Efficient Automated Adaptive Computation of Protein-Ligand Absolute Binding Free Energies

Finlay Clark[1]; Graeme Robb[2]; Daniel Cole[3]; Julien Michel[1]

[1]*EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom*

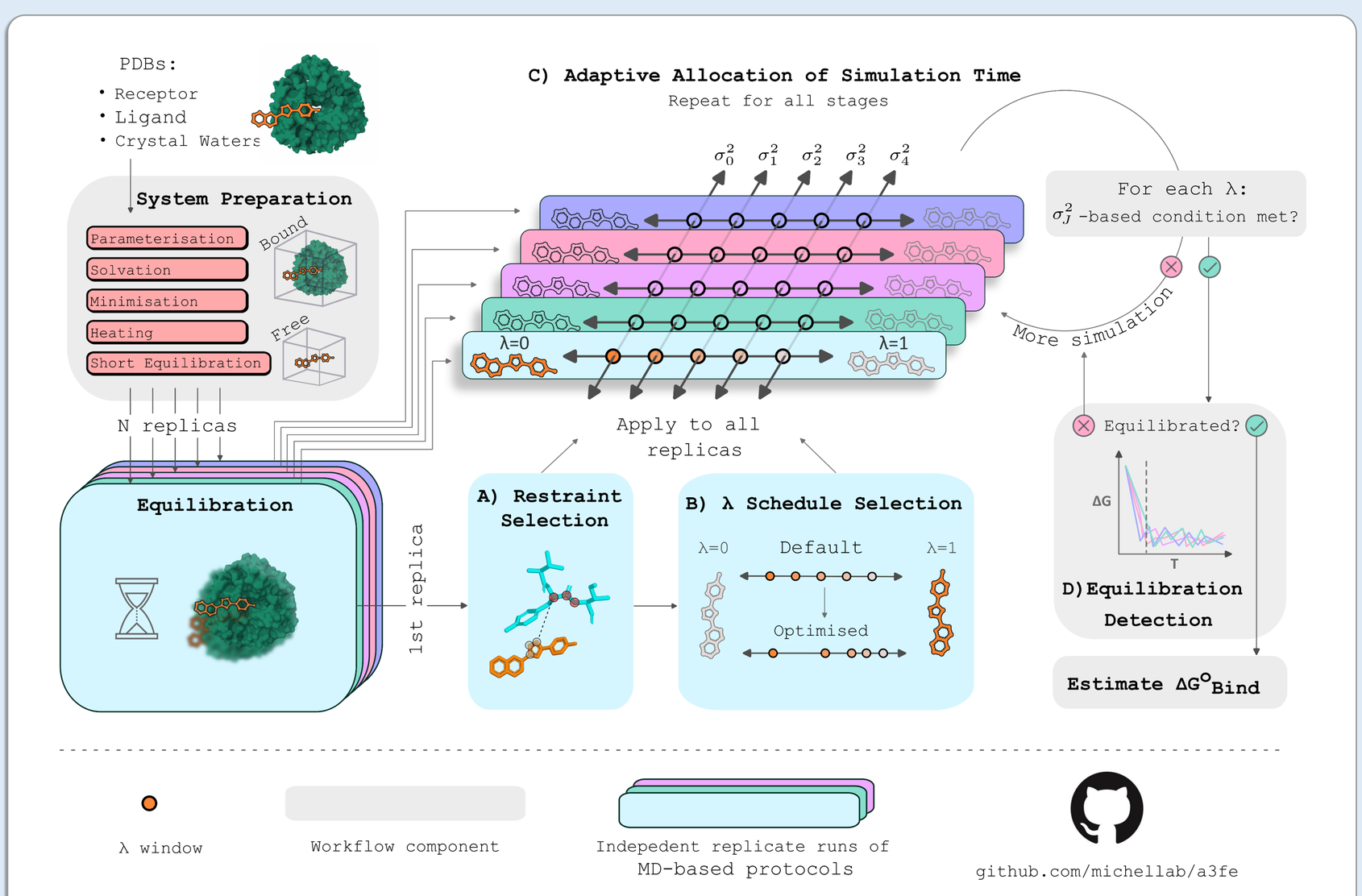
[2]*Oncology R&D, AstraZeneca, Cambridge CB4 0WG, United Kingdom*

[3]*School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom*

Corresponding author: ***finlay.clark@ed.ac.uk***

Alchemical absolute binding free energy calculations must become more efficient

- We implement and test an **open-source fully-automated adaptive workflow for efficient ABFE calculations**
- Based on BioSimSpace (system preparation) and SOMD (Sire/ OpenMM, free energy calculations)



A) Restraints are selected to mimic strong receptor-ligand interactions

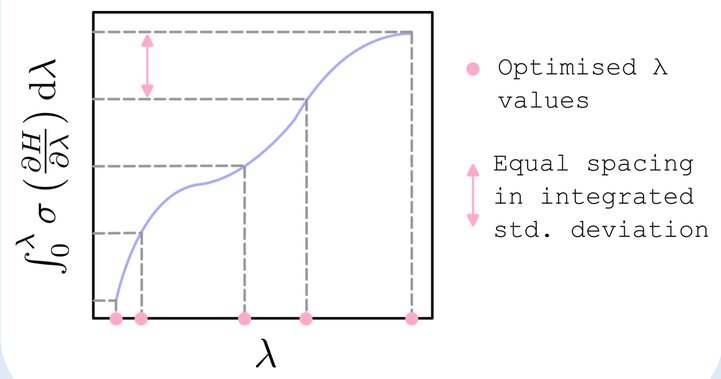
Algorithm

- Candidate sets of anchor points selected from simulation of receptor-ligand complex[1]
- Force constants fit to fluctuations observed
- Pick the stable restraints[2] which most strongly restrict the configurational volume accessible to the ligand once decoupled[3]

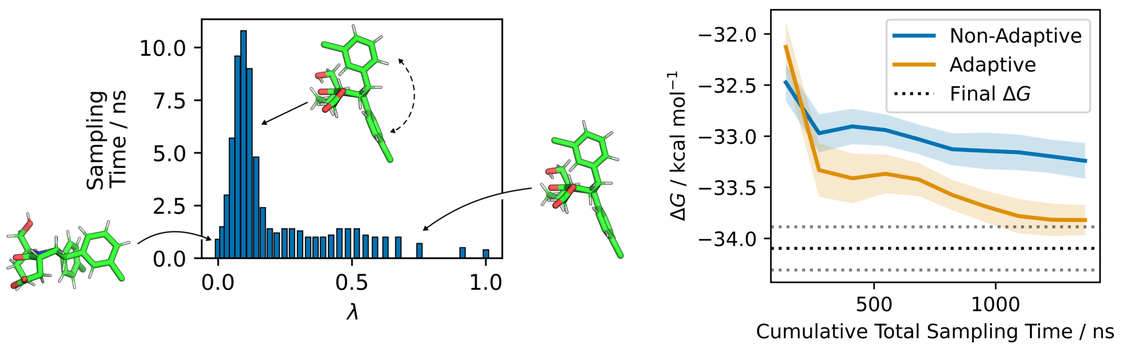
Ensures only two windows required to introduce restraints and free energy cost of introduction is always ~ 1.2 kcal mol^{-1}

B) Lambda windows are spaced based on the standard deviation of the free energy change[4,5]

Standard deviation of $\partial H / \partial \lambda$ as a function of λ estimated from very short initial simulations



C) Sampling time is allocated to minimise inter-replicate uncertainty



1) Simulation time is allocated according to:

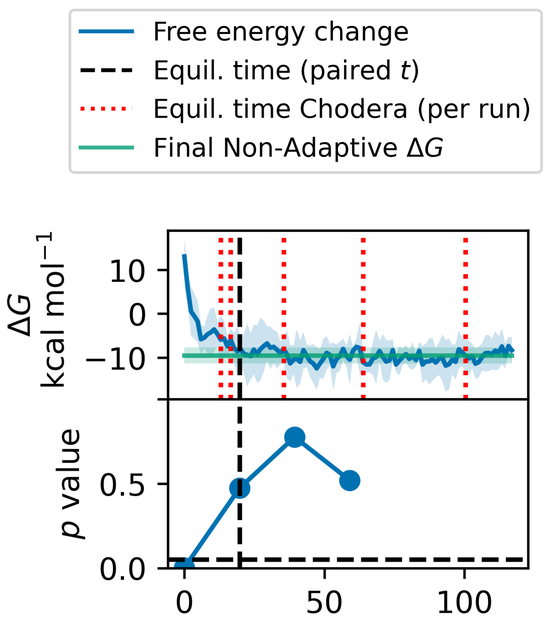
$$t_{\mathrm{Optimal},\lambda} = \sqrt{\frac{t_{\mathrm{Current},\lambda}}{C}}\,\sigma(\Delta G)_{\mathrm{Current},\lambda}$$

Predicted optimal simulation time — Current simulation time — Pre-specified constant — Current uncertainty of per-λ ΔG

2) This can accelerate convergence

Accelerates convergence if the sampling issues occur on a timescale comparable to the simulation duration

D) Equilibration is detected based on an ensemble of replicates



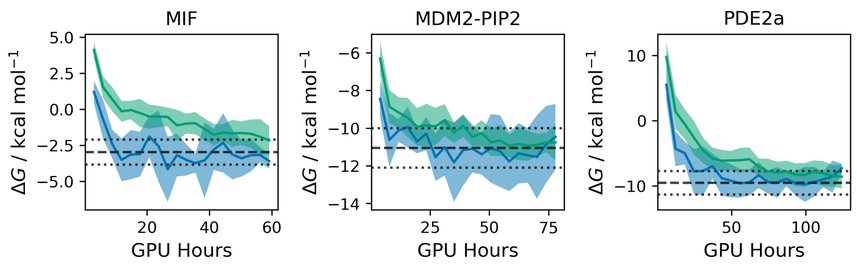
Algorithm

- For each replicate run:
  - Calculate free energy change using first 10 % and last 50 % of data
- Perform a paired *t*-test on the paired differences between the first and last portions of data
- If $p < 0.05$:
  - Not equilibrated. Discard data from start of runs and repeat
- Else:
  - Assume equilibrated

Paired nature of test increases sensitivity to ensemble trend by ignoring systematic differences between replicas

Overall protocol accelerates equilibration

Free energy changes shown for the Lennard-Jones term removal stage, which dominates overall cost and equilibration:

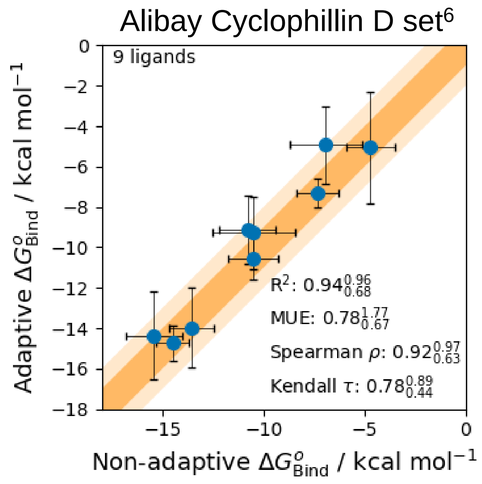


Adaptive protocol:

- Produces equivalent results to non-adaptive protocol
- Often produces faster convergence to long-time result, likely mainly due to wide spacing of λ windows

Overall protocol is robust

Alibay Cyclophillin D set[6]



- Adaptive protocol: 1/3 less compute time
- Adaptive r^2 to experiment: 0.75 [0.37, 0.91]
- Non-adaptive r^2 : 0.81 [0.46, 0.94]

Conclusions

- Restraint selection, window spacing, and equilibration detection algorithms are simple but robust
- Allocating sampling time according to inter-replicate uncertainty only rarely provides an advantage
- Overall **workflow is fully automated, robust and can accelerate equilibration**

References

[1] I. Alibay, IAlibay/MDRestraintsGenerator (version 0.1.0) Zenodo 2021.
[2] S. Boresch et al., *J. Phys. Chem. B*, 2003, **107**, 9535–9551.
[3] F. Clark, G. Robb, D. J. Cole and J. Michel, *J. Chem. Theory Comput.*, 2023, **19**, 3686–3704.
[4] D. D. L. Minh, *J. Comp. Chem.*, 2020, **41**, 715–730.
[5] A. Rizzi, Ph.D. Thesis, Weill Medical College of Cornell University, 2020.
[6] I. Alibay et al., *Commun. Chem.*, 2022, **5**, 105.