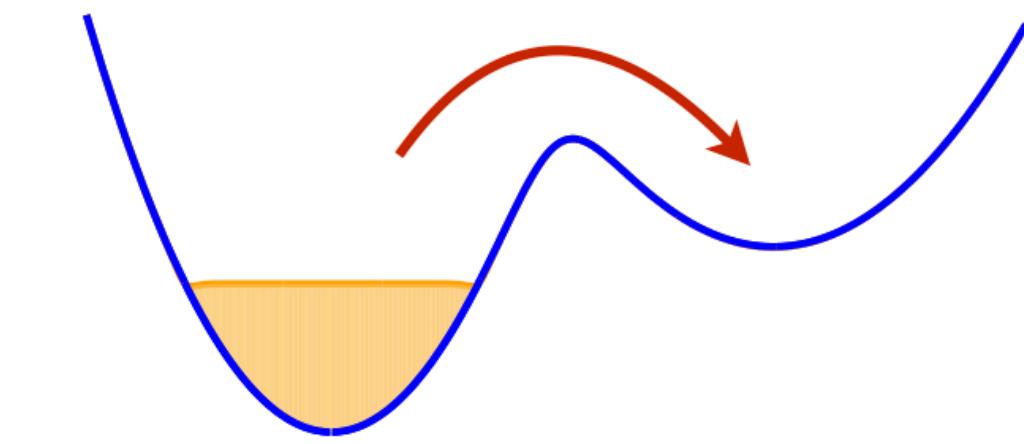
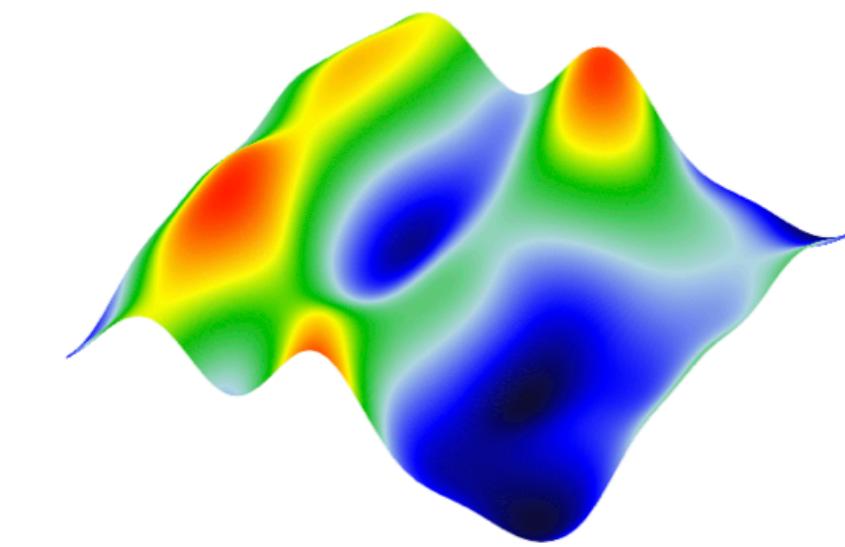
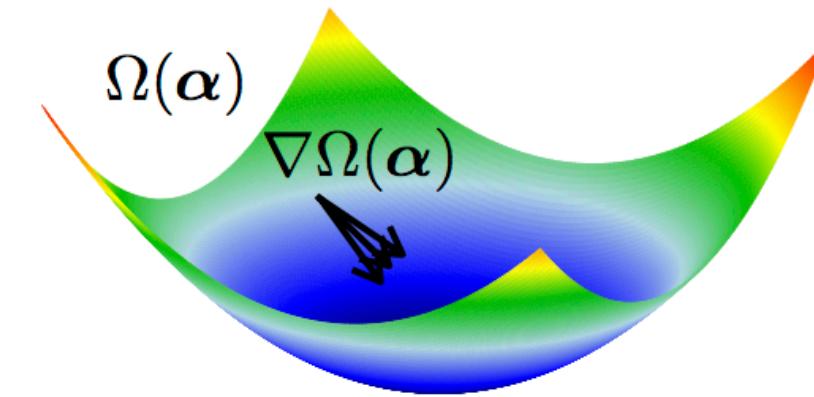


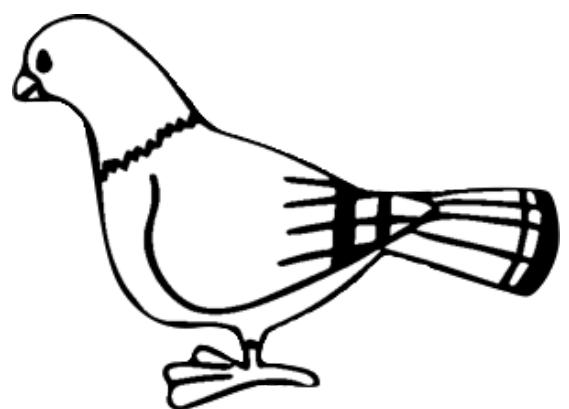
Variationally Enhanced Sampling

$$\Omega[V] = \frac{1}{\beta} \log \frac{\int d\mathbf{s} e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s} e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s} p(\mathbf{s})V(\mathbf{s})$$

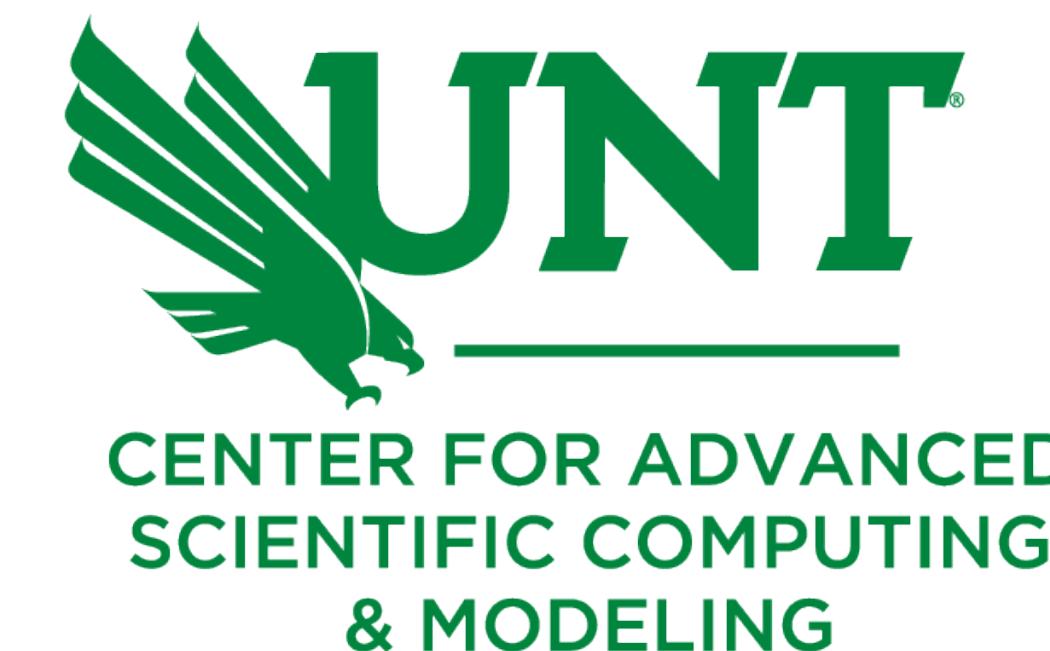


Dr. Omar Valsson

University of North Texas, Denton, TX, USA

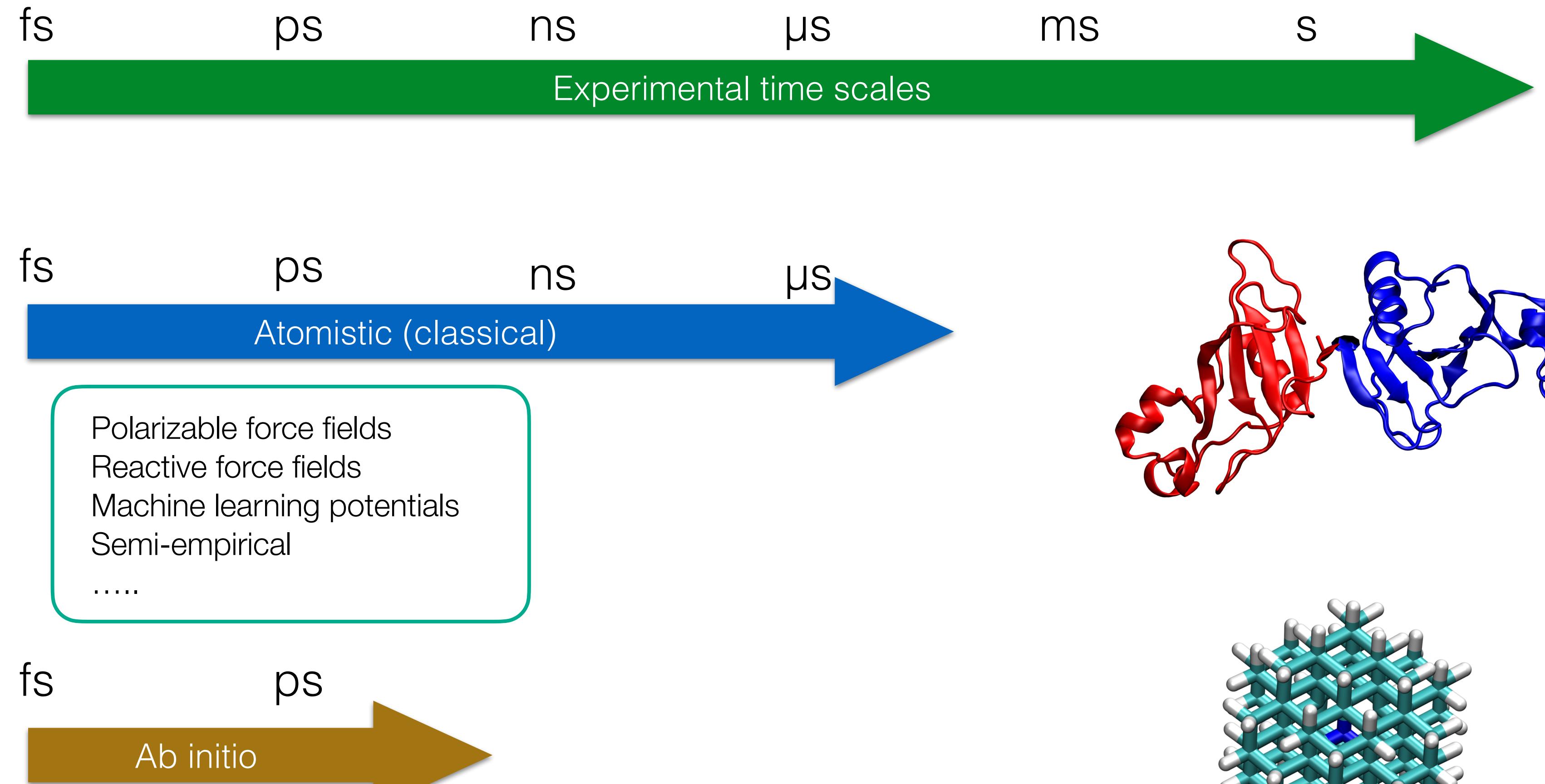


PLUMED Masterclass-22-11
July 4, 2022



www.valsson.info
[@OmarValsson](https://twitter.com/OmarValsson)
omar.valsson@unt.edu

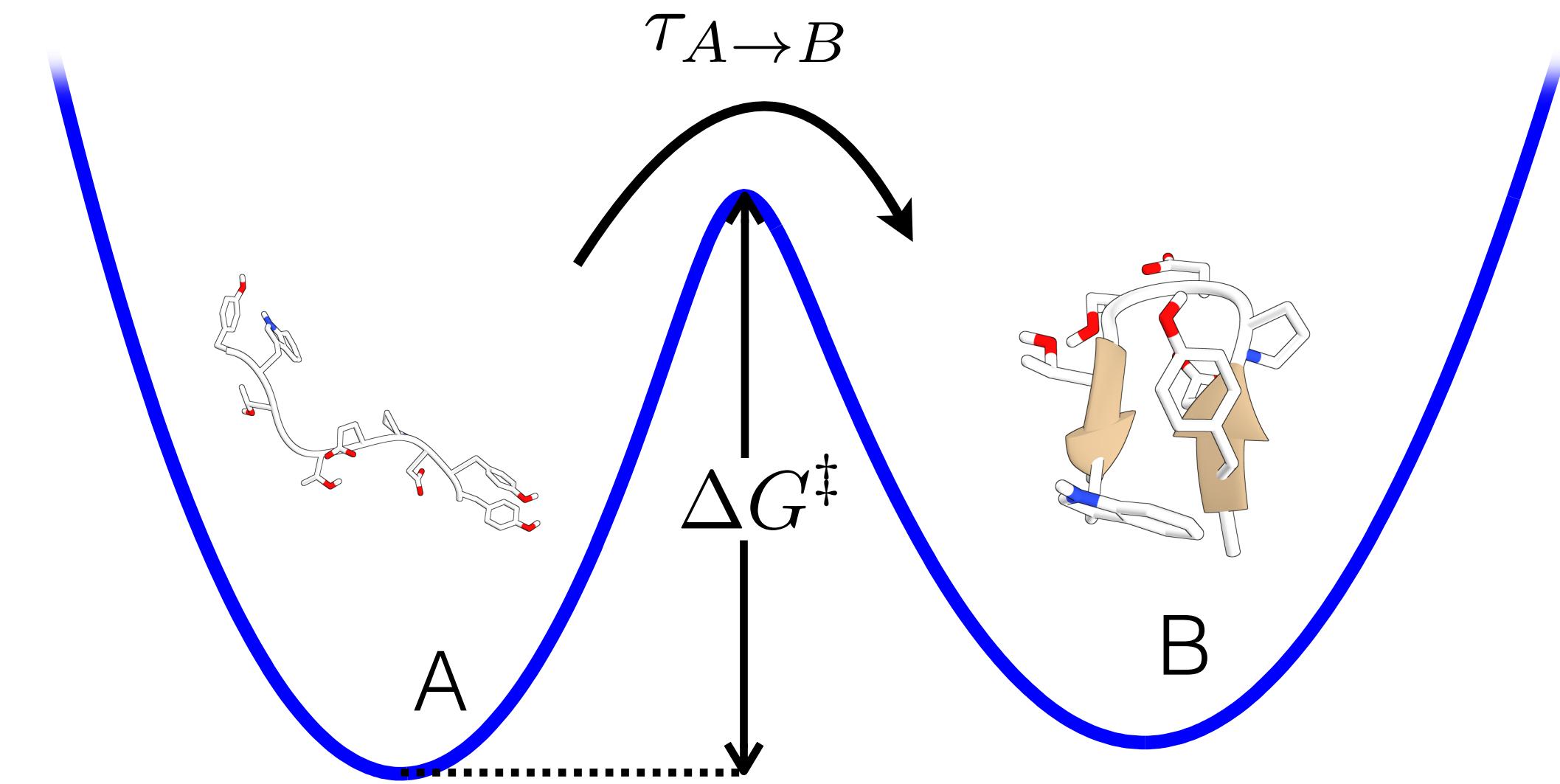
Time Scale Problem in Atomistic Simulations



Disparity between the fastest processes and the long-time scale phenomenon we are interesting in

Time Scale Problem in Atomistic Simulations

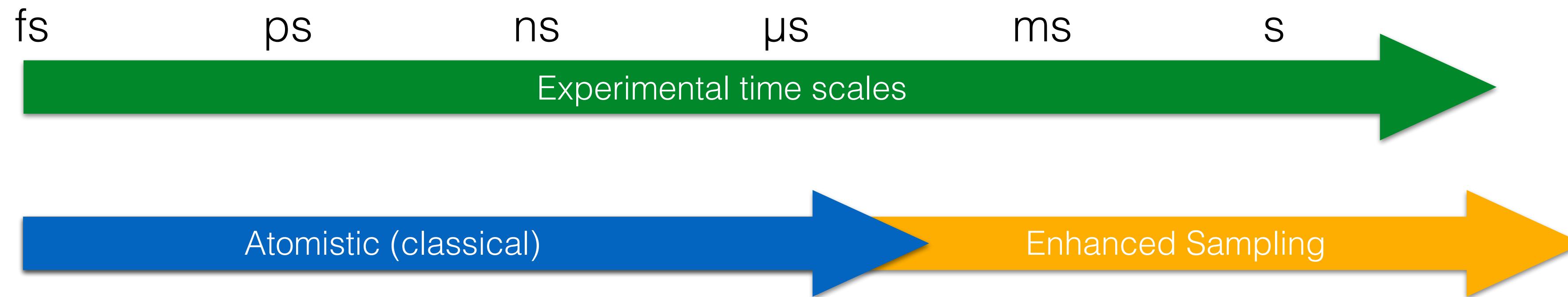
Physical system characterized by many metastable states
separated by high free energy barriers



$$\tau_{A \rightarrow B} = \frac{1}{\nu_0 e^{-\Delta G^\ddagger/k_B T}} \propto e^{+\Delta G^\ddagger/k_B T}$$

$\Delta G^\ddagger \gg k_B T \rightarrow$ trapped in a metastable state

Advanced sampling method needed to overcome barriers and bridge time scales

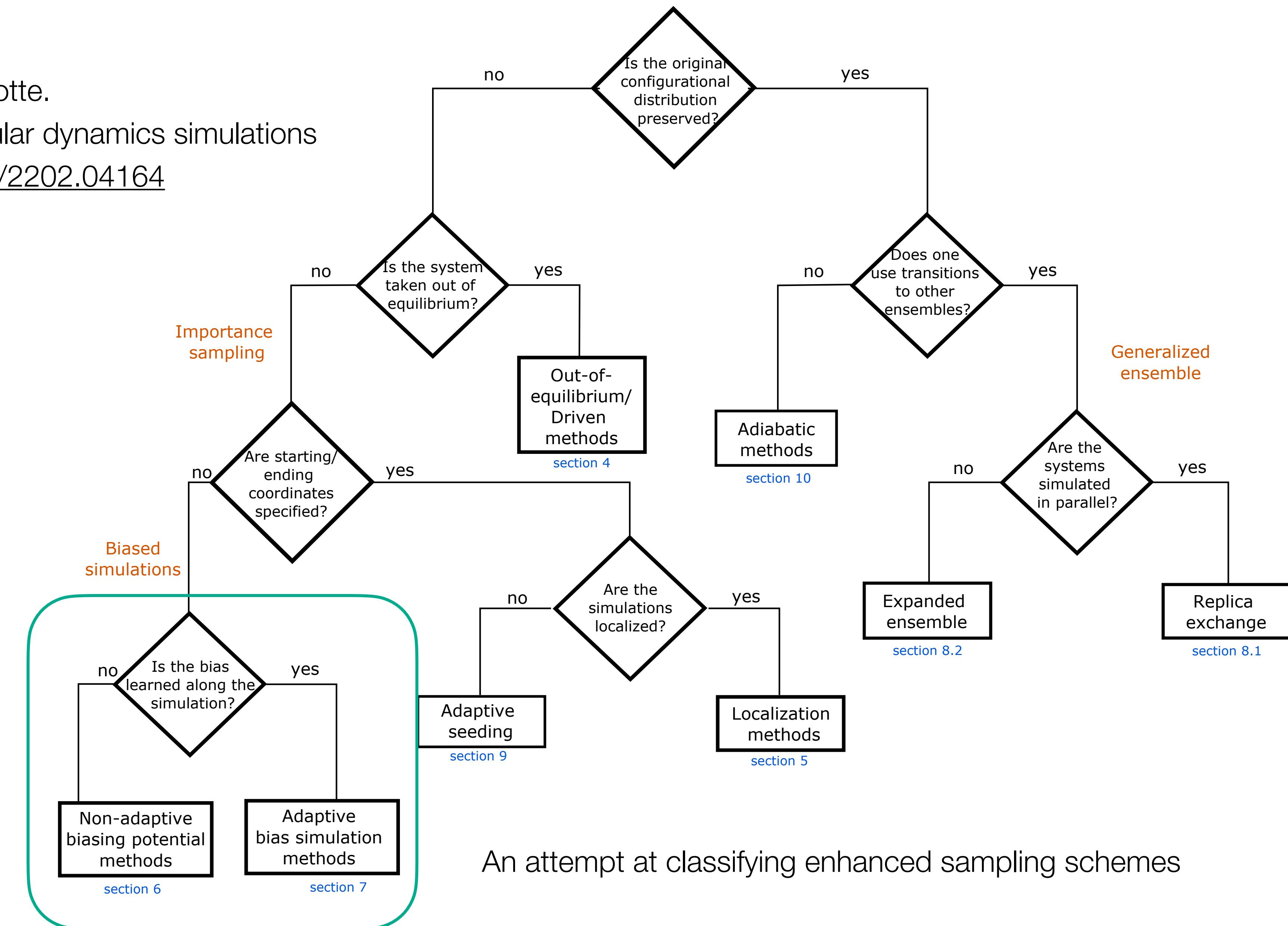


Wide range of enhanced sampling methods available

Hénin, Lelièvre, Shirts, Valsson & Deleuze.

Enhanced sampling methods for molecular dynamics simulations

arXiv:2202.04164 - <https://arxiv.org/abs/2202.04164>



Mapping the Problem to a Lower Dimension

Identify few important coarse-grained variables that describe the problem

- Distinguish the relevant metastable states
- Include the slow degrees of freedom

Large number of generally applicable CVs available
e.g. distances, dihedral angles, coordination numbers, contacts maps, RMSD, path CV,

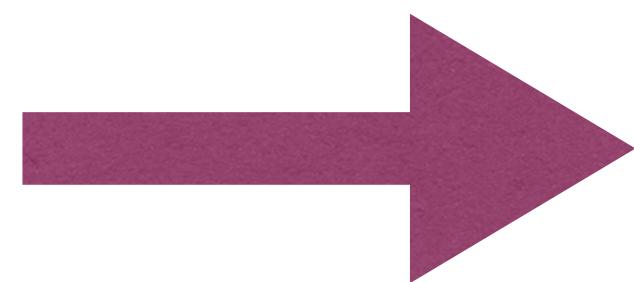
So-called **Collective Variables (CVs)** (or reaction coordinates)

$$\mathbf{s}(\mathbf{R}) = [s_1(\mathbf{R}), s_2(\mathbf{R}), \dots, s_d(\mathbf{R})]$$

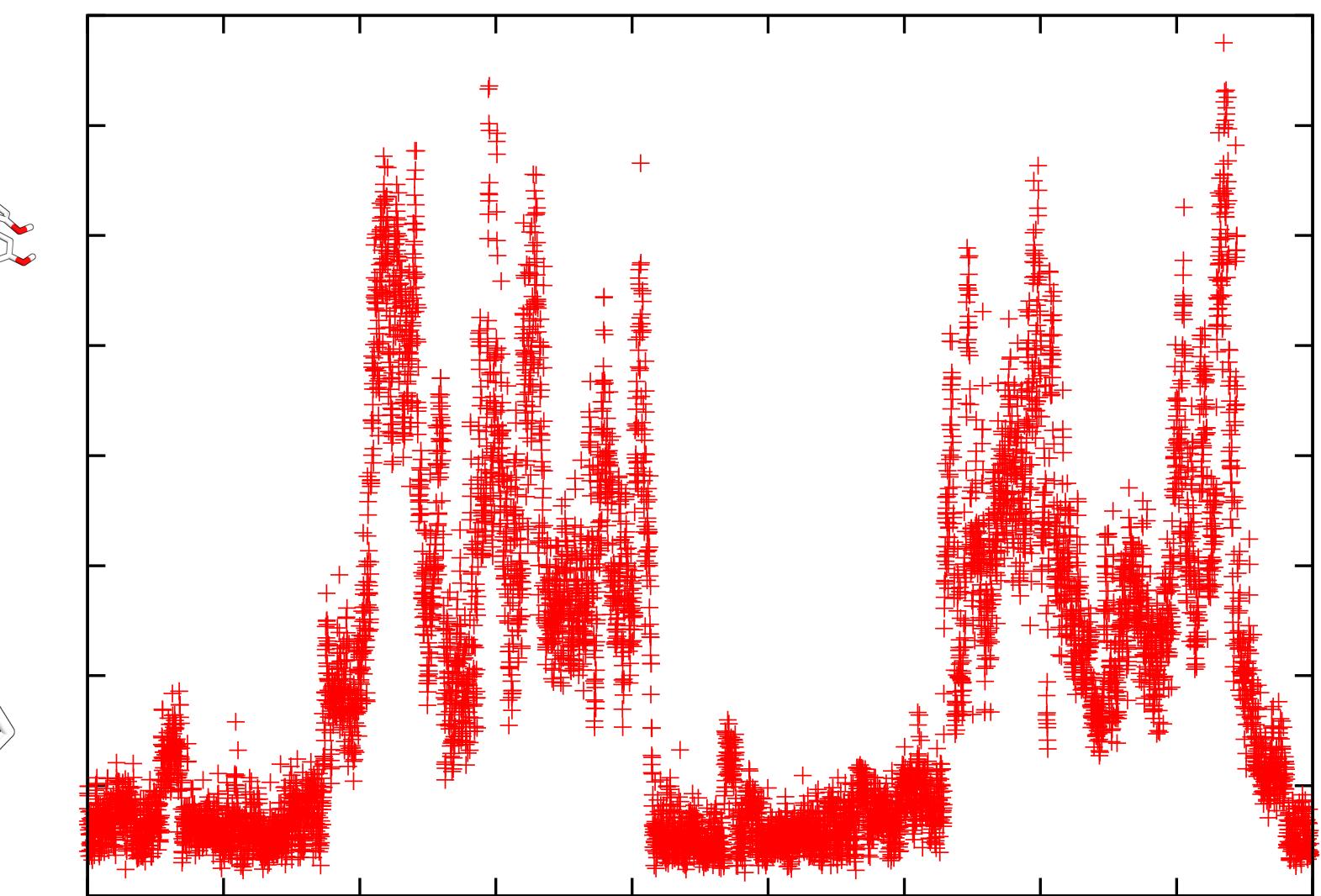
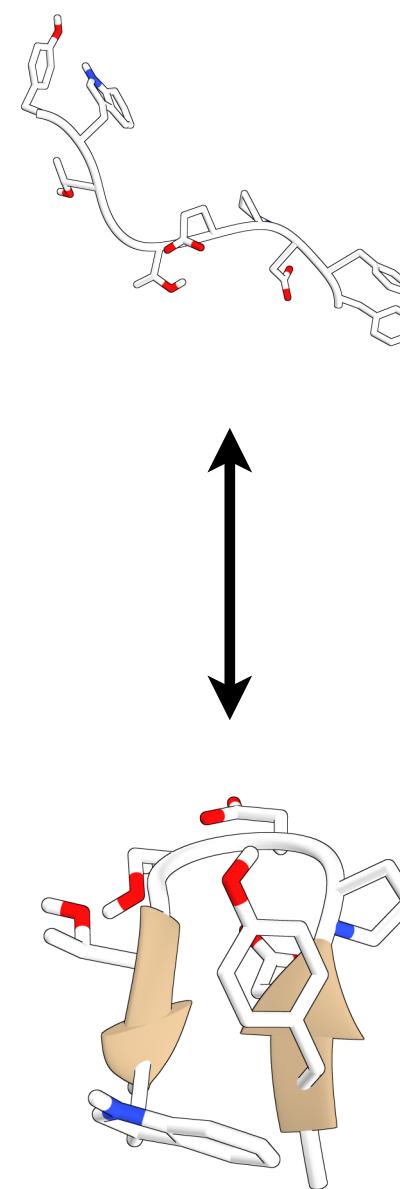
functions of the atomistic coordinates \mathbf{R}



Collective Variables

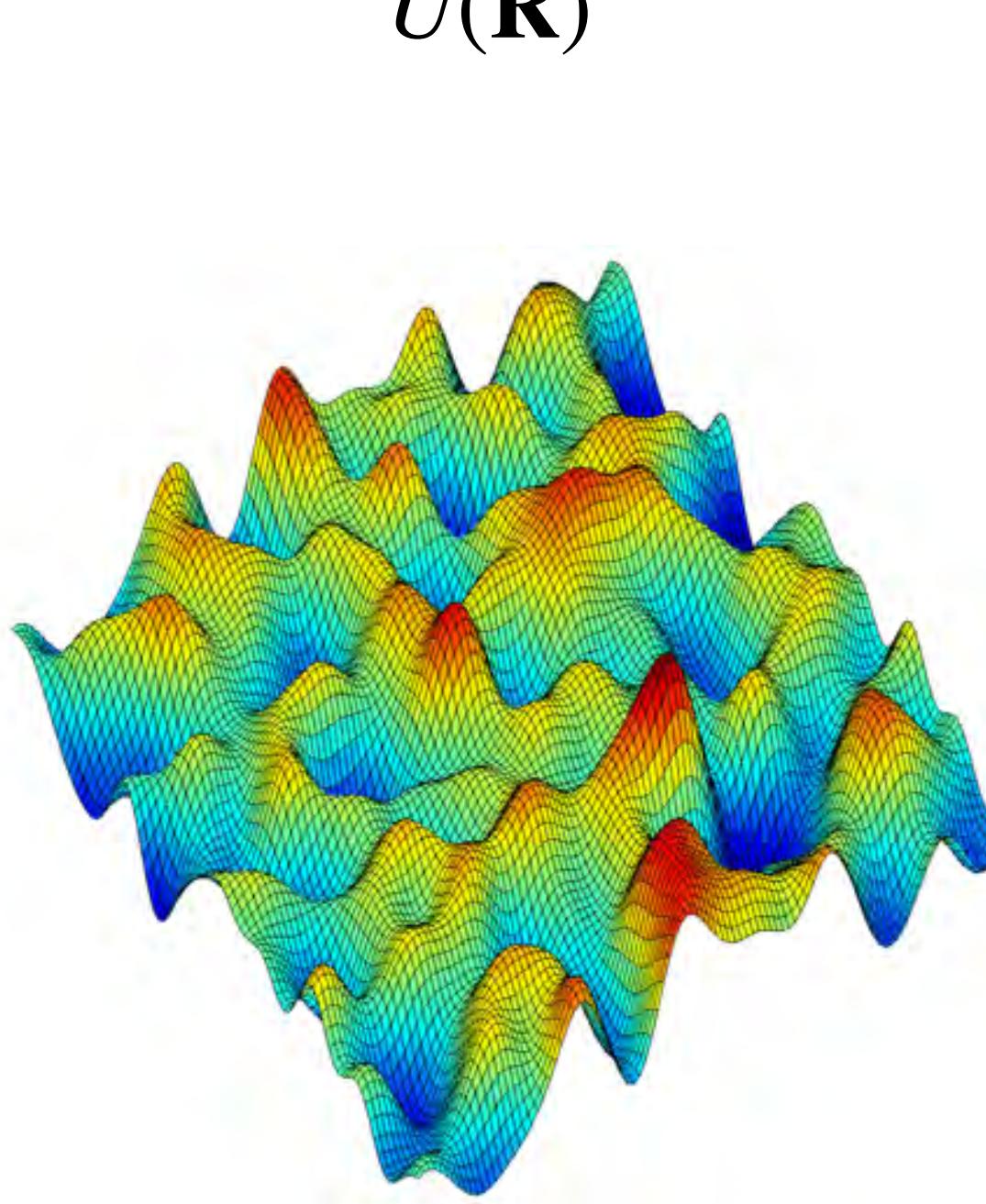


$$\mathbf{s} = [s_1, \dots, s_d]$$



Mapping the Problem to a Lower Dimension

Potential Energy Surface



High-dimensional / Rugged

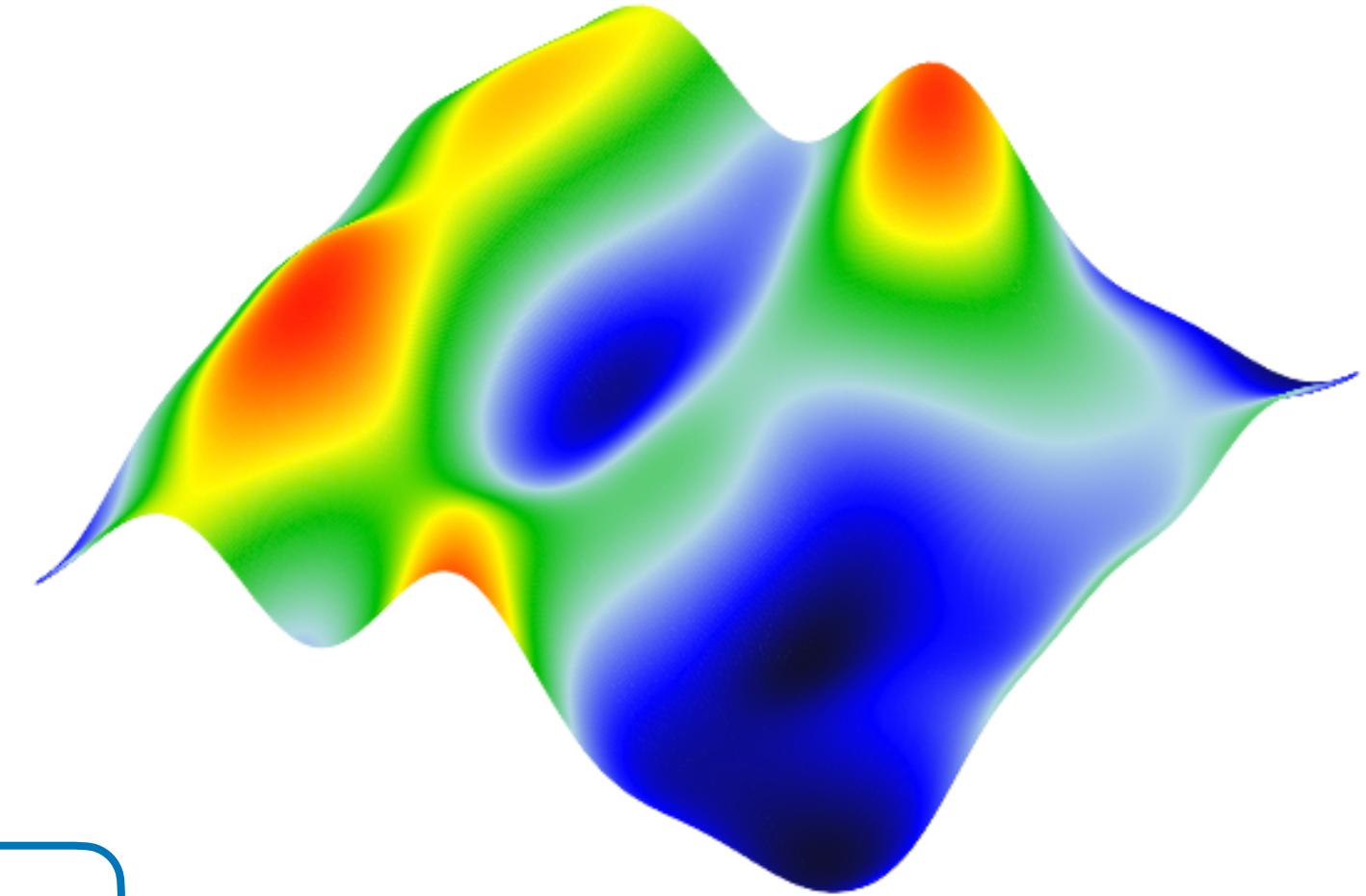
Free Energy Landscape/Surface
(i.e. logarithm of the marginal distribution)

$$F(\mathbf{s}) = -\frac{1}{\beta} \log P(\mathbf{s}) = -\frac{1}{\beta} \log \int d\mathbf{R} \delta [\mathbf{s} - \mathbf{s}(\mathbf{R})] e^{-\beta U(\mathbf{R})}$$

Collective Variables



$$\mathbf{s}(\mathbf{R}) = [s_1(\mathbf{R}), s_2(\mathbf{R}), \dots, s_d(\mathbf{R})]$$

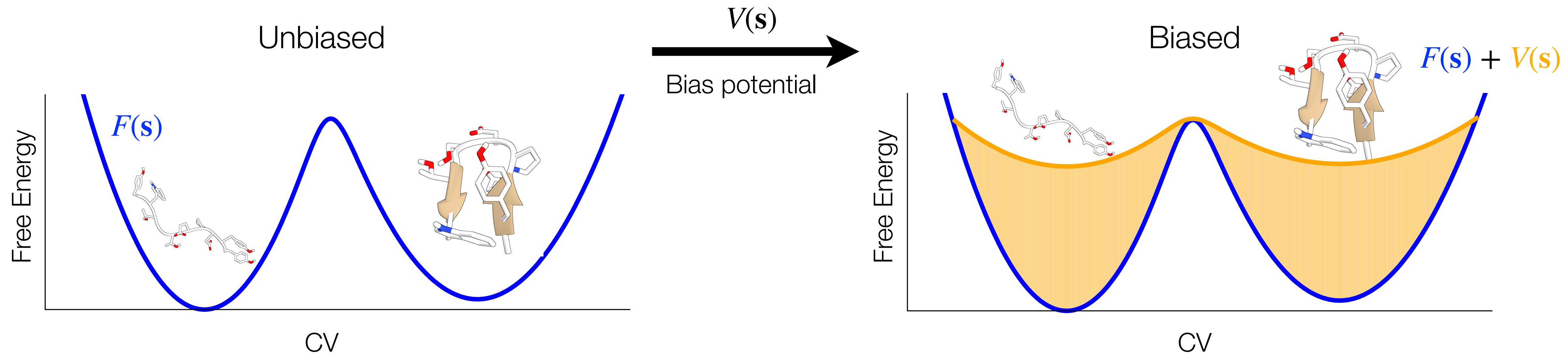


Use the collective variables to both
represent and **drive the sampling**
of the free energy landscape

Low-dimensional / Smooth

CV-based Based Enhanced Sampling Method

Introduce a bias potential $V(\mathbf{s})$ that acts in the space spanned by the **slow** CVs
(originally suggested in umbrella sampling[@])

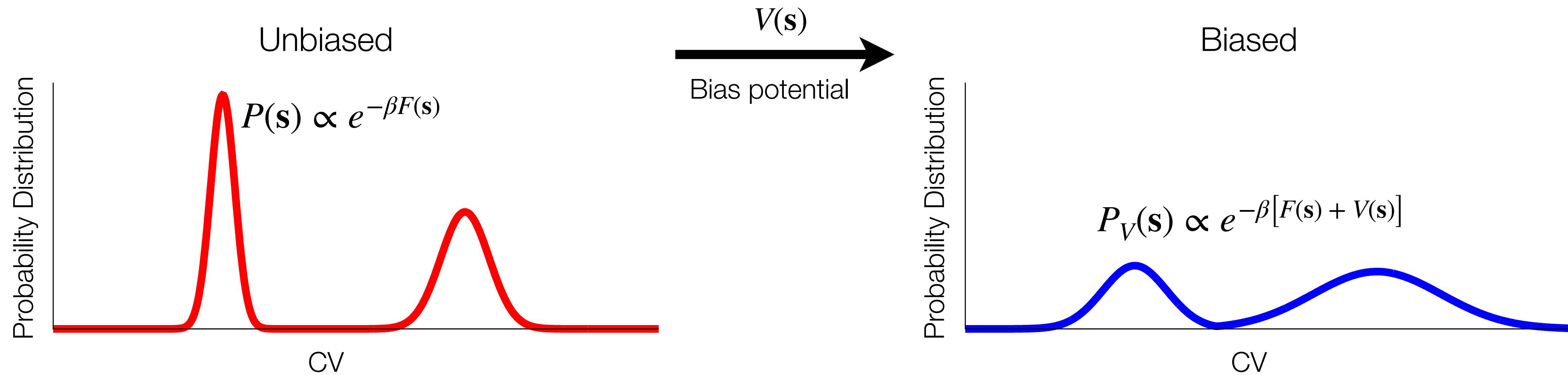


Reduce free energy barriers → easier to move between metastable states

CV-based Based Enhanced Sampling Method

Can also be viewed as enhancing CV fluctuations

Go to a biased distribution that is easier to sample



Various methods that adaptively build the bias potential on-the-fly during the simulation
(e.g. Metadynamics, Adaptive Biasing Force, ...)

Variationally Enhanced Sampling (VES)

Collective variable based enhanced sampling method

Based on a variational principle

$V(\mathbf{s})$ constructed by minimizing a convex functional

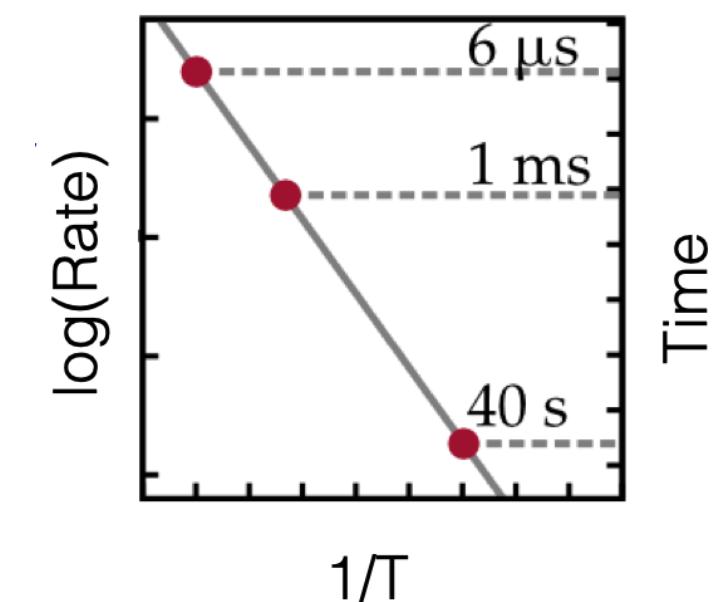
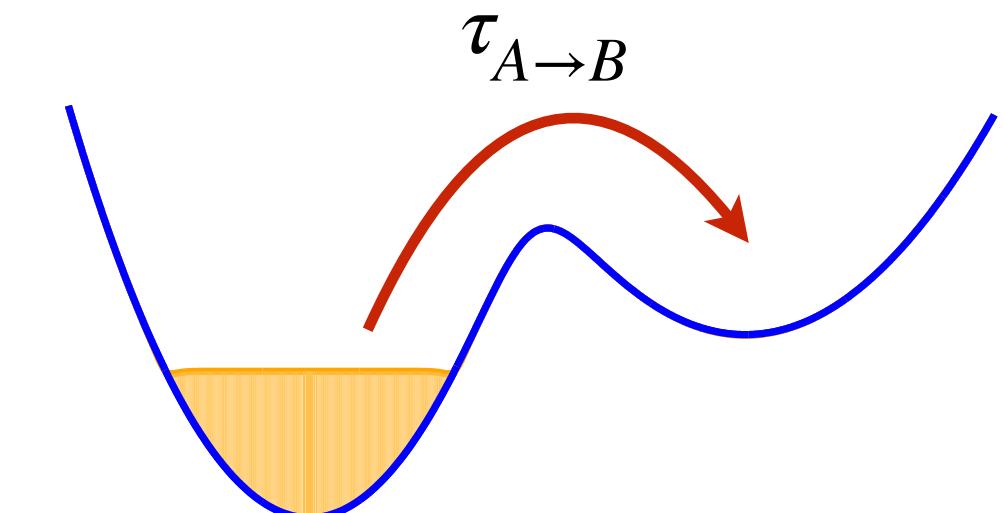
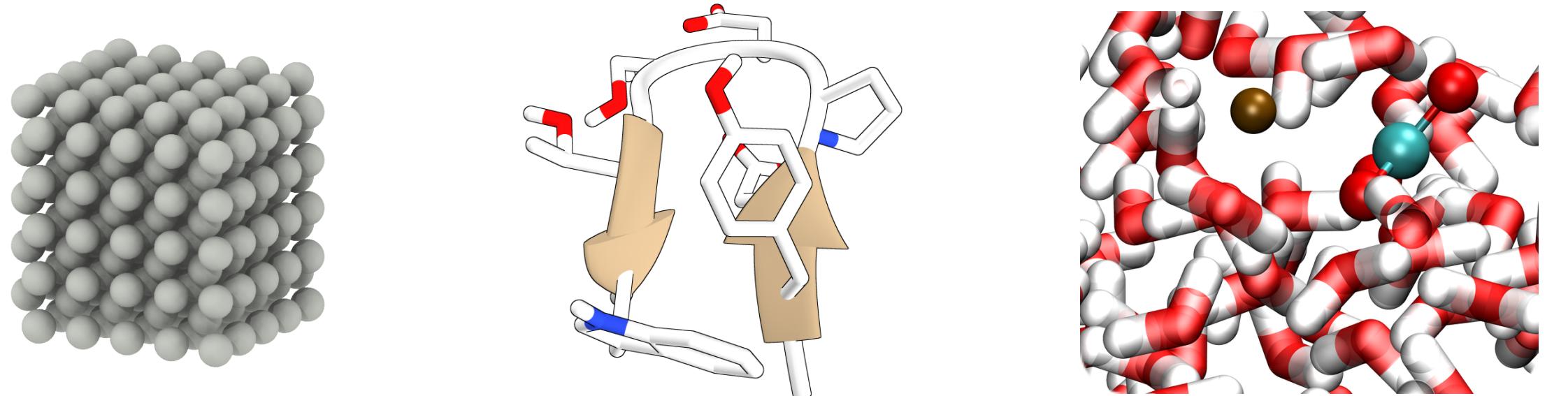
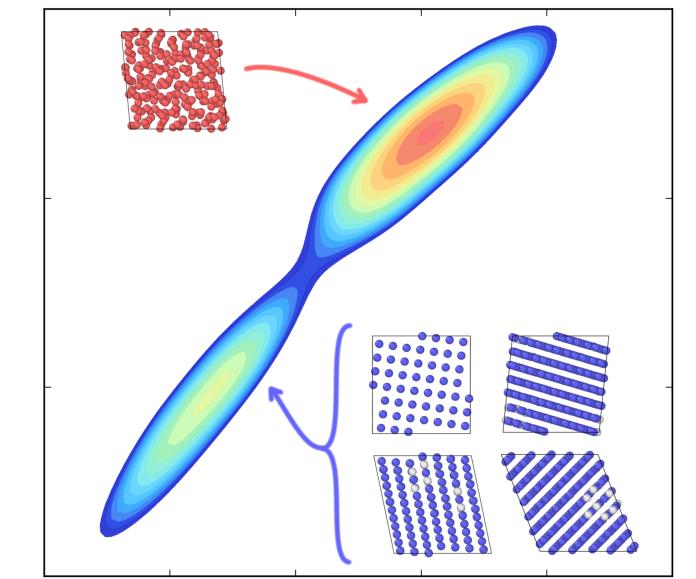
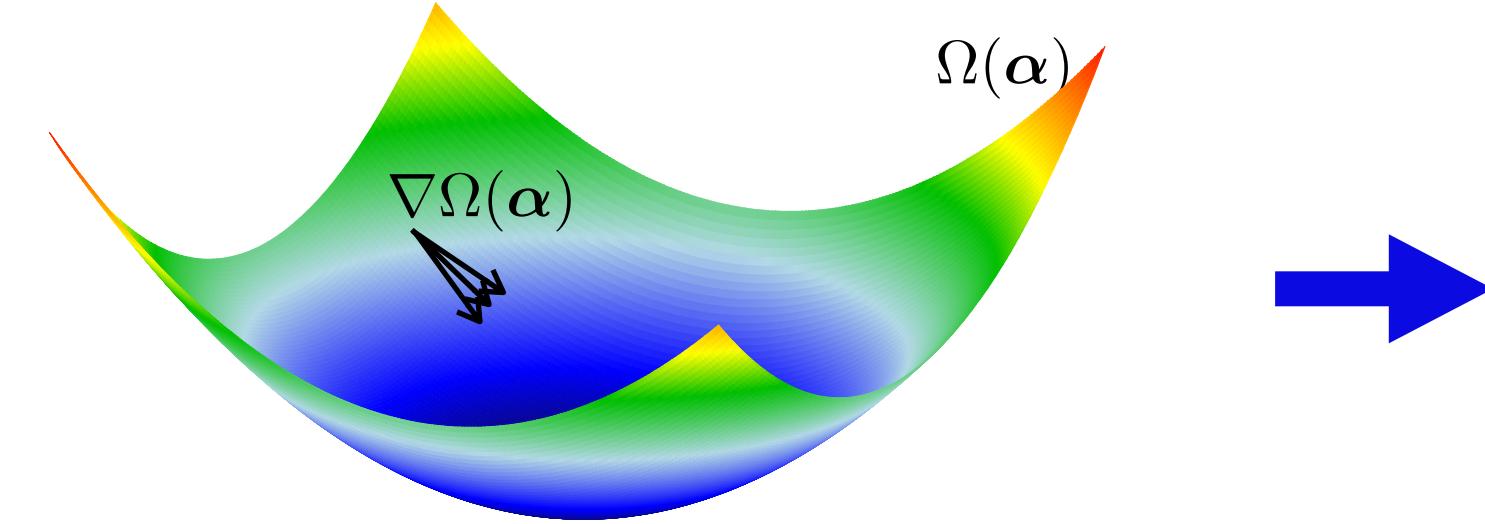
Offers many interesting possibilities in the form of $V(\mathbf{s})$

Can tailor the sampling of the CVs

Framework for developing new ideas
for free energy calculations

Completely general: applicable anywhere one can
define appropriate/good CVs

Can give access to kinetics of rare events



Variational Principle to Enhanced Sampling

Based on a functional $\Omega[V]$ of an external bias potential $V(\mathbf{s})$

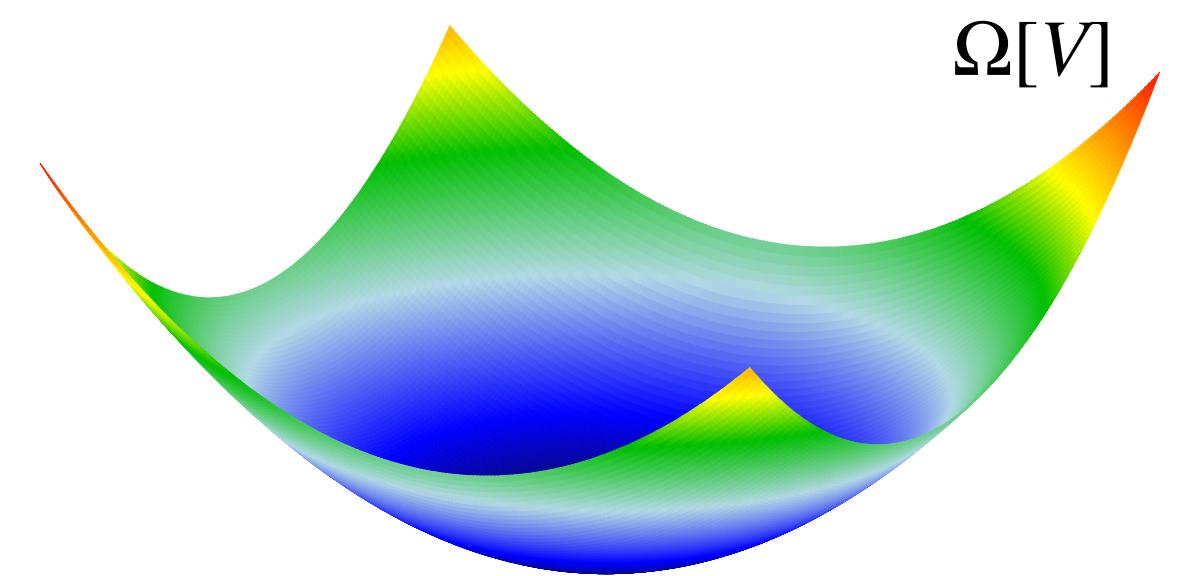
$$\Omega[V] = \frac{1}{\beta} \log \frac{\int d\mathbf{s} e^{-\beta[F(\mathbf{s}) + V(\mathbf{s})]}}{\int d\mathbf{s} e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s} p_{tg}(\mathbf{s})V(\mathbf{s})$$

↑
predefined probability distribution

$\Omega[V]$ is related to the Kullback-Leibler divergence (or cross/relative entropy)

Can show that $\Omega[V]$ is minimized by

$$\min_V \Omega[V] \quad \longrightarrow \quad V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta} \log p_{tg}(\mathbf{s}) + C$$



Global minimum $\Omega[V]$ as is convex

So, by minimizing $\Omega[V]$ we can obtain the FES $F(\mathbf{s})$ and enhance the sampling

The resulting CV distribution determined by the target distribution $p_{tg}(\mathbf{s})$, that we can choose freely

Variational Principle to Enhanced Sampling

The resulting CV distribution determined by the target distribution $p_{\text{tg}}(\mathbf{s})$

$$V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta} \log p_{\text{tg}}(\mathbf{s}) + C \quad \rightarrow$$

$$P_V(\mathbf{s}) = \frac{e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s} e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}} = p_{\text{tg}}(\mathbf{s})$$

Can choose the target distribution $p_{\text{tg}}(\mathbf{s})$ as we want

- can precisely tune the sampling of phase space
- brings a lot of flexibility to the method

Doesn't really need to be known a-priori → can be iteratively updated^a

Variational Principle to Enhanced Sampling

Can be shown that[#]

$$\beta\Omega[V] = D_{\text{KL}}(p_{\text{tg}} \parallel P_V) - D_{\text{KL}}(p_{\text{tg}} \parallel P) = H_x(p_{\text{tg}} \parallel P_V) - H_x(p_{\text{tg}} \parallel P)$$

where D_{KL} is the Kullback-Leibler divergence (relative entropy) and H_x is the cross entropy

$$D_{\text{KL}}(p_{\text{tg}} \parallel P_V) = \int d\mathbf{s} p_{\text{tg}}(\mathbf{s}) \log \frac{p_{\text{tg}}}{P_V(\mathbf{s})}$$
$$(= 0 \text{ if } p_{\text{tg}}(\mathbf{s}) = P_V(\mathbf{s}))$$
$$H_x(p_{\text{tg}} \parallel P_V) = - \int d\mathbf{s} p_{\text{tg}}(\mathbf{s}) \log P_V(\mathbf{s})$$

minimizing $\Omega[V]$ equivalent to minimizing the KL divergence (or cross entropy)

between the target distribution $p_{\text{tg}}(\mathbf{s})$ and the biased distribution $P_V(\mathbf{s}) \propto e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}$

Closely related to cross entropy method of Rubinstein^a and other similar ideas^{b,c,d,e}

^a Rubinstein, Methodol. Comput. Appl. Probab. 1999

^b Shell, J. Chem. Phys. 2008

^c Bilionis and Koutsourelakis, J. Comp. Phys. 2012

^d Zhang, Wang, Hartmann, Weber, and Schütte, SIAM J. Sci. Comput. 2014

^e Cesari, Rei̘ser and Bussi, Computation 2018

[#] Invernizzi, Valsson, Parrinello, PNAS 2017

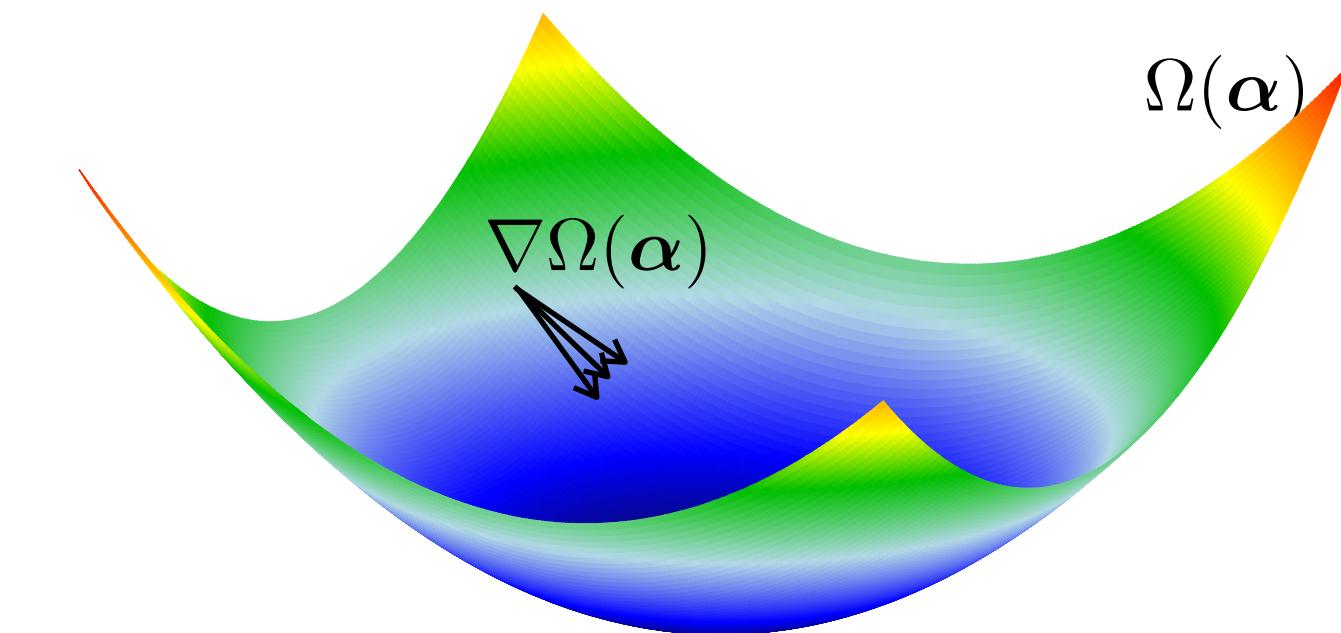
Variationally Enhanced Sampling in Practice

Assume some functional form for the bias $V(\mathbf{s}; \boldsymbol{\alpha})$ that depends on a set of variational parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)$ and minimize $\Omega(\boldsymbol{\alpha}) = \Omega[V(\boldsymbol{\alpha})]$ (convex optimization problem!)

the gradient $\nabla \Omega(\boldsymbol{\alpha})$ is defined as

$$\frac{\partial \Omega(\boldsymbol{\alpha})}{\partial \alpha_i} = - \left\langle \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i} \right\rangle_{V(\boldsymbol{\alpha})} + \left\langle \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i} \right\rangle_{p_{tg}}$$

averages sampled in a
biased MD simulation averages over $p_{tg}(\mathbf{s})$



Noisy optimization problem ➔ have to employ stochastic optimization methods

averaged stochastic gradient decent^a has proven an efficient choice so far

$$\boldsymbol{\alpha}^{(n+1)} = \boldsymbol{\alpha}^{(n)} - \mu \left[\nabla \Omega(\bar{\boldsymbol{\alpha}}^{(n)}) + \mathbf{H}(\bar{\boldsymbol{\alpha}}^{(n)}) [\boldsymbol{\alpha}^{(n)} - \bar{\boldsymbol{\alpha}}^{(n)}] \right] \quad \text{where} \quad \bar{\boldsymbol{\alpha}}^{(n)} = \frac{1}{n+1} \sum_{k=0}^n \boldsymbol{\alpha}^{(k)}$$

Bias Potential Representation

Linear Basis Set Expansion

e.g. plane waves, Legendre polynomials, splines, wavelets,

$$V(\mathbf{s}; \boldsymbol{\alpha}) = \sum_{\mathbf{k}} \alpha_{\mathbf{k}} \cdot f_{\mathbf{k}}(\mathbf{s})$$

Most general \rightarrow can represent any FES

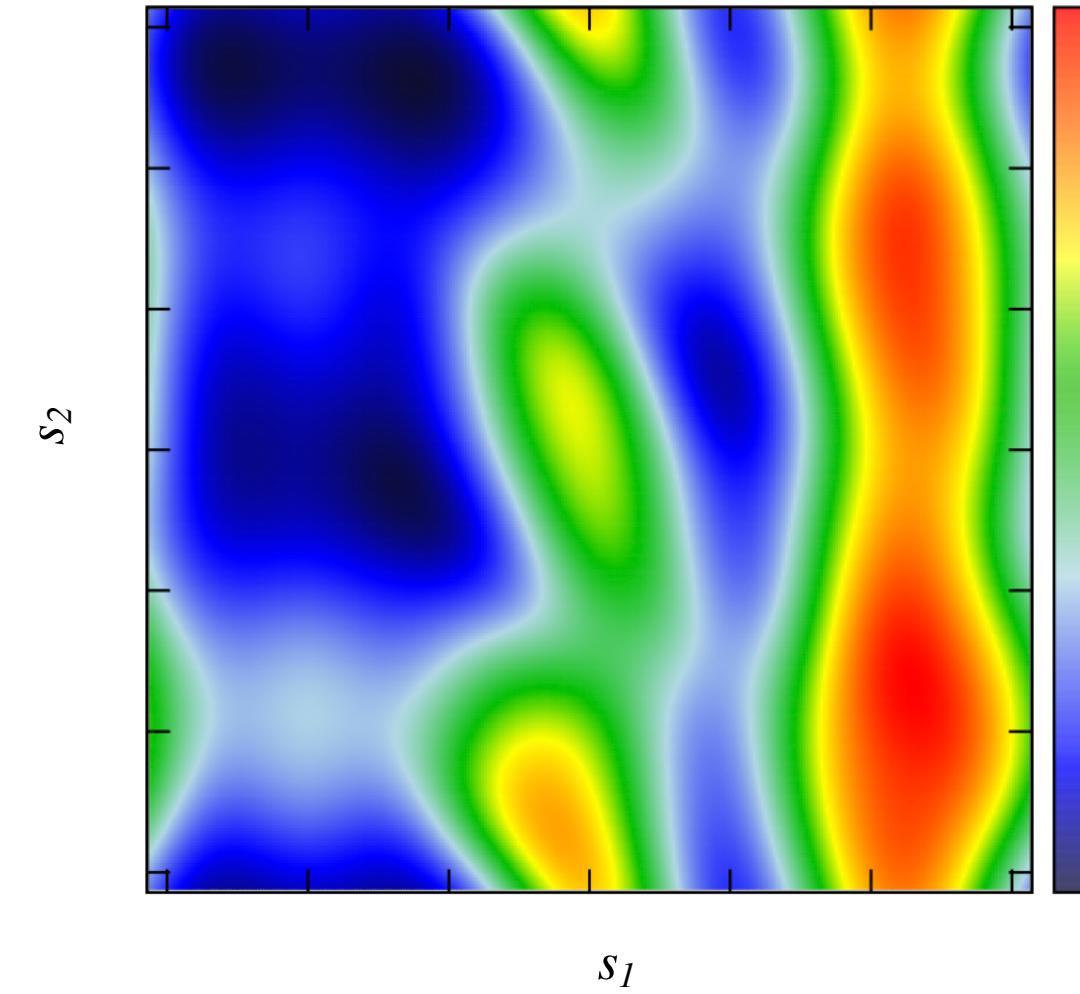
The gradient then simplifies to averages of the basis functions

$$\frac{\partial \Omega(\boldsymbol{\alpha})}{\partial \alpha_i} = - \langle f_i(\mathbf{s}) \rangle_{V(\boldsymbol{\alpha})} + \langle f_i(\mathbf{s}) \rangle_{P_{tg}}$$

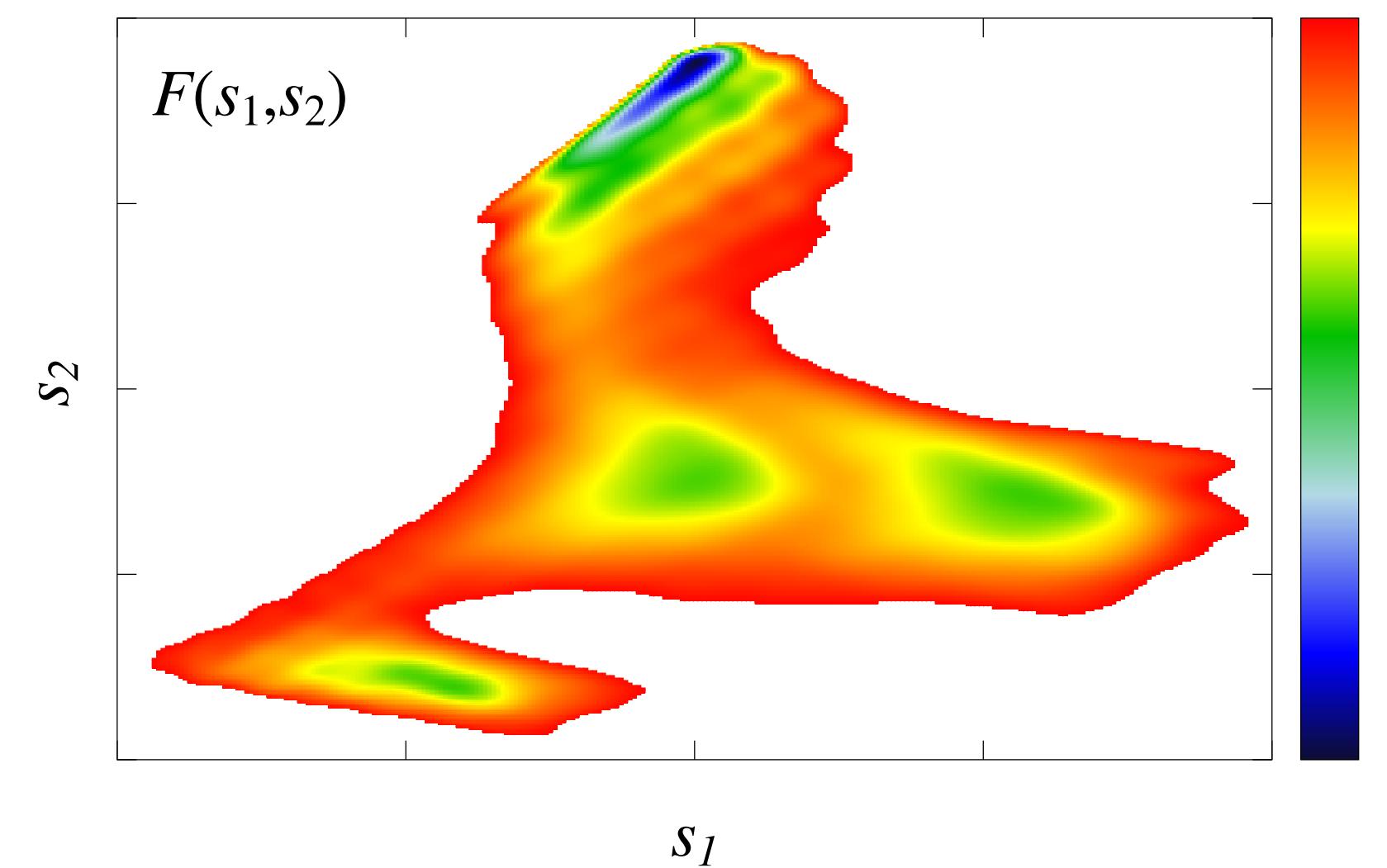
For two CVs or more: tensor product of one-dimensional basis functions

$$V(s_1, s_2; \boldsymbol{\alpha}) = \sum_{k_1, k_2} \alpha_{k_1, k_2} \cdot T_{k_1}(\tilde{s}_1) \cdot T_{k_2}(\tilde{s}_2)$$

Periodic CVs



Non-Periodic CVs



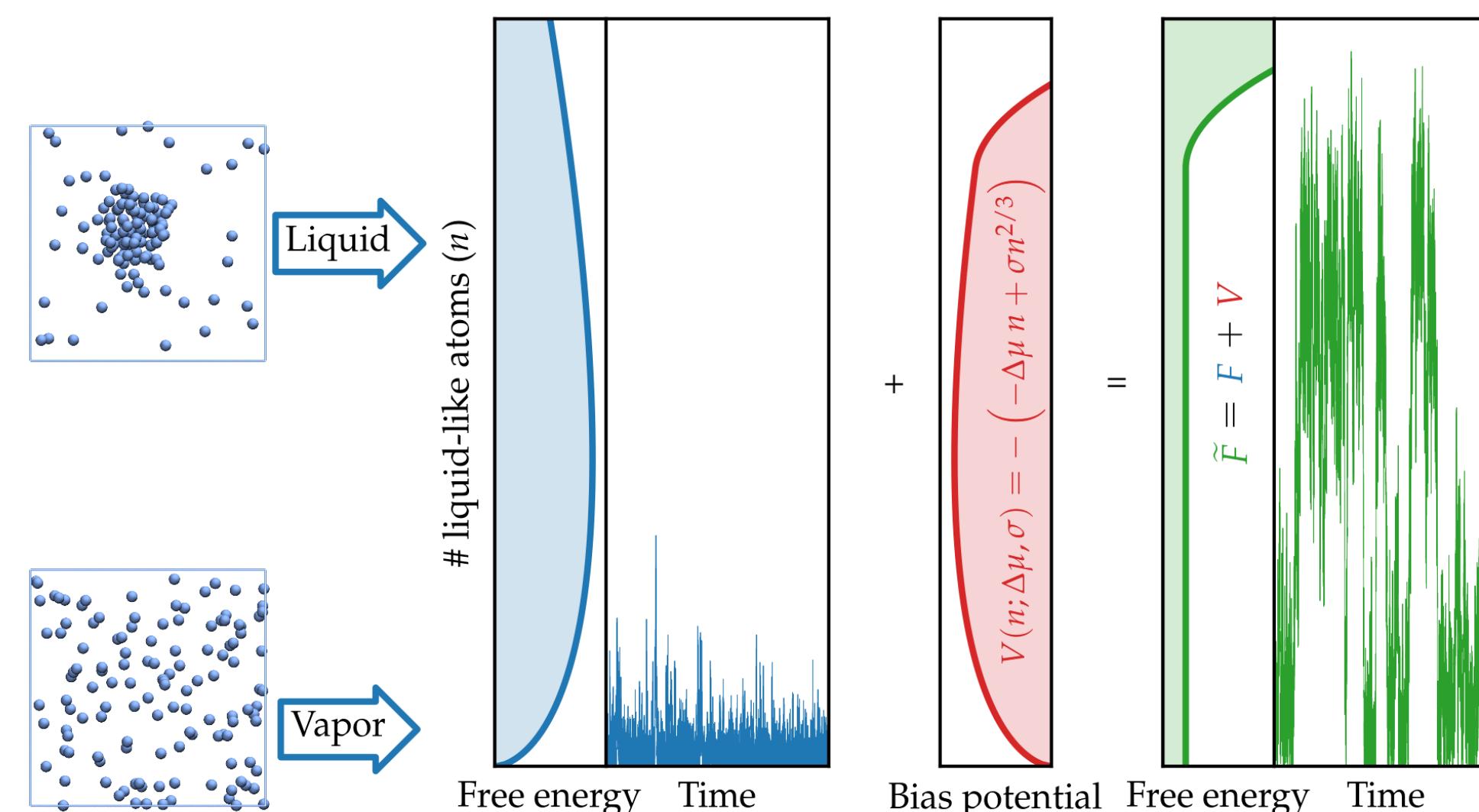
Bias Potential Representation

Bespoke Bias Potentials

e.g., a known free-energy model

Nucleation

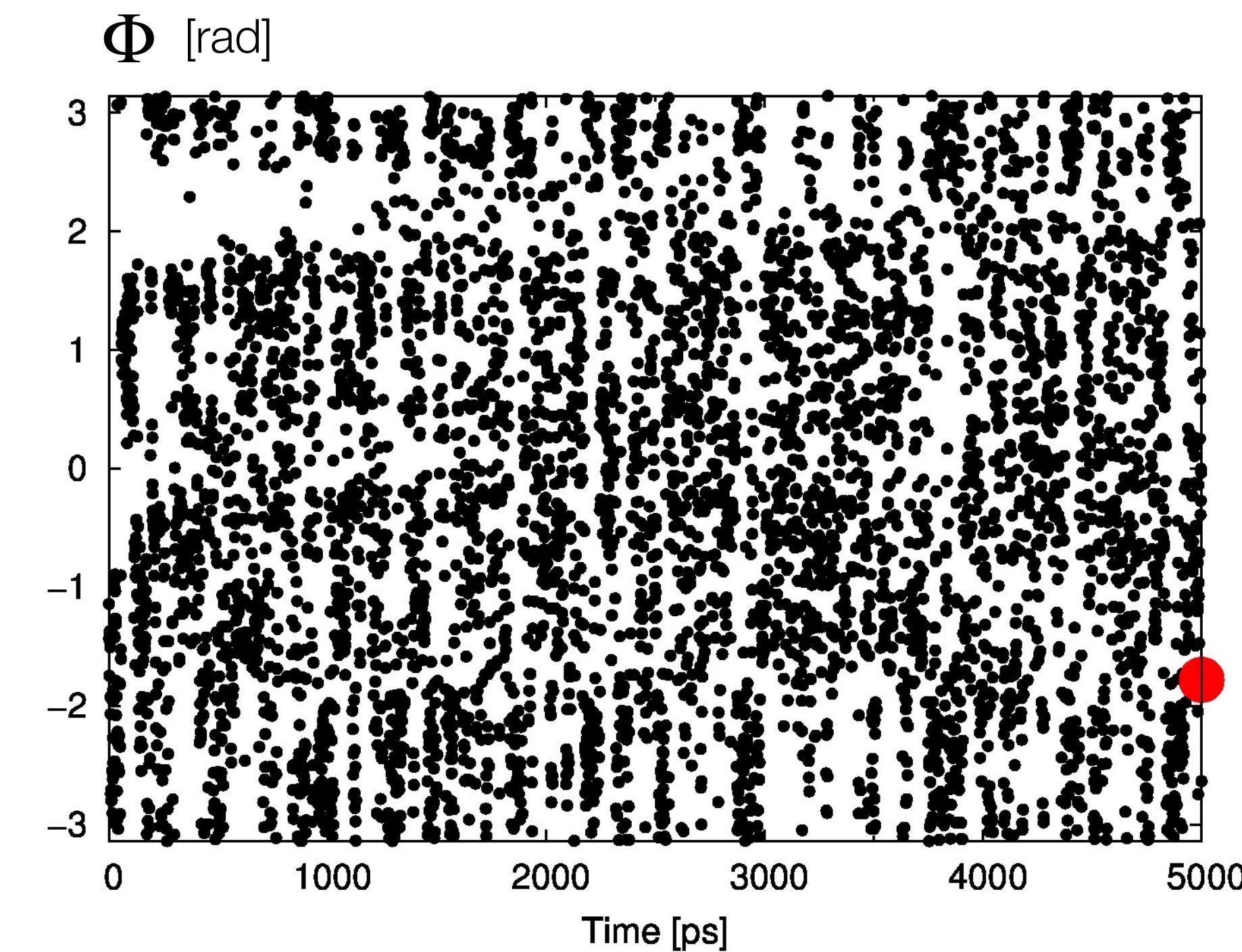
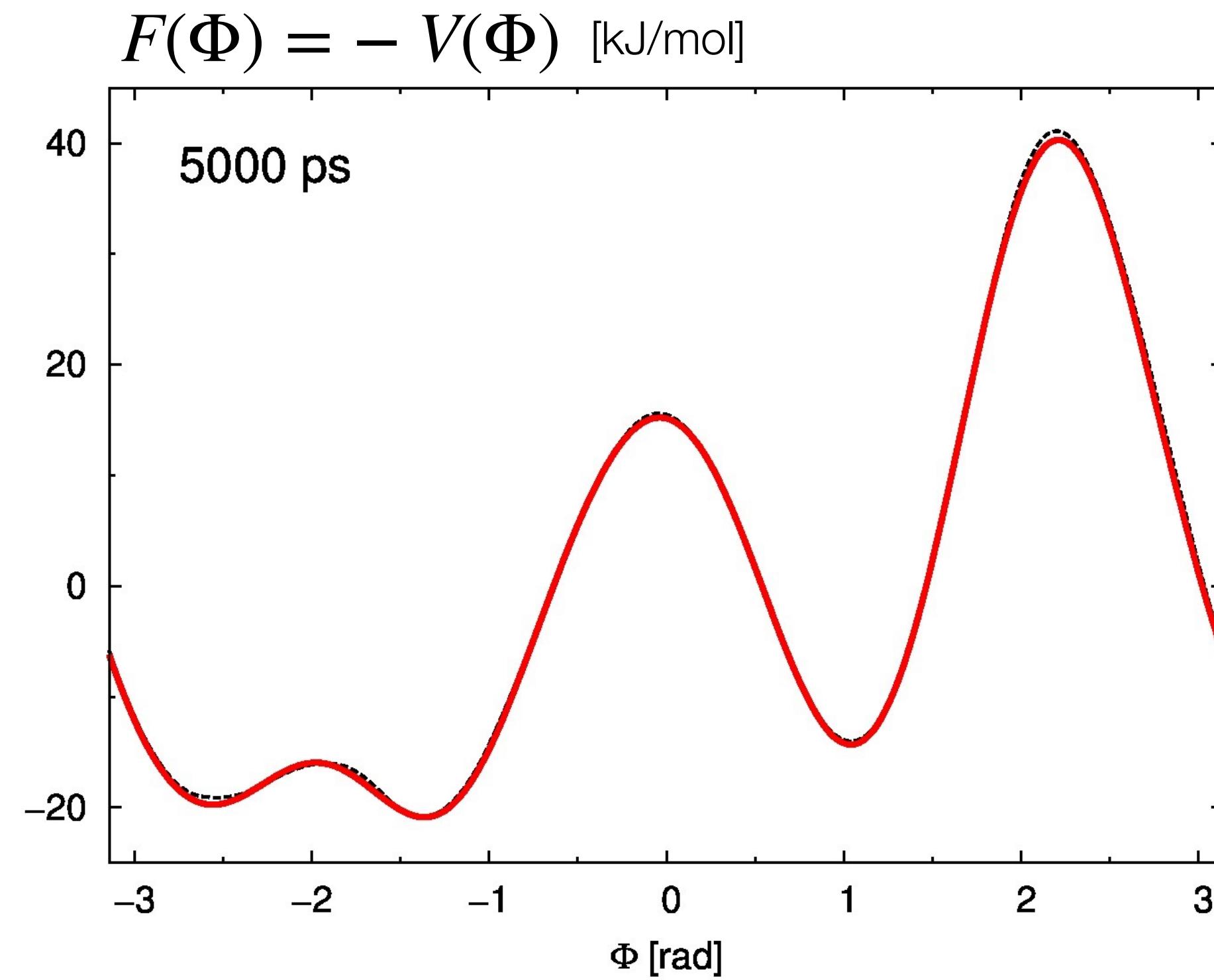
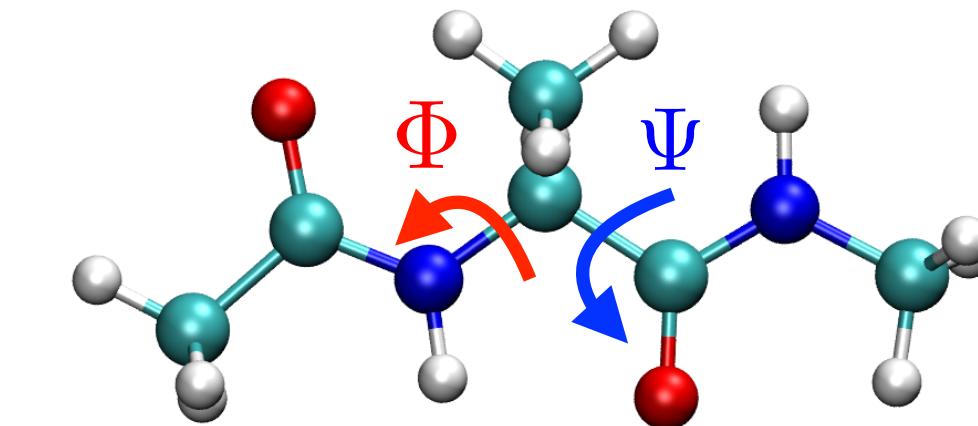
$$F(n) = -\Delta\mu n + \sigma n^{2/3}$$



Piaggi, **Valsson**, and Parrinello, Farad. Discuss. 2016

Example of Time Evolution

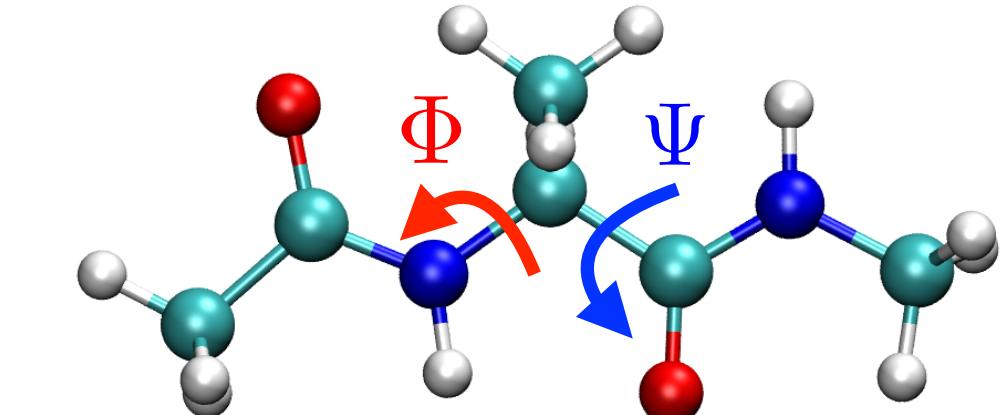
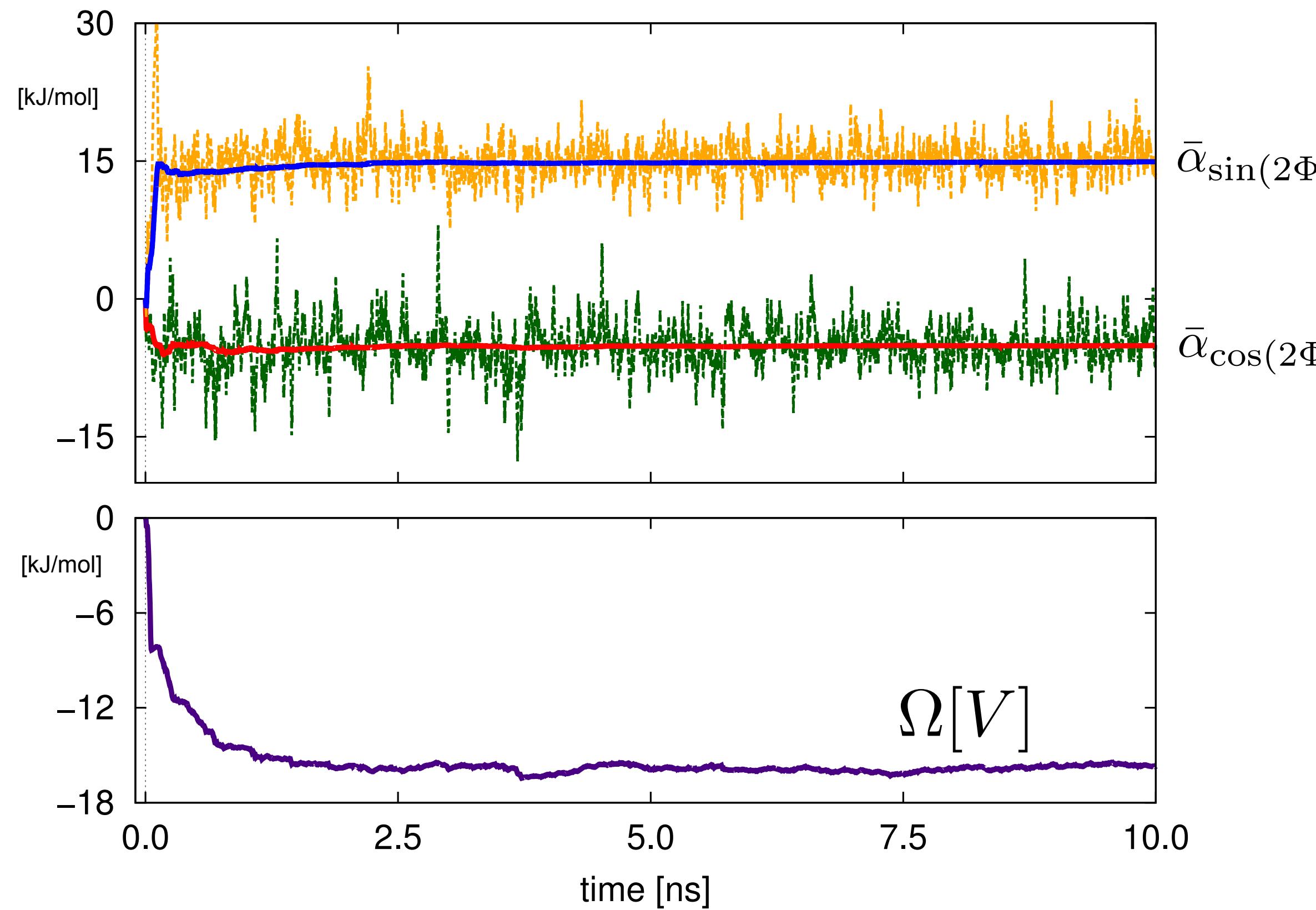
Alanine dipeptide, biasing only the Φ dihedral angle



Example of Time Evolution

Alanine dipeptide, biasing only the Φ dihedral angle

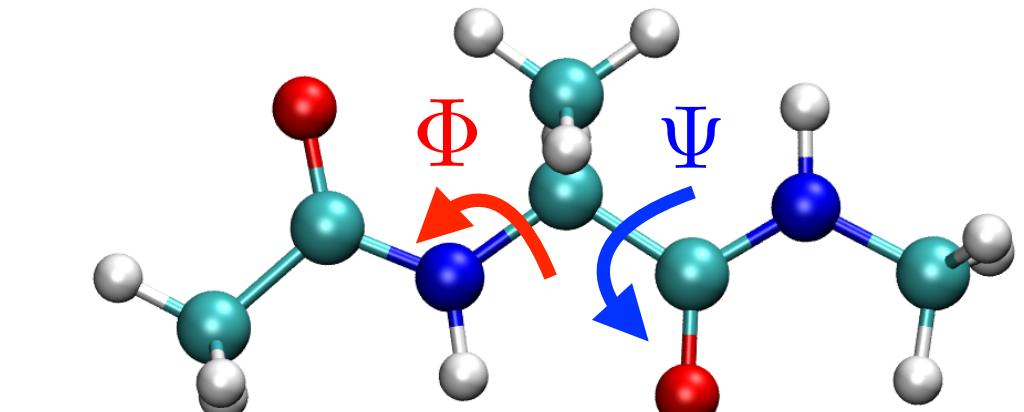
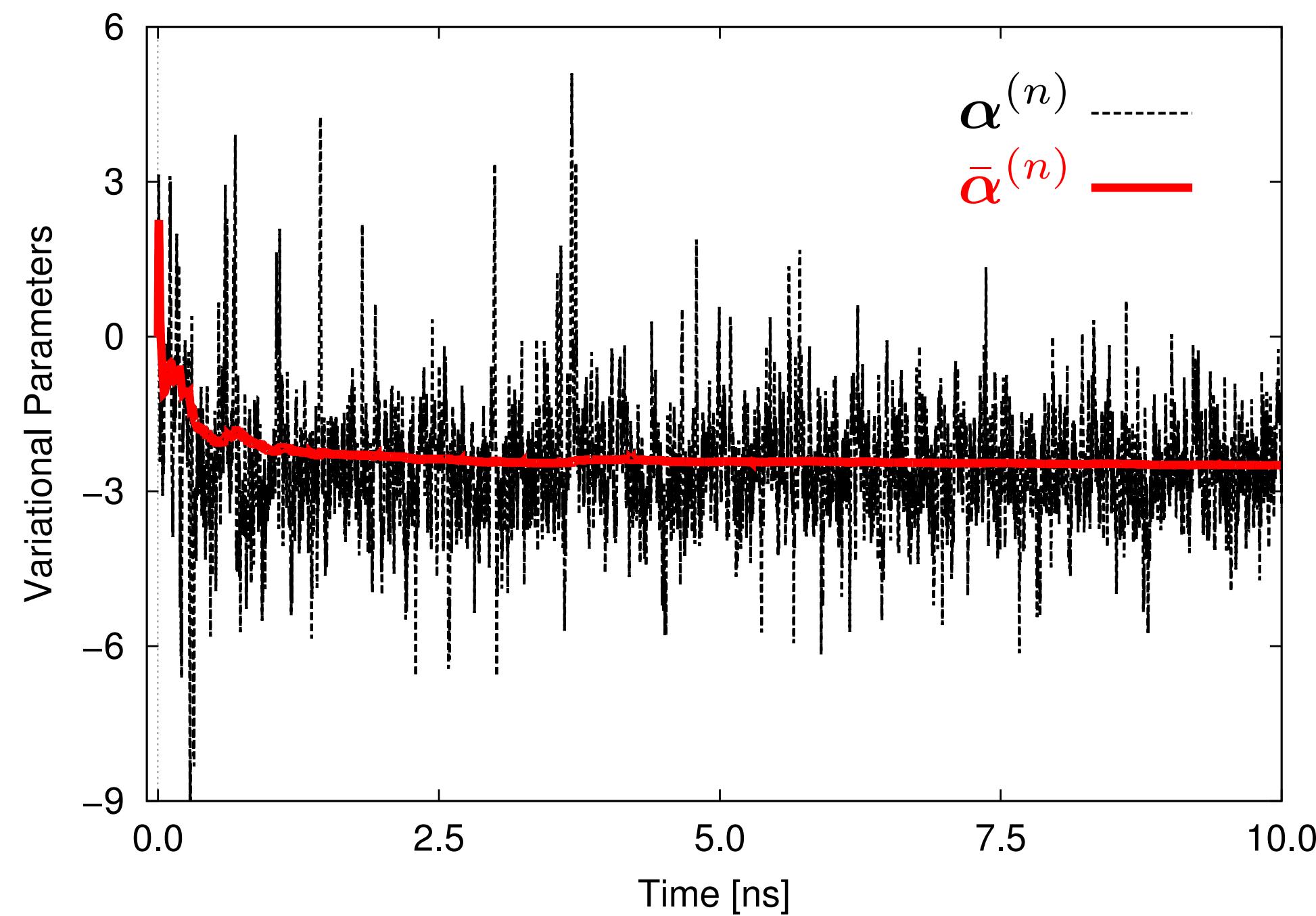
Clear convergence behavior observed for variational parameters and $\Omega[V]$



Example of Time Evolution

Alanine dipeptide, biasing only the Φ dihedral angle

instantaneous parameters fluctuate a lot, but averages converge smoothly



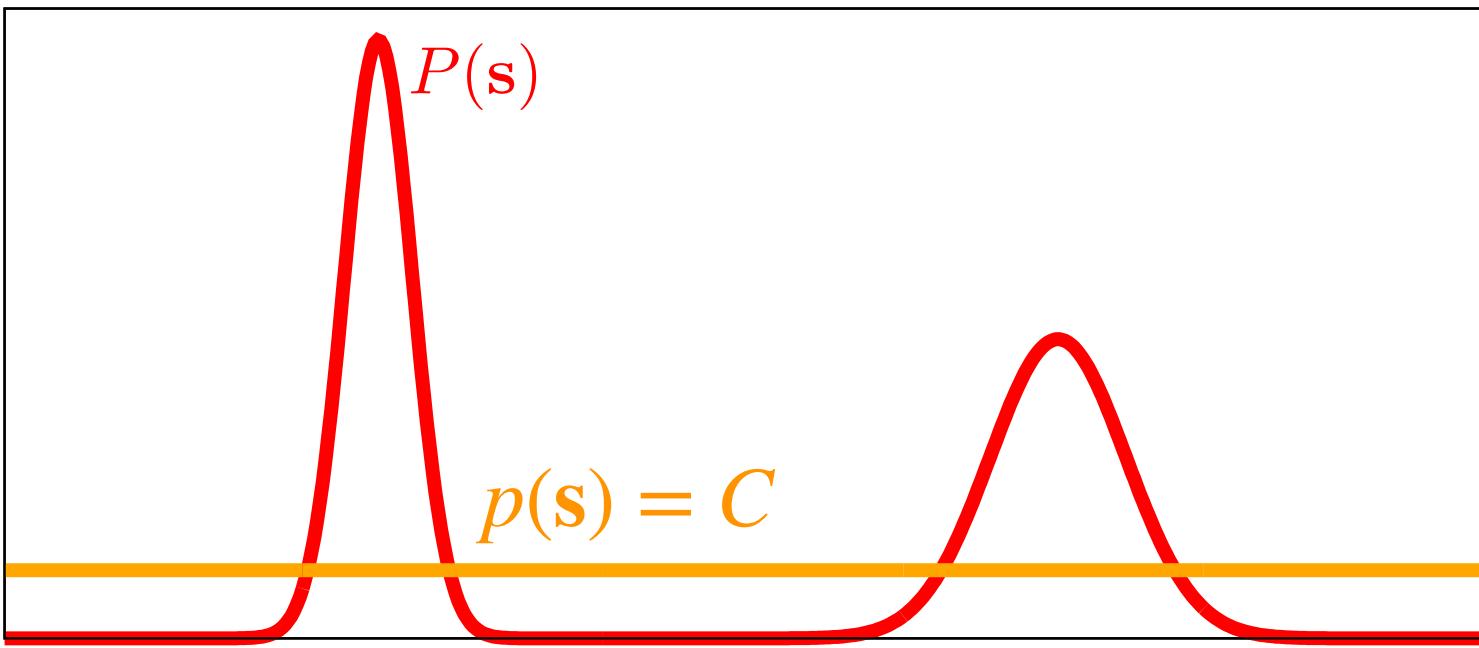
short iterations so the gradient is never really zero, even when converged
→ but averages to zero

Choice of the Target Distribution

The target distribution should generally be easier to sample than the equilibrium distribution

Uniform target distribution

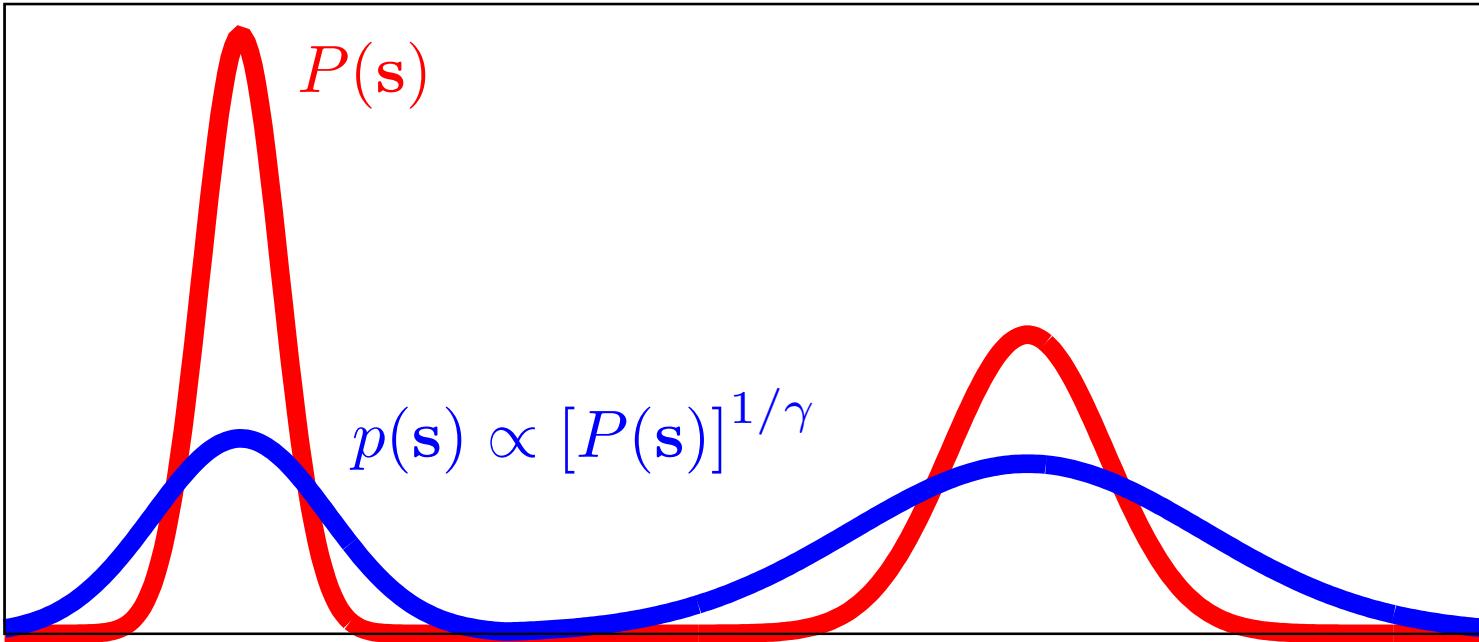
- the simplest choice
- but generally non-optimal



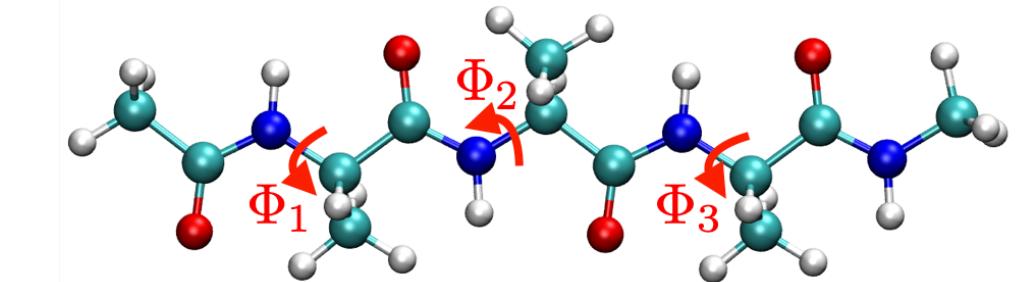
Well-tempered target distribution

- enhancing CV fluctuations in a controlled way

$$p_{\text{tg}}(s) \propto [P(s)]^{1/\gamma}$$

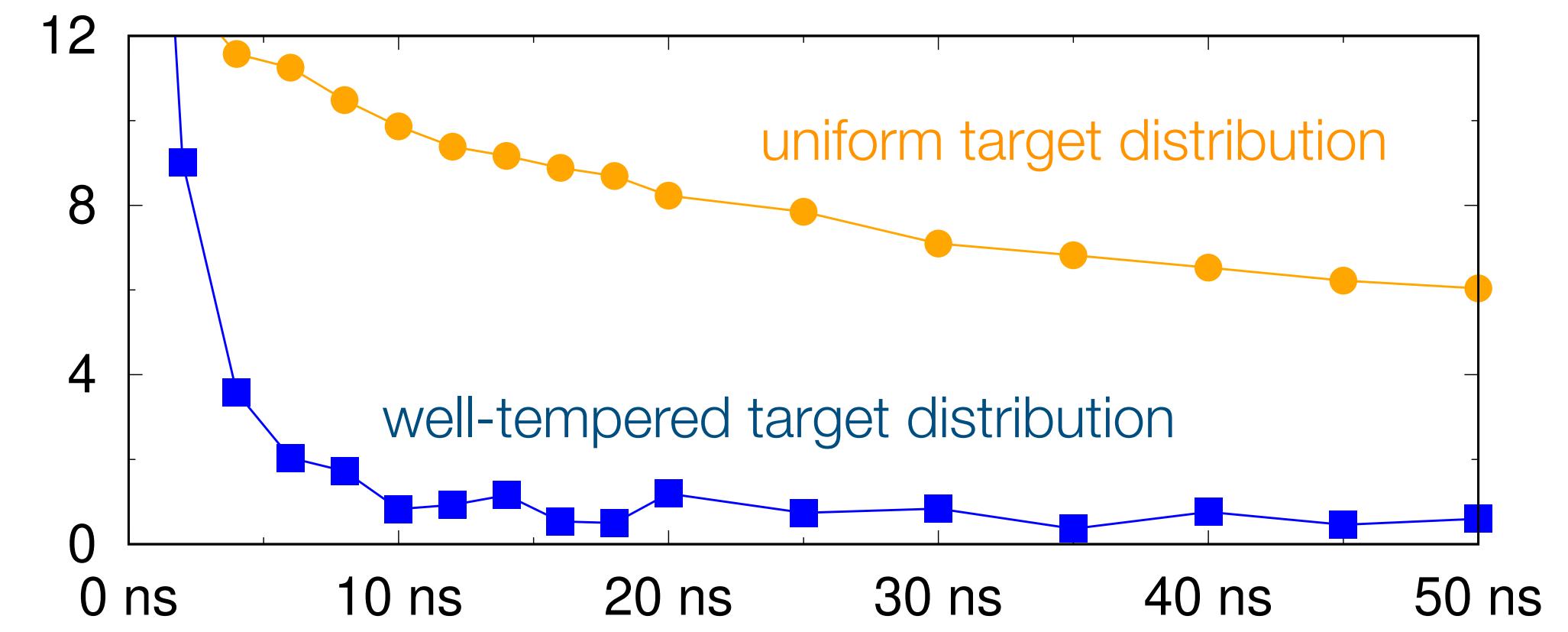


Alanine tetrapeptide

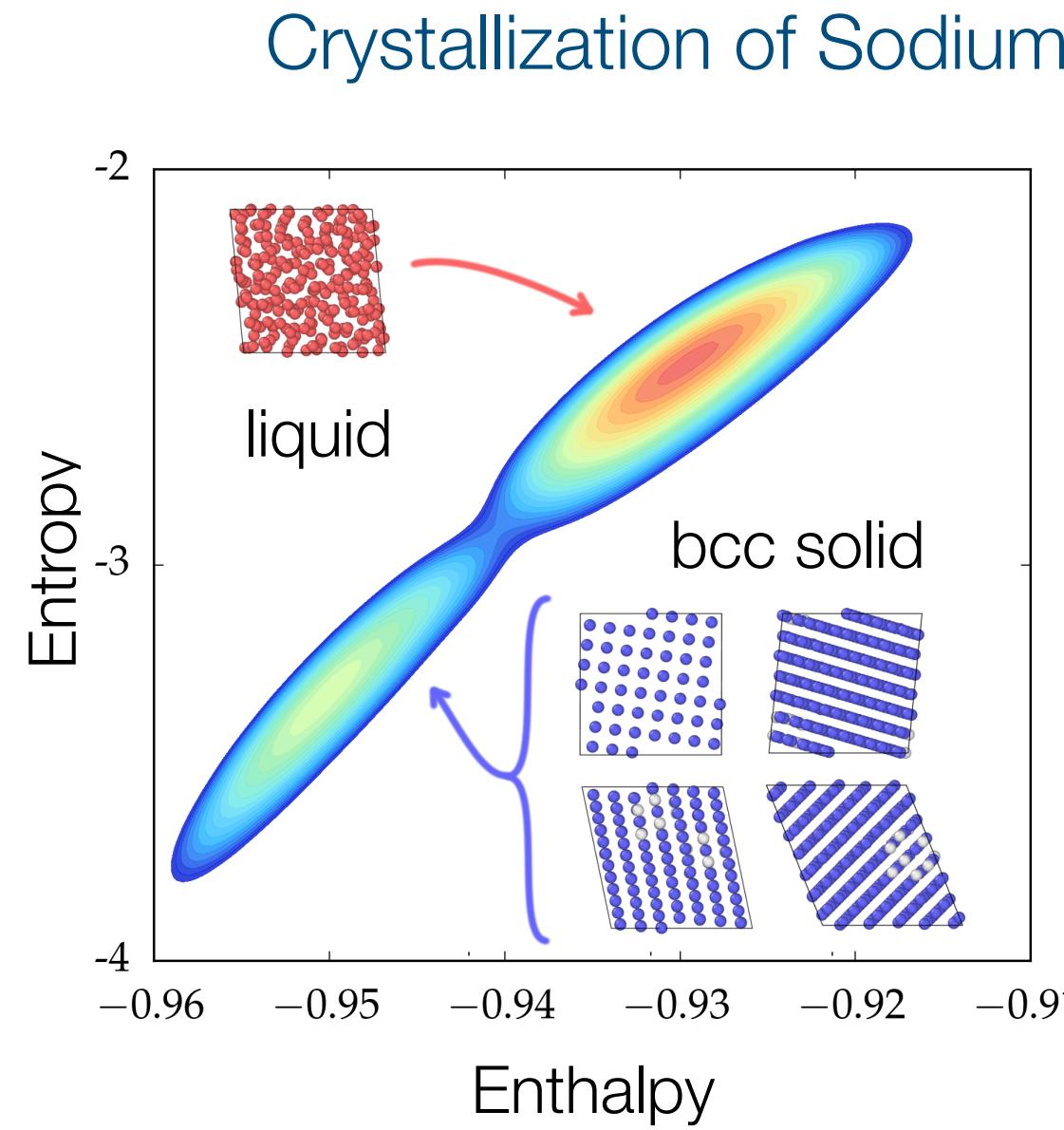


3 CVs \rightarrow ~2000-10000 basis functions

FES Convergence

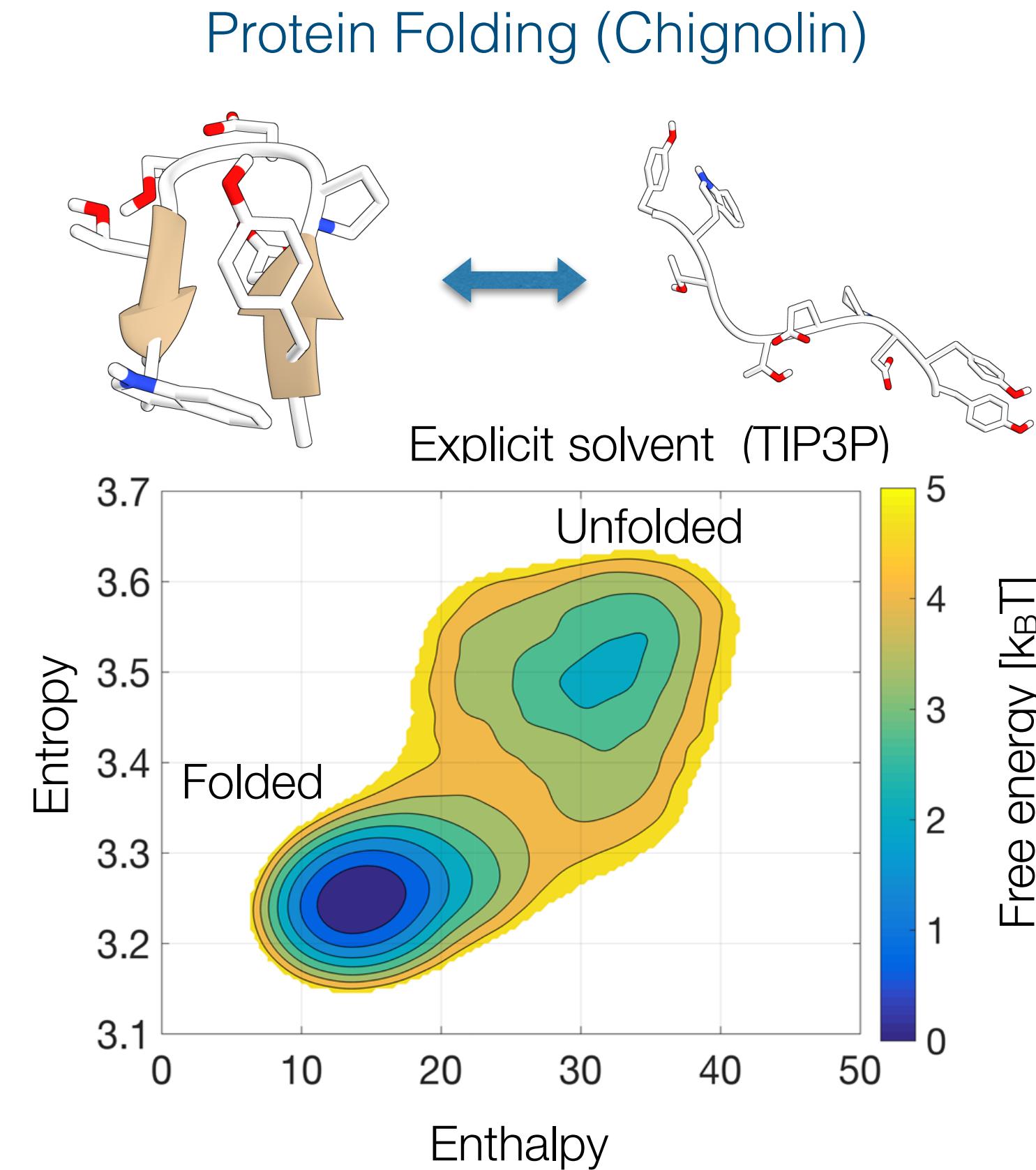


Examples of Applications of VES



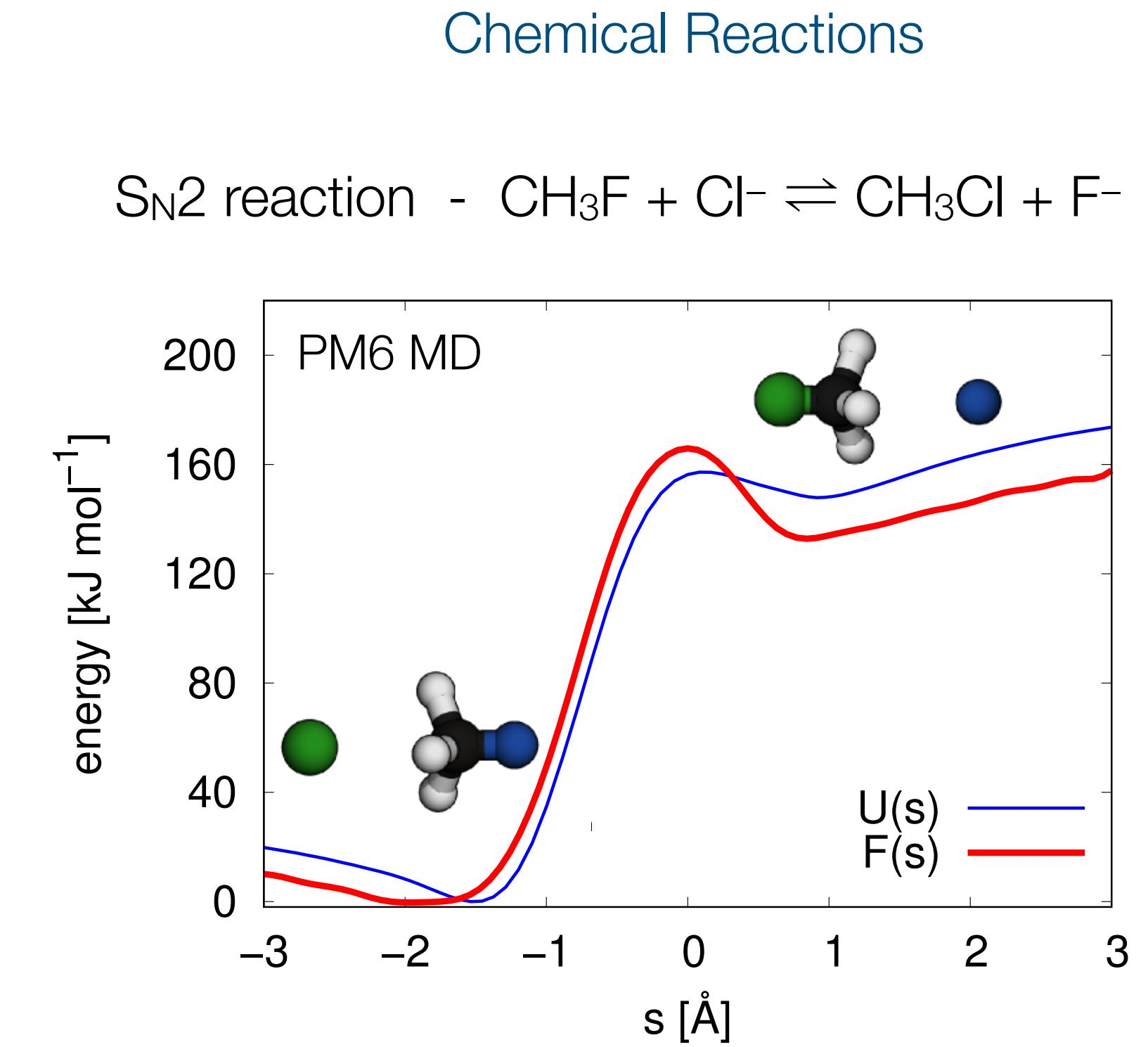
Development of an entropy-based collective variable for crystallization

Piaggi, **Valsson**, Parrinello, Phys. Rev. Lett. 2017



Development of an entropy/NMR-based collective variable for protein dynamics

Palazzesi, **Valsson**, Parrinello, J. Phys Chem. Lett. 2017



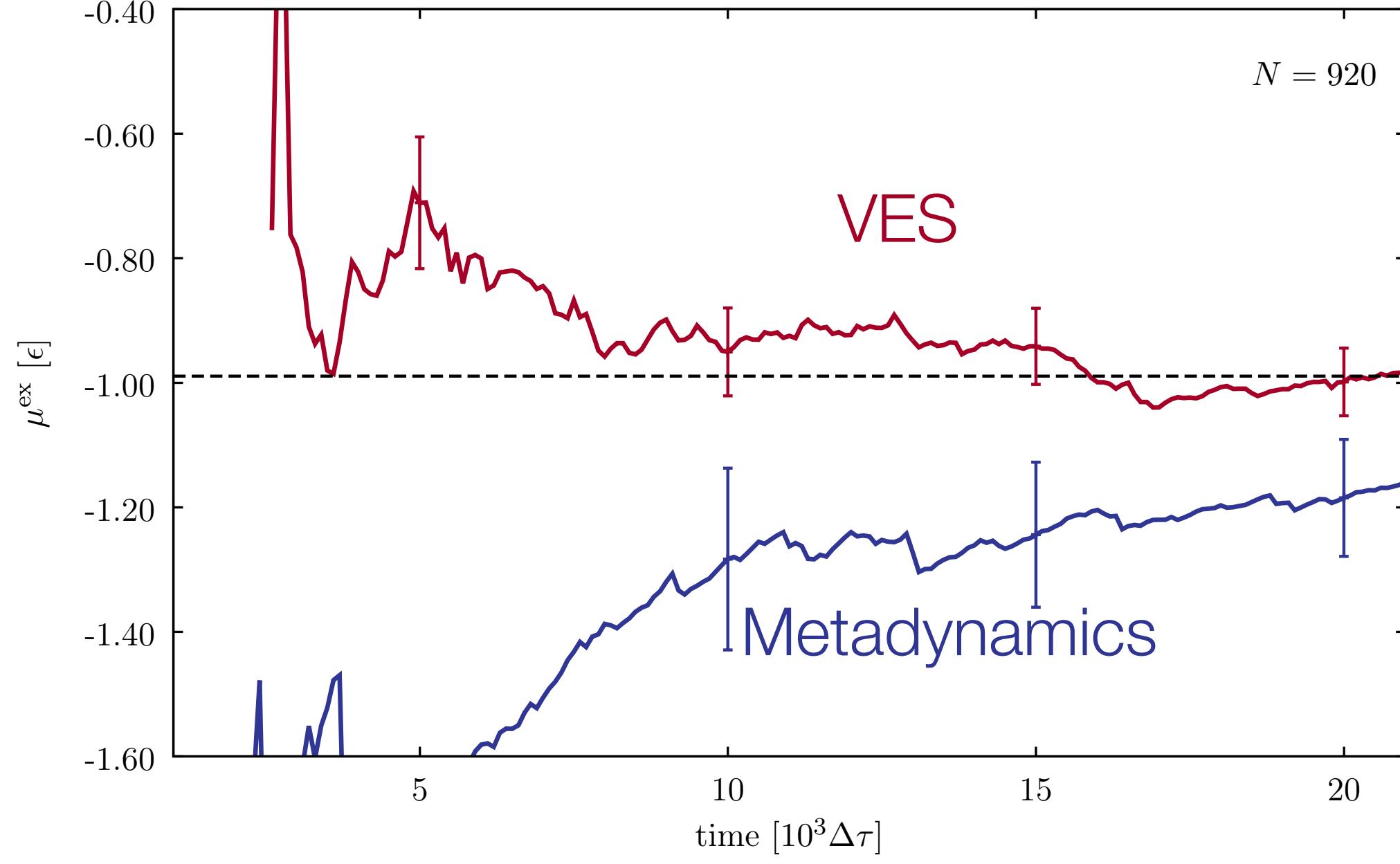
Piccini, McCarty, **Valsson**, Parrinello, J. Phys. Chem. Lett., 2017

Examples of Applications of VES

Chemical Potential Calculations in Non-Homogeneous Liquids

$$s(\mathbf{R}) = -\beta^{-1} \log \left[\frac{1}{M} \sum_{i=1}^M \frac{N+1}{V\rho_{N+1}(\mathbf{R}_i^*)} \exp(-\beta\Delta U(\mathbf{R}_i^*; \mathbf{R})) \right]$$

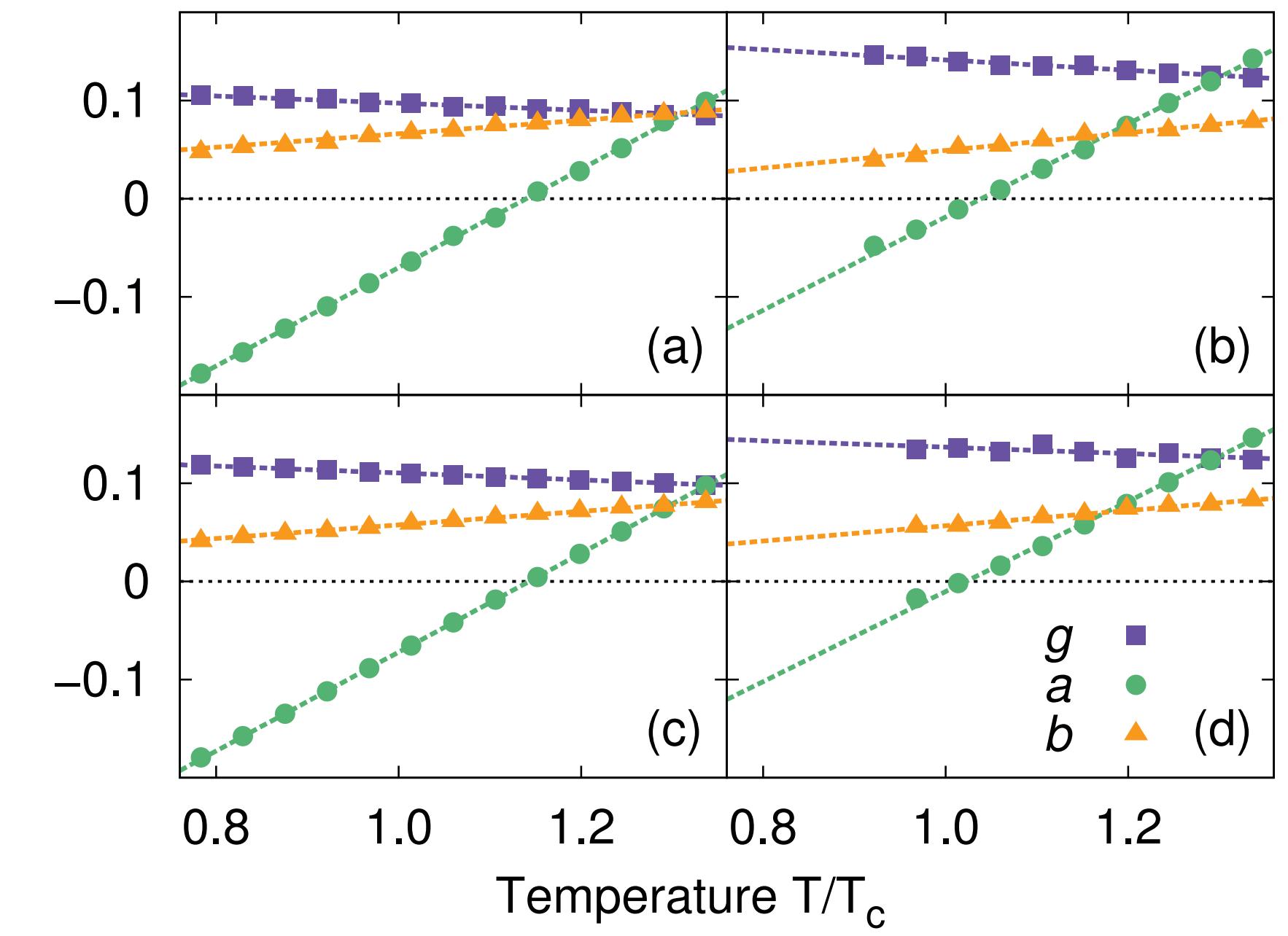
$$\mu^{\text{ex}} = -\beta^{-1} \log \int e^{-\beta s} P(s) ds$$



Calculations of Parameters for Phenomenological Coarse-Grained Models

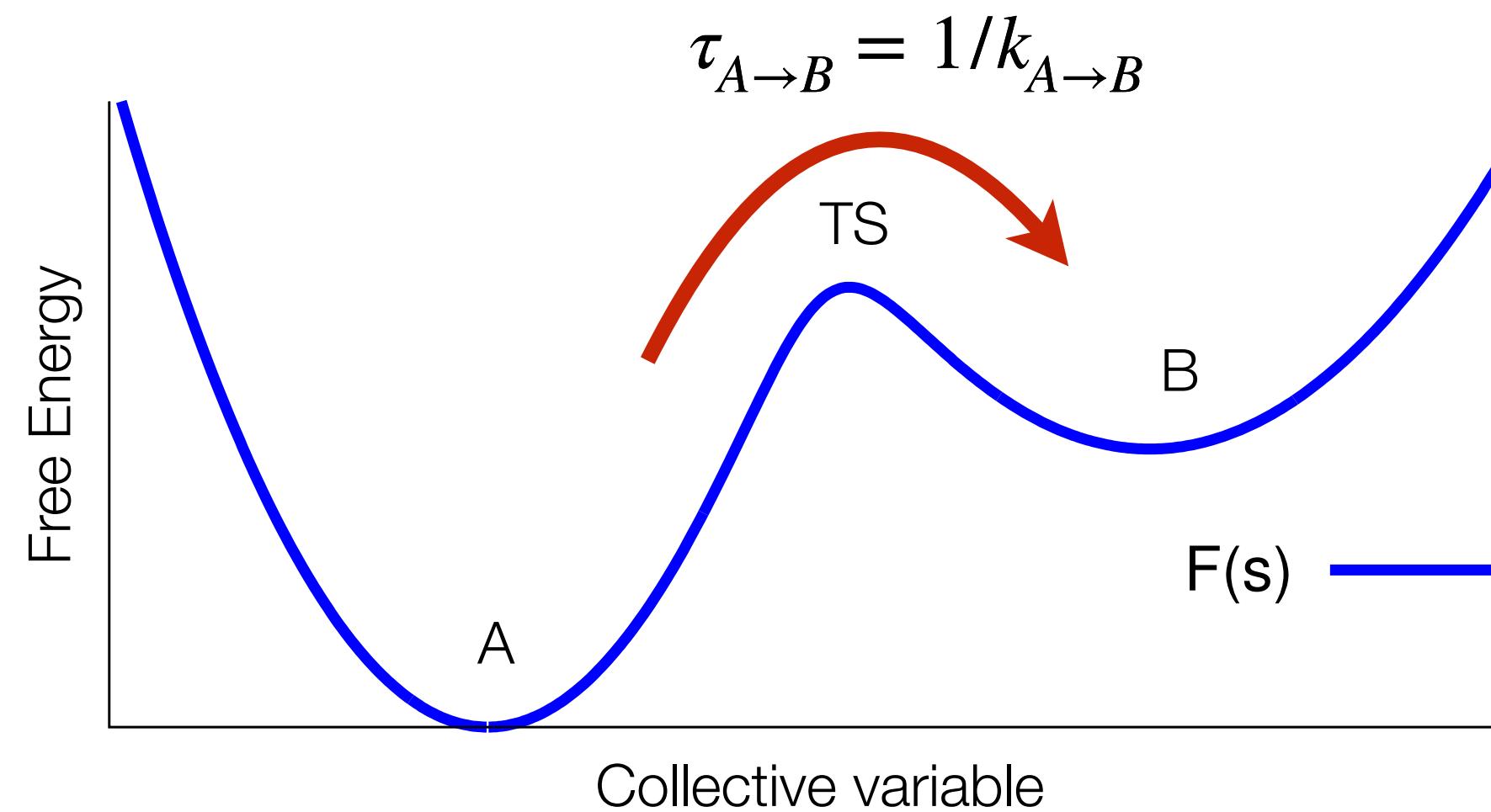
$$F[\psi] = g \int |\nabla \psi(\vec{r})|^2 d^3r + a \int \psi^2(\vec{r}) d^3r + b \int \psi^4(\vec{r}) d^3r$$

Ginzburg-Landau model for 2nd order phase transitions



Invernizzi, Valsson, Parrinello, PNAS 2017

Kinetics of Rare Events from Biased Simulations



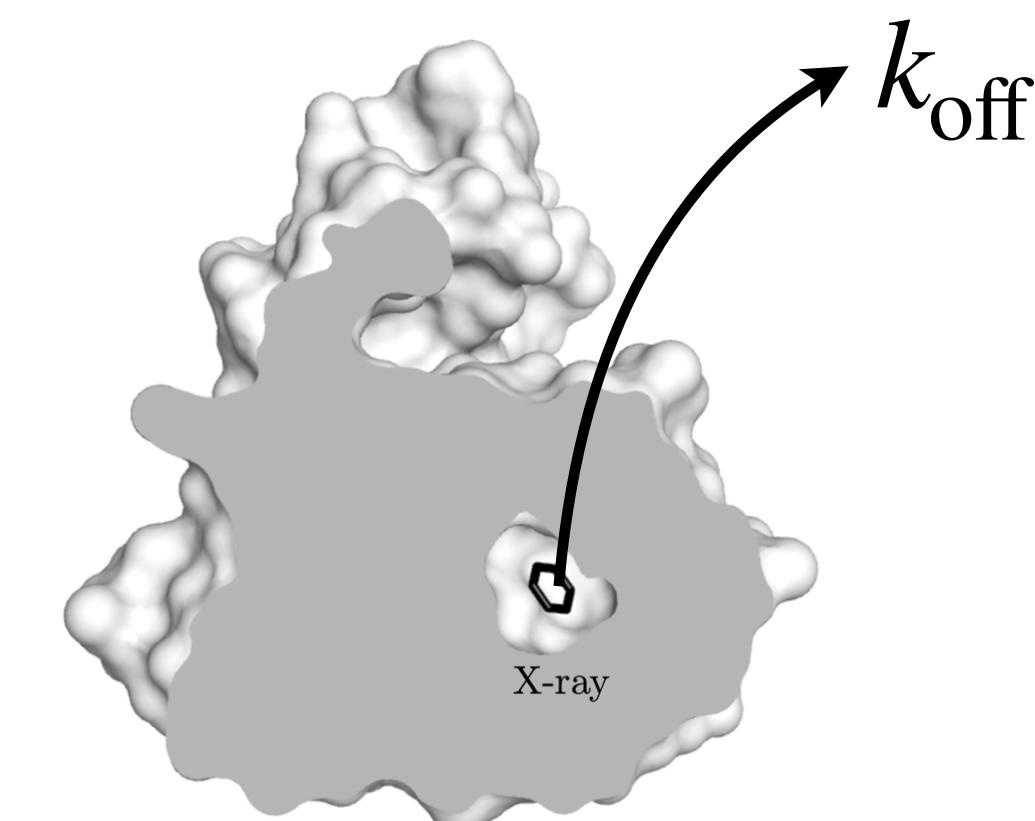
Often interested in kinetics of rare events

e.g. how much time (on average) does it take to go between metastable states A and B

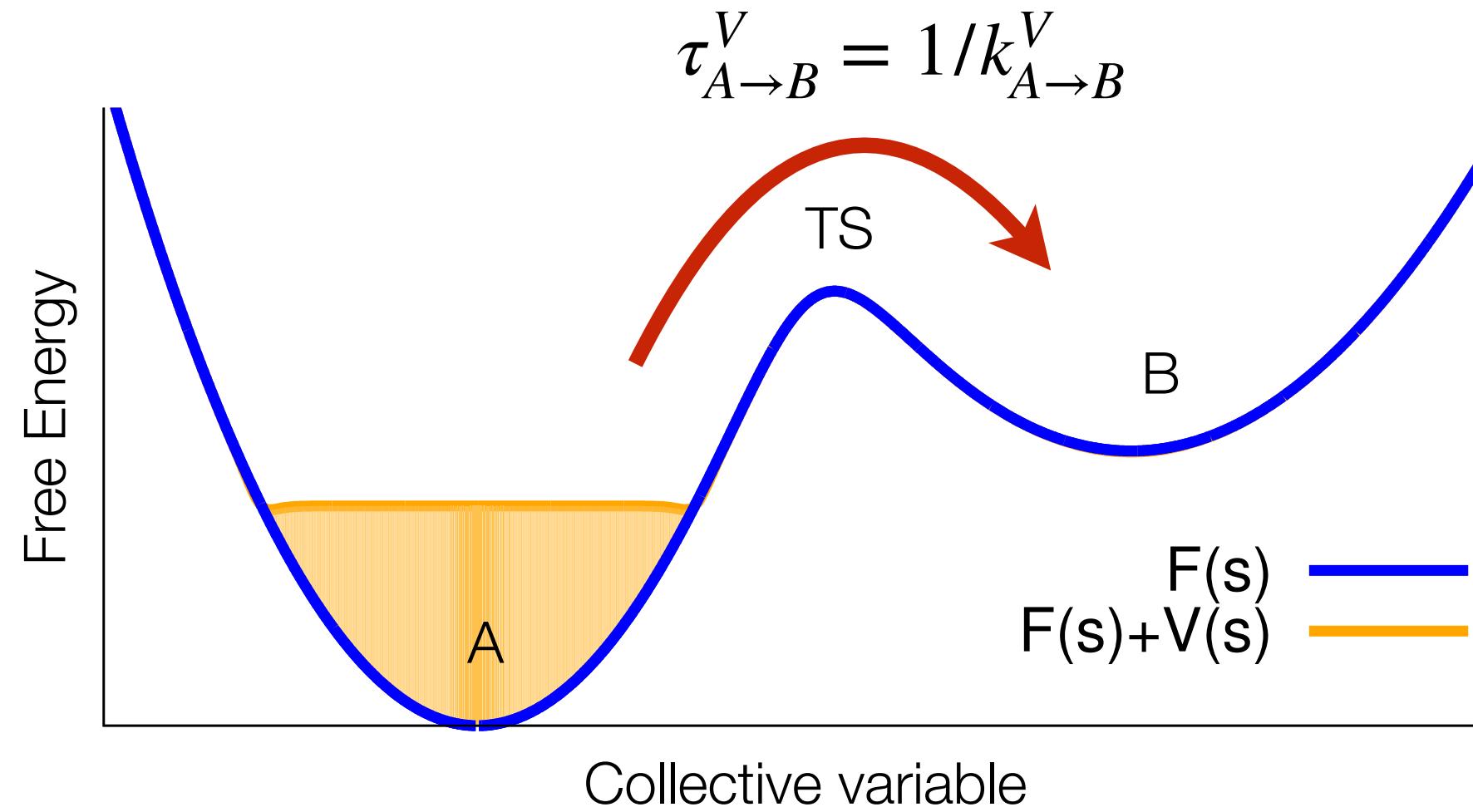
Important example: ligand-protein interactions

In vivo efficacy correlated to the residence time^a

→ unbinding time (k_{off})



Kinetics of Rare Events from Biased Simulations



Often interested in kinetics of rare events

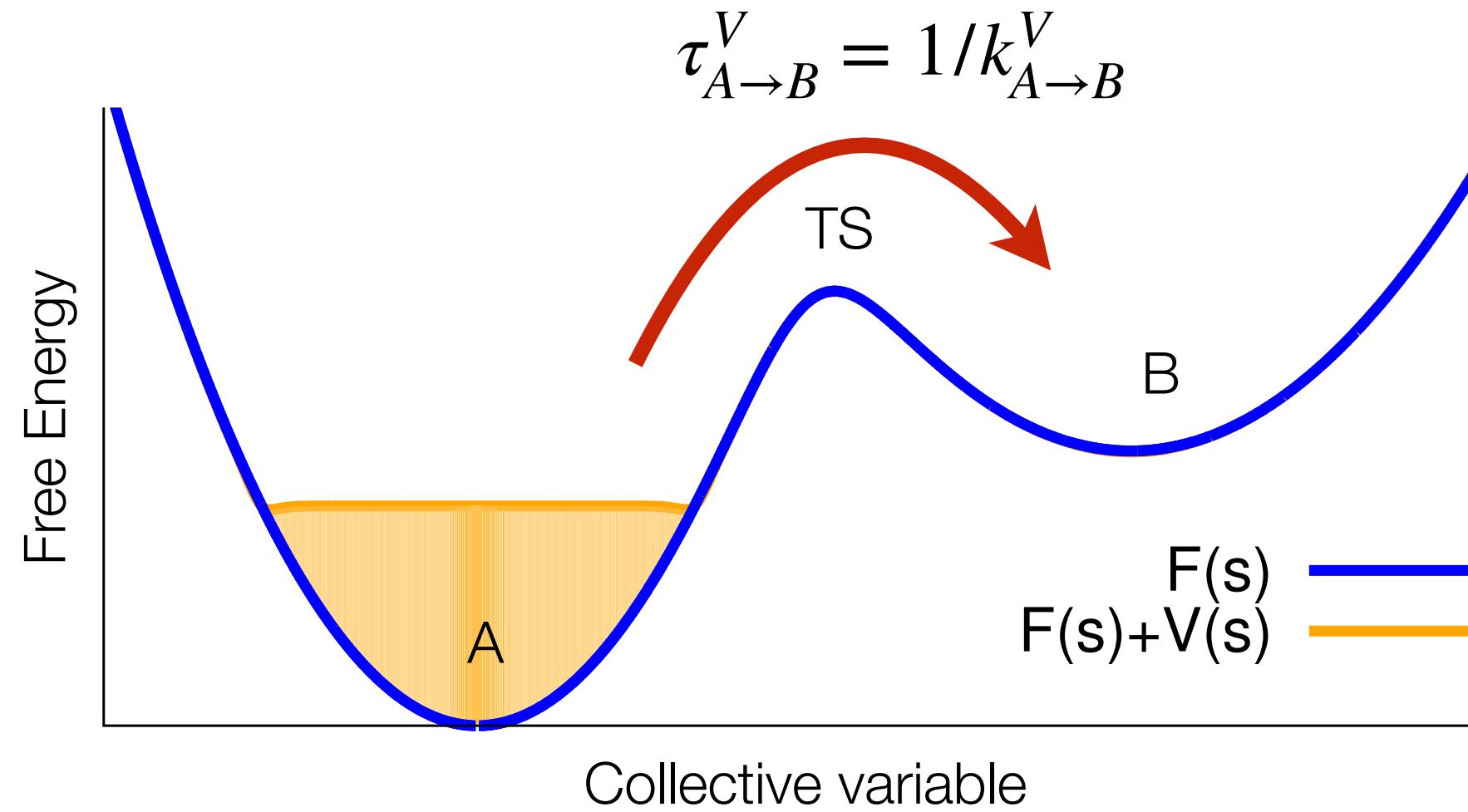
e.g. how much time (on average) does it take to go between metastable states A and B

Biasing will accelerate the kinetics (of course what we want)

$$\tau_{A \rightarrow B}^V \ll \tau_{A \rightarrow B}$$

But how can we relate the biased kinetics to the real unbiased kinetics?

Kinetics of Rare Events from Biased Simulations

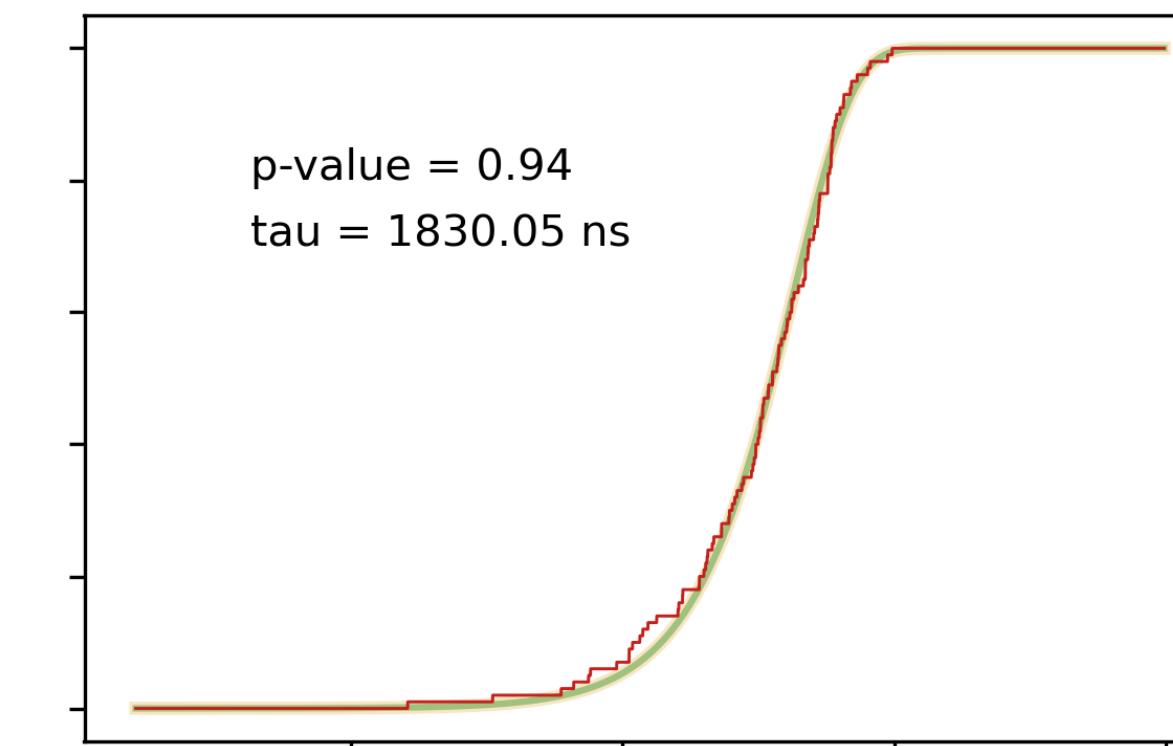


Real kinetic can obtained from bias simulations via time rescaling^{a,b} (hyperdynamics)

$$\tau_{A \rightarrow B} = \tau_{A \rightarrow B}^V \cdot \langle e^{\beta V(s)} \rangle_V$$

valid if the transition state (TS) is **bias-free**,
e.g. can be achieved with infrequent metadynamics^b

can check the validity of the results via p-value test^c

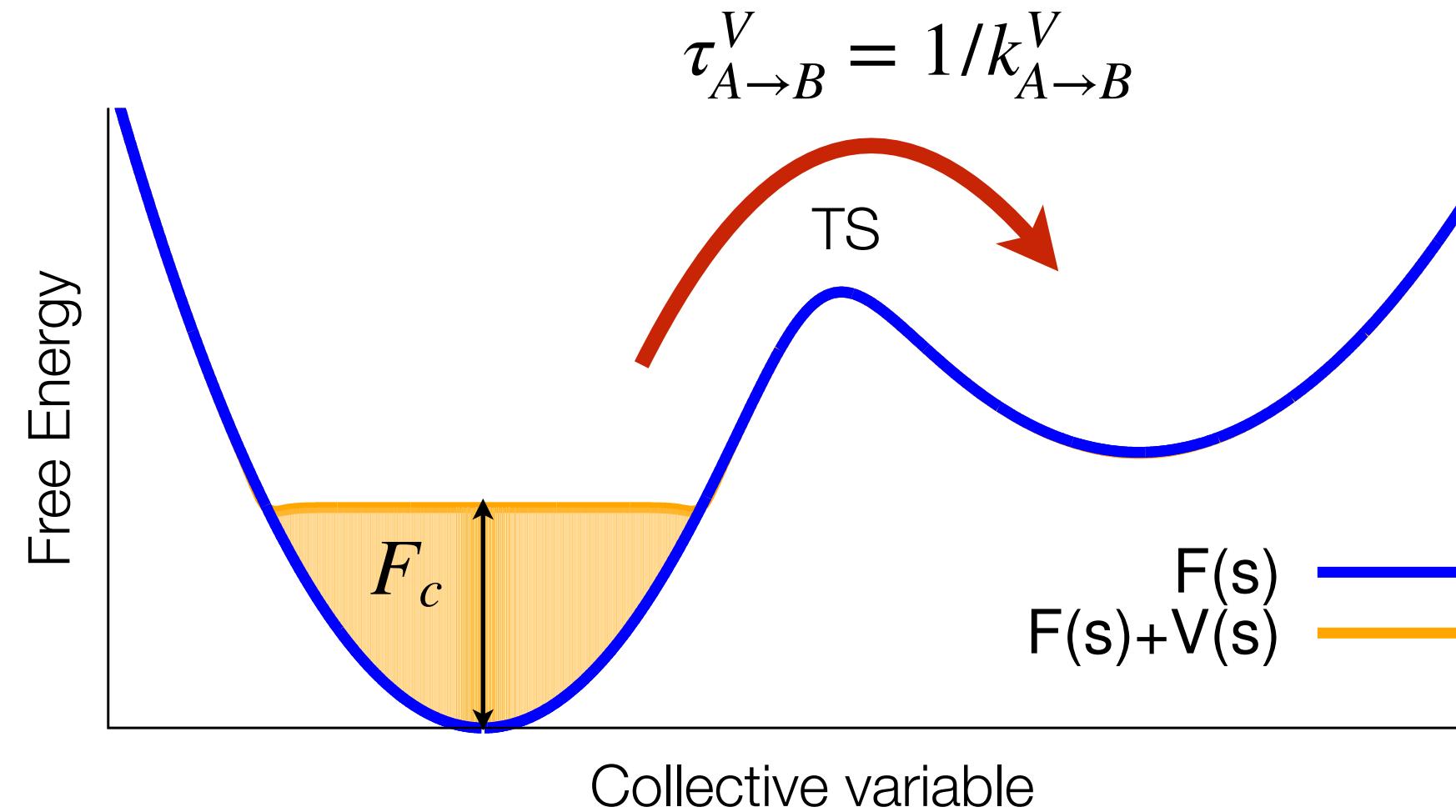


^a Voter, PRL 1997; Grubmueller PRE 1995

^b Tiwary and Parrinello, PRL 2013

^c Salvalaglio, Tiwary and Parrinello, JCTC 2014

Kinetics of Rare Events from Variationally Enhanced Sampling



Real kinetic can obtained from bias simulations via time rescaling (hyperdynamics)

$$\tau_{A \rightarrow B} = \tau_{A \rightarrow B}^V \cdot \langle e^{\beta V(s)} \rangle_V \quad \text{valid if the transition state (TS) is \b{bias-free},}$$

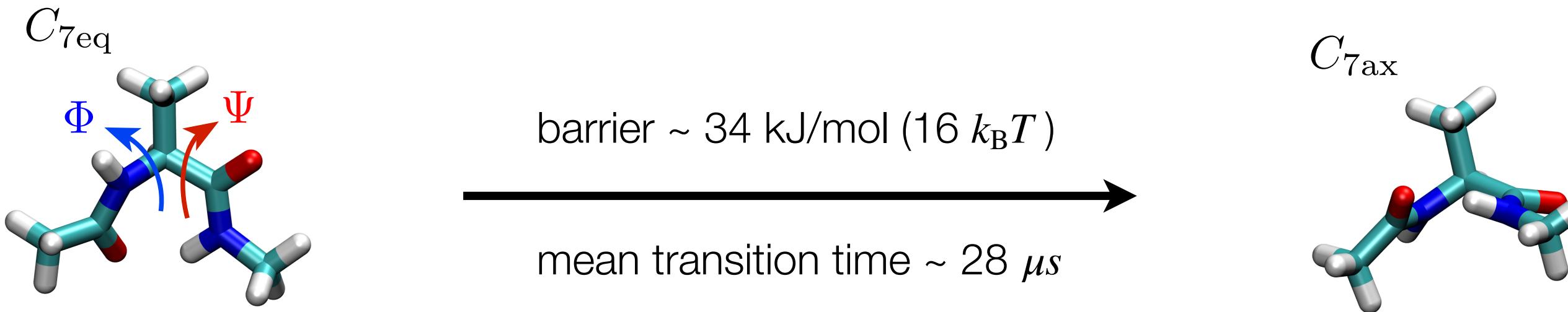
Use VES to construct a bias that only fills the free energy landscape up to a given level (F_c)

$$V(\mathbf{s}; \boldsymbol{\alpha}) = \boxed{v(\mathbf{s}; \boldsymbol{\alpha})} \cdot \boxed{S(-v(\mathbf{s}; \boldsymbol{\alpha}) - F_c)}$$

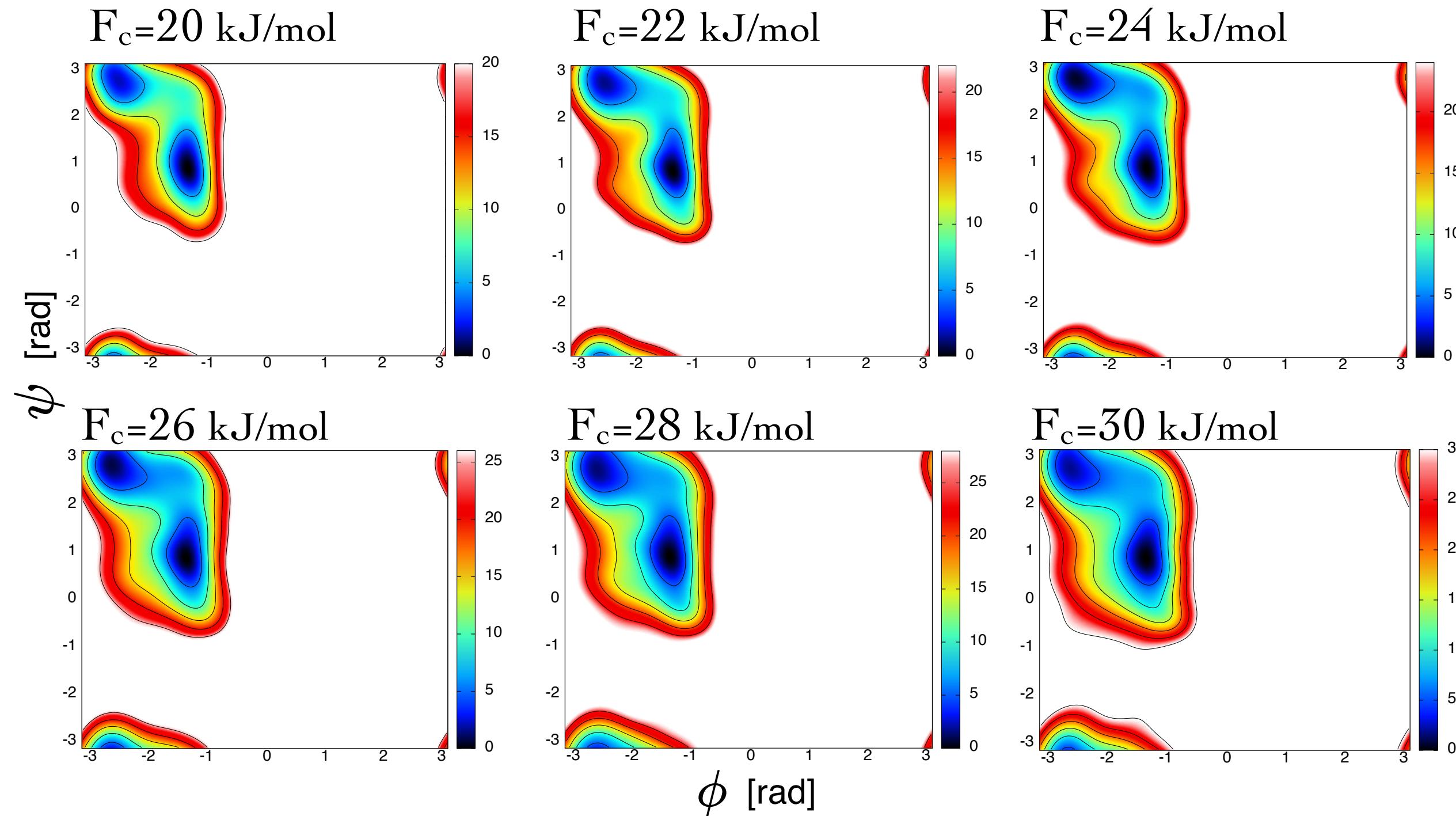
linear expansion switching function

optimized using appropriate target distribution that flattens sampling

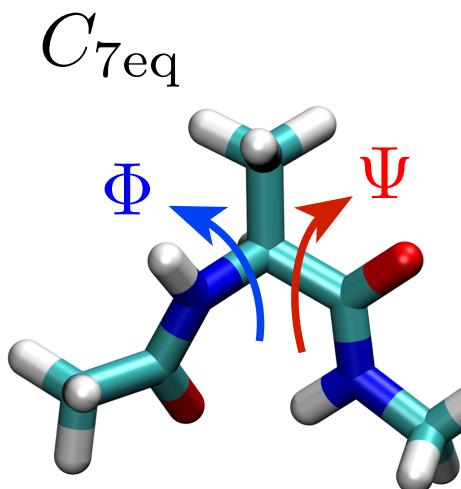
Kinetics from VES: Simple Example



Generate bias potentials with different F_c values from 20 to 30 kJ/mol

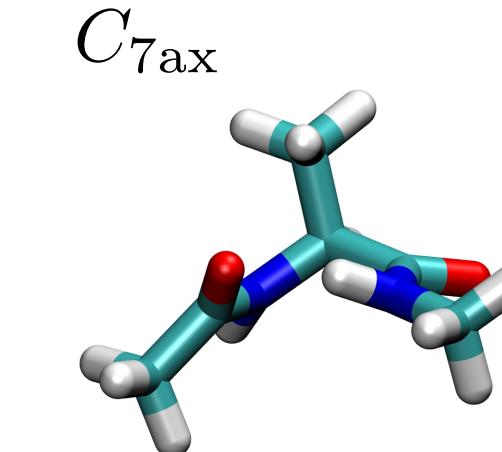


Kinetics from VES: Simple Example

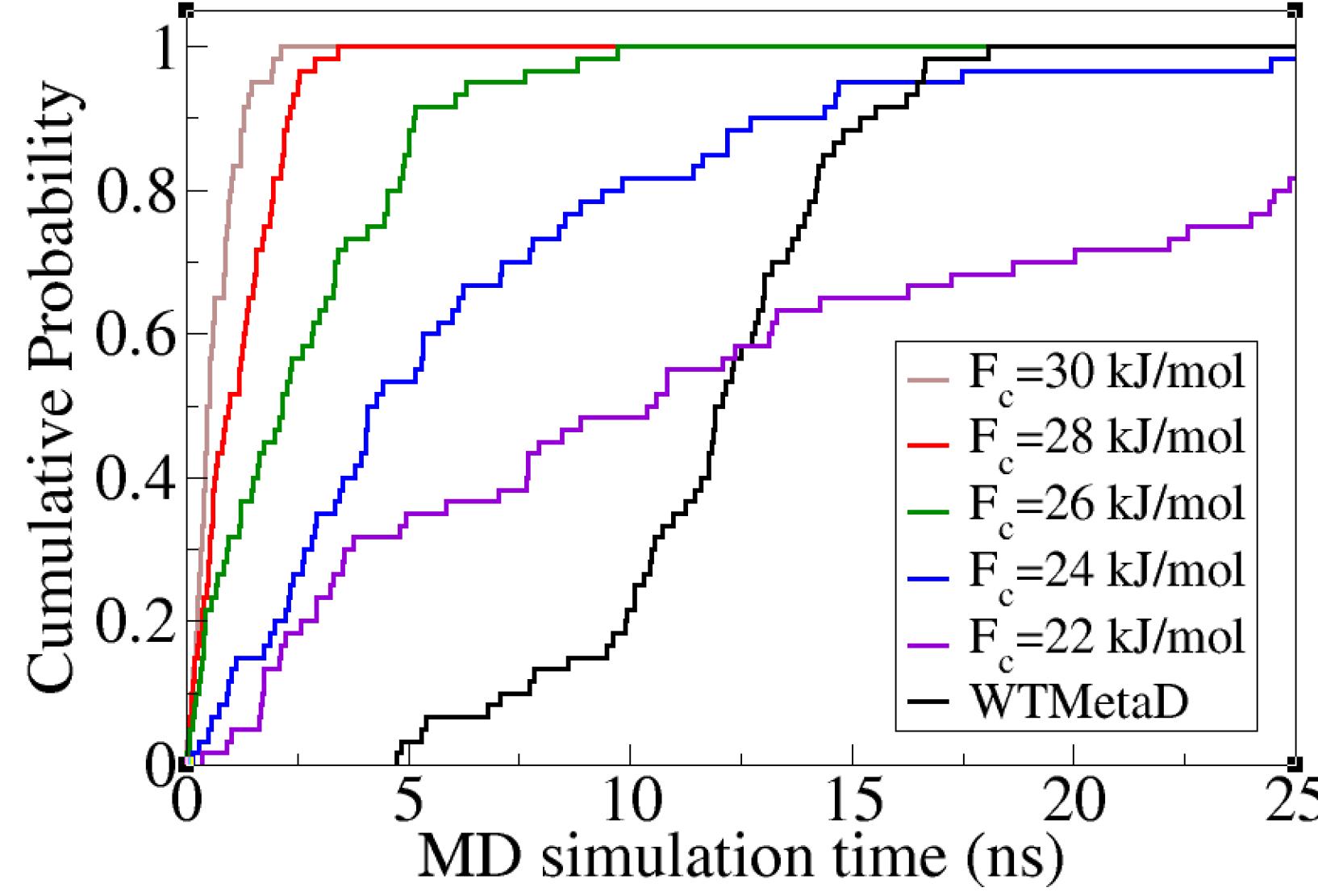


barrier $\sim 34 \text{ kJ/mol} (16 k_B T)$

mean transition time $\sim 28 \mu\text{s}$

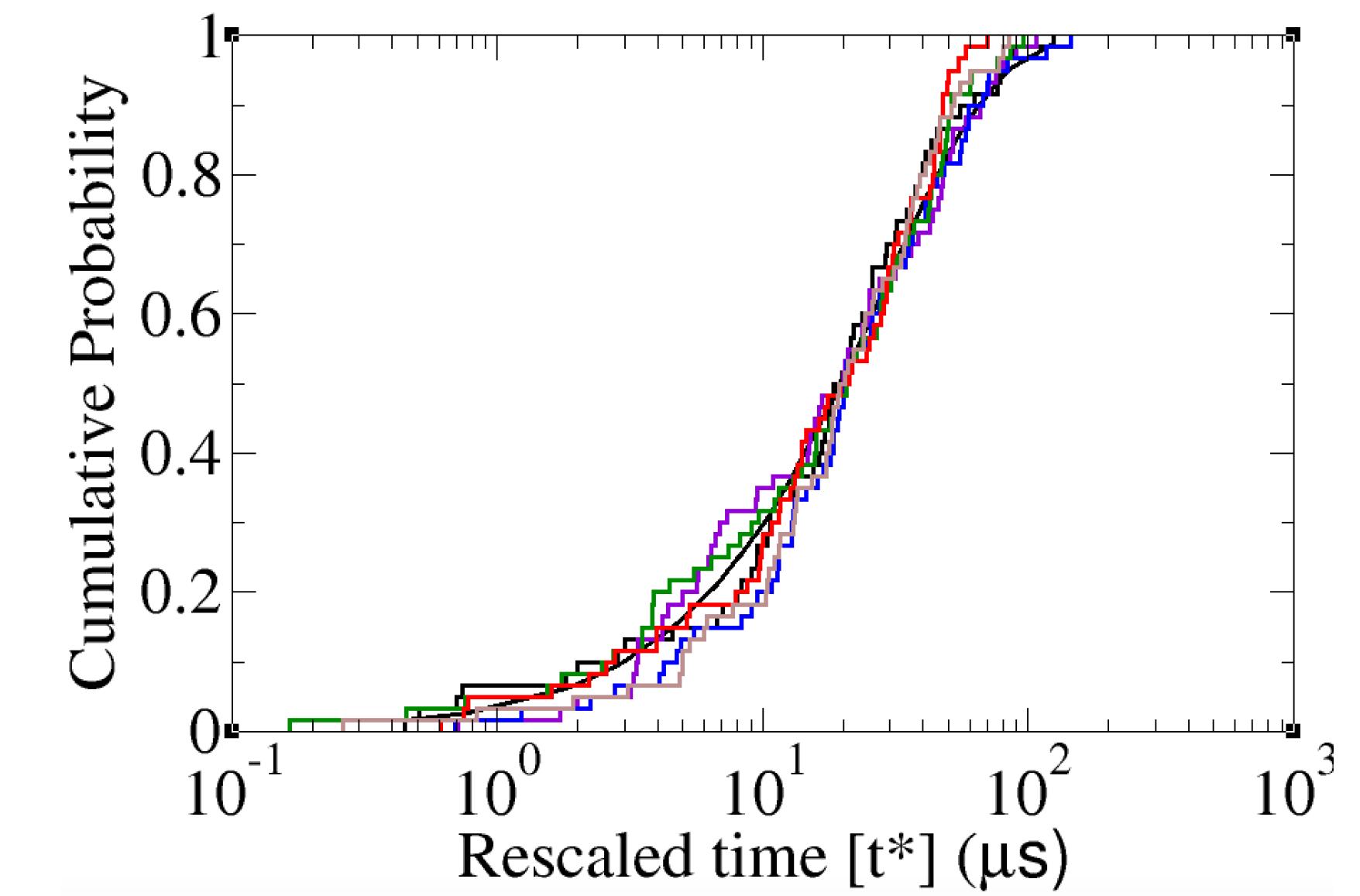


Multiple independent runs for each F_c values (static bias potentials)

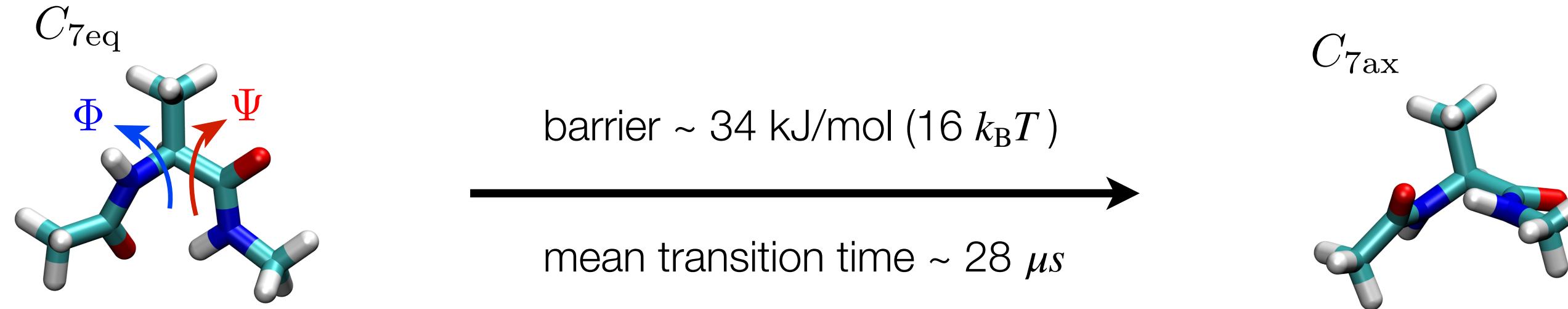


Rescale times

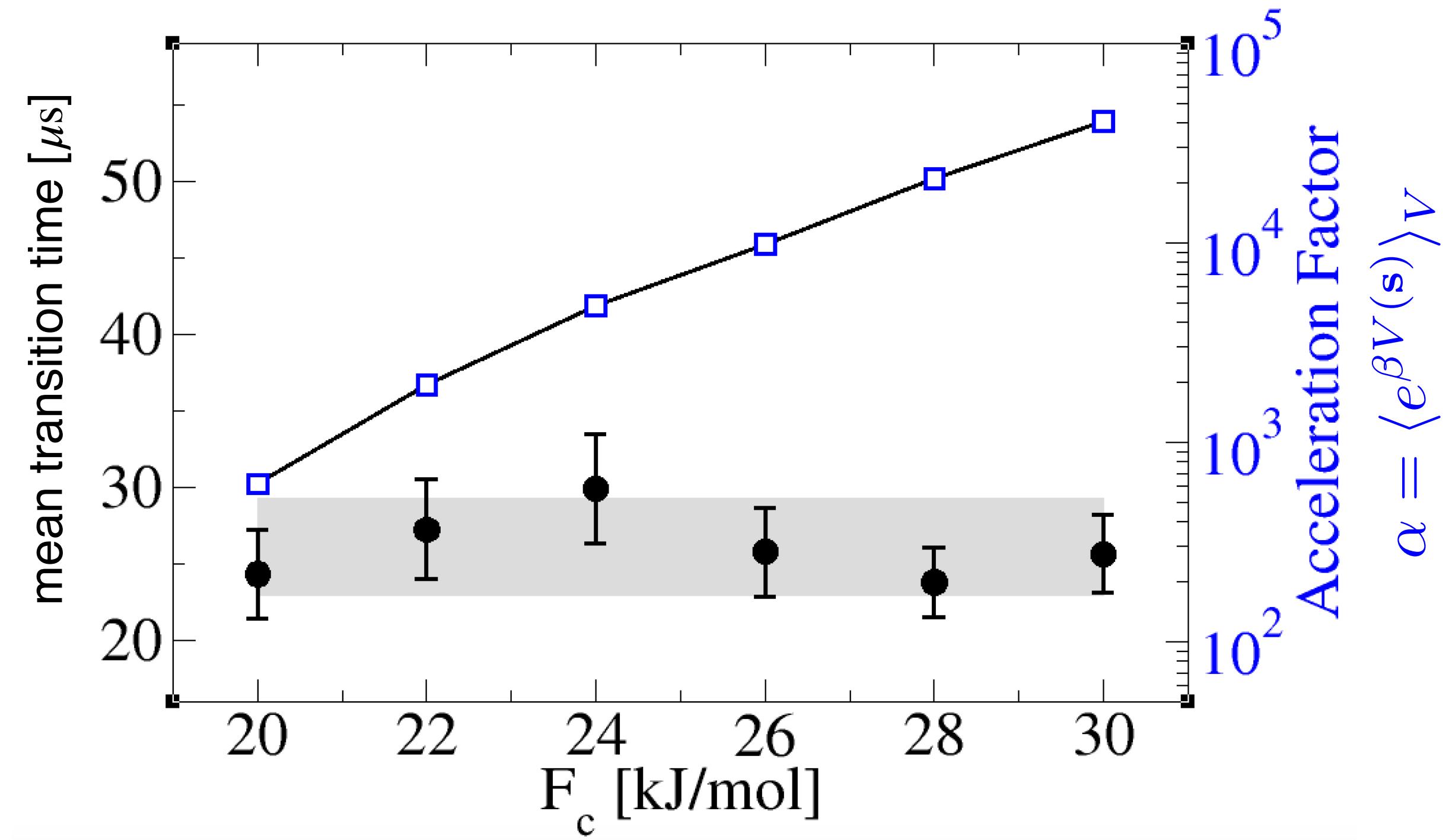
$$\tau_{A \rightarrow B} = \tau_{A \rightarrow B}^V \cdot \langle e^{\beta V(s)} \rangle_V$$



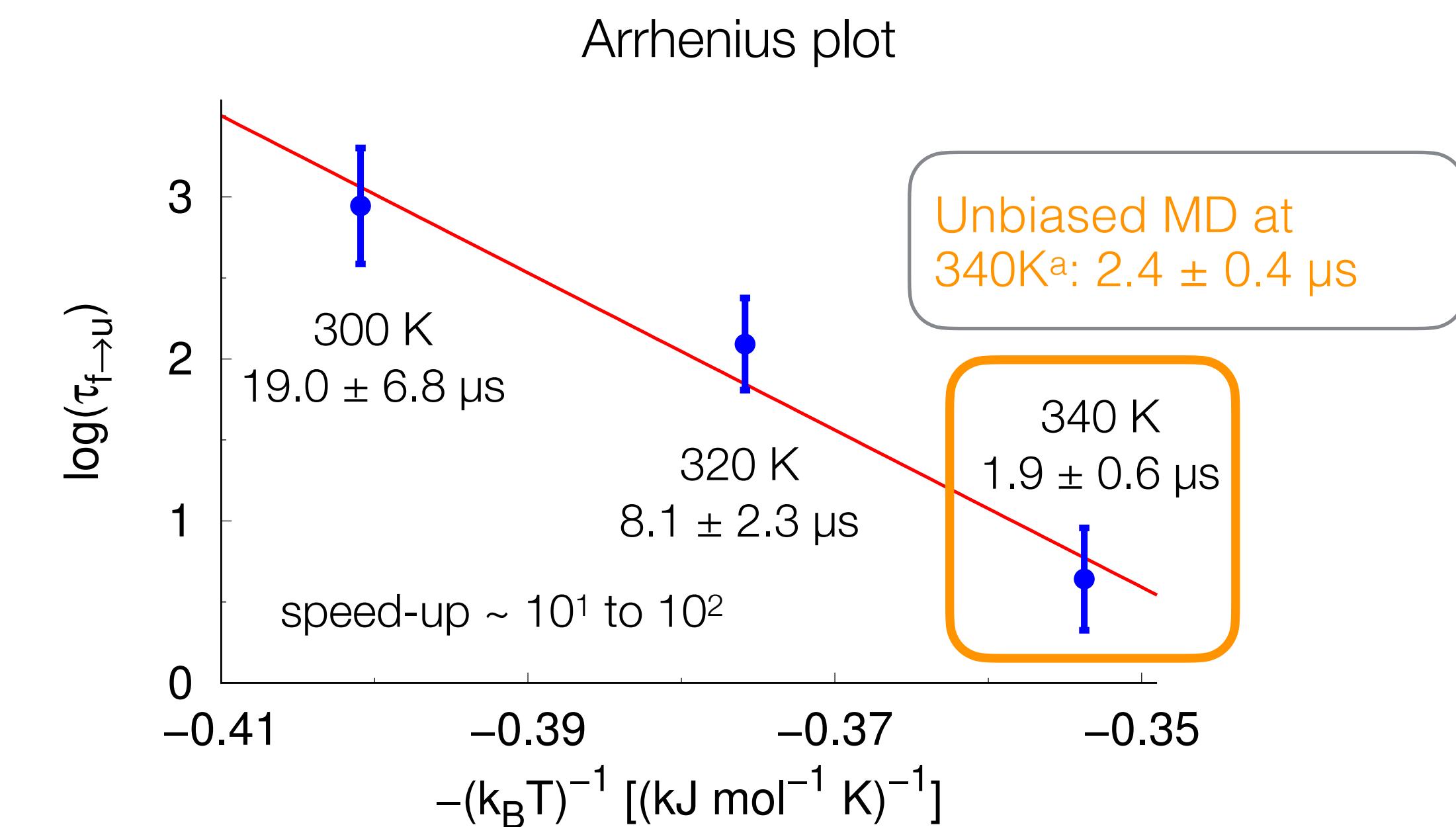
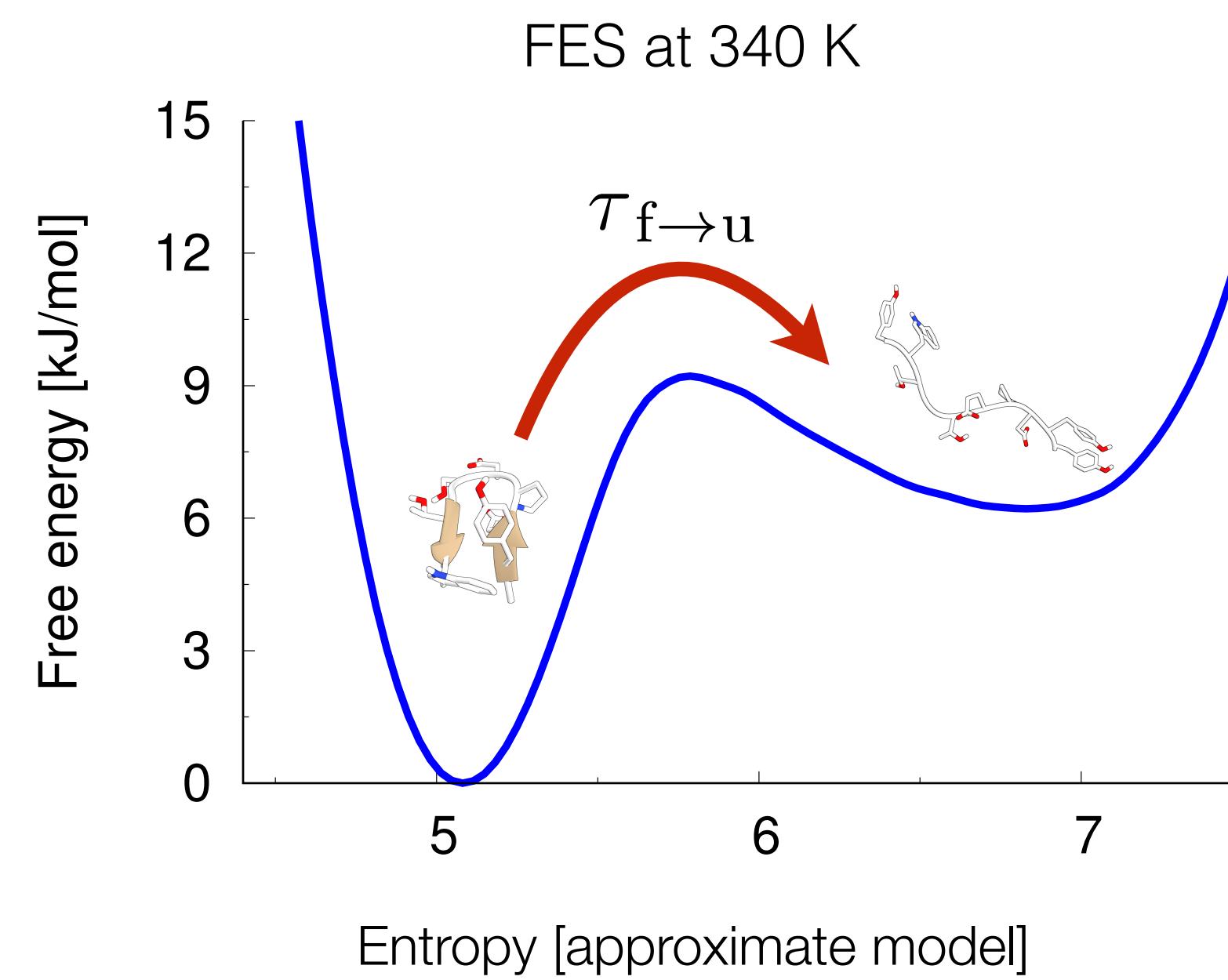
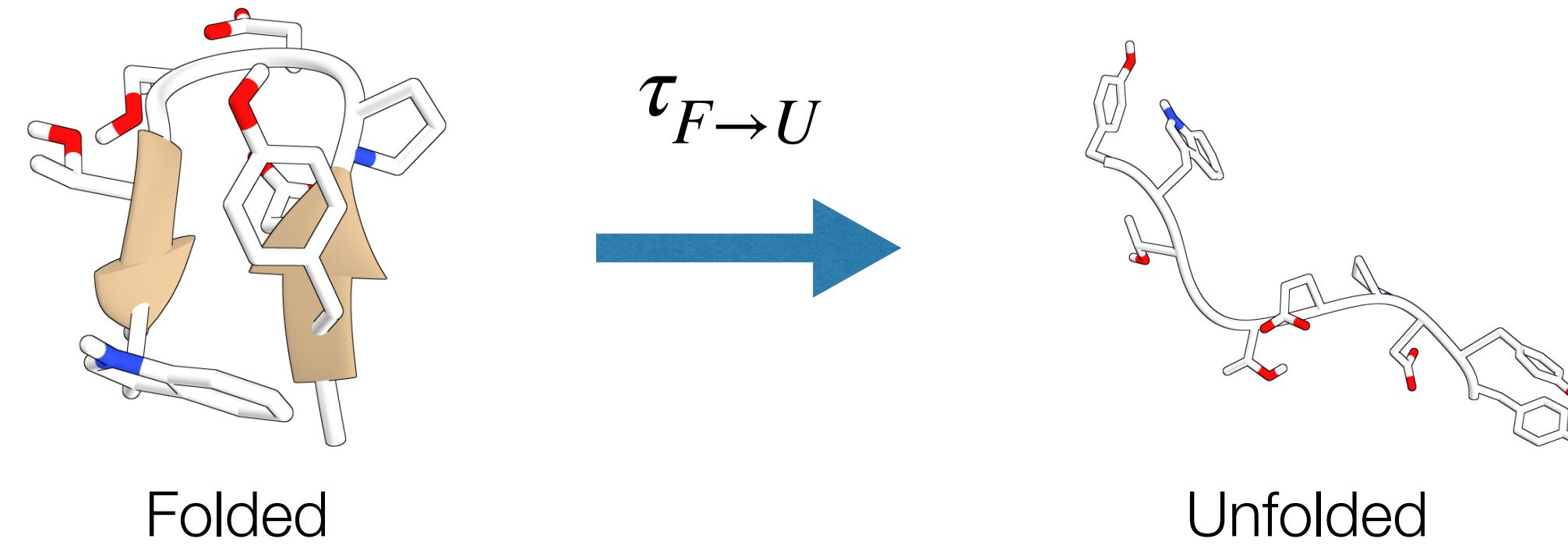
Kinetics from VES: Simple Example



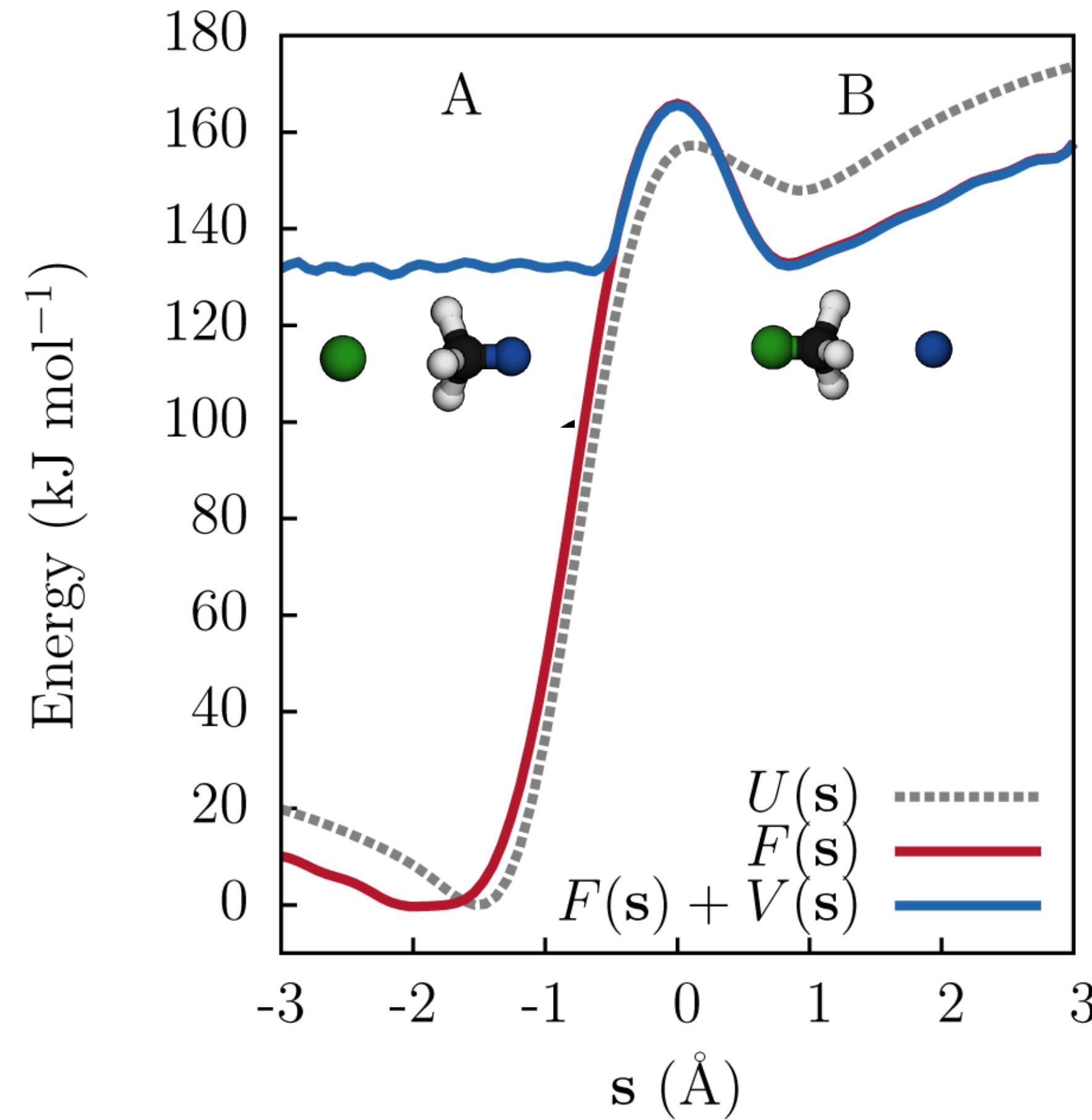
Speedup of up to 10^4 - 10^5 compared to unbiased simulations



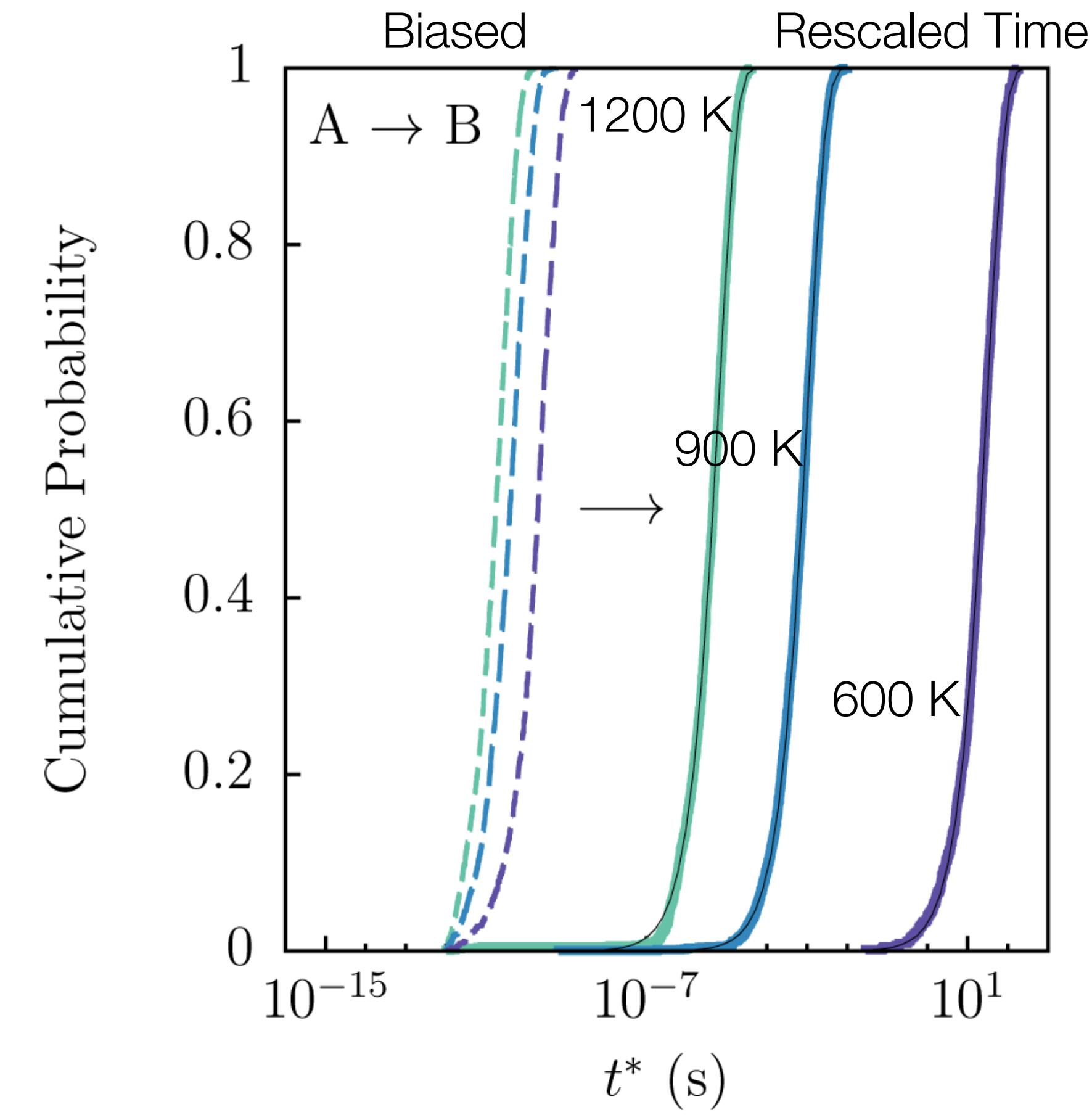
Unfolding Times of Chignolin



Chemical Reactions



PM6 Semi-empirical MD



Wavelet-based Bias Potentials

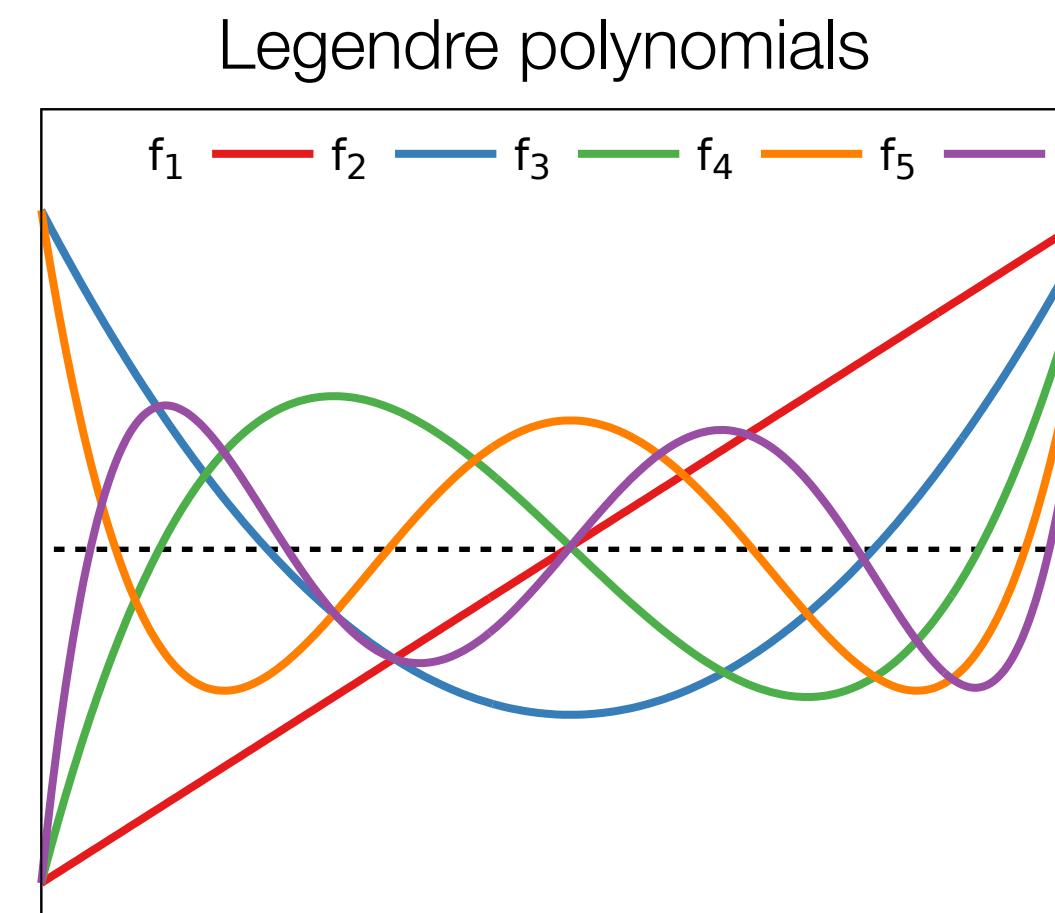
Linear bias expansion in a set of basis functions

$$V(\mathbf{s}; \boldsymbol{\alpha}) = \sum_{\mathbf{k}} \alpha_{\mathbf{k}} \cdot f_{\mathbf{k}}(\mathbf{s})$$

So far: delocalized (and orthogonal) basis functions
(e.g. plane waves, Chebyshev or Legendre polynomials)

But, behavior not always robust

Might localized basis functions perform better?



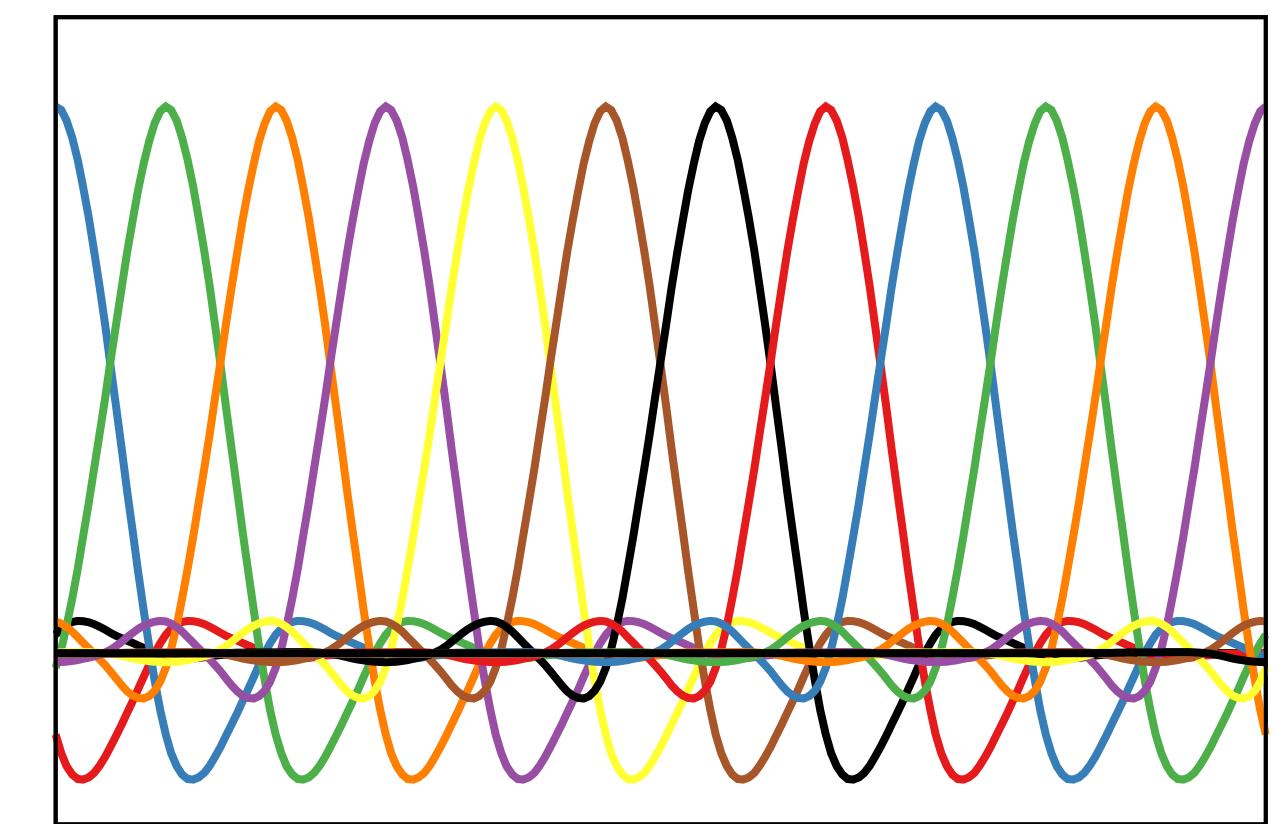
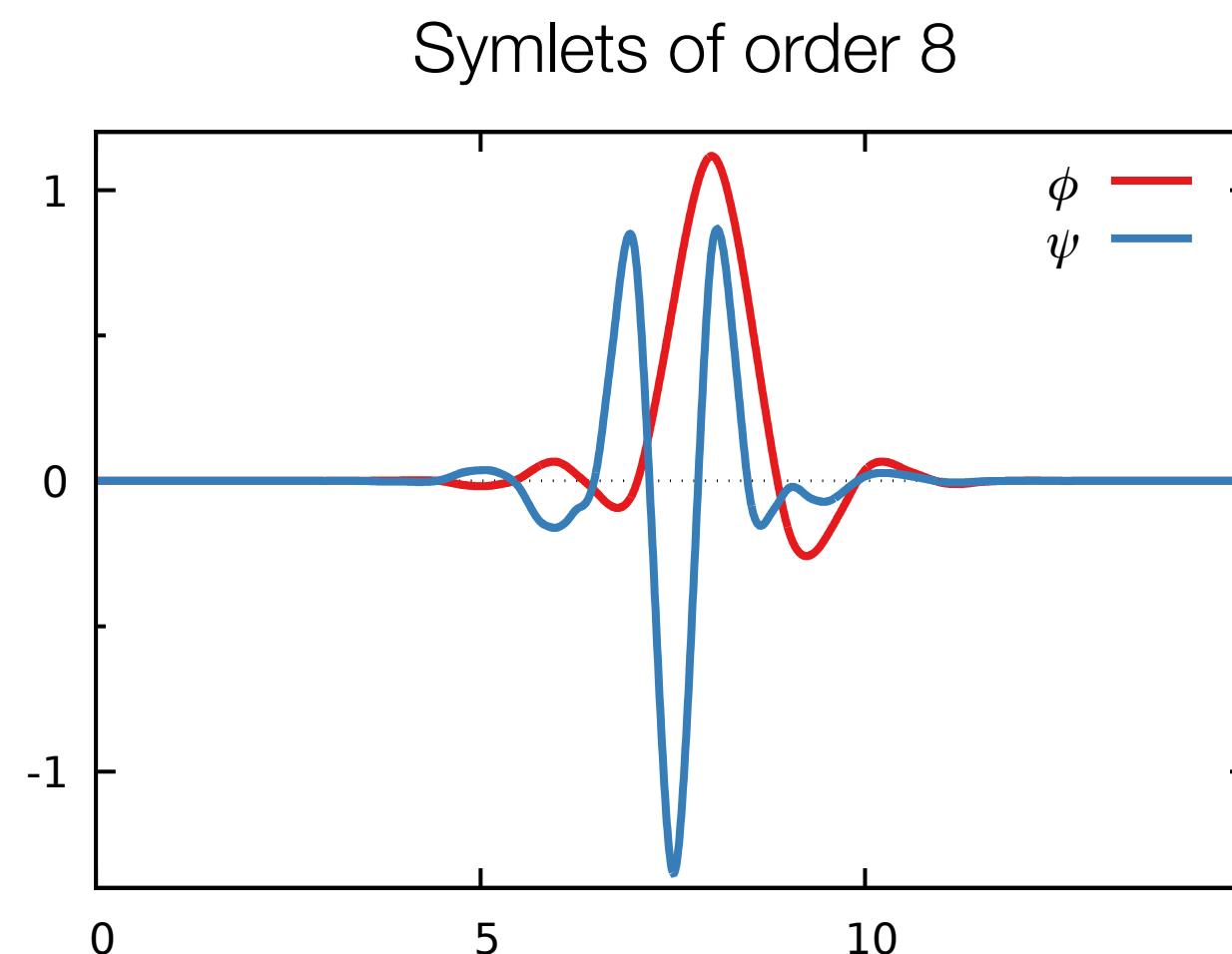
Daubechies Wavelets

Form an orthogonal basis set
(also between resolutions)

Multiresolution property: can increase accuracy
by introducing another level of wavelets

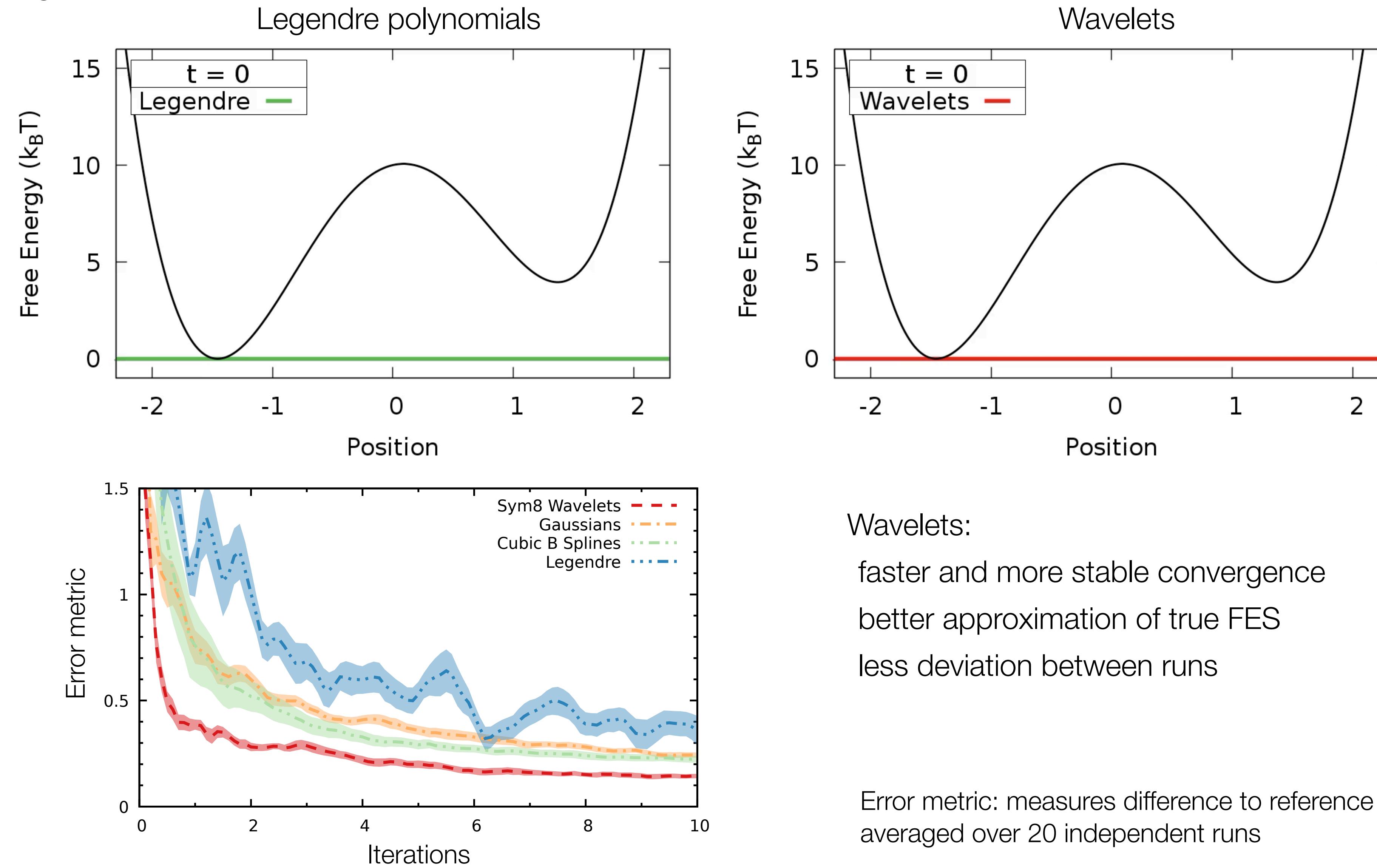
$$f(x) = \sum_k \alpha_k \phi_k^j(x) + \sum_k \sum_{l \geq j} \alpha_{l,k} \psi_k^l(x)$$

Better to use symmetric wavelets (symlets)



Wavelet-based Bias Potentials

Single particle moving
in a 1D potential



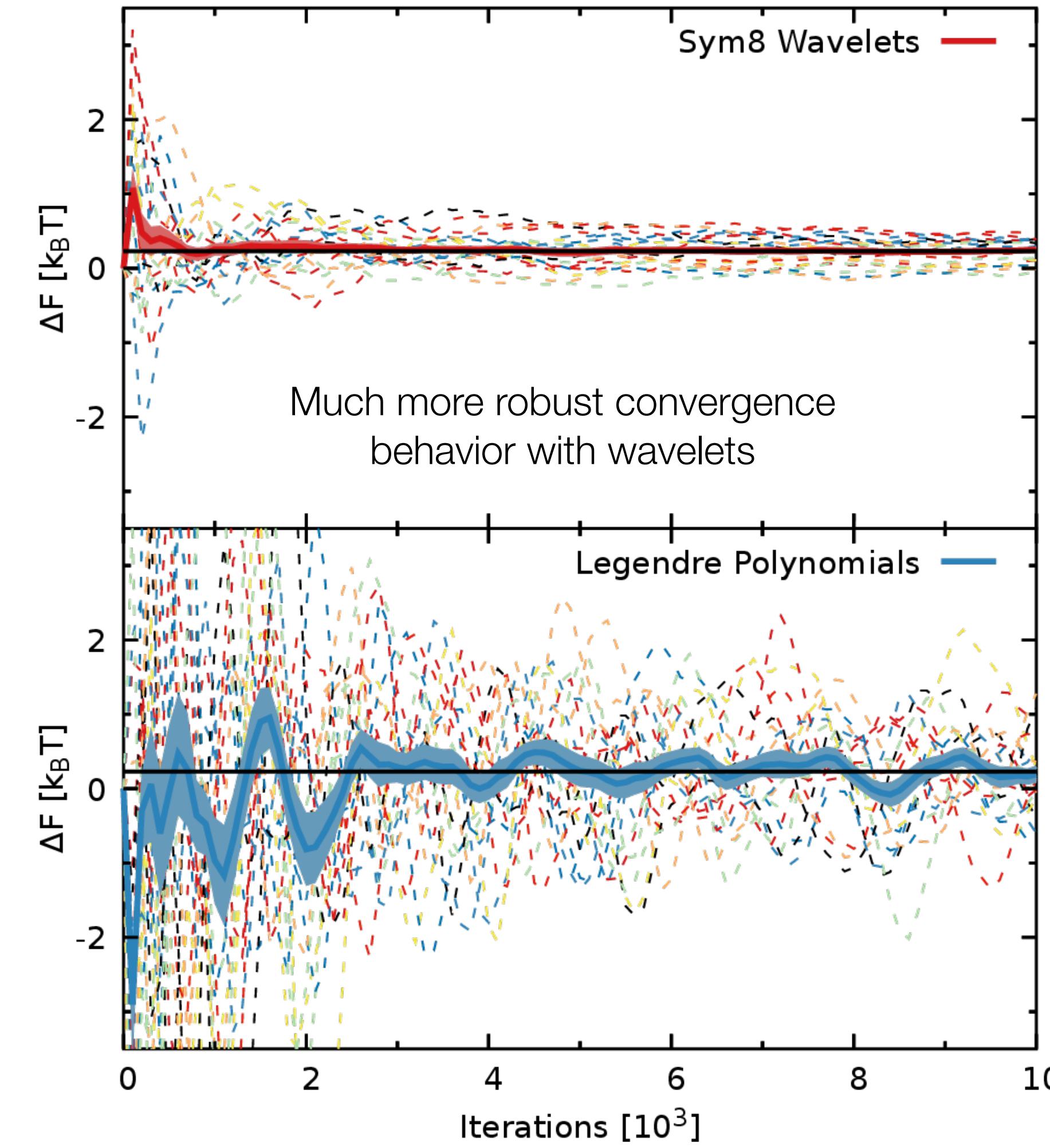
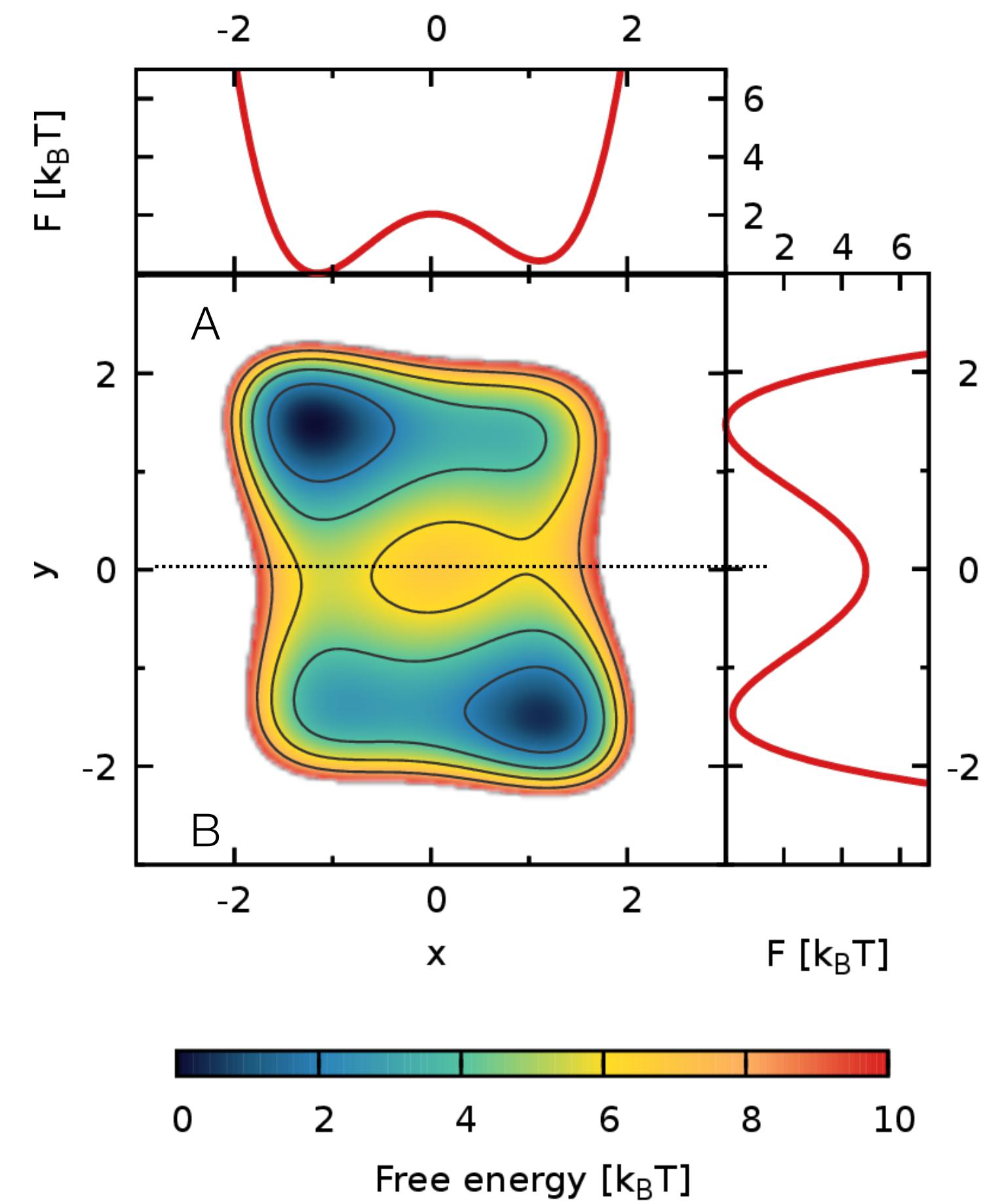
Wavelet-based Bias Potentials

Wolff-Quapp potential

Tensor product of wavelets

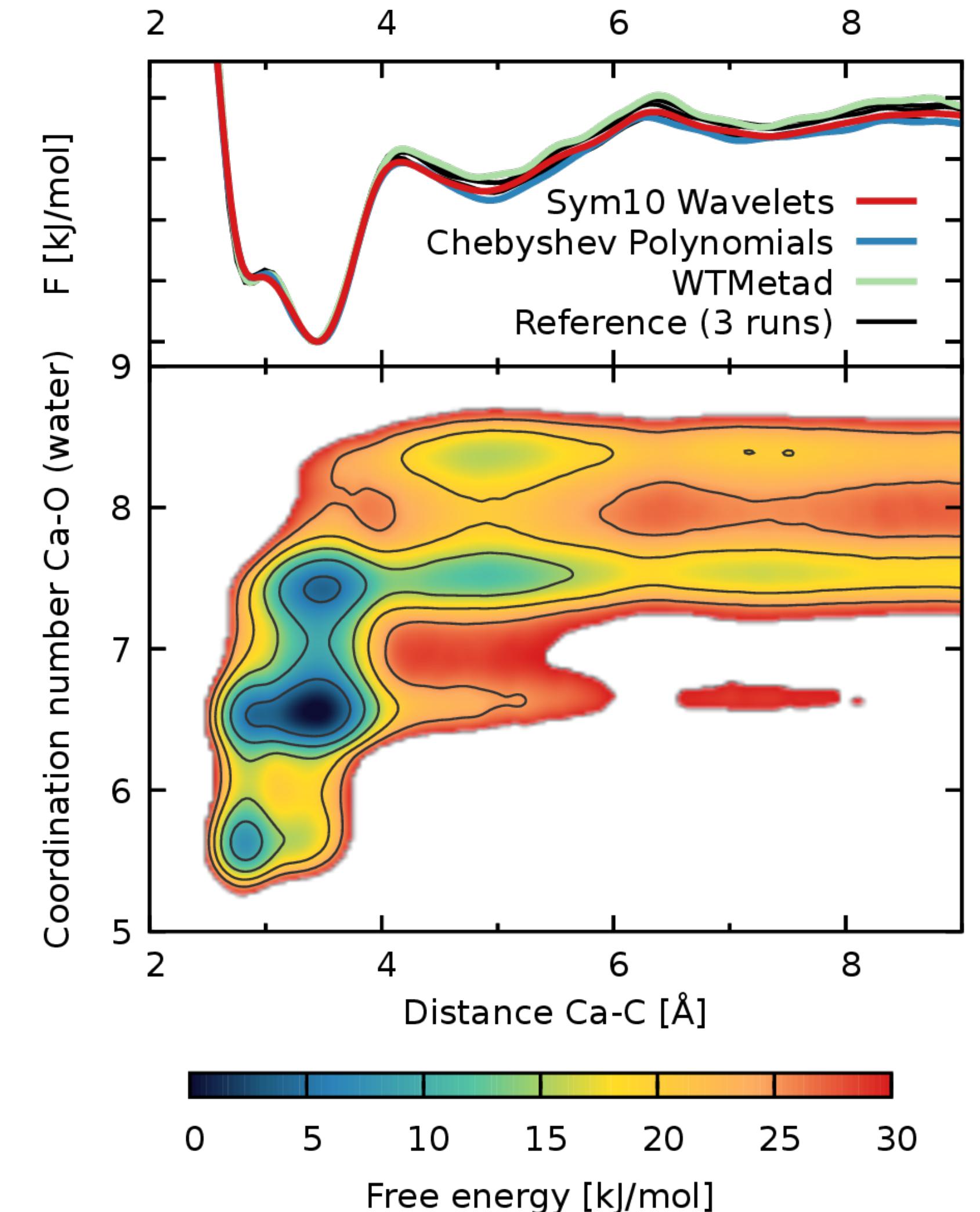
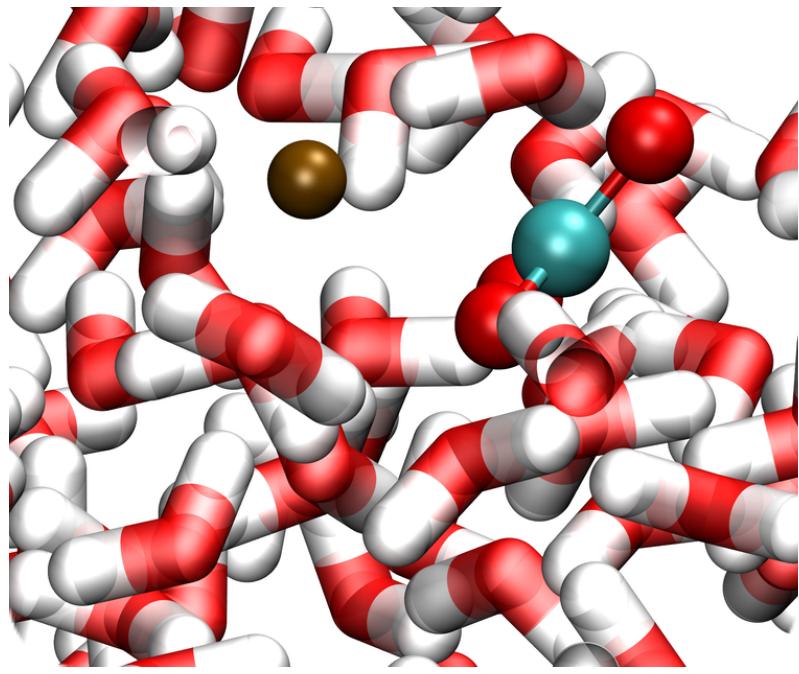
Free energy difference

$$\Delta F_{A,B} = -\frac{1}{\beta} \log \frac{\int_A ds \exp[-\beta F(s)]}{\int_B ds \exp[-\beta F(s)]}$$



Wavelet-based Bias Potentials

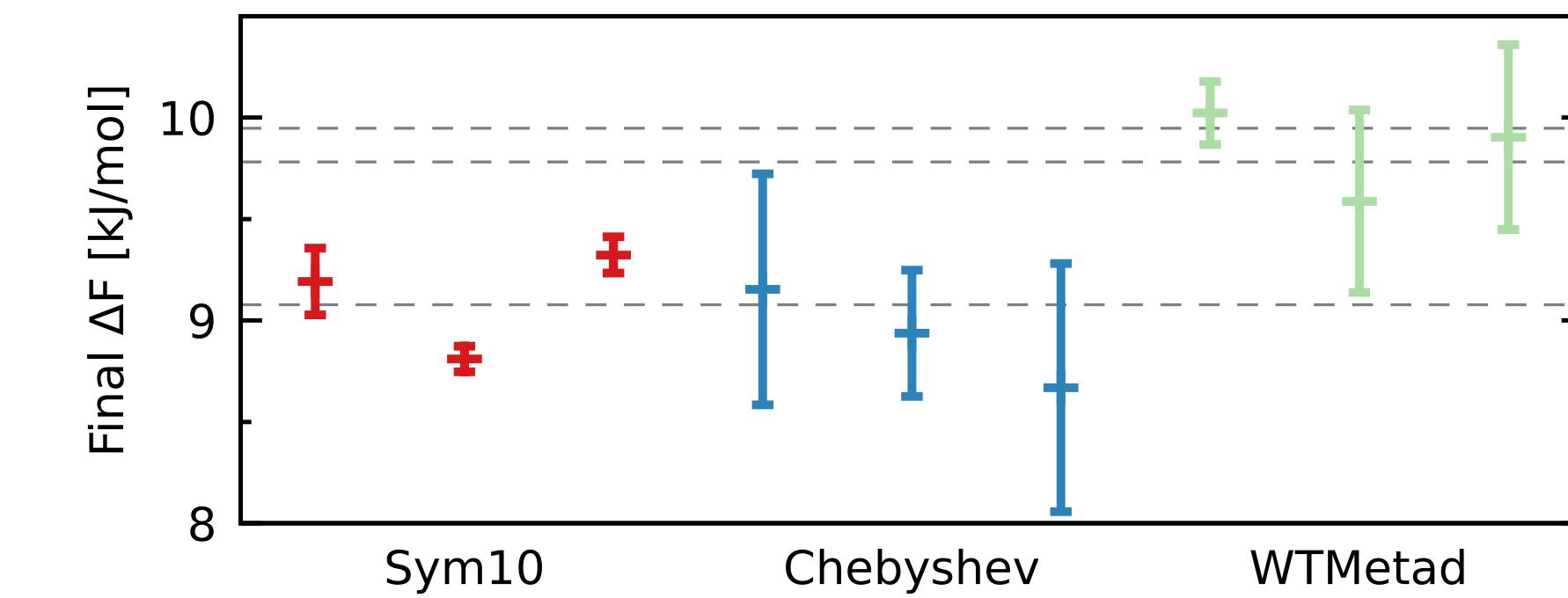
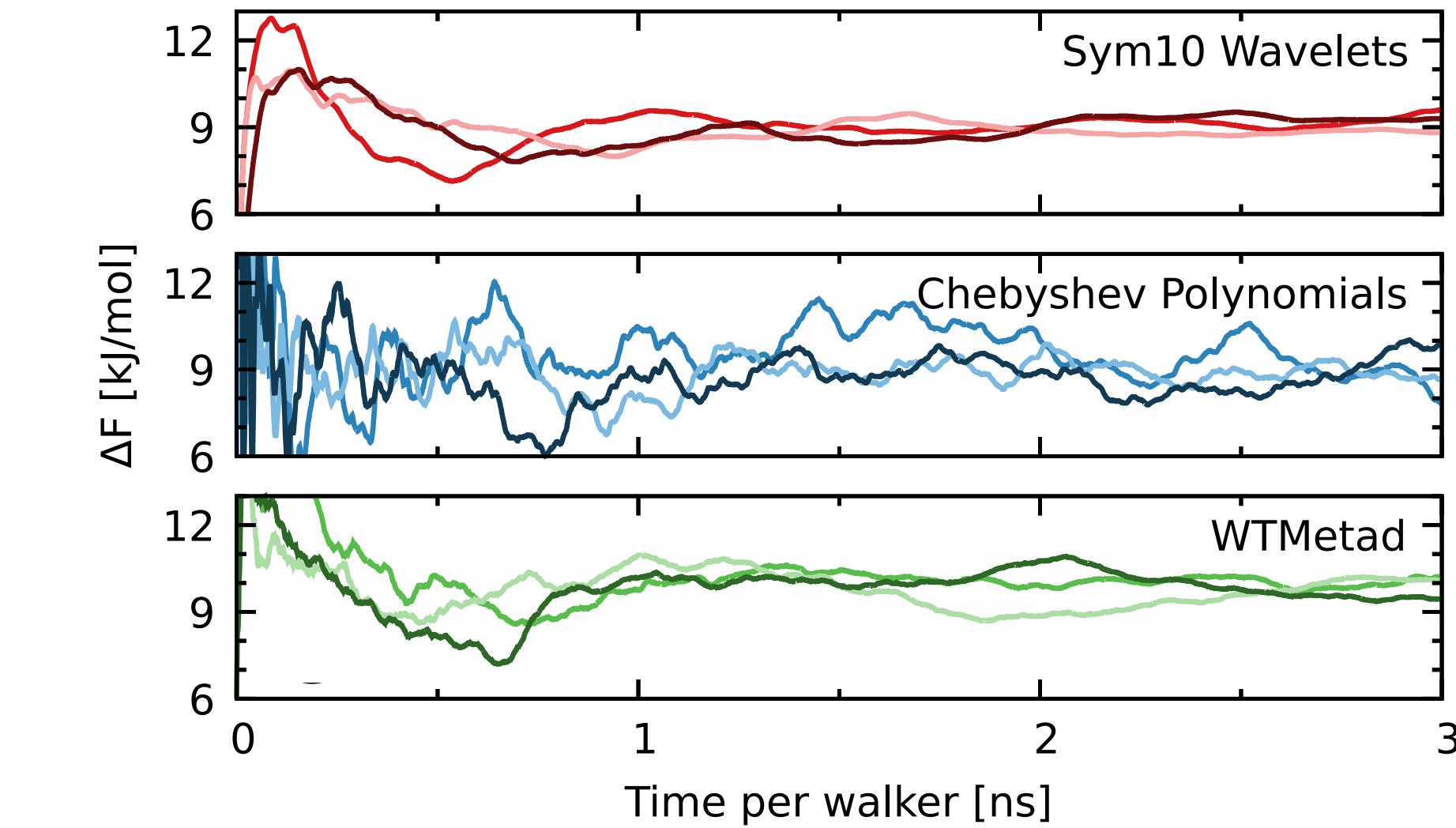
Dissociation of calcium carbonate in water



Free energy difference

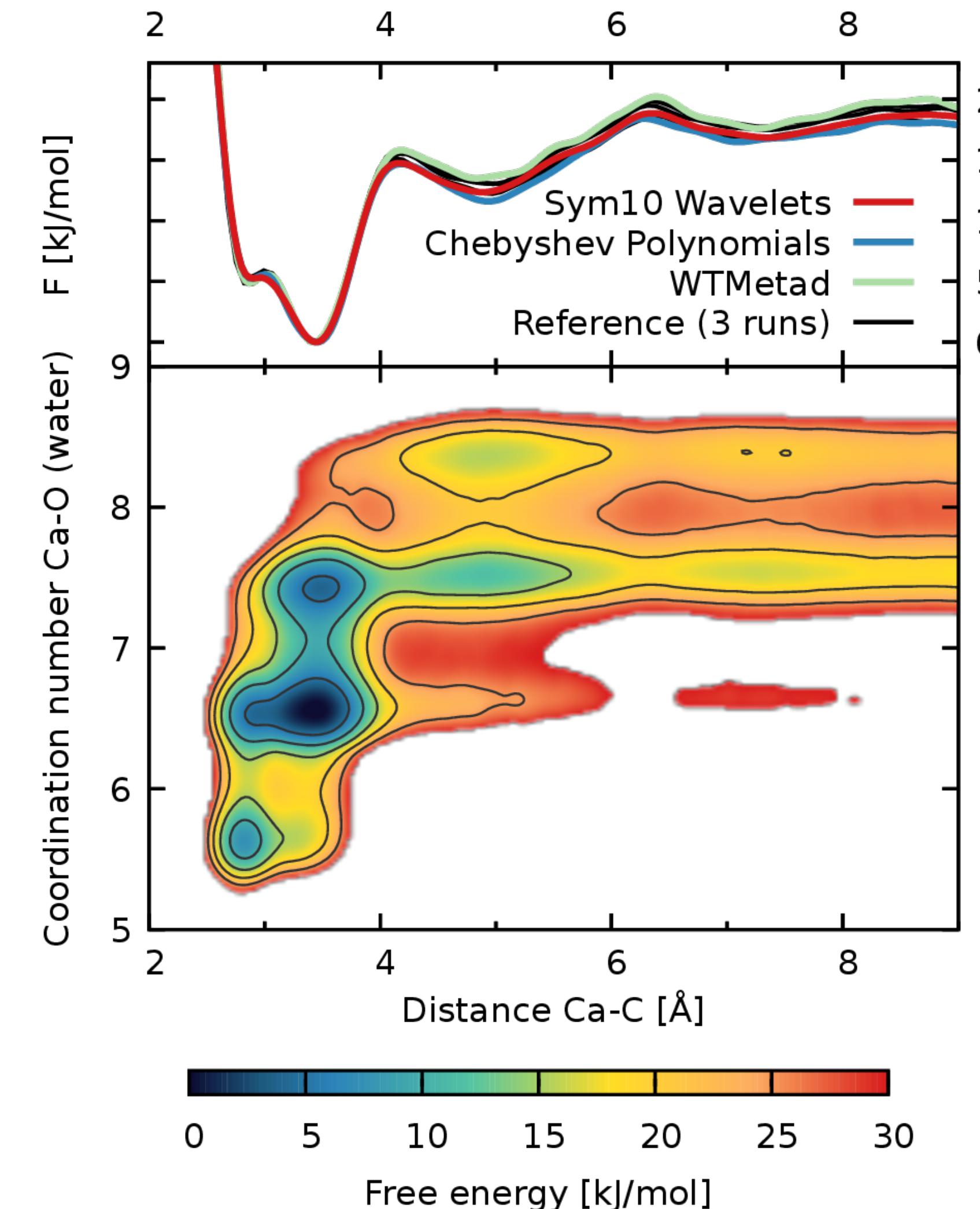
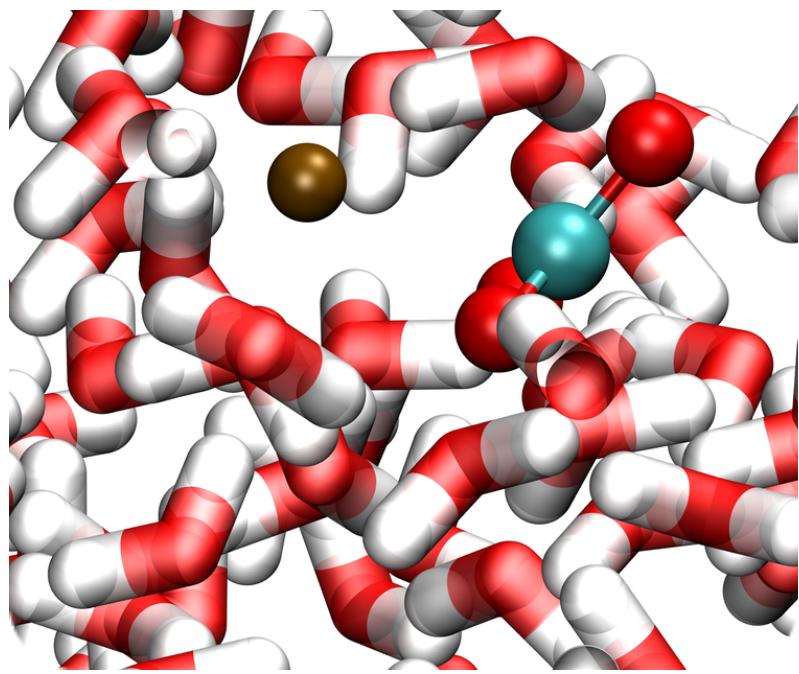
Wavelets much more stable

FES directly from bias



Wavelet-based Bias Potentials

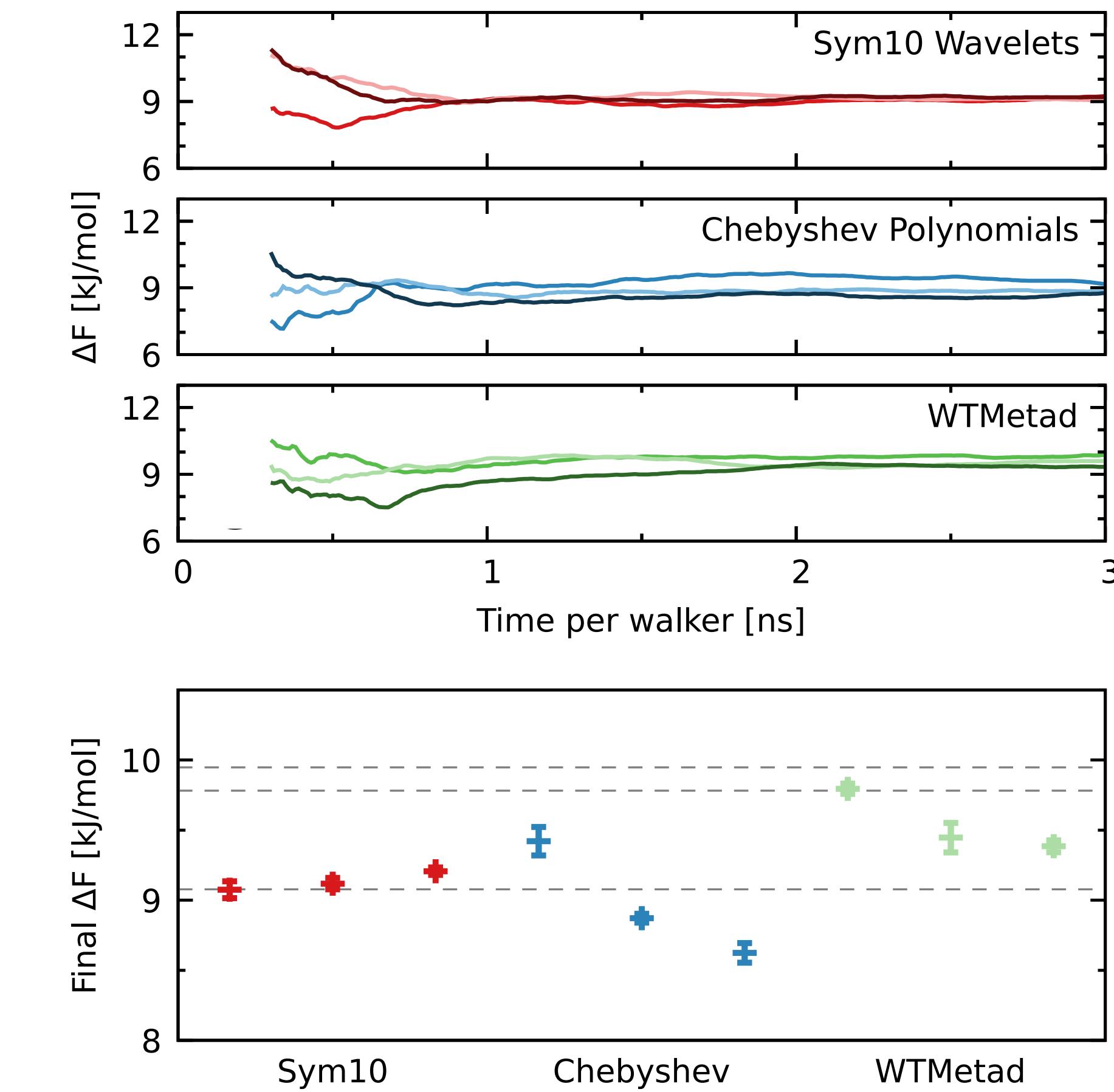
Dissociation of calcium carbonate in water



Free energy difference

Wavelets much more stable

FES from reweighting



Implementation: PLUMED + VES Code

MD Code

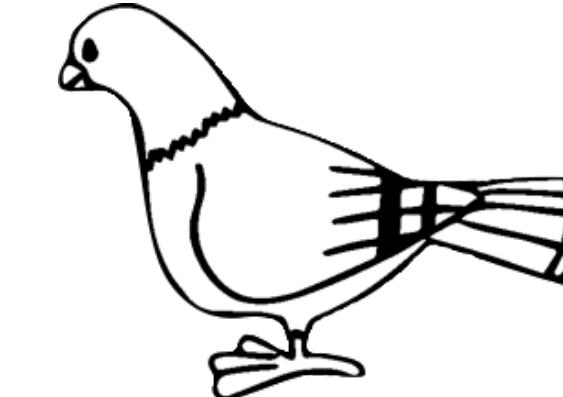
e.g. Gromacs, LAMMPS, CP2K, ESPResSo++, Amber, NAMD, ...



PLUMED 2

www.plumed.org

PLUgin for **Mo**Ecular **D**ynamics (open source)
wide range of collective variables available
metadynamics, umbrella sampling, steered MD, ...
easy to use, extensive tutorials, ...



+

VES Code

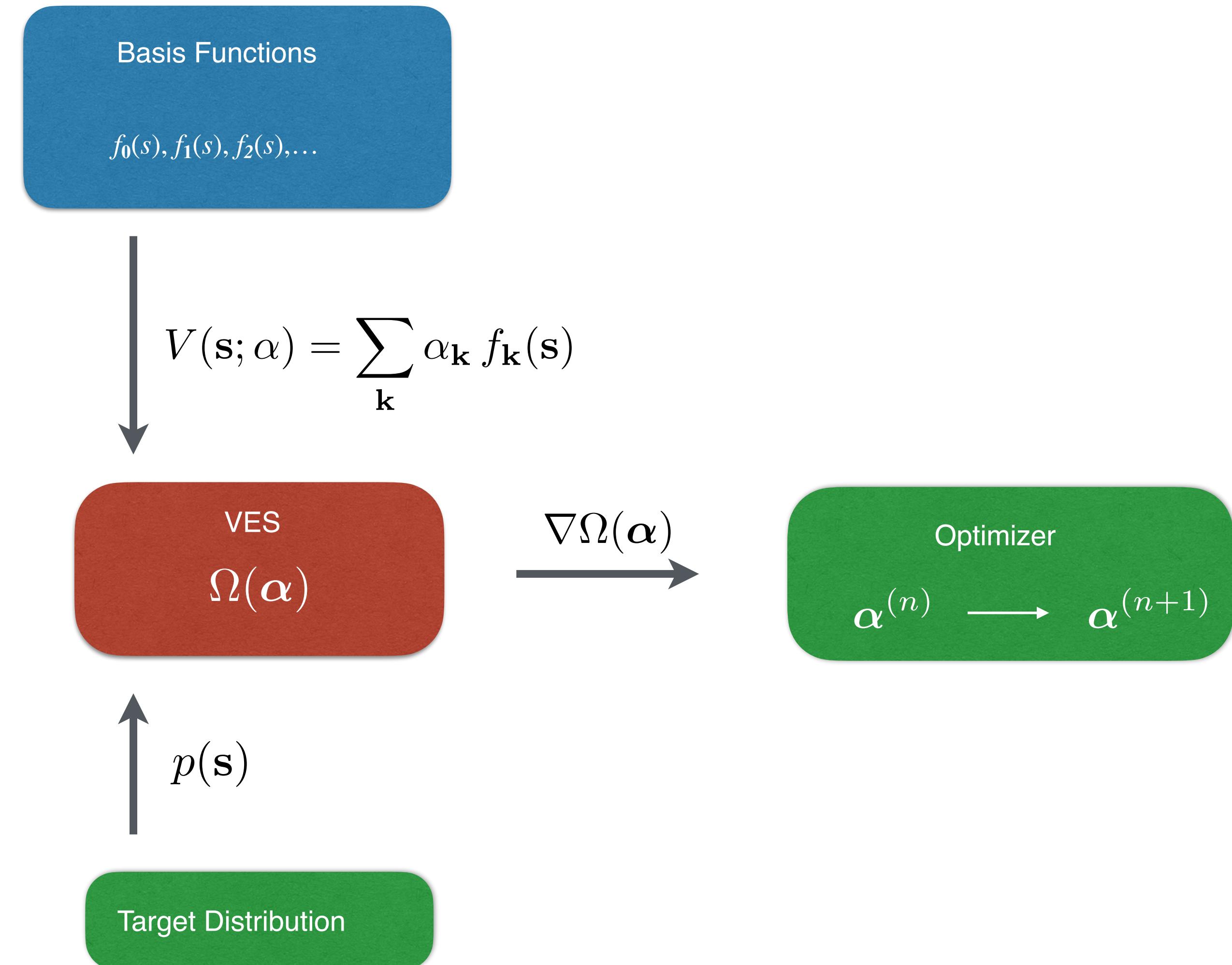
library for PLUMED 2 that implements methods based on VES
modular design - easy to add new features
(basis functions, target distribution, optimizers, ...)
tutorials available
open-source (development on Github)
available in latest version of PLUMED

VES Simulations in PLUMED

Three main ingredients

- Basis set
- Target distribution
- Optimization algorithm

See tutorial



VES Simulations in PLUMED

Can employ multiple walkers

share bias $V(\mathbf{s})$ and cooperatively sample the averages needed for the gradient $\nabla\Omega(\boldsymbol{\alpha})$

observed to significantly improve convergence for difficult cases

Can perform reweighting to obtain equilibrium properties

Can be used to check convergence by reweighing on the CV biased and comparing to results

obtained directly from bias potential (good practice)

Can also reweight for CV not biased

References

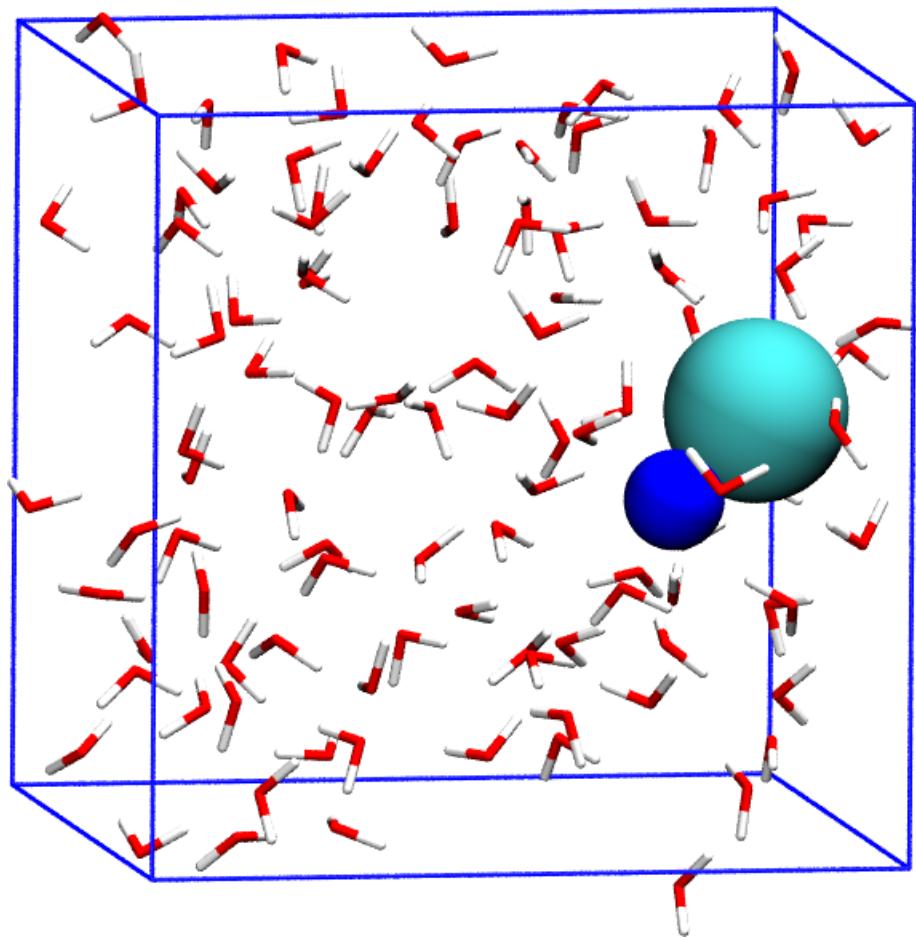
DOI: <https://doi.org/10.1103/PhysRevLett.113.090601>

DOI: <https://doi.org/10.1146/annurev-physchem-040215-112229>

DOI: [10.1007/978-3-319-44677-6_50](https://doi.org/10.1007/978-3-319-44677-6_50)

arXiv:2202.04164 - <https://arxiv.org/abs/2202.04164>

Tutorial: <https://www.plumed.org/doc-master/user-doc/html/masterclass-22-11.html>



Thanks for your attention!
Questions?

Recap of the Main Idea Behind VES

Variational principle to enhanced sampling

based on a functional $\Omega[V]$

$$\Omega[V] = \frac{1}{\beta} \log \frac{\int d\mathbf{s} e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s} e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s} p(\mathbf{s})V(\mathbf{s})$$

which is minimized by

$$V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta} \log p(\mathbf{s}) + C$$

can enhance the sampling and obtain the FES $F(\mathbf{s})$ by
finding the $V(\mathbf{s})$ that minimizes $\Omega[V]$

the resulting CV distribution determined by $p(\mathbf{s})$, which we can choose freely

convex functional \rightarrow easy optimization problem