ACMM

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

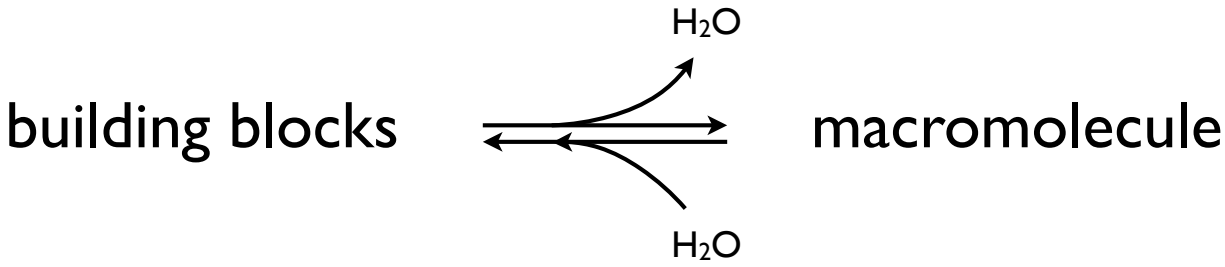
Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

Part 2: Biased sampling

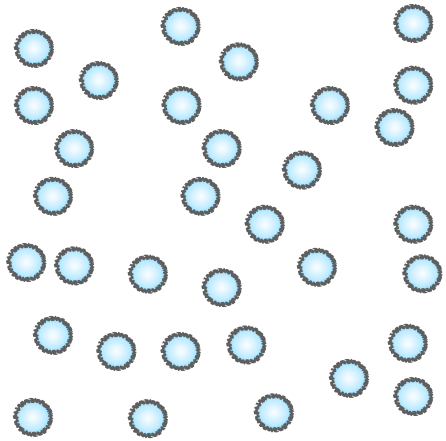
Part 3: Rare events

Biomolecules

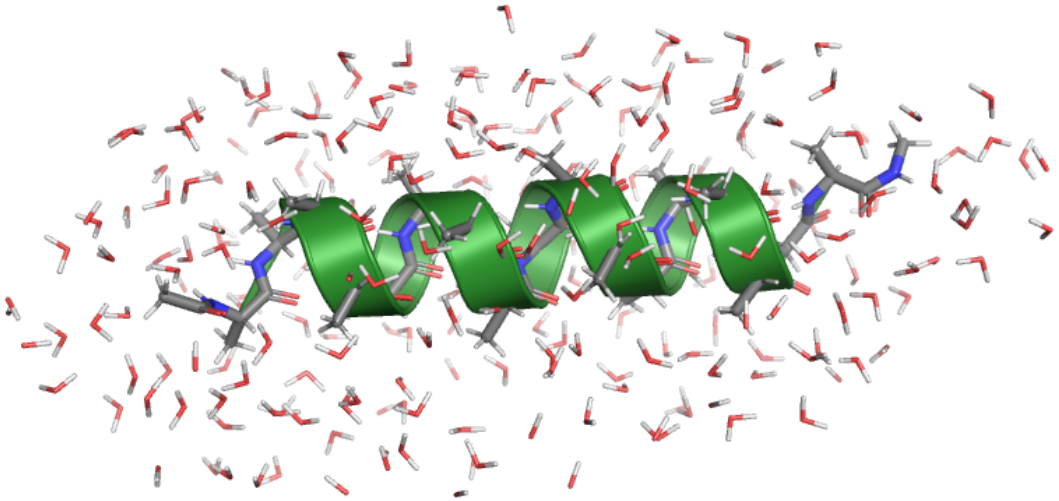


building block	**macromolecule**
amino acid	protein
nucleotide	nucleic acids
monosaccharide	cellulose, starch
lipid	membrane

Interactions in a biomolecular system



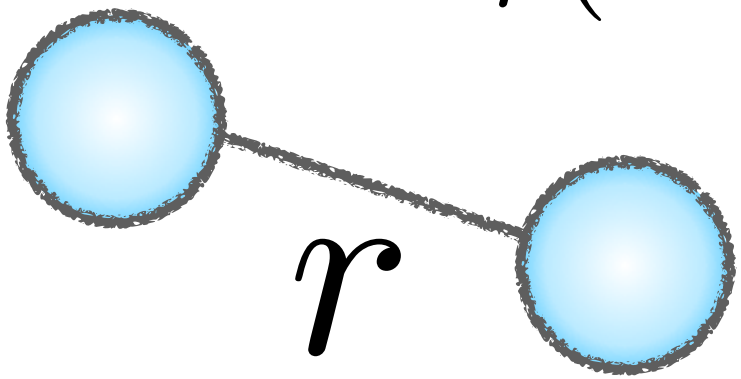
Lennard-Jones particles

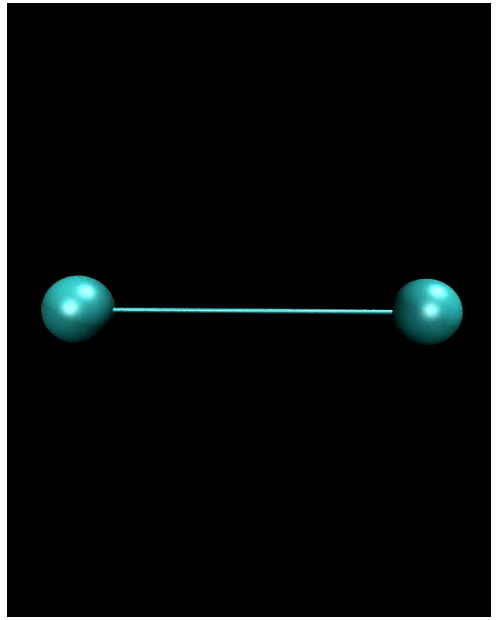


α-helix in water

Bonded interactions

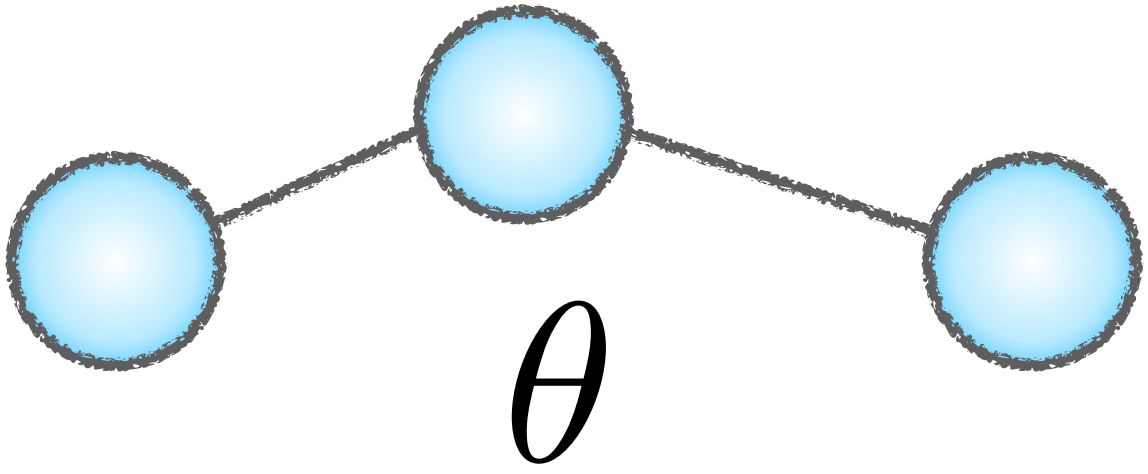
$$k_r(r - r_{eq})^2$$

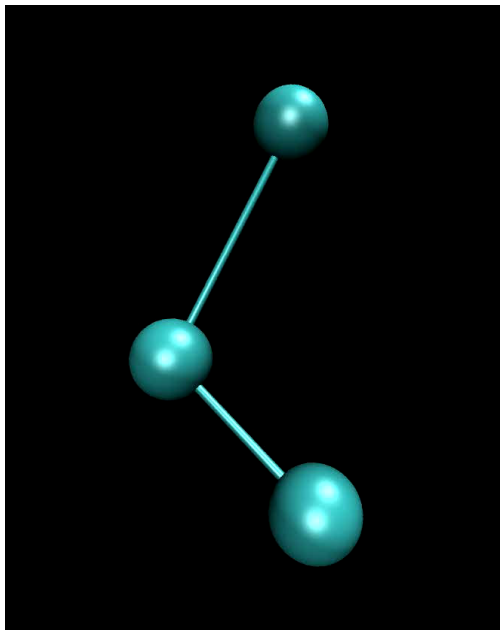




Bonded interactions

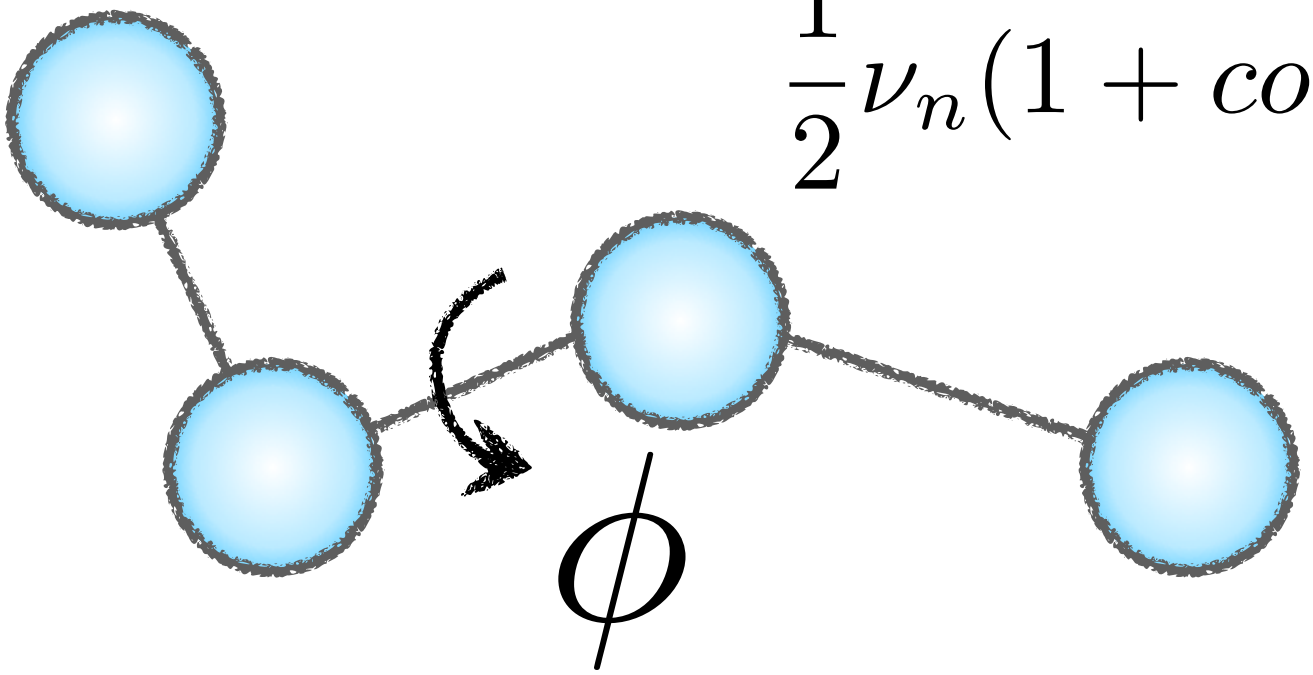
$$k_\theta(\theta - \theta_{eq})^2$$

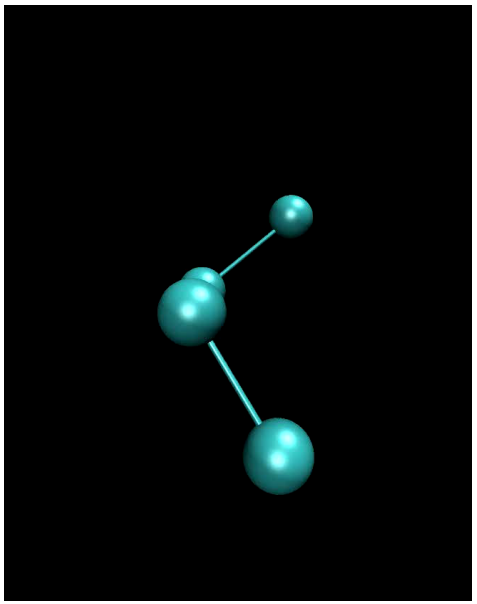




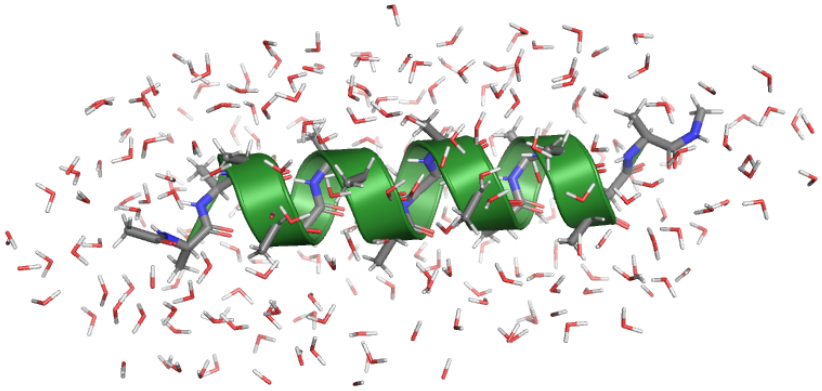
Bonded interactions

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





Bonded interactions



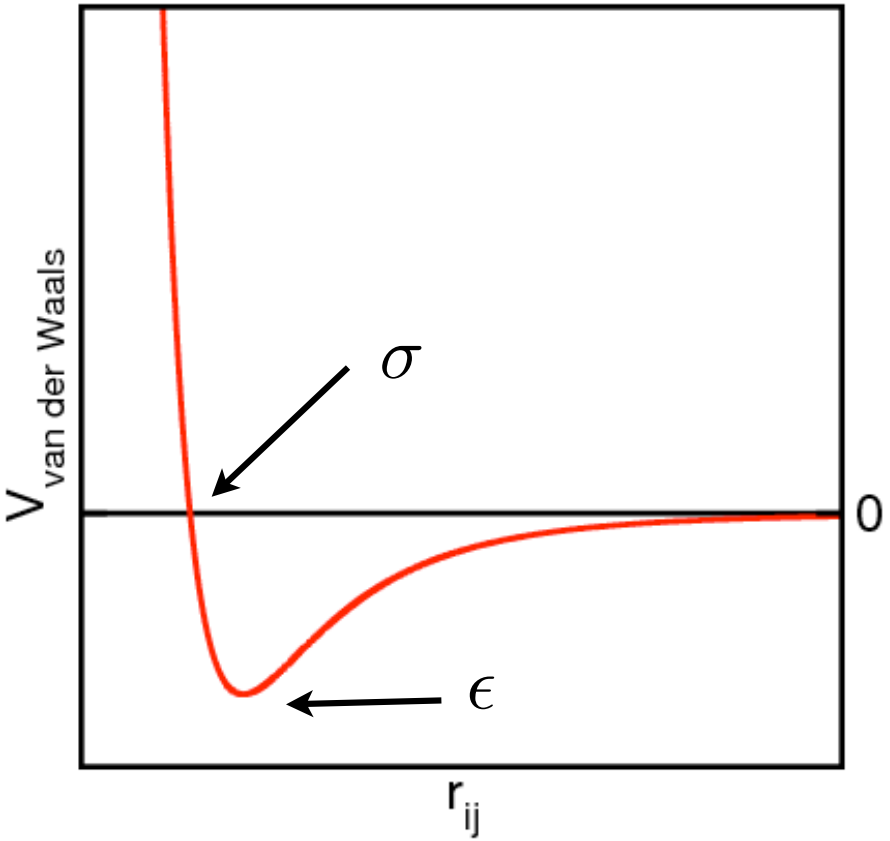
$$V(r) = \sum_{bonds} k_r(r - r_{eq})^2 +$$

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

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

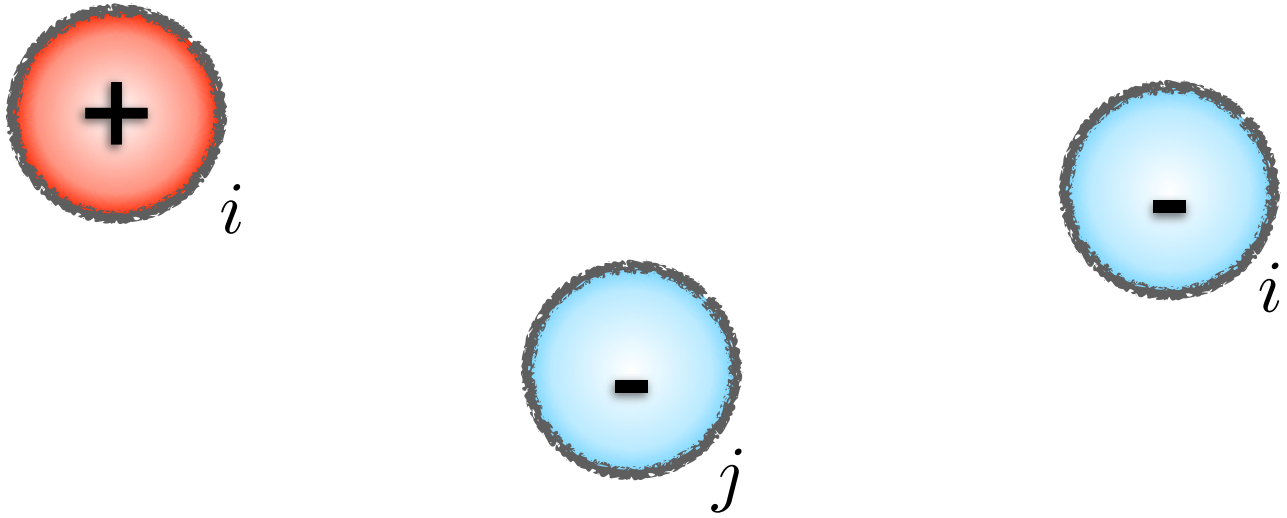
Non-bonded interactions

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



Lennard-Jones potential

Electrostatic interactions



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

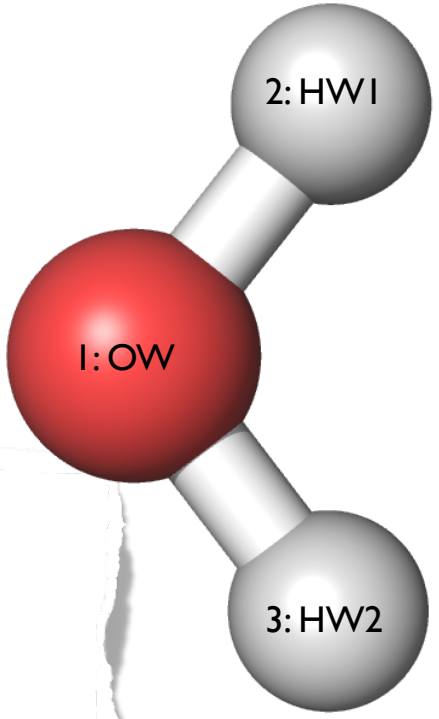
Coulomb's law

Force field

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

Topology

- List of atoms
- Connectivity

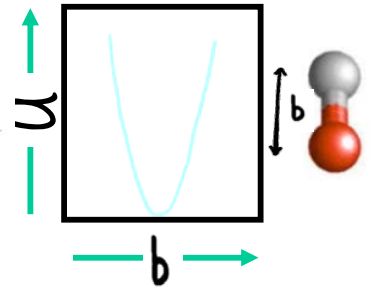


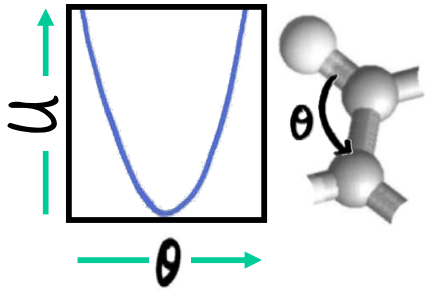

```
[ moleculetype ]
; molname       nrexcl
SOL             2

[ 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
```

MOLECULAR POTENTIAL ENERGY

$$U = \sum \tfrac{1}{2} K_b (b - b_0)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_0)^2$$

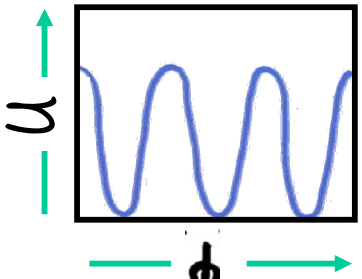
All Bonds — Hooke 1635 — All Angles

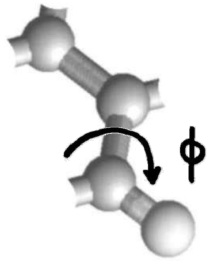




$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

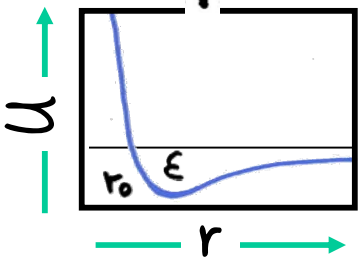
All Torsion Angles — Fourier 1768

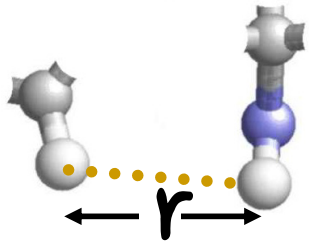




$$+ \sum \varepsilon \left[\left(r_0 / r \right)^{12} - 2 \left(r_0 / r \right)^{6} \right]$$

All Nonbonded pairs — Van der Waals 1837

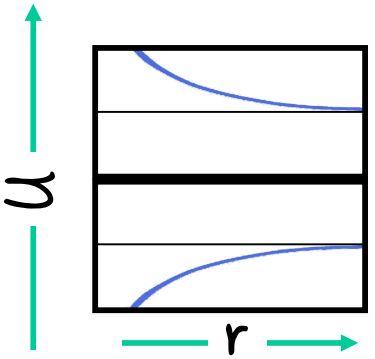


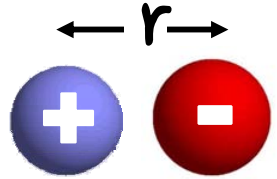


$$+ \sum 332 q_i q_j / r$$

All partial charges — Coulomb 1736

Simple sum over many terms





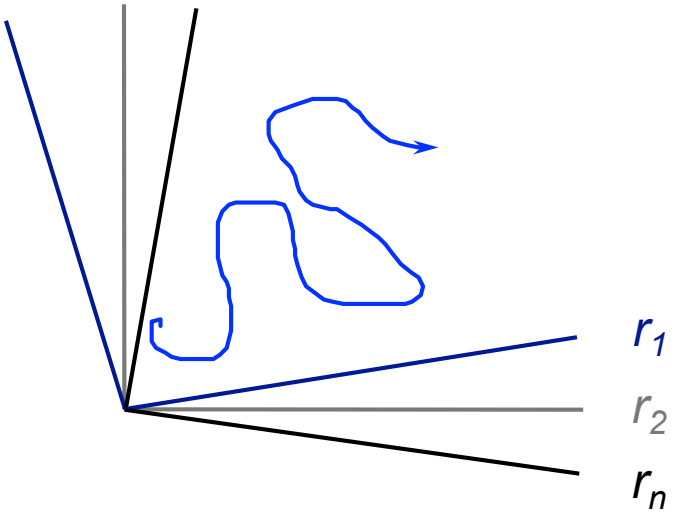
Force field

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

Sampling

Molecular dynamics:

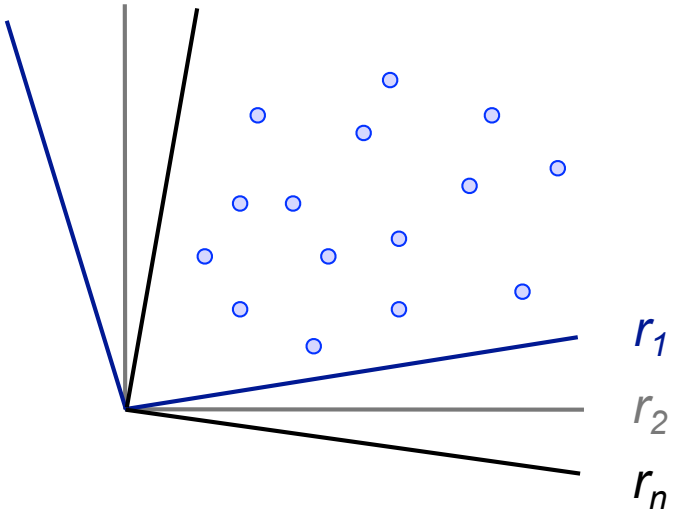
solve equations of motion



MD

Monte Carlo:

importance sampling



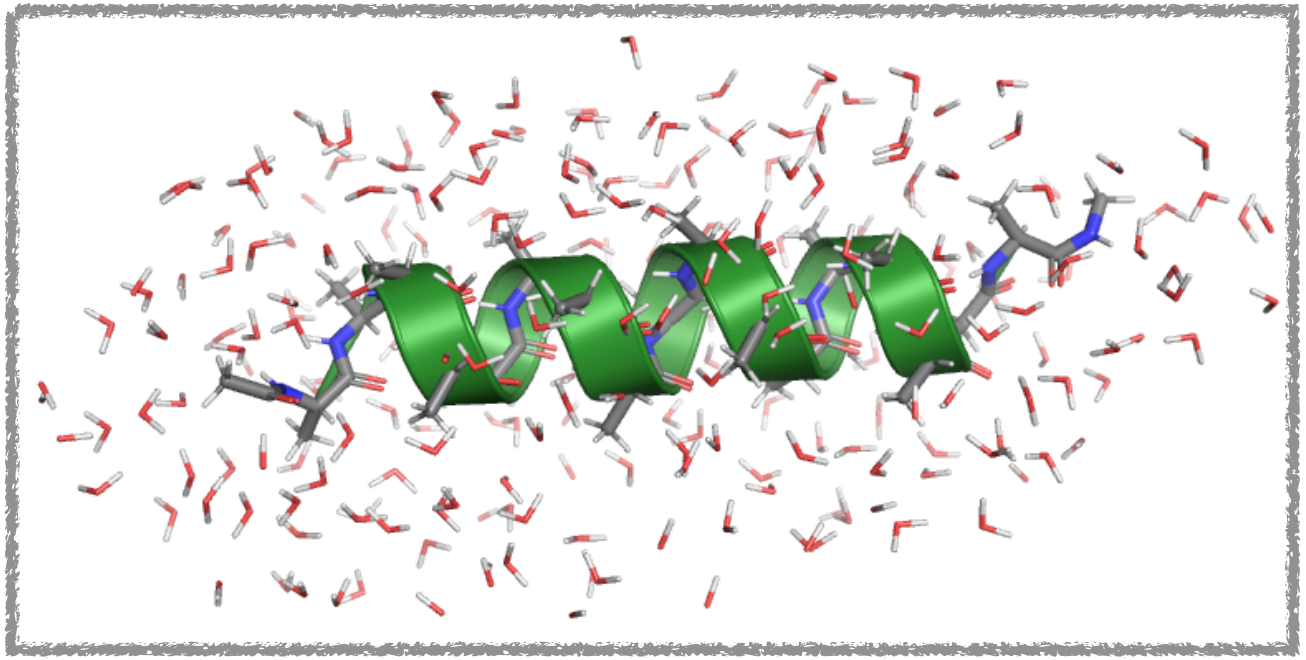
MC

Molecular dynamics

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

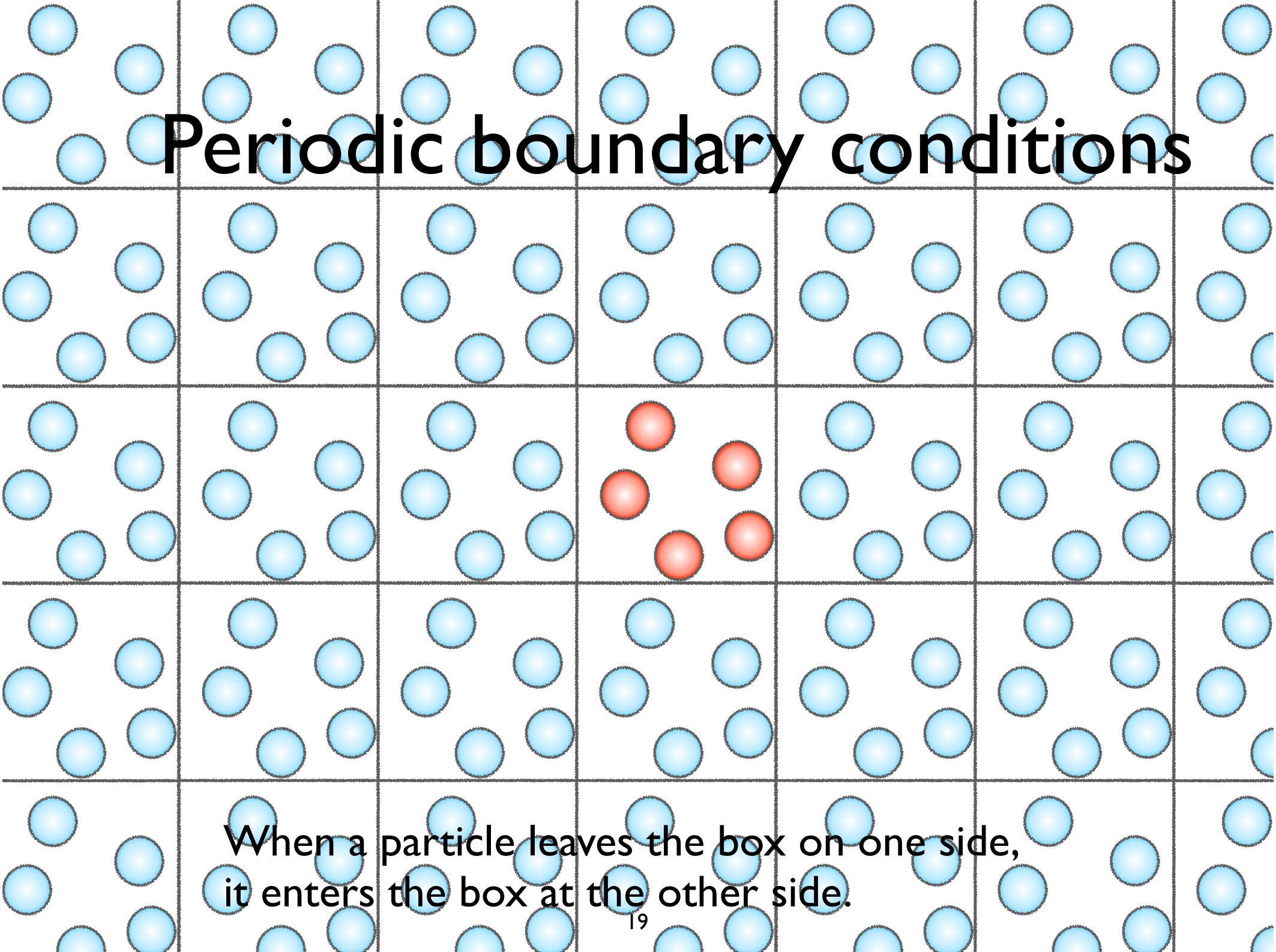
} repeat

Simulating an alpha-helix



constant N, p, T

Periodic boundary conditions



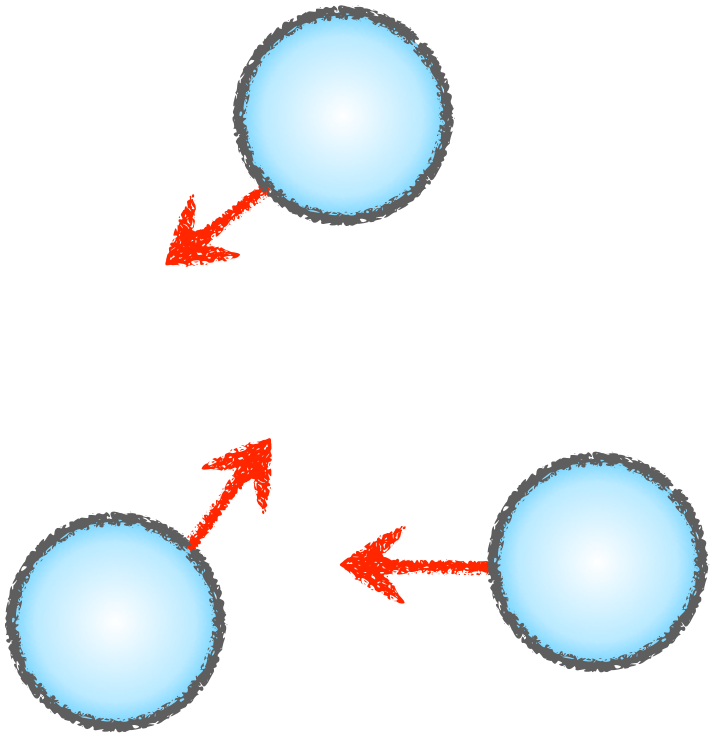
When a particle leaves the box on one side,
it enters the box at the other side.

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

$$v^2 \backsim T$$



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

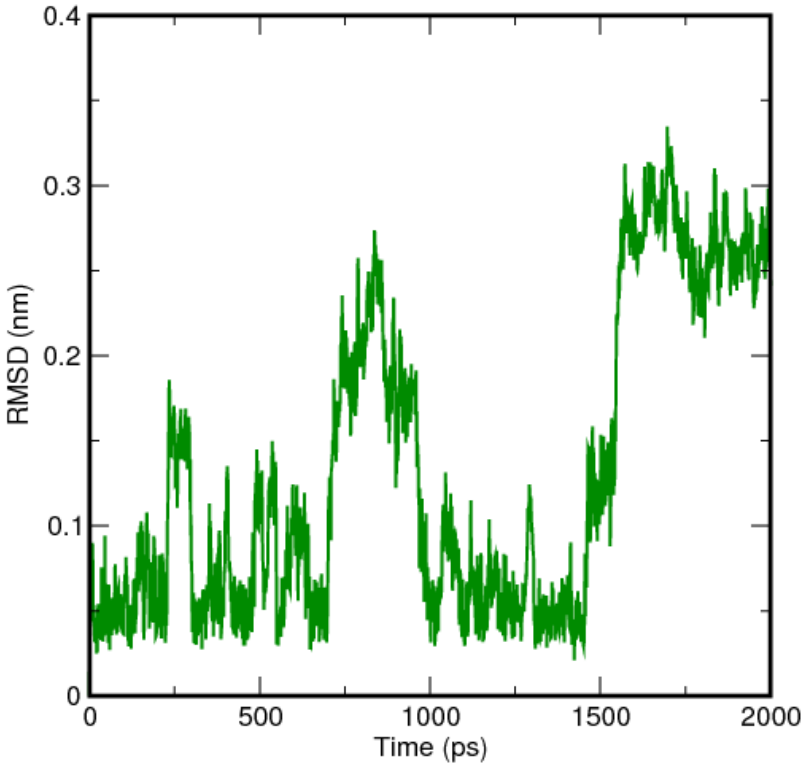
$$\sqrt{\frac{\sum_i^N (r_i - \hat{r}_i)^2}{N}}$$

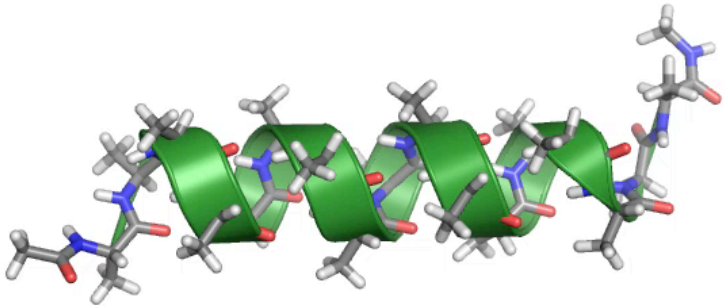
N - number of atoms

i - atom index

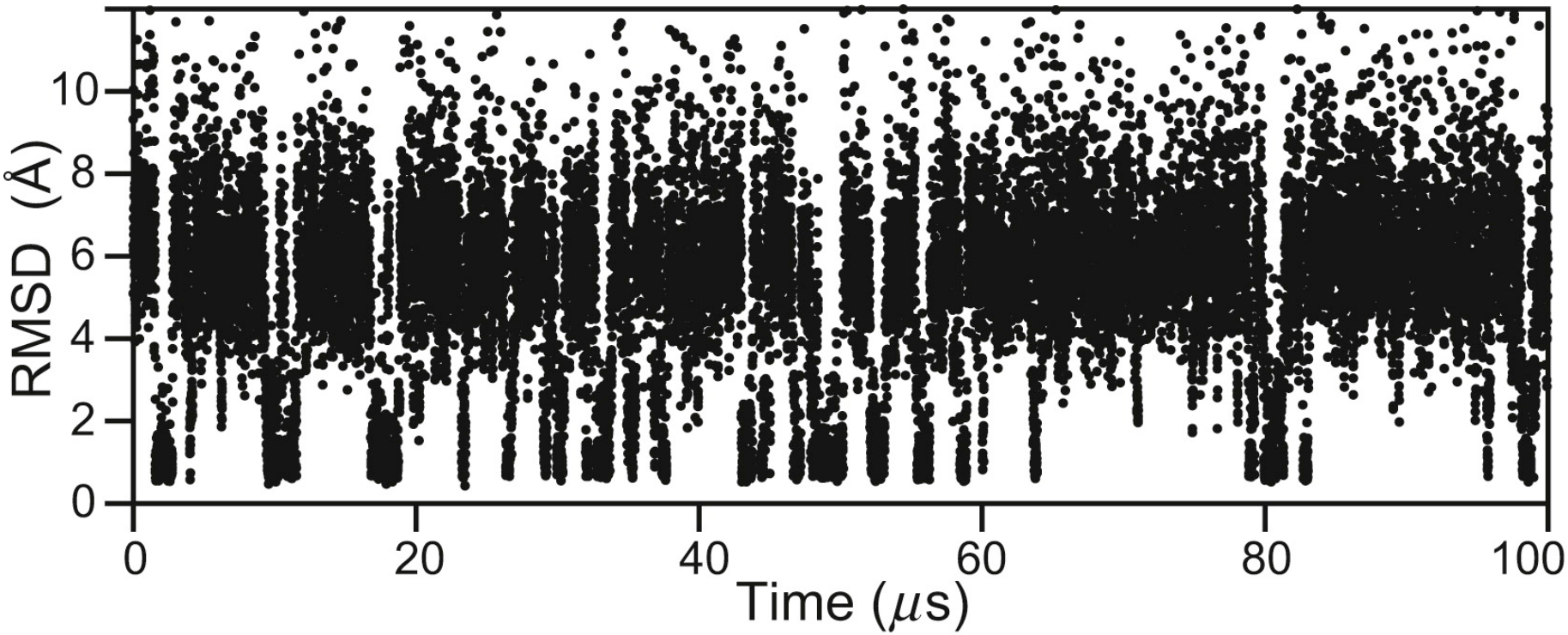
r - position

$\hat{r}$ - 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 $\mathbf{\Gamma} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ 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

$$\langle A \rangle = \int P(\mathbf{\Gamma}) A(\mathbf{\Gamma}) d\mathbf{\Gamma}$$

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

Ergodicity theorem

Suppose we have an ensemble average of a system defined by U(**Γ**) obtained by MC.

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

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

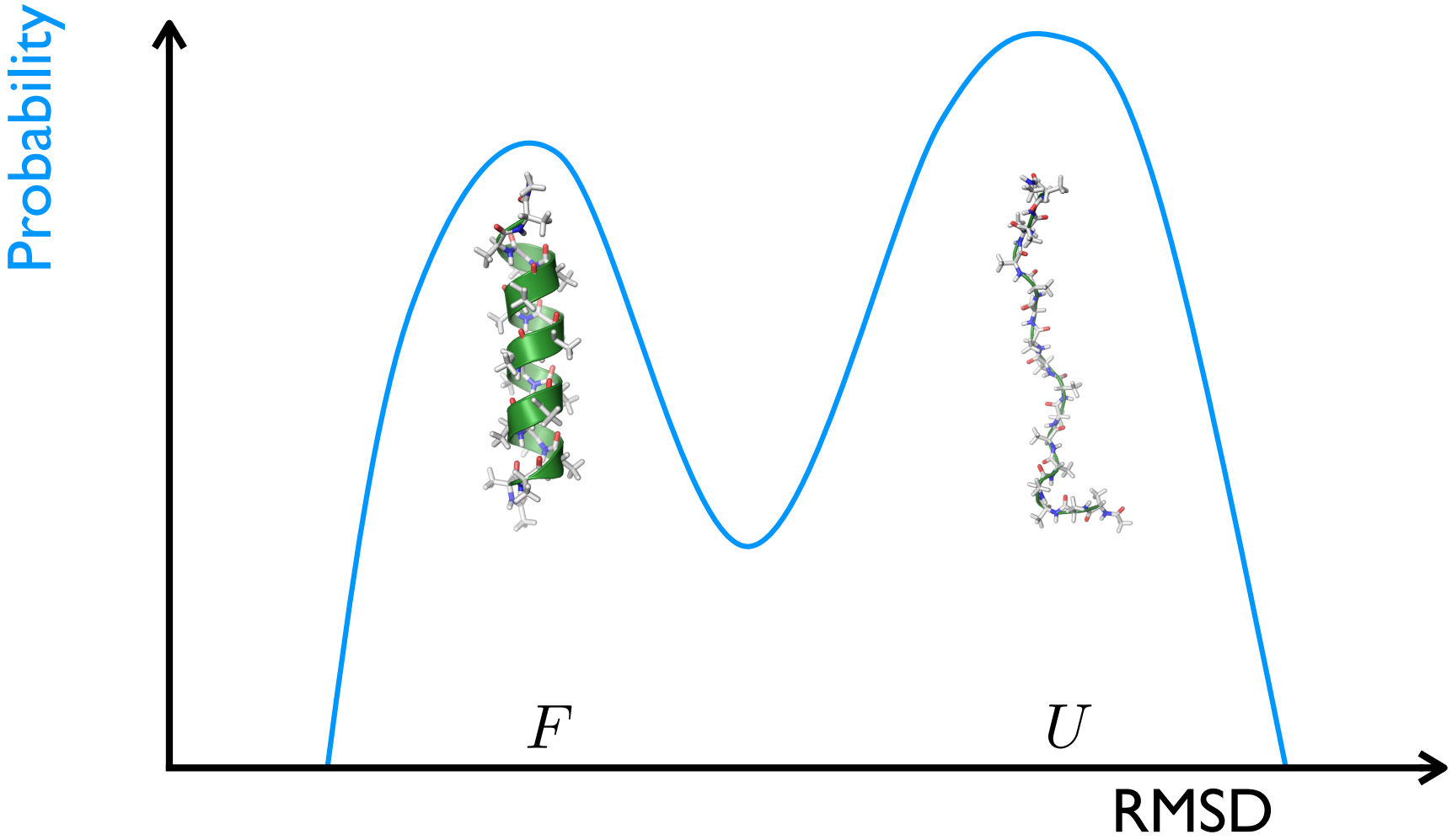
$$\bar{A} = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} A(t) dt$$

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

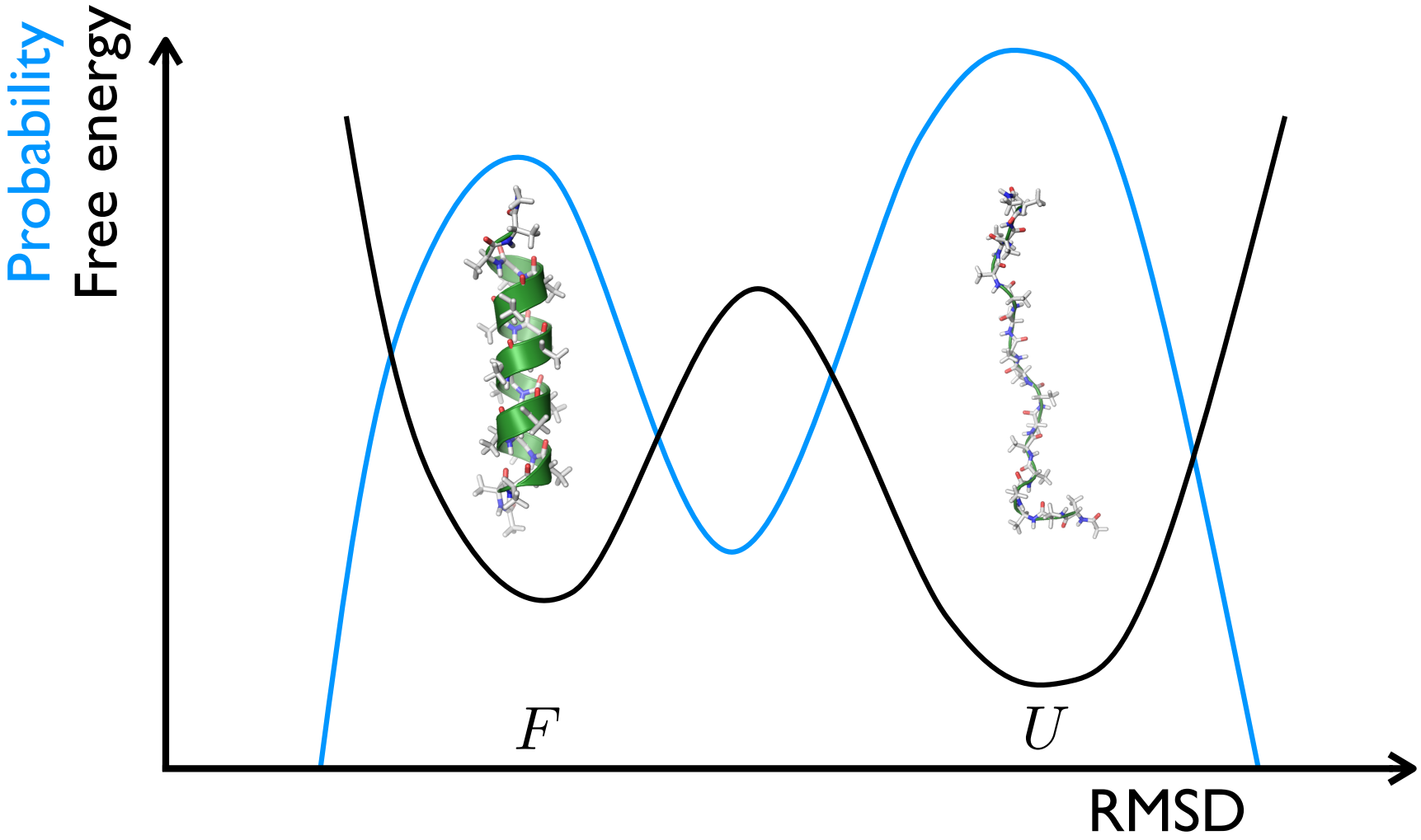
$$\bar{A} = \langle A \rangle$$ MC and MD give the same averages!

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

Probability histogram

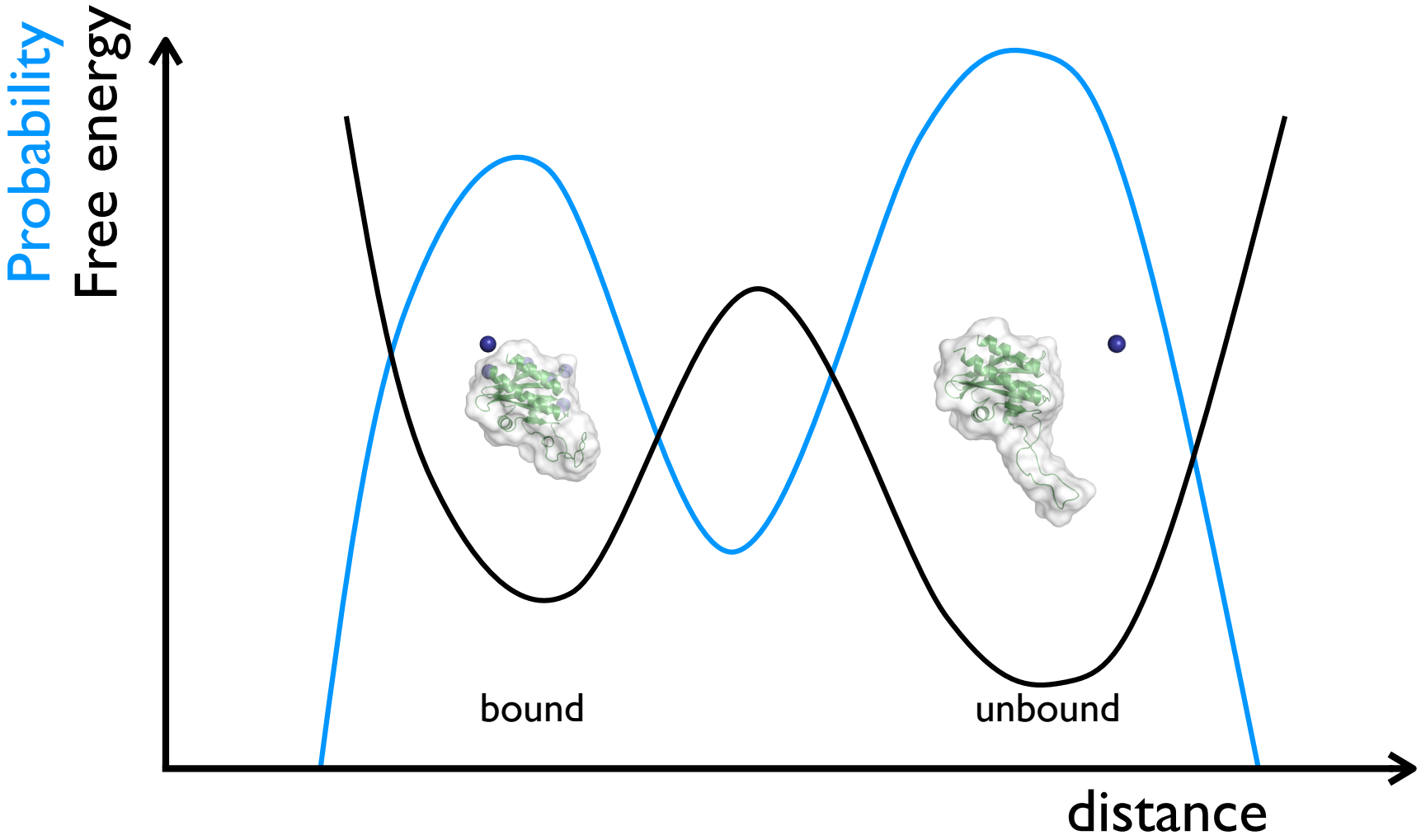


Free energy profile



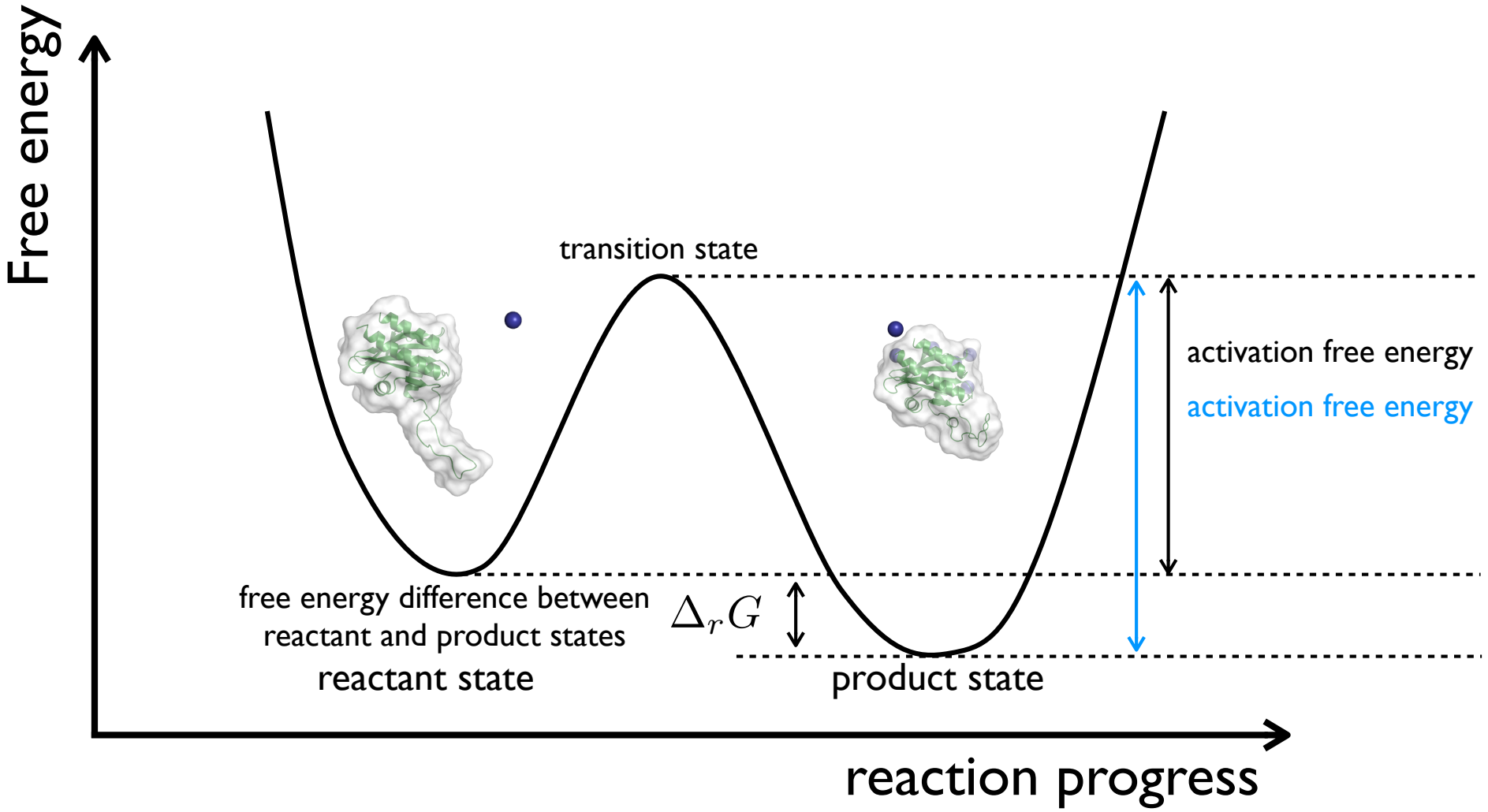
Free energy = -ln(Probability) in units of k_BT

Free energy profile

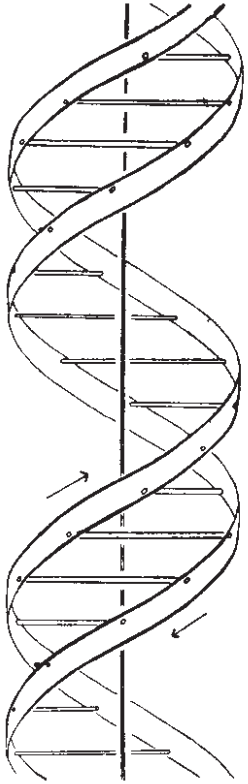


Free energy = -ln(Probability) in units of k_BT

Free energy profile

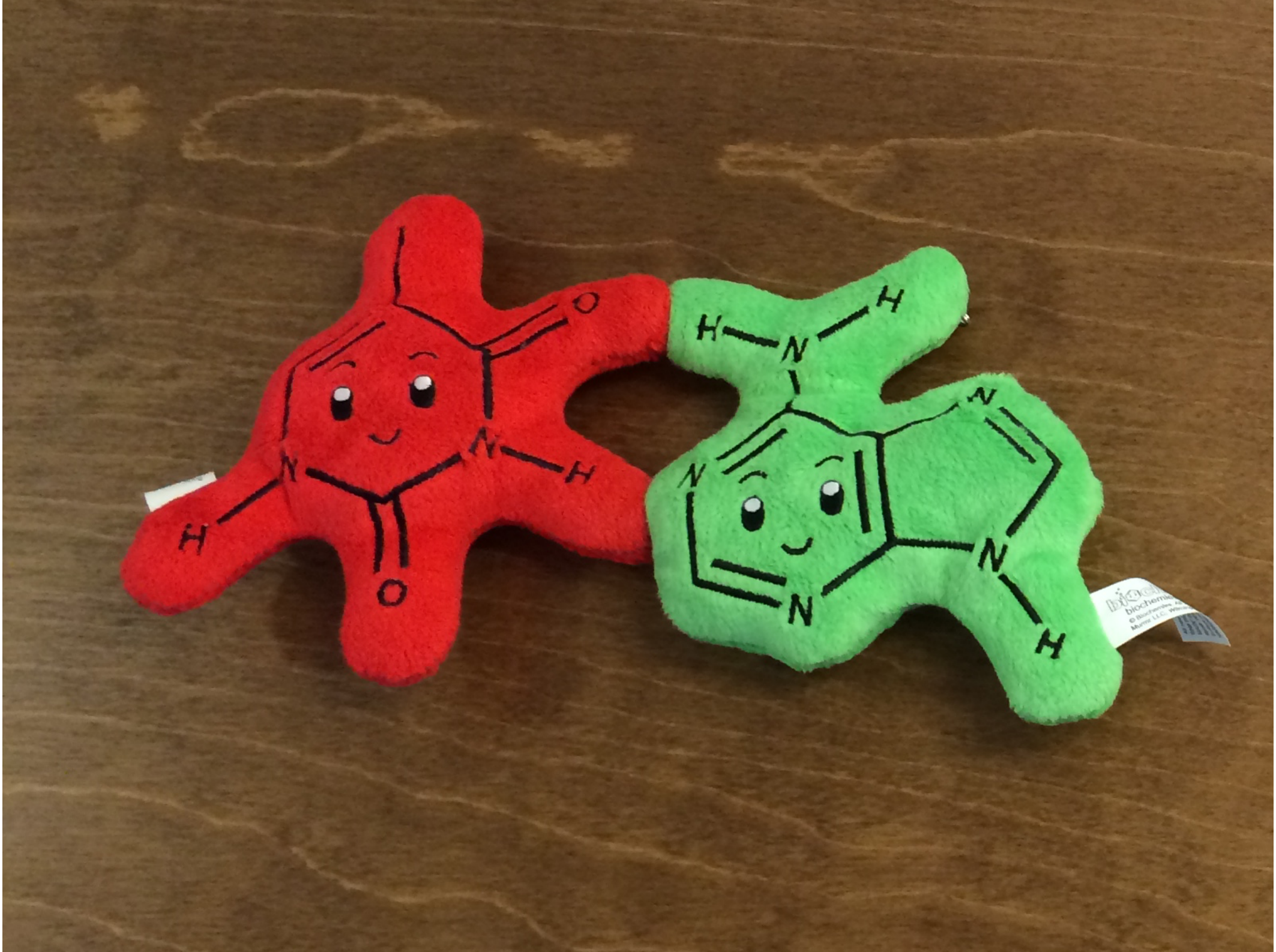


Structure of DNA

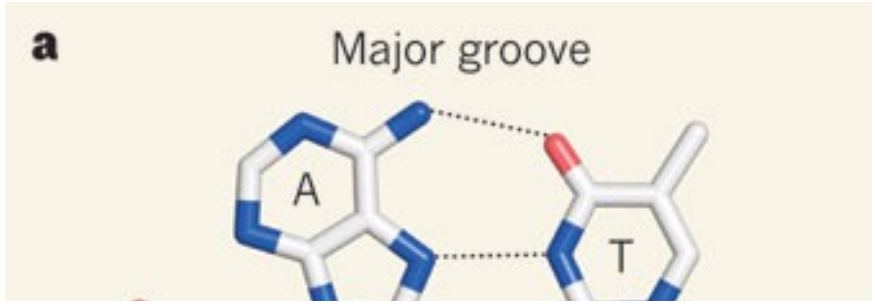


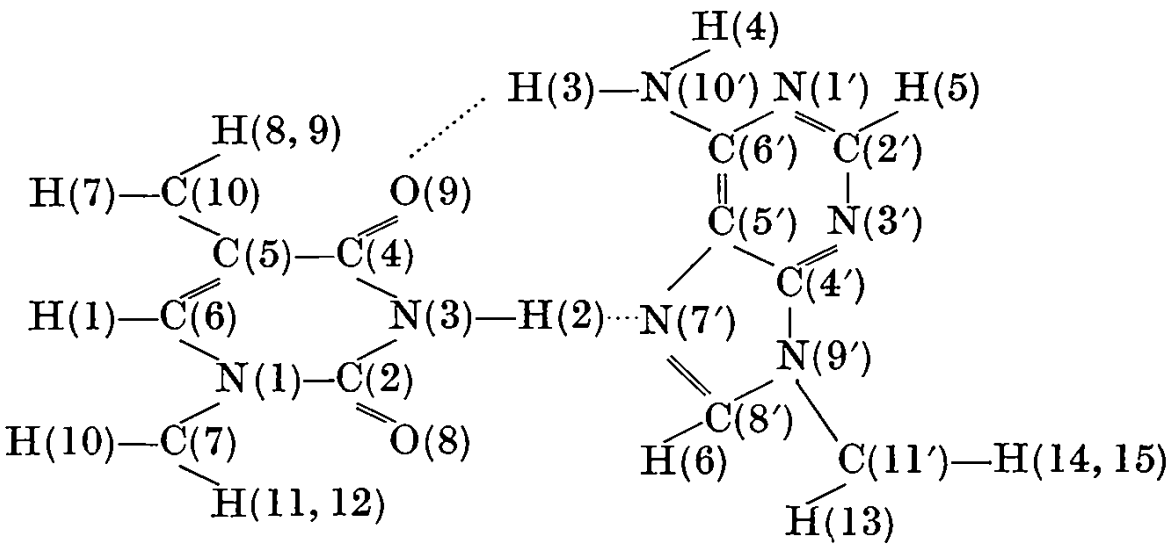
This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

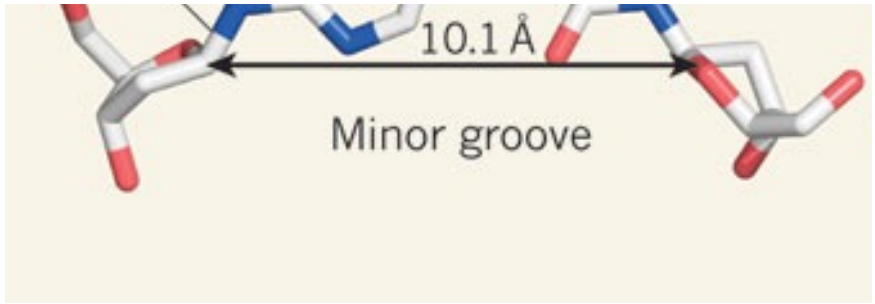
Watson & Crick *Nature* 1953 v. 171 p. 737


O
N
H
H
O
H
N
H
N
N
N
N
H

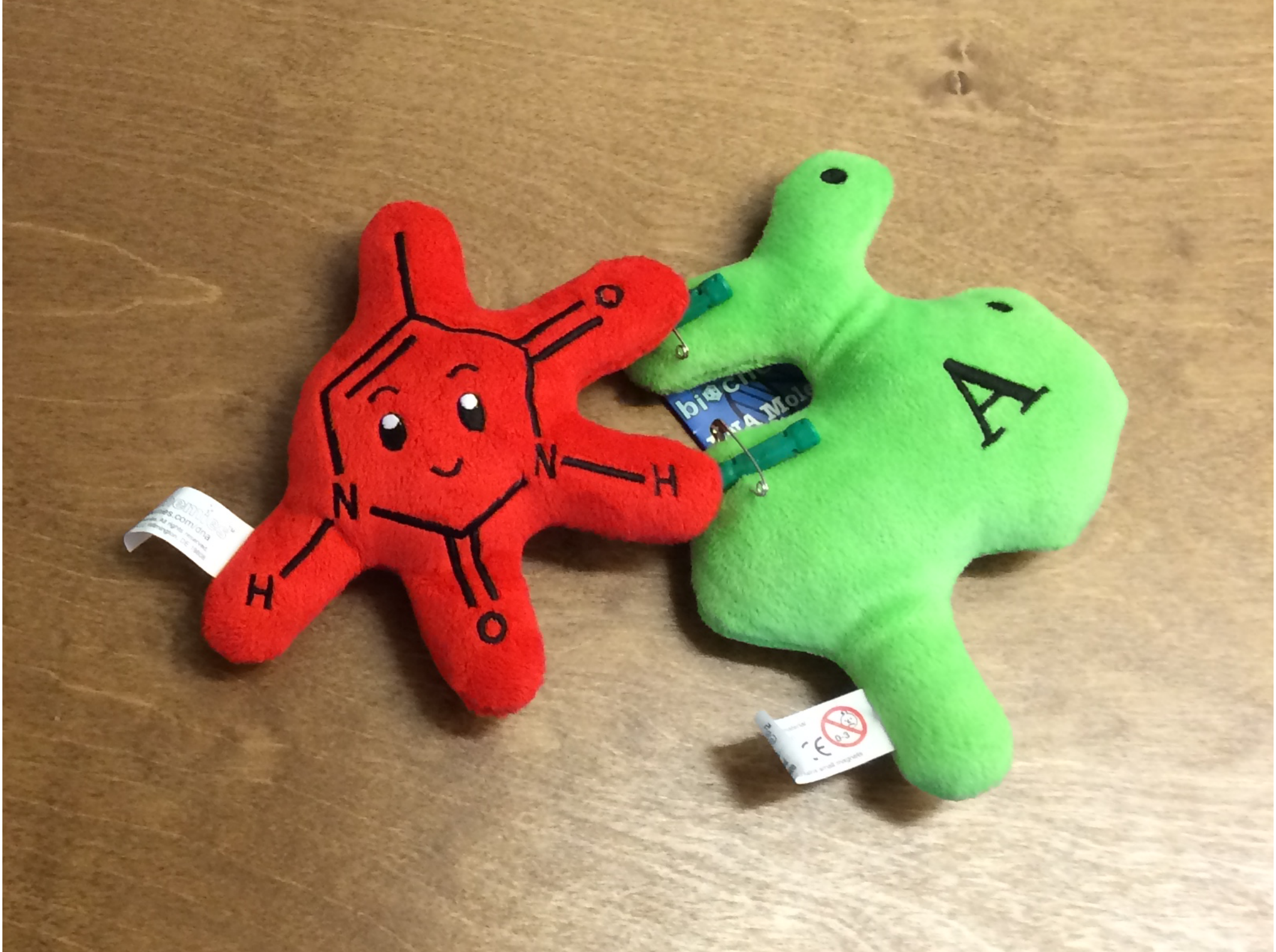
Hoogsteen basepairing



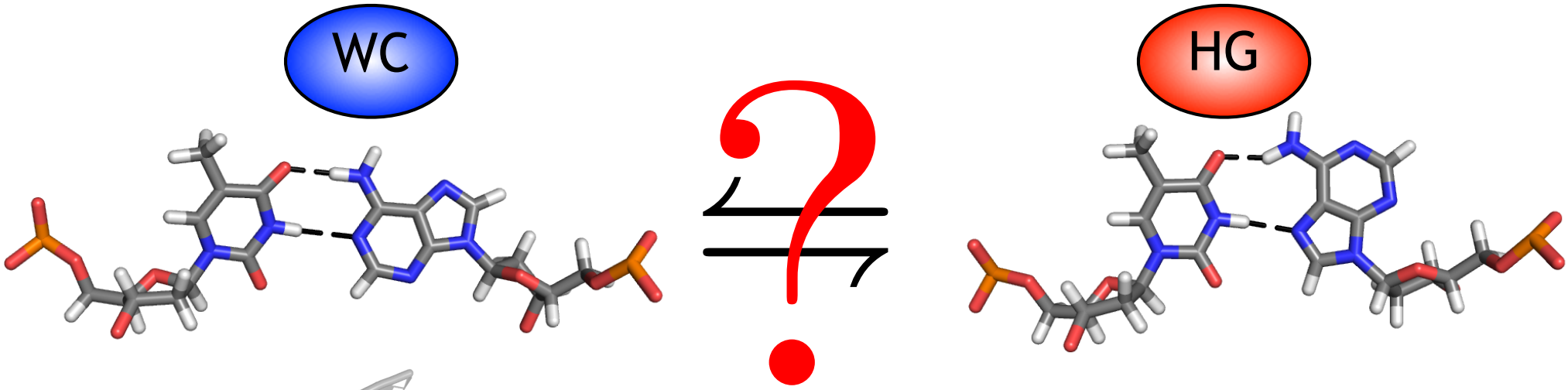


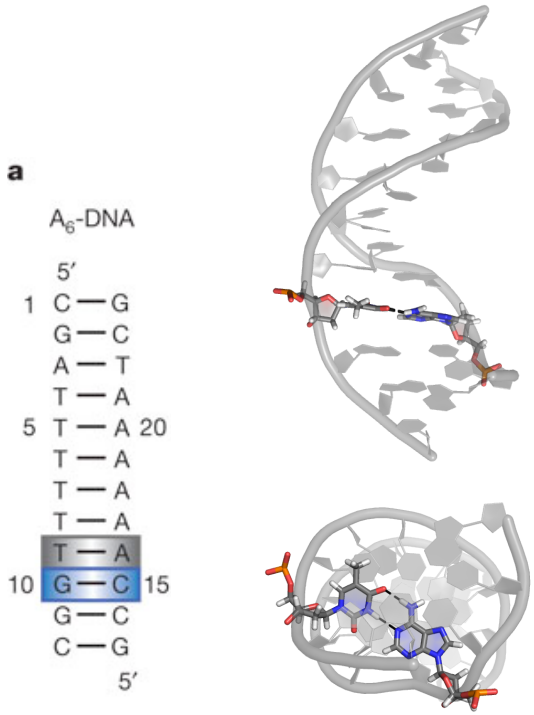


Hoogsteen *Acta Cryst.* 1963
Honig & Rohs *Nature* 2011


O
N
H
N
H
O
A

What is the mechanism?



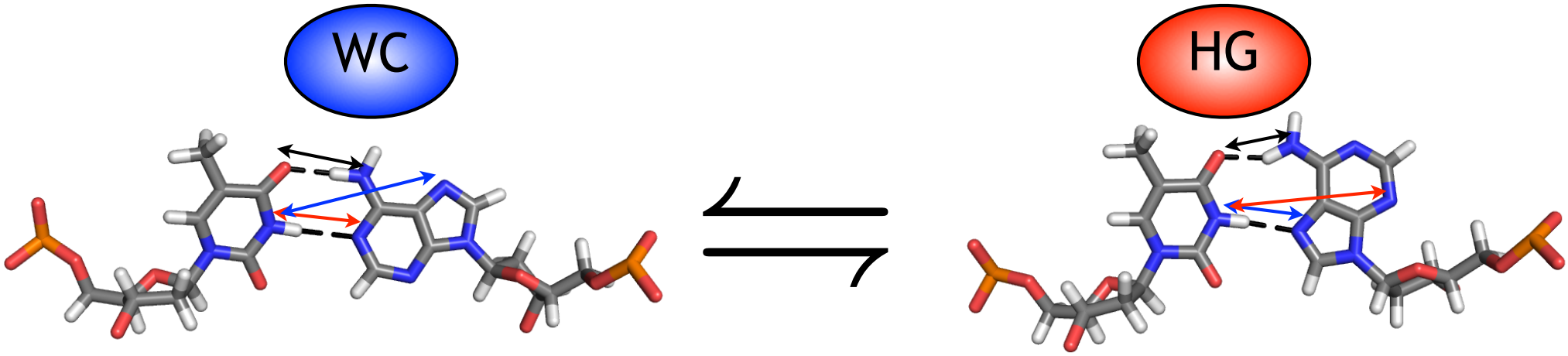


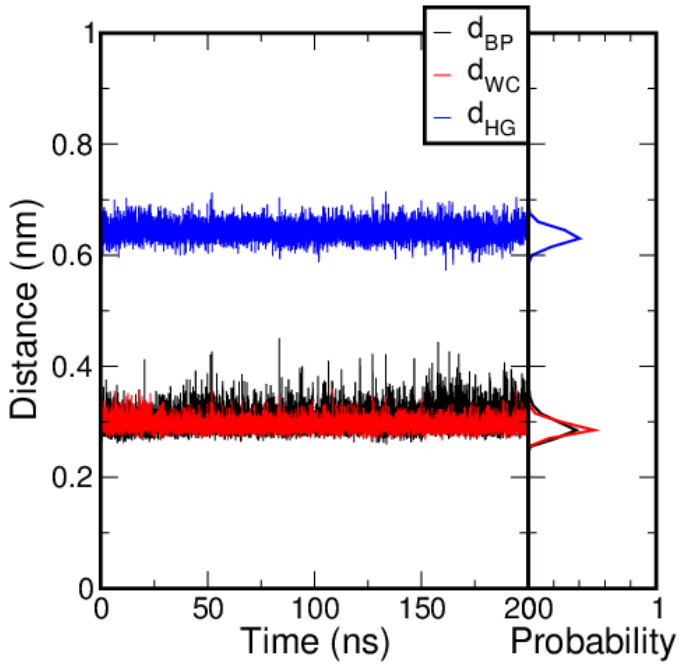
Molecular Dynamics:
5′–GGATTTTTTGGC–3′
3′–CCTAAAAAACCG–5′
AMBER03
TIP3P water + 25 mM NaCl
NpT ensemble, T = 300 K
20868 atoms
gromacs v4.6.1

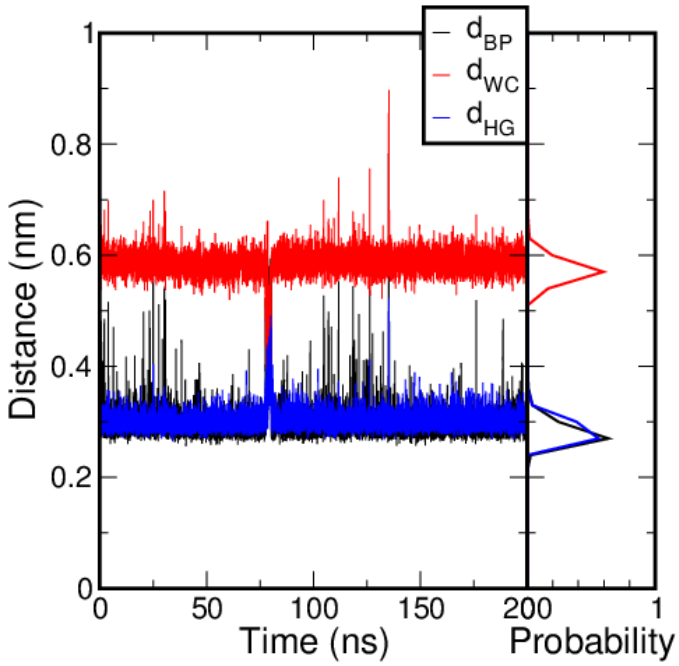
Nikolova et al. Nature 2011

Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

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

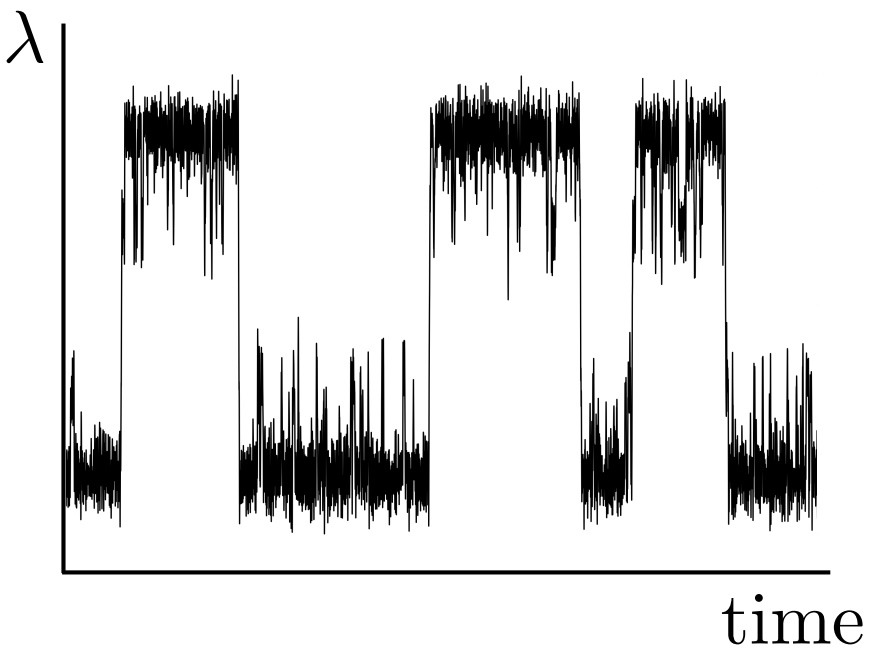


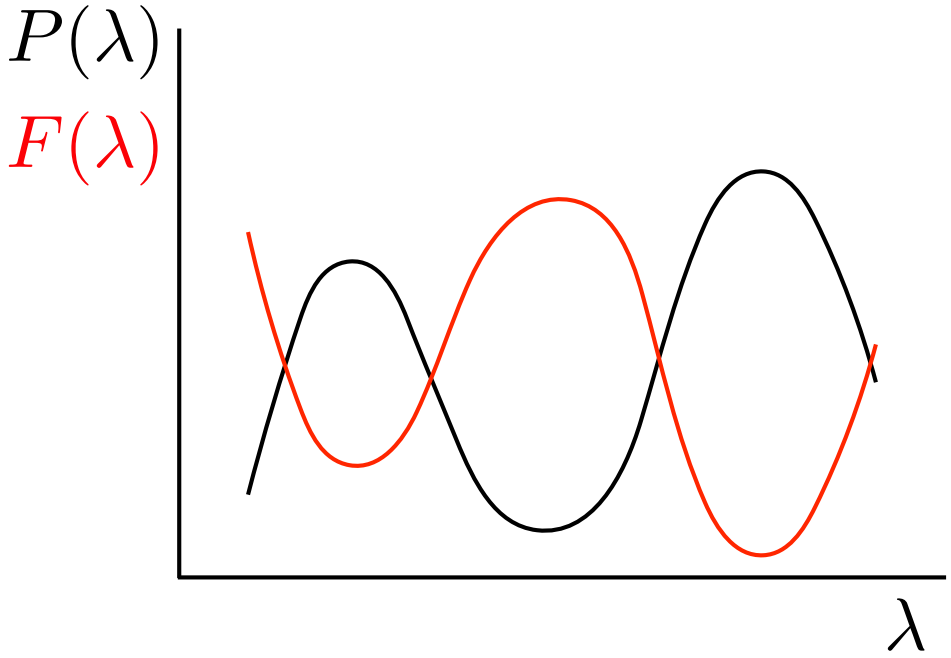




Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Computing a free energy profile from a molecular dynamics simulation

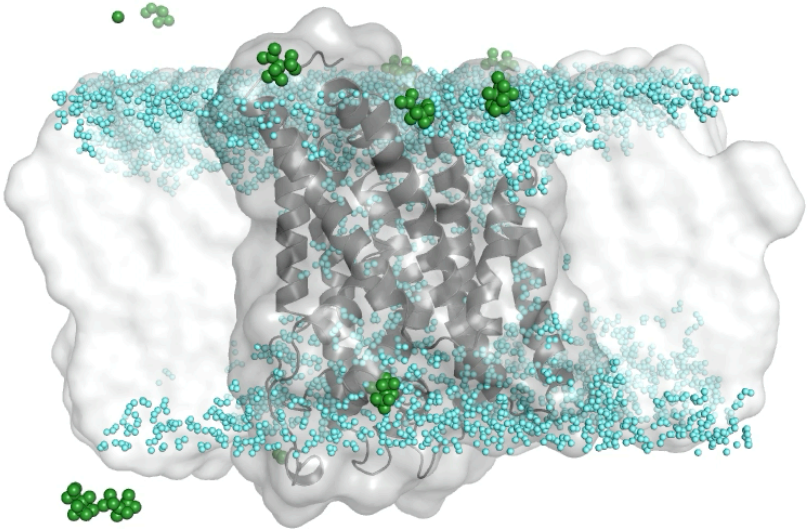




$$F(\lambda) = -k_B T \ln(P(\lambda))$$

Problem: little sampling except in minima of F(λ)

All-atom force field molecular dynamics of biomolecular systems



Blinker *et al.* IJMS 2021

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

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

numerically

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

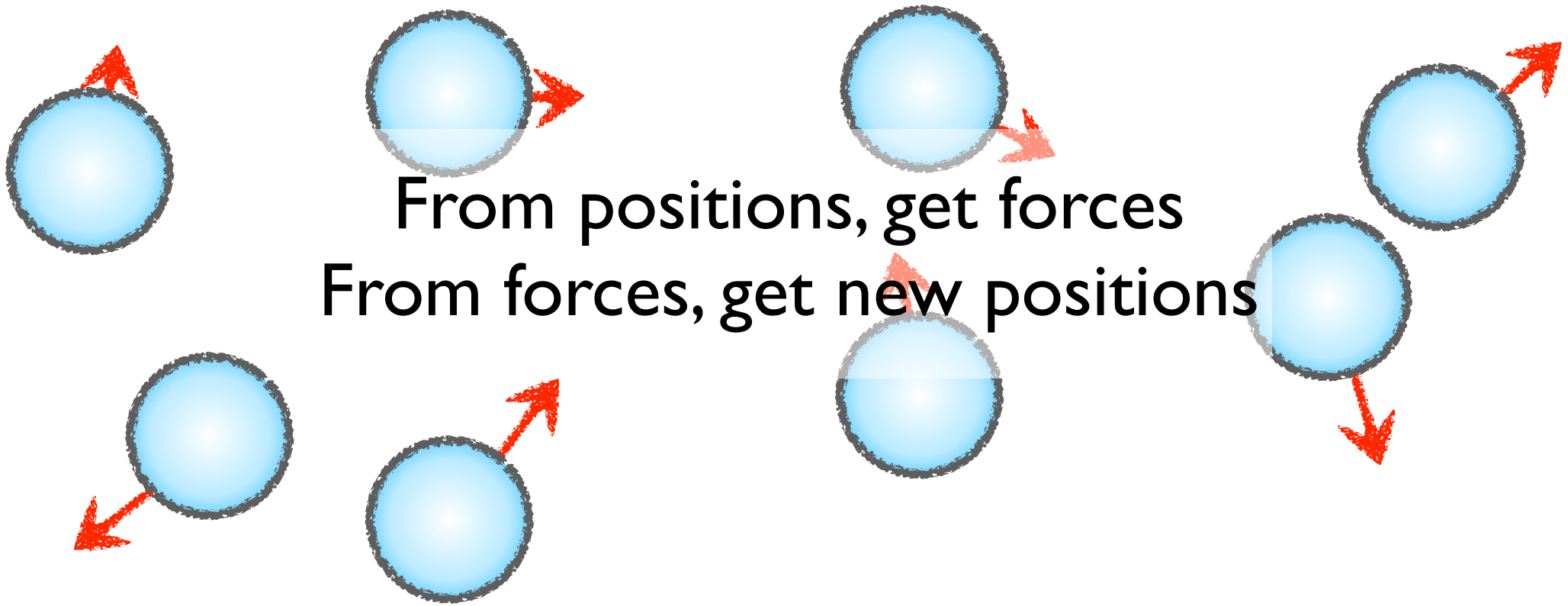
The force F is given by the gradient of the potential *V(r)*.

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

bonds bends torsions non-bonded

Numerical approach

Snapshots



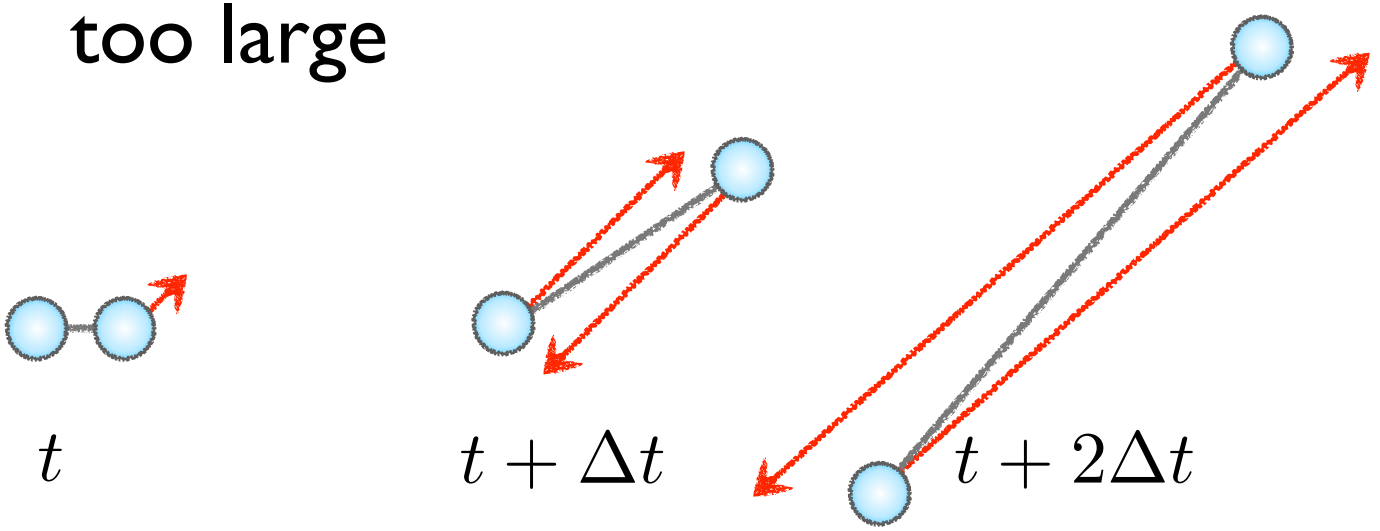
From positions, get forces

From forces, get new positions

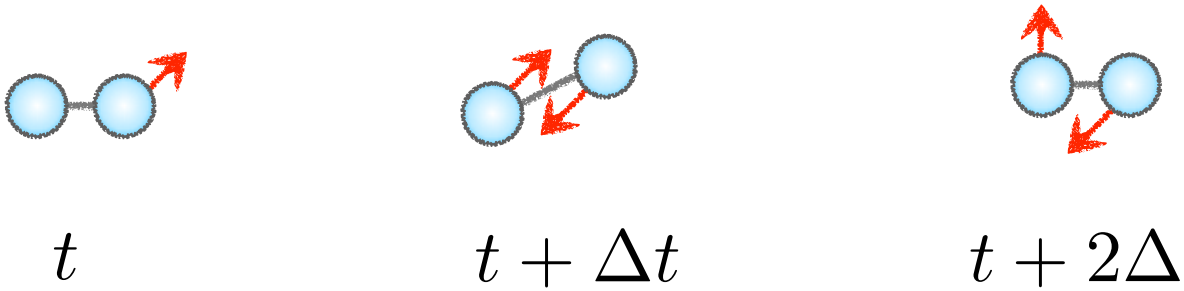
t $t + \Delta t$ $t + 2\Delta t$ $t + 3\Delta t$

Choosing the time step

Δt too large

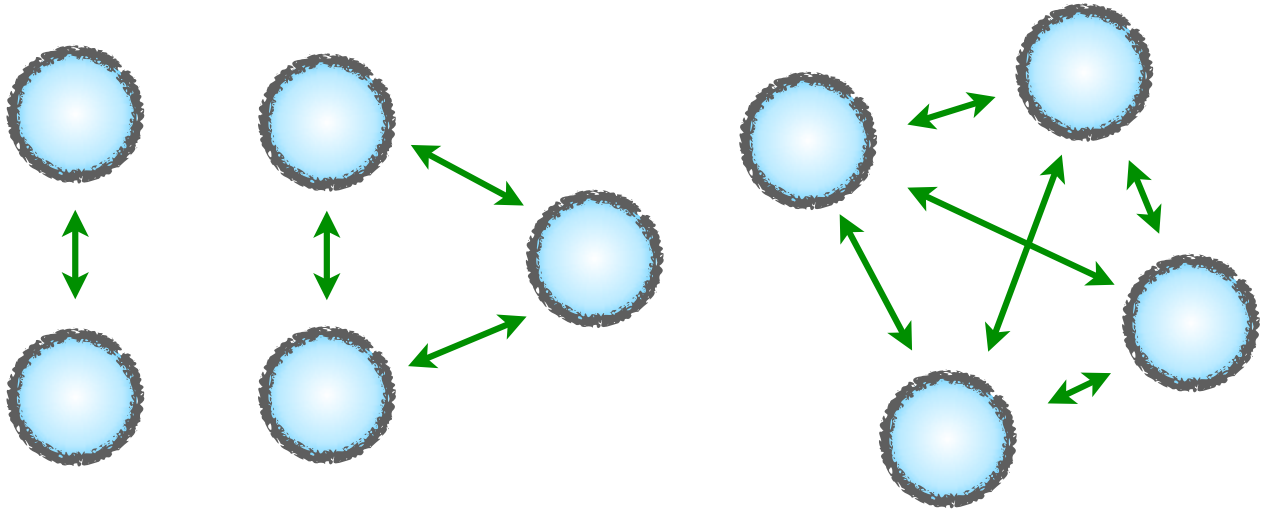


Δt just fine (2 fs for force field MD)



Force computations

Assumption: pairwise interactions



system

number of force computations

1 3 6

A system of a 150 residue protein in water

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

$\Delta t = 2\ \mathrm{fs} = 2 \cdot 10^{-15}\ \mathrm{s} = 0.000000000000002\ \mathrm{s}$

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

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

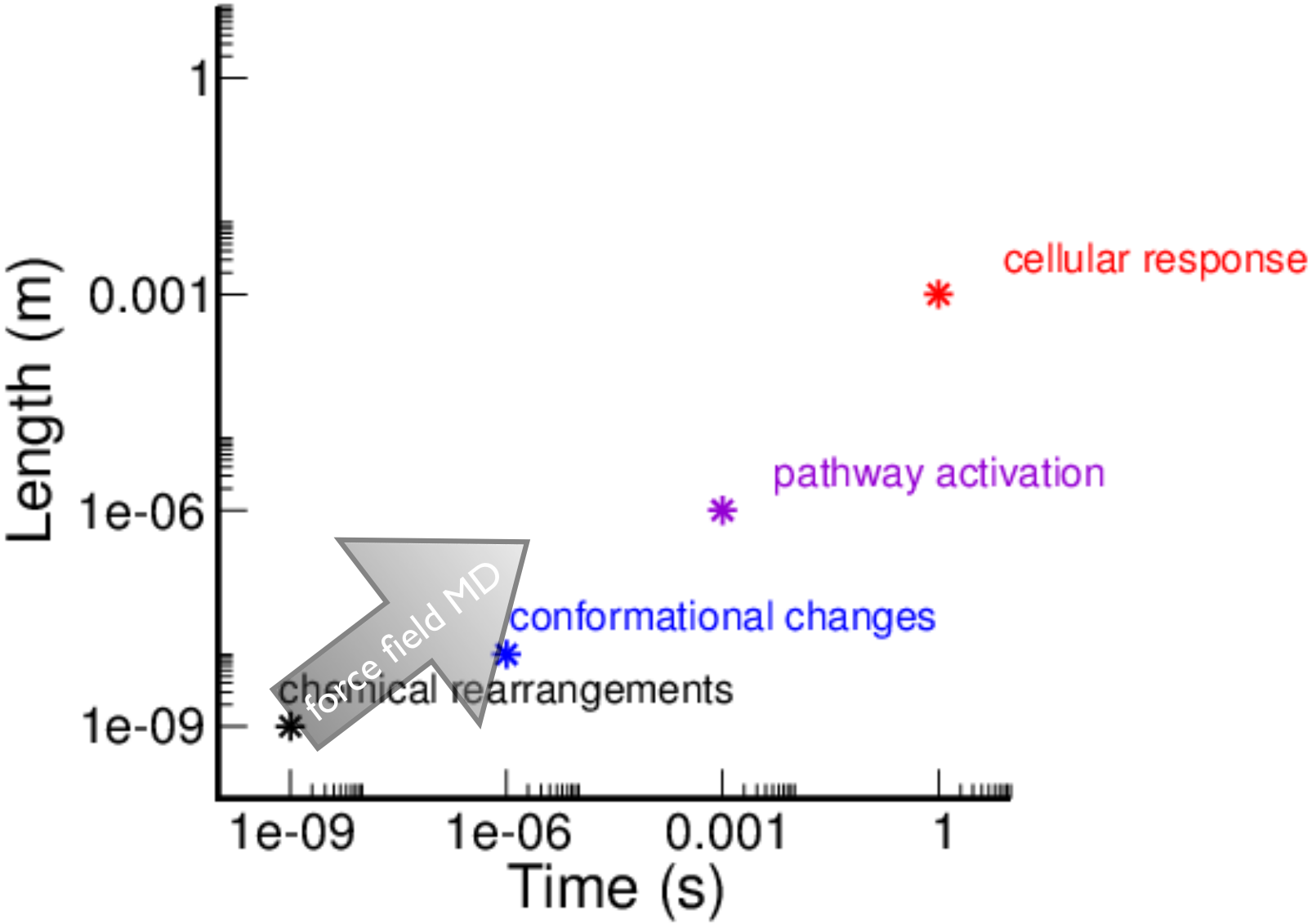
1 ns = 500.000 steps = 13.000 s ≈ 22 minutes

1 s = 500.000.000.000.000 steps
≈ 42.000 year

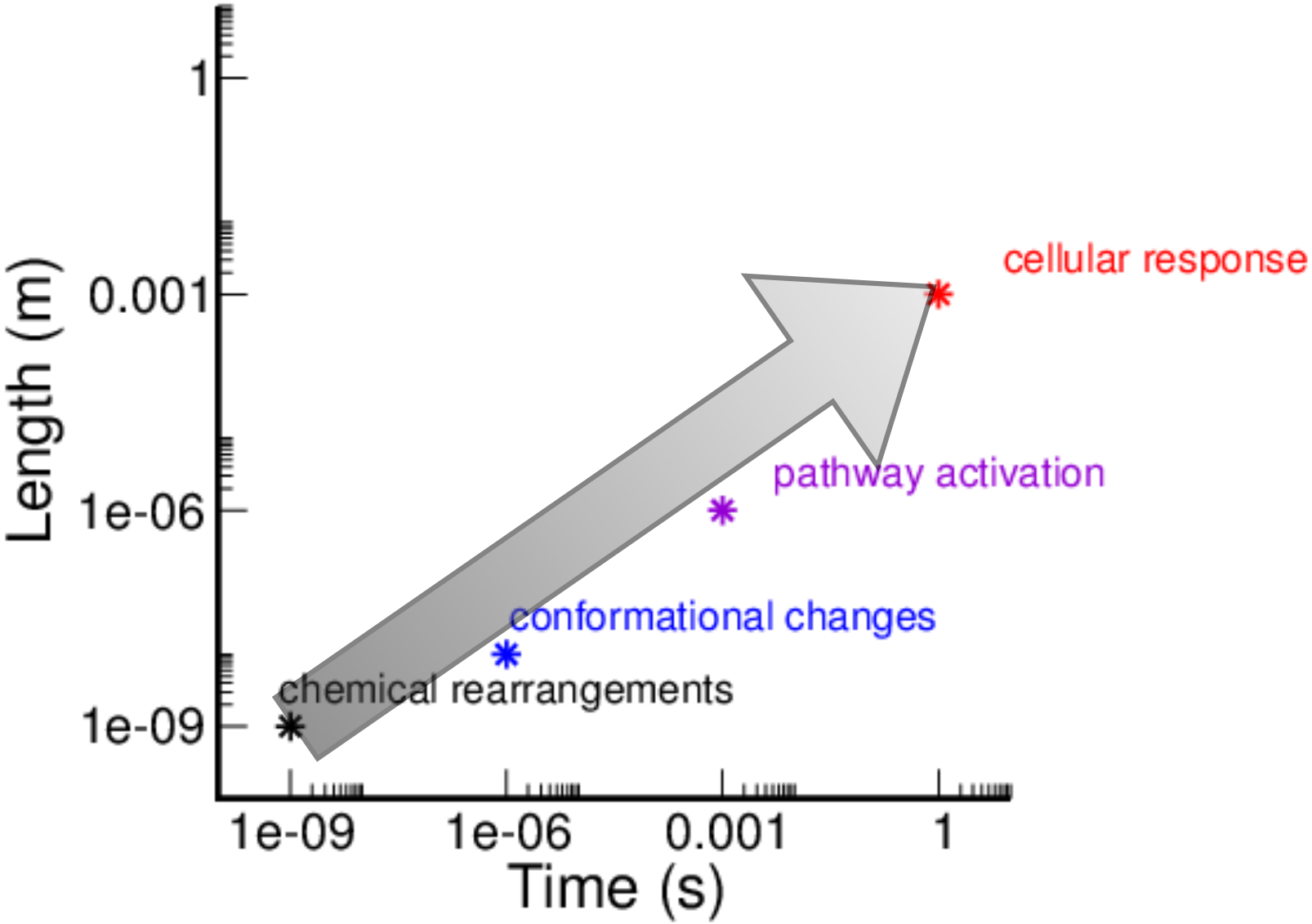
A protein in water

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

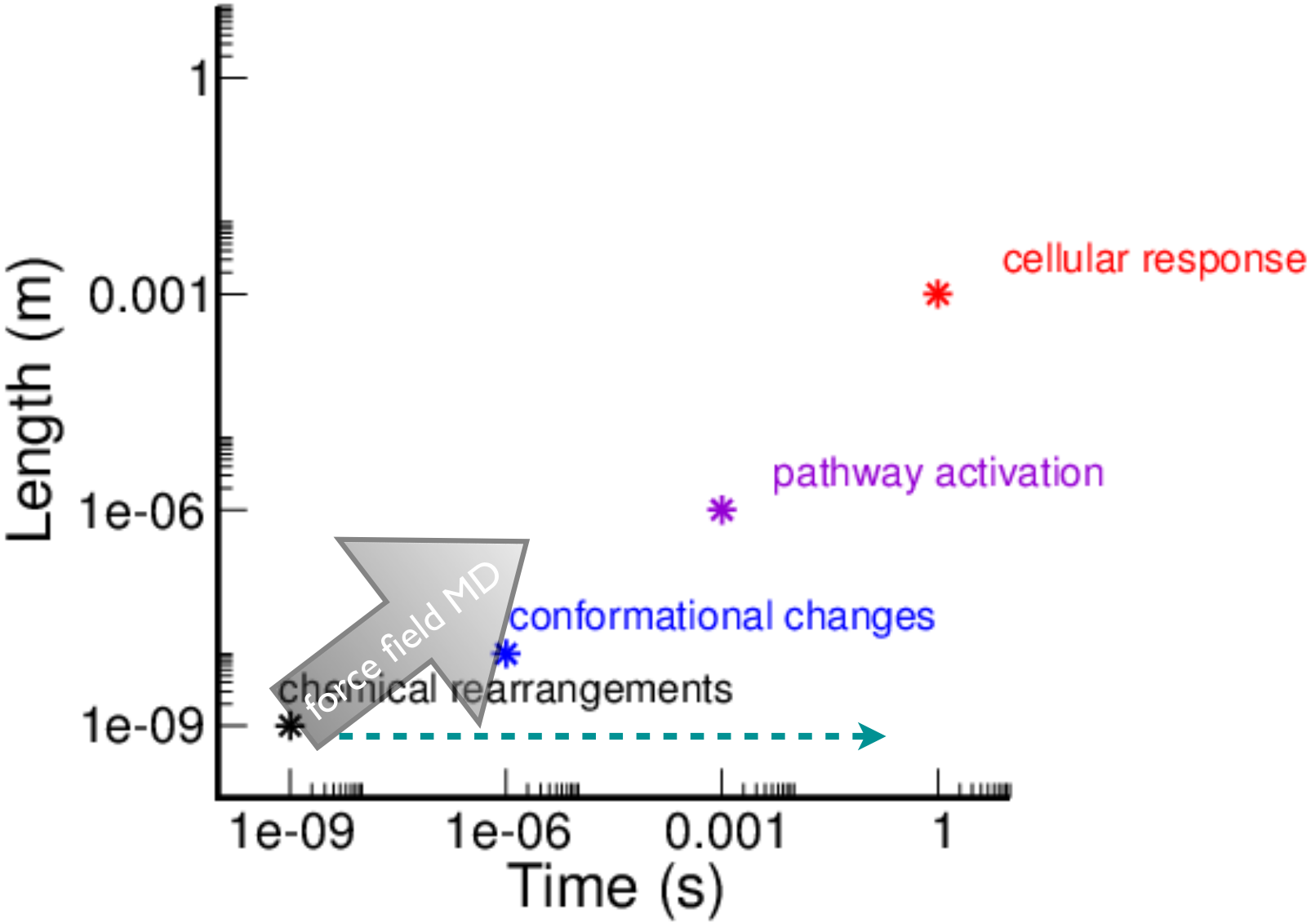
Where we are…


Length (m)
1
0.001
1e-06
1e-09
cellular response
pathway activation
conformational changes
chemical rearrangements
force field MD
1e-09
1e-06
0.001
1
Time (s)

And where we want to go!



Bridging the time gap



Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

Part 2: Biased sampling

Part 3: Path sampling

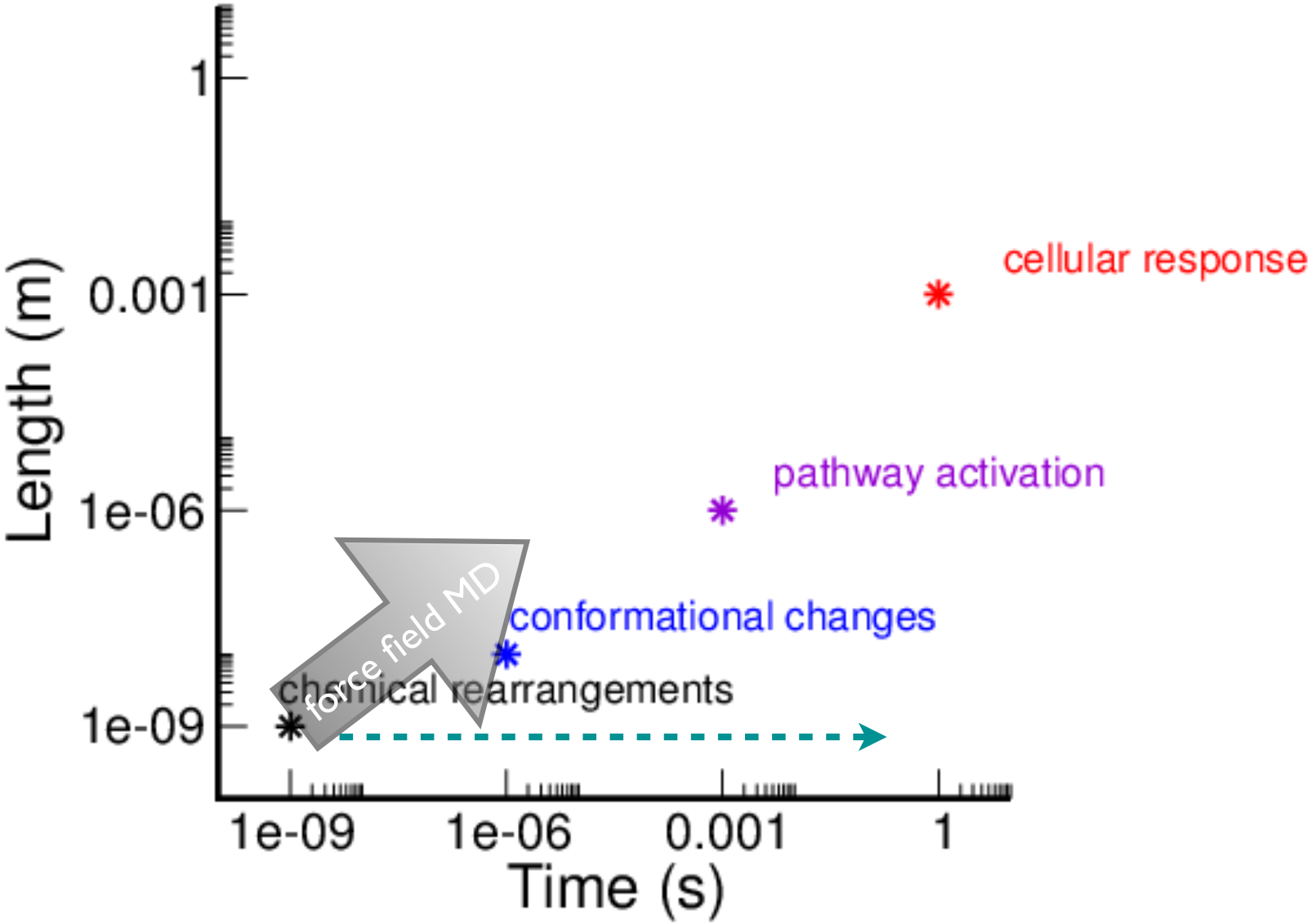
Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

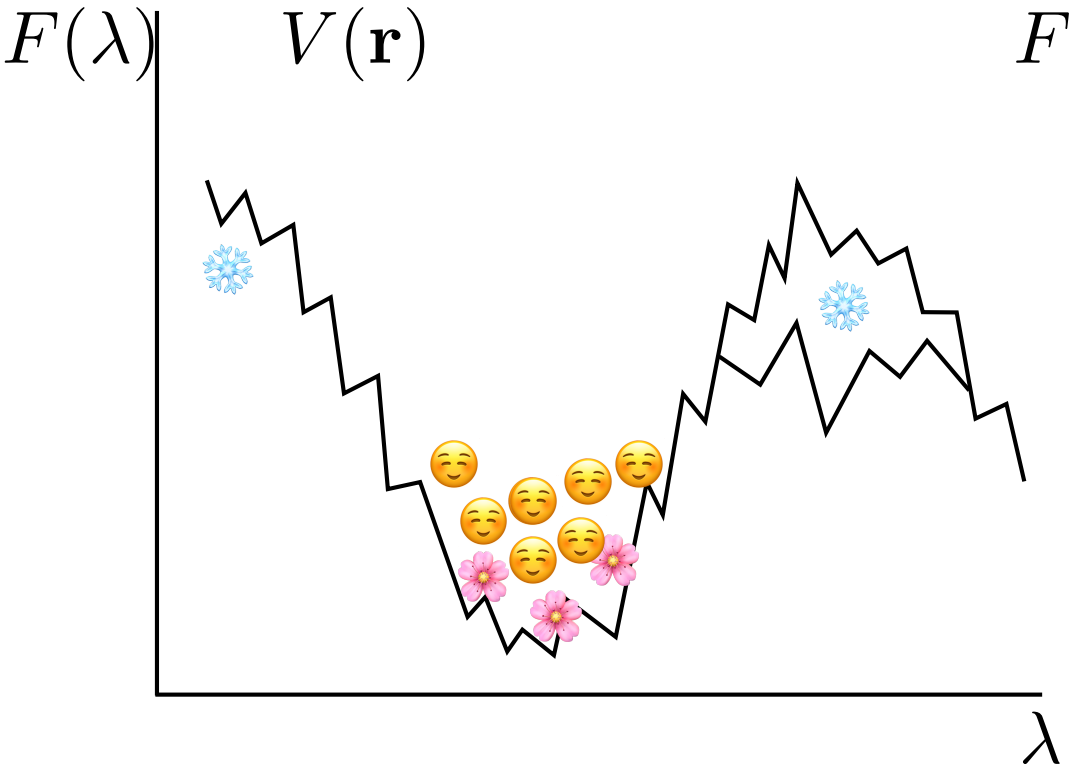
Part 2: Biased sampling

Part 3: Path sampling

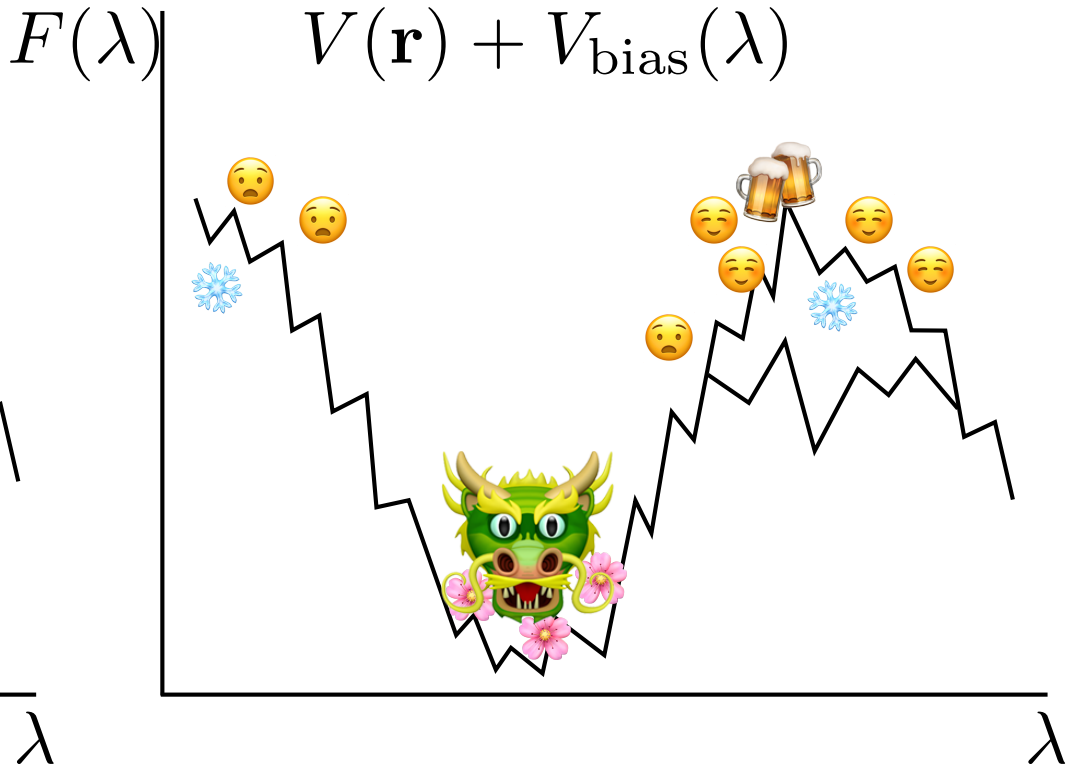
Bridging the time gap



Biased sampling



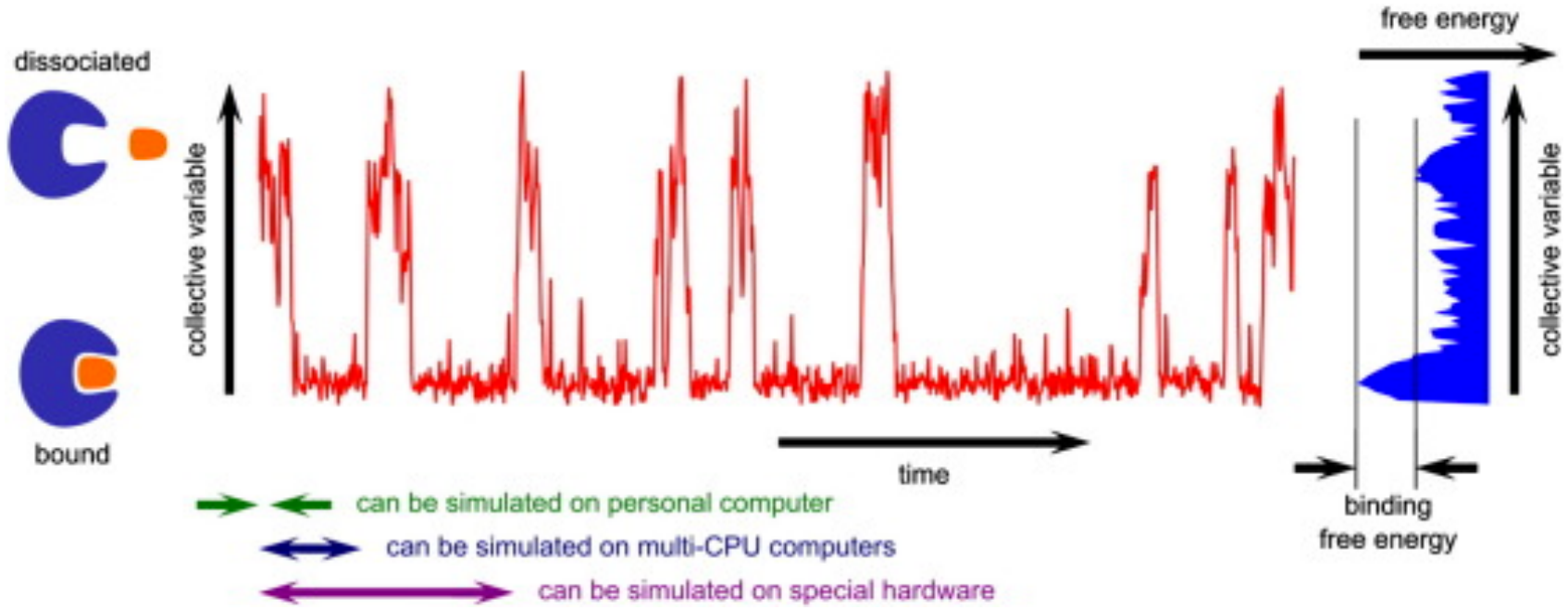
Straightforward MD



Biased sampling

Adapted from lecture Daan Frenkel
Understanding Molecular Simulation course 2014

Reaction coordinate

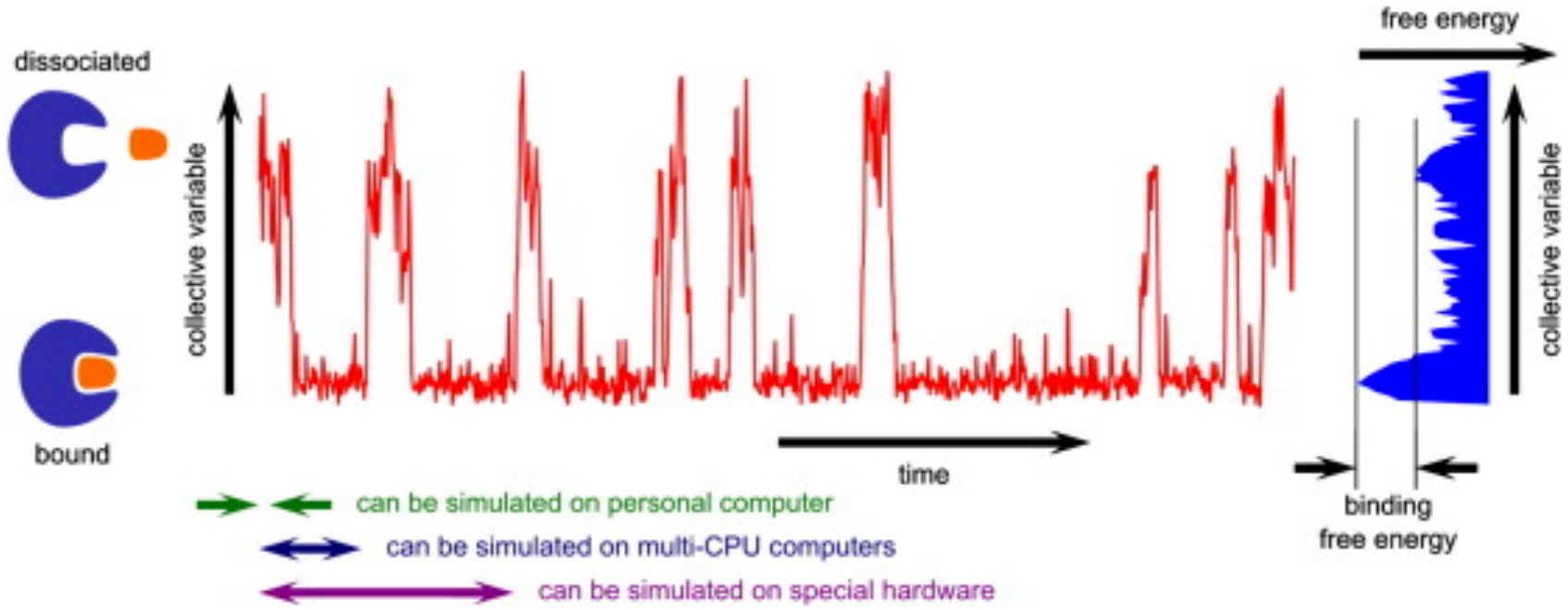


A collective variable (CV) is a function λ of the 3N-dimensional configuration **r**.
λ maps **r** onto an M-dimensional space **z**.
M << 3N

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

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

Reaction coordinate



A collective variable (CV) is a function Q of the 3N-dimensional configuration **r**.
Q maps r onto an M-dimensional space **z**.
M << 3N

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

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

Spiwok et al. Biotechn. Adv. 2015

Biased sampling

potential

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

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

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

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

distribution

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

mulitply 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(**r**)

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 Δ*V* 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.

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

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

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

$$P_b(\mathbf{z}) = \frac{1}{\langle \exp(\beta F(\mathbf{z}) \rangle} = \frac{1}{\int d\mathbf{z} \exp(-\beta F) \exp(\beta F)}$$

$$= \frac{1}{\int d\mathbf{z}} \qquad \langle \exp(\beta F(\mathbf{z}) \rangle = \int d\mathbf{z} P(\mathbf{z}) \exp(\beta F)$$

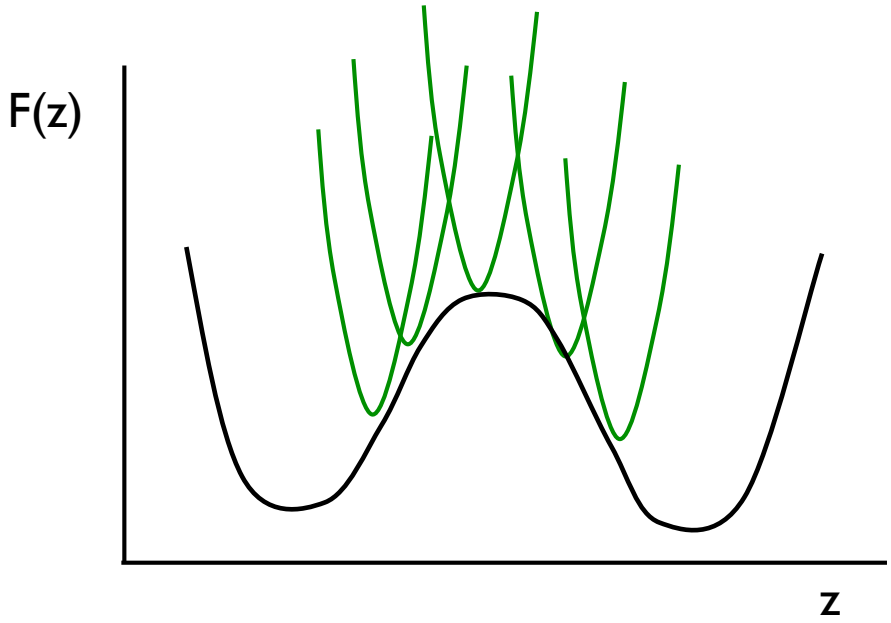
All states z become equiprobable

Umbrella sampling

Harmonic spring keeps the trajectory close to z_i

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

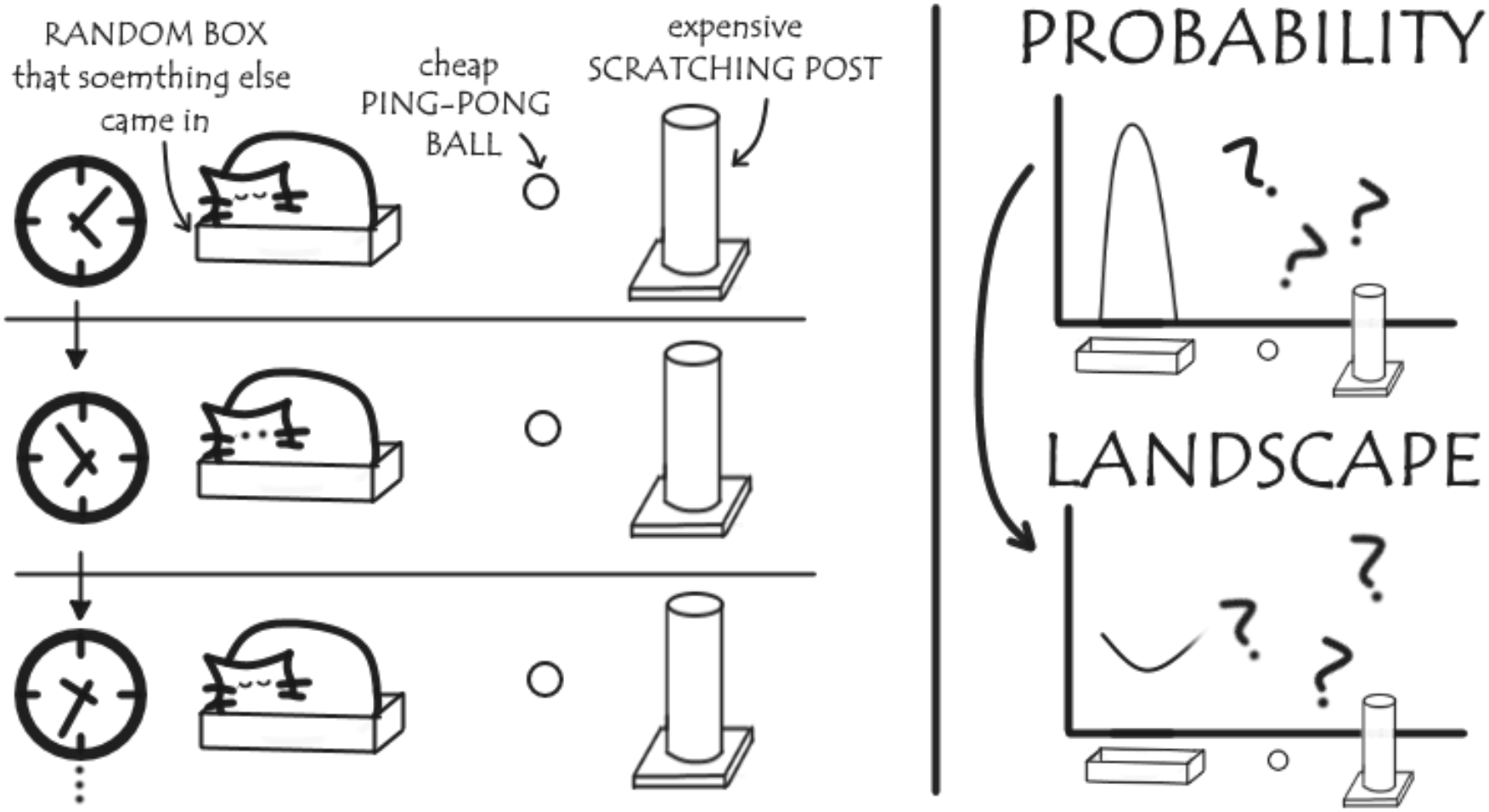
The points z_i and the value of κ must be chosen such that $Q(\mathbf{r}(t))$ makes excursions into the window of its neighbor.



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

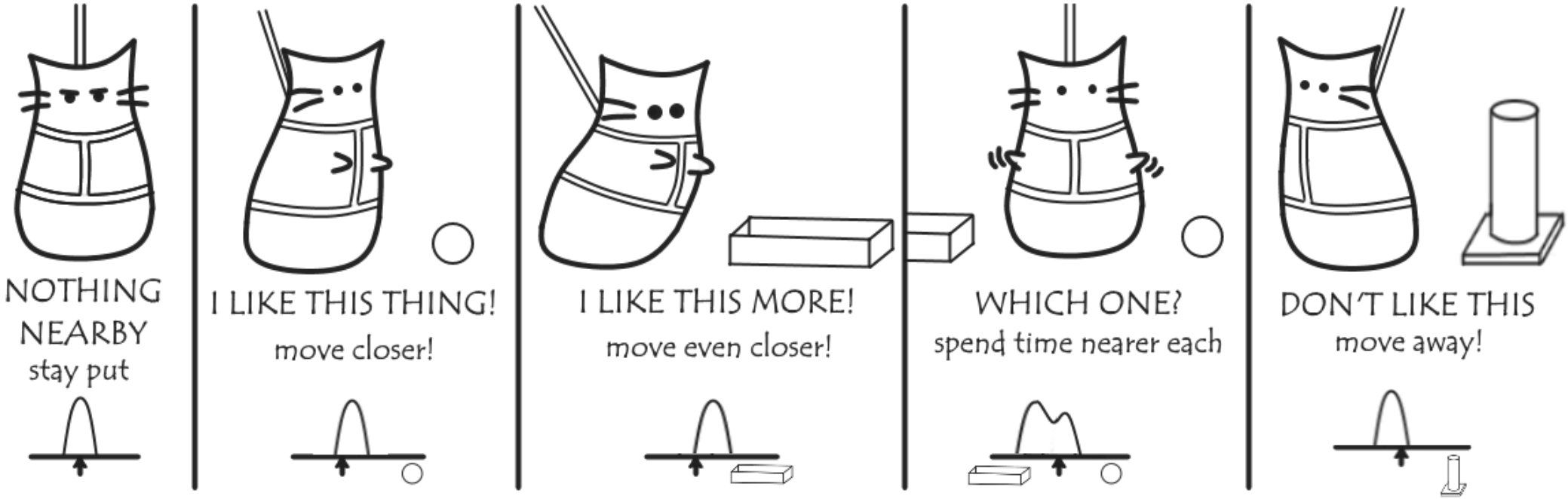
$$P_i(\mathbf{z}) = P_{b,i}(\mathbf{z}) \exp(\frac{1}{2}\beta\kappa|\mathbf{z} - \mathbf{z_i}|^2) \left\langle \exp(-\beta\frac{1}{2}\kappa|Q(\mathbf{r}) - \mathbf{z_i}|^2) \right\rangle$$

Molecular dynamics



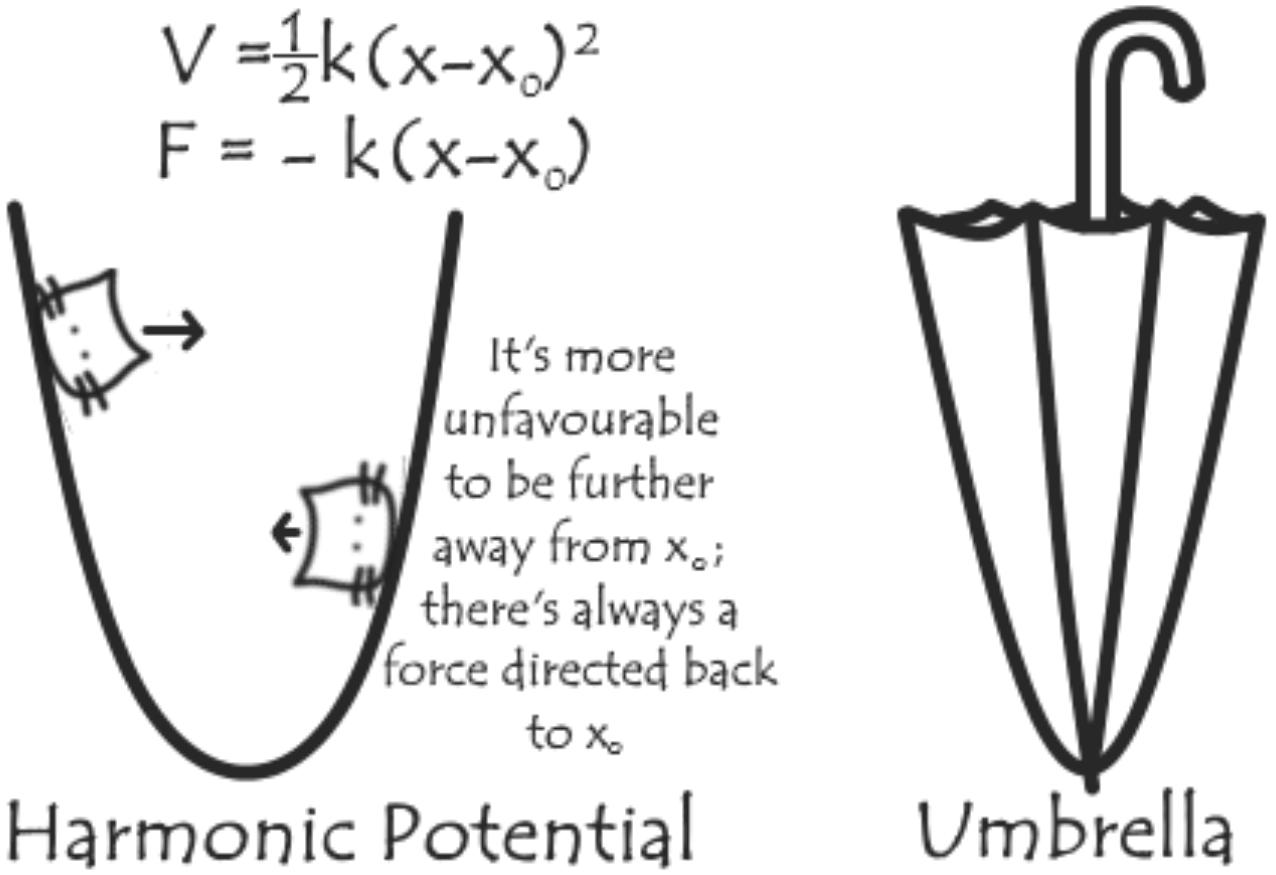
http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

Umbrella sampling



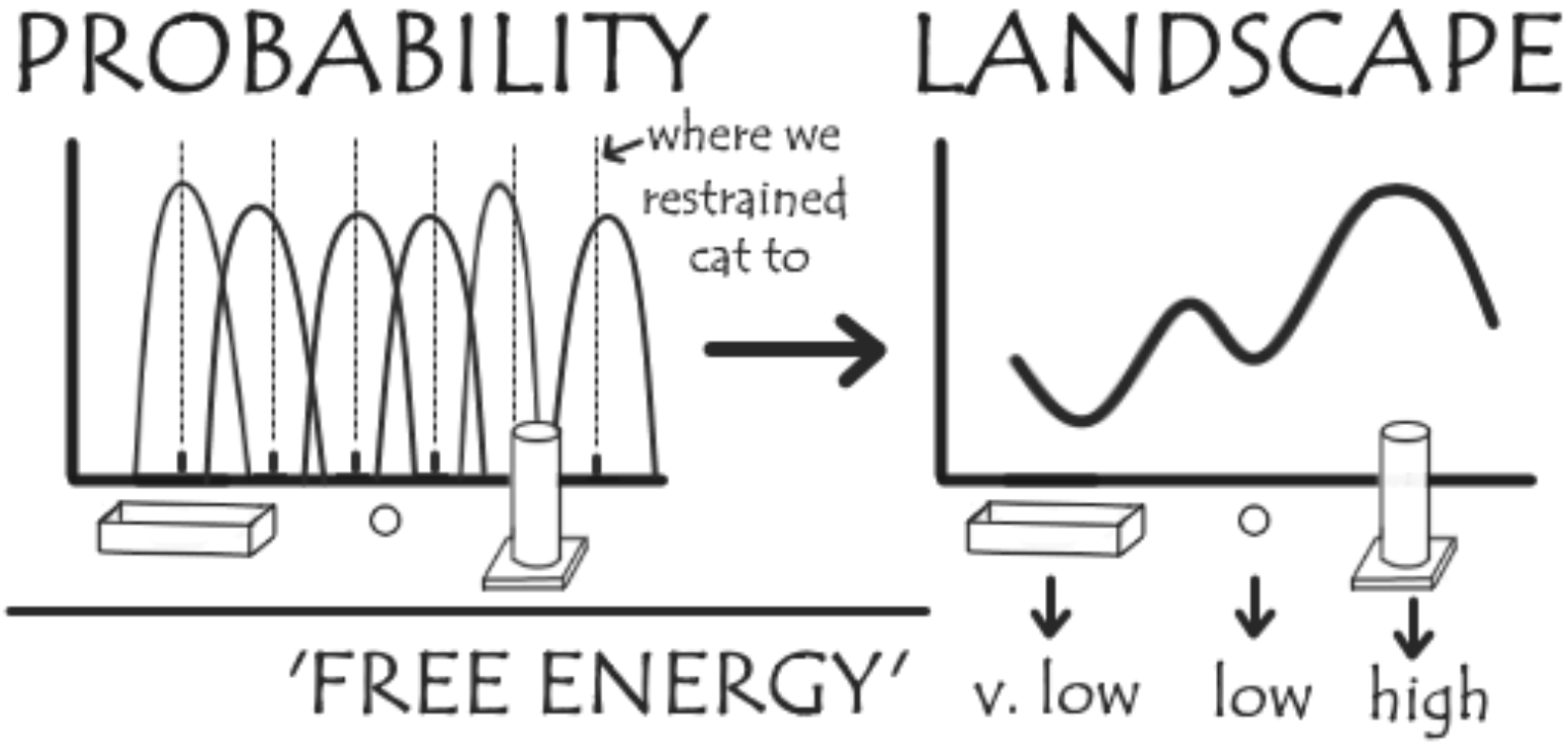
http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

Umbrella sampling



http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

Umbrella sampling

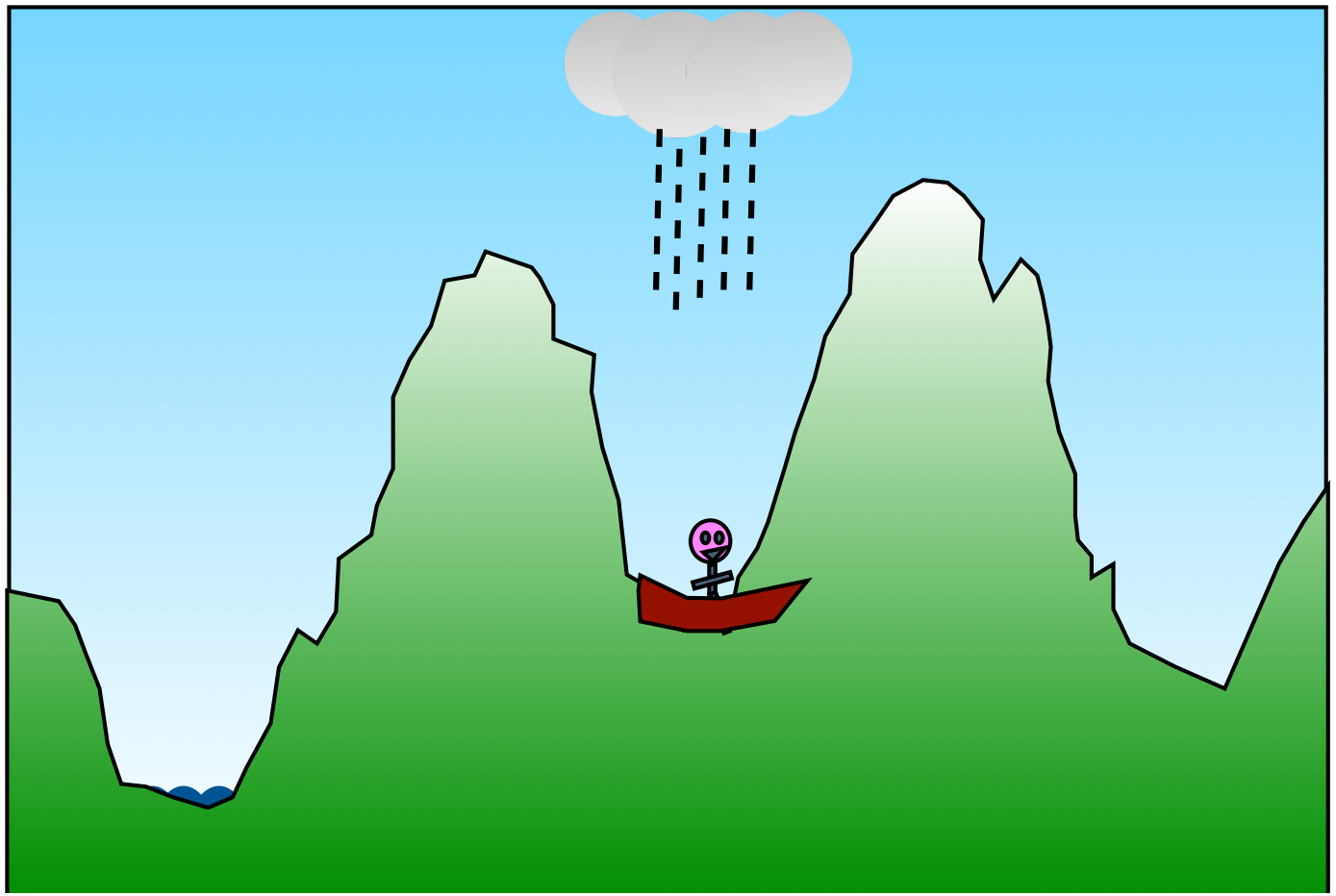


http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

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



Animation by Bernd Ensing
Laio and Parrinello, PNAS (2002)

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.

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

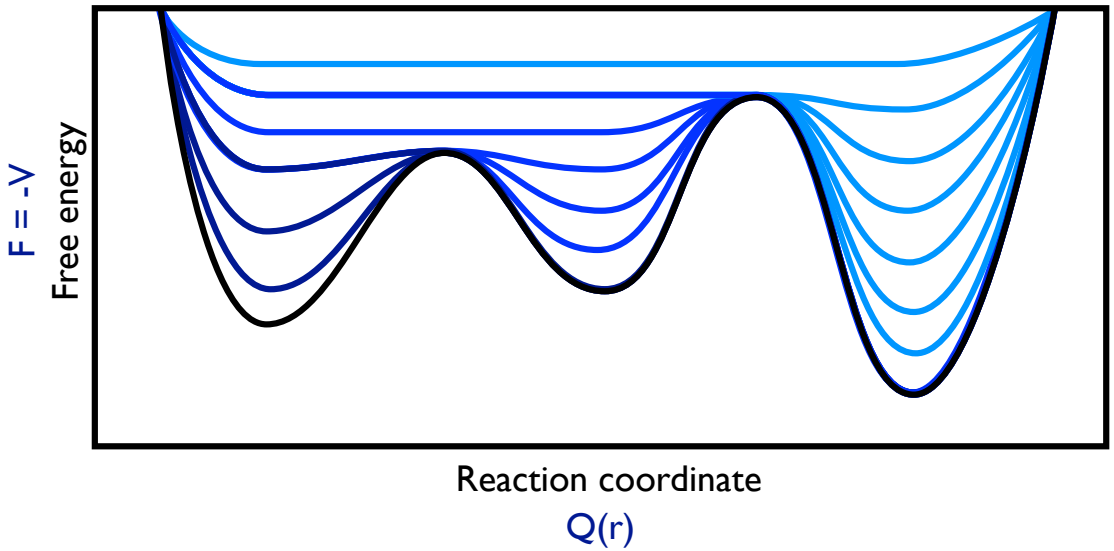
$$t' = \tau_G, 2\tau_G, ...$$

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.

Laio & Parrinello *Proc. Natl. Acad. Sci USA* 2002

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.

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

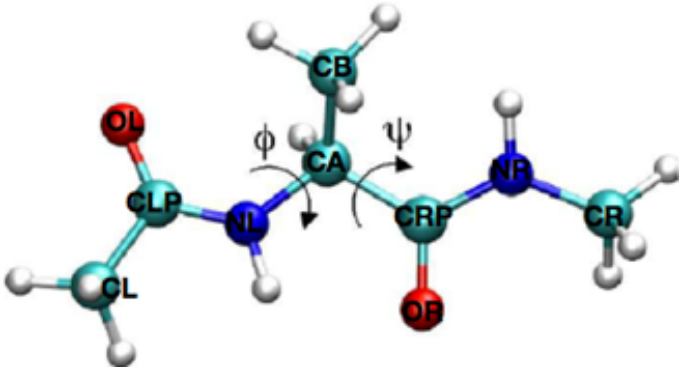
$$t' = \tau_G, 2\tau_G, \ldots$$

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

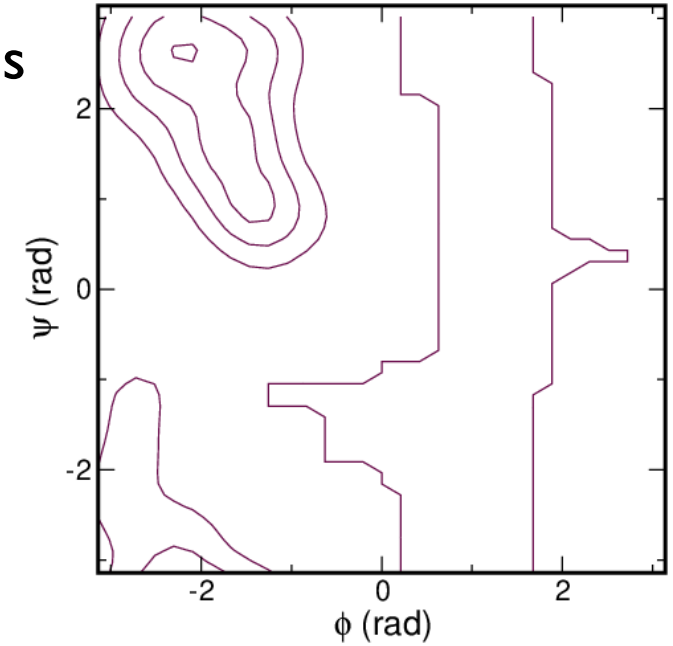
w height of the Gaussian
τ_G time interval between depositions
δQ Gaussian width

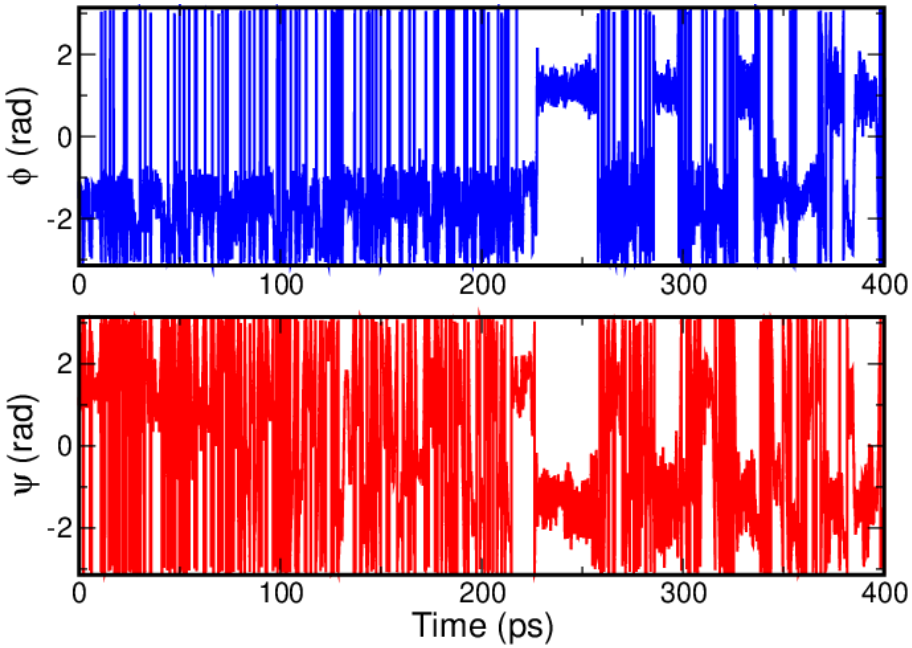
Laio & Parrinello *Proc. Natl. Acad. Sci USA* 2002

Alanine dipeptide in vacuum

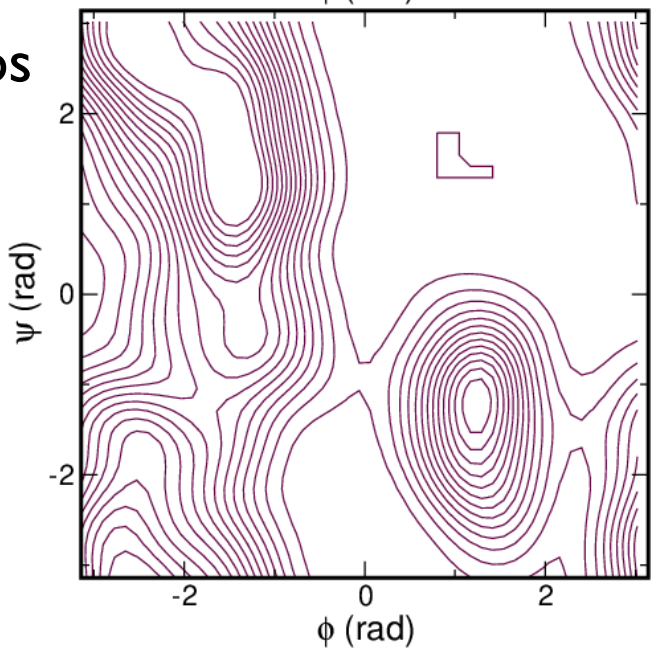


40ps

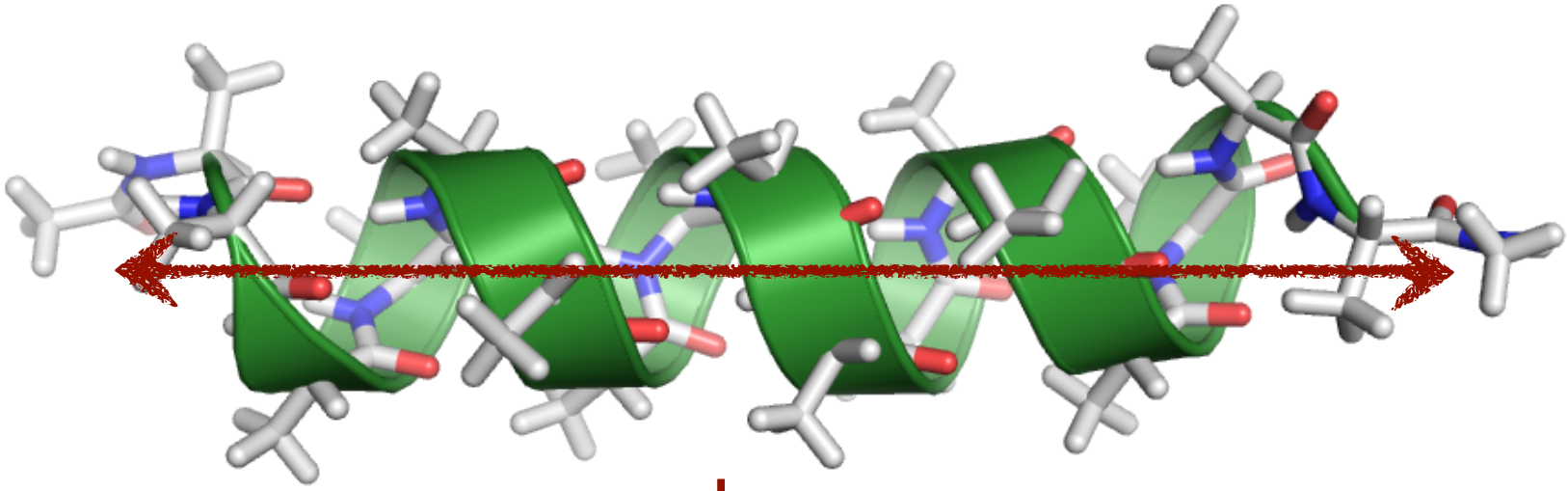




400ps



Setup of a metadynamics simulation

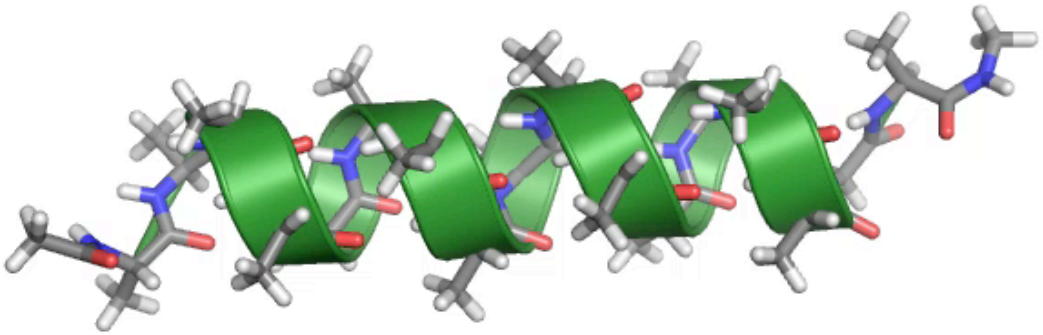


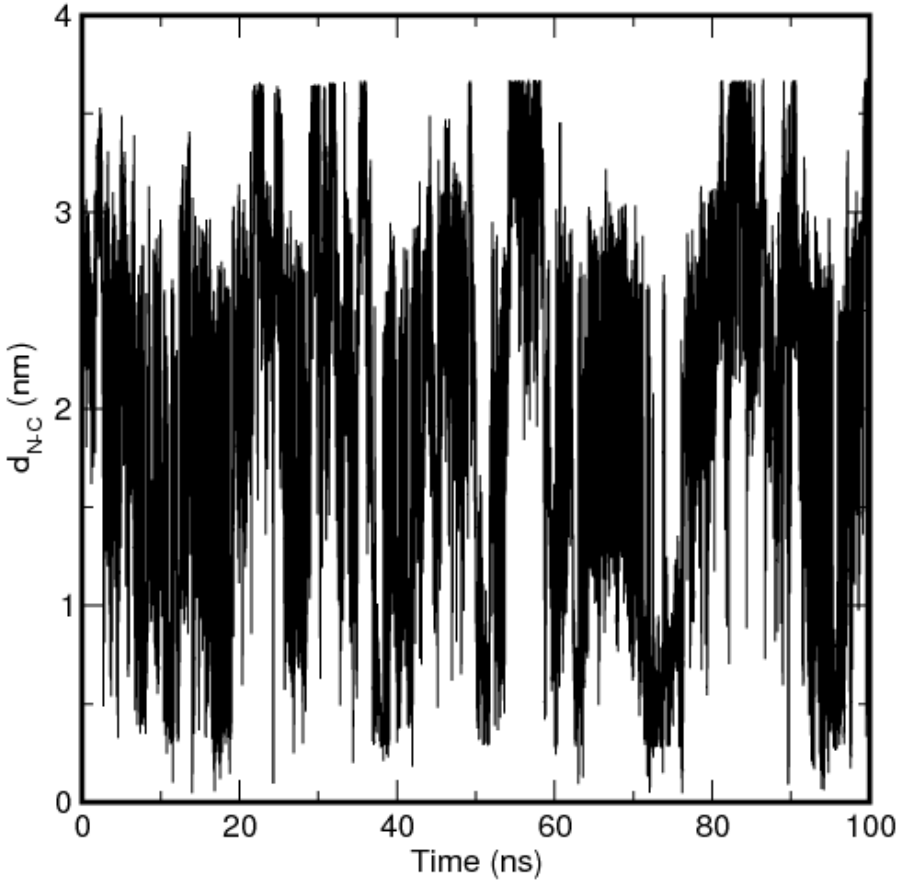
$d_{N\text{-}C}$

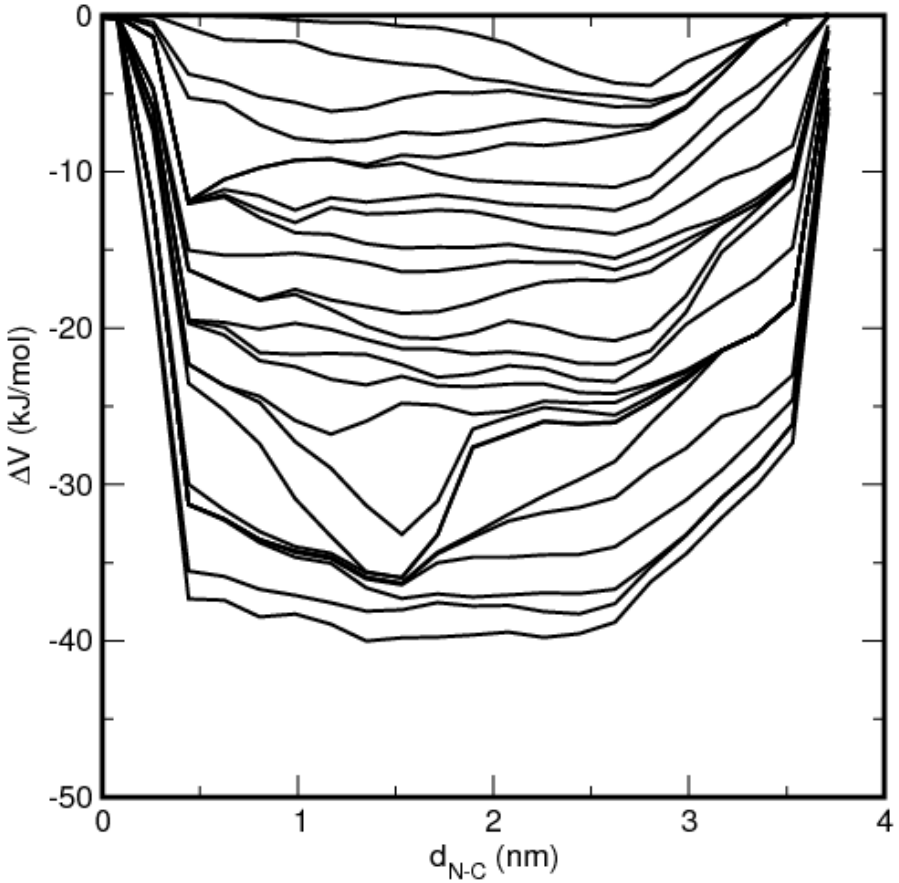
$$\Delta V[Q(\mathbf{r}),t] = w \sum_{t'<t} \exp\left(-\frac{|Q[\mathbf{r}(t)-\mathbf{r}(t')]|^2}{2\delta Q^2}\right)$$

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

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

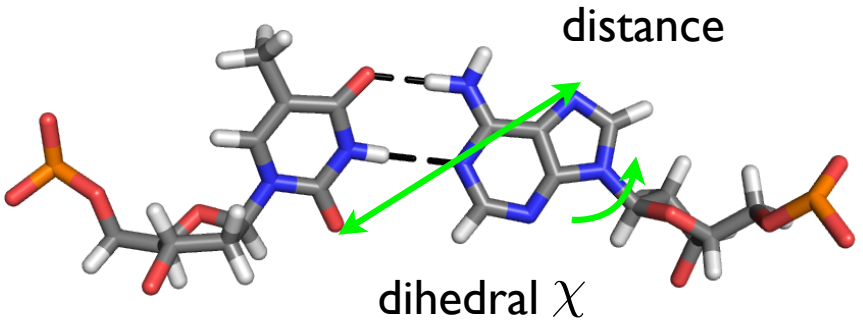



4
3
2
1
0
0
20
40
60
80
100
Time (ns)
d_{N-C} (nm)

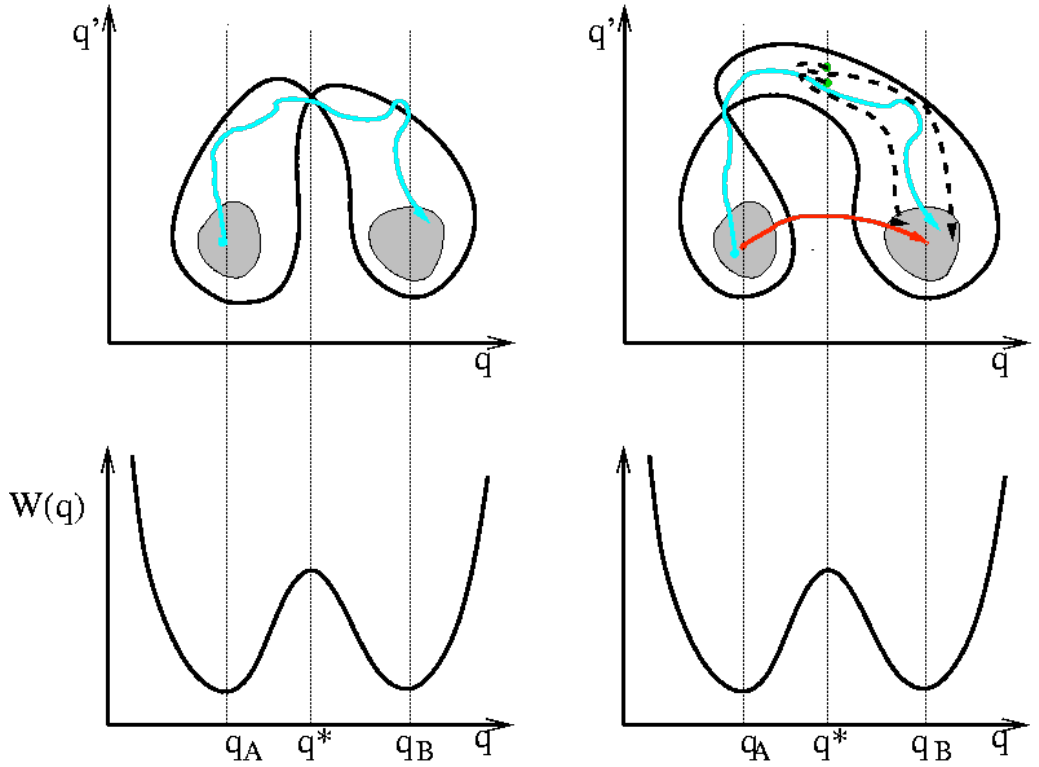

0
-10
-20
-30
-40
-50
0
1
2
3
4
d_{N-C} (nm)
ΔV (kJ/mol)

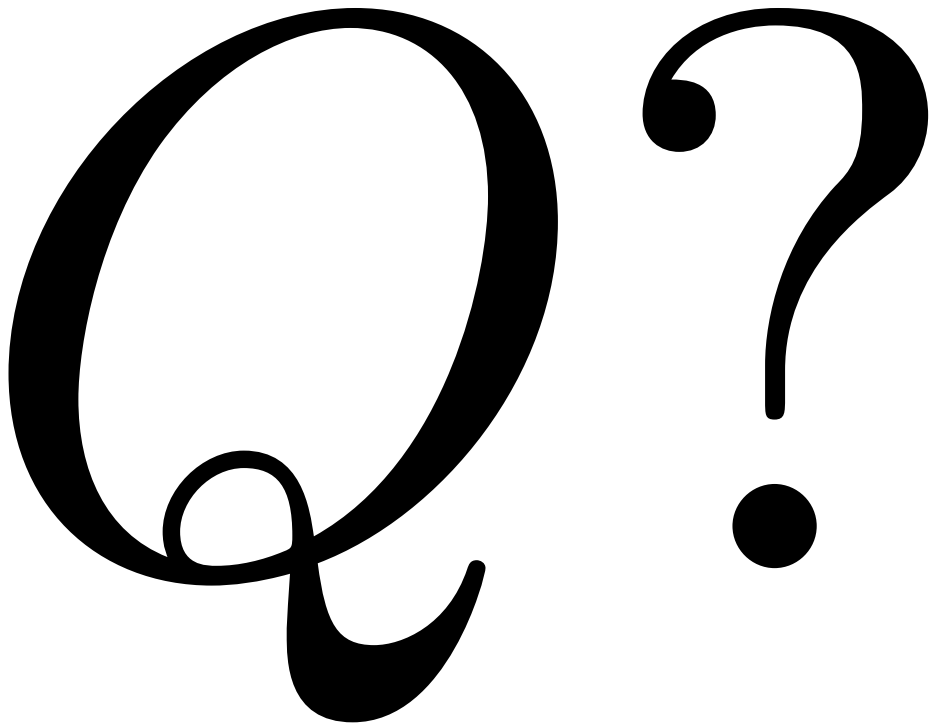
Setup of a metadynamics simulation

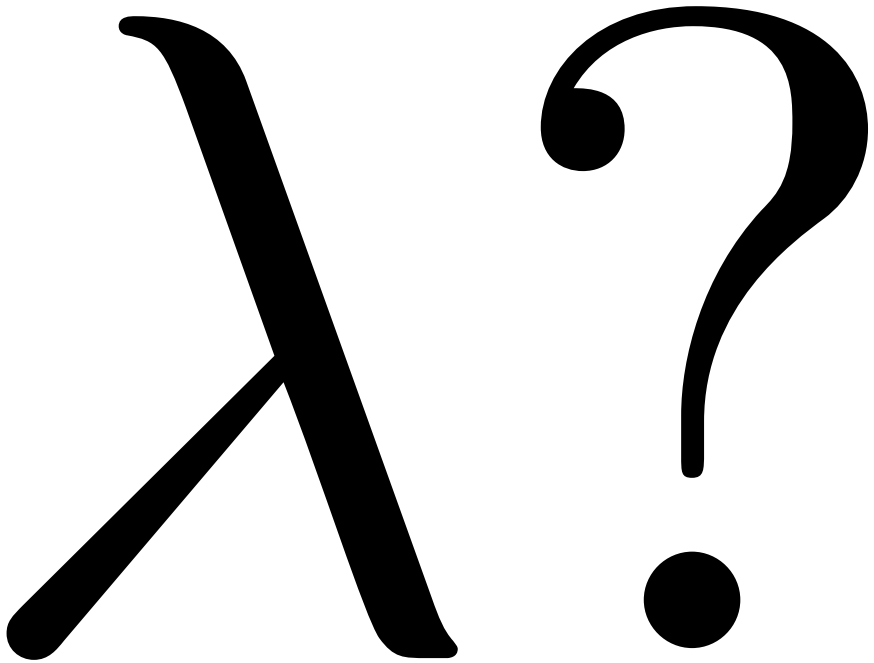
No convergence
Instead, distortion and unfolding
of the DNA fragment
These metadynamics attempts failed...



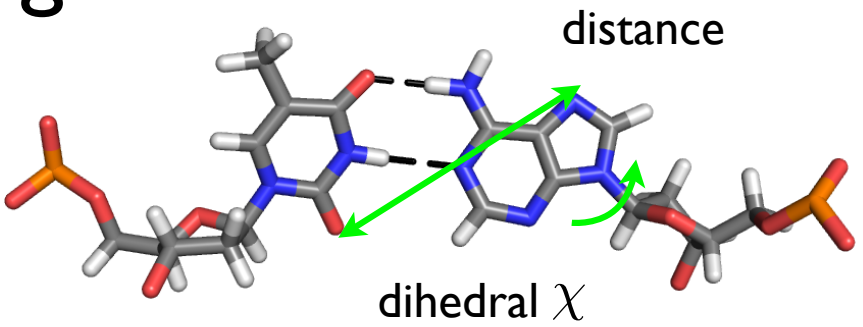
The reaction coordinate problem







Metadynamics can give free energy, if converged



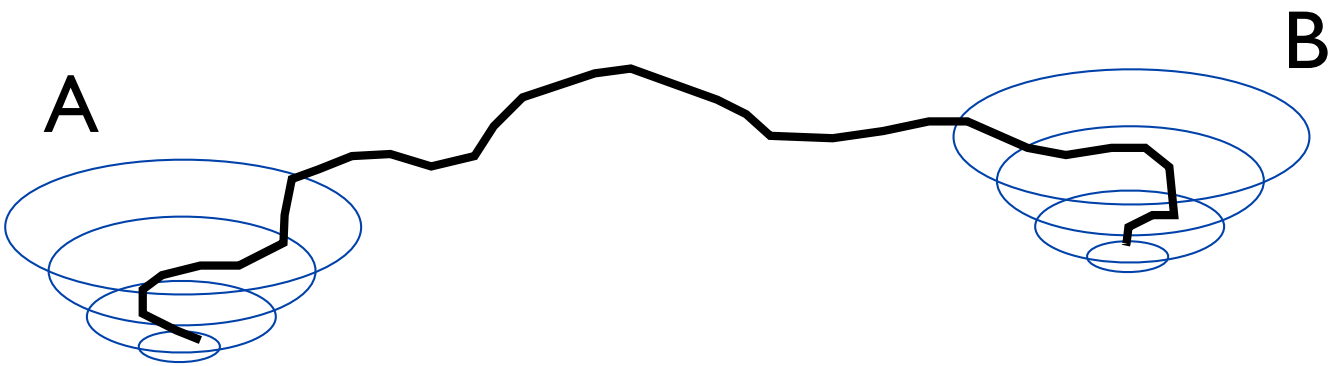
No convergence
Instead, distortion and unfolding
of the DNA fragment
These metadynamics attempts failed...

… 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

Diaz Leines & Ensing *Phys Rev Lett* 2012
Pérez de Alba Ortíz, Vreede & Ensing, Methods in Molecular Biology 2019

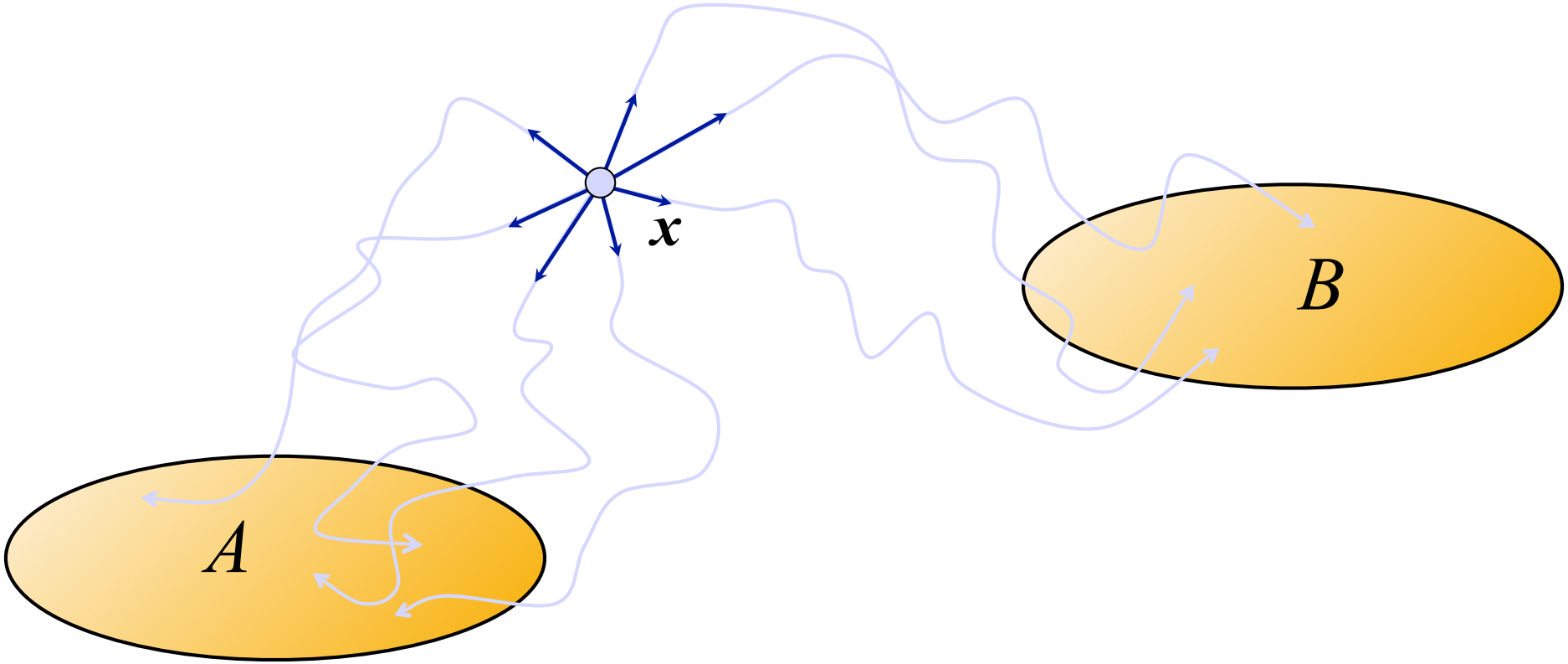
Two state system in a CV space



σ - path in CV space

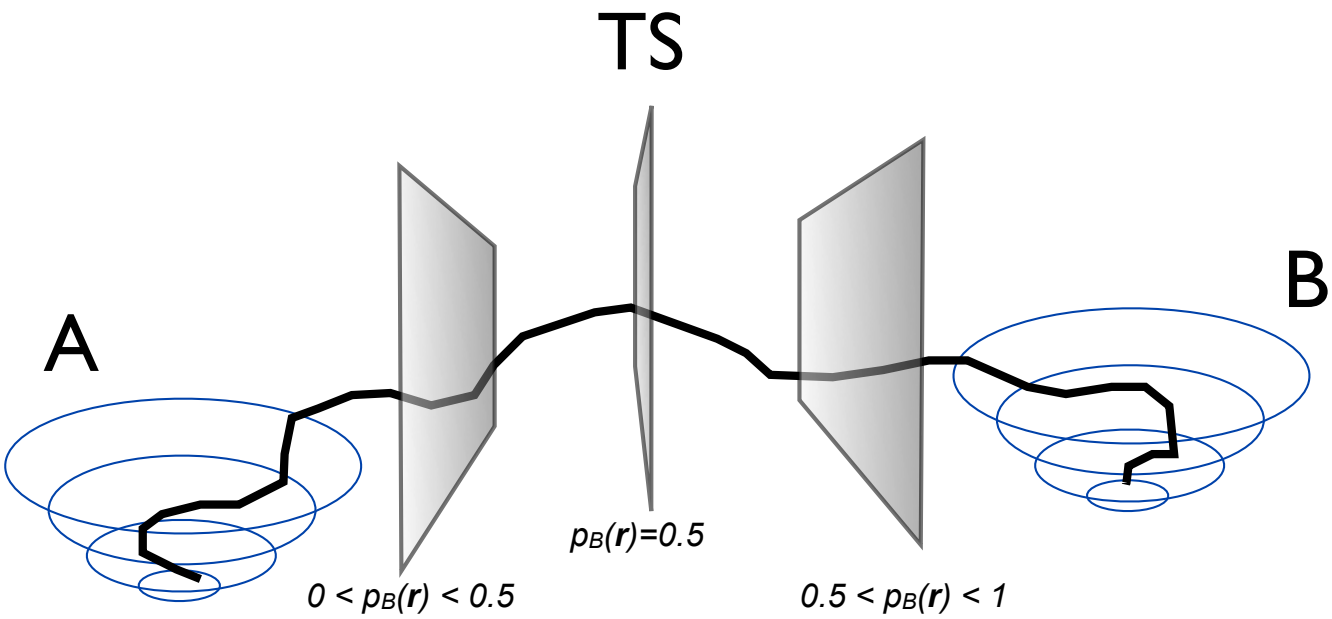
Diaz Leines & Ensing *Phys Rev Lett* 2012

Committor probability p_B



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

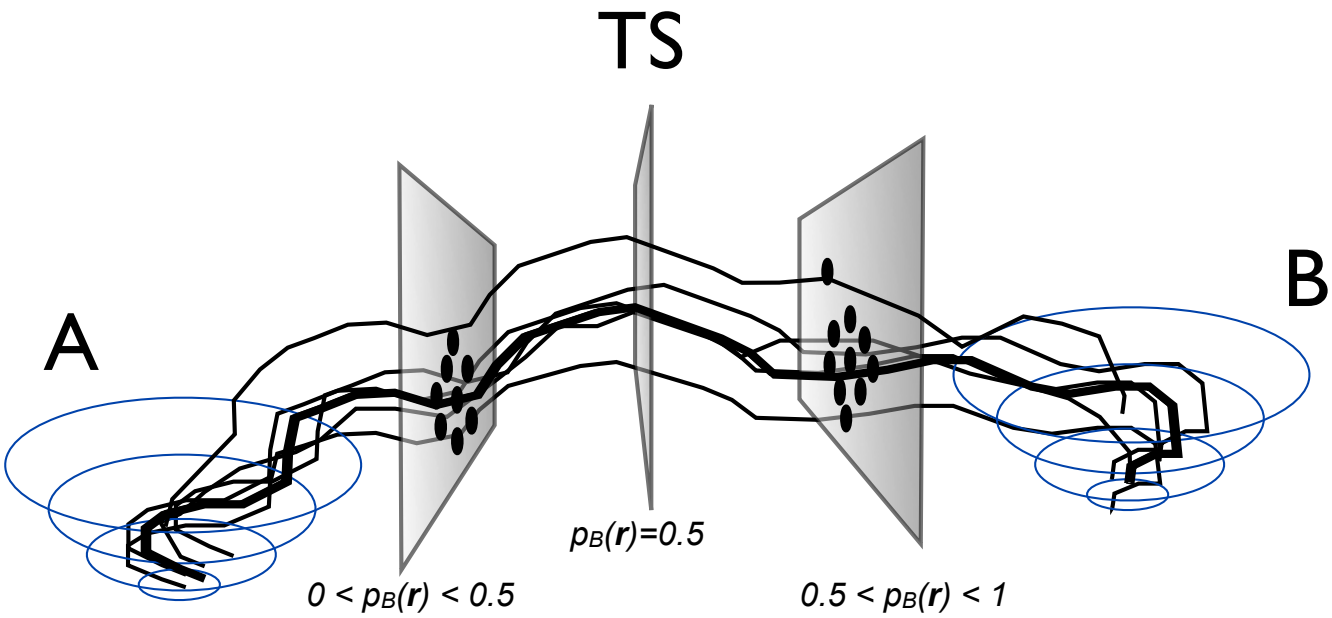
Iso-committor surfaces



σ - path in CV space

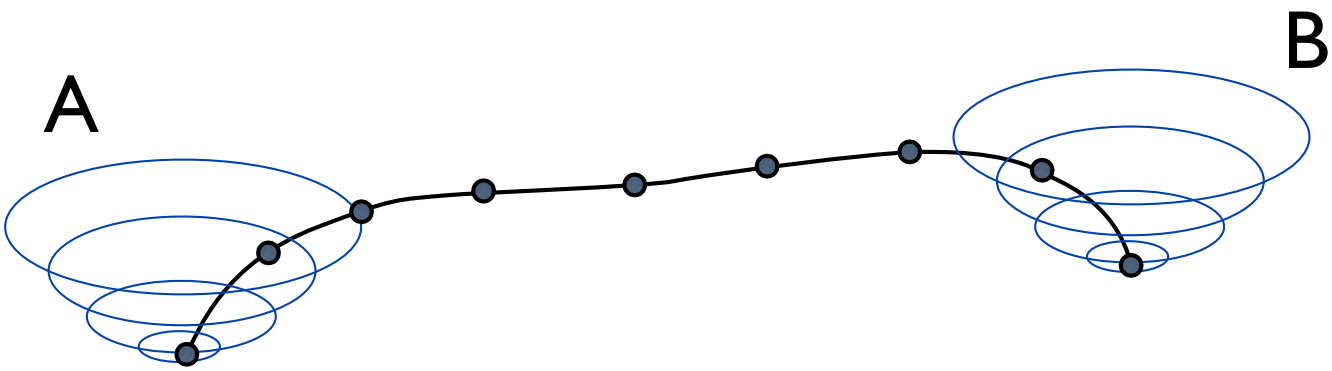
Diaz Leines & Ensing *Phys Rev Lett* 2012

The average transition pathway



Diaz Leines & Ensing *Phys Rev Lett* 2012

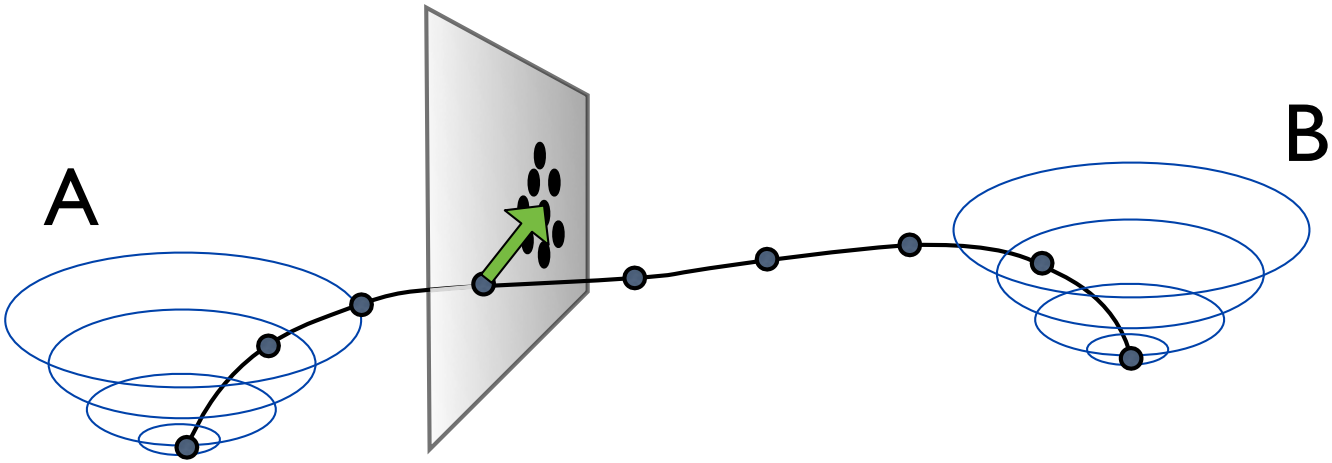
Path metadynamics (i)



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

Diaz Leines & Ensing *Phys Rev Lett* 2012

Path-metadynamics (ii)

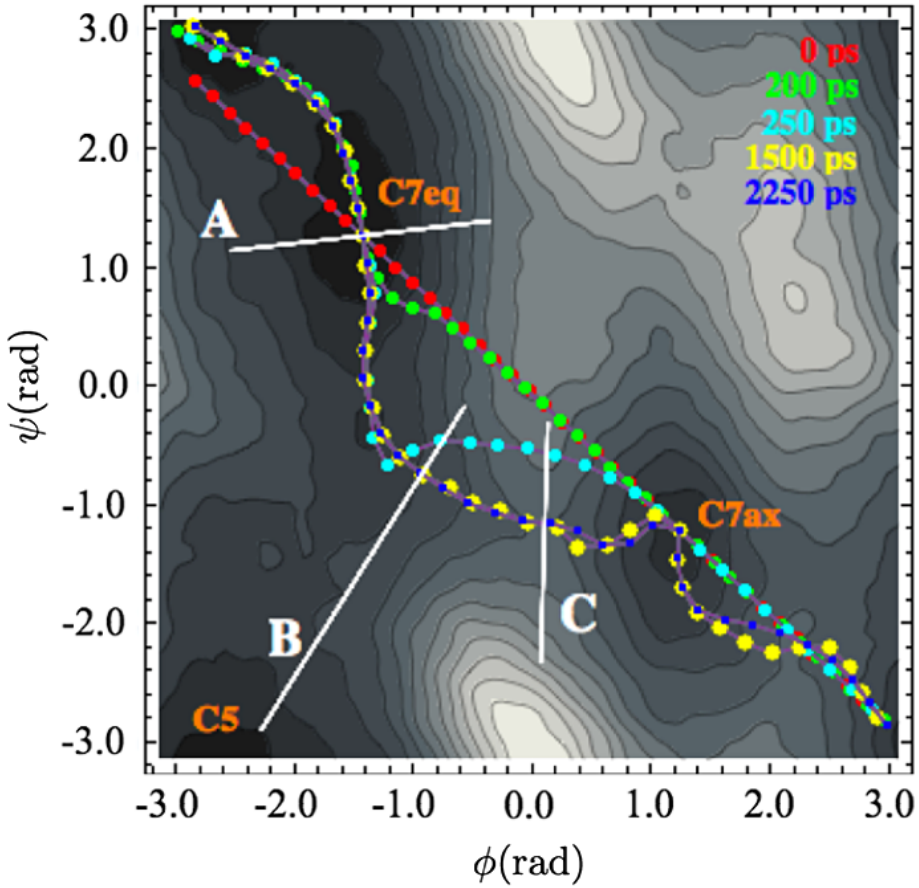


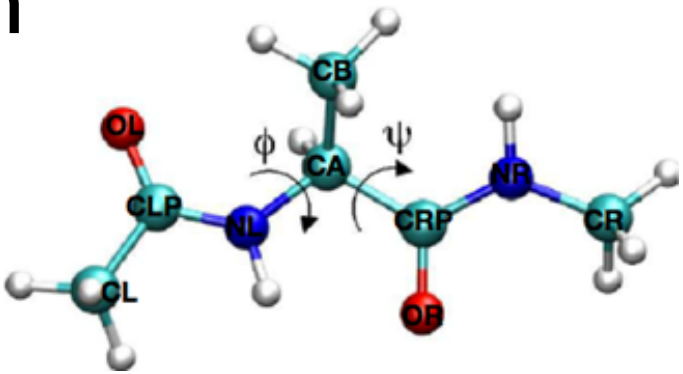
determine distance to mean density
move nodes to mean density

Diaz Leines & Ensing *Phys Rev Lett* 2012

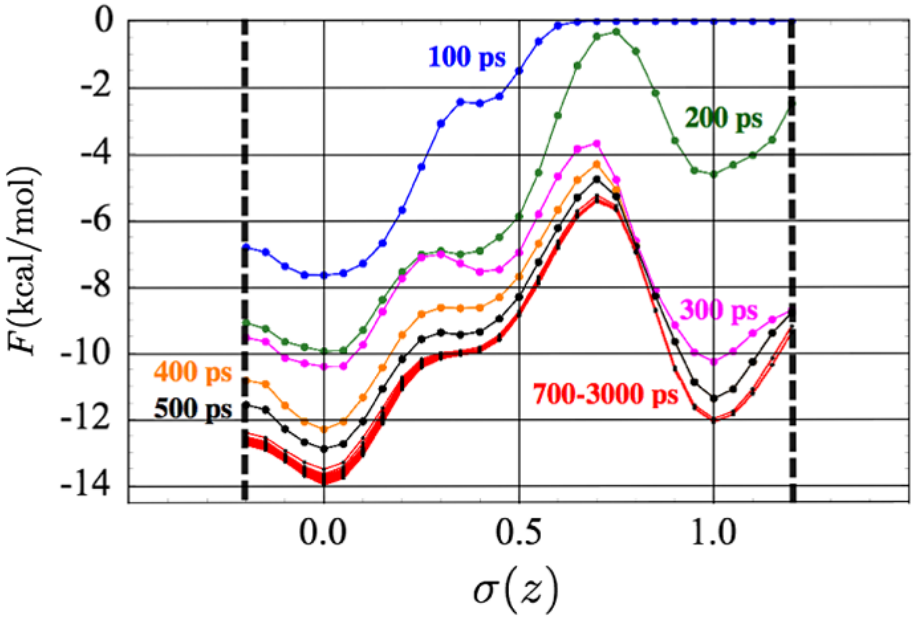
Path-metadynamics of alanine dipeptide in vacuum

path in cv space



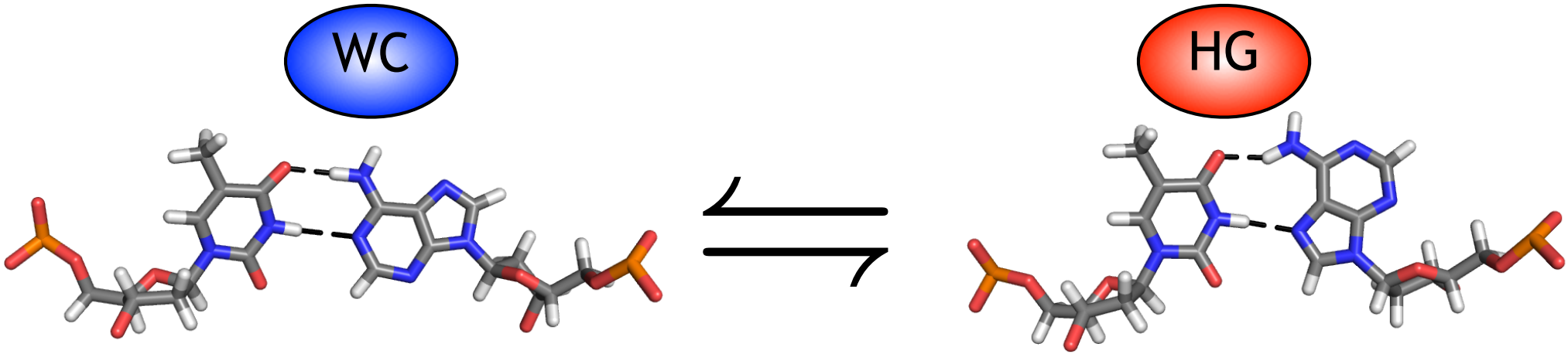


free energy along the path



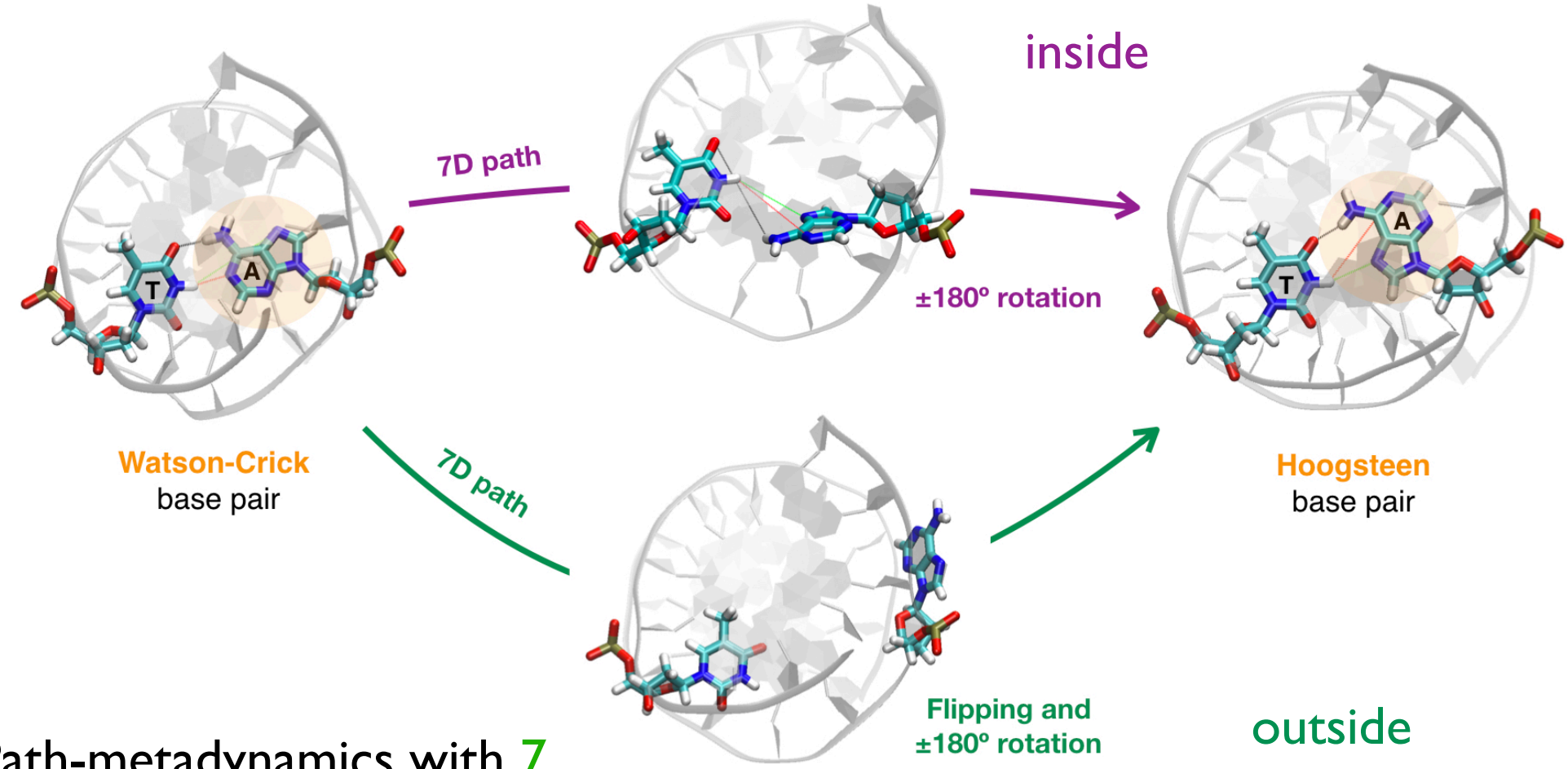
Diaz Leines & Ensing *Phys Rev Lett* 2012

Back to DNA baserolling



Pérez de Alba Ortíz, Vreede & Ensing, Methods in Molecular Biology 2019

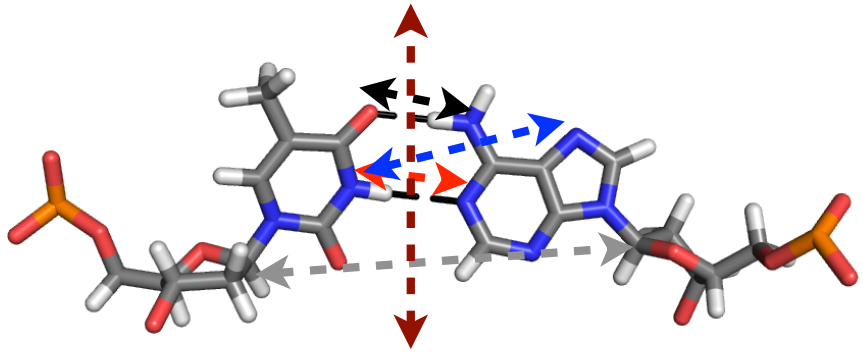
Adaptive path collective variable



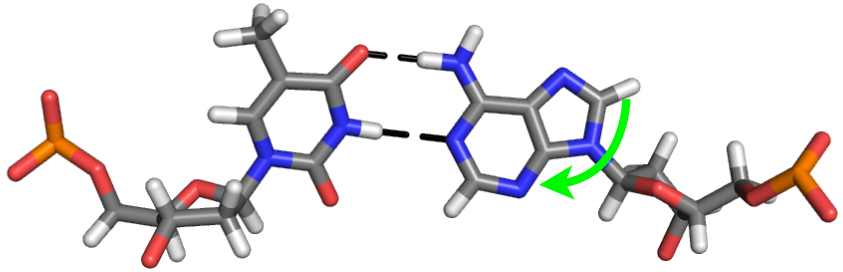
Path-metadynamics with 7 CVs including H-bonding, rolling, flipping, breathing...

Pérez de Alba Ortíz, Vreede & Ensing, Methods in Molecular Biology 2019

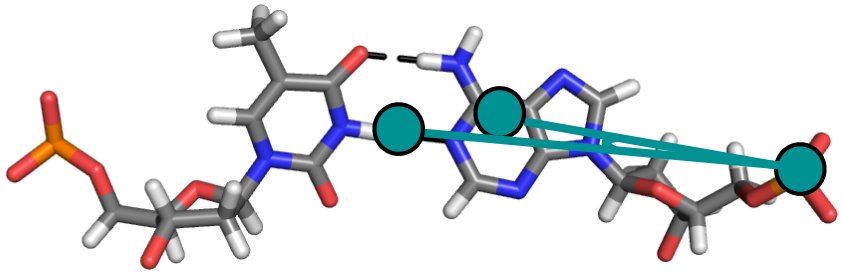
Biasing DNA baserolling defining CVs



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



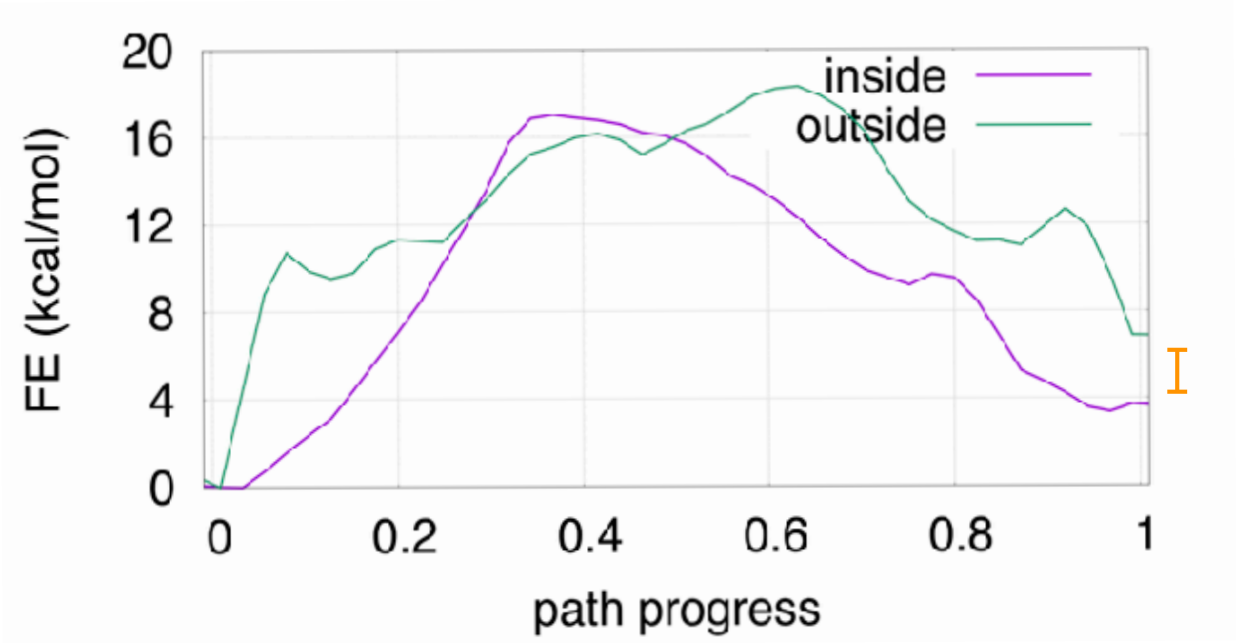
base rolling angle



base flipping angle

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

Free energy profiles

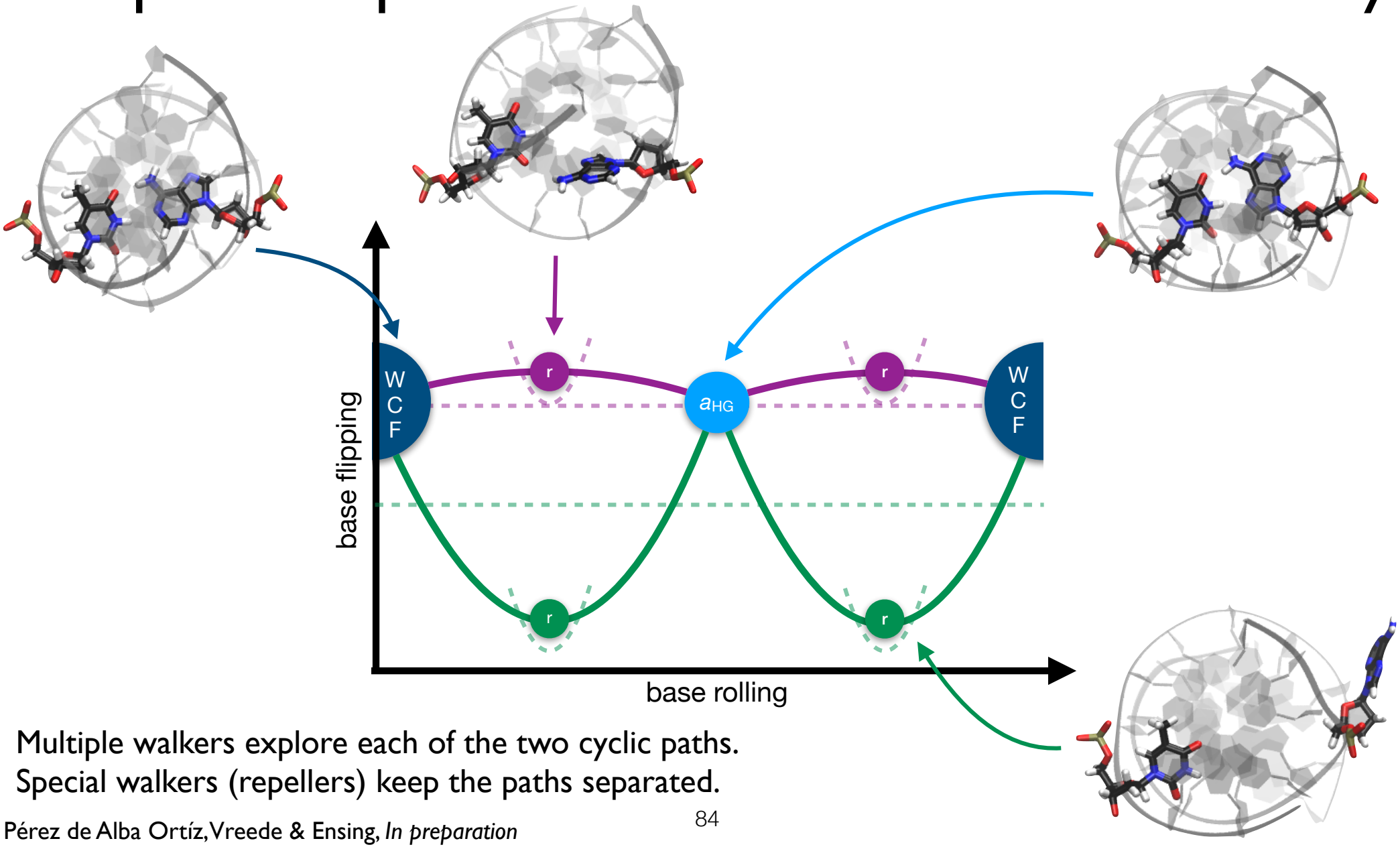


- Path-metadynamics with 7 CVs including H-bonding, rolling, flipping, breathing.
- No clear preference for **inside** or **outside**

- **Lessons in path-metadynamics:**
  - Difficulty in capturing one out of several competing mechanisms
  - Entropic penalty due to the tube potential acting in unnecessary CVs

Pérez de Alba Ortíz, Vreede & Ensing, Methods in Molecular Biology 2019

Multi-path-metadynamics: sample multiple reaction channels simultaneously

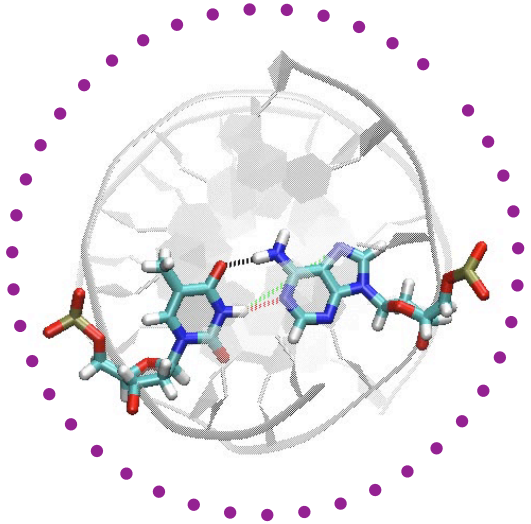


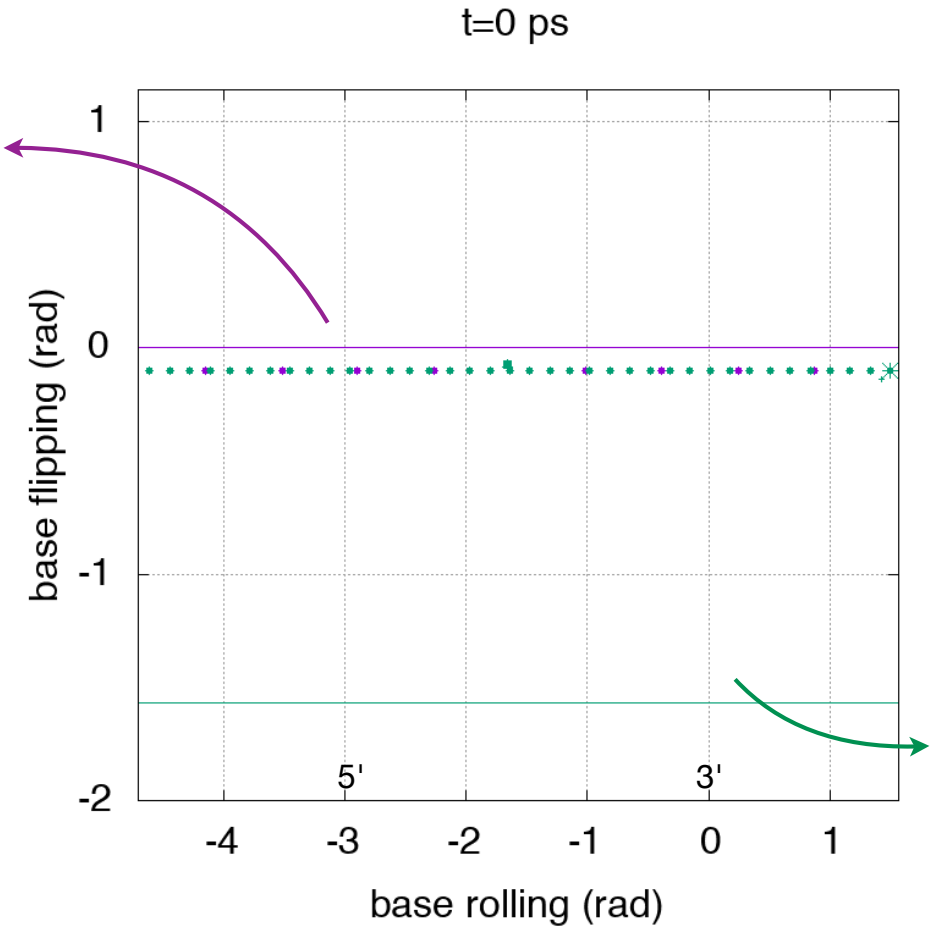
Multiple walkers explore each of the two cyclic paths.
Special walkers (repellers) keep the paths separated.

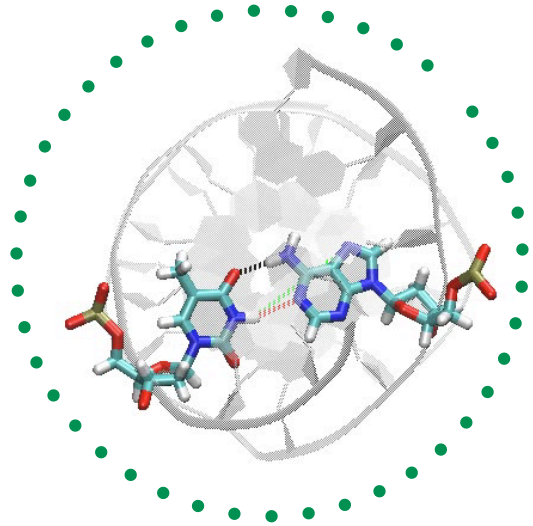
Pérez de Alba Ortíz, Vreede & Ensing, *In preparation*

DNA base rolling with PMD

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

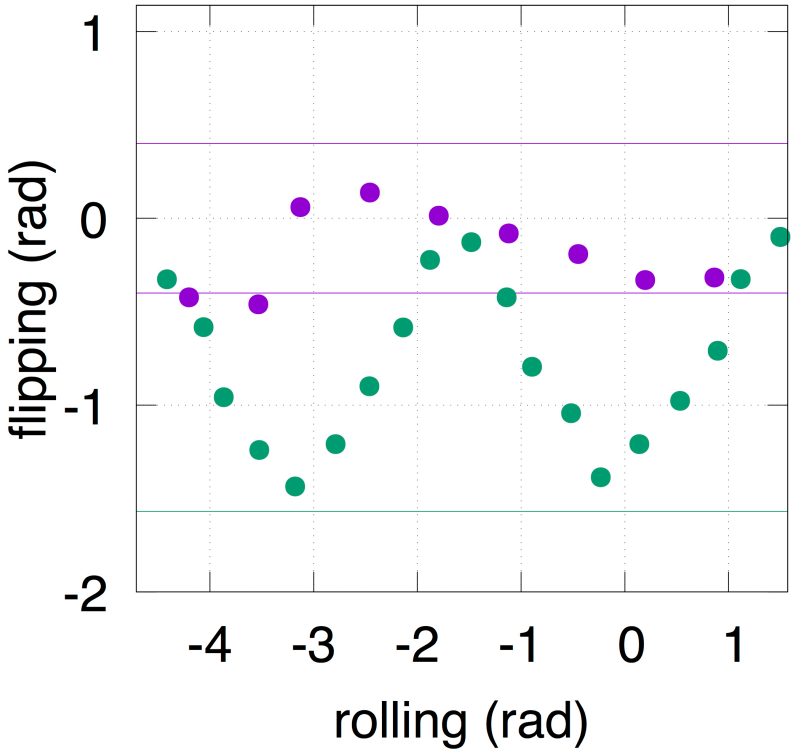


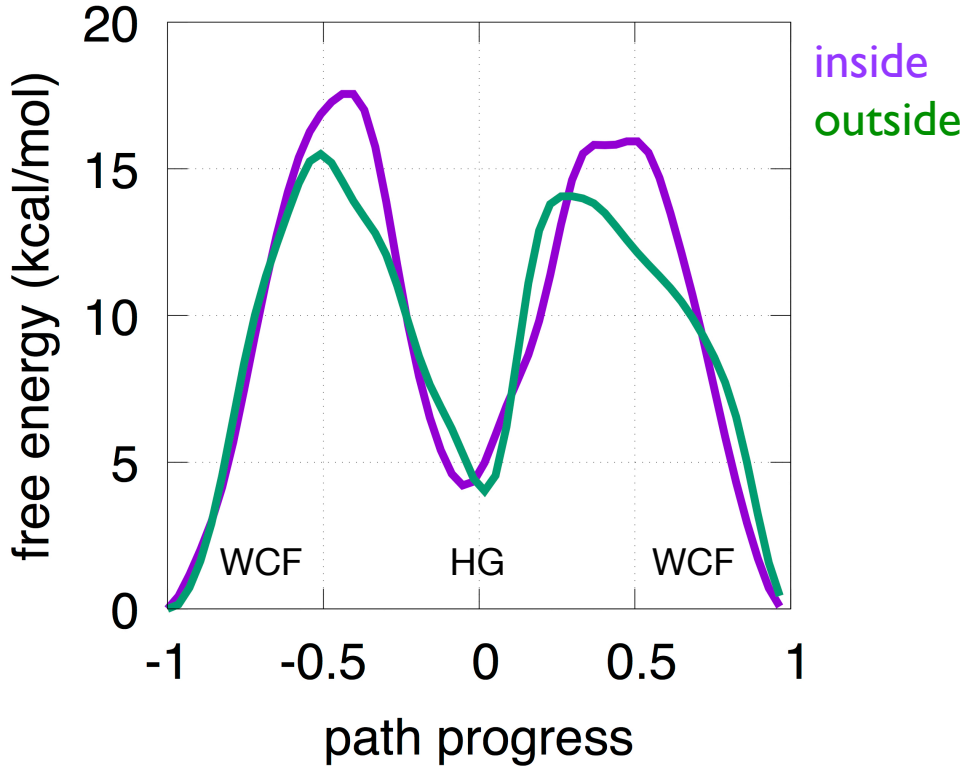




Pérez de Alba Ortíz, Vreede & Ensing, *In preparation*

DNA baserolling with path-metadynamics

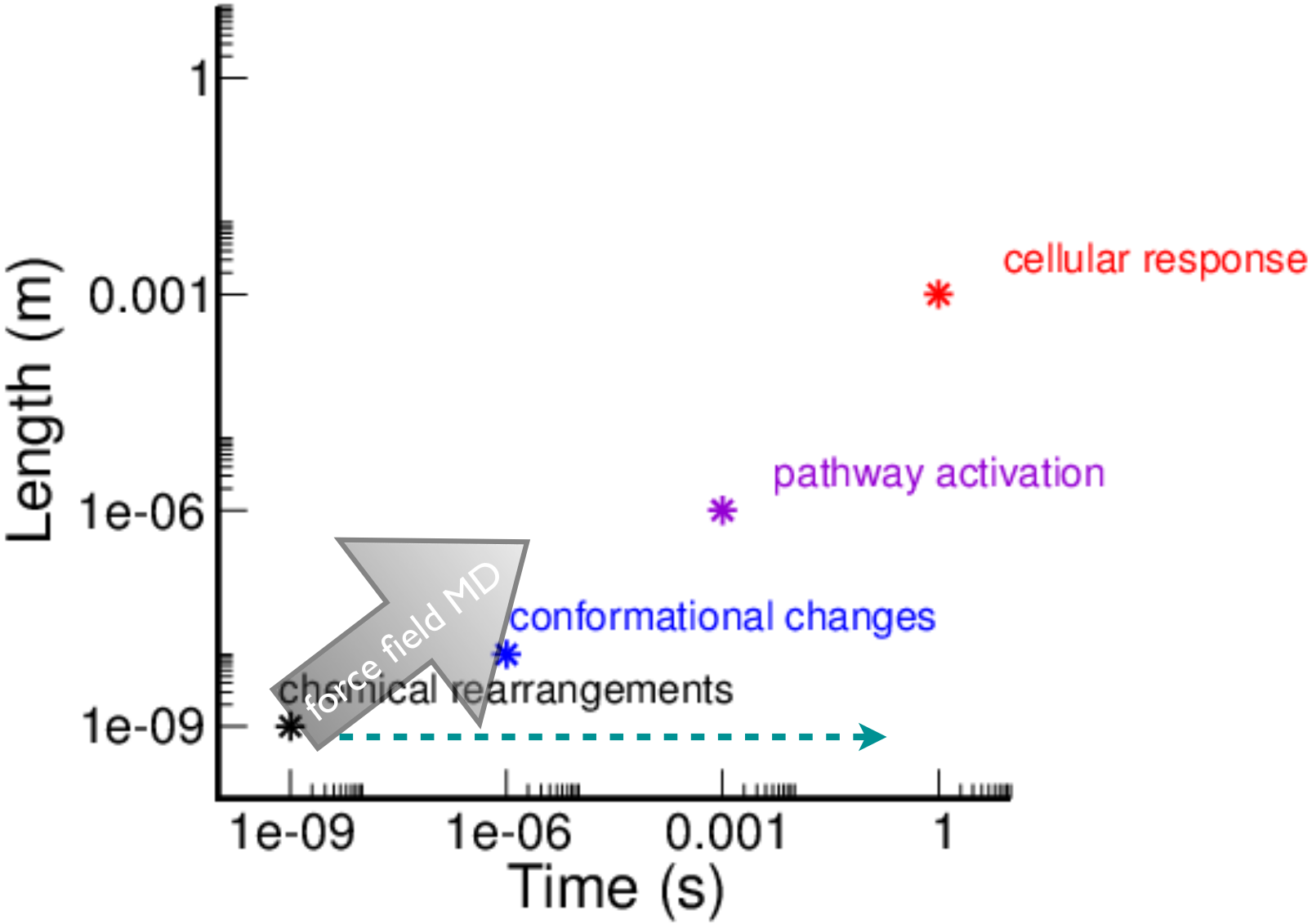




Free energy difference between WCF and HG close to experimental value.
Outside path has lower barriers, similar to TPS results.

Pérez de Alba Ortíz, Vreede & Ensing, Methods in Molecular Biology 2019
Pérez de Alba Ortíz, Vreede & Ensing, in preparation

Bridging the time gap



Biomolecular simulation Outline

Part 1: Simulating biomolecular systems

Part 2: Biased sampling

Part 3: Path sampling

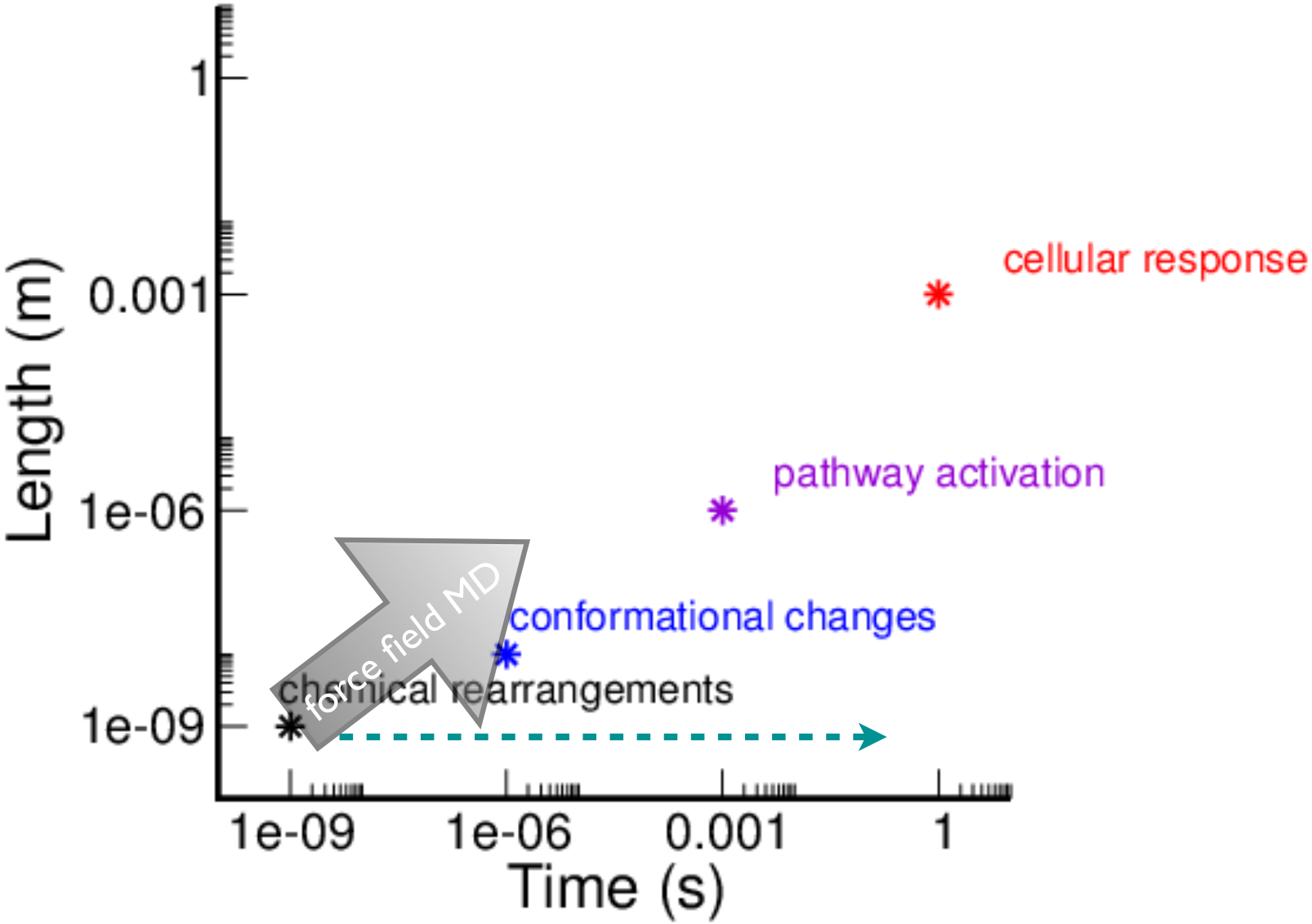
Biomolecular simulation Outline

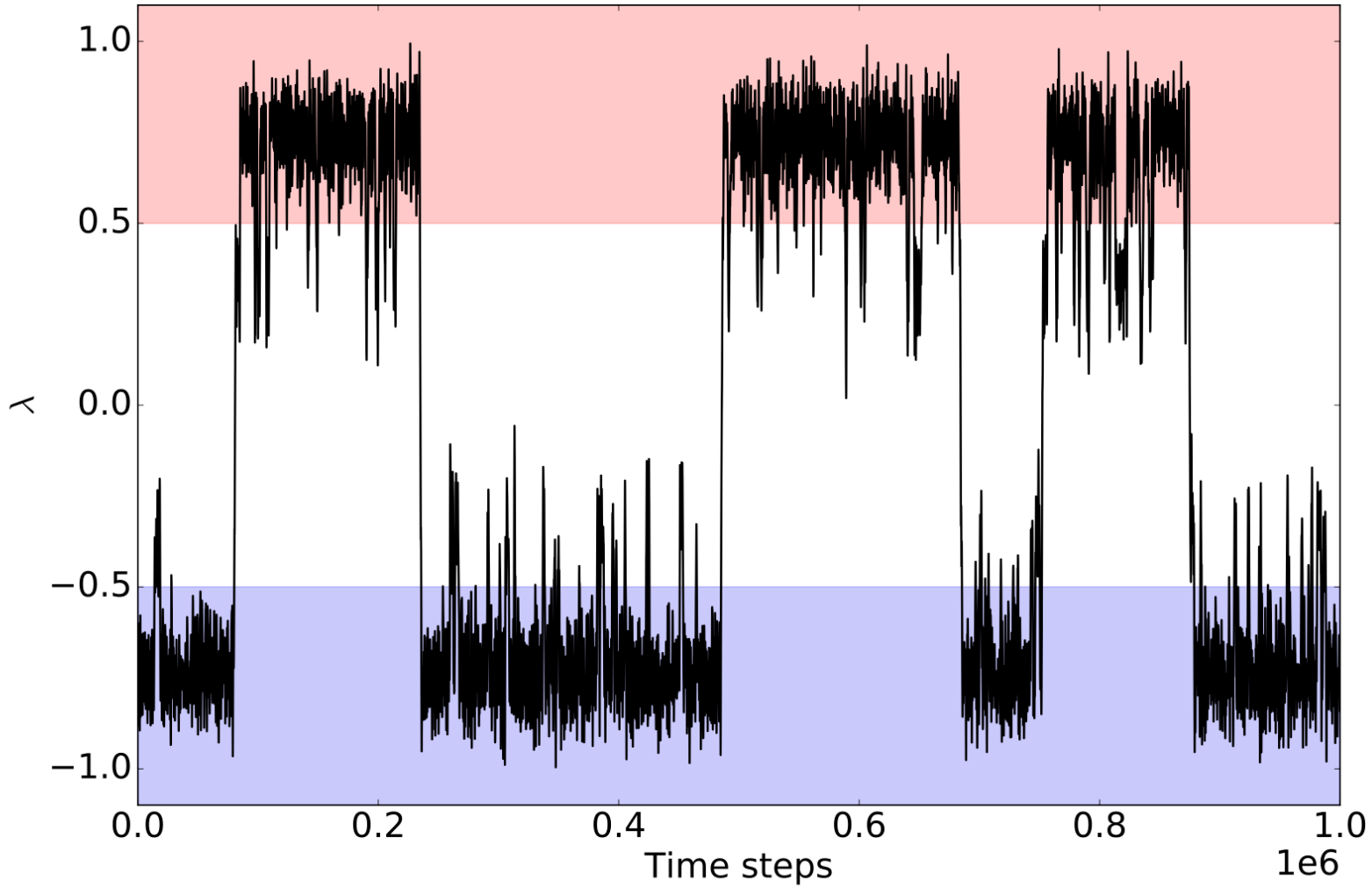
Part 1: Simulating biomolecular systems

Part 2: Biased sampling

Part 3: Path sampling

Bridging the time gap

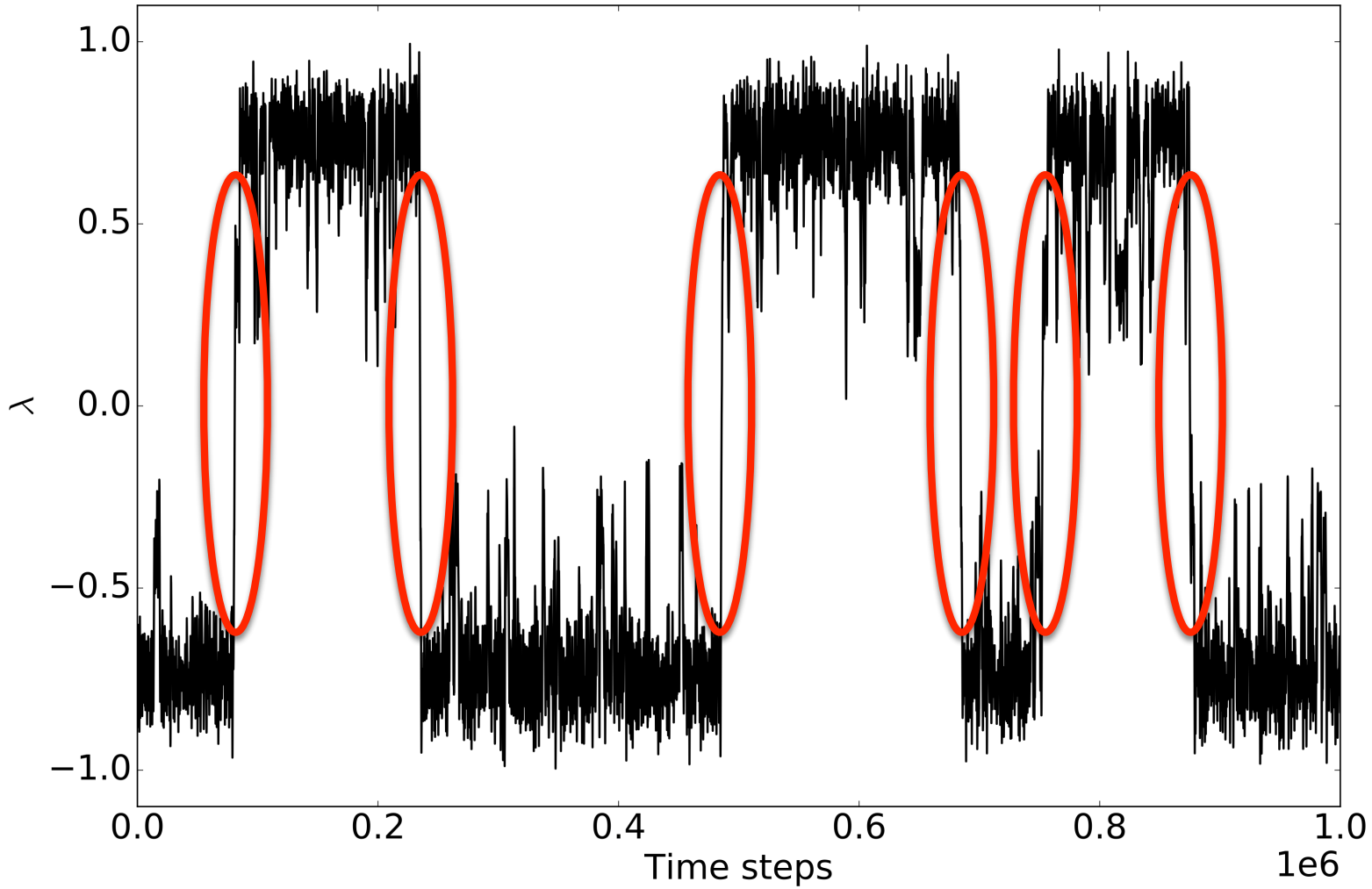




Separation of time scales: fundamental time scale vs. frequency of event

Slide by David Swenson

Path sampling



Focus on the transition regions

Monte Carlo in path space

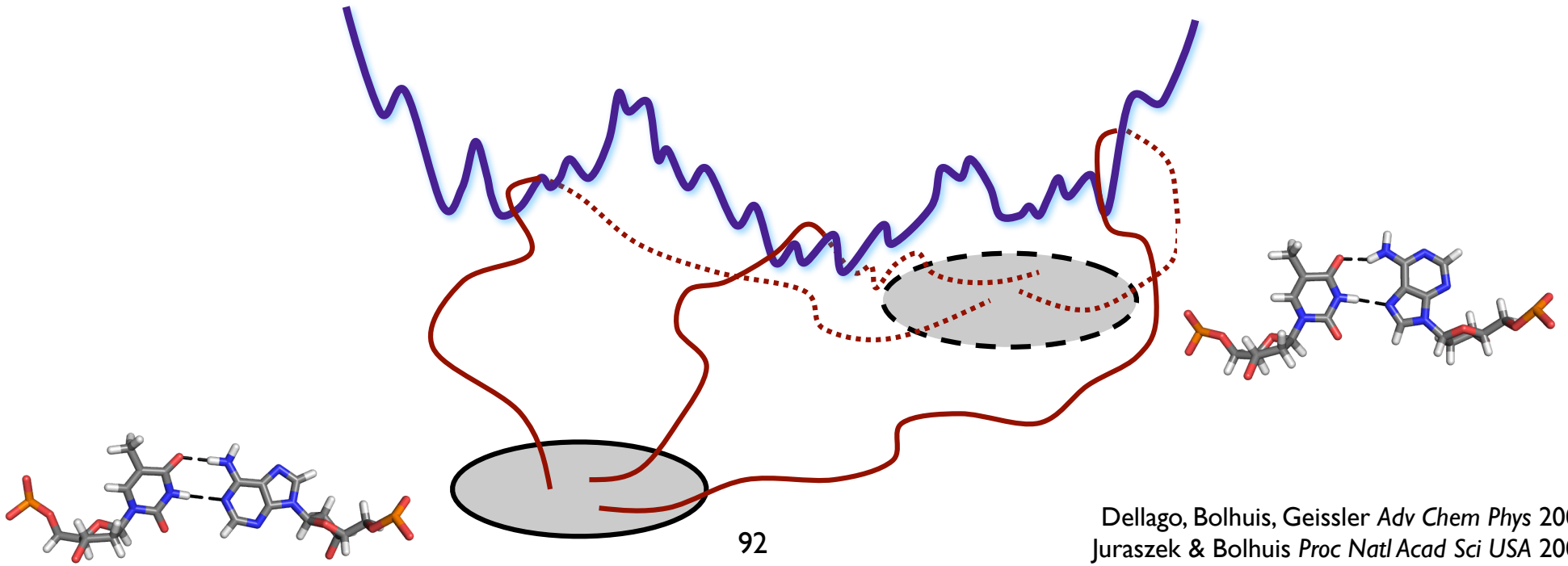
Slide by David Swenson

Sampling rare events: transition path sampling

Input

Result: ensemble of transition paths

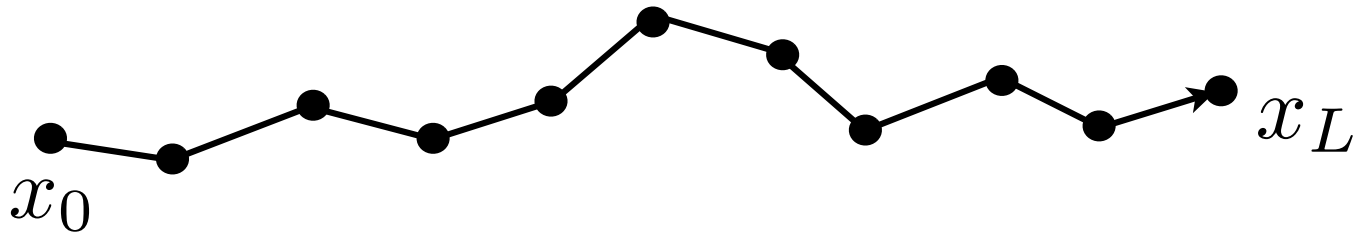
Algorithm



Dellago, Bolhuis, Geissler *Adv Chem Phys* 2002
Juraszek & Bolhuis *Proc Natl Acad Sci USA* 2006

Transition Path

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



Path probability:

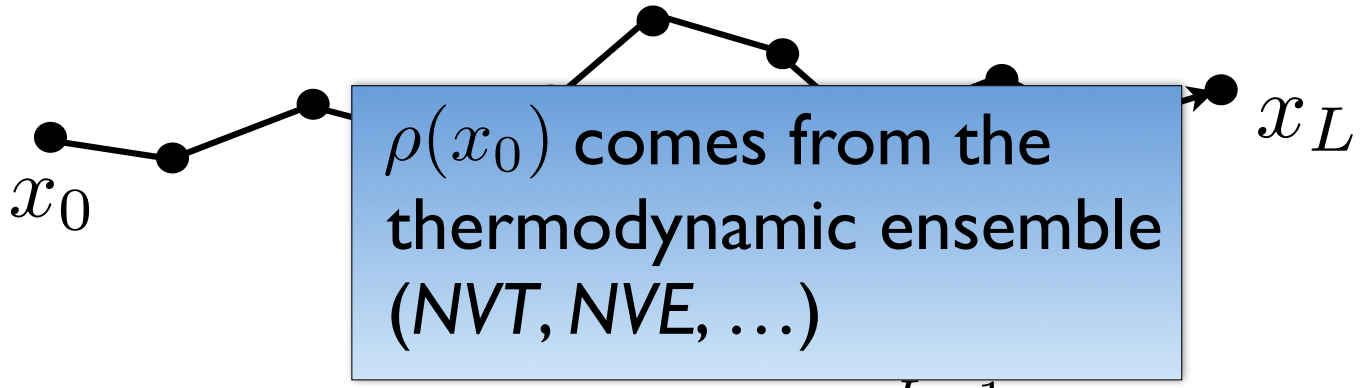
$$\mathcal{P}[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} P(x_i \to x_{i+1}) / Z(L)$$

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:

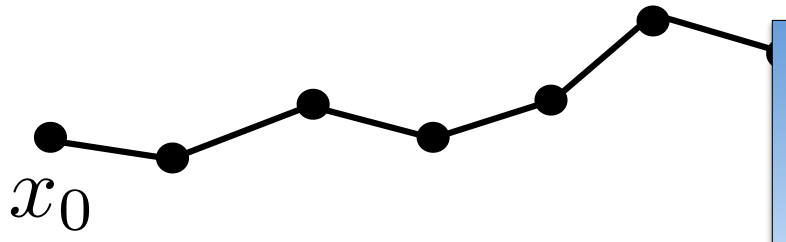
$$\mathcal{P}[\mathbf{x}] = \rho(x_0) \prod_{i=0}^{L-1} P(x_i \to x_{i+1}) / Z(L)$$

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



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

Path probability:

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

normalization constant (partition function)

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

Detailed Balance (Paths)

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

prob. of old state | prob of moving old → new | prob. of new state | prob of moving new → old

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

transition probability | generation probability | acceptance probability

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

Slide by David Swenson

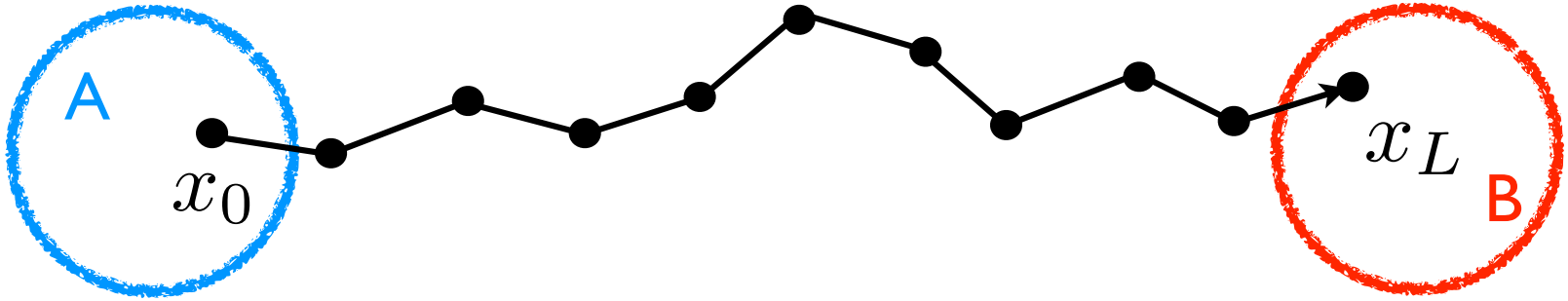
Transition Path

How do we focus on the transition region?

Indicator Function

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

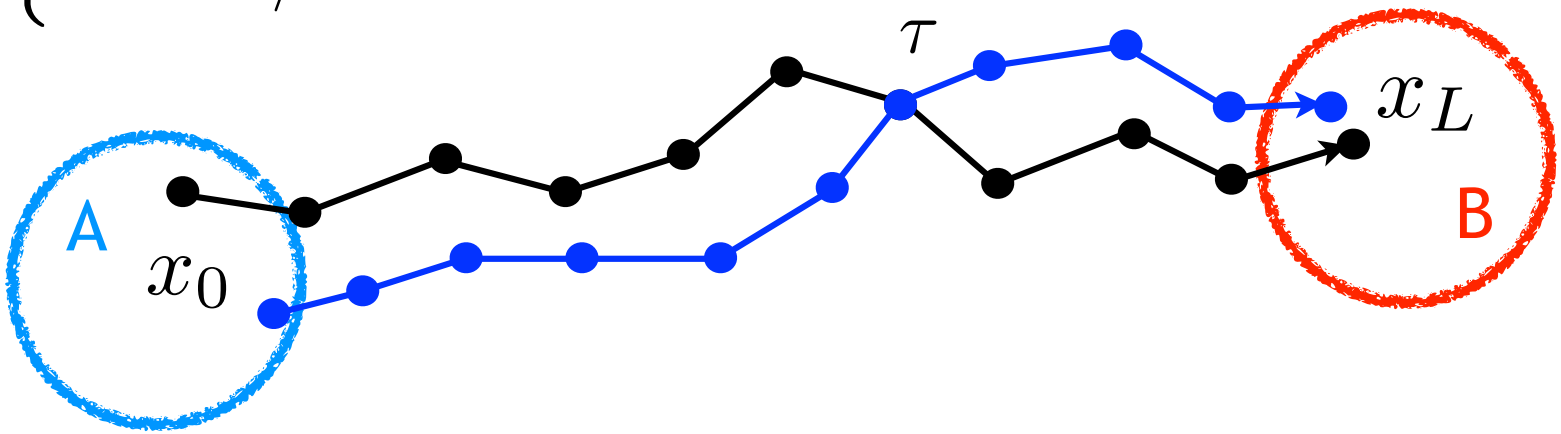
Note: $h_A(x) + h_B(x) \neq 1$!



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

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)

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

Shooting Algorithm: Detailed Balance

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

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

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

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

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

Slide by David Swenson

Shooting Move Acceptance

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

$$\frac{P_{\mathrm{gen}}^{\mathrm{shoot}}\left(\mathbf{x}^{(n)} \to \mathbf{x}^{(o)}\right)}{P_{\mathrm{gen}}^{\mathrm{shoot}}\left(\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}\right)} = \frac{P_{\mathrm{sel}}(x_\tau^{(n)}; \mathbf{x}^{(n)})}{P_{\mathrm{sel}}(x_\tau^{(o)}; \mathbf{x}^{(o)})} \frac{P_{\mathrm{mod}}(x_\tau^{(n)} \to x_\tau^{(o)})}{P_{\mathrm{mod}}(x_\tau^{(o)} \to x_\tau^{(n)})} \frac{P_{\mathrm{fwd}}(\mathbf{x}_{\tau\ldots L}^{(o)}) P_{\mathrm{bkwd}}(\mathbf{x}_{0\ldots\tau}^{(o)})}{P_{\mathrm{fwd}}(\mathbf{x}_{\tau\ldots L}^{(n)}) P_{\mathrm{bkwd}}(\mathbf{x}_{0\ldots\tau}^{(n)})}$$

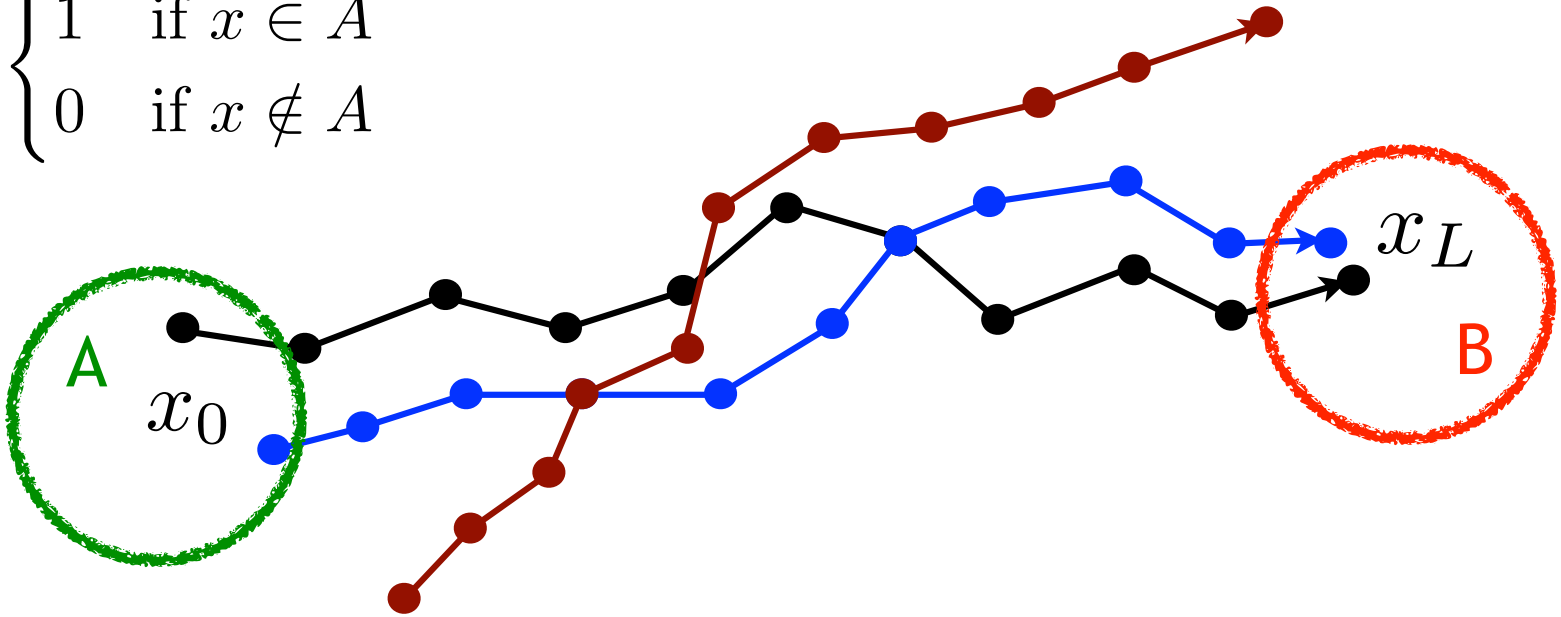
Assumptions:

- 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

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

Transition Path Sampling

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



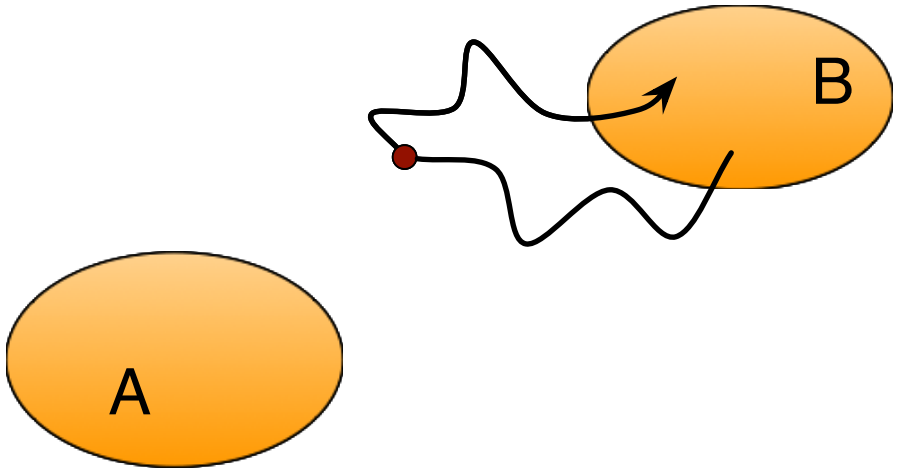
1. Select a point to "shoot" from (frame τ)
2. Modify the velocities at that point
3. Run forward (from τ to *T*) and backward (from τ to 0)
4. Accept/reject

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

What if the modification didn't matter?

Velocity is randomized after ~1ps in biomolecules.

Start with a shooting point at the transition state.



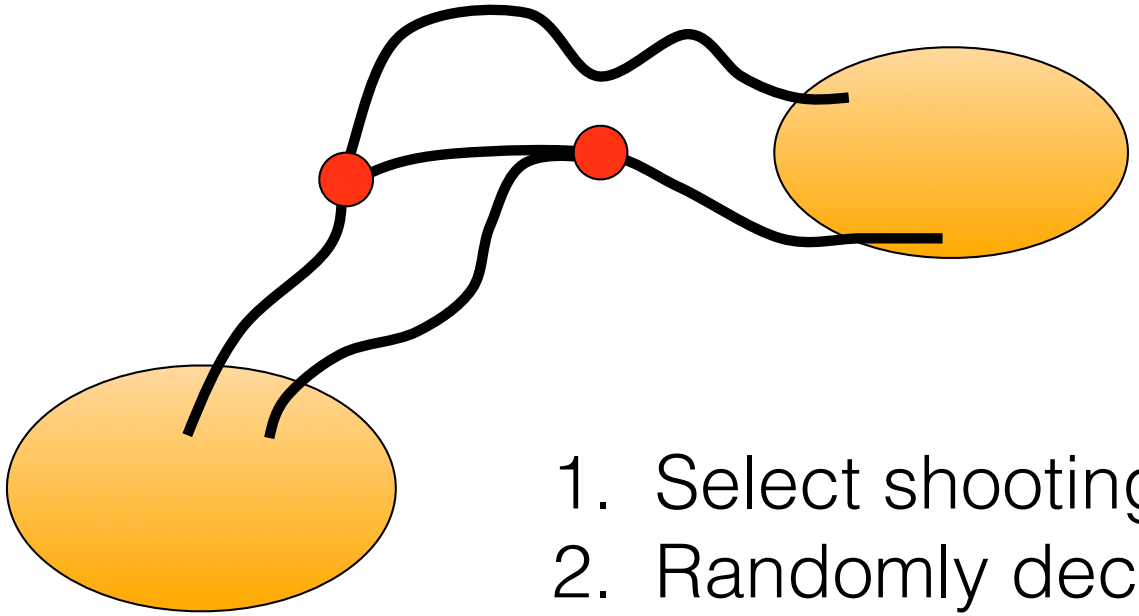
Prob. forward path hits B: 50%

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.

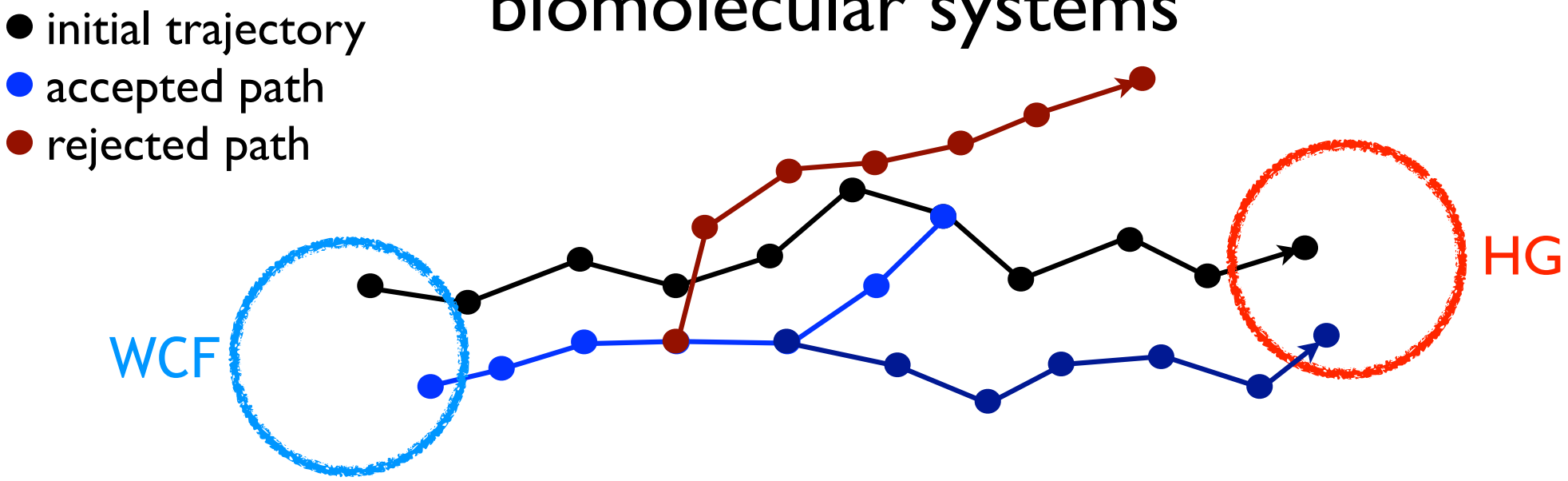
Let's design a shooting move that stops as soon as it enters the state. Accepted paths will have one frame in each state (first frame and last frame), and no other frames will be in any state. The path length will be allowed to vary.

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

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

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

Transition Path Sampling optimised for biomolecular systems



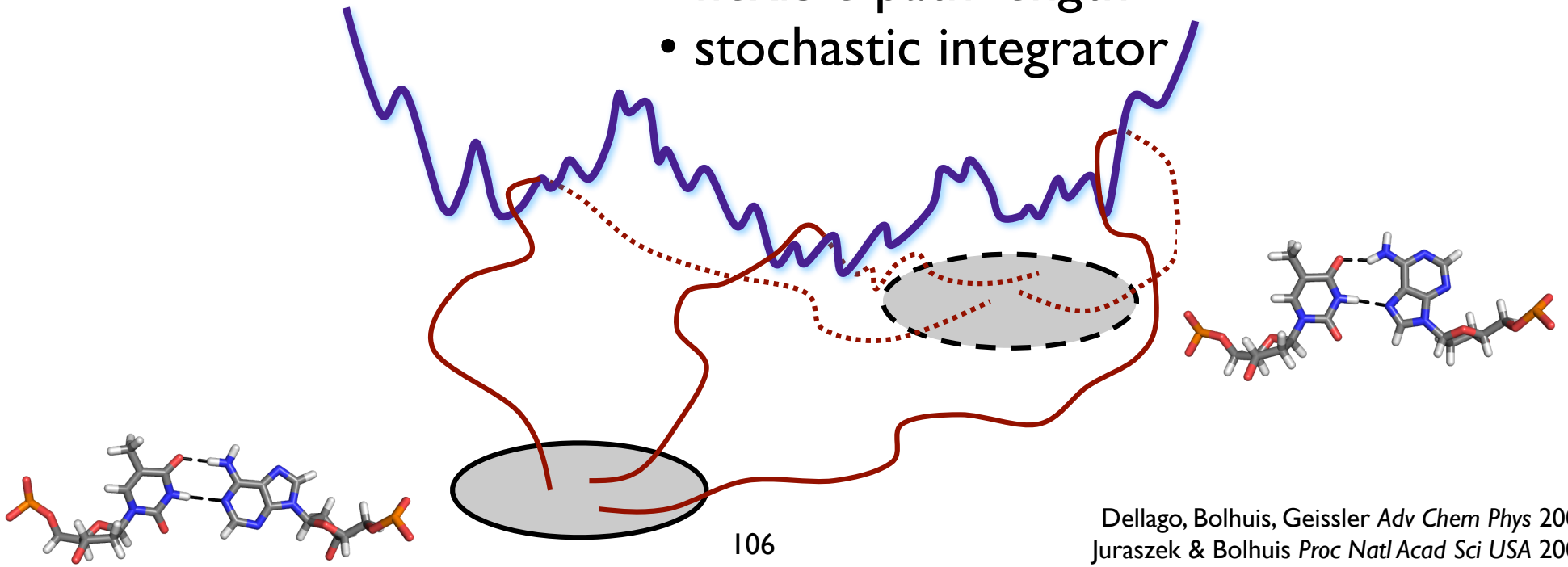
1. Select a point to "shoot" from
2. Randomly decide to shoot forward or backward
3. Run molecular dynamics with stochastic integrator until reaching the state, or reaching maximum length
4. Trial path consists of old and new frames
5. Decide
   a. accept if trial path connects both states
   b. reject if trial path does not connect both states

Sampling rare events: transition path sampling

Input

Result: ensemble of transition paths

Algorithm
- one way shooting
- flexible path length
- stochastic integrator

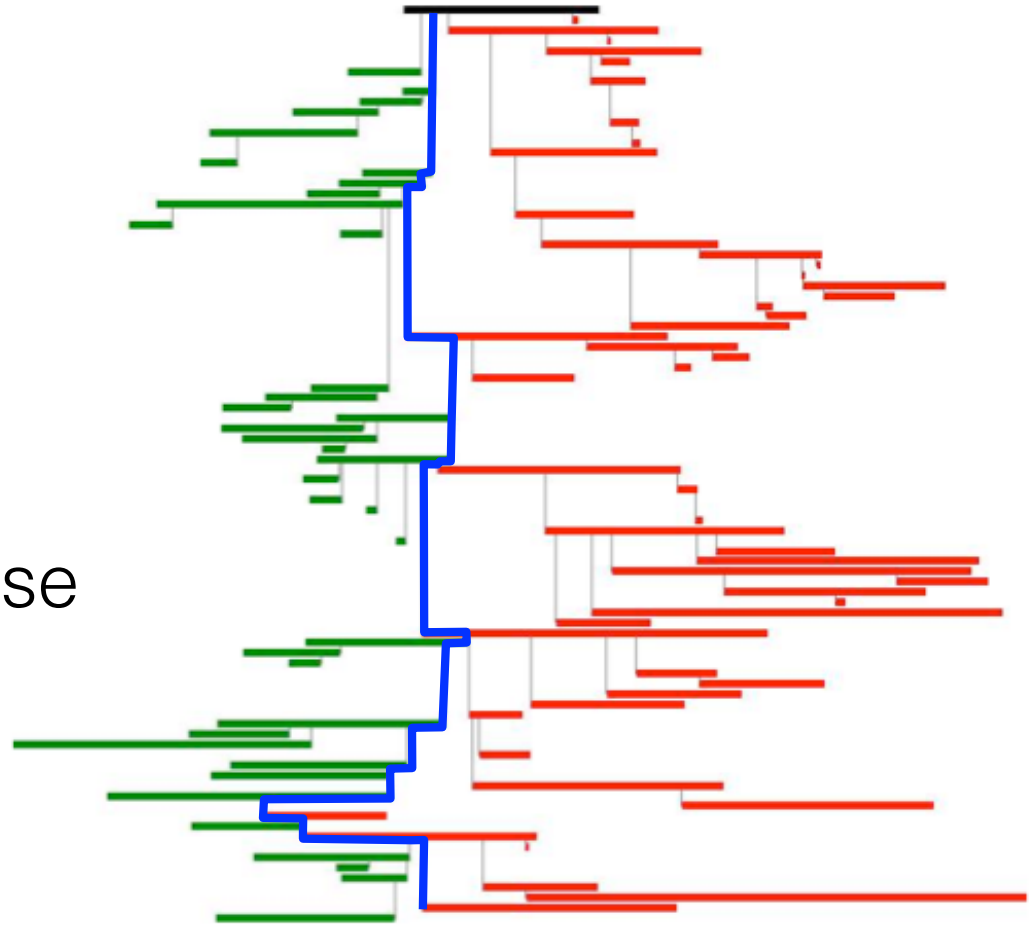


Dellago, Bolhuis, Geissler *Adv Chem Phys* 2002
Juraszek & Bolhuis *Proc Natl Acad Sci USA* 2006

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



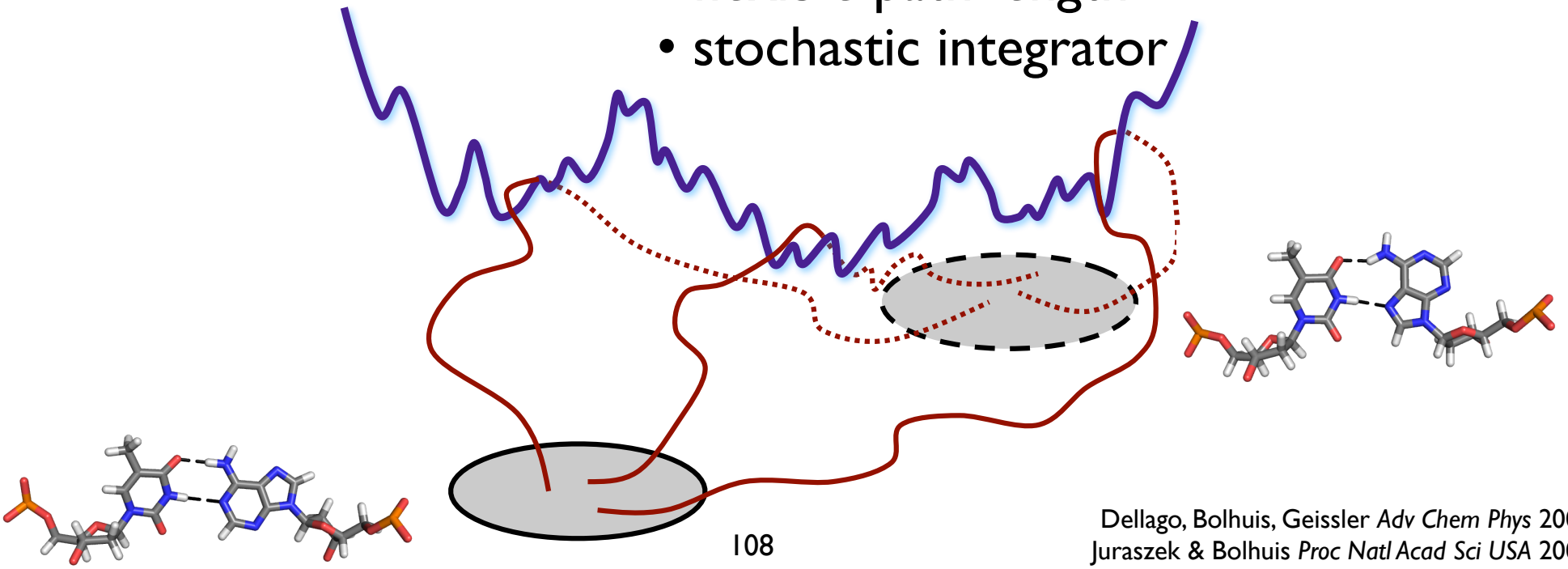
Least-changed path: Stays near top of barrier; much easier than transition state ensemble!

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



Dellago, Bolhuis, Geissler *Adv Chem Phys* 2002
Juraszek & Bolhuis *Proc Natl Acad Sci USA* 2006

How do we define states?

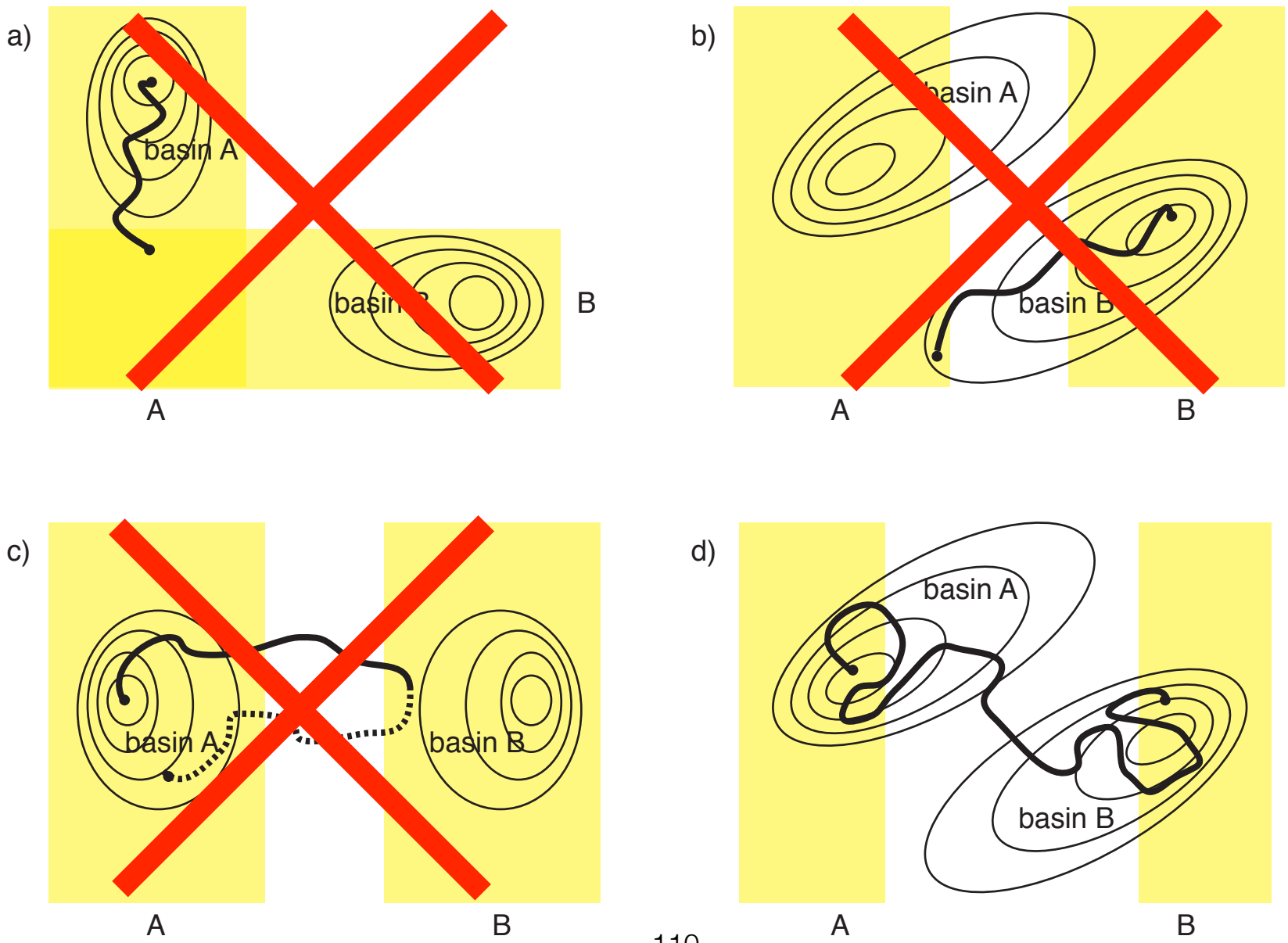
Take something that you can measure (distance, dihedral angle) and say it has to be within some range of values.

Examples:

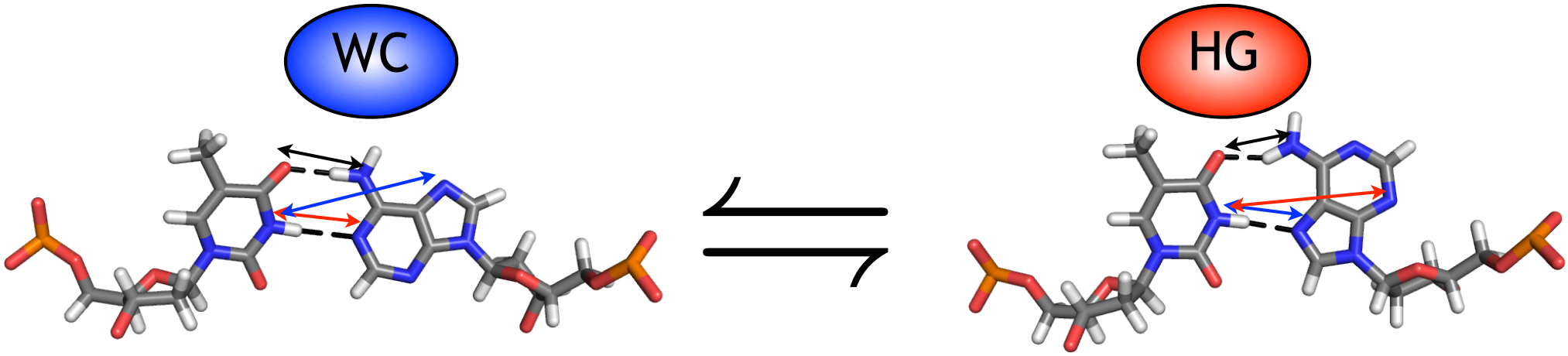
- Hydrogen bond formed ($d_{donor\text{-}acceptor} < 3.5$ Å)
- Specific value of (several) dihedral angles

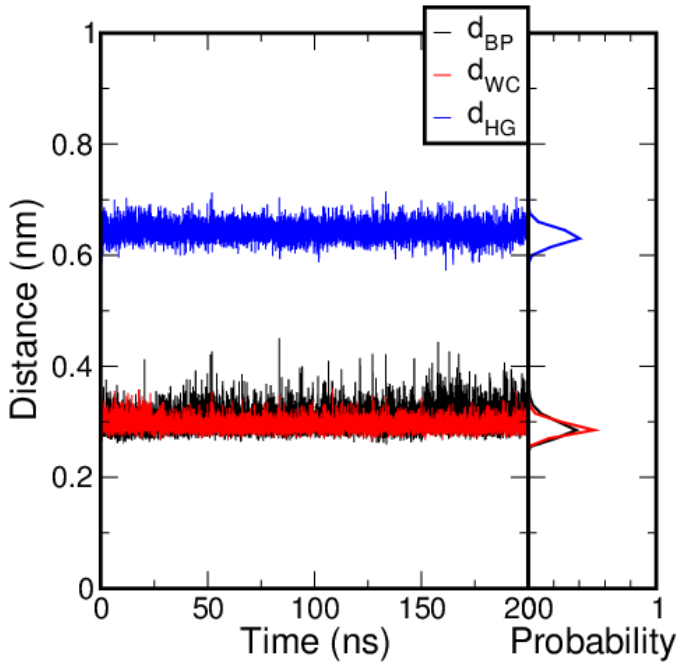
What makes a good definition?

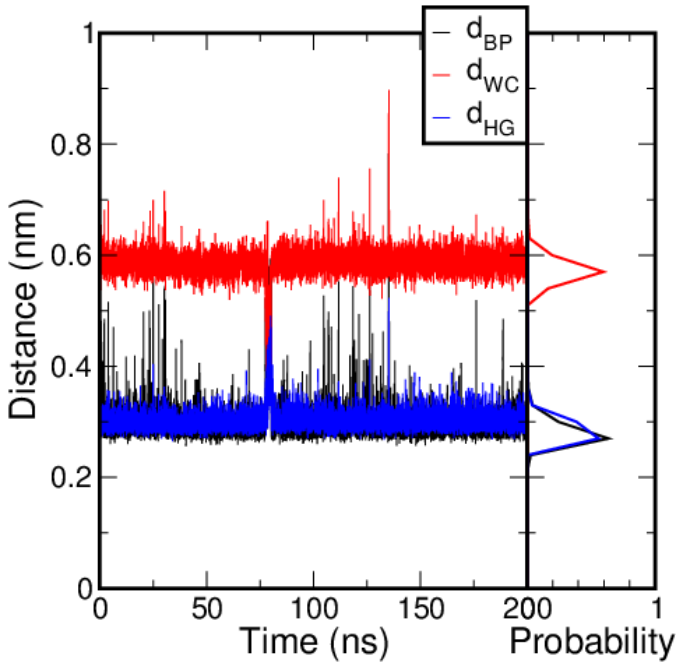
How do we define states?



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







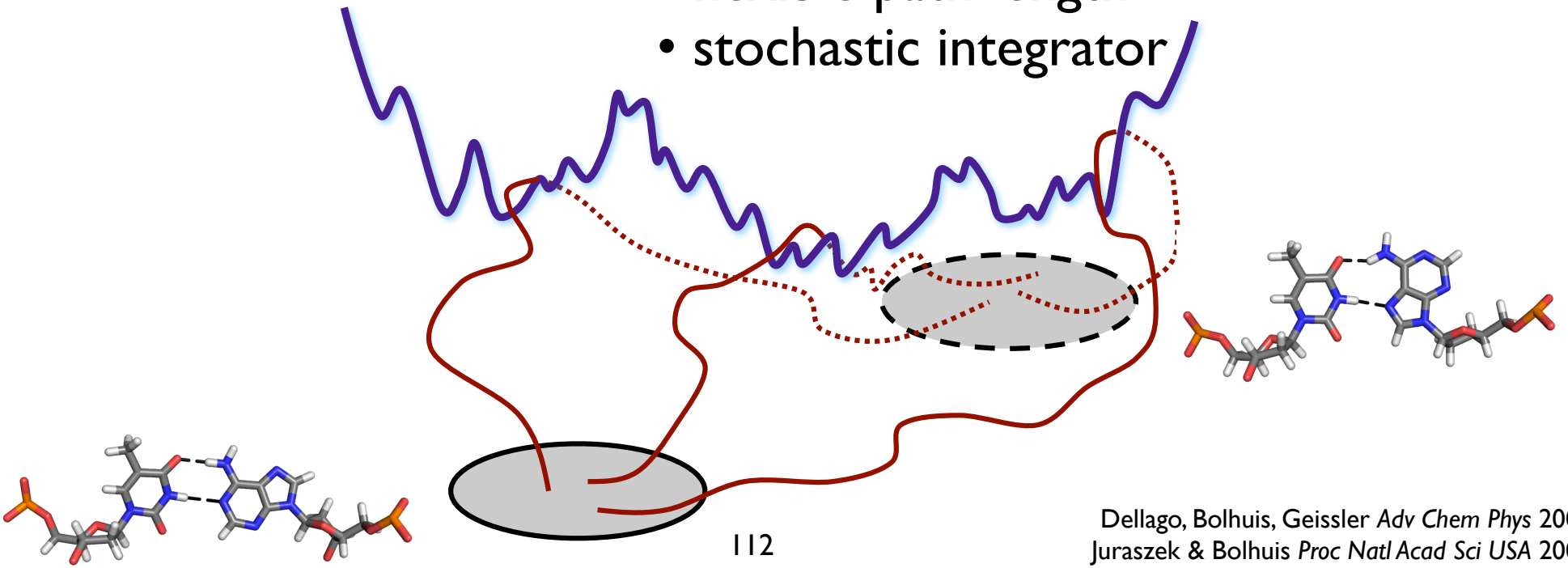
Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Sampling rare events: transition path sampling

Input
- initial reactive path
- definitions of stable states

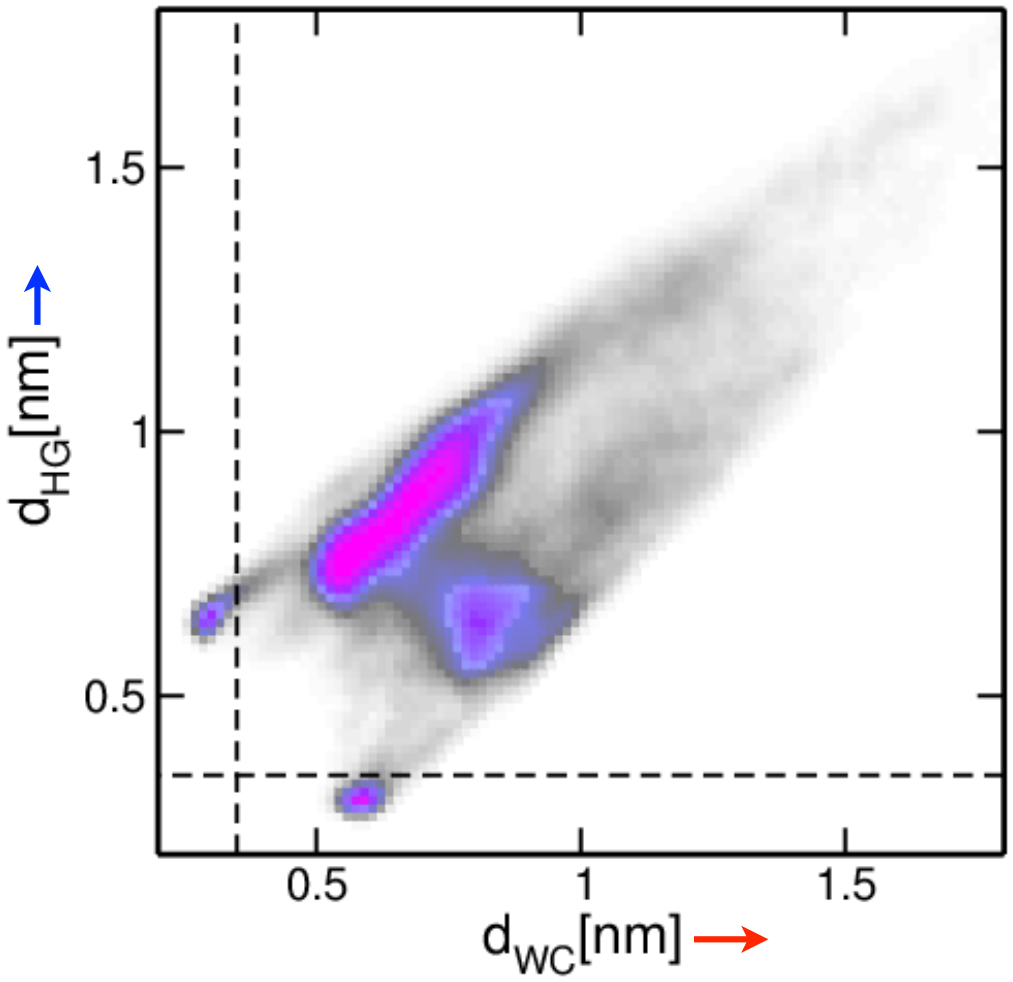
Result: ensemble of transition paths
- mechanism
- relevant reaction coordinates

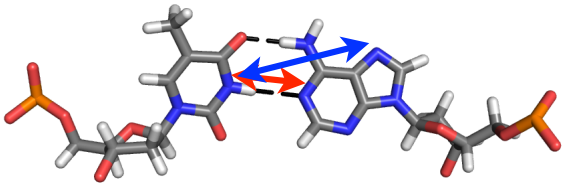
Algorithm
- one way shooting
- flexible path length
- stochastic integrator



Dellago, Bolhuis, Geissler *Adv Chem Phys* 2002
Juraszek & Bolhuis *Proc Natl Acad Sci USA* 2006

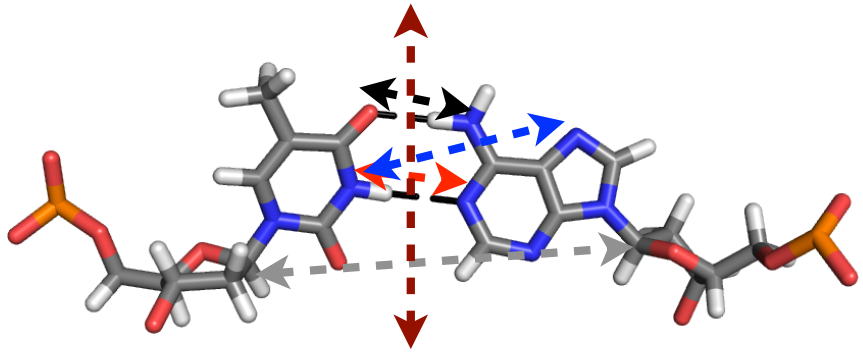
Generate reactive trajectories using transition path sampling



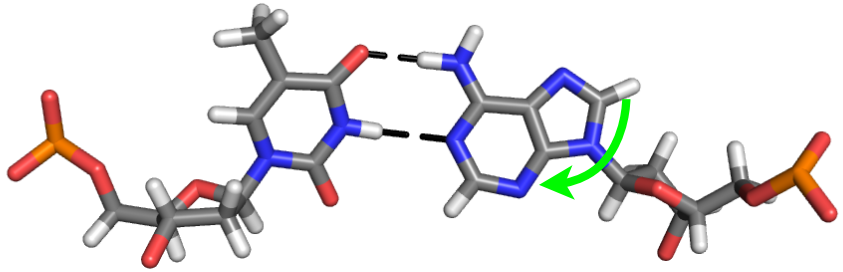


Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

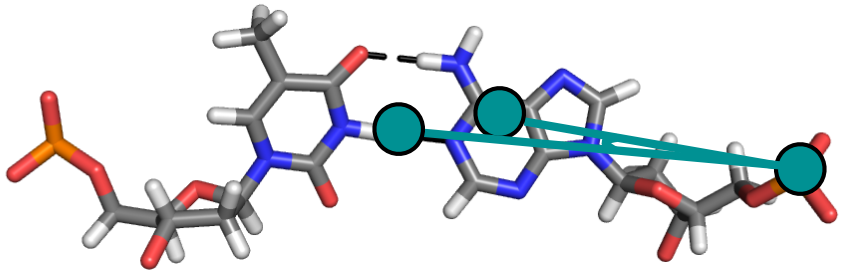
Biasing DNA baserolling defining CVs



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



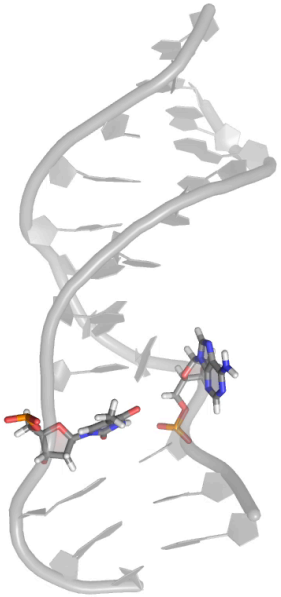
base rolling angle



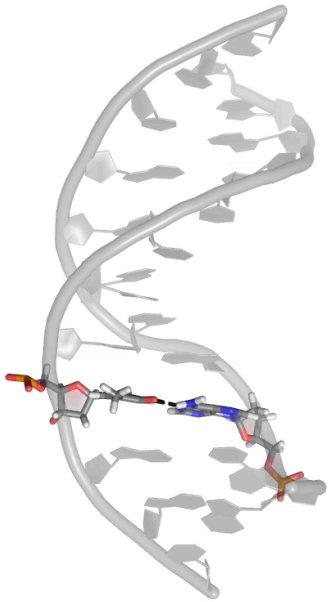
base flipping angle

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

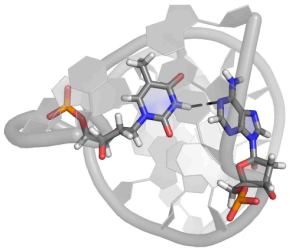
Two types of transitions

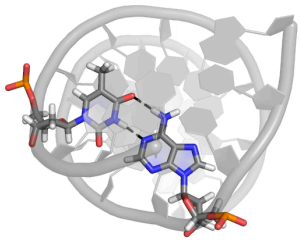


outside

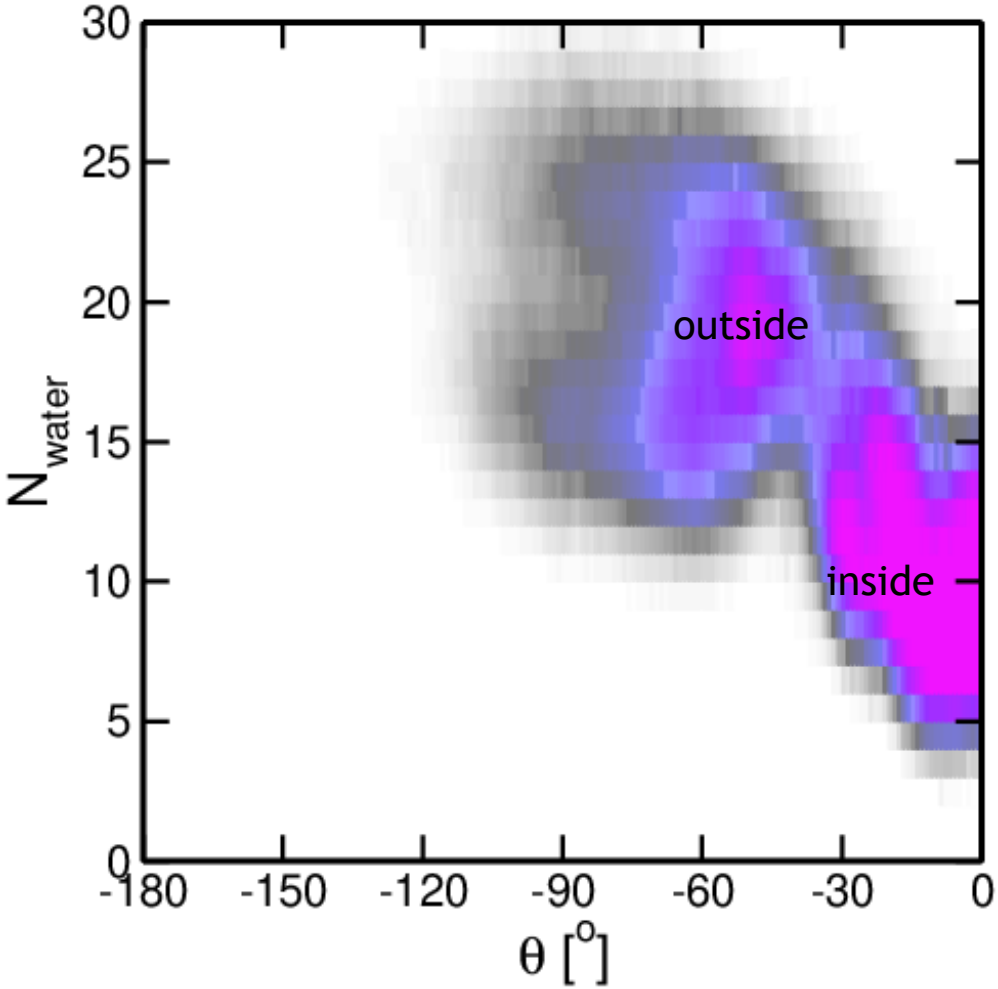


inside

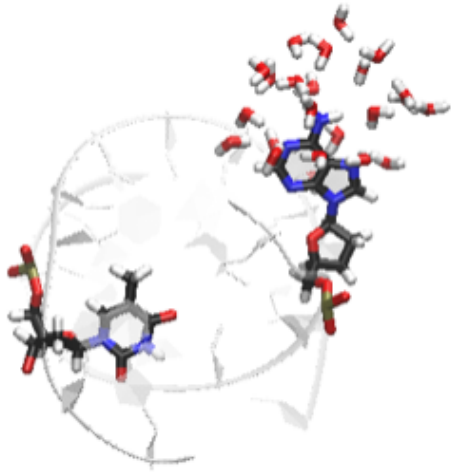




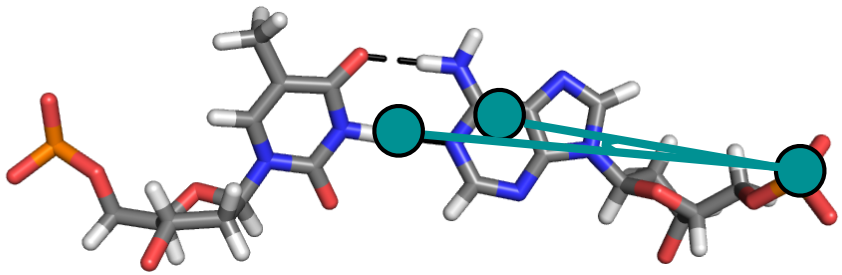
Two reaction channels: inside and outside



N_{water}

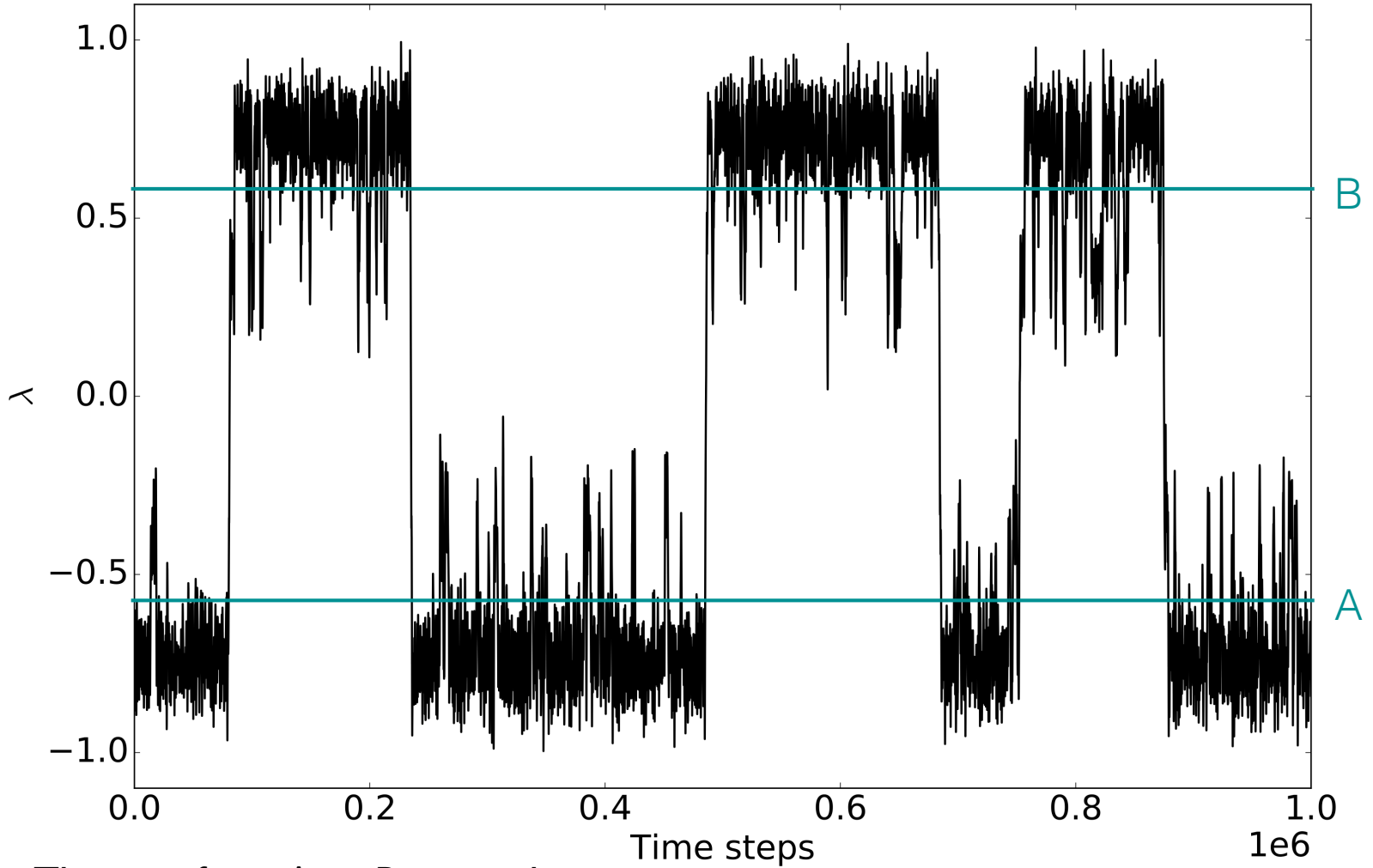


θ base flipping angle



Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Calculate the rate



The rate from A to B is equal to

the number of transitions per time unit

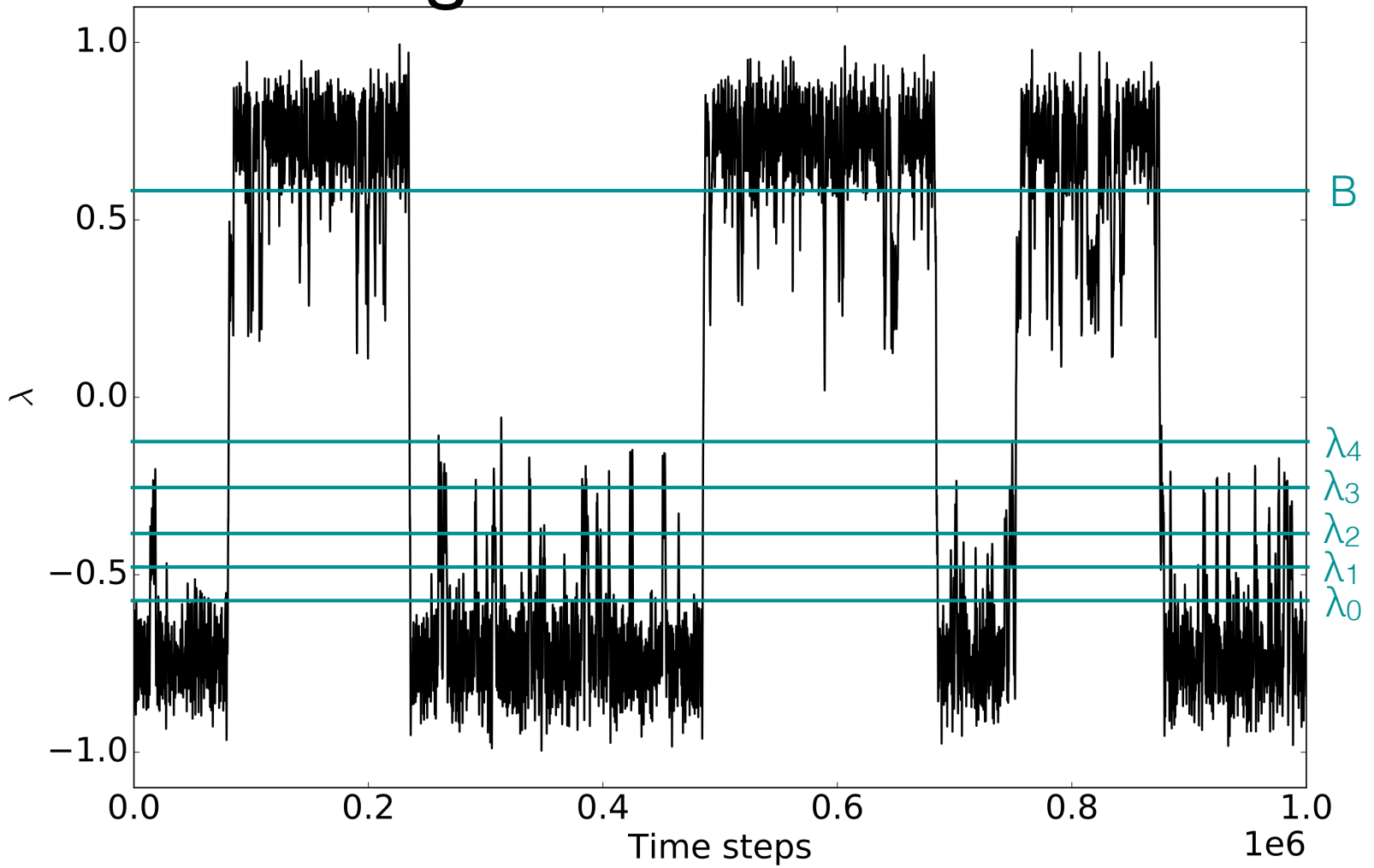
divided by the time spent in state A

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

Slide design by David Swenson

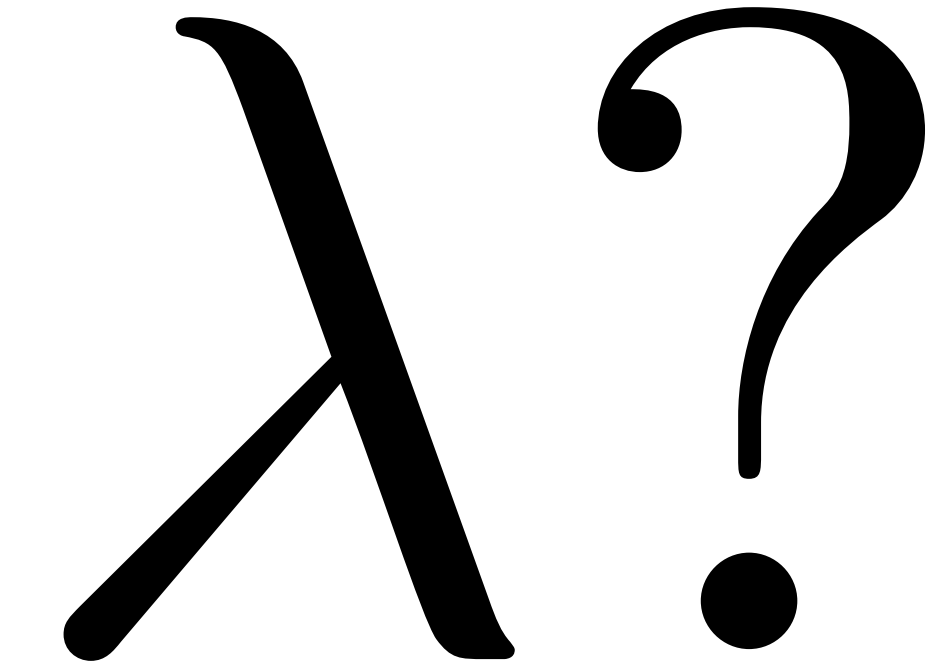
van Erp et al. J. Chem. Phys 2003

Transition Interface Sampling: calculating the rate with interfaces

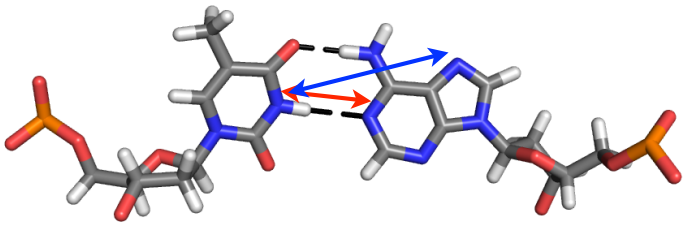


$$k_{AB} = \phi_{A,0} P(B|\lambda_m) \prod_{i=0}^{m-1} P(\lambda_{i+1}|\lambda_i)$$

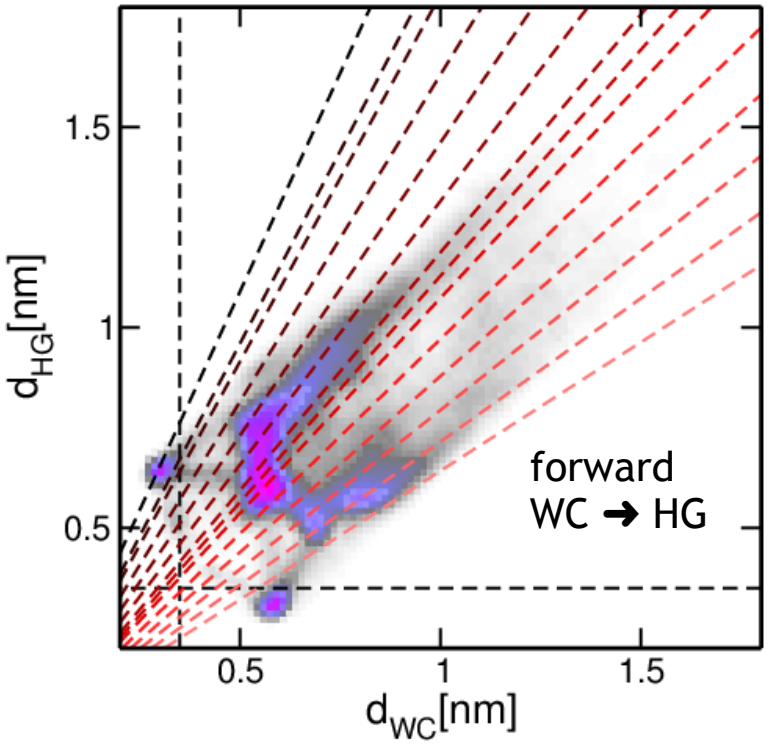
Slide design by David Swenson
van Erp et al. J. Chem. Phys 2003

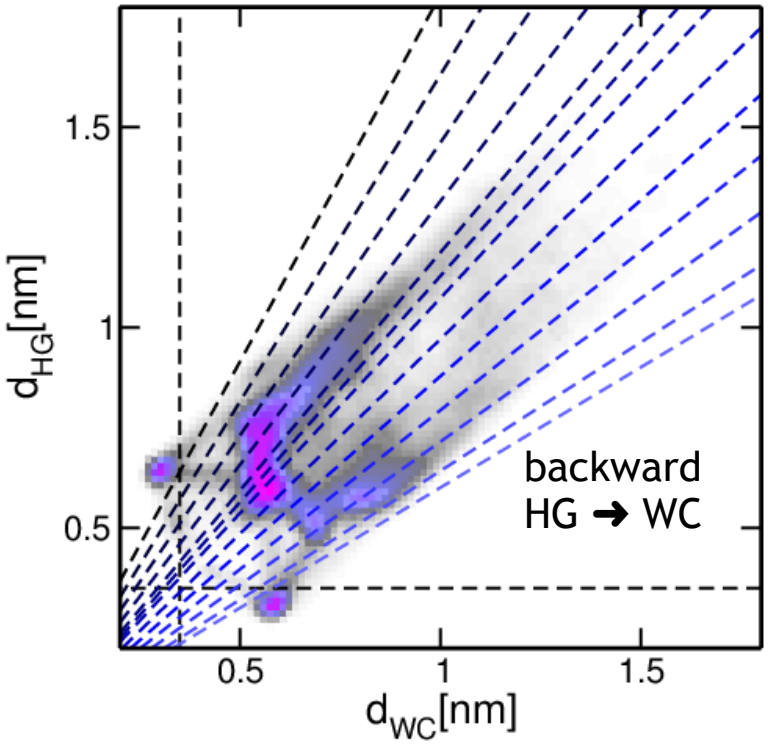


Interfaces



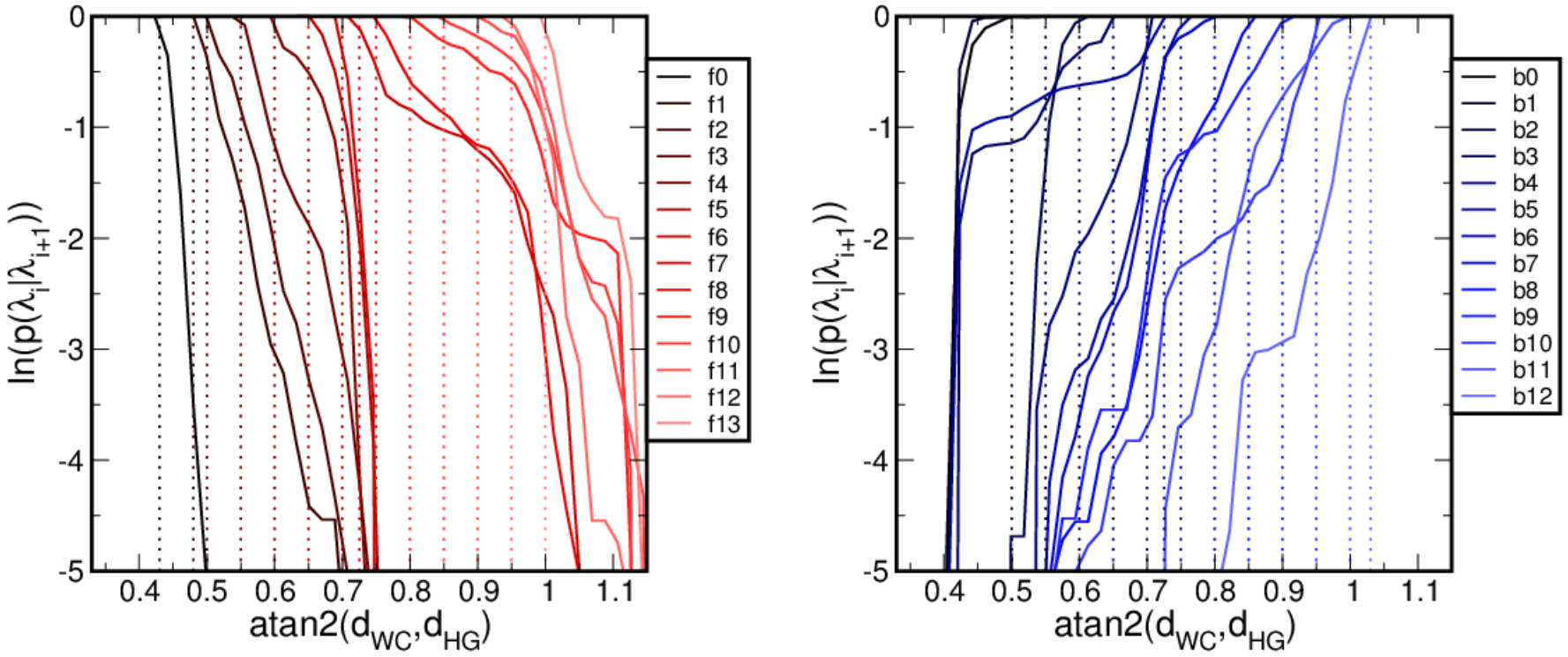
$$\lambda = \arctan(d_{WC}, d_{HG})$$





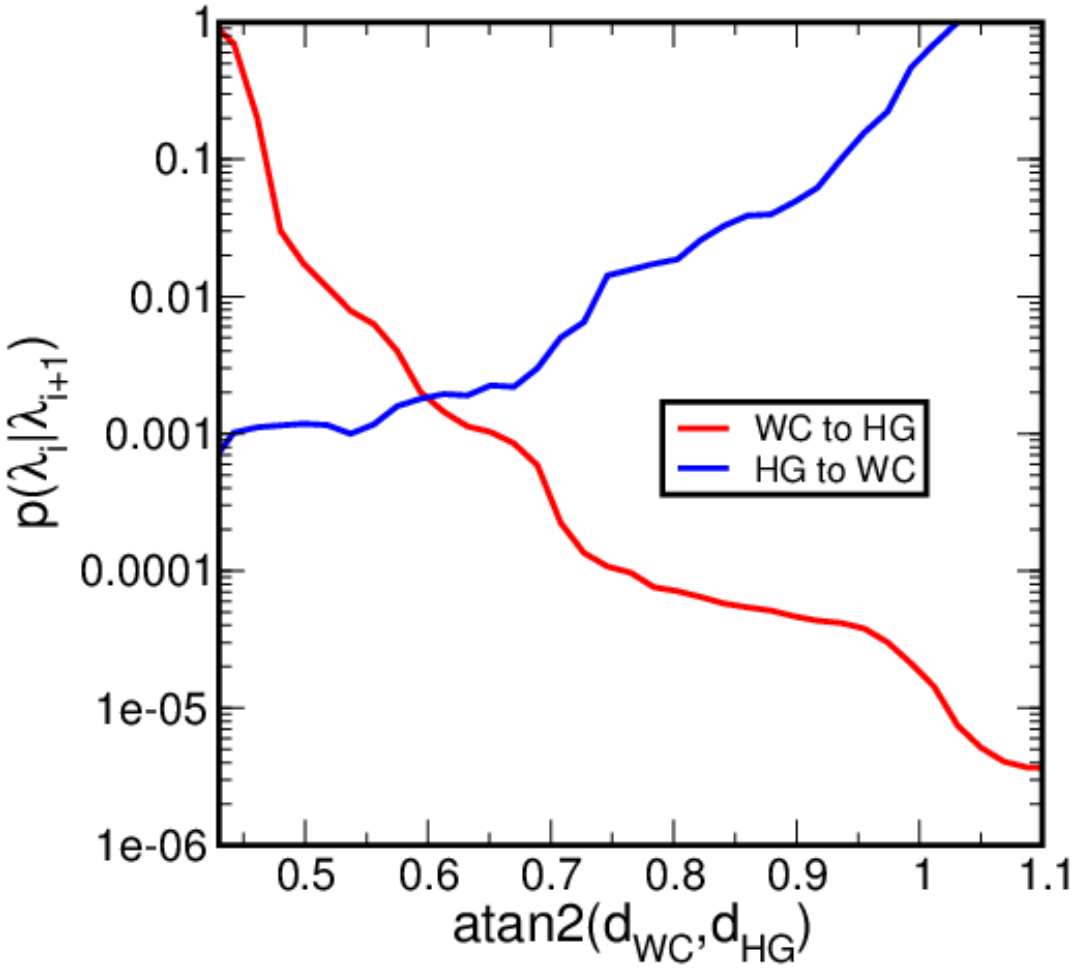
Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Crossing probabilities



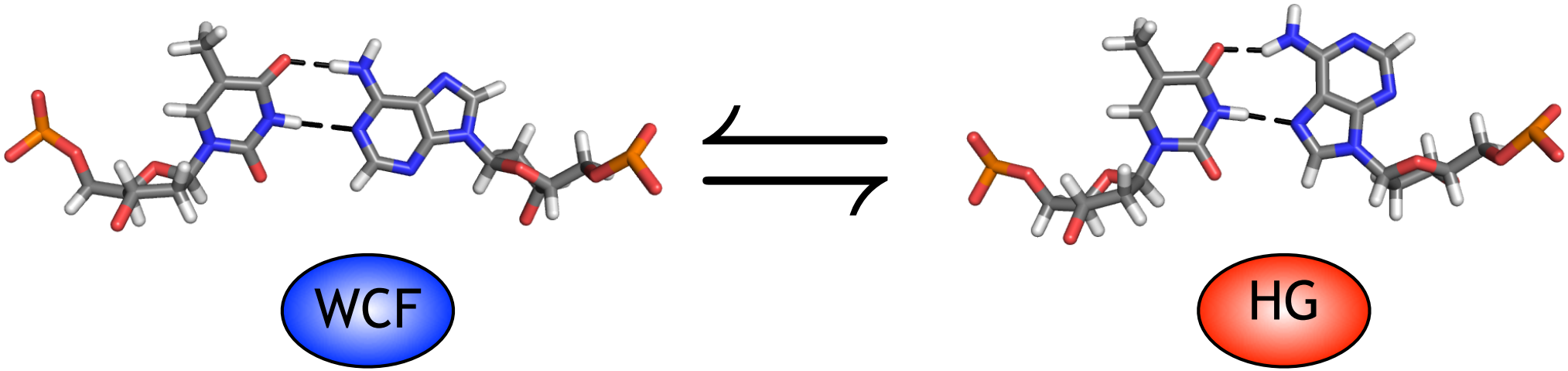
Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Crossing probabilities



Vreede, Pérez de Alba Ortíz, Bolhuis & Swenson, NAR 2019

Rates



	expt*	TIS**
$k_{WC \to HG}$ (s^{-1})	14.2 ± 1.03	742
$k_{HG \to WC}$ (s^{-1})	3670 ± 200	$1.6 \cdot 10^5$
ΔG ($k_B T$)	5.5	5.4
	at 26.0°C	at 300K

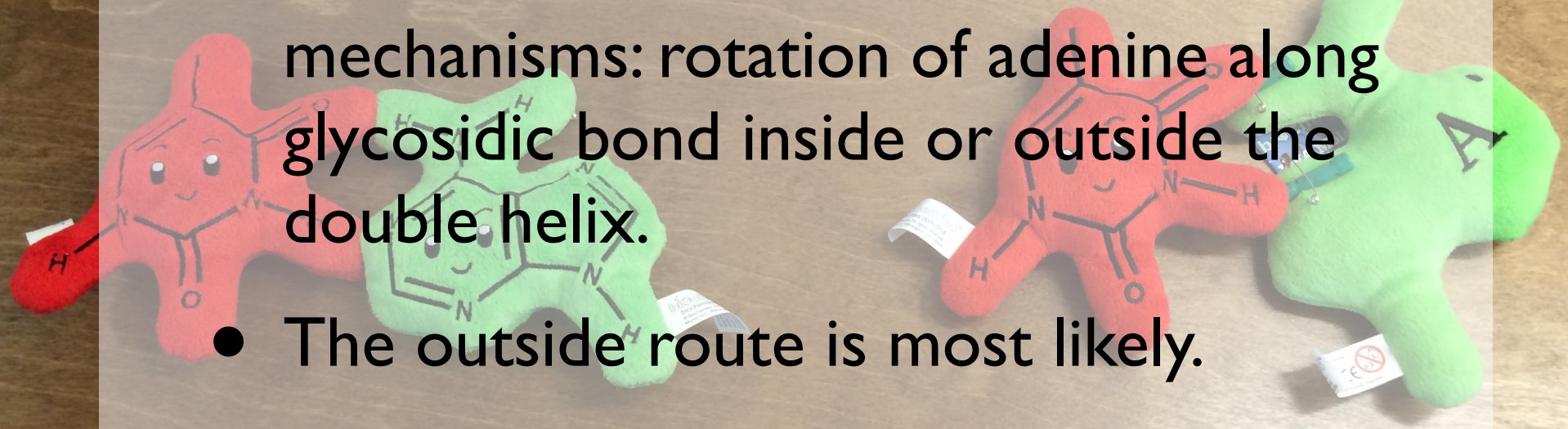
TIS:
one-way shooting
flexible path length
500 accepted paths per interface
13 interfaces

*Nikolova, Kim, Wise, O'Brien, Andricioaei and Al-Hashimi
Nature 2011 v. 470, p. 498
**Vreede, Ortiz de Alba Perez, Bolhuis & Swenson, NAR 2019

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

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