Overview of the SAMPLing challenge results

Andrea Rizzi

D3R 2018 Workshop

La Jolla, 2/22/2018

The SAMPLing challenge moves the focus from accuracy to convergence properties of statistical methods

	Target	Reference	Cost
Host-guest	Model + method	Experiments	/
SAMPLing	Method	Converged free energy calculation	Computational cost

Main questions that the SAMPLing challenge attempts to answer

- Do different methods converge to the same answer?
- How quickly do calculations converge?

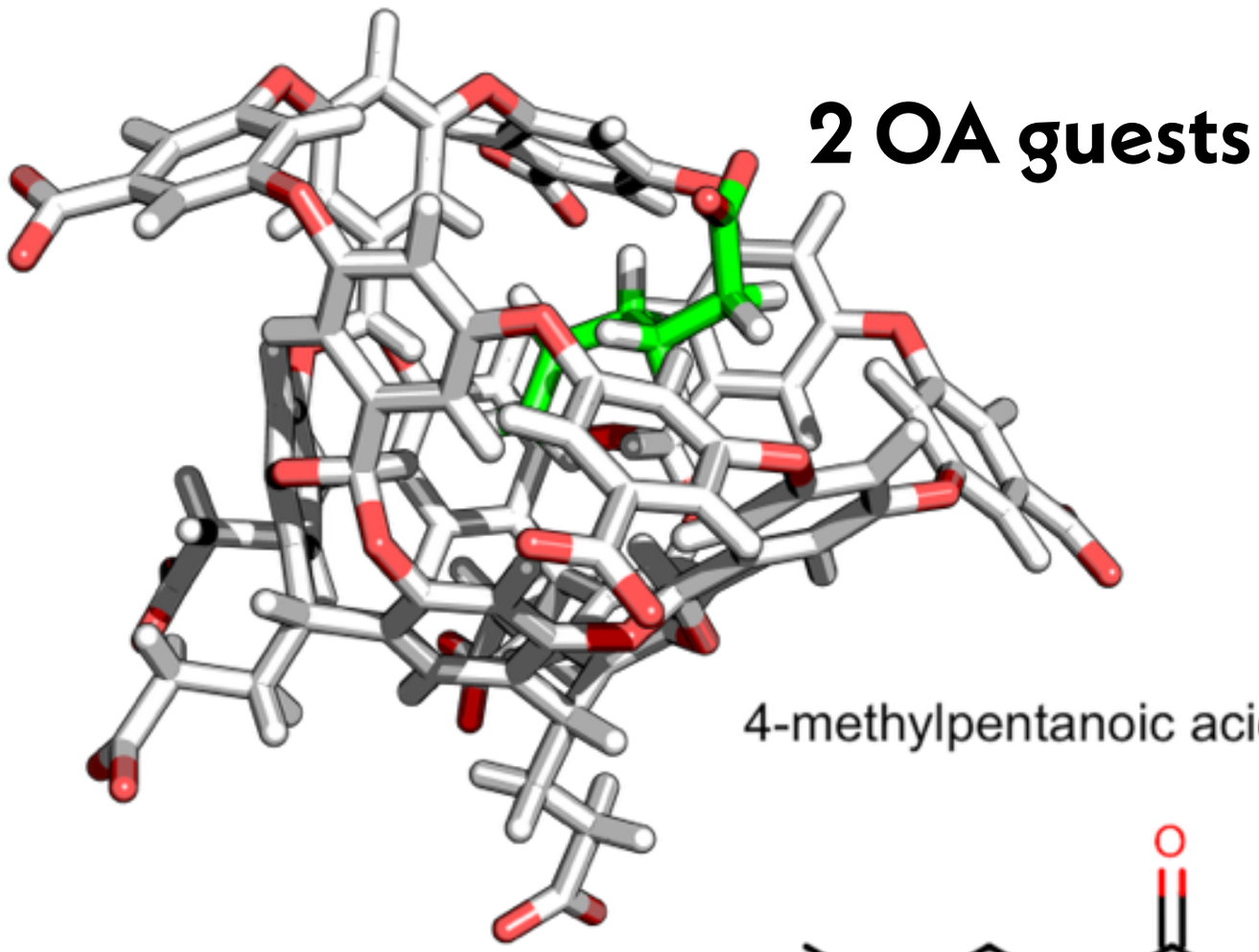
Outline

- Challenge description
- Reference calculations
- Overview of results

Measuring the "cost" of a *method* is not trivial

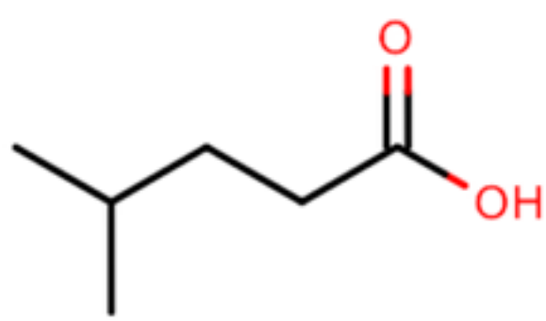
- **CPU/GPU time** affected by hardware/implementation
- Total ns/μs of MD depends on time step and lose meaning with MC
- **Number of energy evaluations**
  - Depend on system size (solvent complex)
  - Energy can often be just updated (MC on a subset of atoms, multiple time scale MD, switch Hamiltonian)

Few targets selected from host-guest challenge mixing fragment-like and drug-like molecules

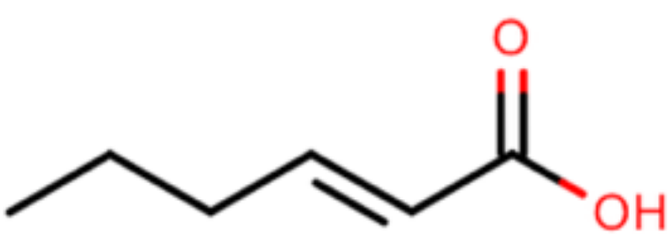


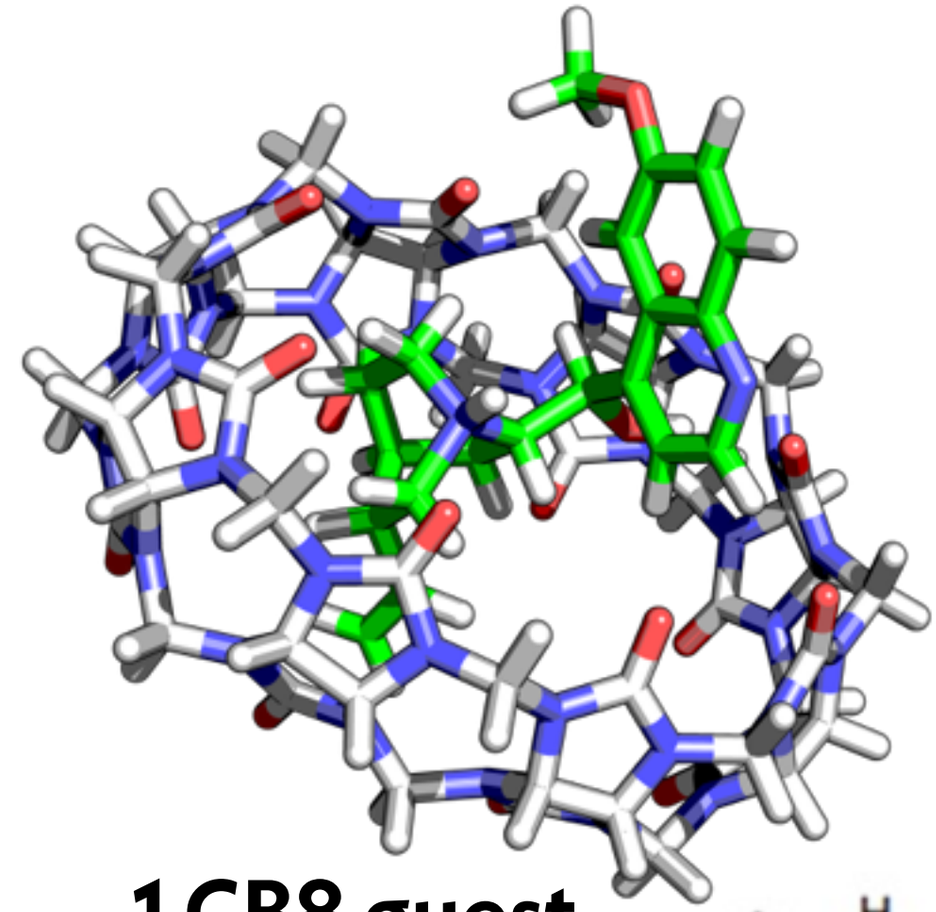
2 OA guests

4-methylpentanoic acid

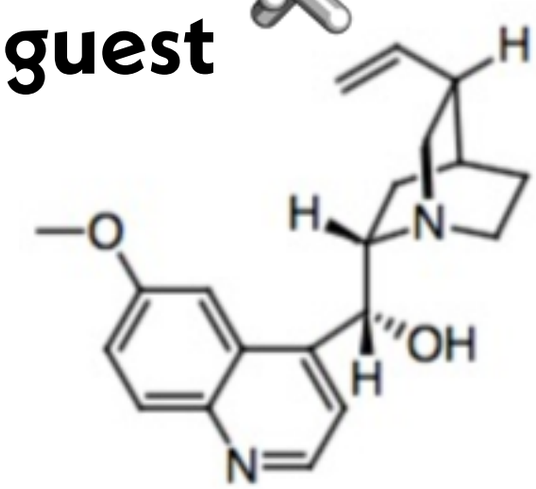


trans-2-hexenoic acid



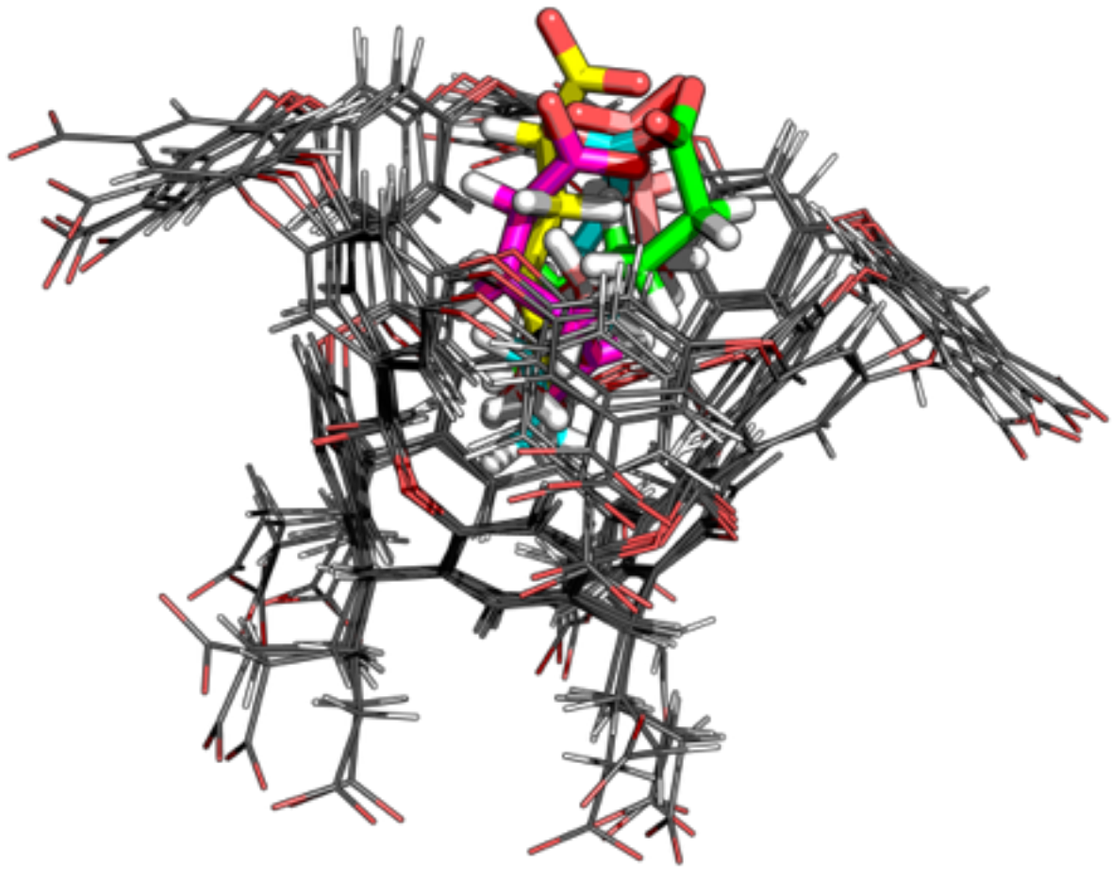


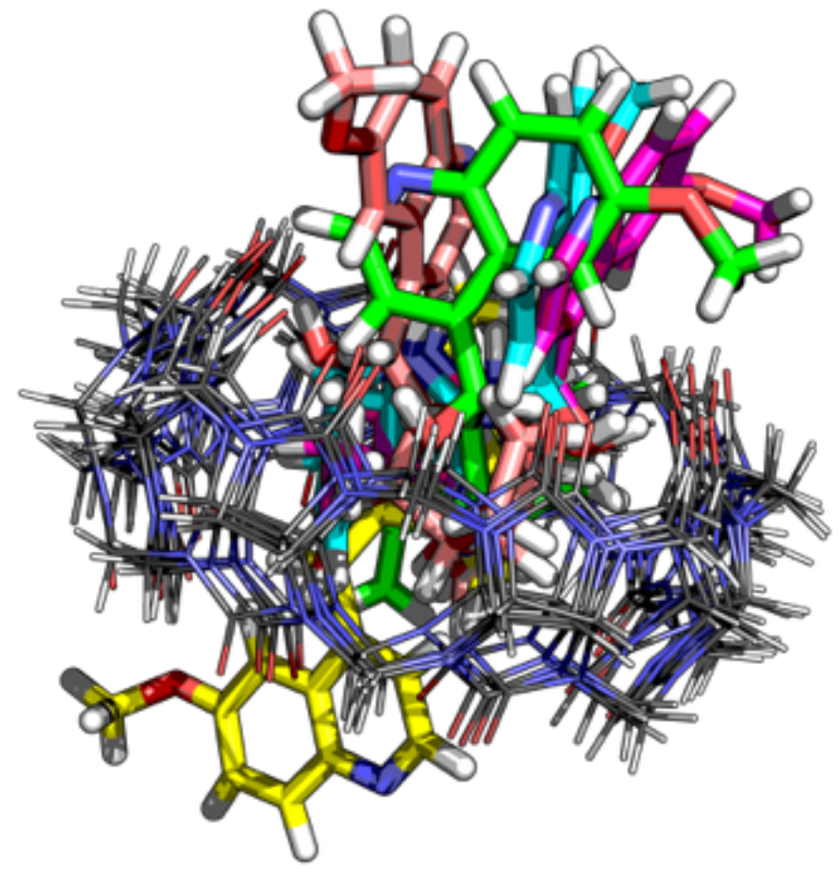
1 CB8 guest



Quinine
CAS 130-95-0
OakWood Chemical

Initial configurations and parameters are shared among all participants



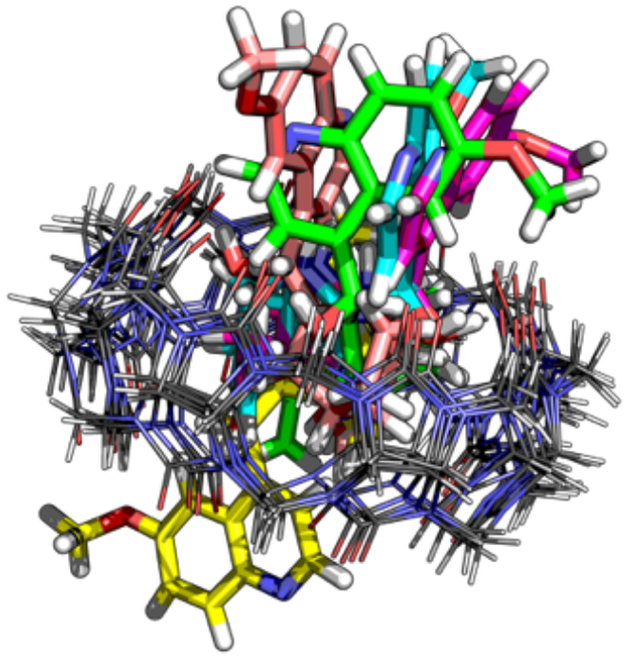


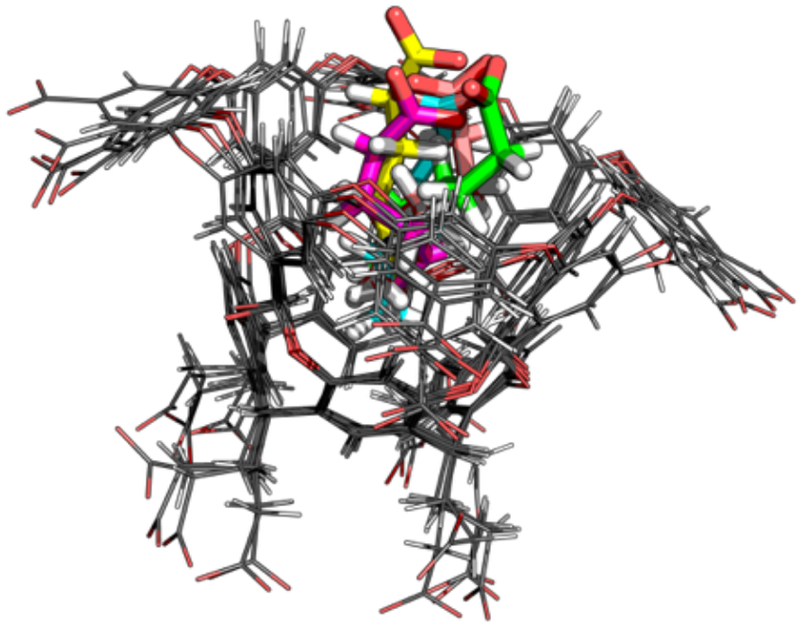
- A “common” setup: AM1-BCC charges / GAFF / TIP3P cubic water box + neutralizing and buffer Cl- Na+ ions

- Files converted in many formats (Michael Shirts)

Long-range treatment parameters can only be suggested

- Long-range treatment is not encoded in parameter file (although important part of the model)
- Reference calculations
  - PME
  - Cutoff electrostatics and VdW at 10Å
  - Switching function for VdW at 9Å





Free energy estimates of replicates at multiple time points allow us to observe the estimate variance in time

- Free energy estimates after 1, 2, ..., 100% of calculation
- Total cost in energy evaluations and time
- Replicates that belong to the same system must have same total cost

Outline

- Challenge description
- Reference calculations
- Overview of results

YANK: A GPU-accelerated Python platform for absolute free energy calculations

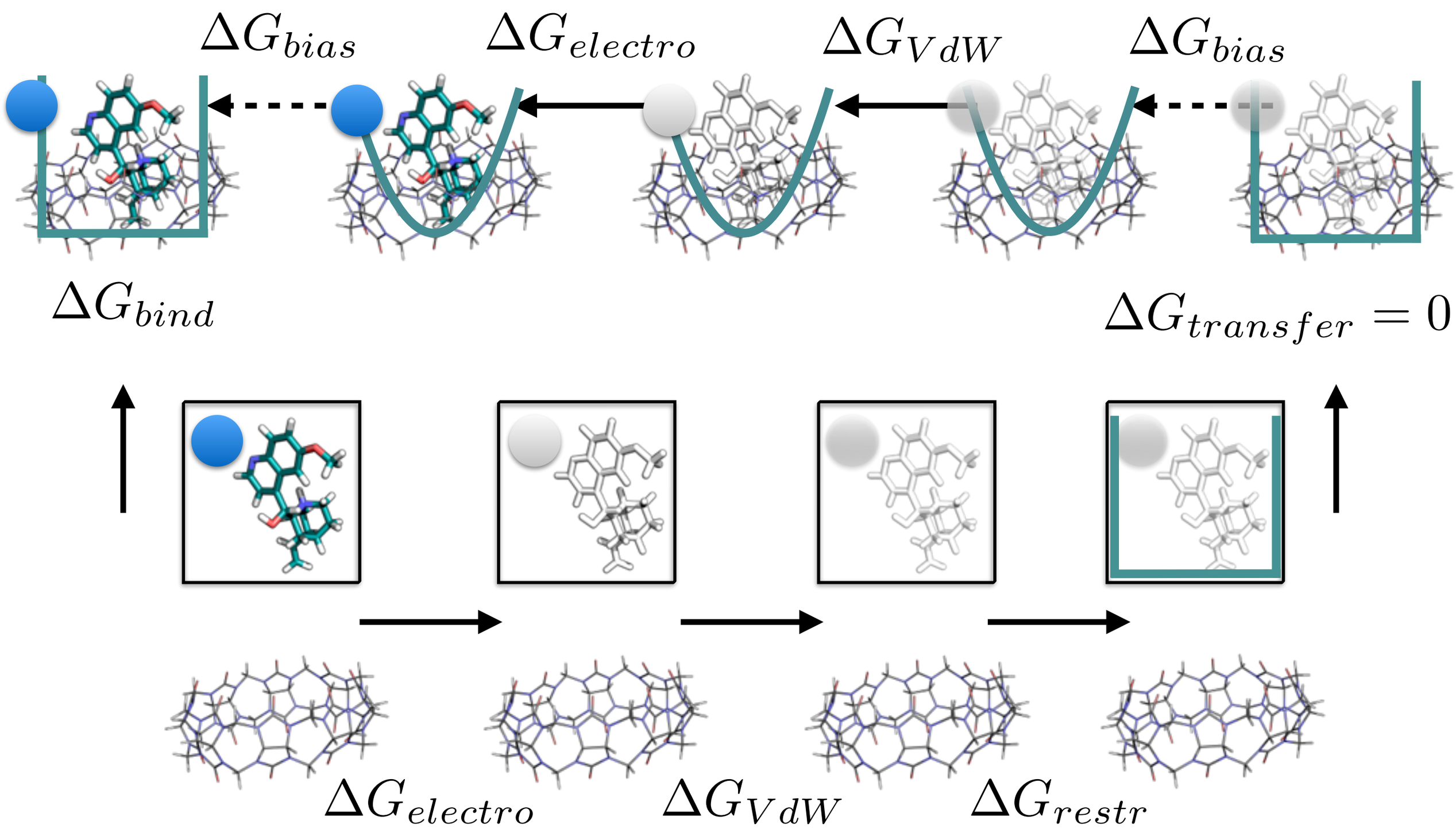
Docs: http://getyank.org/latest/

Built on:

- OpenMMTools
- OpenMM
- AmberTools
- Parmed
- MDTraj
- ...

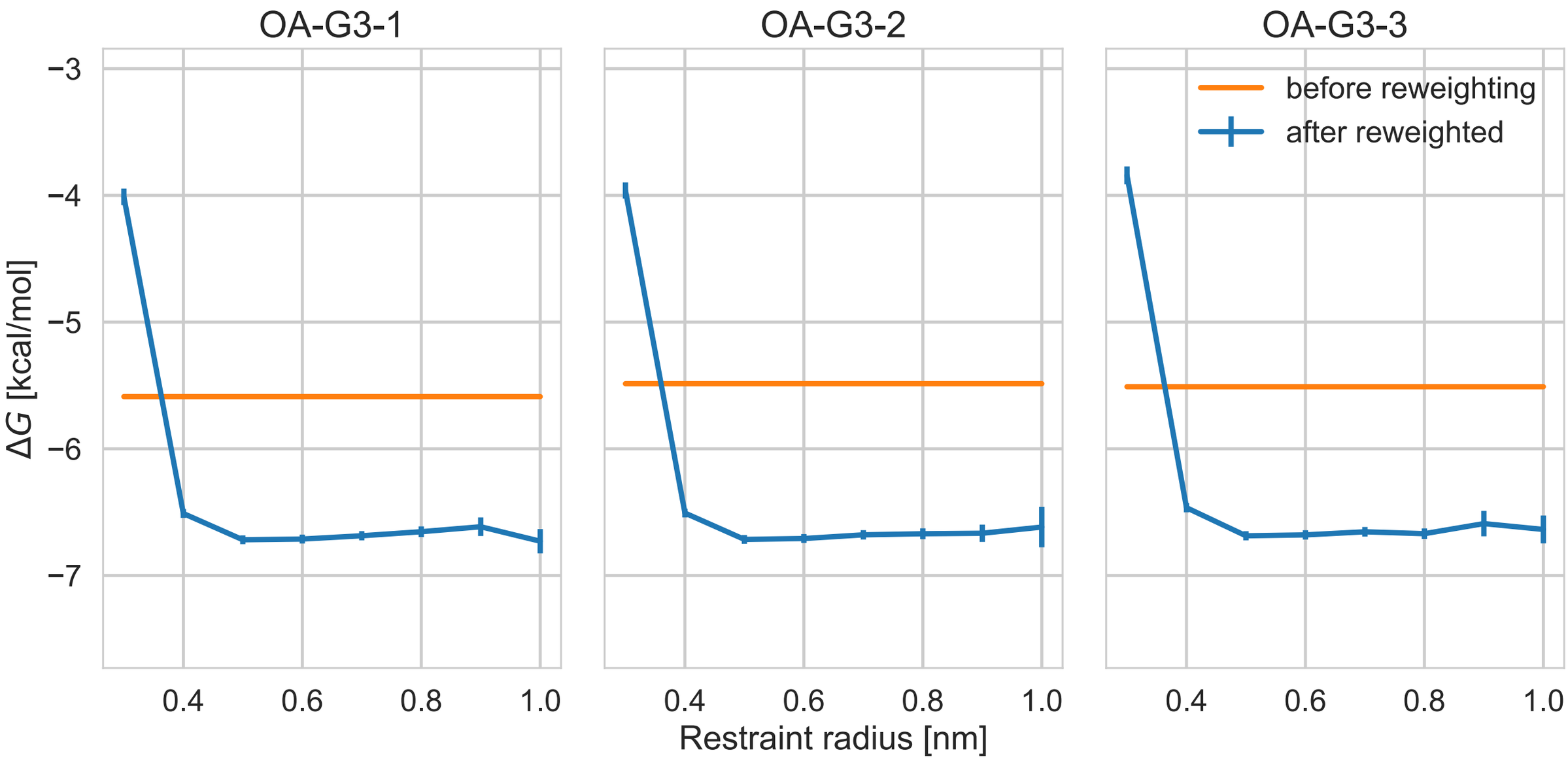
Levi Naden

Thermodynamic cycle



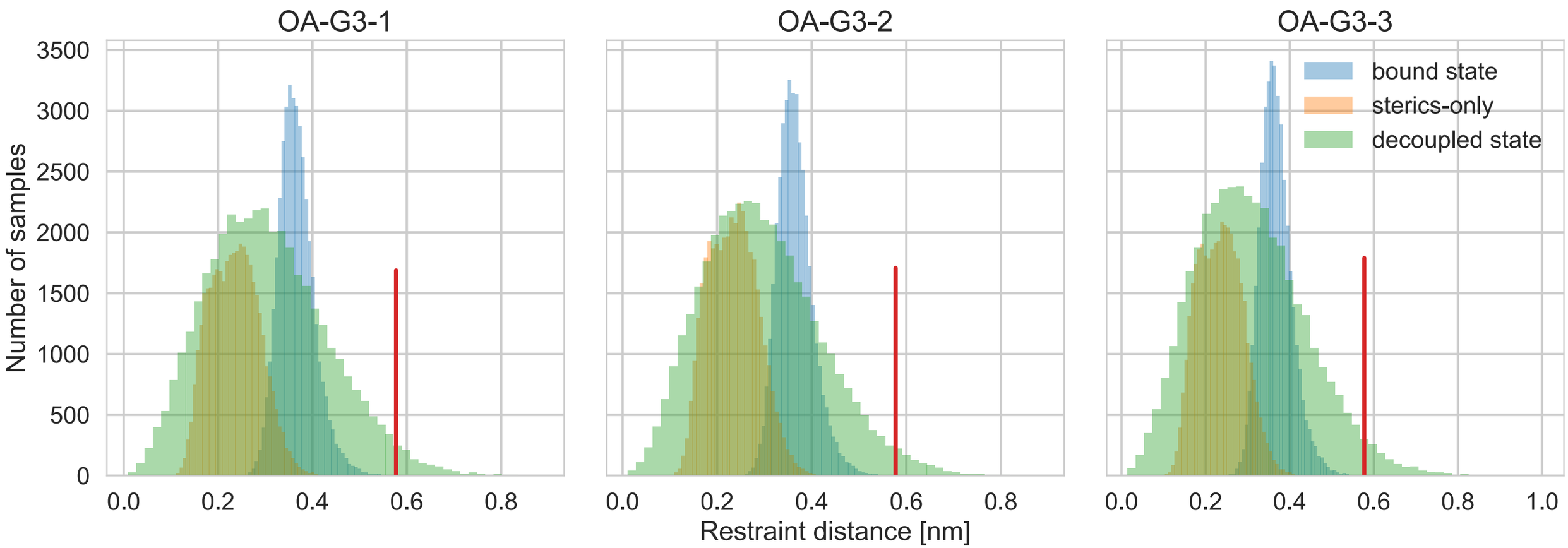
Gilson MK, Given JA, Bush BL, McCammon JA. *Biophysical Journal*. 1997 Mar 72:1047-69.

The harmonic restraint introduces a non-negligible bias



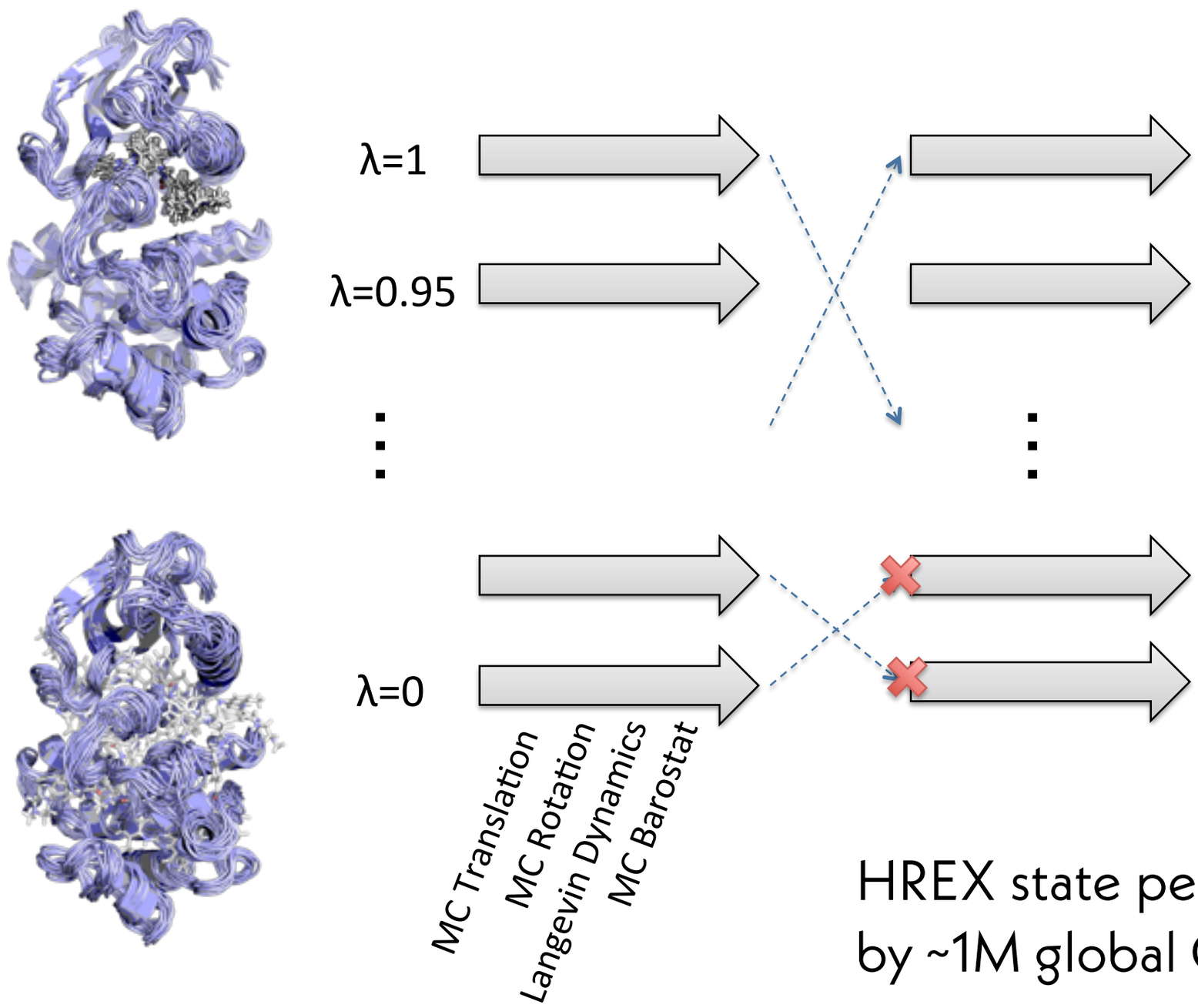
Restraint spring constant ~0.17 kcal/(mol Å^2)

Let the simulation define the binding site and integration volume



Radius of the square well distance determined as 99.99-percentile of bound state harmonic restraint distance distribution.

We mix HREX, MD and Monte Carlo moves to decrease correlation times



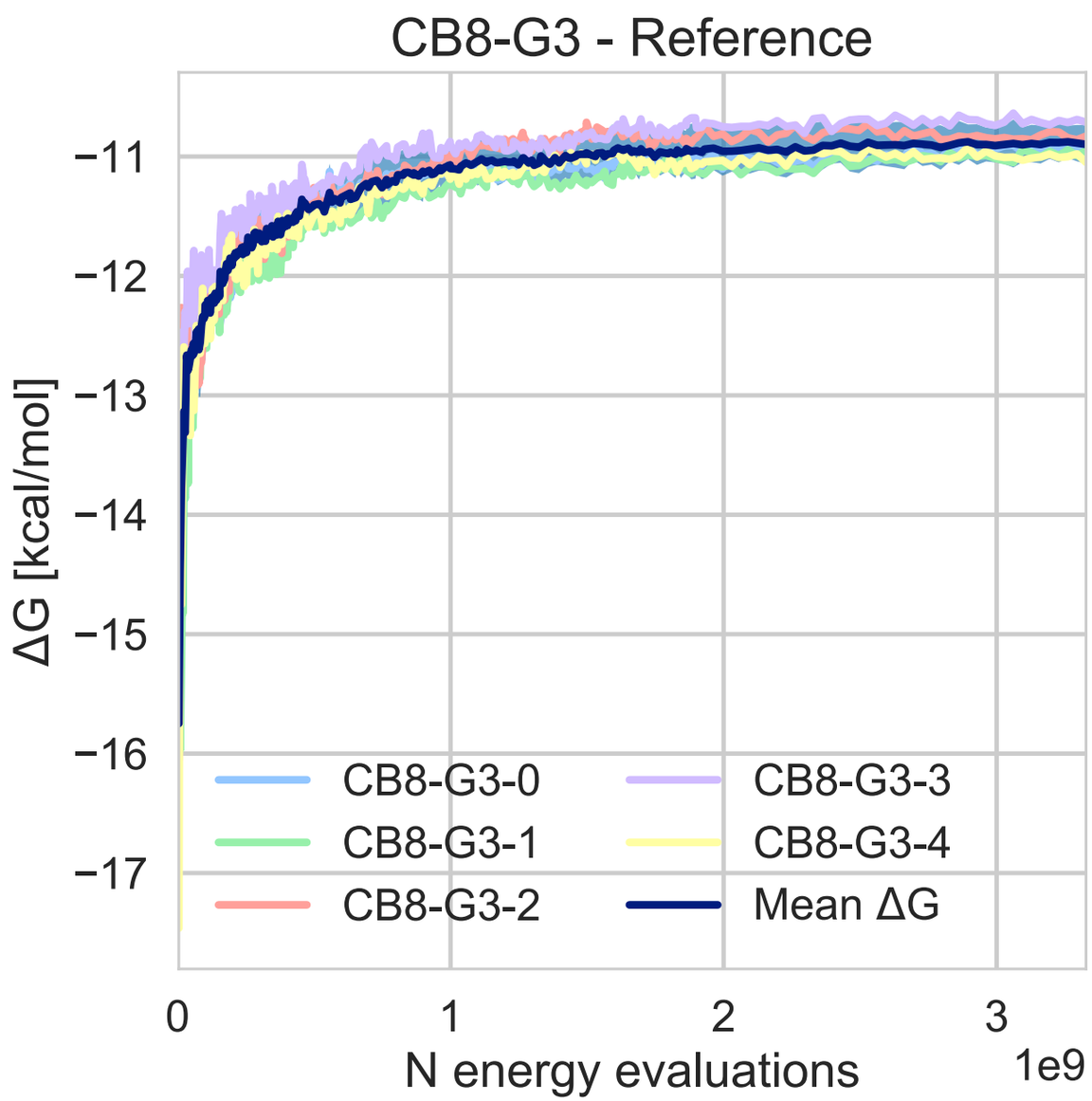
HREX state permutation obtained by ~1M global Gibbs swaps.

Chodera JD, Shirts MR. The Journal of chemical physics. 2011 Nov 21;135(19):194110.

Outline

- Challenge description
- Reference calculations
- Overview of results

The 5 replicates converged to the same value

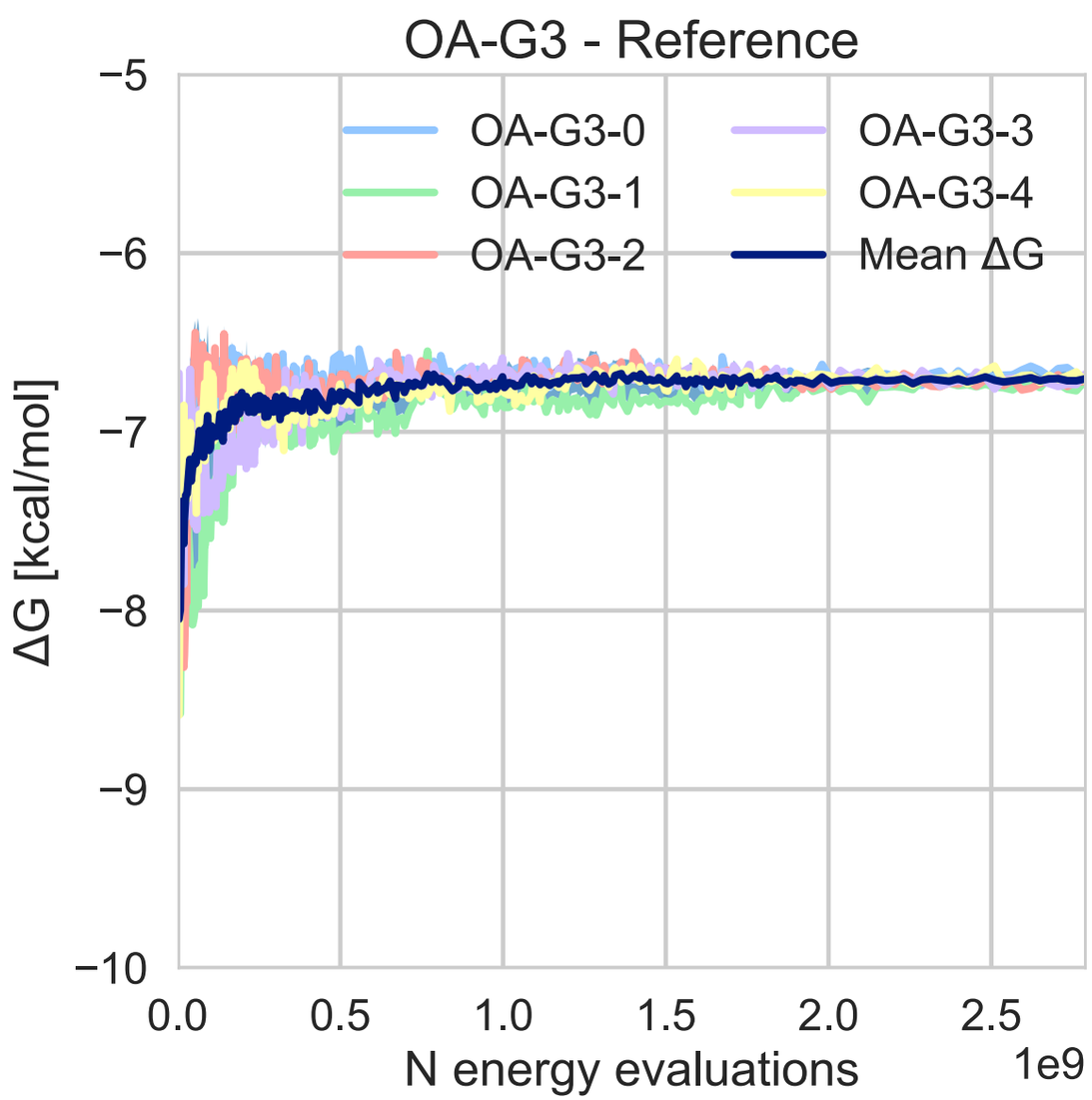


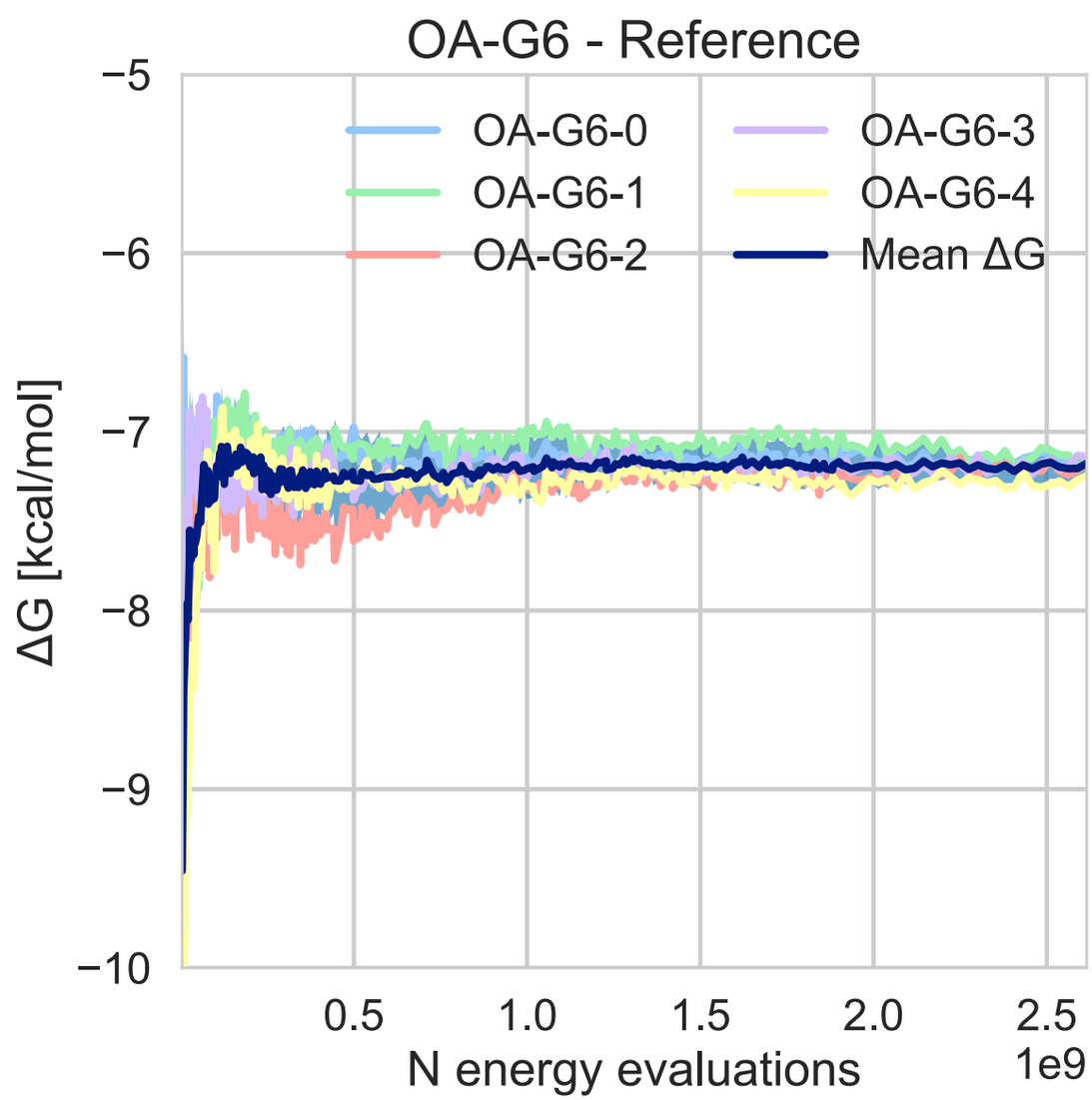
$$\Delta G_{mean} = -10.9 \pm 0.1 \ kcal/mol$$

(t-based 95% confidence interval)

N energy evaluations include calculation of (#states × #states) energy matrix for MBAR and Gibbs sampling.

The 5 replicates converged to the same value

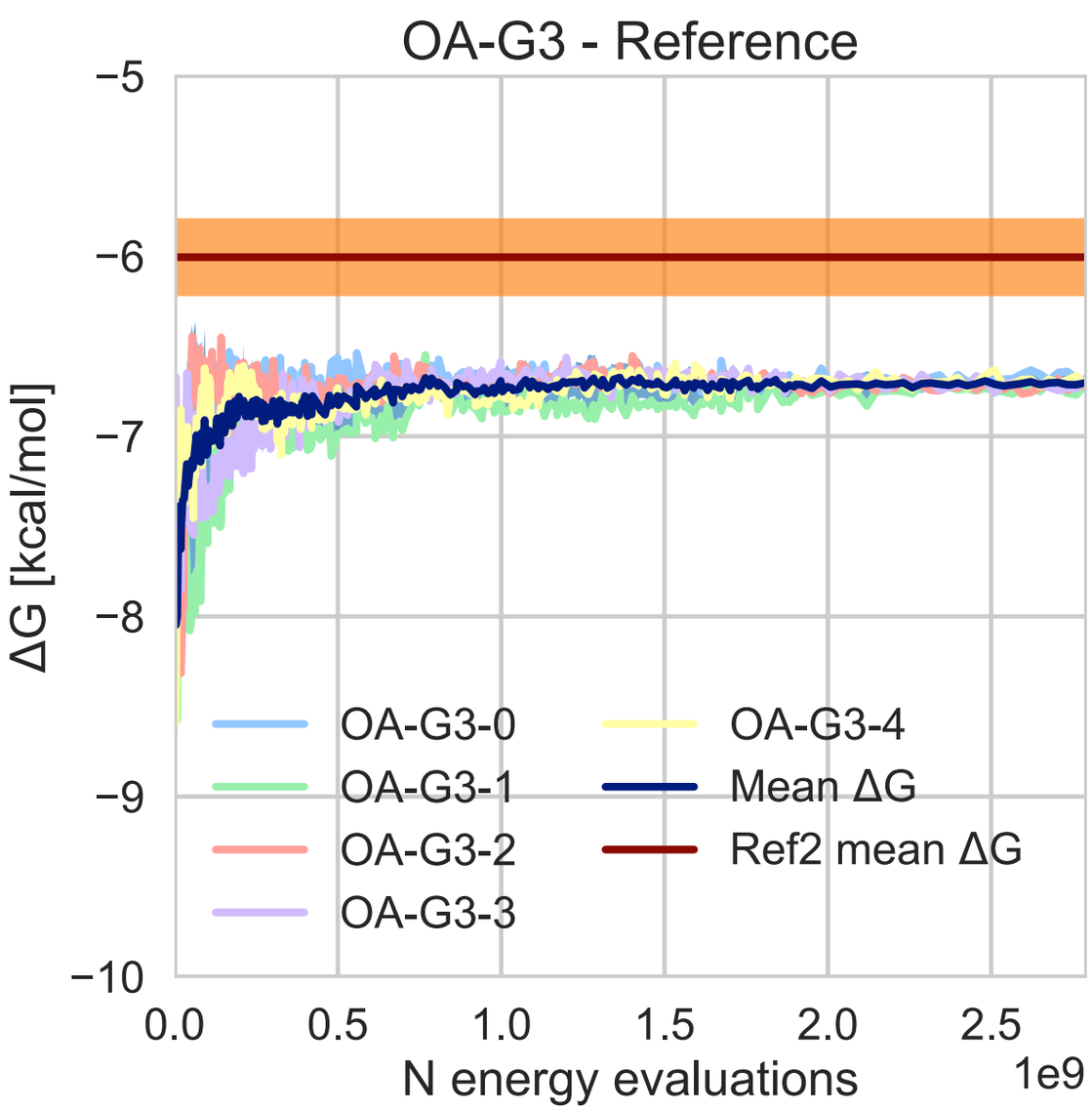


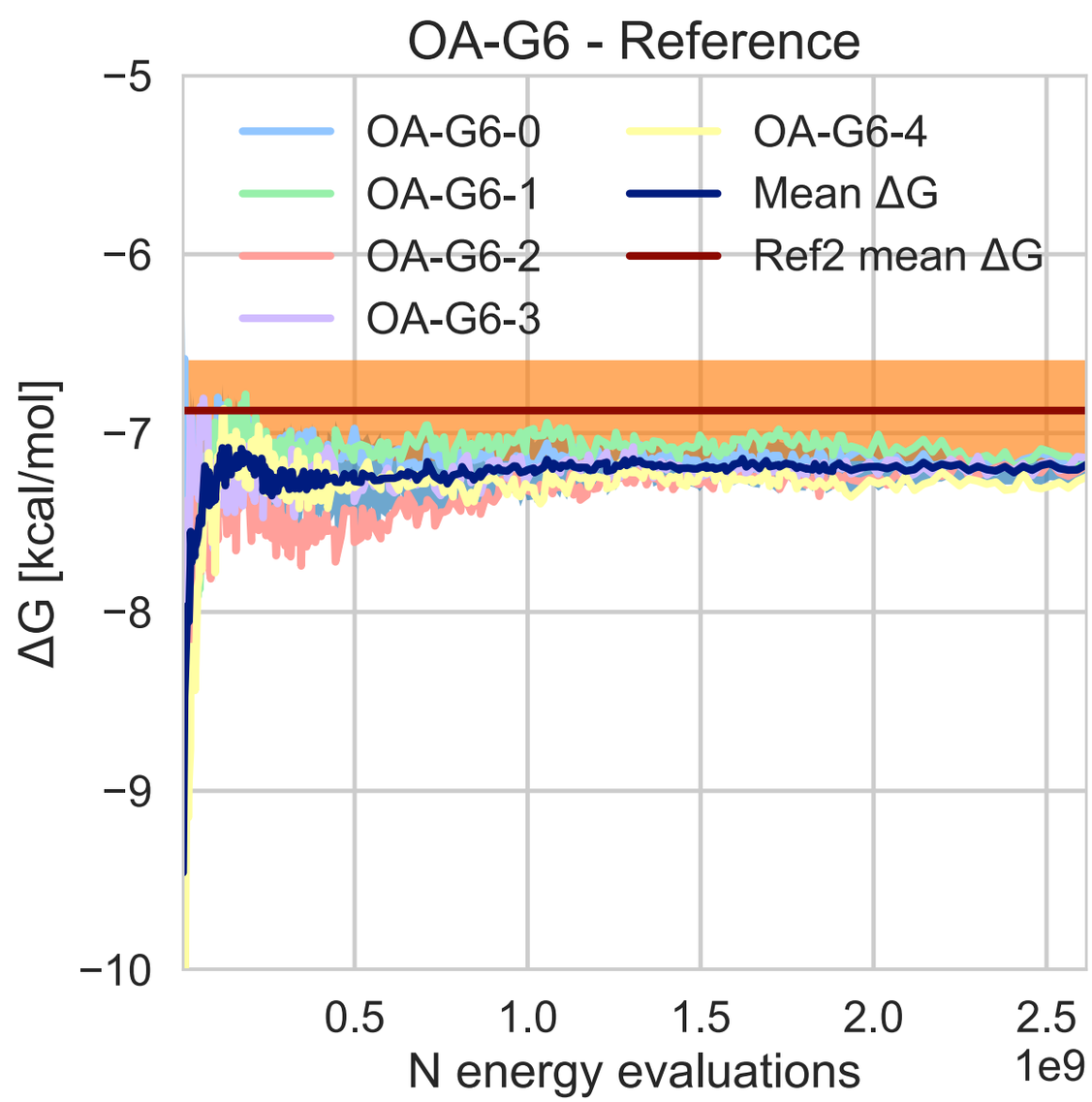


$$\Delta G_{mean} = -6.70 \pm 0.02 \ kcal/mol$$

$$\Delta G_{mean} = -7.17 \pm 0.05 \ kcal/mol$$

There are discrepancies between the two sets of reference calculations on the order of 1kcal/mol





$$\Delta G_{mean} = -6.70 \pm 0.02\ kcal/mol$$

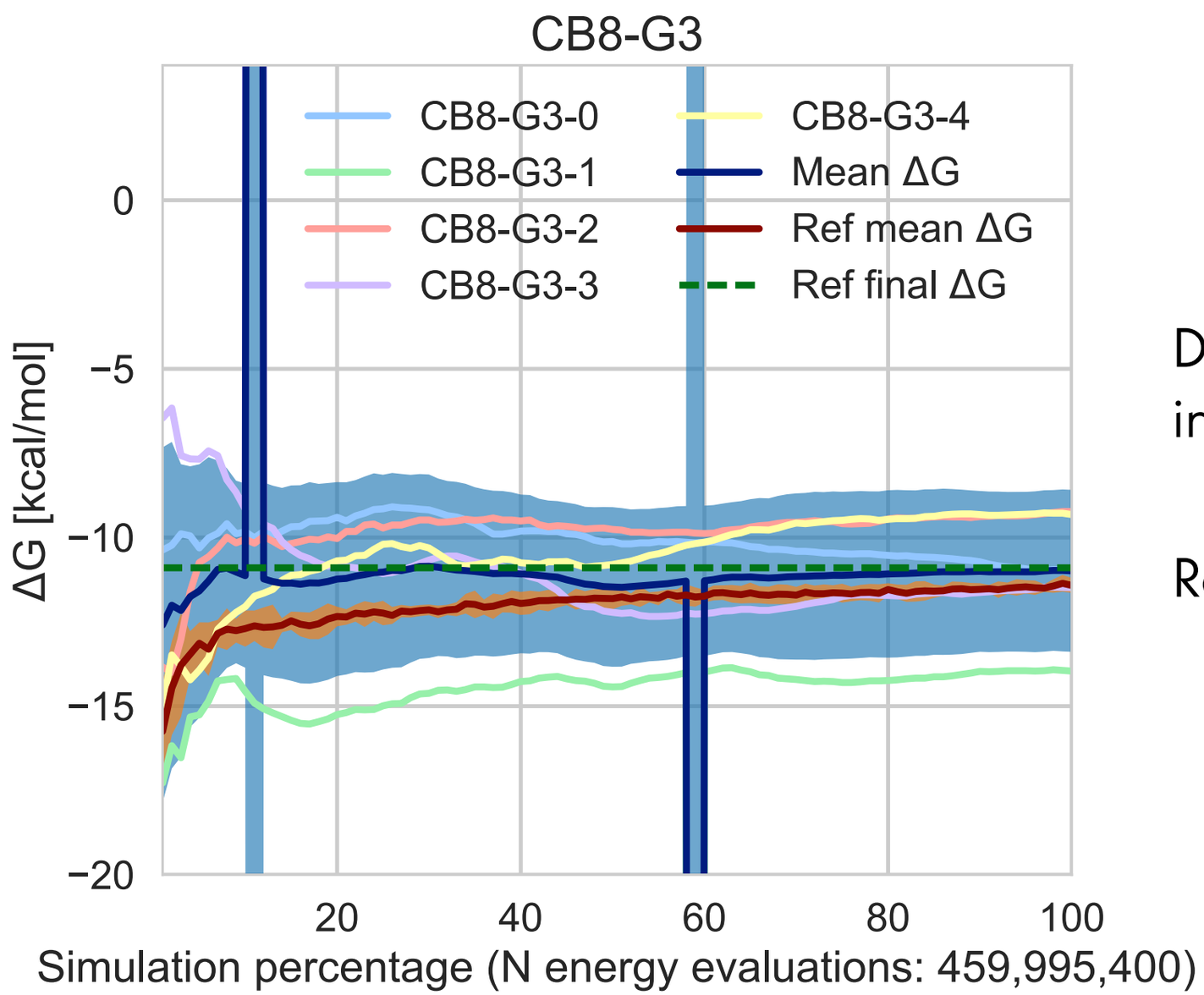
$$\Delta G_{mean\ ref2} = -6.0 \pm 0.2\ kcal/mol$$

$$\Delta G_{mean} = -7.17 \pm 0.05\ kcal/mol$$

$$\Delta G_{mean\ ref2} = -6.9 \pm 0.2\ kcal/mol$$

Travis Jensen, Michael Shirts

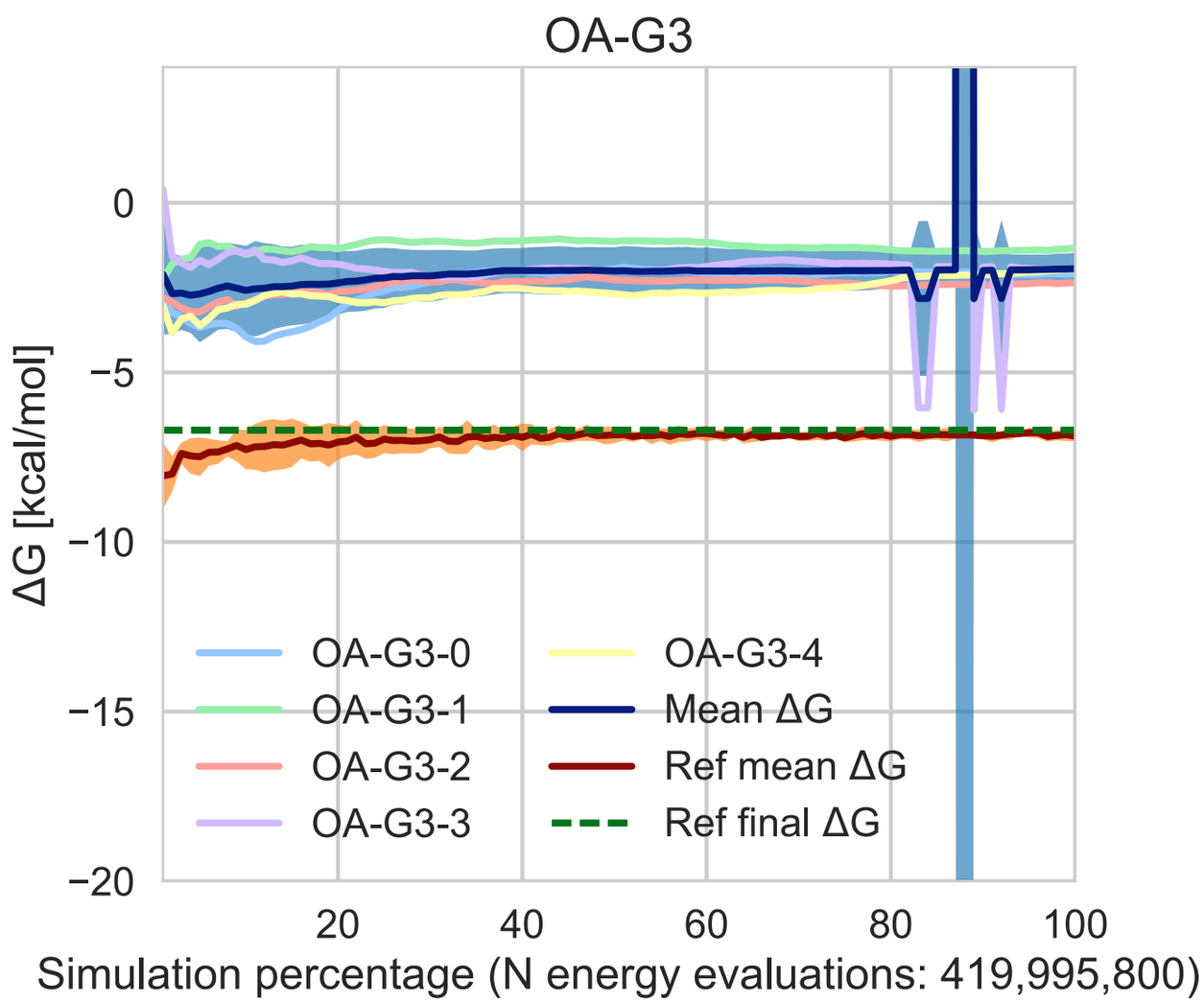
Despite starting from same input files, getting different methods to agree is non-trivial

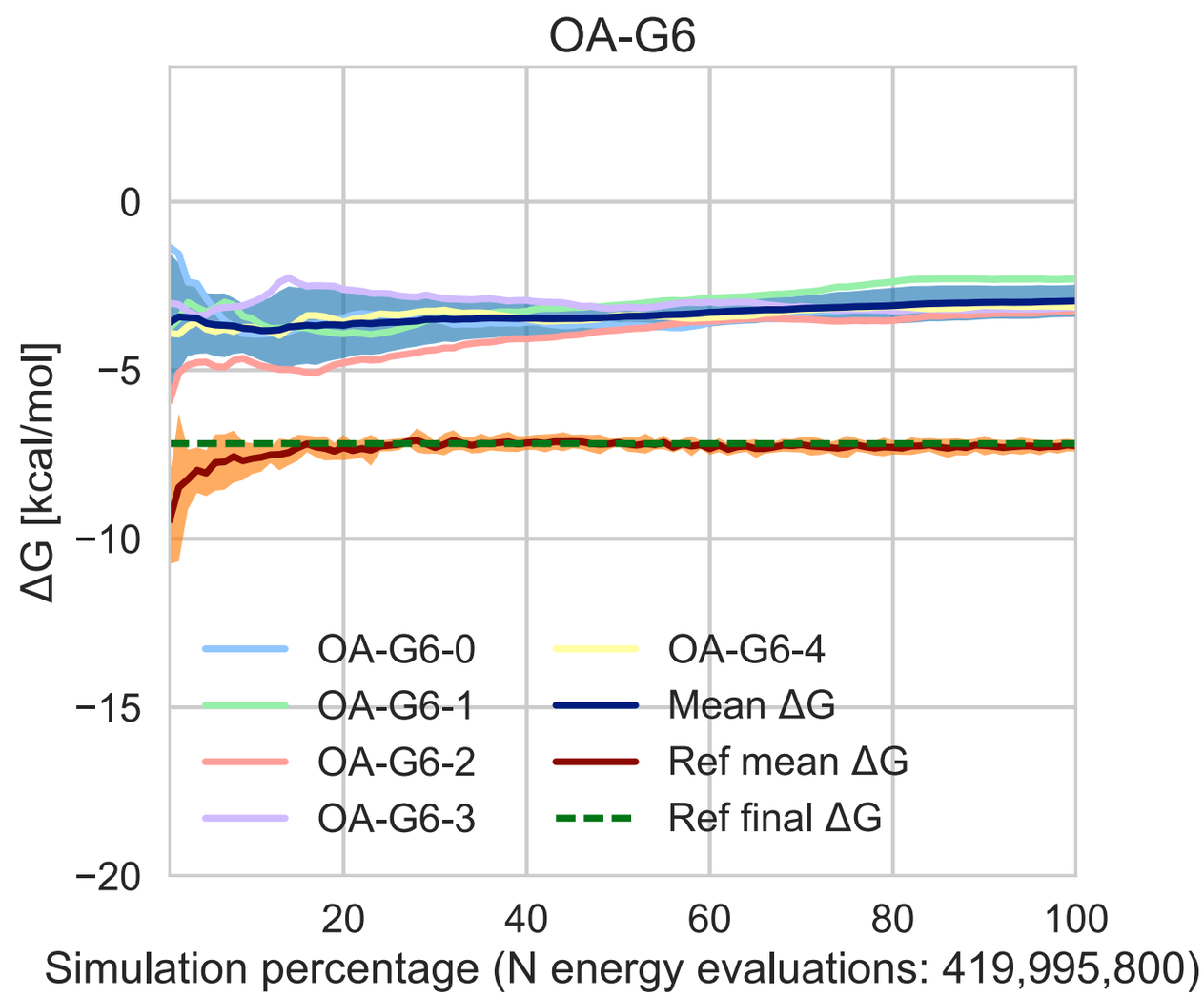


Double annihilation method, independent replicas, 20ns/replica

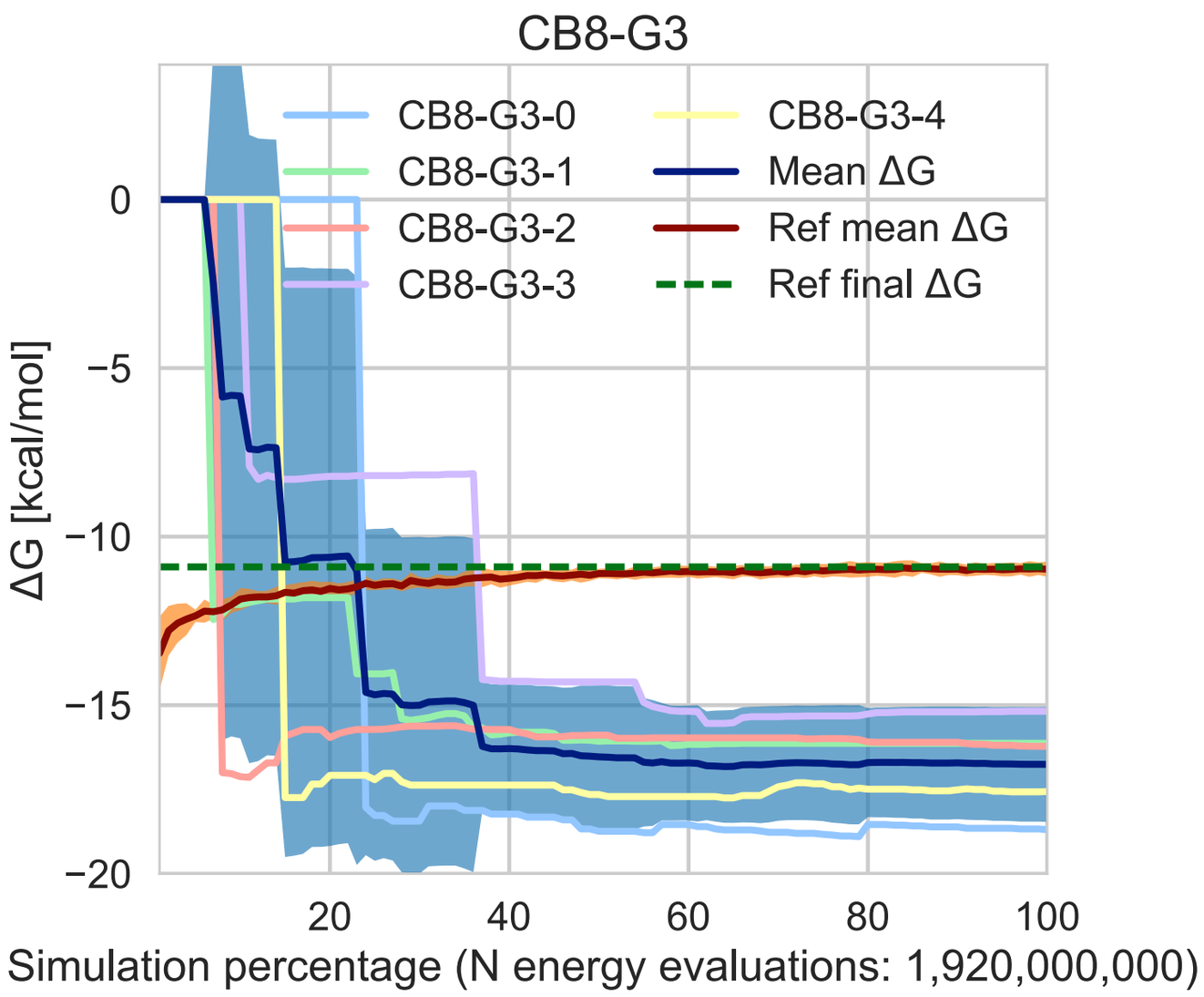
Reaction Field and 12Å cutoff

Despite starting from same input files, getting different methods to agree is non-trivial



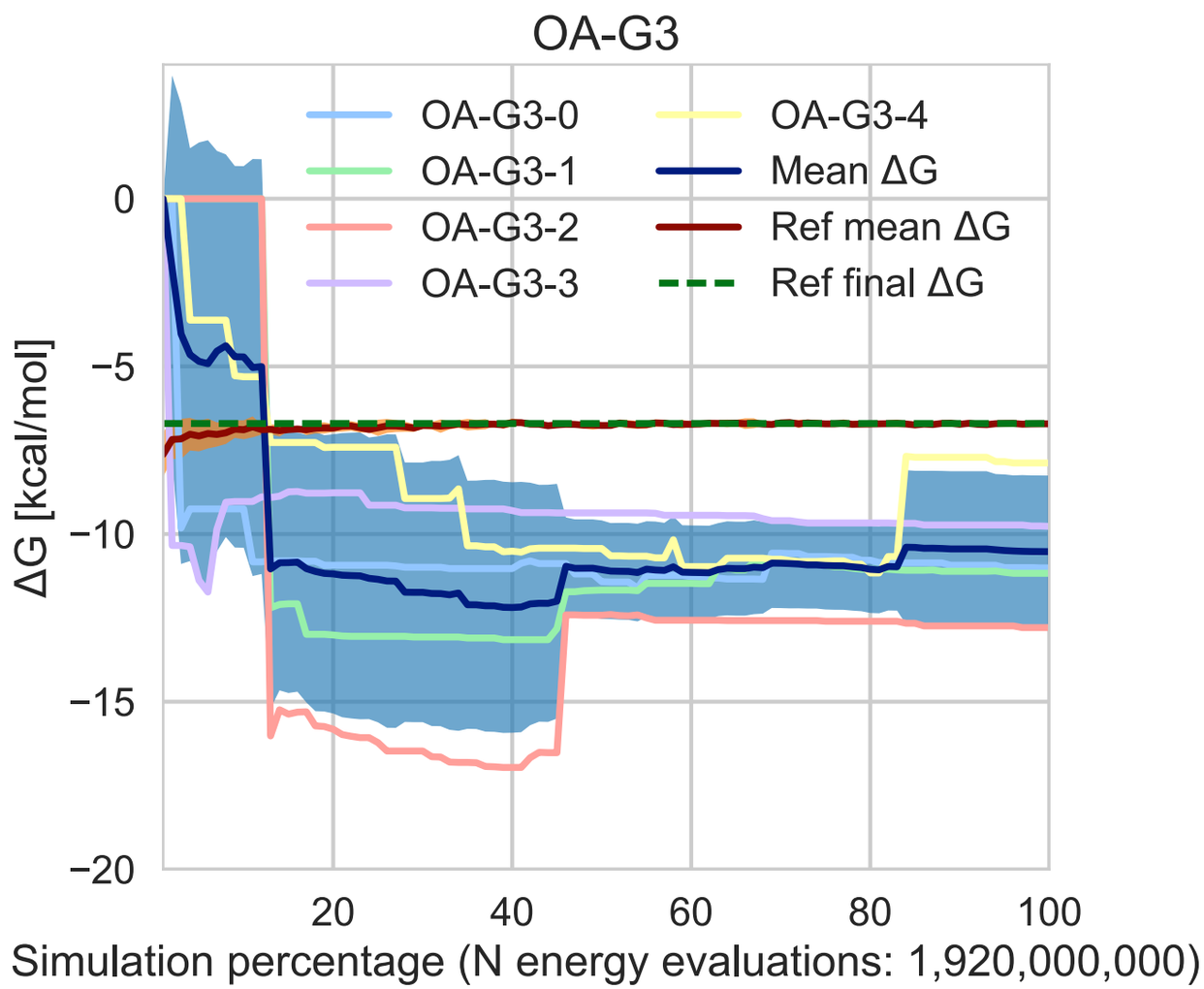


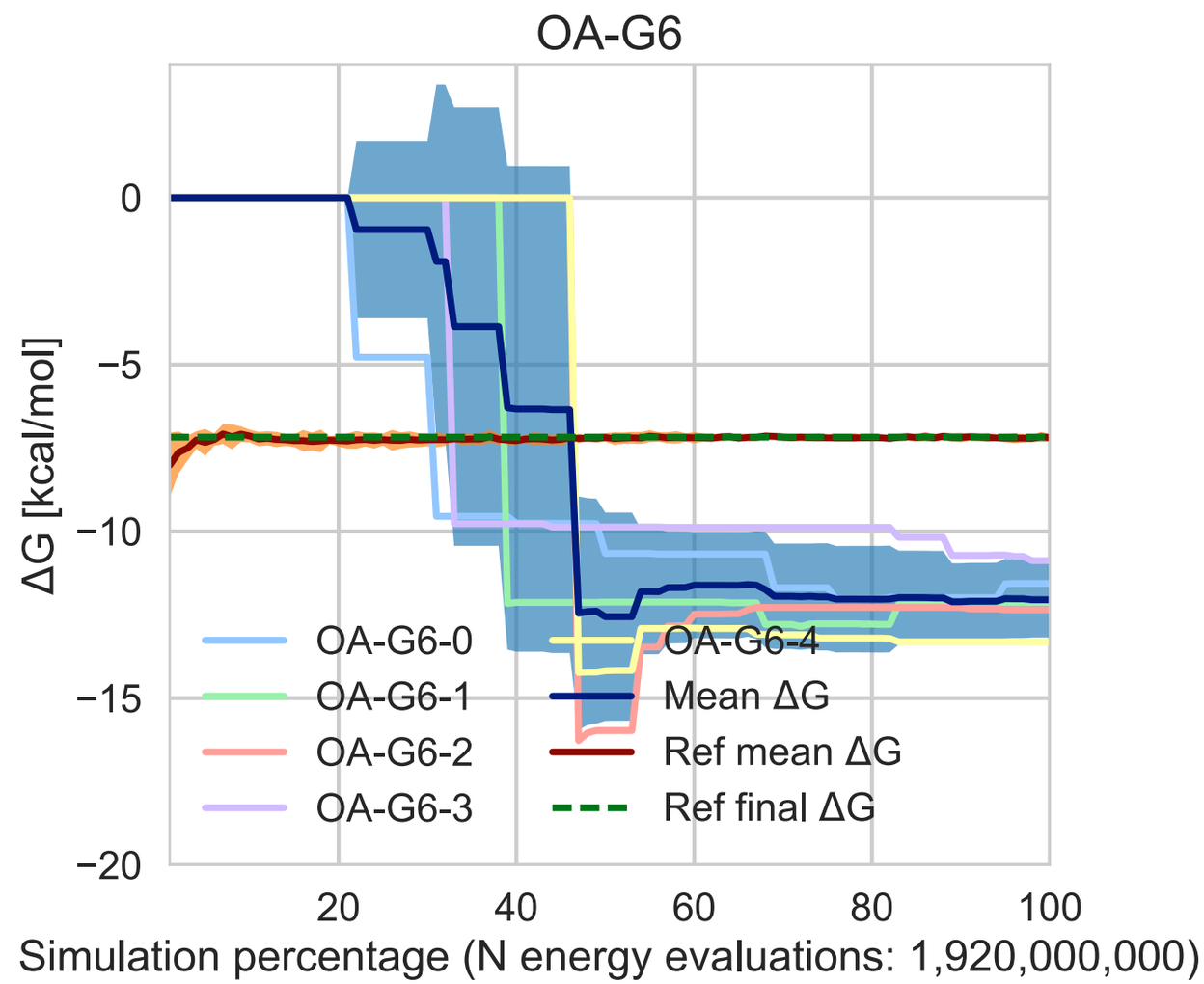
Despite starting from same input files, getting different methods to agree is non-trivial



Estimate free energy from k_{on}/k_{off}

Despite starting from same input files, getting different methods to agree is non-trivial





Next steps

- Get reference calculations to agree
  - Restraint handling
  - Different barostat and PME parameters
  - Missing conformational space (water binding)
- Understand the causes of the discrepancies for the submissions

Conclusions and prospects

- Very different results even when starting from same input files!
  - It might be easier to zero-in on methodological issues.
- Lower the barrier to participate in this type of study. We'd love to hear your feedback!

Acknowledgements

Participants

Julien Michel (Edinburgh)

Stefano Bosisio (Edinburgh)

Michail Papadourakis (Edinburgh)

Alex Dickson (MSU)

Reference calculations

Travis Jensen (CU Boulder)

Michael Shirts (CU Boulder)

Chodera lab

John Chodera

Mehtap Isik

Ariën Sebastiaan (Bas) Rustenburg

Levi Naden

D3R/SAMPL6 Organizers

David Mobley (UCI)

John Chodera (MSKCC)

Michael Shirts (CU Boulder)

Michael Chiu (UCSD)

Michael Gilson (UCSD)

Rommie Amaro (UCSD)

Tri-I CBM PhD Program

David Christini

Christina Leslie

Kathleen Pickering

Margie Hinonangan-Mendoza