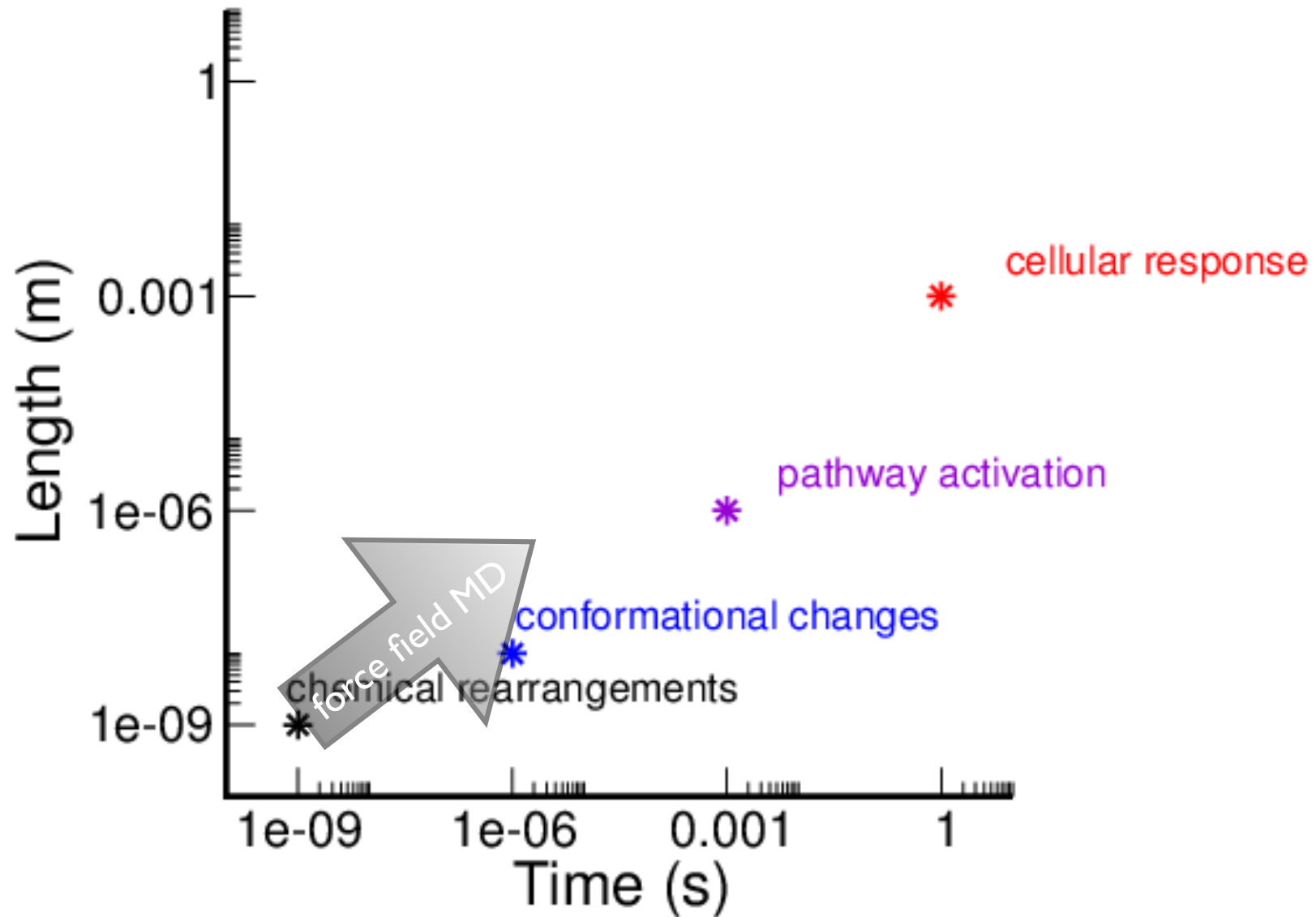
Force computations

Assumption: pairwise interactions

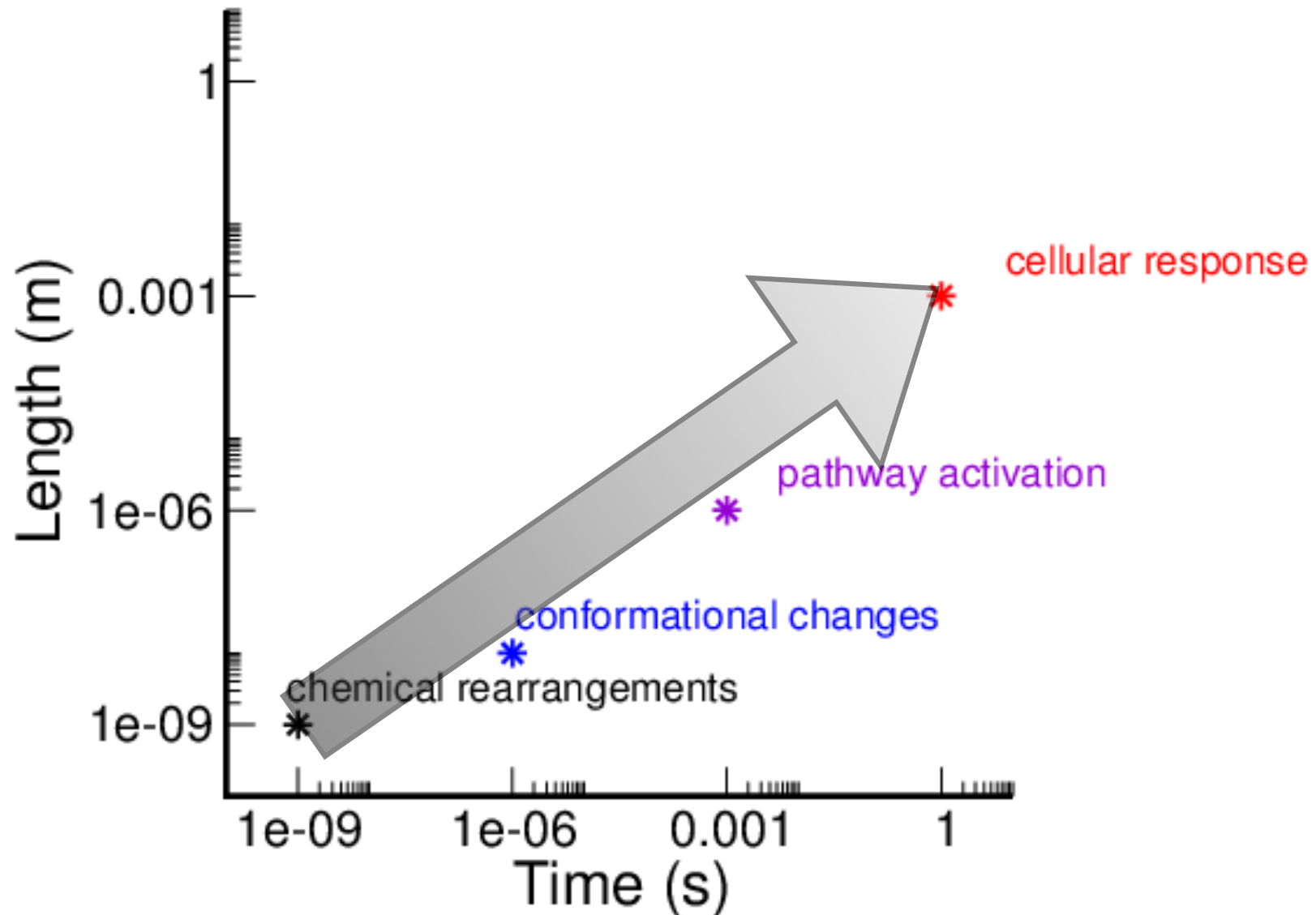


A system of N particles requires $N(N-1)/2$ force computations.

Where we are...



And where we want to go!



$$\Delta t = 2 \text{ fs} = 2 \cdot 10^{-15} \text{ s} = 0.0000000000000002 \text{ s}$$

A time step(Δt) takes about 0.0026 seconds = 2.6 ms on a supercomputer.

$$1 \text{ ns} = 10^{-9} \text{ s}$$

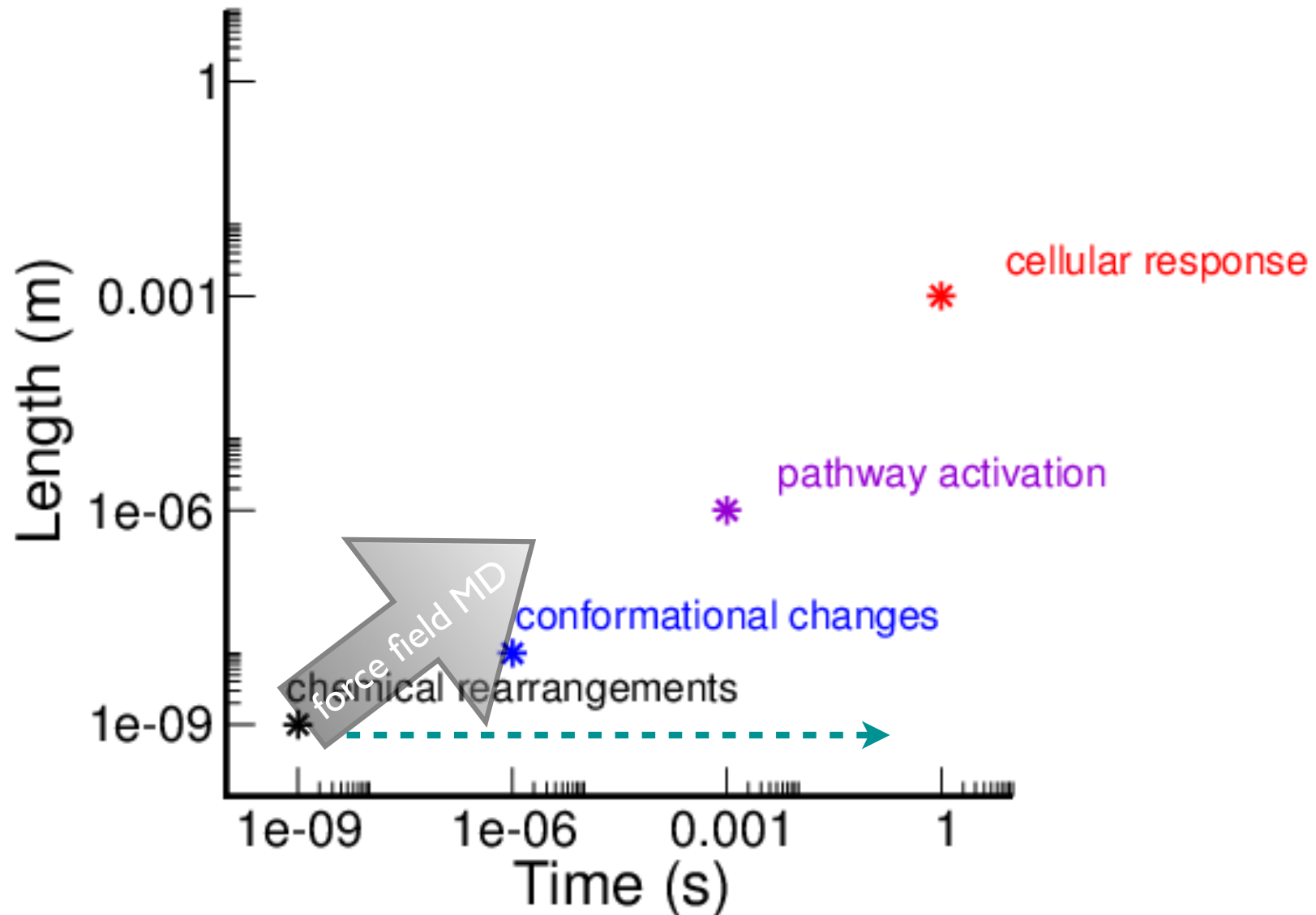
$$1 \text{ ns} = 500.000 \text{ steps} = 13.000 \text{ s} \approx 22 \text{ minutes}$$

$$1 \text{ s} = 500.000.000.000.000 \text{ steps} \\ \approx 42.000 \text{ year}$$

A protein in water

How long does it take to calculate the movements of the protein for 1 s?

Bridging the time gap



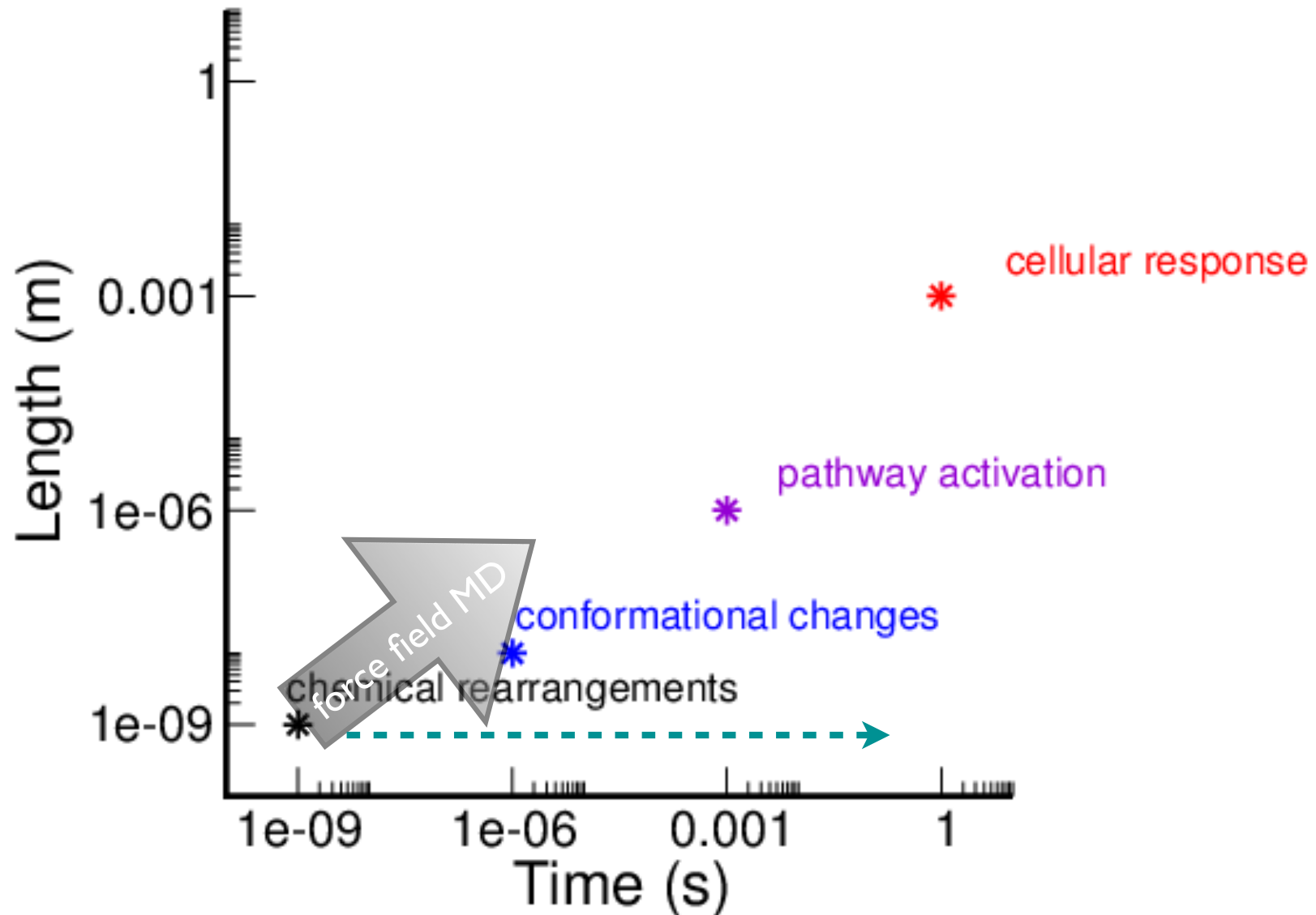
Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

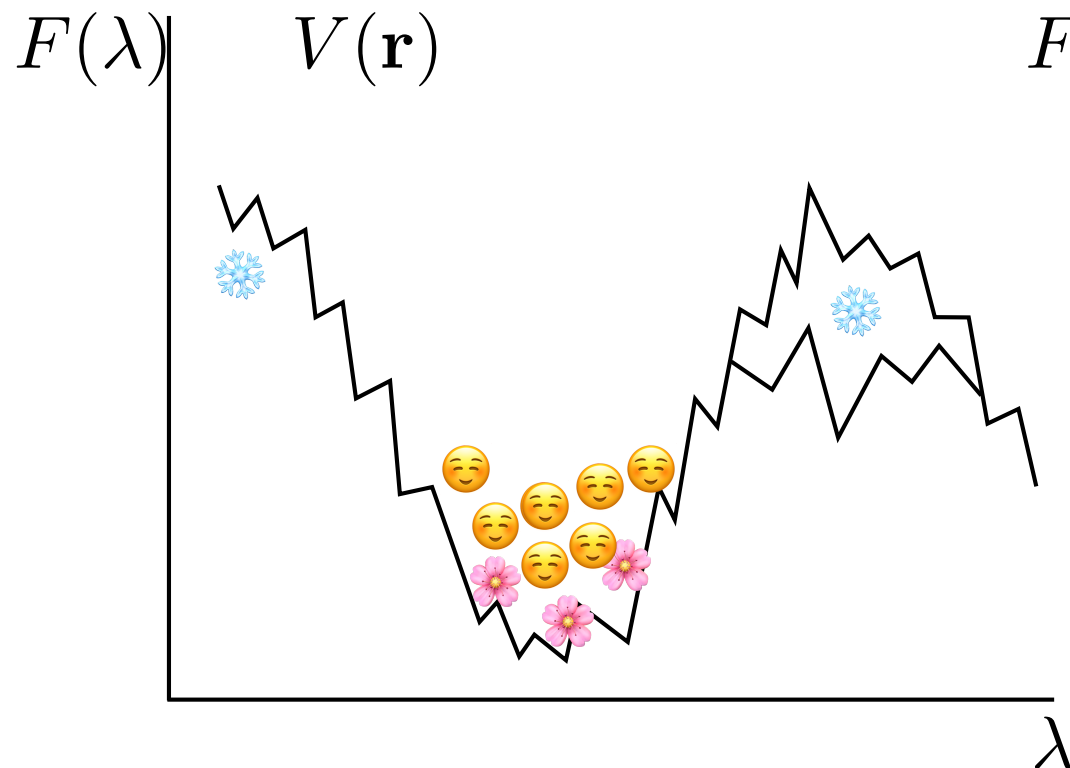
Part 2: Biased sampling

Part 3: Path sampling

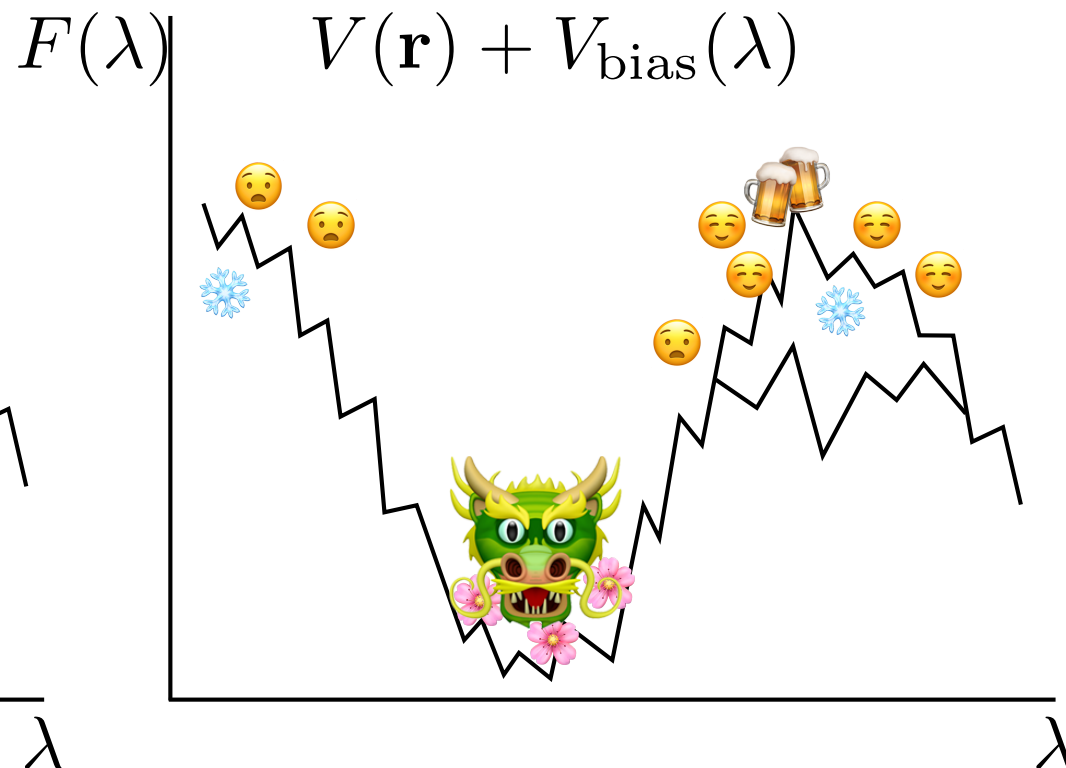
Bridging the time gap



Biased sampling

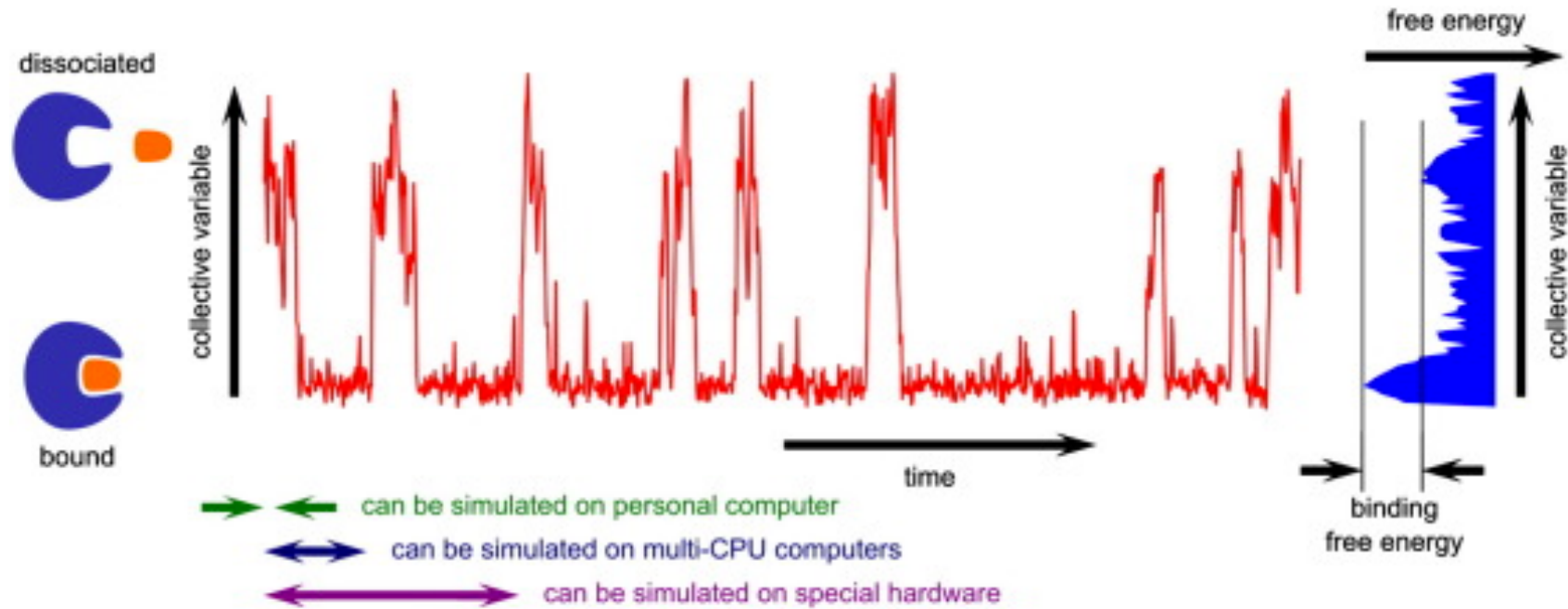


Straightforward MD



Biased sampling

Reaction coordinate



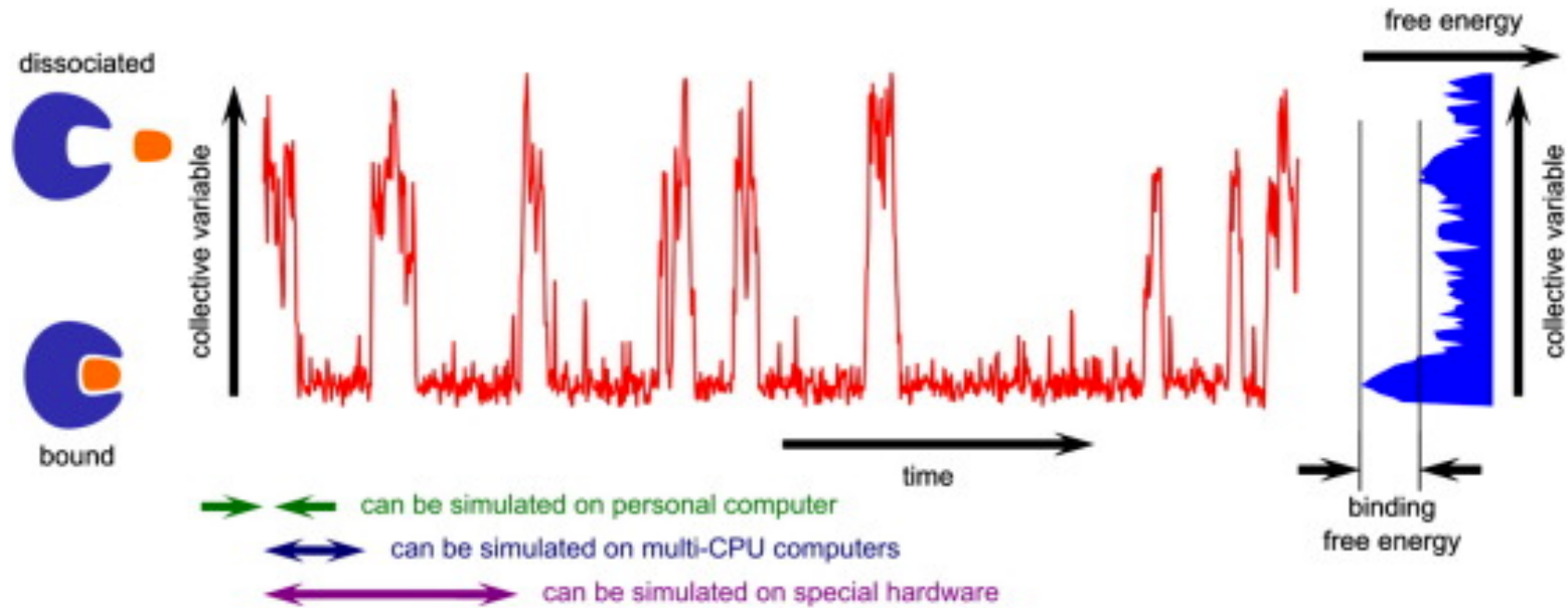
A collective variable (CV) is a function λ of the $3N$ -dimensional configuration $\mathbf{r}$.
 λ maps $\mathbf{r}$ onto an M -dimensional space $\mathbf{z}$.

$$M \ll 3N$$

$$P(\mathbf{z}) = \langle \delta[\lambda(\mathbf{r}) - \mathbf{z}] \rangle$$

$$F(\mathbf{z}) = -k_B T \ln \langle \delta[\lambda(\mathbf{r}) - \mathbf{z}] \rangle$$

Reaction coordinate



A collective variable (CV) is a function Q of the $3N$ -dimensional configuration $\mathbf{r}$.
 Q maps $\mathbf{r}$ onto an M -dimensional space $\mathbf{z}$.

$M \ll 3N$

$$P(\mathbf{z}) = \langle \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle$$

$$F(\mathbf{z}) = -k_B T \ln \langle \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle$$

Biased sampling

potential

$$V_b(\mathbf{r}) = V(\mathbf{r}) + \Delta V(\mathbf{r})$$

$V(\mathbf{r})$ interatomic potential, dependent on positions $\mathbf{r}$
 $\Delta V(\mathbf{r})$ bias potential, dependent on $\mathbf{r}$

Biased sampling

Derive statistics on a system with different energetics from the energetics used to perform the sampling

potential

$$V_b(\mathbf{r}) = V(\mathbf{r}) + \Delta V(\mathbf{r})$$

$V(\mathbf{r})$ interatomic potential, dependent on positions $\mathbf{r}$
 $\Delta V(\mathbf{r})$ bias potential, dependent on $\mathbf{r}$

distribution

$$P_b(\mathbf{z}) = \frac{\int d\mathbf{r} \exp(-\beta V(\mathbf{r})) \exp(-\beta \Delta V(\mathbf{r})) \delta[Q(\mathbf{r}) - \mathbf{z}]}{\int d\mathbf{r} \exp(-\beta V(\mathbf{r})) \exp(-\beta \Delta V(\mathbf{r}))}$$

multiply by 1

$$= \frac{\int d\mathbf{r} \exp(-\beta V(\mathbf{r})) \exp(-\beta \Delta V(\mathbf{r})) \delta[Q(\mathbf{r}) - \mathbf{z}]}{\int d\mathbf{r} \exp(-\beta V(\mathbf{r}))} \frac{\int d\mathbf{r} \exp(-\beta V(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V(\mathbf{r})) \exp(-\beta \Delta V(\mathbf{r}))}$$

$$= \frac{\langle \exp(-\beta \Delta V(\mathbf{r})) \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle}{\langle \exp(-\beta \Delta V(\mathbf{r})) \rangle}$$

ensemble averaged on
interatomic potential $V(\mathbf{r})$

$\mathbf{z}$ - reaction coordinate space

Q - collective variable

Biased sampling

$$P_b(\mathbf{z}) = \frac{\langle \exp(-\beta \Delta V(\mathbf{r})) \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle}{\langle \exp(-\beta \Delta V(\mathbf{r})) \rangle}$$

bias $\Delta V(\mathbf{z})$ is a function of the collective variables $Q(\mathbf{r})$

$$P_b(\mathbf{z}) = \frac{\exp(-\beta \Delta V(\mathbf{z})) \langle \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle}{\langle \exp(-\beta \Delta V(Q(\mathbf{r}))) \rangle}$$

unbiased statistics $P(\mathbf{z}) = \langle \delta[Q(\mathbf{r}) - \mathbf{z}] \rangle$

$$P(\mathbf{z}) = P_b(\mathbf{z}) \exp(\beta \Delta V(\mathbf{z})) \langle \exp(-\beta \Delta V(Q(\mathbf{r}))) \rangle$$

An ergodic MD simulation on ΔV can provide the statistics of Q as if generated with the unbiased potential V .

Choosing ΔV

The closer the bias potential is to the negative free energy $-F(z)$ the more uniform the sampling of Q will be.

$$\text{if } \Delta V(Q(\mathbf{r})) = -F(Q(\mathbf{r}))$$

$$\text{then } \exp(\beta \Delta V(\mathbf{z})) = \exp(-\beta F(\mathbf{z})) = P(\mathbf{z})$$

$$P(\mathbf{z}) = P_b(\mathbf{z}) \exp(\beta \Delta V(\mathbf{z})) \langle \exp(-\beta \Delta V(Q(\mathbf{r}))) \rangle$$

$$\begin{aligned} P_b(\mathbf{z}) &= \frac{1}{\langle \exp(\beta F(\mathbf{z})) \rangle} = \frac{1}{\int d\mathbf{z} \exp(-\beta F) \exp(\beta F)} \\ &= \frac{1}{\int d\mathbf{z}} \quad \langle \exp(\beta F(\mathbf{z})) \rangle = \int d\mathbf{z} P(\mathbf{z}) \exp(\beta F) \end{aligned}$$

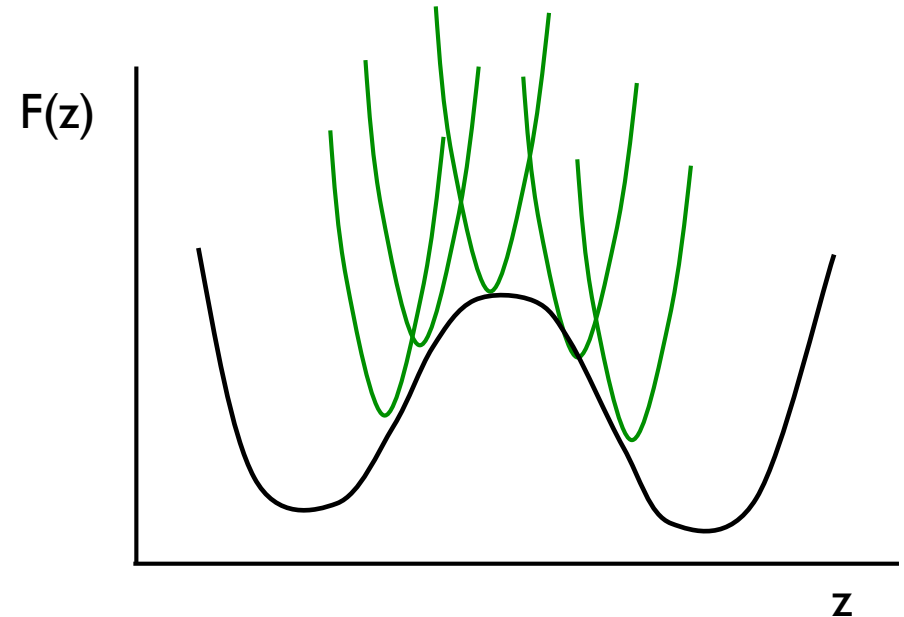
All states $\mathbf{z}$ become equiprobable

Umbrella sampling

Harmonic spring keeps the trajectory close to $\mathbf{z}_i$

$$\Delta V_i(\mathbf{r}) = \frac{1}{2} \kappa [Q(\mathbf{r}) - \mathbf{z}_i]^2$$

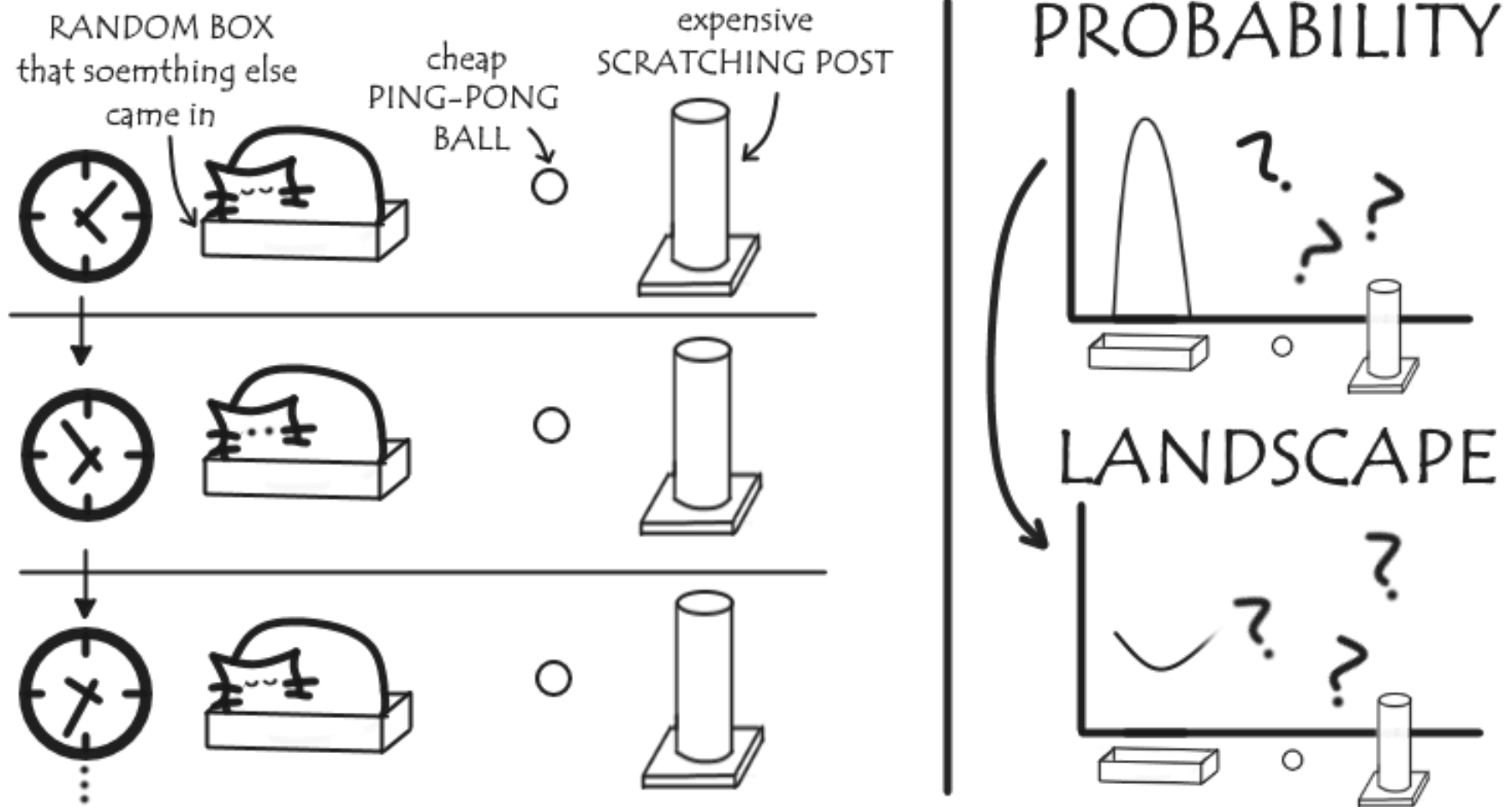
The points $\mathbf{z}_i$ and the value of κ must be chosen such that $Q(\mathbf{r}(t))$ makes excursions into the window of its neighbor.



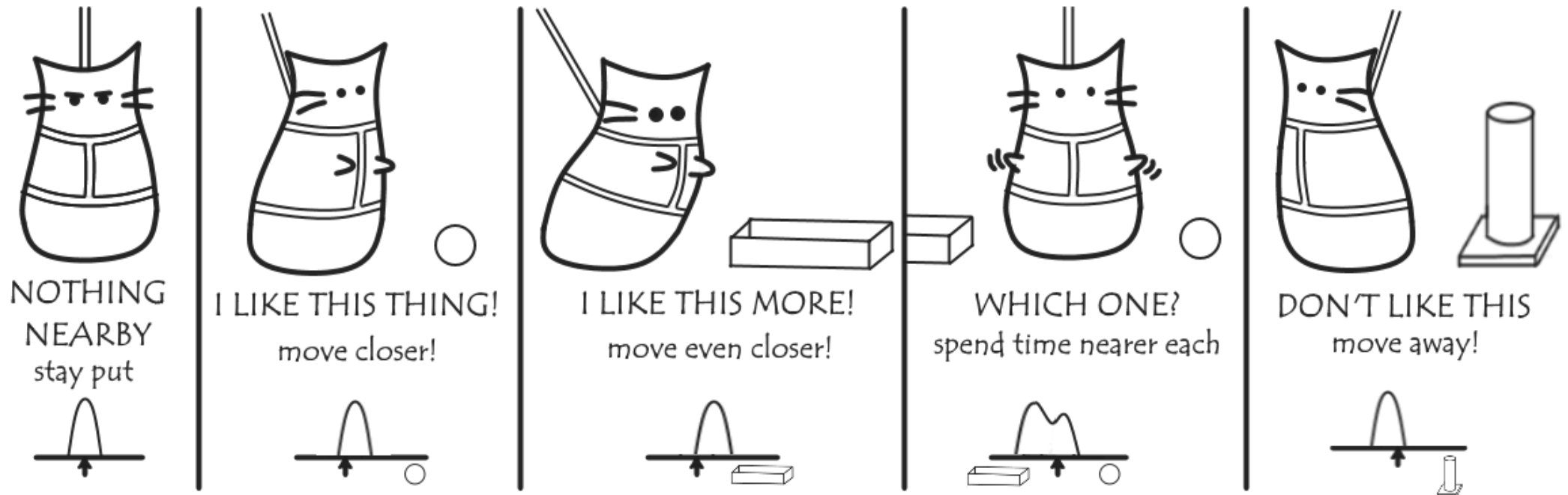
$$P(\mathbf{z}) = P_b(\mathbf{z}) \exp(\beta \Delta V(\mathbf{z})) \langle \exp(-\beta \Delta V(Q(\mathbf{r}))) \rangle$$

$$P_i(\mathbf{z}) = P_{b,i}(\mathbf{z}) \exp\left(\frac{1}{2} \beta \kappa |\mathbf{z} - \mathbf{z}_i|^2\right) \left\langle \exp\left(-\beta \frac{1}{2} \kappa |Q(\mathbf{r}) - \mathbf{z}_i|^2\right) \right\rangle$$

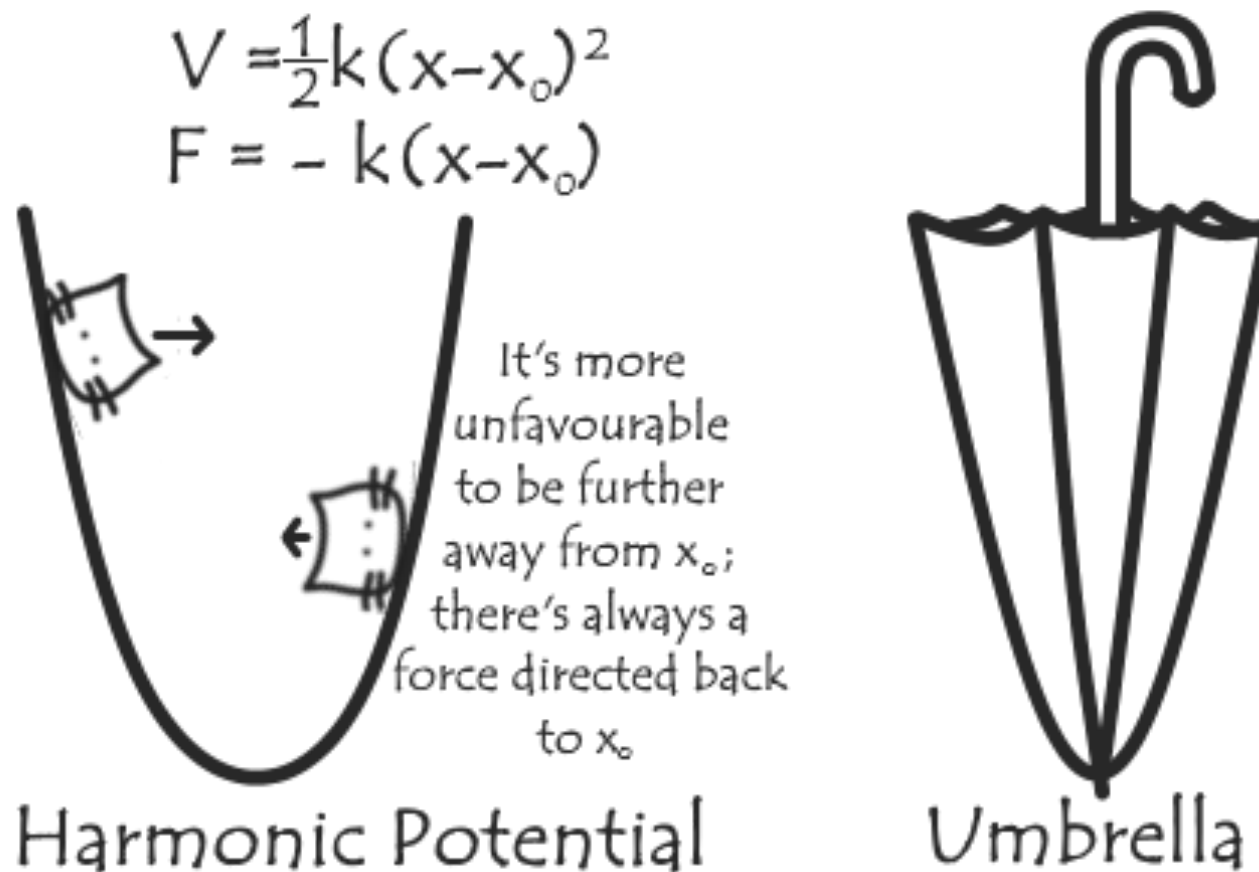
Molecular dynamics



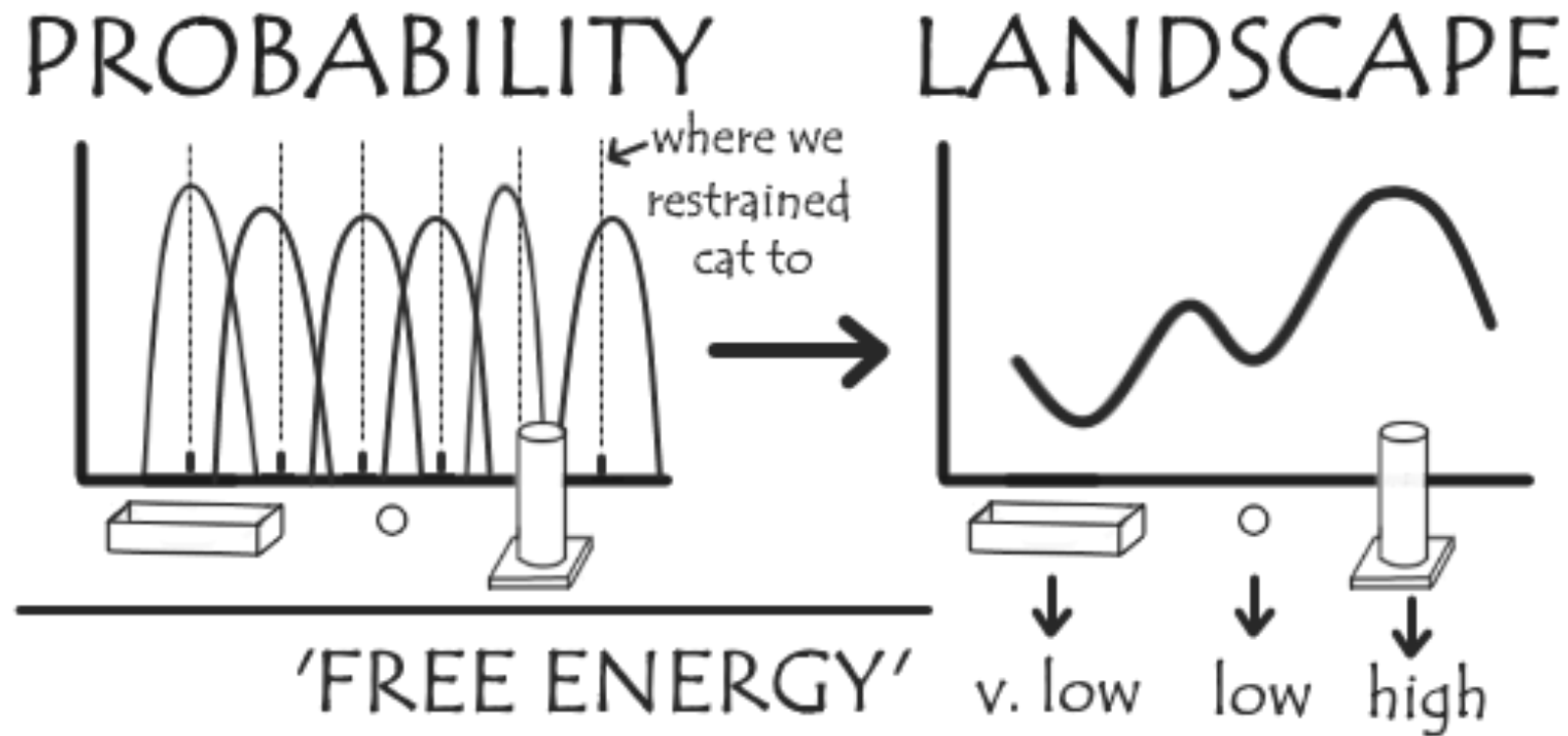
Umbrella sampling



Umbrella sampling



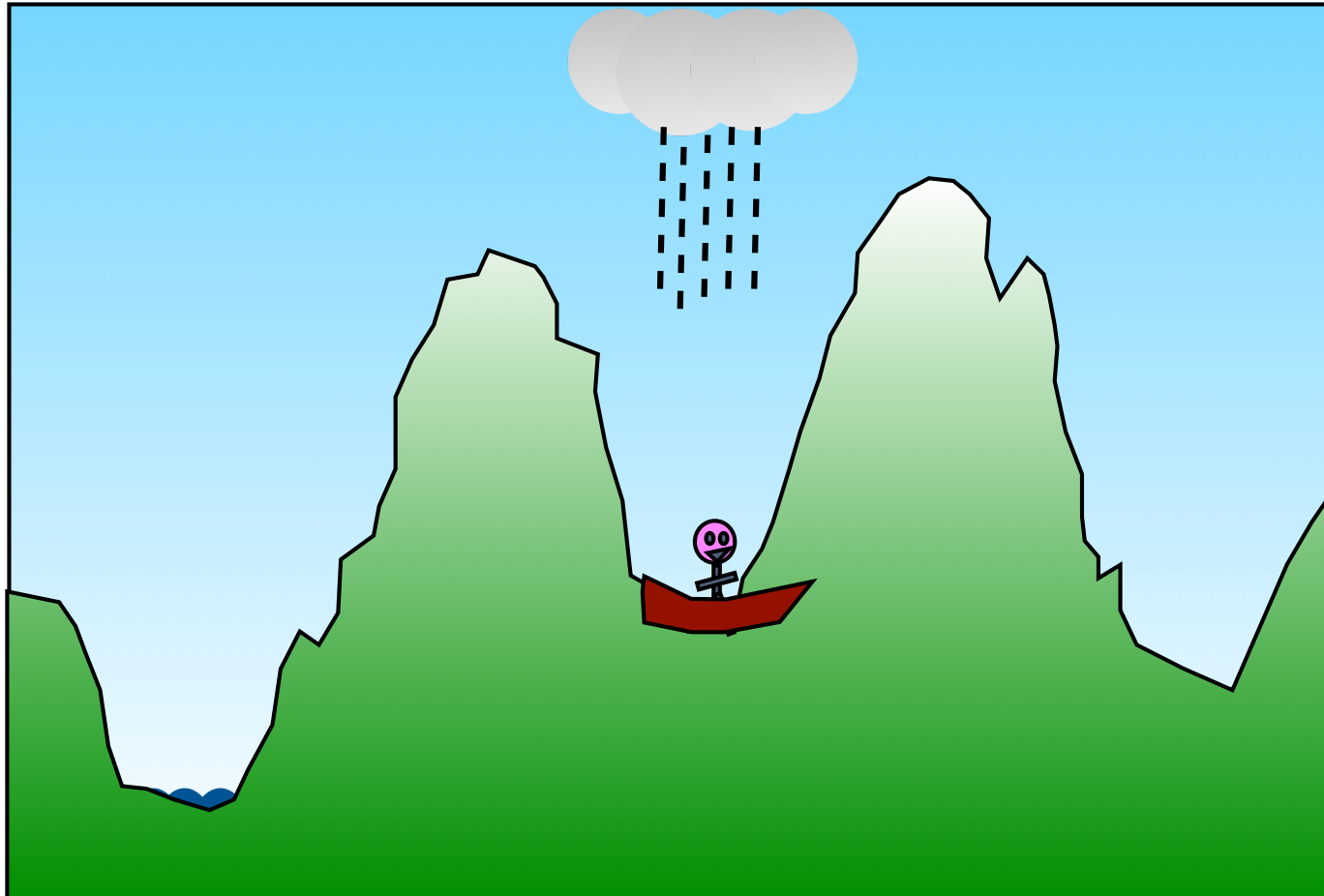
Umbrella sampling



Choosing ΔV

The closer the bias potential is to the negative free energy $-F(z)$ the more uniform the sampling of Q will be.

Metadynamics



Metadynamics

$$V(\mathbf{r}) = V_0(\mathbf{r}) + \Delta V(Q(\mathbf{r}), t)$$

force field

bias potential

The bias potential acts on collective variables Q , which approximate the reaction coordinate.

$$\Delta V[Q(\mathbf{r}), t] = w \sum_{t' < t} \exp \left(-\frac{|Q[\mathbf{r}(t) - \mathbf{r}(t')]|^2}{2\delta Q^2} \right)$$

$t' = \tau_G, 2\tau_G, \dots$

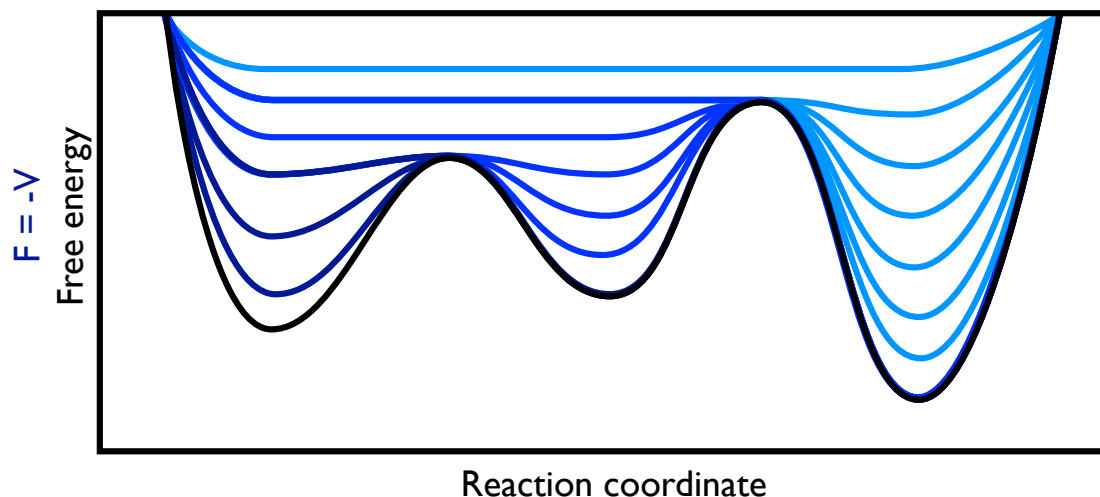
w height of the Gaussian

τ_G time interval between depositions

δQ Gaussian width

The bias is built as a sum of repulsive Gaussian functions, centered on the points in collective variable space already visited.

Metadynamics



$$V(\mathbf{r}) = V_0(\mathbf{r}) + \Delta V(\mathbf{r}, t)$$

The bias is built as a sum of Gaussian functions centered on the points in CV space already visited.

$$\Delta V[Q(\mathbf{r}), t] = w \sum_{t' < t} \exp \left(-\frac{|Q[\mathbf{r}(t) - \mathbf{r}(t')]|^2}{2\delta Q^2} \right)$$

$$t' = \tau_G, 2\tau_G, \dots$$

w height of the Gaussian

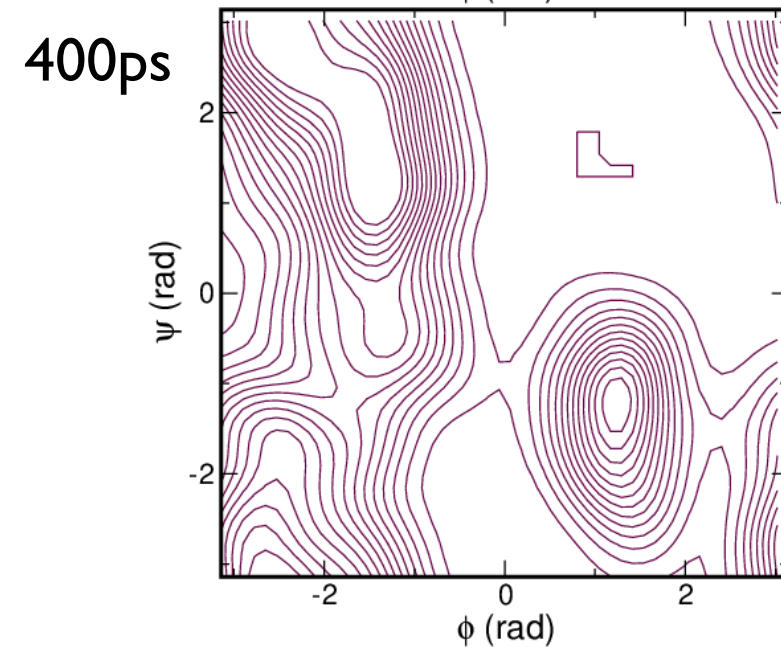
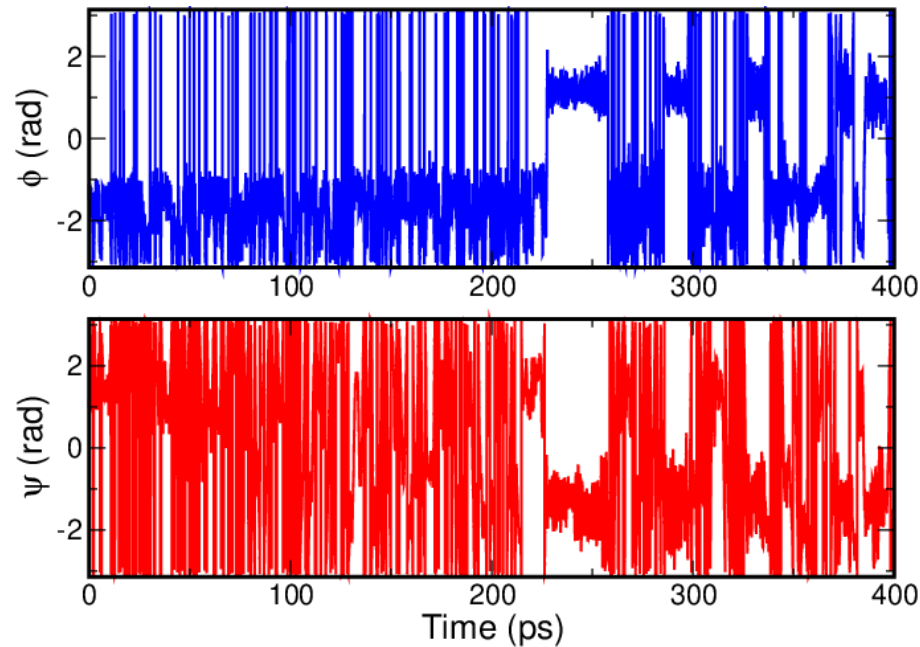
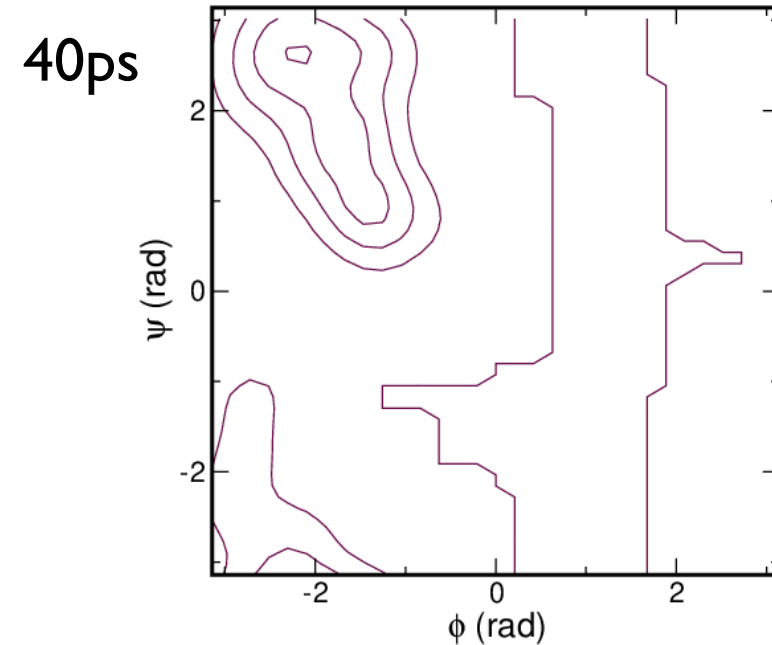
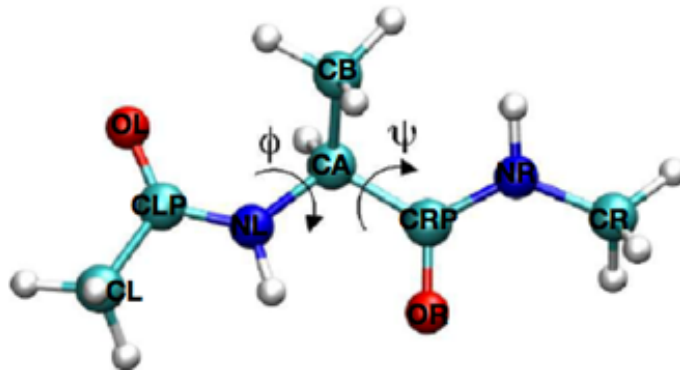
τ_G time interval between depositions

δQ Gaussian width

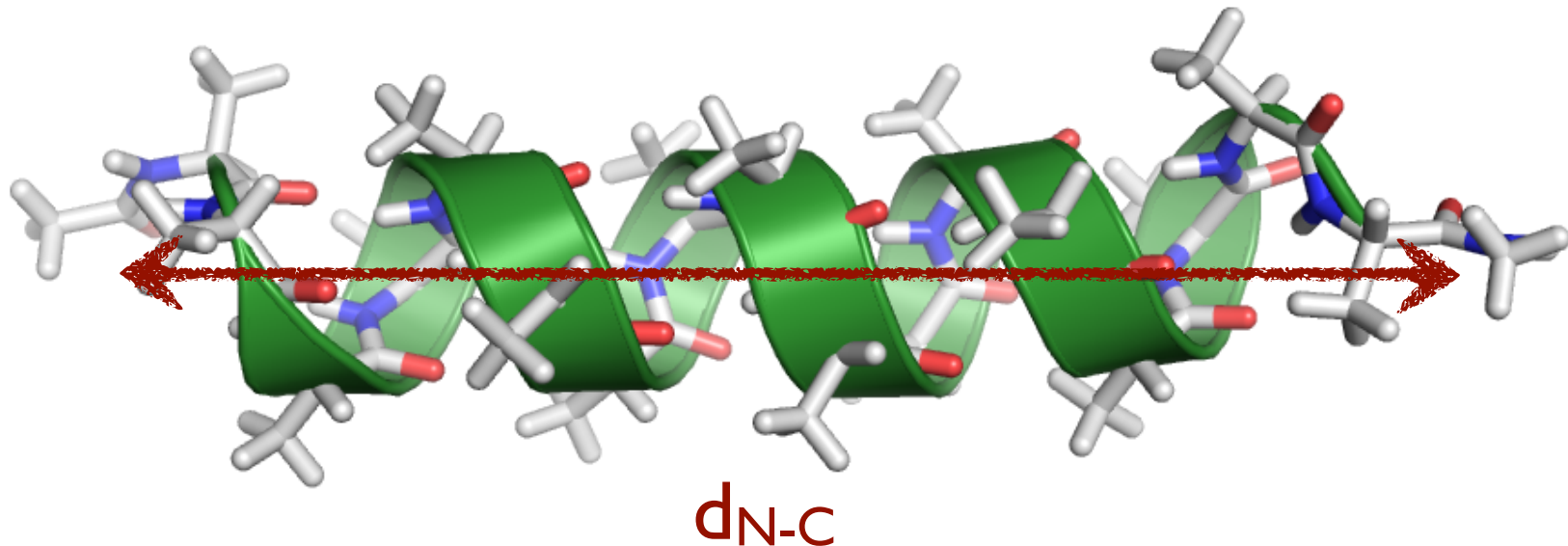
$$F = - \lim_{t \rightarrow \infty} \Delta V(Q(\mathbf{r}), t)$$

64

Alanine dipeptide in vacuum

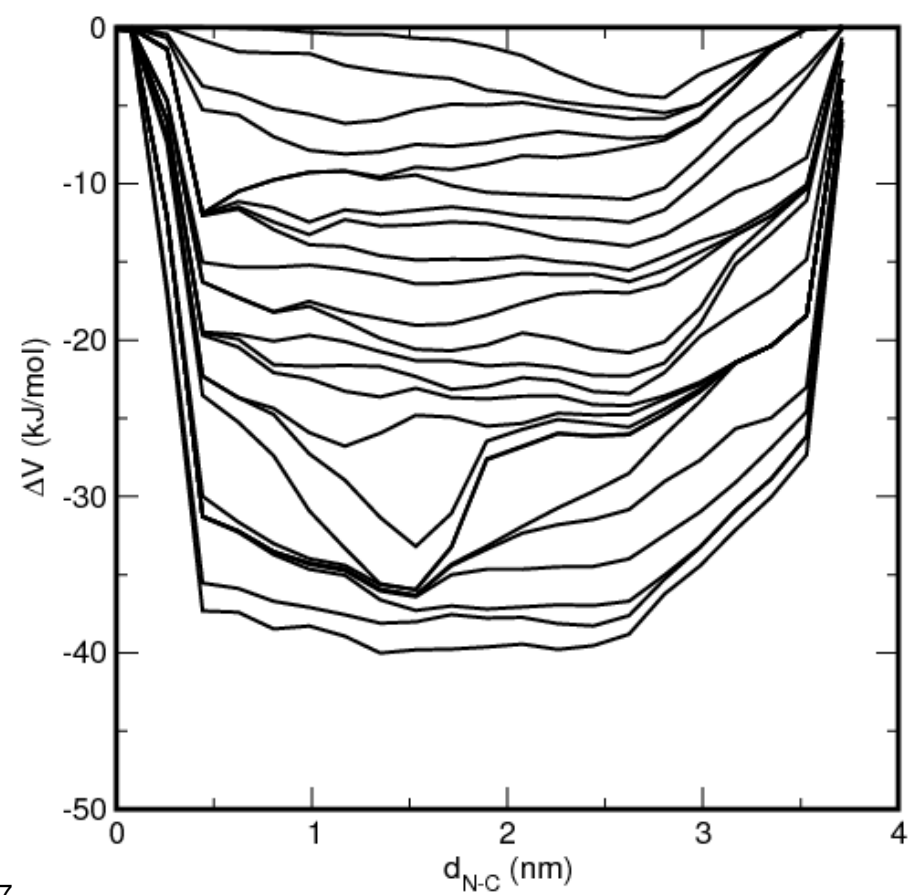
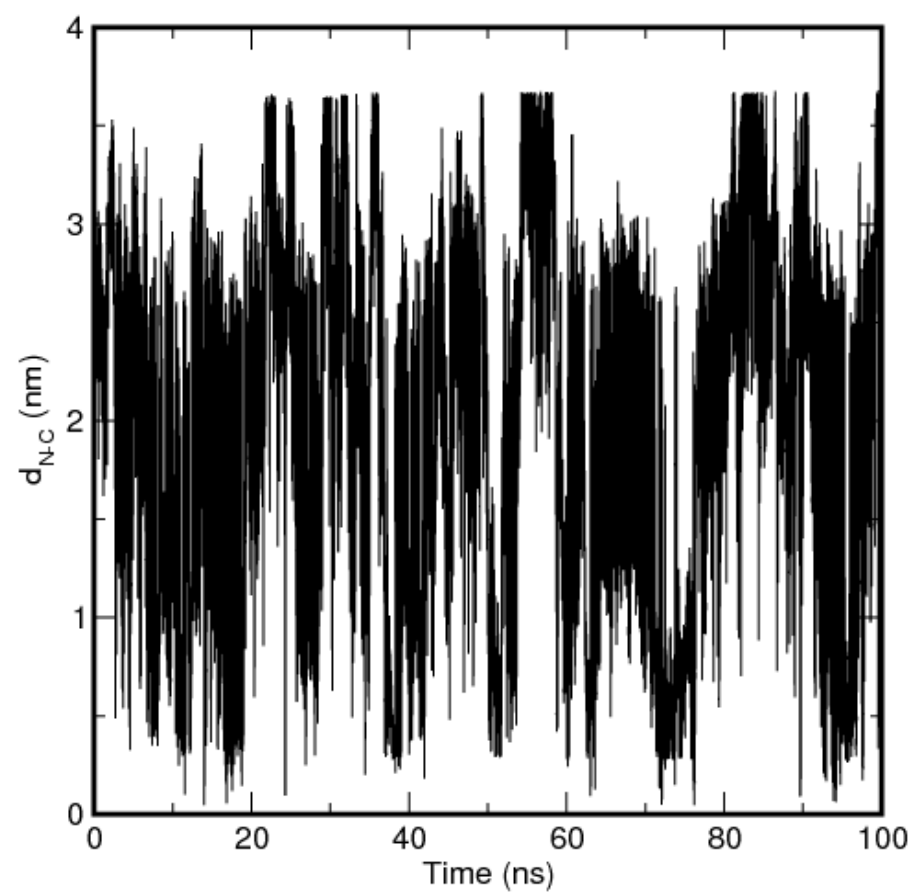
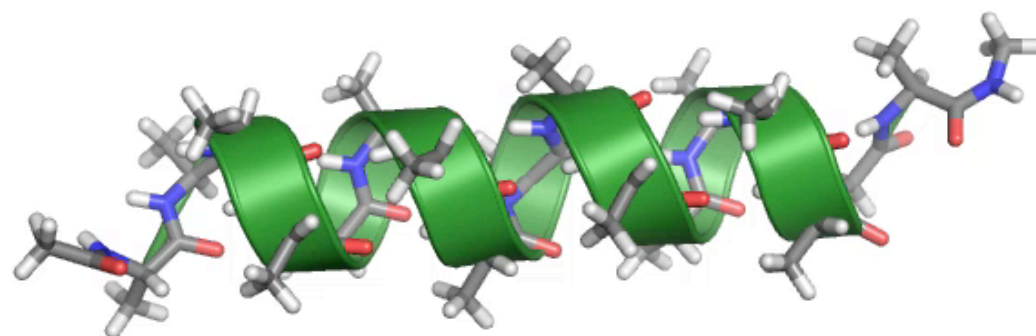


Setup of a metadynamics simulation



$$\Delta V[Q(\mathbf{r}), t] = w \sum_{t' < t} \exp \left(- \frac{|Q[\mathbf{r}(t) - \mathbf{r}(t')]|^2}{2\delta Q^2} \right)$$
$$t' = \tau_G, 2\tau_G, \dots$$

w height of the Gaussian
 τ_G time interval between depositions
 δQ Gaussian width

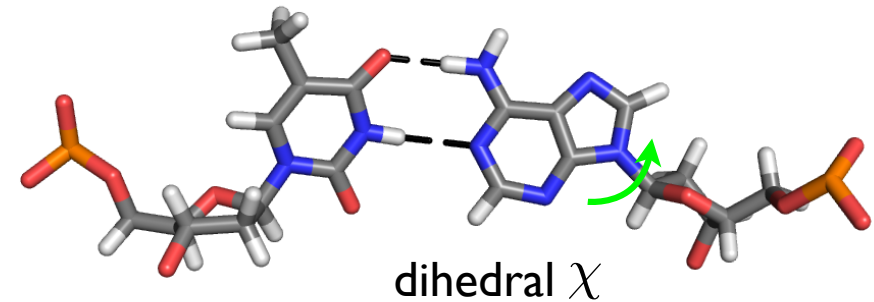


Setup of a metadynamics simulation

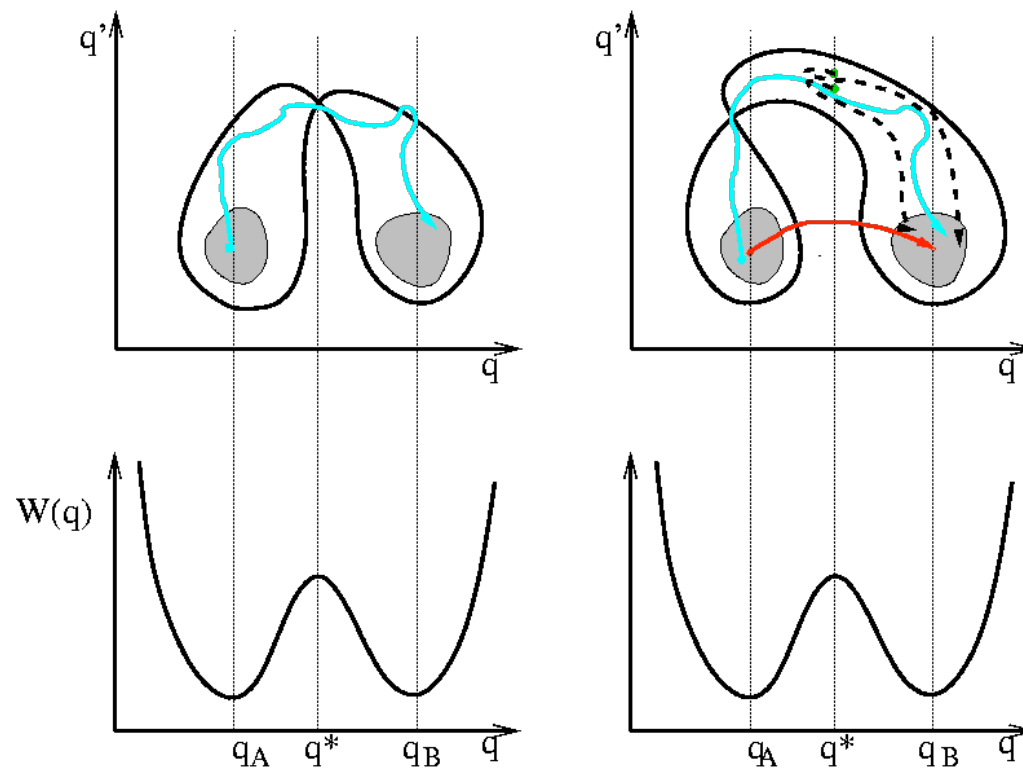
No convergence

Instead, distortion and unfolding
of the DNA fragment

These metadynamics attempts failed...



The reaction coordinate problem



Q?

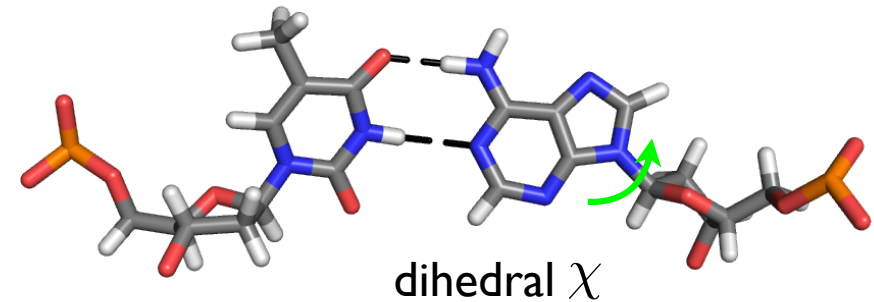
λ ?

Metadynamics can give free energy, if converged

No convergence

Instead, distortion and unfolding
of the DNA fragment

These metadynamics attempts failed...

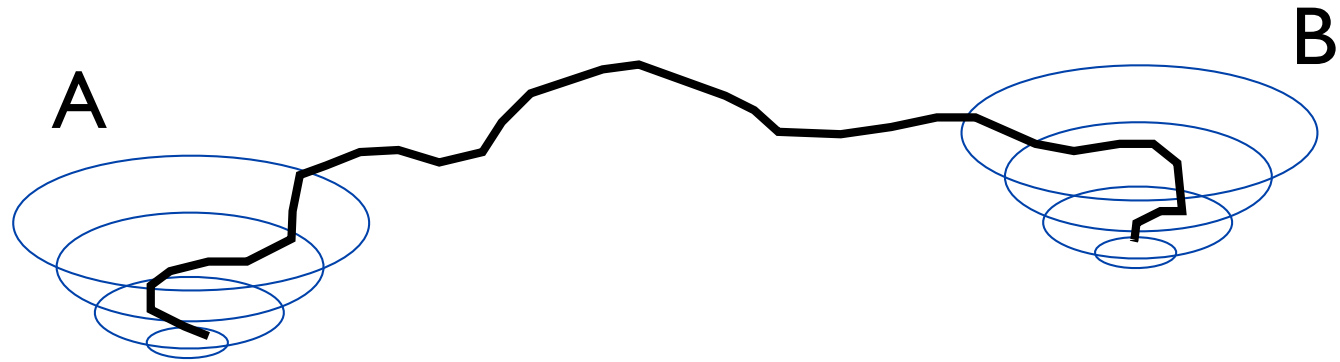


... instead, use the adaptive path collective variable (path cv)

- a function of other collective variables
- bias potential works on the path cv
- adapts during the simulation

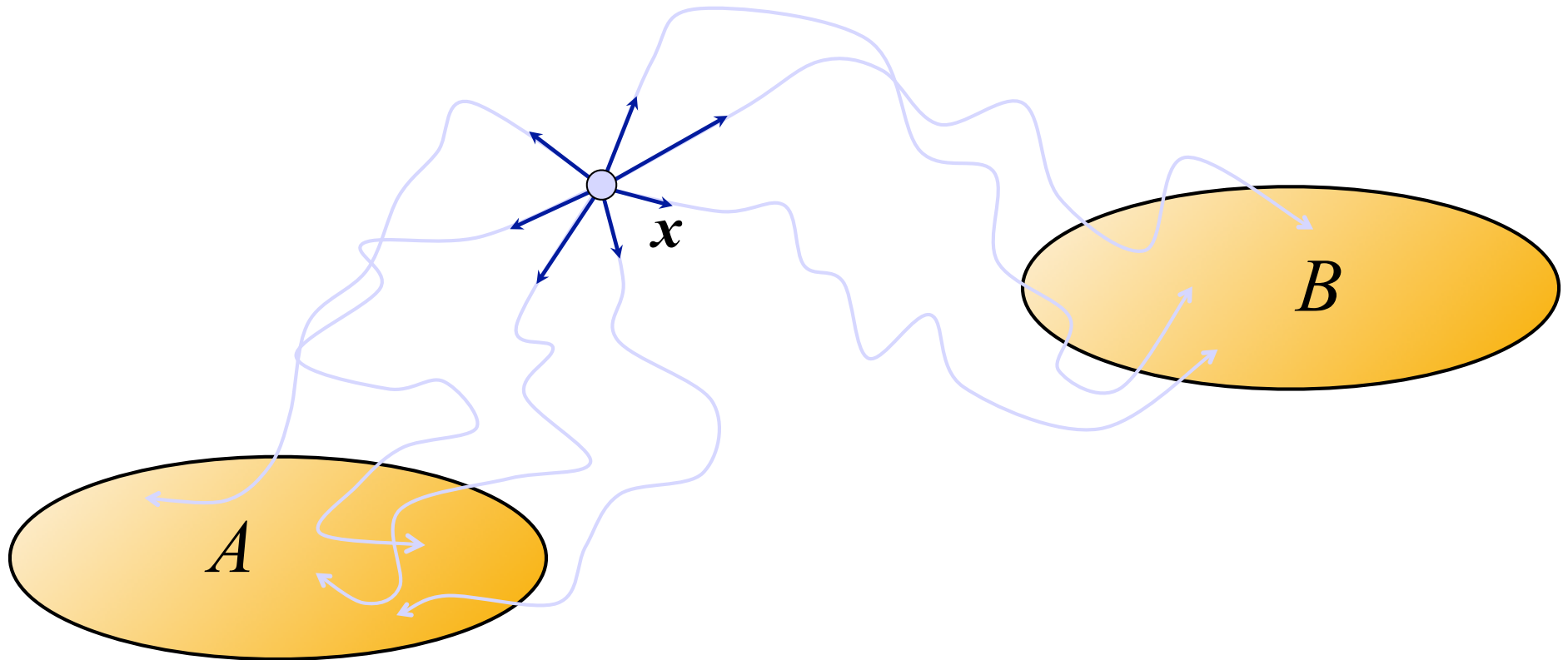
path-metadynamics

Two state system in a CV space



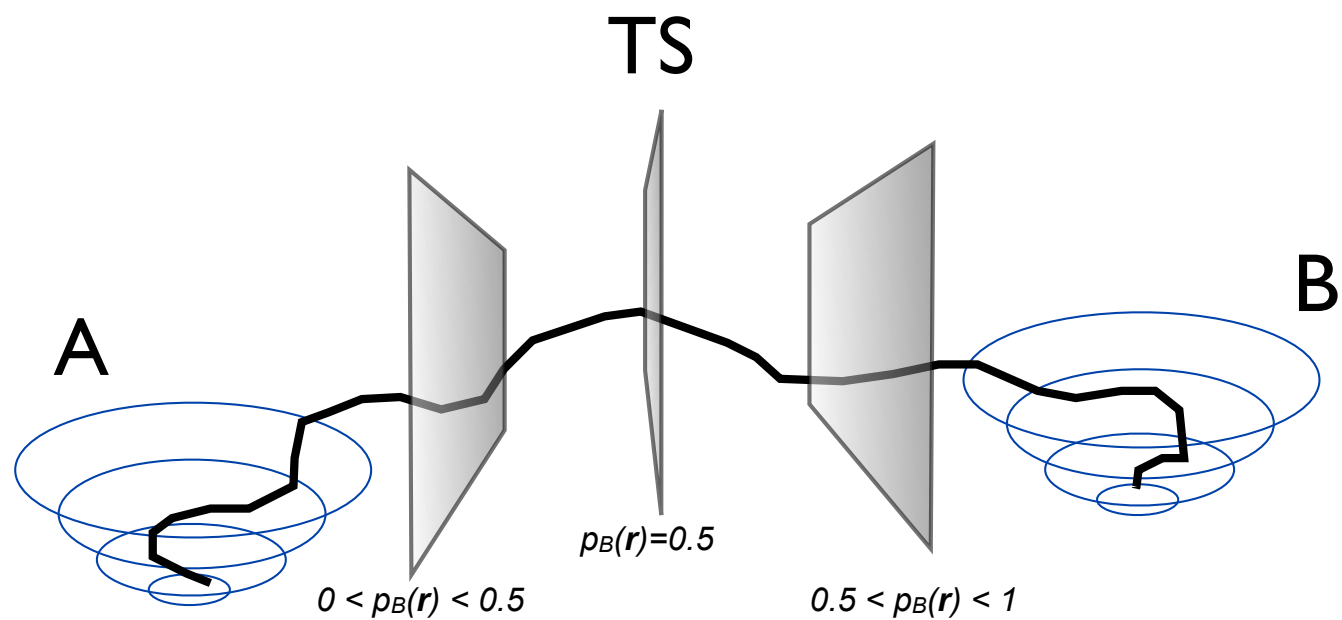
σ - path in CV space

Committer probability p_B



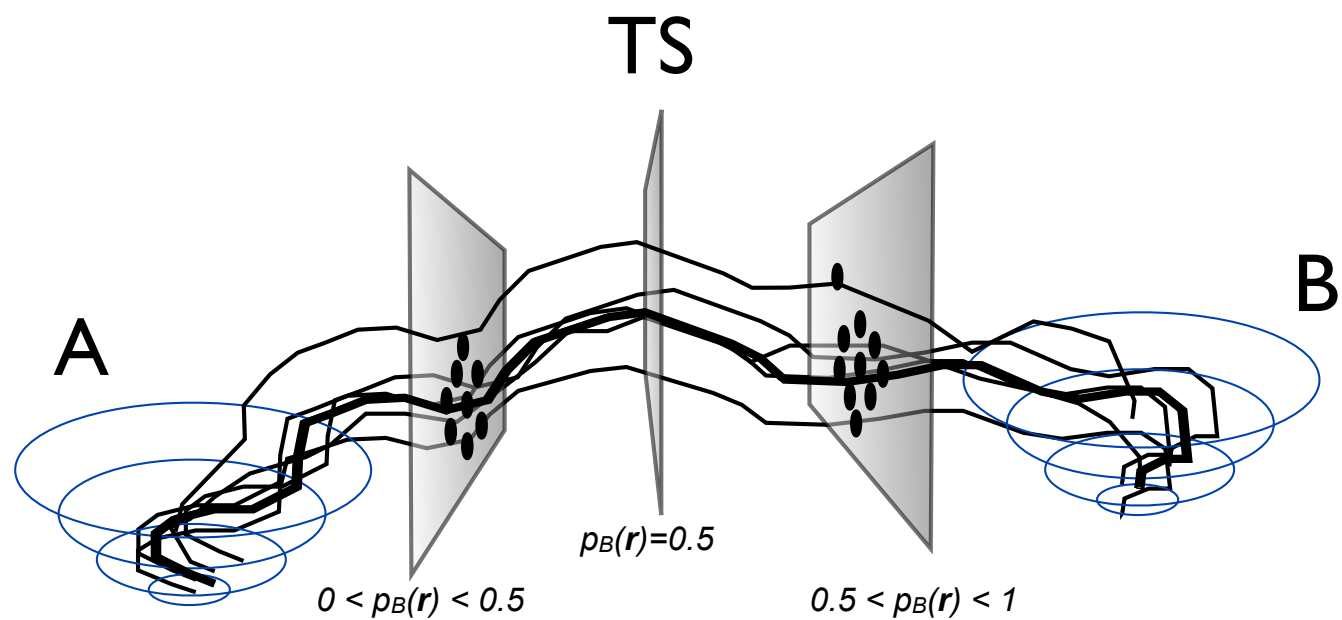
Probability that a trajectory initiated at x
(random velocities) ends in state B

Iso-committor surfaces

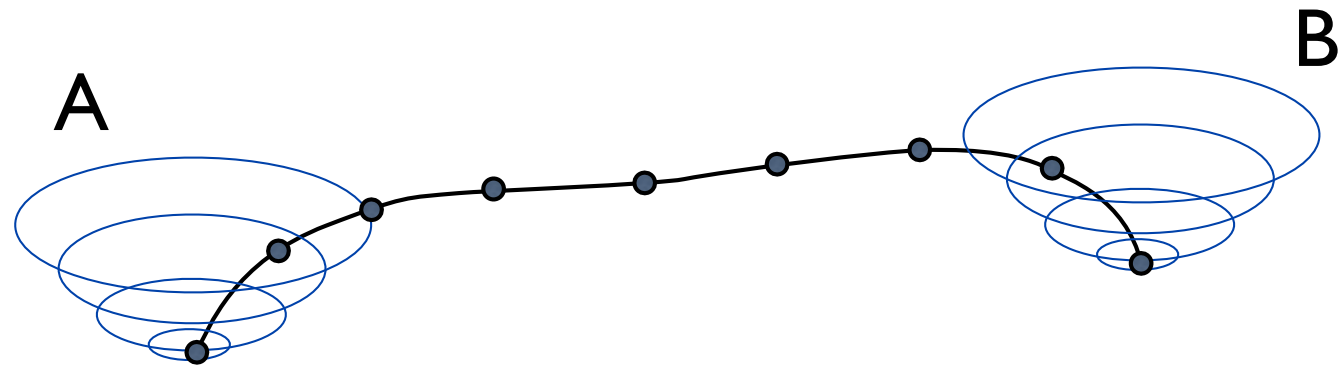


σ - path in CV space

The average transition pathway

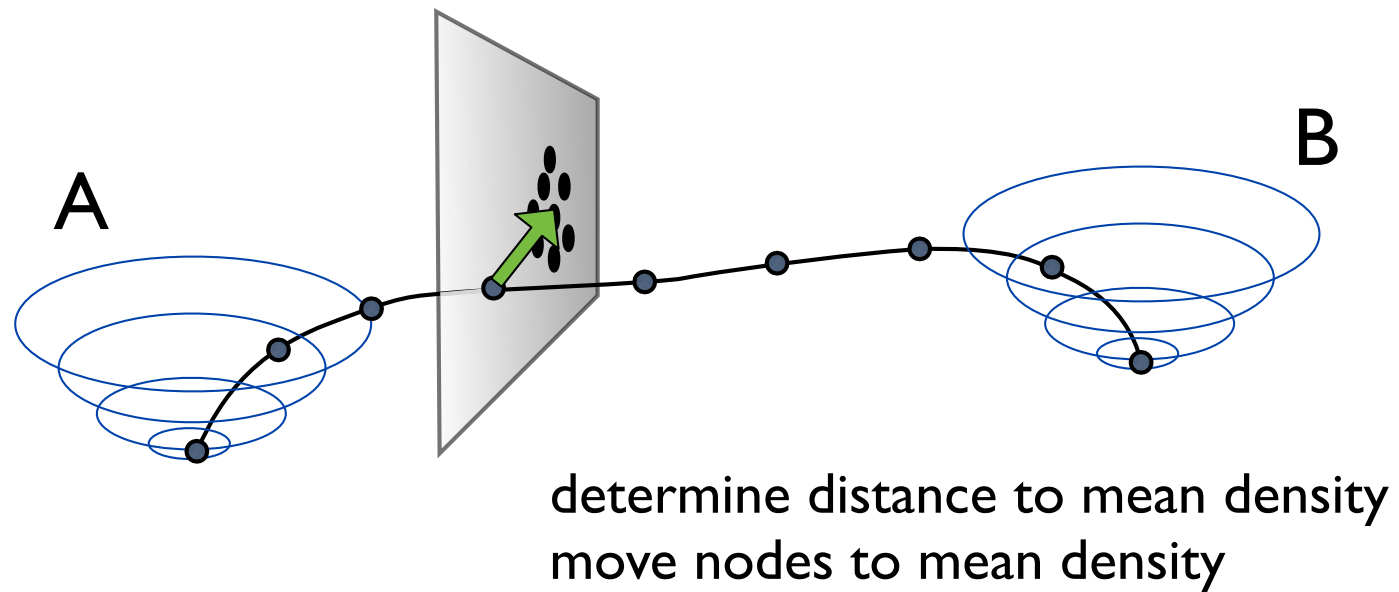


Path metadynamics (i)



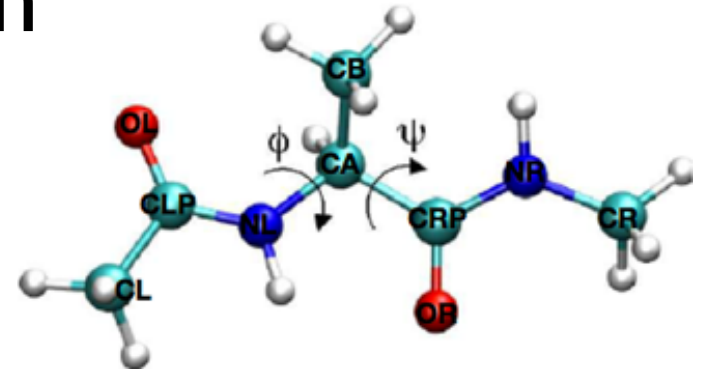
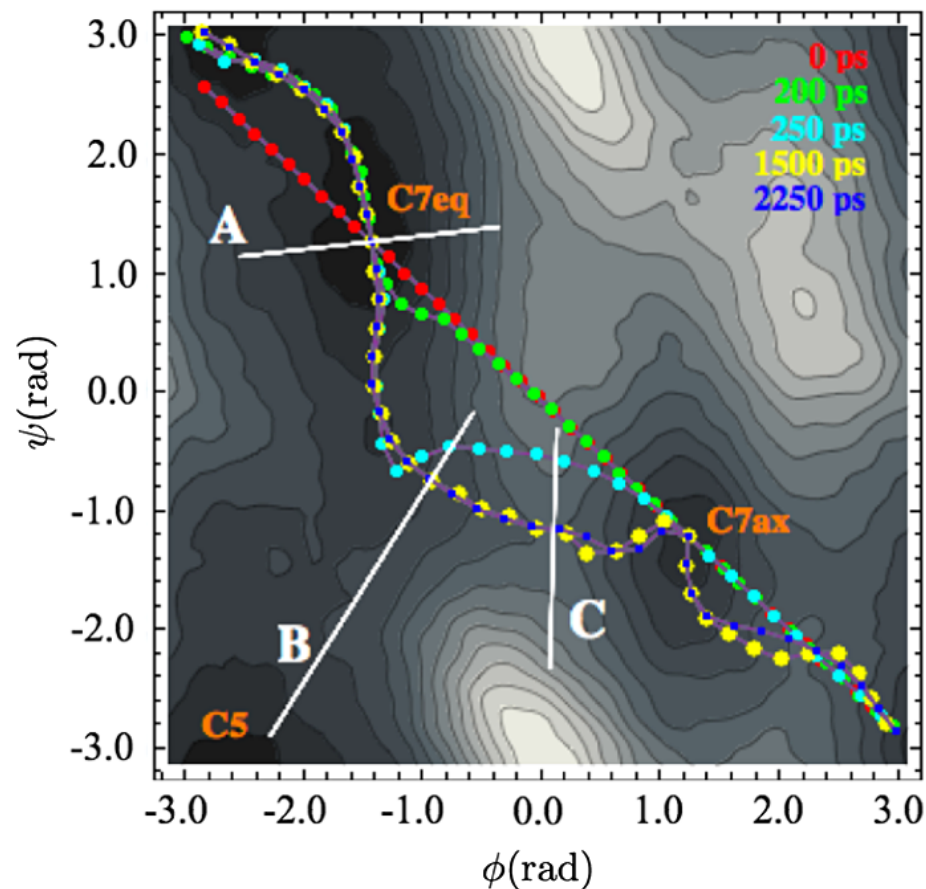
construct a path in CV space by placing nodes
perform metadynamics along the path

Path-metadynamics (ii)

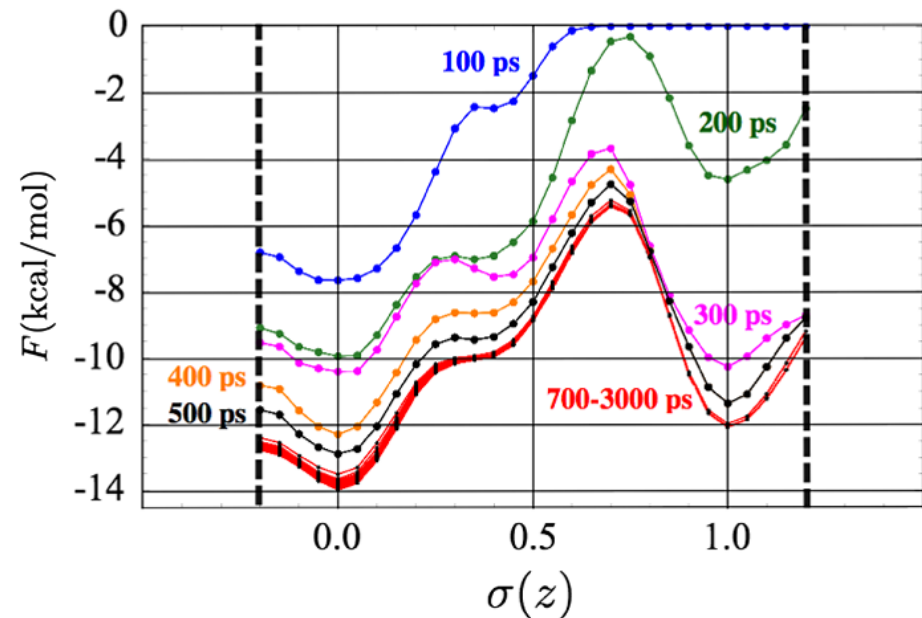


Path-metadynamics of alanine dipeptide in vacuum

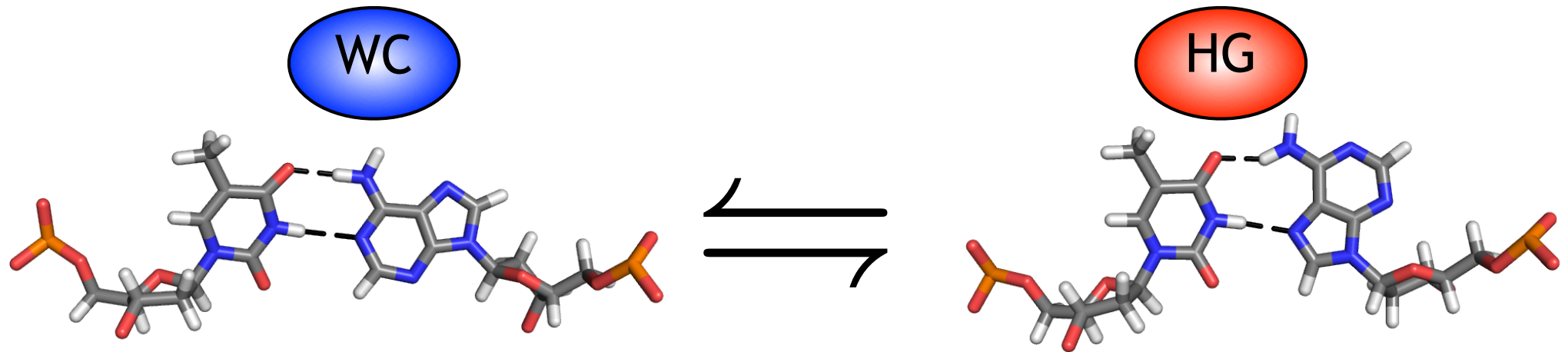
path in cv space



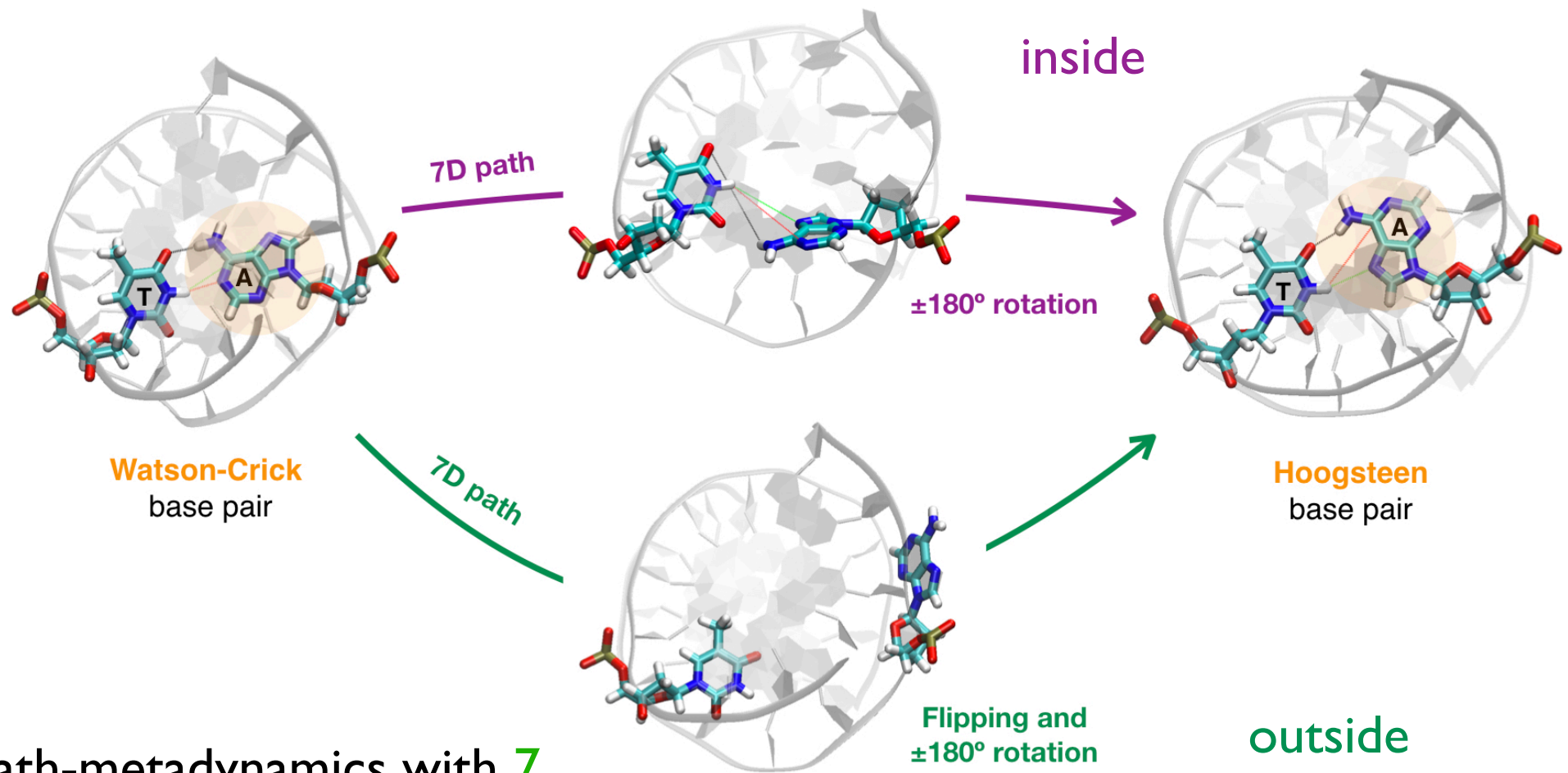
free energy along the path



Back to DNA baserolling

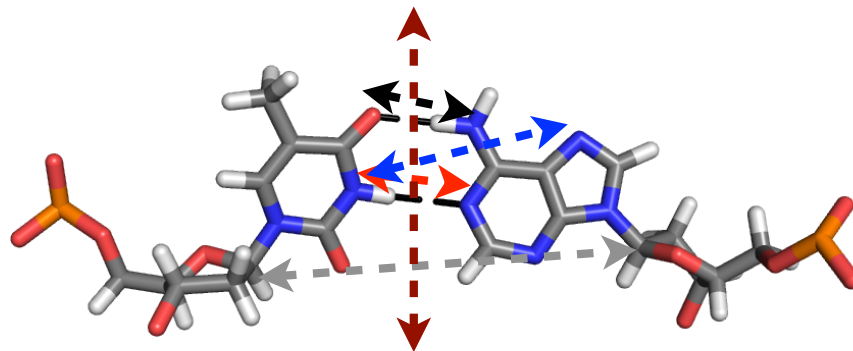


Adaptive path collective variable



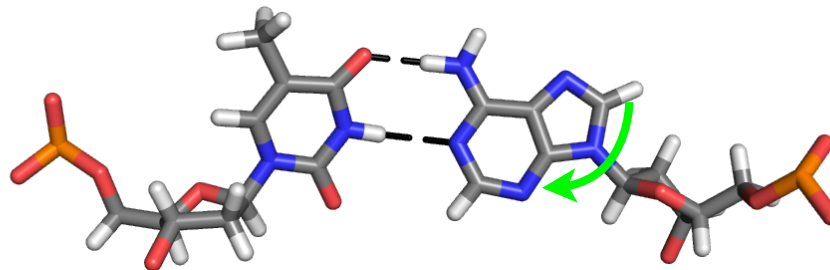
Path-metadynamics with 7 CVs including H-bonding, rolling, flipping, breathing...

Biasing DNA baserolling defining CVs

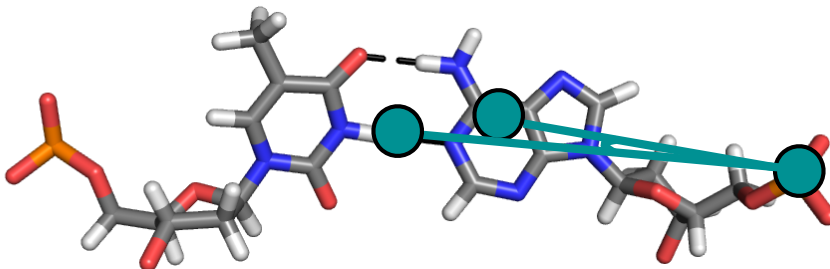


distances:

d_{WC} , d_{HG} , d_{HB} , d_{CC} , d_{NB}



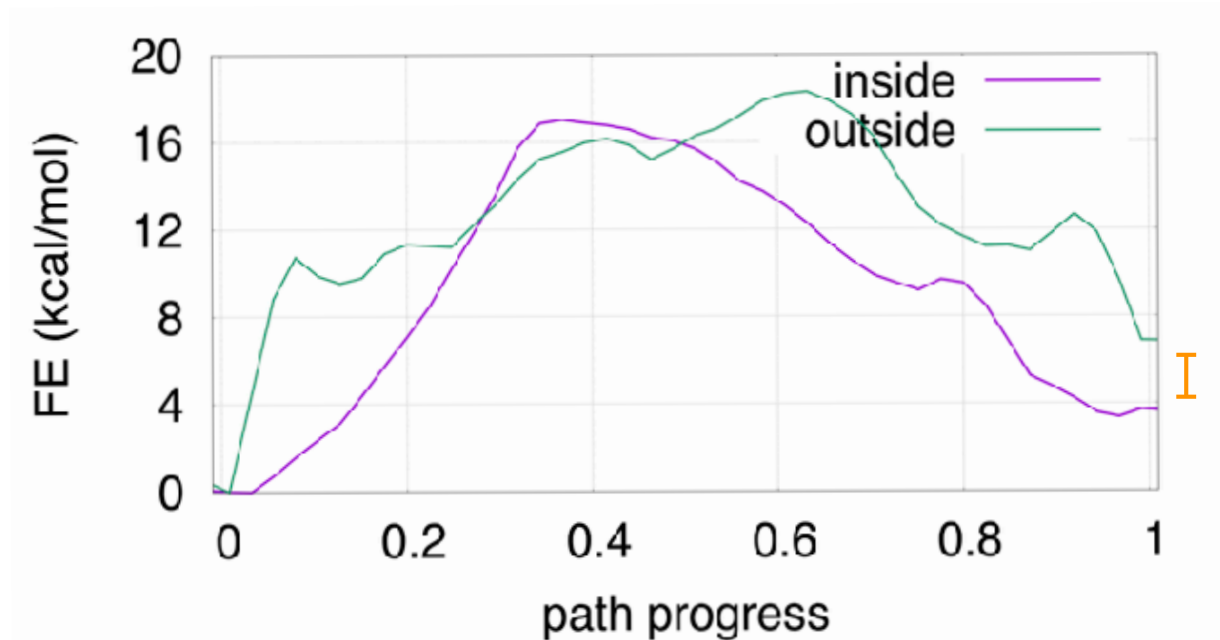
base rolling angle



base flipping angle

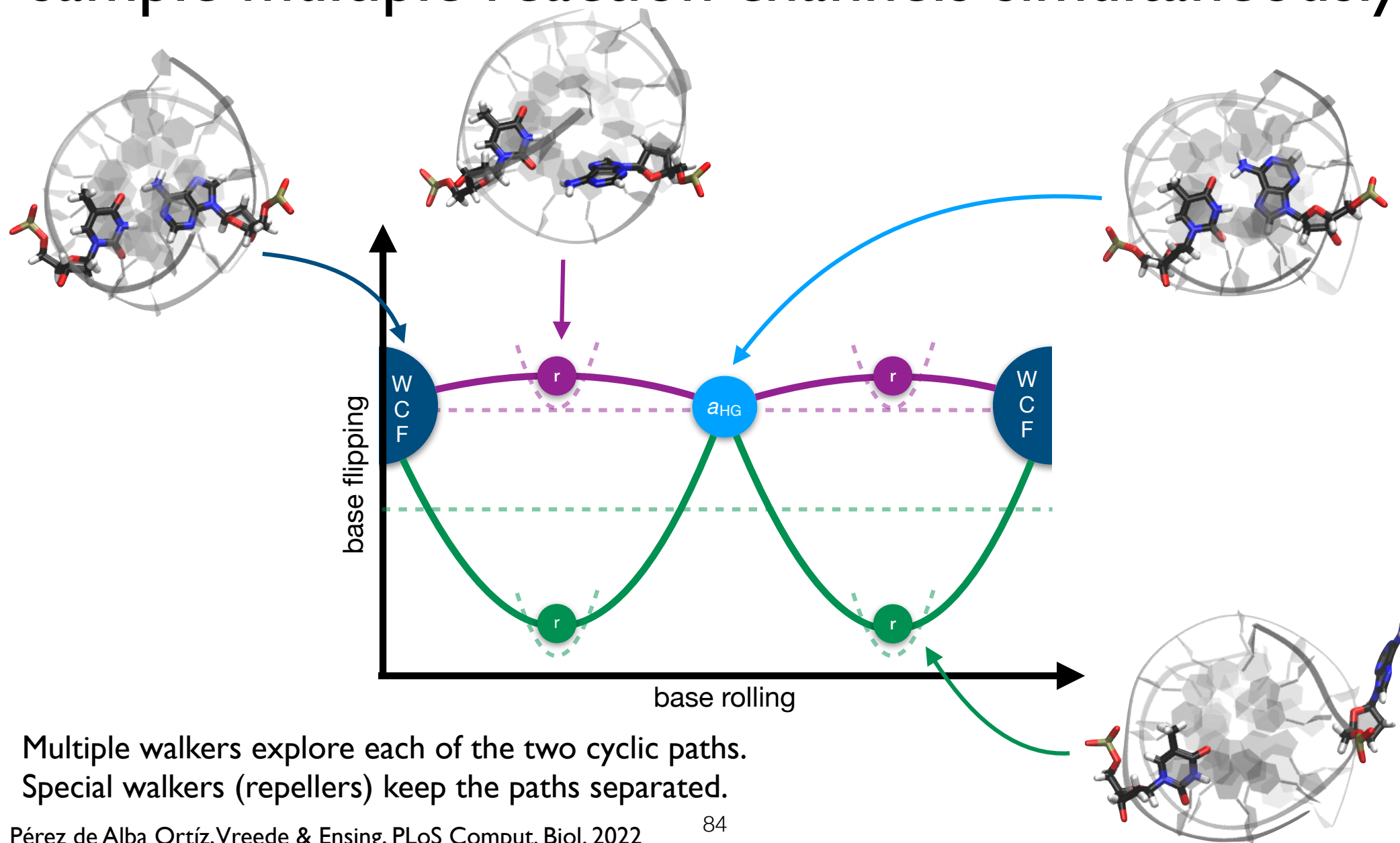
Path-metadynamics with 7 CVs including H-bonding, rolling, flipping and breathing.

Free energy profiles



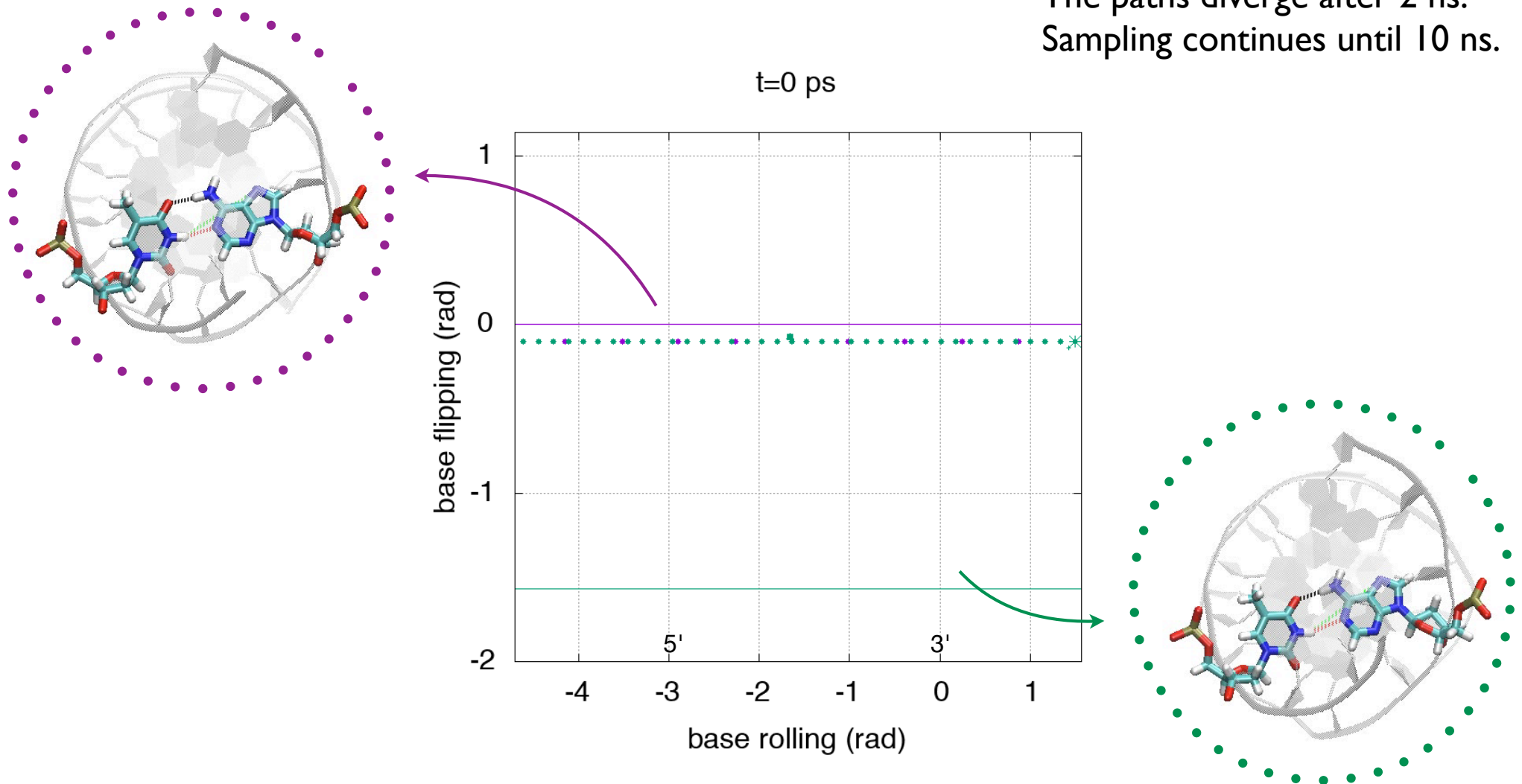
- Path-metadynamics with 7 CVs including H-bonding, rolling, flipping, breathing.
- No clear preference for inside or outside
- **Lessons in path-metadynamics:**
 - Difficulty in capturing one out of several competing mechanisms
 - Entropic penalty due to the tube potential acting in unnecessary CVs

Multi-path-metadynamics: sample multiple reaction channels simultaneously

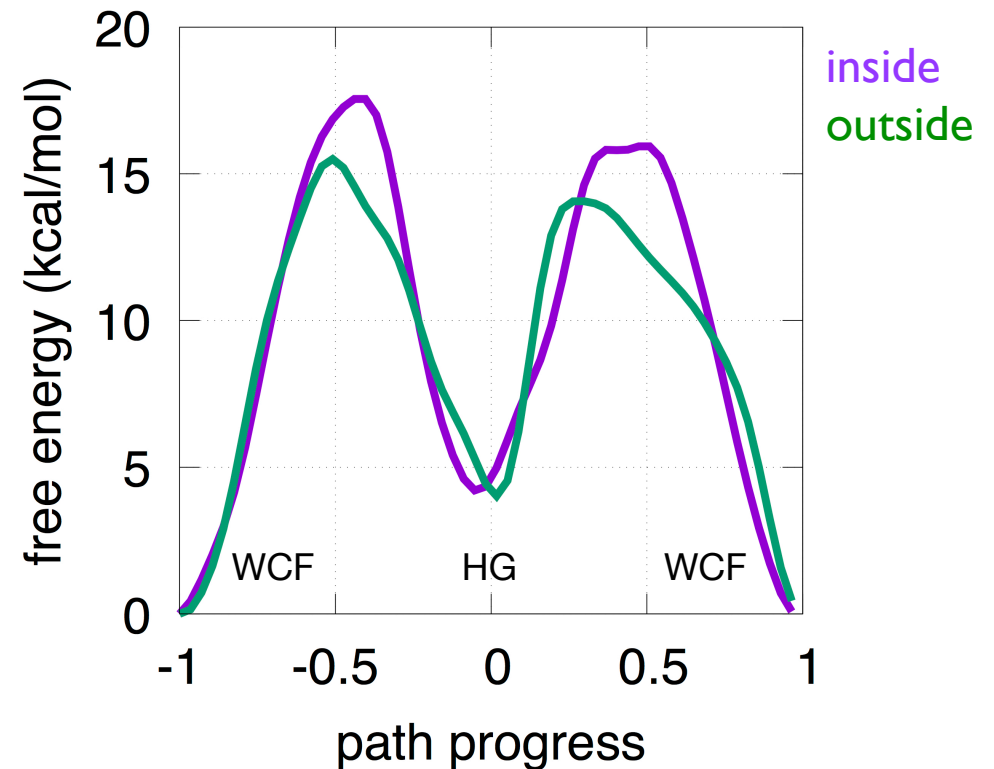
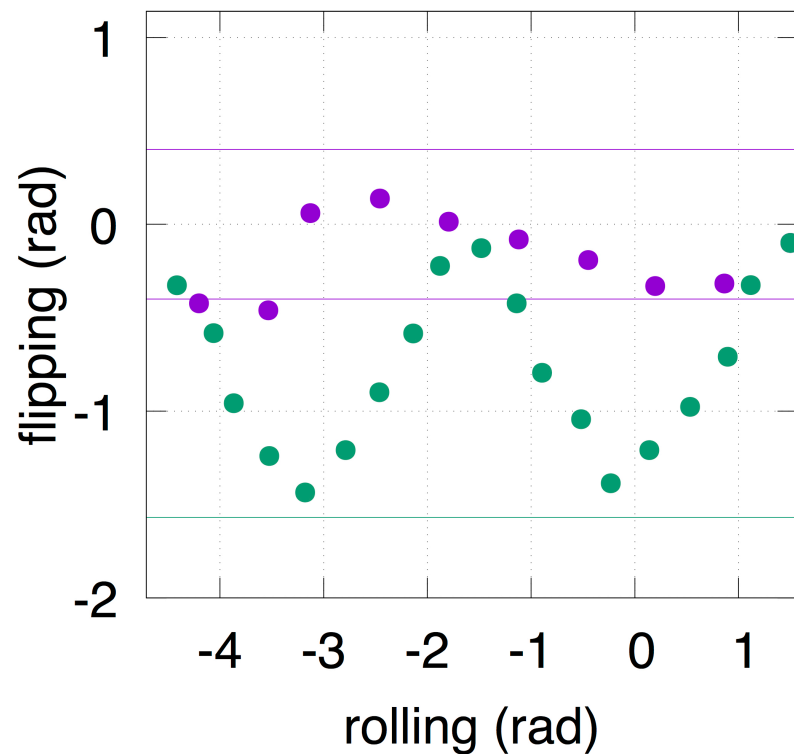


DNA base rolling with PMD

The paths diverge after 2 ns.
Sampling continues until 10 ns.

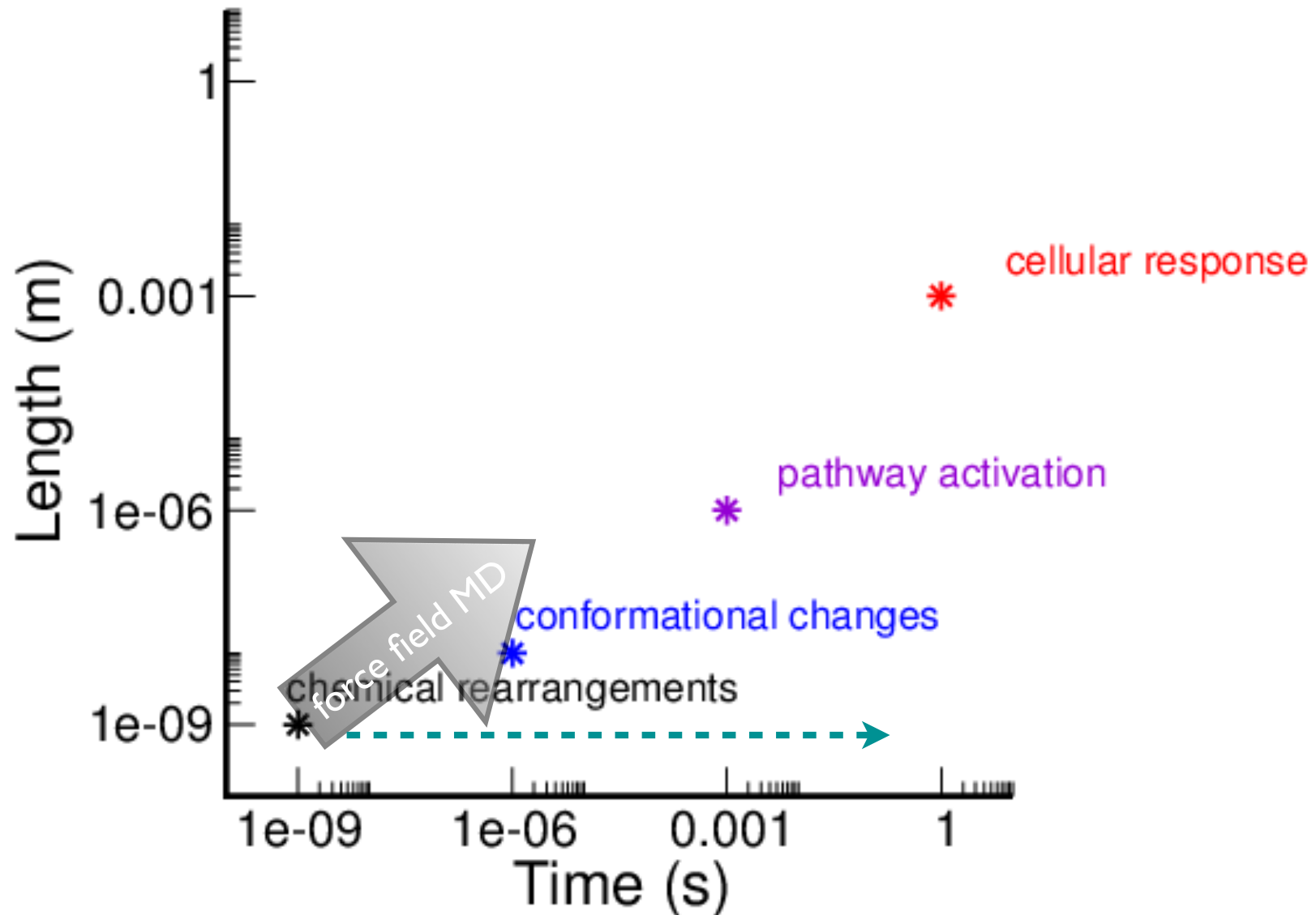


DNA baserolling with path-metadynamics



Free energy difference between WCF and HG close to experimental value.
Outside path has lower barriers, similar to TPS results.

Bridging the time gap



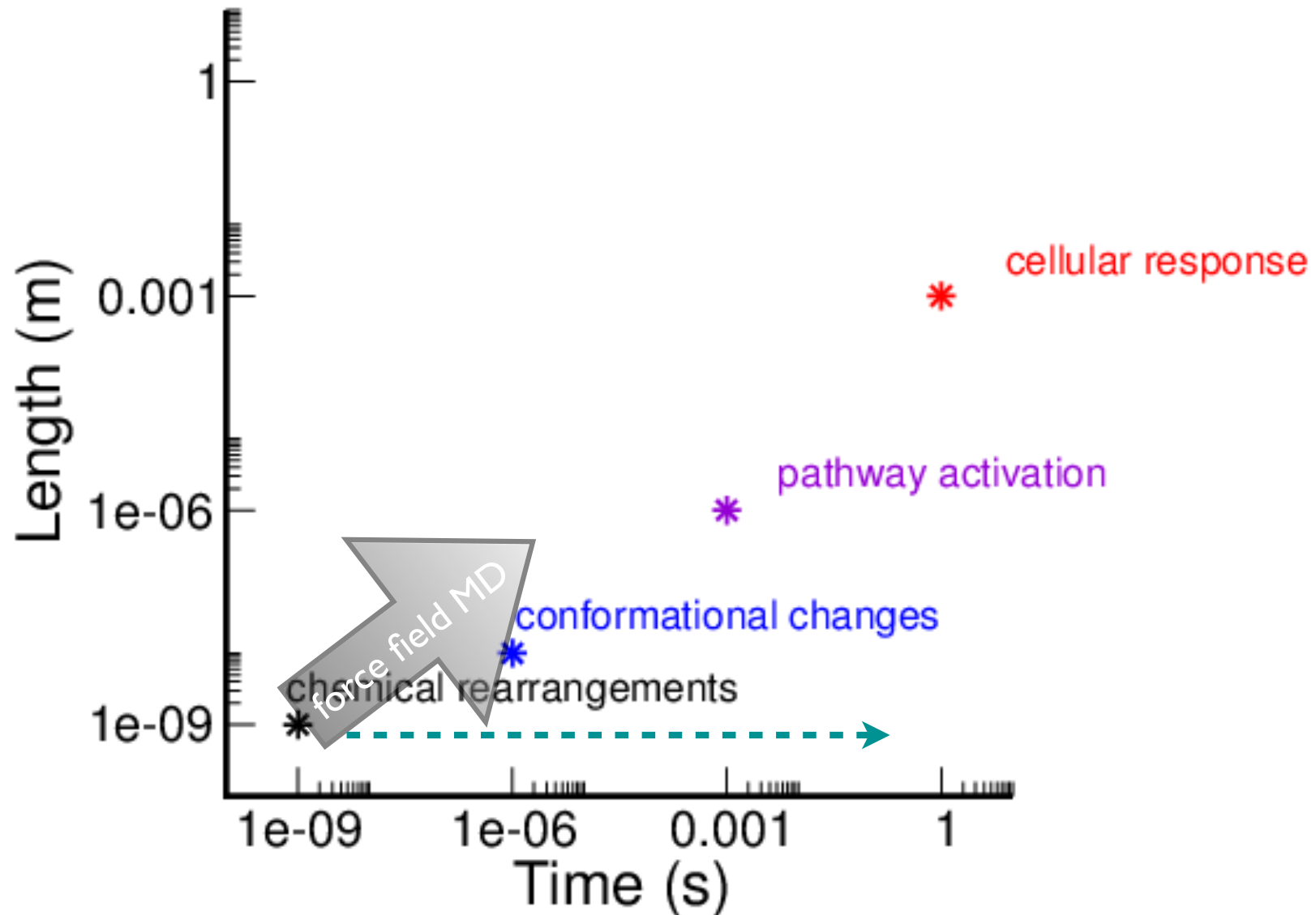
Biomolecular simulation Outline

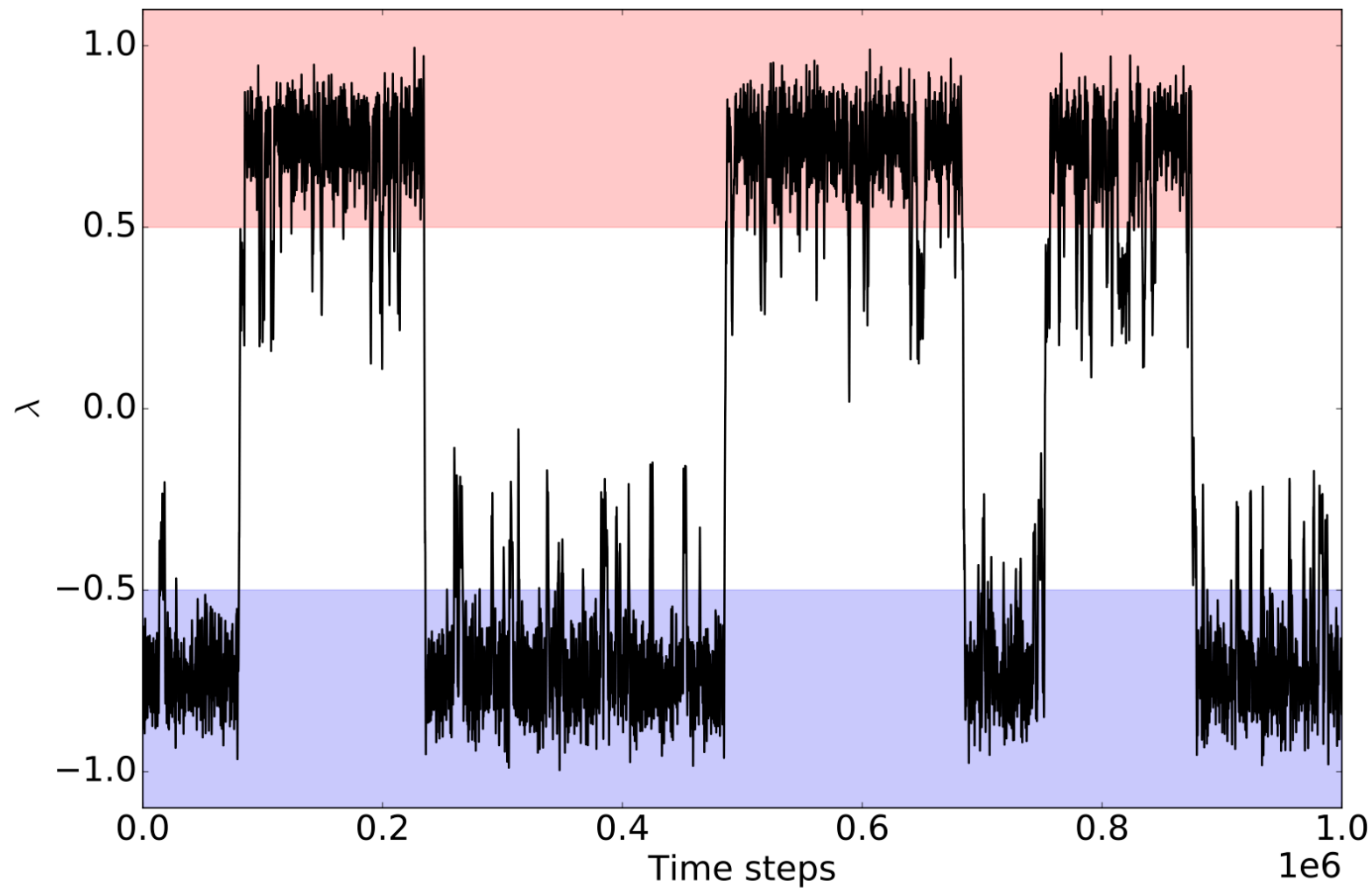
Part 1: Simulating biomolecular systems

Part 2: Biased sampling

Part 3: Path sampling

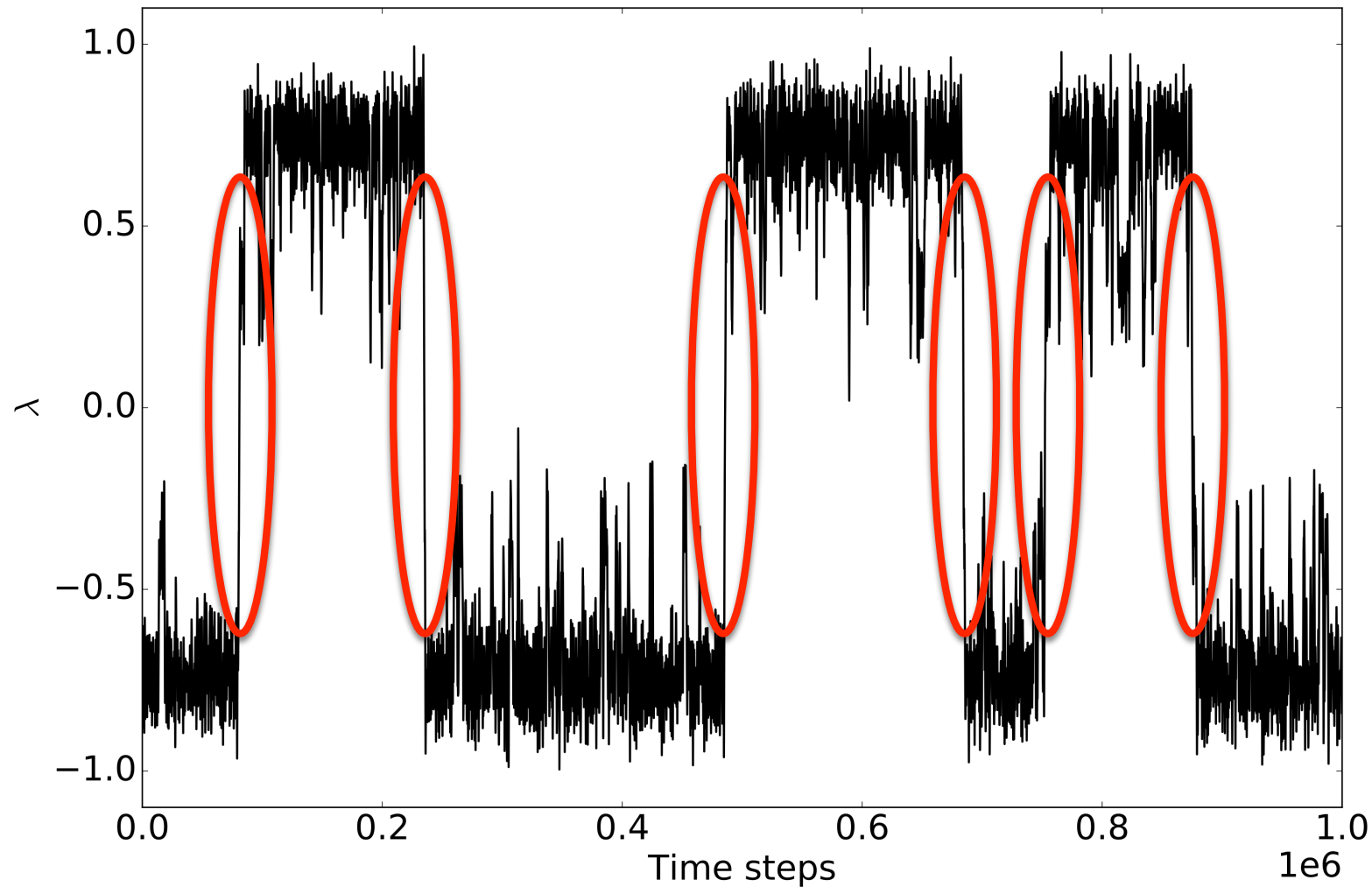
Bridging the time gap





Separation of time scales:
fundamental time scale₉₀ vs. frequency of event

Path sampling



Focus on the **transition** regions

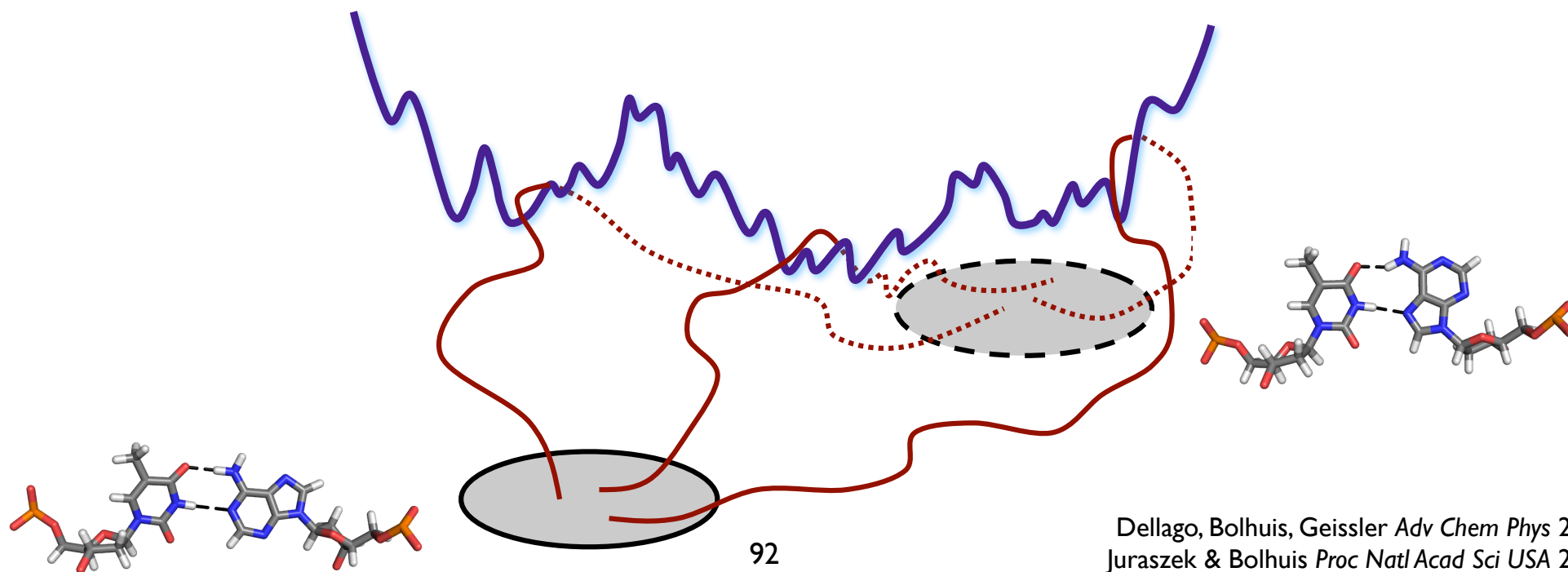
Monte Carlo in path space

Sampling rare events: transition path sampling

Input

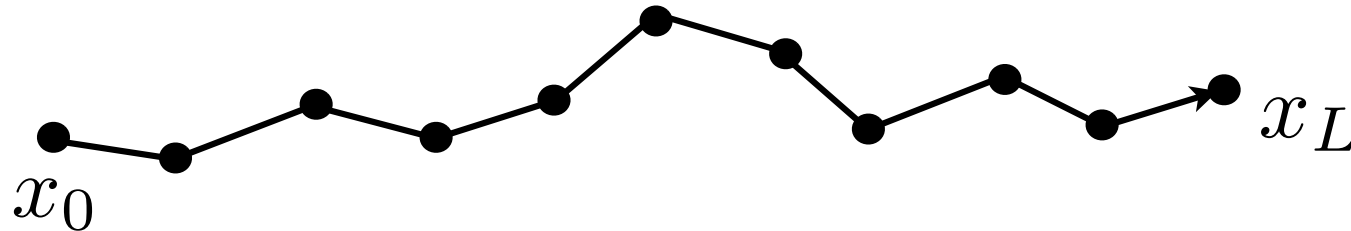
Result: ensemble of transition paths

Algorithm



Transition Path

A path (MD trajectory) is
a sequence of discrete
time snapshots



Path probability:

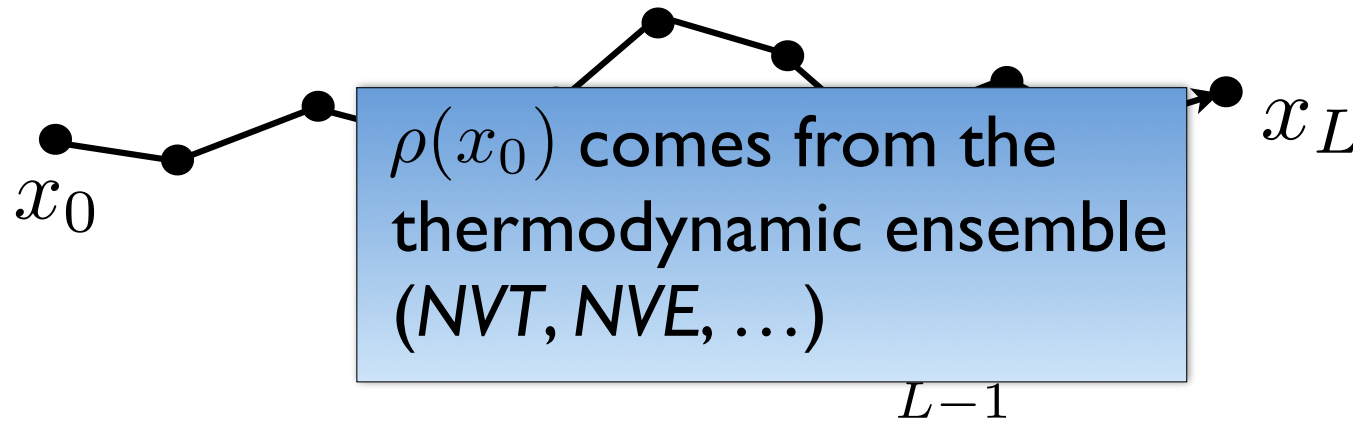
$$\mathcal{P}[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} P(x_i \rightarrow x_{i+1}) / Z(L)$$

normalization constant
(partition function)

$$Z(L) = \int \mathcal{D}\mathbf{x} \mathcal{P}[\mathbf{x}(L)] Z(L)$$

Transition Path

A path (MD trajectory) is a sequence of discrete time snapshots



Path probability:

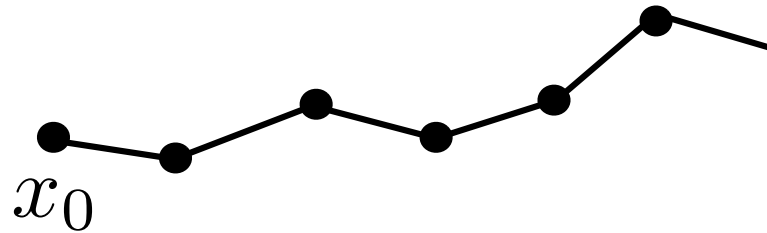
$$\mathcal{P}[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} P(x_i \rightarrow x_{i+1}) / Z(L)$$

normalization constant
(partition function)

$$Z(L) = \int \mathcal{D}\mathbf{x} \mathcal{P}[\mathbf{x}(L)] Z(L)$$

Transition Path

A path (MD trajectory) is a sequence of discrete time snapshots



$P(x_i \rightarrow x_{i+1})$ comes from the molecular dynamics algorithm (integrator)

Path probability:

$$\mathcal{P}[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} P(x_i \rightarrow x_{i+1}) / Z(L)$$

normalization constant
(partition function)

$$Z(L) = \int \mathcal{D}\mathbf{x} \mathcal{P}[\mathbf{x}(L)] Z(L)$$

Detailed Balance (Paths)

$$\mathcal{P}[\mathbf{x}^{(o)}] \pi \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right) = \mathcal{P}[\mathbf{x}^{(n)}] \pi \left(\mathbf{x}^{(n)} \rightarrow \mathbf{x}^{(o)} \right)$$

prob. of
old state

prob of moving
old $\rightarrow$ new

prob. of
new state

prob of moving
new $\rightarrow$ old

$$\pi \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right) = P_{\text{gen}} \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right) P_{\text{acc}} \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right)$$

transition probability

generation probability

acceptance probability

$$P_{\text{acc}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}) = \min \left(1, \frac{\mathcal{P}[\mathbf{x}^{(n)}] P_{\text{gen}} \left(\mathbf{x}^{(n)} \rightarrow \mathbf{x}^{(o)} \right)}{\mathcal{P}[\mathbf{x}^{(o)}] P_{\text{gen}} \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right)} \right)$$

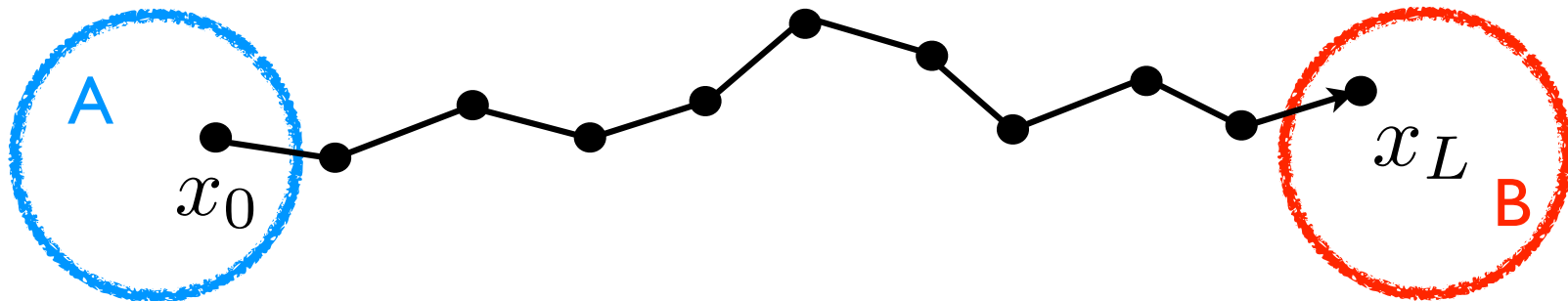
Transition Path

How do we focus on the transition region?

Indicator
Function

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

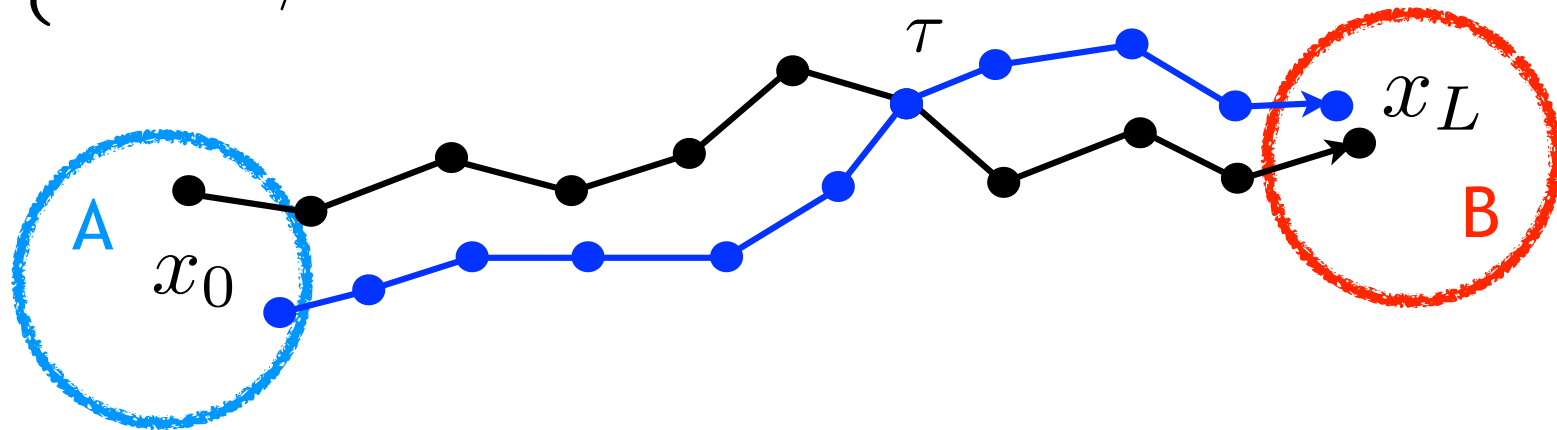
Note: $h_A(x) + h_B(x) \neq 1$!



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

Change the path

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



1. Select a point to “shoot” from (frame τ)
2. Modify the velocities at that point
3. Run forward (from τ to T) and backward (from τ to 0)

$$P_{\text{gen}}^{\text{shoot}} \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right) = P_{\text{sel}}(x_{\tau}^{(o)}; \mathbf{x}^{(o)}) P_{\text{mod}}(x_{\tau}^{(o)} \rightarrow x_{\tau}^{(n)}) P_{\text{fwd}}(\mathbf{x}_{\tau \dots L}^{(n)}) P_{\text{bkwd}}(\mathbf{x}_{0 \dots \tau}^{(n)})$$

Shooting Algorithm: Detailed Balance

$$P_{\text{gen}}^{\text{shoot}} \left(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)} \right) = P_{\text{sel}}(x_{\tau}^{(o)}; \mathbf{x}^{(o)}) P_{\text{mod}}(x_{\tau}^{(o)} \rightarrow x_{\tau}^{(n)}) P_{\text{fwd}}(\mathbf{x}_{\tau \dots L}^{(n)}) P_{\text{bkwd}}(\mathbf{x}_{0 \dots \tau}^{(n)})$$

$$P_{\text{sel}}(x_{\tau}^{(o)}; \mathbf{x}^{(o)}) = 1/(L^{(o)} - 1) \quad \text{select shooting point}$$

$$P_{\text{mod}}(x_{\tau}^{(o)} \rightarrow x_{\tau}^{(n)}) = P_{\text{mod}}(x_{\tau}^{(n)} \rightarrow x_{\tau}^{(o)}) \quad \text{modify it (symmetric!)}$$

$$P_{\text{fwd}}(\mathbf{x}_{\tau \dots L}^{(n)}) = \prod_{i=\tau}^{L-1} P(x_i \rightarrow x_{i+1}) \quad \text{make forward segment}$$

$$P_{\text{bkwd}}(\mathbf{x}_{0 \dots \tau}^{(n)}) = \prod_{t=1} \bar{P}(x_t \rightarrow x_{t-1}) \quad \text{make backward segment}$$

Shooting Move Acceptance

$$P_{\text{acc}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}) = \min \left(1, \frac{\mathcal{P}_{AB}(\mathbf{x}^{(n)}) P_{\text{gen}}^{\text{shoot}}(\mathbf{x}^{(n)} \rightarrow \mathbf{x}^{(o)})}{\mathcal{P}_{AB}(\mathbf{x}^{(o)}) P_{\text{gen}}^{\text{shoot}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)})} \right)$$

$$\frac{P_{\text{gen}}^{\text{shoot}}(\mathbf{x}^{(n)} \rightarrow \mathbf{x}^{(o)})}{P_{\text{gen}}^{\text{shoot}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)})} = \frac{P_{\text{sel}}(x_{\tau}^{(n)}; \mathbf{x}^{(n)}) P_{\text{mod}}(x_{\tau}^{(n)} \rightarrow x_{\tau}^{(o)}) P_{\text{fwd}}(\mathbf{x}_{\tau \dots L}^{(o)}) P_{\text{bkwd}}(\mathbf{x}_{0 \dots \tau}^{(o)})}{P_{\text{sel}}(x_{\tau}^{(o)}; \mathbf{x}^{(o)}) P_{\text{mod}}(x_{\tau}^{(o)} \rightarrow x_{\tau}^{(n)}) P_{\text{fwd}}(\mathbf{x}_{\tau \dots L}^{(n)}) P_{\text{bkwd}}(\mathbf{x}_{0 \dots \tau}^{(n)})}$$

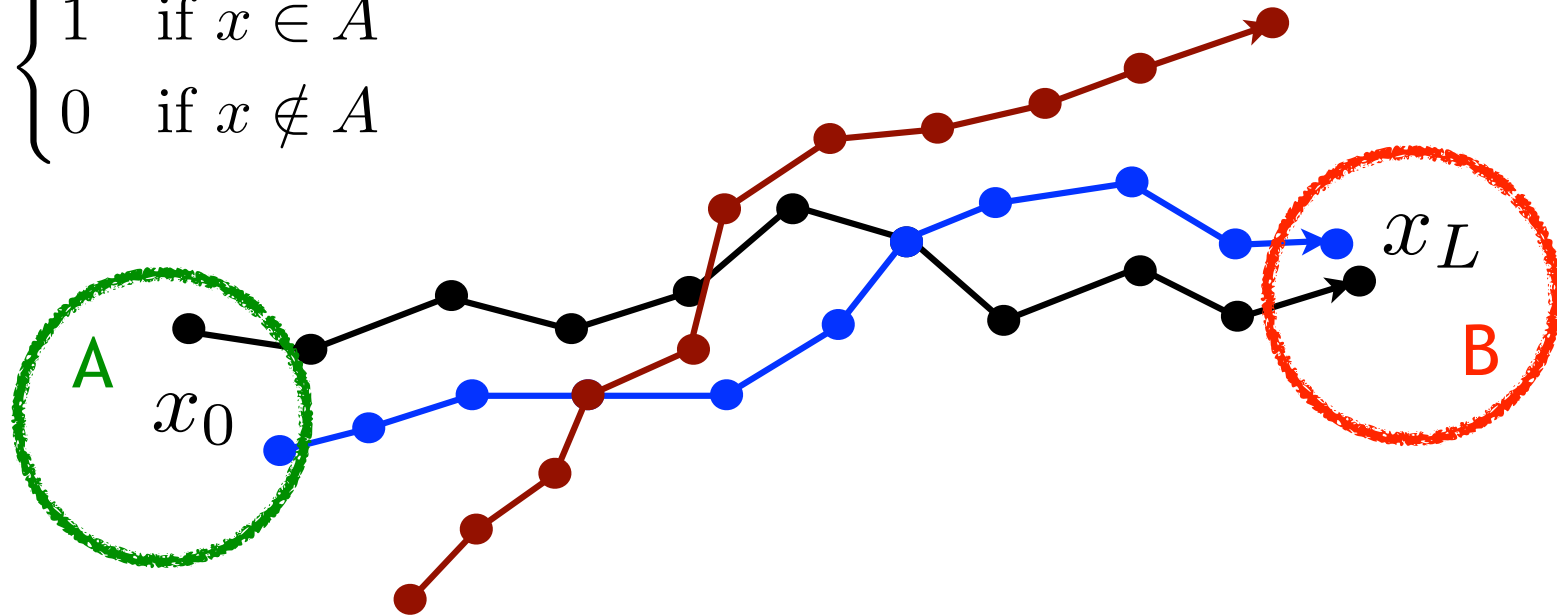
Assumptions:

- symmetric, distribution-preserving modification of the shooting point
- reversible (and distribution-preserving) dynamics
- fixed number of frames in each trajectory, equal probability of selecting any as a shooting point

$$P_{\text{acc}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}) = h_A \left(x_0^{(n)} \right) h_B \left(x_L^{(n)} \right)$$

Transition Path Sampling

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



1. Select a point to “shoot” from (frame τ)
2. Modify the velocities at that point
3. Run forward (from τ to T) and backward (from τ to 0)
4. Accept/reject

$$P_{\text{acc}}(\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}) = h_A \left(x_0^{(n)} \right) h_B \left(x_L^{(n)} \right)$$

What if the modification didn't matter?

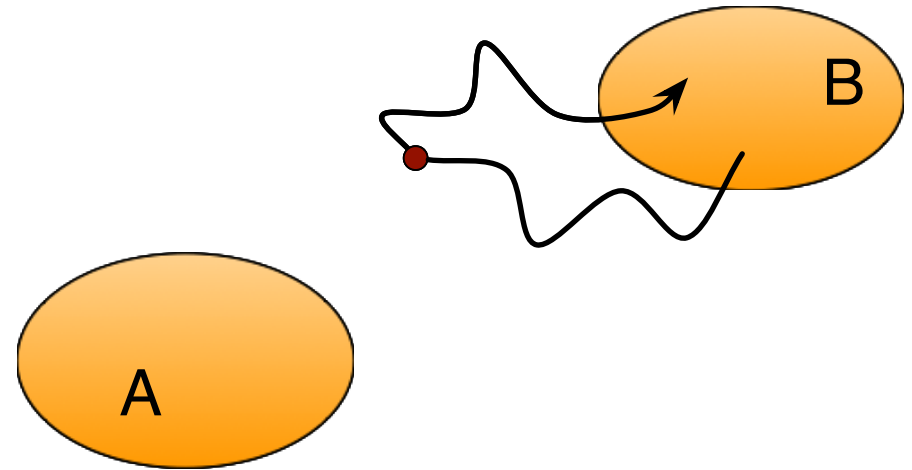
Velocity is randomized after
~1ps in biomolecules.

Start with a shooting point at
the transition state.

Prob. forward path hits B: 50%

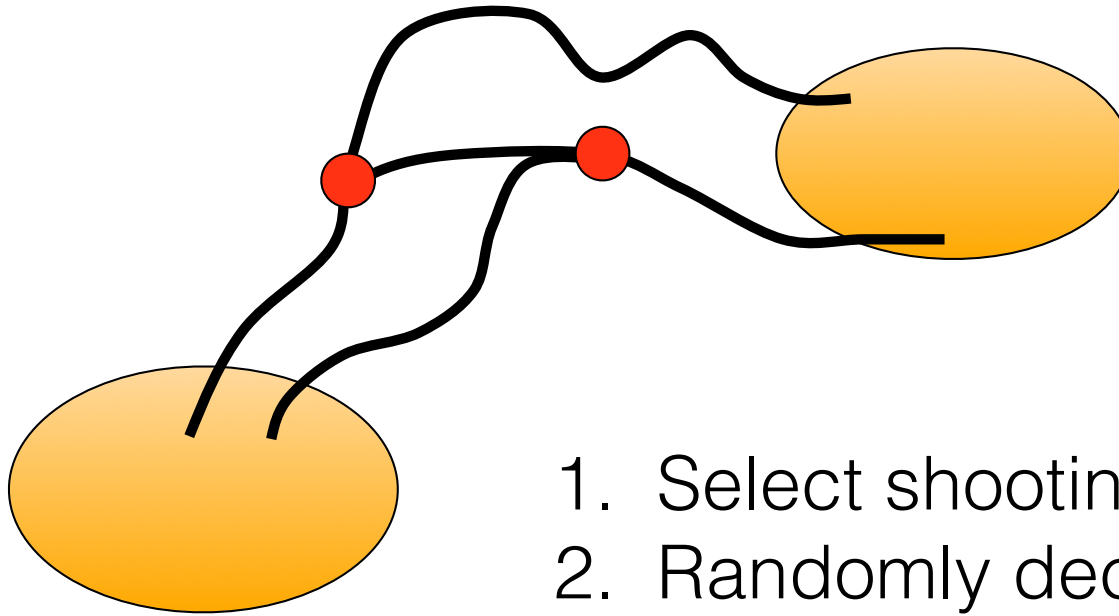
Prob. backward path hits A: 50%

Best acceptance: 25%. Can we do better?



If we only run one direction, we don't have a physical path ...
... unless we use a stochastic integrator!

Stochastic (One-Way) Shooting



1. Select shooting point
2. Randomly decide to shoot either forward or backward
3. The new trajectory has some frames from the old trajectory

No modification of the selected snapshot!

Flexible-Length Shooting

The time spent in the barrier region has a wide variance. The average time can be 1/5 as long as the time of the longest 1% of trajectories. If you always have to go to the longest time, you end up doing 5 times as much work.

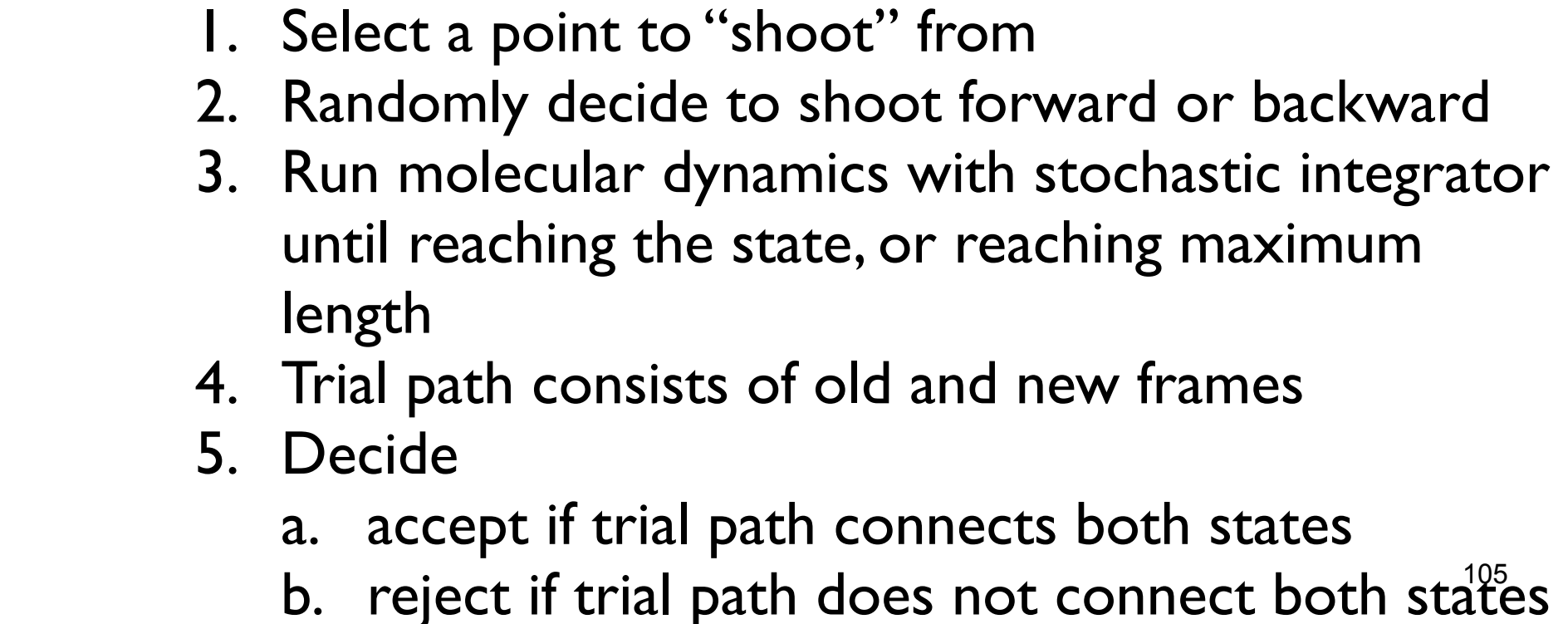
Let's design a shooting move that stops as soon as it enters the state. Accepted paths will have one frame in each state (first frame and last frame), and no other frames will be in any state. The path length will be allowed to vary.

$$P_{\text{sel}}(x_\tau; \mathbf{x}^{(o)}(L^{(o)})) = 1/(L^{(o)} - 1)$$

$$\frac{P_{\text{sel}}(x_\tau; \mathbf{x}^{(n)}(L^{(n)}))}{P_{\text{sel}}(x_\tau; \mathbf{x}^{(o)}(L^{(o)}))} = \frac{1/(L^{(n)} - 1)}{1/(L^{(o)} - 1)} = \frac{L^{(o)} - 1}{L^{(n)} - 1}$$

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

- initial trajectory
- accepted path
- rejected path



- 105

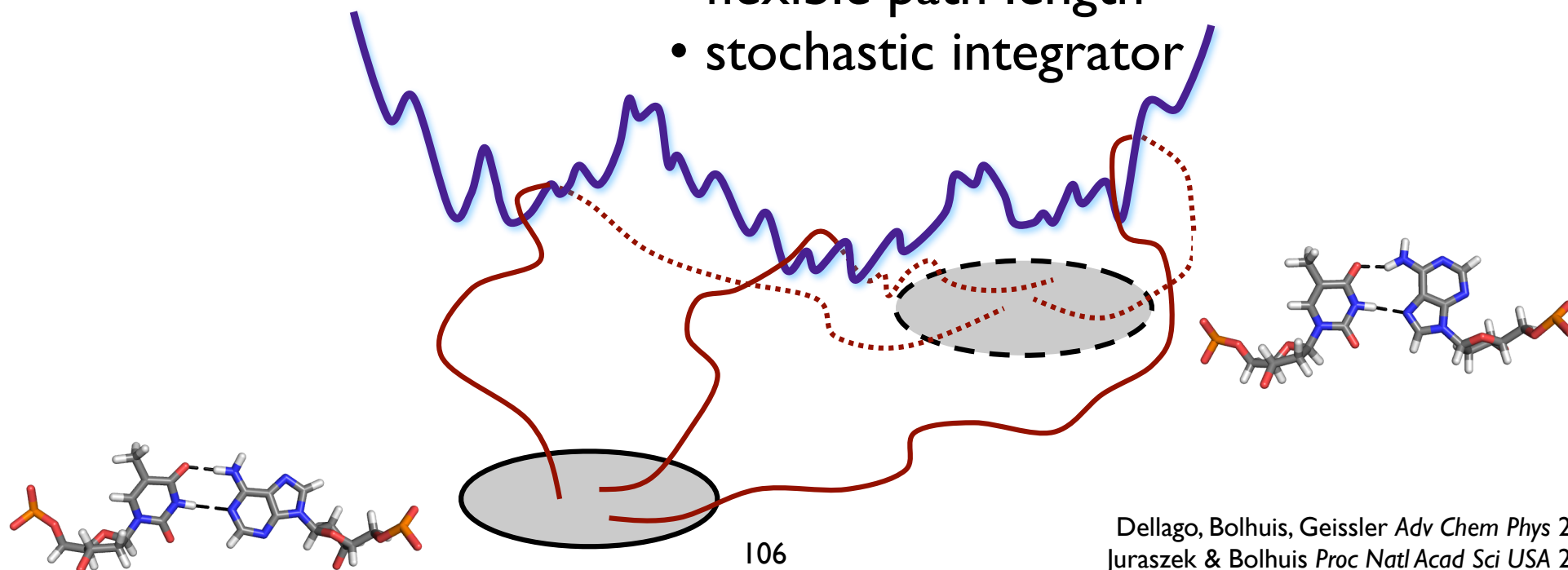
Sampling rare events: transition path sampling

Input

Result: ensemble of transition paths

Algorithm

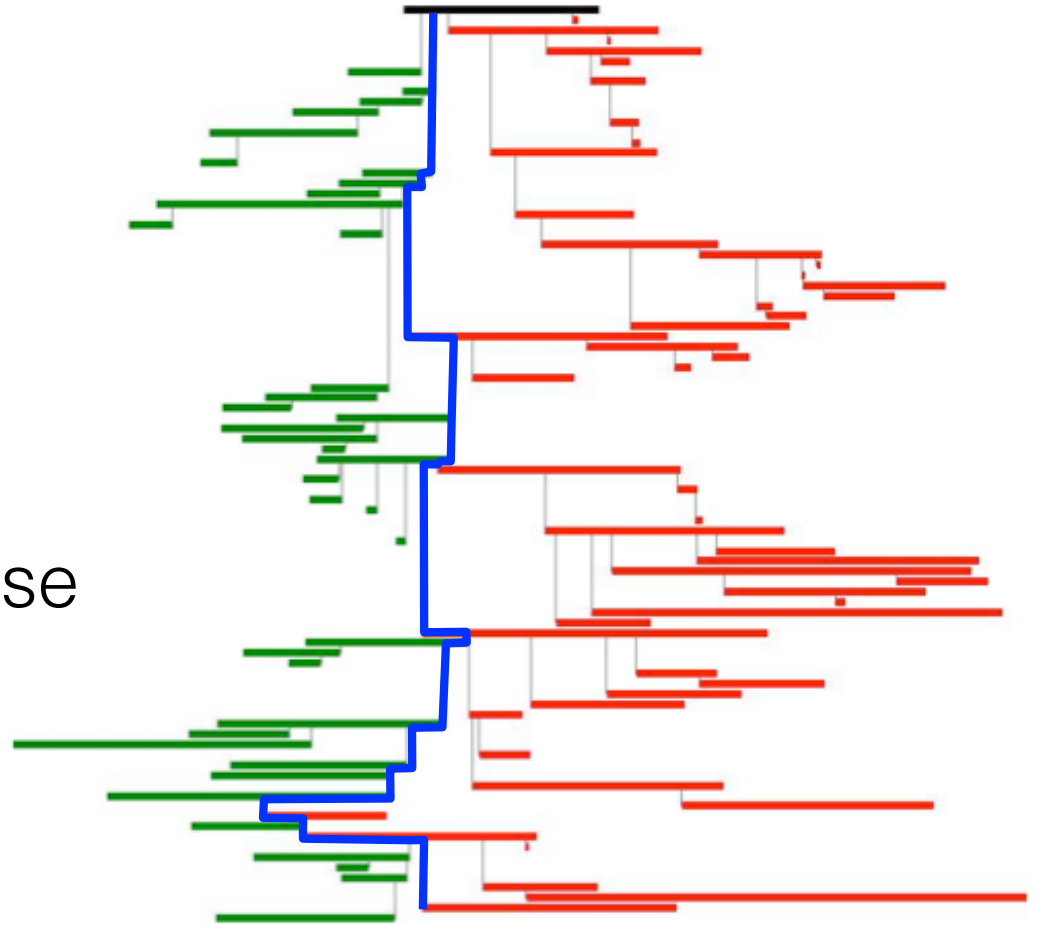
- one way shooting
- flexible path length
- stochastic integrator



Path Trees and Transition Region

Path tree is a check of good behavior in simulation:

- Both forward and backward shots accepted with similar frequency
- See decorrelated paths: those with no frames in common



Least-changed path: Stays near top of barrier;
Approximation of transition state ensemble

Sampling rare events: transition path sampling

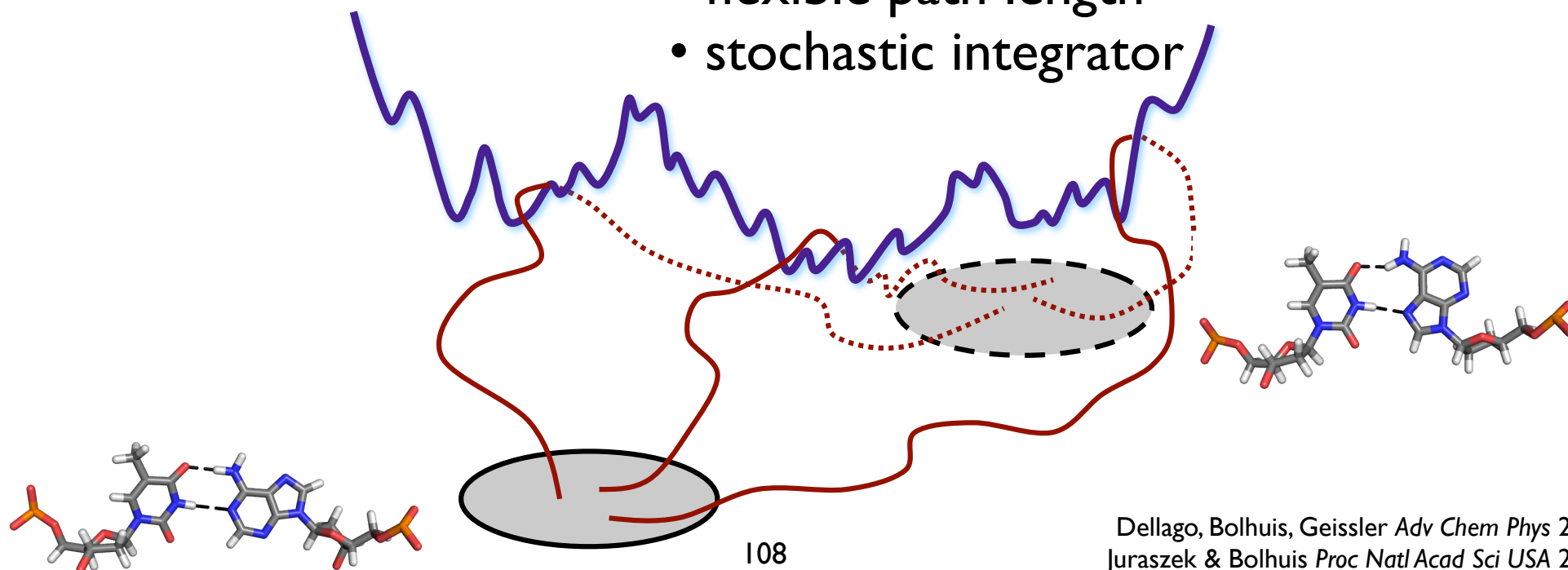
Input

- initial reactive path
- definitions of stable states

Result: ensemble of transition paths

Algorithm

- one way shooting
- flexible path length
- stochastic integrator



How do we define states?

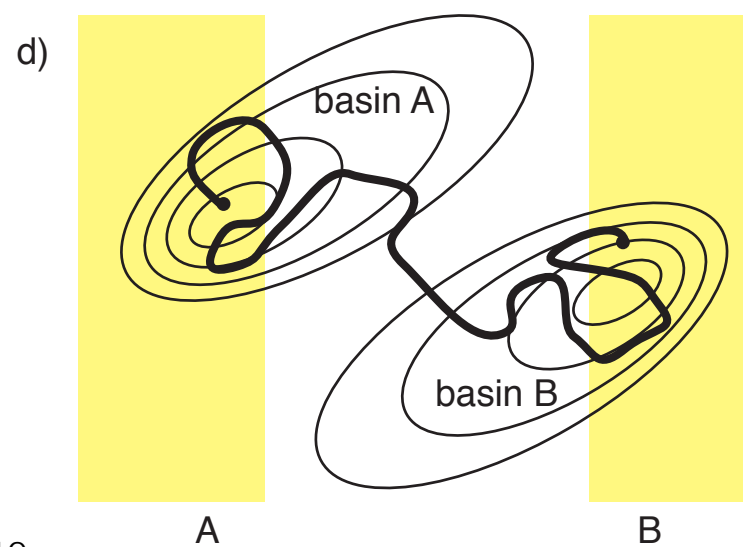
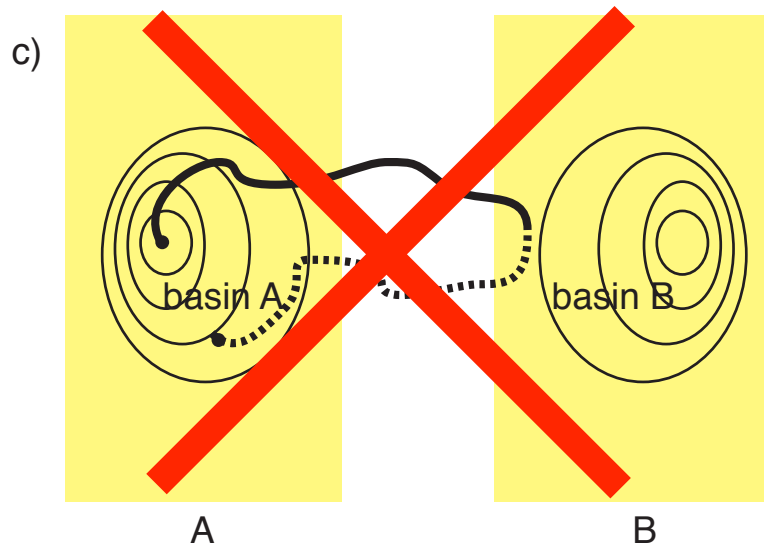
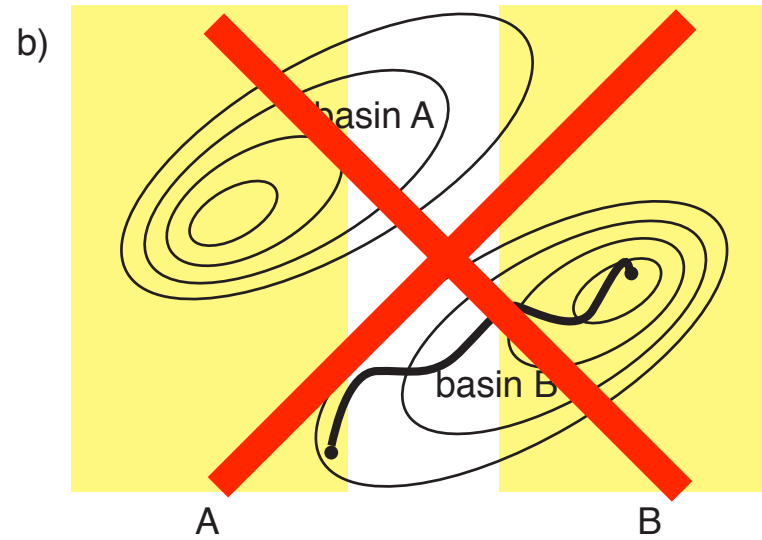
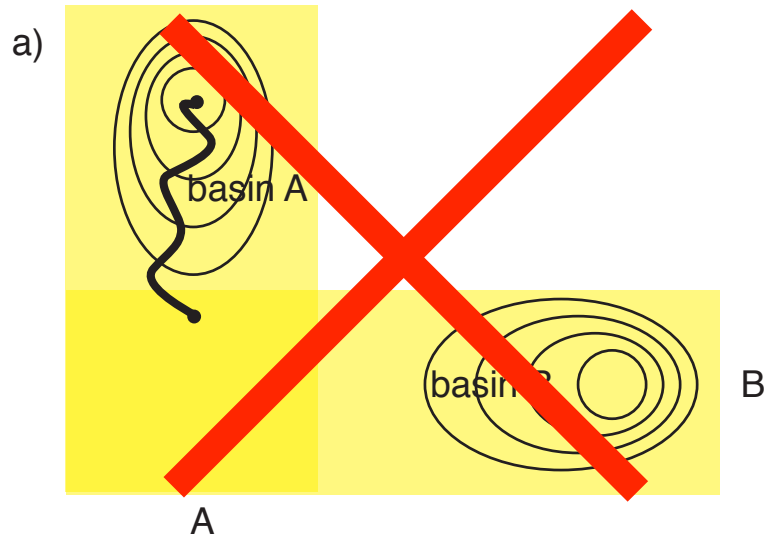
Take something that you can measure (distance, dihedral angle) and say it has to be within some range of values.

Examples:

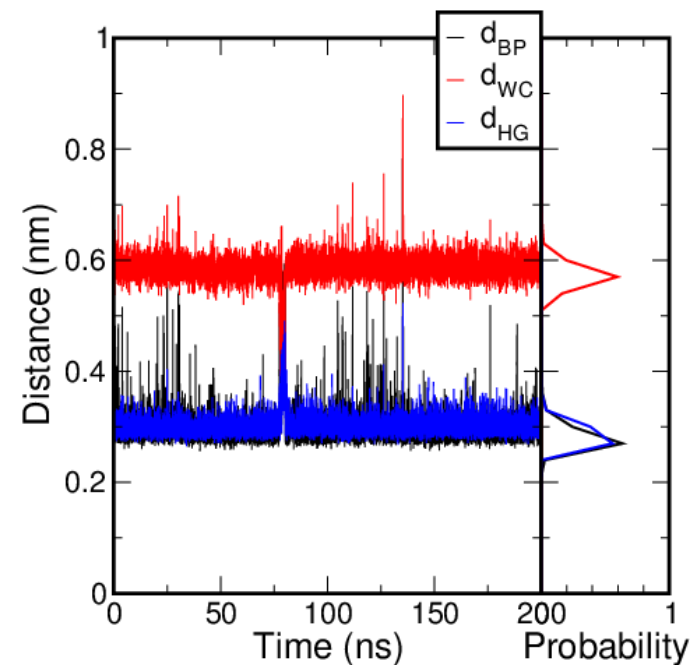
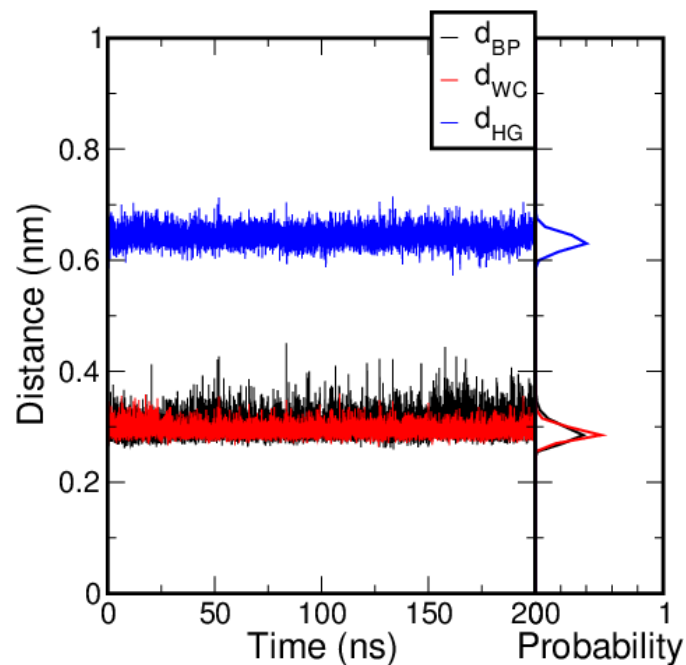
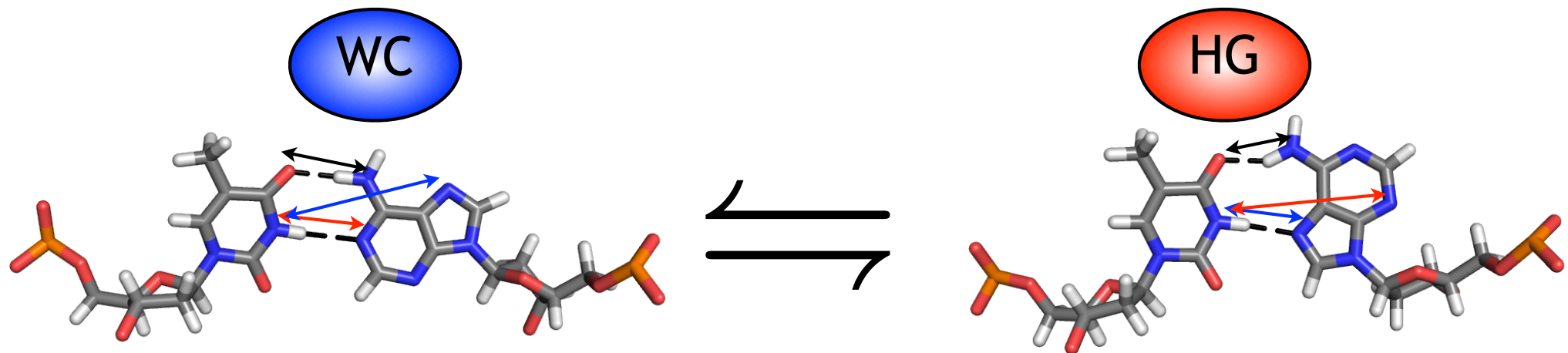
- Hydrogen bond formed ($d_{\text{donor-acceptor}} < 3.5 \text{ \AA}$)
- Specific value of (several) dihedral angles

What makes a good definition?

How do we define states?



The transition between WC and HG does not occur within 200 ns.



Sampling rare events: transition path sampling

Input

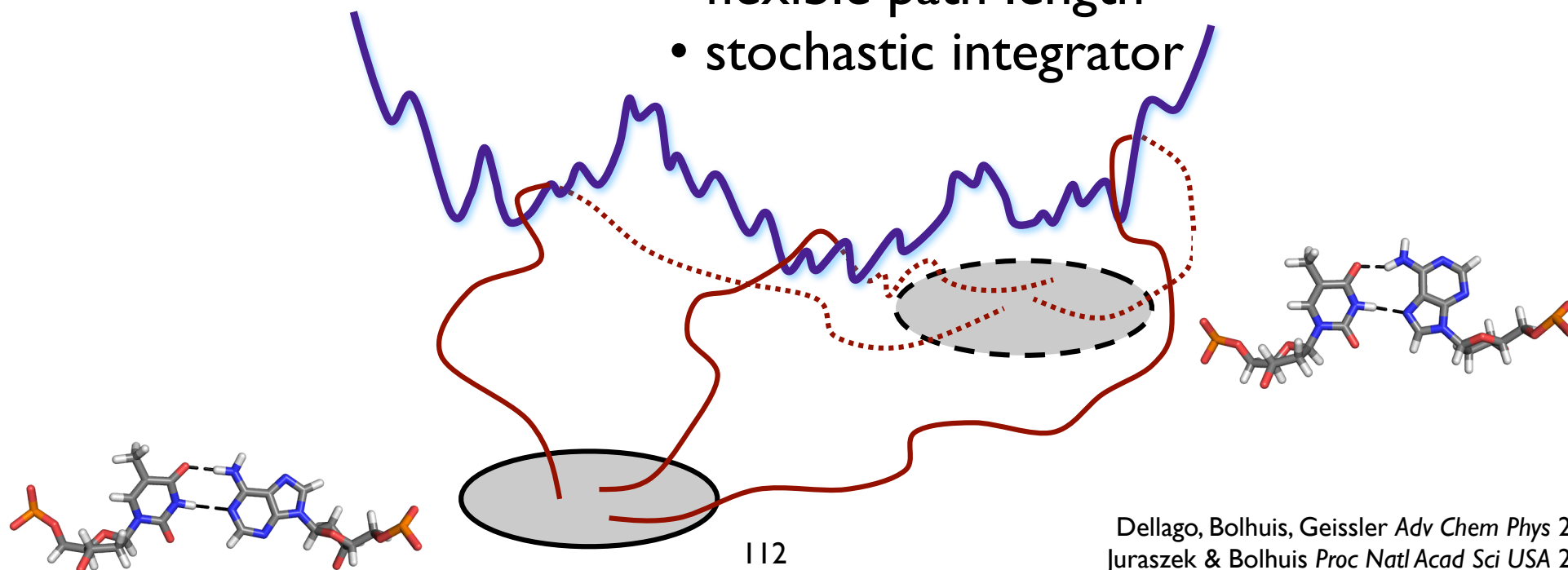
- initial reactive path
- definitions of stable states

Result: ensemble of transition paths

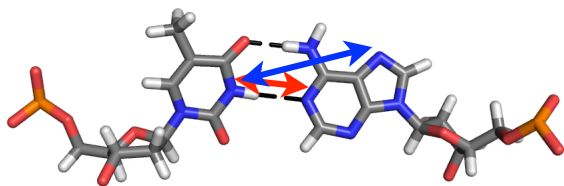
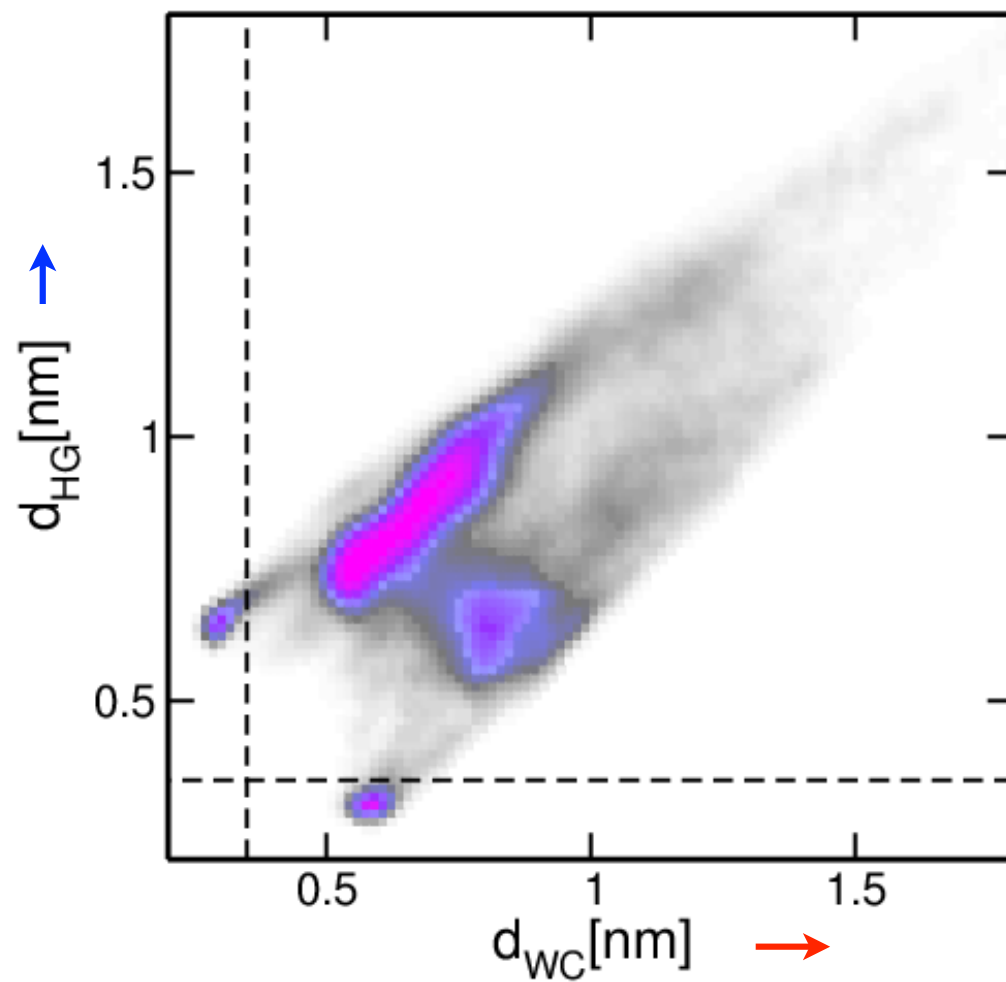
- mechanism
- relevant reaction coordinates

Algorithm

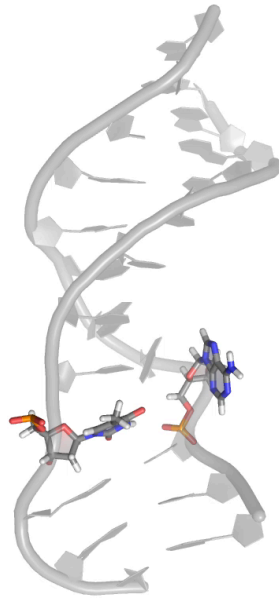
- one way shooting
- flexible path length
- stochastic integrator



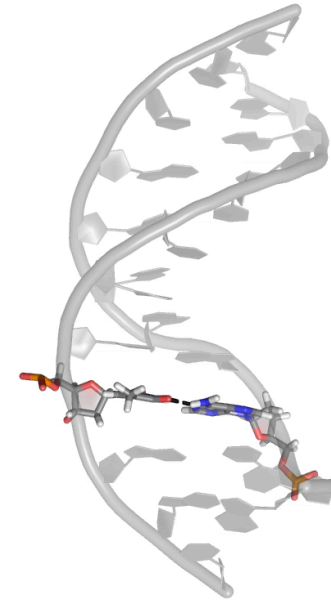
Generate reactive trajectories using transition path sampling



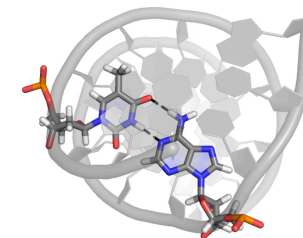
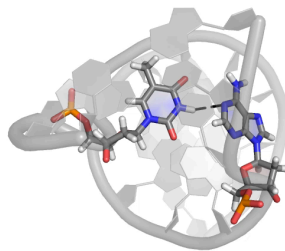
Two types of transitions



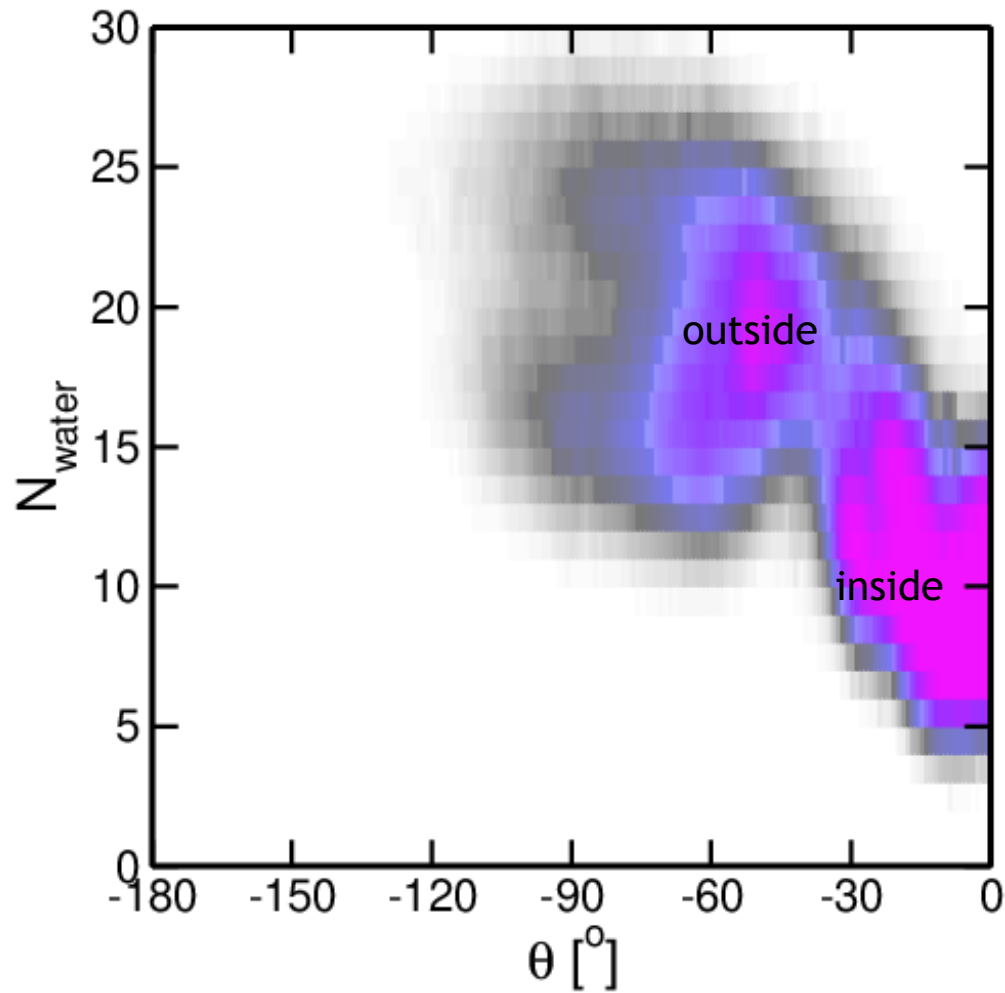
outside



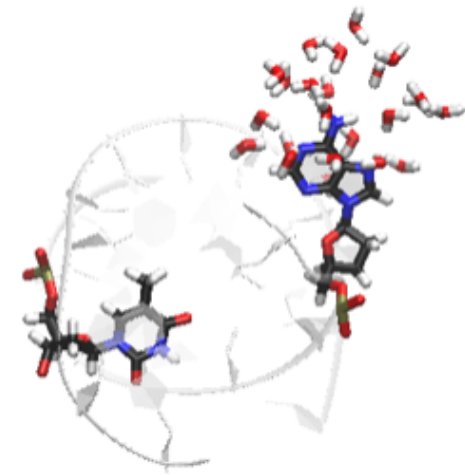
inside



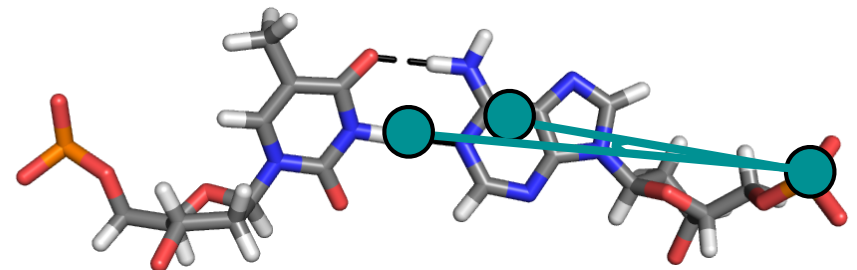
Two reaction channels: inside and outside



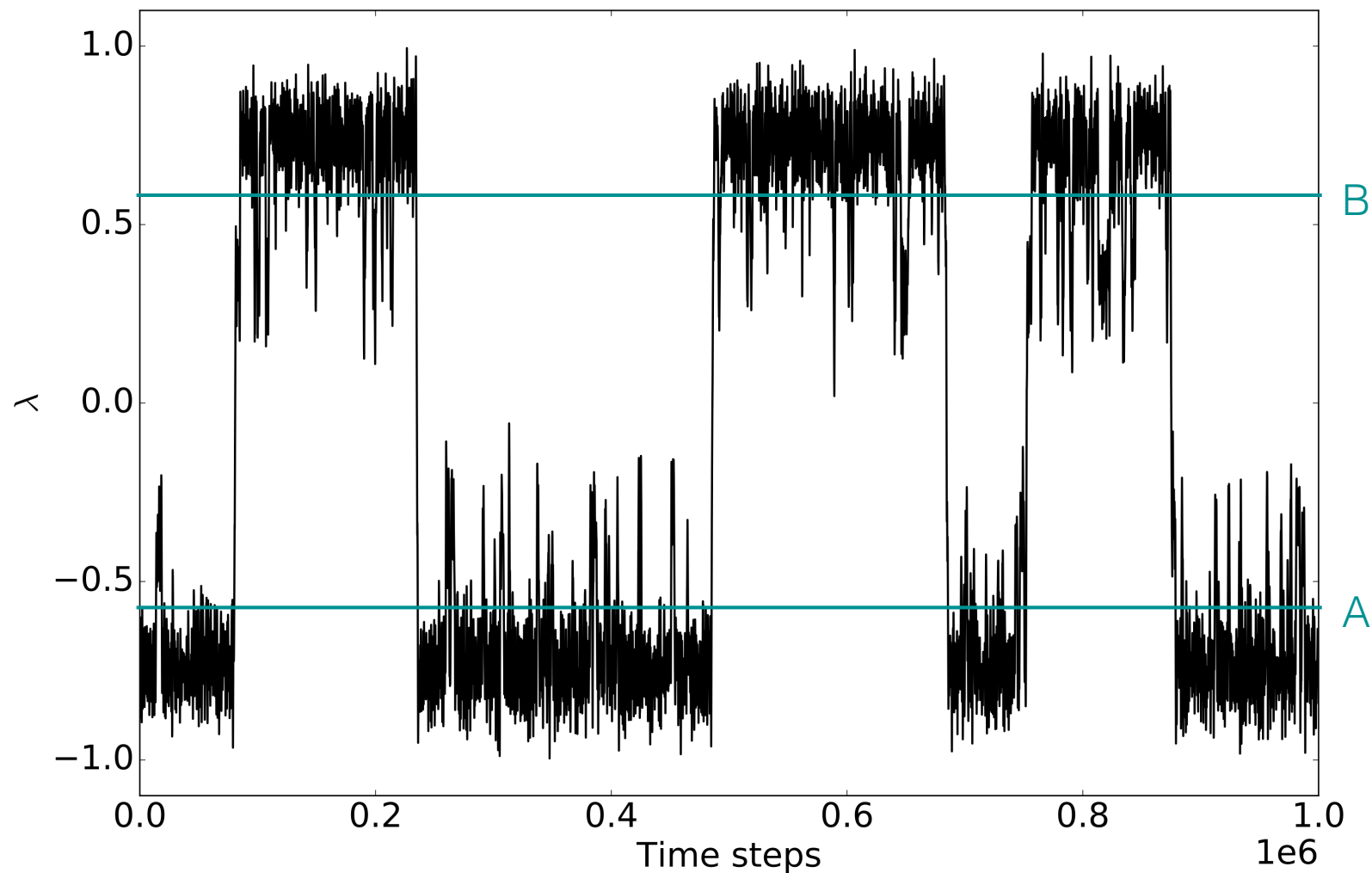
N_{water}



θ base flipping angle



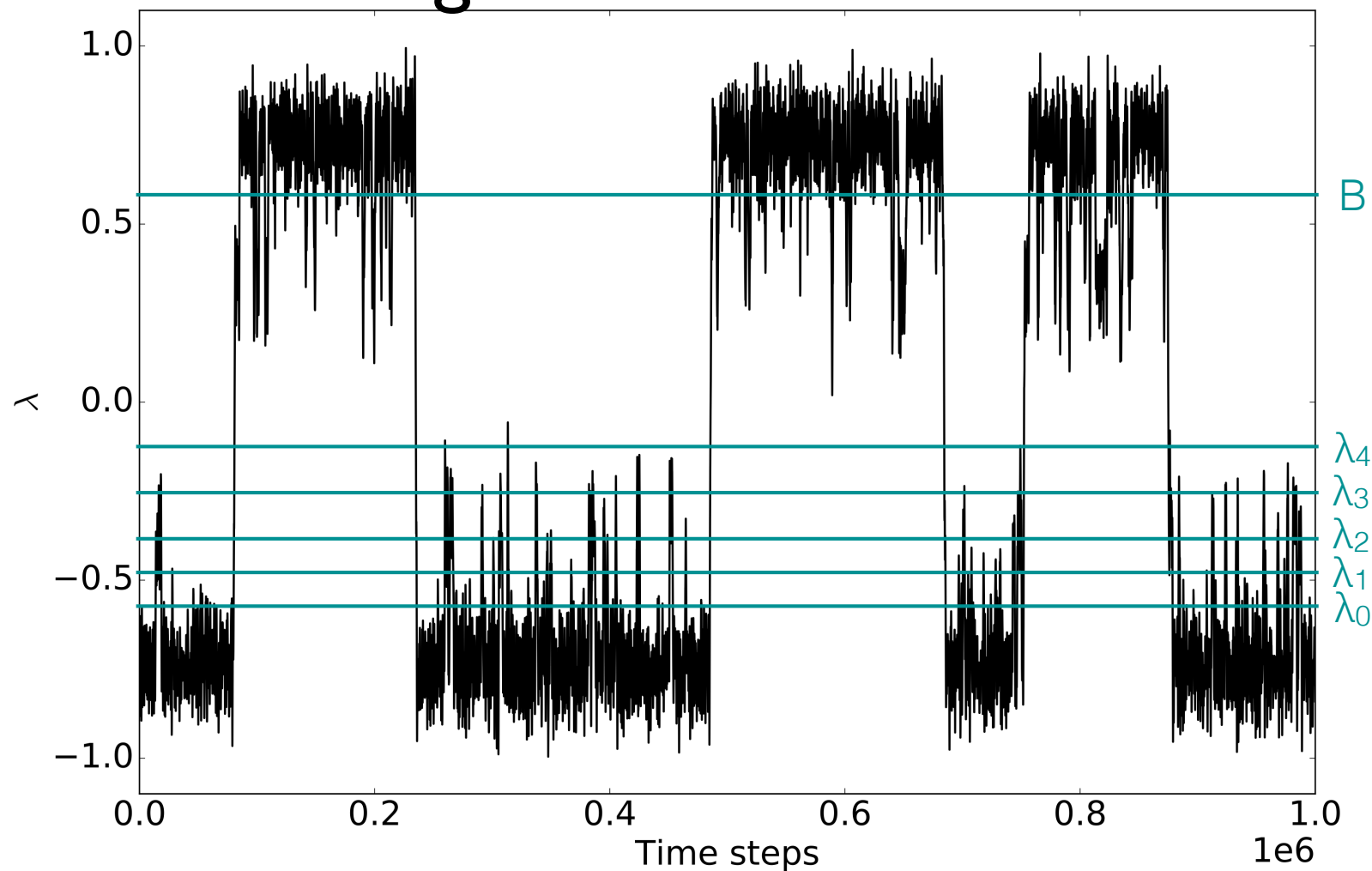
Calculate the rate



The rate from A to B is equal to
the number of transitions per time unit
divided by the time spent in state A

$$k_{AB} = \lim_{\mathcal{T} \rightarrow \infty} \frac{N_{A \rightarrow B}(\mathcal{T})}{t_{tot}^A(\mathcal{T})}$$

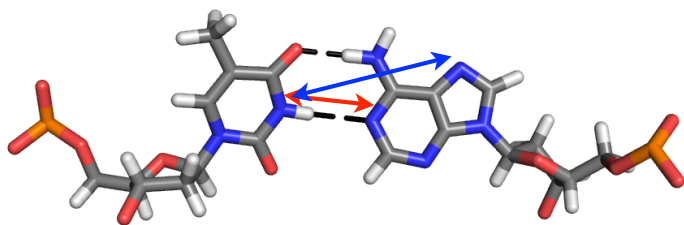
Transition Interface Sampling: calculating the rate with interfaces



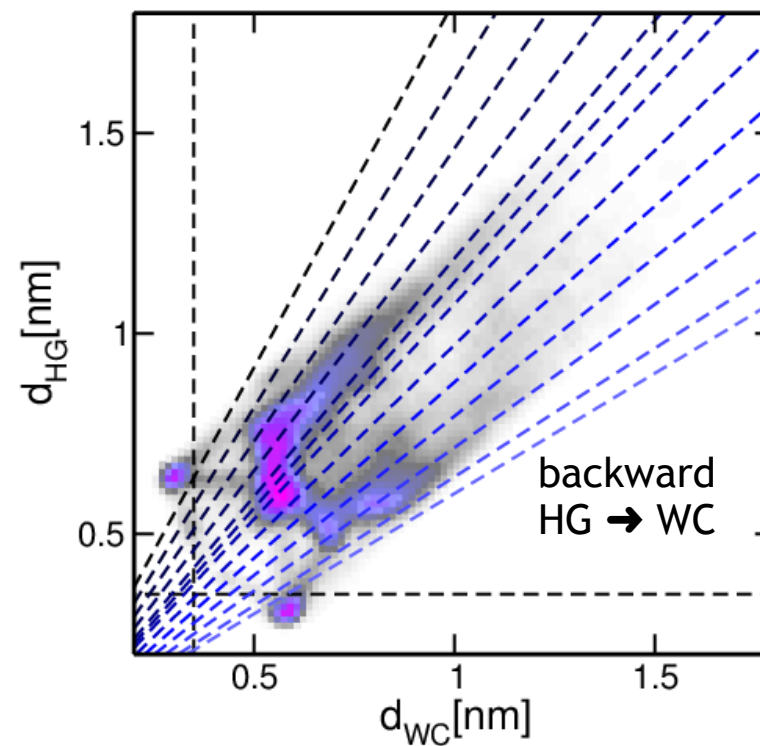
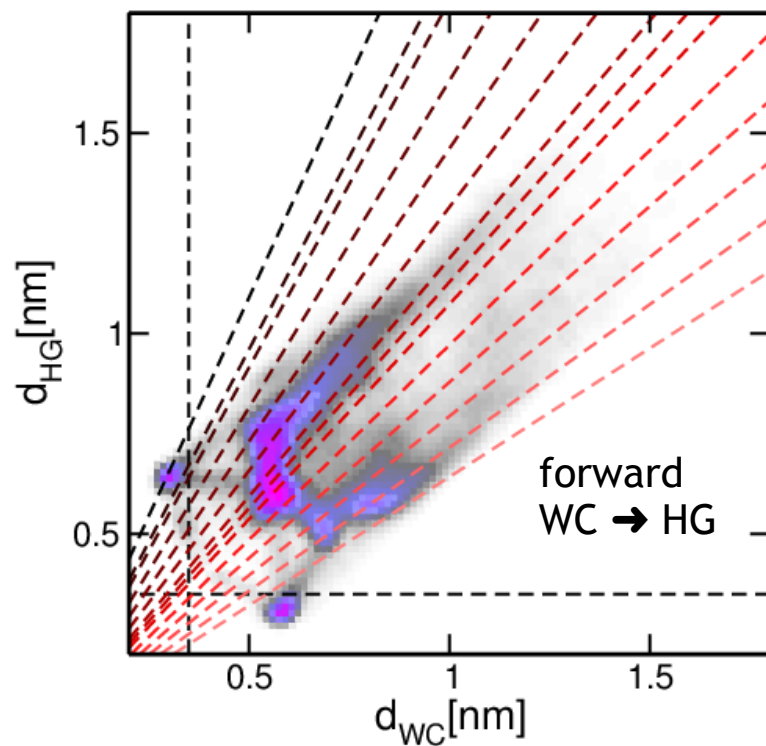
$$k_{AB} = \phi_{A,0} P(B|\lambda_m) \prod_{i=0}^{m-1} P(\lambda_{i+1}|\lambda_i)$$

λ ?

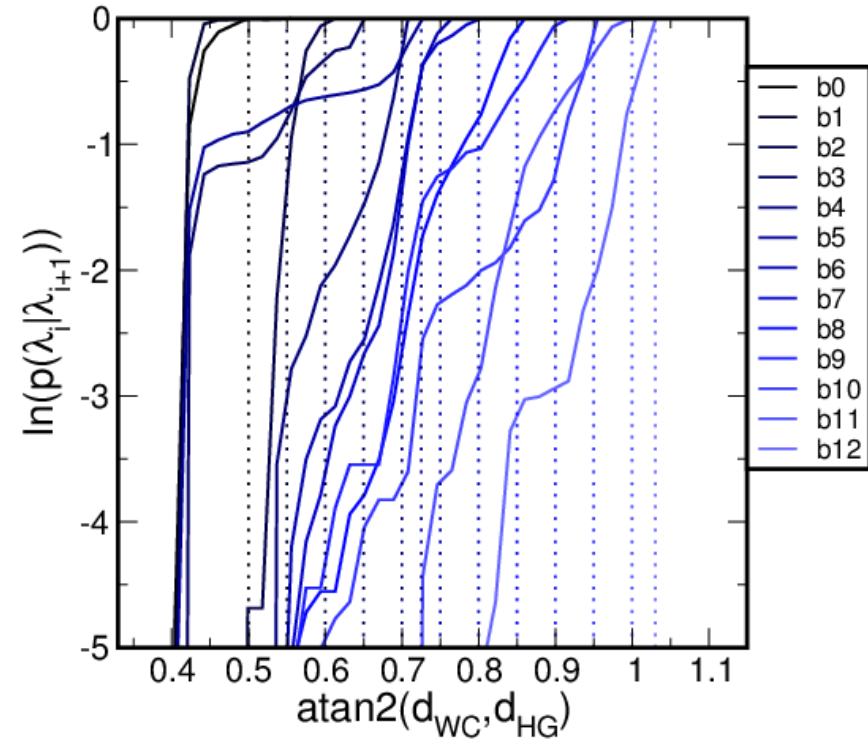
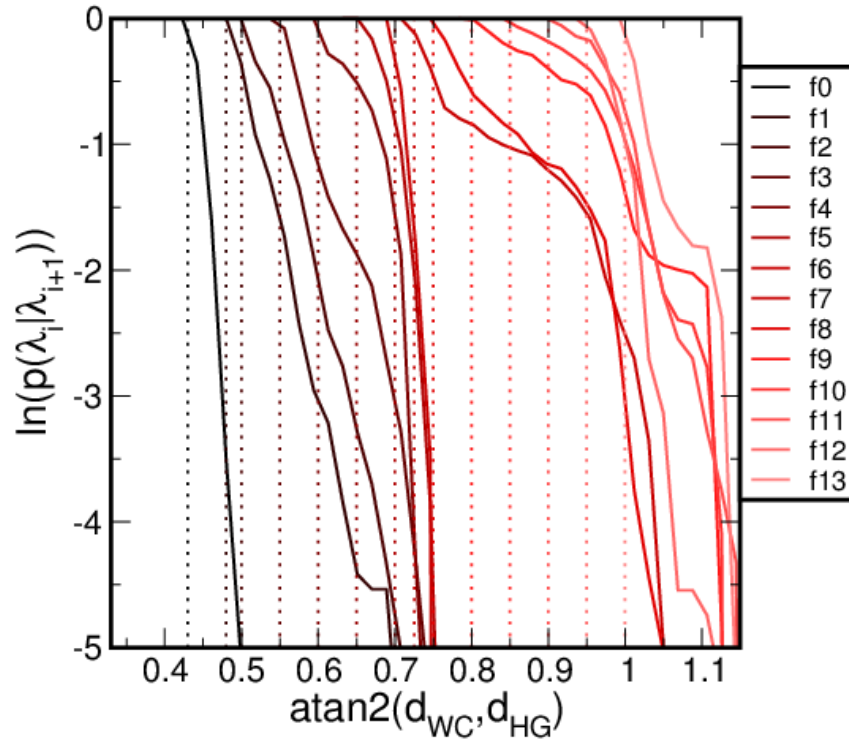
Interfaces



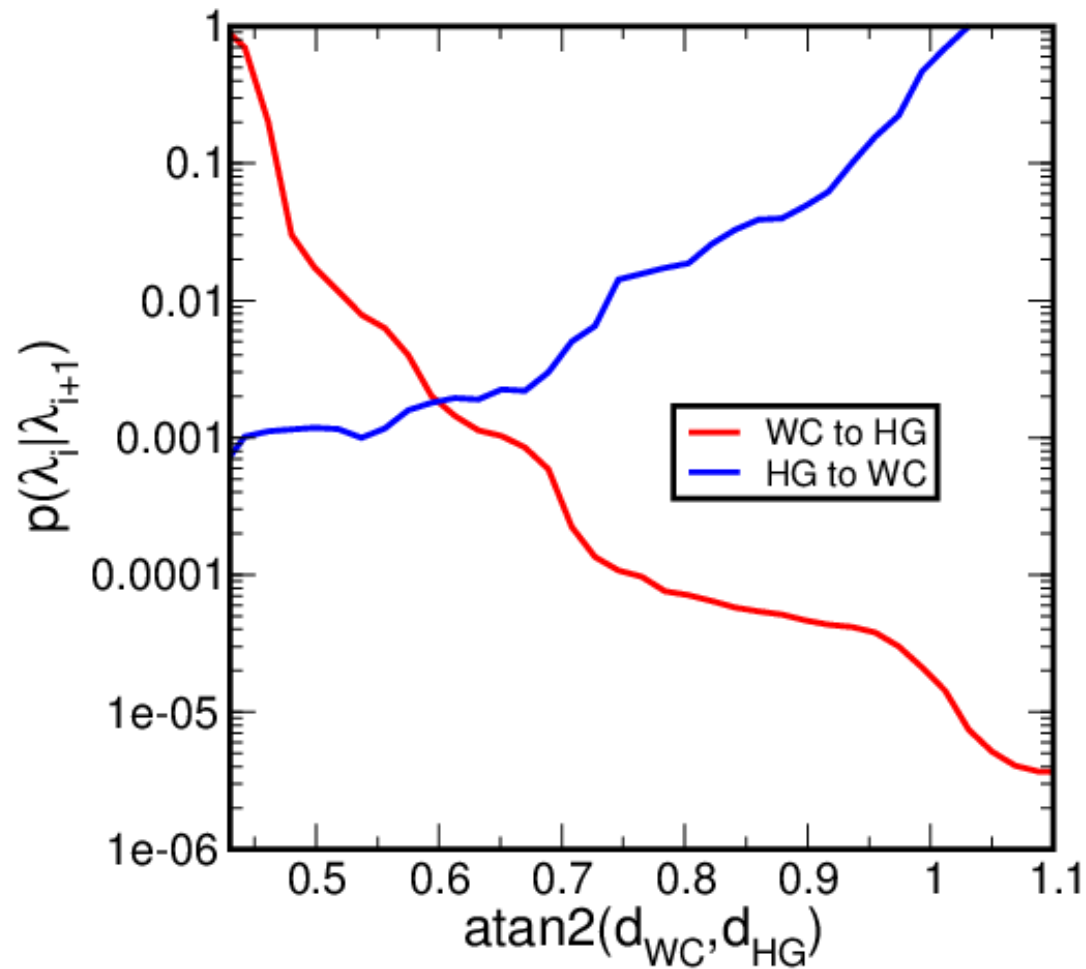
$$\lambda = \arctan(d_{WC}, d_{HG})$$



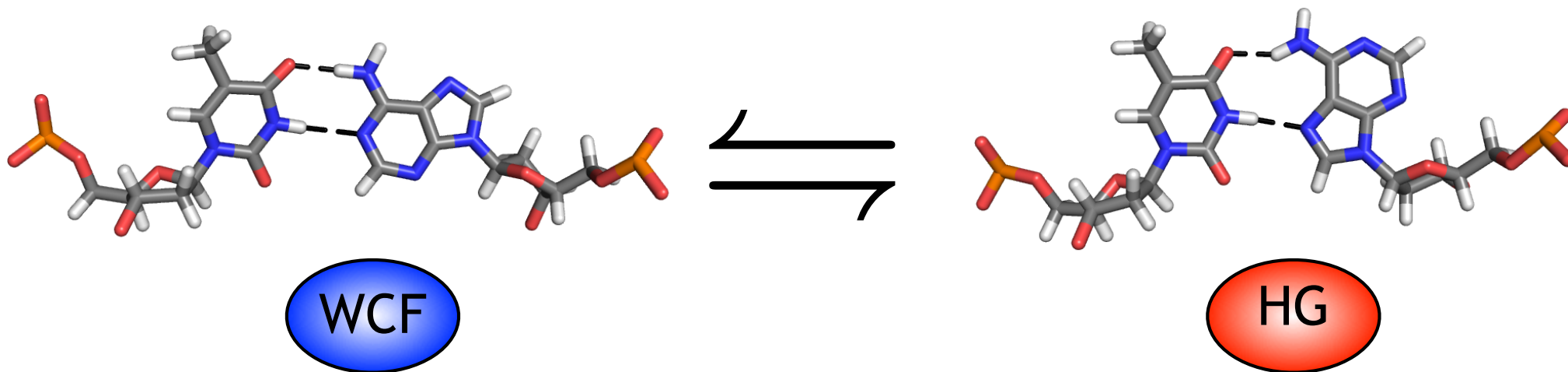
Crossing probabilities



Crossing probabilities



Rates



	expt [*]	TIS ^{**}
$k_{WC \rightarrow HG} \text{ (s}^{-1}\text{)}$	14.2 ± 1.03	742
$k_{HG \rightarrow WC} \text{ (s}^{-1}\text{)}$	3670 ± 200	$1.6 \cdot 10^5$
$\Delta G \text{ (k}_B\text{T)}$	5.5	5.4

at 26.0°C

at 300K

TIS:
 one-way shooting
 flexible path length
 500 accepted paths per interface
 13 interfaces

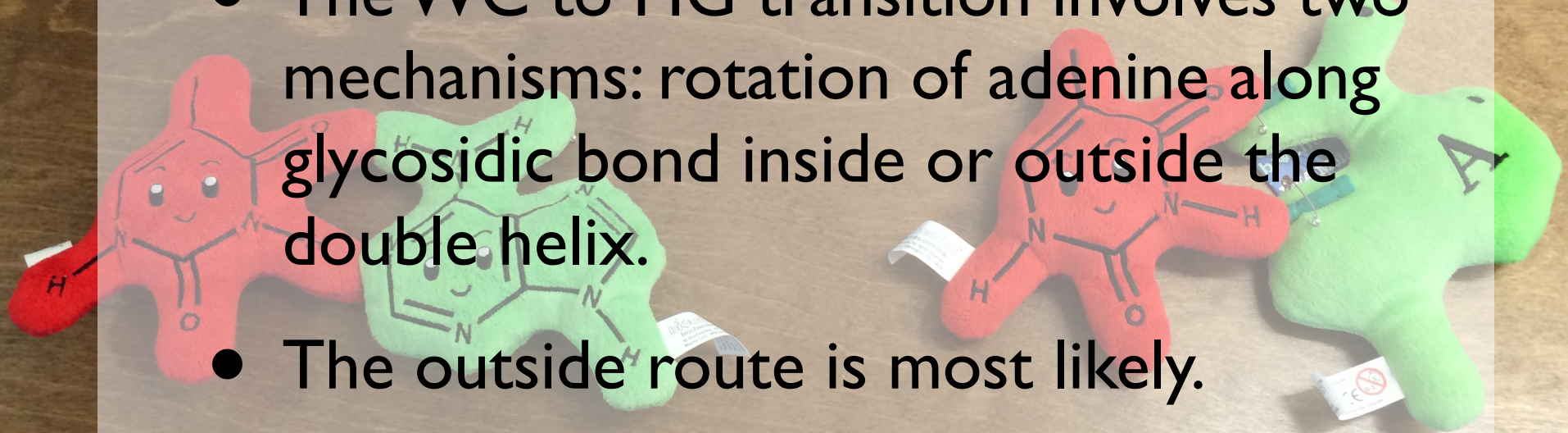
*Nikolova, Kim, Wise, O'Brien, Andricioaei and Al-Hashimi
 Nature 2011 v. 470, p. 498

**Vreede, Ortiz de Alba Perez, Bolhuis & Swenson, NAR 2019

$$\Delta G = -\ln \frac{k_{WC \rightarrow HG}}{k_{HG \rightarrow WC}}$$

In summary

- The WC to HG transition involves two mechanisms: rotation of adenine along glycosidic bond inside or outside the double helix.
- The outside route is most likely.
- TIS results qualitatively agree with experimental free energy difference.



Bridging the time gap

