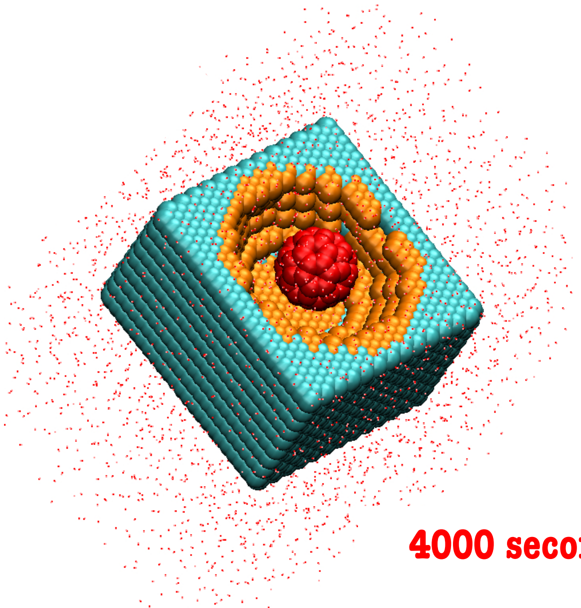
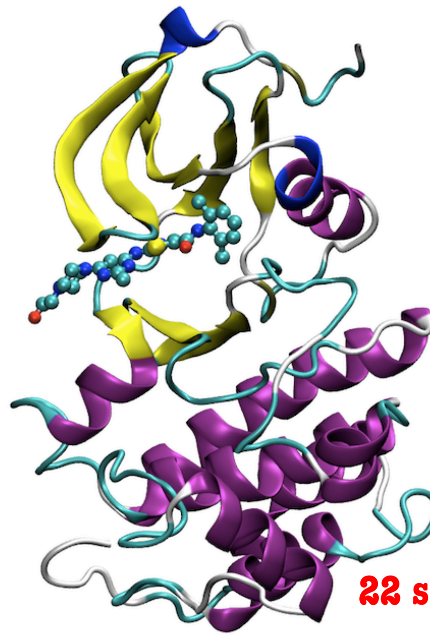
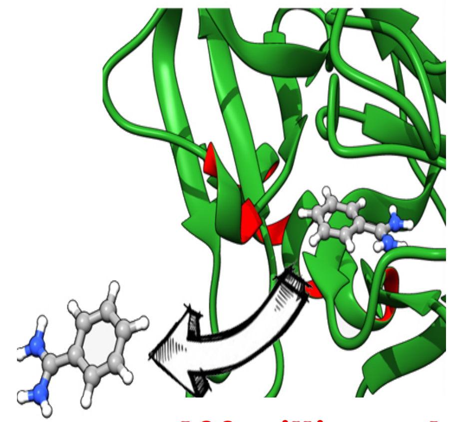
Identifying and enhancing important fluctuations for sampling molecular systems with rare events



4000 seconds



22 seconds



100 milliseconds

Pratyush Tiwary

Department of Chemistry, Columbia University, New York

Valsson, Tiwary & Parrinello ***Ann. Rev. Phys. Chem. 2016***
Tiwary & Berne ***Proc. Natl. Acad. Sci.*** **2016;** ***J. Chem. Phys.*** **2016**

Outline

- Drug unbinding kinetics – a grand challenge for atomistic simulations and enhanced sampling

- Key issues in enhanced sampling:
  - Which collective variables to bias ?
  - How to get unbiased kinetics ?

- Spectral gap optimization of order parameters (SGOOP)

- Applications to fully atomistic simulations of unbinding in explicit water:
  - hydrophobic buckyball and cavity
  - FDA-approved anti-cancer drug and Src kinase
  - BIRB analog and p38 kinase*

- Summary and outlook

Acknowledgments

Bruce Berne (Columbia)
Michele Parrinello (ETH)
Richard Friesner (Columbia)
Jagannath Mondal (TIFR)
Jim Pfaendtner (UW)

Joe Morrone (IBM)
Matteo Salvalaglio (UCL)
Rodrigo Casasnovas (Julich)
Paolo Carloni (Julich)
Axel van de Walle (Caltech/Brown)

SCHRÖDINGER.

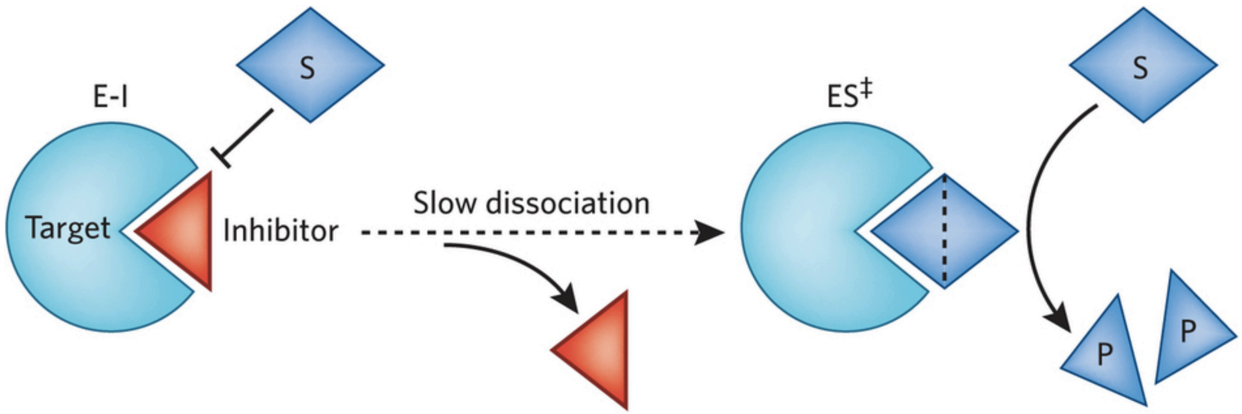
XSEDE
Extreme Science and Engineering
Discovery Environment

NIH

BILL & MELINDA GATES foundation

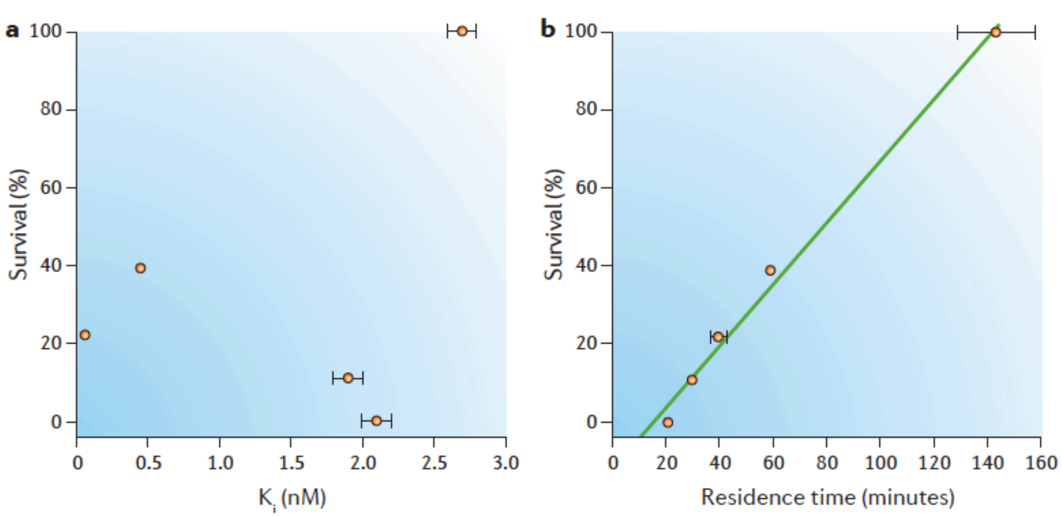
PLUMED

Why care for drug unbinding kinetics



k_{off} = 1/residence time

$K_i = k_{off} / k_{on}$

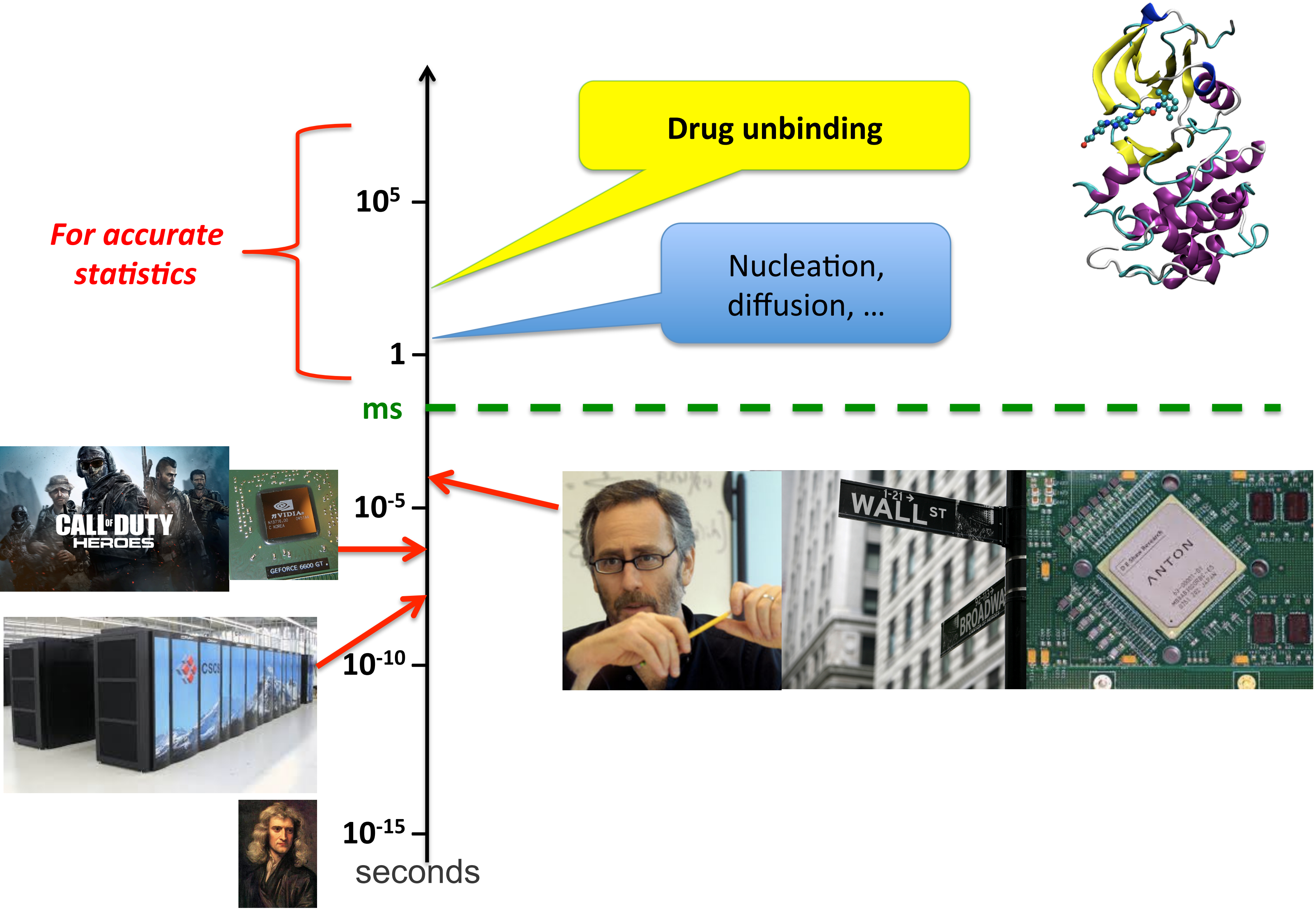


Copeland,
Nature Chem Bio 2015

Atomistic simulations could rationalize molecular determinants of unbinding kinetics and assist tailor improved drugs

Pathways, roles of protein flexibility and water ***BUT…***

The grand challenge in MD: timescales beyond milliseconds

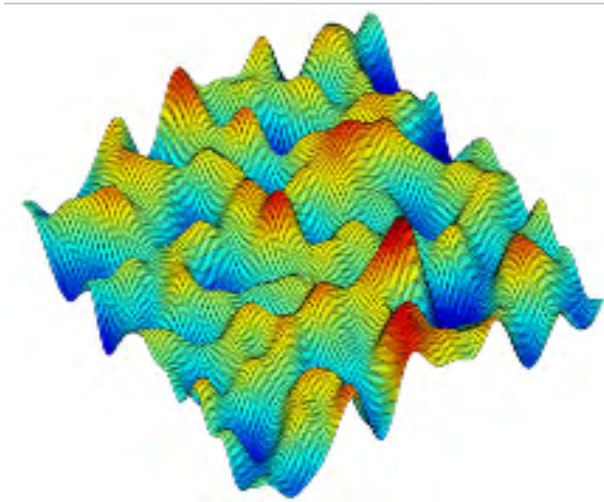


Fluctuations that matter: Collective Variables (CV)

Potential Energy Surface $U(\mathbf{R})$

➡ High-dimensional,

➡ Rugged



Introduce CVs that demarcate relevant stable states

- Need not be perfect RC

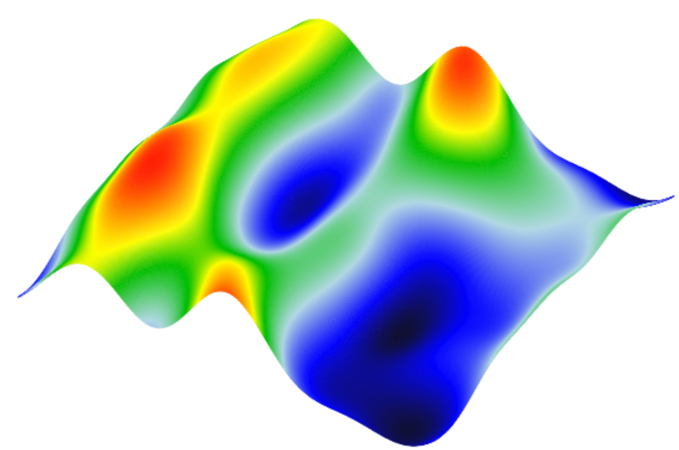
$$\mathbf{s}(\mathbf{R}) = (s_1(\mathbf{R}), s_2(\mathbf{R}), \ldots, s_d(\mathbf{R}))$$

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

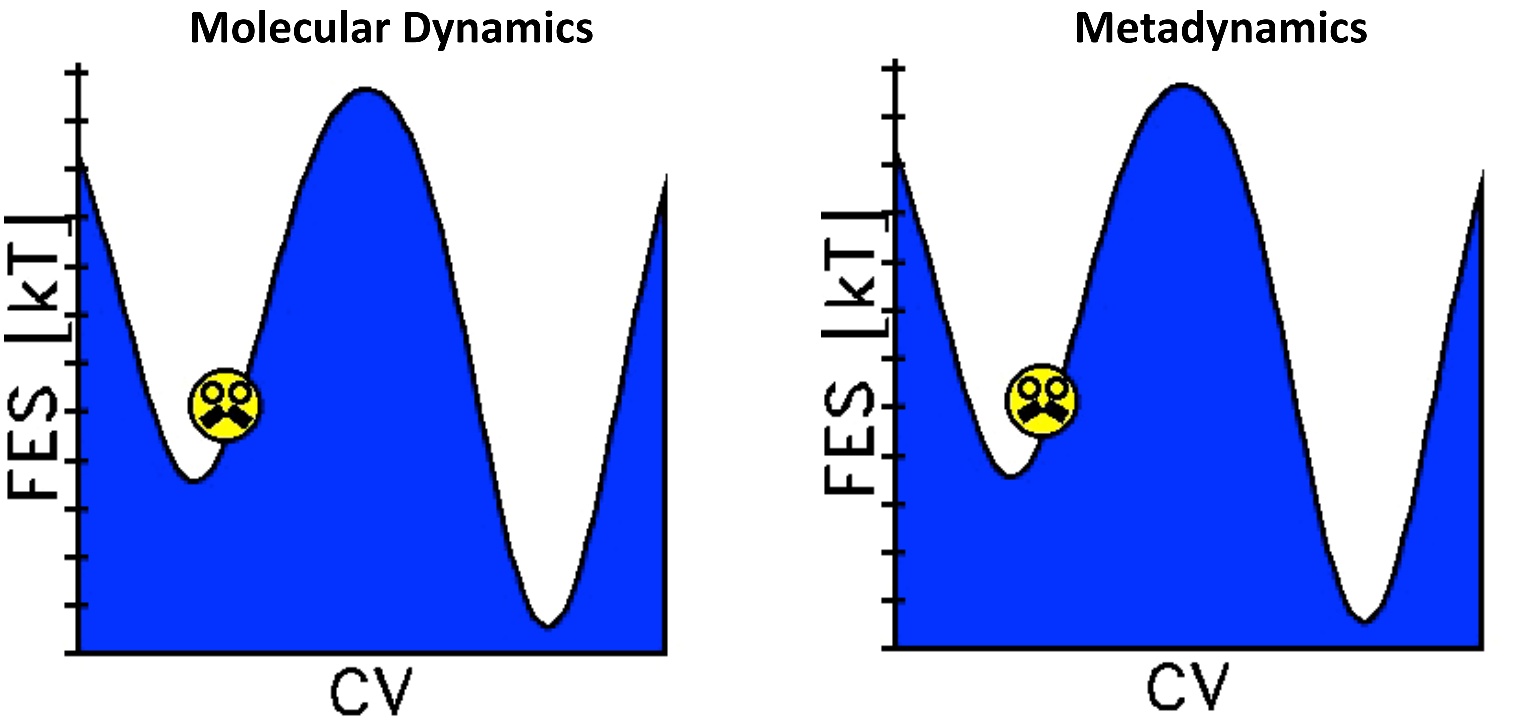
Free Energy Surface $F(\mathbf{s})$

➡ Low-dimensional,

➡ Smooth



Time-dependent enhancement of important fluctuations: Metadynamics



Add repulsive gaussian where you go, as function of chosen collective variable (CV)

Recover all sorts of thermodynamic averages as function of ***any*** CVs

Movie credit: G Bussi

Valsson, Tiwary, Parrinello, ARPC 2016

Two big problems with enhanced sampling

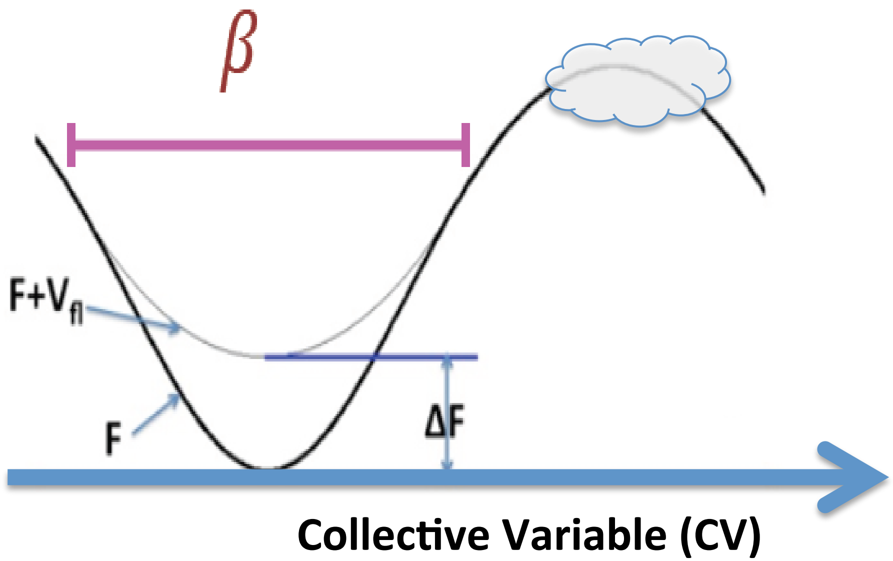
1. Need to presciently identify relevant low-dimensional collective variables

Solution : "Spectral gap optimization of order parameters (SGOOP)"
Tiwary and Berne, Proc. Natl. Acad. Sci. 2016

2. True dynamical information lost during the course of biasing

Solution : "Infrequent metadynamics"
Tiwary and Parrinello, Phys. Rev. Lett. 2013

From metadynamics to dynamics



$$k_{\beta\to\alpha} \propto \kappa \frac{Z_{TS}}{Z_{\beta}}$$ without bias

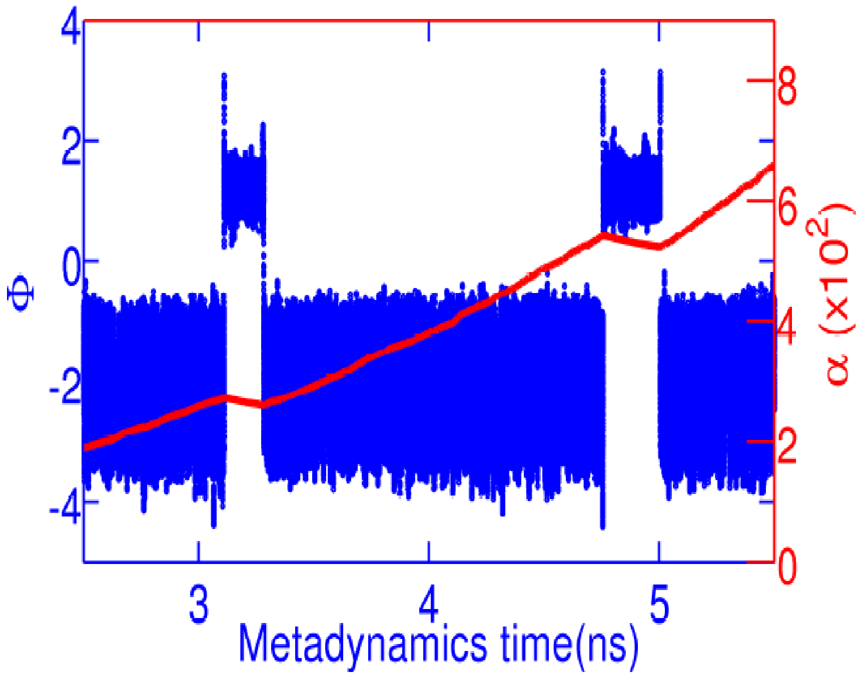
$$k^{*}_{\beta\to\alpha} \propto \kappa \frac{Z_{TS}}{Z^{*}_{\beta}}$$ with bias

Acceleration factor of rates

$$\frac{k^{*}_{\beta\to\alpha}}{k_{\beta\to\alpha}} = \frac{Z_{\beta}}{Z^{*}_{\beta}} = \left\langle e^{\beta V_{fl}} \right\rangle^{*}$$

Grubmueller PRE 1995
Voter, JCP 1997 + PRL 1997
Tiwary & Parrinello, PRL 2013

Doing away with TS knowledge : Infrequent Metadynamics

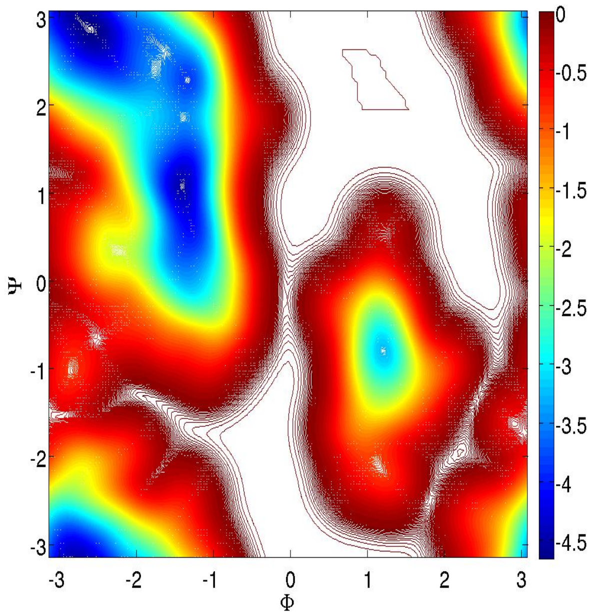


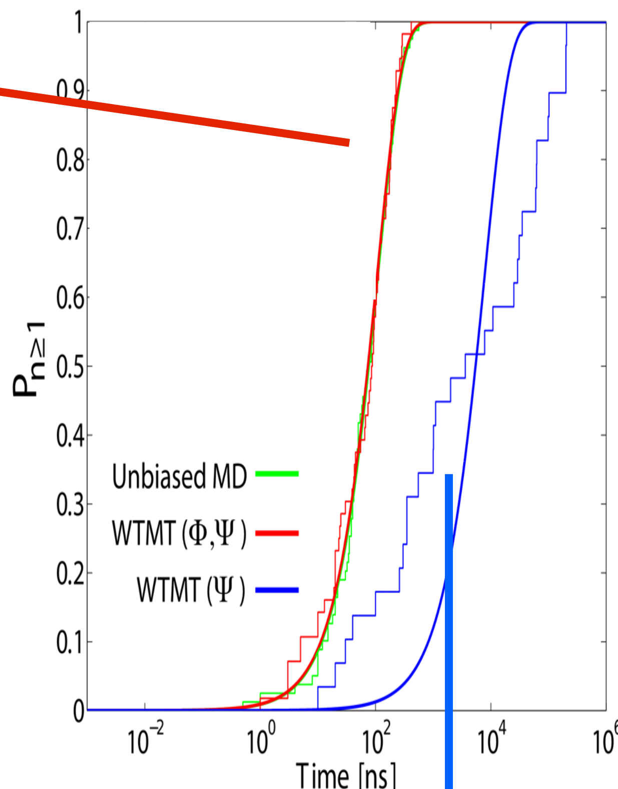
- Transitions are rare-but-fast: make bias deposition slower than time in bottleneck
- Can identify *a posteriori* if requirements met – **TS corruption will introduce memory into dynamics**
- Other approaches can complement this: Mccarty, Valsson, Tiwary, Parrinello PRL 2015

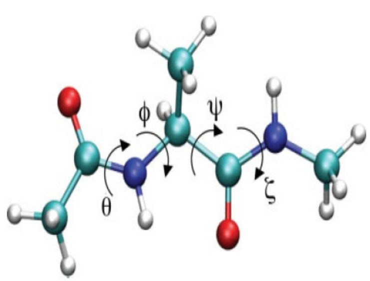
Tiwary & Parrinello, PRL 2013

Check memoryless-ness through Poisson behavior *a posteriori*

small KS statistic
large *p-value* (~0.9)
uncorrupted dynamics







large KS statistic
small *p-value* (~1E-4)
corrupted barriers and dynamics

Salvalaglio, Tiwary and Parrinello, J. Chem. Theor. & Comp (2014)

Problems with enhanced sampling: #1

Need to presciently identify
relevant low-dimensional collective variables
(possibly from a larger dictionary of choices)

Solution :
"Spectral gap optimization of order parameters (SGOOP)"
Tiwary and Berne, Proc. Natl. Acad. Sci. 2016

Is it even possible to solve this problem?

Statistical mechanics does not tell us what the relevant variables are. This is our choice. If we choose well, the results may be useful; if we choose badly, the results (while still formally correct) will probably be useless.

- Robert Zwanzig

Nonequilibrium Statistical Mechanics

(Ch. 8, "Projection Operators")

Quantifying the importance of low-dimensional CV through spectral gap of unbiased dynamics

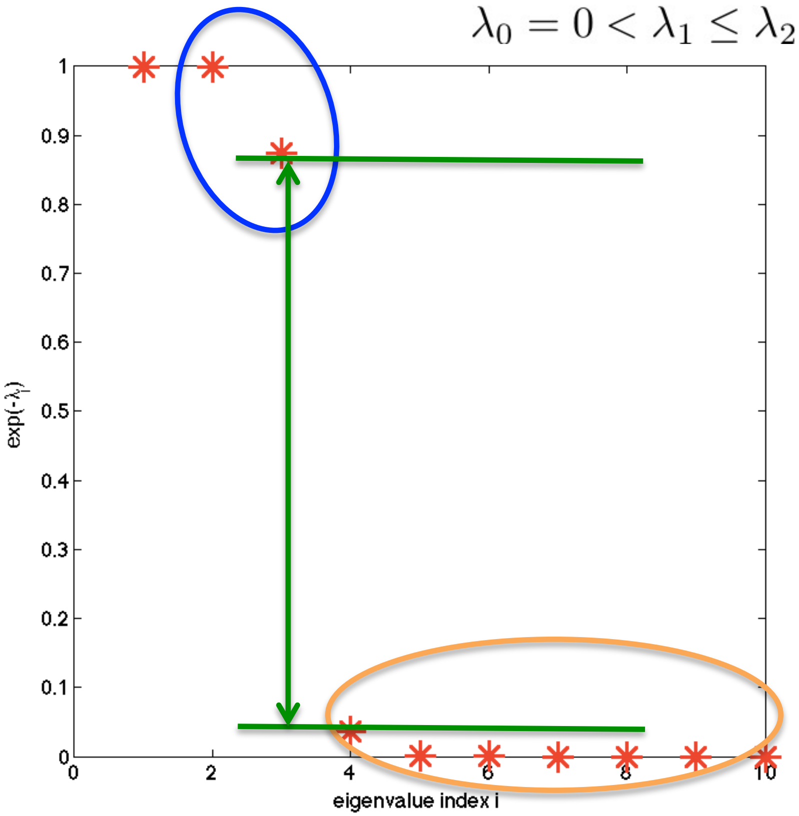
$$\partial_t p(\vec{r}, t) = \mathcal{L}(\vec{r})\, p(\vec{r}, t)$$

← How (unbiased) probability of being in CV region *(r,r+dr)* evolves in space and time

$$p(r,t) = \phi_0(r) + \sum_{n \neq 0} c_n \phi_n(r) e^{-\lambda_n t}$$

$$\lambda_0 = 0 < \lambda_1 \leq \lambda_2$$

← Solution



Motivation :
When projected on low dimensional CVs, the best CV will show best **timescale separation** into visible **visible slow** and **orthogonal fast** processes – highest **spectral gap**

Challenge :
Given **limited stationary and dynamical knowledge** in rare event molecular systems, how to quantify spectral gap for various CVs?

Maximum entropy over paths and Caliber

ET Jaynes, ARPC 1980
Presse, Ghosh, Lee, Dill RMP 2013

Entropy over states = $-\Sigma_a \; p_a \log \, p_a$

Entropy over paths = $-\Sigma_\tau \; p_\tau \log \, p_\tau \qquad = -\Sigma_{a,b} \; p_a \; k_{a,b} \; \log \; k_{a,b}$

Caliber $\quad = -\Sigma_{a,b} \; p_a \; k_{a,b} \; \log \; k_{a,b} + \Sigma \rho_i * \left[\langle A^i \rangle - A^i_{obs} \right]$

Path entropy *Constraints*

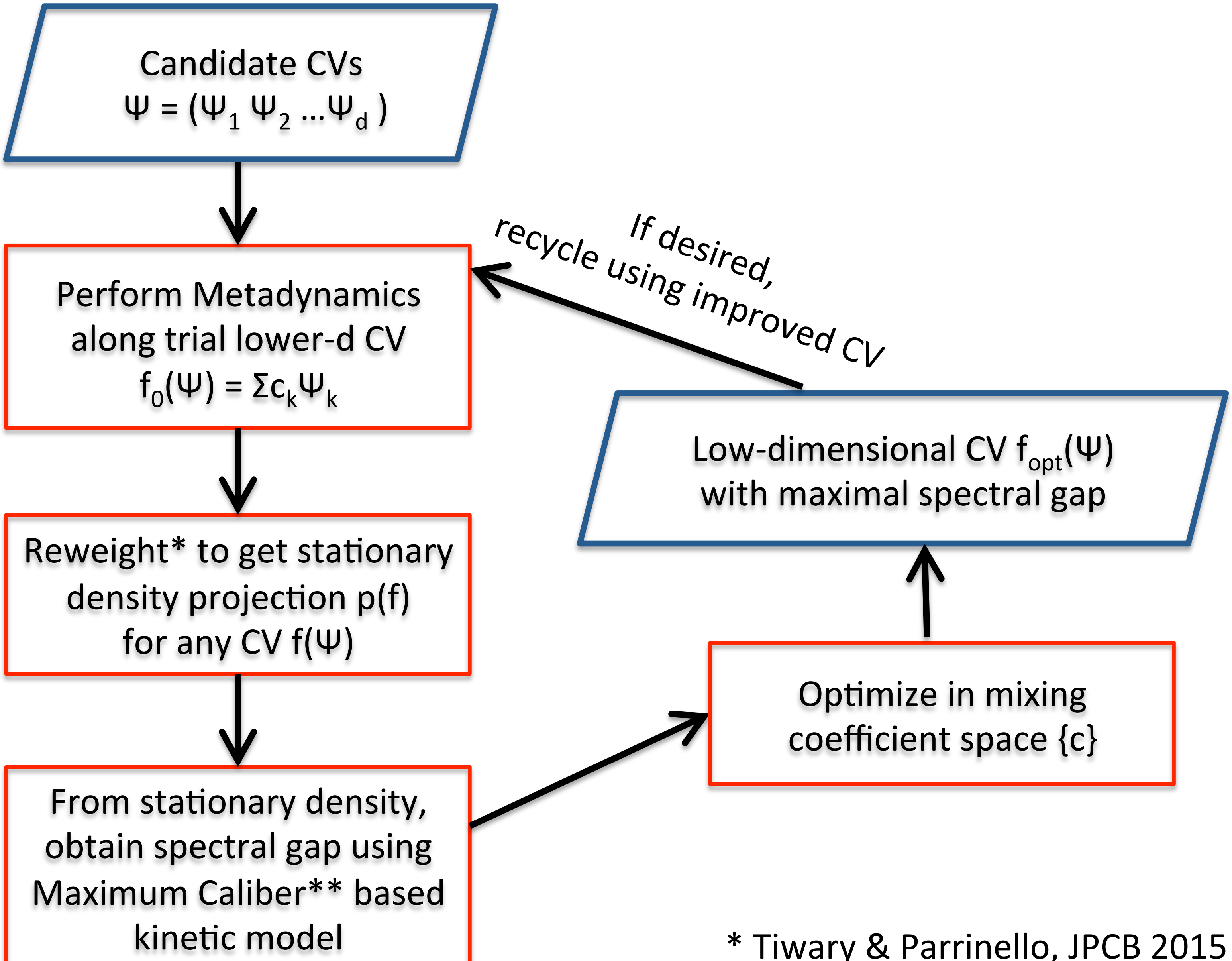
Maximum caliber relation between transfer matrix and stationary probabilities

$$k_{a,b} = \sqrt{\frac{p_b}{p_a}} e^{-\Sigma \rho_i A^i_{a,b}}$$

Also Bicout & Szabo JCP 1998
Hummer NJP 2005

Spectral gap optimization of order parameters: SGOOP

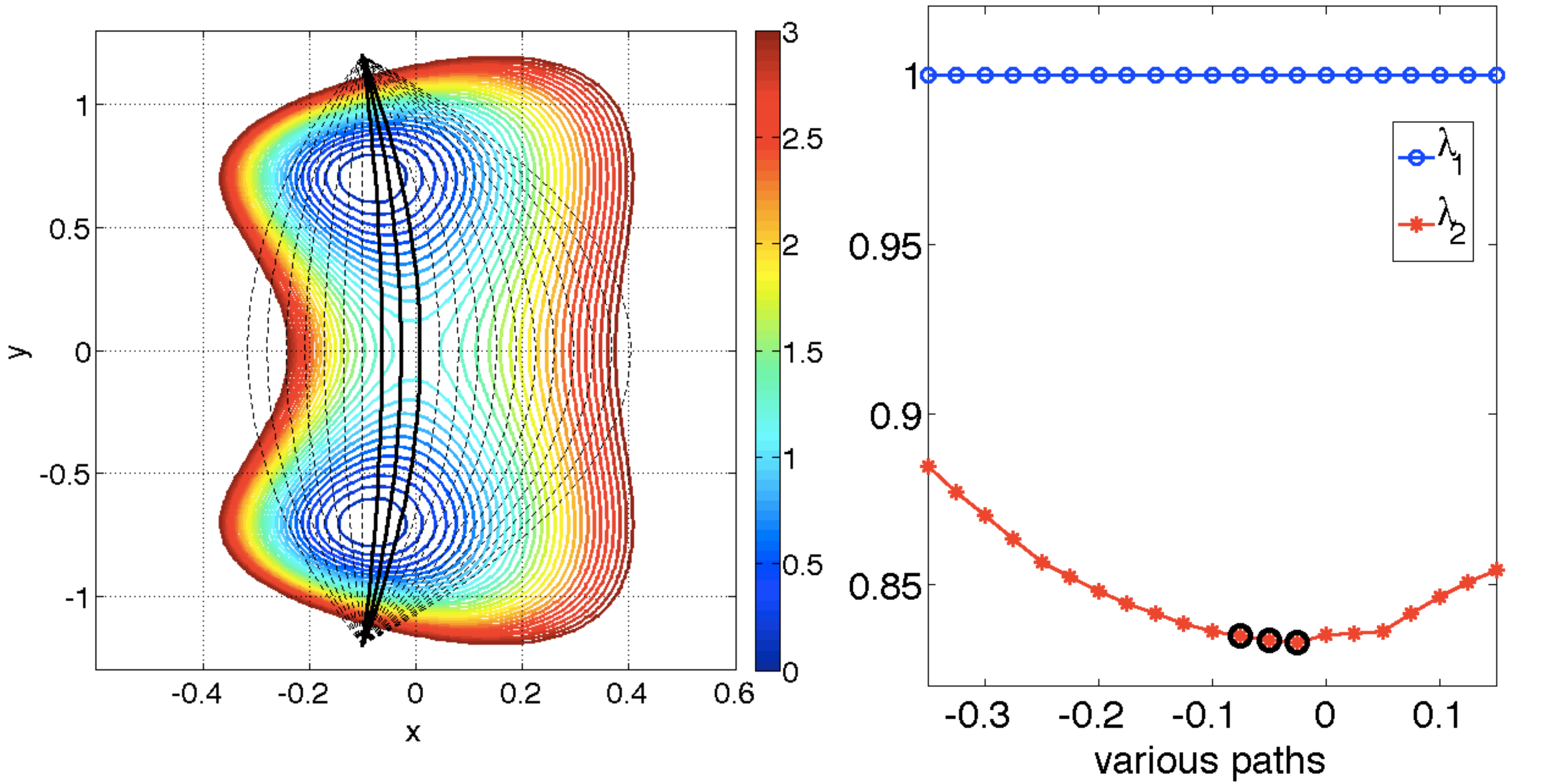
Tiwary and Berne, PNAS 2016



$$k_{a,b} = \sqrt{\frac{p_b}{p_a}} e^{-\rho}$$

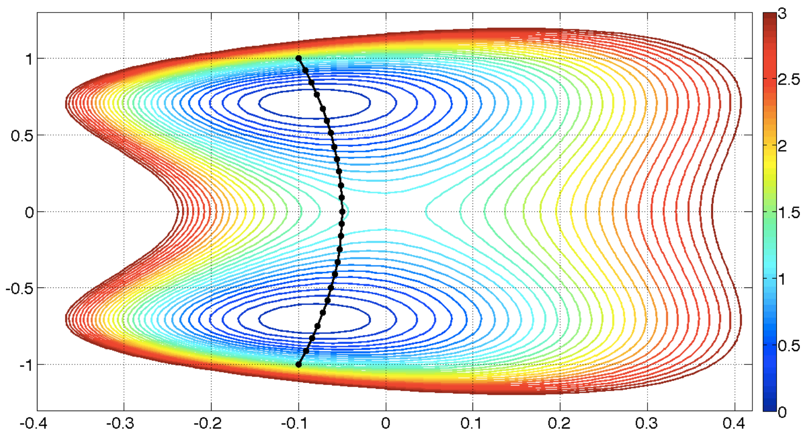
* Tiwary & Parrinello, JPCB 2015

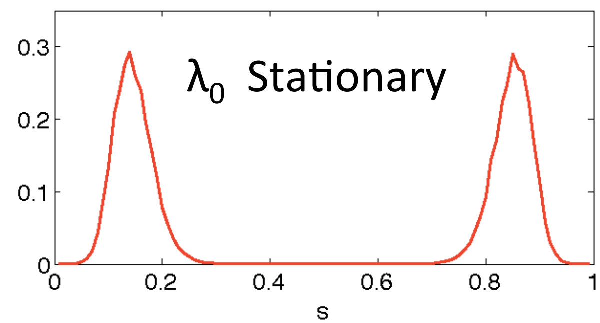
2-d model potential with path CVs
Largest spectral gaps = minimum energy paths

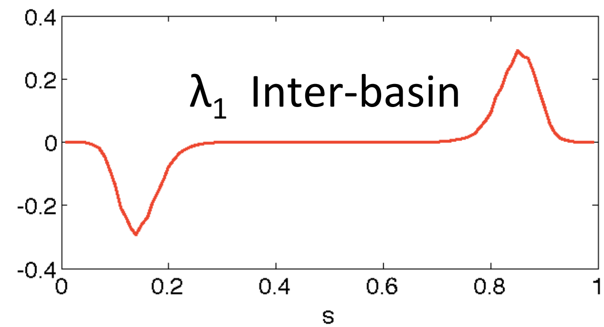


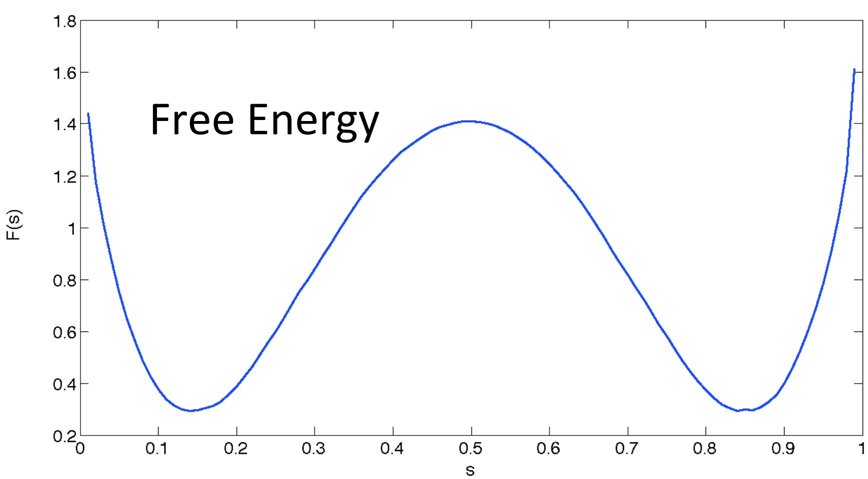
Model potential: De Leon, Berne JCP 1981
Path CVs: Branduardi, Gervasio, Parrinello JCP 2007

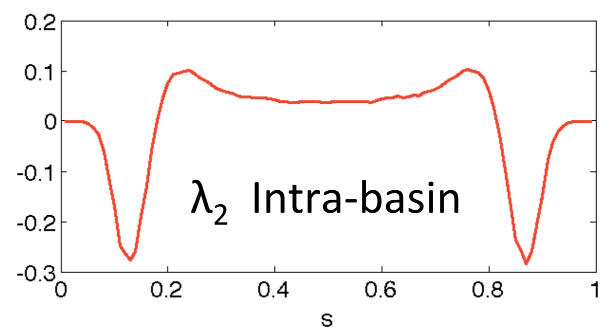
Inter-basin and intra-basin eigenvectors



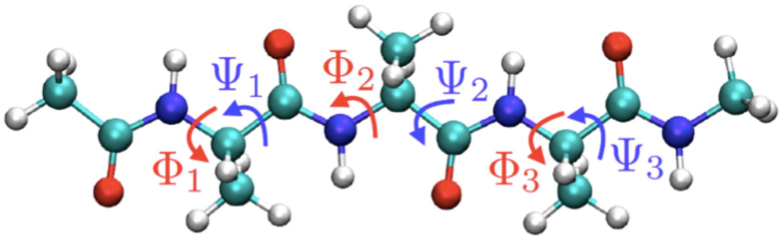




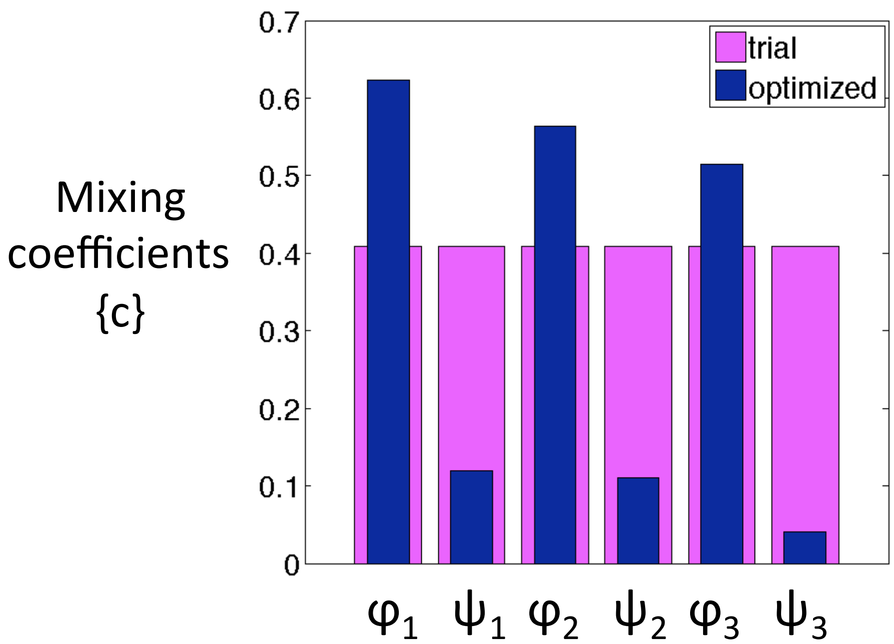


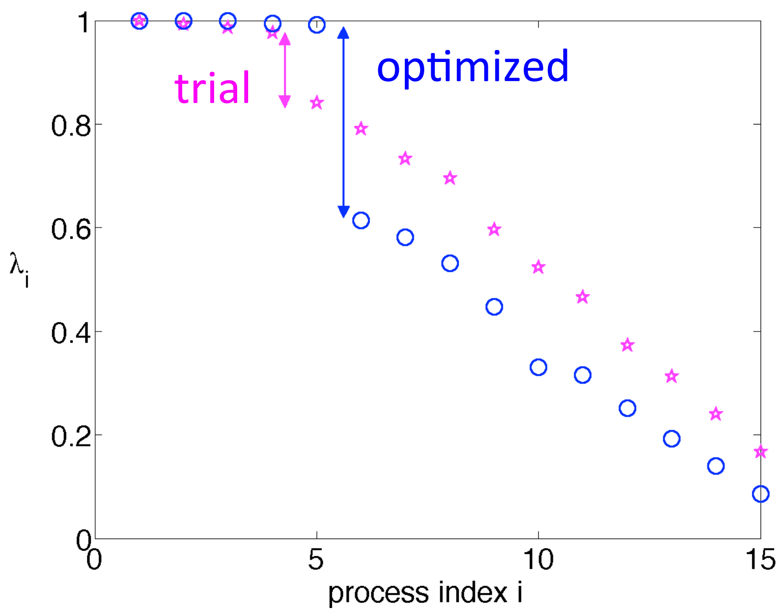


SGOOP in action for free energies: Ace-Ala$_3$-Nme

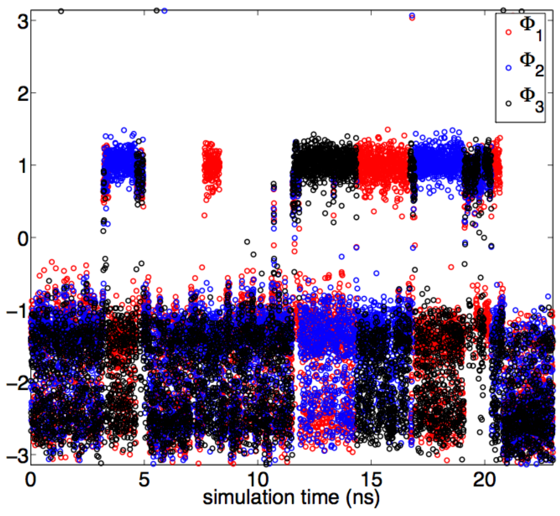


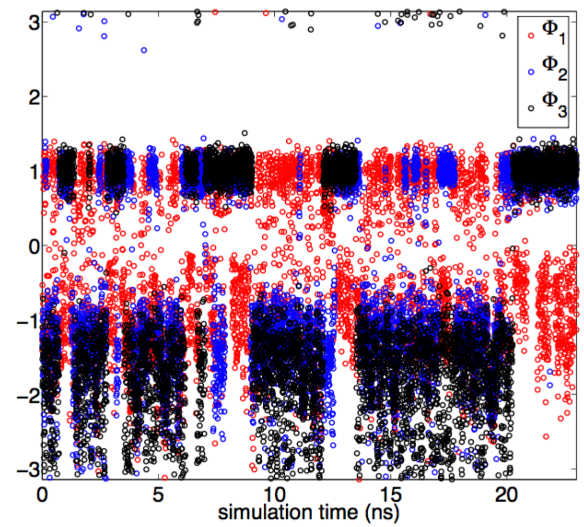
- **Objective:** find 1-d CV biasing which will maximize 6-d exploration





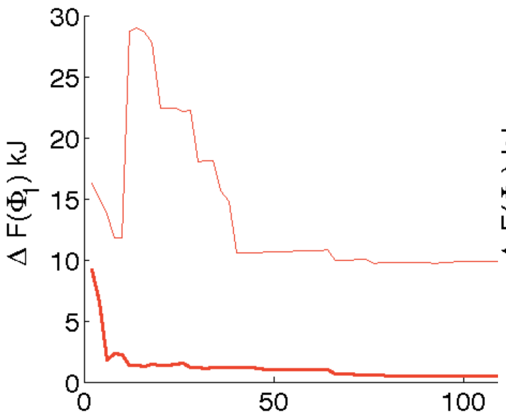
Several orders of magnitude improvement in FES convergence speed

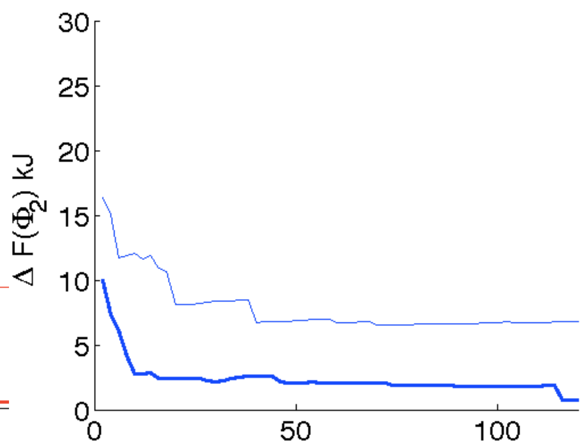


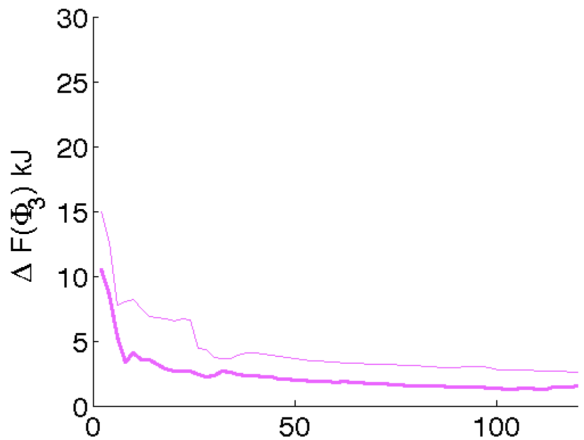


Trajectories biasing trial CV

Trajectories biasing optimized CV

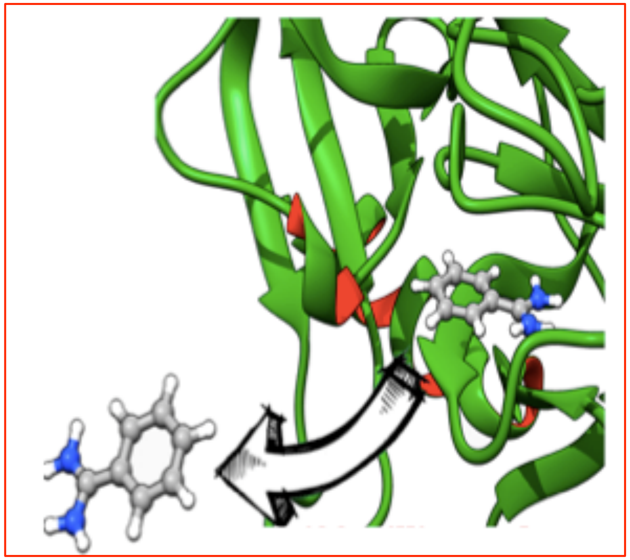




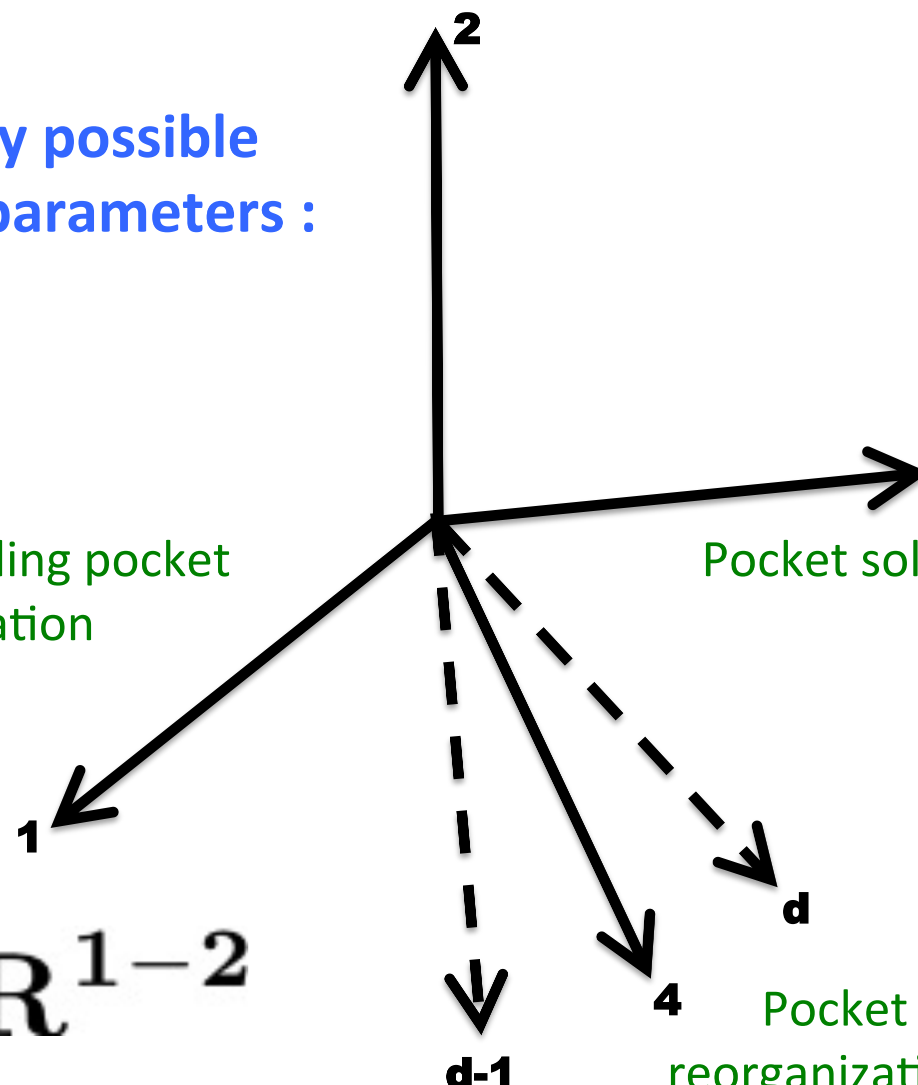


Error relative to reference FES for trial CV (thin line) and optimized CV (thick line)

SGOOP for ligand unbinding



Many possible order parameters :

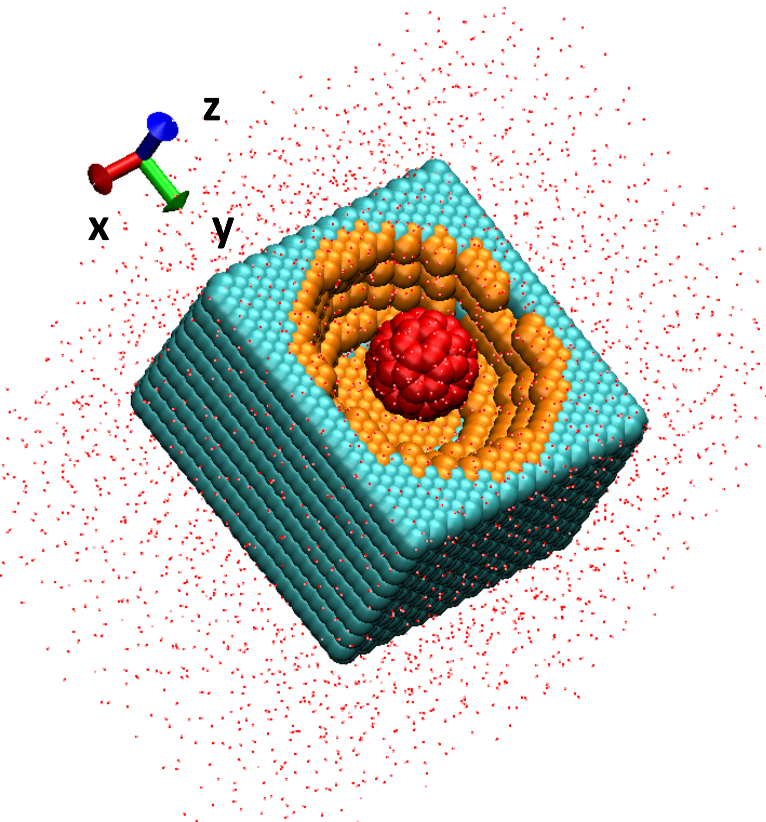


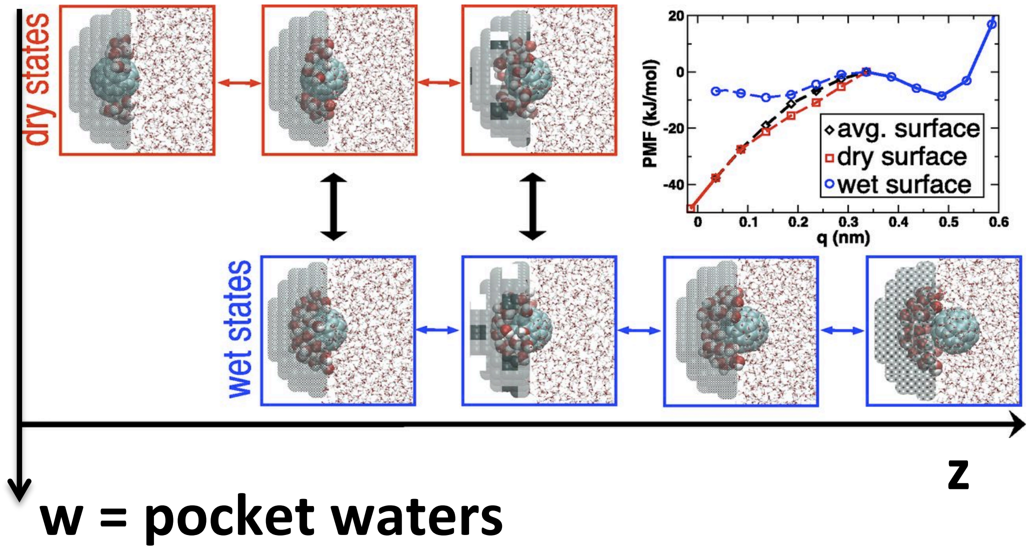
$$\mathrm{R}^{3\mathrm{N}} \mapsto \mathrm{R}^{\mathrm{d}} \mapsto \mathrm{R}^{1-2}$$

Intuition / non-equilibrium MD

SGOOP

Buckyball-cavity unbinding
in explicit TIP4P water





Objective: find optimal 1-d CV of the form

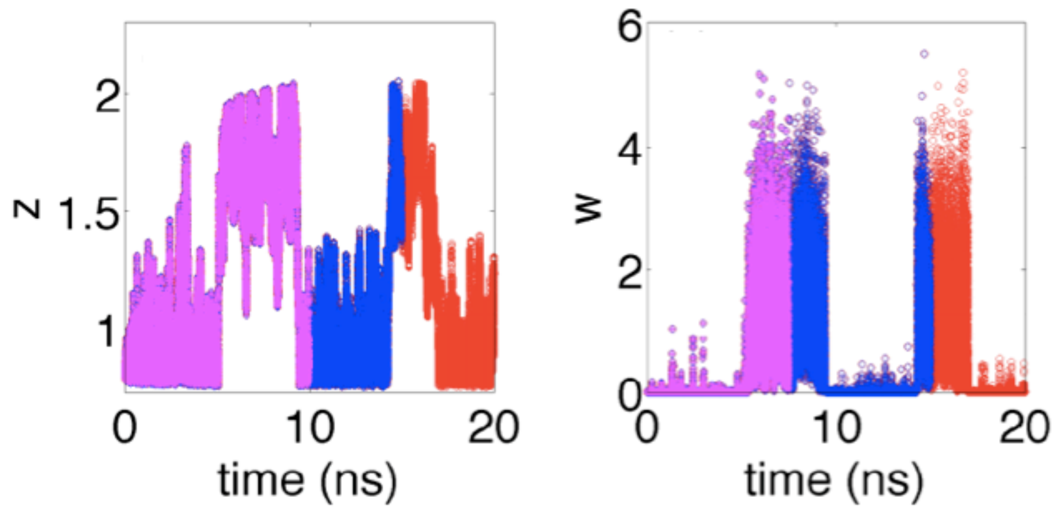
$$\Psi(z, w) = \{z + m_w w;\ m_w \geq 0\}$$

m_w quantifies the "wetness" of CV

Mondal, Morrone, Berne PNAS 2013
Tiwary, Mondal, Morrone, Berne PNAS 2015
Tiwary and Berne JCP 2016

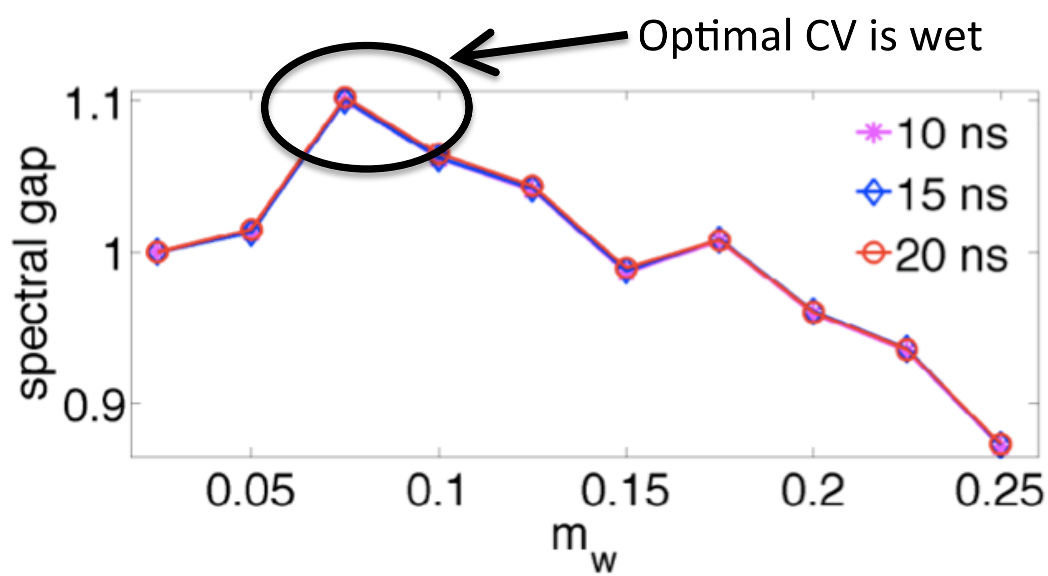
SGOOP: How much do cavity water fluctuations really matter for driving unbinding?

Short metadynamics with trial CV = z



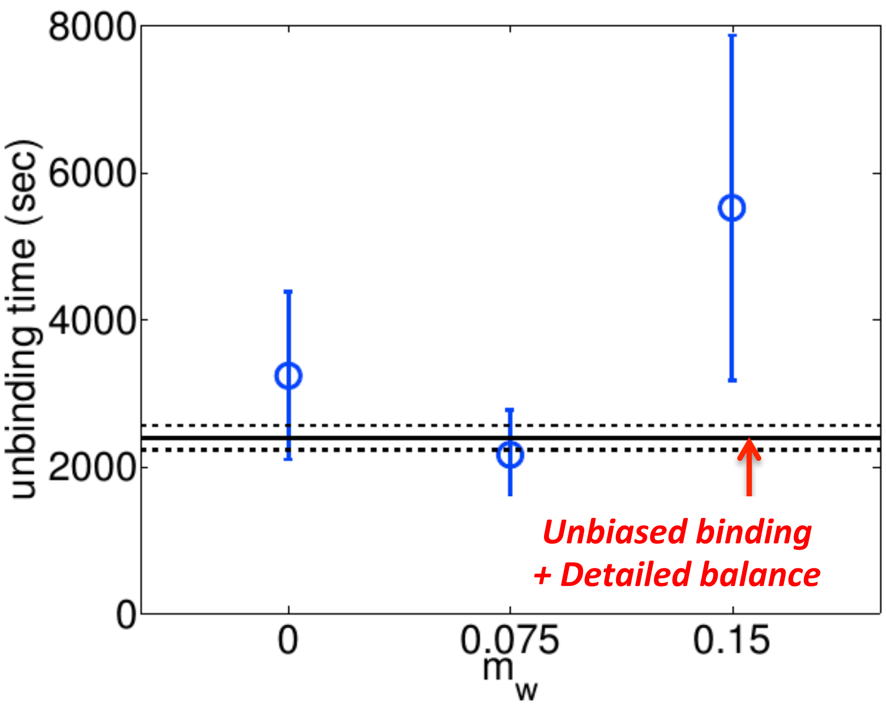
Robust spectral gap estimates for different CVs

$$\Psi(z, w) = \{z + m_w w;\ m_w \geq 0\}$$



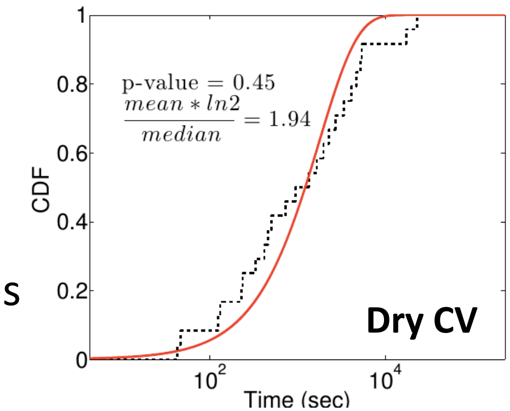
Tiwary and Berne JCP 2016

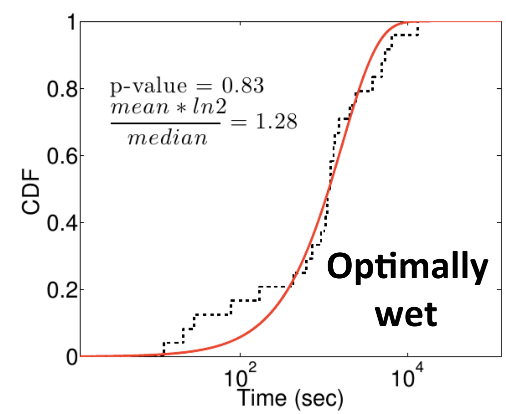
Optimally wet CV gives the most accurate unbinding time through infrequent metadynamics

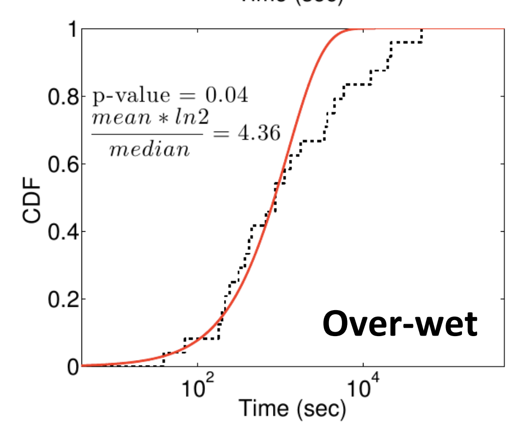


Infrequent metadynamics on optimally wet CV gives best agreement with detailed balance and unbiased binding time estimate

Infrequent metadynamics on optimally wet CV gives best Poisson fits:

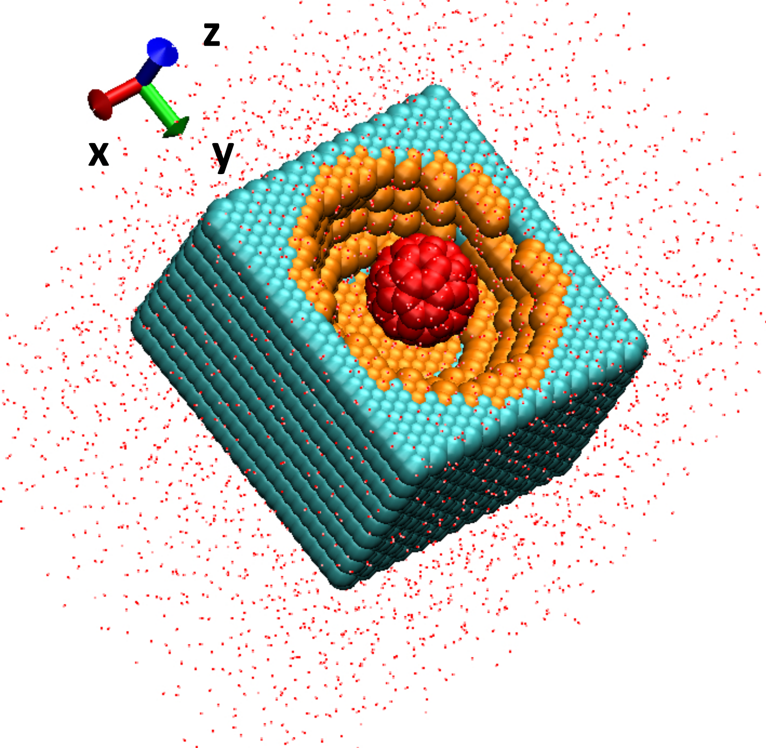




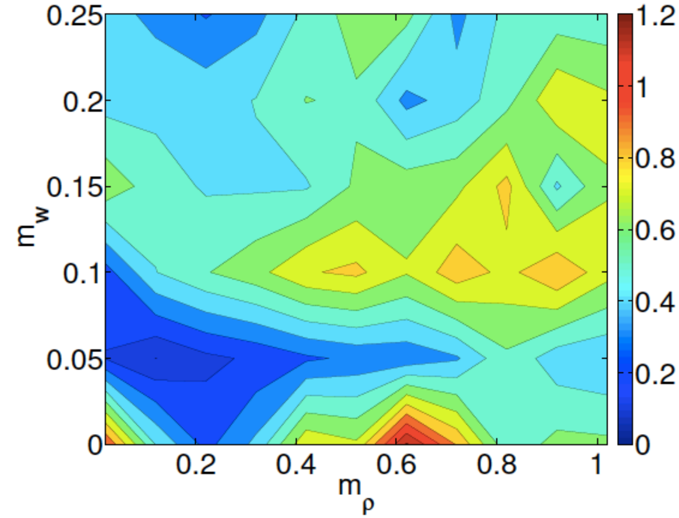


What if buckyball has no steric constraints?

$$\rho = \sqrt{x^2 + y^2}$$



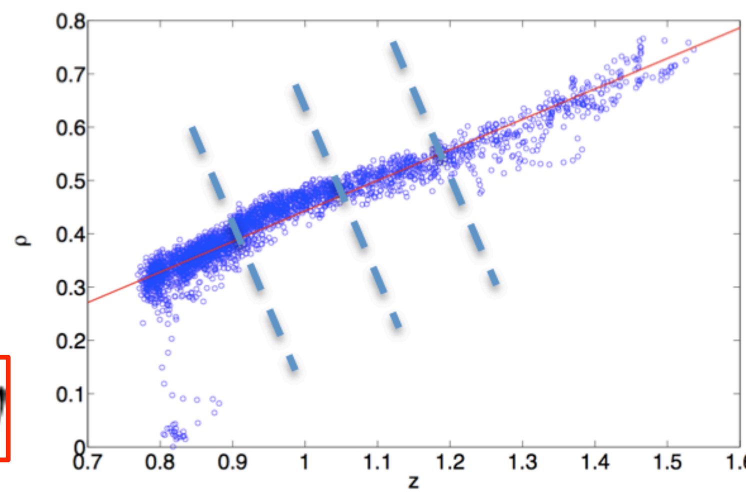
Contour map
of spectral gaps



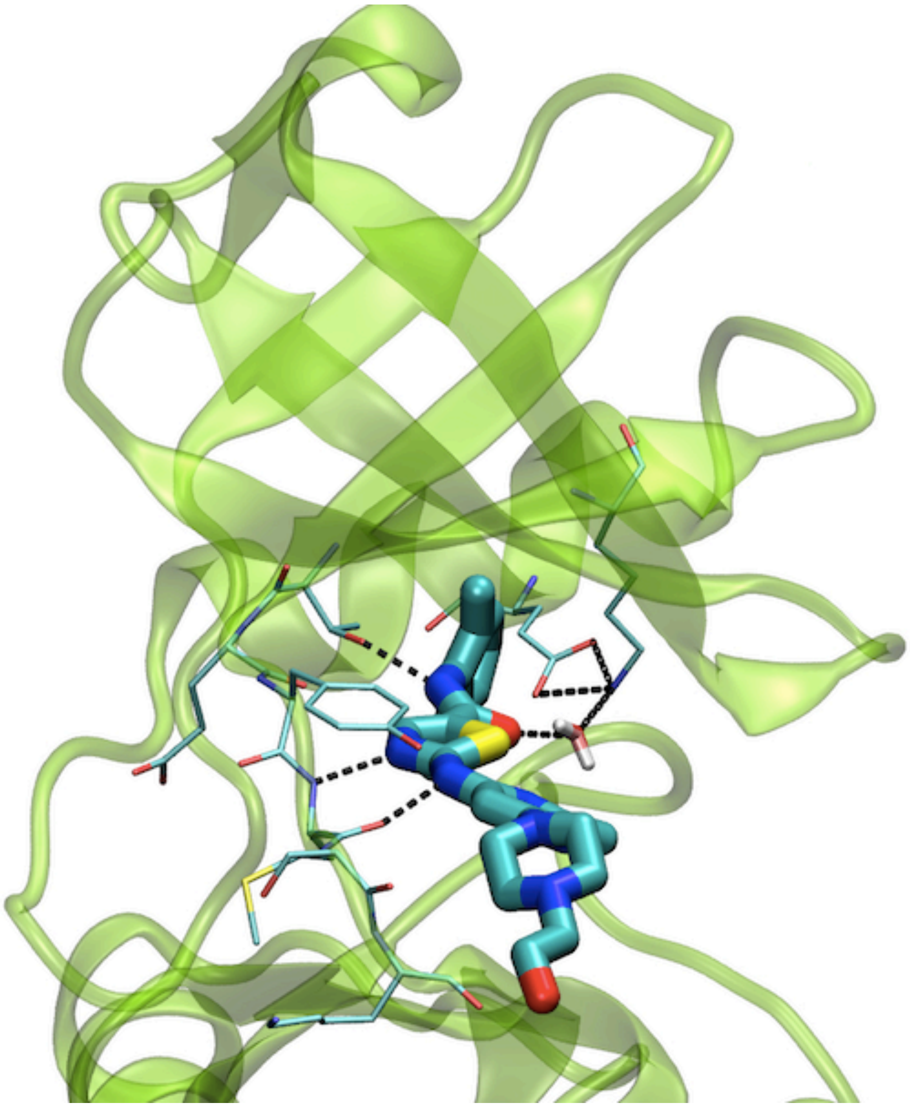
Water density fluctuations now become a driven rather than driving variable -
Buckyball rolls along the sides, water follows

Objective: find optimal 1-d CV of the form

$$\boxed{\Psi_{ideal} \approx z + 0.6\rho + 0.0w}$$



Anti-cancer drug Dasatinib (Sprycel®, Bristol-Myers Squibb) from Src kinase



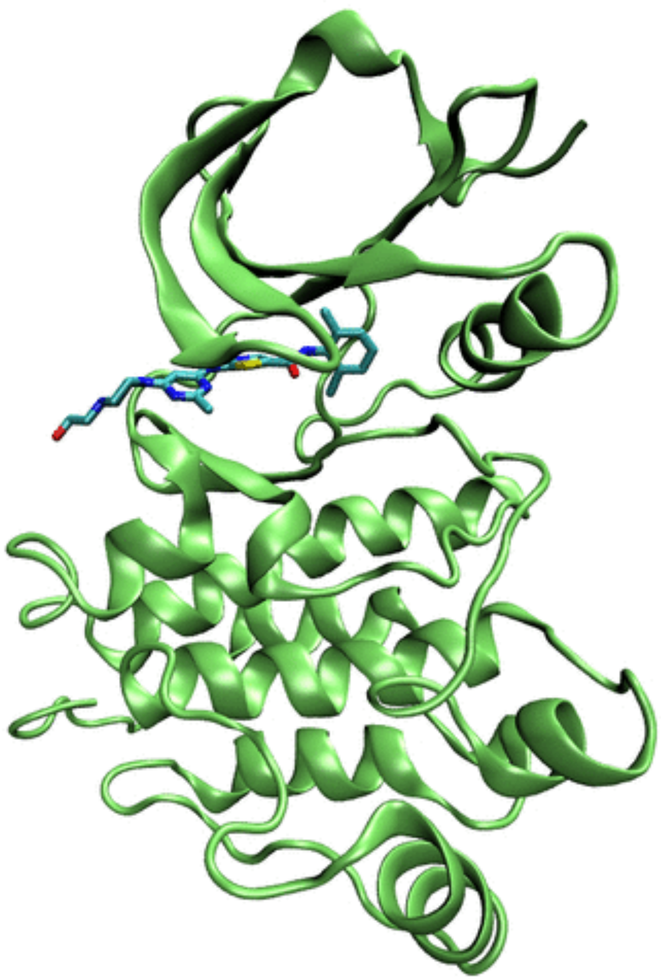
- ATP = energy currency
  Kinase = ATM card

- Previous works[1,2] studying binding

- 2 CVs:
  (1) distance between H-bond formers
  (2) solvation state of binding pocket

1. Shan, Shaw and co-workers, JACS 2011
2. Mondal, Friesner, Berne JCTC 2014
3. Shan, Kuriyan, Shaw et al PNAS 2009

Anti-cancer drug Dasatinib (Sprycel®, Bristol-Myers Squibb) from Src kinase



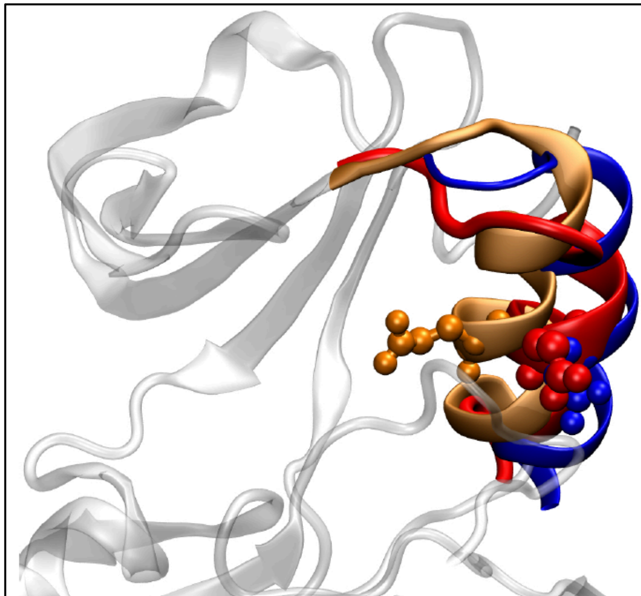
Using acceleration factor, our residence time estimate[1] = 20 +/-9 sec

Experiments[2] = 18-900 sec

Variety of other detailed, intriguing findings – stay tuned for our paper[1]

1. Tiwary, Mondal & Berne, under preparation
2. Shaw, Seeliger et al PNAS 2009; Seeliger, Soellner ACS Chem Bio 2016

Coupled water-protein fluctuations matter

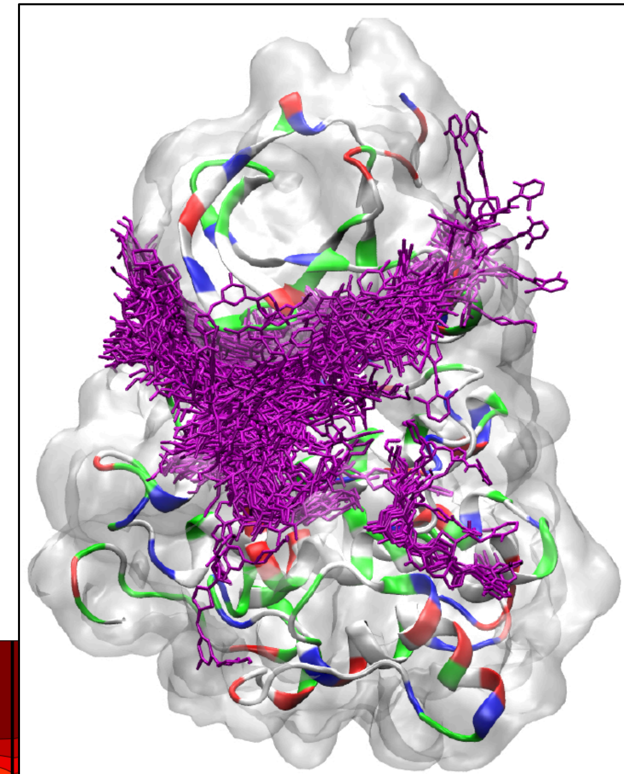


States #2-4

Orange = αC helix in

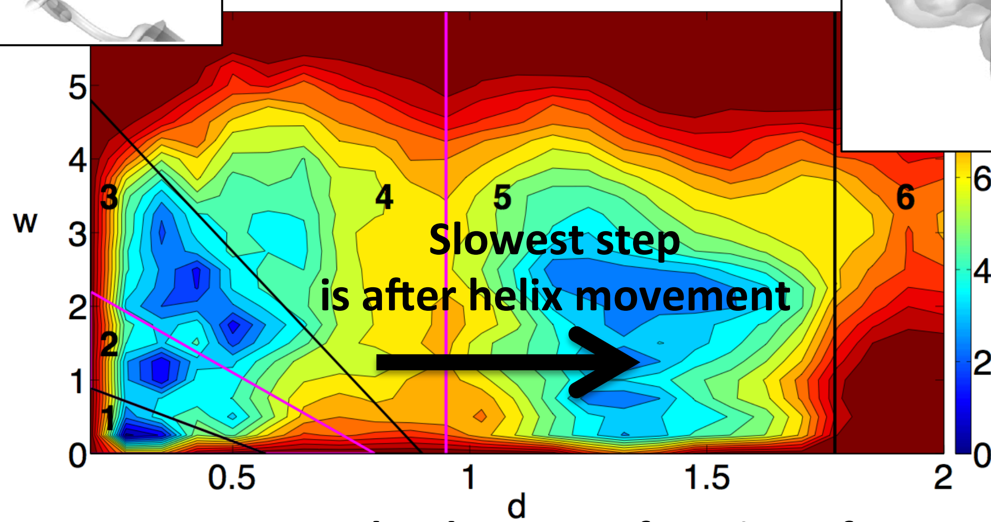
Red = αC helix out (this work)

Blue = αC helix out (expt.)



State #6

Surface and bulk diffusion



Free energy landscape as function of d (Ligand-protein separation) and w (pocket hydration)

Tiwary, Mondal & Berne, under preparation

Summary and outlook

- Recent progress in enhanced sampling with CVs allows obtaining dynamics and thermodynamics of rare-event (bio)molecular systems with statistical accuracy

- Successful applications to ligand unbinding kinetics – buckyball, trypsin-benzamidine, Src-Dasatinib, p38-BIRB

- Mixing order parameters into low-dimensional form is critical for methods such as metadynamics and umbrella sampling – SGOOP helps with quick and systematic refinement of CVs in hard to sample rare event systems

- Collective variables as tools to study molecular systems, that can be iteratively refined – **they are both input and output for enhanced sampling**