Tackling sampling and accuracy issues in biomolecular simulations

Max Bonomi
University of Cambridge
mb2006@cam.ac.uk

Outline

- Molecular Dynamics as a computational microscope
  - sampling problems
  - accuracy of force fields
- Enhanced sampling with biased MD
  - umbrella sampling
  - metadynamics
  - recent metadynamics developments
- The open source library PLUMED
- Combining simulations with experiments: Metainference
- Addressing sampling and accuracy issues: M&M
- Towards modelling of cry-electron microscopy data

Tackling sampling and accuracy issues in biomolecular simulations

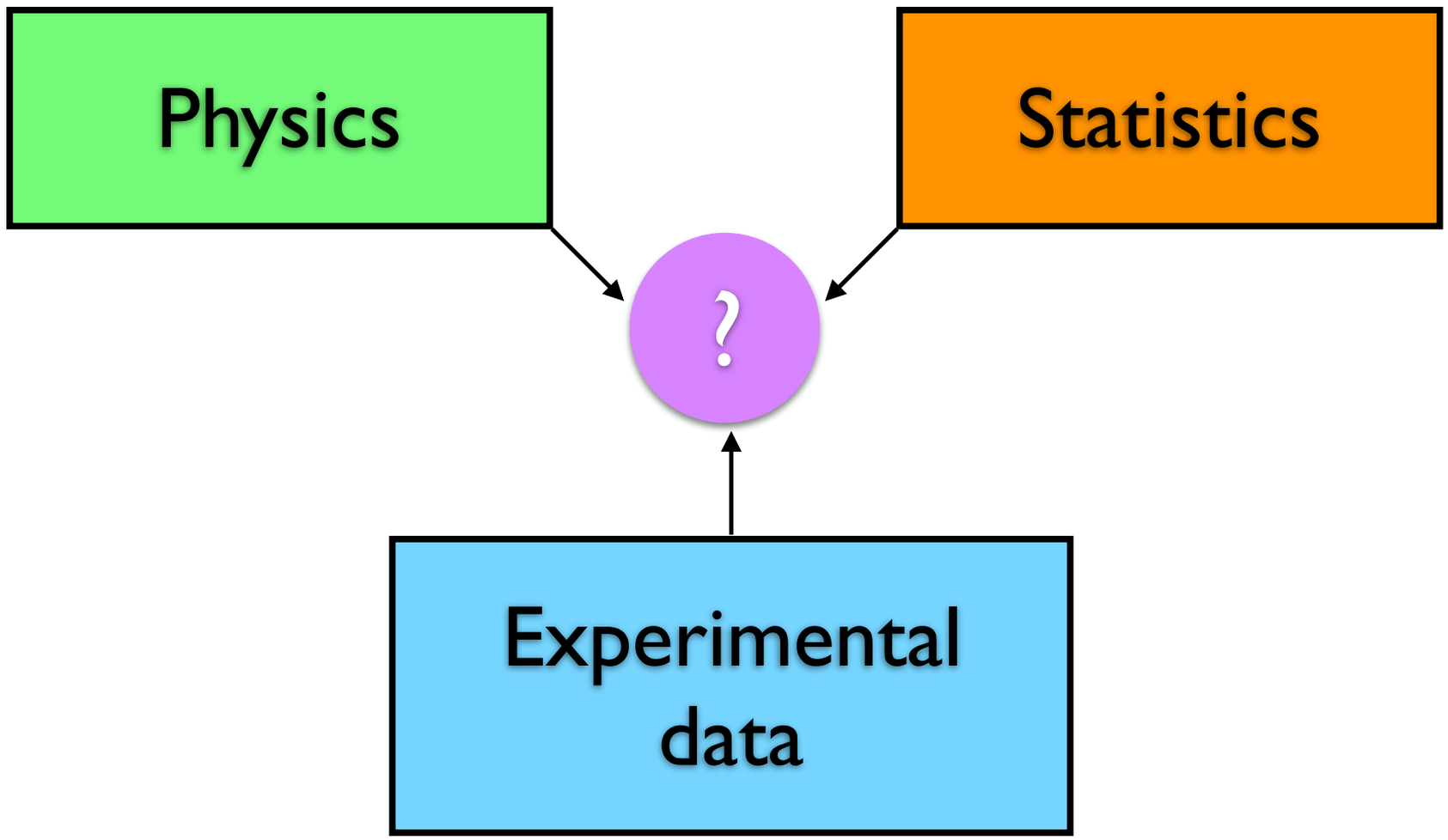
Max Bonomi
University of Cambridge
mb2006@cam.ac.uk

Outline

- Molecular Dynamics as a computational microscope
  - sampling problems
  - accuracy of force fields
- Enhanced sampling with biased MD
  - umbrella sampling
  - metadynamics
  - recent metadynamics developments
- The open source library PLUMED
- Combining simulations with experiments: Metainference
- Addressing sampling and accuracy issues: M&M
- Towards modelling of cry-electron microscopy data

Improving the accuracy of force fields

A more accurate description of a system can be achieved if we combine all the sources of information available

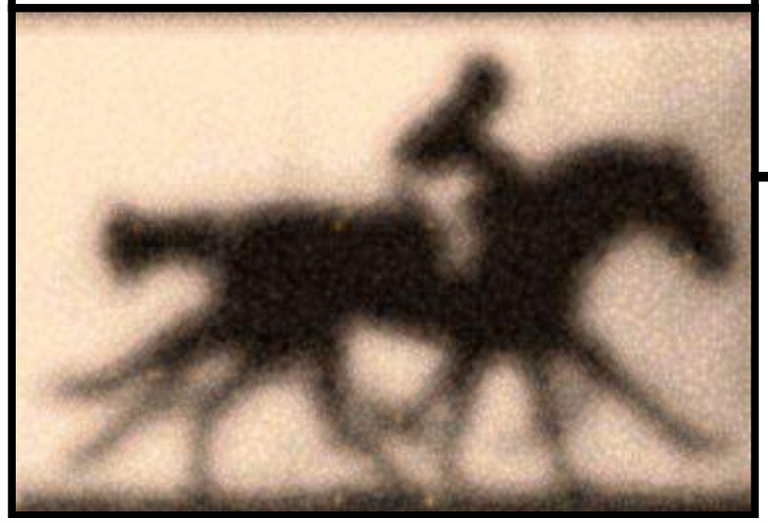


How can we properly combine them?

The challenges of data modelling

a) Input

Measurement+Prior

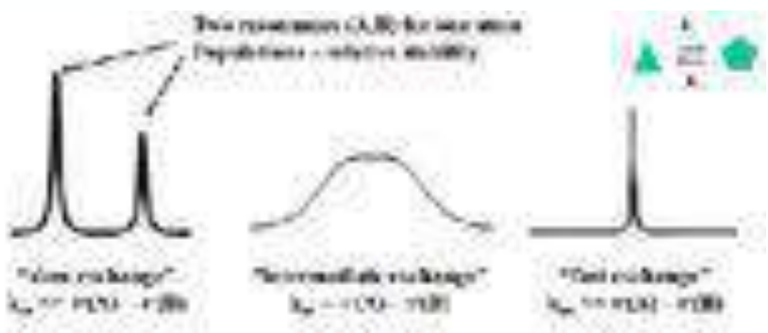


b) Errors

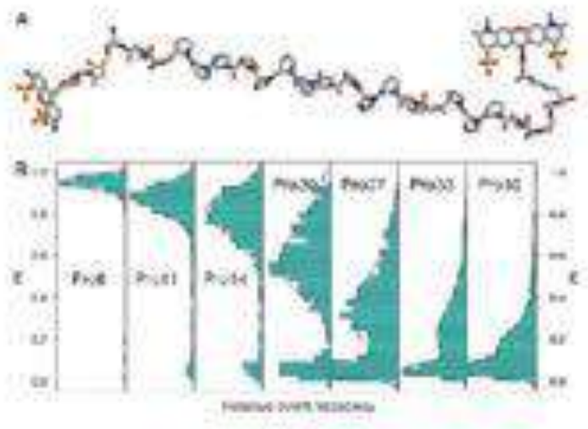
Mixing	
Random	
Systematic	
Forward model	

EXAMPLES

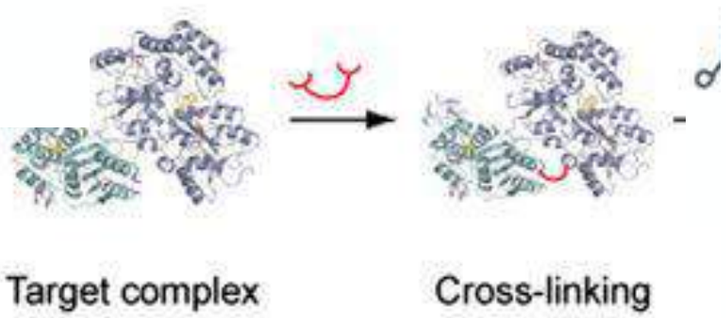
NMR



FRET distributions



XL/MS



The challenges of data modeling

Mixing		Replica-Averaged Modelling+ +Camilloni *et al.* JACS 2012
Random		Bayesian Modelling* *Rieping *et al.* Science 2005
Systematic		
Forward model		

Replica-Averaged Modelling

Find the minimal perturbation of the prior information that reproduces exactly the experimental data, assuming that data are averaged over multiple conformations

The maximum entropy principle (MEP) recipe is to add an harmonic restraint between experimental and predicted data:

$$E_{RAM}(\mathbf{X}) = E_{MD}(\mathbf{X}) + \frac{1}{2}k(d - f(\mathbf{X}))^2$$

where predicted data is averaged over multiple replicas of the system:

$$f(\mathbf{X}) = \frac{1}{N}\sum_{r=1}^{N} f(X_r)$$

The intensity of the restraint should be as strong as possible and should scale more than linearly with the number of replicas *N*.

Camilloni *et al.* JACS 2012

Bayesian Modelling

$$p(X, \sigma|d) \propto p(d|X, \sigma) \cdot p(X) \cdot p(\sigma)$$

posterior **likelihood** **priors**

The **model** comprises the structure coordinates and additional parameters (noise level, calibration...).

The **likelihood** function encodes the agreement with the data *d* (through a **forward model**) and provides a model for the noise.

The **priors** define the probability of model, given any knowledge other than the data.

The **Bayesian score** is: $E_{Bayes}(X, \sigma) = -k_B T \cdot \log p(X, \sigma|d)$

Rieping *et al.* Science 2005

Addressing these challenges

Mixing	
Random	
Systematic	
Forward model	

Metainference*

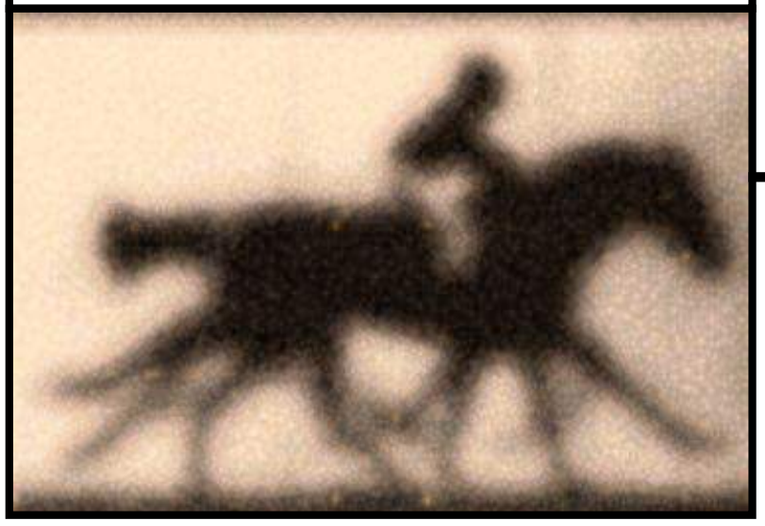
*Bonomi *et al.* Science Advances 2016

To produce ensemble of models

and determine their populations

a) Input

Measurement+Prior



b) Errors

Mixing

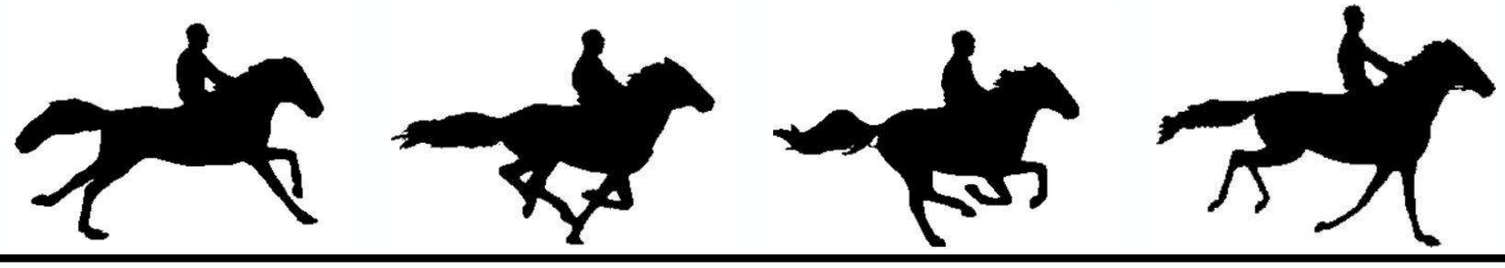
Random

Systematic

Forward model

c) Output

Metainference: Ensemble of models



30% 20% 45% 5%

Metainference

We want to determine to which extent a prior distribution of models is modified by the introduction of exp data

We model a sample of the distribution of models, made by *N* "replicas":

$$f(\mathbf{X}) = \frac{1}{N}\sum_{r=1}^{N} f(X_r)$$

The central limit theorem tells us that the error in calculating average quantities with a finite sample is:

$$p(\tilde{f}|\mathbf{X},\sigma^{SEM}) = \frac{1}{\sqrt{2\pi}\sigma^{SEM}}\exp\left[-\frac{(\tilde{f}-f(\mathbf{X}))^2}{2(\sigma^{SEM})^2}\right]$$

with the standard error of the mean decreasing with the dimension of the sample:

$$\sigma^{SEM} \propto 1/\sqrt{N}$$

Metainference

The Metainference posterior probability in the case of a single data point is:

$$p(\mathbf{X}, \tilde{\mathbf{f}}, \sigma^{\mathbf{B}}, \sigma^{\mathbf{SEM}}|d) \propto \prod_{r=1}^{N} p(d|\tilde{f}_r, \sigma_r^B) \cdot p(\tilde{f}_r|\mathbf{X}, \sigma_r^{SEM}) \cdot p(\sigma_r^B) \cdot p(X_r) \cdot p(\sigma_r^{SEM})$$

where:

$p(d\|\tilde{f}_r, \sigma_r^B)$	data likelihood: exp and theo errors σ_r^B
$p(\tilde{f}_r\|\mathbf{X}, \sigma_r^{SEM})$	CLT: statistical error in calculating averages
$p(\sigma_r^{SEM})$	CLT: fix σ_r^{SEM} dependence on *N*
$p(\sigma_r^B)$	prior on exp and theo errors
$p(X_r)$	prior on structure

One step beyond

In the case of Gaussian noise model, we can marginalize $\tilde{f}$

The Metainference energy function (or score) for the general case of N_d independent data points:

$$E_{MI}(\mathbf{X},\sigma) = k_B T \cdot \sum_{r=1}^{N} \left\{ -\log p(X_r) + \sum_{i=1}^{N_d} (d_i - f_i(\mathbf{X}))^2 \frac{1}{2\sigma_{r,i}^2} + \log \sigma_{r,i} - \log p(\sigma_{r,i}) \right\}$$

where $\sigma_{r,i}$ includes all sources of errors:

$$\sigma_{r,i} = \sqrt{(\sigma_{r,i}^{SEM})^2 + (\sigma_{r,i}^{B})^2}$$

and $\sigma_{r,i}^{SEM} \propto 1/\sqrt{N}$

errors are negligible

Replica-Averaged Modelling

data is not generated by an ensemble

Bayesian Modelling

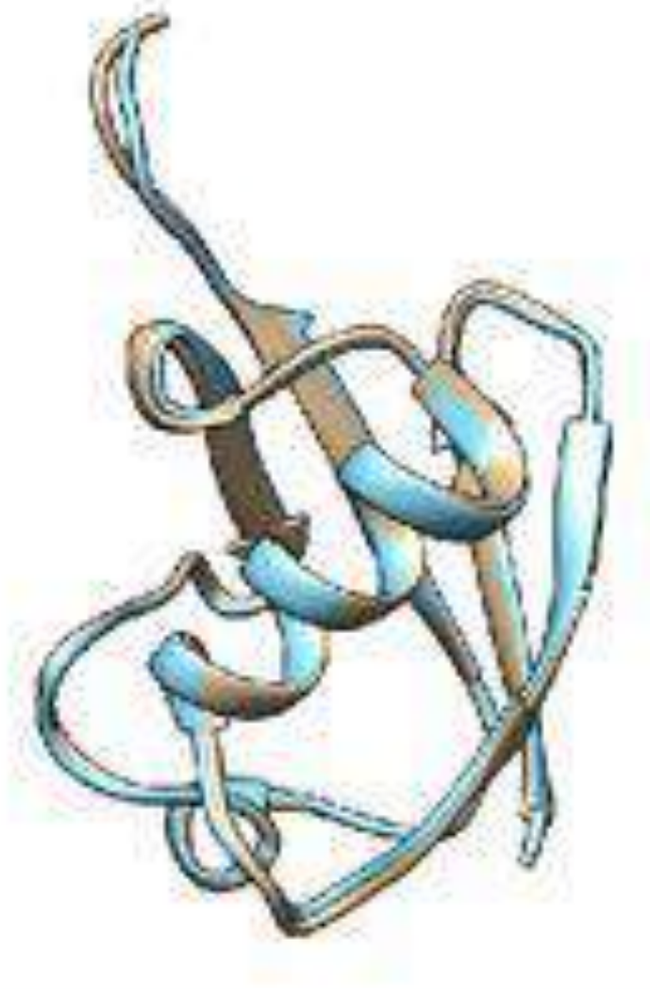
Integrative Dynamical Biology

We compare Metainference and replica-averaged modeling with real experimental data collected on ubiquitin:

- Chemical Shifts + RDCs

We also compare the Metainference ensemble with single structures:

- X-ray (1UBQ)
- NMR (1D3Z)



Cα-RMSD = 0.52 Å

and with the ensemble generated by standard MD

Models are evaluated by fit with other exp data (RDCs, J3)

Bonomi *et al.* Science Advances 2016

Technical details

- CHARMM22* with TIP3P explicit solvent, ~25000 atoms

  Piana *et al.* Biophys. J. 2011 Jorgensen *et al.* JCP 1983

- GROMACS 4.6.7 + PLUMED 2 (development branch)

  Hess *et al.* JCTC 2008 Tribello *et al.* CPC 2014

- Double parallelization: 8 replicas (ensemble modelling) x 8 cores per replica

- Non-bonded interactions cutoff at 0.9 nm + PME

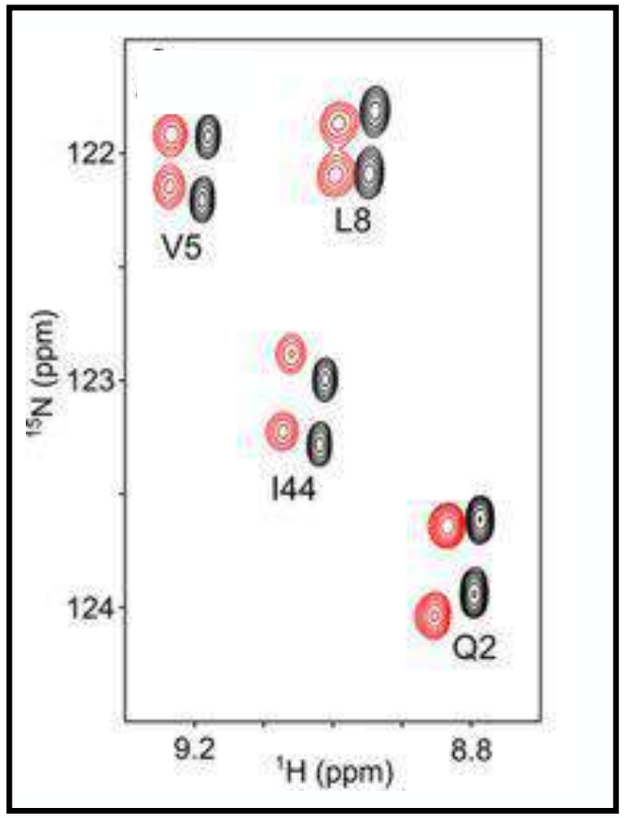
- NVT with Bussi-Donadio-Parrinello thermostat Bussi *et al.* JCP 2007

- Chemical Shifts predicted by Camshifts Kohlhoff *et al.* JACS 2009

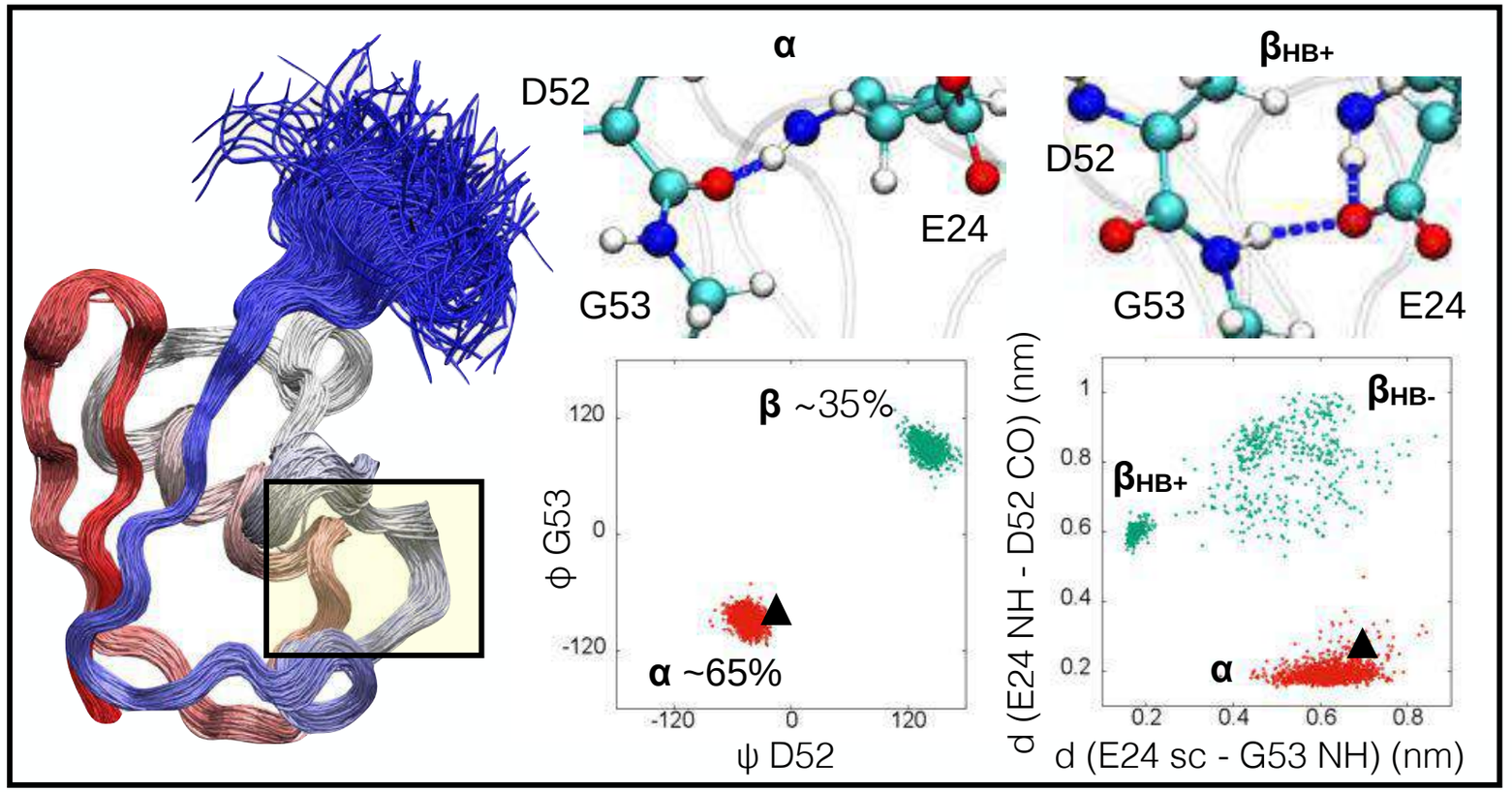
- RDC predicted by θ-method Camilloni & Vendruscolo JPCB 2015

Ubiquitin ensembles

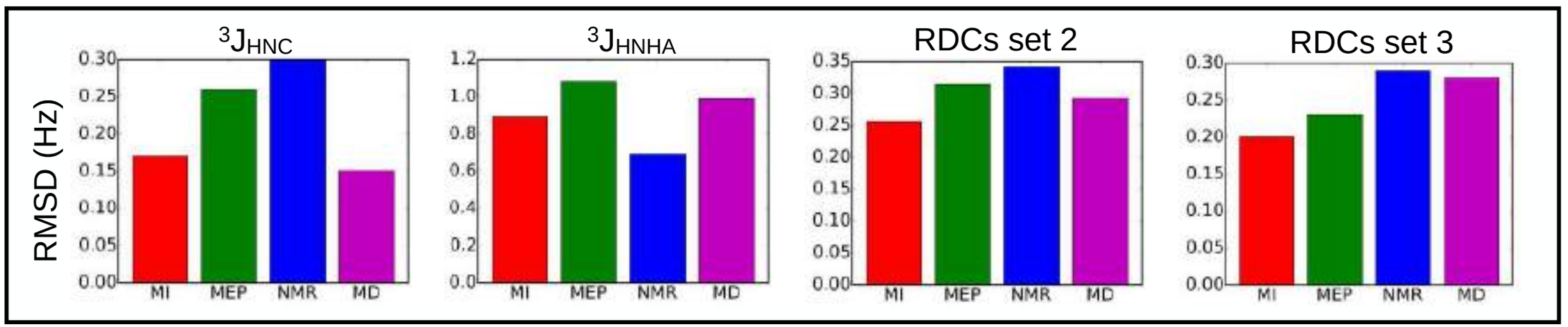
a) Input



b) Ensemble

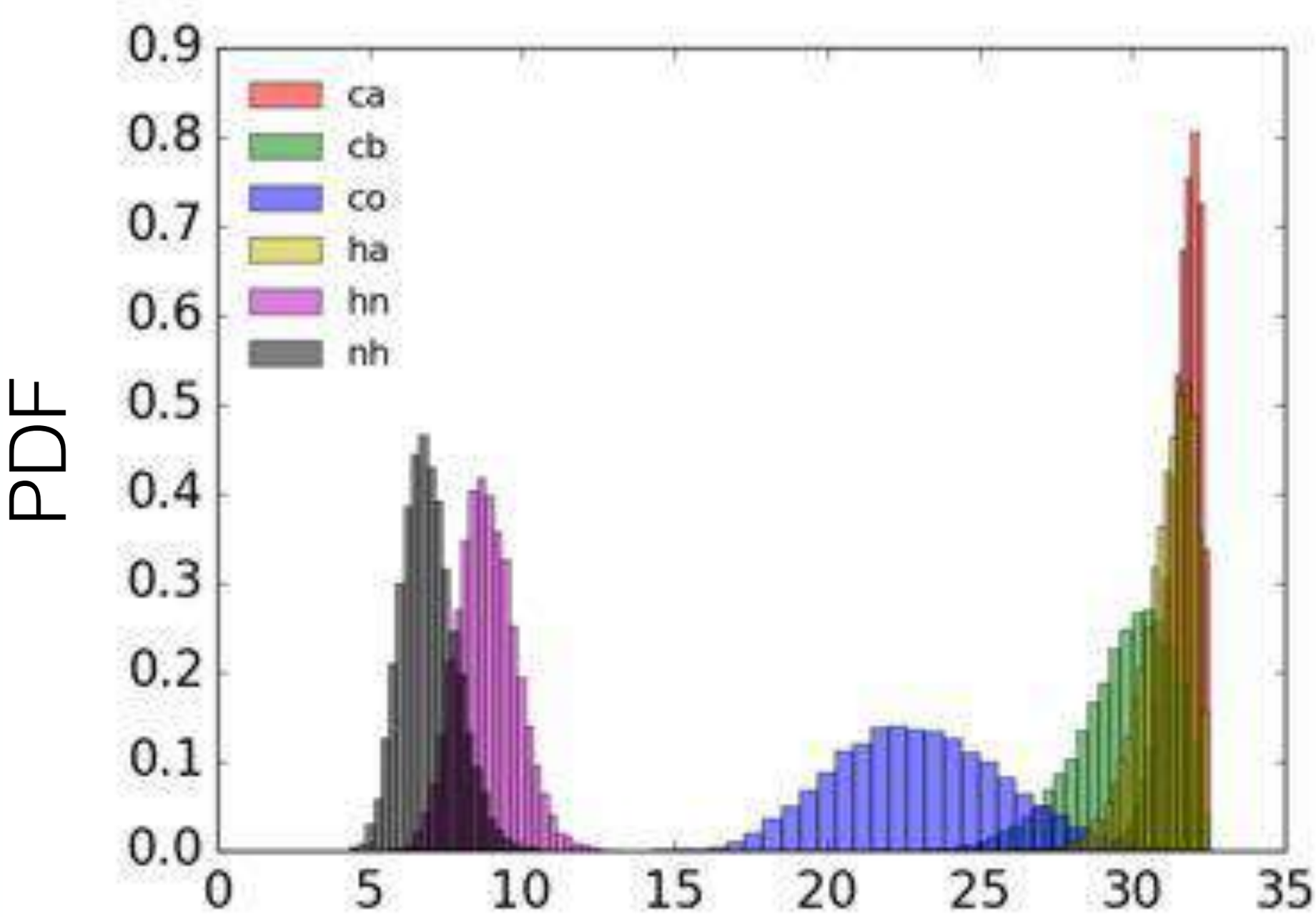


c) Validation



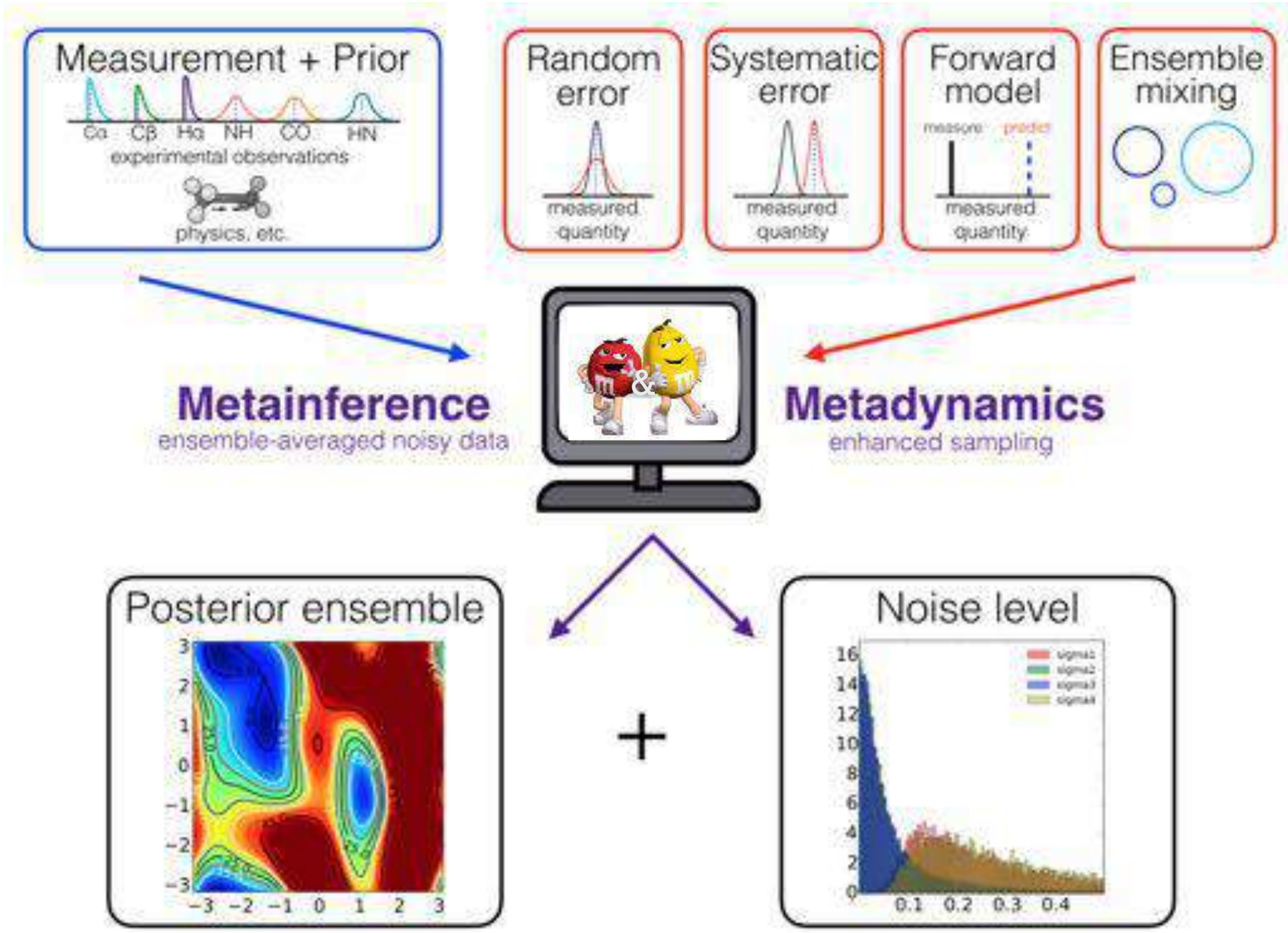
Bonomi *et al.* Science Advances 2016

Chemical Shifts weights



Bonomi *et al.* Science Advances 2016

Metadynamic Metainference



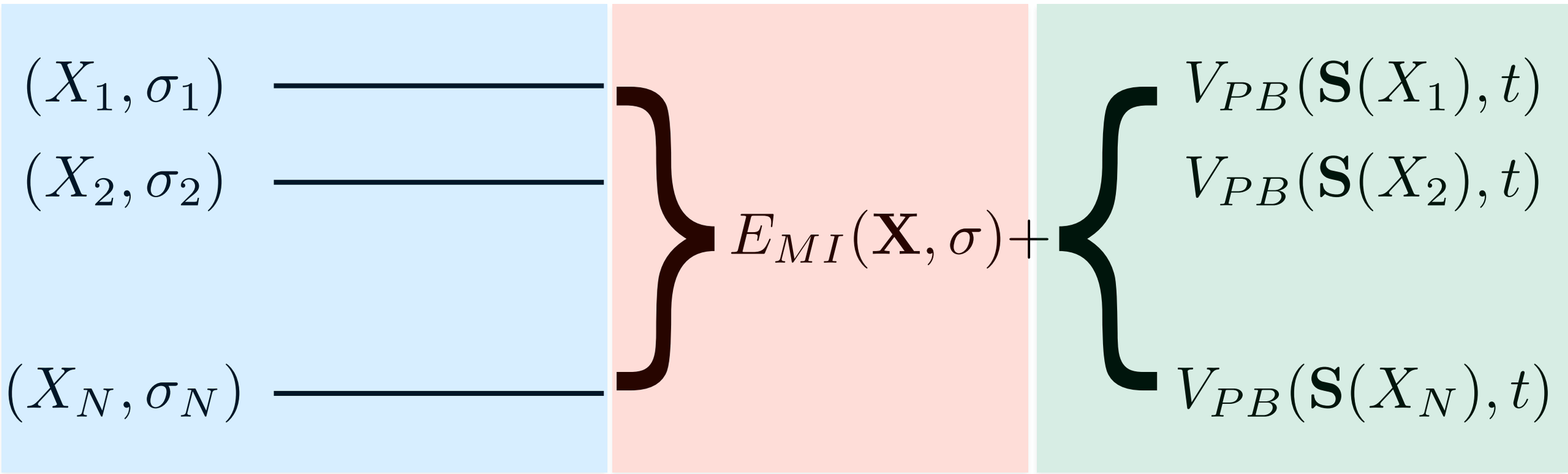
Bonomi *et al.* Scientific Reports. 2016

Metadynamics Metainference

Ensemble of replicas

Metainference energy function

PBMetaD bias



with these additional tricks:

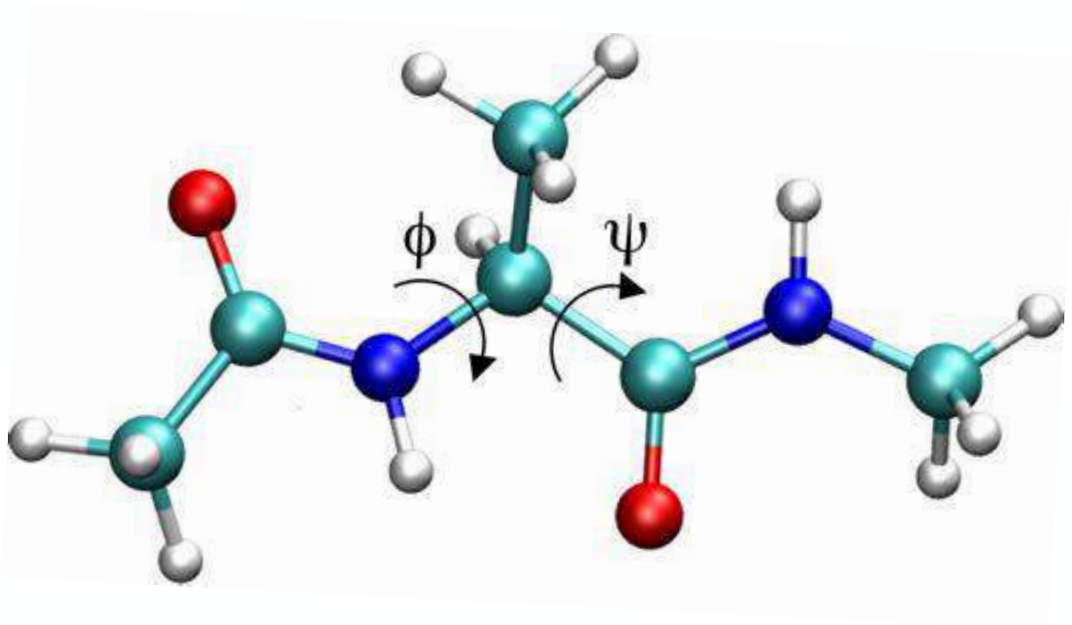
- replicas share the bias, as in multiple-walkers MetaD*
- need to reweigh to calculate averages in the unbiased ensemble

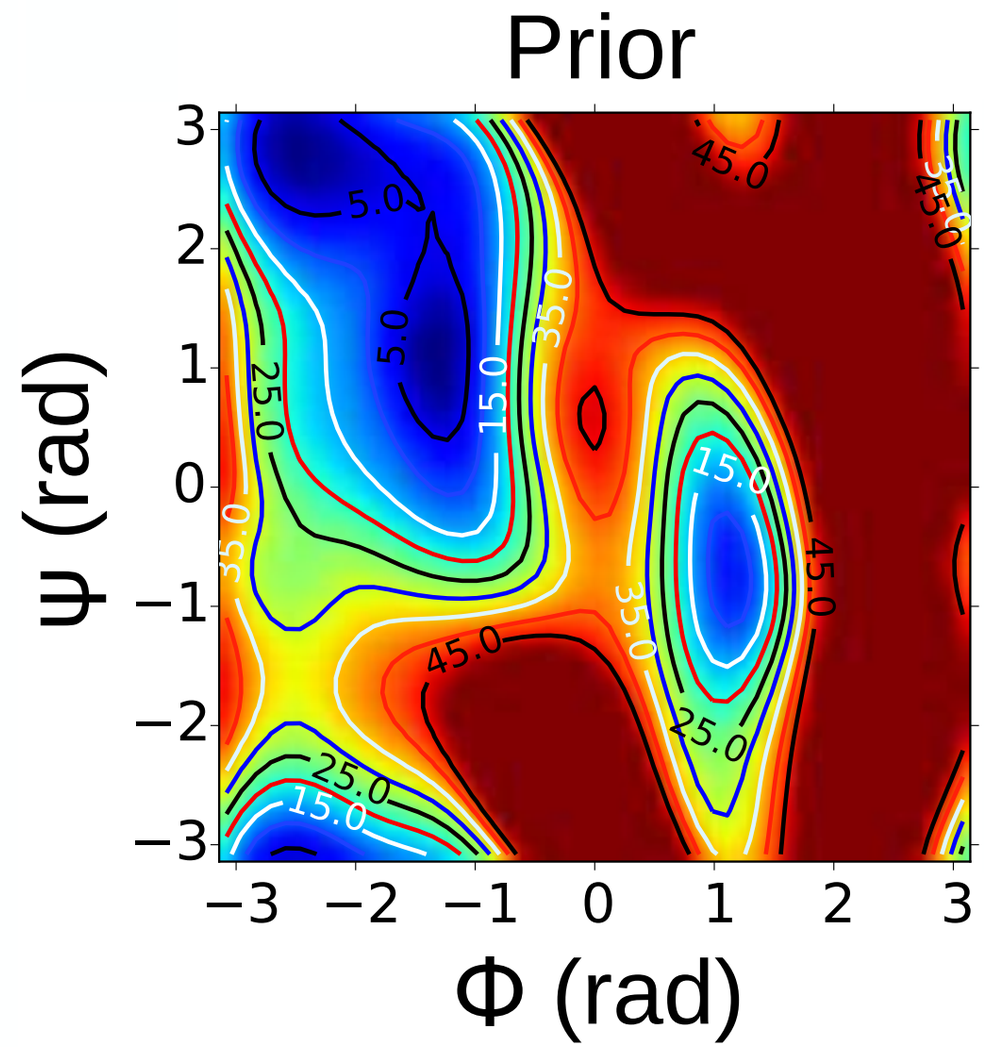
*Raiteri *et al.* JPCB 2006

Benchmark

Our favorite test case: alanine dipeptide in vacuum

The prior information is the AMBER99SB-ILDN force field*

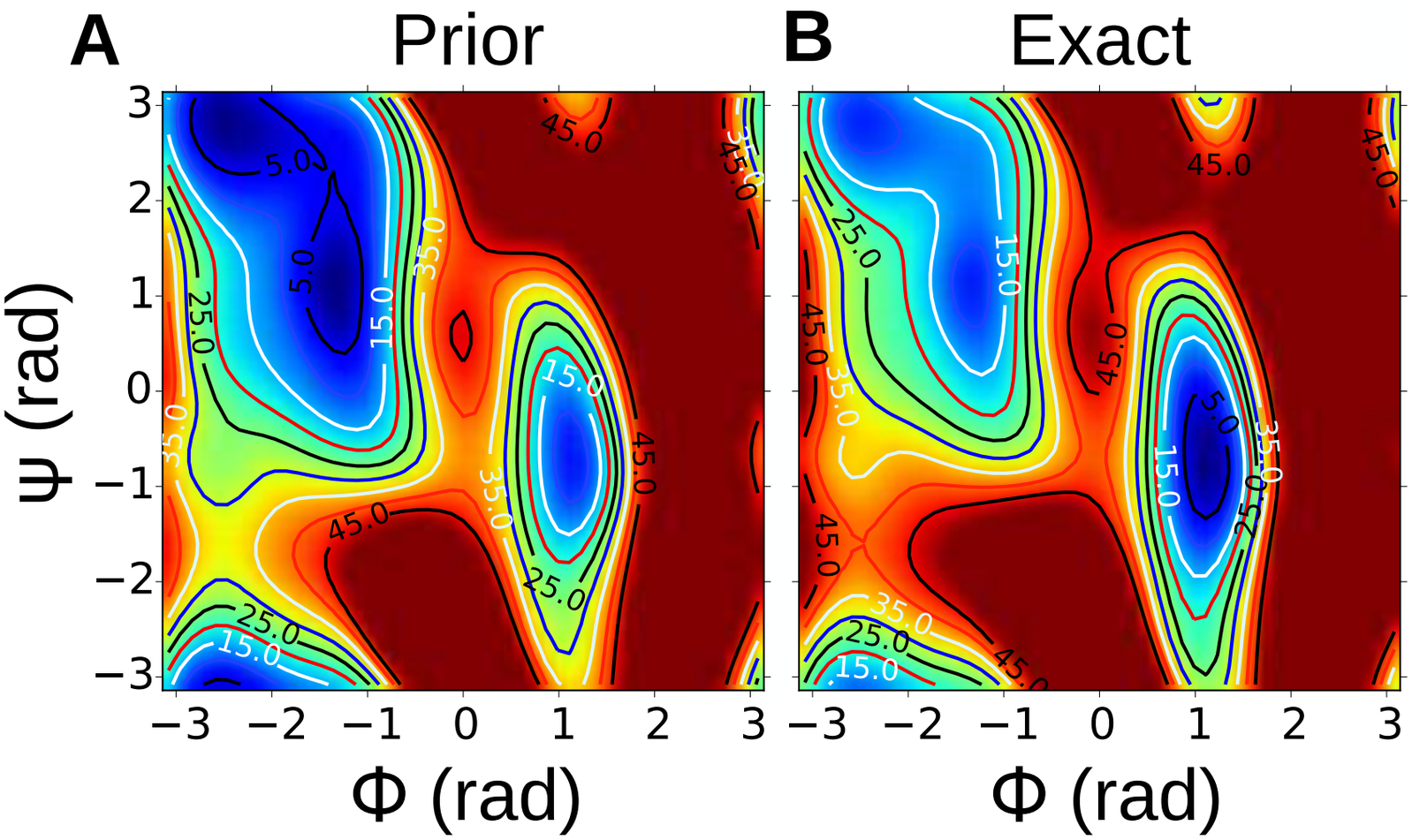




*Lindorff-Larsen *et al.* Proteins 2010

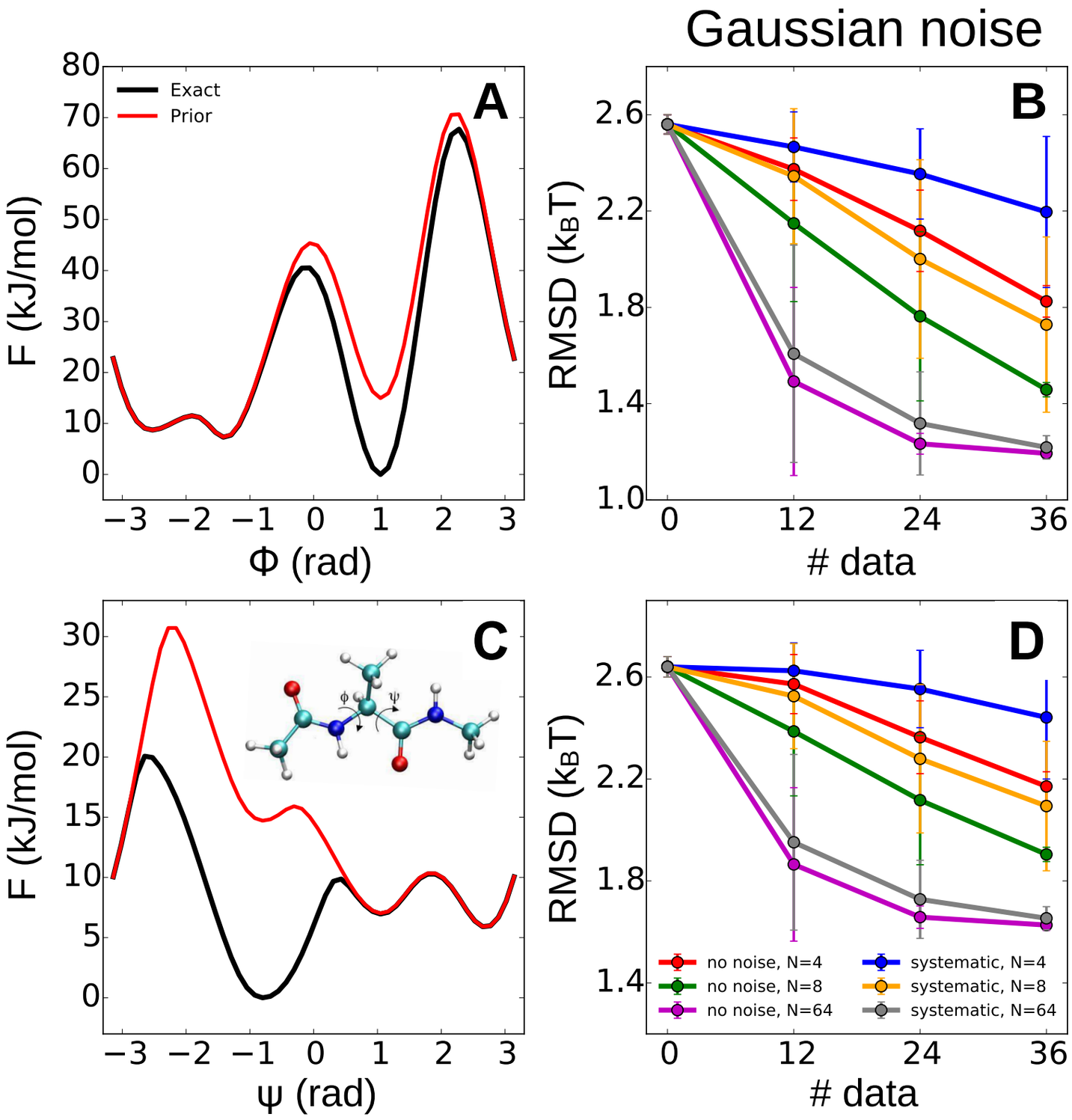
Benchmark

We assume that the prior is inaccurate and that in the real distribution the relative weight of the two minima is different:



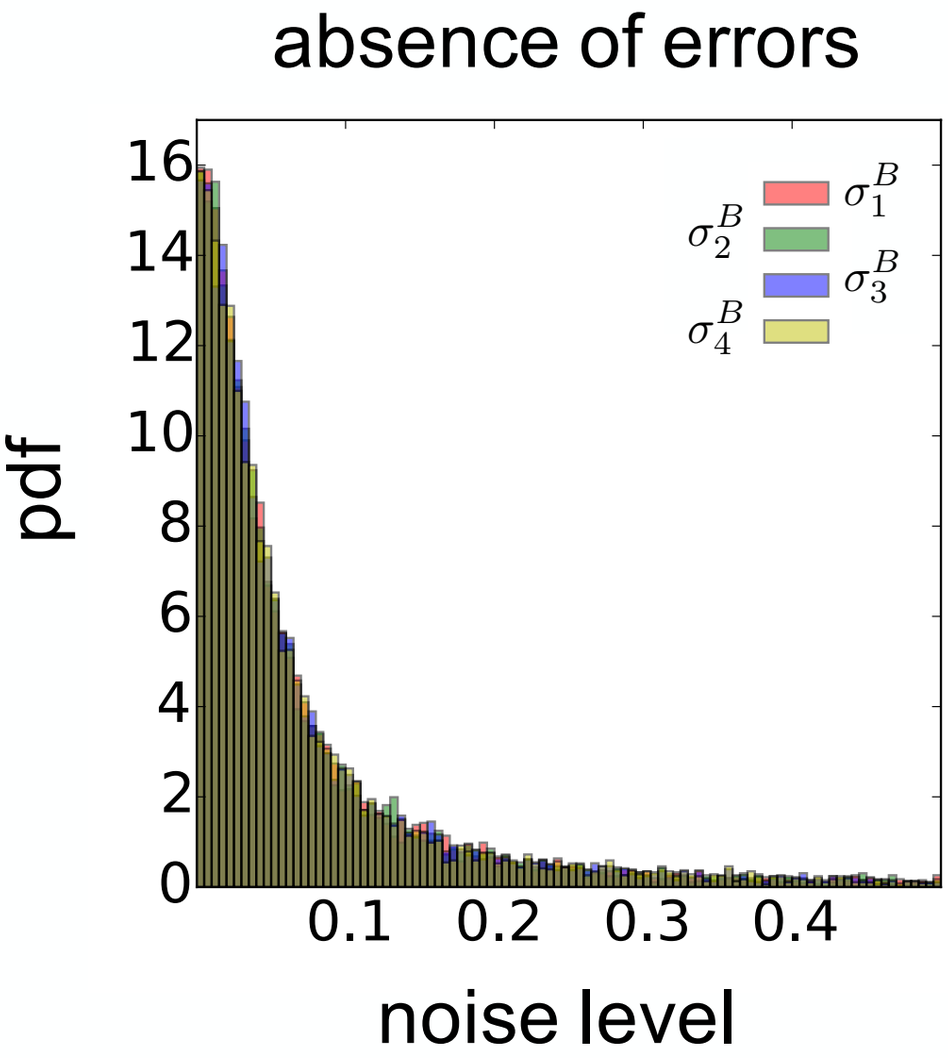
We introduce synthetic experimental data as average distances between heavy atoms, calculated in the <u>exact</u> ensemble, + **noise**

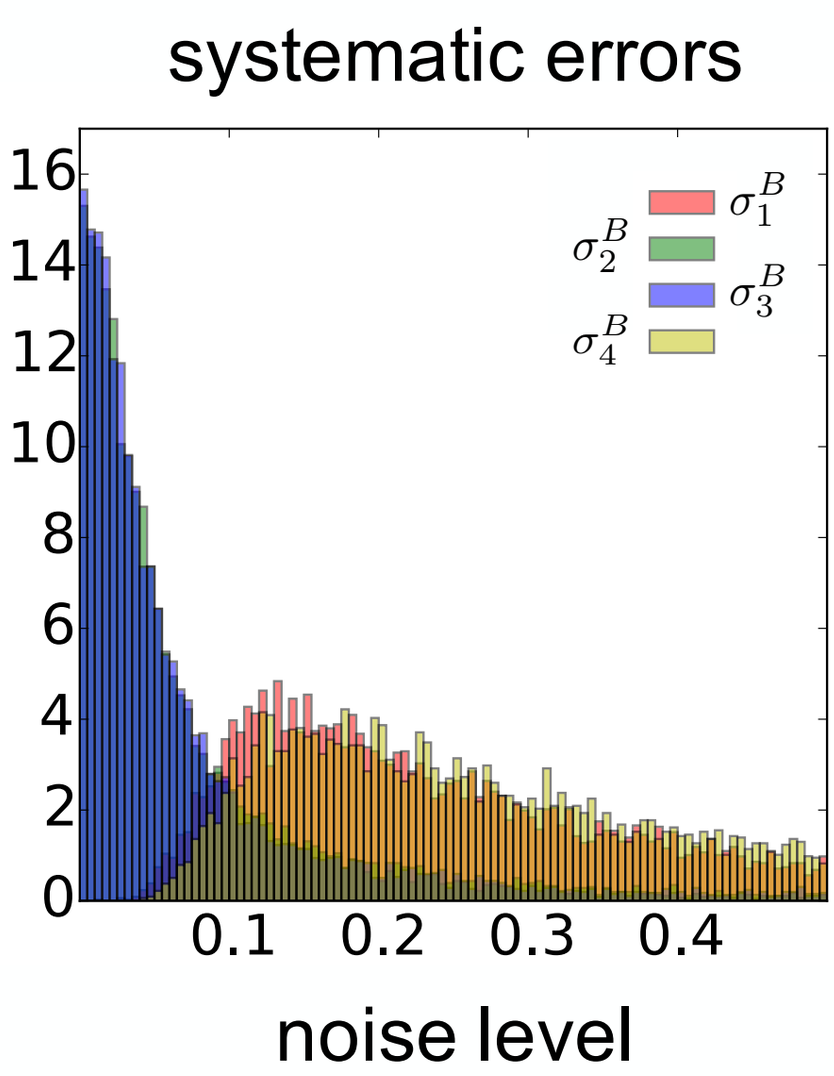
Results



Bonomi *et al.* Scientific Reports. 2016

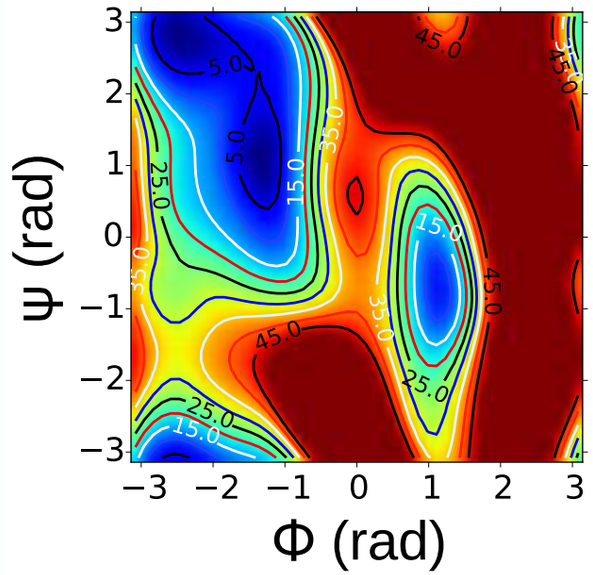
Noise inference



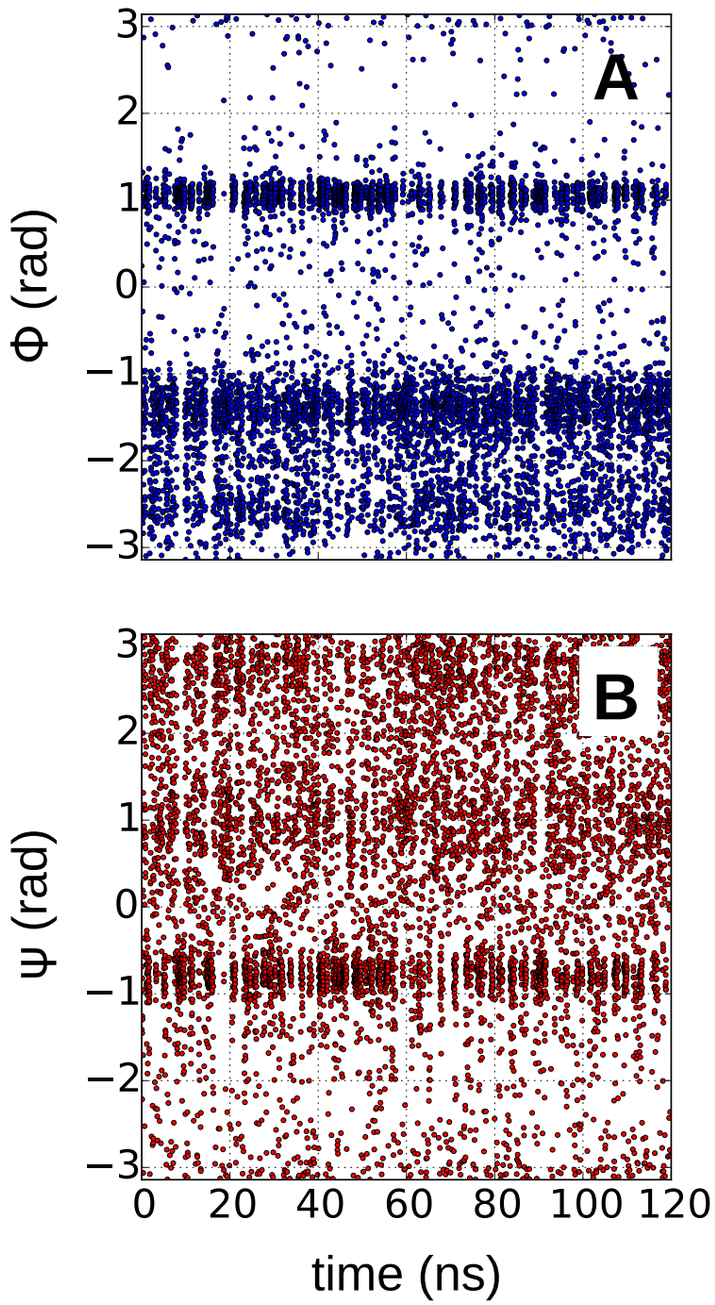


Bonomi *et al.* Scientific Reports. 2016

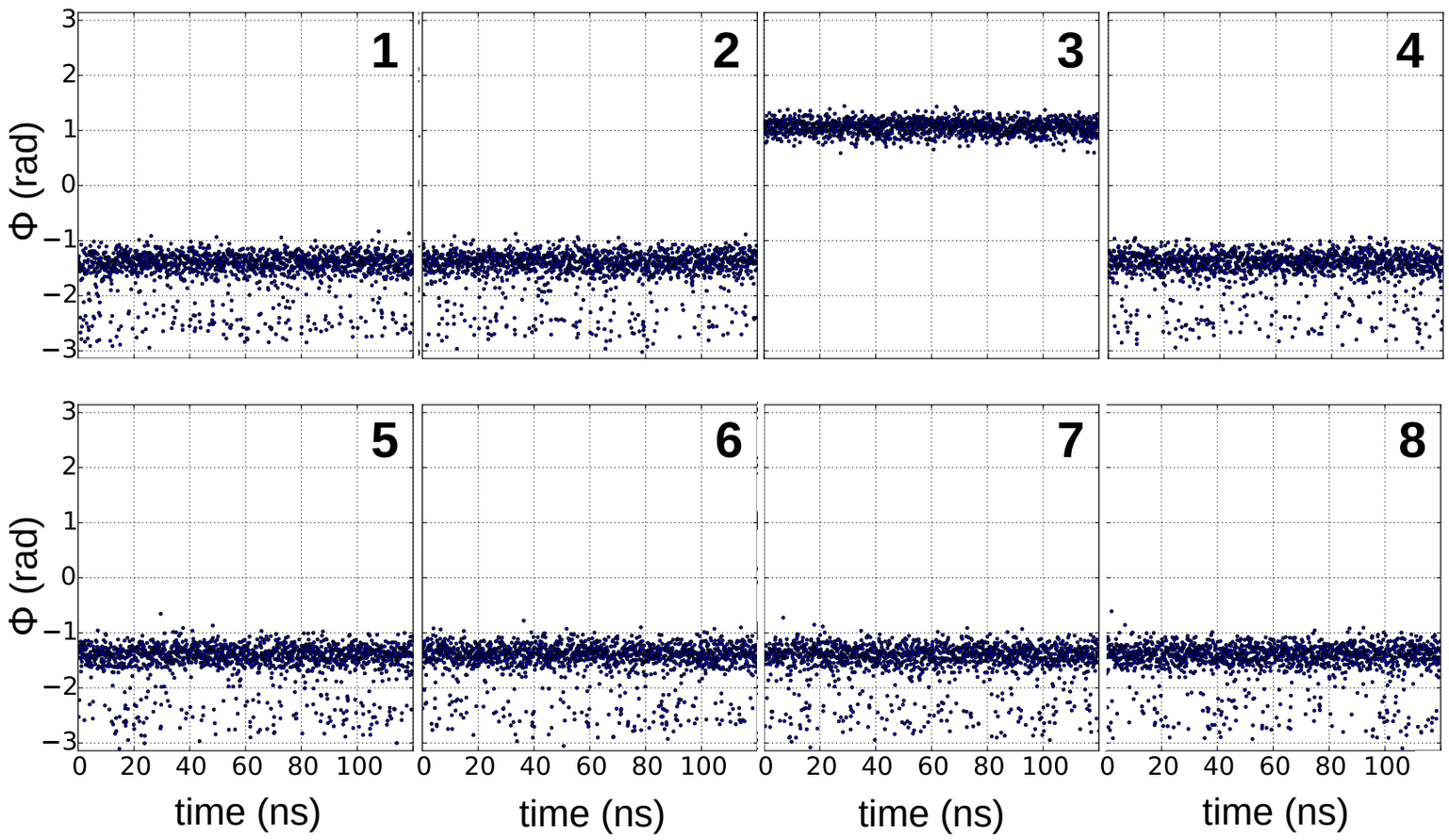
Metainference alone



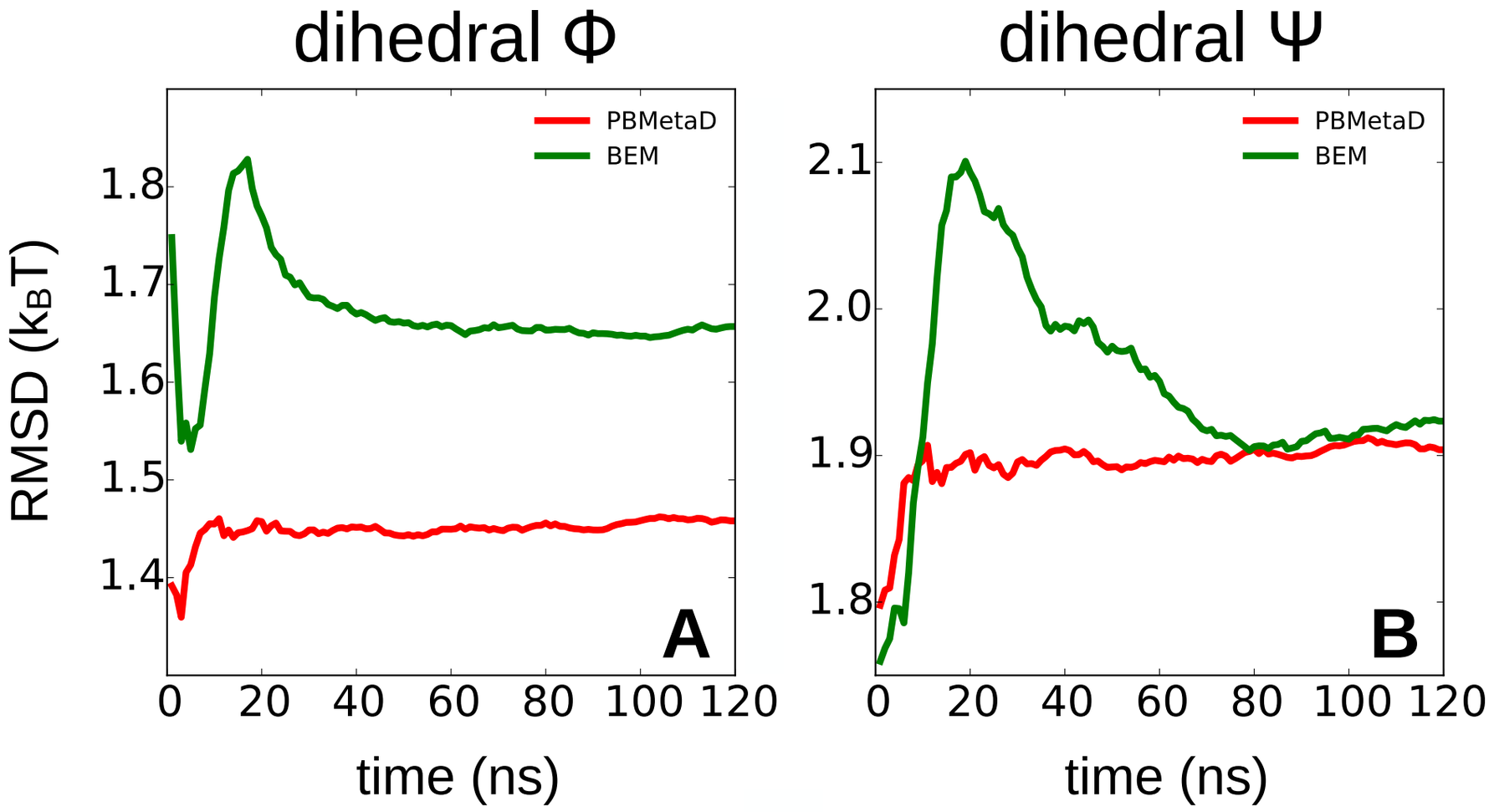
with PBMetaD



without PBMetaD



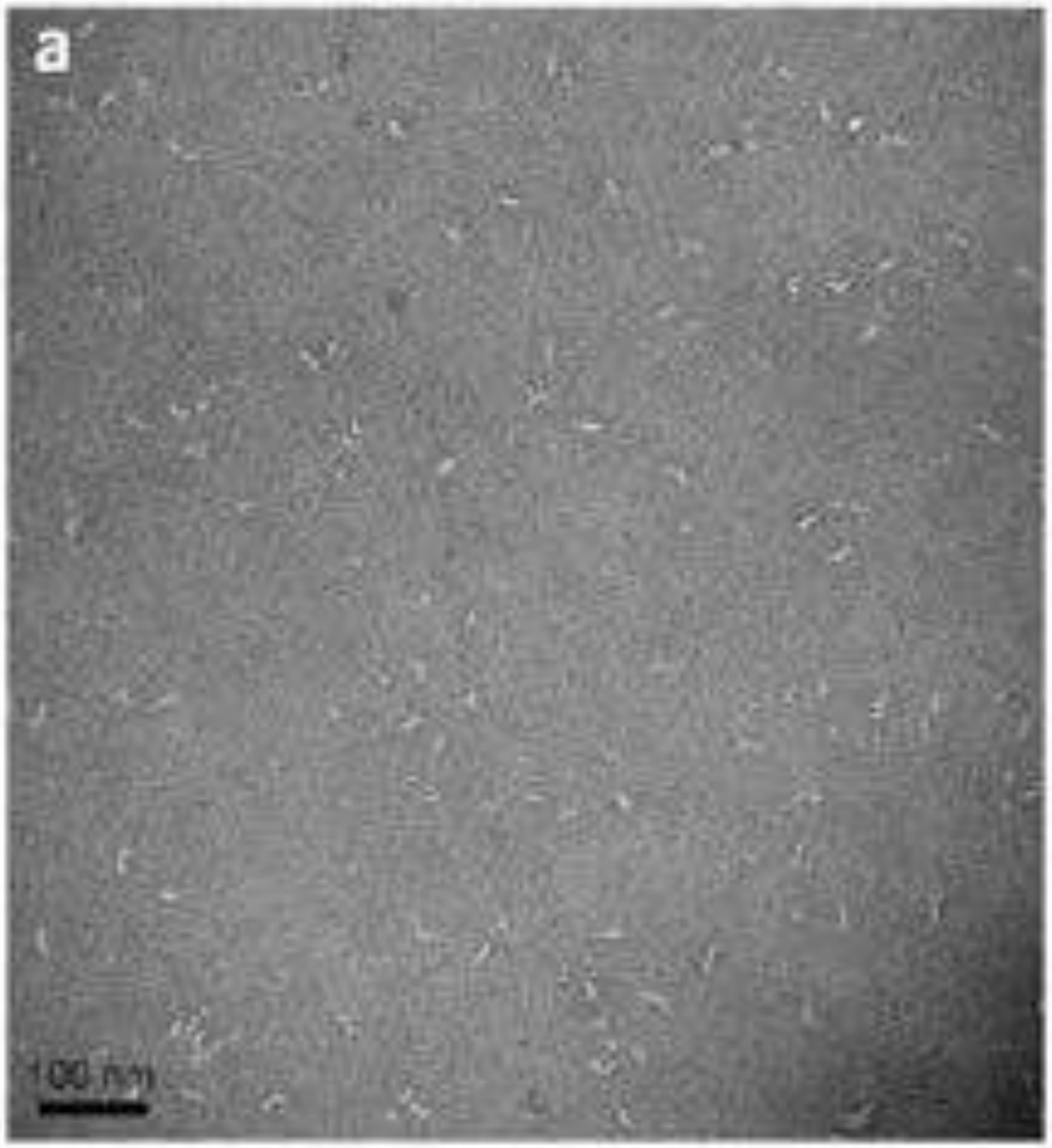
BEM+Metainference



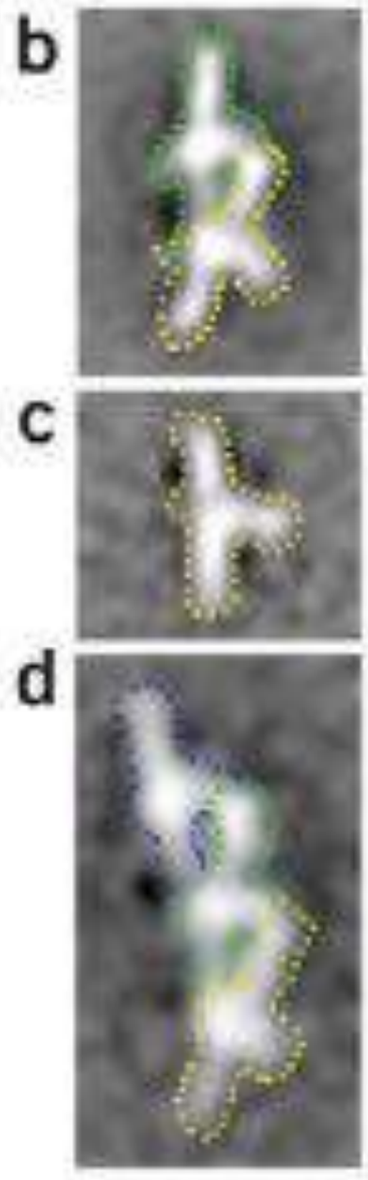
- + <u>efficiency</u>: replicas share the bias
- + <u>accuracy</u>: averages are calculated in the unbiased ensemble by on-the-fly reweighing

3D EM reconstructions

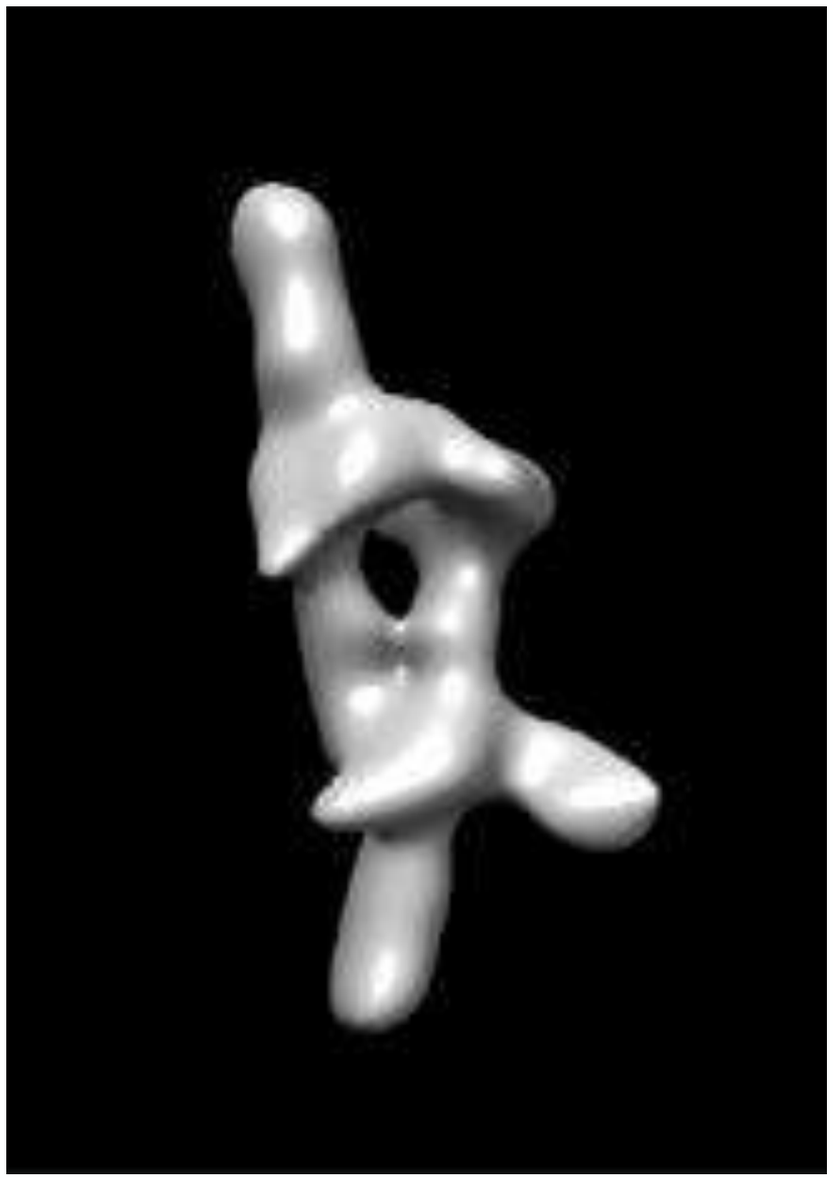
Negative-stain EM of Dam1



Class averages



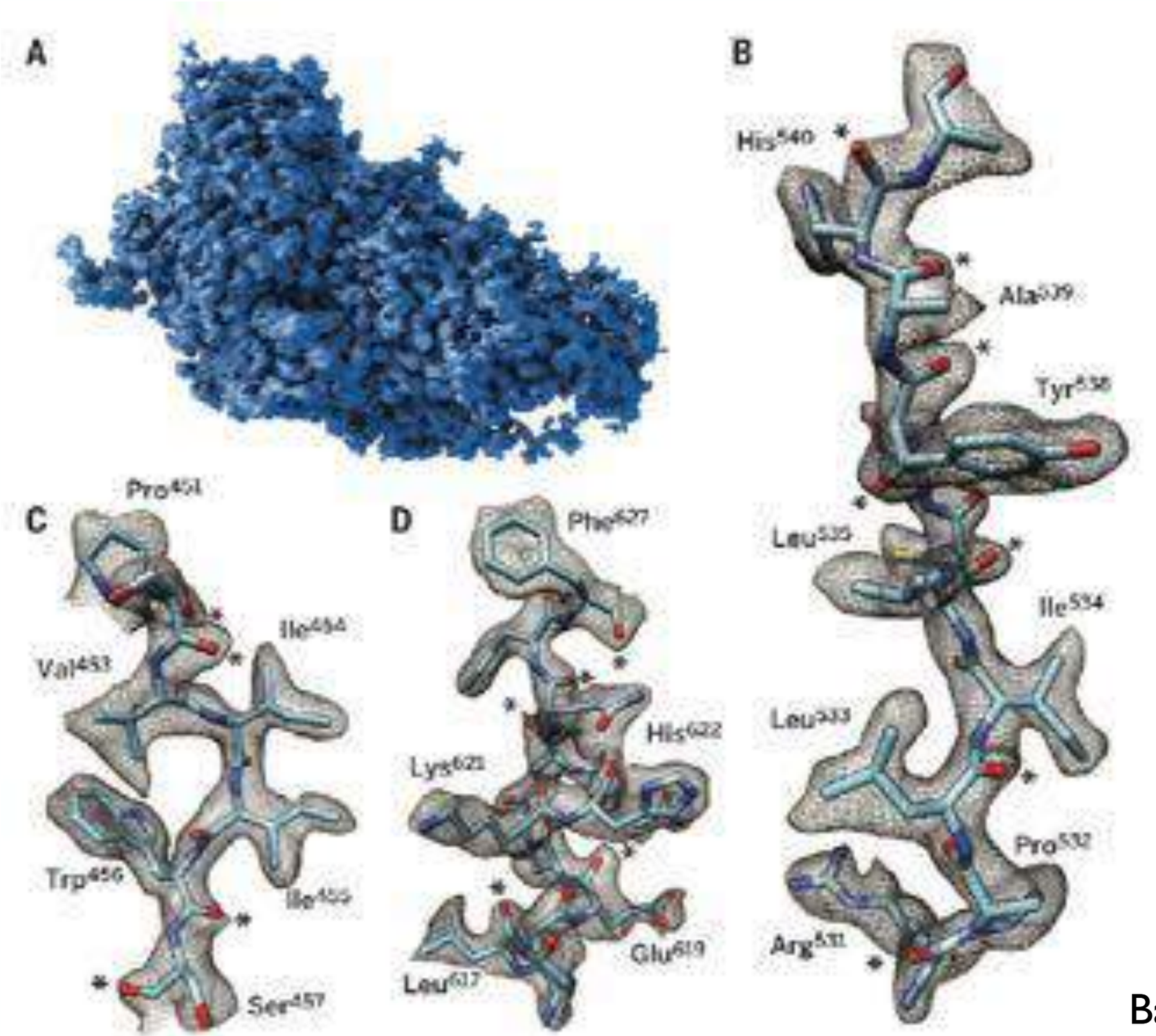
Single-particle 3D reconstruction



resolution = 28 Å

Nogales *et al.* NSMB 2015

Approaching the resolution of X-ray crystallography



Cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor

Resolution = 2.2 Å

Bartesaghi *et al.* Science 2015

Bayesian modelling of EM data

To use 3D EM reconstructions with metainference we need:

- a forward model, *i.e.* a predictor of the EM map from a single structure
- a model of noise

The forward model is a Gaussian Mixture Model*, with one Gaussian centered on each atom (or coarse-grained bead) $\mathbf{R}_i$:

$$f_M(\mathbf{x}) = \sum_{i=1}^{N} \pi_i \, f_{M,i}(\mathbf{x}|\mathbf{R}_i, \sigma_i)$$

with:

- π_i = relative mass of the component
- $\sigma_i \propto$ radius of the component

*Robinson *et al.* eLife 2015

Bayesian modelling of EM data

The data *D* is also represented by a GMM f_D :

- optimal number of Gaussians to represent the data from resolution of the experimental map

- components of data GMM $f_{D,k}$ are treated as independent data points

The posterior distribution is then:

$$p(\mathbf{R}, \sigma | D) \propto p(f_D | \mathbf{R}, \sigma)\, p(\mathbf{R})\, p(\sigma)$$

and from the independence of the data GMM components:

$$p(\mathbf{R}, \sigma | D) \propto p(\mathbf{R}) \prod_{k=1}^{N_D} p(f_{D,k} | \mathbf{R}, \sigma_k)\, p(\sigma_k)$$

Noise models

We use a lognormal likelihood:

$$p(f_{D,k}|\mathbf{R},\sigma_k) = \frac{1}{\sqrt{2\pi}\, d_{DD,k}\, \sigma_k} \cdot exp\left[-0.5 \log\left(\frac{d_{DD,k}}{d_{MD,k}}\right)^2 / \ \sigma_k^2\right]$$

where $d_{MD,k}$ measures the overlap between mixture models:

$$d_{MD,k} = \int d\mathbf{x}\, f_M(\mathbf{x})\, f_{D,k}(\mathbf{x})$$

In metainference, we simulate an ensemble of replicas and the forward model is averaged over the replicas:

$$d_{MD,k} = \int d\mathbf{x} \left(\frac{1}{N}\sum_{r=1}^{N} f_M^r(\mathbf{x})\right)\ f_{D,k}(\mathbf{x}) = \frac{1}{N}\sum_{r=1}^{N} d_{MD,k}^r = \langle d_{MD,k}\rangle$$

Noise models

We will test different two assumptions:

1) same level of noise in all parts of the map, which can then be marginalized:

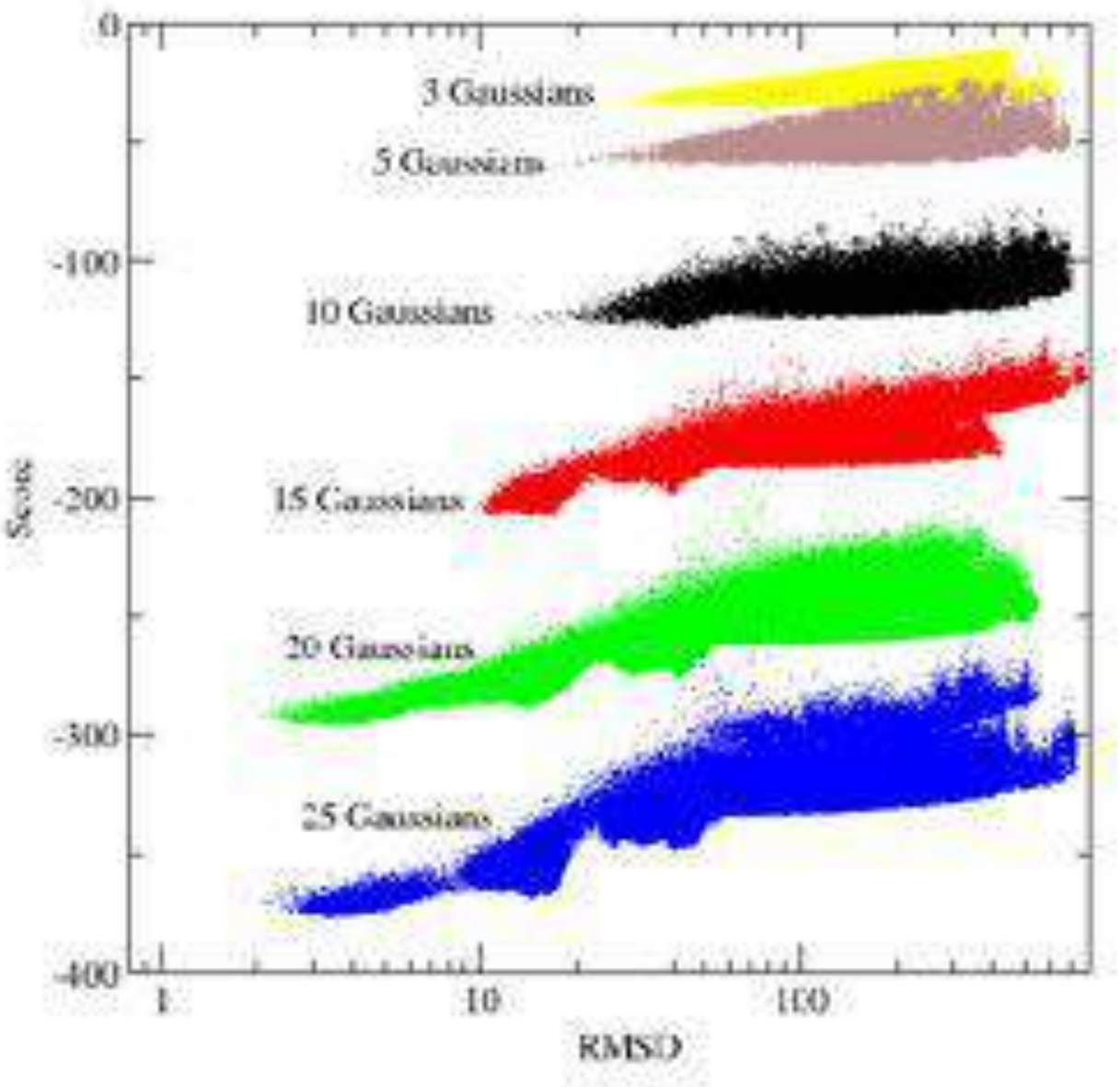
$$p(f_D|\mathbf{R}) \propto \frac{\left(\sum_{k=1}^{N_D} \log^2\left(d_{MD,k}/\ d_{DD,k}\right)\right)^{-N_D/2}}{\prod_{k=1}^{N_D} d_{DD,k}}$$

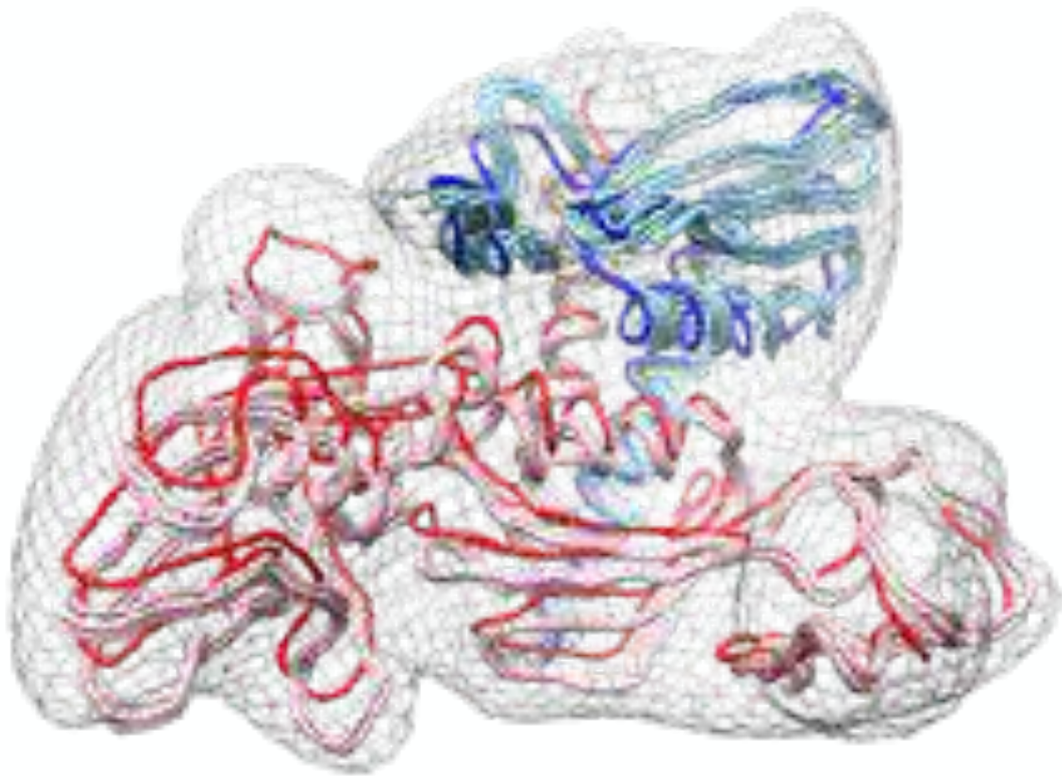
2) different levels of noise, but distributed around a typical value σ_0 , with long tail to tolerate outliers:

$$p(f_D|\mathbf{R},\sigma_0) \propto \prod_{k=1}^{N_D} \frac{\sigma_0}{d_{DD,k}} \cdot \frac{1}{\log^2\left(\frac{d_{DD,k}}{d_{MD,k}}\right) + 2\sigma_0^2}$$

Results I

Rigid body docking of two subunits of RNA polymerase II, as a function of the number of Gaussians used to represent the (synthetic) data



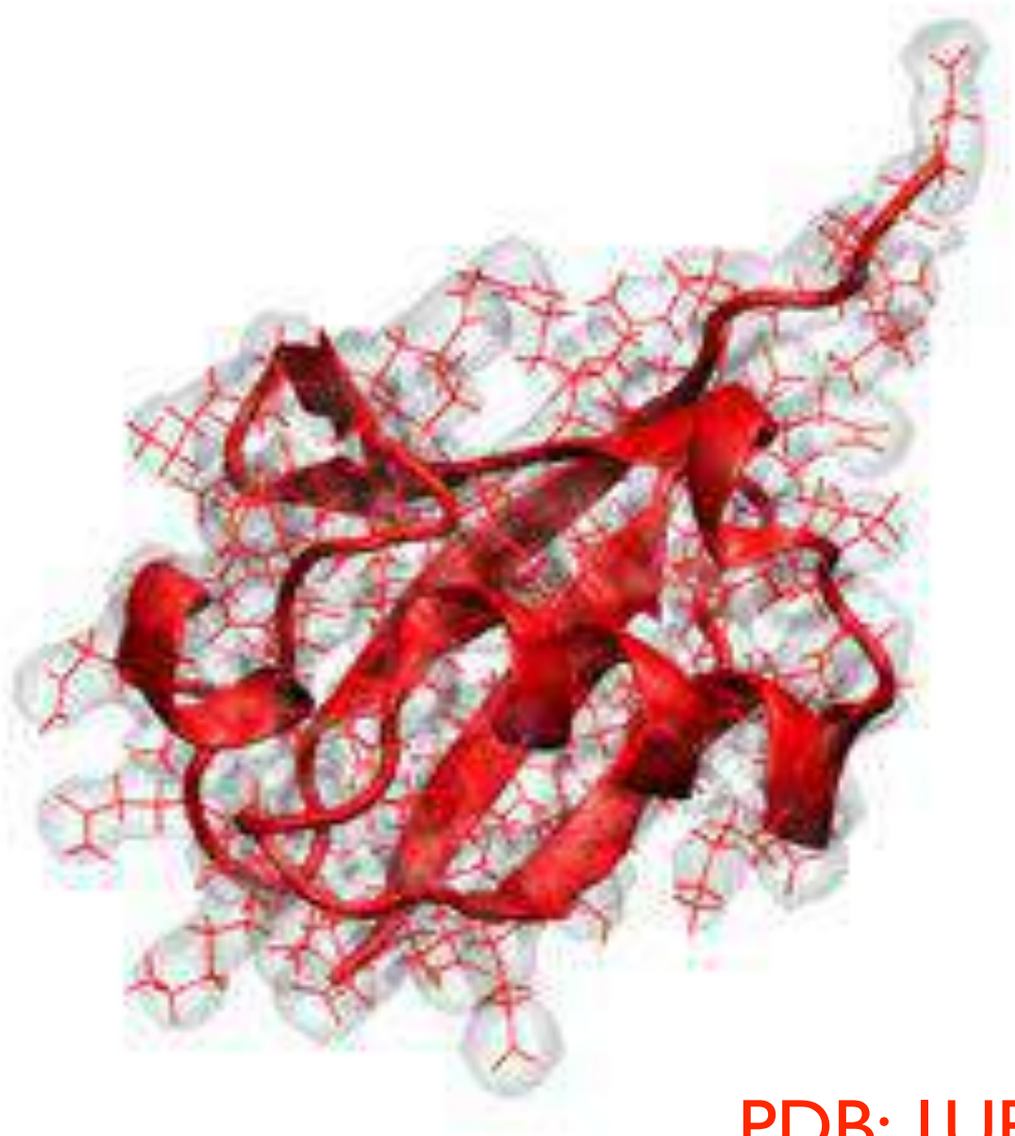


with S. Hanot, R. Pellarin

Results II

atomistic refinement of Ubiquitin

- high-resolution synthetic map
- absence of noise in the data
- one noise parameter
- AMBER99SB*-ILDN
- Implicit solvent
- Sampling with PT



PDB: 1UBQ

with S. Hanot, R. Pellarin

Results II

atomistic refinement of Ubiquitin

- high-resolution synthetic map
- absence of noise in the data
- one noise parameter
- AMBER99SB*-ILDN
- Implicit solvent
- Sampling with PT

RMSD [Å]	backbone	all-H
initial	4.4	5.2
refined	0.4	1.4

PDB: 1UBQ

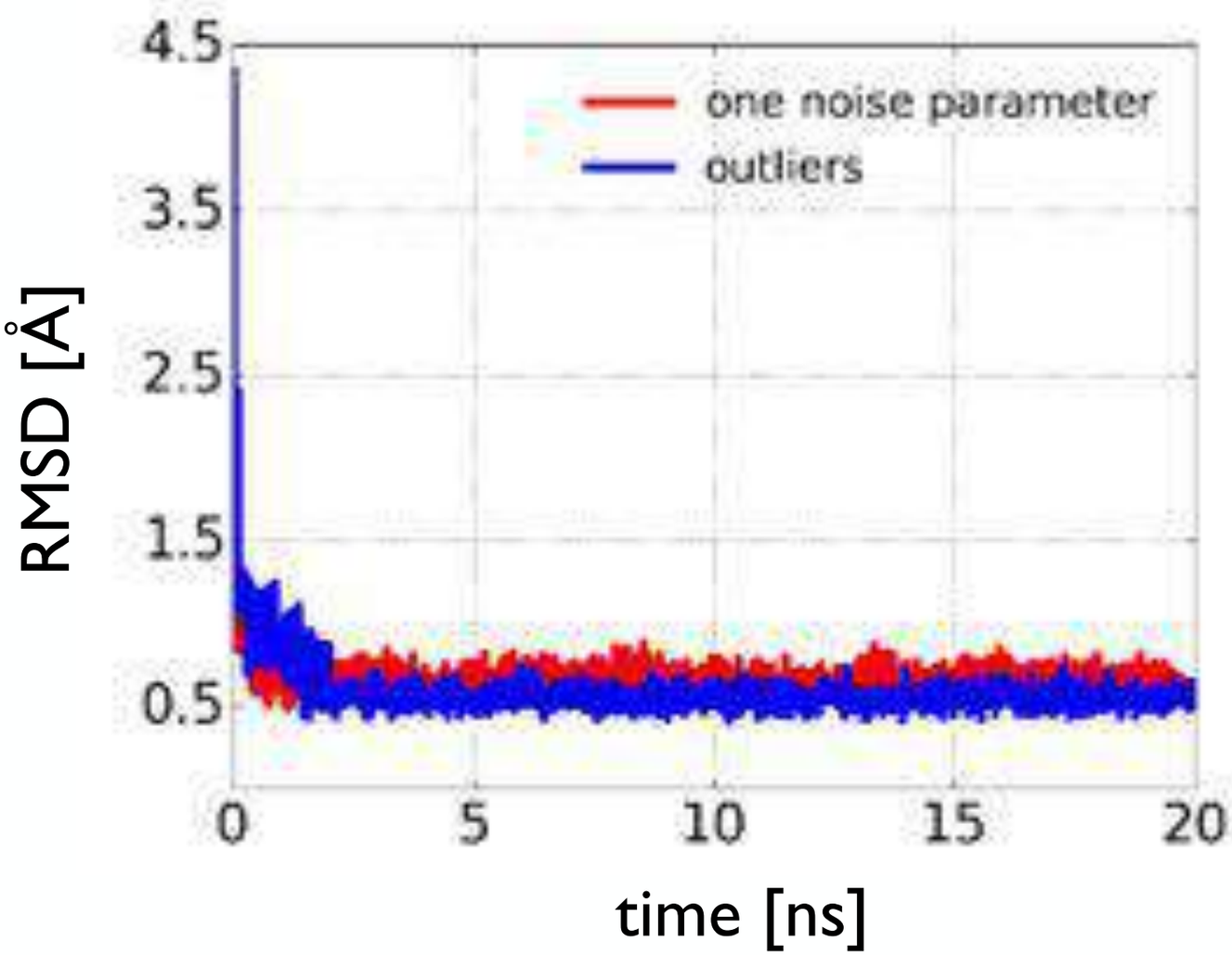
MI

with S. Hanot, R. Pellarin

Comparing noise models

atomistic refinement of Ubiquitin

- high-resolution synthetic map
- Gaussian noise to 20% of data GMM components
- Noise models with one parameter and outliers
- AMBER99SB*-ILDN
- Implicit solvent
- Sampling with PT



Conclusions

Metainference integrates noisy data collected on heterogeneous systems into MD simulations to improve the accuracy of force fields

M&M enables modelling ensemble of states separated by high free-energy barriers, using noisy and ensemble-average data

Applications of M&M include:

- ensemble structural determination from NMR data
- modelling of cryo-electron microscopy data
- analyze microfluidic data on polydisperse mixtures
- modulation of IDPs landscape by small molecules

Acknowledgments

Jim Pfaendtner

Carlo Camilloni
Andrea Cavalli
Gabi Heller
Francesco Aprile
Michele Vendruscolo

Samuel Hanot
Riccardo Pellarin

Paolo Arosio
Thomas Muller
Tuomas Knowles

Davide Branduardi
Giovanni Bussi
Gareth Tribello

All of you for your attention!

Tutorial instructions

plumed.github.io/doc-v2.3/user-doc/html/cineca.html

A computational microscope

Molecular Dynamics (MD) evolves a system in time under the effect of a potential energy function

How? By integrating Newton's equations of motion

$$m_i \ddot{\mathbf{R}}_i = -\nabla_{\mathbf{R}_i} V$$

The potential (or force field) is derived from

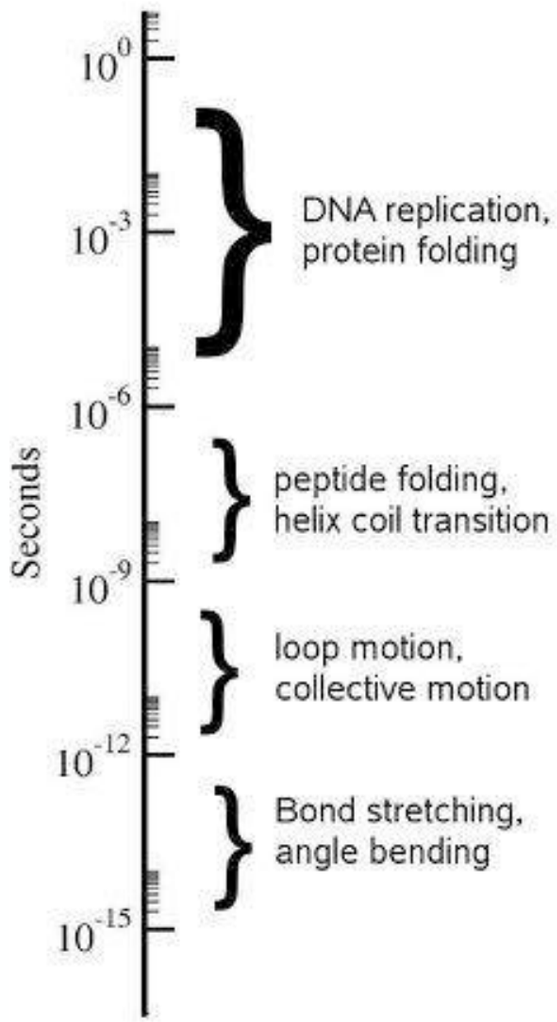
- Higher accuracy calculations
- Fitting experimental observables

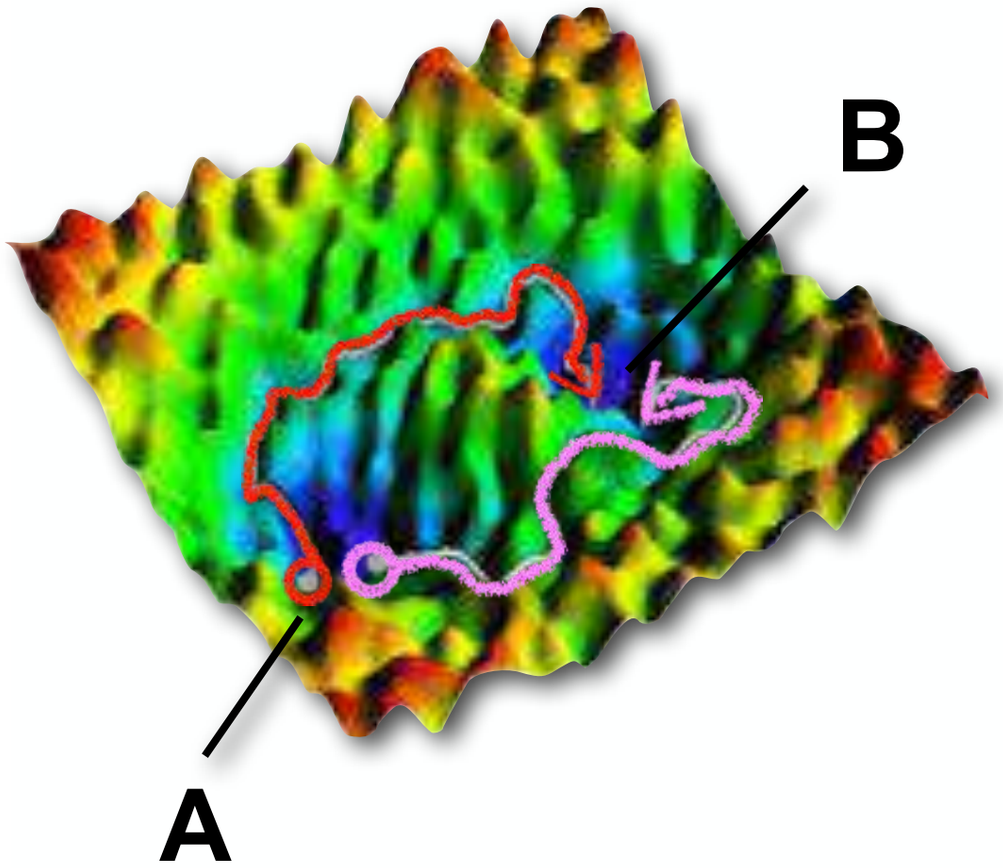
Limitations:

- time scale accessible in standard MD
- accuracy of classical force fields

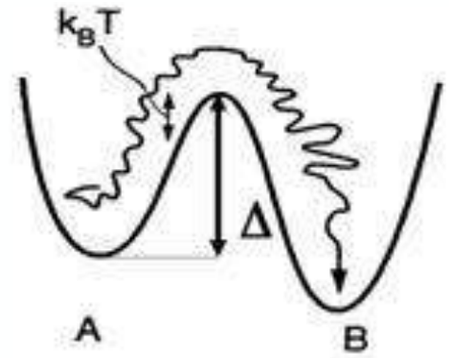
The time scale problem

In MD, sampling efficiency is limited by the time scale accessible in typical simulations:





★ Activated events



★ Slow diffusion

Dimensional reduction

It is often possible to describe a physical/chemical process in terms of a small number of coarse descriptors of the system:

$$\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{R}) = (S_1(\boldsymbol{R}), \ldots, S_d(\boldsymbol{R}))$$

Key quantity of thermodynamics is the free energy as a function of these variables:

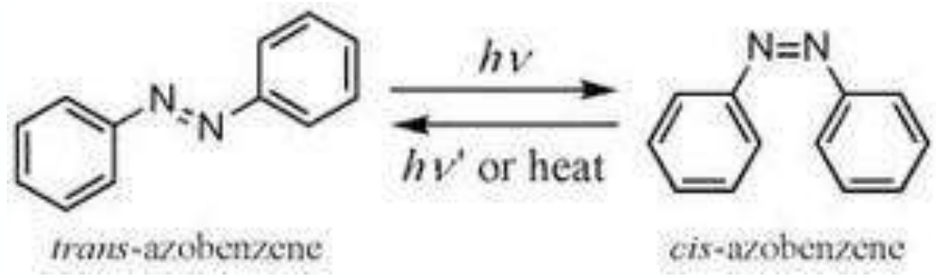
$$F(\boldsymbol{S}) = -\frac{1}{\beta} \ln P(\boldsymbol{S}) \quad \text{where} \quad \beta = \frac{1}{k_B T}$$

canonical ensemble

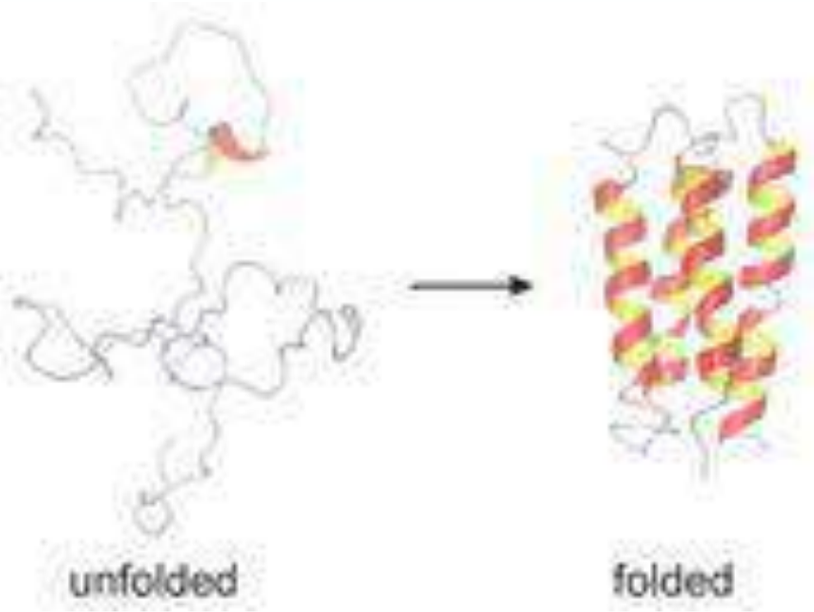
$$P(\boldsymbol{S}) = \frac{\int d\boldsymbol{R}\, \delta(\boldsymbol{S} - \boldsymbol{S}(\boldsymbol{R}))\, e^{-\beta U(\boldsymbol{R})}}{\int d\boldsymbol{R}\, e^{-\beta U(\boldsymbol{R})}}$$

Examples

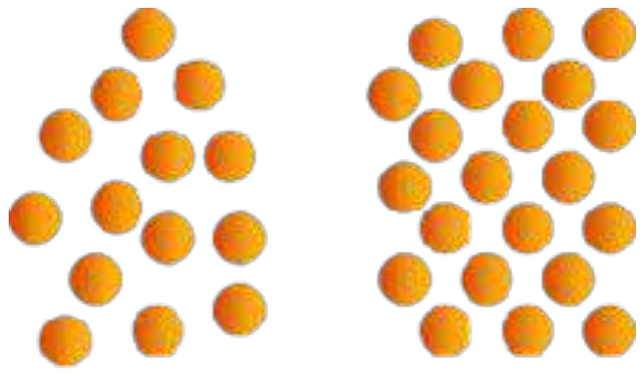
Isomerization:
dihedral angle



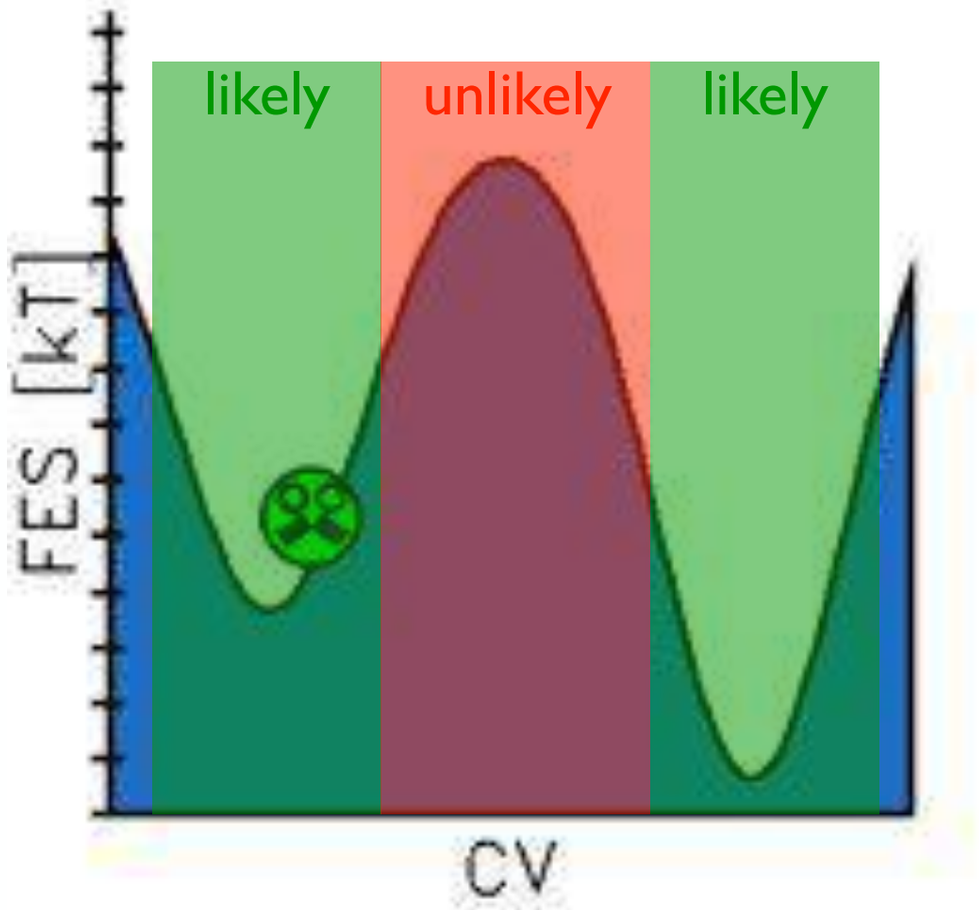
Protein folding:
gyration radius,
number of contacts,
...



Phase transitions:
lattice vectors,
bond order parameters,
...



Rare events simplified



How can we estimate a free energy difference if we never see a transition?

$$F(A) - F(B) = -k_B T \ln \frac{N_A}{N_B}$$

Biased sampling

The idea is to add a bias potential that acts on the collective variables:

$$U(\mathbf{R}) \rightarrow U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))$$

In this biased ensemble the free energy becomes:

$$F'(\mathbf{S}) = -\frac{1}{\beta} \ln P'(\mathbf{S}) + C$$

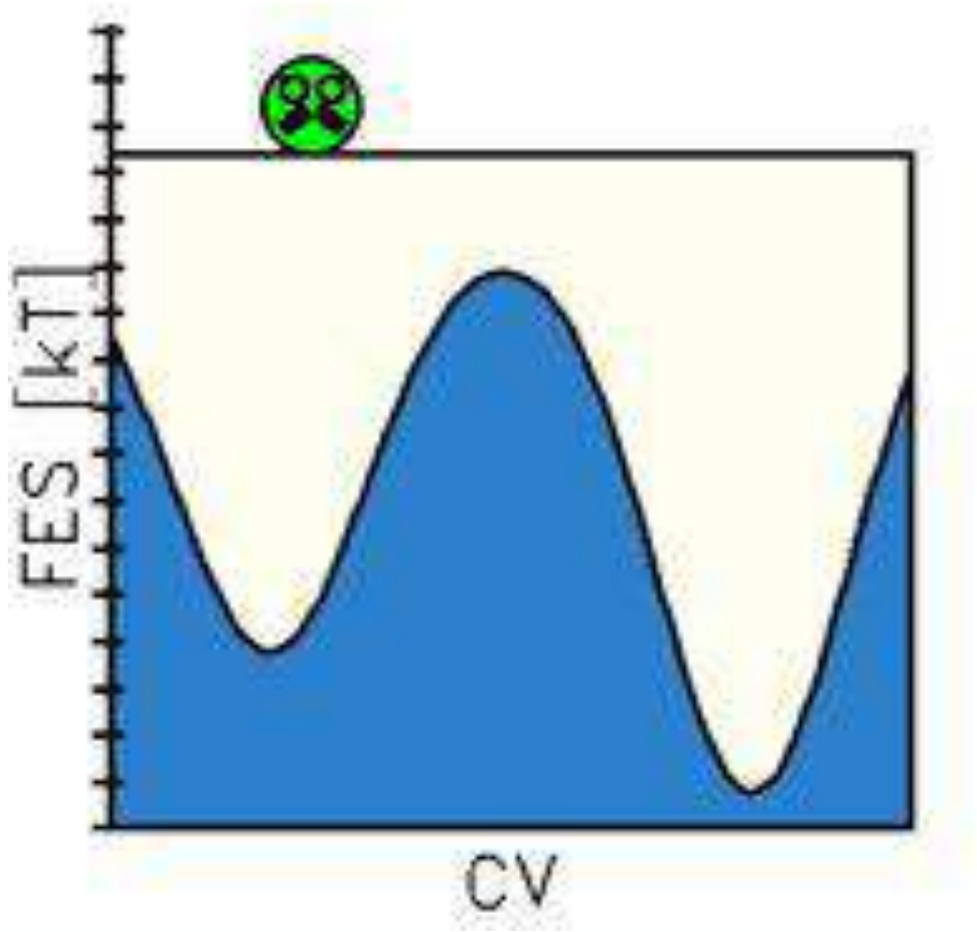
where $$P'(\mathbf{S}) = \frac{\int d\mathbf{R}\, \delta(\mathbf{S} - \mathbf{S}(\mathbf{R}))\, e^{-\beta[U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))]}}{\int d\mathbf{R}\, e^{-\beta[U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))]}}$$

which leads to:

$$F'(\mathbf{S}) = F(\mathbf{S}) + V(\mathbf{S})$$

Umbrella sampling

What is a good choice of bias potential?



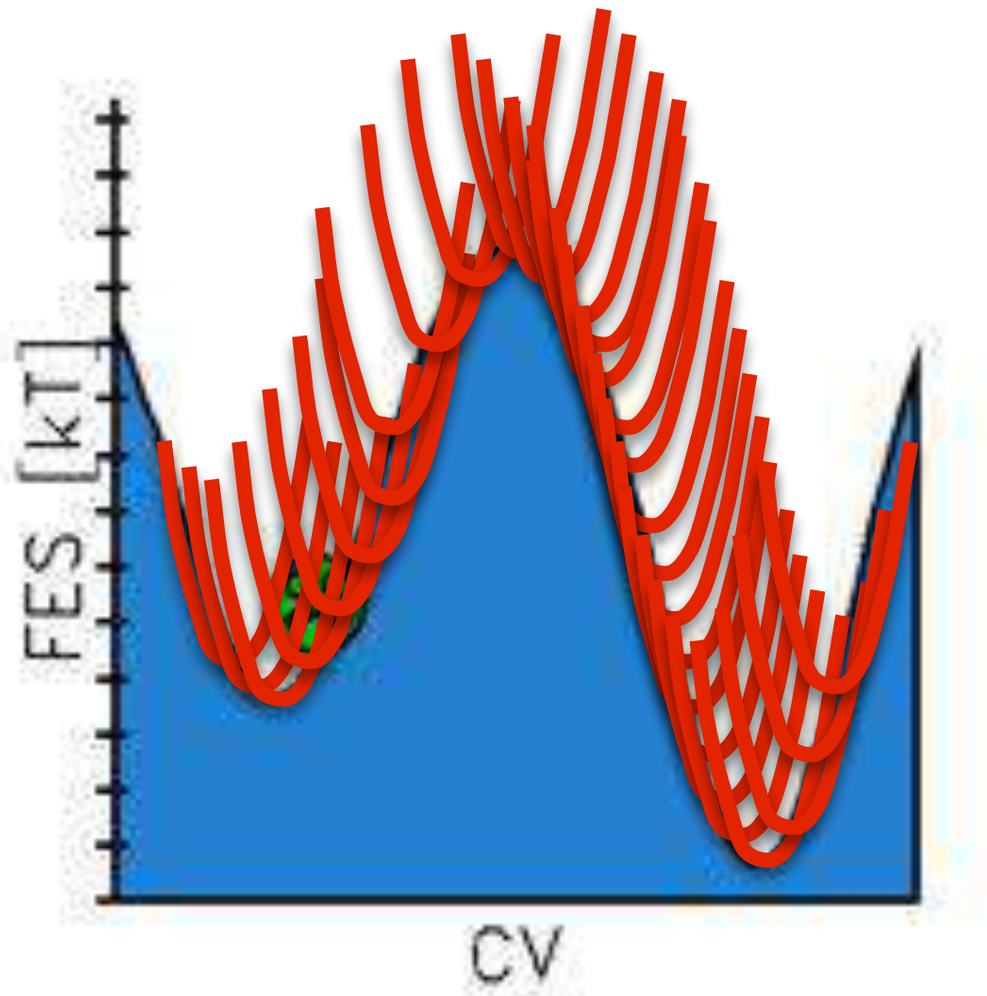
The one that leads to $F'(\mathbf{S}) = 0 \longrightarrow V(\mathbf{S}) = -F(\mathbf{S})$

Let's use an approximation of the free energy as bias potential

Torrie & Valleau JCP 1977

Multiple restraints + WHAM

The idea is to do multiple umbrella sampling calculations using harmonic restraints as bias potentials

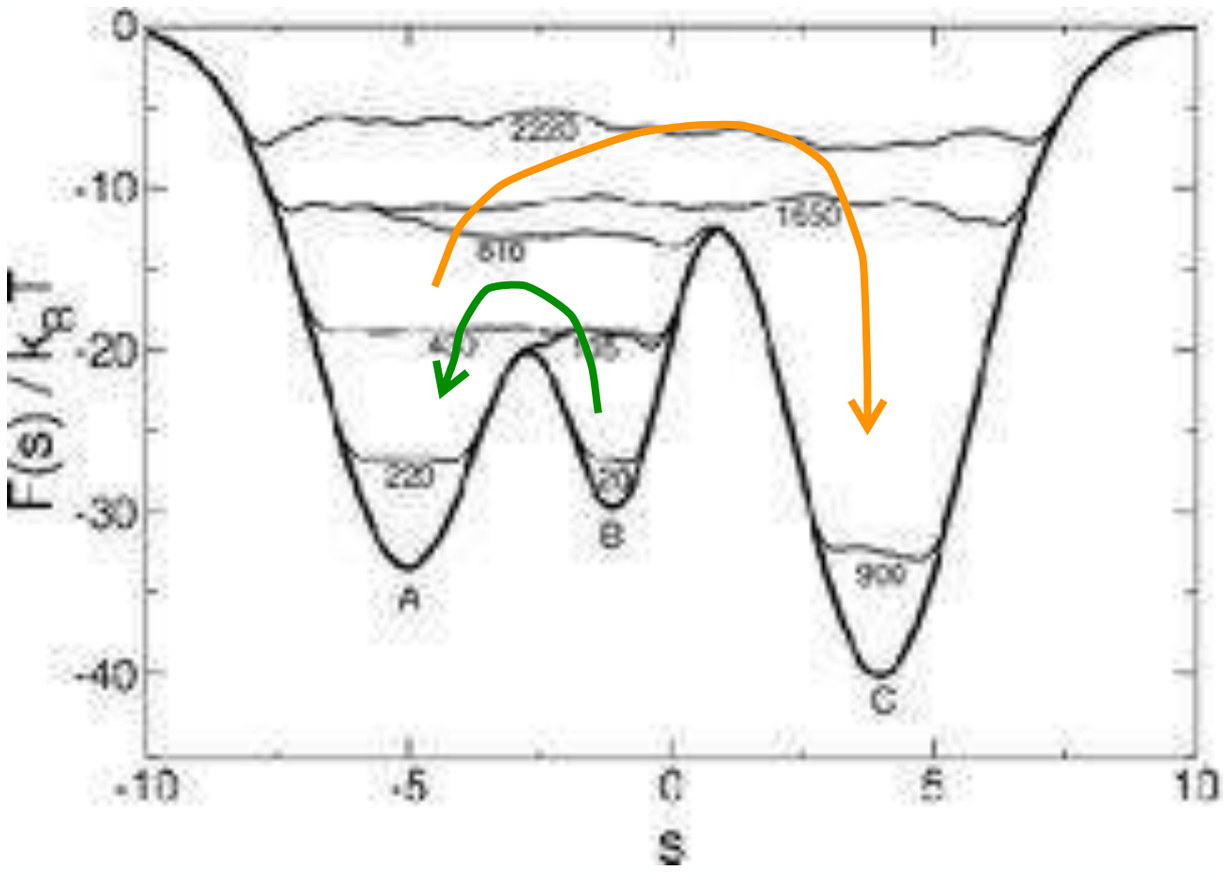


And use WHAM* to merge the biased simulations

*Ferrenberg & Swendsen PRL 1989

Metadynamics

History-dependent bias potential acting on selected degrees of freedom or Collective Variables (CVs)



$$\boldsymbol{S} = (S_1(\boldsymbol{R}), ..., S_d(\boldsymbol{R}))$$

$$V_G(\mathbf{S}, t) = W \sum_{t'=\tau_G, 2\tau_G, \ldots}^{t'<t} exp\left(-\sum_{i=1}^{d} \frac{(S_i - S_i(\mathbf{R}(t')))^2}{2\sigma_i^2}\right)$$

$$V_G(\boldsymbol{S}, t \to \infty) = -F(\boldsymbol{S}) + C$$

Laio & Parrinello PNAS 2002

REVIEW: Barducci, Bonomi, Parrinello WIREs Comput Mol Sci 2011

Pros and Cons

Advantages

- Enhanced sampling along the CVs
- Reconstruction of the FES:

$$V_G(\boldsymbol{S}, t \to \infty) = -F(\boldsymbol{S}) + C$$ Bussi, Laio, Parrinello PRL 2006

- *A priori* knowledge of the landscape not required

Disadvantages

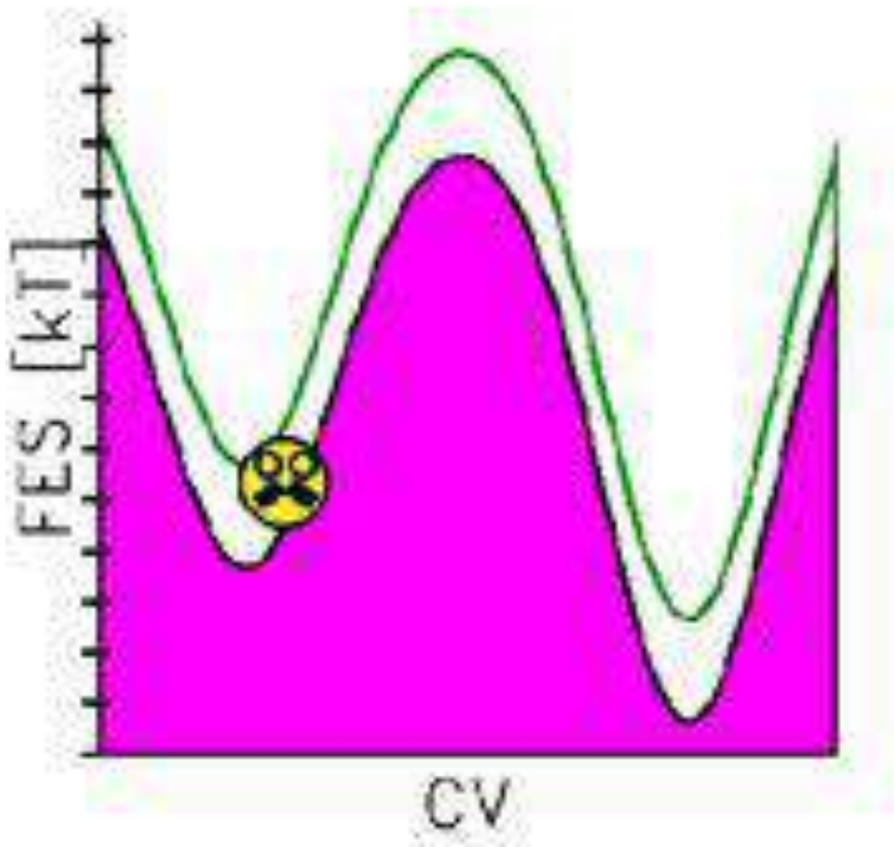
- Lack of convergence in a single run
- Overfilling
- The choice of the CVs is not trivial

Well-Tempered Metadynamics

The initial Gaussian height w_0 is rescaled during the simulation:

$$w = w_0 \, e^{-\frac{V(\boldsymbol{s},t)}{k_B \Delta T}}$$

where $T + \Delta T$ is a fictitious CV temperature.



- Convergence and overfilling issues solved:

$$V(\mathbf{s},t) \rightarrow -\frac{\Delta T}{T + \Delta T} F(\mathbf{s})$$

- ΔT used to tune the extent of exploration

Barducci, Bussi, Parrinello PRL 2008

Choosing the right CVs

A good set of CVs for metadynamics (and other biasing techniques) should:

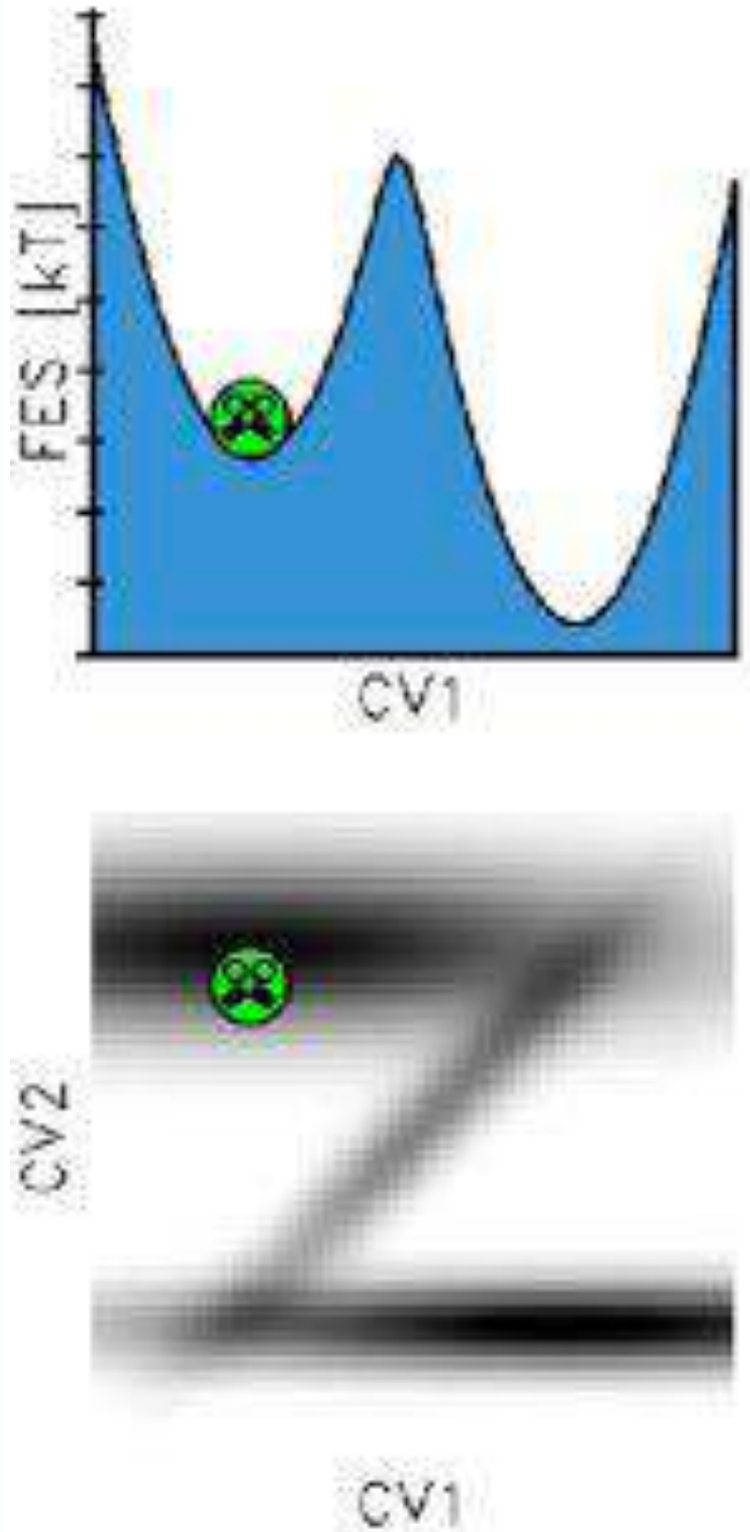
- Discriminate between initial and final states
- Be as small as possible
- Include all the slow modes of a process

Metadynamics is inefficient with a large number of CVs.

Possible strategies:

- devise automatic protocols to find good CVs
- improve metadynamics to deal with a large number of CVs
- couple metadynamics with other methods, such as REM

Hidden degrees of freedom



Bias Exchange Metadynamics

- N replicas at the same temperature T
- Different CVs and bias potentials

- Exchange probability

$$\Delta_{j,k} = \cancel{(\beta_j - \beta_k)(U(\boldsymbol{R}_j) - U(\boldsymbol{R}_k))}$$
$$+ \beta_j [V_G^{(j)}(\boldsymbol{S}(\boldsymbol{R}_j), t) - V_G^{(j)}(\boldsymbol{S}(\boldsymbol{R}_k), t)]$$
$$+ \beta_k [V_G^{(k)}(\boldsymbol{S}(\boldsymbol{R}_k), t) - V_G^{(k)}(\boldsymbol{S}(\boldsymbol{R}_j), t)]$$

Piana and Laio, J.Phys. Chem. B (2007)

Parallel Bias Metadynamics

Biasing a large number of CVs with WTMetaD is inefficient

In PBMetaD we apply multiple low-dimensional bias potentials:

$$V(S_1, t), ..., V(S_N, t)$$

one at a time:

$$P_t(\mathbf{R}, \eta) \propto \exp\left[-\beta\left(U(\mathbf{R}) + \sum_i \eta_i V(S_i, t)\right)\right]$$

where $\eta = (\eta_1, ..., \eta_N)$ switches on and off (and allows updating) one bias potential at a time

Each bias potential converges to the corresponding free energy:

$$V(S_i, t) \to -\frac{\Delta T}{T + \Delta T} F(S_i)$$

Pfaendtner & Bonomi JCTC 2015

Parallel Bias Metadynamics

Since we are not interested in the η-distribution, we can marginalize this variable:

$$P_t(\mathbf{R}) = \int d\eta P_t(\mathbf{R}, \eta) \propto \exp\left[-\beta \left(U(\mathbf{R}) + V_{PB}(\mathbf{S}, t)\right)\right]$$

where:

$$V_{PB}(\mathbf{S}, t) = -\frac{1}{\beta} \log \sum_{i=1}^{N} \exp - \left[\beta V(S_i, t)\right]$$

In order for each bias potential to converge to the corresponding free energy, we need a new rescaling rule:

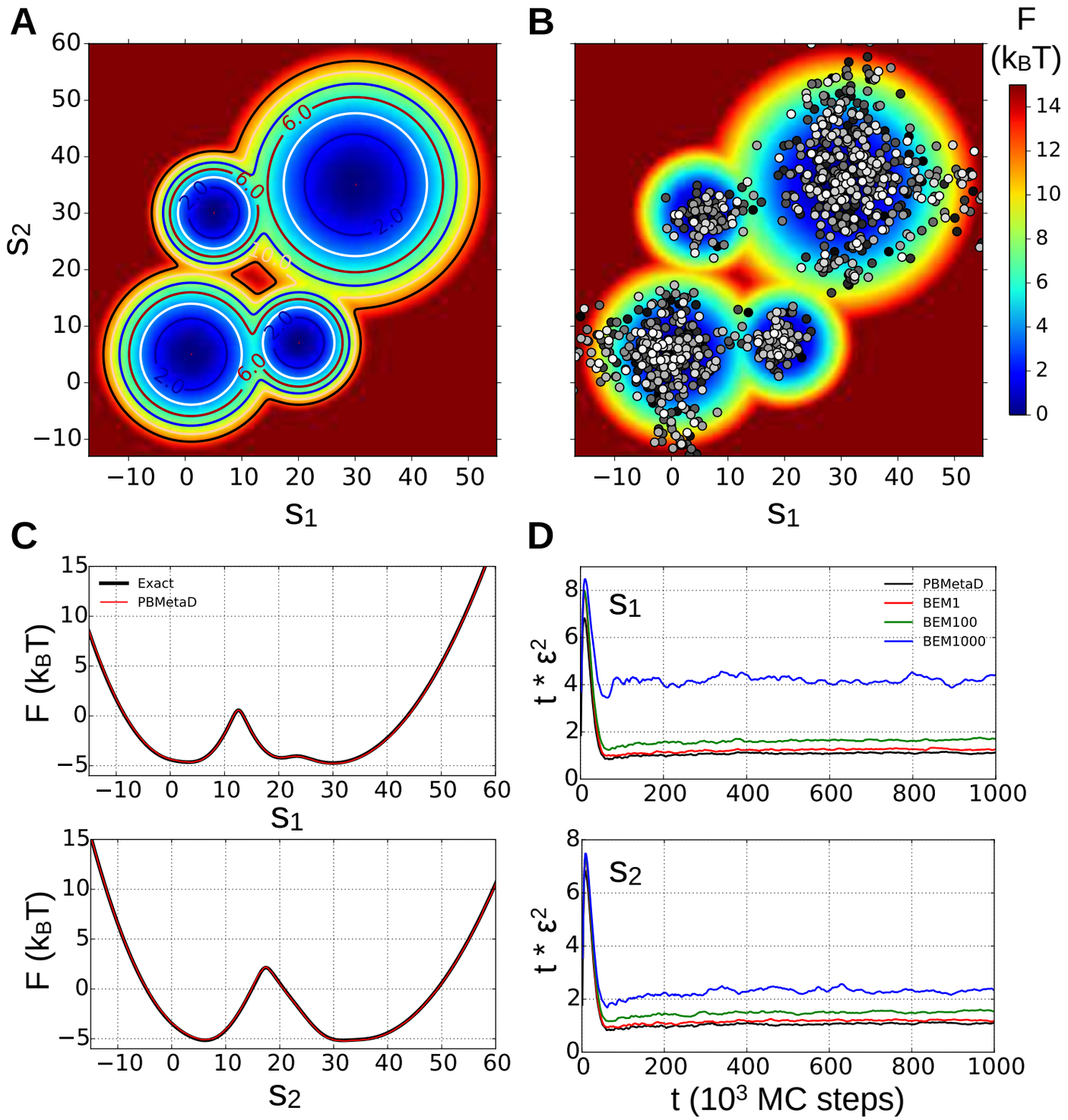
$$\omega_i = \omega_{0,i}\, e^{-\frac{V(S_i,t)}{k_B \Delta T_i}} P(\eta_i = 1|\mathbf{R})$$

where:

$$P(\eta_i = 1|\mathbf{R}) = \frac{\exp\left[-\beta V(S_i, t)\right]}{\sum_{j=1}^{N} \exp\left[-\beta V(S_j, t)\right]}$$

Pfaendtner & Bonomi JCTC 2015

Benchmark on a model system



Pfaendtner & Bonomi JCTC 2015

Convergence and reweighting

We can now demonstrate analytically that the individual bias potentials converge to the correspondent free energies:

$$\beta \widetilde{F}_1(s_1) = -\frac{\gamma V_1(s_1,\tau)}{\Delta T} + \log \frac{\tau \gamma}{\Delta T Z_1}$$

and that we can adapt a reweighting technique developed for well-tempered metadynamics* to recover the unbiased Boltzmann distribution:

$$P(\mathbf{R},t) = e^{-\beta\{V_{PB}(s_1(\mathbf{R}),s_2(\mathbf{R}),t)-c(t)\}}.P_0(\mathbf{R})$$

$$e^{\beta c(t)} \propto \int ds_1 \left[e^{\frac{\gamma V_1(s_1,t+\Delta t)}{\Delta T}} - e^{\frac{\gamma V_1(s_1,t)}{\Delta T}} \right]$$

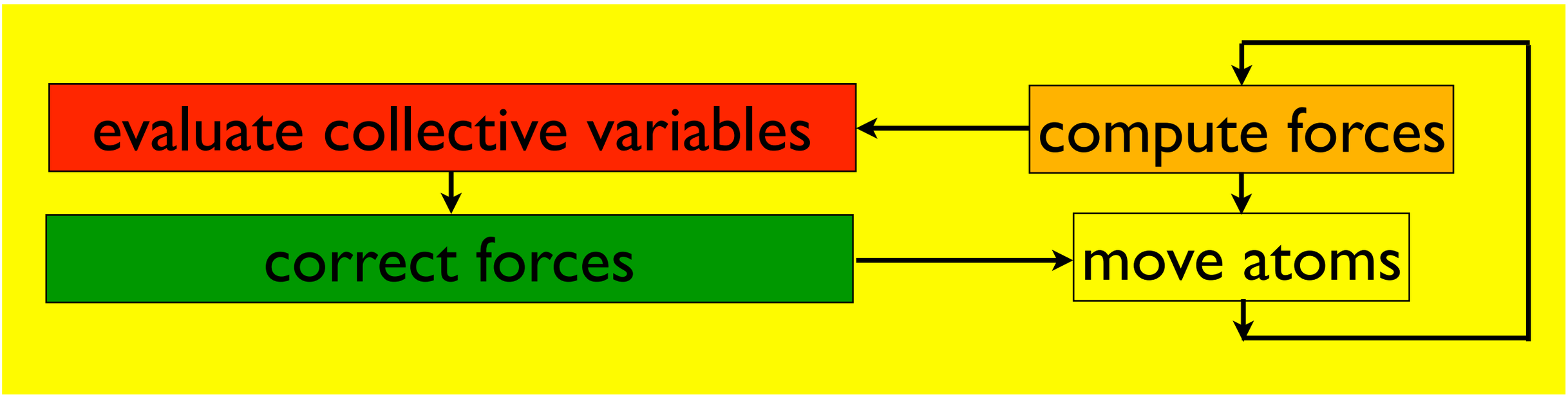
*Tiwary, Parrinello JPCB 2015

Tiwary, Pfaendtner, Bonomi. In preparation

The implementation

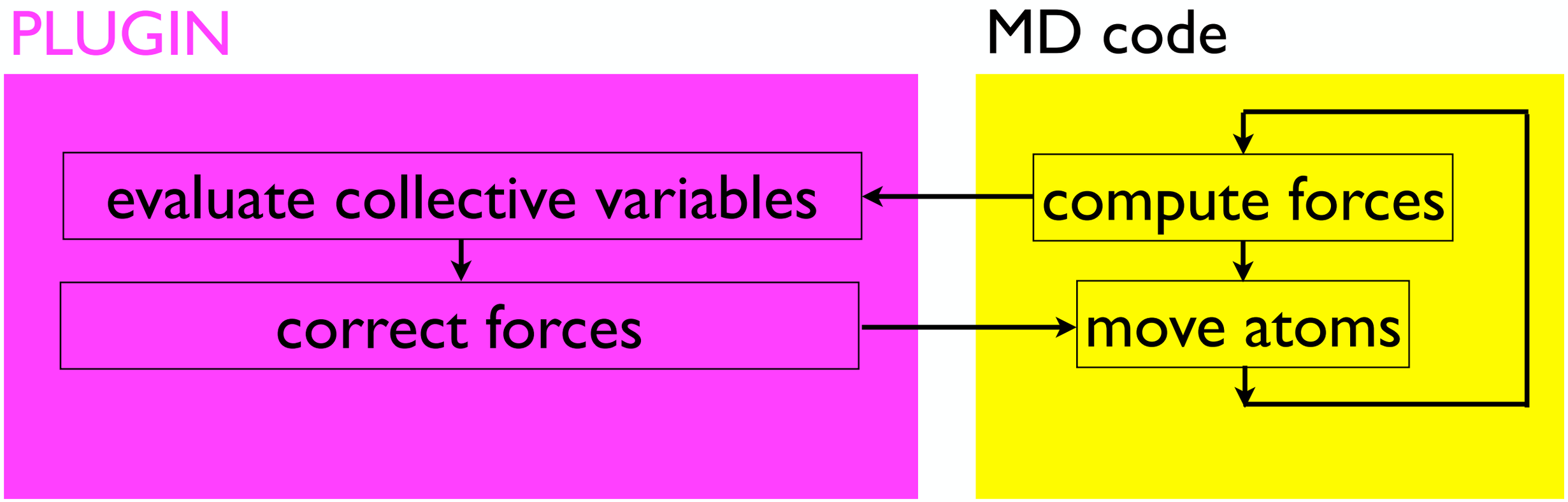
depending on the physical problem:
distances, angles, ...

depending on physical
problem/type of machine/...



several possible algorithms
e.g. umbrella sampling, metadynamics, ...

PLUMED



Bonomi *et al.* CPC 2008
Tribello *et al.* CPC 2014

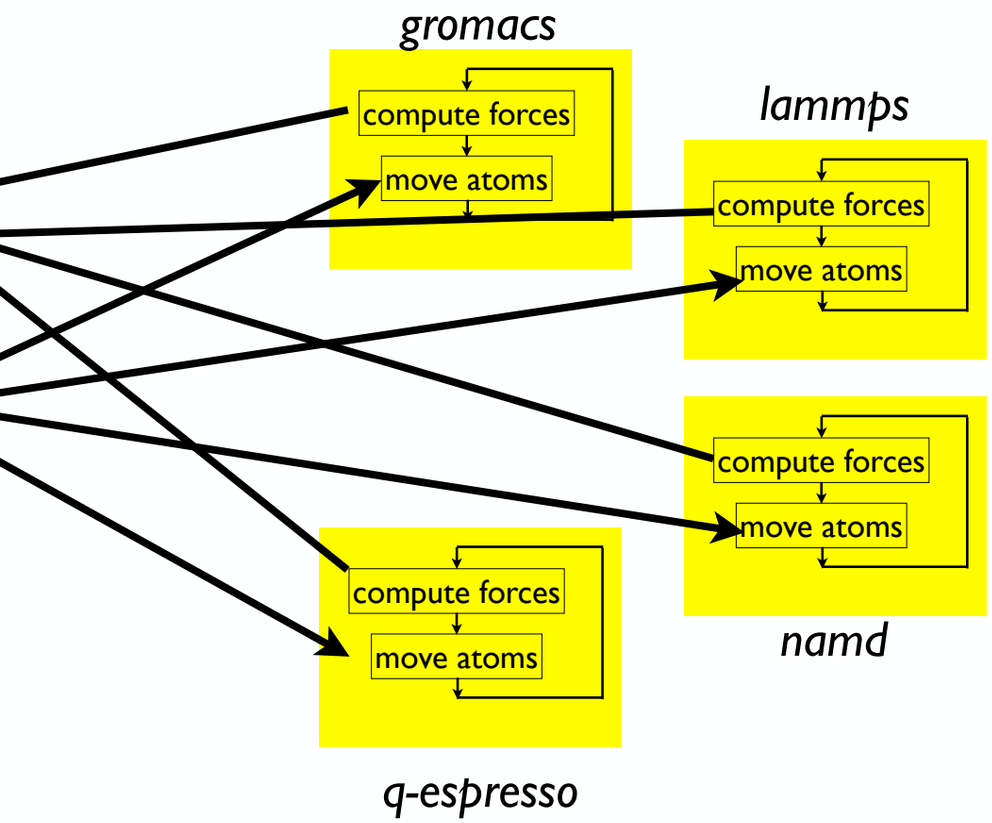
PLUMED

PLUGIN

evaluate collective variables

correct forces

MD codes



One open source pluging
for several MD codes!

Why **PLUMED**?

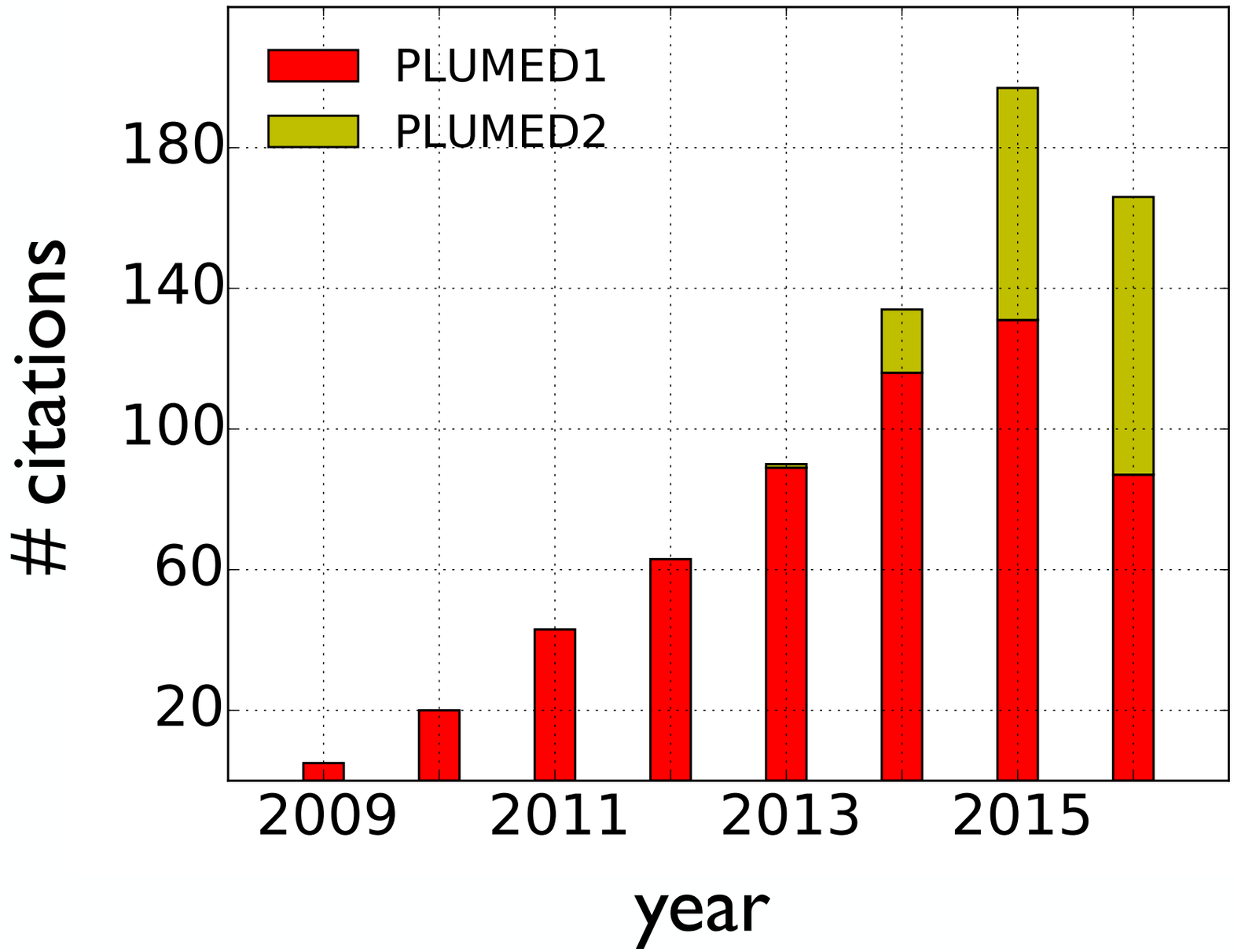
PLUgin for **ME**ta**D**ynamics

PLUgin for free-energy **ME**tho**D**s

PLUgin for **M**ol**E**cular **D**ynamics

Bonomi *et al.* CPC 2008
Tribello *et al.* CPC 2014

A quickly growing community



PLUMED1 = Bonomi *et al.* CPC 2008
PLUMED2 = Tribello *et al.* CPC 2014

Source: Google Scholar (Sep 2016)

What can you do with PLUMED?

Analyze trajectories[$]

```
                    # using plumed as a standalone tool
plumed driver --igro traj.gro --plumed plumed.dat
```

Analyze simulations on the fly*

```
# e.g. using gromacs:
mdrun -plumed plumed.dat
```

Bias simulations on the fly*

```
# e.g. using gromacs:
mdrun -plumed plumed.dat
```

[$]from command line or from VMD - Giorgino, CPC (2014), http://github.com/tonigi/vmd_plumed

*used in combination with a supported MD engine, e.g. GROMACS, NAMD, LAMMPS, Q-ESPRESSO, AMBER + others

PLUMED+MD

PLUMED

MD code

read from a separate file

initialization

evaluate collective variables

correct forces

initialization

compute forces

move atoms

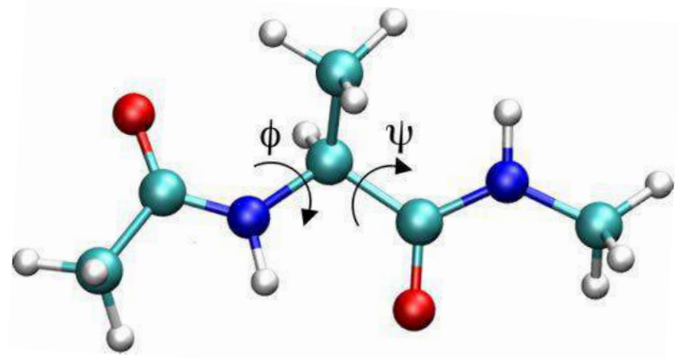
also derivatives w.r.t. atom positions

sometime using history-dependent schemes

Example of PLUMED input file

CV

```
# collective variables definition
phi:    TORSION ATOMS=5,7,9,15
psi:    TORSION ATOMS=7,9,15,17
```



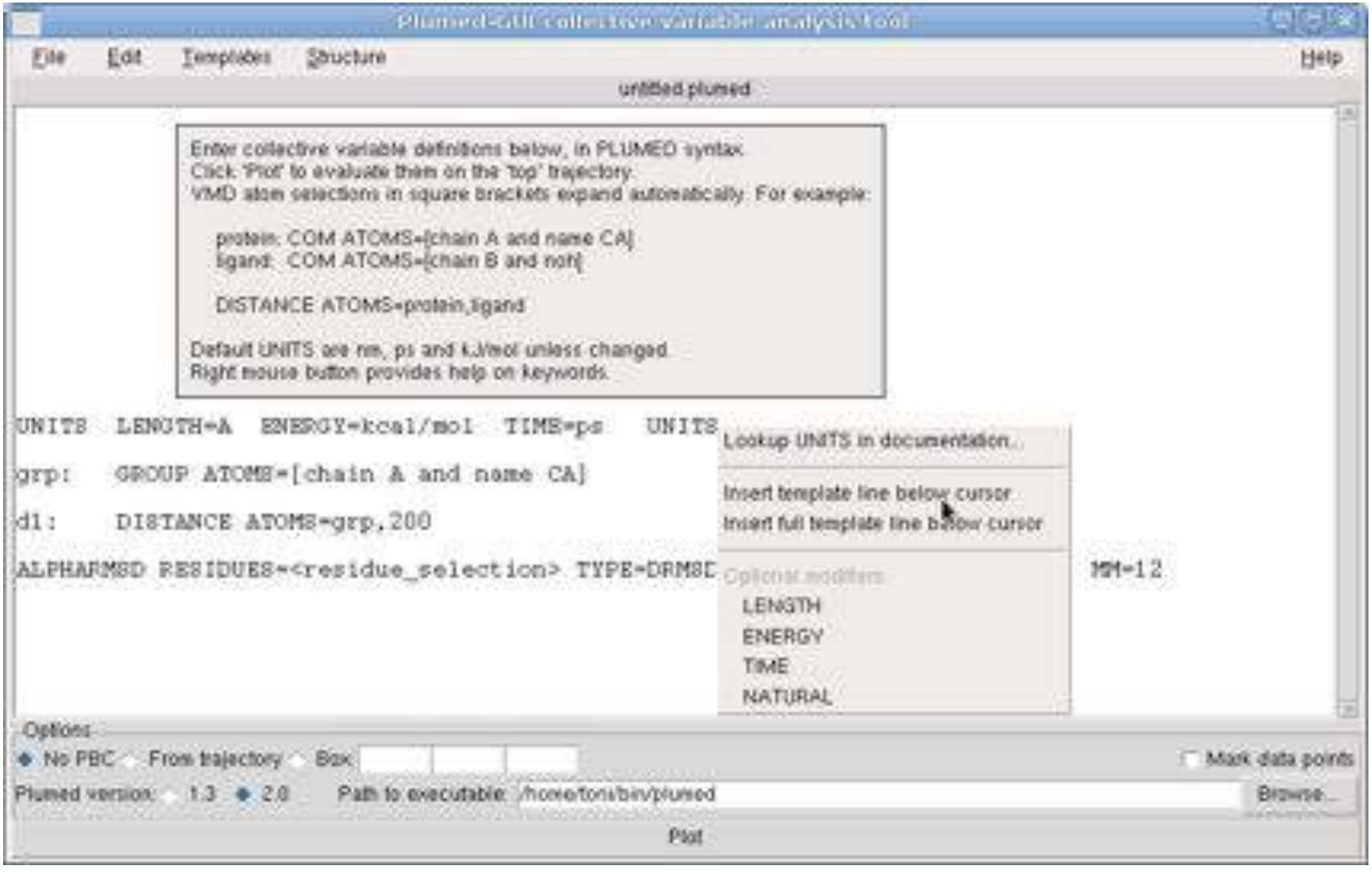
BIAS

```
# activate Parallel Bias Metadynamics
PBMETAD ...
 ARG=phi,psi
 PACE=500 HEIGHT=1.2
 SIGMA=0.35,0.35
 FILE=HILLS_PHI,HILLS_PSI
 BIASFACTOR=8.0
... PBMETAD
```

OUTPUT

```
# printout
PRINT ARG=phi,psi,pbmetad.bias STRIDE=500 FILE=COLVAR
```

PLUMED + VMD (GUI)



http://www.ks.uiuc.edu/Research/vmd

Giorgino, CPC (2014) - see http://github.com/tonigi/vmd_plumed

MD codes supported

GROMACS - fast, tuned for biomolecules, open source
NAMD - fast, tuned for biomolecules, scalable
LAMMPS - very general and scalable, open source
QuantumESPRESSO - DFT, open source
AMBER/sander, many force methods (QMMM, semi-empirical,...)

+ some code has PLUMED support out-of-the-box
  - CP2K
  - ESPResSo
  - PINY-MD
  - IPHIGENIE

PLUMED is a library with a documented API
thus, you can easily add your own code!

http://www.gromacs.org
http://www.ks.uiuc.edu/Research/namd
http://lammps.sandia.gov
http://www.quantum-espresso.org
http://ambermd.org

On the WEB

Website: http://www.plumed.org/

Github: http://github.com/plumed/plumed2

User & developer mailing lists

User & developer manuals + tutorials

Conclusions

MD simulations suffer from limitations in sampling capabilities and accuracy of empirical force fields

A wide variety of enhanced sampling methods are based on the idea of adding a bias potentials on selected degrees of freedom, or Collective Variables (CVs)

Metadynamics is a powerful enhanced sampling method, but its efficiency does not optimally scale with the number of CVs used

PBMetaD is an efficient way to enhance sampling using a large number of CVs

PLUMED is a open source library:

- to analyze MD simulations, on-the-fly and a posteriori
- to bias MD simulations and accelerate sampling
- compatible with many popular MD codes

Acknowledgments



UNIVERSITY of WASHINGTON

Jim Pfaendtner

Carlo Camilloni
Andrea Cavalli
Gabi Heller
Francesco Aprile
Michele Vendruscolo

Institut Pasteur

Samuel Hanot
Riccardo Pellarin

Paolo Arosio
Thomas Muller
Tuomas Knowles

PLUMED

Davide Branduardi
Giovanni Bussi
Gareth Tribello

All of you for your attention!