Adaptive free energy biases:

doesn't matter what way you push

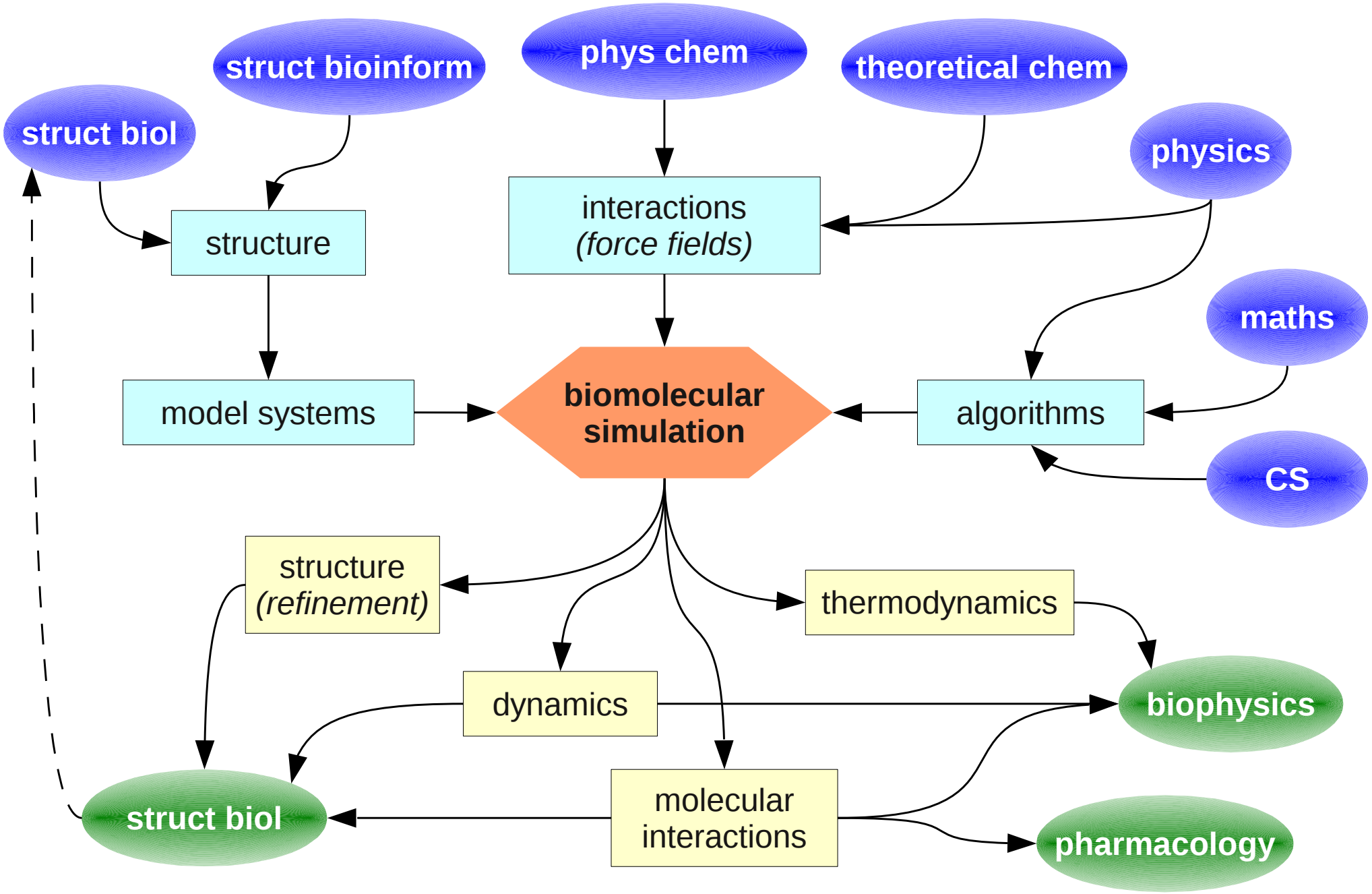
as long as you push hard enough

Jérôme Hénin

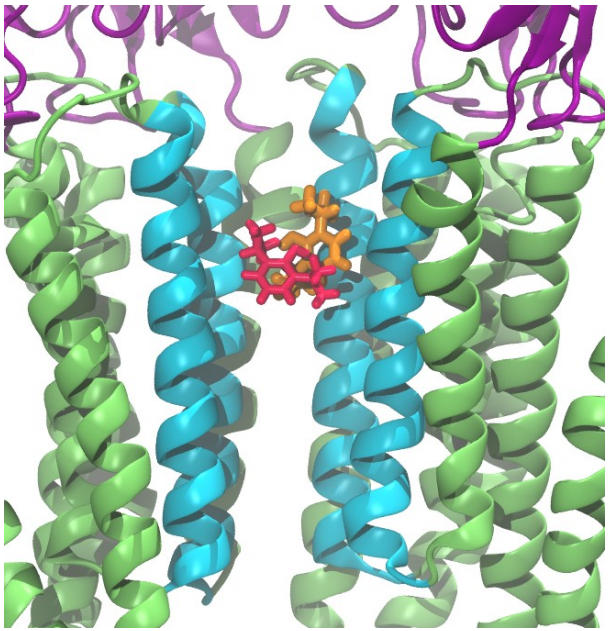
CNRS, Aix-Marseille University

CECAM workshop, Paris
6 June 2012

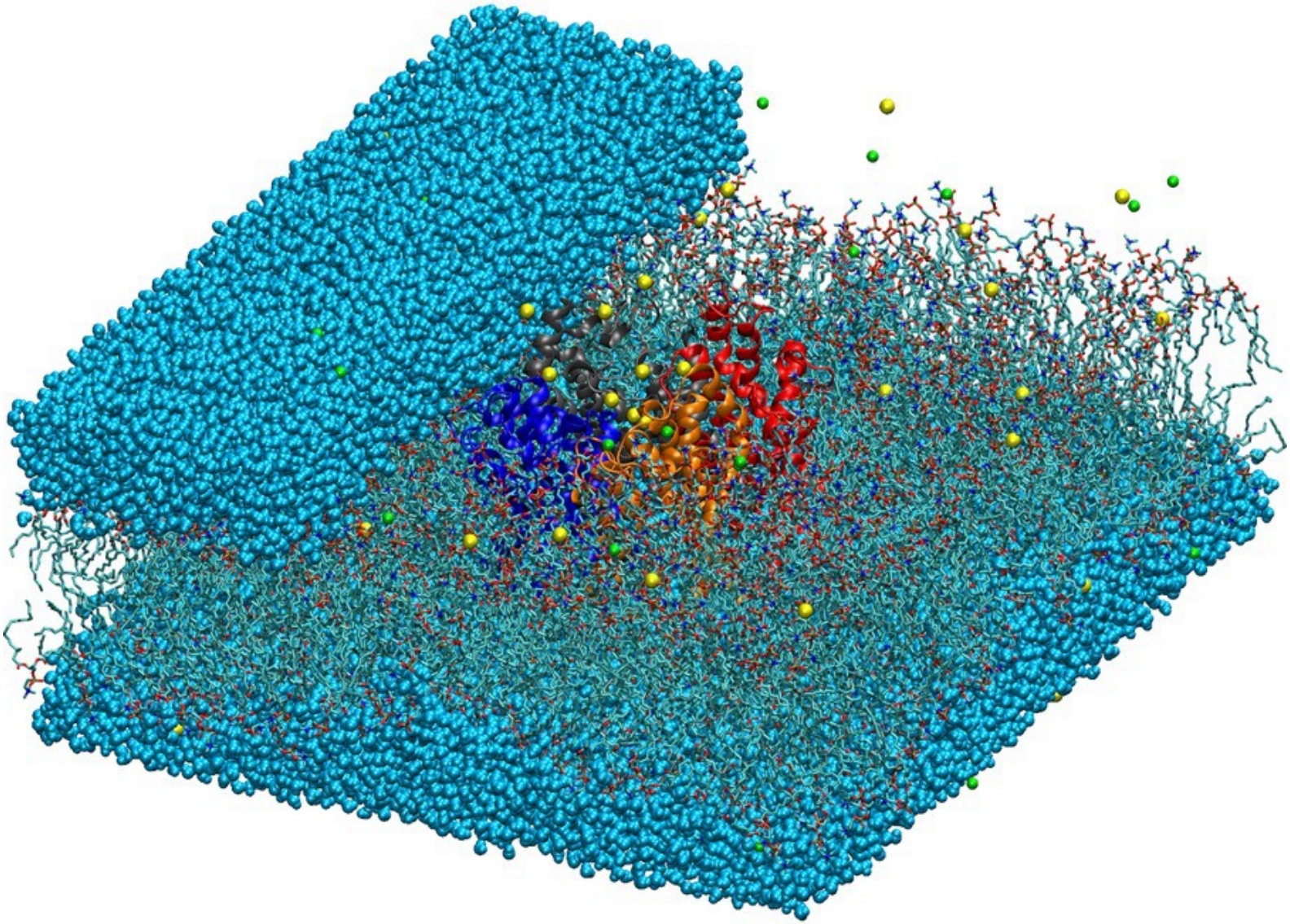
we talk to everyone...



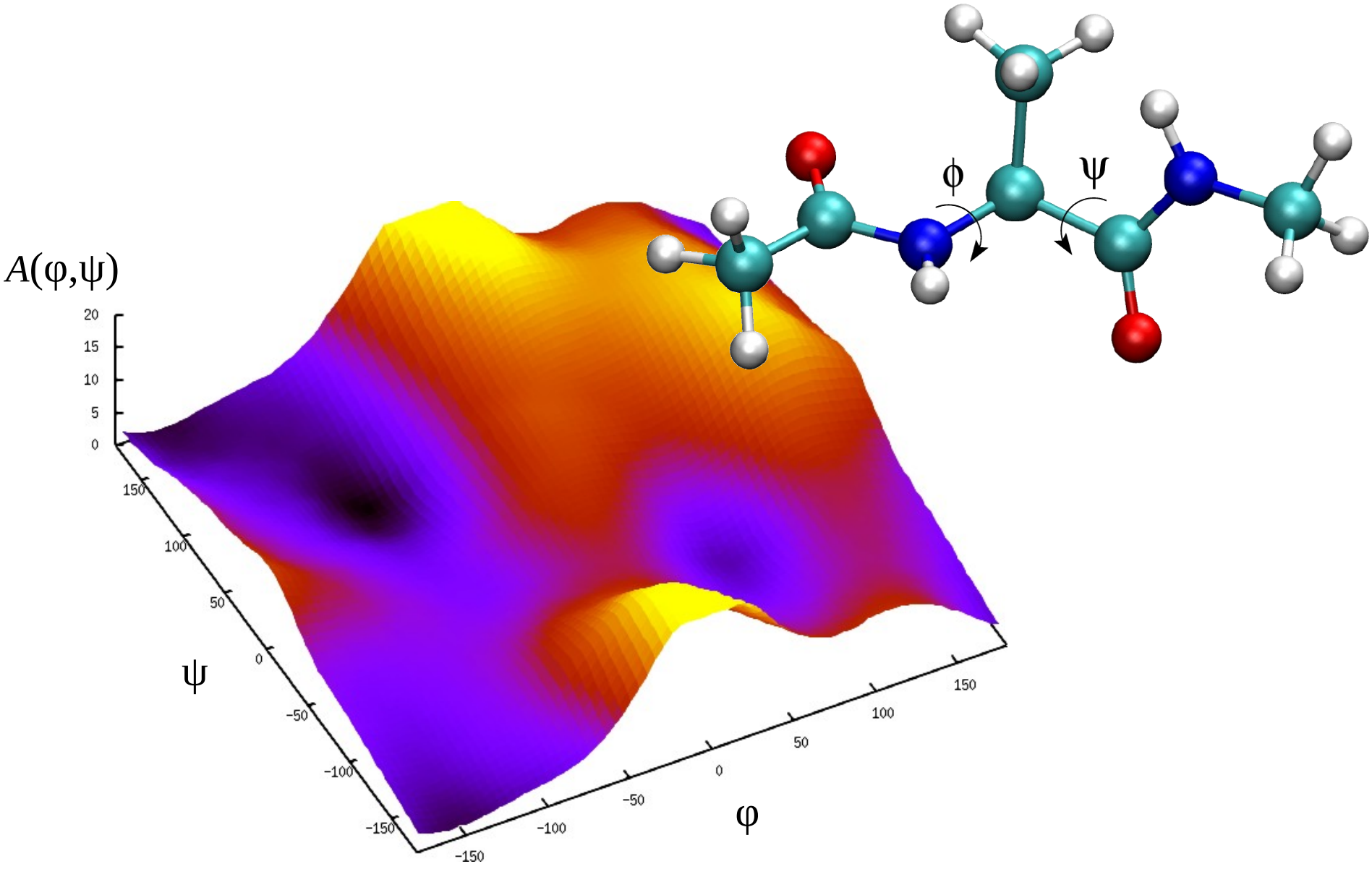
...including biologists



Biology has too many degrees of freedom



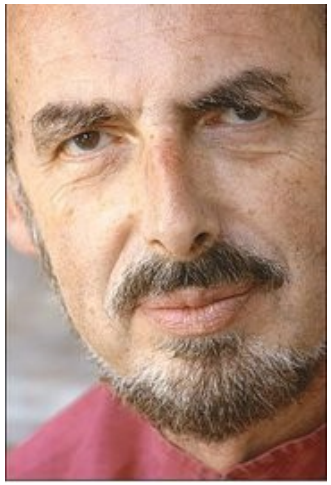
Reduced representation



“Have a bias toward action - let's see something happen now.”
Indira Gandhi

Adaptive Biasing Force:

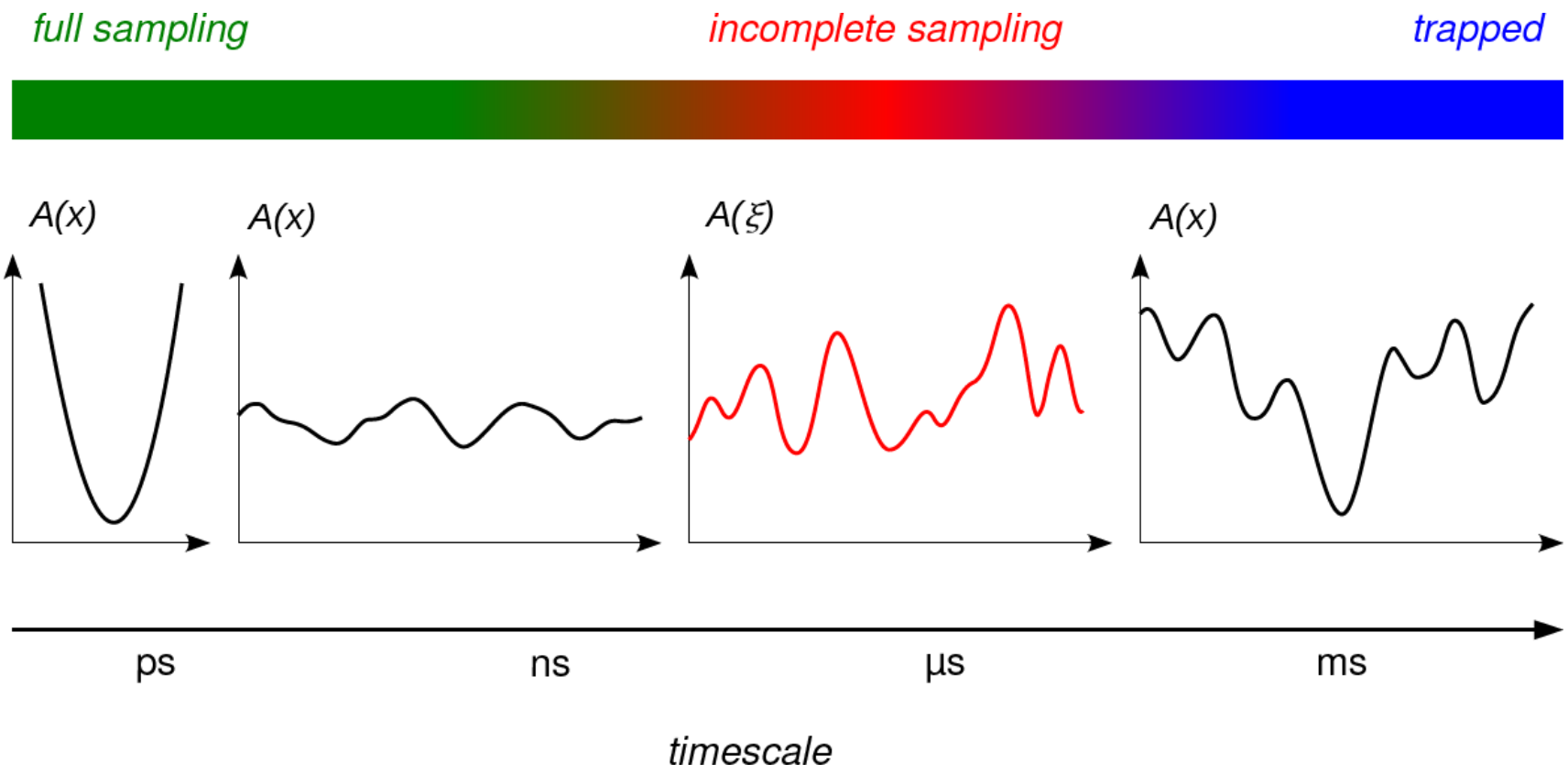
making things happen



Andrew Pohorille

Chris Chipot

The problem: partial sampling



Thermodynamic integration

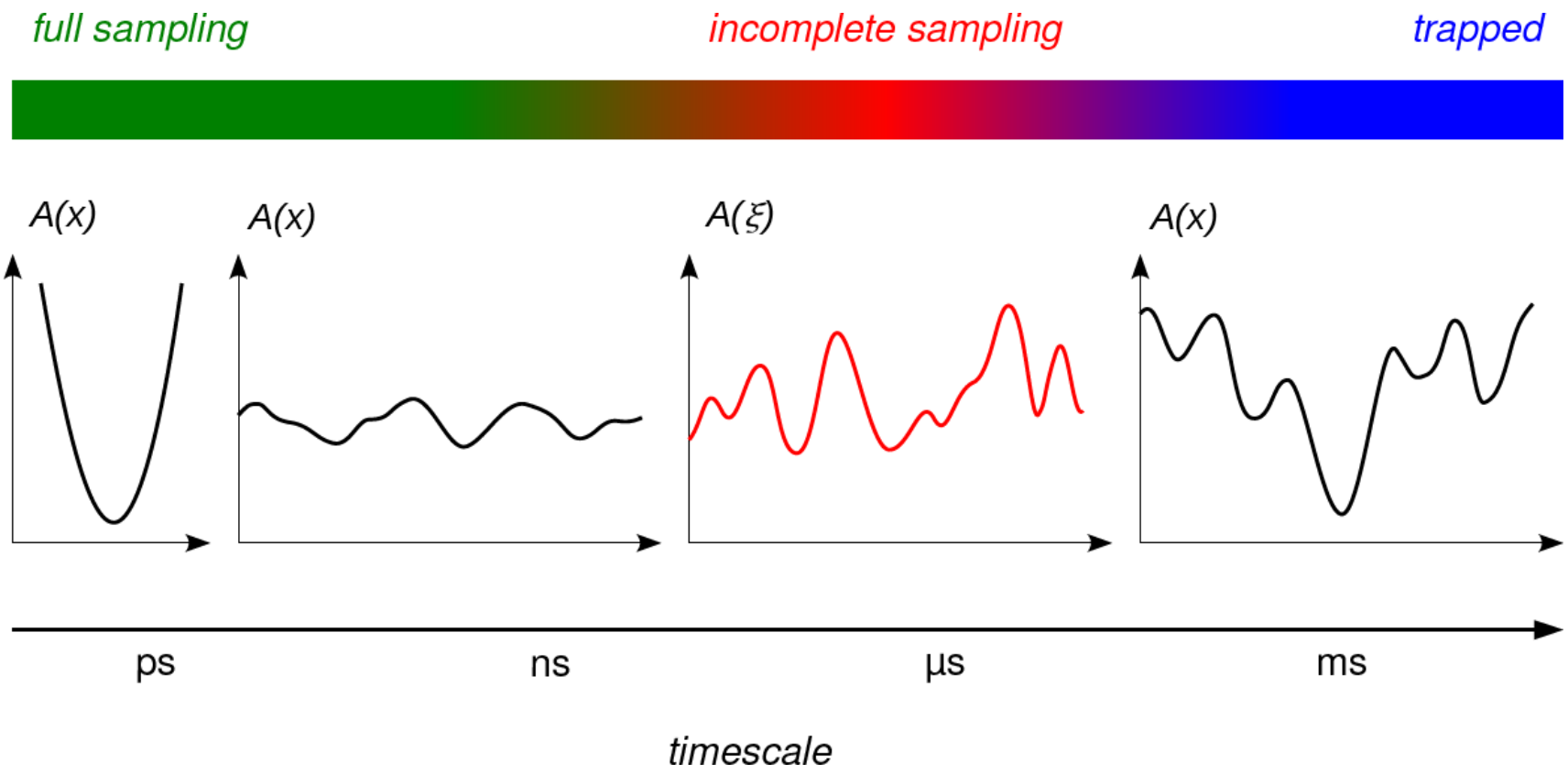
- Basic expression:

$$\frac{dA}{d\xi} = \left\langle \frac{\partial \mathcal{V}}{\partial \xi} - k_B T \, \frac{\partial \ln |J|}{\partial \xi} \right\rangle_{\xi}^{\text{conf}}$$

  ($\mathcal{V}$: potential energy; J: Jacobian)

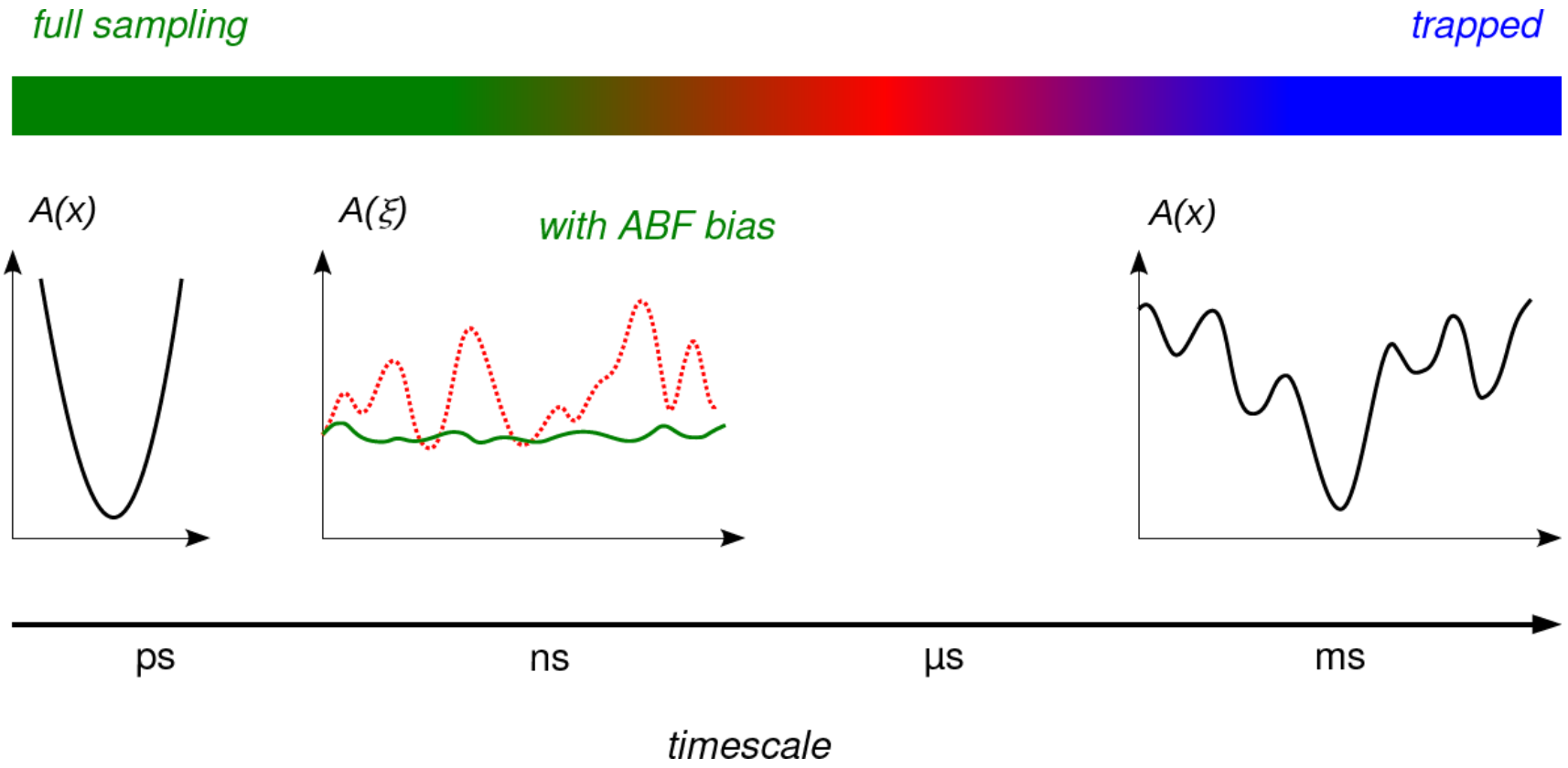
- $\frac{\partial \mathcal{V}}{\partial \xi}$ is a force along ξ
- $-k_B T \, \frac{\partial \ln |J|}{\partial \xi}$: correction for the dependence on ξ of the phase space volume element

Timescale separation



Timescale separation

ABF: Darve and Pohorille, 2001



Intermission: statistical error analysis

- often depends on effective sampling: number of uncorrelated samples $N_{eff} = N / t_{corr}$
- biomolecules: trapped DOFs, hence $t_{corr} >= N$
- hence $N_{eff} <= 1$, everything depends on initial conditions

"We work with models of the simulation process, sidestepping the ***tricky and computationally expensive problem*** *of* ***relying on simulations to provide their own error statistics****."*

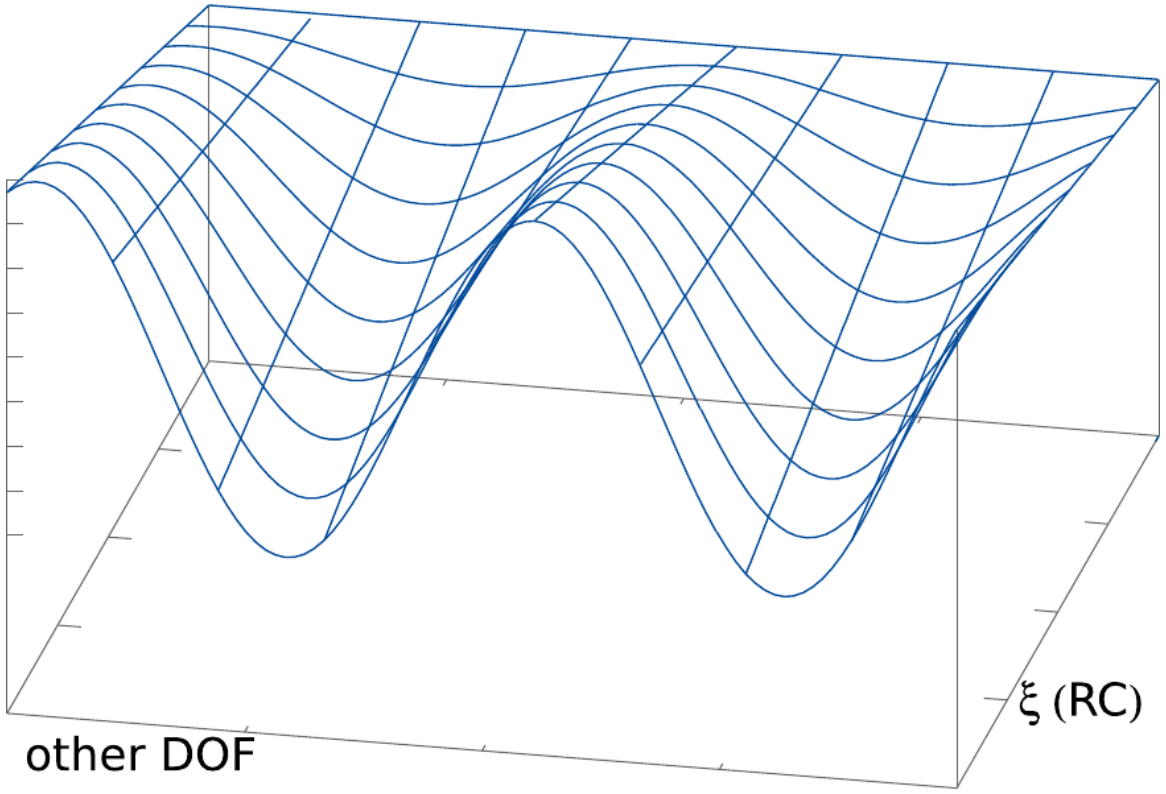
D. Kofke and P. Cummings, Mol. Phys. 1997

"We assume that this problem has ***already been solved****"*

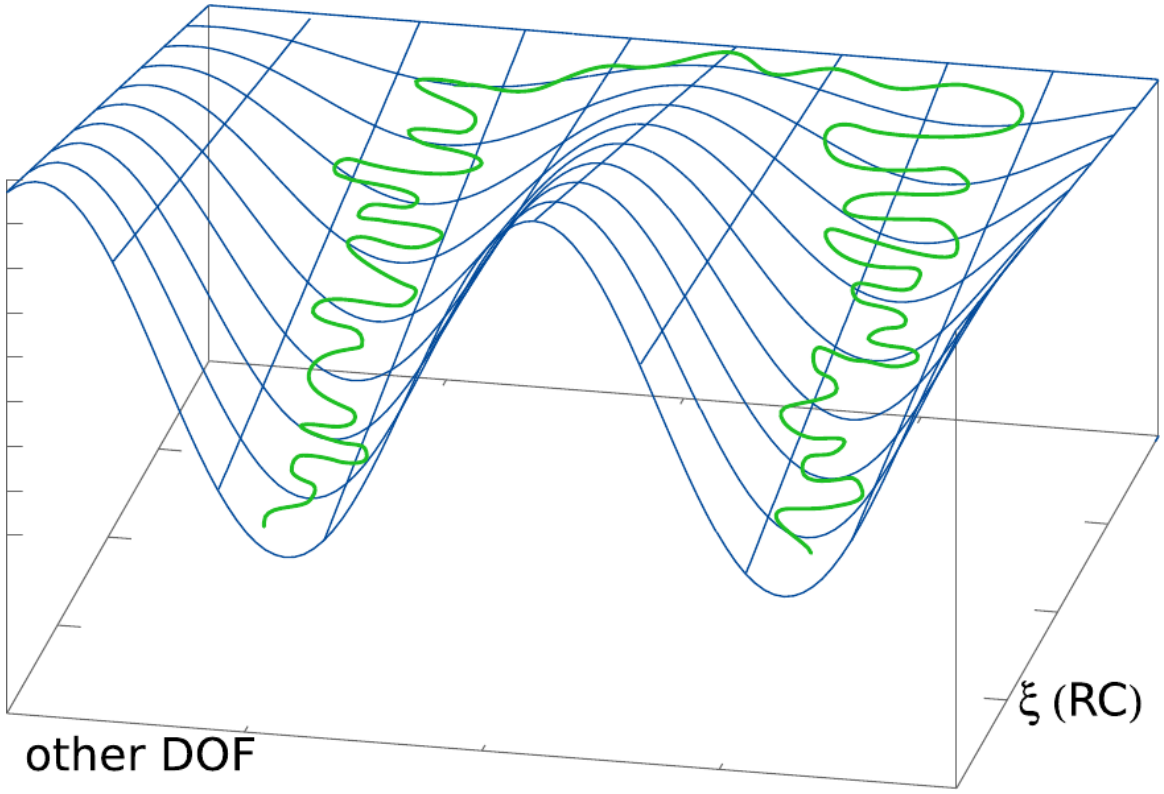
D. Kofke, CECAM 2012

- some cases are better behaved than others (e.g. neglected tail model)
- block averaging among acceptable options?

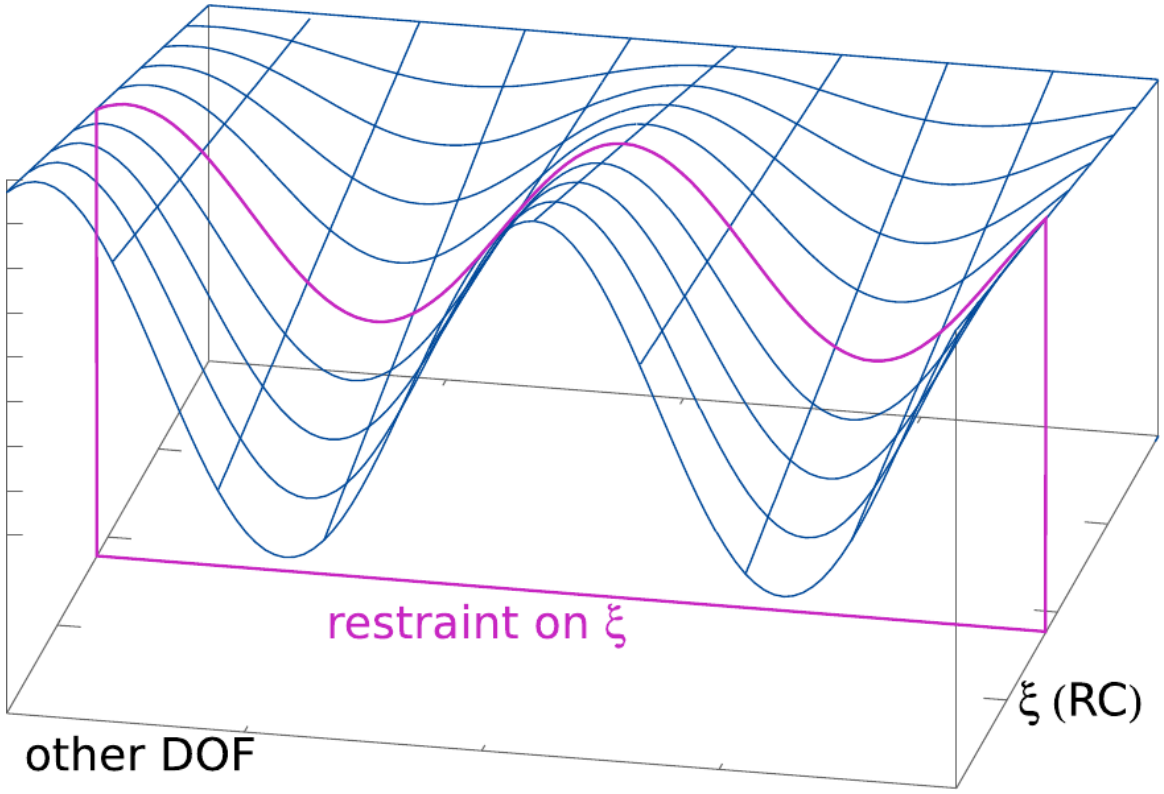
Sampling orthogonal DOFs



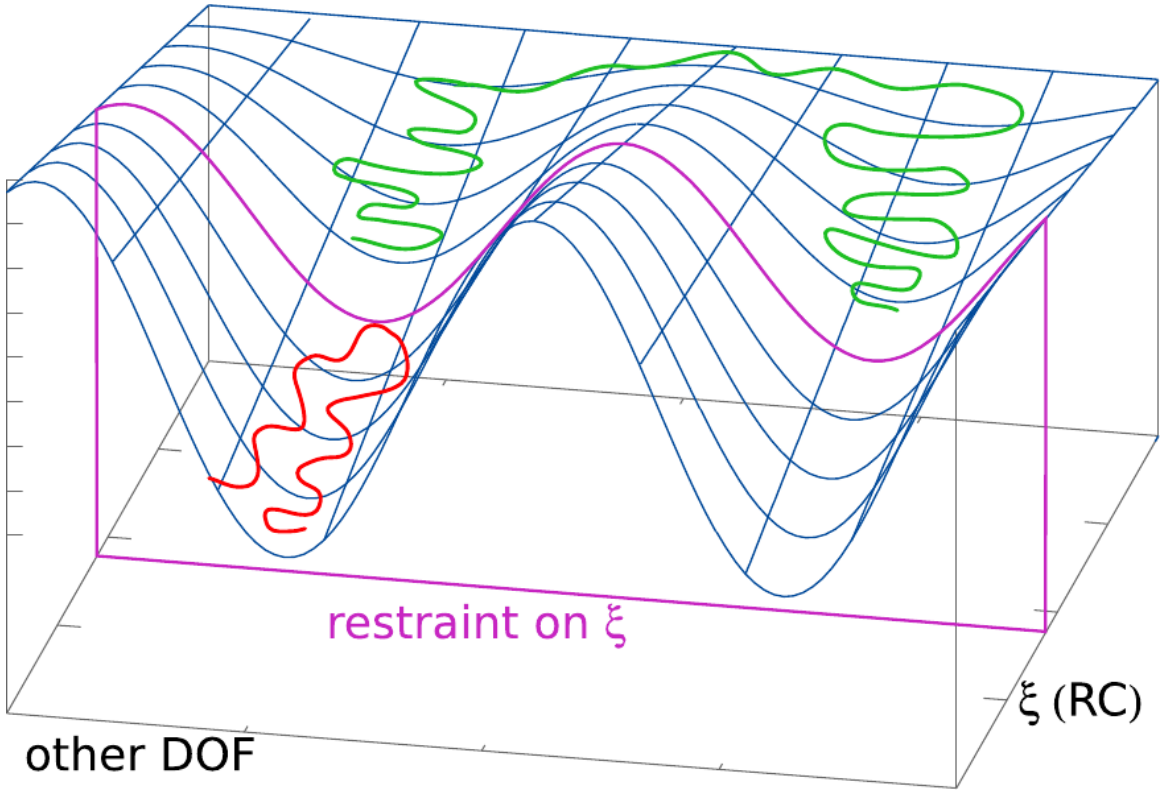
Sampling orthogonal DOFs



Sampling orthogonal DOFs



Sampling orthogonal DOFs



Implementation for large biomolecules:

"collective variables module"

Giacomo Fiorin

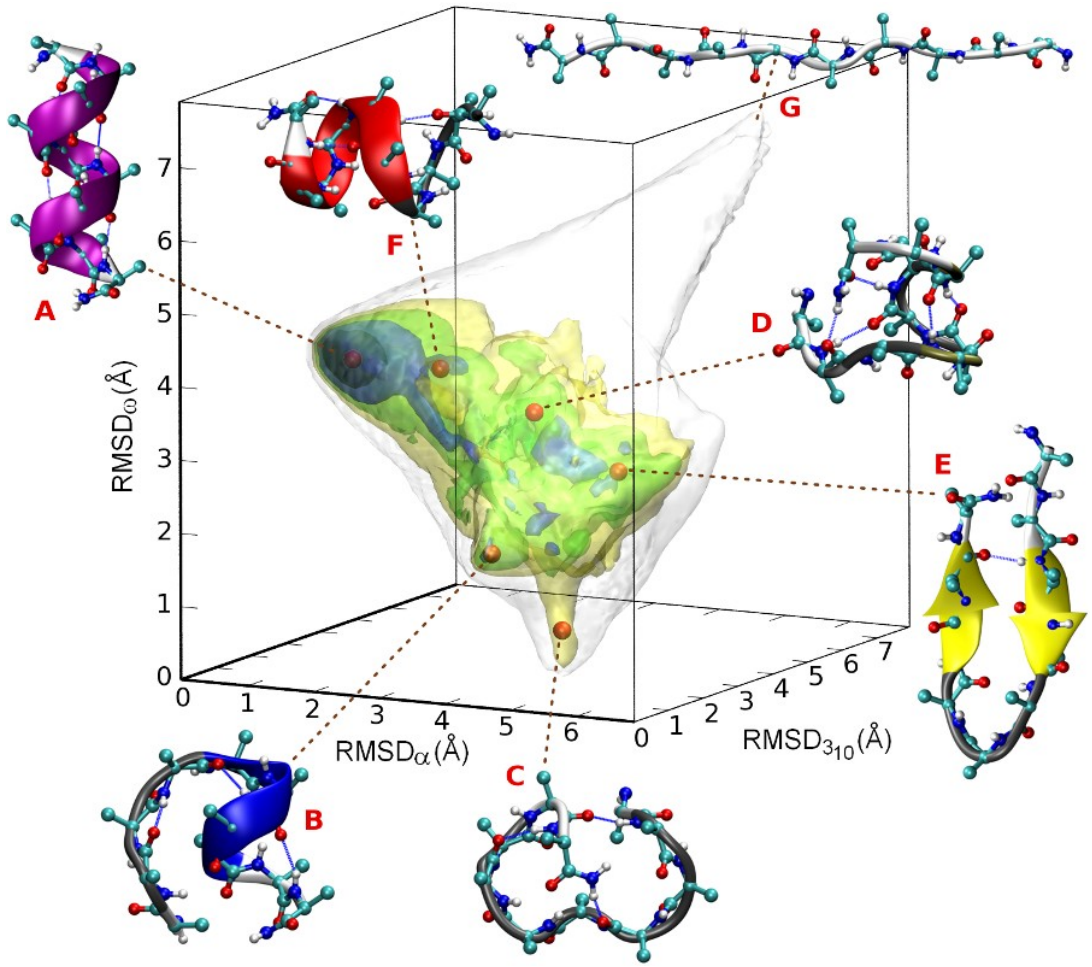
(Klein lab)

Colvars: flexible generalized coord. biases

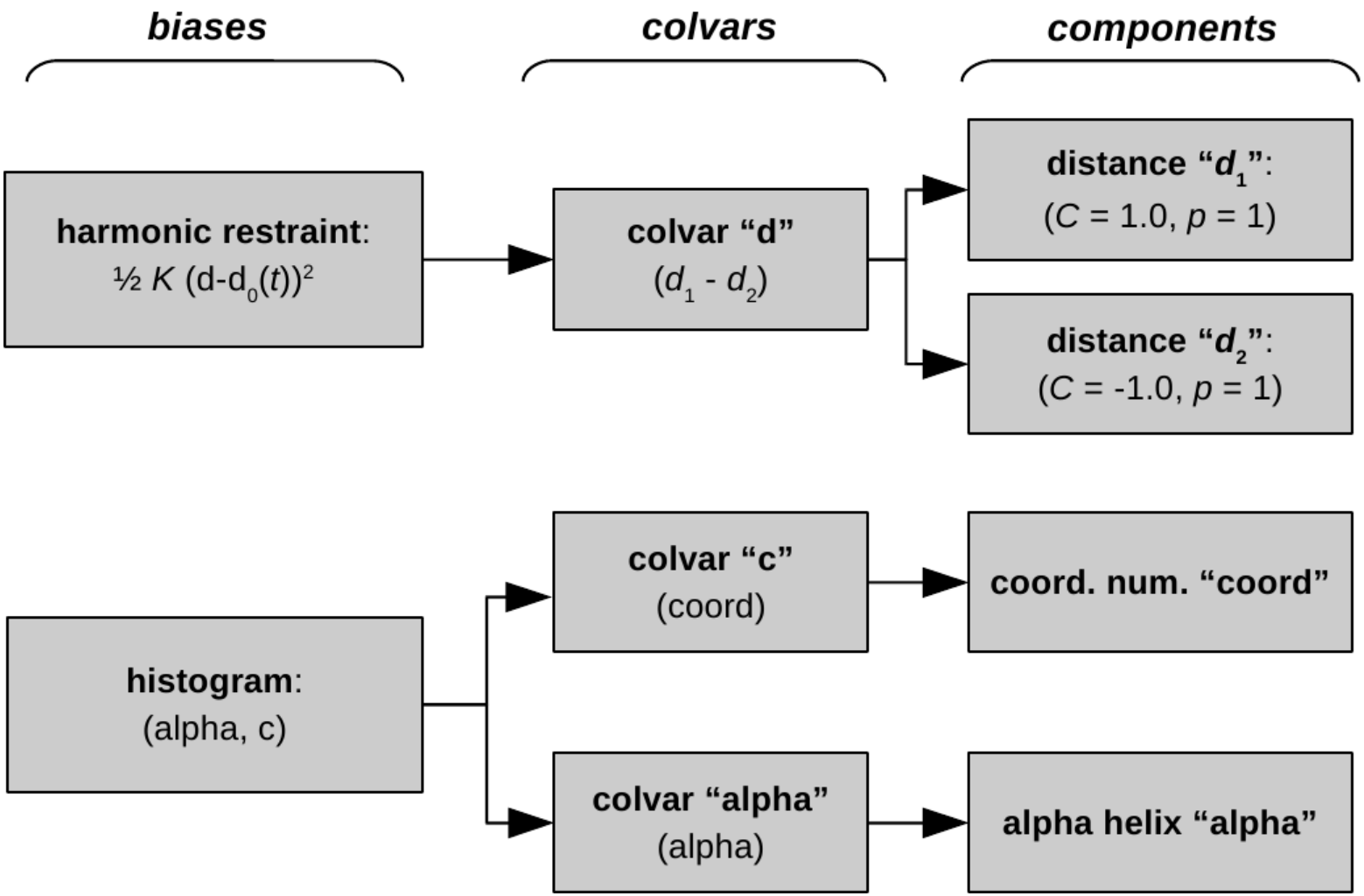
- arbitrary dimension
- run-time combination of variables
- sophisticated variables available
- ABF, ABP (metadynamics)
- moving restraints:

  steered MD, targeted MD
- C++, designed for extensibility

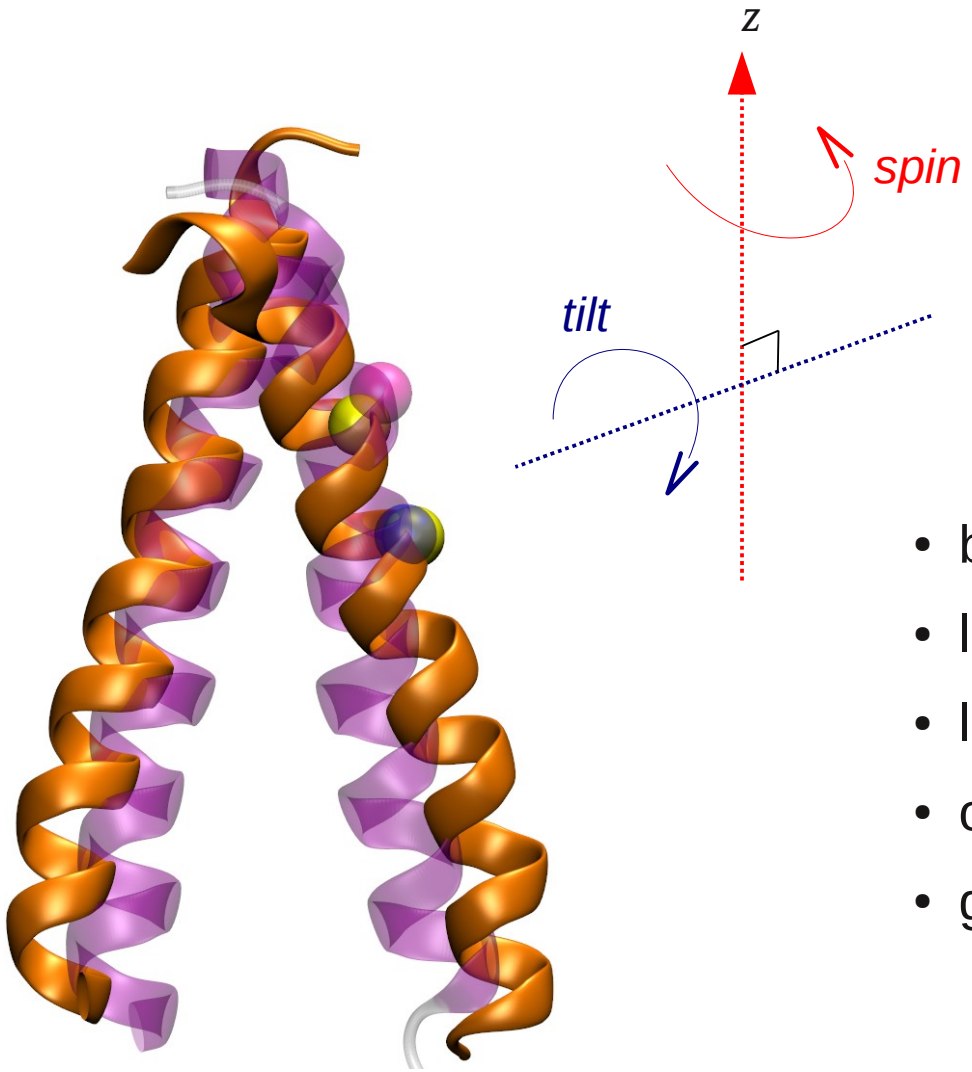
- included in **NAMD**
- available for **LAMMPS**
- generic interface: may be ported to any MD software



Two-sided modularity: variables, algorithms



Rotation angles around preferred axes



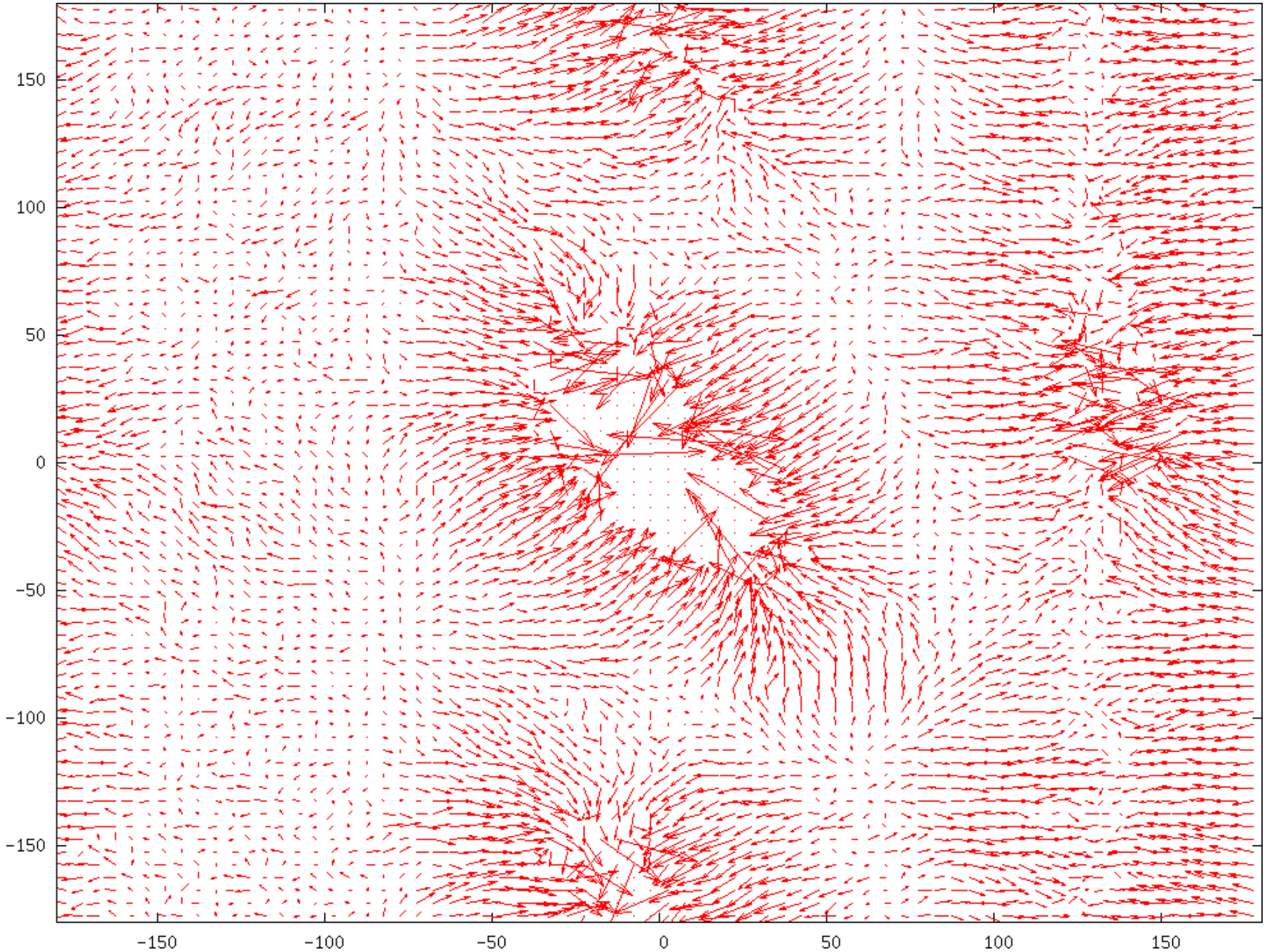
- based on reference dimer structure
- least-square fit of whole dimer
- least-square fit of each helix
- optimal rotation split into two rotations
- gives two rotation angles

Dimer of helical transmembrane protein segments

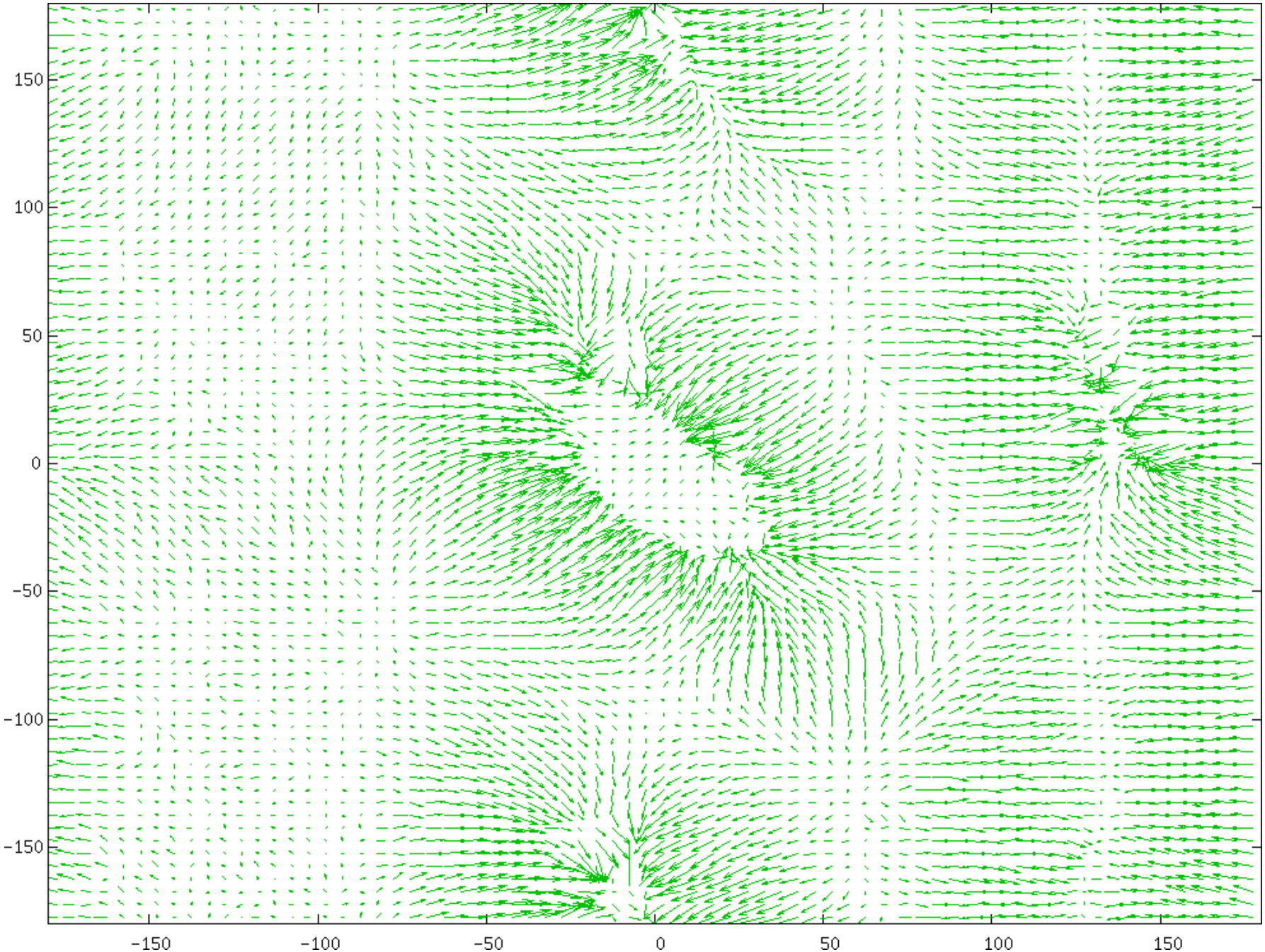
ABF: thermodynamic force in dim > 1

- for each variable $\boldsymbol{\xi}_i$, force is measured along arbitrary vector field $\boldsymbol{v}_i$ *(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: $$\frac{\partial A}{\partial \xi_i} = \left\langle \boldsymbol{v}_i \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} V - k_B T \boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \boldsymbol{v}_i \right\rangle_{\boldsymbol{\xi}}$$
- divergence of $\boldsymbol{v}_i$ gives geometric correction (ideal gas entropy term)

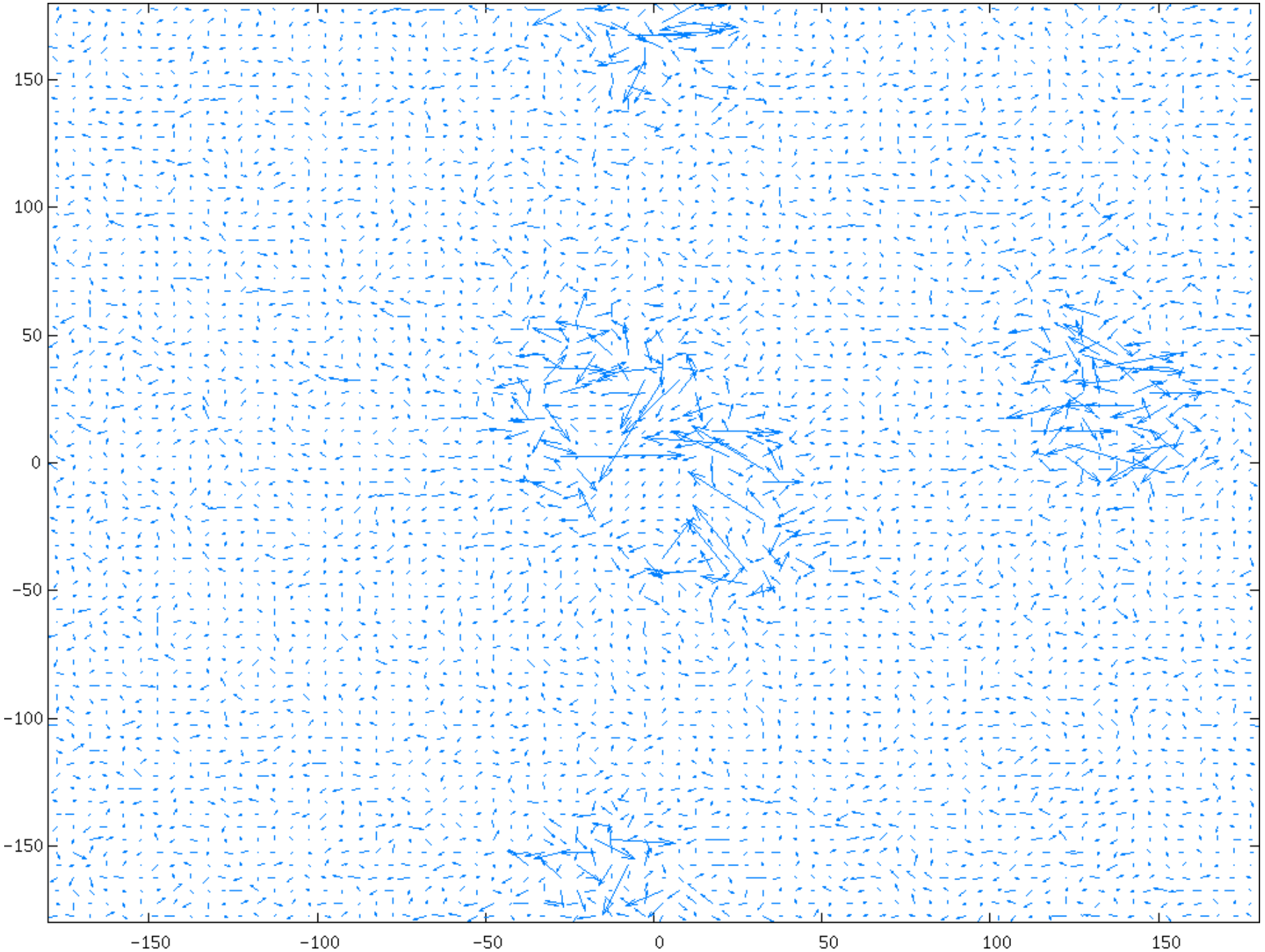
Integrating n-dimension “gradients”



Helmholtz decomposition

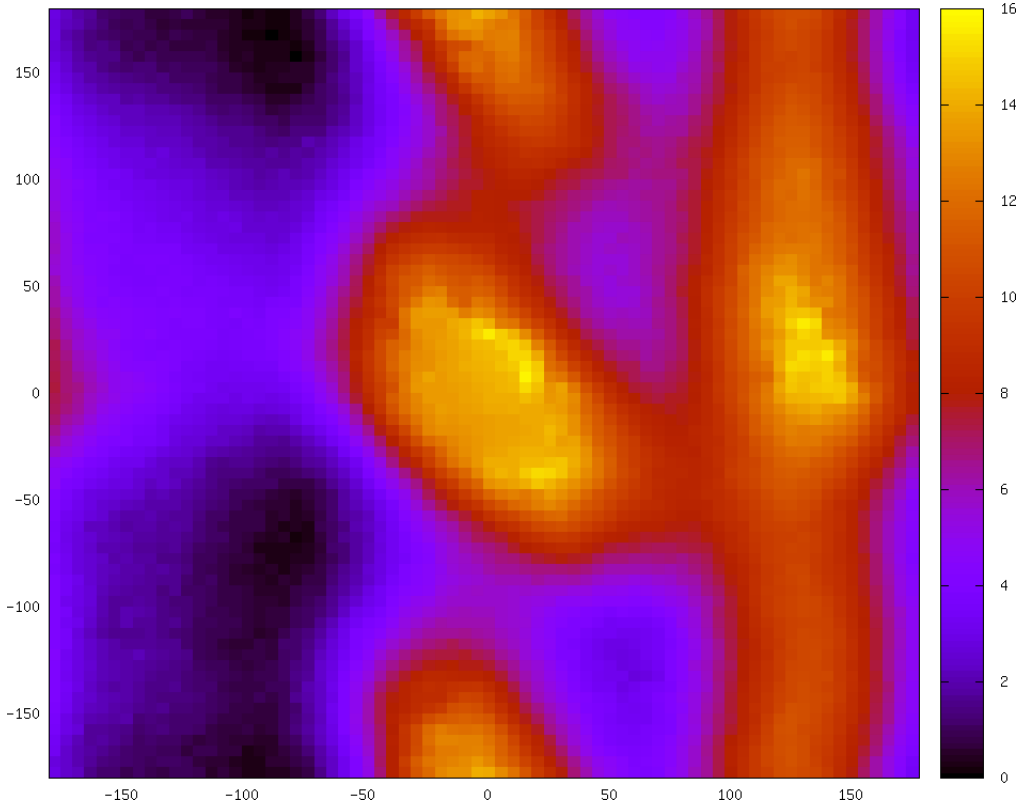


Helmholtz decomposition

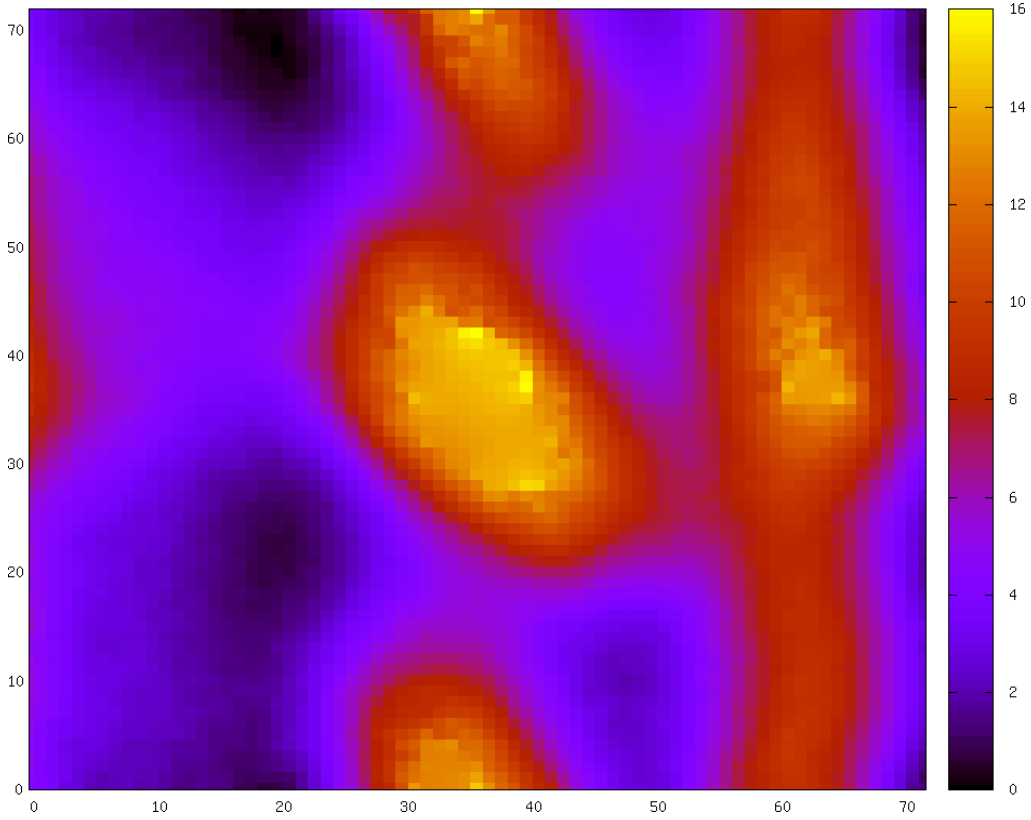


On-the-fly Poisson integration

with Tony Lelièvre (ENPC, Paris)



Monte-Carlo, PBC



Poisson, Neumann BC

ABF: cumbersome requirements

- calculate gradients (OK)
- differentiate Jacobian determinant (second derivatives):
  - design explicit generalized coordinates
  - choose “inverse gradient” field
    - calculate divergence of this field
- mutually orthogonal RCs
- RCs orthogonal to constraints

→ not always applicable to complicated variables

Extended-system ABF (eABF)

Tony Lelièvre

Extended-system ABF (eABF)

- for each collective variable $\xi_i(x)$, add extended coordinate q_i
- coupled by harmonic spring: $V_h = \frac{1}{2} k \left(\xi_i(x) - q_i\right)^2$

- separate Hamiltonian integrator using fictitious mass

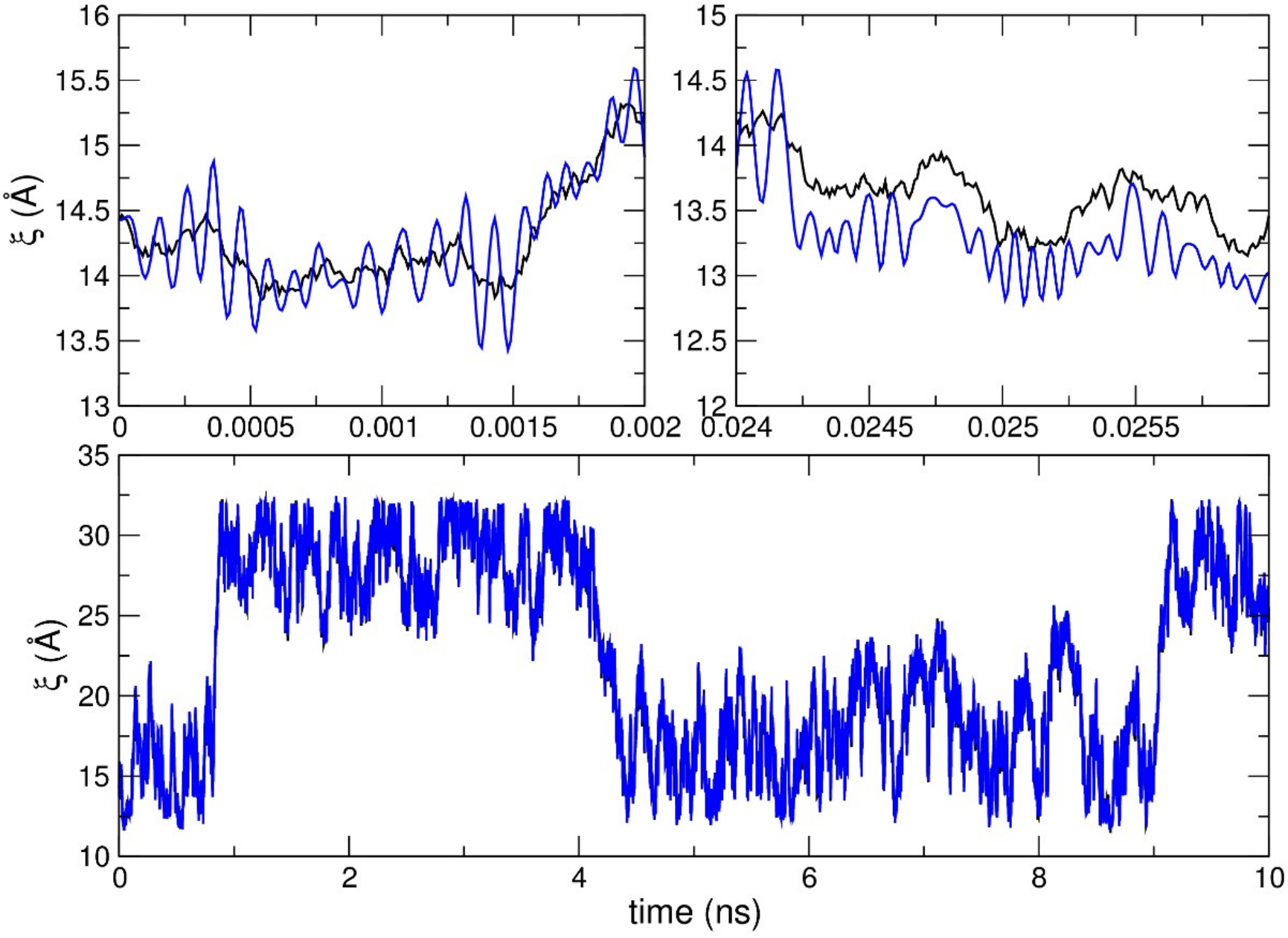
  Lelièvre et al. JCP 2007, Zheng and Yang JCTC 2012

- pick mass and force constant based on desired fluctuation and time constant:

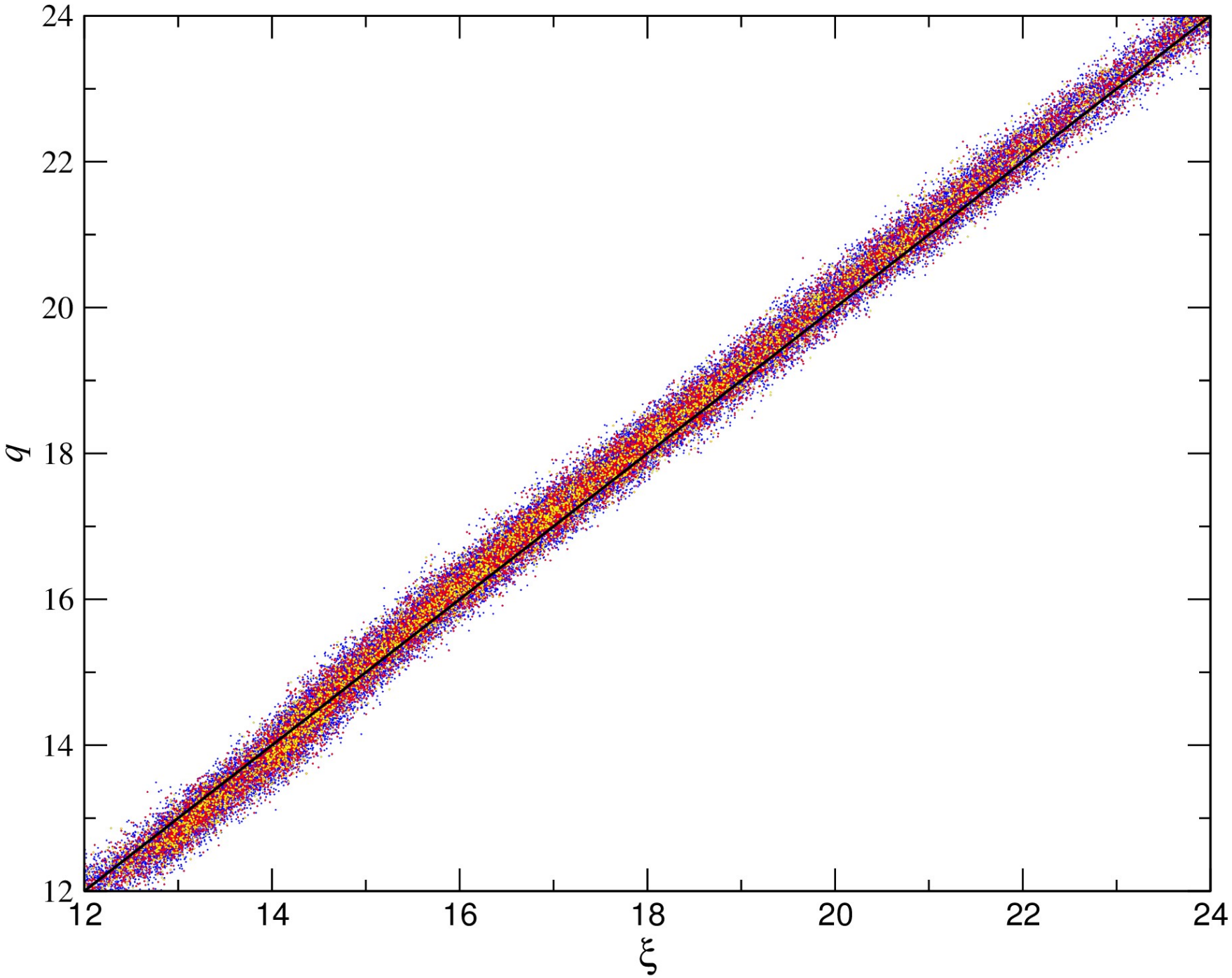
$$\sigma = \sqrt{\frac{k_B T}{k}}$$

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

eABF: fluctuations



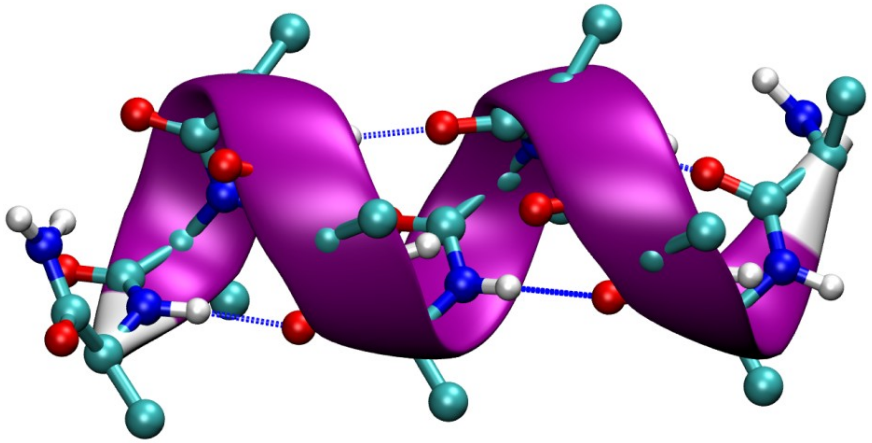
eABF: fluctuations

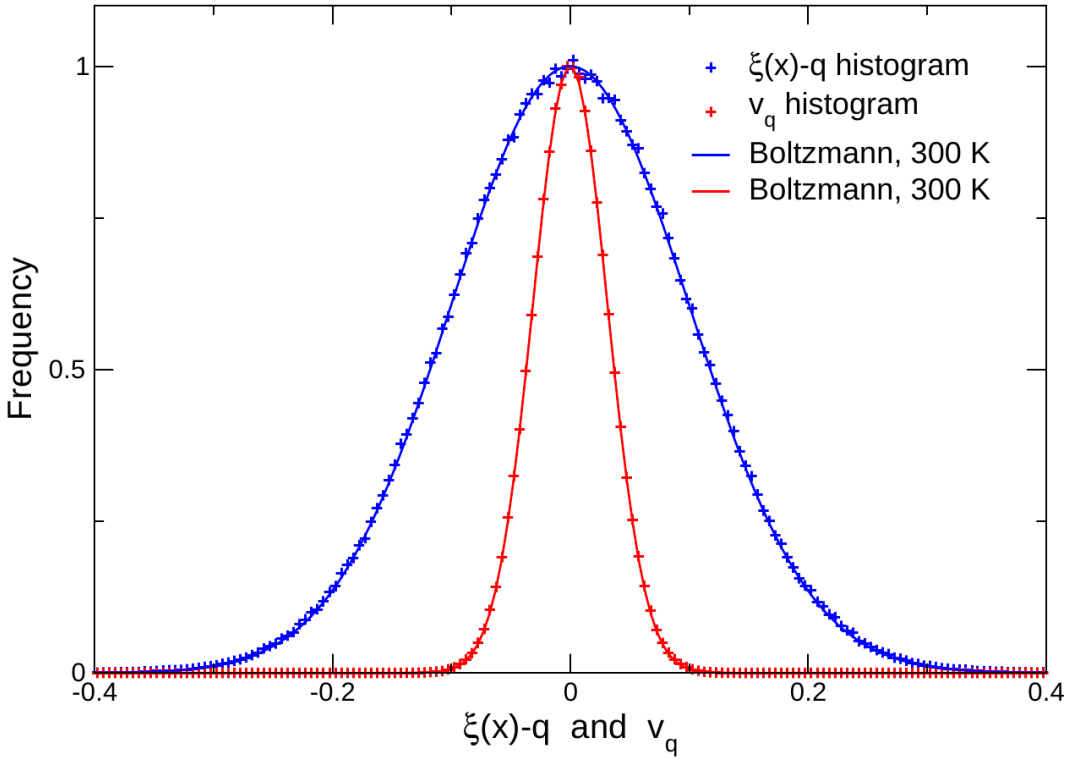


eABF: extended DOF thermalization

Do we need to thermostat the extended DOF?

- deca-alanine peptide in vacuum
- Langevin on atom DOFs, 300 K
- extended coordinate:
  peptide length



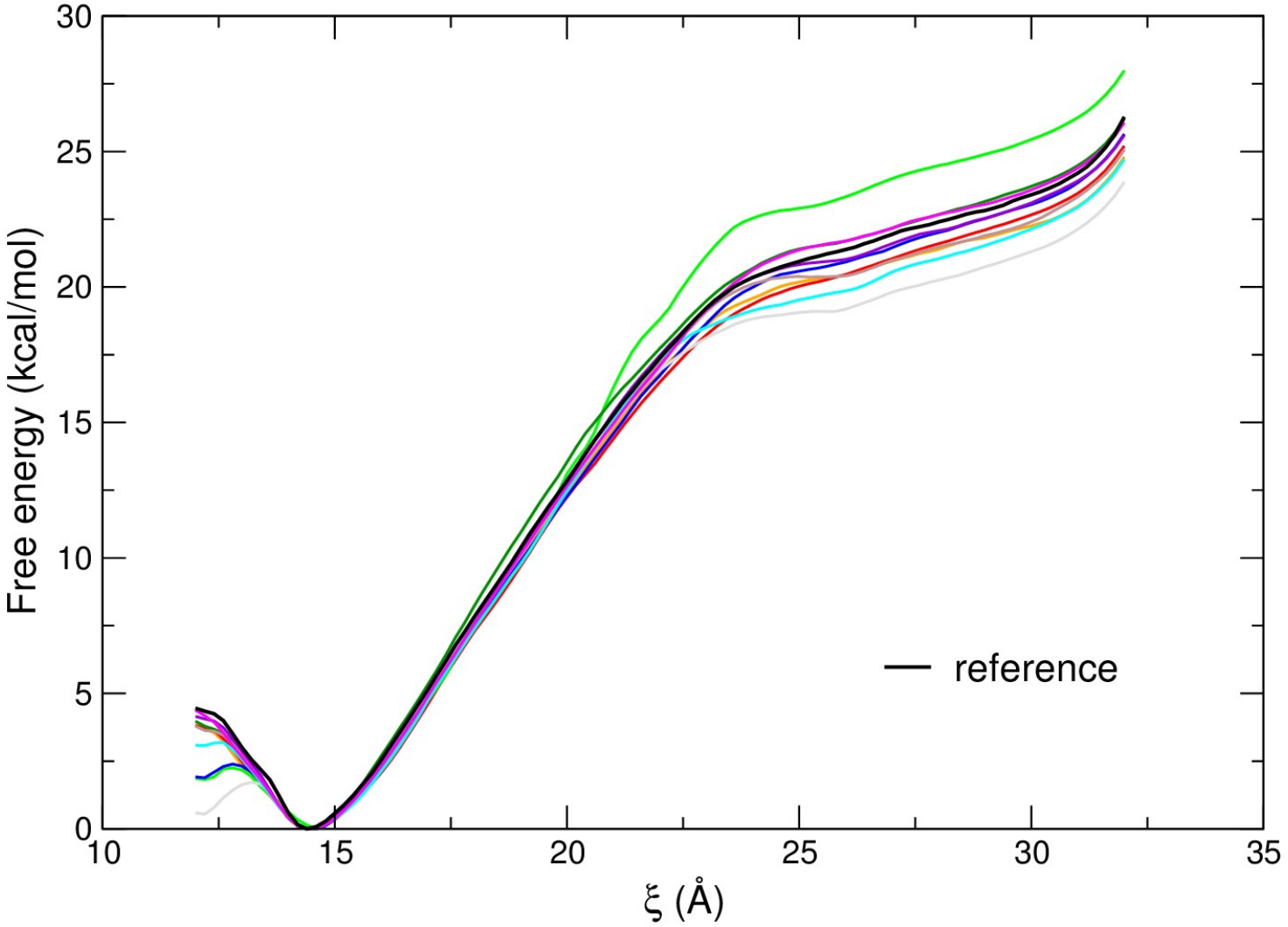


eABF: selling points

- Technical requirements: just the gradients (same as metadynamics)
  - No need to calculate Jacobian or second derivatives
  - No need to design explicit generalized coordinates
  - No need to choose “inverse gradient” field
  - No need to be orthogonal to constraints
  - No need for mutual orthogonality of variables

→ easily applicable to any combination of sophisticated variables

eABF PMF from naïve estimator



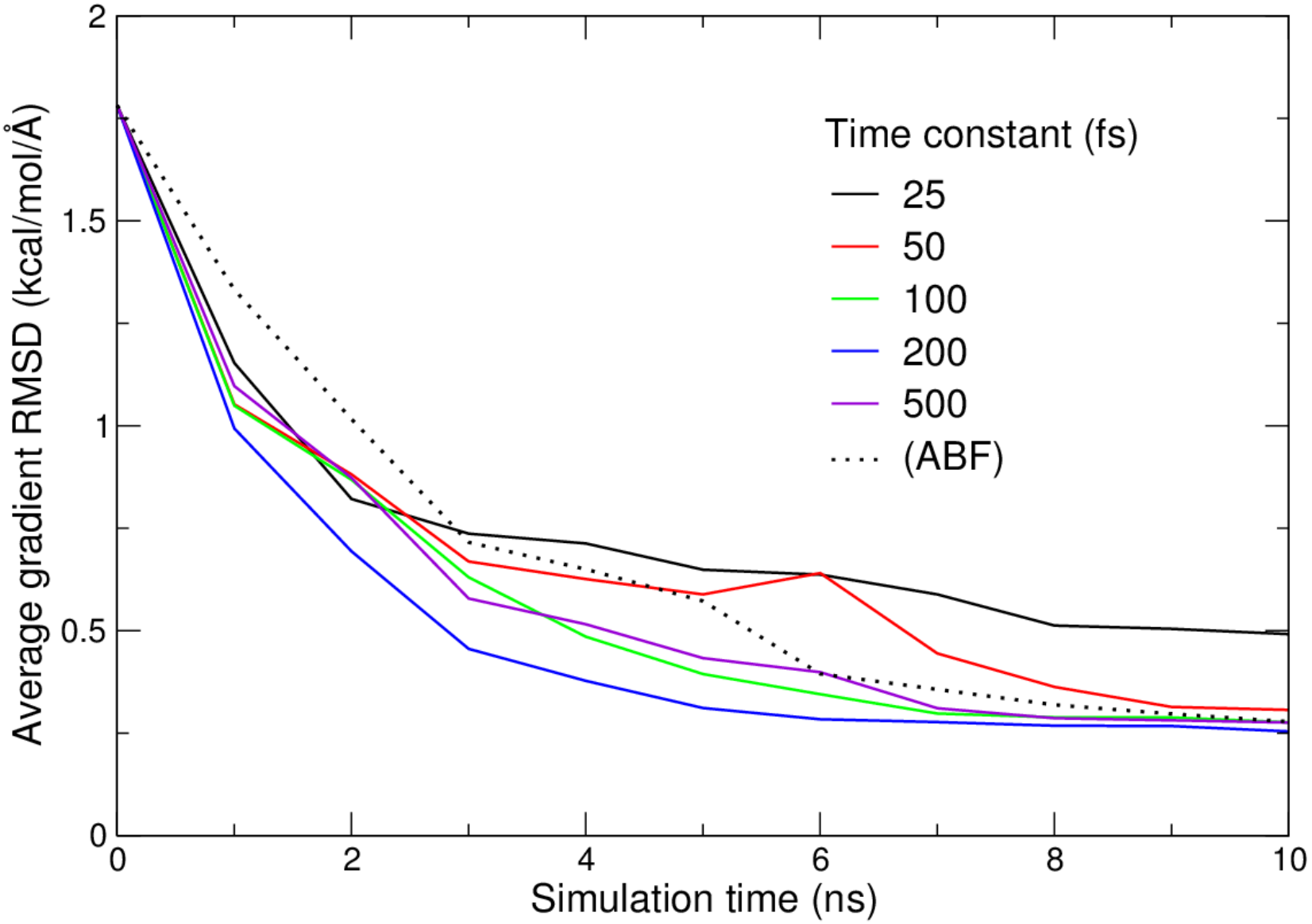
eABF: possible issues

- does the fictitious mass slow down diffusion?
- *not if oscillator time scale is small (typically less than 1 ps)*

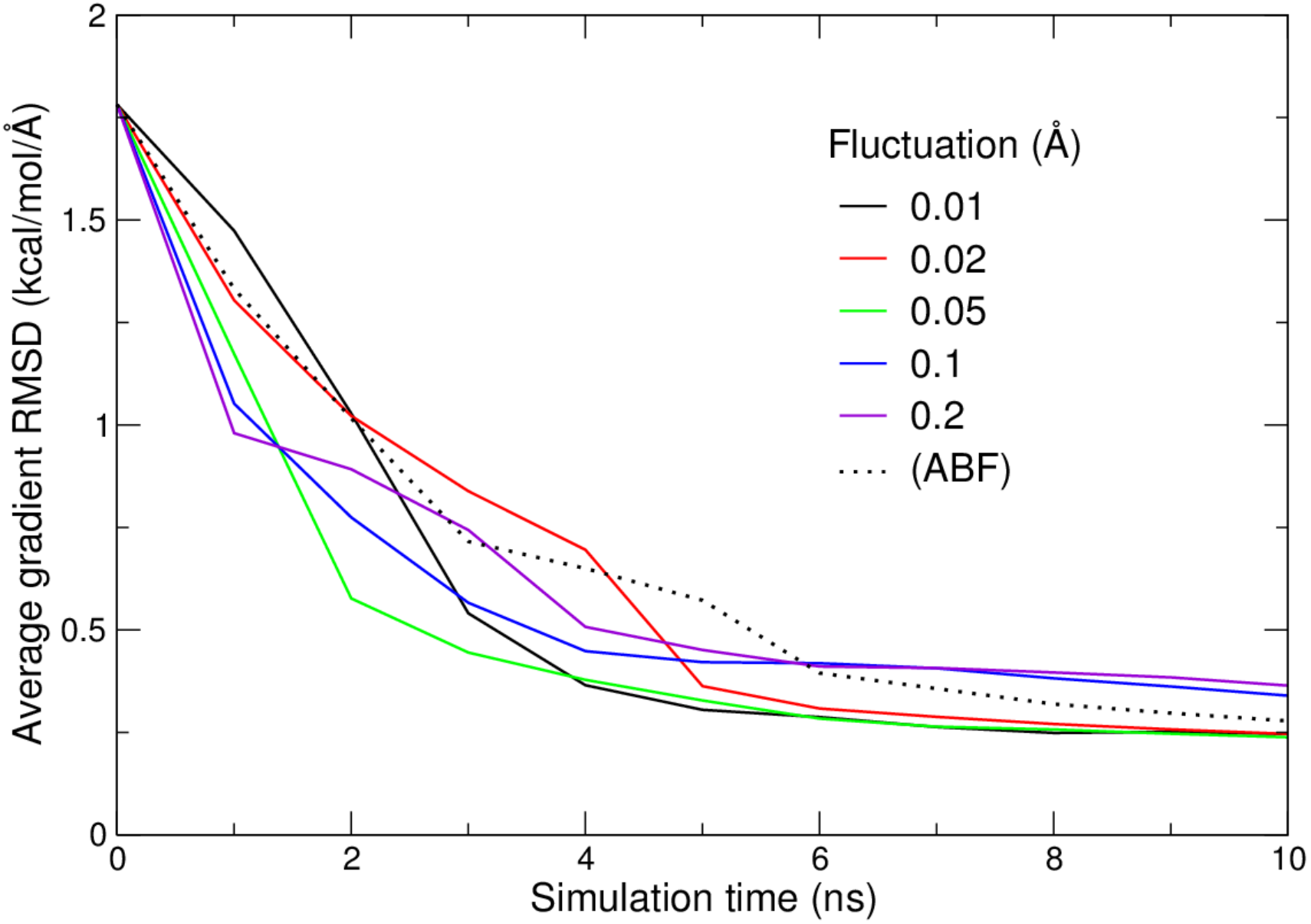
- is the PMF inaccurate?
- *it is a biased estimator, <u>but:</u>*
  - *the bias can be made very small with reasonable values of σ*
  - *there are other estimators*

- beneficial effect: force smoothing (variance reduction)

eABF: time scale and convergence



eABF: length scale and convergence



An unbiased estimator

- can we correct the measured PMF based on actual sampling?

$$\rho(\xi, q) = \rho^0(\xi) e^{-\frac{\beta k}{2}(\xi - q)^2} e^{-\beta \tilde{A}(q)}$$

$$\rho(\xi) \propto \rho^0(\xi) \int e^{-\frac{\beta k}{2}(\xi - q\prime)^2} e^{-\beta \tilde{A}(q\prime)} dq\prime$$

$$\frac{d \ln \rho(\xi)}{d\xi} = \frac{d \ln \rho^0(\xi)}{d\xi} - \frac{\int \rho(\xi, q\prime) k(\xi - q\prime) dq\prime}{\int \rho(\xi, q\prime) dq\prime}$$

$$\frac{dA}{d\xi} = k_B T \frac{d \ln \rho(\xi)}{d\xi} - k(\xi - \bar{q}(\xi))$$

- asymptotically unbiased
- convergence is poorer than naïve estimator!

A *q*-centric perspective

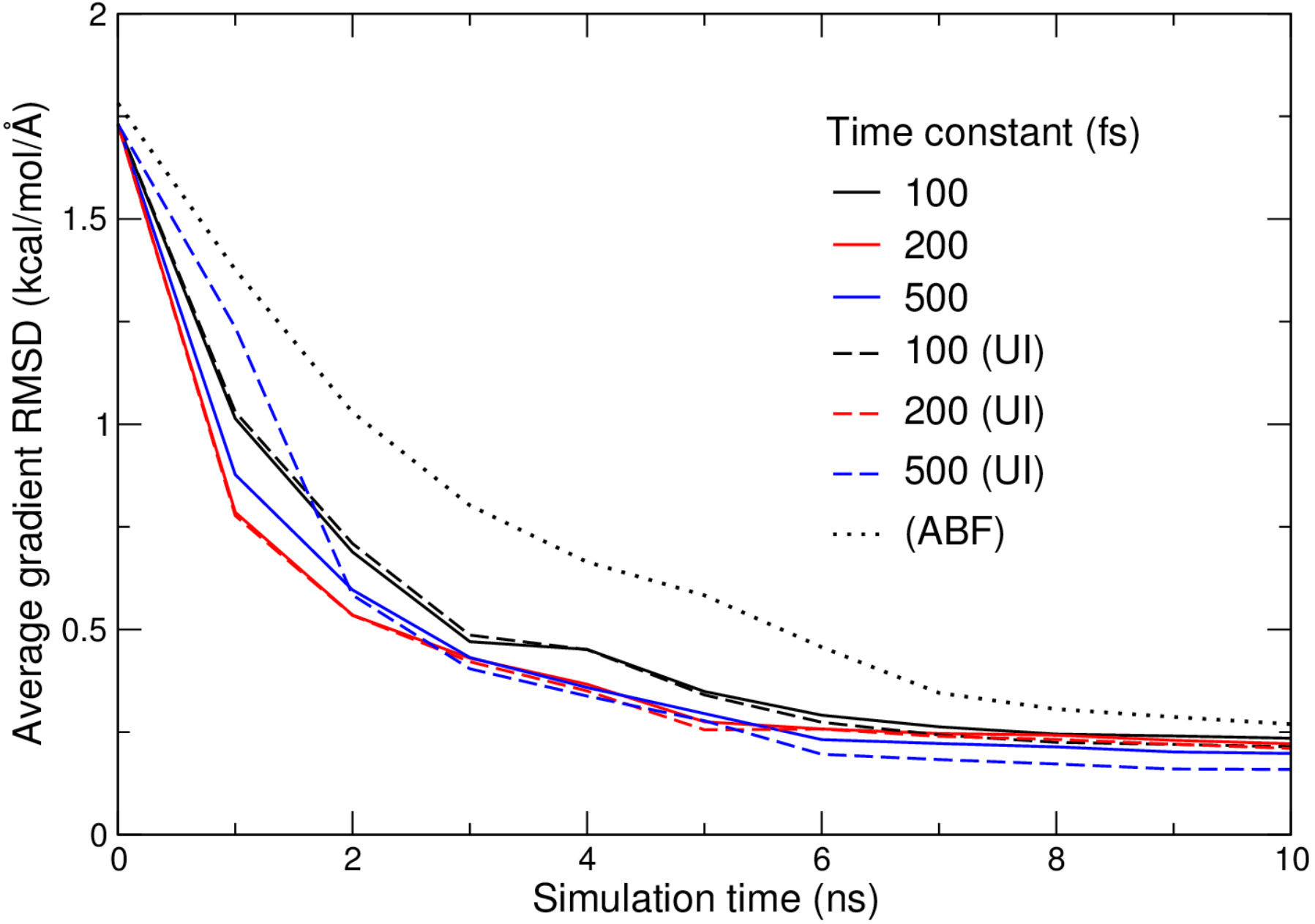
- proposed by Wei Yang, based on Umbrella Integration (Kästner and Thiel)
- distribution of ξ at each q value viewed as an umbrella sampling histogram

$$\rho(\xi, q) = \rho^0(\xi) e^{-\frac{\beta k}{2}(\xi - q)^2} e^{-\beta \tilde{A}(q)}$$

$$\forall q, \quad \frac{dA}{d\xi} = -k_B T \frac{d \ln \rho(\xi, q)}{d\xi} - k(\xi - q)$$

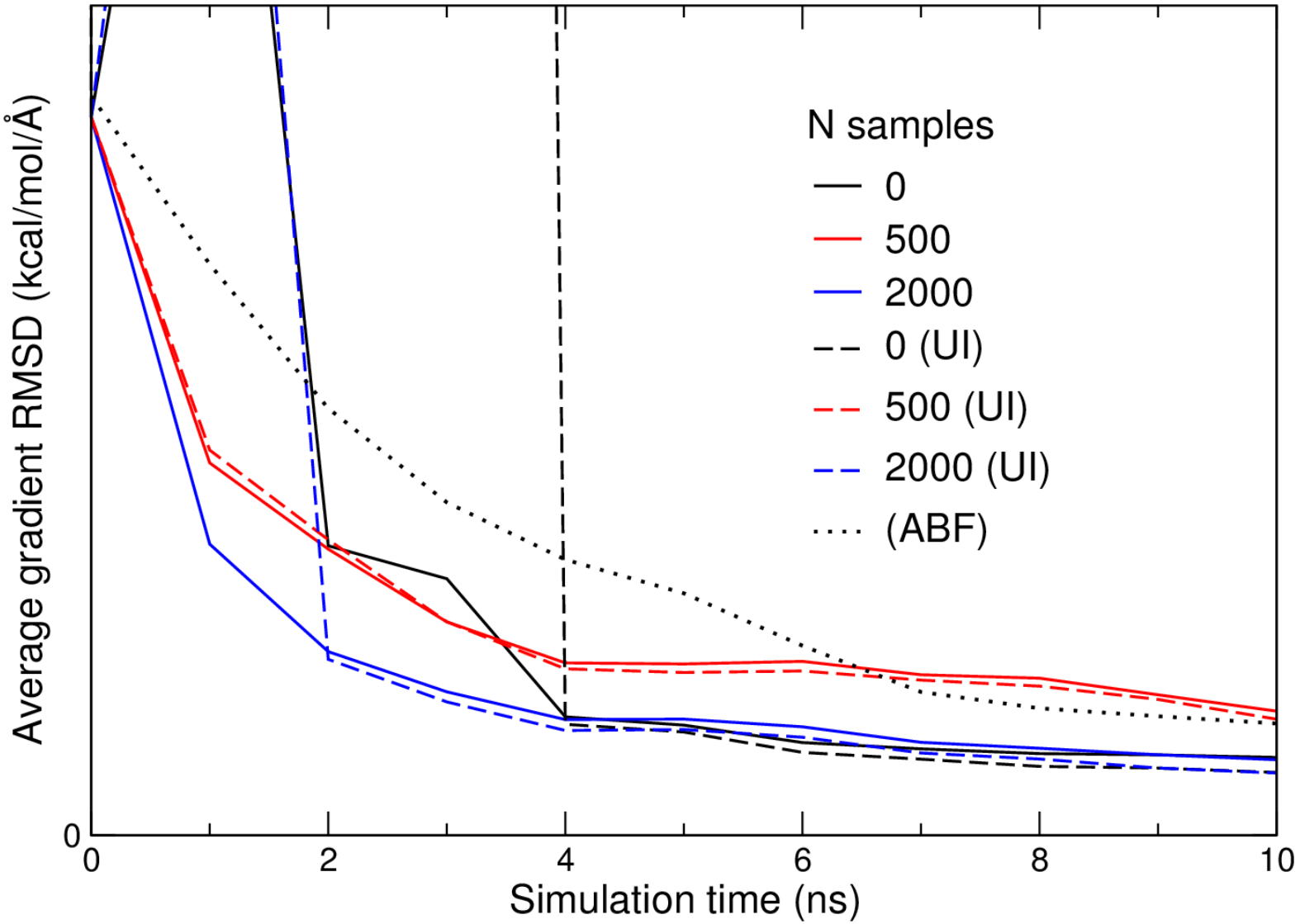
- exploit approximately Gaussian distribution: $\rho(\xi, q) \approx e^{-\frac{1}{2}\left(\frac{\xi - \bar{\xi}_q}{\sigma_q}\right)^2}$
- combine histograms for all q values

UI estimator convergence: time scale

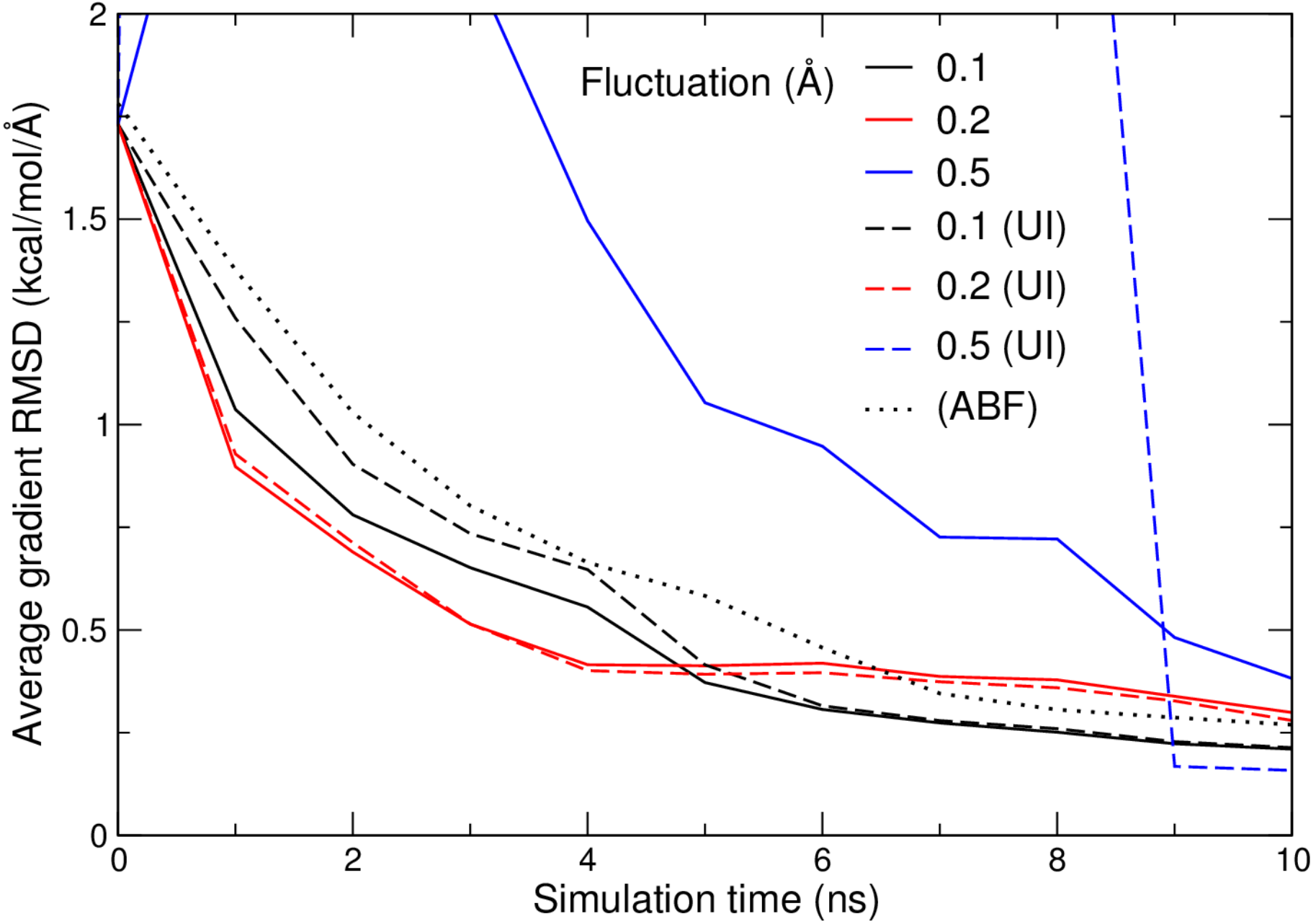


UI estimator convergence: delayed bias

Number of samples per bin before full ABF bias is applied



UI estimator convergence: length scale

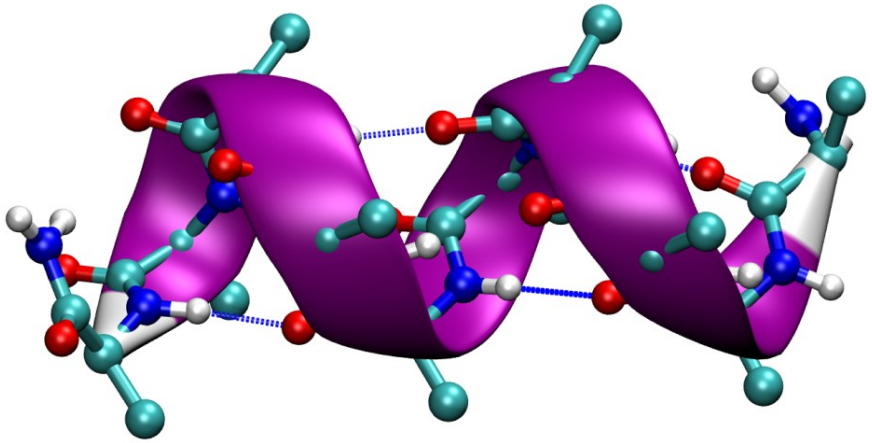


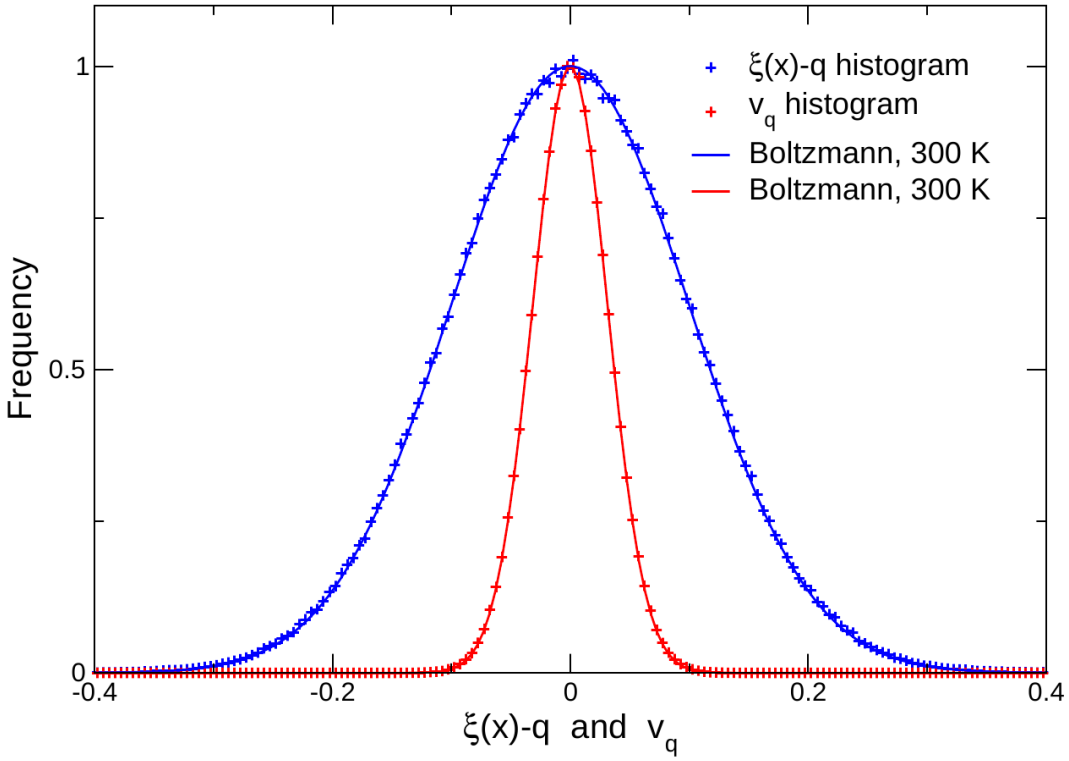
eABF: extended DOF thermalization

Do we need to thermostat the extended DOF?

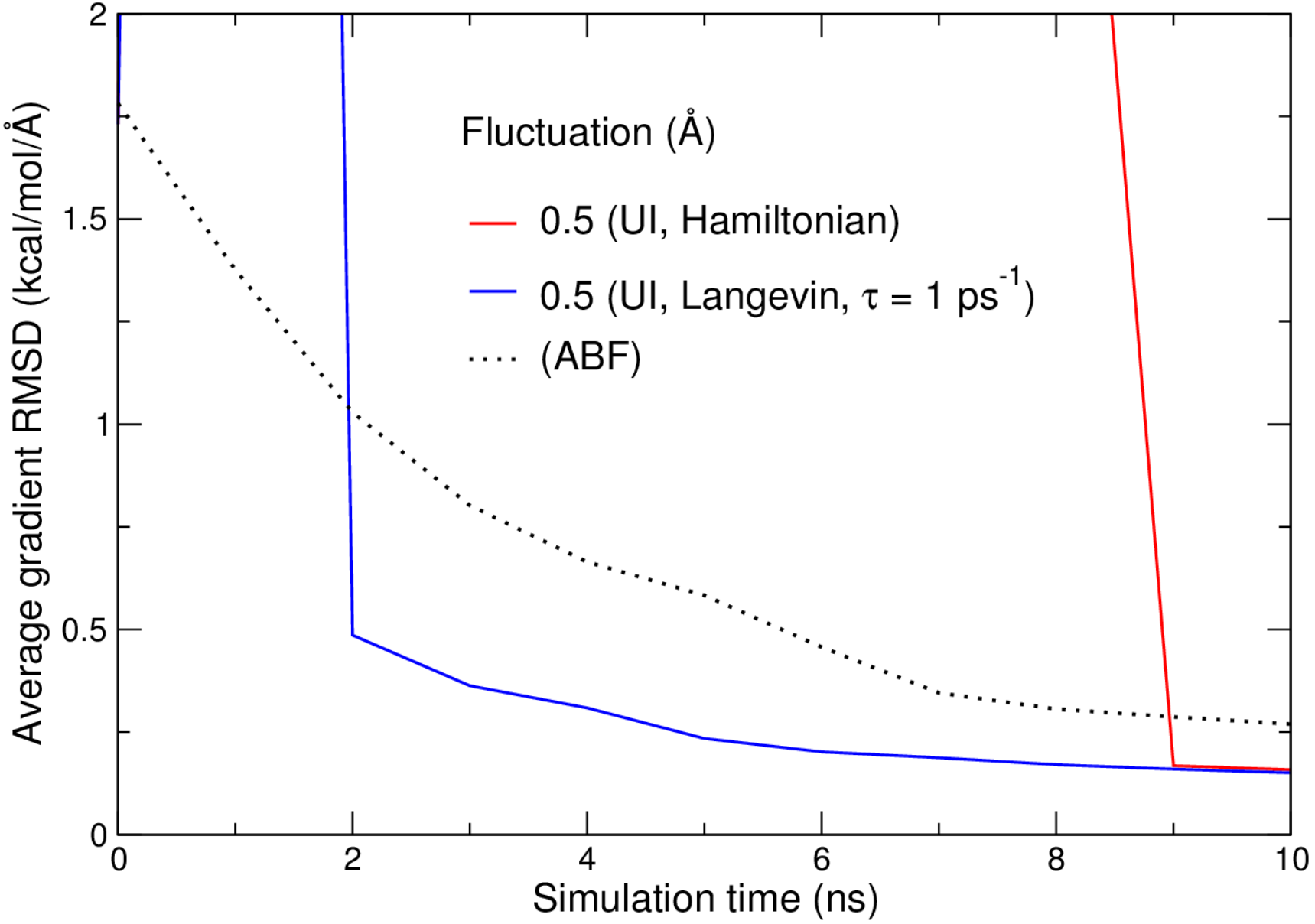
Maybe we do.

- deca-alanine peptide in vacuum
- Langevin on atom DOFs, 300 K
- extended coordinate:
  peptide length





eABF with Langevin dynamics



Dimension reduction: the next frontier?

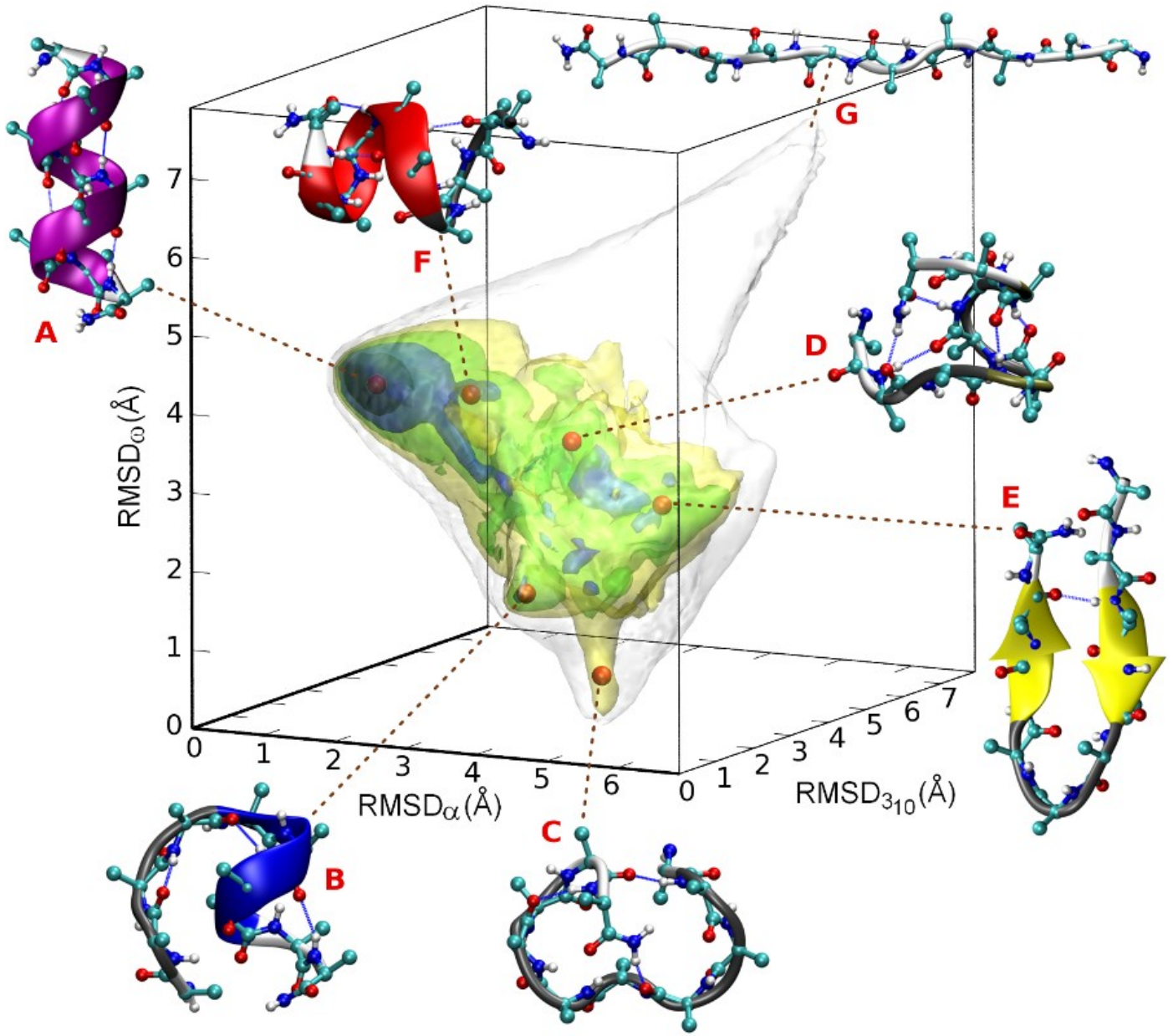
Designing a low-dimension model is hard

Sometimes intuitive coordinates are just not good enough.

(Hénin et al. JCTC 2010)

How many are needed?

(intrinsic dimension)



Discovering descriptive coordinates

- target: describe peptide/protein conformation changes
- principal components analysis (PCA)

- normal mode analysis from harmonic model: easy, no prior sampling
- dihedral PCA (Altis et al. 2007)
  - problem: loss of resolution when bonded distance increases
- distance map PCA
  - problem: high dimension!
- contact map PCA
  - worth trying?

Algorithms for dimension reduction

Assumption: low-dimension object embedded in high-dimension space

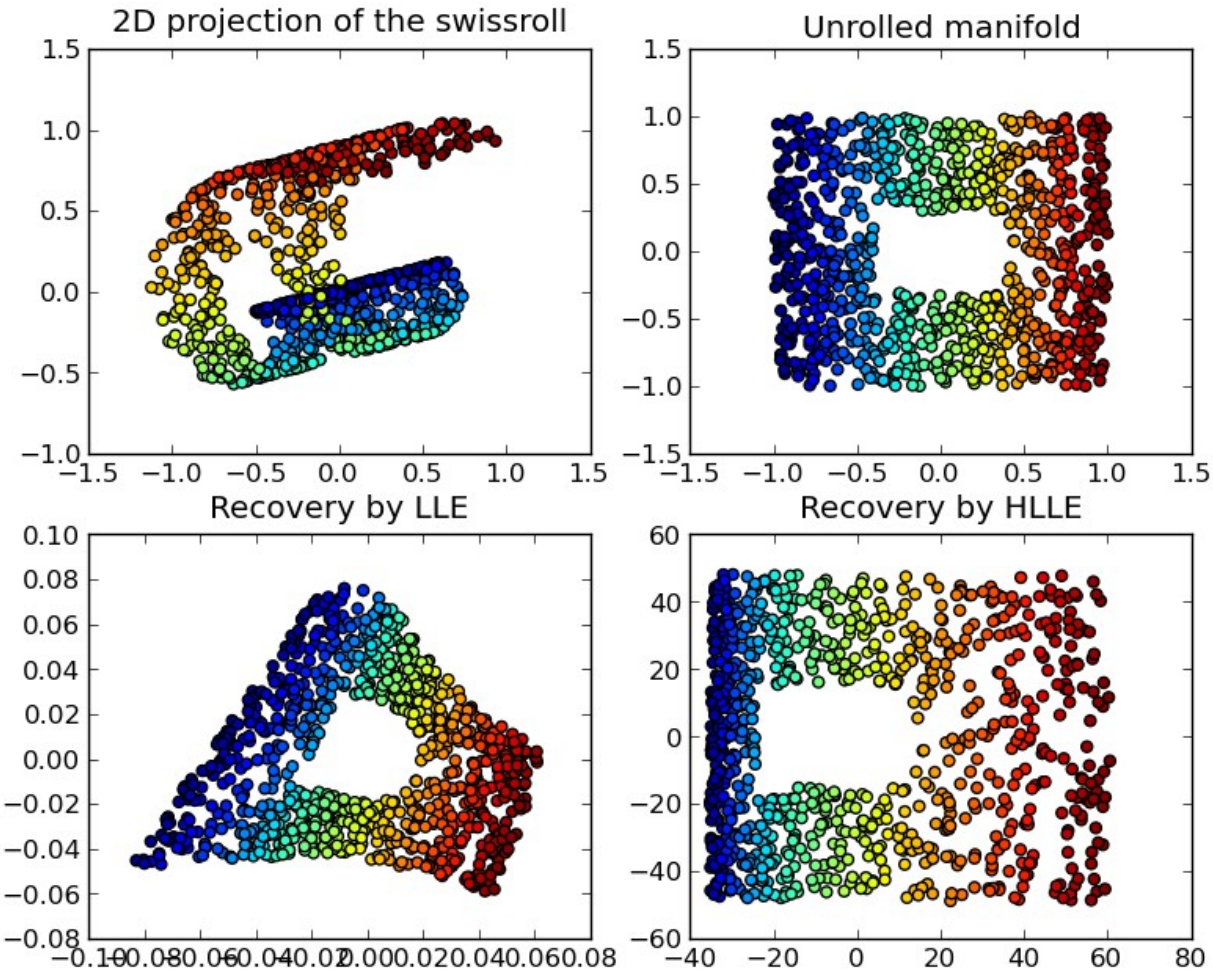


image by Olivier Grisel

THE JOURNAL OF CHEMICAL PHYSICS **134**, 124116 (2011)

Determination of reaction coordinates via locally scaled diffusion map

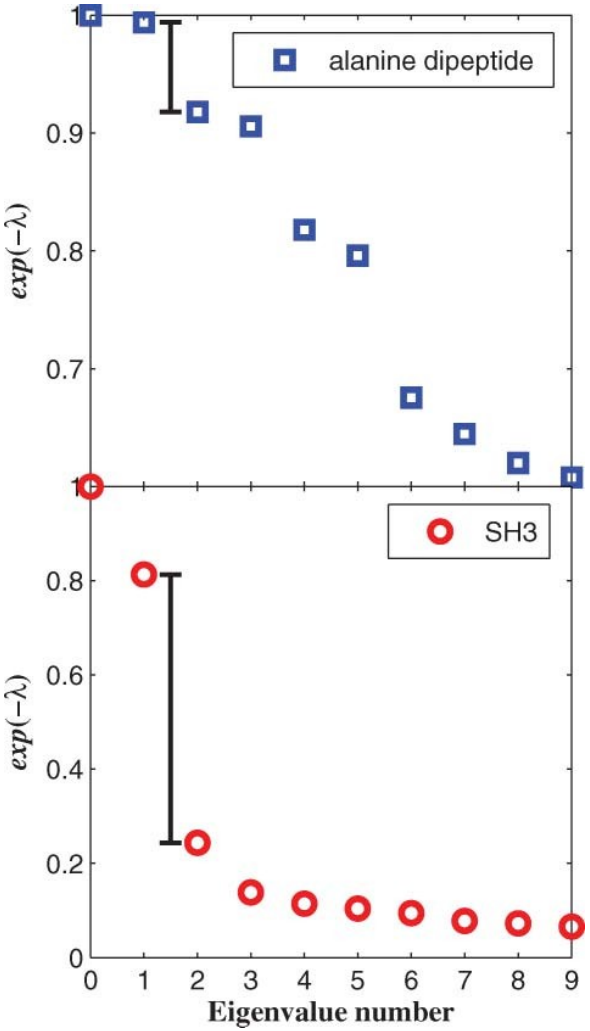
Mary A. Rohrdanz,[1] Wenwei Zheng,[1] Mauro Maggioni,[2] and Cecilia Clementi[1,a)]
[1]*Rice University, Department of Chemistry, Houston, Texas 77005, USA*
[2]*Duke University, Department of Mathematics, Durham, North Carolina 27708, USA*

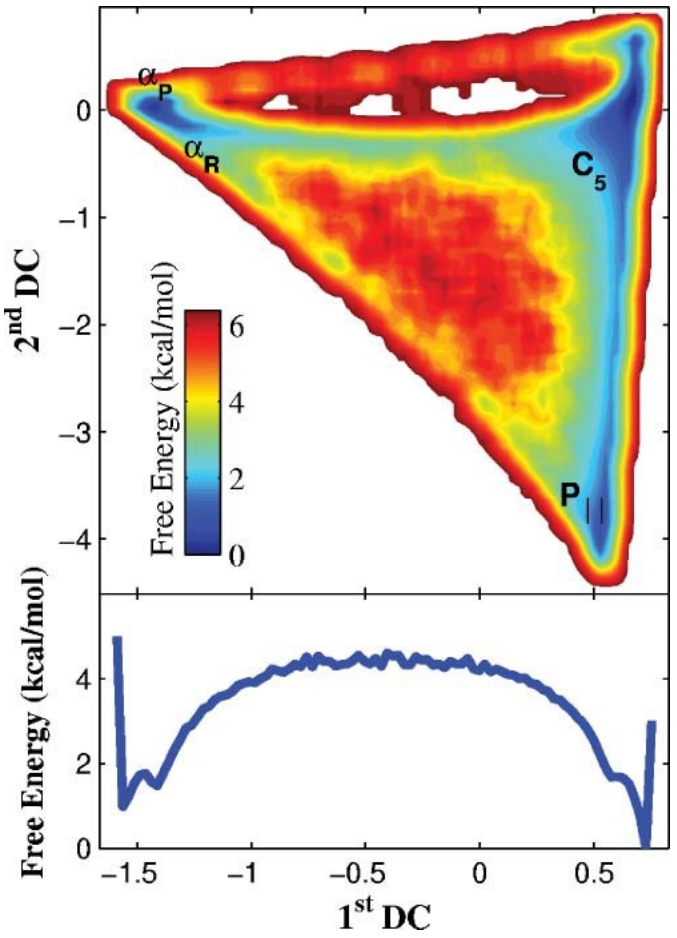
- Goal: approximate solution to the Fokker-Planck equation, as:

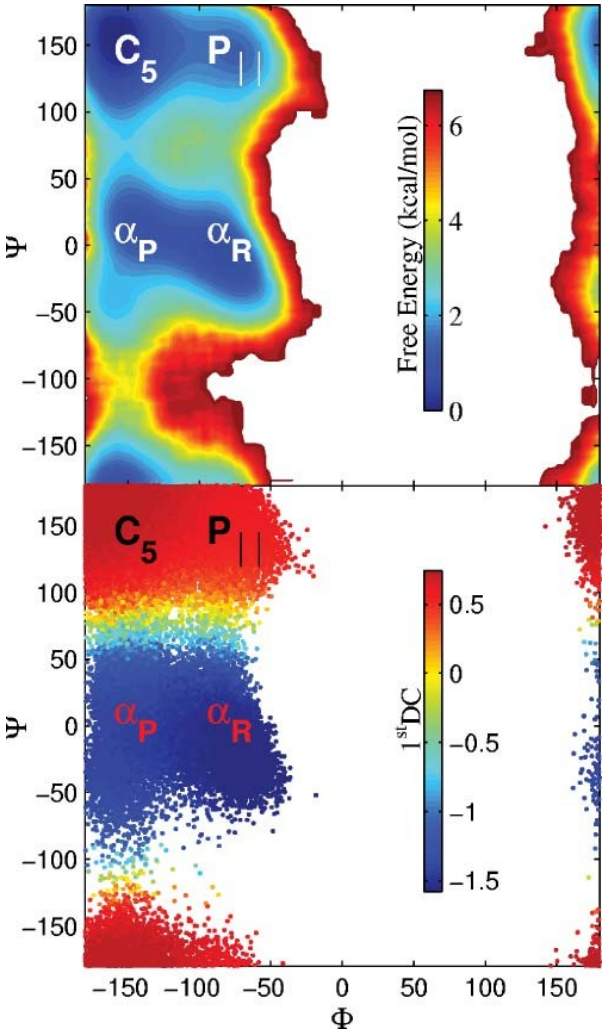
$$p(\boldsymbol{x}, t) = \phi_0(\boldsymbol{x}) + \sum_{i=1}^{k} c_i \phi_i(\boldsymbol{x}) e^{-\lambda_i t}$$

- start from Boltzmann-distributed samples
- calculate transition matrix (with local scale): $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\varepsilon_i \varepsilon_j}\right)$
- diagonalize
- eigenvectors $\phi_i(\boldsymbol{x})$ are *diffusion coordinates*

Alanine dipeptide example

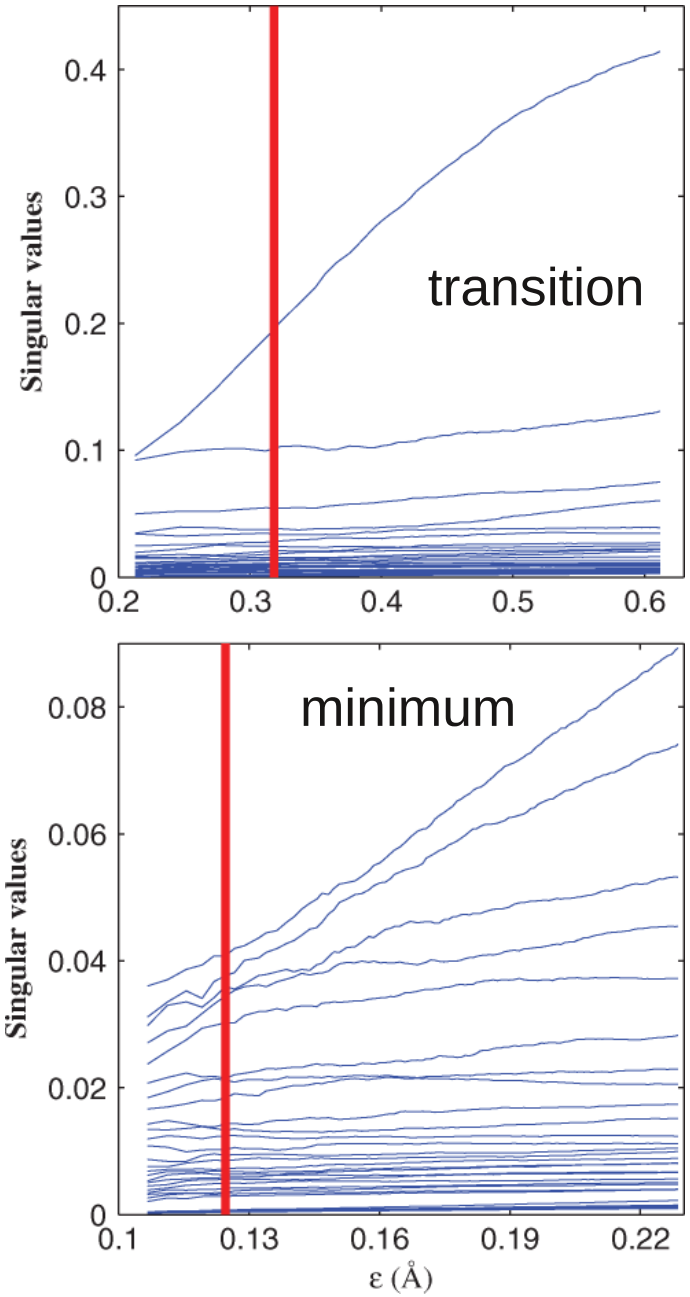


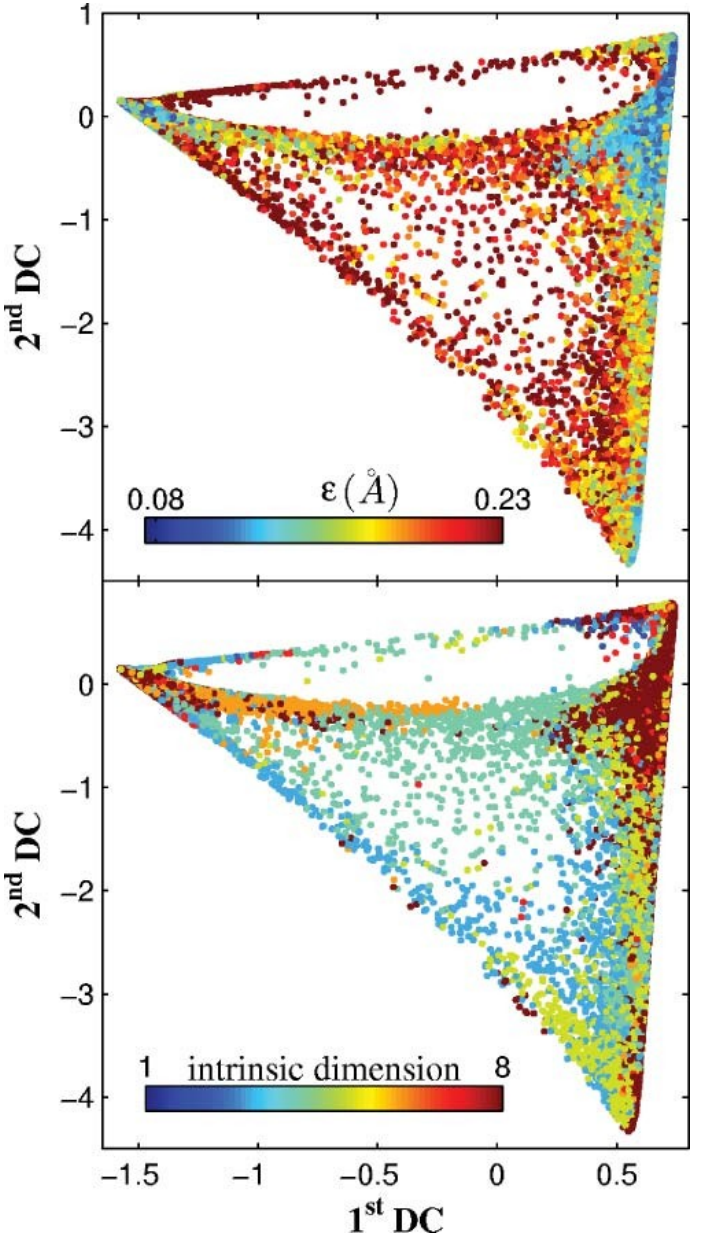




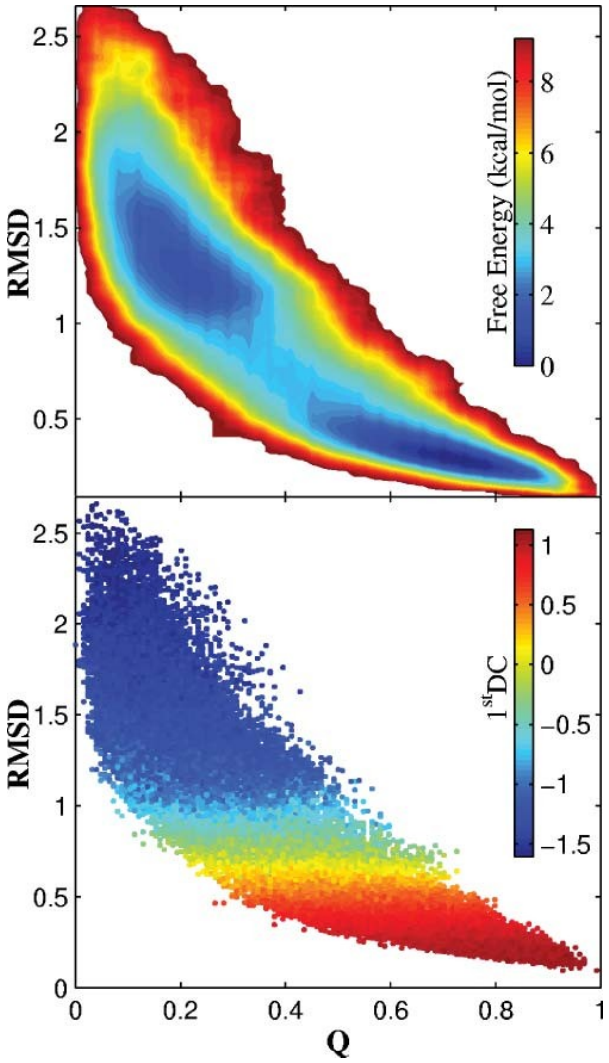
Local intrinsic dimension

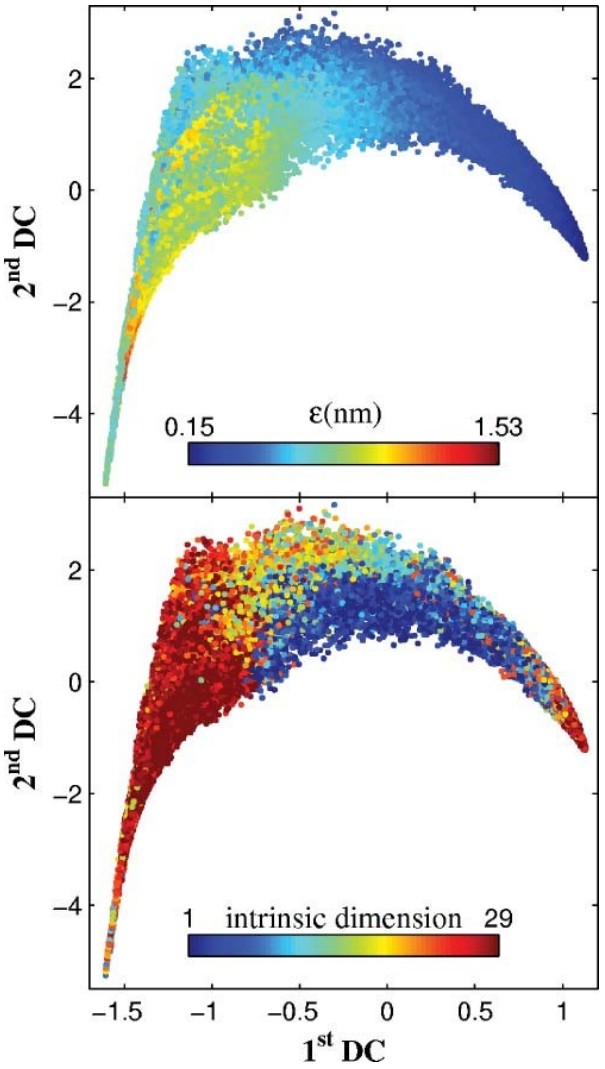
multidimensional scaling (MDS)

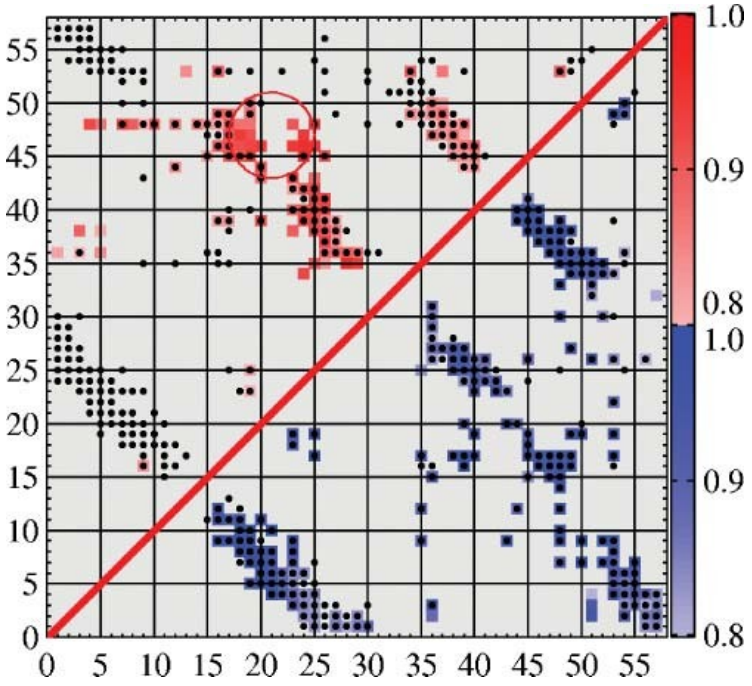




SH3 domain example







correlation with native contacts

Applicability of LSDMap?

- only an analysis of previous sampling
- recursive approach possible (bias, sample, analyze, repeat)

- diffusion coordinates are defined implicitly on sampled points
- differentiable extension?


Thank you