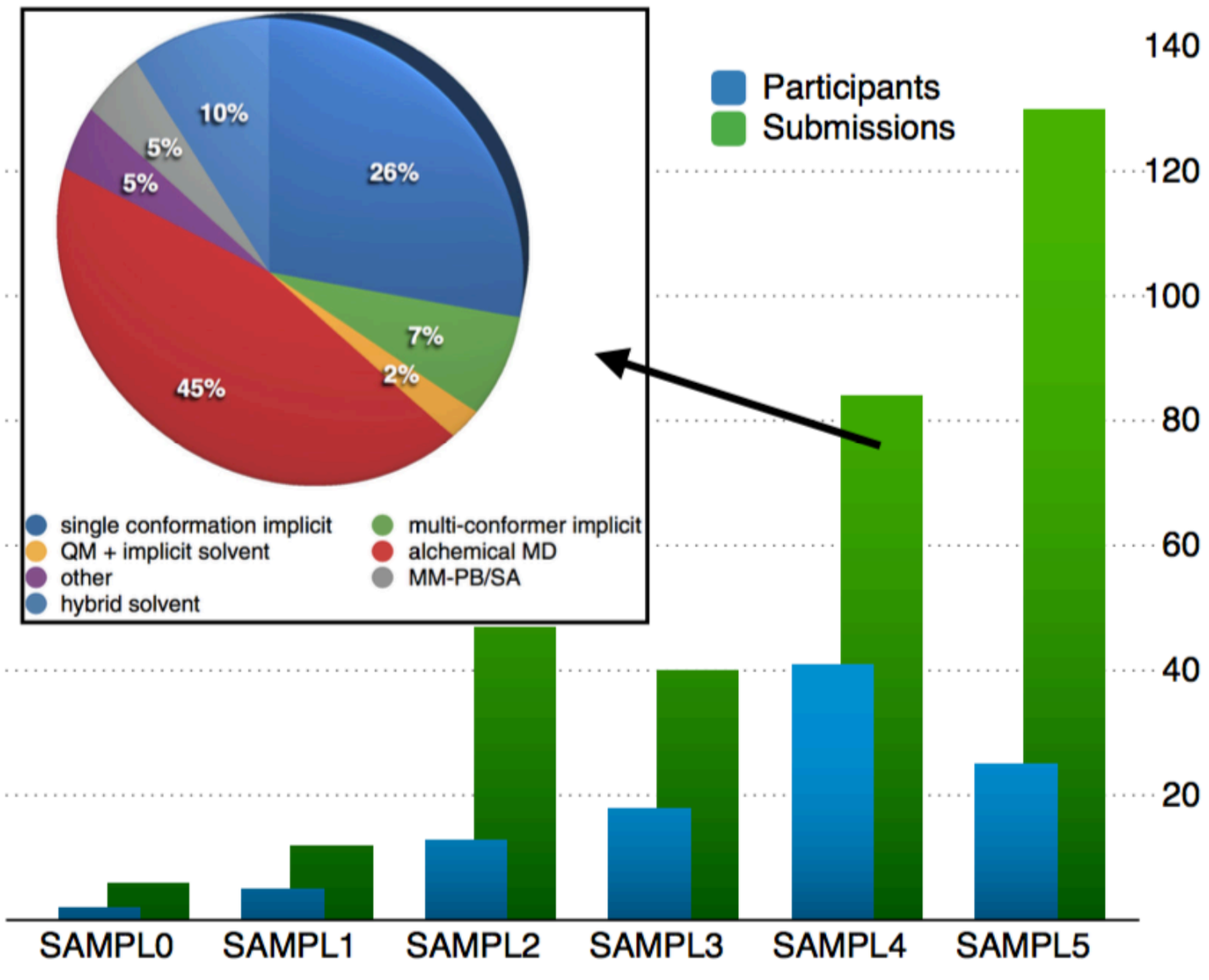
SAMPL: Past, present and future

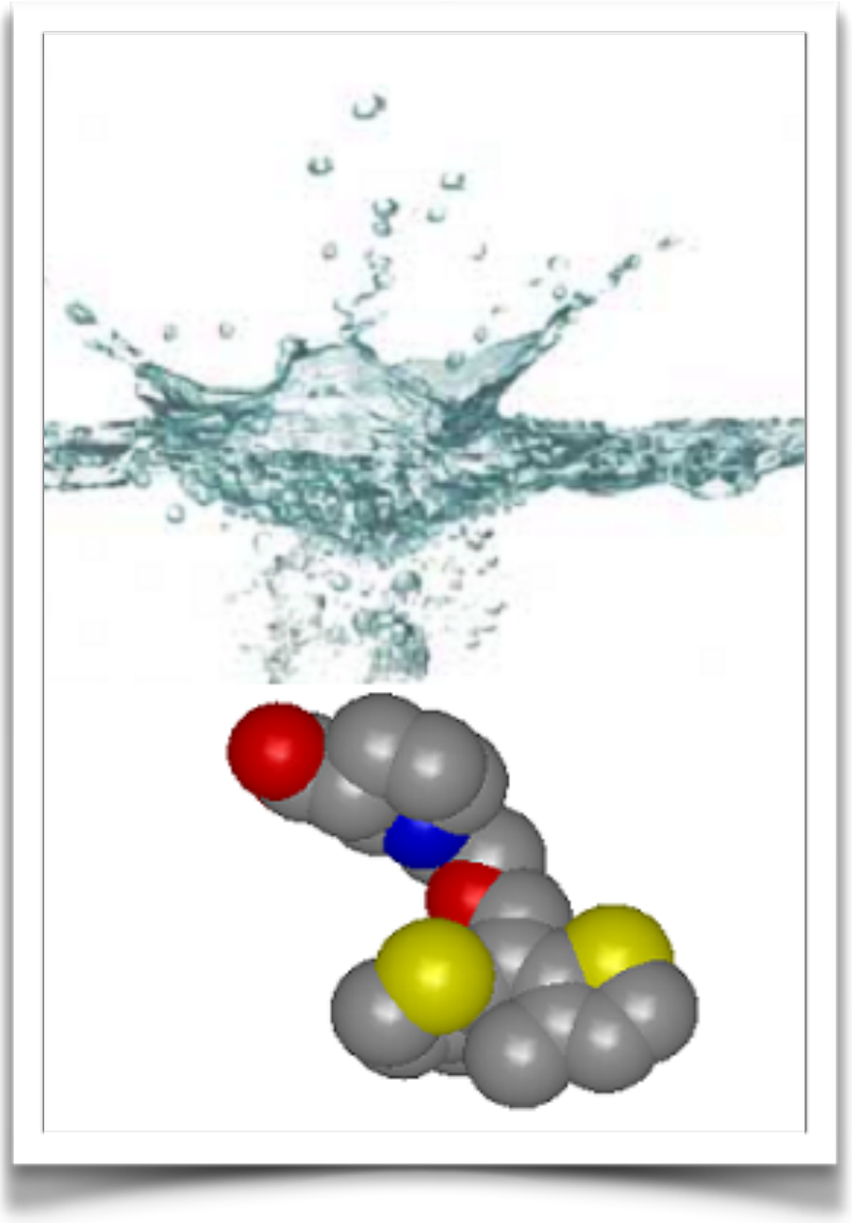
David Mobley

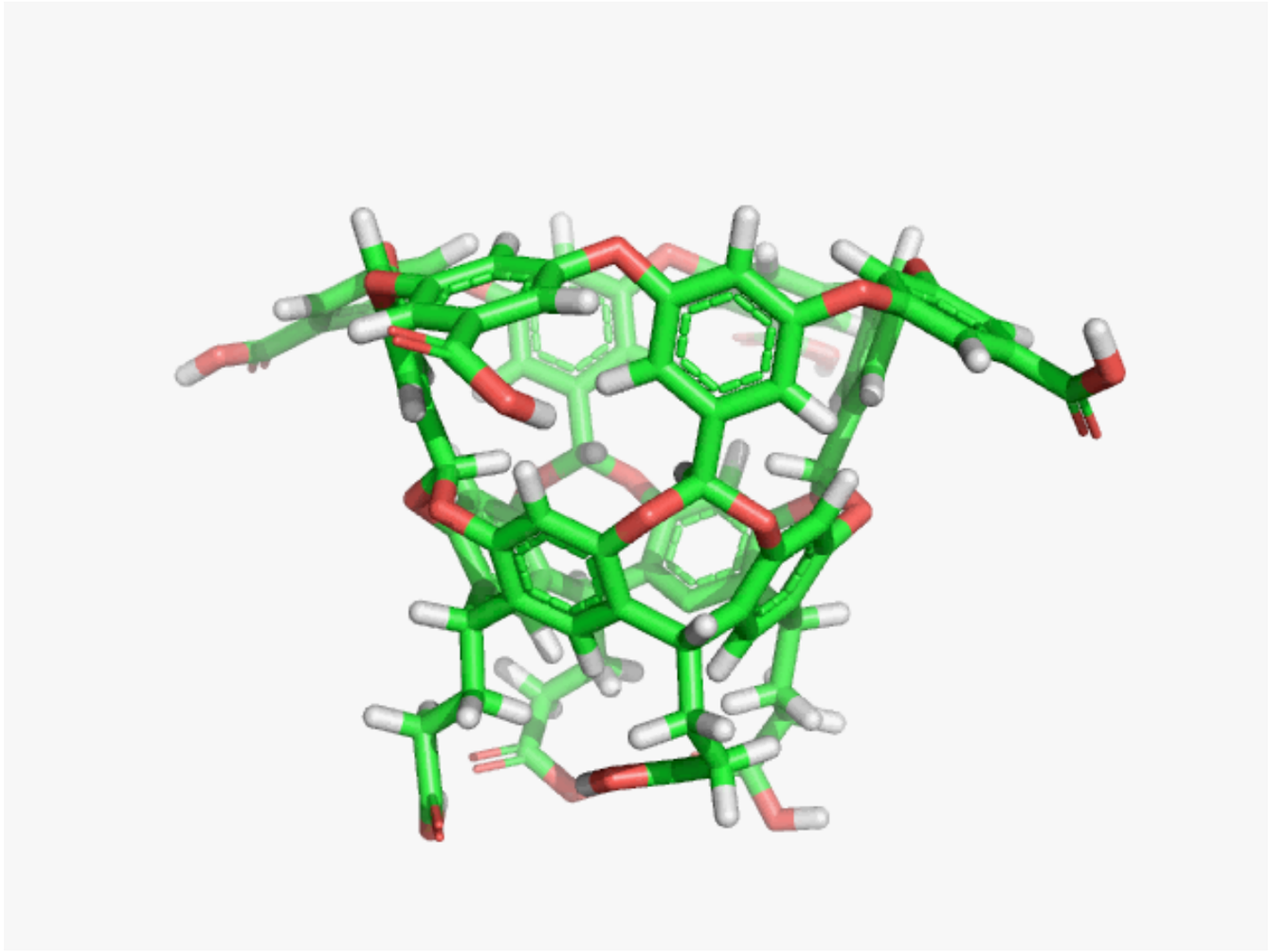
SAMPL was motivated by the need to compare predictive accuracy on a level playing field



- Originated by Nicholls at OpenEye ’07-’08
- Unfunded academic collaboration since SAMPL3/2012
- More than 100 publications to date

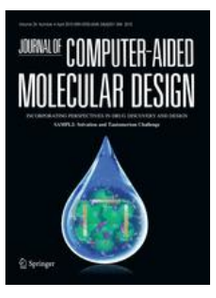
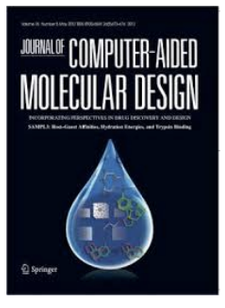
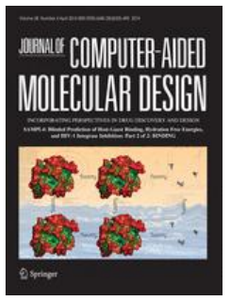
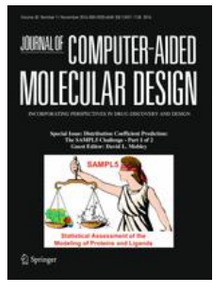
The two most frequent kinds of challenge involve solvation and (simple) binding





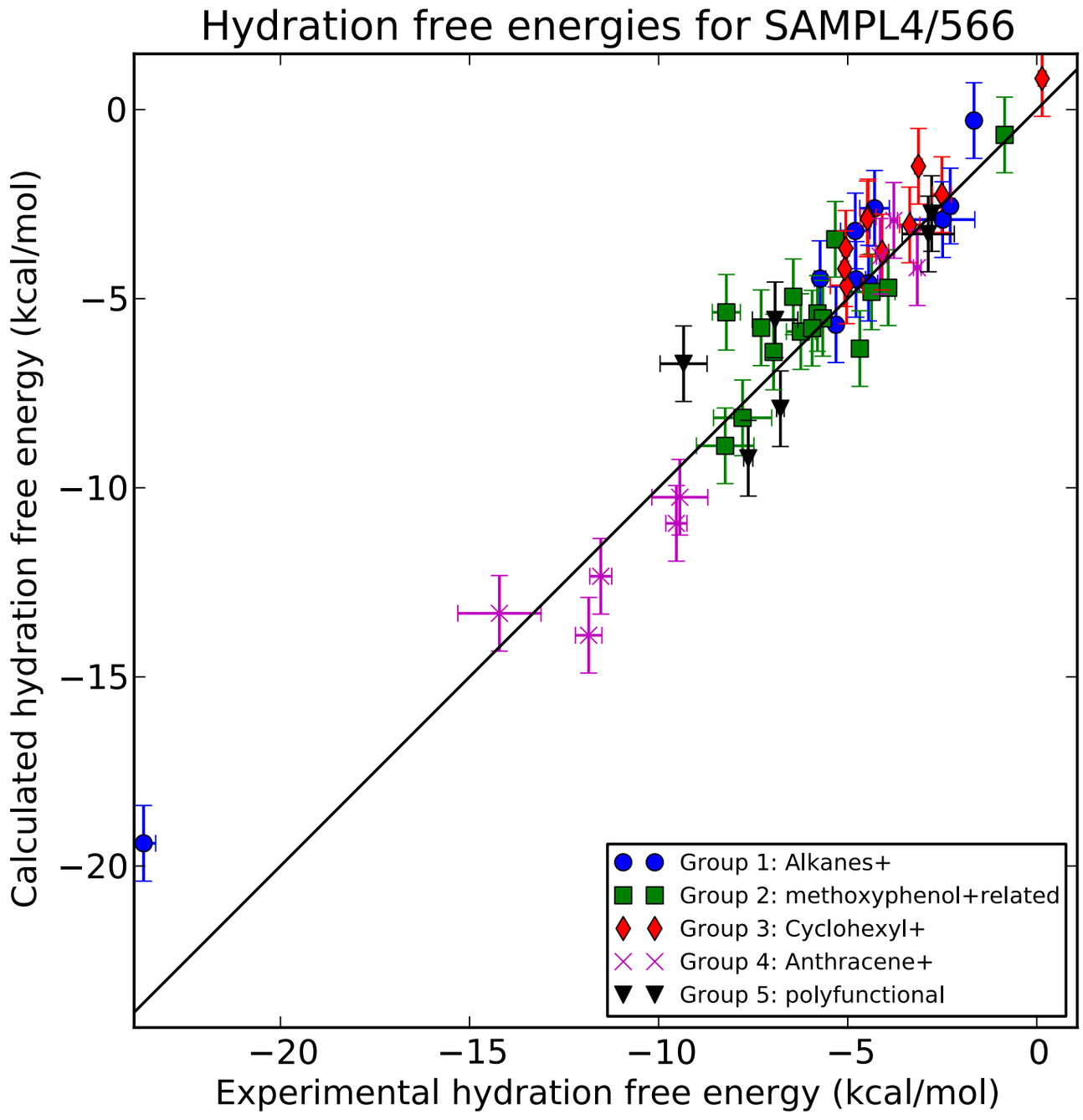
SAMPL focuses on challenges which are at the borderline of tractability

SAMPL0	SAMPL1	SAMPL2	SAMPL3	SAMPL4	SAMPL5
2007	2008	2009	2011	2013	2016
JNK3 kinase inhibitors	CDK2 kinase inhibitors	hydration free energies	trypsin inhibitors	HIV-1 integrase inhibitors	distribution coefficients
hydration free energies	hydration free energies	tautomer ratios	hydration free energies	hydration free energies	CBClip host-guest
				octa acid host-guest	CB7 host-guest
				CB7 host-guest	


- Hydration free energies thru SAMPL4
- Protein-ligand binding thru SAMPL4 (then D3R)
- Host-guest SAMPL3 on
- Usually relies on donated data and/or industry internships

Hydration free energy predictions have improved greatly over the years



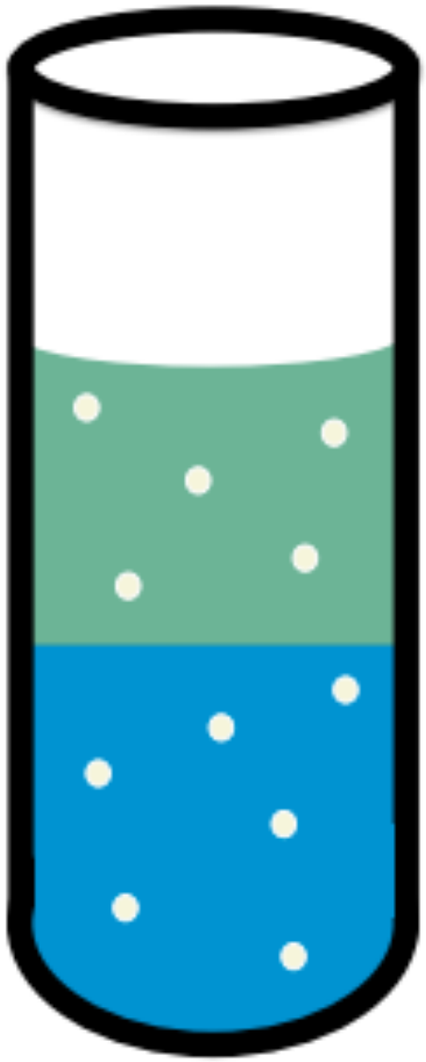
- By around SAMPL4, many methods had improved dramatically
- Successful methods saw broad adoption
- Could detect experimental problems

In SAMPL5, we moved away from hydration free energies to related properties which are more easily* measurable

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

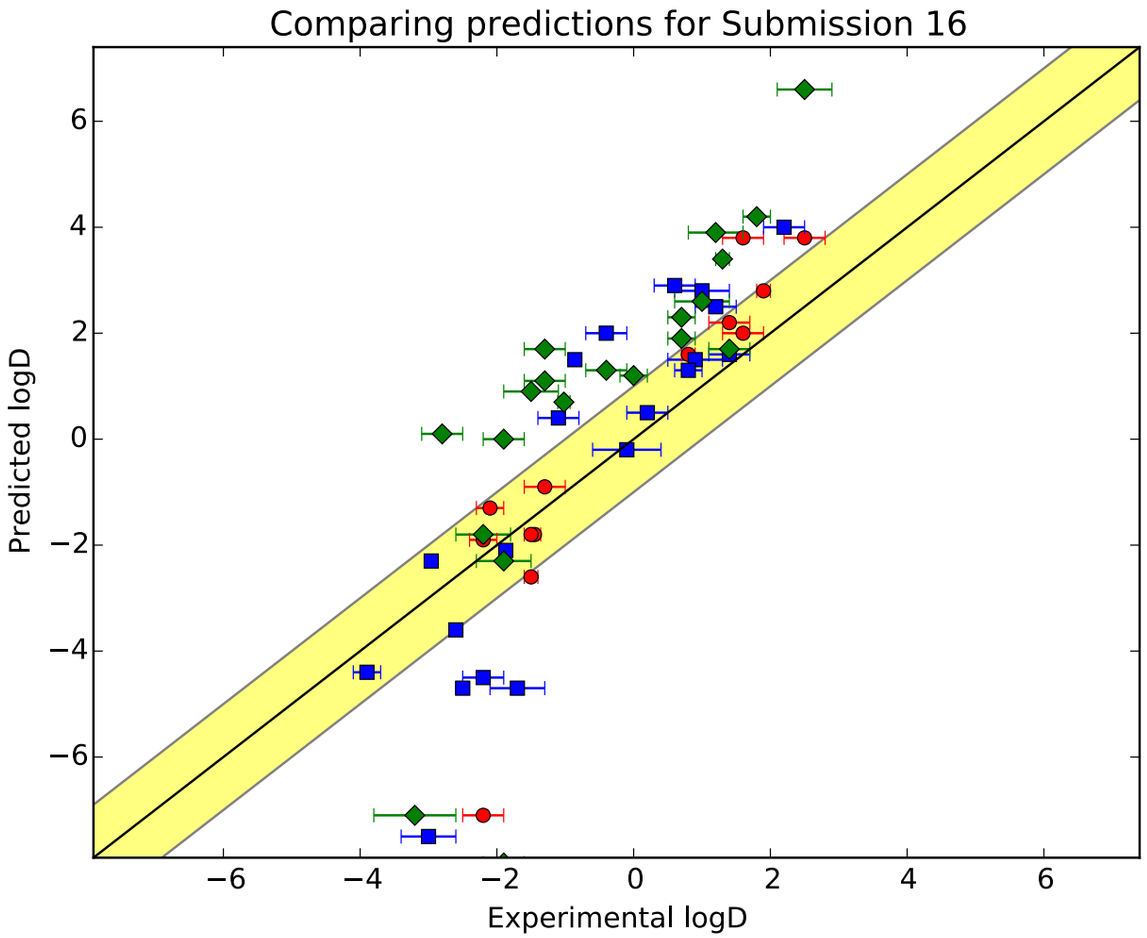
$$P_{cyc} = \frac{[Neutral\ solute\ in\ cyclohexane]}{[Neutral\ solute\ in\ water]}$$

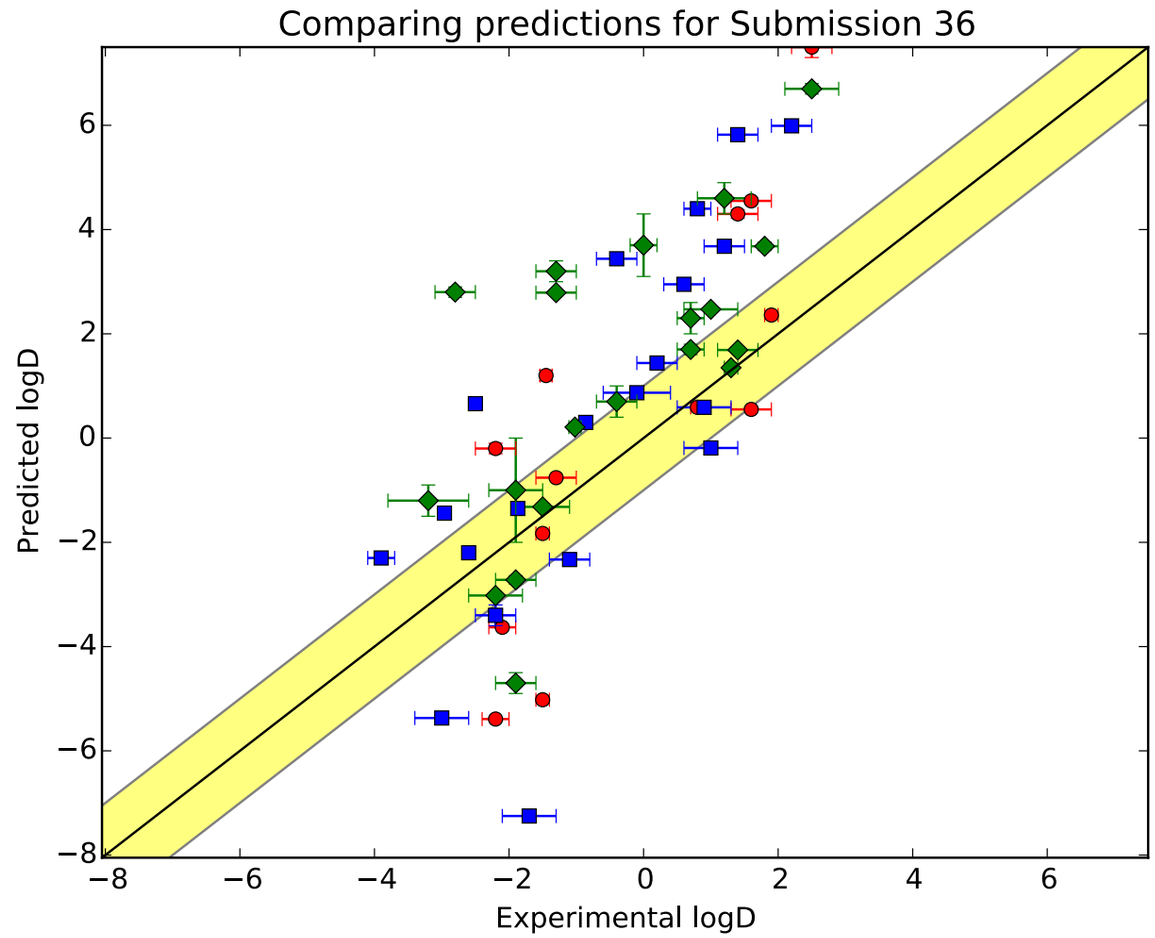
$$D_{cyc} = \frac{[Solute\ in\ cyclohexane]}{[Solute\ in\ water]}$$



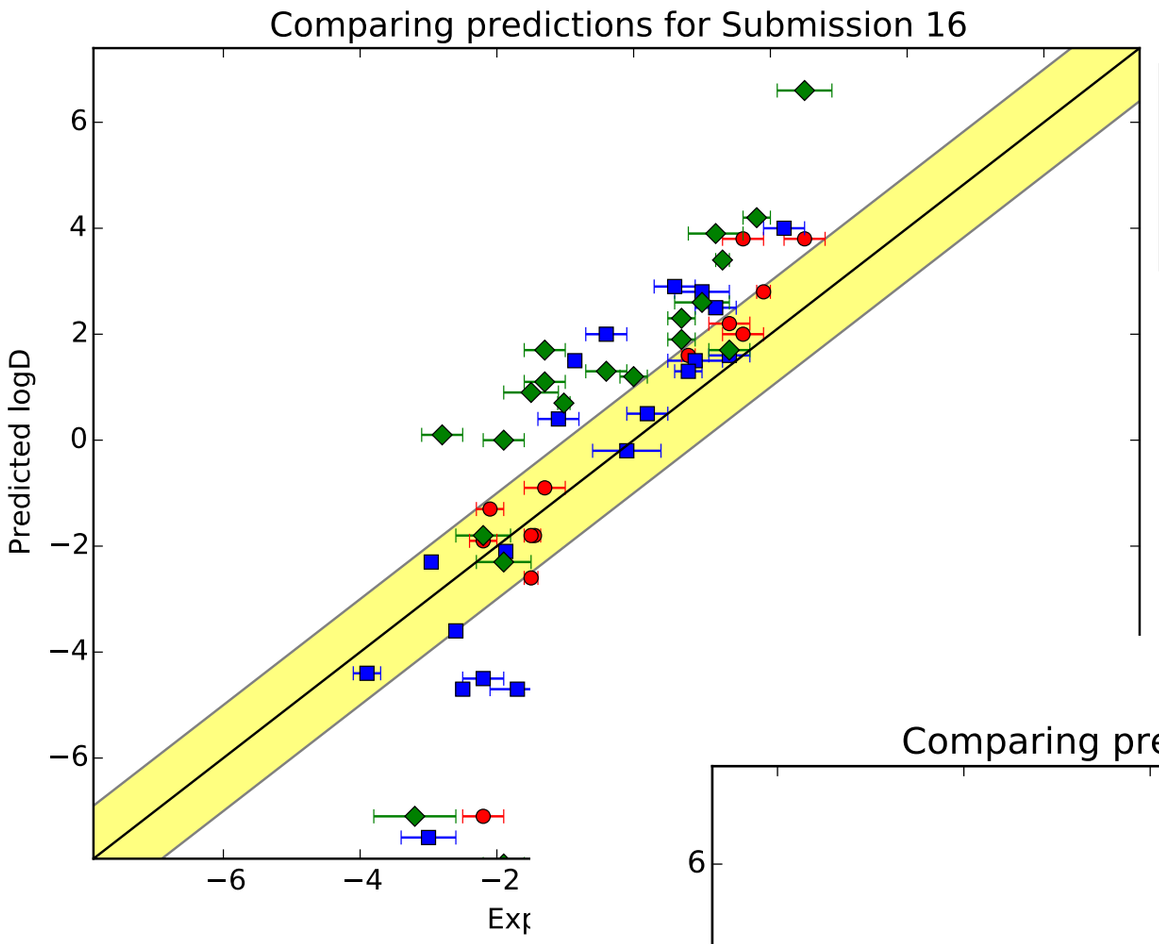
pH dependent, so we report $\log D_{7.4}$ at pH 7.4

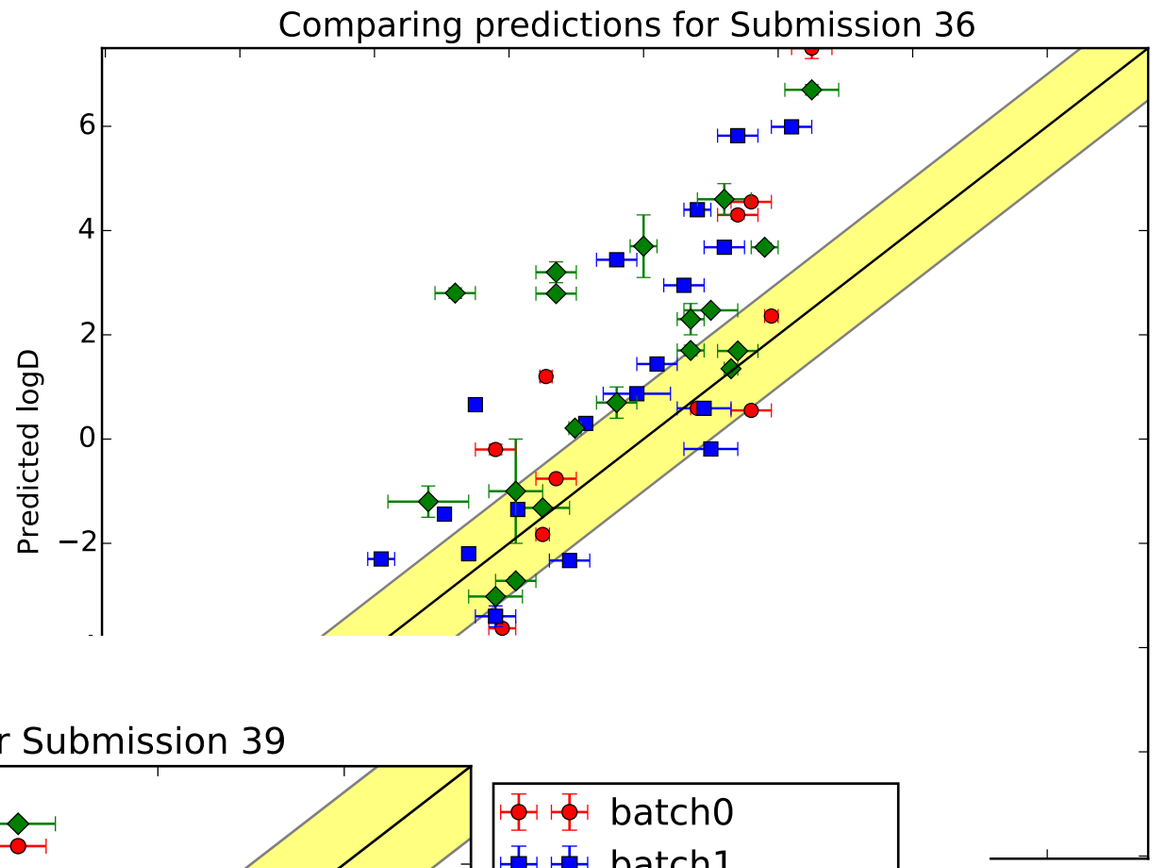
People were able to do fairly well

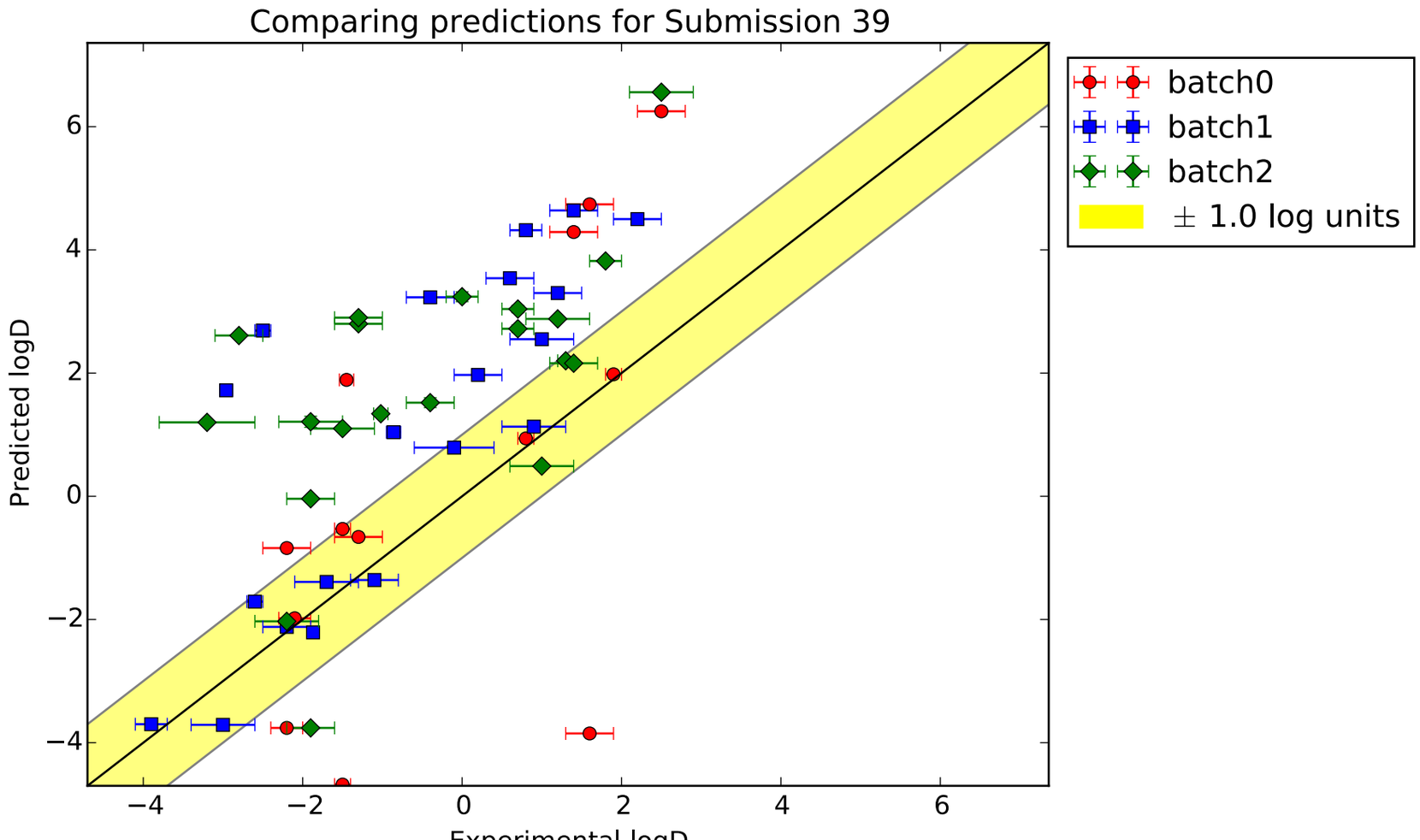




People were able to do fairly well

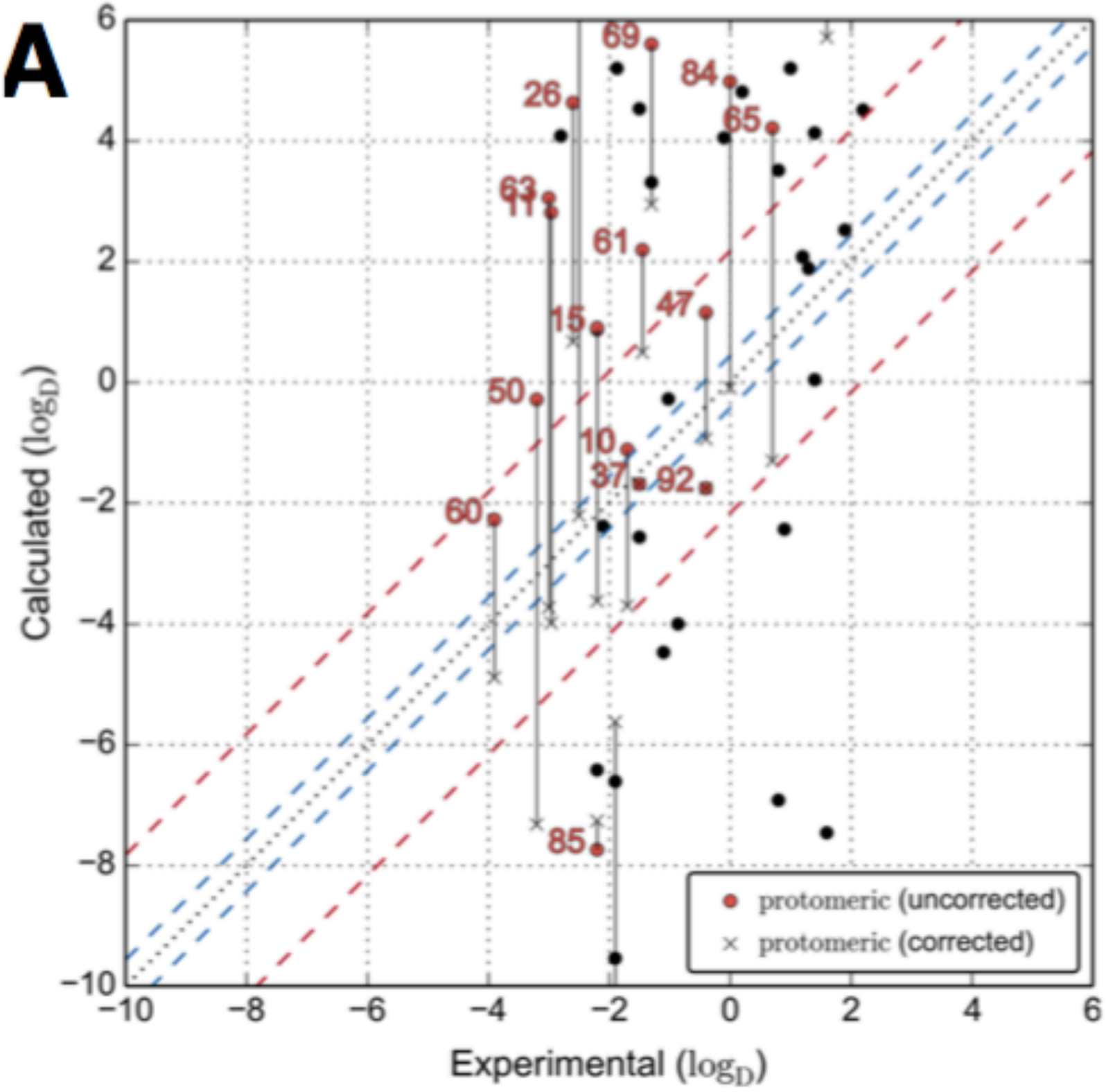






But it turns out here, the difference between logP and logD is crucial

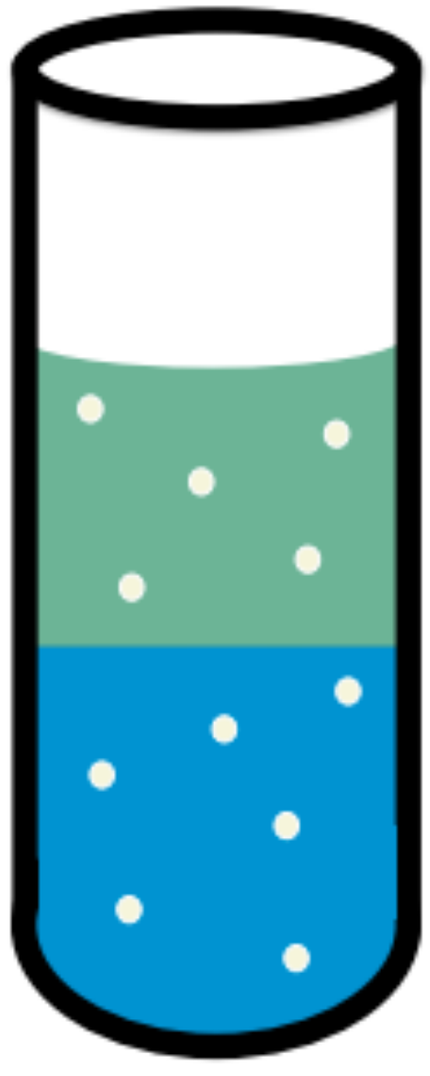
Including predicted pKa's dramatically improves predicted values



Data from Pickard et al.

This led us to plan SAMPL6 to focus on predicting logD *given* pKa values

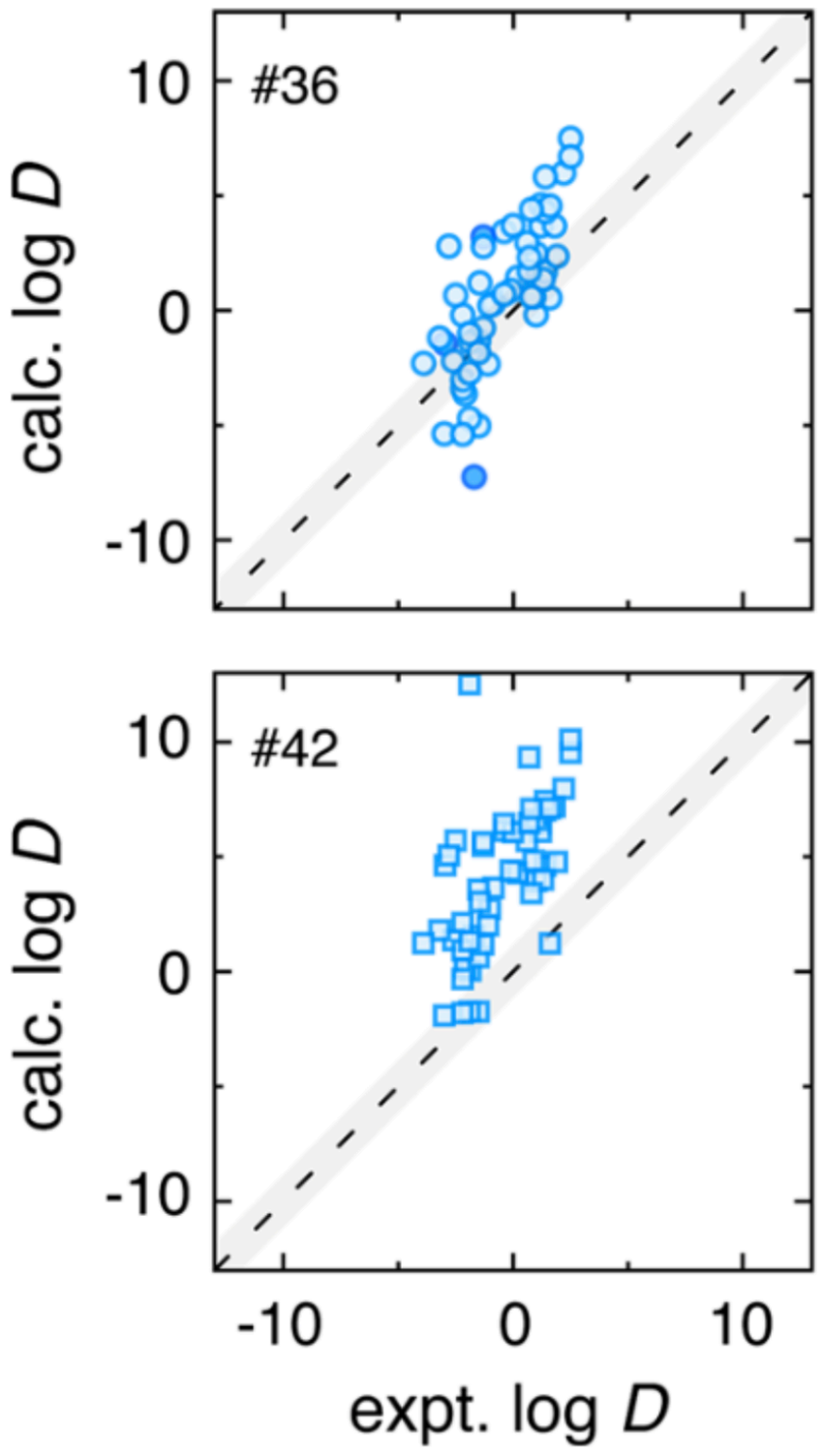
However, collection of the combined data set took too long, so we decided to run an interim challenge on pKa while continuing logD measurements



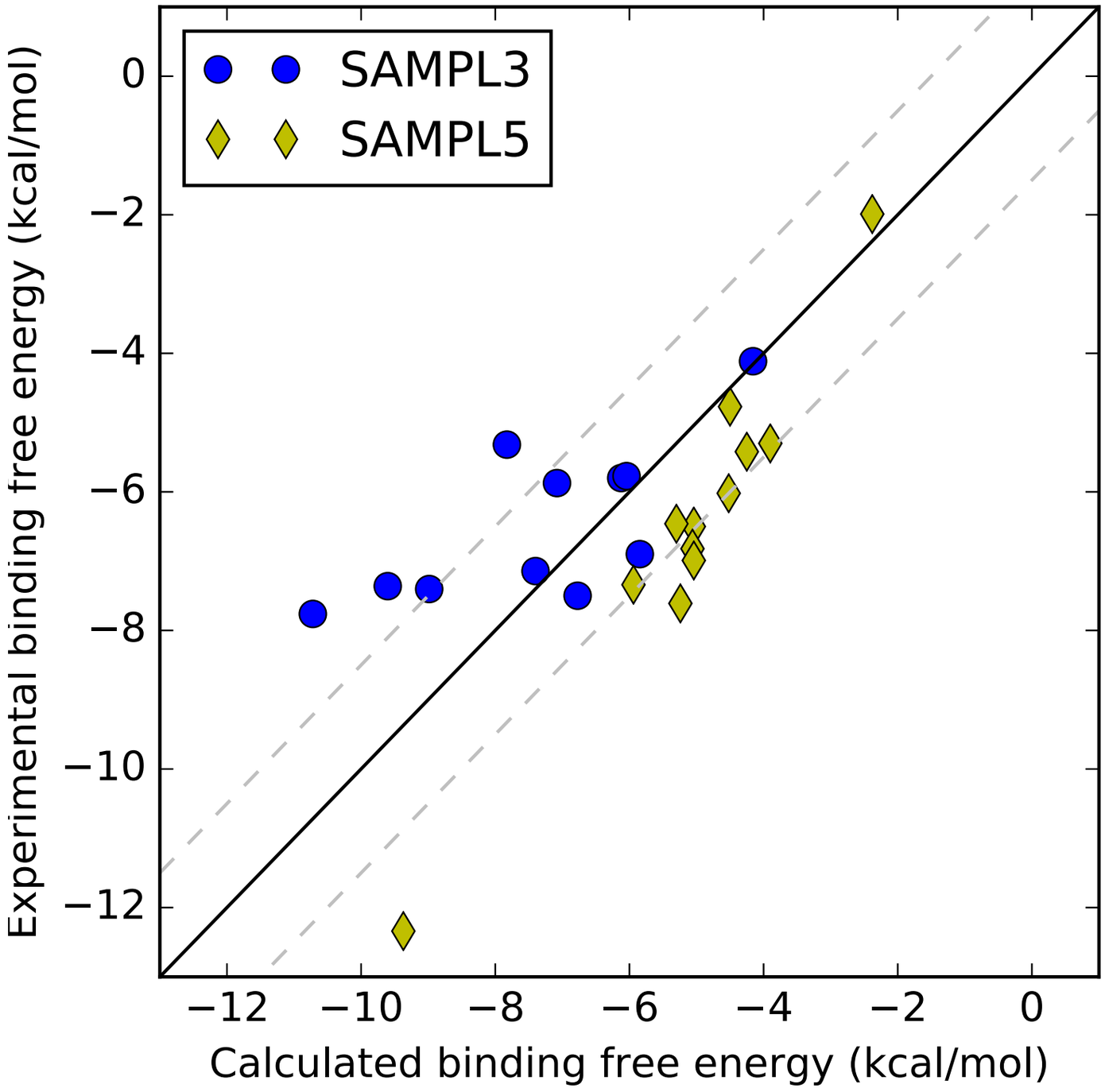
This type of data has been extremely useful in driving and testing new developments

Re-parameterization of GAFF to improve pure solvent dielectric constants resulted in dramatically better predictions (Fennell et al.)

For more on the history of SAMPL and future plans, see our grant proposal at https://escholarship.org/uc/item/7cf8c6cr

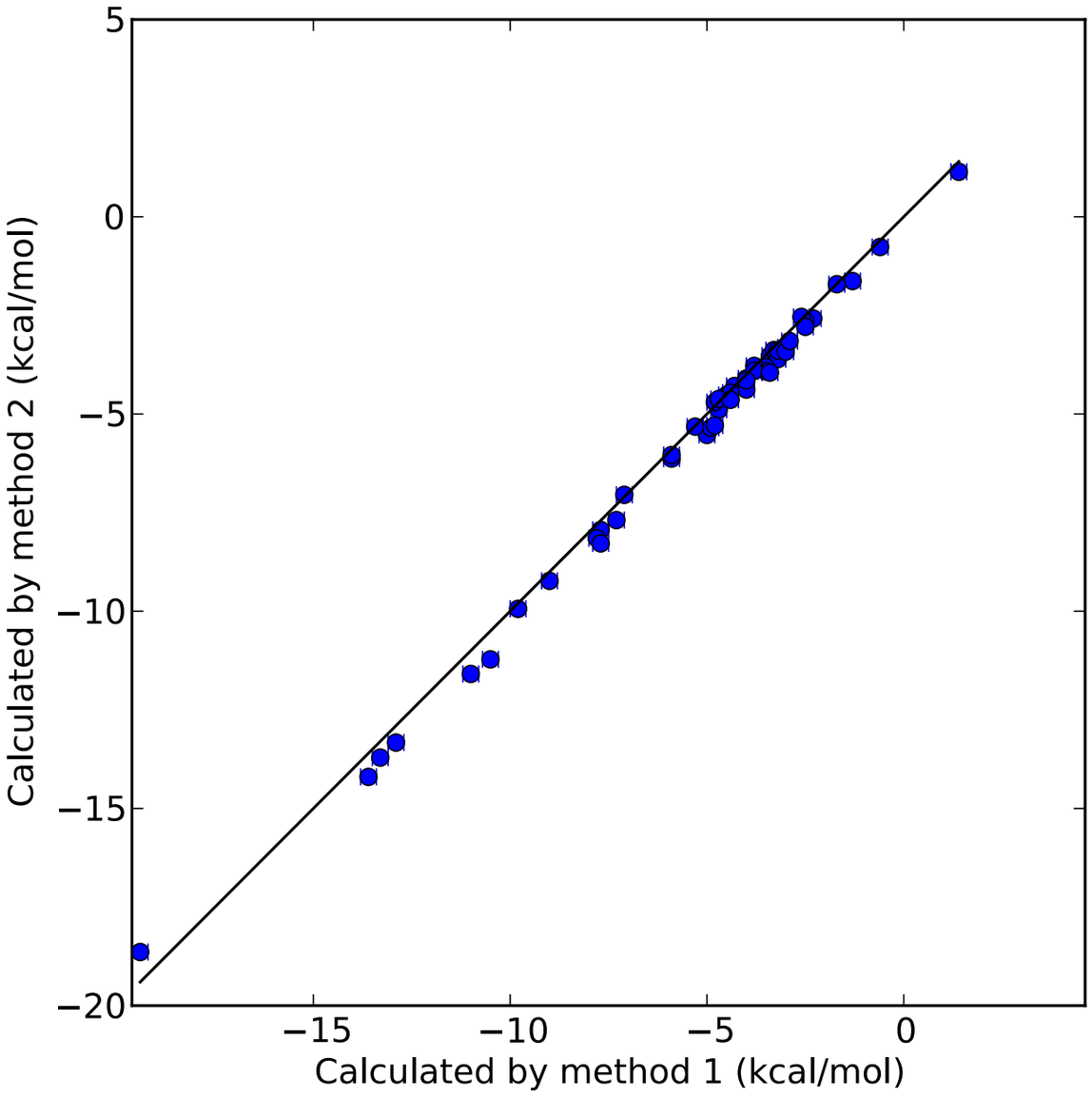


It’s also driven progress in host-guest binding prediction

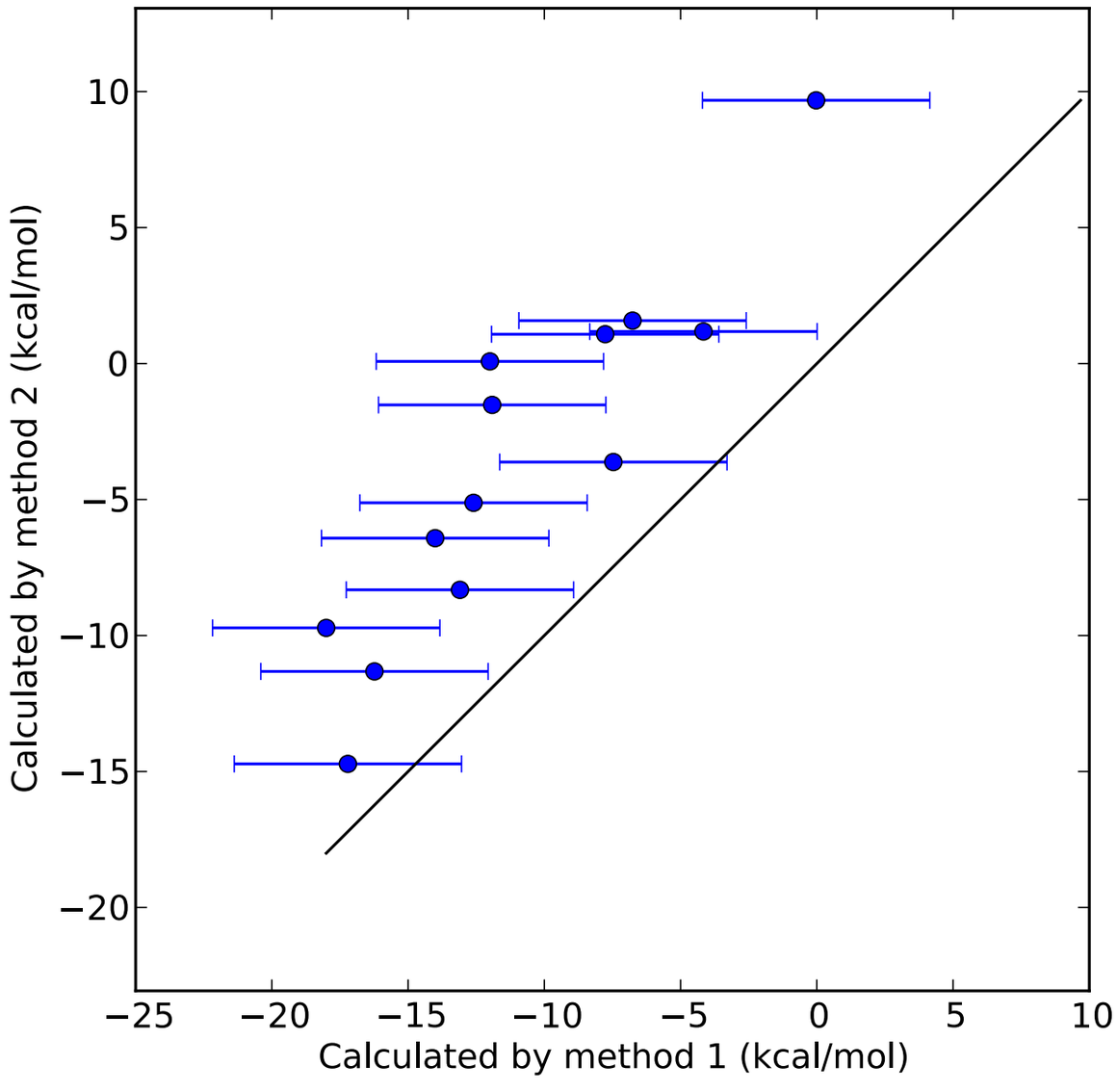


We’re trying to make more use of reference calculations to allow convergence, efficiency tests

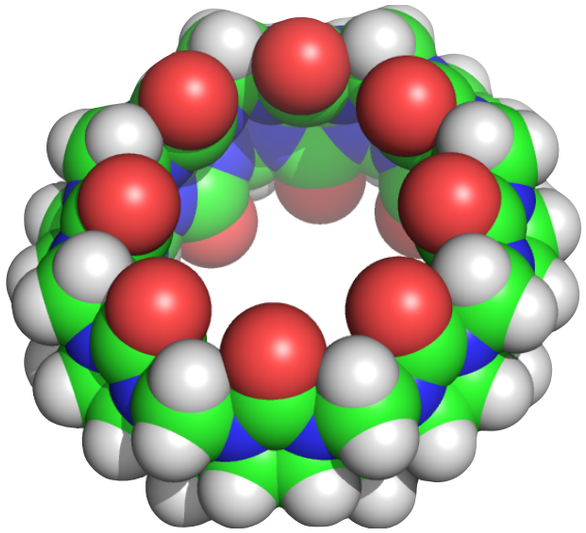
For SAMPL4 hydration, methods which are the same agree

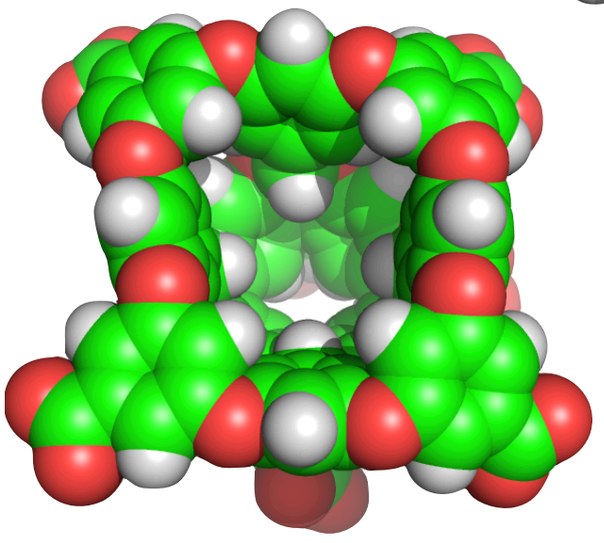


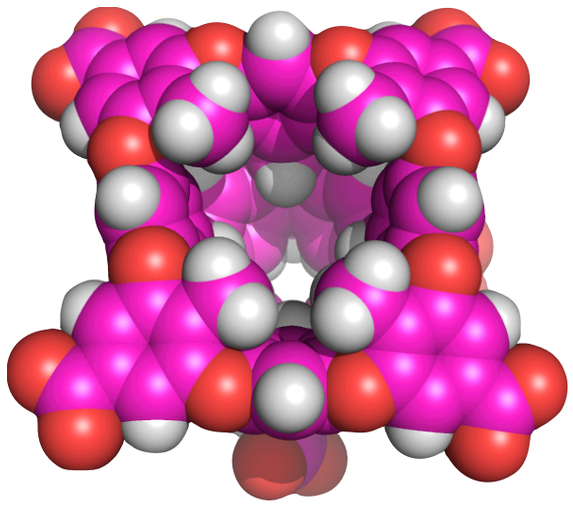
For HG systems, things are not necessarily so simple

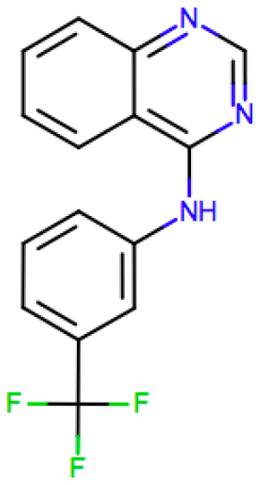


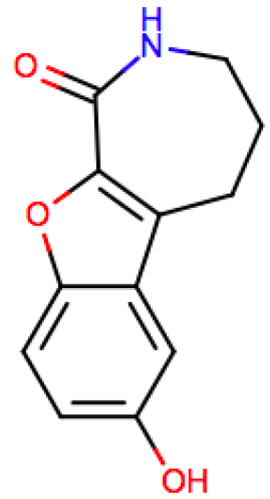
SAMPL6 focused on pKa and host-guest binding, trying to work towards protein-ligand binding

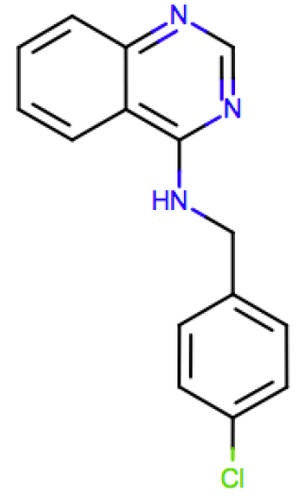


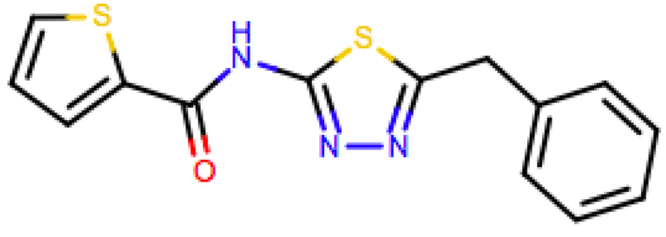




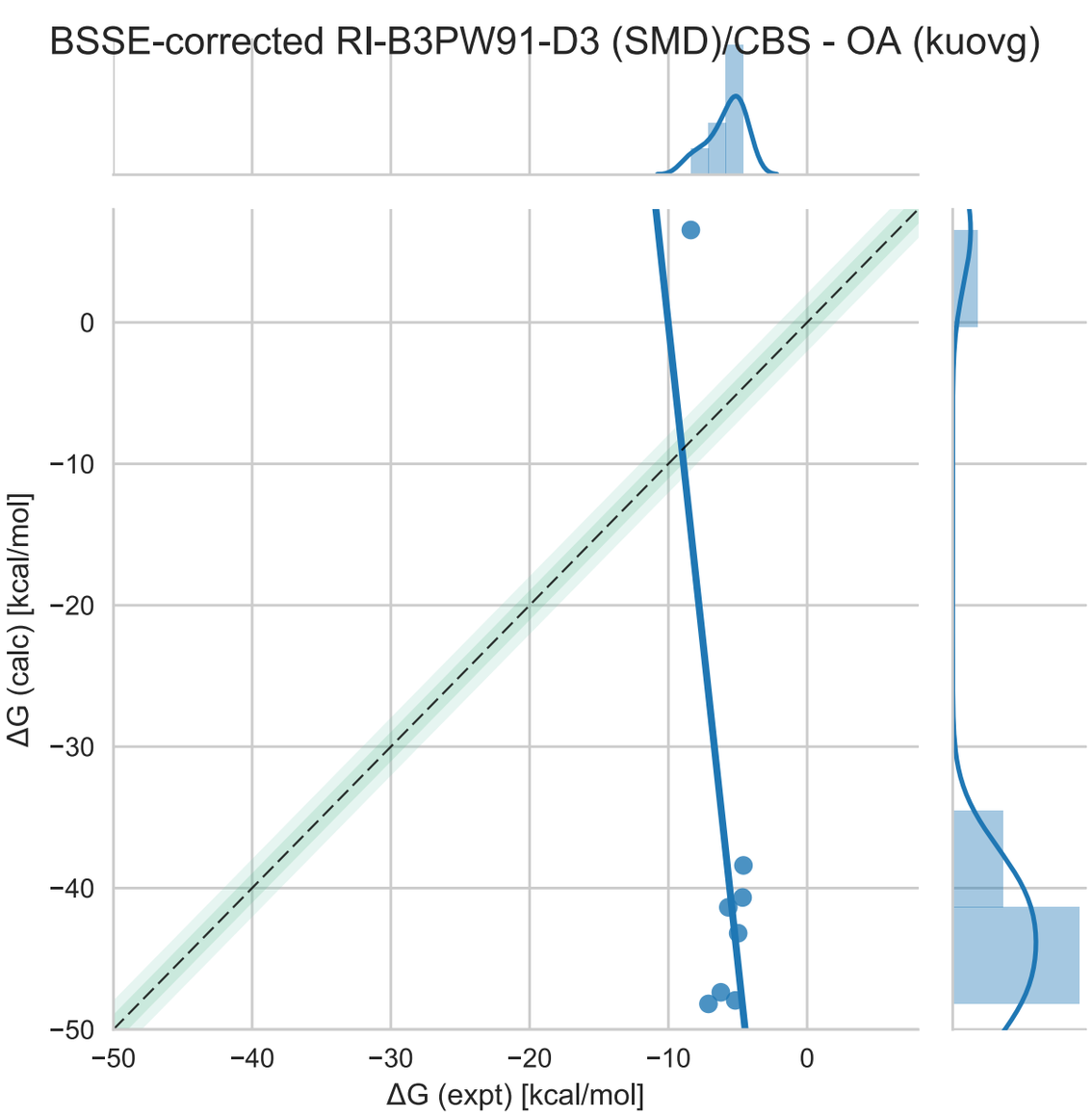


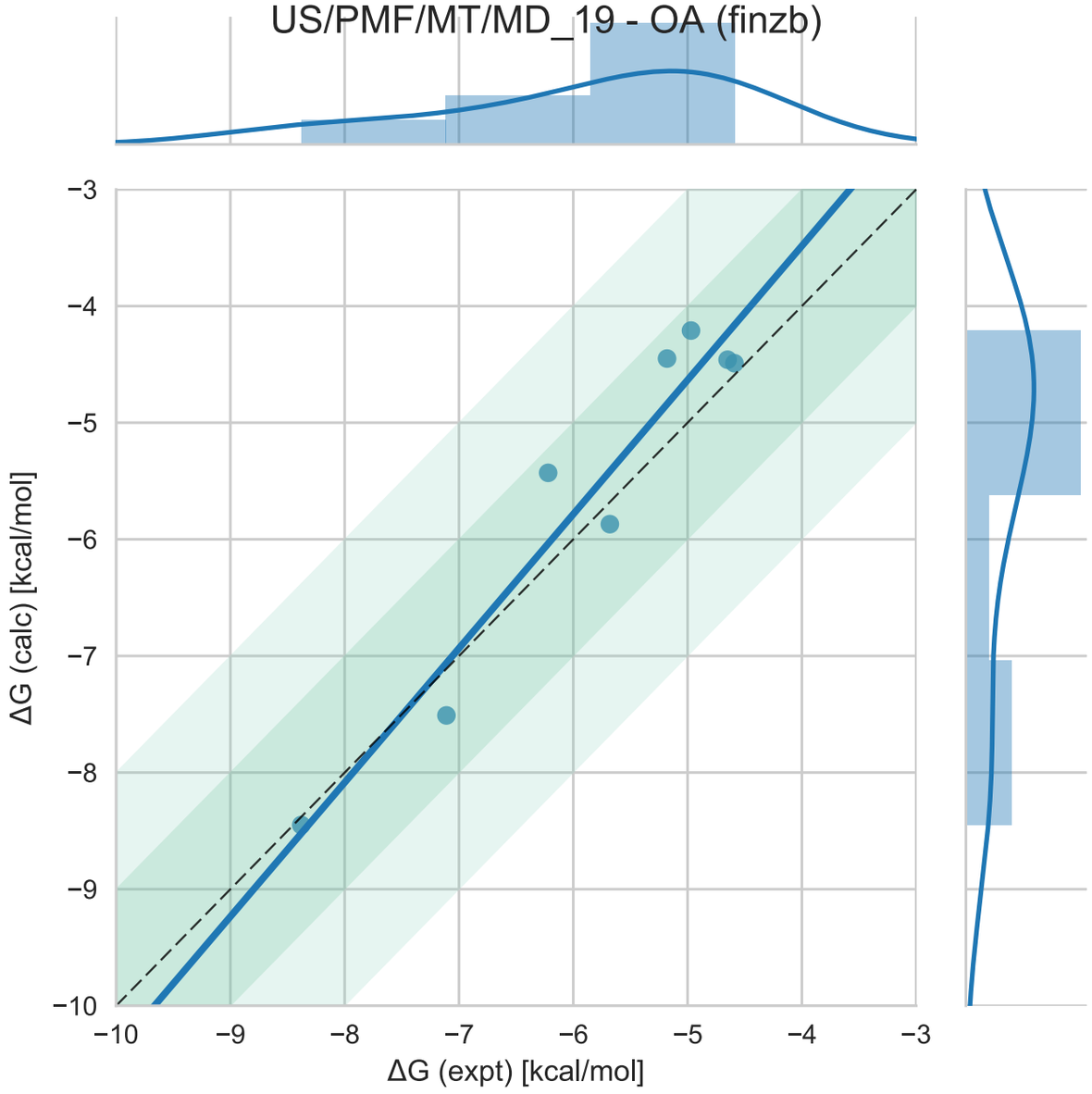






There’s still a long ways to go, and people are still learning big lessons

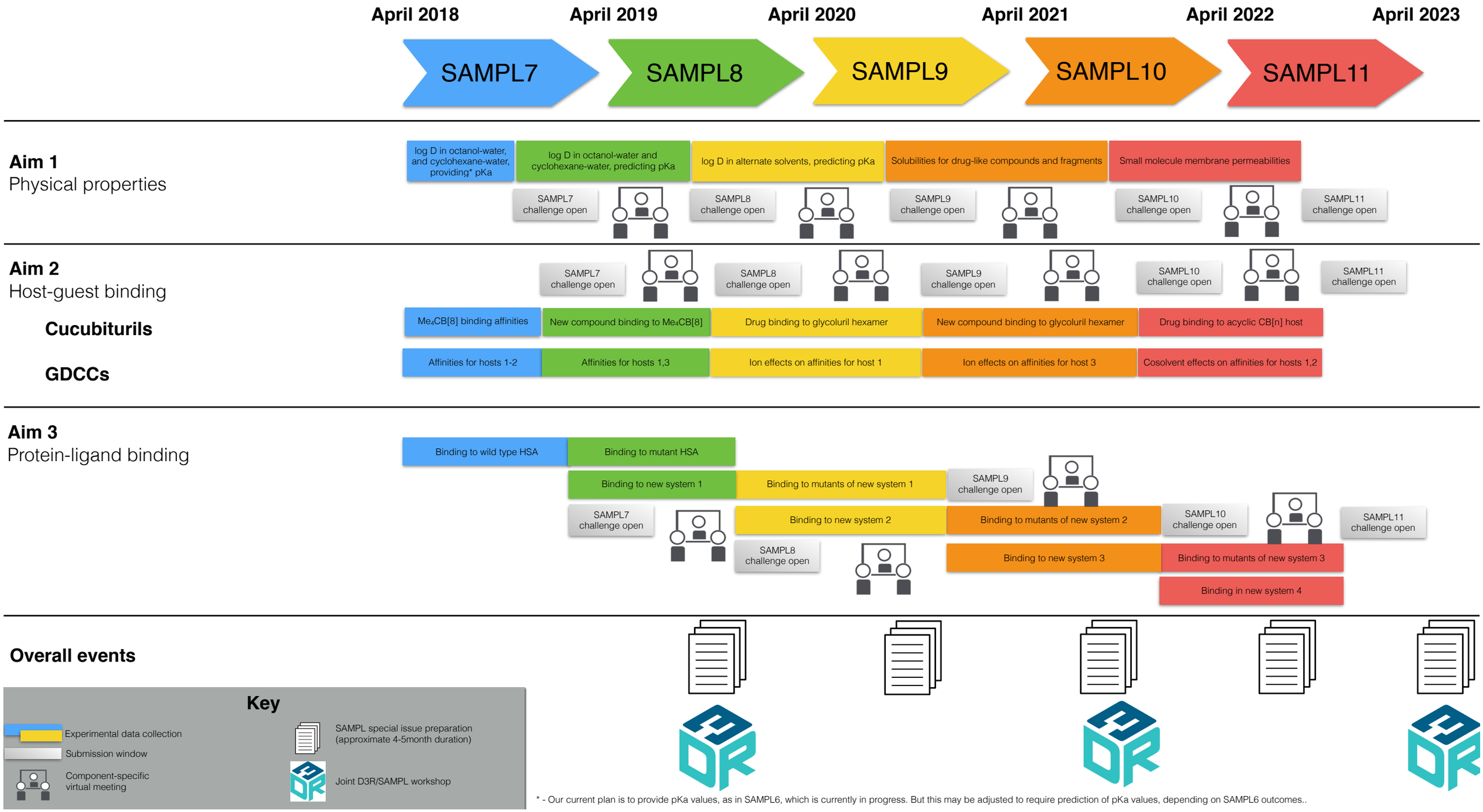




SAMPL6 papers are encouraged:
Special issue, JCAMD

- Firm June 1 submission deadline
- Review process will be as normal, but if you submit, expect to serve as a reviewer
- Paper order strongly influenced by submission order
- Paper titles should include “SAMPL6”
- Cover art selected from among first few submissions, so submit early!
- Online publication as papers are ready
- Will include work on pKa experiments; host-guest work published separately

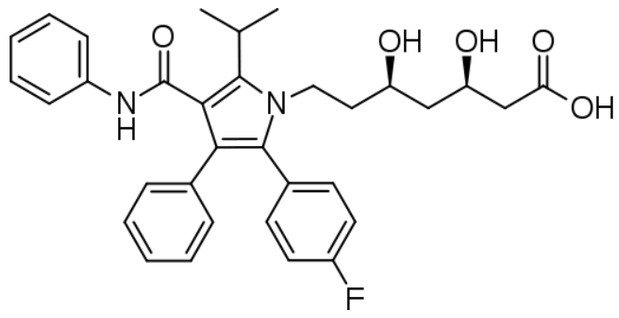
Our plan for future SAMPL challenges involves three tracks: Phys props., HG binding, P-L binding



* - Our current plan is to provide pKa values, as in SAMPL6, which is currently in progress. But this may be adjusted to require prediction of pKa values, depending on SAMPL6 outcomes..

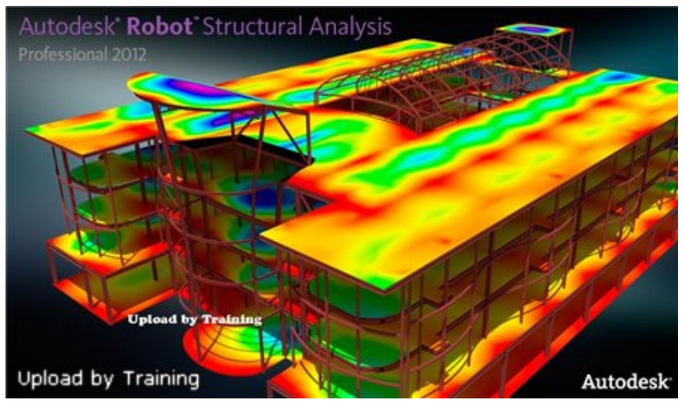
https://escholarship.org/uc/item/7cf8c6cr

HOW DO WE TURN DRUG DISCOVERY FROM RESEARCH



("it sometimes works")

INTO AN ENGINEERING SCIENCE?



("it actually works")

Aim 1

Physical properties

log D in octanol-water, and cyclohexane-water, providing* pKa

log D in octanol-water and cyclohexane-water, predicting pKa

log D in alternate solvents, predicting pKa

Solubilities for drug-like compounds and fragments

Small molecule membrane permeabilities

Aim 2

Host-guest binding

Cucubiturils

- Me_4CB[8] binding affinities
- Drug binding to glycoluril hexamer
- Drug binding to acyclic CB[n] host
- New compound binding to Me_4CB[8]
- New compound binding to glycoluril hexamer

GDCCs

- Affinities for hosts 1-2
- Ion effects on affinities for host
- Cosolvent effects on affinities for
- Affinities for hosts 1,3
- Ion effects on affinities for host 3

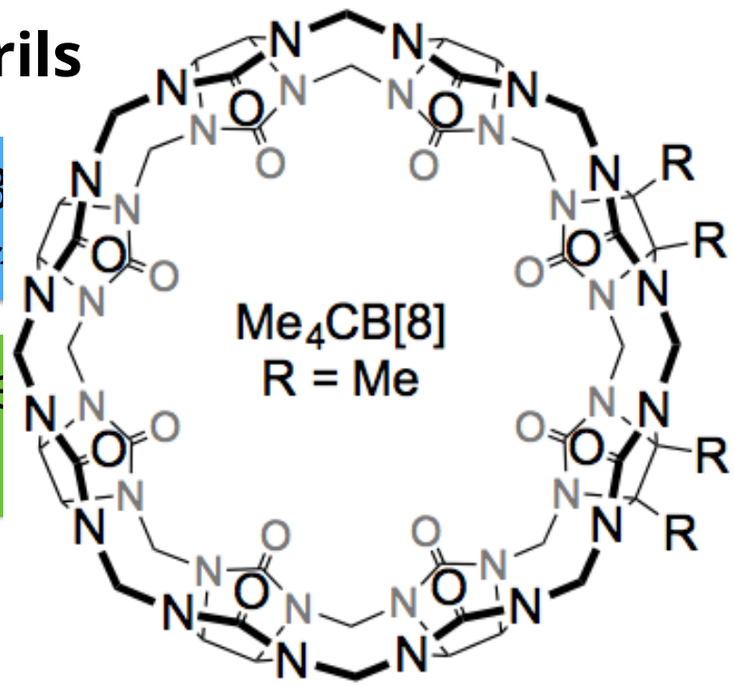
Aim 2

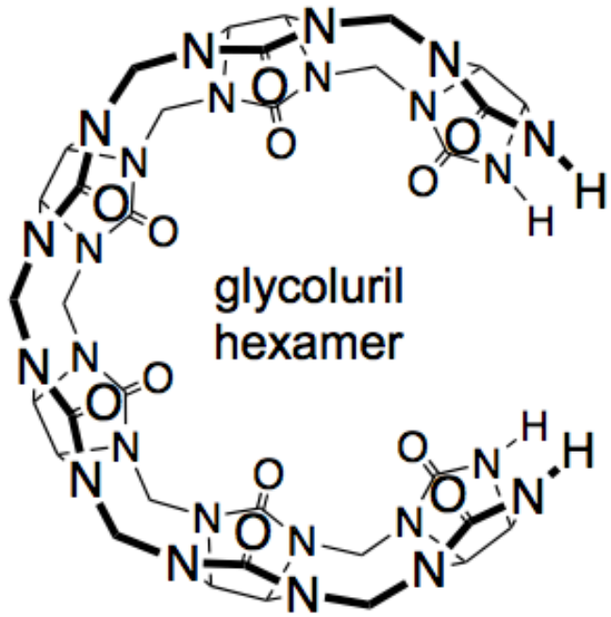
Host-guest binding

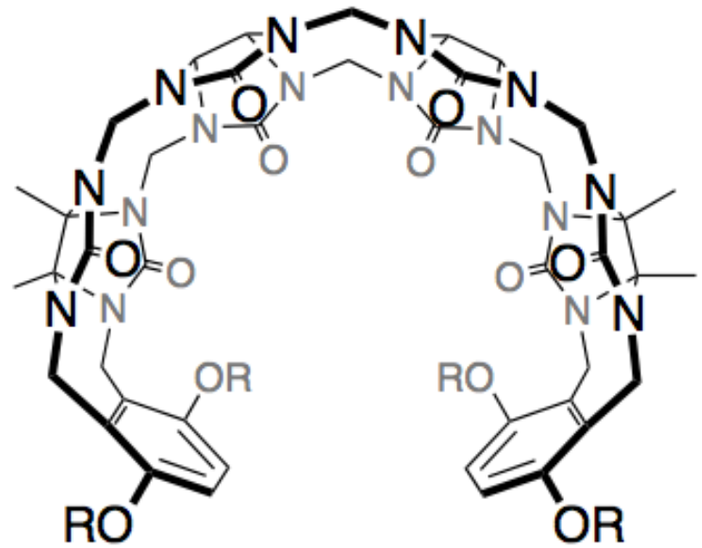
Cucubiturils

Me$_4$CB
af

Ne







M1C$_n$ R=O(CH$_2$)$_n$SO$_3$Na (n=0,3)
M1PC$_0$ R=OPO$_3$H$_2$

GDCCs

Affinities for hosts 1-2

Ion effects on affinities for host

Cosolvent effects on affinities for

Affinities for hosts 1,3

Ion effects on affinities for host 3

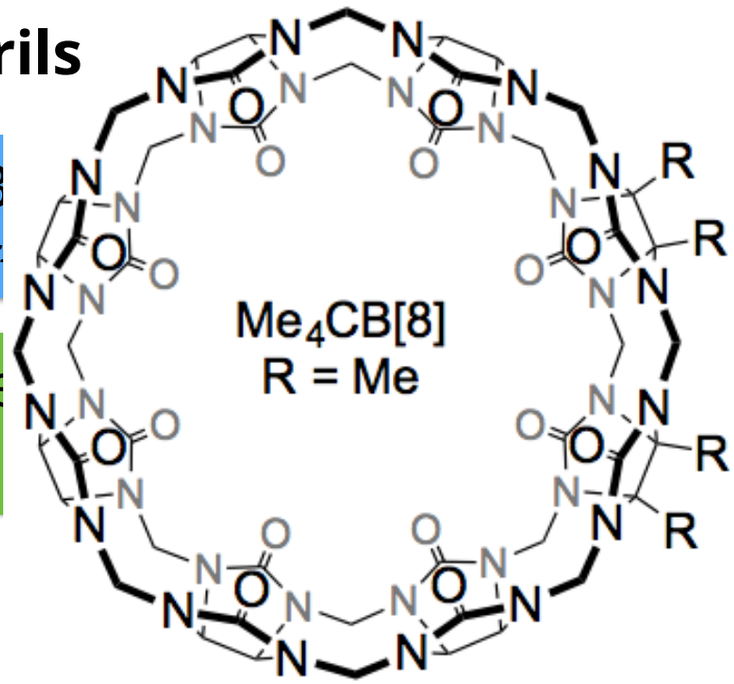
Aim 2

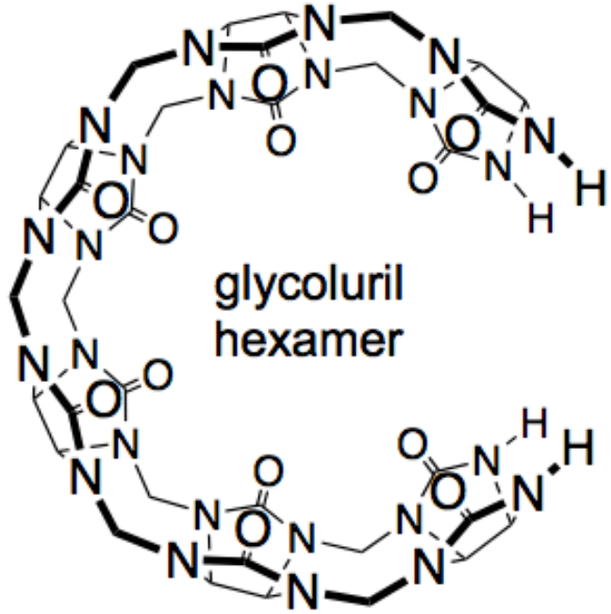
Host-guest binding

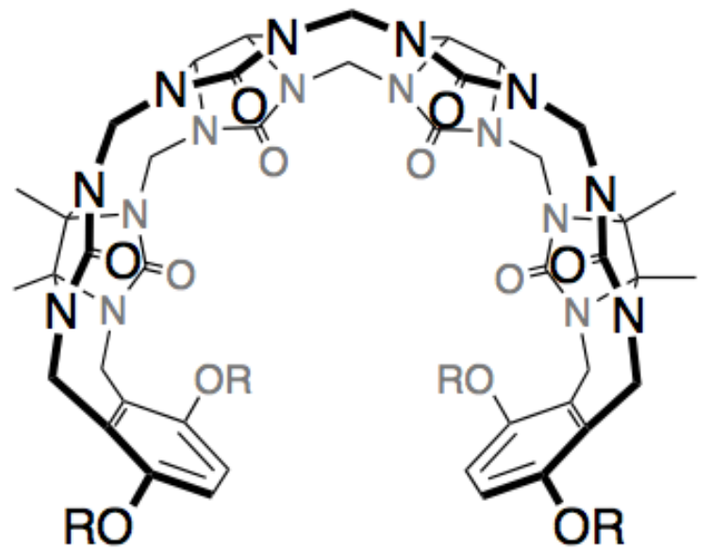
Cucubiturils

Me$_4$CB
af

Ne

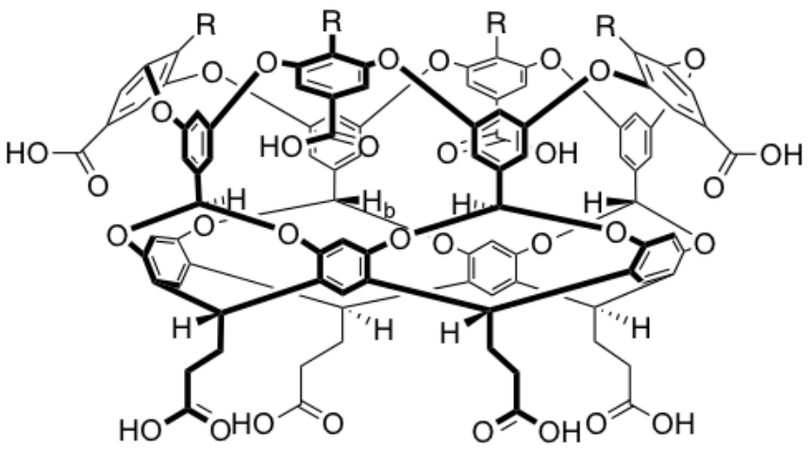




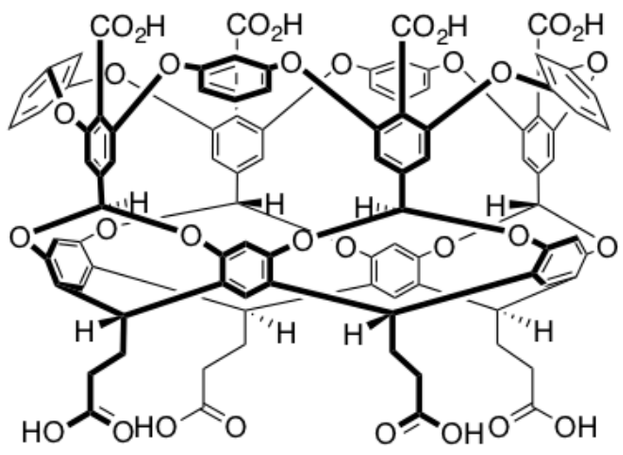


M1C$_n$ R=O(CH$_2$)$_n$SO$_3$Na (n=0,3)
M1PC$_0$ R=OPO$_3$H$_2$

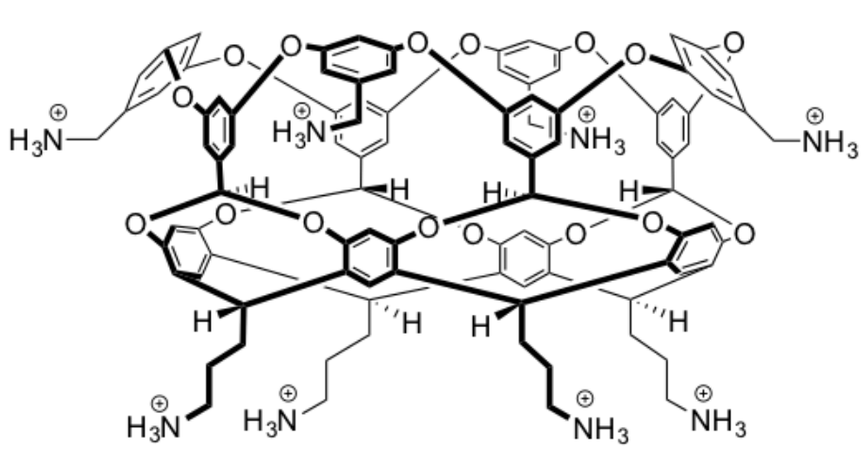
GDCCs



1



2



3

Aim 3

Protein-ligand binding

Binding to wild type HSA

Binding to mutant HSA

Binding to new system 1

Binding to mutants of new system 1

Binding to new system 2

Binding to mutants of new

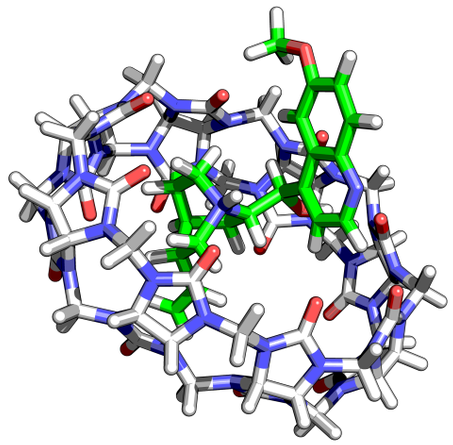
Binding to new system 3

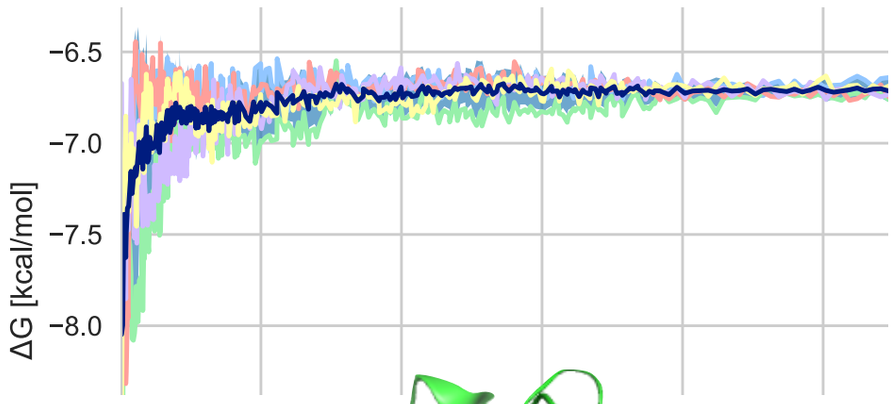
Binding to mutants of new system 3

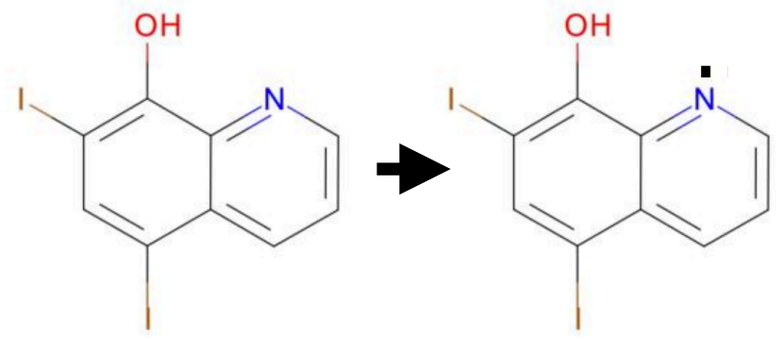
Binding in new system 4

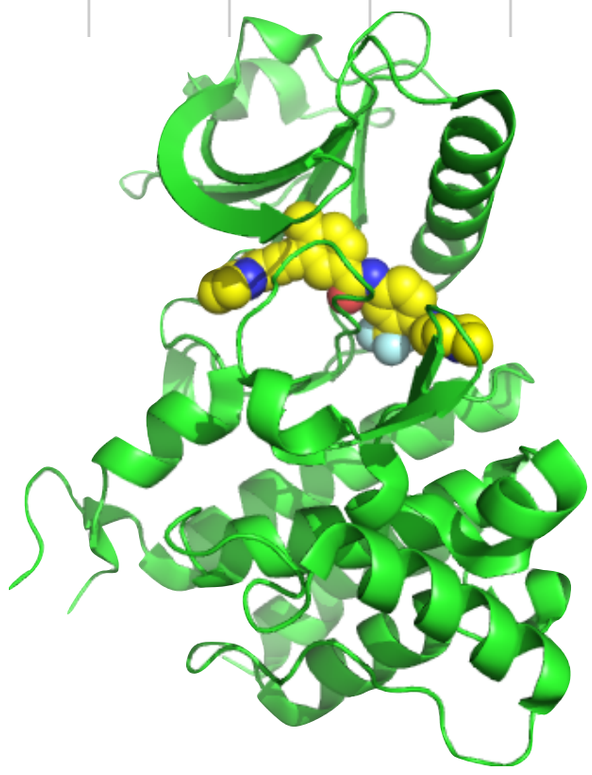
If we could design the ideal SAMPL challenge, what would it look like? Data would be tailored for maximum learning

- Protein-ligand, pKa, logD, and maybe even host-guest data on same compounds
- If your binding prediction is wrong — is it because you got the pKa wrong? Or the solvation free energy/partitioning?
- Is it because of sampling?





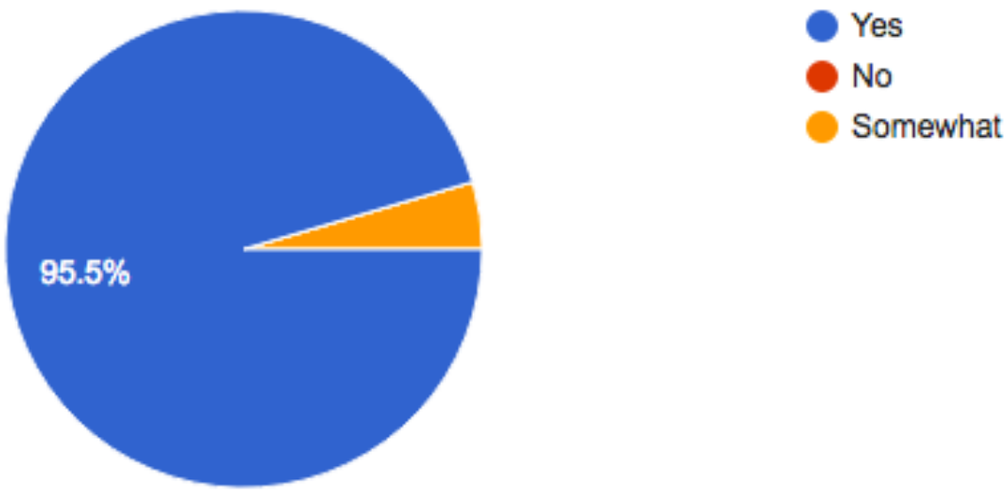




Input we've received on SAMPL has been very supportive

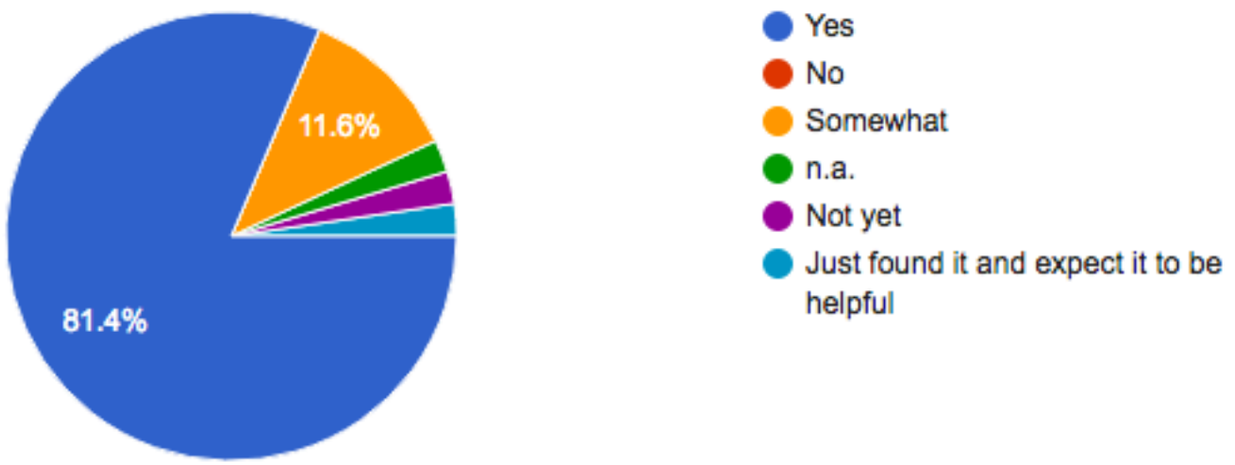
7. Do you see SAMPL as a valuable resource to the community?

44 responses



8. Has SAMPL been important in driving progress in the field or in your group?

43 responses



People were enthusiastic about the benefits of SAMPL

https://escholarship.org/uc/item/2jq8s2zr

Still, the future is uncertain; sustainability requires funding or manpower

Acknowledgments

- D3R for help with SAMPL submission handling and support of SAMPL via this workshop; special thanks to Mike Chiu for submission handling
- Prior SAMPL participants, collaborators, and data providers (too many to list)
- SAMPL6:
  - HG experiments: Bruce Gibb (Tulane), Lyle Isaacs (Maryland) and their group members, including Steven Murkli and Nolan McNeill at Maryland.
  - pKa experiments: Mehtap Isik (Chodera lab) with Dorothy Levorse and Tim Rhodes at Merck; support from Brad Sherborne at Merck; Caitlin Bannan for knowledge transfer and pKa staging help
  - Merck for donation of instrument time, supplies, …
  - Analysis: Mehtap Isik (pKa), Andrea Rizzi (host-guest)
  - Michael Shirts for host-guest input file conversion; Andrea Rizzi, Travis Jensen, Michael Shirts for SAMPLing runs
- All the participants