Rare event simulation

Macroscopic phenomenological theory

Chemical reaction: A ⟷ B

$$\frac{dc_A(t)}{dt} = -k_{A\to B}c_A(t) + k_{B\to A}c_A(t)$$

$$\frac{dc_B(t)}{dt} = +k_{A\to B}c_A(t) - k_{B\to A}c_A(t)$$

Total number of molecules:

$$\frac{d[c_A(t) + c_B(t)]}{dt} = 0$$

Equilibrium:

$$\frac{dc_A(t)}{dt} = \frac{dc_B(t)}{dt} = 0 \qquad \frac{\langle c_A \rangle}{\langle c_B \rangle} = \frac{k_{B\to A}}{k_{A\to B}}$$

Rare event simulation

Macroscopic phenomenological theory

Make a small perturbation:

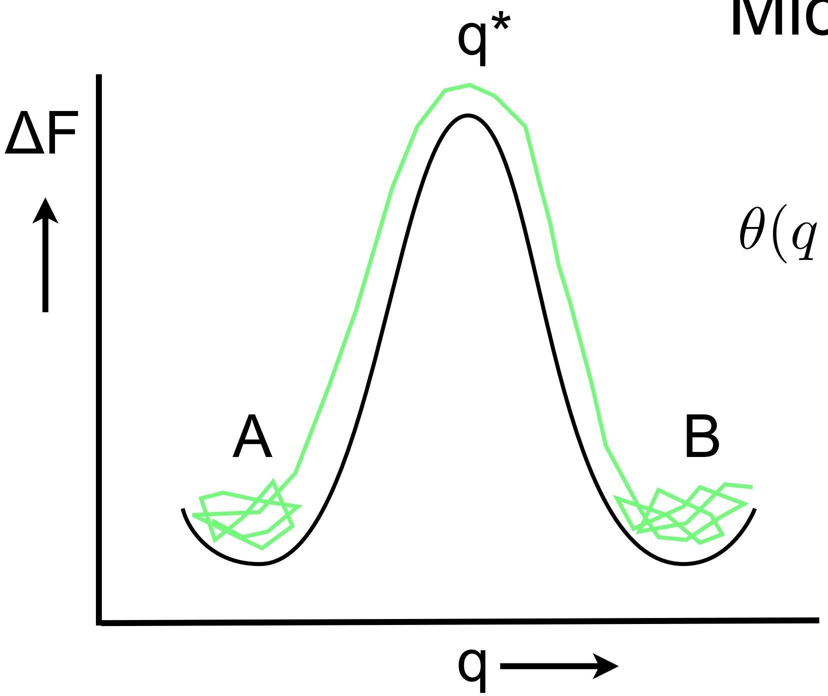
$$c_A(t) = \langle c_A \rangle + \Delta c_A(t) \qquad c_B(t) = \langle c_B \rangle + \Delta c_A(t)$$

$$\frac{d\Delta c_A(t)}{dt} = -k_{A\rightarrow B}\Delta c_A(t) - k_{B\rightarrow A}\Delta c_A(t)$$

$$\begin{aligned} \Delta c_A(t) &= \Delta c_A(0)\exp[-(k_{A\rightarrow B} + k_{B\rightarrow A})t] \\ &= \Delta c_A(0)\exp[-t/\tau] \end{aligned}$$

$$\begin{aligned} \tau &= \left(k_{A\rightarrow B} + k_{B\rightarrow A}\right)^{-1} \\ &= k_{A\rightarrow B}^{-1}\left(1 + \langle c_A \rangle / \langle c_B \rangle\right)^{-1} = \frac{\langle c_B \rangle}{k_{A\rightarrow B}} \end{aligned}$$

Rare event simulation



Microscopic linear response theory

$$\theta(q-q^*) = \begin{cases} 0 \quad \text{if} \quad q-q^* < 0 \quad \text{(Reactant A)} \\ 1 \quad \text{if} \quad q-q^* > 0 \quad \text{(Product B)} \end{cases}$$

Perturbation

add bias to increase concentration c_A

$$H = H_0 - \epsilon g_A(q-q^*)$$

$$g_A(q-q^*) = 1 - \theta(q-q^*) = \theta(q^*-q)$$

$$\Delta c_A = \langle c_A \rangle_\epsilon - \langle c_A \rangle_0$$

$$\Delta c_A = \langle g_A \rangle_\epsilon - \langle g_A \rangle_0$$

$\langle g_A \rangle \rightarrow$ probability to be in state A

Very small perturbation: linear response theory

$$\Delta c_A = \langle g_A \rangle_\epsilon - \langle g_A \rangle_0 \qquad H = H_0 - \epsilon g_A(q - q^*)$$

How does the response (Δc) depend on the perturbation (Δε)?

$$\begin{aligned}\frac{d\Delta c_A}{d\epsilon} &= \beta\Big(\langle (g_A)^2 \rangle_0 - \langle g_A \rangle_0^2\Big) \\ &= \beta\Big(\langle g_A \rangle_0 \big(1 - \langle g_A \rangle_0\big)\Big) \\ &= \beta\Big(\langle c_A \rangle_0 \big(1 - \langle c_A \rangle_0\big)\Big) = \beta \langle c_A \rangle_0 \langle c_B \rangle_0\end{aligned}$$

Outside the barrier

$g_A = 0 \text{ or } 1$

$g_A(x) g_A(x) = g_A(x)$

Switch of the perturbation: dynamic linear response

$$\begin{aligned}\Delta c_A(t) &= \Delta c_A(0) \frac{\Delta g_A(0) \Delta g_A(t)}{\langle c_A \rangle \langle c_B \rangle} \\ &= \Delta c_A(0) \exp[-t/\tau]\end{aligned}$$

holds for sufficiently long times

Linear response theory: static

$$\langle \Delta A \rangle = \langle A \rangle - \langle A \rangle_0 \qquad H = H_0 - \epsilon B$$

$$\langle A \rangle_0 = \frac{\int d\Gamma A \exp[-\beta(H_0)]}{\int d\Gamma \exp[-\beta(H_0)]} \qquad \langle A \rangle = \frac{\int d\Gamma A \exp[-\beta(H_0 - \epsilon B)]}{\int d\Gamma \exp[-\beta(H_0 - \epsilon B)]}$$

$$\begin{aligned}\left\langle \frac{\partial \Delta A}{\partial \epsilon} \right\rangle &= \frac{\int d\Gamma \beta A B \exp[-\beta(H_0 - \epsilon B)] \int d\Gamma \exp[-\beta(H_0 - \epsilon B)]}{\left\{ \int d\Gamma \exp[-\beta(H_0 - \epsilon B)] \right\}^2} \\ &- \frac{\int d\Gamma A \exp[-\beta(H_0 - \epsilon B)] \int d\Gamma \beta B \exp[-\beta(H_0 - \epsilon B)]}{\left\{ \int d\Gamma \exp[-\beta(H_0 - \epsilon B)] \right\}^2} \\ &= \beta \left\{ \langle AB \rangle_0 - \langle A \rangle_0 \langle B \rangle_0 \right\}\end{aligned}$$

$$\exp[-t/\tau] = \frac{\left\langle \Delta g_A(0)\Delta g_A(t)\right\rangle}{\left\langle c_A\right\rangle\left\langle c_B\right\rangle}$$

Derivative $$-\frac{1}{\tau}\exp[-t/\tau] = \frac{\left\langle g_A(0)\dot{g}_A(t)\right\rangle}{\left\langle c_A\right\rangle\left\langle c_B\right\rangle} = \frac{\left\langle \dot{g}_A(0)g_A(t)\right\rangle}{\left\langle c_A\right\rangle\left\langle c_B\right\rangle}$$

Δ has disappeared
because of derivative

For sufficiently short *t*

$$k_{A\rightarrow B}(t) = \frac{\left\langle \dot{g}_A(0)g_A(t)\right\rangle}{\left\langle c_A\right\rangle}$$

$$\dot{g}_A(q-q^*) = \dot{q}\frac{\partial g_A(q-q^*)}{\partial q} = -\dot{q}\frac{\partial g_B(q-q^*)}{\partial q}$$

$$k_{a\rightarrow B}(t) = \frac{\left\langle \dot{q}(0)\frac{\partial g_B(q(0)-q^*)}{\partial q}g_B(t)\right\rangle}{\left\langle c_A\right\rangle}$$

Stationary

$$\frac{d}{dt}\langle A(t)B(t+t')\rangle = 0$$

$$\langle A(t)\dot{B}(t+t')\rangle + \langle \dot{A}(t)B(t+t')\rangle = 0$$

$$\langle A(t)\dot{B}(t+t')\rangle = -\langle \dot{A}(t)B(t+t')\rangle$$

Eyring's transition state theory

$$\begin{aligned} k_{a\rightarrow B}(t) &= \frac{\left\langle \dot{q}(0)\frac{\partial g_B(q(0)-q^*)}{\partial q} g_B(t)\right\rangle}{\left\langle c_A \right\rangle} \\ &= \frac{\left\langle \dot{q}(0)\delta(q(0)-q^*)\theta(q(t)-q^*)\right\rangle}{\left\langle \theta(q^*-q)\right\rangle} \end{aligned}$$

Correlation between velocity of states that are at the top of the barrier at *t=0* and in the product state B some time *t* later.

Let us consider the limit $t \rightarrow 0^+$:

$$\begin{aligned} \lim_{t\rightarrow 0^+} &= \theta\big(q(t)-q^*\big) = \theta\big(\dot{q}(t)\big) \\ k^{\mathrm{TST}}_{a\rightarrow B}(t) &= \frac{\left\langle \dot{q}(0)\delta(q(0)-q^*)\theta(\dot{q})\right\rangle}{\left\langle \theta(q^*-q)\right\rangle} \end{aligned}$$

Bennett-Chandler approach

(or Reactive flux method)

$$k_{a\to B}(t) = \frac{\left\langle \dot{q}(0)\delta(q(0)-q^*)\theta(q(t)-q^*)\right\rangle}{\left\langle \theta(q^*-q)\right\rangle}$$

$$k_{a\to B}(t) = \frac{\left\langle \dot{q}(0)\delta(q(0)-q^*)\theta(q(t)-q^*)\right\rangle}{\left\langle \delta(q(0)-q^*)\right\rangle} \times \frac{\delta(q(0)-q^*)}{\theta(q^*-q)}$$

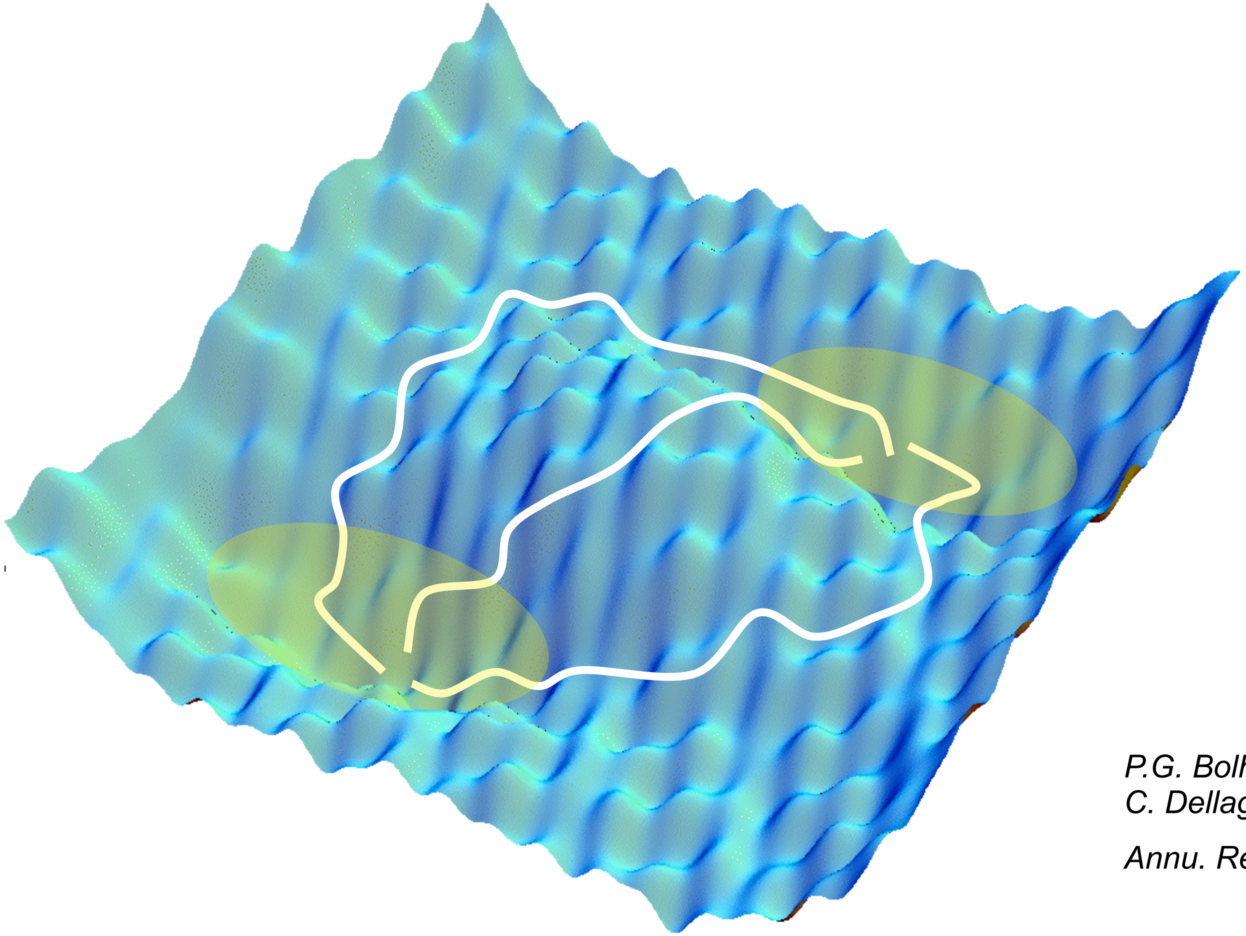
Conditional average: $\dot{q}(0)\theta(q(t)-q^*)$
given that we start on top of barrier

Probability to find *q*
on barrier top

Computational scheme:

- Determine the probability with free energy calculation
- Compute conditional average from “shooting” trajectories from barrier top

Transition path sampling

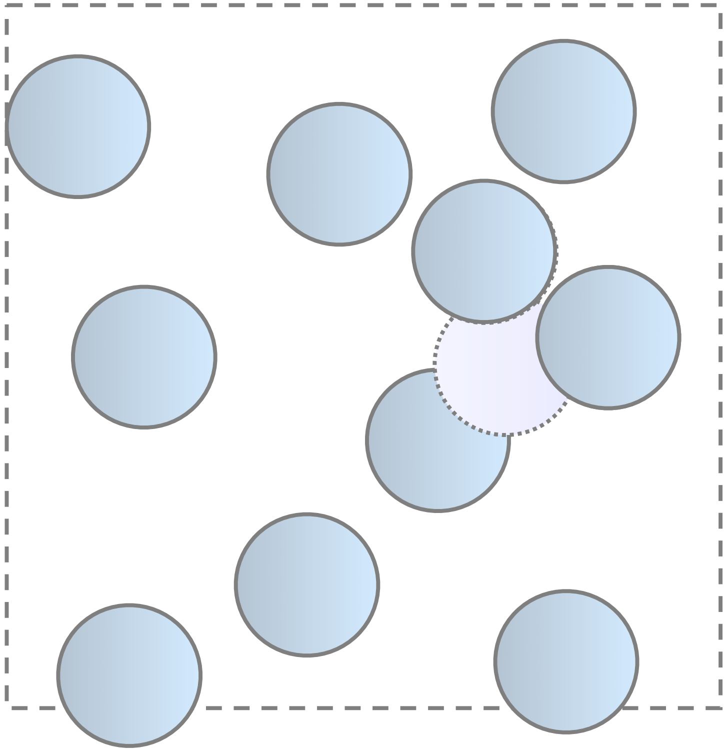


P.G. Bolhuis, D. Chandler,
C. Dellago, P.L. Geissler

Annu. Rev. Phys. Chem 2002

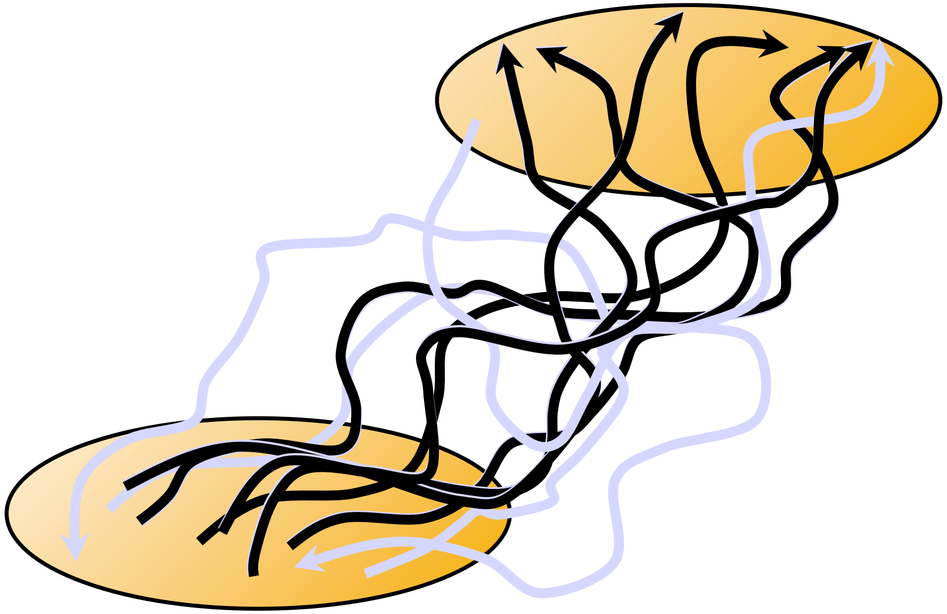
Importance sampling of the path ensemble:
all trajectories that lead over barrier and connect stable states.

Transition path sampling



Monte Carlo simulation:

trials by random moves

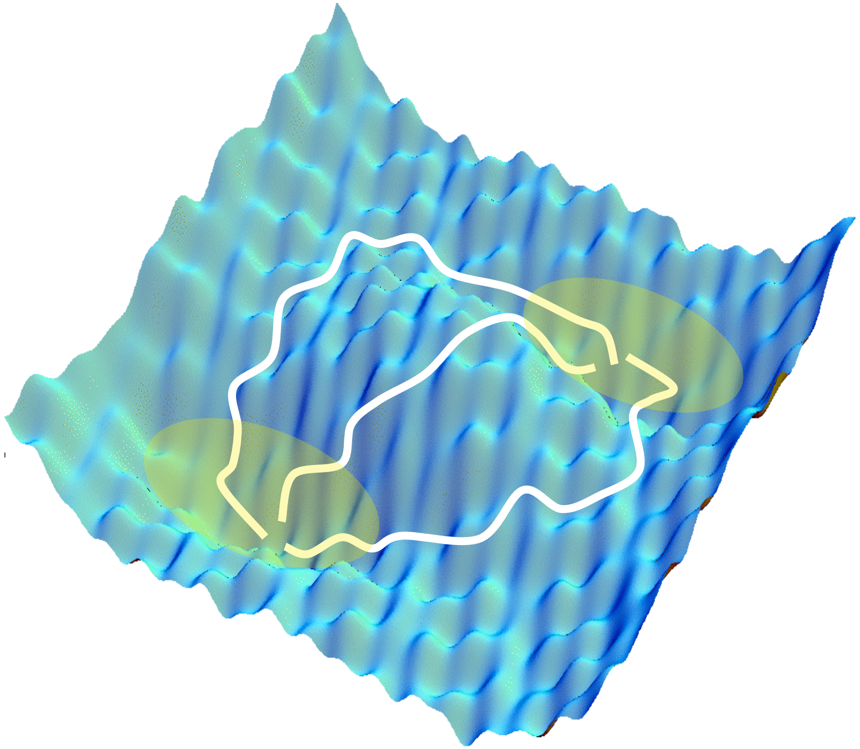


Transition path sampling:

trial paths by shooting move

TPS is Monte Carlo sampling of dynamical trajectories

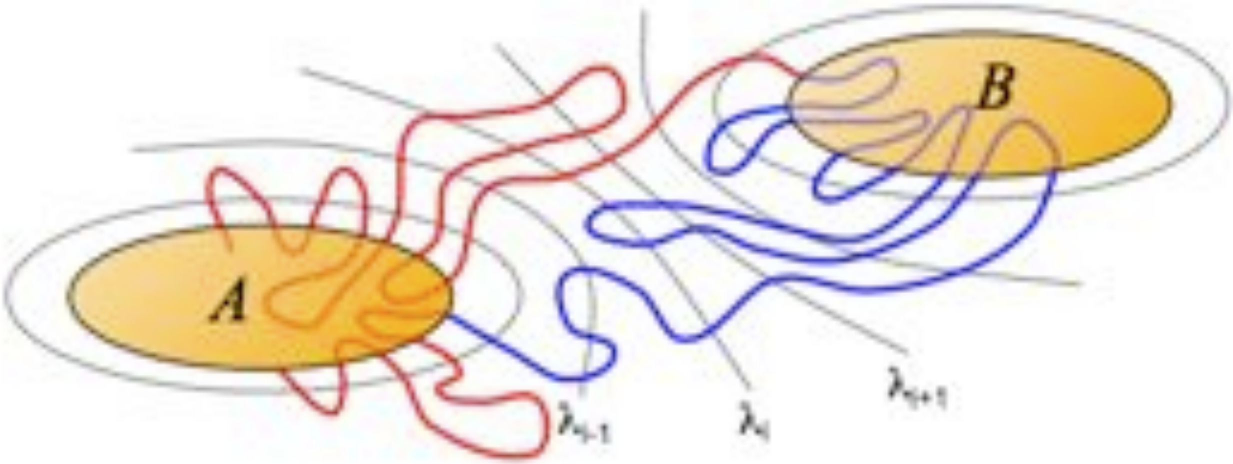
Transition path sampling



- No reaction coordinate (only state definition)
- True dynamics
- Works well for 2-state problems
- Needs an initial pathway
- Analysis required
- Transition state ensemble

Transition interface sampling

Diffusive processes



- Introduce interfaces between the stable states
- Sample pathways between the interfaces
- Committor surfaces (probability to end up in *B*)

fin

Free energy methods

It is very (very) difficult to compute (or measure) absolute thermodynamic properties, such as free energy and entropy, that depend on the size of the phase space.

But we can compute relative free energies, in particular free energy differences between thermodynamic states.

Reaction equilibrium constants $A \longleftrightarrow B$

$$K = \frac{[B]}{[A]} = \frac{p_B}{p_A} = [-\beta(G_B - G_A)]$$

Examples:

- Chemical reactions, catalysis, isomerization, etc...
- Protein folding, ligand binding affinity, protein-protein association
- Phase diagrams, coexistence lines, critical points, transitions

Free energy perturbation

$$\Delta\beta F = -\ln(Q_B/Q_A) = -\ln\left(\frac{\int ds^N \exp(-\beta U_B)}{\int ds^N \exp(-\beta U_A)}\right)$$

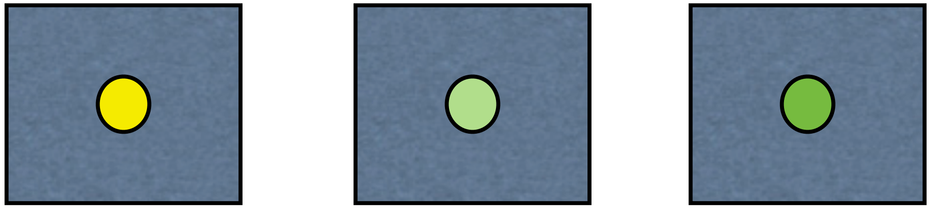
$$\begin{aligned}\Delta F &= -k_B T \ln\left(\frac{\int ds^N \exp(-\beta U_A)\exp(-\beta\Delta U)}{\int ds^N \exp(-\beta U_A)}\right)\\ &= -k_B T \ln\left\langle \exp(-\beta\Delta U)\right\rangle_A\end{aligned}$$

$$\Delta F = -k_B T \ln\left\langle \exp(-\beta\Delta U)\right\rangle_A = -k_B T \ln\left\langle \exp(\beta\Delta U)\right\rangle_B$$

Sampling problems may lead to hysteresis between the two samples

Alchemical transformation

E.g. solvation free energy difference between ions

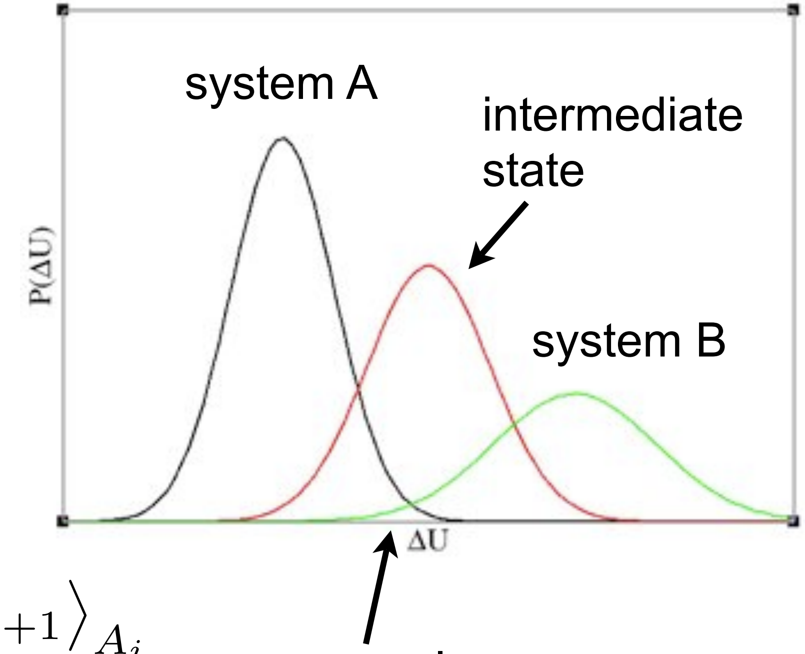


Cl^- intermediate stage(s) Br^-

To improve the sampling,
- define *n-2* fictitious intermediate states
- compute ΔF between state *i* and *i+1*

$$H(\lambda_i) = \lambda_i H_A + (1-\lambda_i) H_B$$

$$\begin{aligned}\Delta F &= \sum_{i=1}^{N-1} \Delta F_{i,i+1} \\ &= -k_B T \sum_{i=1}^{N-1} \ln \left\langle \exp -\beta \Delta U_{i,i+1} \right\rangle_{A_i}\end{aligned}$$



poor overlap

Umbrella Sampling

Bias the samping along an order parameter *q*

Add and subtract bias potential *w(q)*:

$$P(q) = \frac{\int d\mathbf{r}^N \delta(q'(\mathbf{r}^N) - q) \exp\left[-\beta(U(\mathbf{r}^N) + w(q') - w(q'))\right]}{\int d\mathbf{r}^N \exp\left[-\beta(U(\mathbf{r}^N + w(q') - w(q'))\right]}$$

$$P(q) = \frac{\int d\mathbf{r}^N \delta(q'(\mathbf{r}^N) - q) \exp\left[-\beta(U(\mathbf{r}^N) + w(q'))\right] \exp(\beta w(q'))}{\int d\mathbf{r}^N \exp\left[-\beta(U(\mathbf{r}^N + w(q'))) \exp(\beta w(q'))\right]}$$

$$P(q) = \frac{\left\langle \delta(q'(\mathbf{r}^N) - q) \exp(\beta w(q')) \right\rangle_{\mathrm{biased}}}{\left\langle \exp(\beta w(q')) \right\rangle_{\mathrm{biased}}}$$

$$P(q) = \frac{\exp(\beta w(q))}{\left\langle \exp(\beta w(q')) \right\rangle_{\mathrm{biased}}} P_{\mathrm{biased}}(q)$$

- Let w(q) be a good guess of minus the free energy F(q), or
- Choose w(q) to confine sampling to a specific window along q.
- Or do both.

$$F(q) = k_B T \ln P(q) = -k_B T \ln P_{\mathrm{biased}}(q) - w(q) + \mathrm{const}$$

Constrained MD

The derivative of the free energy F(λ) with respect to λ can be written as an ensemble average.

$$\left(\frac{\partial F(\lambda)}{\partial \lambda}\right)_{\mathrm{NVT}} = \frac{\int d\mathbf{r}^N (\partial U(\lambda)/\partial\lambda) \exp[-\beta U(\lambda)]}{\int d\mathbf{r}^N \exp[-\beta U(\lambda)]}$$

$$= \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

The free energy difference between states A and B can then be obtained by thermodynamic integration

$$F(\lambda_B) - F(\lambda_A) = \int_{\lambda_A}^{\lambda_B} d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

In the case "hard" constraints are used, additional corrections are needed to unbias for sampling in a constraint ensemble (instead of the actual NVT ensemble)

Metadynamics

Escaping free-energy minima,
Laio and Parrinello, PNAS (2002)

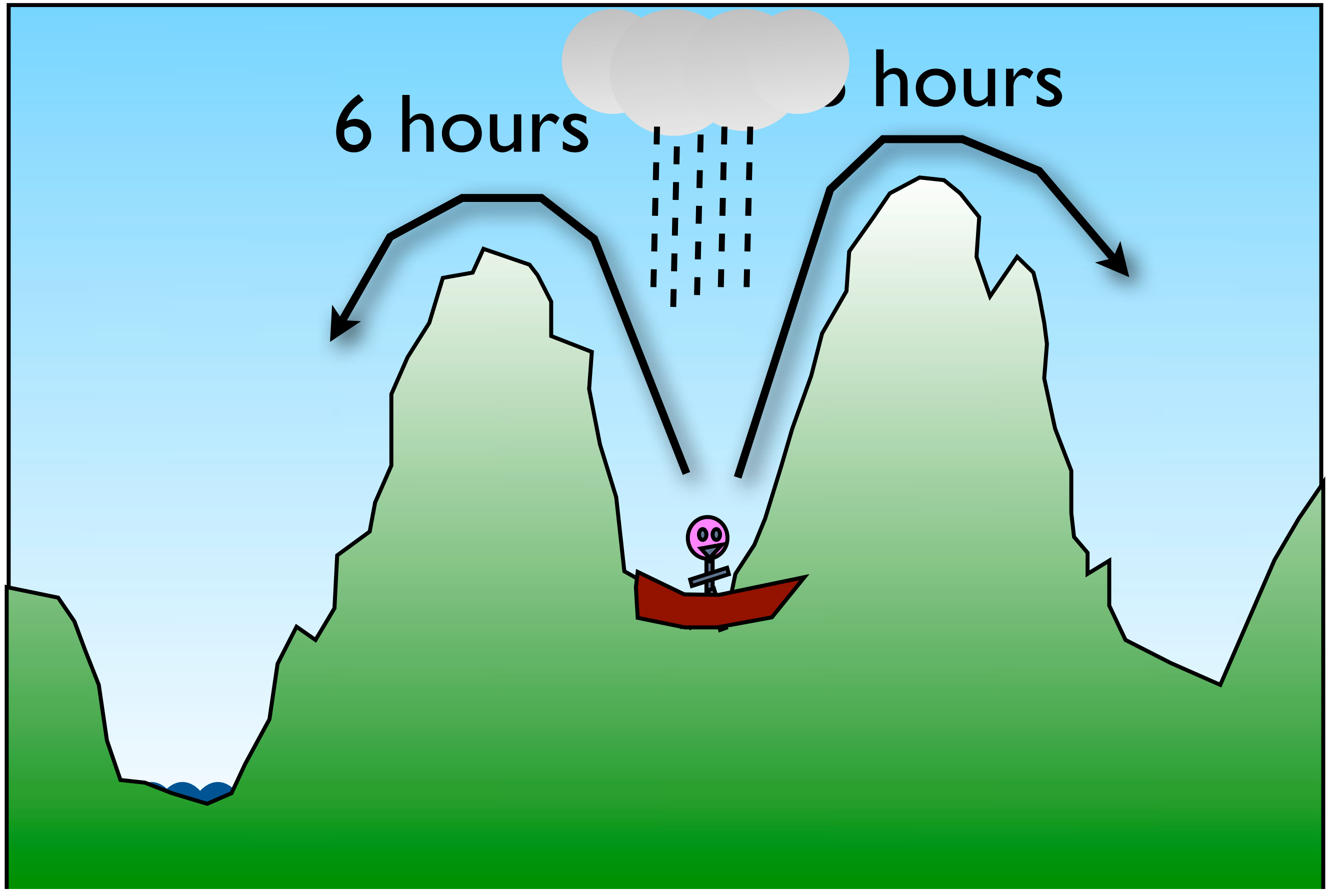
$$V(t,s) = \sum_{t'<t} H_{t'} \prod_{\alpha} \exp\left[\frac{-(s_\alpha - s_\alpha^{t'})^2}{2\delta_\alpha^2 W^2} \right]$$

- The metadynamics biasing potential “grows” with time, by adding relatively small repulsive Gaussian potentials.
- The potentials placed at visited points in the space of order parameters (collective variables) enhance sampling of unexplored regions.
- The Gaussian “hills” accumulate in the free energy minima, until the counter-balance the basins and allow the system to escape to product states, where the process repeats
- The biasing potential is an estimator of the free energy.

The error depends on the height, width, and time interval of the added Gaussians, and on the diffusion, temperature and order parameter space of the system.

$$\epsilon = C_d \sqrt{\frac{HWS}{D\Delta\tau\beta}}$$

flooding the landscape with hills



Steered MD

Mechanical work to bring the system from state A to state B

$$W_{A\to B} \geq \Delta F_{A\to B}$$

Jarzynski's equality

$$\left\langle \exp[-\beta W_{A\to B}] \right\rangle_A = \exp[-\beta \Delta F_{A\to B}]$$

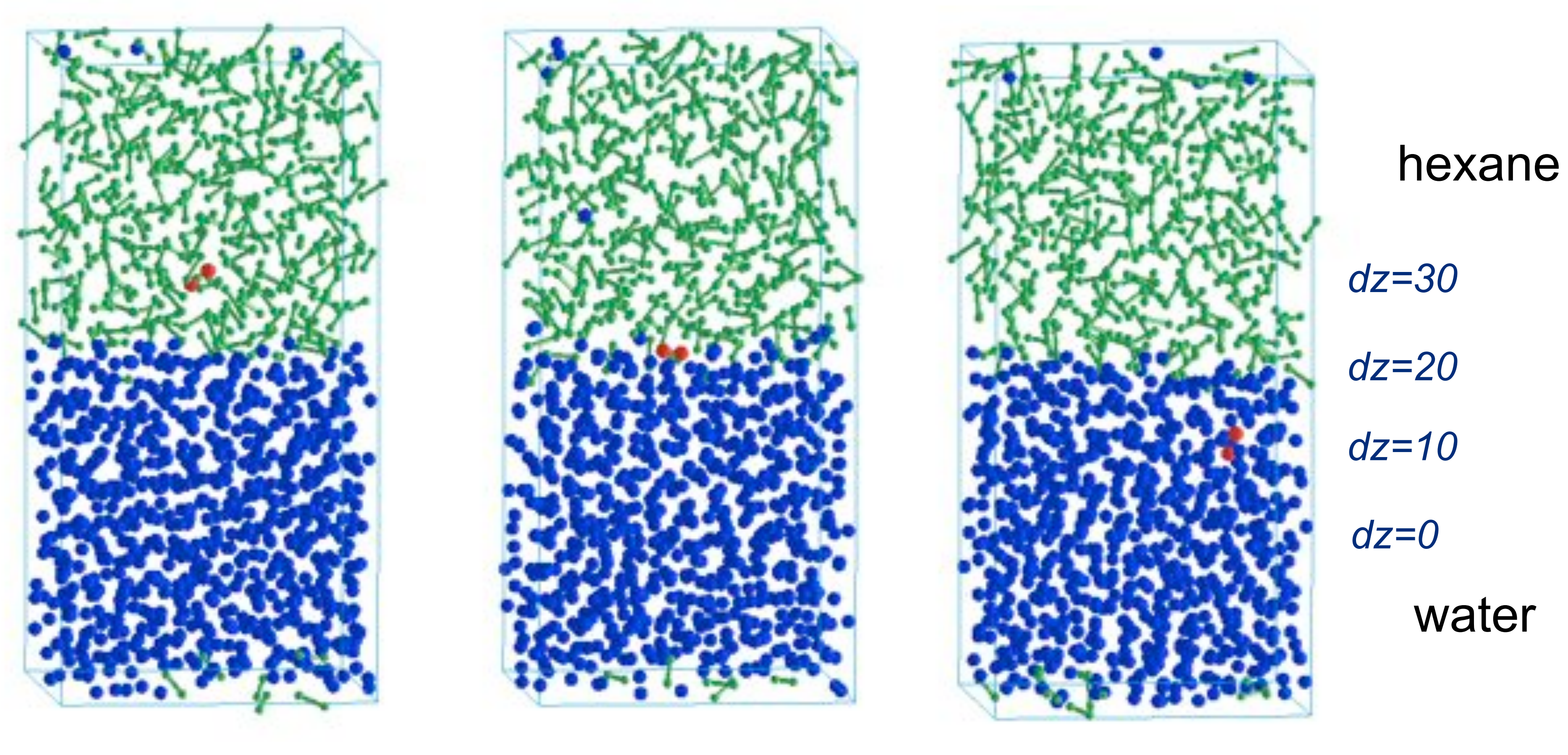
Surprisingly, we can obtain the equilibrium free energy difference from a non-equilibrium simulation, in which we force the system in a finite time to move from A to B.

Although, this may sound as a *free lunch* method, note that it requires sampling an exponential distribution of the work. For infinitely slow switching from A to B, the system is always in equilibrium so that a single simulation gives ΔF. But the faster the switching the more rare are the important low-work contributions to the average, so that many steered simulations are required for convergence.

Example

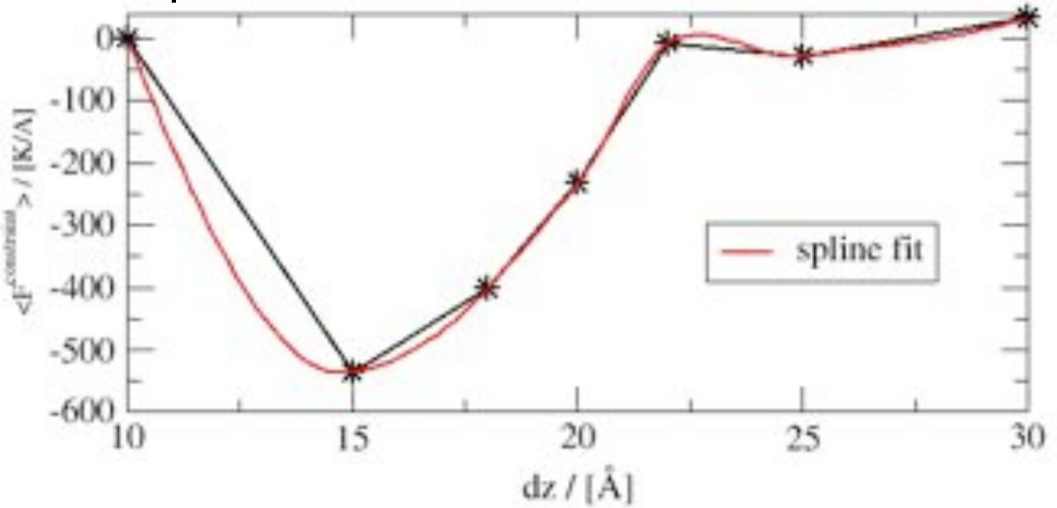
Relative solvation free energy of hexane molecule

The solvation free energy free difference of a solute in different solvents is here used a target property to parameterized a coarse-grain forcefield.

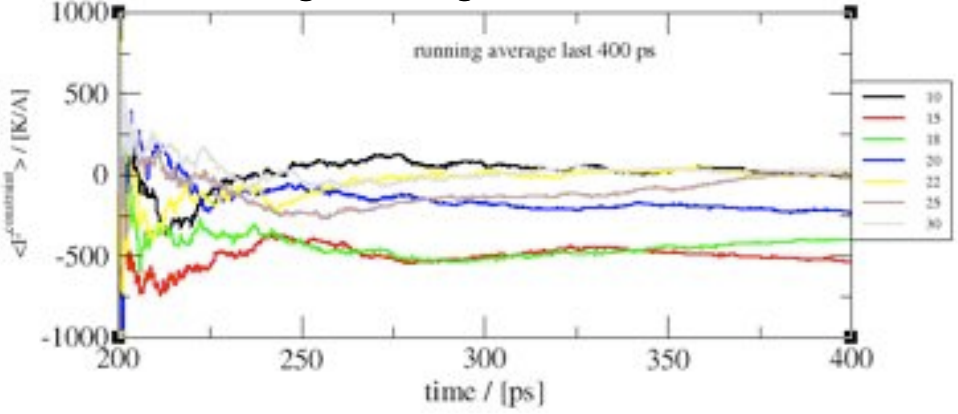


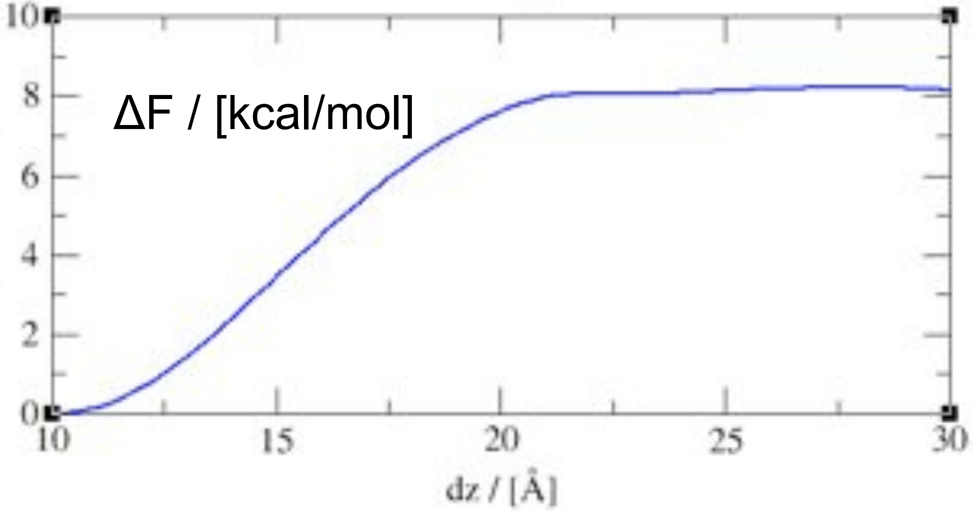
Constrained MD

Spline fit of constraint force measurements

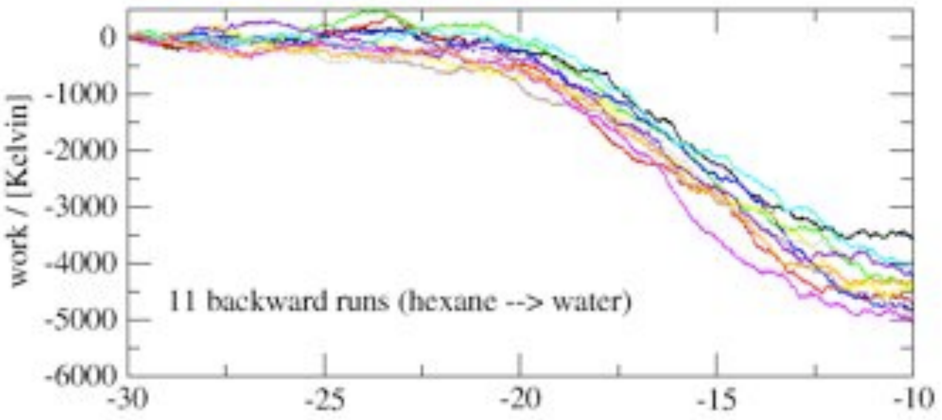


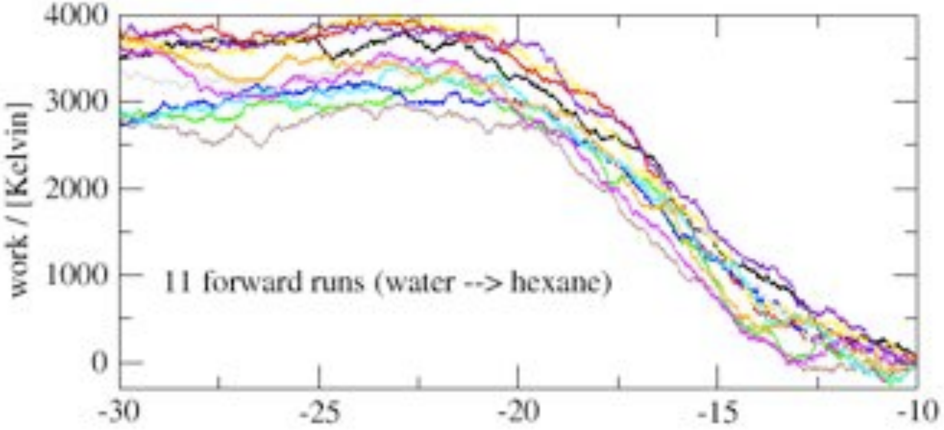
Running average constraint force

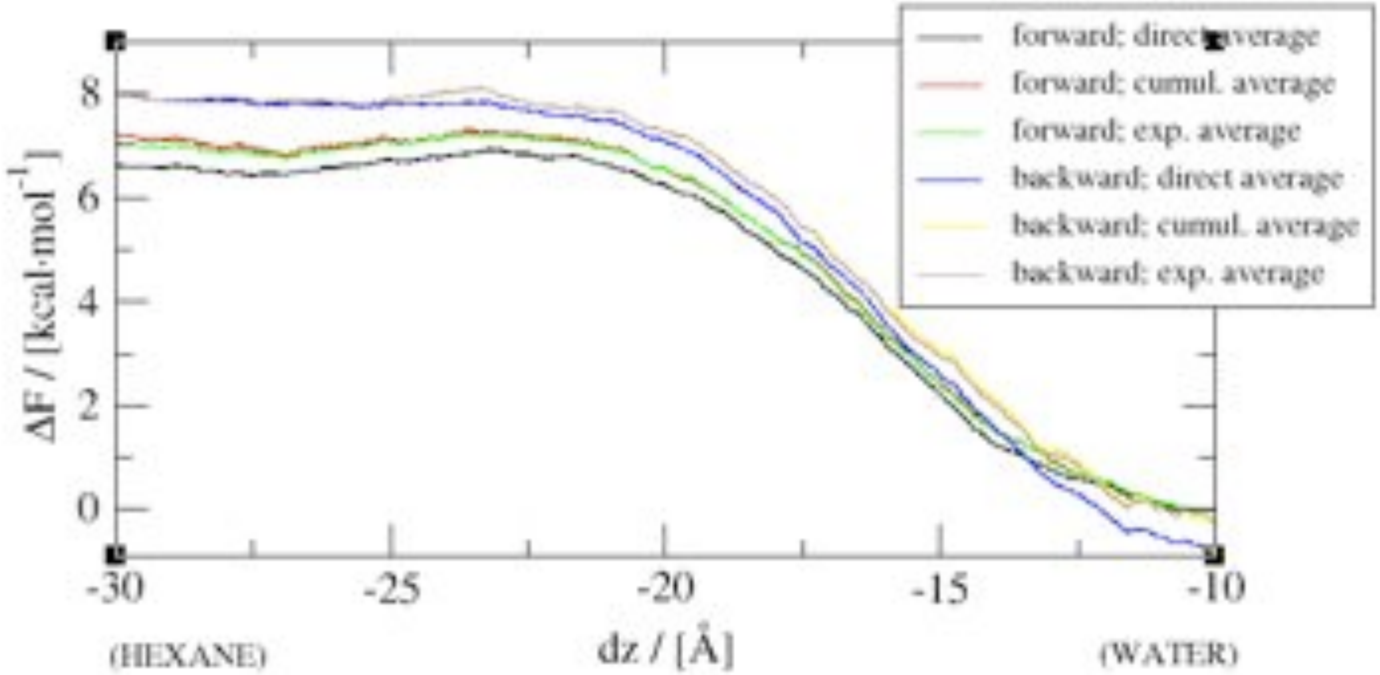




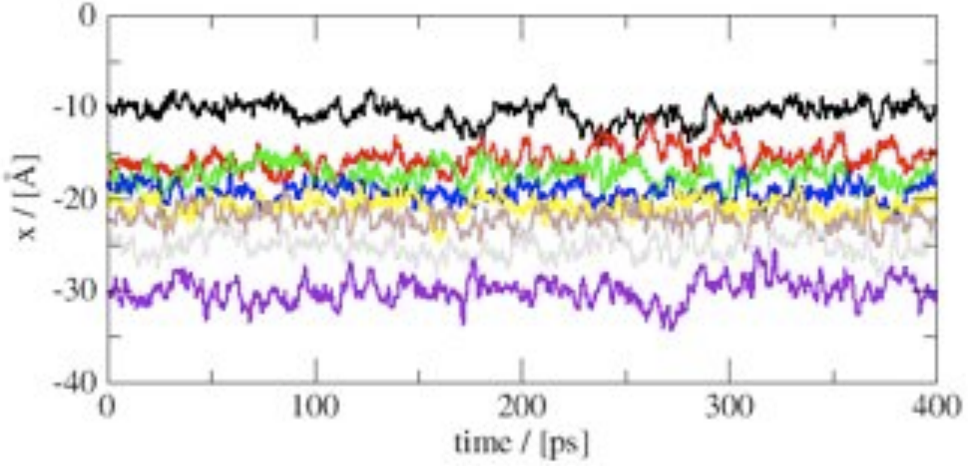
Steered MD

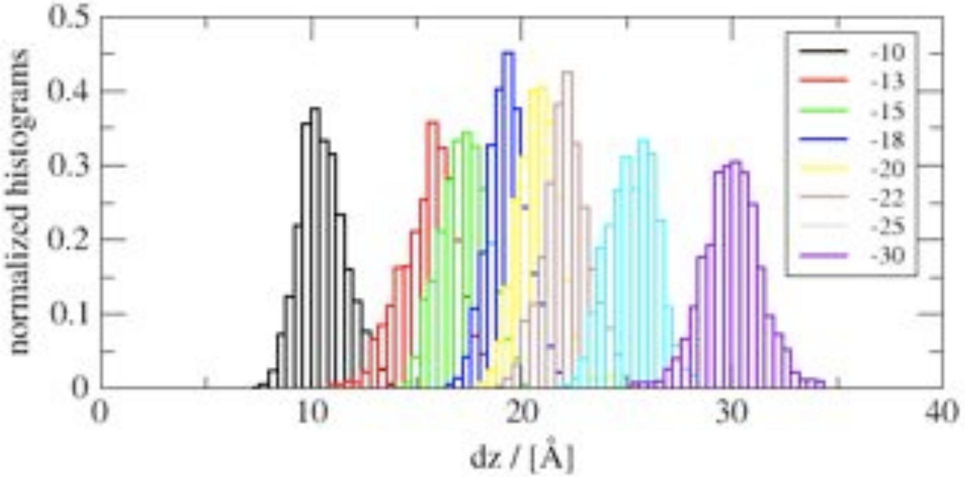

work / [Kelvin]
0
-1000
-2000
-3000
-4000
-5000
-6000
-30
-25
-20
-15
-10
11 backward runs (hexane --> water)

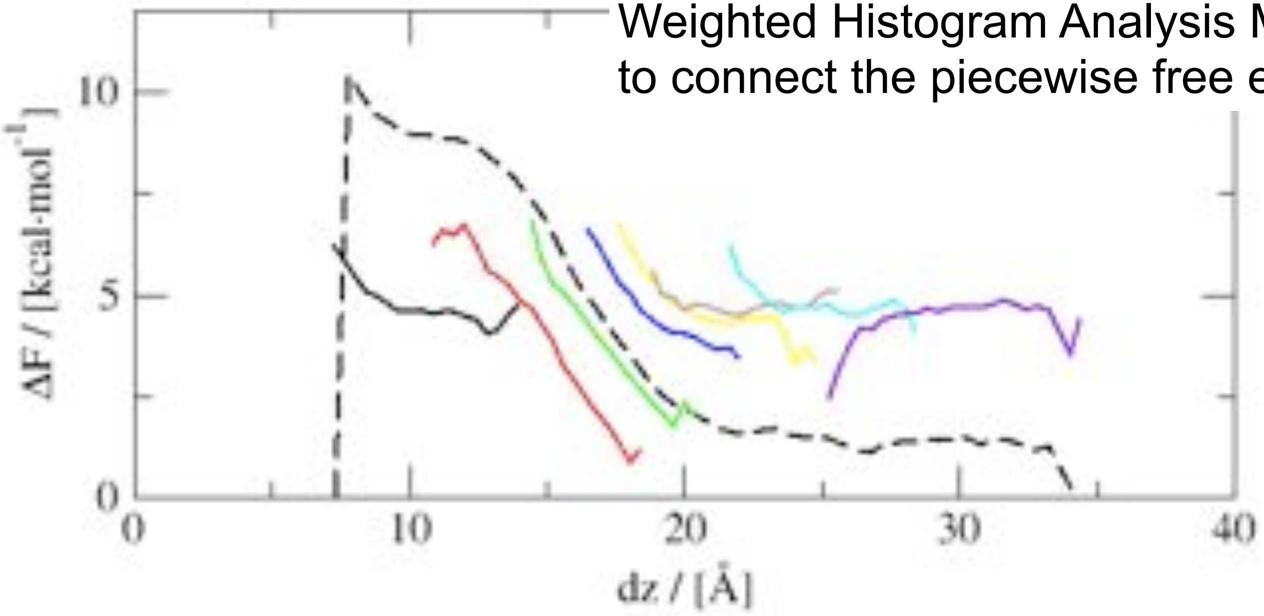

work / [Kelvin]
4000
3000
2000
1000
0
-30
-25
-20
-15
-10
11 forward runs (water --> hexane)


ΔF / [kcal·mol⁻¹]
8
6
4
2
0
-30
-25
-20
-15
-10
(HEXANE)
dz / [Å]
(WATER)
forward; direct average
forward; cumul. average
forward; exp. average
backward; direct average
backward; cumul. average
backward; exp. average

Umbrella Sampling





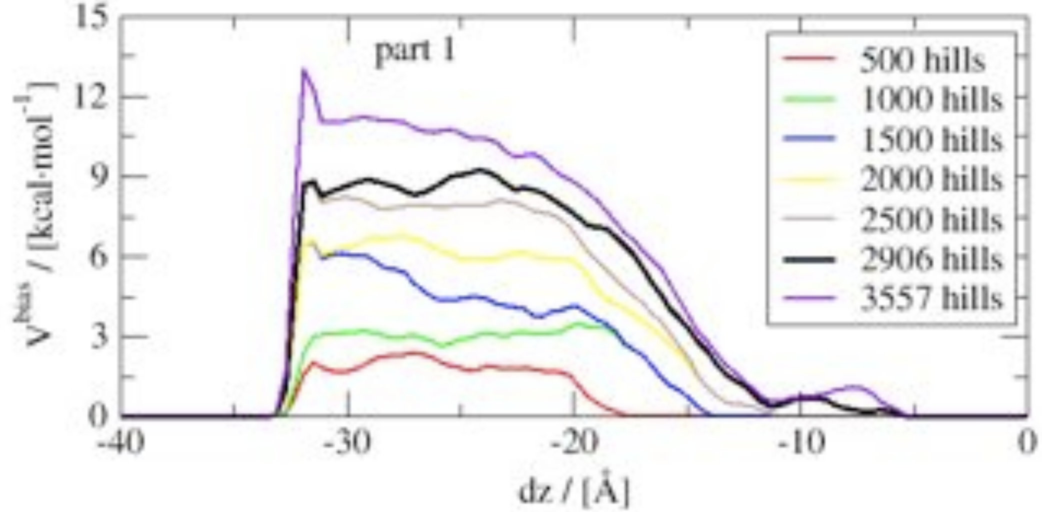


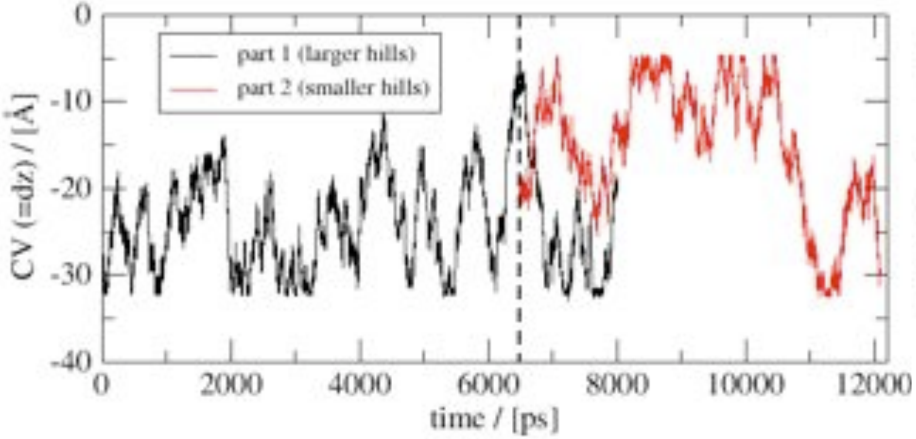
Weighted Histogram Analysis Method (WHAM)
to connect the piecewise free energy curves

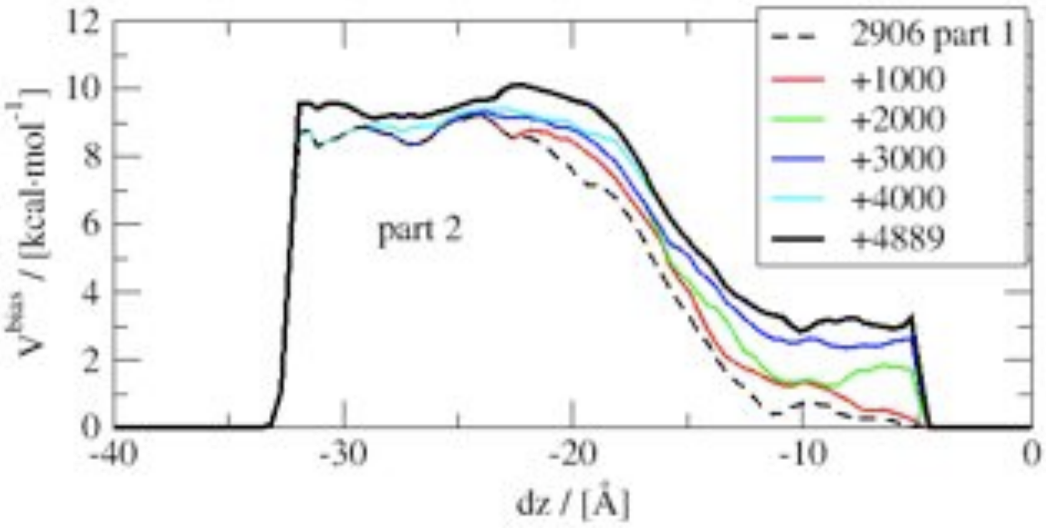
Metadynamics

Simulation in two parts with different size of the Gaussian “hills”:

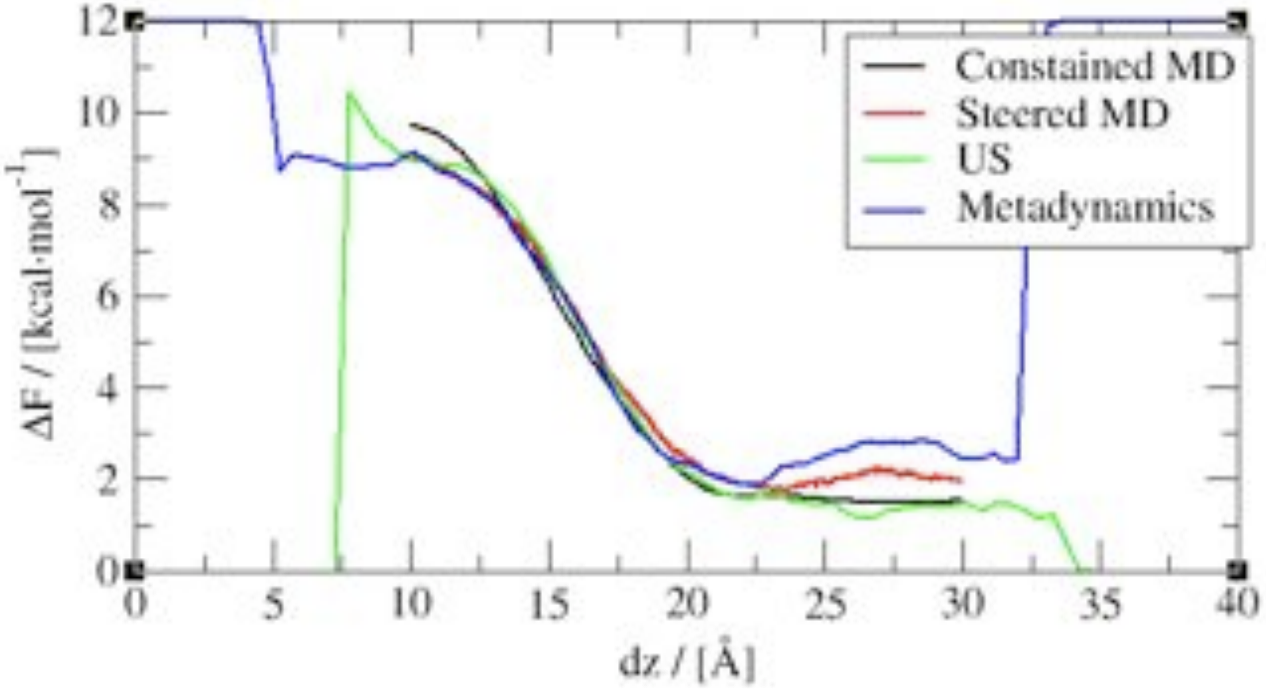
- Part 1: H=0.25 K, W=0.4 Å
- Part 2: H=0.10 K, W=0.2 Å







Comparisson



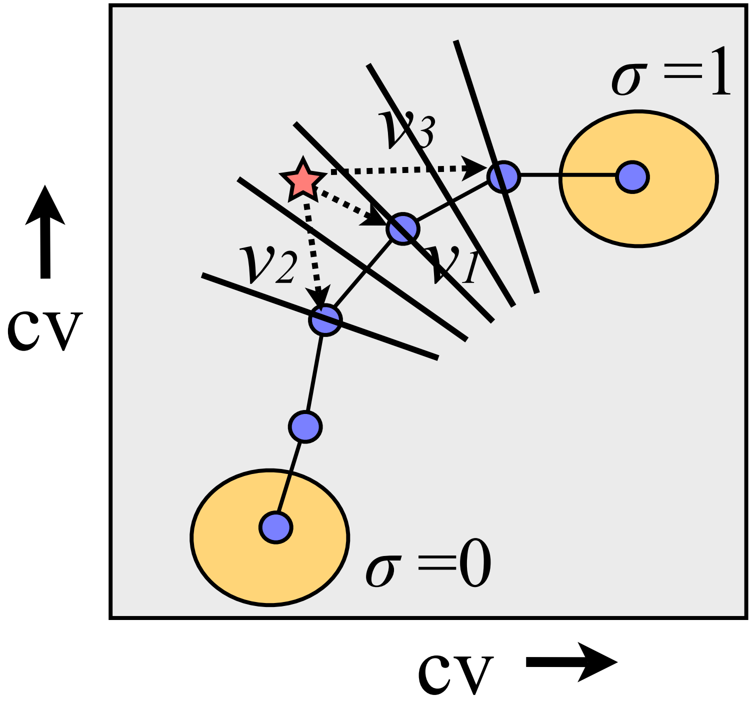
Path-metadynamics

Combine metadynamics with reaction path finding methods such as the “string method” or the “nudge elastic band” method

Start with a guess reaction path from A to B.
While biasing the system along the guess path, evolve the path to find the minimal free energy path

- add extra collective variable: σ
- σ is a function of all other coll. variables
- biasing potential is only working on σ
- σ function adapts on the fly

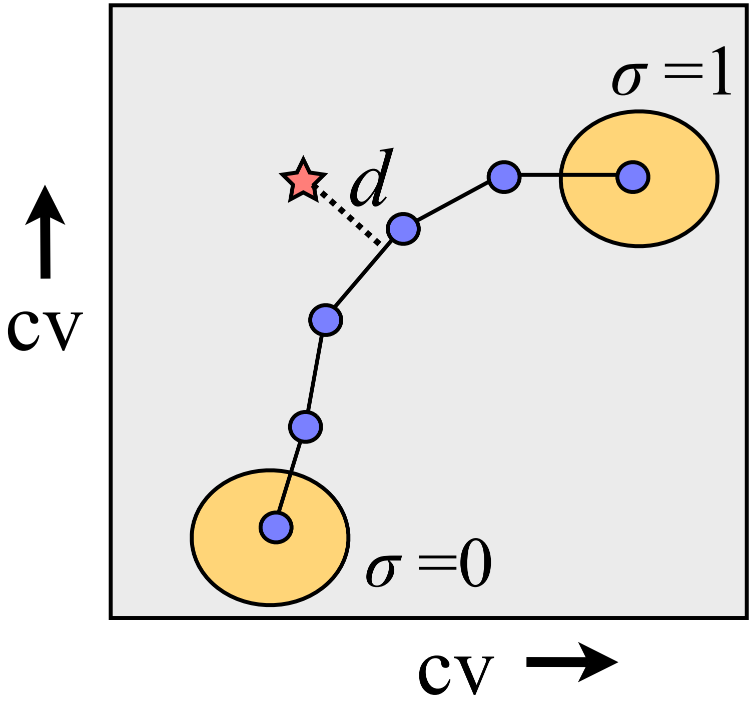
path as the collective variable



path CV

$$\sigma = n^{-1}\left(i_{min} \pm \frac{\sqrt{(\bar{v}_1\bar{v}_3)^2 - \bar{v}_3\bar{v}_3(\bar{v}_1\bar{v}_1 - \bar{v}_2\bar{v}_2)} - \bar{v}_1\bar{v}_3}{2(\bar{v}_3\bar{v}_3 - 1)}\right)$$

path as the collective variable



histogram distance to the path for each node
maximum of *P(d)* should be at *d*=0

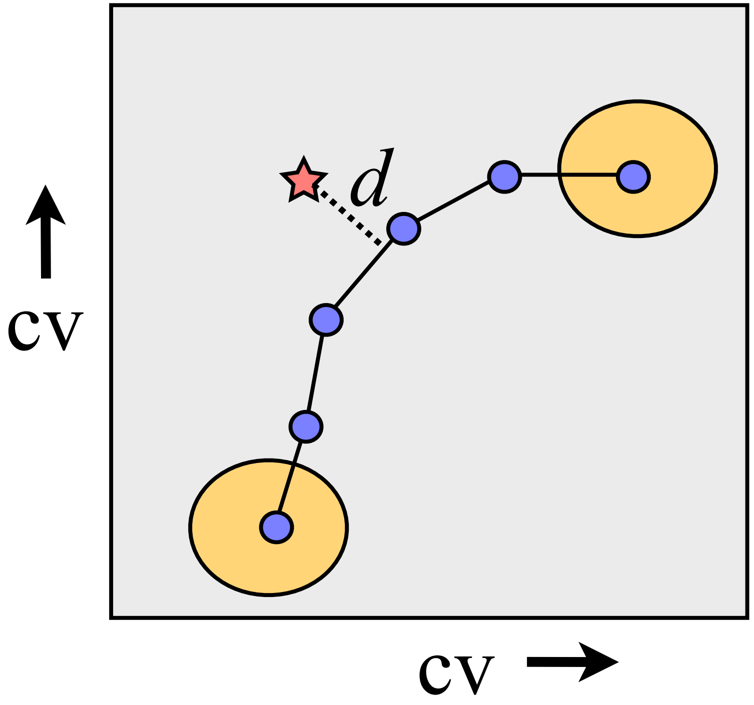
or accumulate average *d for each node*

path as the collective variable

$$< d_i >= \frac{\sum_0^\tau w_i * d_i}{\sum_0^\tau w_i}$$

$$w_i(t) = (\sigma(n-1) - i_{min}) f^{\tau - t}$$

$$f = \exp[\tau^{-1} * \ln \frac{1}{2}]$$



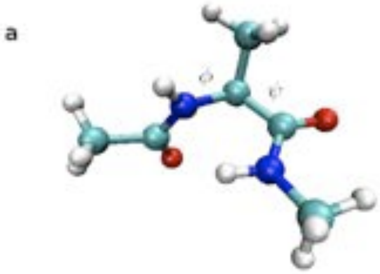
move nodes to *d* (every step)
set *d*=0
redistribute nodes along path

- Zhiwei Liu
- Preston Moore

University of the Sciences, Philadelphia

Alanine dipeptide Ramachandran plot

a



b

Classical Molecular Dynamics
(CM3D code by Preston Moore, USP)

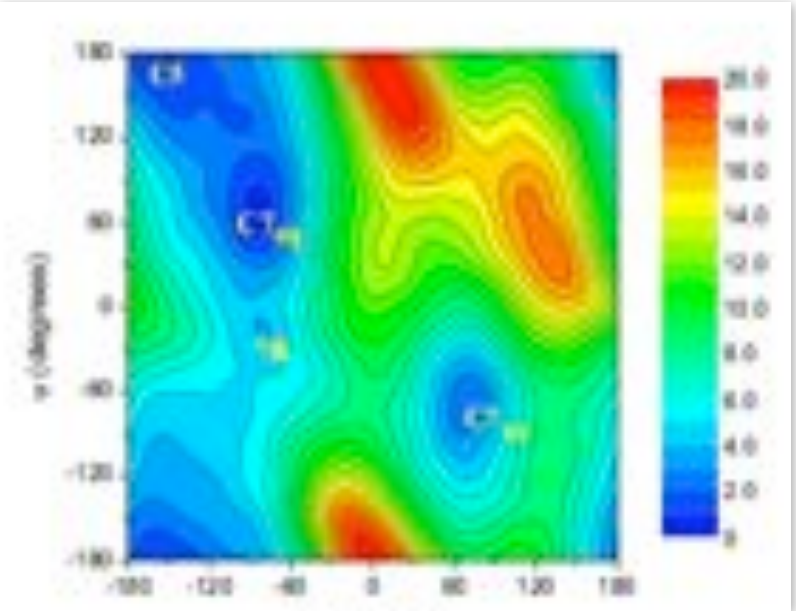
Alanine dipeptide + 216 water
CHARMM27 forcefield (modified internal params water)
Cubic periodic box L=18.8 Angstrom
NVT ensemble, T=298K

Hill size: W=0.2 rad, H=0.02 kcal/mol
Hill stride: 100 MD steps (=50 fs)

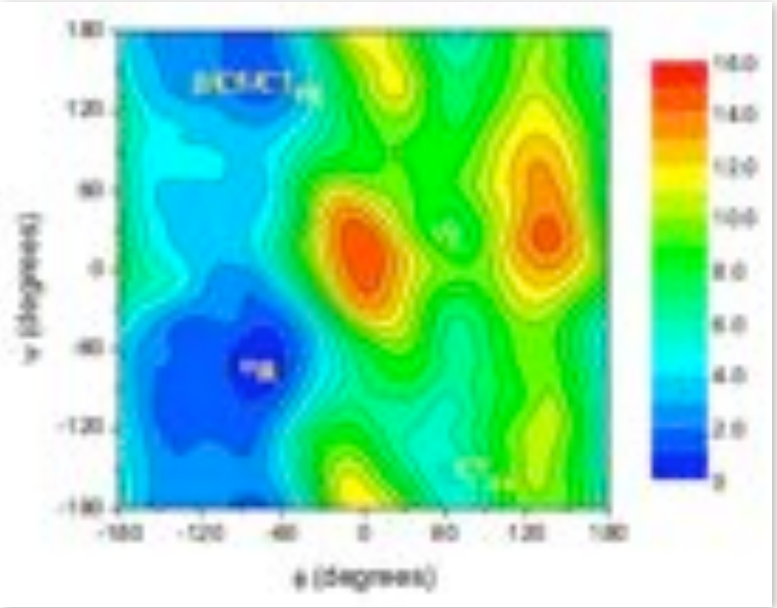
Total simulation time: ca. 5 ns

Error in energy: 0.3-0.5 kcal/mol
Error in angles: 3-10 degrees

Gas-phase



Aqueous solution



Bernd Ening, M DeVivo, ZW Liu, PB Moore, ML Klein, *Acc. Chem. Res.* 2006, 39, 73.

Alanine dipeptide

Parameters:

T = 300 K

$H_{gaussian}$ = 10 K

$W_{gaussian}$ = 0.05

$\Delta t_{gaussian}$ = 100 MD steps

n = 20 + 20 + 20

τ = 1000 MD steps

every recrossing:

H x 50%

W x 50%

τ x 100