



# Metadynamics Metainference with *PLUMED*

Max Bonomi  
University of Cambridge  
[mb2006@cam.ac.uk](mailto: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

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

*V<sub>eff</sub>*

The potential (or force field) is derived from

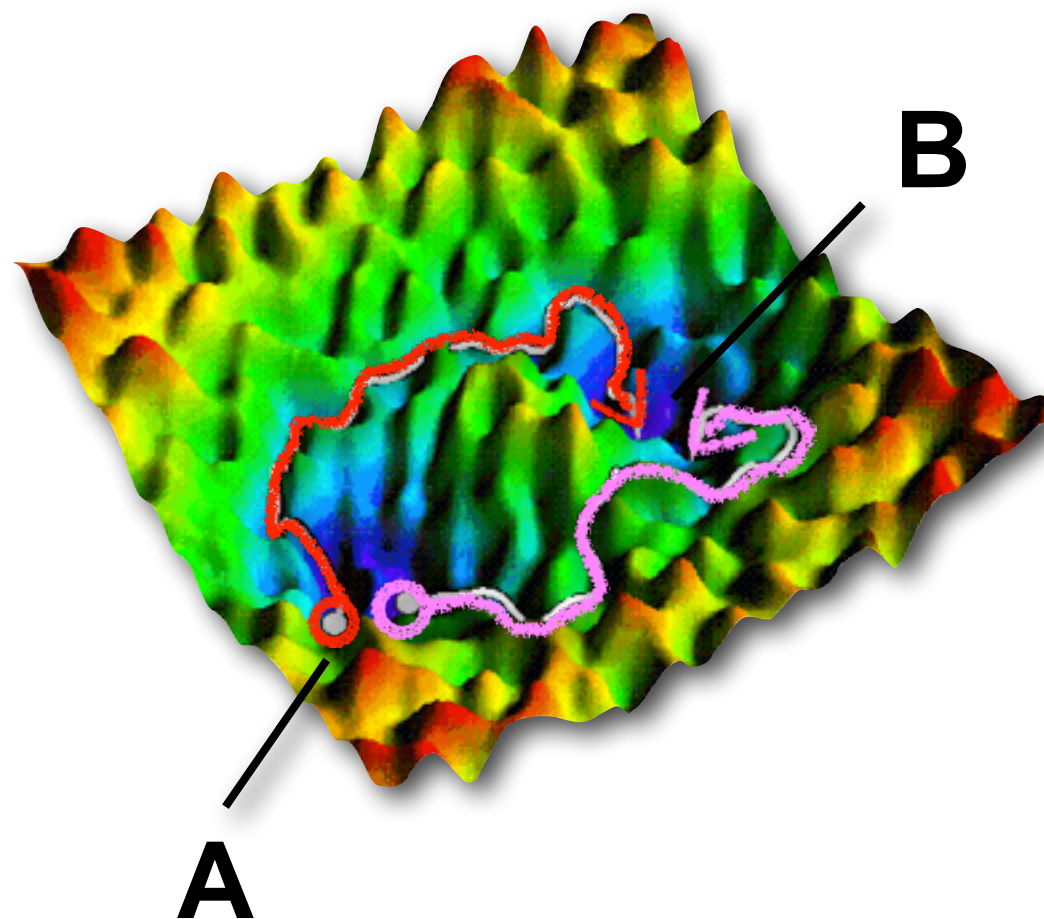
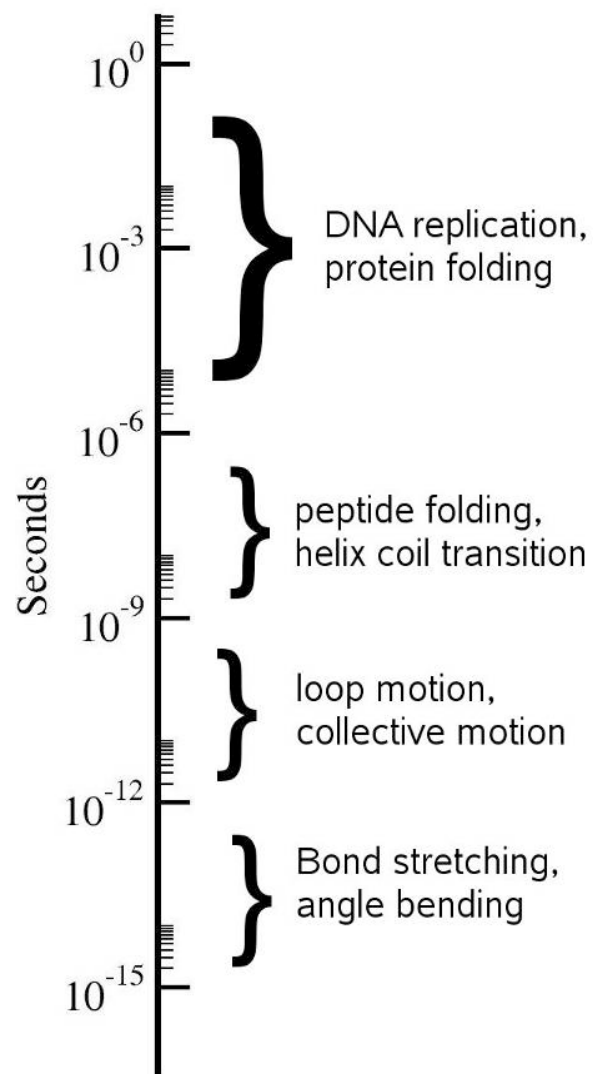
- Higher accuracy calculations
- Fitting experimental observables

Limitations:

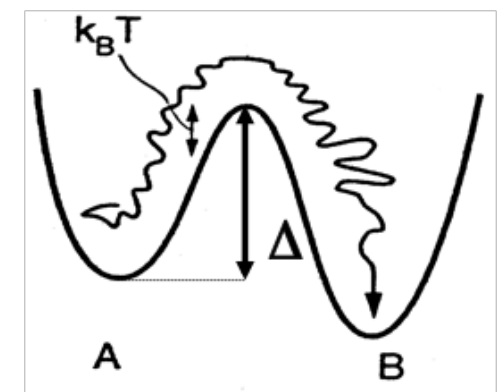
- time scale accessible in standard MD
- accuracy of classical force fields

# 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:

$$\mathbf{S} = \mathbf{S}(\mathbf{R}) = (S_1(\mathbf{R}), \dots, S_d(\mathbf{R}))$$

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

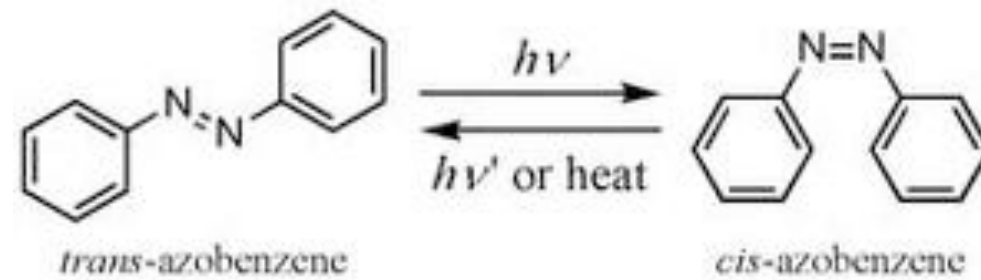
$$F(\mathbf{S}) = -\frac{1}{\beta} \ln P(\mathbf{S}) \quad \text{where} \quad \beta = \frac{1}{k_B T}$$

canonical  
ensemble

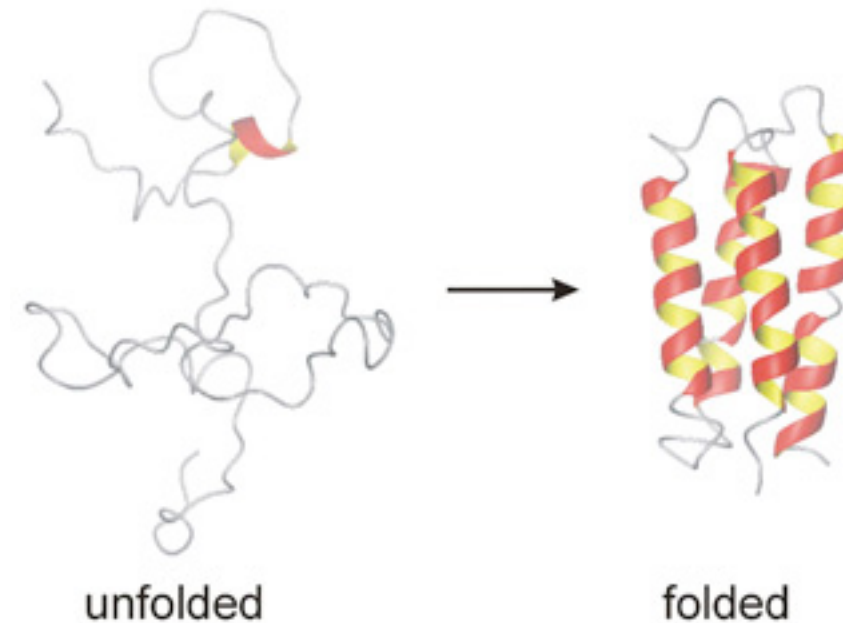
$$P(\mathbf{S}) = \frac{\int d\mathbf{R} \delta(\mathbf{S} - \mathbf{S}(\mathbf{R})) e^{-\beta U(\mathbf{R})}}{\int d\mathbf{R} e^{-\beta U(\mathbf{R})}}$$

# Examples

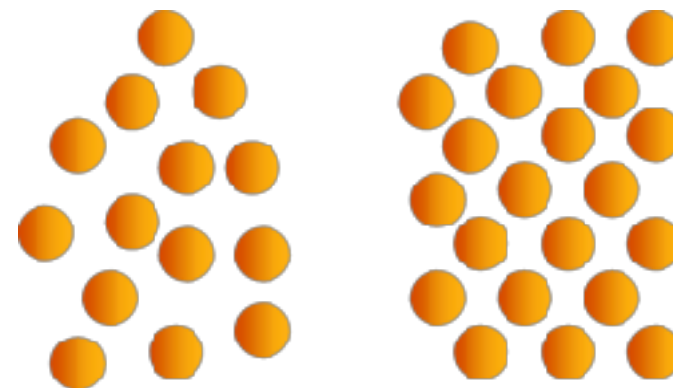
Isomerization:  
dihedral angle



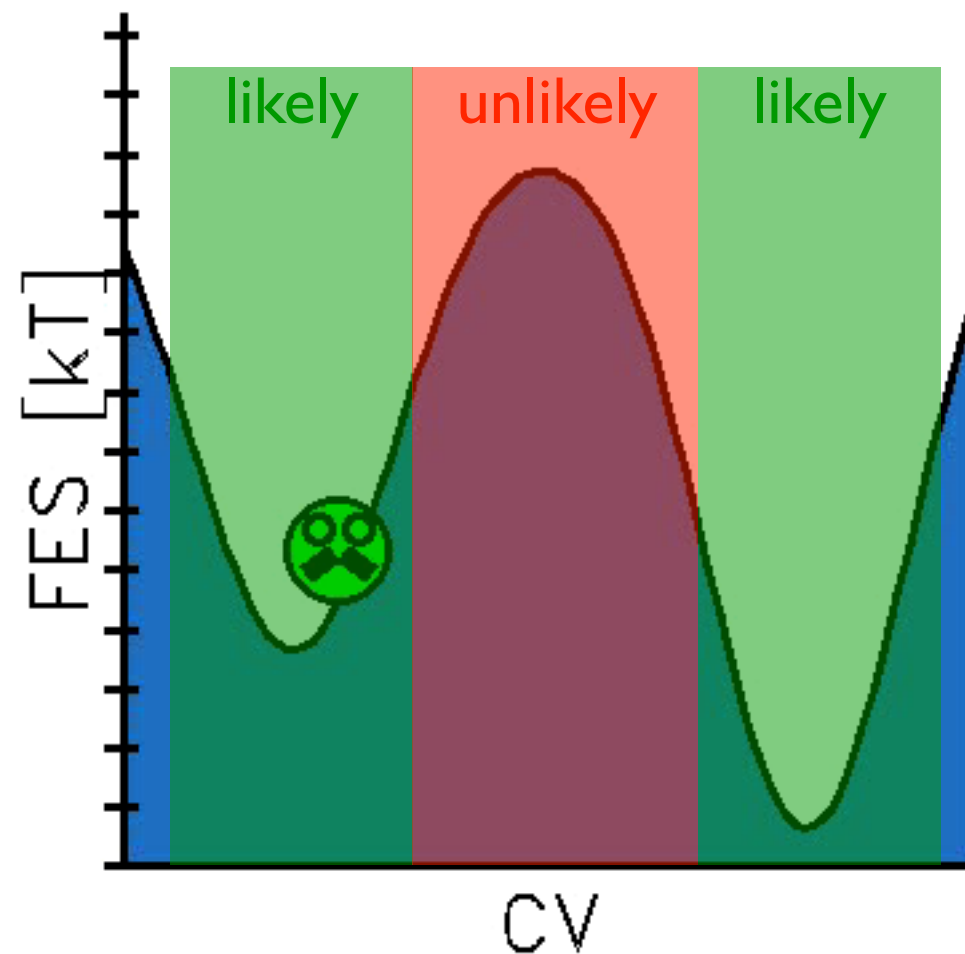
Protein folding:  
gyration radius,  
number of contacts,  
...



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



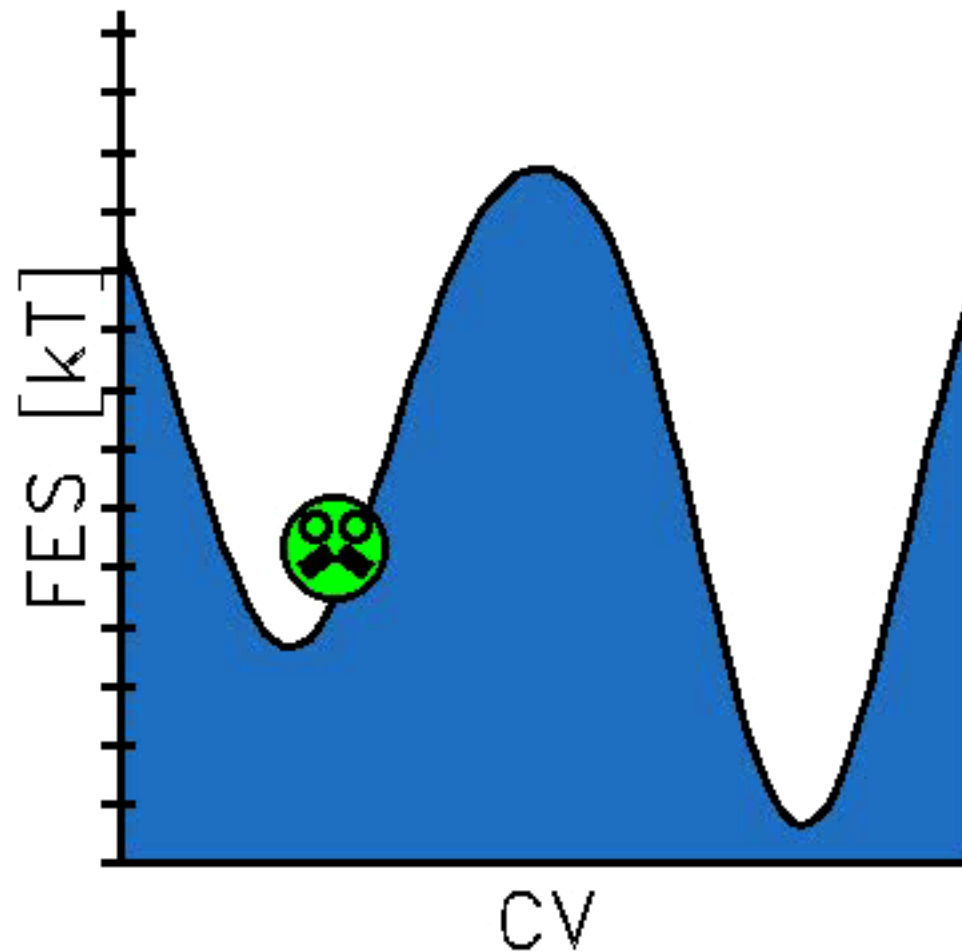
# Rare events simplified



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

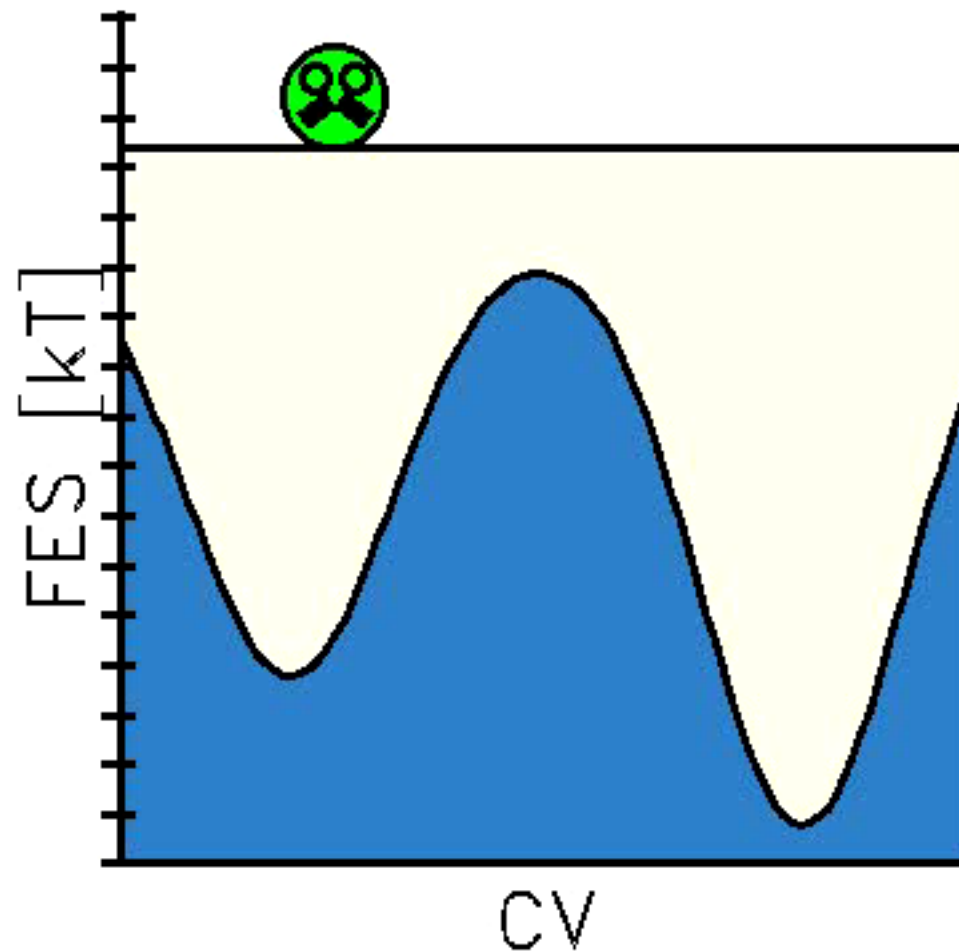


The idea is to add a bias potential that acts on the collective variables:

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

What is a good choice of bias potential?

# Biased sampling



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

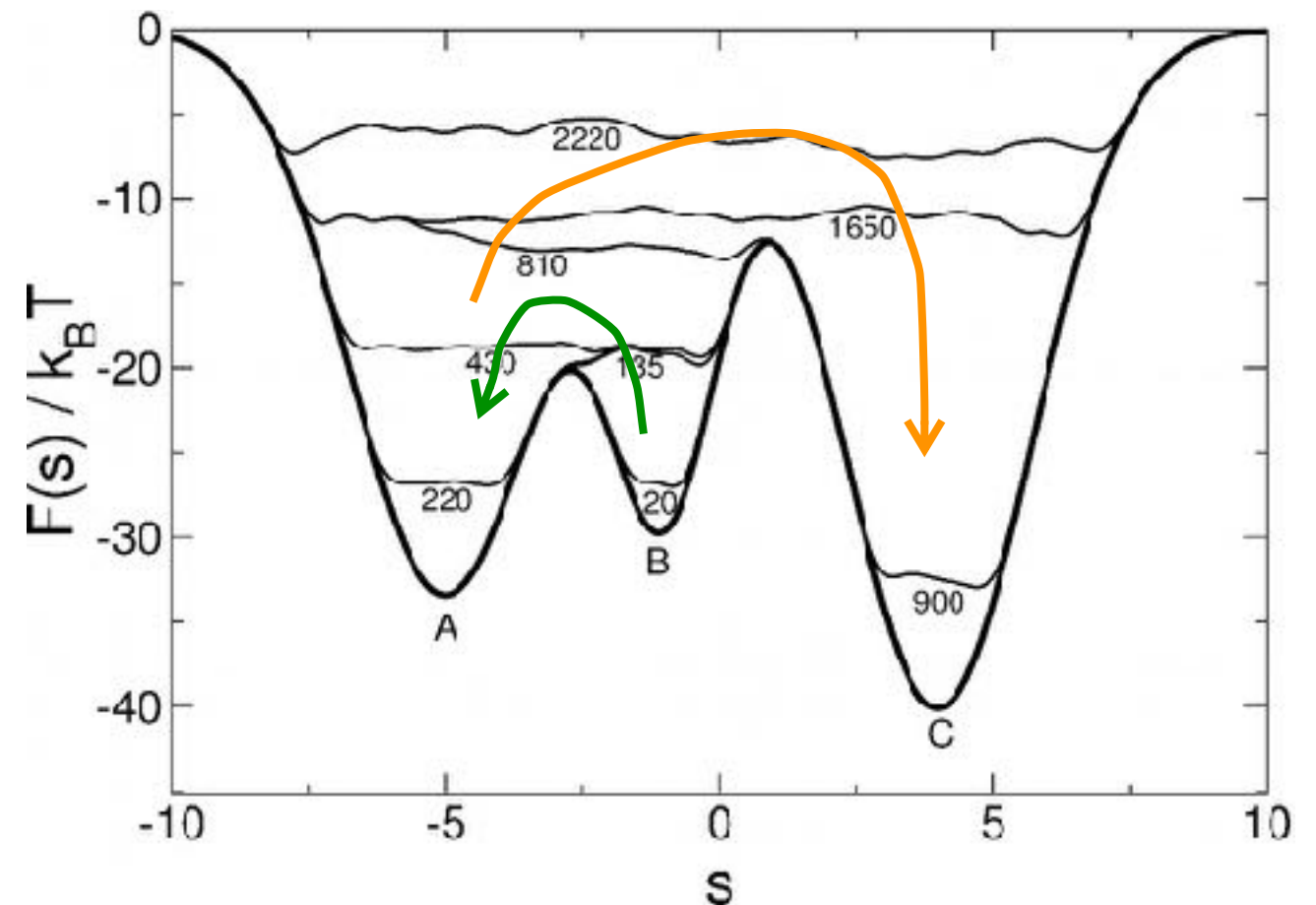
$$U(\mathbf{R}) \rightarrow U(\mathbf{R}) + V(\mathbf{S}(\mathbf{R}))$$

What is a good choice of bias potential?

# Metadynamics

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

$$\mathbf{S} = (S_1(\mathbf{R}), \dots, S_d(\mathbf{R}))$$



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

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

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:

$$V_G(\mathbf{S}, t \rightarrow \infty) = -F(\mathbf{S}) + C \quad \text{Bussi, Laio, Parrinello PRL 2006}$$

- *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

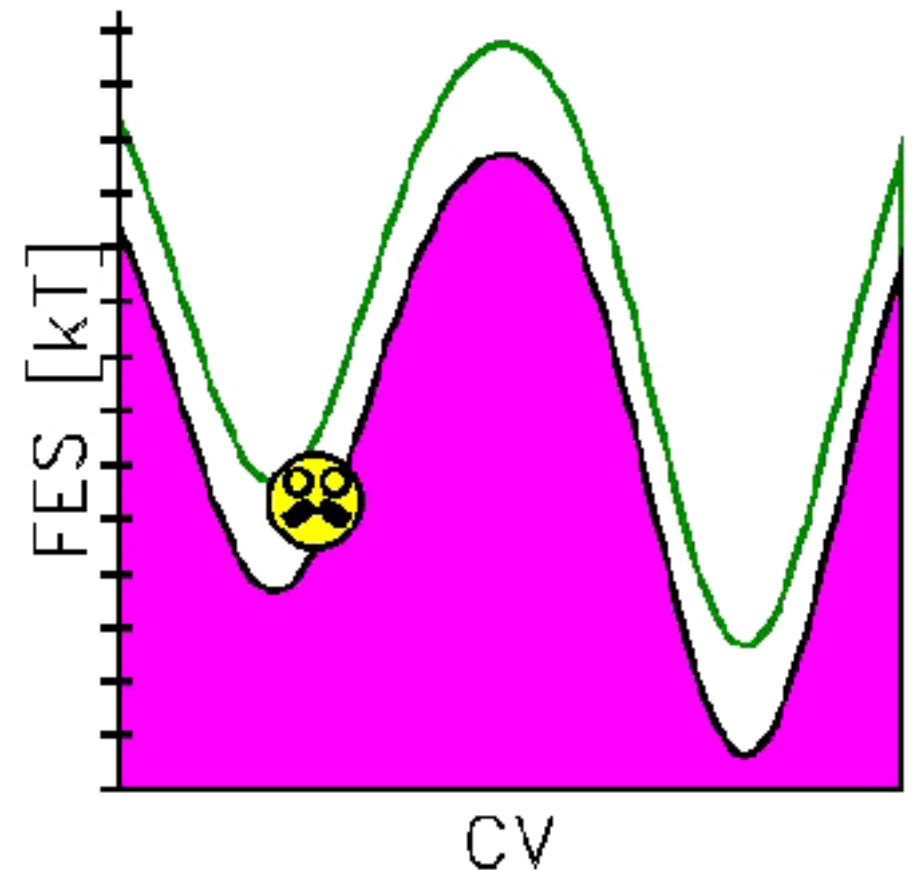


# Well-Tempered Metadynamics

The initial Gaussian height  $w_0$  is rescaled during the simulation:

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

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



- Convergence and overfilling issues solved:

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

- $\Delta T$  used to tune the extent of exploration

# 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:

$$V(S_1, t), \dots, V(S_N, t)$$

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

Each bias potential converges to the corresponding free energy:

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

# Parallel Bias Metadynamics

Since we are not interested in the  $\eta$ -distribution, we can marginalize this variable:

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

where:

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

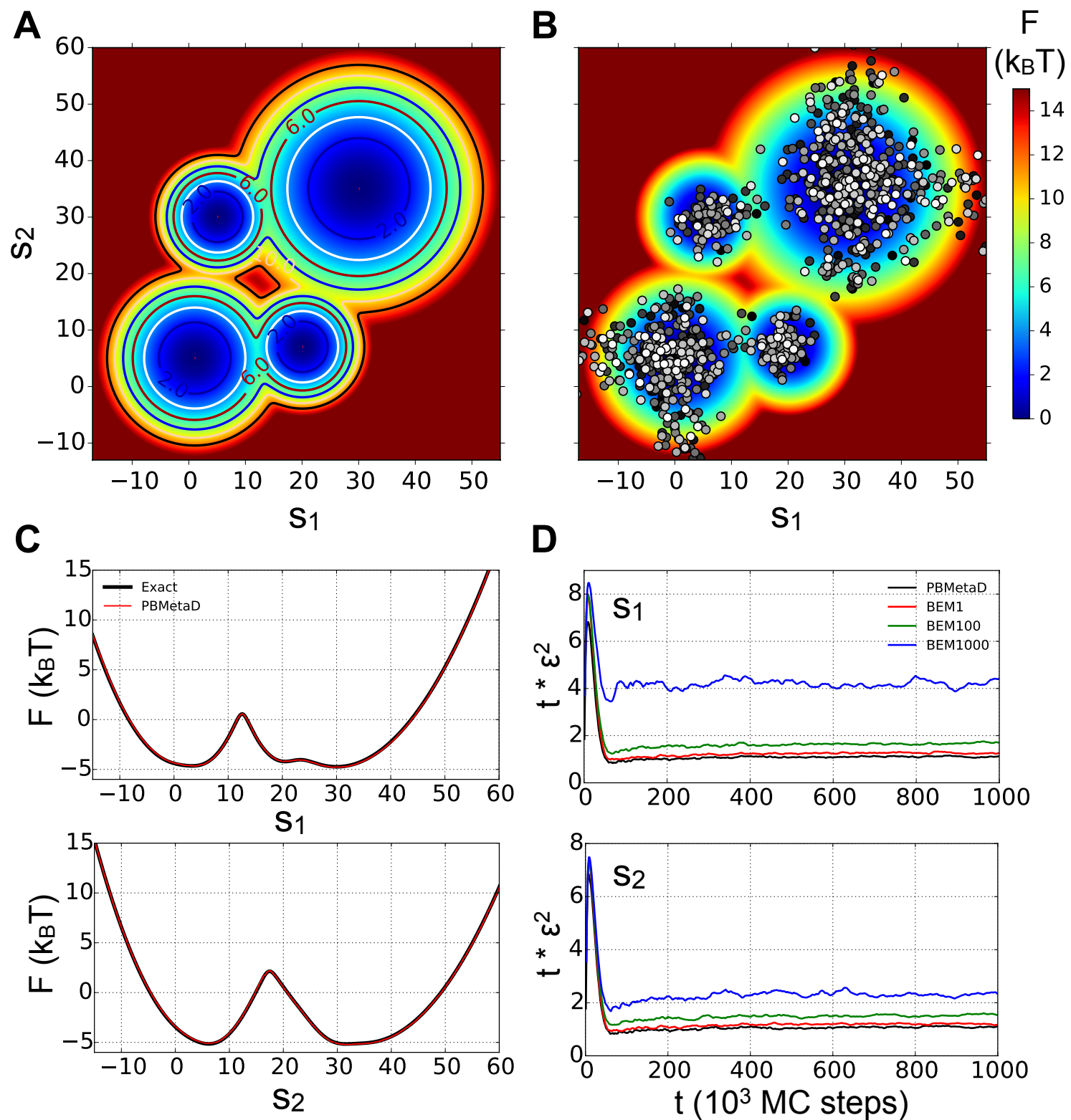
In order for each bias potential to converge to the corresponding free energy, we need a new rescaling rule:

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

where:

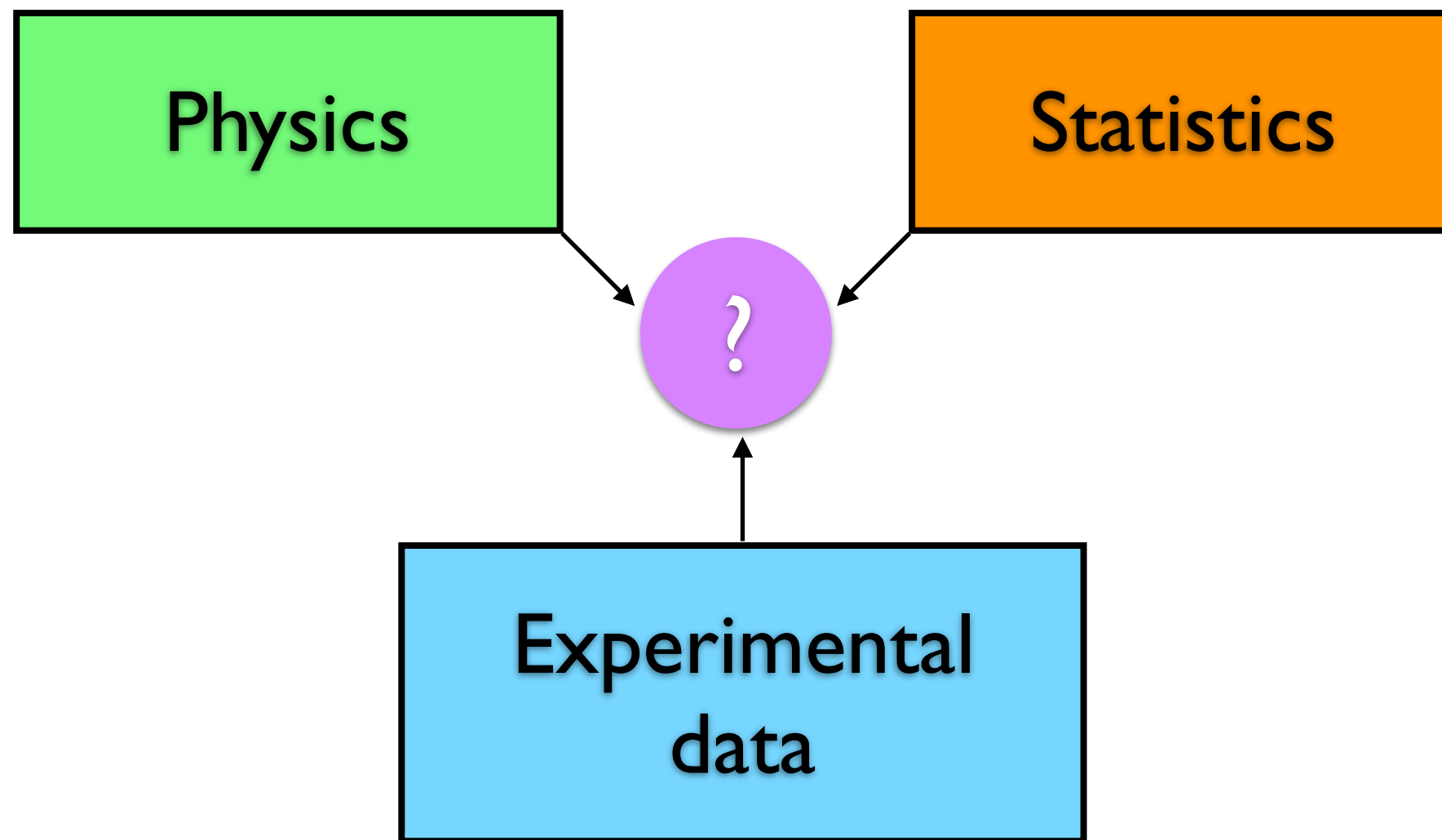
$$P(\eta_i = 1 | \mathbf{R}) = \frac{\exp[-\beta V(S_i, t)]}{\sum_{j=1}^N \exp[-\beta V(S_j, t)]}$$

# Benchmark on a model system







# 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 <sup>+</sup> <sup>+</sup> Camilloni <i>et al.</i> JACS 2012
Random		Bayesian Modelling <sup>*</sup> <sup>*</sup> Rieping <i>et al.</i> Science 2005
Systematic		
Forward model		

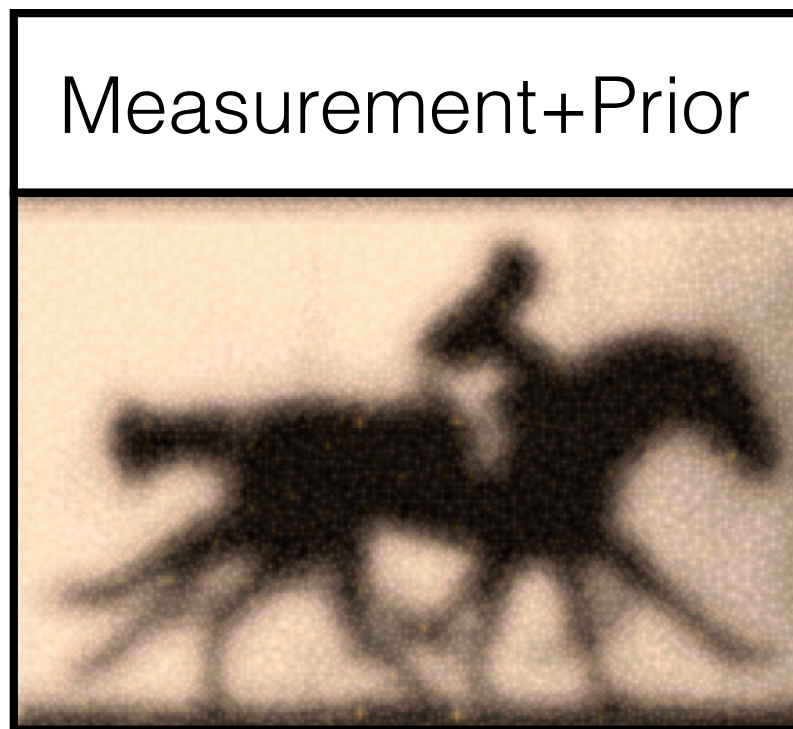


# Addressing these challenges

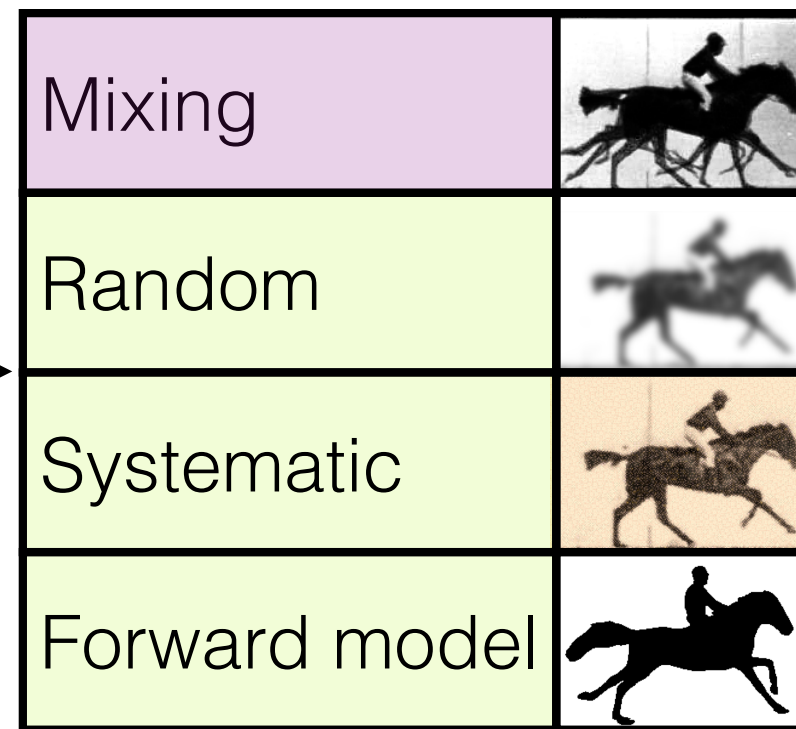
Mixing		<p><b>Metainference*</b></p> <p><i>*Bonomi et al. Science Advances 2016</i></p>
Random		
Systematic		
Forward model		

# To produce ensemble of models and determine their populations

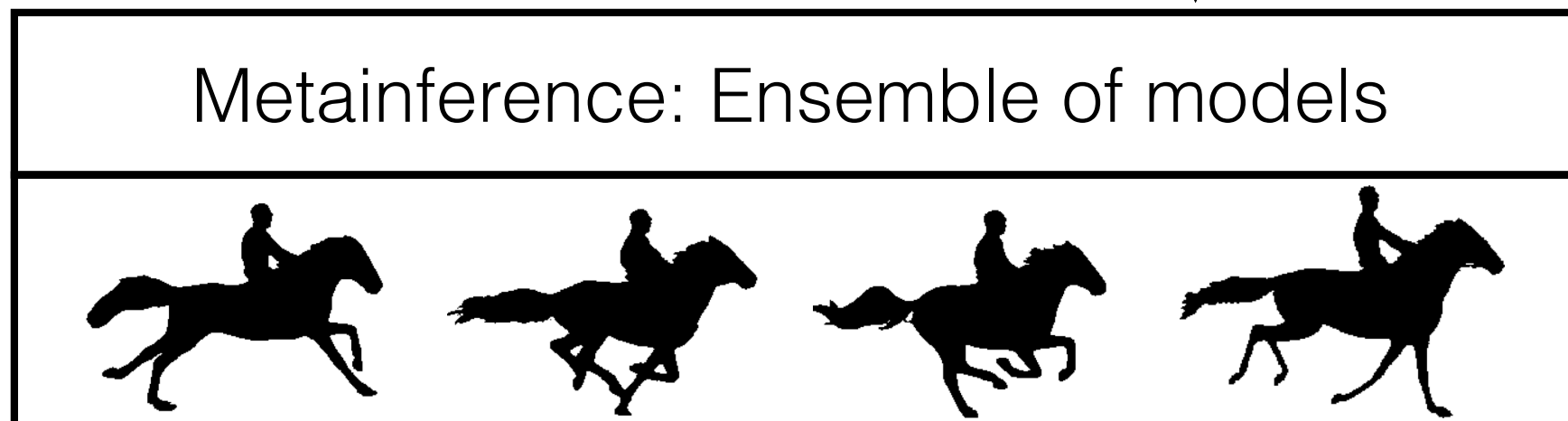
a) Input



b) Errors



c) Output



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

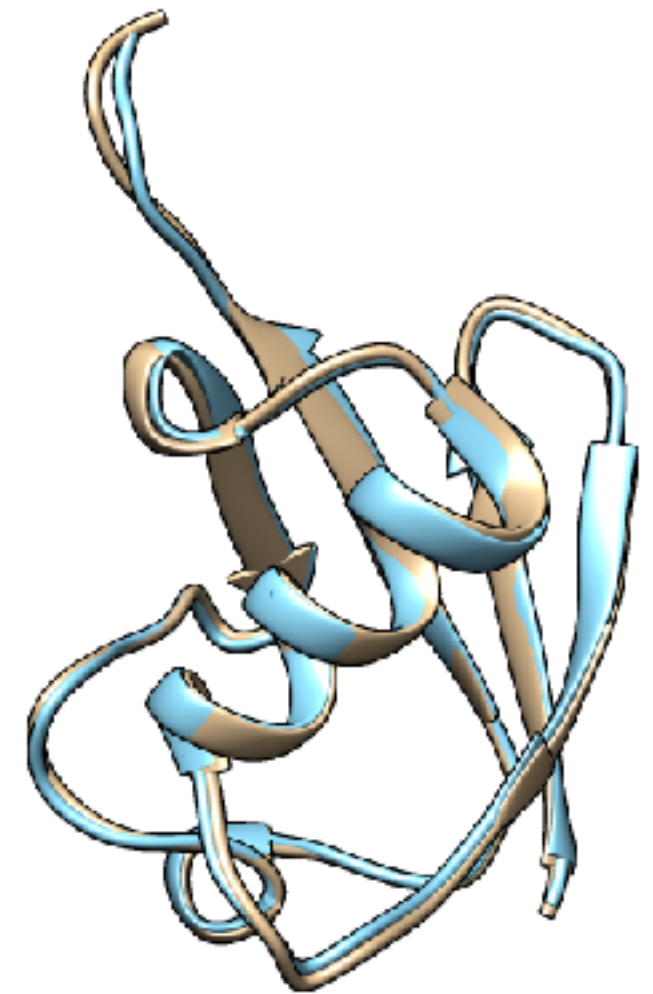
# Integrative Dynamical Biology

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)



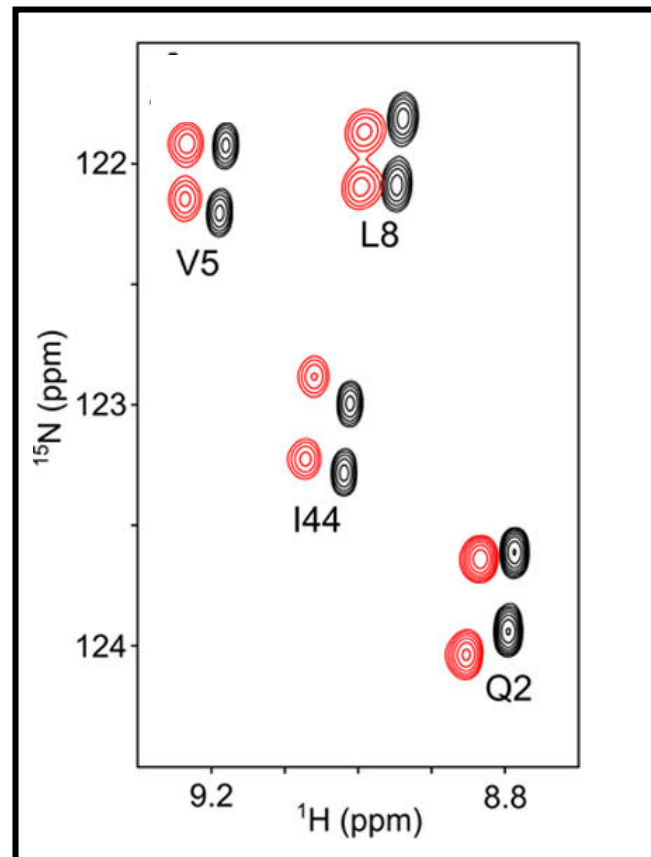
$C\alpha$ -RMSD = 0.52 Å

and with the ensemble generated by standard MD

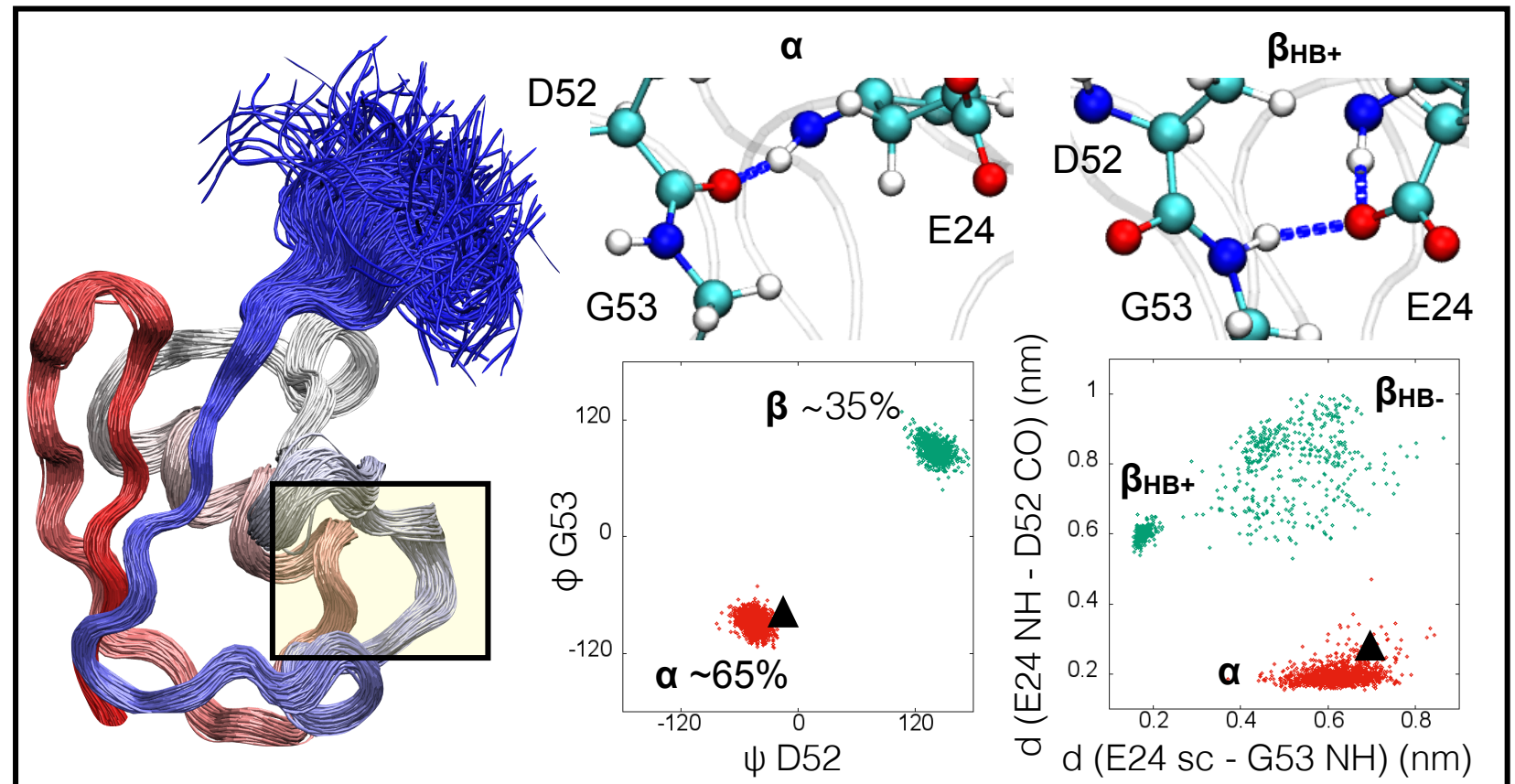
Models are evaluated by fit with other exp data (RDCs, J3)

# Ubiquitin ensembles

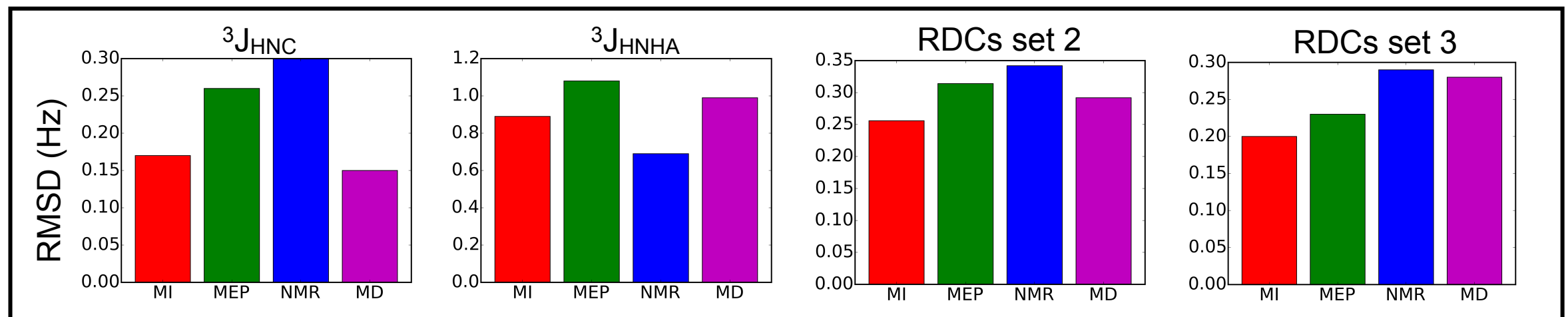
a) Input



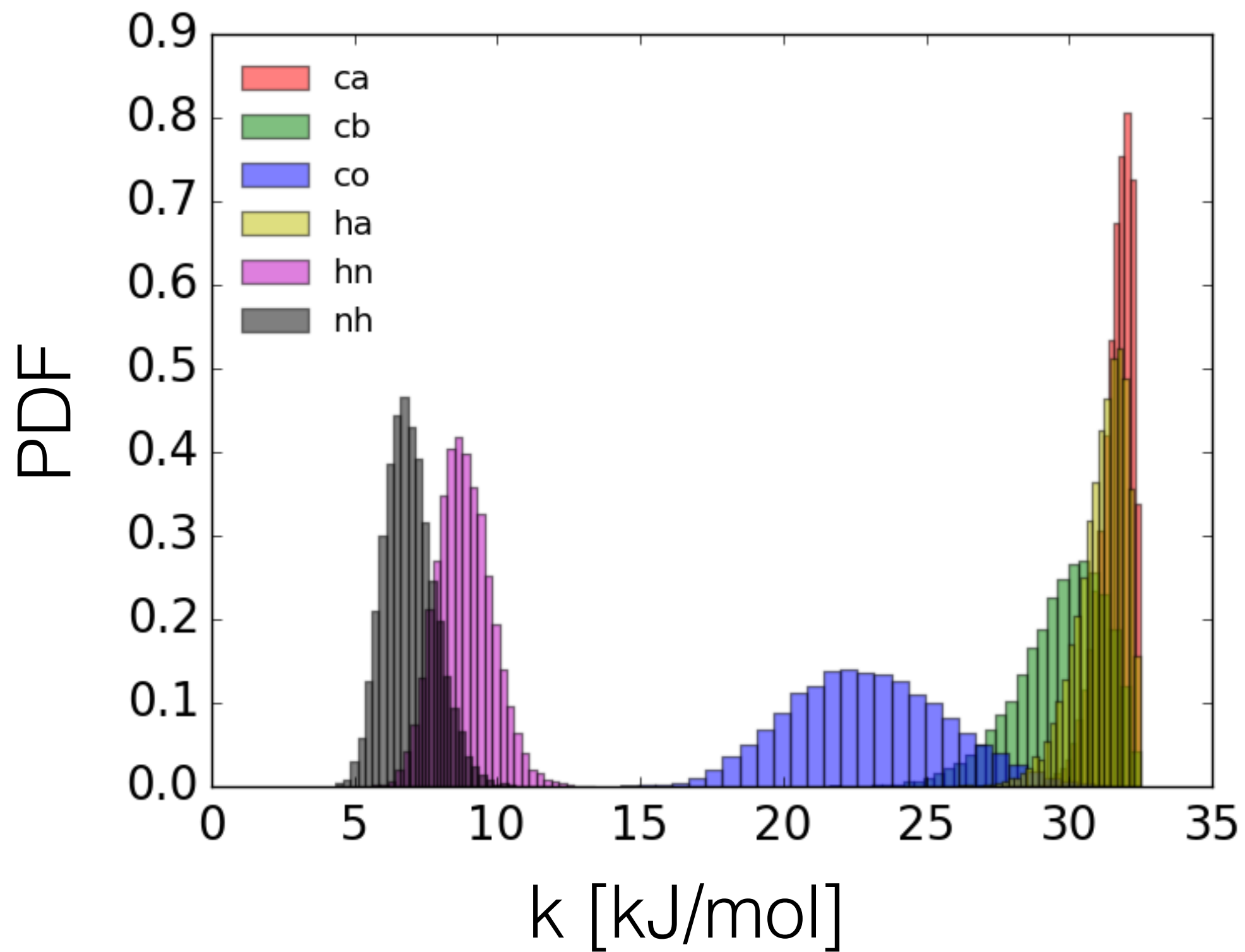
b) Ensemble



c) Validation

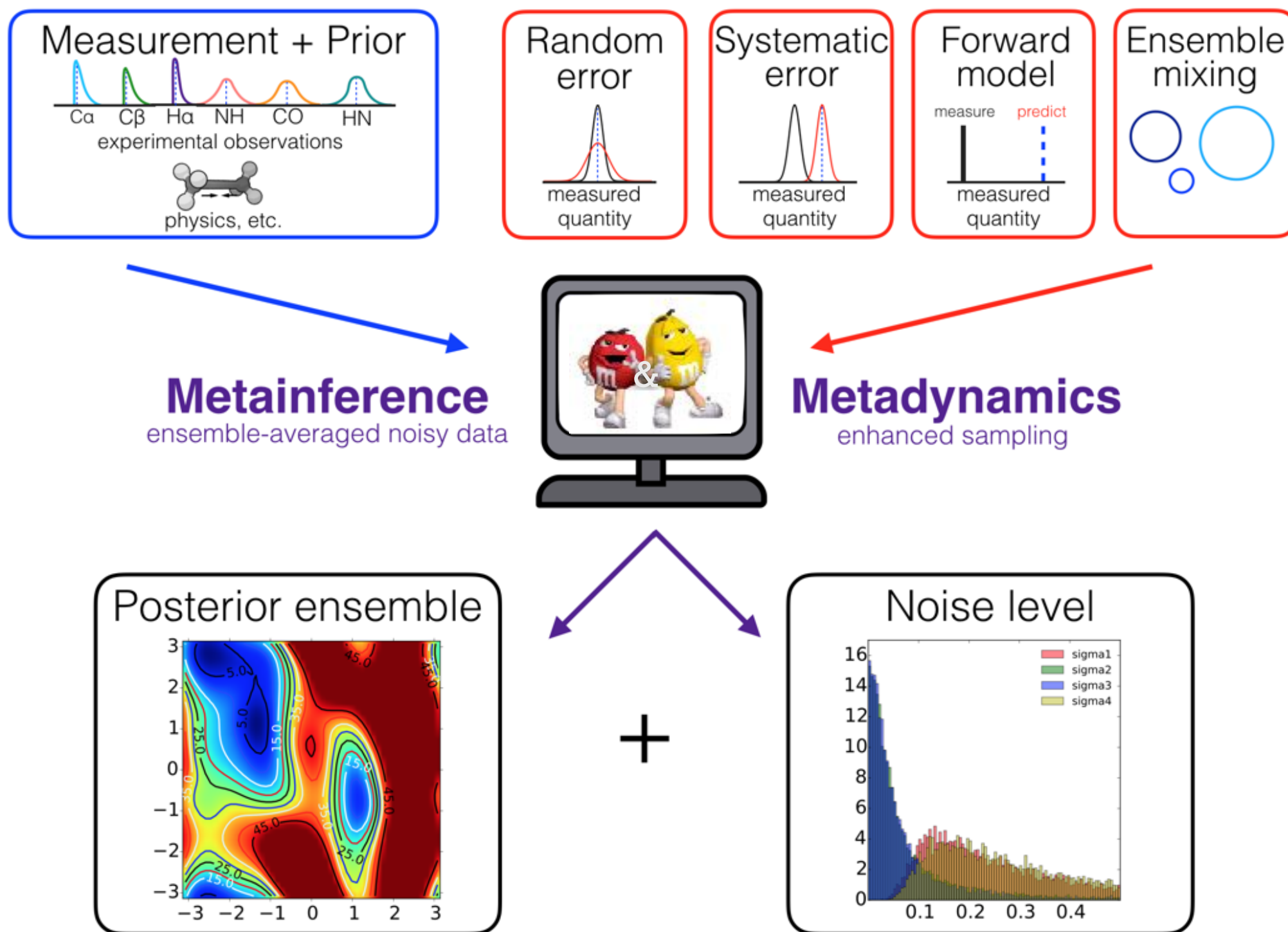


# Chemical Shifts weights





# Metadynamic Metainference



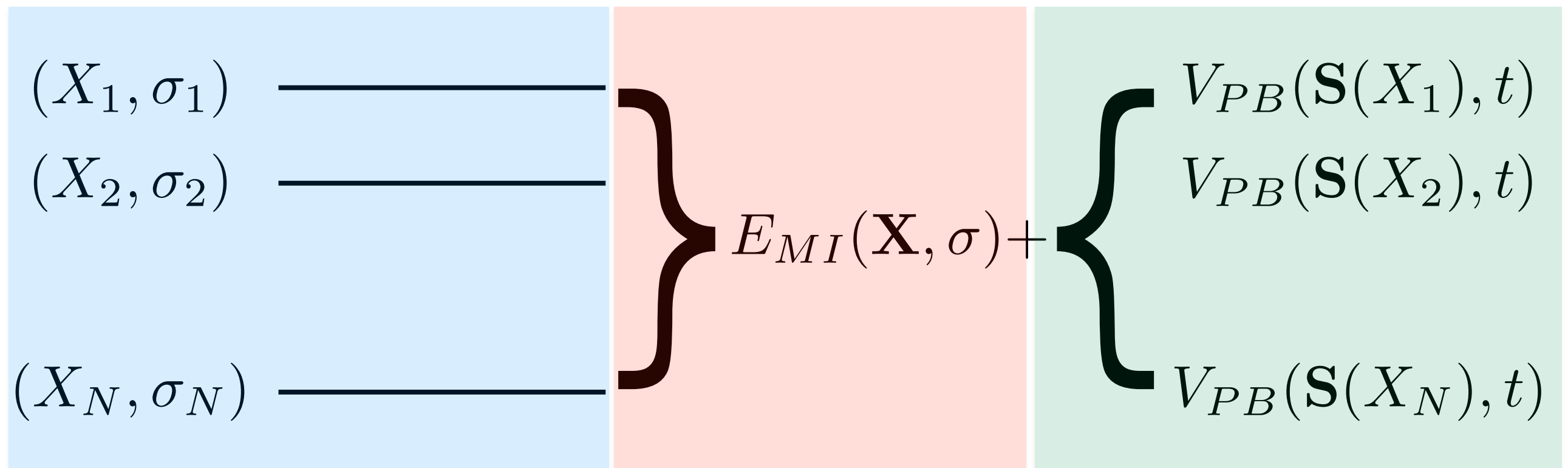


# Metadynamics Metainference

Ensemble  
of replicas

Metainference  
energy function

PBMetaD  
bias



with these additional tricks:

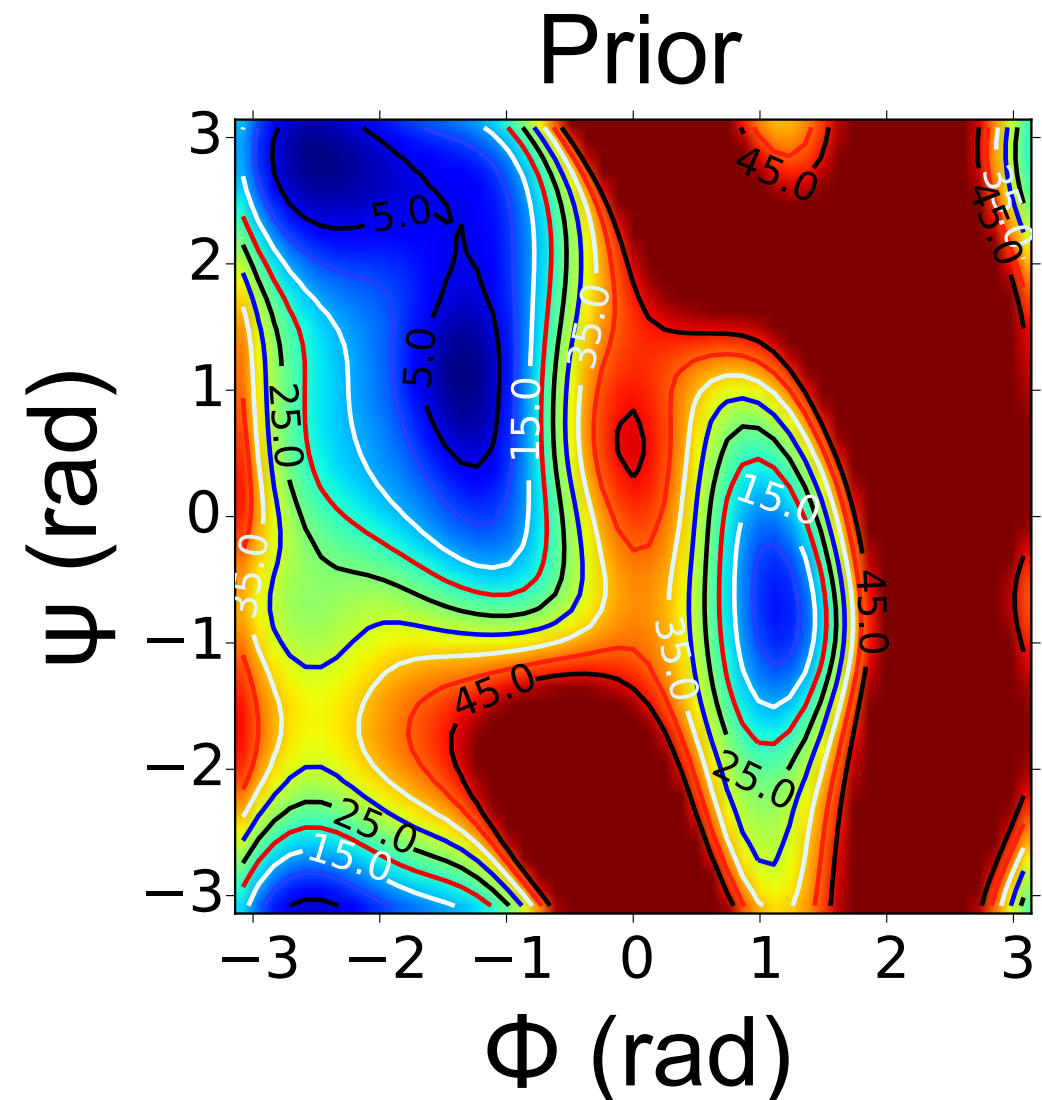
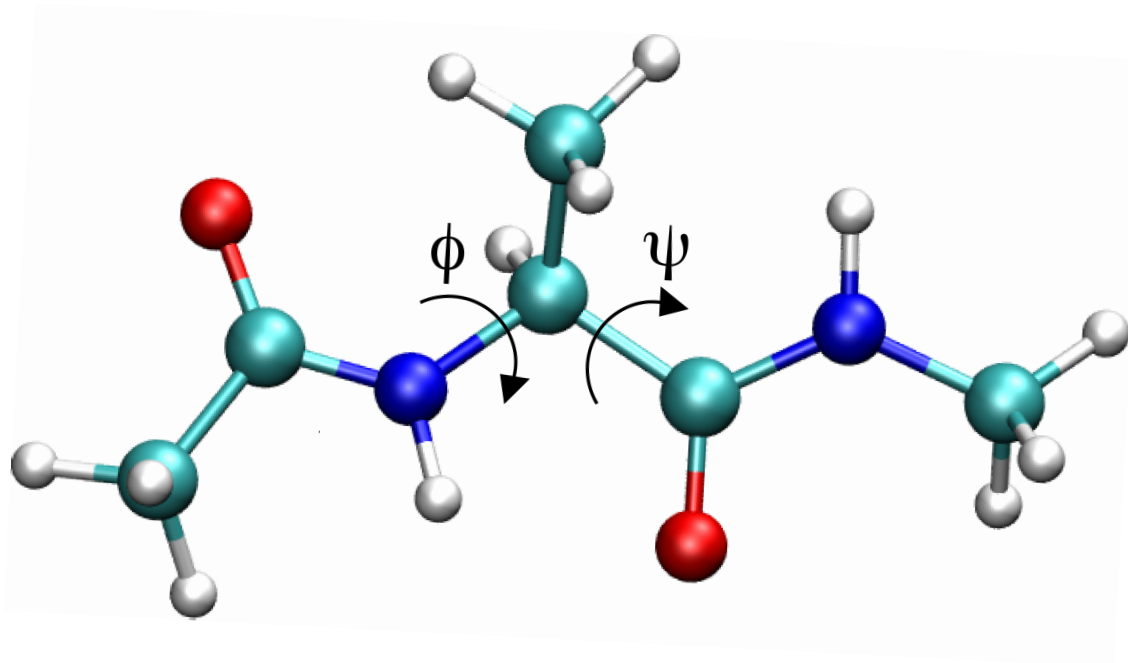
- replicas share the bias, as in multiple-walkers MetaD\*
- need to reweigh to calculate averages in the unbiased ensemble

\*Raiteri *et al.* JPCB 2006

# Benchmark

Our favorite test case: alanine dipeptide in vacuum

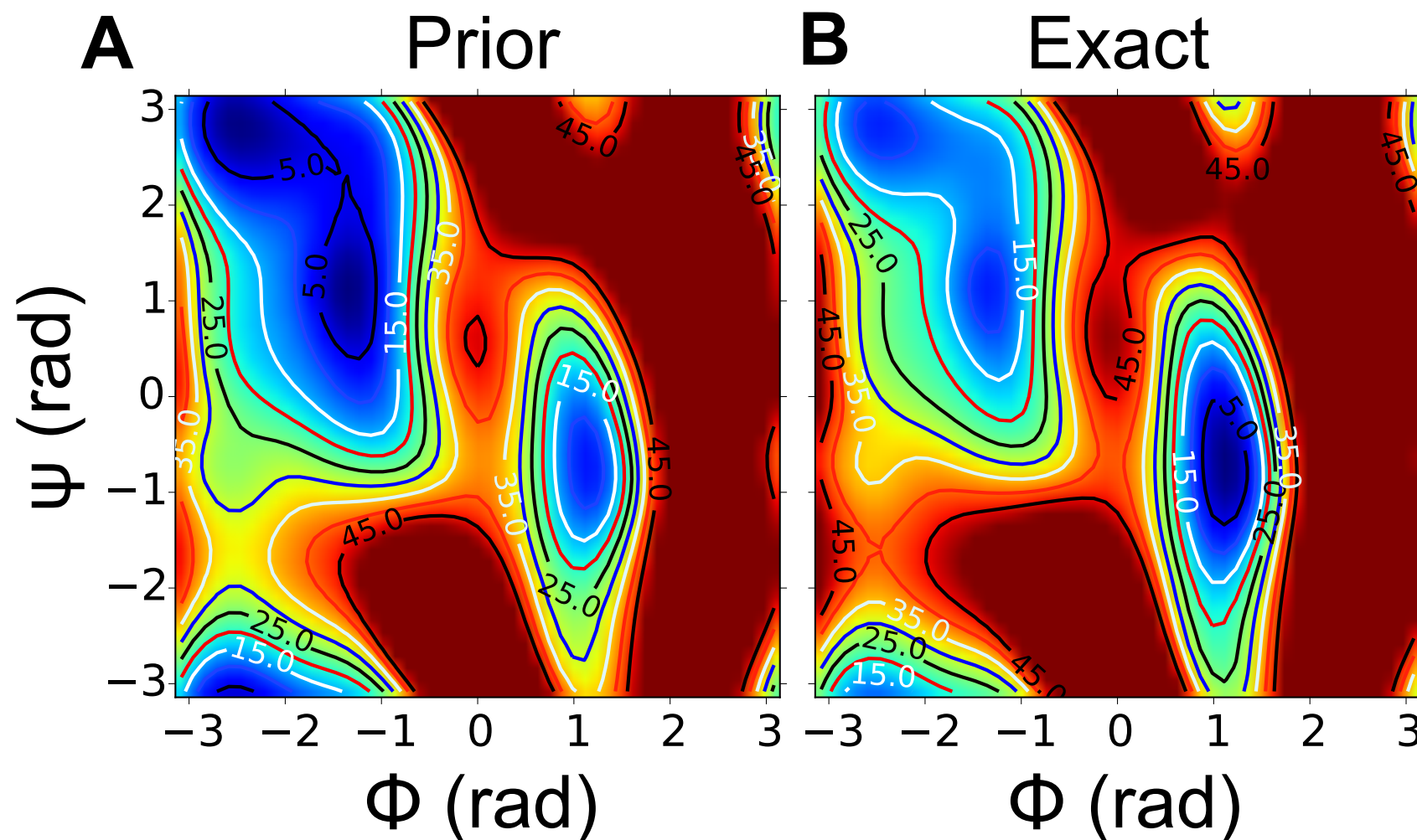
The prior information is the AMBER99SB-ILDN force field\*



\*Lindorff-Larsen et al. Proteins 2010

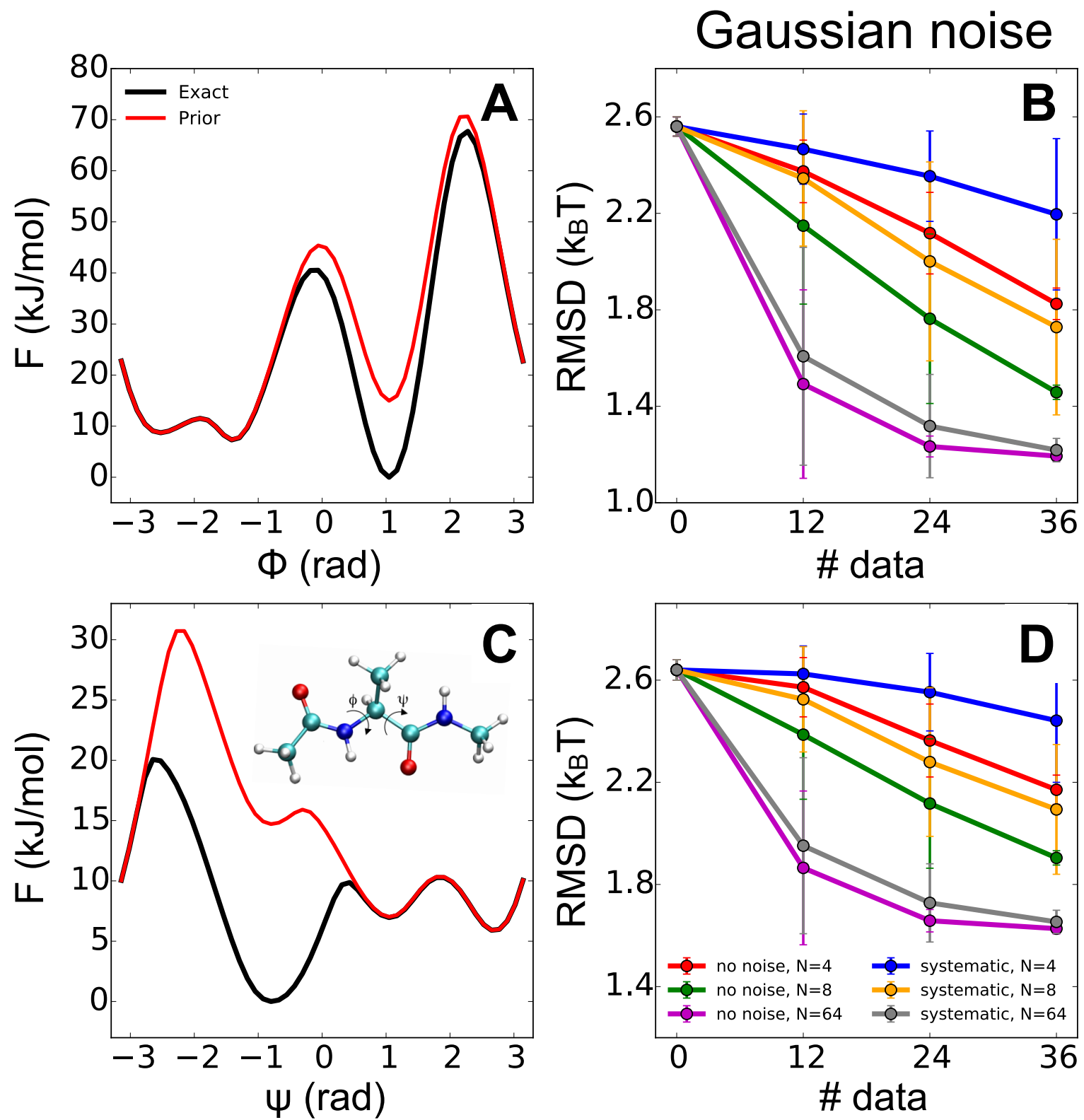
# Benchmark

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

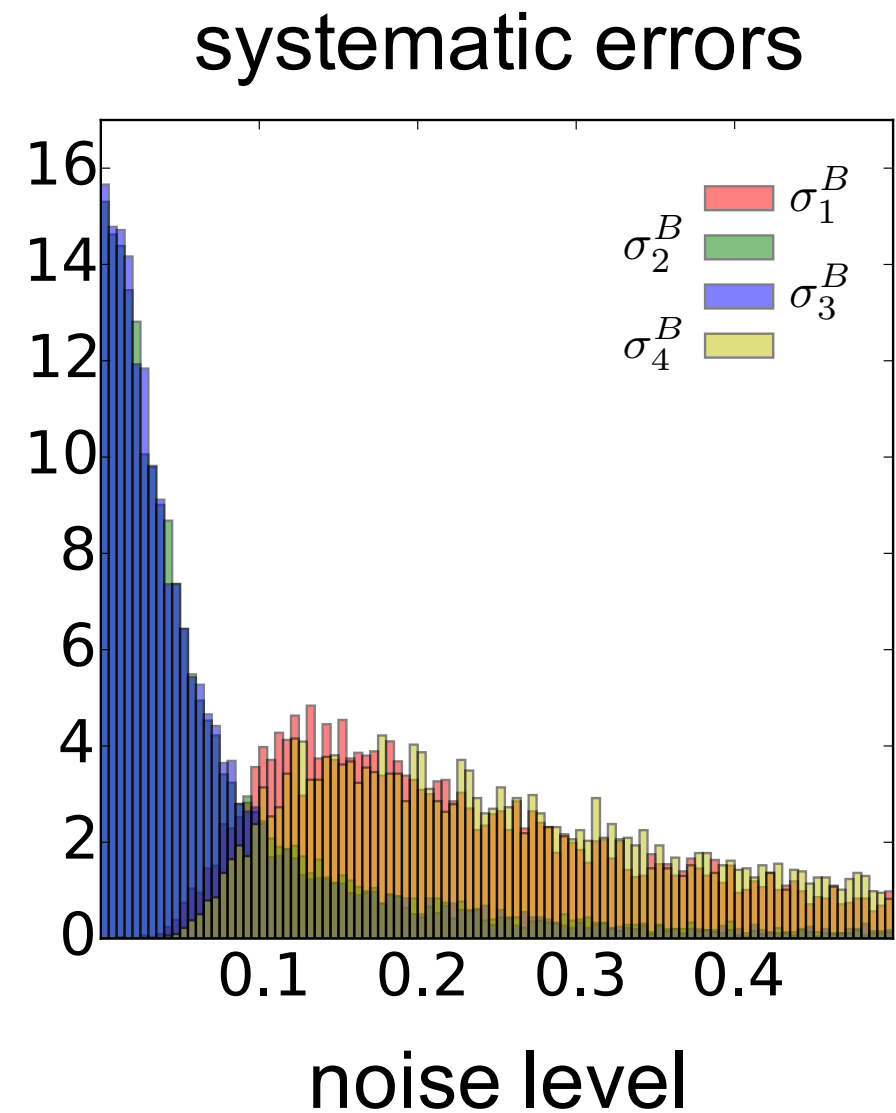
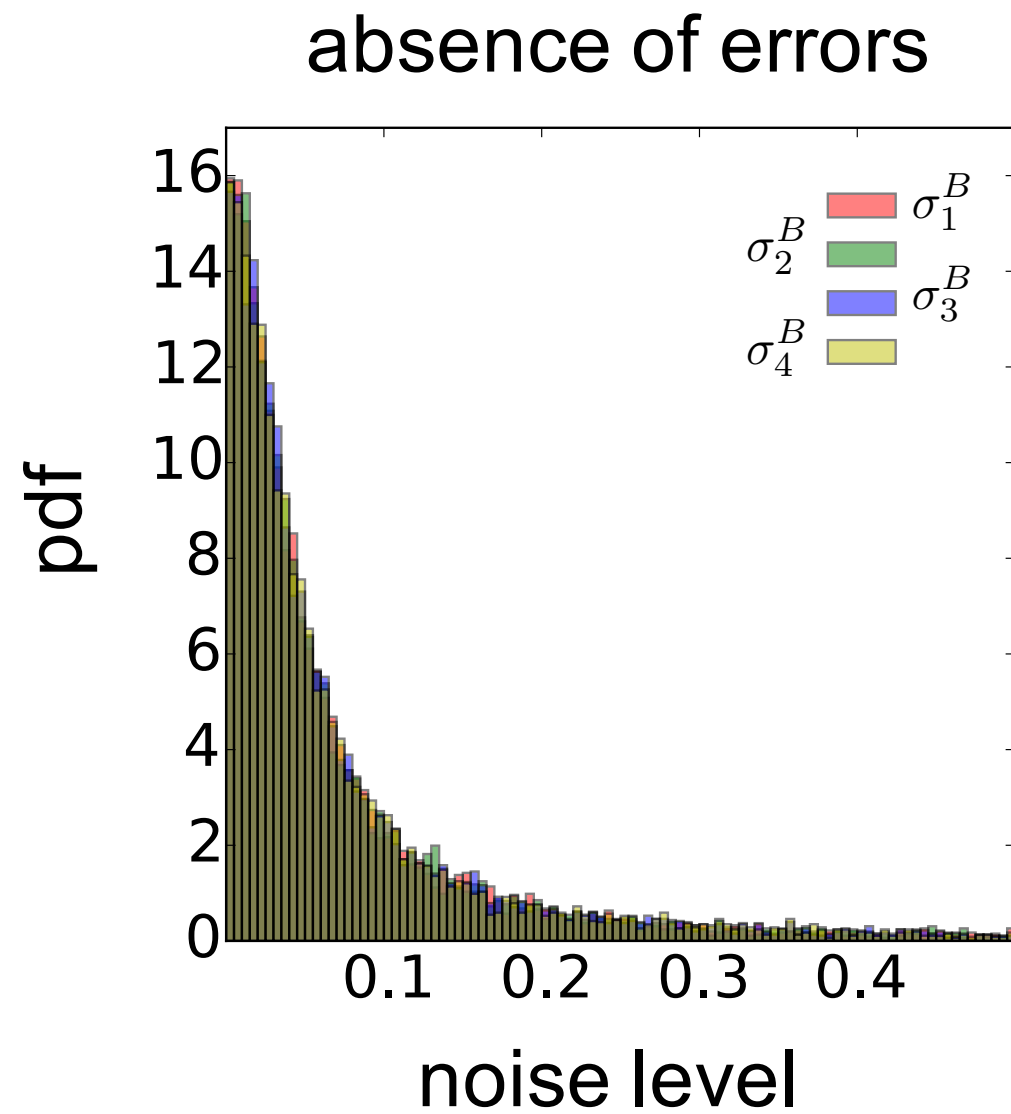


We introduce synthetic experimental data as average distances between heavy atoms, calculated in the exact ensemble, + **noise**

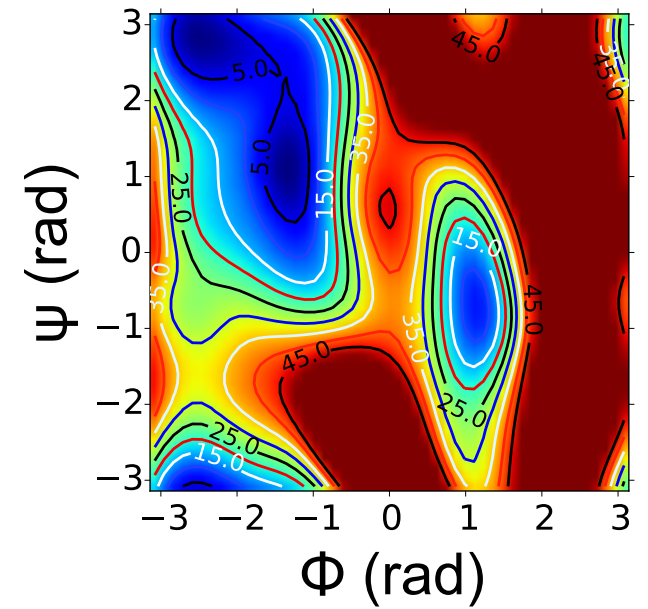
# Results



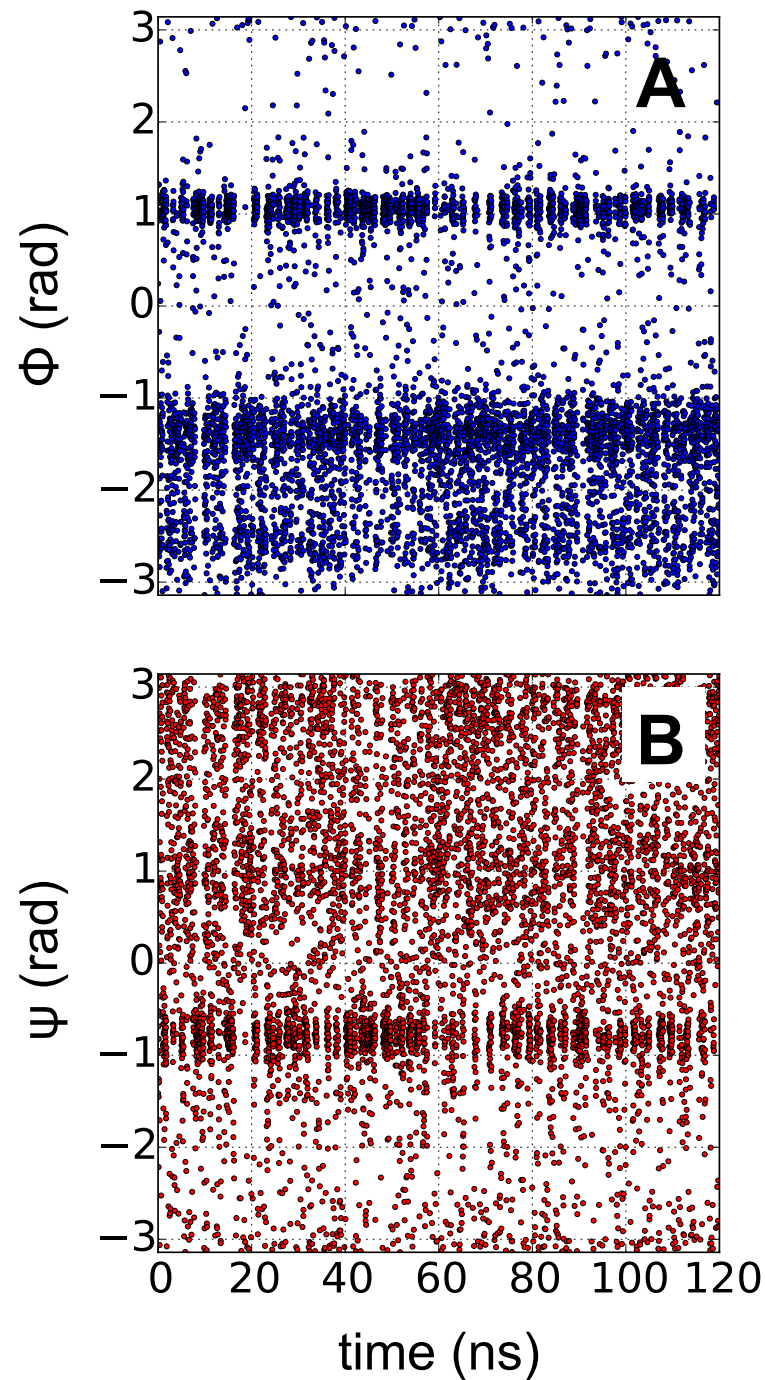
# Noise inference



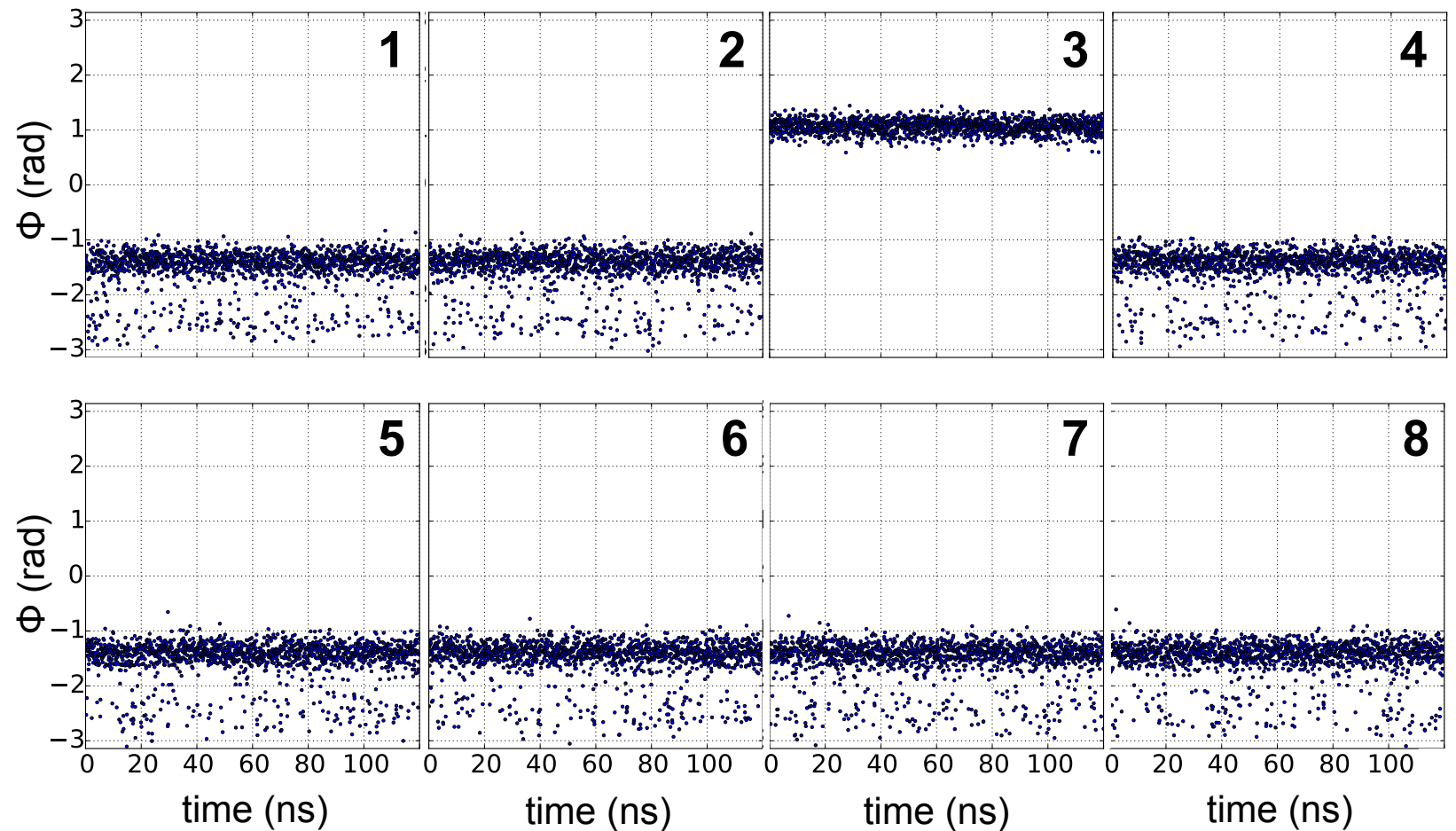
# Metainference alone



with PBMetaD



without PBMetaD

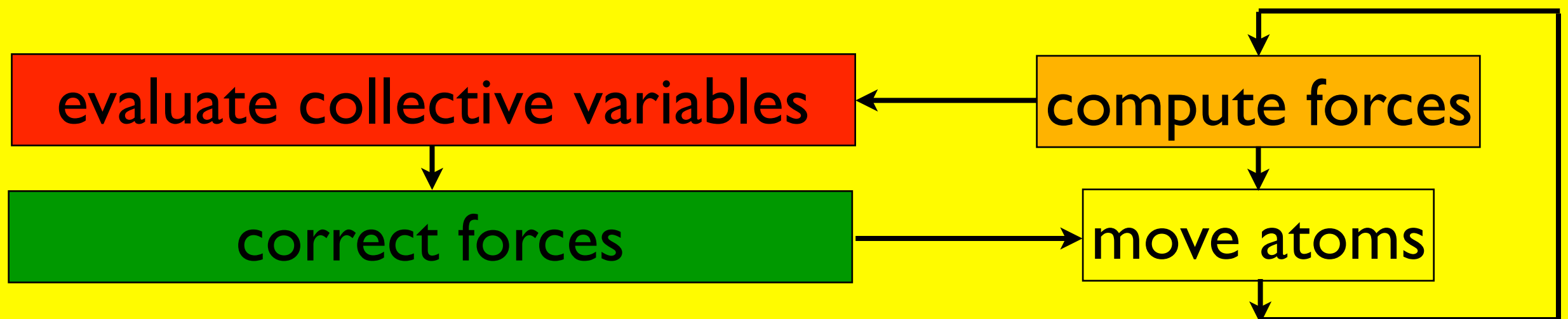




# The implementation

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

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



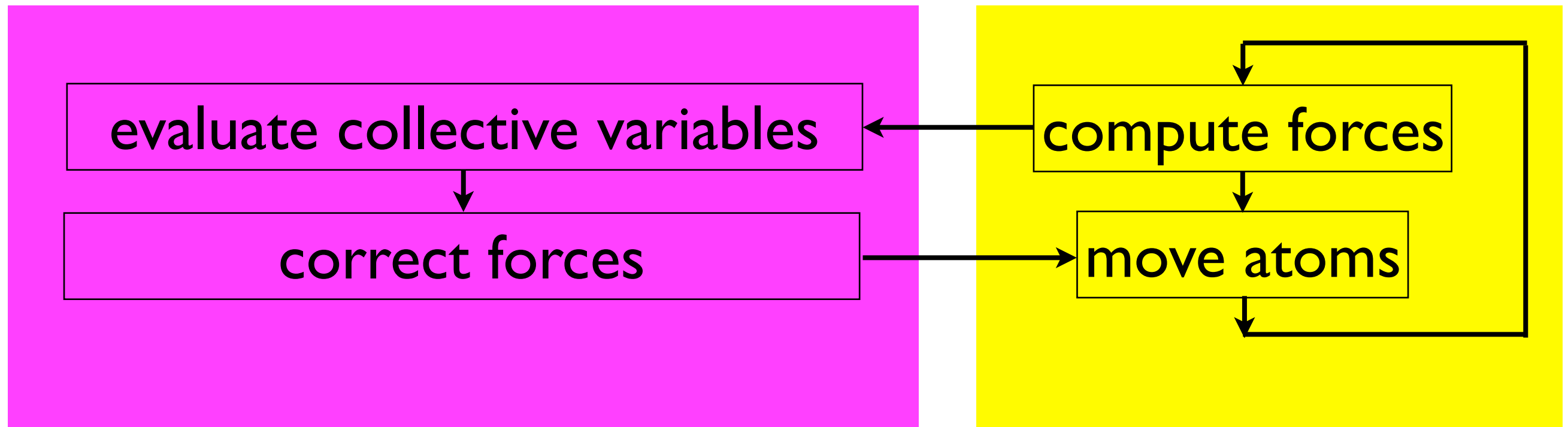
several possible algorithms  
e.g. umbrella sampling, metadynamics, ...



# PLUMED

## PLUGIN

## MD code

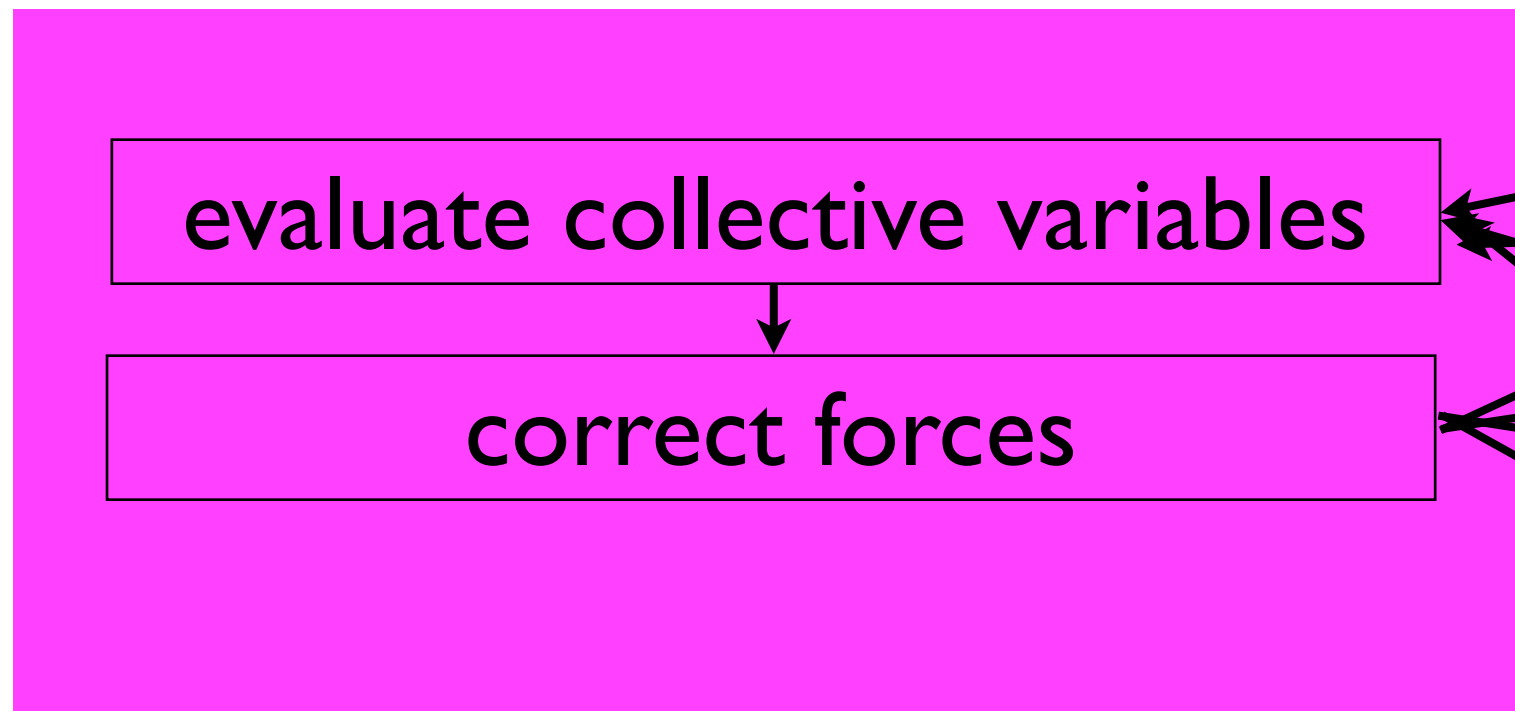


Bonomi *et al.* CPC 2008

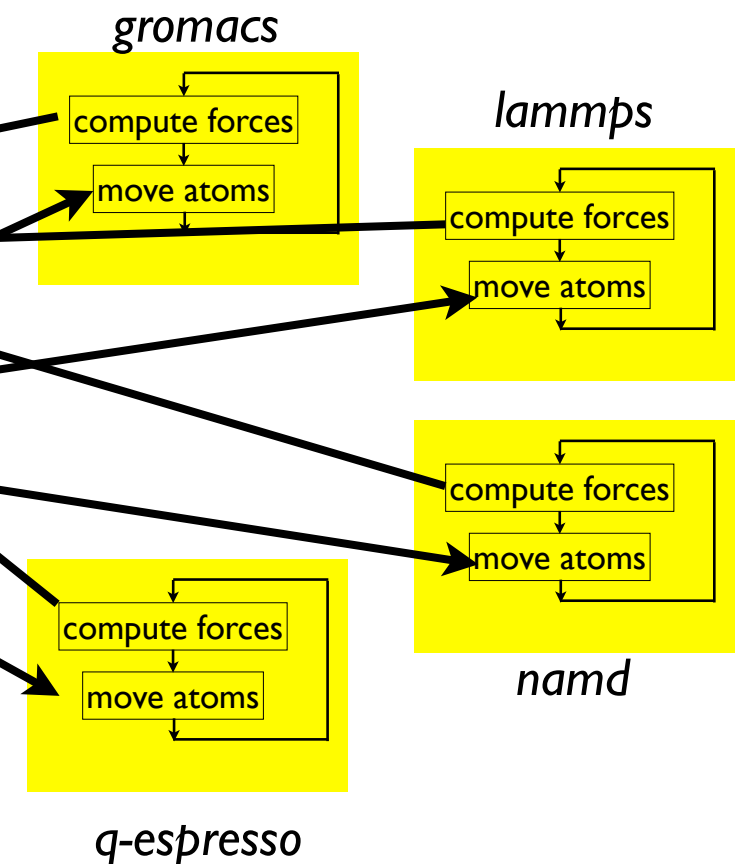
Tribello *et al.* CPC 2014

# PLUMED

## PLUGIN



## MD codes



One open source plugging  
for several MD codes!

Why **PLUMED**?

**PLU**gin for **ME**ta**D**ynamics

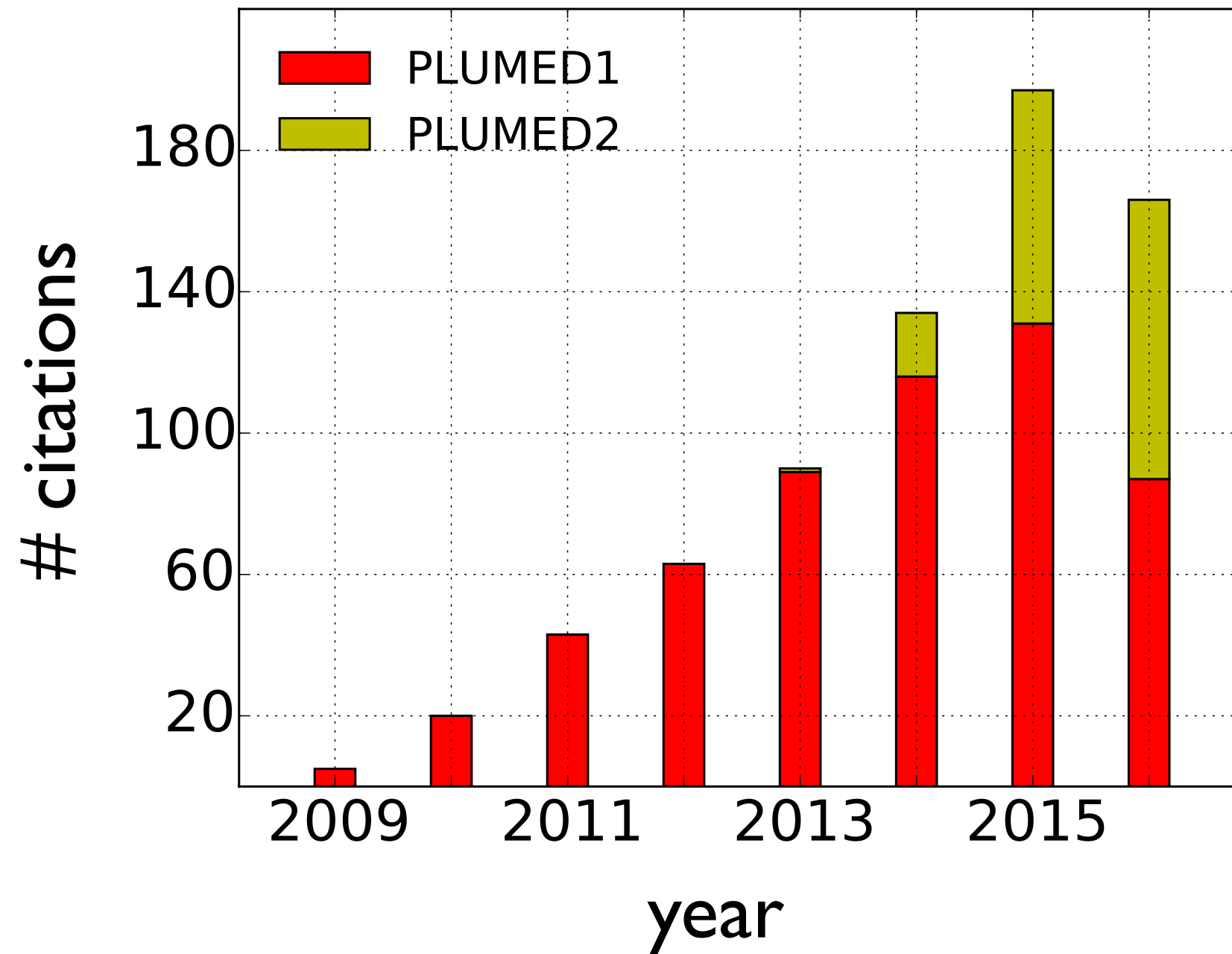
**PLU**gin for free-energy **ME**tho**D**s

**PLU**gin for **MO**lecular **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<sup>\$</sup>

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

## Analyze simulations on the fly<sup>\*</sup>

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

## Bias simulations on the fly<sup>\*</sup>

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

<sup>\$</sup>from command line or from VMD - Giorgino, CPC (2014), [http://github.com/tonigi/vmd\\_plumed](http://github.com/tonigi/vmd_plumed)

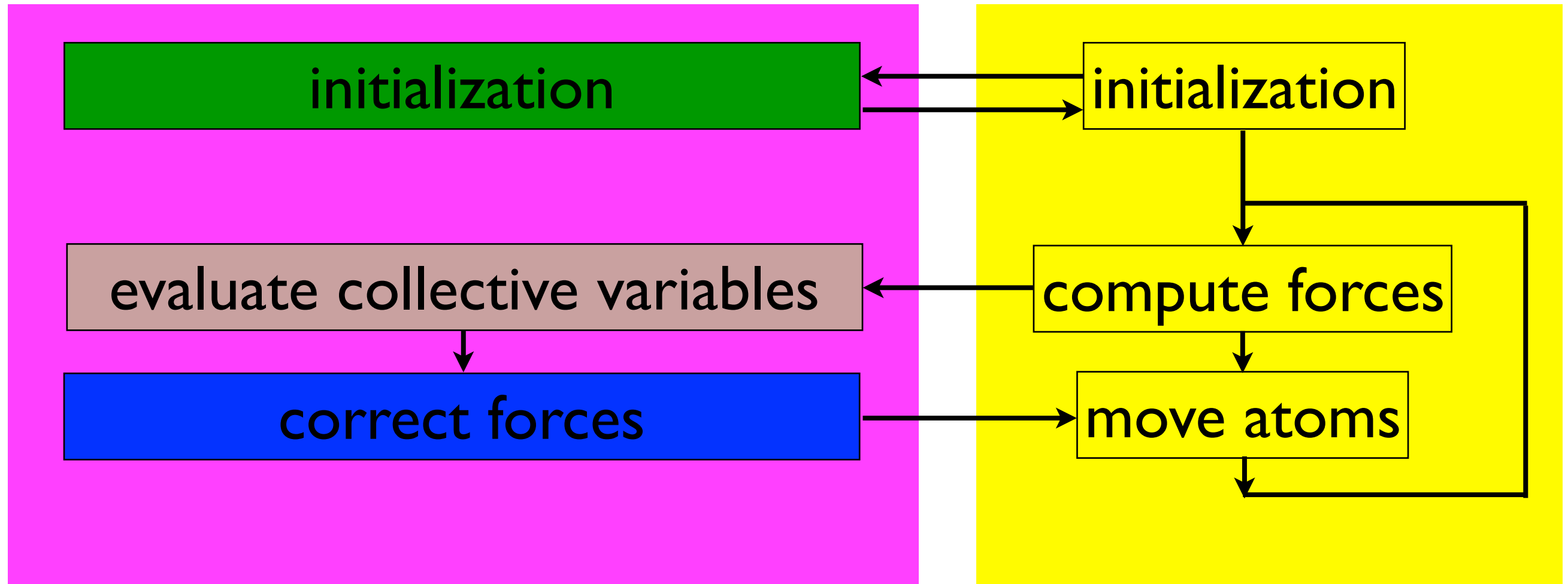
<sup>\*</sup>used in combination with a supported MD engine, e.g.  
GROMACS, NAMD, LAMMPS, Q-ESPRESSO, AMBER + others

# PLUMED+MD

PLUMED

read from a separate file

MD code



also derivatives w.r.t. atom positions

sometime using history-dependent schemes

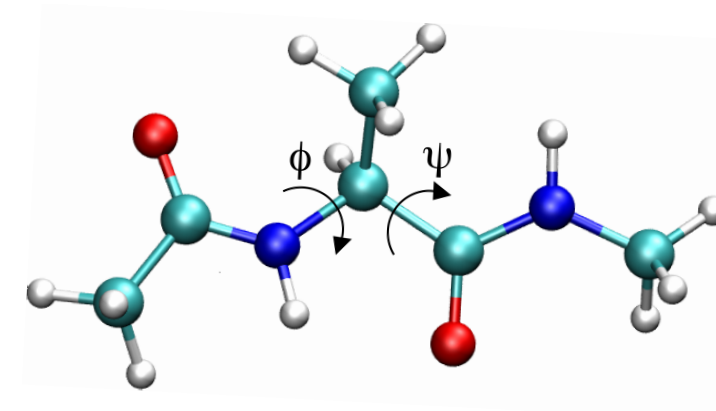
# Example of PLUMED input file

CV

```
# collective variables definition
```

```
phi:  TORSION ATOMS=5,7,9,15
```

```
psi:  TORSION ATOMS=7,9,15,17
```



```
# 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
```

```
# printout
```

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

BIAS

OUTPUT

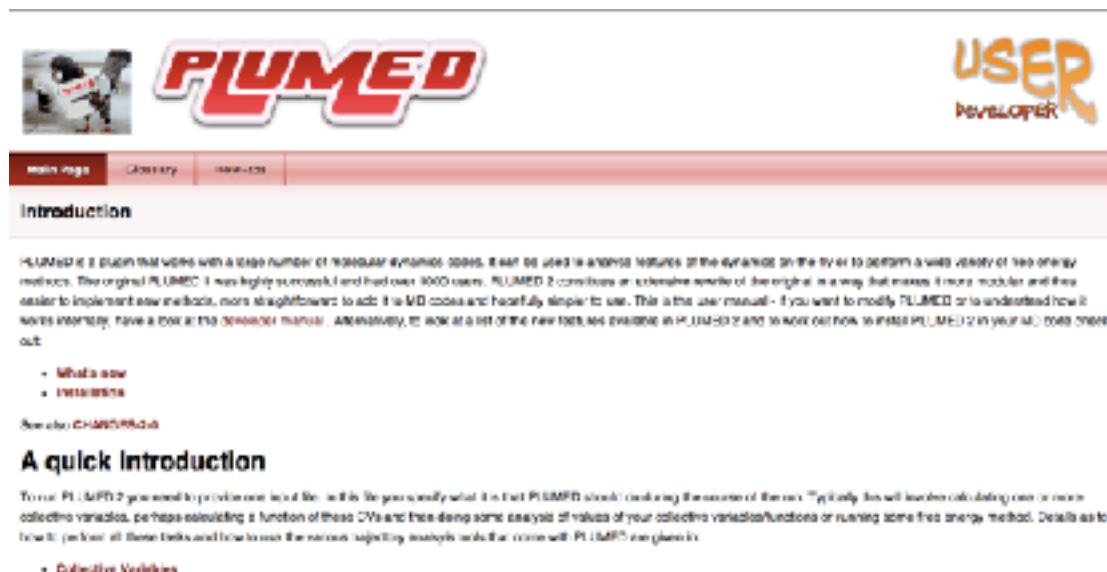
# On the WEB

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

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

User & developer mailing lists

User & developer manuals + tutorials



The screenshot shows the PLUMED User Manual website. At the top, there is a navigation bar with links for "main page", "About", and "Contact". The main content area is titled "Introduction" and contains a paragraph about PLUMED 2.0, its features, and how to use it. Below the introduction, there is a section titled "A quick introduction" which provides a brief overview of the software's capabilities and how to get started.

**PLUMED** USER DEVELOPER

main page About Contact

### Introduction

PLUMED 2.0 is a plugin that works with a large number of molecular dynamics codes. It can be used to analyse features of the dynamics on the fly or to perform a wide variety of free energy methods. The original PLUMED 1 was highly successful and had over 1000 users. PLUMED 2 provides an extensive rewrite of the original in a way that makes it more modular and thus easier to implement new methods, more straightforward to add it to MD codes and hopefully simpler to use. This is the user manual - if you want to modify PLUMED or to understand how it works internally, have a look at the [developer manual](#). Alternatively, to look at a list of the new features available in PLUMED 2 and to work out how to install PLUMED 2 in your MD code check out:

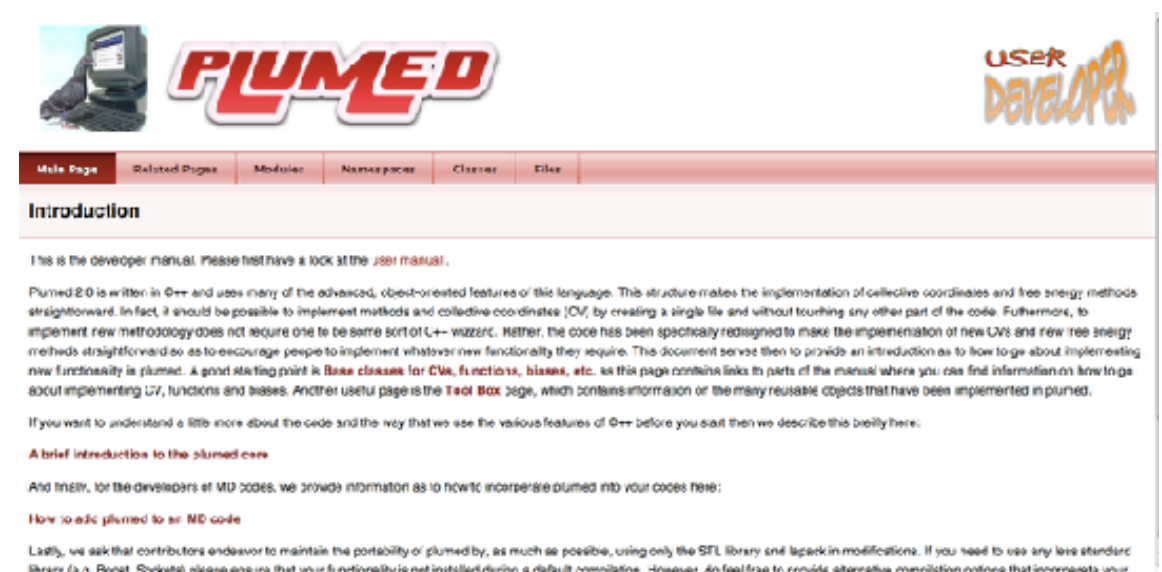
- [What's new](#)
- [Installation](#)

Author: CHANGPENG

### A quick introduction

To use PLUMED 2 you need to provide your input file. In this file you specify what it is that PLUMED should calculate (the name of the method) and how to calculate it (the name of the method). For example, to calculate the free energy of a collective variable, perhaps calculating a function of these CVs and then doing some analysis of values of your collective variables/functions or running some free energy method. Details as to how to perform all these tasks and how to use the various sampling analysis modules come with PLUMED are given in:

- [Collective Variables](#)



The screenshot shows the PLUMED Developer Manual website. At the top, there is a navigation bar with links for "Main Page", "Related Pages", "Module", "Namespace", "Class", and "File". The main content area is titled "Introduction" and contains a paragraph about the developer manual, its purpose, and how to use it. Below the introduction, there is a section titled "A brief introduction to the plumed core" which provides a brief overview of the software's internal structure and how to get started.

**PLUMED** USER DEVELOPER

Main Page Related Pages Module Namespace Class File

### Introduction

This is the developer manual. Please first have a look at the [JSDP manual](#).

Plumed 2.0 is written in C++ and uses many of the advanced, object-oriented features of this language. This structure makes the implementation of collective coordinates and free energy methods straightforward. In fact, it should be possible to implement methods and collective coordinates (CV) by creating a single file and without touching any other part of the code. Furthermore, to implement new methodology does not require one to be some sort of C++ wizard. Rather, the code has been specially redesigned to make the implementation of new CVs and new free energy methods straightforward so as to encourage people to implement whatever new functionality they require. This document serves then to provide an introduction as to how to go about implementing new functionality in plumed. A good starting point is [Base classes for CVs, functions, biases, etc.](#) as this page contains links to parts of the manual where you can find information on how to go about implementing CV, functions and biases. Another useful page is the [Tool Box](#) page, which contains information on the many reusable objects that have been implemented in plumed.

If you want to understand a little more about the code and the way that we use the various features of C++ before you start then we describe this briefly here:

### A brief introduction to the plumed core

And finally, for the developers of MD codes, we provide information as to how to incorporate plumed into your codes here:

### How to add plumed to an MD code

Lastly, we ask that contributors endeavor to maintain the portability of plumed by, as much as possible, using only the STL library and to keep in mind that modifications. If you need to use any less standard library (e.g. Boost, Sockets) please ensure that your functionality is not installed during a default compilation. However, do feel free to provide alternative compilation options that incorporate your

# 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

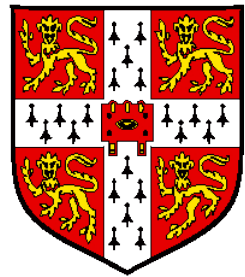
M&M enables modelling ensemble of states separated by high free-energy barriers, using noisy and ensemble-average data

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



Carlo Camilloni  
Andrea Cavalli  
Gabi Heller  
Francesco Aprile  
Michele Vendruscolo

Paolo Arosio  
Thomas Muller  
Tuomas Knowles

All of you for your attention!



UNIVERSITY *of* WASHINGTON

Jim Pfaendtner



Institut Pasteur

Samuel Hanot  
Riccardo Pellarin



Davide Branduardi  
Giovanni Bussi  
Gareth Tribello

# Tutorial instructions



[plumed.github.io/doc-v2.3/user-doc/html/cineca.html](https://plumed.github.io/doc-v2.3/user-doc/html/cineca.html)