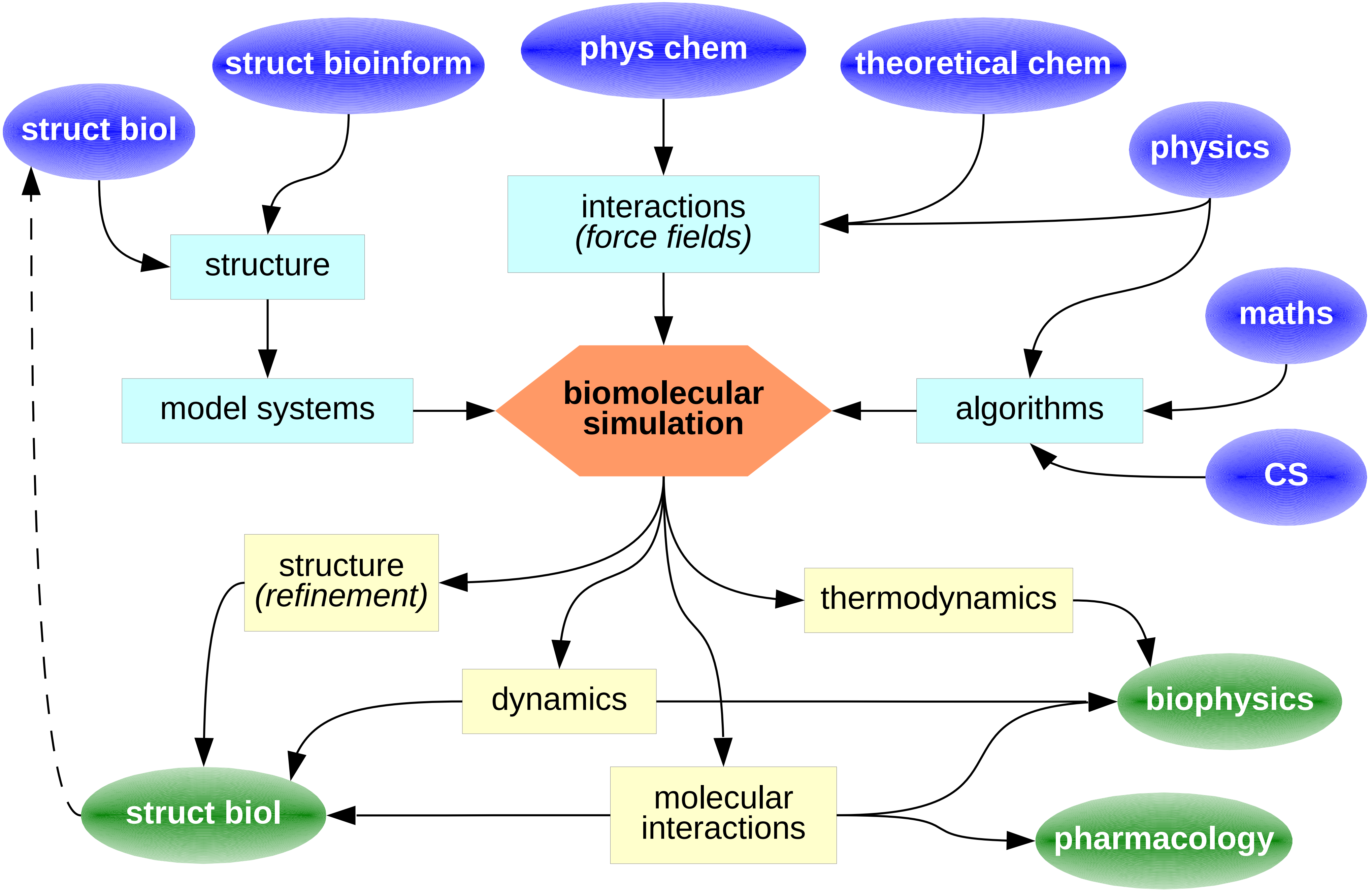
Sampling the free energy surfaces of collective variables

Jérôme Hénin

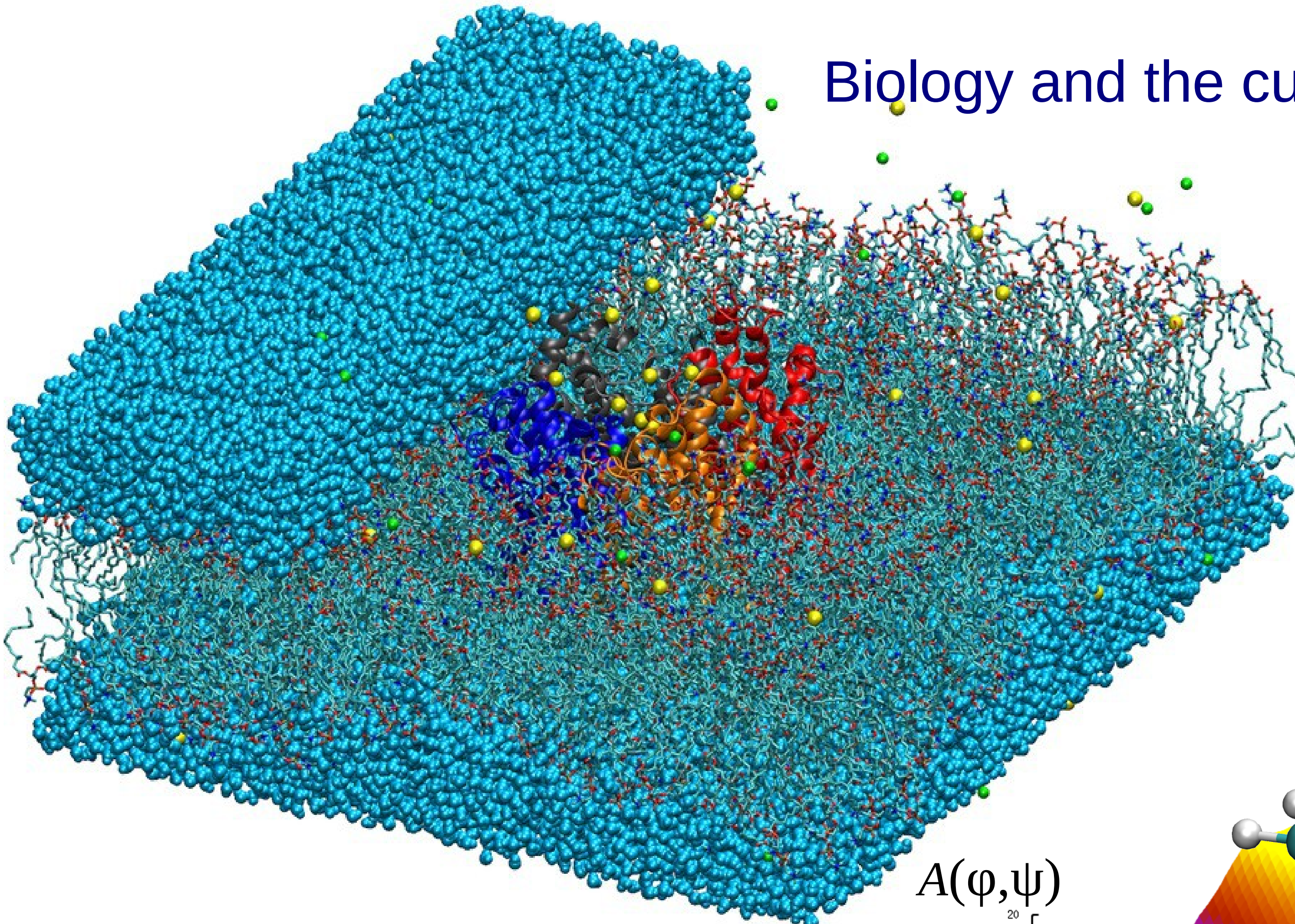
Enhanced Sampling
and Free-Energy Calculations
Urbana, 12 September 2018

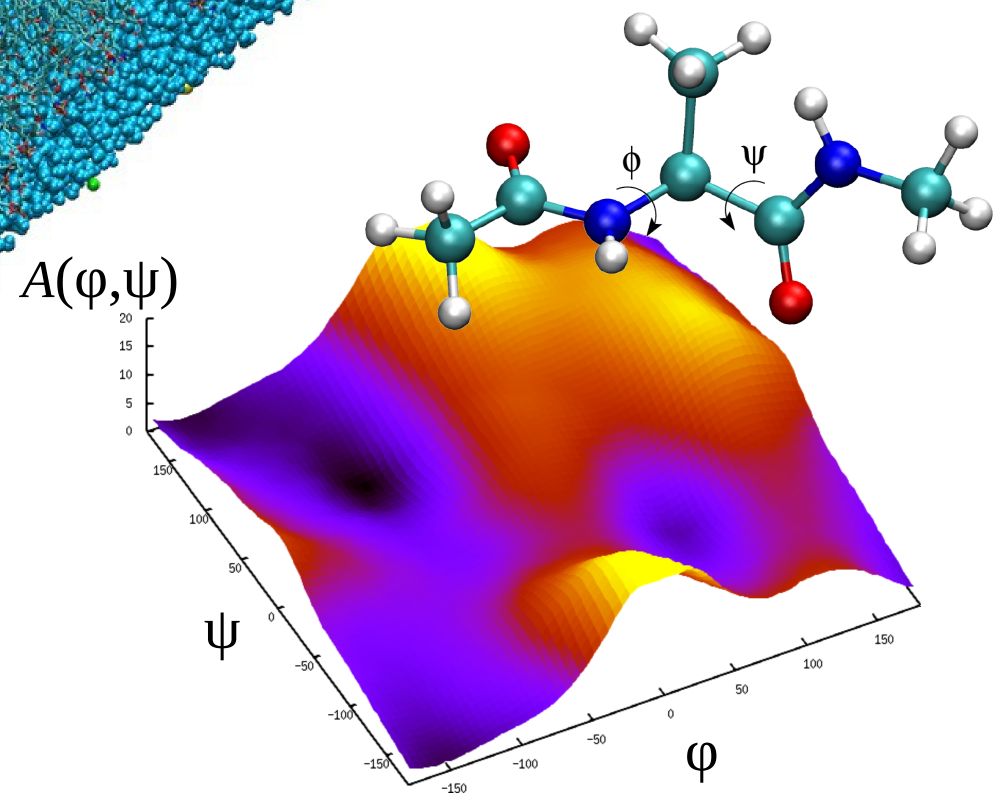
Please interrupt!



struct bioinform
phys chem
theoretical chem
struct biol
physics
interactions
(force fields)
structure
maths
model systems
biomolecular
simulation
algorithms
CS
structure
(refinement)
thermodynamics
dynamics
biophysics
molecular
interactions
struct biol
pharmacology

Biology and the curse of dimensionality





we need reduced representations made of few selected coordinates

- for human intuition
- for importance sampling

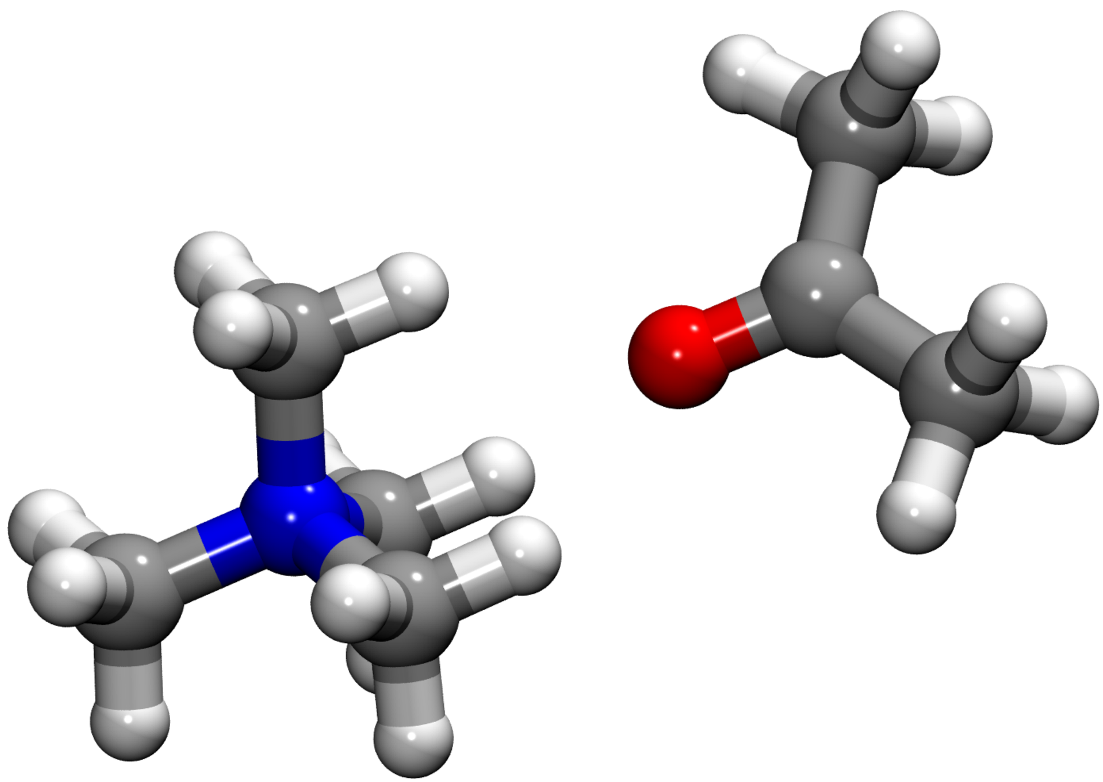
Outline

- Free energy
- Collective variables
- Free energy landscapes

- Methods to compute (estimate) FE landscapes
  - from probability distribution (histograms)
  - from forces (thermodynamic integration)
  - from adapted biasing potential (metadynamics)

- Methods to sample FE landscapes
  - umbrella sampling
  - metadynamics : adaptive biasing potential
  - adaptive biasing force

Tetramethylammonium – acetone binding



Free energy

- **free energy differences ↔ probability ratios**

$$\Delta F_{AB} = F_B - F_A = -k_B T \ln\left(\frac{P_B}{P_A}\right)$$

$$\Delta F_{AB} > 0 \iff P_B < P_A$$

$$\Delta F_{AB} = 0 \iff P_B = P_A$$

- ***macrostates*** (A, B) are collections of ***microstates*** (atom coordinates *x*)
- →probabilities of macrostates are sums (integrals) over microstates

$$P_A = \int_A p(x)dx$$

- probabilities of microstates follow ***Boltzmann distribution***

$$p(x) = \frac{\exp(-V(x)/k_B T)}{Z} = \frac{1}{Z}e^{-\beta V(x)}$$

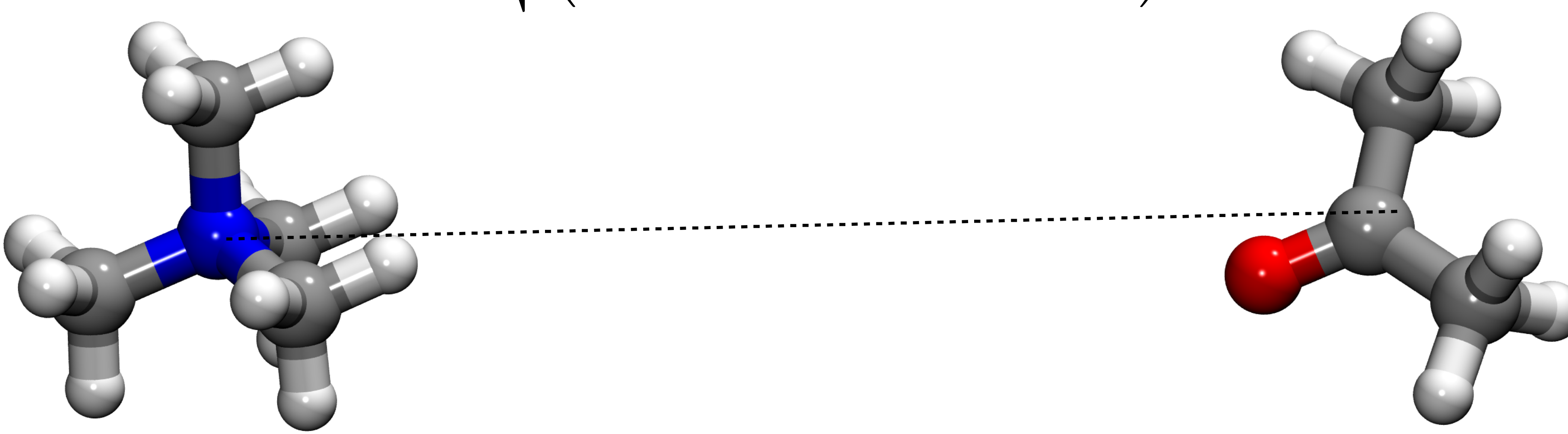
Collective variables

- geometric variables that depend on the positions of several atoms (hence "collective")
- mathematically: functions of atomic coordinates $z = \xi(x_i, y_i, z_i ...)$
- example: distance between two atoms

$$d_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} = \sqrt{(r_2 - r_1)^2}$$

- distance between the centers of mass of groups of atoms G_1, G_2

$$d_{G1,G2} = \sqrt{\left(\frac{\sum_{i \in G_2} m_i r_i}{\sum_{i \in G_2} m_i} - \frac{\sum_{i \in G_1} m_i r_i}{\sum_{i \in G_1} m_i} \right)^2}$$



Probability distribution of a collective variable

- we know the 3N-dimensional probability distribution of atom coordinates *x:*

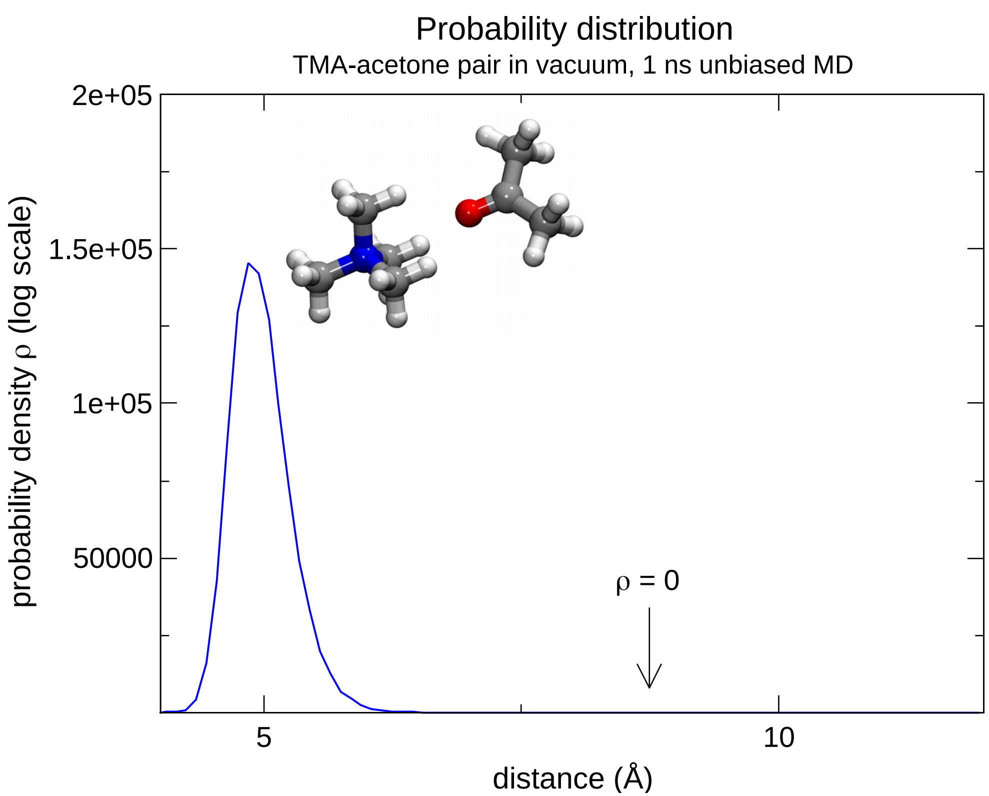
$$p(x) = \frac{\exp(-V(x)/k_B T)}{Z}$$

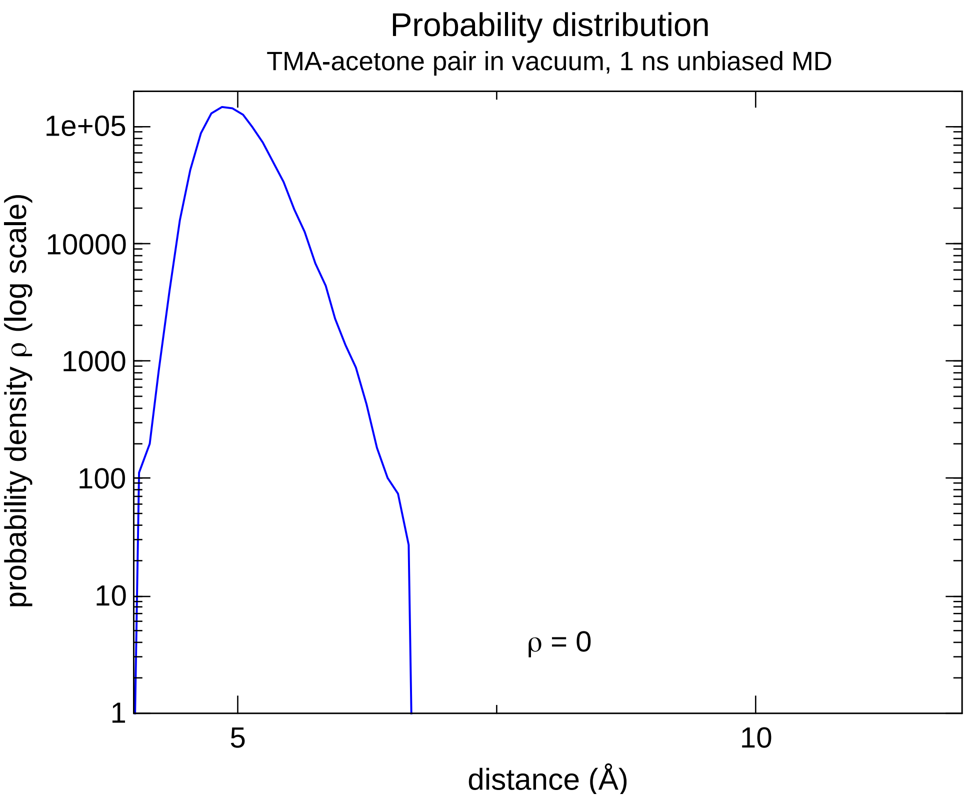
- what is the probability distribution of $z = \xi(x)$
- theory: sum (integral) over all the values of *x* corresponding to a value of *z*

$$\rho(z) = \int p(x)\delta(\xi(x) - z)dx$$

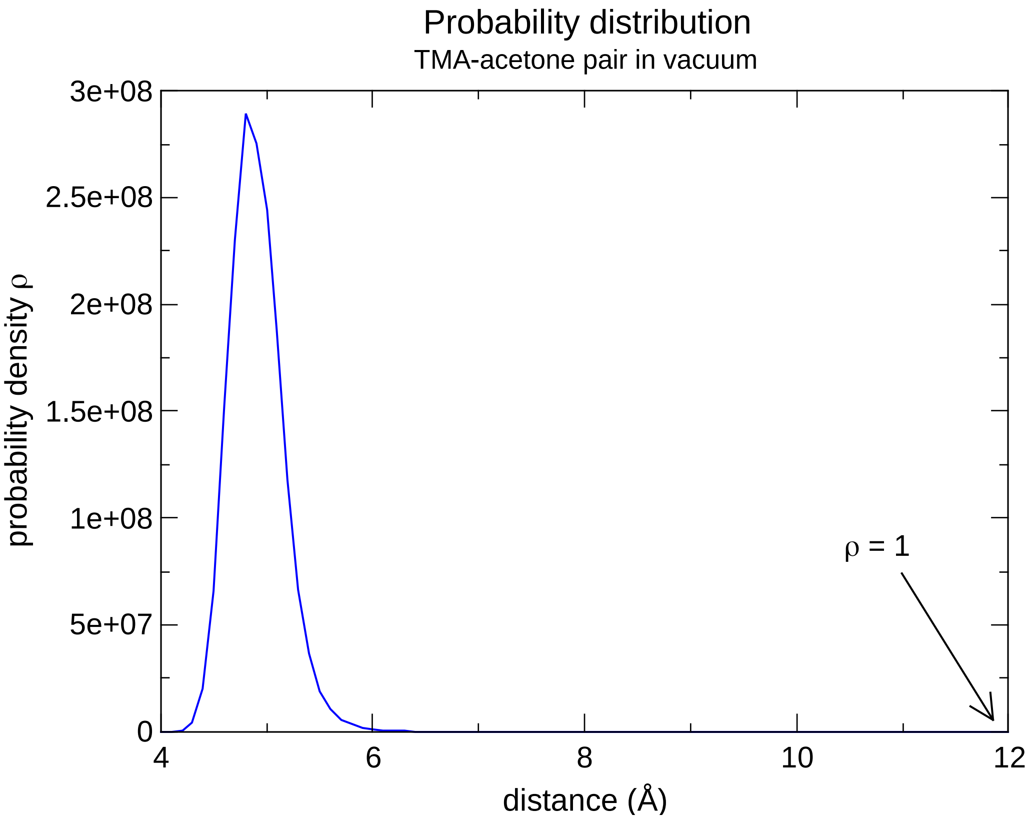
- in simulations: sample and calculate a **histogram** of coordinate *z*

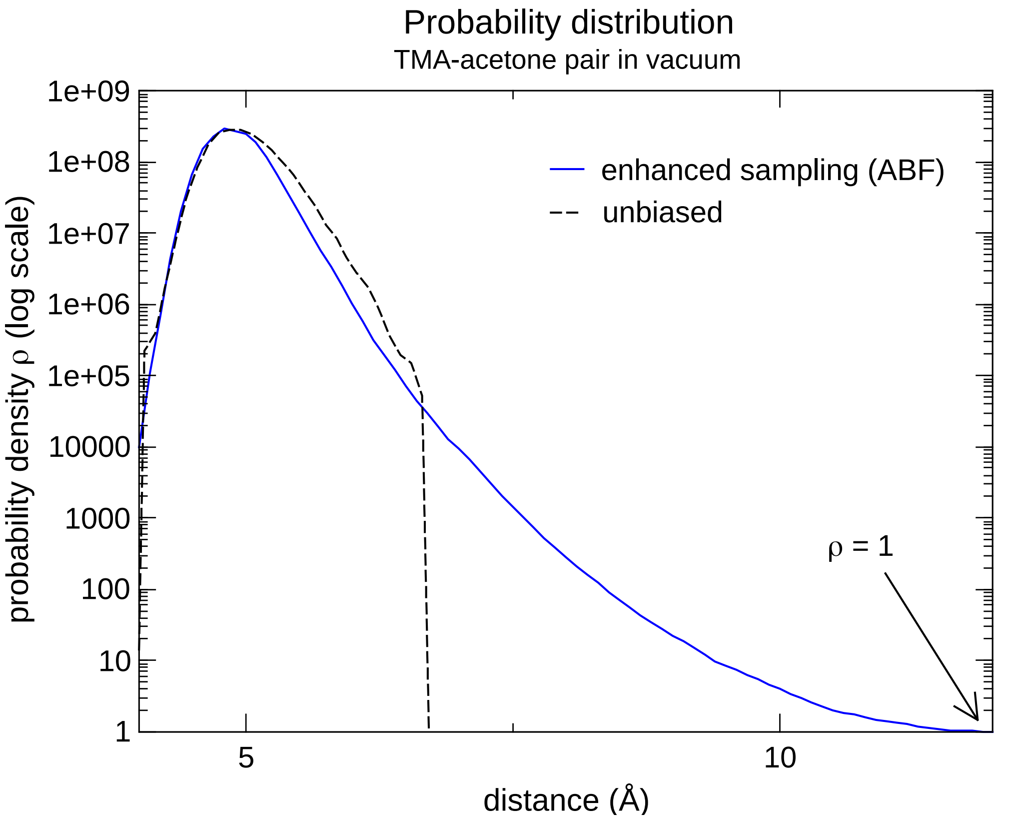
Probability distribution of a collective variable
(1) from unbiased simulation



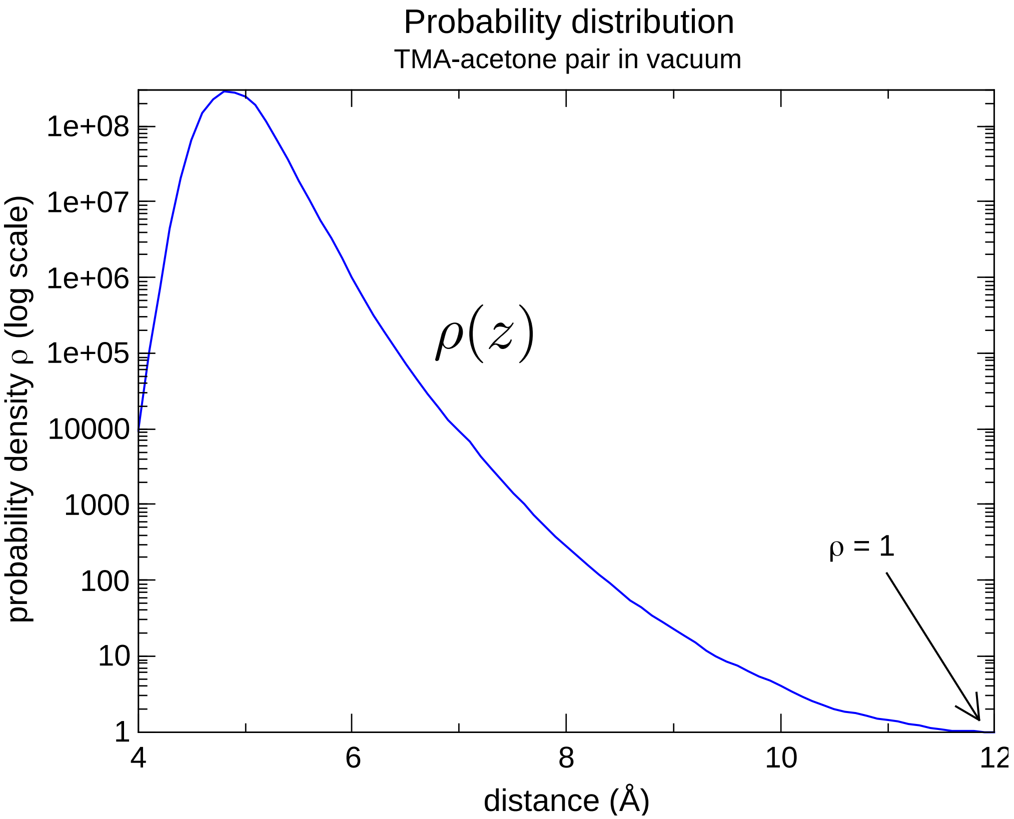


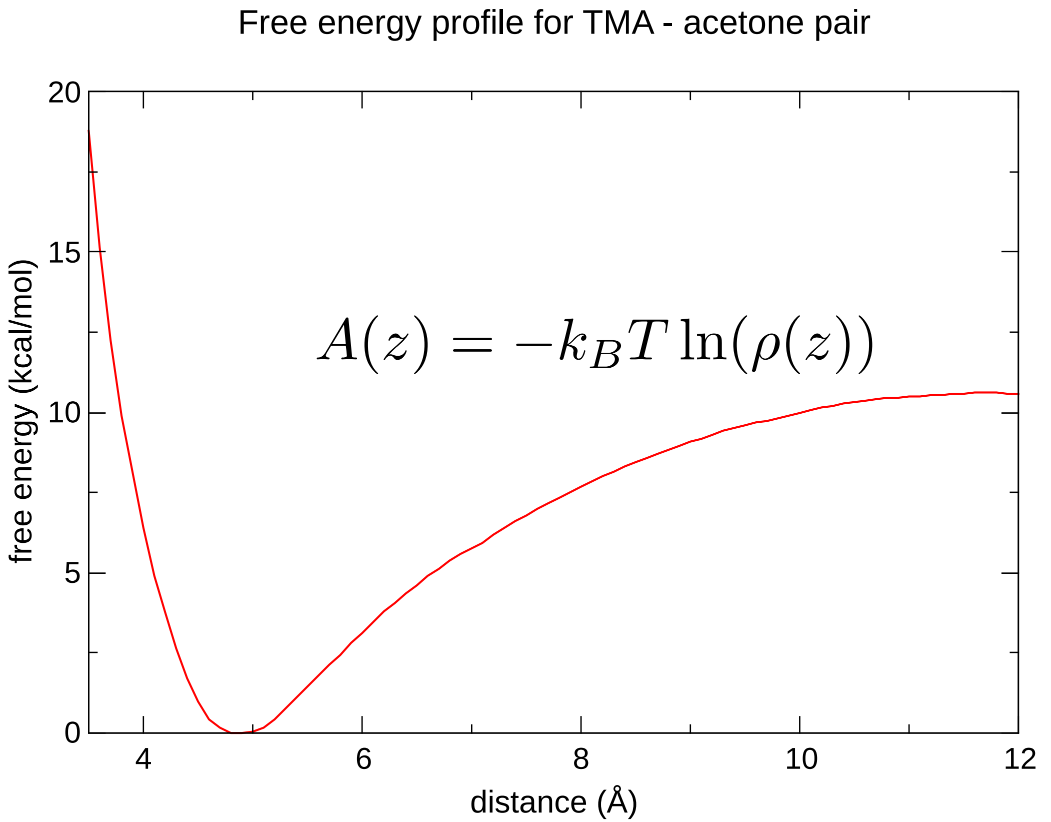
Probability distribution of a collective variable (2) with enhanced sampling





From probability to free energy





Ways to calculate the free energy

- from unbiased histogram $A(z) = -k_B T \ln(\rho(z)) + C$

- from biased histogram (*importance sampling*) with bias $V^{bias}(z)$

$$A(z) = -k_B T \ln(\rho(z)) - V^{\mathrm{bias}}(z) + C$$

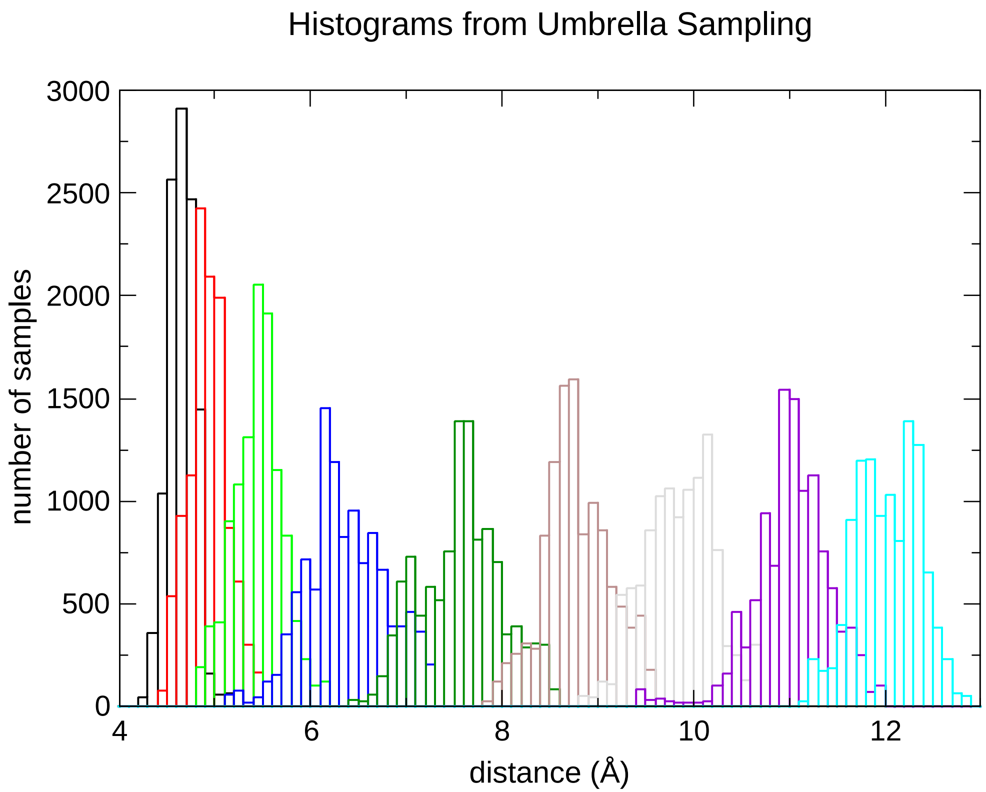
  - in Umbrella Sampling, need to find values of *C!*

- estimate and integrate free energy derivative (gradient): *Thermodynamic Integration*

$$A(z) = \int_0^z A'(z^*) dz^* + A(0)$$

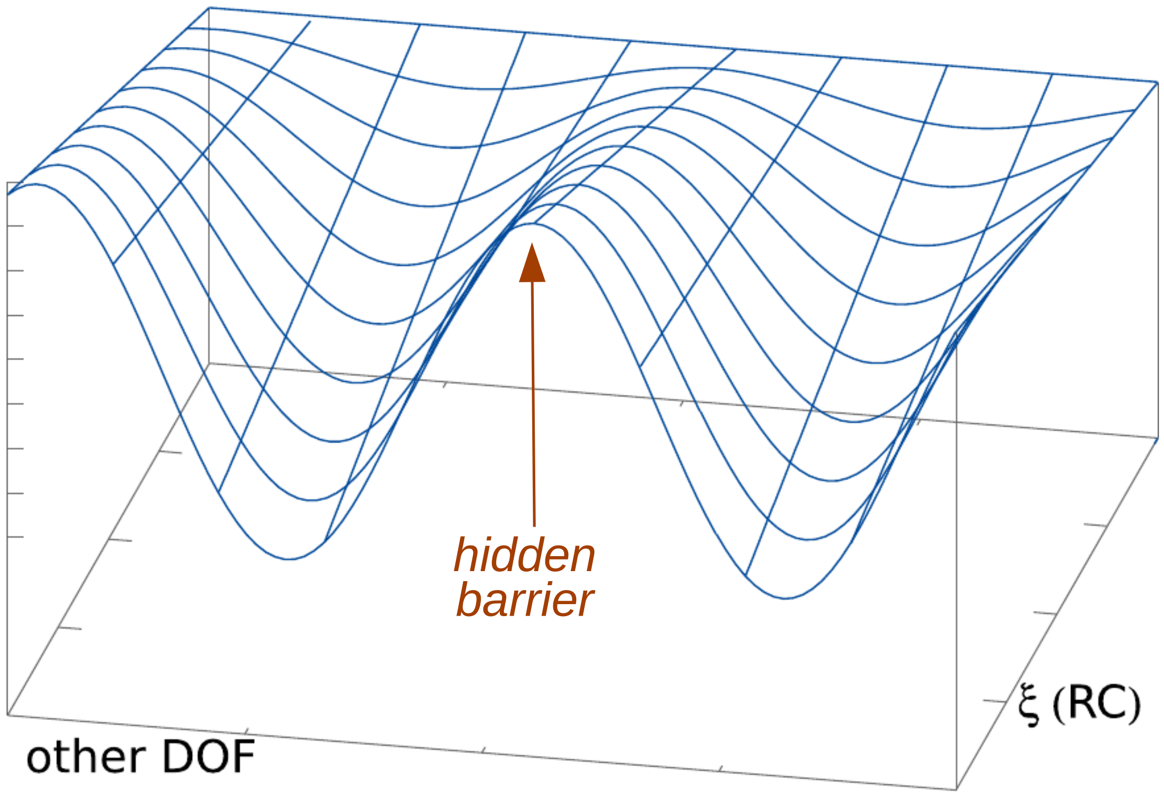
$$A'(z) = \left\langle \frac{\partial V}{\partial z} - k_B T \frac{\partial \ln |J|}{\partial z} \right\rangle_{\xi(x)=z}$$

Umbrella sampling

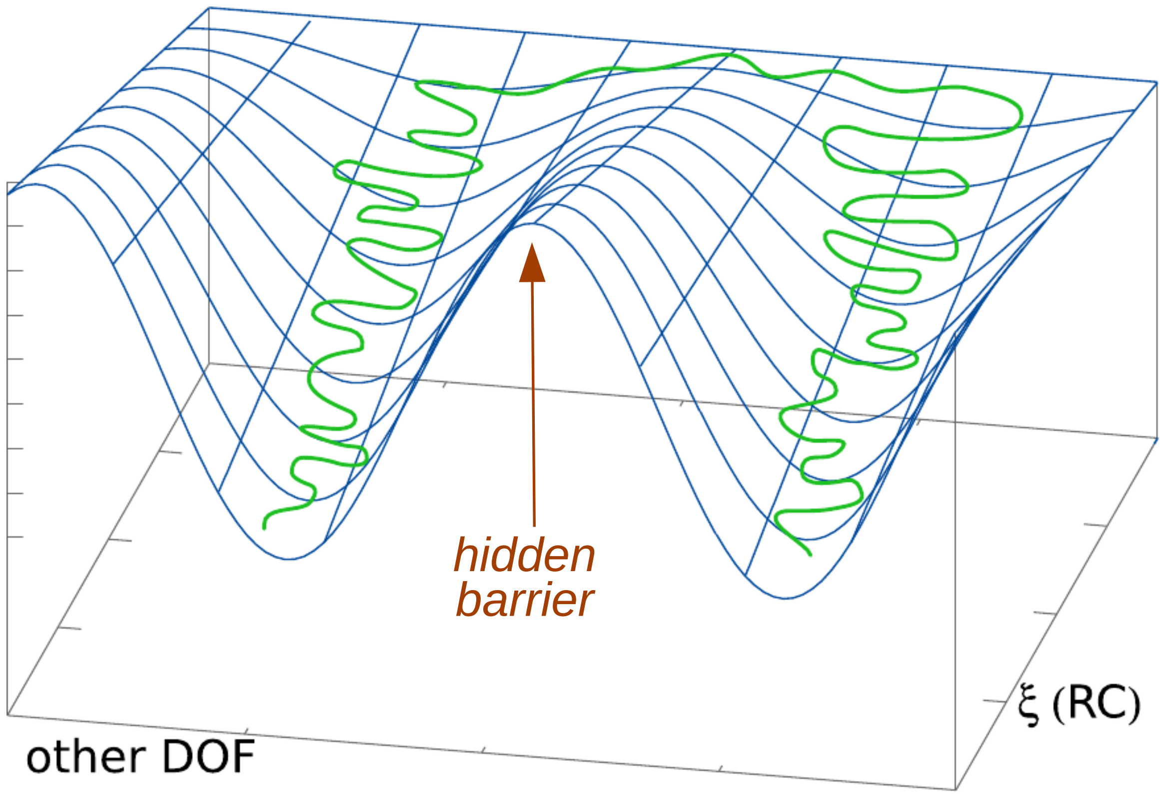


- distribute (*stratify*) sampling using multiple confinement restraints
- combine partial information of each histogram by computing relative free energies
  - WHAM (weighted histogram analysis method)
  - MBAR (multistate Bennett's acceptance ratio)
- requires overlap between sampling in adjacent windows

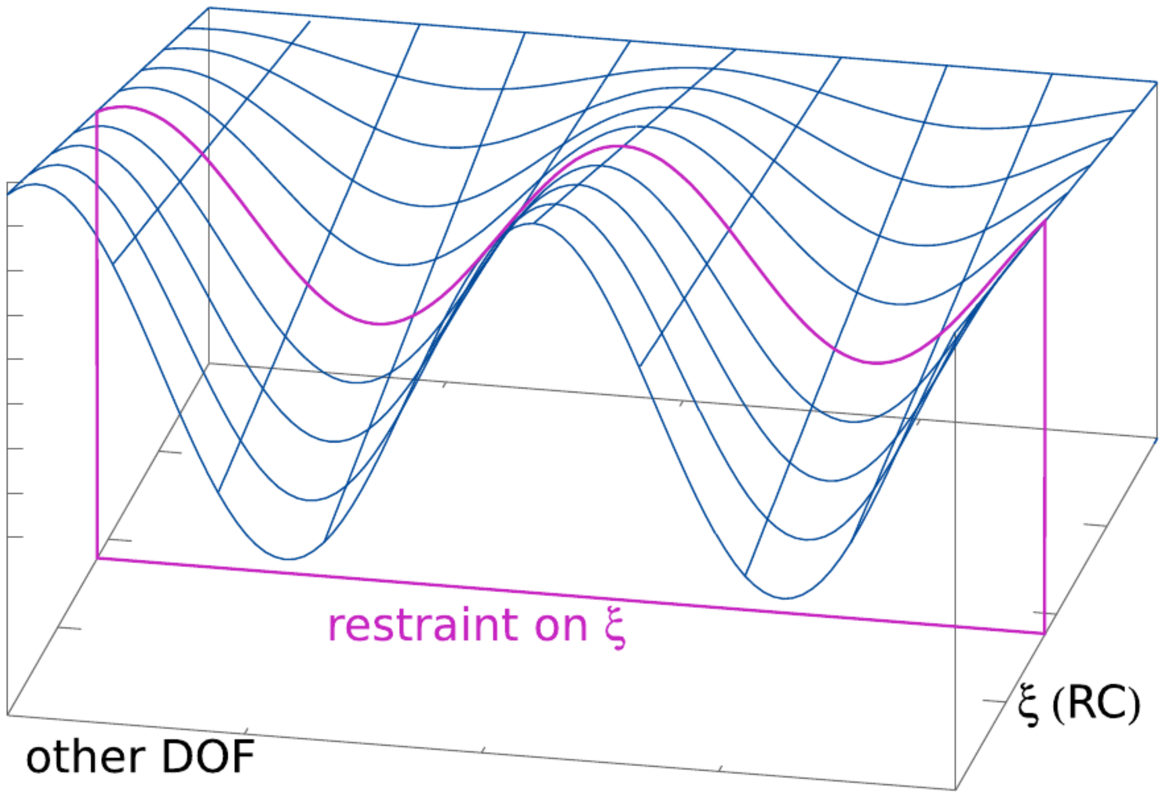
Multi-channel free energy landscape



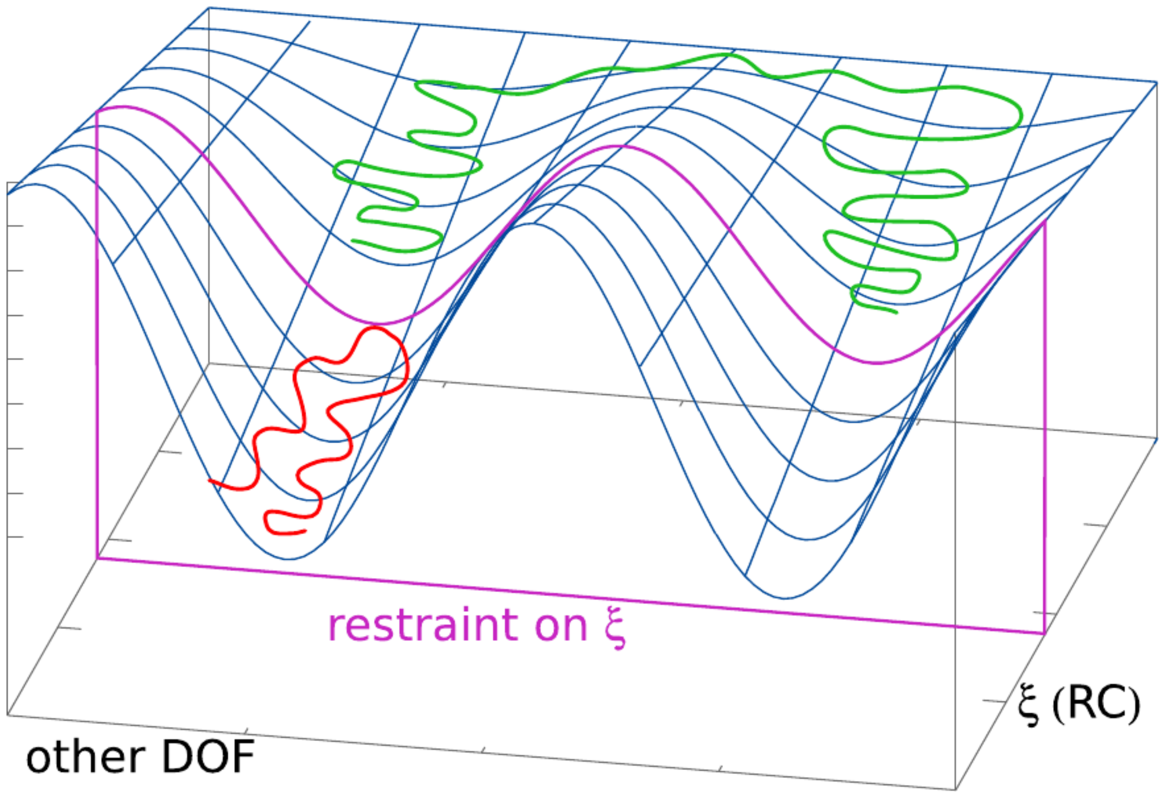
Multi-channel free energy landscape



Umbrella Sampling: *stratification*

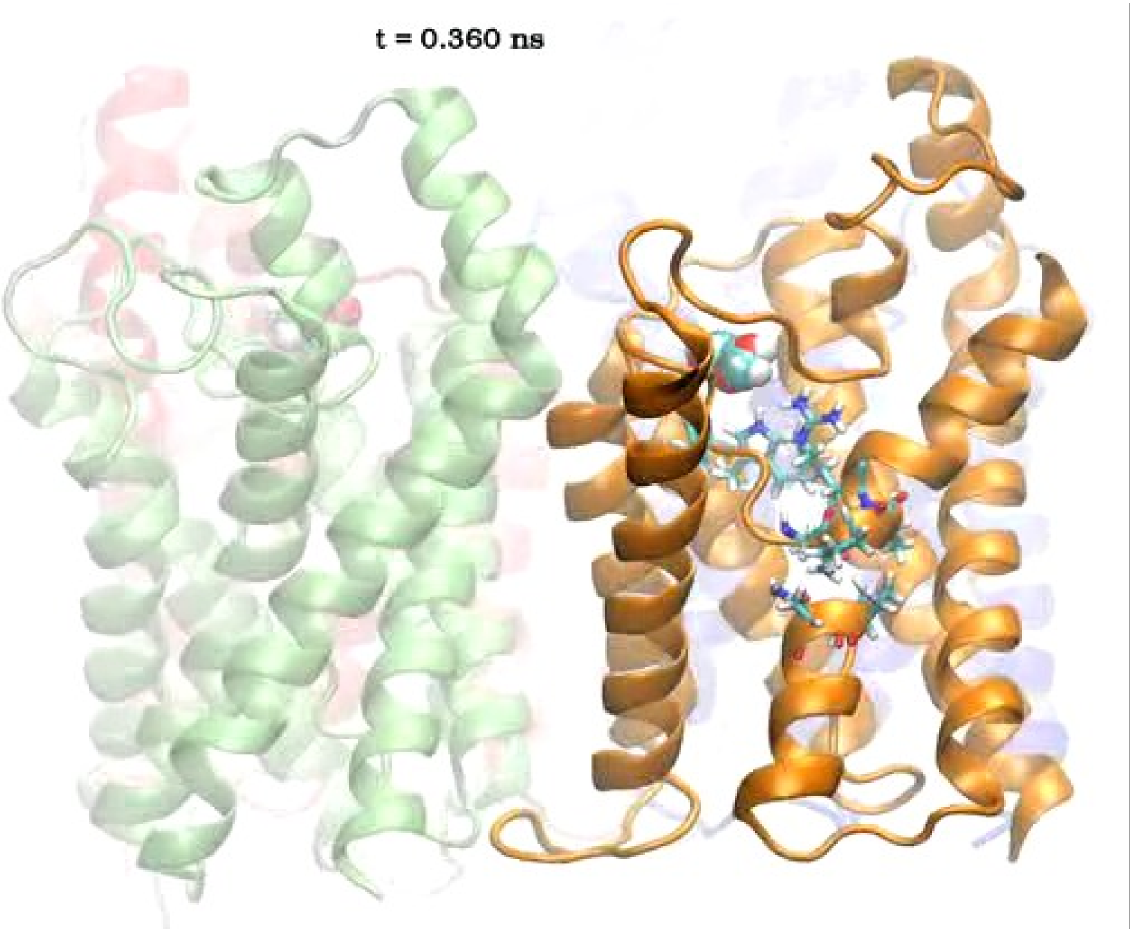


Umbrella *Sampling or Not Sampling?*



benefit of adaptive sampling methods: no stratification needed

Orthogonal relaxation in ABF



Hénin, Tajkhorshid, Schulten & Chipot, Biophys J. 2008

Adaptive sampling 1: adaptive biasing potential

Free energy profile *A(z)* is linked to distribution of transition coordinate:

$$e^{-\beta A(z)} \propto \rho(z) \propto \int e^{-\beta V(x)} \delta(\xi(x) - z)\, dx$$

ABP: time-dependent biased potential

$$\tilde{V}_t(x) = V(x) - A_t(\xi(x))$$ where A_t converges to A

Long-time biased distribution: $\tilde{\rho}_\infty(z) \propto e^{-\beta(A(z) - A_\infty(z))}$

that is, a **uniform distribution**.

Adaptive Biasing Potential : Metadynamics

- adaptive bias is sum of Gaussian functions created at current position
- pushes coordinate away from visited regions
- convergence requires careful tuning of time dependence of the bias (“well-tempered” metadynamics)

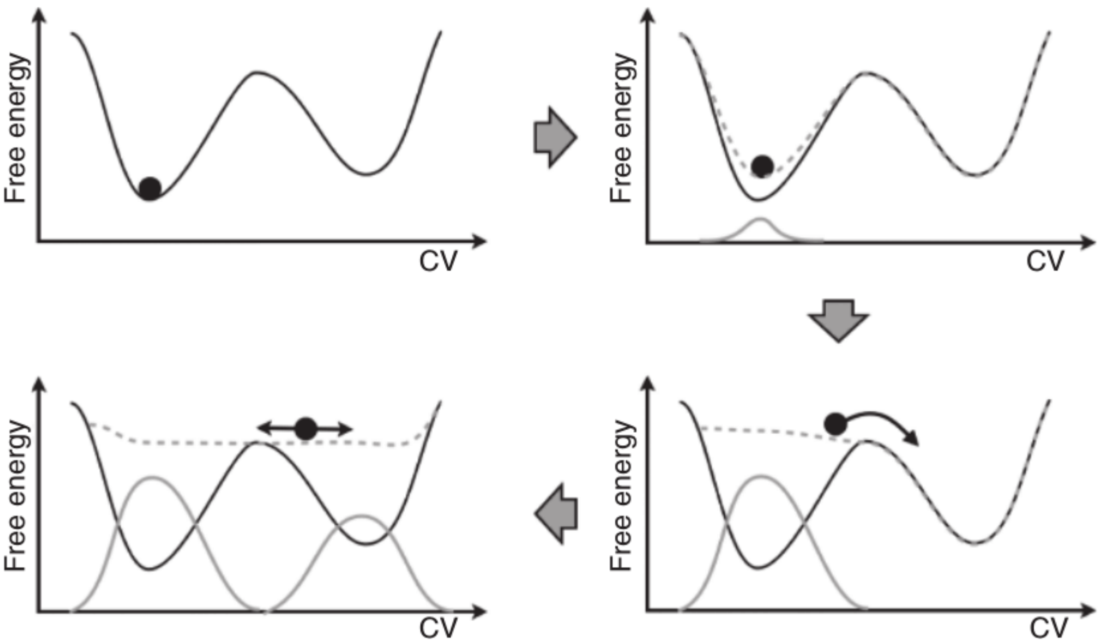


Illustration: Parrinello group, ETH Zürich

Adaptive sampling 2: Adaptive Biasing Force (ABF)

- ABF: time-dependent biasing force

$$\tilde{F}_t(x) = -\nabla V(x) + A'_t(\xi(x))\nabla\xi$$ where A'_t converges to A'

- long-time biased distribution is uniform, as in ABP
- how do we estimate A'?

Free energy derivative is a mean force

$$A(z) = -k_B T \ln \left(\int e^{-\beta V(x)} \delta(\xi(x) - z) dx \right)$$

$$A'(z) = \left\langle \frac{\partial V}{\partial z} - k_B T \frac{\partial \ln |J|}{\partial z} \right\rangle_{\xi(x)=z}$$

$-\dfrac{\partial V}{\partial z}$ is a projected force (defined by coordinate transform)

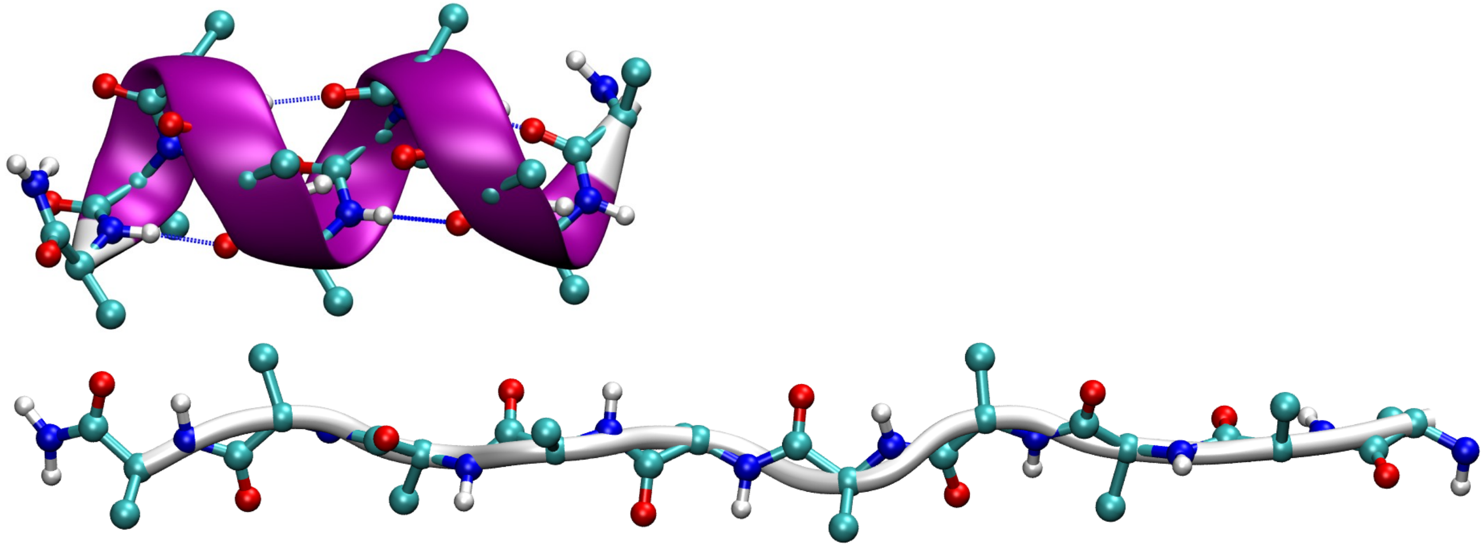
$\dfrac{1}{\beta} \dfrac{\partial \ln |J|}{\partial z}$ is a geometric (entropic) term

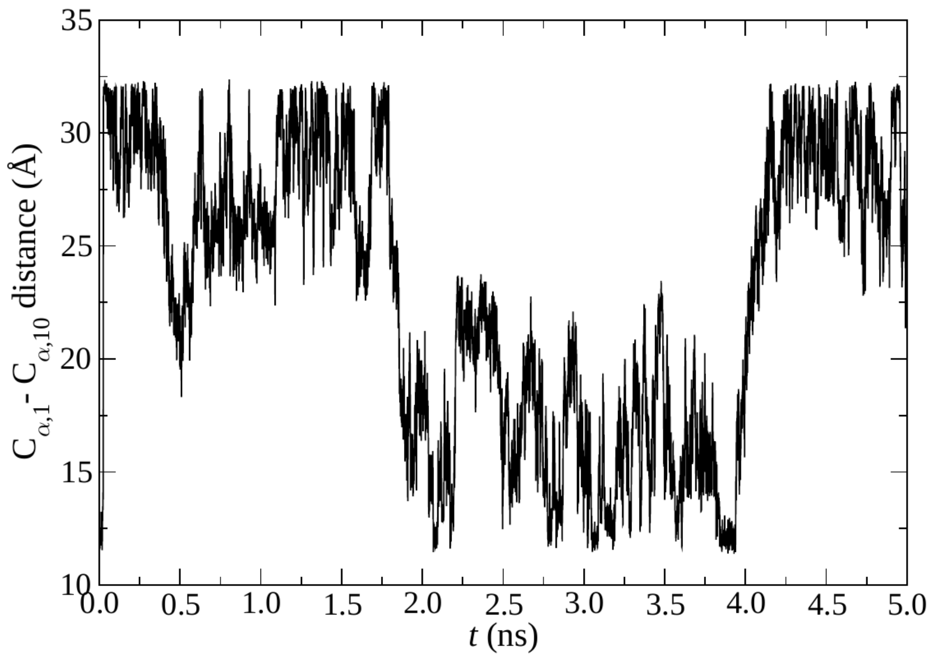
den Otter J. Chem. Phys. 2000

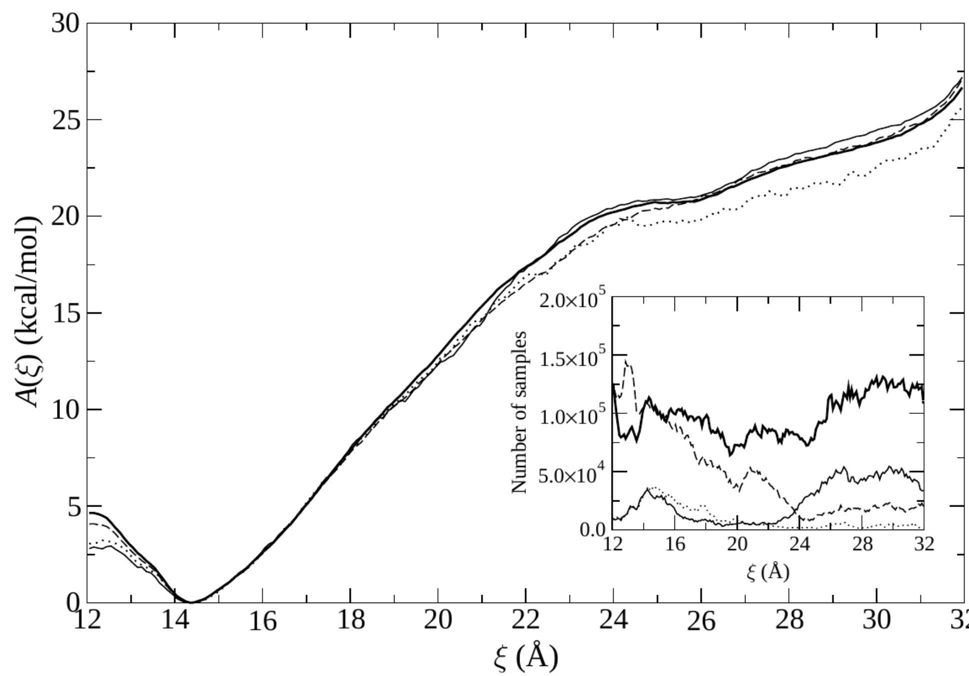
Simpler estimator of free energy gradient

- for each variable ξ_i, force is measured along arbitrary vector field $v_i(x)$ *(Ciccotti et al. 2005)*
- orthogonality condition: $v_i \cdot \nabla_x \xi_j = \delta_{ij}$
- free energy gradient: $\partial_i A(z) = \langle v_i \cdot \nabla_x V - k_B T \nabla_x \cdot v_i \rangle_{\xi(x)=z}$
- there are other estimators:
  - from constraint force (original ABF, Darve & Pohorille 2001)
  - from time derivatives of coordinate (Darve & Pohorille 2008)

1. Stretching deca-alanine

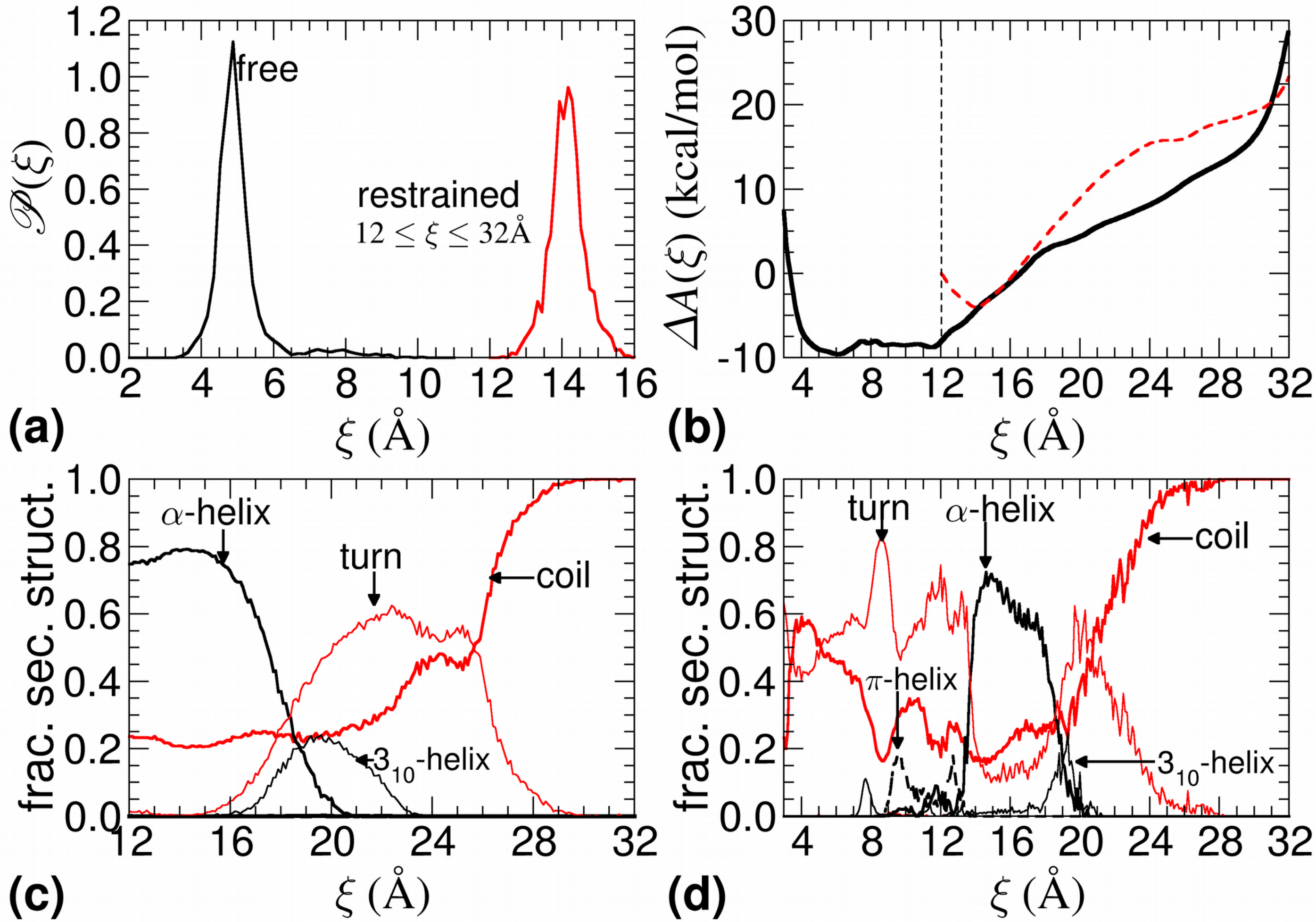


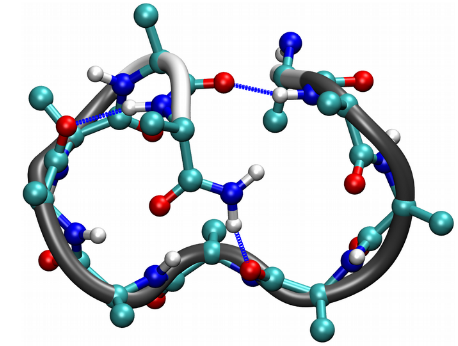


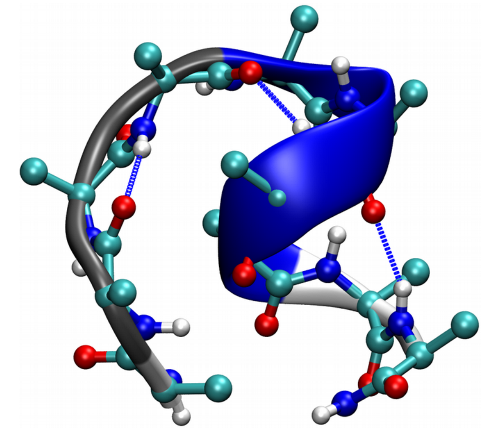


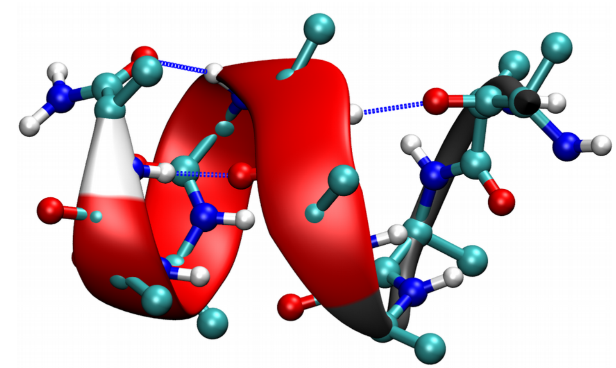
Hénin & Chipot JCP 2004

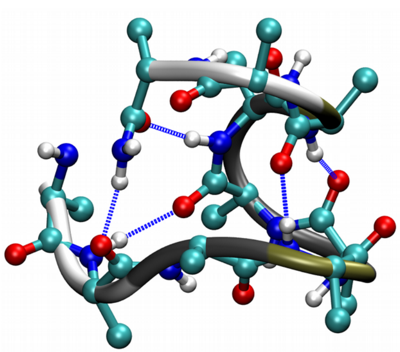
2. Sampling deca-alanine?

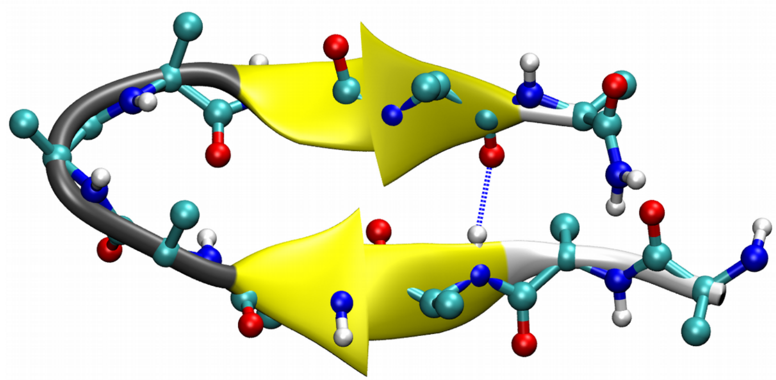






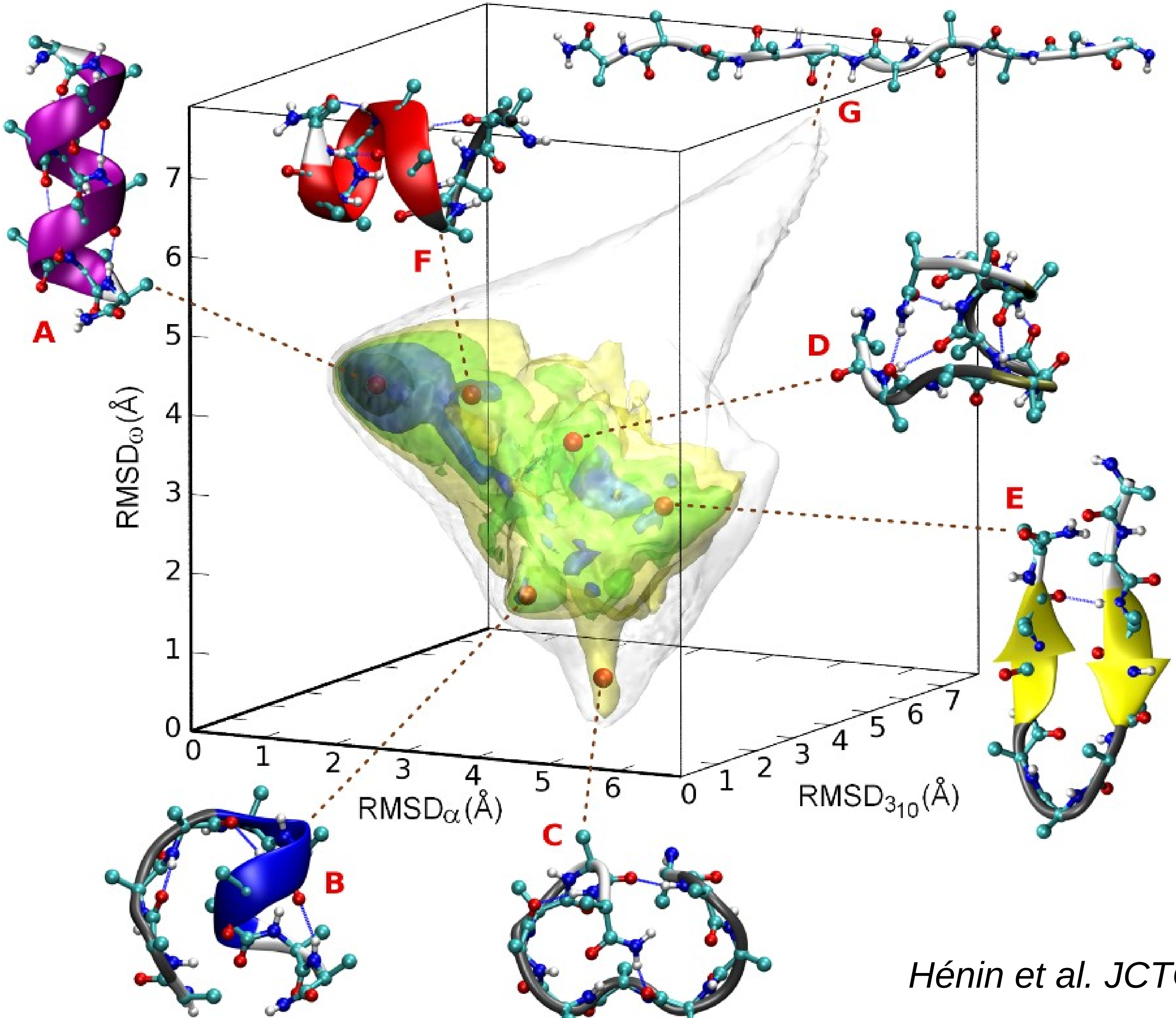






Chipot & Hénin JCP 2005

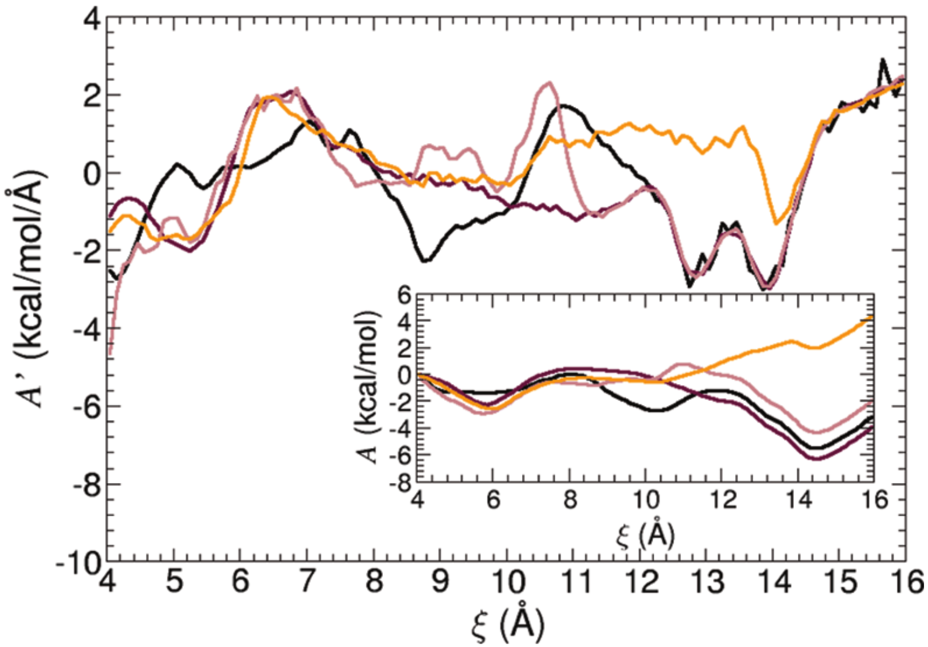
3. Sampling in higher dimension



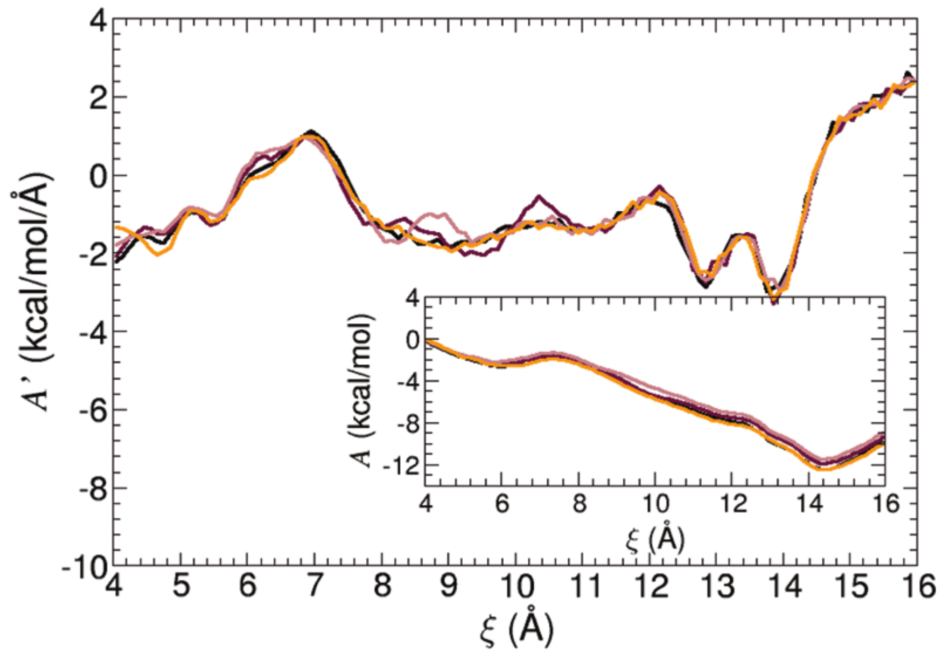
Hénin et al. JCTC 2010

4. More robust sampling for poor coordinates: Multiple-Walker ABF

- good performance with *hidden barriers* (Minoukadeh, Chipot, Lelièvre 2010)
- can sample systems using incomplete set of collective variables?



ABF, 1 x 100 ns



MW-ABF, 32 x 3 ns

ABF: a tale of annoying geometry

Estimator of free energy gradient:

- for each variable ξ_i, force is "measured" along arbitrary vector field v_i *(Ciccotti et al. 2005)*
- orthogonality conditions: $\begin{cases} \boldsymbol{v}_i \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \xi_j = \delta_{ij} \\ \boldsymbol{v}_i \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \sigma_k = 0 \end{cases}$

- free energy gradient: $\partial_i A(z) = \langle v_i \cdot \nabla_x V - k_B T \nabla_x \cdot v_i \rangle_{\xi(x)=z}$

- geometric calculations are sometimes intractable
(e.g. second derivatives of elaborate coordinates)
- orthogonality conditions are additional constraints
- in practice, many cases where **ABF is unavailable**

extended-system Adaptive Biasing Force (eABF)

- idea: Lelièvre, Rousset & Stoltz 2007
- implementation: Fiorin, Klein & Hénin 2013

Get rid of geometry by watching an unphysical variable λ ,
harmonically coupled to our geometric coordinate:

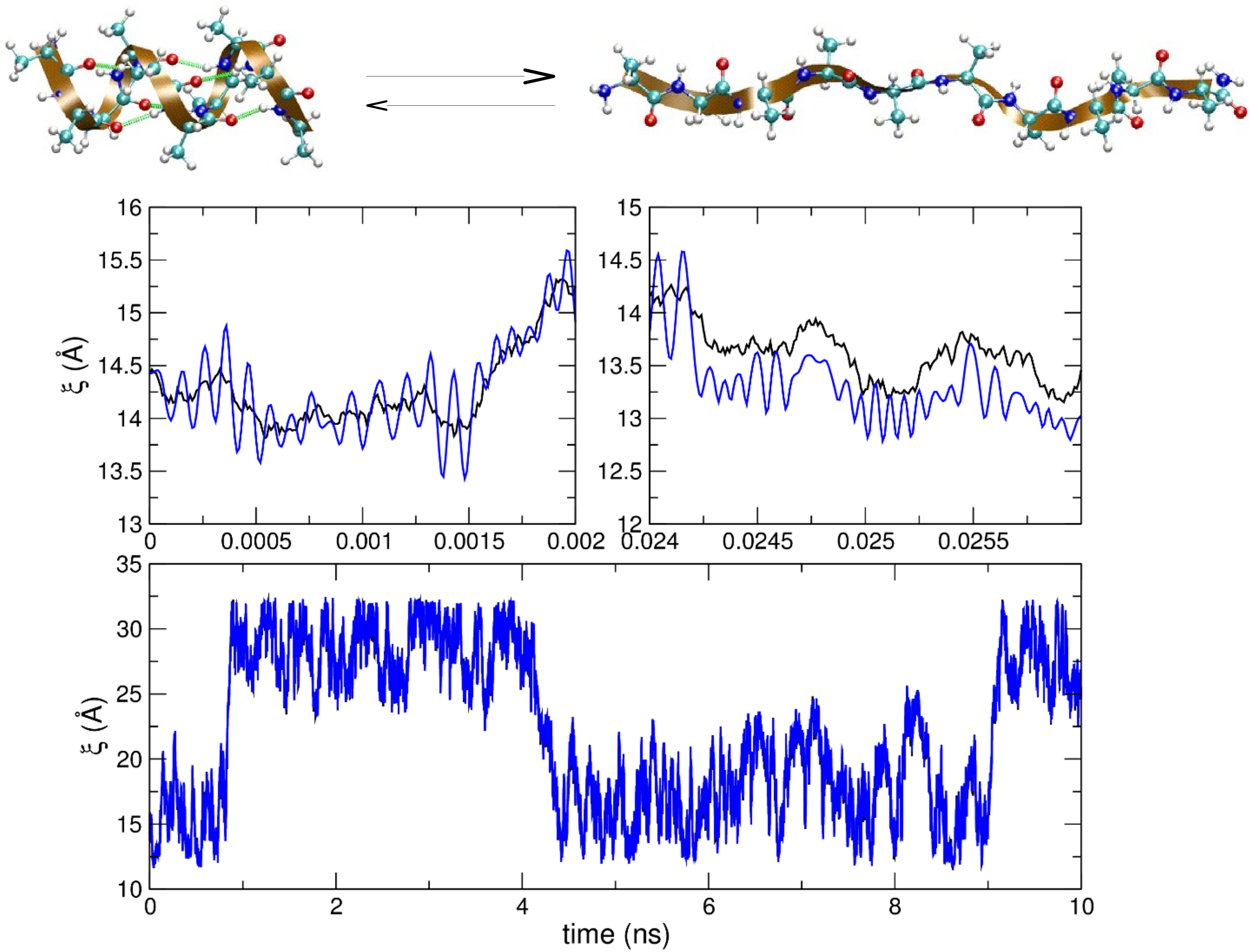
$$V^k(x,\lambda) = V(x) + \frac{1}{2}k\left(\xi(x) - \lambda\right)^2$$

λ undergoes Langevin dynamics with mass m.
Mass and force constant based on desired fluctuation and period:

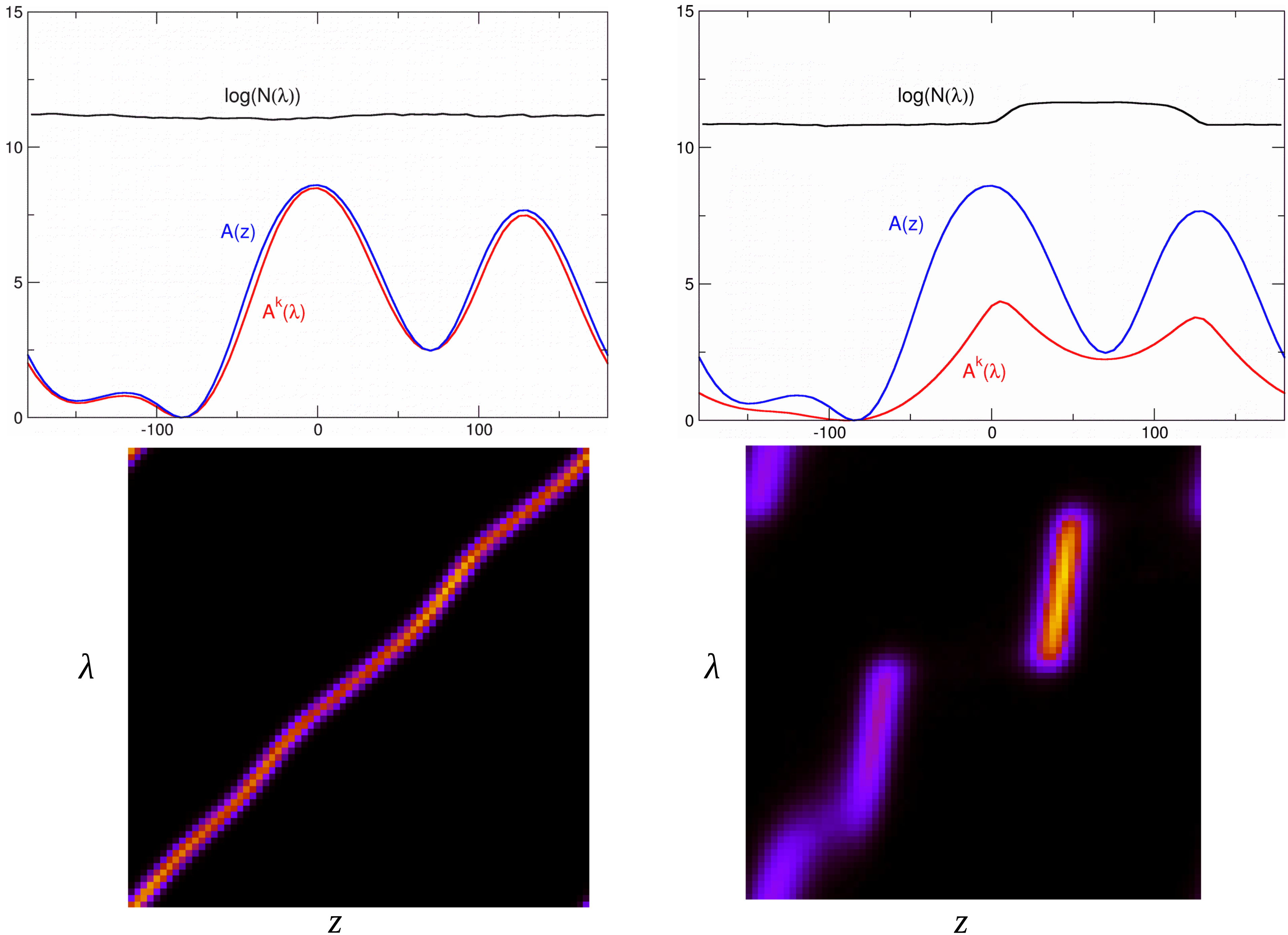
$$\sigma = \sqrt{\frac{k_B T}{k}}$$

$$\tau = 2\pi\sqrt{\frac{m}{k}}$$

eABF trajectories



Tight vs. loose coupling



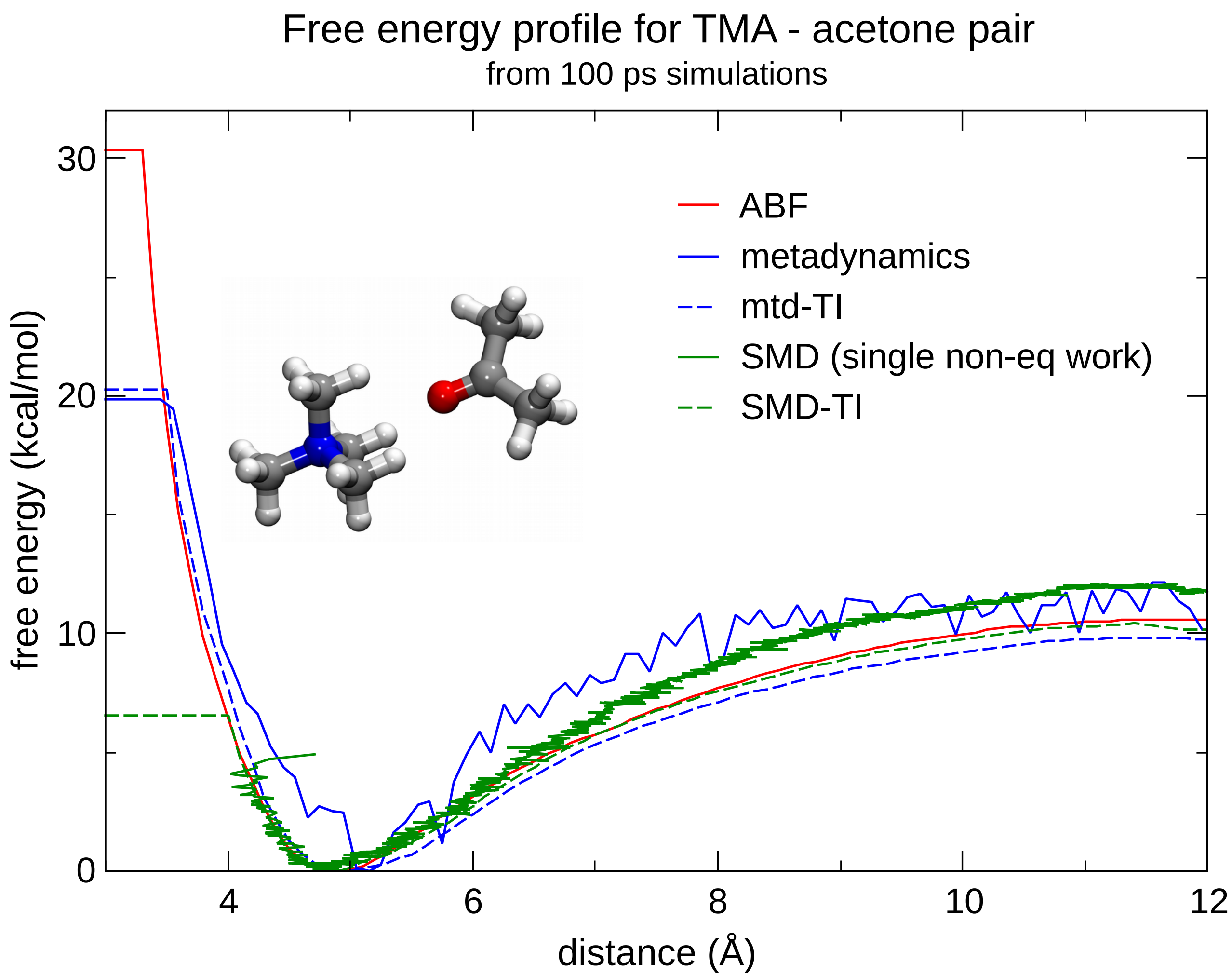
Free energy estimators for eABF

- A^k is an estimator of free energy A, asymptotically accurate **for high *k***
- other estimators lift this “stiff spring” requirement:
  - umbrella integration (Kästner & Thiel 2005, Zheng & Yang 2012, Fu, Shao, Chipot & Cai 2016)
  - CZAR (Lesage, Lelièvre, Stoltz & Hénin 2017)
- using these estimators, eABF is a **hybrid adaptive method** (free energy estimate is separate from bias)

Hybrid methods

- adaptive sampling combines free energy estimation and enhanced sampling
- hybrid methods: bias based on one estimator, use another estimator to compute final free energy
- examples:
  - unbiased sampling with thermodynamic integration
  - metadynamics with thermodynamic integration
  - eABF dynamics with UI or CZAR estimator

Different estimates at very short sampling times



- same long-time results, but different short-time convergence!
- caution: may be system-dependent
- efficiency of sampling vs. biases in short-time estimates

  → benefit of hybrid methods

Thank you!

Questions?