Metadynamics Metainference with PLUMED

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:
  - metadynamics
  - recent developments

- Combining simulations with experiments: Metainference

- Addressing sampling and accuracy issues: M&M

- The open source library PLUMED

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

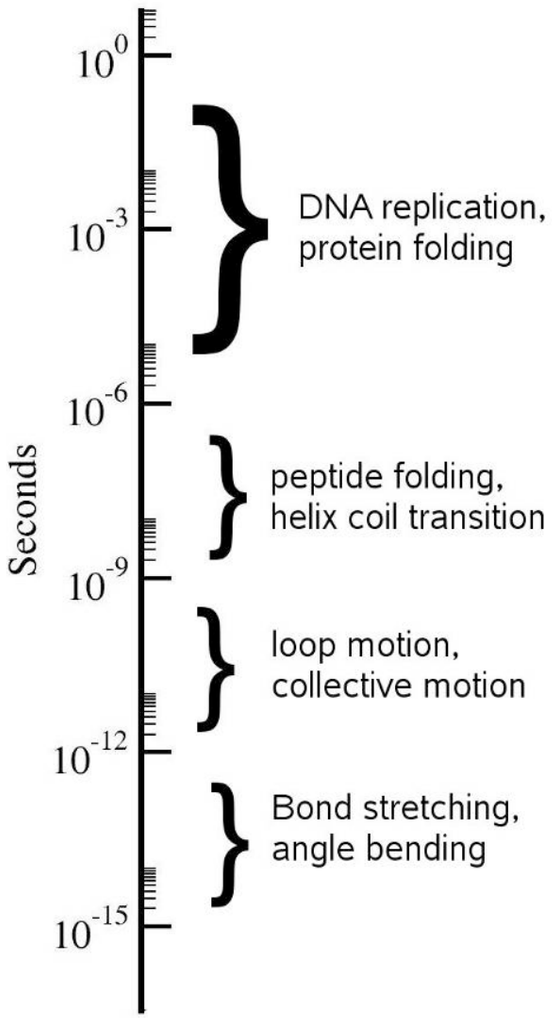
- Higher accuracy calculations
- Fitting experimental observables

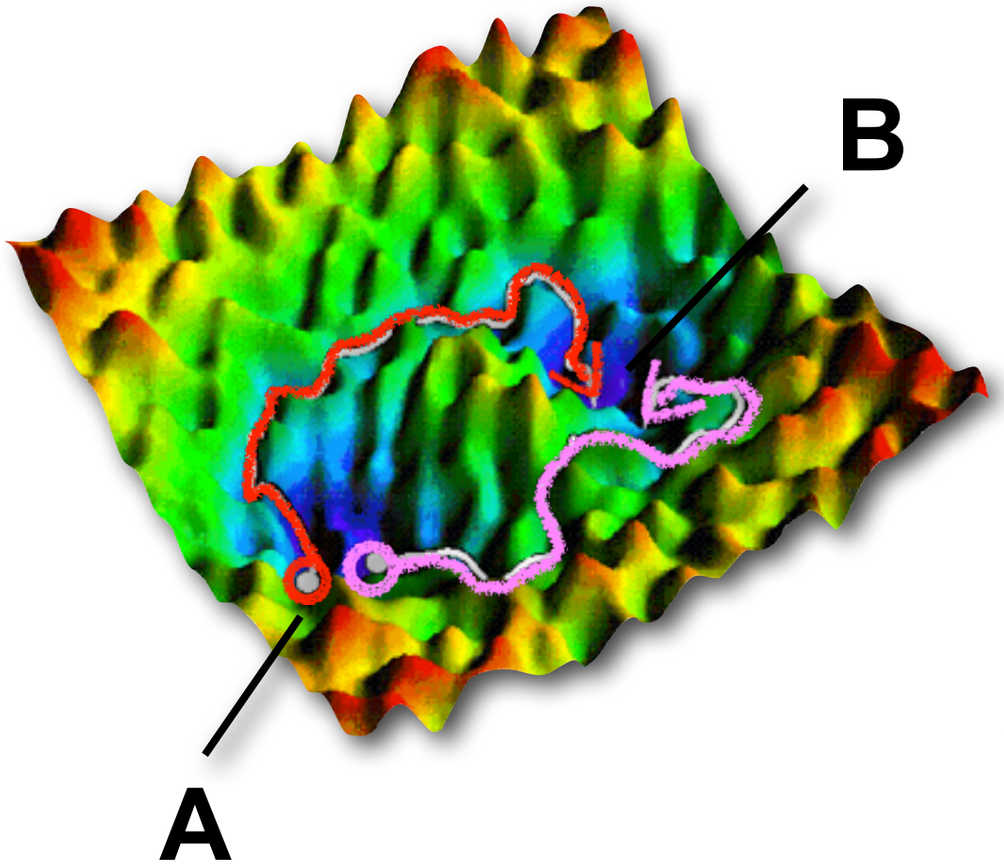
Limitations:

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

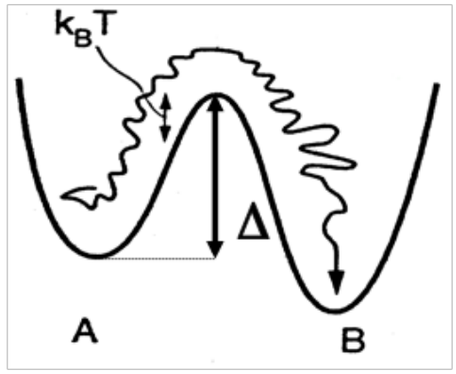
The time scale problem

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:

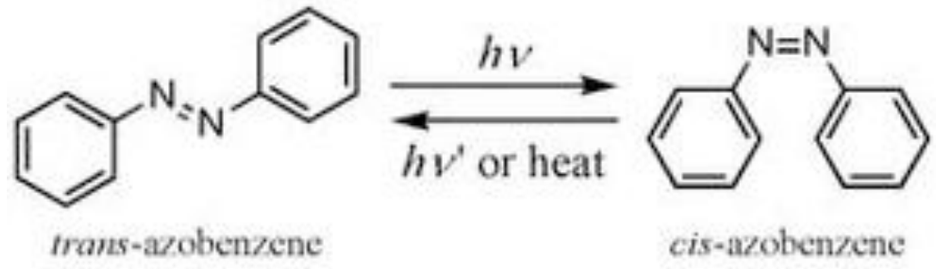
$$F(\boldsymbol{S}) = -\frac{1}{\beta} \ln P(\boldsymbol{S}) \quad \text{where} \quad \beta = \frac{1}{k_B T}$$

canonical ensemble

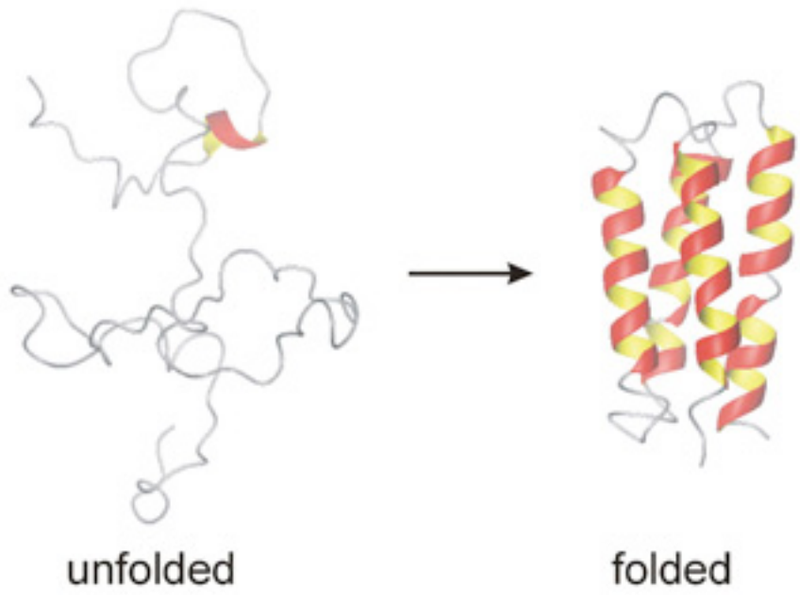
$$P(\boldsymbol{S}) = \frac{\int d\boldsymbol{R}\, \delta(\boldsymbol{S} - \boldsymbol{S}(\boldsymbol{R}))\, e^{-\beta U(\boldsymbol{R})}}{\int d\boldsymbol{R}\, e^{-\beta U(\boldsymbol{R})}}$$

Examples

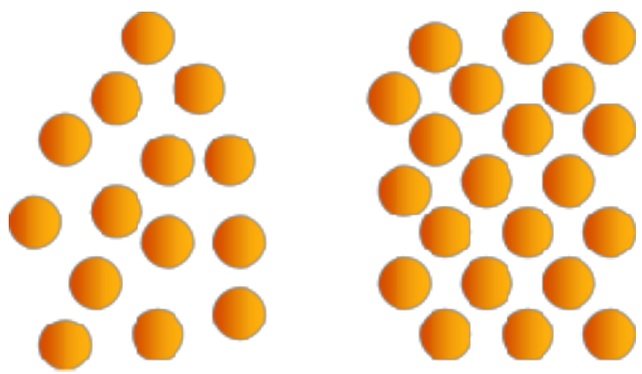
Isomerization:
dihedral angle



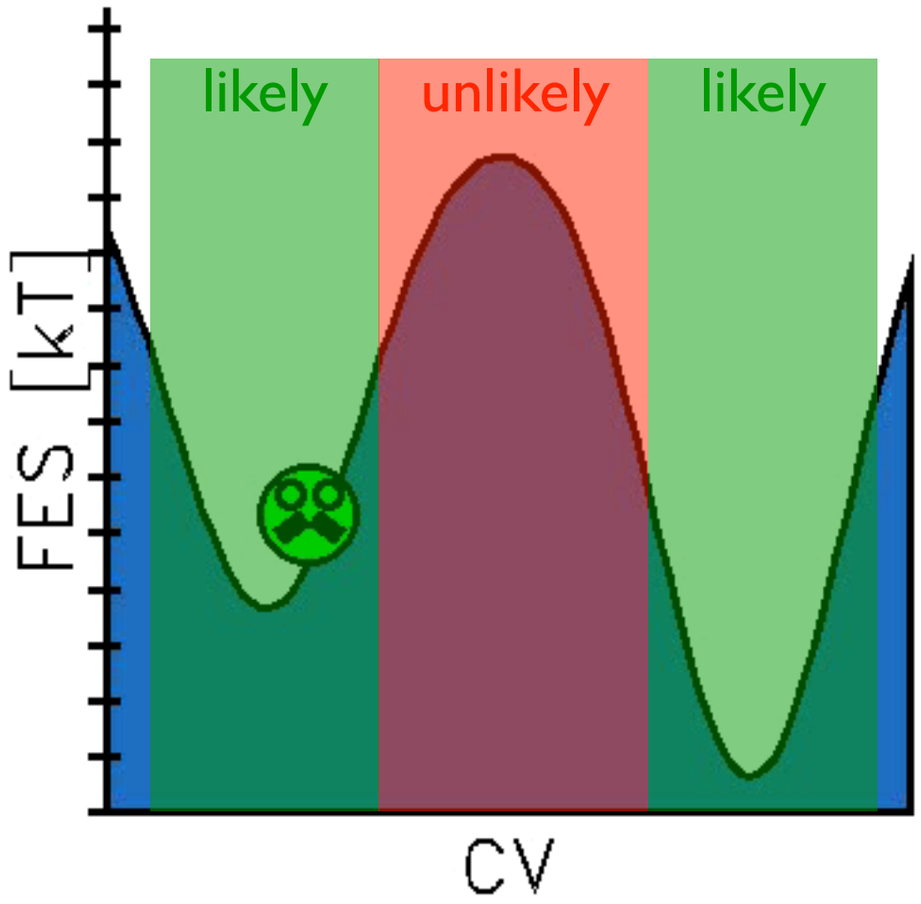
Protein folding:
gyration radius,
number of contacts,
...



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



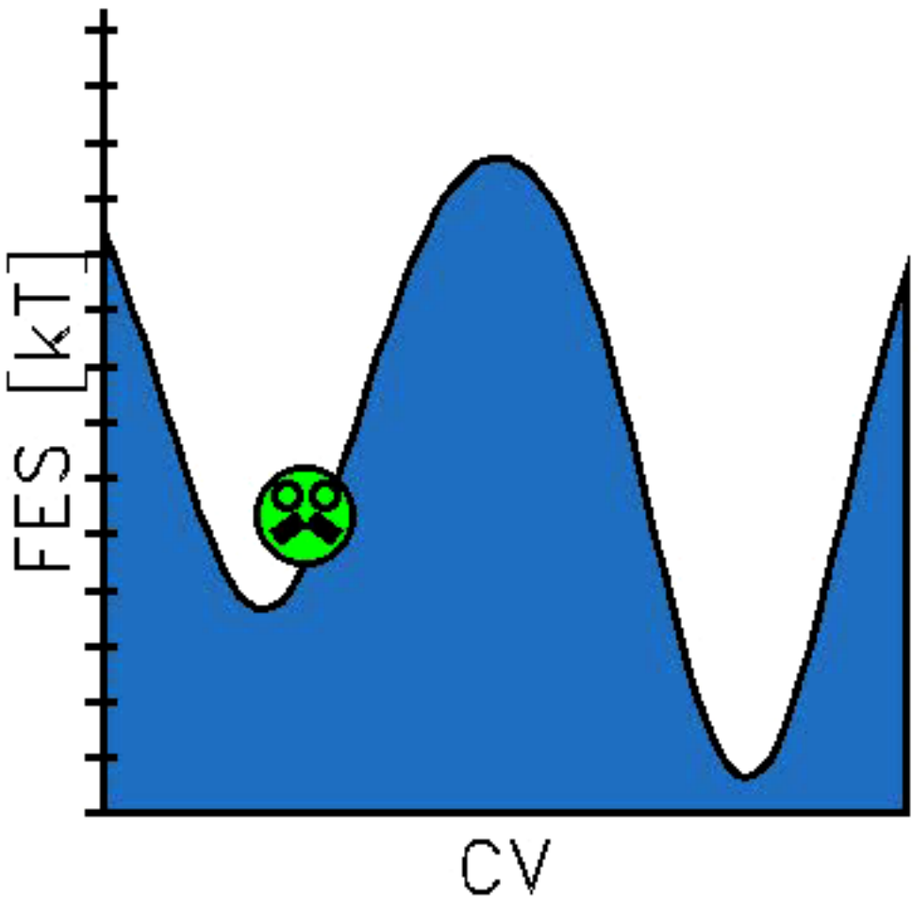
Rare events simplified



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

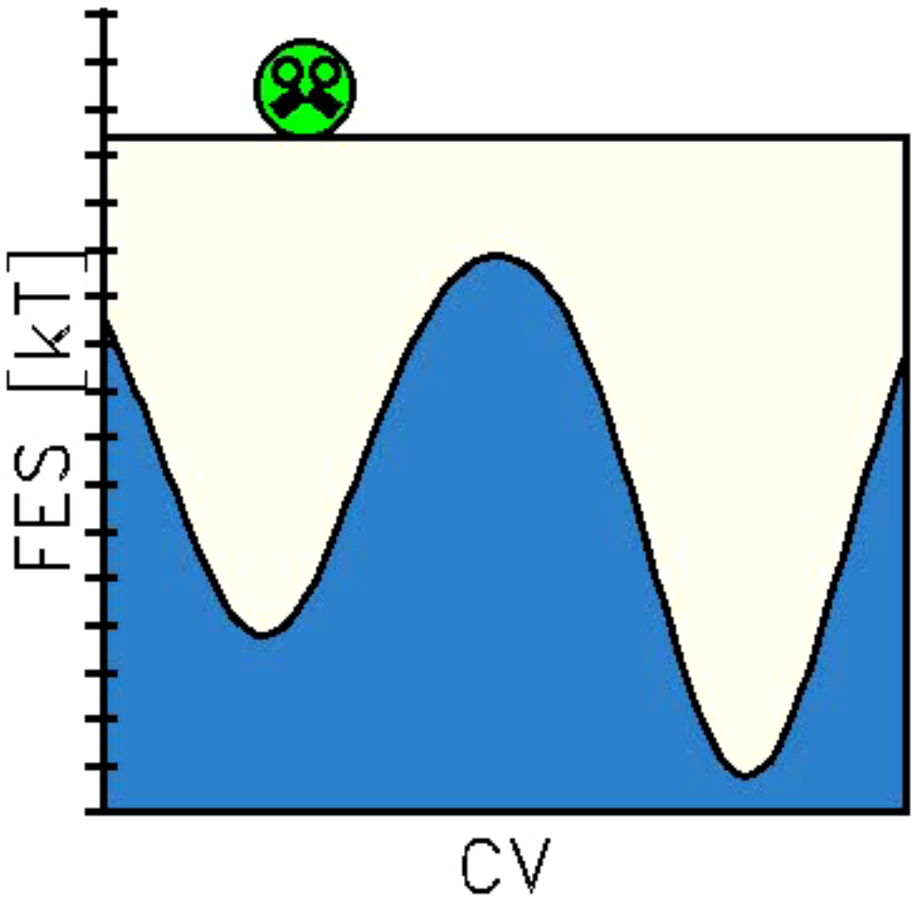


The idea is to add a bias potential that acts on the collective variables:

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

<u>What is a good choice of bias potential?</u>

Biased sampling



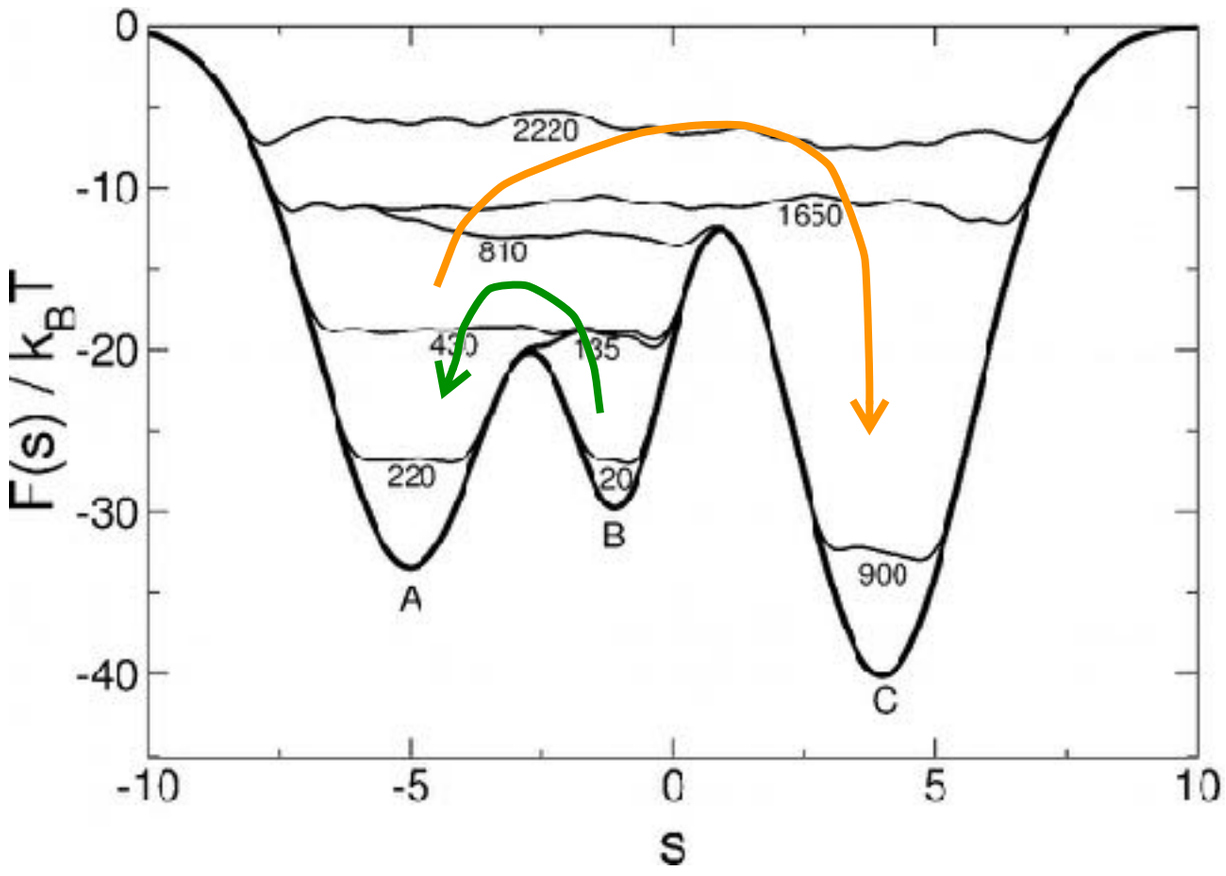
The idea is to add a bias potential that acts on the collective variables:

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

What is a good choice of bias potential?

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

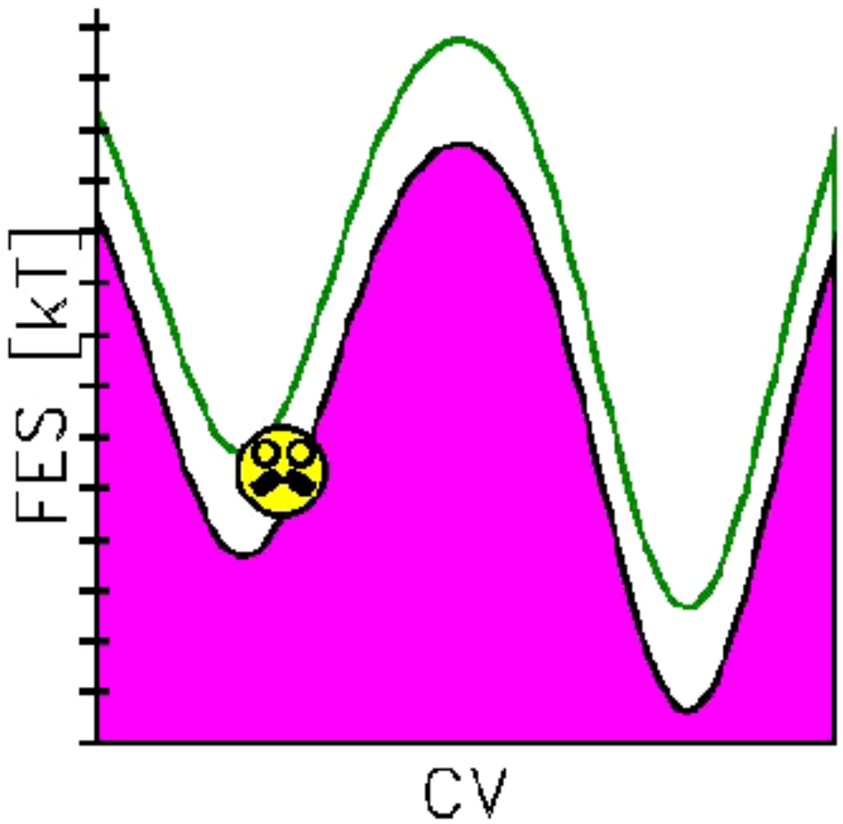
- Lack of convergence in a single run
- Overfilling
- The choice of the CVs is not trivial

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:

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

- 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

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:

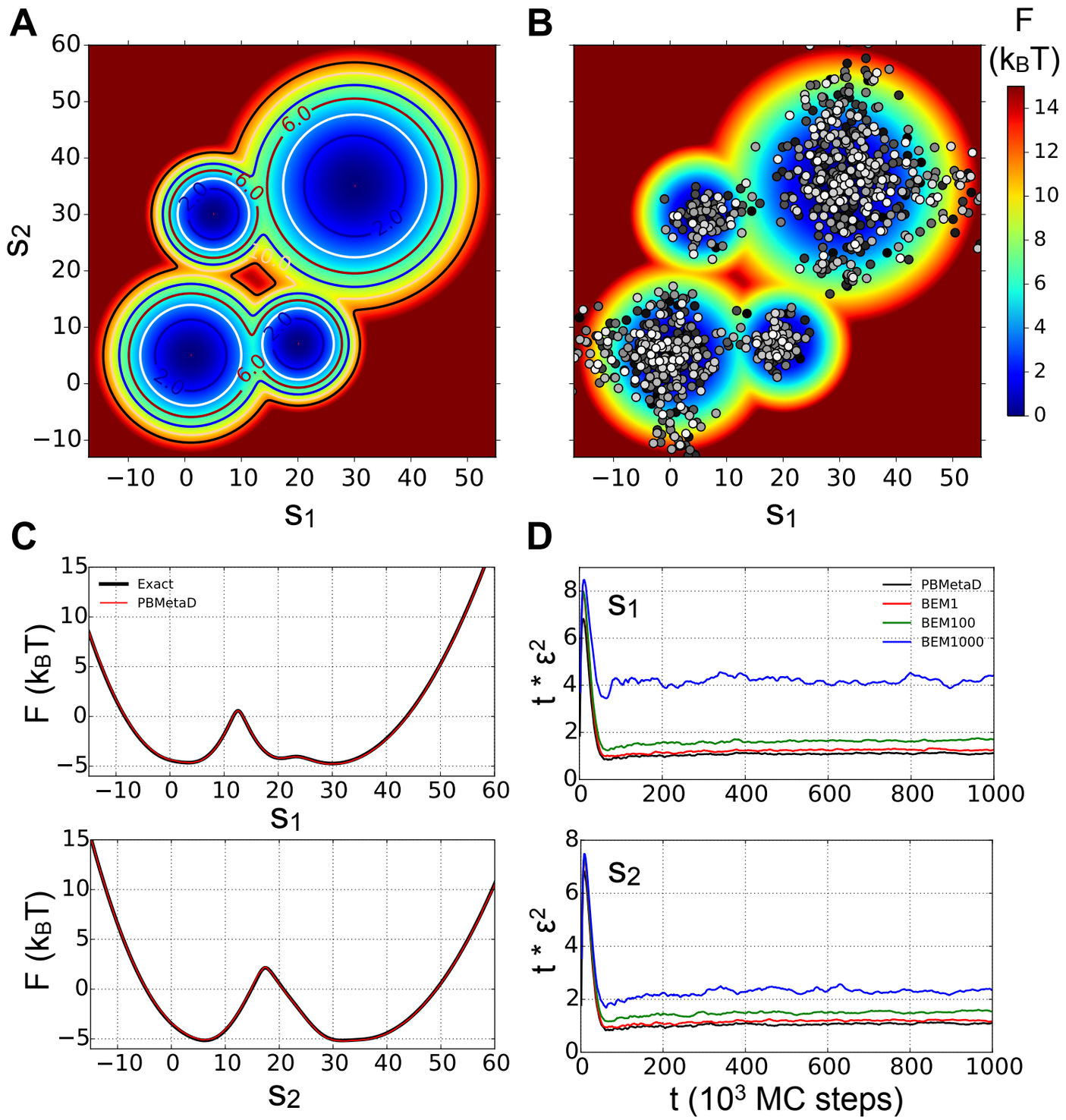
$$\omega_i = \omega_{0,i}\, e^{-\frac{V(S_i,t)}{k_B \Delta T_i}}\, P(\eta_i = 1|\mathbf{R})$$

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

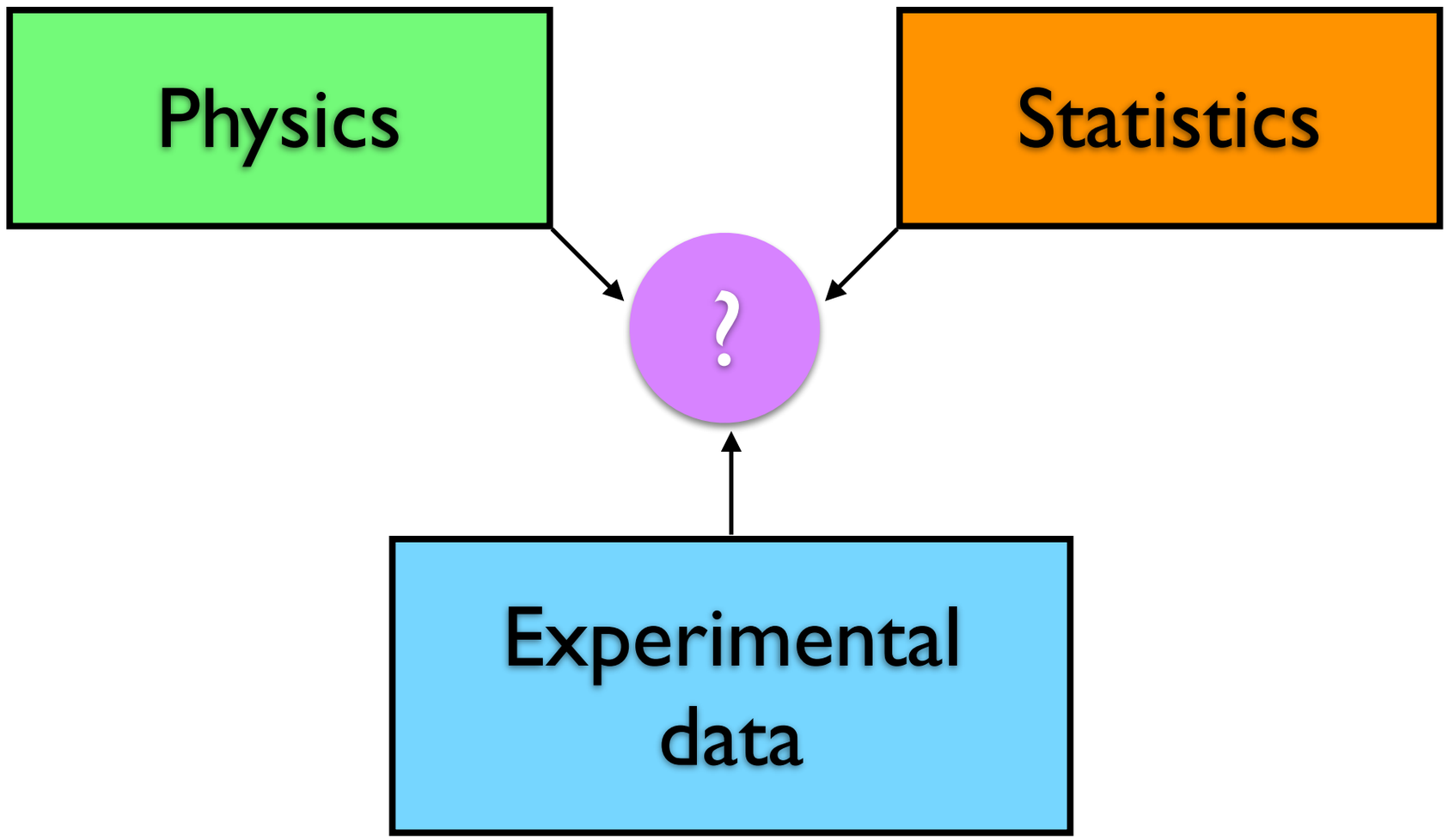
Benchmark on a model system



Pfaendtner & Bonomi JCTC 2015

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

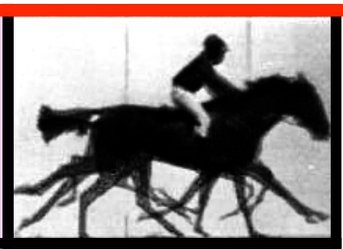


How can we properly combine them?

The challenges of data modeling

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

Addressing these challenges

Mixing	
Random	
Systematic	
Forward model	

Metainference*

*Bonomi *et al.* Science Advances 2016

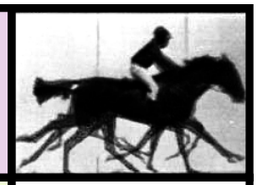
To produce ensemble of models

and determine their populations

a) Input

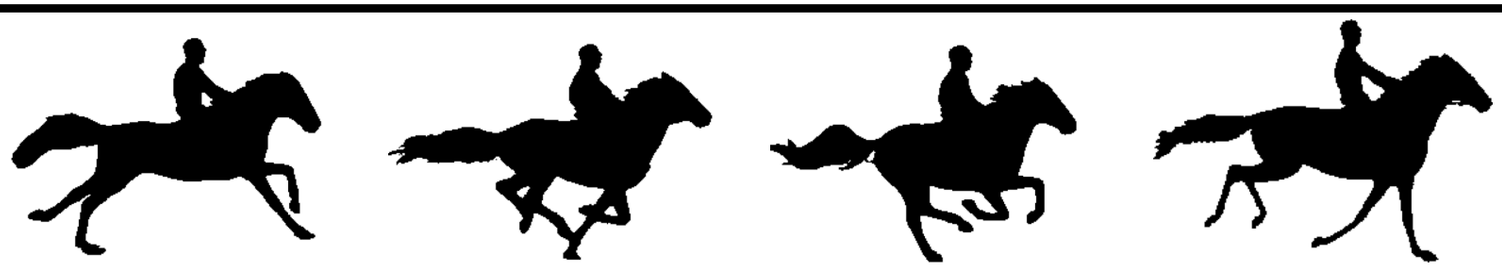
Measurement+Prior

b) Errors

Mixing	
Random	
Systematic	
Forward model	

c) Output

Metainference: Ensemble of models



30% 20% 45% 5%

Metainference

Inspired by replica-averaged modelling, we consider a finite sample of the distribution of models (N replicas):

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

The Metainference energy function (or score) is:

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

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

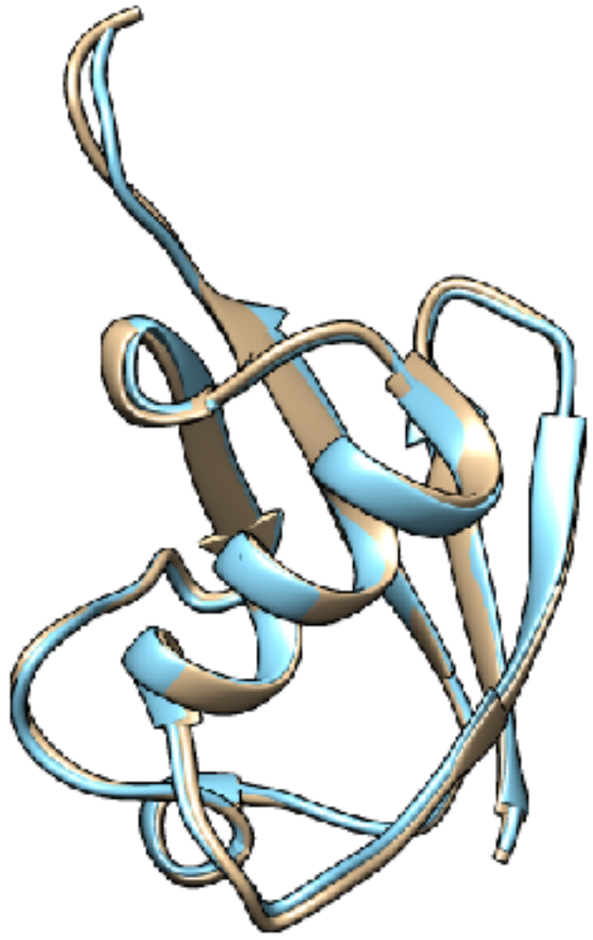
Integrative Dynamical Biology

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

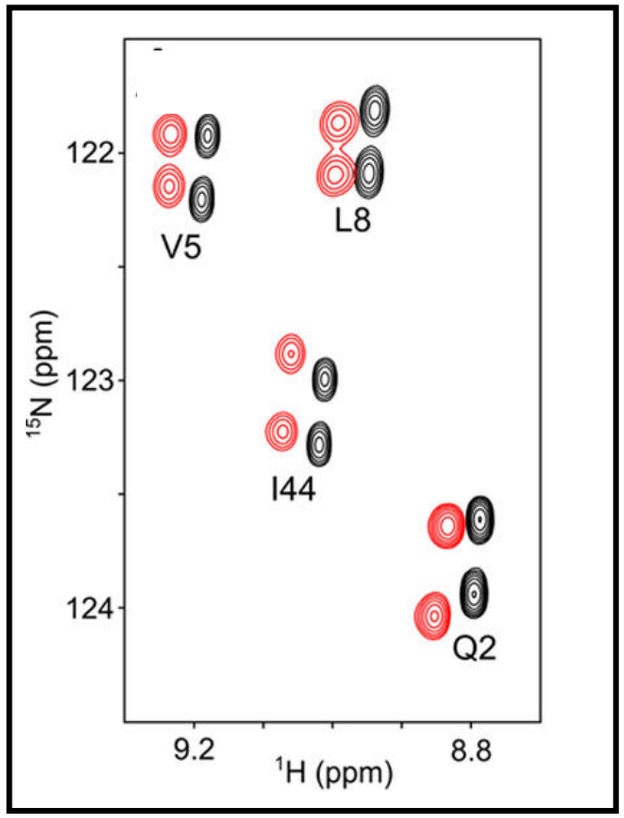
and with the ensemble generated by standard MD

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

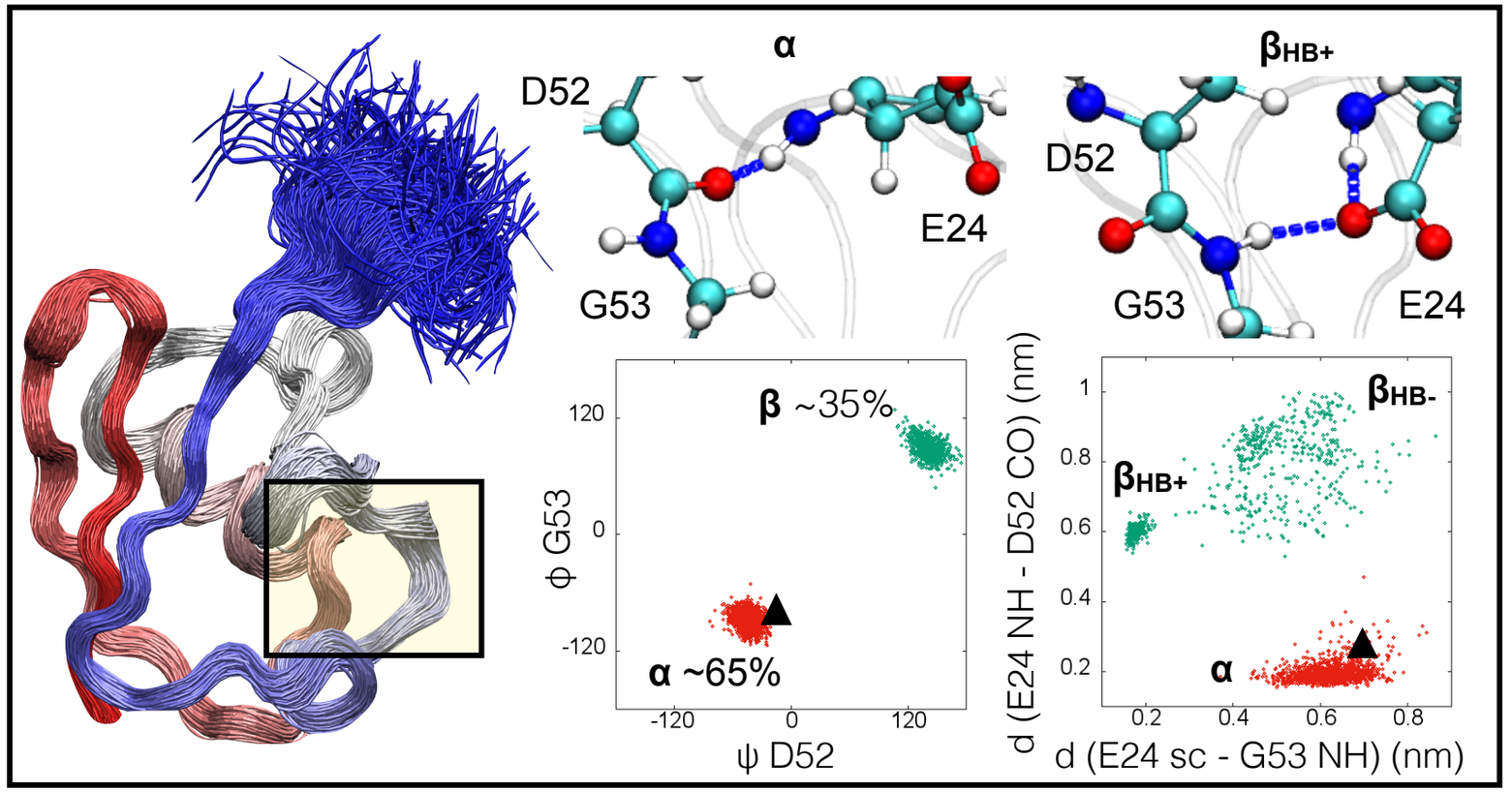
Bonomi *et al.* Science Advances 2016

Ubiquitin ensembles

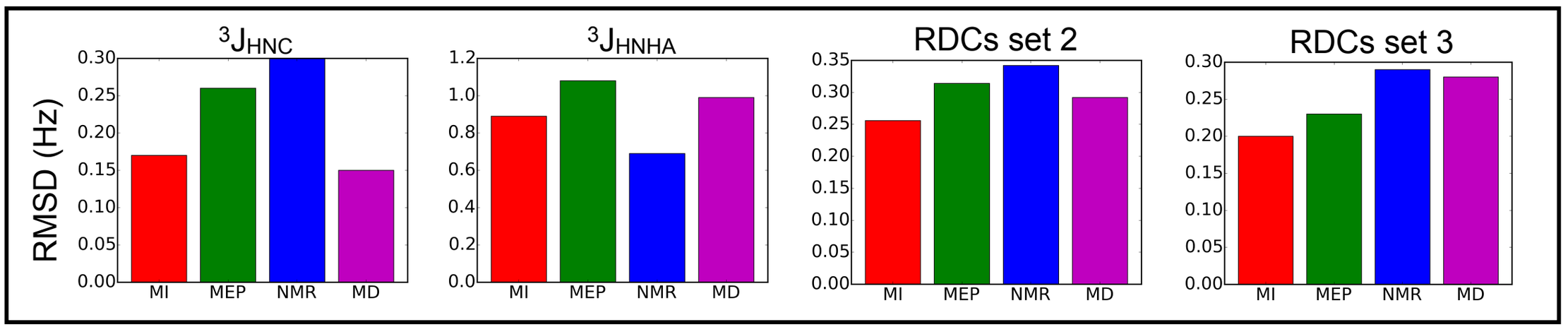
a) Input



b) Ensemble

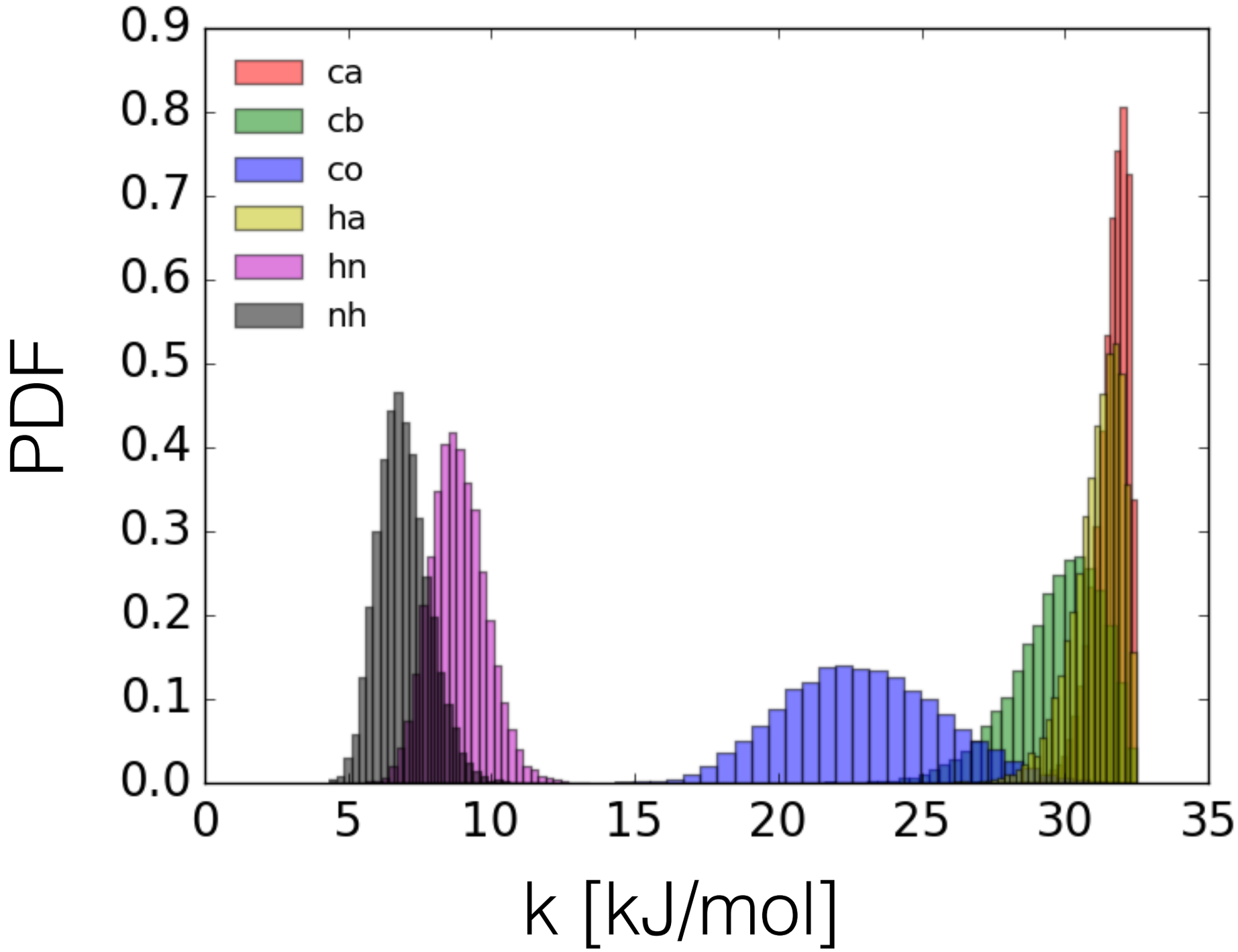


c) Validation



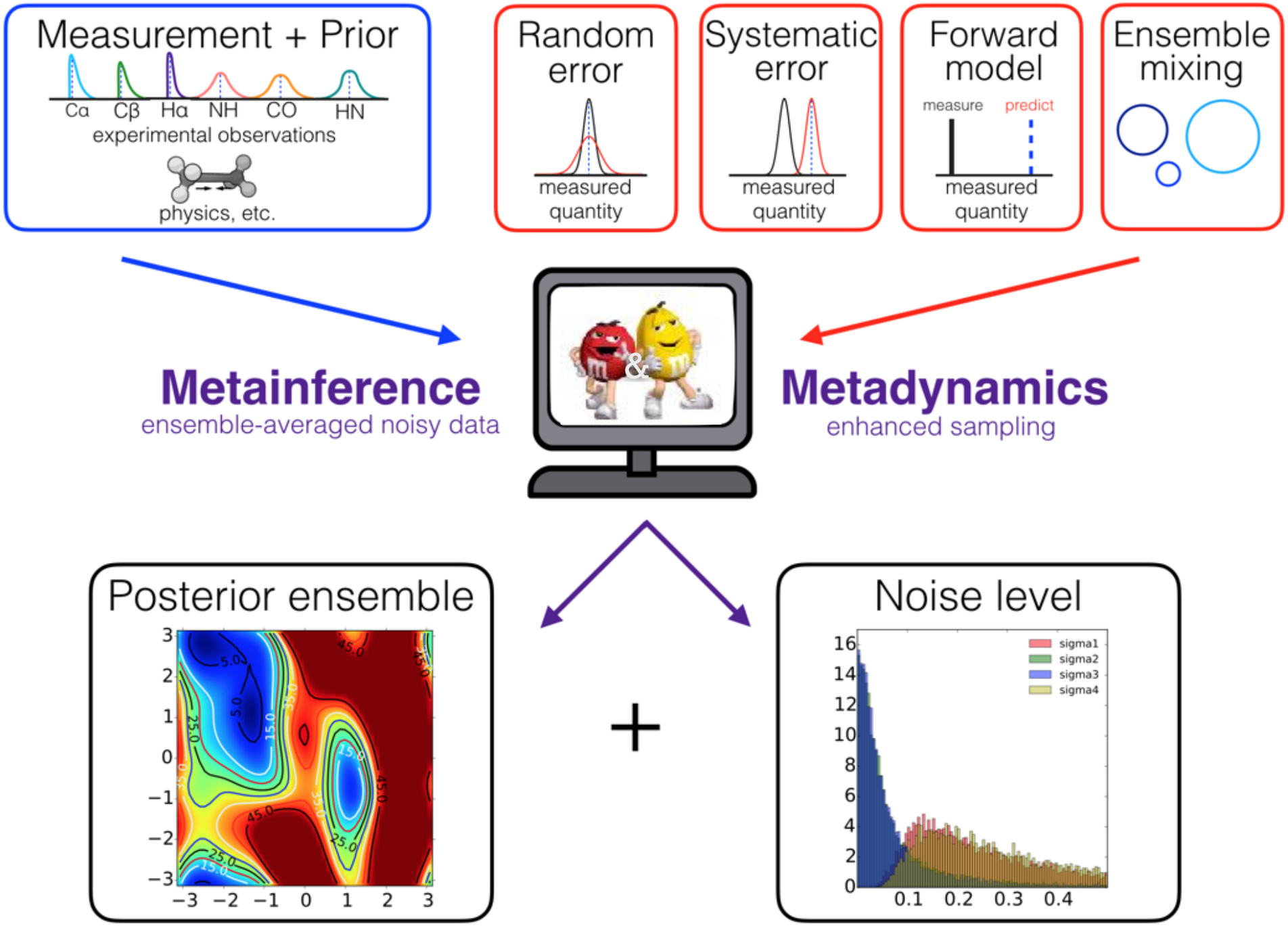
Bonomi *et al.* Science Advances 2016

Chemical Shifts weights



Bonomi *et al.* Science Advances 2016

Metadynamic Metainference



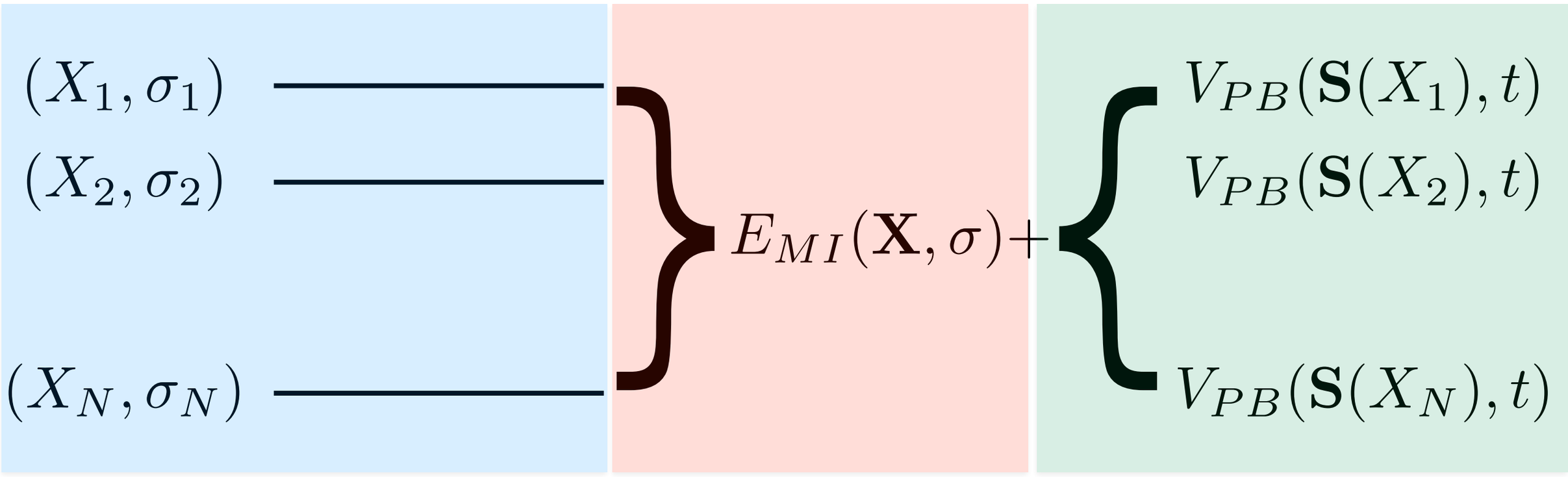
Bonomi *et al.* Scientific Reports. 2016

Metadynamics Metainference

Ensemble of replicas

Metainference energy function

PBMetaD bias



with these additional tricks:

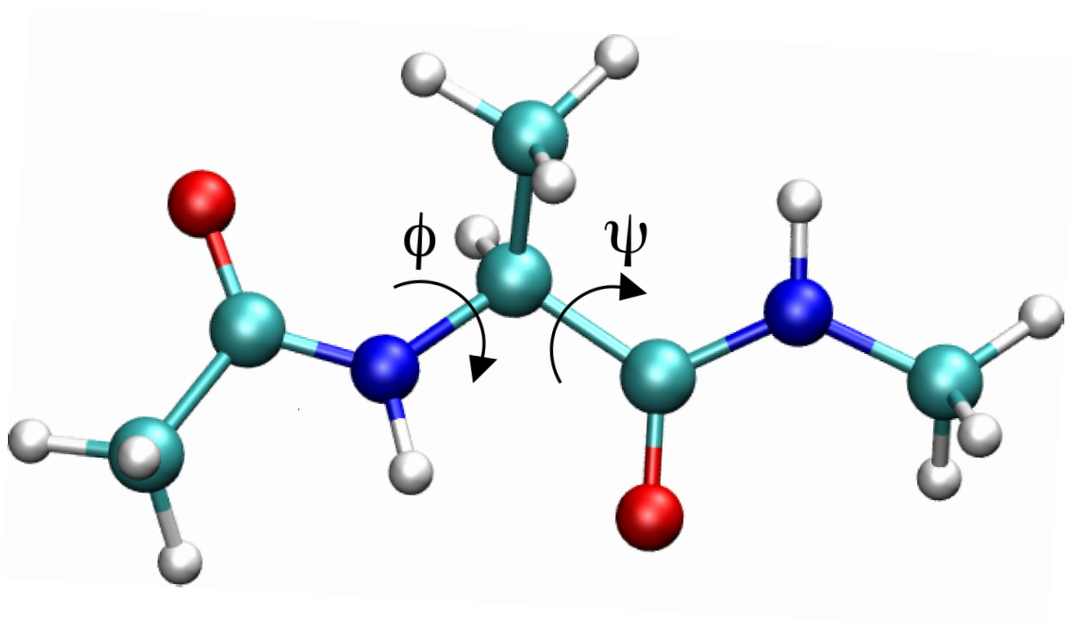
- replicas share the bias, as in multiple-walkers MetaD*
- need to reweigh to calculate averages in the unbiased ensemble

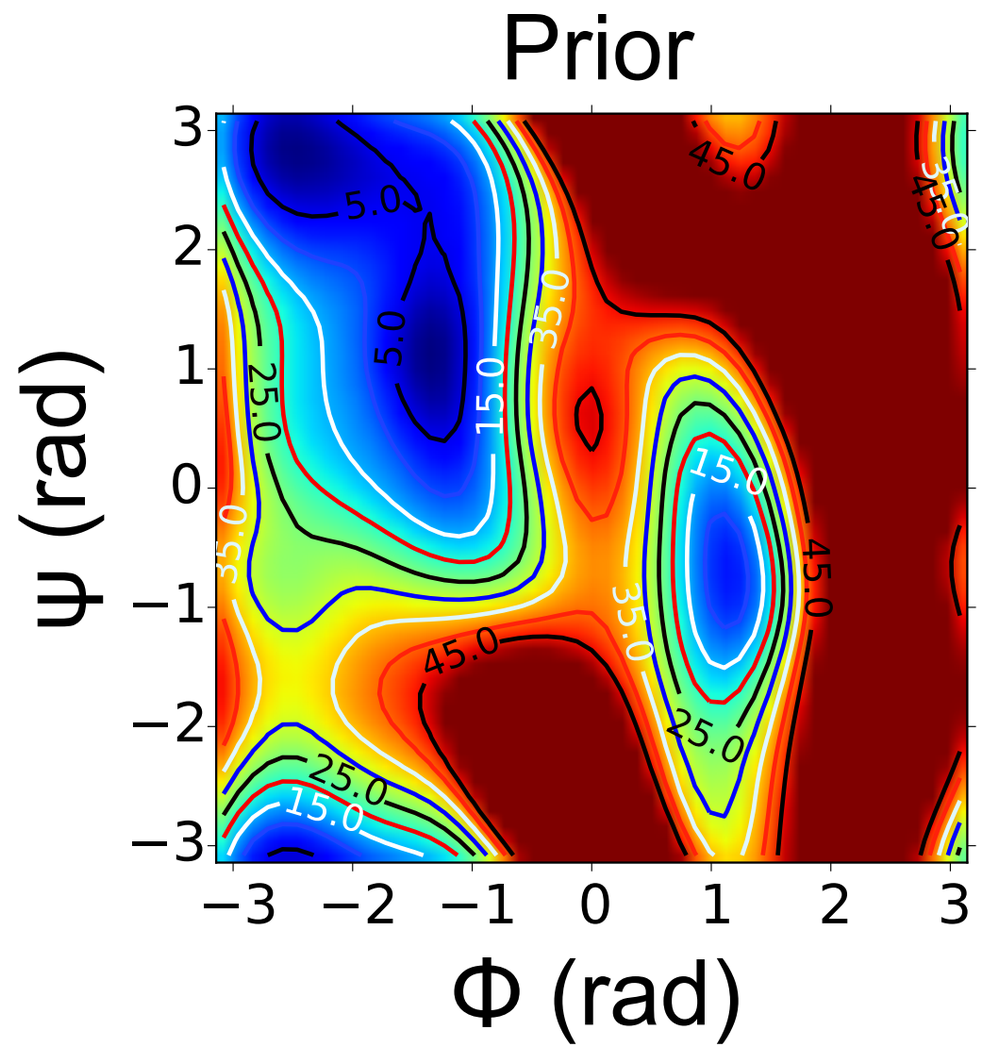
*Raiteri *et al.* JPCB 2006

Benchmark

Our favorite test case: alanine dipeptide in vacuum

The prior information is the AMBER99SB-ILDN force field*

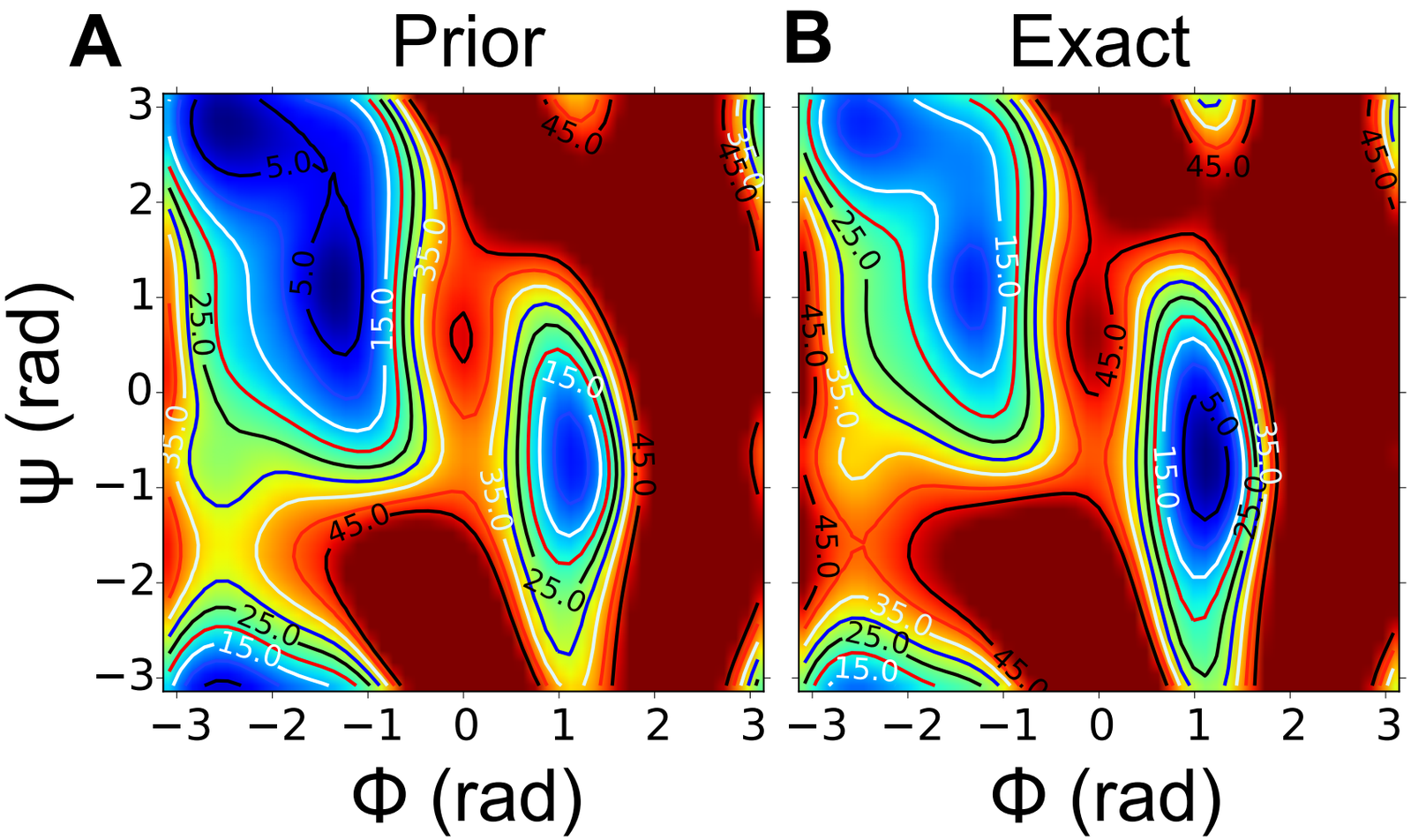




*Lindorff-Larsen *et al.* Proteins 2010

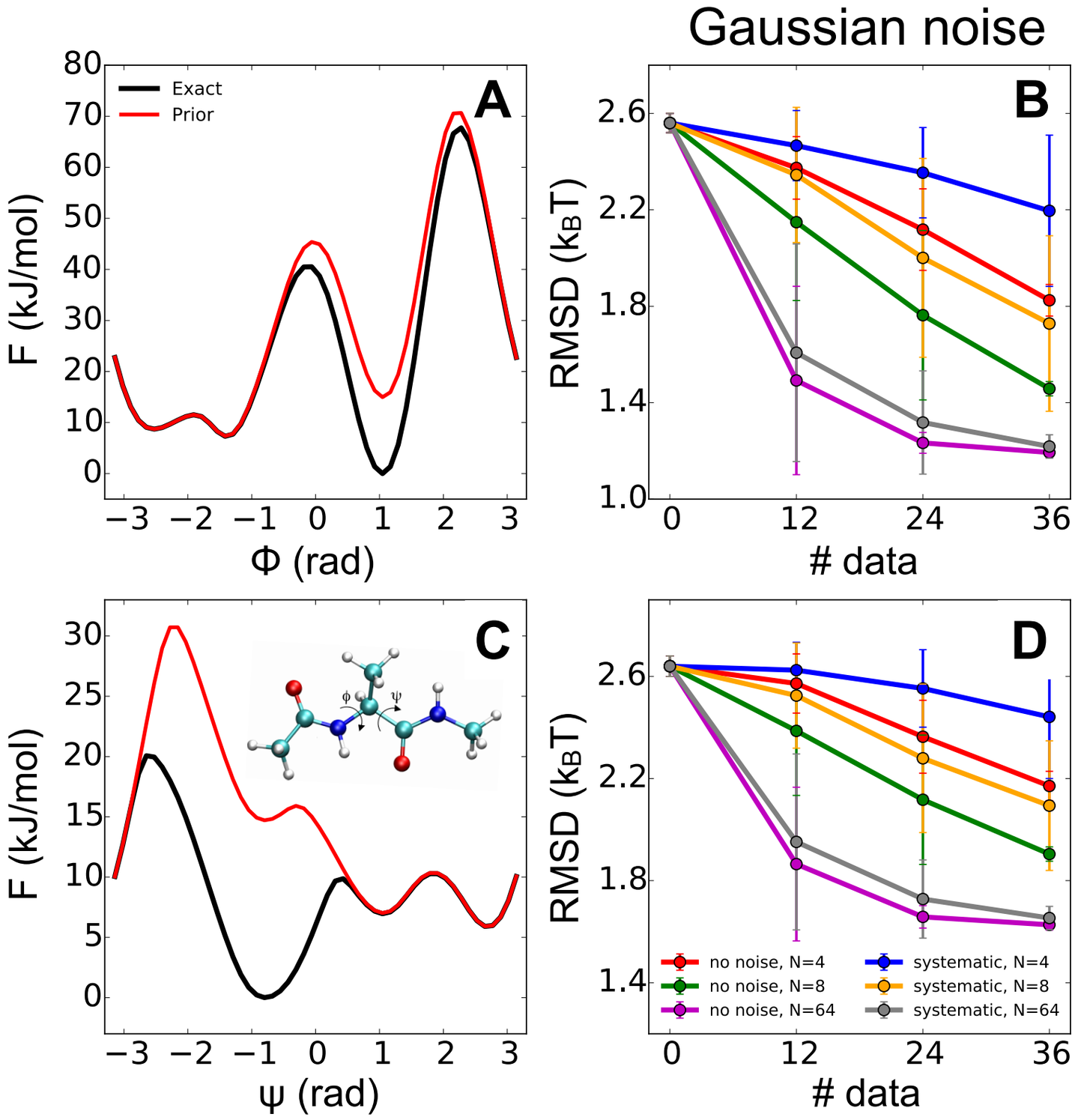
Benchmark

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



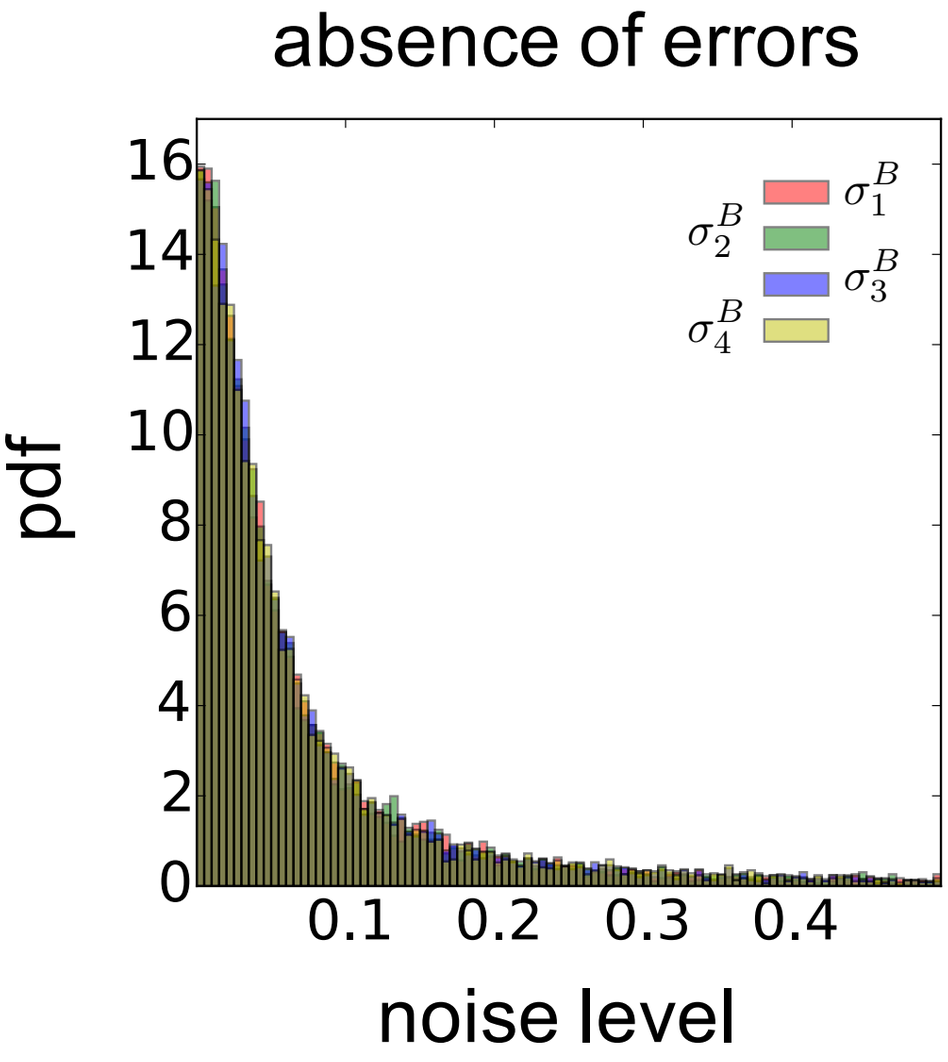
We introduce synthetic experimental data as average distances between heavy atoms, calculated in the <u>exact</u> ensemble, + **noise**

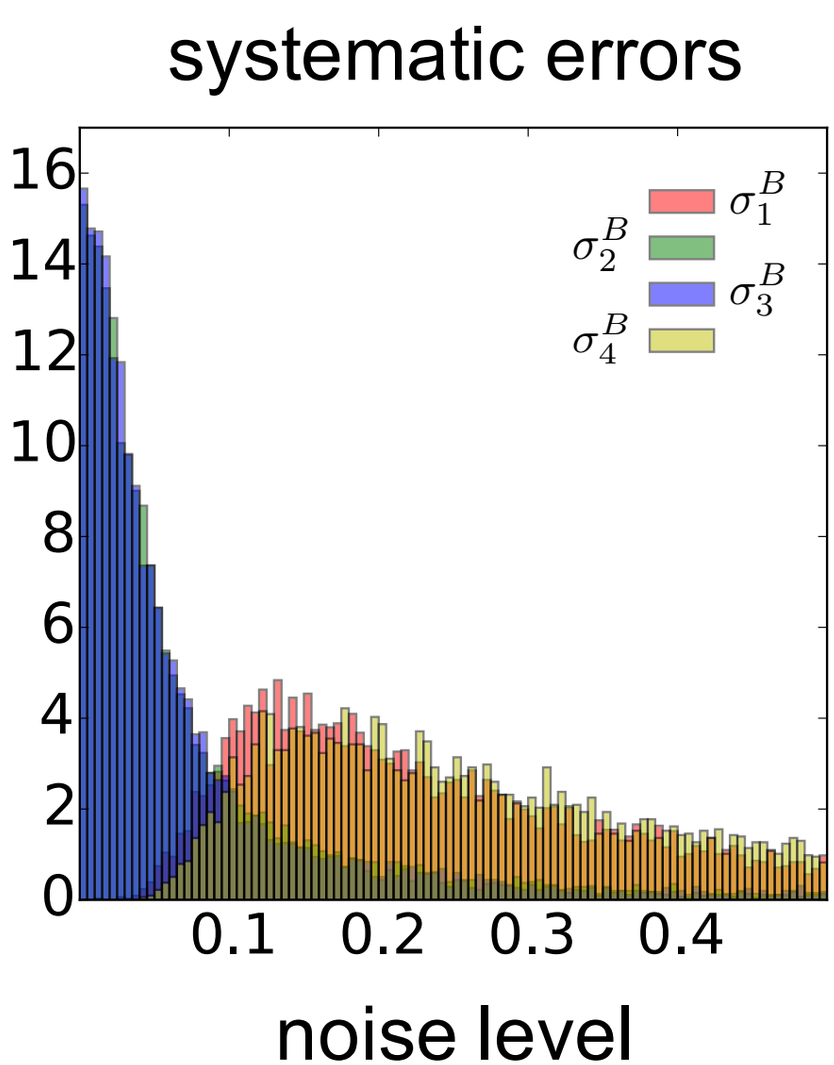
Results



Bonomi *et al.* Scientific Reports. 2016

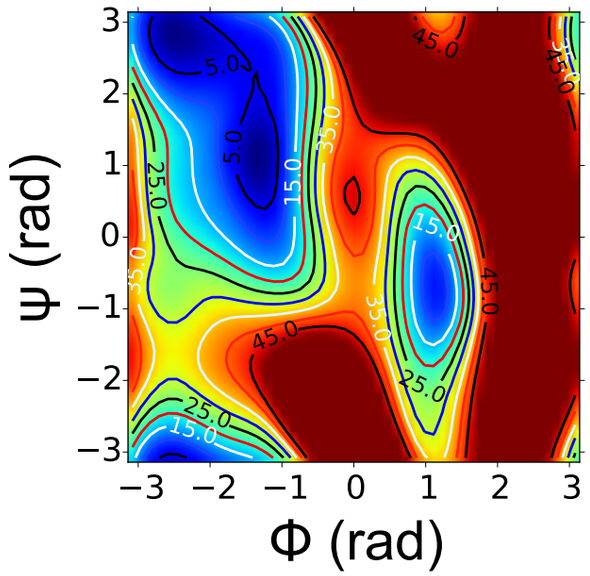
Noise inference



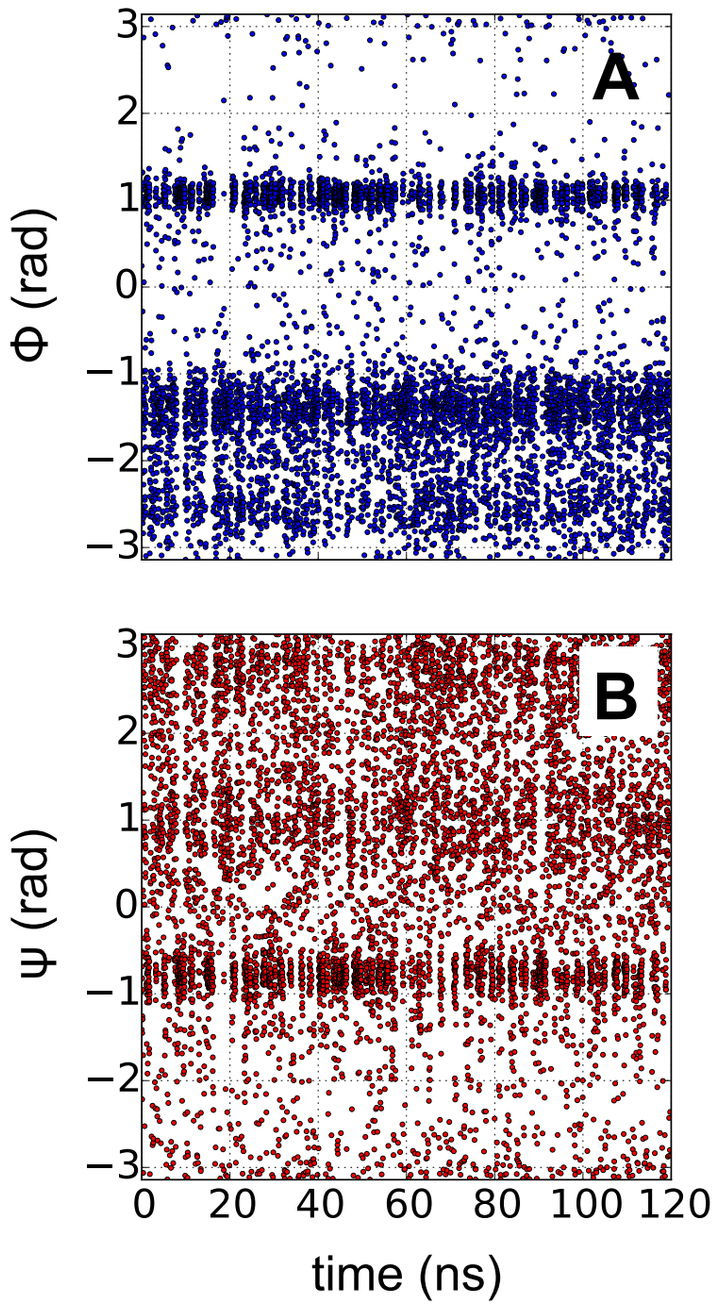


Bonomi *et al.* Scientific Reports. 2016

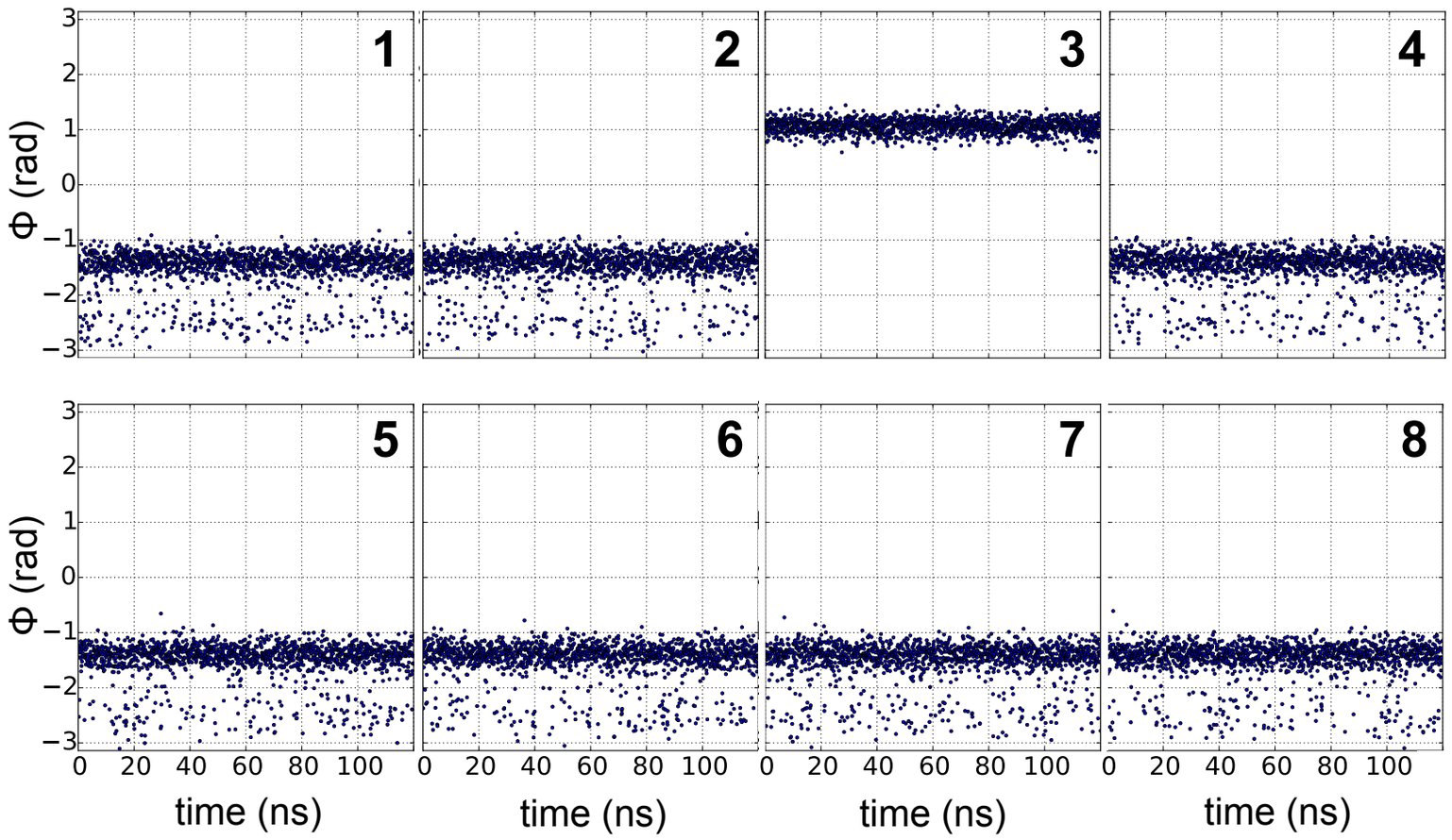
Metainference alone



with PBMetaD



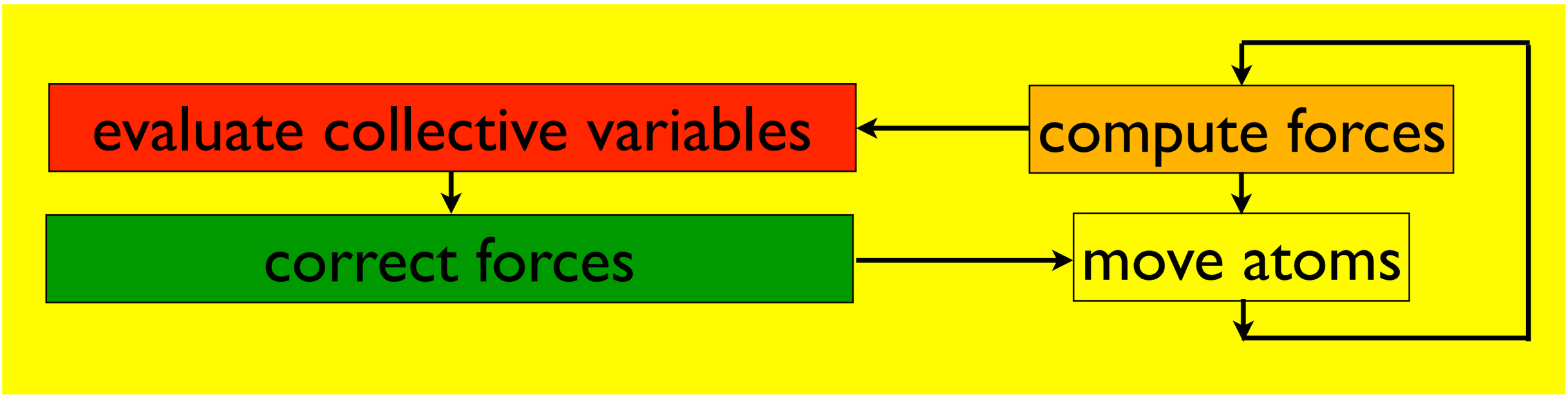
without PBMetaD



The implementation

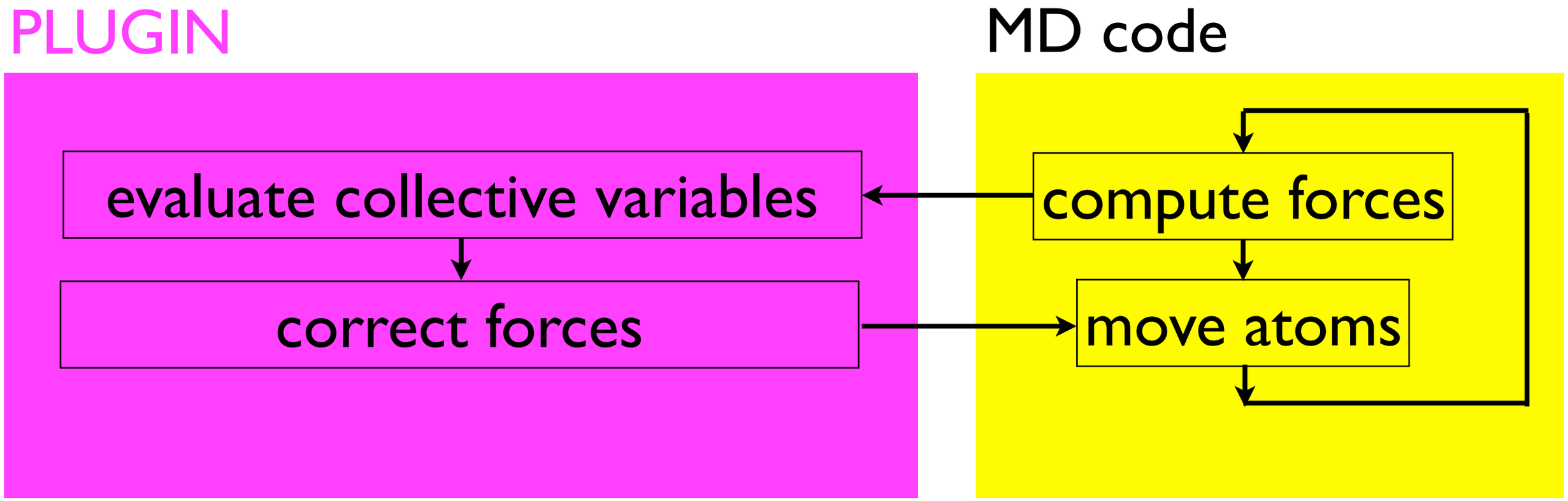
depending on the physical problem:
distances, angles, ...

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

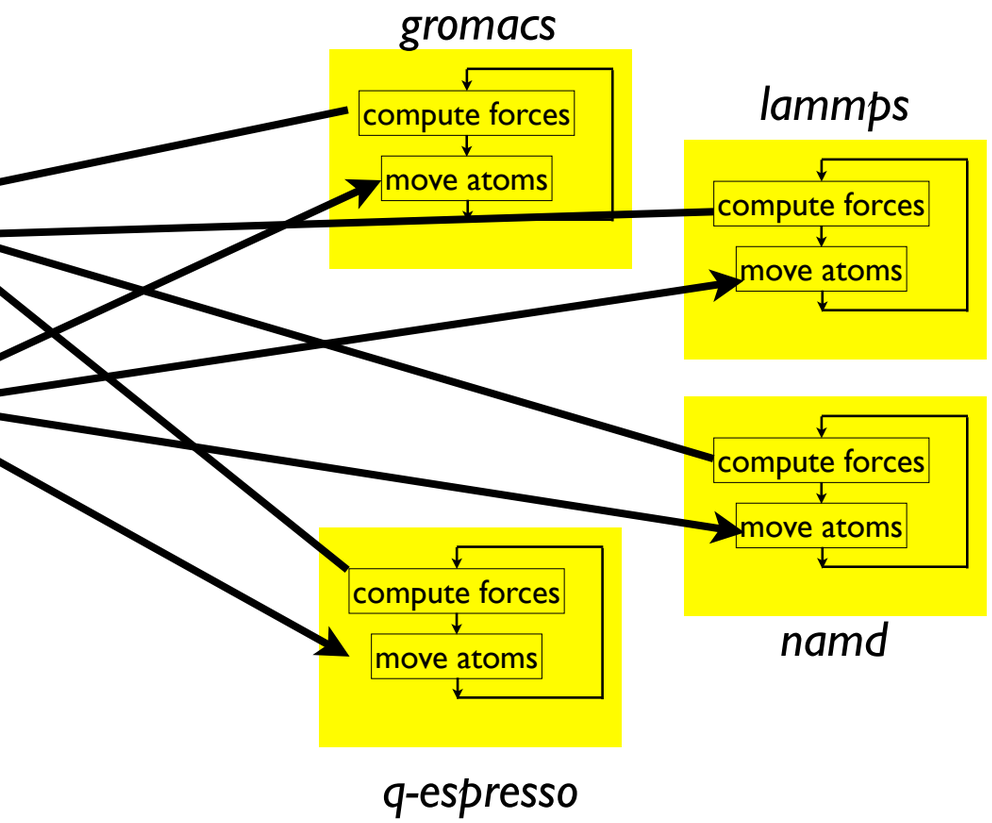
PLUMED

PLUGIN

evaluate collective variables

correct forces

MD codes



One open source pluging
for several MD codes!

Why **PLUMED**?

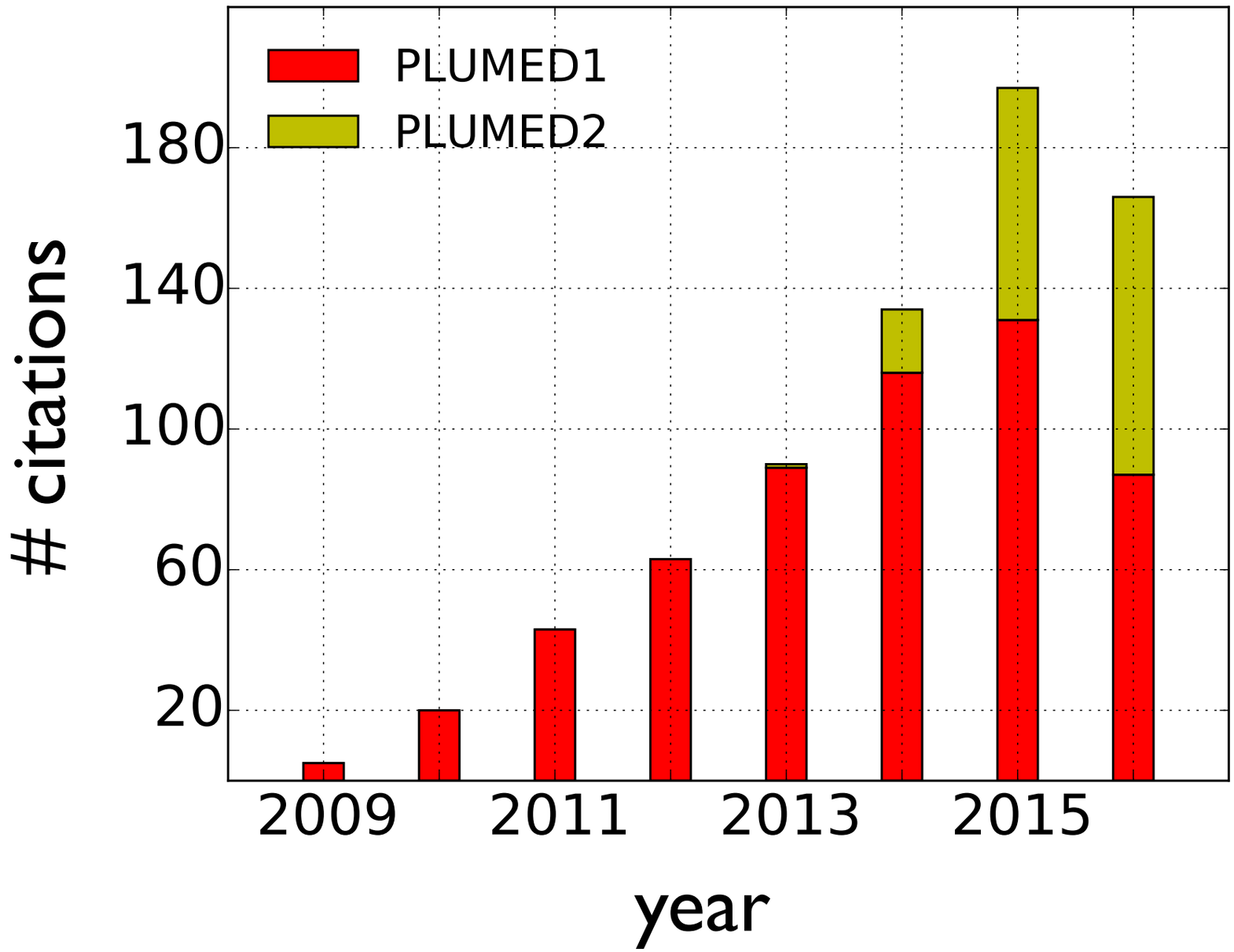
PLUgin for **ME**ta**D**ynamics

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

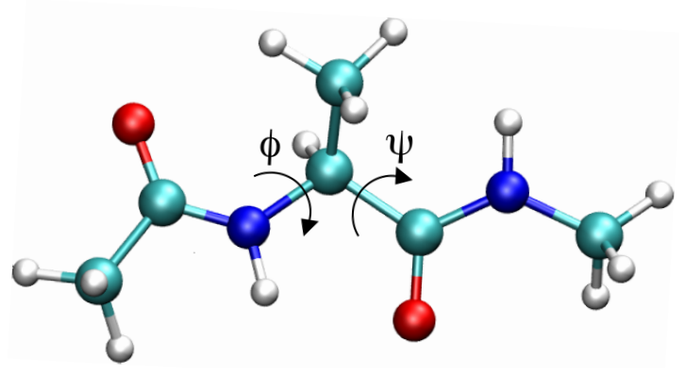
also derivatives w.r.t. atom positions

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
```

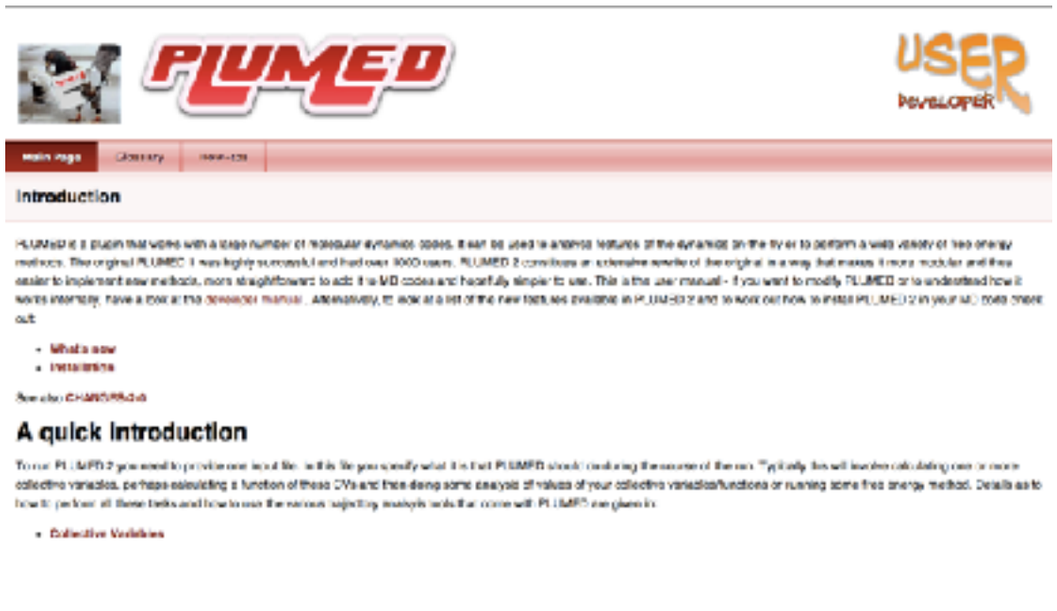
On the WEB

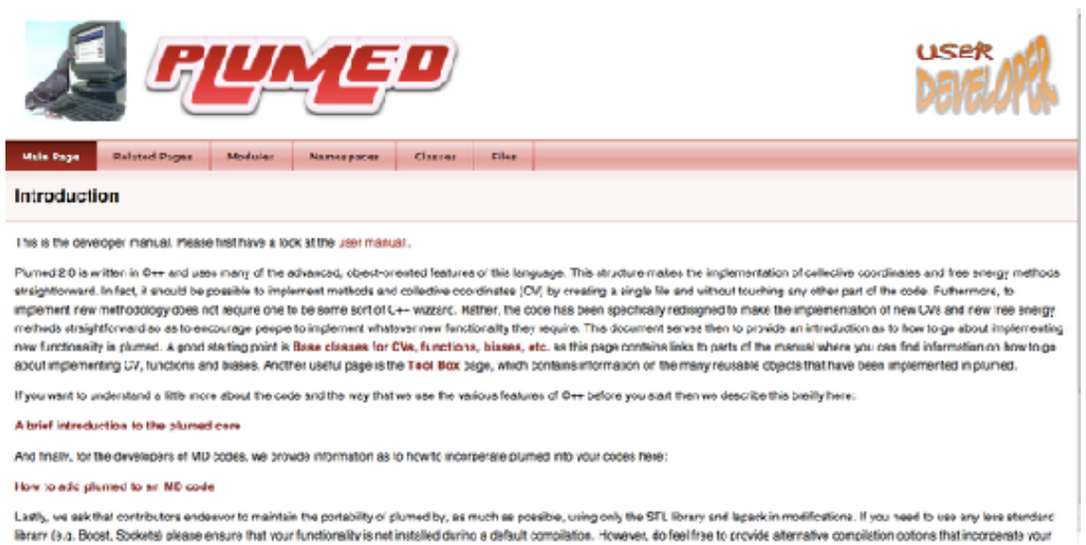
Website: http://www.plumed.org/

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

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

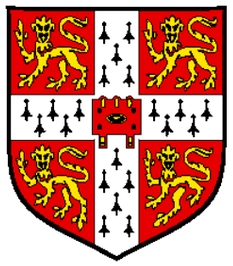
Metainference integrates noisy data collected on heterogeneous systems into MD simulations to improve the accuracy of force fields

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

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



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