

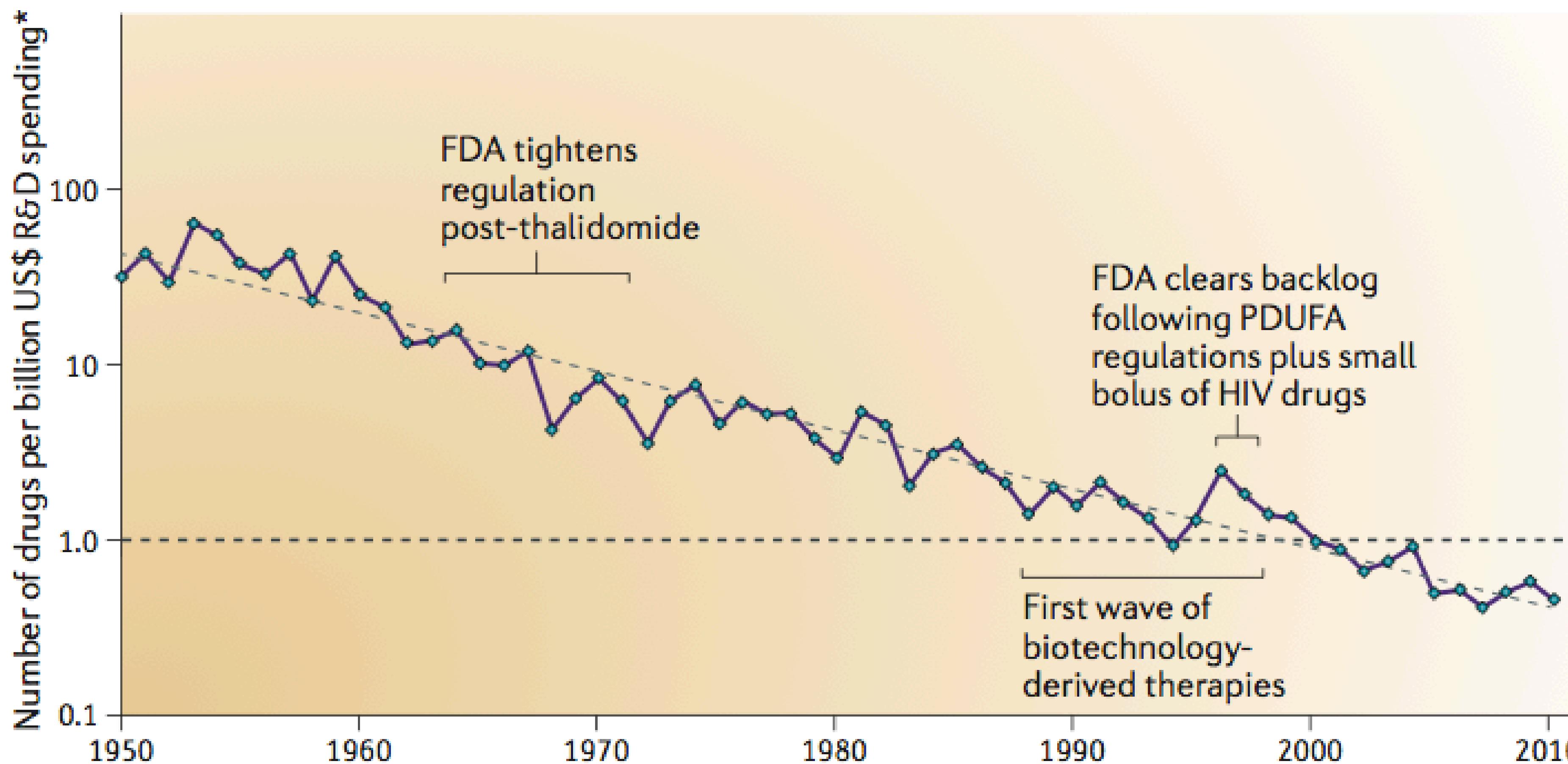


Free energy calculations: What do we want them, what do they do, and how do they work?

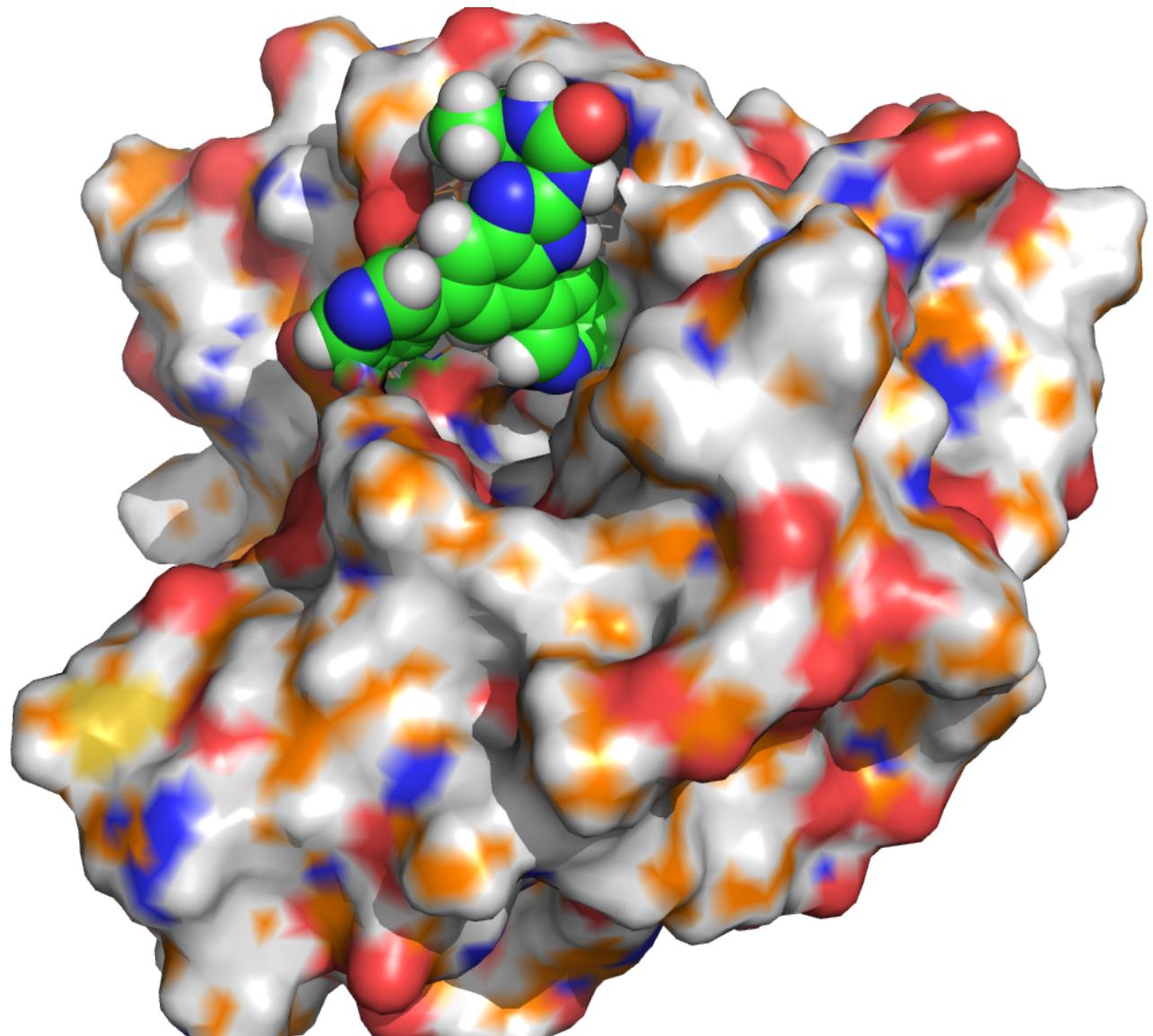
David Mobley
UC Irvine

Drug discovery is hard and expensive; we want to help change that

a Overall trend in R&D efficiency (inflation-adjusted)



Computation could guide drug discovery



Seven years from now, a chemist working on drug lead optimization generates ideas for 100 new compounds, plugs them into their computer before leaving work, and arrives back in the morning to find:

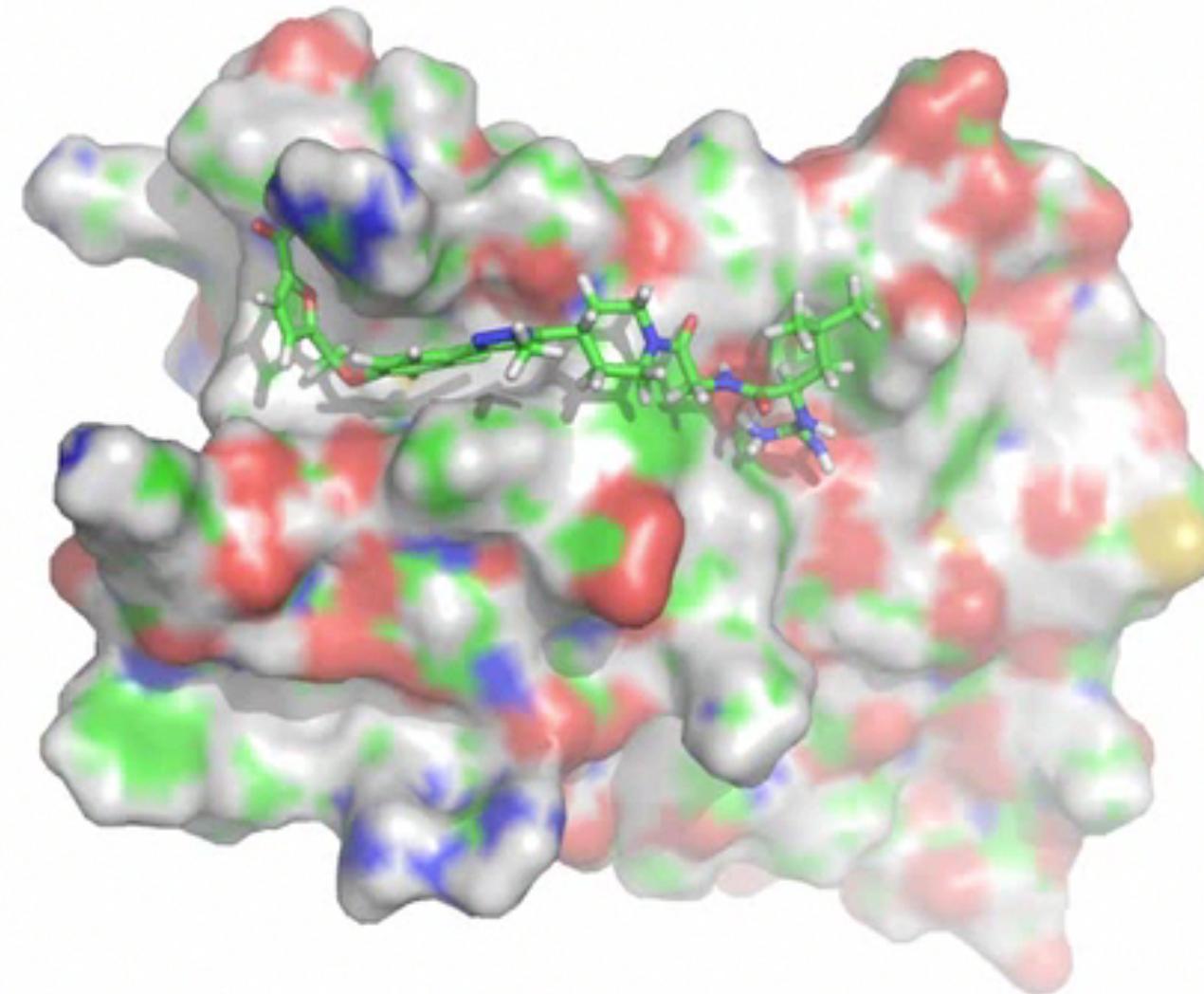
- Predicted affinity to target
- Predicted selectivity
- Likely drug resistance profile
- Predicted solubility
- Solvent conditions for purification
- Additional modifications to consider

Binding free energies involve
a ratio of partition functions

$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$

Binding free energies involve
a ratio of partition functions

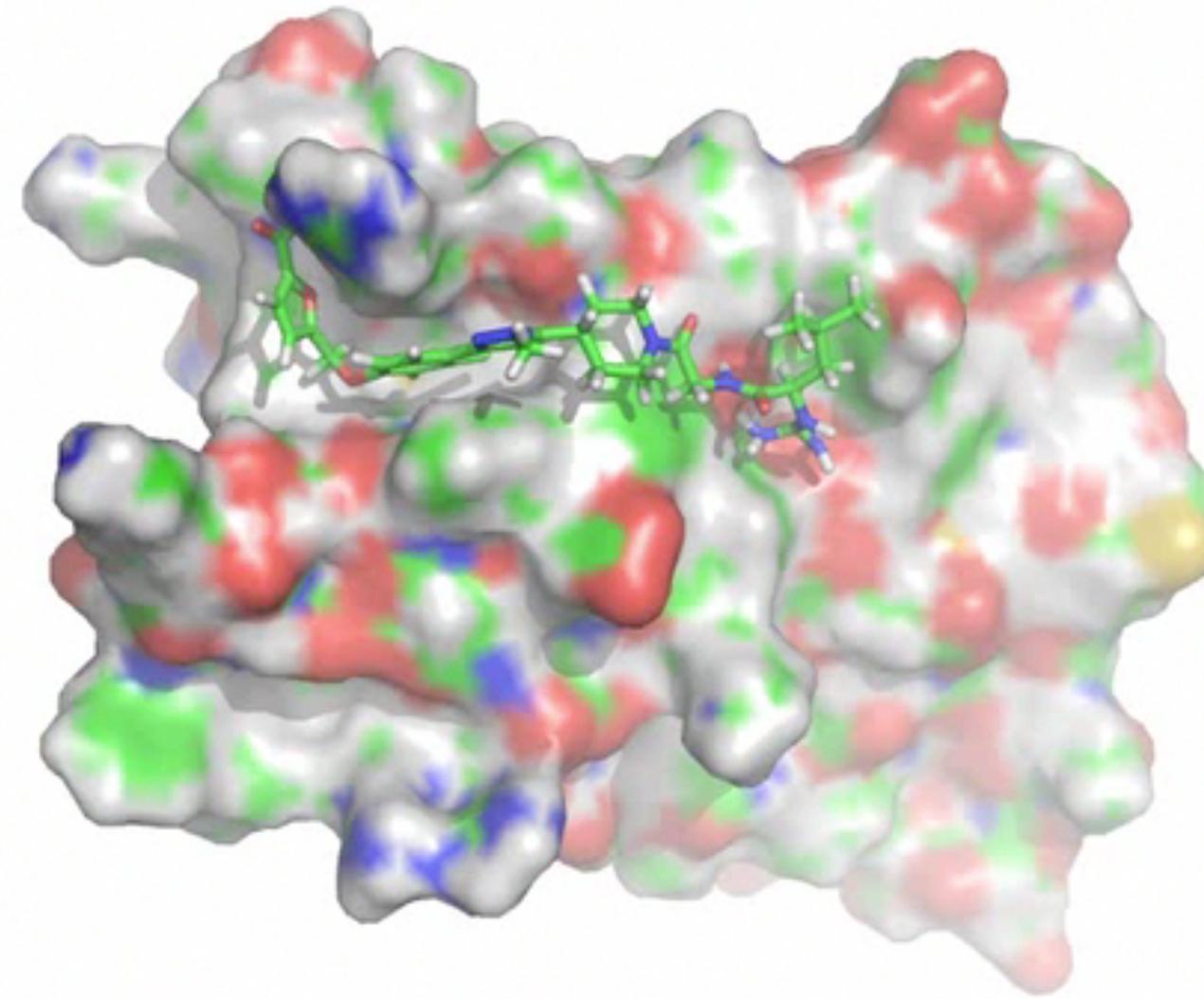
$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$



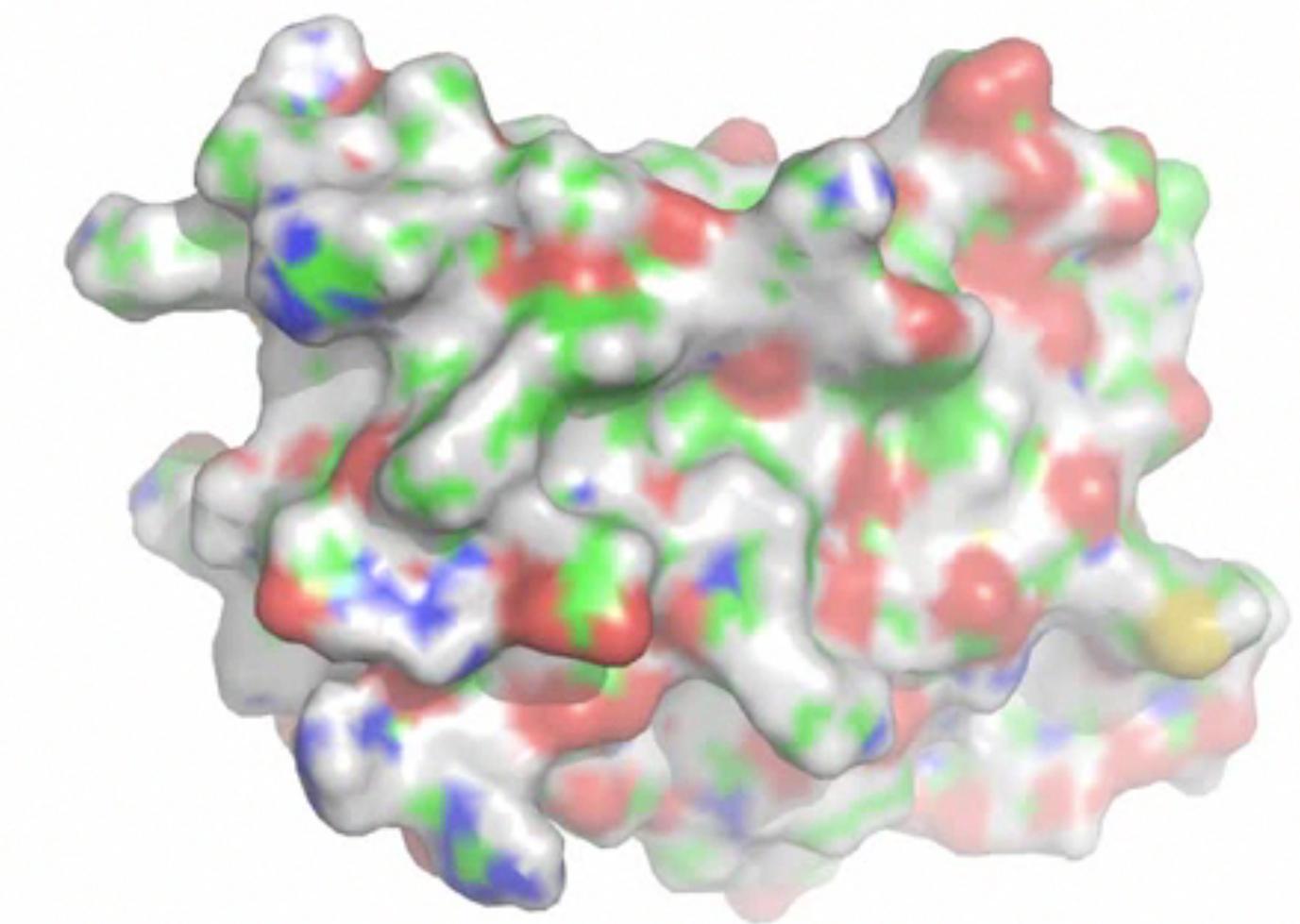
$$Q_{PL}$$

Binding free energies involve
a ratio of partition functions

$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$



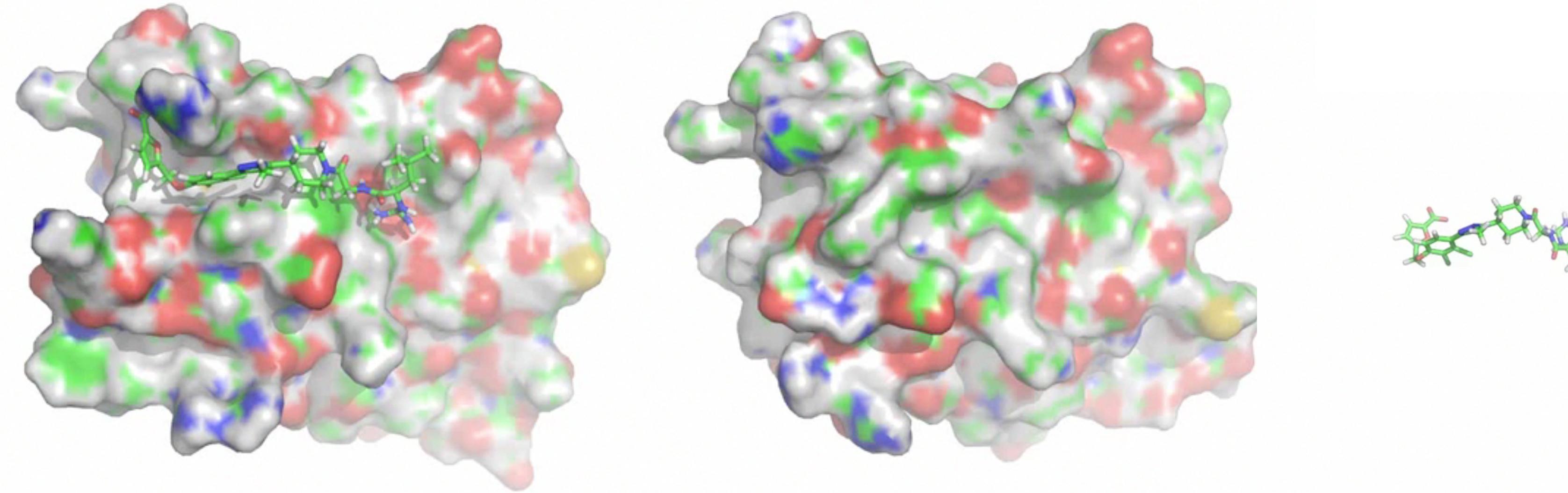
Q_{PL}



Q_P

Binding free energies involve
a ratio of partition functions

$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$

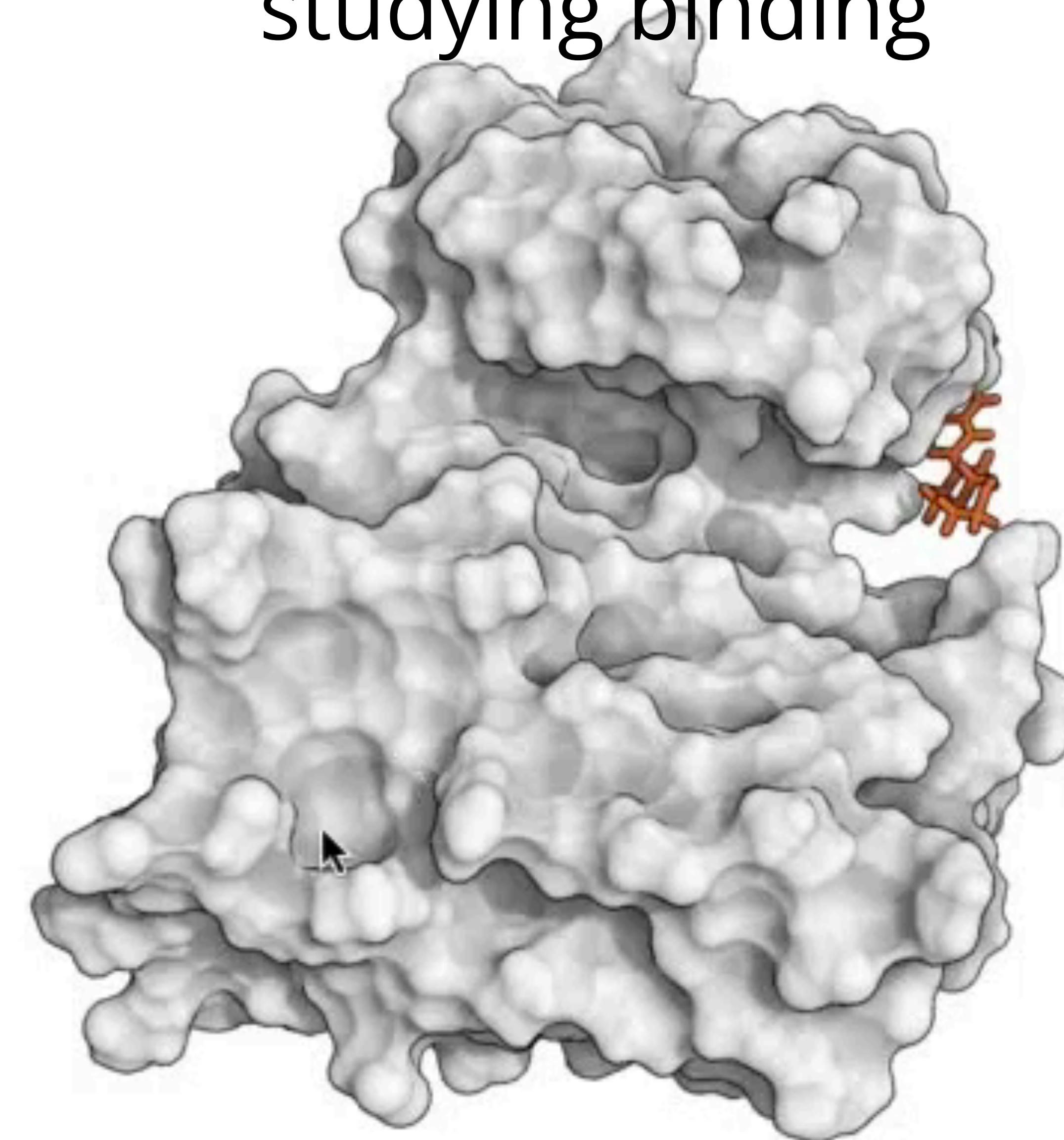


Q_{PL}

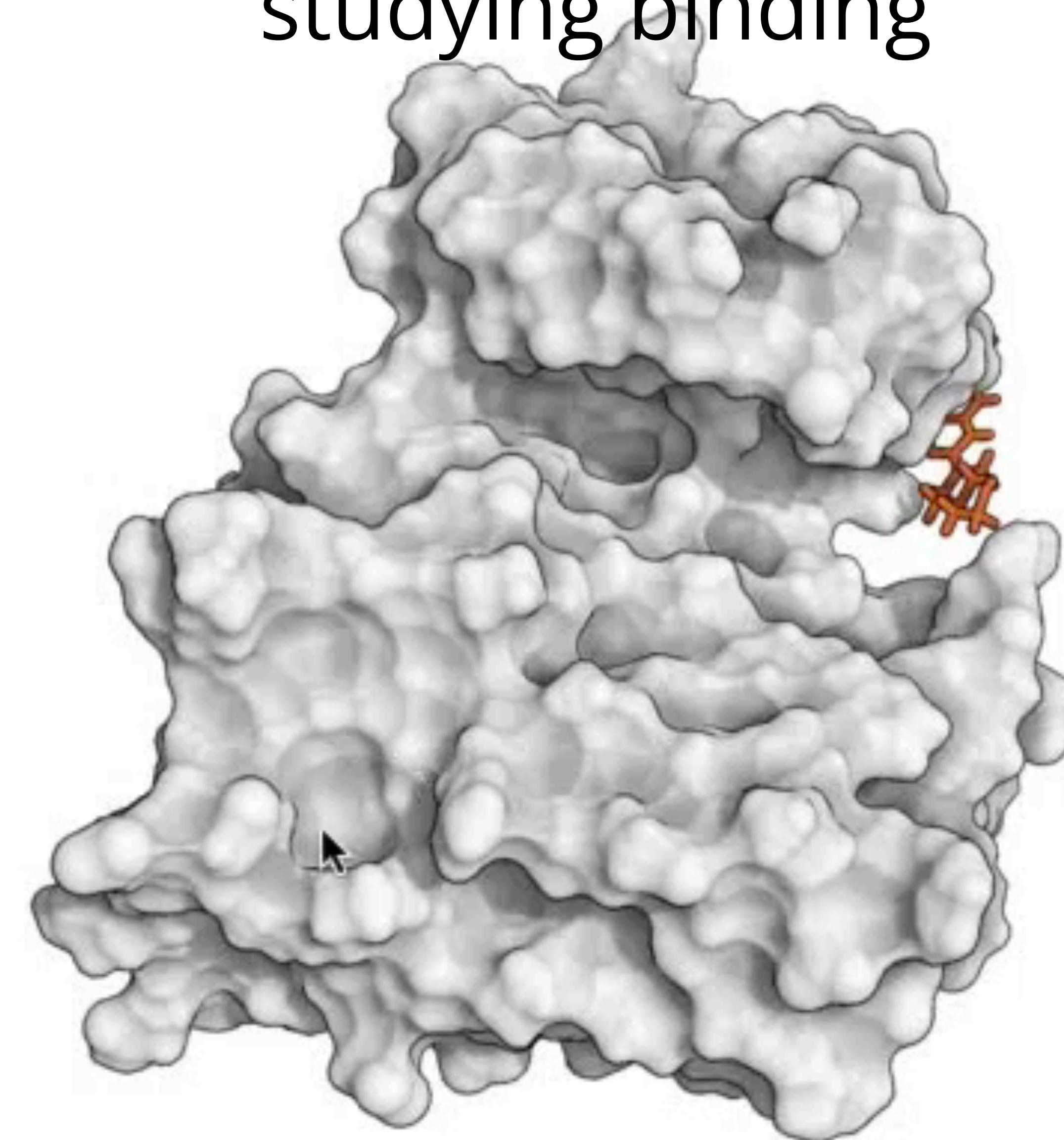
Q_P

Q_L

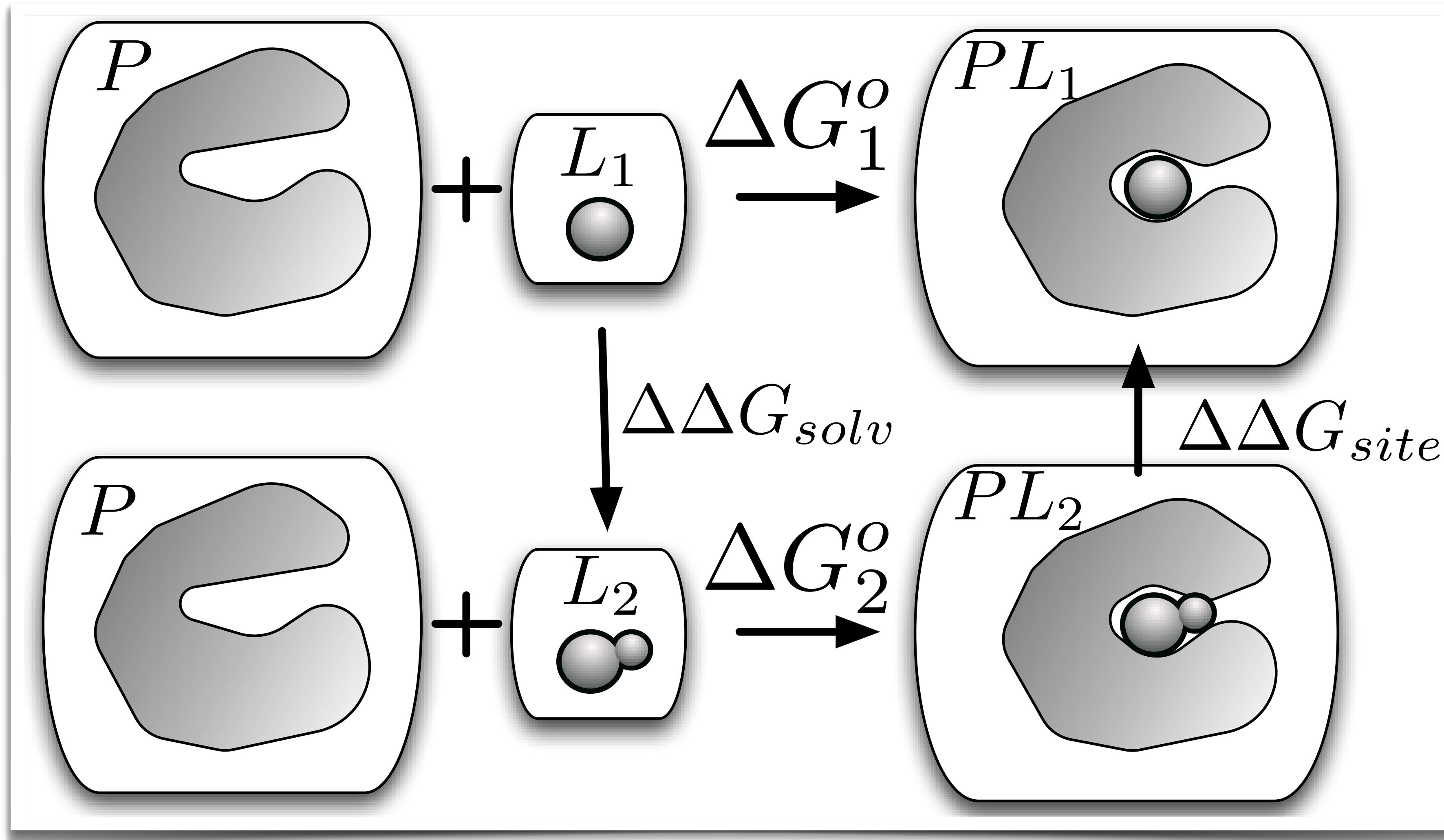
We use molecular dynamics simulations as a tool for
studying binding



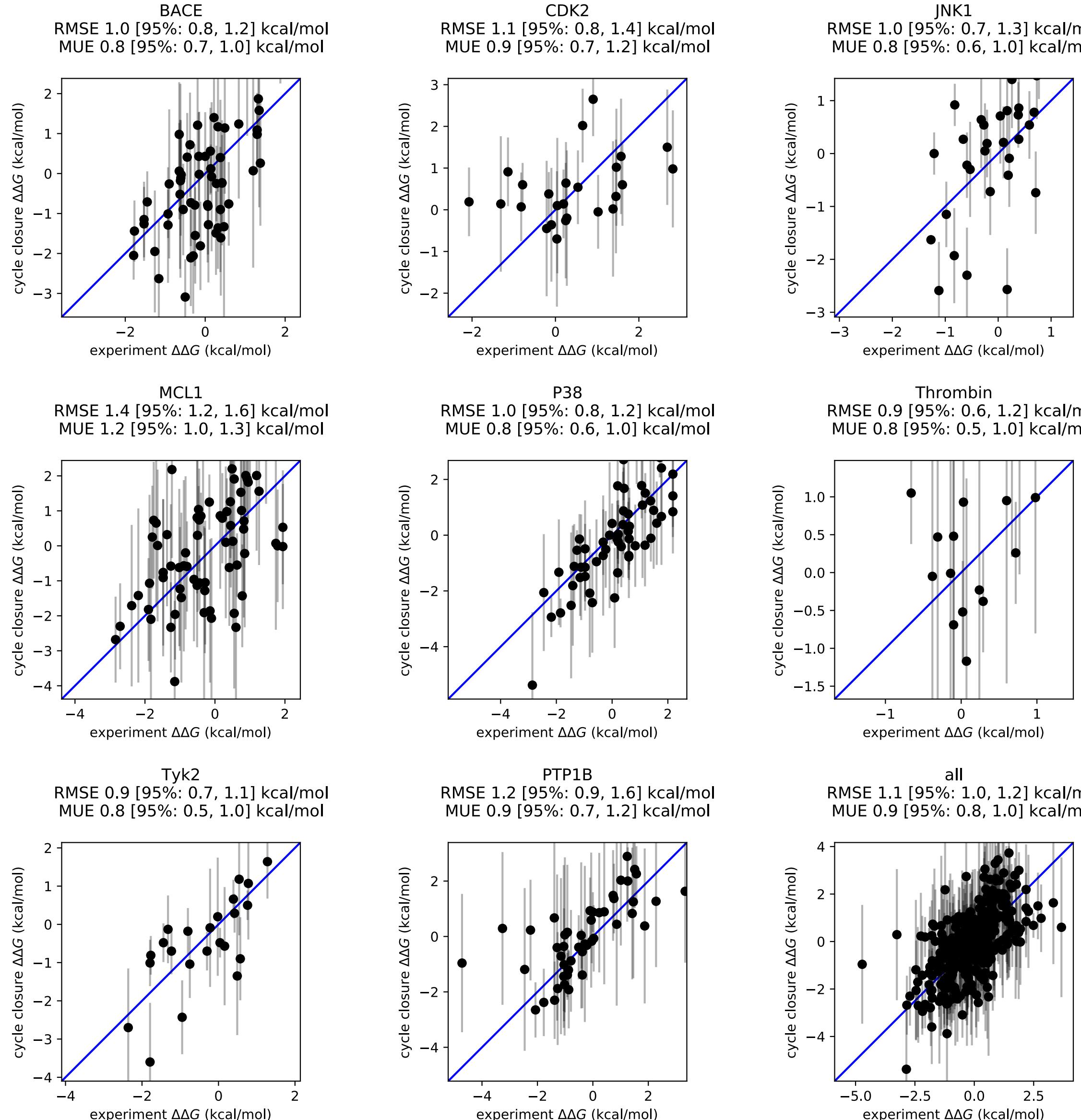
We use molecular dynamics simulations as a tool for
studying binding



Relative free energy calculations compare binding of different ligands



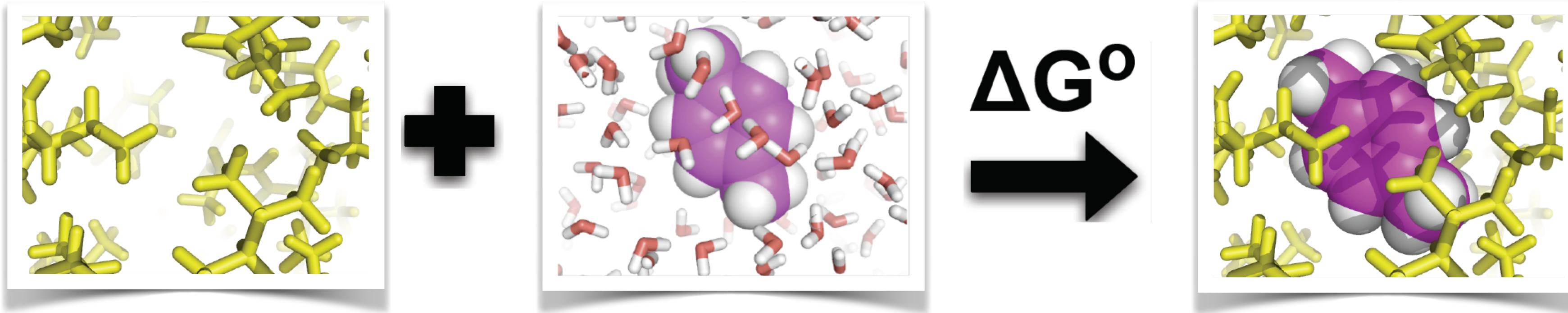
This approach has allowed large scale tests and applications of relative free energy calculations



Data JACS 2015, DOI 10.1021/ja512751q

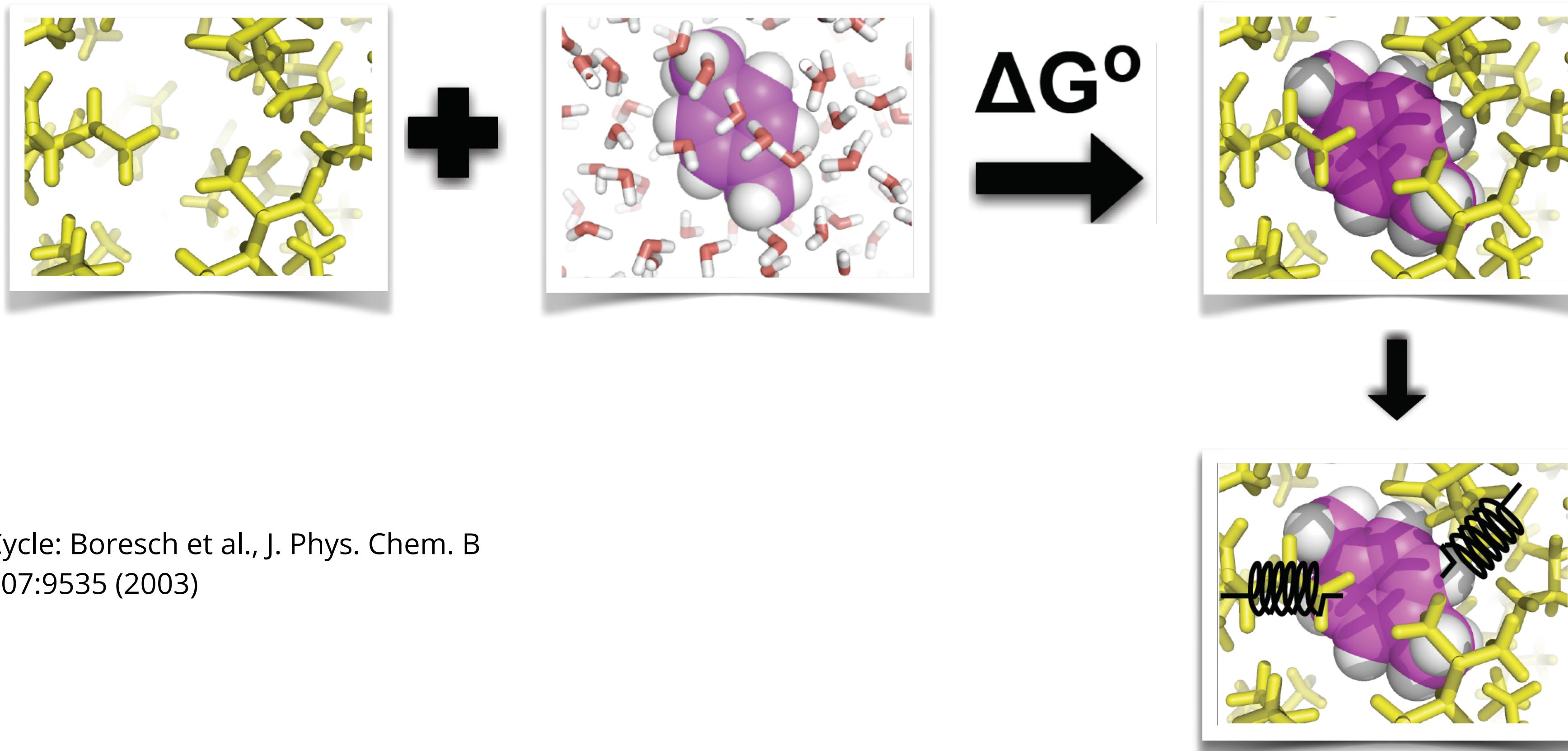
Re-plotted by John Chodera on a by-target basis

“Absolute” calculations are an alternate approach



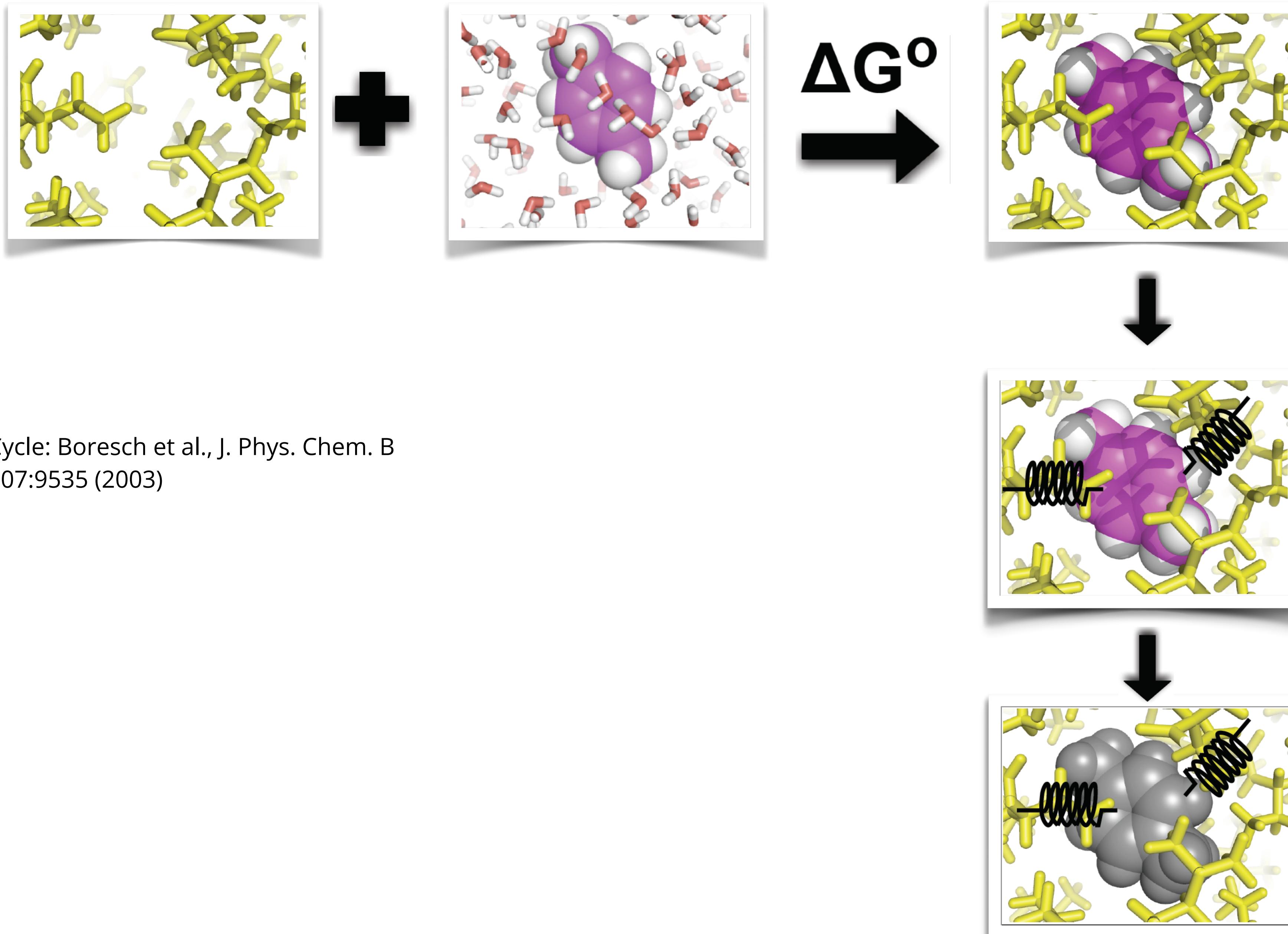
Cycle: Boresch et al., J. Phys. Chem. B
107:9535 (2003)

“Absolute” calculations are an alternate approach

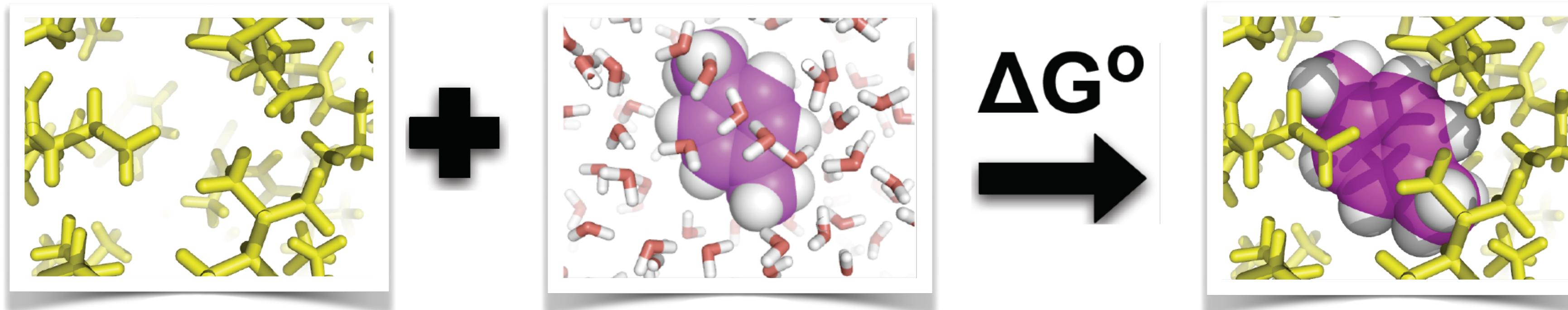


Cycle: Boresch et al., J. Phys. Chem. B
107:9535 (2003)

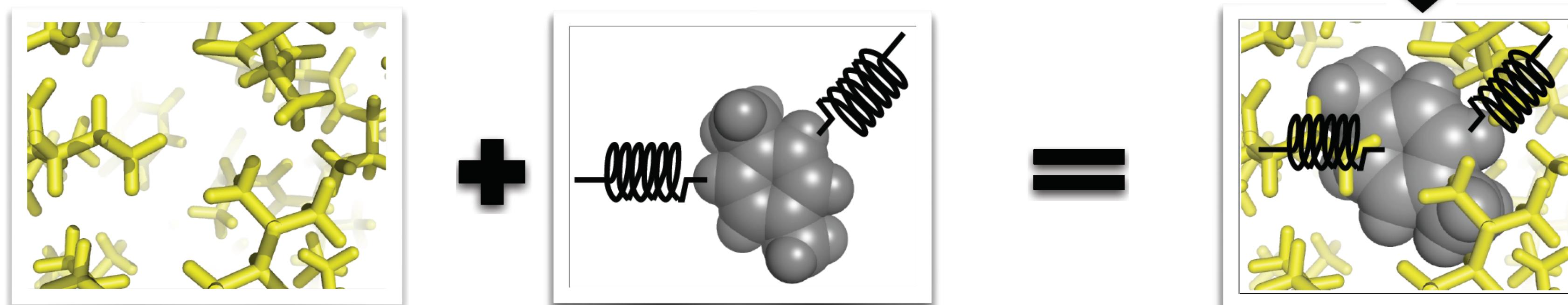
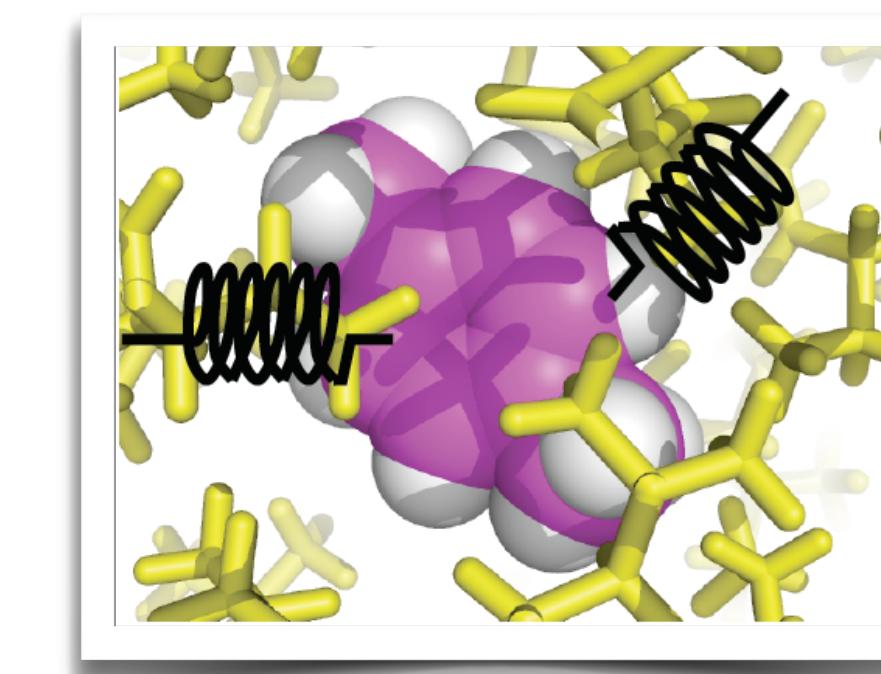
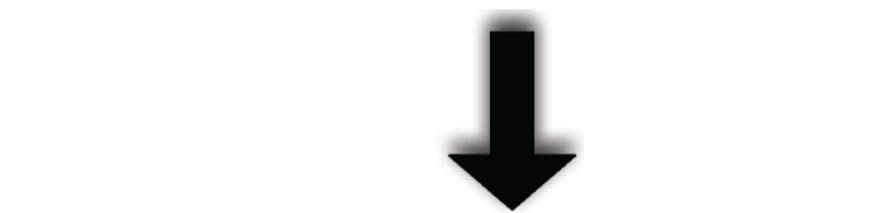
“Absolute” calculations are an alternate approach



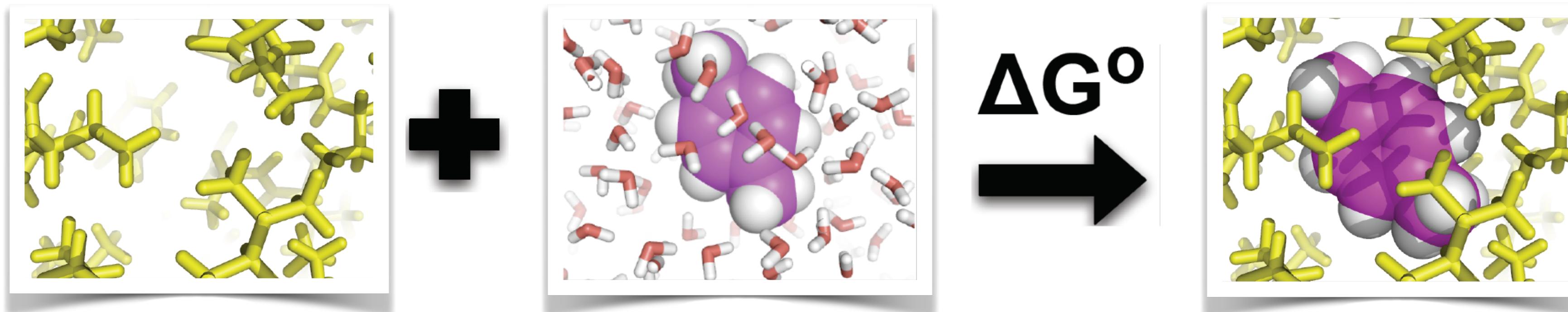
“Absolute” calculations are an alternate approach



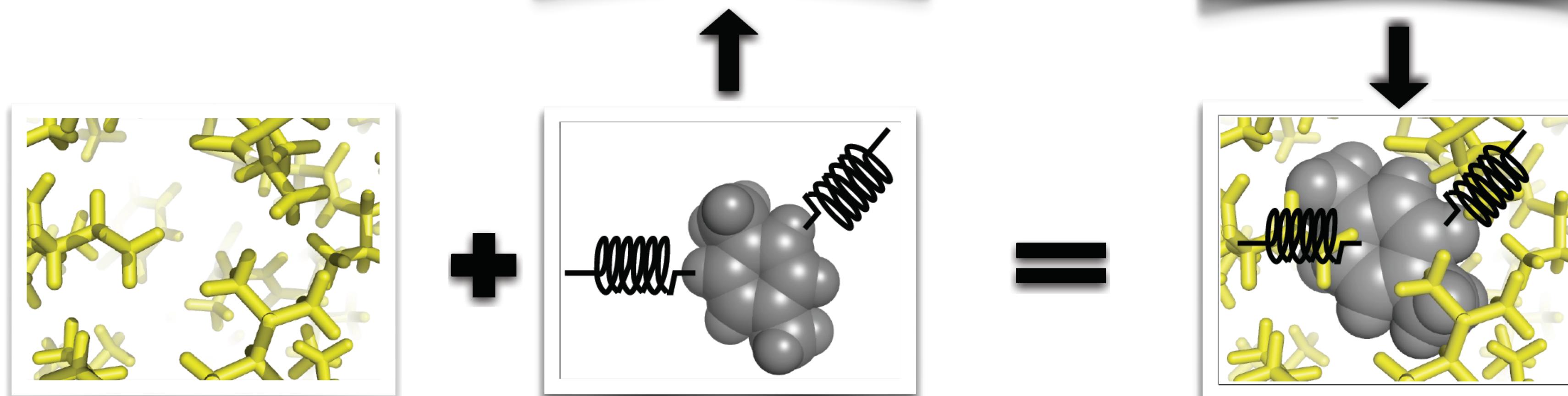
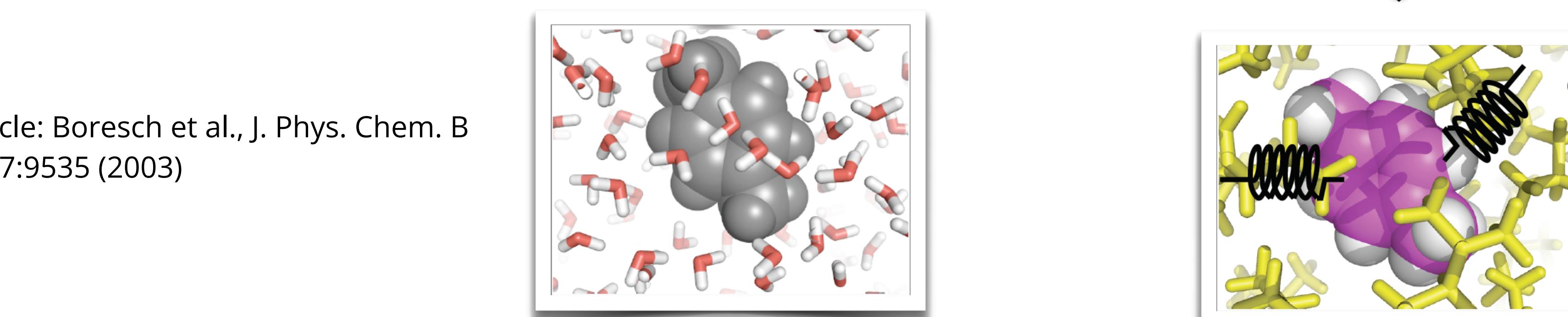
Cycle: Boresch et al., J. Phys. Chem. B
107:9535 (2003)



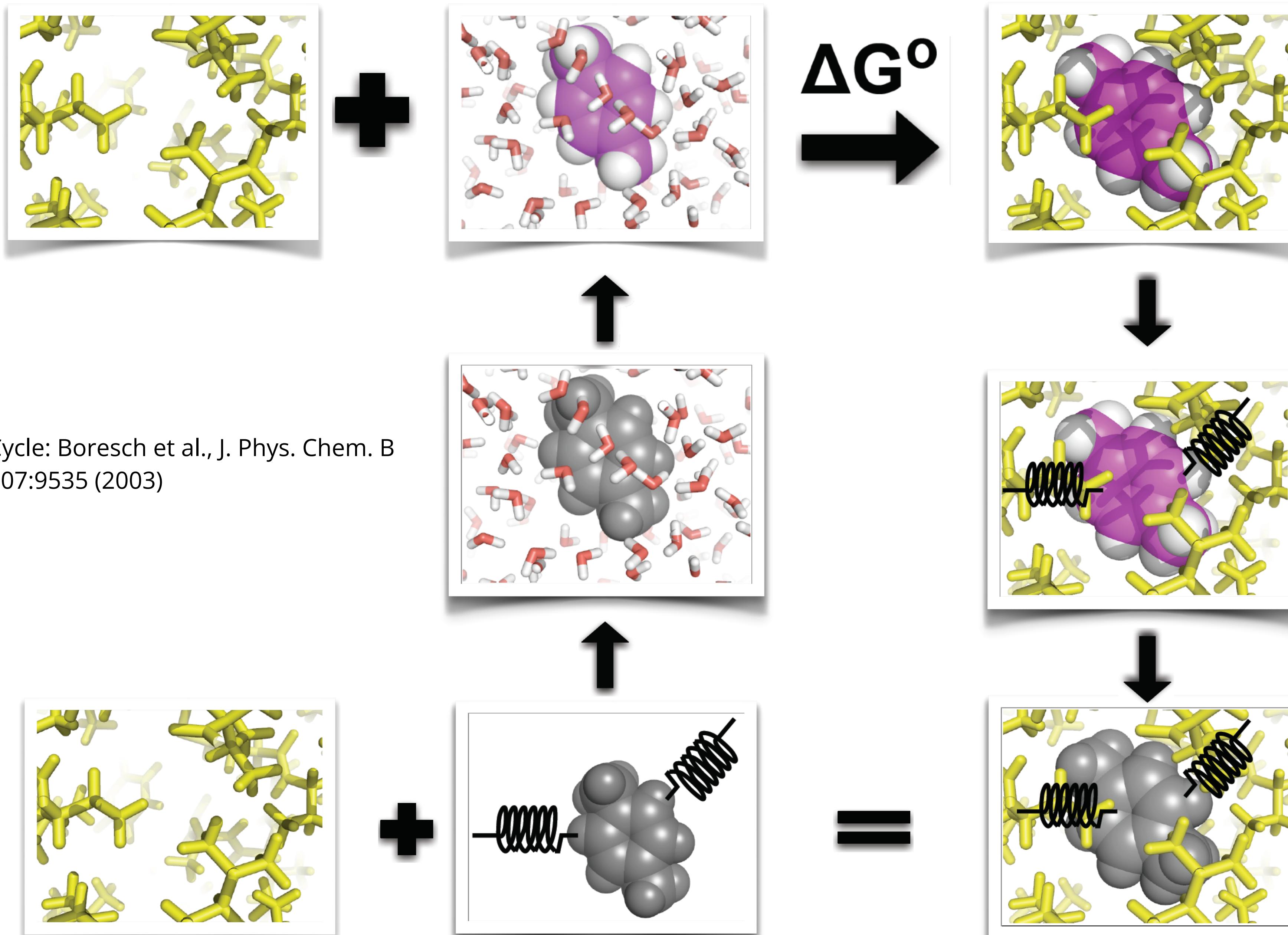
“Absolute” calculations are an alternate approach



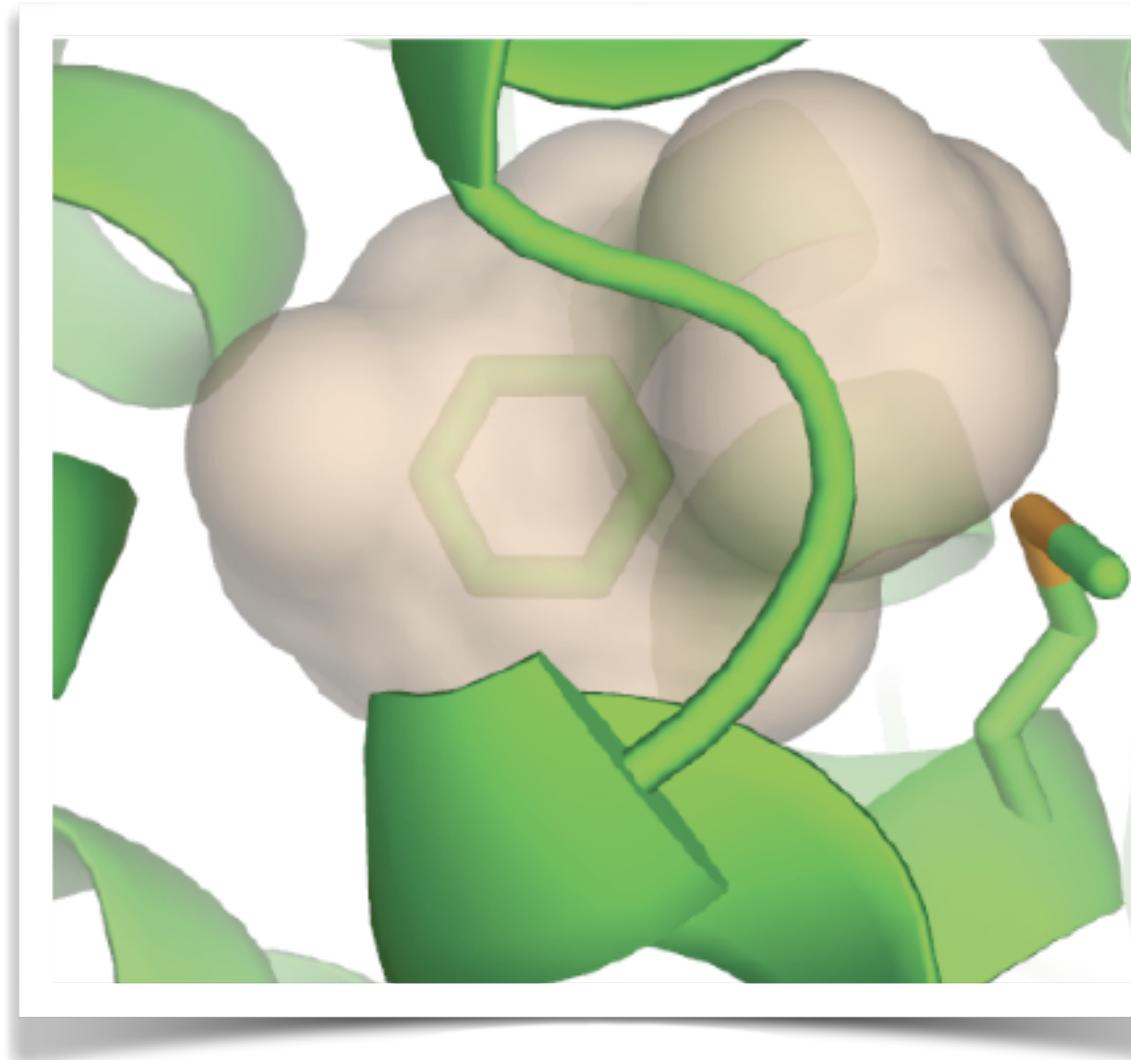
Cycle: Boresch et al., J. Phys. Chem. B
107:9535 (2003)



“Absolute” calculations are an alternate approach

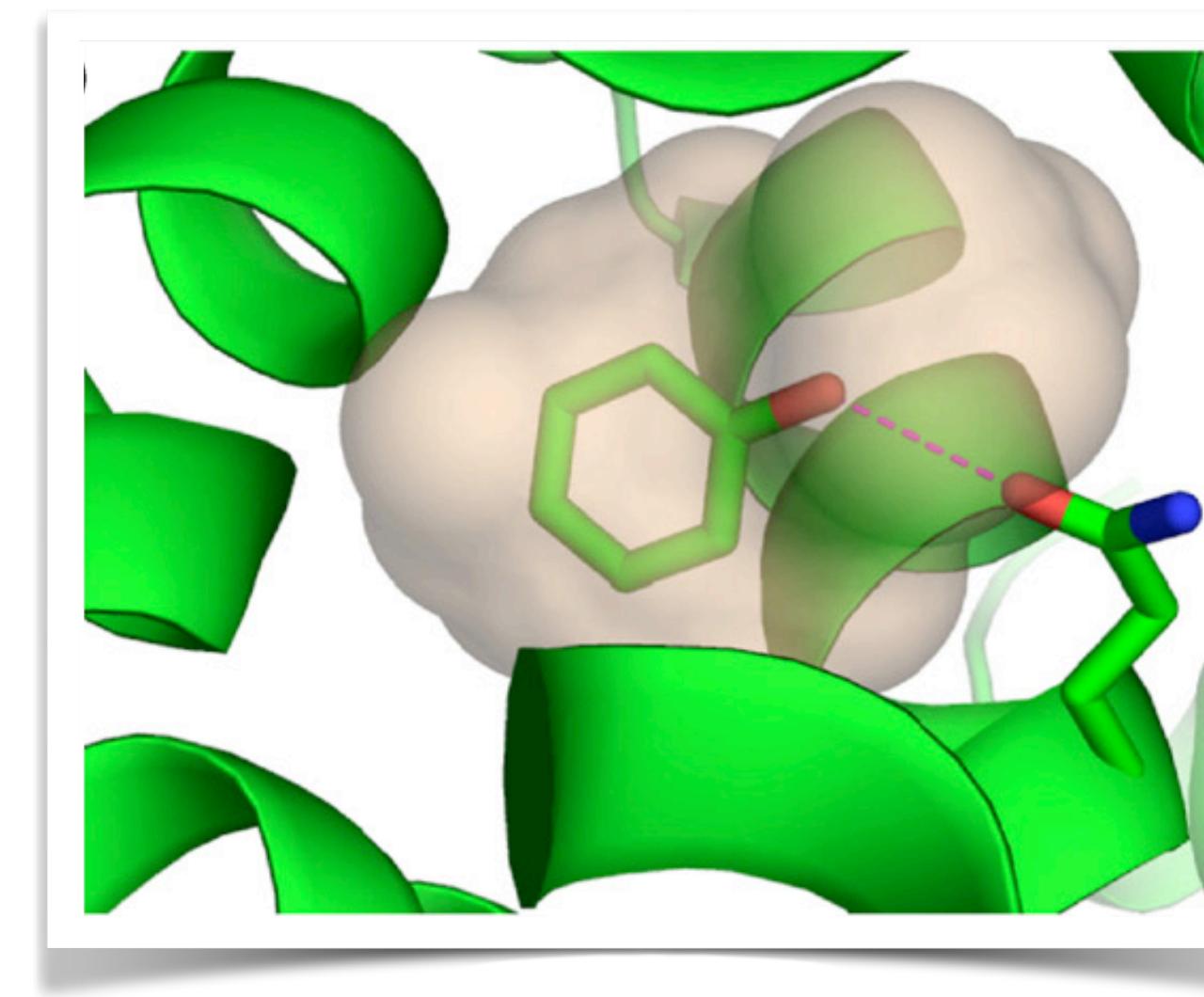


We've used a progression of model binding sites for free energy method development



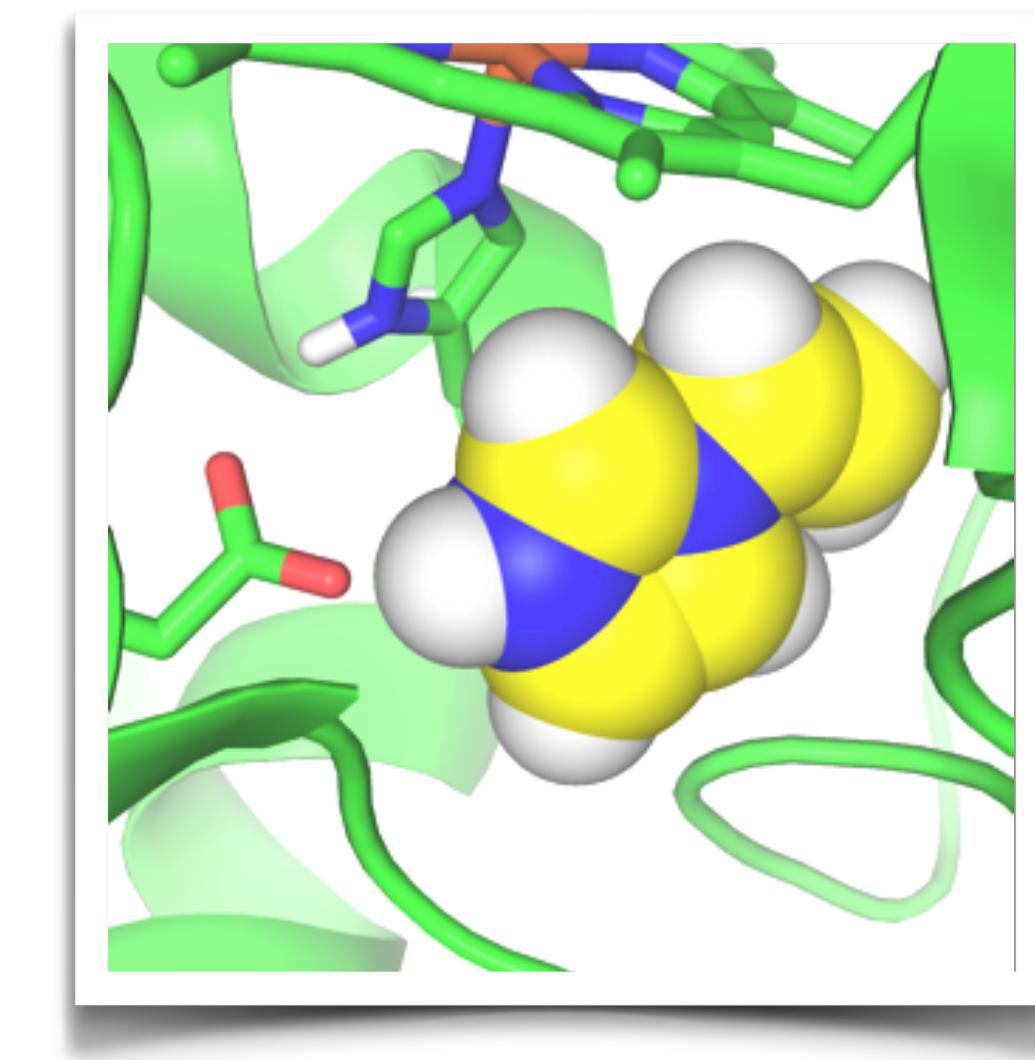
Lysozyme L99A

- Simple
- Nonpolar
- Dry



Lysozyme L99A/M102Q

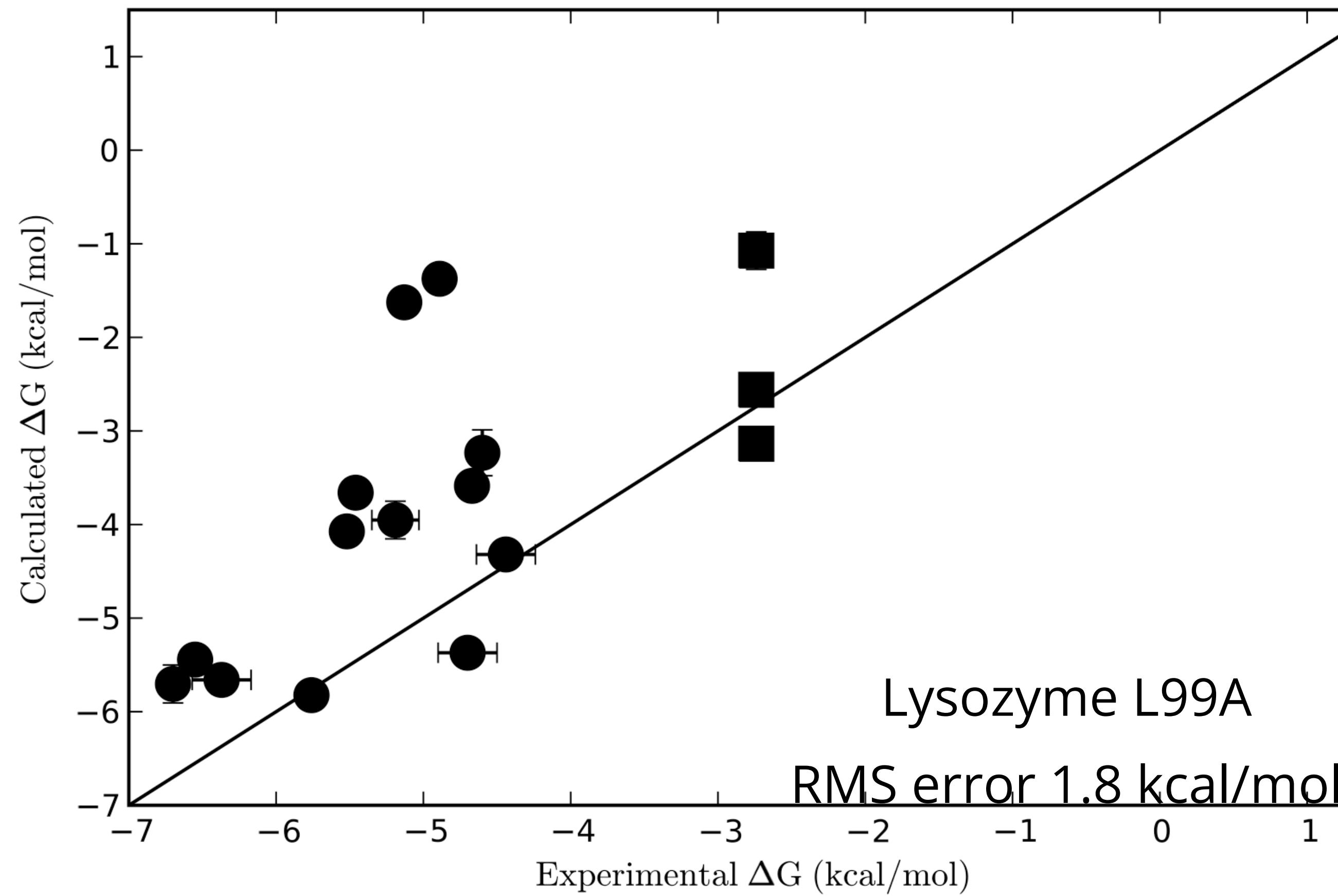
- Simple
- Polar
- Dry
- Additional stable binding modes



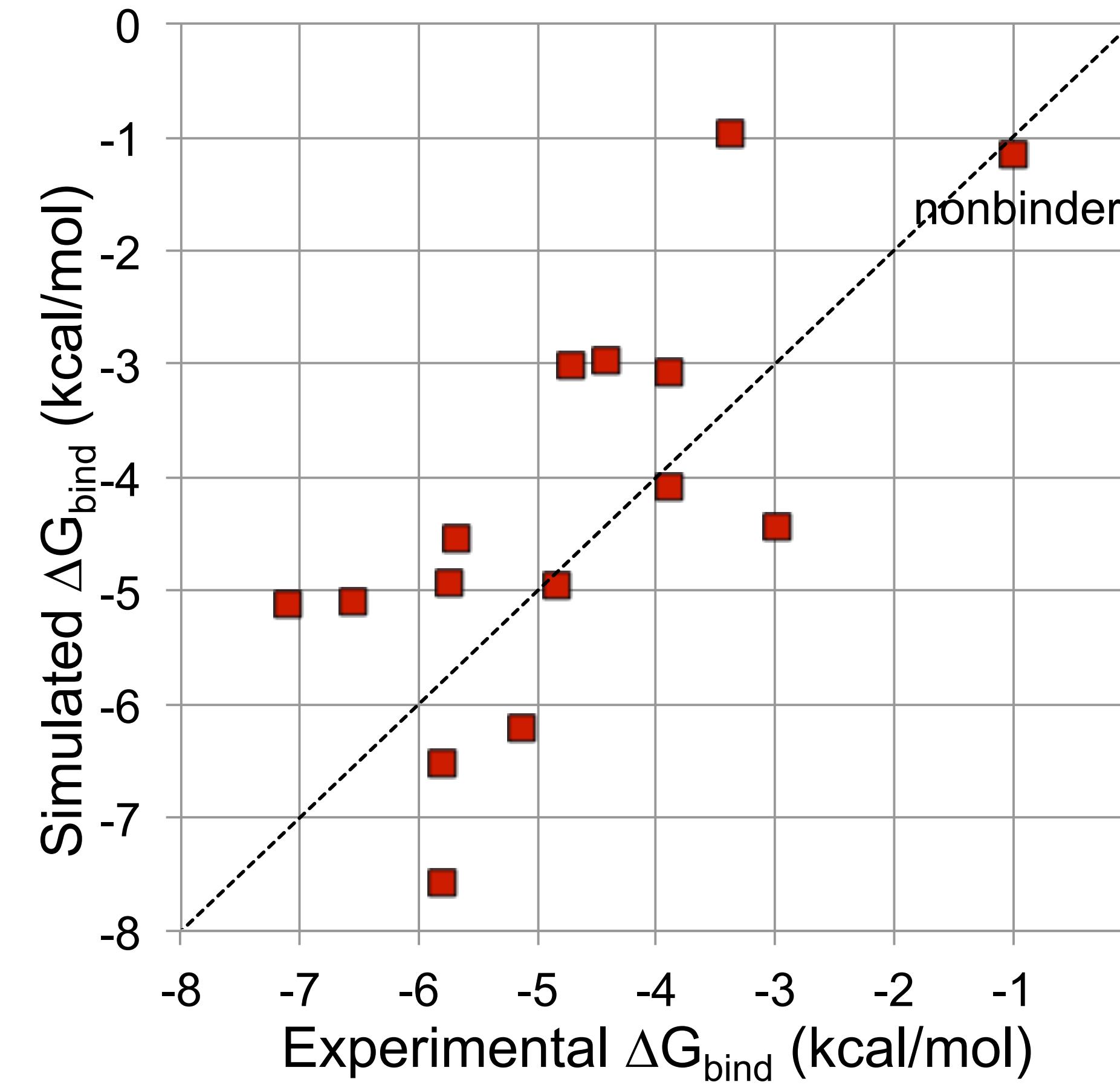
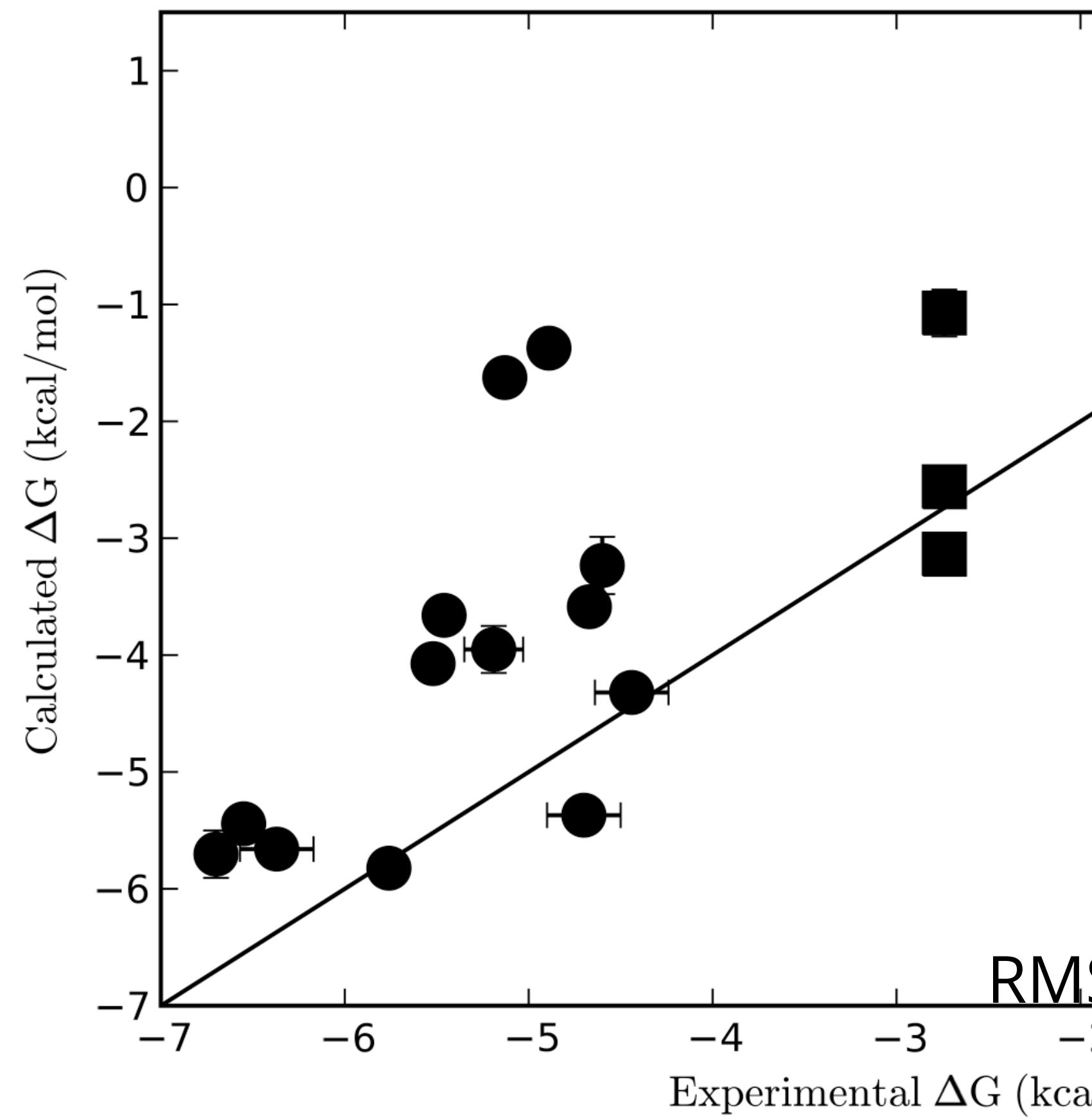
Cytochr. C Peroxidase

- Simple (?)
- Polar, Charged
- Wet
- Additional stable binding modes
- Force field issues?

Absolute free energy calculations on these sites
are tractable and have taught us a great deal



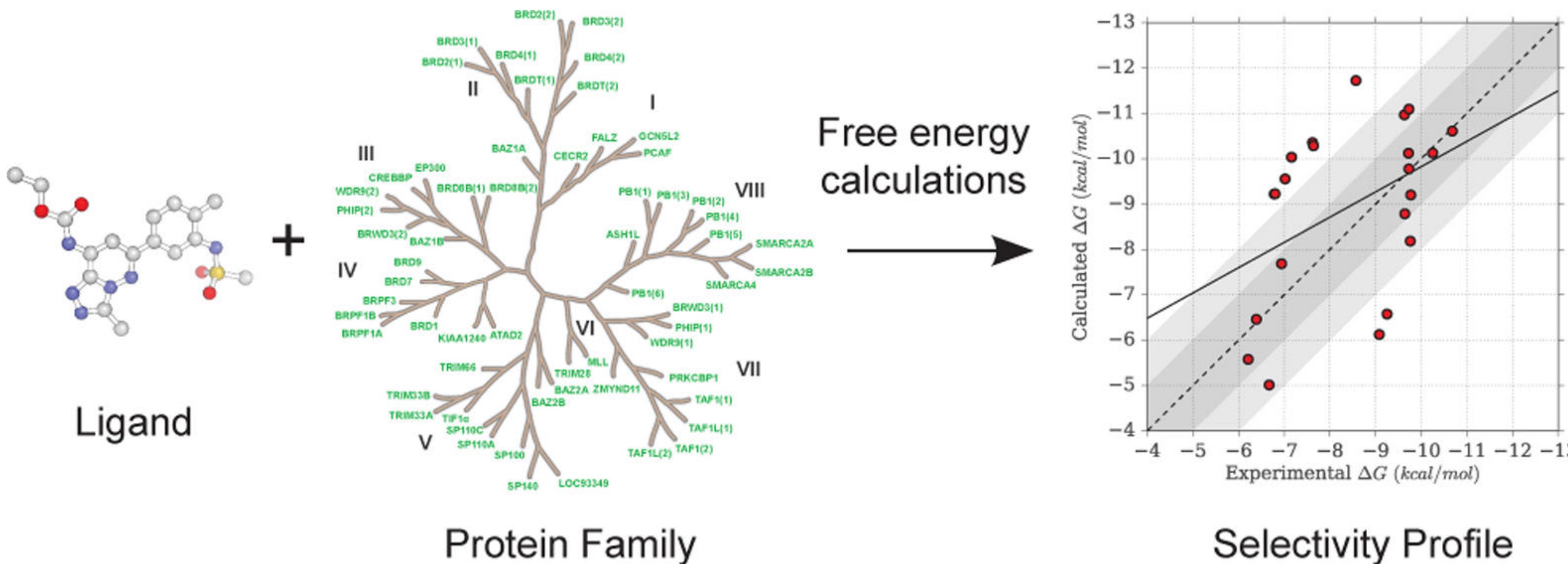
Absolute free energy calculations on these sites
are tractable and have taught us a great deal



Cytochr. C Peroxidase

RMS error 1.4 kcal/mol

The same technology can be used to compute selectivity across a family



Aldeghi et al., "Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations," JACS 2017 <http://dx.doi.org/10.1021/jacs.6b11467>

Focus today is on calculating free energies, especially differences

- Free energy differences between thermodynamic states are very important
 - Protein-ligand binding
 - Addition of particle (computing chemical potential)
 - Changes in temperature
 - Change in potential energy function
 - ...
- Starting off: Perturbation-theory based approach
 - Free Energy Perturbation (FEP)
 - “Zwanzig relation” is my preferred term

The free energy difference is related to the ratio of partition functions

- Free energy difference between states 0 and 1
 - Described by potential energies U_0 and U_1
 - We are taking the system and modifying the energy function
 - For example, turning off or turning on a particle
 - Allows us to move between states
- The free energy difference is related to the ratio of partition functions (here, canonical ensemble):

$$\beta A_1 - \beta A_0 = -\ln \frac{Q_0}{Q_1}$$

This ratio can be rewritten in a simple, useful way

$$\begin{aligned}\beta A_1 - \beta A_0 &= -\ln \frac{Q_0}{Q_1} \\ &= -\ln \frac{\left(\frac{Z_1}{\Lambda(T)^{3N} N!} \right)}{\left(\frac{Z_0}{\Lambda(T)^{3N} N!} \right)} \\ &= -\ln \frac{\int e^{-\beta U_1(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N}\end{aligned}$$

Let's apply the “multiply by 1” trick:

$$\begin{aligned}\beta A_1 - \beta A_0 &= -\ln \frac{\int e^{-\beta U_1(\mathbf{r}^N) + \beta U_0(\mathbf{r}^N) - \beta U_0(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N} \\ &= -\ln \frac{\int e^{-\beta \Delta U(\mathbf{r}^N) - \beta U_0(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N} \quad \text{where} \quad \Delta U(\mathbf{r}^N) = U_1(\mathbf{r}^N) - U_0(\mathbf{r}^N)\end{aligned}$$

This looks kind of like the probability distribution for state 0:

$$\wp_0(\mathbf{r}^N) = \frac{e^{-\beta U_0(\mathbf{r}^N)}}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N} \quad , \text{ so making this substitution:} \quad \beta A_1 - \beta A_0 = -\ln \int \wp_0(\mathbf{r}^N) e^{-\beta \Delta U(\mathbf{r}^N)} d\mathbf{r}^N$$

We usually write this as an ensemble average

Instead of writing $\beta A_1 - \beta A_0 = -\ln \int \wp_0(\mathbf{r}^N) e^{-\beta \Delta U(\mathbf{r}^N)} d\mathbf{r}^N$

We usually write $\beta A_1 - \beta A_0 = -\ln \langle e^{-\beta \Delta U} \rangle_0$

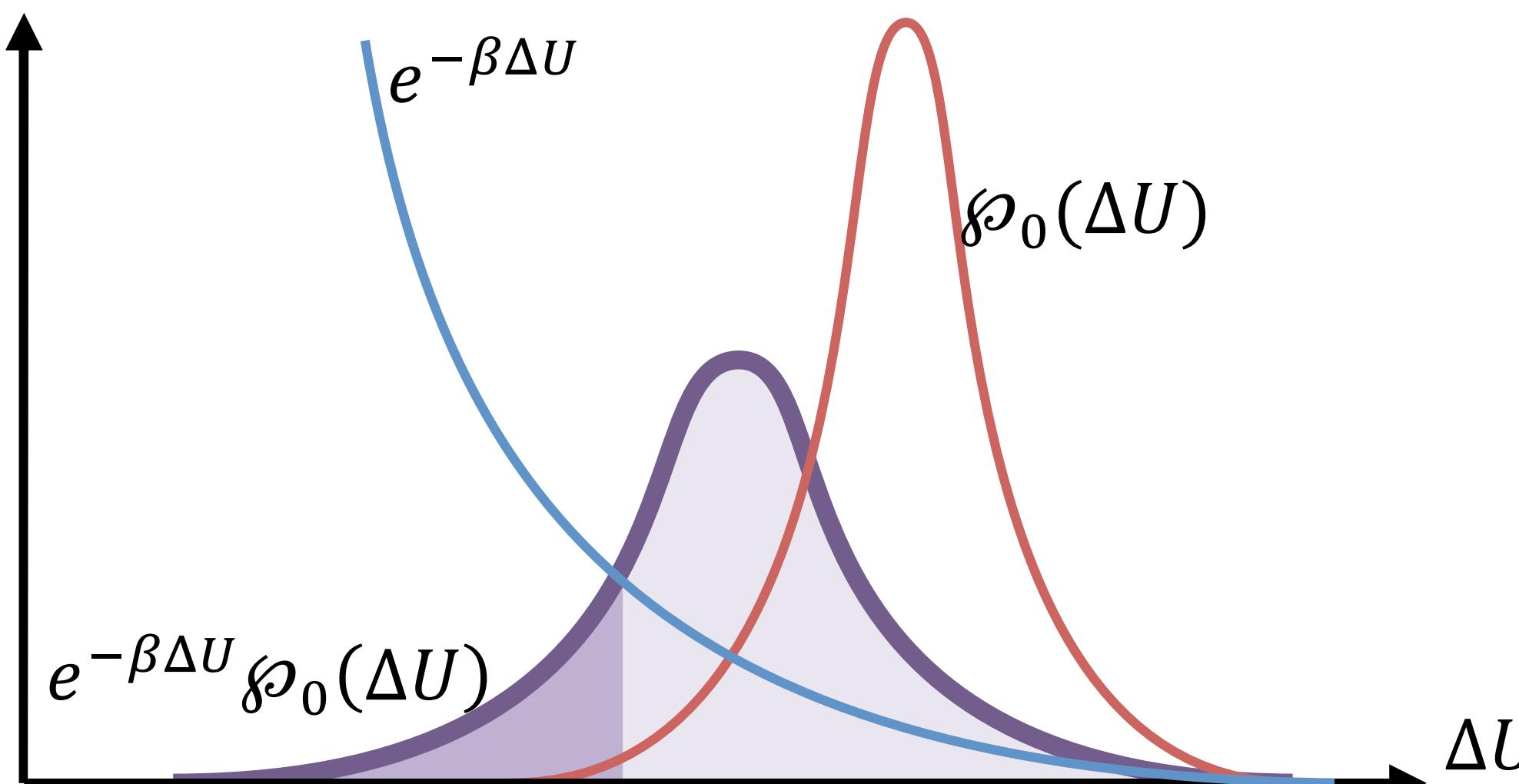
It is also true that $\beta A_1 - \beta A_0 = \ln \langle e^{\beta \Delta U} \rangle_1$

- The above, the Zwanzig relation, is important:
 - The free energy difference can be computed based on an average over configurations taken from one of the states of interest
 - We can generate these configurations with MC or MD
 - The free energy comes from evaluating the energies of these configurations in both potentials U_0 and U_1 , and taking an appropriate average of the energy difference

As usual, distribution tails can lead to convergence problems

- We can rewrite the expression in a way that informs about the likely errors:

$$\begin{aligned}\beta A_1 - \beta A_0 &= -\ln\langle e^{-\beta\Delta U}\rangle_0 \\ &= -\ln \int e^{-\beta\Delta U} \wp_0(\Delta U) d\Delta U\end{aligned}$$



This places a practical limit on the types of changes that can be done

- The Zwanzig relations sensitivity to the tails of the distribution means only very small perturbations can be done reliably
 - i.e. where ΔU near 0
 - Can be achieved by introducing many intermediate states, but better alternatives are available
- One other aspect of Zwanzig relation: One can, from two simulations, obtain two different estimates of the free energy difference
 - Presumably the best free energy estimate is some combination of these
 - But what combination?

The Bennett acceptance ratio minimizes the expected error in the free energy difference

- Bennett's approach modified the free energy equation:

$$\begin{aligned}\beta A_1 - \beta A_0 &= -\ln \frac{\int e^{-\beta U_1(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N} \\ &= -\ln \left(\frac{\int e^{-\beta U_1(\mathbf{r}^N)} d\mathbf{r}^N}{\int w(\mathbf{r}^N) e^{-\beta U_0(\mathbf{r}^N) - \beta U_1(\mathbf{r}^N)} d\mathbf{r}^N} \frac{\int w(\mathbf{r}^N) e^{-\beta U_0(\mathbf{r}^N) - \beta U_1(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U_0(\mathbf{r}^N)} d\mathbf{r}^N} \right) \\ &= -\ln \frac{\langle we^{-\beta U_1} \rangle_0}{\langle we^{-\beta U_0} \rangle_1}\end{aligned}$$

where w is some arbitrary weighting function that will be chosen to reduce the error

- Bennett minimized the error and found:

$$w(\mathbf{r}^N) \propto (n_0^{-1} e^{-\beta A_0 - \beta U_1(\mathbf{r}^N)} + n_1^{-1} e^{-\beta A_1 - \beta U_0(\mathbf{r}^N)})^{-1}$$

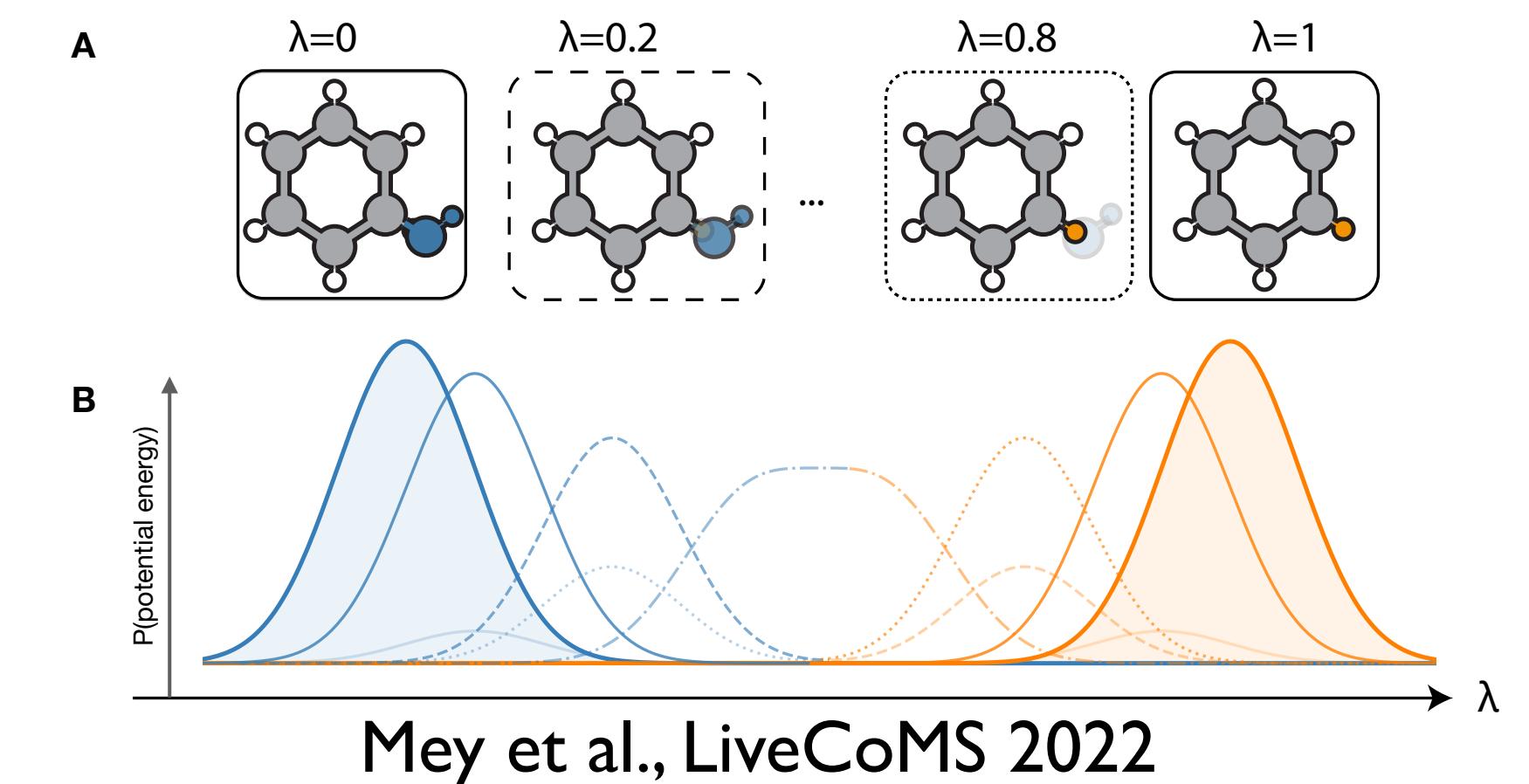
where n_0 and n_1 are the number of trajectory configurations used in the averages
(see Frenkel and Smit for derivation)

The Bennett acceptance ratio (BAR) gives a simple expression for the free energy difference

- The final result, with the weighting function:

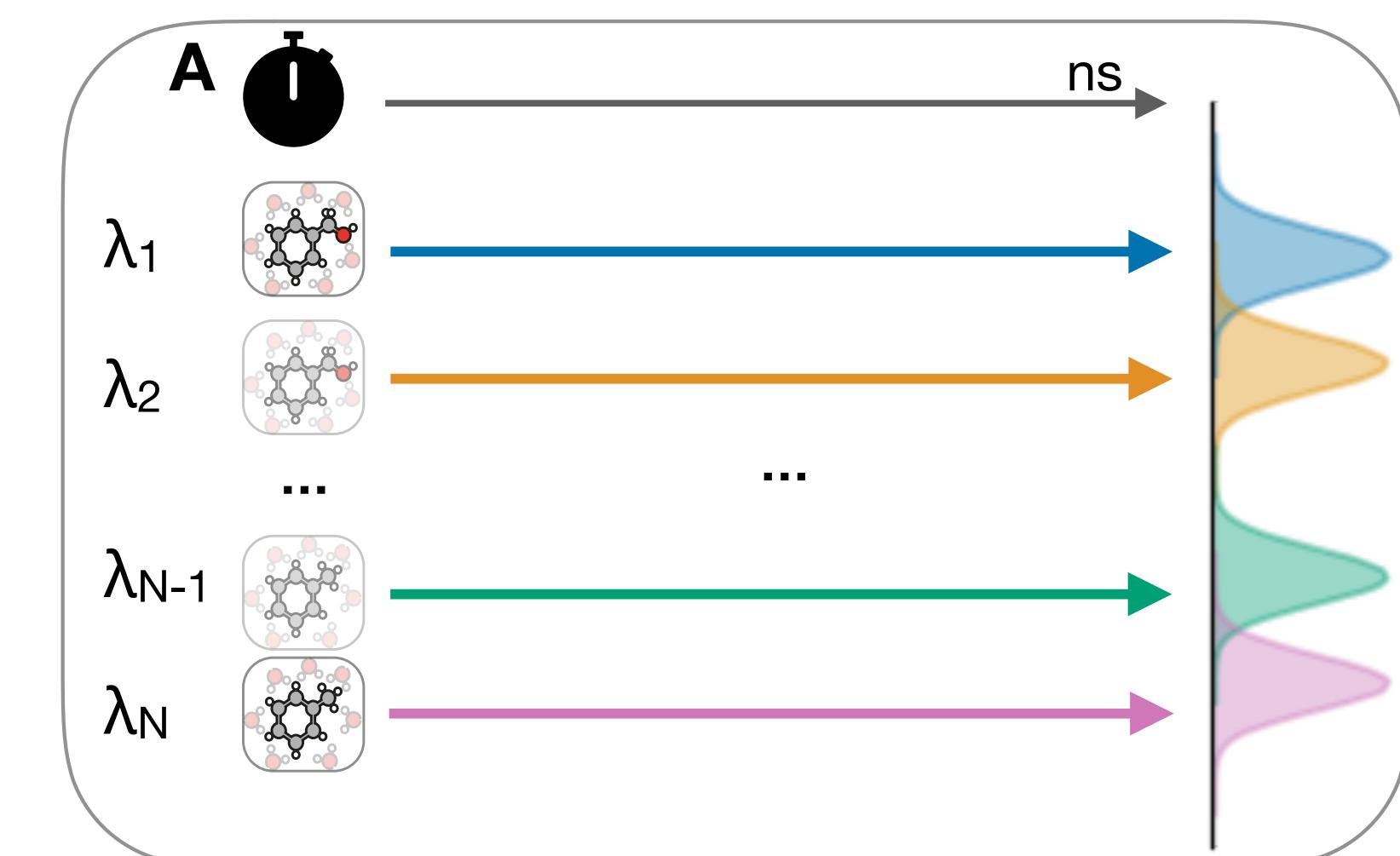
$$\beta\Delta A = \ln \frac{\left\langle \frac{1}{1 + e^{-\beta\Delta U + \beta\Delta A}} \right\rangle_0}{\left\langle \frac{1}{1 + e^{\beta\Delta U - \beta\Delta A}} \right\rangle_1}$$

- This is solved self-consistently
- Additional nuances:
 - Now we need the energies of each configuration and can't evaluate the averages on the fly; often trajectories are just stored and energies evaluated later
 - ΔU is $U_1 - U_0$ for state 0, $U_0 - U_1$ for state 1



Still, if the states are dissimilar, we may need intermediate states: alchemical states

- If the potentials are dissimilar enough, overlap may be poor
- Important regions for one potential are unimportant for the other, and vice versa
- Introduce intermediate states, often using a scaling parameter:
$$U = (1 - \lambda)U_0 + \lambda U_1$$
- λ runs from 0 to 1 and we run intermediate simulations
 - Compute pairwise free energy differences, and the total free energy difference is the sum



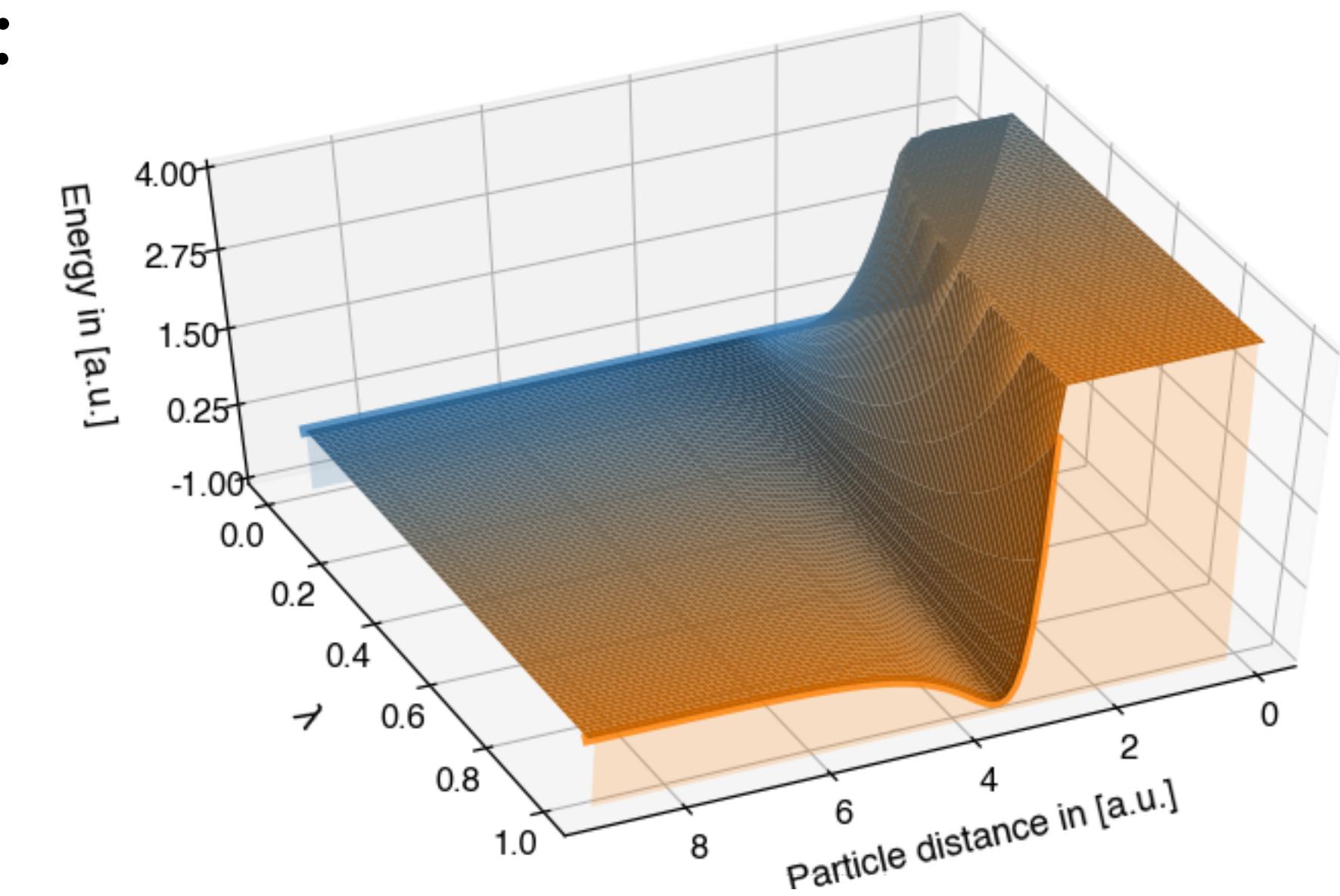
Mey et al., LiveCoMS 2022

For insertions or deletions of particles, another functional form is preferable

- Lennard-Jones interactions may be scaled with:

$$V_{\text{"softcore" vDW}} = 4\epsilon(1 - \lambda) \left[\frac{1}{[\alpha\lambda + (r/\sigma)^6]^2} - \frac{1}{\alpha\lambda + (r/\sigma)^6} \right]$$

- Avoids large forces and numerical instabilities associated with deletions involving linear scaling



Mey et al., LiveCoMS 2022

Computing the chemical potential can be simple

- Chemical potential is $\mu = \left(\frac{\partial A}{\partial N}\right)_{T,V}$

Really, since the particle number is discrete:

$$\mu = A(T, V, N + 1) - A(T, V, N)$$

- Using the connection with the partition function:

$$\beta\mu = \mu_{ig} + \ln \frac{V \int e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U(\mathbf{r}^{N+1})} d\mathbf{r}^{N+1}}$$

Where μ_{ig} is an ideal gas component; we are interested in the excess (not including this)

- Letting the top integral contain a “dummy particle”:

$$\beta\mu_{ex} = \ln \frac{\int e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^{N+1}}{\int e^{-\beta U(\mathbf{r}^{N+1})} d\mathbf{r}^{N+1}}$$

- We can do this by just turning on the interactions of the dummy particle

$$\beta\mu_{ex} = -\ln \langle e^{-\beta\Delta U} \rangle_0$$

“Widom insertion” applies this for insertions of particles

- How to actually make this work?
 - Widom proposed insertion:
 - Run a simulation with N particles
 - Periodically try inserting a particle at a random location
 - Compute the change in energy due to the insertion
 - Remove the particle and continue running the simulation
 - Evaluate $\langle e^{-\beta \Delta U} \rangle$ over the simulation and use it to get the chemical potential
 - This works well if the particle is small and/or system is not very dense
 - Fails for highly directional interactions

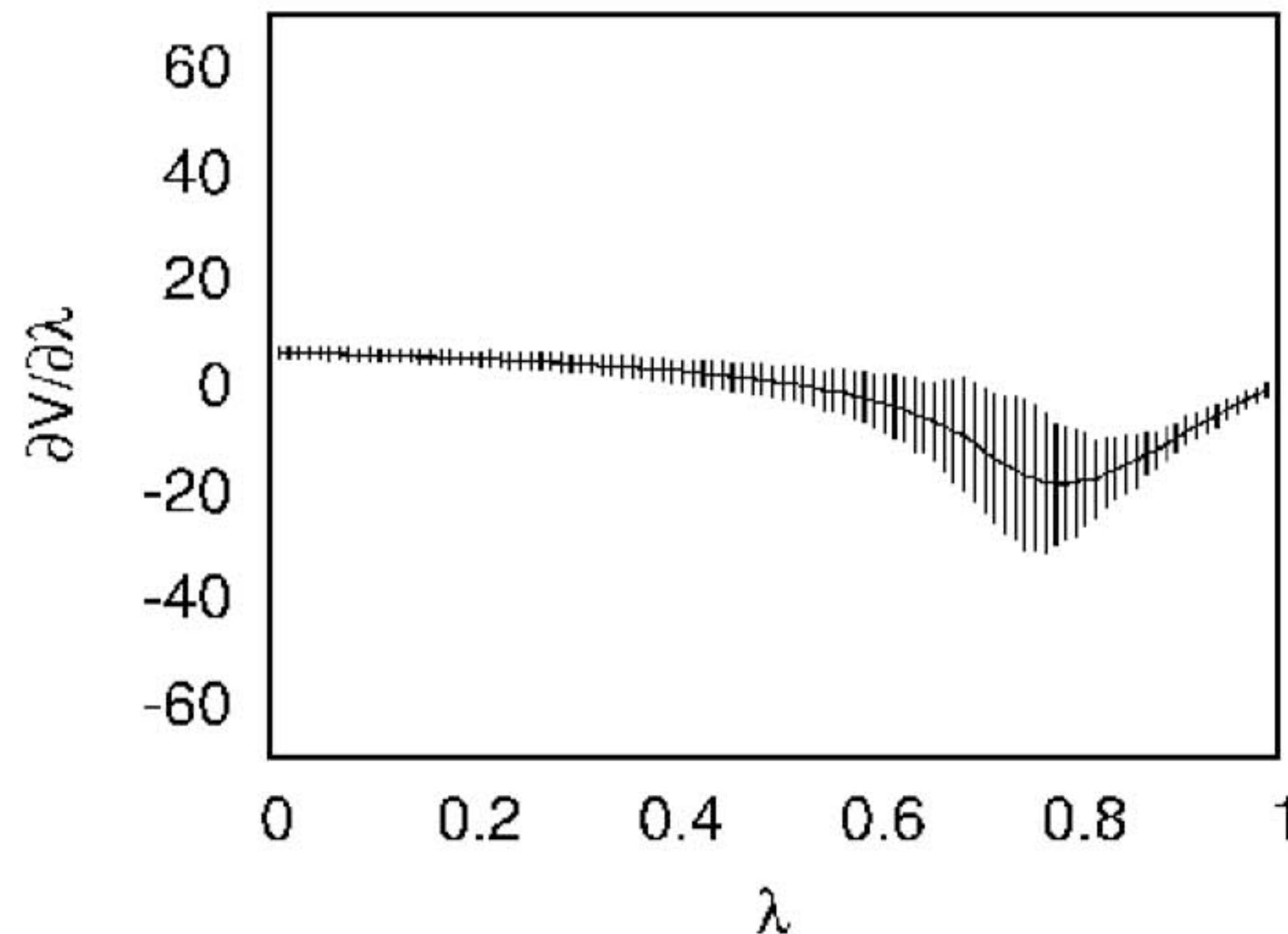
Thermodynamic integration is an alternate approach to free energy calculations

- Idea: Sometimes it is easy to compute the derivative of the free energy with respect to some parameter
 - Suppose U is also a function of λ :

$$A = -k_B T \ln \frac{1}{N! \Lambda(T)^{3N}} \int e^{-\beta U(\mathbf{r}^N; \lambda)} d\mathbf{r}^N$$

- Let's take the derivative:
$$\begin{aligned}\frac{dA}{d\lambda} &= -k_B T \frac{d}{d\lambda} \ln \int e^{-\beta U(\mathbf{r}^N; \lambda)} d\mathbf{r}^N \\ &= -k_B T \frac{\int -\beta \left(\frac{dU}{d\lambda} \right) e^{-\beta U(\mathbf{r}^N; \lambda)} d\mathbf{r}^N}{\int e^{-\beta U(\mathbf{r}^N; \lambda)} d\mathbf{r}^N} \\ &= \left\langle \frac{dU}{d\lambda} \right\rangle_\lambda\end{aligned}$$
- Now, we can integrate to get the free energy difference $A(\lambda_1) - A(\lambda_0) = \int_{\lambda_0}^{\lambda_1} \left\langle \frac{dU}{d\lambda} \right\rangle_\lambda d\lambda$

Thermodynamic integration (TI) involves numerical integration

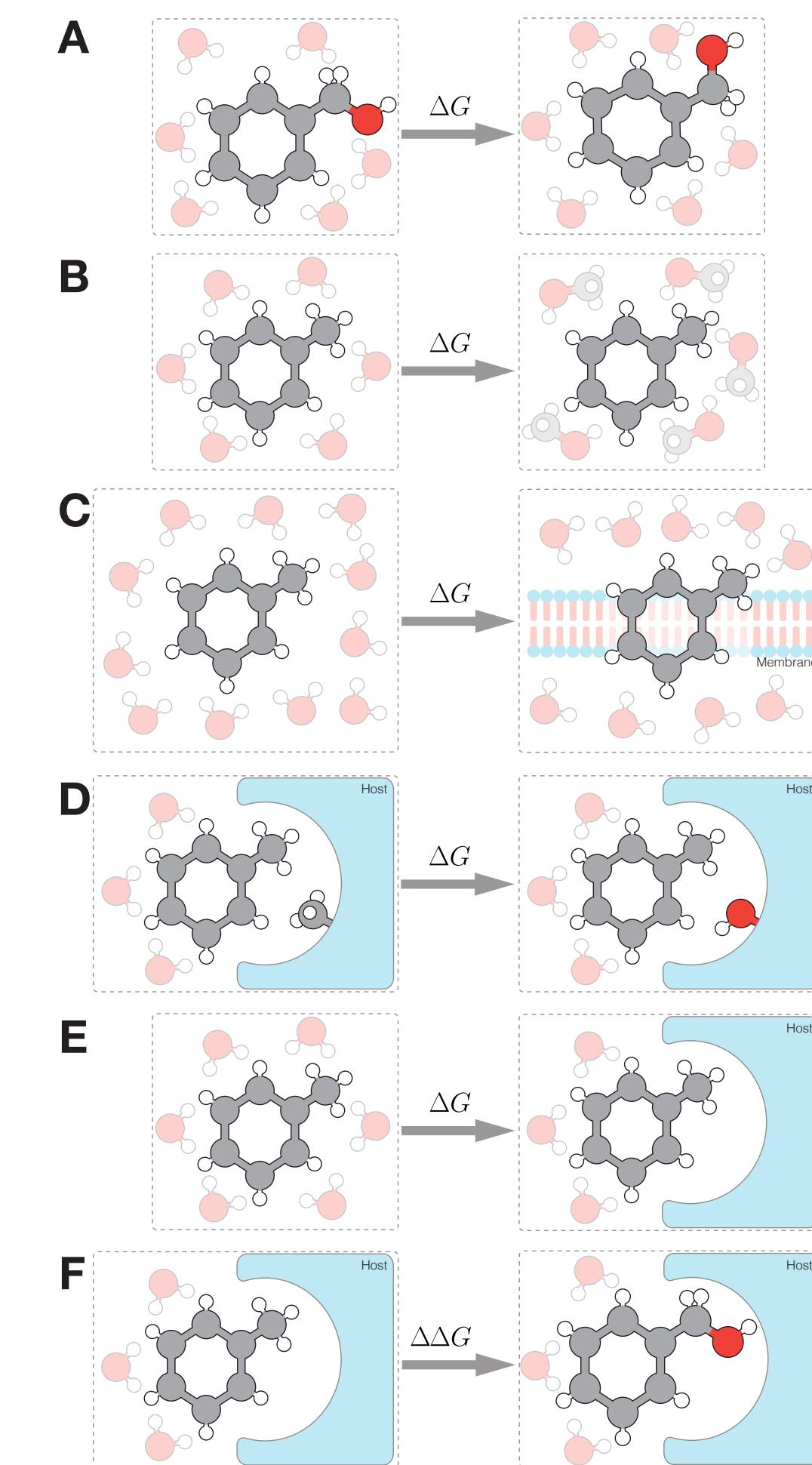


(Error bars enlarged by a factor of 30 for visibility)

Data from Steinbrecher, Mobley, Case (JCP 2007)

Overall, BAR (or multistate generalization, MBAR) should be preferred

- Provides minimum-variance free energy estimates
 - Can be derived from maximum-likelihood methods
 - Statistically optimal use of the available data
- Zwanzig relation is useful for computing nontraditional free energy differences, when intermediate states are impossible or not easily constructed
 - Free energy differences between force fields
 - Free energy difference of switching to a QM approach
 - Between different solvent models



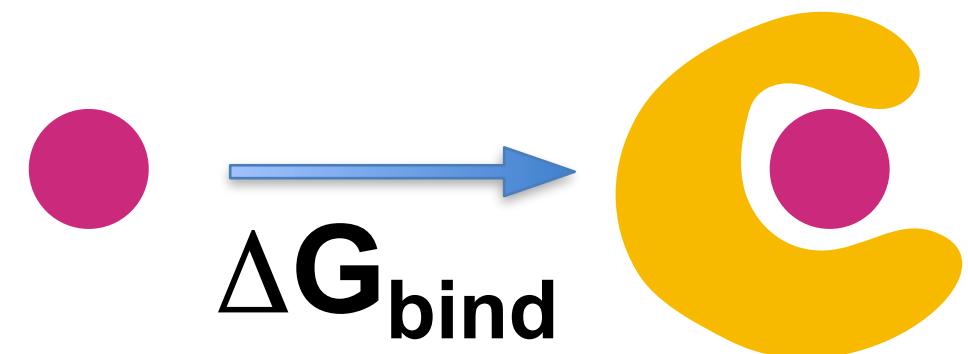
A practical note on why free energy calculations can work

- It is often not possible to calculate the average enthalpy of a reasonably large (protein+water) system with ~kcal/mol accuracy
- Yet free energy calculations can get to higher accuracy than enthalpy differences
 - This is because fluctuations in the system which lead to noise in the potential energy cancel out
 - Anything not affecting the perturbed region will cancel out when taking the energy differences or derivatives that go into the averages

Absolute free energies can be calculated relative to a known reference state

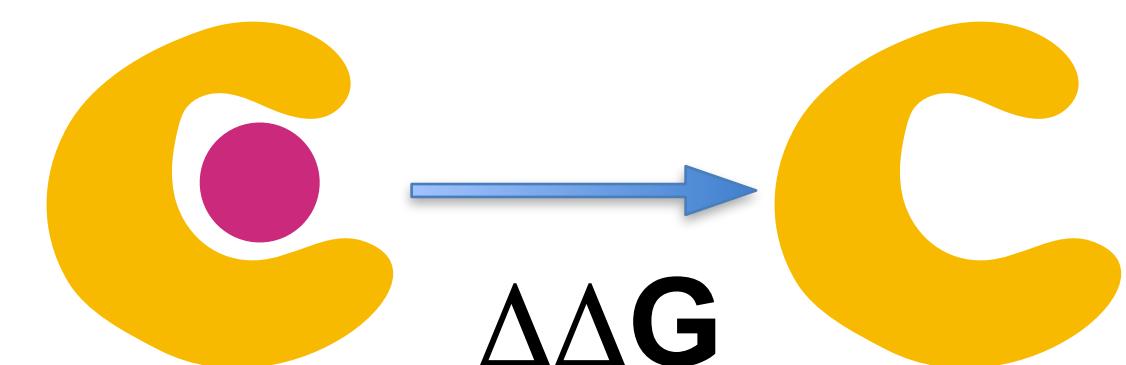
- We cannot typically directly calculate the absolute free energy of a system
 - But we can connect the system to reference states for which the free energy can be calculated analytically
 - For a liquid/gas, can transform the system to an ideal gas reference state
 - Compute the free energy of the transformation
 - Free energy of an ideal gas is known
 - For a solid, transform it to an ideal harmonic (Einstein) crystal
 - where the atoms are not interacting and are harmonically restrained to their positions
 - Again, free energy known

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



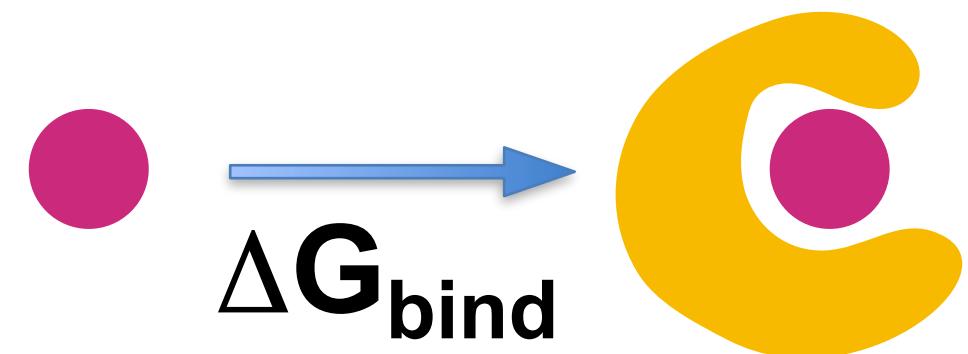
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



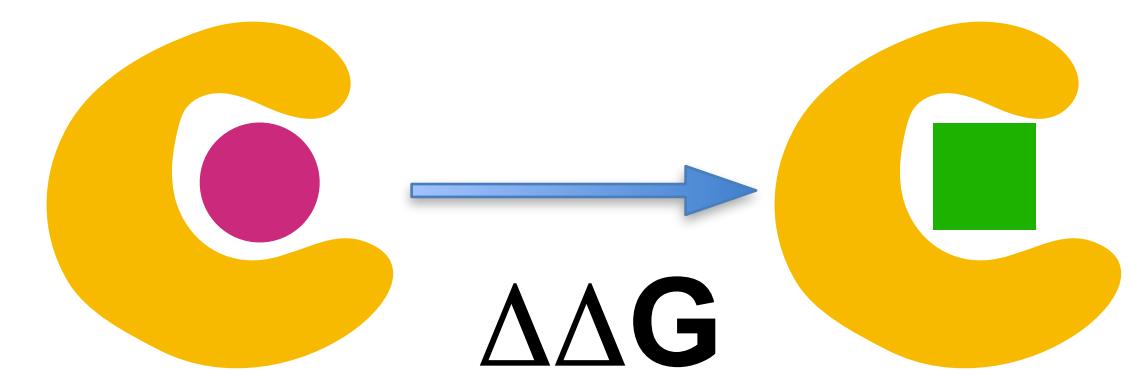
Enhanced sampling methods for binding free energies

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



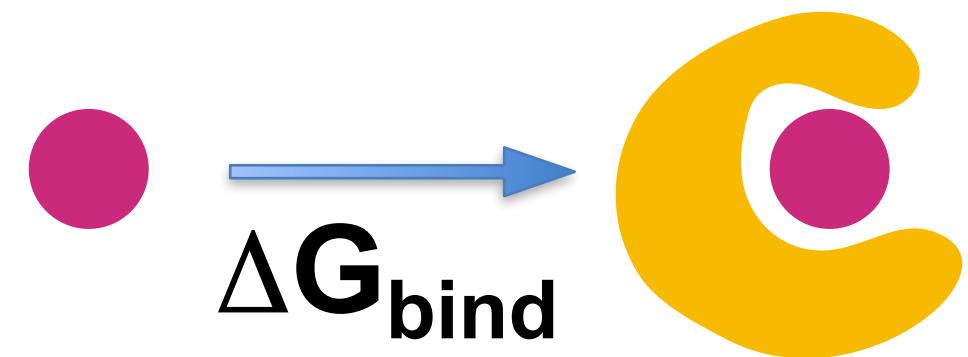
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



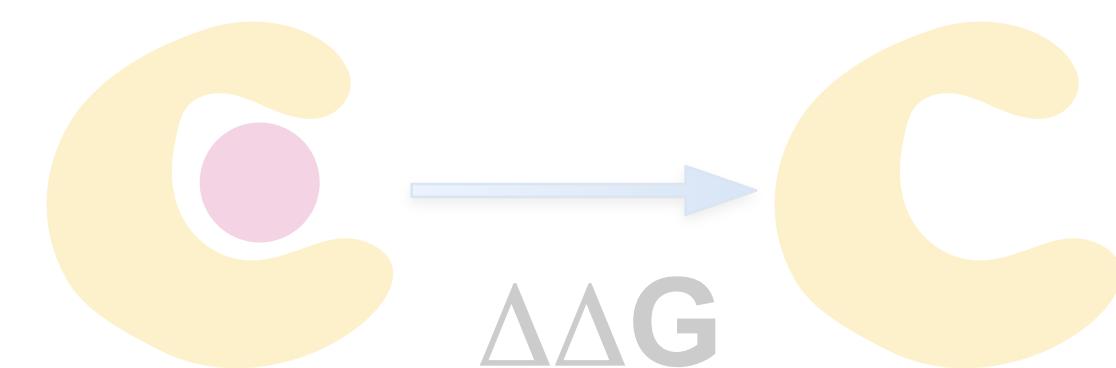
Enhanced sampling methods for binding free energies

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



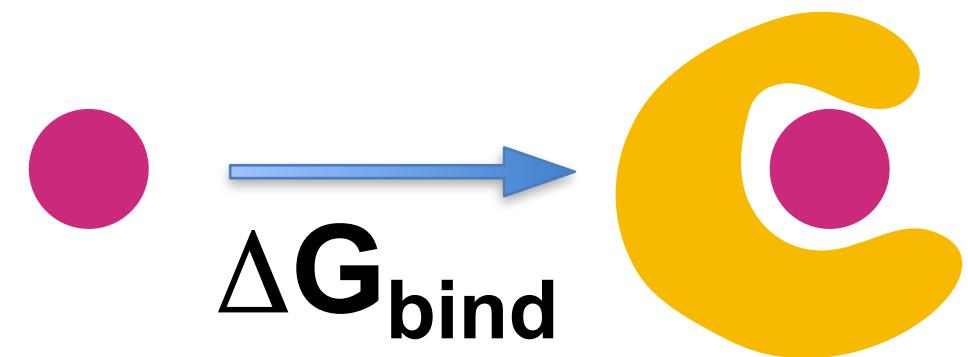
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



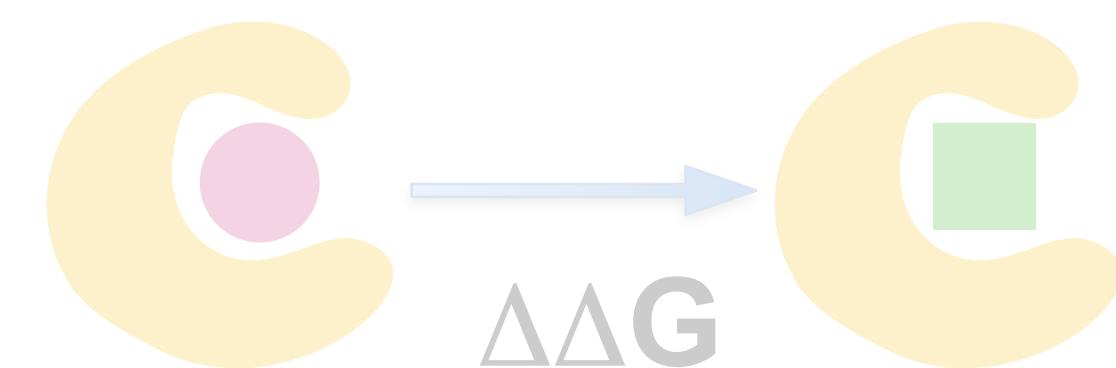
Enhanced sampling methods for binding free energies

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



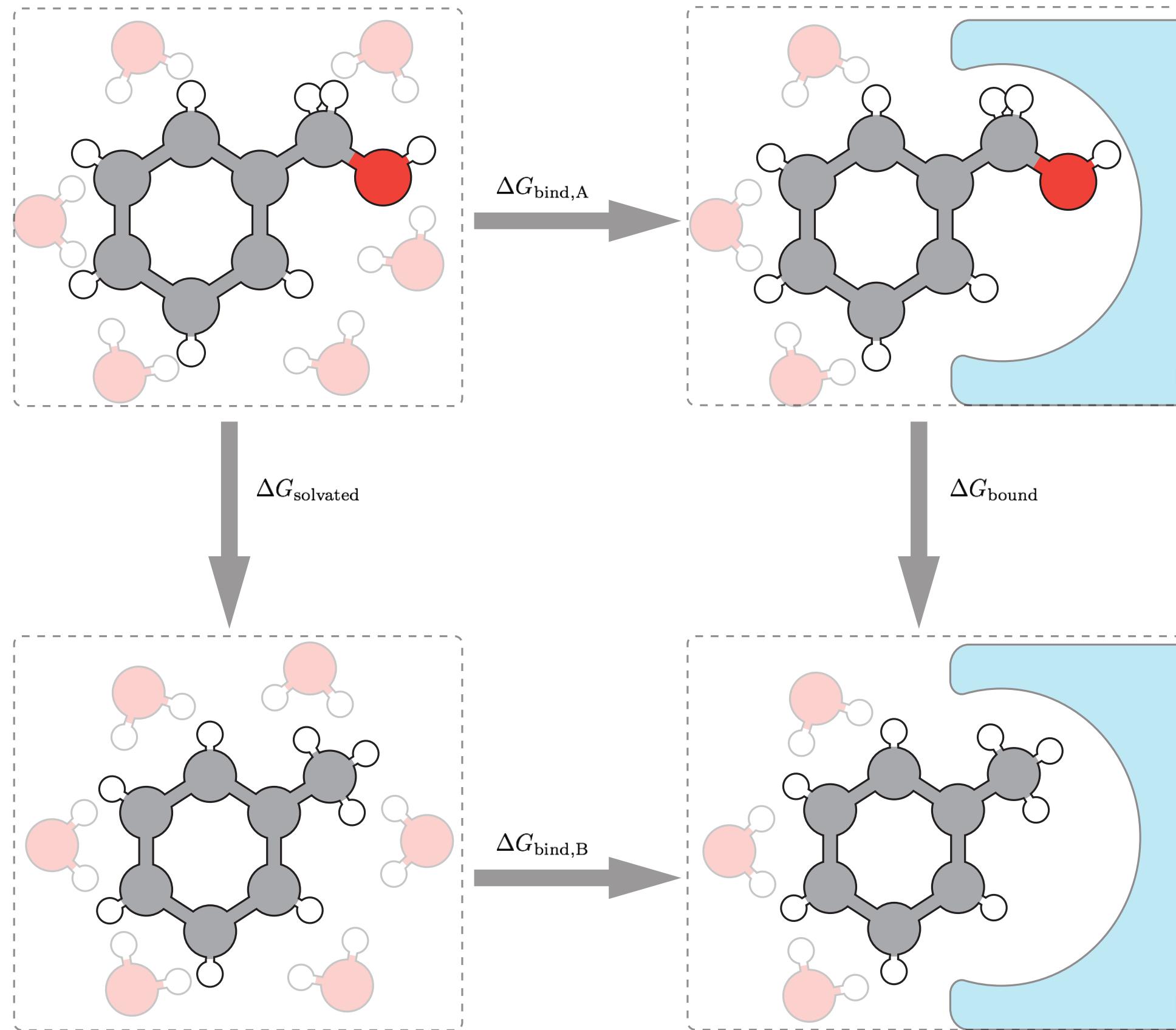
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods

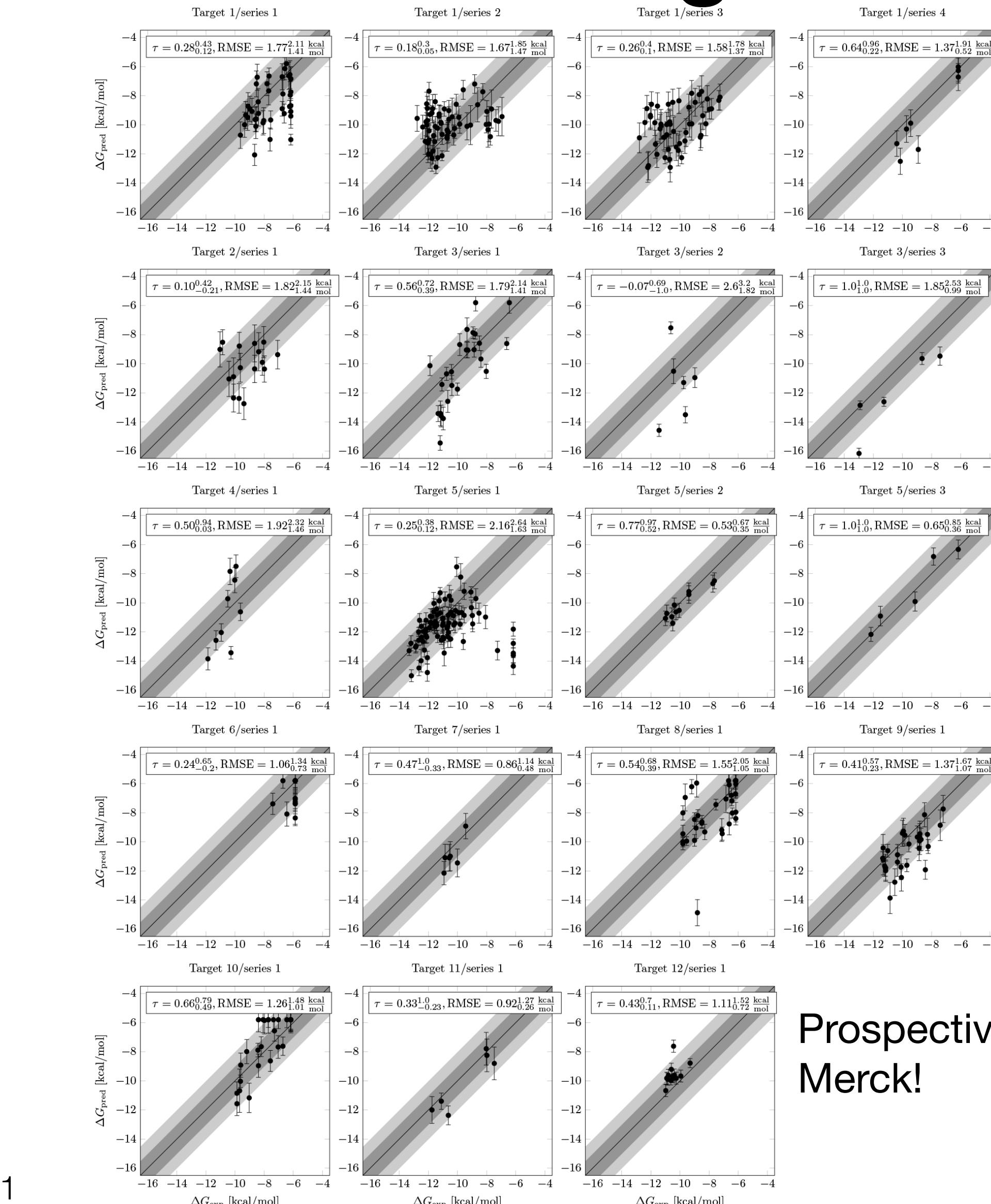


Enhanced sampling methods for binding free energies

Free energy calculations now see broad industry adoption in structure-based design

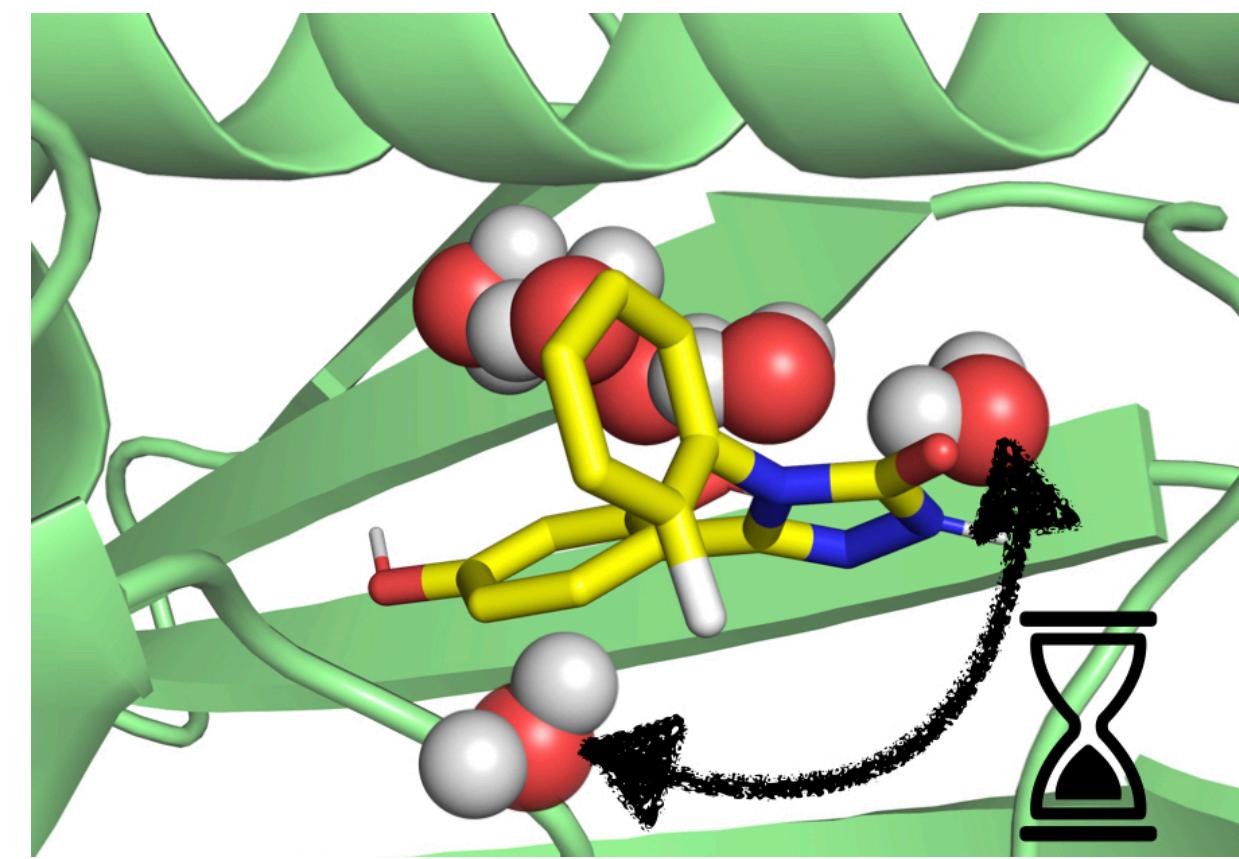
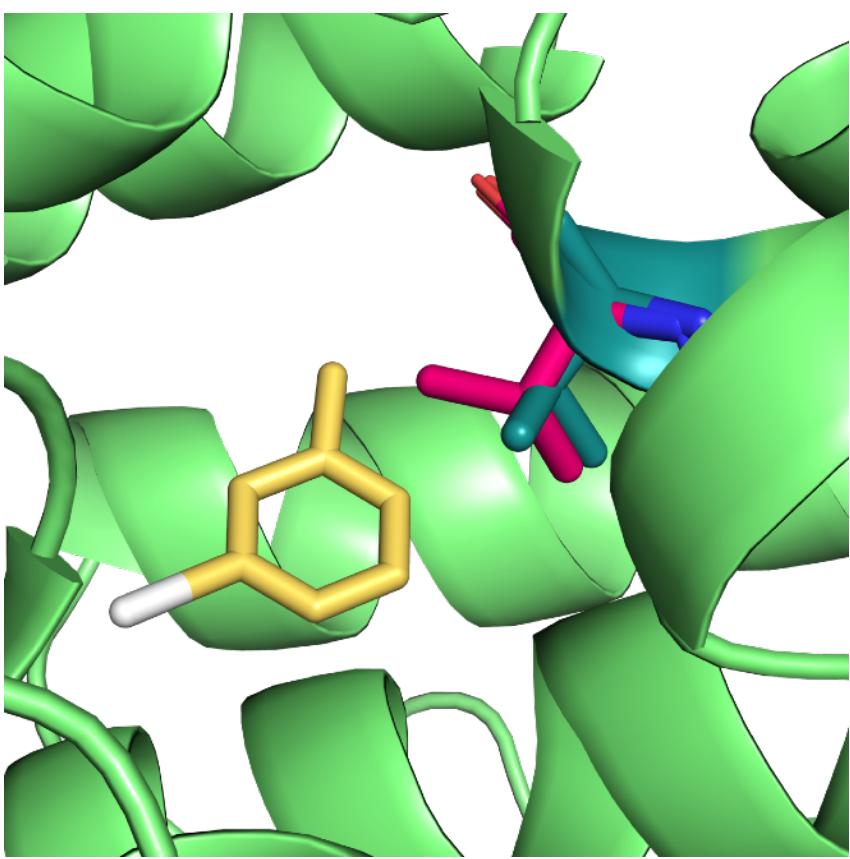


10.33011/livecoms.2.1.18378
10.1021/acs.jcim.0c00900

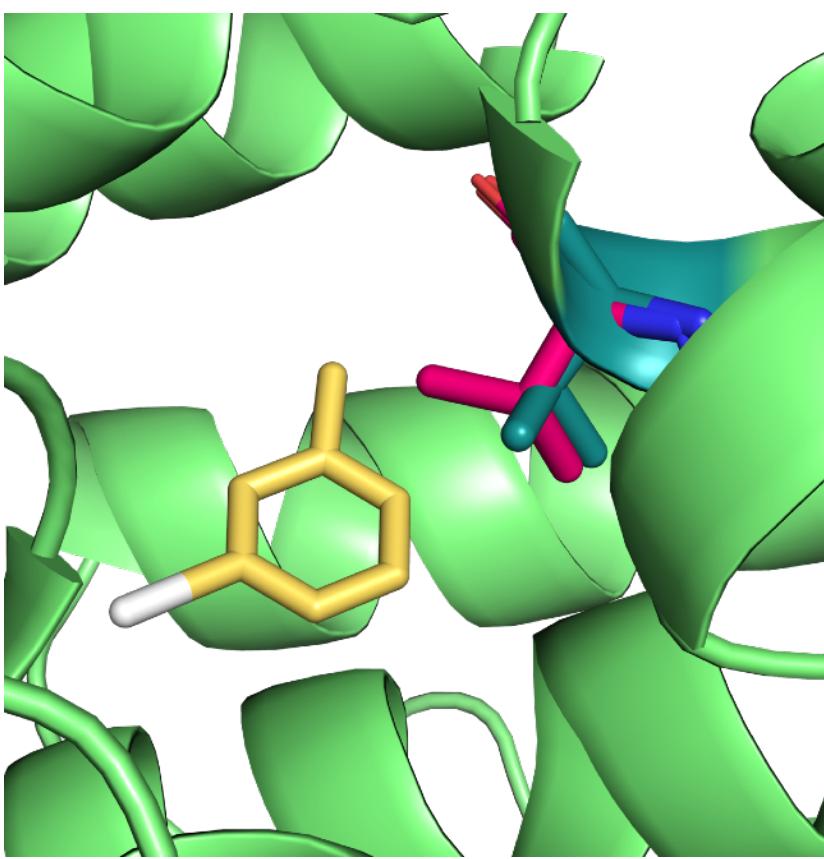


Prospective work from
Merck!

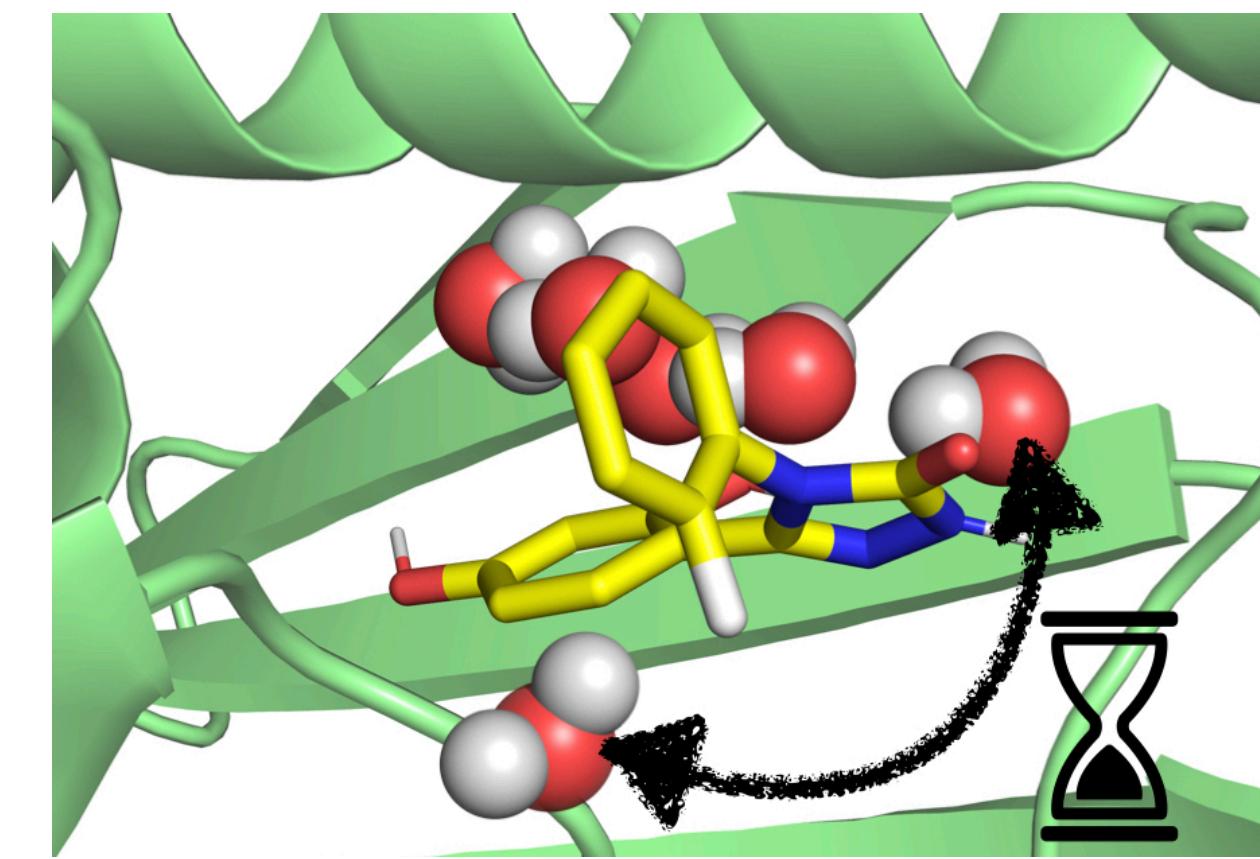
Free energy calculations are challenging



Free energy calculations are challenging



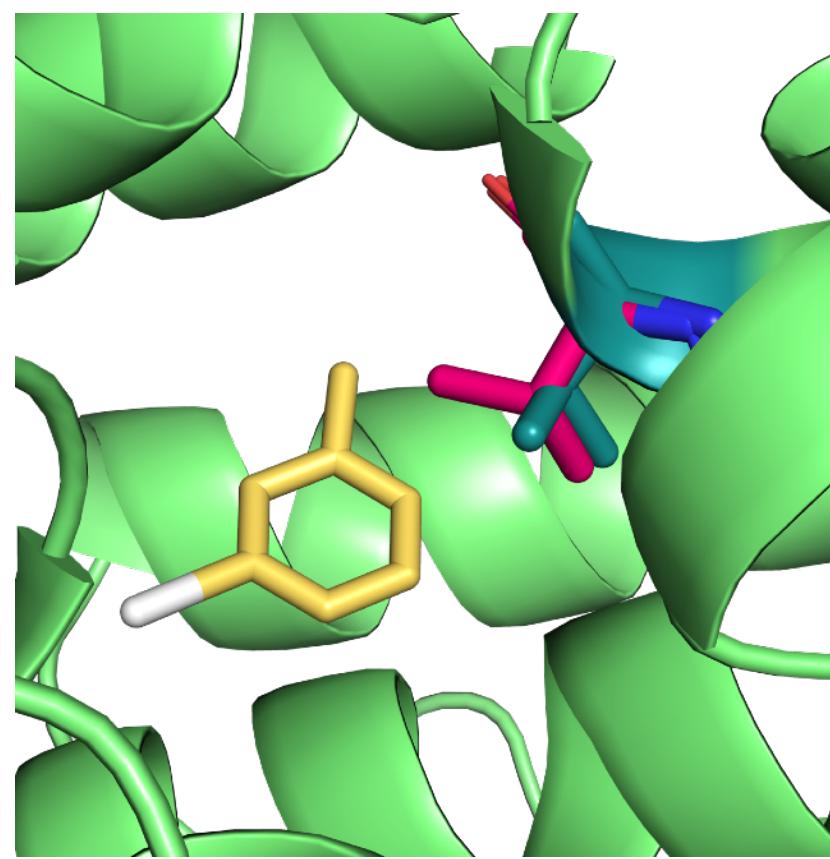
Slow protein conformational changes upon binding



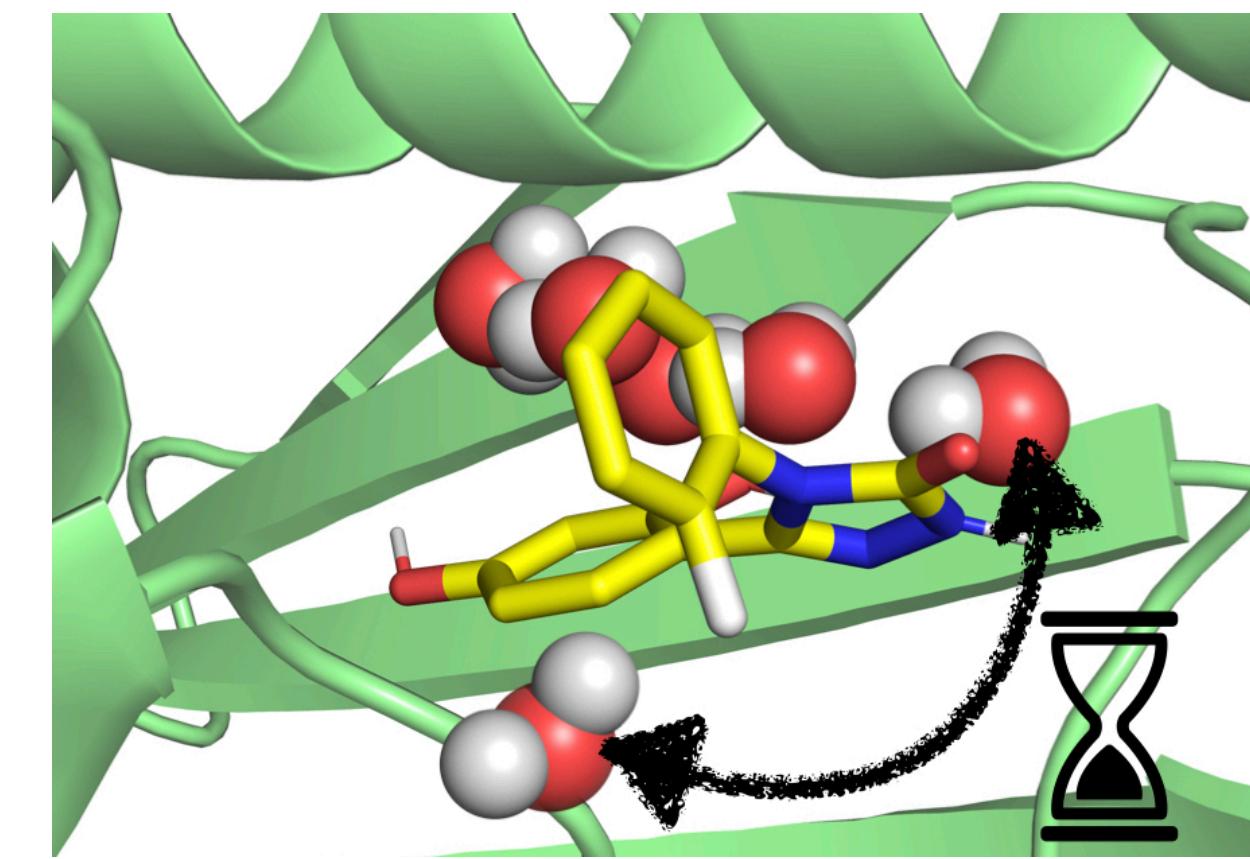
Buried waters in the binding site

How do we address these issues?

Free energy calculations are challenging



Slow protein conformational changes upon binding

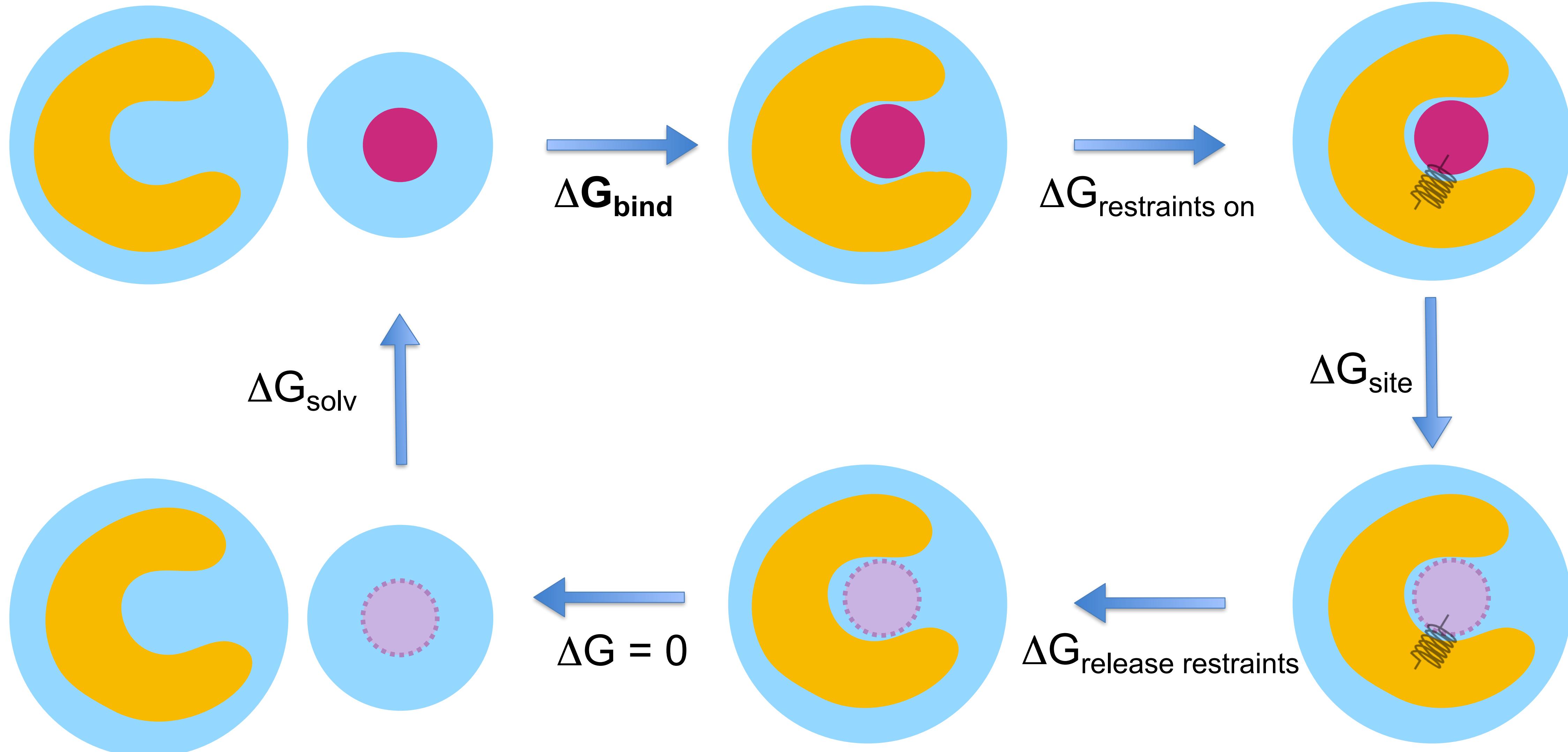


Buried waters in the binding site

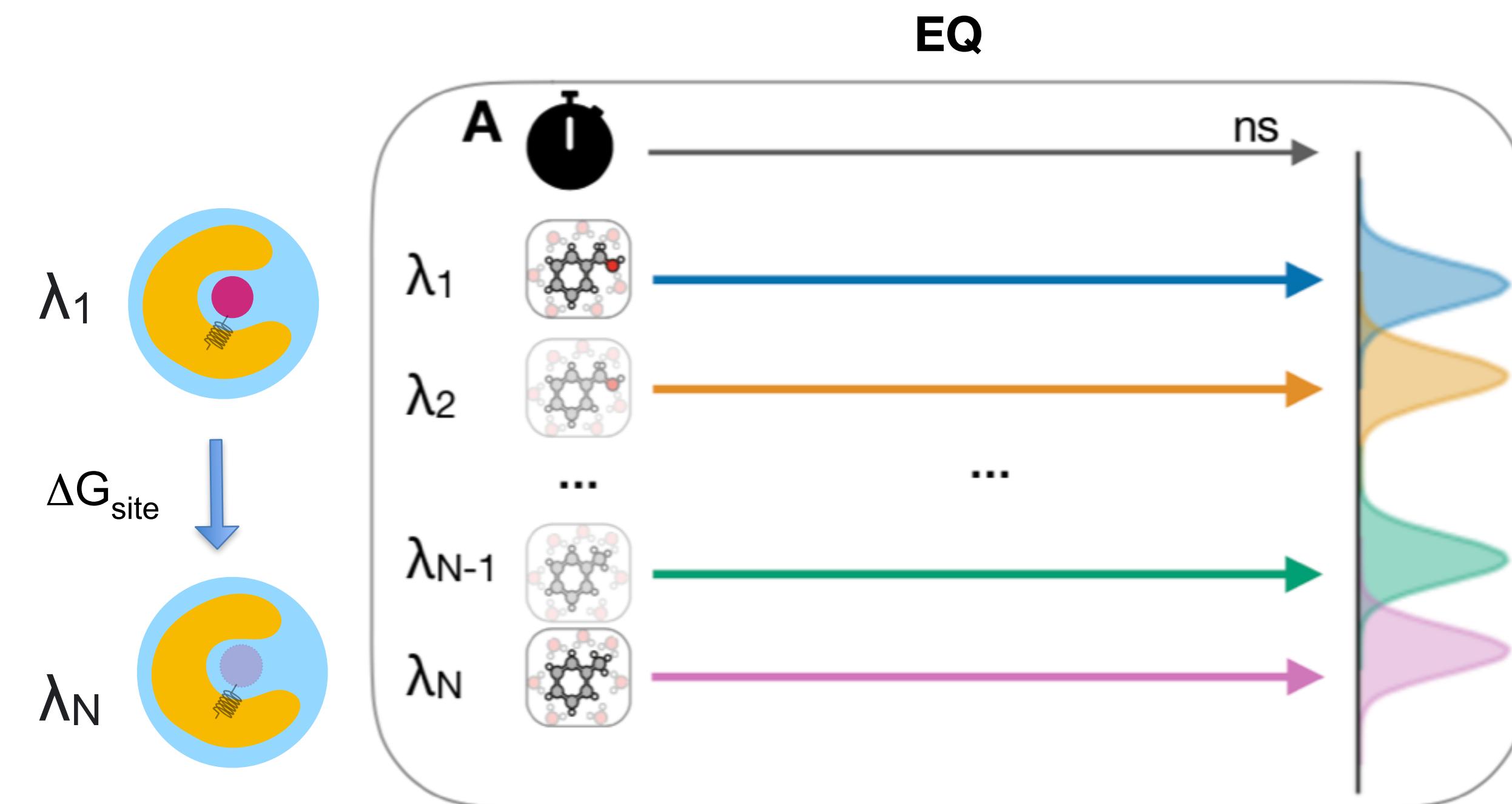
How do we address these issues?

How do we detect the problems in the first place?

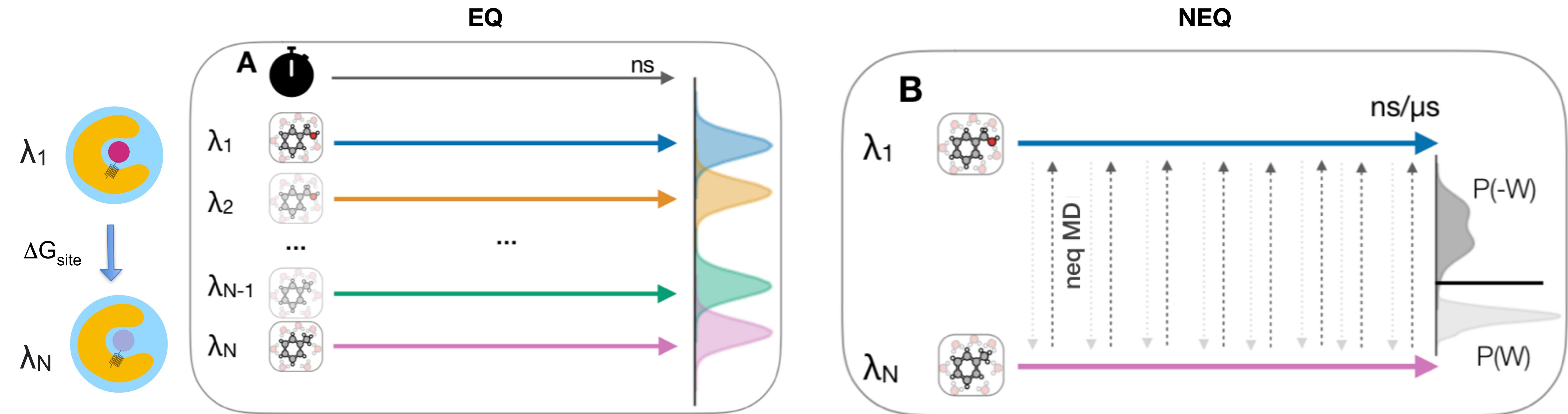
Binding free energies can be obtained through a thermodynamic cycle



We compared equilibrium and non-equilibrium approaches



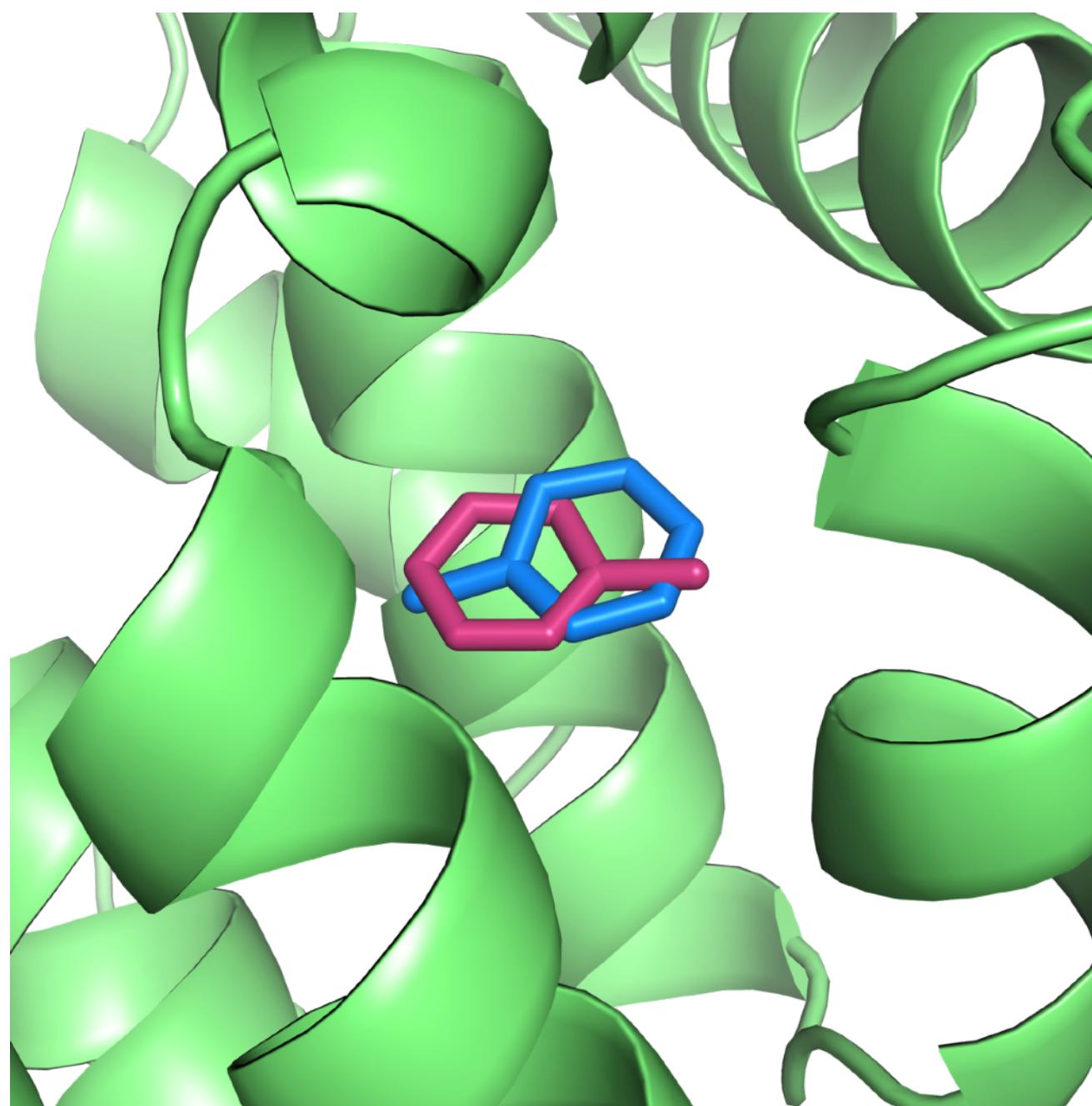
We compared equilibrium and non-equilibrium approaches



Crooks Fluctuation Theorem:
$$\frac{P_F(W)}{P_R(-W)} = e^{\beta(W - \Delta G)}$$

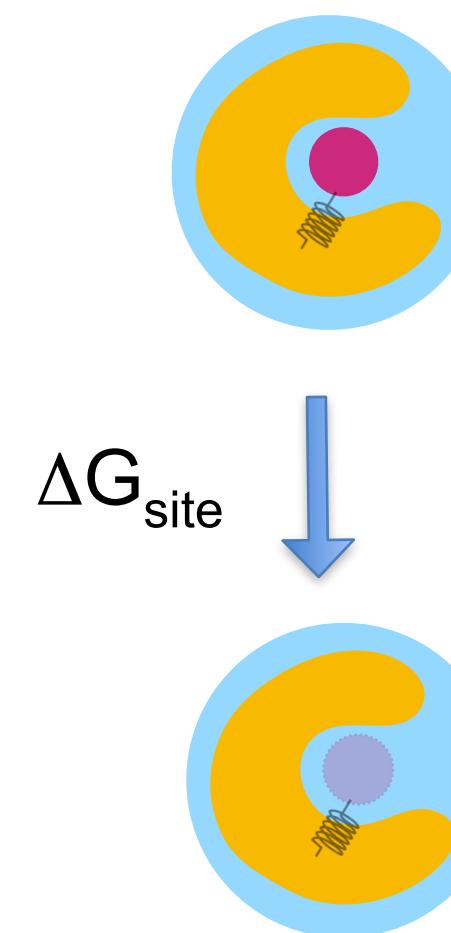
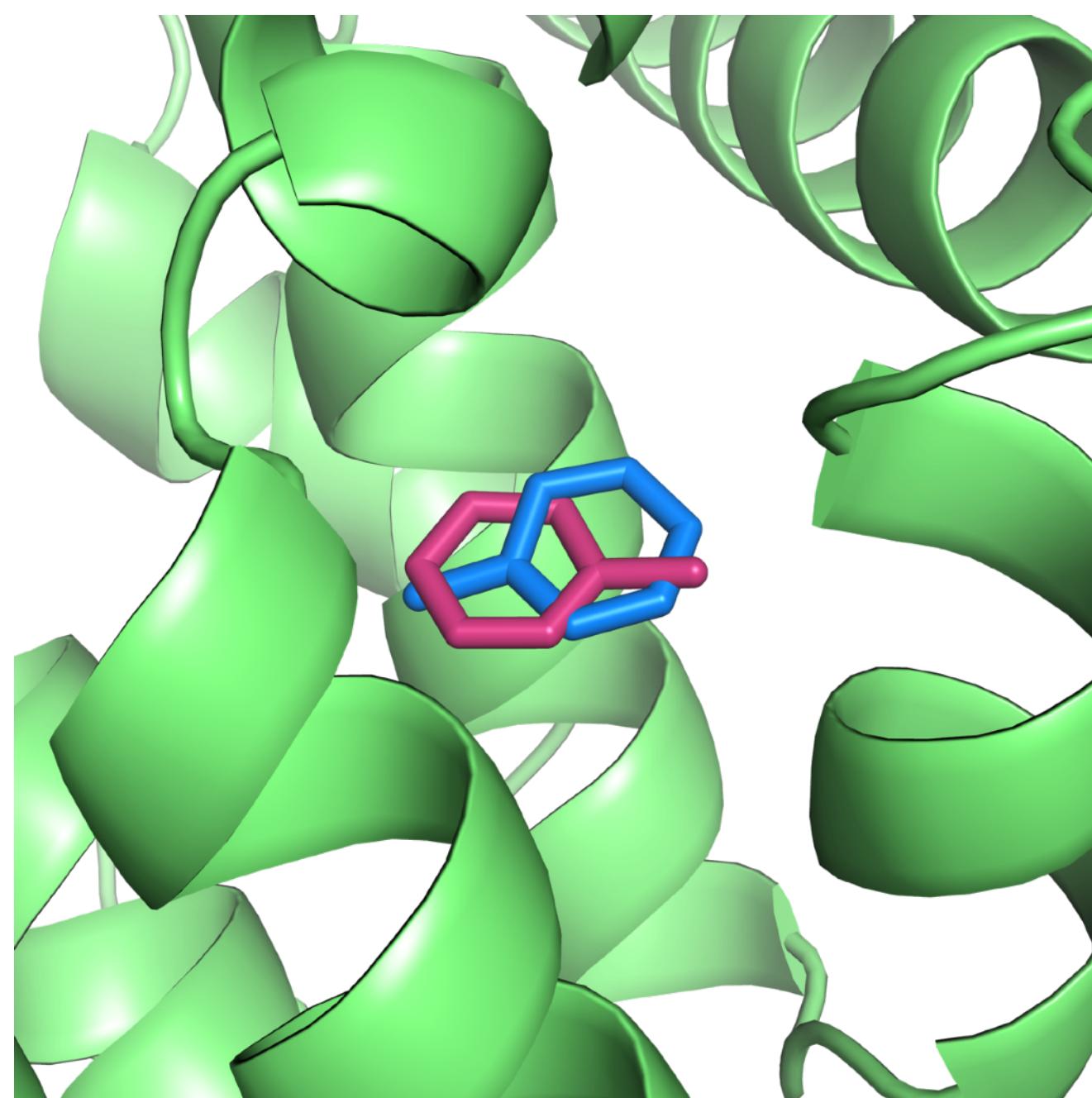
Toluene/T4 lysozyme L99A - Both approaches converge to statistically the same result

Two binding modes of toluene

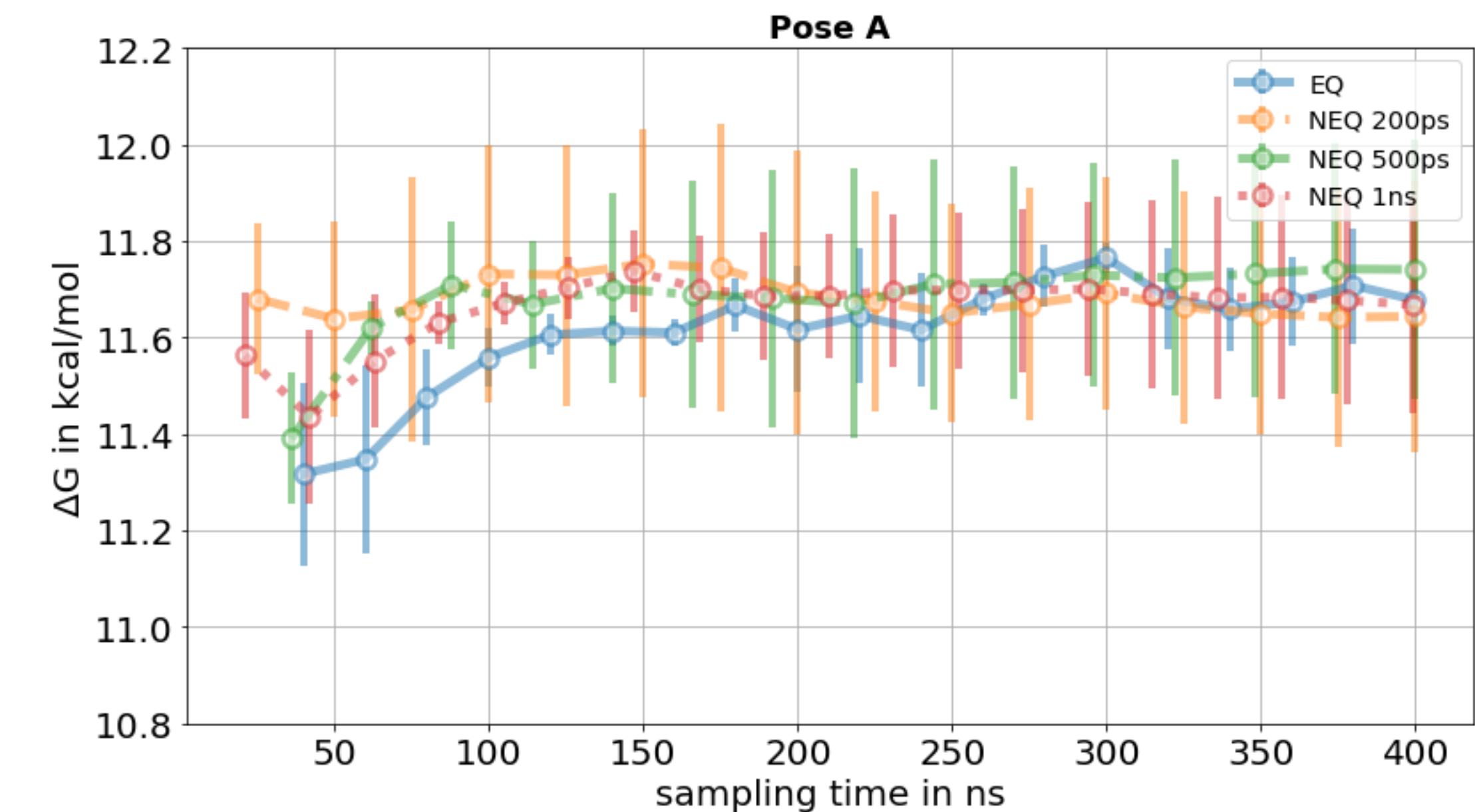


Toluene/T4 lysozyme L99A - Both approaches converge to statistically the same result

Two binding modes of toluene



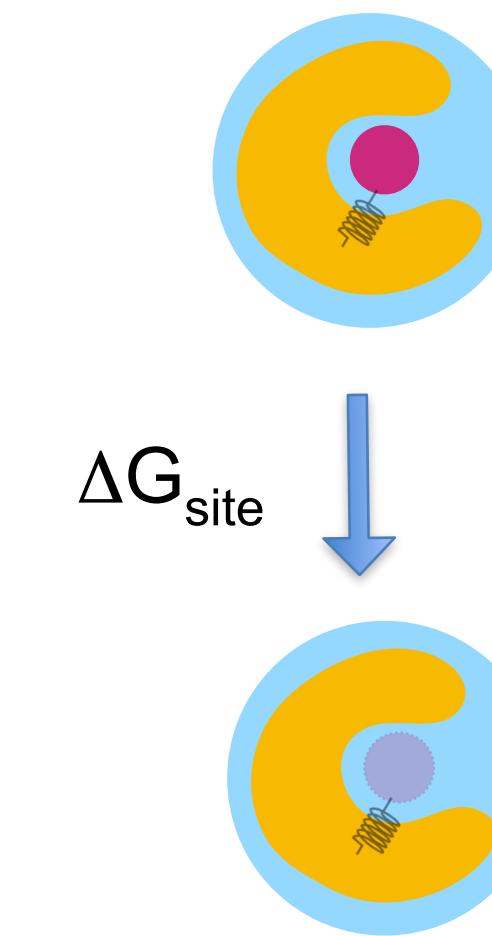
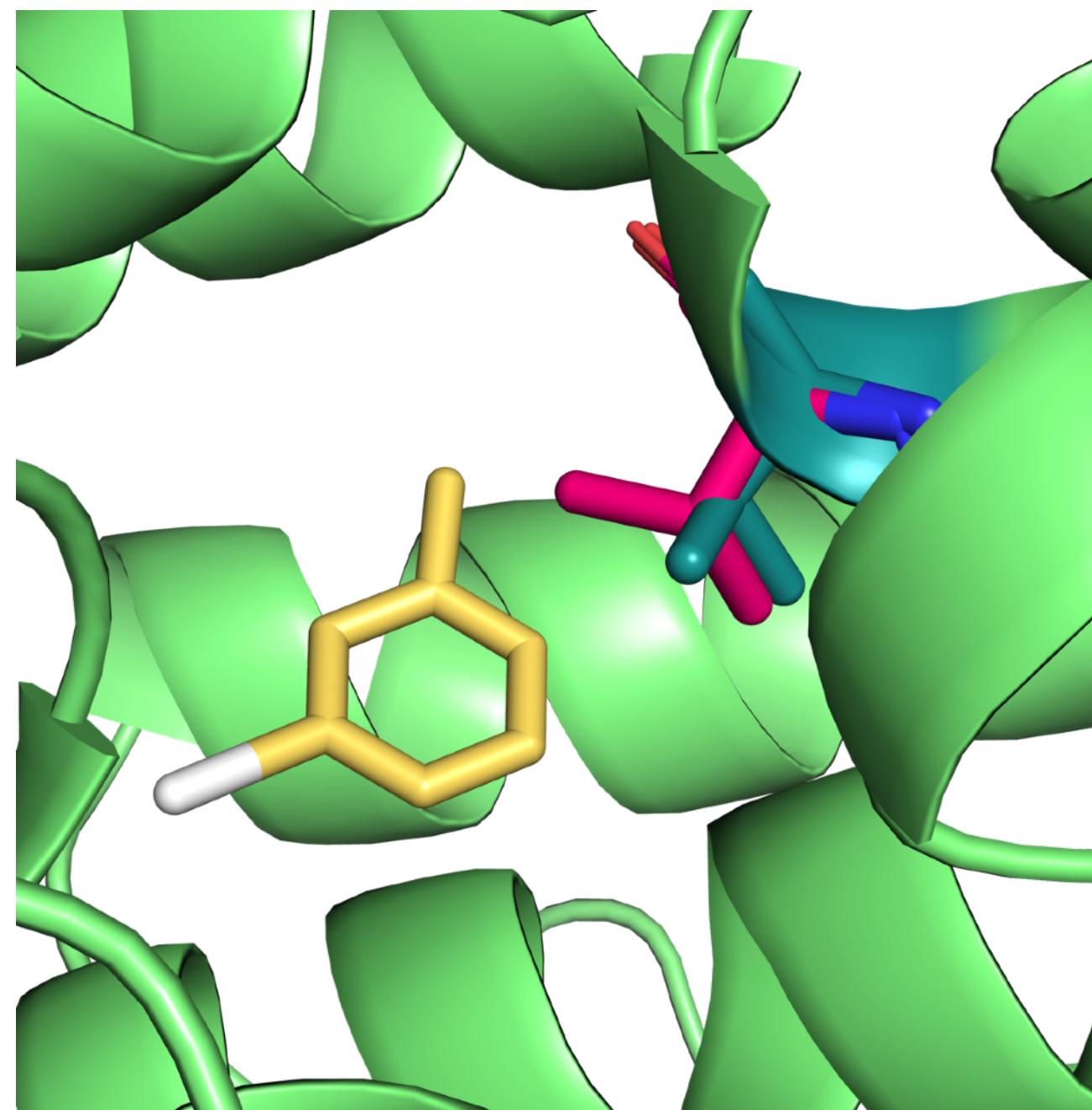
Both EQ and NEQ approaches converged to statistically the same free energy difference



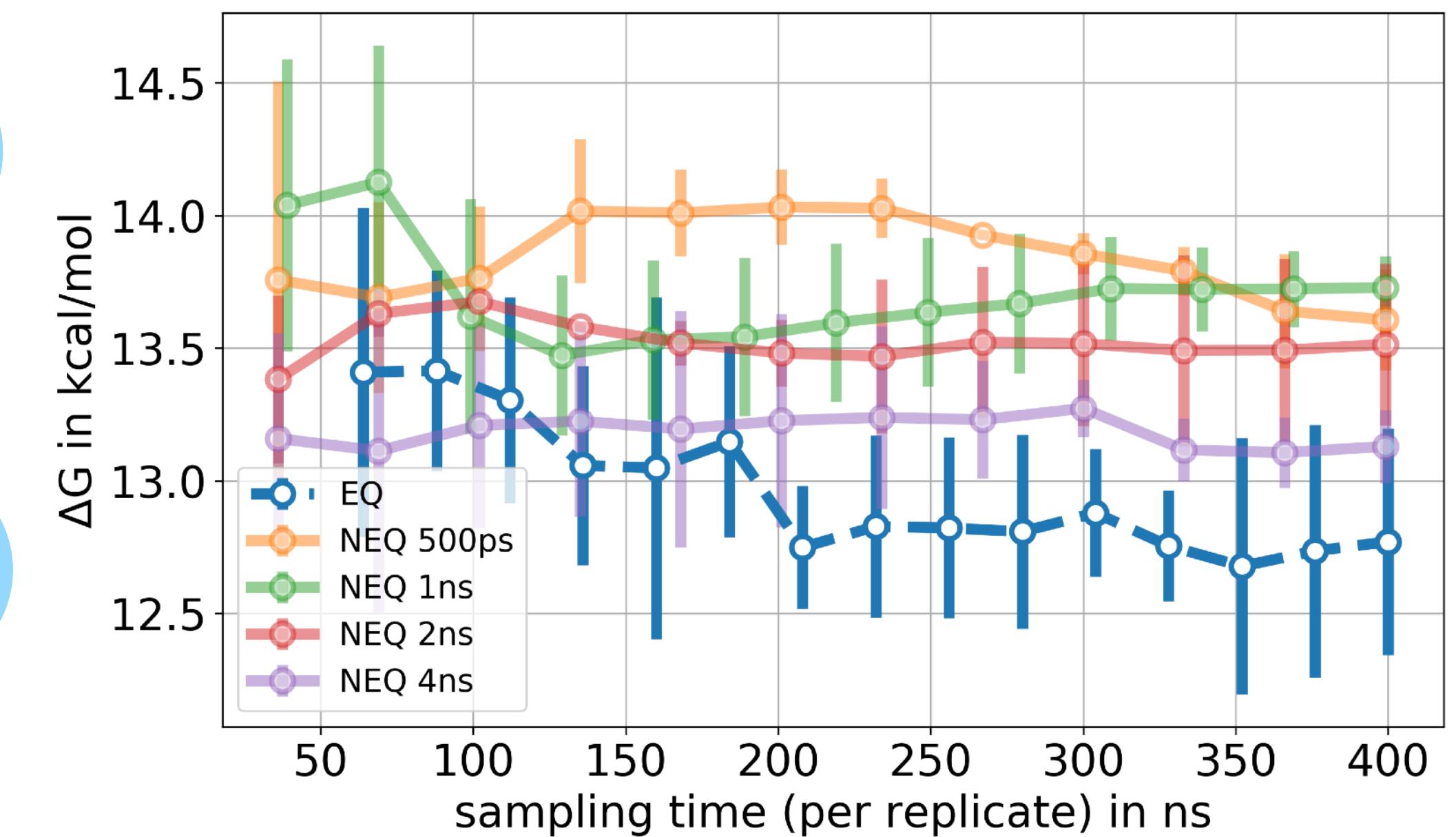
We ran three independent repeats to better assess the uncertainty

3-iodotoluene/T4 lysozyme L99A - EQ and NEQ approaches did not converge to the same free energy difference indicating sampling problems

Side chain rearrangement upon binding in the 3-iodotoluene/T4 lysozyme L99A system

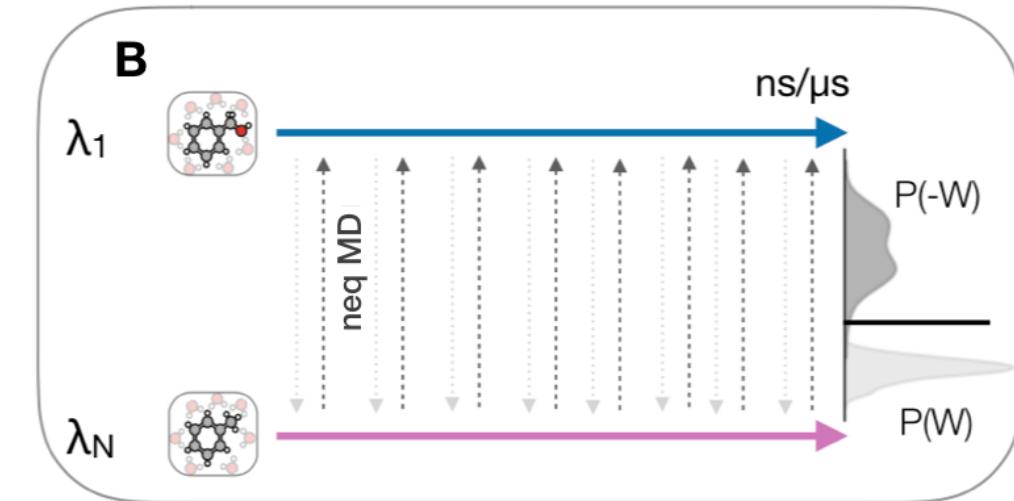
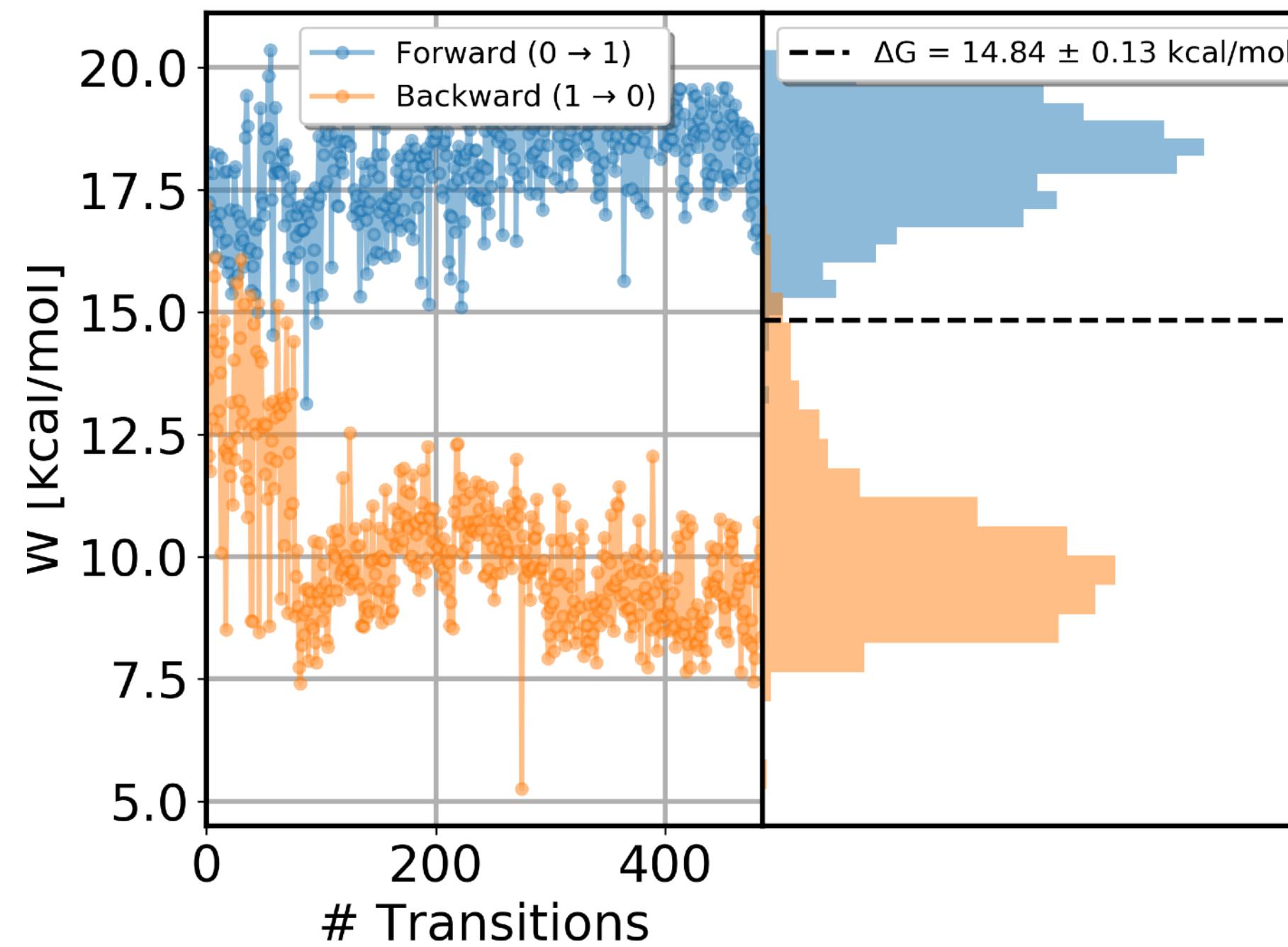


Different protocols did not converge to the same ΔG

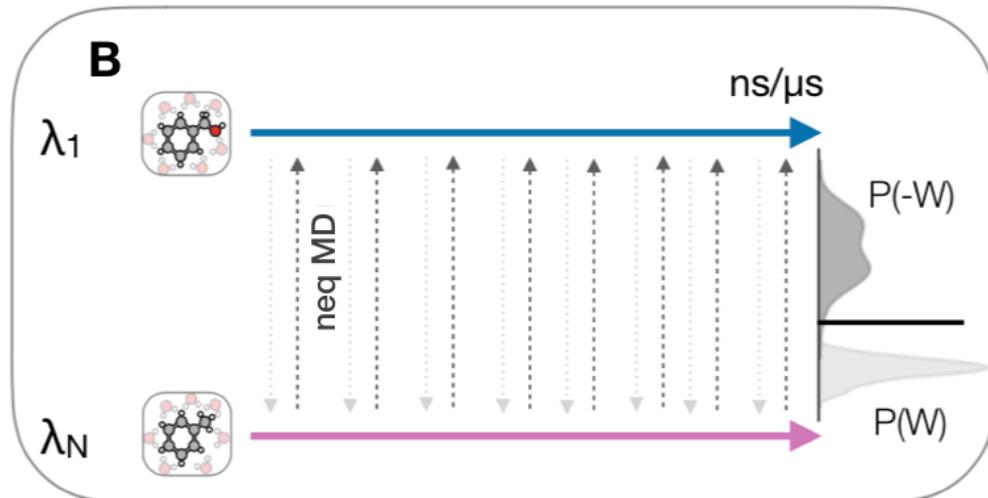


Calculating the Pearson Correlation Coefficient between work values and side chain dihedral angles helped identify sampling problems in the NEQ approach

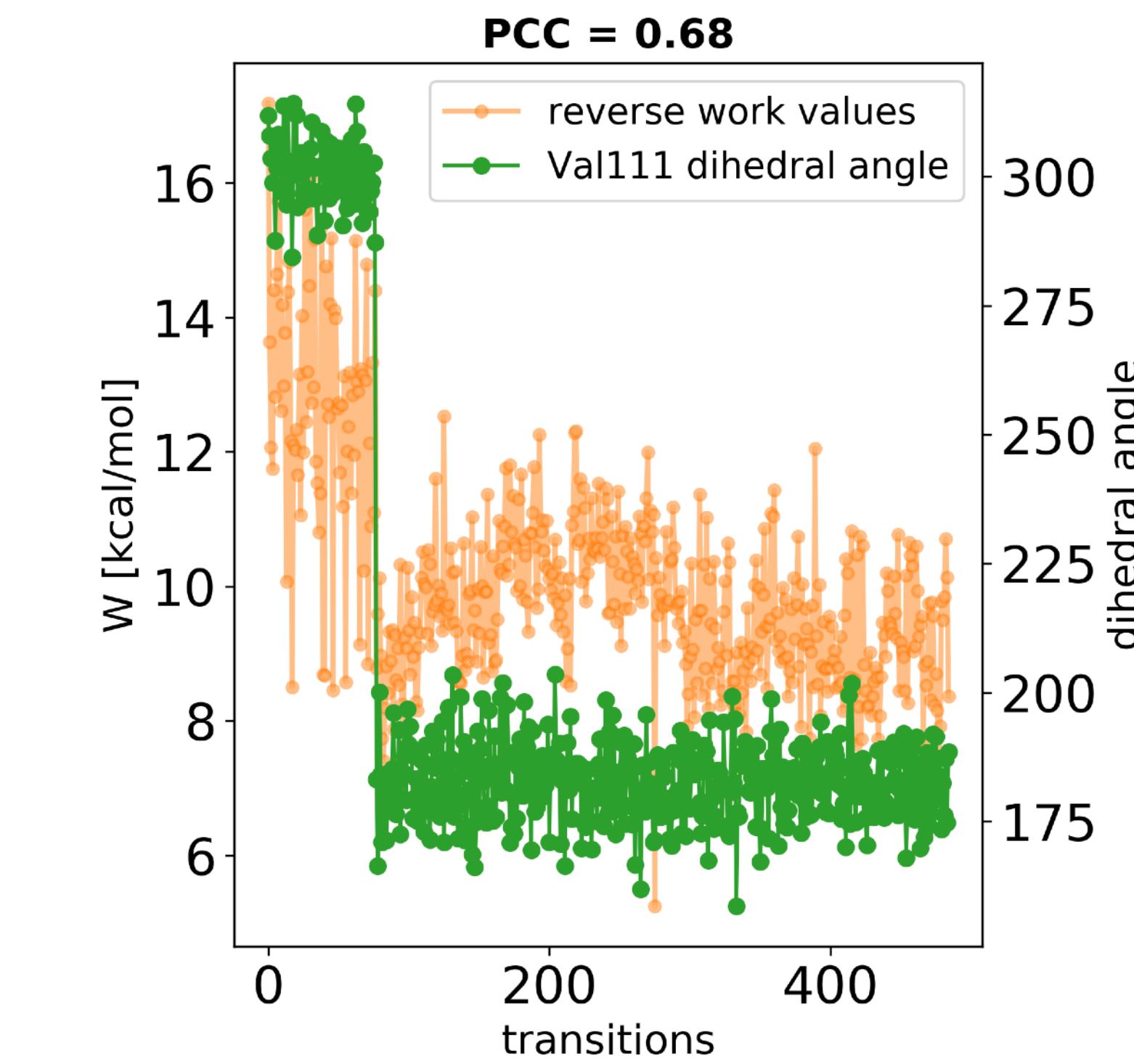
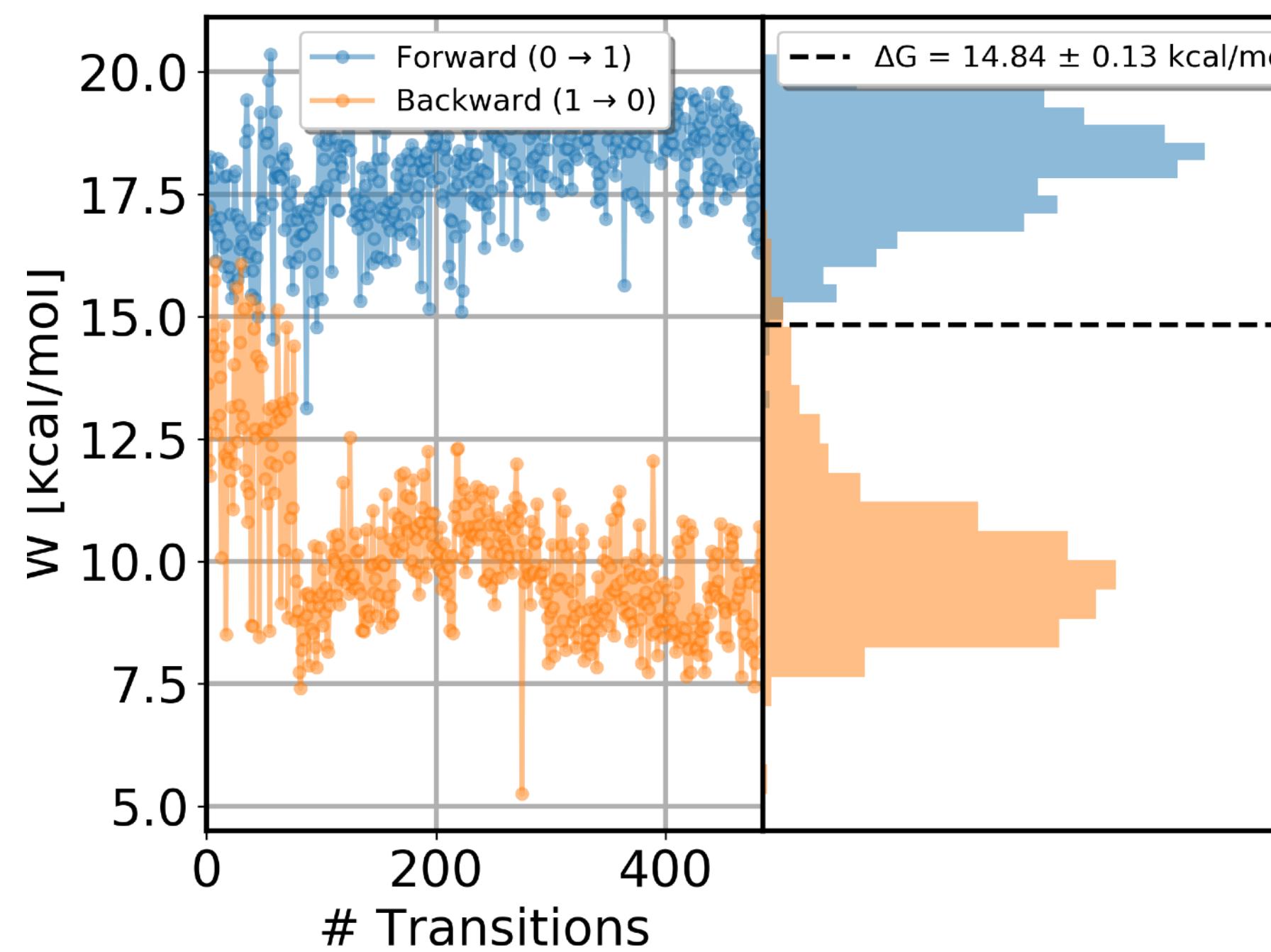
In NEQ poor overlap of work distributions indicates sampling issues



Calculating the Pearson Correlation Coefficient between work values and side chain dihedral angles helped identify sampling problems in the NEQ approach

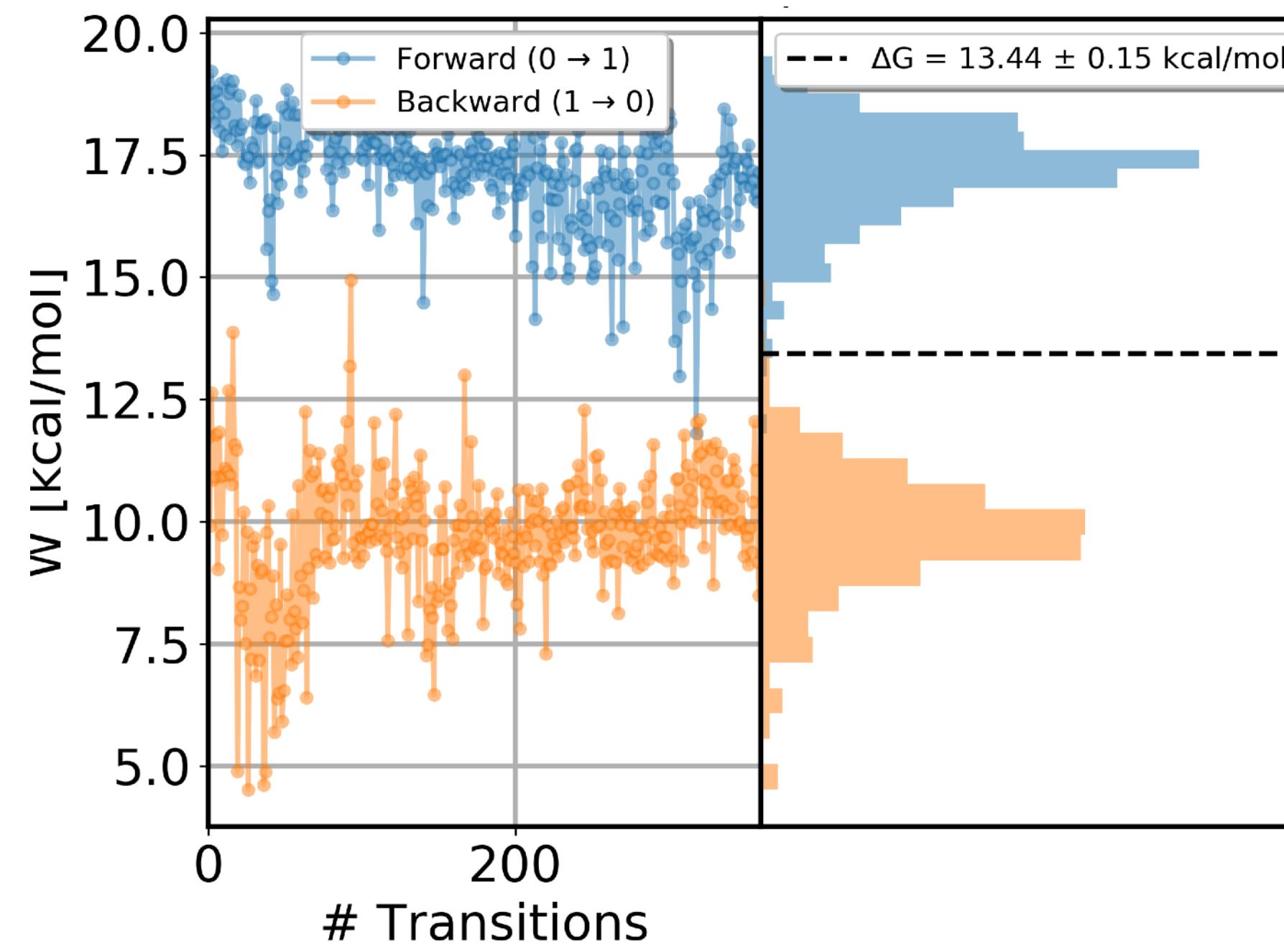


In NEQ poor overlap of work distributions indicates sampling issues

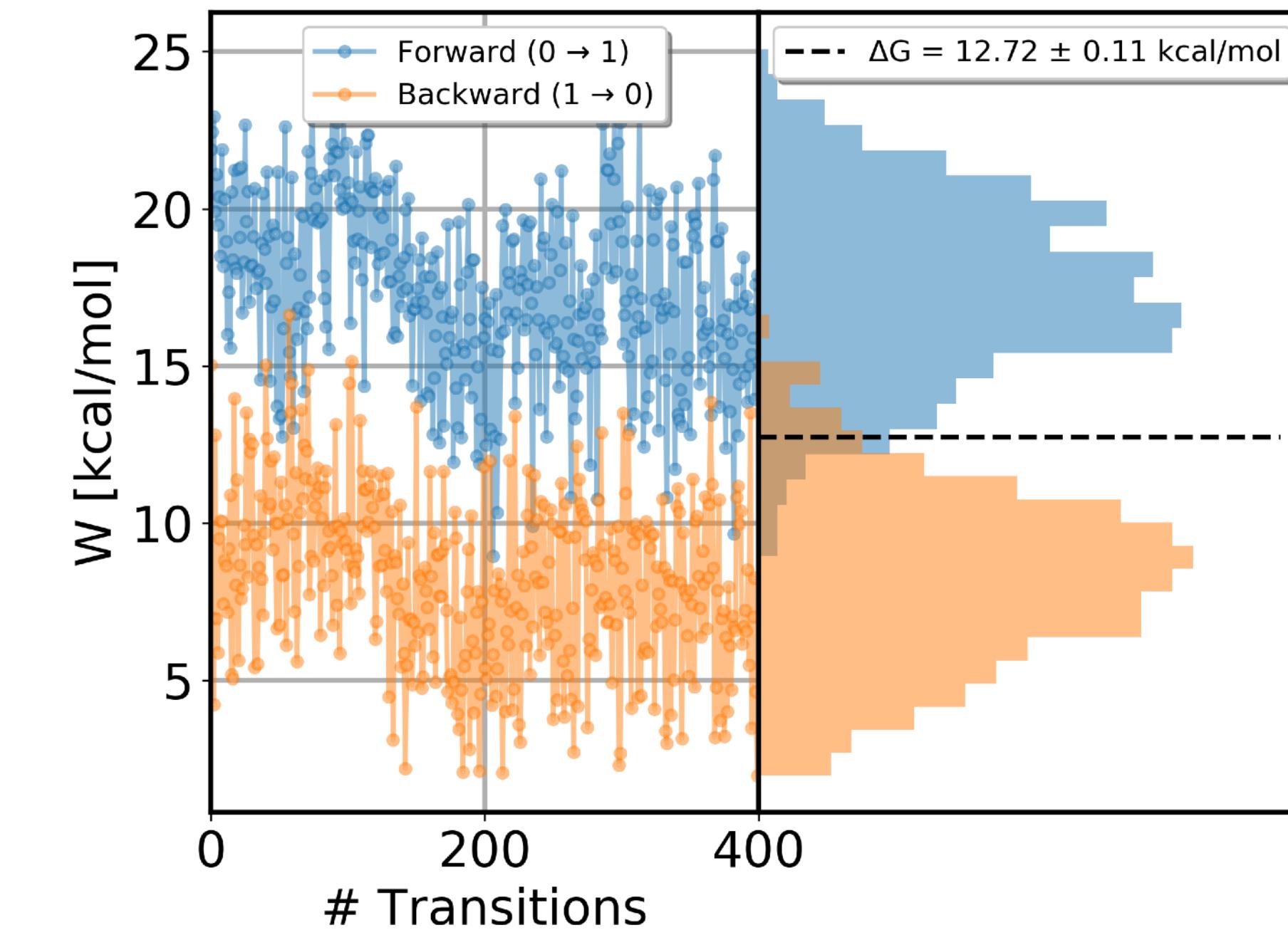


Restraining V111 leads to a better overlap of NEQ work distributions

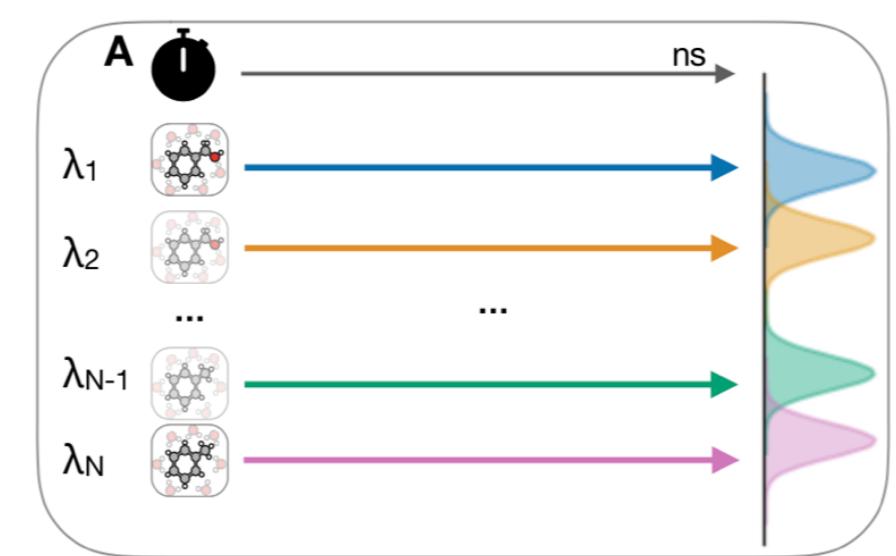
Non overlapping work distributions
can indicate sampling problems



Overlapping work distributions after
restricting the Val111 side chain

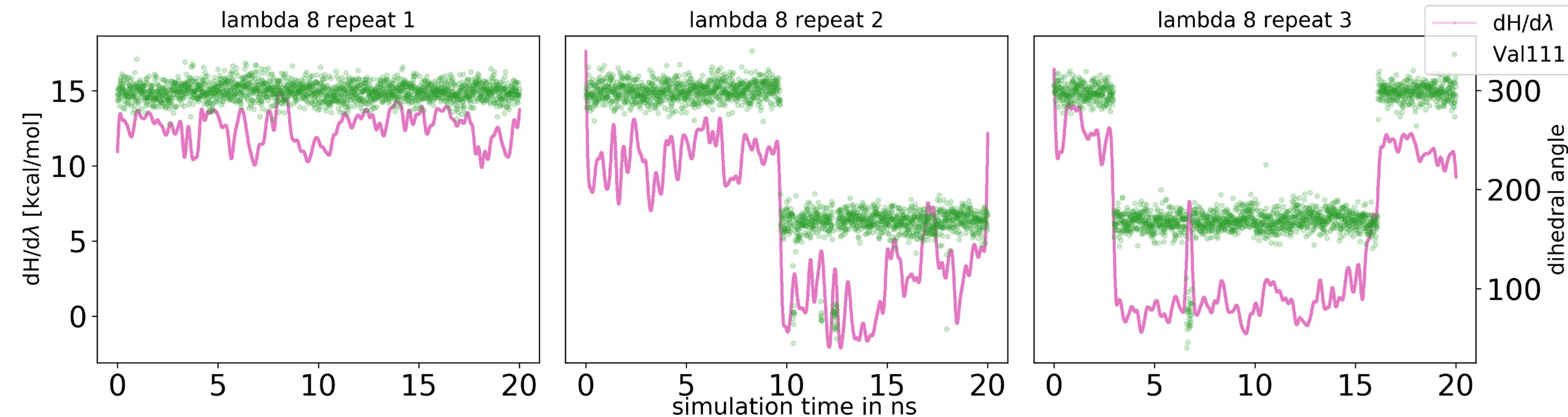
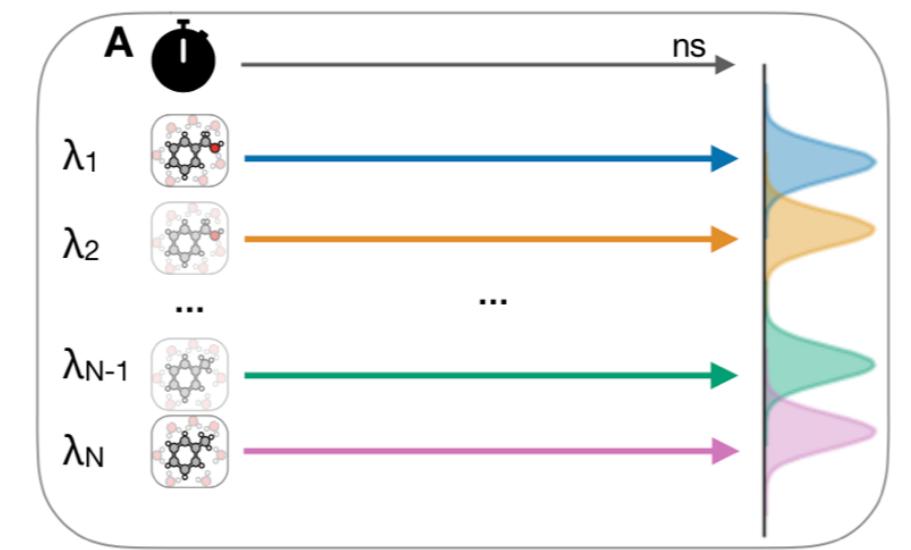


Analyzing sudden changes in $dH/d\lambda$ helped identify sampling problems in the EQ approach

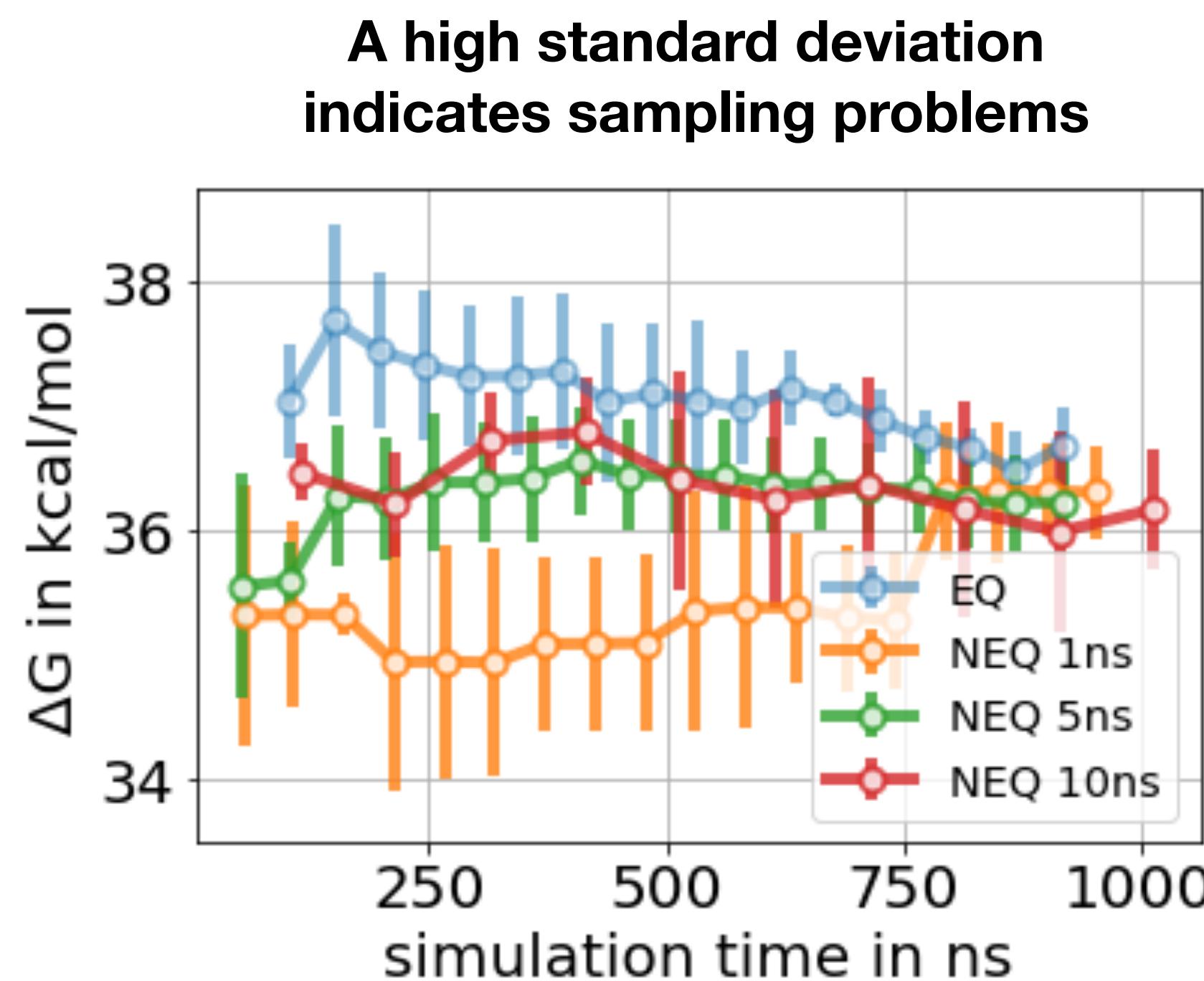


Analyzing sudden changes in $dH/d\lambda$ helped identify sampling problems in the EQ approach

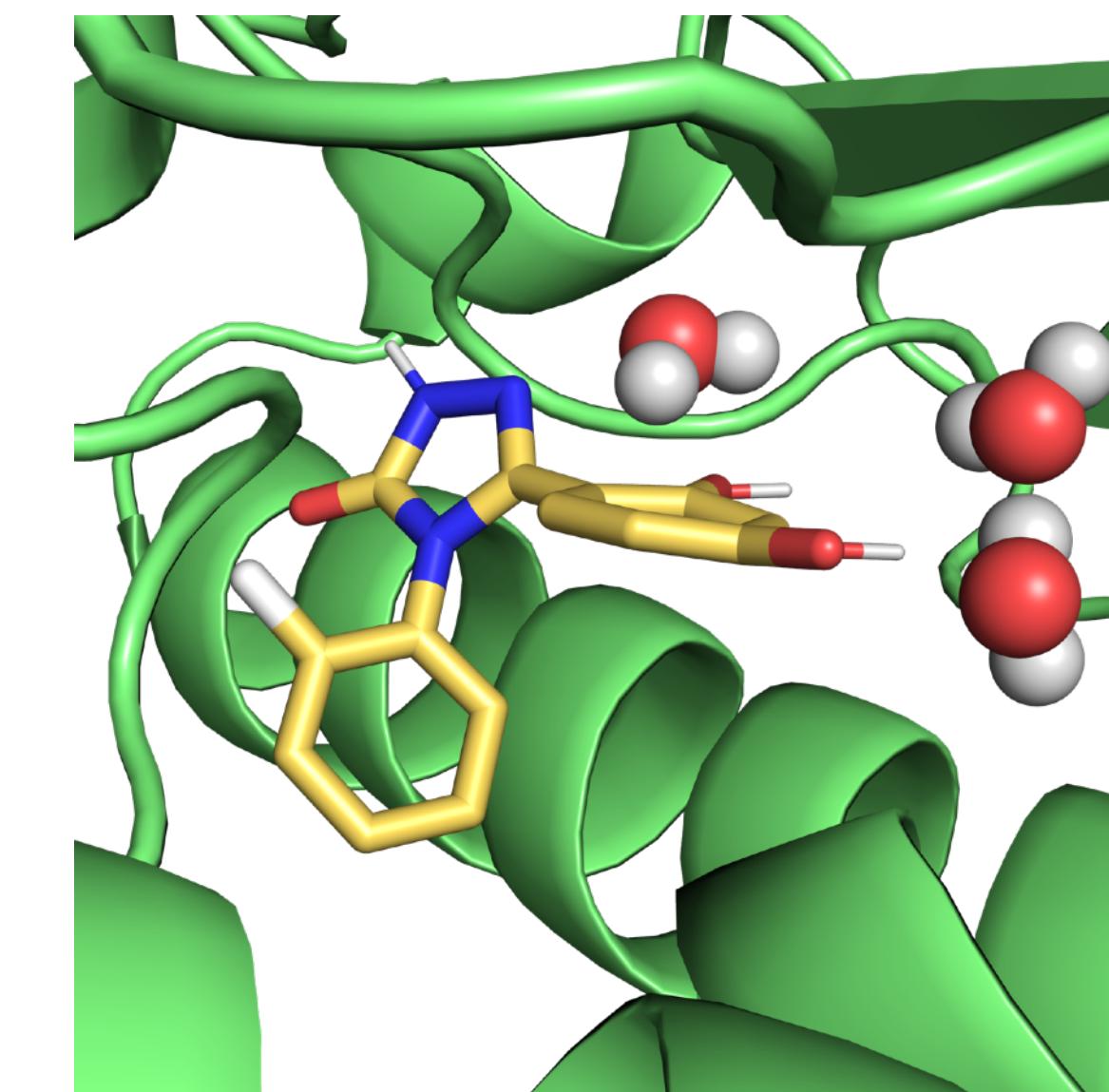
Thermodynamic integration: $\Delta A = \int_0^1 \left\langle \frac{dU(\lambda, q)}{d\lambda} \right\rangle_\lambda d\lambda.$



Slow water rearrangement upon ligand binding can lead to sampling problems in HSP90 systems

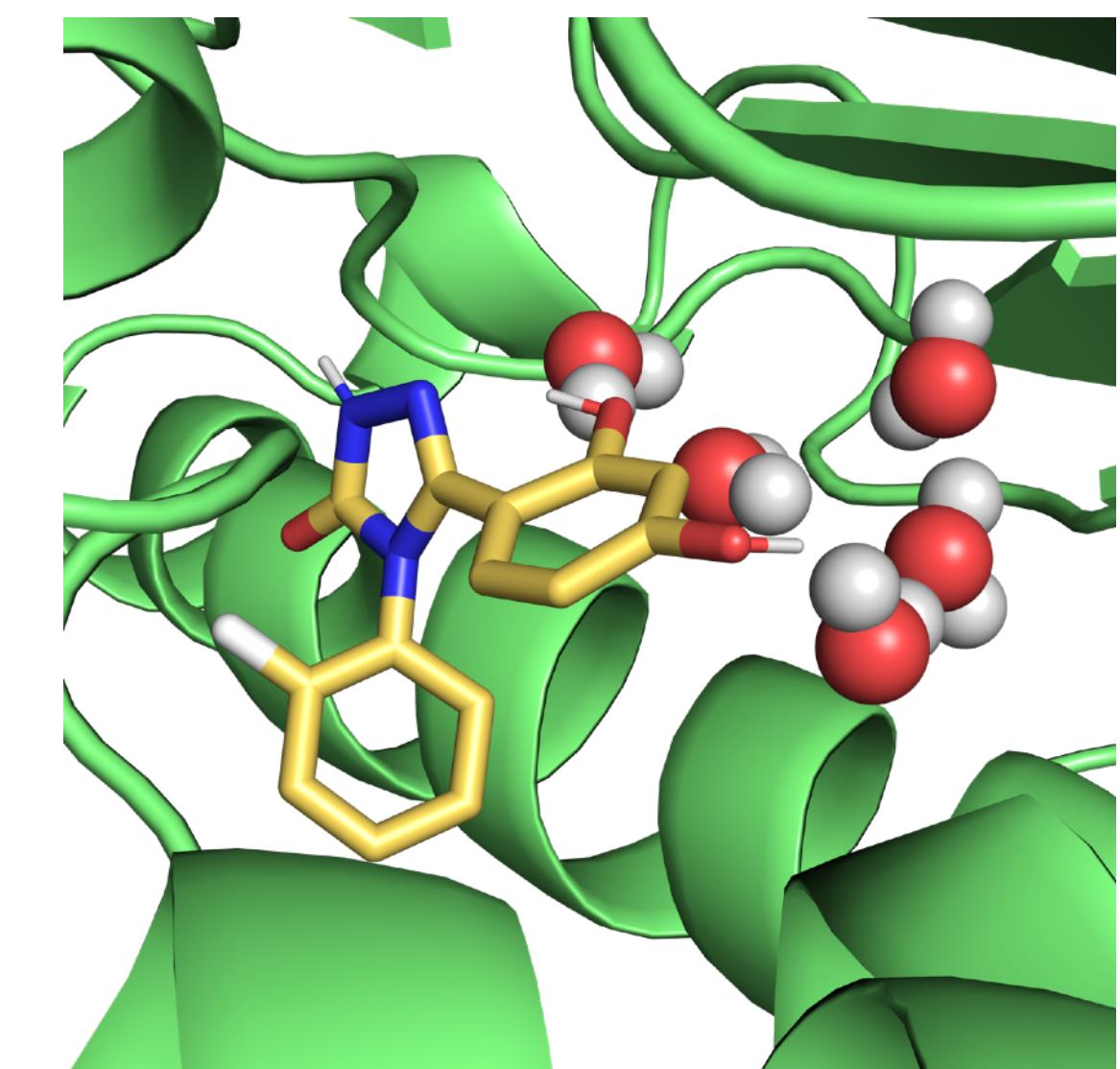
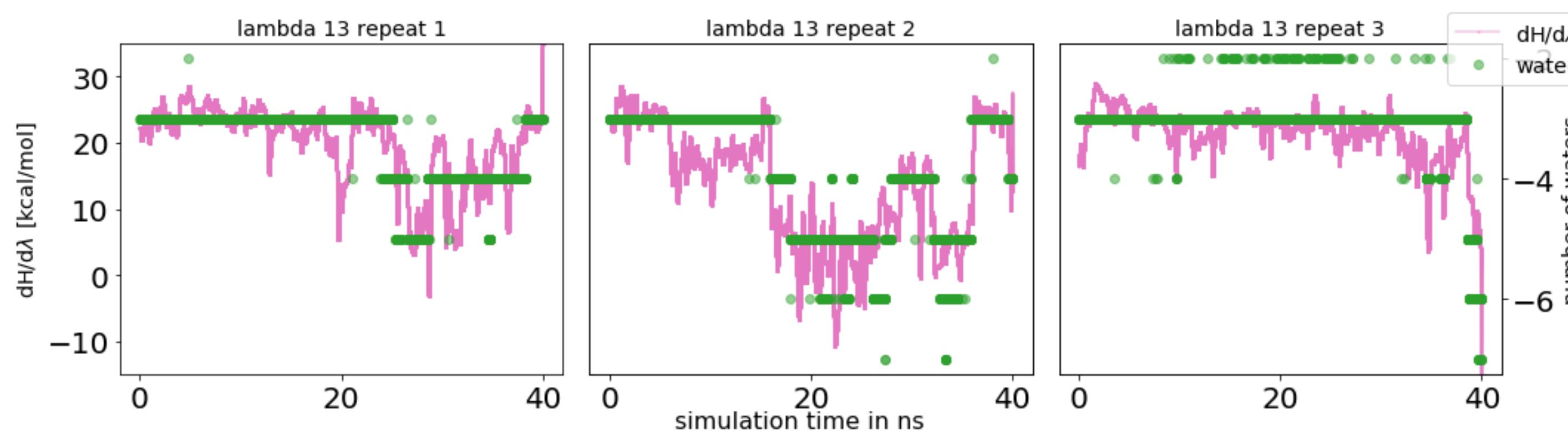


Water displacement upon binding led to sampling issues in FEC in the HSP90 system.



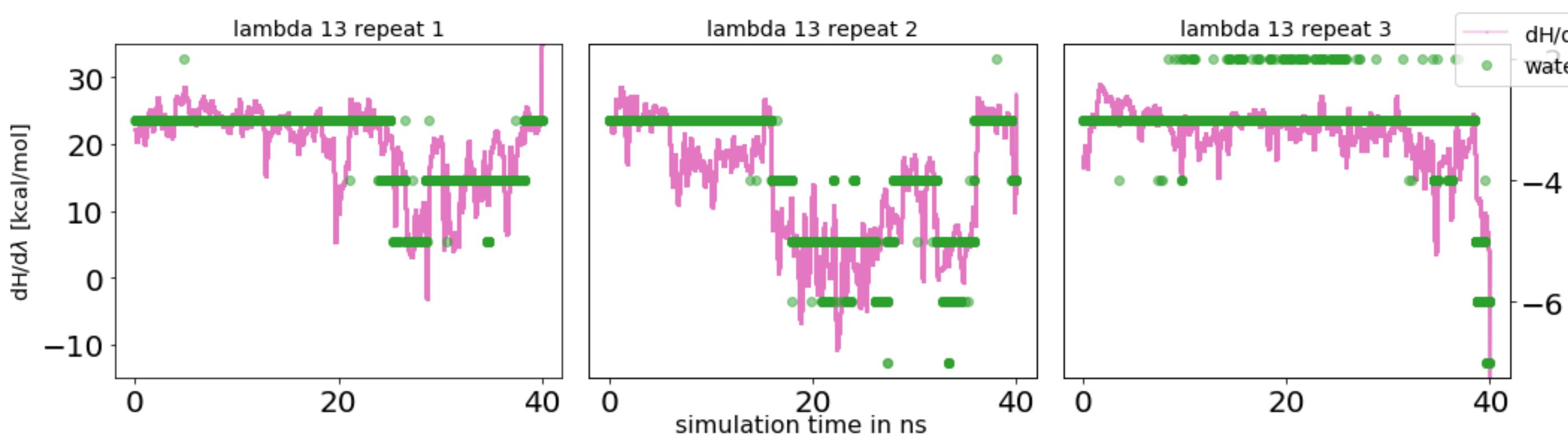
Water sampling problems manifested differently in EQ and NEQ approaches

Water problems in intermediate states in EQ

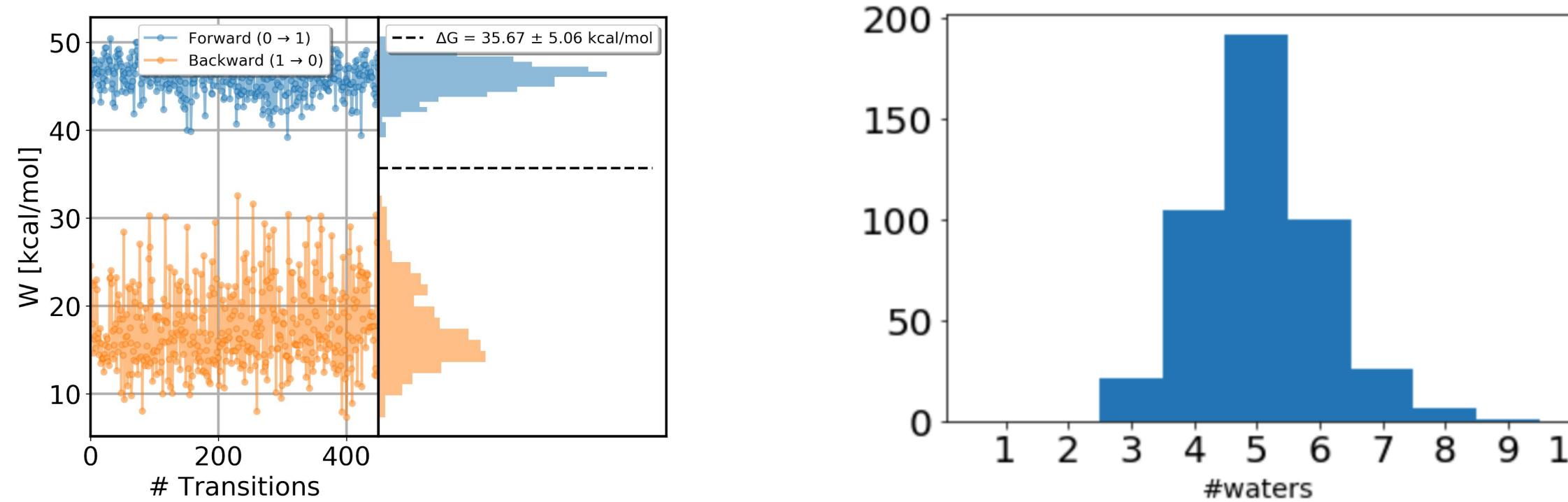


Water sampling problems manifested differently in EQ and NEQ approaches

Water problems in intermediate states in EQ

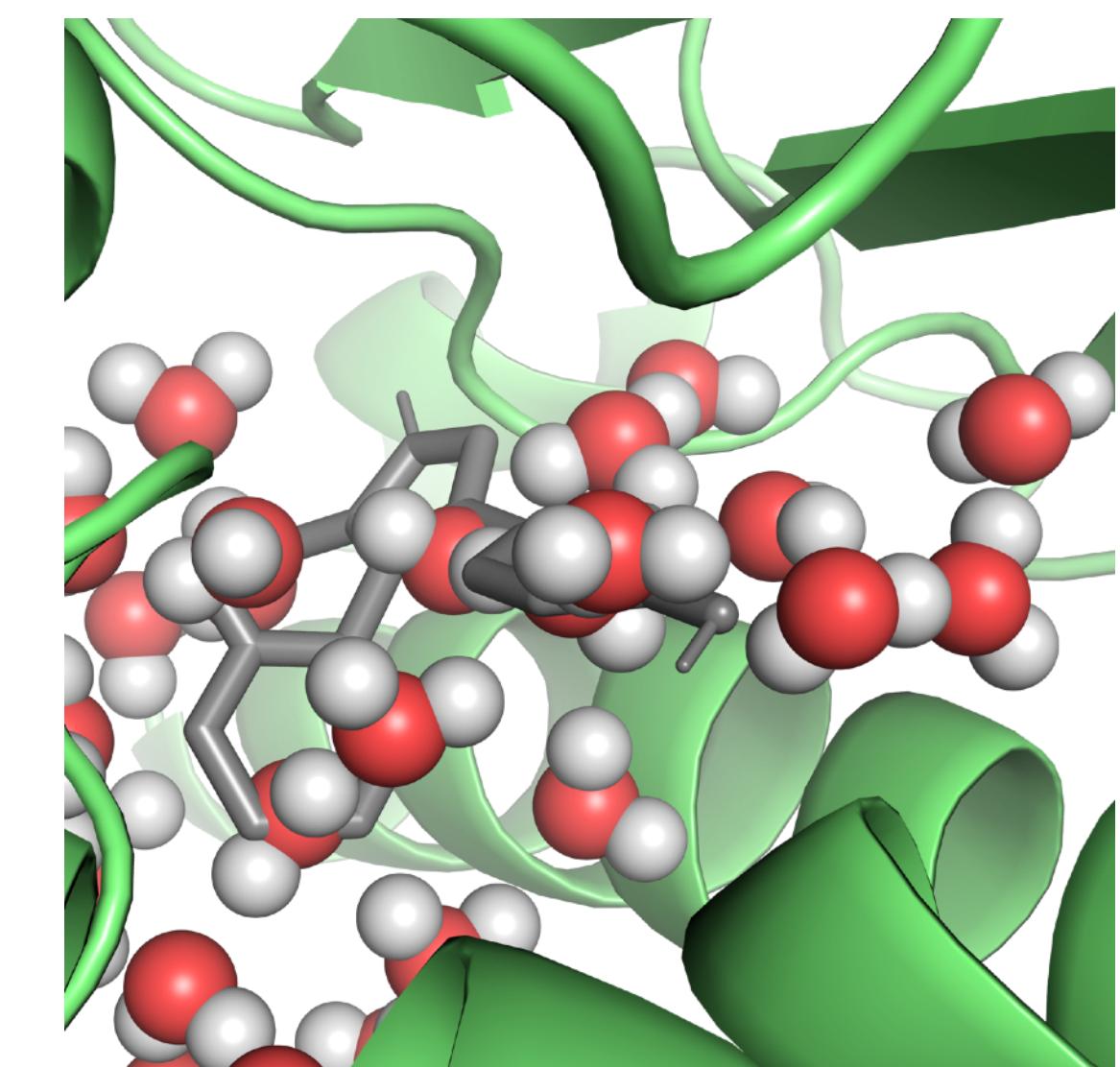
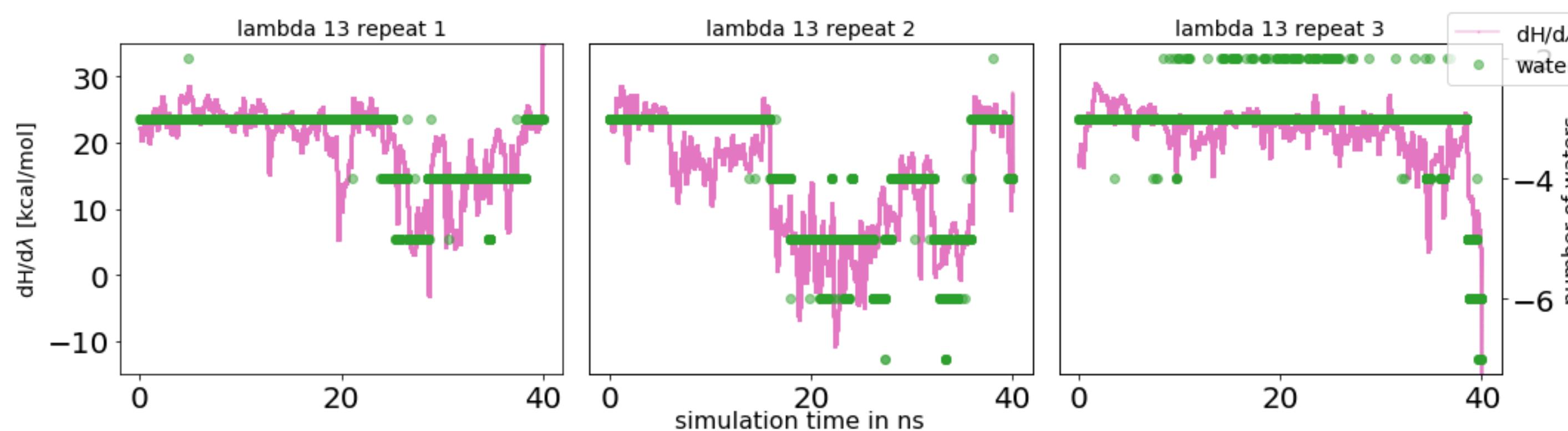


Water molecules get trapped in the binding site
which leads to non-overlapping work distributions in NEQ

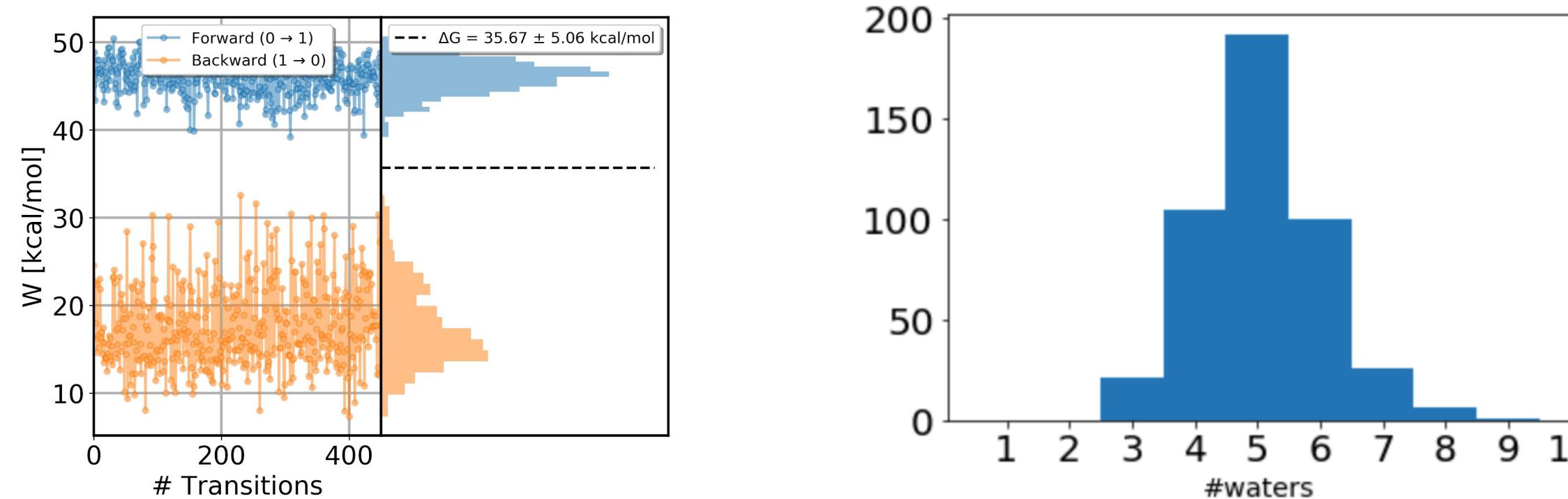


Water sampling problems manifested differently in EQ and NEQ approaches

Water problems in intermediate states in EQ

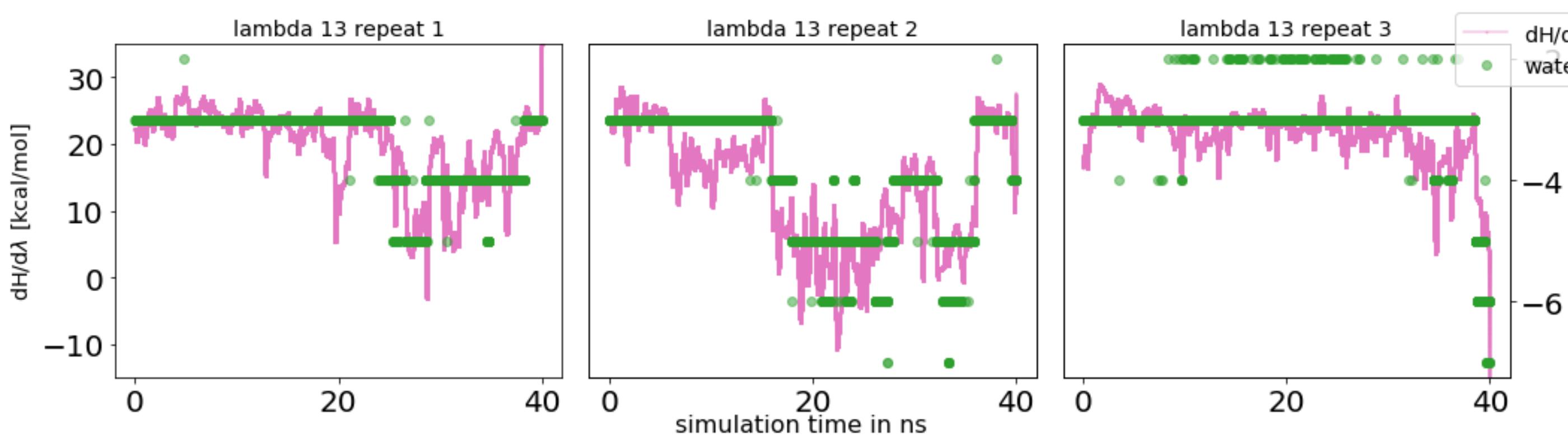


Water molecules get trapped in the binding site
which leads to non-overlapping work distributions in NEQ

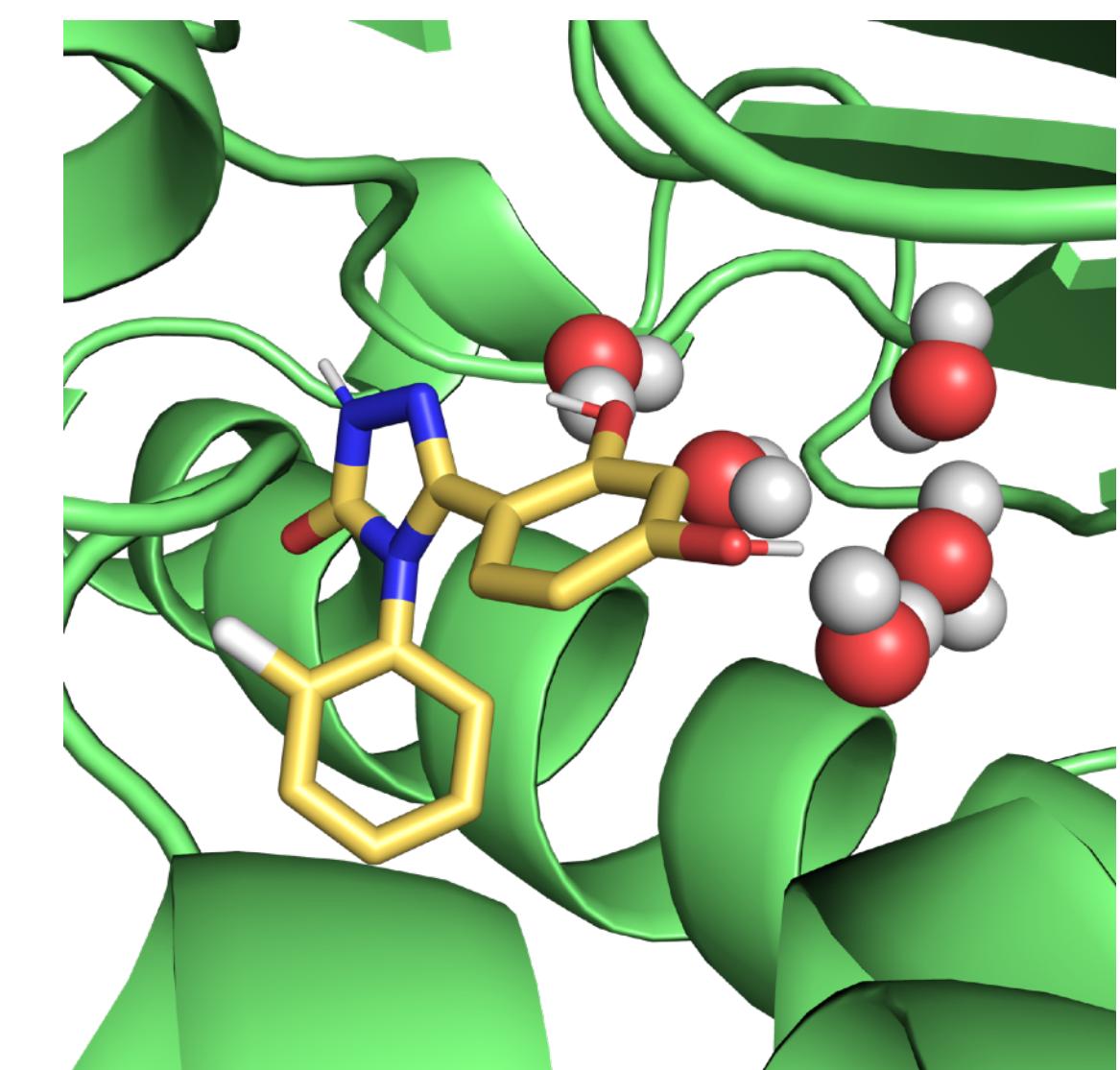
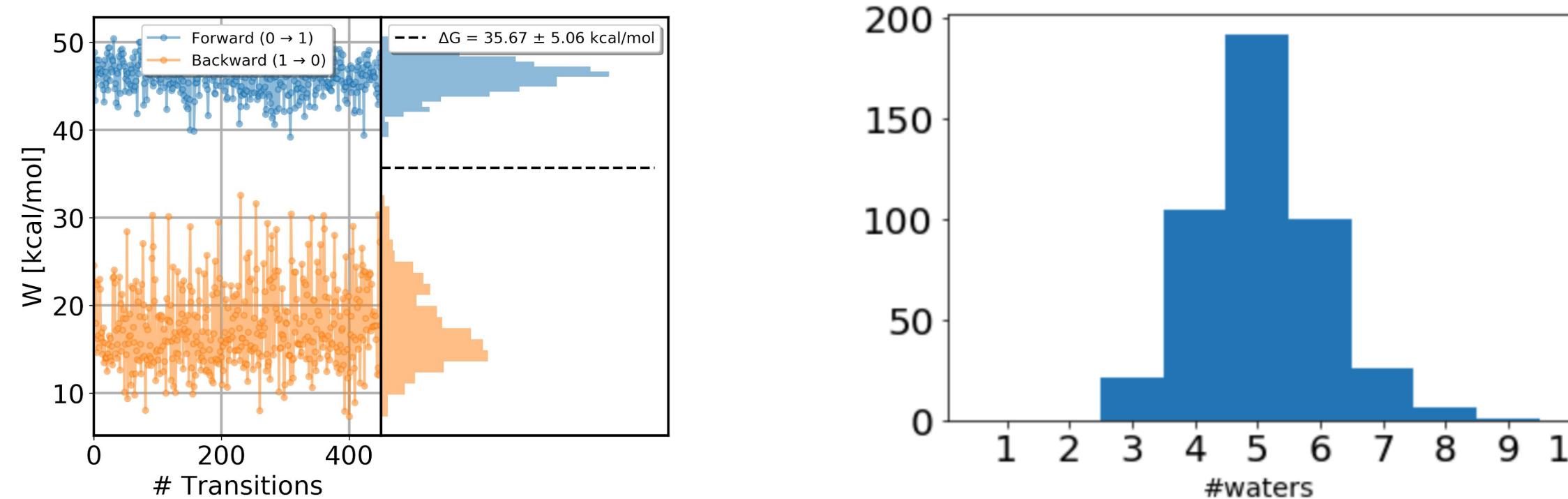


Water sampling problems manifested differently in EQ and NEQ approaches

Water problems in intermediate states in EQ

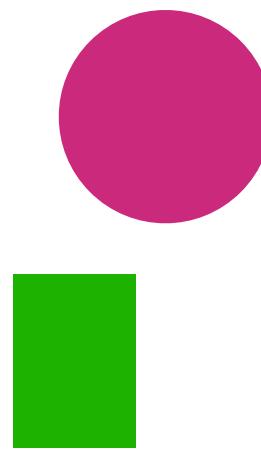


Water molecules get trapped in the binding site
which leads to non-overlapping work distributions in NEQ



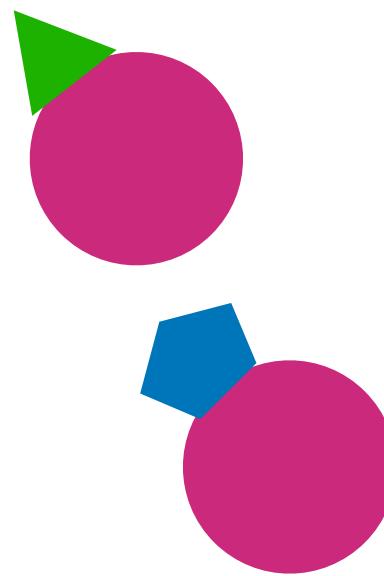
ABFE and RBFE have different benefits and weaknesses

ABFE



- No restriction on ligand structure
- No restriction on ligand binding mode

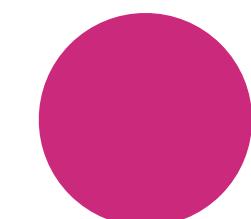
RBFE



- Ligands have to be similar
- Ligands have to have same binding mode
- Mostly R group modifications

ABFE and RBFE have different benefits and weaknesses

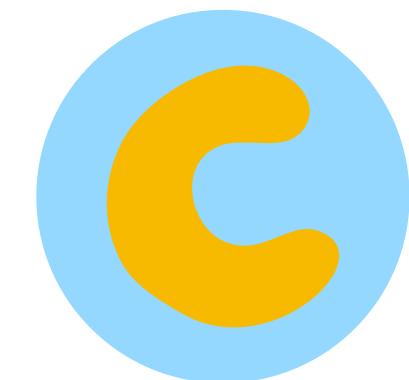
ABFE



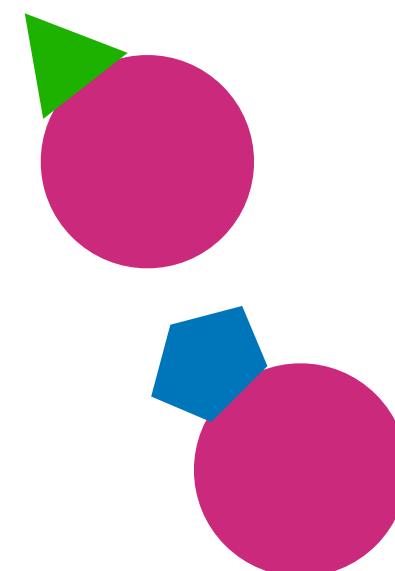
- No restriction on ligand structure
- No restriction on ligand binding mode



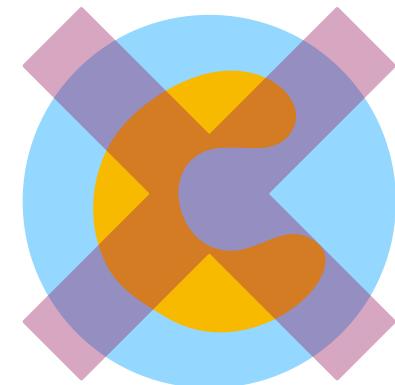
- Sampling issues can be more severe than in RBFE



RBFE



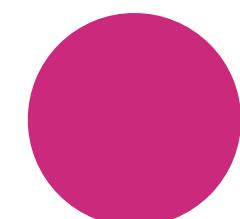
- Ligands have to be similar
- Ligands have to have same binding mode
- Mostly R group modifications



- Sampling of the apo system not necessary

ABFE and RBFE have different benefits and weaknesses

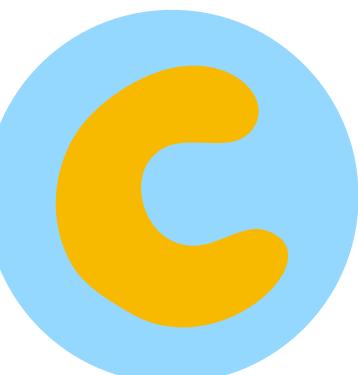
ABFE



No restriction on ligand structure
No restriction on ligand binding mode

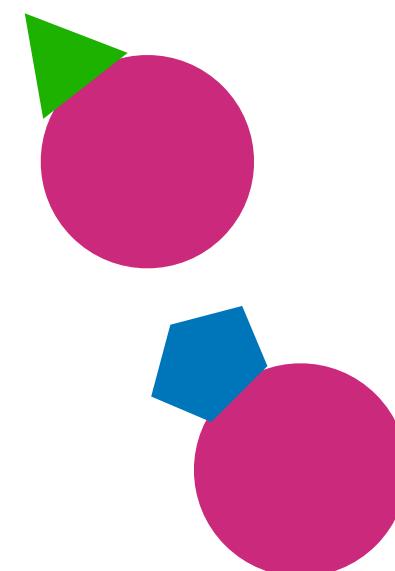


Sampling issues can be more severe than in RBFE

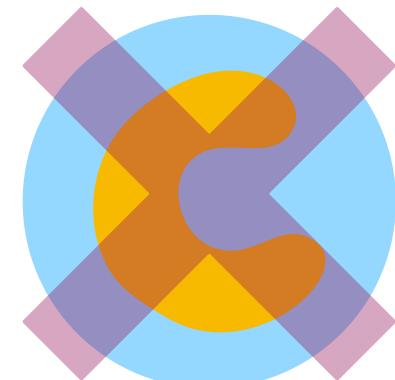


computationally expensive

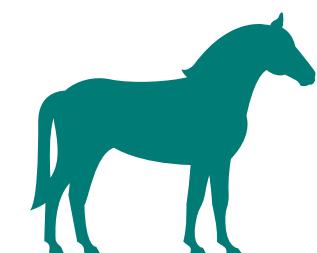
RBFE



Ligands have to be similar
Ligands have to have same binding mode
Mostly R group modifications

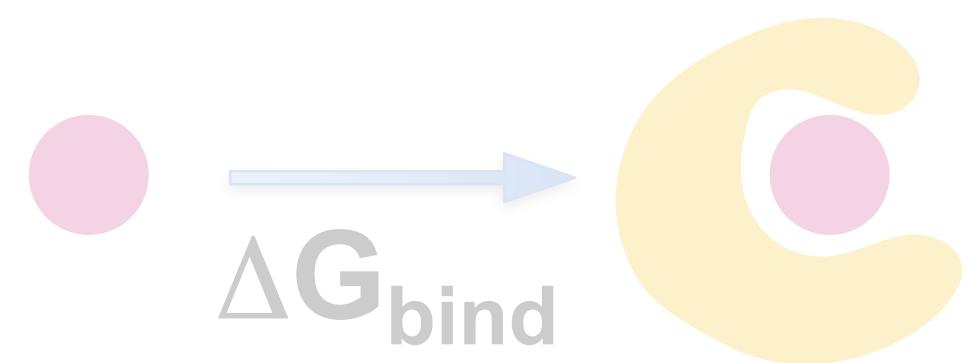


Sampling of the apo system not necessary



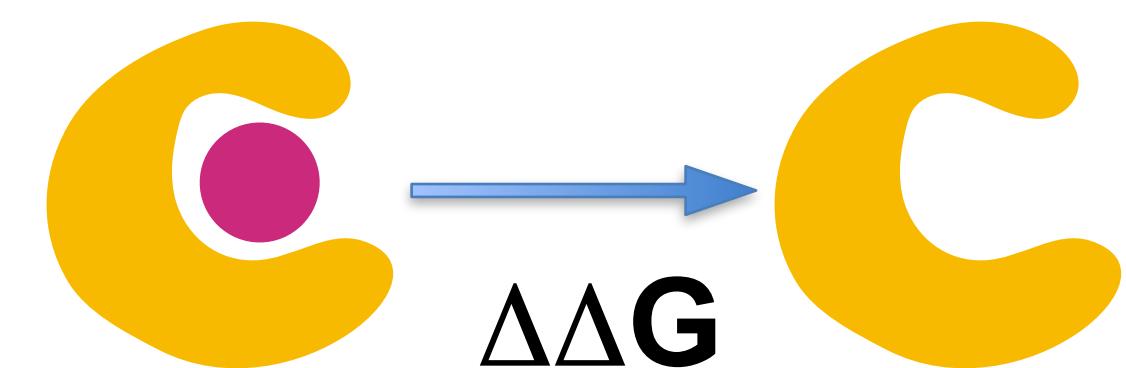
computationally efficient

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



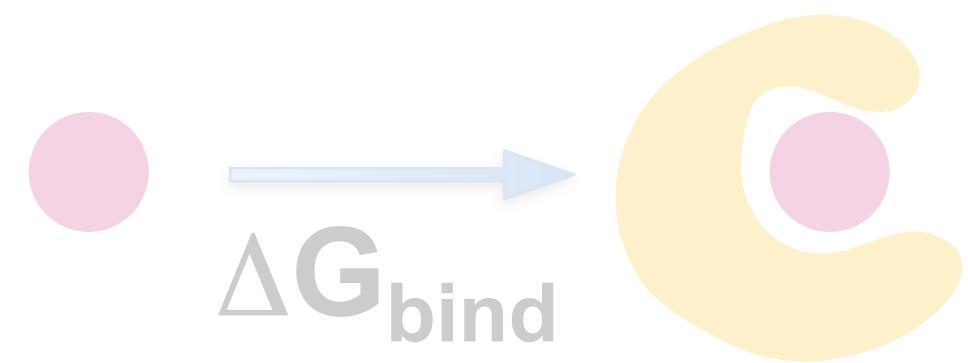
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



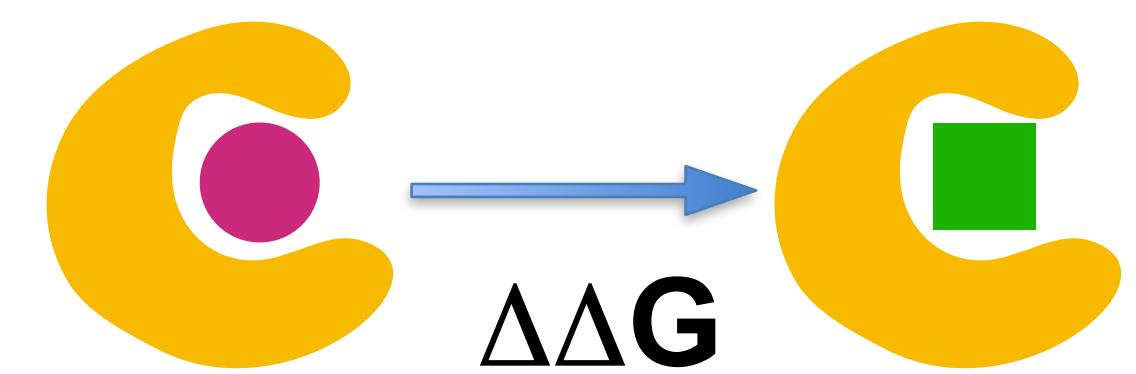
Enhanced sampling methods for binding free energies

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



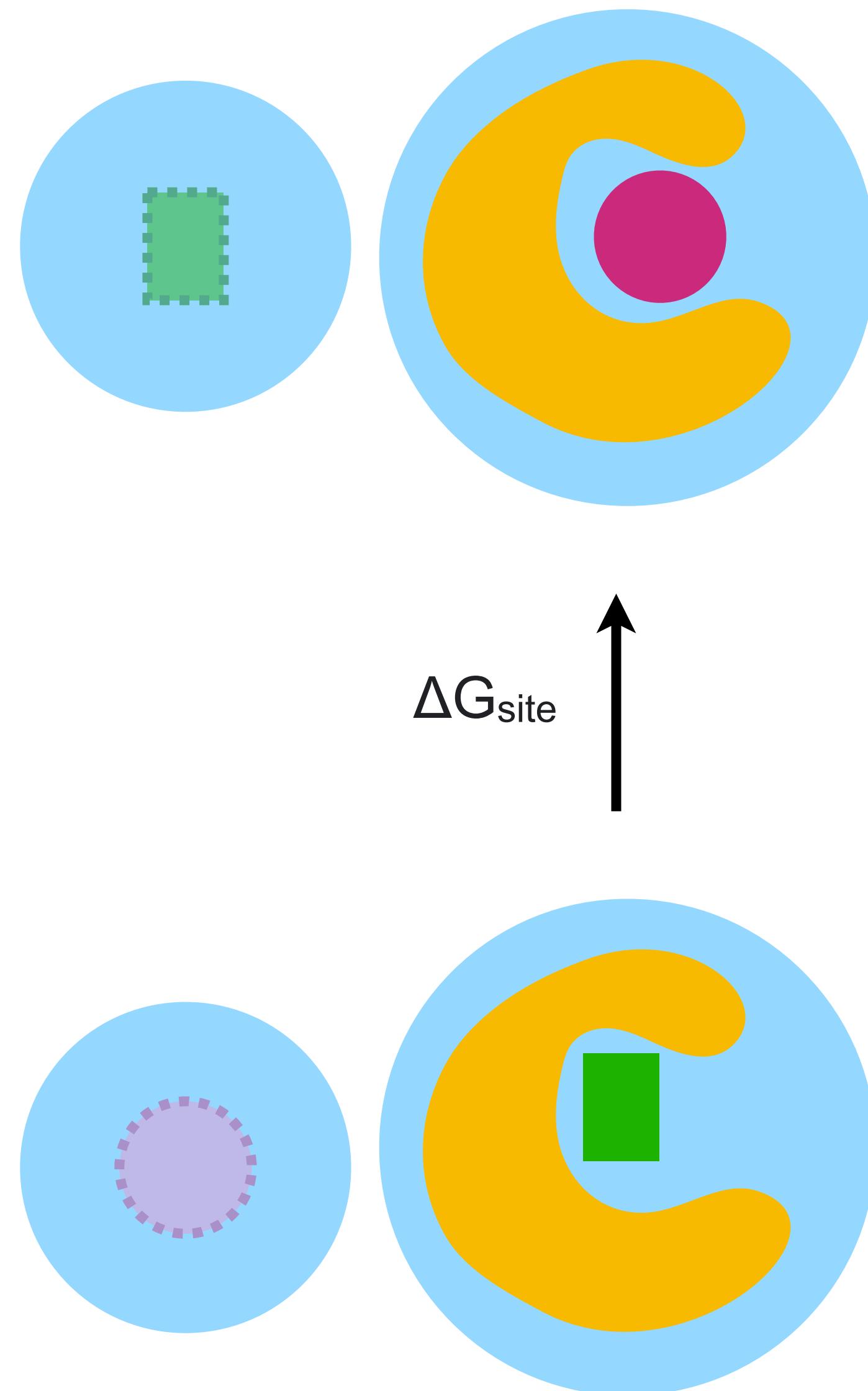
Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods

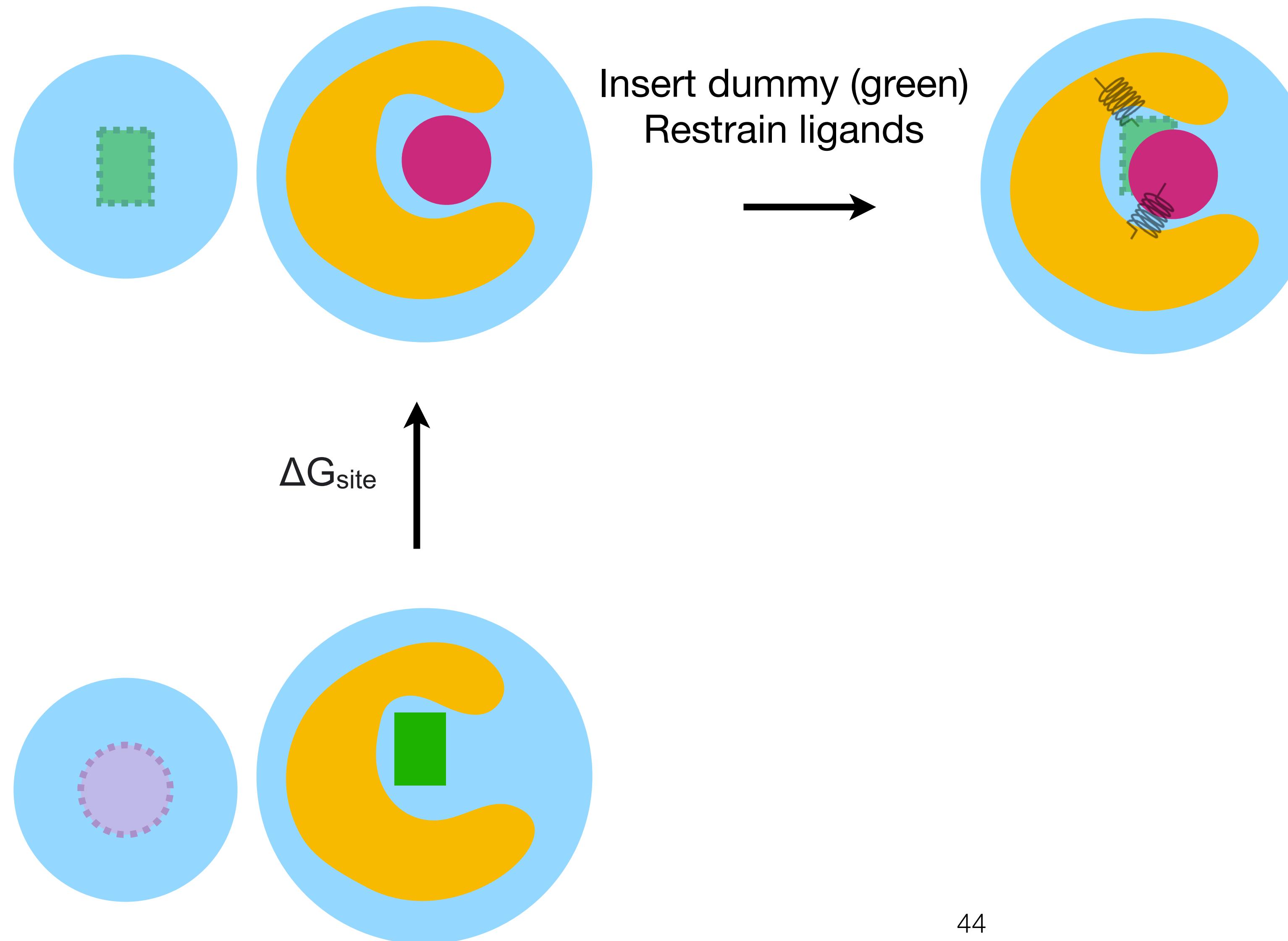


Enhanced sampling methods for binding free energies

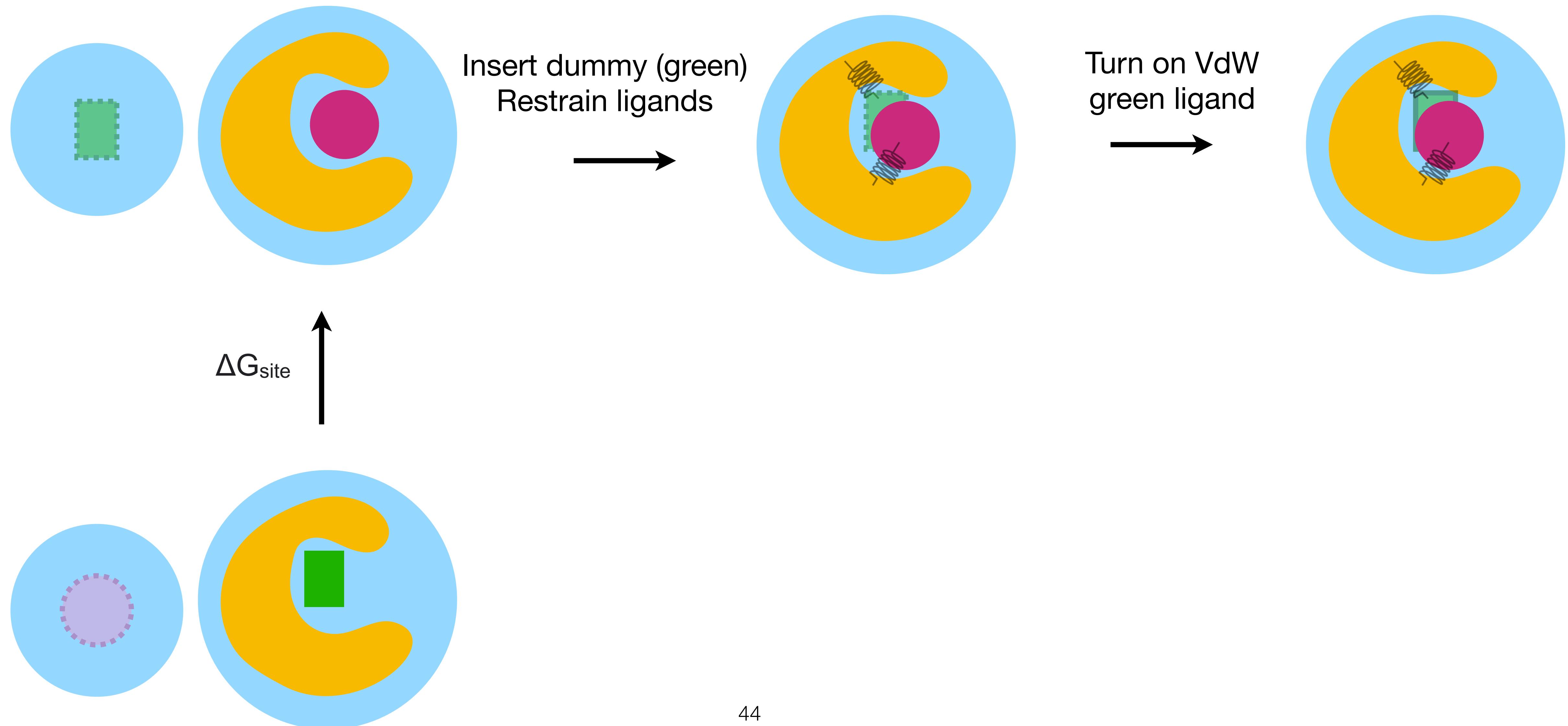
Separated Topologies approach does
RBFE by running two ABFE in opposite directions



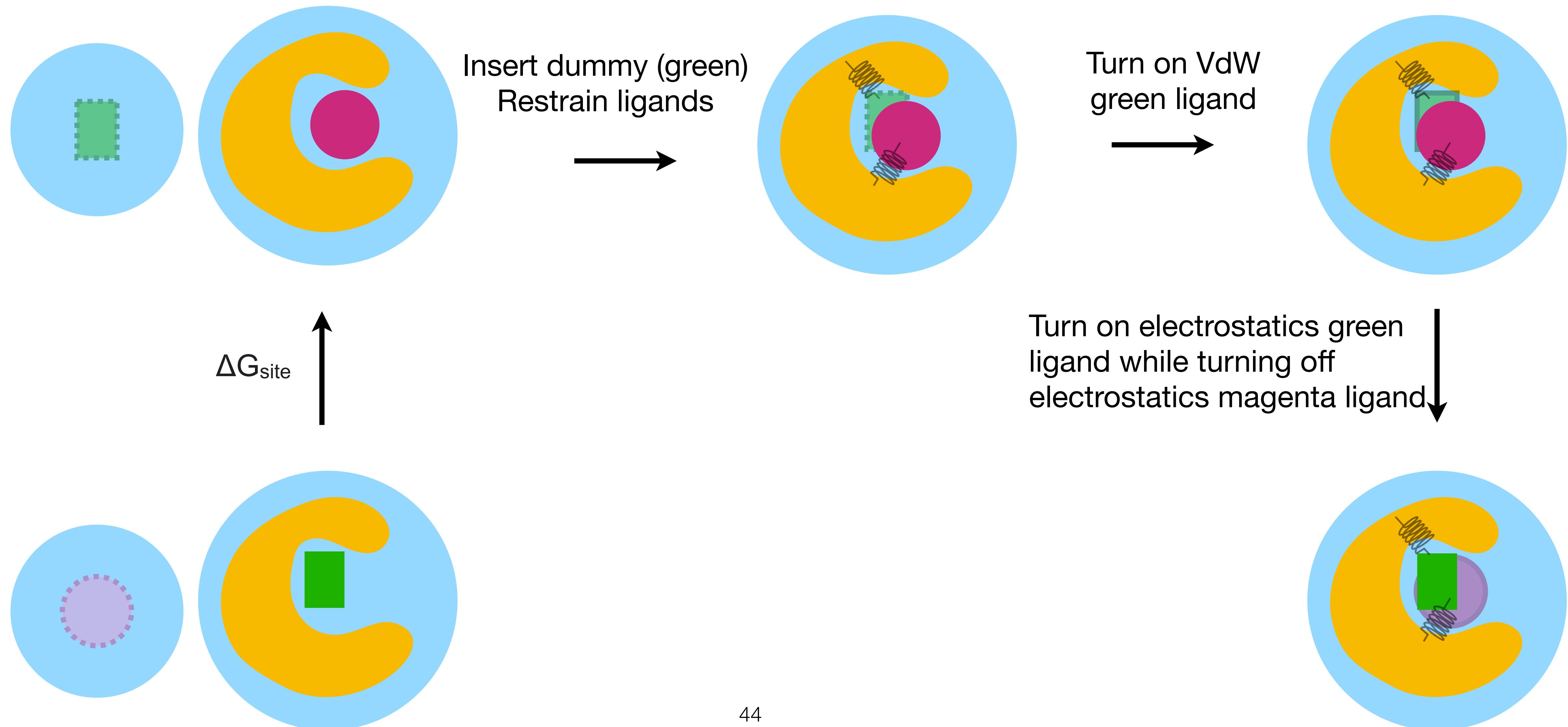
Separated Topologies approach does RBFE by running two ABFE in opposite directions



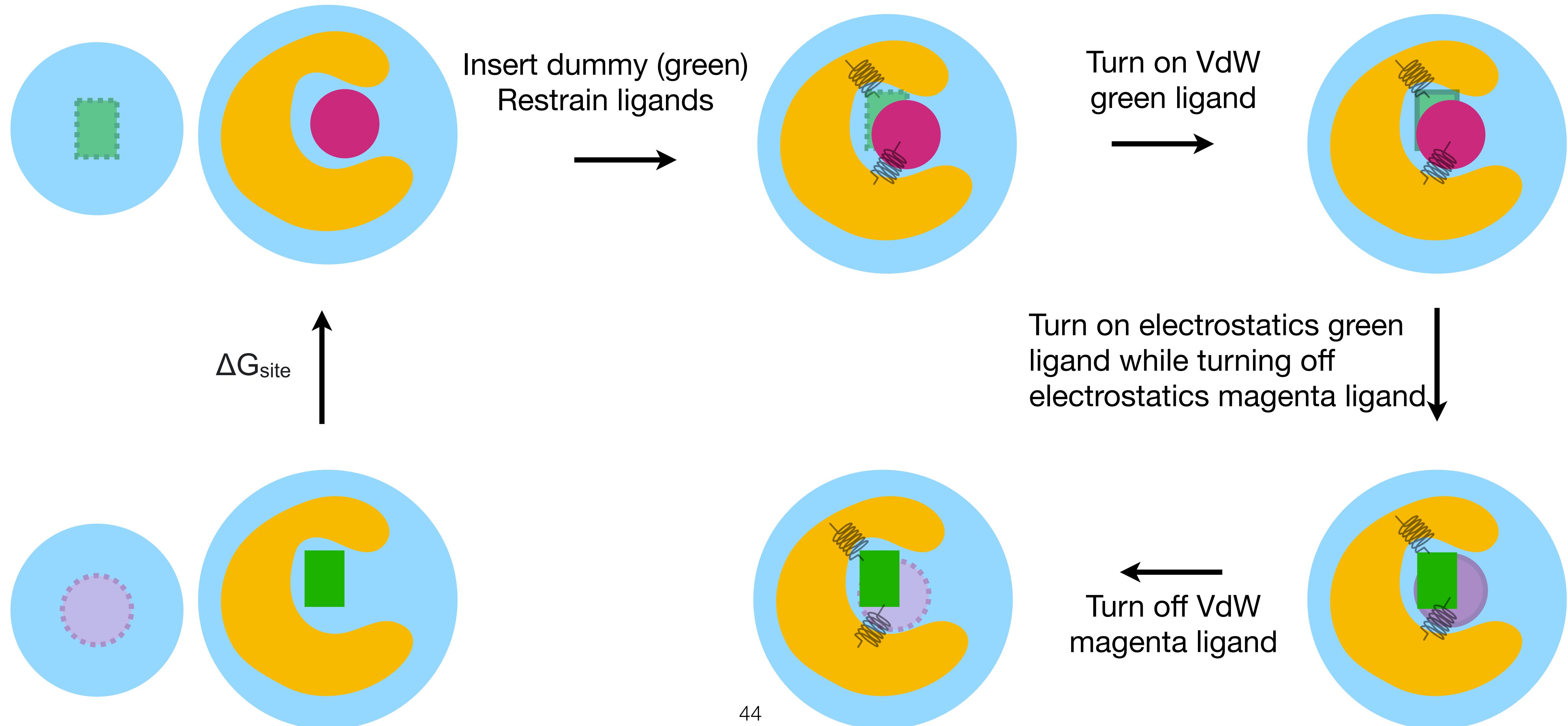
Separated Topologies approach does RBFE by running two ABFE in opposite directions



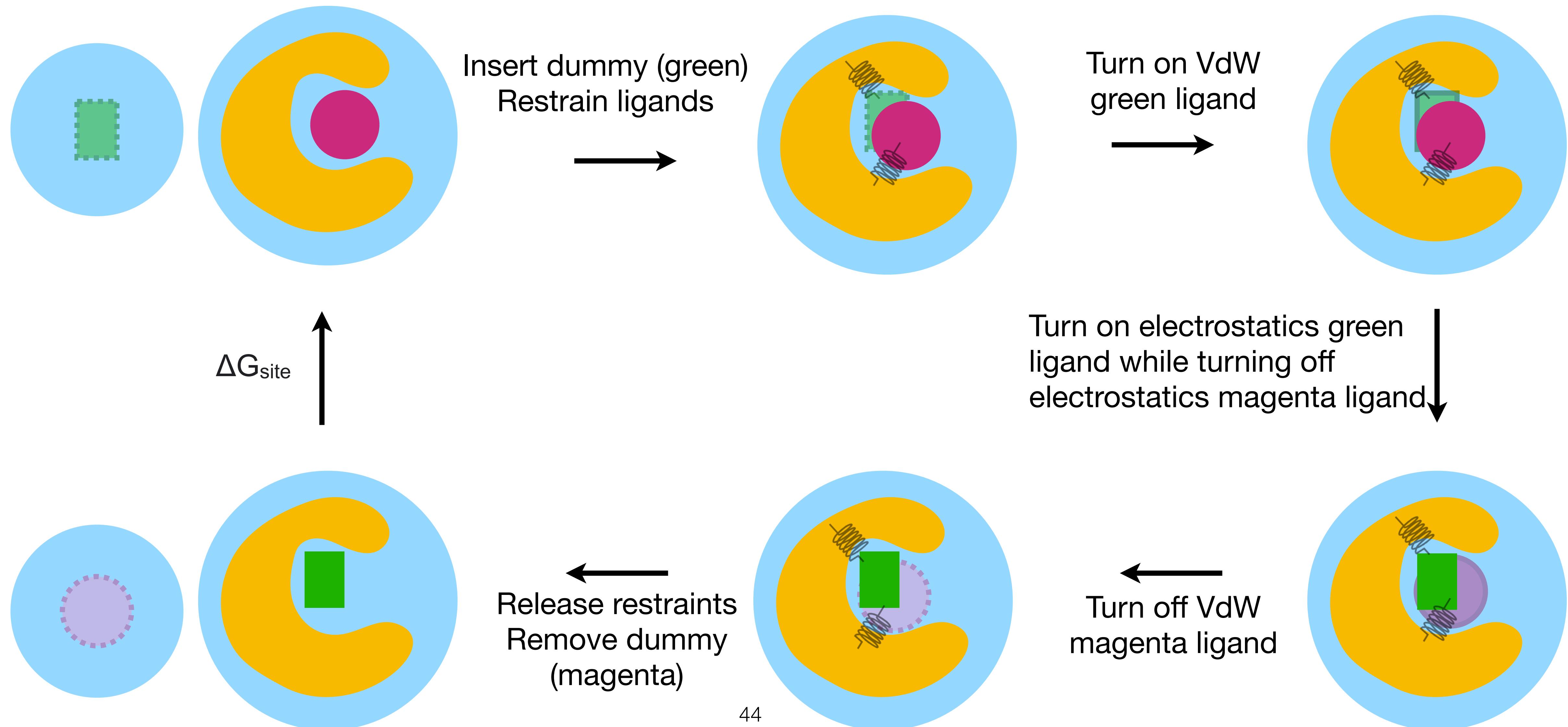
Separated Topologies approach does RBFE by running two ABFE in opposite directions



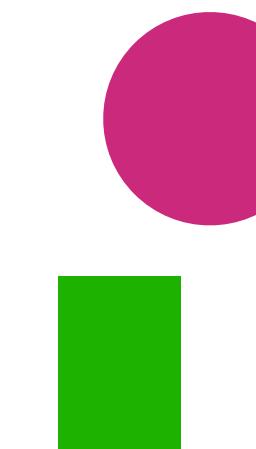
Separated Topologies approach does RBFE by running two ABFE in opposite directions



Separated Topologies approach does RBFE by running two ABFE in opposite directions

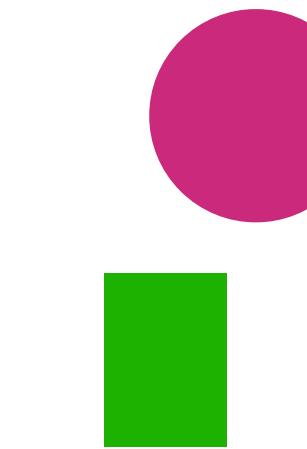


Separated Topologies approach combines benefits of RBFE and ABFE

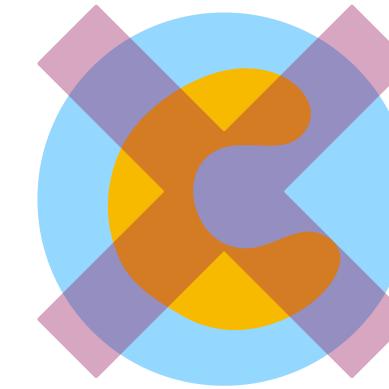


- No restriction on ligand structure
- No restriction on ligand binding mode

Separated Topologies approach combines benefits of RBFE and ABFE

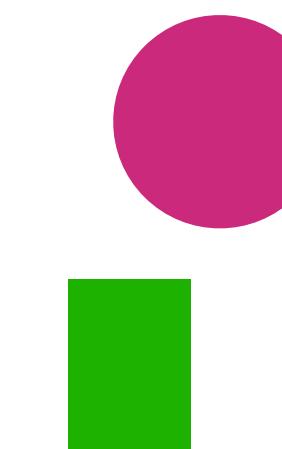


No restriction on ligand structure
No restriction on ligand binding mode

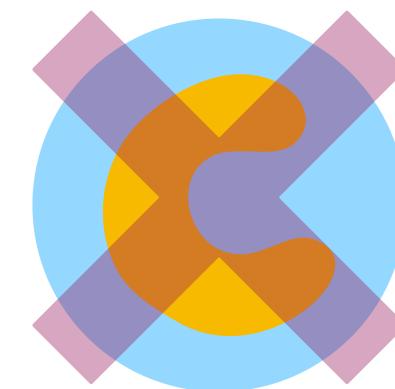


Sampling of the apo system not necessary

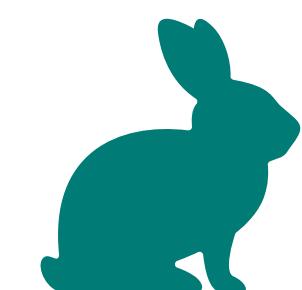
Separated Topologies approach combines benefits of RBFE and ABFE



No restriction on ligand structure
No restriction on ligand binding mode



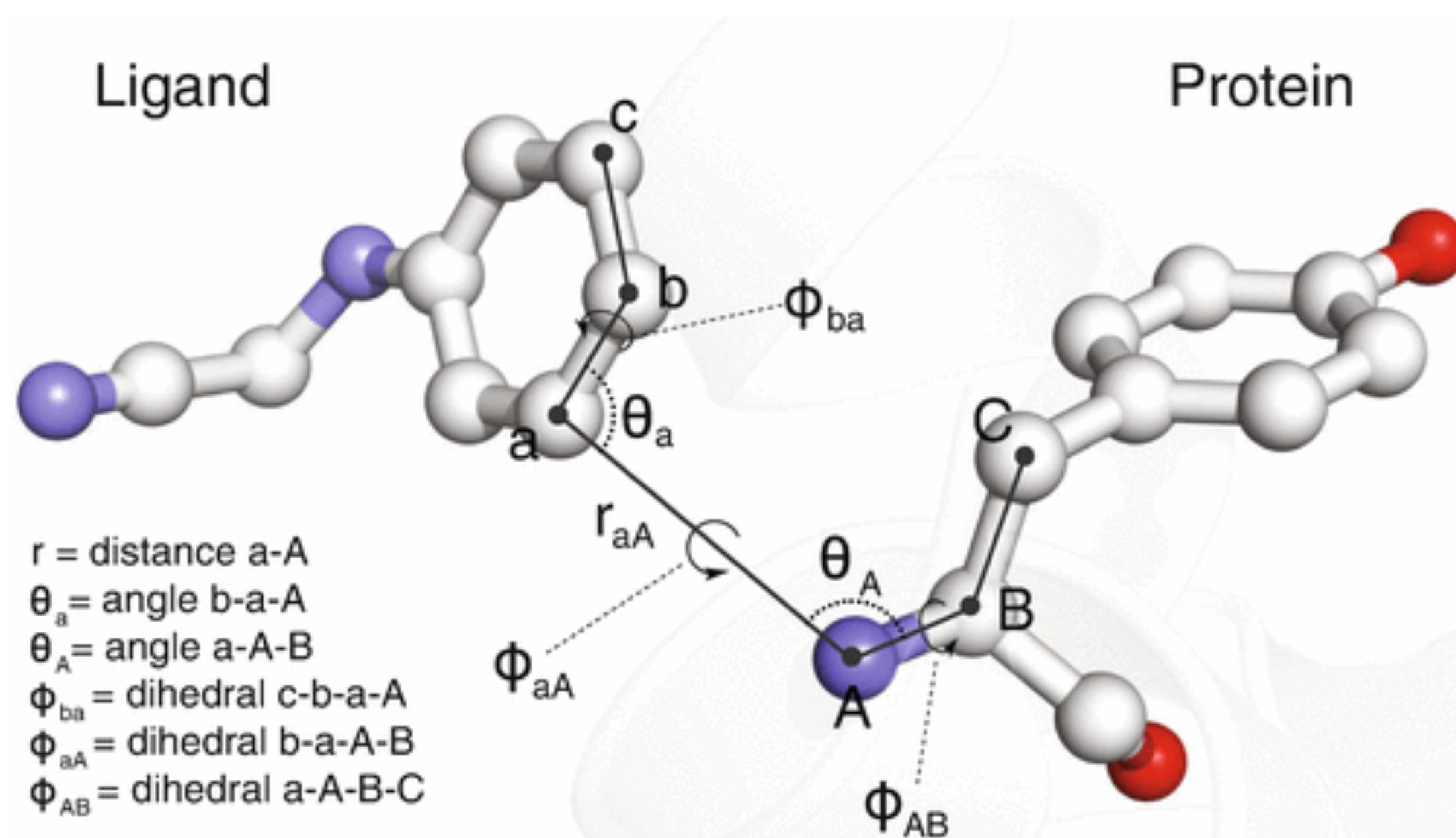
Sampling of the apo system not necessary



Computational efficiency still has to be investigated
but will probably be somewhere between RBFE and ABFE

We tested different types of restraints and decided to use orientational (Boresch-style) restraints

Restraints are defined between 3 protein and 3 ligand atoms



Other restraints tested:

- Position restraints
- Restraining the ligands to the center of the binding site using a virtual site
- Restraining the two ligands to each other using distance restraints

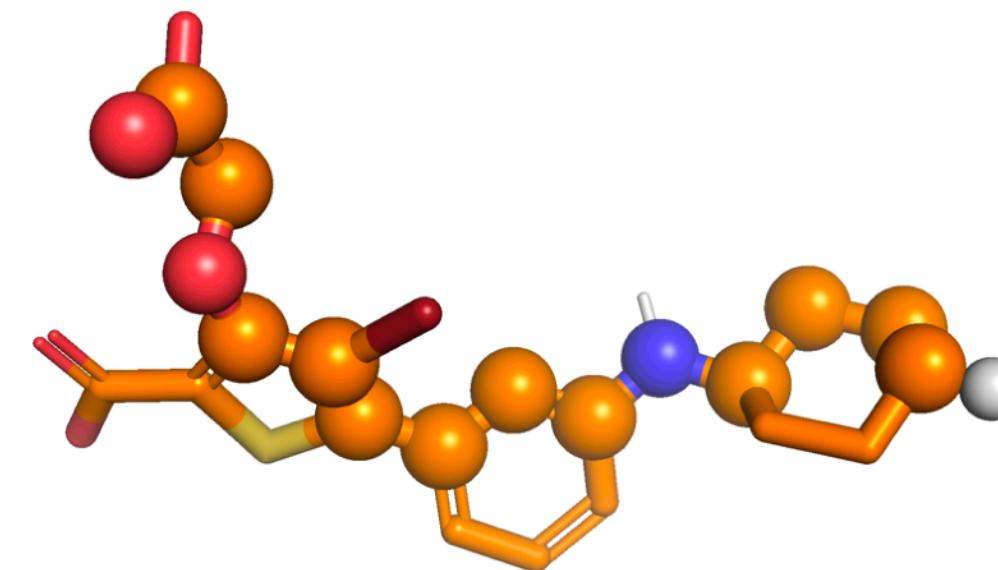
We found Boresch-style restraints to be most efficient.

We developed heuristics for automatically picking suitable atoms for Boresch-style restraints

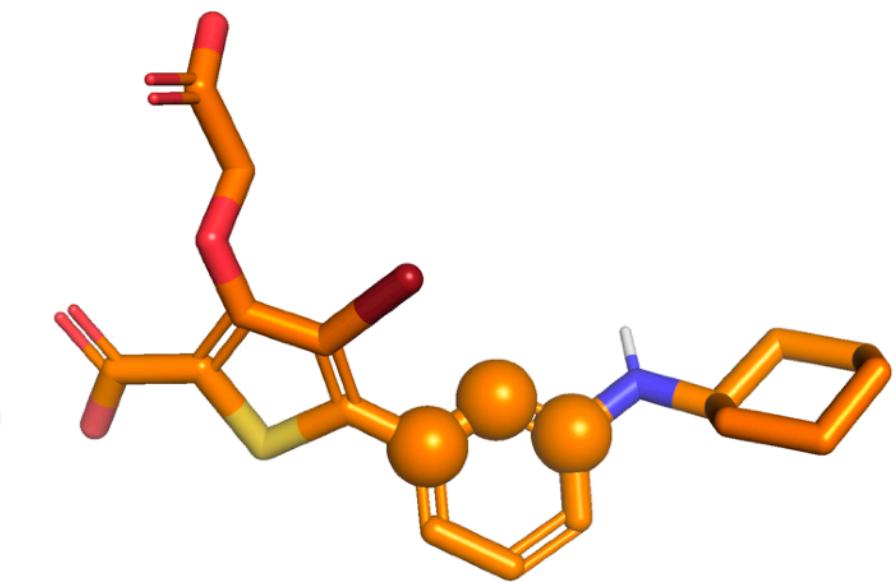
Developed robust new approaches for picking ligand atoms for restraints:

- Pick central atoms by taking the longest path and then picking ring atoms closest to the middle of that path
- If trajectories are available, exclude atoms with a large RMSF

Longest path



Central atoms



Select stable protein atoms:

- Backbone atoms in helix or beta sheet

Tested and ensured that these approaches successfully pick restraint atoms in all cases across the 20 targets in the protein ligand benchmark datasets (Hahn et al.)

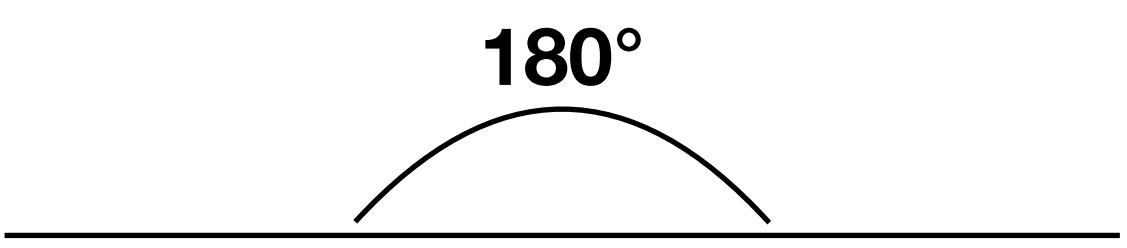
Automatic selection as well as user defined selection of atoms possible

Code available on GitHub (<https://github.com/MobleyLab/SeparatedTopologies/tree/latest>)

We identified criteria necessary to ensure stable simulations

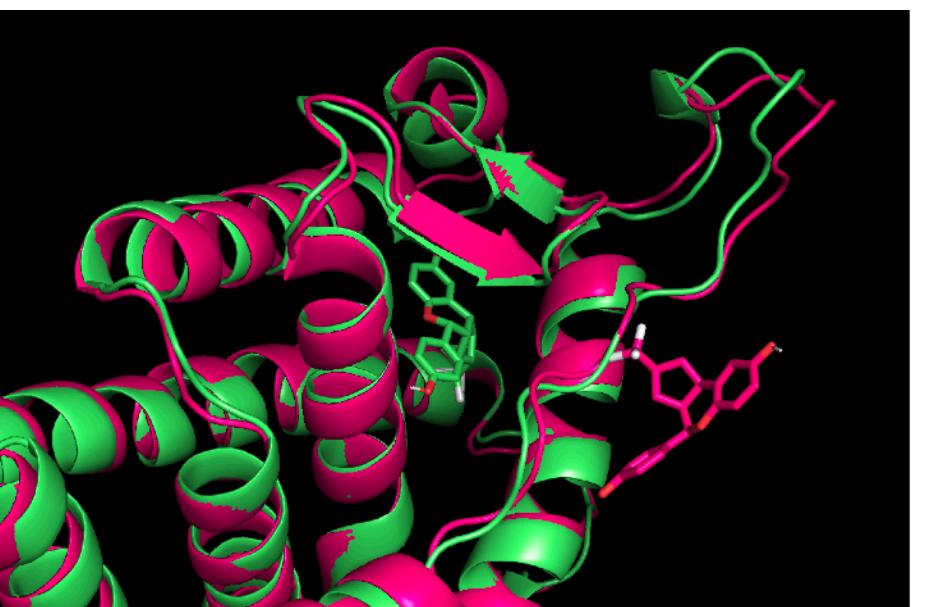
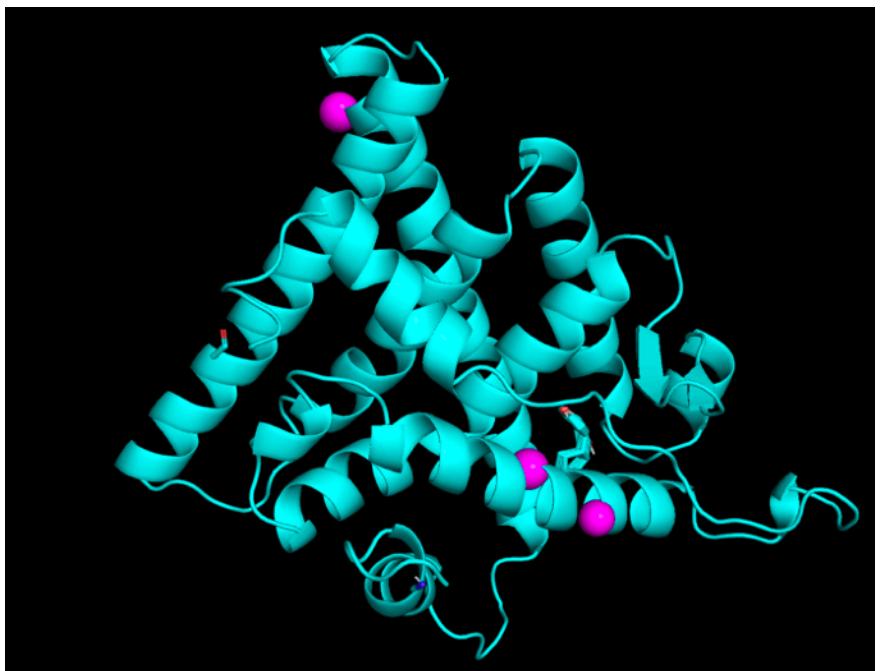
Identified restraint values that lead to numerical instabilities:

- Simulations crash if angles are restrained too close to 0° and 180°

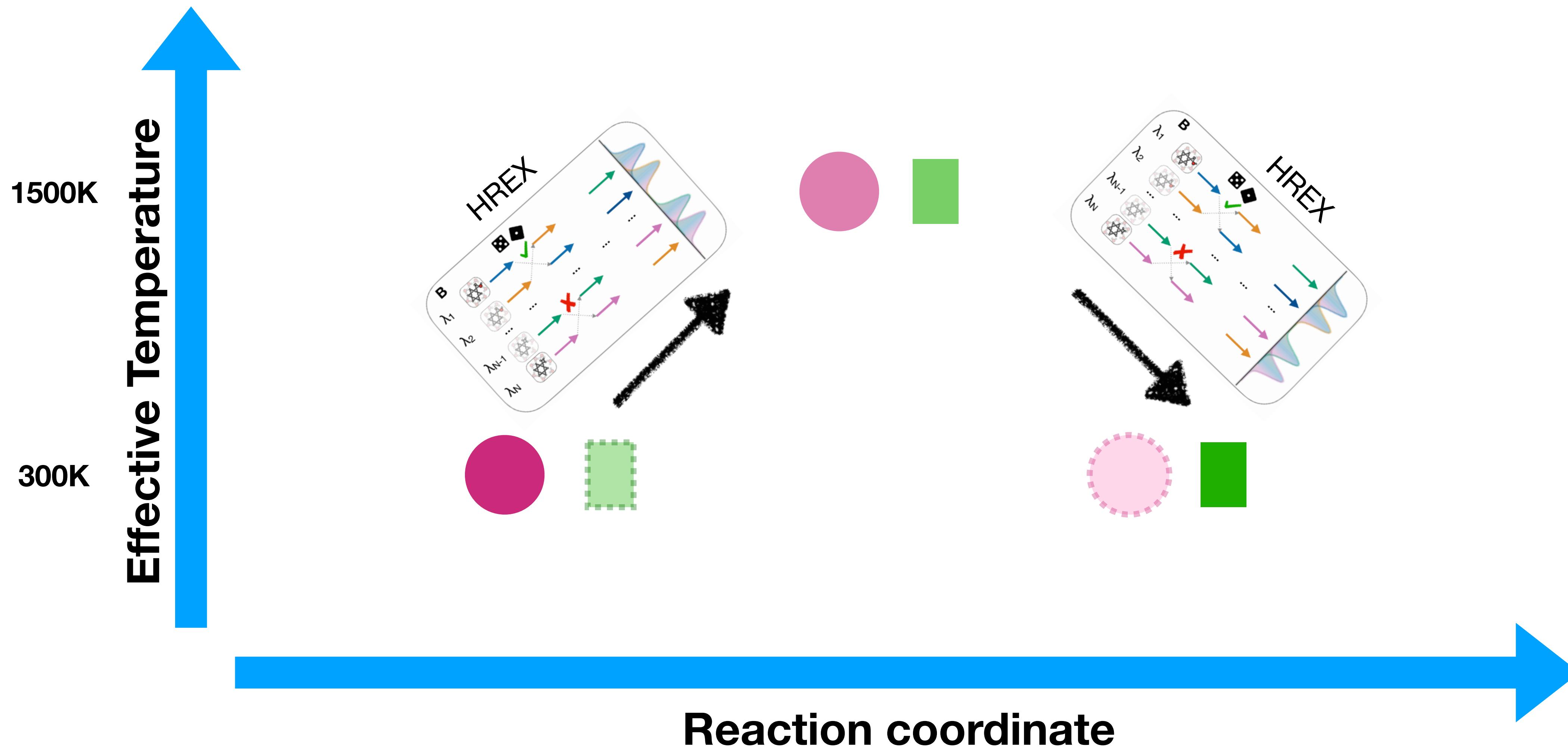


Identified problems due to periodic boundary conditions:

- If protein atoms are too far from each other, restraints can be unstable due to PBC and the minimum-image convention
- This can lead to the ligand leaving the binding site in the alchemical intermediates
- To avoid this, protein atoms have to be less than half the box size away from one another



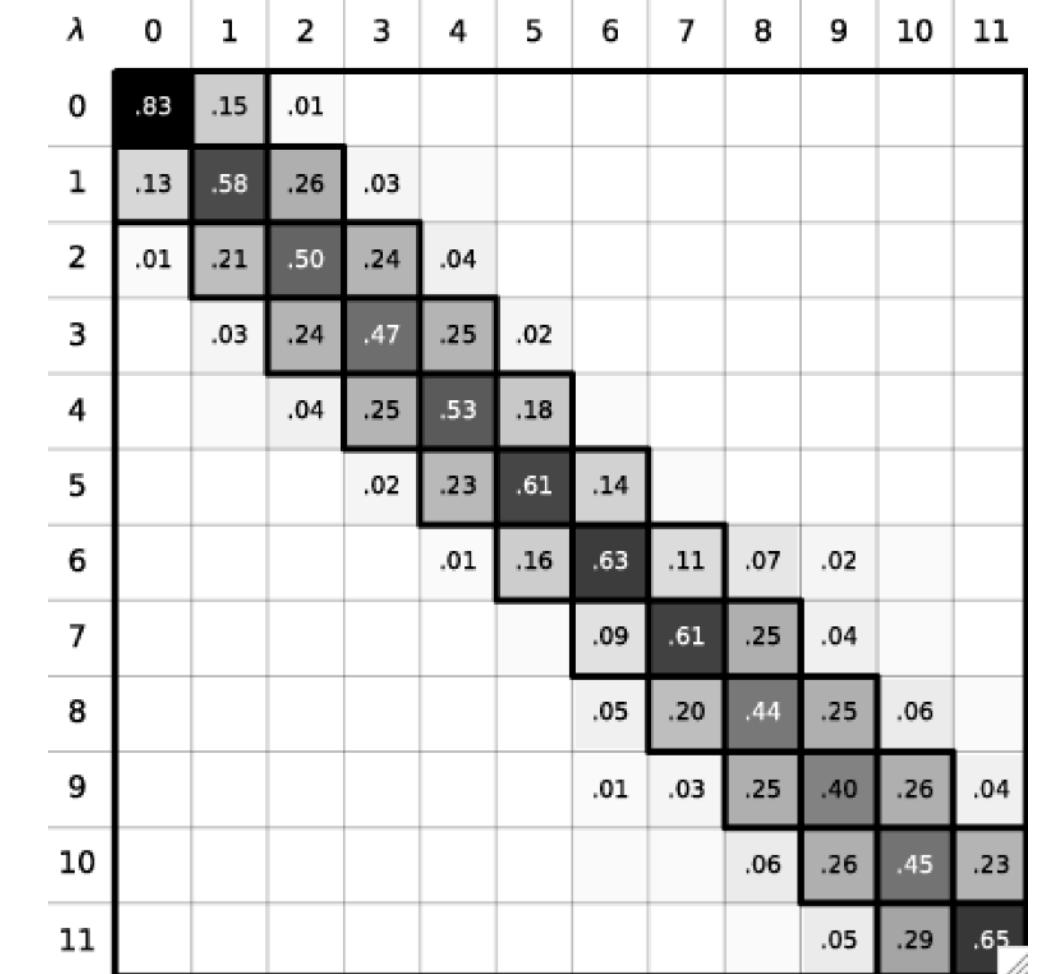
We use replica exchange with solute tempering (REST) to enhance sampling



We tested different protocols to improve the efficiency of the approach

Example from TYK2 dataset

| protocol | ddG in kcal/mol |
|---|-----------------|
| No enhanced sampling, 45 lambda windows | -1.2 +/- 0.3 |
| HREX, 45 lambda windows | -1.6 +/- 0.2 |
| REST scaling, 20 lambda windows | -1.3 +/- 0.2 |



We were able to reduce the number of lambda windows (in the binding site) from 45 to 20

- Finding an efficient alchemical pathway by looking at the overlap matrix

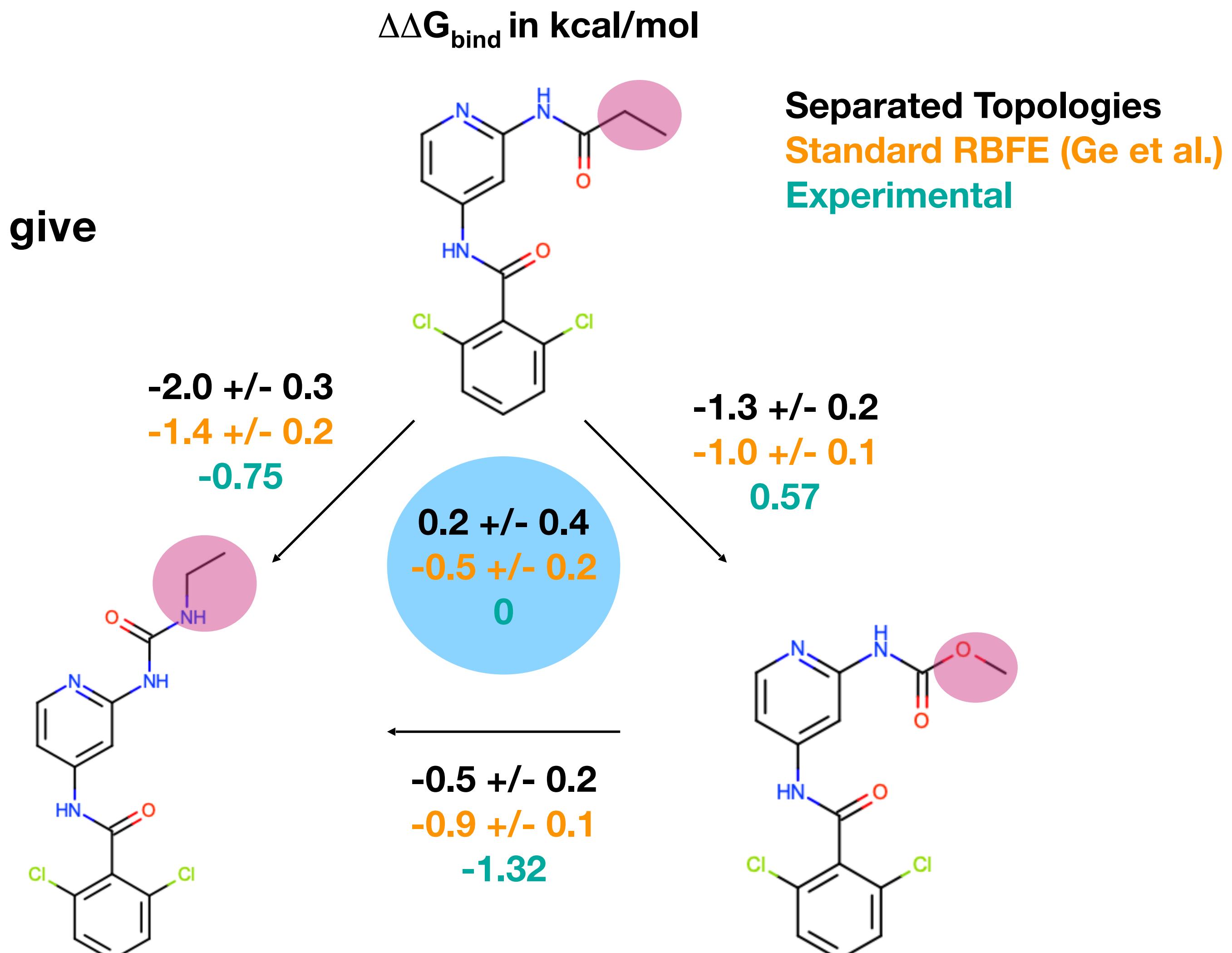
We implemented Replica Exchange with solute tempering (REST) scaling

- Standard deviation across three replicates was reduced using enhanced sampling

1) TYK2: We are able to obtain good convergence with the Separated Topologies approach

Separated Topologies and standard RBFE give comparable results

Good cycle closure suggests convergence



2) Estrogen receptor alpha: Scaffold hopping transformations

$\Delta\Delta G_{\text{bind}}$ in kcal/mol

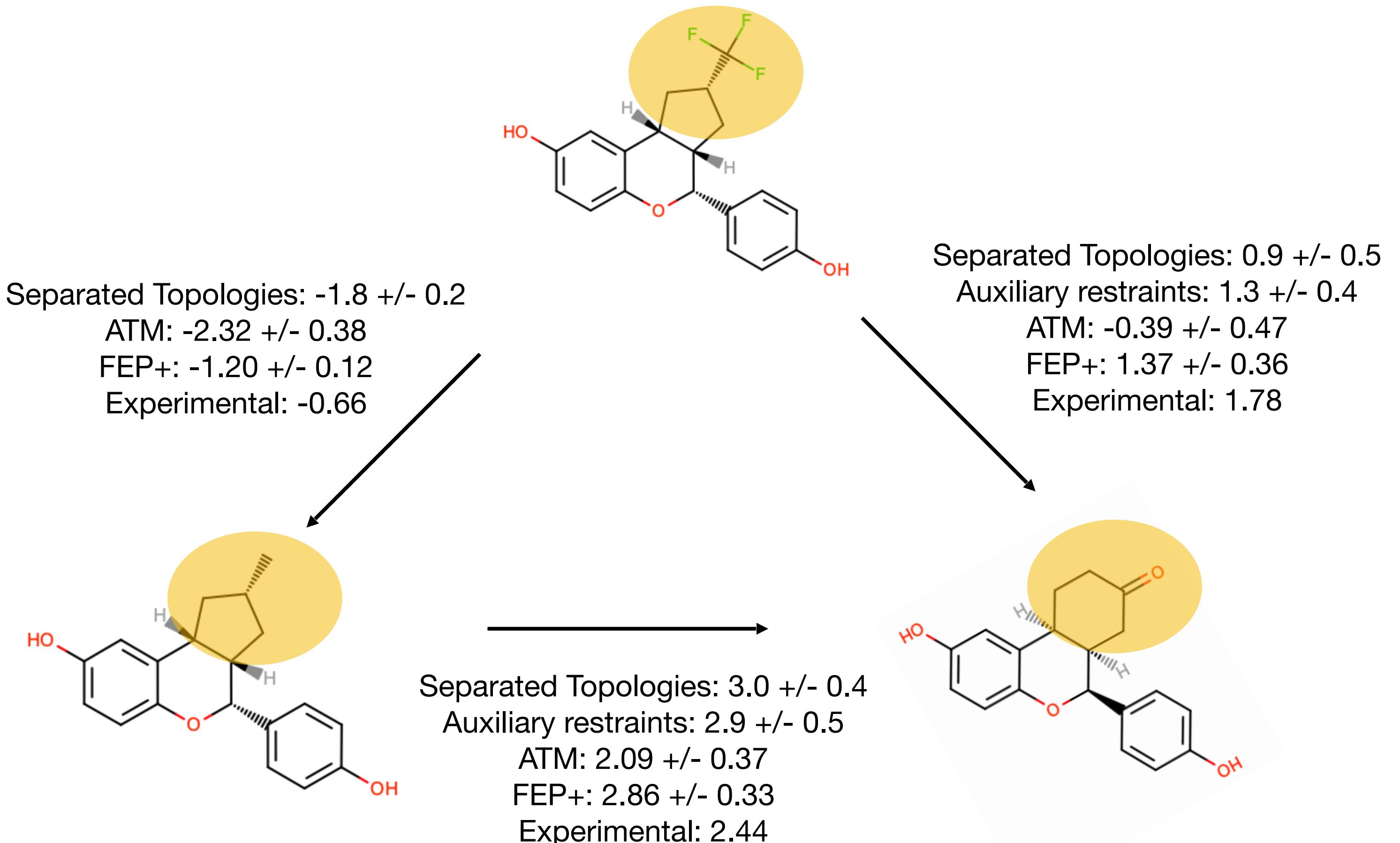
The Separated Topologies and other scaffold hopping approaches give comparable results

Other scaffold hopping approaches:

- Using soft-bond potential (FEP+)
- Using auxiliary restraints
- Alchemical Transfer method (ATM)

We used the same input files and force field parameters provided as in the auxiliary restraints paper

- Force field: GAFF2, Amber19SB, TIP3P



Scaffold Hopping Transformations Using Auxiliary Restraints for Calculating Accurate Relative Binding Free Energies, Zou et al., 2021

Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method, Azimi et al., 2021

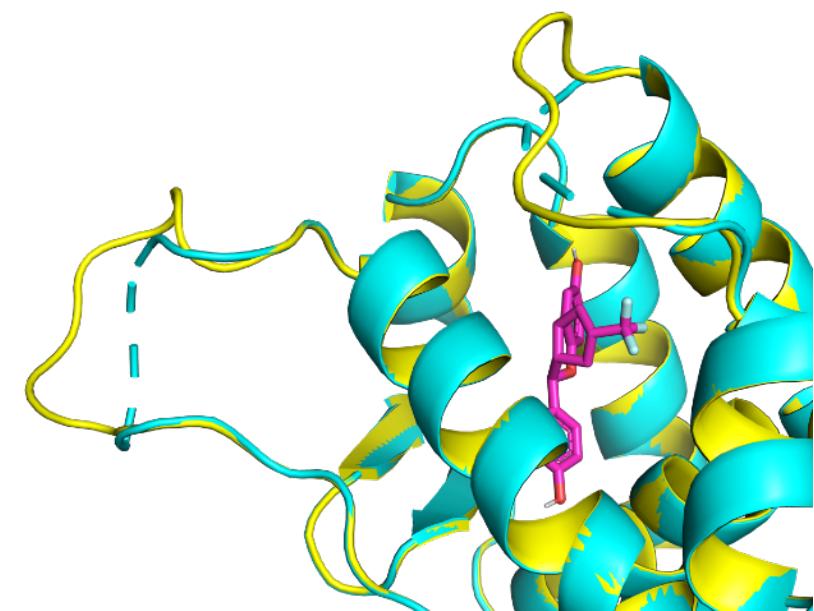
Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery, Wang et al., 2017

Comparison with other methods challenging due to different system preparation by different groups

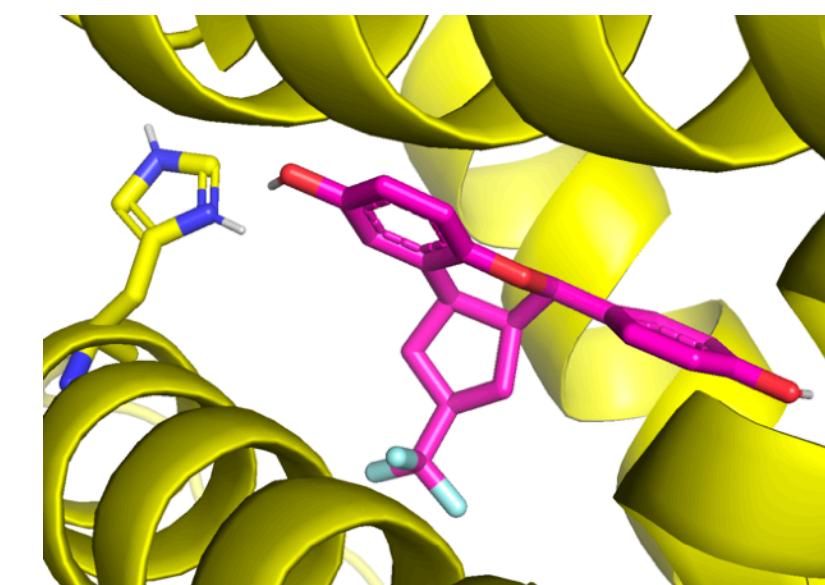
We ran calculations using different input structures:

- Prepared by us using OpenEye Spruce to prepare the protein and model missing loops
- From the publication using auxiliary restraints
- From the publication using the alchemical transfer method

Missing loops modeled in some studies,
but not in others



Different protonation state of histidine residues has
an impact on the predicted binding free energy



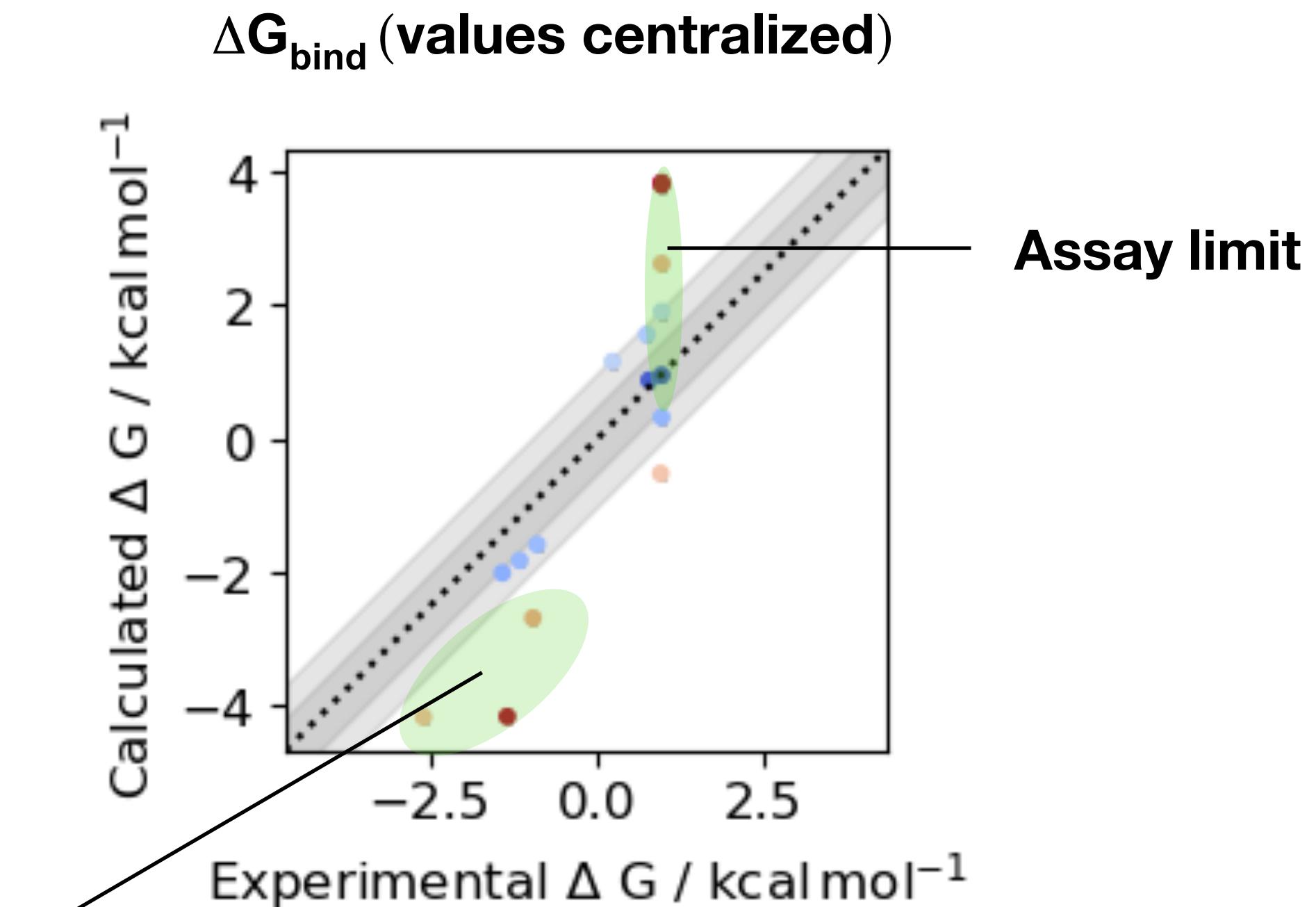
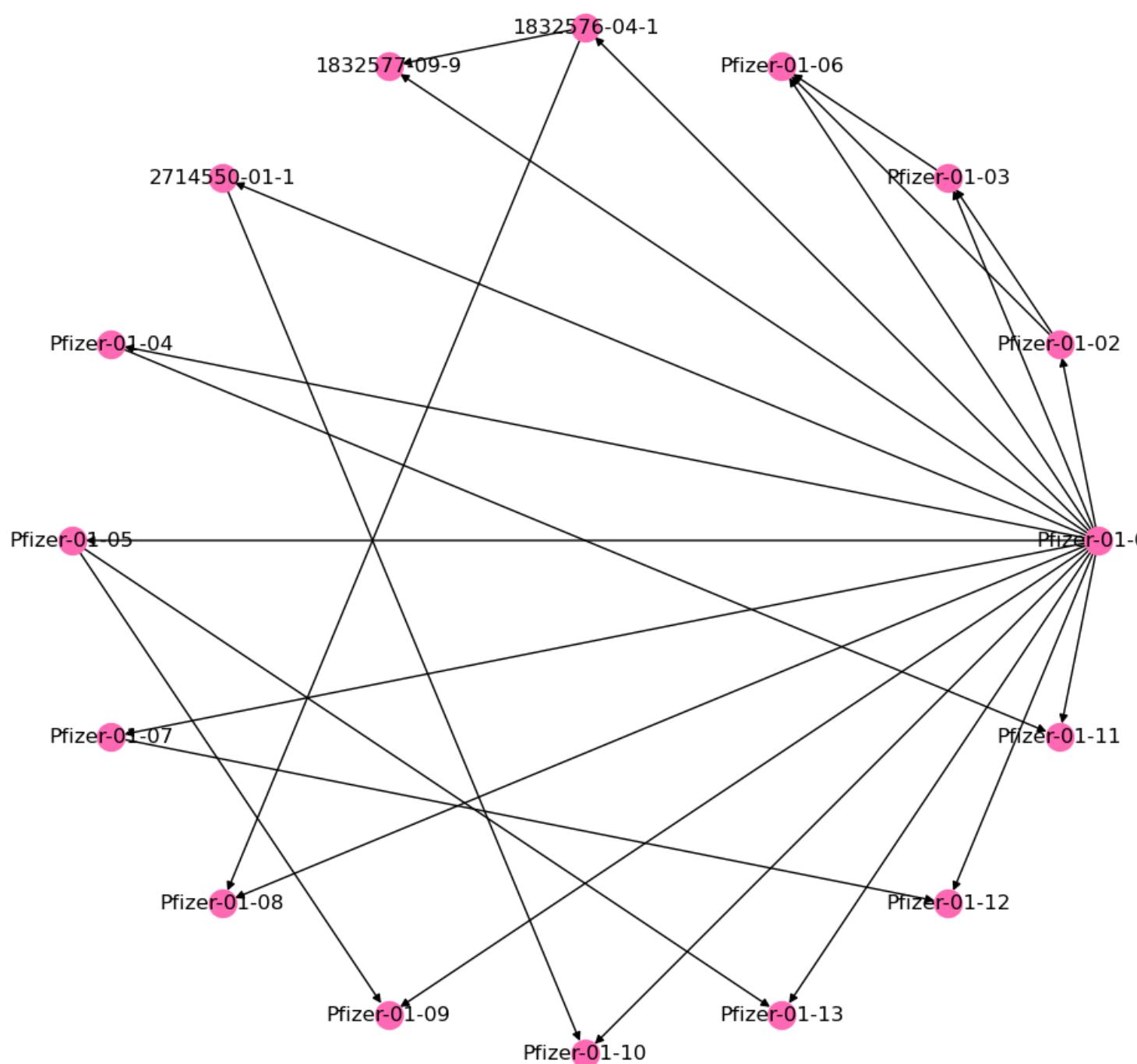
Example impact of protonation state HIS220 in one transformation

- Neutral: $\Delta\Delta G = -1.4 \pm 0.4$
- Protonated (+1): $\Delta\Delta G = -0.2 \pm 0.4$

Results suggest that modeling of missing loops and the protonation state of the histidine residues have an impact on the predicted binding free energy

3) Pfizer dataset: Preliminary data look promising, but outliers still need to be investigated

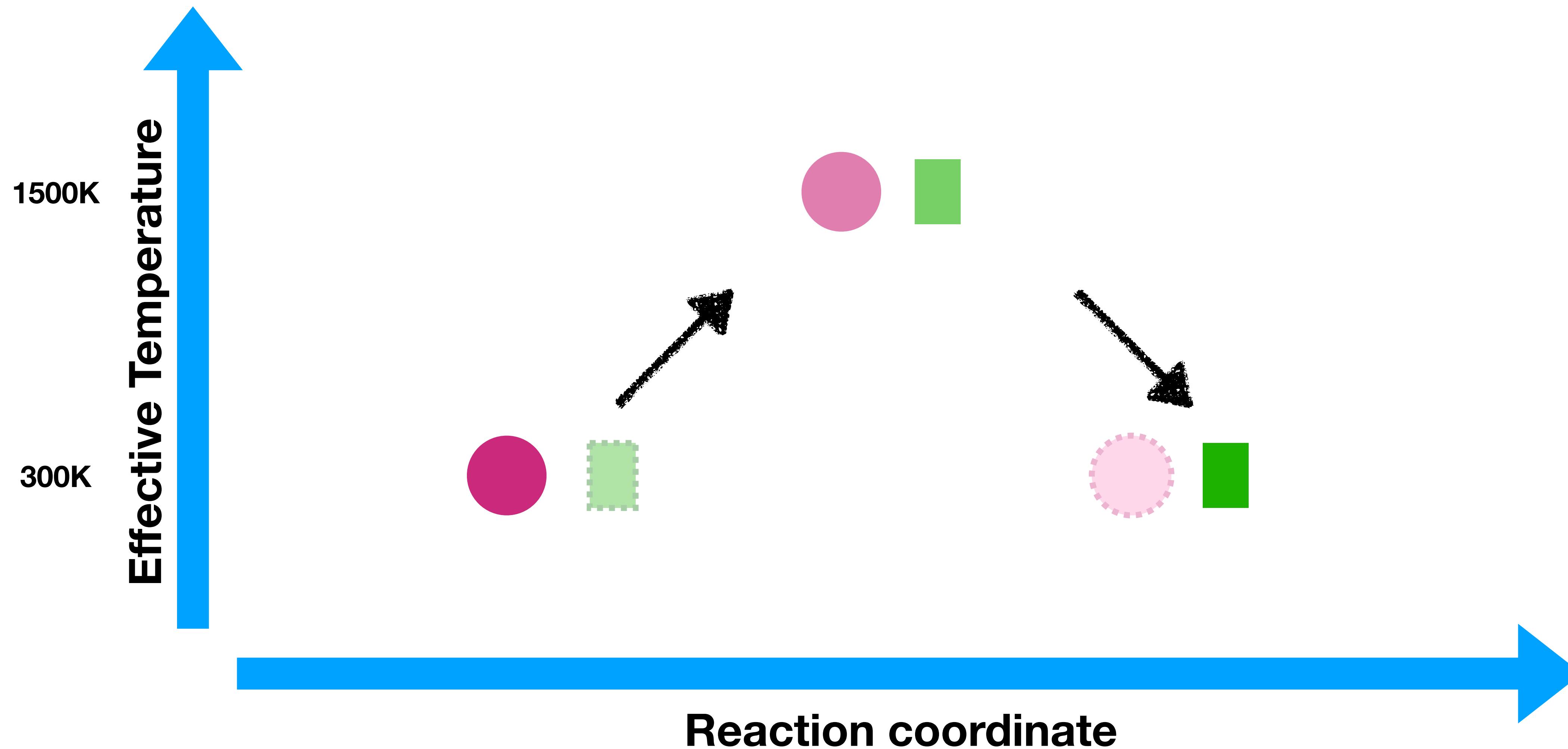
Testing the method on a larger dataset



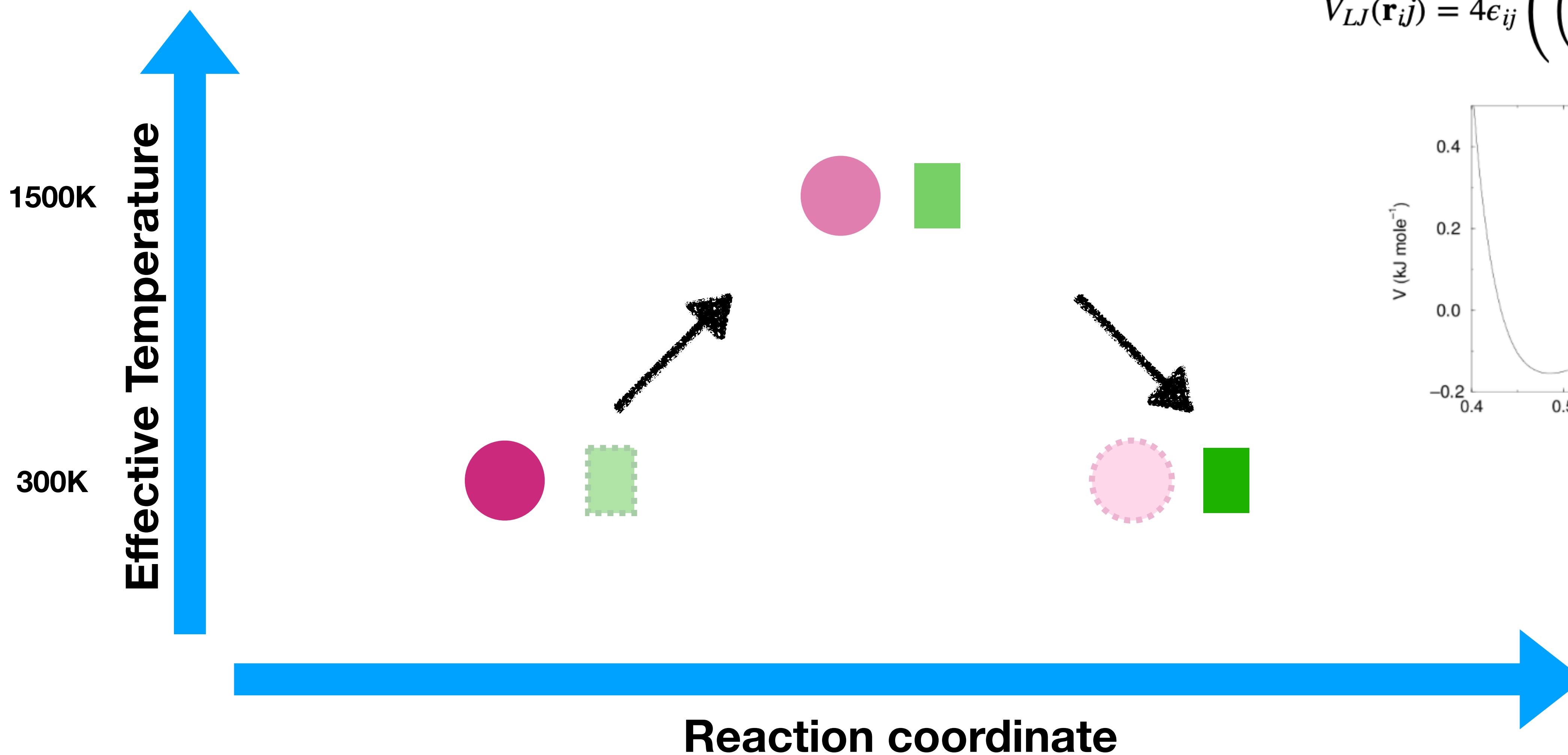
(N = 16)

| | | |
|-------|------|-------------------|
| RMSE: | 1.55 | [95%: 1.06, 1.99] |
| MUE: | 1.27 | [95%: 0.87, 1.74] |
| R2: | 0.77 | [95%: 0.58, 0.91] |
| rho: | 0.88 | [95%: 0.73, 0.95] |

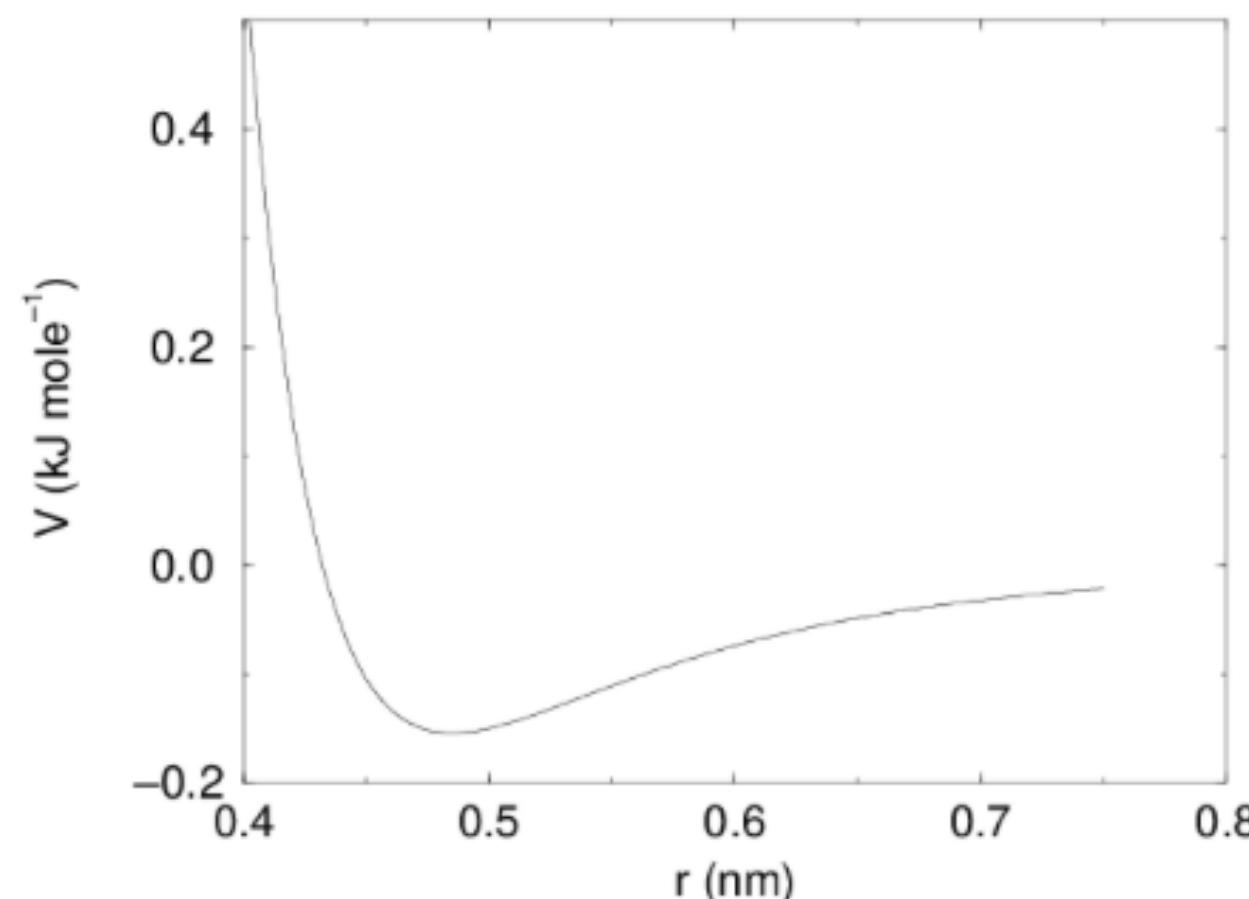
We use Replica exchange with solute tempering (REST) to enhance sampling



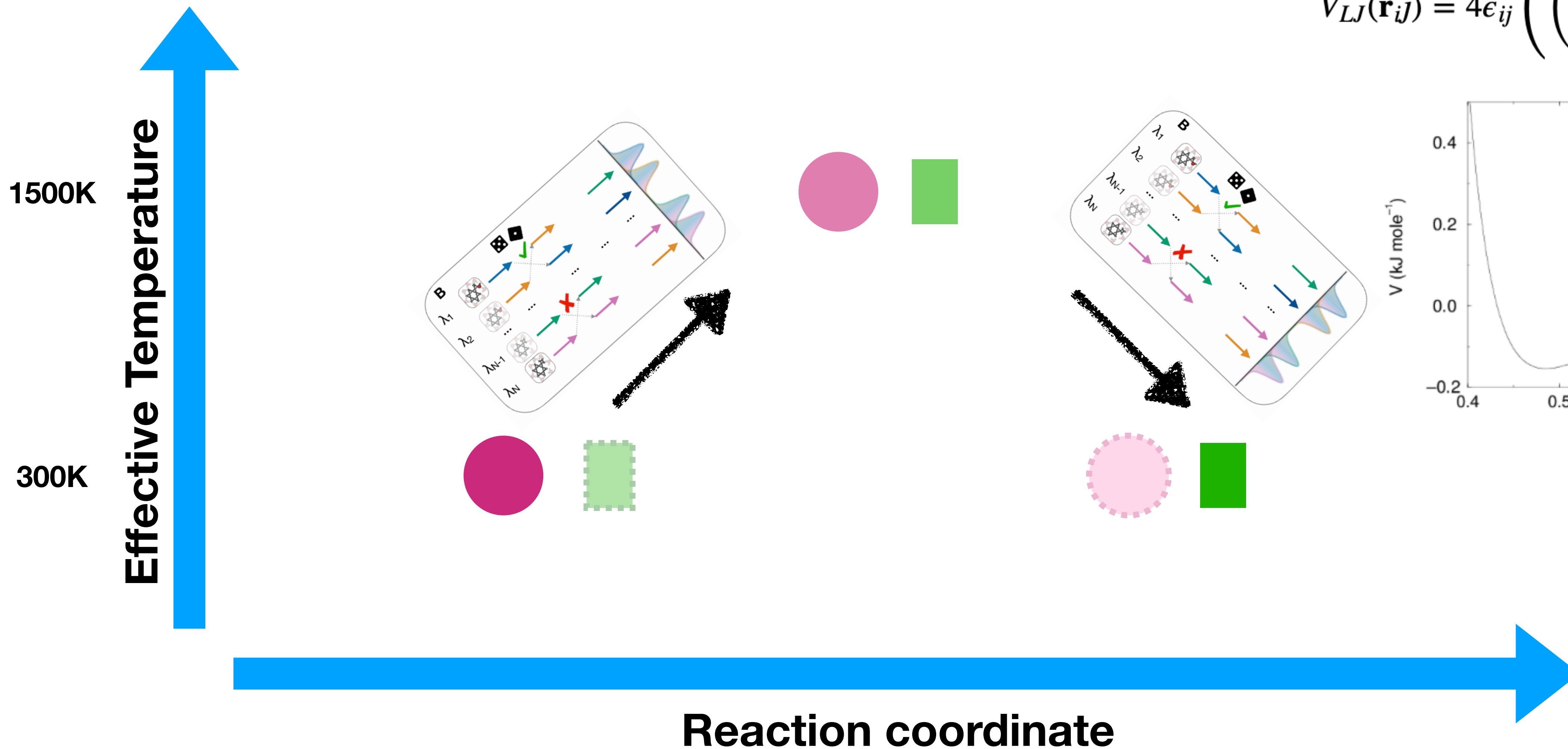
We use Replica exchange with solute tempering (REST) to enhance sampling



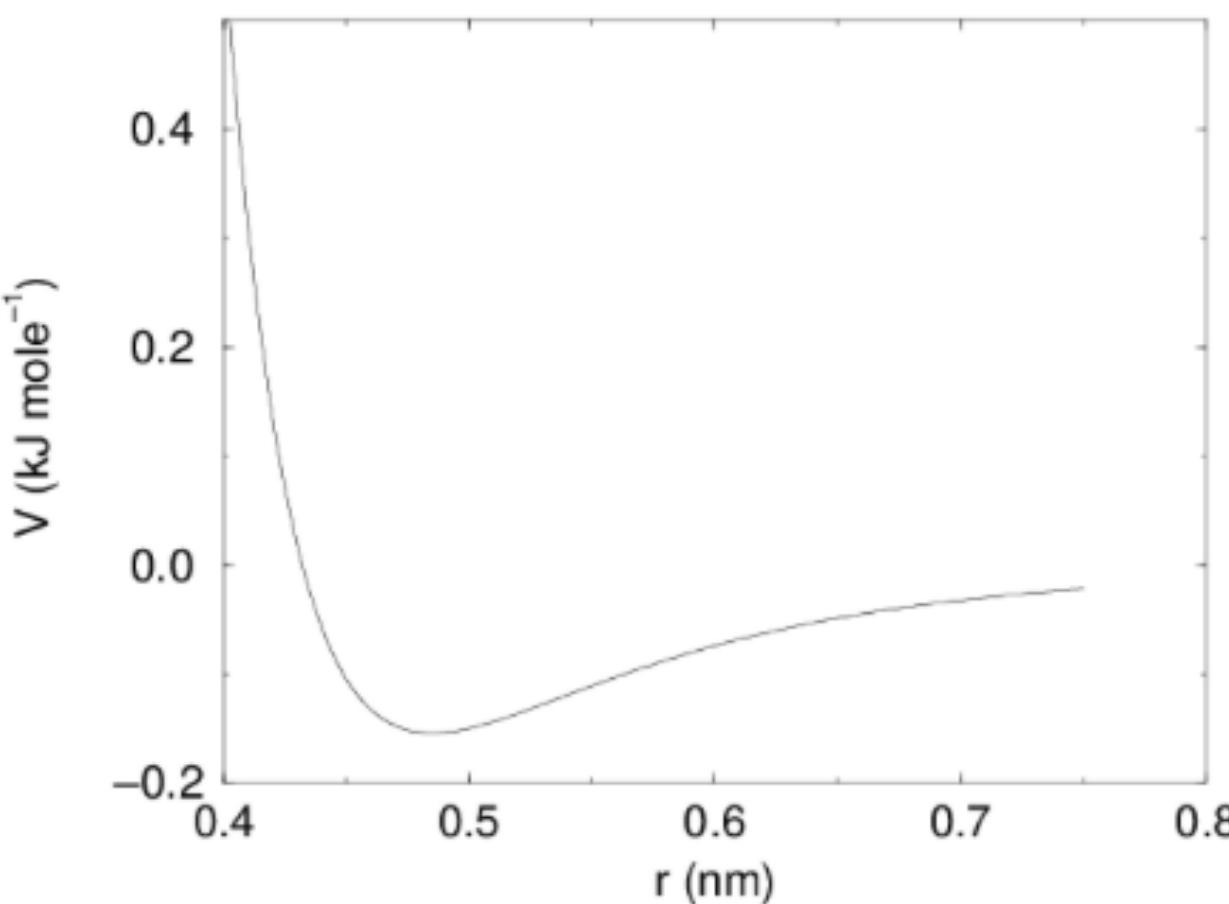
$$V_{LJ}(\mathbf{r}_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$



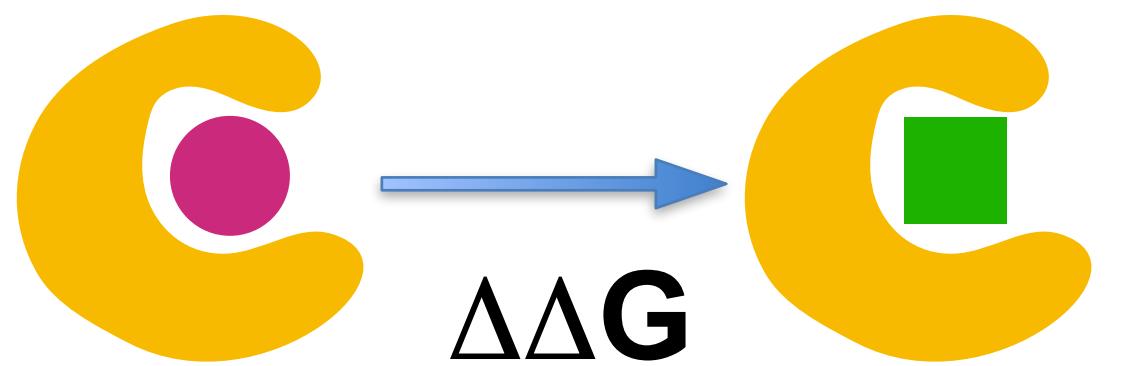
We use Replica exchange with solute tempering (REST) to enhance sampling



$$V_{LJ}(\mathbf{r}_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$

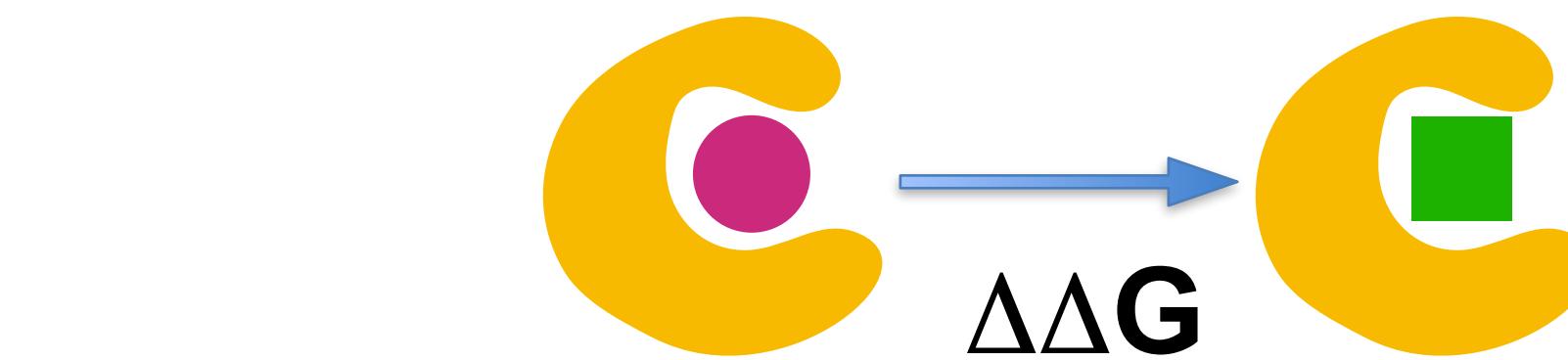


Outlook

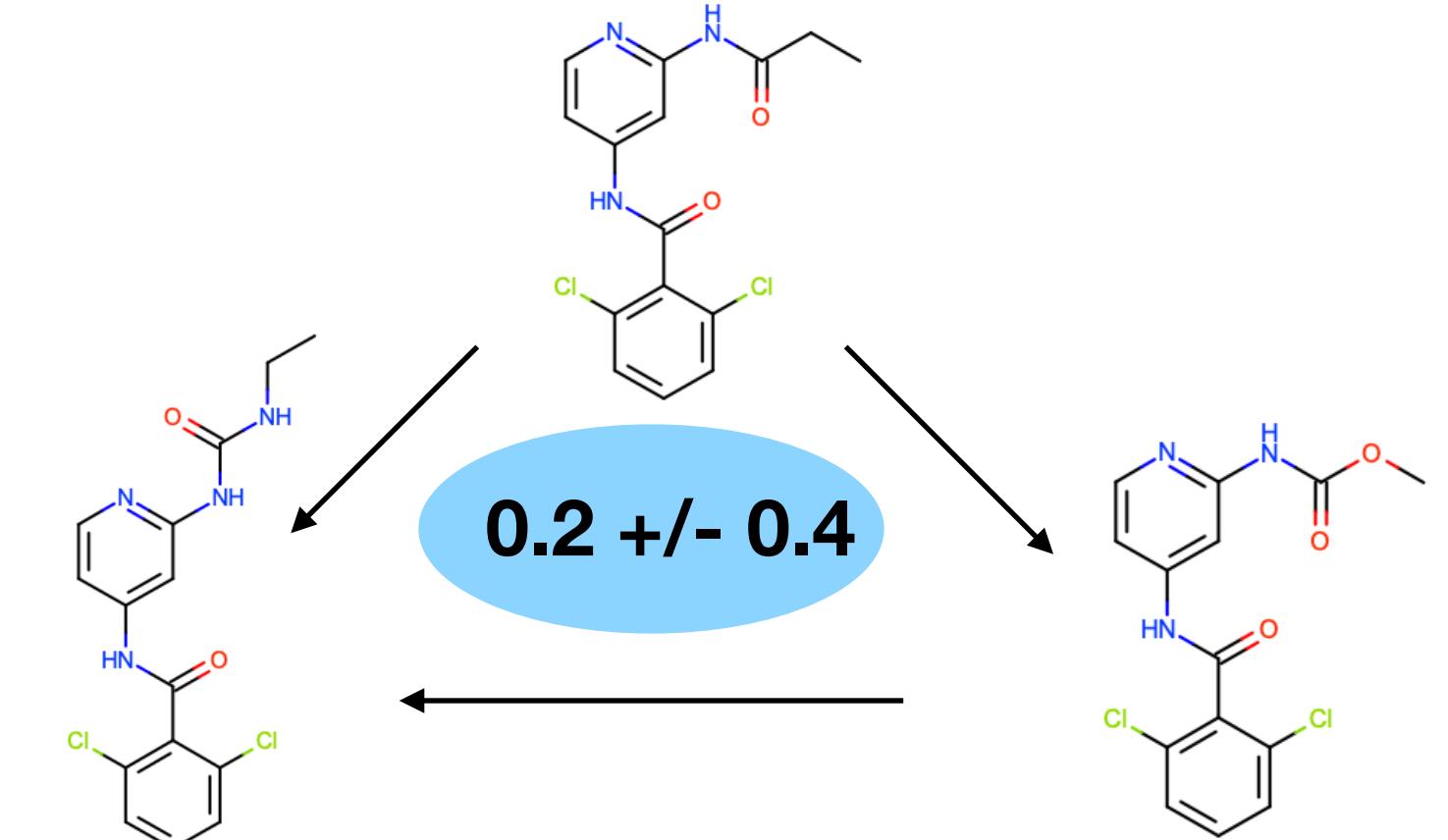


We implemented an alternative approach to RBFE, Separated Topologies, which allows the comparison of structurally diverse ligands

Outlook

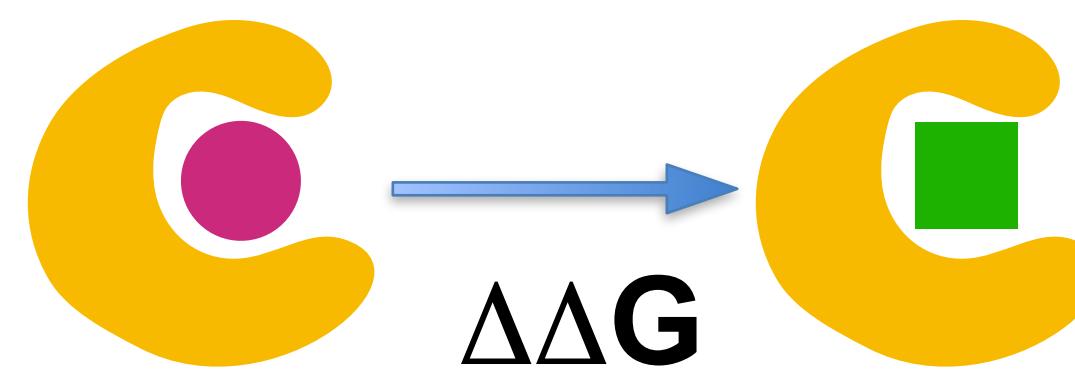


We implemented an alternative approach to RBFE, Separated Topologies, which allows the comparison of structurally diverse ligands

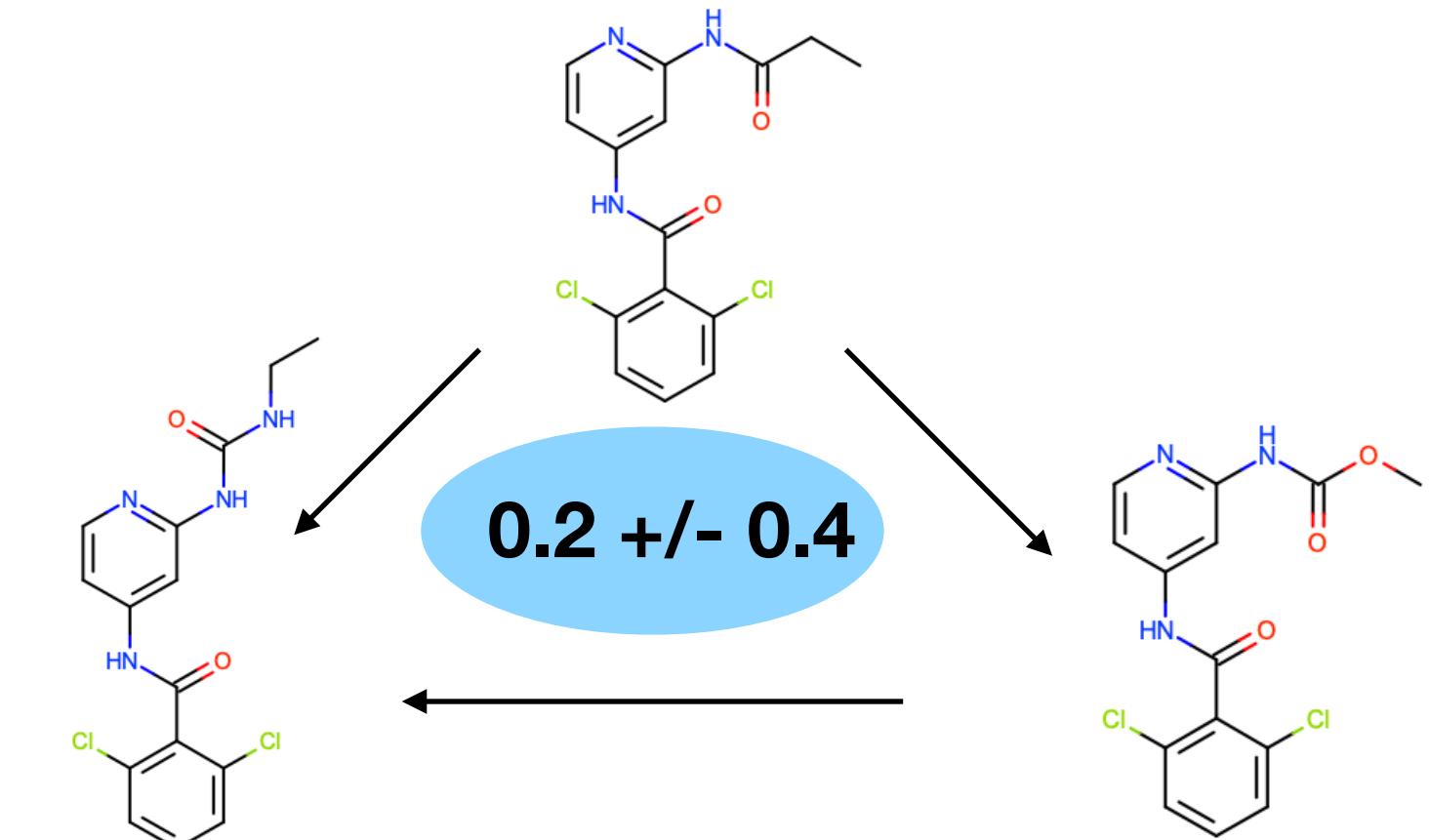


Good convergence and comparable results to standard RBFE were obtained on a TYK2 test system

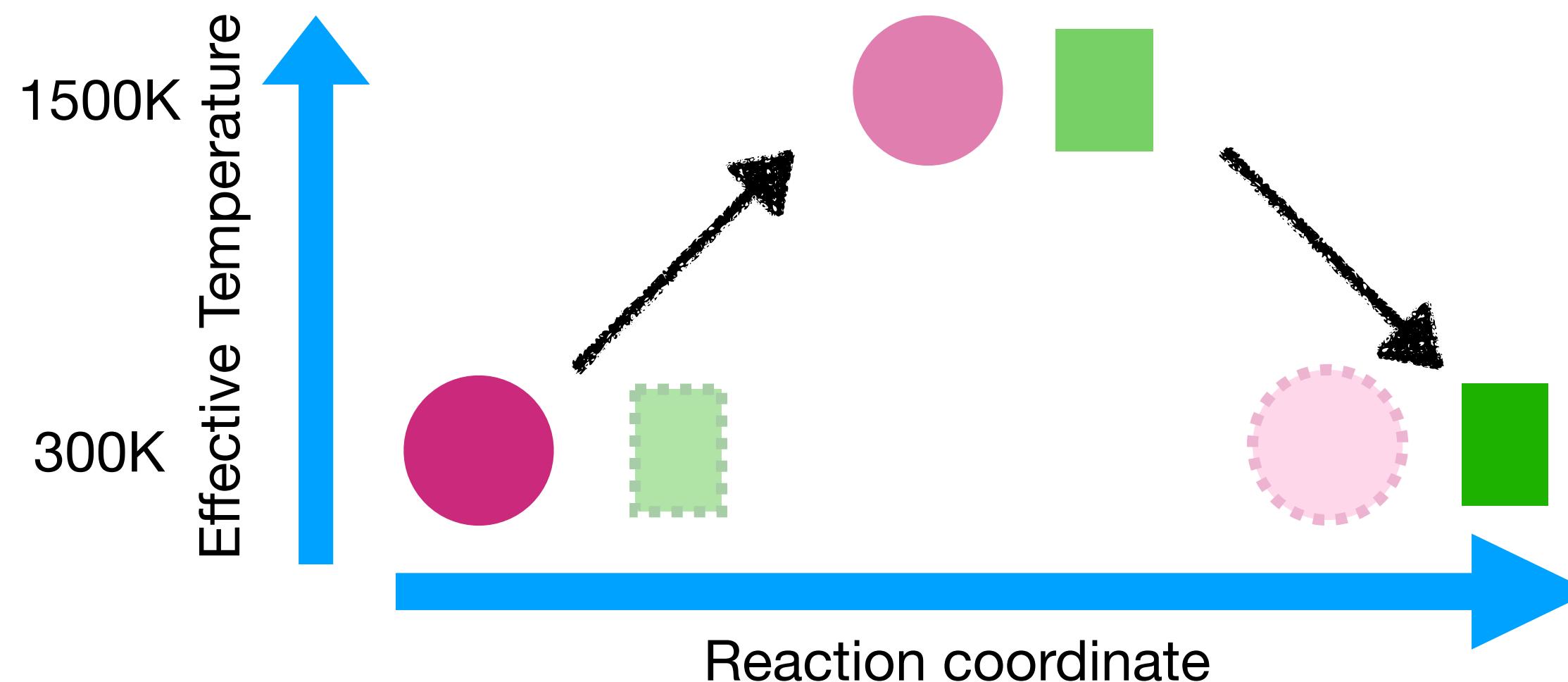
Outlook



We implemented an alternative approach to RBFE, Separated Topologies, which allows the comparison of structurally diverse ligands

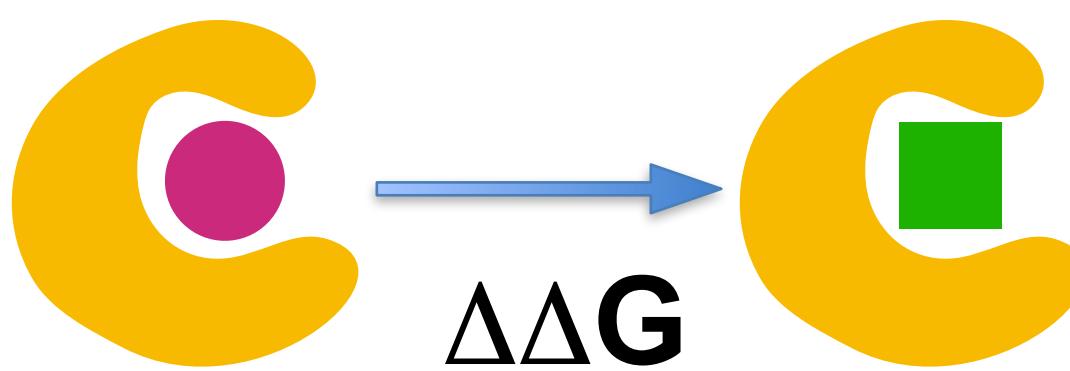


Good convergence and comparable results to standard RBFE were obtained on a TYK2 test system

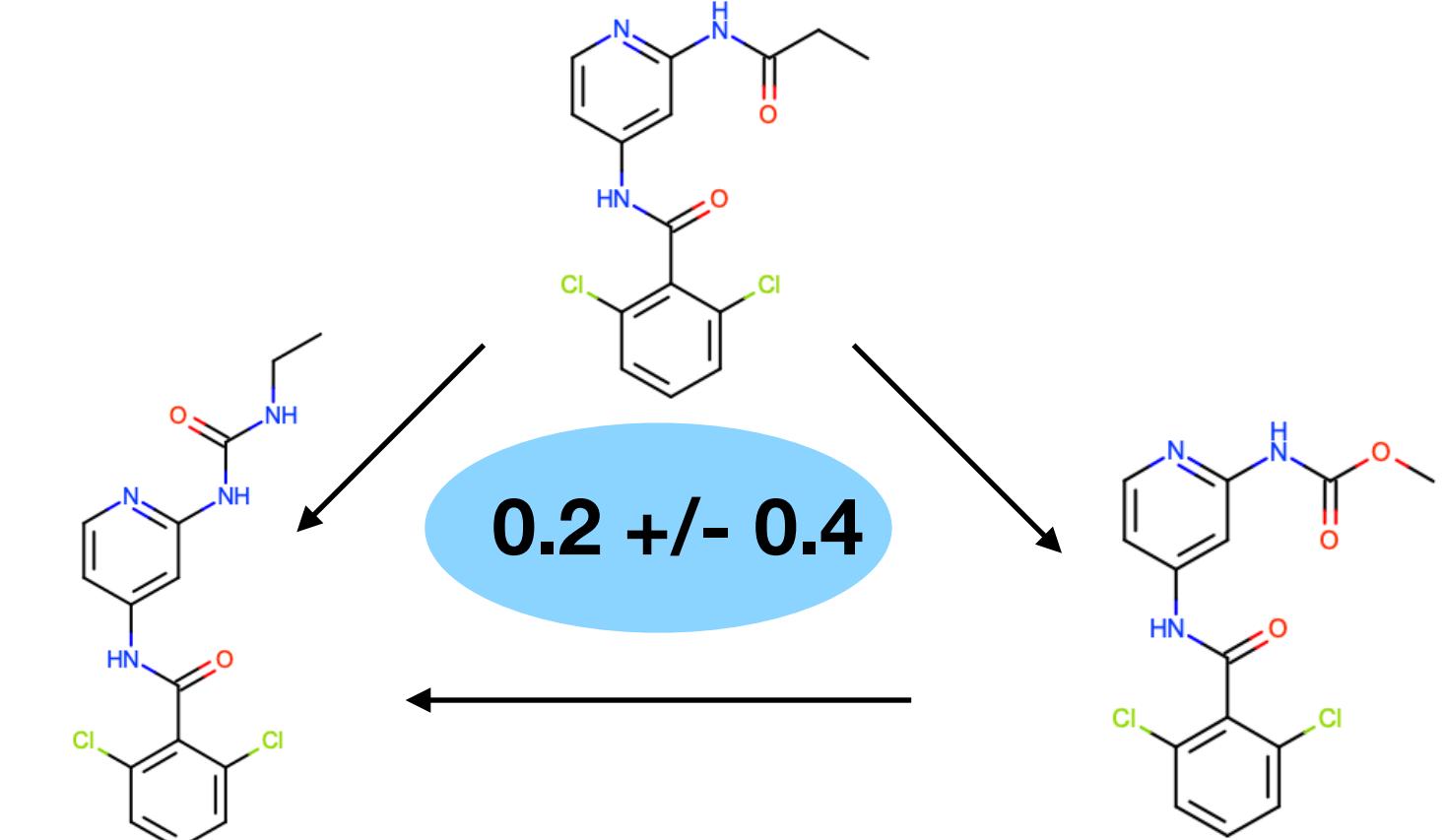


We use enhanced sampling to improve the efficiency of the approach

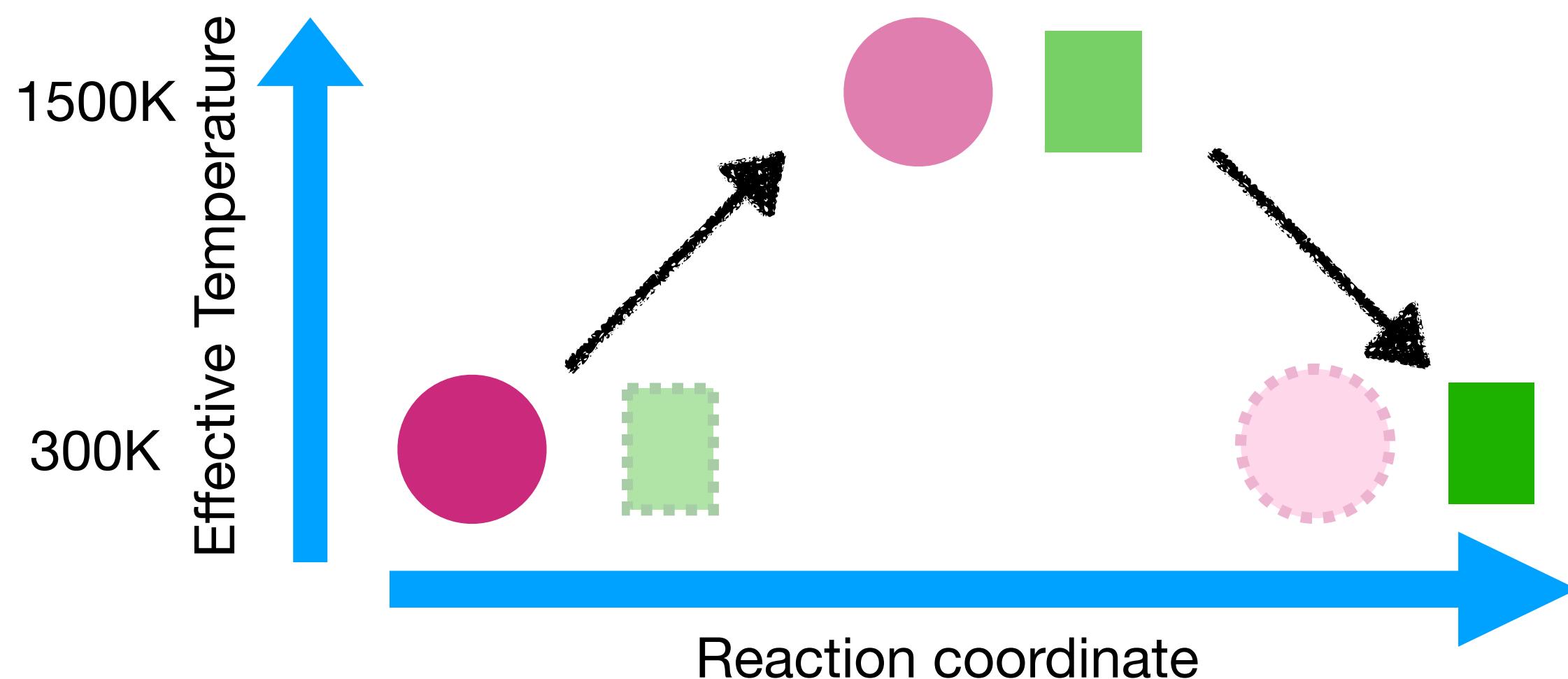
Outlook



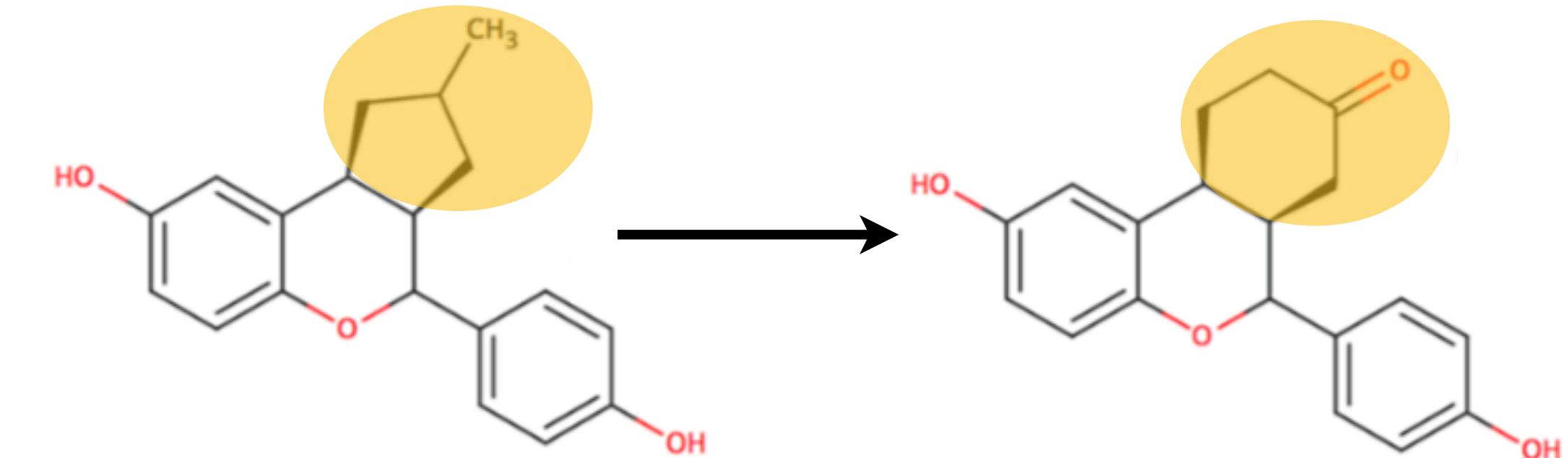
We implemented an alternative approach to RBFE, Separated Topologies, which allows the comparison of structurally diverse ligands



Good convergence and comparable results to standard RBFE were obtained on a TYK2 test system

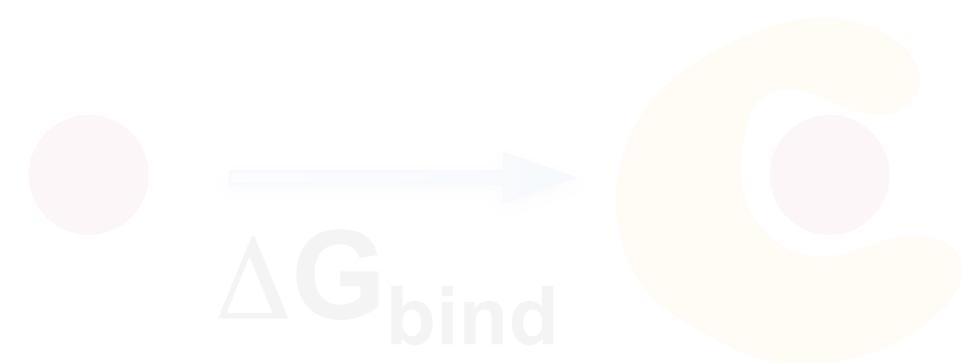


We use enhanced sampling to improve the efficiency of the approach



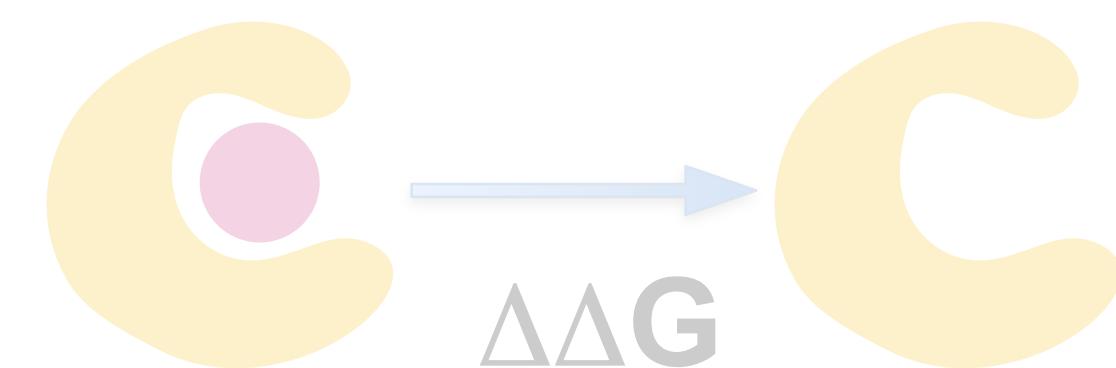
We will further test the approach on systems with structurally diverse ligands

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



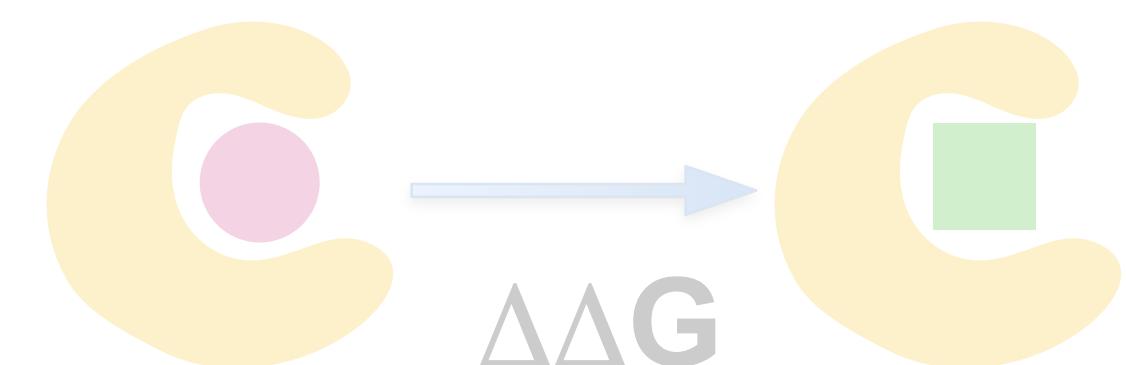
Enhanced sampling methods for binding free energies

Outline



Absolute binding free energy calculations

Challenges encountered applying equilibrium and non-equilibrium approaches



Separated Topologies

Overcoming limitations and broadening the scope of standard binding free energy methods



Enhanced sampling methods for binding free energies



Sam Gill



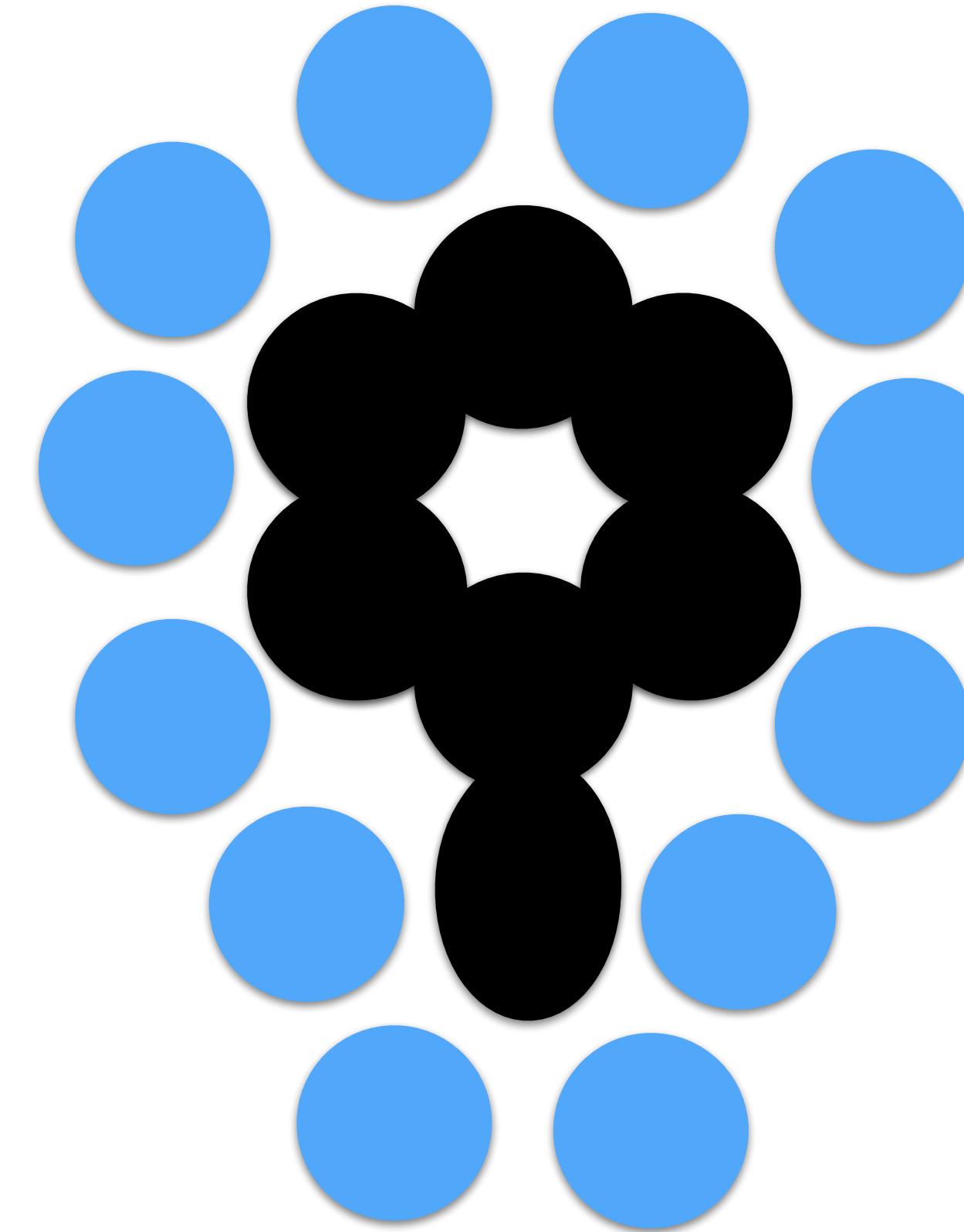
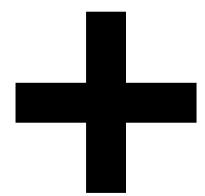
BLUES

Binding modes of
Ligands
Using
Enhanced
Sampling

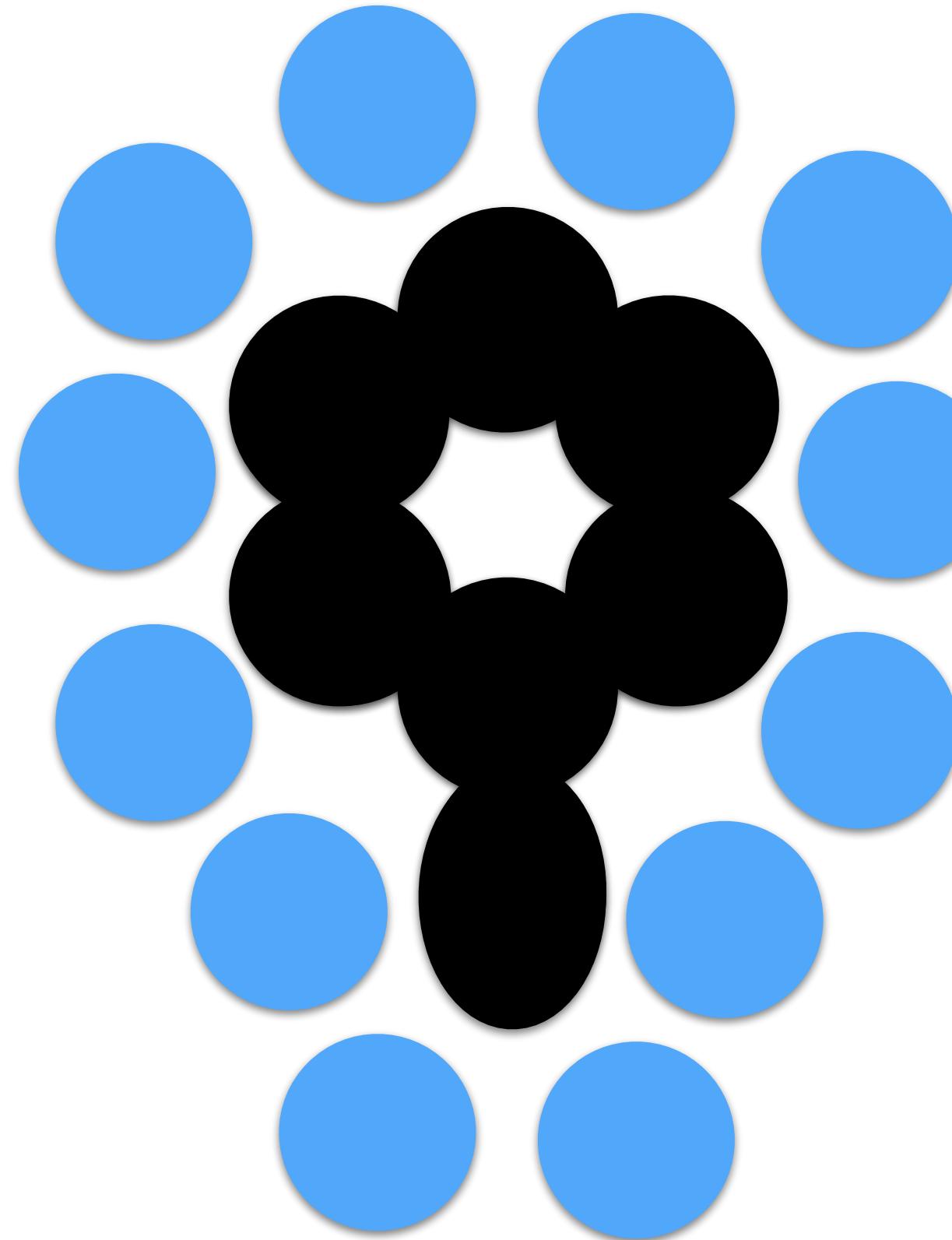
BLUES is a hybrid approach that combines:
Nonequilibrium Candidate Monte Carlo moves with MD



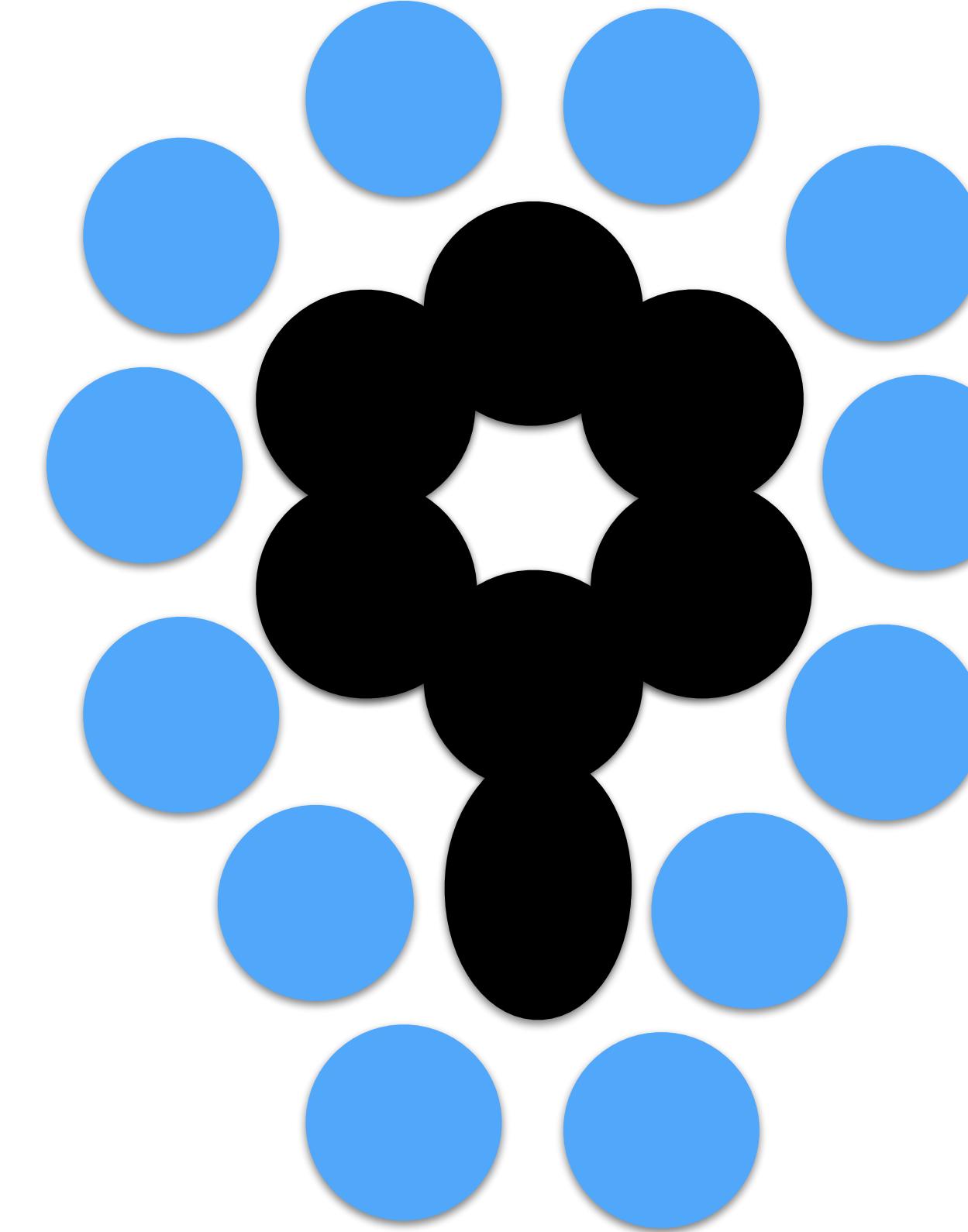
BLUES



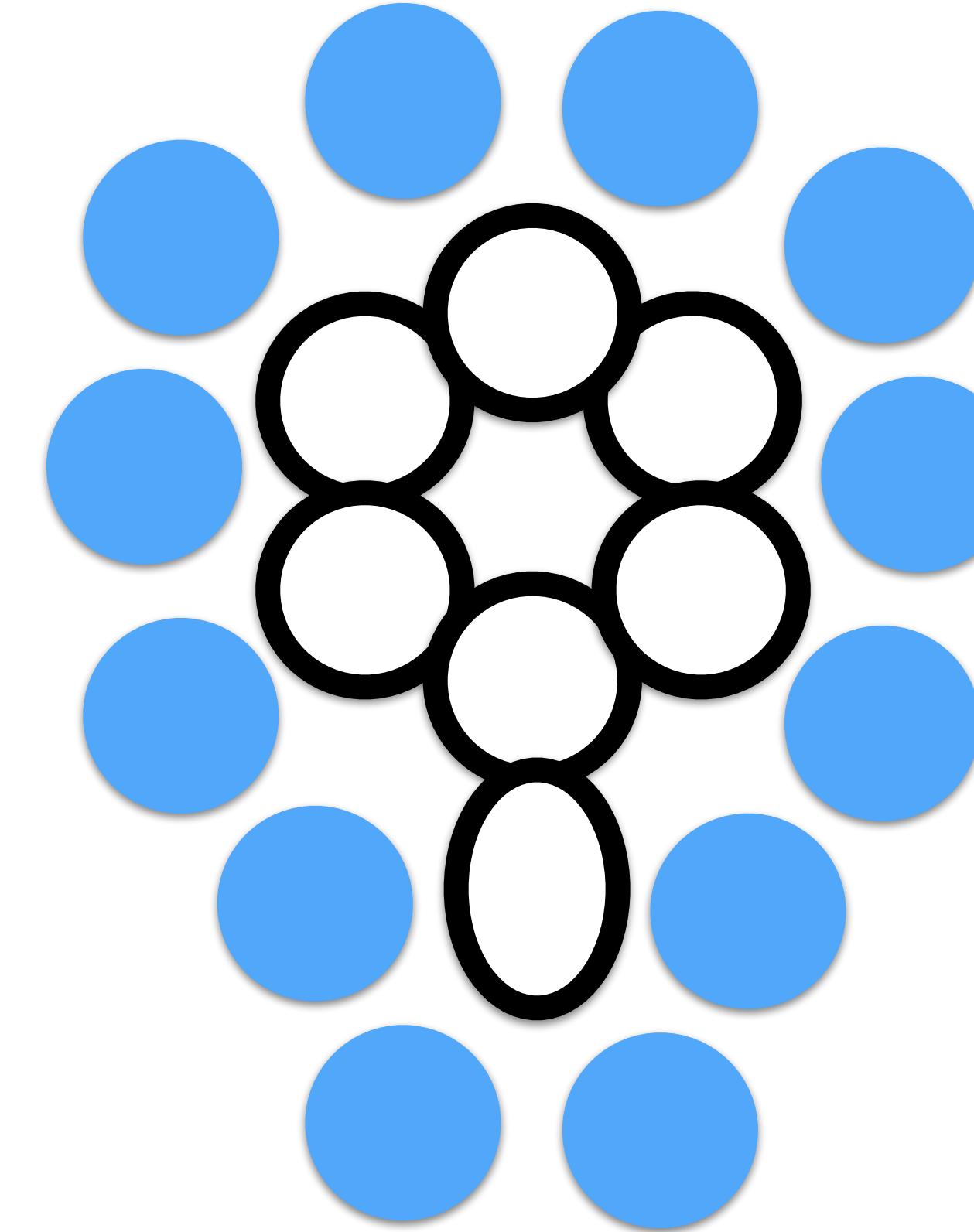
BLUES is a hybrid approach that combines:
Nonequilibrium Candidate Monte Carlo moves with MD



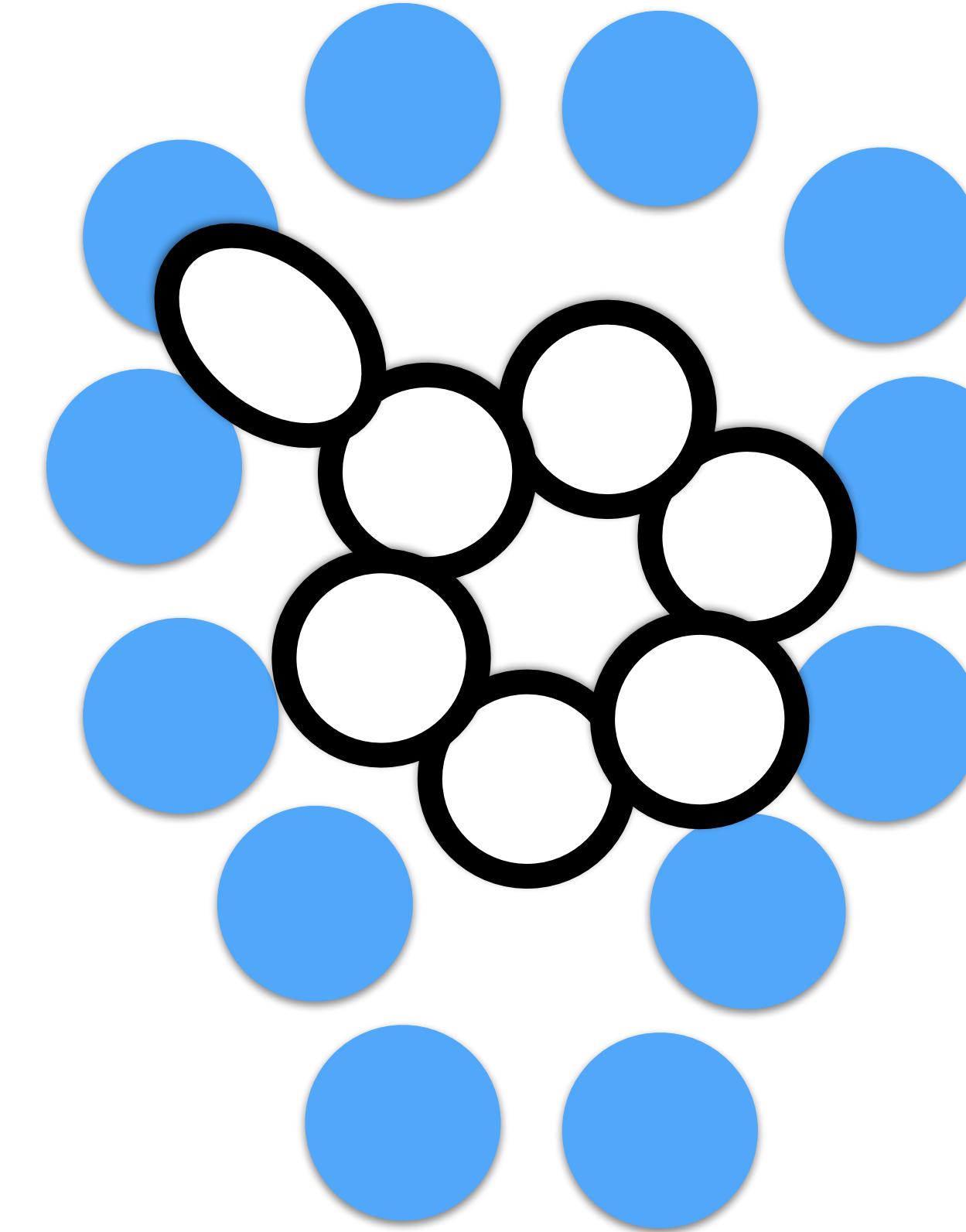
Nonequilibrium Candidate Monte Carlo (NCMC)
allows us to propose a random rotational move to the ligand



Nonequilibrium Candidate Monte Carlo (NCMC)
allows us to propose a random rotational move to the ligand



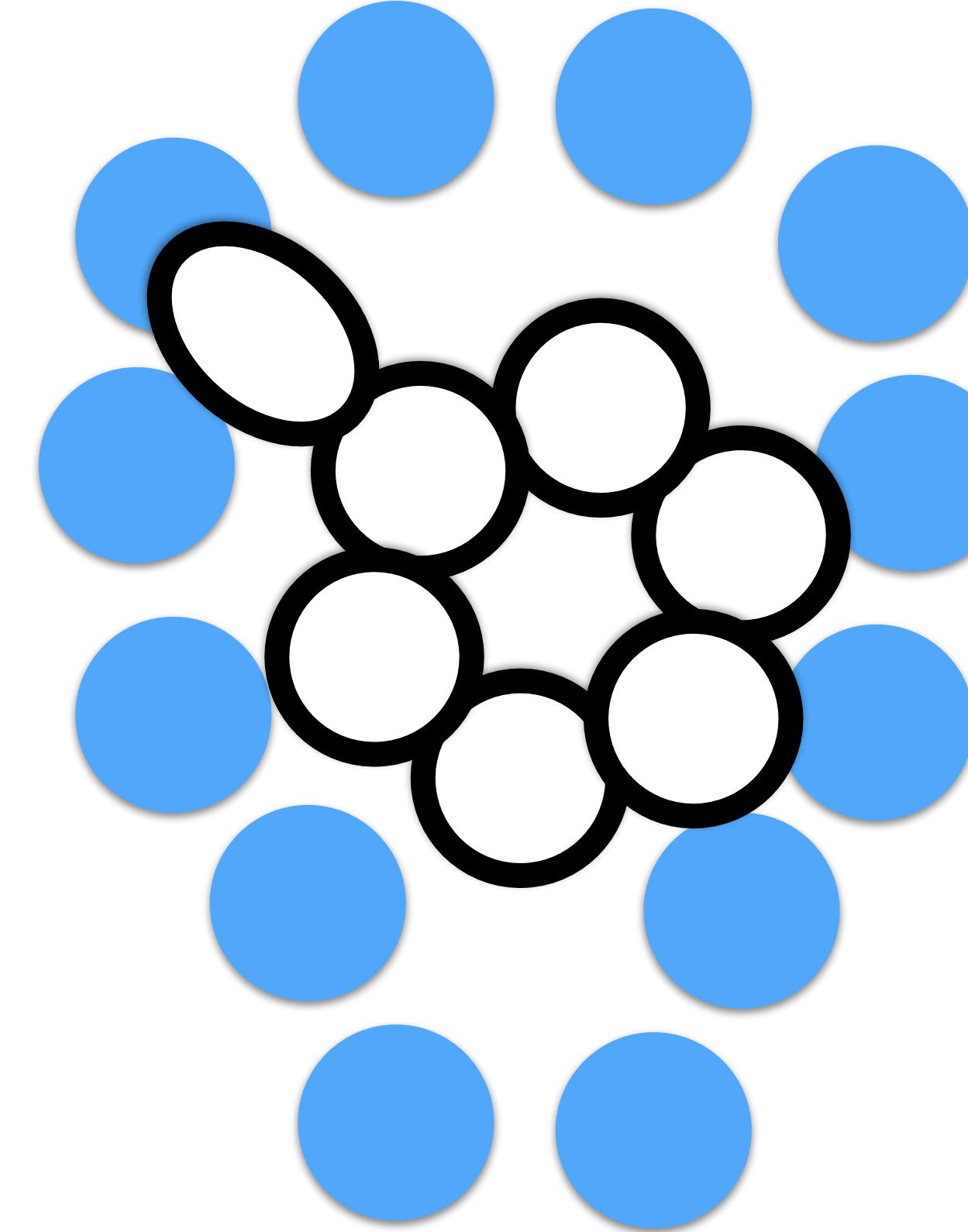
Nonequilibrium Candidate Monte Carlo (NCMC)
allows us to propose a random rotational move to the ligand



Nonequilibrium Candidate Monte Carlo (NCMC)
allows us to propose a *random rotational* move to the ligand



- + Random rotation may identify
new possible binding modes

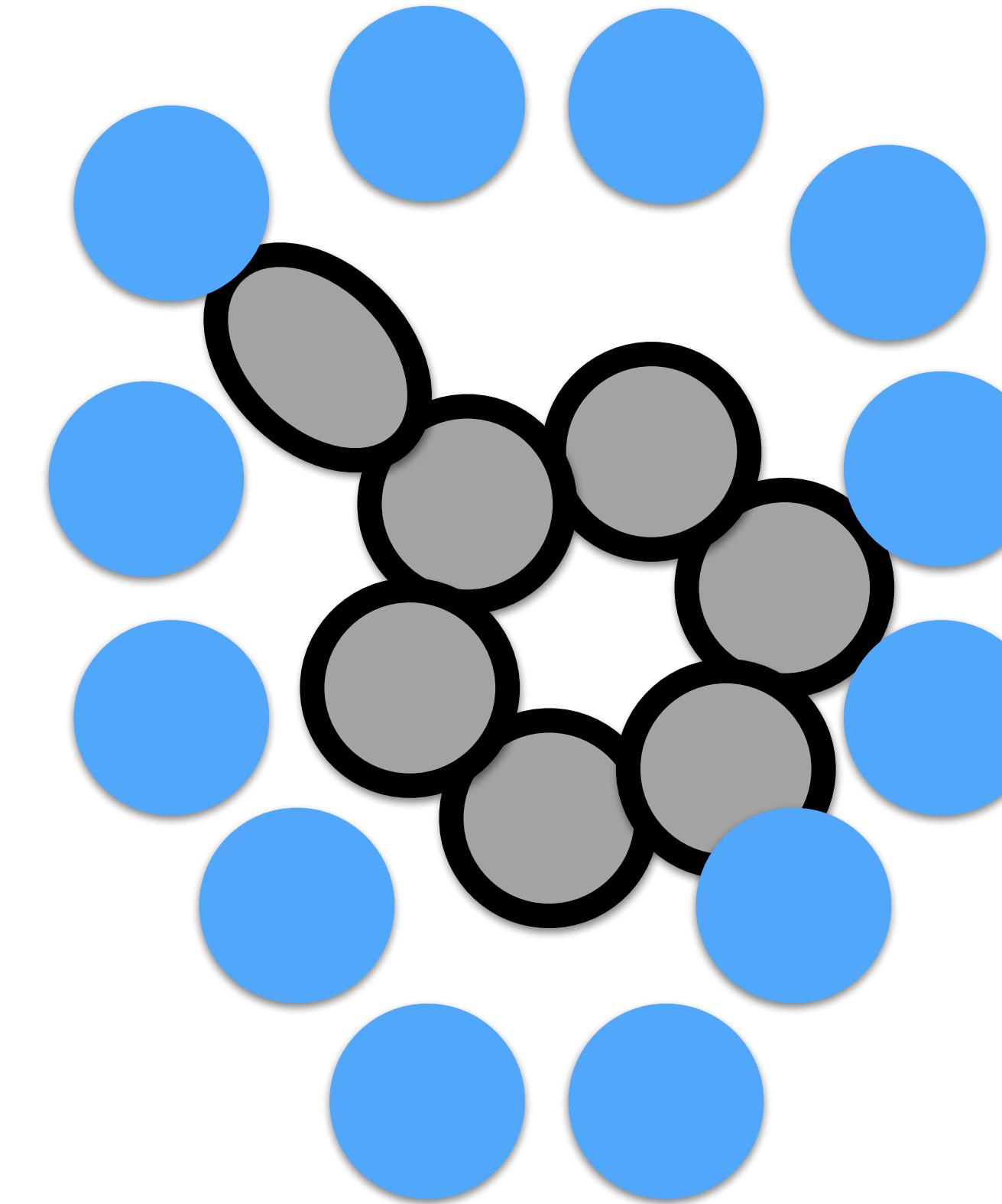


NCMC move is divided into smaller steps to help resolve potential clashes



BLUES

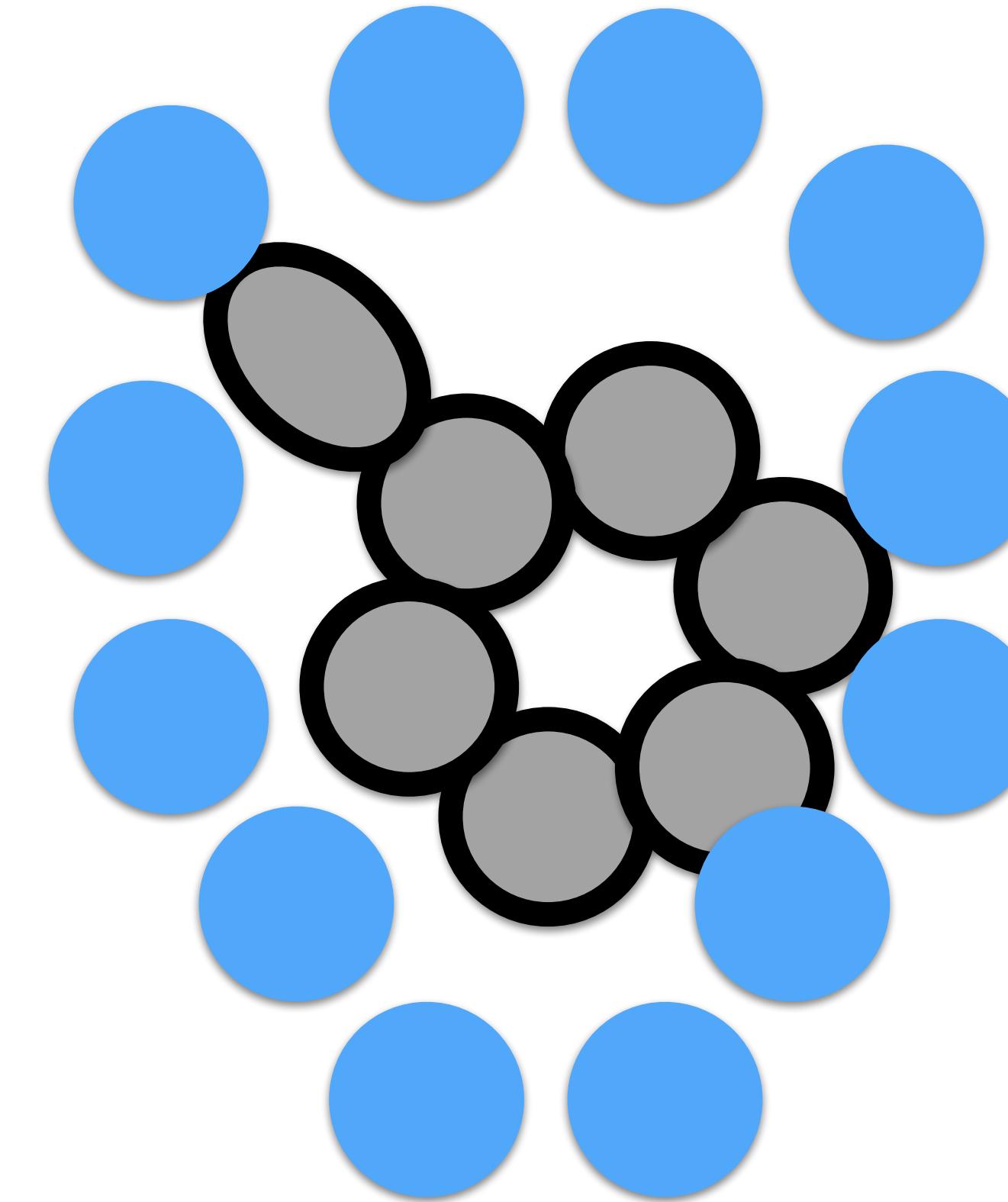
- + *Random rotation may identify new possible binding modes*



NCMC move is divided into smaller steps to help resolve potential clashes



- + *Random rotation may identify new possible binding modes*

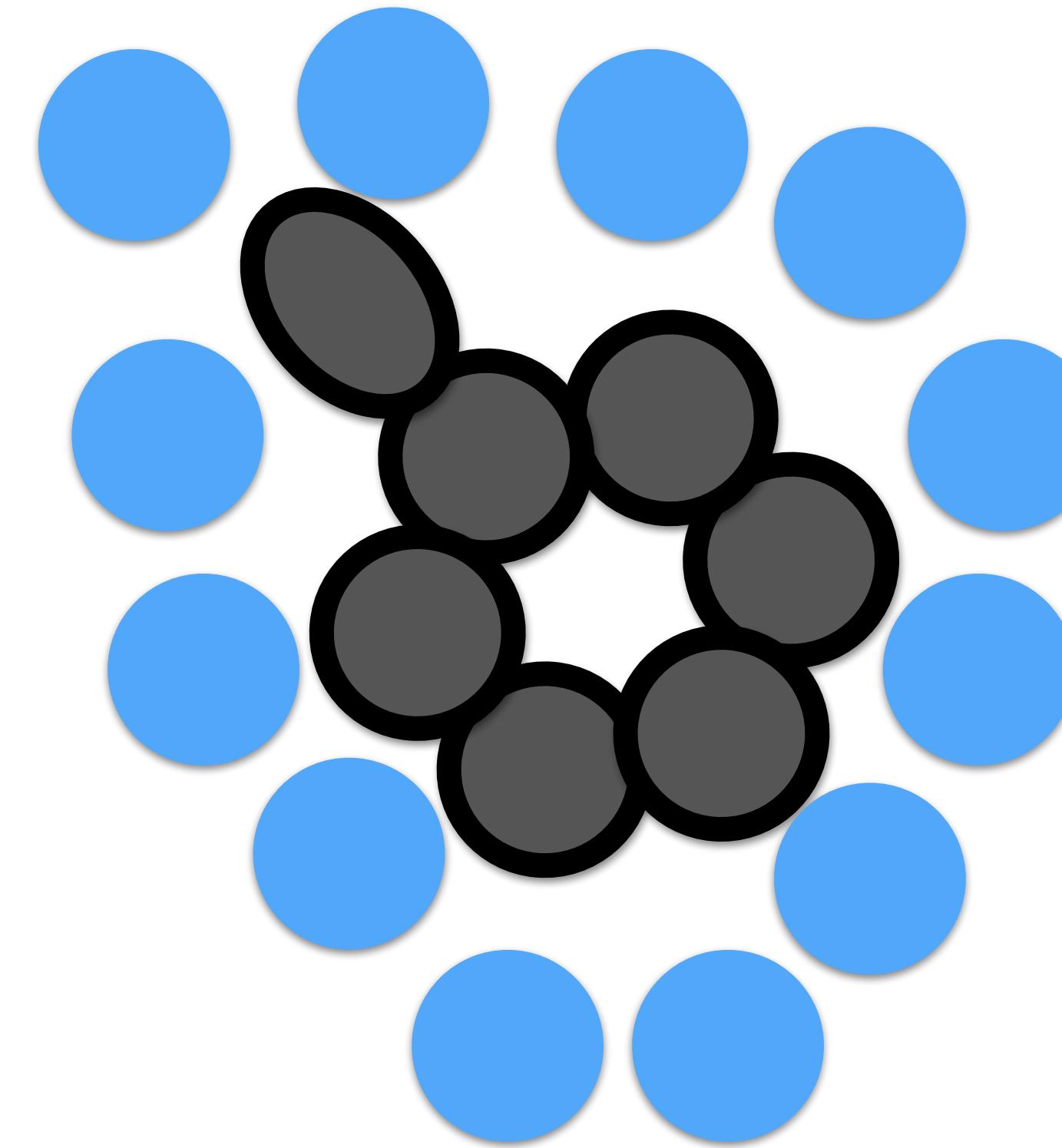


- Change ligand interactions
- Relaxation

NCMC move is divided into smaller steps to help resolve potential clashes



- + *Random rotation may identify new possible binding modes*

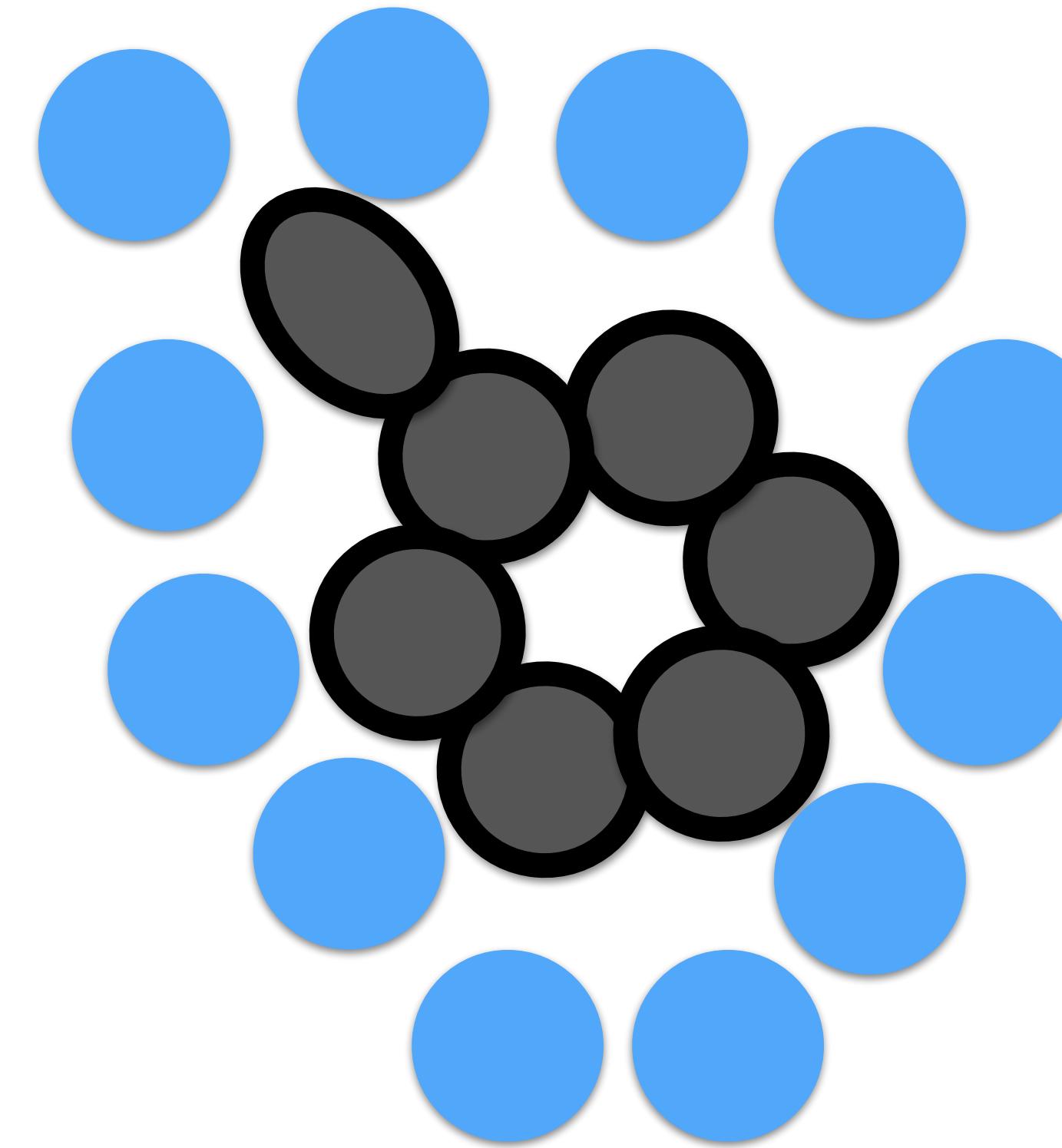


- Change ligand interactions
- Relaxation

NCMC move is divided into smaller steps to help resolve potential clashes



- + *Random rotation may identify new possible binding modes*
- + **Relaxation gives higher move acceptance than traditional Monte Carlo**



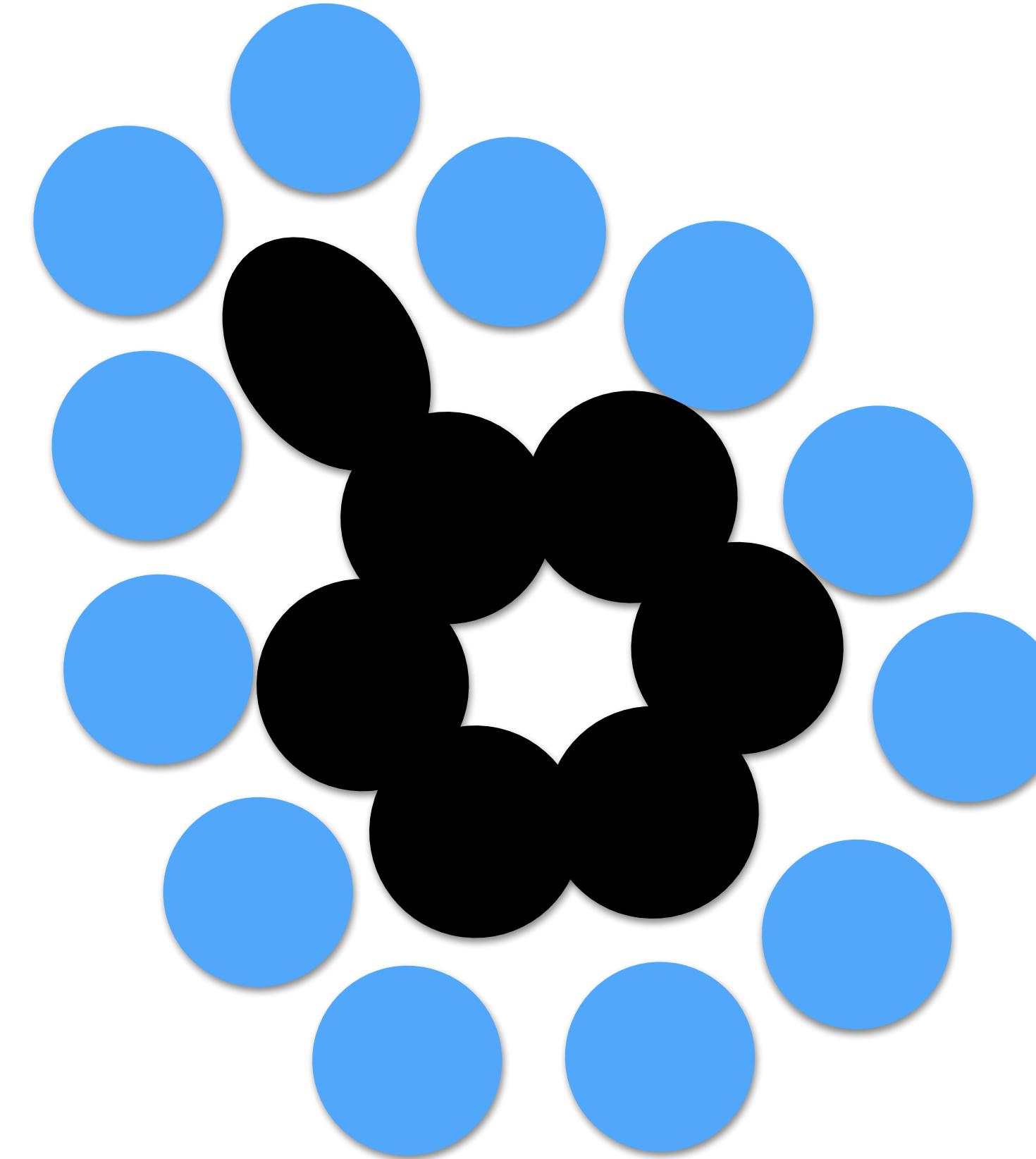
- Change ligand interactions
- Relaxation

NCMC move is accepted/rejected based upon the total work done



BLUES

- + *Random rotation may identify new possible binding modes*
- + *Relaxation gives higher move acceptance than traditional Monte Carlo*

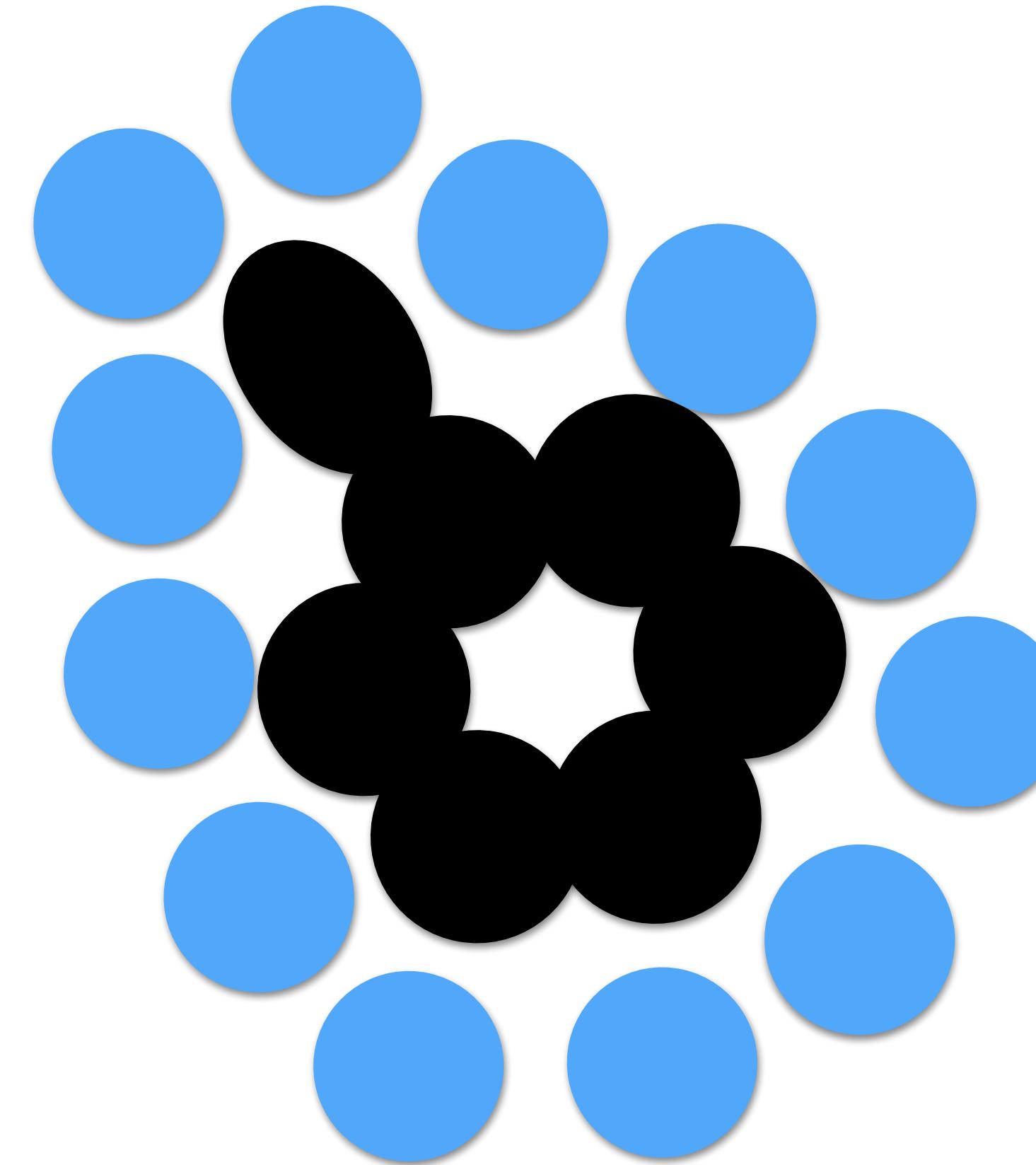


NCMC move is accepted/rejected based upon the total work done

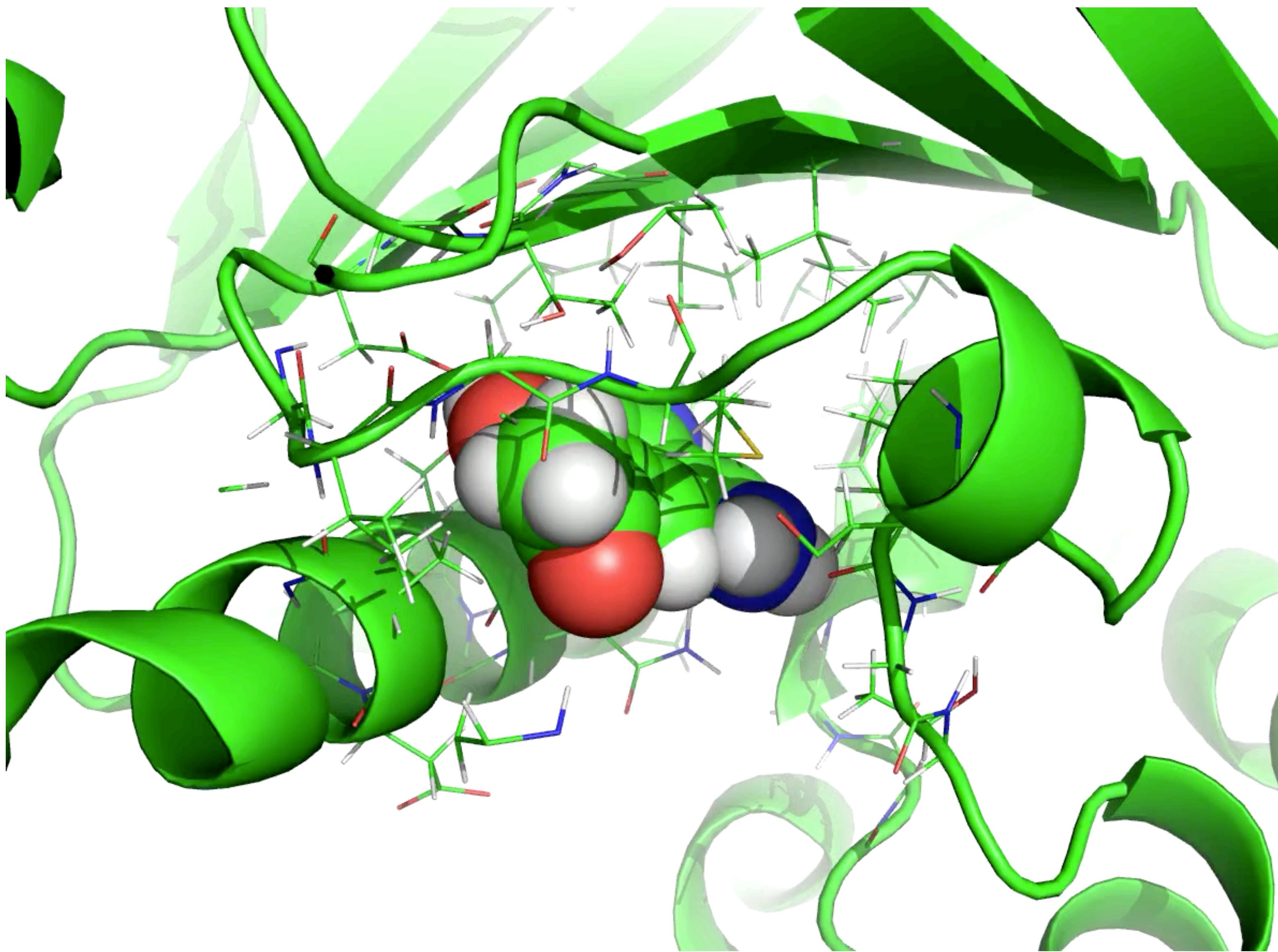


BLUES

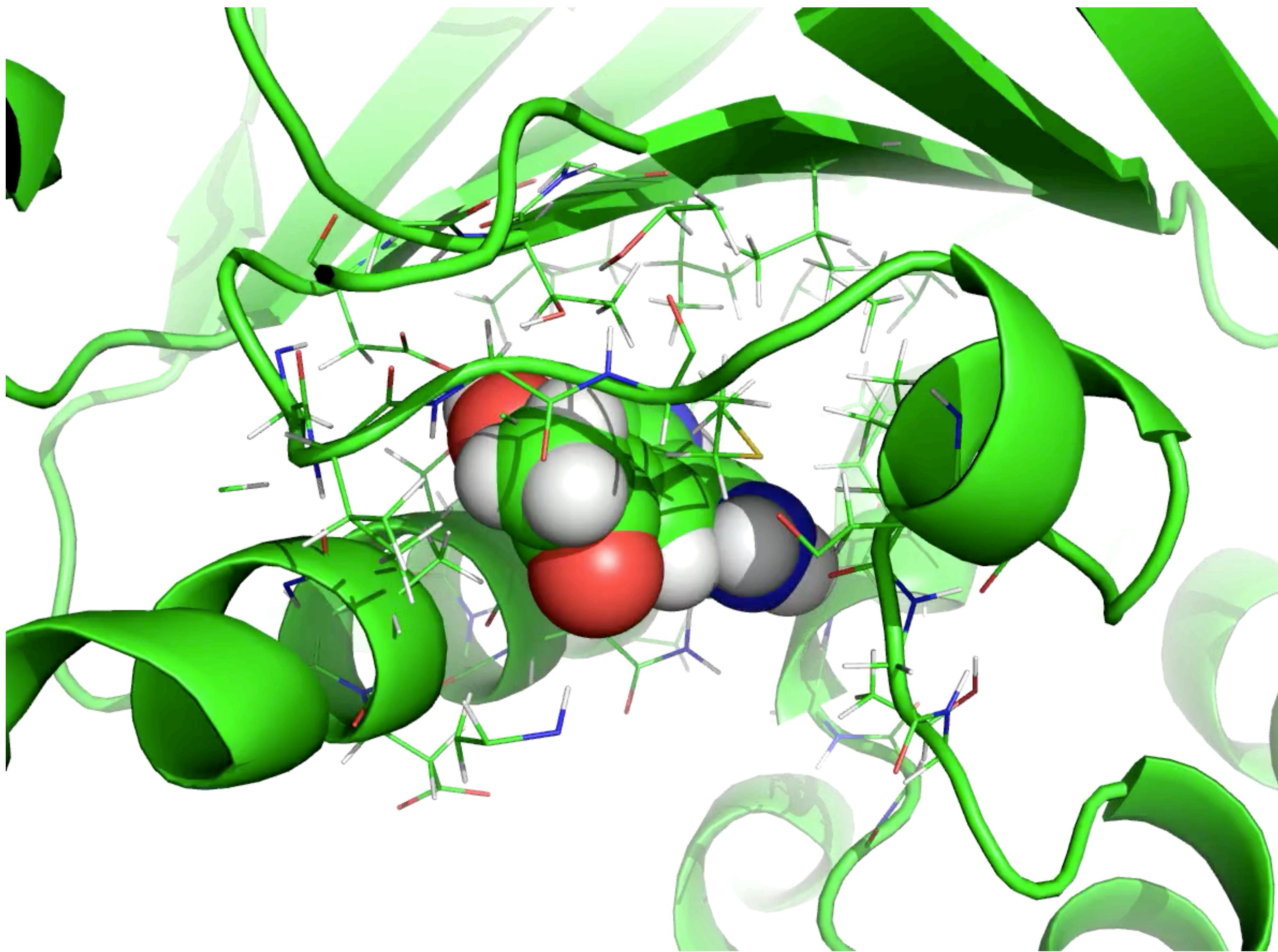
- + *Random rotation may identify new possible binding modes*
- + *Relaxation gives higher move acceptance than traditional Monte Carlo*
- + **Move acceptance/rejection is followed by conventional MD.**



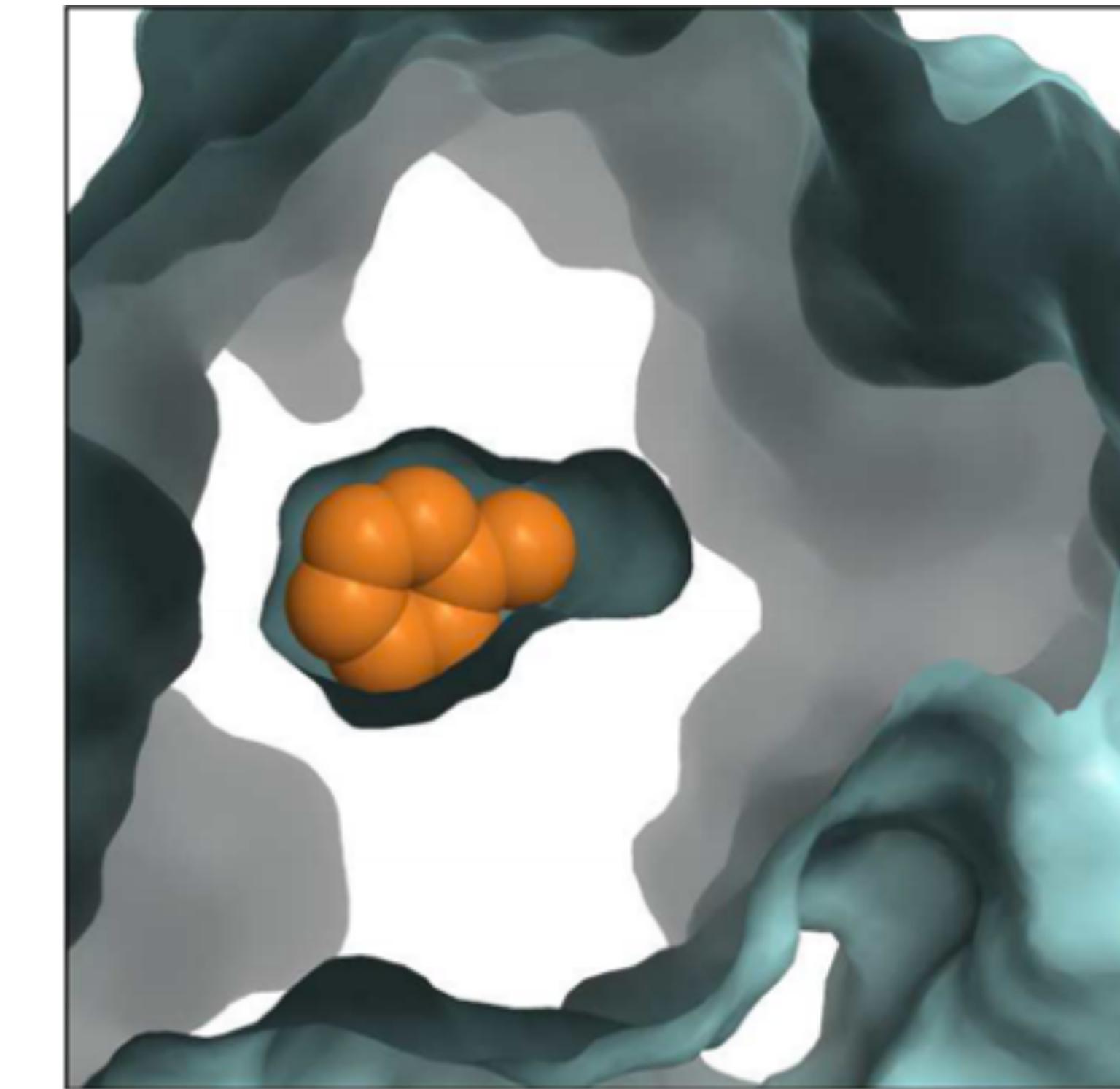
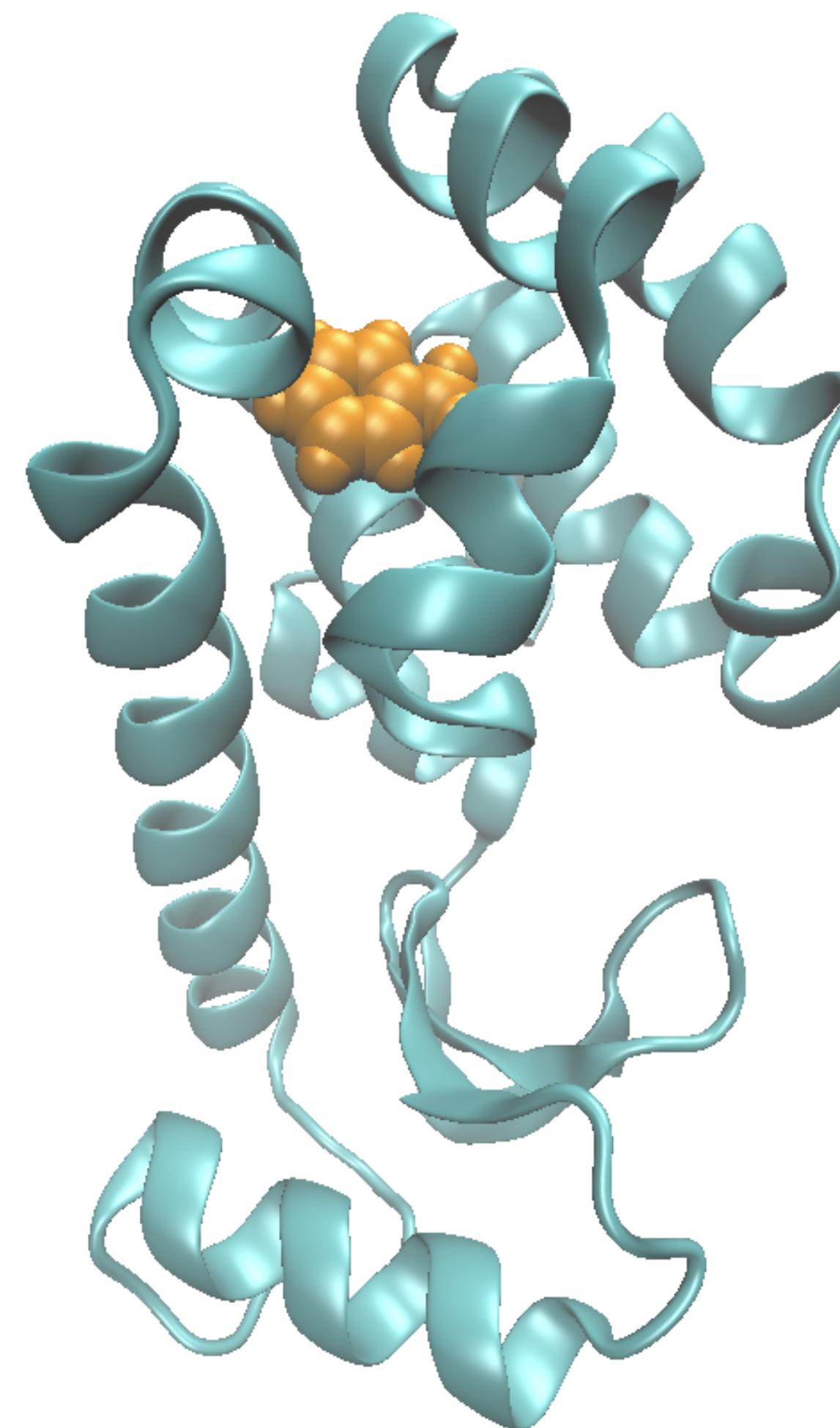
Here is what BLUES looks like:



Here is what BLUES looks like:

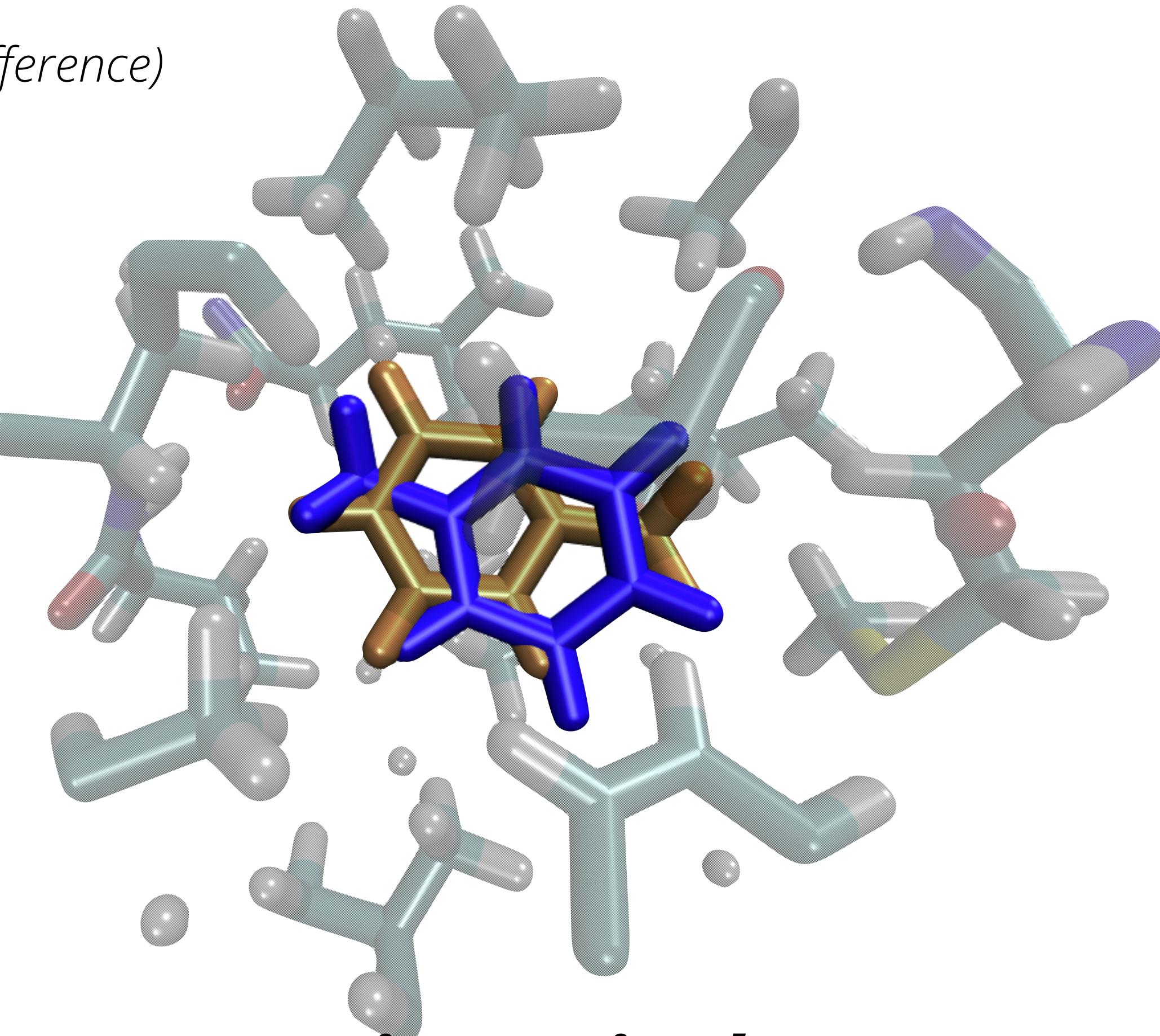
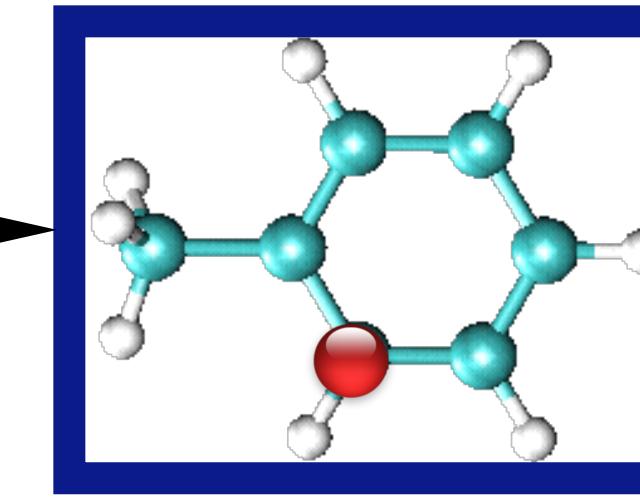
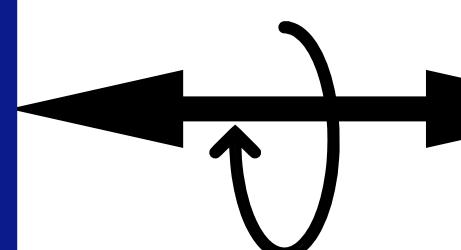
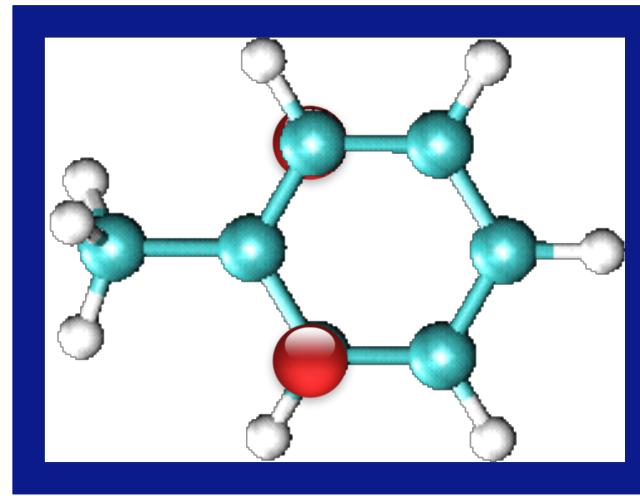
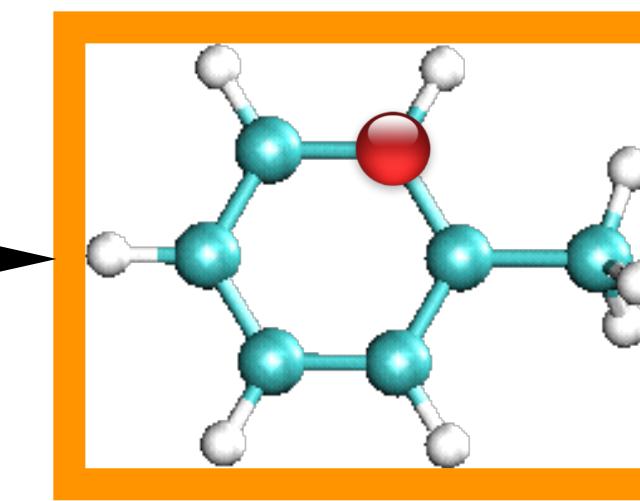
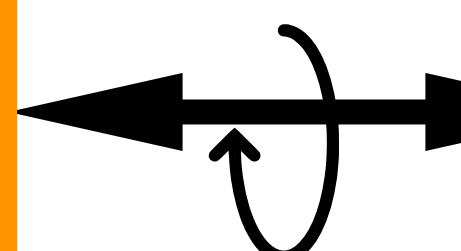
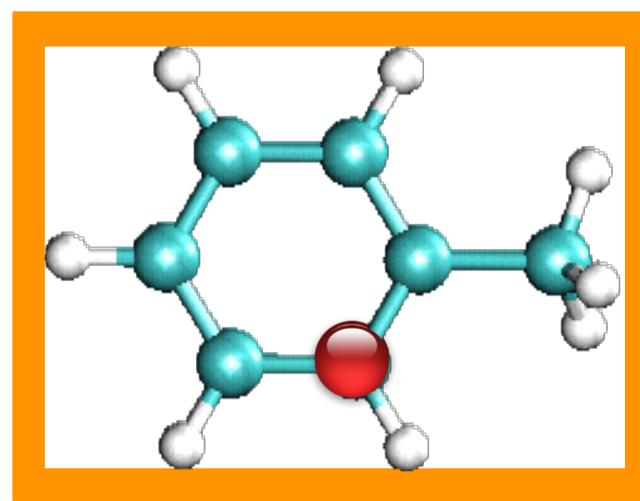


T4 lysozyme with toluene serves as a model binding system and our initial test case for BLUES



From $1.3\mu\text{s}$ of simulation data we found: 4 ligand binding modes *and their occupancies*

Toluene binding modes: (red dot to highlight difference)

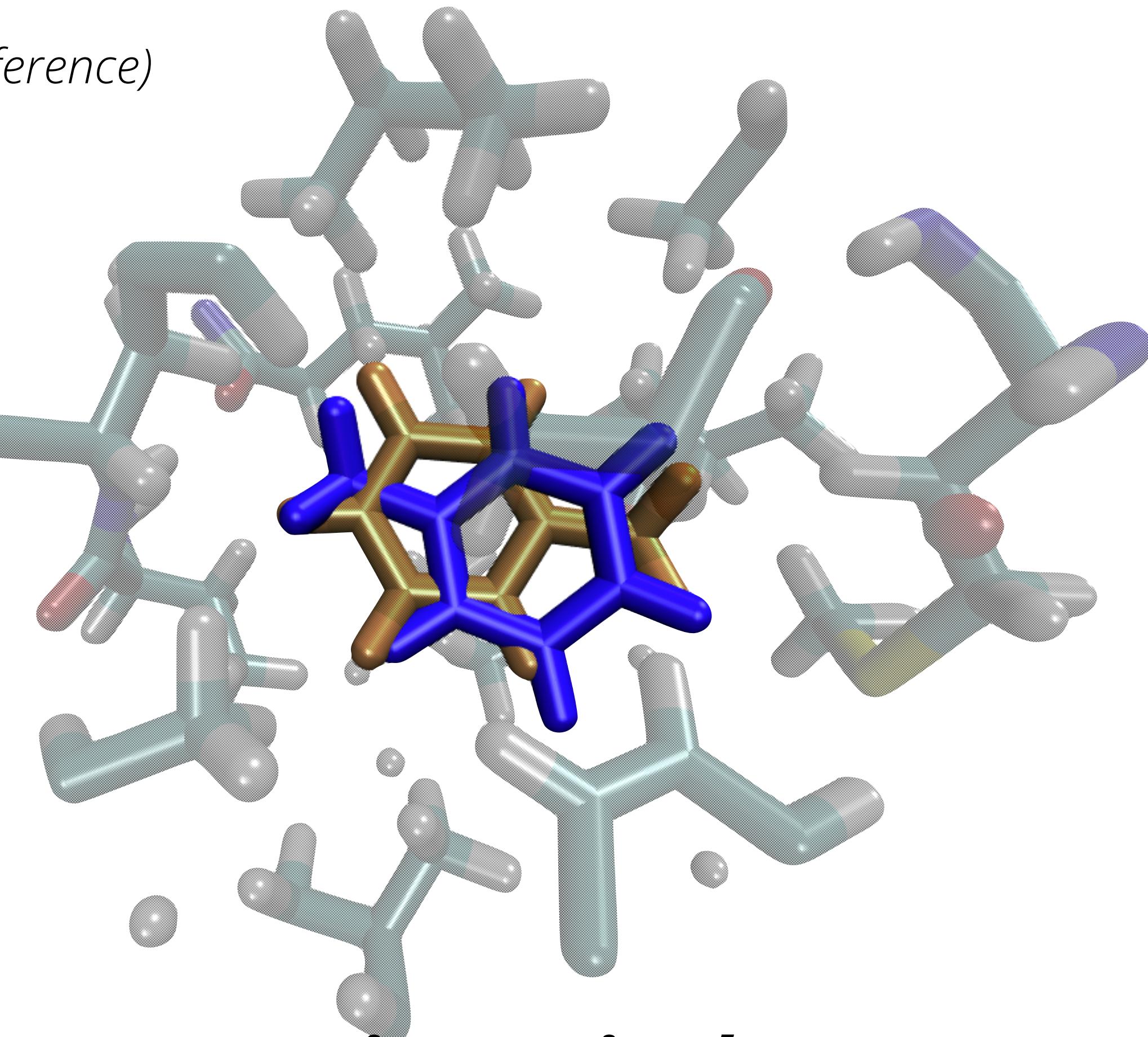
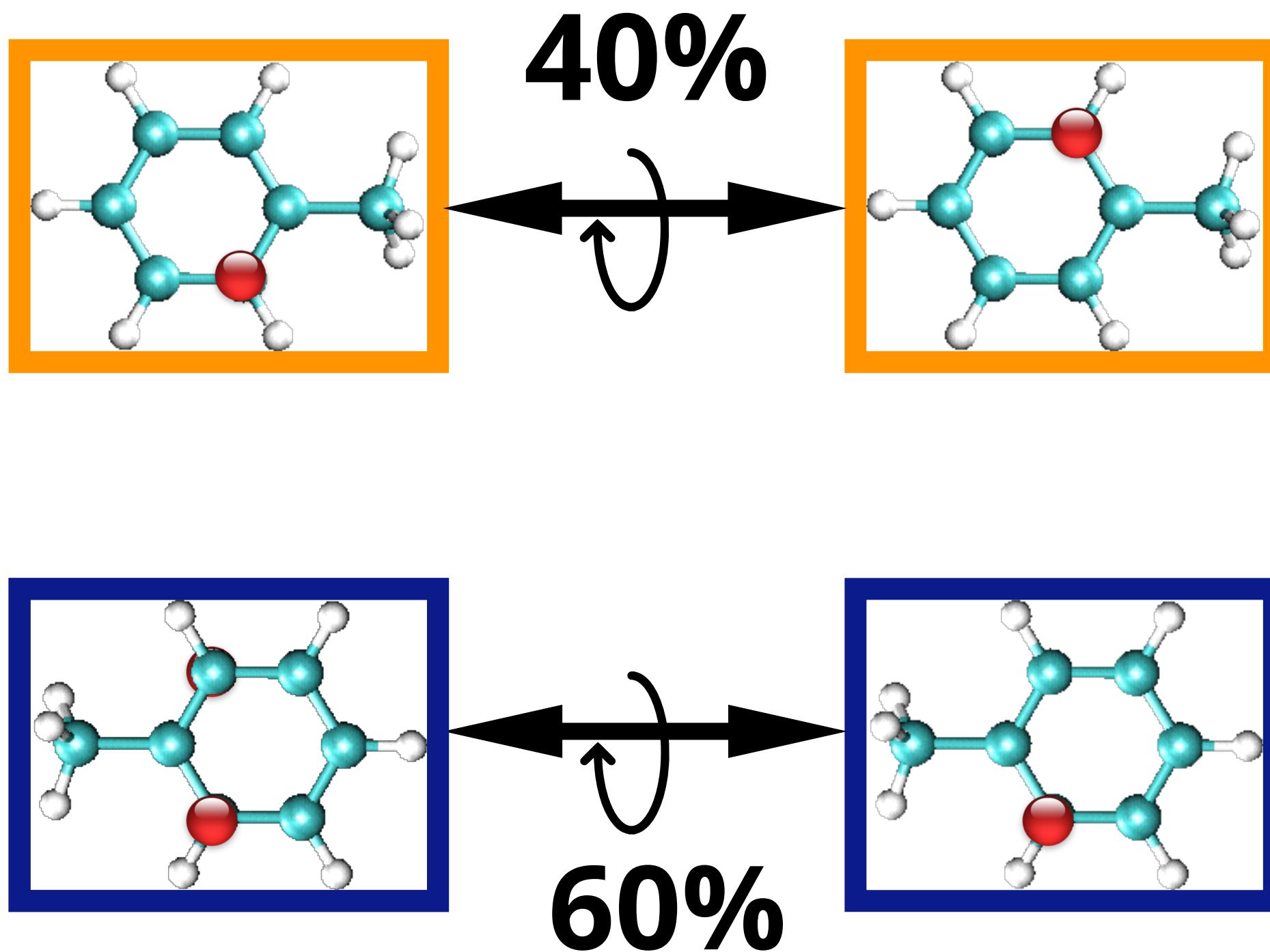


Each binding mode has 1 symmetric equivalent:

Out-of-plane flip

From $1.3\mu\text{s}$ of simulation data we found: 4 ligand binding modes *and their occupancies*

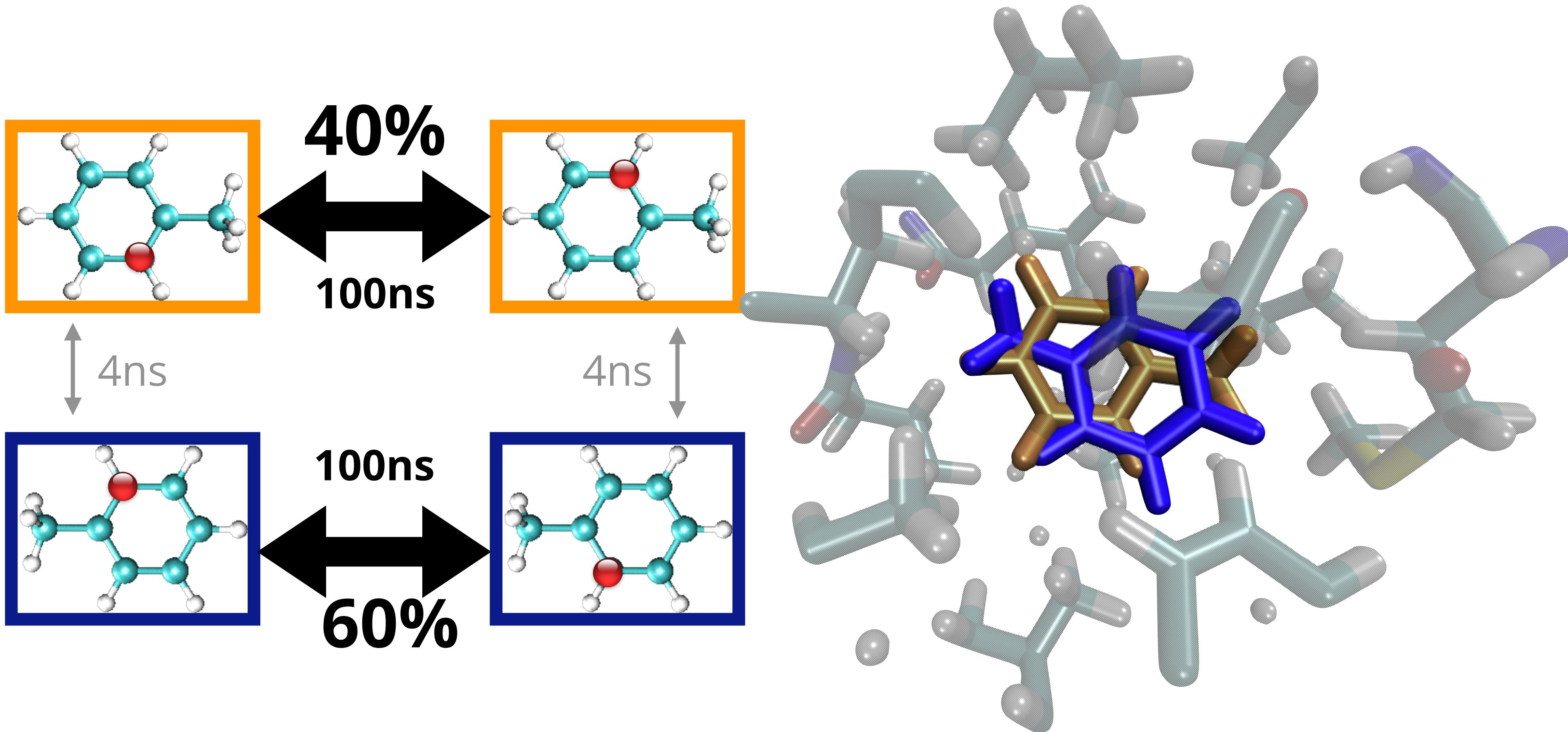
Toluene binding modes: (red dot to highlight difference)



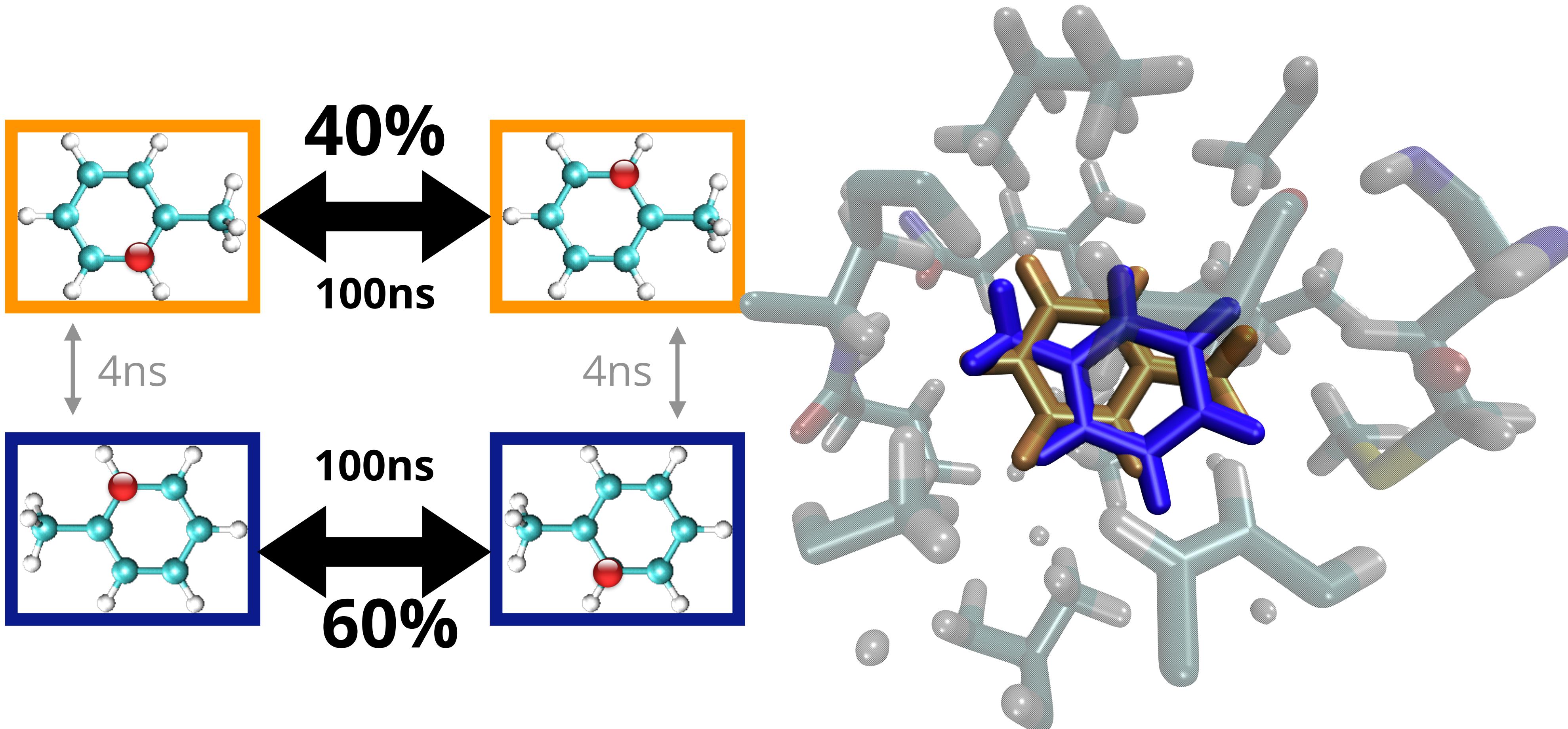
Each binding mode has 1 symmetric equivalent:

Out-of-plane flip

MSM suggested 100ns to transition between symmetry equivalent modes

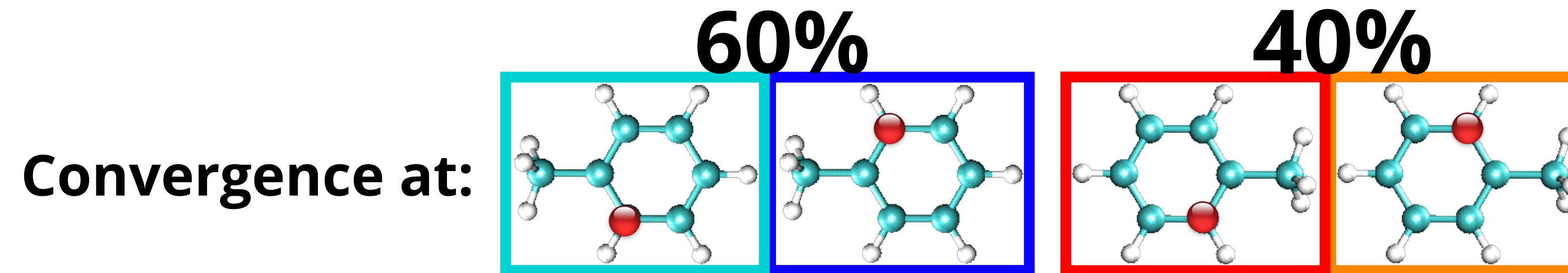


MSM suggested 100ns to transition between symmetry equivalent modes

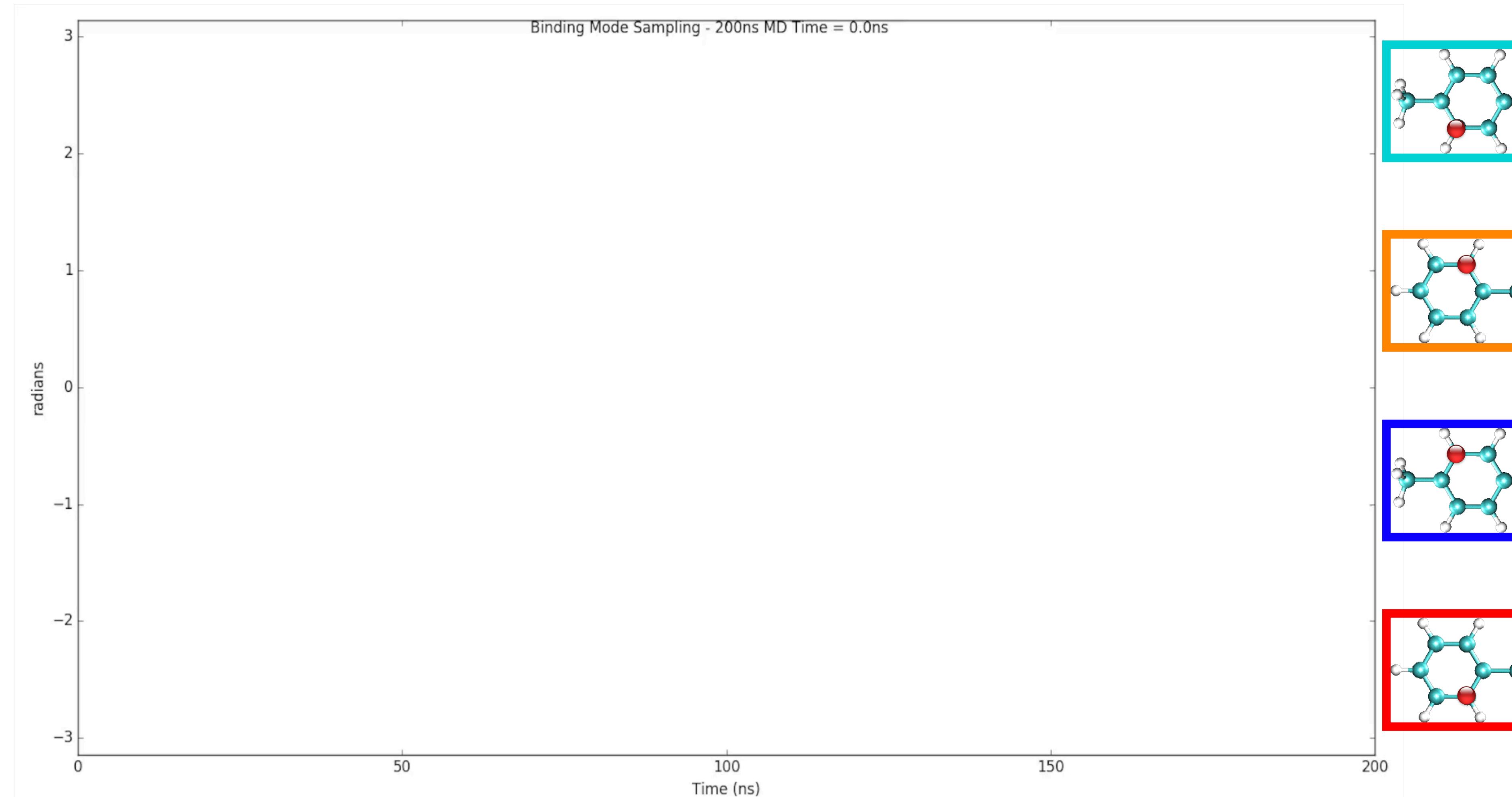


Even simple binding mode transitions can be slow!

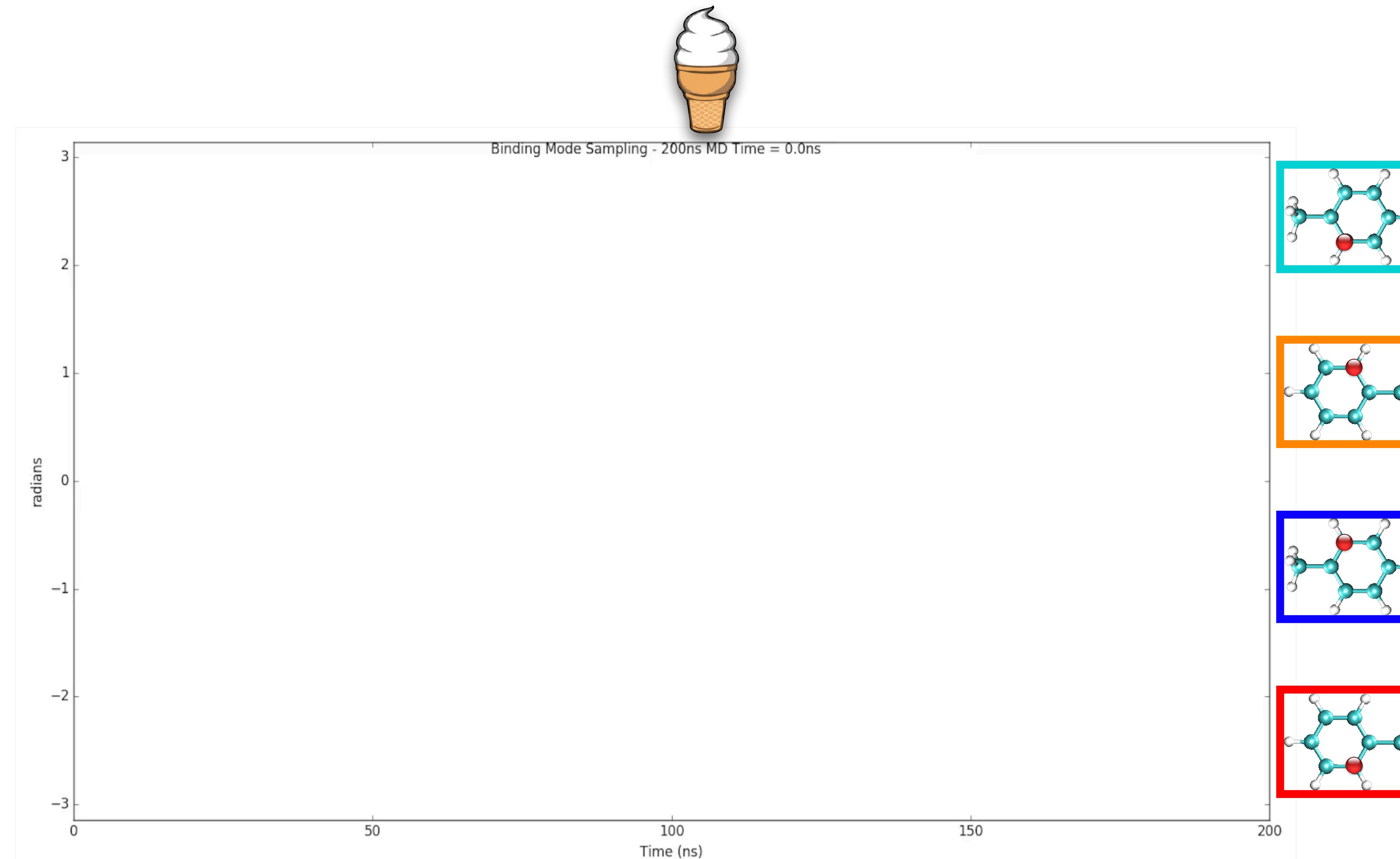
With enough simulation time, symmetric binding modes *should* have equivalent populations.



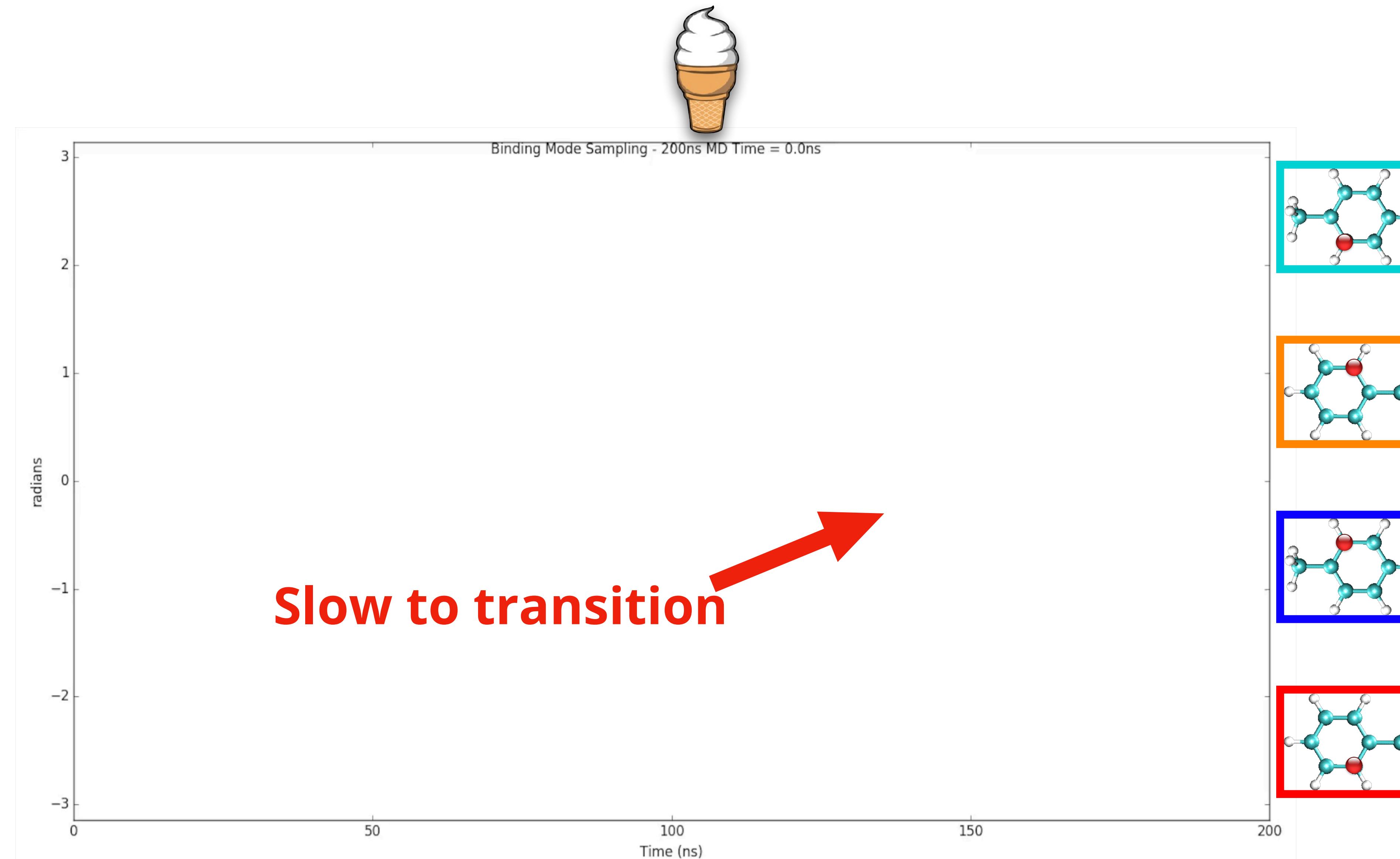
In 200ns of “vanilla” MD, symmetric binding modes do NOT have equivalent populations.



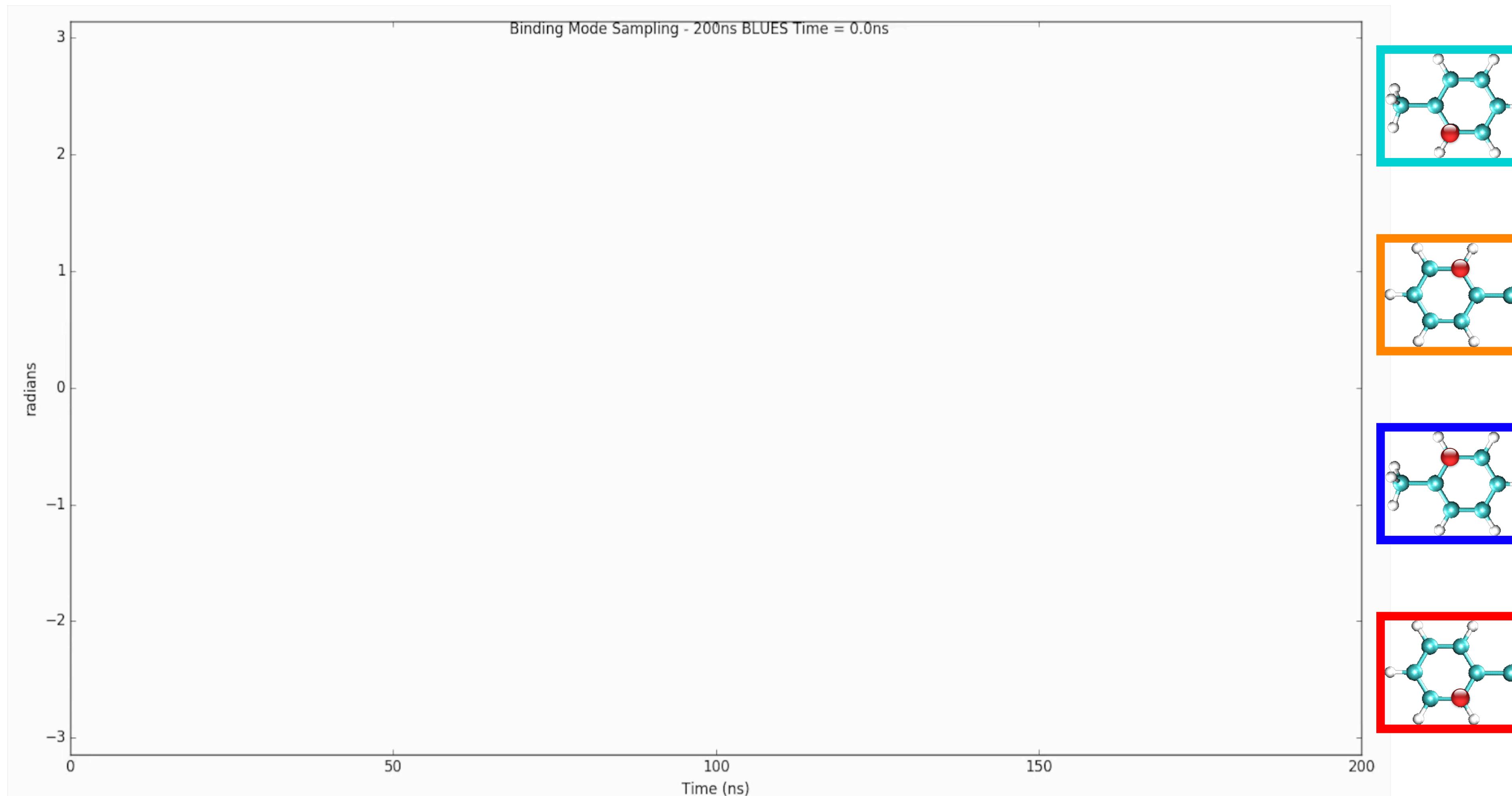
In 200ns of “vanilla” MD, symmetric binding modes do NOT have equivalent populations.



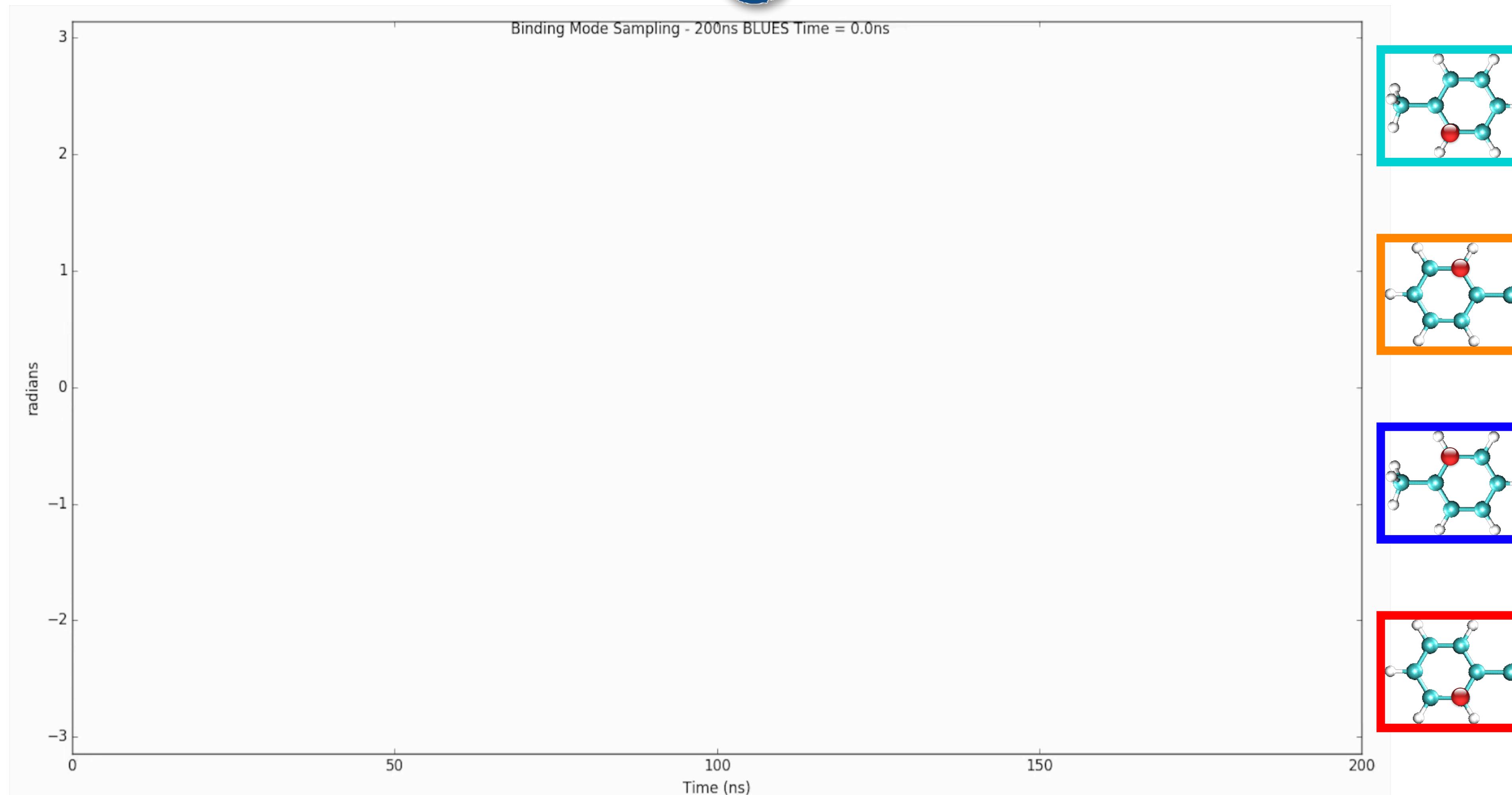
In 200ns of “vanilla” MD, symmetric binding modes do NOT have equivalent populations.



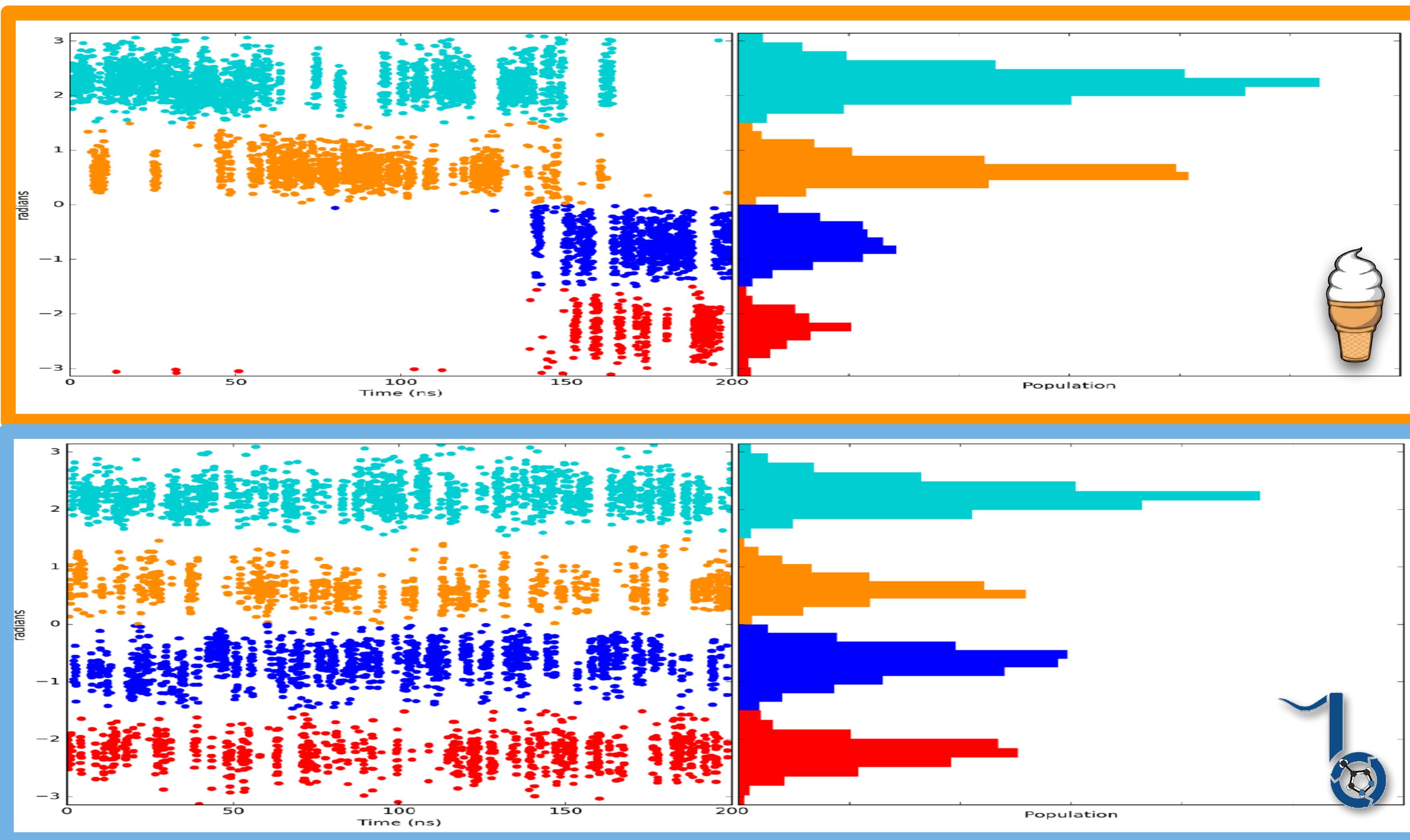
BLUES rapidly transitions between binding modes,
producing the correct populations



BLUES rapidly transitions between binding modes, producing the correct populations



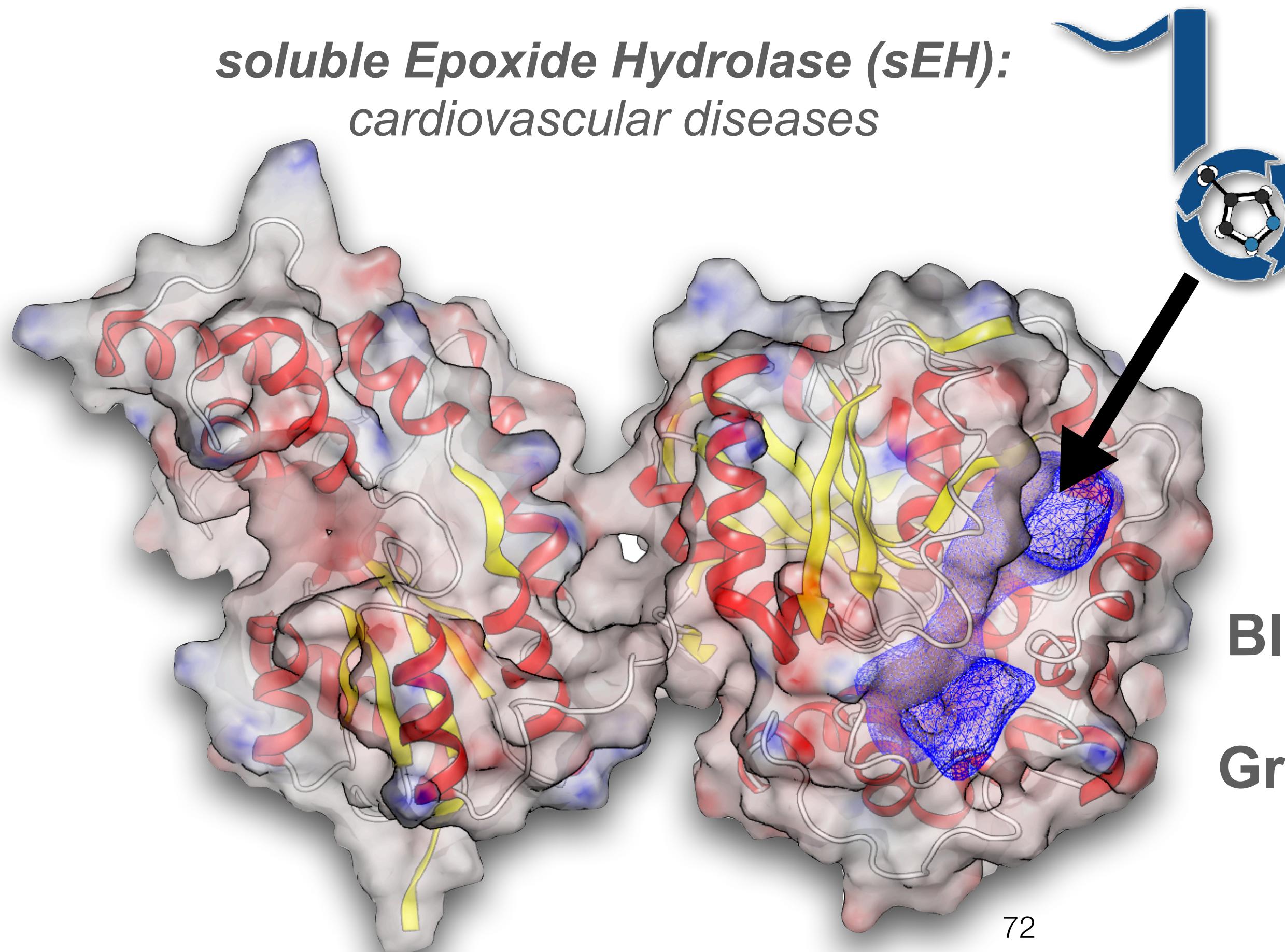
BLUES rapidly transitions between binding modes, producing the correct populations



We're beginning to test BLUES on a more pharmaceutically interesting target...

Given the binding site, can we predict:

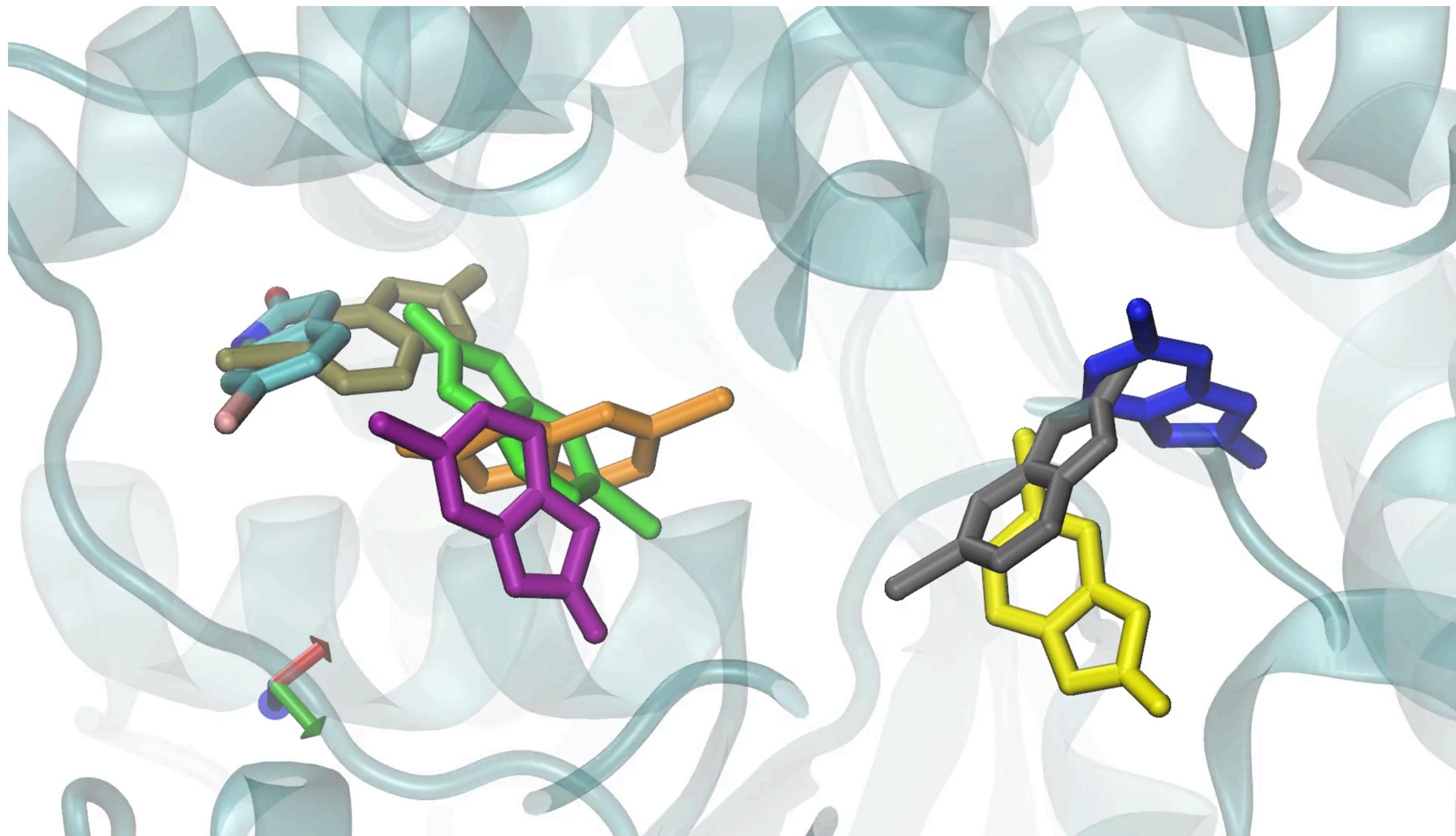
ligand binding modes for **59 fragments?**



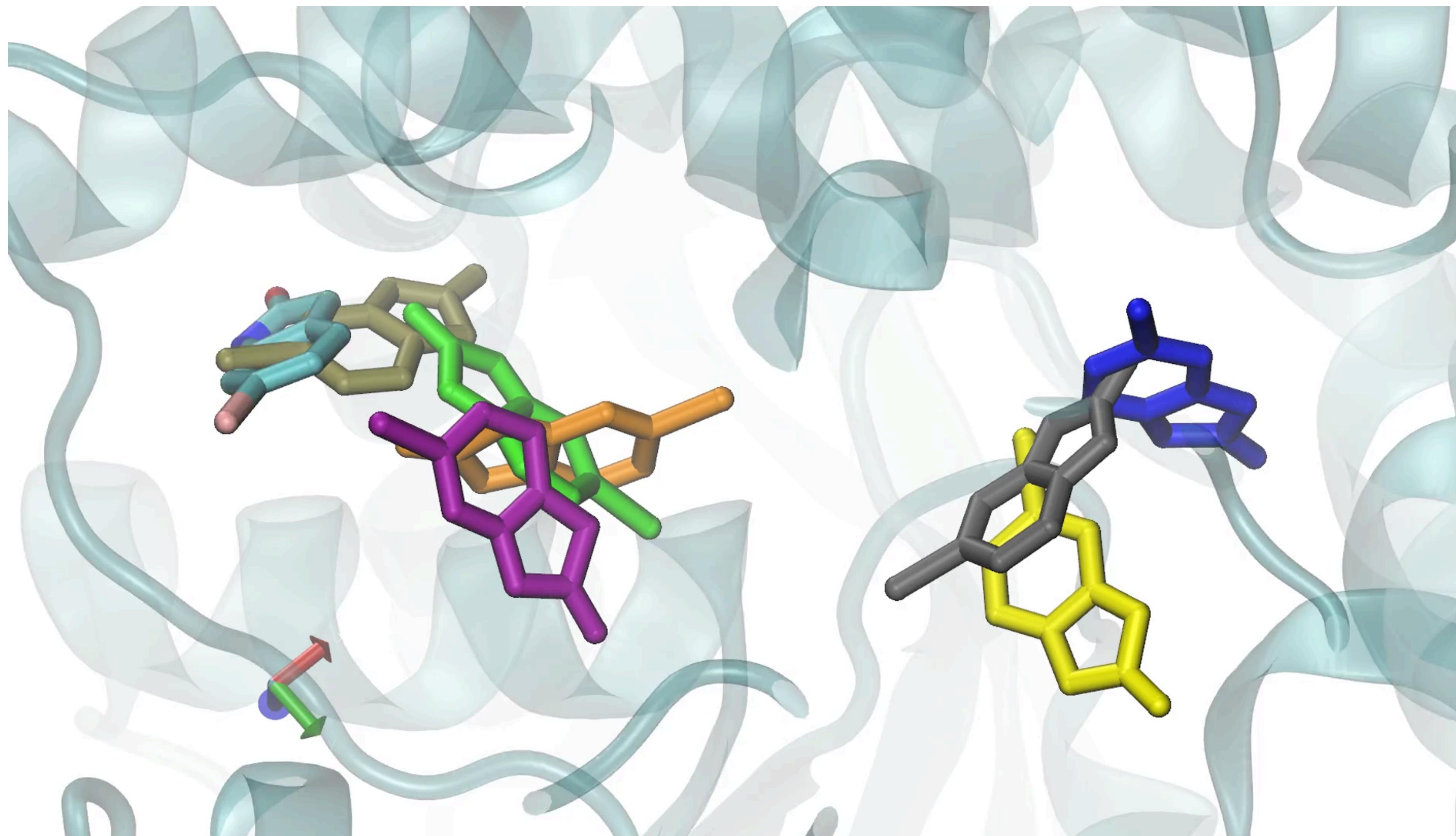
Blind prediction challenge
from
Greg Warren & Chris Bayly



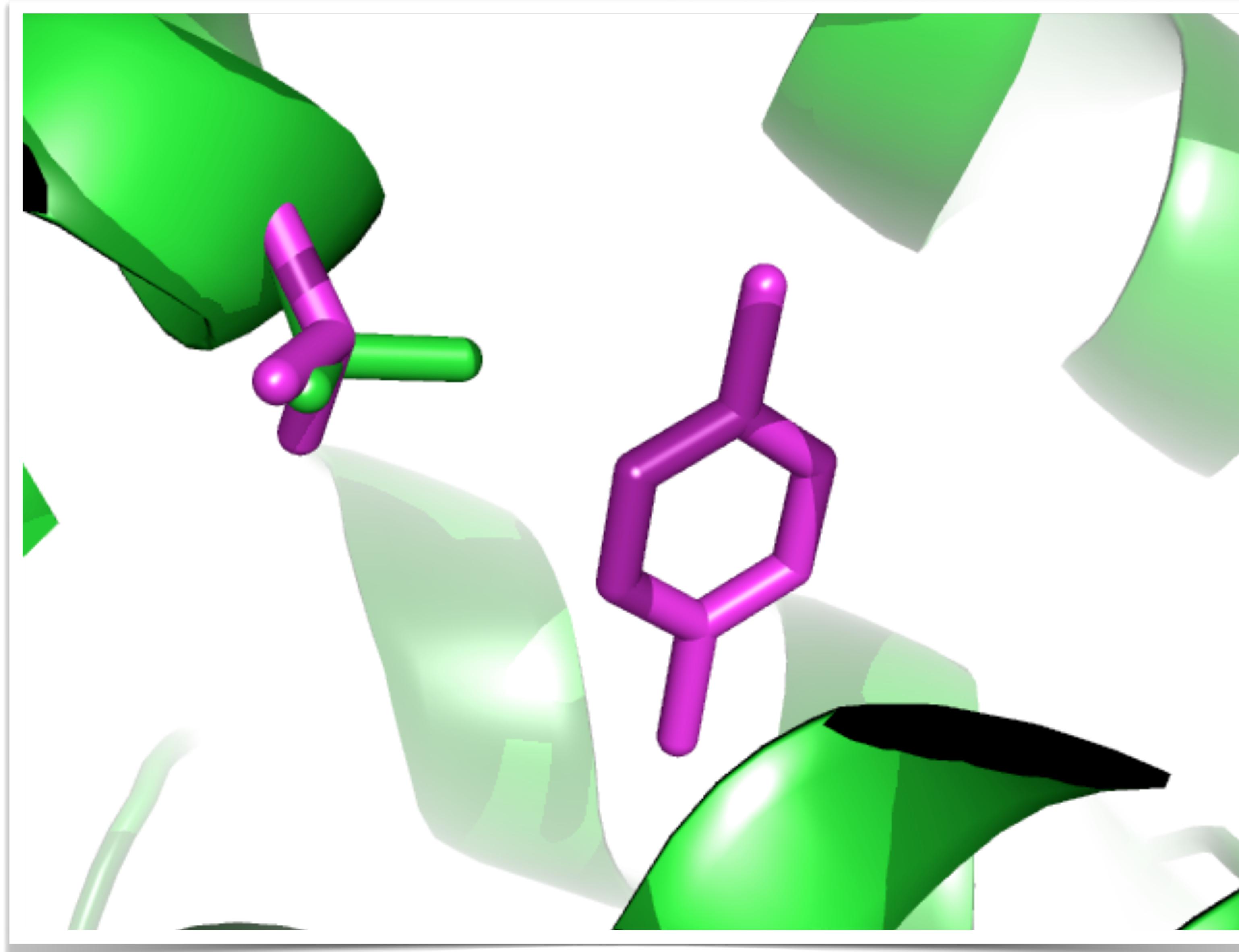
We tried to generalize this via a molecular darting approach, but it's proven challenging



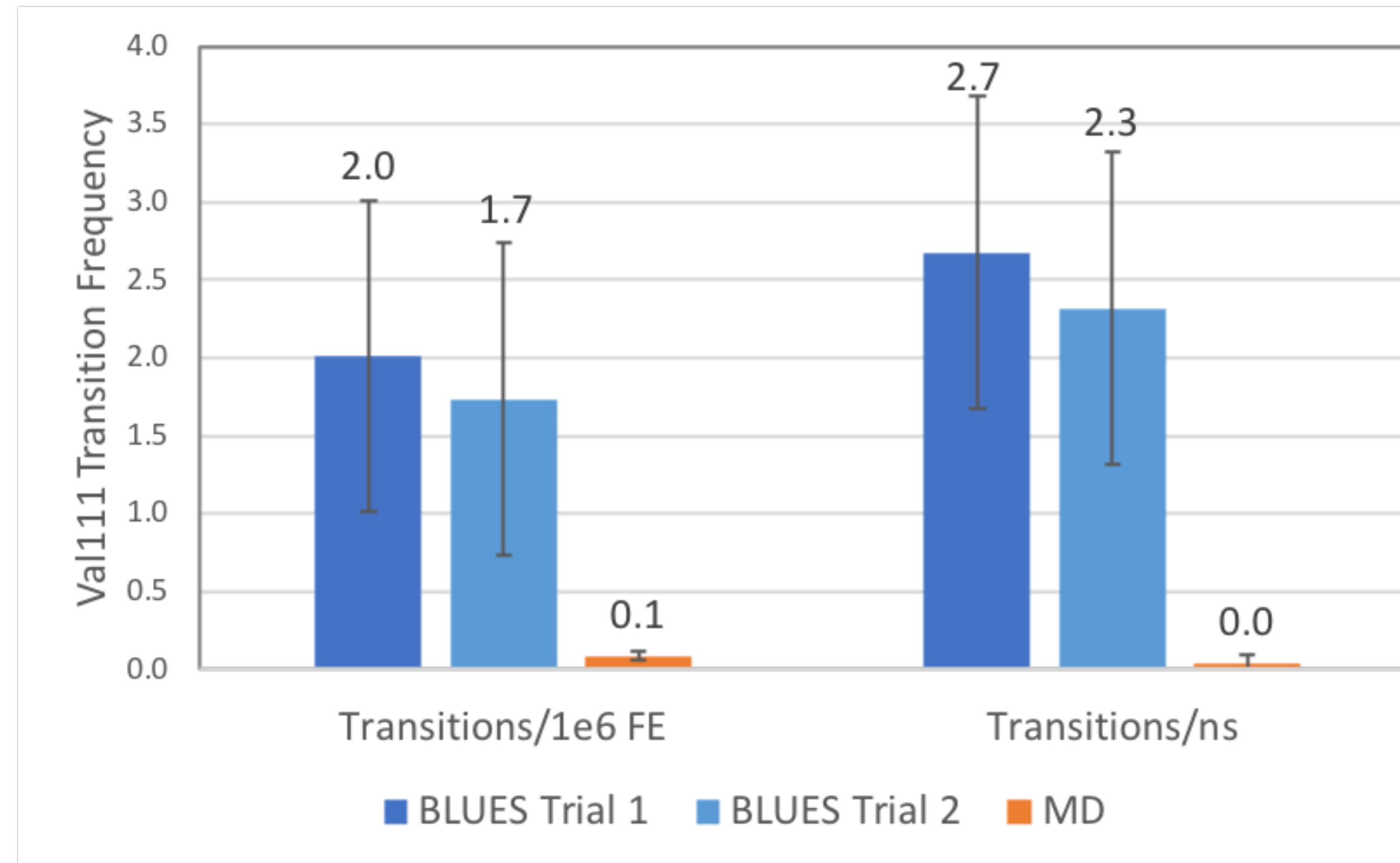
We tried to generalize this via a molecular darting approach, but it's proven challenging



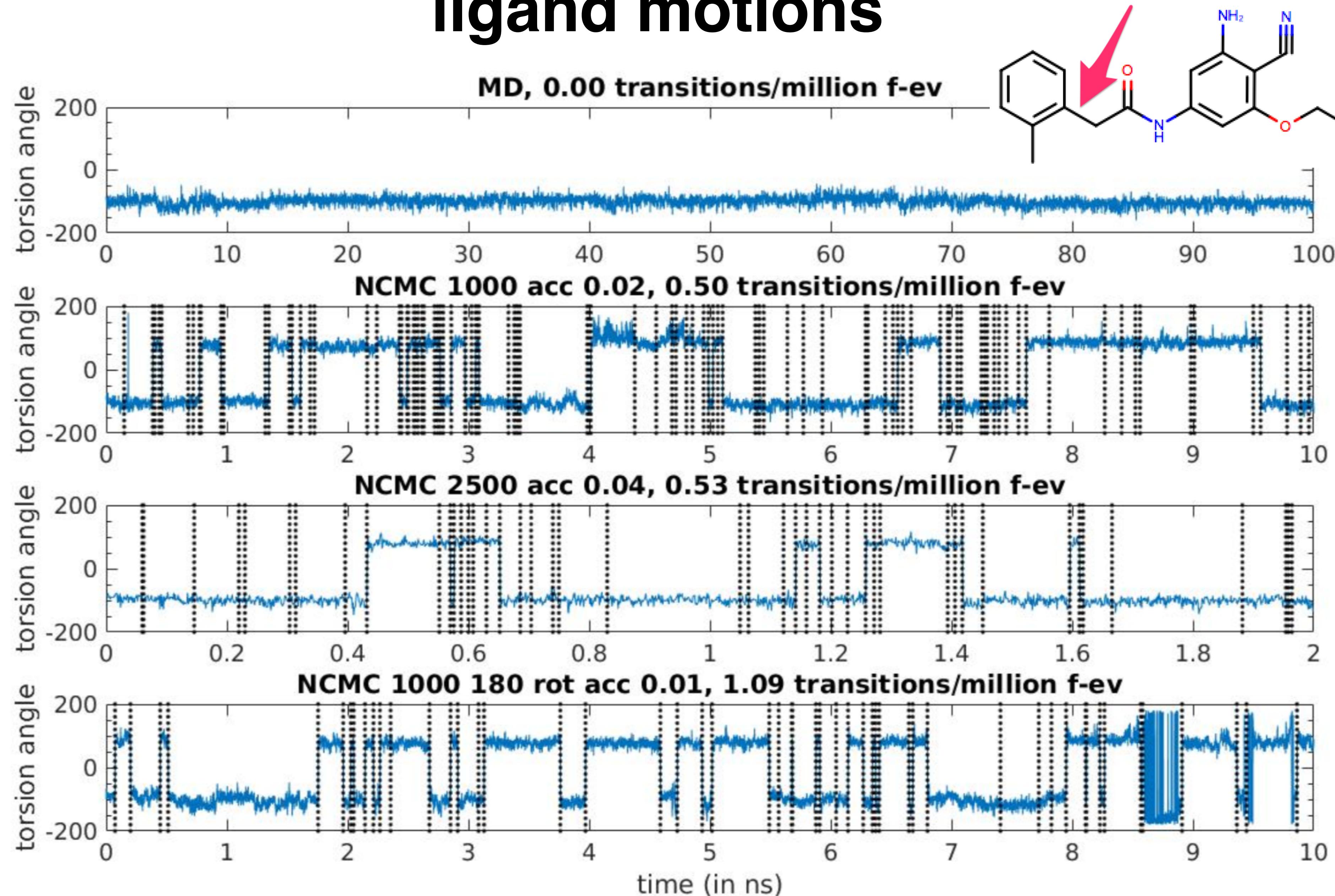
**Some other move types are rather useful,
such as sidechain moves**



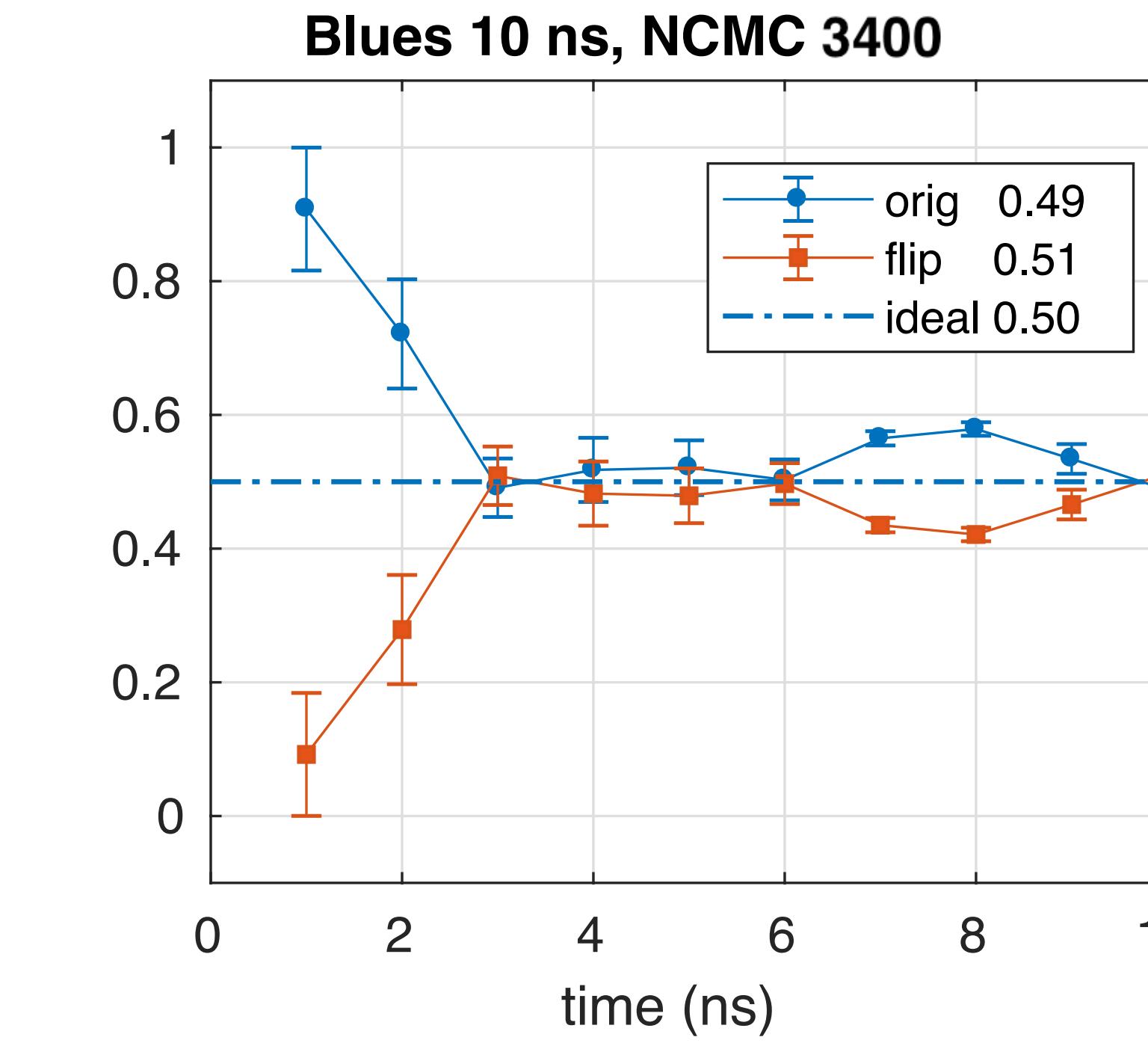
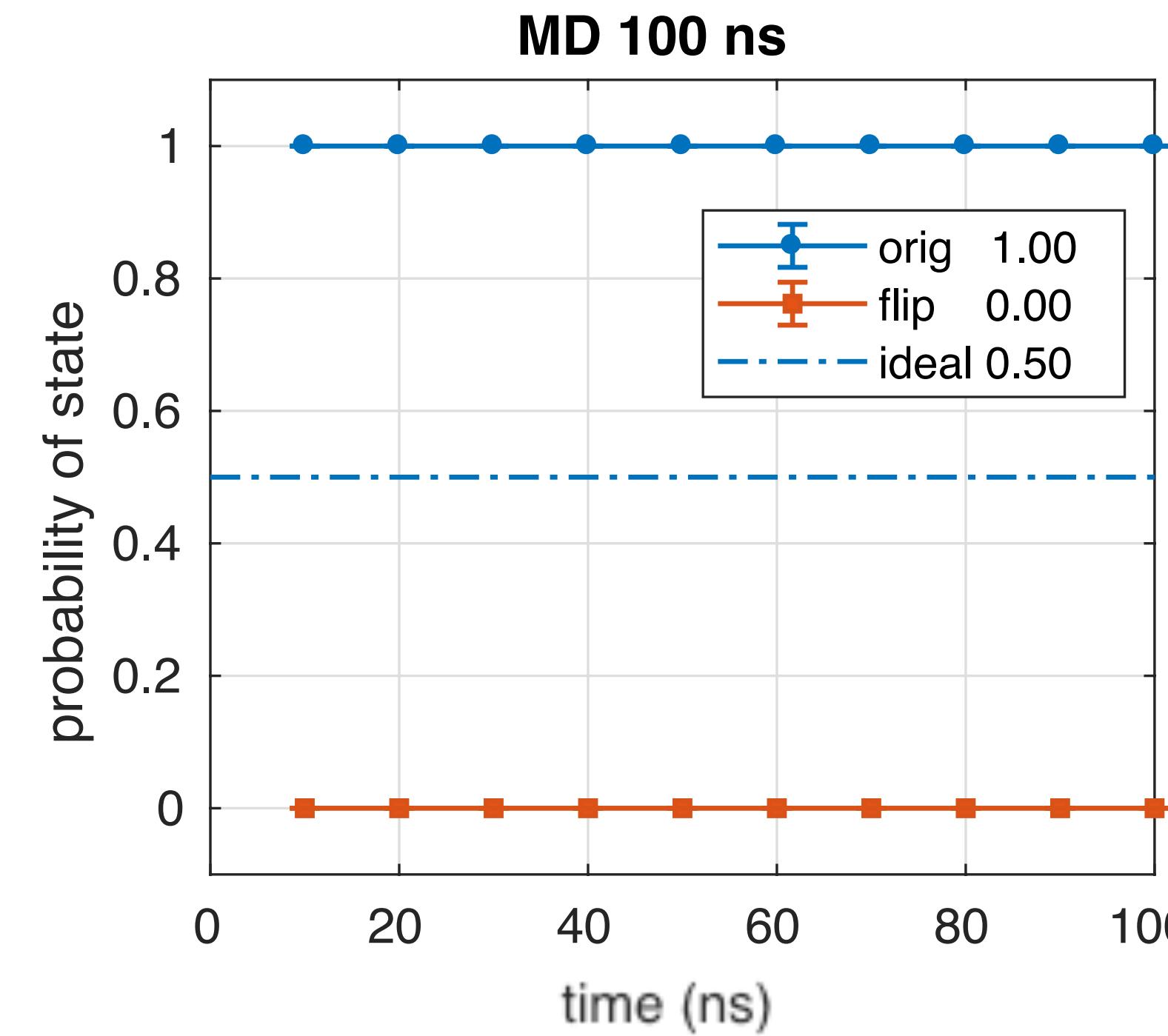
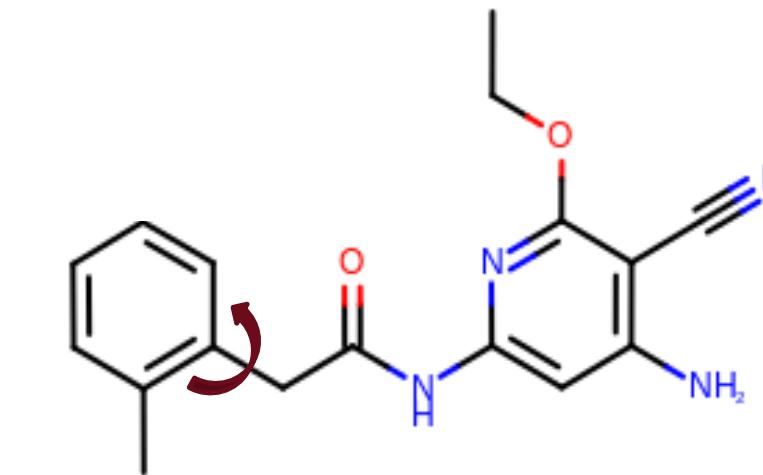
Some other move types are rather useful, such as sidechain moves



... and enhanced sampling of internal ligand motions

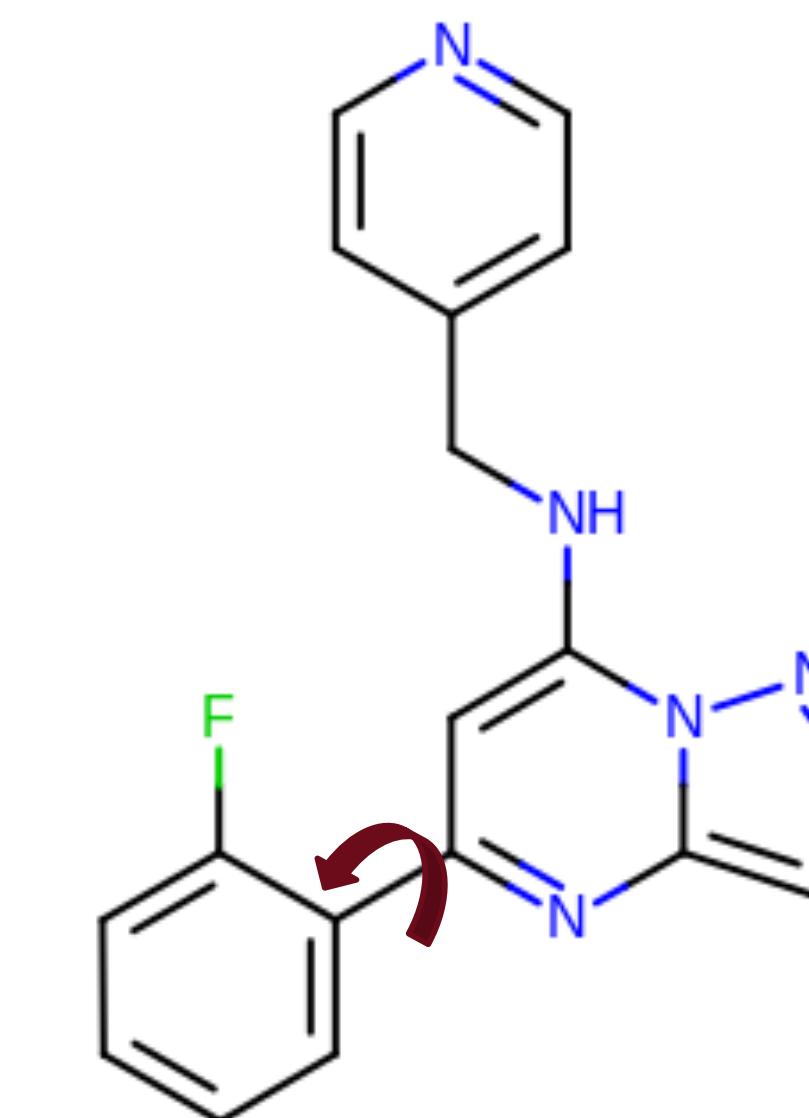
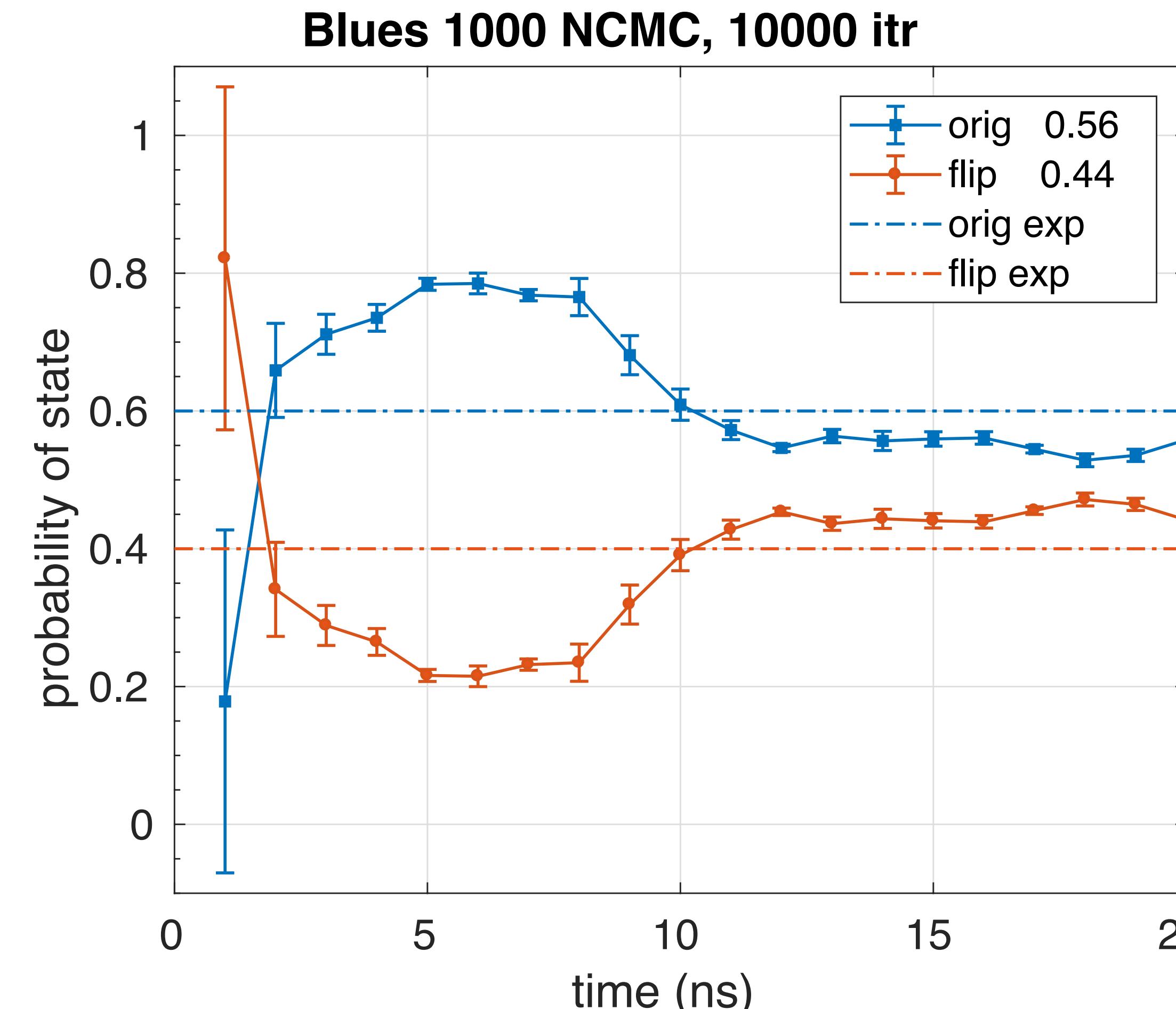


This results in faster convergence



NCMC converges to the correct population!

A CDK2 inhibitor has known binding modes with experimental occupancies

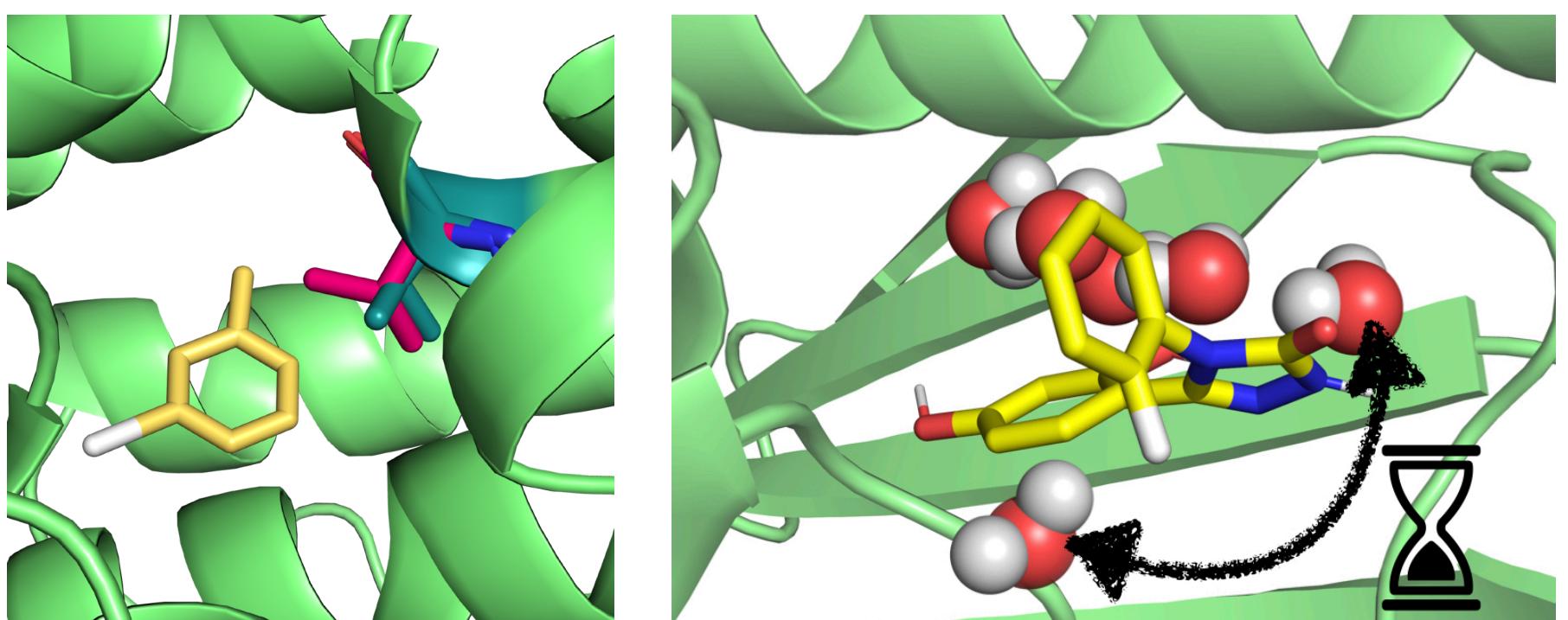


Experimental error bar $\pm 10\%$

Summary

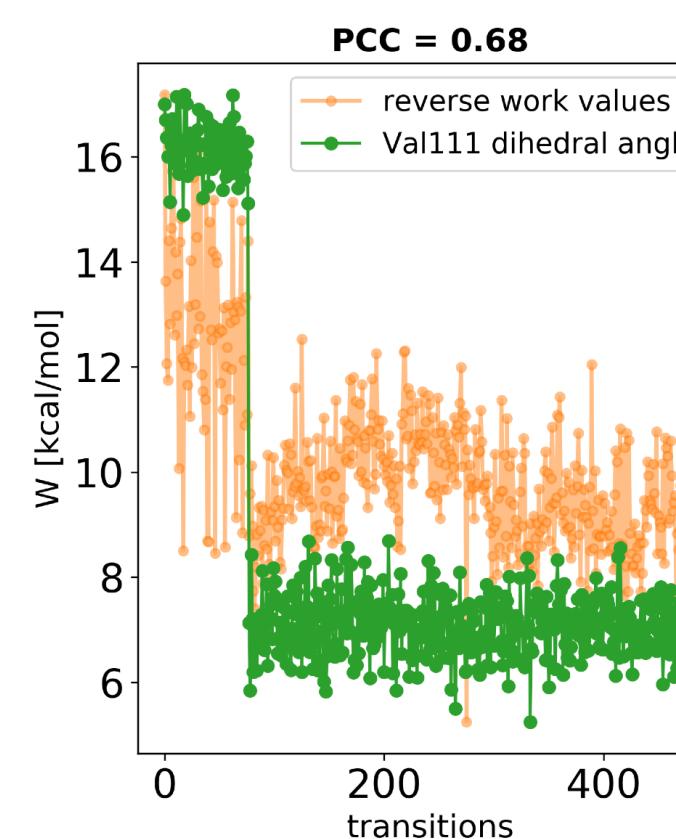
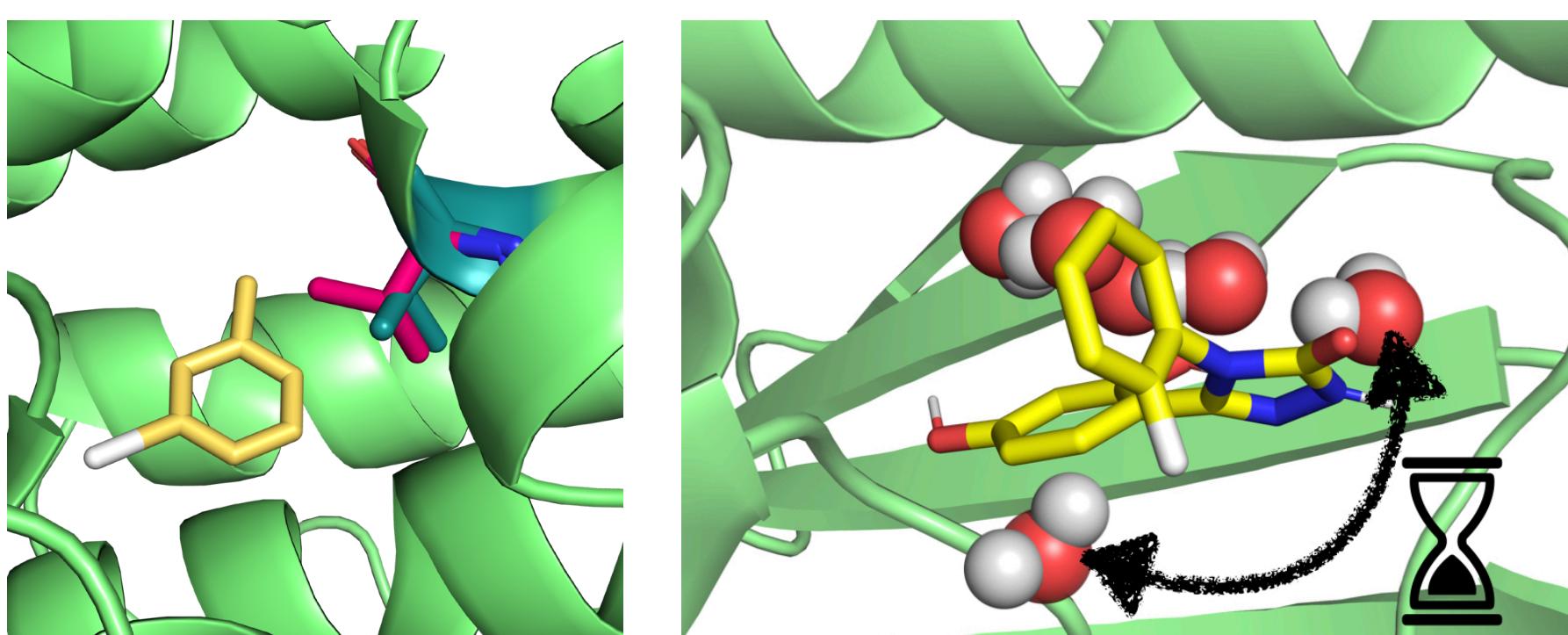
Summary

We find slow sampling of
various degrees of freedom
in FEC

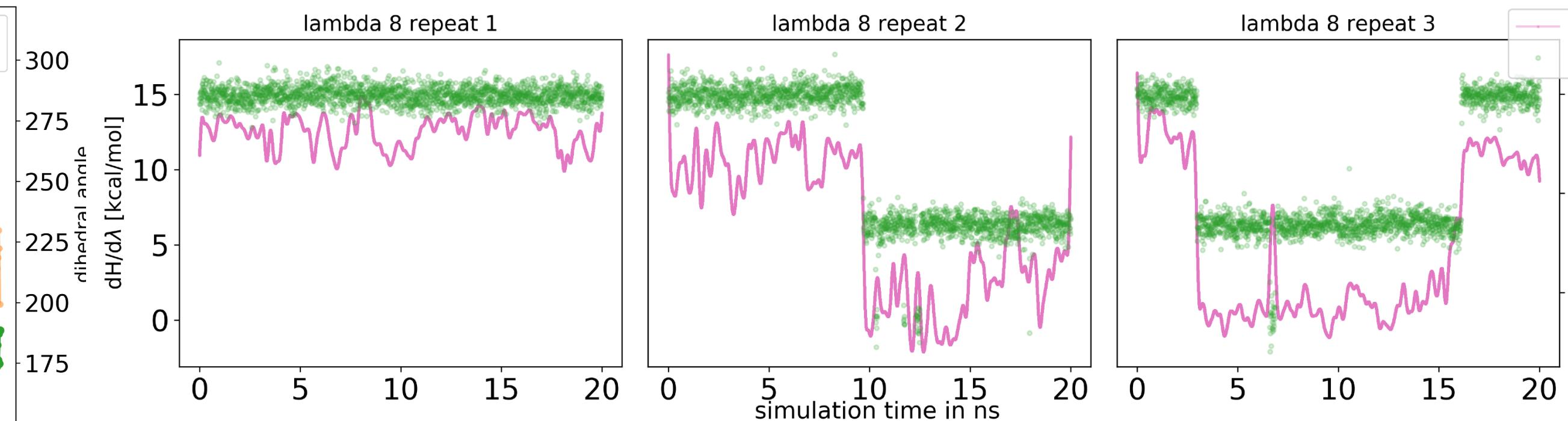


Summary

We find slow sampling of various degrees of freedom in FEC

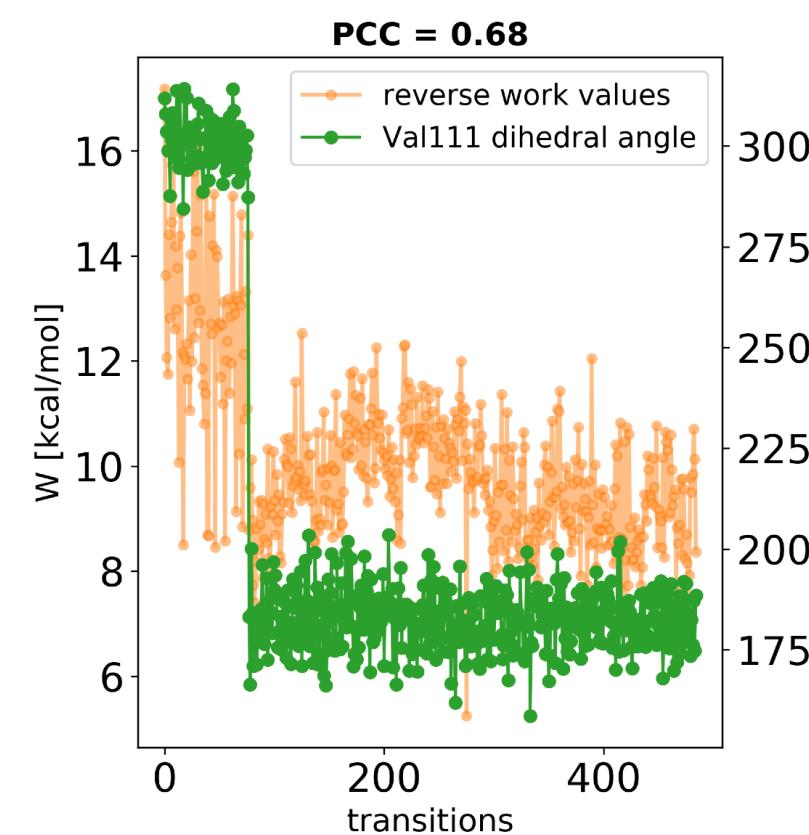
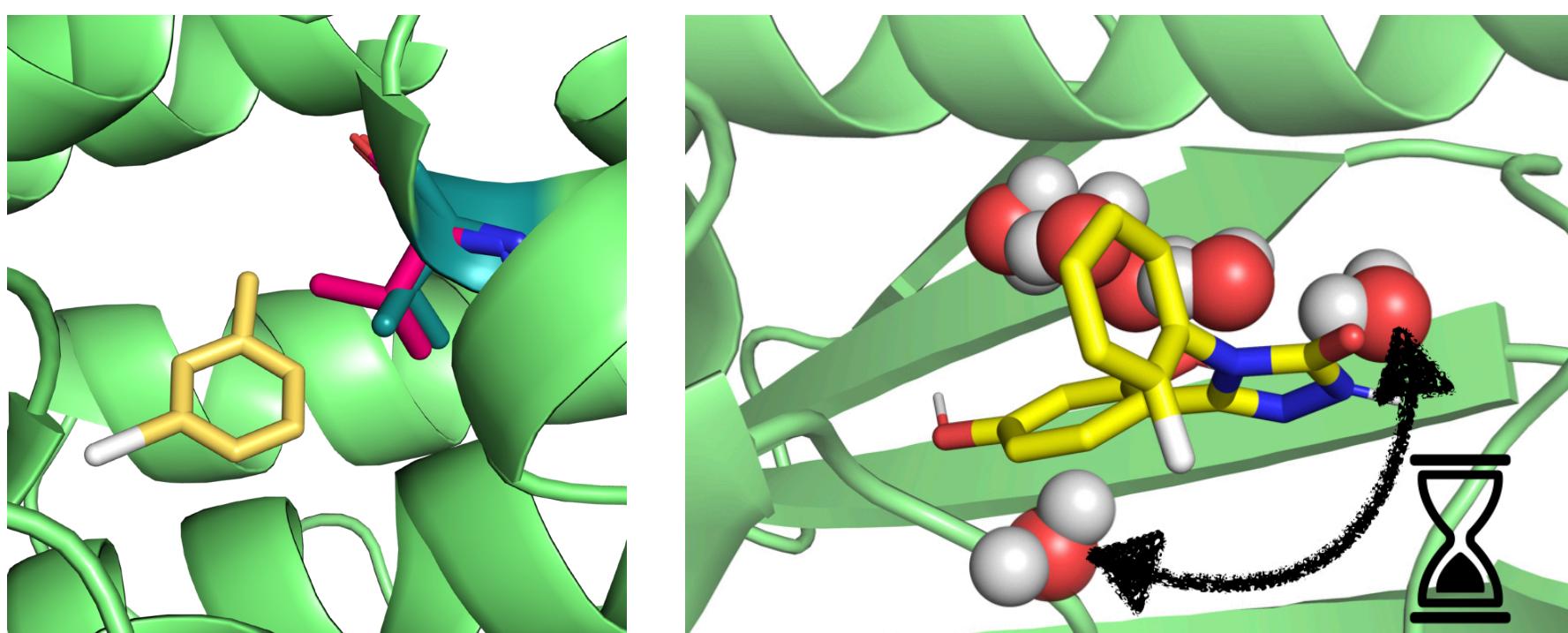


The Pearson correlation coefficient and dH/dl values can help identify sampling problems

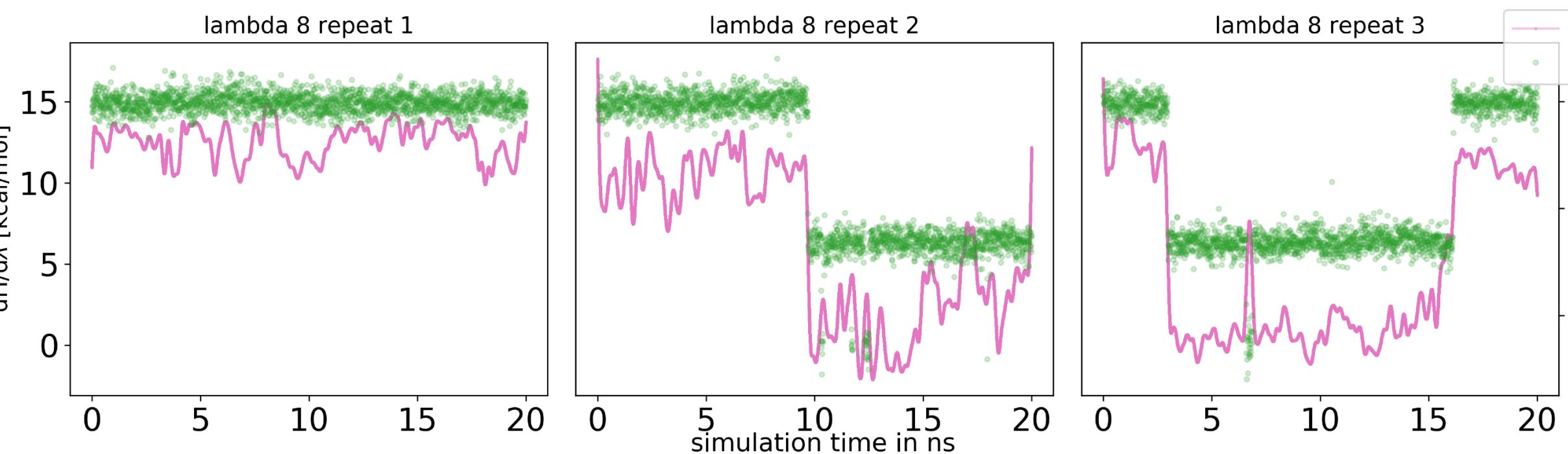


Summary

We find slow sampling of various degrees of freedom in FEC

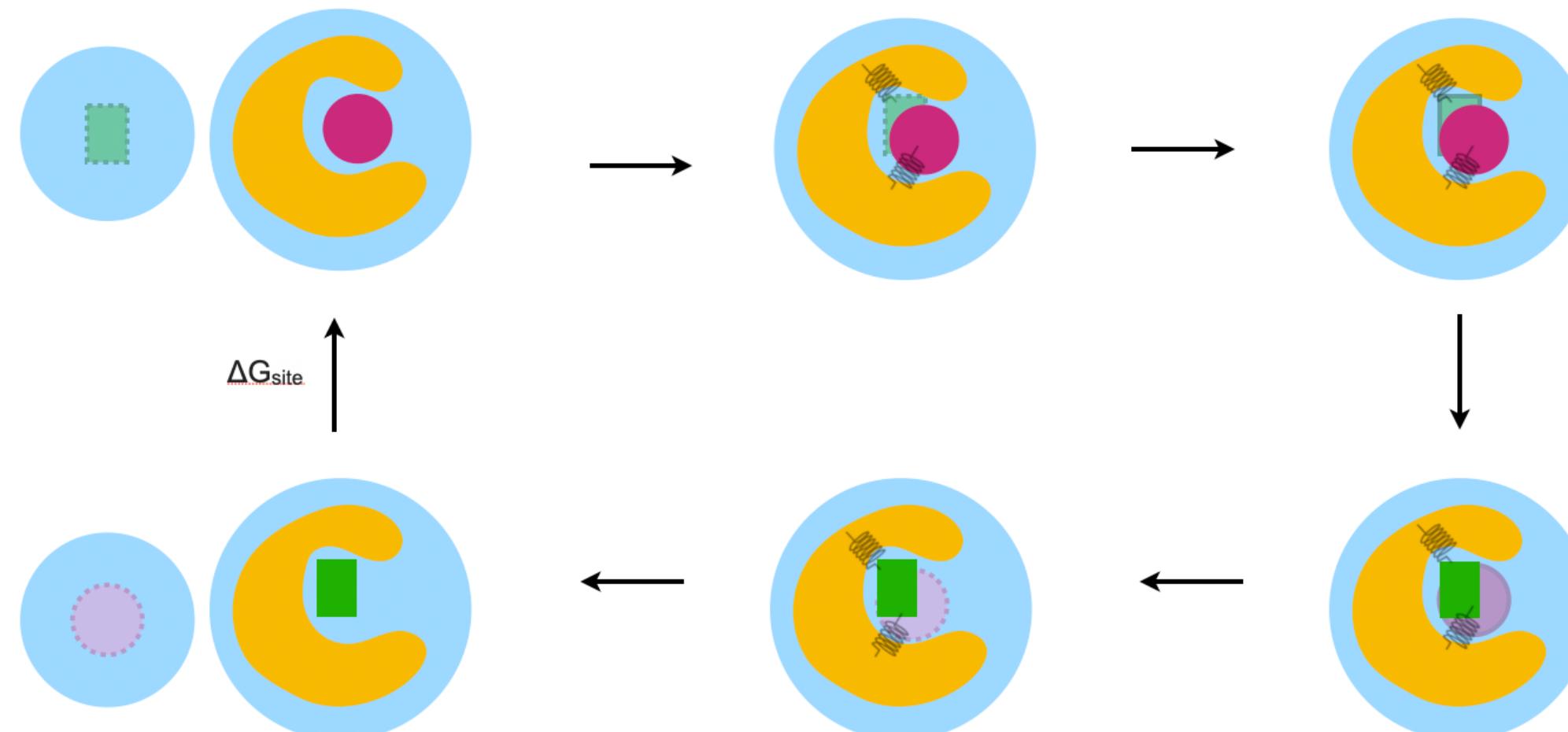


The Pearson correlation coefficient and dH/dl values can help identify sampling problems



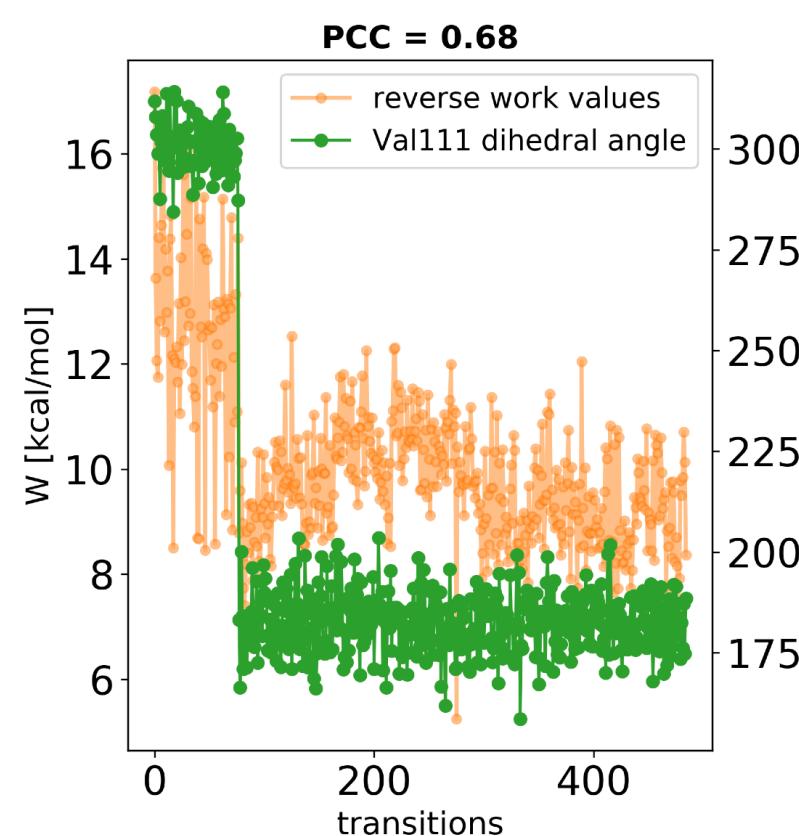
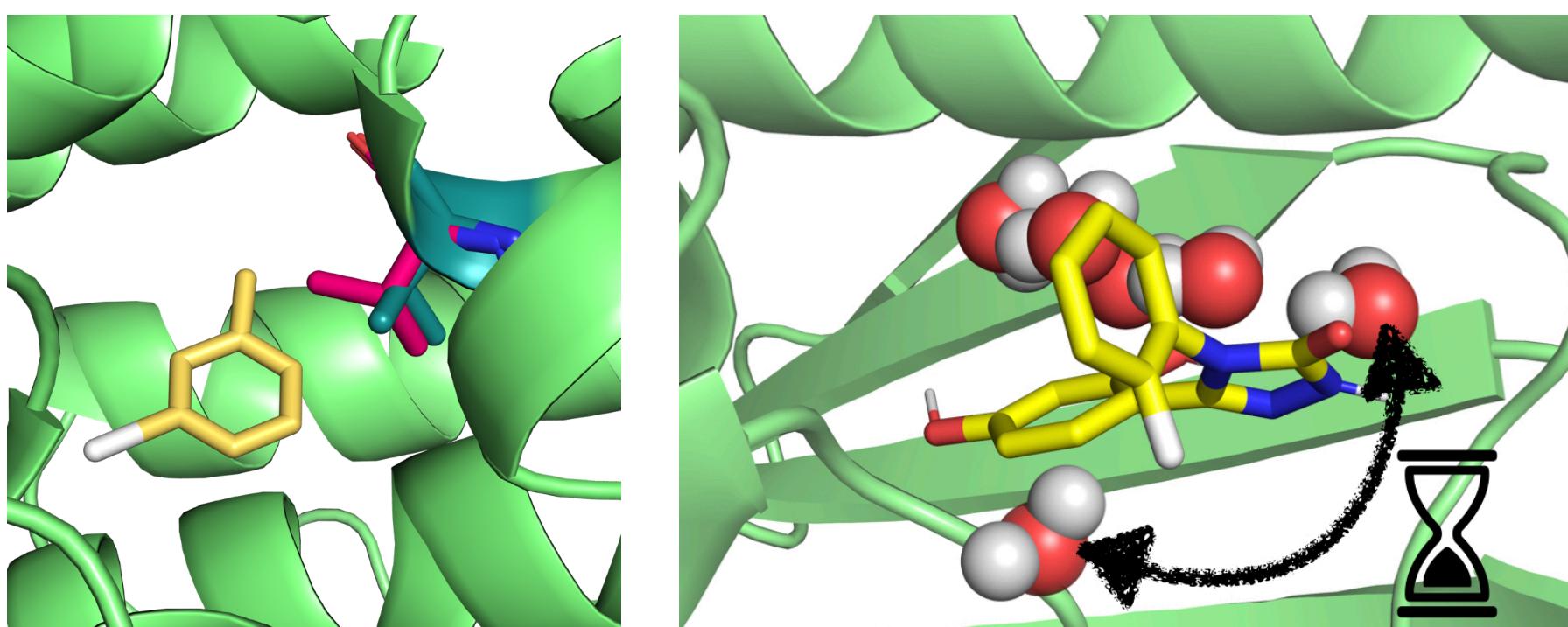
We're exploring separated topologies

as a middle ground between relative and absolute

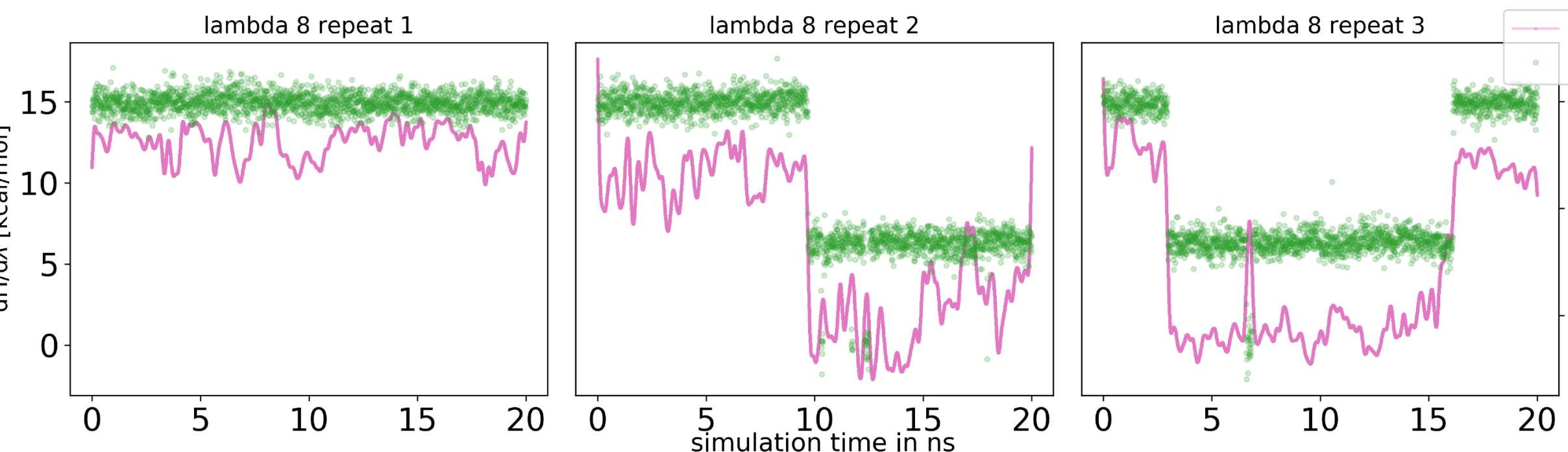


Summary

We find slow sampling of various degrees of freedom in FEC

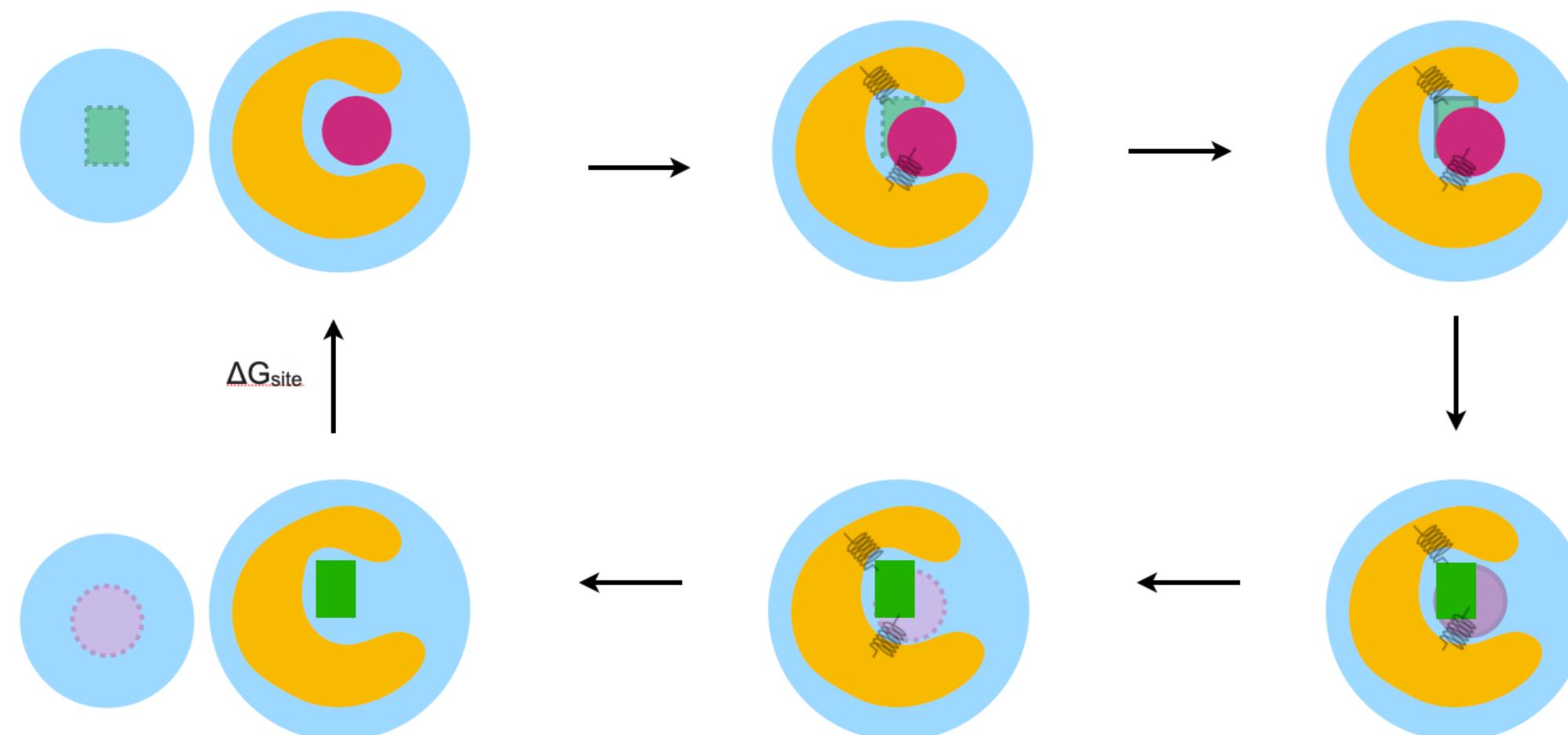


The Pearson correlation coefficient and dH/dl values can help identify sampling problems



We're exploring separated topologies

as a middle ground between relative and absolute



There's still more work to be done on sampling of binding modes

Acknowledgements



Hannah Baumann
David Wych
Yunhui Ge
Ana Caldaruse
Jessica Maat
Chris Zhang
Martin Amezcue

Oanh Tran
Trevor Gokey
Anjali Dixit
Danielle Bergazin
Pavan Behara
Lily Wang



Pfizer for funding & collaboration:
Eric Dybeck
Vishnu Sresht

Bert de Groot
Vytautas Gapsys



David Hahn
Lorenzo D'Amore