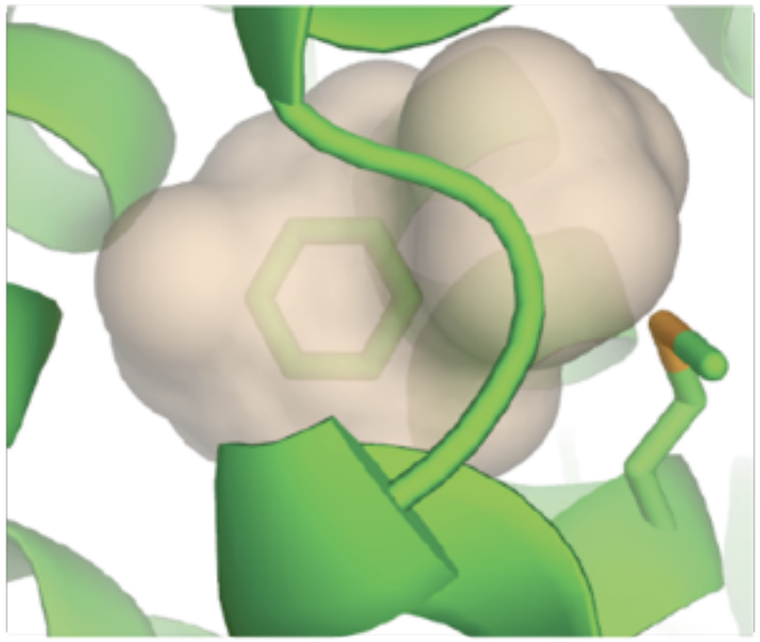
Using solution-phase free energy calculations to improve binding free energies

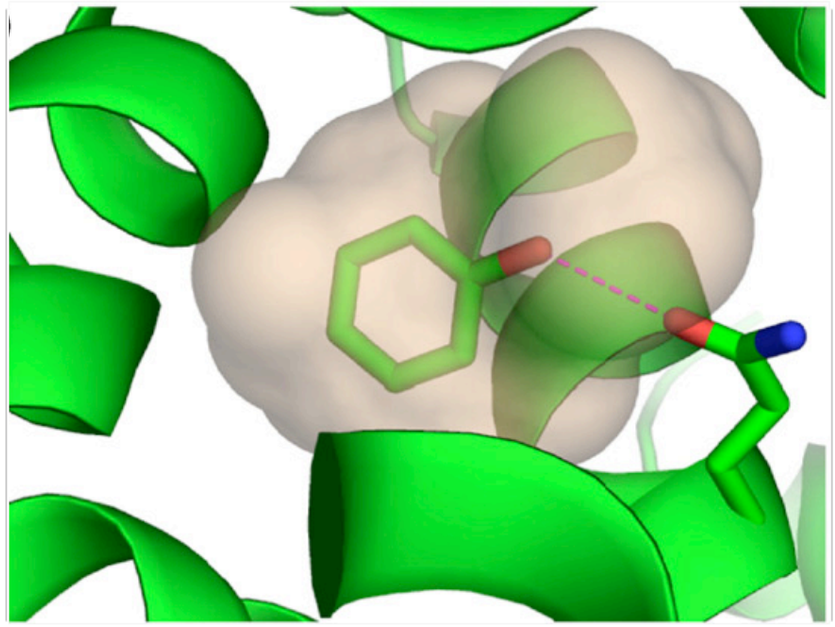
David Mobley

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



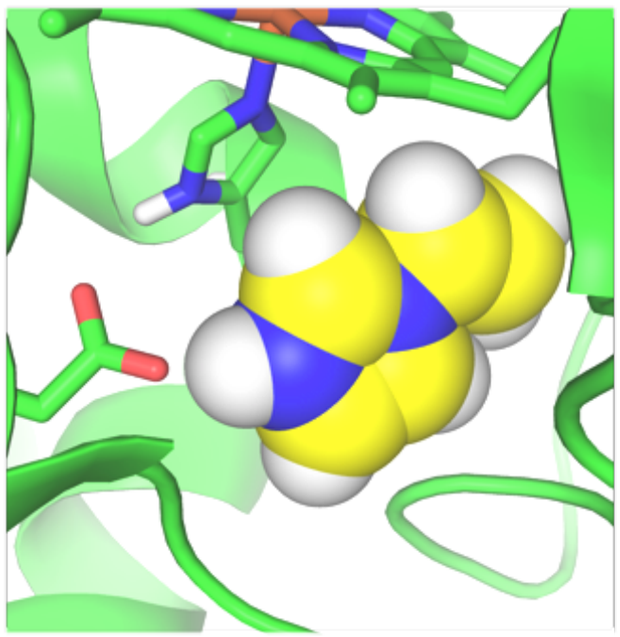
Lysozyme L99A

- Simple
- Nonpolar
- Dry



Lysozyme L99A/M102Q

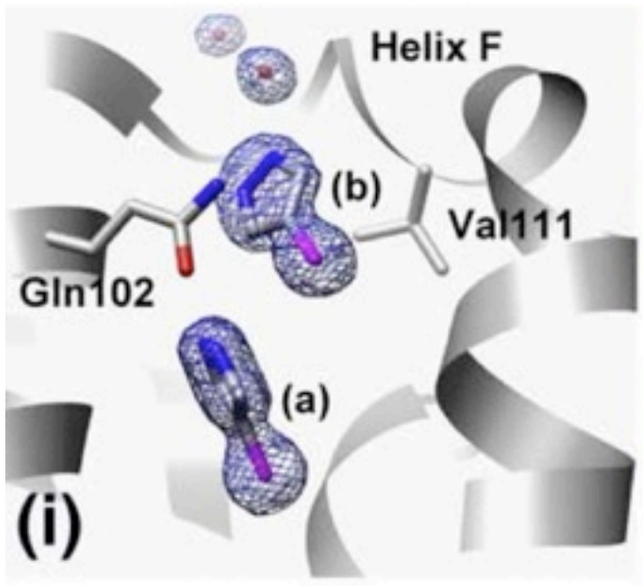
- Simple
- Polar
- Dry
- Additional stable binding modes

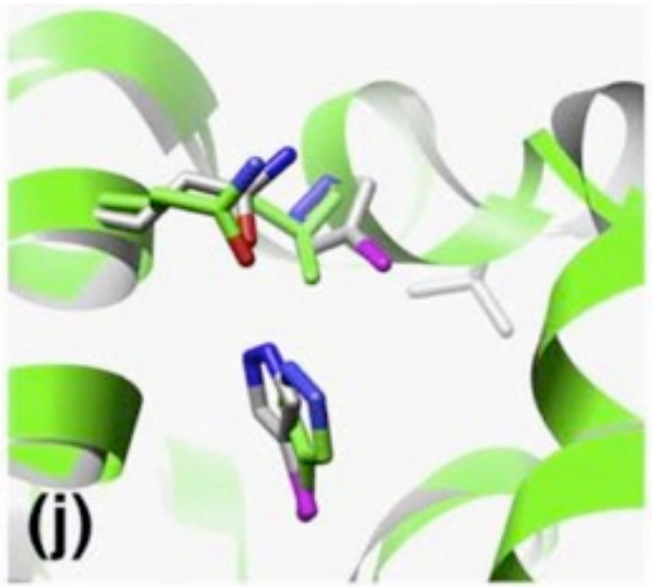


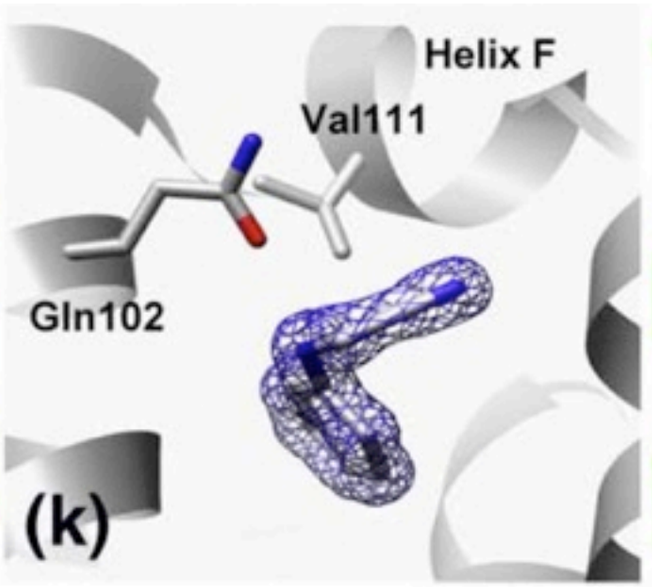
Cytochr. C Peroxidase

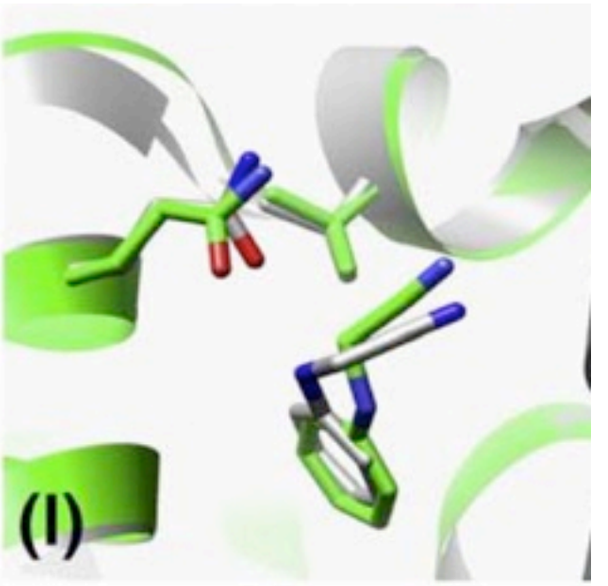
- Simple (?)
- Polar, Charged
- Wet
- Additional stable binding modes
- Force field issues?

In the lysozyme sites, we typically end up with ~1.5 kcal/mol RMS errors and substantial predictive power



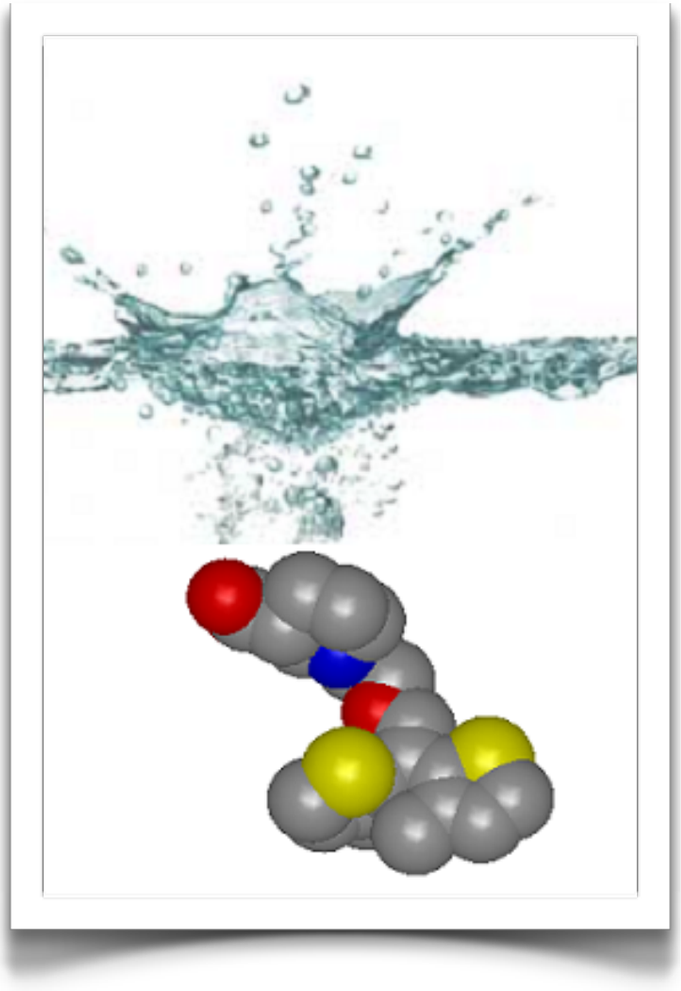






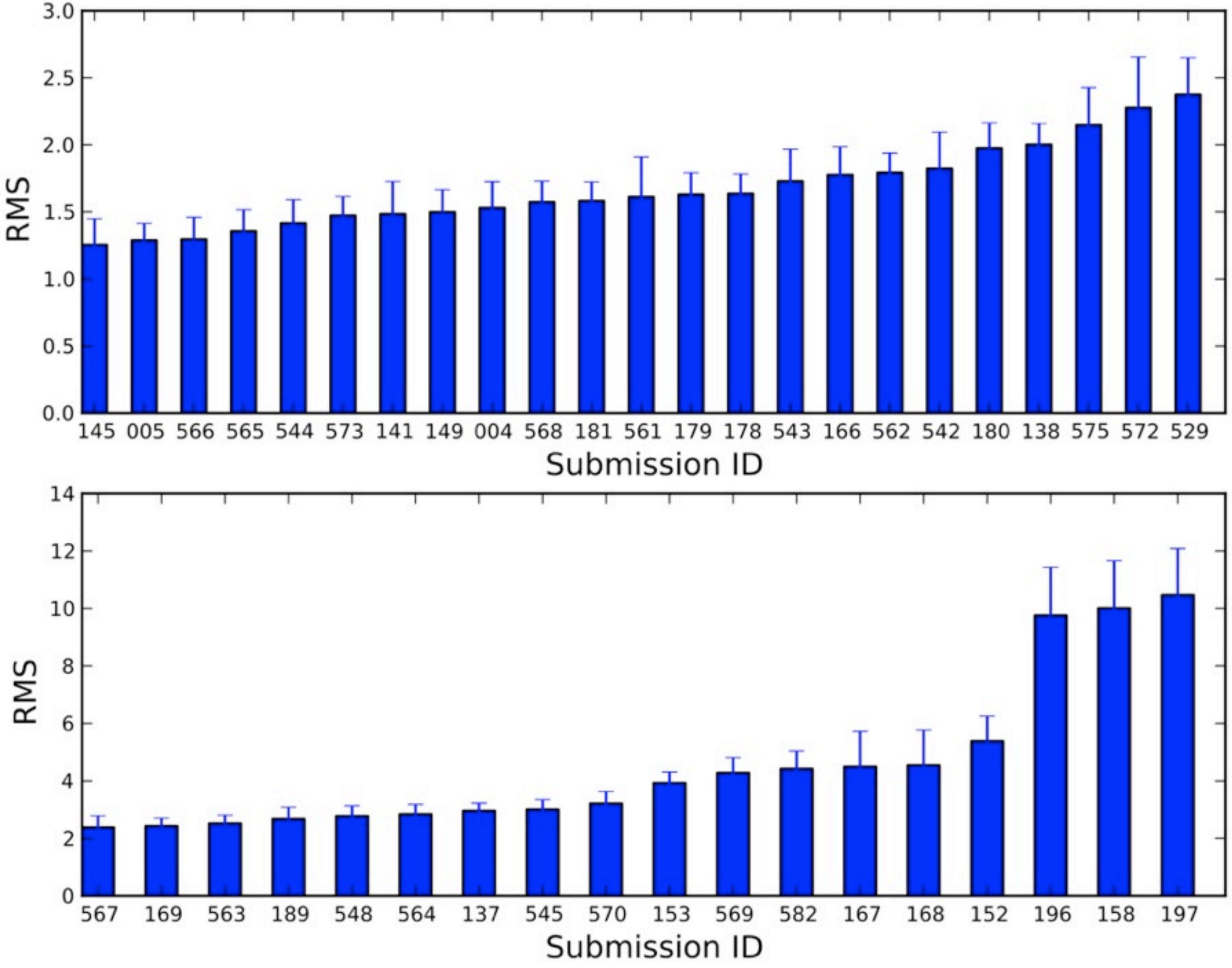
Hydration: It’s what we do every day*

- Two subsets: Blind and supplementary
- Most people did all of both
- Started with 52 compounds
- Post-SAMPL, cut to 49 due to human error

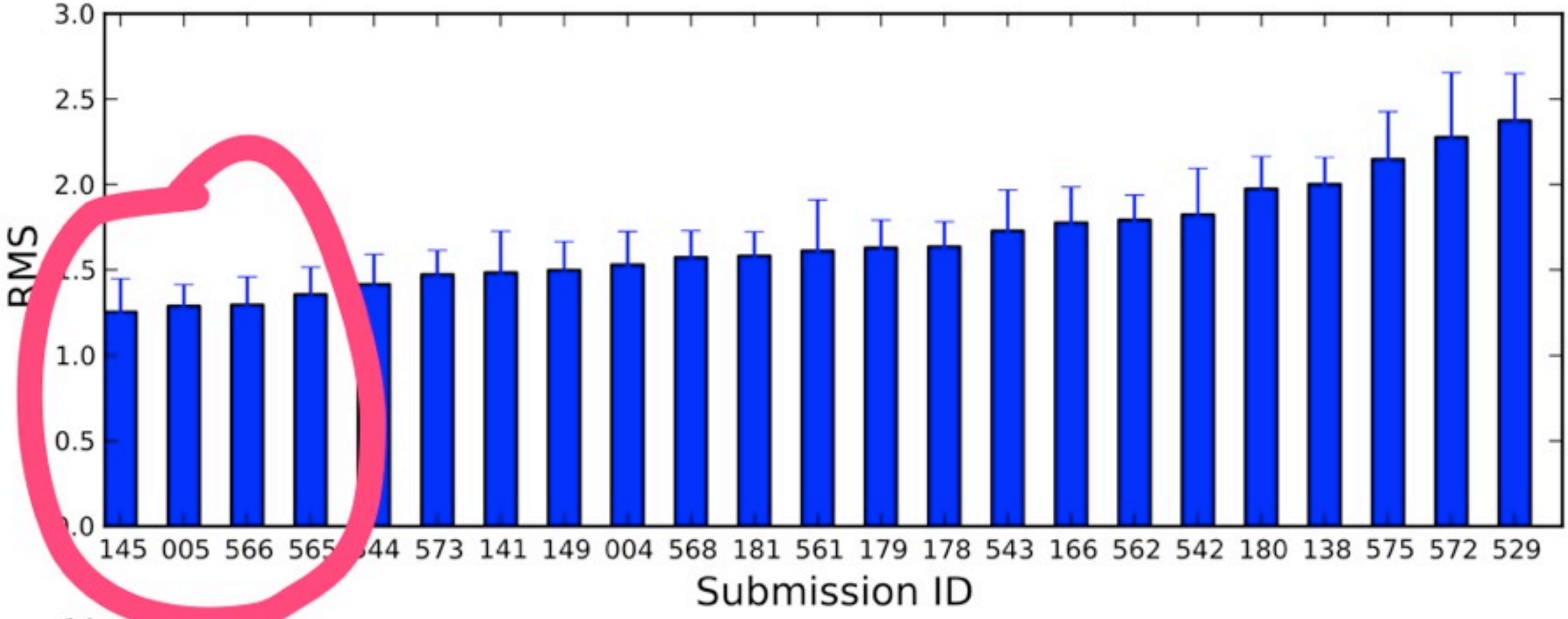


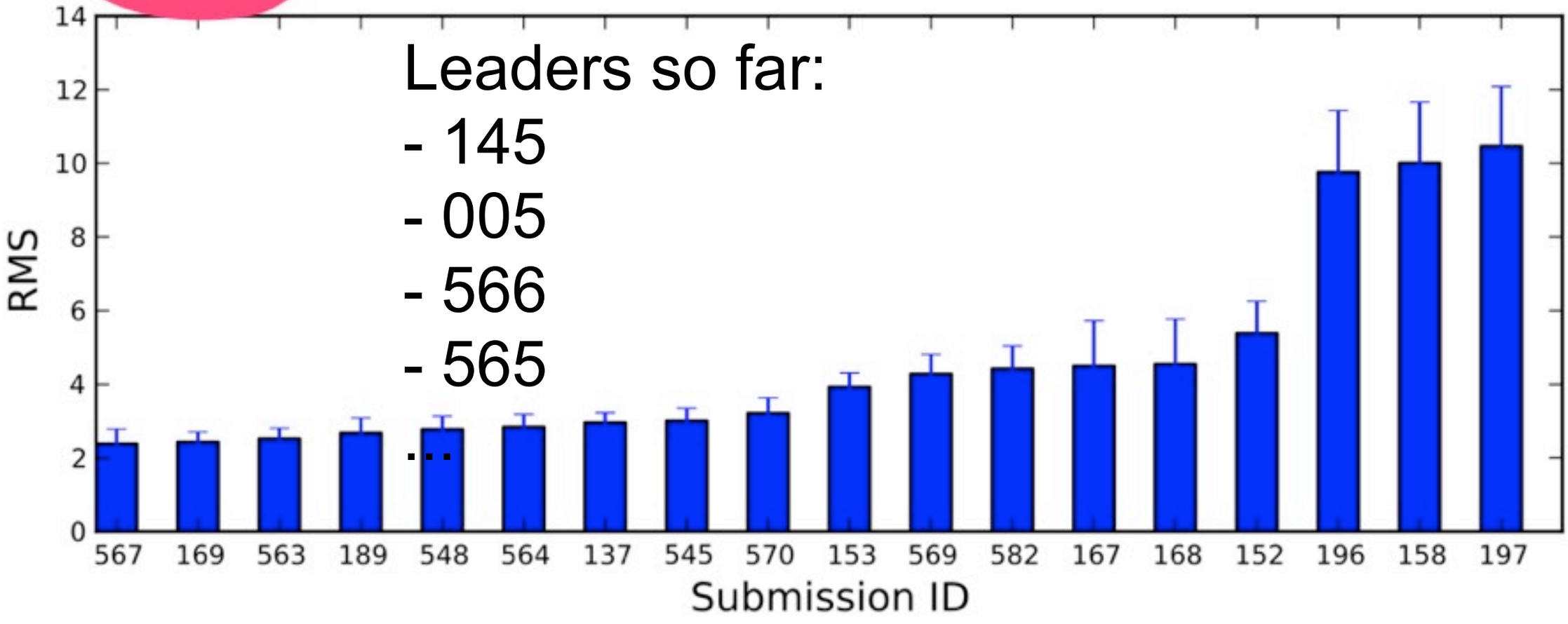
* - Every SAMPL

For SAMPL4, we ran a bunch of different metrics...



For SAMPL4, we ran a bunch of different metrics...





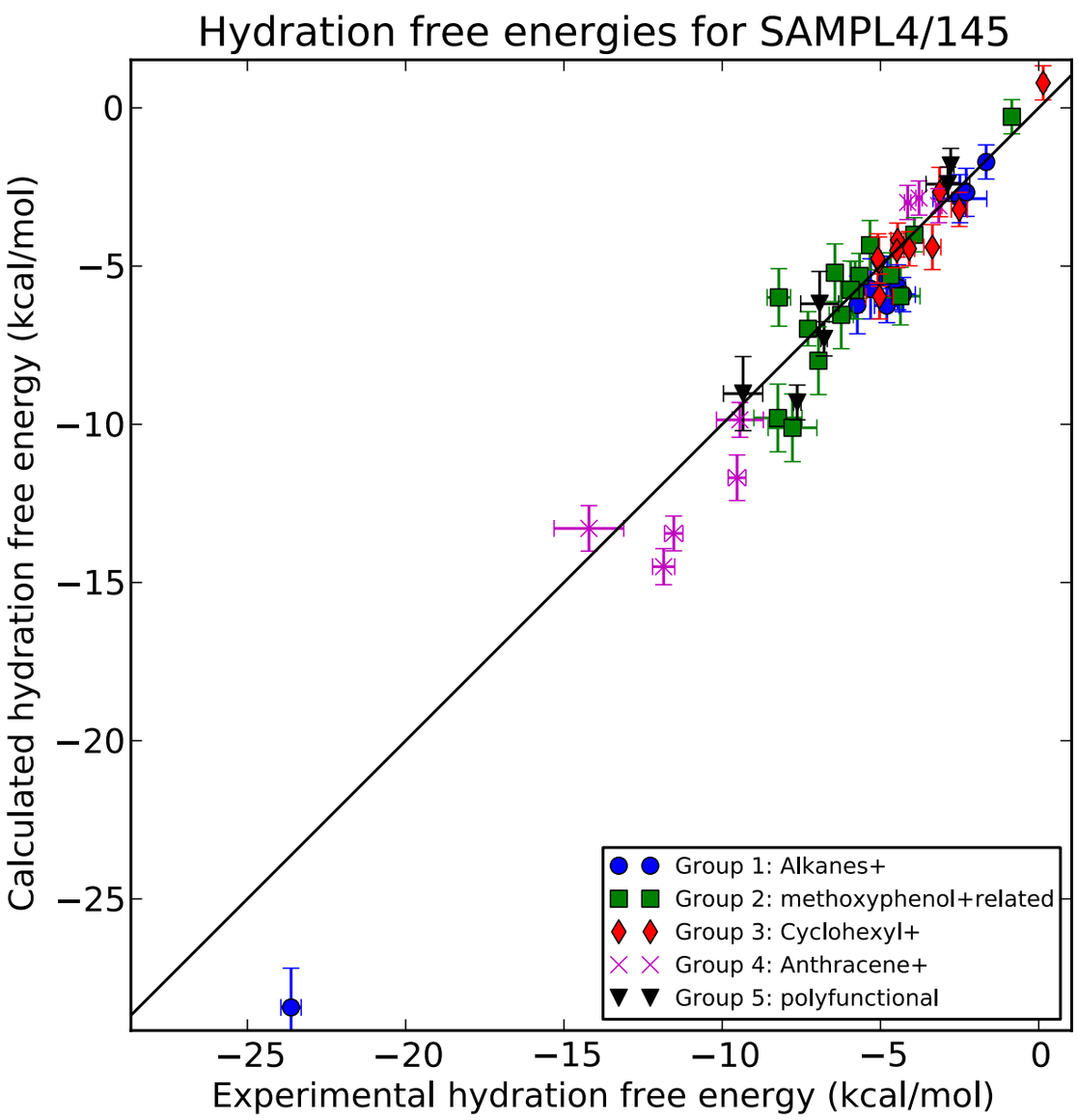
Leaders so far:

- 145
- 005
- 566
- 565
- ...

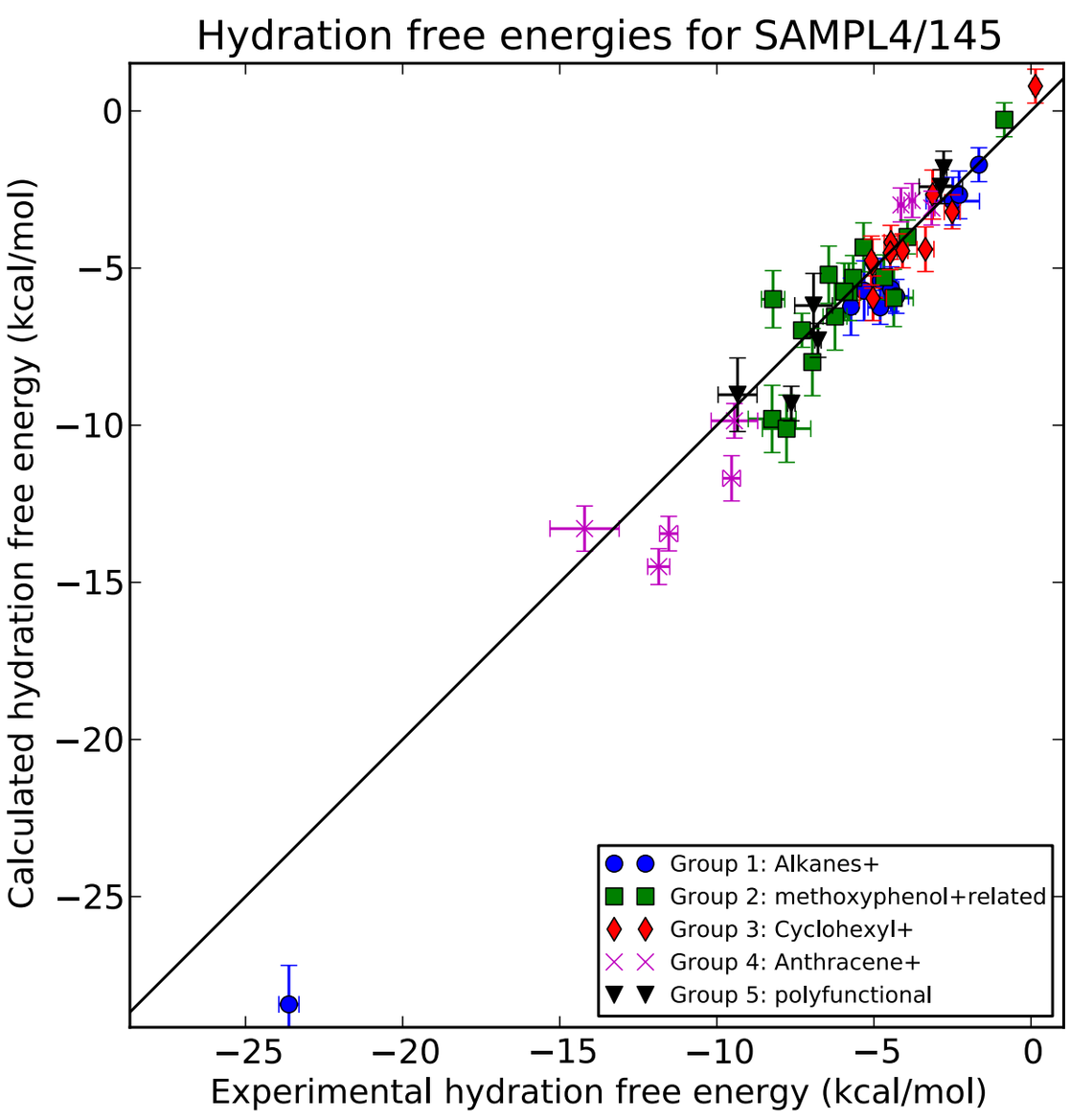
What are these methods?

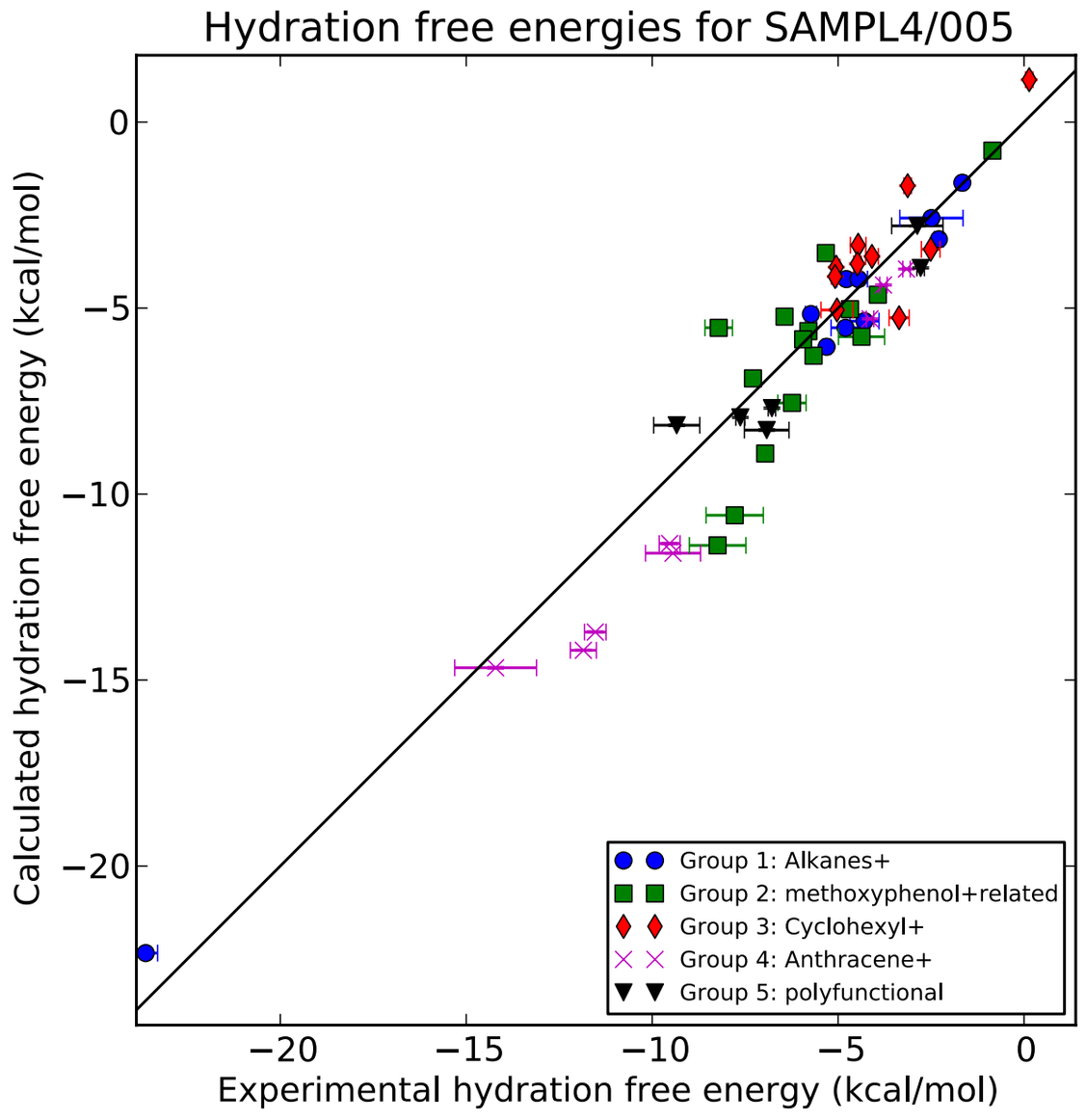
- 145: QM + implicit solvent + funct. group corrections
  - Lars Sandberg, University of Dundee
  - Conformational search with Schrödinger tools, then geometry optimization (QM, implicit solvent)
  - Separate polarization, electrostatic, dispersion, repulsion, cavity formation components
  - Empirical functional group corrections for charge transfer to water
- 005: Explicit solvent alchemical MD
  - Mobley lab (Karisa Wymer)
  - Standard approach, new hydroxyl parameters (with Chris Fennell)
- 566: PB single conformer
  - Matt Geballe, OpenEye
  - Omega, then gas phase minimization; pick low energy conformation
  - AM1-BCC charges (symmetrized), ZAP (ZAP9 radii)
  - Like SAMPL2 (Nicholls et al. JCAMD 6:293)

Several disparate methods did comparably well

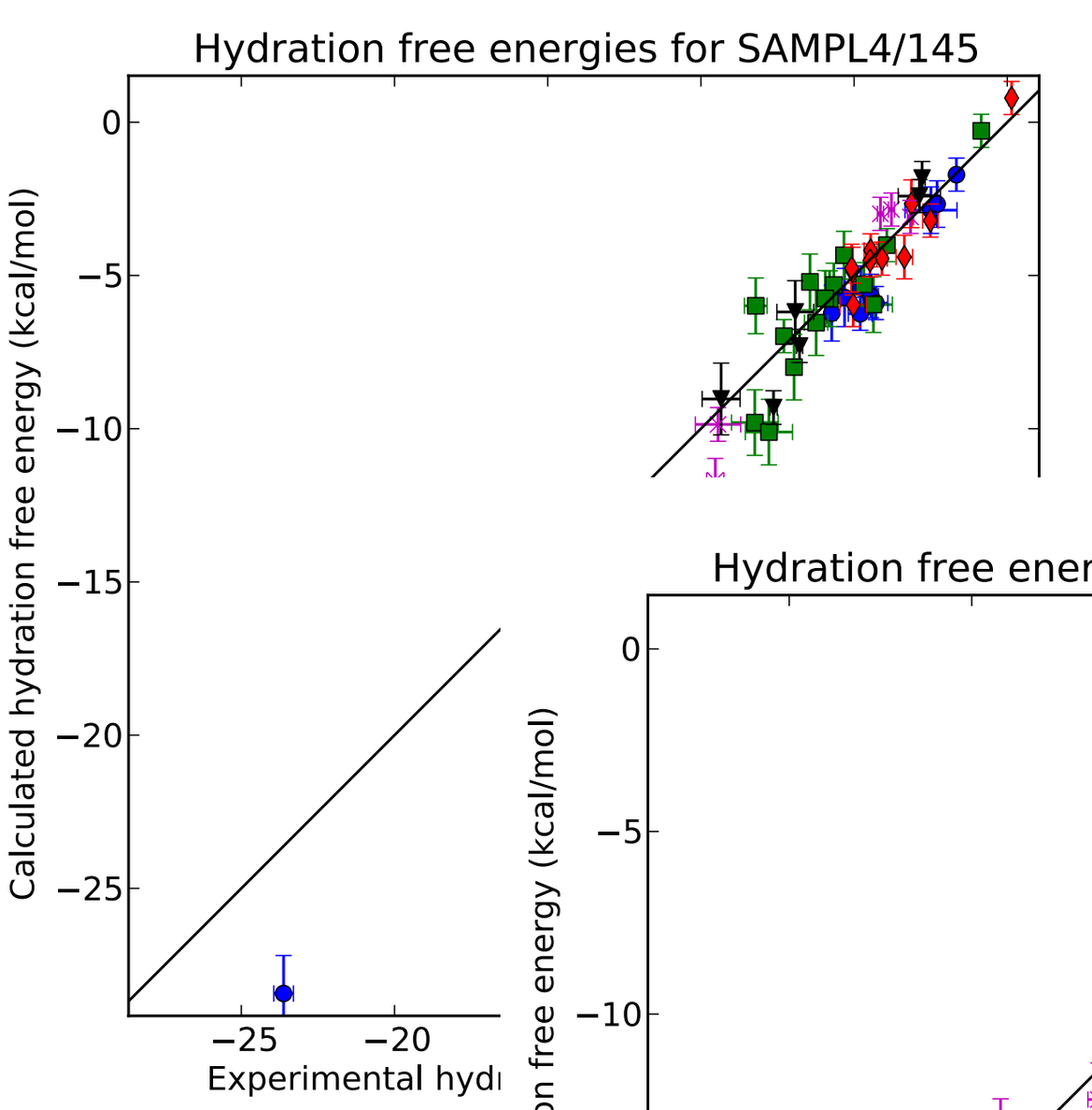


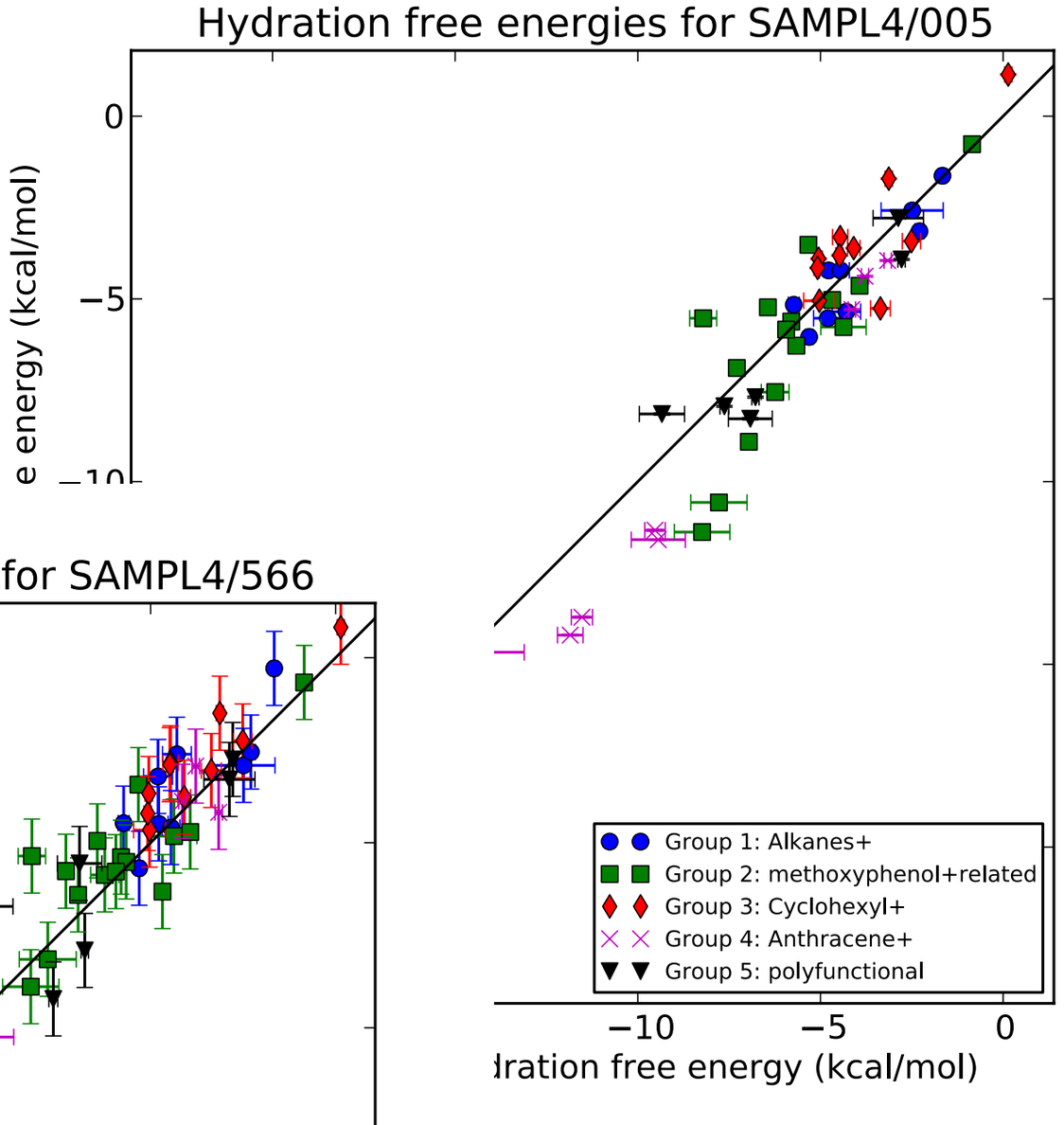
Several disparate methods did comparably well

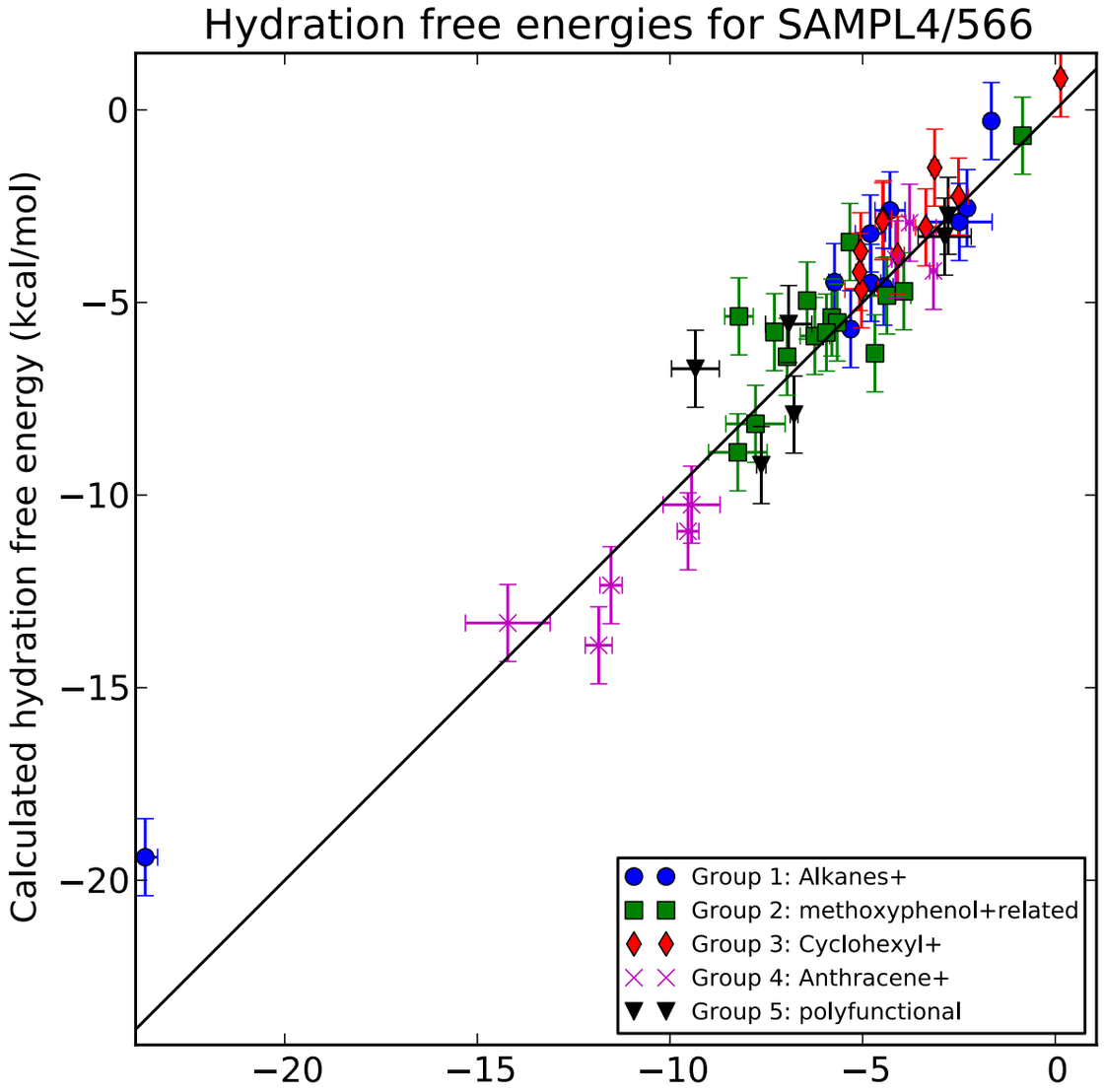




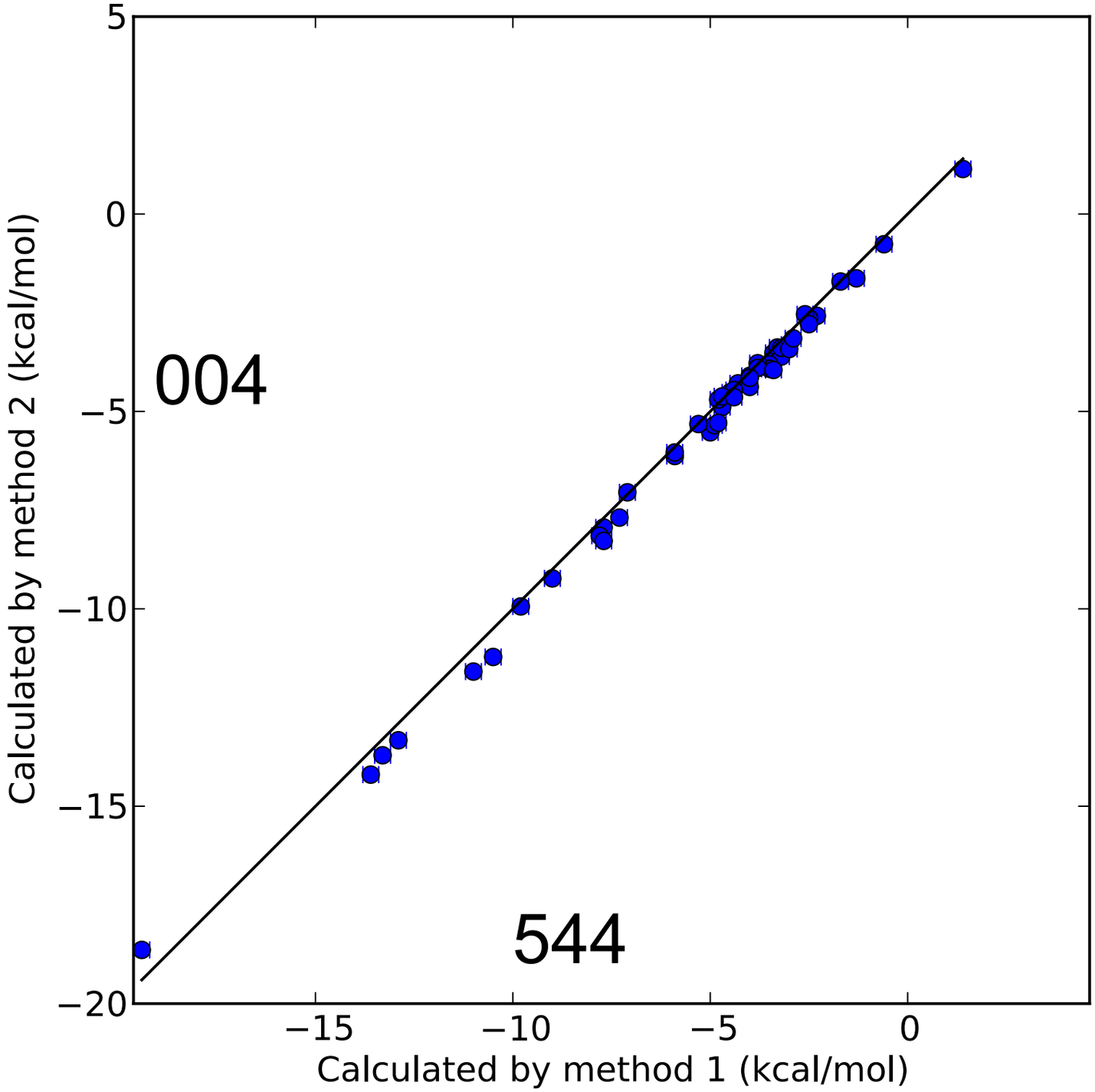
Several disparate methods did comparably well





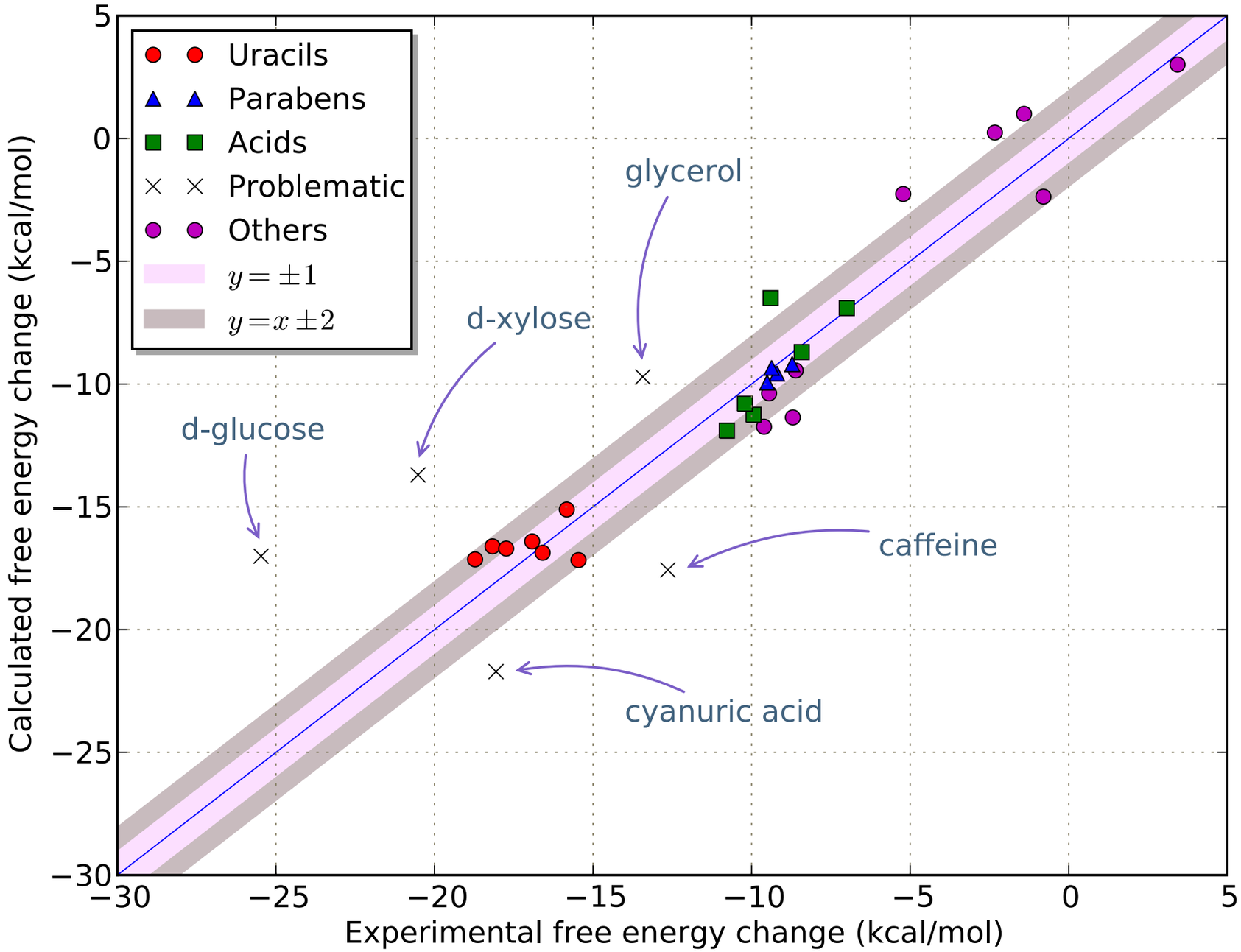


Good news: Methods which are the same agree

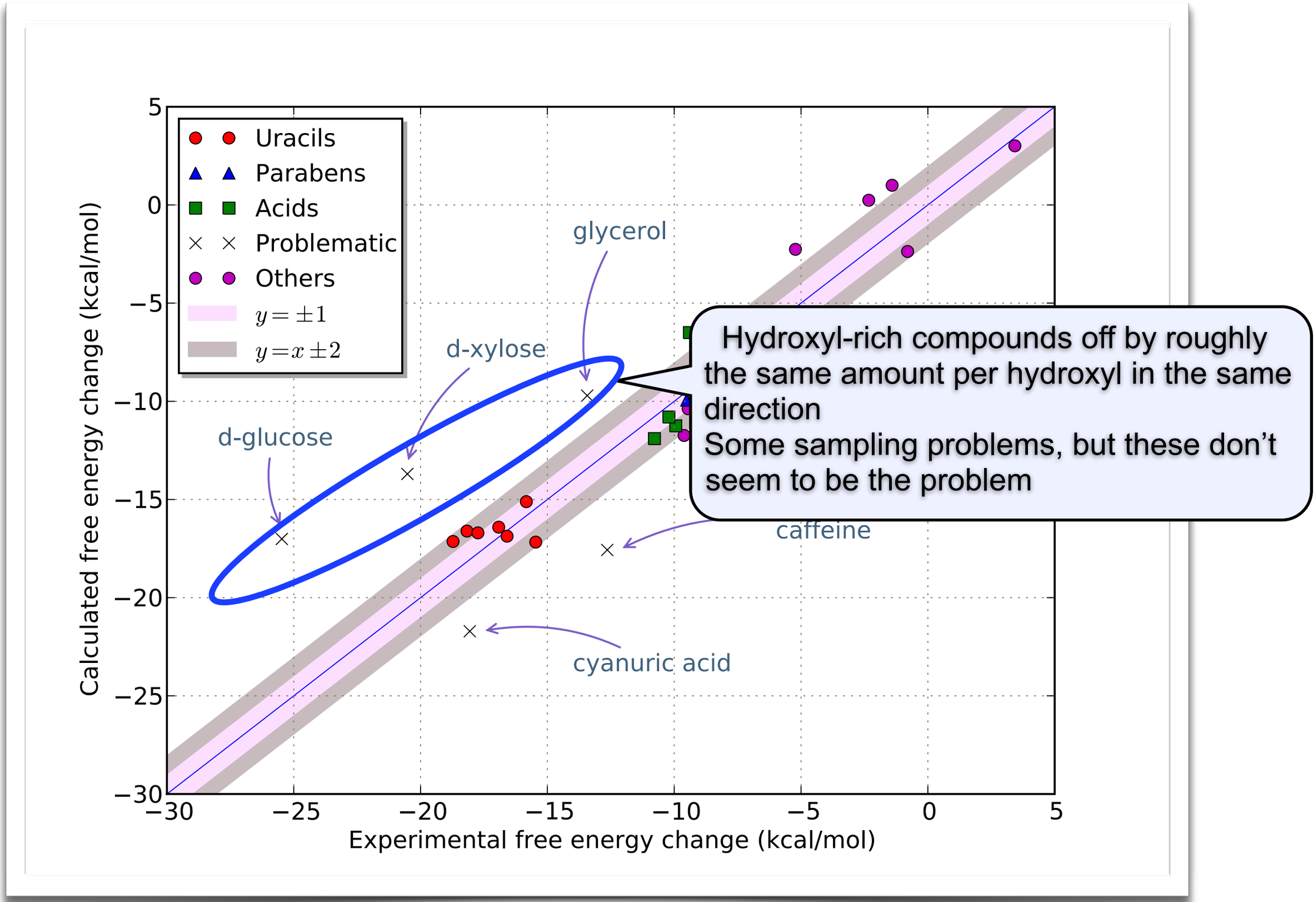


AM1-BCC GAFF (Gilson and Mobley labs)

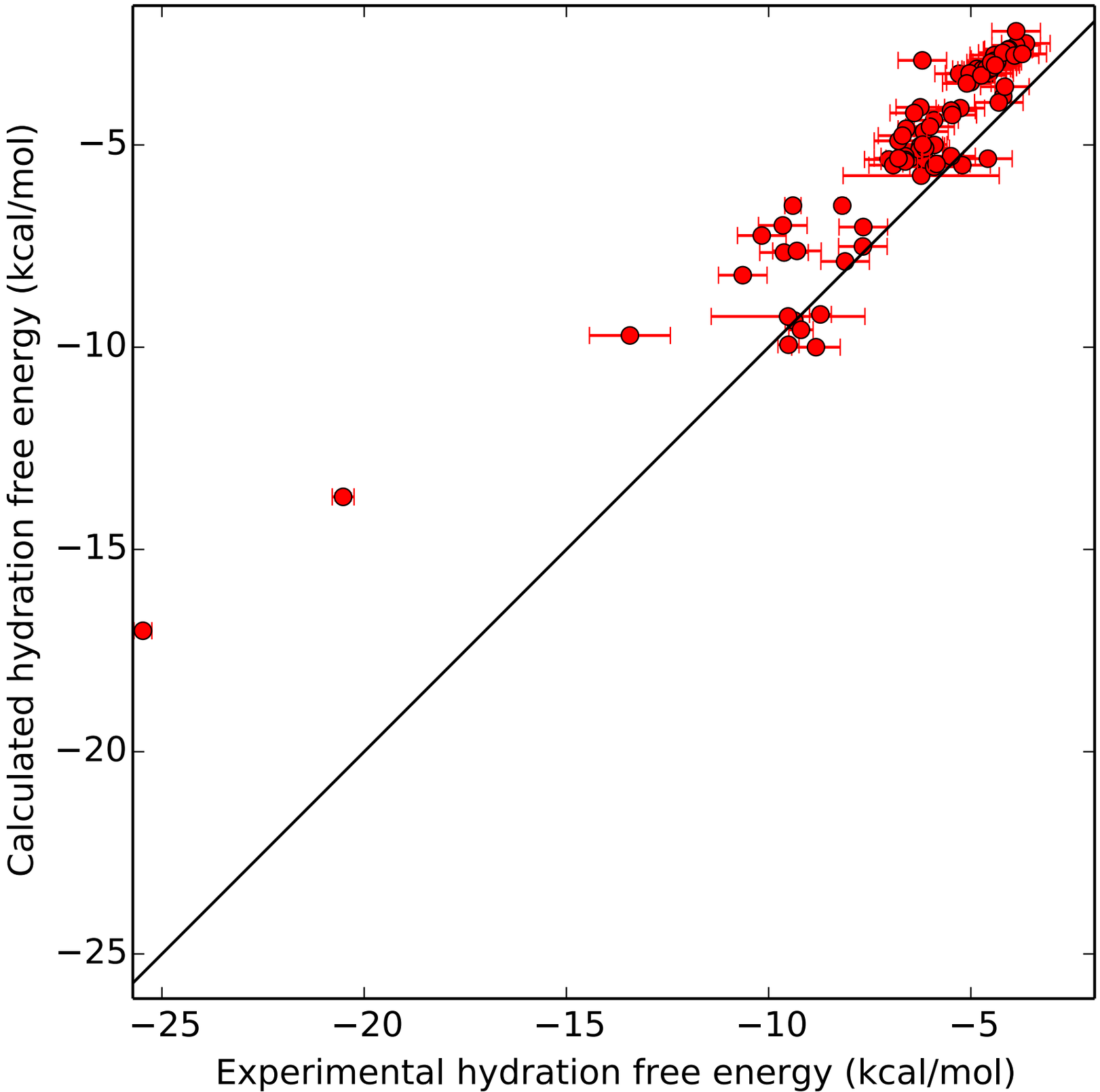
Sometimes, doing predictions highlights issues we hadn't noticed before



Sometimes, doing predictions highlights issues we hadn't noticed before



Focusing in on hydroxyls in our large set, there was a systematic error

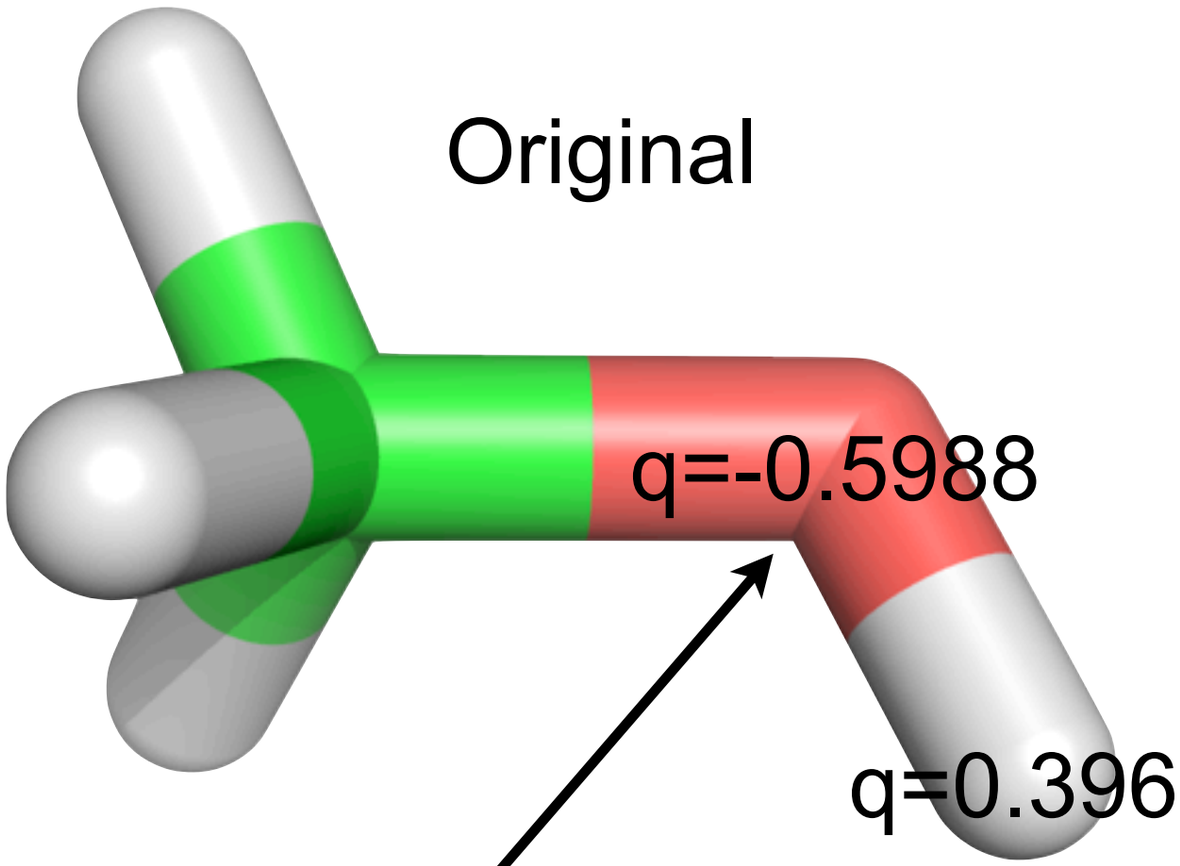


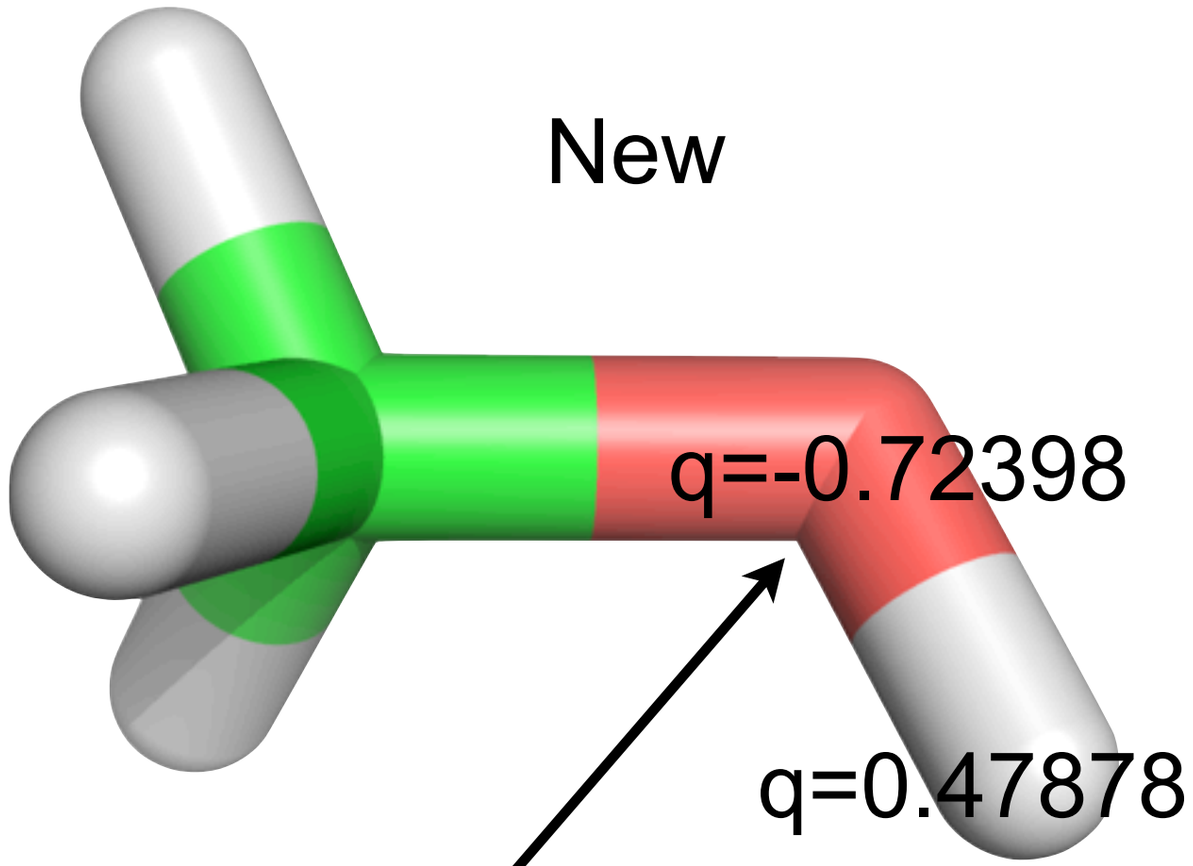
Idea: Re-fit parameters for hydroxyls based on neat liquid properties?

- Start with some initial force field and optimize parameters automatically to reproduce measured properties
- Density, heat of vaporization, dielectric constant

- As a starting point, we took methanol, ethanol, butanol, and propanol, and optimized beginning from OPLS and GAFF
- Hydroxyl parameters end up in a similar place regardless of starting FF and across molecules

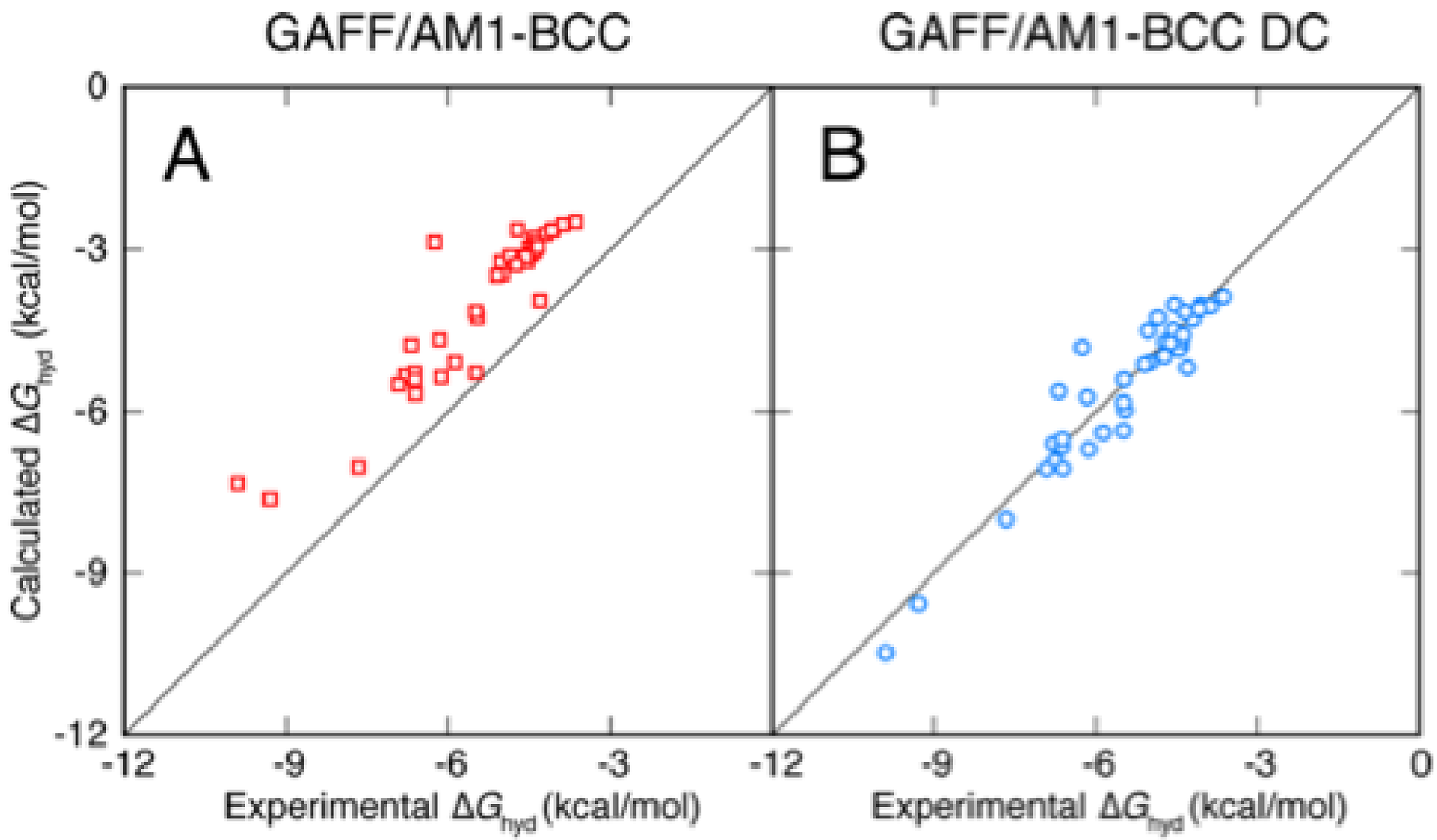
We don’t want to do new neat liquid simulations for each new molecule, so we use the methanol hydroxyl parameters for a large set





In practice, this is a new σ and ε, plus a hydroxyl charge scaling factor of 1.20905

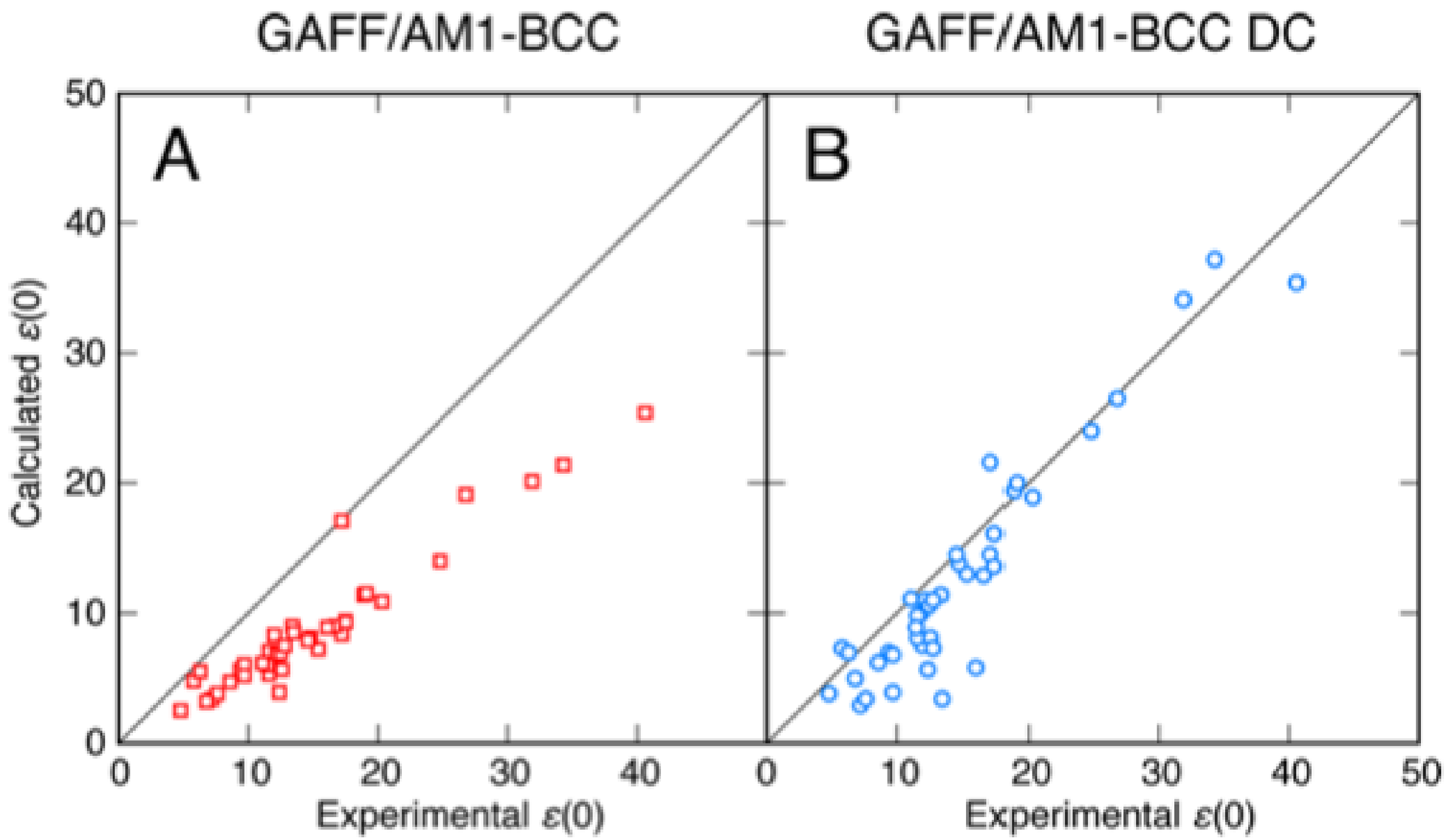
The new parameters dramatically improved performance on a large test set



RMSE 1.52
MSE 1.43
(units kcal/mol)

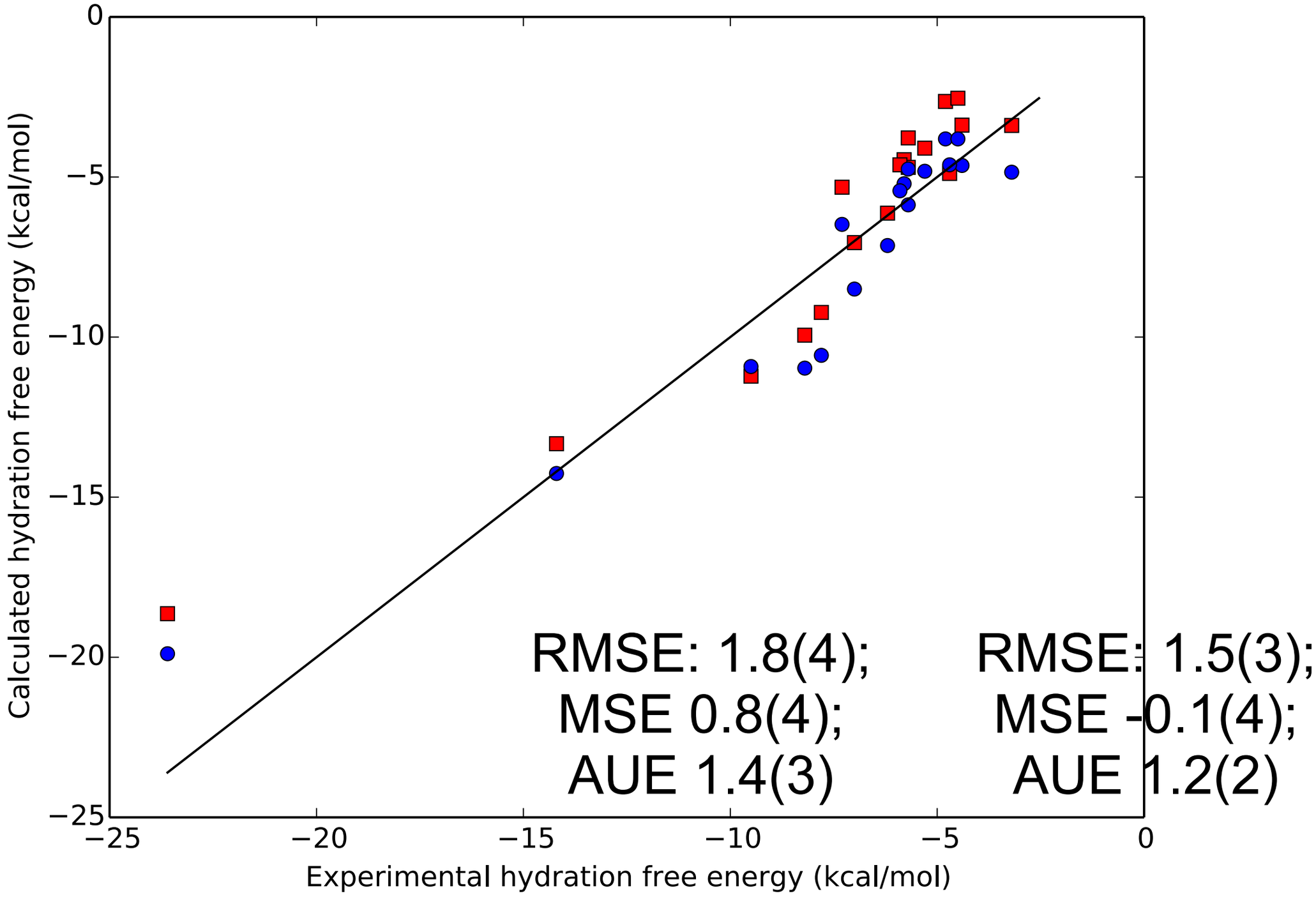
RMSE: 0.68
MSE -0.5

Dielectric-corrected GAFF does dramatically better at dielectric constants as well

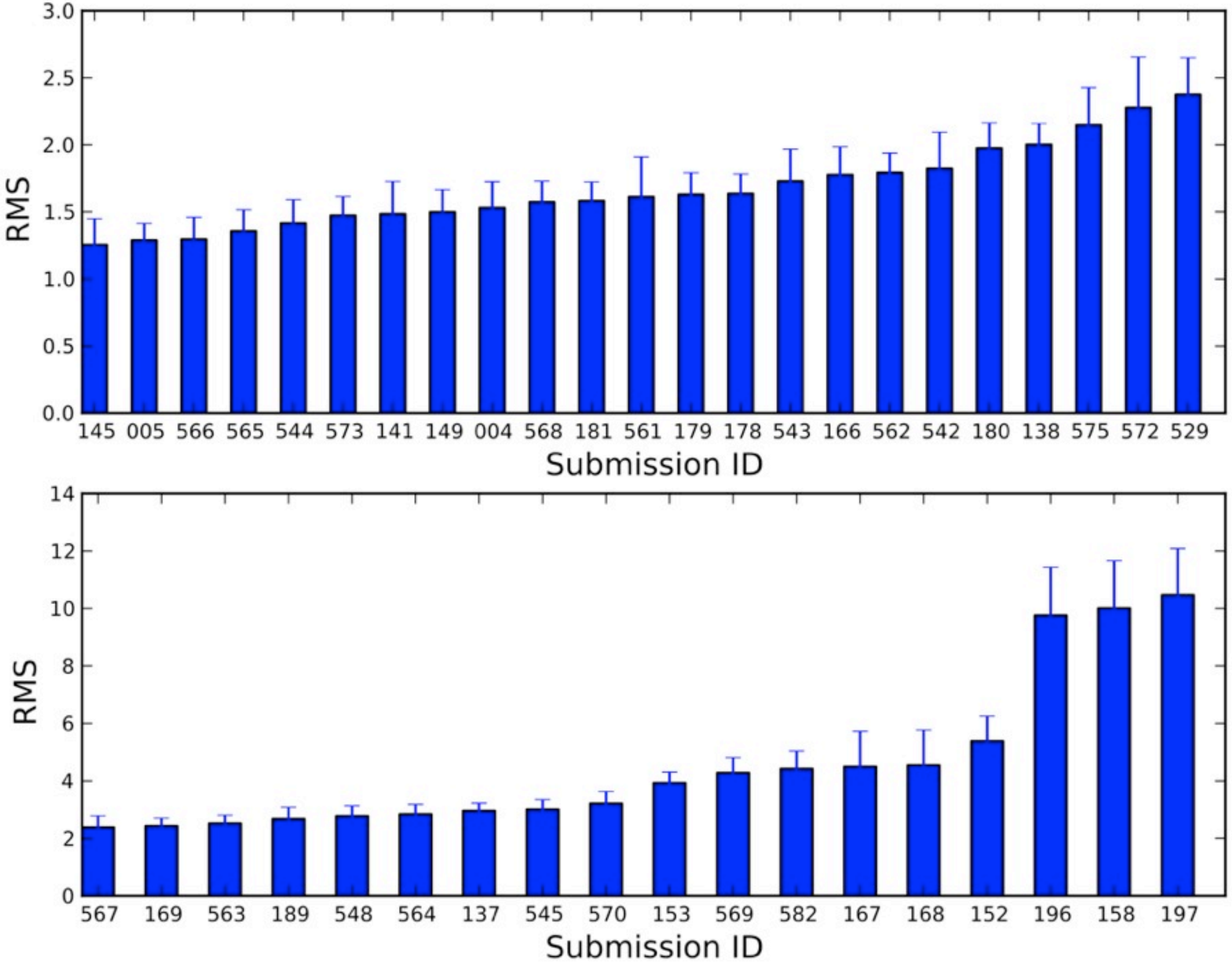


Fennell, Wymer and Mobley, JPCB 2014 (DOI 10.1021/jp411529h)

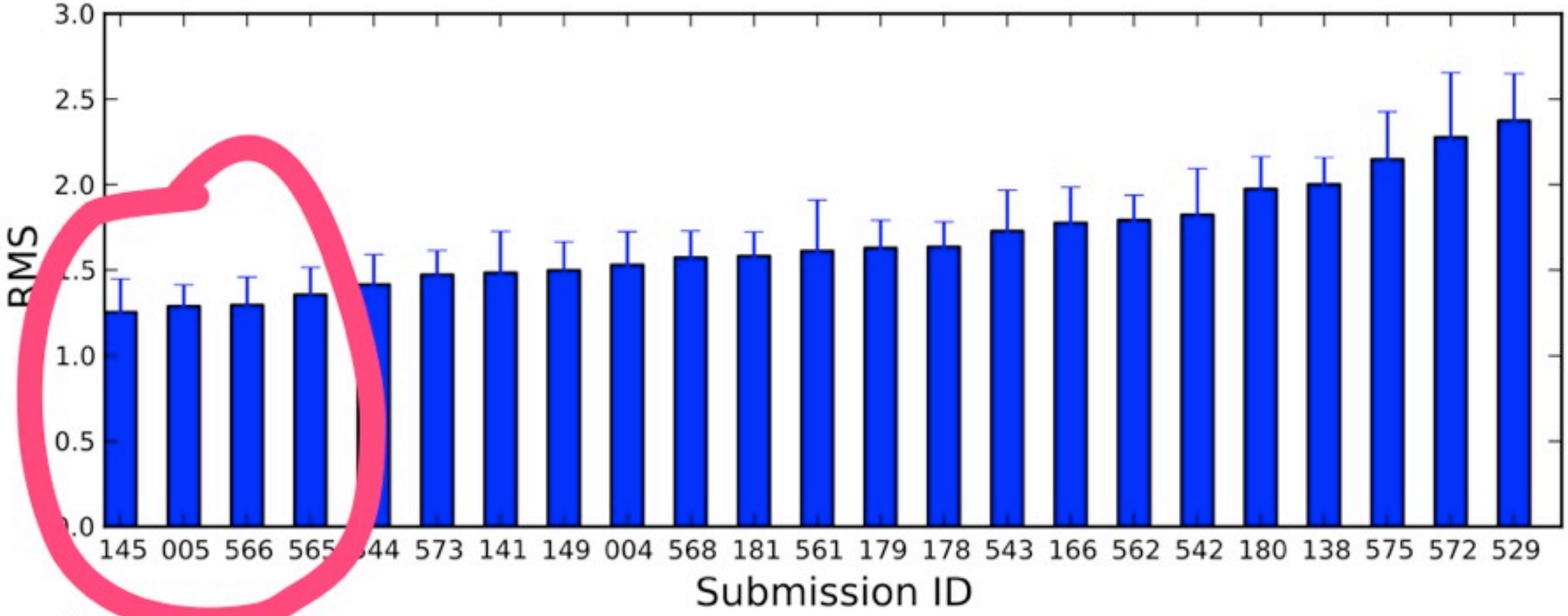
For the SAMPL4 hydroxyl-containing compounds, there is statistically significant improvement, though modest

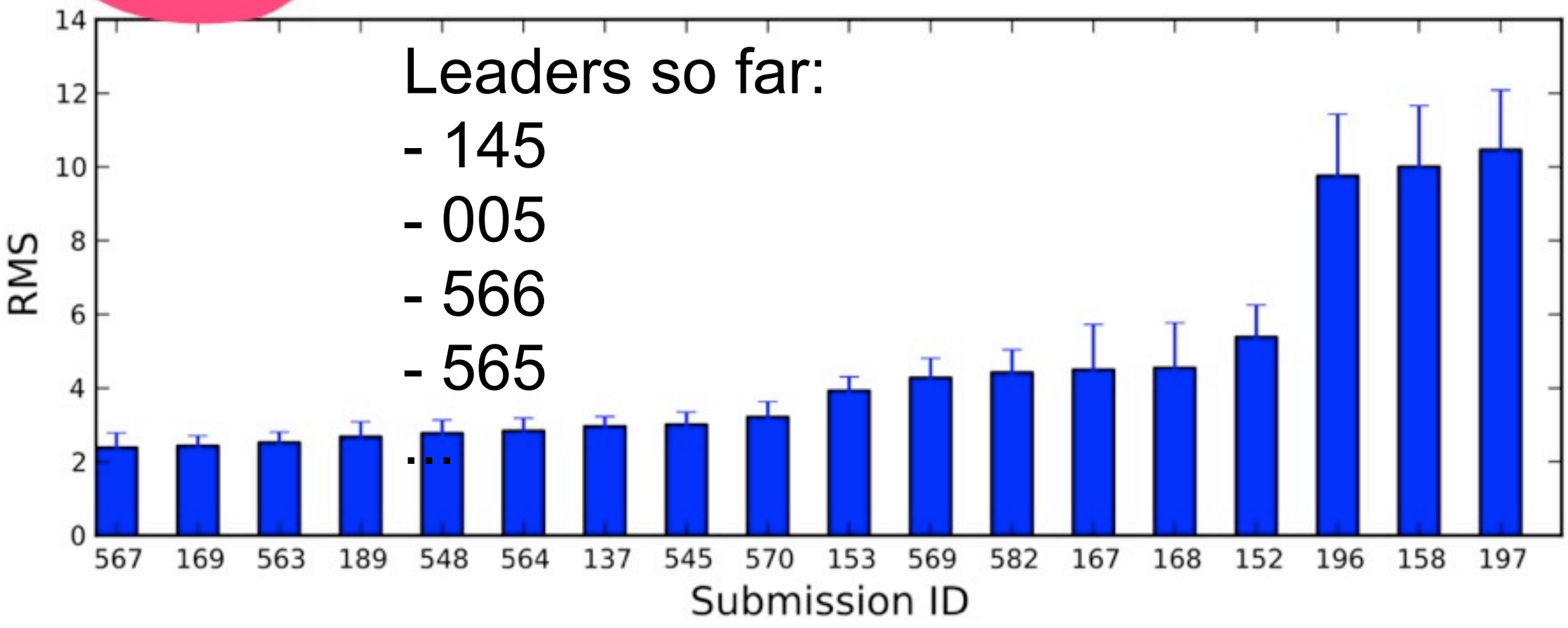


This was one of the top methods at SAMPL4

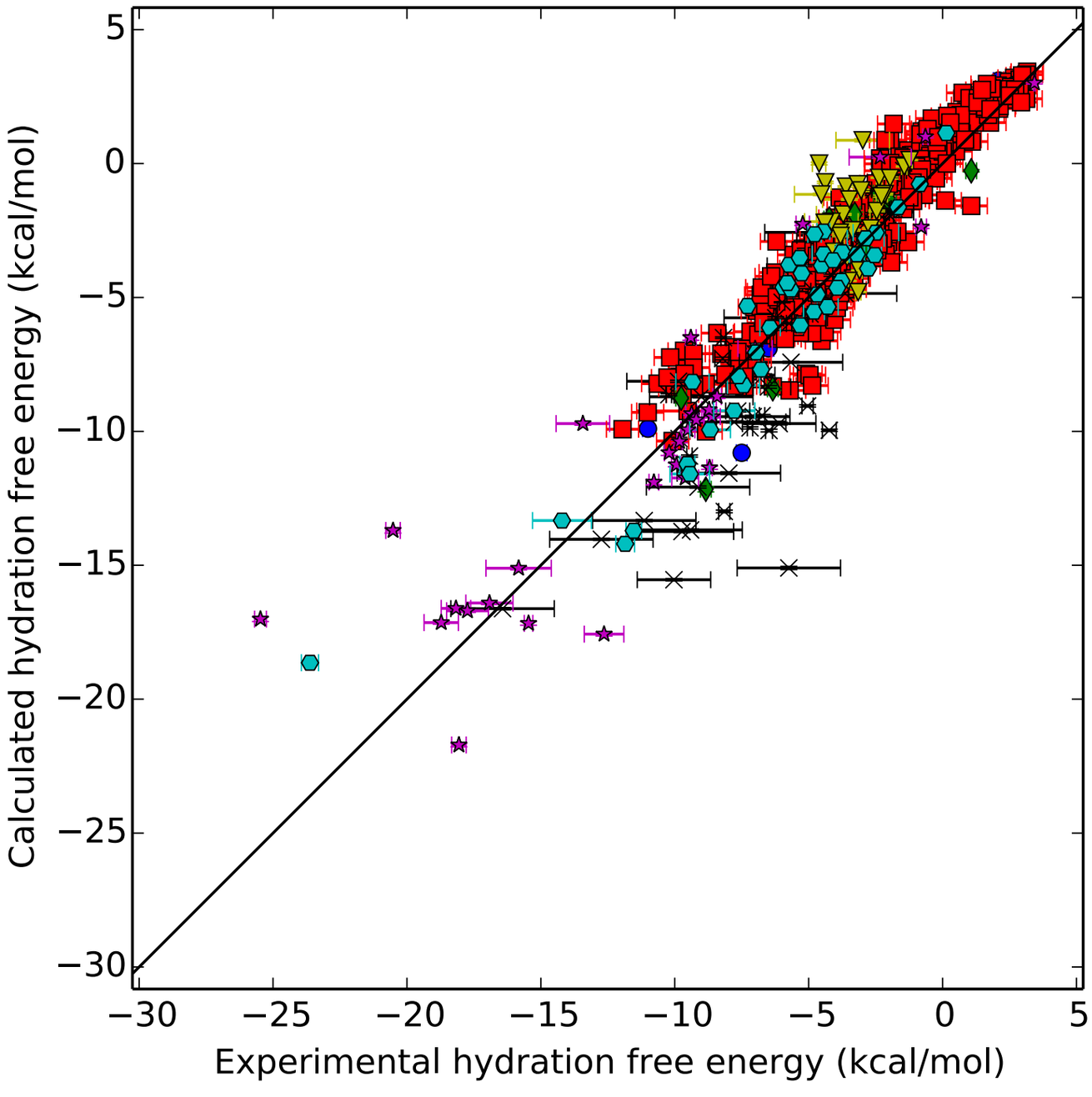


This was one of the top methods at SAMPL4



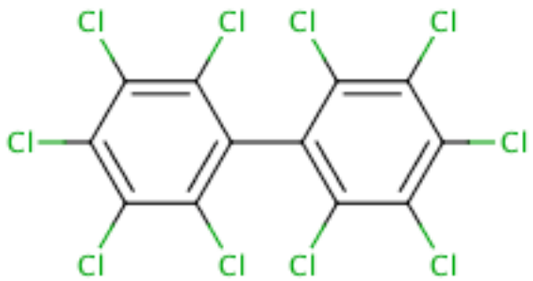


Hydration free energies have been helpful for a variety of purposes, so we updated our “504 molecule set”

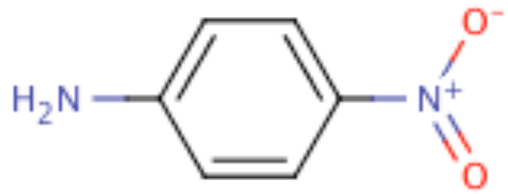


- 643 molecules
- Expt. & calc. values (GAFF)
- Structures, parameters, input files
- Literature citations
- (Curation ongoing)
- FreeSolv
- Permanent cite-able URL,
- http://www.escholarship.org/uc/item/6sd403pz

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

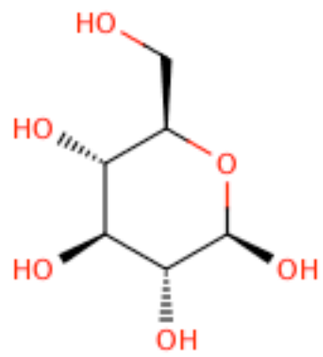


Largest MW



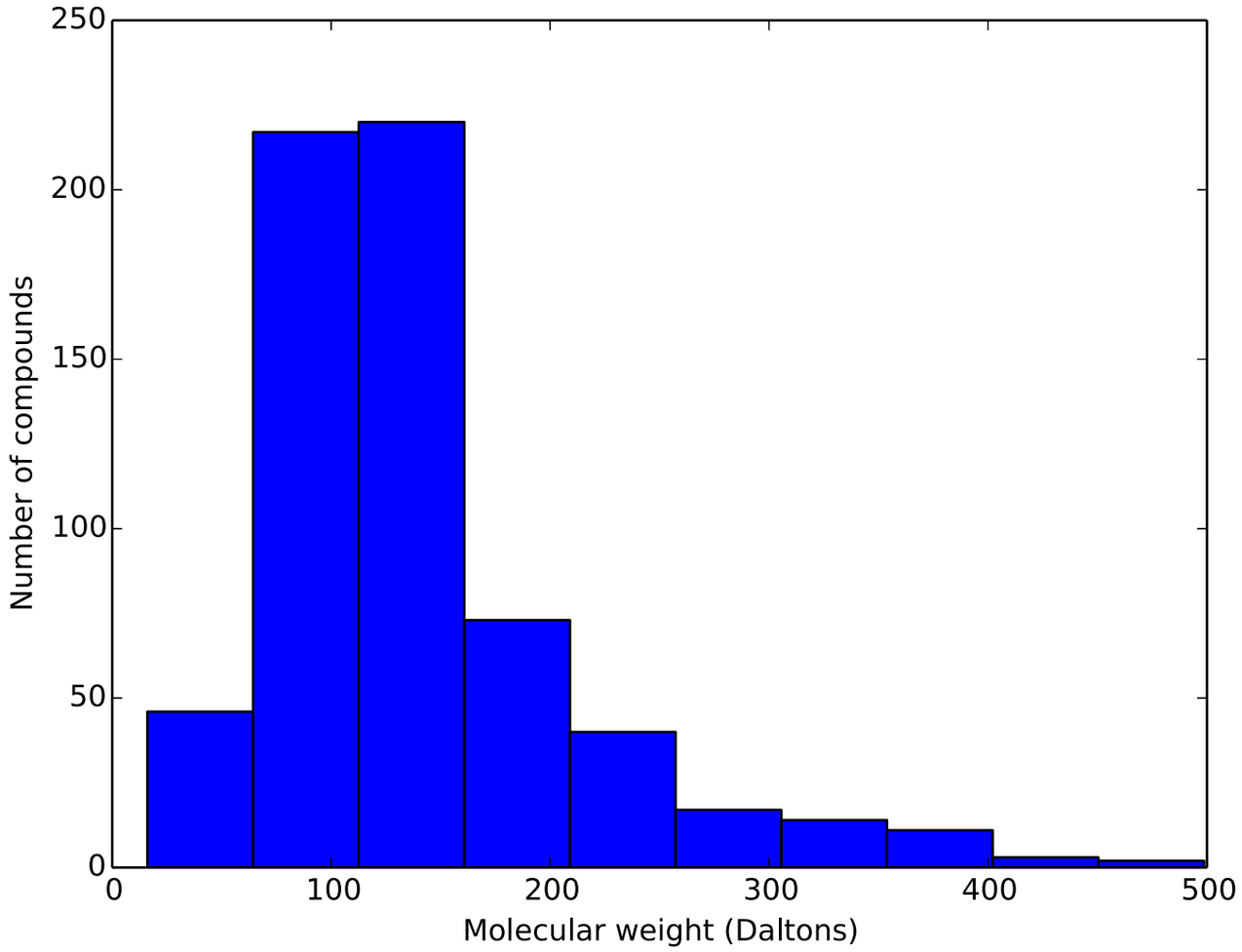
4-nitroaniline: Largest dipole

Octafluorocyclobutane
Most hydrophobic

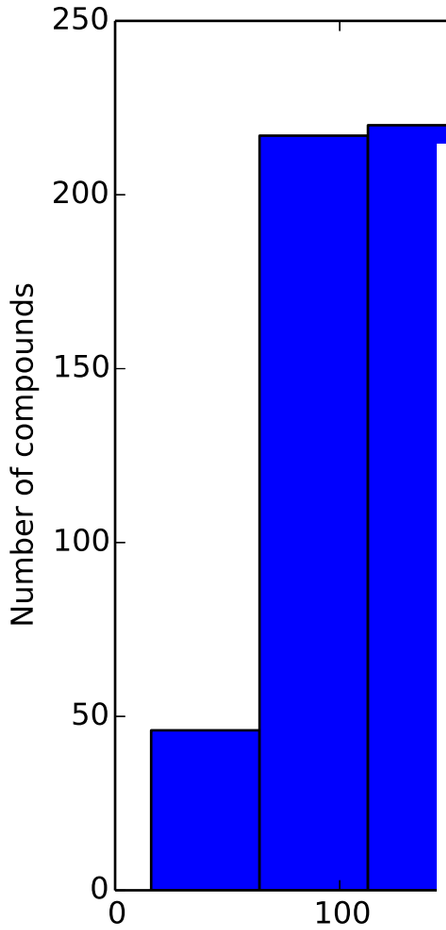


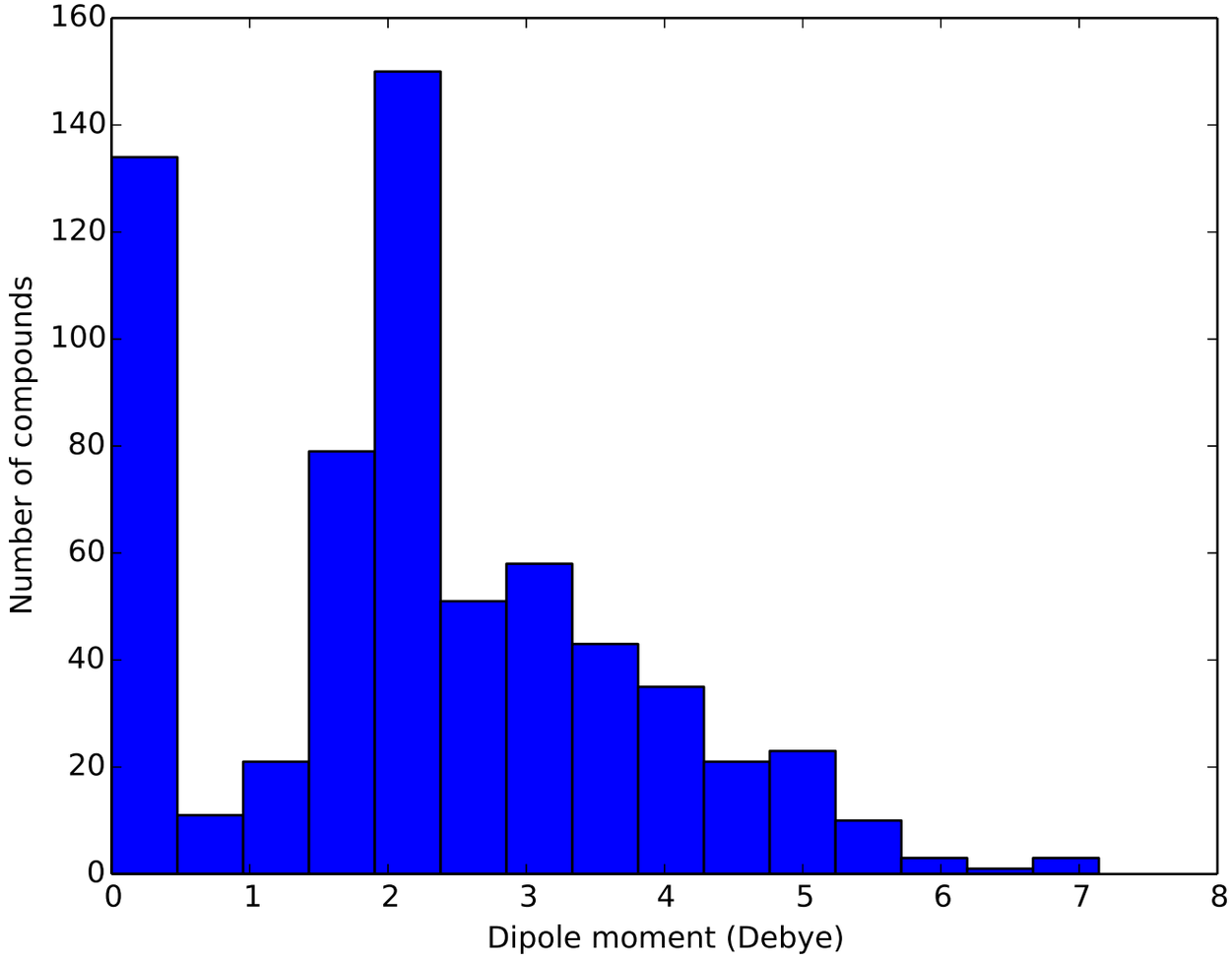
Most negative experimental value

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

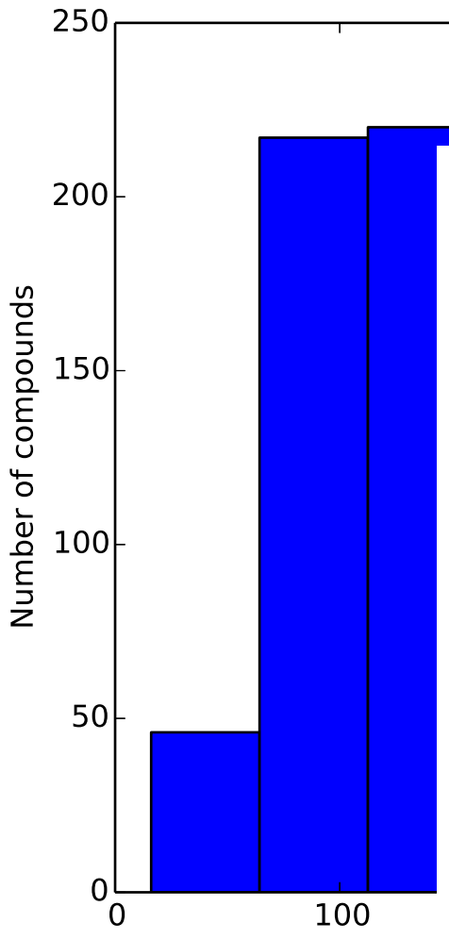


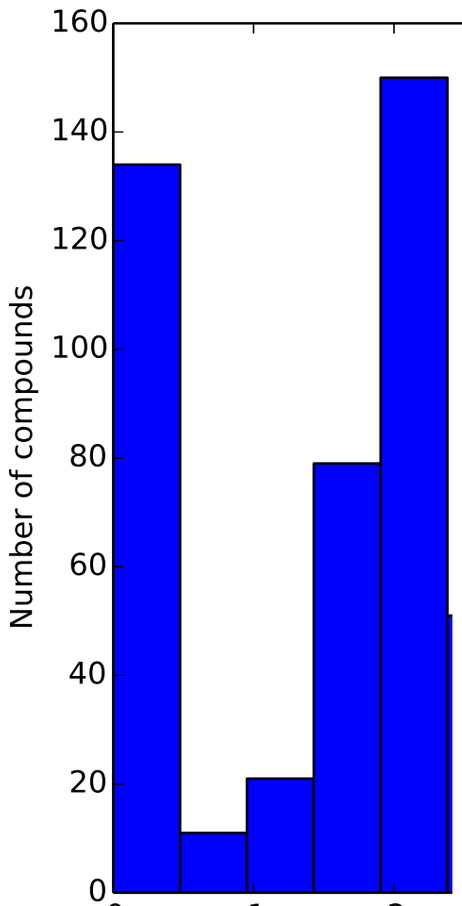
From the standpoint of testing force fields for drug-like molecules, though, we have a ways to go

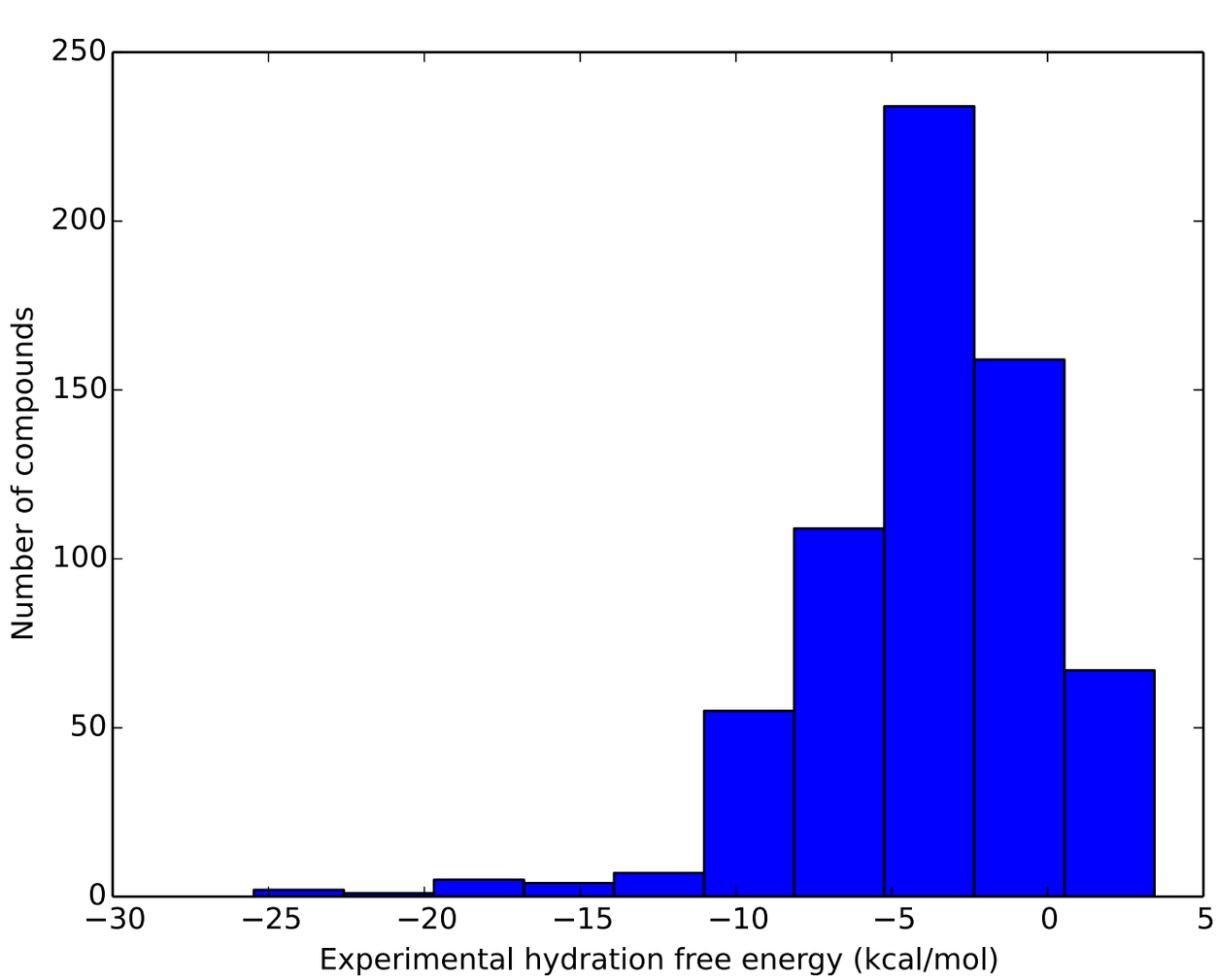




From the standpoint of testing force fields for drug-like molecules, though, we have a ways to go







Various functional groups are underrepresented or not represented compared to drugs

DrugBank

FreeSolv

DrugBank

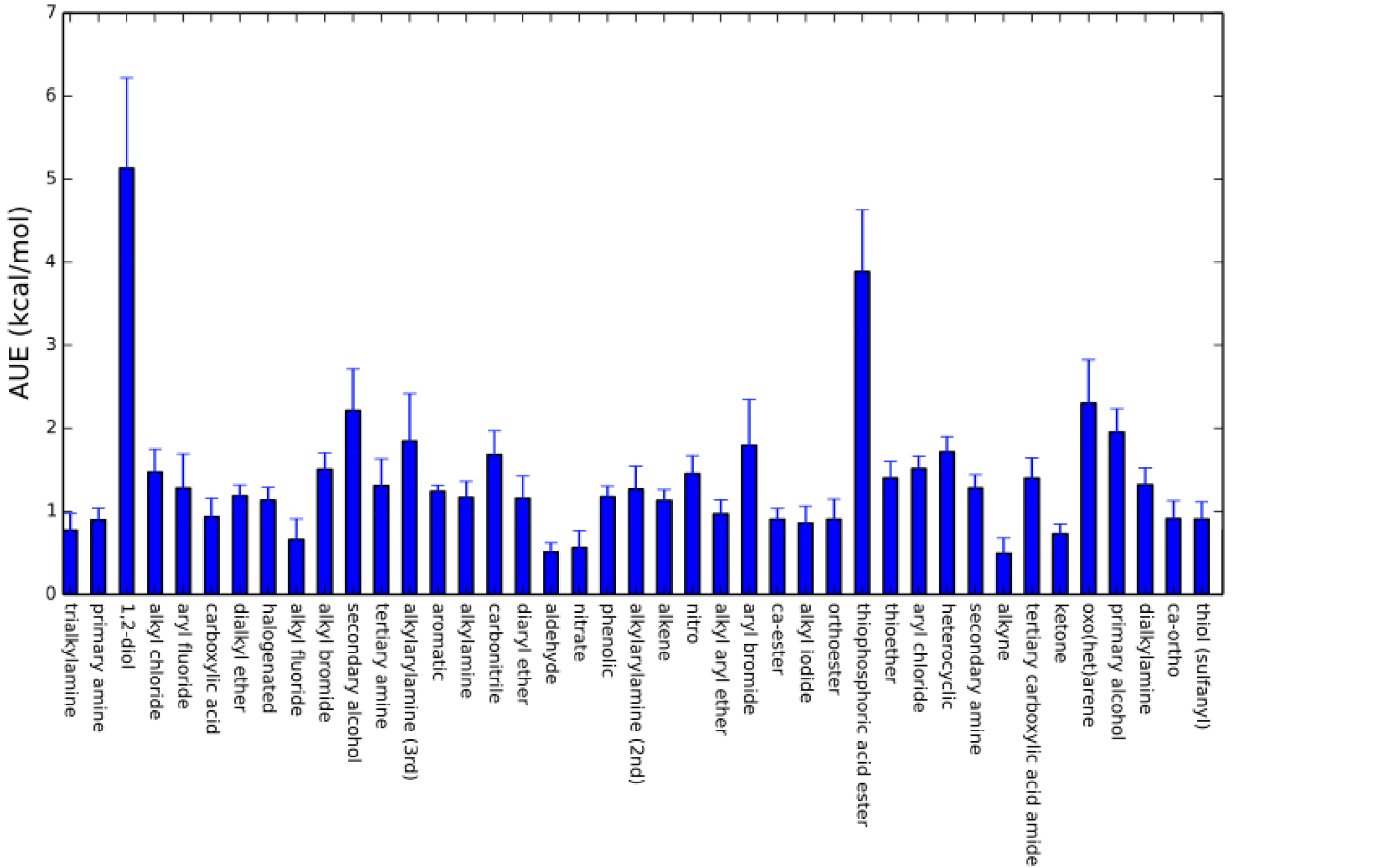
FreeSolv

thiol (sulfanyl)
thiol
thioether
thiocarboxylic acid ester
sulfuric acid monoester
sulfonic acid
sulfone
sulfonamide
quaternary ammonium salt
phenolic
oxo(het)arene
oxime/oxime ether
nitro
lactone
ketone
imino(het)arene
imine
hydroxylamine
hydrazine derivative
hemithioaminal
halogenated
guanidine
ether
enol ether
enol
enamine
carbonitrile
ca-ortho
ca-imide
ca-ester
ca-amidine
ca
boronic acid & derivs
aryl iodide
aryl bromide
amine, tertiary
amine, secondary
amine, primary
amine, alkyl
aminal or hemiaminal
alkyne
alkyl/aryl fluoride
alkyl/aryl chloride
aldehyde
alcohol, tertiary
alcohol, secondary
alcohol, primary
alcohol, 1,2-amino
acetal or hemiacetal
1,2-diol

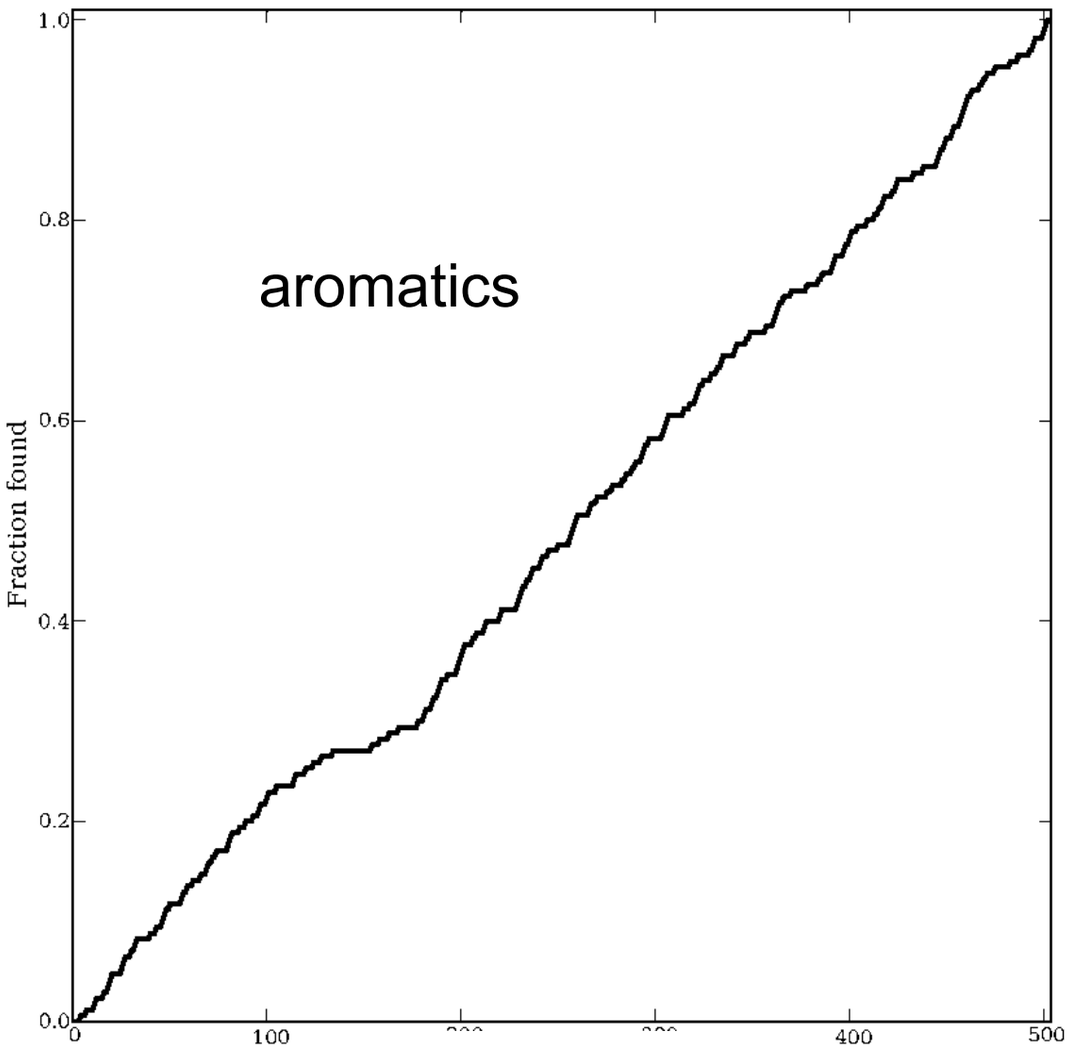
0.10 0.08 0.06 0.04 0.02 0.00

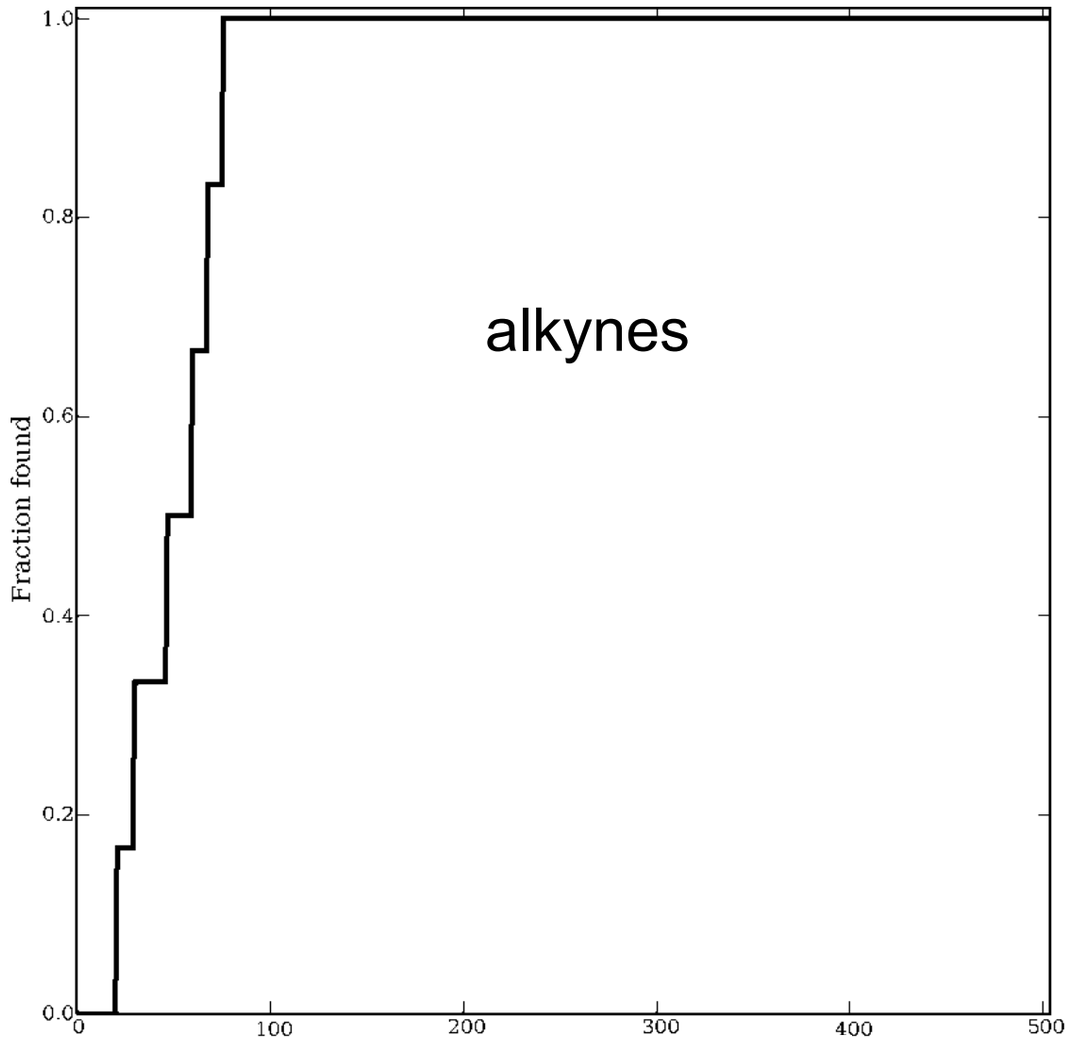
0.10 0.08 0.06 0.04 0.02 0.00

Certain functional groups appear to still be particularly problematic

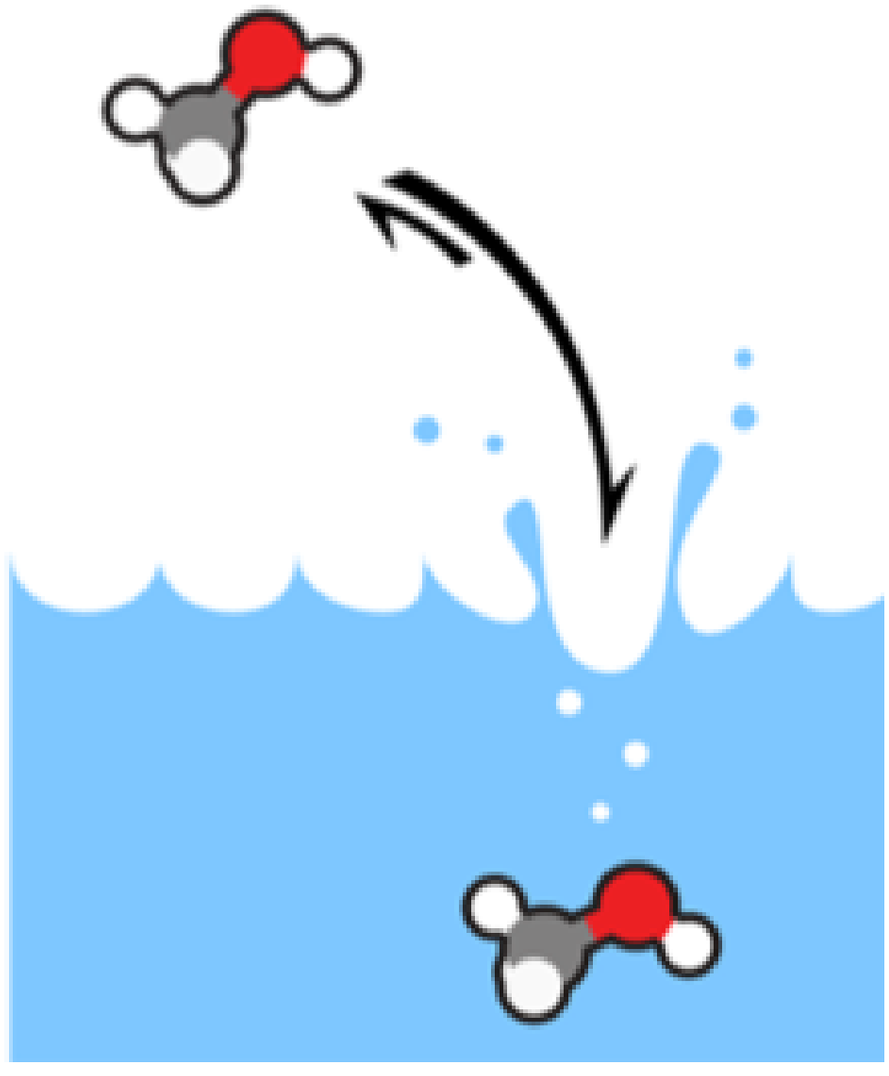


Reminder: We can look at functional groups which are overrepresented at high error





While hydration free energies have been extremely useful, we are simply running out



- Probably ~3000 in total
- Not commonly measured
- Not enough coverage of drug-like molecules
- What if we we get hydration right at the expense of other properties?

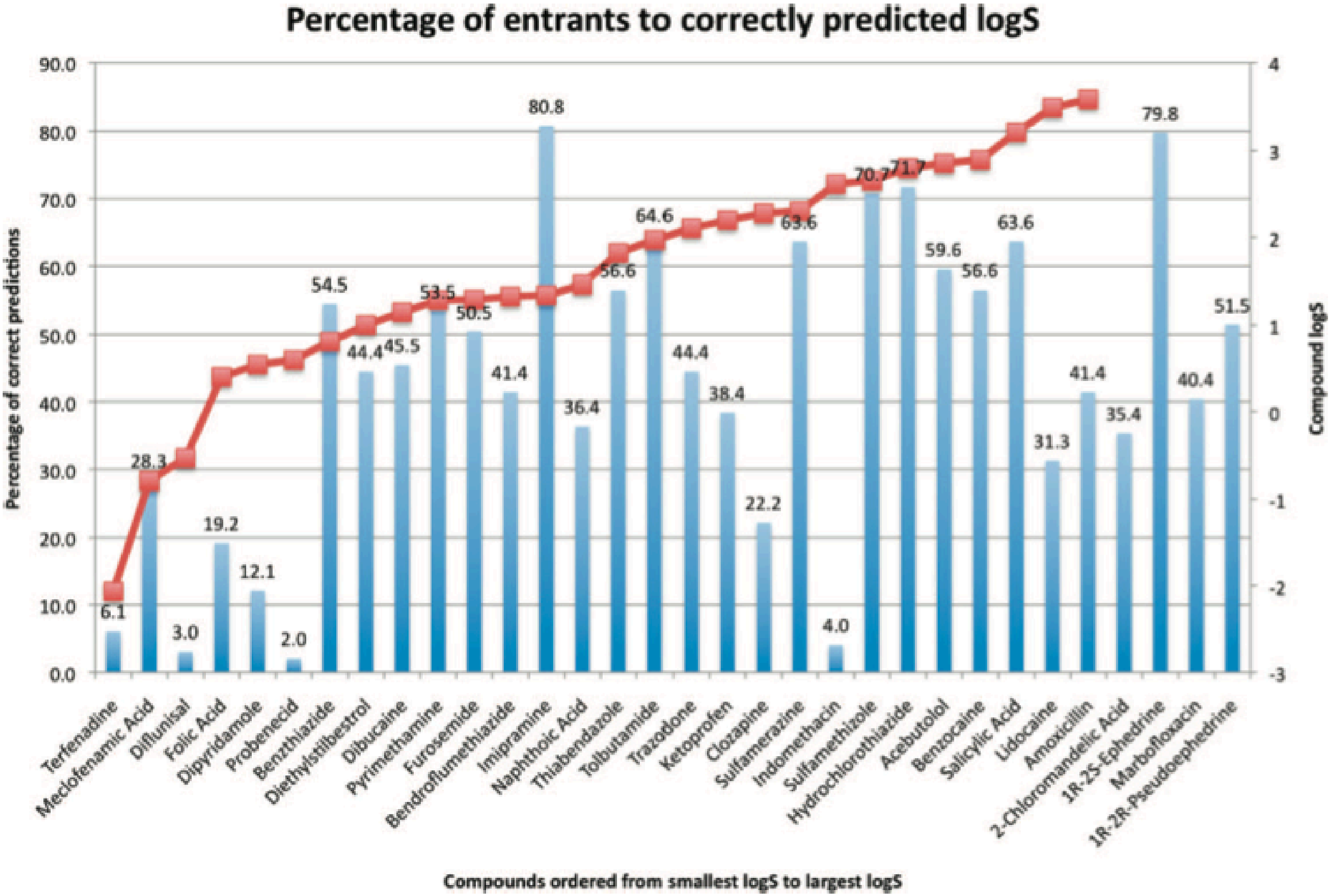
Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Reliable Measurements?

Antonio Llinàs,* Robert C. Glen, and Jonathan M. Goodman*

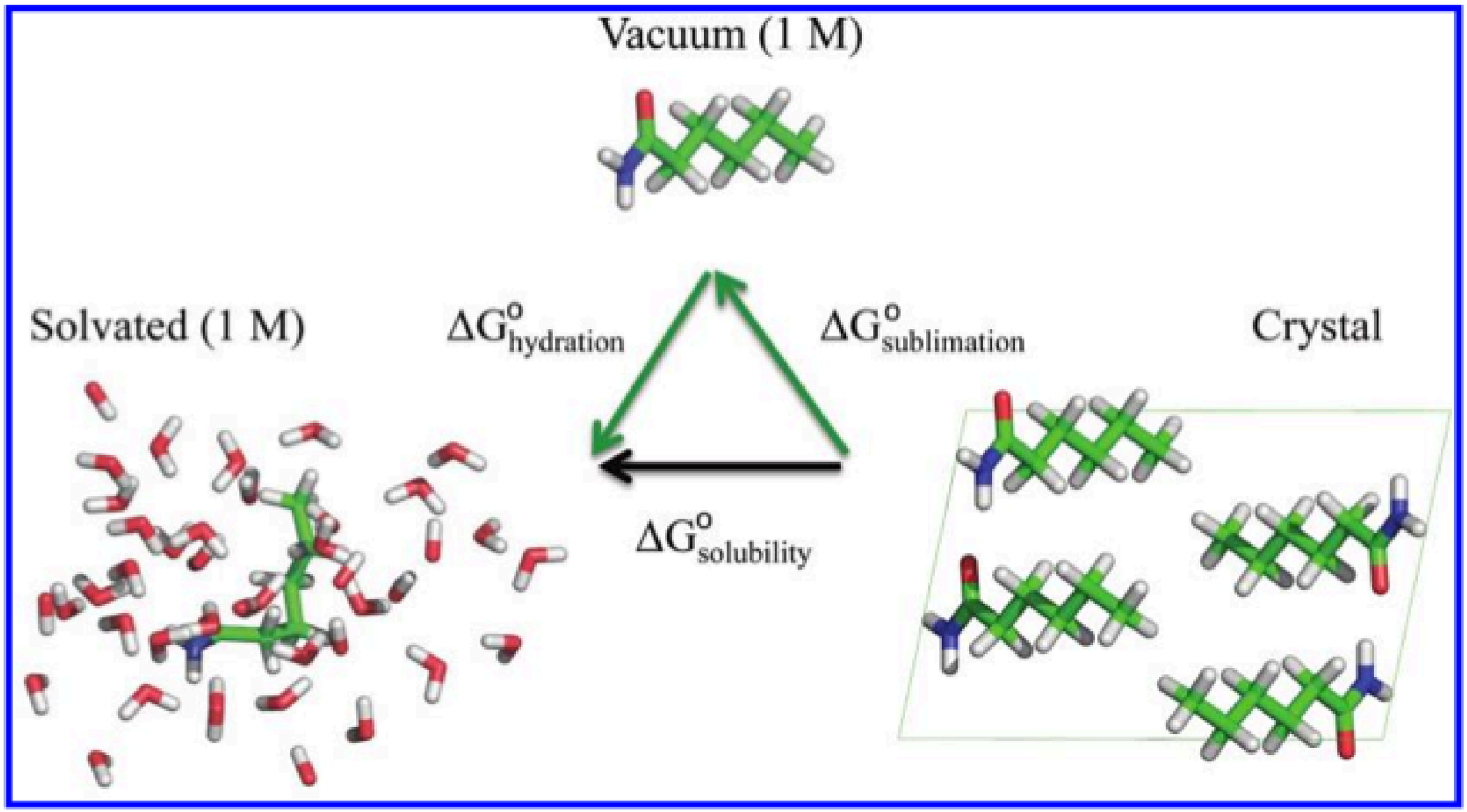
Pfizer Institute for Pharmaceutical Materials Science & Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received February 15, 2008

Solubility is a key physicochemical property of molecules. Serious deficiencies exist in the consistency and reliability of solubility data in the literature. The accurate prediction of solubility would be very useful. However, systematic errors and lack of metadata associated with measurements greatly reduce the confidence in current models. To address this, we are accurately measuring intrinsic solubility values, and here we report results for a diverse set of 100 druglike molecules at 25 °C and an ionic strength of 0.15 M using the CheqSol approach. This is a highly reproducible potentiometric technique that ensures the thermodynamic equilibrium is reached rapidly. Results with a coefficient of variation higher than 4% were rejected. In addition, the Potentiometric Cycling for Polymorph Creation method, $[PC]^2$, was used to obtain multiple polymorph forms from aqueous solution. We now challenge researchers to predict the intrinsic solubility of 32 other druglike molecules that have been measured but are yet to be published.

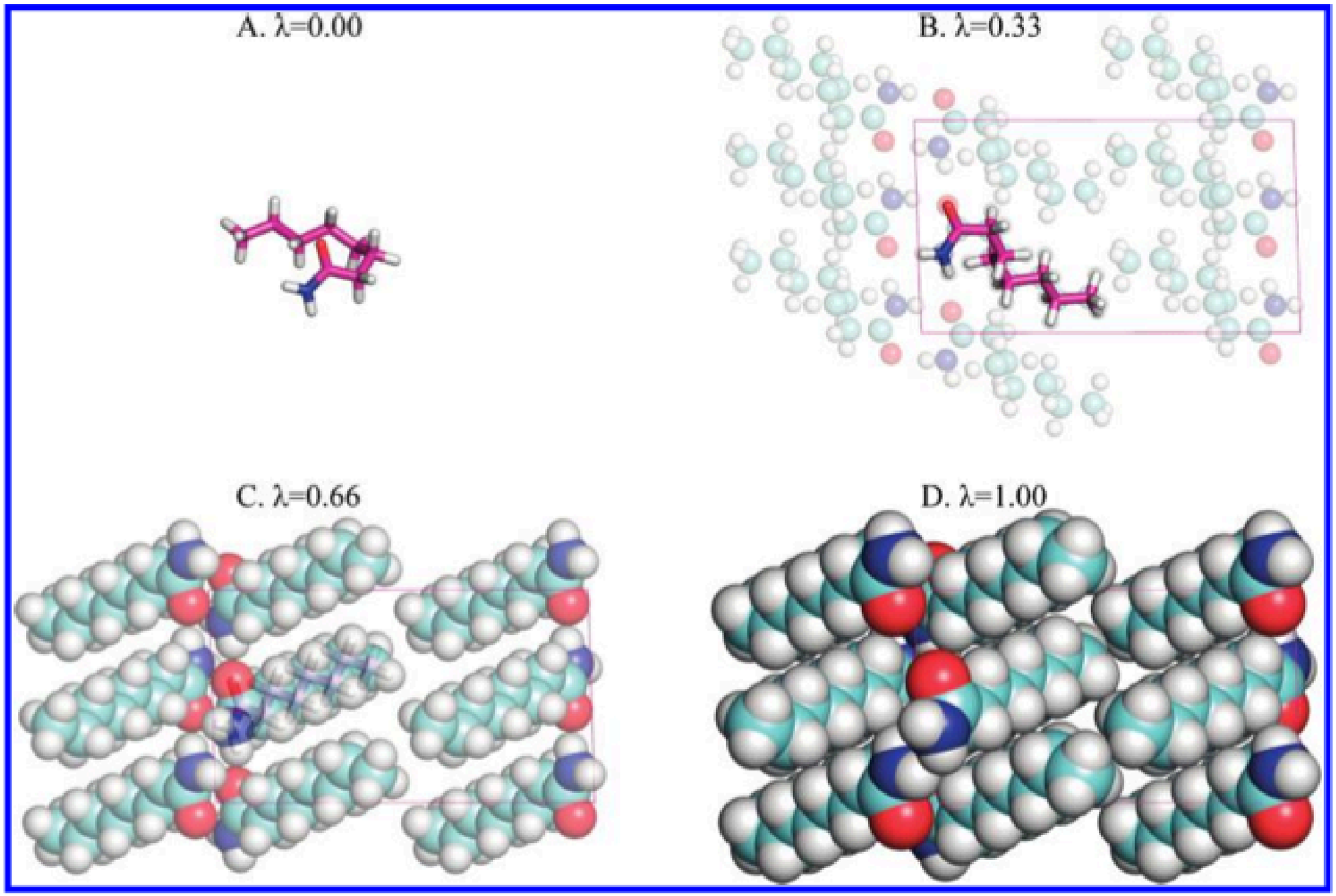

Percentage of entrants to correctly predicted logS
Percentage of correct predictions
Compound logS
Compounds ordered from smallest logS to largest logS
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
4
3
2
1
0
-1
-2
-3
28.3
6.1
3.0
19.2
12.1
2.0
54.5
44.4
45.5
53.5
50.5
41.4
80.8
36.4
56.6
64.6
44.4
38.4
22.2
63.6
4.0
70.7
71.7
59.6
56.6
63.6
31.3
41.4
35.4
79.8
40.4
51.5
Terfenadine
Mefenamic Acid
Diflunisal
Folic Acid
Dipyridamole
Probenecid
Benzthiazide
Diethylstilbestrol
Dibucaine
Pyrimethamine
Furosemide
Bendroflumethiazide
Imipramine
Naphthoic Acid
Thiabendazole
Tolbutamide
Trazodone
Ketoprofen
Clozapine
Sulfamerazine
Indomethacin
Sulfamethizole
Hydrochlorothiazide
Acebutolol
Benzocaine
Salicylic Acid
Lidocaine
Amoxicillin
2-Chloromandelic Acid
1R-2S-Ephedrine
Marbofloxacin
1R-2R-Pseudoephedrine

Some efforts are taking solubility prediction in more physical directions



Schnieders et al., JCTC 8:1721-1736 (2012)

Sublimation is calculated via alchemical techniques, as is solvation



Schnieders et al., JCTC 8:1721-1736 (2012)

Results on an initial series appear promising without any empirical tuning

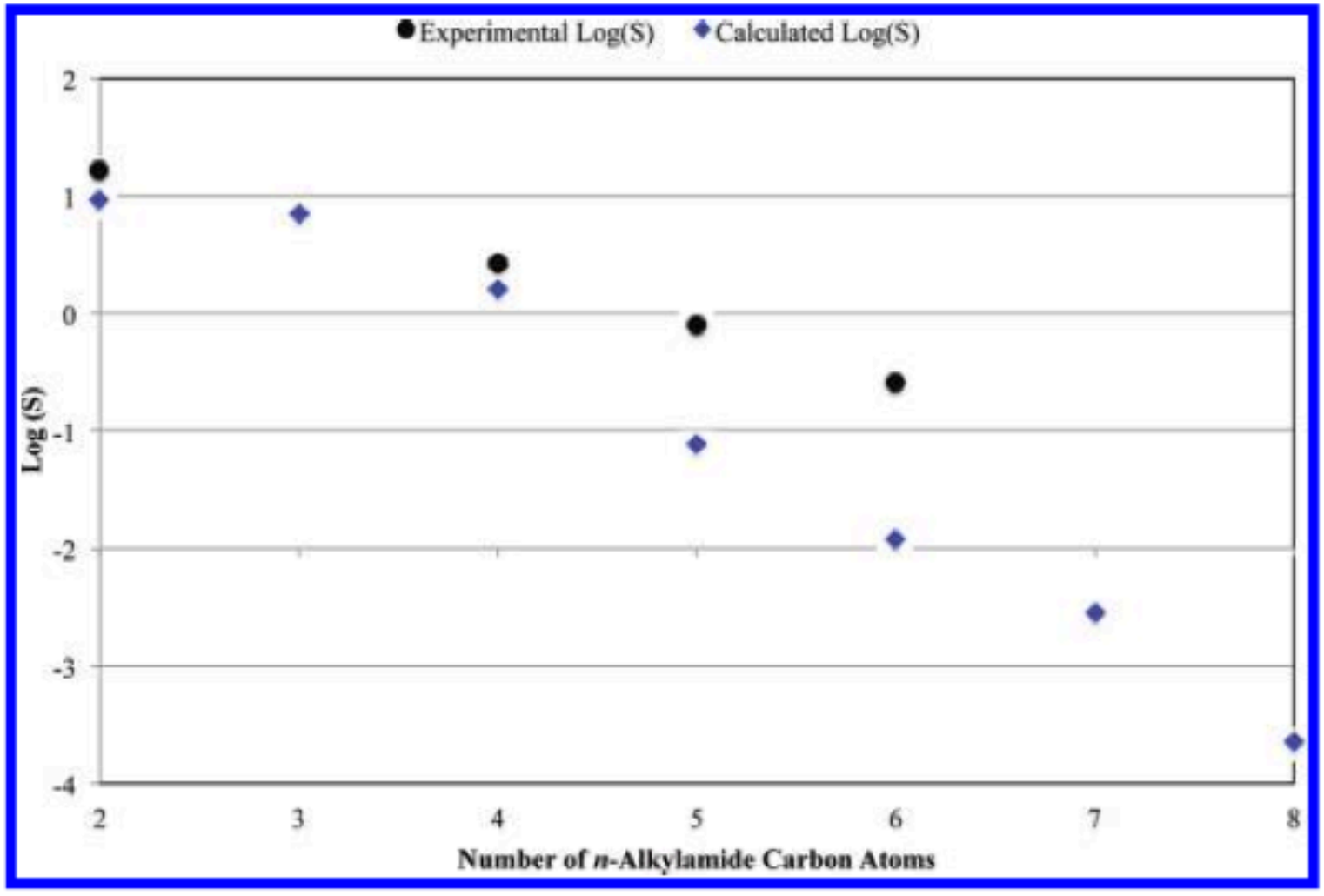
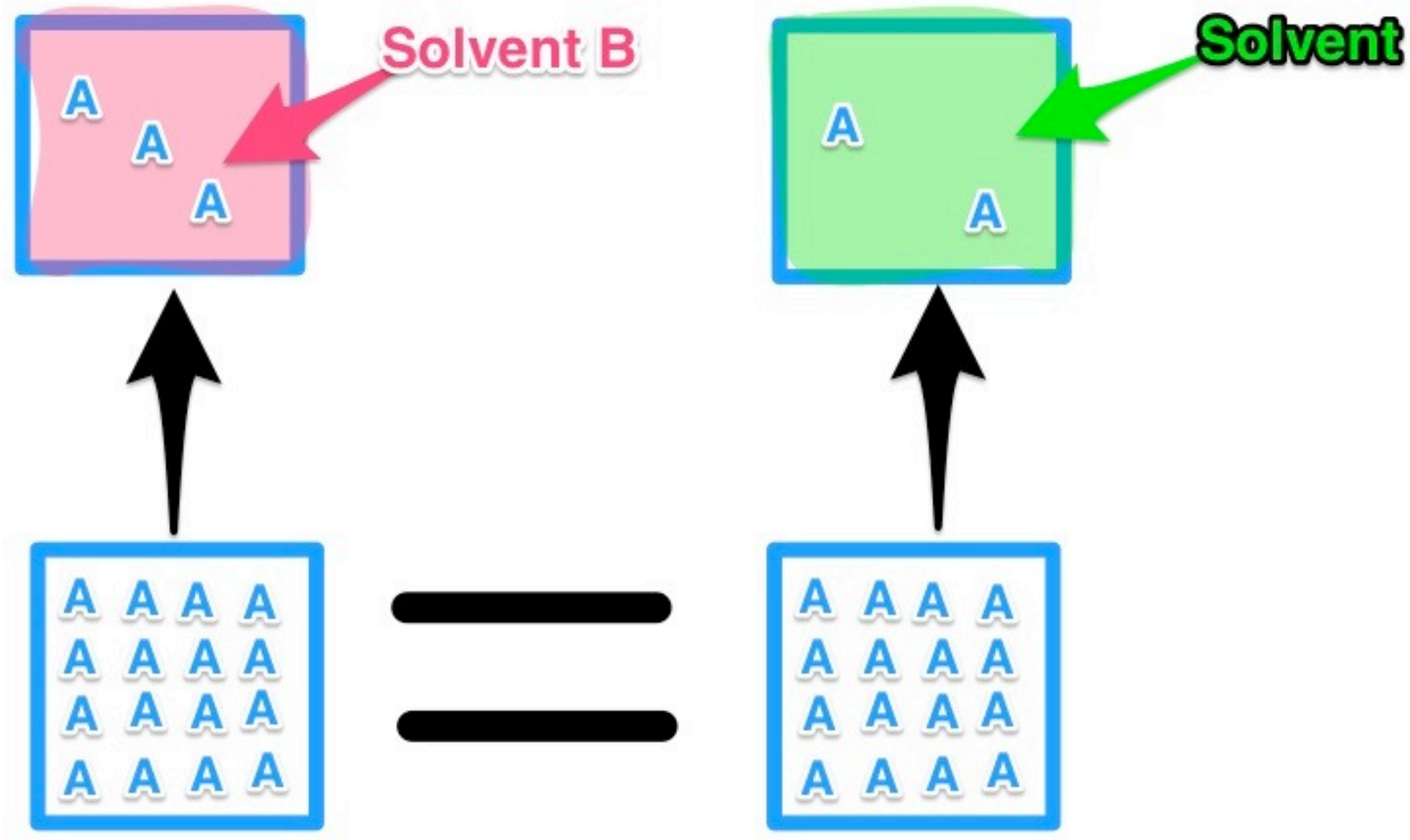


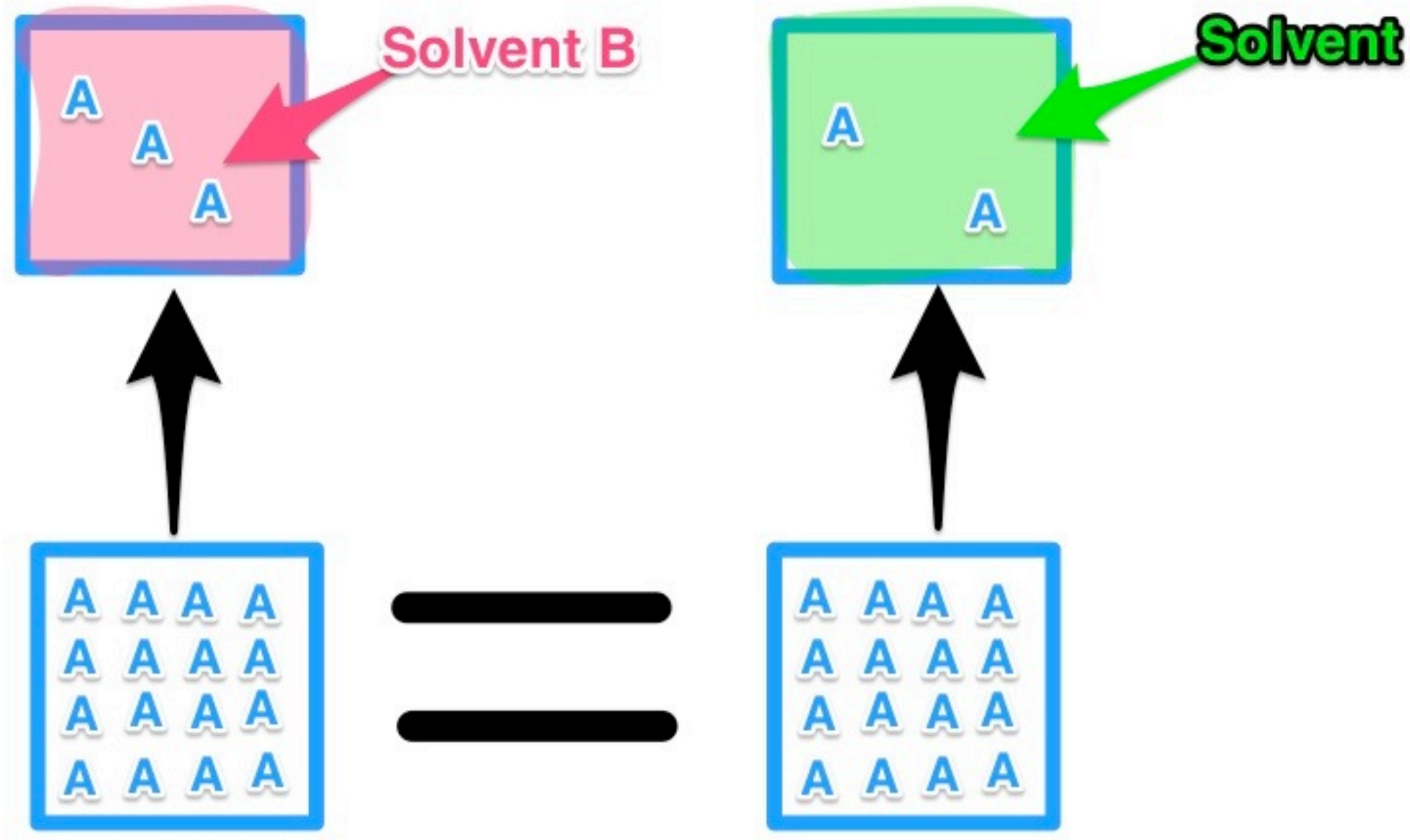
Figure 5. Shown are experimental and calculated log(*S*) values for the *n*-alkylamides (*S* has units of mol/L) from acetamide to octanamide. There is a monotonic trend in both the experimental and calculated values toward lower solubility with each additional CH_2 group due to increasingly favorable deposition and to a lesser extent from unfavorable solvation.

Schnieders et al., JCTC 8:1721-1736 (2012)

We took a different angle: What if we want to avoid the solid phase?

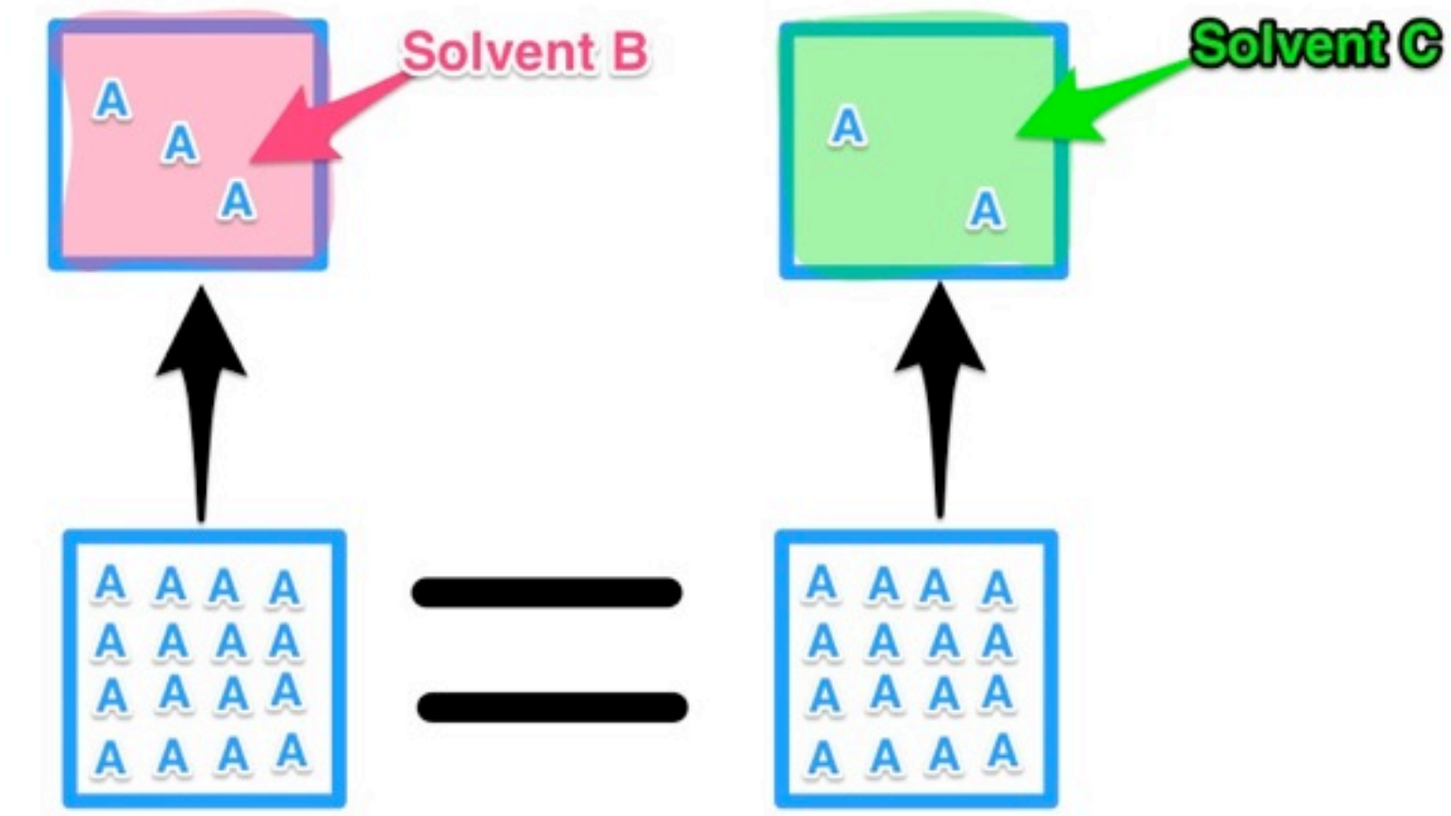


We took a different angle: What if we want to avoid the solid phase?



Additional plus: Not biased by water

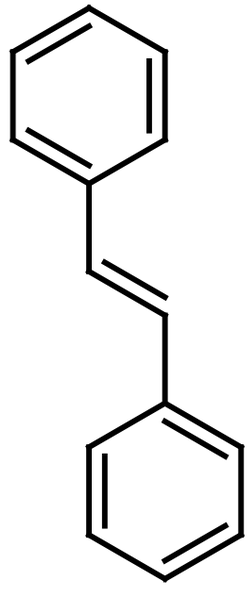
At infinite dilution, a relative solubility calculation is two solvation free energy calculations



$$\ln \frac{c_1^\alpha}{c_1^\zeta} = \ln \left(\frac{x_1^\alpha}{x_1^\zeta} \frac{v_\zeta(T,p)}{v_\alpha(T,p)} \right) = \beta\mu_1^{\zeta,\mathrm{res},\infty}(T,p) - \beta\mu_1^{\alpha,\mathrm{res},\infty}(T,p)$$

Small problem: There is some arbitrariness in how we analyze

trans-stilbene



2,2,4-trimethylpentane vs tert-butylcyclohexane

expt. -0.6; calc. -0.2(2)

benzene vs tetrahydrofuran

expt. -0.6; calc. -1.2(2)

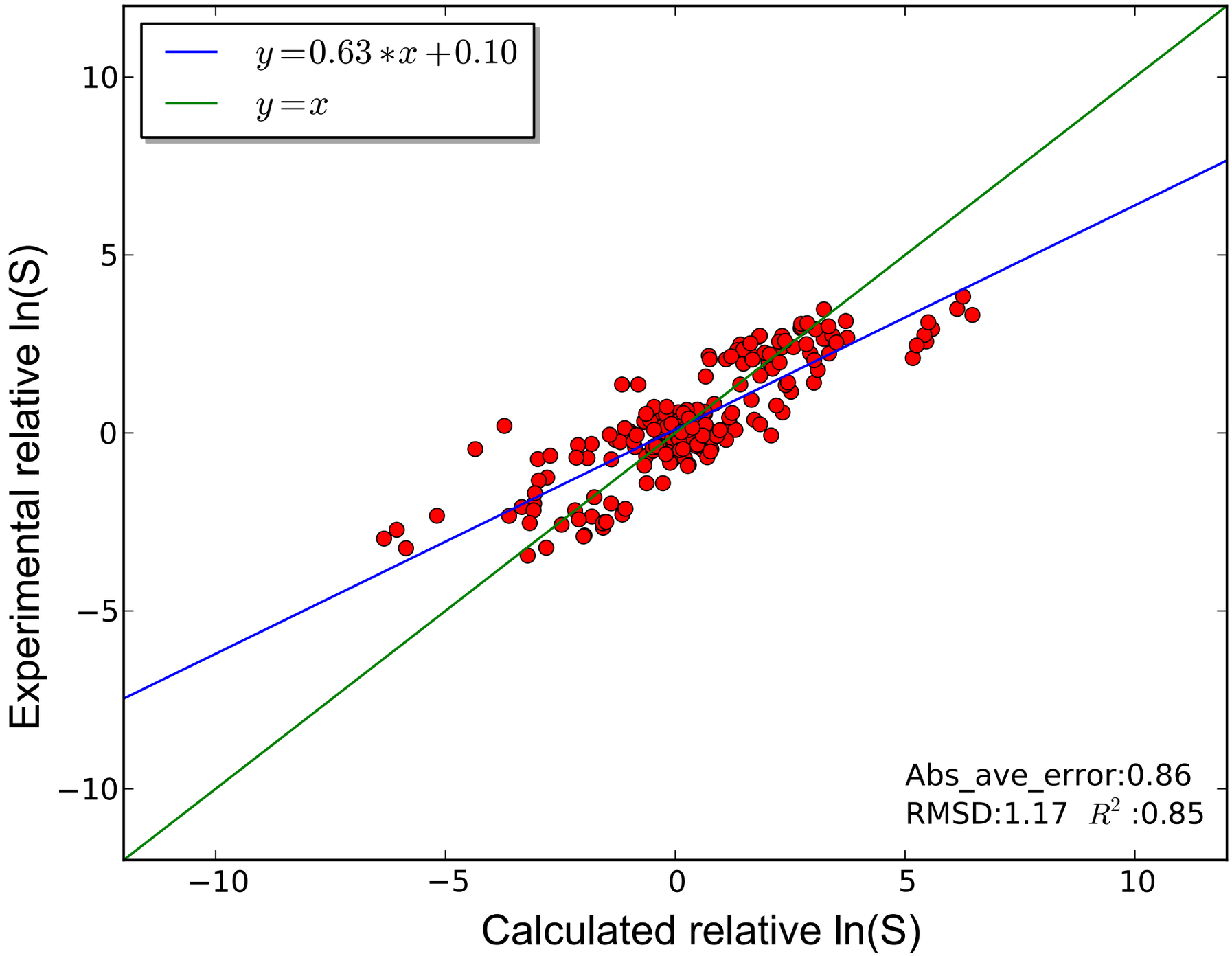
benzene vs 2,2,4-trimethylpentane

expt. 2.58; calc. 2.5(2)

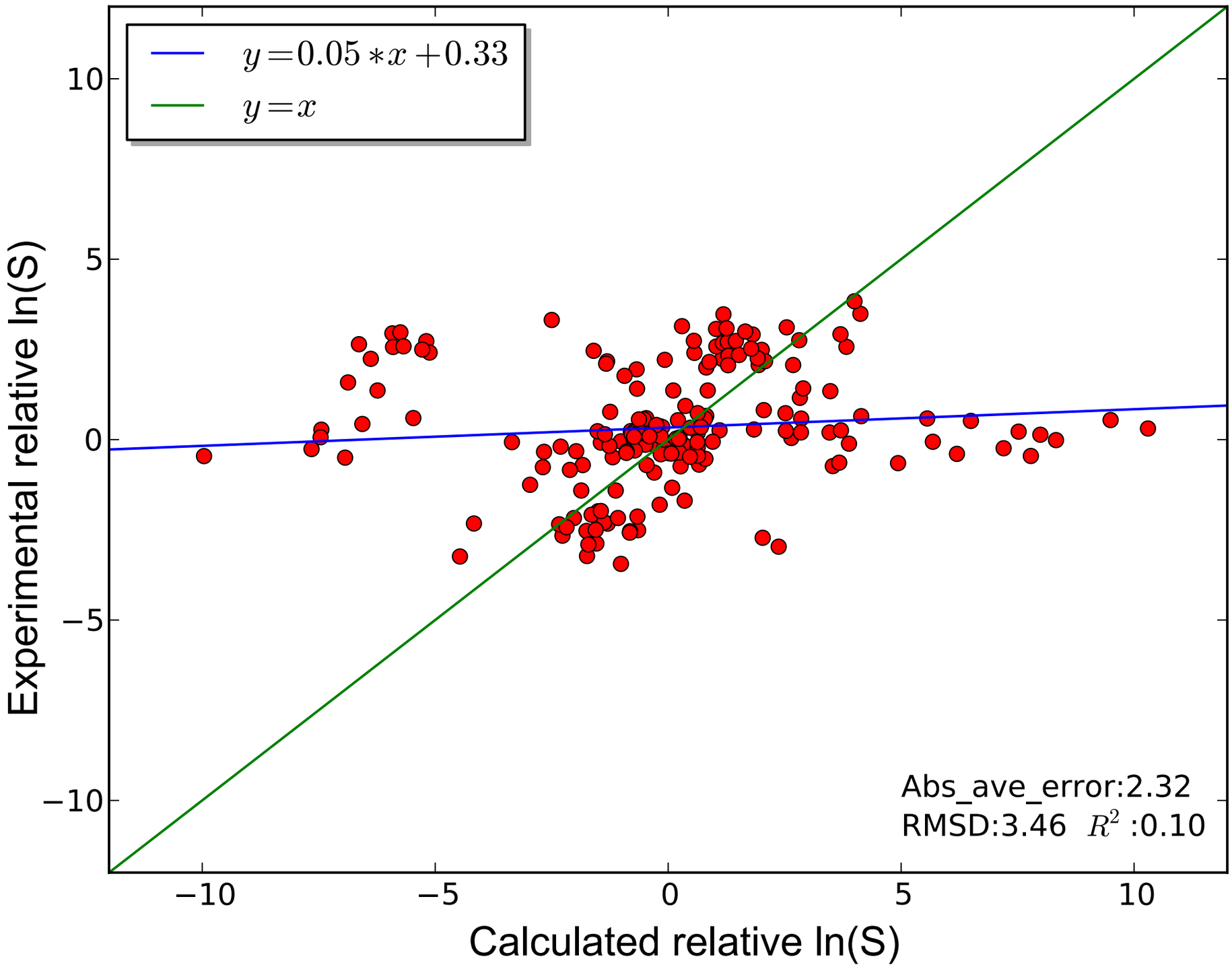
Here, we remove the arbitrariness by considering all possible pairs

(8 solutes, 29 solvents, 55 combinations)

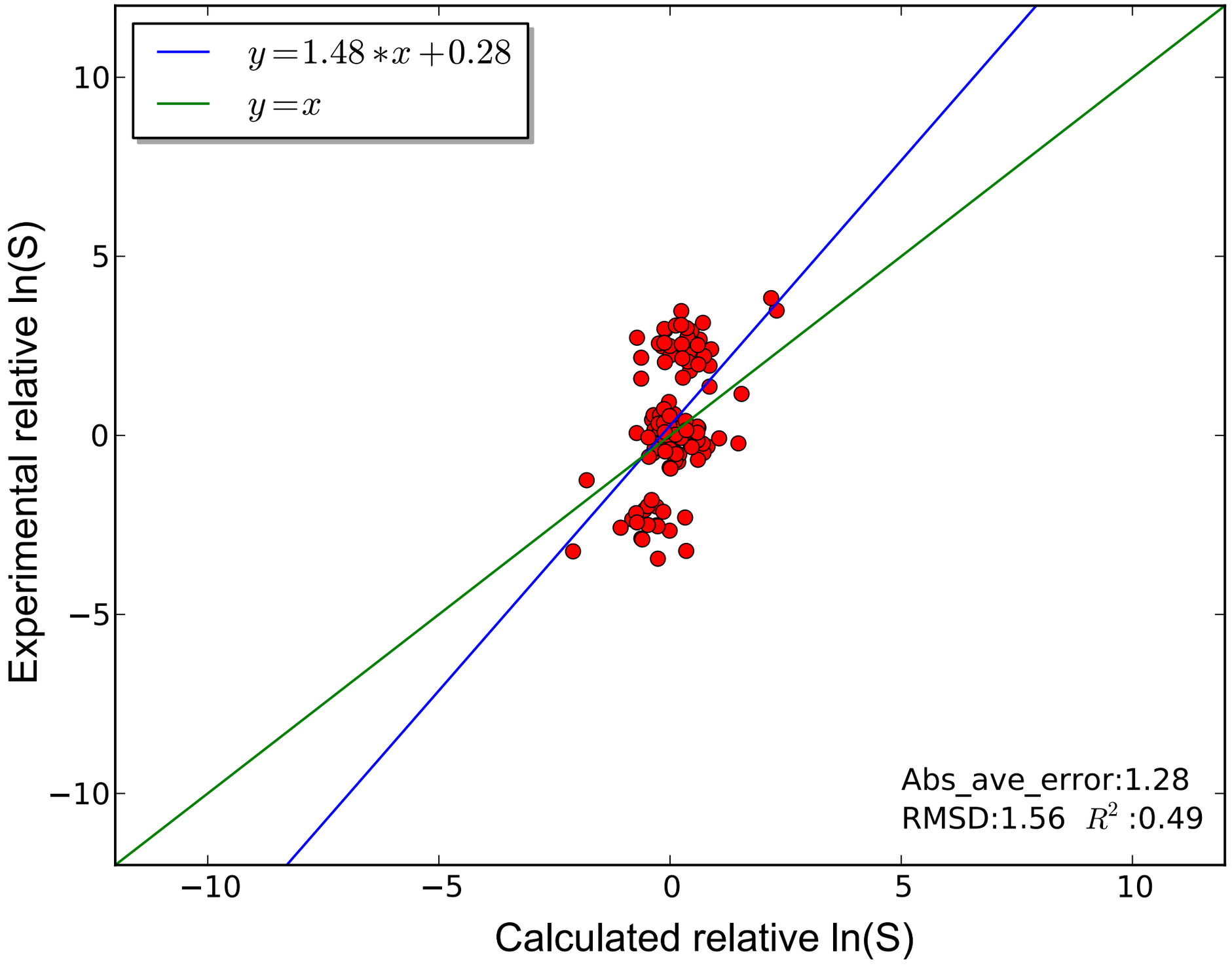
On the whole calculated and experimental values agree rather well



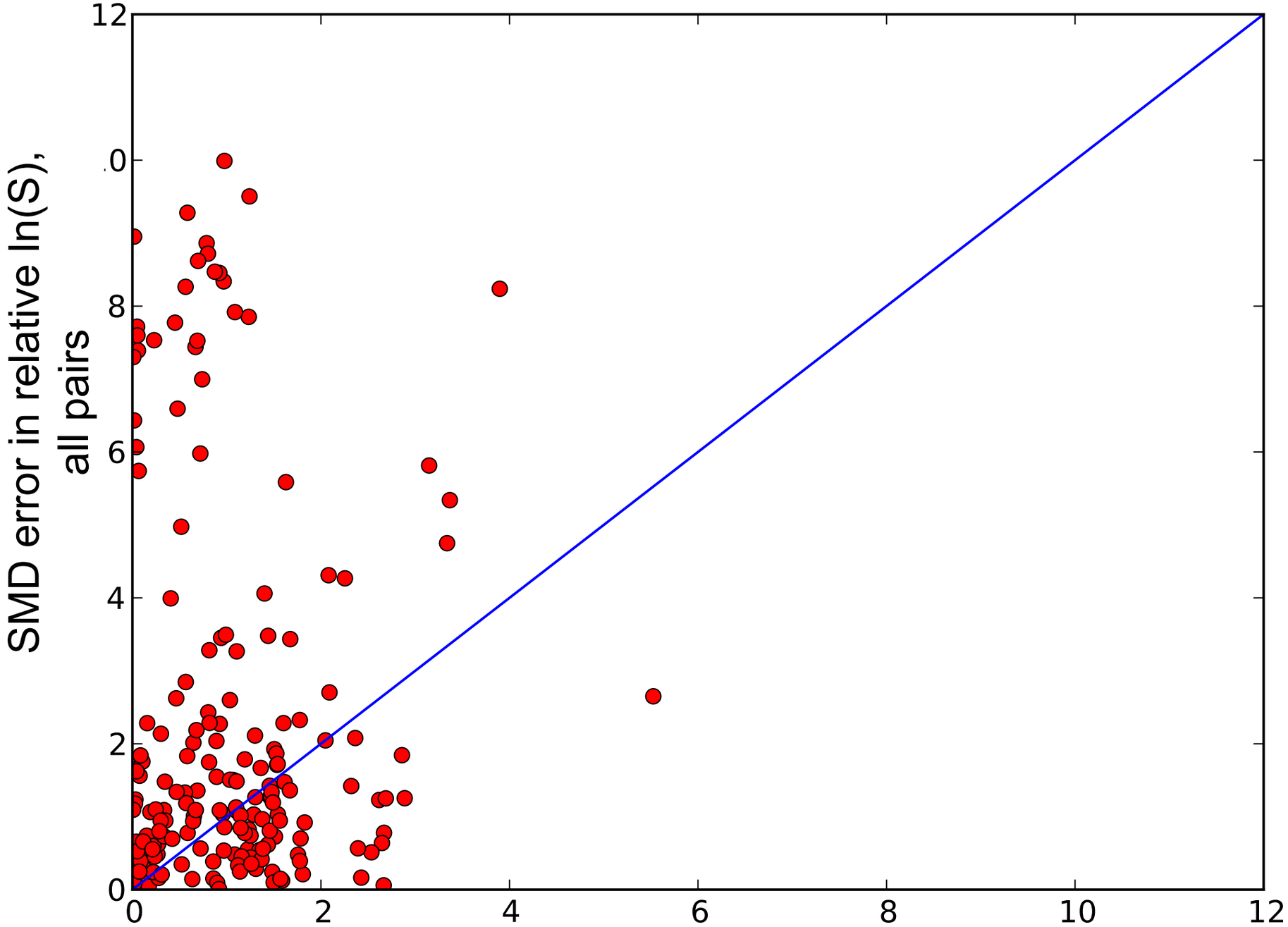
For comparison, SMD is a QM-based solvation model with empirical solvation parameters



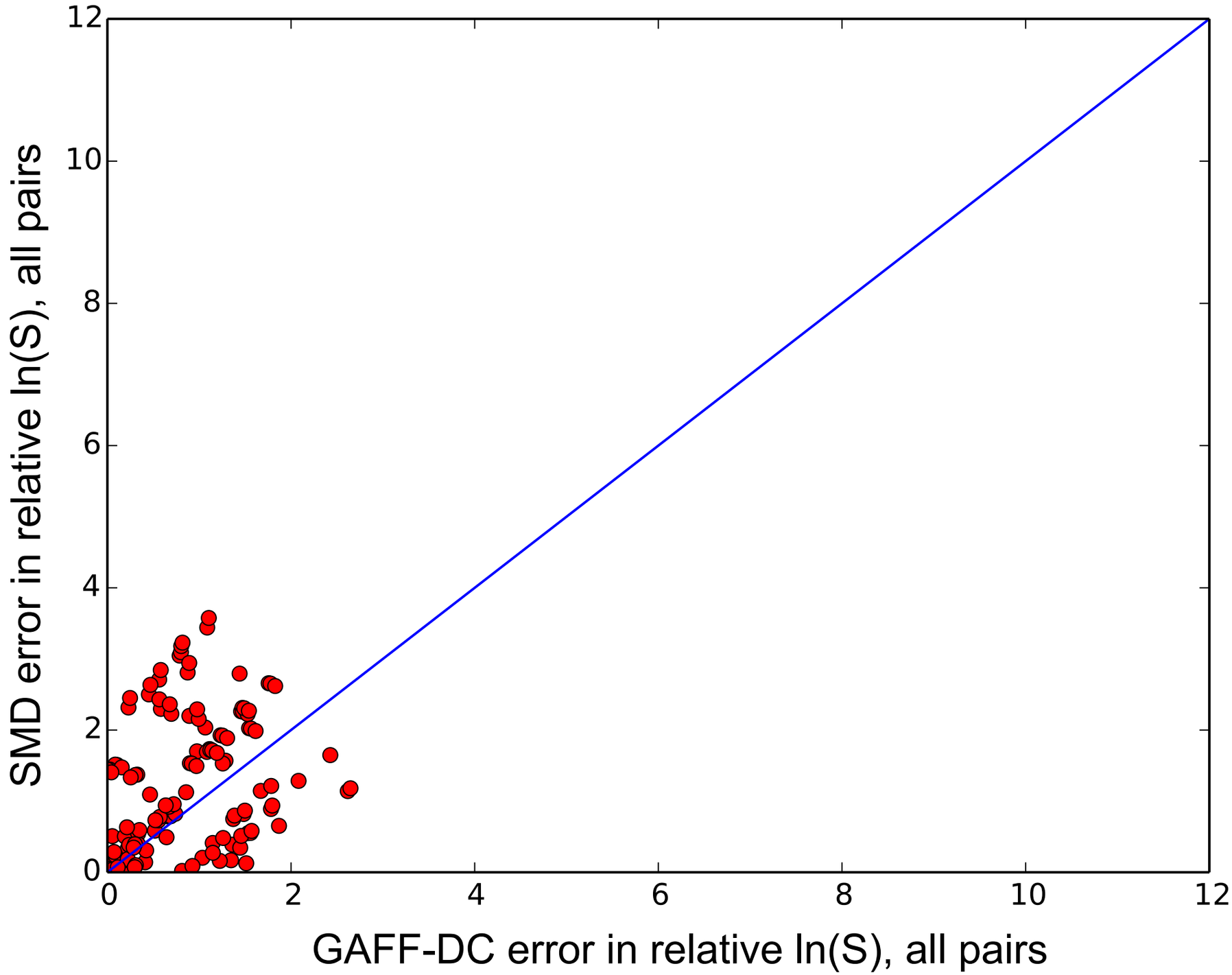
UNIFAC does somewhat better but does not cover the whole set



Another way of looking at this is a parity plot of errors, GAFF vs SMD

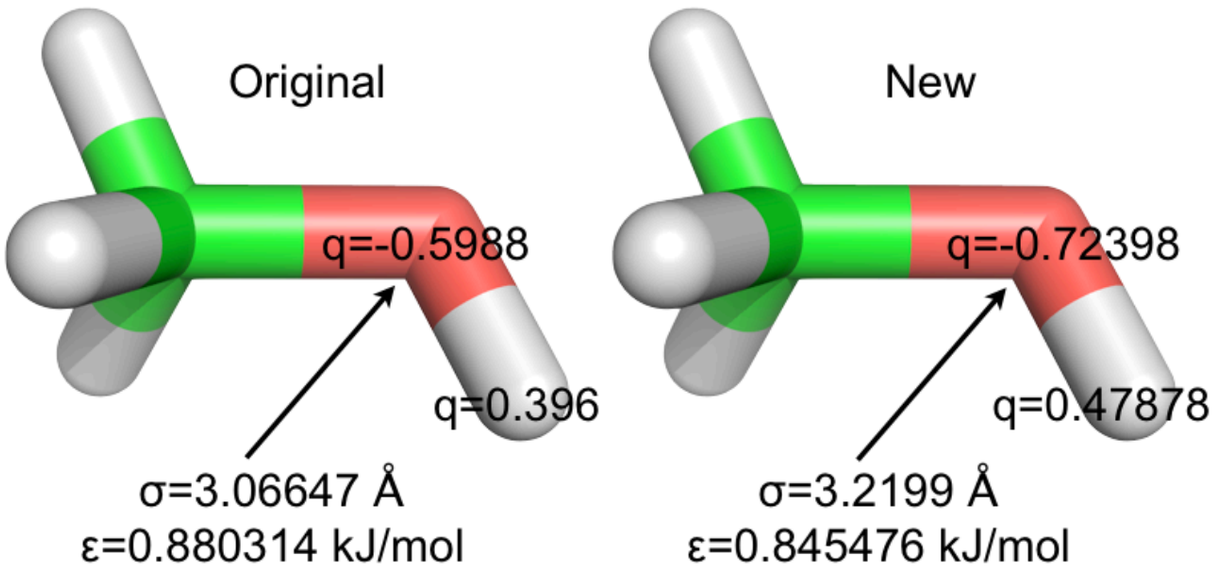


Another way of looking at this is a parity plot of errors, GAFF vs UNIFAC



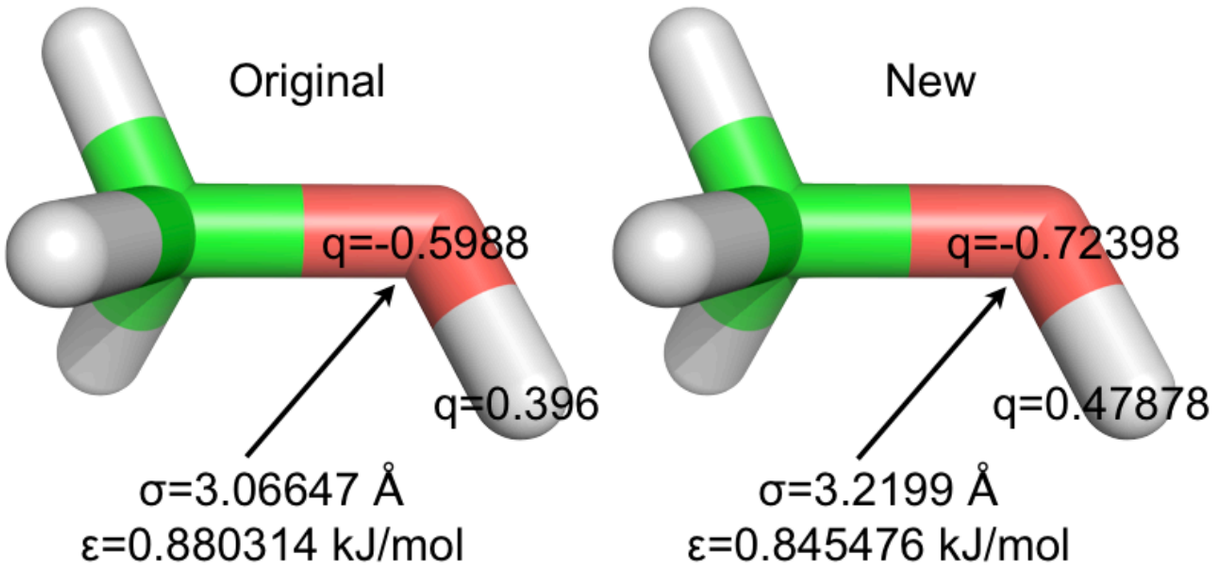
Conclusions

Conclusions

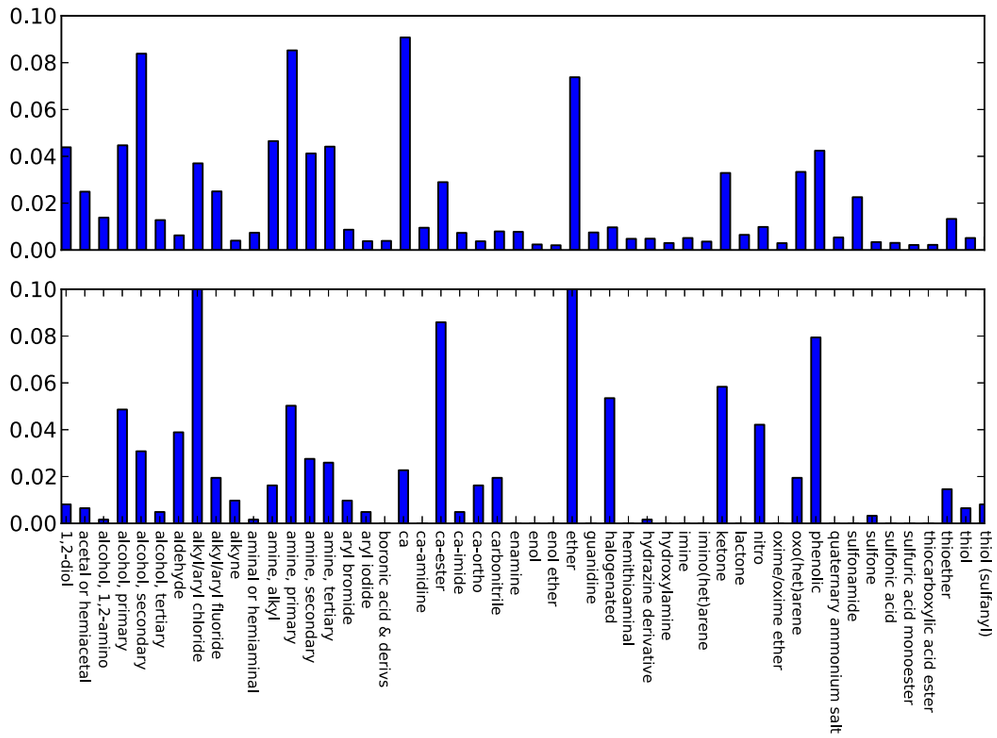


We have new GAFF-DC hydroxyl parameters

Conclusions

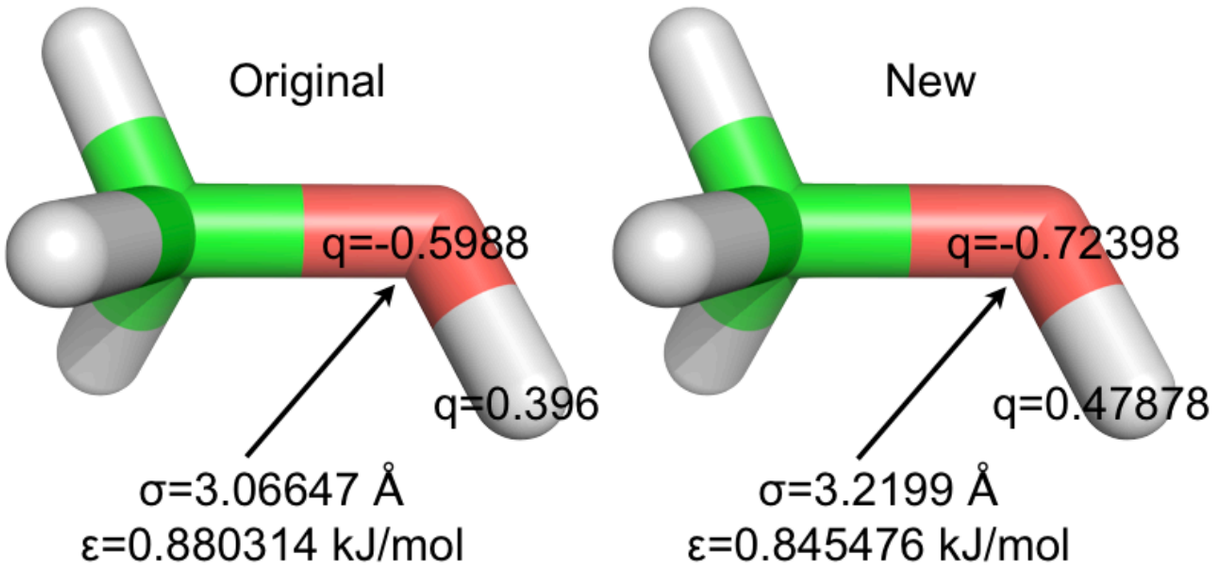


We have new GAFF-DC hydroxyl parameters

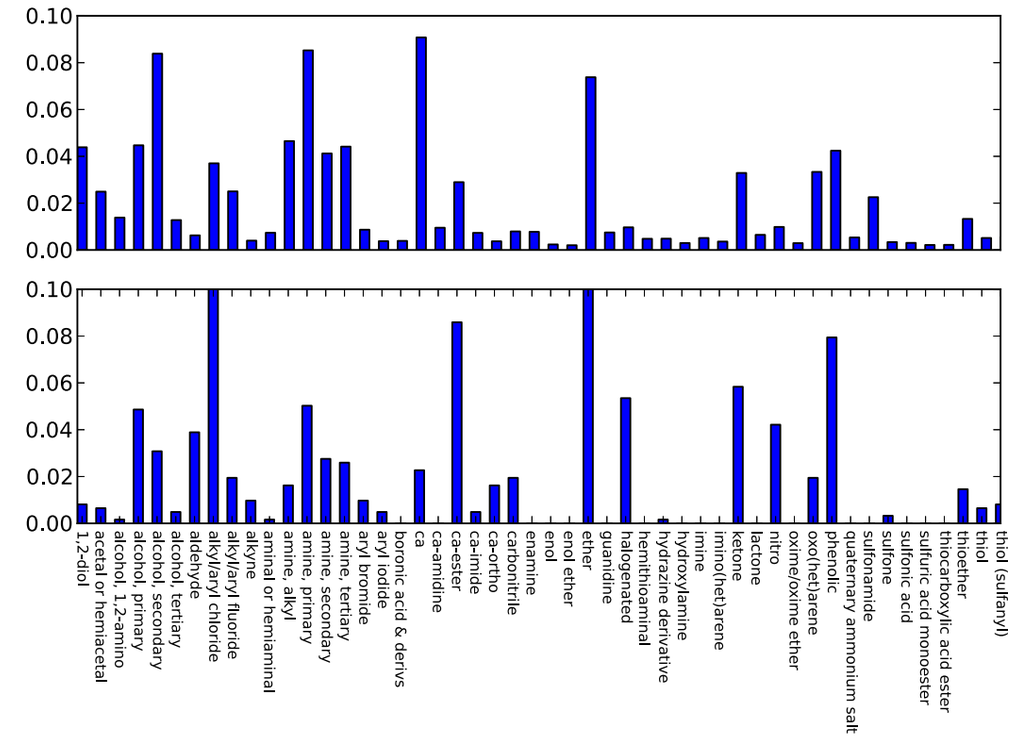


The FreeSolv database is available, but we need more

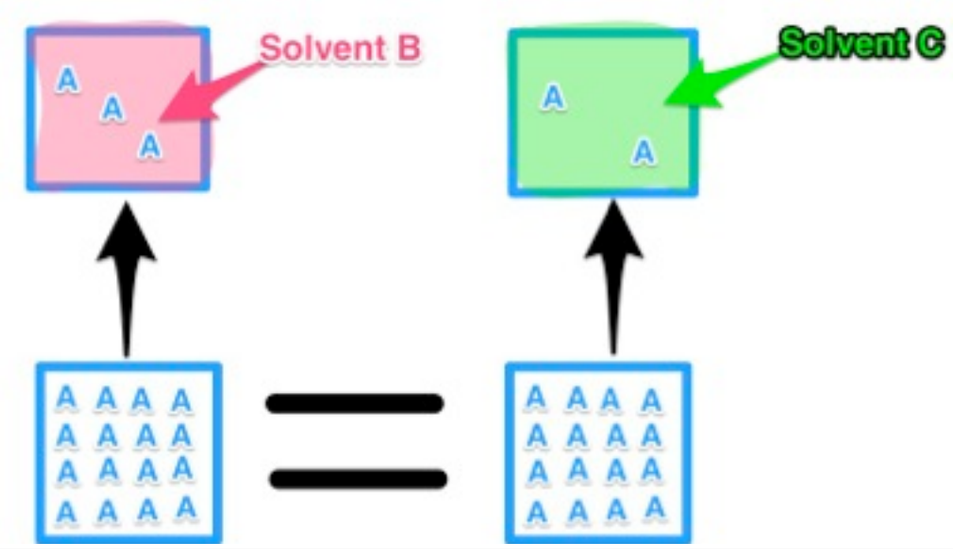
Conclusions



We have new GAFF-DC hydroxyl parameters



The FreeSolv database is available, but we need more

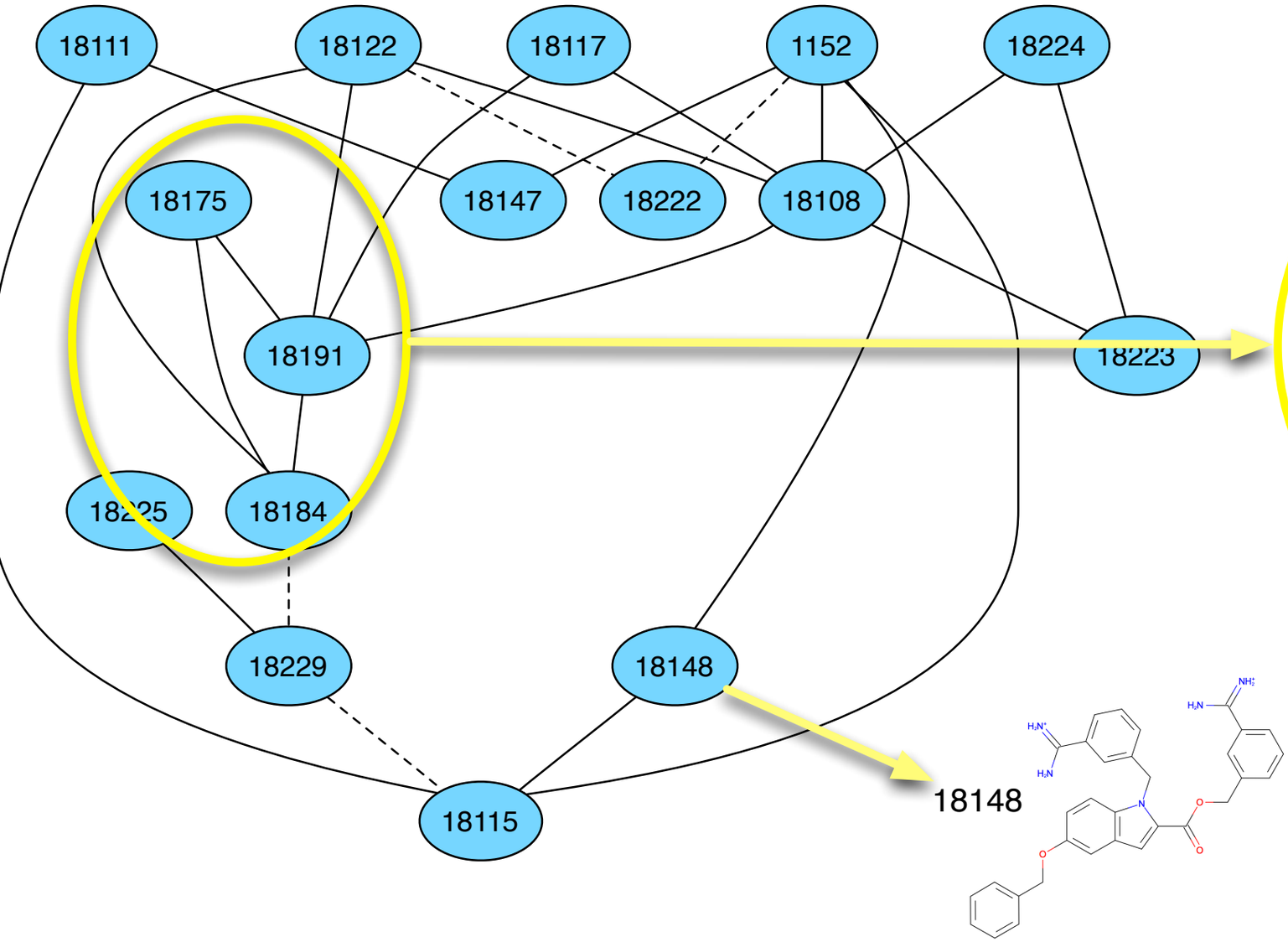


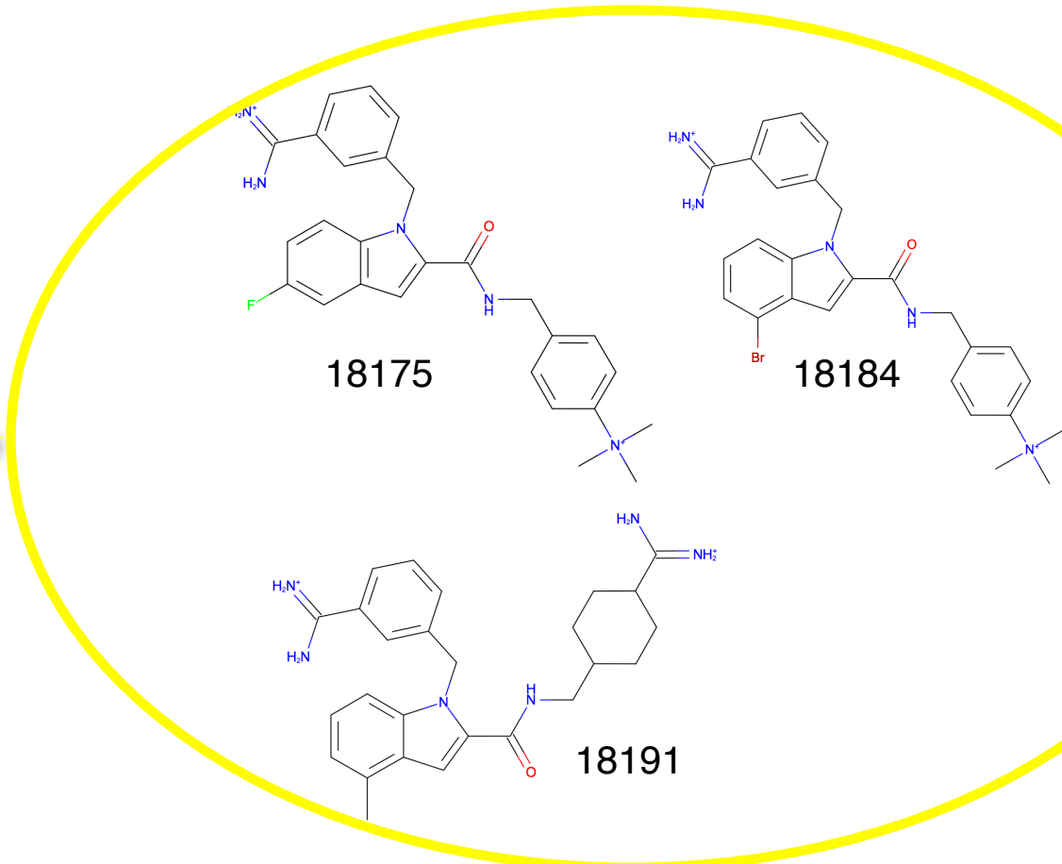
Relative solubility calculations look like an exciting source of data

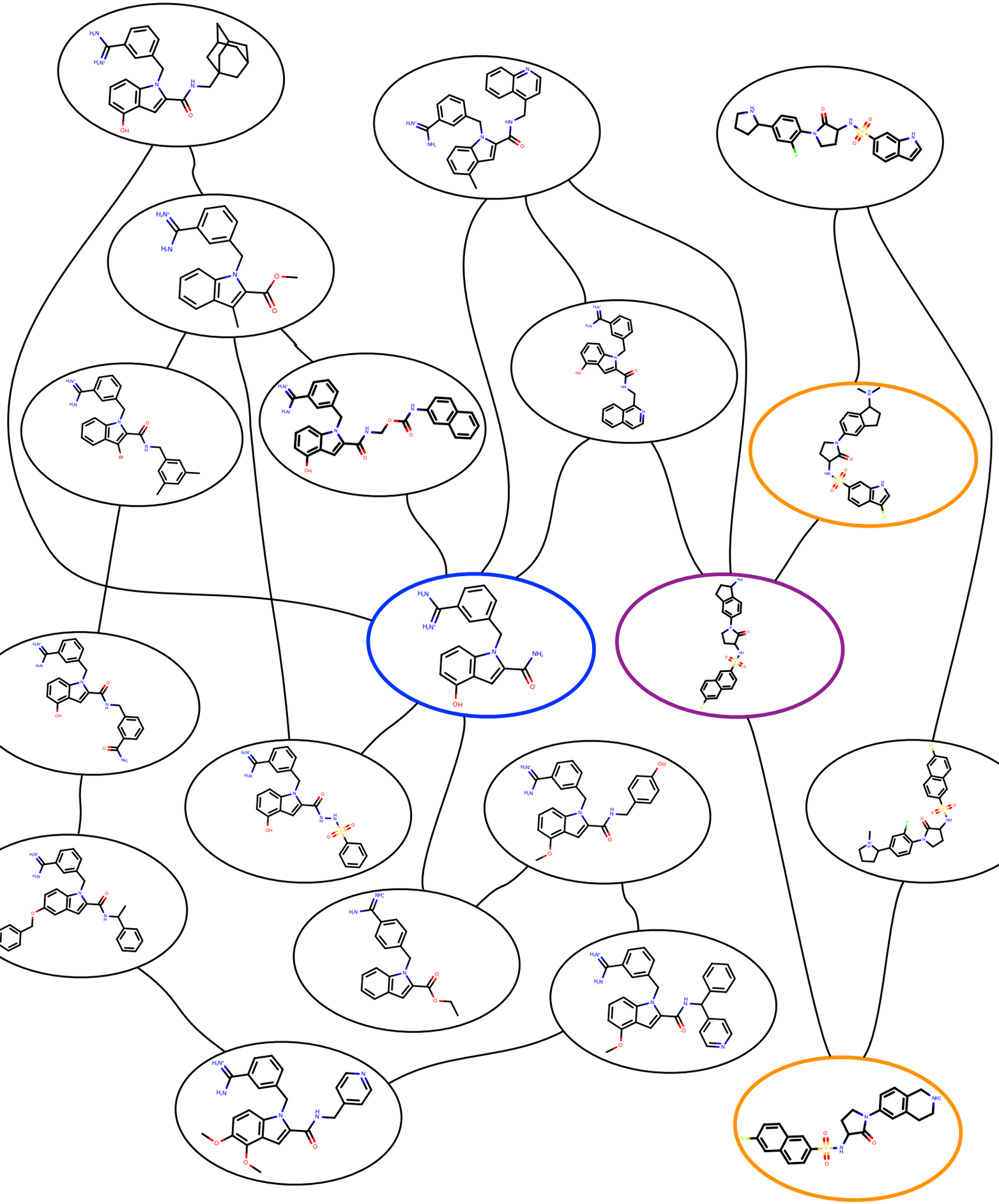
Acknowledgments

- Hydration: J. Peter Guthrie (Univ. Western Ontario), Shuai Liu, Nathan Lim, Karisa Wymer, and many others in my group (UCI); many past and present collaborators including Christopher Bayly and John Chodera and many others
- Hydroxyl: Chris Fennell (Oklahoma State), Karisa Wymer (UCI)
- Relative solubility: Andrew Paluch (Miami, OH), Shannon Cao (UCI), Kevin Hoang (UCI), Shuai Liu (UCI)
- Relative free energy planning: Shuai Liu (UNO/UCI), folks at Schrödinger including Teng Lin, Yujie Wu, Robert Abel. Graph theory (UNO): Chris Summa, J. Redmann
- Binding work: Especially Gabe Rocklin, Ken Dill, Alan Graves, Sarah Boyce, Brian Shoichet, and many other people in experiment and modeling over the years
- Funding: UCI, UNO, NIH, Louisiana Board of Regents, NSF EPS-1003897
- PyMol

We built LOMAP for automated planning of relative free energy calculations

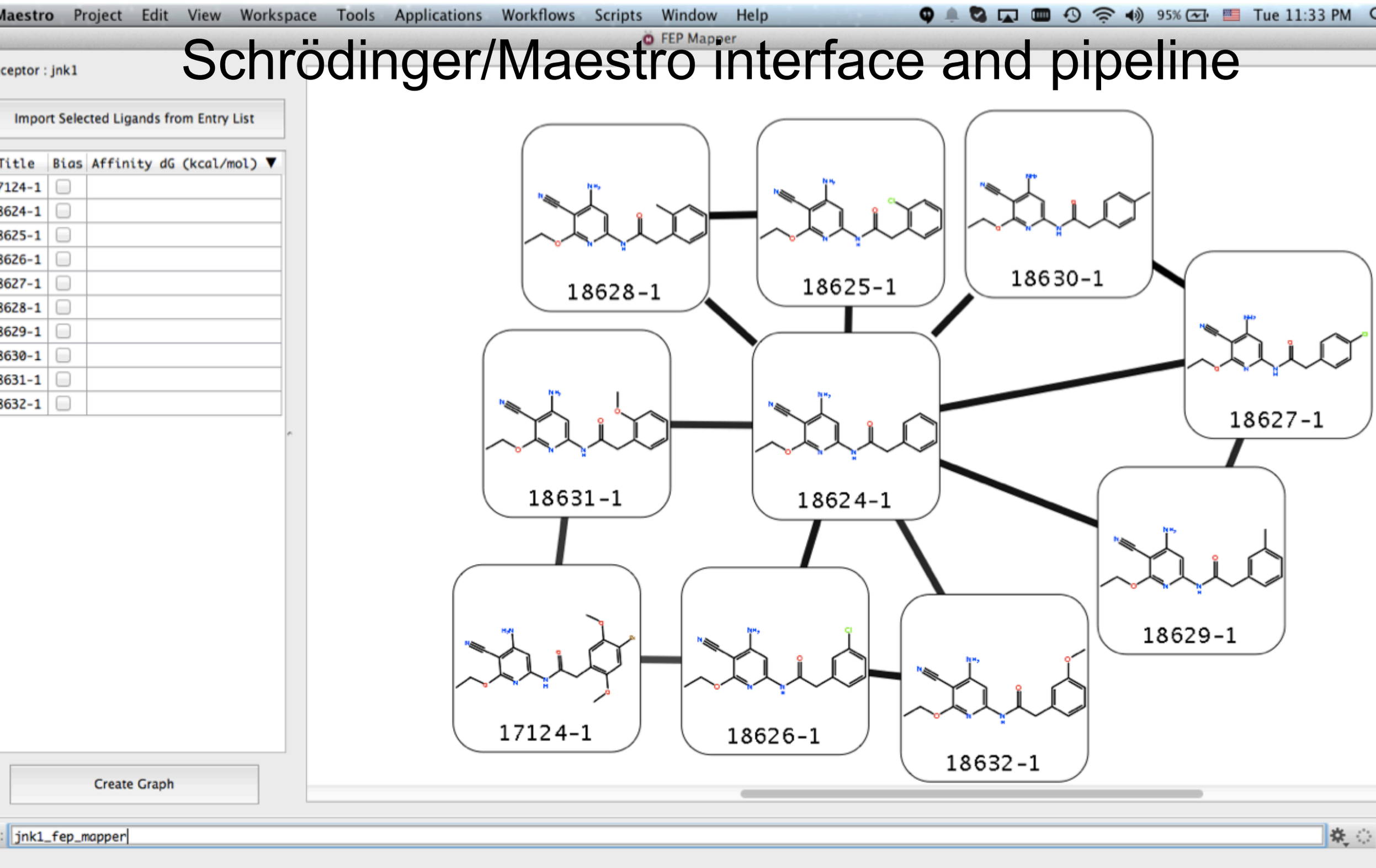




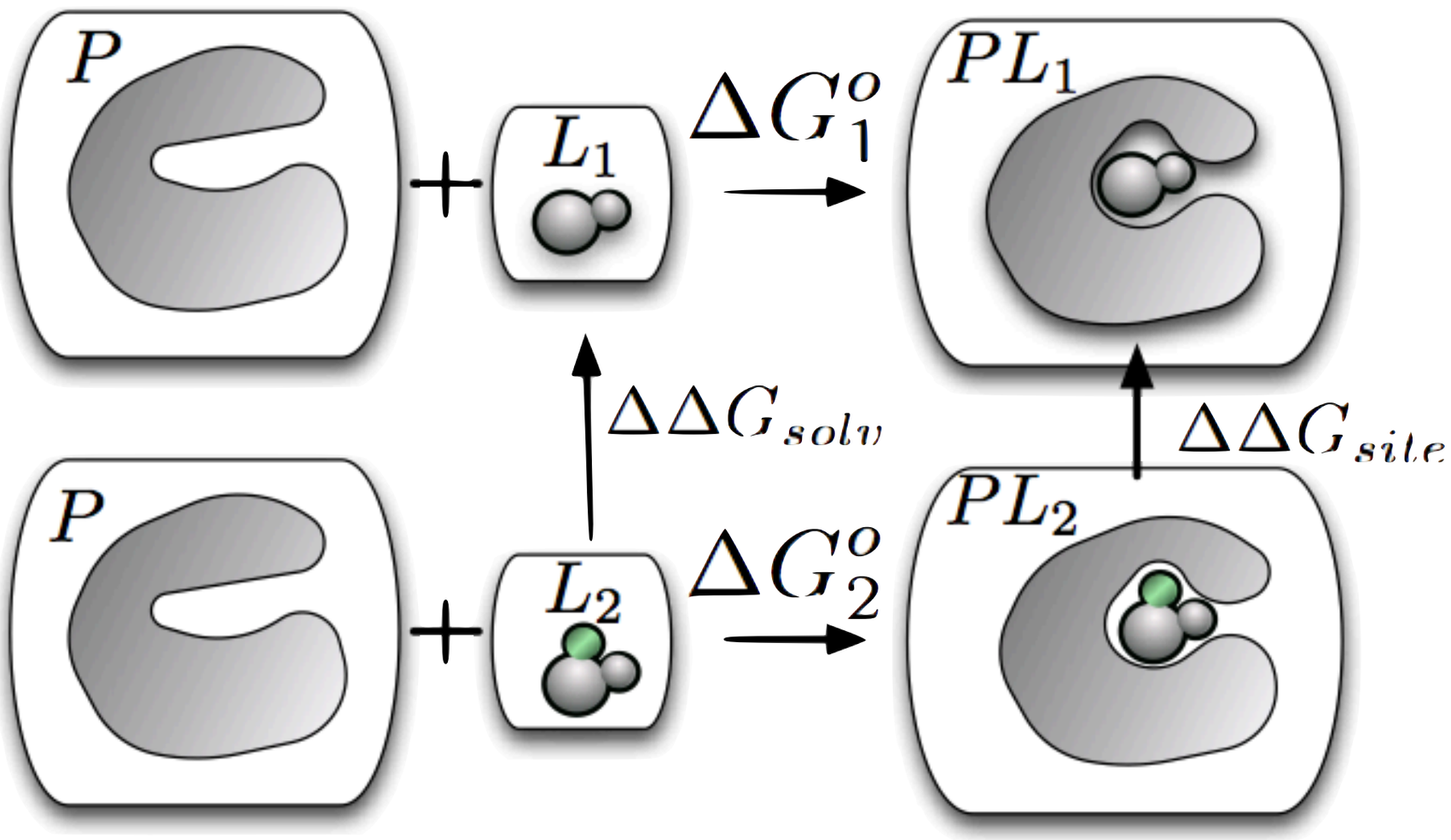


Called Lead Optimization Mapper (LOMAP); available under BSD

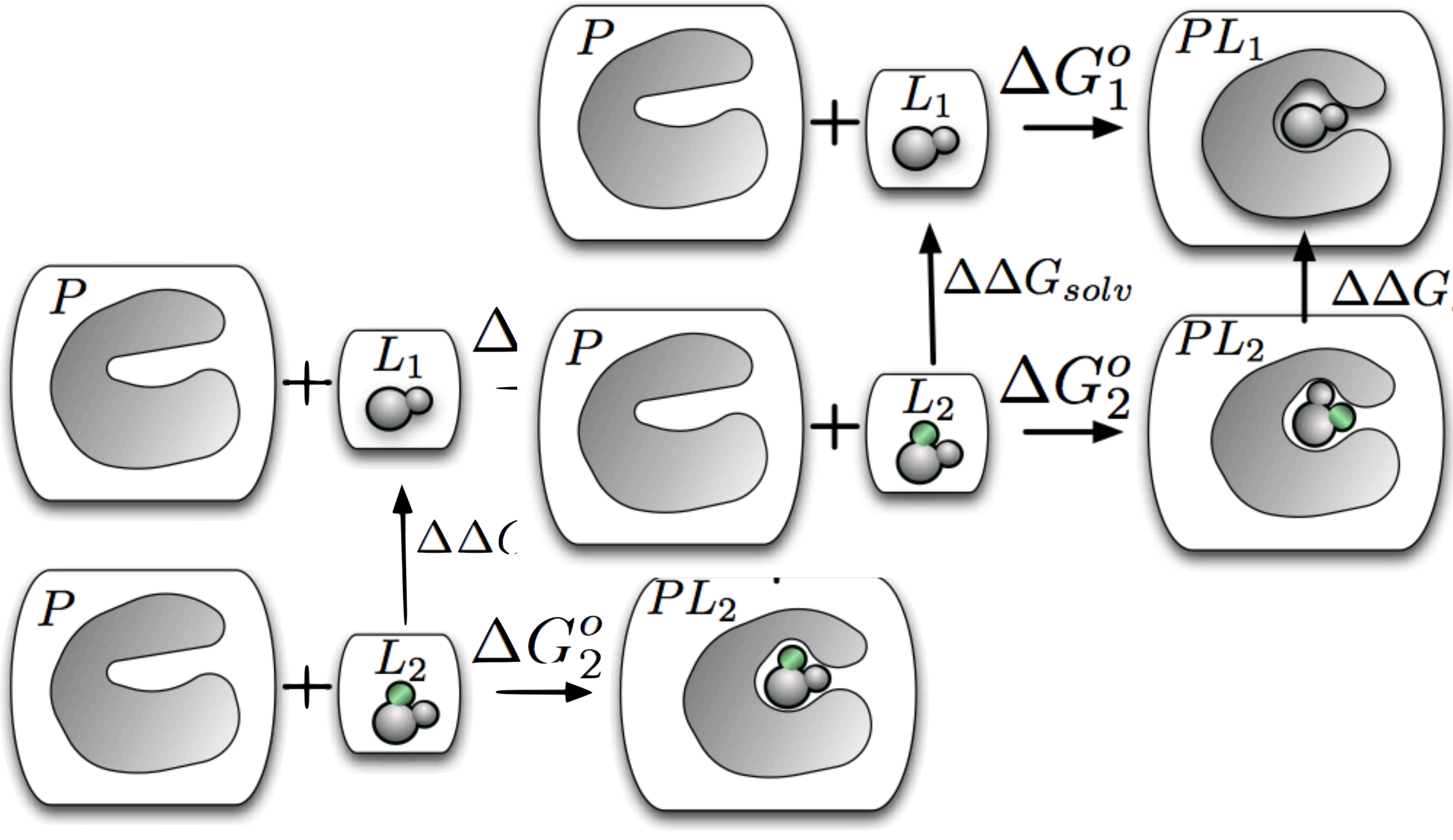
Schrödinger/Maestro interface and pipeline



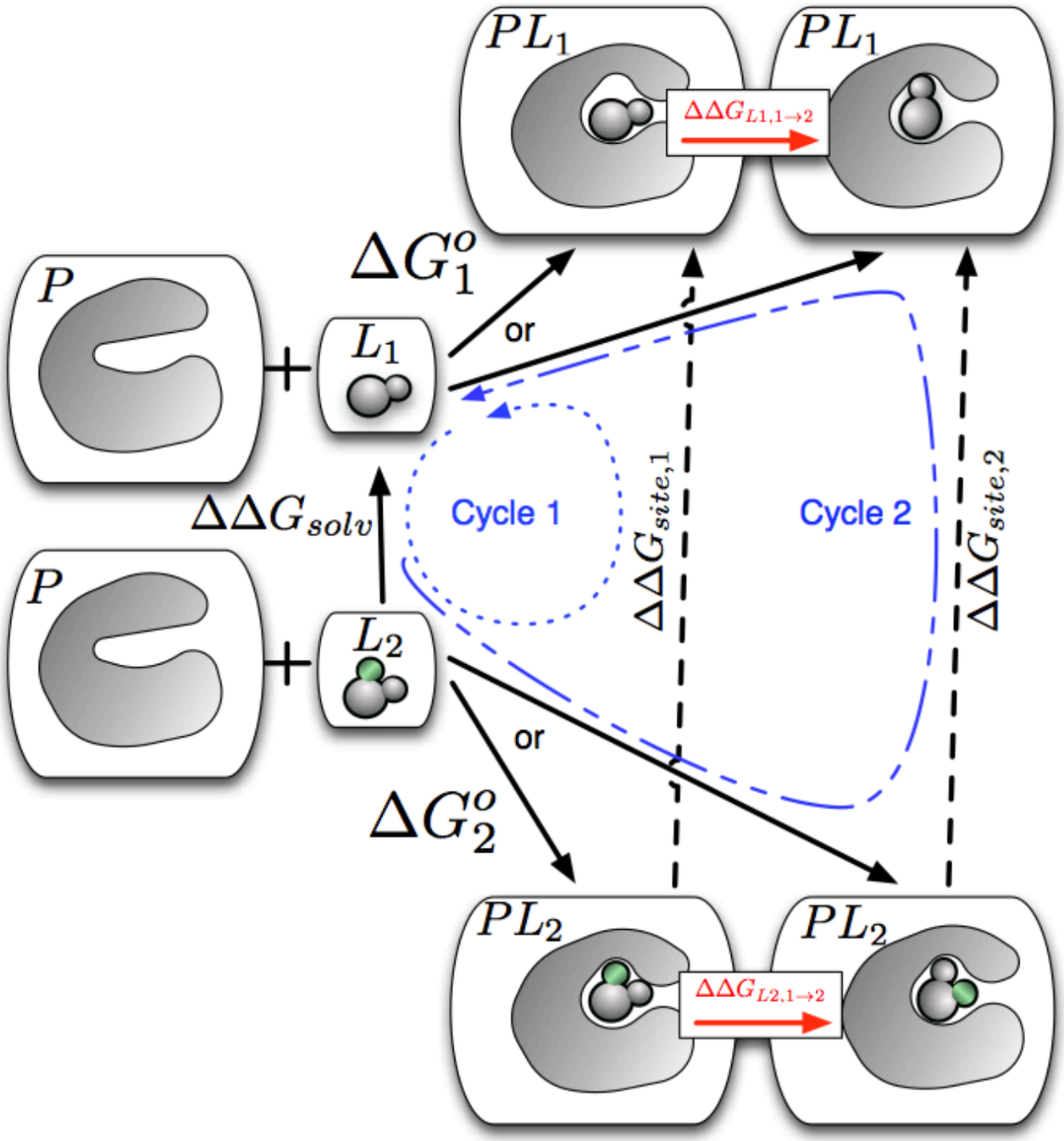
There's just one problem: What if we don't know the binding mode?



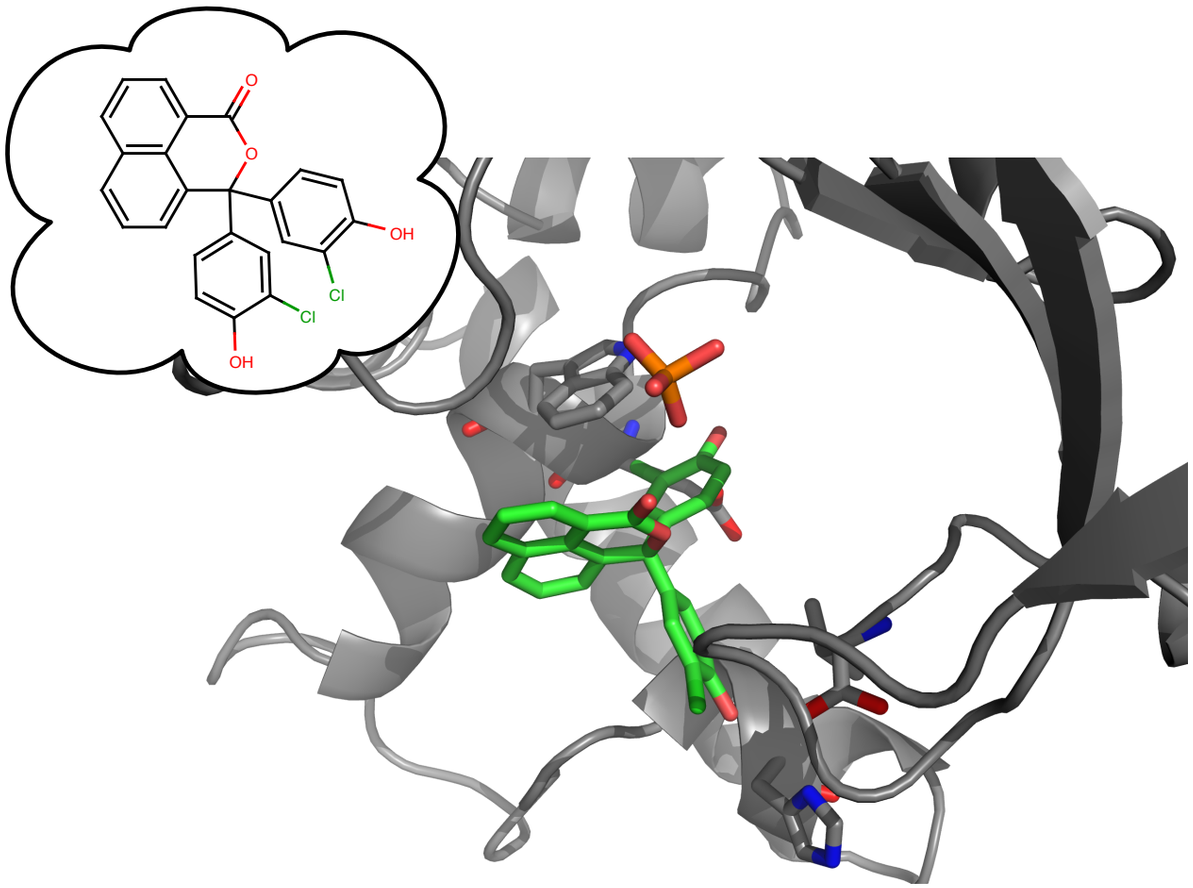
There's just one problem: What if we don't know the binding mode?

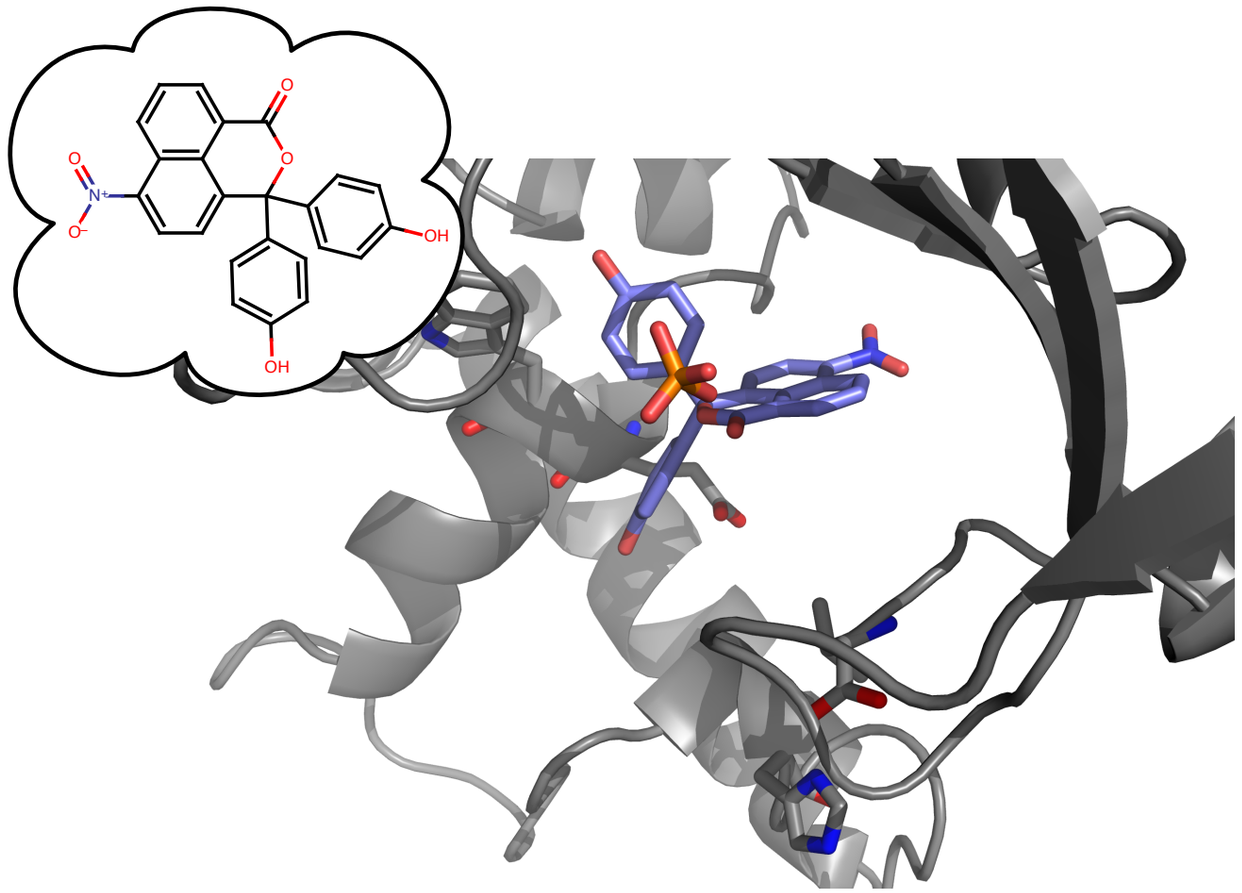


Additional information is needed: the relative free energies or populations of the different potential binding modes



Does this ever happen in real life? Yes! Examples are easy to come by, but we don't really know how often

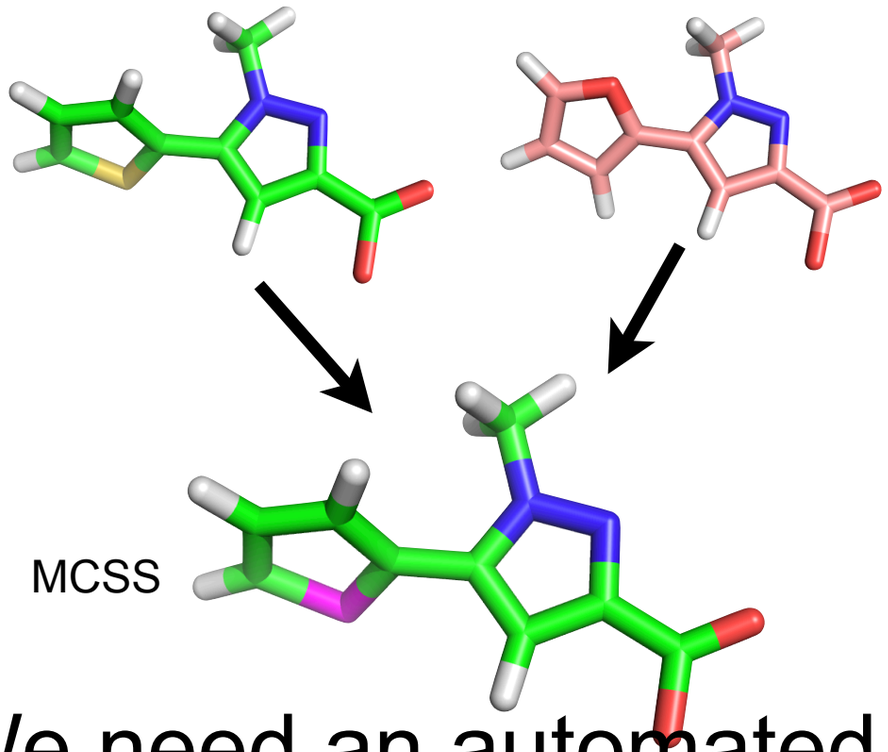




Stout et al., Biochem. 38:1607 (1999)

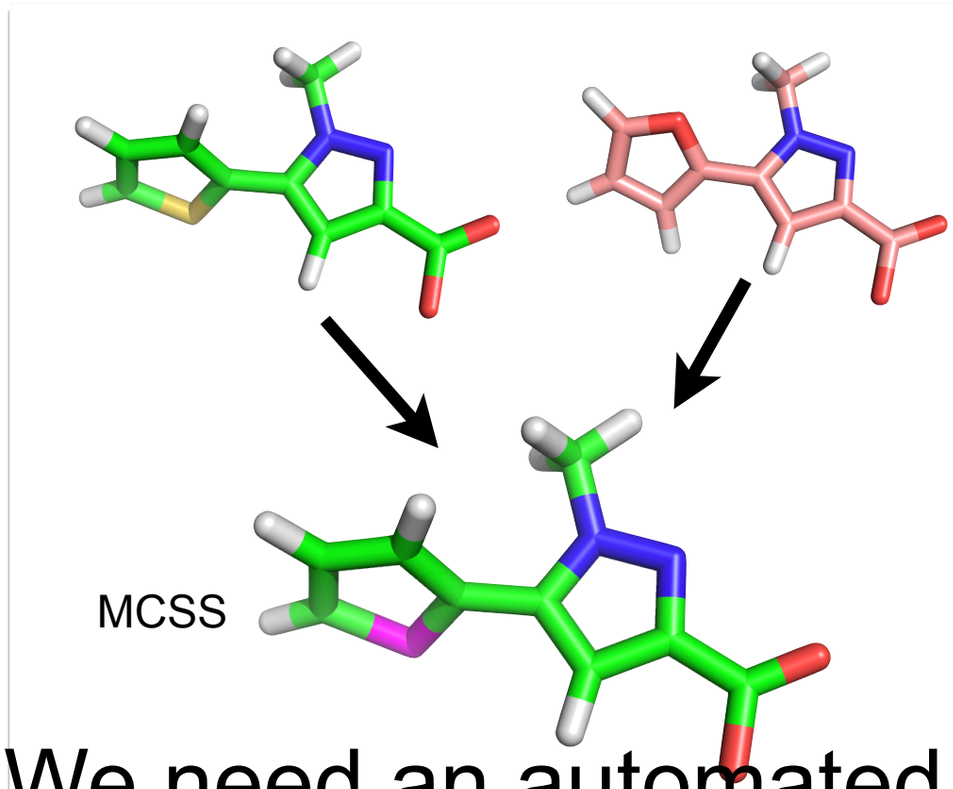
To sum up...

To sum up...

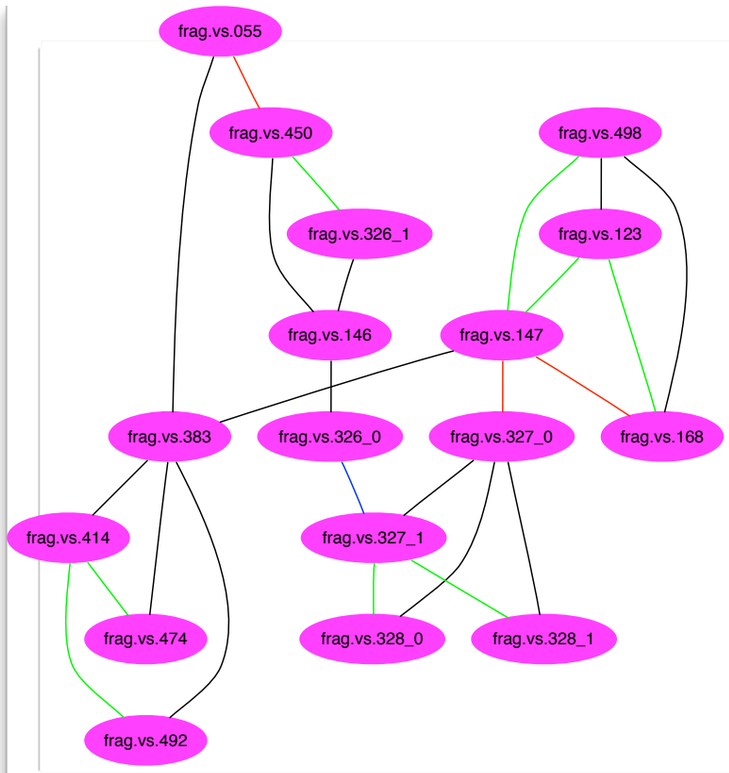


We need an automated way to plan which calculations to do

To sum up...

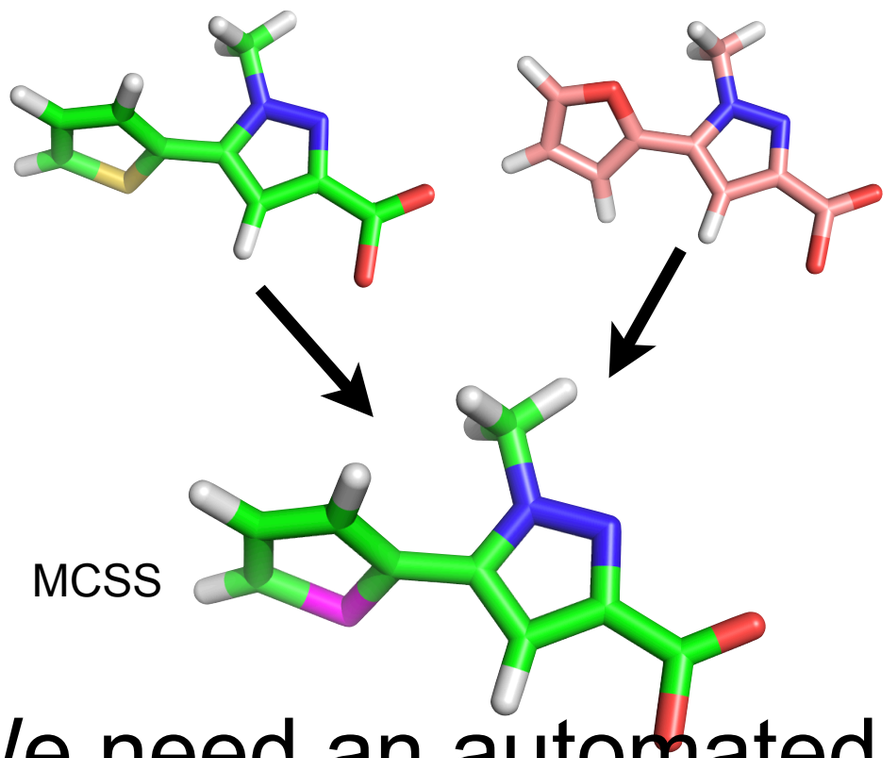


We need an automated way to plan which calculations to do

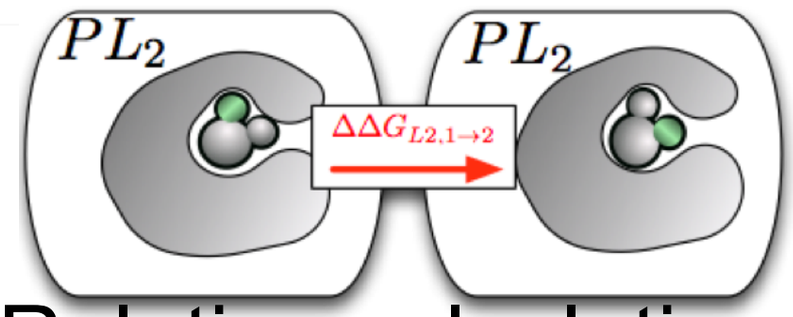


We have a new tool for automated planning of calculations

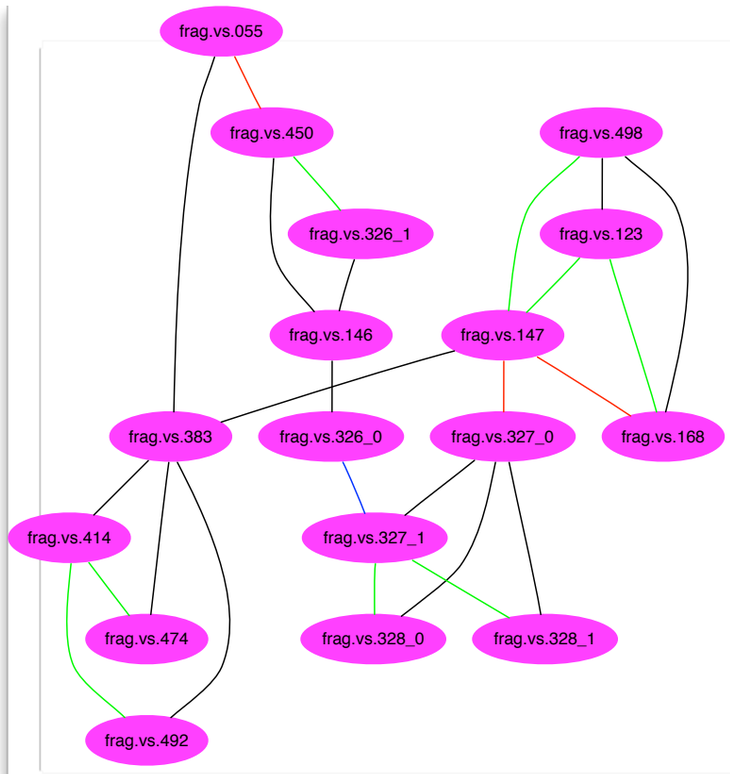
To sum up...



We need an automated way to plan which calculations to do

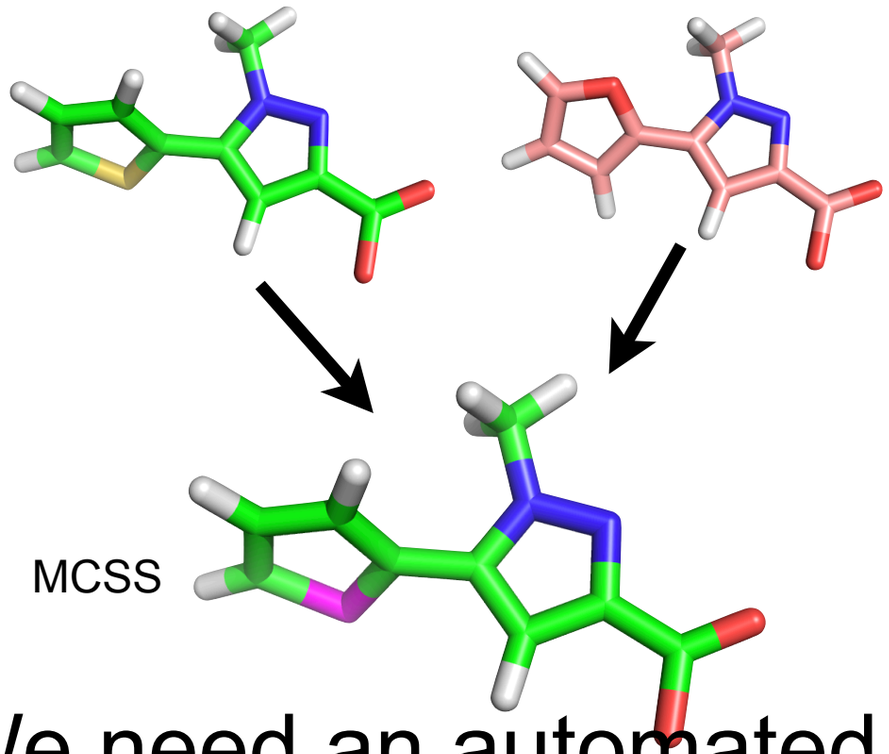


Relative calculations may need careful orientational sampling

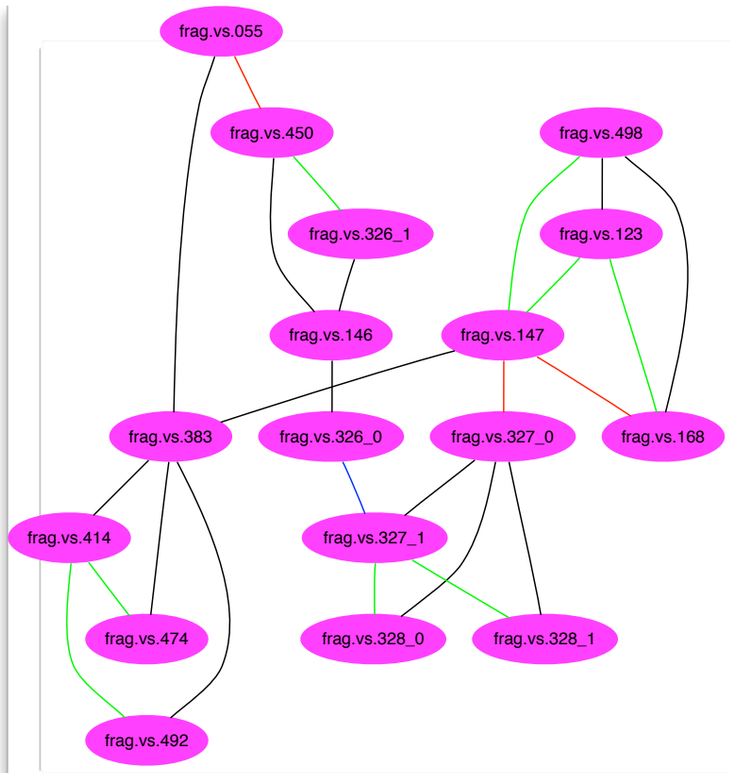


We have a new tool for automated planning of calculations

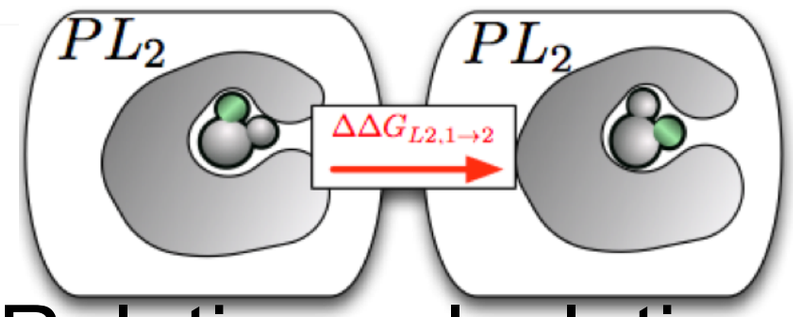
To sum up...



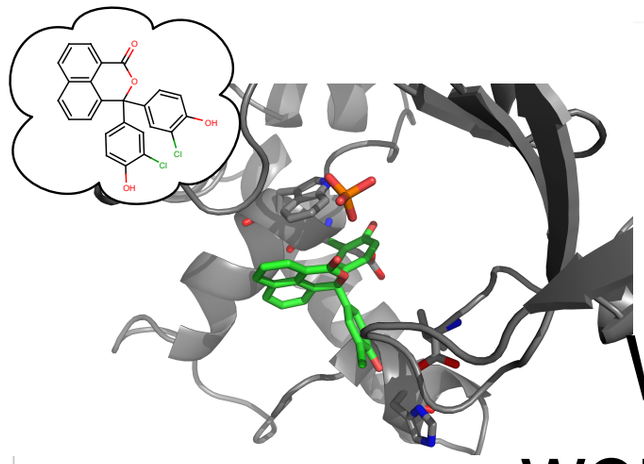
We need an automated way to plan which calculations to do



We have a new tool for automated planning of calculations

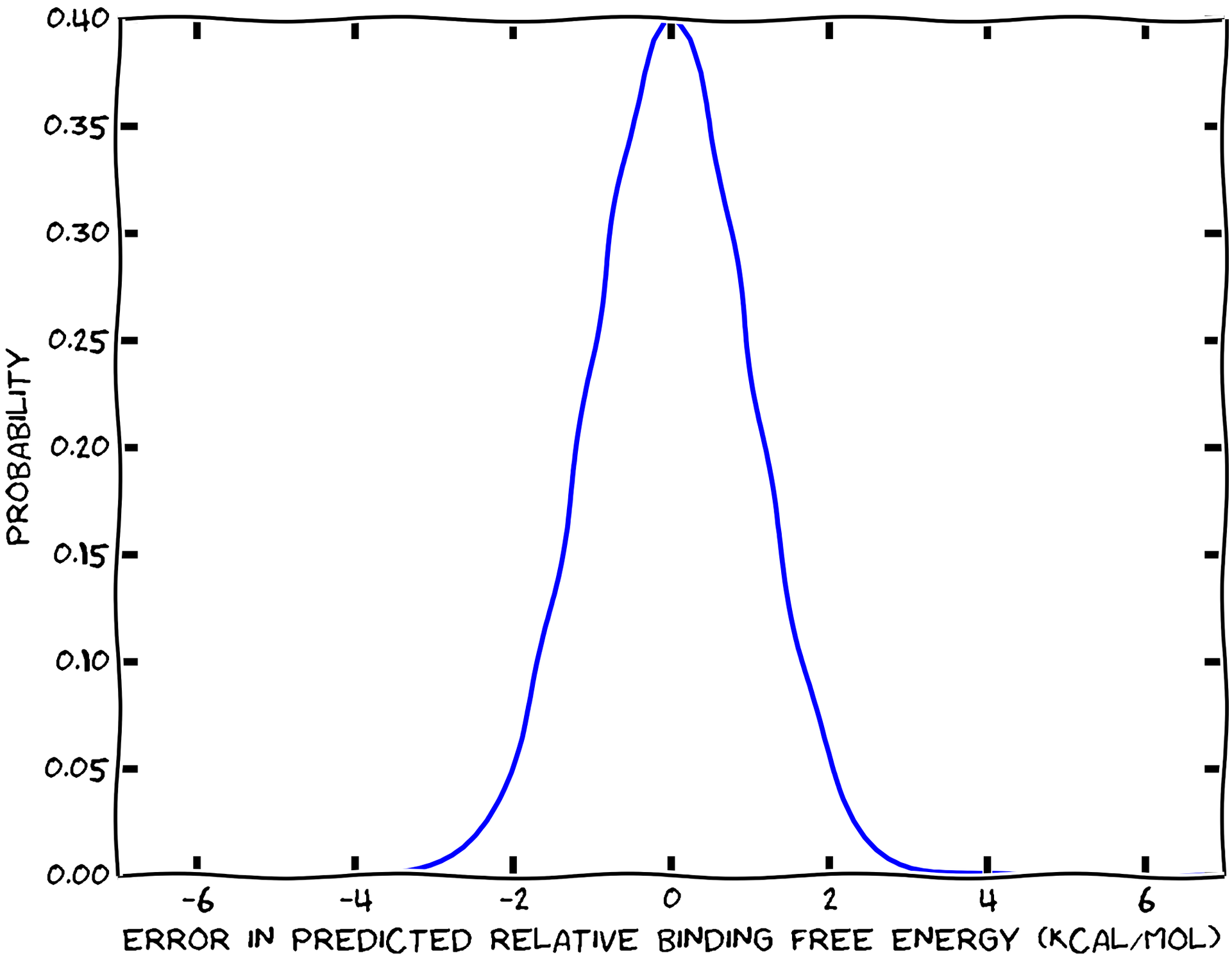


Relative calculations may need careful orientational sampling

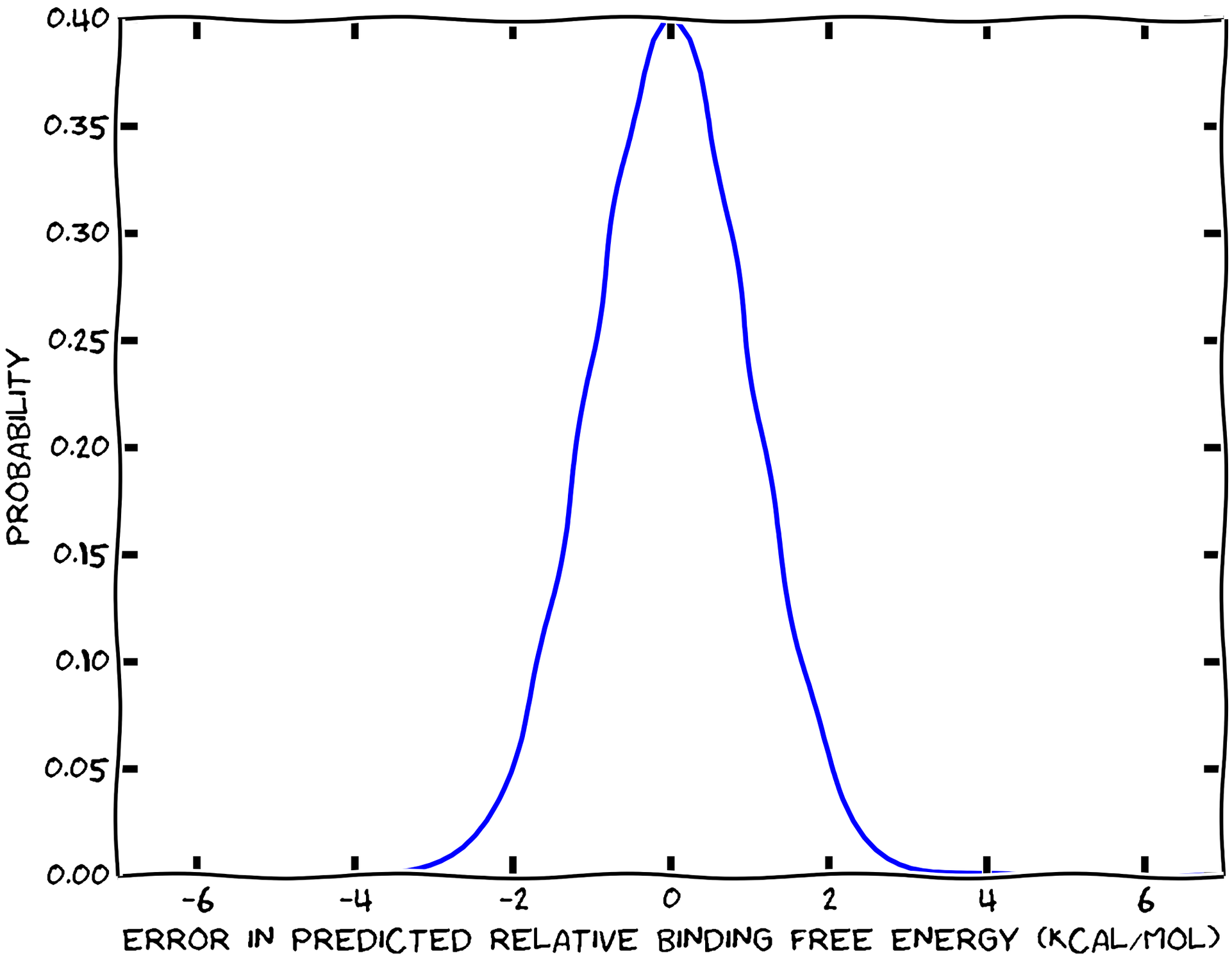


We do have to worry about binding mode changes

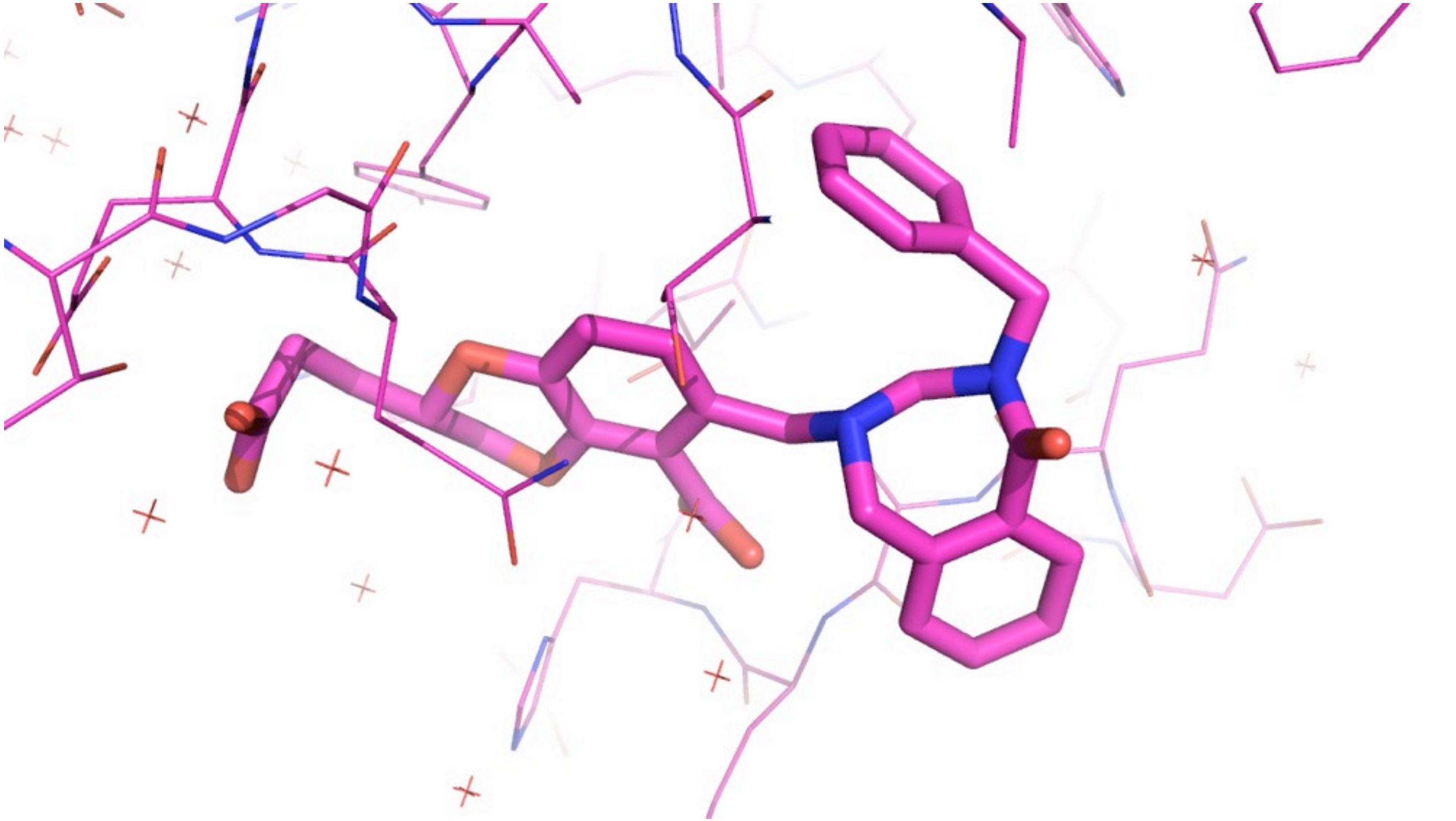
What error distribution should we see from a predictive method?



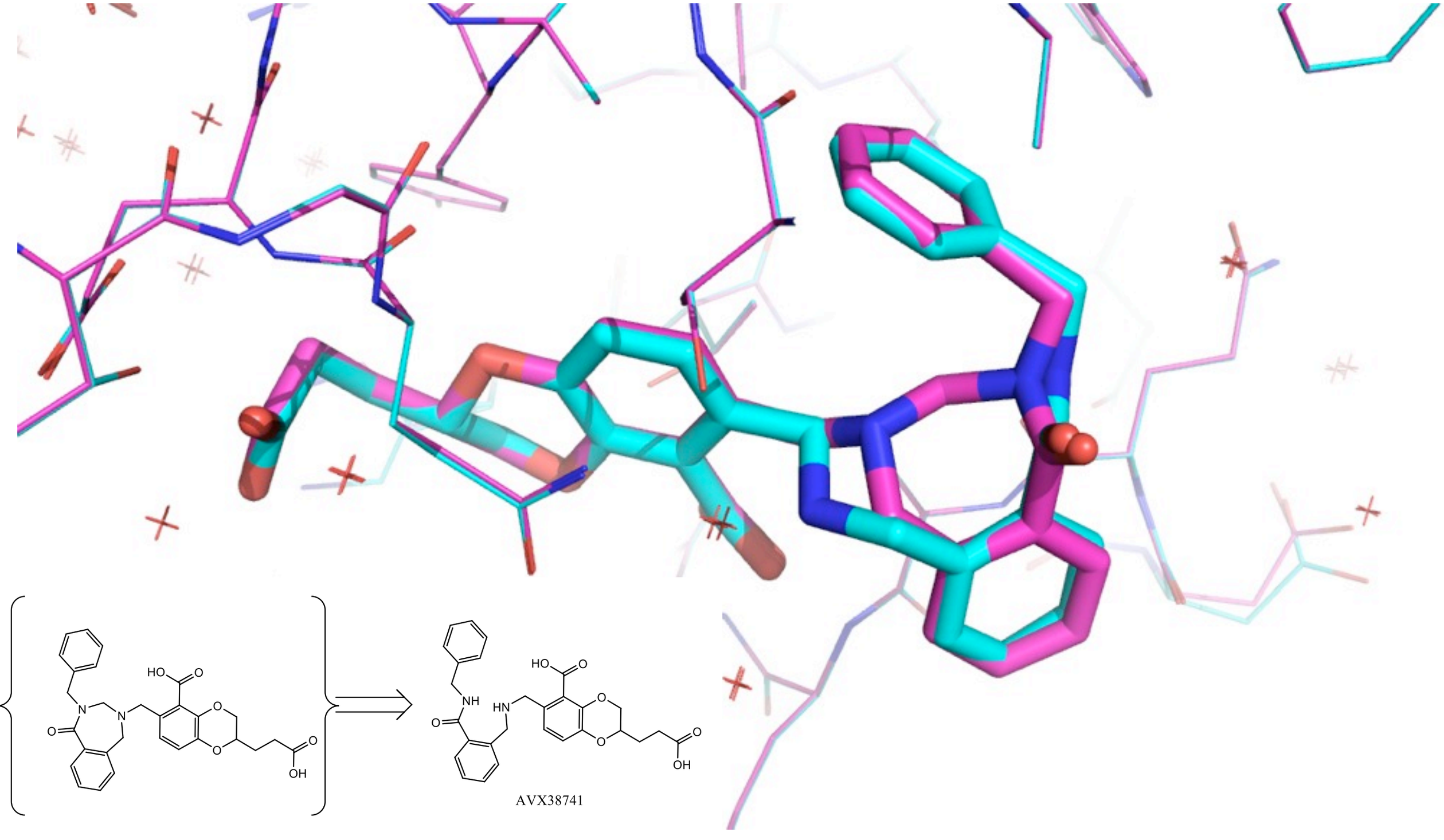
What error distribution should we see from a predictive method? Not Gaussian...



Why not?
We may not have the right system

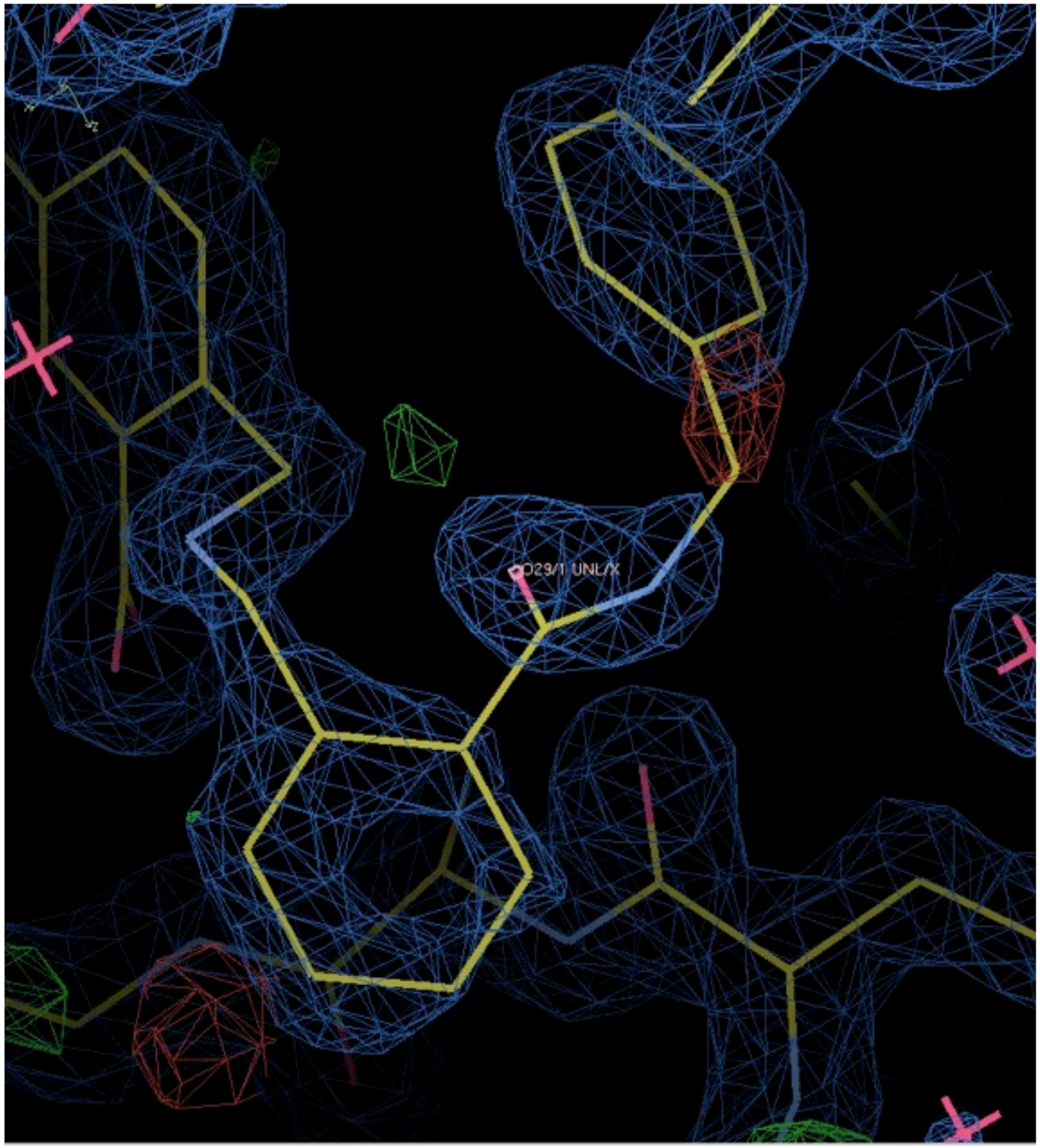


Why not?
We may not have the right system

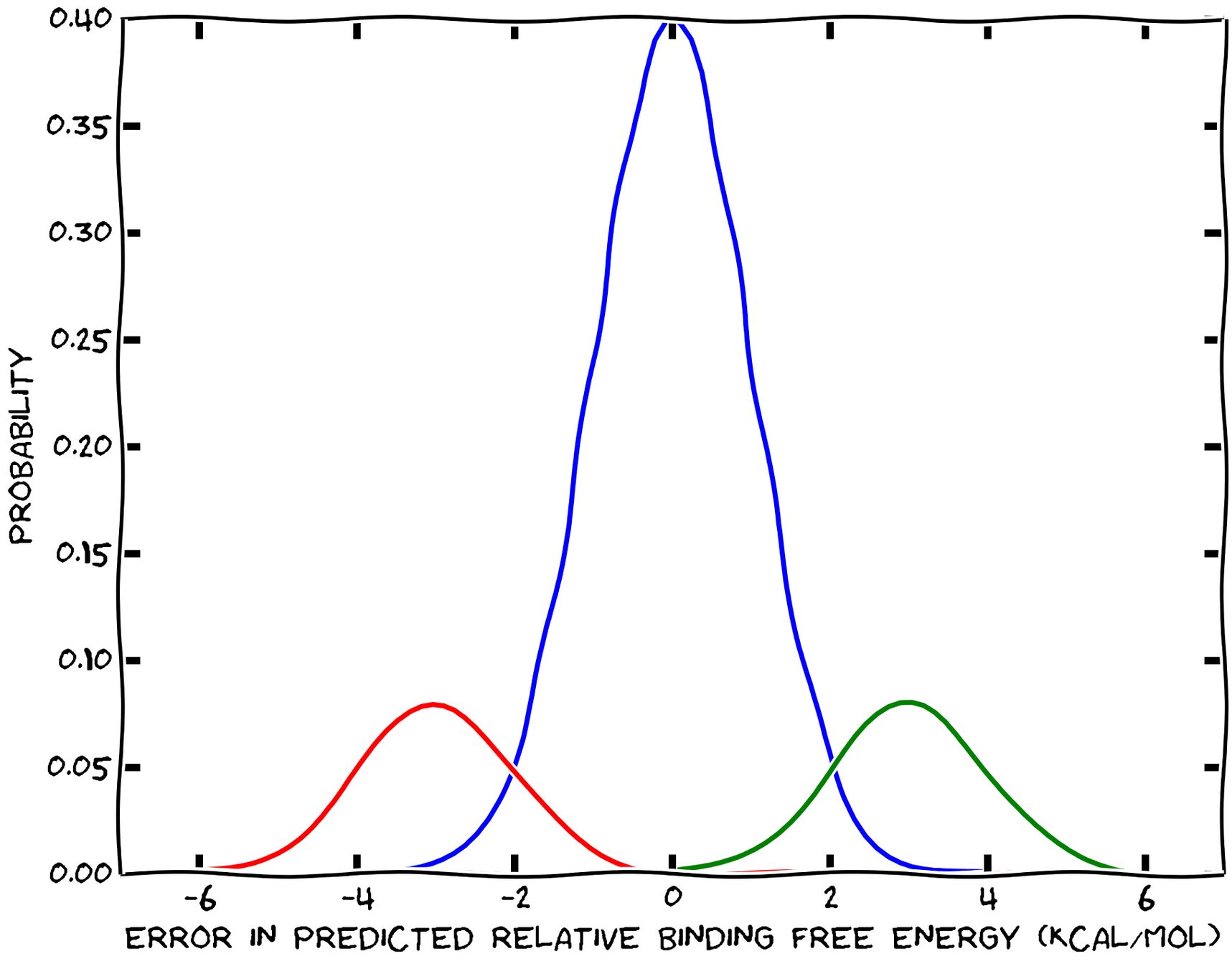


AVX38741

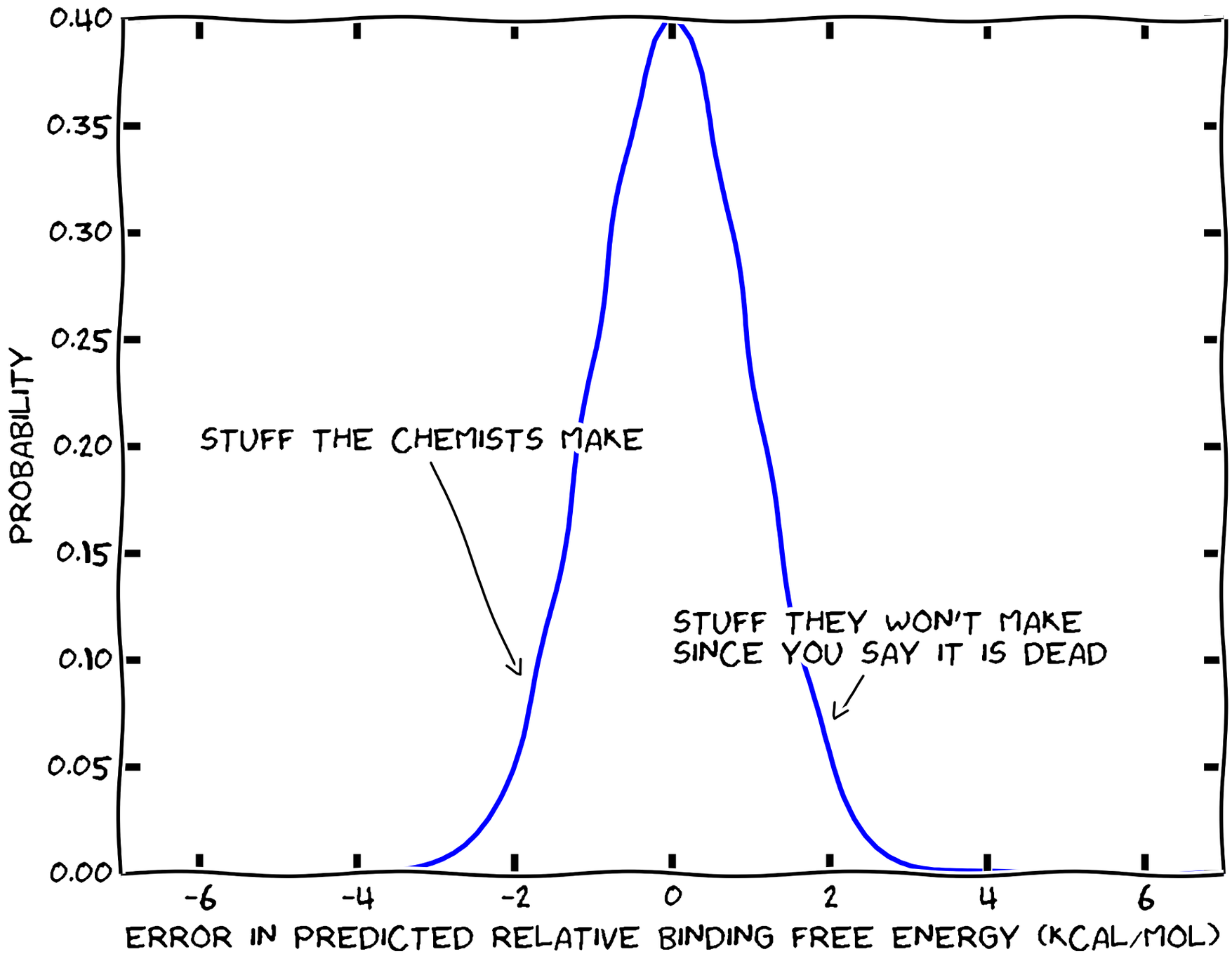
Why not?
We may not have the right system



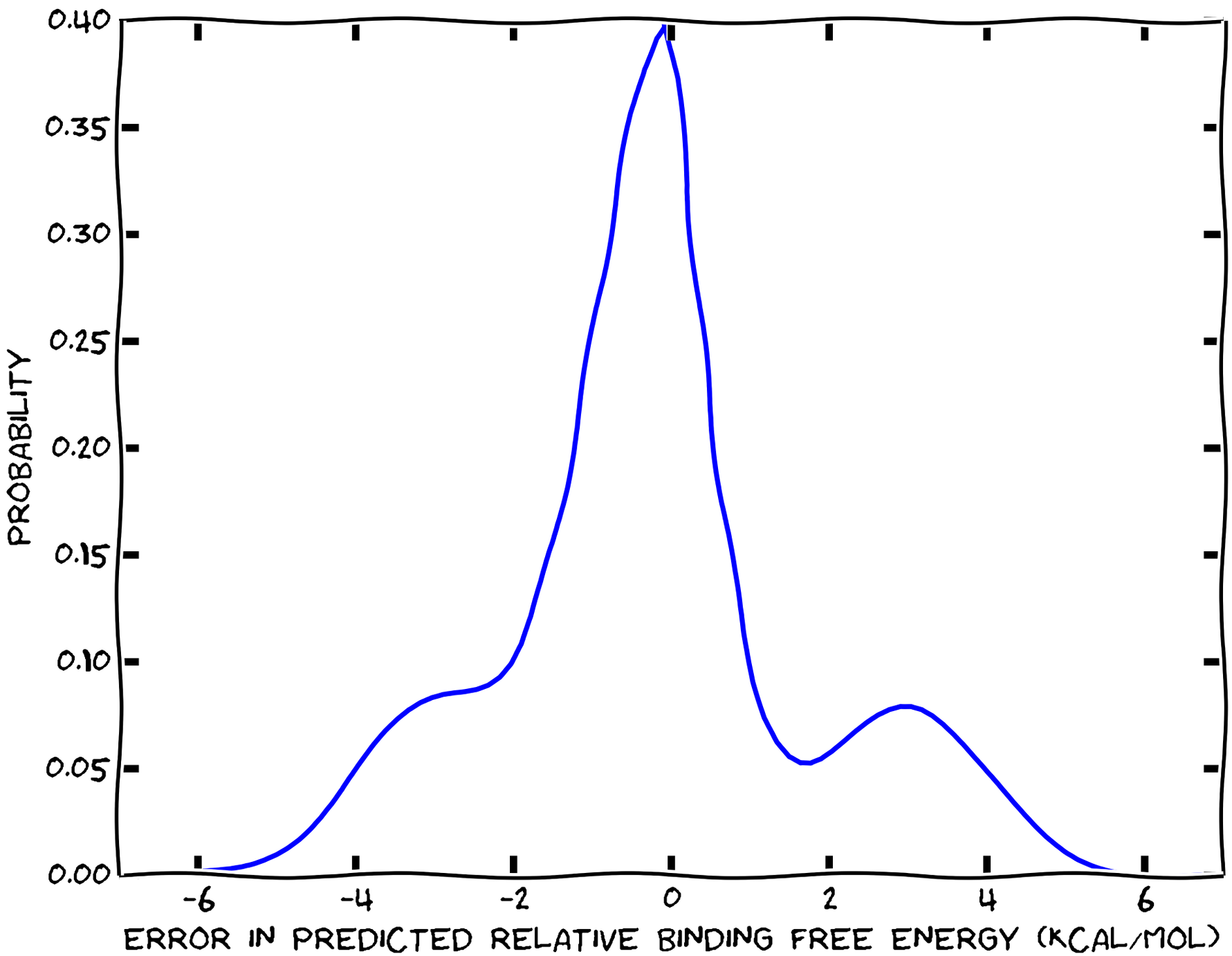
So, what should we have then?
Something non-Gaussian at least



There is another complexity -- which compounds will they make when you make predictions?



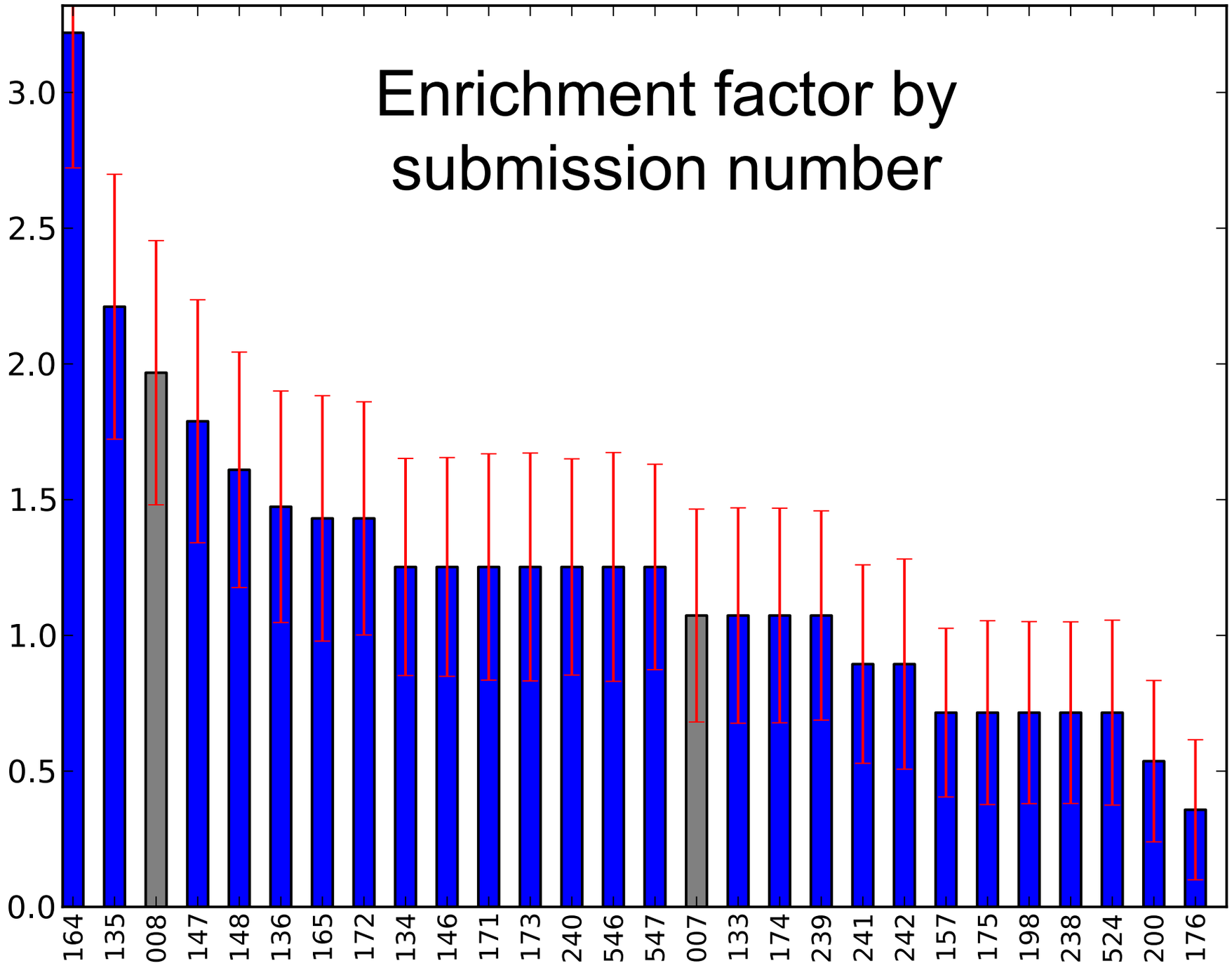
So, in the end, our reasonable method looks pretty dismal



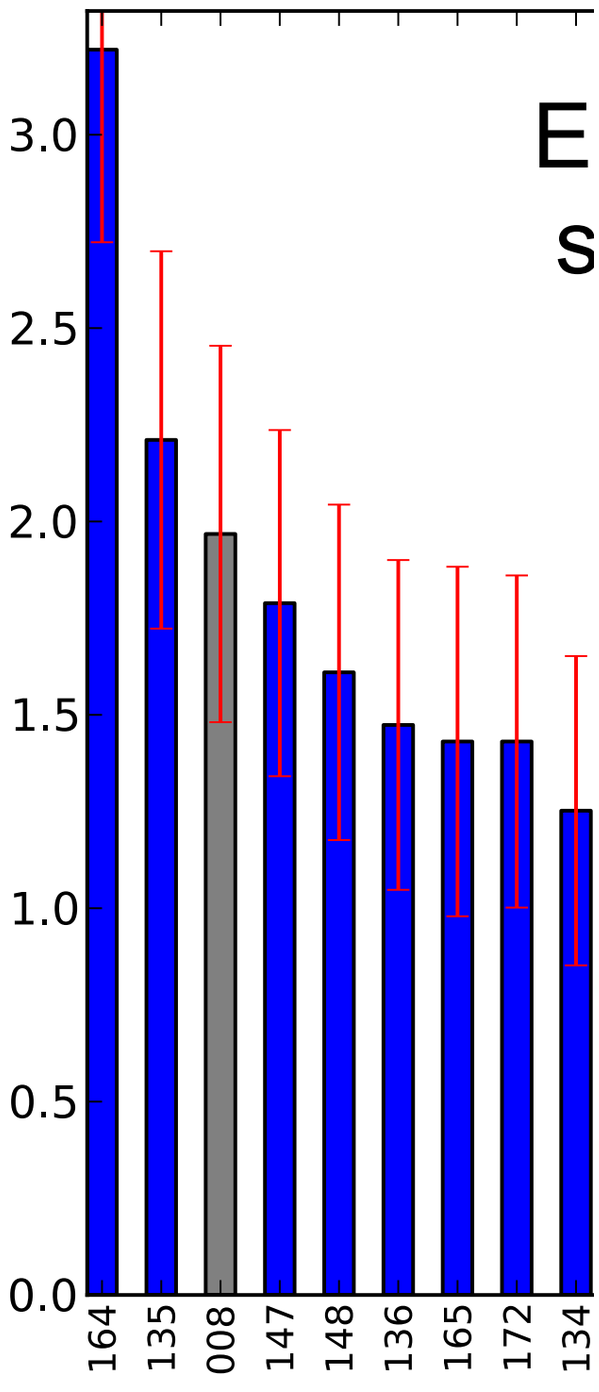
Major take-away: “Application is not validation”

Apparent performance in application may be substantially different from true average performance due to selection bias/issues

The SAMPL4 challenge included an HIV-1 integrase virtual screening challenge



The SAMPL4 challenge included an HIV-1 integrase virtual screening challenge



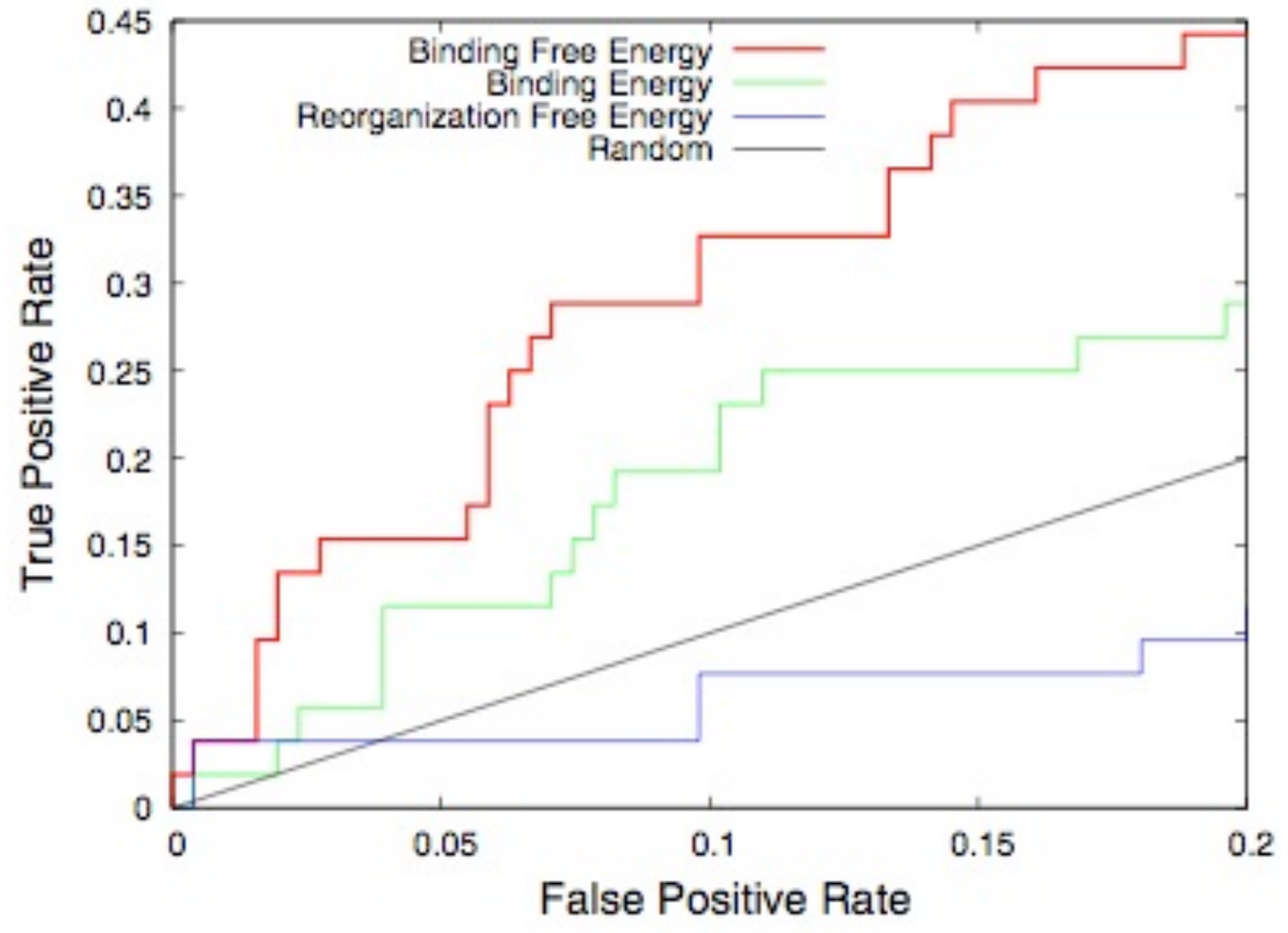


Fig. 2 Early ROC curve for the classification of binders and non-binders to the LEDGF site of integrase. The ROC curve based on binding free energy rankings is in *red* (*top*), the ones based on binding energy rankings and reorganization free energy rankings are in *green* (*middle*) and *blue* (*bottom*) respectively. The *straight black line* is the 1:1 line corresponding to random picking. Results for sorting reorganization free energies from low to high are shown here, the reverse rankings yielded poorer performance

What must be done for these to become a routine part of lead optimization?

- Can probably be used now/soon when system is “well behaved” (but how do you know?)
- But in general:
  - Better sidechain sampling
  - Handling of missing residues/loops
  - Ligand binding mode sampling
  - Slow protein conformational changes
- Failure prediction, especially:
  - Sampling failure
  - Force field failure
- Need to validate, not just apply...
  - Need follow-up calcs AND experiments when calculation, experiment disagree