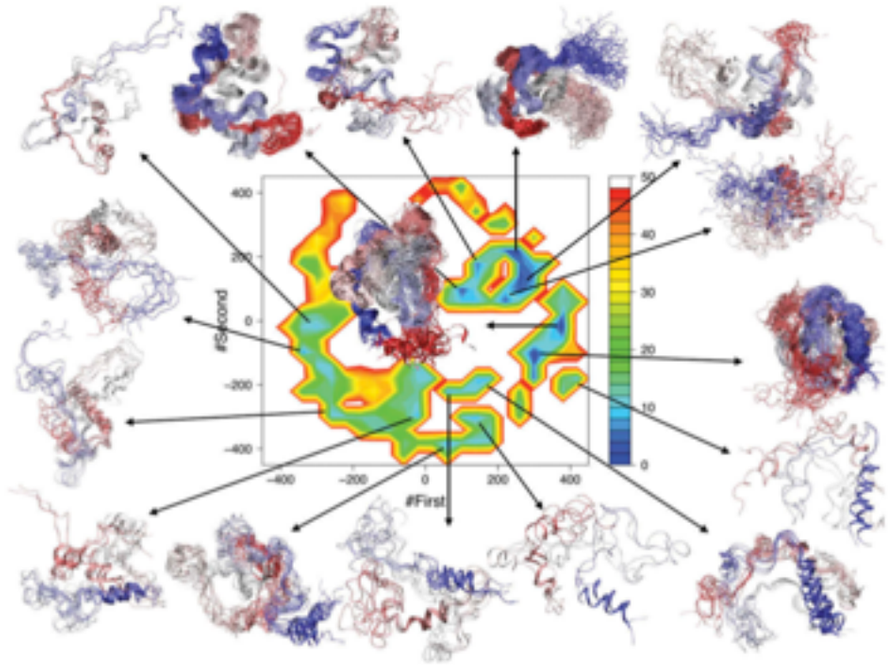
General principles of enhanced-sampling methods based on bias potentials

Carlo Camilloni



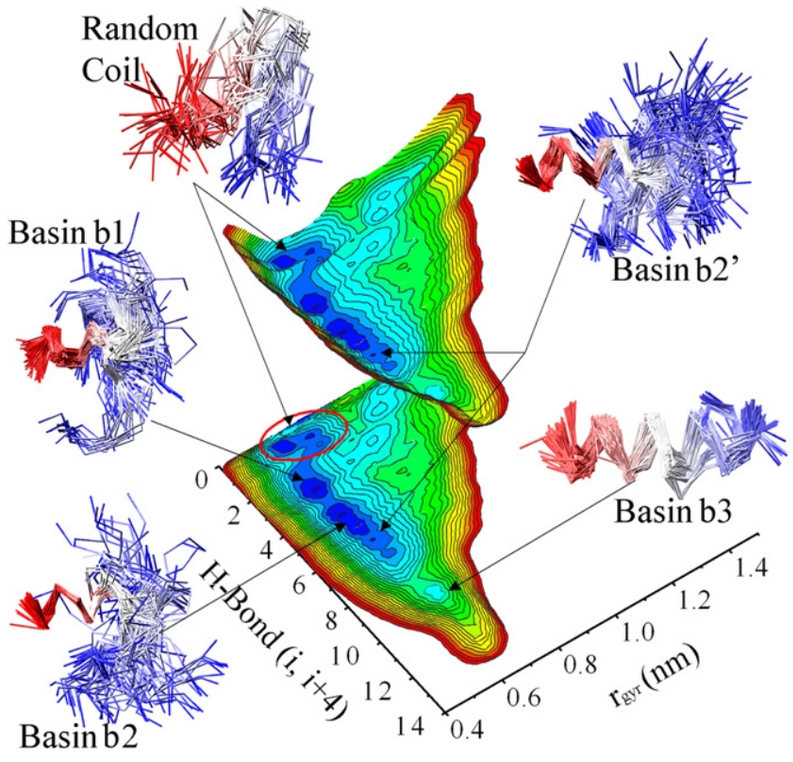
Literature

1. Bussi, G.; Tribello, G. **Analyzing and biasing simulations with PLUMED.** In: Bonomi, M.; Camilloni, C. (editors). Biomolecular Simulations. Springer Methods in Molecular Biology; 2019.
2. Camilloni, C.; Pietrucci, F. **Advanced Simulation Techniques for the Thermodynamic and Kinetic Characterization of Biological Systems.** Advances in Physics: X 2018, 3 (1), 1477531.
3. Kästner, J. **Umbrella Sampling.** *WIREs Comput Mol Sci* 2011, *1* (6), 932–942.
4. Roux, B.; Souaille, M. **Extension of the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations.** *Comput. Phys. Commun.* 2001, *135* (1), 40–57.
5. Torrie, G.; Valleau, J. **Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation-Umbrella Sampling.** *J. Comput. Phys.* 1977, *23*, 187–199.

Sampling

Sampling allows to learn some information on a large population by merely looking at a comparably small number of individuals.

In the case of a protein for example we would like to now the distribution of the secondary structure populations and that of the radius of gyration and RMSD.



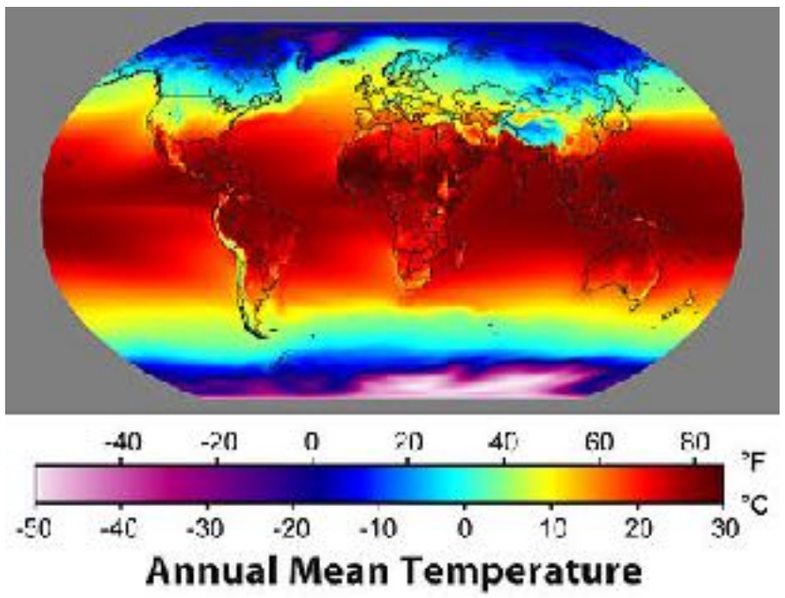
Importance vs Random Sampling

Let's say that we would like to estimate some property of the human population as a function of the latitude and longitude. By random sampling we would select a latitude and longitude value at random (uniform distribution) and check how many people are there. Most of the time the pair of numbers will point to uninhabited regions.

$$P(\theta, \phi) \propto \frac{1}{4\pi}$$

Alternatively we can use a different probability distribution to extract our latitude and longitude values, for example we could weigh more regions with an average temperature around 20C, or we could weigh more regions with an average surface density close to 3000Kg/m^3

$$P(\theta, \phi) \propto e^{-\frac{(T(\theta,\phi)-20)^2}{2\sigma_T^2}}$$



The first probability distribution will help us avoiding deserts, glaciers etc, while the second would avoid the oceans and seas. In principle we could use both probability distributions together.

More on statistical equilibrium

If a close system is in equilibrium here we mean that the average properties are the same if observed over time or over space (ensemble) or over both. Furthermore the behaviour of a single molecule in time (if the time is long enough) is the same of the behaviour of all the molecules at a given time. This is call ERGODICITY and is the basic principle that makes computer simulation useful!

This notion of equilibrium (if you want intra-molecular equilibrium) is also valid if we look at molecular interactions (inter-molecular equilibrium), that is for example the fraction of bound and unbound molecules in a solution should be the same of the fraction of time spent bound and free for a single couple of molecules.

Although dynamical measurements (movies) must agree with ensemble measurements (snapshots) the latter lack dynamical information: for example you can say in both case what is the fraction of a conformational state or of a bound state but in the first case you can also say how long molecules are in a given state, I.e. the life time or the rate of a reaction.

Non-equilibrium measurements refer to studies where the system is suddenly perturbed (by temperature, chemical agents, etc) and so average quantities will change over time.

Statistics and Mechanics

What is the connection between statistics and mechanics?

That is the Boltzmann equation: $$pdf(x,v) \equiv \rho(x,v) \propto \exp\left[\frac{-E(x,v)}{k_B T}\right] = \exp\left[\frac{-U(x)}{k_B T}\right] \exp\left[\frac{-K(v)}{k_B T}\right]$$

Let's look at this pdf in more detail:

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

$$pdf(v) \equiv \rho(v) \propto \exp\left[\frac{-K(v)}{k_B T}\right] = \exp\left[\frac{-\frac{1}{2}mv^2}{k_B T}\right]$$

Conformations and velocities are independent. This means that we can study them separately. The distribution of the velocities does not affect the distribution of the configurations. Furthermore the distribution of the velocity is Gaussian, so it is easy to integrate it.

So for all practical purposes we can just ignore the velocity:

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

This gives a link between the energy of a conformation and the probability of observing it

Statistics and Mechanics

What is the connection between statistics and mechanics?

That is the Boltzmann equation: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

The probability of a conformation (microstate) is <u>proportional</u> (exponentially) to minus its energy and inversely proportional to k_BT (Boltzmann constant times Temperature).

T (K)	T (C)	k_BT (kJ/mol)	k_BT (kcal/mol)
273.15	0	2.27	0.54
293.15	20	2.44	0.58
300.00	26.85	2.49	0.60
310.15	37	2.58	0.62

1. **Lower energy conformations are more likely than high energy ones.**
2. **Relative probabilities become more equal as temperature increases.**

In principle we can normalise it and get the complete probability distribution, but let's think of a force field, to calculate the normalisation we need to calculate the energy for all possible conformations.

We have a problem of SAMPLING.

…Mechanics

The mechanical problem of the accessible time-scales in MD is translated in statistical terms in a SAMPLING problem. To estimate probabilities we need many many conformations. Convergence becomes the problem of understanding whether the conformations we got are enough to say something relevant.

$$pdf(x) \equiv \rho(x) = \frac{\exp\left[-U(x)/k_B T\right]}{\int_V \exp\left[-U(x)/k_B T\right] dx}$$

Given a sampling we can always estimate the pdf by normalising it over the sampled conformations

$$\hat{Z} = \int_V \exp\left[-U(x)/k_B T\right] dx$$

The normalisation constant is usually called (configurational) Partition Function

$$< g >= \frac{\int_V g(x) \exp\left[-U(x)/k_B T\right] dx}{\int_V \exp\left[-U(x)/k_B T\right] dx}$$

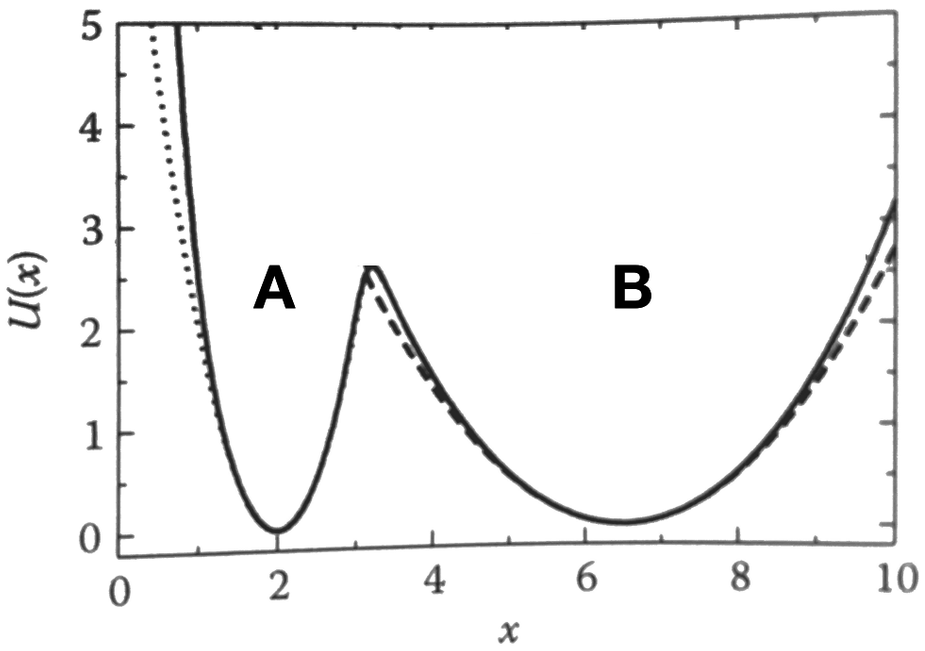
Given a sampling it is possible to calculate averages (e.g. the average RMSD, …)

In practice in MD/MC we don't need to calculate any integral. What we should know is what is the pdf we are trying to estimate (the Boltzmann distribution).

States, Probabilities and Free-Energy

State: is a collection of configurations (microstates), ideally belonging to the same potential energy basin.

A simple 1D potential energy



Here we can visually define two states A and B for the two basins, e.g. all the conformations belonging to A and to B

$$p_A = \int_{V_A} \rho(x)dx \propto \int_{V_A} \exp[-U(x)/k_BT]dx$$

$$p_B = \int_{V_B} \rho(x)dx \propto \int_{V_B} \exp[-U(x)/k_BT]dx$$

The Free-Energy is the "effective" energy of a state, i.e. the energy that will give the same probability

$$\frac{p_A}{p_B} = \frac{\int_{V_A} \exp[-U(x)/k_BT]dx}{\int_{V_B} \exp[-U(x)/k_BT]dx} \equiv \frac{\exp(-F_A/k_BT)}{\exp(-F_B/k_BT)}$$

$$F_i \propto -k_BT \ln\left(\int_{V_i} \exp[-U(x)/k_BT]dx\right)$$

Importance Sampling

$$< O(q) >_p = \int P(q)O(q)dq = \int \frac{P(q)O(q)\rho(q)}{\rho(q)}dq = < \frac{O(q)P(q)}{\rho(q)} >_\rho$$

In order to sample the former integrals we can in principle:

1. **Enumerate all the possible conformations and then calculate the needed function**
2. **Generate random conformations (with a uniform probability distribution)**
3. **Generate conformations from any other known probability distribution**

We have 'importance sampling' every time we are not using the uniform probability distribution

MC and MD are special case of importance sampling.

Importance Sampling in MD/ MC

In MC with metropolis and MD with a thermostat we are in the condition for which in the limit t->Infinity the system is sampling

$$< O(q) >_p = \int \frac{P(q)O(q)P^{MD}(q)}{P^{MD}(q)} dq = < \frac{O(q)P(q)}{P^{MD}(q)} >_{P^{MD}} = < O(q) >_{P^{MD}}$$

This assumption is correct only if the MD algorithm is correct

$$P(q) \propto \exp\left[\frac{-U(q)}{k_B T}\right]$$

Adding a bias: doing importance sampling on top of MD/MD

$$P(q) = \frac{e^{-U(q)/kT}}{\int e^{-U(q)/kT} dq}$$

Configurational probability

The problem of this quantity is that it is a bit highly dimensional (3N atoms) so not exactly intuitive.

We are usually interested in more intuitive informations like the probability to observe a radius of gyration, or an angle, or a RMSD, etc. These properties, that are functions of the positions are generically called **collective variables s(q)**

$$P(s) \propto \int dq e^{-\frac{U(q)}{k_B T}} \delta(s - s(q))$$

From this mono or few dimensional probability distribution we can now define a free energy landscape

$$F(s) = -k_B T \log P(s)$$

If the above relation is used then the free energy is estimate but for an additive constant

$$P'(s) \propto \int dq e^{-\frac{U(q)+V(s(q))}{k_B T}} \delta(s - s(q)) \propto e^{-\frac{V(s(q))}{k_B T}} P(s)$$

$$F'(s) = -k_B T \log P'(s) = F(s) + V(s) + C$$

$$P(q) \propto P'(q) e^{\frac{V(s(q))}{k_B T}} \qquad w \propto e^{\frac{V(s(q))}{k_B T}}$$

The addition of a biasing potential reweighs the probability of observing all conformations and has an additive effect on the free energy

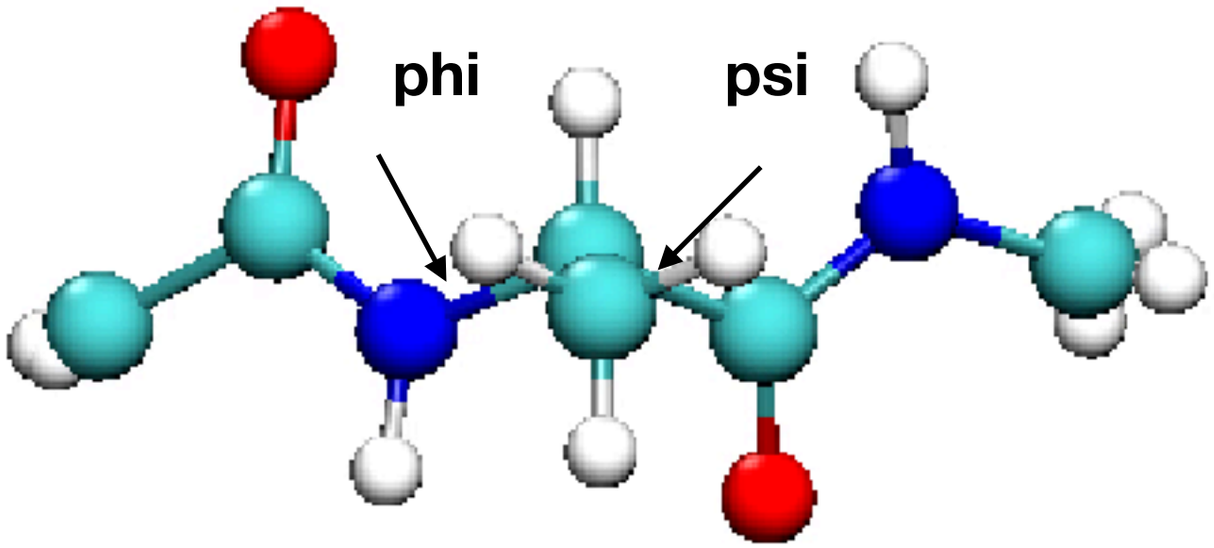
If the bias is constant in time and the sampling is complete it is possible to remove its effect by assigning a new weight to each sampled frame.

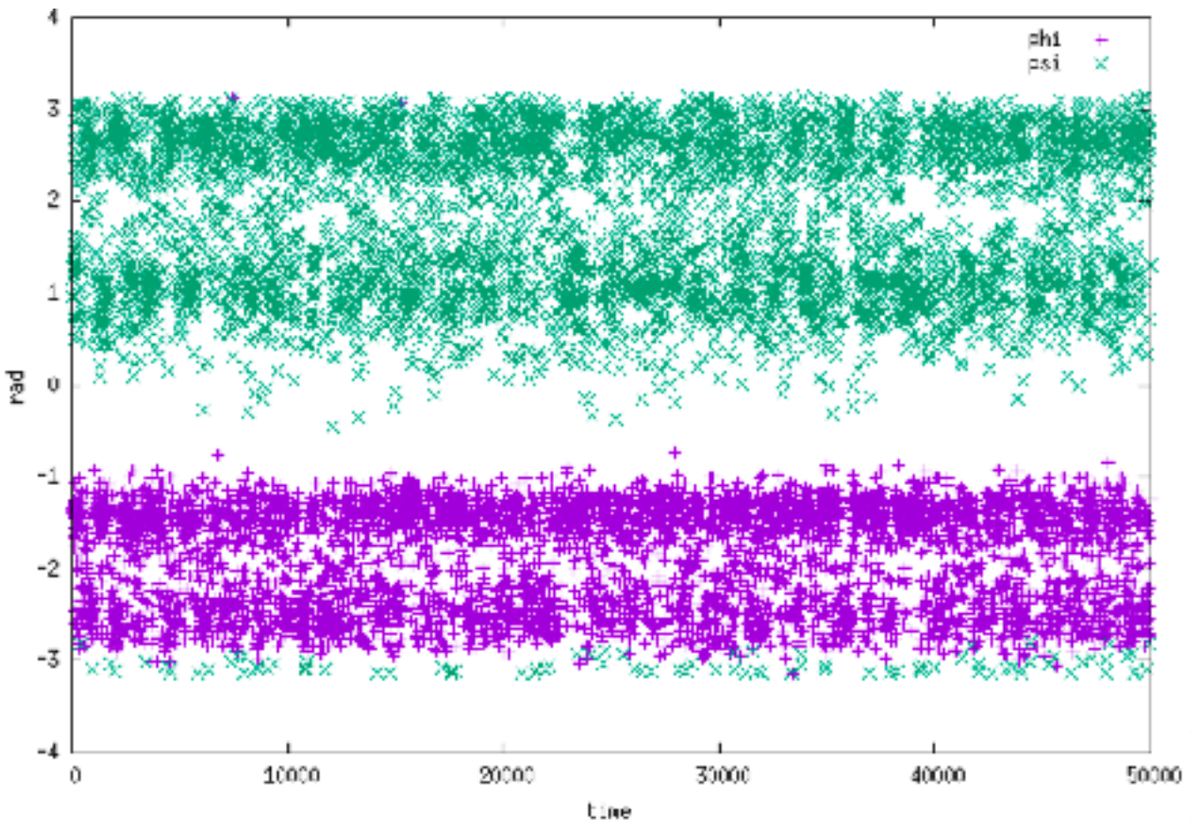
Umbrella Sampling

The original idea of umbrella sampling is that of building a probability distribution that will help sampling the wanted integral. Such a probability should bridge the unknown free energy with an optimal distribution keeping a good overlap with both. The optimal solution is to use the free energy itself.

In this way one could do random sampling in optimally selected regions of the phase space. For example within a range of free energies high enough to just overcome a free energy barrier.

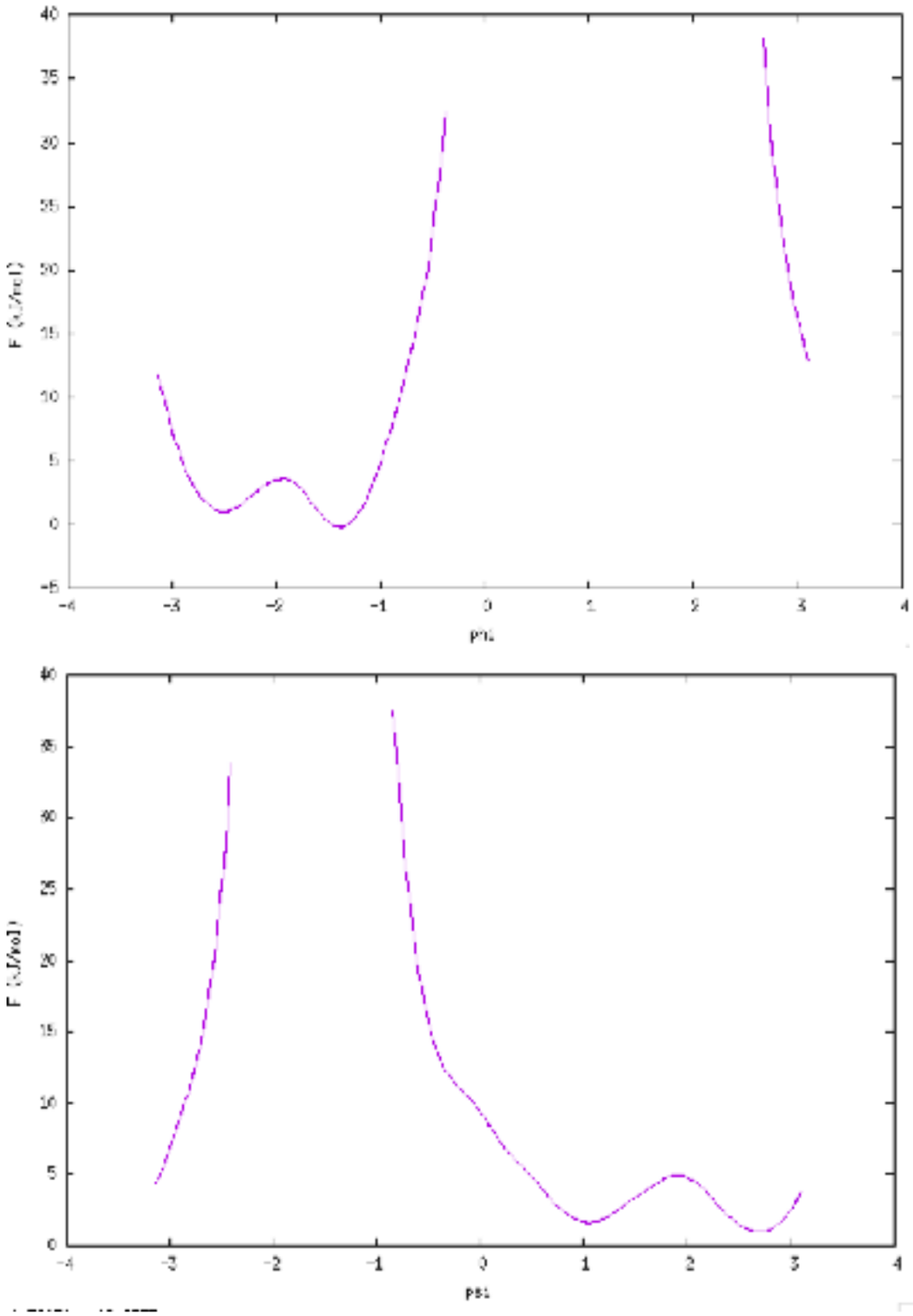
Alanine Dipeptide

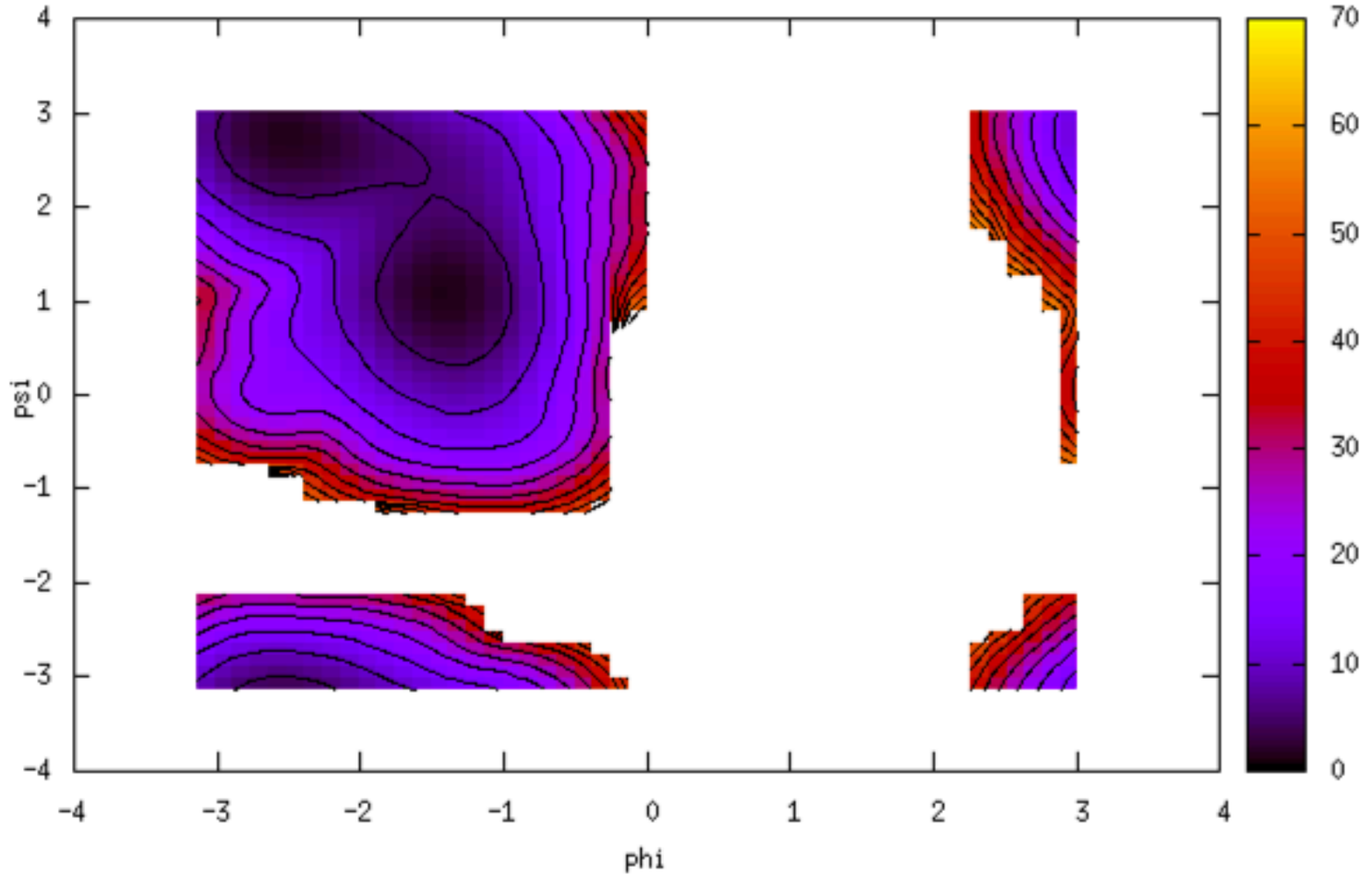





```
# vim:ft=plumed
MOLINFO STRUCTURE=../aladip.pdb

phi: TORSION ATOMS=@phi-2
psi: TORSION ATOMS=@psi-2
hhphi: HISTOGRAM ARG=phi STRIDE=10 GRID_MIN=-pi GRID_MAX=pi GRID_BIN=200 BANDWIDTH=0.1
hhpsi: HISTOGRAM ARG=psi STRIDE=10 GRID_MIN=-pi GRID_MAX=pi GRID_BIN=200 BANDWIDTH=0.1
hh: HISTOGRAM ARG=phi,psi STRIDE=10 GRID_MIN=-pi,-pi GRID_MAX=pi,pi GRID_BIN=50,50 BANDWIDTH=0.2,0.2
ffphi: CONVERT_TO_FES GRID=hhphi TEMP=298
ffpsi: CONVERT_TO_FES GRID=hhpsi TEMP=298
ff: CONVERT_TO_FES GRID=hh TEMP=298
DUMPGRID GRID=ffphi FILE=ffphi.dat
DUMPGRID GRID=ffpsi FILE=ffpsi.dat
DUMPGRID GRID=ff FILE=ff2d.dat
PRINT ARG=phi,psi FILE=colvar.dat STRIDE=10
```

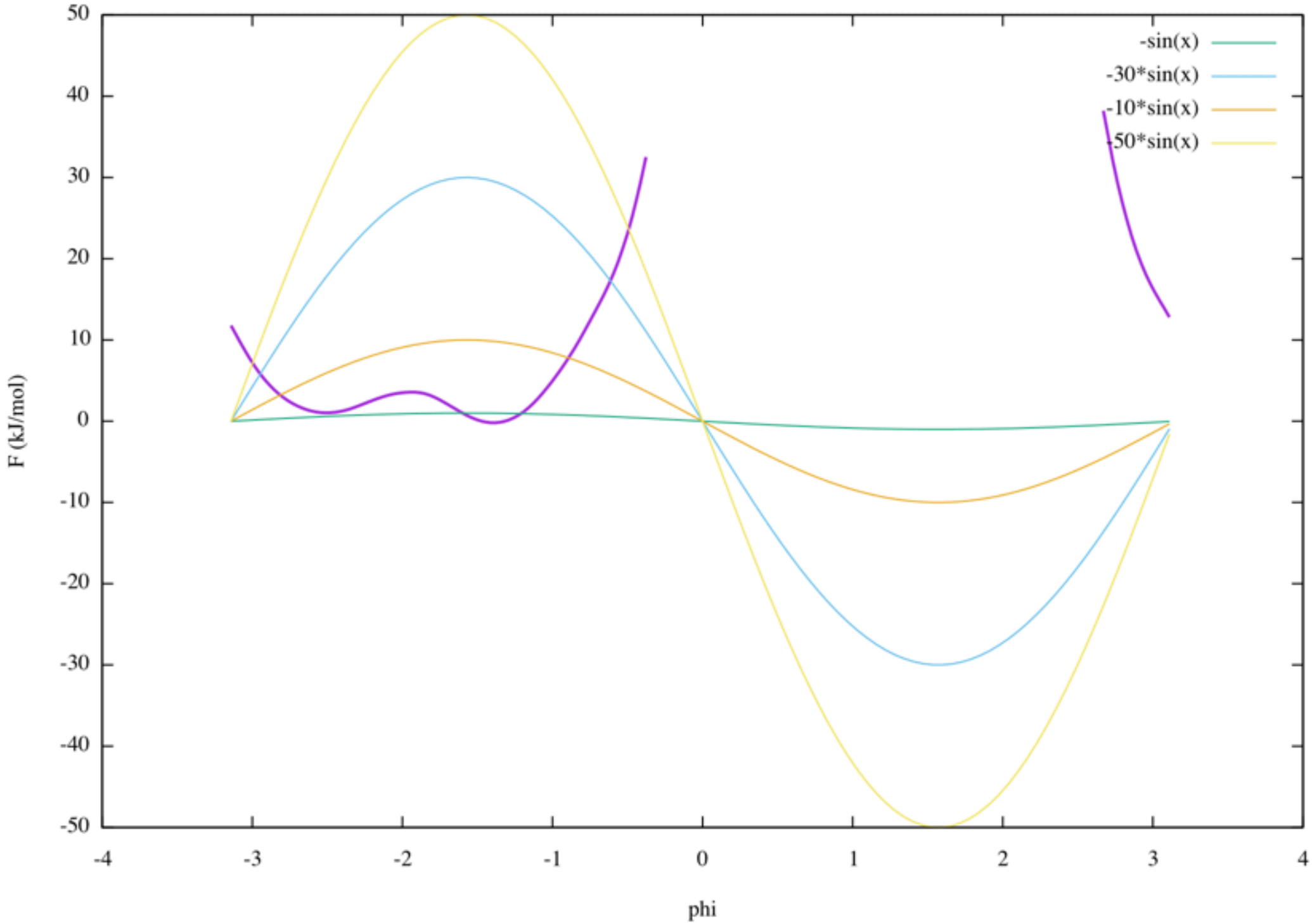
Alanine Dipeptide





Since these are periodic collective variables we could use a new ‘importance sampling’ defined as -sin(phi)

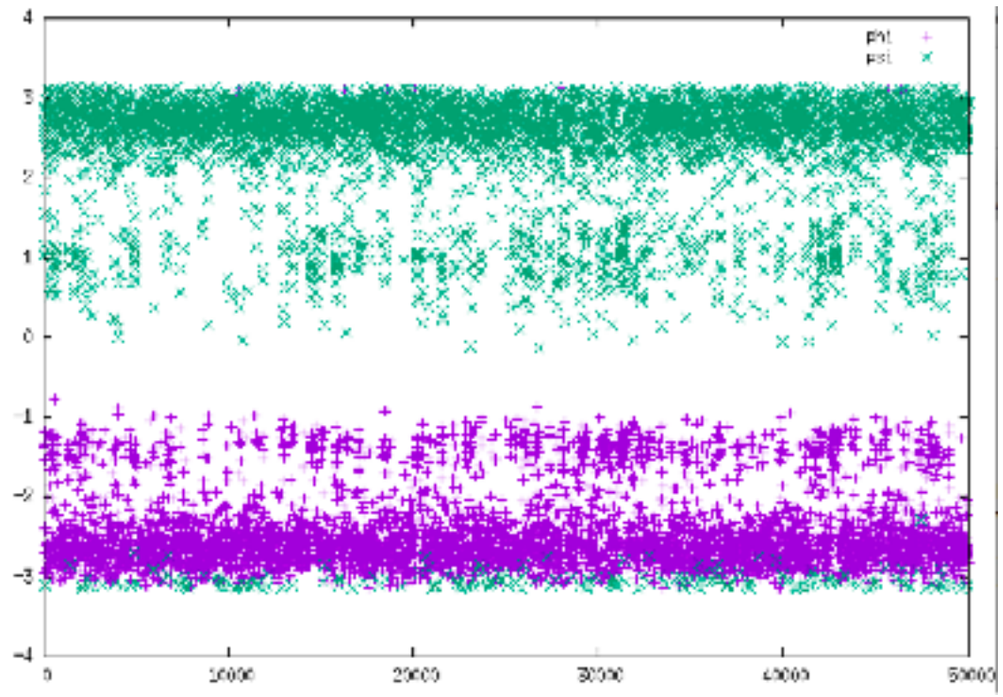
Alanine Dipeptide



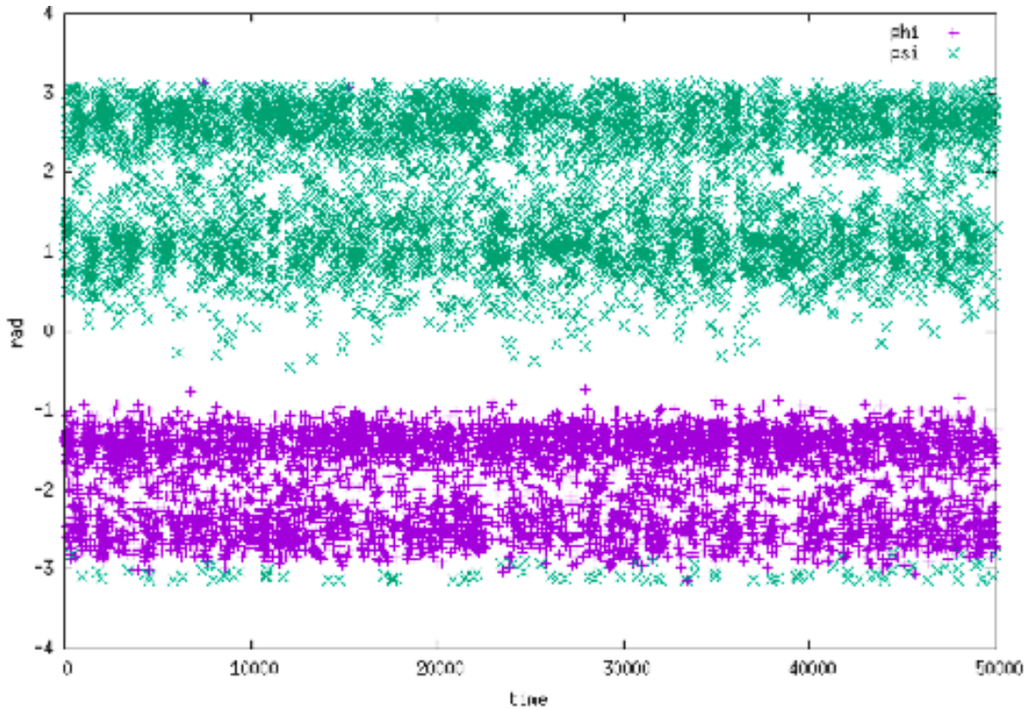
Adding a bias:

```
# vim:ft=plumed
MOLINFO STRUCTURE=aladip.pdb

phi: TORSION ATOMS=@phi-2

MATHEVAL ...
ARG=phi
LABEL=doubleg
FUNC=-10*sin(x)
PERIODIC=NO
... MATHEVAL

b: BIASVALUE ARG=doubleg

PRINT ARG=phi FILE=phi.dat STRIDE=10
```

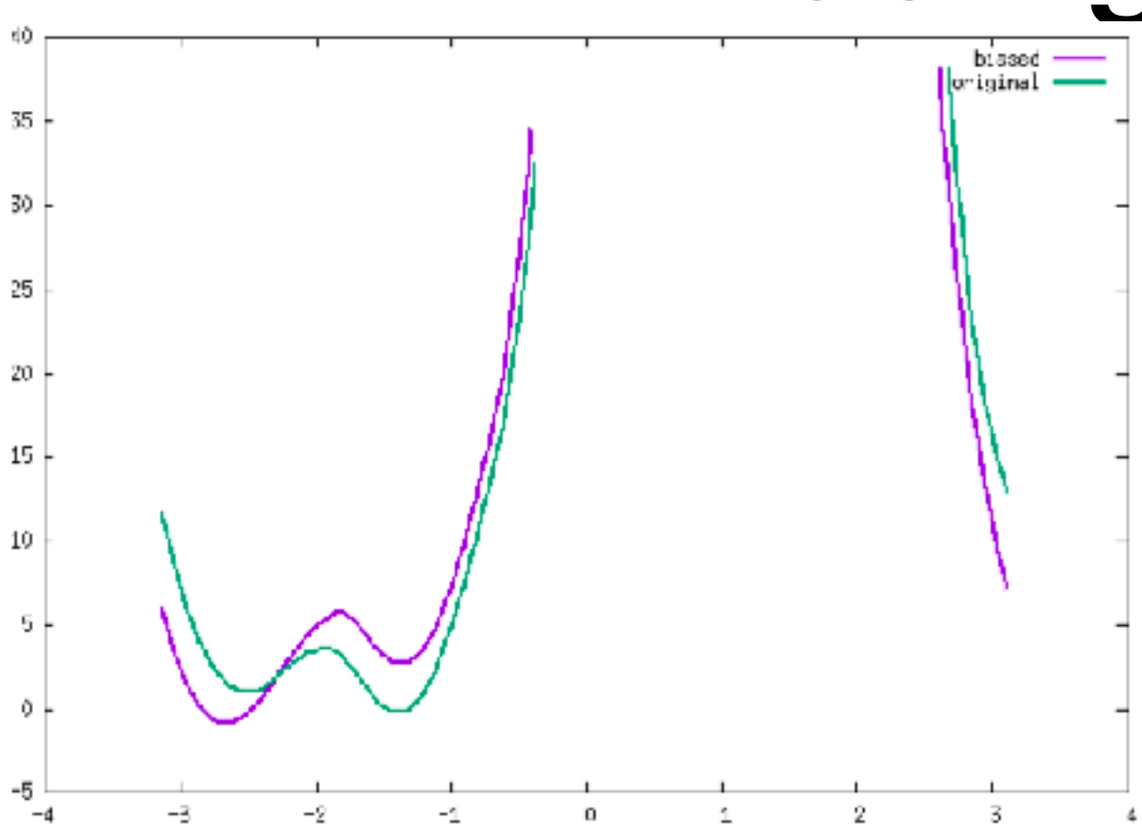
Biased

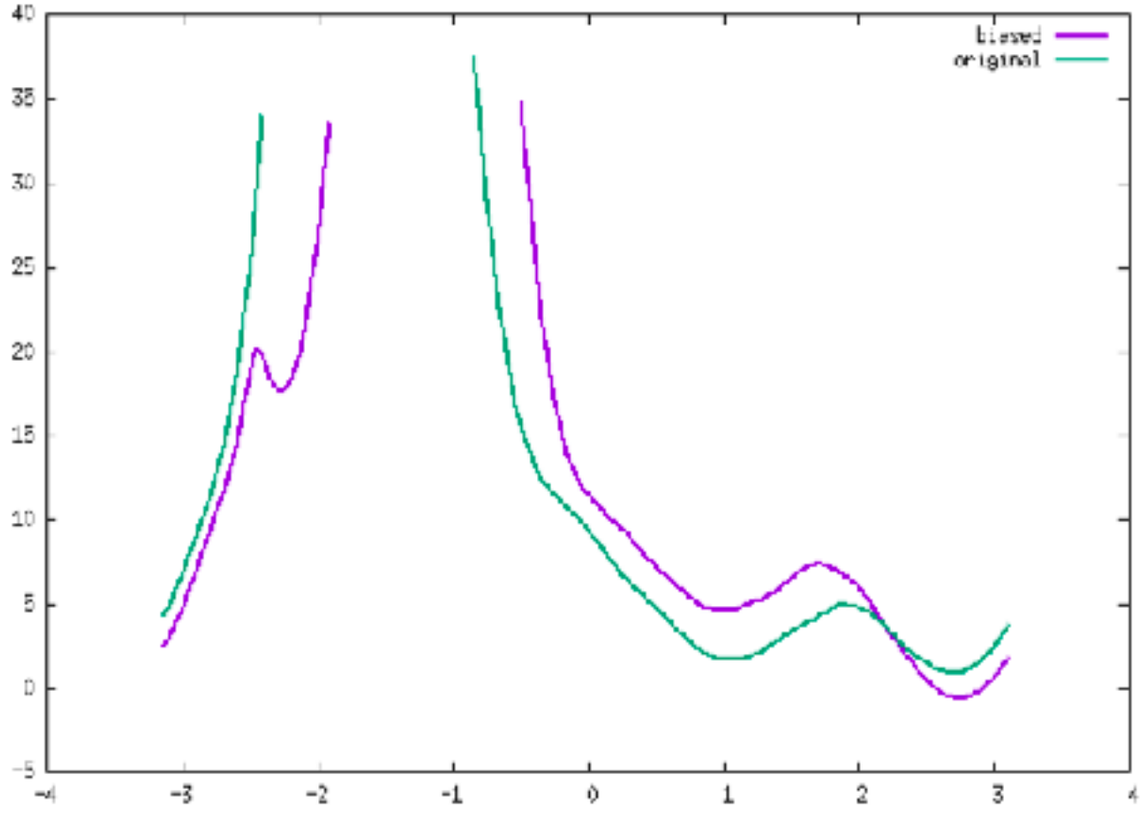


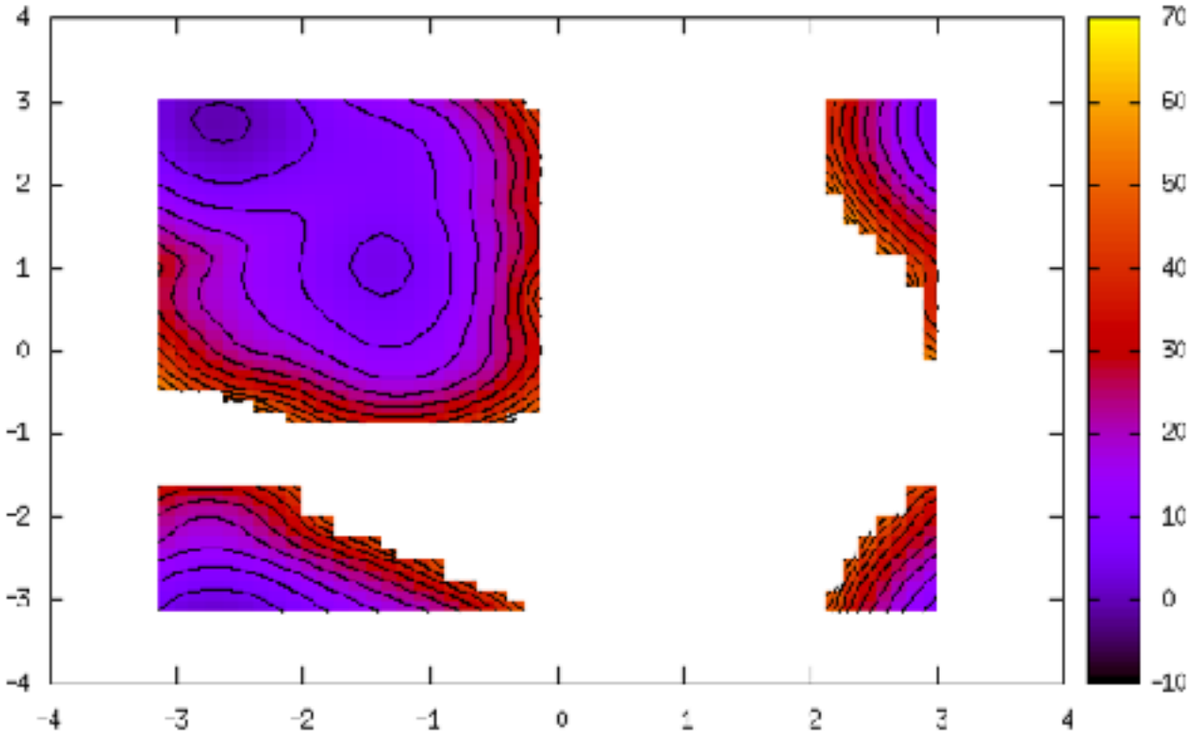
Original

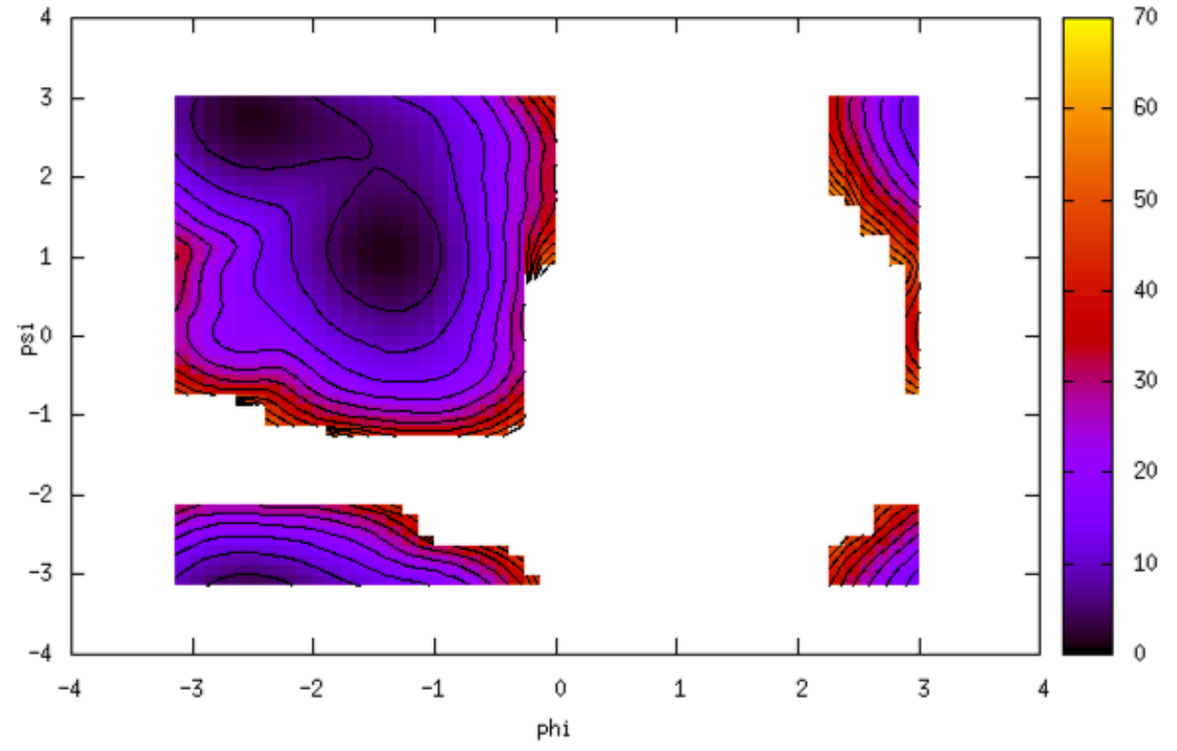


Adding a bias:





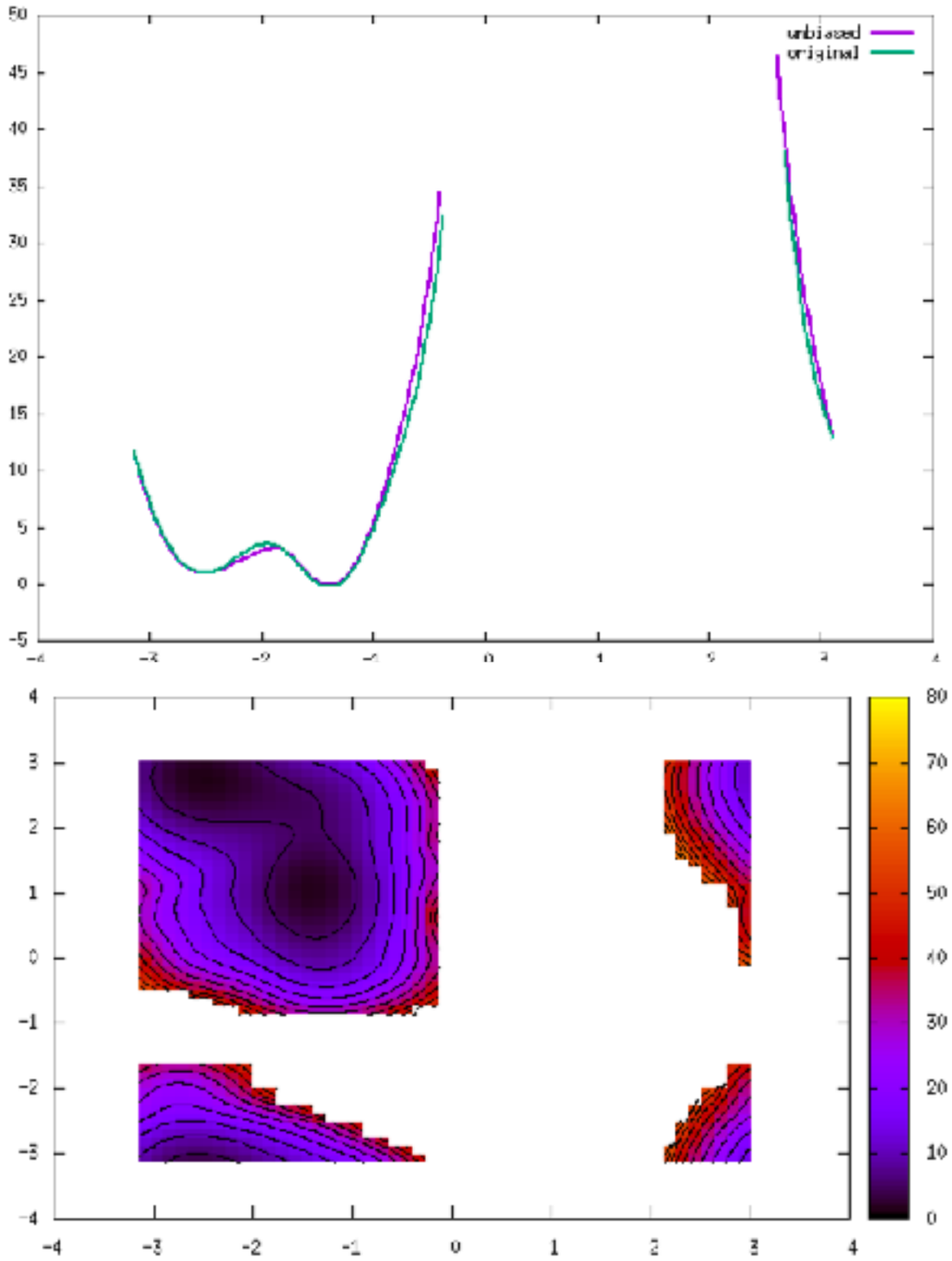


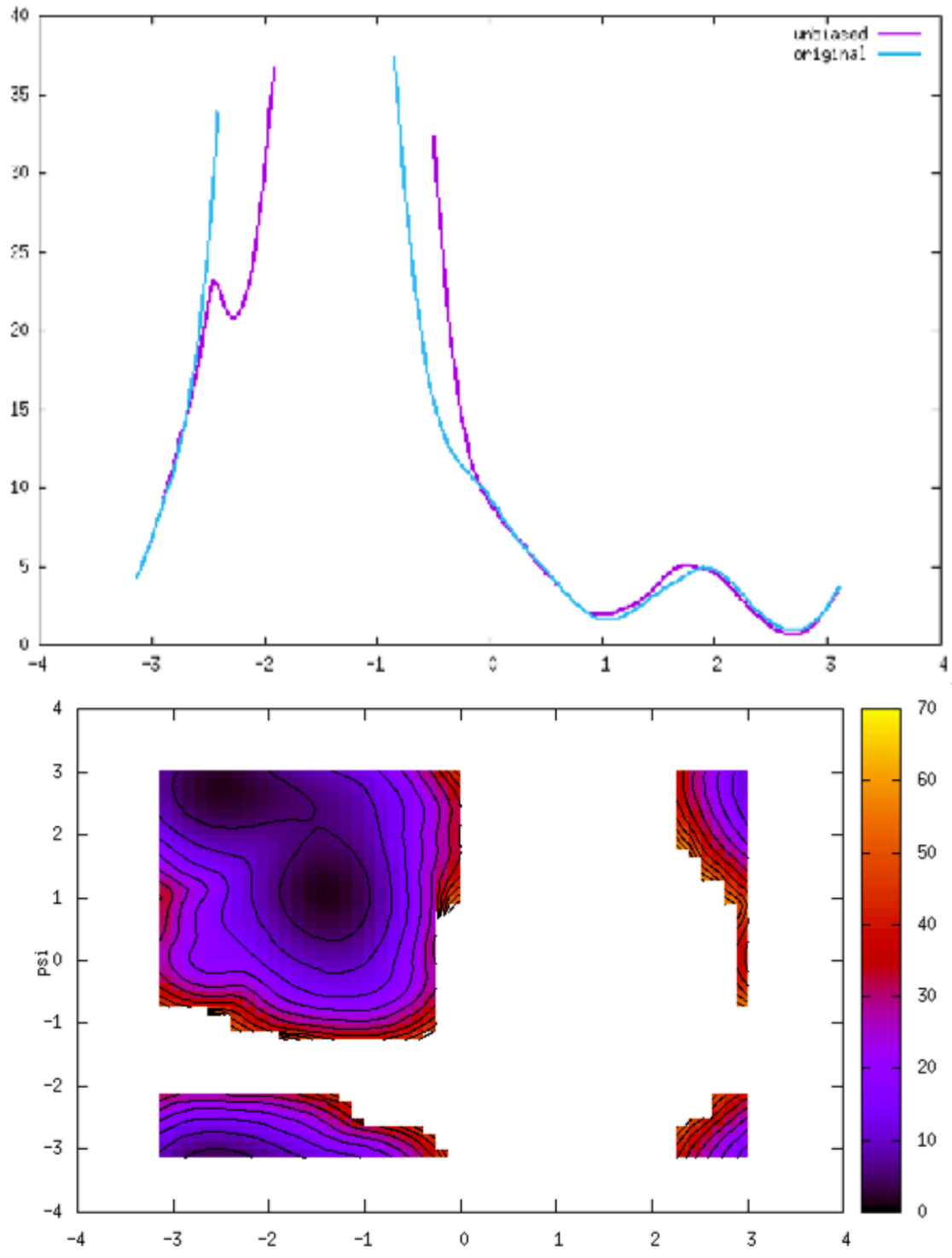


Removing a bias:

```
hh: HISTOGRAM ARG=phi,psi STRIDE=10 GRID_MIN=-pi,-pi GRID_MAX=pi,pi GRID_BIN=50,50 BANDWIDTH=0.2,0.2
ffphi: CONVERT_TO_FES GRID=hhphi TEMP=298
ffpsi: CONVERT_TO_FES GRID=hhpsi TEMP=298
ff: CONVERT_TO_FES GRID=hh TEMP=298
DUMPGRID GRID=ffphi FILE=ffphi.dat
DUMPGRID GRID=ffpsi FILE=ffpsi.dat
DUMPGRID GRID=ff FILE=ff2d.dat

MATHEVAL ...
ARG=phi
LABEL=doubleg
FUNC=-10*sin(x)
PERIODIC=NO
... MATHEVAL

b: BIASVALUE ARG=doubleg

as: REWEIGHT_BIAS TEMP=298

hhphiu: HISTOGRAM ARG=phi STRIDE=10 GRID_MIN=-pi GRID_MAX=pi GRID_BIN=200 BANDWIDTH=0.1 LOGWEIGHTS=as
hhpsiu: HISTOGRAM ARG=psi STRIDE=10 GRID_MIN=-pi GRID_MAX=pi GRID_BIN=200 BANDWIDTH=0.1 LOGWEIGHTS=as
hhu: HISTOGRAM ARG=phi,psi STRIDE=10 GRID_MIN=-pi,-pi GRID_MAX=pi,pi GRID_BIN=50,50 BANDWIDTH=0.2,0.2 LOGWEIGHTS=as
ffphiu: CONVERT_TO_FES GRID=hhphiu TEMP=298
ffpsiu: CONVERT_TO_FES GRID=hhpsiu TEMP=298
ffu: CONVERT_TO_FES GRID=hhu TEMP=298
DUMPGRID GRID=ffphiu FILE=ffphiu.dat
DUMPGRID GRID=ffpsiu FILE=ffpsiu.dat
DUMPGRID GRID=ffu FILE=ff2du.dat

PRINT ARG=phi,psi,b.bias FILE=colvar.dat STRIDE=10
```

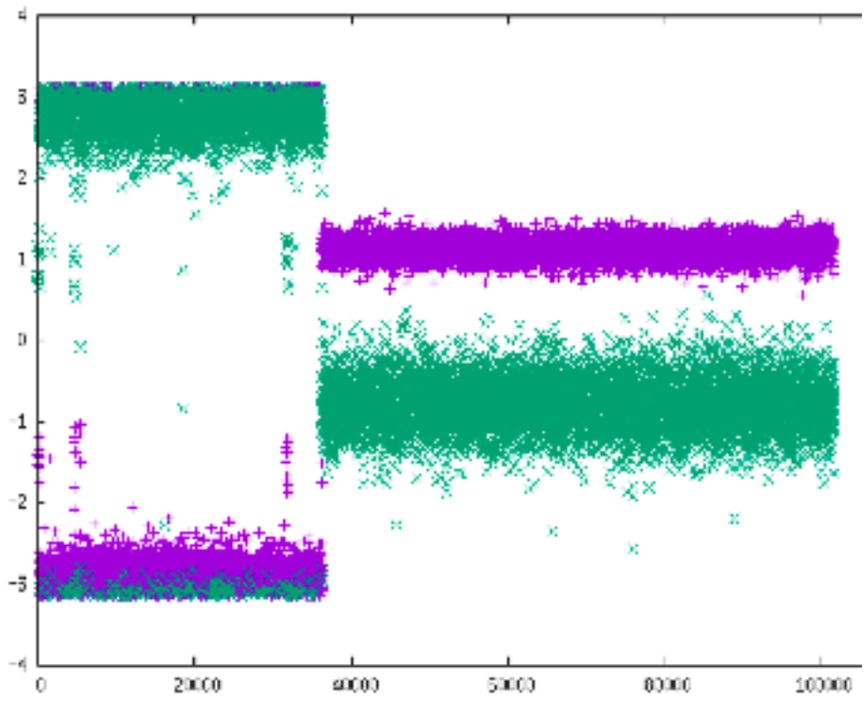
Removing a bias:

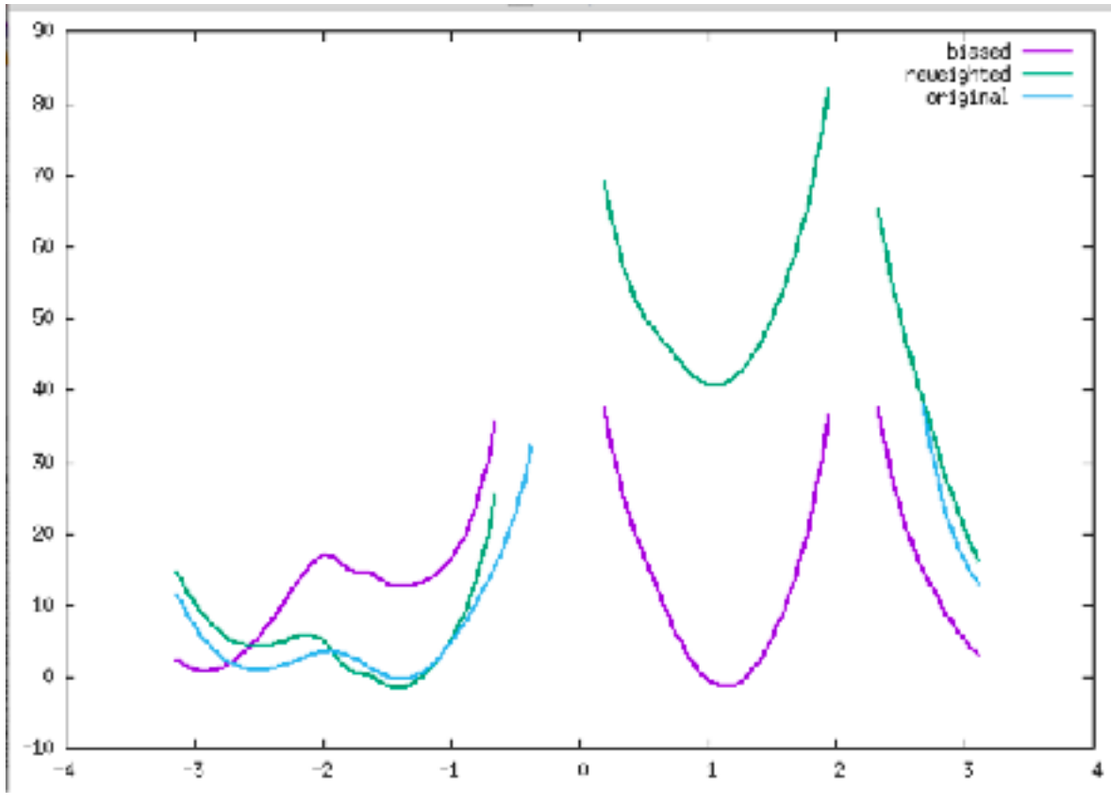




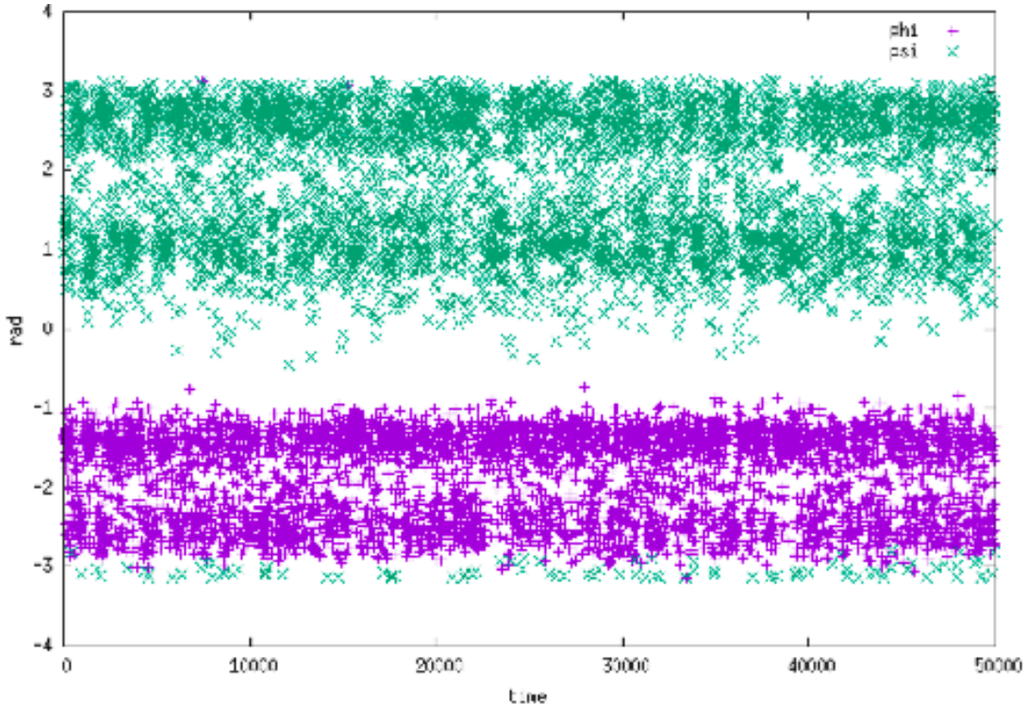
Adding a bias:

if we now use a stronger bias we can leave the local minimum but even doubling the sampling time we are just trapped in a new minimum. The two distributions don't overlap.





Original

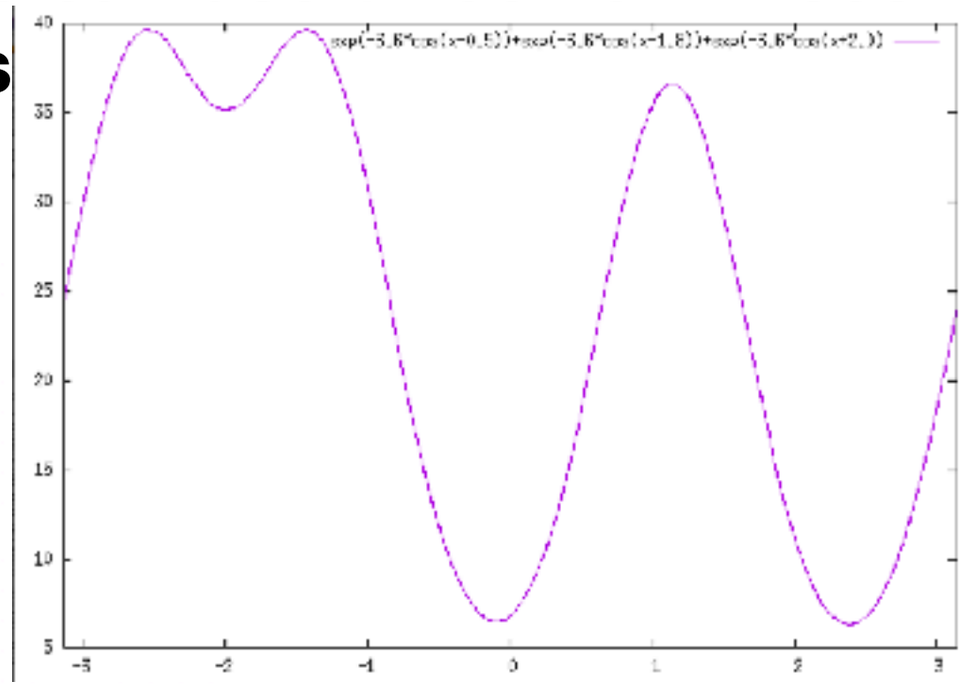


Adding a bias:

So the main issue of using 'umbrella sampling' is how to build a biasing potential that can bridge the real unknown probability distribution and a target distribution easy to sample.

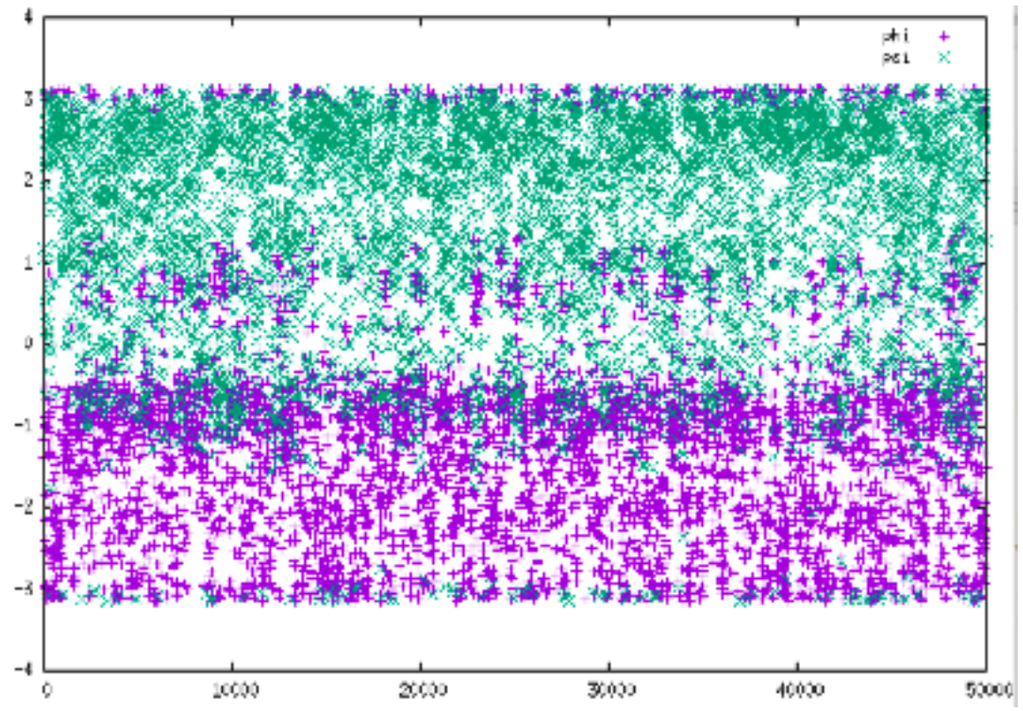
In the case of alanine dipeptide we can now add an additional bias in the new minimum and hope to be able to sample almost uniformly everywhere. Since we have three basins we can use three von Mises functions (circular Gaus

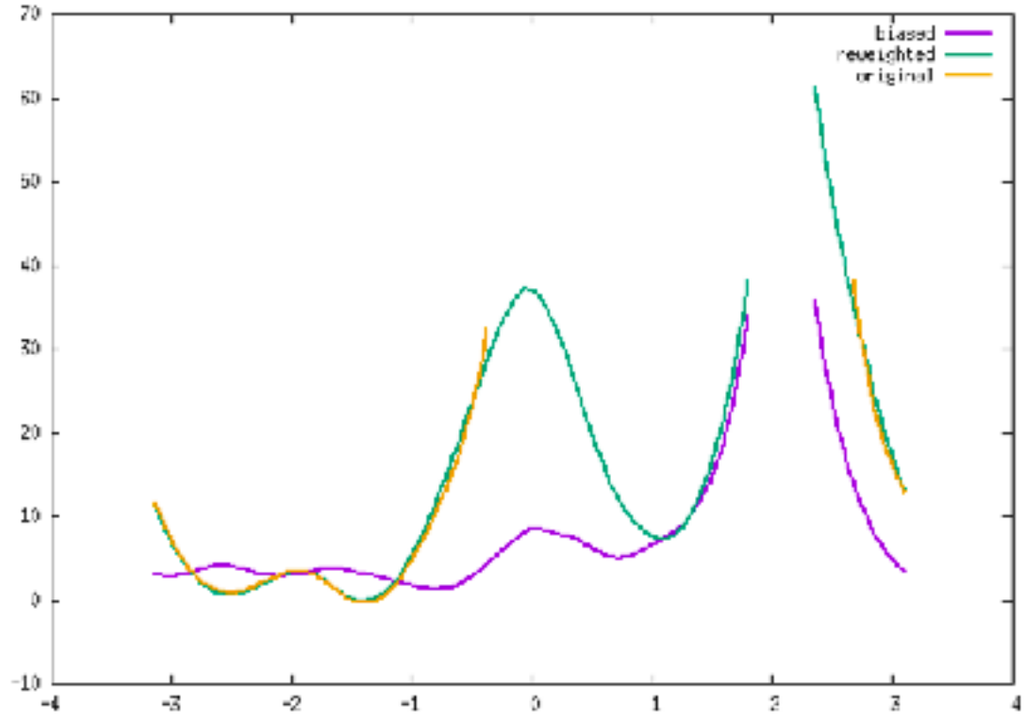
$$e^{k\cos(x-\mu)}$$

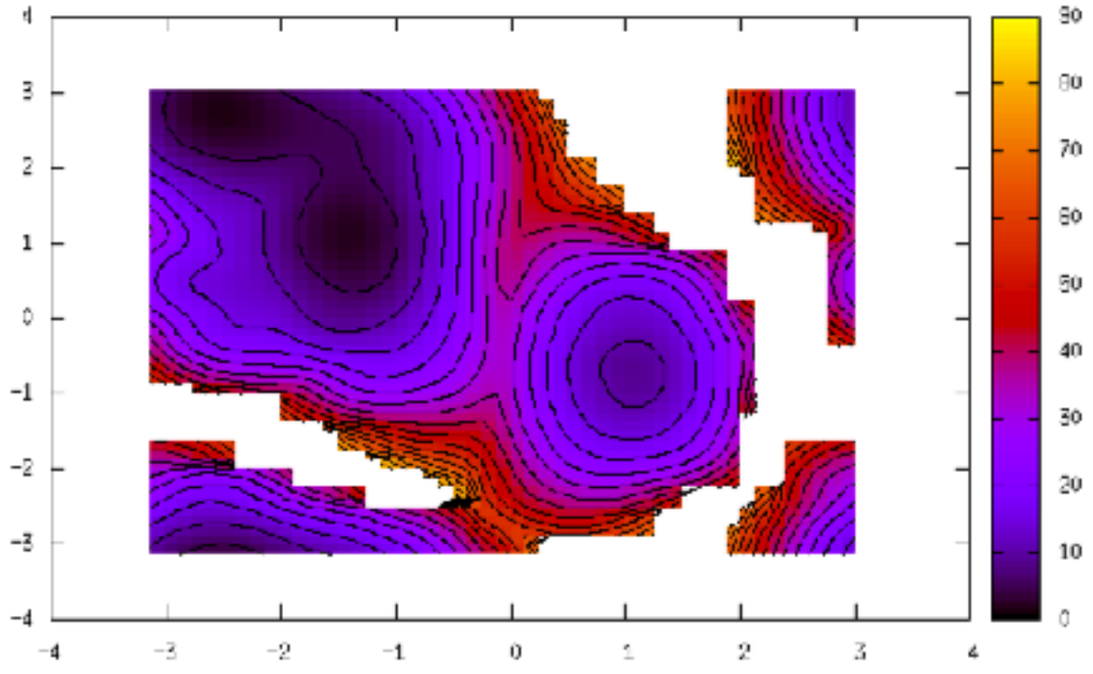


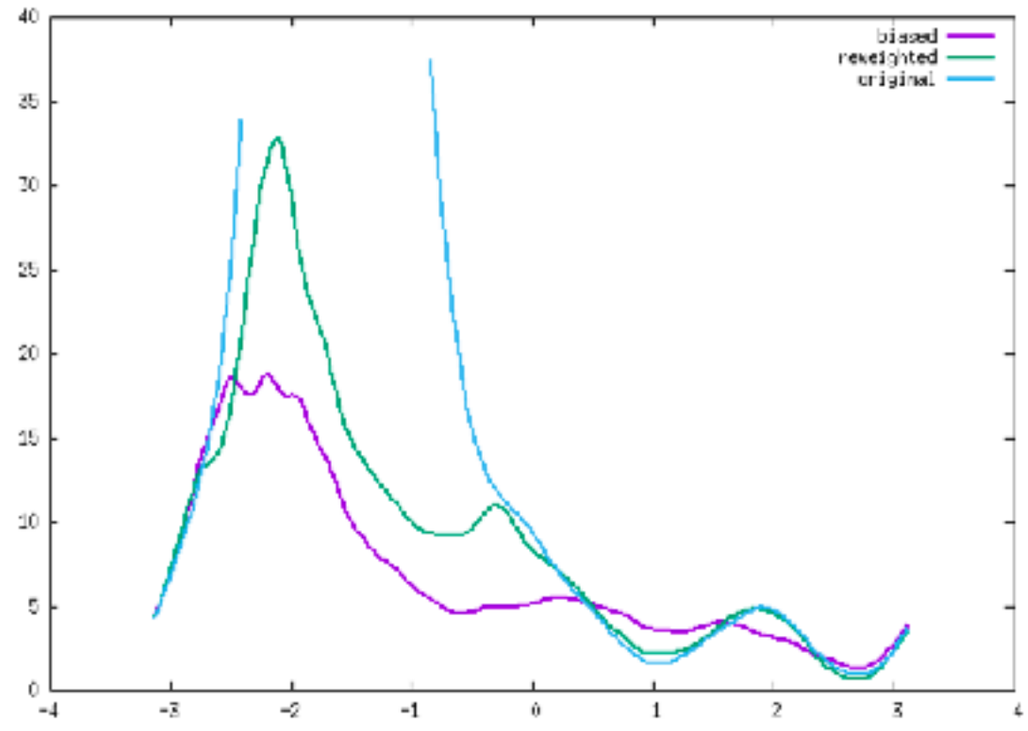

```
# vim:ft=plumed

MOLINFO STRUCTURE=aladip.pdb

phi: TORSION ATOMS=@phi-2

MATHEVAL ...
ARG=phi
LABEL=doubleg
FUNC=exp(-3.6*cos(x-0.5))+exp(-3.6*cos(x-1.8))+exp(-3.6*cos(x+2.))
PERIODIC=NO
... MATHEVAL

b: BIASVALUE ARG=doubleg

PRINT ARG=phi FILE=phi.dat STRIDE=10
```

Adding a bias:





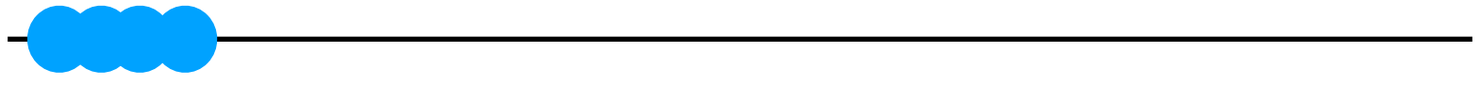




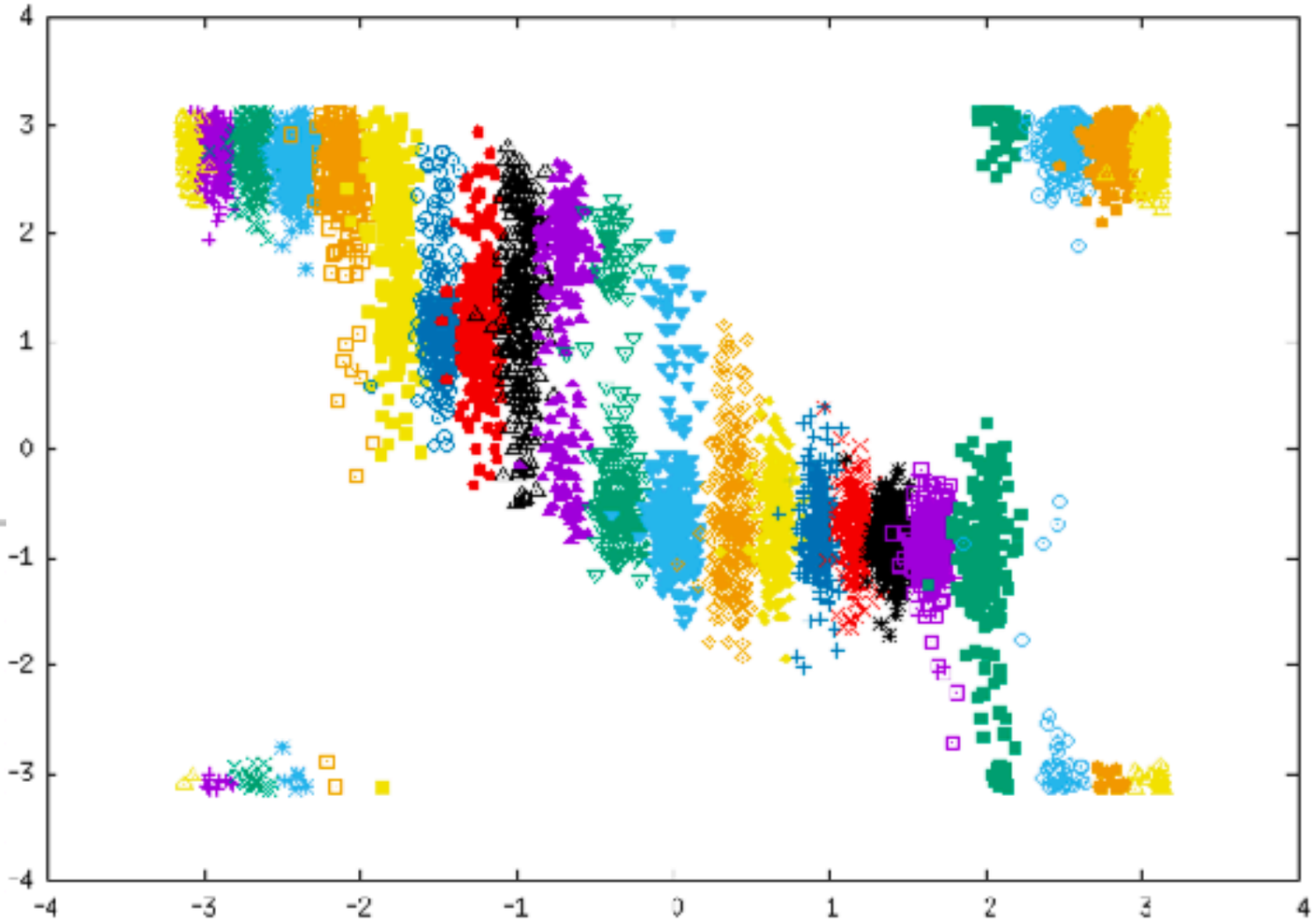
Umbrella Sampling: windows implementation

The WHAM solution to the problem of umbrella sampling consists to force the sampling to be uniform along one or more collective variables, with the hypothesis that the sampling in all other directions is converged.

By adding a sufficiently strong quadratic potential it is possible to restrain the sampling in the surrounding of a point in the CV space. Neighbour window should be close enough that there is overlap in the sampling.
(if the value of CV is constrained like with shake this is the blue moon ensemble)

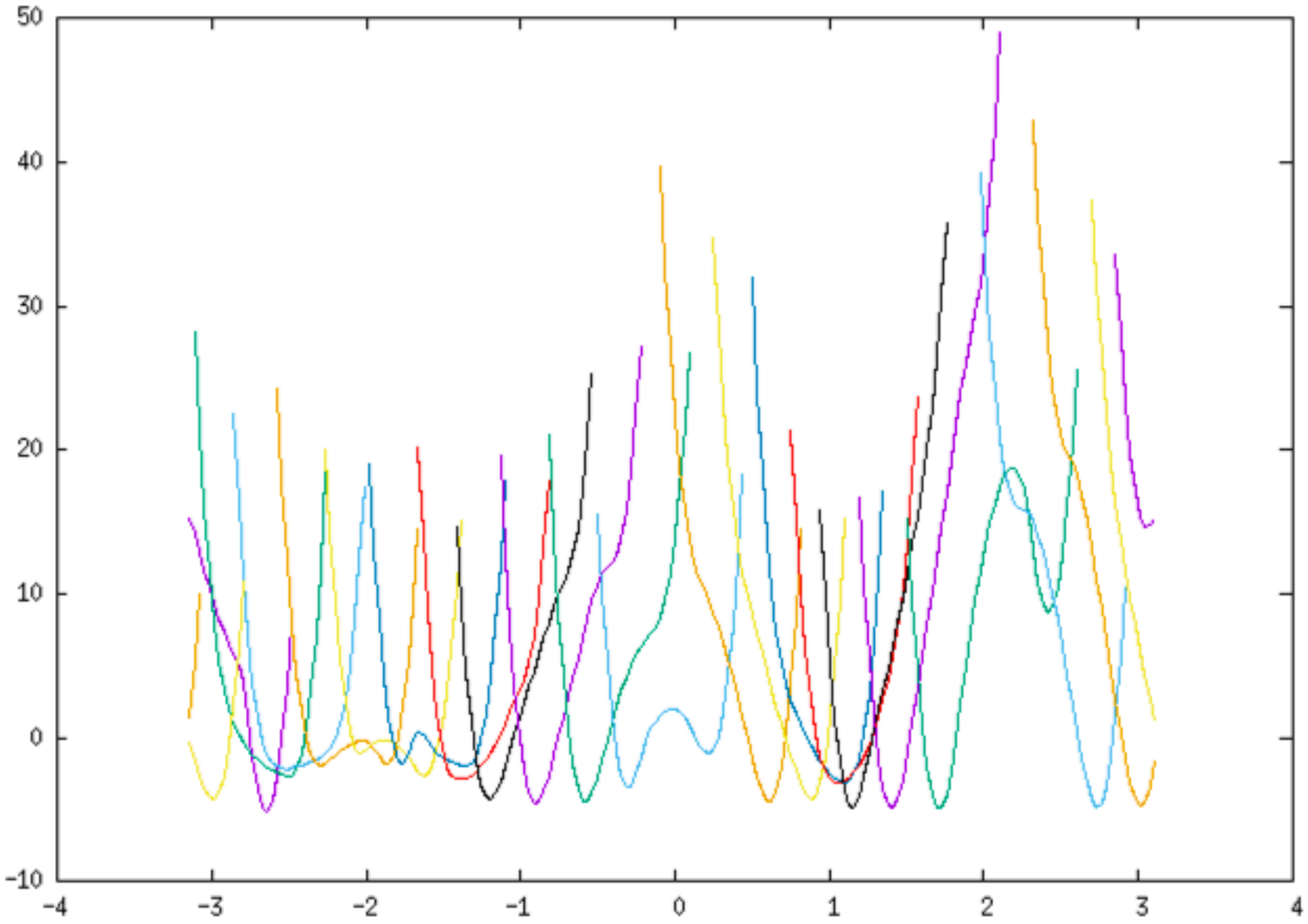


Umbrella Sampling: windows implementation




```
AT=-3
for i in `jot - 0 20`; do

cat >plumed.$i.dat << EOF

MOLINFO STRUCTURE=aladip.pdb

phi: TORSION ATOMS=@phi-2
psi: TORSION ATOMS=@psi-2

restraint-phi: RESTRAINT ARG=phi KAPPA=500.0 AT=$AT

PRINT ARG=phi,psi FILE=colvar.dat STRIDE=10

EOF

gmx_mpi mdrun -plumed plumed.$i.dat -nsteps 250000 -x traj$i.xtc -nb cpu -v -s short.tpr -cpi state -noappend

AT=`echo 'scale=1; '$AT'+0.3' | bc`;
done
```

Merging windows



WHAM allow recovering the overall probability distribution.

The problem is that each simulation samples a different probability distribution associated with a free-energy with an unknown additive constant

The probability of observing all the configurations sampled in our N biased trajectories is:

$$P(\mathbf{T}) \propto \prod_{j=1}^{M} \prod_{k=1}^{N} (c_k w_{kj} p_j)^{t_{kj}}$$

M is the number of bin used to make the histogram of the CV. t_{kj} is the number of frames from the trajectory k in the bin j. w_{kj} account for the effect of the bias of simulation k on the configurations belonging to bin j.

$$e^{+\frac{V(\mathbf{q})}{k_B T}}$$

Ck is a free parameter to enforce normalisation and pj is the unknown unbiased probability.

WHAM allow recovering the overall probability distribution.

$$\mathscr{L} = \sum_{j=1}^{M} \sum_{k=1}^{N} t_{kj} \ln c_k w_{kj} p_j + \sum_{k}^{N} \lambda_k \left(\sum_{j=1}^{M} c_k w_{kj} p_j - 1 \right)$$

Here we took the ln of the P(T) and we add a constraint to impose normalisation for each trajectory. One can find optimum for this equation by searching for the derivatives with respect to pj and lambdak and imposing them equal to zero.

$$p_j \propto \frac{\sum_{k=1}^{N} t_{kj}}{\sum_k c_k w_{kj}}$$

$$c_k = \frac{1}{\sum_{j=1}^{M} w_{kj} p_j}$$

Iteratively, one guess a set of ck, get pj and so on. Replica specific information is lost. So

$$\tilde{w}_n = \frac{1}{\sum_k c_k w_{kn}}$$

Is the weight of frame n of the concatenated trajectory

Umbrella sampling

- **The sampling of each window should be tested for**
  - **de-correlation**
  - **convergence**
  - **Structural overlap with neighbour windows**
  - **If successive windows starting configurations have been generated going forward one should ideally repeat the procedure in the opposite direction**

- **Umbrella sampling can be coupled to other methods to enhance the sampling:**
  - **Replica-exchange (the simplest is exchange between neighbour windows, this is essentially free)**
  - **Parallel Tempering (Run each window at many temperature)**
  - **Metadynamics**

Principles for biased simulations

MD/MC are importance sampling algorithm that can be tweaked by using additional probability distributions.

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

Change the temperature
Scale the force-field
Add a potential, also time-dependent as long as it converges to something, on a function f(x)

MaxLL methods like WHAM (DRAM, TRAM, xxAM, etc) can be employed to recover the unbiased original distribution