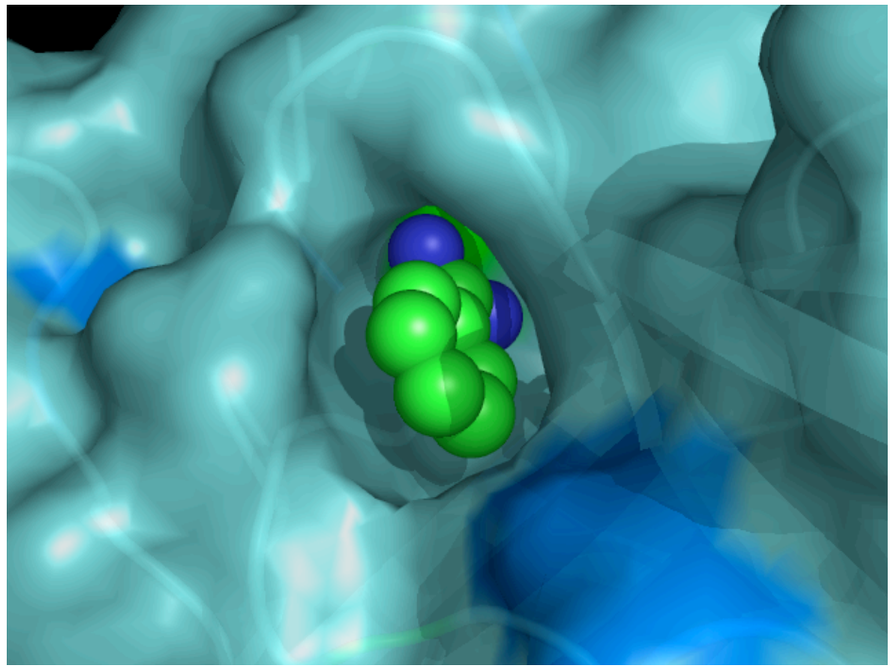
Using molecular simulations to predict protein-ligand binding and solvation

David Mobley

UCSF
University of California
San Francisco

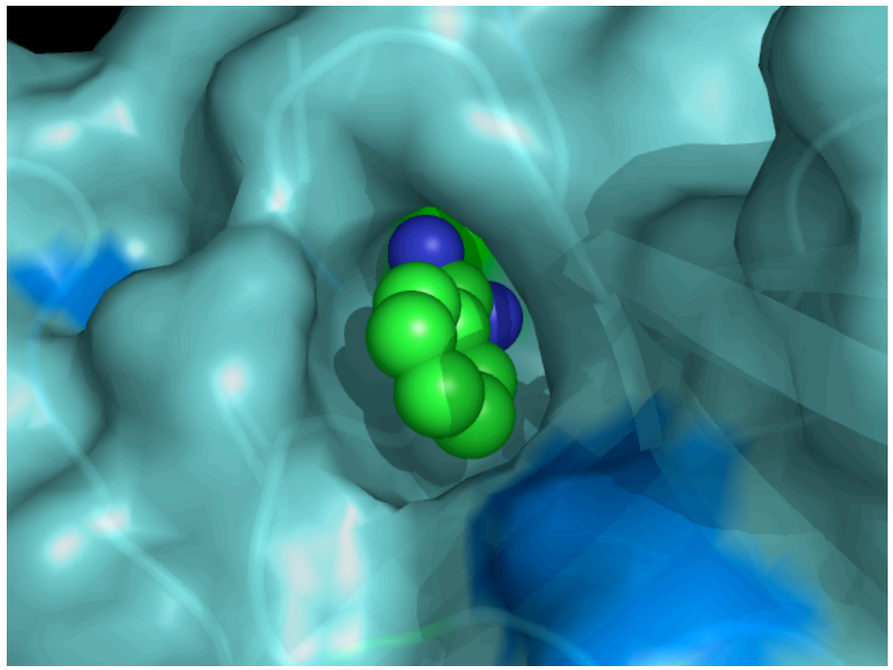
uno
THE UNIVERSITY of NEW ORLEANS

What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

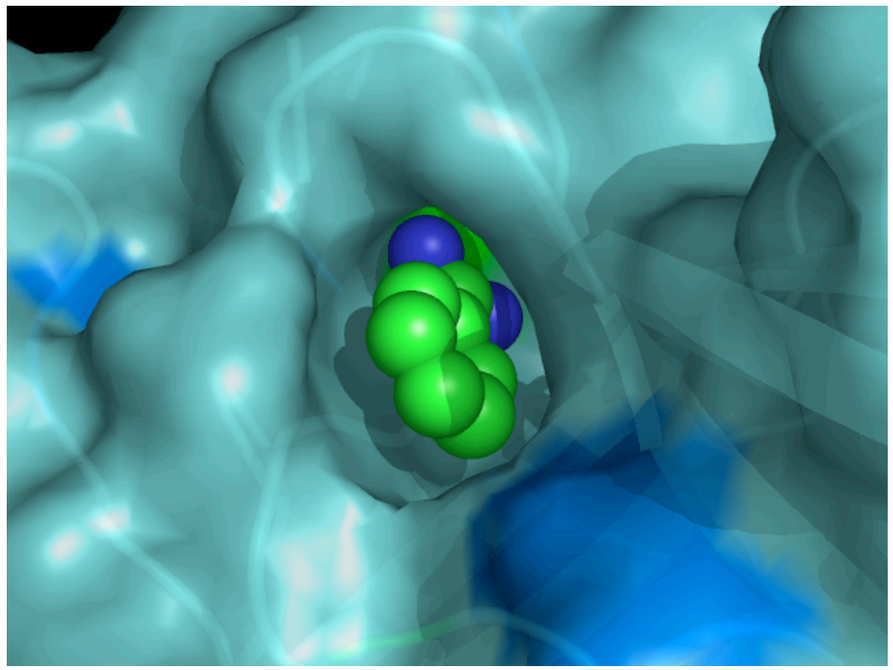
What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

- Design molecules to manipulate protein function

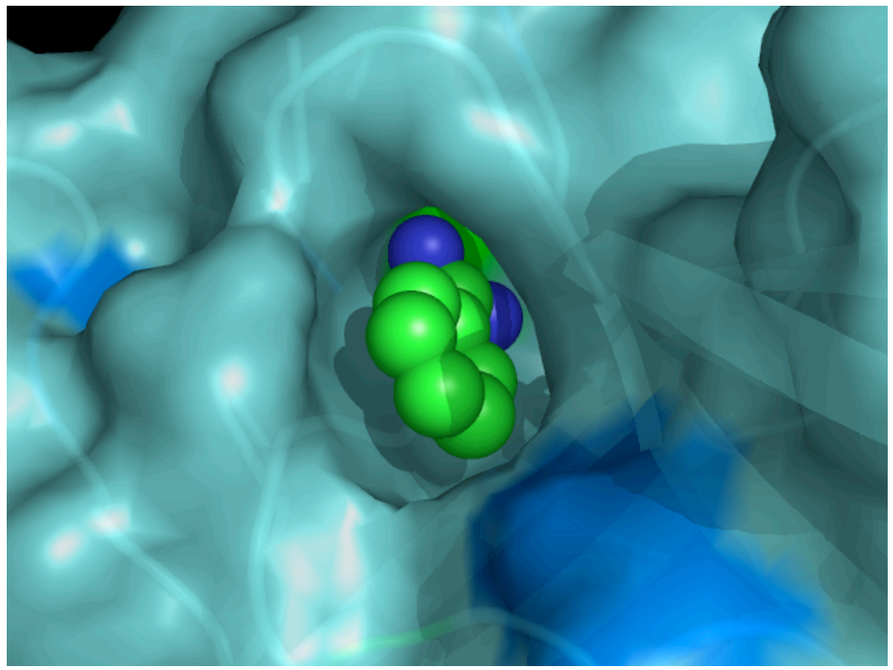
What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

- Design molecules to manipulate protein function
- Recognize toxins

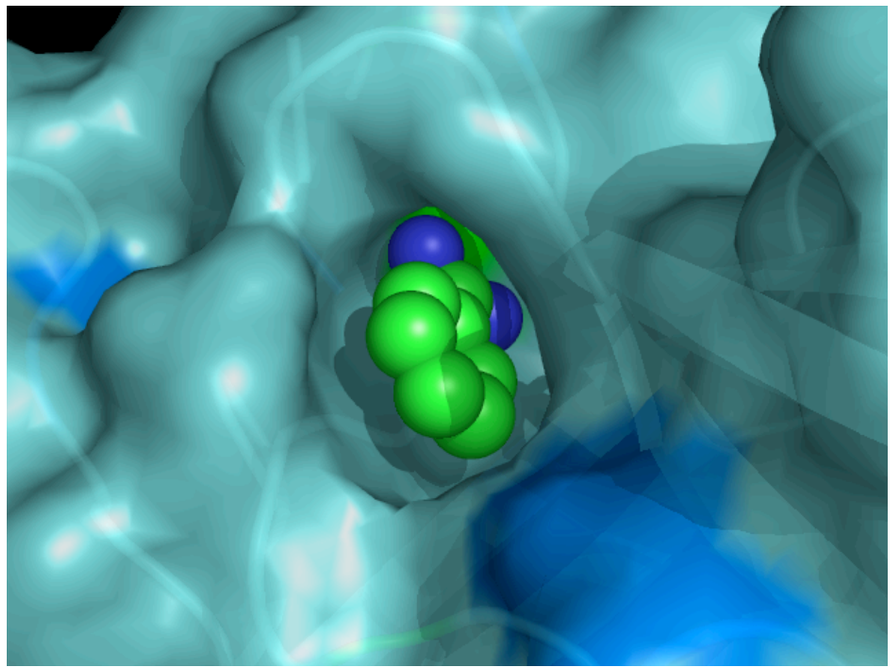
What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

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

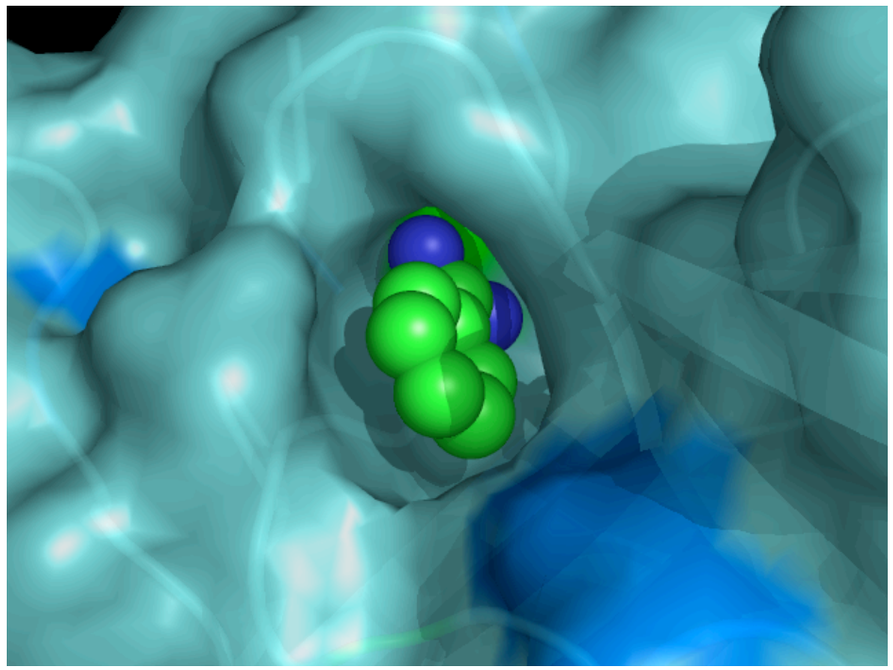
What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

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

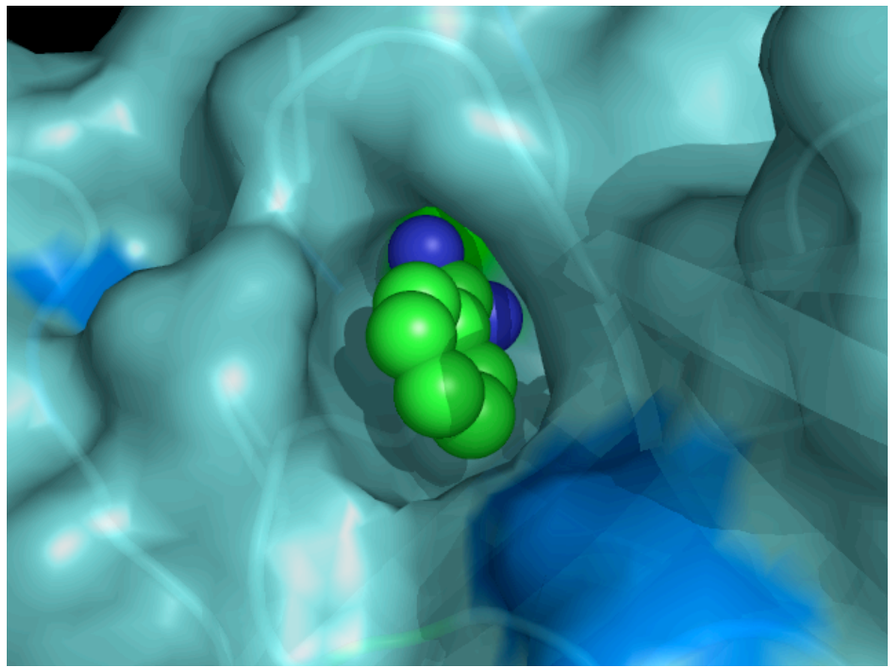
What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

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

What could we do with accurate binding free energies?



$$P + L \xrightarrow{\Delta G^0} PL$$

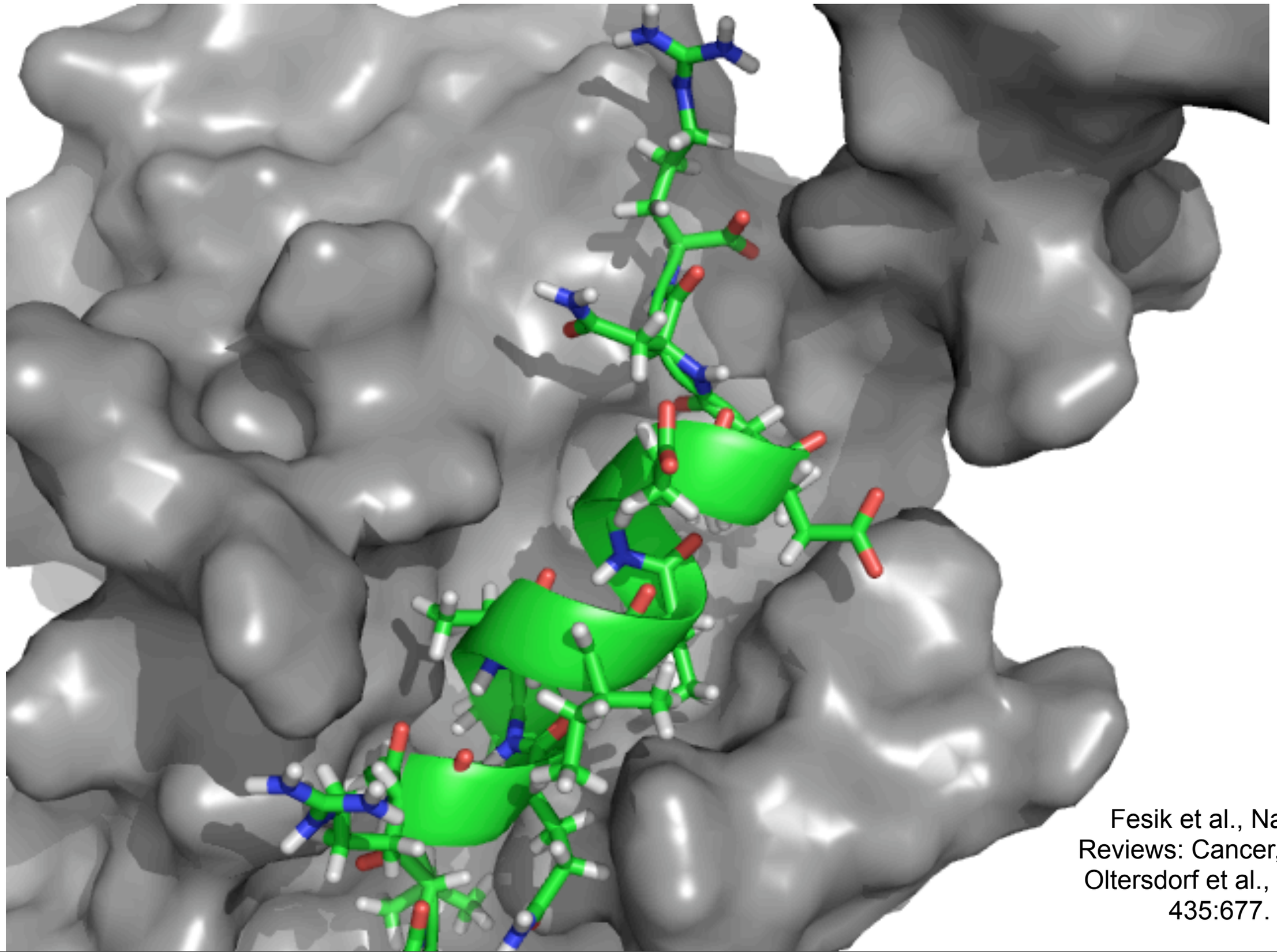
- Design molecules to manipulate protein function
- Recognize toxins
- Identify enzyme functions
- Protein design: Design binders to target molecule
- Aid medicinal chemistry
- Finally enable cold fusion

Drug design:

Drug discovery is hard and expensive

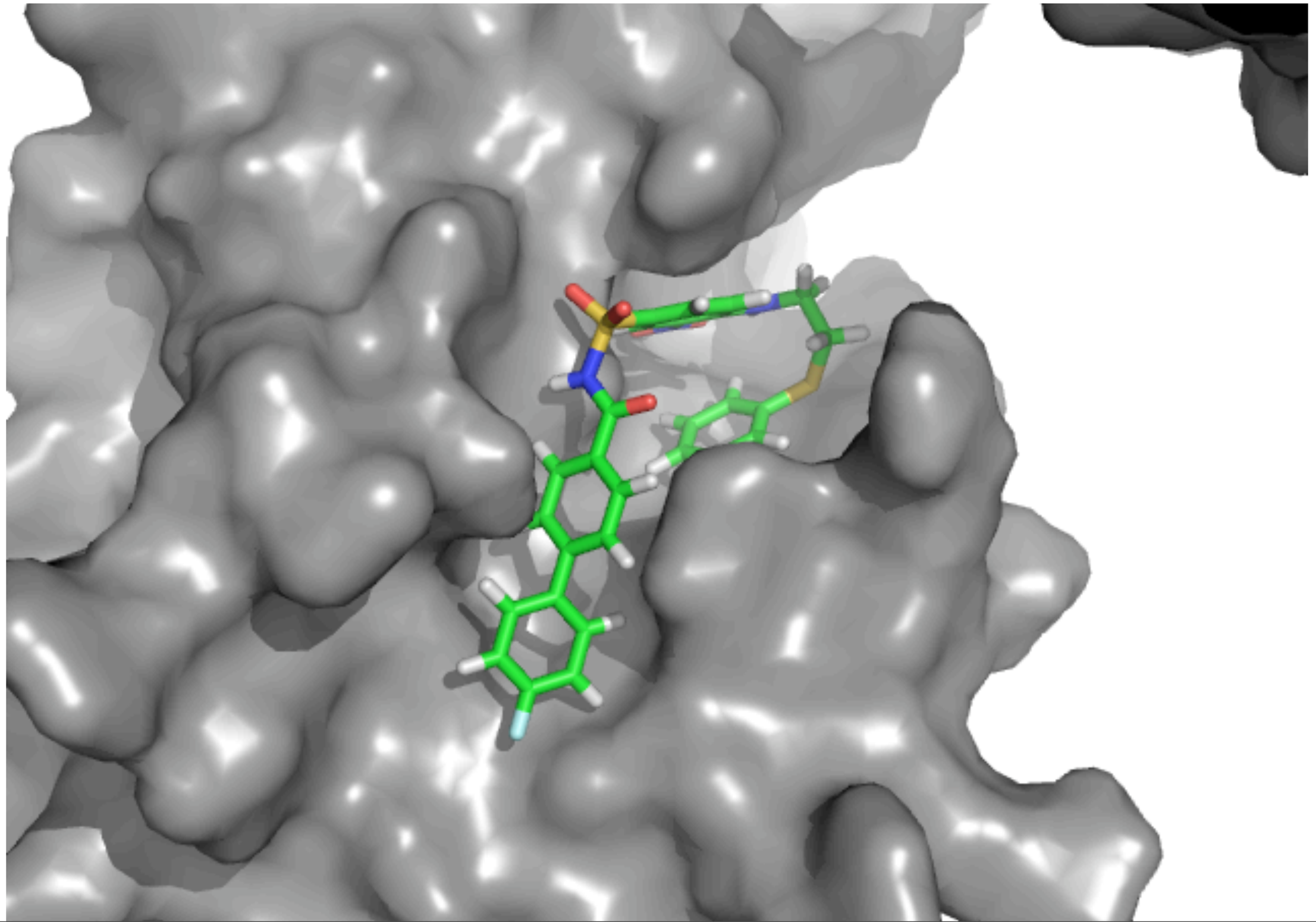
- $50B /yr in Pharma, billions at NIH
- Long, costly pipeline
  - –Screening: 10^6 compounds
  - –Takes 12-15 years
  - –Average ~ $1 billion per drug

Inhibitors can make good drugs



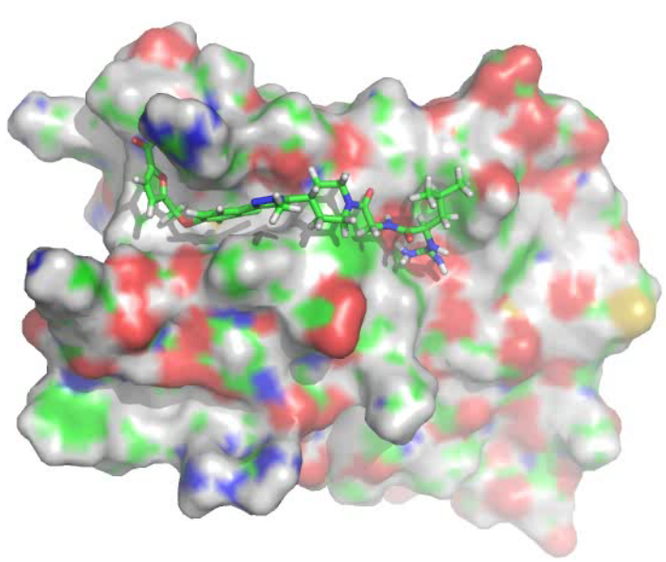
Fesik et al., Nature Reviews: Cancer, 5:876; Oltersdorf et al., Nature 435:677.

Small molecules can mimic binding partners

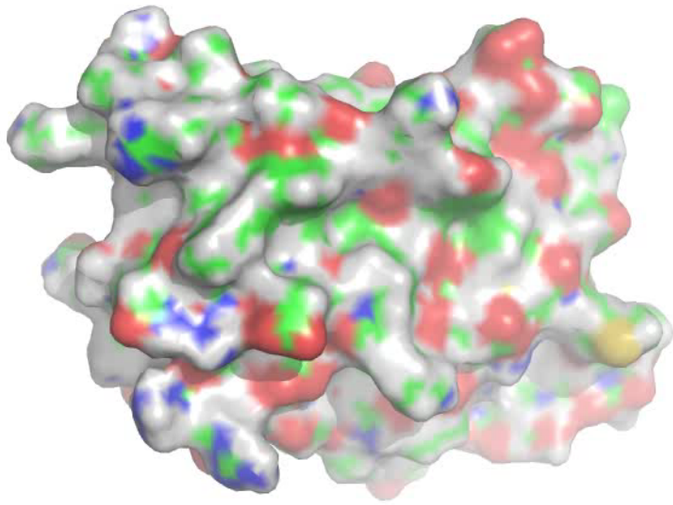


Binding free energies involve a ratio of partition functions

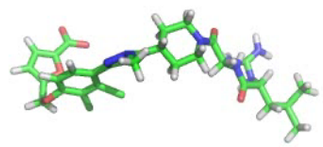
$$\Delta G = -k_B T \; ln \; Q_{PL}/Q_P Q_L$$



Q_{PL}



Q_P



Q_L

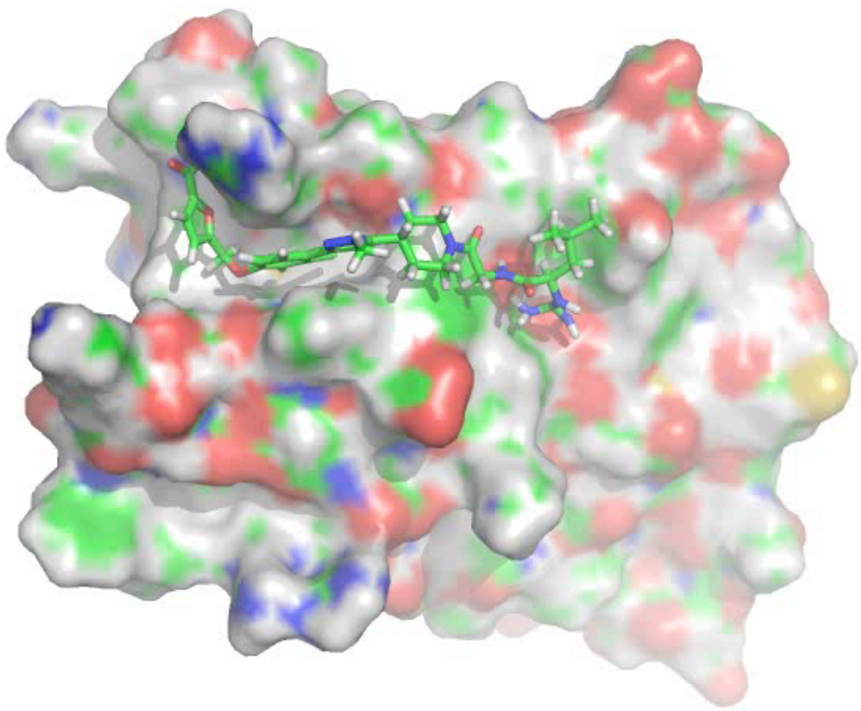
Binding free energies involve a ratio of partition functions

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

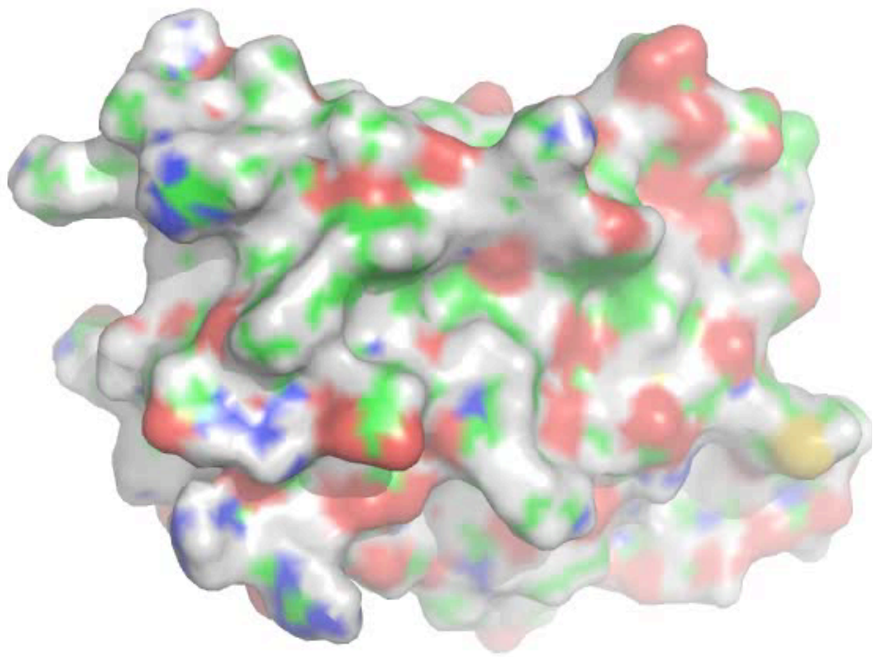
Q_{PL} Q_P Q_L

Binding free energies involve a ratio of partition functions

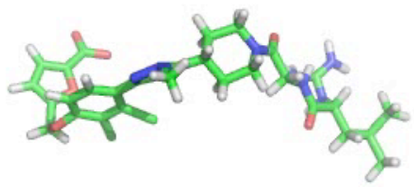
$$\Delta G = -k_B T \; ln \; Q_{PL}/Q_P Q_L$$



Q_{PL}

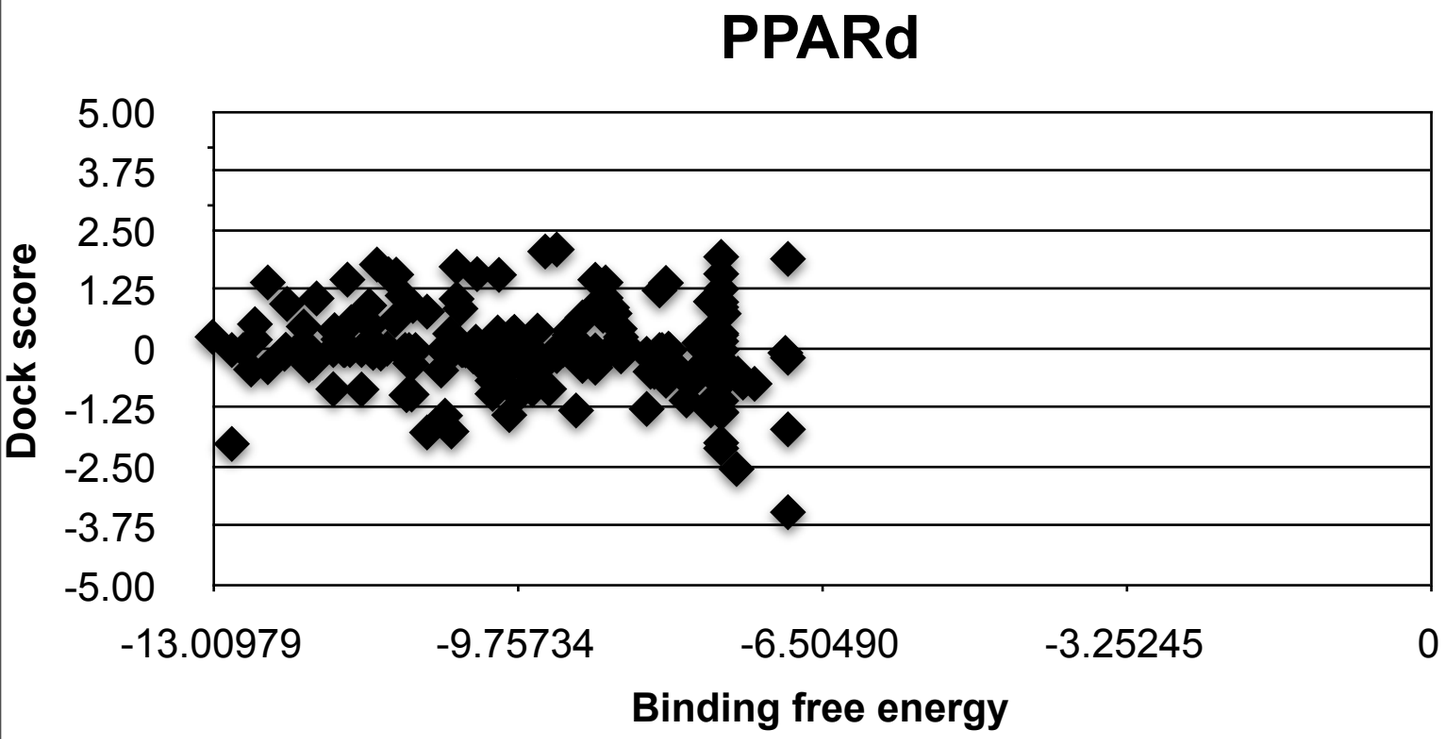


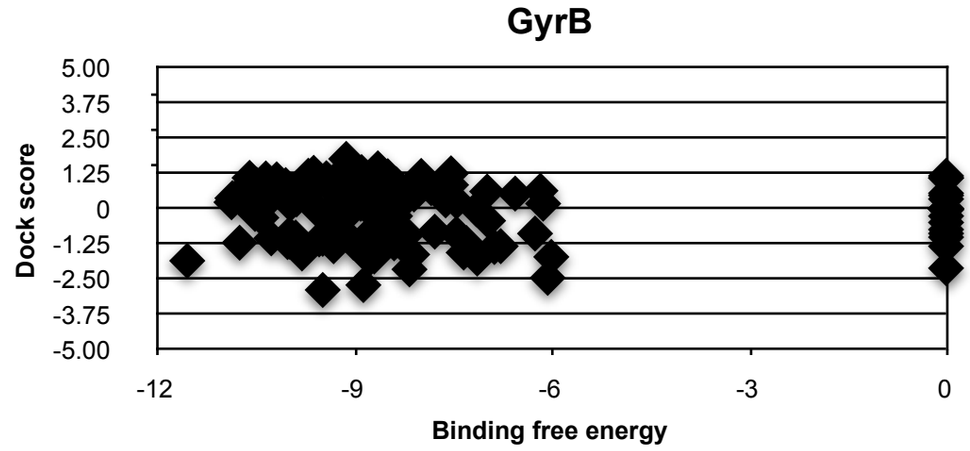
Q_P

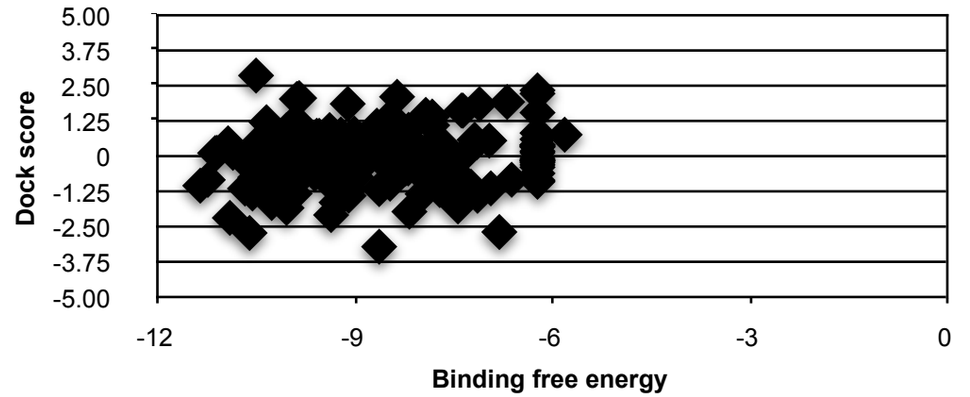


Q_L

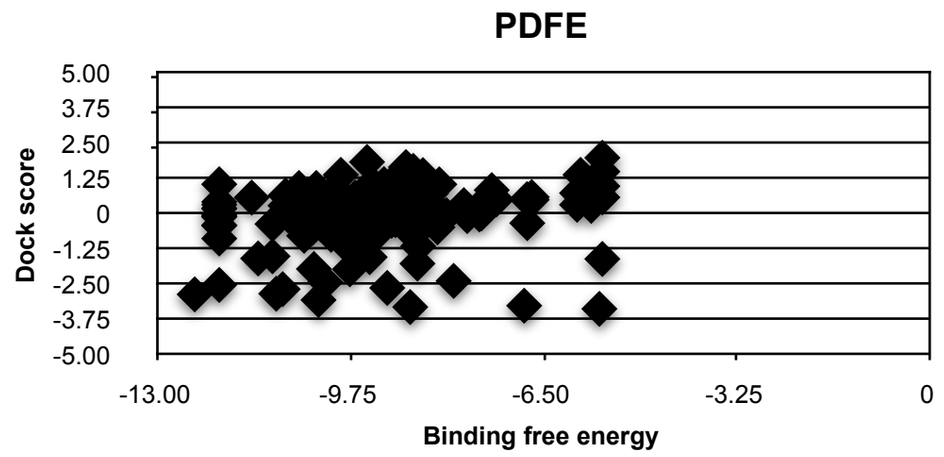
Existing methods for predicting binding need improvement





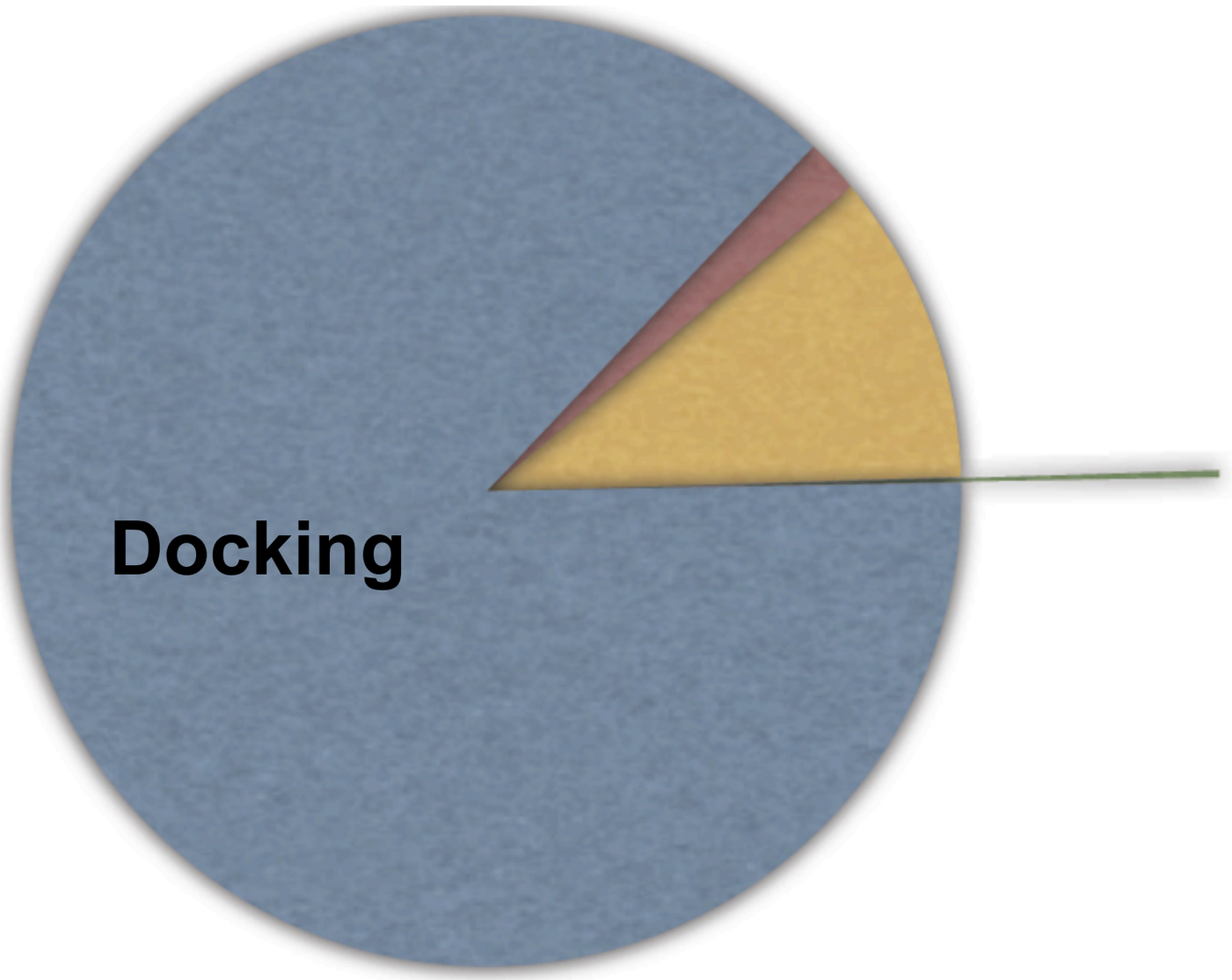


- Docking can't calculate binding free energies, or even relative binding strengths



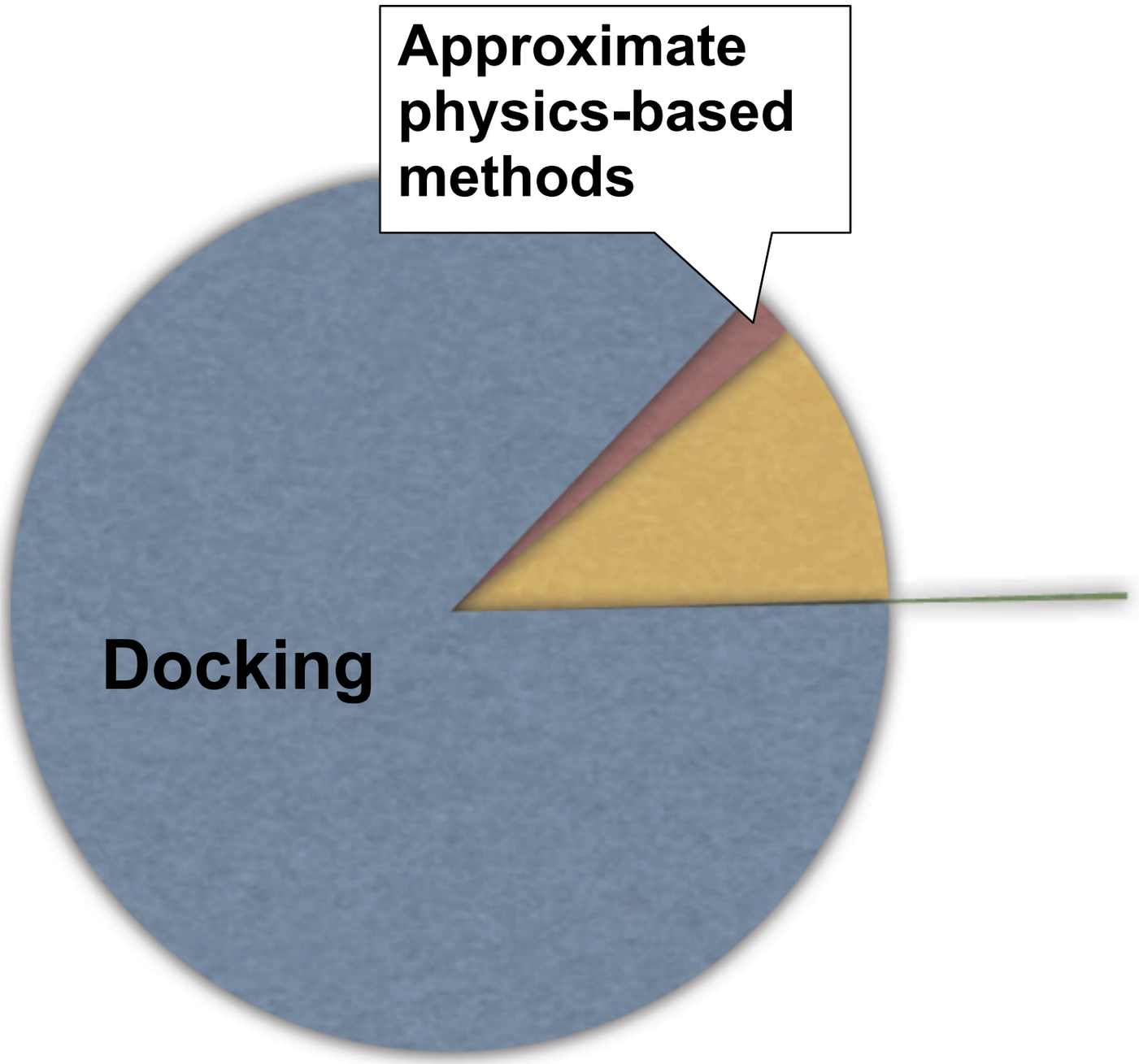
Warren et al, JMC. 49:5912 (2005); also Velec et al., JMC. 48:6296 (2005), Huang et al., JMC. 49:6789 (2006)

Docking is tremendously popular because it's simple and fast, with few alternatives



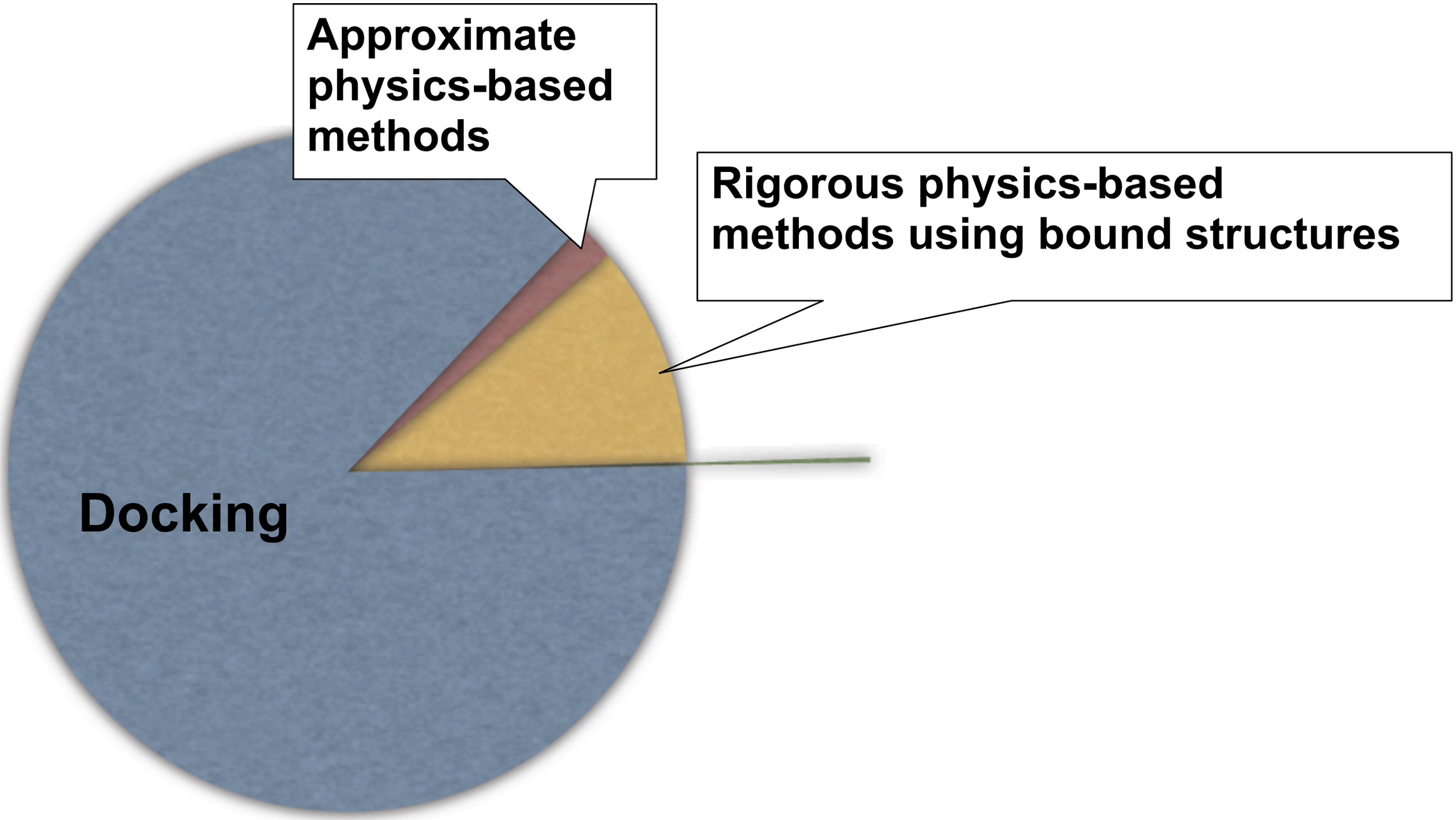
Relative publication numbers from Google scholar, 11/2007

Docking is tremendously popular because it's simple and fast, with few alternatives



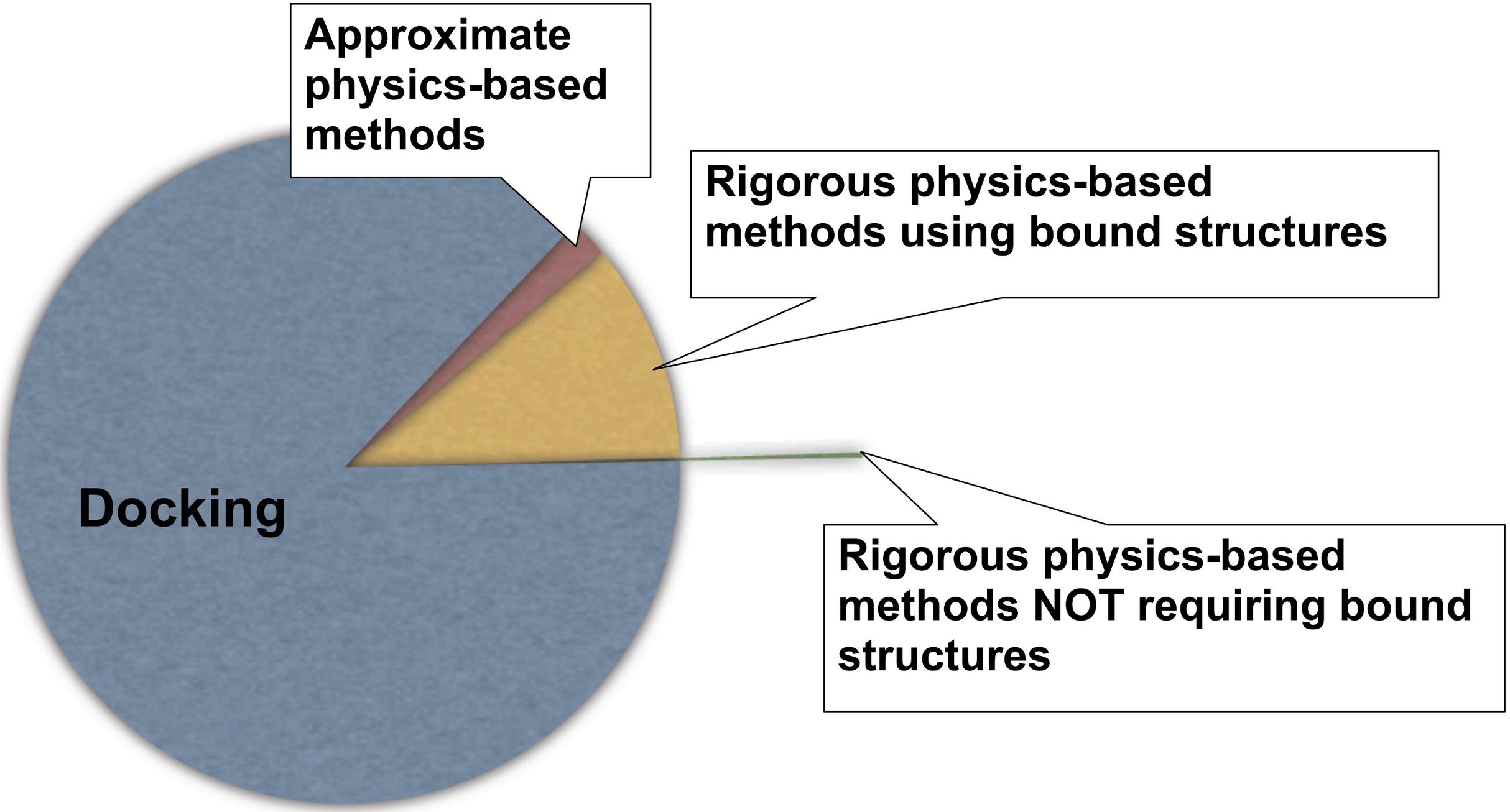
Relative publication numbers from Google scholar, 11/2007

Docking is tremendously popular because it's simple and fast, with few alternatives



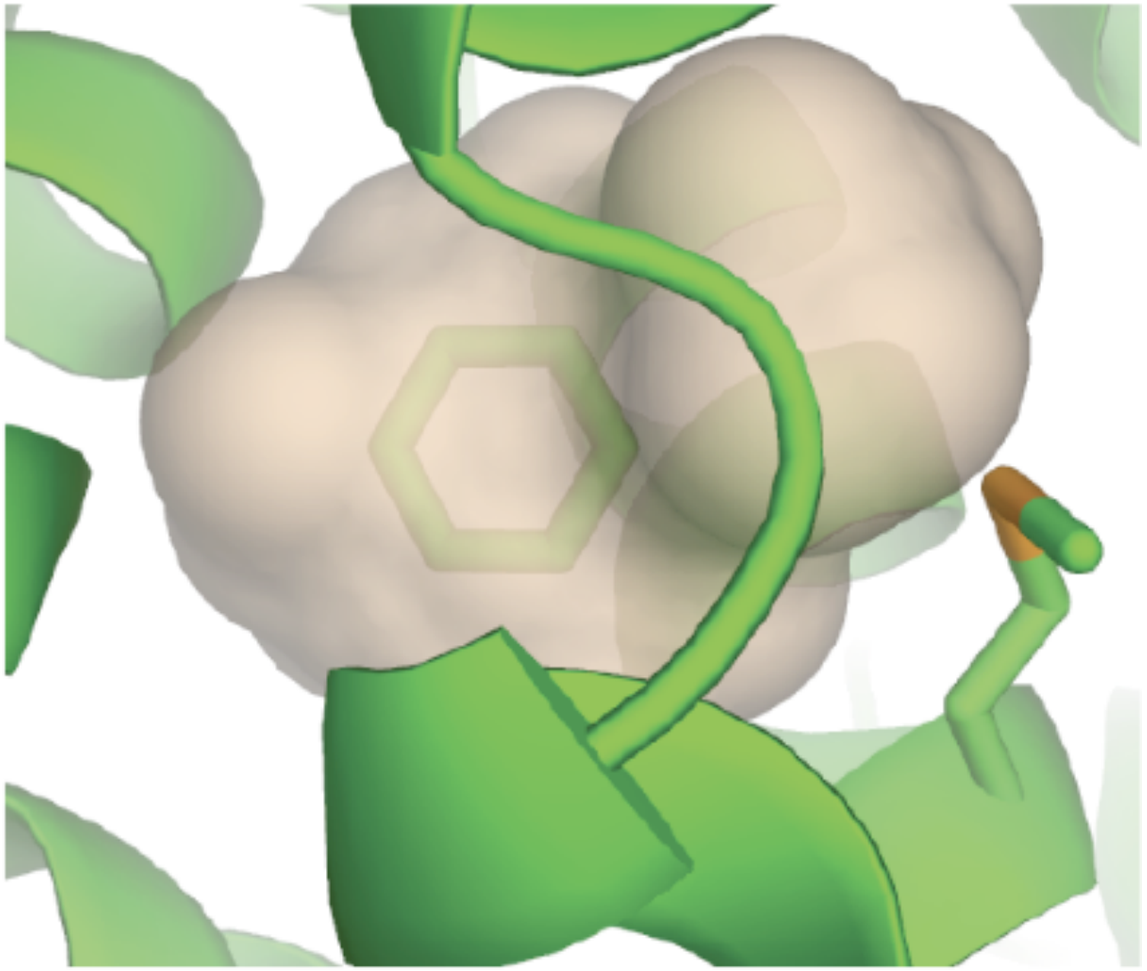
Relative publication numbers from Google scholar, 11/2007

Docking is tremendously popular because it's simple and fast, with few alternatives



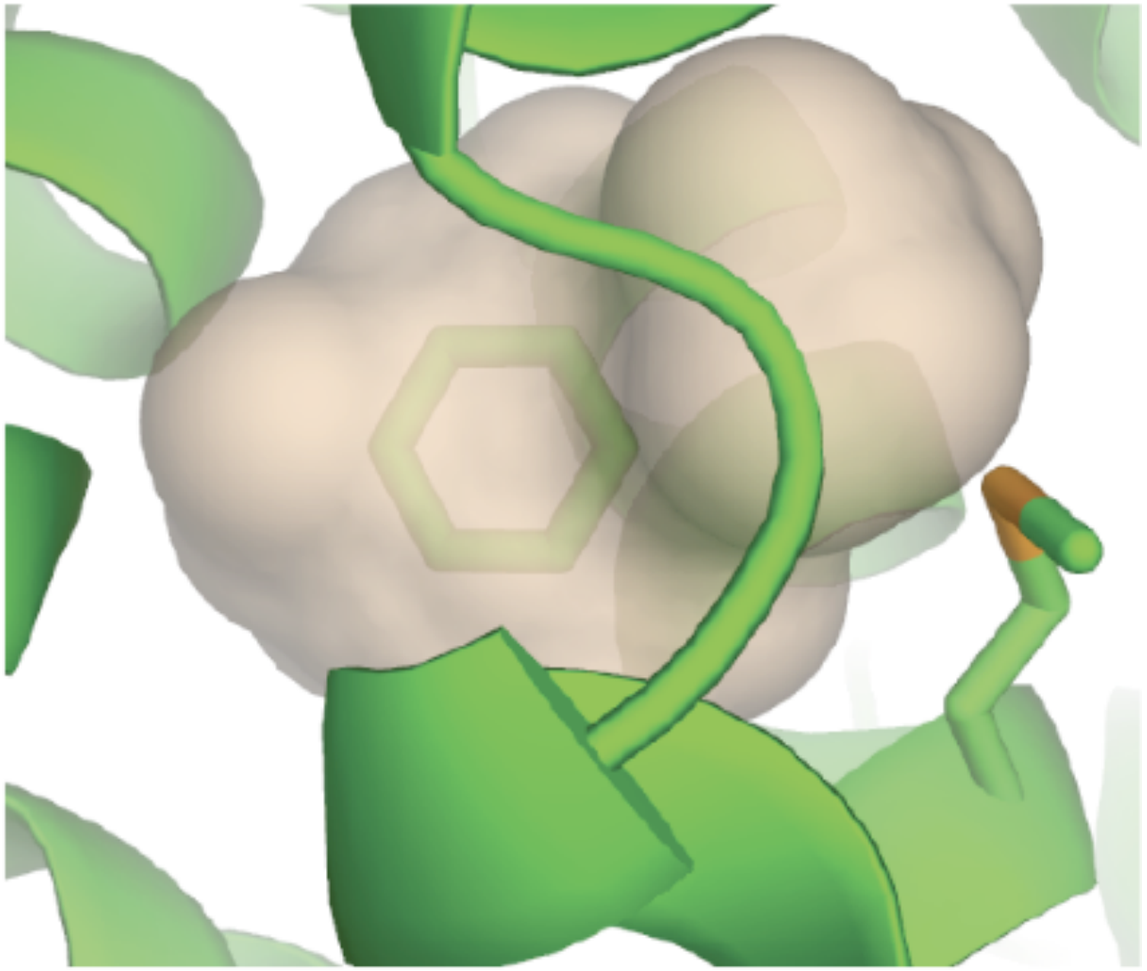
Relative publication numbers from Google scholar, 11/2007

Wonderful experimental model system to improve binding calculations



- Simple nonpolar cavity
- Well characterized
- Easy to get structural data
- Experimental collaborators -- Shoichet
- Opportunity for predictions
- Transferable insights

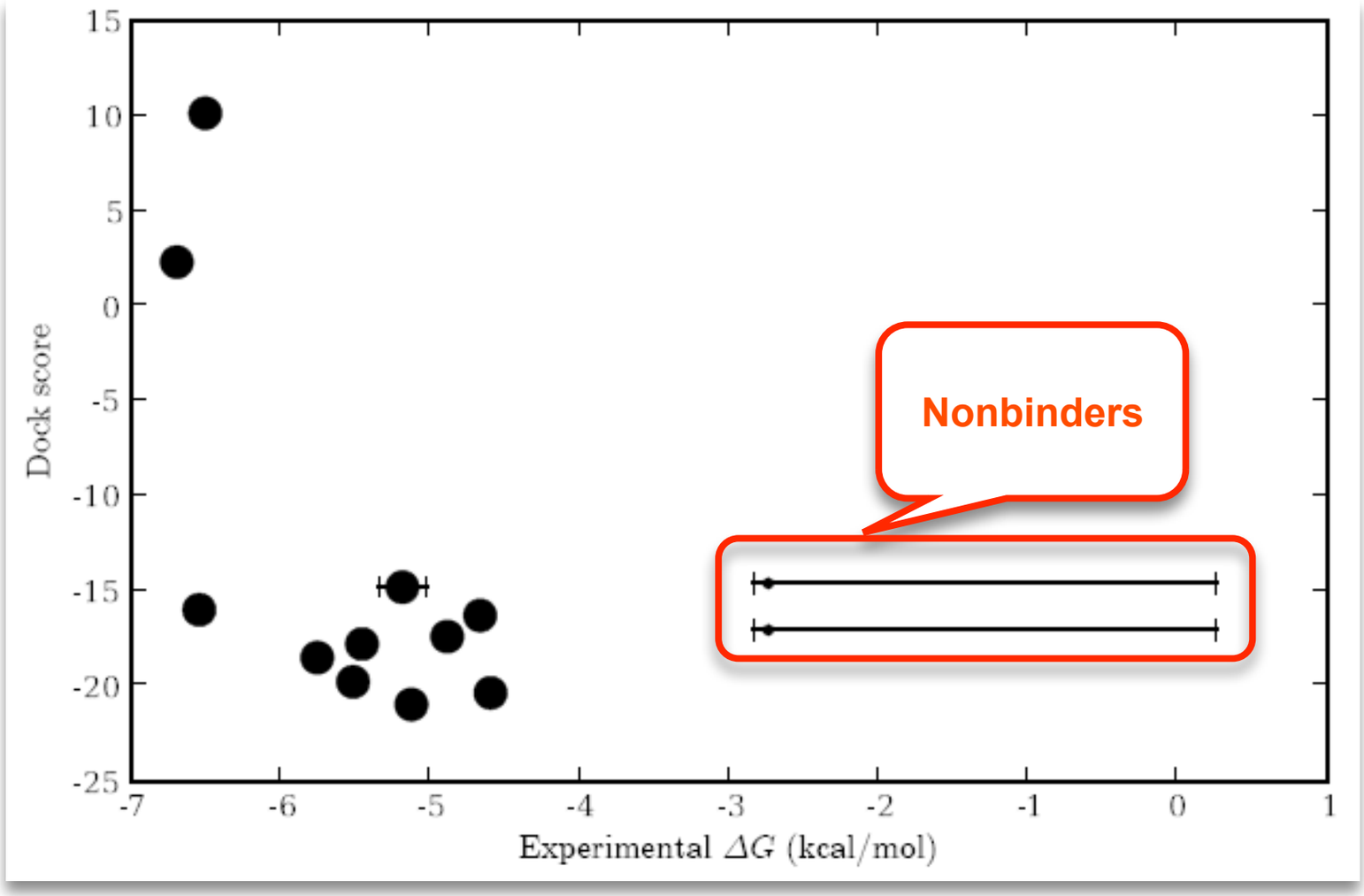
Wonderful experimental model system to improve binding calculations



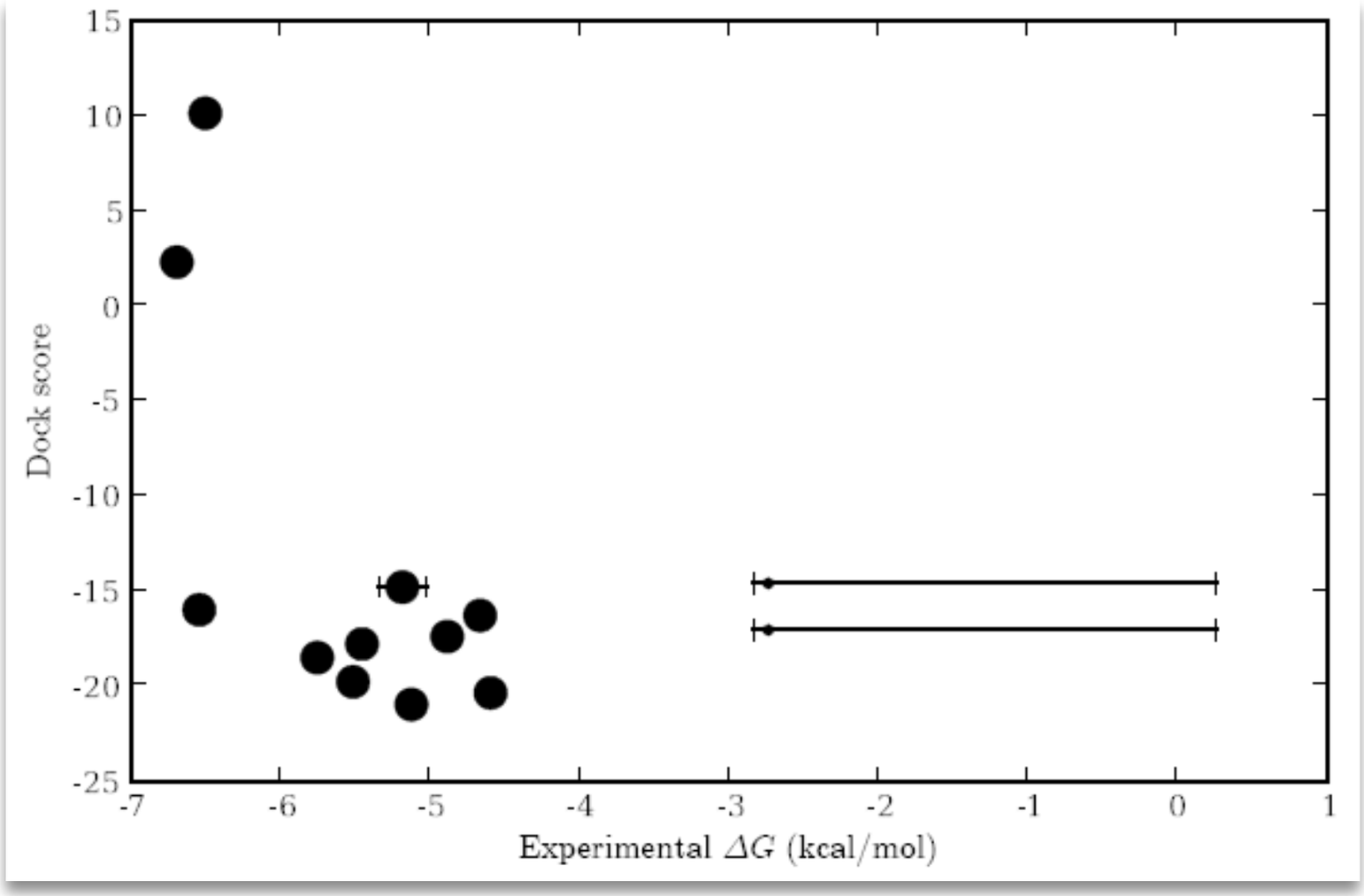
- Simple nonpolar cavity
- Well characterized
- Easy to get structural data
- Experimental collaborators -- Shoichet
- Opportunity for predictions
- Transferable insights

- **If this won't work on a simple binding site…**

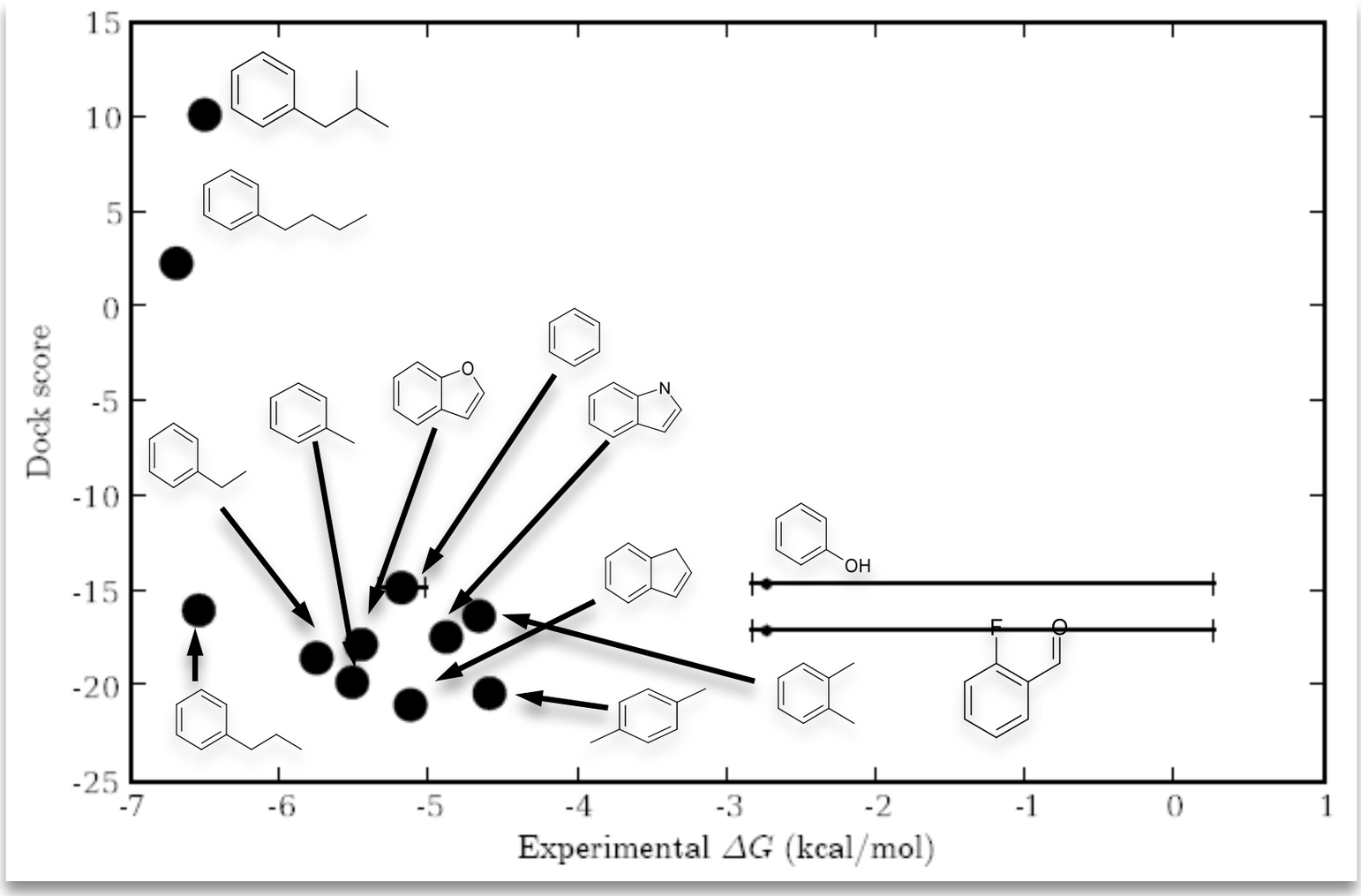
Here, docking performs poorly for relative binding strengths



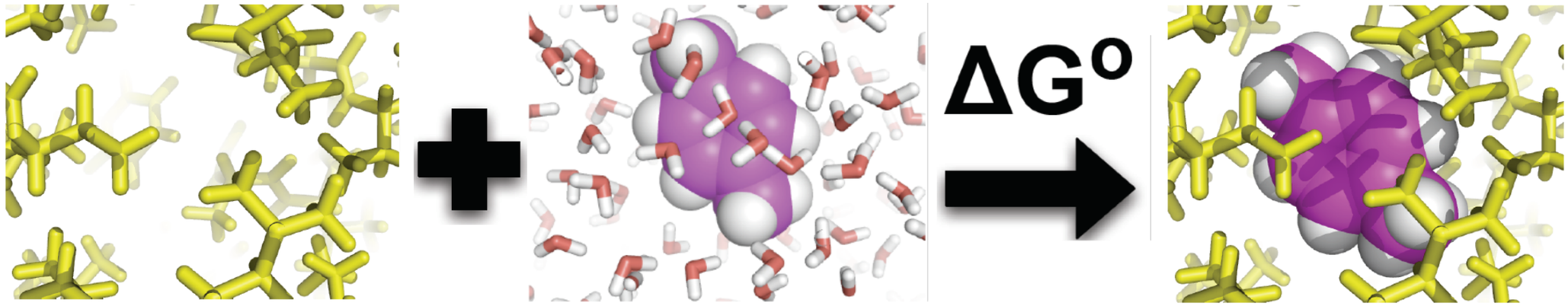
Here, docking performs poorly for relative binding strengths



Here, docking performs poorly for relative binding strengths

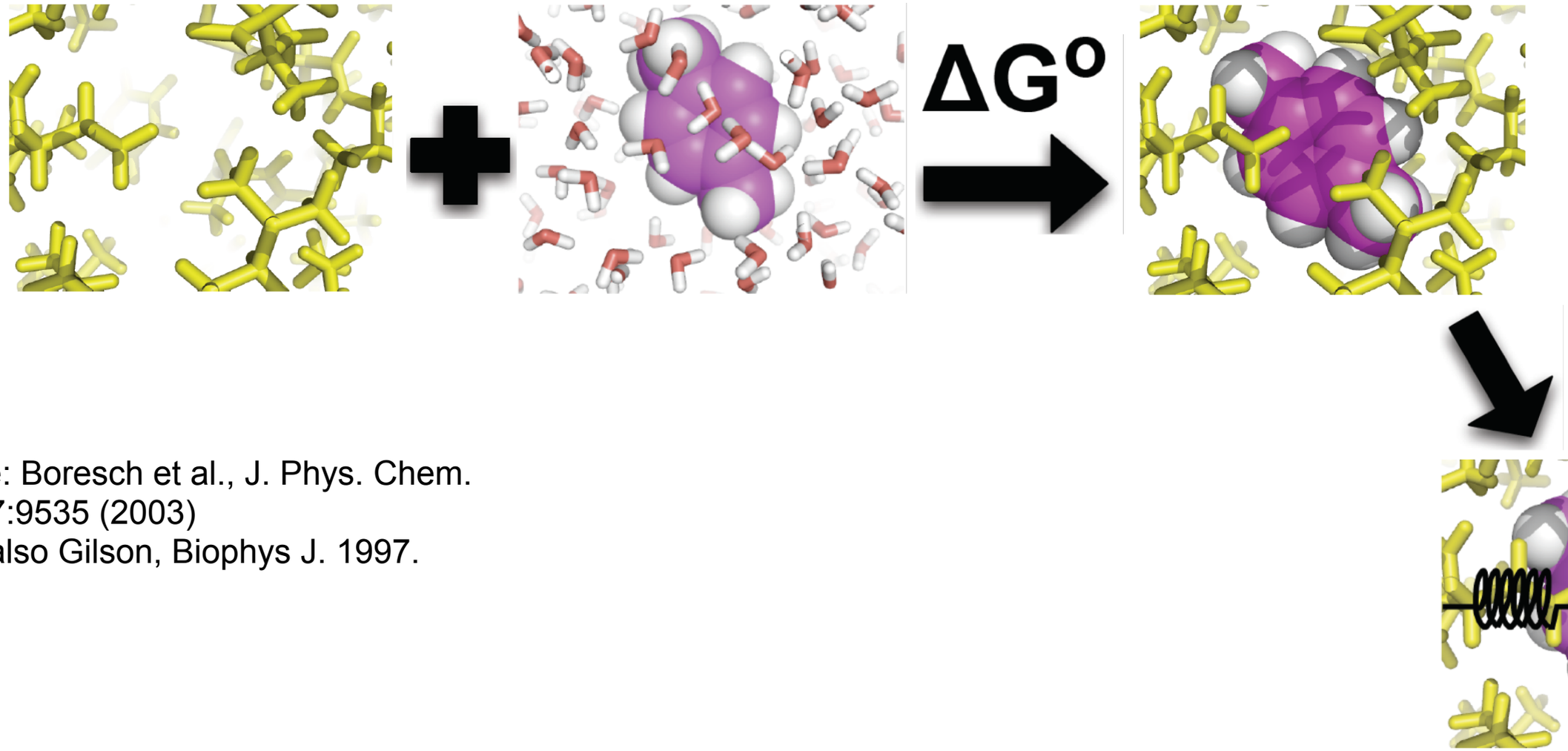


Thermodynamic cycle



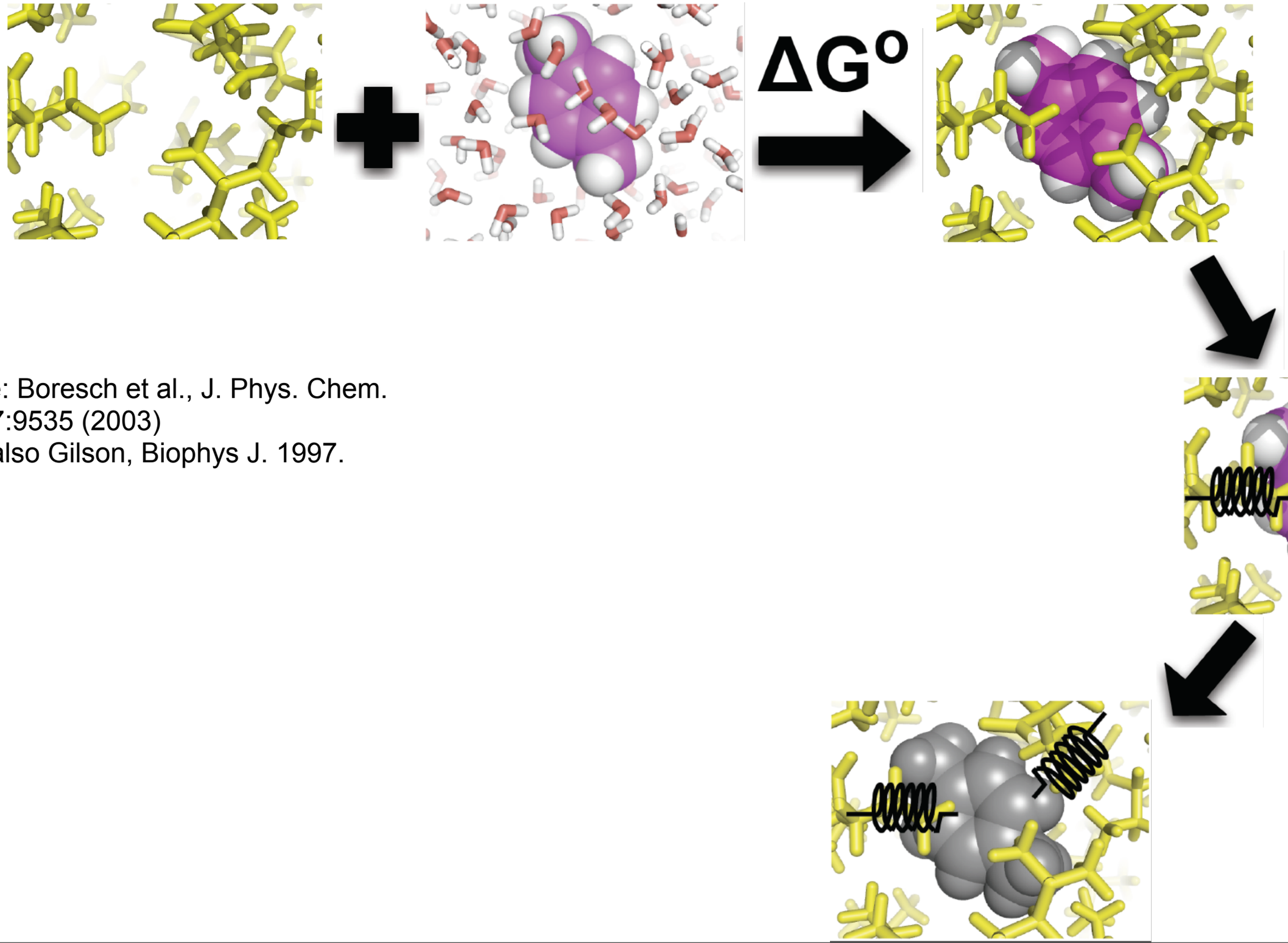
Cycle: Boresch et al., J. Phys. Chem. B 107:9535 (2003)
See also Gilson, Biophys J. 1997.

Thermodynamic cycle



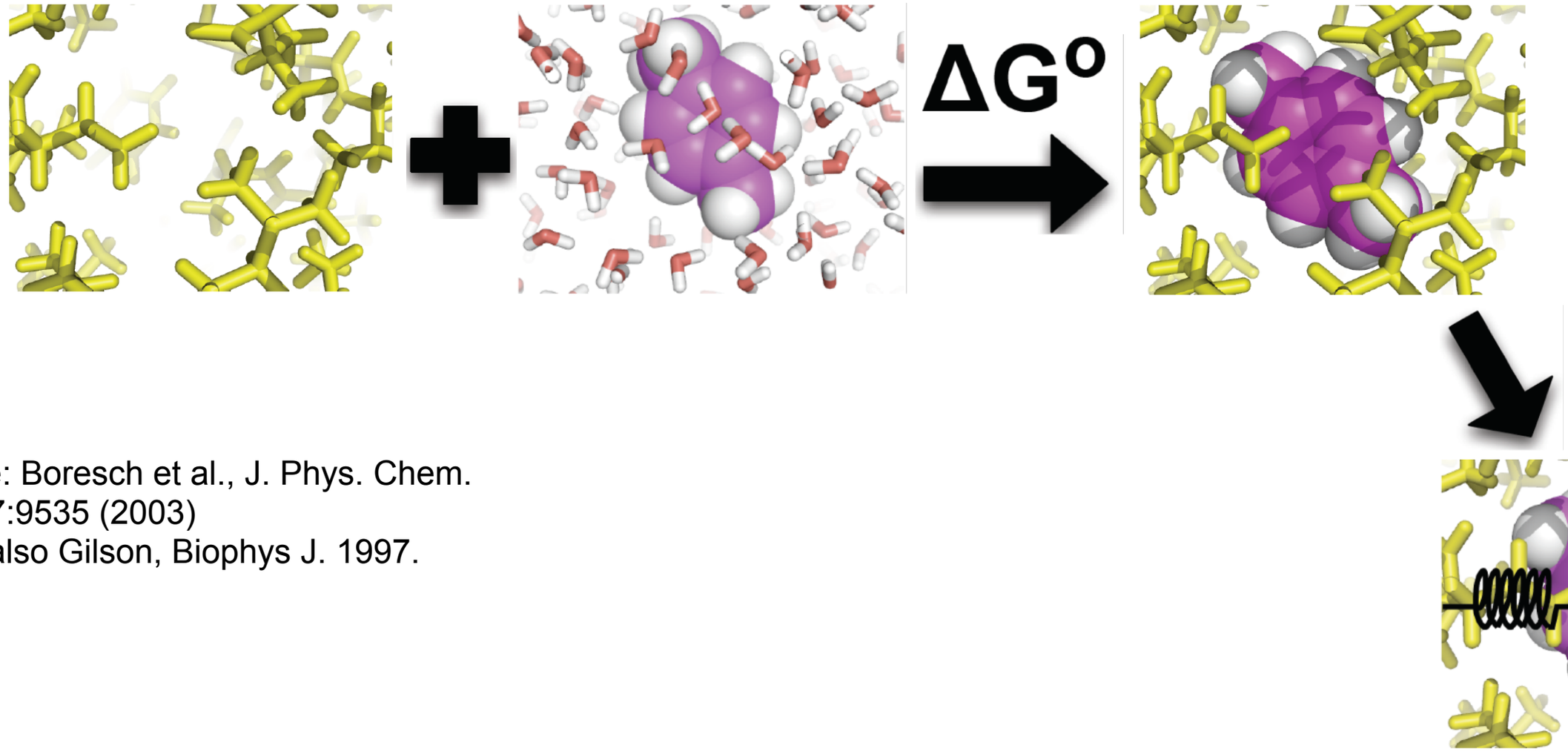
Cycle: Boresch et al., J. Phys. Chem. B 107:9535 (2003)
See also Gilson, Biophys J. 1997.

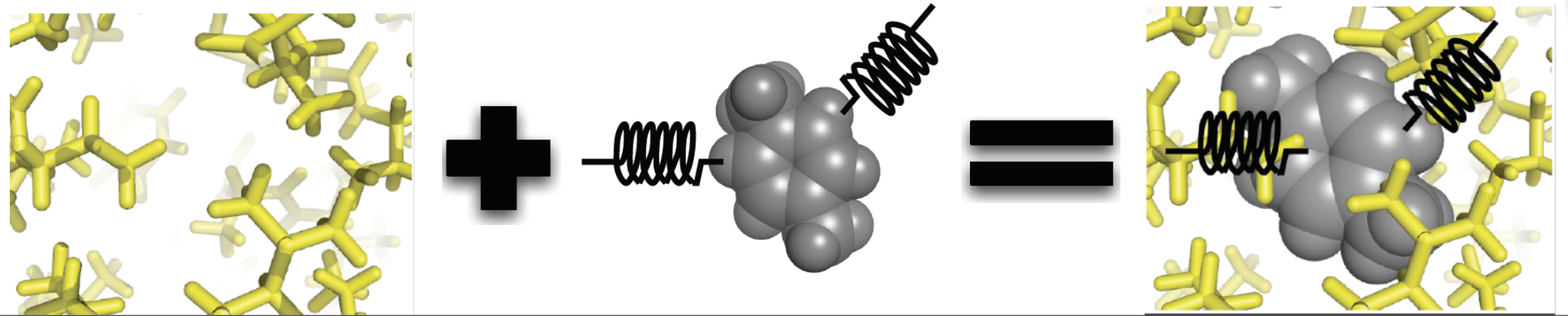
Thermodynamic cycle



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

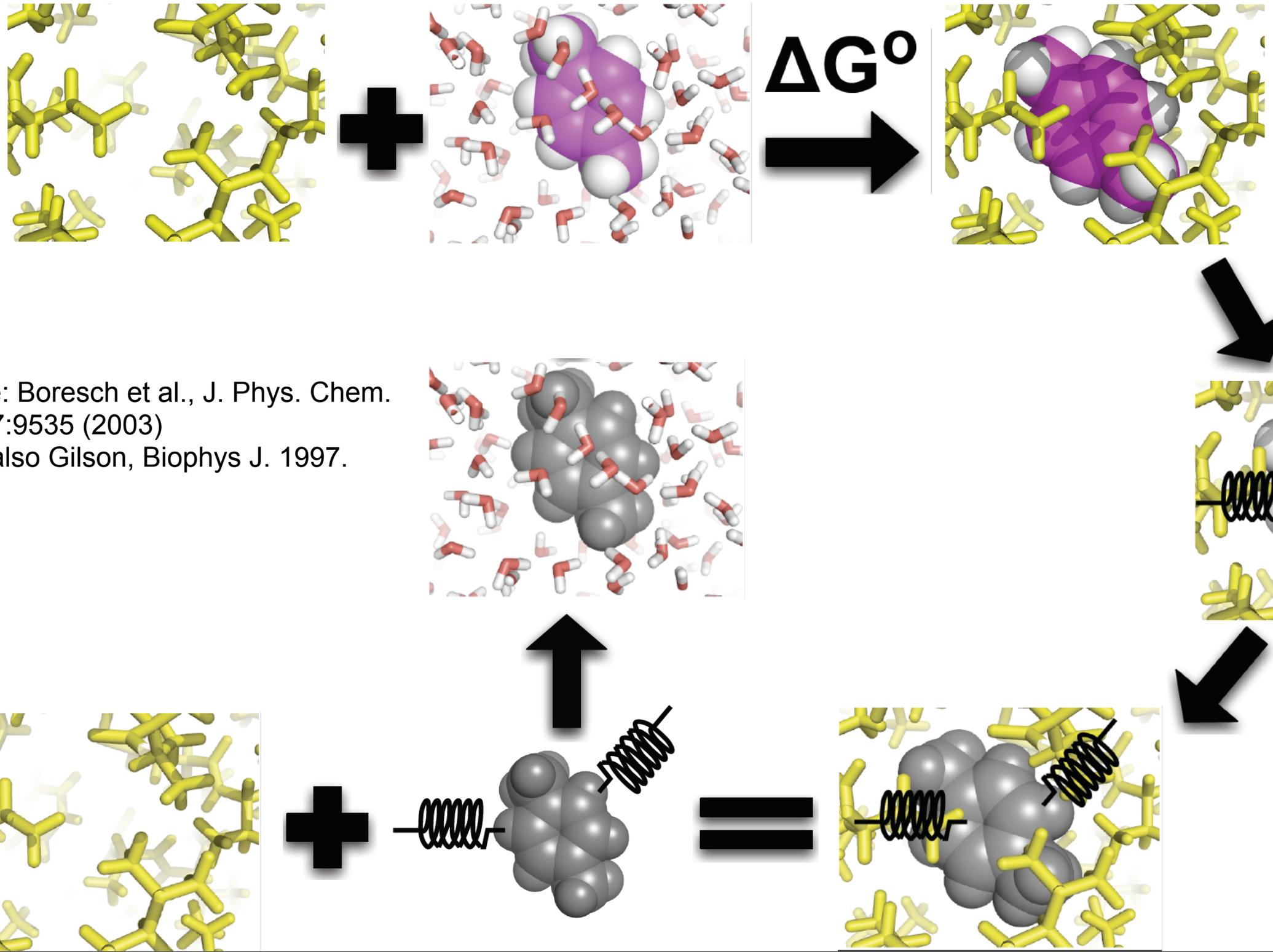
Thermodynamic cycle





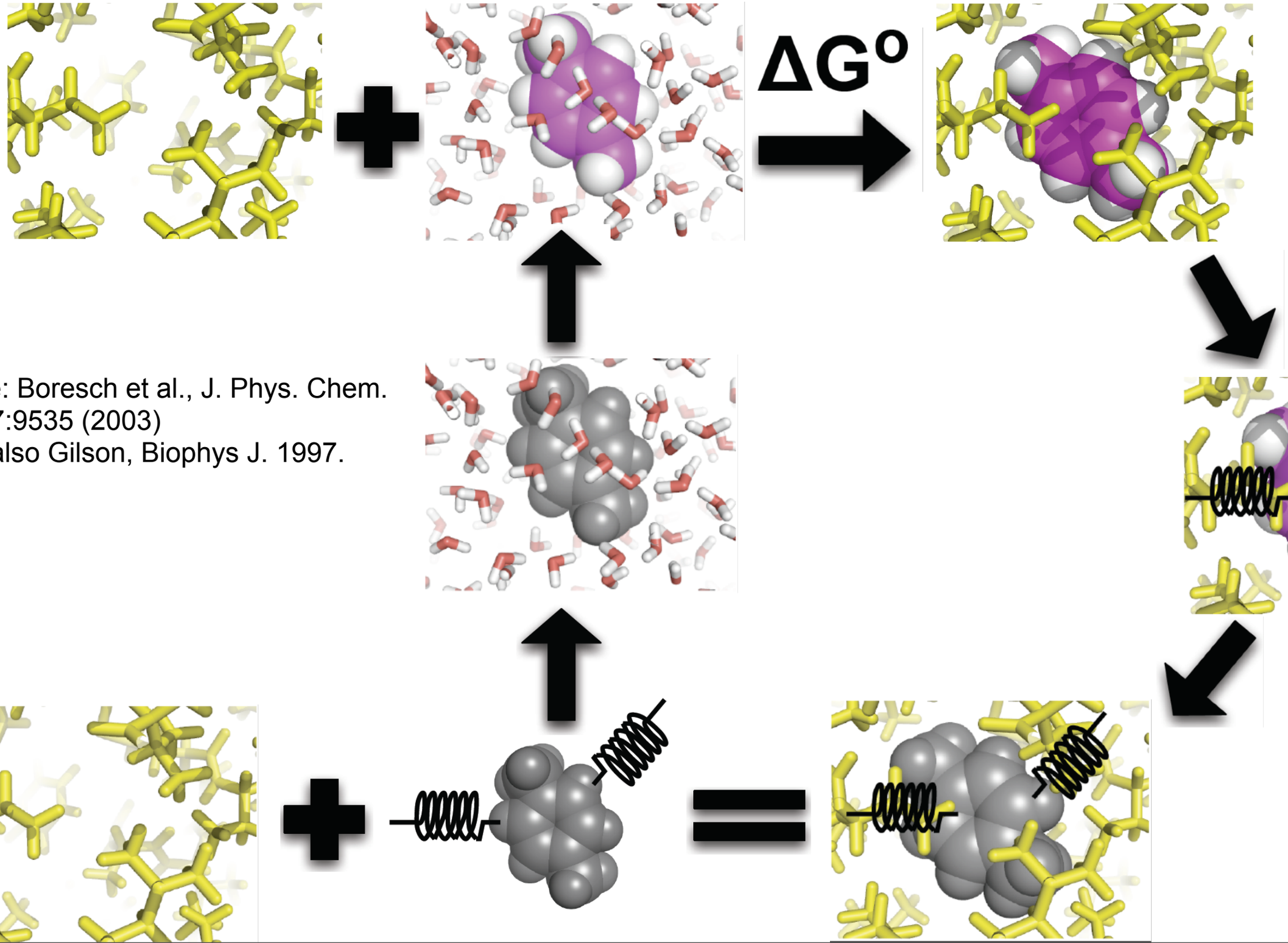
Cycle: Boresch et al., J. Phys. Chem. B 107:9535 (2003)
See also Gilson, Biophys J. 1997.

Thermodynamic cycle



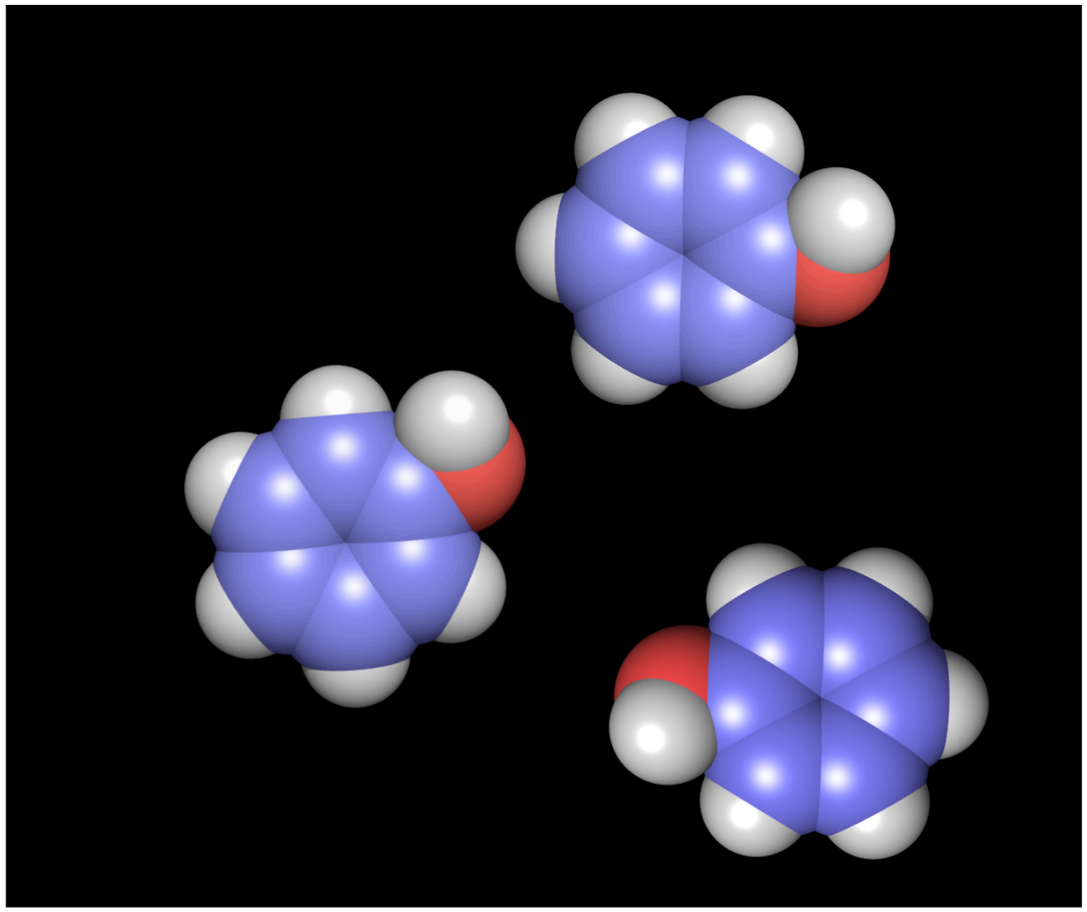
Cycle: Boresch et al., J. Phys. Chem. B 107:9535 (2003)
See also Gilson, Biophys J. 1997.

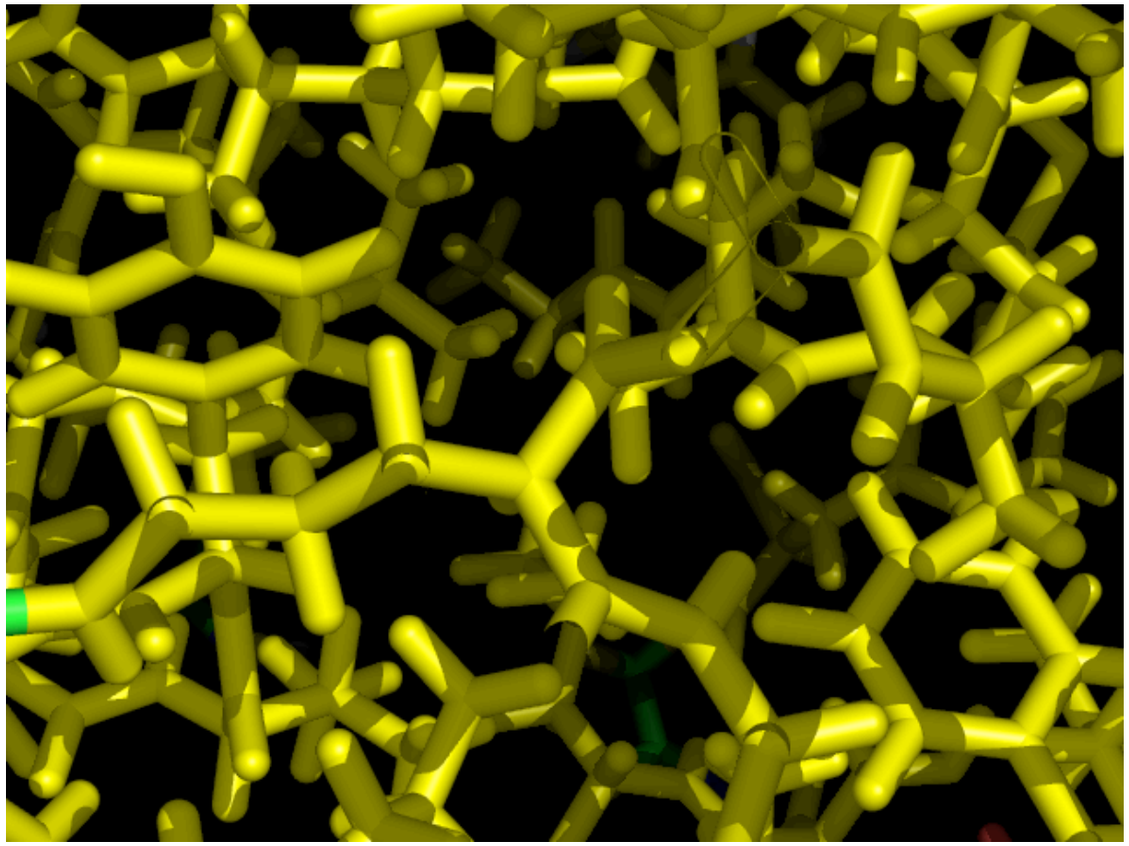
Thermodynamic cycle



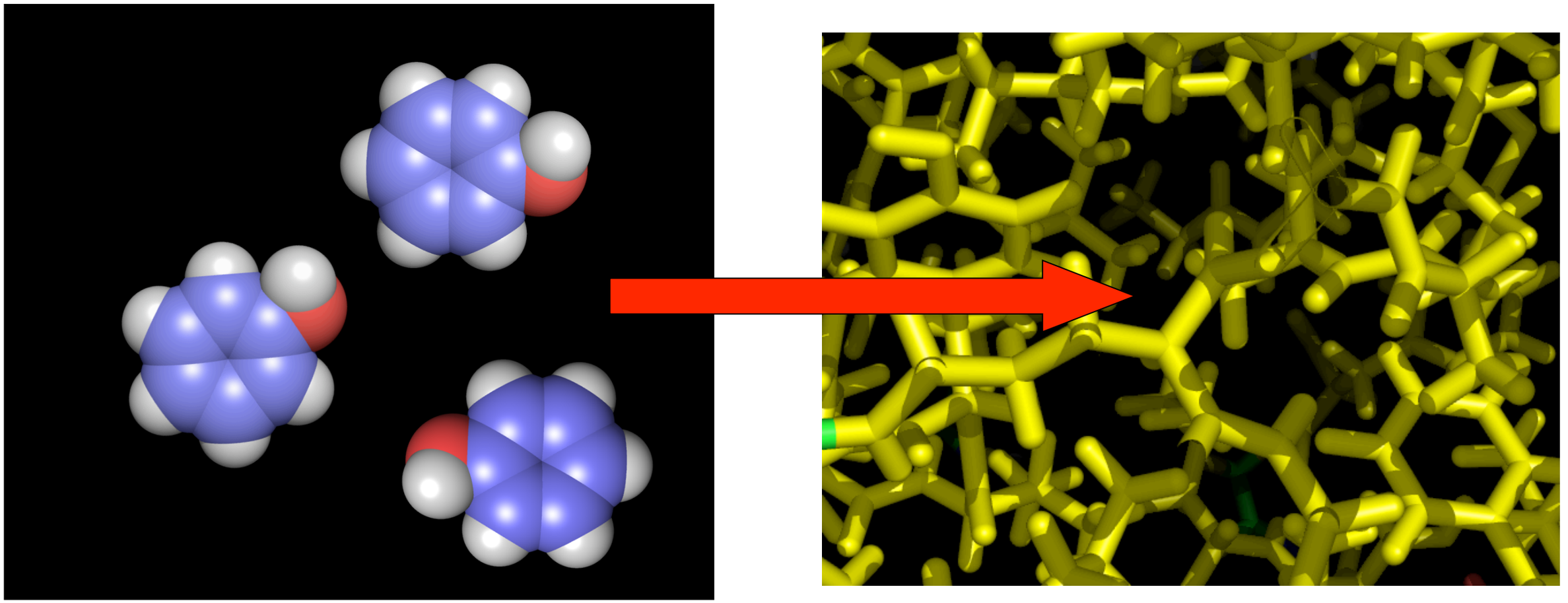
Cycle: Boresch et al., J. Phys. Chem. B 107:9535 (2003)
See also Gilson, Biophys J. 1997.

Our approach requires no knowledge of bound structure



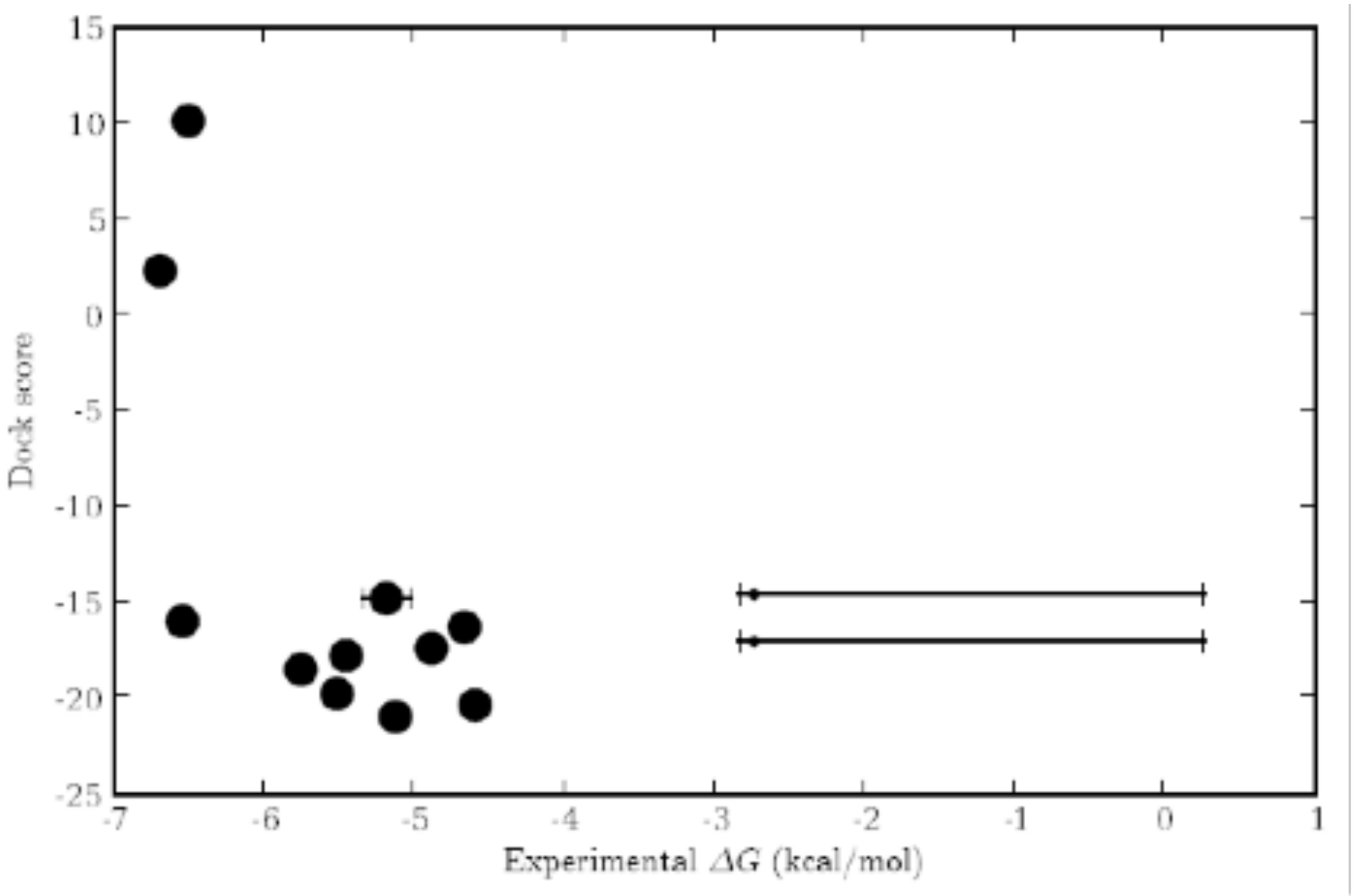


Our approach requires no knowledge of bound structure

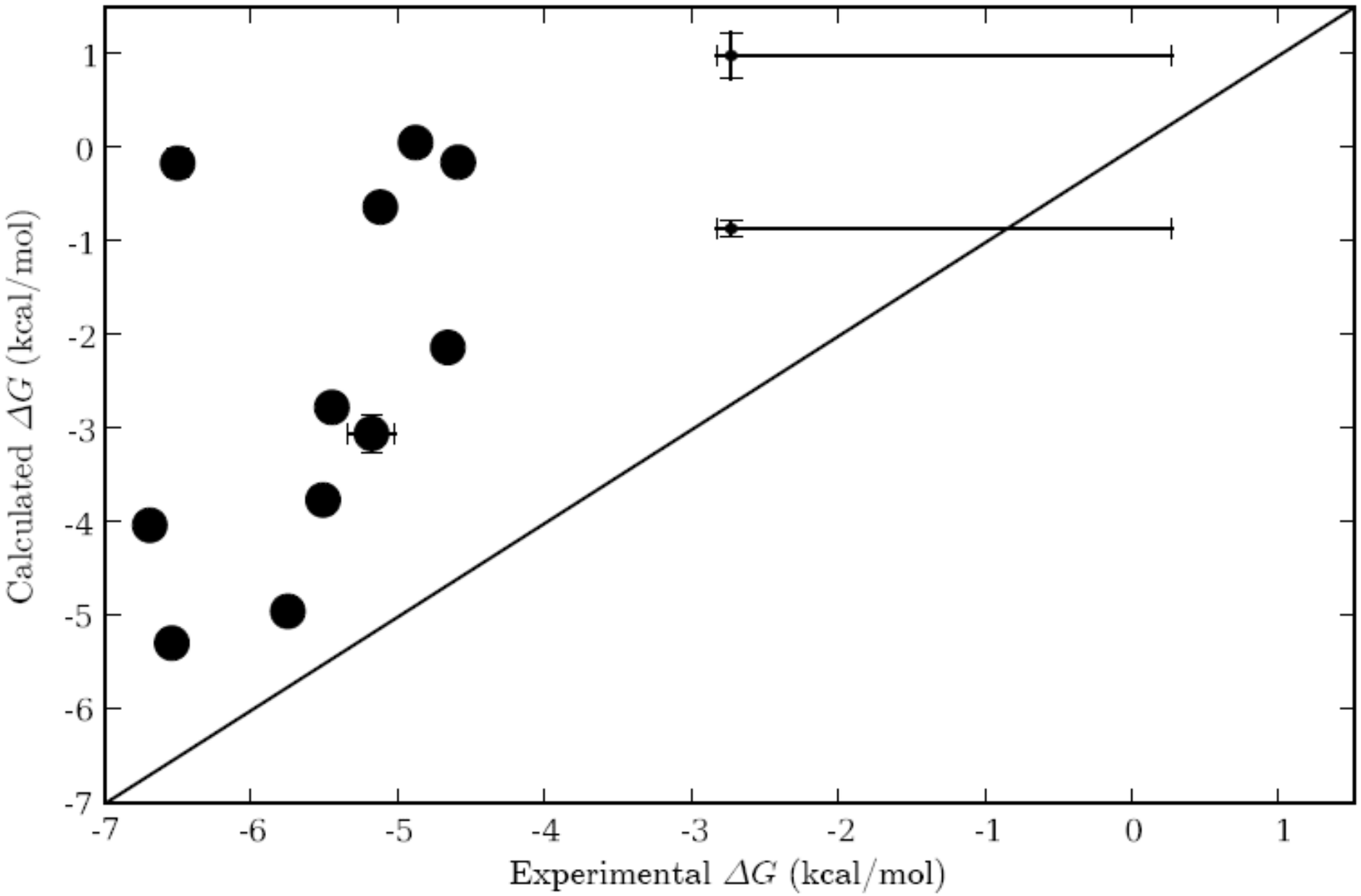


- Generate starting conformations using docking

Test: In apolar cavity, docking performs poorly

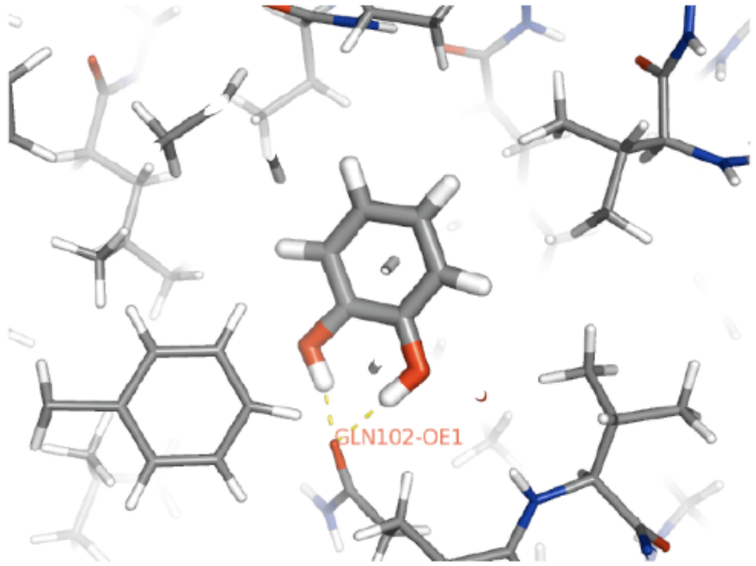


Step 1: Free energy calculations do work better than docking

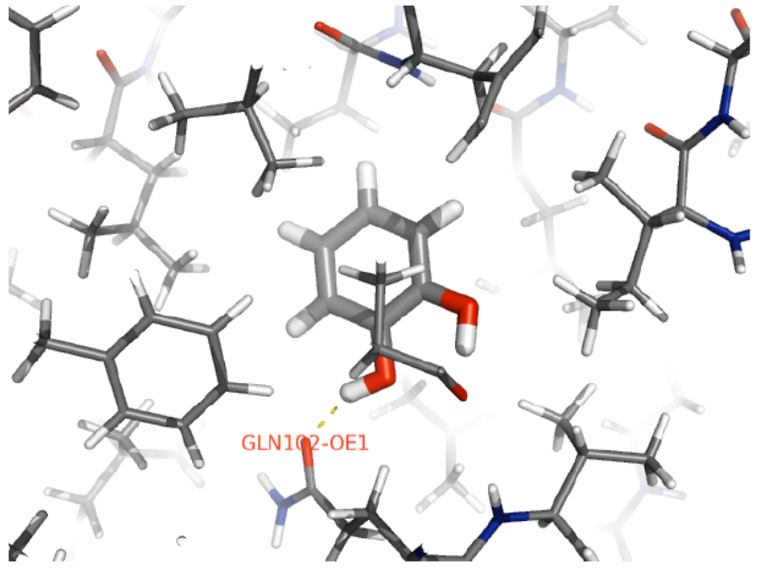


RMS error 3.5 kcal/mol

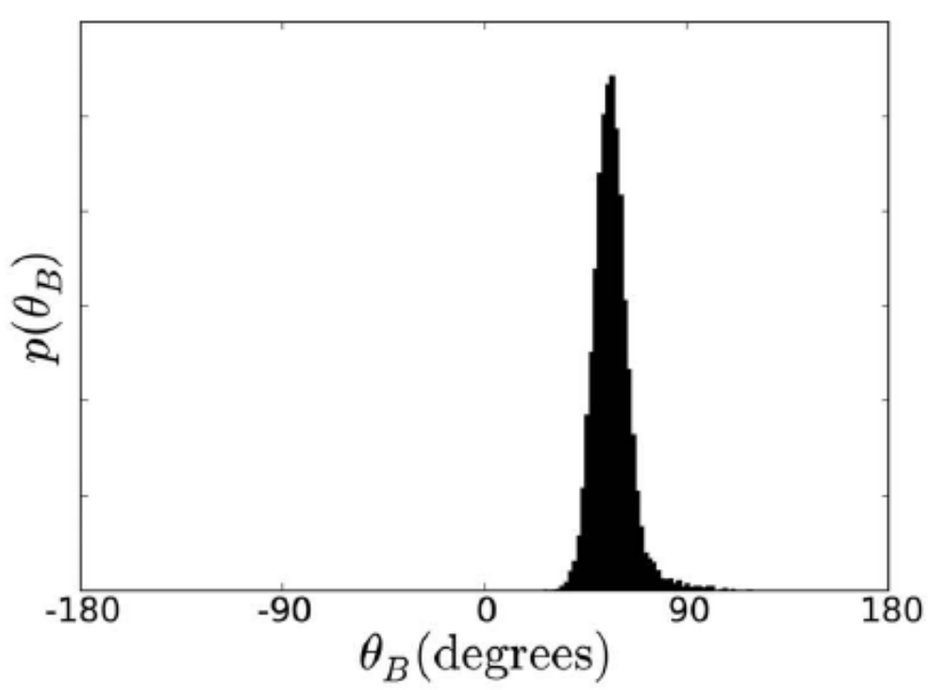
Problem: Multiple ligand orientations are hard to sample

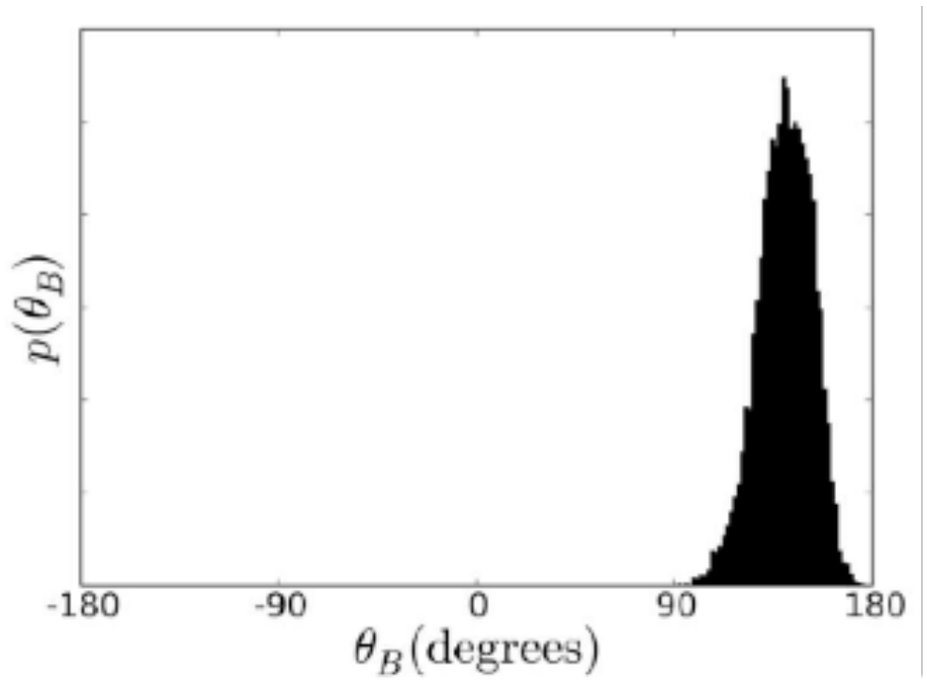


(a) Restrained orientation 1



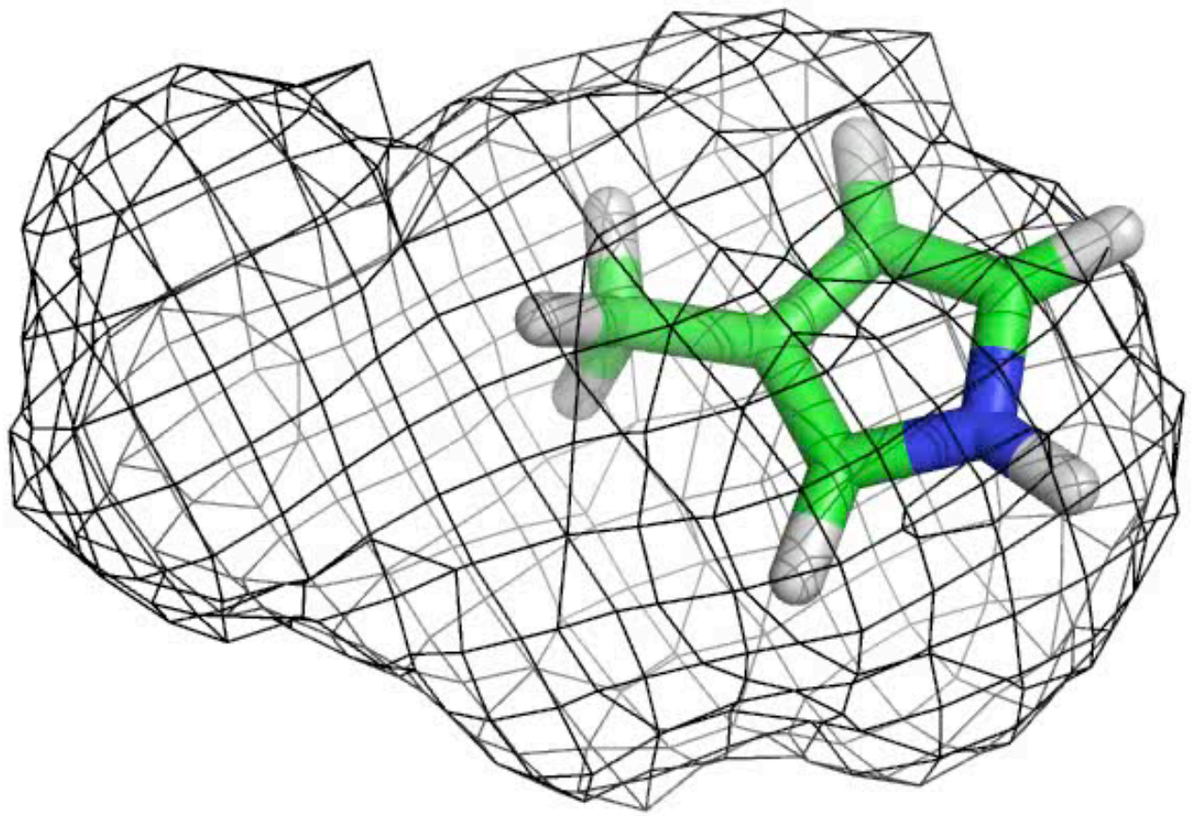
(b) Restrained orientation 2

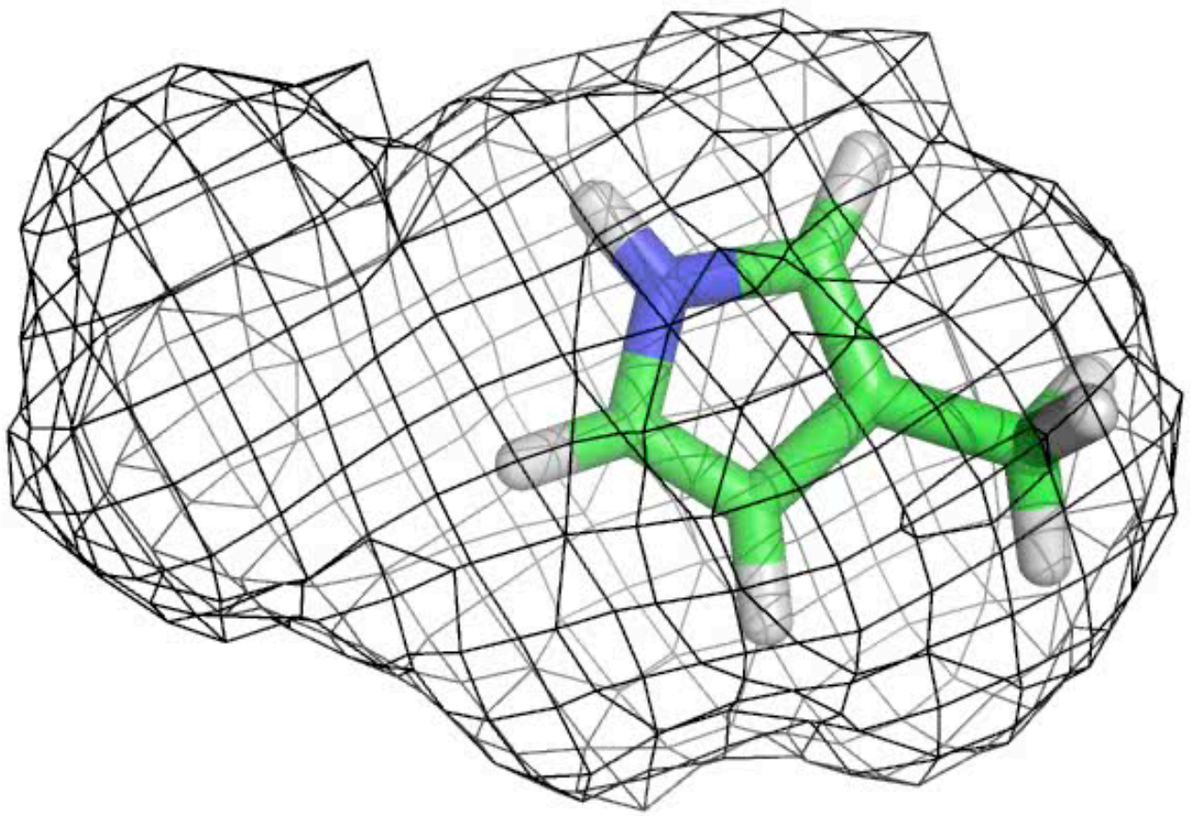




- Even 5 ns for each simulation is not enough

Solution: Separate calculations for different orientations

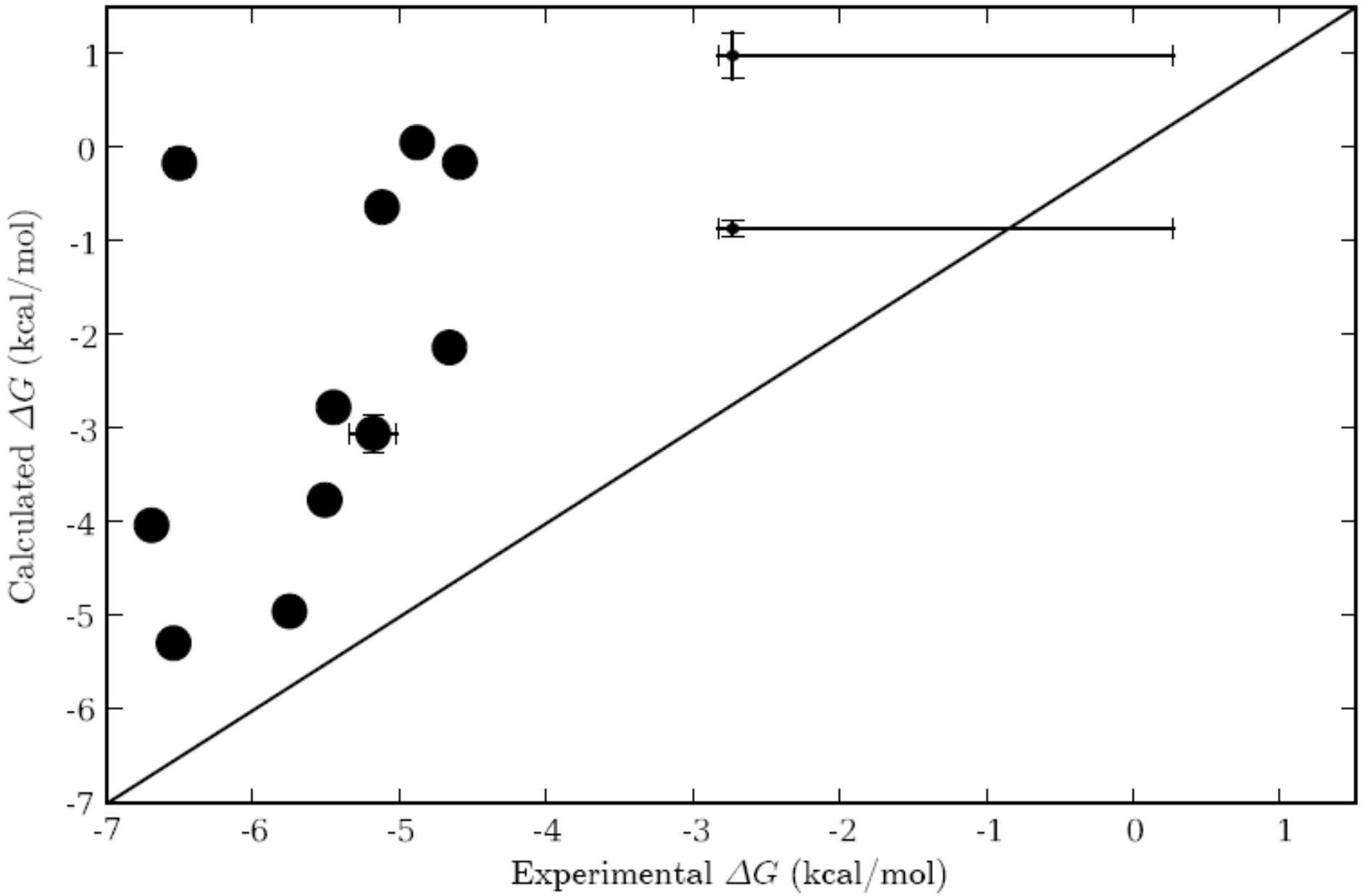




$$\Delta G^o = -k_B T \; ln \; (e^{-\frac{\Delta G_1^o}{k_B T}} + e^{-\frac{\Delta G_2^o}{k_B T}})$$

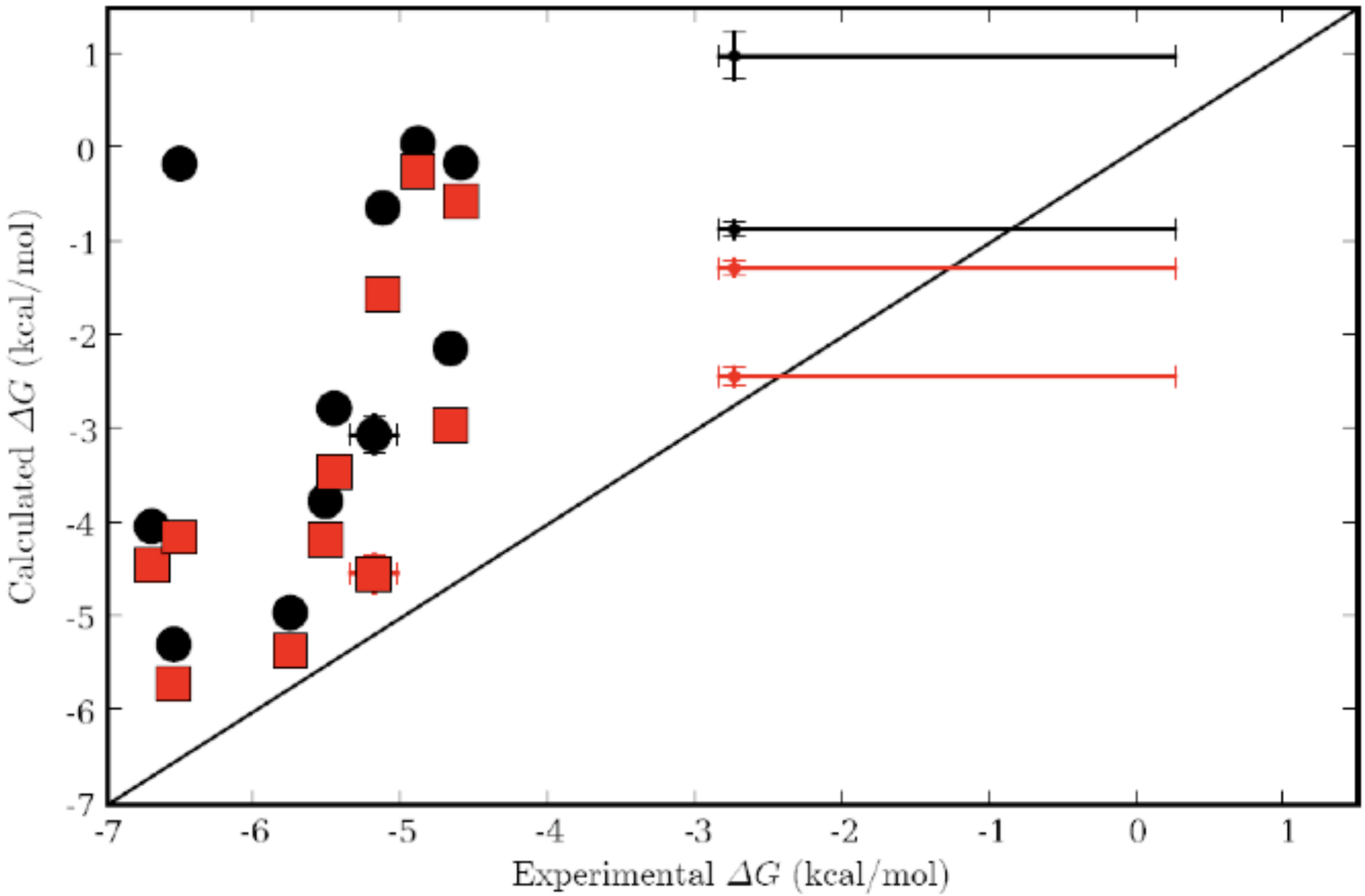
Mobley et al, *J. Chem. Phys.* 125:084902 (2006)

Step 1: Docking + free energy calculations was promising



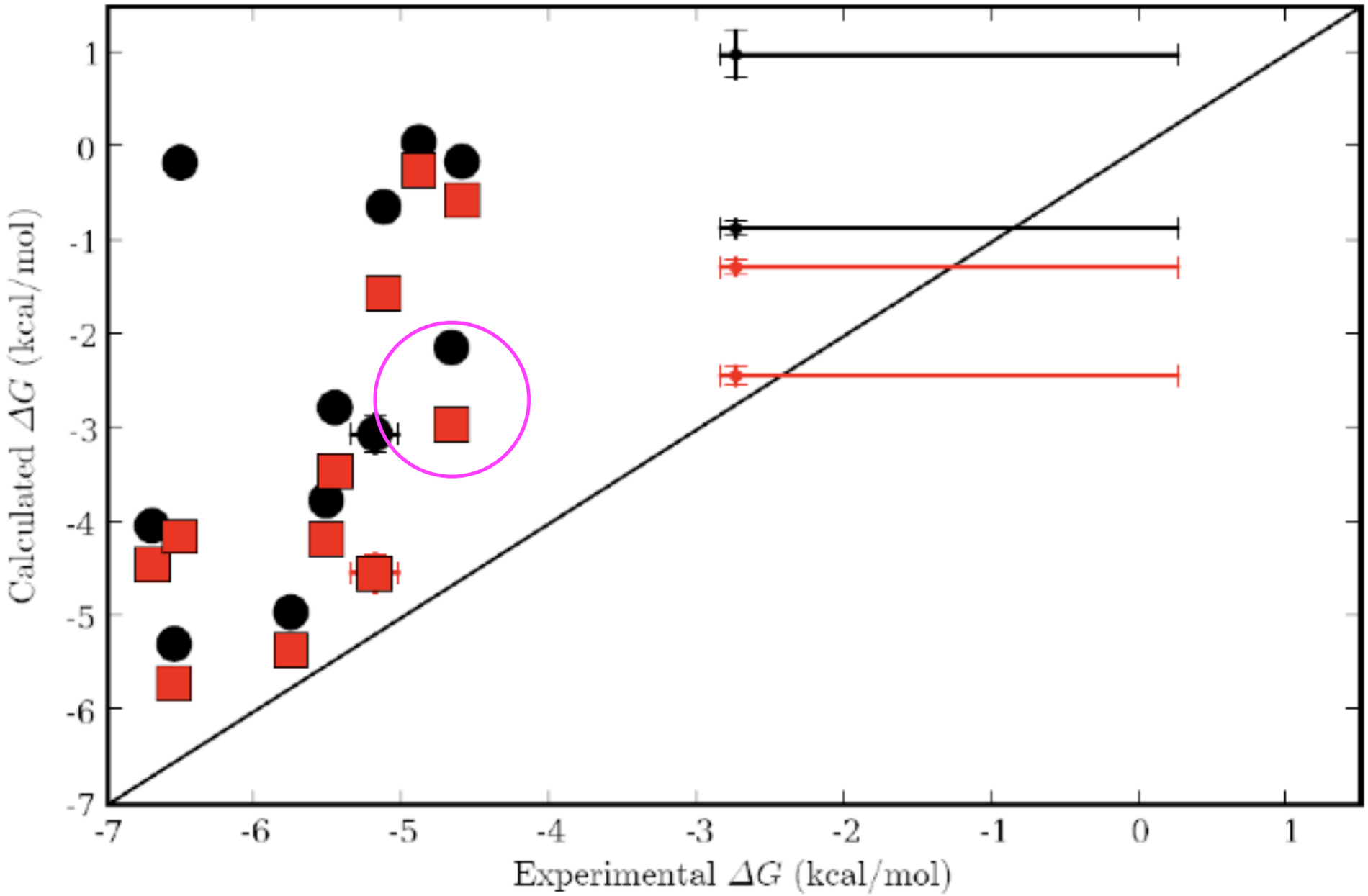
RMS error 3.5 kcal/mol

Step 2: Free energies improve when multiple orientations are included



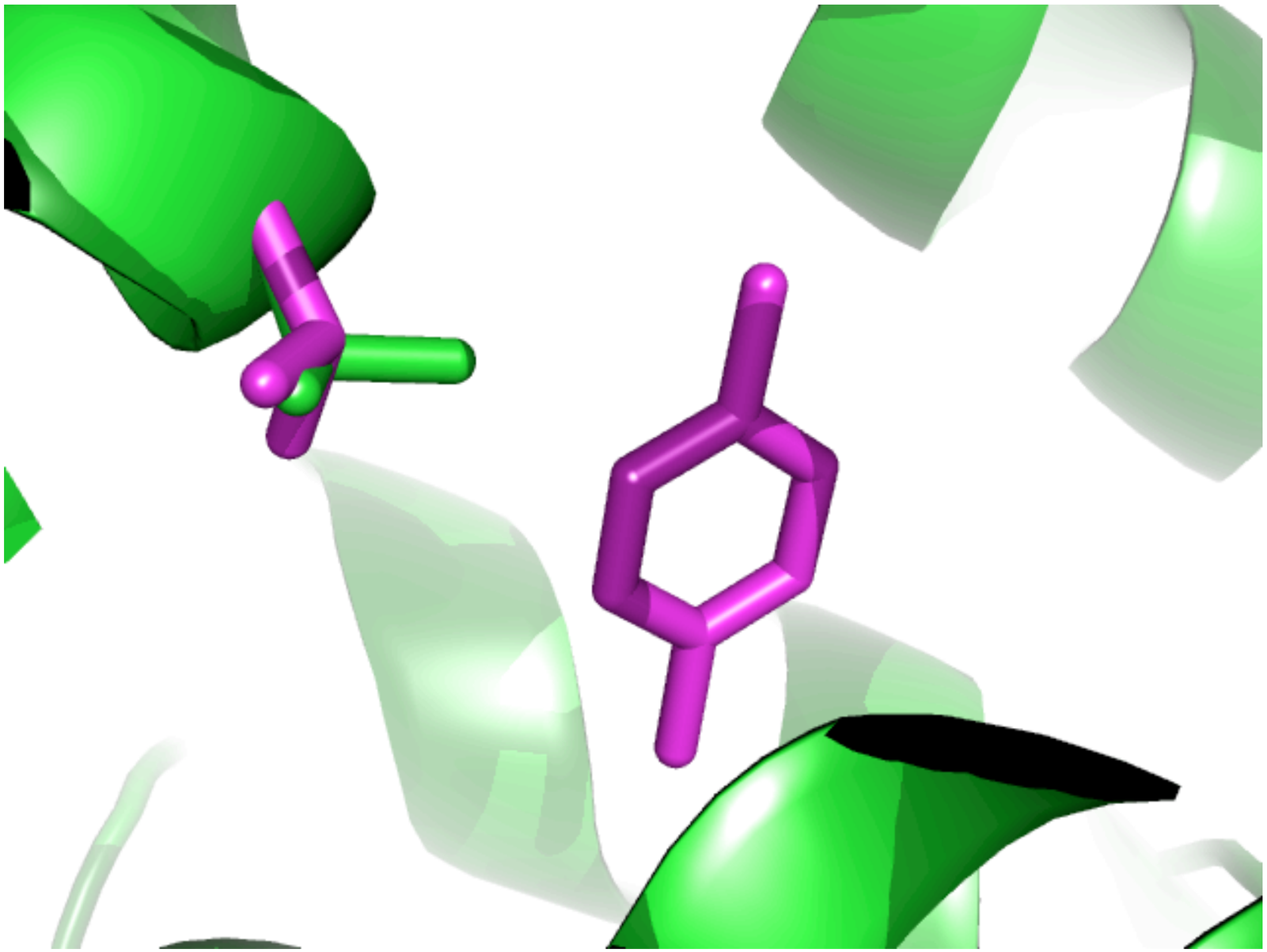
RMS error 2.5 kcal/mol

Step 2: Free energies improve when multiple orientations are included



RMS error 2.5 kcal/mol

Remaining problems partly due to conformational change

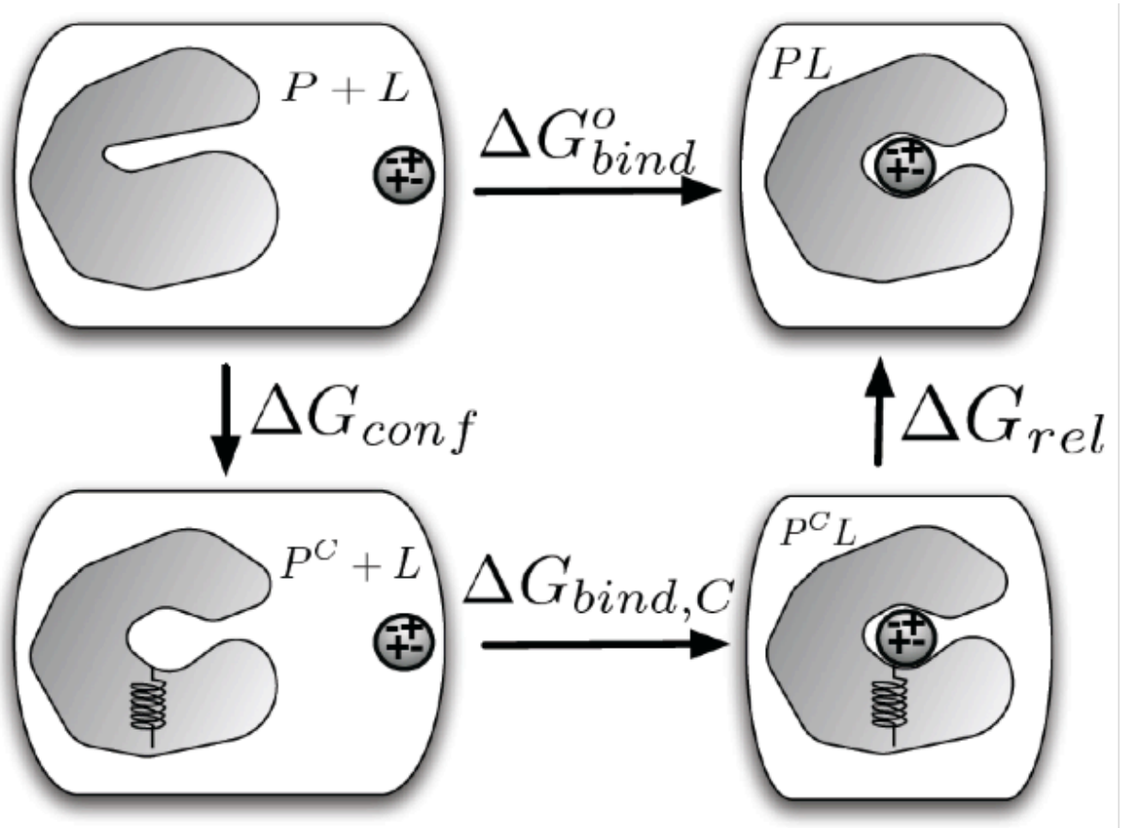


From apo structure: ΔG=-3.0+/-0.1 kcal/mol

From holo structure: ΔG=-7.3+/-0.1 kcal/mol

Experiment: -4.6 kcal/mol

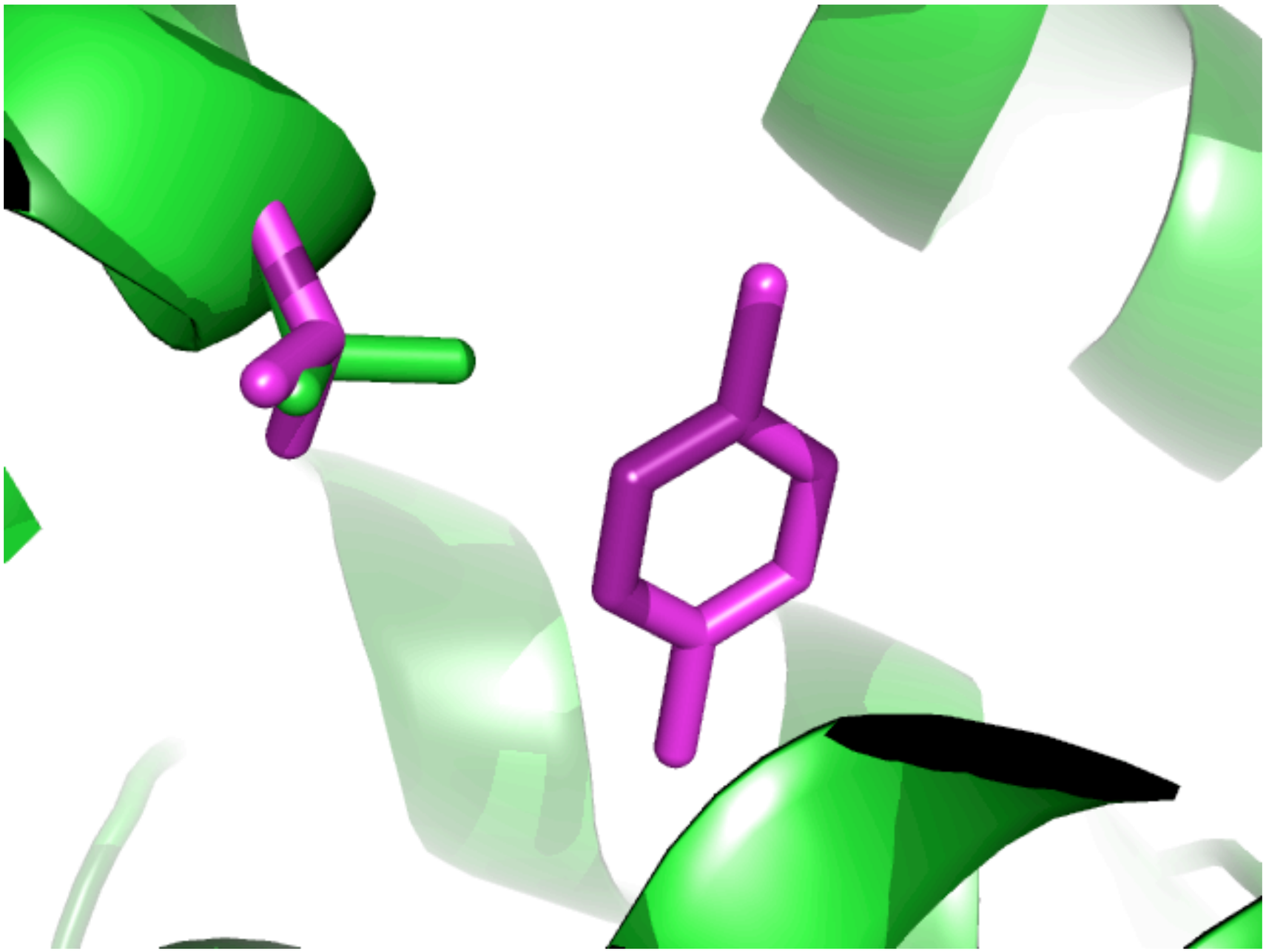
Solution: Confine-and-release



- Restrict protein
- Bind ligand
- Release protein

D. Mobley, J. Chodera, K. Dill, *J. Chem. Theory. Comput.* **3**: 1231 (2007).

Confine-and-release approach works

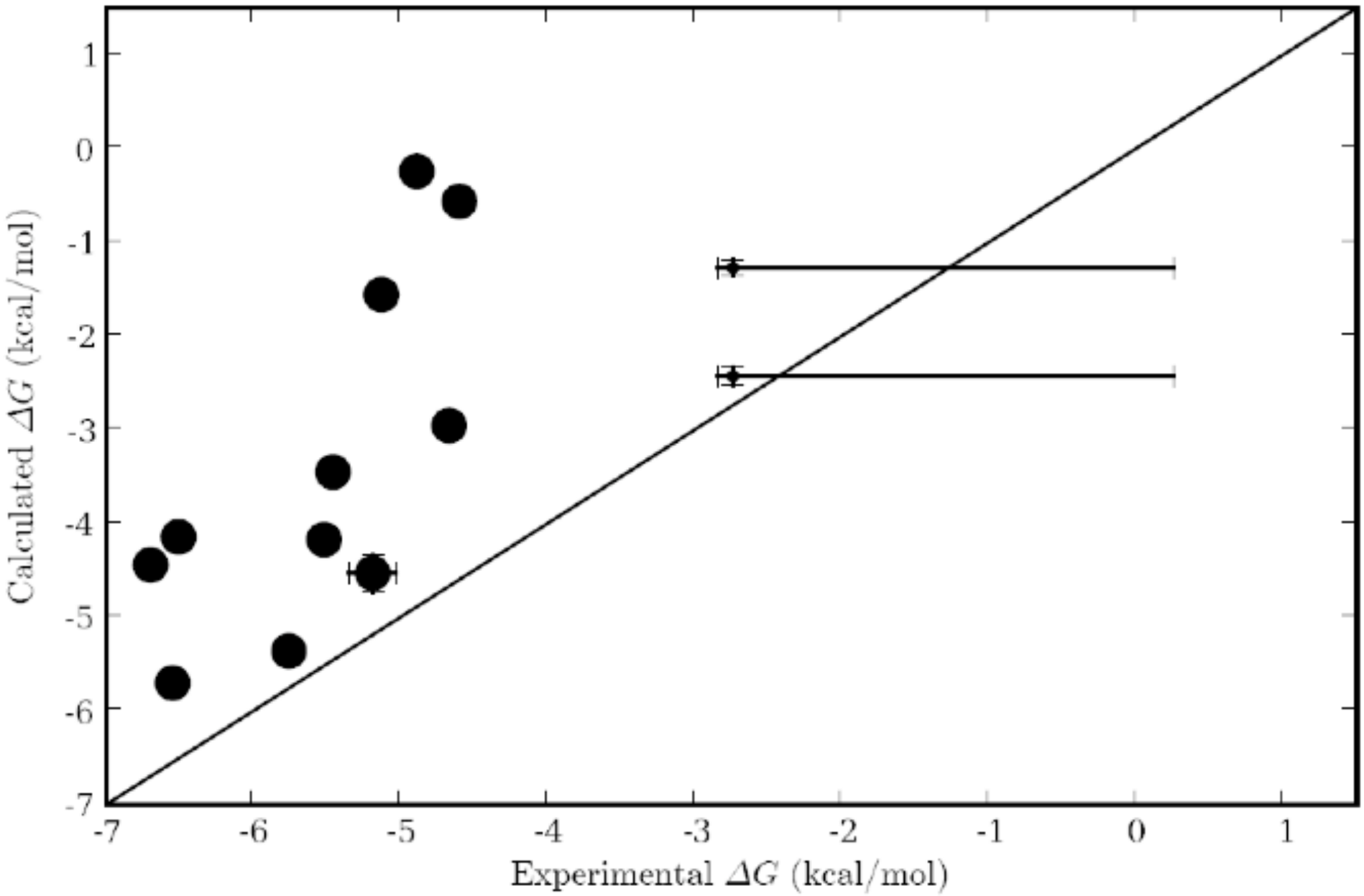


From apo structure: ΔG=-3.5+/-0.2 kcal/mol

From holo structure: ΔG=-3.4+/-0.2 kcal/mol

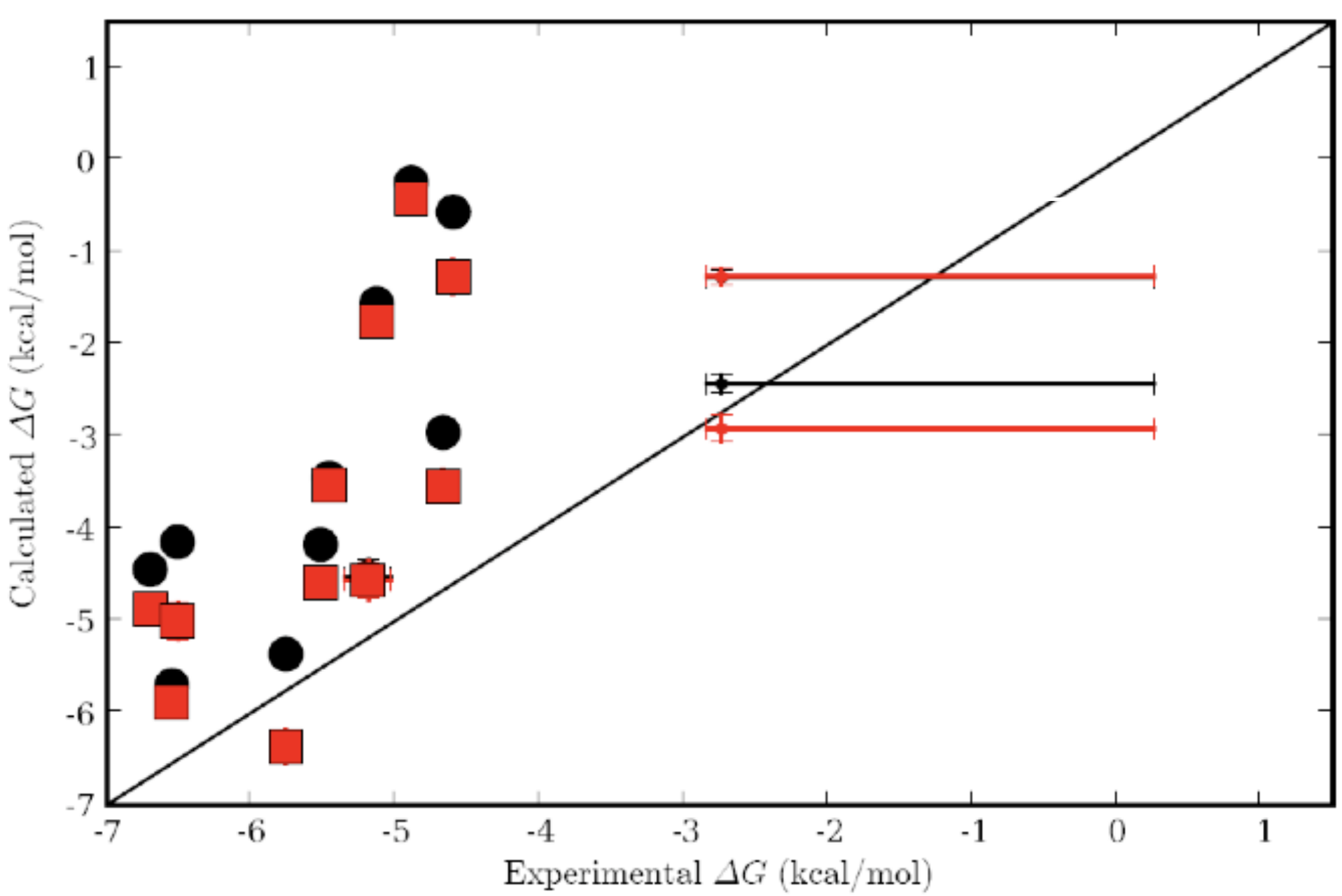
Experiment: -4.6 kcal/mol

Step 2: Without confine-and-release



RMS error 2.5 kcal/mol

Step 3: Confine-and-release helps



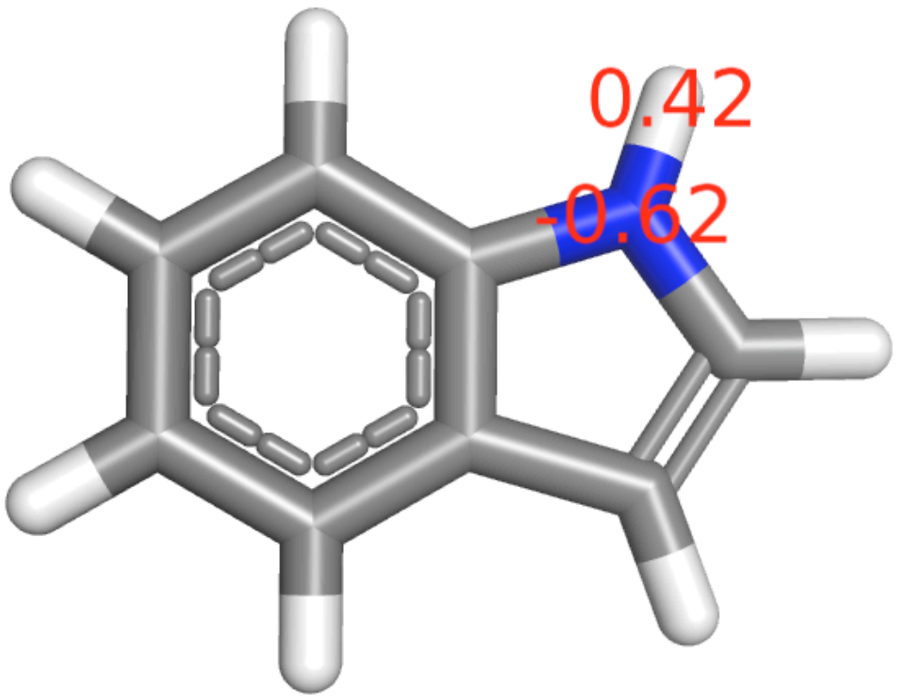
RMS error 2.2 kcal/mol

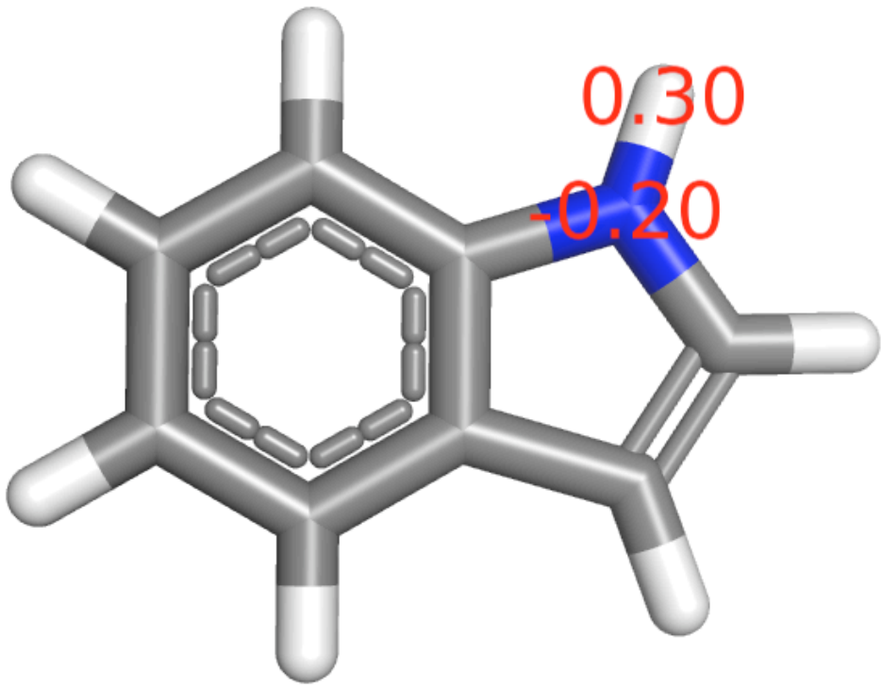
The remaining problems aren’t from sampling…
What else could cause them?

The remaining problems aren't from sampling…
What else could cause them?

Parameters?

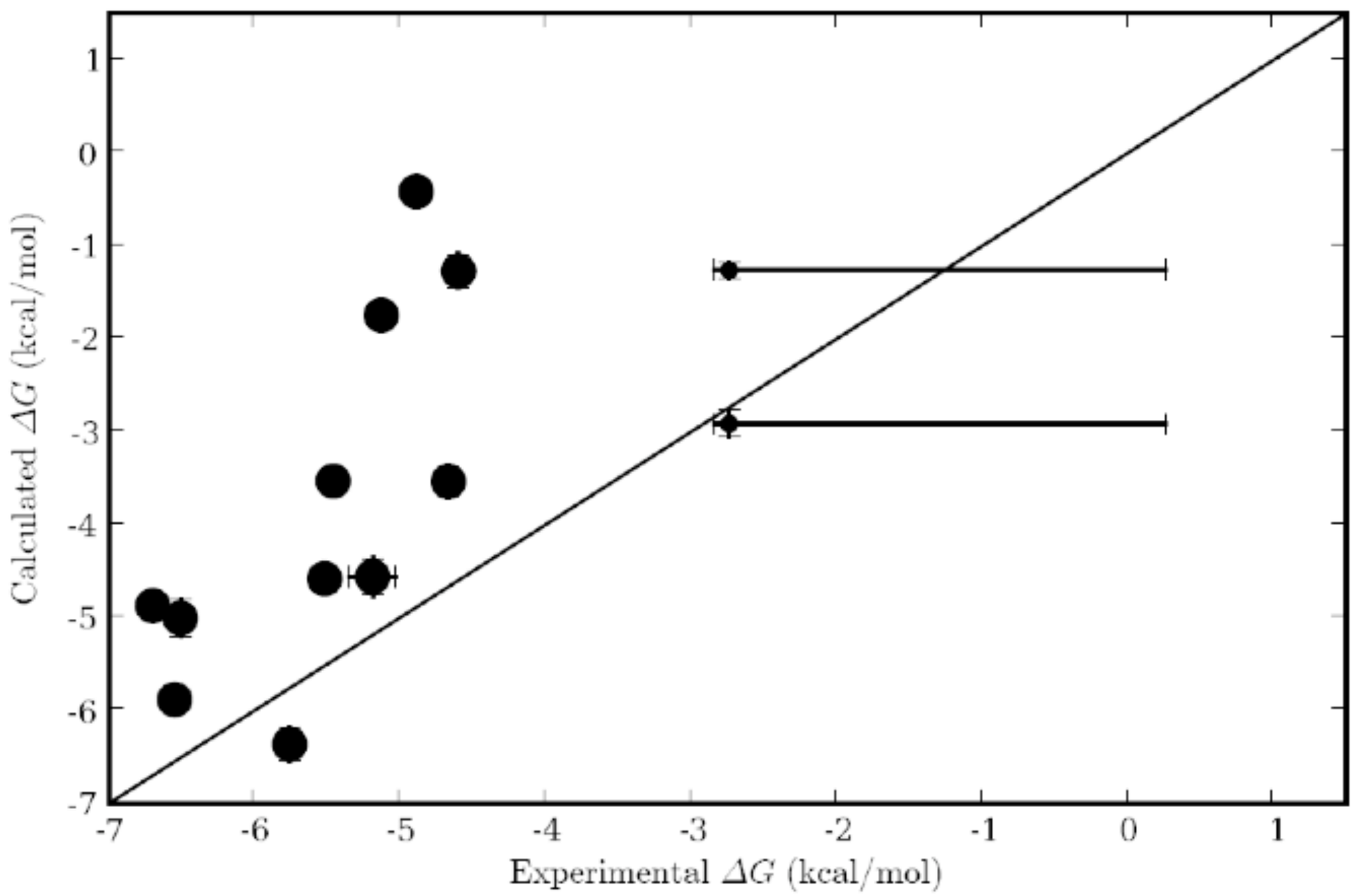
Partial charges are important





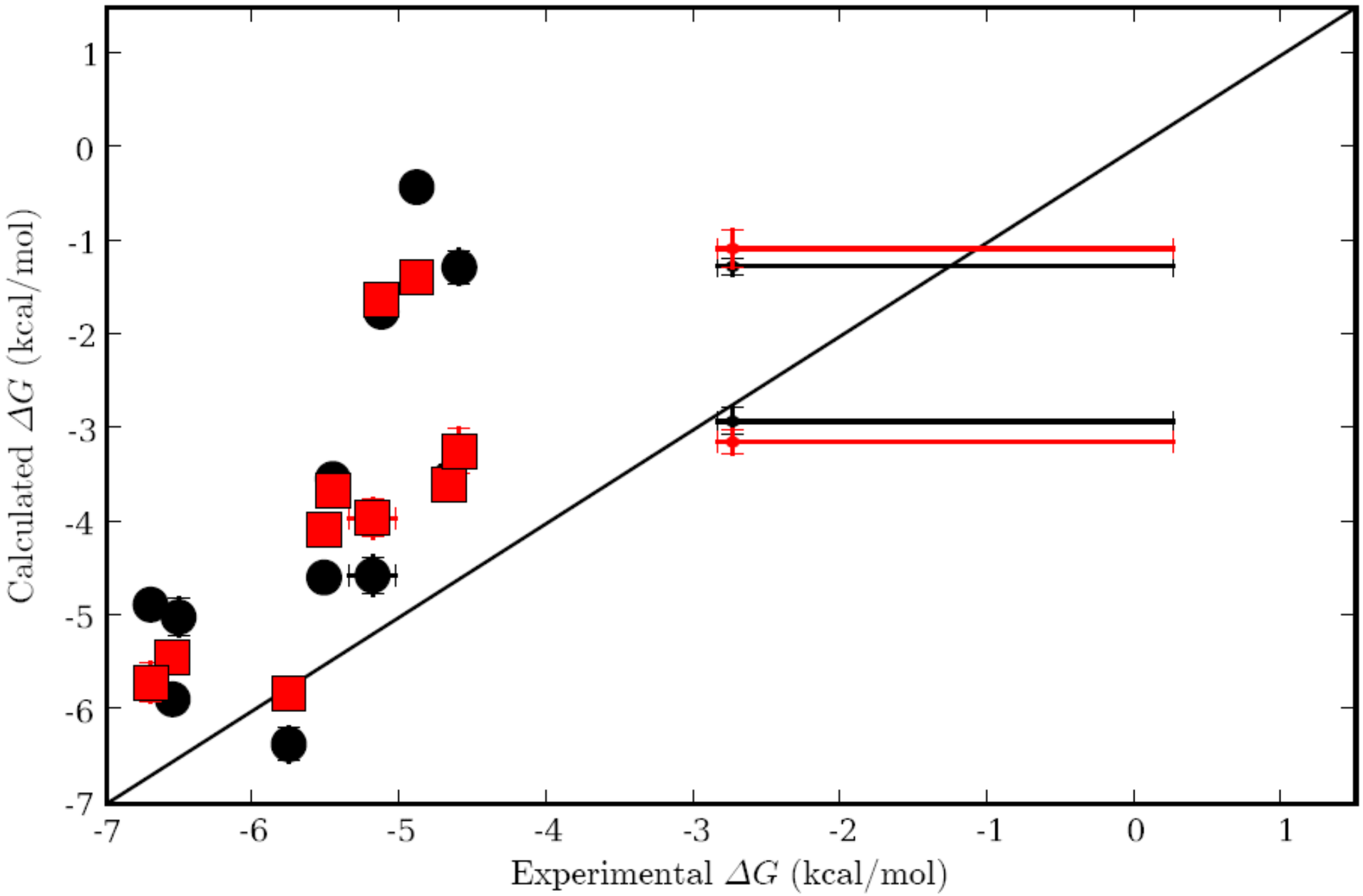
Mobley et al, *J. Phys. Chem. B.* 125:084902 (2007)

Step 3: With confine-and-release and original charges



RMS error 2.2 kcal/mol

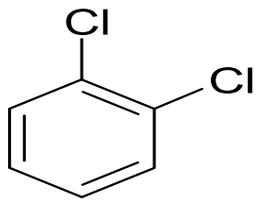
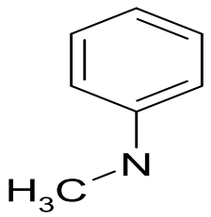
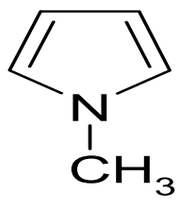
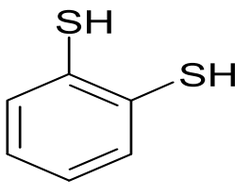
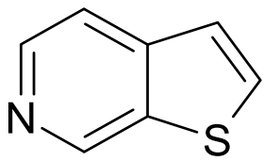
Step 4: Better partial charges improve agreement with experiment



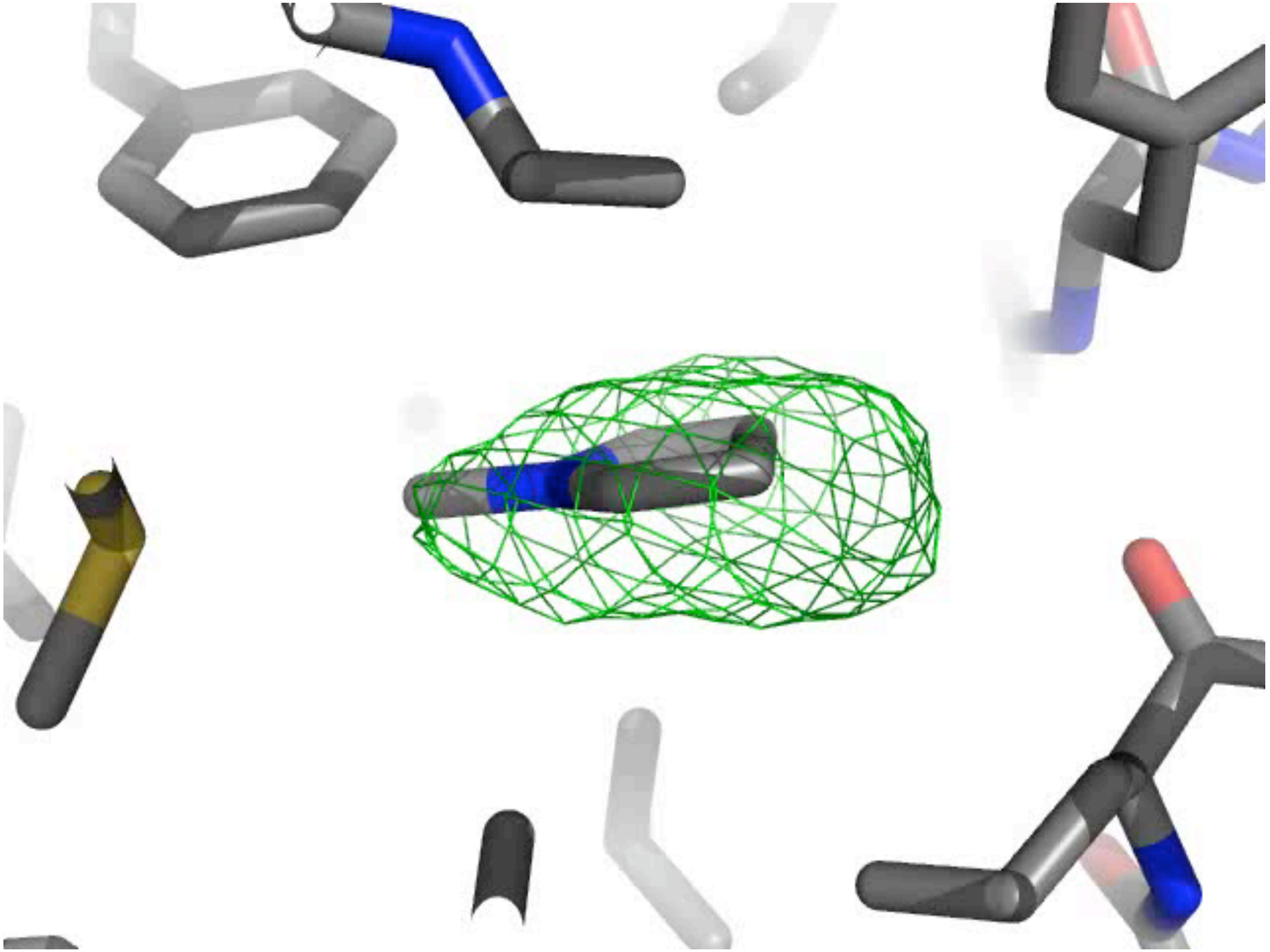
RMS error 1.8 kcal/mol

Putting it all together:

A blind test

Name	**Predicted ΔG (kcal/mol)**	**Expt. ΔG (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

Successfully predicts bound orientations

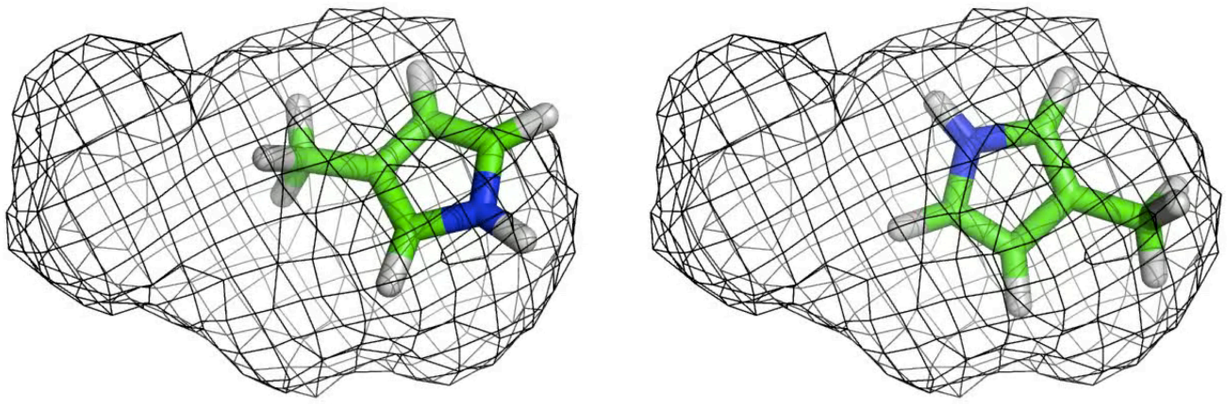


1-Methylpyrrole

Mobley et al., JMB, 2007

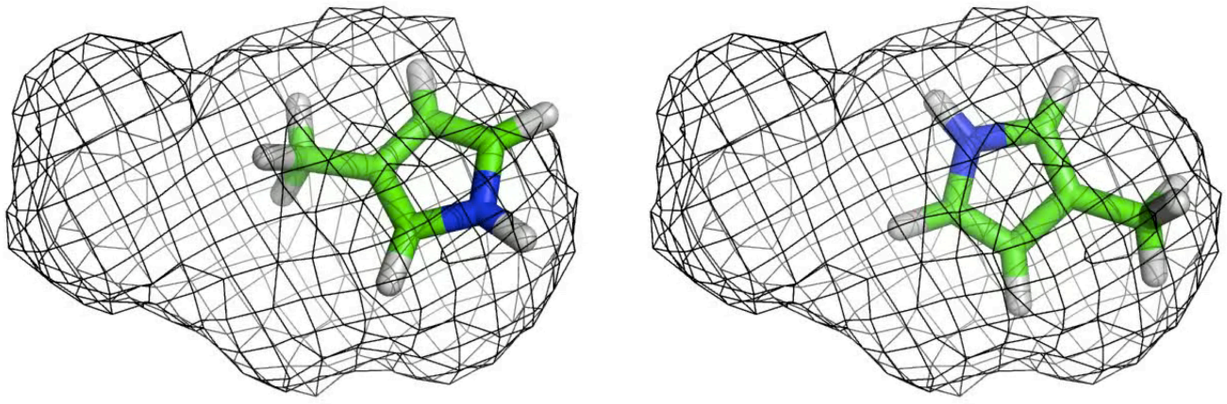
What did we learn?

What did we learn?

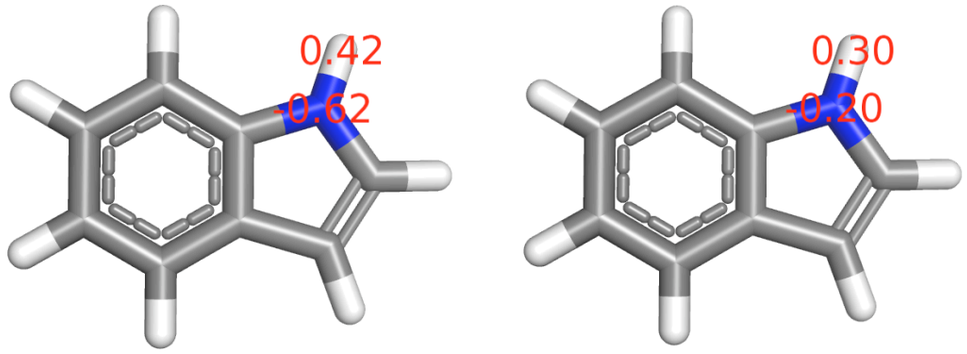


Consider multiple orientations

What did we learn?

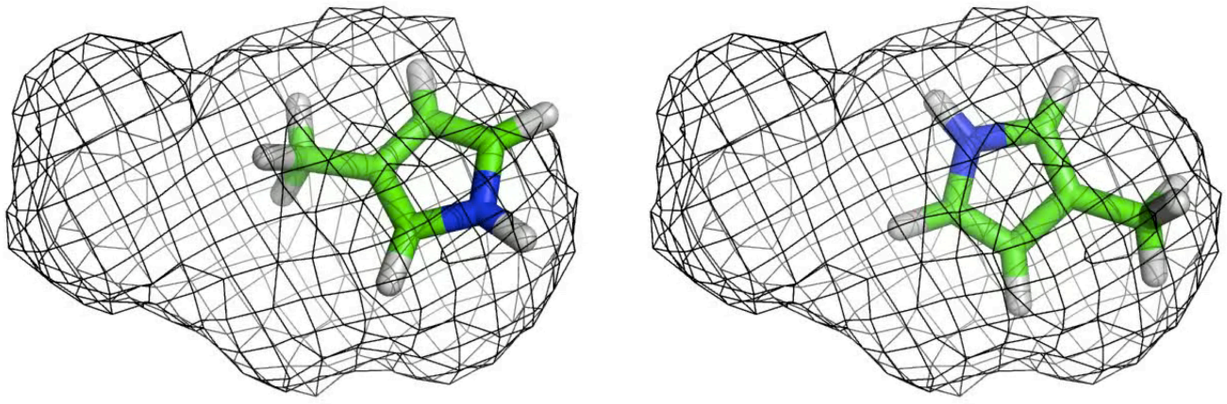


Consider multiple orientations

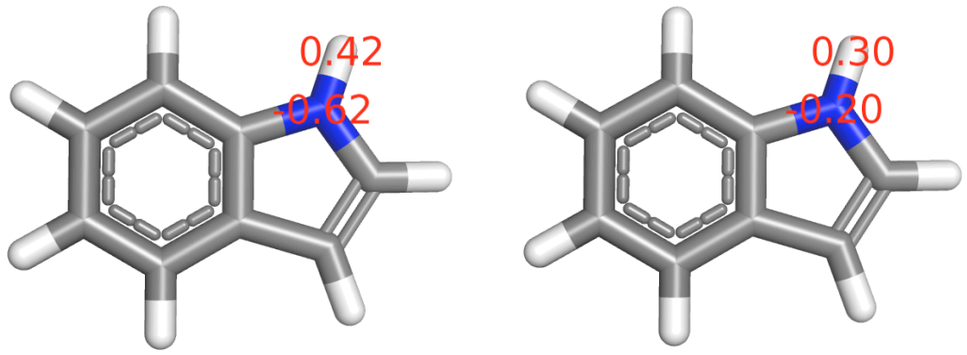


Charge model is important

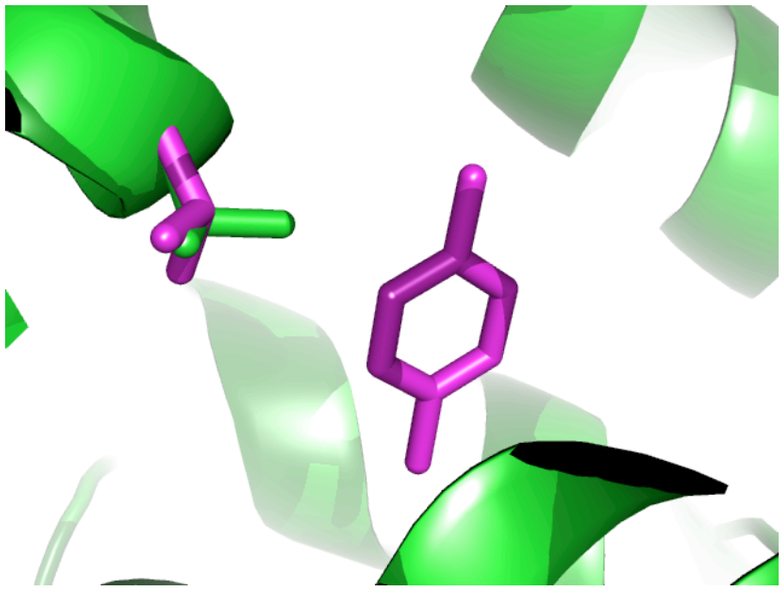
What did we learn?



Consider multiple orientations

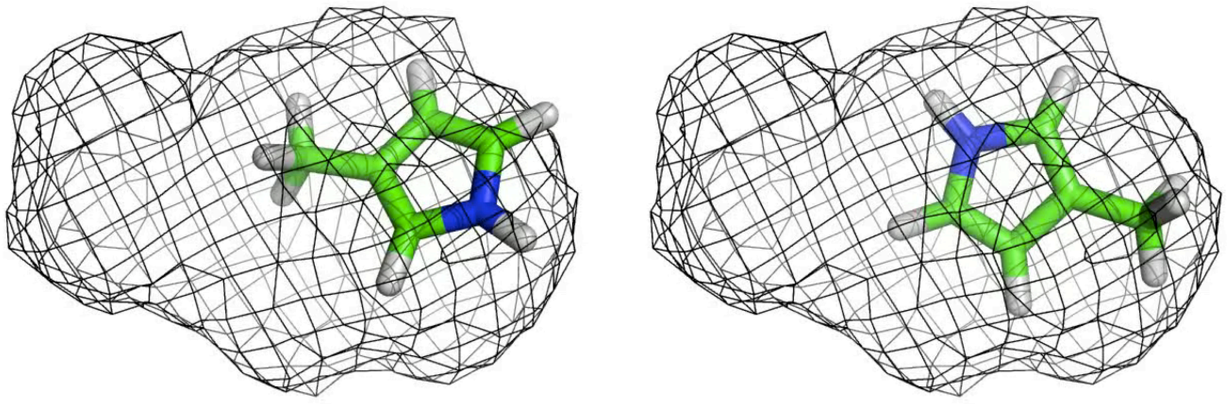


Charge model is important

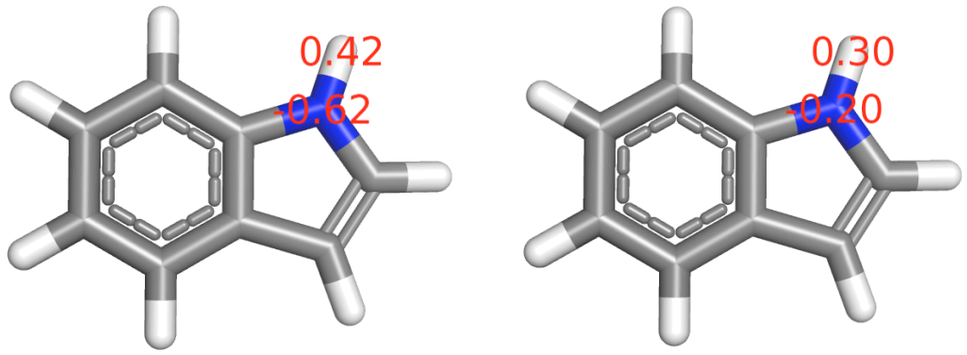


Conformational change is key

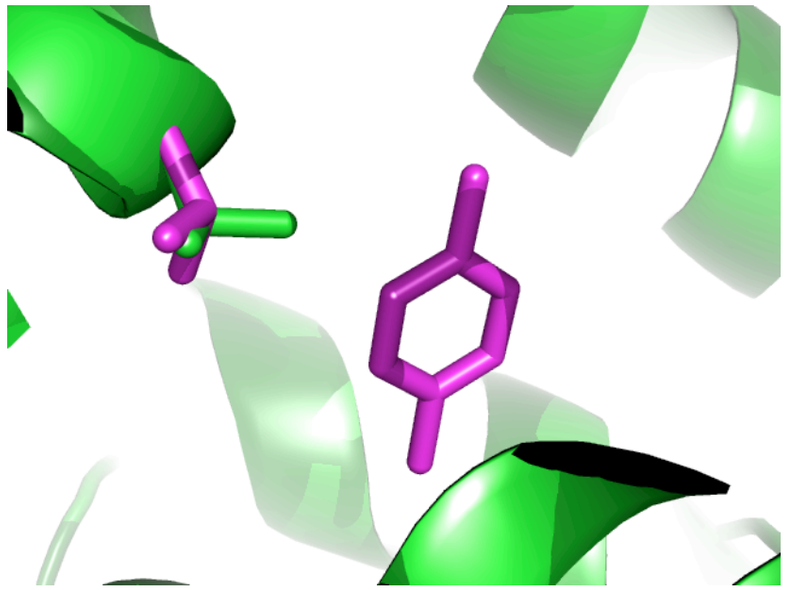
What did we learn?



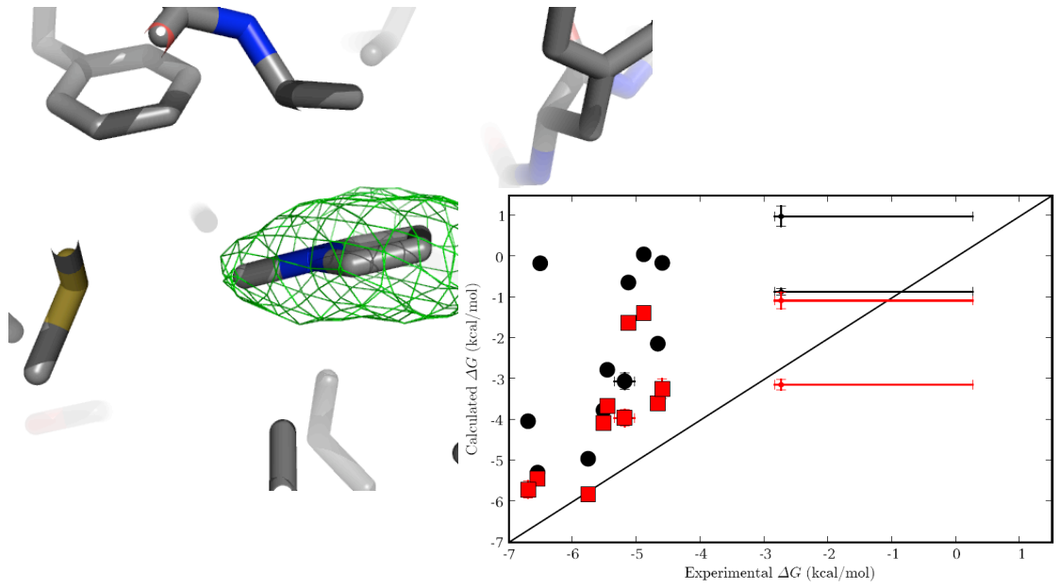
Consider multiple orientations



Charge model is important

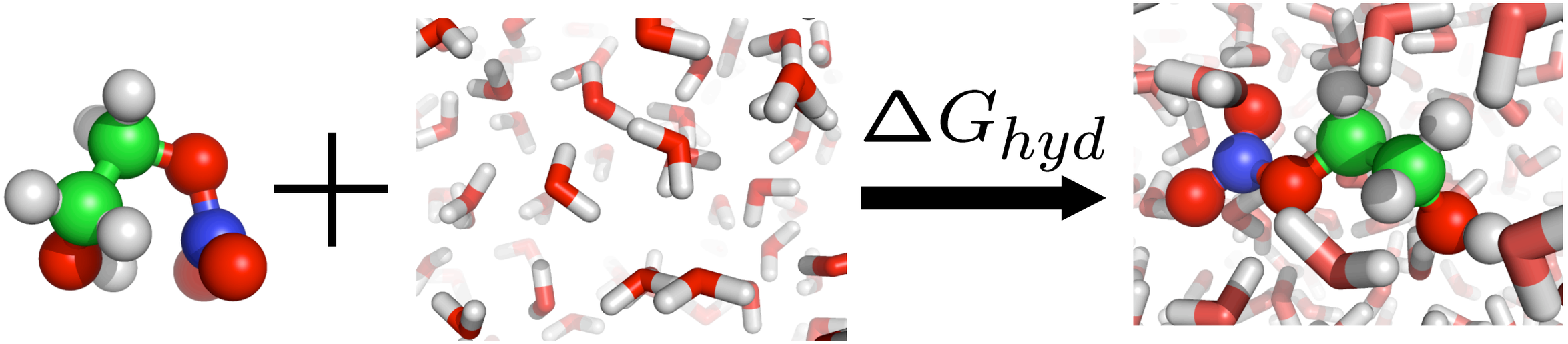


Conformational change is key

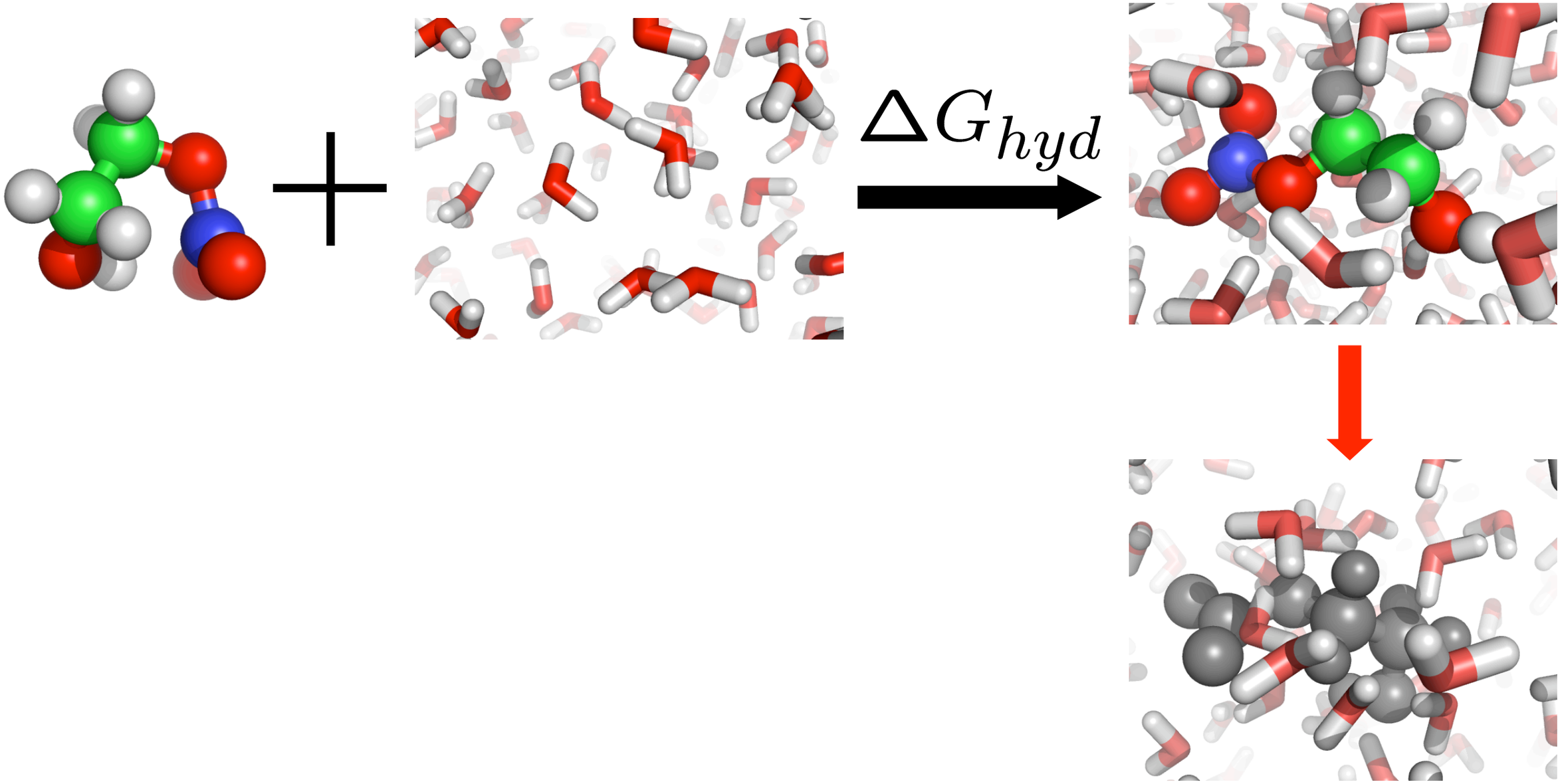


Systematic improvements possible with physics-based modeling

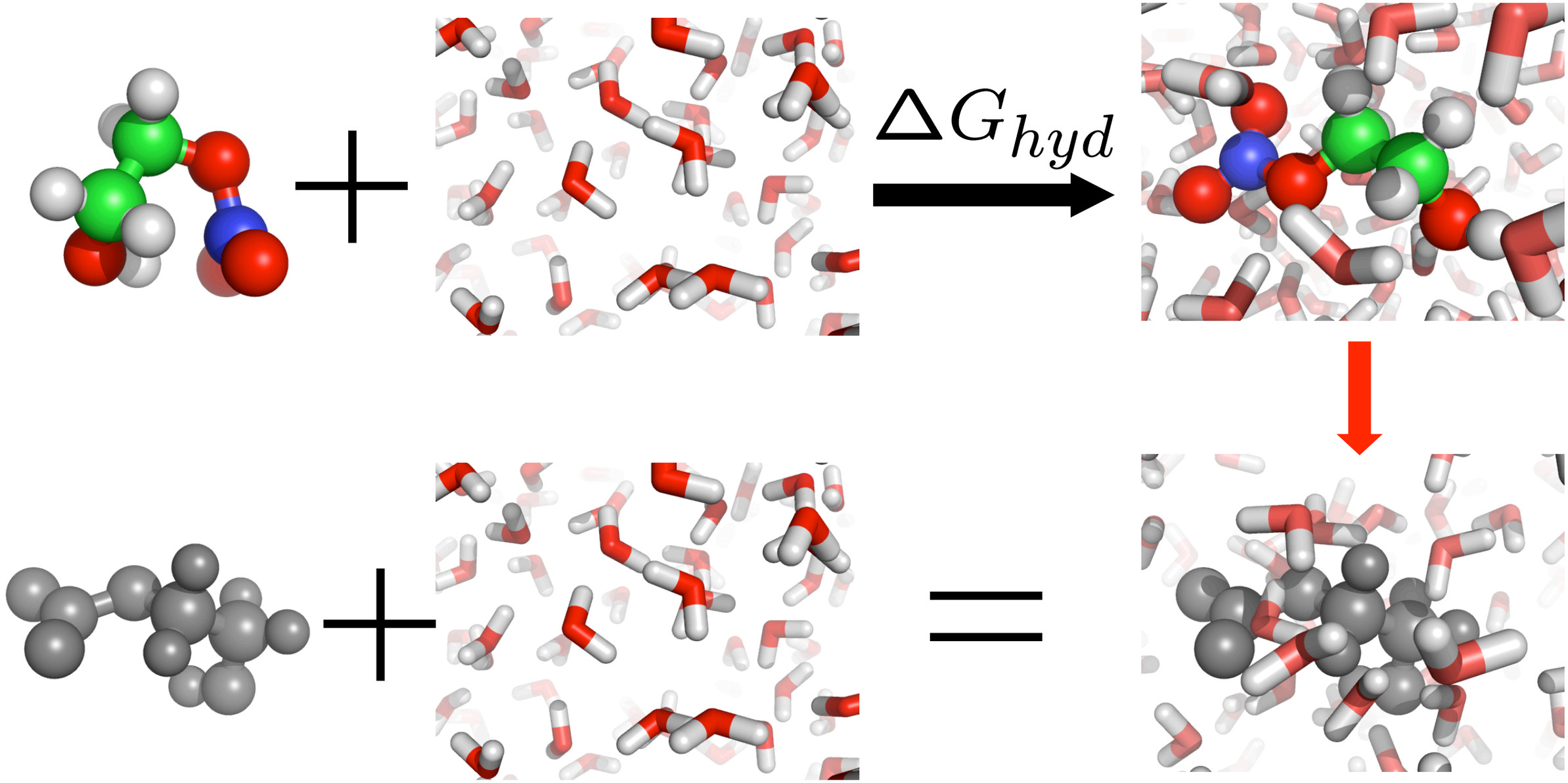
Alchemical calculations yield rigorous hydration free energies



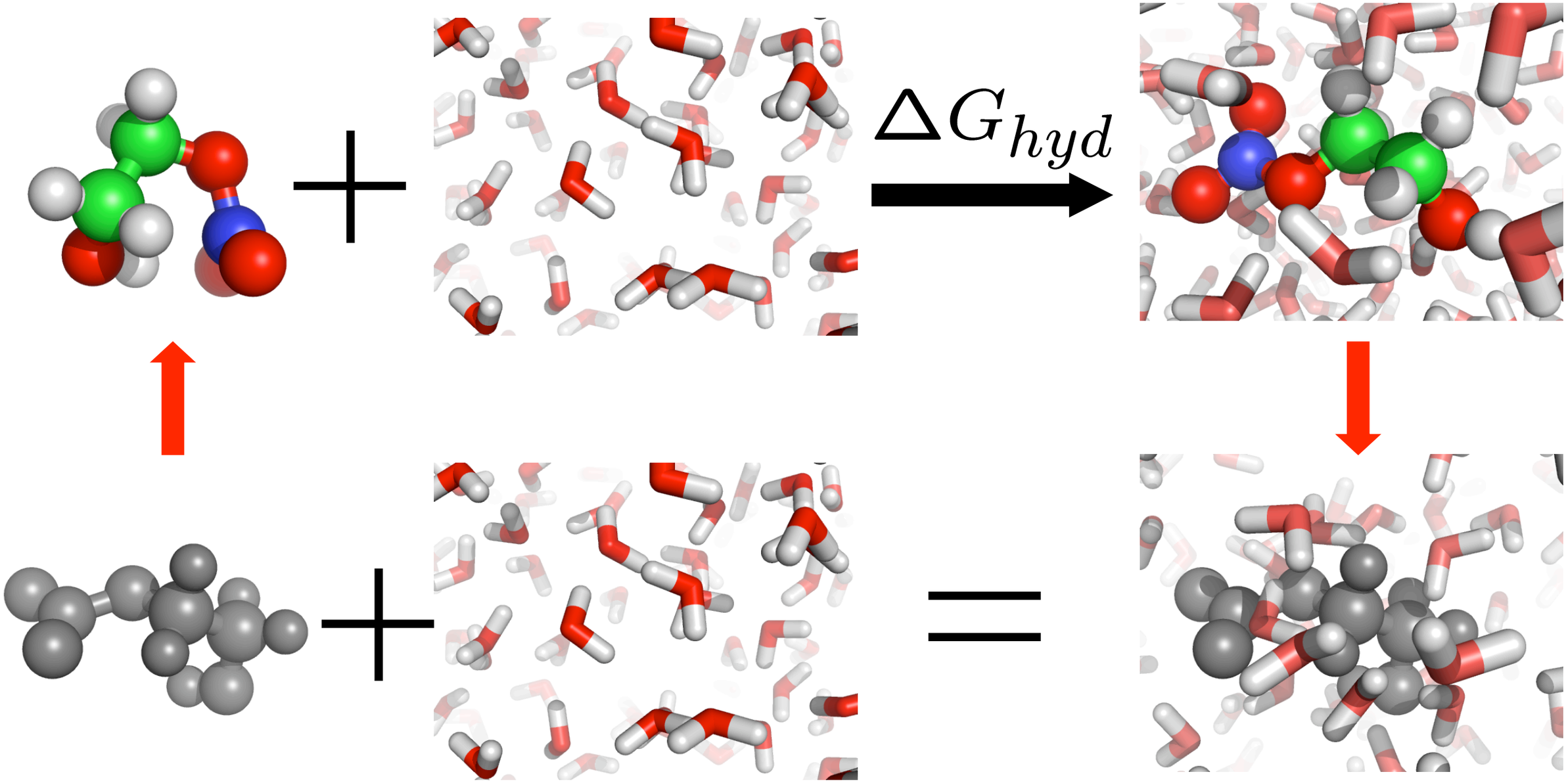
Alchemical calculations yield rigorous hydration free energies



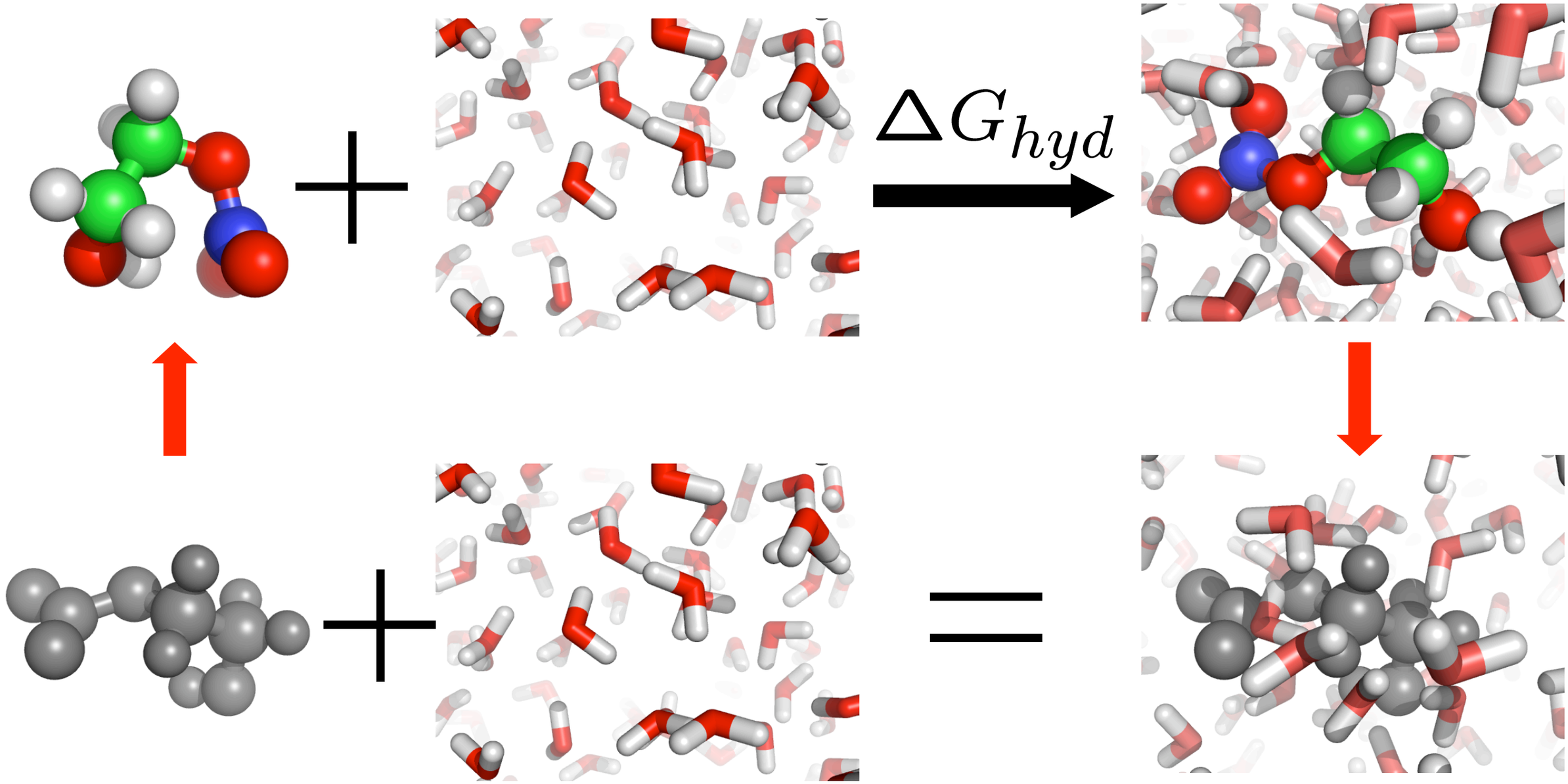
Alchemical calculations yield rigorous hydration free energies



Alchemical calculations yield rigorous hydration free energies

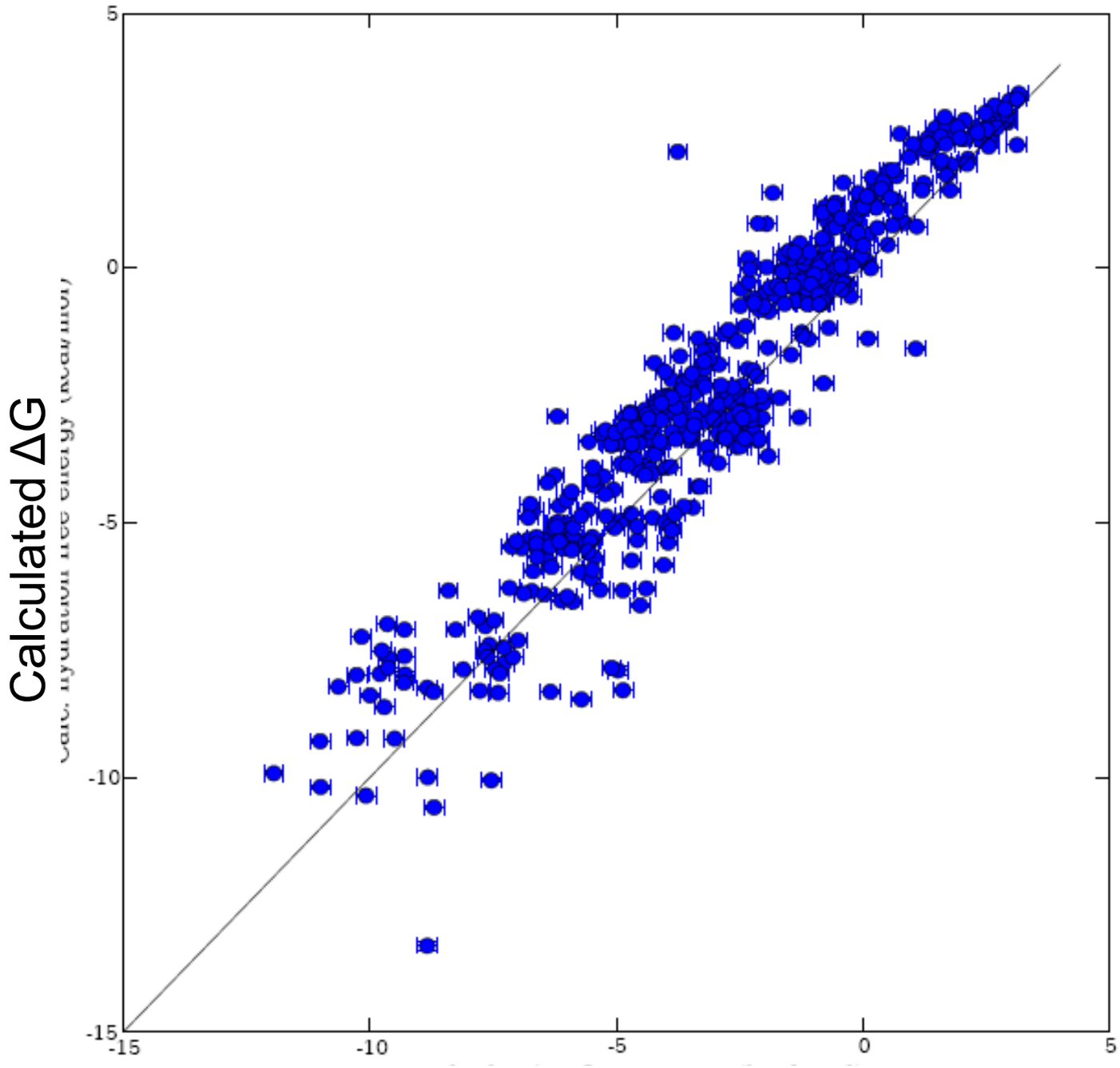


Alchemical calculations yield rigorous hydration free energies



- We use explicit solvent MD and alchemical free energy calculations analyzed with BAR

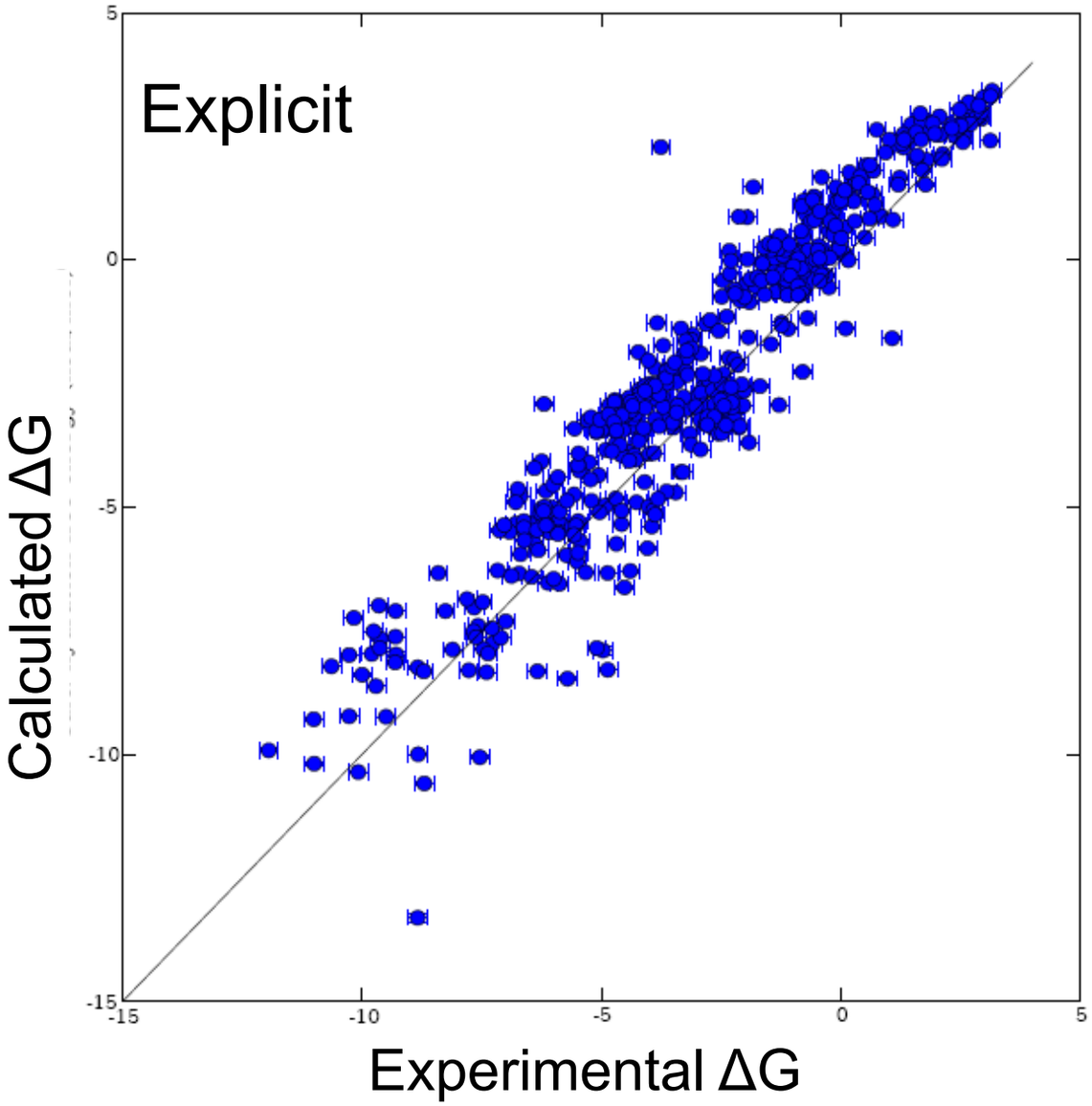
Calculated hydration free energies correlate well with experiment



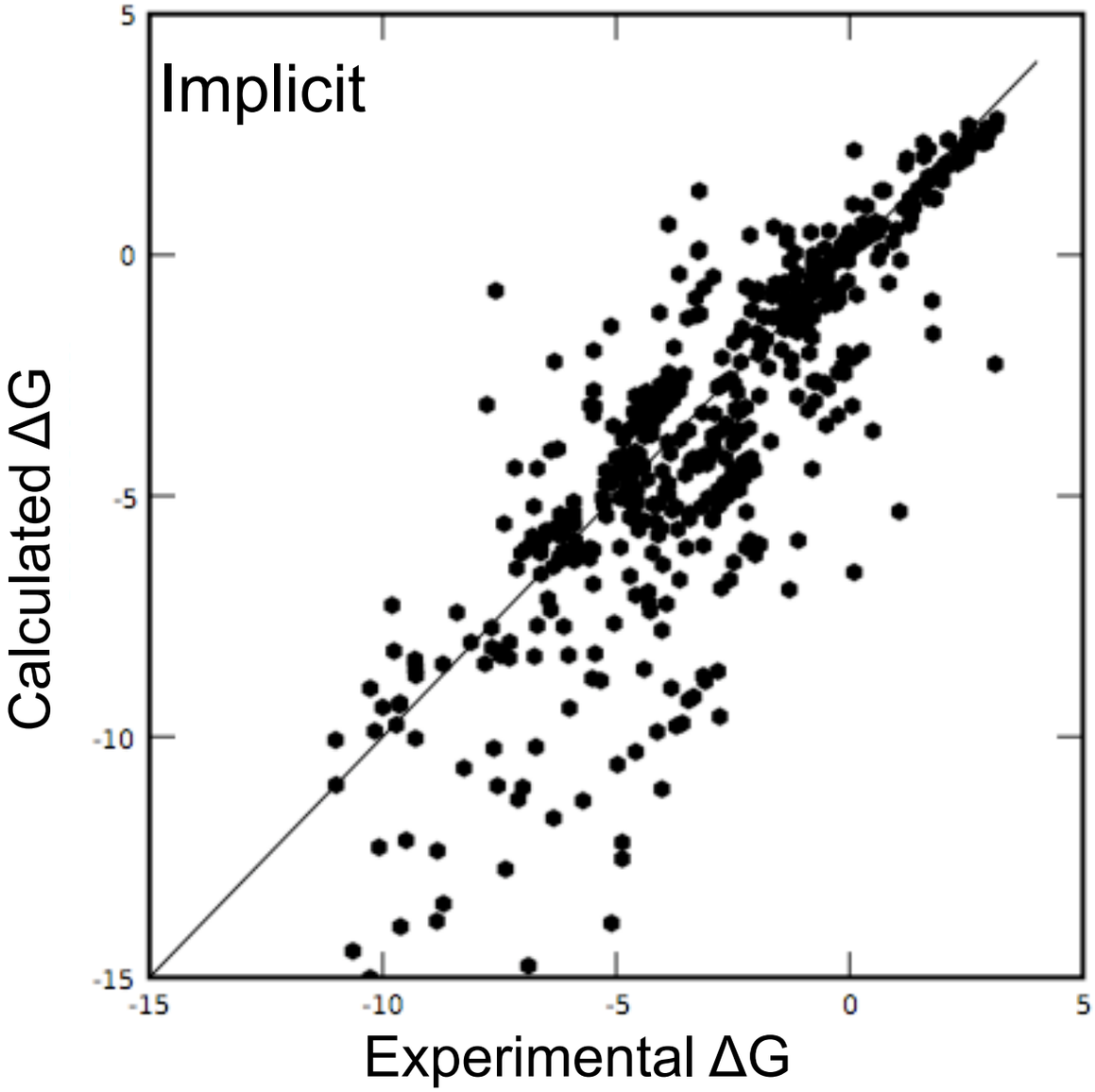
- (AM1-BCC v1 charges)
- Statistics:
  - RMS error 1.23+/-0.01 kcal/mol
  - R^2 = 0.89+/-0.01
  - Mean error 0.651+/-0.002 kcal/mol

Mobley, Cooper, Bayly, Shirts, and Dill, submitted, 2008.

Explicit solvent gives more accurate results than implicit solvent



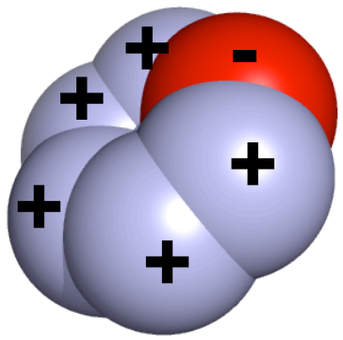
RMS error 1.23+/-0.01 kcal/mol
R^2 = 0.89+/-0.01



RMS errors 2.0 to 2.4 kcal/mol
R^2 = 0.68 to 0.77

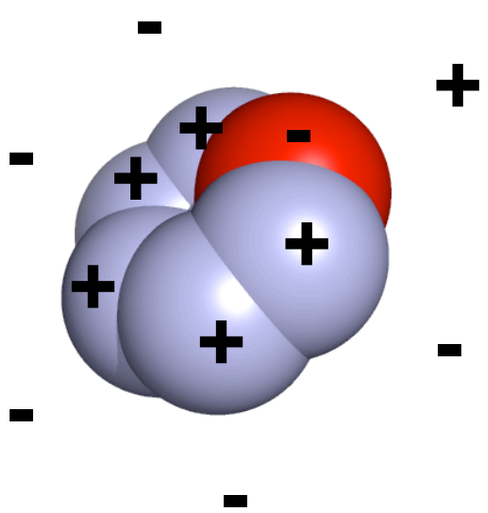
Mobley, Chodera and Dill, J. Phys. Chem. B. **112: 938-946 (2008).**

Hydration free energies provide a probe of the underlying physics



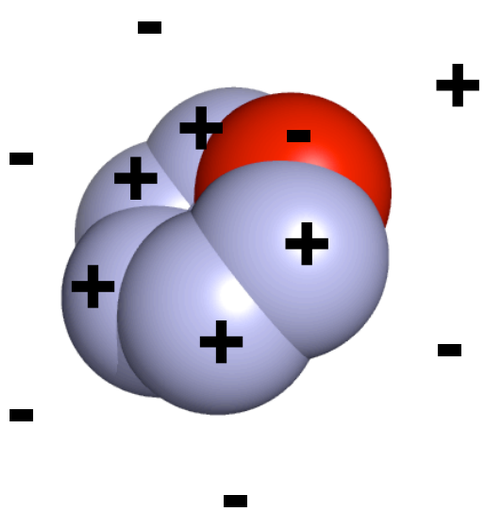
- Continuum solvent models

Hydration free energies provide a probe of the underlying physics



- Continuum solvent models

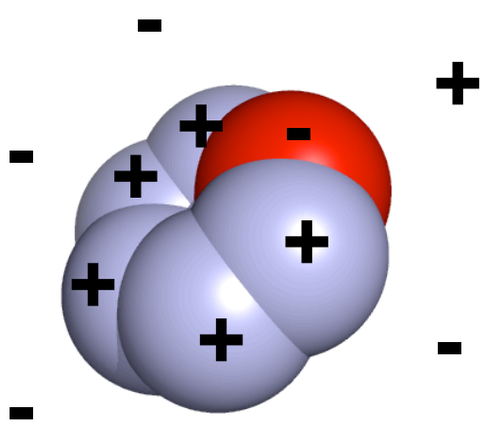
Hydration free energies provide a probe of the underlying physics



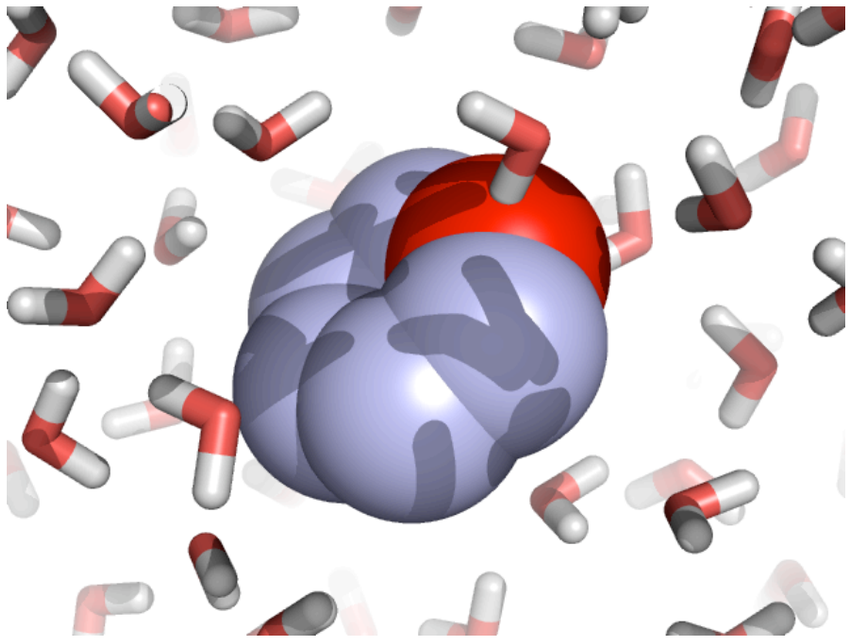
- Continuum solvent models

- Explicit solvent
  - Does the “average” response look like that of a continuum solvent model?
  - What about the nonpolar part?

Hydration free energies provide a probe of the underlying physics



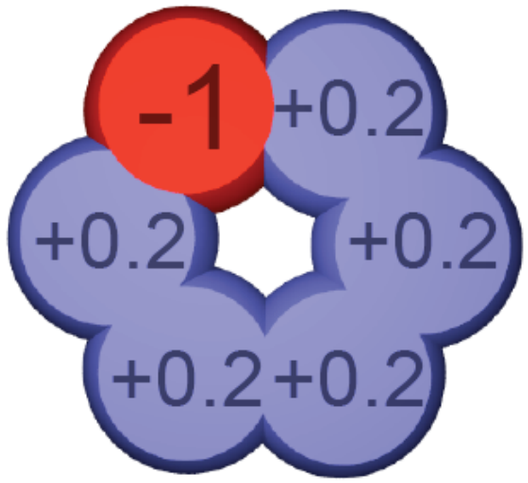
- Continuum solvent models



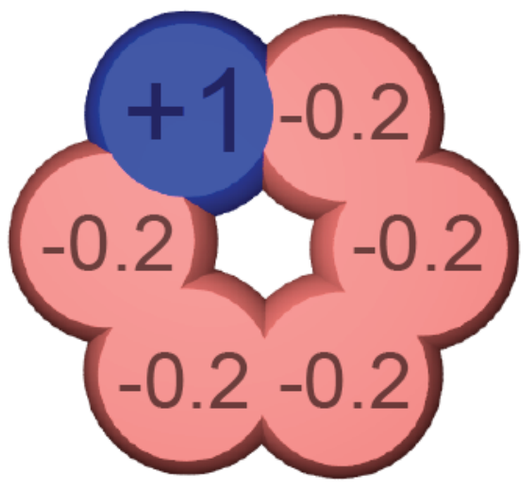
- Explicit solvent
  - Does the “average” response look like that of a continuum solvent model?
  - What about the nonpolar part?

Artificial solutes provide a sensitive probe of water's electrostatic response

- Consider two artificial ring-shaped solutes:



Hydration free energy:
-26.01 kcal/mol

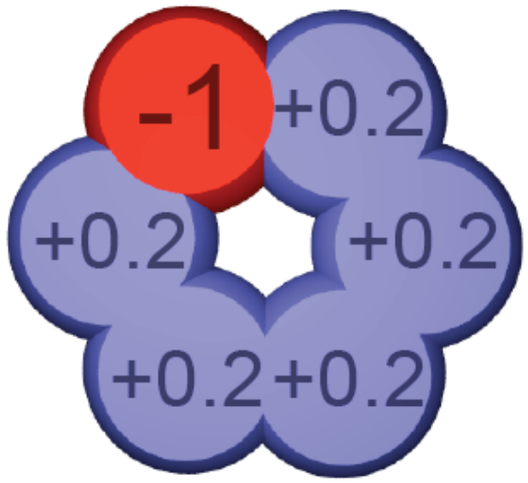


Hydration free energy:
-16.38 kcal/mol

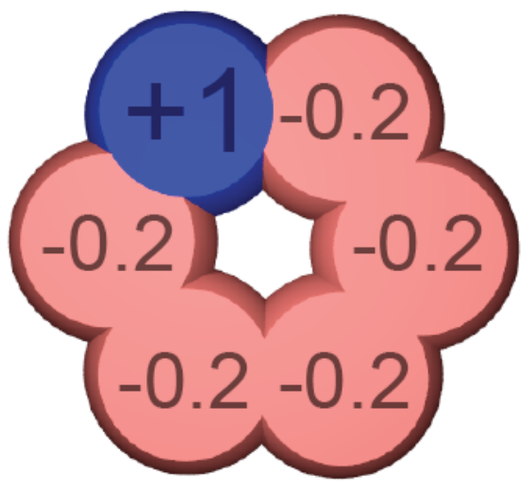
Mobley, Chodera and Dill, J. Phys. Chem. B. **112:2405-2414 (2008).**

Artificial solutes provide a sensitive probe of water's electrostatic response

- Consider two artificial ring-shaped solutes:



Hydration free energy:
-26.01 kcal/mol



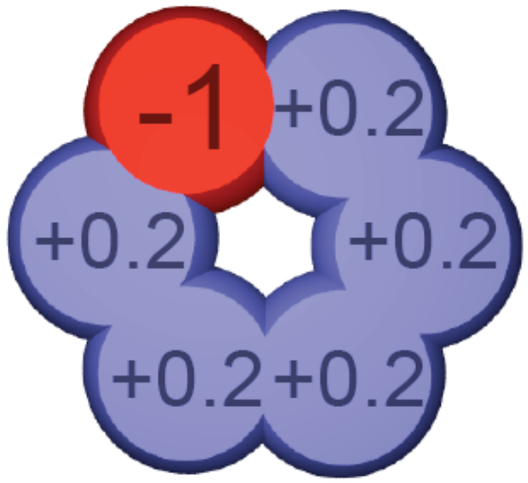
Hydration free energy:
-16.38 kcal/mol

Difference (hydration free energy asymmetry):
-9.63 kcal/mol

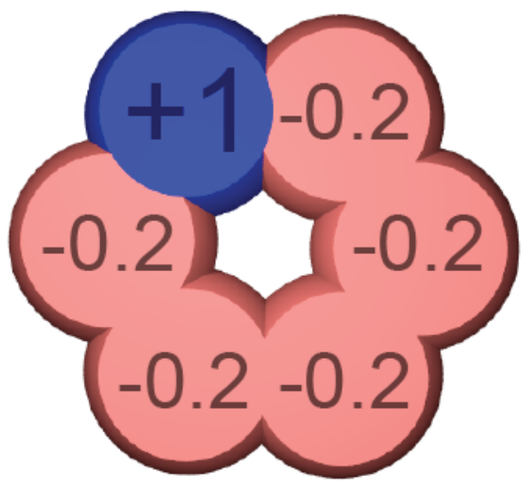
Mobley, Chodera and Dill, J. Phys. Chem. B. **112:2405-2414 (2008).**

Artificial solutes provide a sensitive probe of water's electrostatic response

- Consider two artificial ring-shaped solutes:



Hydration free energy:
-26.01 kcal/mol



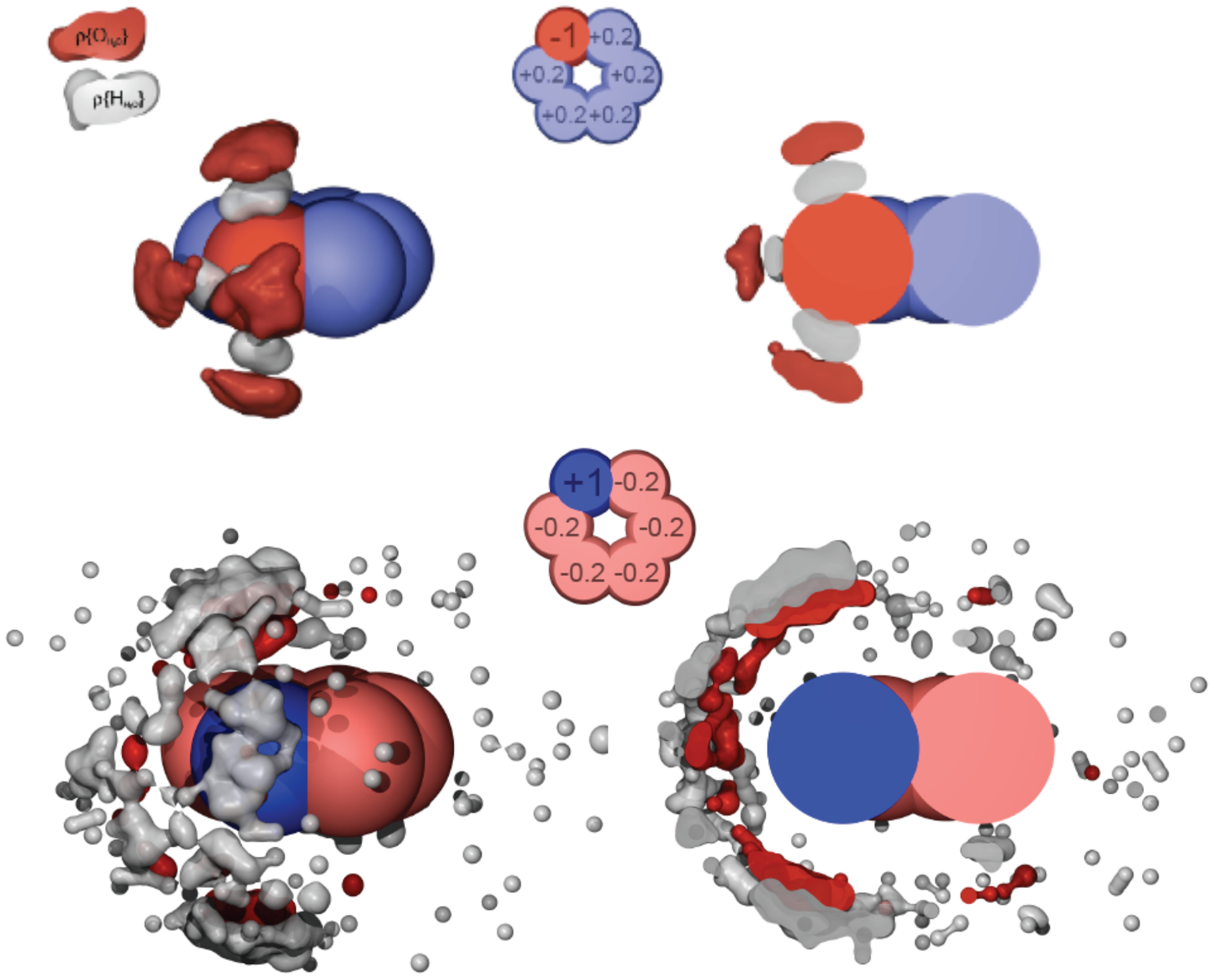
Hydration free energy:
-16.38 kcal/mol

Difference (hydration free energy asymmetry):
-9.63 kcal/mol

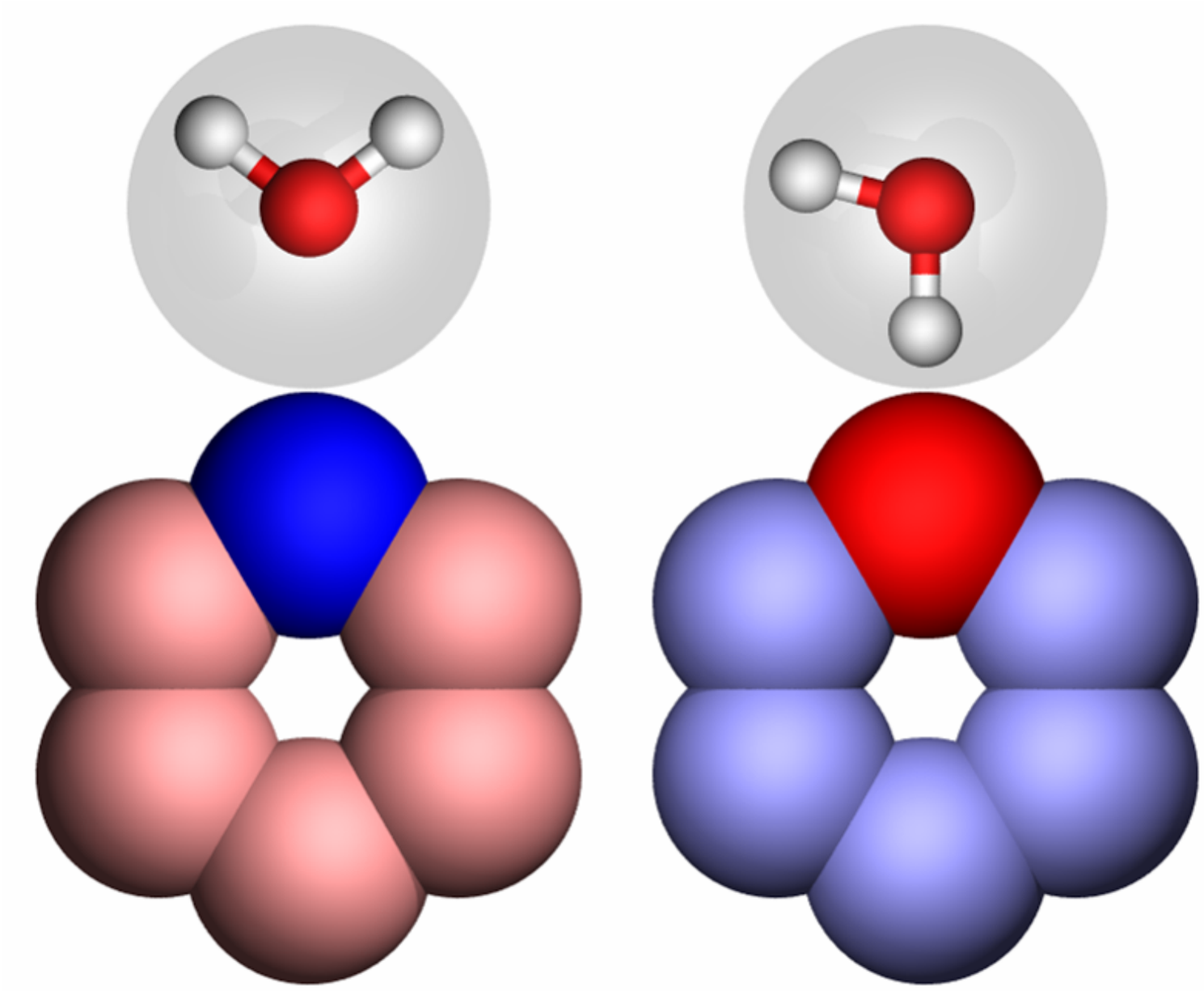
- In implicit solvent, the hydration free energies are equal!

Mobley, Chodera and Dill, J. Phys. Chem. B. **112:2405-2414 (2008).**

Water does *not* respond as a dielectric continuum – structure is crucial



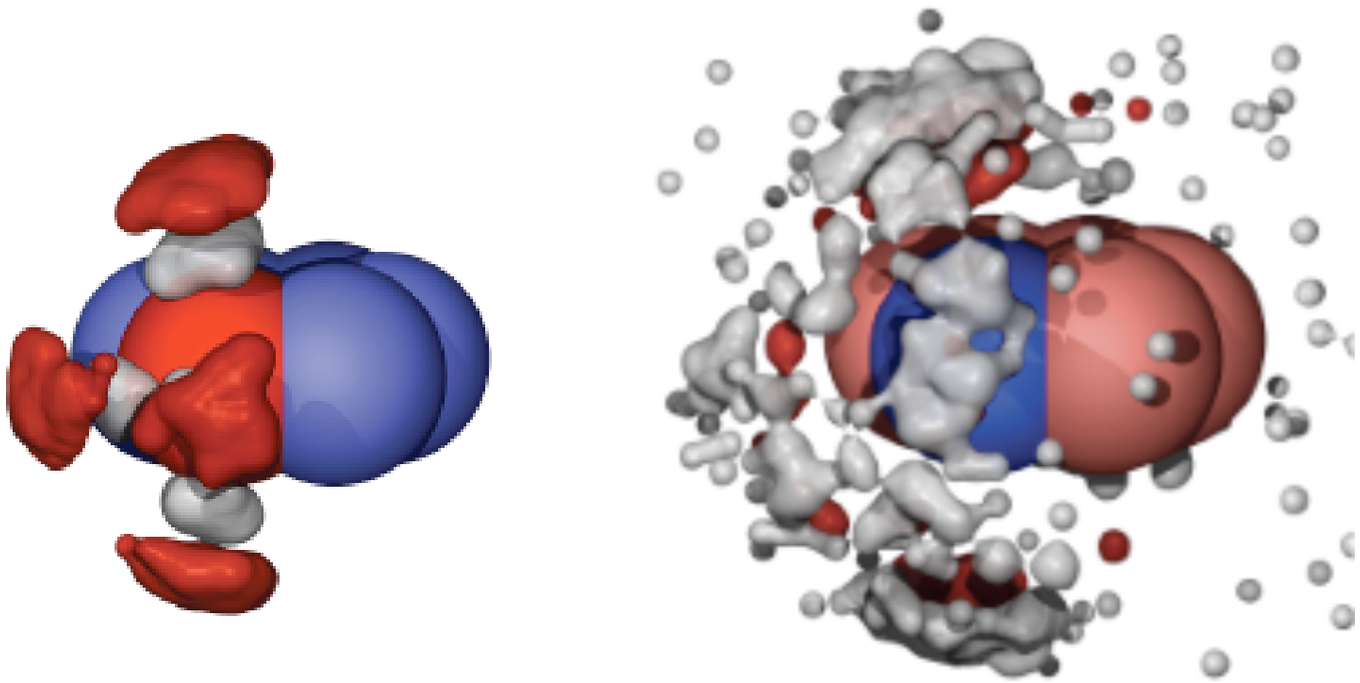
Hydration asymmetry is driven by the structure of water



Mobley, Chodera and Dill, J. Phys. Chem. B. **112:2405-2414 (2008).**

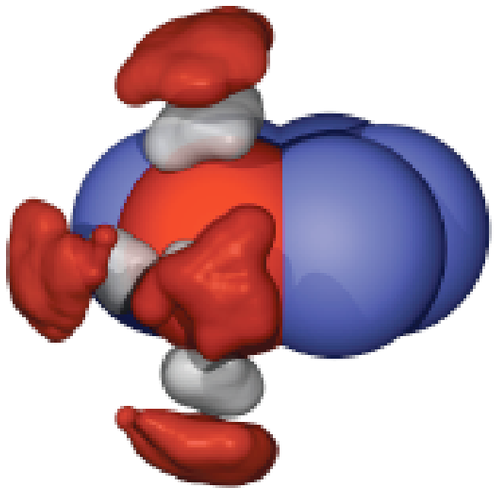
What did we learn about water?

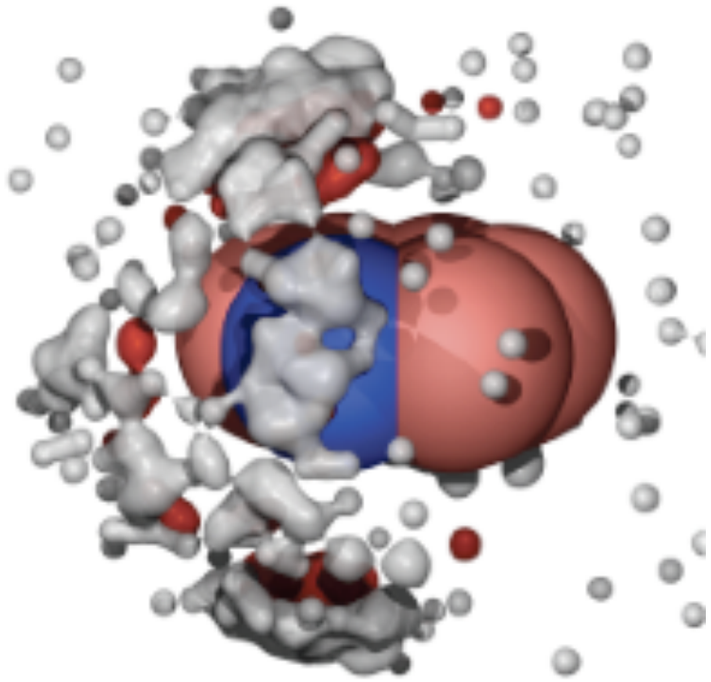
What did we learn about water?



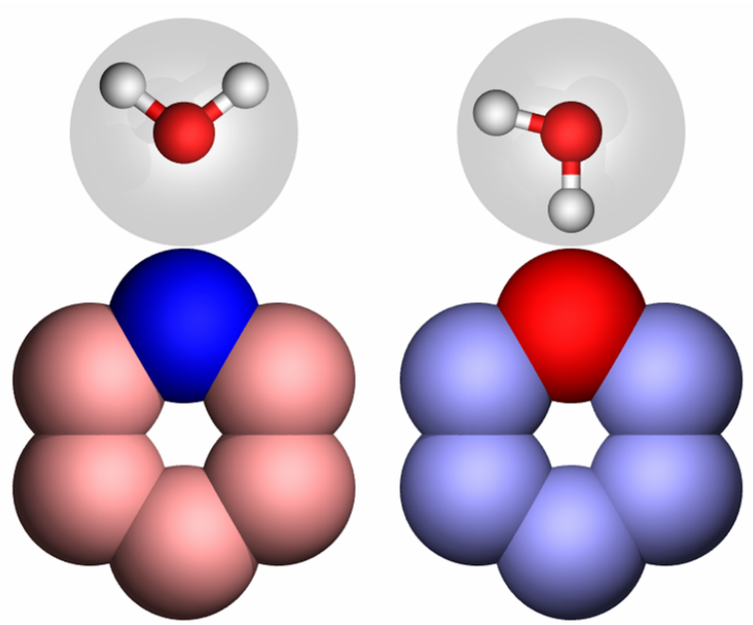
Water’s electrostatic response is asymmetric with respect to charge

What did we learn about water?





Water’s electrostatic response is asymmetric with respect to charge



This asymmetry is due to the inherent asymmetry of the water molecule

Possible points of collaboration

- Computational studies of:
  - Biomolecular binding/interactions
  - Thermodynamic properties (transfer free energies, solubility, etc.)
  - Proteins/nucleic acids
  - Organic solvents
- Testing/improving molecular dynamics simulations and algorithms

Acknowledgments

- Lysozyme:
  - Apolar: Alan Graves, John Chodera, Andrea McReynolds, Brian Shoichet, Ken Dill
  - Polar: Sarah Boyce, Brian Shoichet, Ken Dill
- 504 molecule set:
  - Christopher Bayly, Matthew Cooper, Michael Shirts (WCA separation), Ken Dill
- Asymmetry:
  - Alan Barber II, Christopher Fennell, Ken Dill