GRAND CANONICAL MONTE CARLO

Hannah Bruce Macdonald
Dr. Christopher Cave-Ayland
Prof. Jonathan Essex
University of Southampton
20th February 2016

GCMC

GCMC background

Simulation set-up

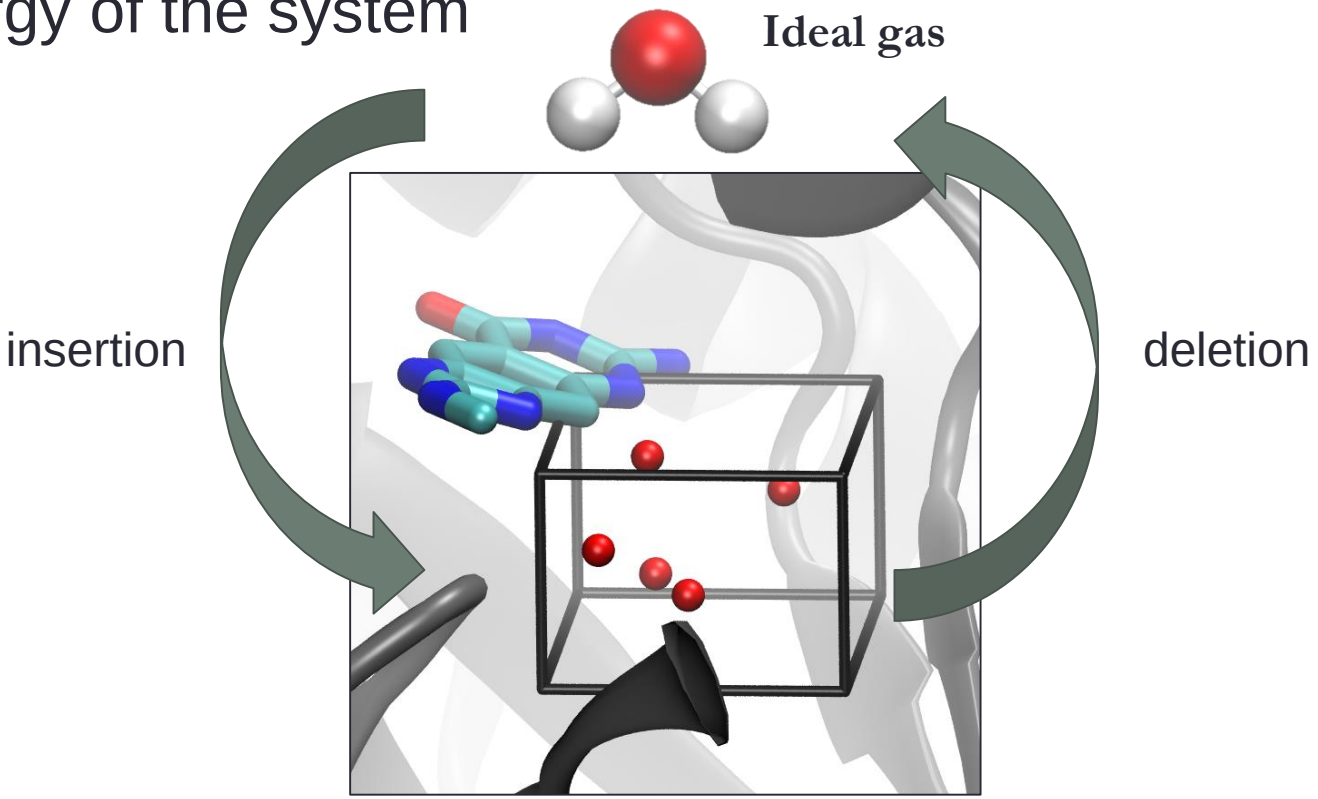
Results and Analysis

Replica exchange

Summary

Grand Canonical Monte Carlo

- μVT ensemble, where μ is chemical potential
- N varies
- Find the number of waters which minimises the free energy of the system



Varying N

- **Inserting waters**
- B = μ□β + ln(N)
- B proportional to μ□

1) Generate random water position and orientation

2) Calculate energy of inserting

3) Accept or reject insertion

Low B High B

Insertion

$$P_{insert} = \min\left[1, \frac{1}{N+1} e^{B} e^{-\beta \Delta E}\right]$$

Varying N

- **Deleting waters**

1) Pick a GCMC water molecule at random
2) Calculate energy of removing
3) Accept or reject deletion

Low B High B

Deletion

$$P_{remove} = \min\left[1, Ne^{-B}e^{-\beta\Delta E}\right]$$

- **Moving GC waters**
- Sample the ‘on’ GC water molecules at random with normal acceptance criteria

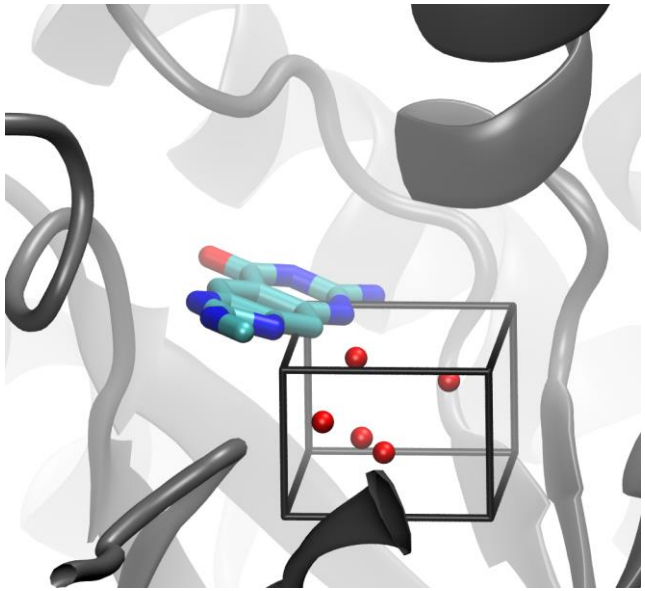
Simulation set up

- Same ‘normal’ MC steps, but now GCMC sampling, insertion and deletion included.
- GC insertion must equal GC deletion (detailed balance)
- Two equilibration steps
  - GCMC only
  - full system sampling
- Production simulation of 40 M simulation steps

GCMC box

Issue

- GCMC can have issues with acceptance rates
- Inserting into a dense system



Solution

- Limit GCMC to a specific region
  - Over a water site of interest
  - Around a group of a ligand
  - Over a ligand
  - Complete active site
- Remove any known water molecules in the region
- Box can cover ligand or protein atoms
- TIP4P used as default

B values

Single B

- One processor
- Water locations
- N*
- Simulate at B_{equil}

B_{equil}

- $B_{equil} = \mu \beta + \ln(V_{box}/V^{o})$

Excess chemical potential of water in water **-6.2 kcal mol^{-1}**

Thermodynamic beta **1.69 kcal^{-1} mol**

Standard volume of a water molecule in bulk **30.0 A^{3}**

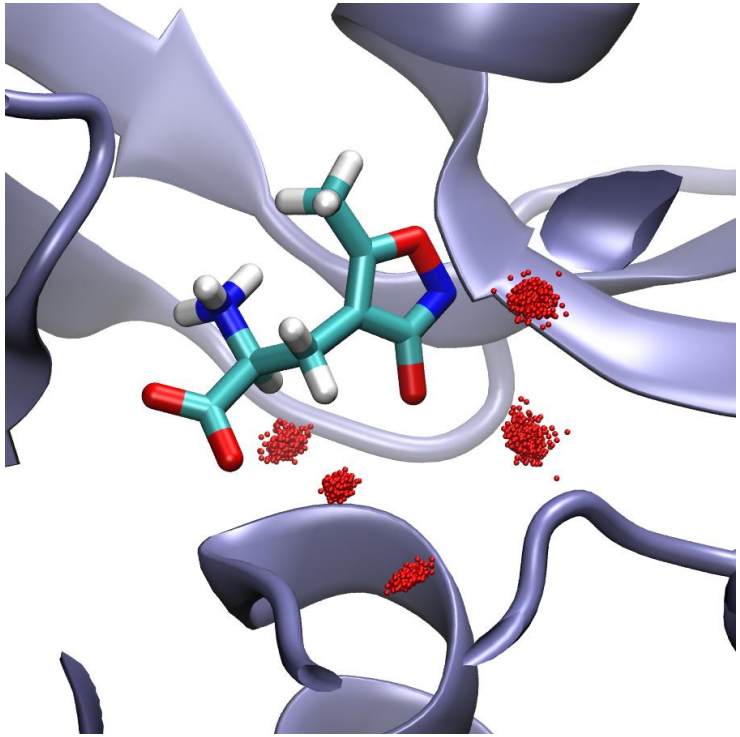
Volume of GCMC region – set a priori

Titration

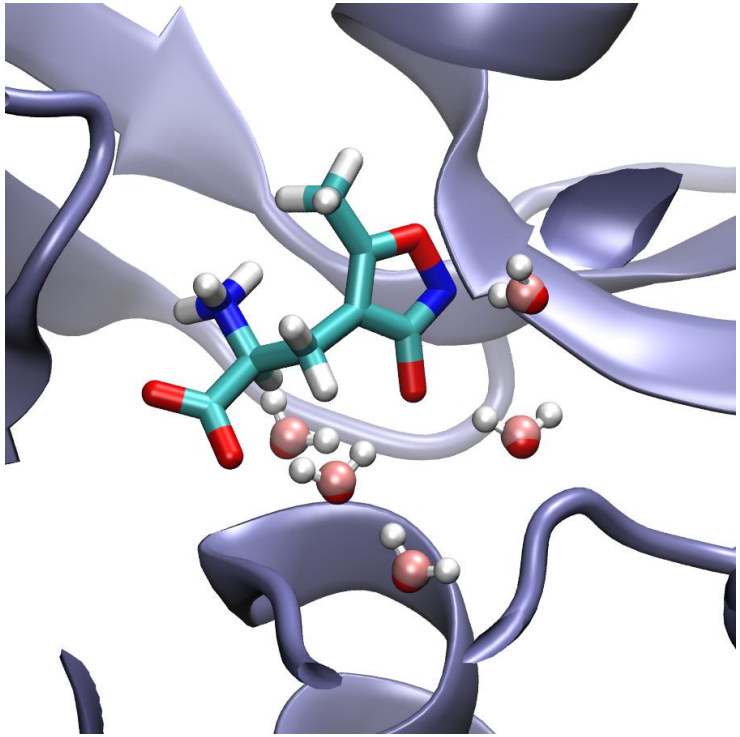
- Multiple processors
- Water locations
- N*
- Binding free-energy of water molecules and networks
- Improved sampling from replica exchange
- Simulate at a range of B values
- B_{equil}-21 to B_{equil}+2

Water Locations

- B_{equil} should simulate N*
- All of the GCMC water locations from the simulations can be looked at

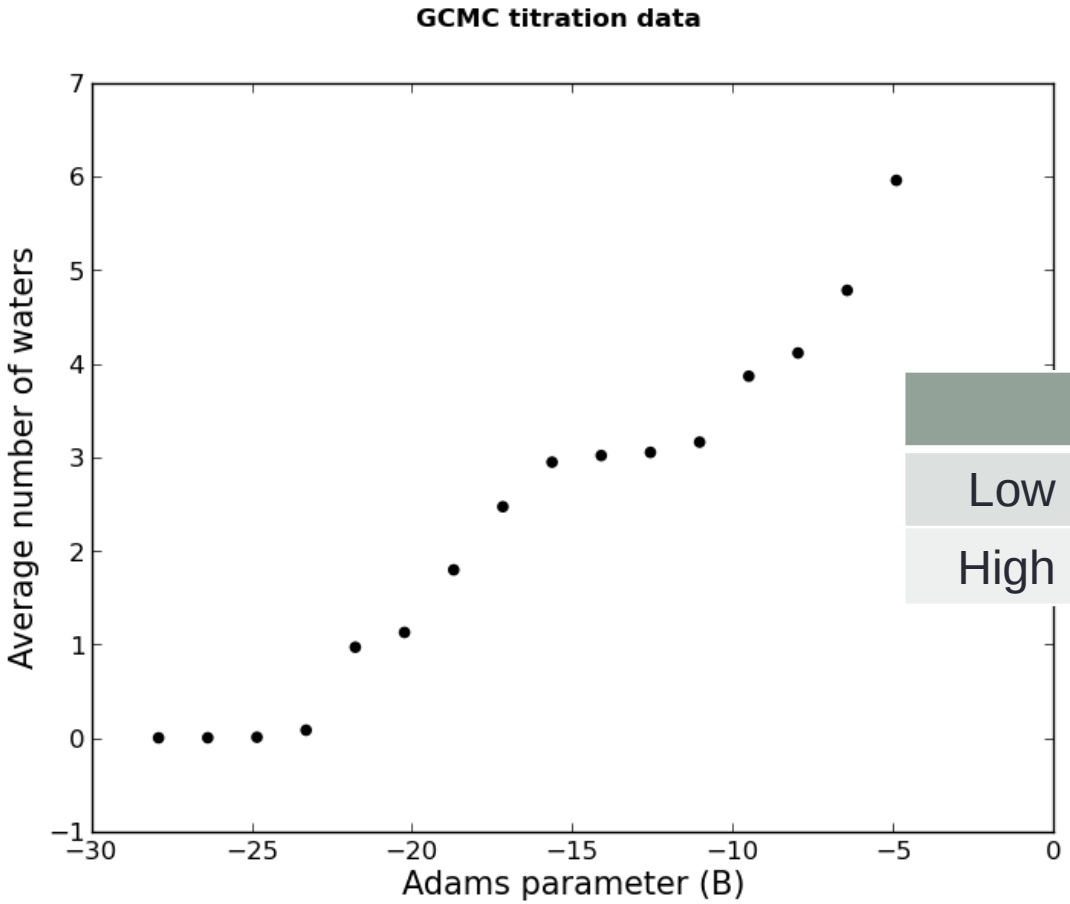


- The locations can be clustered to give the water network



Titrations

- Simulations of multiple B values will have a range of water occupancies.

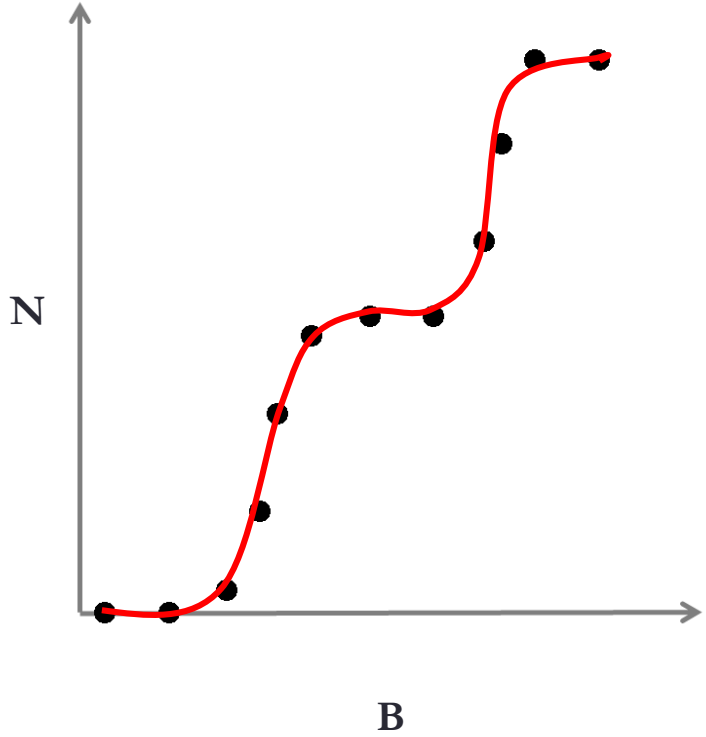


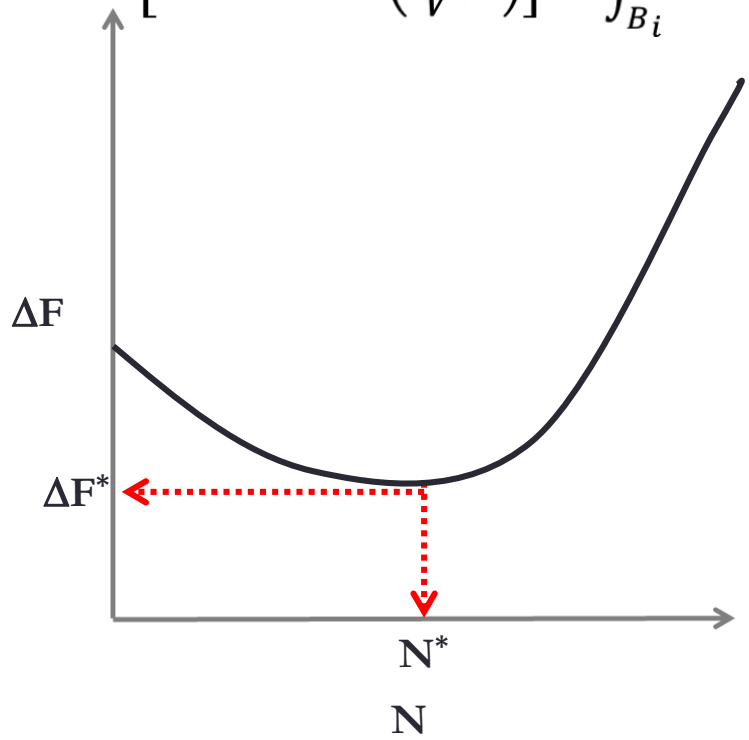
	Insertion	Deletion	<N>
Low B	↓	↑	↓
High B	↑	↓	↑

Calculating binding free energies

- Binding free energy of the water networks can be calculated from the titration.
- Lowest energy network will be N*

$$\beta\Delta G^{o}_{bind}(N_i - N_f) = N_f B_f - N_i B_i - (N_f - N_i)\left[\beta\mu'_{sol} + \ln\left(\frac{V_{box}}{V^o}\right)\right] - \int_{B_i}^{B_f} N(B)dB$$

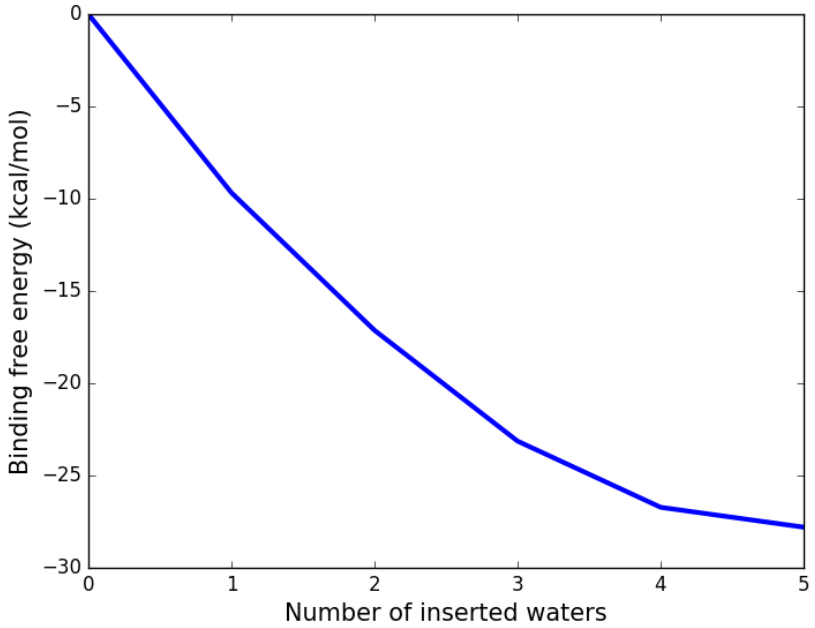




Binding free energies

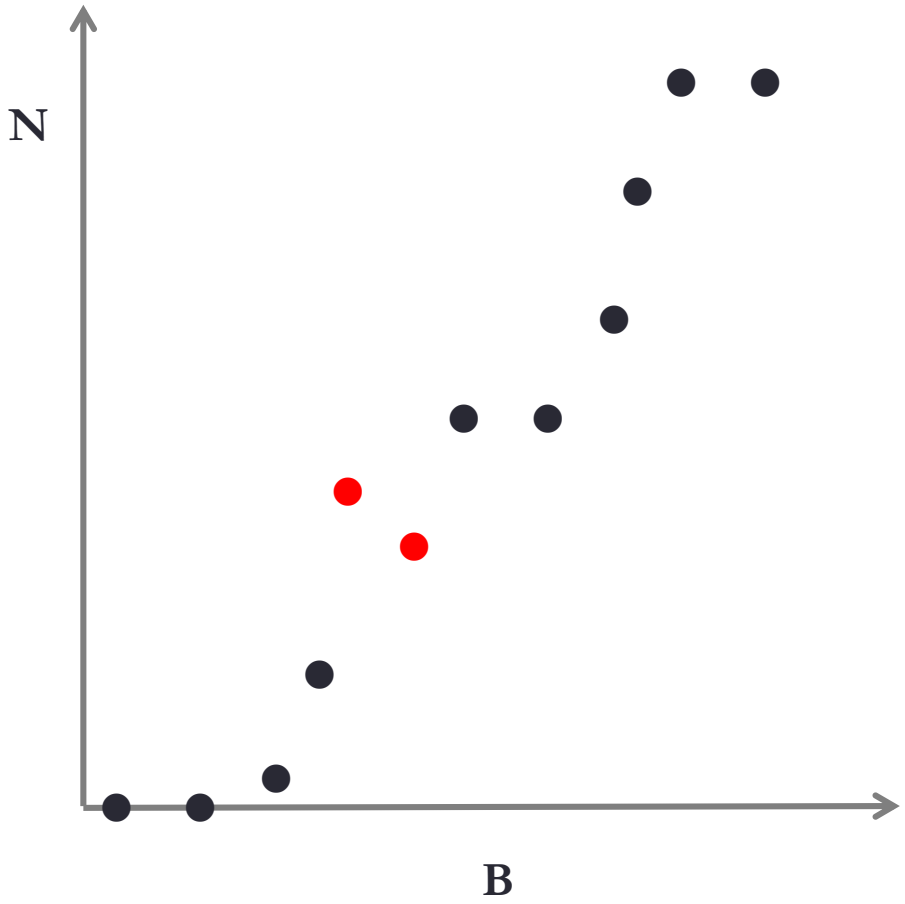
FREE ENERGIES:
Quoted errors are from multiple repeats of the fitting.

	IDEAL GAS TRANSFER FREE ENERGIES					BINDING FREE ENERGIES	
'# Waters'	'Mean'	'Std. dev.'	'25th Percentile'	'Median'	'75th Percentile'	'Mean'	'Median'
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	-15.89	0.03	-15.89	-15.88	-15.87	-9.69	-9.68
2.00	-29.53	0.06	-29.56	-29.52	-29.48	-17.13	-17.12
3.00	-41.73	0.06	-41.75	-41.73	-41.68	-23.13	-23.13
4.00	-51.51	0.03	-51.54	-51.52	-51.50	-26.71	-26.72
5.00	**-58.78**	**0.06**	**-58.81**	**-58.77**	**-58.74**	**-27.78**	**-27.77**



Replica exchange

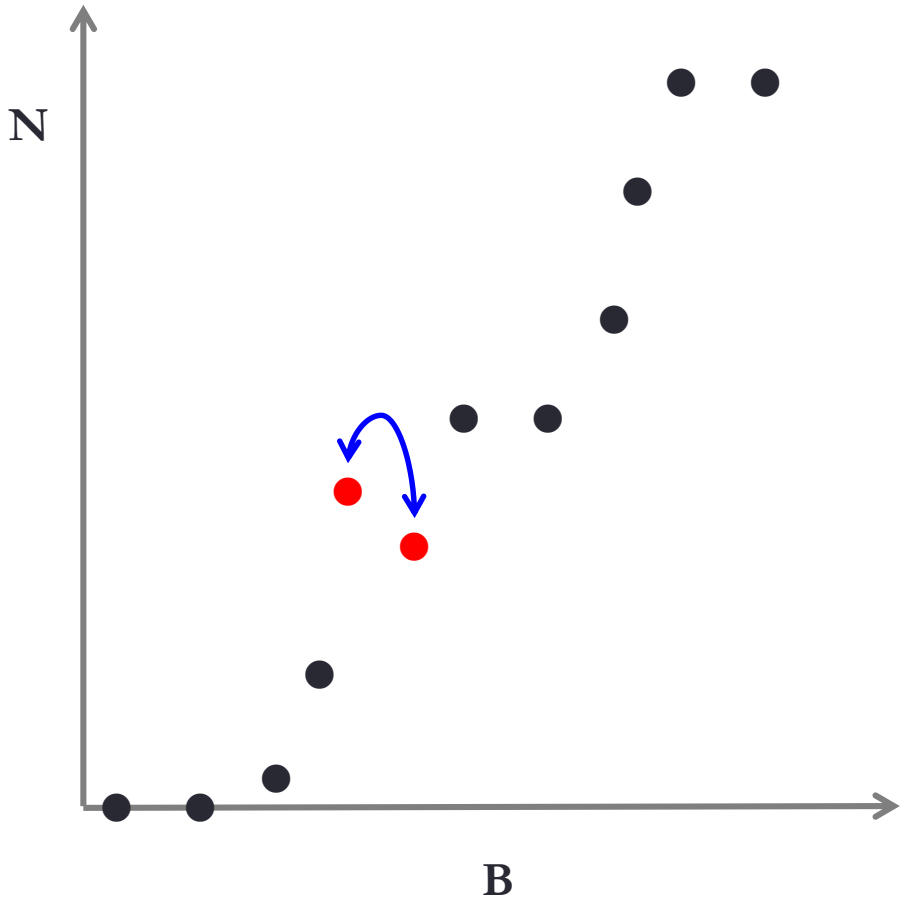
- Results should be monotonic, but sometimes are not due to convergence.
- Attempt RE in B values to help get more monotonic results
- Easier to fit



$$P_{swap} = \min\left[1, e^{\Delta B \Delta E}\right]$$

Replica exchange

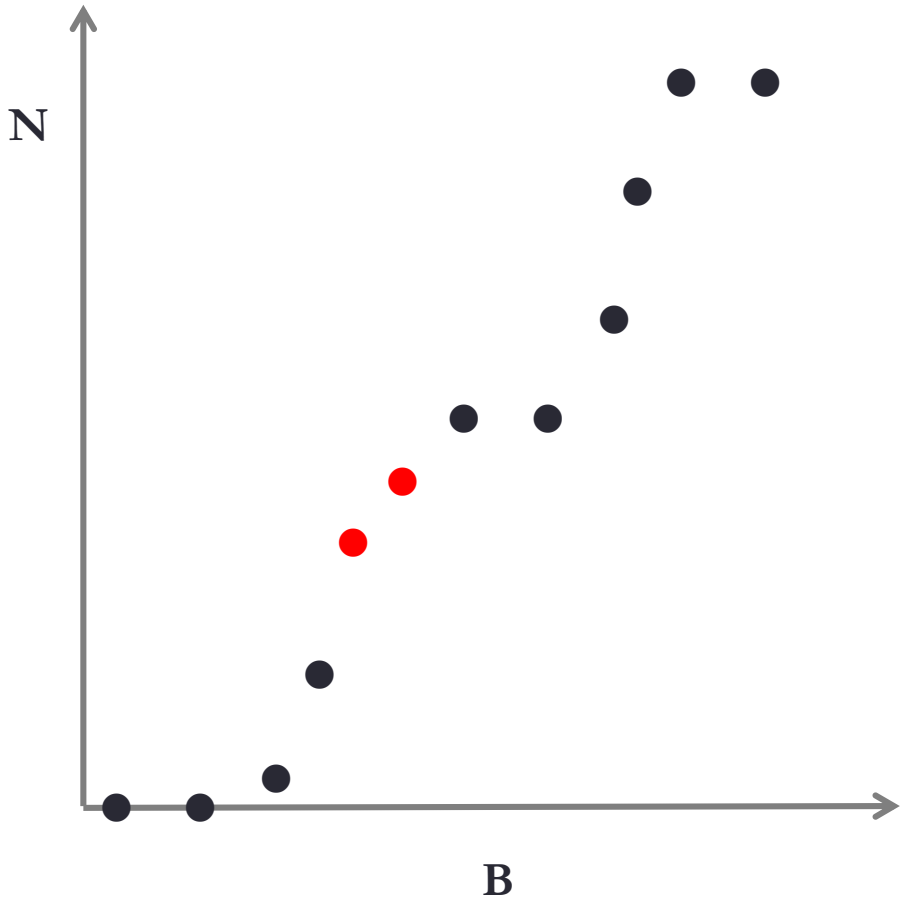
- Results should be monotonic, but sometimes are not due to convergence.
- Attempt RE in B values to help get more monotonic results
- Easier to fit



$$P_{swap} = \min\left[1, e^{\Delta B \Delta E}\right]$$

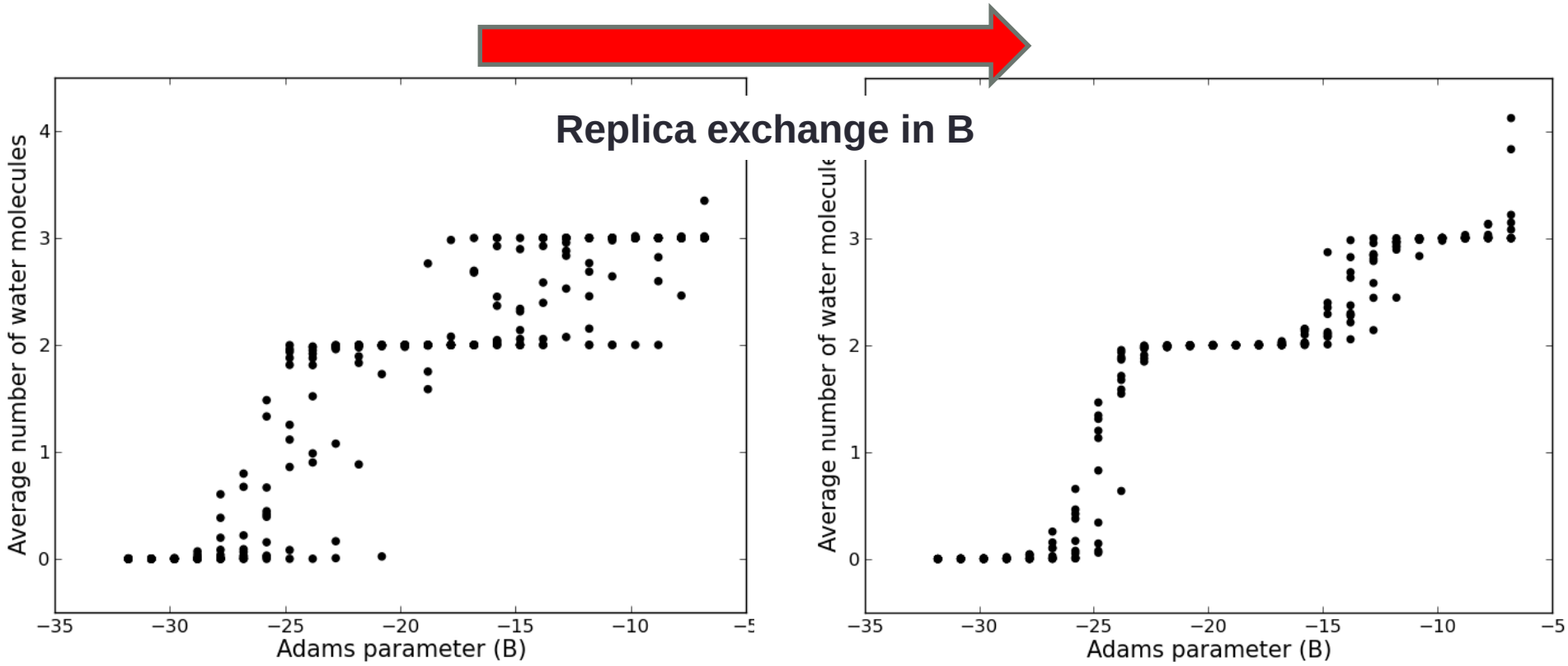
Replica exchange

- Results should be monotonic, but sometimes are not due to convergence.
- Attempt RE in B values to help get more monotonic results
- Easier to fit

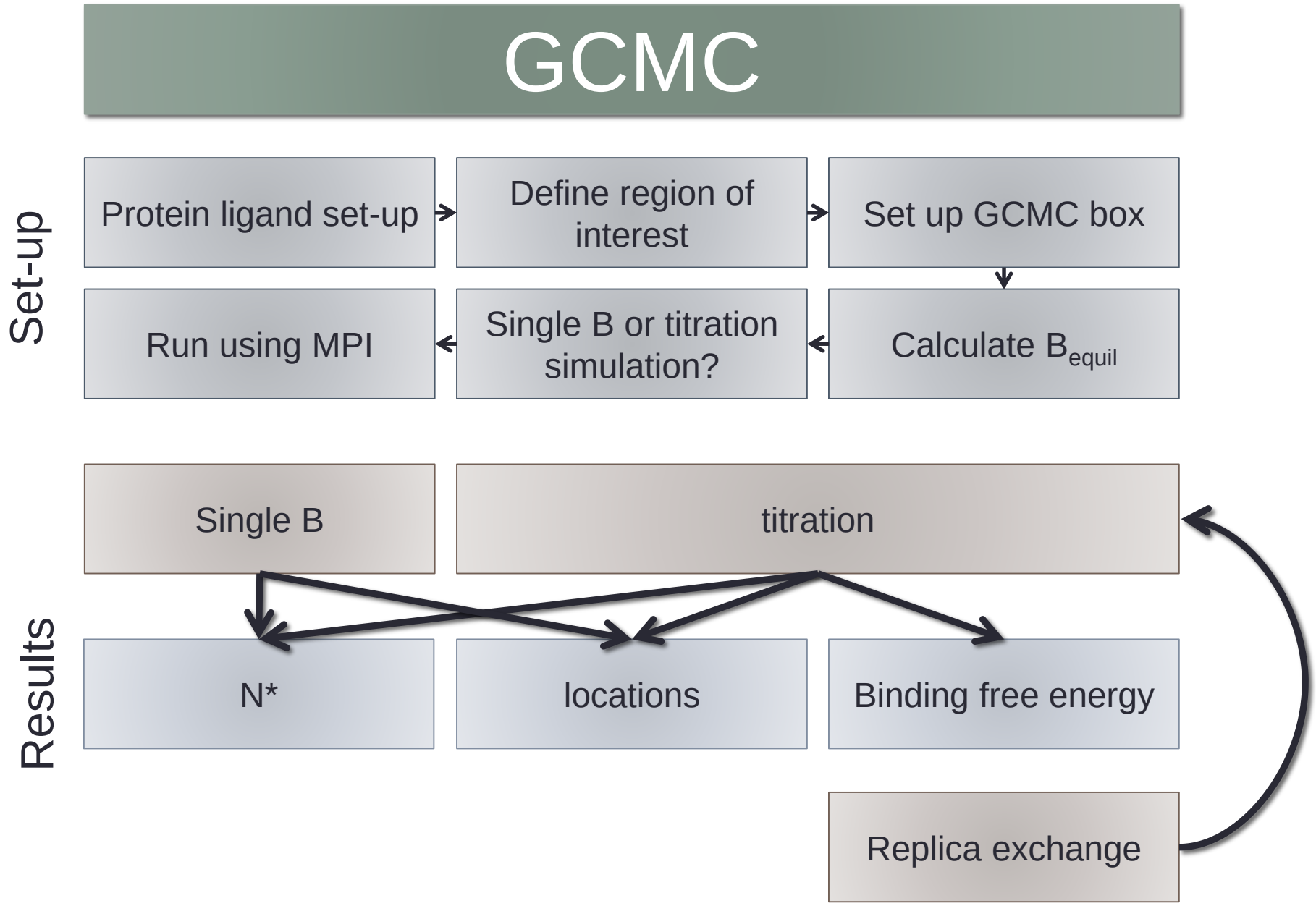


$$P_{swap} = \min\left[1, e^{\Delta B \Delta E}\right]$$

Replica exchange



GCMC


Set-up
Protein ligand set-up
Define region of interest
Set up GCMC box
Calculate B_{equil}
Single B or titration simulation?
Run using MPI
Results
Single B
titration
N*
locations
Binding free energy
Replica exchange

Summary

- No knowledge of water molecules needed (occupancy or location)
- Two types of simulation
  - Single B
  - titration
- Can calculate optimal occupancy and cluster locations
- Titrations can calculate binding free energies of water molecules and networks using GCI equation
- Improved reliability with replica exchange