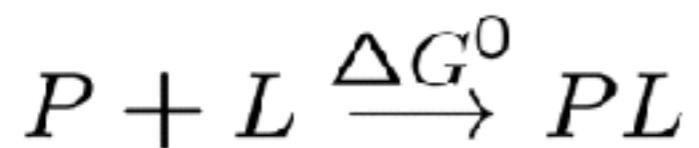
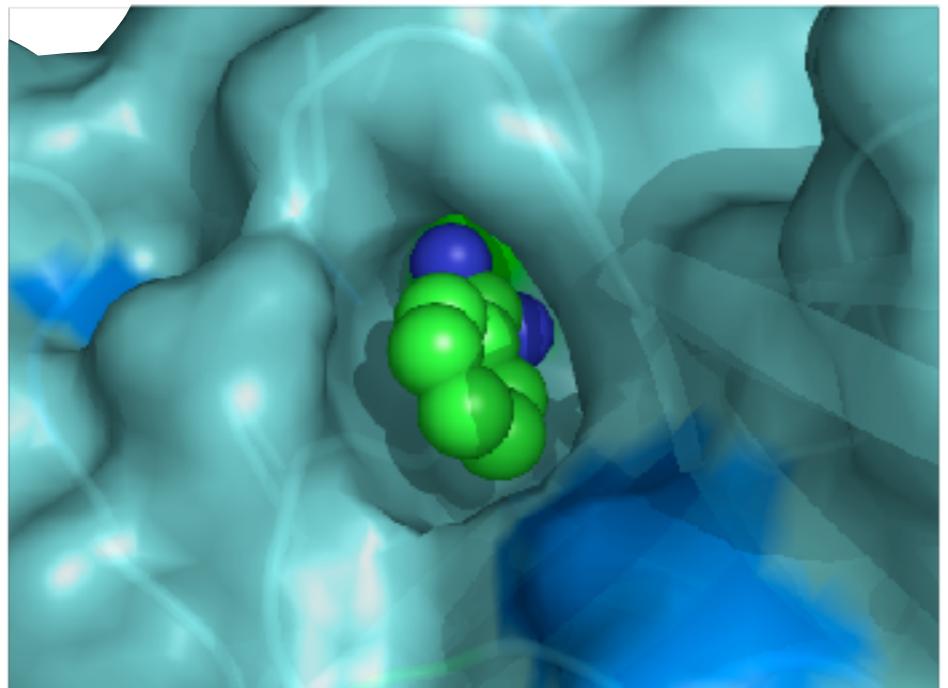


# Free energy calculations: What we want them for, what they can do, and how they work

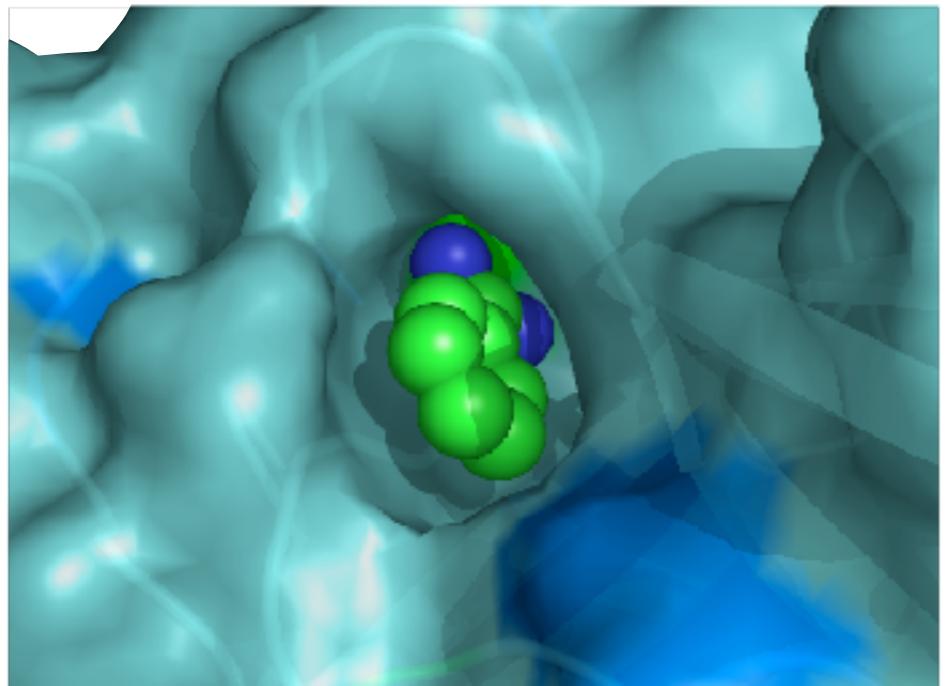
David Mobley



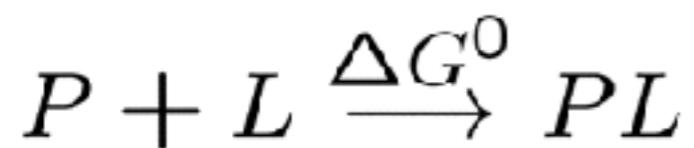
# What could we do with accurate binding free energies?



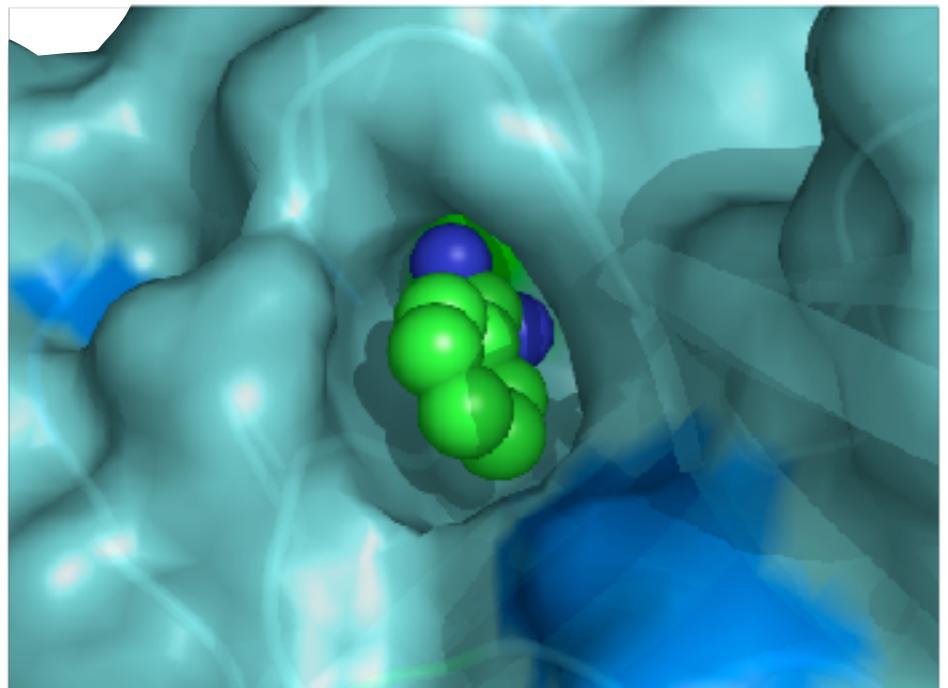
# What could we do with accurate binding free energies?



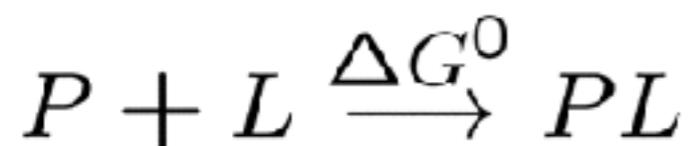
- Design molecules to manipulate protein function



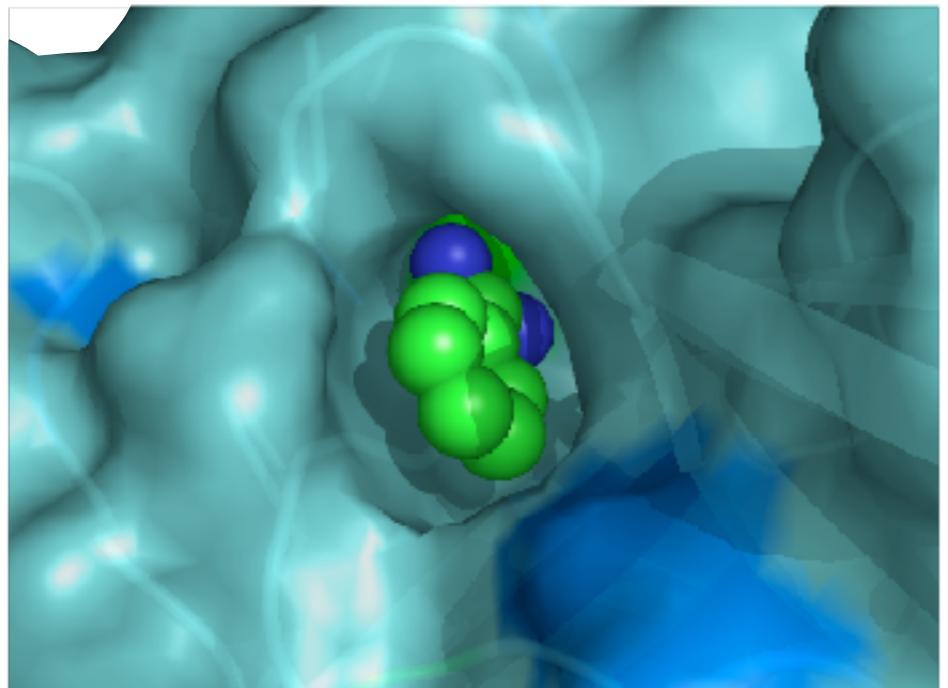
# What could we do with accurate binding free energies?



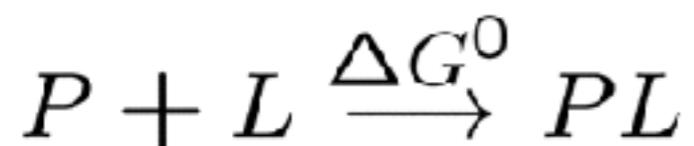
- Design molecules to manipulate protein function
- Recognize toxins



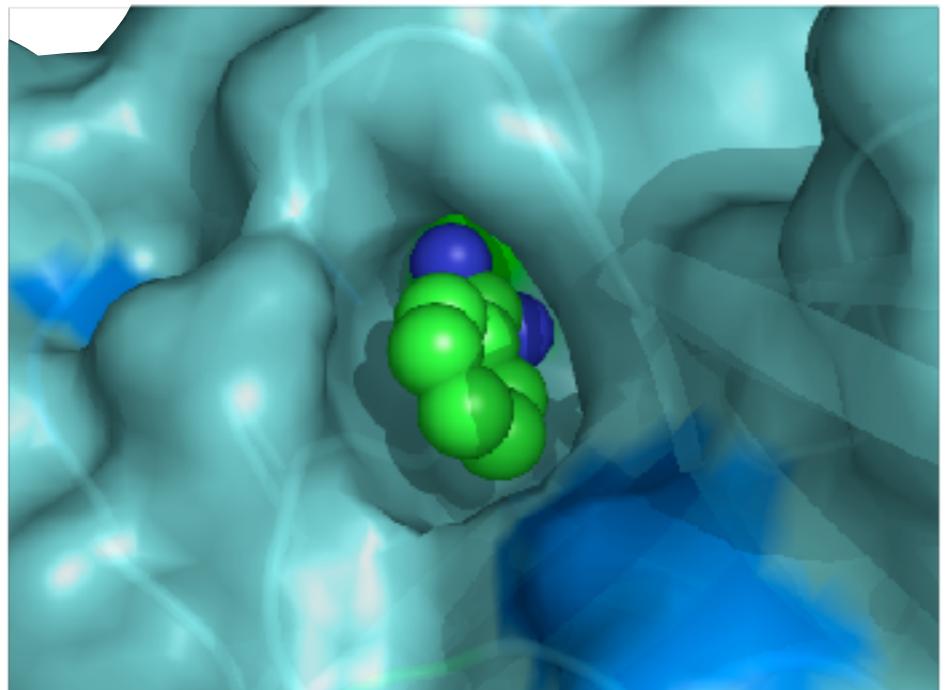
# What could we do with accurate binding free energies?



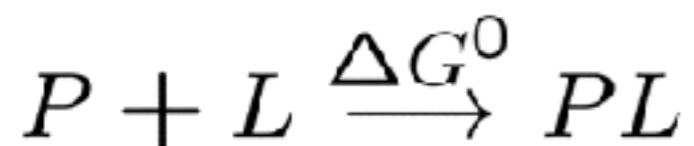
- Design molecules to manipulate protein function
- Recognize toxins
- Identify enzyme functions



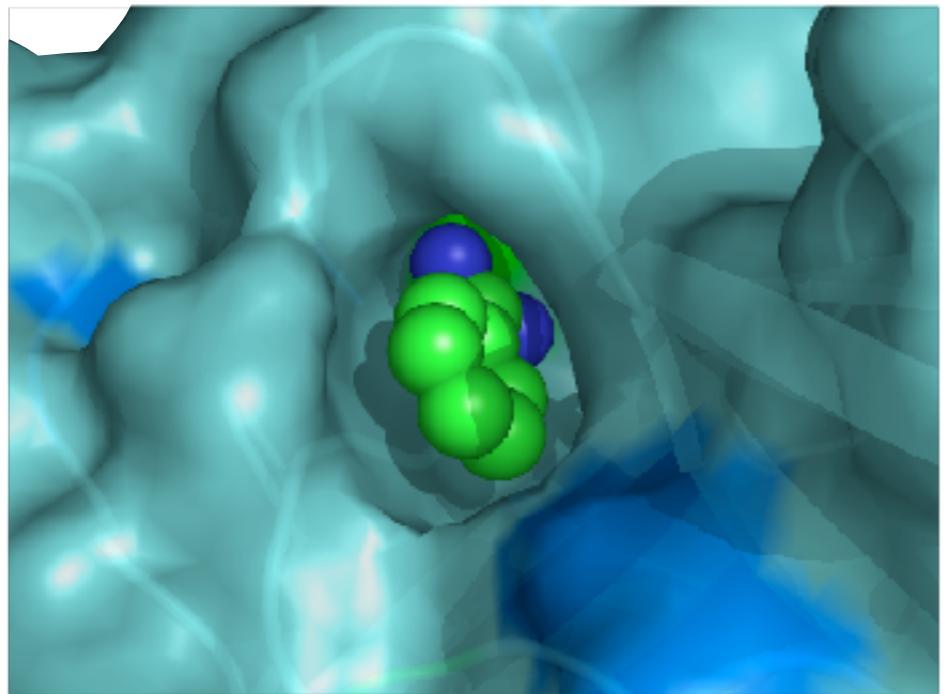
# What could we do with accurate binding free energies?



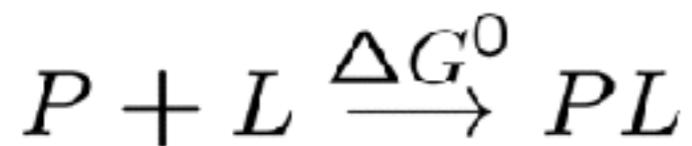
- Design molecules to manipulate protein function
- Recognize toxins
- Identify enzyme functions
- Protein design: Design binders to target molecule



# What could we do with accurate binding free energies?



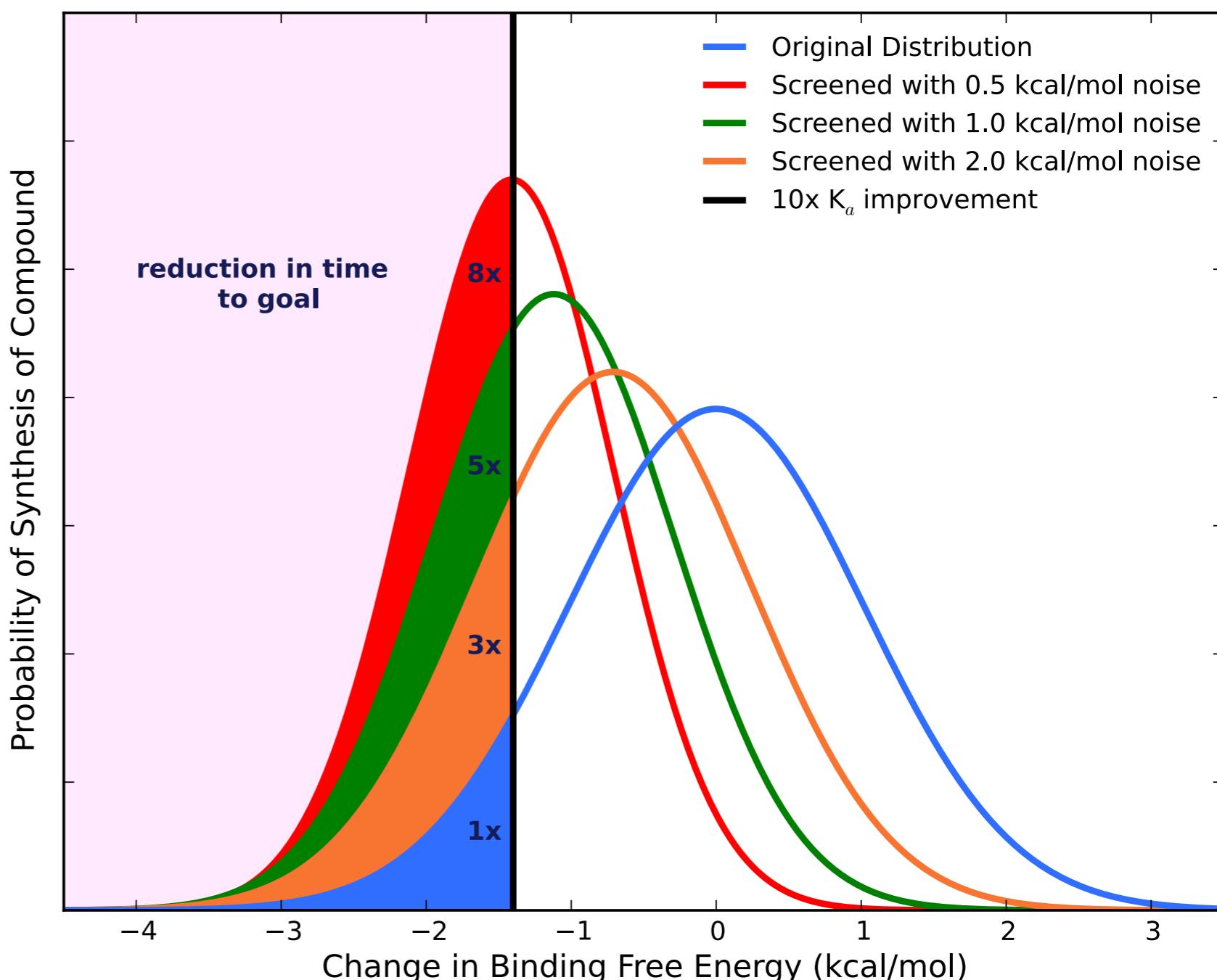
- Design molecules to manipulate protein function
- Recognize toxins
- Identify enzyme functions
- Protein design: Design binders to target molecule
- Aid medicinal chemistry, guide synthesis



# Even modest accuracies in calculated binding free energies can have significant benefits

## Hypothetical pipeline:

- Medicinal chemist suggests 100 derivatives or compounds per week
- Your job is to pick the top 10 to carry forward



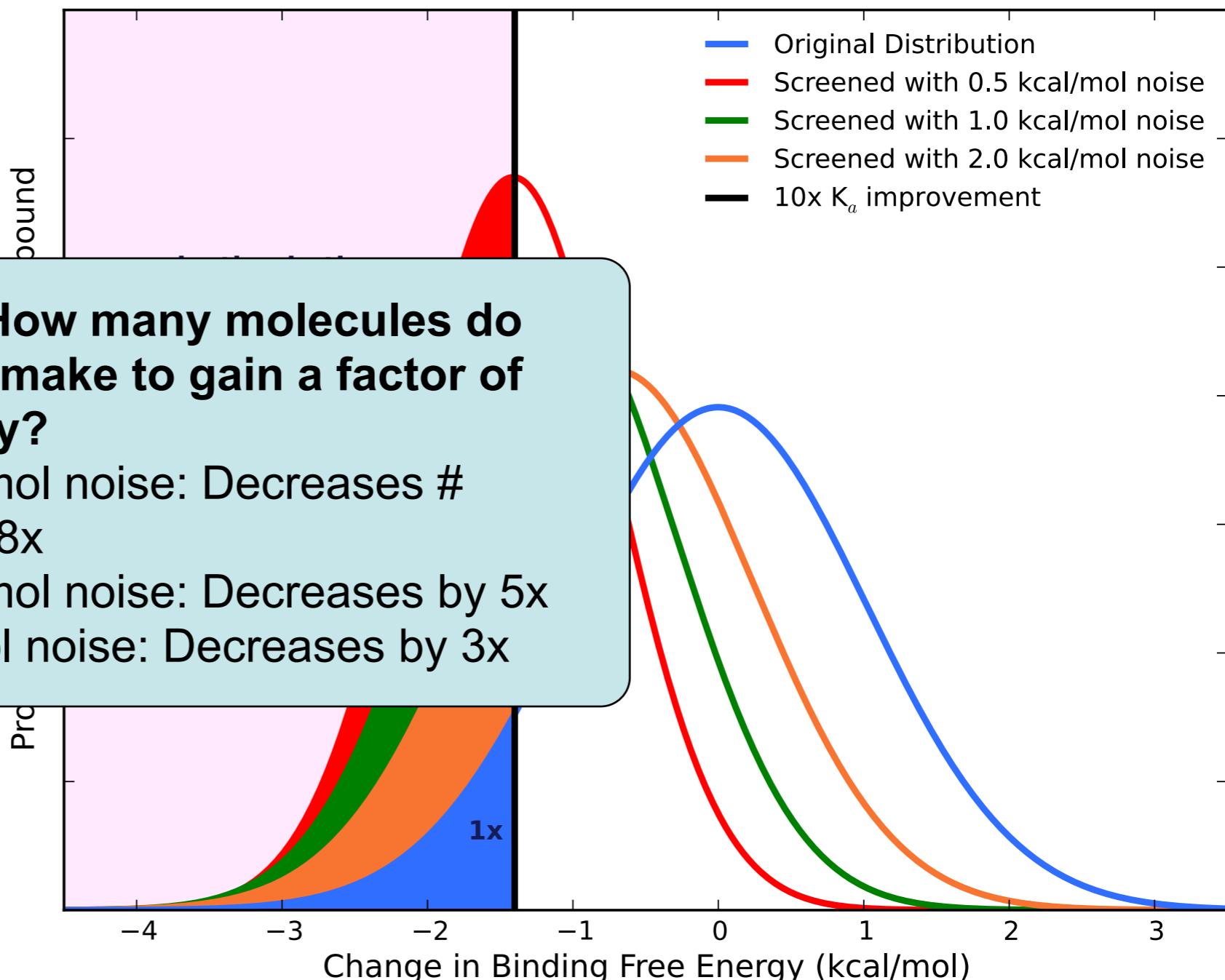
# Even modest accuracies in calculated binding free energies can have significant benefits

## Hypothetical pipeline:

- Medicinal chemist suggests 100 derivatives or compounds per bound state
- Your job is to reduce the number to carry forward

**Question: How many molecules do we have to make to gain a factor of 10 in affinity?**

- 0.5 kcal/mol noise: Decreases # required by 8x
- 1.0 kcal/mol noise: Decreases by 5x
- 2 kcal/mol noise: Decreases by 3x

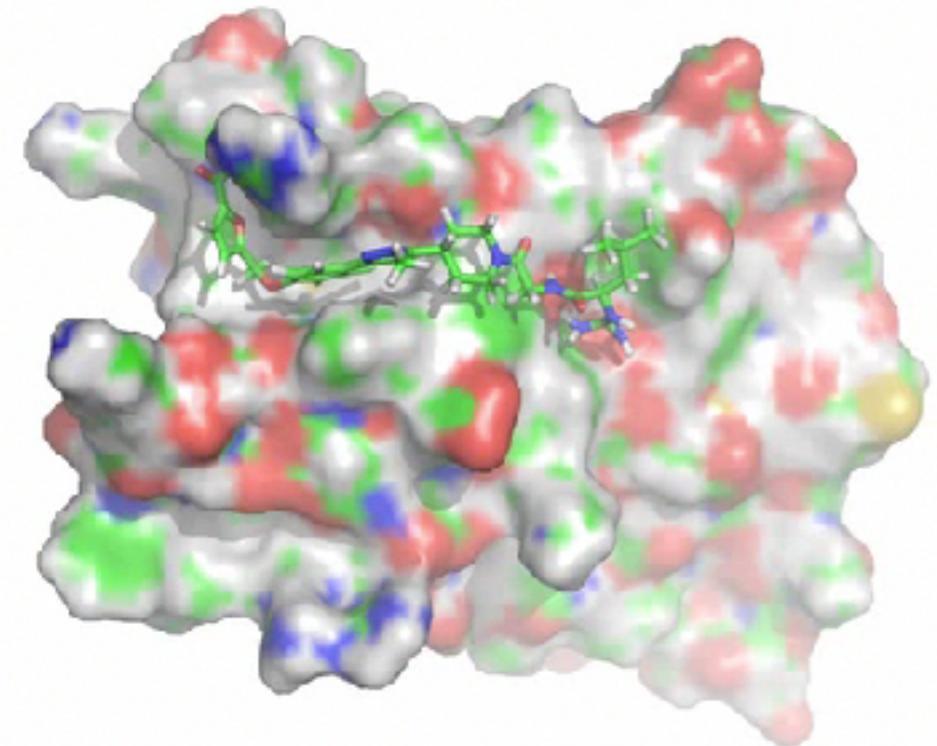


Binding free energies involve  
a ratio of partition functions

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

Binding free energies involve  
a ratio of partition functions

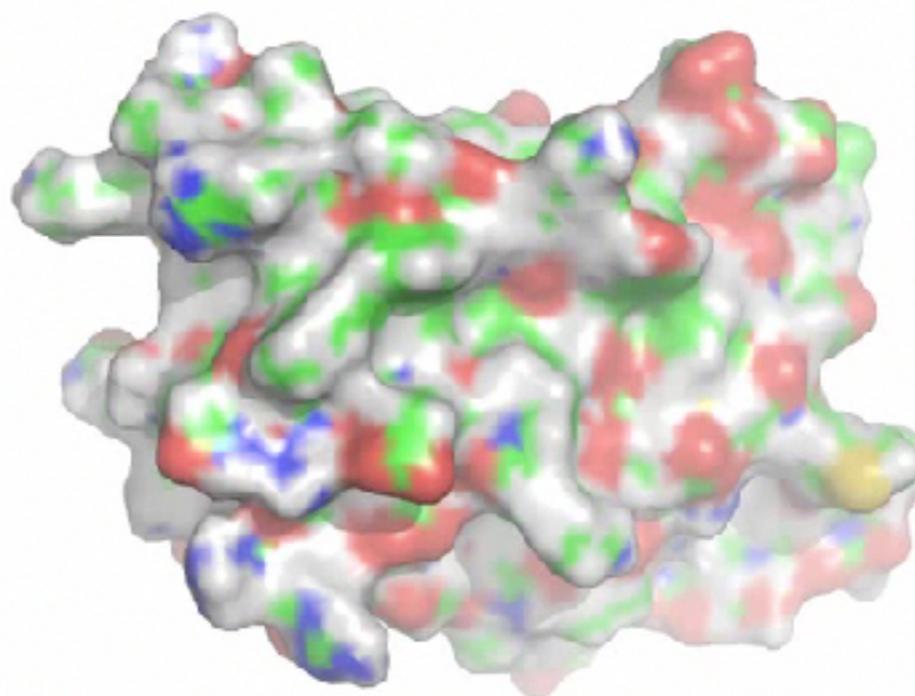
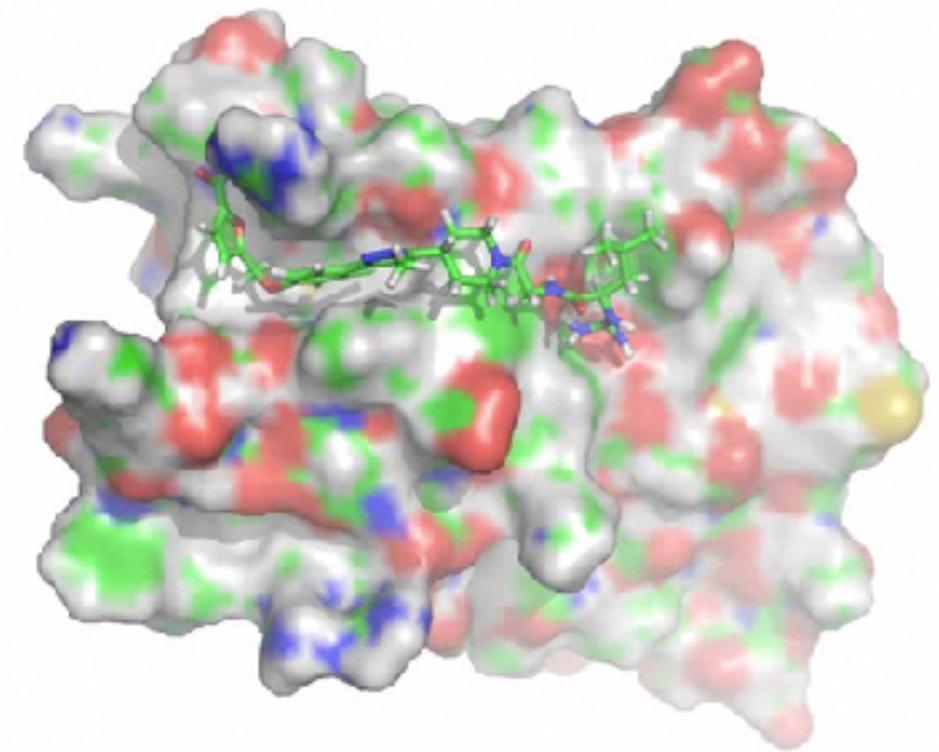
$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$



$$Q_{PL}$$

Binding free energies involve  
a ratio of partition functions

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

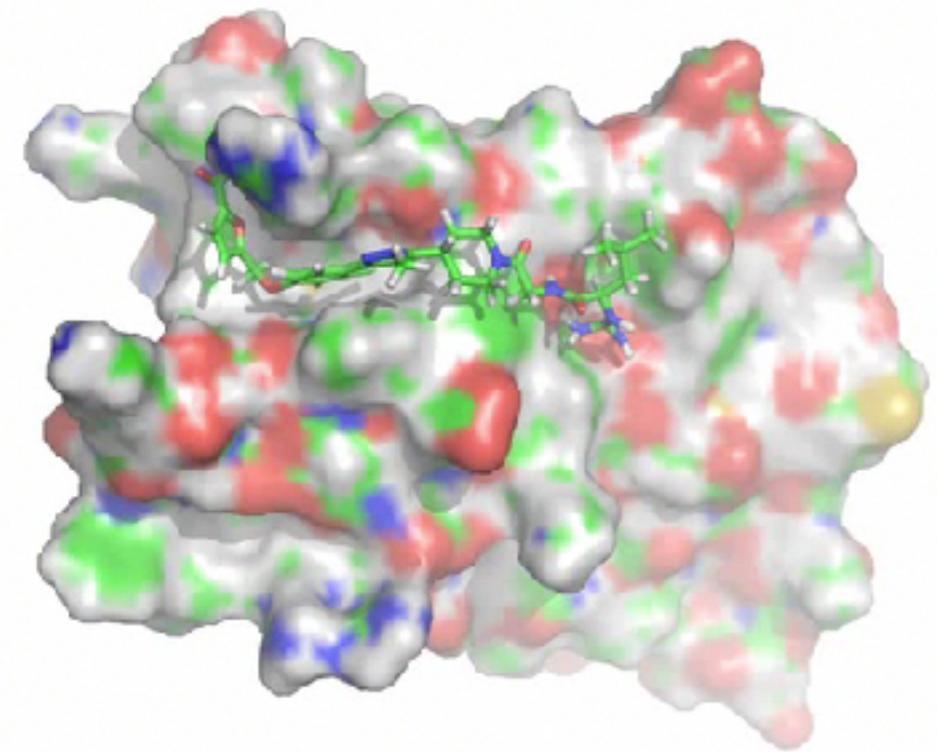


$$Q_{PL}$$

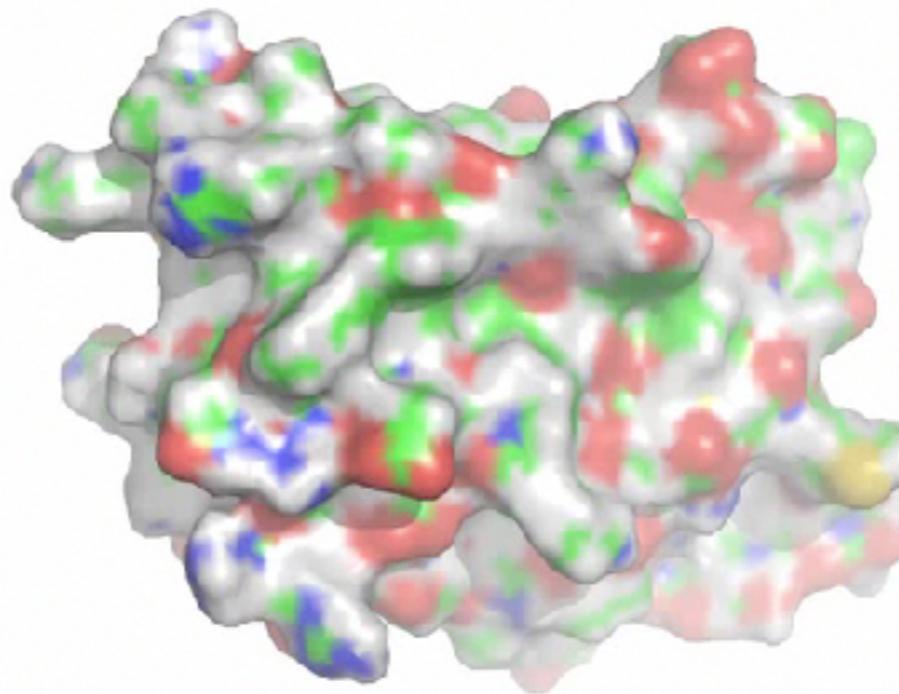
$$Q_P$$

Binding free energies involve  
a ratio of partition functions

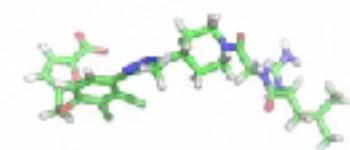
$$\Delta G = -k_B T \ln Q_{PL}/Q_P Q_L$$



$Q_{PL}$



$Q_P$



$Q_L$

# Docking is approximate

$$\Delta G = \Delta H - T\Delta S$$

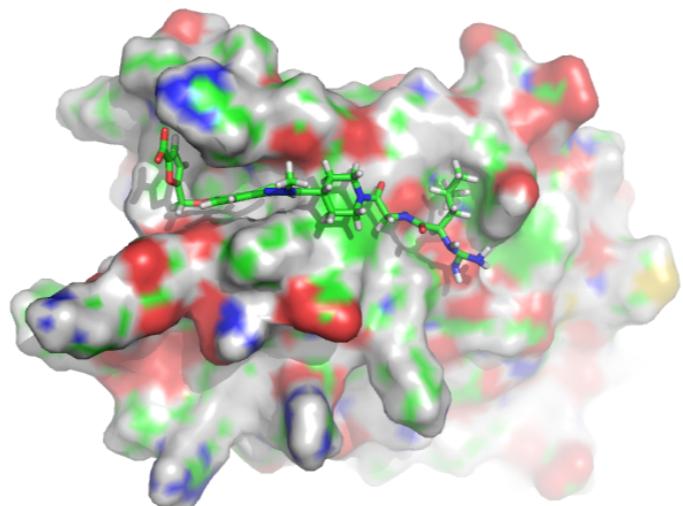
# Docking is approximate

$$\Delta G = \Delta H - T\cancel{\Delta S}$$

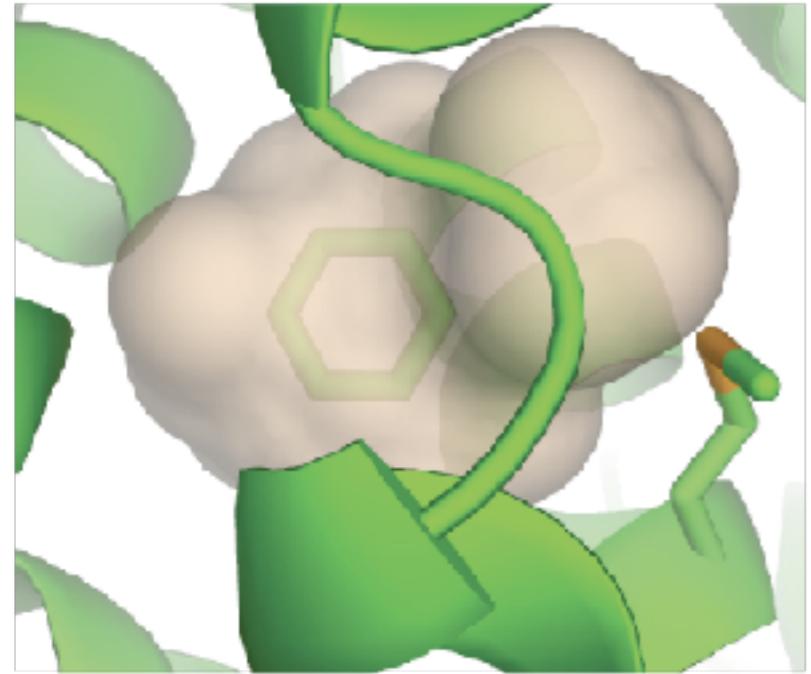
# Docking is approximate

$$\Delta G = \Delta H - \cancel{T\Delta S}$$

- Score  $\sim \Delta H$
- Sometimes add  $-\Delta G_{solv}$  to score
- Uses “optimal” single orientations

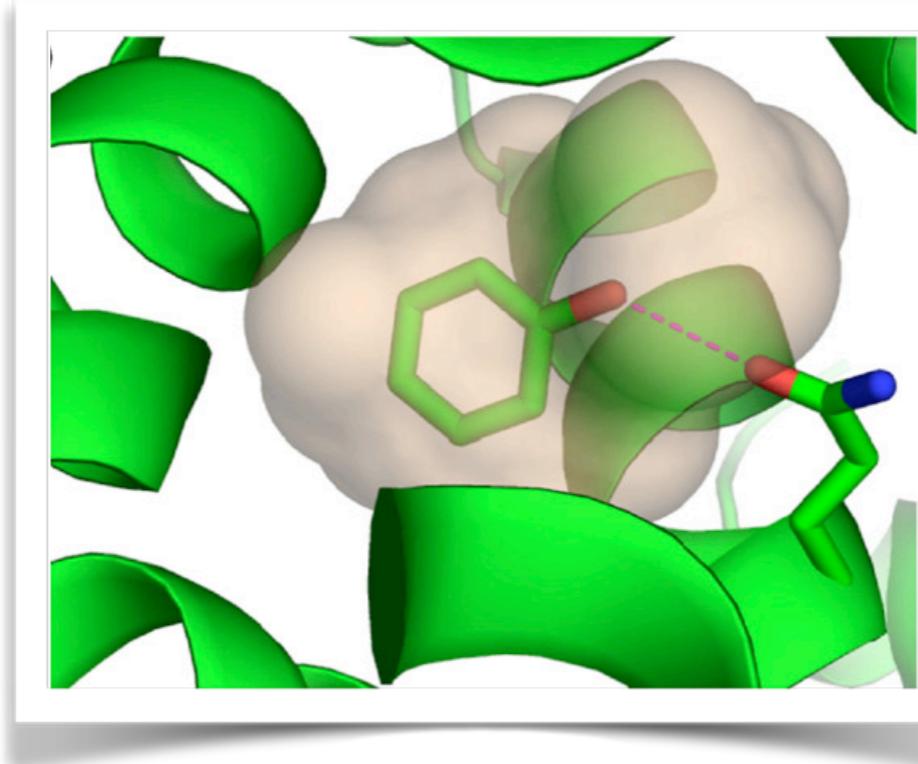


# We've been using a progression of model binding sites to test and improve these methods



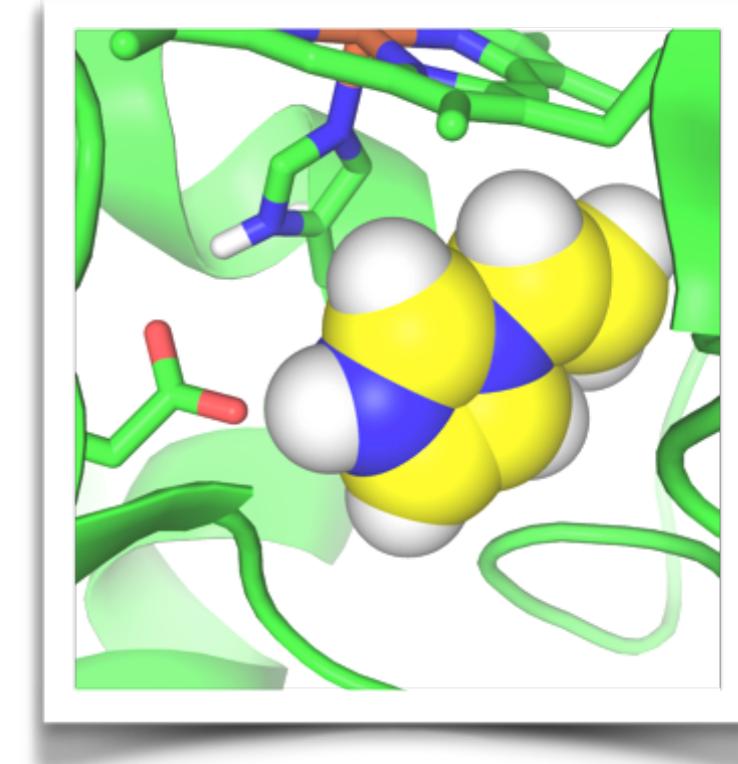
Lysozyme L99A

- Simple
- Nonpolar
- Dry



Lysozyme L99A/M102Q

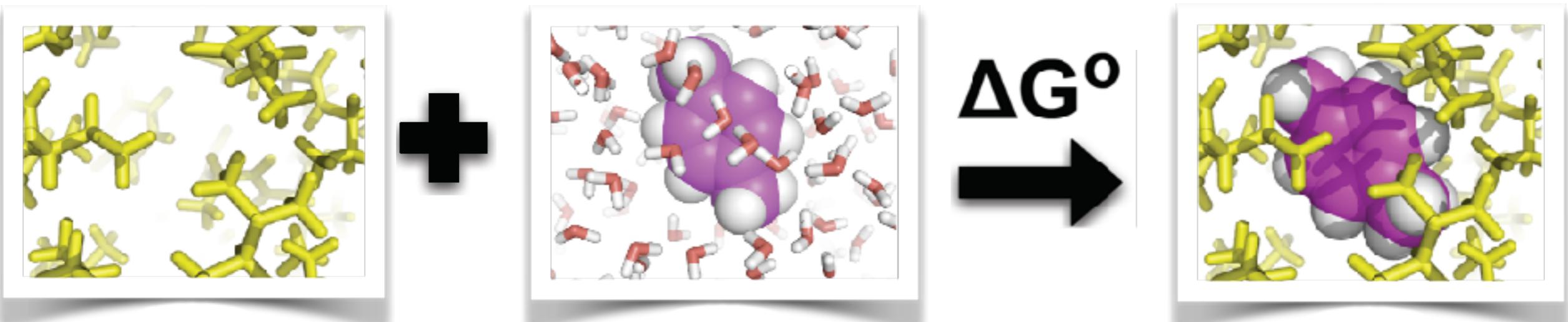
- Simple
- Polar
- Dry
- Additional stable binding modes



Cytochr. C Peroxidase

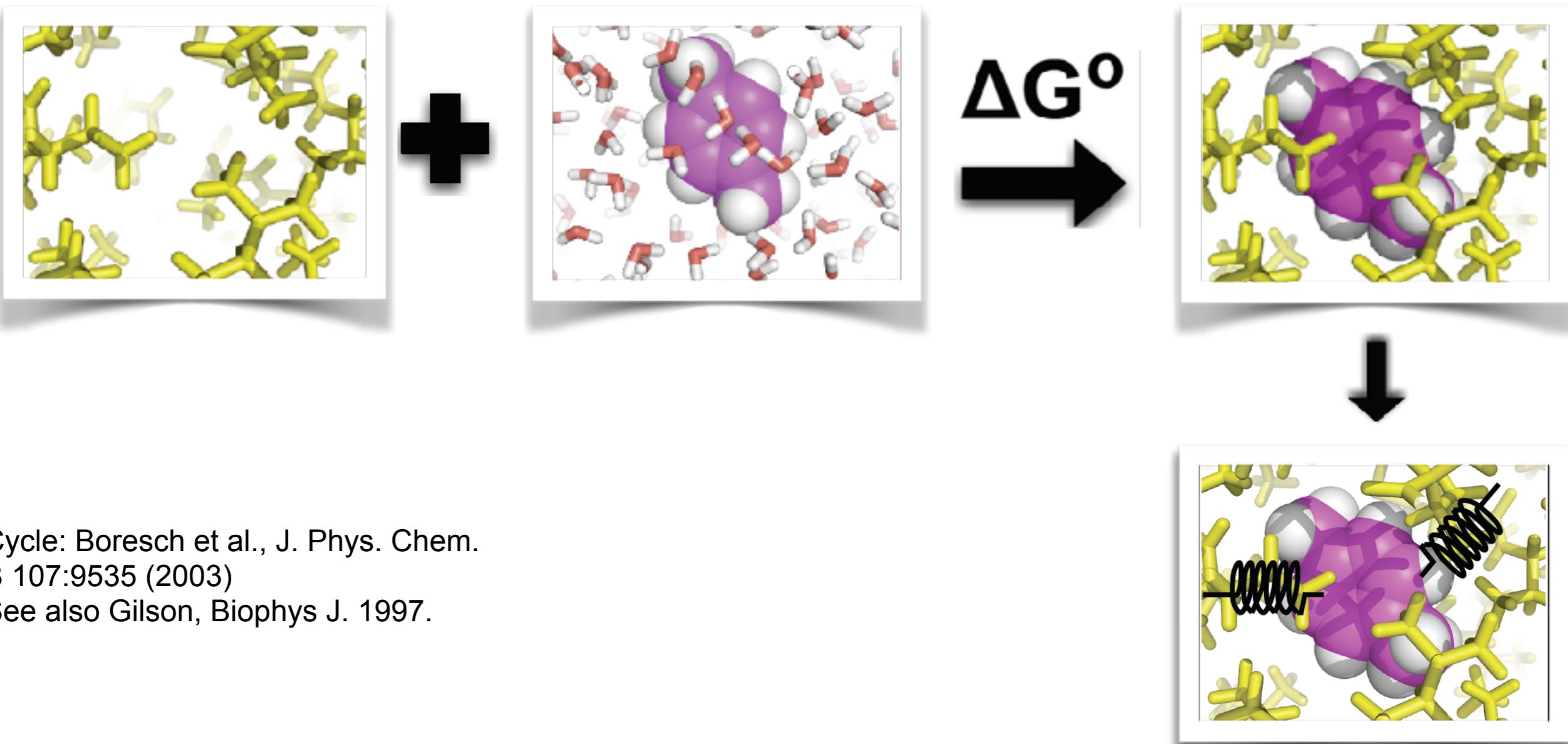
- Simple (?)
- Polar, Charged
- Wet
- Additional stable binding modes
- Force field issues?

# Thermodynamic cycle



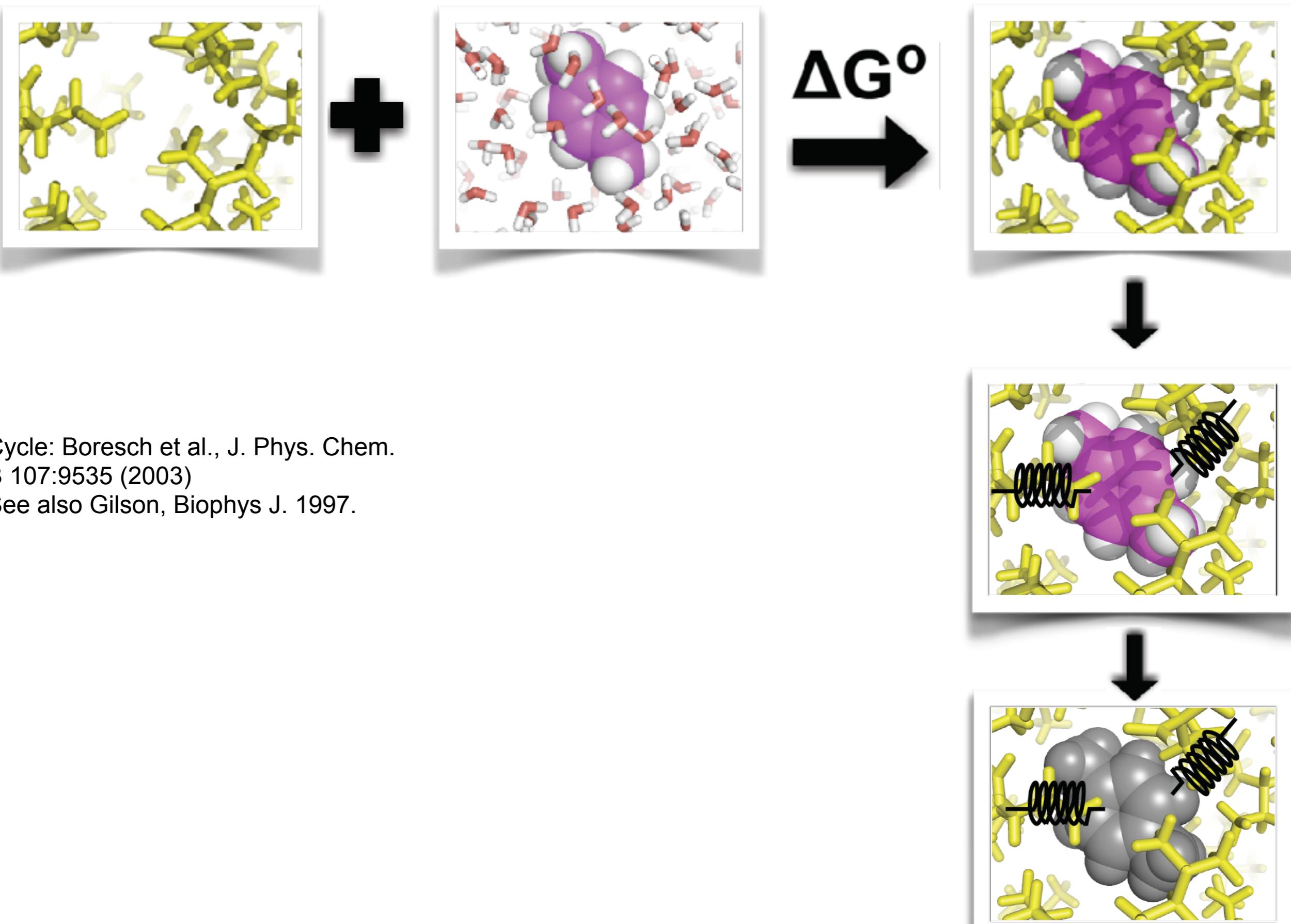
Cycle: Boresch et al., J. Phys. Chem.  
B 107:9535 (2003)  
See also Gilson, Biophys J. 1997.

# Thermodynamic cycle

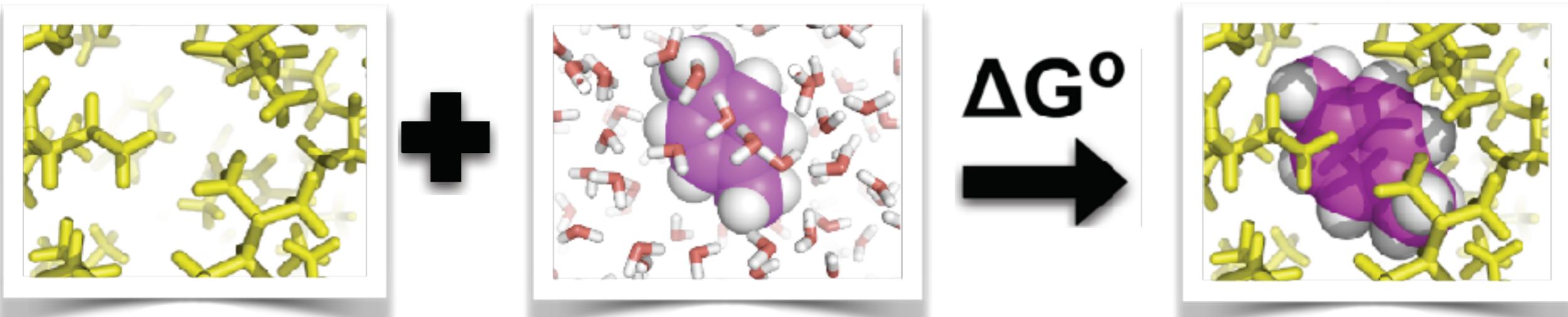


Cycle: Boresch et al., J. Phys. Chem.  
B 107:9535 (2003)  
See also Gilson, Biophys J. 1997.

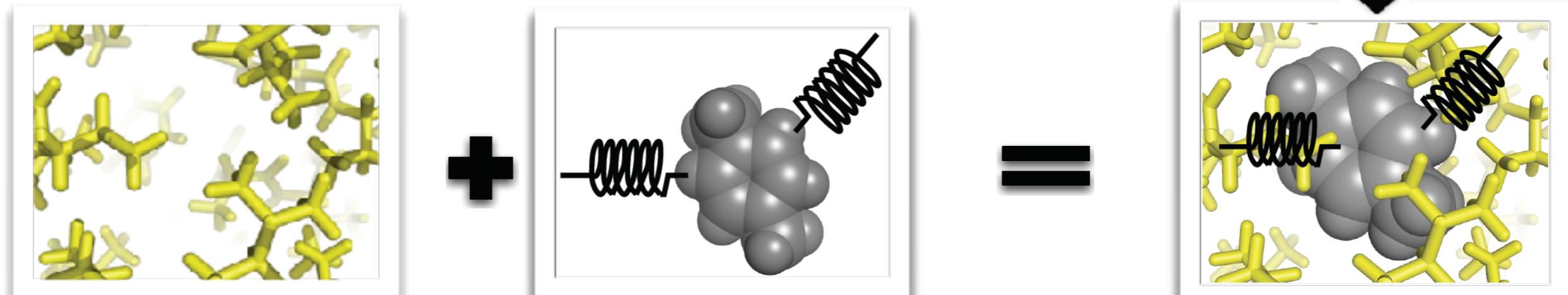
# Thermodynamic cycle



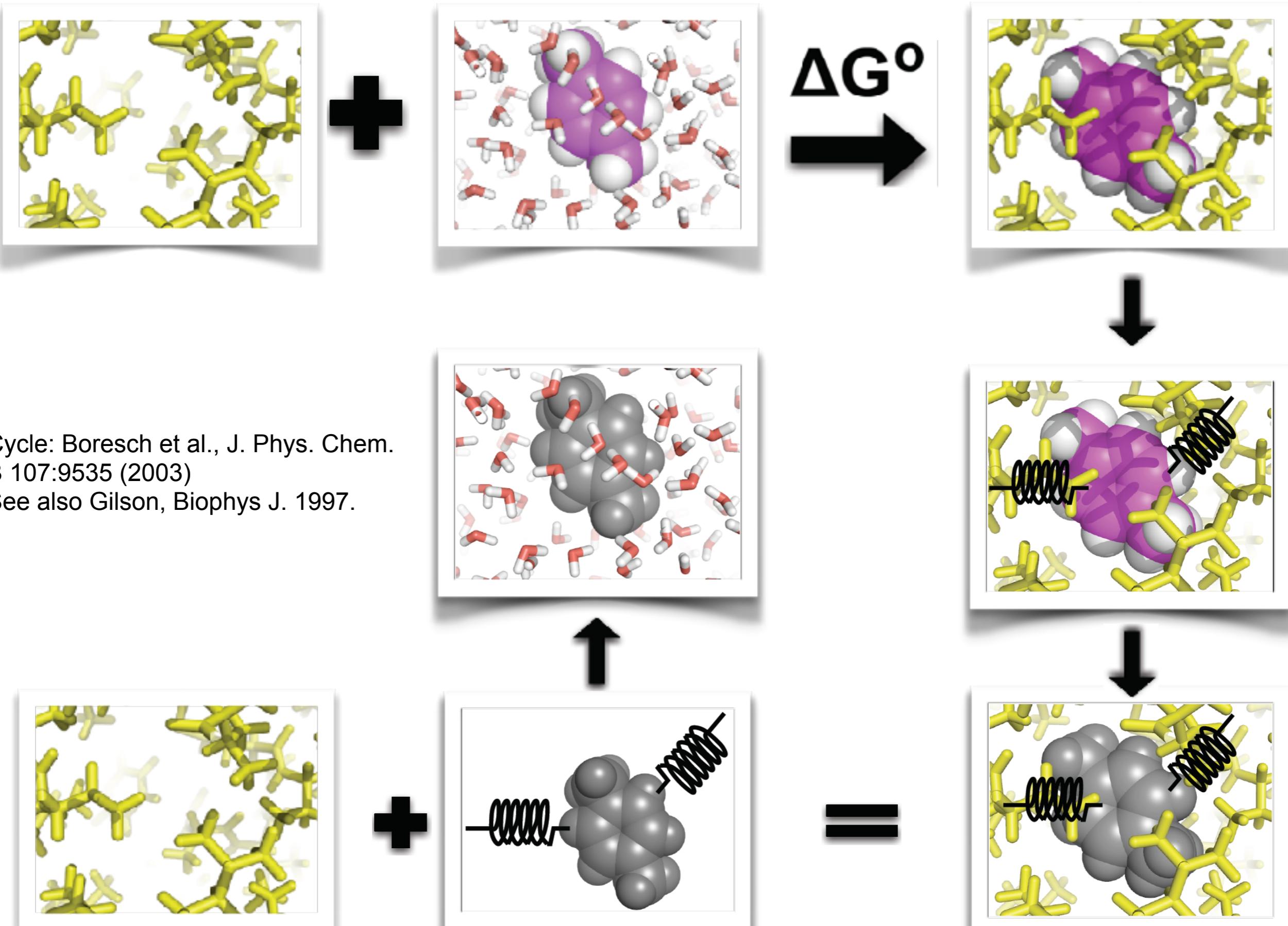
# Thermodynamic cycle



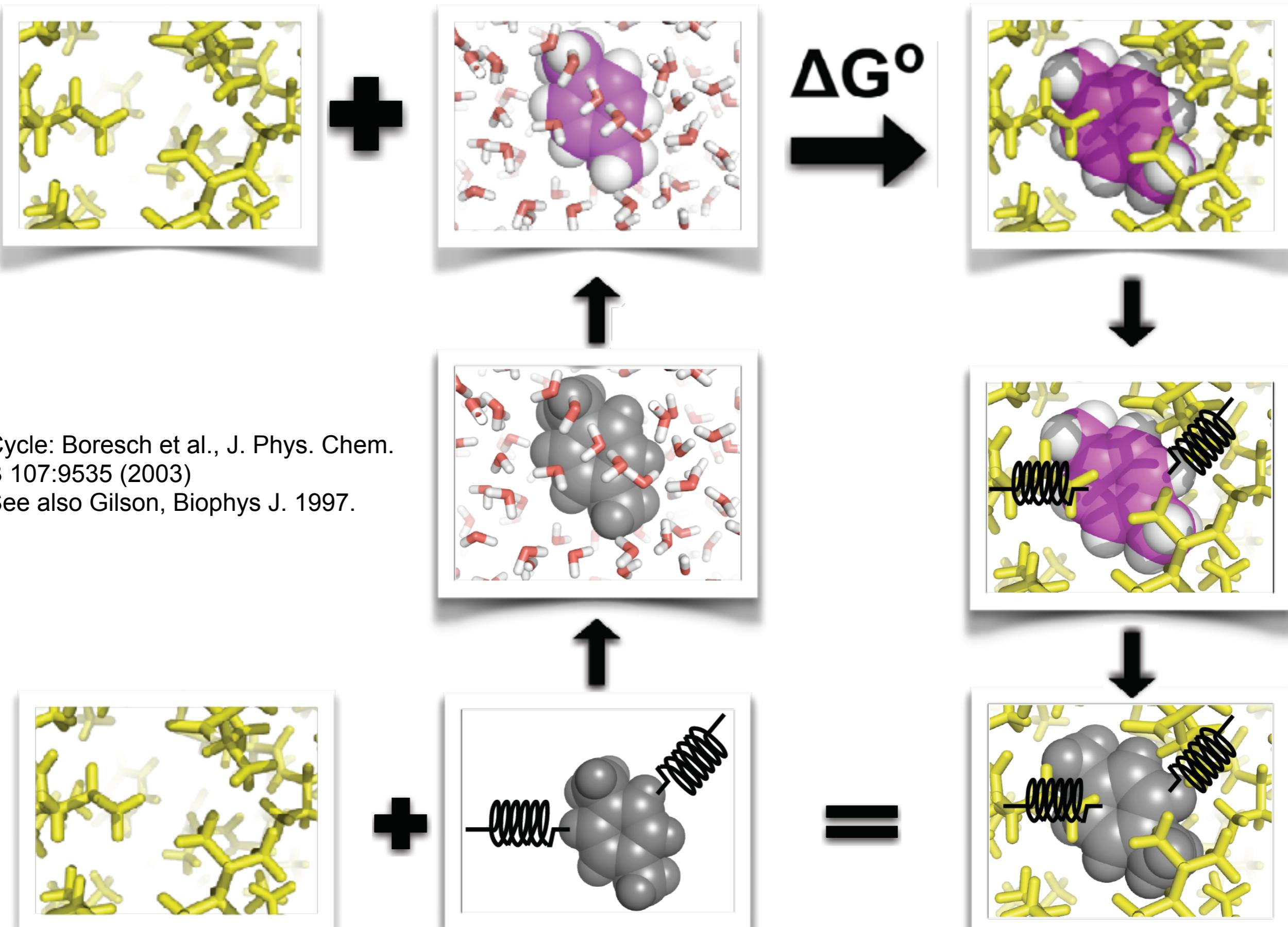
Cycle: Boresch et al., J. Phys. Chem.  
B 107:9535 (2003)  
See also Gilson, Biophys J. 1997.



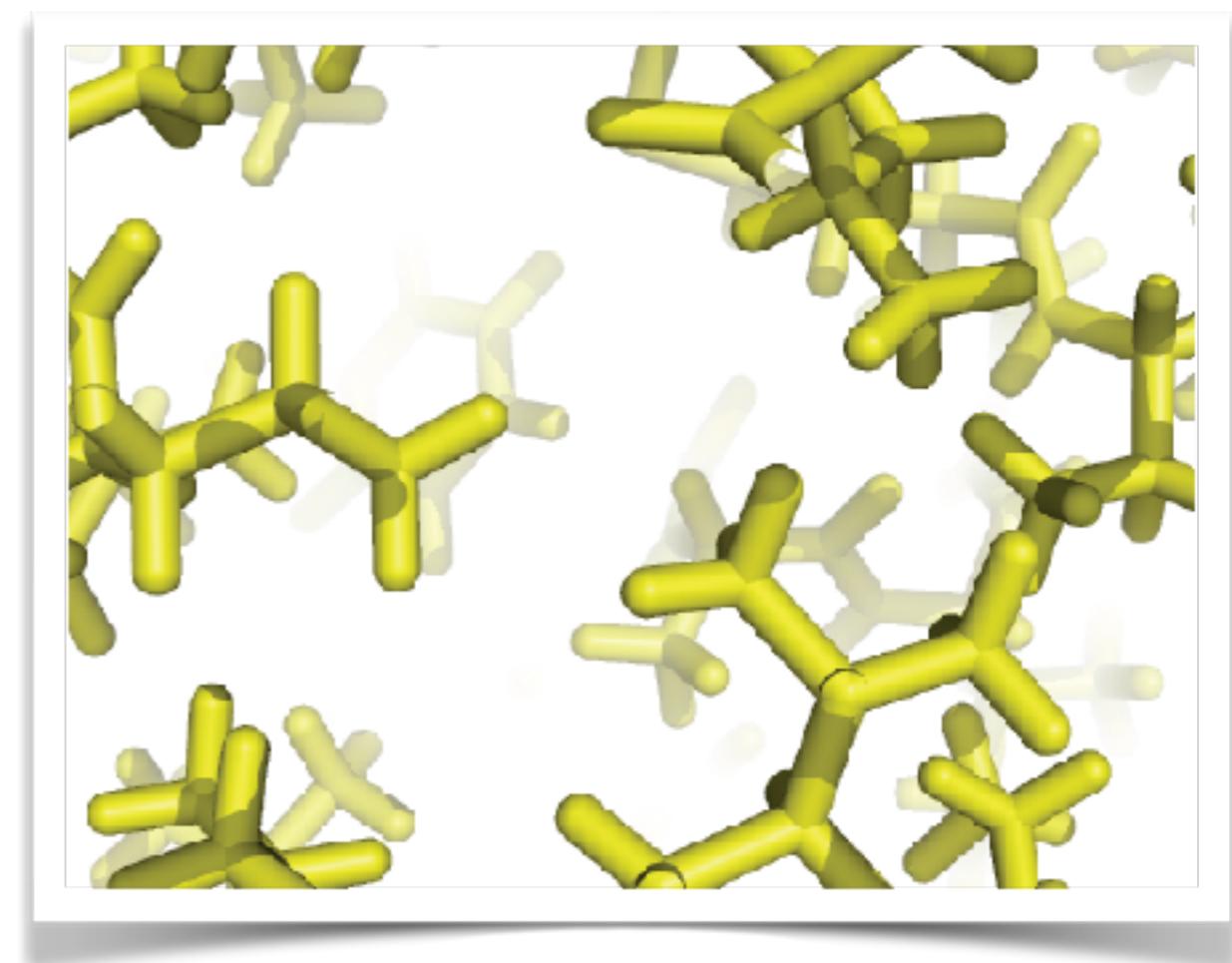
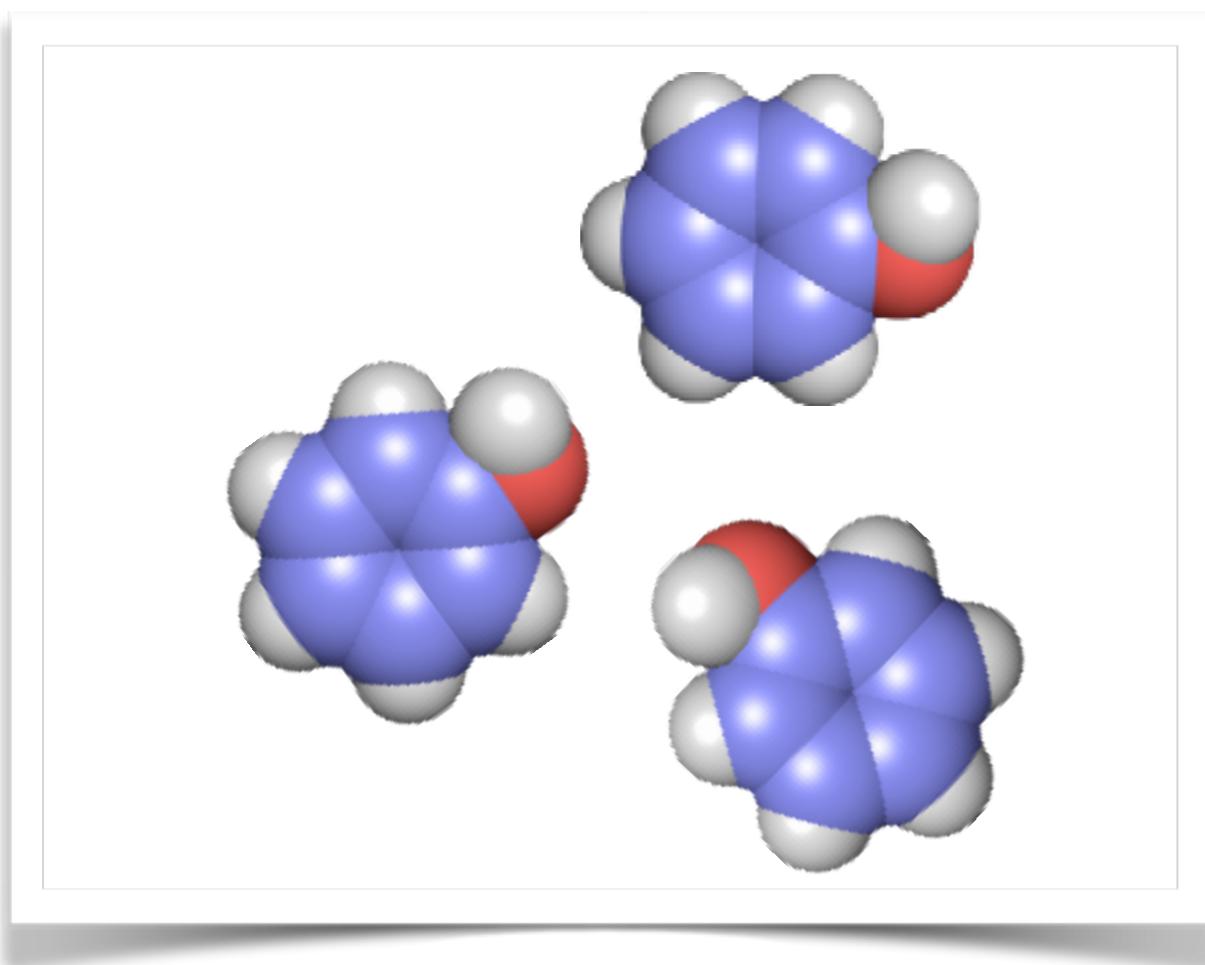
# Thermodynamic cycle



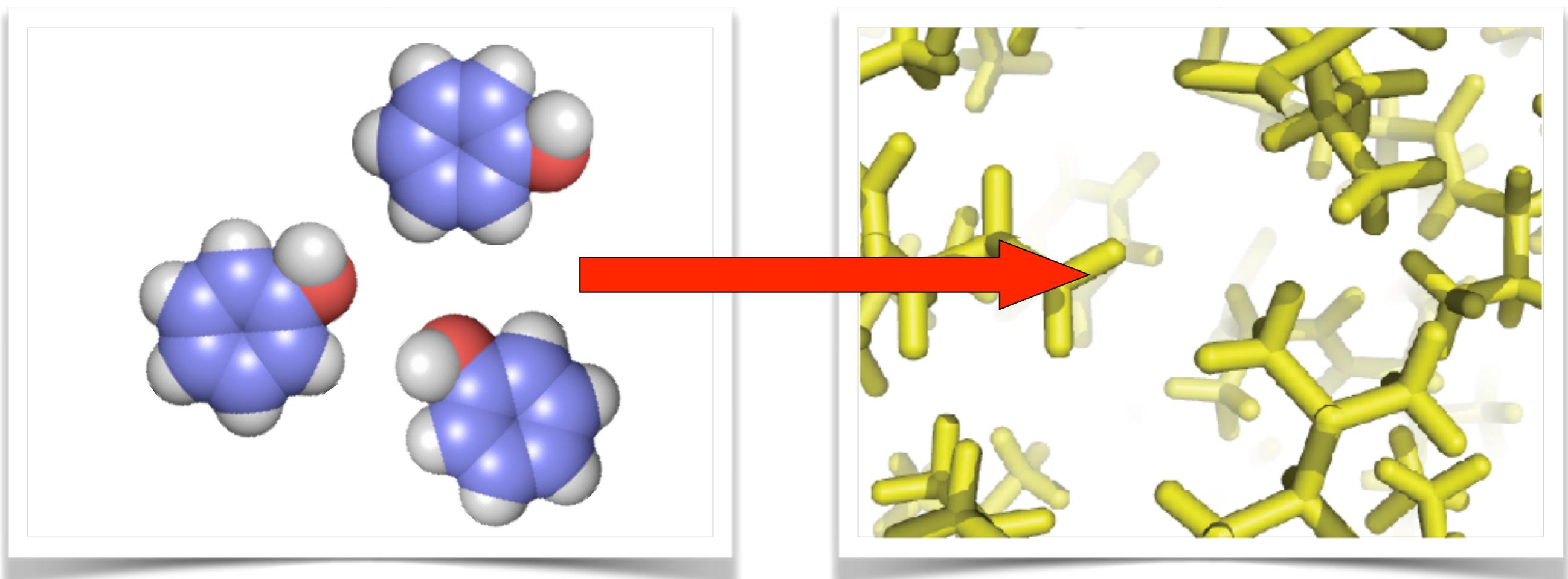
# Thermodynamic cycle



Our approach requires no knowledge of bound structure



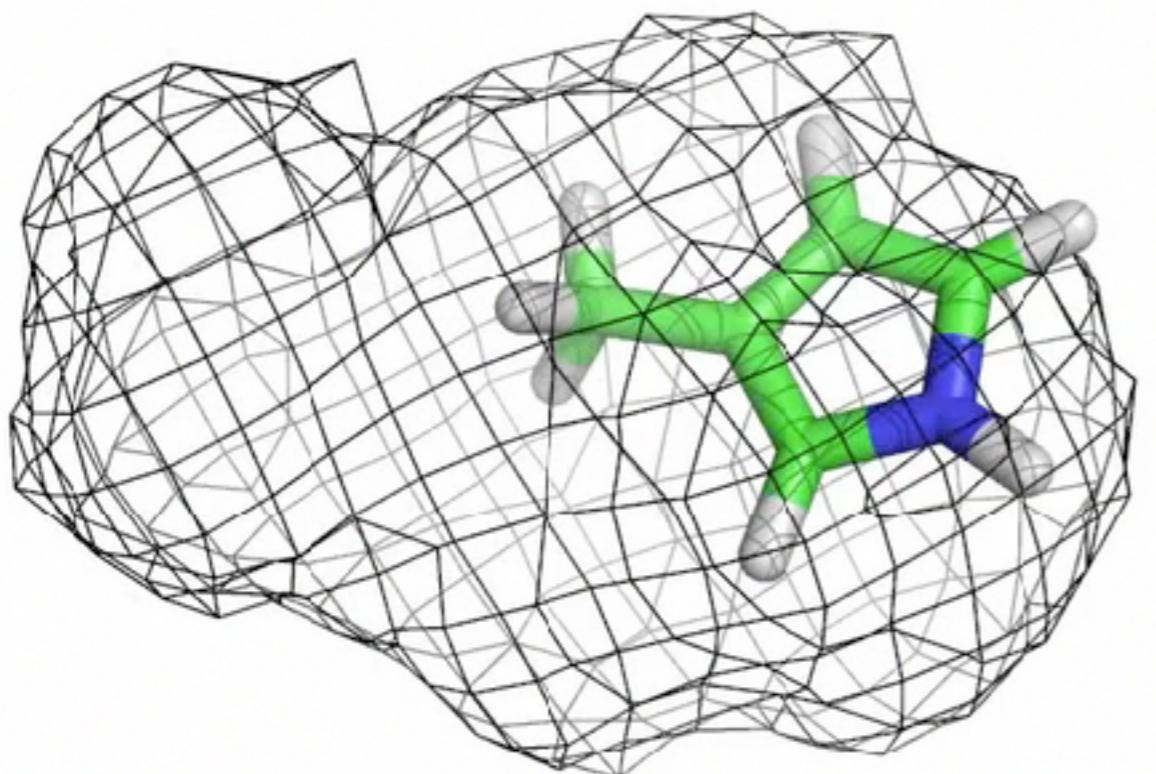
# Our approach requires no knowledge of bound structure



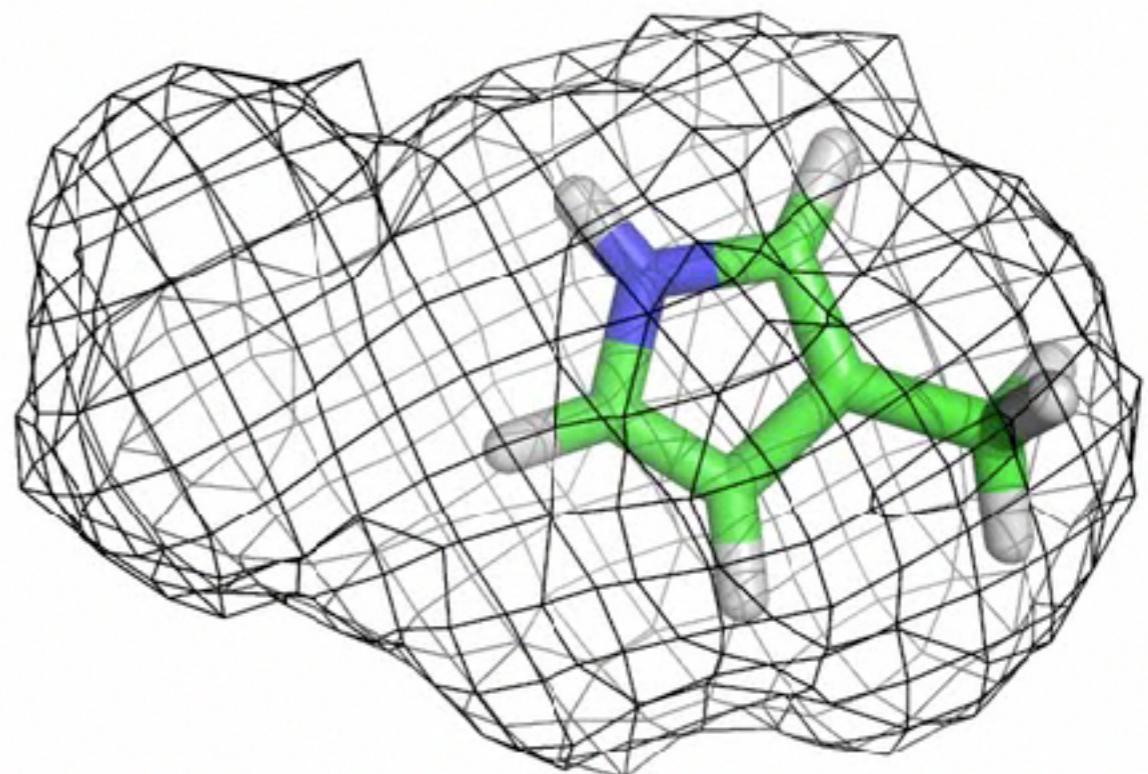
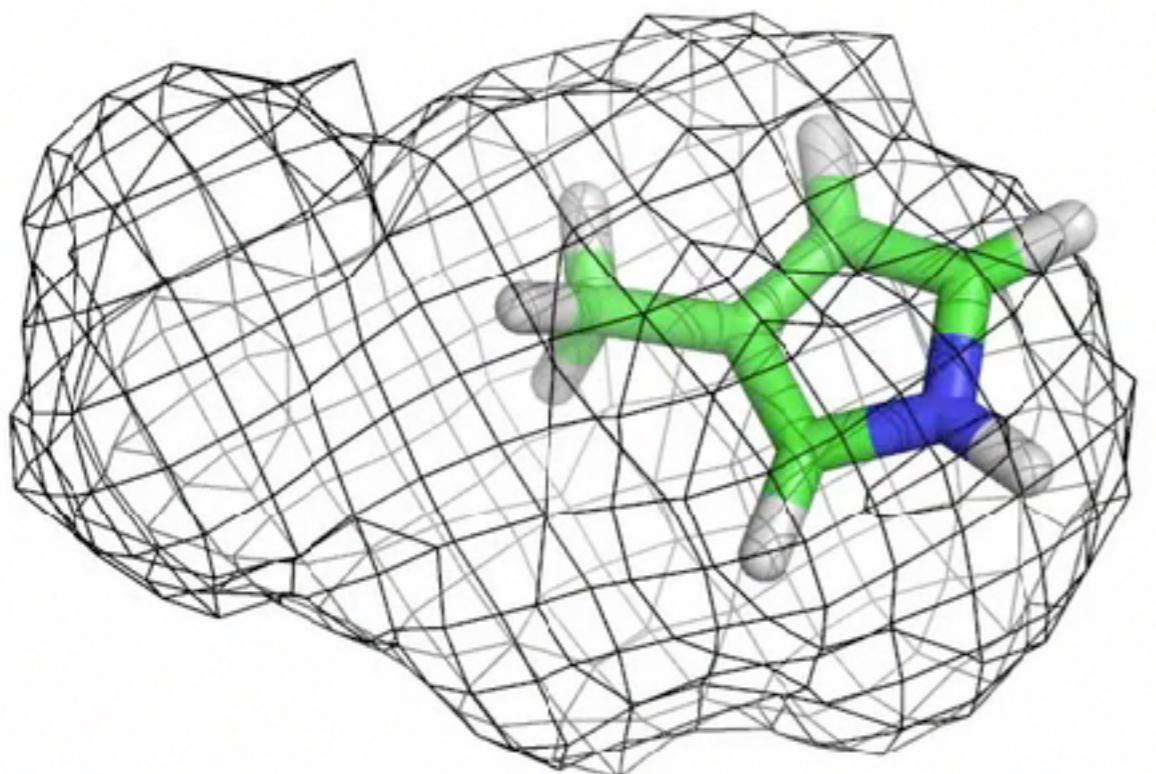
- Generate starting conformations using docking

Separate calculations needed  
for different orientations

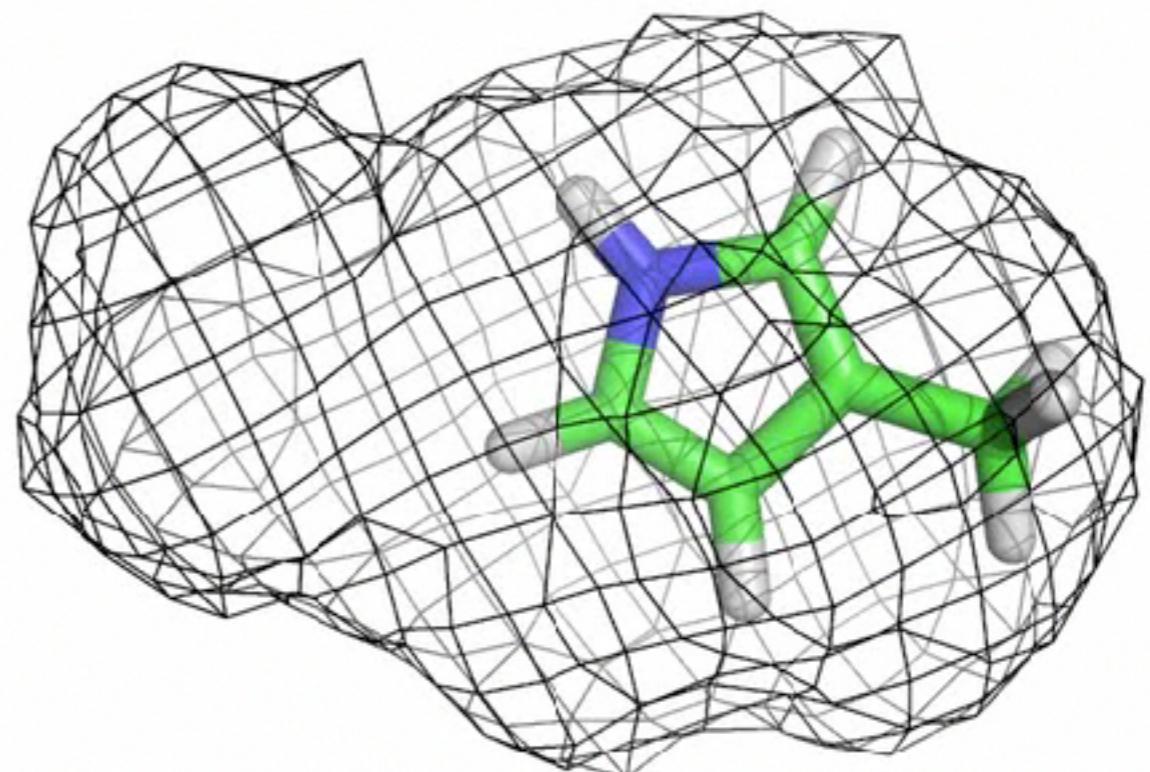
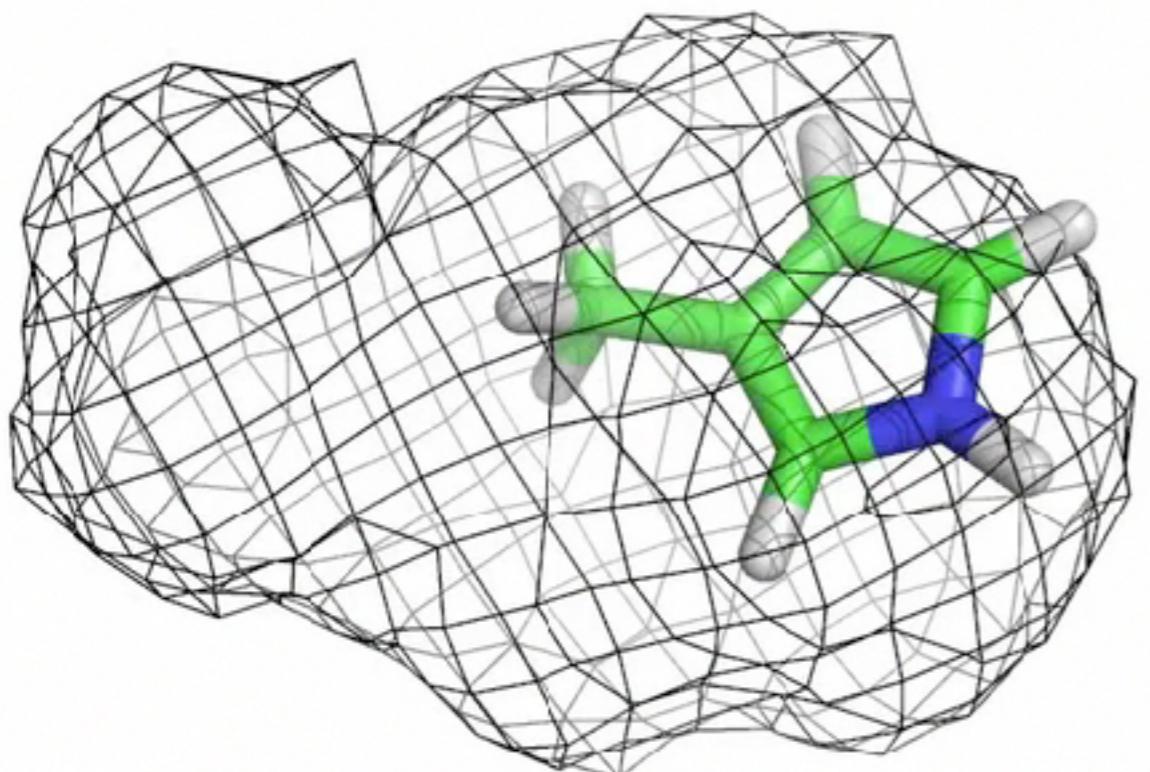
# Separate calculations needed for different orientations



# Separate calculations needed for different orientations

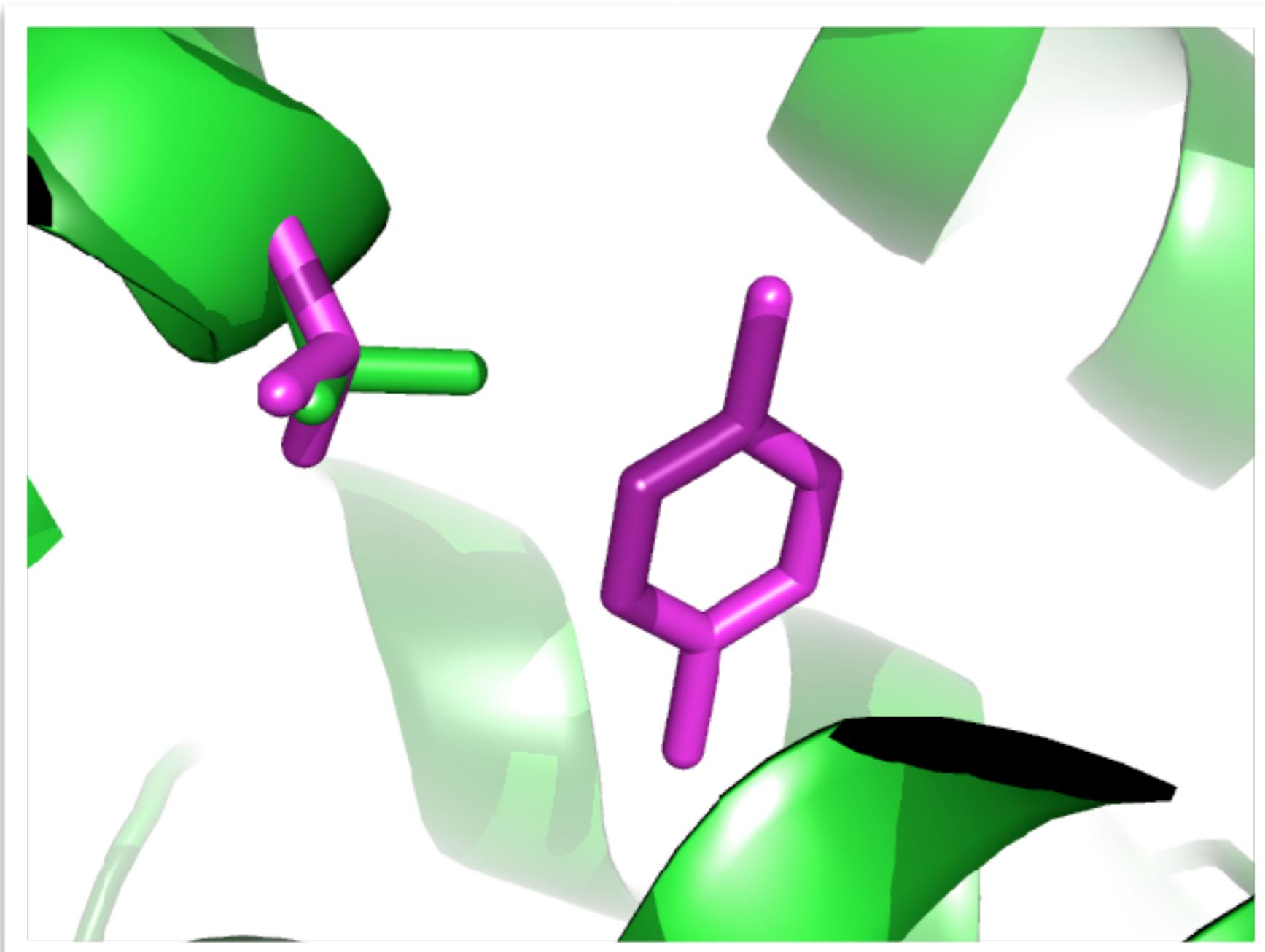


# Separate calculations needed for different orientations



$$\Delta G^o = -k_B T \ln \left( e^{-\frac{\Delta G_1^o}{k_B T}} + e^{-\frac{\Delta G_2^o}{k_B T}} \right)$$

Problem: Some protein conformational changes present sampling problems, bias results

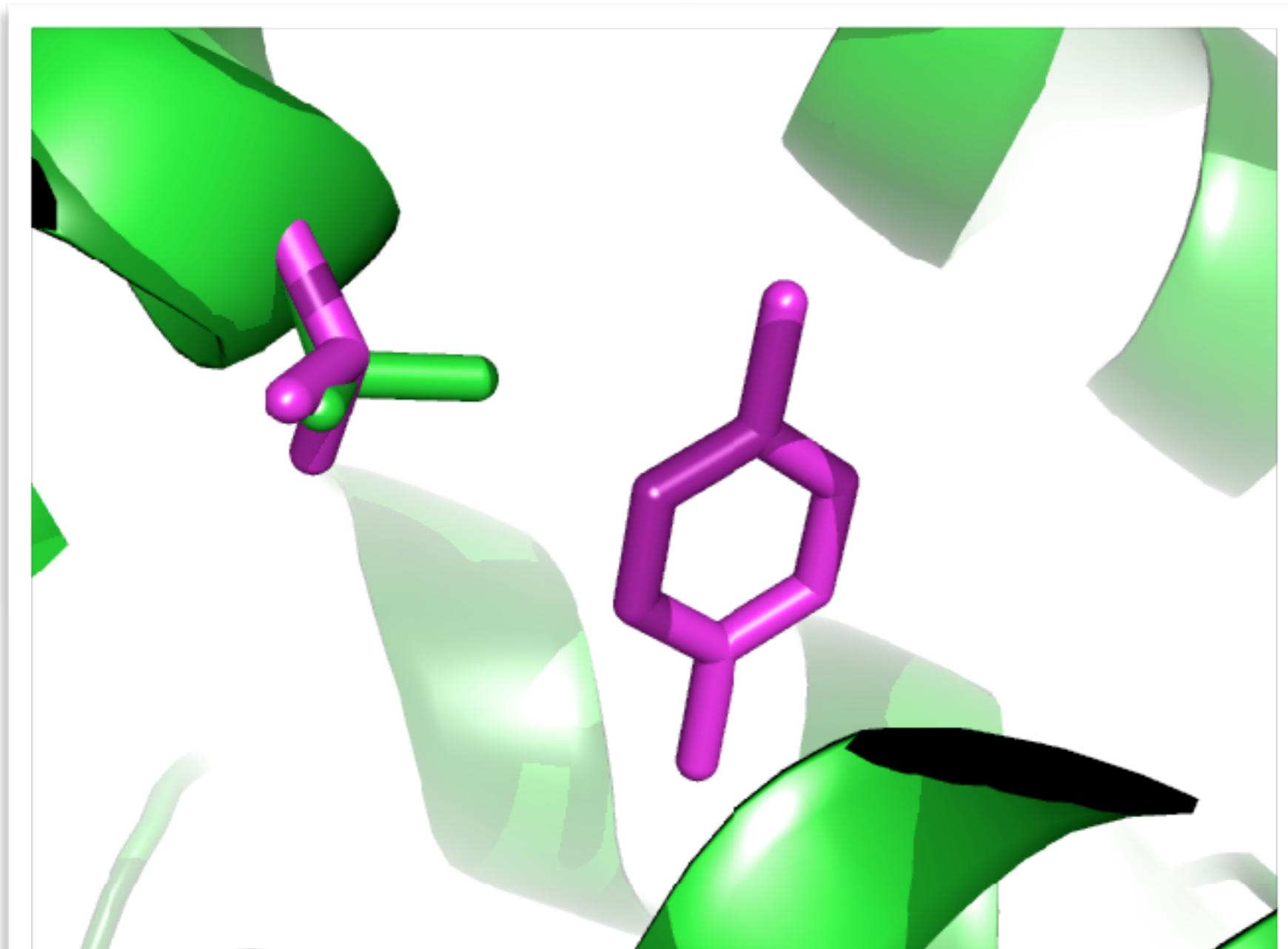


From **apo** structure:  $\Delta G = -3.0 \pm 0.1 \text{ kcal/mol}$

From **holo** structure:  $\Delta G = -7.3 \pm 0.1 \text{ kcal/mol}$

Experiment:  $-4.6 \text{ kcal/mol}$

# It's necessary to include conformational change

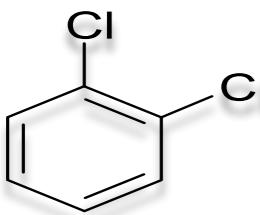
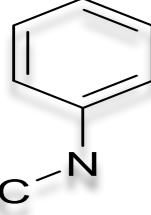
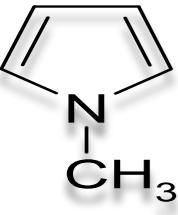
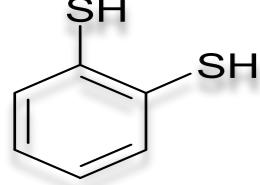
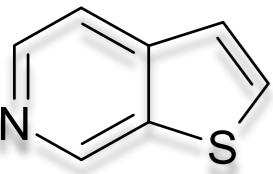


From **apo** structure:  $\Delta G = -3.5 \pm 0.2 \text{ kcal/mol}$

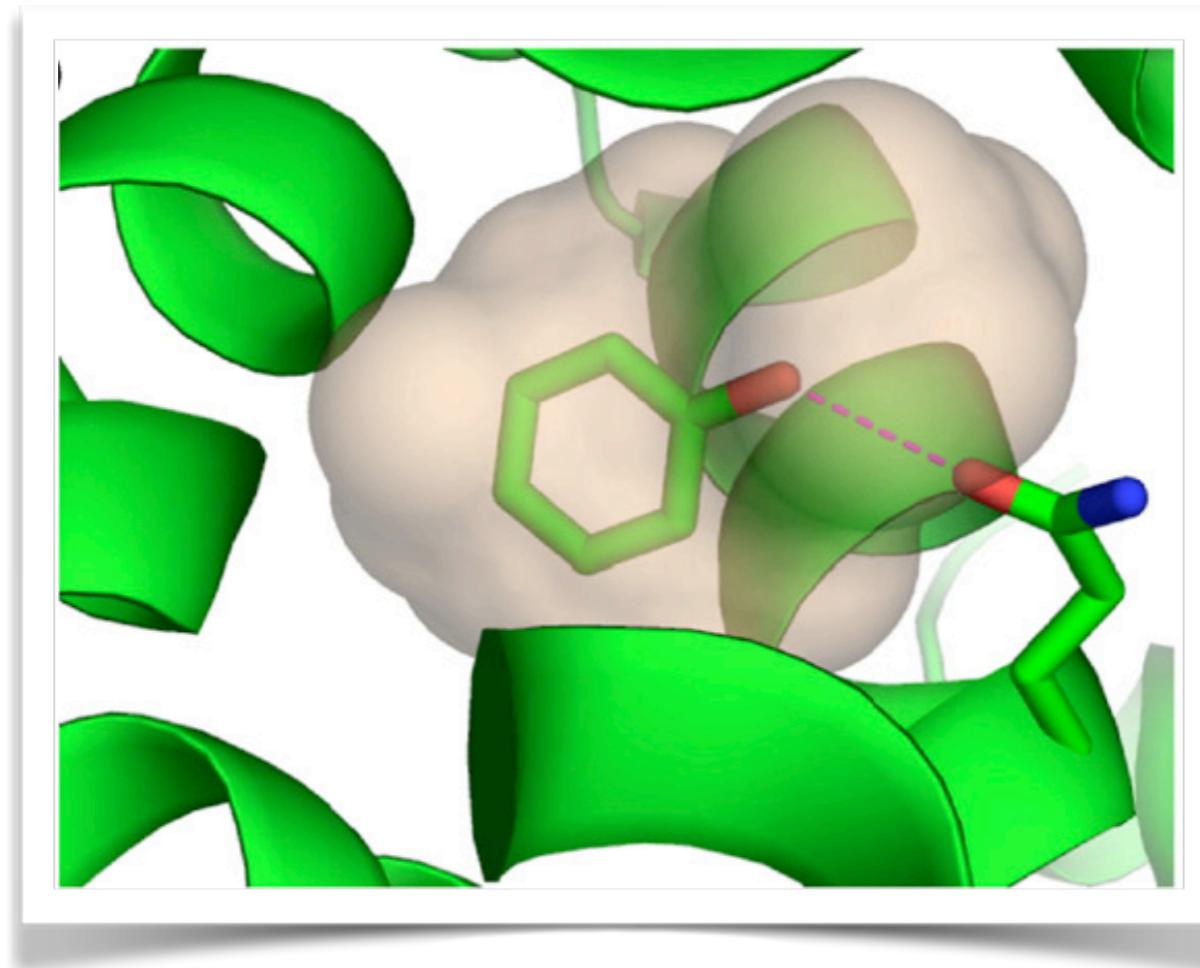
From **holo** structure:  $\Delta G = -3.4 \pm 0.2 \text{ kcal/mol}$

Experiment:  $-4.6 \text{ kcal/mol}$

# Putting it all together: A blind test

	Name	Predicted $\Delta G$ (kcal/mol)	Expt. $\Delta G$ (kcal/mol)
	1,2-dichlorobenzene	-5.6	-6.4
	N-methylaniline	-5.4	-4.7
	1-methylpyrrole	-4.3	-4.4
	1,2-benzenedithiol	-2.8	< -2.7
	thieno[2,3c]-pyridine	-2.6	> -3.6

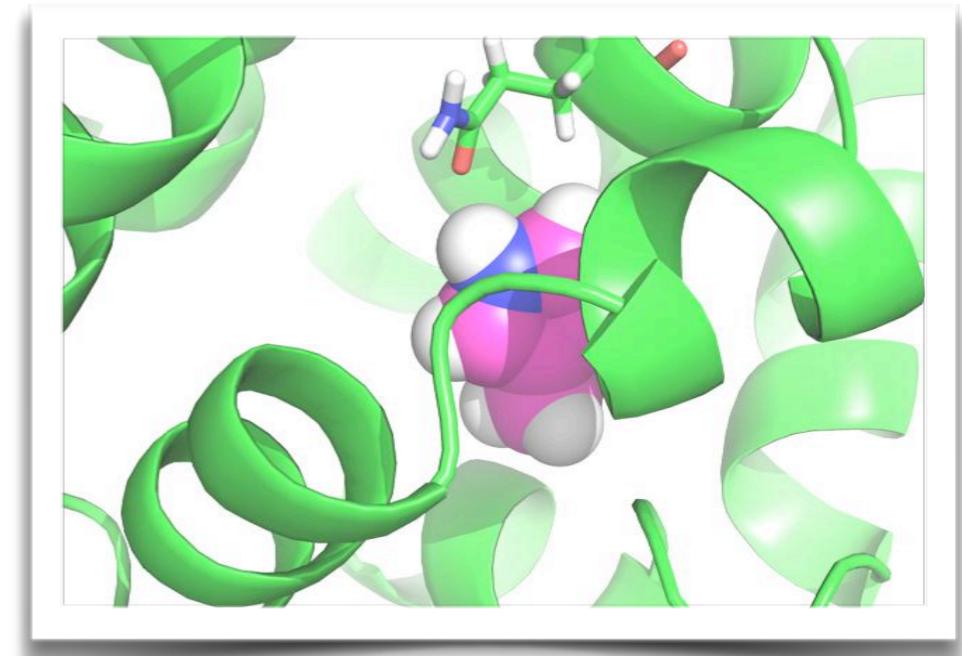
# Polar cavity adds hydrogen bonding



**Lysozyme  
L99A/M102Q**

Here, many affinities are accurately predicted

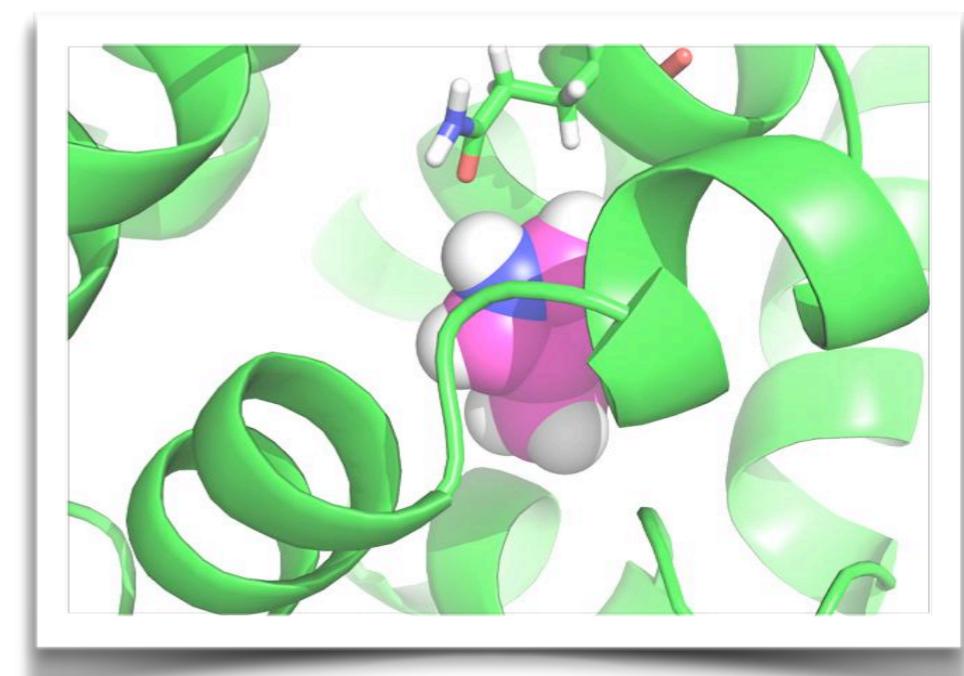
Here, many affinities are accurately predicted



# Here, many affinities are accurately predicted

## Well predicted binding affinities

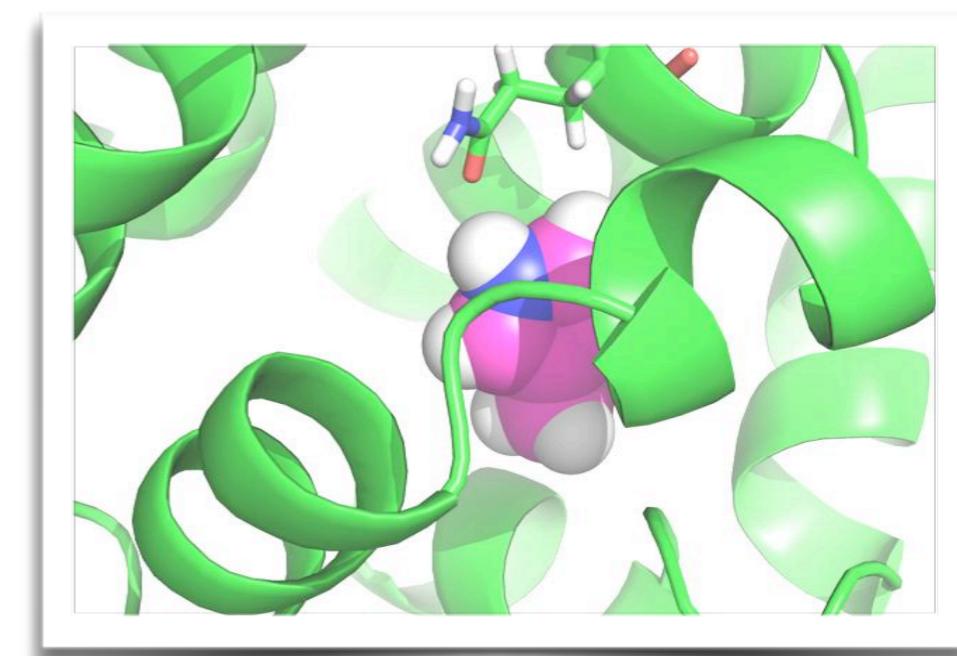
Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error
<chem>c1ccccc1NCC#N</chem>	-5.8	-5.5	-0.3
<chem>c1cc2sc(C=C2)cc1</chem>	-5.4	-4.7	-0.7
<chem>c1ccsc1[N+](=O)[O-]</chem>	-5.3	-4.8	0.5
<chem>Fc1cc(F)c(N)cc1</chem>	-4.6	Binder	~ 1?
<chem>COC1CCCCC1OCC</chem>	-4.5	Weak binder	~ 1-1.5?
<chem>Cc1ccccc1O</chem>	-4.4	-5.2	0.8
<chem>Cc1cc(F)cc(N)cc1</chem>	-3.9	-4.7	0.8
<chem>Cc1cc(O)cc(O)cc1</chem>	-3.1	-4.2	1.0



# Here, many affinities are accurately predicted

## Well predicted binding affinities

Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error
	-5.8	-5.5	-0.3
	-5.4	-4.7	-0.7
	-5.3	-4.8	0.5
	-4.6	Binder	~ 1?
	-4.5	Weak binder	~ 1-1.5?
	-4.4	-5.2	0.8
	-3.9	-4.7	0.8
	-3.1	-4.2	1.0

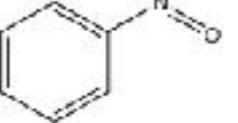


Correctly predicted  
non-binders

Molecule	Calc. $\Delta G$
	-2.5
	-1.9
	-1.9
	+0.2

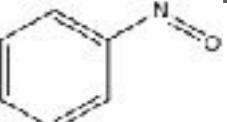
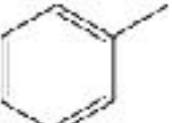
Yet we see many failures, and understand some

# Yet we see many failures, and understand some Near misses

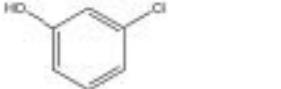
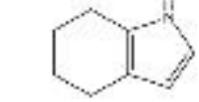
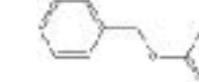
Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

# Yet we see many failures, and understand some

## Near misses

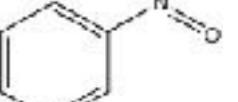
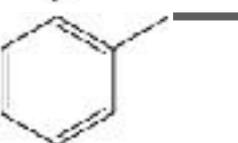
Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

We know what went wrong

	-4.1	-5.5	1.4
	-3.3	-4.9	1.6
	-3.0	-4.6	1.6
	-1.4	-4.5	3.0

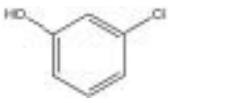
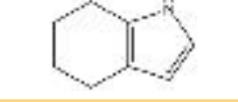
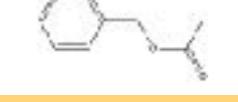
# Yet we see many failures, and understand some

## Near misses

Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

Wrongly trapped water makes binding appear weaker by clashing with ligand

We know what went wrong

	-4.1	-5.5	1.4
	-3.3	-4.9	1.6
	-3.0	-4.6	1.6
	-1.4	-4.5	3.0

# Yet we see many failures, and understand some

## Near misses

Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

Wrongly trapped water makes binding appear weaker by clashing with ligand

We know what went wrong

	-4.1	-5.5	1.4
	-3.3	-4.9	1.6
	-3.0	-4.6	1.6
	-1.4	-4.5	3.0

Initial simulations of ligand did not find dominant binding mode; we missed it

# Yet we see many failures, and understand some

## Near misses

Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

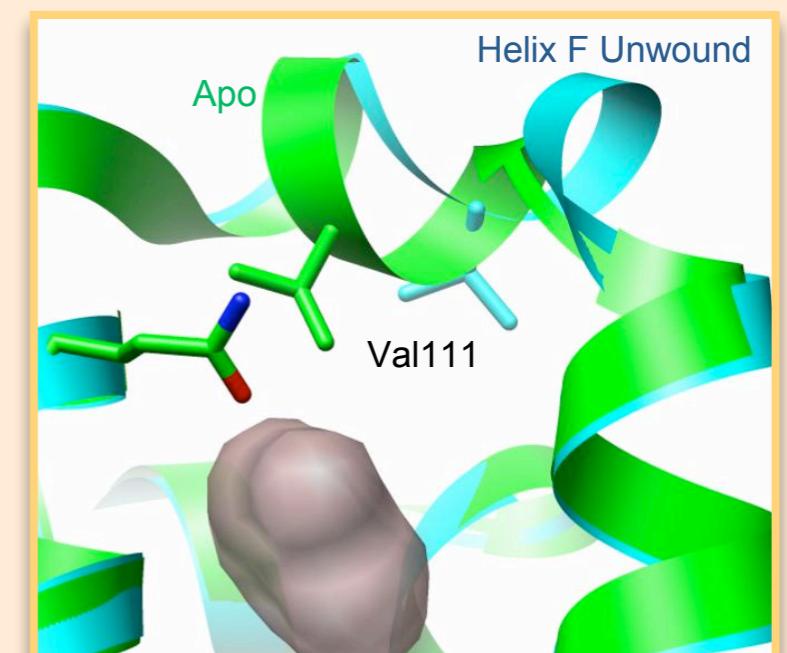
Wrongly trapped water makes binding appear weaker by clashing with ligand

We know what went wrong

	-4.1	-5.5	1.4
	-3.3	-4.9	1.6
	-3.0	-4.6	1.6
	-1.4	-4.5	3.0

Initial simulations of ligand did not find dominant binding mode; we missed it

Slow conformational changes present problems



# Yet we see many failures, and understand some

## Near misses

Molecule	Calc. $\Delta G$	Expt. $\Delta G$	Error (kcal/mol)
	-5.4	Weak binder	~-2?
	-3.5	-4.9	1.4

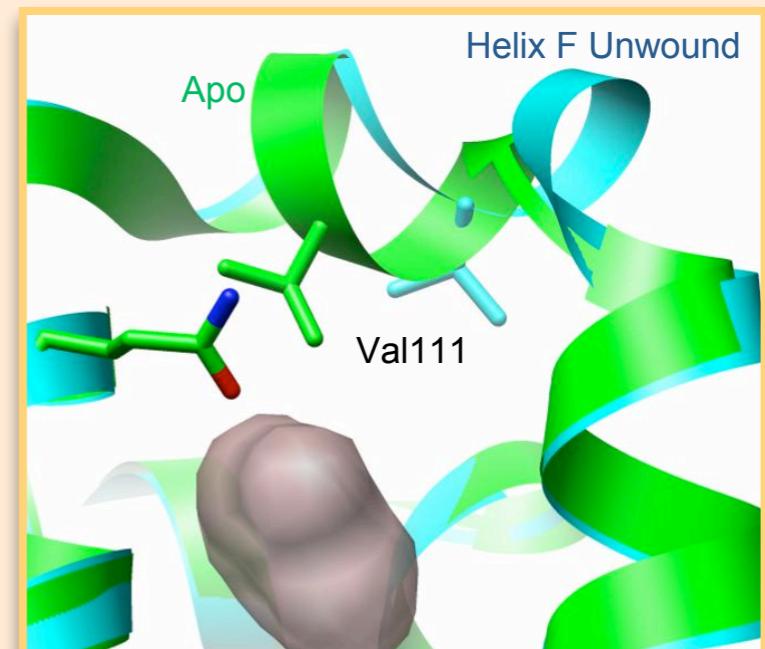
Wrongly trapped water makes binding appear weaker by clashing with ligand

## We know what went wrong

	-4.1	-5.5	1.4
	-3.3	-4.9	1.6
	-3.0	-4.6	1.6
	-1.4	-4.5	3.0

Initial simulations of ligand did not find dominant binding mode; we missed it

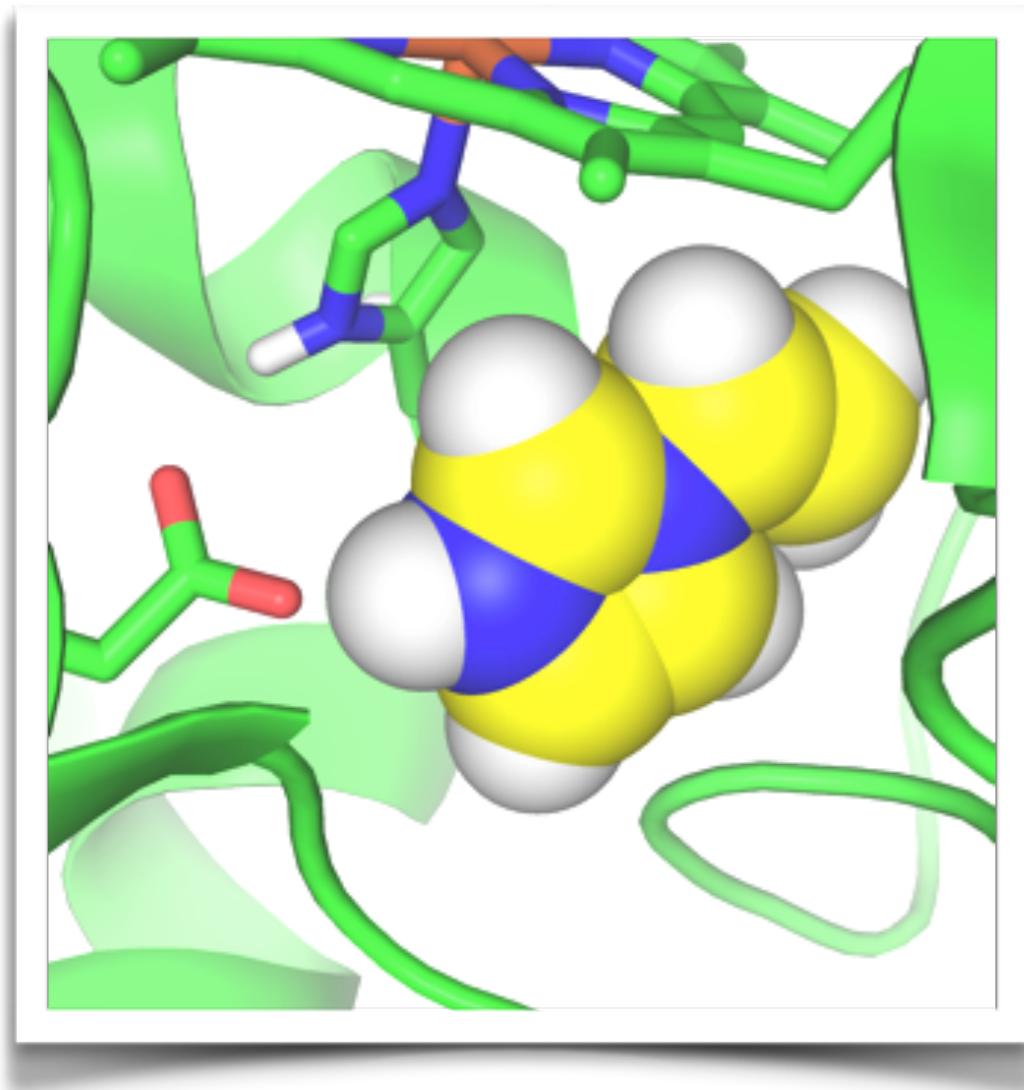
Slow conformational changes present problems



## Mysteries

	-5.7	Nonbinder	Expt.?
	-3.9	Nonbinder	Oxidizes?
	-3.4	Nonbinder	?

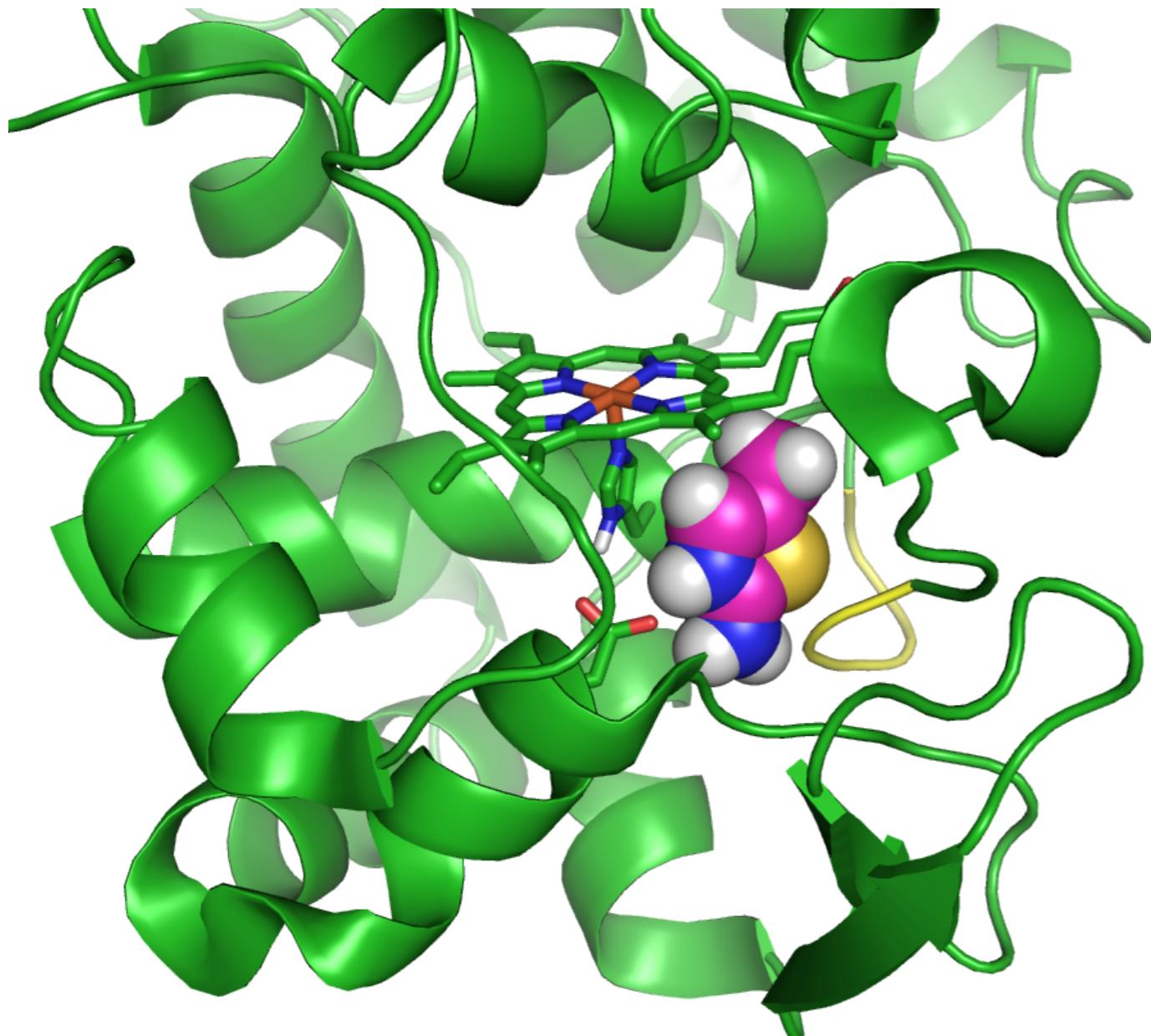
# Cytochrome C Peroxidase model sites are another step towards biological binding sites



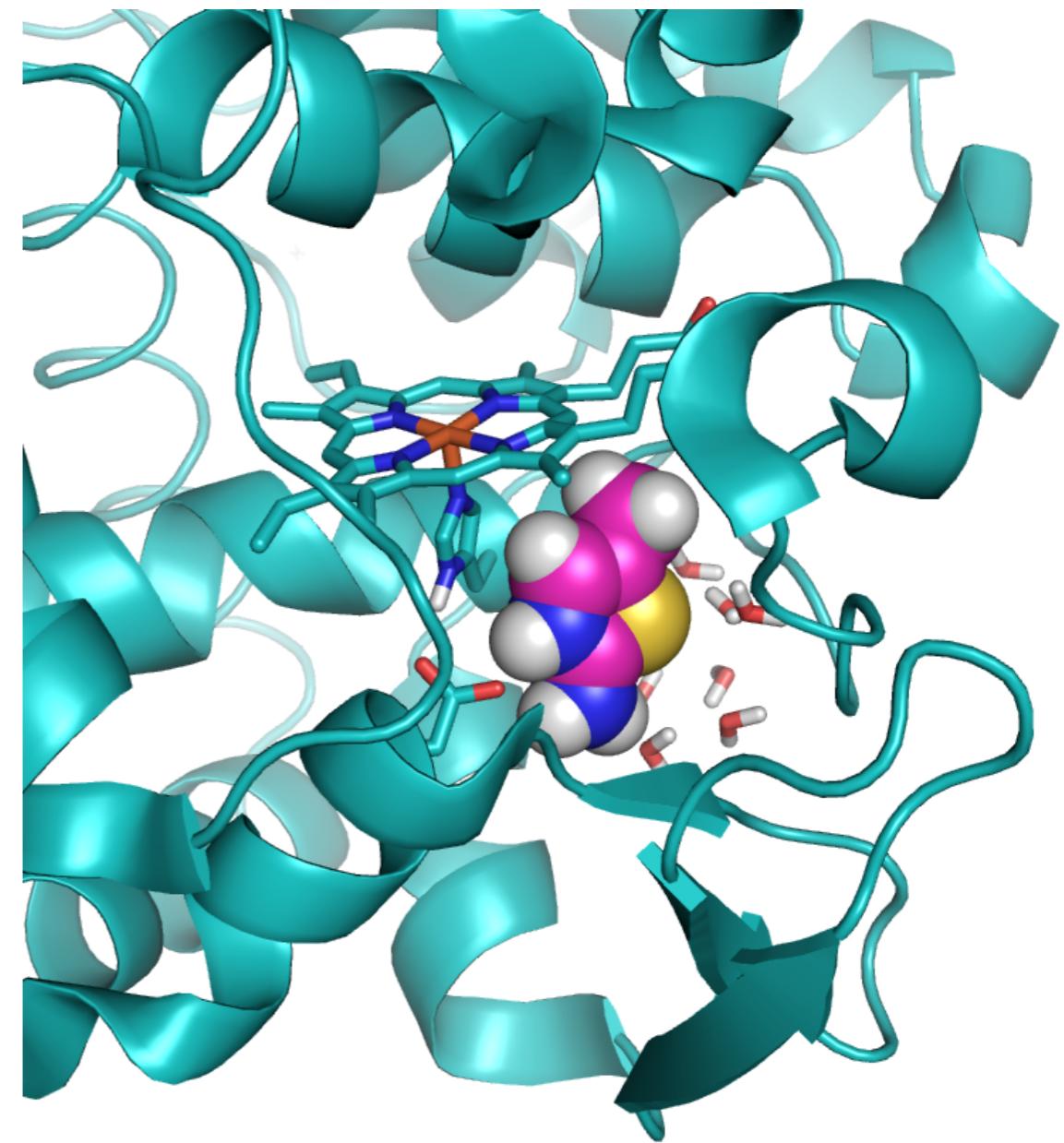
- Binds +1 ligands
- Some neutrals
- Affinities -7.1 to -2.9 kcal/mol
- 1-2 ordered waters with ligands
- Empty site: ordered waters

We consider two versions of the binding site

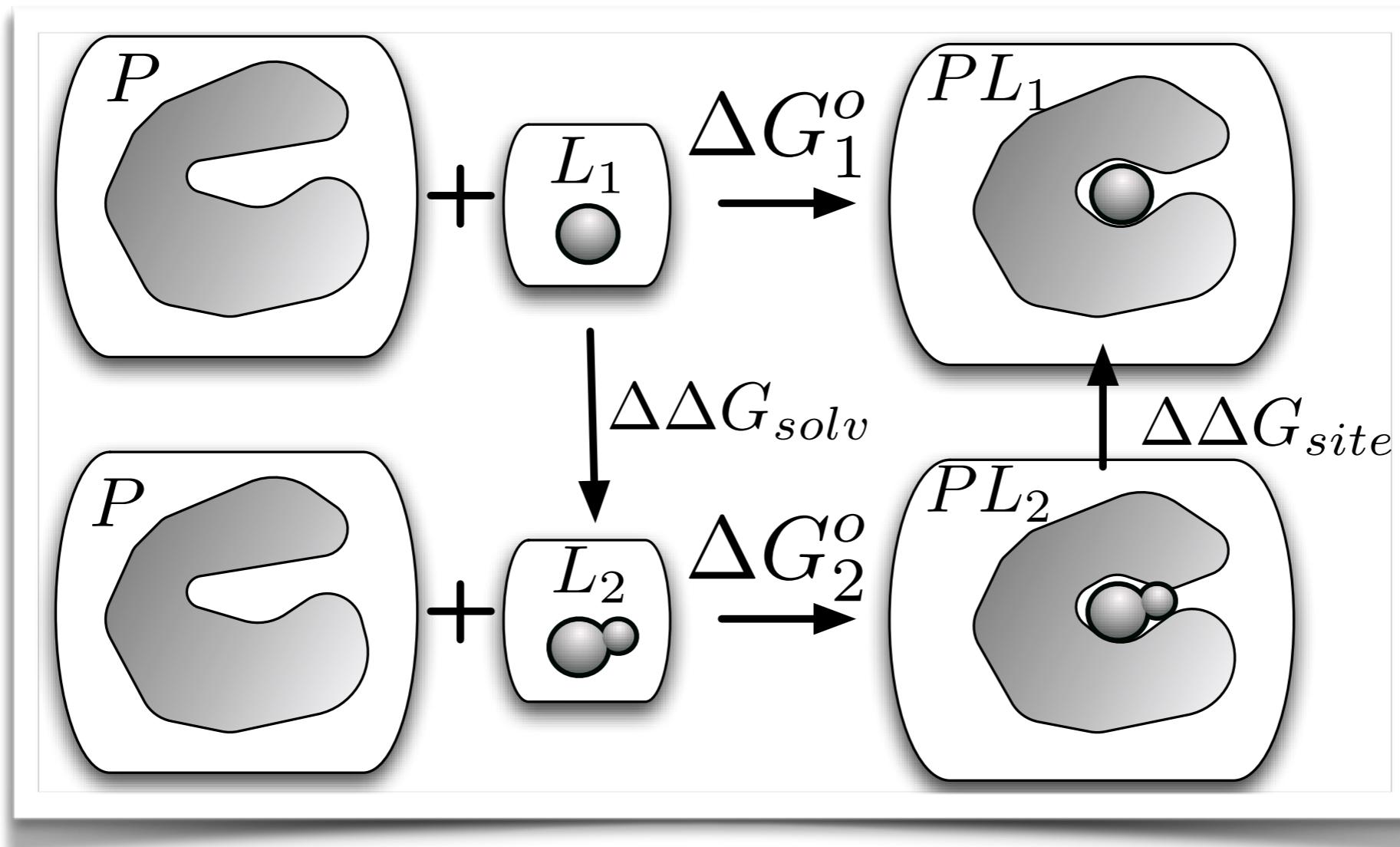
CCP W191G “WT”



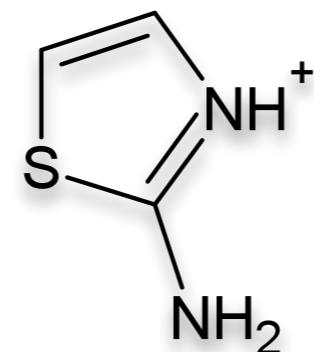
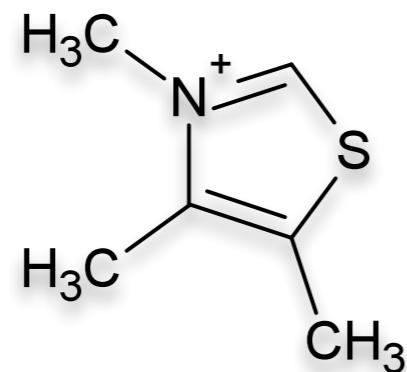
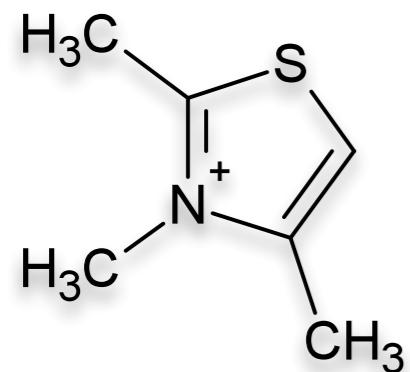
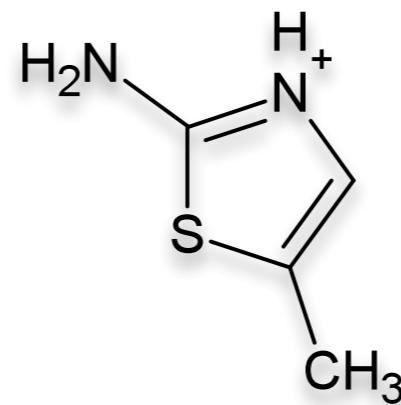
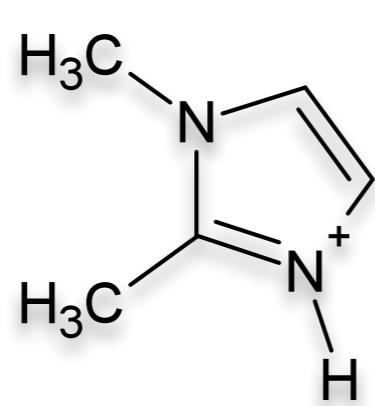
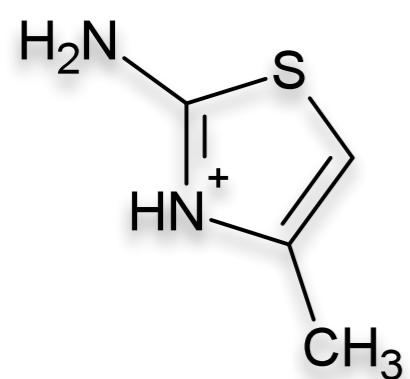
CCP W191G Gateless



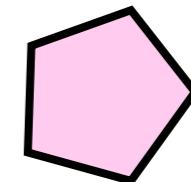
Here, we do both absolute  
and relative free energies



We actually compute relative free energies relative to a scaffold, which simplifies the calculations

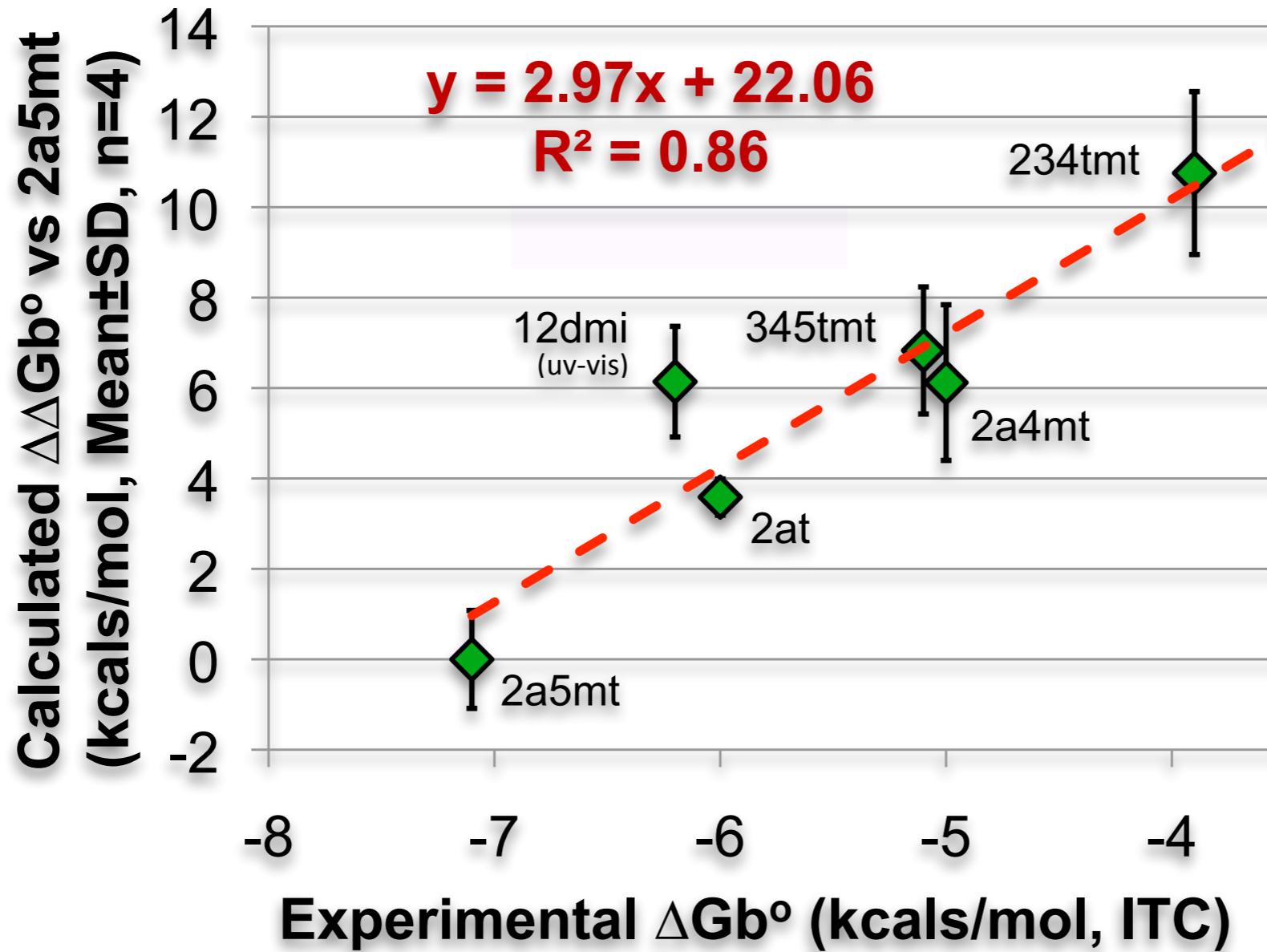


Ligands

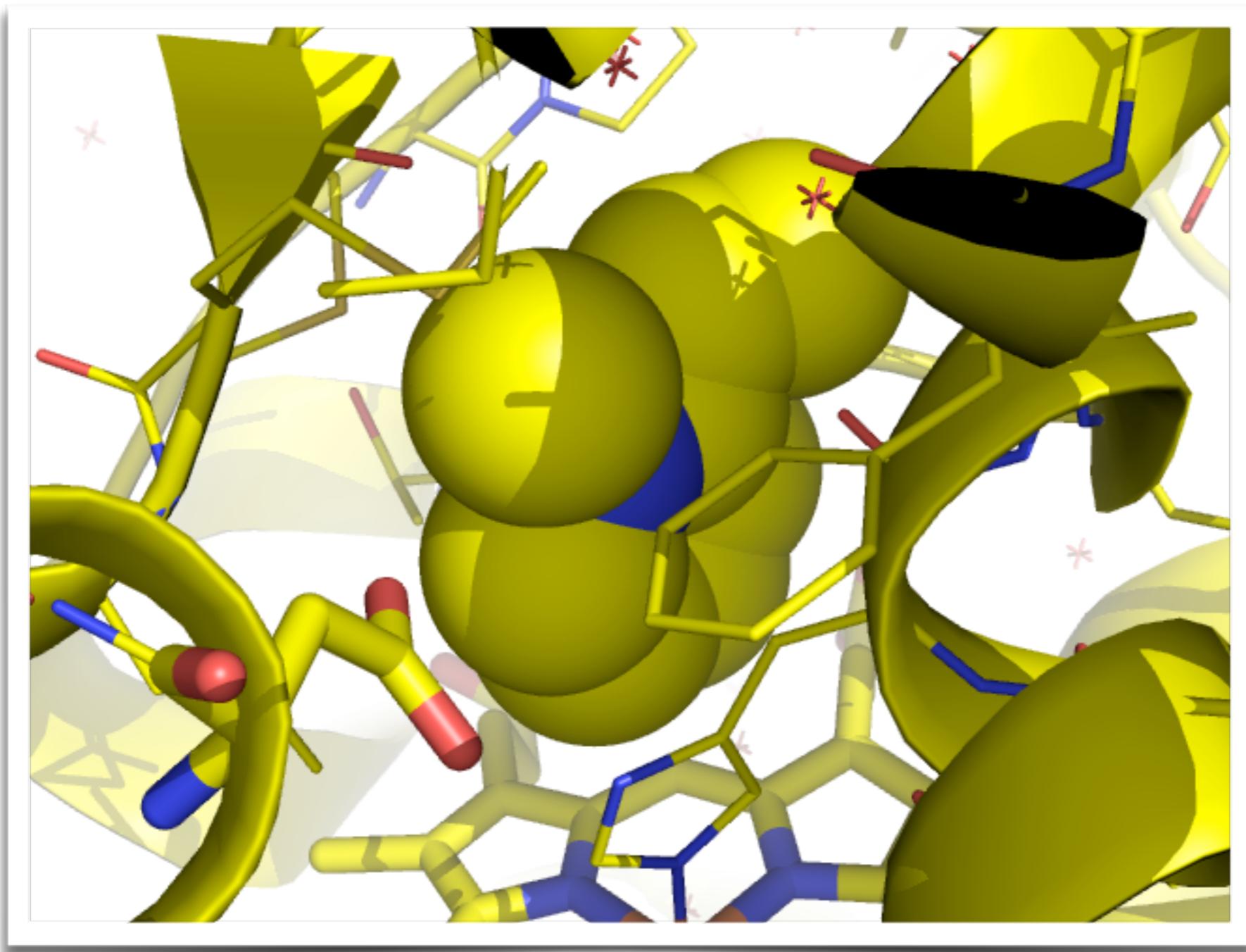


Artificial  
scaffold

Our initial tests, and previous work, gets the trend right but the slope wrong

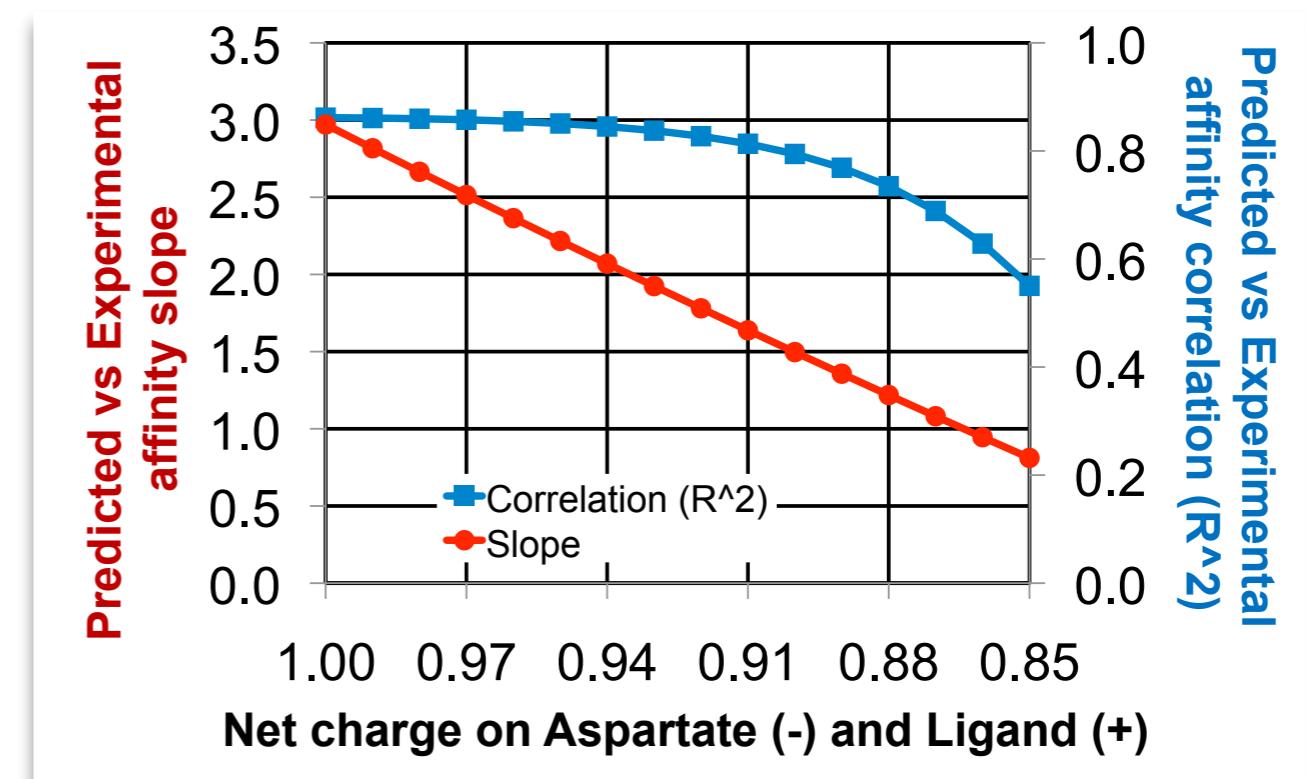


Overpolarization of the force field could lead to charge-charge interactions that are too strong

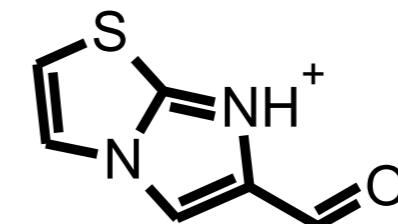
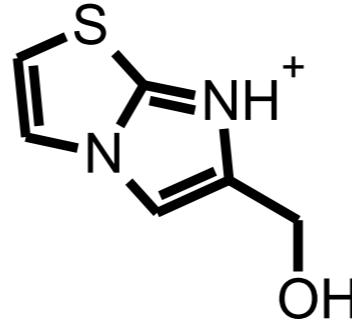
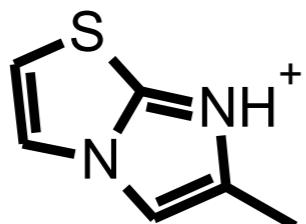
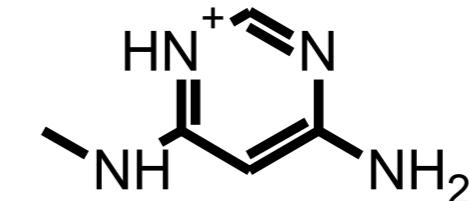
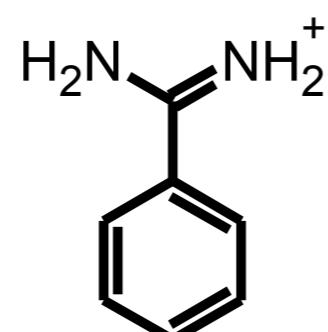
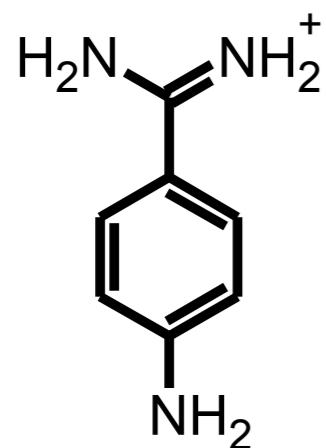
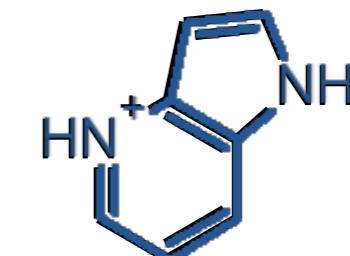
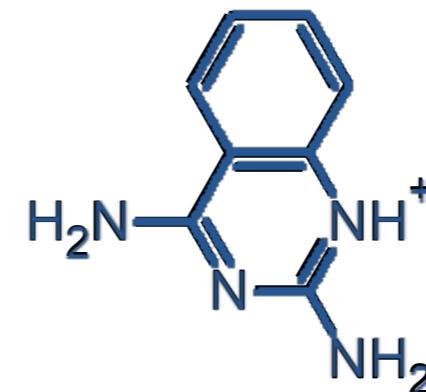
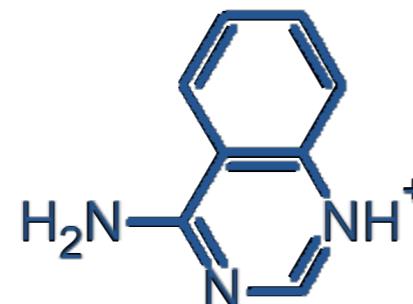
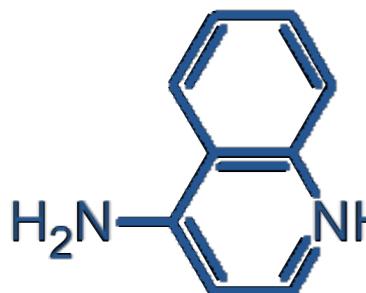


# To make good predictions, we also consider absolute free energies

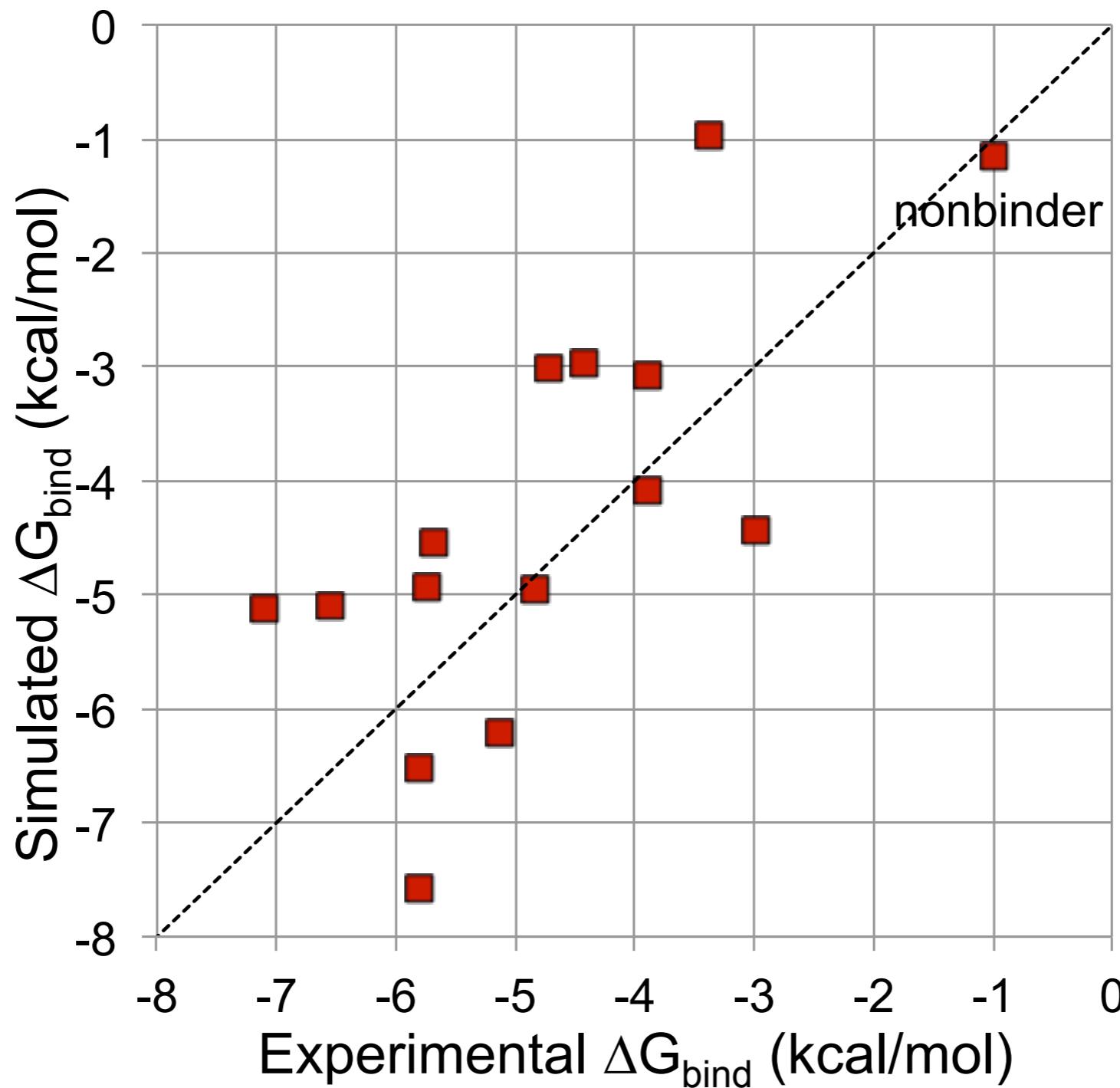
- Slopes are better with slightly smaller charges
- If charges are too small, absolute free energy in closed site is too positive
- Too negative if charges too big
- Settled on reducing force field charges by 2%



# Next challenge: Predict binding of 11 diverse ligands to CCP gateless

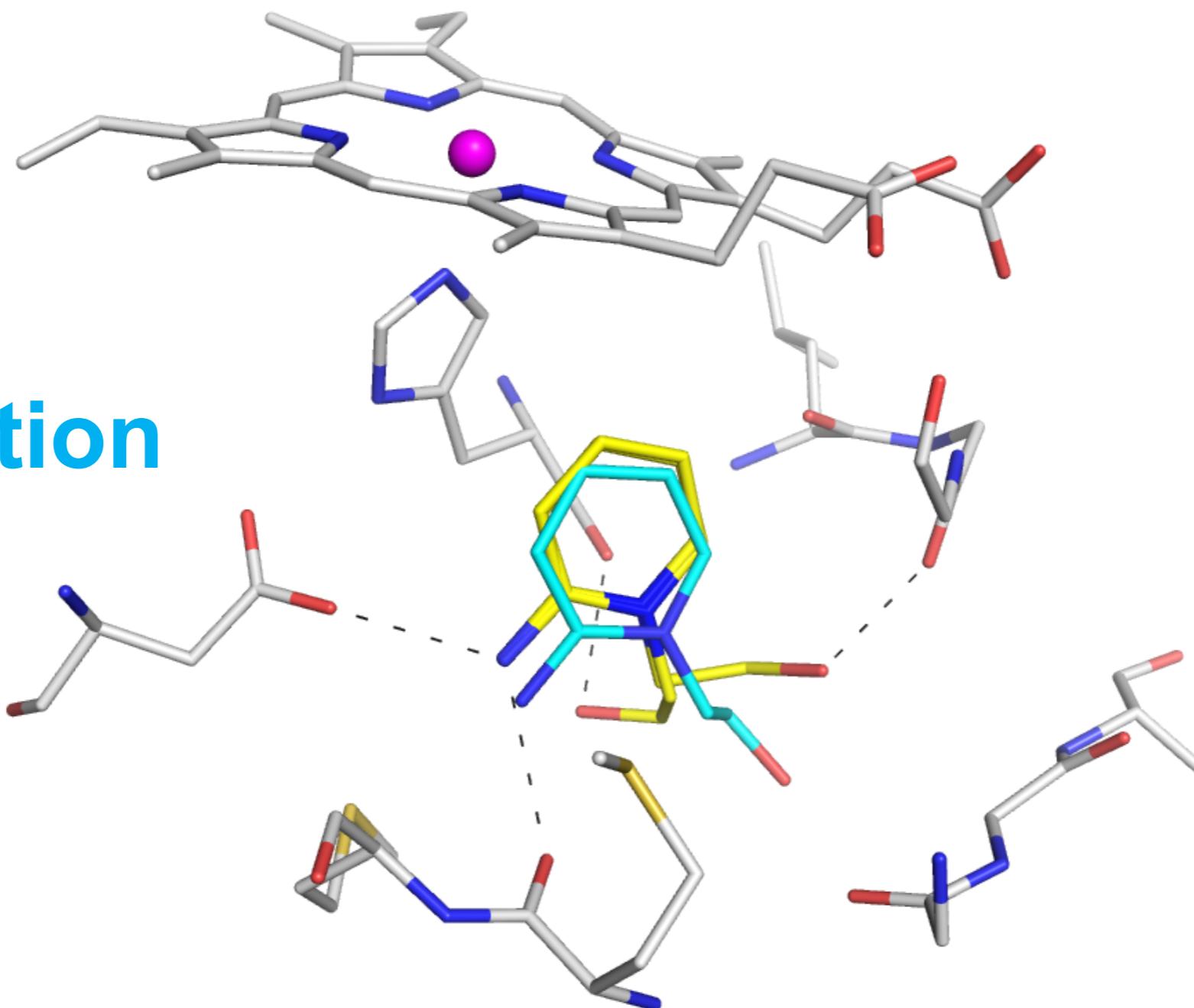


With reduced charges, we have good agreement with experiment

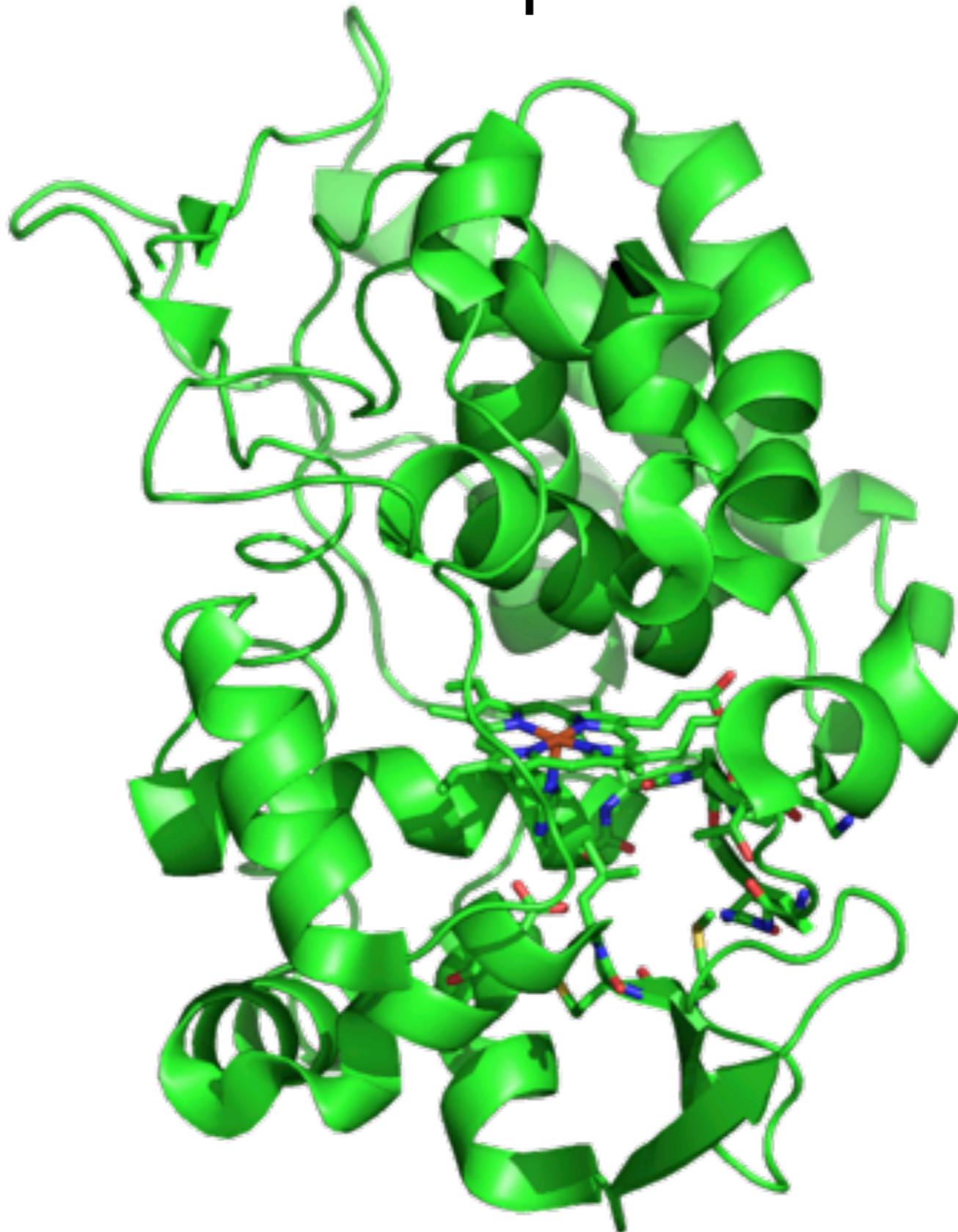


Binding mode prediction works out well here also

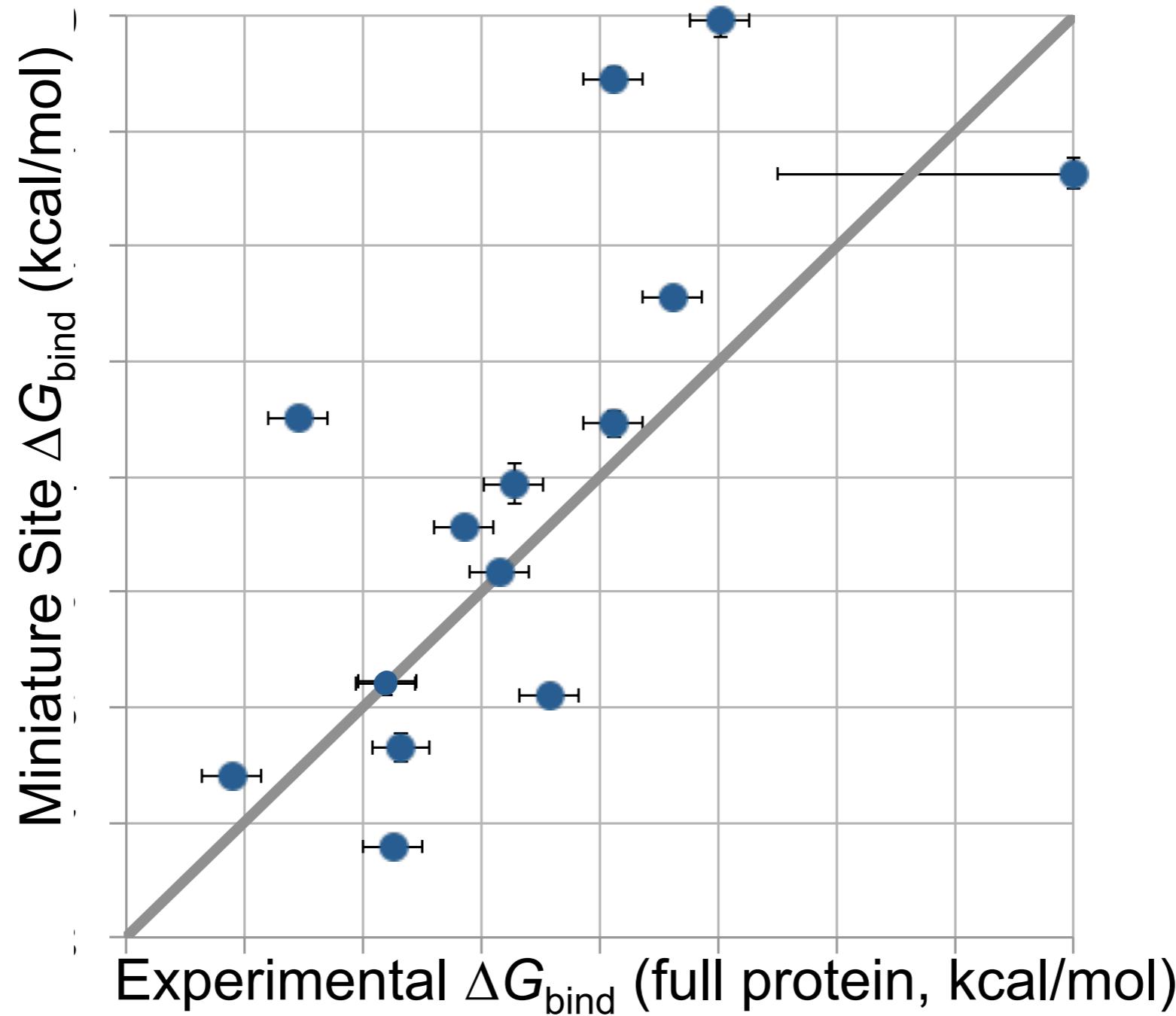
X-Ray  
Prediction



We used a “miniature binding site” of just 14 residues to test protein effects

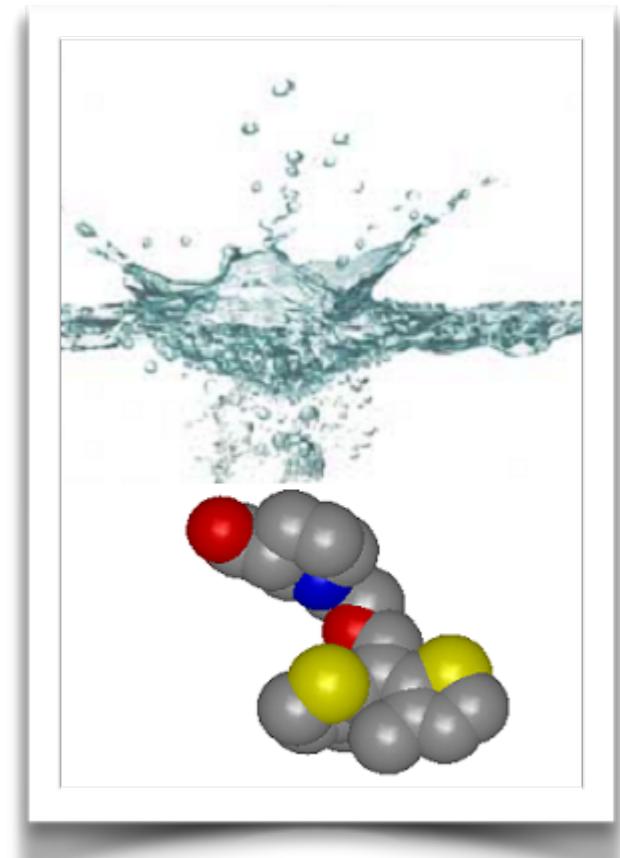


# Affinities in the miniature site are much improved even without charge scaling

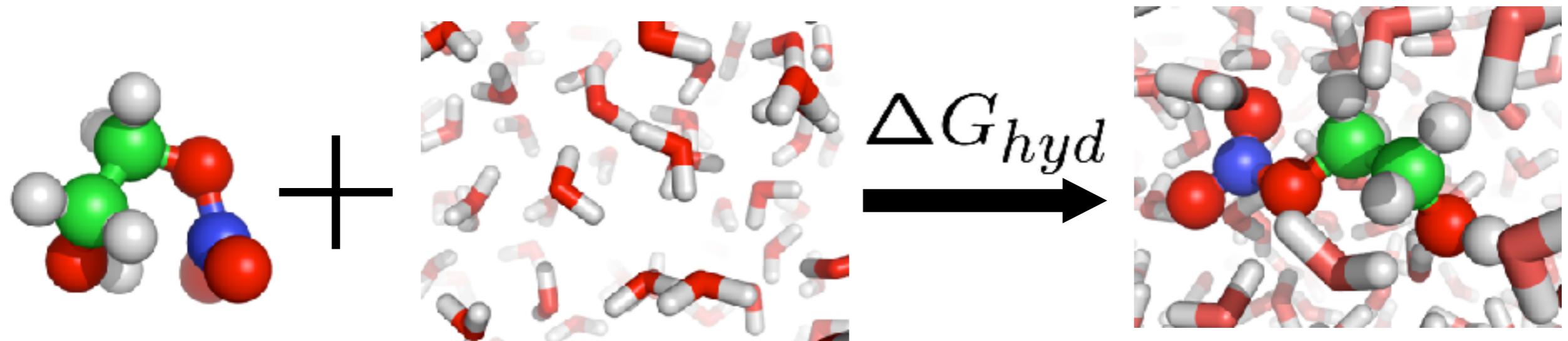


# Hydration free energies provide a test case where we can handle sampling in the right system

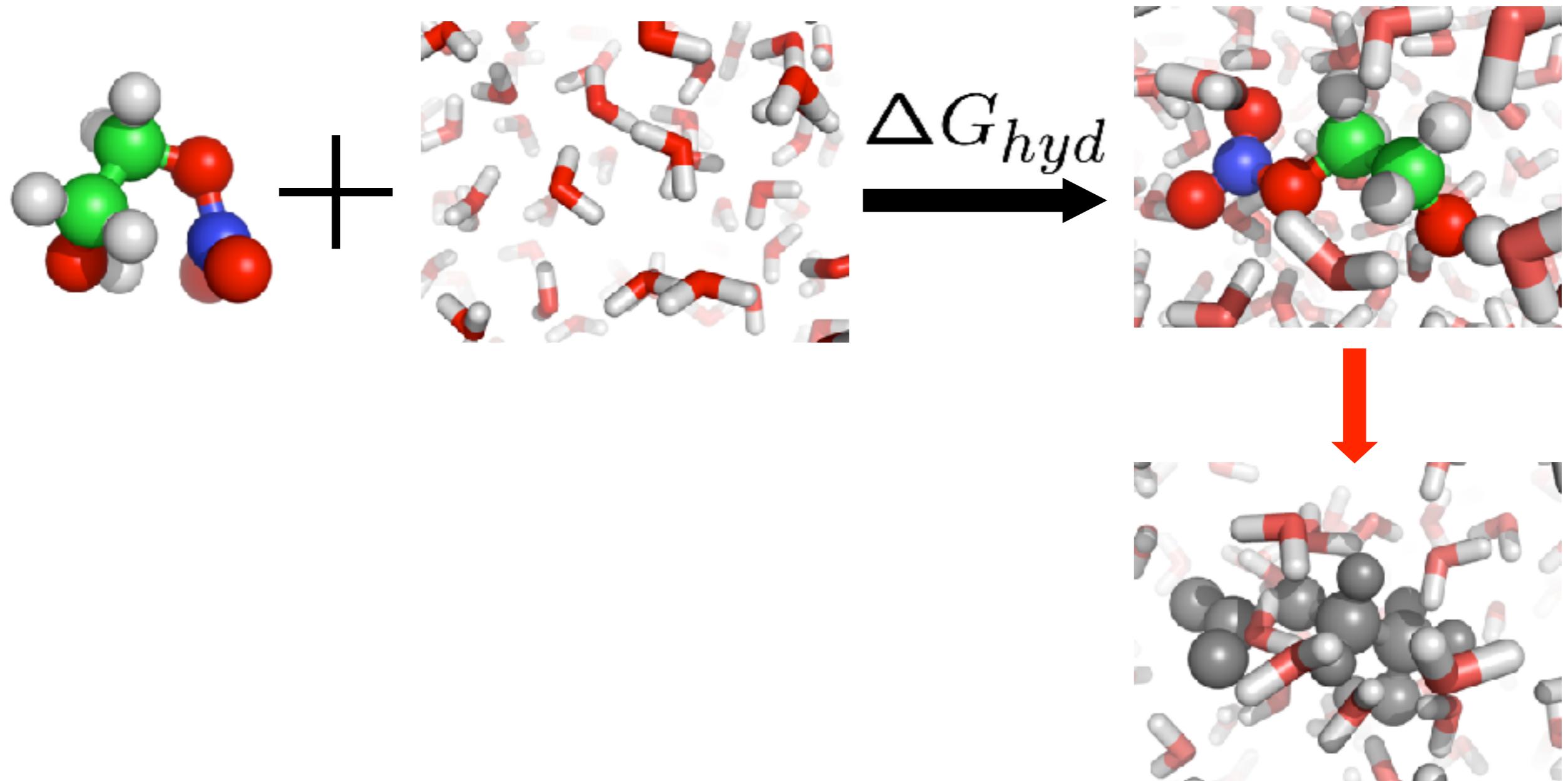
- Precise calculation straightforward
- Enables quantitative comparison with experiment
- Can help improve force fields
- Proxy for accuracy we might expect in drug discovery applications



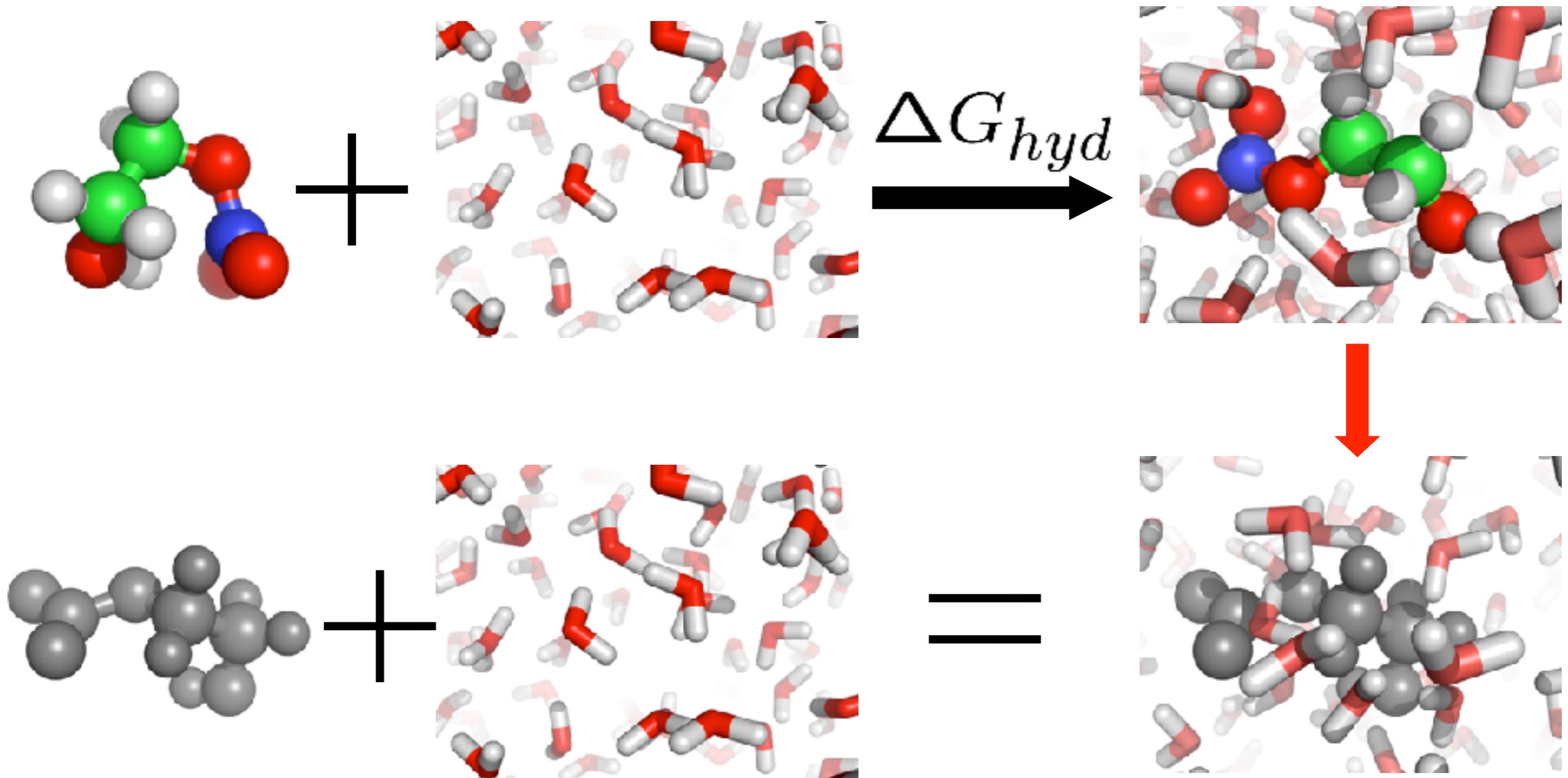
# Alchemical calculations yield rigorous hydration free energies



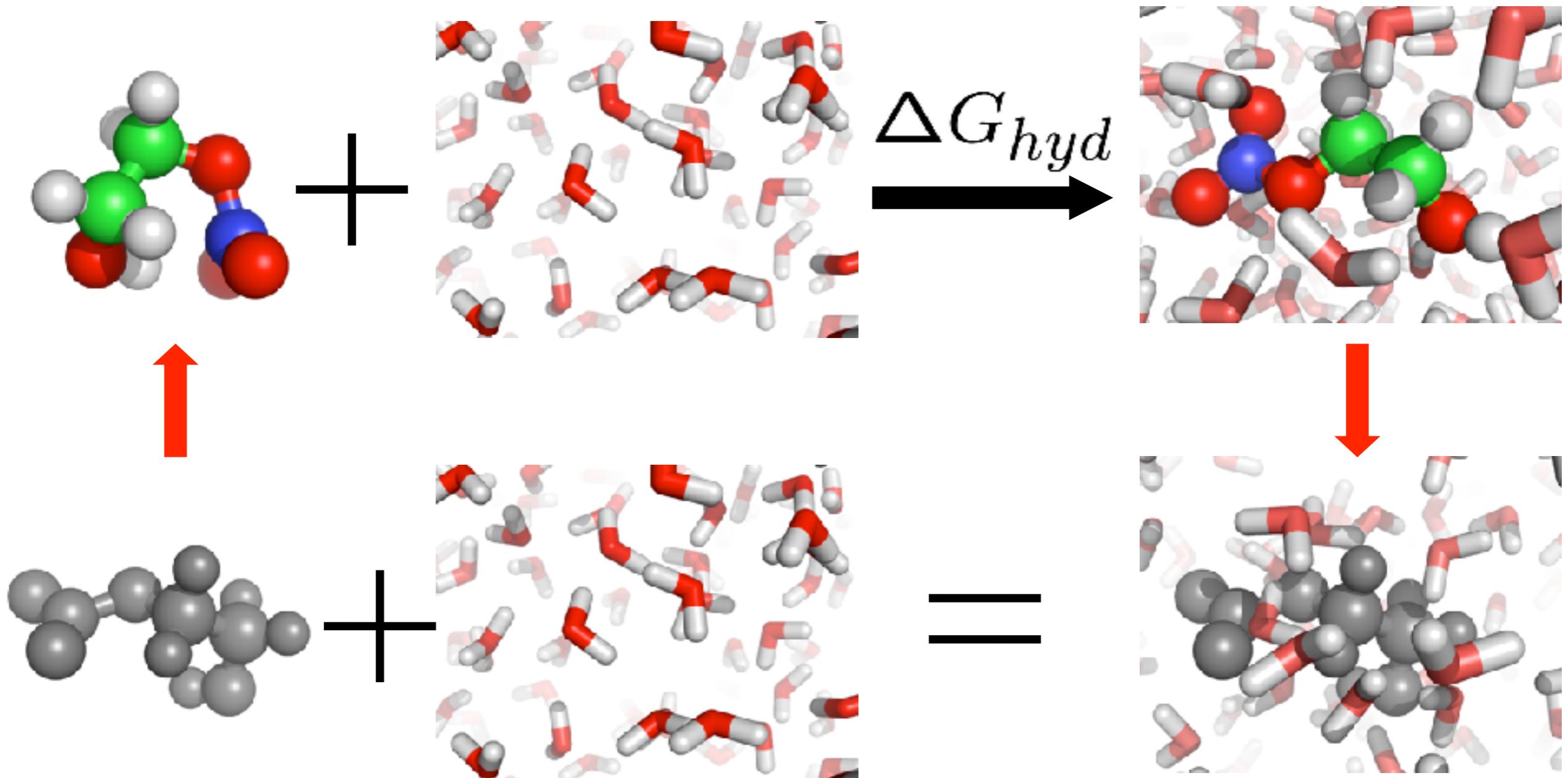
# Alchemical calculations yield rigorous hydration free energies



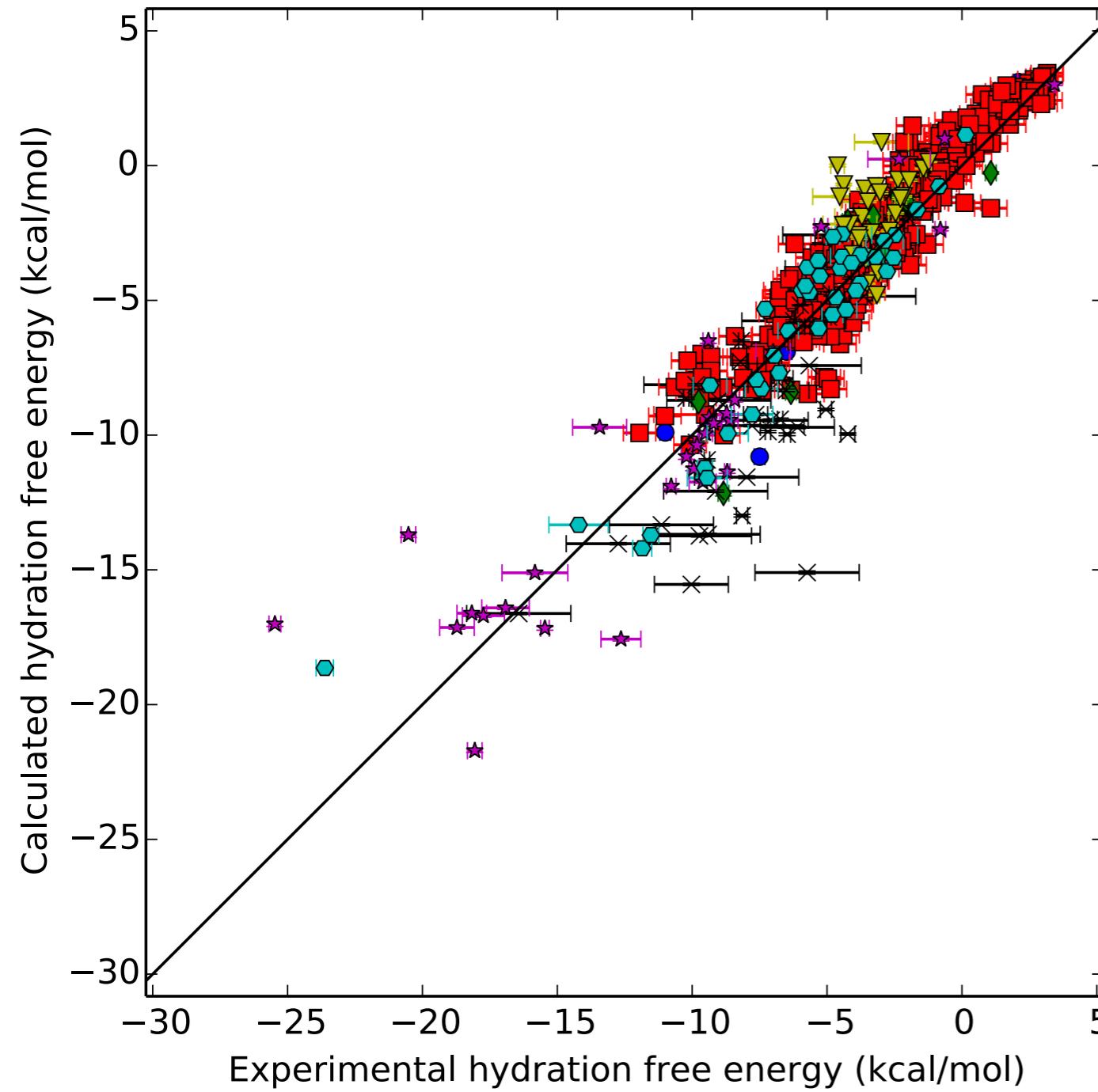
# Alchemical calculations yield rigorous hydration free energies



# Alchemical calculations yield rigorous hydration free energies

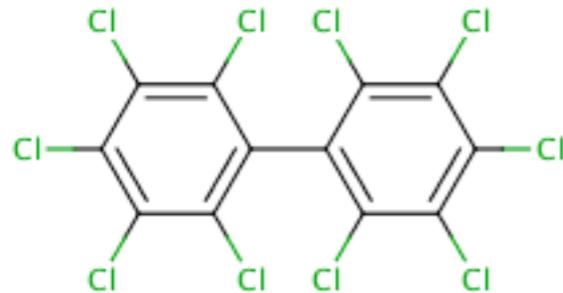


# Hydration free energies have been helpful for a variety of purposes, so we updated our “504 molecule set” as FreeSolv

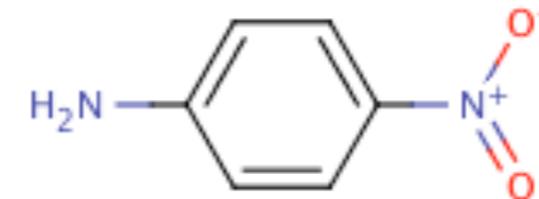


- 643 molecules
- Expt. & calc. values (GAFF)
- Structures, parameters, input files
- Literature citations
- (Curation ongoing)
- FreeSolv
- Permanent cite-able URL,
- <http://www.escholarship.org/uc/item/6sd403pz>

From the standpoint of testing force fields for drug-like molecules, though, we have a ways to go



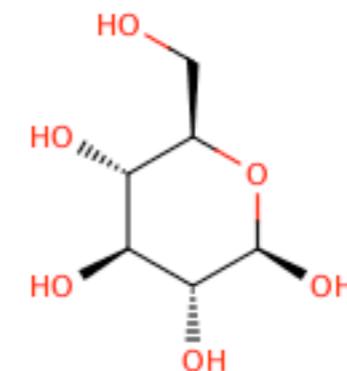
Largest MW



4-nitroaniline: Largest dipole



Octafluorocyclobutane  
Most hydrophobic

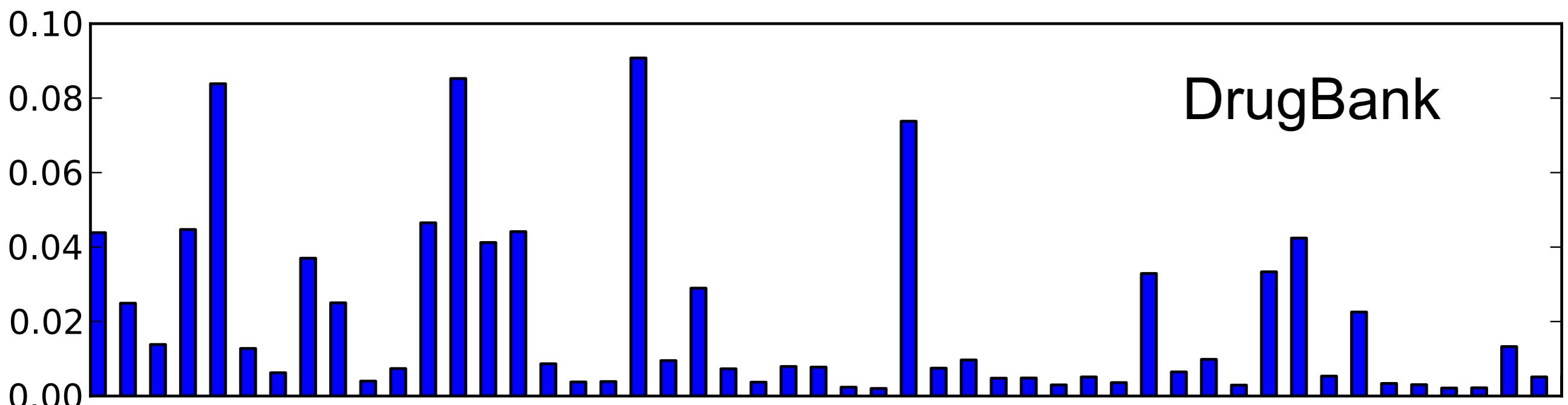


Most negative  
experimental value

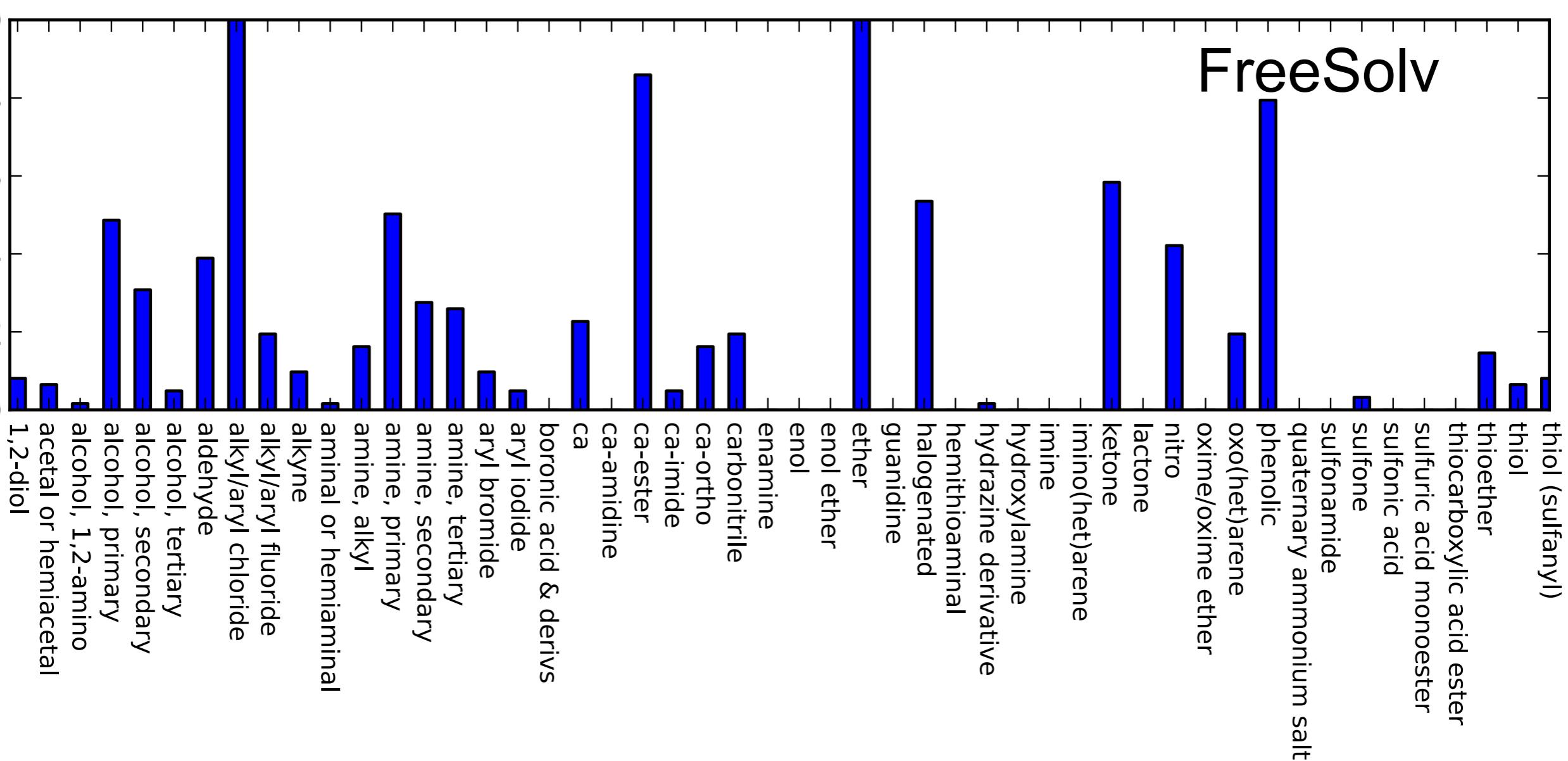
Various functional groups are underrepresented or  
not represented compared to drugs

DrugBank  
FreeSolv

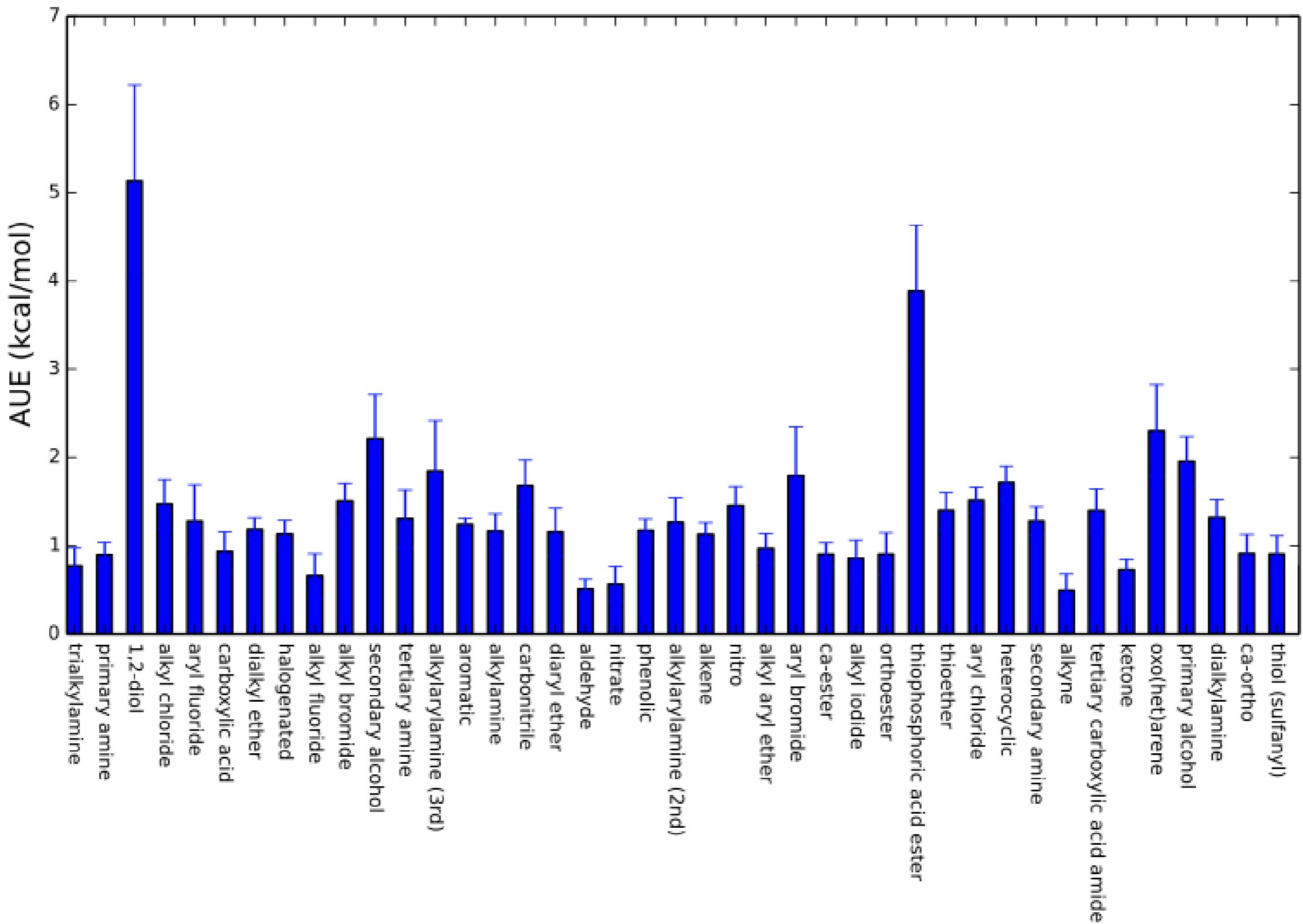
# DrugBank



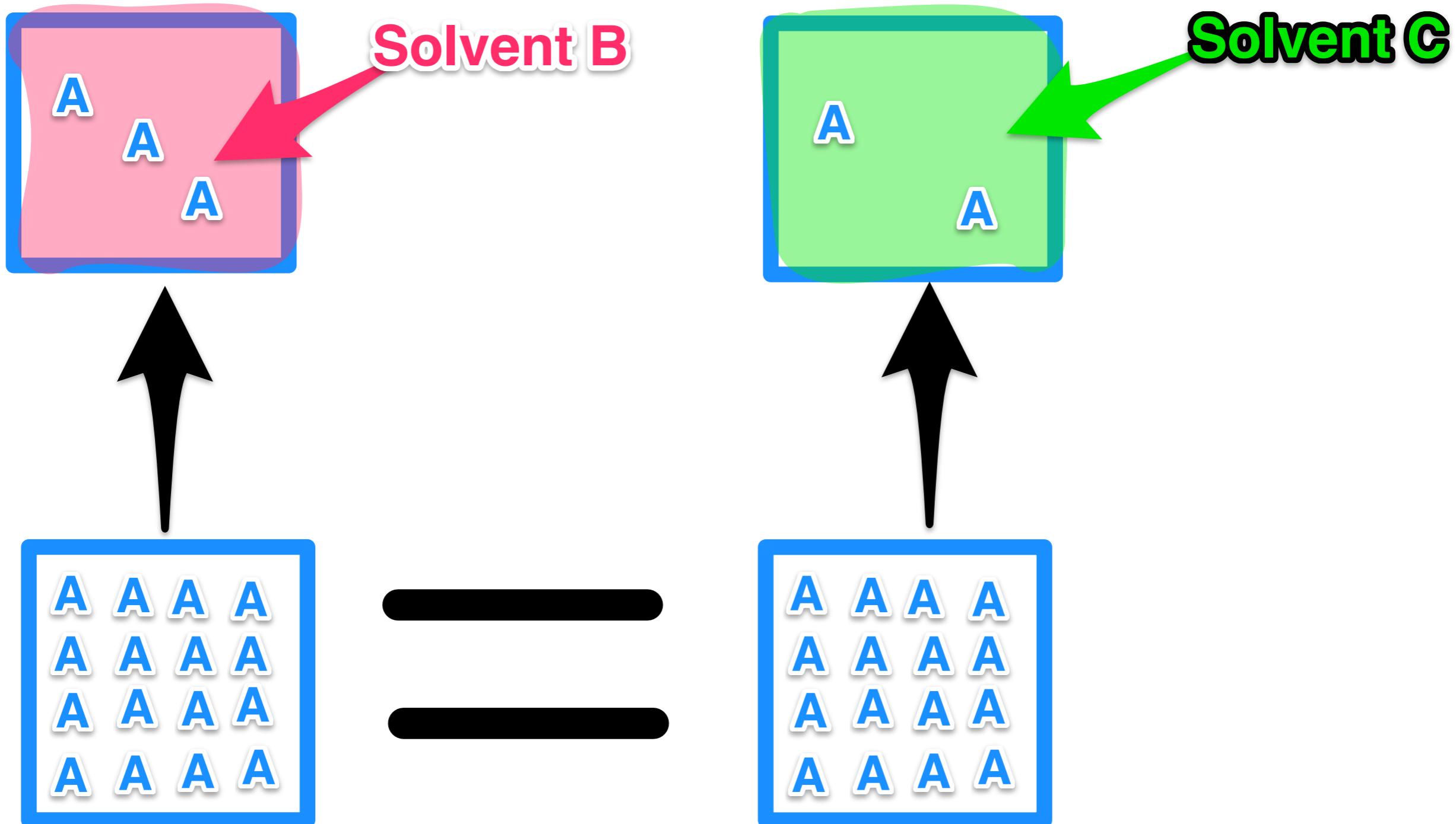
# FreeSolv



# Certain functional groups appear to still be particularly problematic

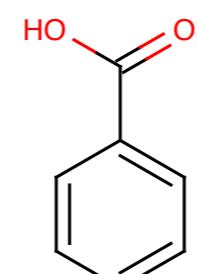
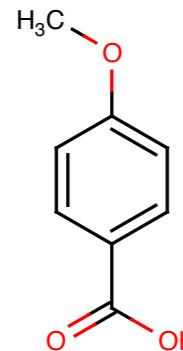
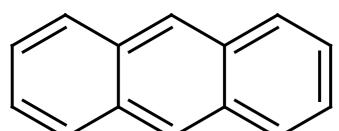
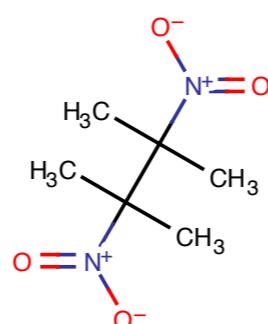
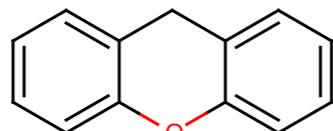
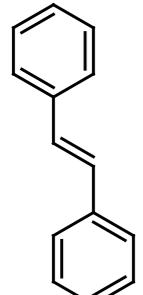
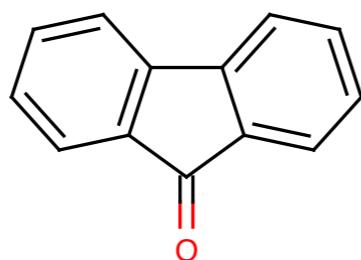
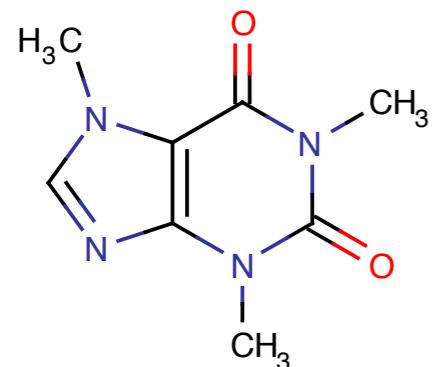


# We're out of hydration free energies. What's next? Relative solubilities

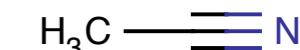
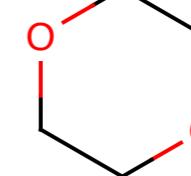
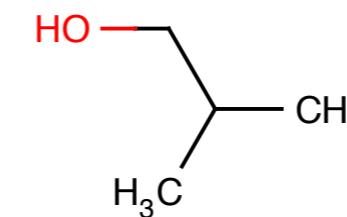
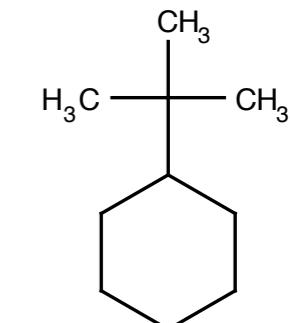
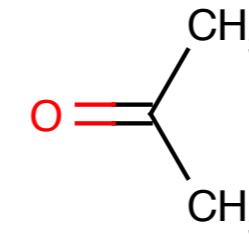
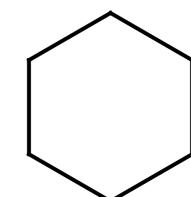
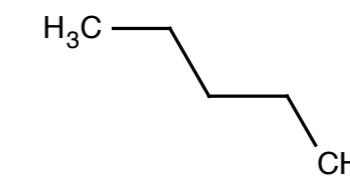
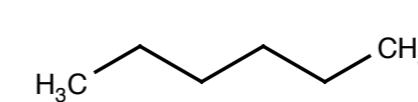
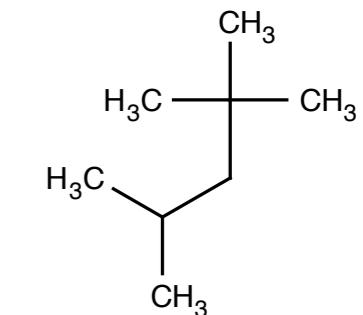
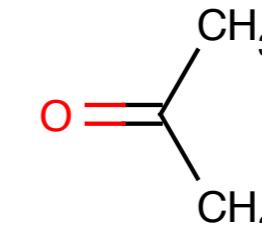
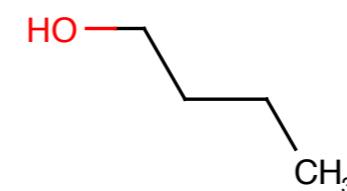


We calculated relative solubilities for a few compounds - 10+ solutes, 10+ solvents

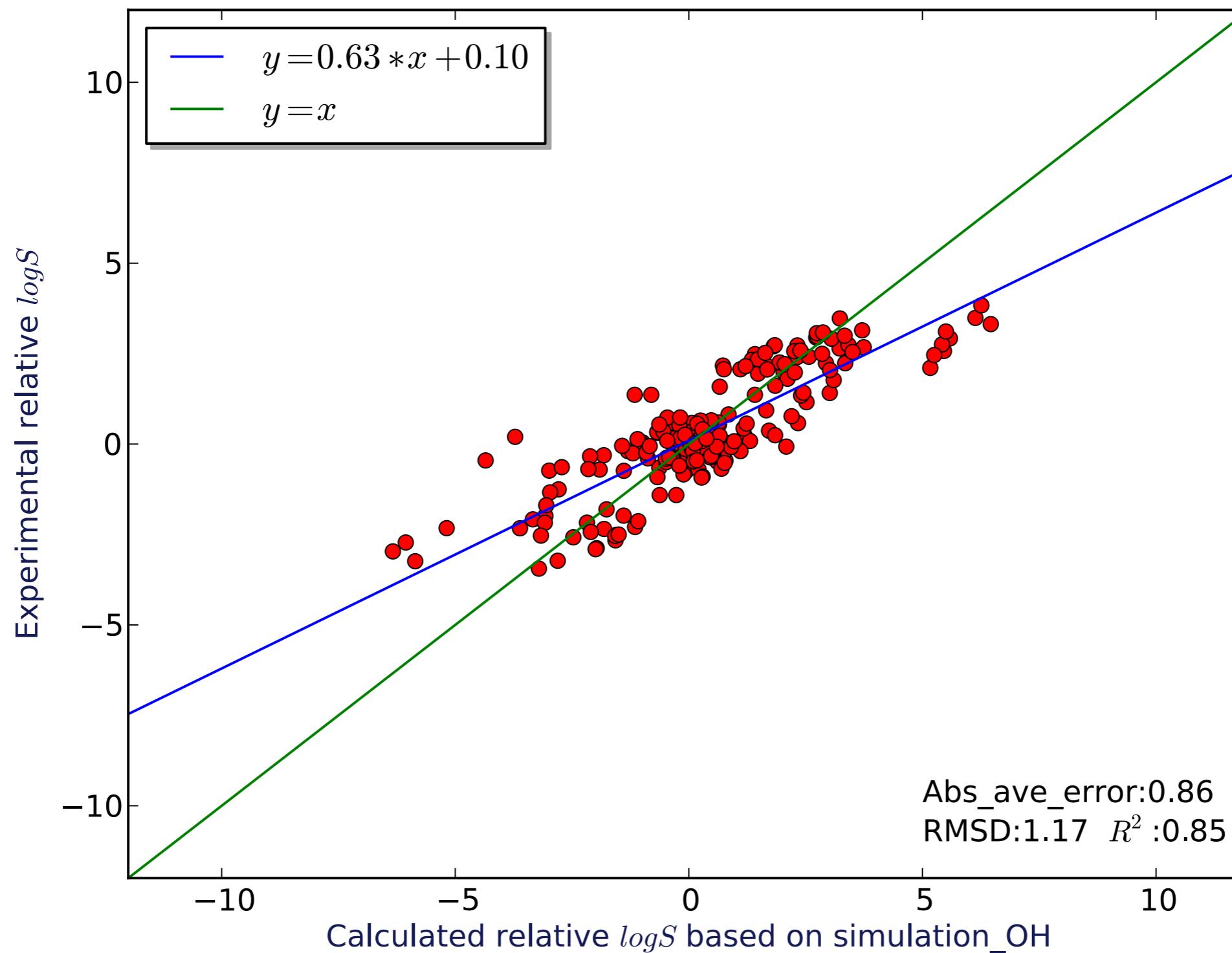
Solutes



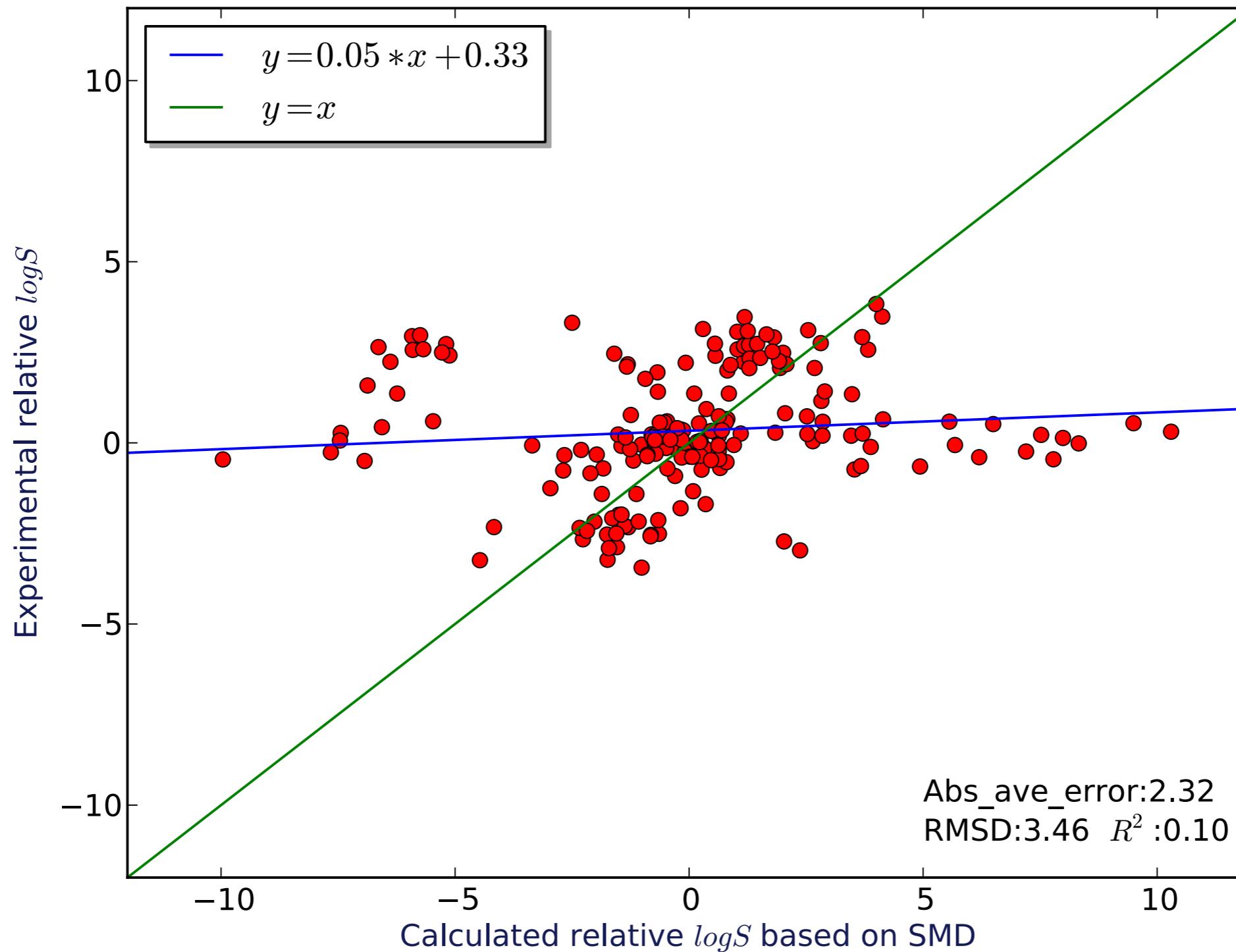
Solvents



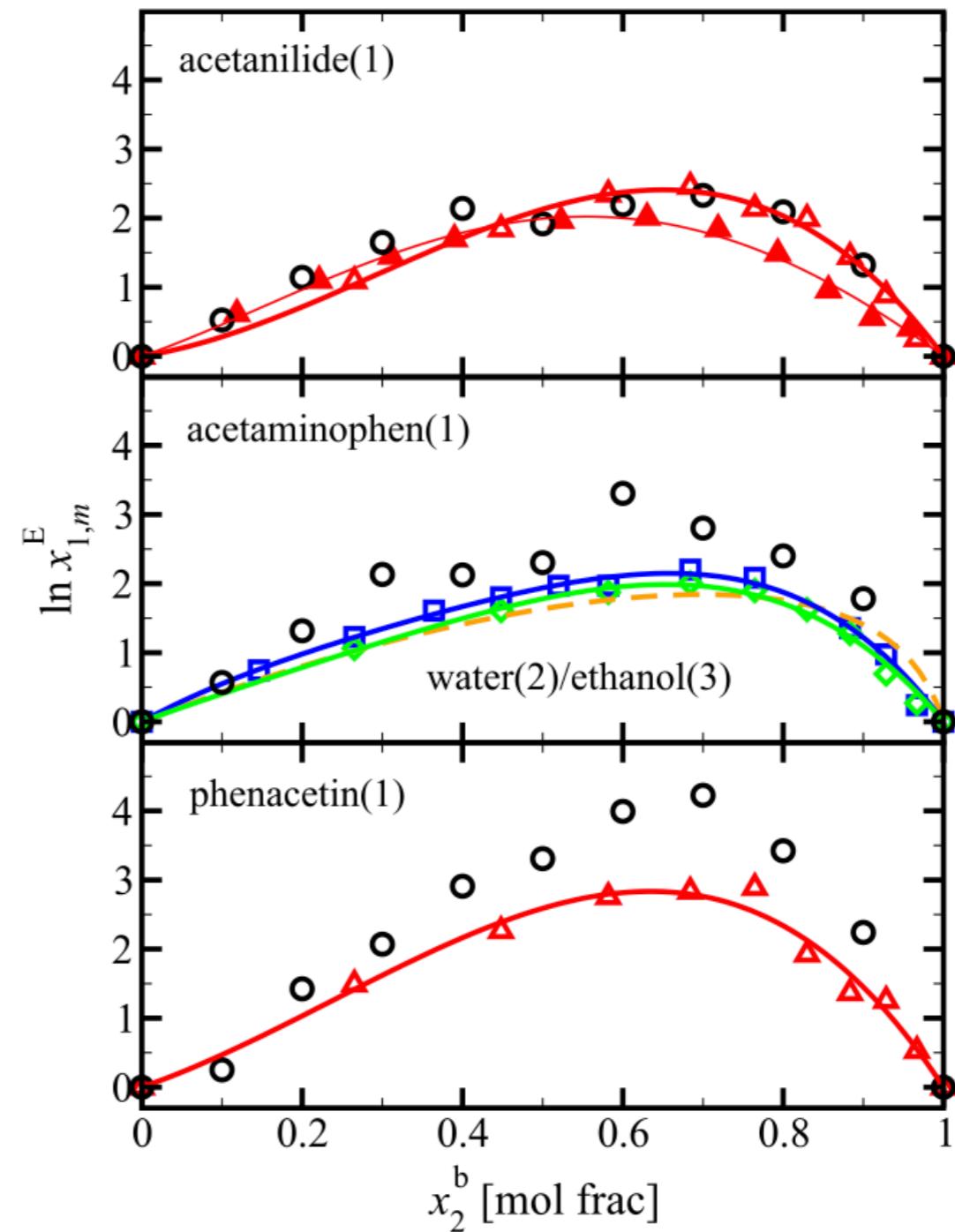
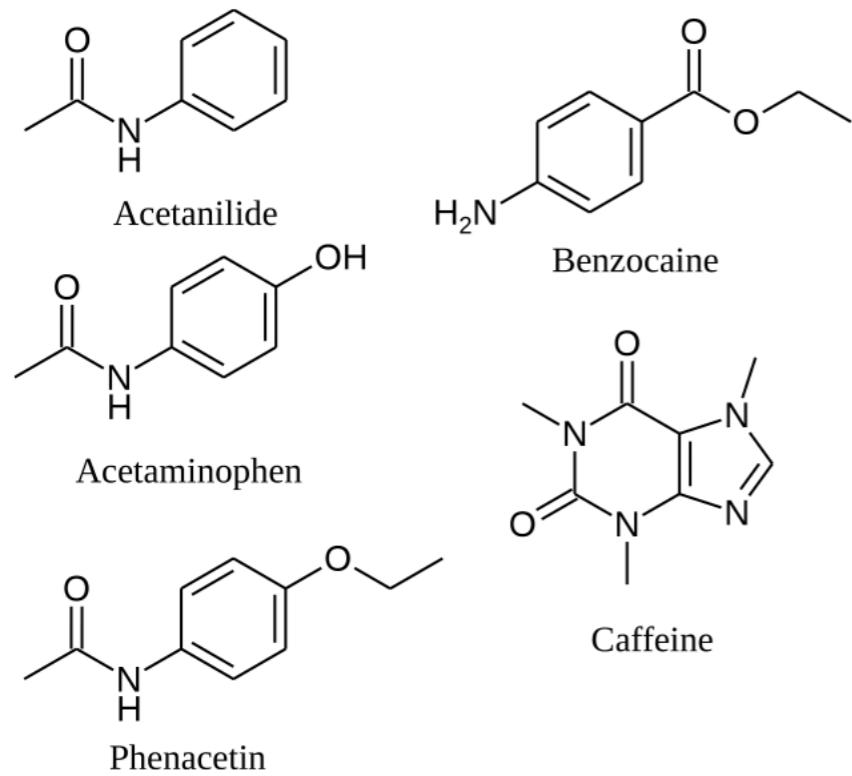
Initial results for a few relative solubilities appear promising, even compared to more highly parameterized models



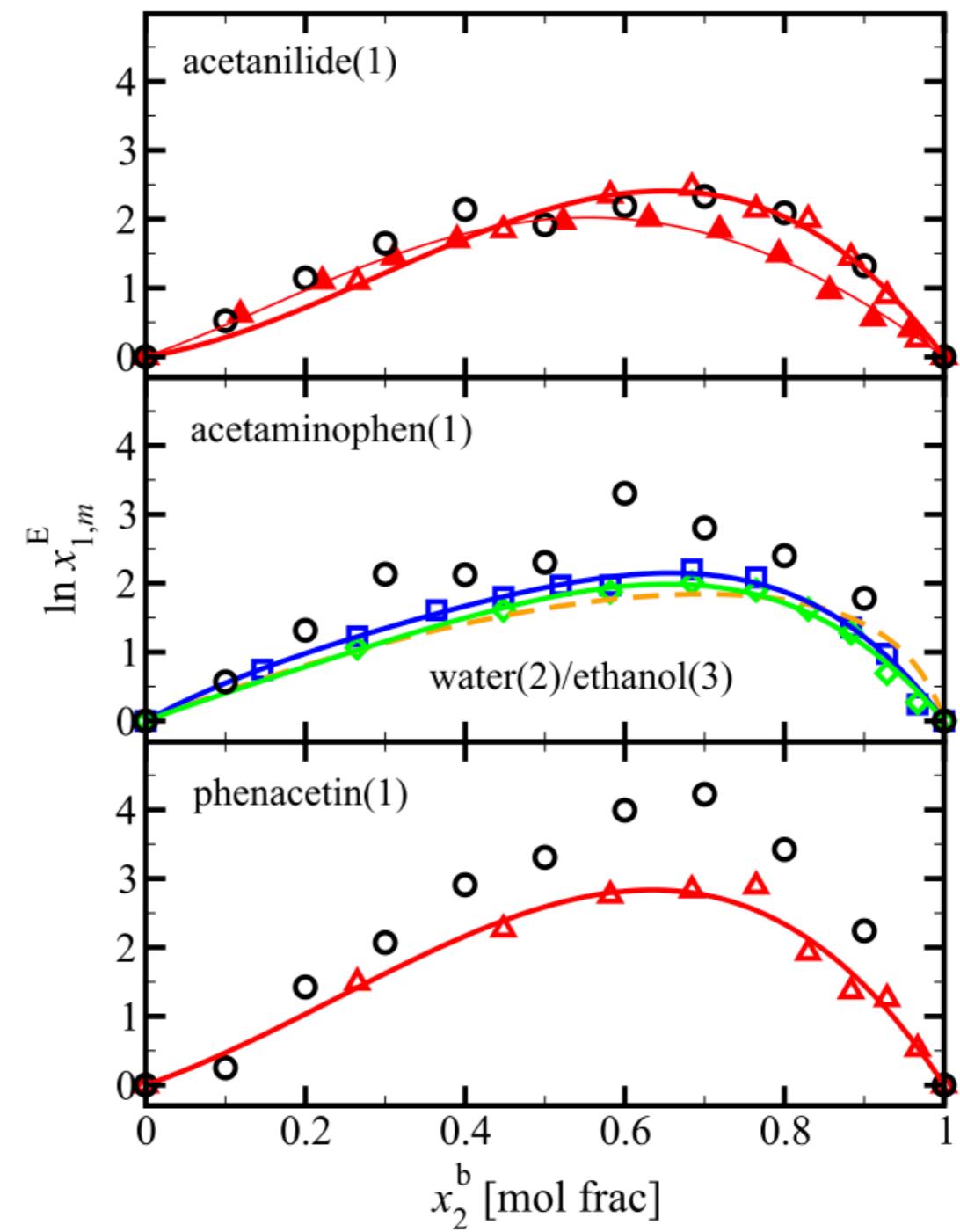
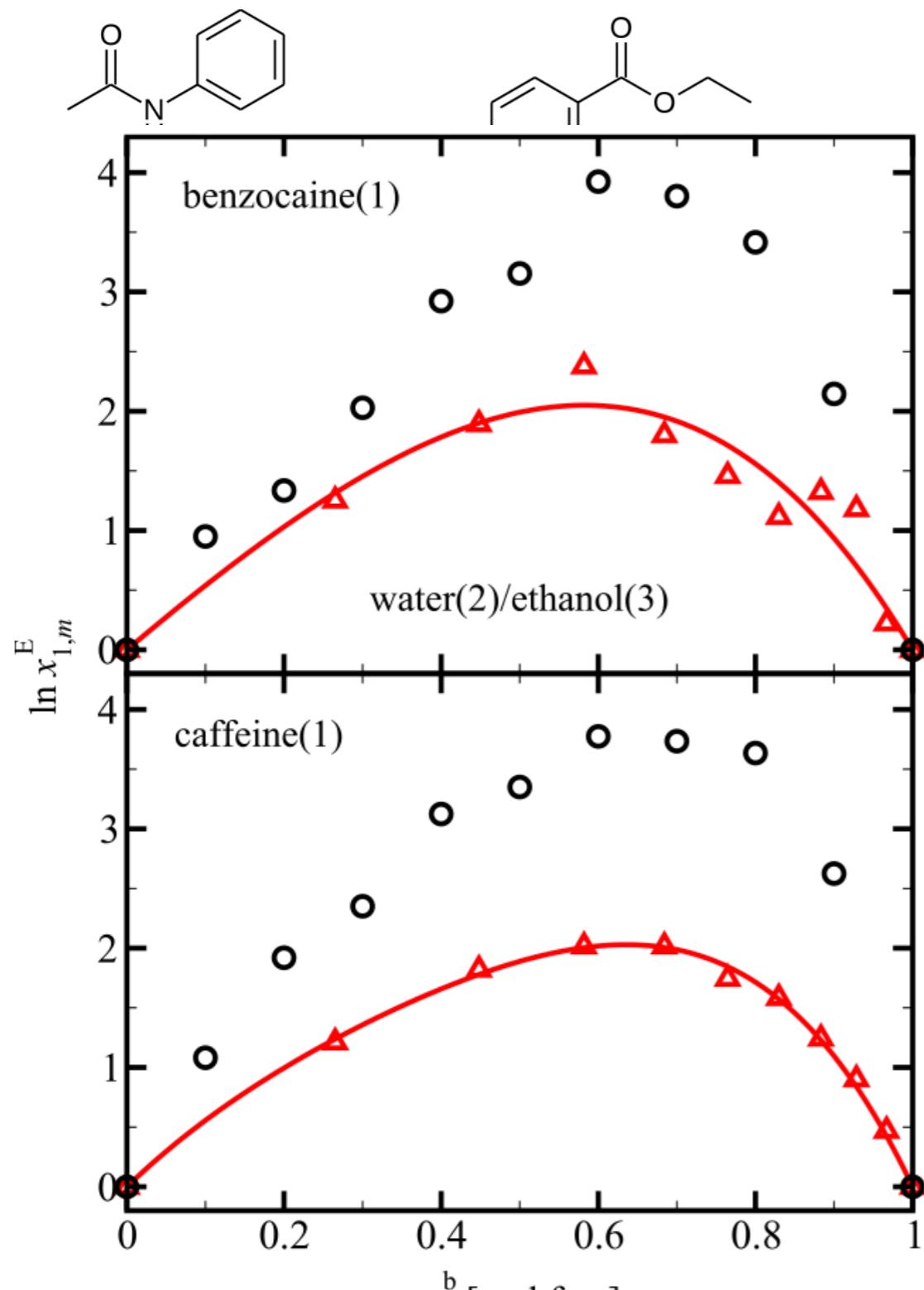
Initial results for a few relative solubilities appear promising, even compared to more highly parameterized models



# We also looked at excess solubility in binary mixtures of water/ethanol



# We also looked at excess solubility in binary mixtures of water/ethanol

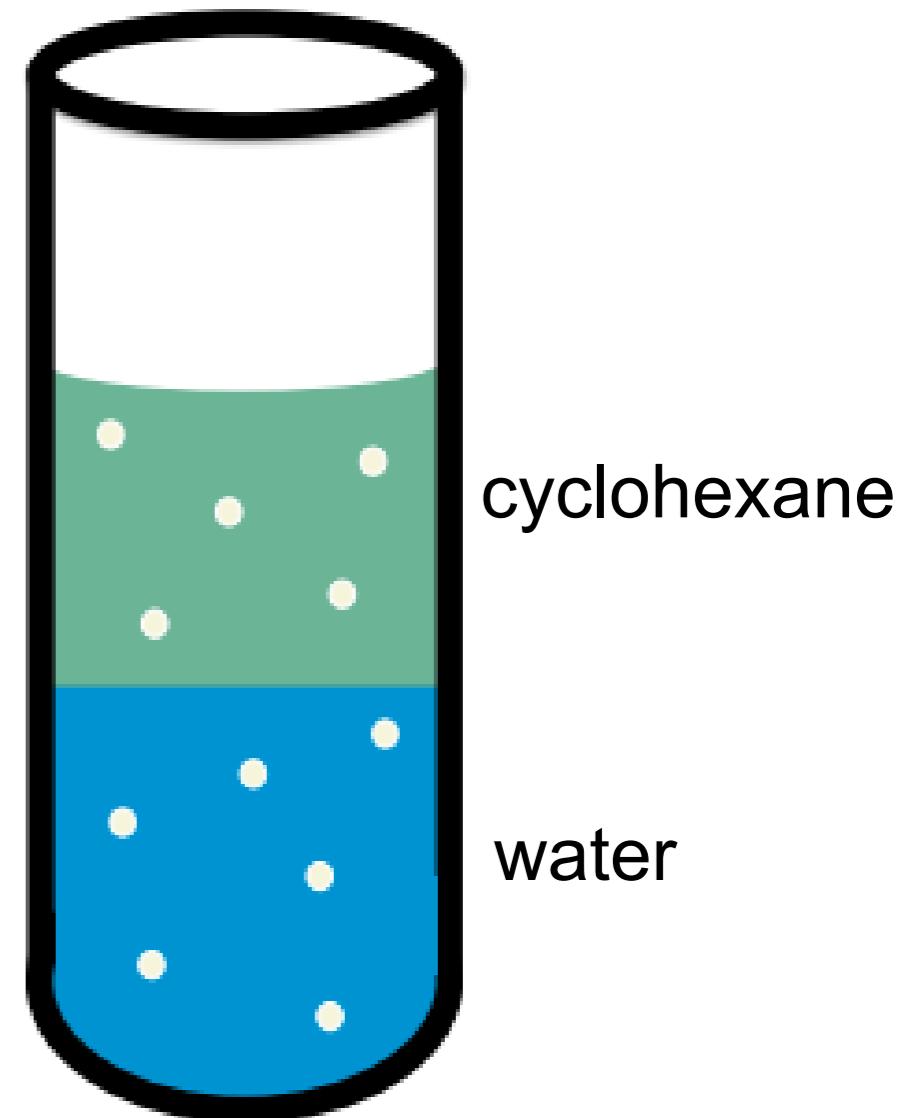


It's also possible to do partition/distribution:  
At Genentech, distribution coefficients were  
obtained for 53 compounds

Partition coefficients and distribution coefficients  
are similar, but the latter includes all species:

$$P_{cyc} = \frac{[\text{Neutral solute in cyclohexane}]}{[\text{Neutral solute in water}]}$$

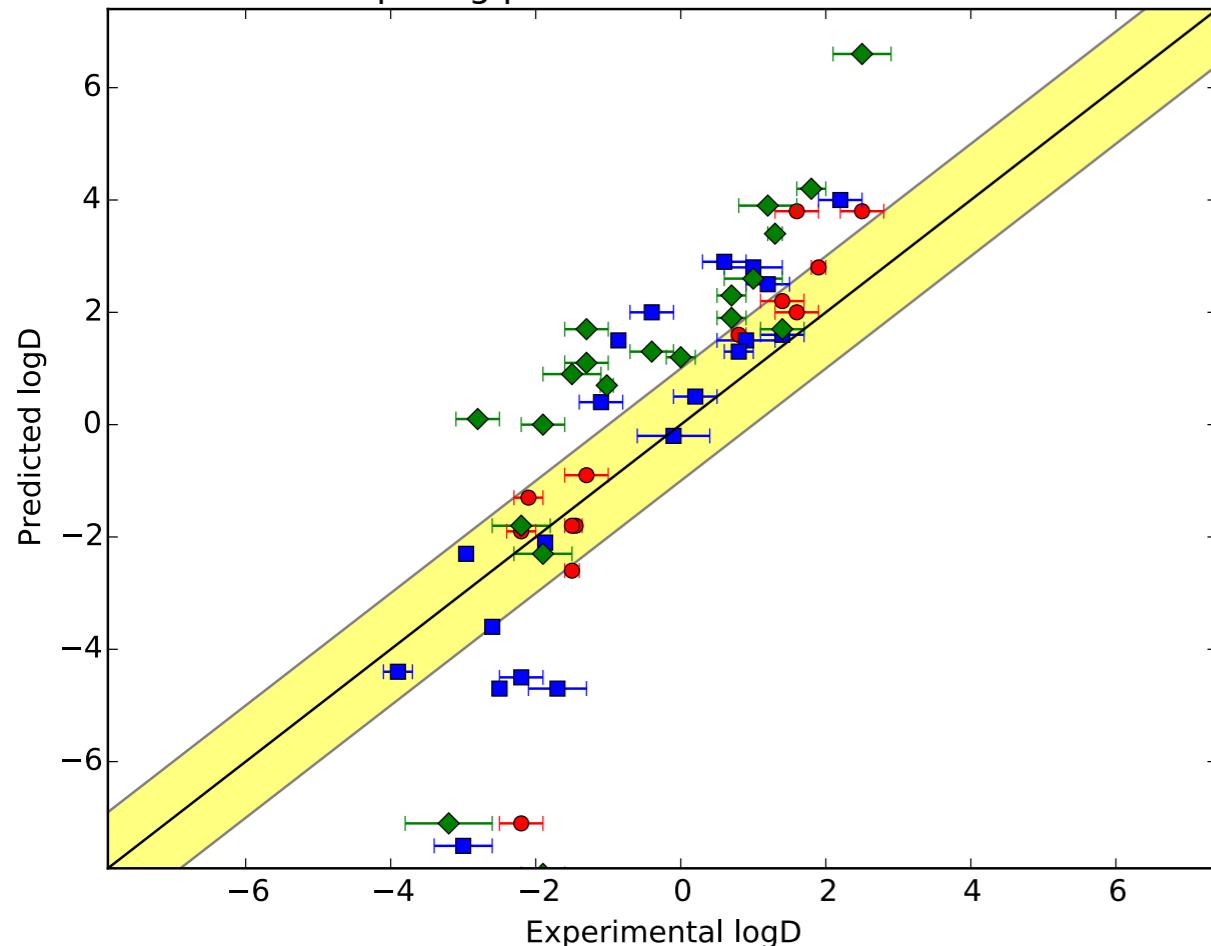
$$D_{cyc} = \frac{[\text{Solute in cyclohexane}]}{[\text{Solute in water}]}$$



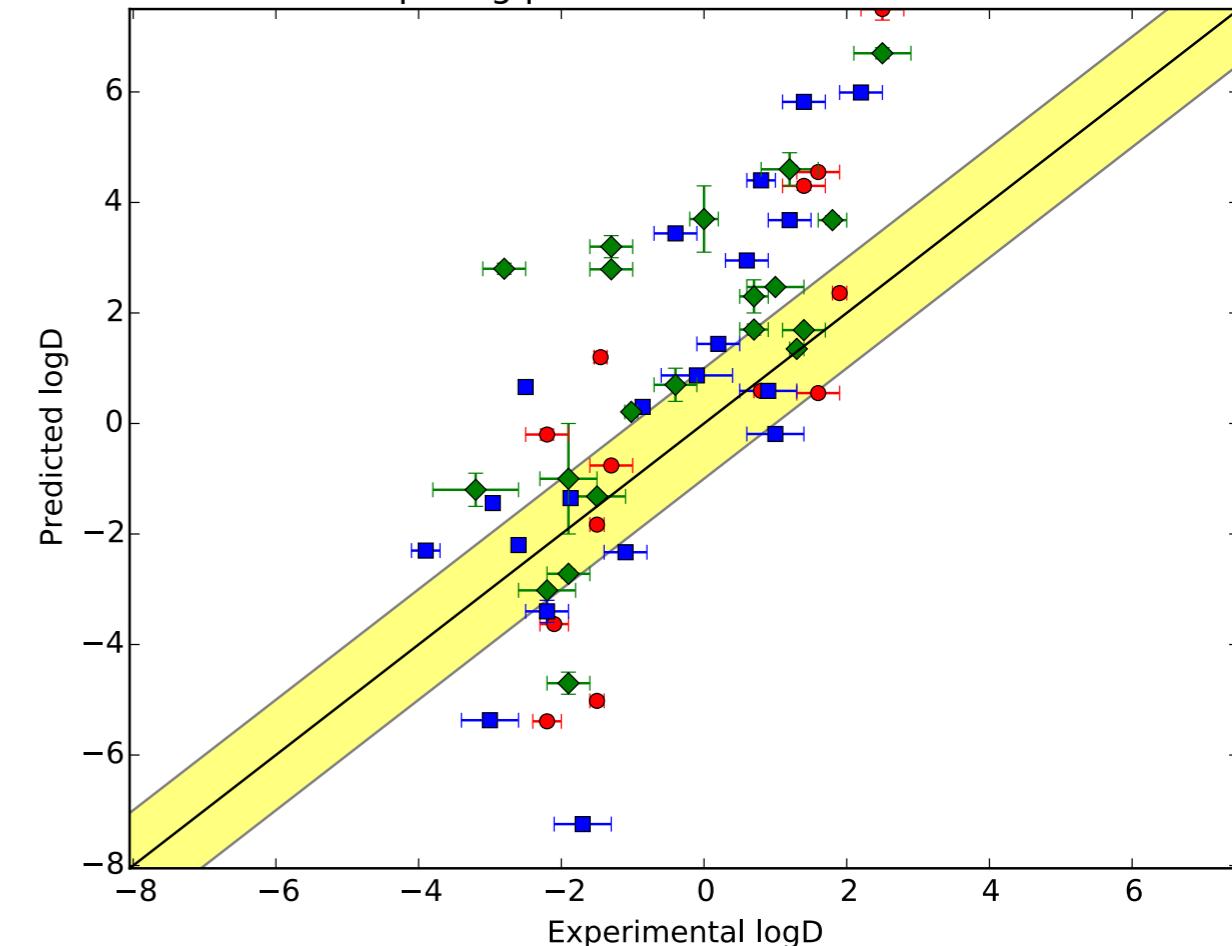
pH dependent, so we report  $\log D_{7.4}$  at pH 7.4

# In the SAMPL5 blind prediction challenge, success was mixed

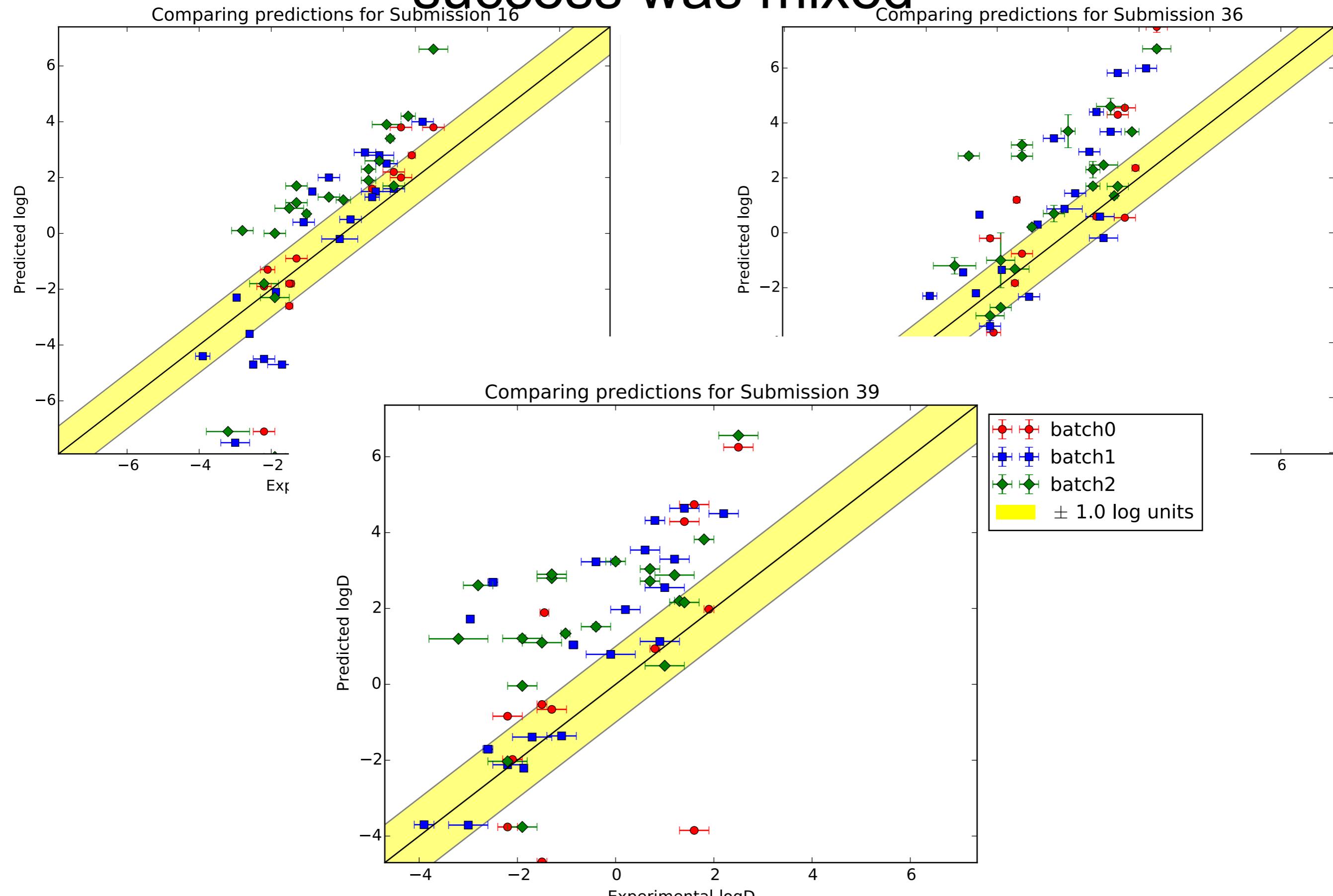
Comparing predictions for Submission 16



Comparing predictions for Submission 36

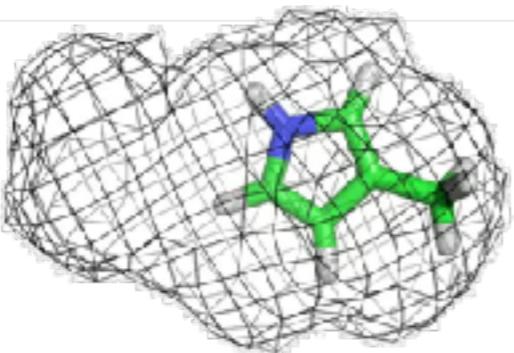
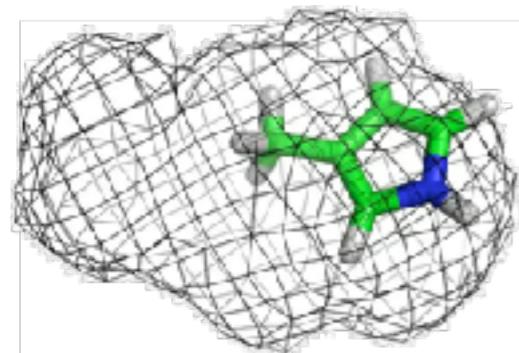


# In the SAMPL5 blind prediction challenge, success was mixed



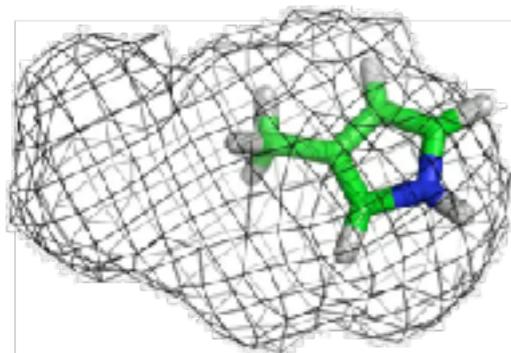
# What did we learn?

# What did we learn?

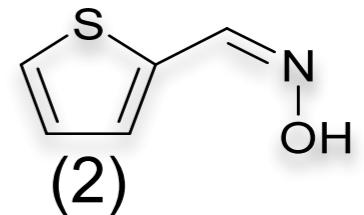
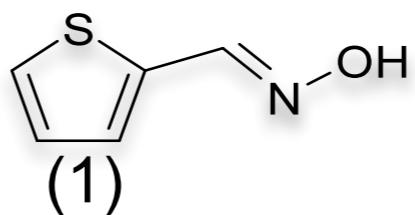
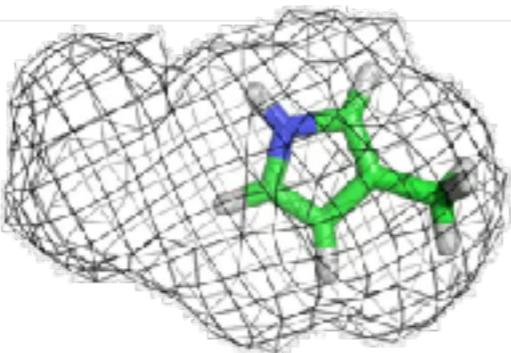


Consider multiple orientations

# What did we learn?

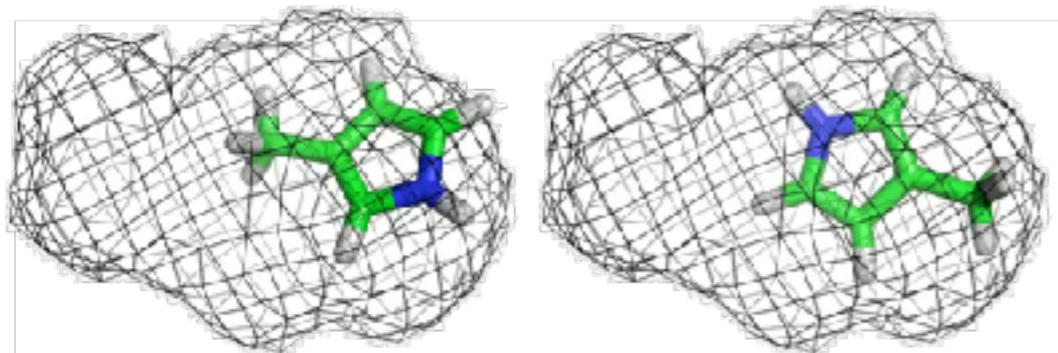


Consider multiple orientations

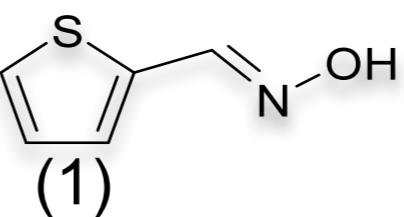


We can make nontrivial predictions

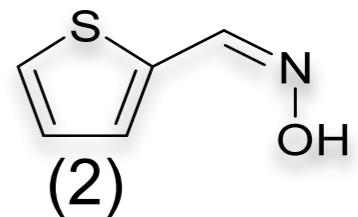
# What did we learn?



Consider multiple orientations

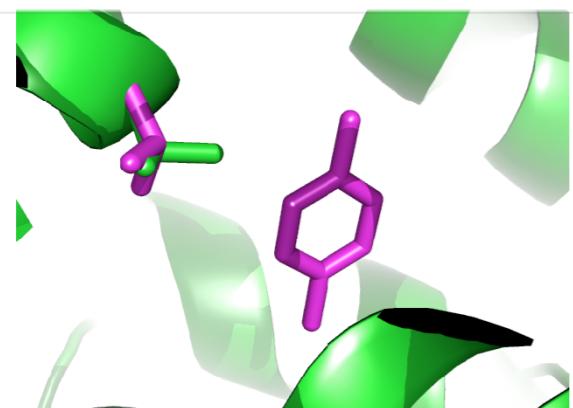


(1)



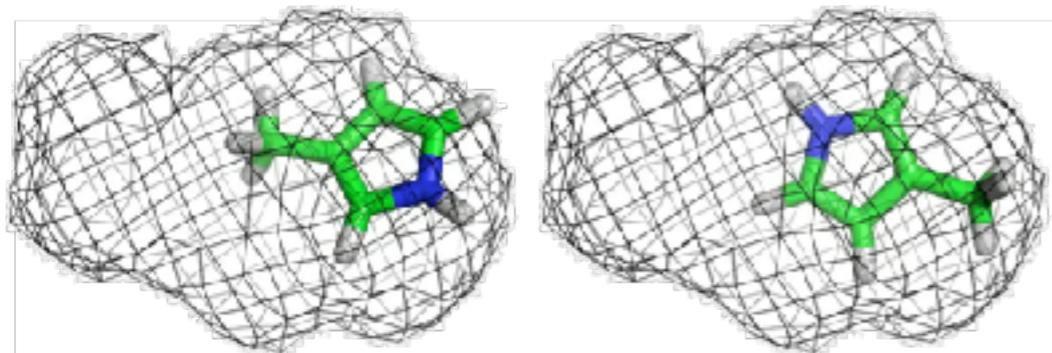
(2)

We can make nontrivial predictions

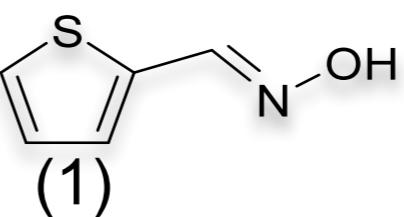


Conformational change is key

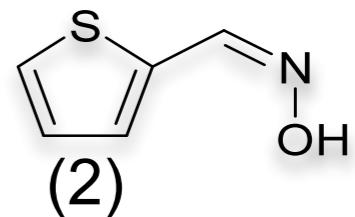
# What did we learn?



Consider multiple orientations

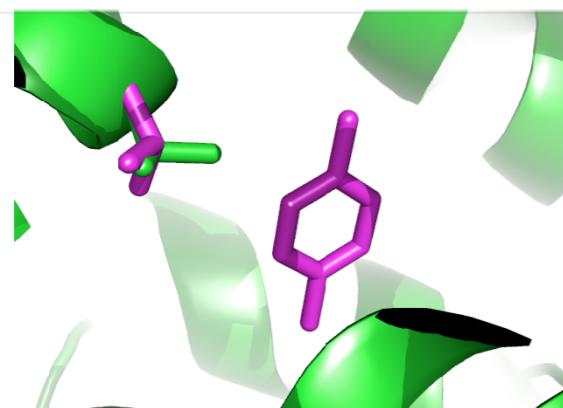


(1)

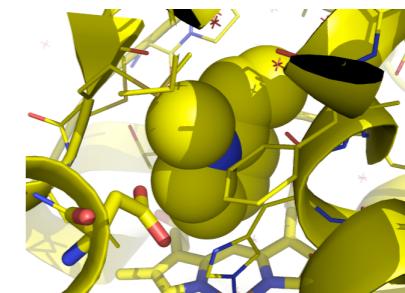
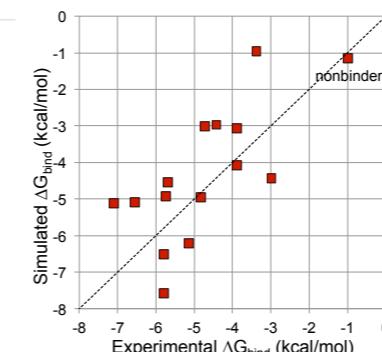


(2)

We can make nontrivial predictions

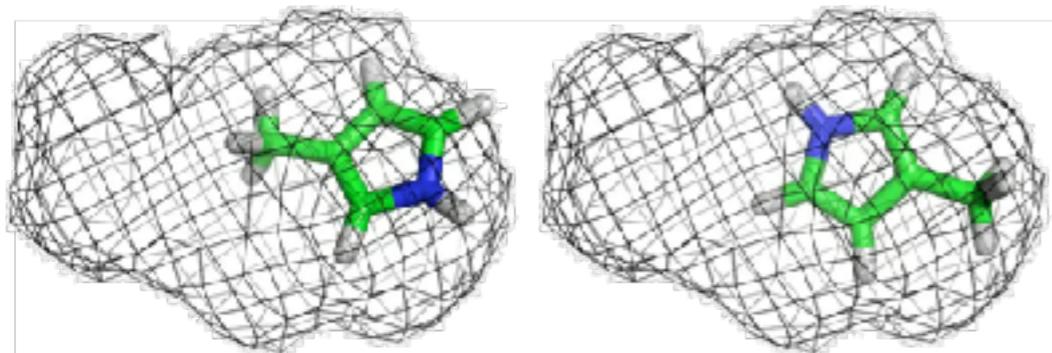


Conformational change is key

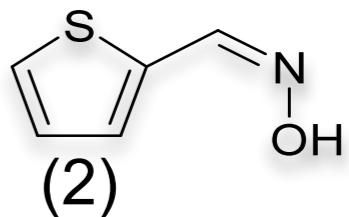
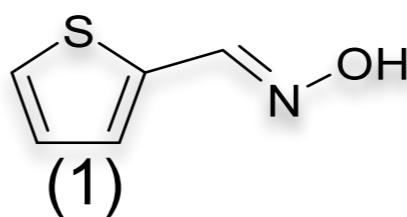


Charged binding sites now tractable

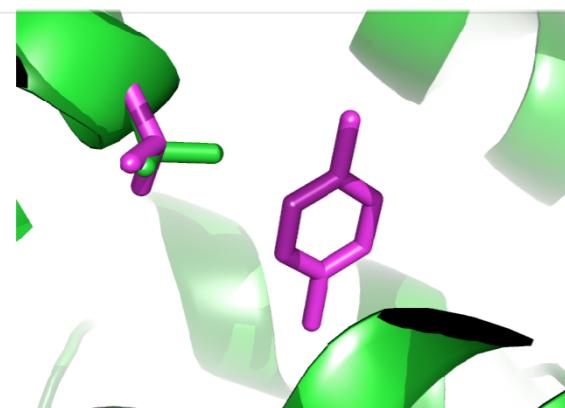
# What did we learn?



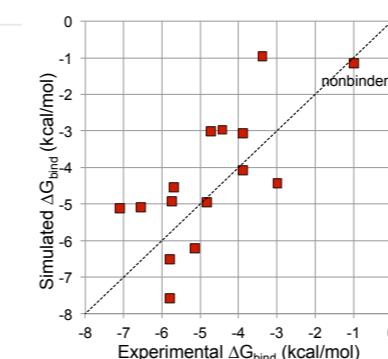
Consider multiple orientations



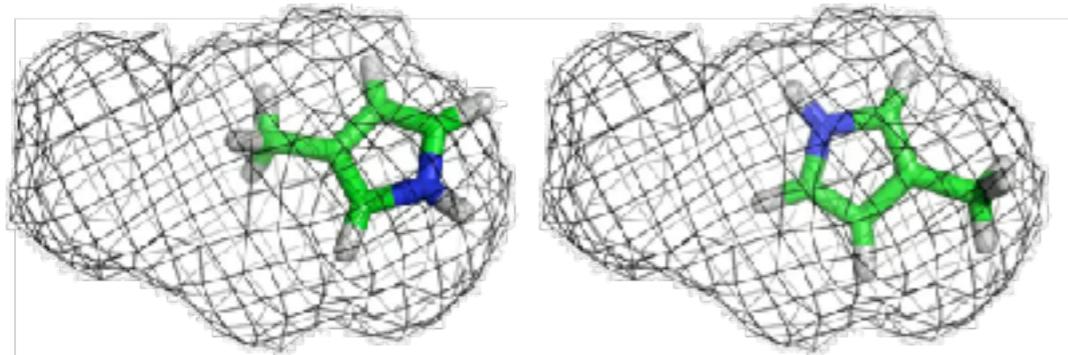
We can make nontrivial predictions



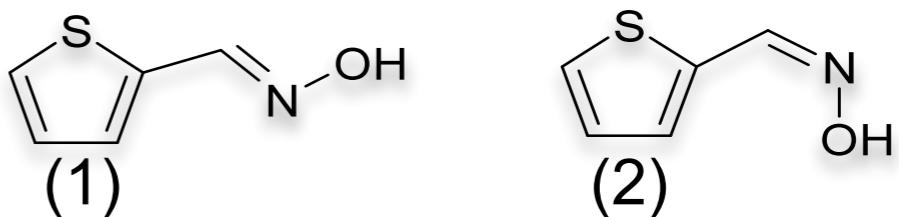
Conformational change is key



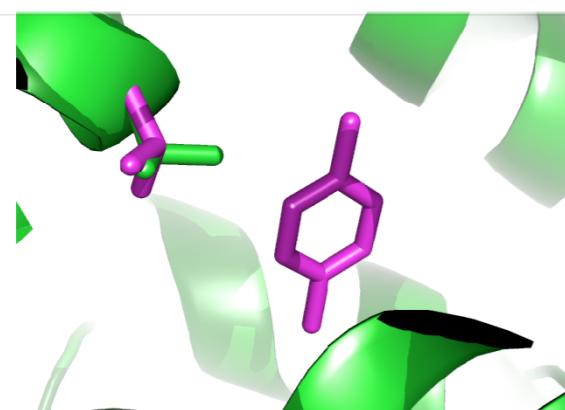
# What did we learn?



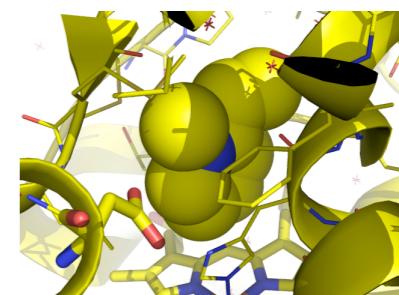
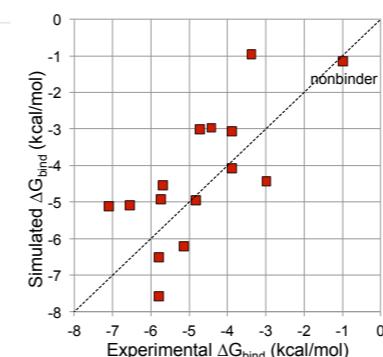
# Consider multiple orientations



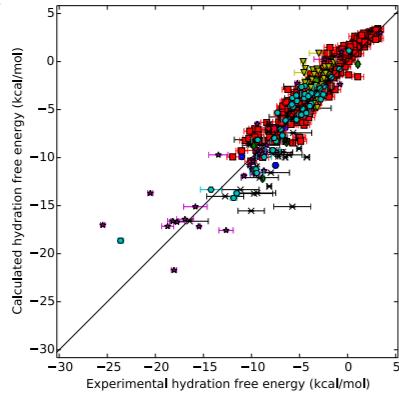
# We can make nontrivial predictions



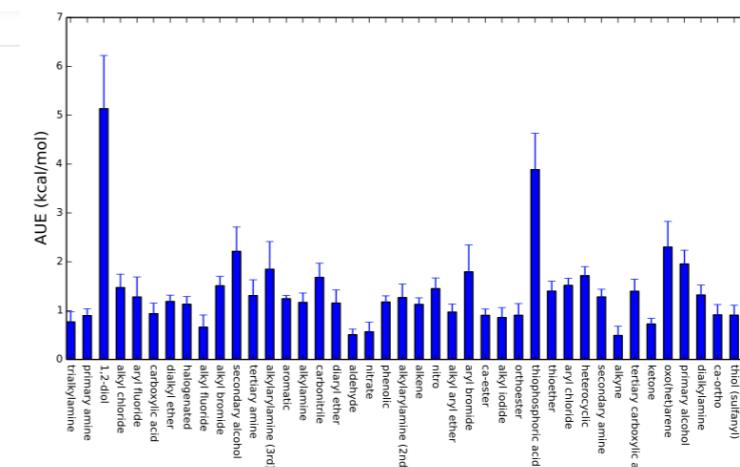
# Conformational change is key



## Charged binding sites now tractable



# Hydration free energies are accurate



# We need more data for FF testing

# 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:

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

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

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

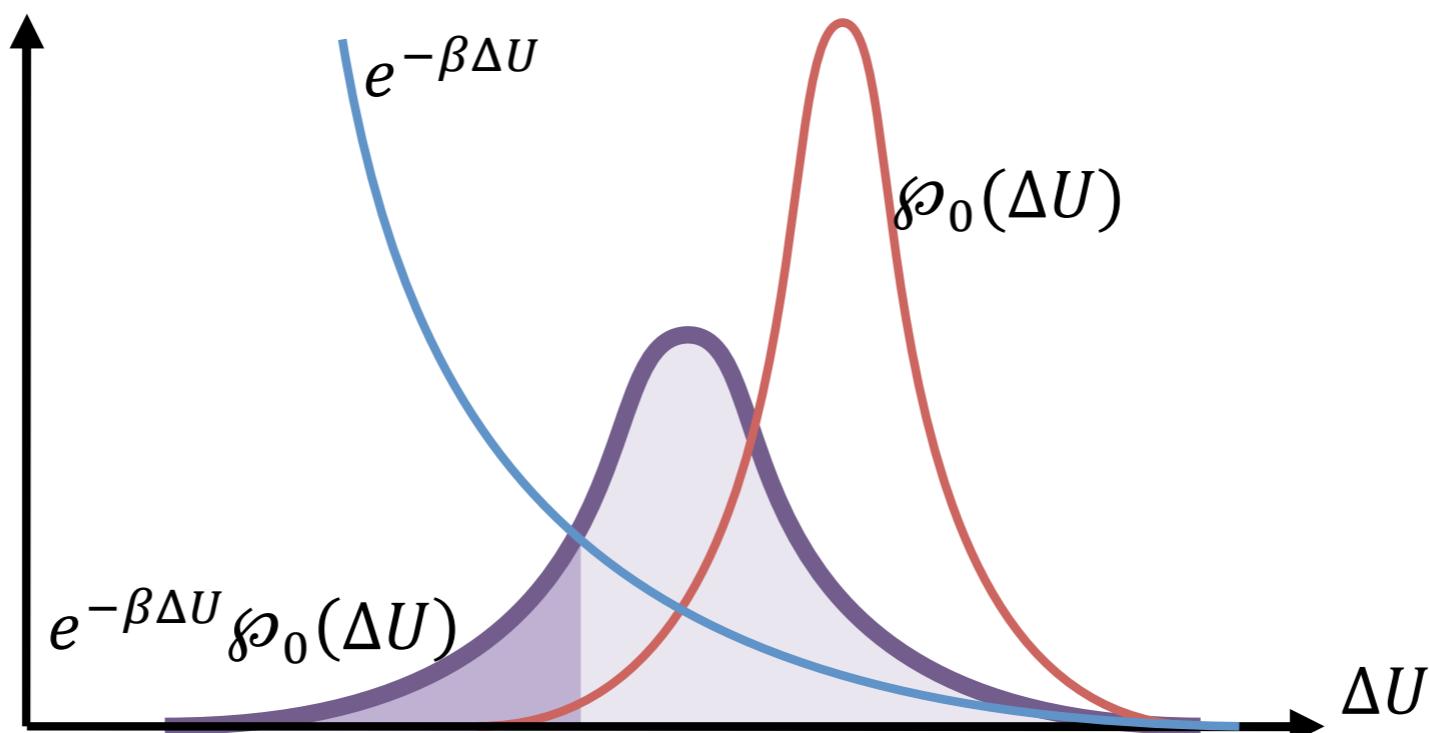
It is also true that  $\beta A_1 - \beta A_0 = \ln \langle e^{\beta \Delta U} \rangle_1$

- 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  $\Delta U$  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

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

- Bennett's approach modified the free energy

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

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
  - $\Delta U$  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:

$$U = (1 - \lambda)U_0 + \lambda U_1$$

- $\lambda$  runs from 0 to 1 and we run intermediate simulations
  - Compute pairwise free energy differences, and the total free energy difference is the sum

# 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

# 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  $\mu_{ig}$  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

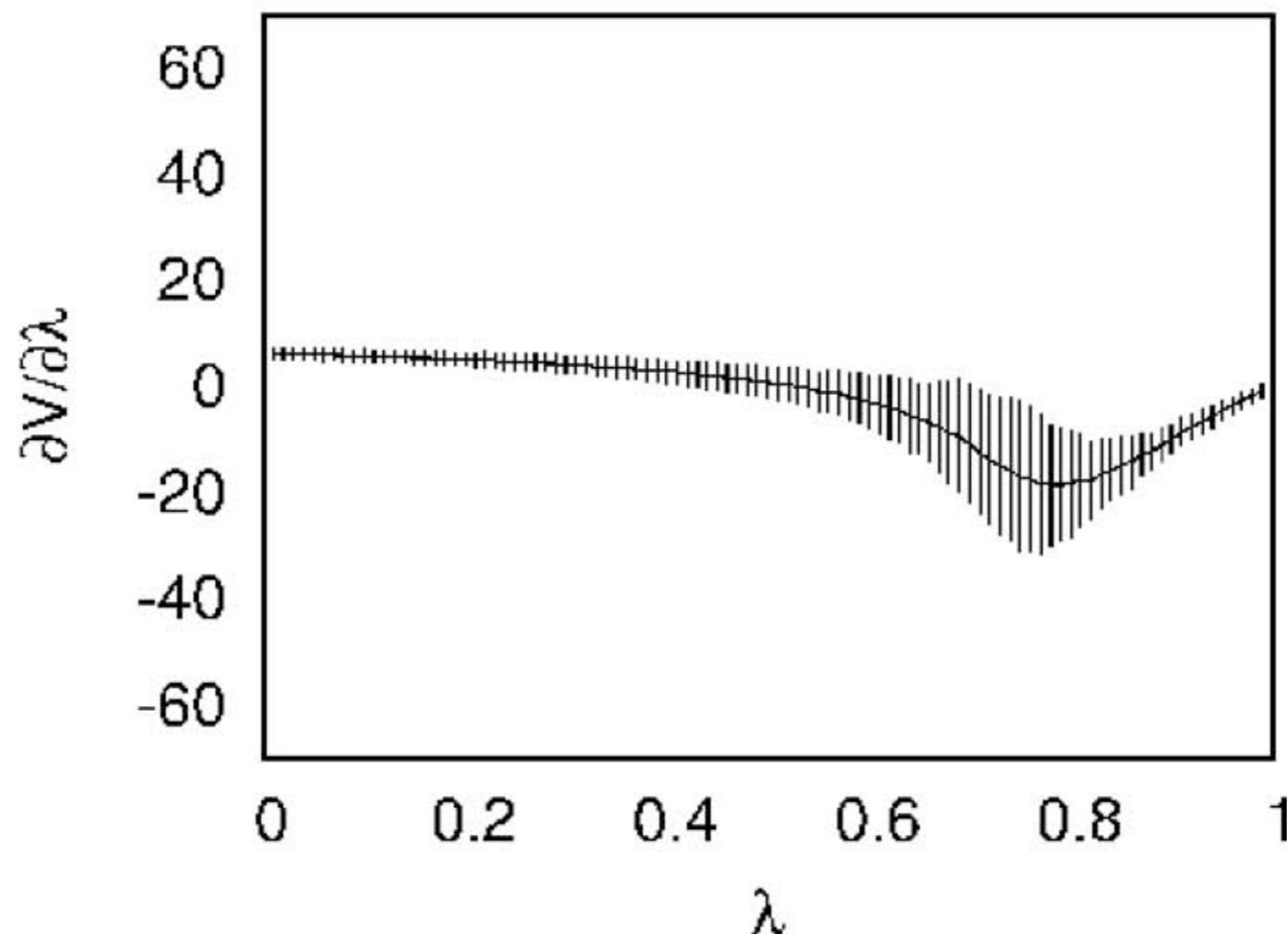
# Thermodynamic integration is an alternate approach to free energy calculations

- 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  $\lambda$ :

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

- 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

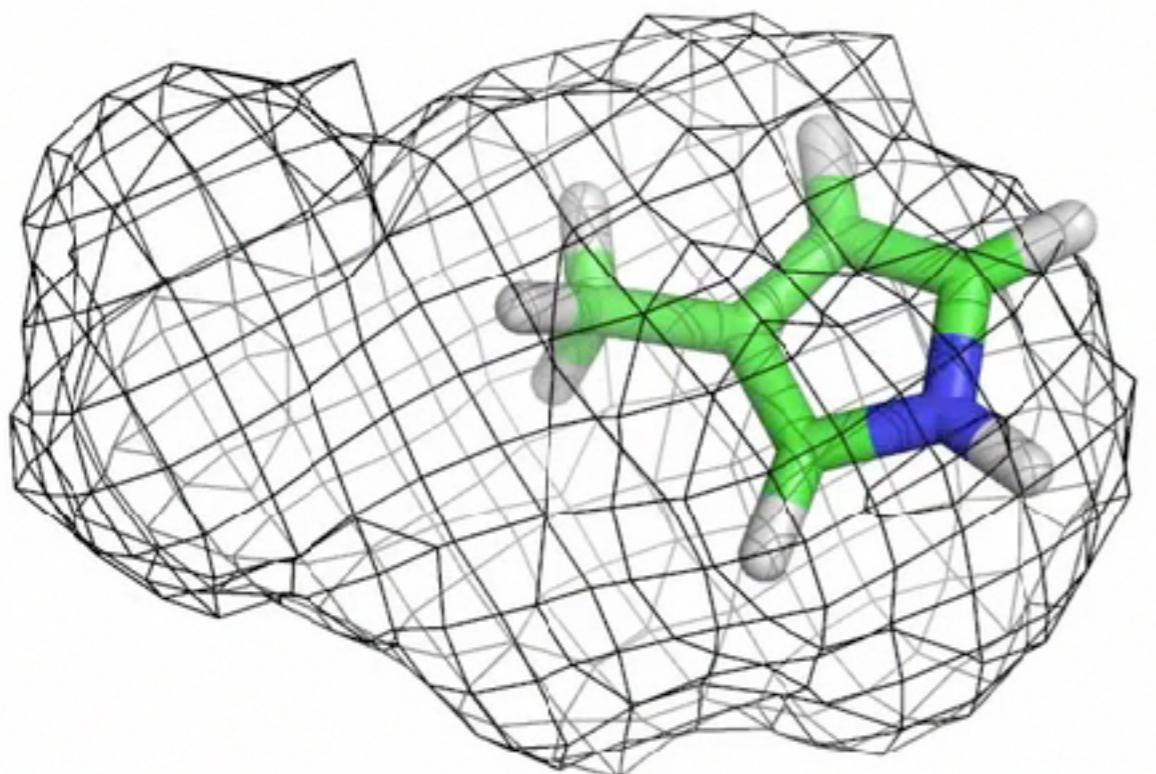
- It is often not possible to calculate the average enthalpy of a reasonably large (protein+water) system with  $\sim$ kcal/mol accuracy
- Yet free energy calculations can get to higher accuracy than enthalpy differences
  - This is because fluctuations in the system which lead to noise in the potential 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

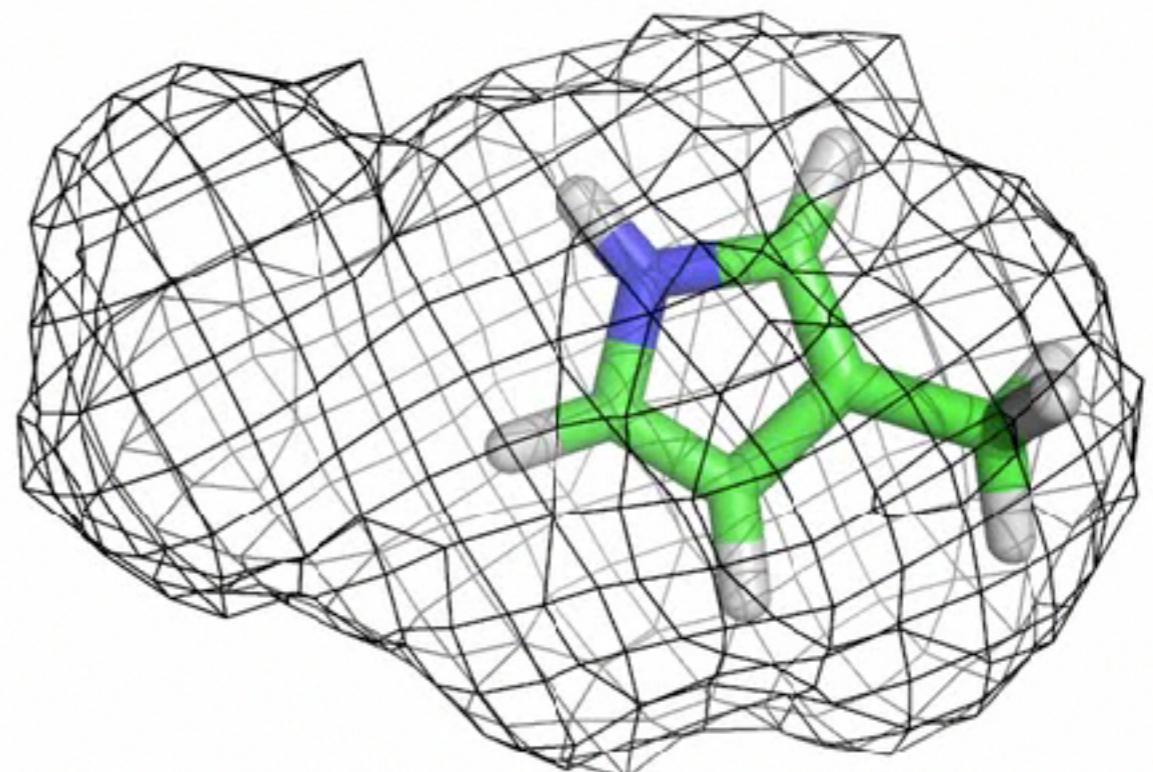
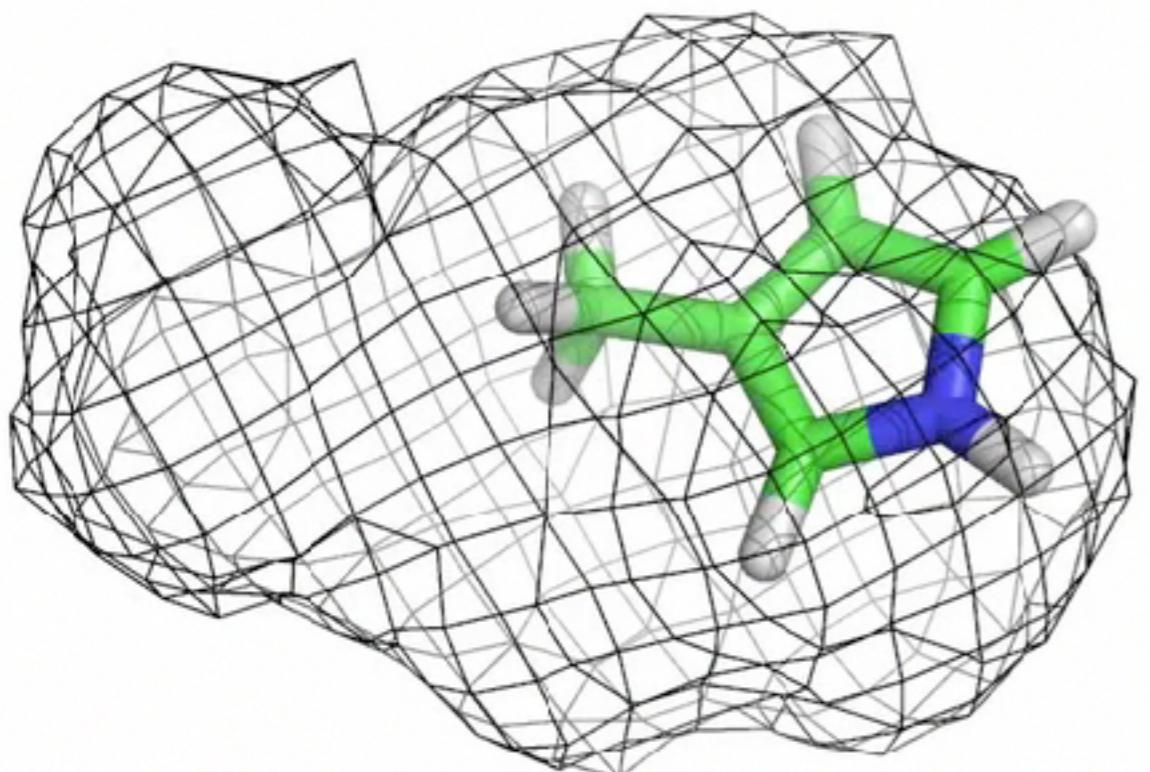
- We cannot typically directly calculate the absolute free energy of a system
  - But we can connect the system to reference states for which the free energy can be calculated analytically
    - For a liquid/gas, can transform the system to an ideal gas reference state
      - Compute the free energy 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

Back to research problems: We still have this issue about slow binding mode sampling

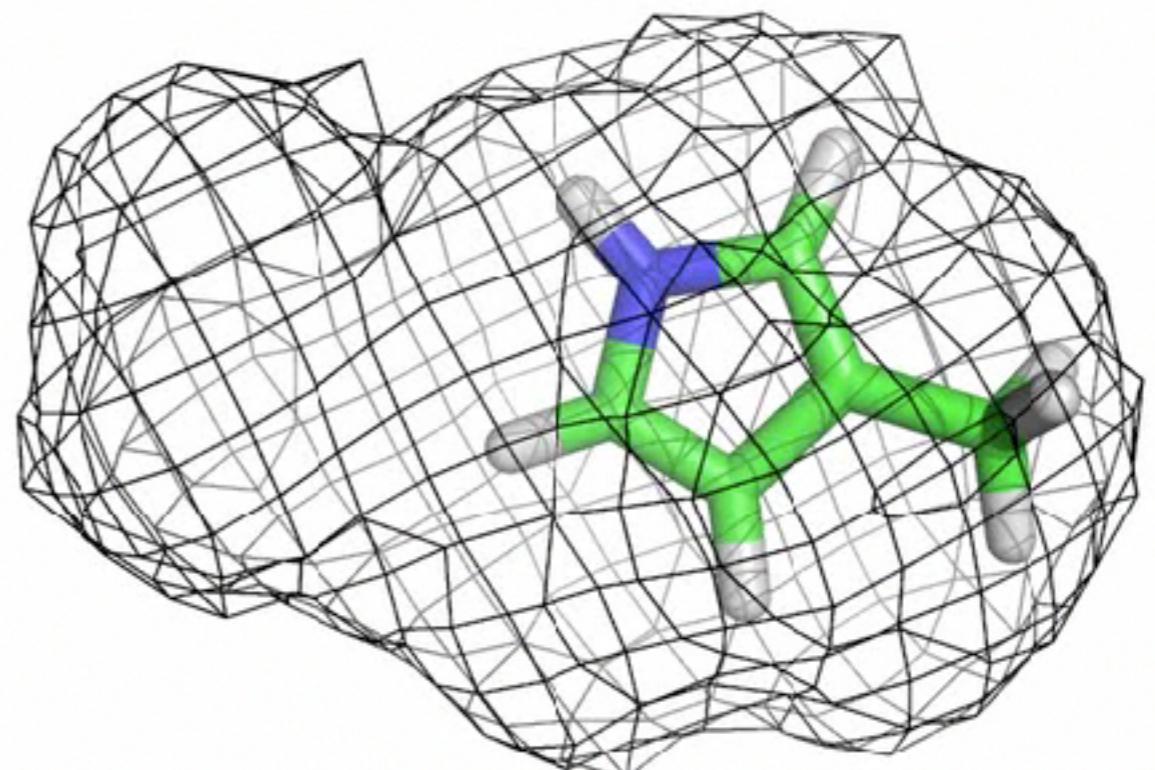
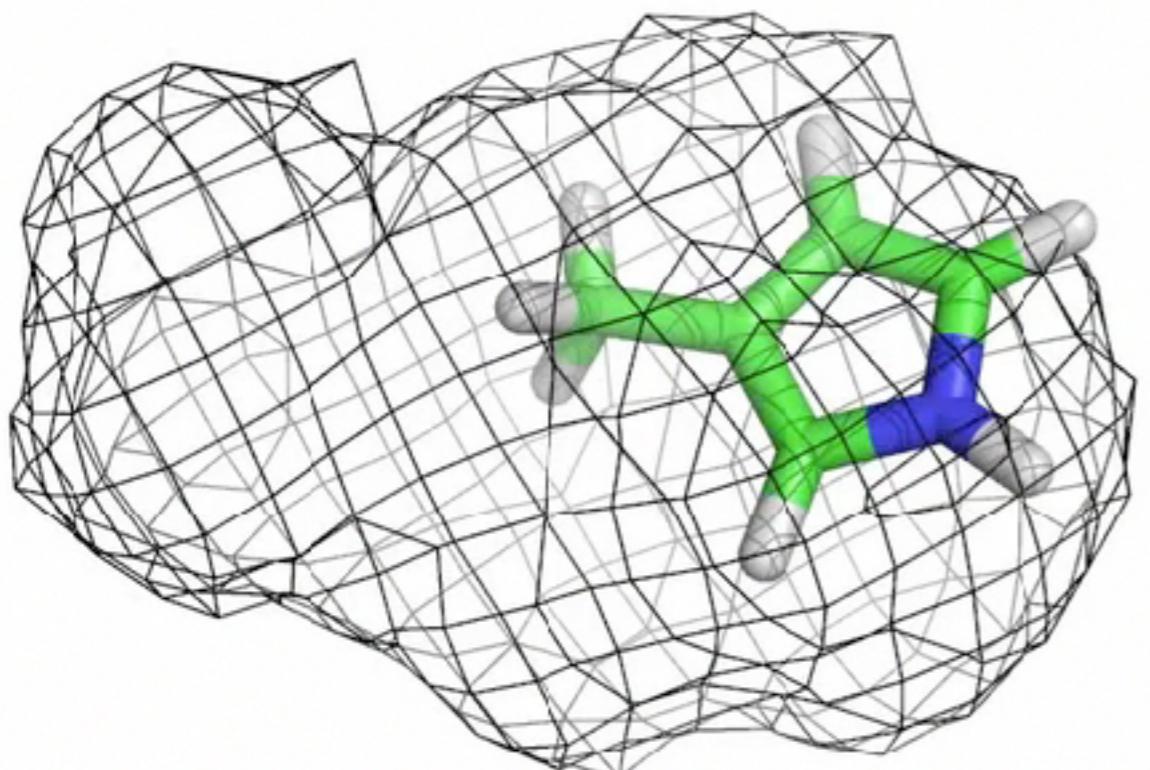
Back to research problems: We still have this issue about slow binding mode sampling



Back to research problems: We still have this issue about slow binding mode sampling

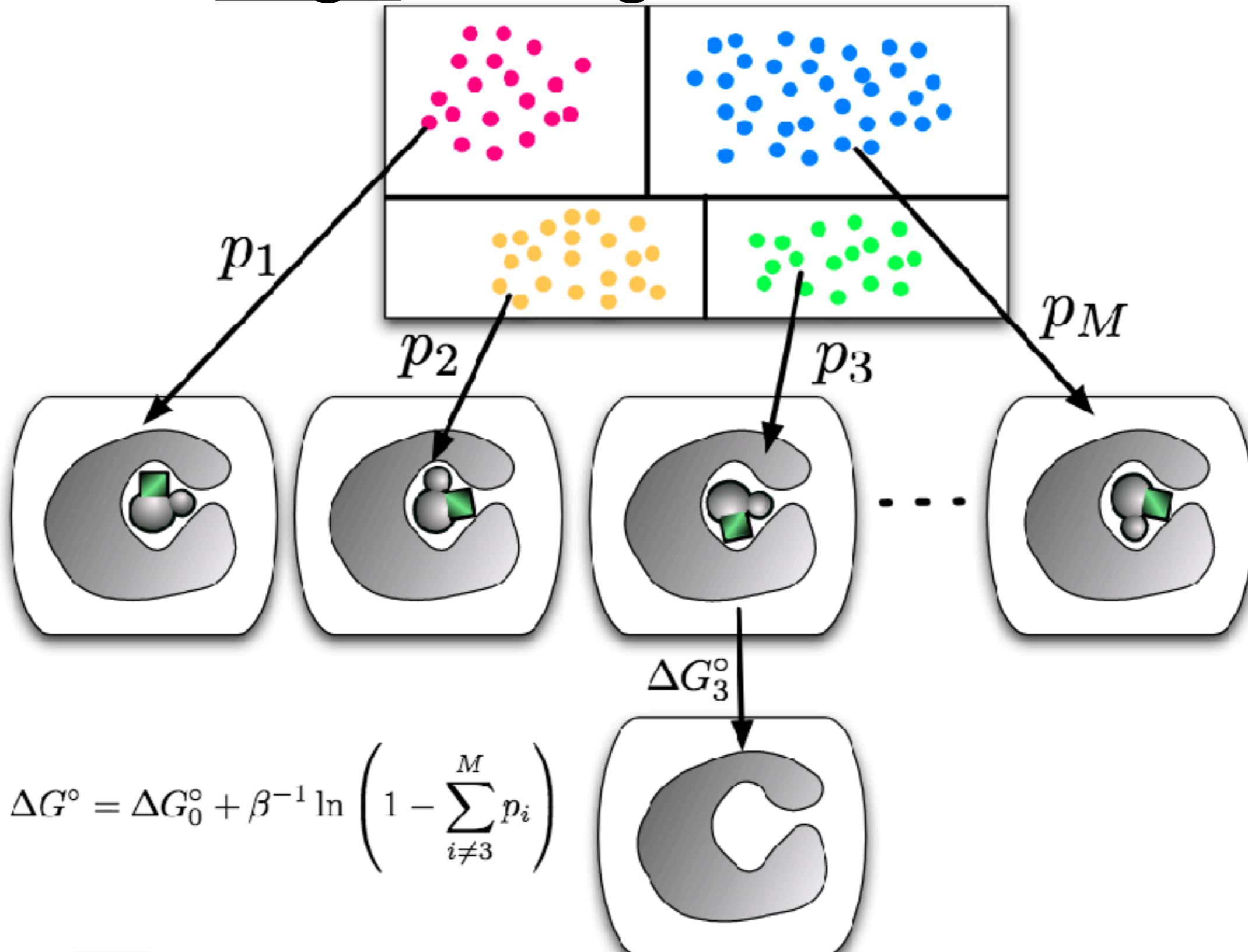


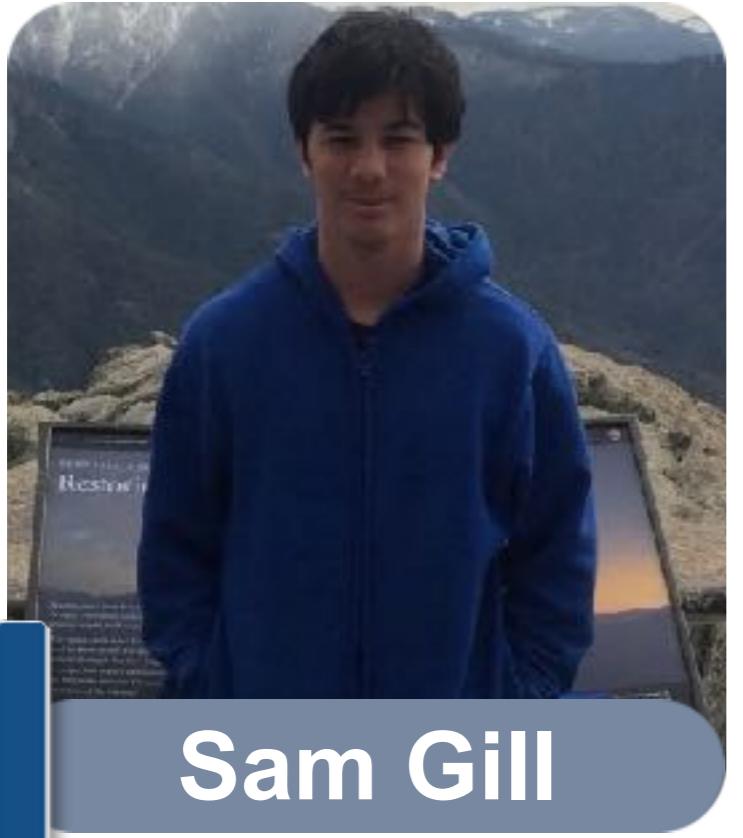
Back to research problems: We still have this issue about slow binding mode sampling



$$\Delta G^o = -k_B T \ln \left( e^{-\frac{\Delta G_1^o}{k_B T}} + e^{-\frac{\Delta G_2^o}{k_B T}} \right)$$

# Full binding free energy can be obtained from the populations explored with BLUES and a single binding calculation





Sam Gill



**BLUES**

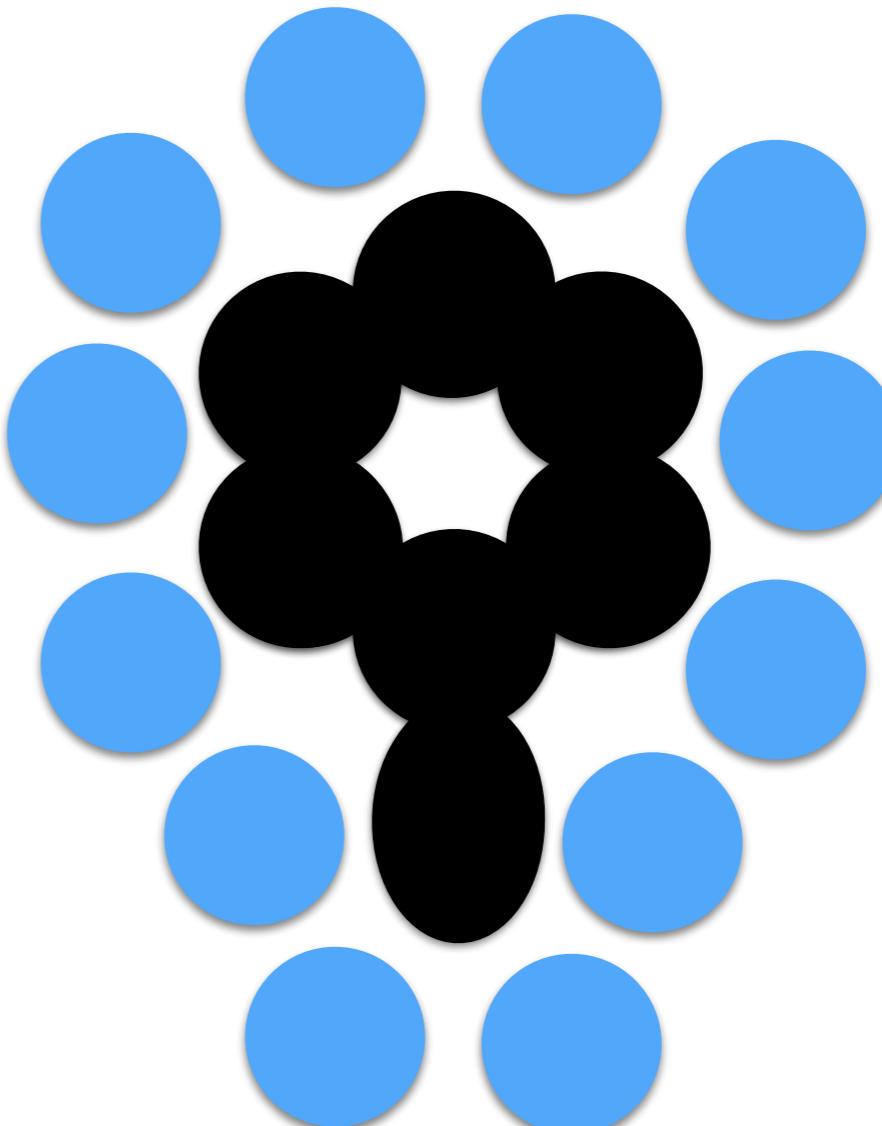
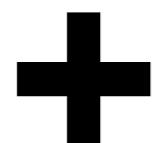
*Binding modes of  
Ligands  
Using  
Enhanced  
Sampling*

[doi: 10.26434/chemrxiv.5406907.v2](https://doi.org/10.26434/chemrxiv.5406907.v2)

**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

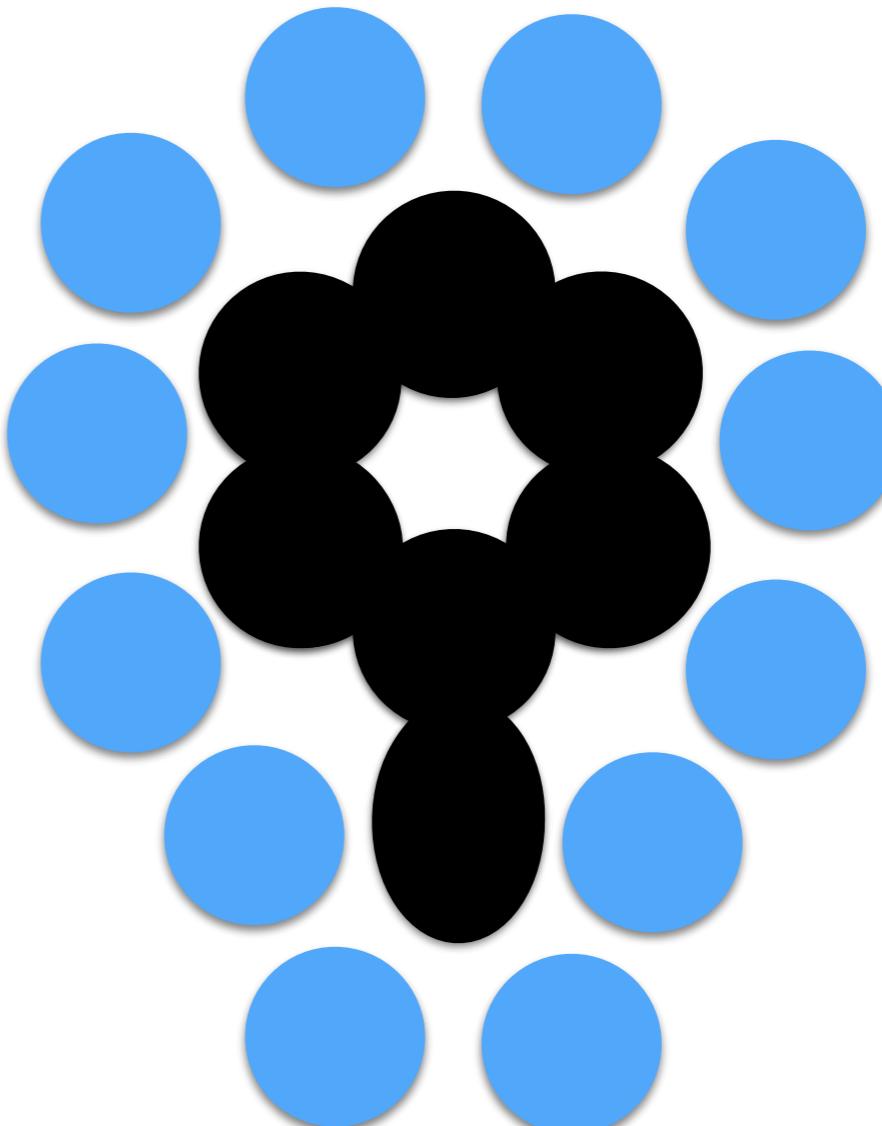


**BLUES**

Nonequilibrium  
Candidate  
Monte Carlo  
(NCMC)



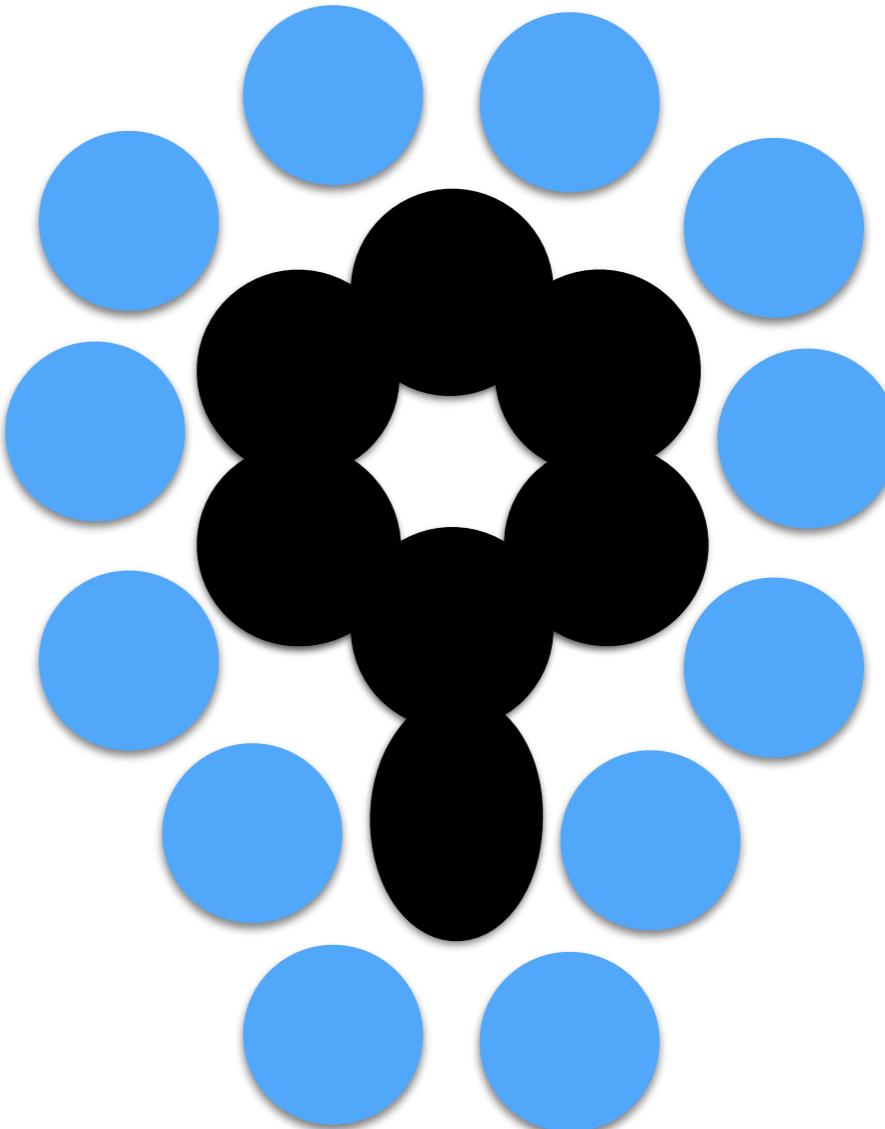
Molecular  
Dynamics (MD)



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



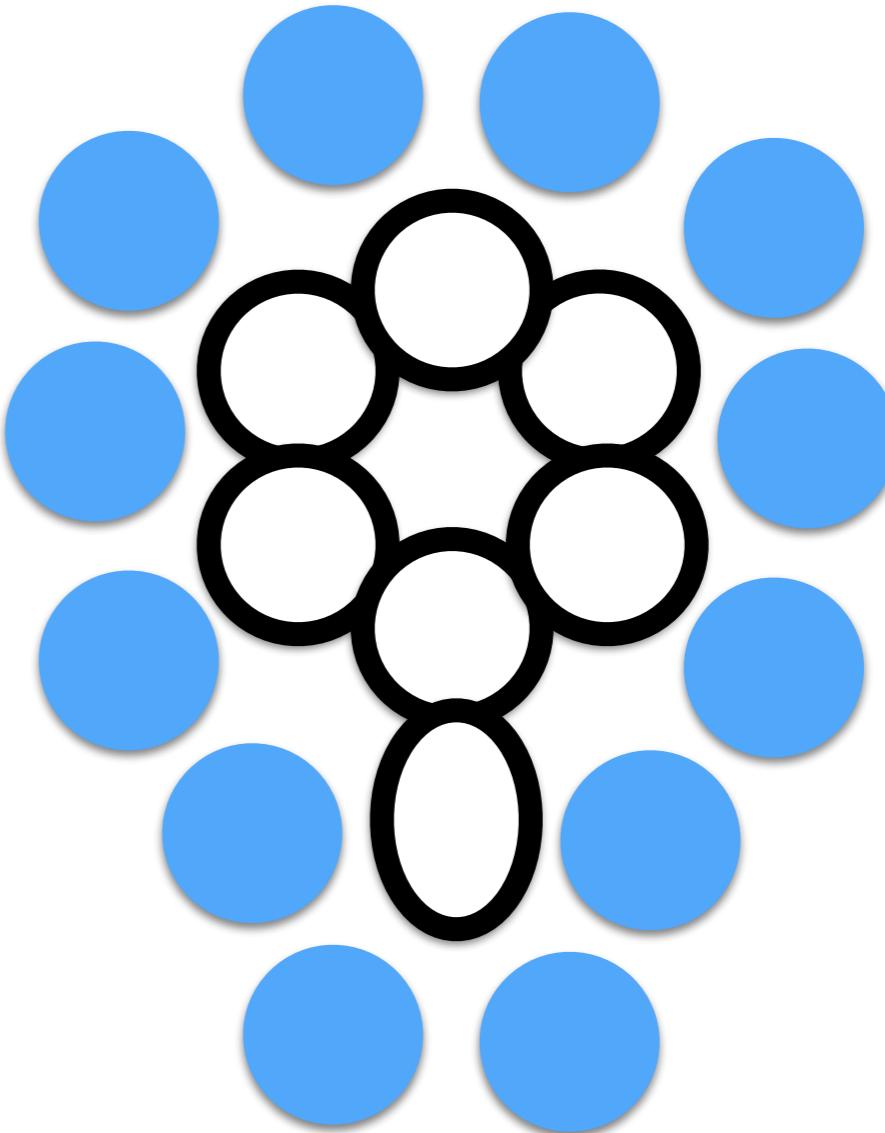
**BLUES**



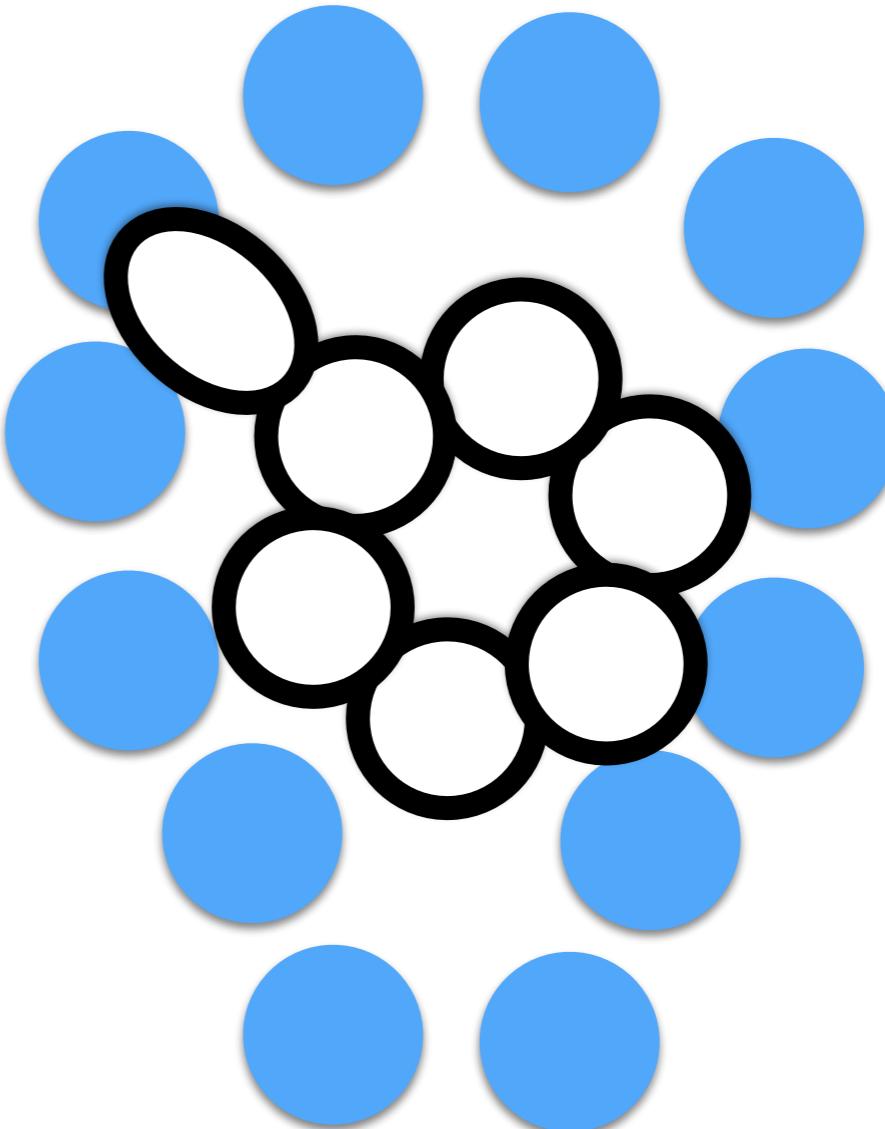
***Nonequilibrium Candidate Monte Carlo (NCMC)***  
allows us to propose a *random rotational* move to the ligand



**BLUES**



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

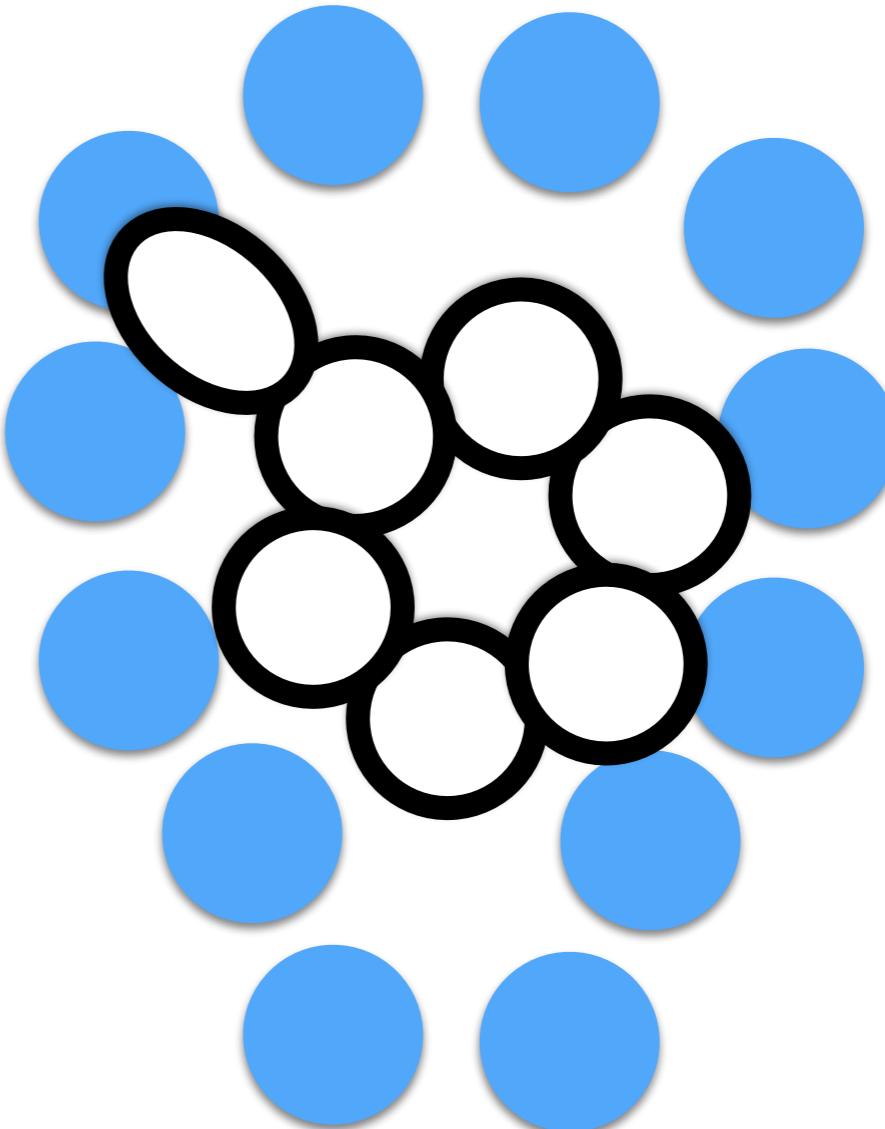


**Nonequilibrium Candidate Monte Carlo (NCMC)**  
allows us to propose a random rotational move to the ligand



## BLUES

- + Random rotation may identify new possible binding modes

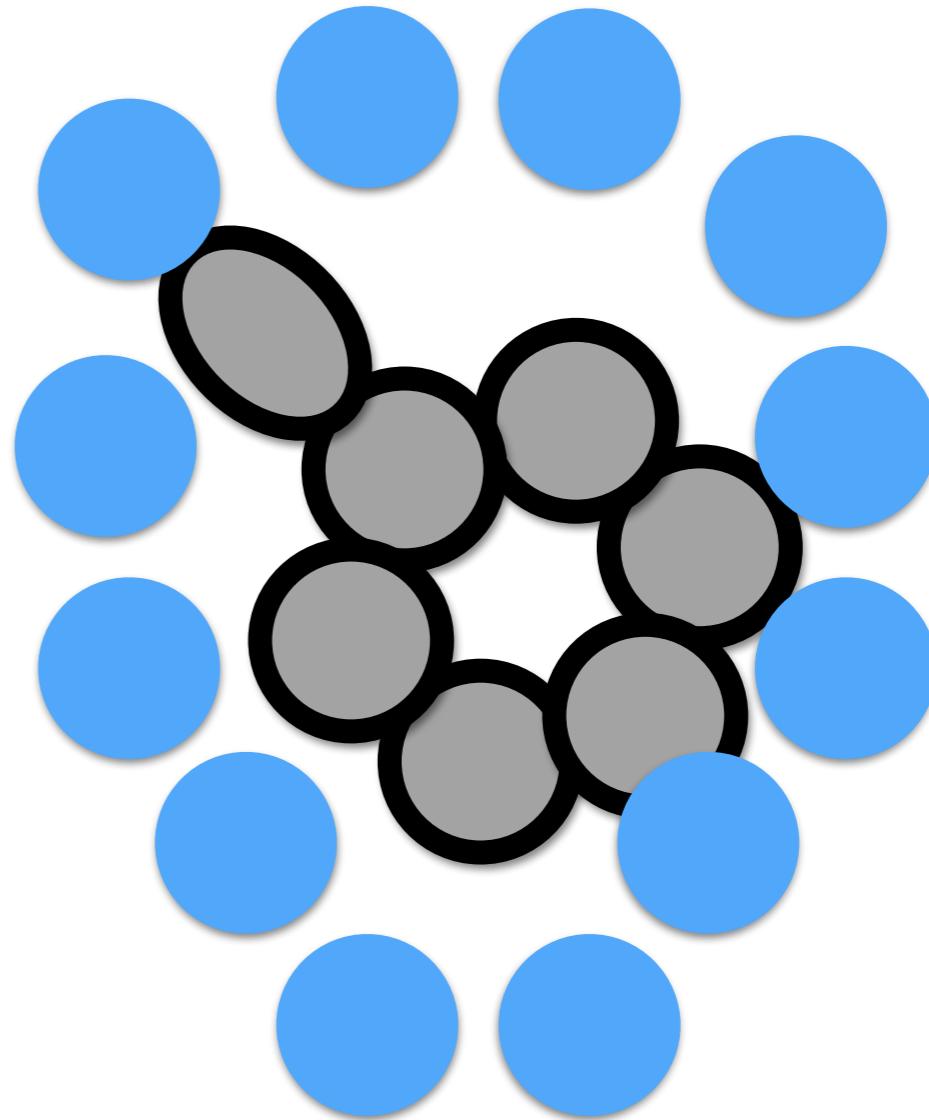


# NCMC move is divided into smaller steps to help resolve potential clashes



## BLUES

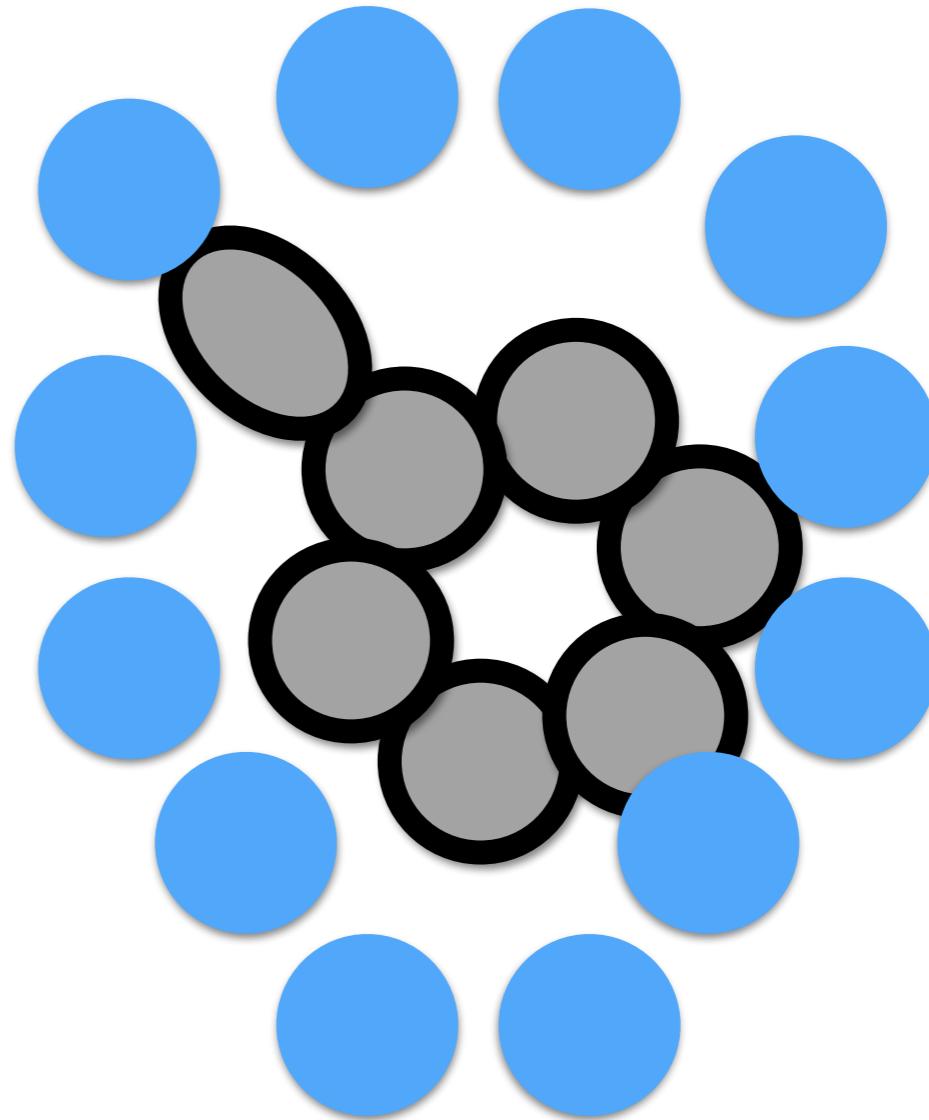
- + *Random rotation may identify new possible binding modes*



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



+ *Random rotation may identify new possible binding modes*



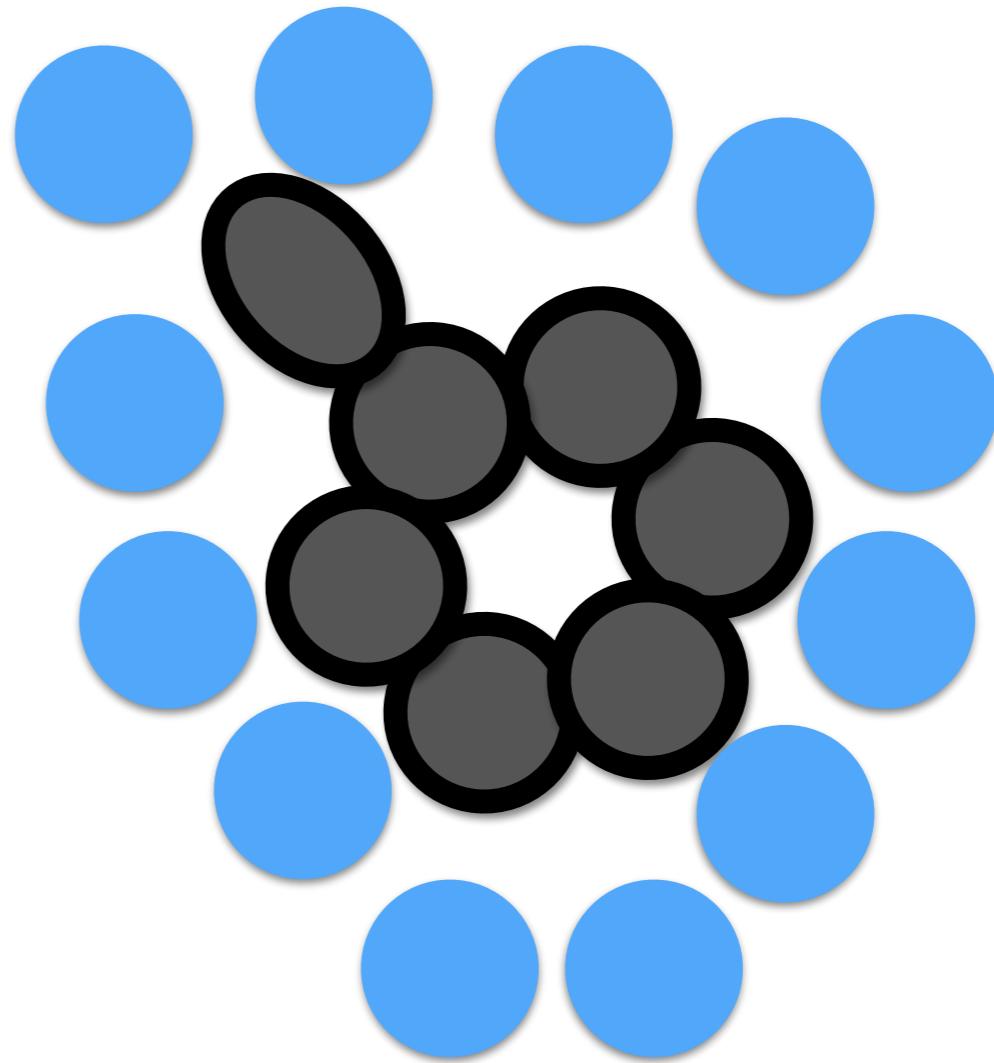
- Change ligand interactions
- Relaxation

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



## BLUES

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



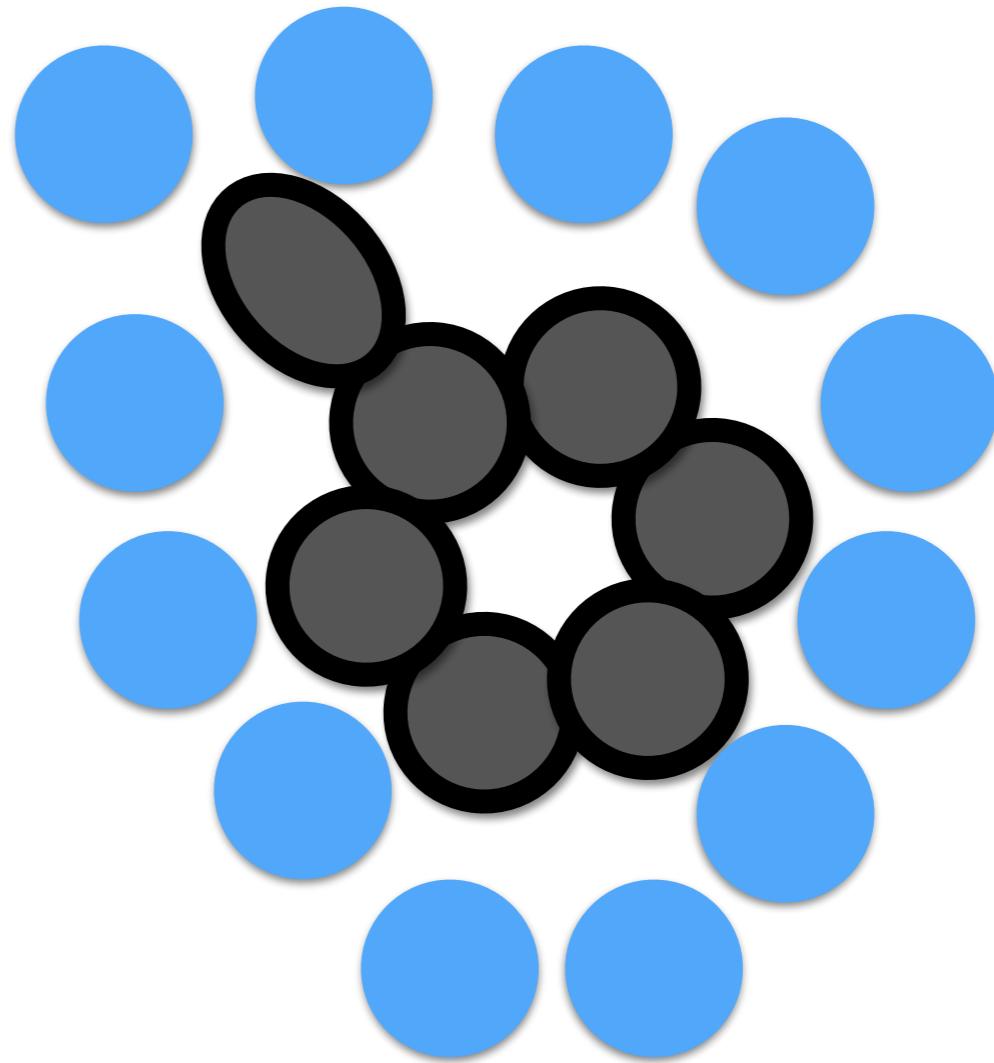
- Change ligand interactions
- Relaxation

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



## BLUES

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



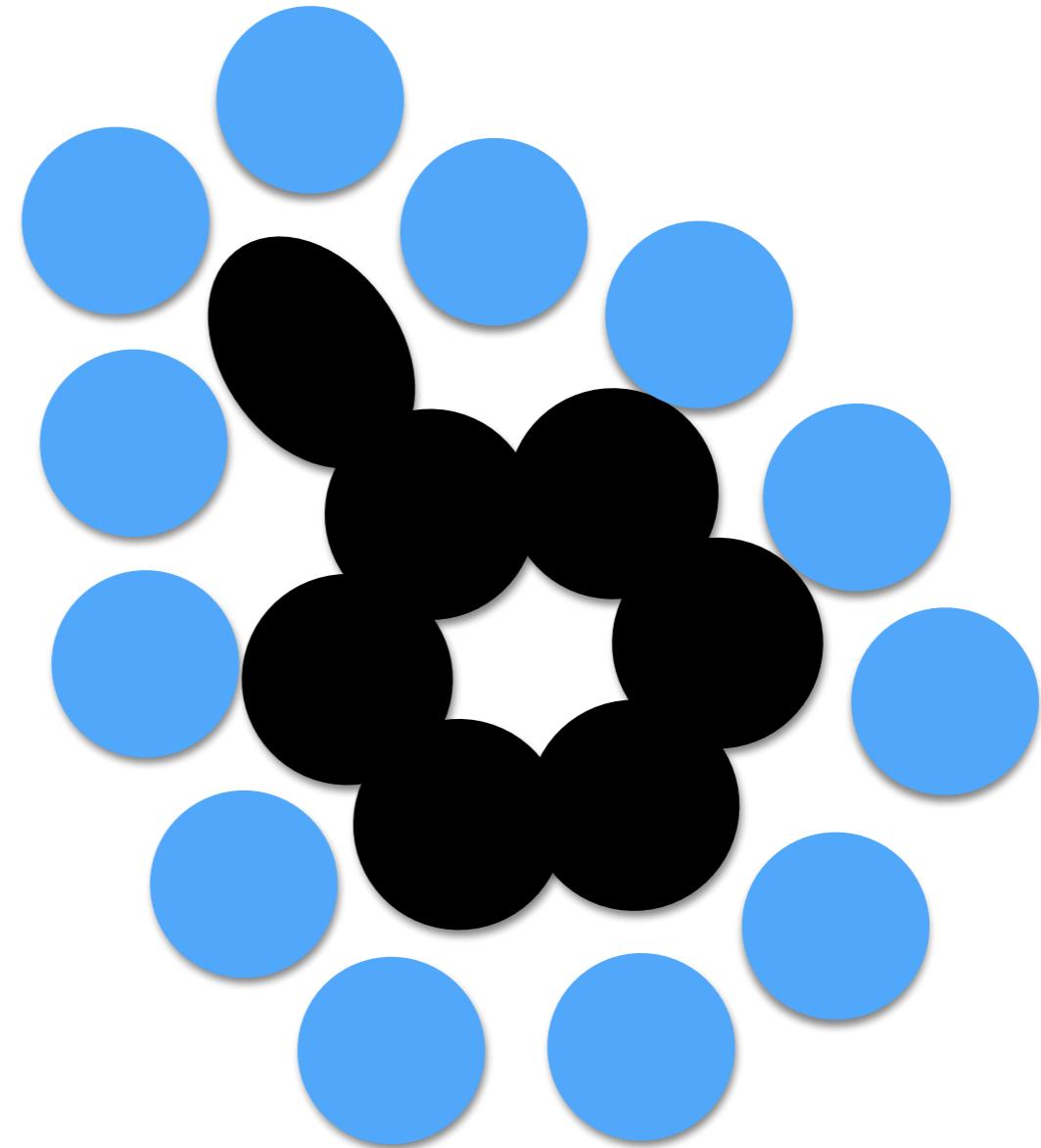
- Change ligand interactions
- Relaxation

# 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*

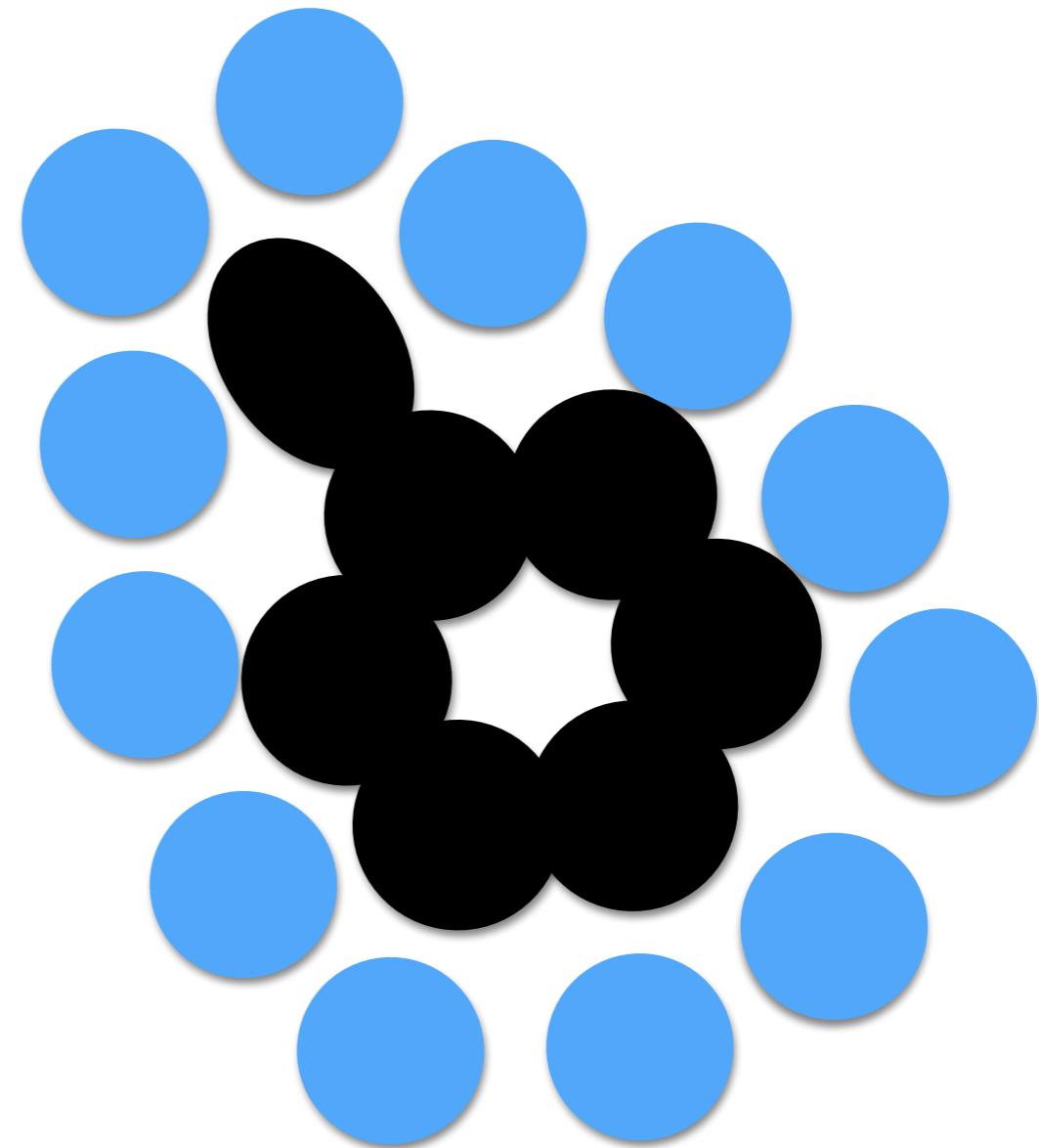


# NCMC move is accepted/rejected based upon the total work done

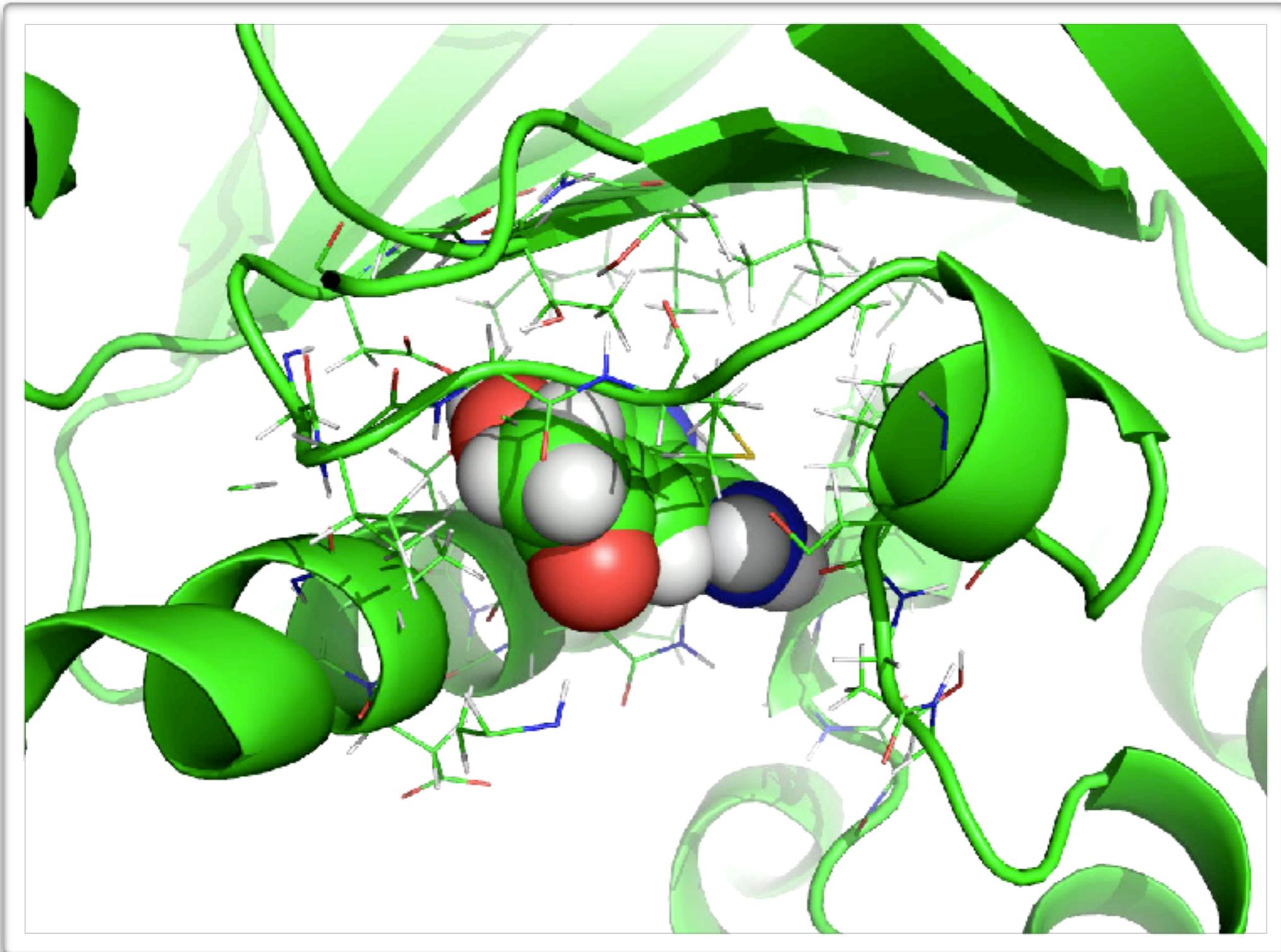


## BLUES

- + *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.**

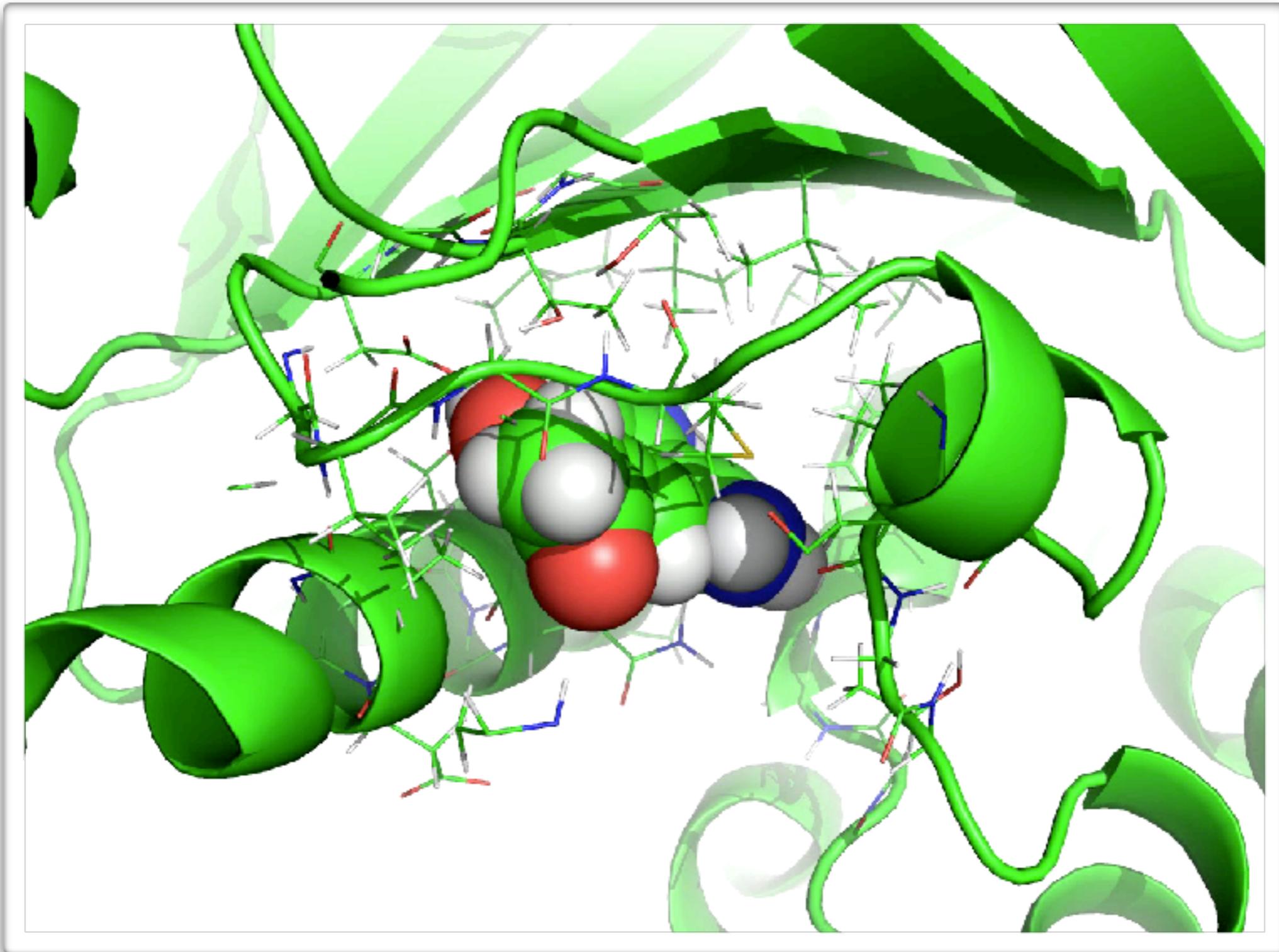


# Here is what BLUES looks like:



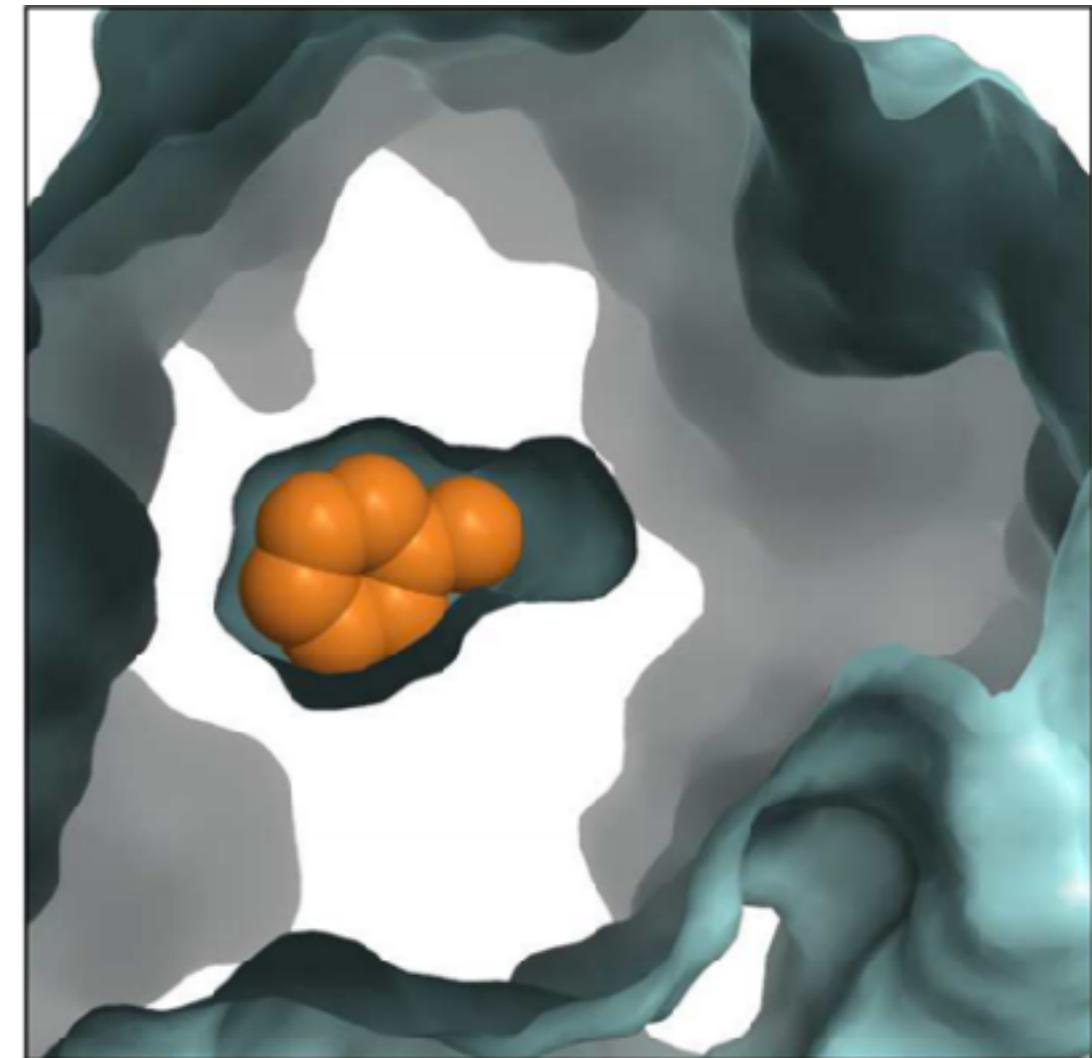
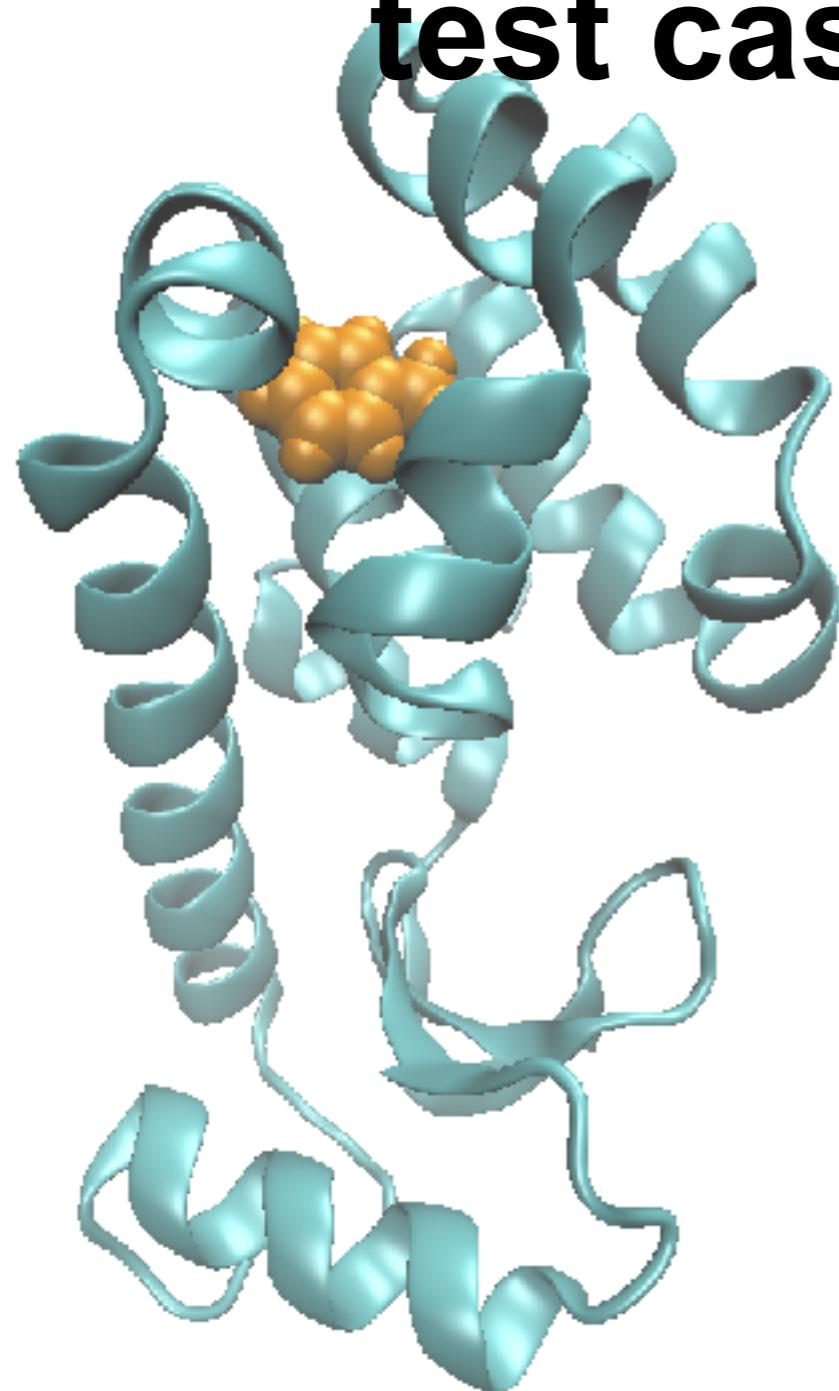
**Scale ligand off, random rotation, scale ligand on**

# Here is what BLUES looks like:

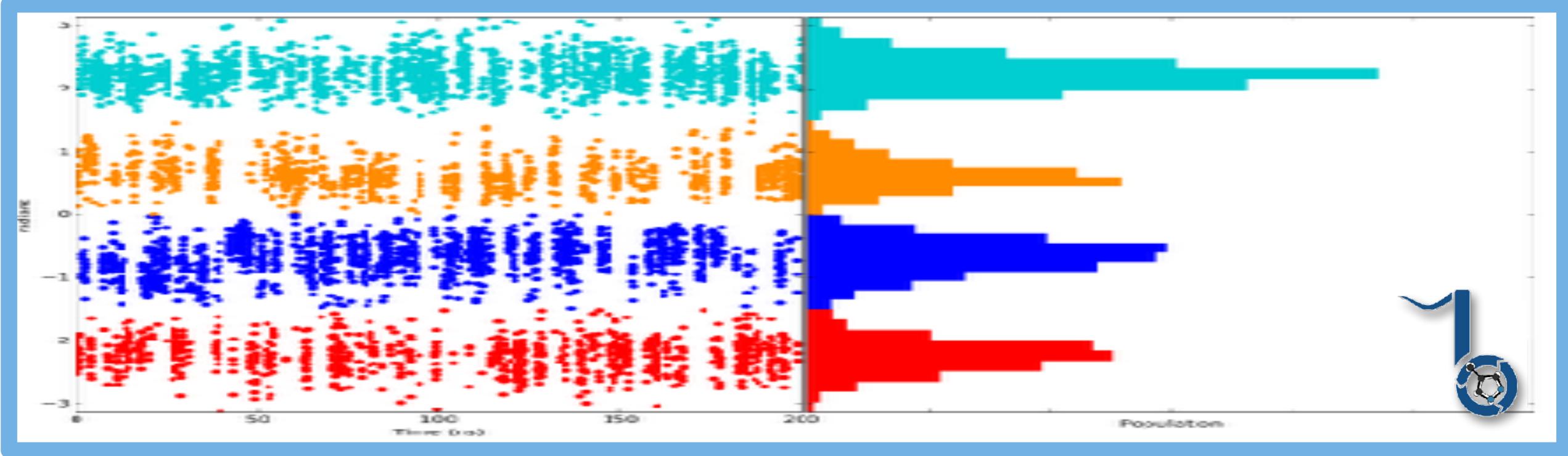
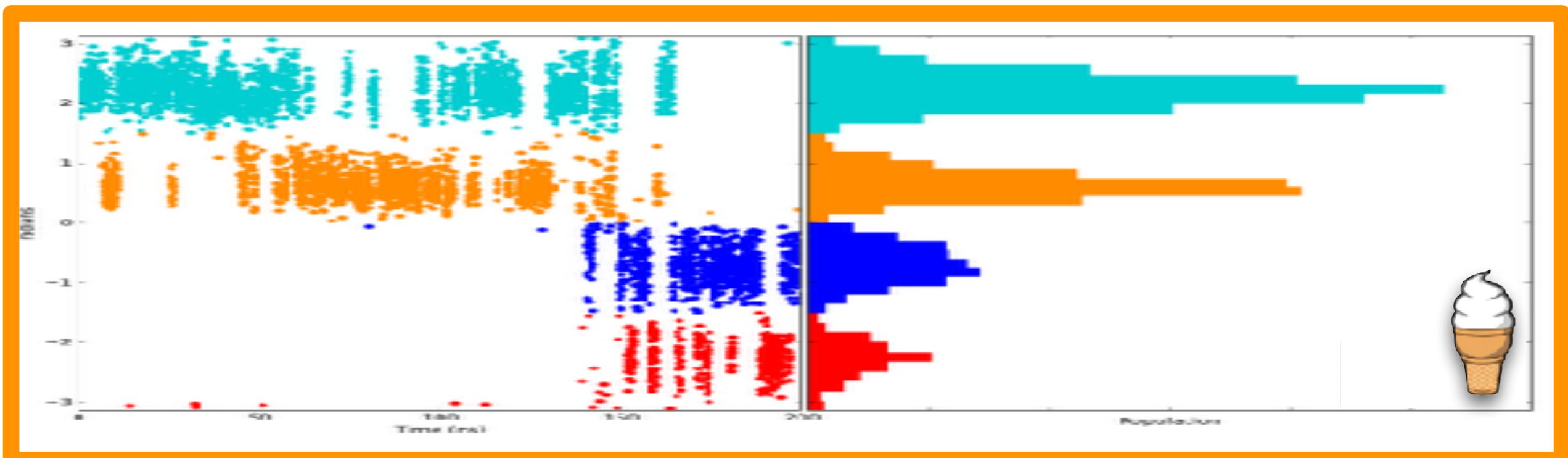


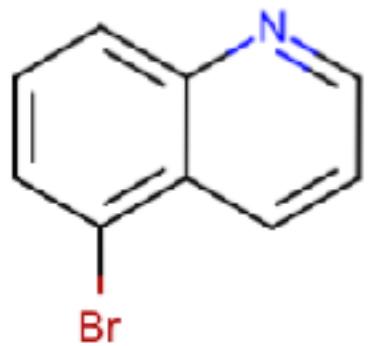
**Scale ligand off, random rotation, scale ligand on**

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



# BLUES rapidly transitions between binding modes, producing the correct populations

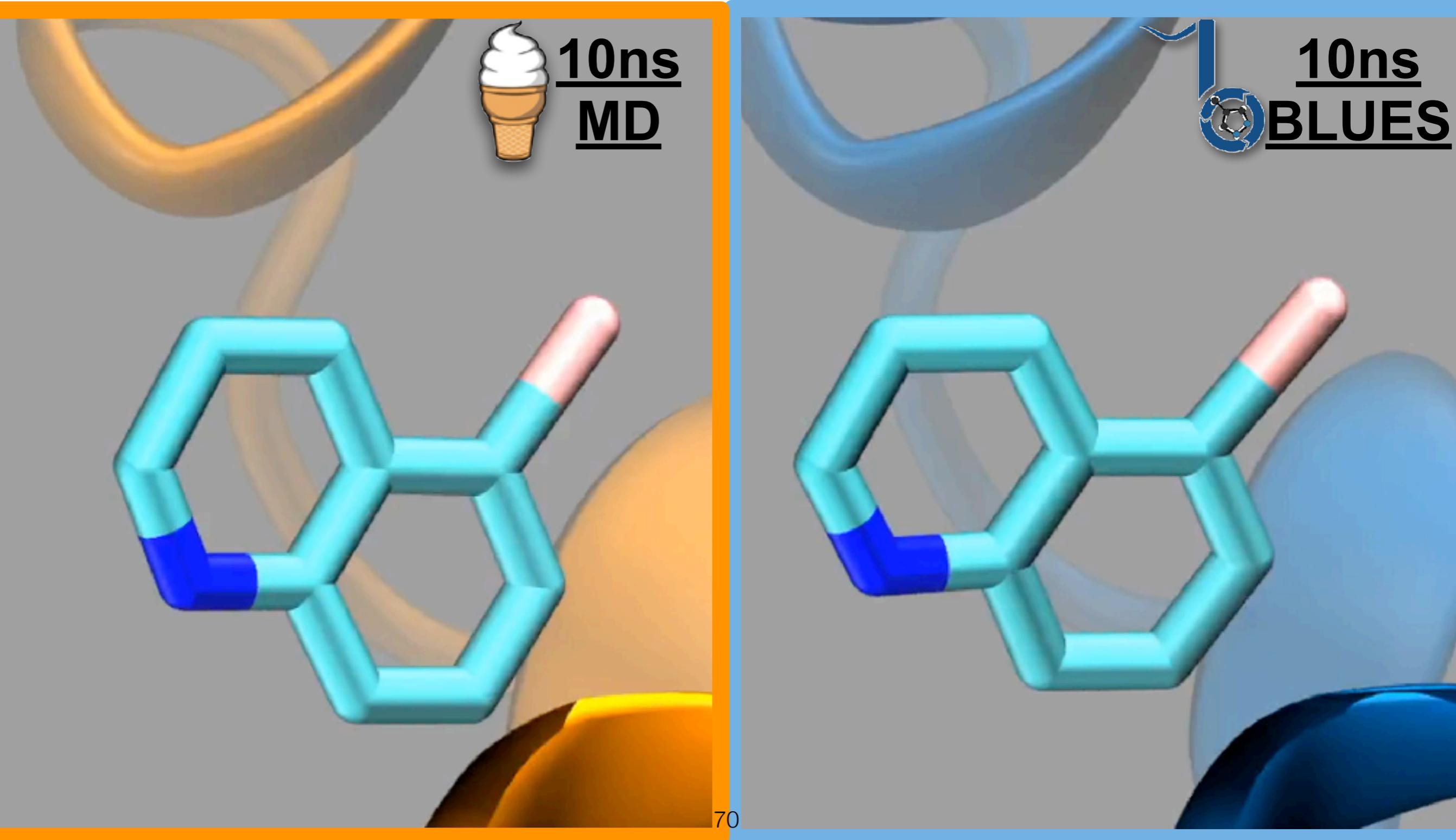


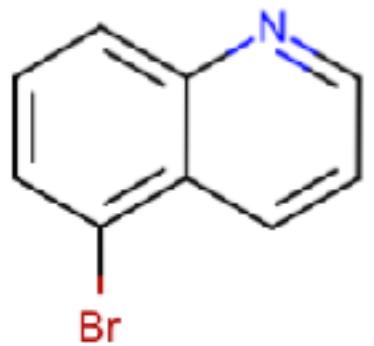


# BLUES shows accelerated sampling over “vanilla” MD simulations



Danielle Bergazin

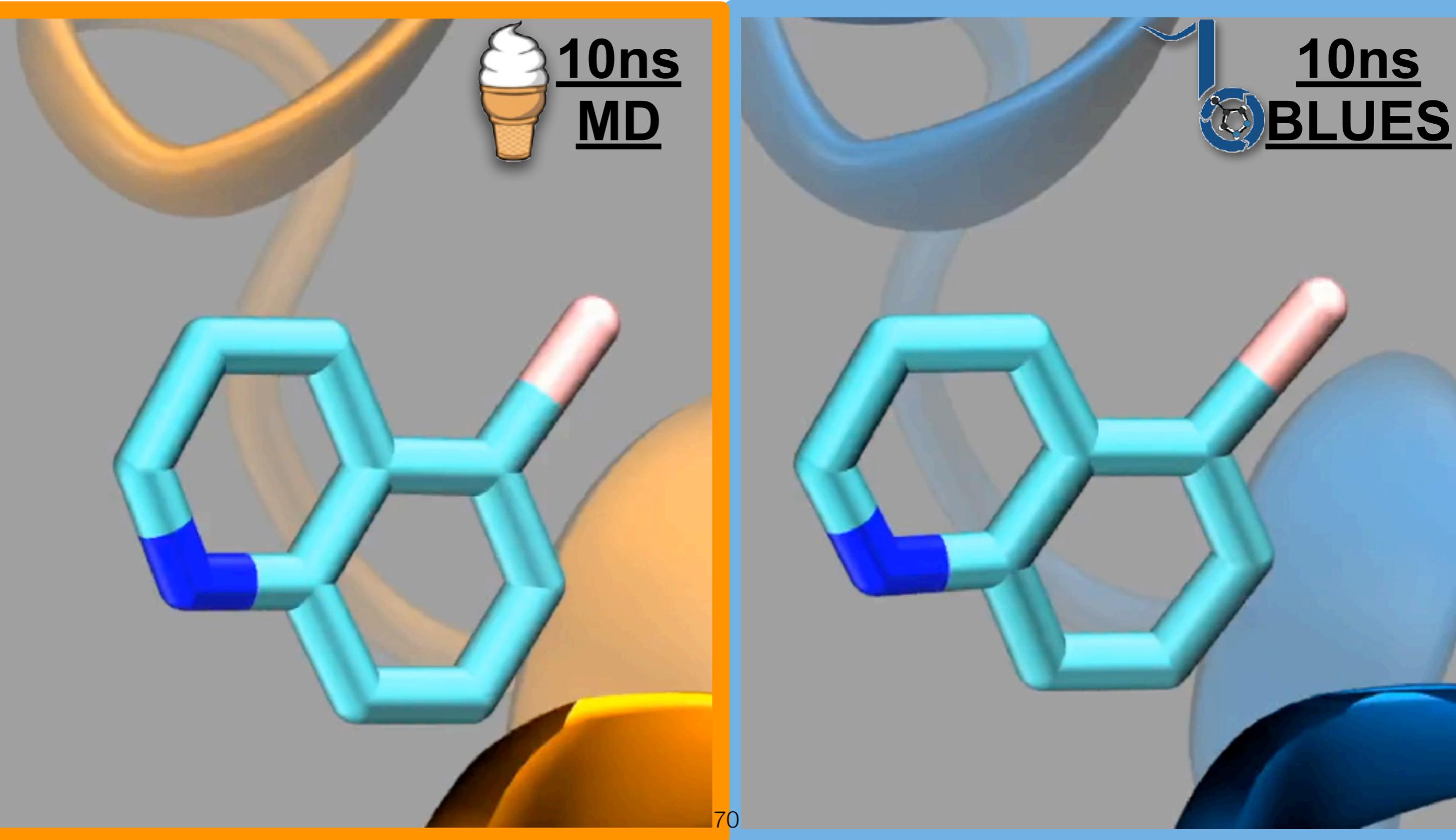




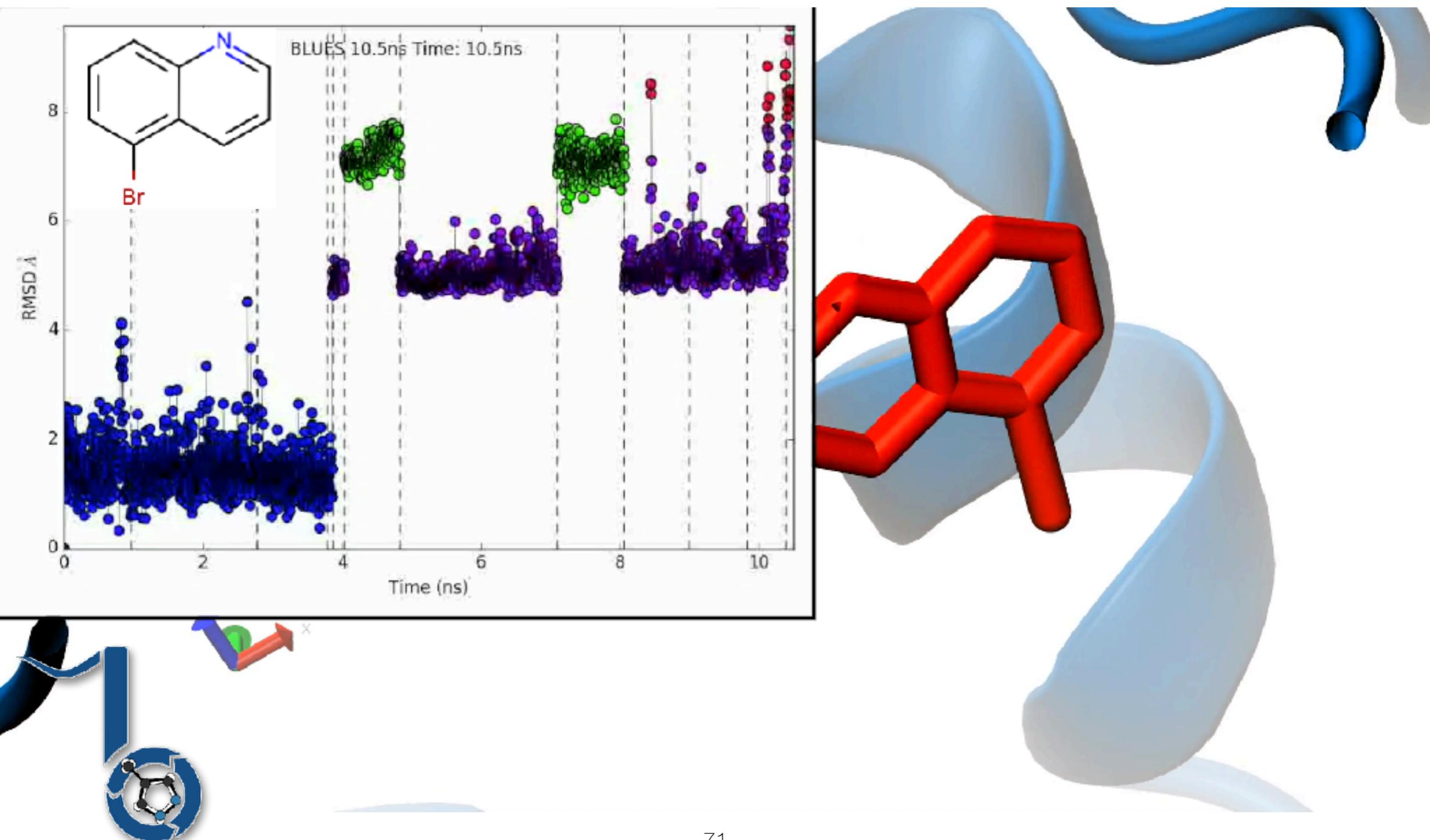
# BLUES shows accelerated sampling over “vanilla” MD simulations



Danielle Bergazin



# We're working on a robust approach to automatically identify the ligand binding modes



# We're working on a robust approach to automatically identify the ligand binding modes

