The Accelerated Weight Histogram method

References

- The GROMACS reference manual

- V. Lindahl, J. Lidmar and B. Hess, *Accelerated weight histogram method for exploring free energy landscapes,* J. Chem. Phys. 141, 044110 (2014)

- V. Lindahl, J. Lidmar and B. Hess, *Sequence dependency of canonical base pair opening in the DNA double helix*, PLOS Computational Biology, 13 (4), e1005463 (2017)

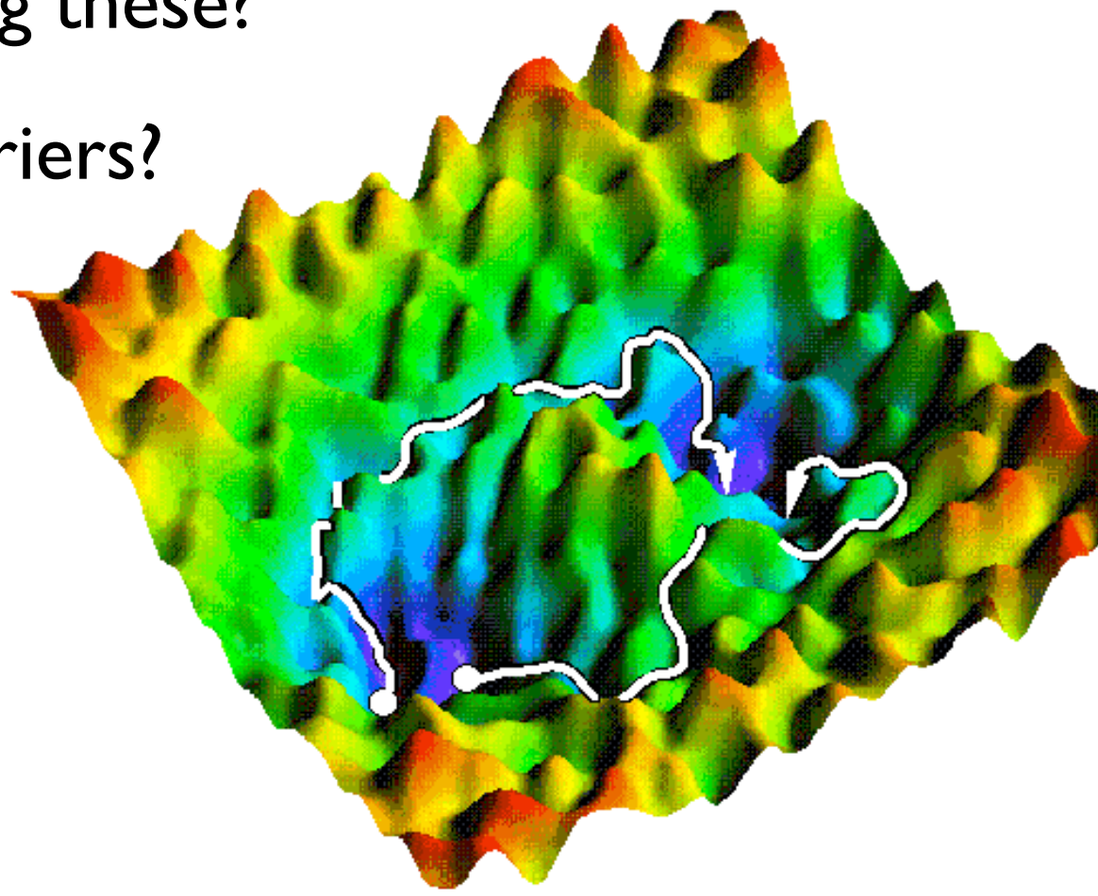
- V. Lindahl, P. Gourdon, M. Andersson and B. Hess, *Permeability and ammonia selectivity in aquaporin TIP2;1: linking structure to function*, Scientific Reports, 8, 2995 (2018)

- V. Lindahl, J. Lidmar and B. Hess, *Riemann metric approach to optimal sampling of multidimensional free-energy landscapes*, Phys. Rev. E, 98, 023312 (2018)

most of the work done by Viveca Lindahl

Sampling free-energy landscapes

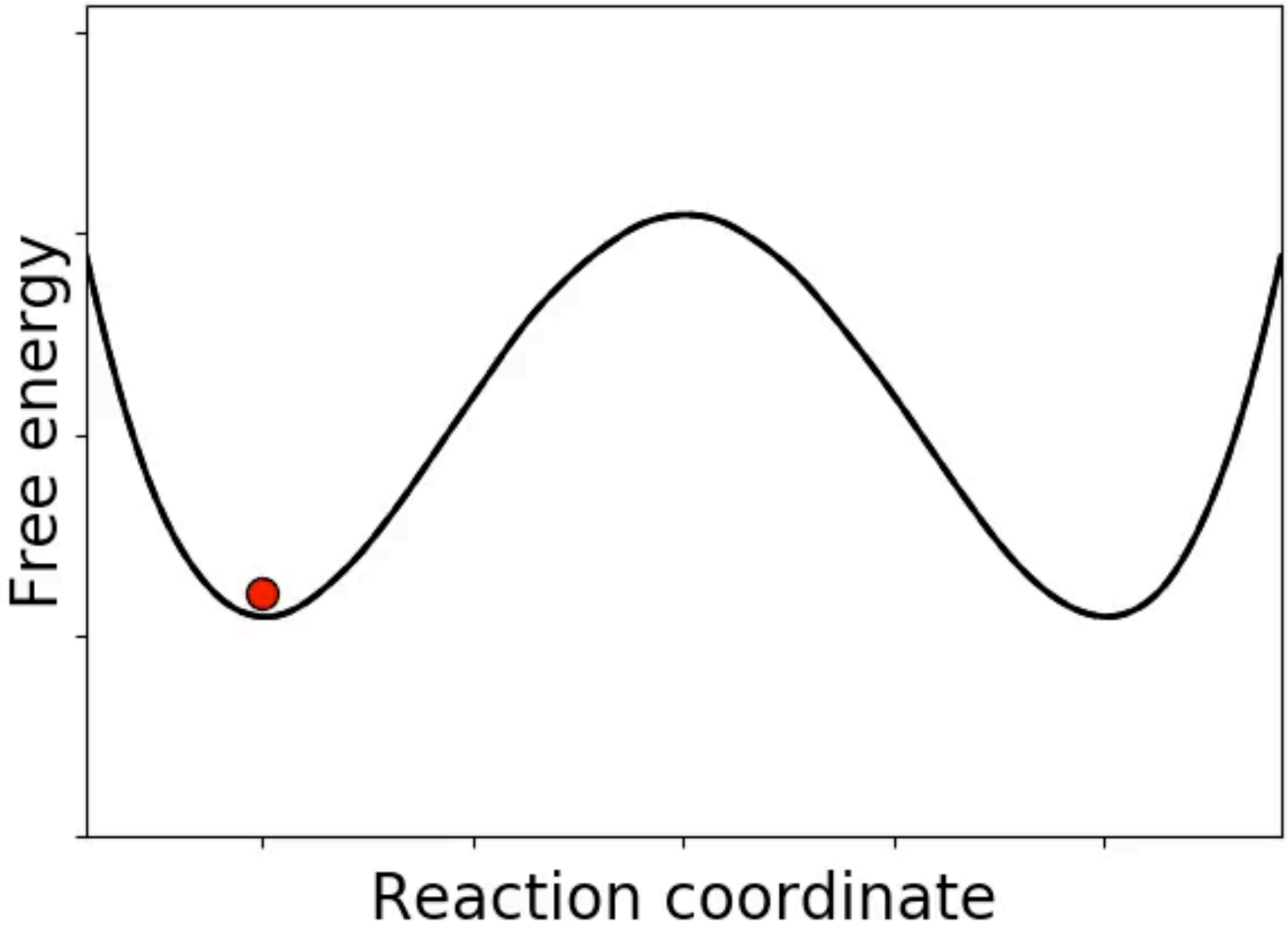
- Energy landscapes of biomolecules are high dimensional and often rough
- Typical situations:
  - You know the beginning and the end state
    - What are the paths connecting these?
    - What are the free-energy barriers?
  - You know the beginning state
    - …
- Can you come up with a reaction coordinate?



Typical issue in molecular dynamics

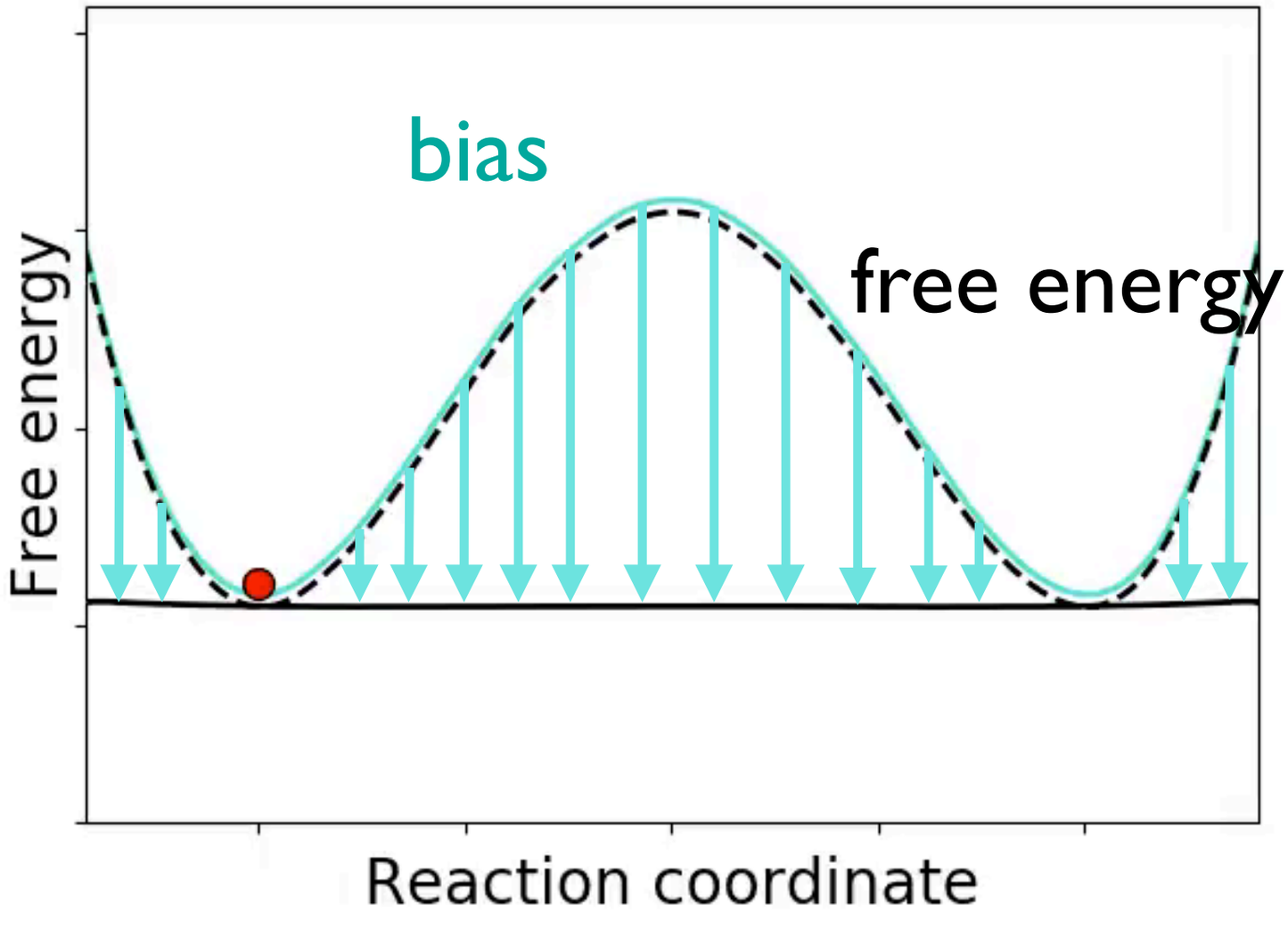
- Interesting events happen on a microsecond timescales
  - thus we needs billions of time steps of 2 fs
- But the event itself is often fast
  - We can (smartly) use many independent simulations
  - We can smartly bias simulations to get more events
- Thus
  - More efficient use of compute time
  - Shorter time to solution

The sampling problem



The solution

Trick: add a bias potential to make the effective potential flat
Issue: the potential (or free-energy) is what we are after!



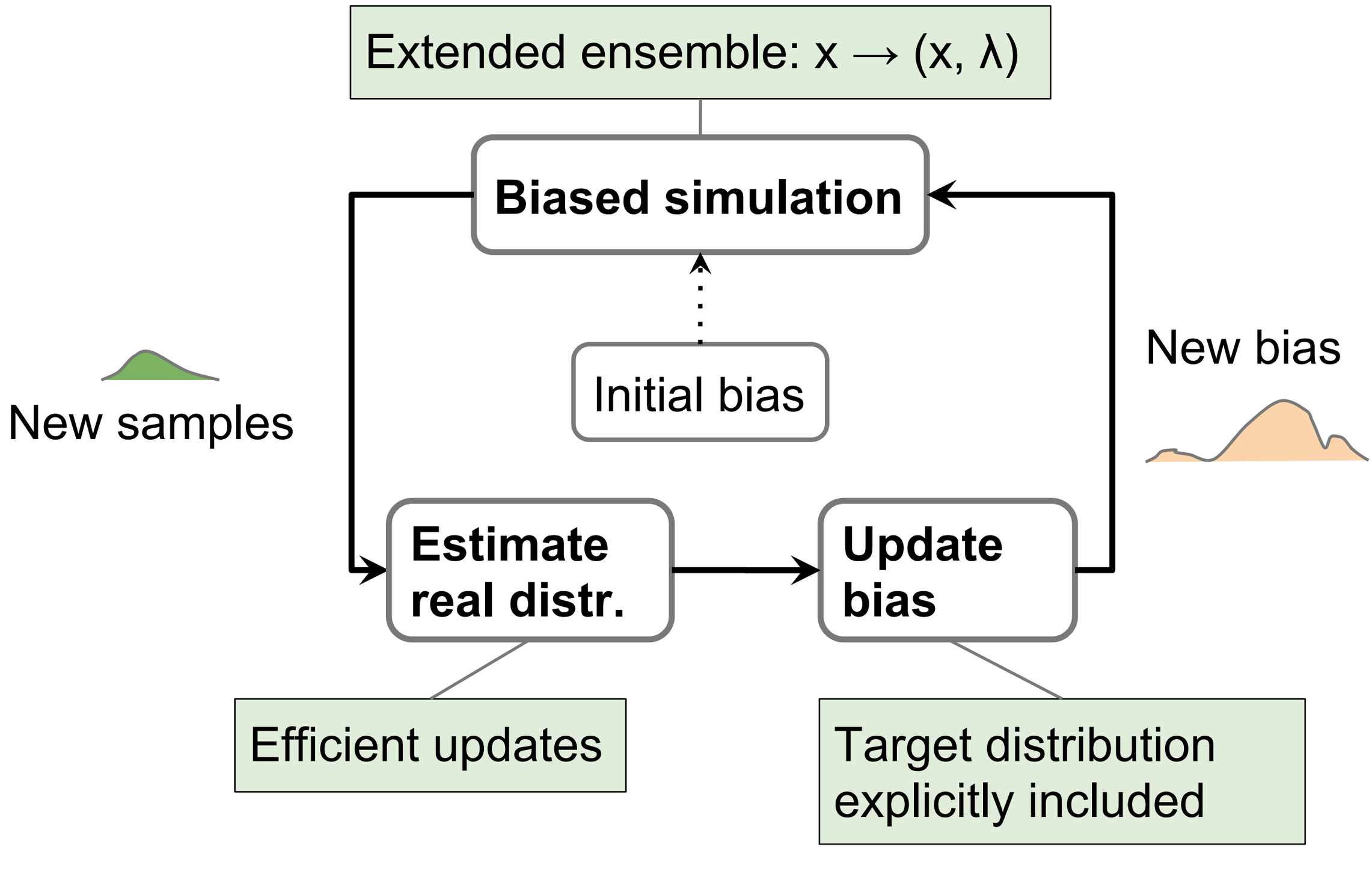
Solving the bias potential problem

- The bias potential is both input and output of the method:
  - an iterative solver is required
- Available solvers:
  - Metadynamics
  - Adaptive biasing force
  - Accelerated weight histogram method
  - …

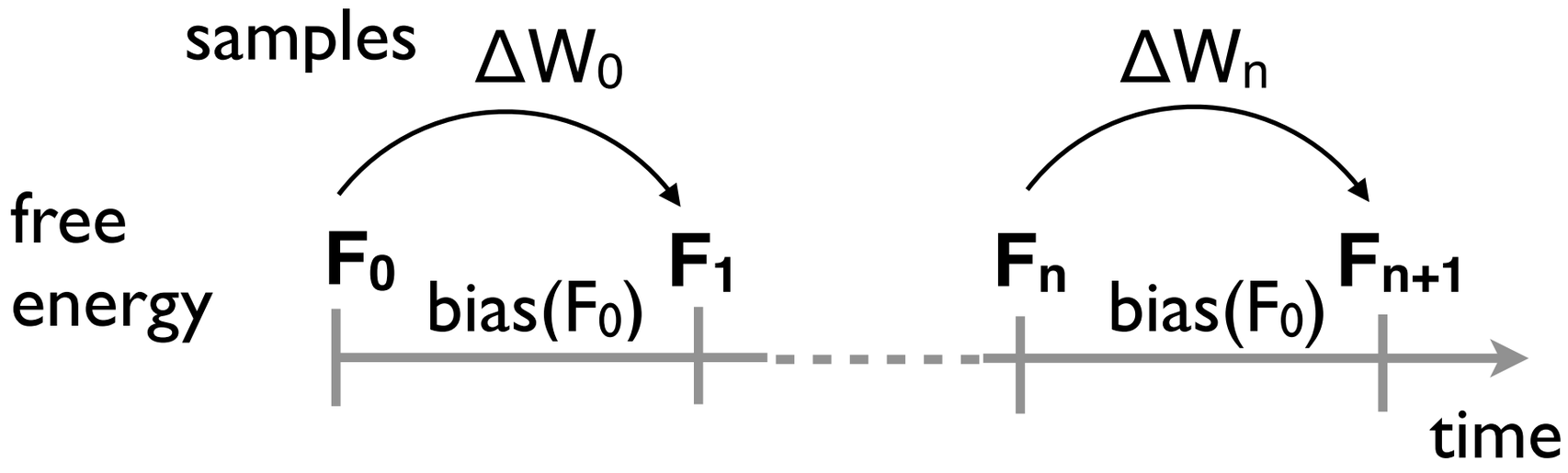
Accelerated weight histogram method

- The target distribution is the central quantity (unlike most other methods)
- The target distribution can be chosen freely and can, but does not have to, depend on the free-energy
- Initial exponential convergence, later sqrt(#samples)
  - automatically controlled
- Only one, uncritical, convergence parameter

AWH schematically



The Accelerated Weight Histogram method



Iterative scheme to solve for the unknown bias / free-energy:

- collect samples (using MD)
- update the free-energy estimate

$$\Delta F_n(\lambda) = -\ln\left(1 + \frac{\Delta W_n(\lambda)}{N_n \pi(\lambda)}\right) \sim \frac{1}{N_n}$$

collected samples

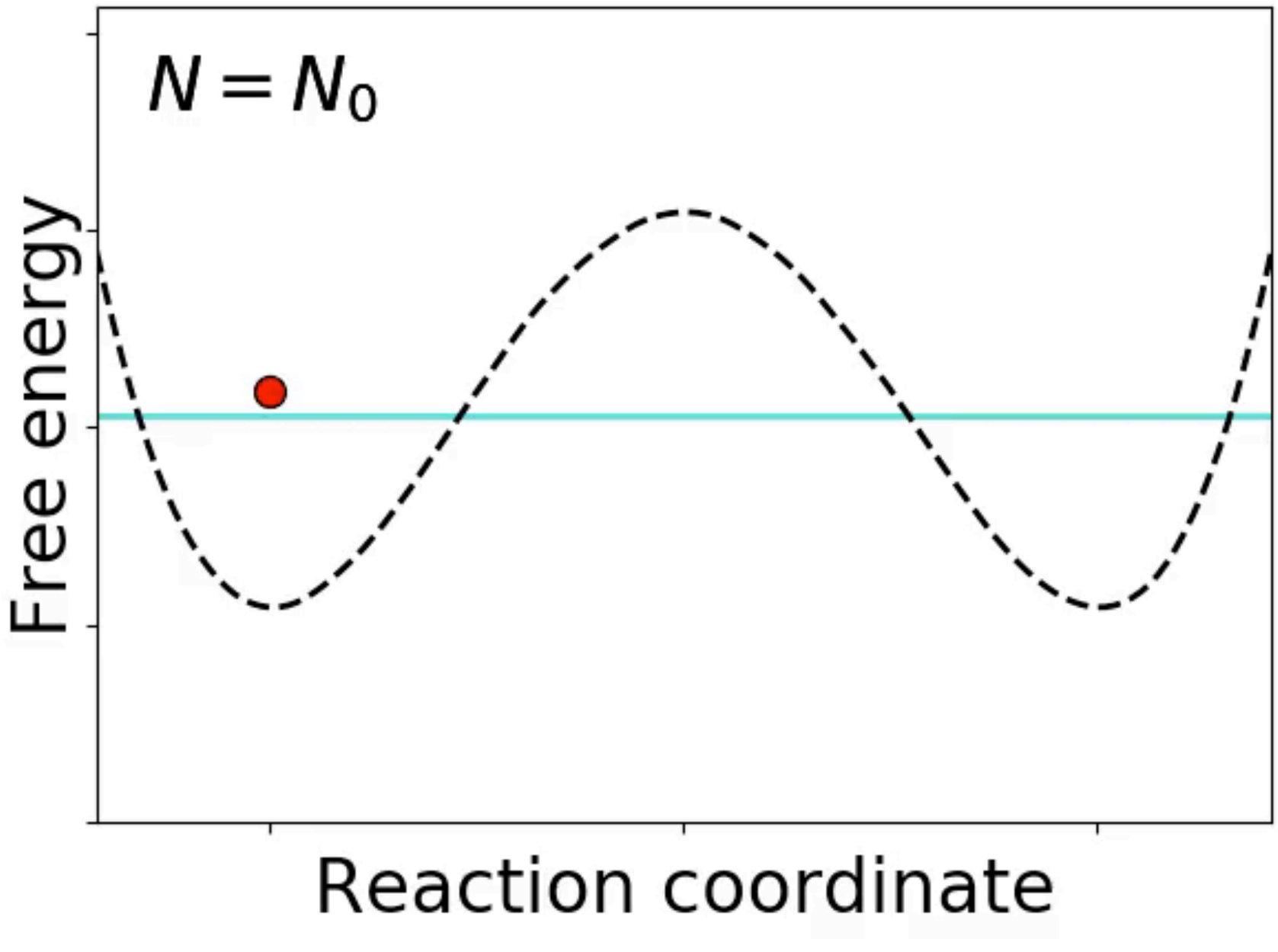
prior number of samples

target distribution (flat)

How does AWH apply the bias?

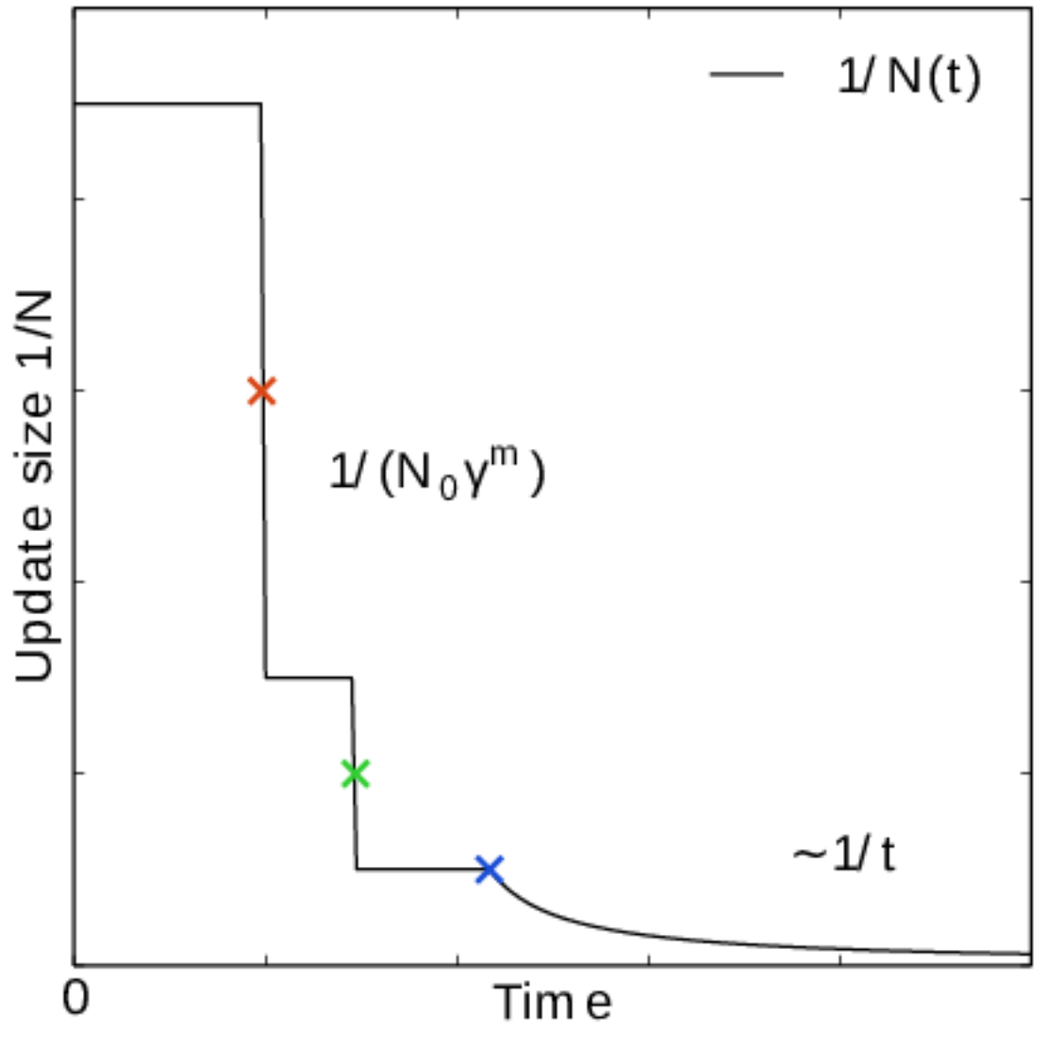
- Initially we used a harmonic umbrella potential which moved using Monte Carlo moves
- Now the default setup is the "Boltzmann inversion" of a convolution of Gaussians, produces by harmonic umbrellas, which creates a smooth bias potential
    - This is done on a regular grid considering only close neighbors
        - Thus no higher cost when using a fine grid / high force constant

AWH in action



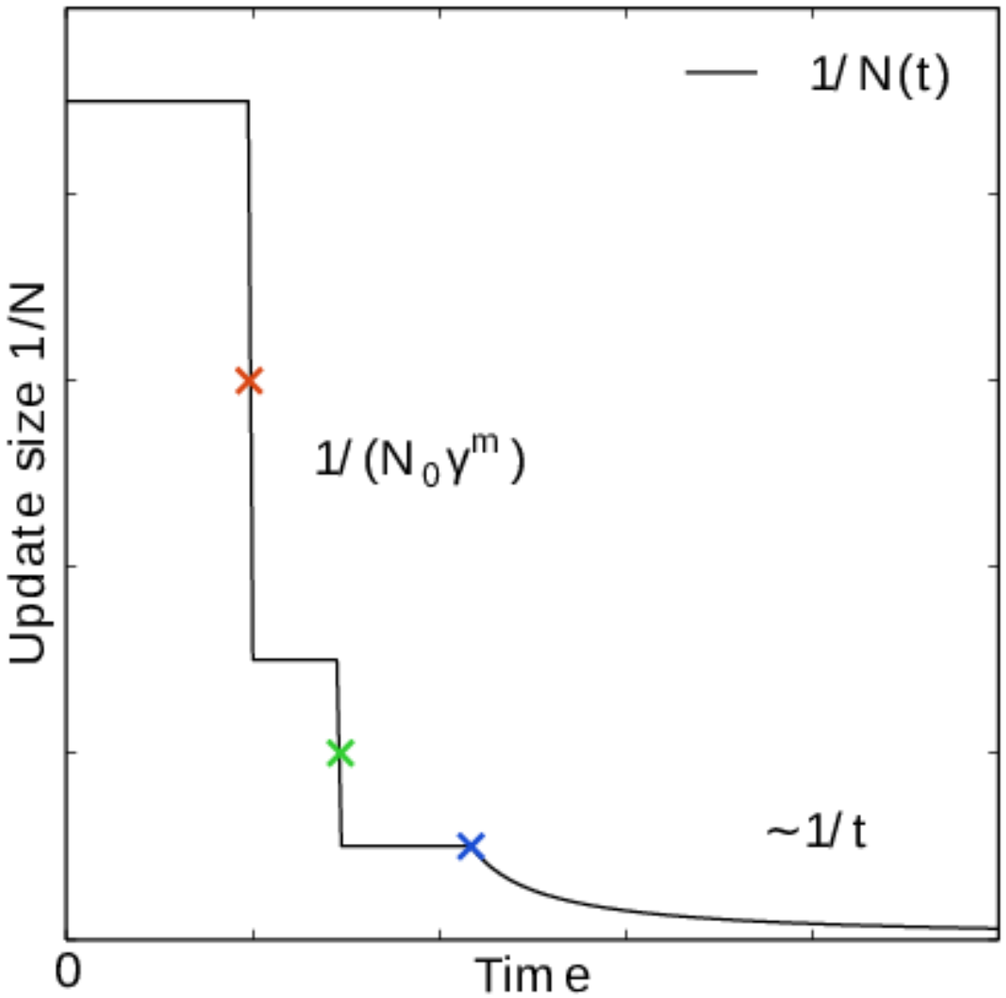
Exponential-linear convergence

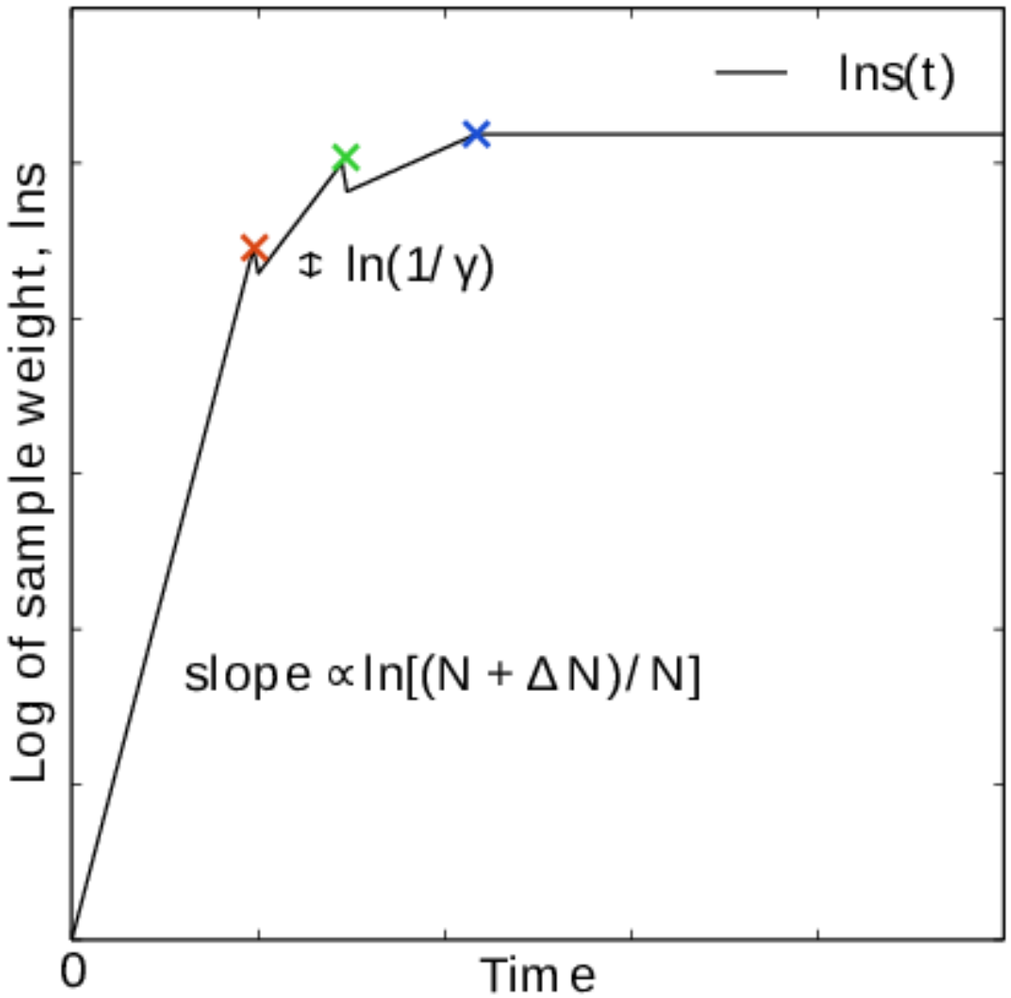
- Initial stage
  - constant update **size**
  - Divide update size by 3 at every covering (cross
- Final stage
  - constant update **weight**
- Switch when initial weight would exceed final weight
  - automated equilibration check



Sample weights

- Initial stage, non-equilibrium samples get weighted down





Example study:

DNA base pair opening

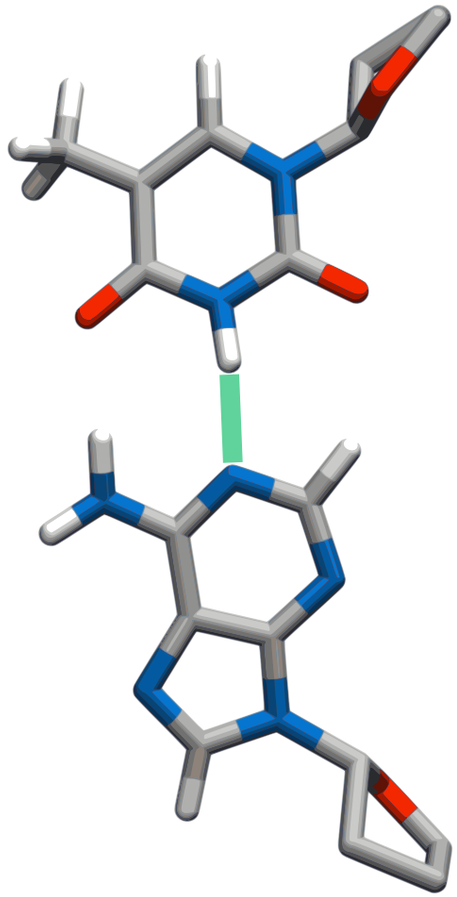
1 ms^{-1}

closing

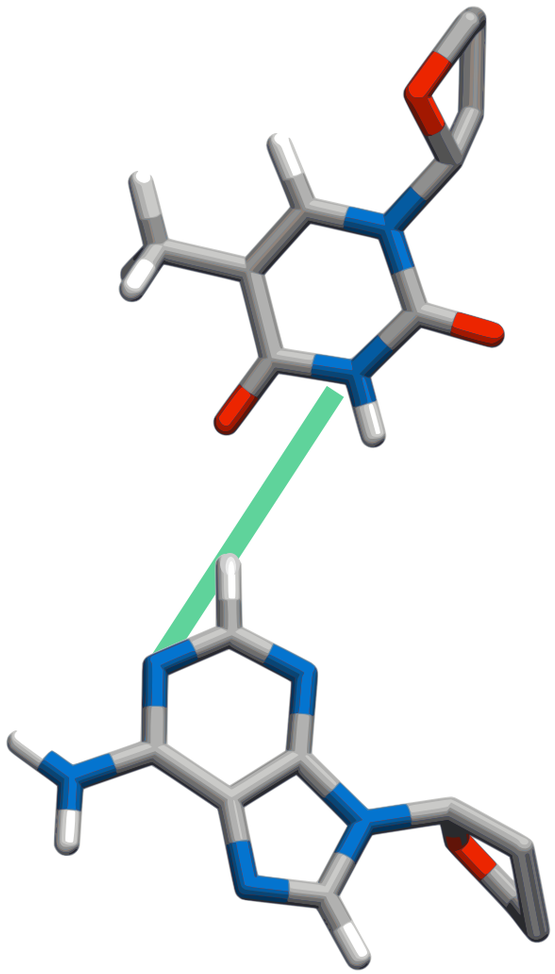
10^4 ms^{-1}

A reaction coordinate

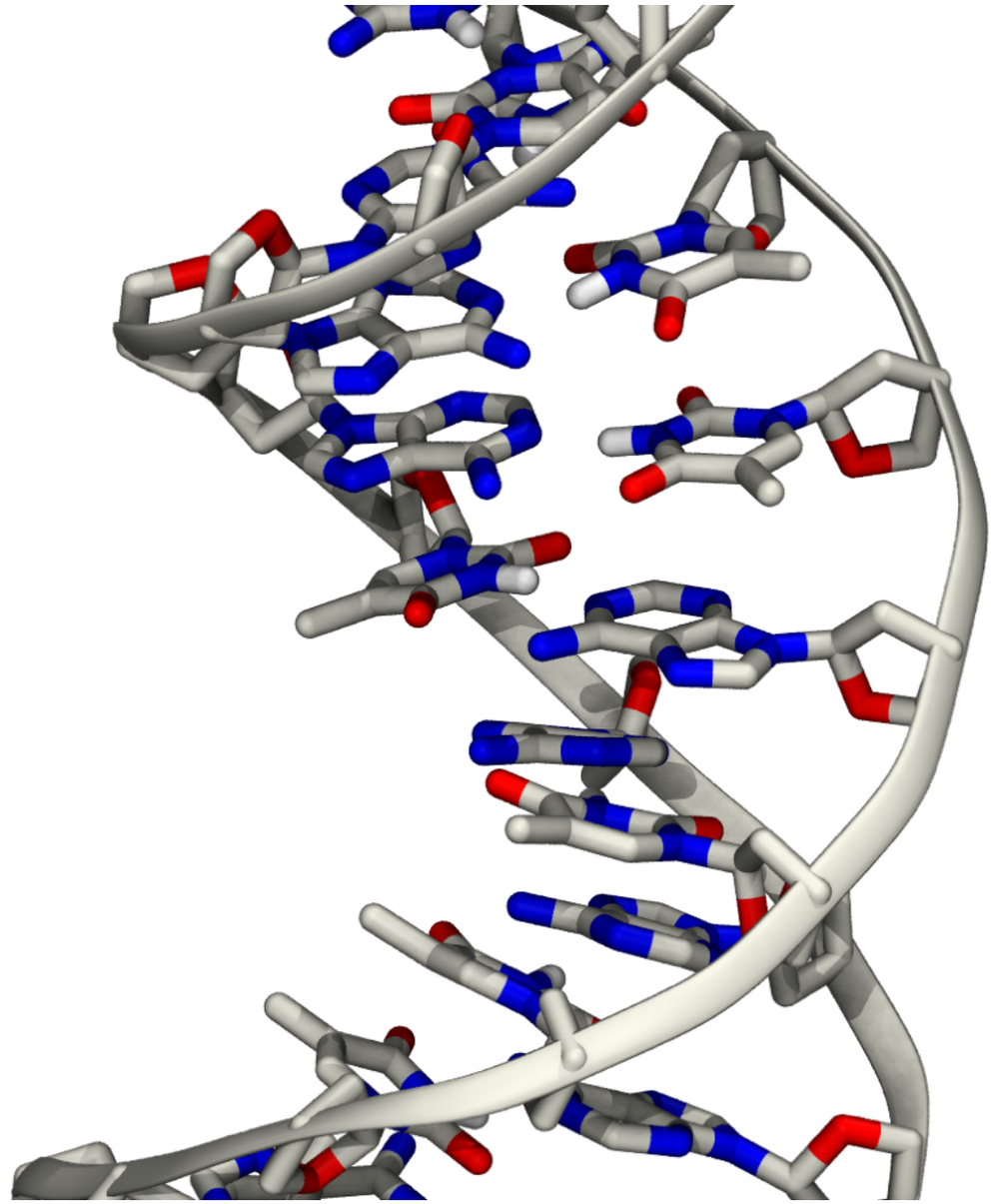
base-pair opening



closed



open



mdp options

```
pull                    = yes
pull-ngroups            = 2
pull-group1-name        = resid8
pull-group2-name        = resid33
pull-ncoords            = 1
pull-coord1-geometry    = distance
pull-coord1-groups      = 1 2
pull-coord1-type        = external-potential
pull-coord1-potential-provider = awh
                    awh                             = yes
                    awh-nbias                       = 1
                    awh1-ndim                       = 1
                    awh1-dim1-coord-index           = 1
                    awh1-dim1-start                 = 0.25
                    awh1-dim1-end                   = 0.60
                    awh1-dim1-force-constant        = 128000
                    awh1-dim1-diffusion             = 5e-5
                    awh1-error-init                 = 5
                    awh-share-multisim              = yes
                    awh1-share-group                = 1
                    awh1-equilibrate-histogram      = yes
```

Initial update size

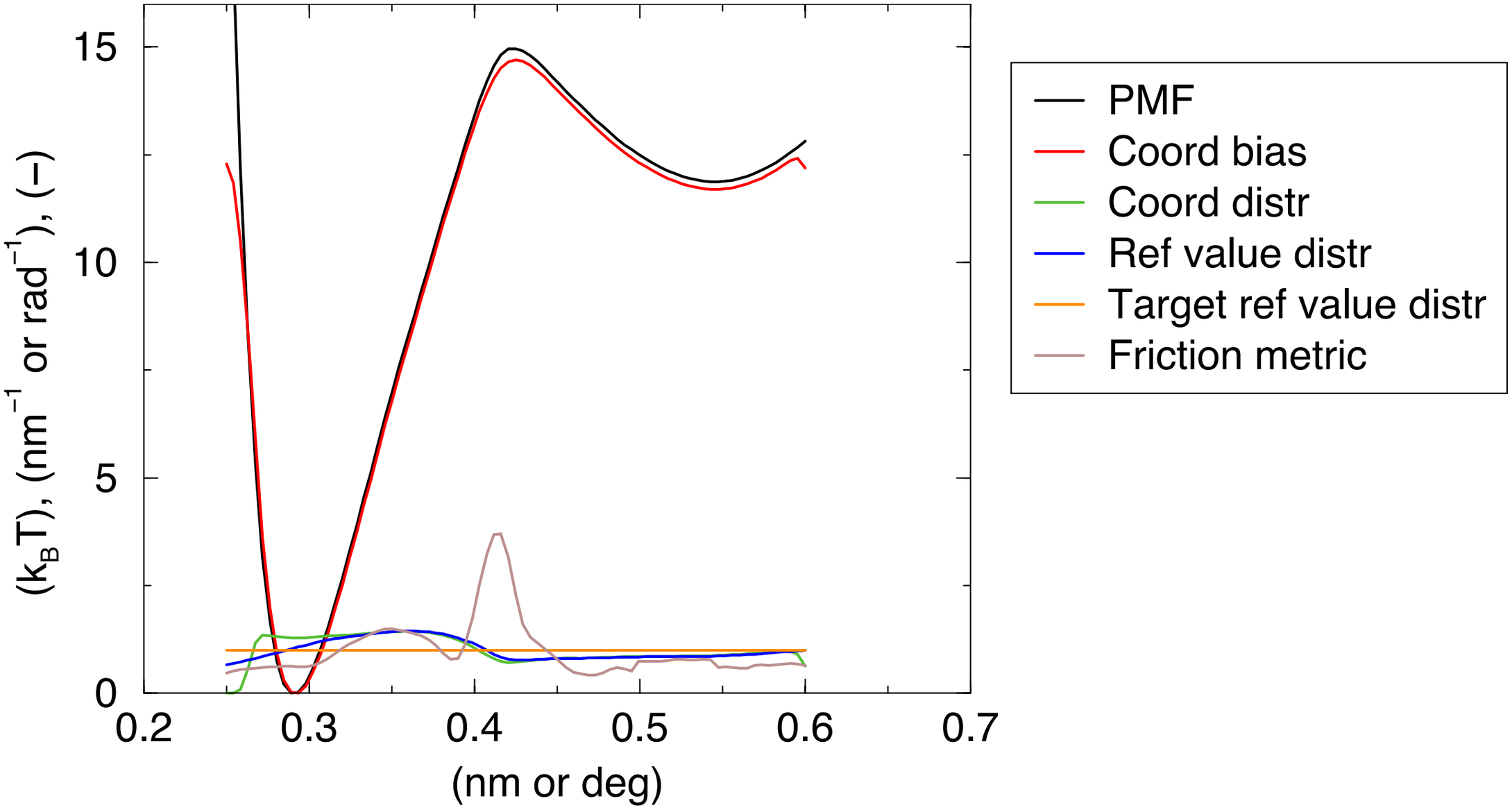
```
awh1-dim1-diffusion = 5e-5 (nm²/ps)
awh1-error-init     = 5    (kJ/mol)
```

- AWH has only one tuning parameter: the initial update size
  - This is set using two meaningful parameters:
    - The initial error in the free-energy ε_0
    - The diffusion along the reaction coordinate *D*

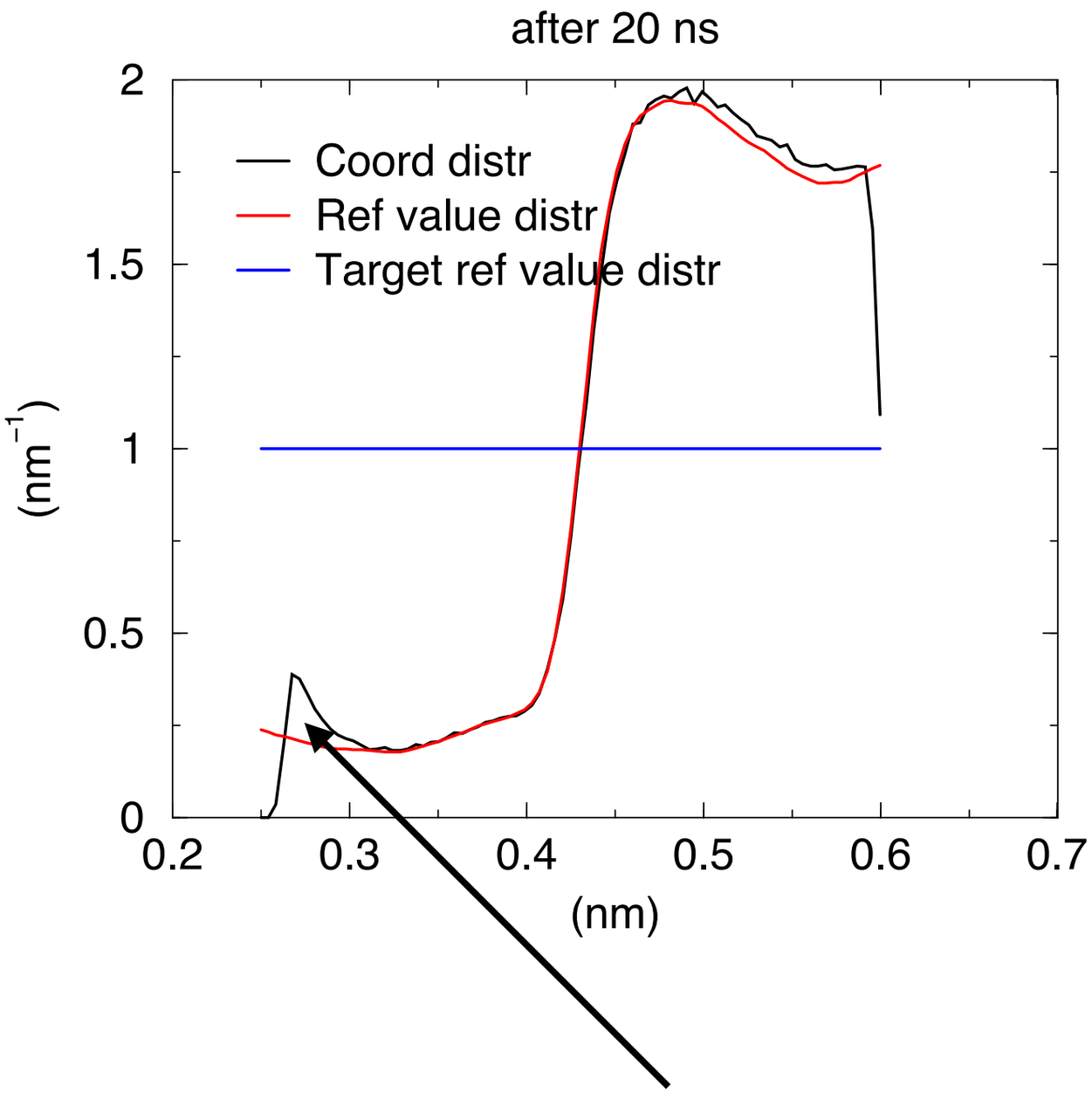
$$\frac{1}{N_0} = \frac{1}{N_0(\epsilon_0, D)} \sim D\epsilon_0^2$$

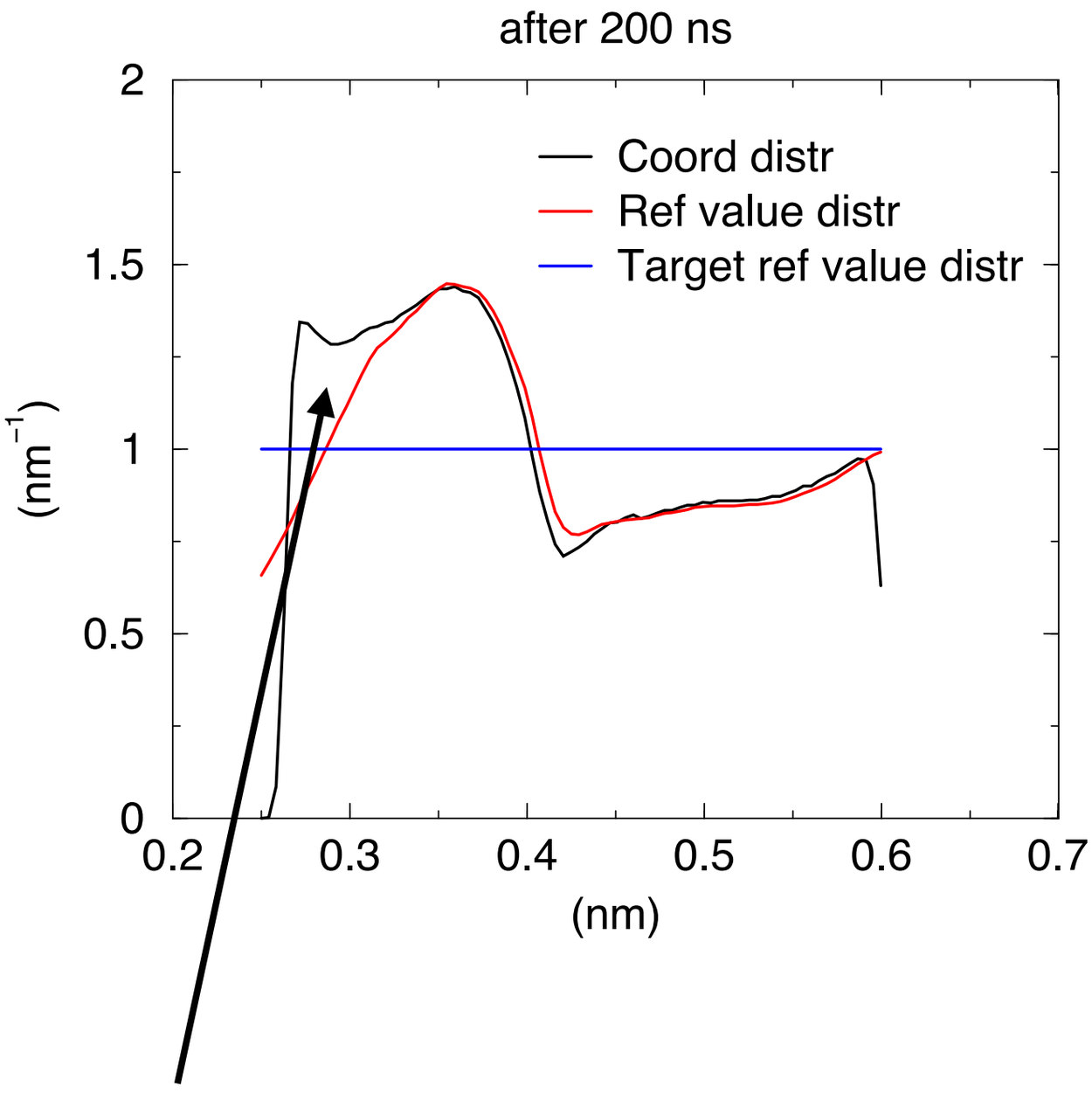
- Much too large parameters: slow convergence
- Much too small parameters: system might be pulled apart

Output of: gmx awh



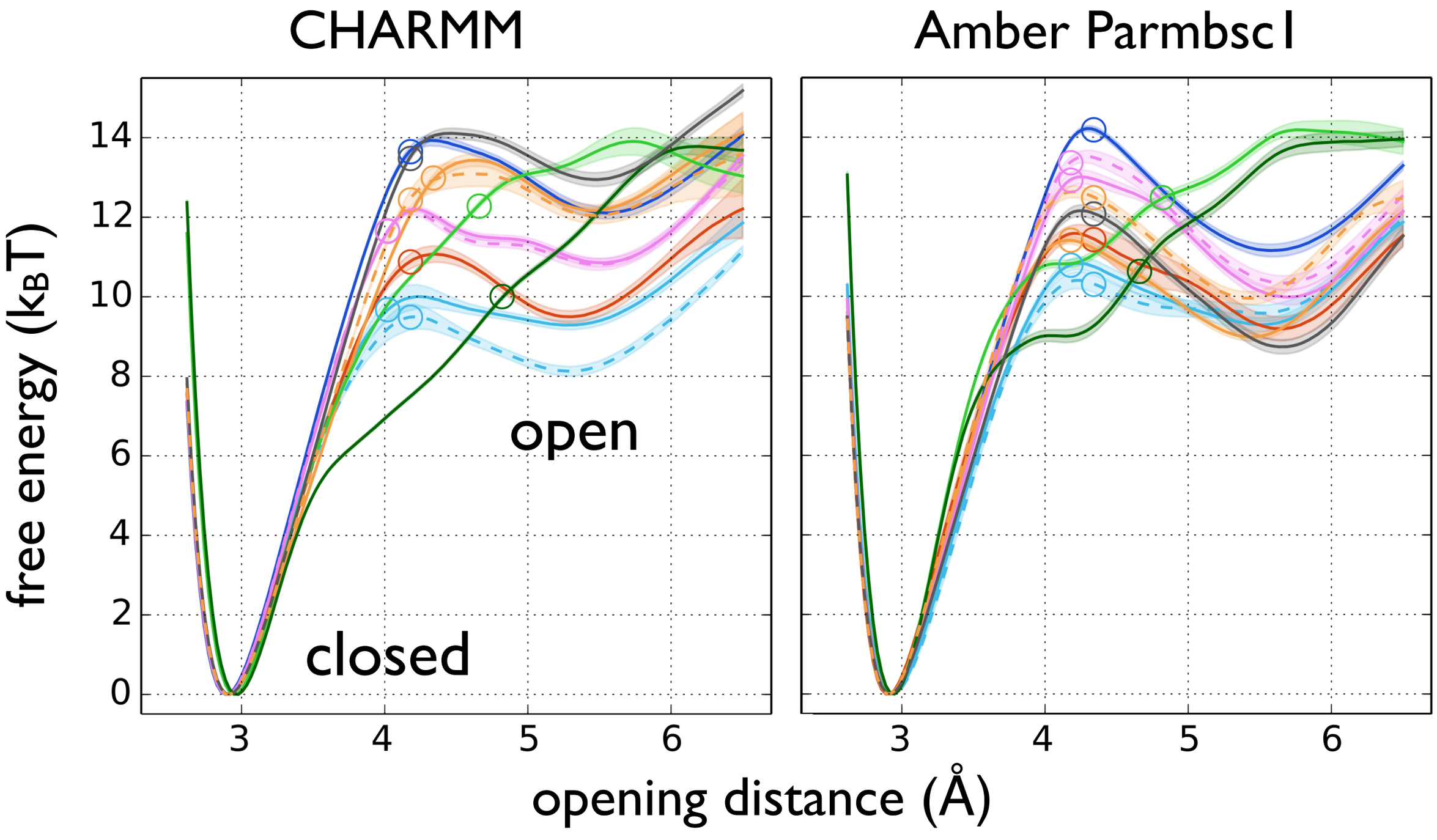
Distributions





typical for the free-energy being steeper
than the umbrella potential

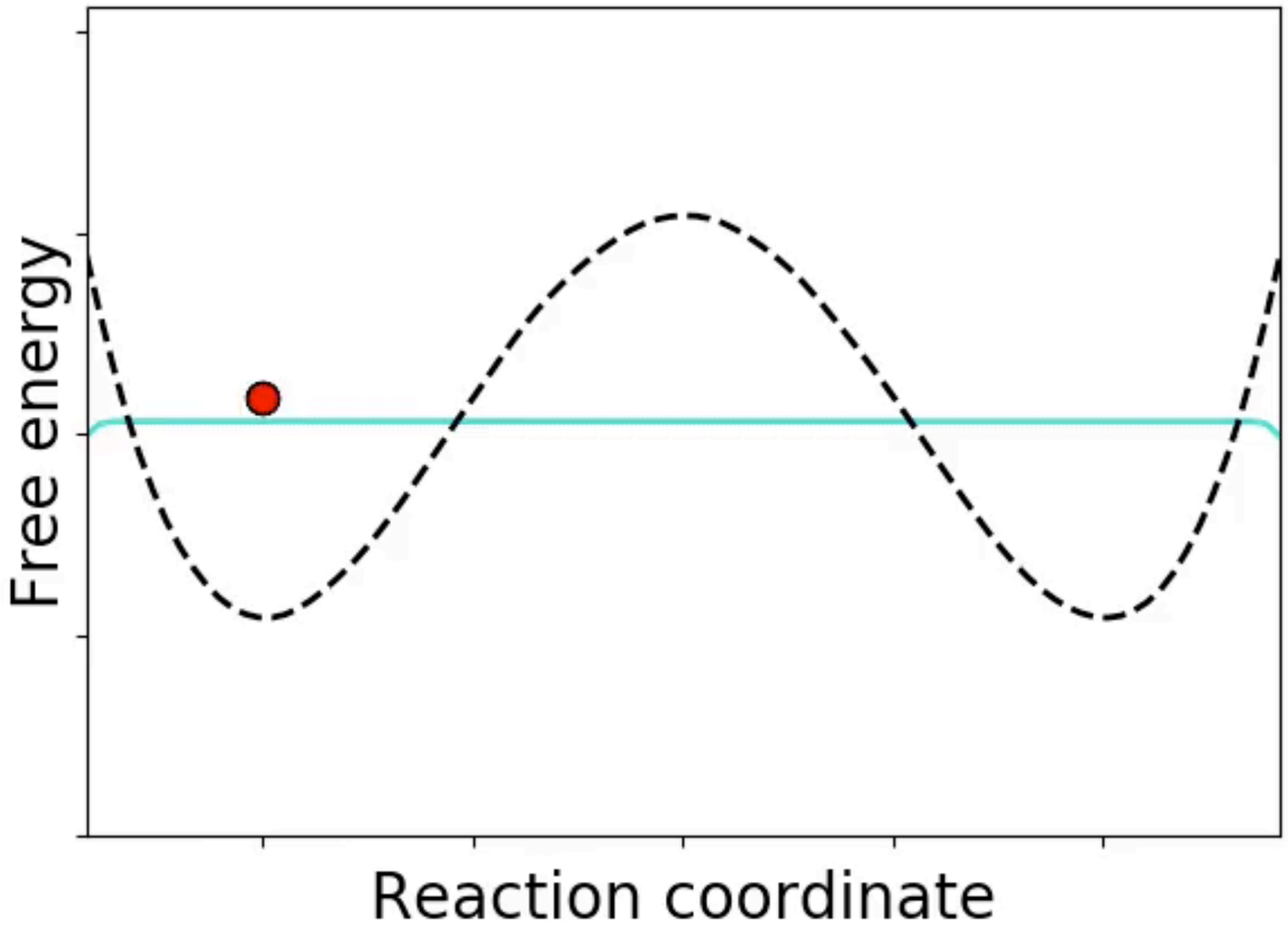
Results for different force fields and base pairs



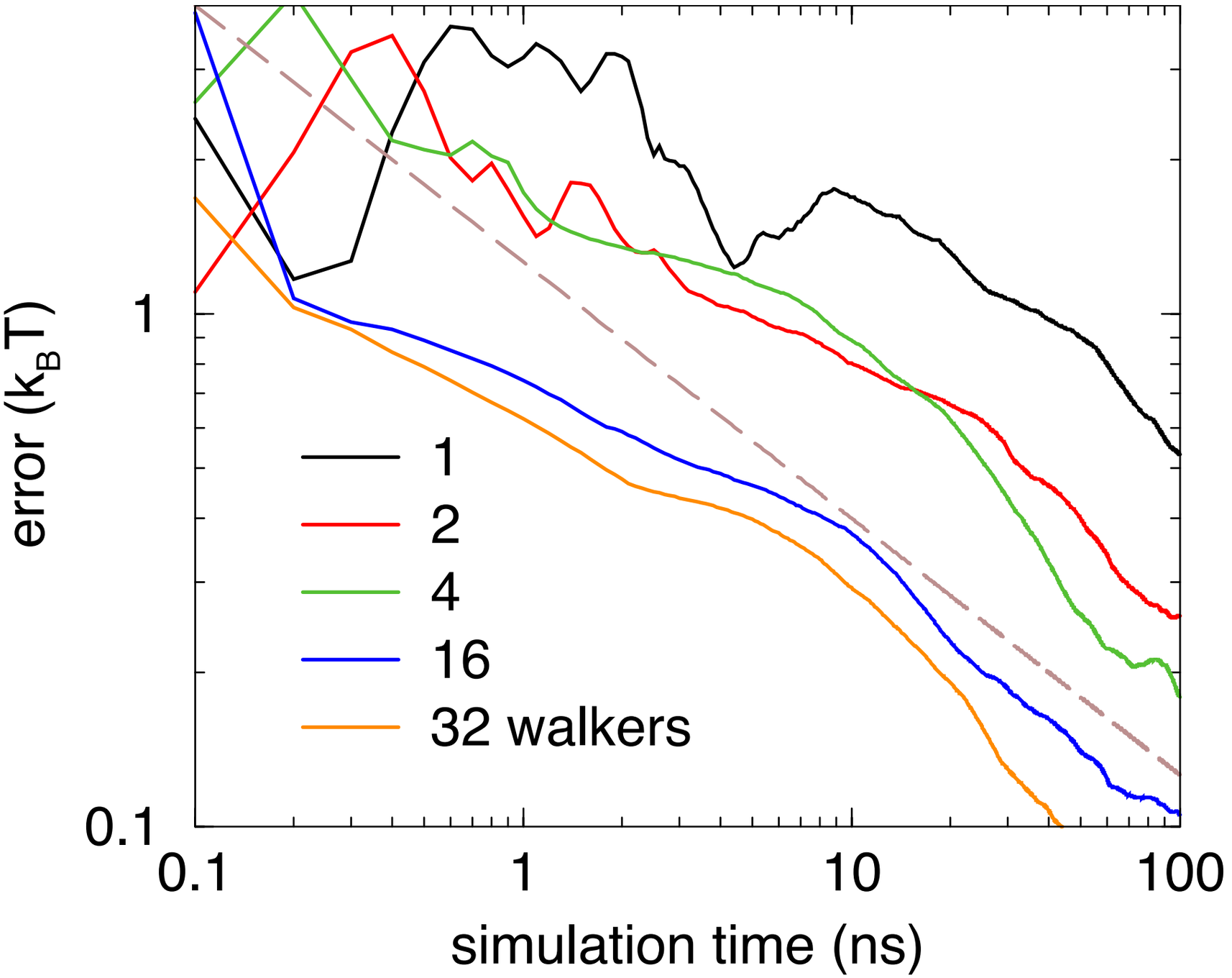
AWH with ensemble parallelism

- There is no reason to limit the sampling phase in AWH to a single simulation
- We can run an ensemble of multiple simulations
- Simulations sample independently and share the bias
  - How many copies can we run?
  - Or: how does the convergence depend on the number of copies?

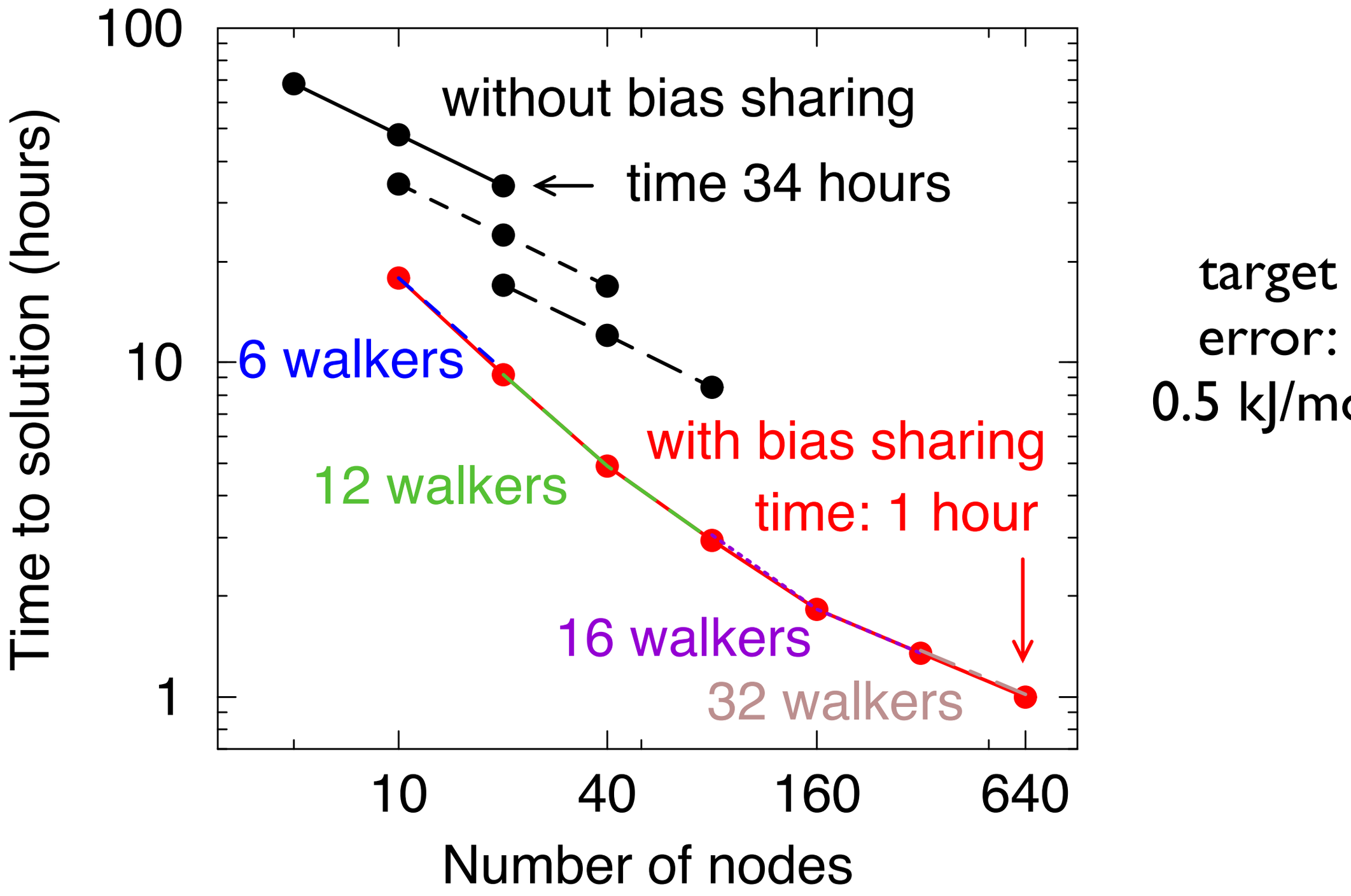
AWH with ensemble parallelism



Converge for #walkers



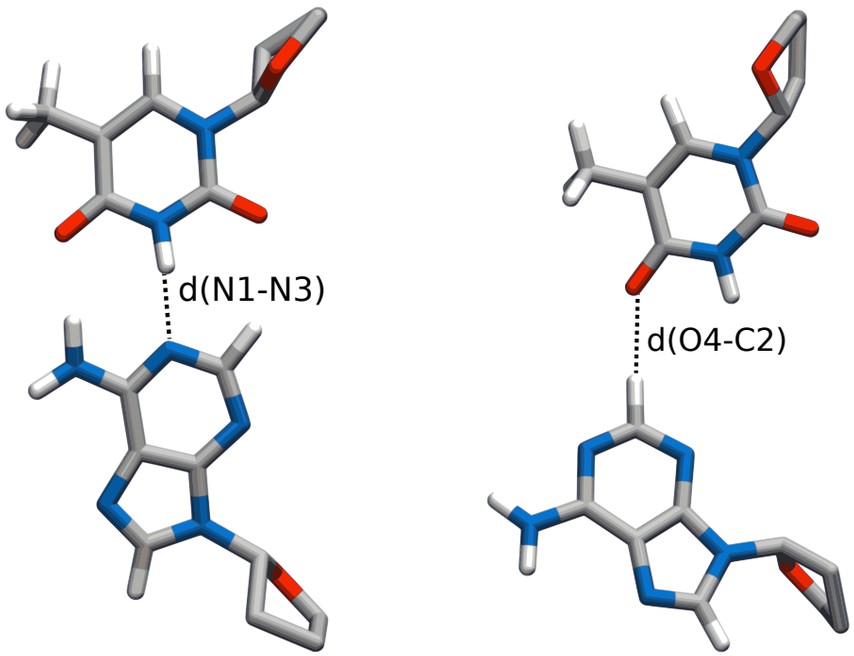
Time to solution for opening 20 base pairs on CSCS Piz Daint CPUs + P100 GPUs



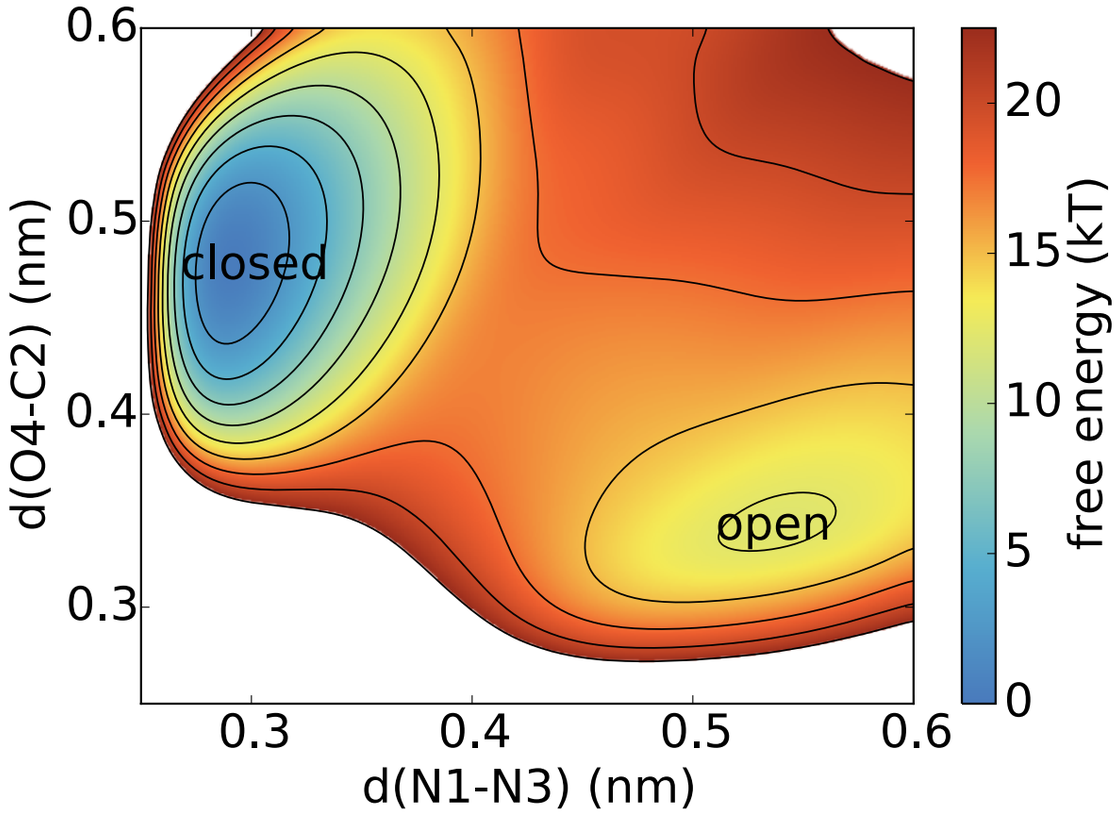
DNA with 2D AWH with FE cut-off of 20 k_BT

distance used in 1D case

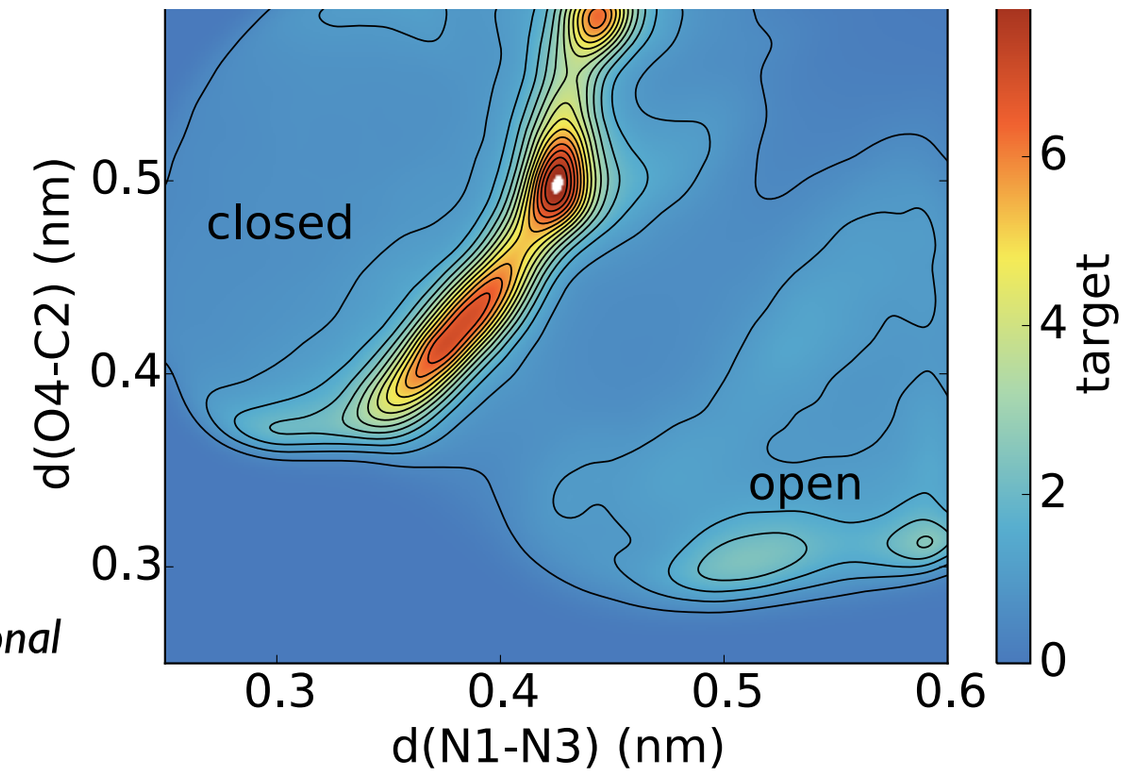
HB distance in open state



PMF

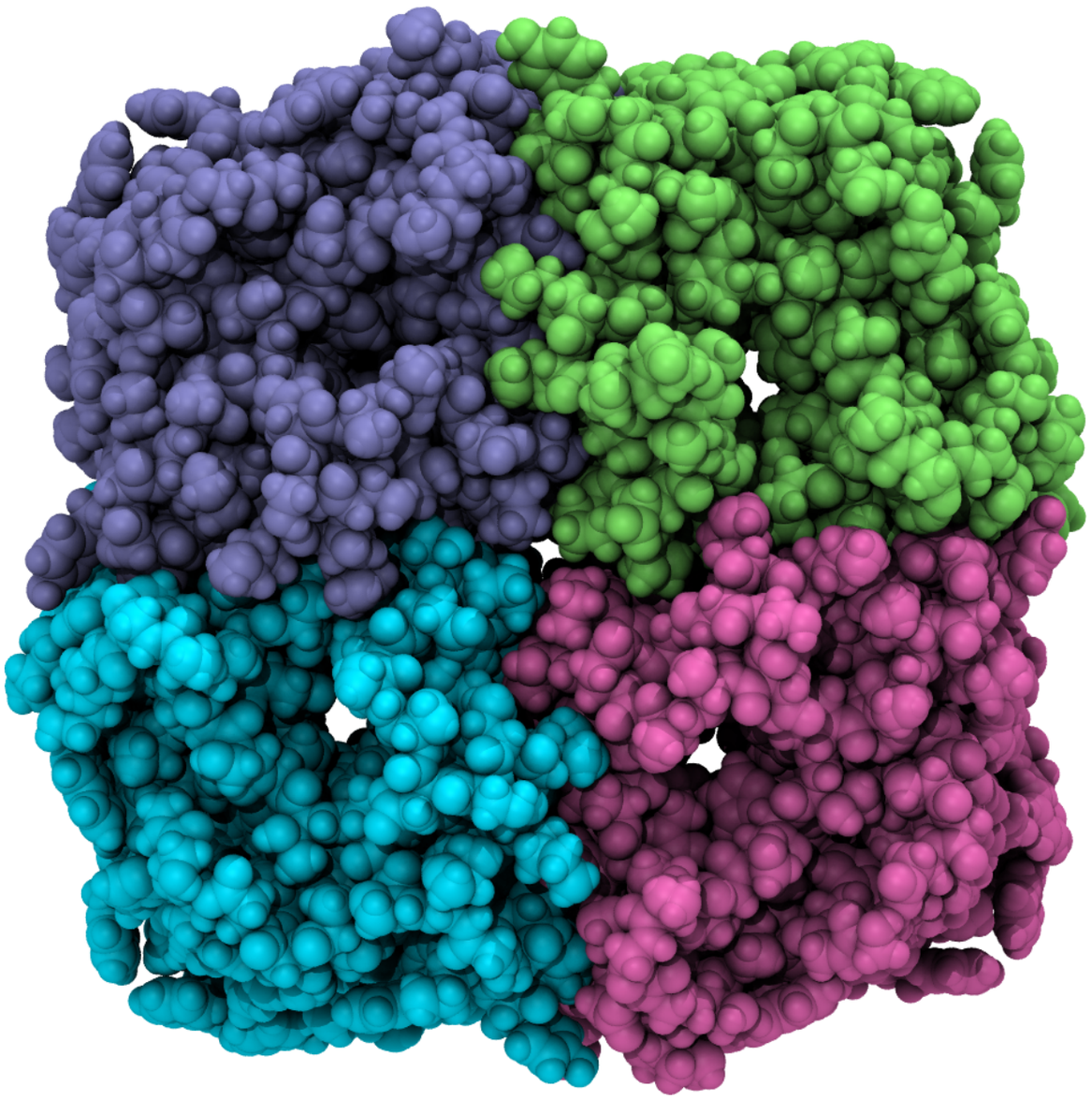


friction metric

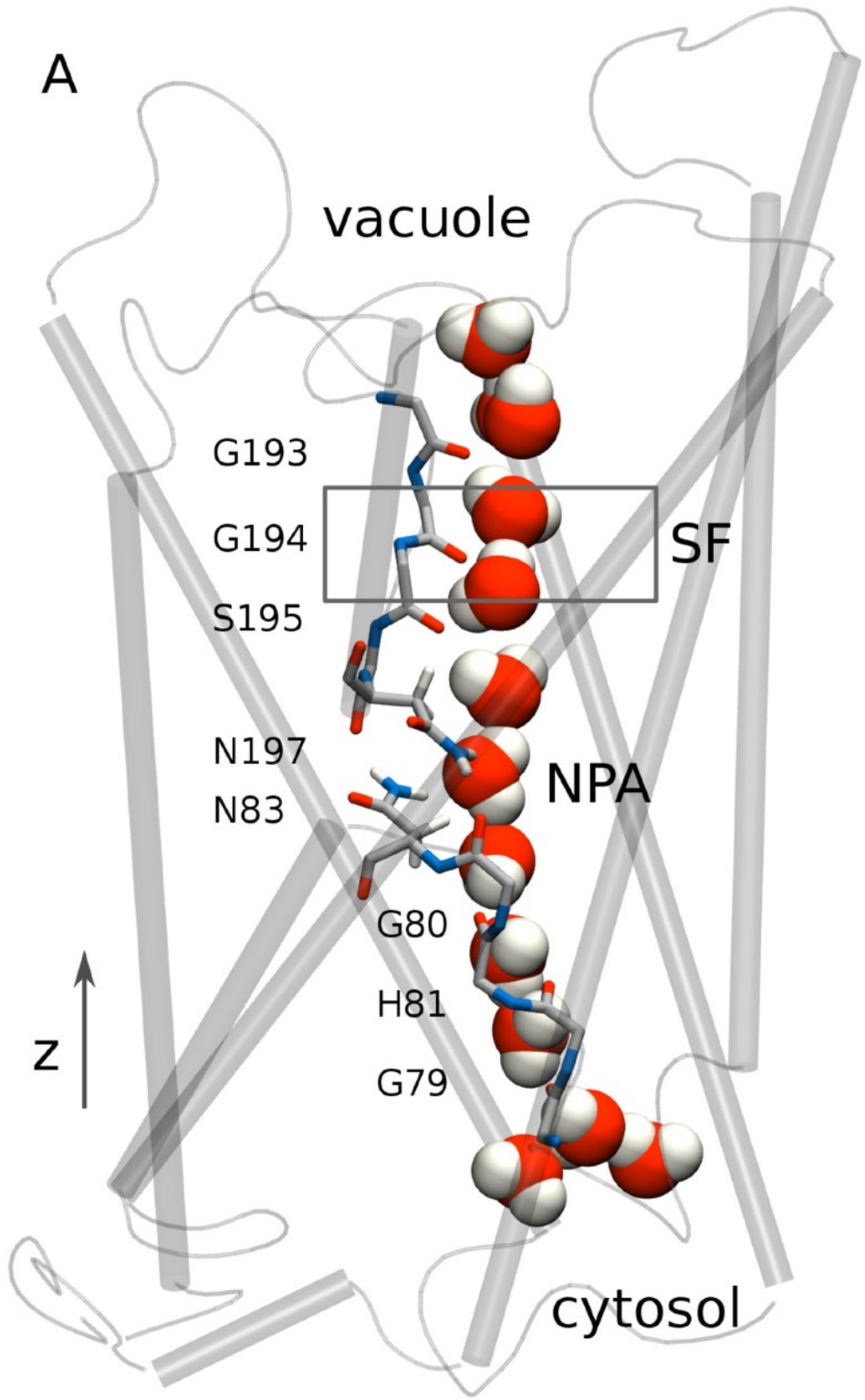


Riemann metric approach to optimal sampling of multidimensional free-energy landscapes, Phys. Rev. E, 98, 023312 (2018)

Aquaporin H_2O vs NH_3

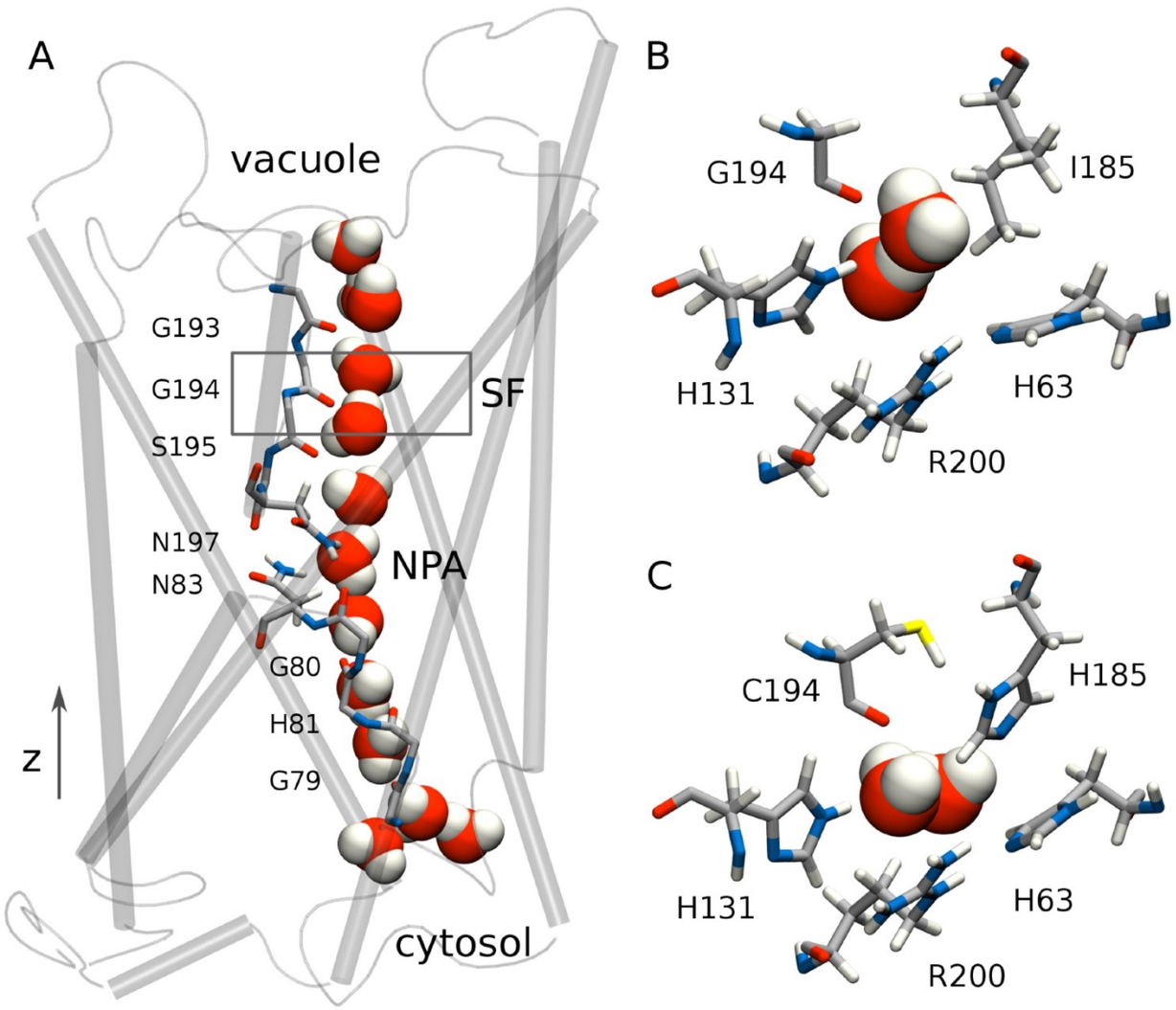


we run 4 independent AWH biases,
1 for each monomer

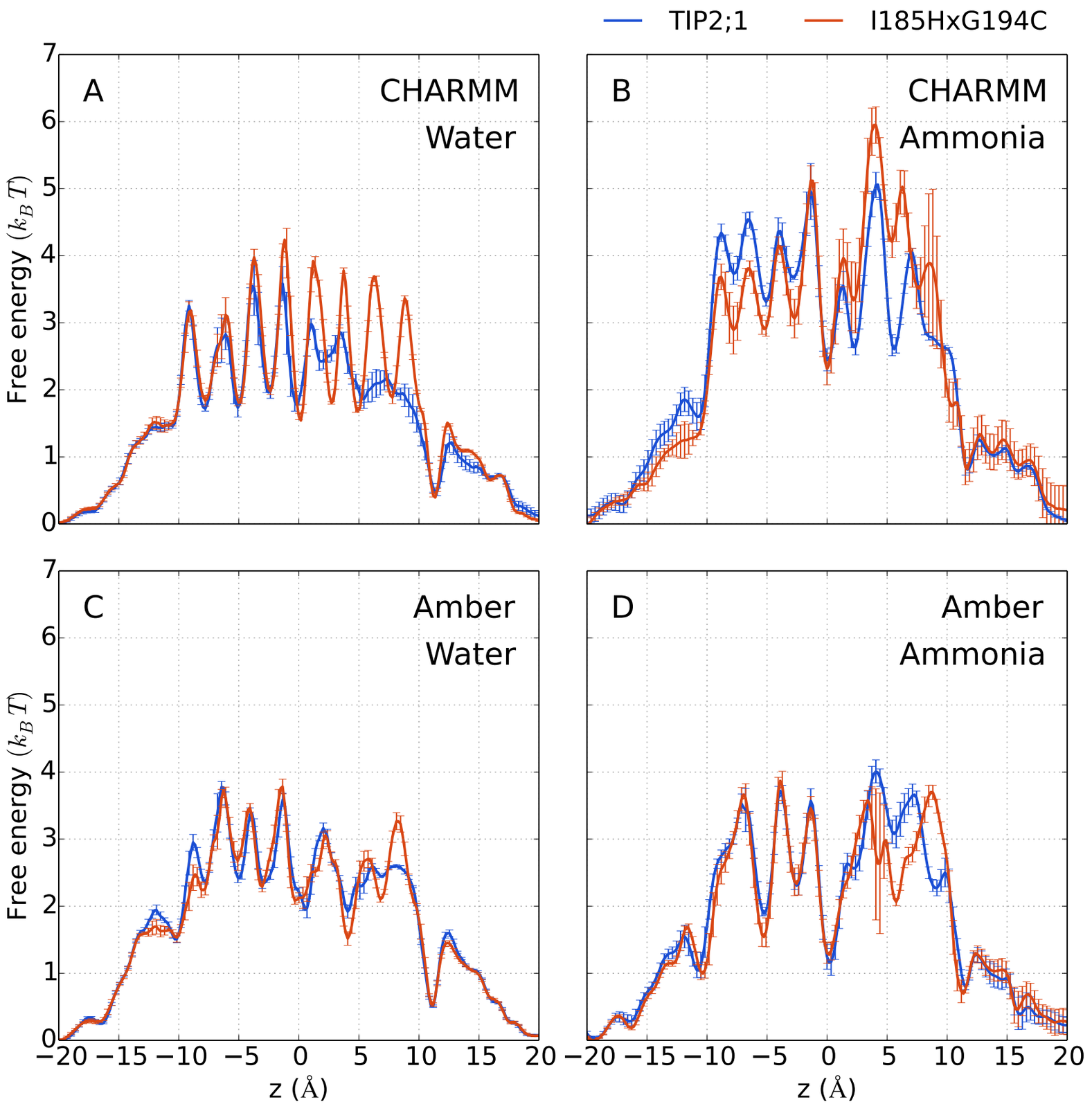


Aquaporin H_2O vs NH_3: methods

- For H_2O permeability we can use equilibrium simulations
- For NH_3 we use AWH
  - Umbrella sampling can block rearrangements of pore lining side chains
  - Forced motion can lead to non-equilibrium artefacts



Aquaporin H_2O vs NH_3: PMFs



AWH vs PLUMED

- AWH is very robust and easy to use
- AWH can currently only act on COM pull coordinates
- AWH is fully integrated into GROMACS
- AWH and COM pulling works efficiently in parallel
- We will extend the pull code with atomic contact and combinations of pull coordinates
  - Suggestions for extensions are welcome
- We will not cover all options range PLUMED supports

AWH conclusions

- Very robust accelerated sampling method
- Target distribution can be chosen freely
- One only parameter that controls convergence
  - and this parameter is not sensitive
- Supports 1D, 2D, 3D and 4D reaction coordinates
- Supports multiple walkers
- Works efficiently in parallel
- Note: The challenge is (still) is coming up with good RCs!