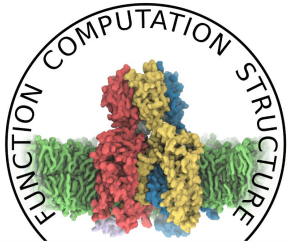
Applying the Accelerated Weight Histogram method to alchemical transformations

Berk Hess (hess@kth.se)
Magnus Lundborg (magnus.lundborg@ercopharma.com)



MOLECULAR BIOPHYSICS STOCKHOLM

ERCO PHARMA

Outline

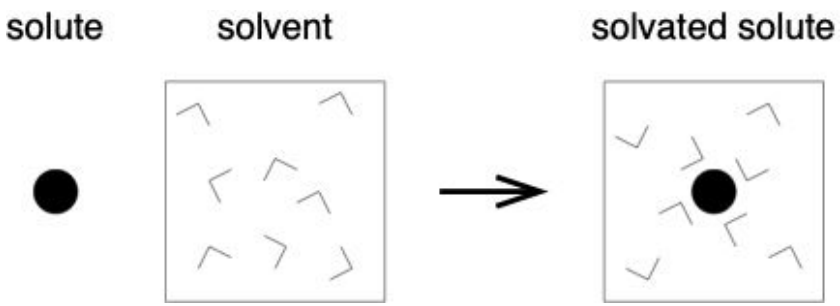
- Short introduction to alchemical free energy calculations
- The basics of AWH
- Examples of applications
- How to set up AWH free energy calculations in practice

Alchemical free energy calculations

Free energies differences give the relative population of states A and B:

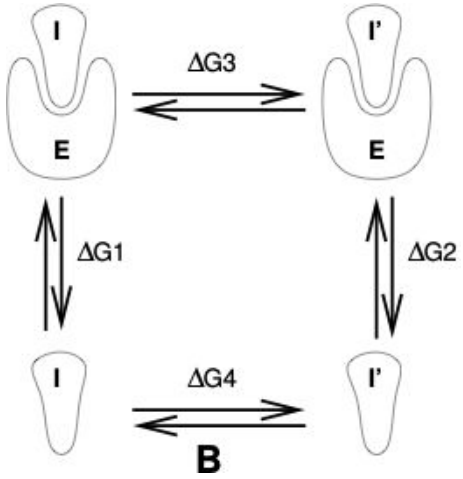
$P_A / P_B = \exp(\Delta G / k_B T)$, free energy difference $\Delta G = G_2 - G_1$

Solvation free energy



Differences in binding free-energy

$\Delta\Delta G = \Delta G_1 - \Delta G_2$
$= \Delta G_3 - \Delta G_4$



How to couple states A and B?

Coupling parameter approach

- Add a coupling parameter λ to the Hamiltonian:

$$H = H(p,q;\lambda)$$

$$H(p,q;0) = H_A(p,q)\ ,\ H(p,q;1) = H_B(p,q)$$

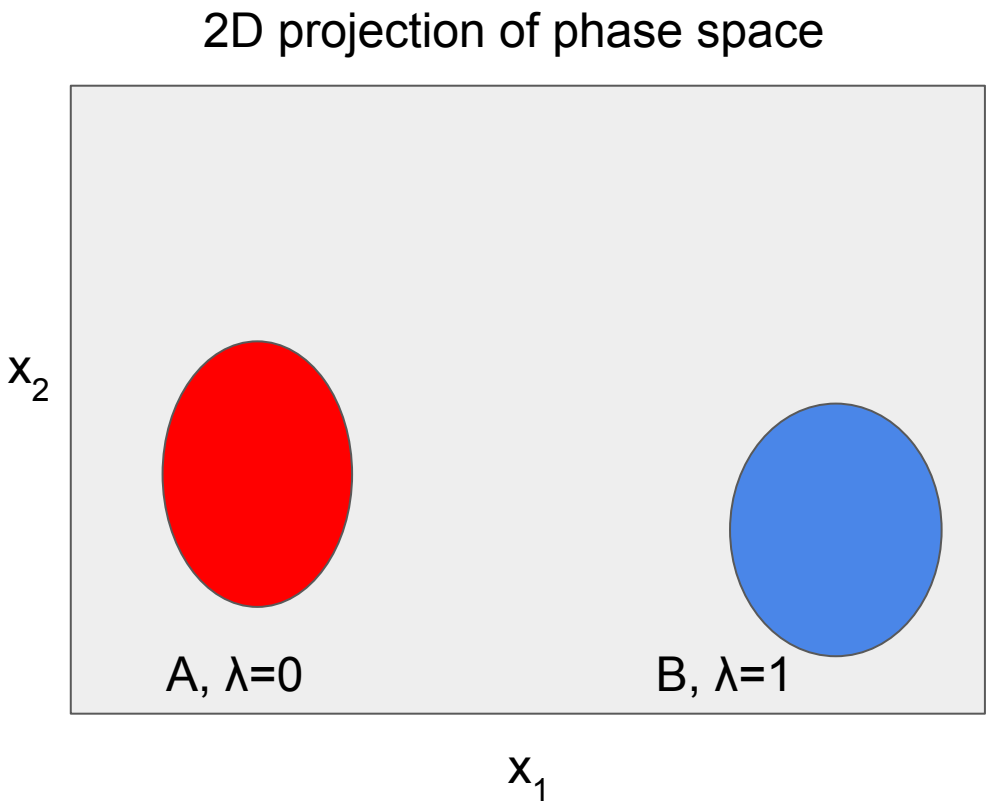
- Free energy is then given by:

$$G_B(p,T) - G_A(p,T) = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{NPT;\lambda} d\lambda$$

There is often no overlap between the end states

In most cases the phase space distributions of states A and B have no overlap

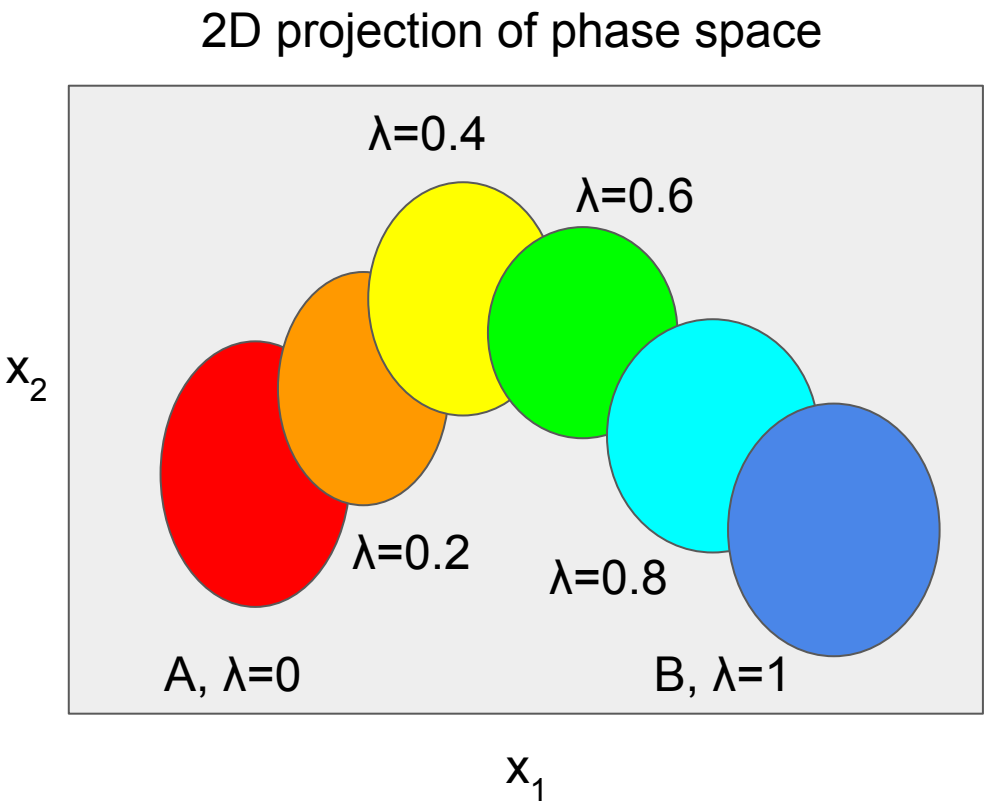
One could compute dH/dλ in A and B, but dH/dλ might vary a lot in between



We need to create overlap

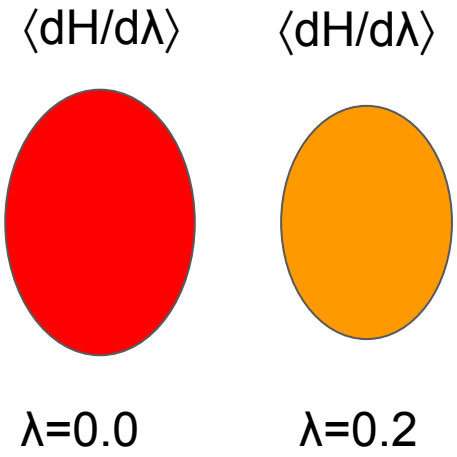
The λ-path is arbitrary, only the end states are given

- We need to choose an efficient path
- We need to choose points along the path

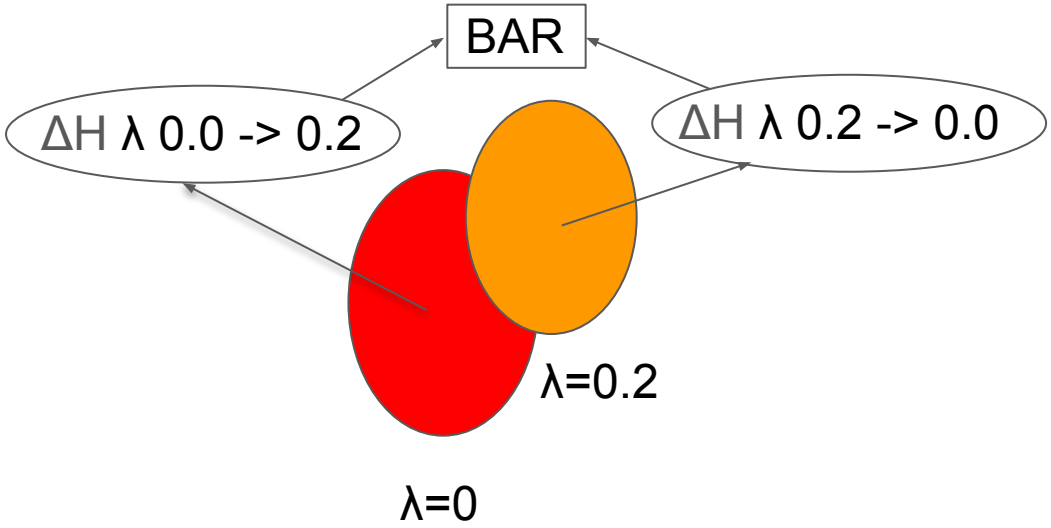


Computing the free energy differences

Thermodynamic integration
of dH/dλ, local operation, quadrature (integration) **error**



Bennett acceptance ratio
Uses Hamiltonian (energy) differences between two λ values both ways simultaneously



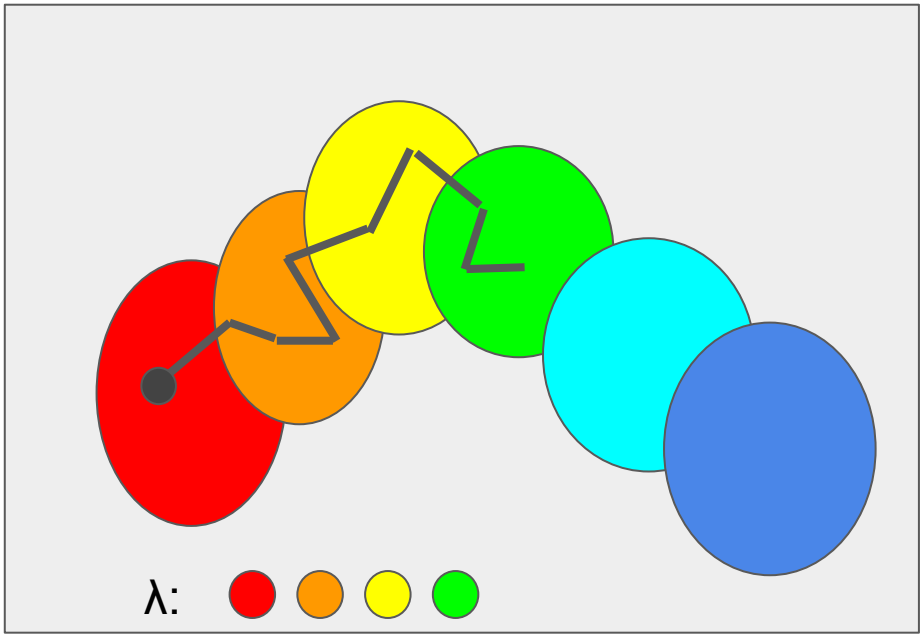
AWH moves λ dynamically

In a single simulation λ moves dynamically using Monte Carlo

A bias potential is added to λ to achieve uniform sampling

Hamiltonian differences are computed to all other λ values (like in MBAR)

2D projection of phase space



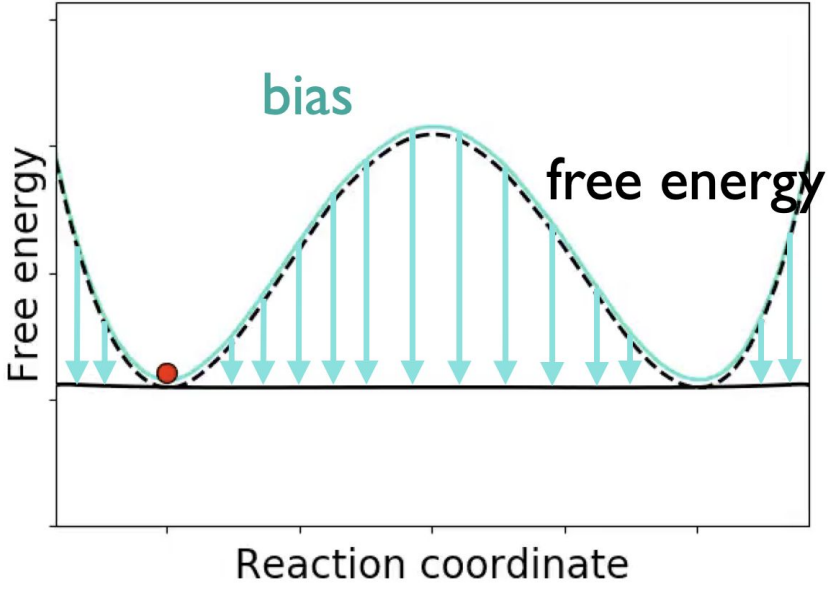
History of the Accelerated Weight Histogram method

- The method was originally developed as a general enhanced sampling method by Jack Lidmar "Improving the efficiency of extended ensemble simulations: The accelerated weight histogram method," Phys. Rev. E **85**, 056708 (2012). doi: 10.1103/PhysRevE.85.056708.
- Adapted for collective coordinates in MD and implemented in GROMACS by Viveca Lindahl V. Lindahl, J. Lidmar, and B. Hess, "Accelerated weight histogram method for exploring free energy landscapes," *The Journal of chemical physics*, **141** [4] 044110 (2014).
  BioExcel webinar in 2020: https://bioexcel.eu/webinar-accelerating-sampling-in-gromacs-with-the-awh-method-2019-09-24/
- A metric was devised for reaction coordinates V. Lindahl, J. Lidmar, and B. Hess, "Riemann metric approach to optimal sampling of multidimensional free-energy landscapes," Phys. Rev. E **98**, 023312 (2018). doi: 10.1103/PhysRevE.98.023312.
- A manuscript on AWH for alchemical free-energy calculations has been submitted (Magnus Lundborg, Jack Lidmar, Berk Hess)

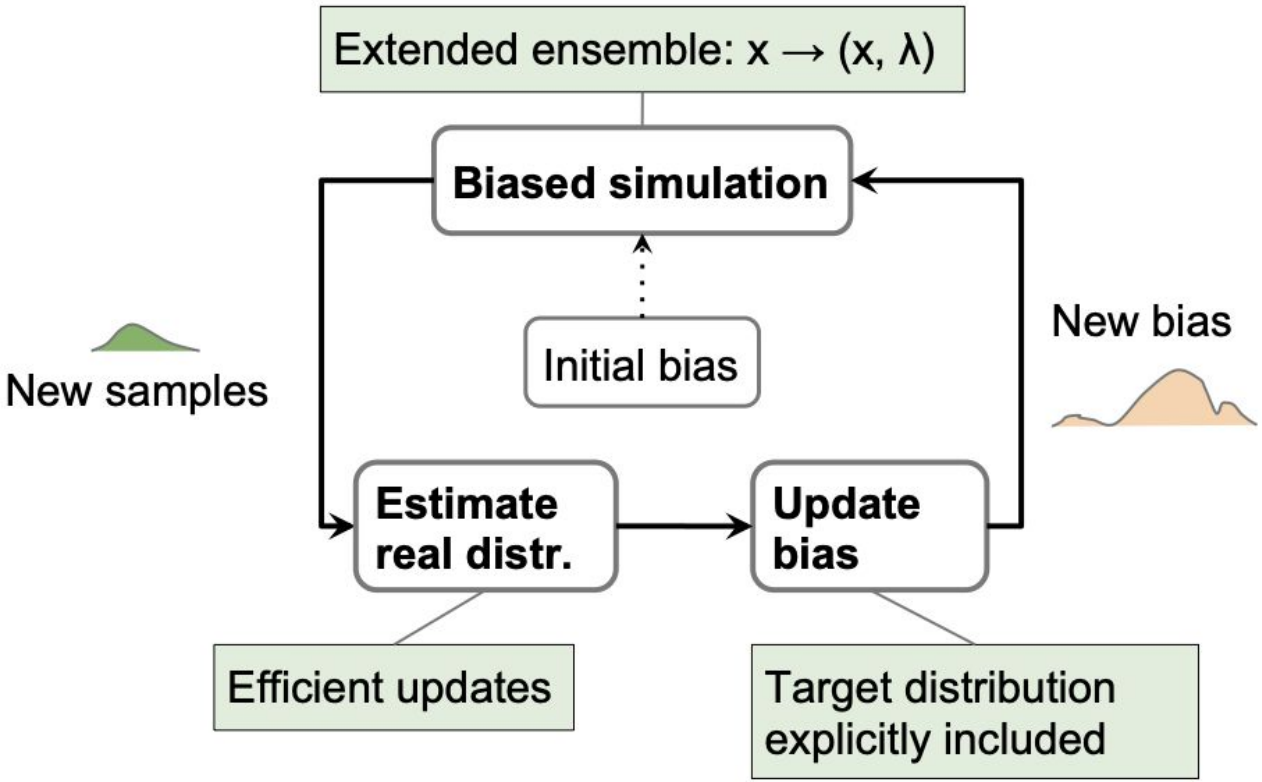
Flatten the free-energy landscape

Trick: add a bias potential to make the effective potential flat

Issue: the potential (or free-energy) is what we are after!



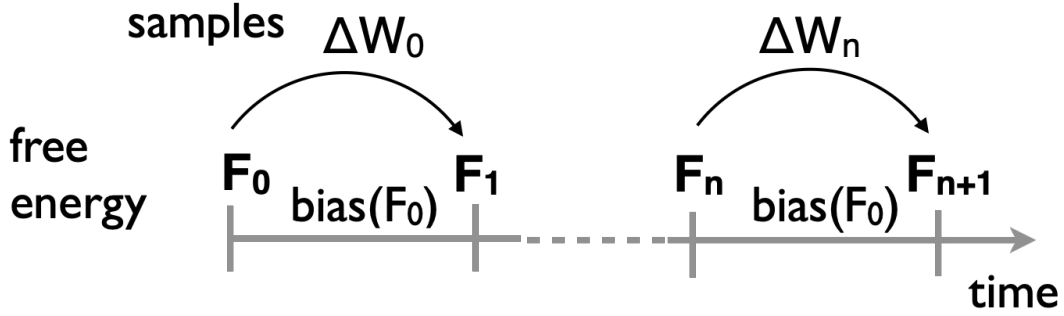
AWH schemetically



The Accelerated Weight Histogram method

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

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



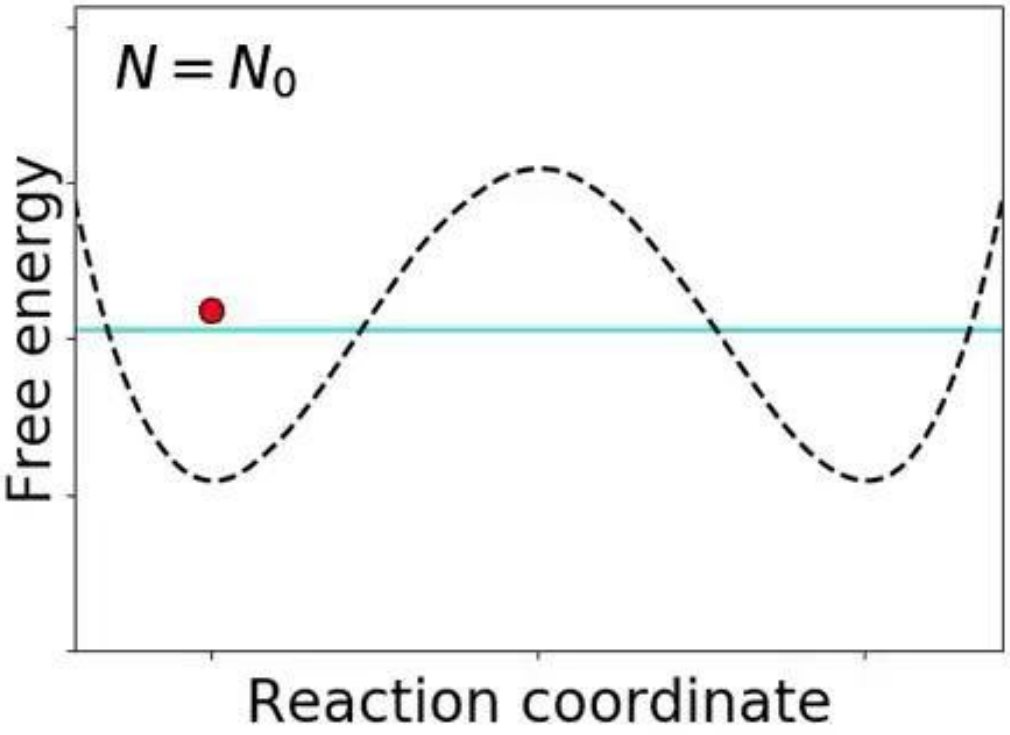
$$\Delta F_n(\lambda) = -\ln\left(1 + \frac{\Delta W_n(\lambda)}{N_n \pi(\lambda)}\right) \sim \frac{1}{N_n}$$

collected samples

prior number of samples

target distribution (flat)

AWH in action (on model double-well potential)



Ensemble parallelism using multiple “walkers”

AWH can be “trivially” parallelized by having multiple copies of the system, so called “walkers”, contribute to the same AWH bias

- Work for any number of walkers
- Reduces time to solution

Choices for AWH free energy calculations

There are few option / parameter choices to make:

- Number of walkers -> choose freely according to your resources
- Number of lambda points -> choose sufficiently many
- AWH diffusion coefficient / initial error -> choose 0.001 ps^{-1} / 10 kJ/mol

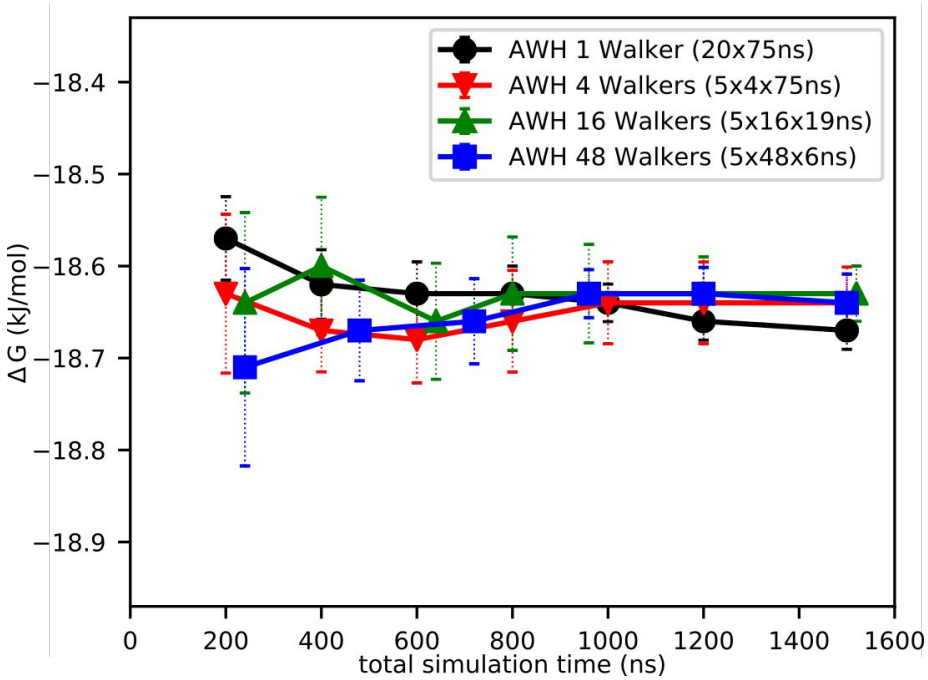
The initial update size for AWH needs to be set, we do this through a diffusion coefficient and initial error. These parameters are not sensitive

Advantages of AWH: easy to set up, no sensitive parameters!

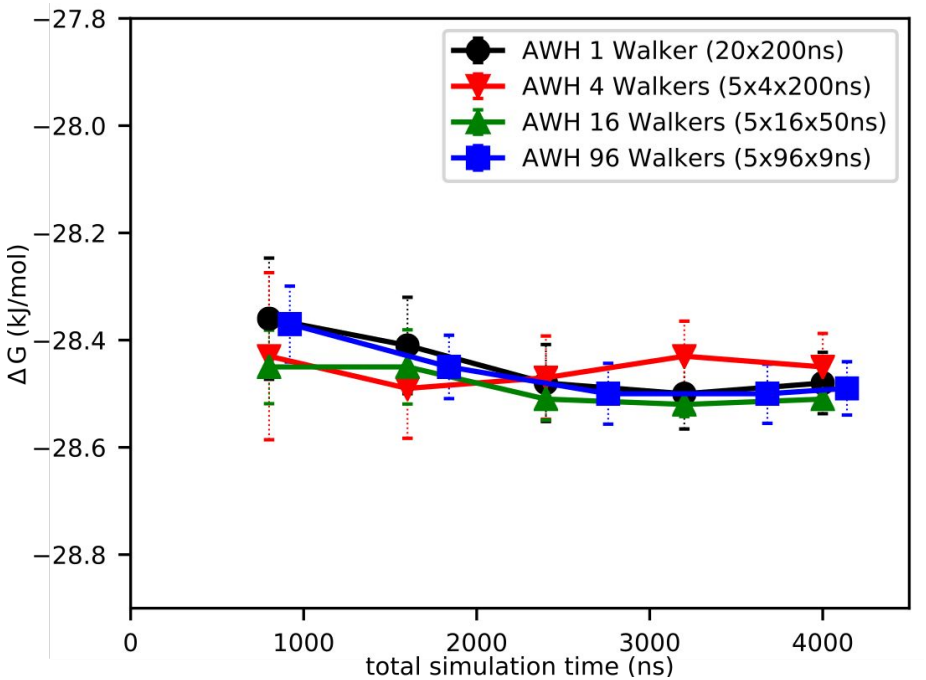
Some examples

Number of walkers

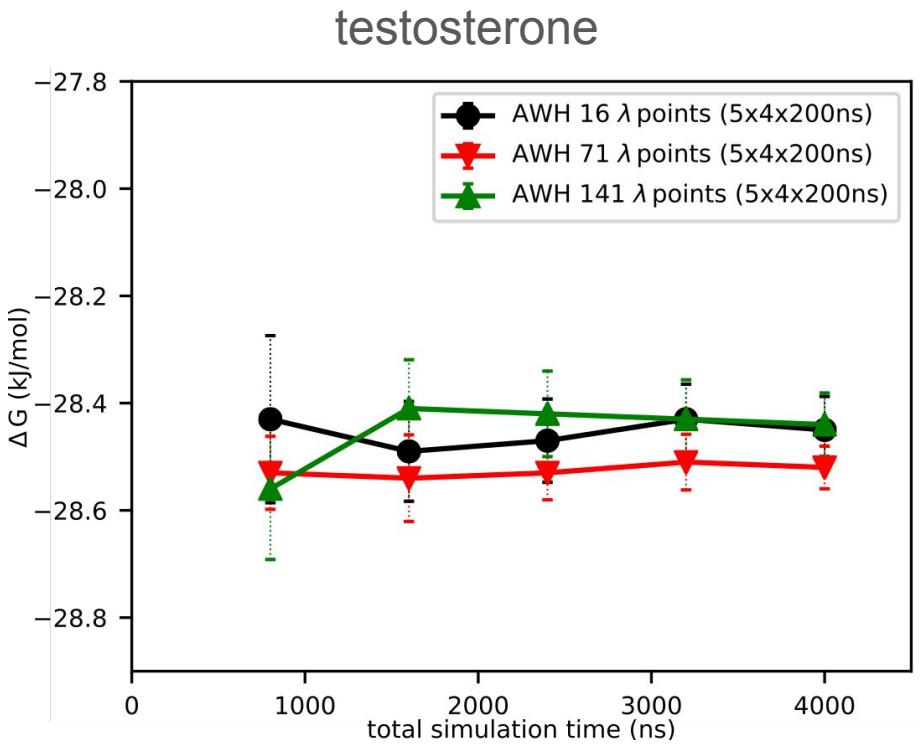
ethanol



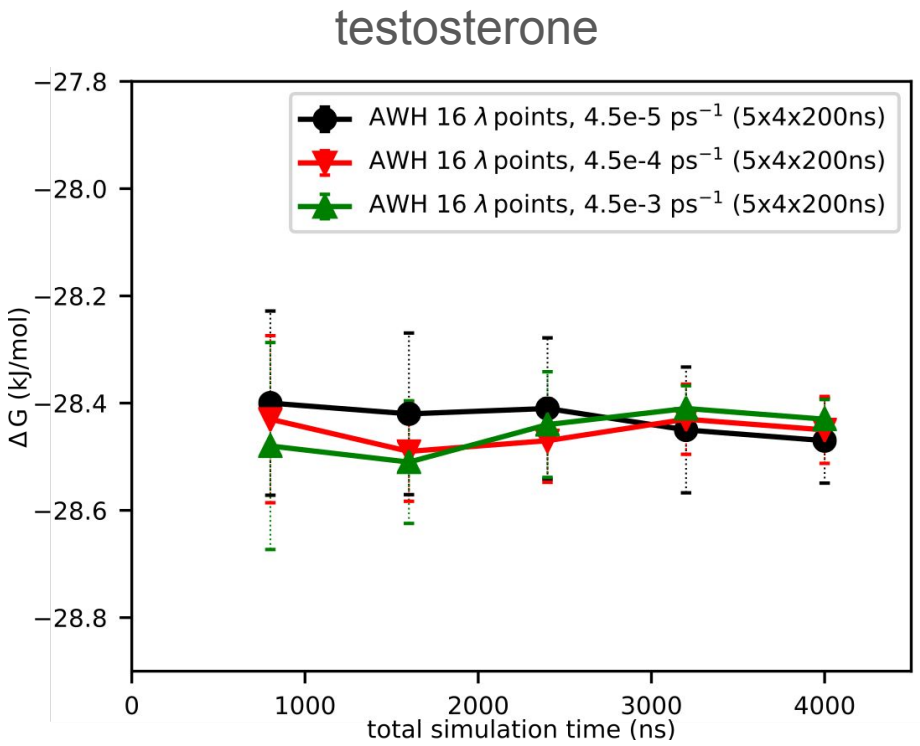
testosterone



Number of lambda points

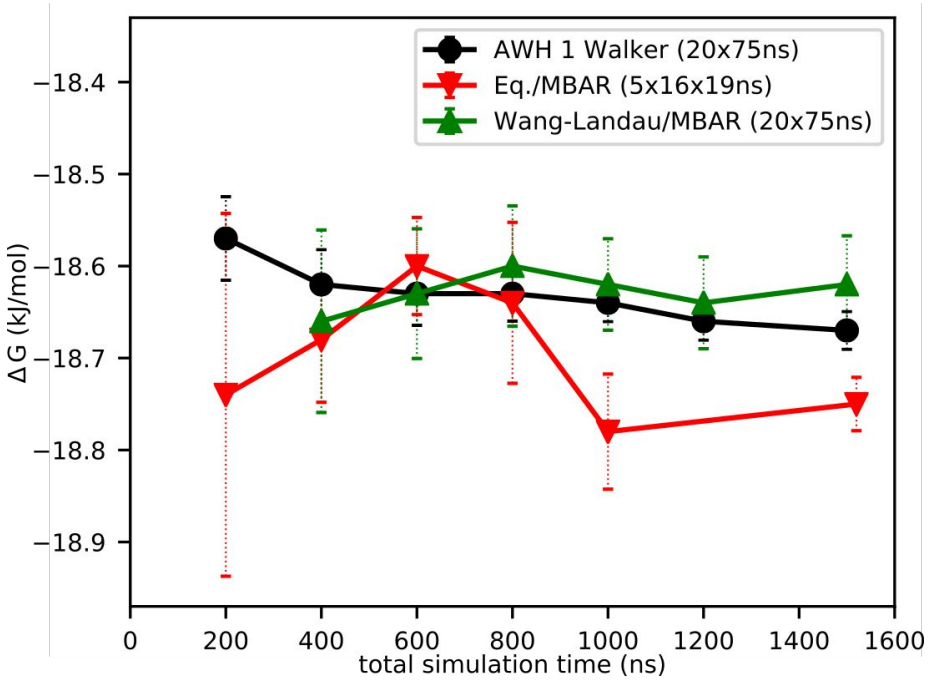


AWH diffusion coefficient

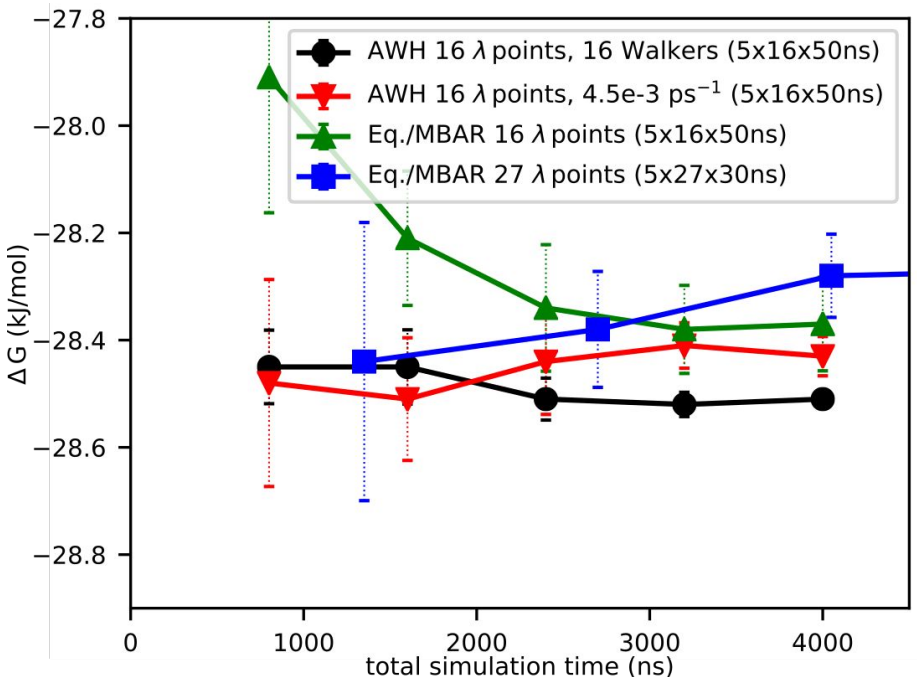


Convergence

ethanol

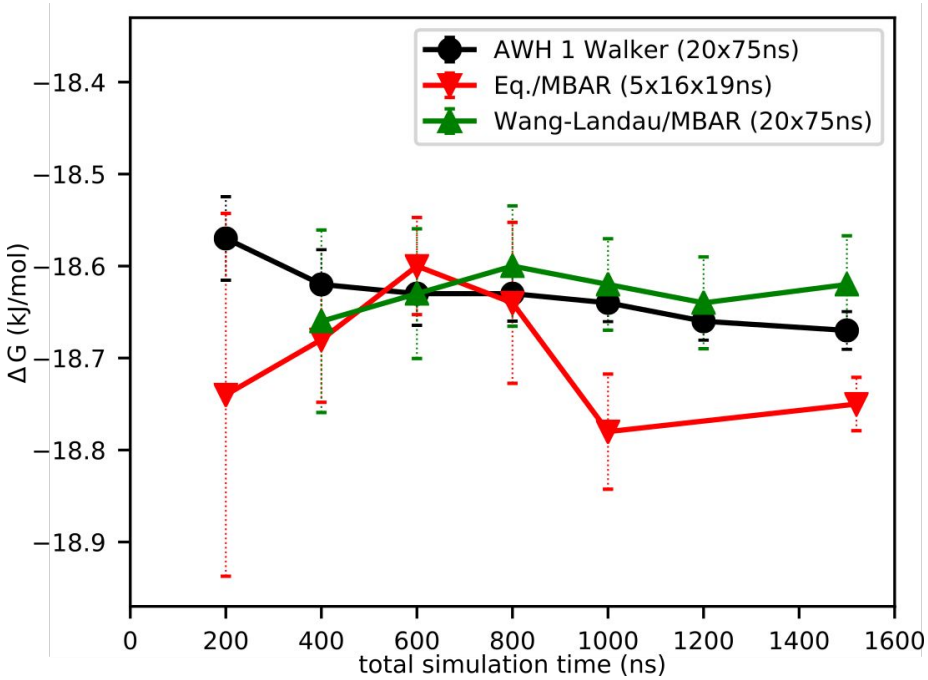


testosterone

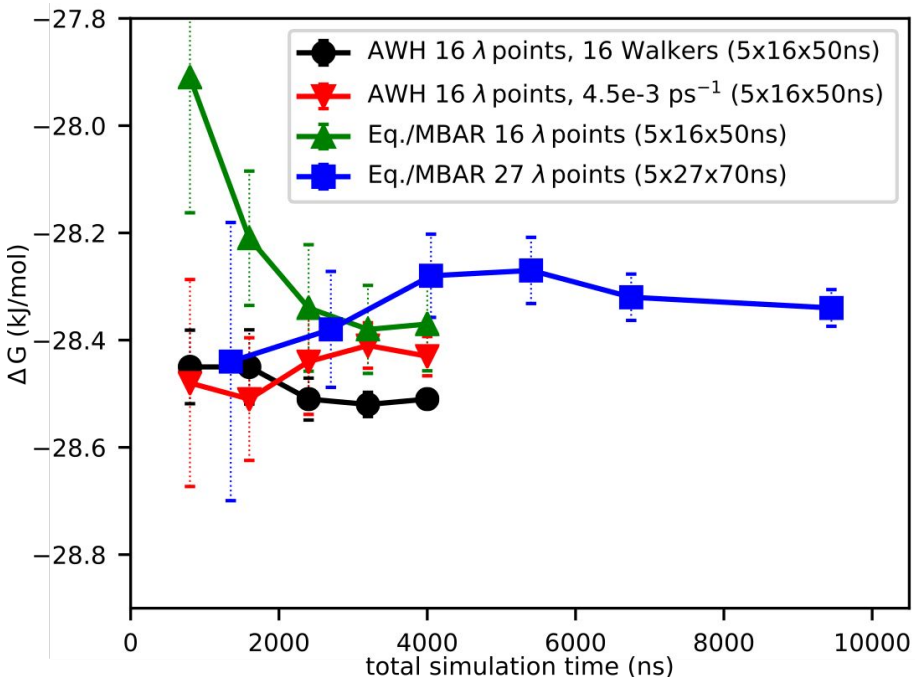


Convergence

ethanol

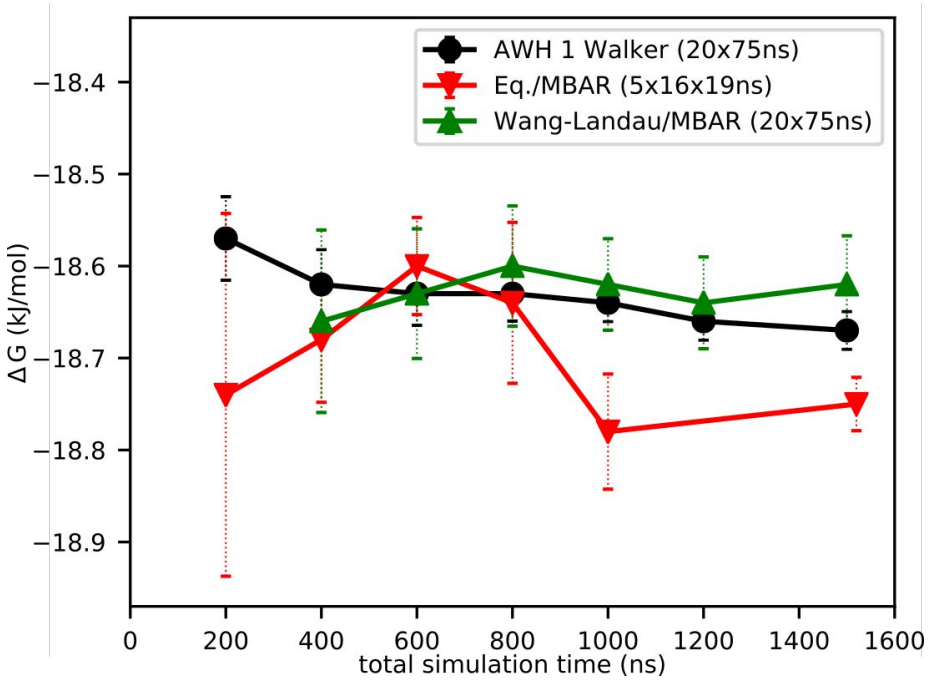


testosterone

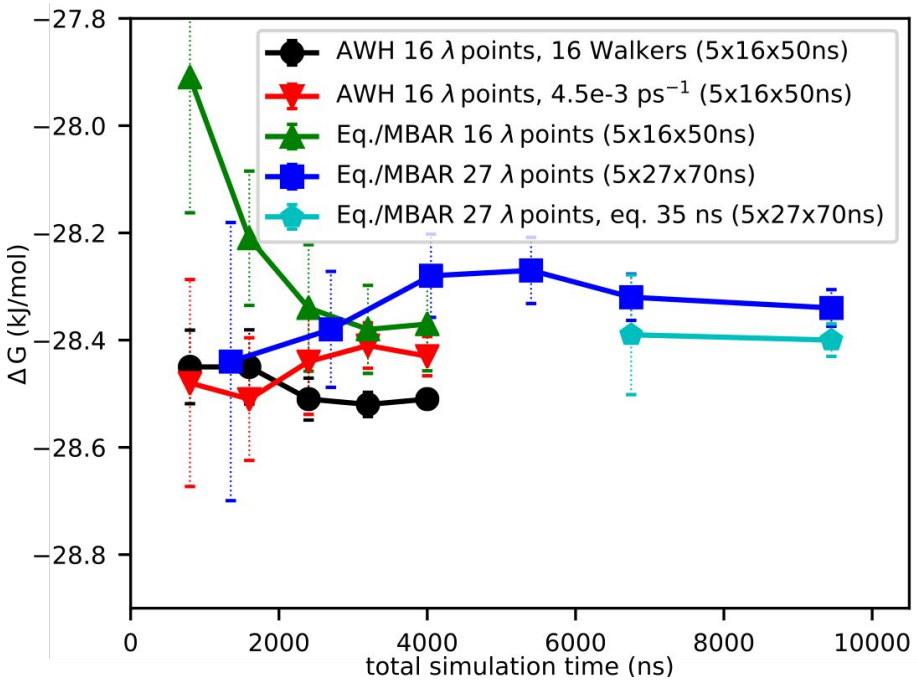


Convergence

ethanol



testosterone



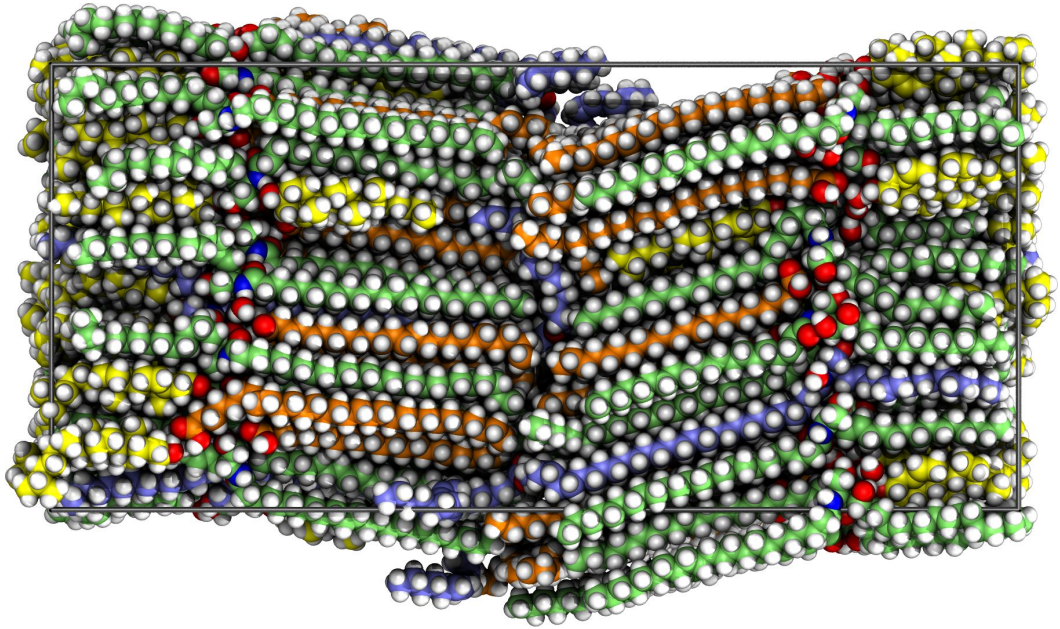
Permeability calculations through skin lipid barrier

- Complex Lipid system in near gel-state - slow diffusion through the system and slow convergence.

Lundborg et al., J. Struct. Biol. 2018, 203, 149-161

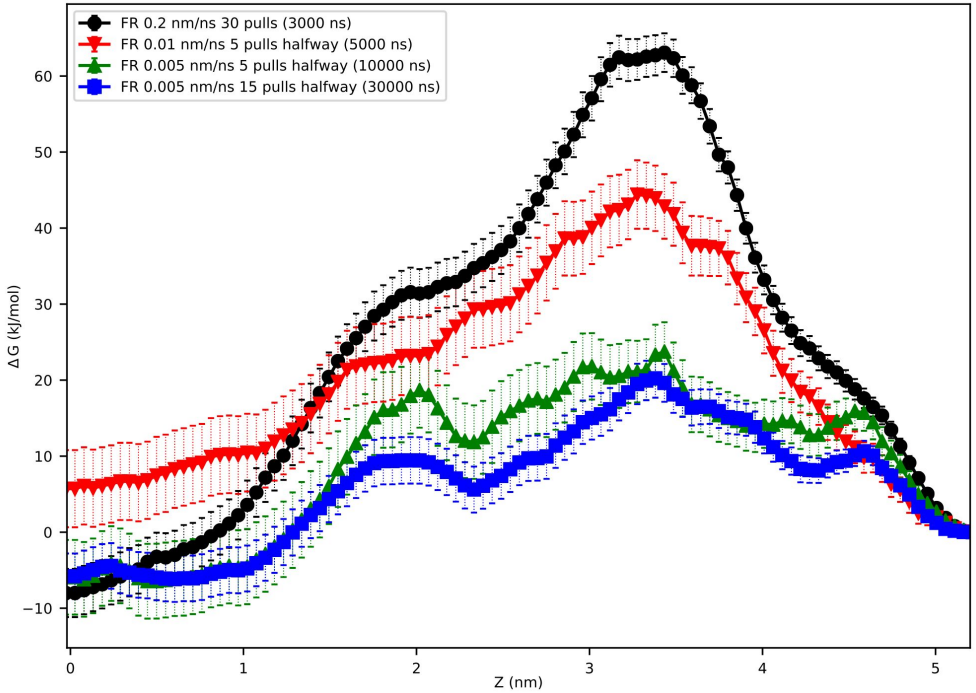
Permeability calculations through skin lipid barrier

- Lipid system in near gel-state - slow diffusion through the system and slow convergence.



Permeability calculations through skin lipid barrier

- Nonequilibrium pulling (FR method) was sensitive to the pulling speed.

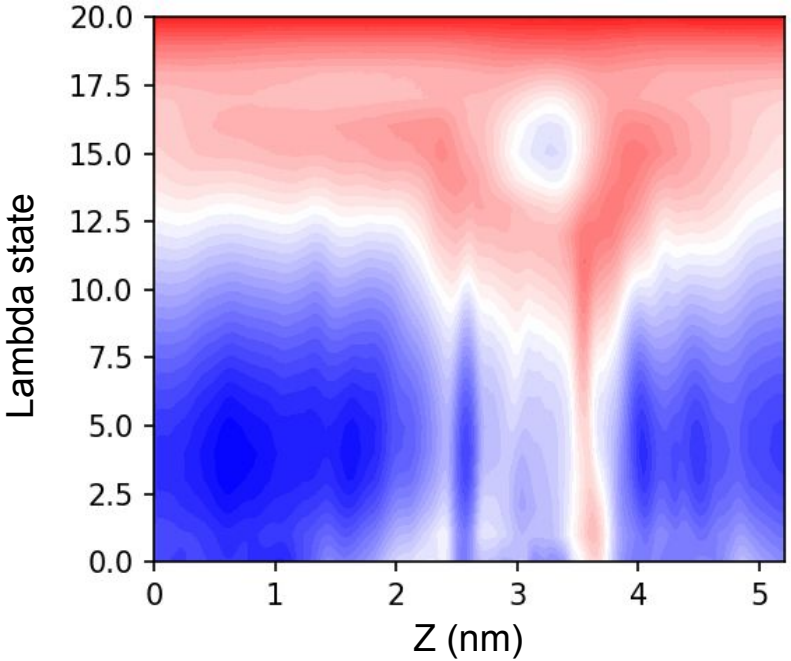


Permeability calculations through skin lipid barrier

- Umbrella sampling required very long simulations and long equilibration times.

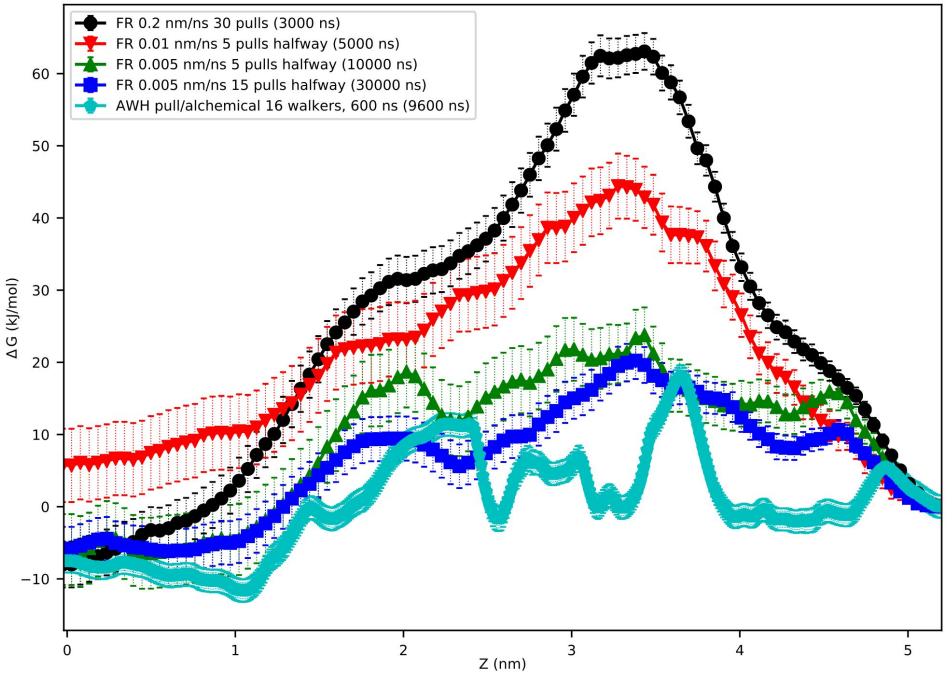
Permeability calculations through skin lipid barrier

- Combining an alchemical reaction coordinate dimension with a pull dimension.



Permeability calculations through skin lipid barrier

- Combining an alchemical reaction coordinate dimension with a pull dimension.



Error estimation

- No error estimation from the analysis.
- Repeat simulations
  - Most certain error estimate for equilibrium simulations as well.

Relevant GROMACS simulation parameters

- free-energy
  - Lambda states
    - Not necessary to optimize the lambda point distribution.
  - calc-lambda-neighbors = -1

Relevant GROMACS simulation parameters

- awh
  - awh-potential = umbrella
  - awh-nstsample : a multiple of nstcalcenergy
  - awh1-dim1-coord-provider = fep-lambda
  - awh1-dim1-start and -end : the indices of the first and last lambda points
  - awh1-dim1-diffusion : as high as “possible” (~10^{-2} to 10^{-4} ps^{-1})
  - awh1-equilibrate-histogram : recommended if running multiple walkers

Example mdp input (free-energy options)

```
free-energy             = yes
couple-lambda0          = none
couple-lambda1          = vdwq
couple-moltype          = ethanol
couple-intramol         = no
init-lambda-state       = 15
vdw-lambdas             = 1.0 1.00 1.00 1.00 1.00 1.00 0.90 0.80 0.70
0.60 0.50 0.40 0.30 0.20 0.10 0.00
coul-lambdas            = 1.0 0.80 0.60 0.40 0.20 0.00 0.00 0.00 0.00
0.00 0.00 0.00 0.00 0.00 0.00 0.00
calc-lambda-neighbors   = -1
separate-dhdl-file      = no
sc_alpha                = 0.5
sc_sigma                = 0.3
sc_power                = 1
sc_coul                 = no
```

Example mdp input (awh options)

```
awh                    = yes
awh-potential          = umbrella
awh-nstout             = 500000
awh-nbias              = 1
awh-nstsample          = 10
awh-nsamples-update    = 10
awh1-error-init        = 10
awh1-equilibrate-histogram = no
awh1-target            = constant
awh1-growth            = exp-linear
awh1-ndim              = 1
awh1-dim1-coord-provider = fep-lambda
awh1-dim1-coord-index = 1
awh1-dim1-start        = 0
awh1-dim1-end          = 15
awh1-dim1-diffusion    = 0.002
```

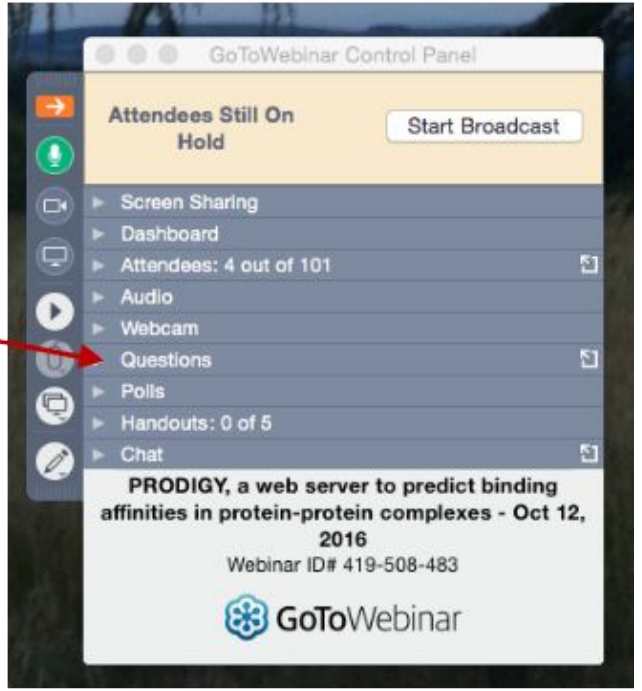
Summary

- AWH can be used for alchemical free energy calculations in GROMACS 2021
- Not very sensitive to the input parameters, such as the distribution, and number, of lambda states.
- Converges at least as quickly as equilibration simulations.
- Trivial parallelization with flexible number of copies of the system.
- Can be combined with other AWH reaction coordinates.

Audience Q&A session

- Please use the Questions function in GoToWebinar application
  - If you *don't have audio*, please mention that in the question.

- Any other questions or points to discuss after the live webinar? Join the discussions at http://ask.bioexcel.eu.



BioExcel Summer School

04 Jun, 2021 – 11 Jun, 2021 - online

See https://bioexcel.eu/