

Story time and puppet show with Mr. Bayes

Data, Models, and Decisions



John D. Chodera

Computational Biology Program, Memorial Sloan-Kettering Cancer Center

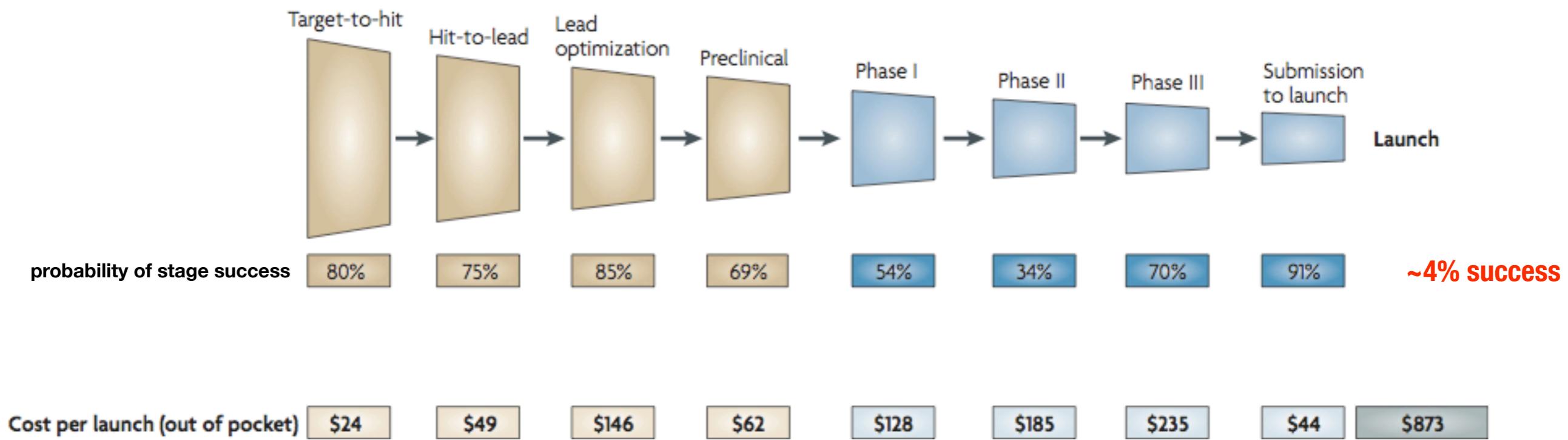
<http://www.choderalab.org>

We are all here because we desperately hope modeling can improve the dismal success rates in drug discovery

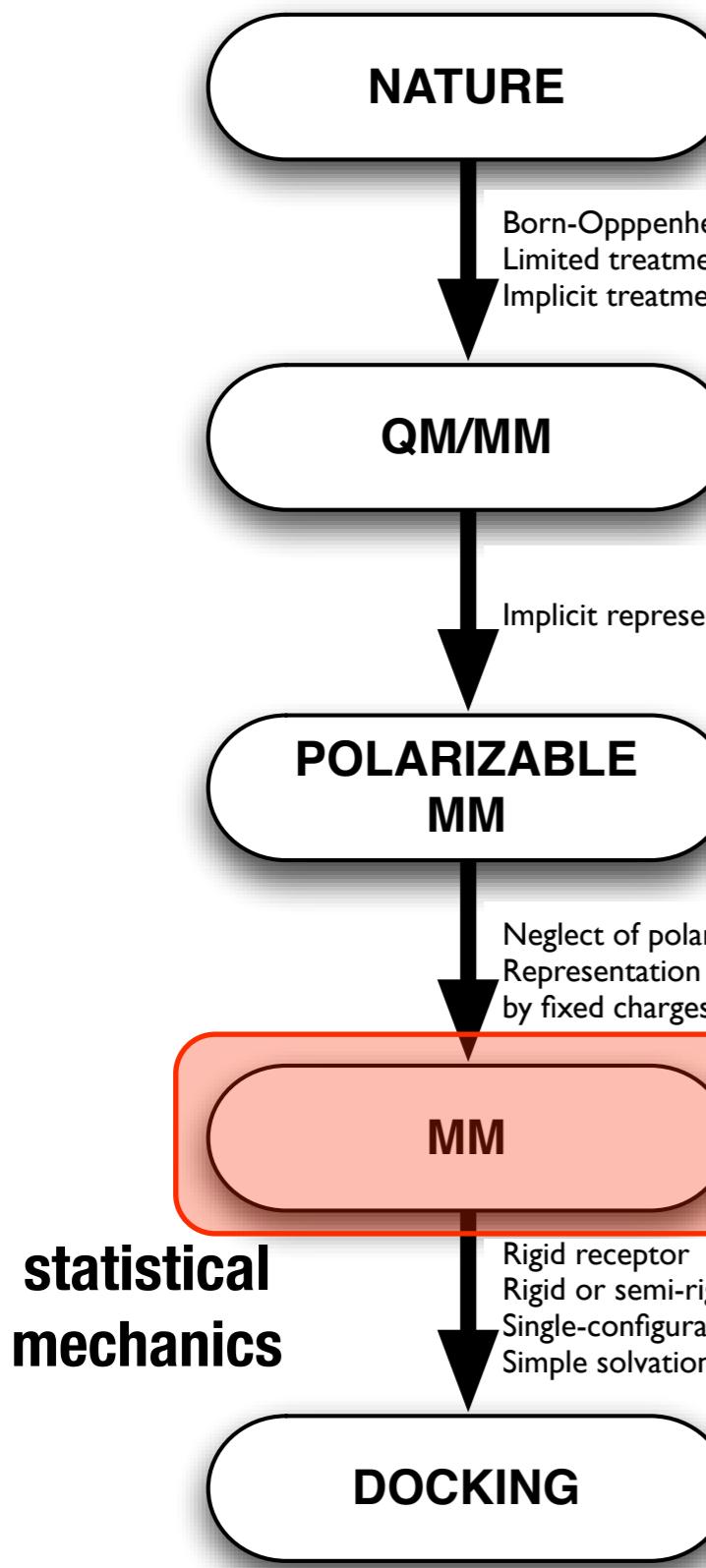
Total pharma research spending has **doubled** to \$65.3B in last decade

Number of new molecular entities approved by FDA 2005-2009 is **half** that from previous five years

Number of truly innovative new molecular entities has **remained constant** at 5-6/year



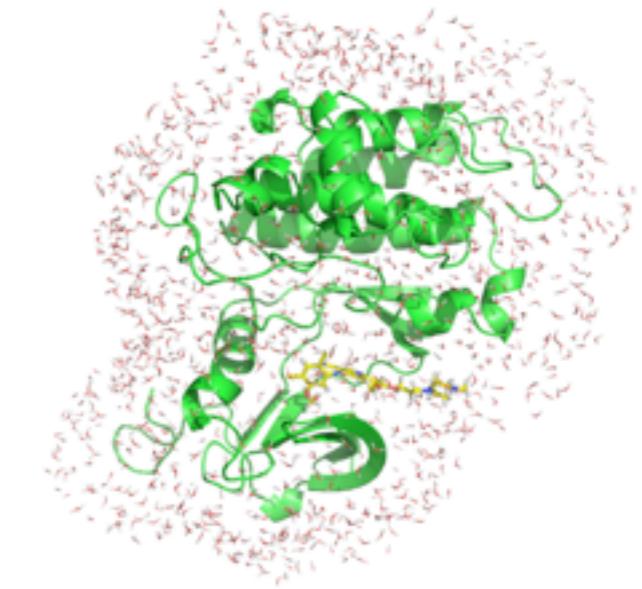
What details are crucial for useful accuracy in rational design? Our lab is trying to find out!



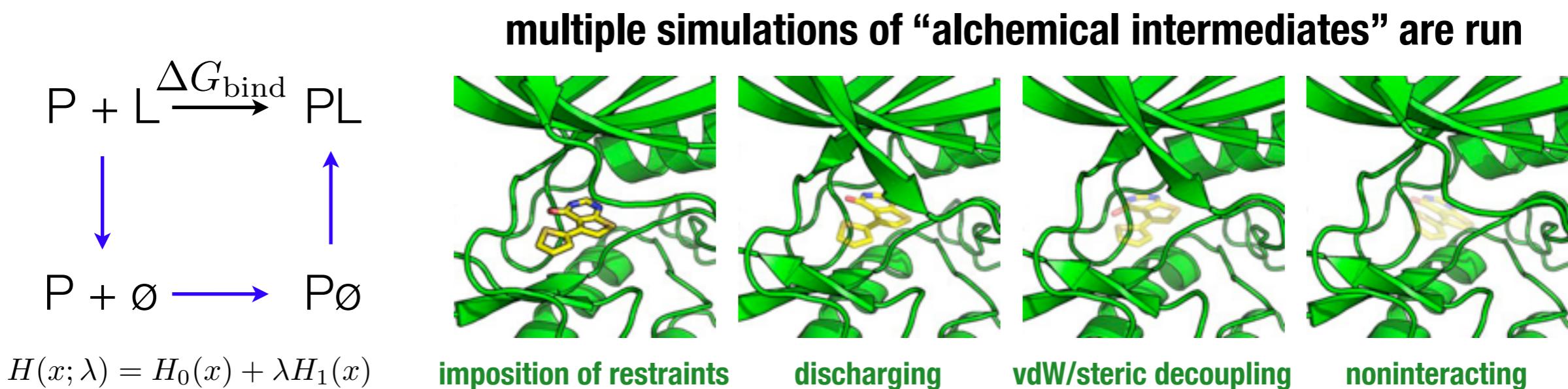
molecular mechanics potential energy function (e.g. AMBER)

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

if accurate enough, systematically remove detail



Alchemical free energy calculations provide a rigorous way to efficiently compute binding affinities for a given forcefield



Requires **orders of magnitude** less effort than simulating direct association process, but includes all enthalpic/entropic contributions to binding free energy.

Not be as fast as OpenEye tools, but will eventually give true binding free energy for a given forcefield (be it bad, or really really bad).

Learning (to not get depressed) from failure: Fail fast, fail cheap, learn something in the process

computational predictions



experimental confirmation

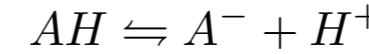
Predicting binding affinities computationally



\$500
1.5 TFLOP
50-500x speedup

GPU acceleration

$$\pi(x)K(x,y) = \pi(y)K(y,x)$$

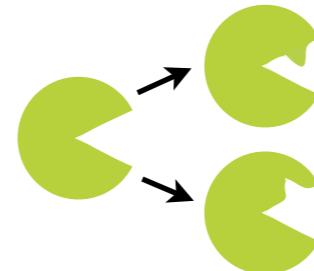


MCMC framework



enhanced algorithms

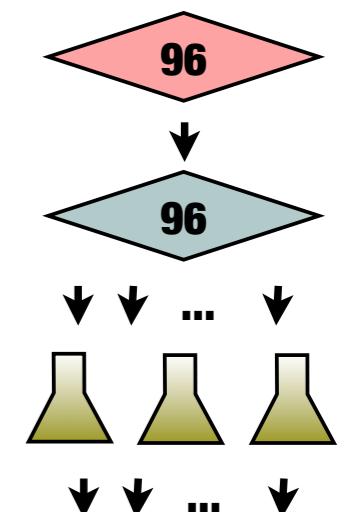
Testing predictions by mutating the target protein



mutate proteins
instead of ligands

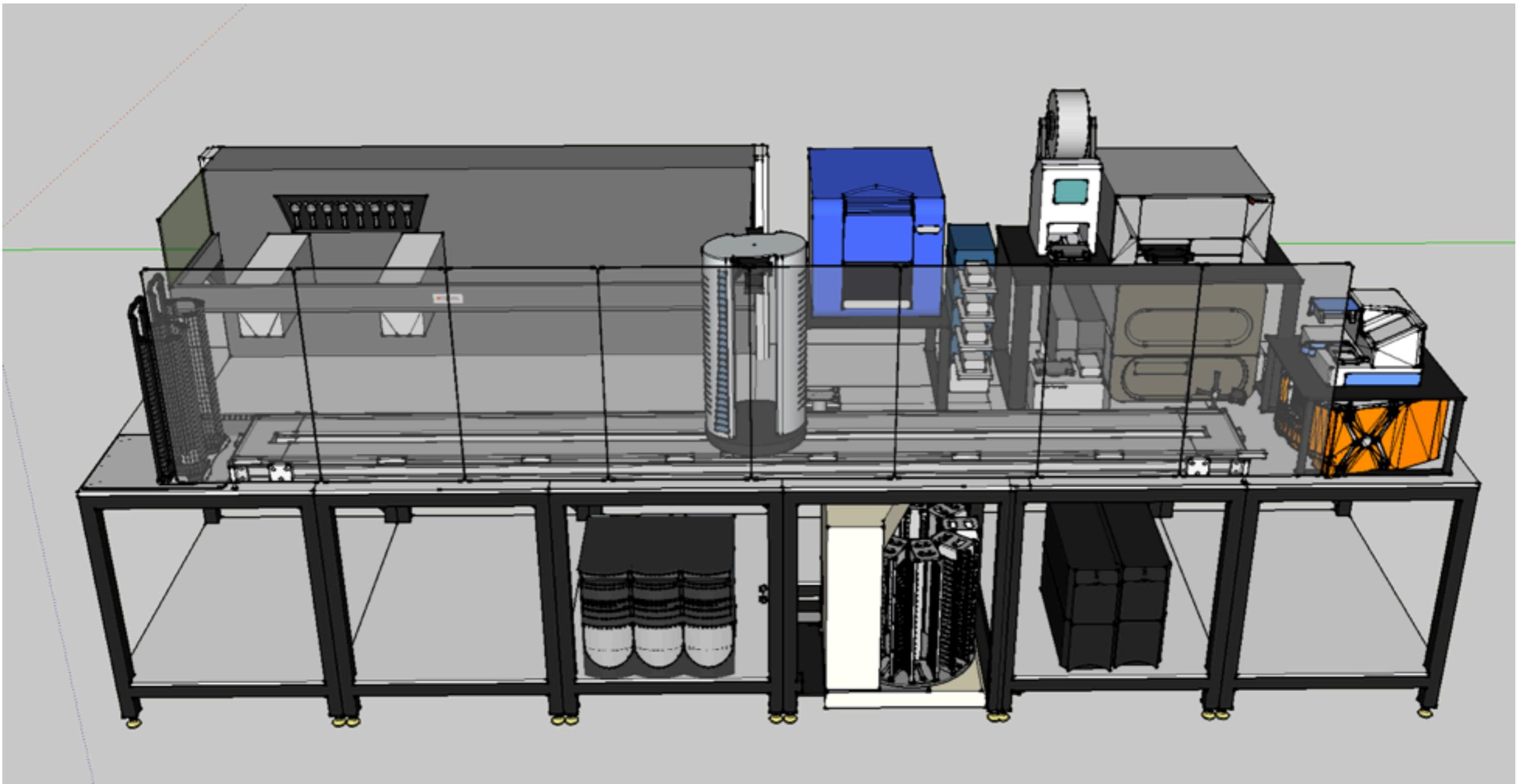


buy inexpensive
ligands



high-throughput
experiments

The last, best hope for data: RUG-1 (robotic undergraduate no. 1)



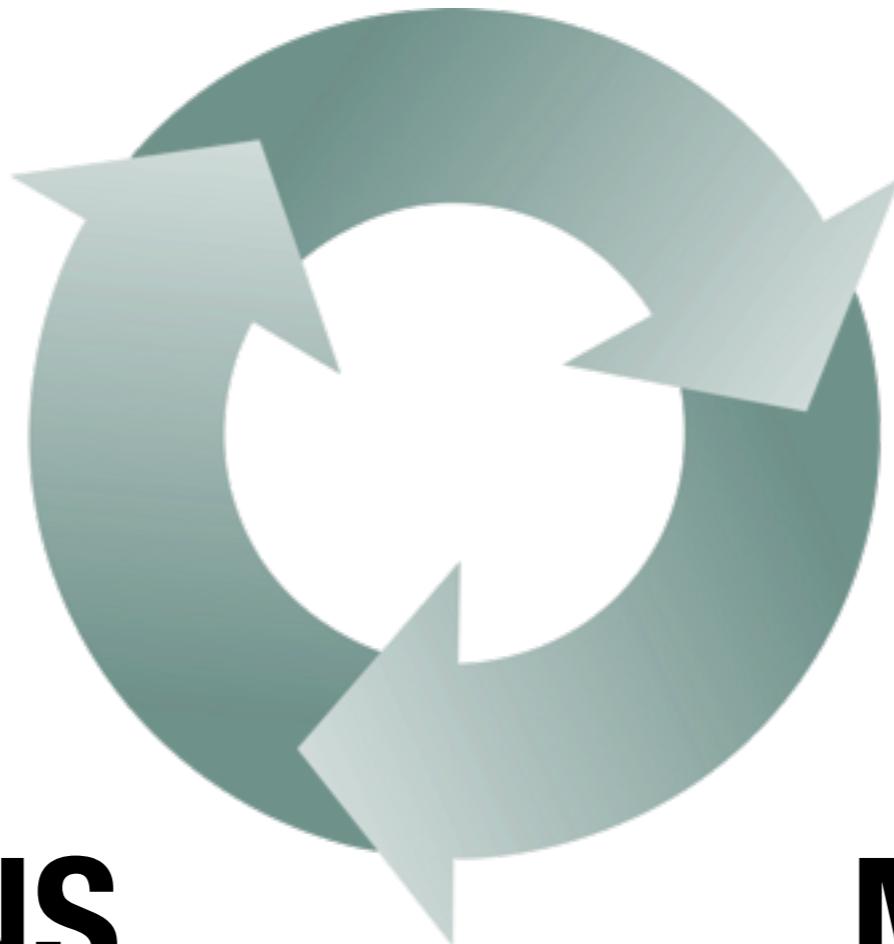
Automated platform for bacterial cloning, mutagenesis, expression, purification, and binding affinity measurement with 24/7 wallet-draining capability

The universal cycle of progress



The universal cycle of progress

DATA



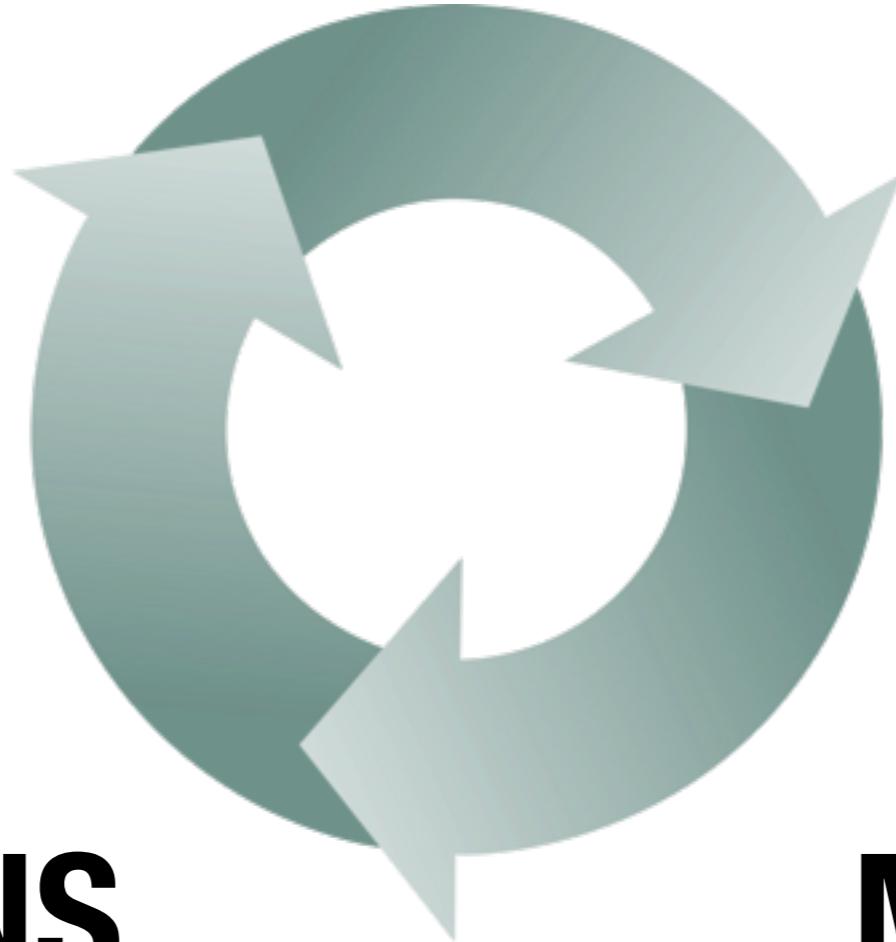
DECISIONS

MODELS



The universal cycle of progress

DATA

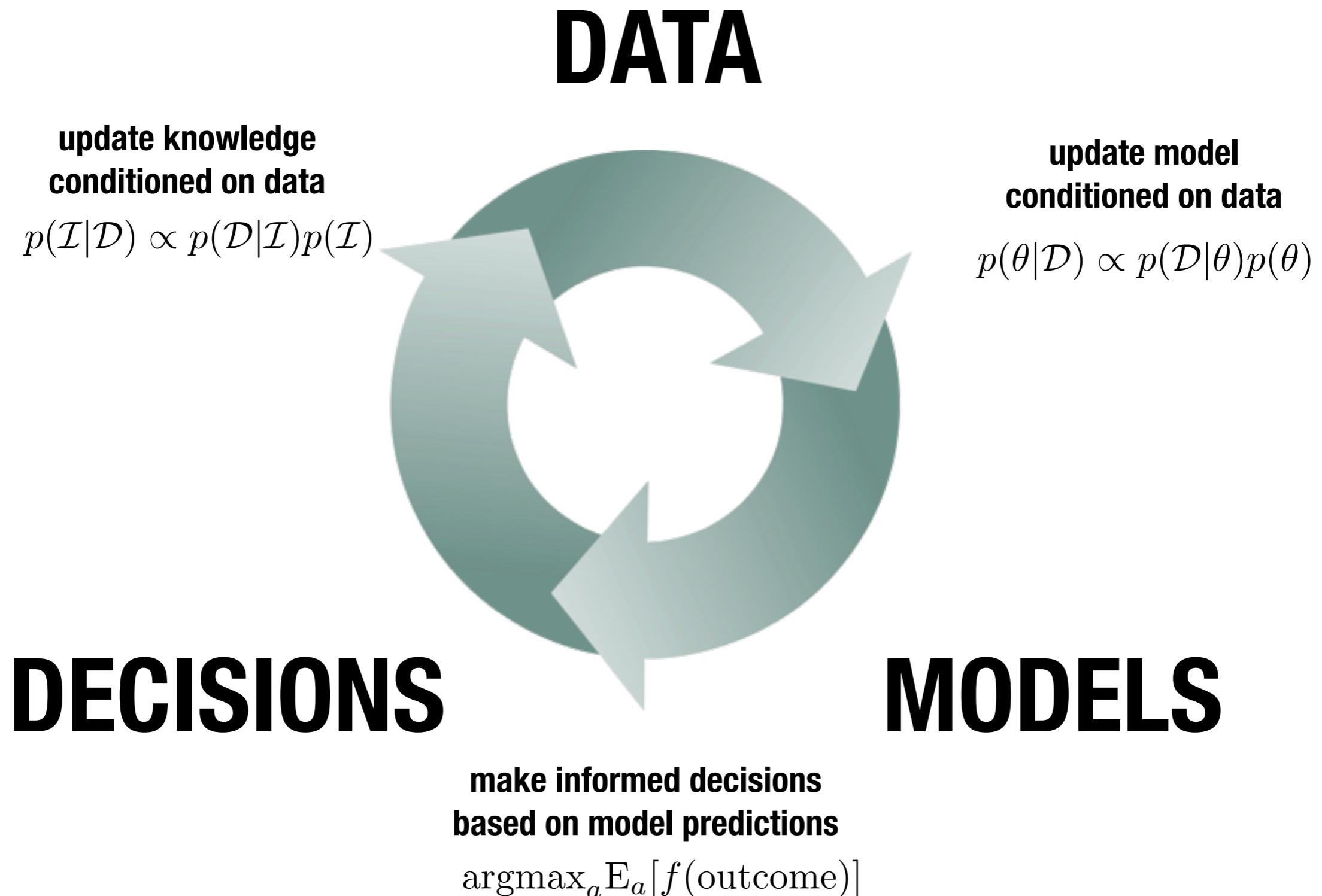


DECISIONS

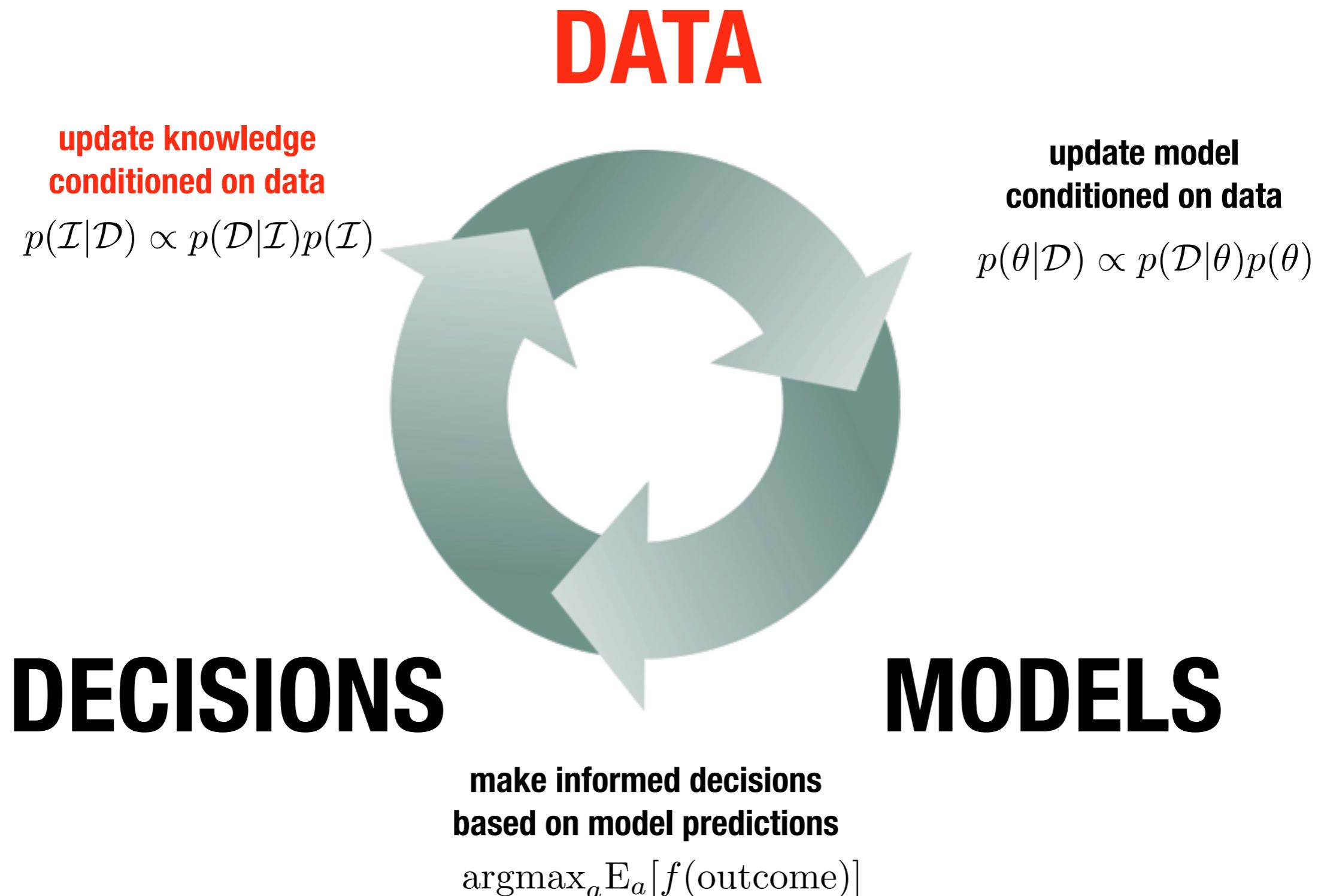
MODELS

How can we do all this without losing our minds?

Bayesian inference can drive progress



Bayesian inference can drive progress

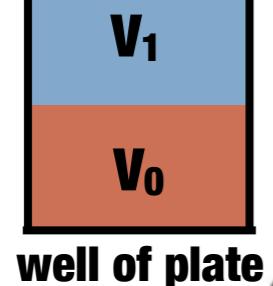


What is experimental error and where does it come from?

fluorescence measurement



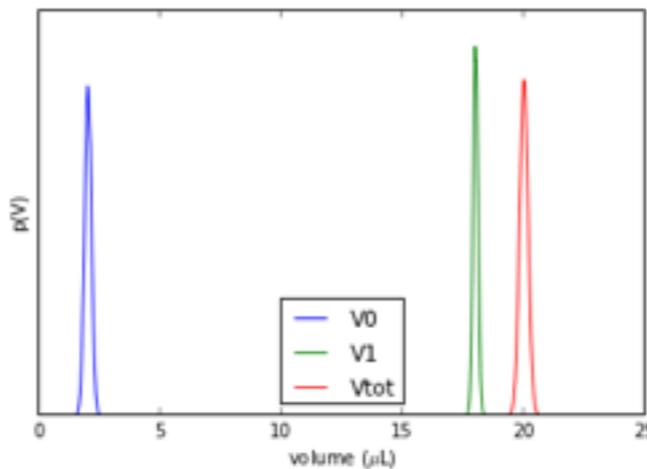
assay mix
compound



$$\begin{aligned}V_0 &= 2 \mu\text{L} \\V_1 &= 18 \mu\text{L} \\K_d &= 100 \text{ nM} \\C_0 &= 10 \text{ mM} \\C_1 &= 1 \mu\text{M}\end{aligned}$$

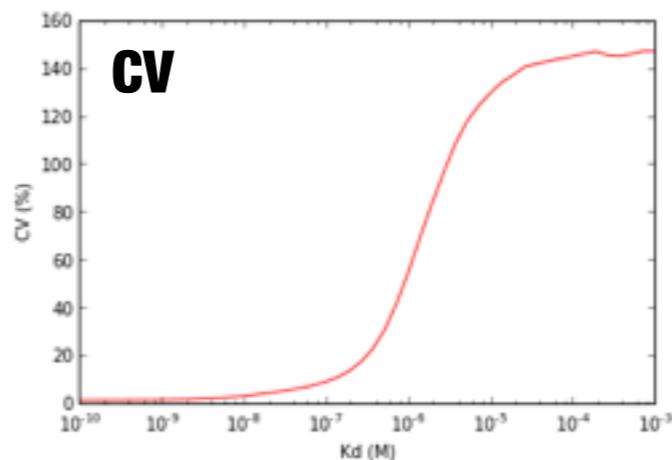


finite pipetting precision

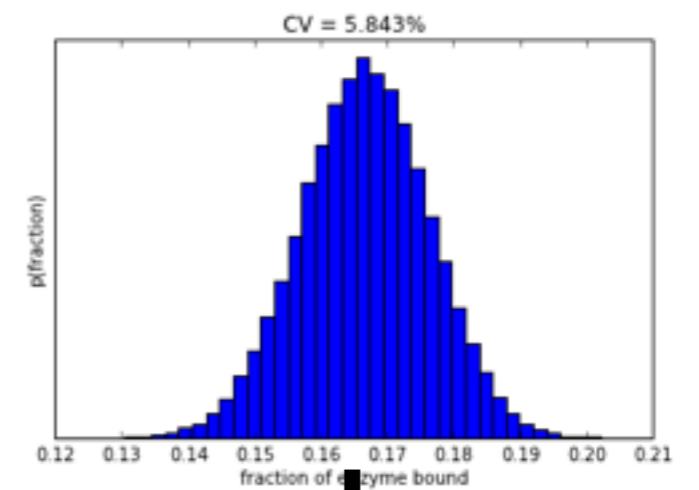


BIOMEK FX® PIPETTING PERFORMANCE SPECIFICATIONS				
SPAN-B SYSTEMS		Tip Types	Accuracy ± %	Precision ε %
Transfer Volume	Span-B Syringe Volume	P20, Fixed 60 mm	5	10
0.5 μL	250 μL	P20, P50, Fixed 60 mm	3	7
1 μL	250, 500, 1000 μL	P20, P50, Fixed 60 mm	3	5
5 μL	250, 500, 1000 μL	P20, Fixed 60 mm	3	5
10 μL	500 μL	P50, P250	3	5
50 μL	500 μL	P250	3	5
100 μL	500 μL	P250, P1000	3	5

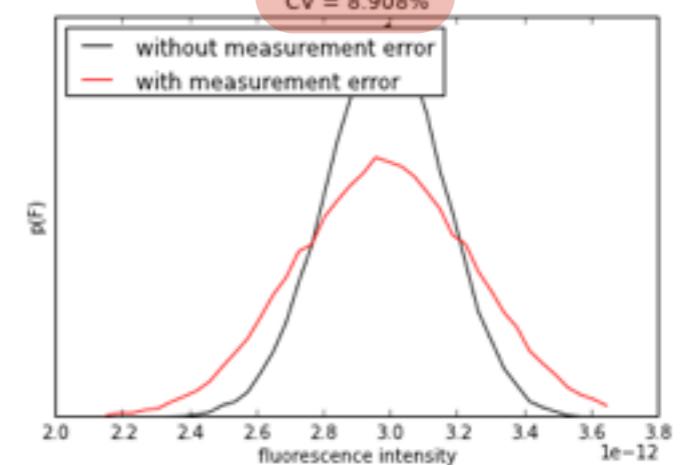
Is the expected CV small enough to be useful?



variation in protein:ligand complex concentration



CV = 8.908%



signal broadening due to measurement error

Modeling experimental error can be a valuable exercise: Biomek dilution series vs Echo acoustic dispensing

In the Pipeline

[« Aveo Gets Bad News on Tivozanib | Main | The Medical Periodic Table »](#)

May 3, 2013

Drug Assay Numbers, All Over the Place

Posted by Derek

There's a [truly disturbing paper](#) out in PLoS ONE with potential implications for a lot of assay data out there in the literature. The authors are looking at the results of biochemical assays as a function of how the compounds are dispensed in them, pipet tip versus **acoustic**, which is the sort of idea that some people might roll their eyes at. But people who've actually done a lot of biological assays may well feel a chill at the thought, because this is just the sort of you're-kidding variable that can make a big difference.

http://pipeline.corante.com/archives/2013/05/03/drug_assay_numbers_all_over_the_place.php

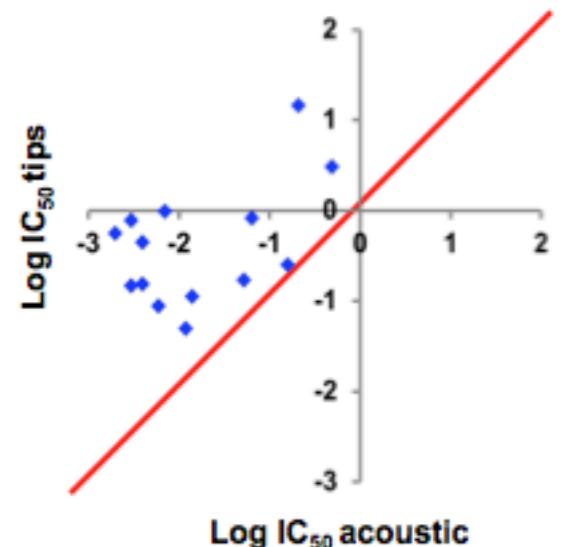
OPEN  ACCESS Freely available online



Dispensing Processes Impact Apparent Biological Activity as Determined by Computational and Statistical Analyses

Sean Ekins^{1*}, Joe Olechno², Antony J. Williams³

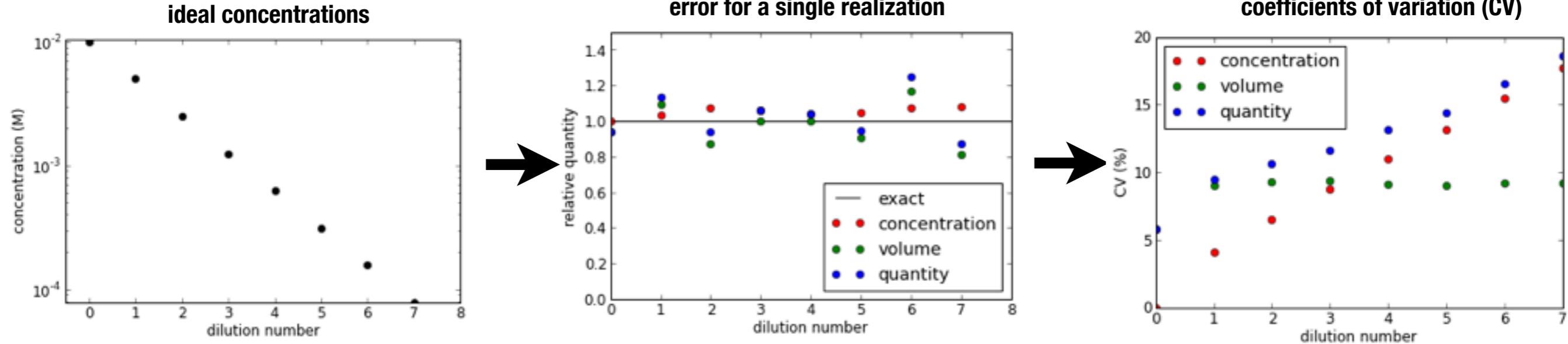
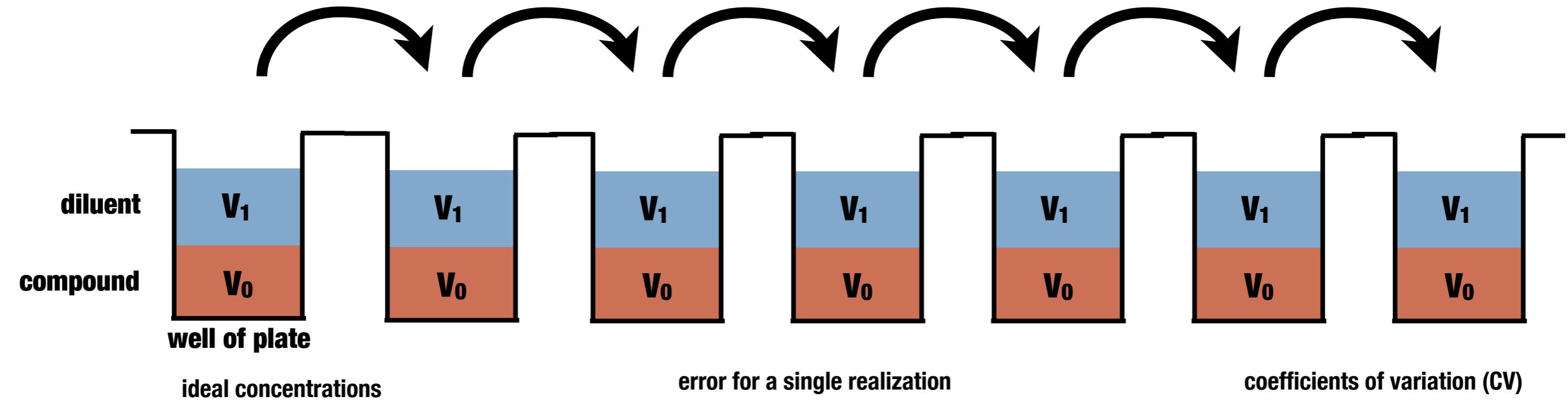
¹ Collaborations in Chemistry, Fuquay-Varina, North Carolina, United States of America, ² Labcyte Inc., Sunnyvale, California, United States of America, ³ Royal Society of Chemistry, Wake Forest, North Carolina, United States of America



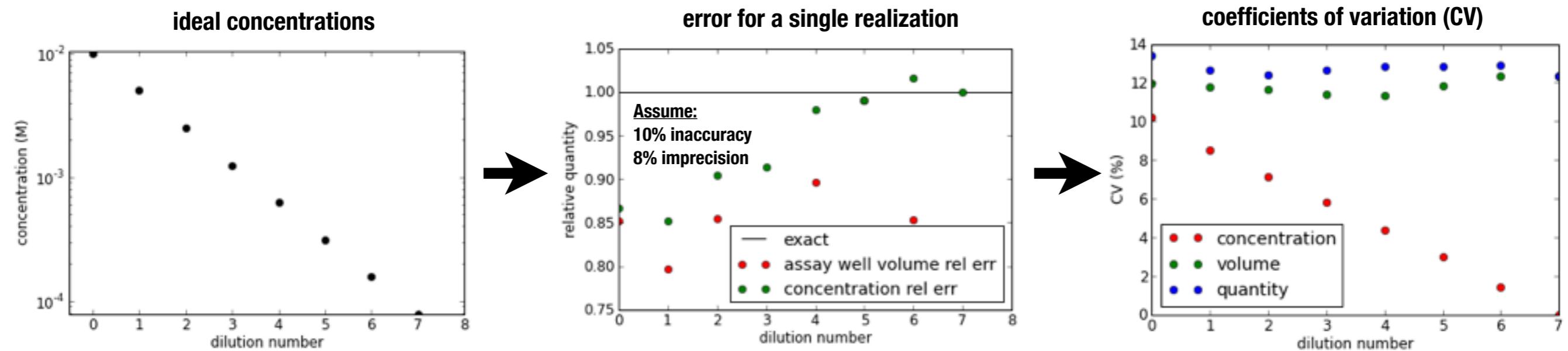
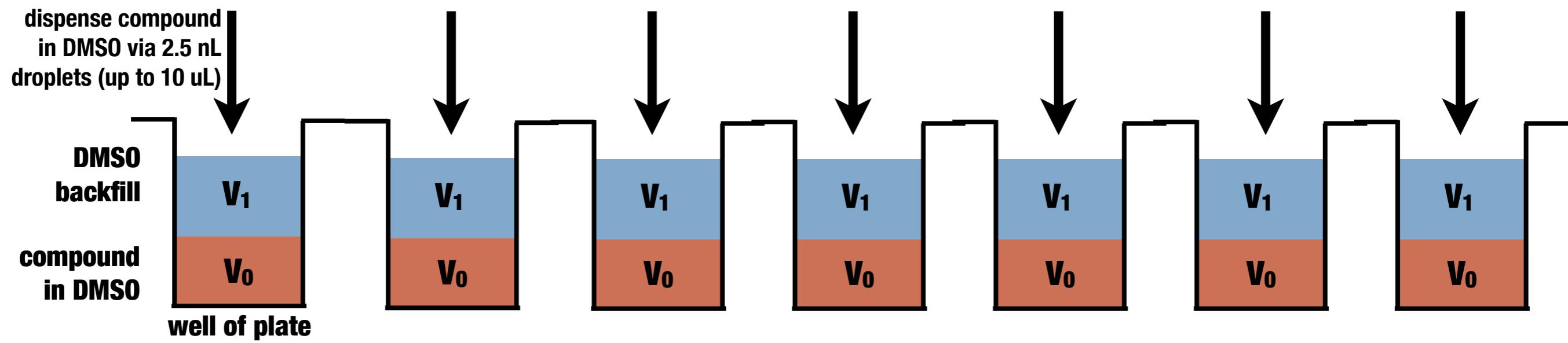
Elkins et al. PLoS One 8:e62325, 2013.

Is this just a result of pipetting accuracy differences?
Let's model the experiment to find out.

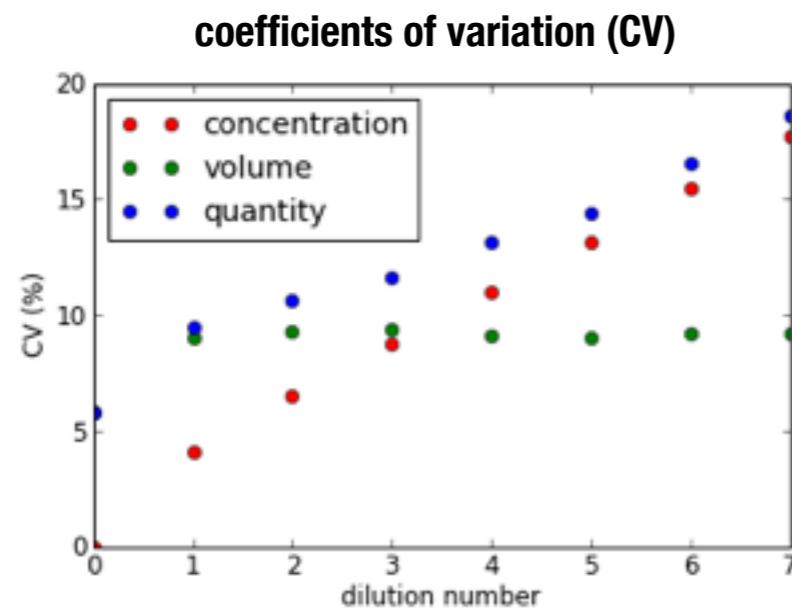
Modeling experimental error can be a valuable exercise: Preparing a dilution series via liquid-handling robot



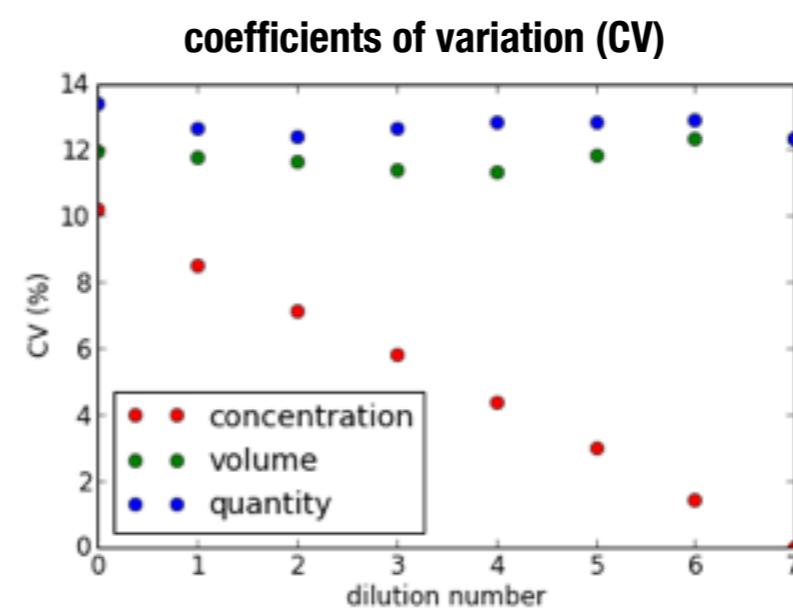
Modeling experimental error can be a valuable exercise: Preparing a dilution series via acoustic dispensing



Modeling experimental error can be a valuable exercise: Comparison of tip-based and acoustic dispensing



tip-based



acoustic

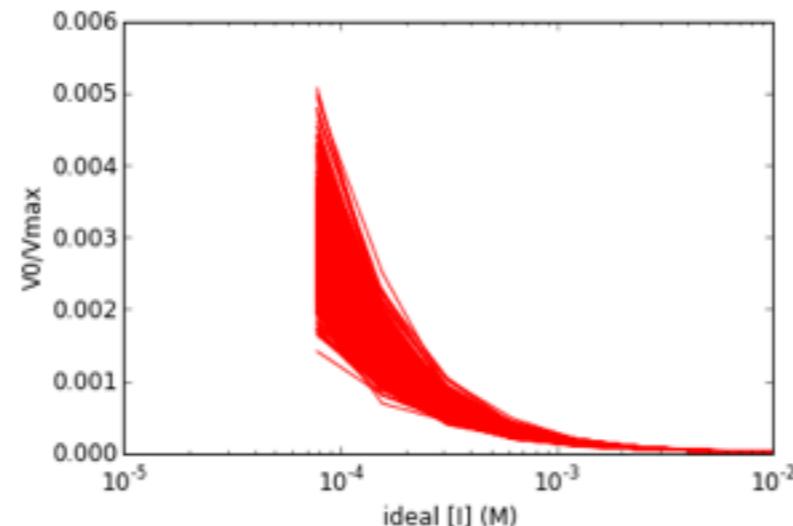
Modeling experimental error can be a valuable exercise: Assay model

↓
Dispense into assay plate:
2 uL compound dilution
10 uL enzyme assay mix



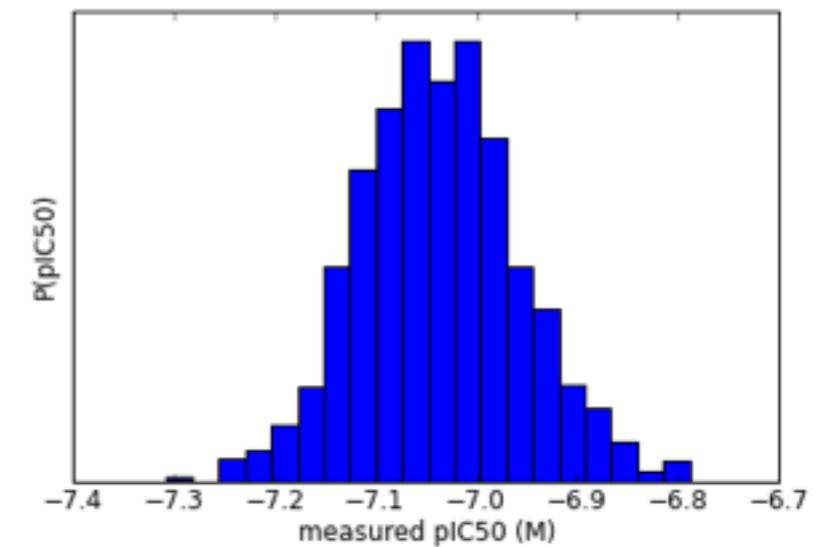
Assume for EphB4 assay:
Michaelis-Menten kinetics
competitive inhibition
[ATP] = 4 uM
[EphB4] = 6 uM
Km(ATP) = 1.71 uM

measure initial reaction velocity



fit it to get IC50

distribution of fit pIC50s



Modeling experimental error can be a valuable exercise: Biomek dilution series vs Echo acoustic dispensing

In the Pipeline

[« Aveo Gets Bad News on Tivozanib | Main | The Medical Periodic Table »](#)

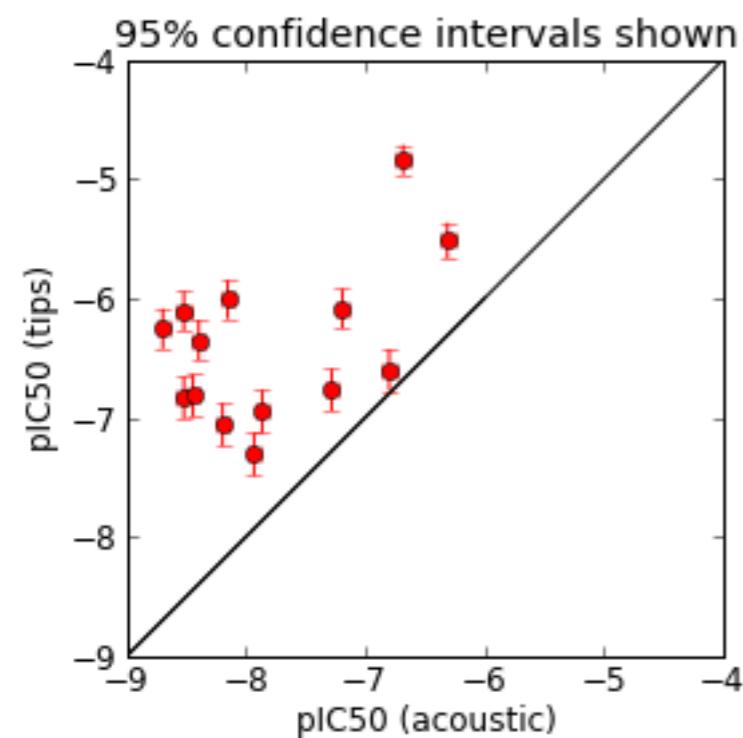
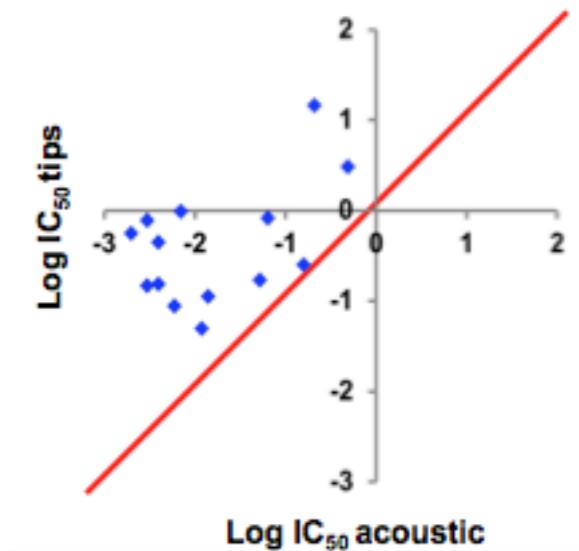
May 3, 2013

Drug Assay Numbers, All Over the Place

Posted by Derek

There's a [truly disturbing paper](#) out in PLoS ONE with potential implications for a lot of assay data out there in the literature. The authors are looking at the results of biochemical assays as a function of how the compounds are dispensed in them, pipet tip versus **acoustic**, which is the sort of idea that some people might roll their eyes at. But people who've actually done a lot of biological assays may well feel a chill at the thought, because this is just the sort of you're-kidding variable that can make a big difference.

http://pipeline.corante.com/archives/2013/05/03/drug_assay_numbers_all_over_the_place.php



Elkins et al. PLoS One 8:e62325, 2013.

OK, that's not sufficient to explain the discrepancy...
What are we missing?

Modeling experimental error can be a valuable exercise: Dilution is a big problem with liquid-based pipetting

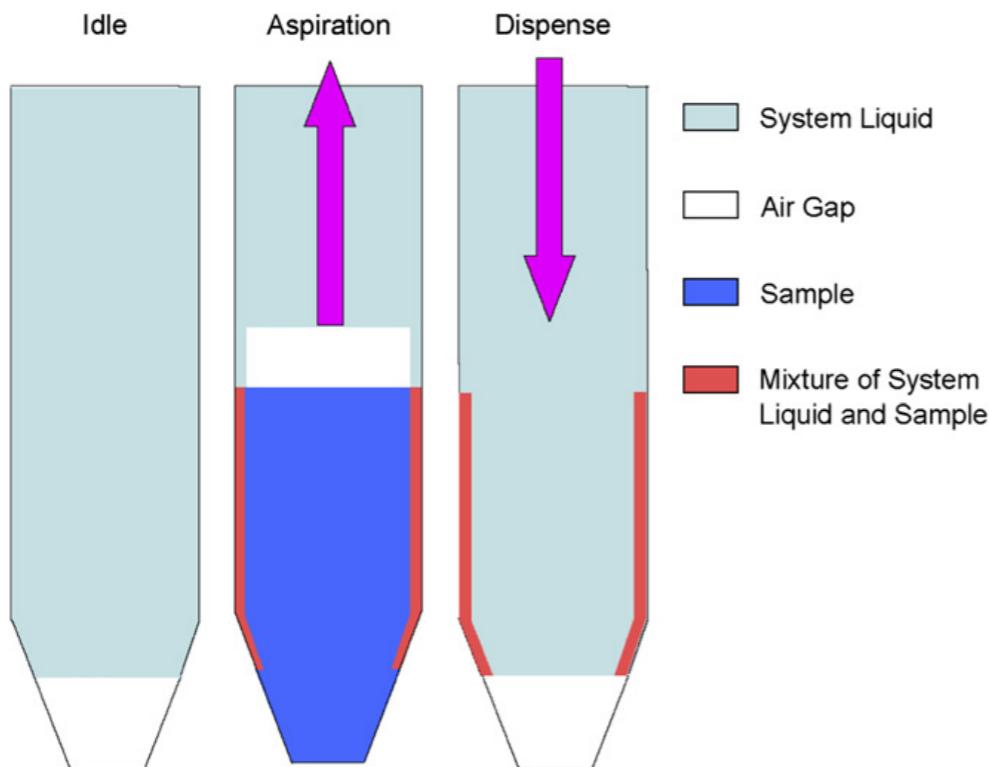


Figure 1. Schematic diagram of the dilution effect.

Table I. Comparison between the MVS method and the gravimetry method—Tecan ALH using water liquid class

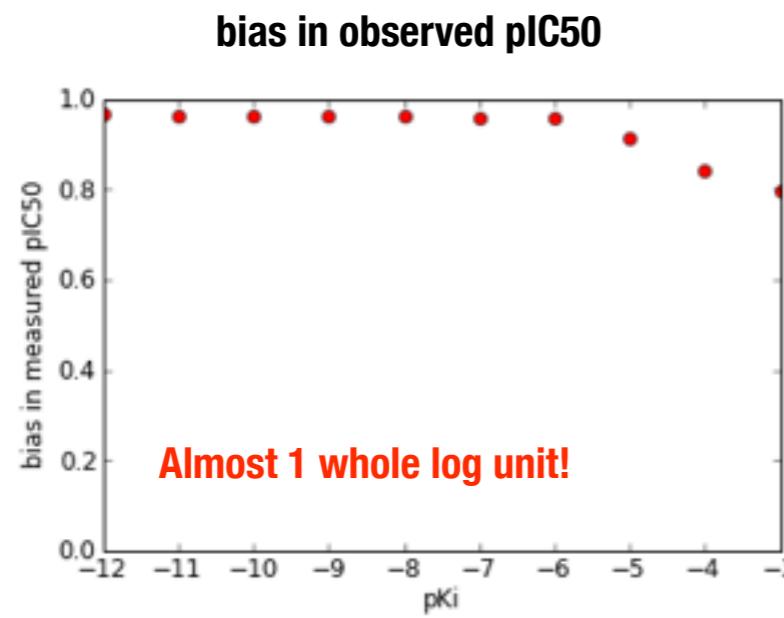
	MVS (μL)	Gravimetry (μL)	
Target volume	20	200	20
Mean volume (μL)	18.74	190.08	20.15
Inaccuracy (%)	-6.30	-4.96	0.75
StDev	0.22	1.74	0.6
CV ($n = 96$) (%)	1.17	0.92	1.94
			0.59

These solutions were delivered by the Tecan ALH using an aspirate/dispense protocol based on a water liquid class. Inaccuracy corresponds to $100 \times (\text{Observed volume} - \text{Target volume})/\text{Target volume}$. Coefficient of variation (CV) corresponds to $100 \times \text{StDev}/\text{mean}$, where StDev is the standard deviation.

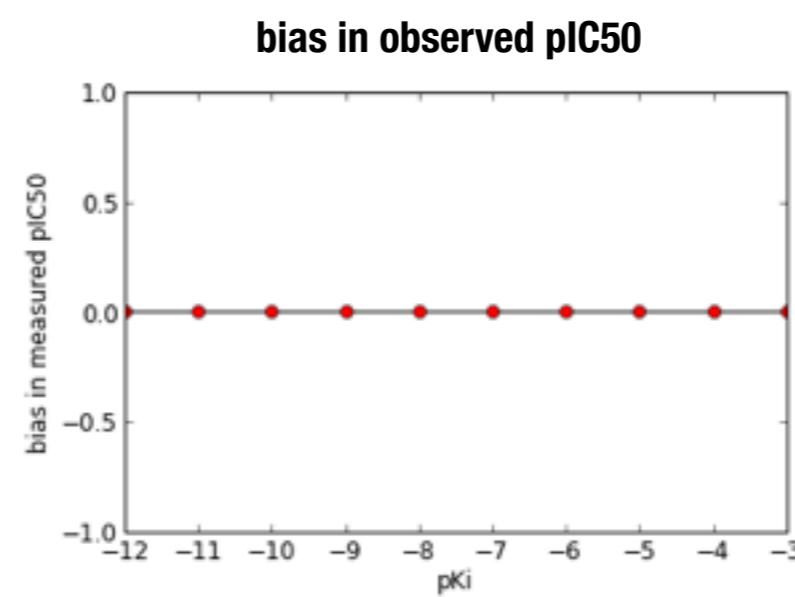


Gu and Deng. JALA 12:355, 2007 (BMS)
Dong, Ouyang, Liu, and Jemal. JALA 11:60, 2006. (BMS)

Modeling experimental error can be a valuable exercise: Comparison of tip-based and acoustic dispensing



tip-based



acoustic

Modeling experimental error can be a valuable exercise: Biomek dilution series vs Echo acoustic dispensing

In the Pipeline

[« Aveo Gets Bad News on Tivozanib | Main | The Medical Periodic Table »](#)

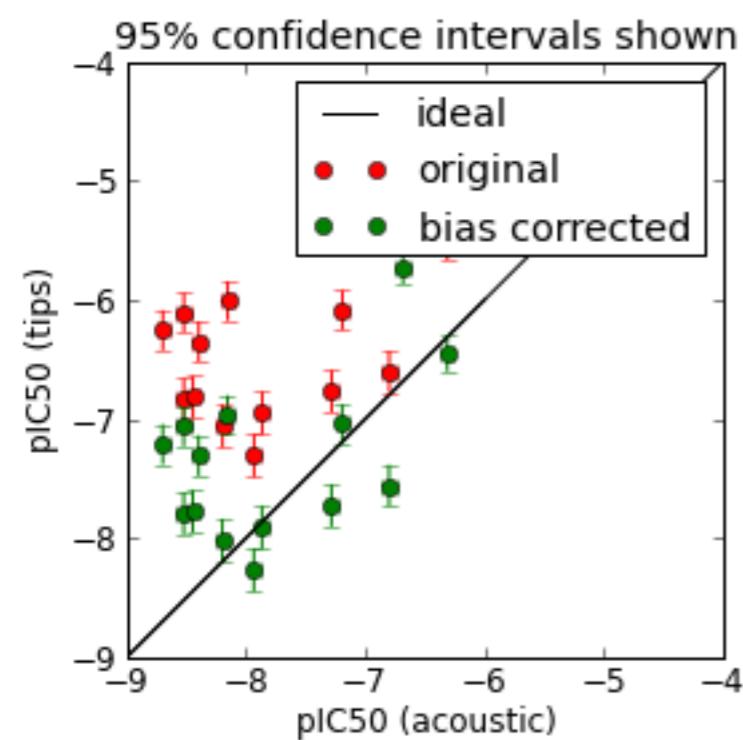
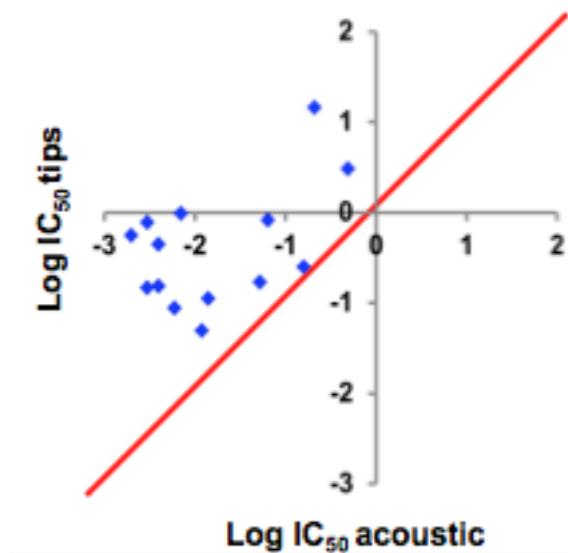
May 3, 2013

Drug Assay Numbers, All Over the Place

Posted by Derek

There's a [truly disturbing paper](#) out in PLoS ONE with potential implications for a lot of assay data out there in the literature. The authors are looking at the results of biochemical assays as a function of how the compounds are dispensed in them, pipet tip versus **acoustic**, which is the sort of idea that some people might roll their eyes at. But people who've actually done a lot of biological assays may well feel a chill at the thought, because this is just the sort of you're-kidding variable that can make a big difference.

http://pipeline.corante.com/archives/2013/05/03/drug_assay_numbers_all_over_the_place.php



Elkins et al. PLoS One 8:e62325, 2013.

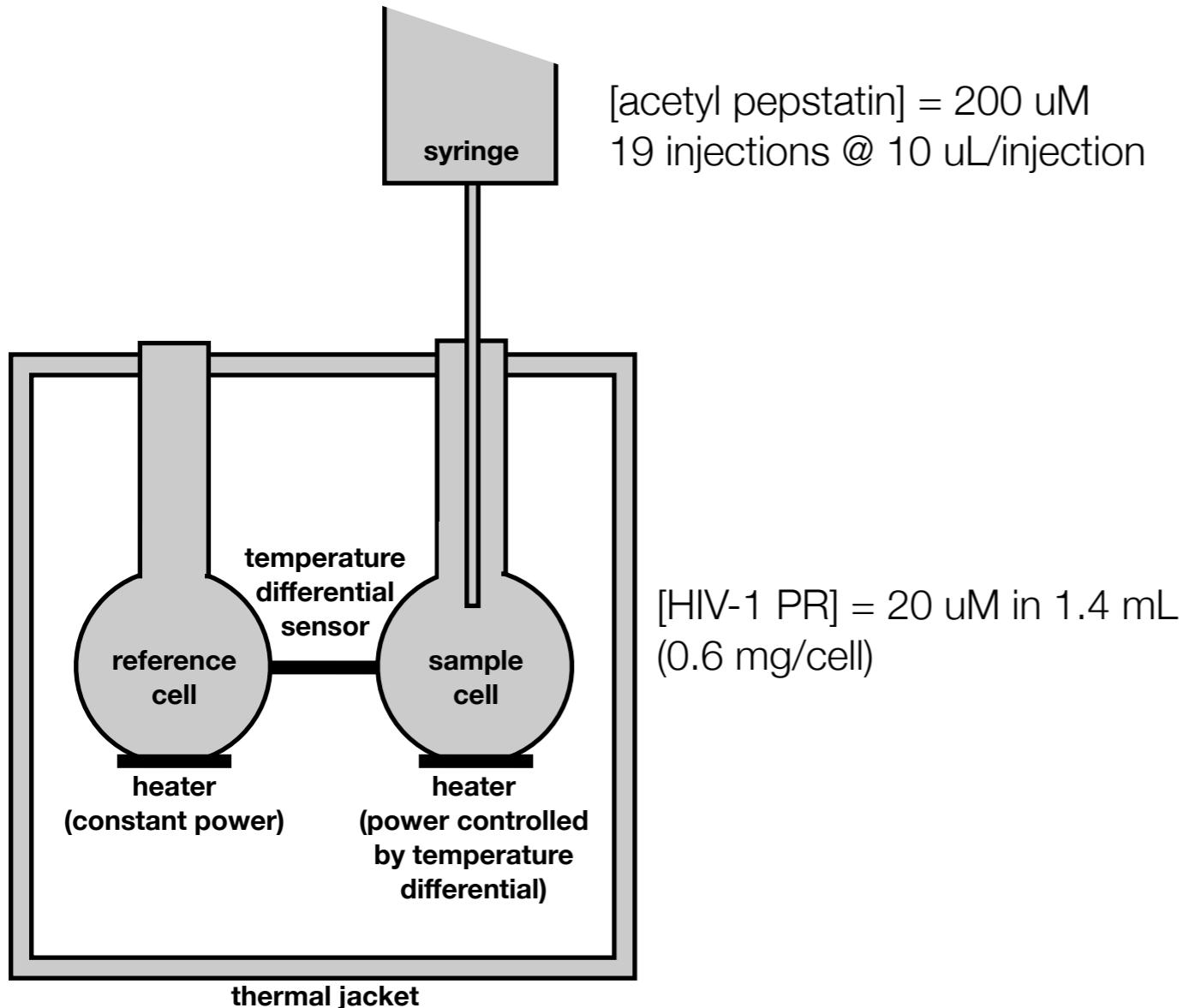
OK, that's all well and good, but how can we use Bayesian techniques to estimate the uncertainty in experimental data?

**I'm glad you asked!
I'll show you!**

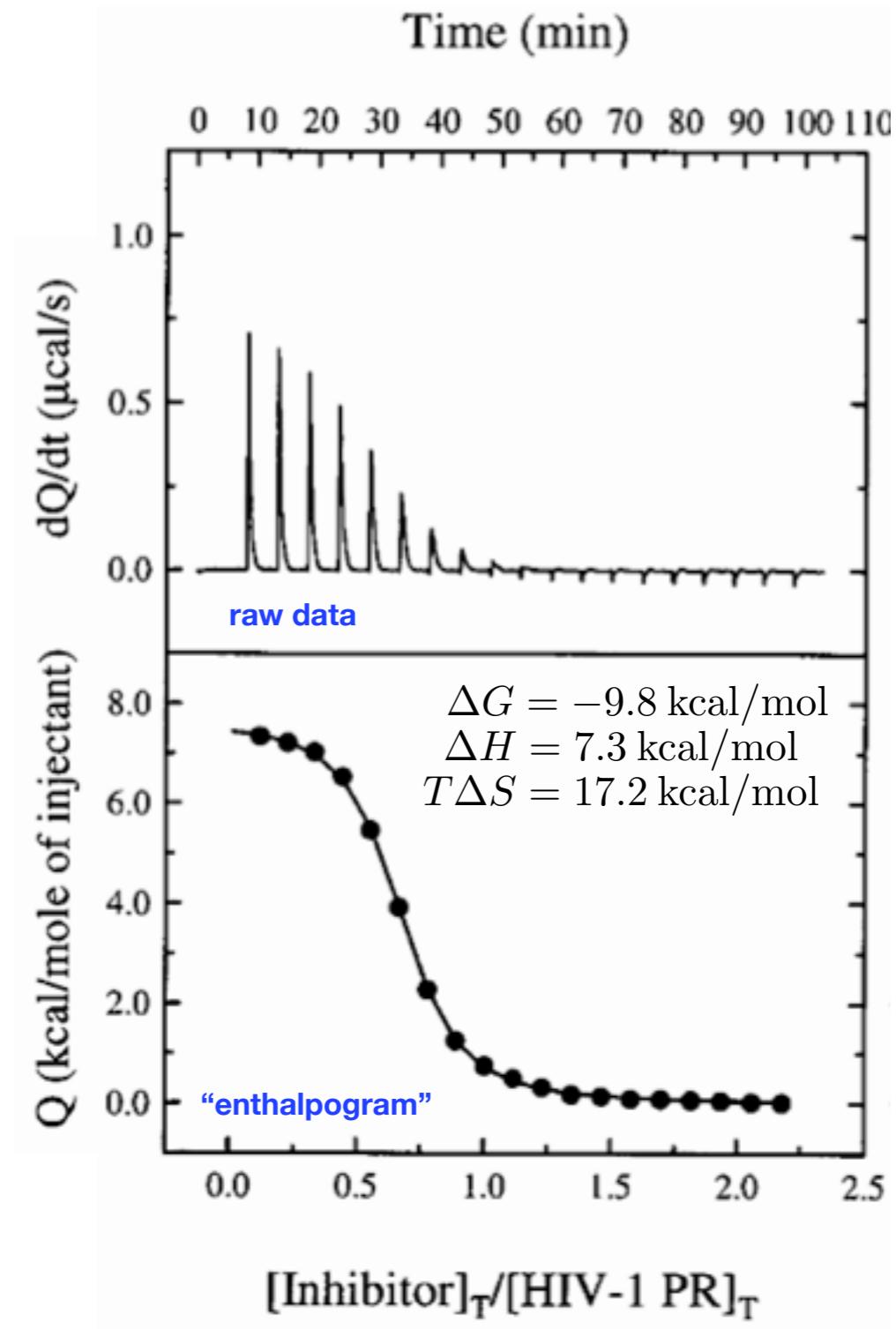


Isothermal titration calorimetry (ITC) can simultaneously interrogate free energies and enthalpies of binding

[illustrative experiment]



[HIV-1 PR] = 20 uM in 1.4 mL
(0.6 mg/cell)



(Note that some reactions have no measurable change in heat, and are not measurable by ITC.)

Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

How reliable is calorimetric data, really?

A test of variation among truly independent experiments

The ABRF-MIRG'02 study:

Send identical aliquots of the **same sample** of protein and ligand to 14 core analysis facilities (experts!) and ask them to report the measured ΔG and ΔH by ITC.

The should get the **same answer**, within error.

The reported errors should match the variation among the reported results.

This experiment is almost never repeated because of the large quantity of protein needed for one ITC experiment, and the undesirability of **repeating the experiment from scratch** multiple times.

This is pretty much the only dataset of its kind reported in the literature.

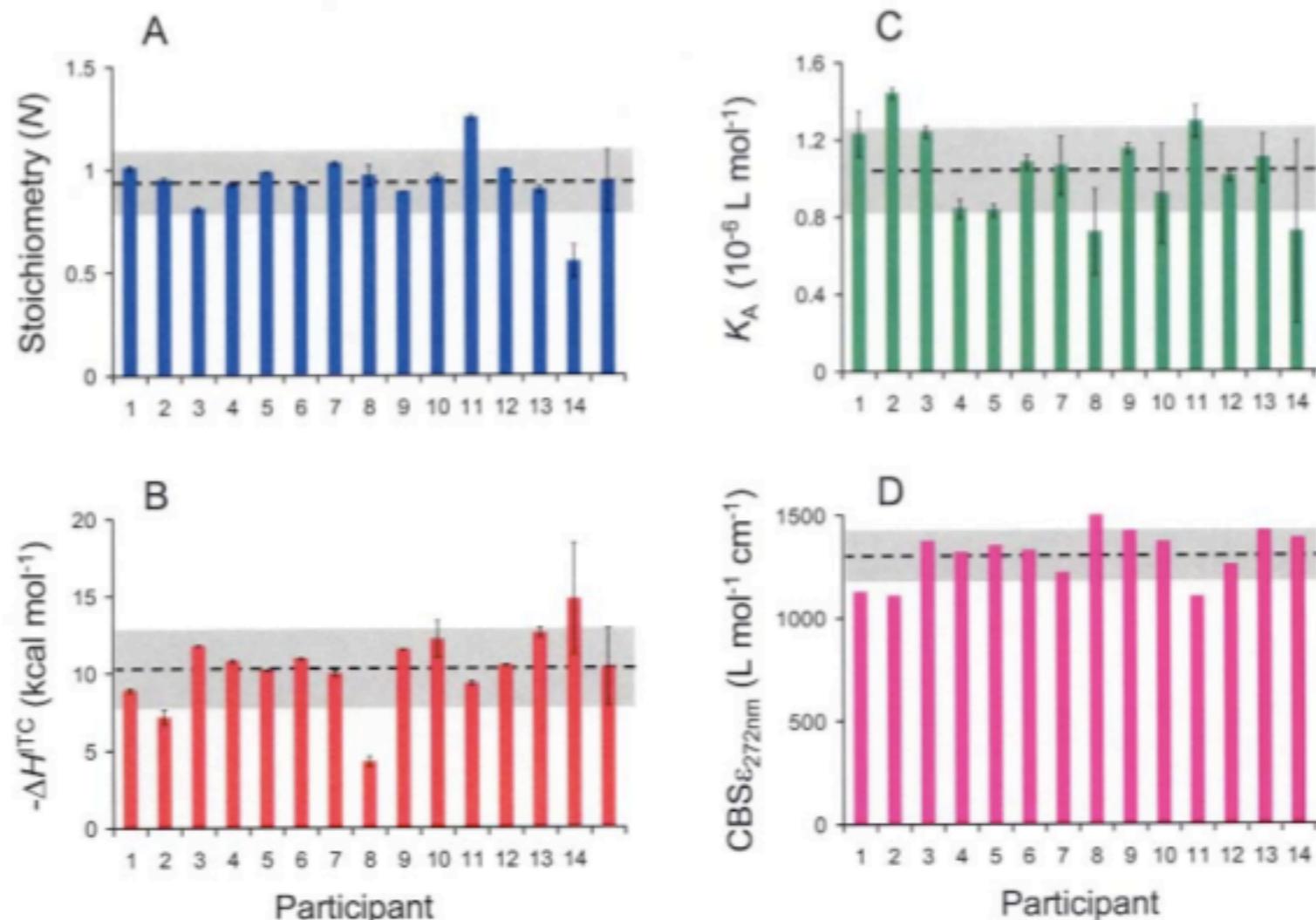


FIGURE 5

ITC characterization of the CBS/CA II interaction (based on Fig. 4 and Table 3). **A:** Stoichiometry (0.94 ± 0.15). **B:** Enthalpy upon binding ($-10.4 \pm 2.5 \text{ kcal mol}^{-1}$). **C:** Affinity [$(1.00 \pm 0.22) \times 10^6 \text{ L mol}^{-1}$]. **D:** molar extinction coefficient for CBS at 272 nm ($1307 \pm 126 \text{ L mol}^{-1} \text{ cm}^{-1}$). Mean values and the standard deviation for 14 determinations are represented by the horizontal dotted lines and gray bands. Error bars denote the standard deviation of nonlinear least squares analysis, except for participants 10 and 14 who reported standard deviations for replicate analyses. No errors were reported for the extinction coefficient determinations.

How reliable is calorimetric data, really? Error is significantly underreported

TABLE 3

Summary of Reported Isothermal Titration Calorimetry Results^a

Participant	Molar Binding Ratio (<i>N</i>)	K_A ($10^{-6} \times L\ mol^{-1}$)	ΔH^{ITC} (kcal mol $^{-1}$)	C-value ^b	Control Titrations
1	1.01 ± 0.01	1.2 ± 0.1	-8.9 ± 0.1	17	CBS into buffer and buffer into CA II. Former subtracted as part of data analysis.
2	0.95 ± 0.01	1.44 ± 0.03	-7.2 ± 0.5	75	None
3	0.81 ± 0.01	1.24 ± 0.03	-11.8 ± 0.02	44	CBS into buffer and buffer into CA II. Former subtracted as part of data analysis.
4	0.929 ± 0.007	0.84 ± 0.05	-10.8 ± 0.1	24	Pilot run.
5	0.987 ± 0.003	0.84 ± 0.03	-10.20 ± 0.04	41	CBS into buffer, subtracted in data analysis.
6	0.921 ± 0.003	1.08 ± 0.04	-10.95 ± 0.05	39	CBS into buffer, subtracted in data analysis.
7	1.03 ± 0.01	1.1 ± 0.2	-10.0 ± 0.2	55	CBS into buffer, subtracted in data analysis.
8	0.97 ± 0.05	0.7 ± 0.2	-4.3 ± 0.3	7.0	CBS into buffer and buffer into CA II. Both subtracted as part of data analysis.
9	0.891 ± 0.002	1.15 ± 0.03	-11.53 ± 0.04	59	CBS into buffer, average value subtracted in analysis.
10 ^c	0.96 ± 0.02	0.9 ± 0.2	-12 ± 1	34	CBS into buffer and buffer into CA II. Average of former used in analysis.
11	1.25 ± 0.01	1.3 ± 0.1	-9.3 ± 0.1	9.0	None
12	1.000 ± 0.003	1.01 ± 0.03	-10.51 ± 0.04	29	Buffer into buffer.
13	0.90 ± 0.02	1.1 ± 0.1	-12.6 ± 0.3	12	CBS into buffer, linear regression subtracted in analysis.
14 ^c	0.55 ± 0.08	0.7 ± 0.3	-15 ± 4	22	CBS into buffer, subtracted in analysis.

Note that observed 20% error in K_A gives only ± 0.1 kcal/mol error in ΔG , while 20% error in ΔH directly impacts ΔH .

This means absolute error in ΔG is actually still small, while absolute error in ΔH is big.

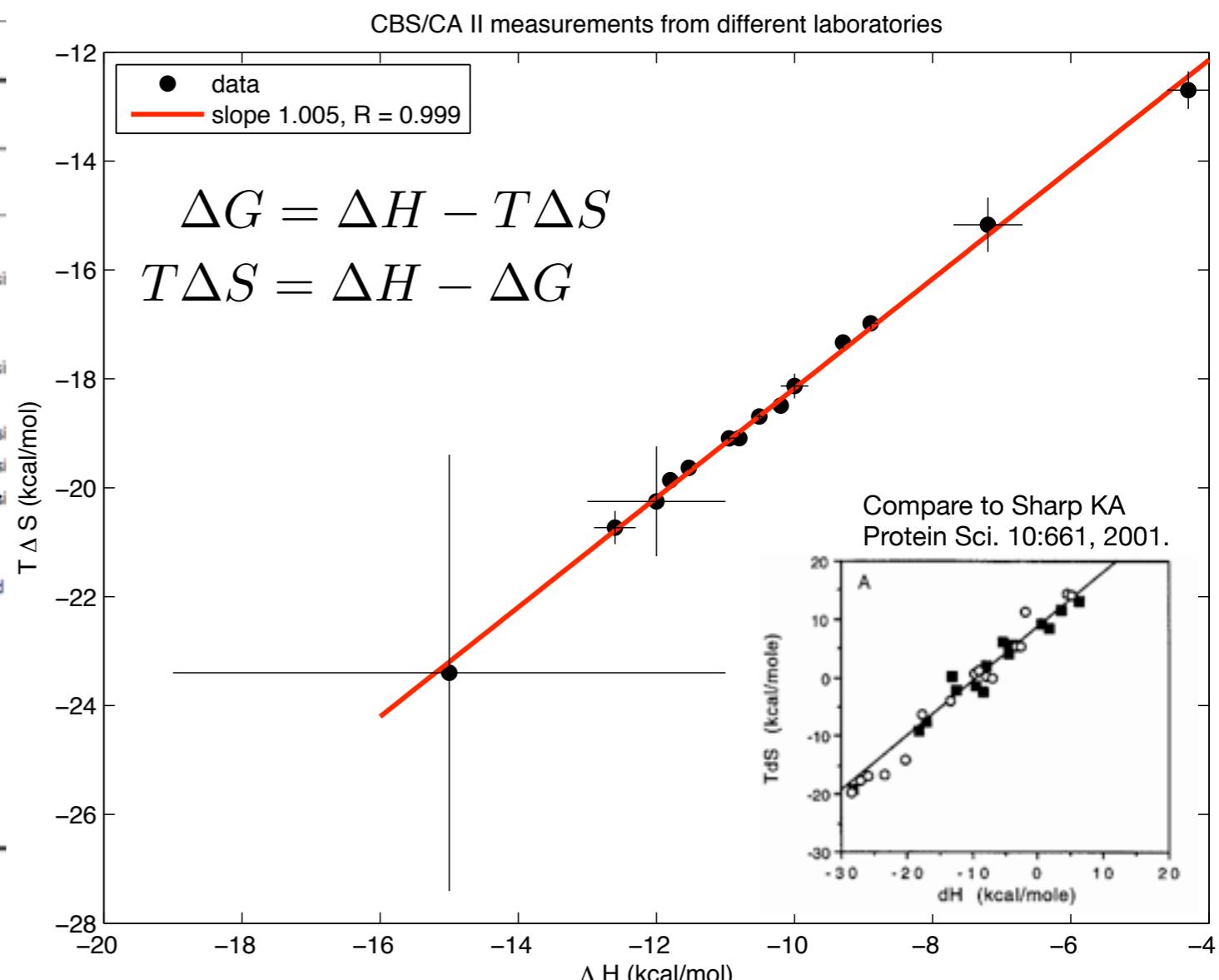
**The reported error bars cannot be trusted.
They're often an order of magnitude (or more!) too small.**

ITC measurements on the **same** system by different laboratories exhibits apparent entropy-enthalpy compensation!

T A B L E 3

Summary of Reported Isothermal Titration Calorimetry Results^a

Participant	Molar Binding Ratio (<i>N</i>)	K_A ($10^{-4} \times \text{L mol}^{-1}$)	ΔH^{ITC} (kcal mol $^{-1}$)	C-value ^b	Control Titrations
1	1.01 ± 0.01	1.2 ± 0.1	-8.9 ± 0.1	17	CBS into buffer and buffer into CA II. Former subtracted as part of data analysis.
2	0.95 ± 0.01	1.44 ± 0.03	-7.2 ± 0.5	75	None
3	0.81 ± 0.01	1.24 ± 0.03	-11.8 ± 0.02	44	CBS into buffer and buffer into CA II. Former subtracted as part of data analysis.
4	0.929 ± 0.007	0.84 ± 0.05	-10.8 ± 0.1	24	Pilot run.
5	0.987 ± 0.003	0.84 ± 0.03	-10.20 ± 0.04	41	CBS into buffer, subtracted in analysis.
6	0.921 ± 0.003	1.08 ± 0.04	-10.95 ± 0.05	39	CBS into buffer, subtracted in analysis.
7	1.03 ± 0.01	1.1 ± 0.2	-10.0 ± 0.2	55	CBS into buffer, subtracted in analysis.
8	0.97 ± 0.05	0.7 ± 0.2	-4.3 ± 0.3	7.0	CBS into buffer, subtracted in analysis. Both subtracted as part of data analysis.
9	0.891 ± 0.002	1.15 ± 0.03	-11.53 ± 0.04	59	CBS into buffer, average value subtracted in analysis.
10 ^c	0.96 ± 0.02	0.9 ± 0.2	-12 ± 1	34	CBS into buffer and buffer into CA II. Average of former used in analysis.
11	1.25 ± 0.01	1.3 ± 0.1	-9.3 ± 0.1	9.0	None
12	1.000 ± 0.003	1.01 ± 0.03	-10.51 ± 0.04	29	Buffer into buffer.
13	0.90 ± 0.02	1.1 ± 0.1	-12.6 ± 0.3	12	CBS into buffer, linear regression subtracted in analysis.
14 ^c	0.55 ± 0.08	0.7 ± 0.3	-15 ± 4	22	CBS into buffer, subtracted in analysis.



Myszka et al. J. Biomol. Tech. 14:247, 2003.

How can these plots be real evidence of compensation if we can generate the same plot from different measurements on same system?

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D} data

θ model parameters

$p(\theta|\mathcal{D})$ posterior

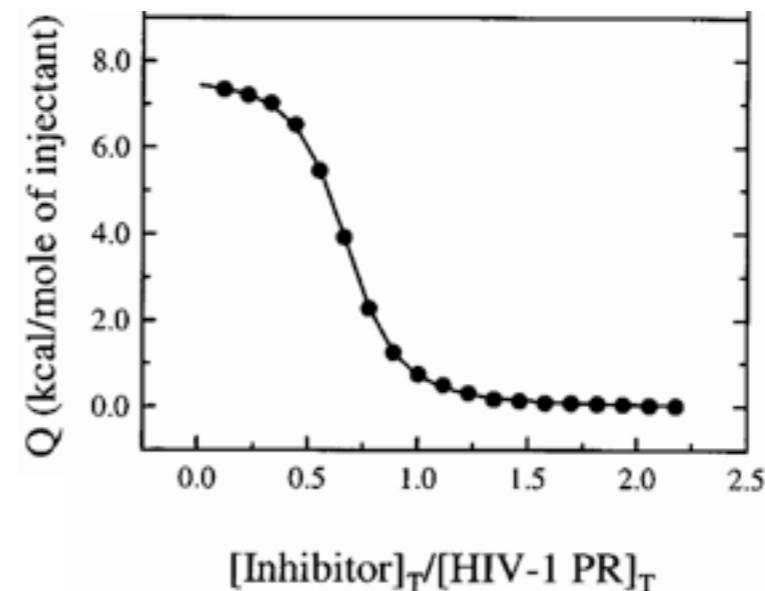
$p(\mathcal{D}|\theta)$ sampling distribution (model)

$p(\theta)$ prior

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D}	data	$\mathcal{D} = \{q_1, q_2, \dots, q_N\}$	measurements of evolved heat
θ	model parameters		
$p(\theta \mathcal{D})$	posterior		
$p(\mathcal{D} \theta)$	sampling distribution (model)		
$p(\theta)$	prior		



Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D}	data	$\theta = \{\Delta G, \Delta H, T\Delta S, \Delta H_0\}$	thermodynamic parameters
θ	model parameters		
$p(\theta \mathcal{D})$	posterior	$\Delta G = -kT \ln K_a$	free energy of binding
$p(\mathcal{D} \theta)$	sampling distribution (model)	ΔH	enthalpy of binding
$p(\theta)$	prior	$T\Delta S$	entropic contribution to binding
		ΔH_0	heat of dilution
		... any additional parameters ...	

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D}	data	$p(\mathcal{D} \theta) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(q_n - q_n^*)^2}{2\sigma^2}\right]$	Gaussian error model
θ	model parameters		
$p(\theta \mathcal{D})$	posterior	q_n	measured heat of injection n
$p(\mathcal{D} \theta)$	sampling distribution (model)	q_n^*	true heat of injection n
$p(\theta)$	prior	σ	std dev of error in measured heat (nuisance parameter)

$$P + L \xrightarrow{\Delta H} PL$$

$$q_n^* = Q_n - Q_{n-1}$$

$$Q_n = \Delta H \cdot V_n [PL]_n + n\Delta H_0 \quad \text{heat potential}$$

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D}	data	$p(\Delta G, \Delta H, \Delta H_0, \sigma) \propto \sigma^{-1}$	prior on measurement noise
θ	model parameters	$[M]_0^* \sim \mathcal{N}([M]_0, \sigma_M^2)$	prior on cell and syringe concentrations
$p(\theta \mathcal{D})$	posterior	$[L]_s^* \sim \mathcal{N}([L]_s, \sigma_L^2)$	
$p(\mathcal{D} \theta)$	sampling distribution (model)		
$p(\theta)$	prior	$\Delta G, \Delta H, \Delta H_0$	can be of any sign and value
		$\sigma > 0$	scale parameter; can be of any magnitude (Later, could build in some <i>a priori</i> knowledge of instrument error or calibration runs.)

Important points:

- * We don't know the size of the measurement error
 - * We don't know the actual ligand and protein concentrations
- No problem: Just make them **nuisance parameters!**

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$$p(\theta|\mathcal{D}) = (2\pi)^{-N/2}\sigma^{-(N+1)} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (q_n - q_n^*)^2\right] \text{ posterior}$$

\mathcal{D} data
 θ model parameters

$p(\theta|\mathcal{D})$ posterior

$p(\mathcal{D}|\theta)$ sampling distribution (model)

q_n^* nonlinear function of thermodynamic parameters

$p(\theta)$ prior

How can we sample from this awful function?

$$p(\theta|\mathcal{D}) = (2\pi)^{-N/2} \sigma^{-(N+1)} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (q_n - q_n^*)^2 \right] \text{ posterior}$$

pymc is like ‘Bayes in a box’!

```
import pymc

# First, find the maximum a posteriori estimate.
map = pymc.MAP(model)
map.fit(iterlim=20000)

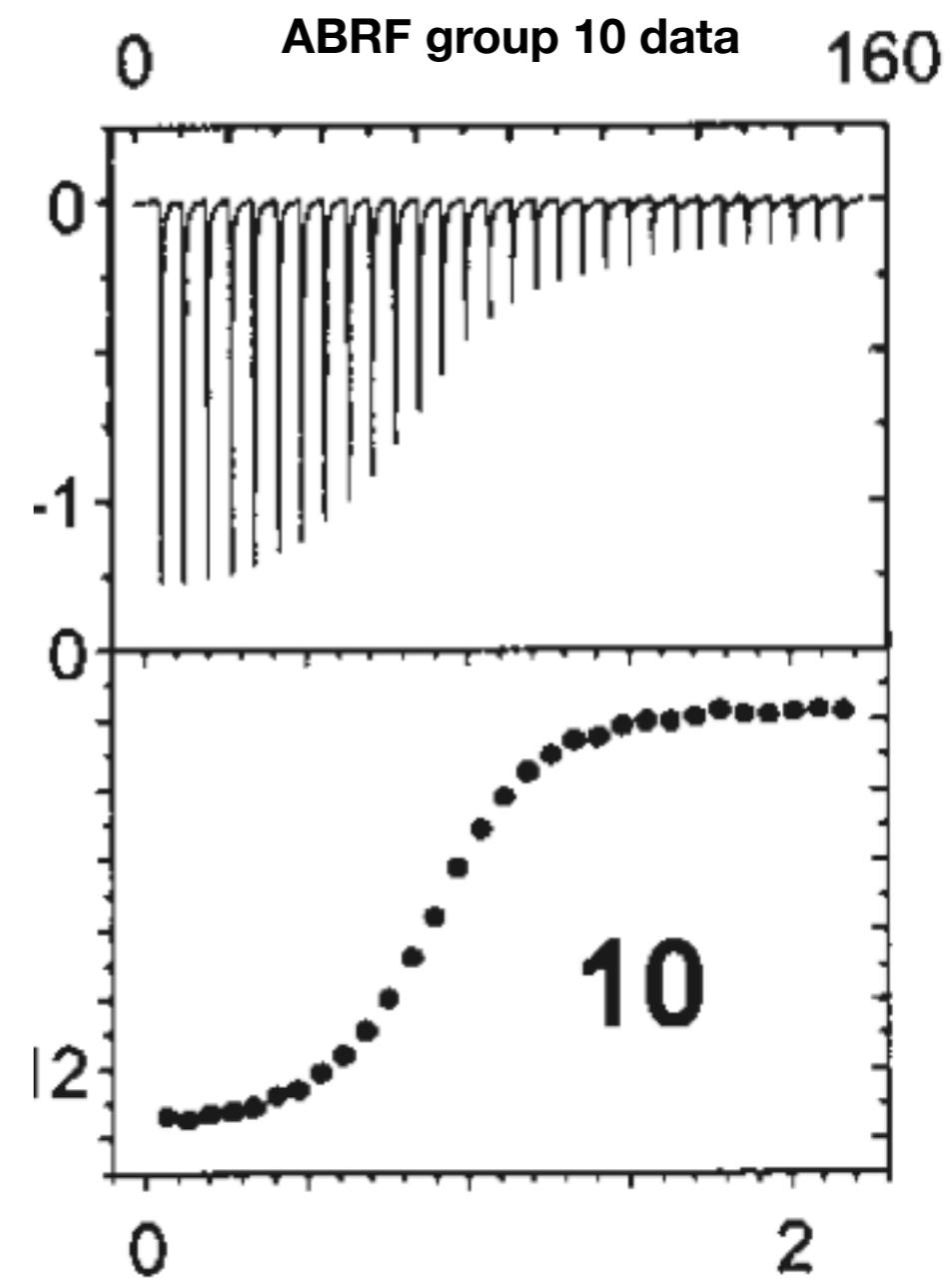
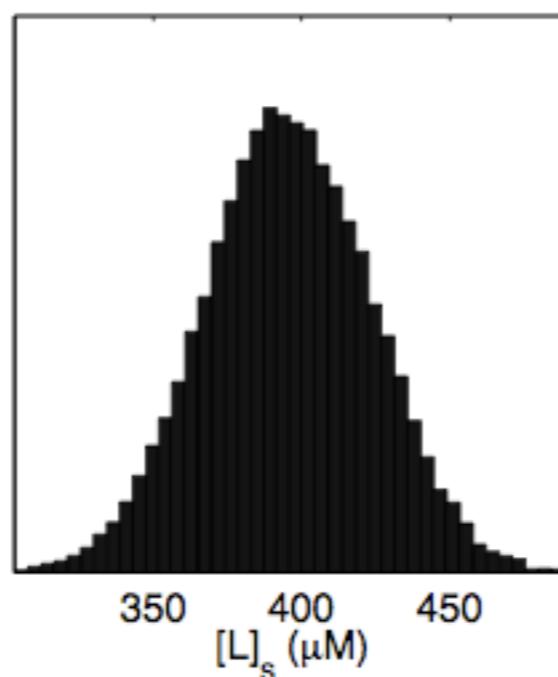
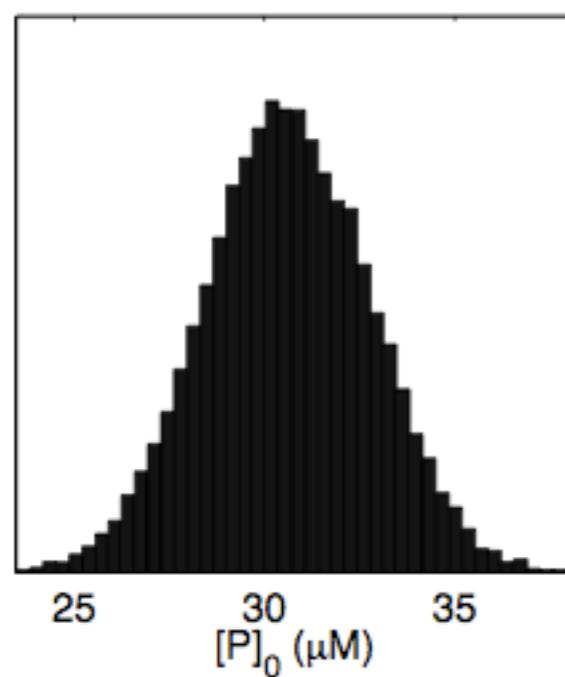
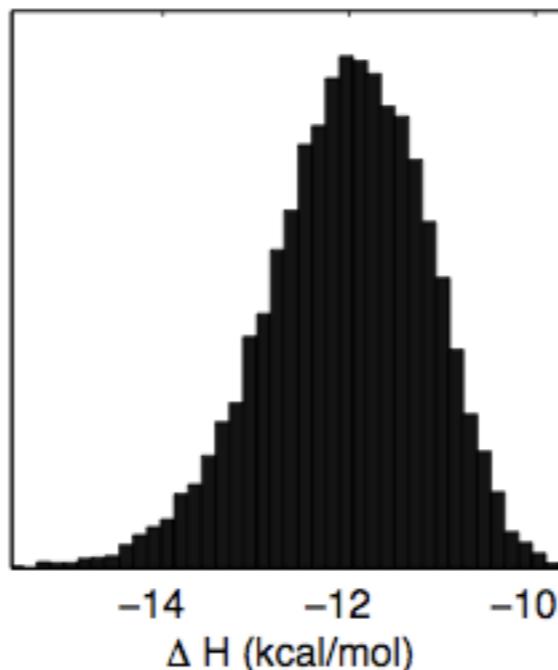
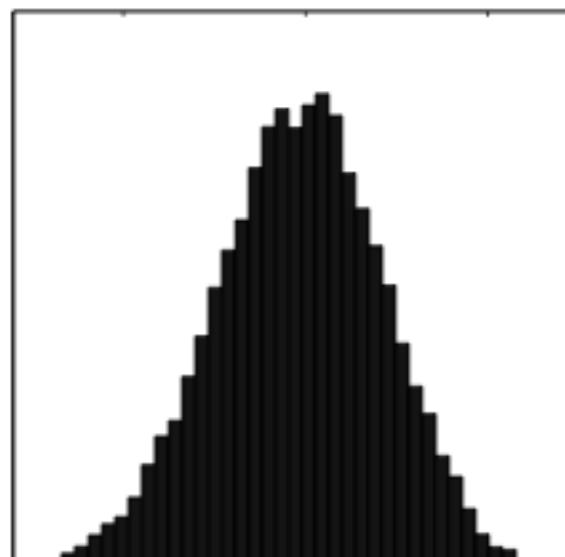
# Now, sample from the posterior.
mcmc = pymc.MCMC(model, db='ram')
mcmc.sample(iter=niters, burn=nburn, thin=nthin, progress_bar=True)
```



<https://pypi.python.org/pypi/pymc>

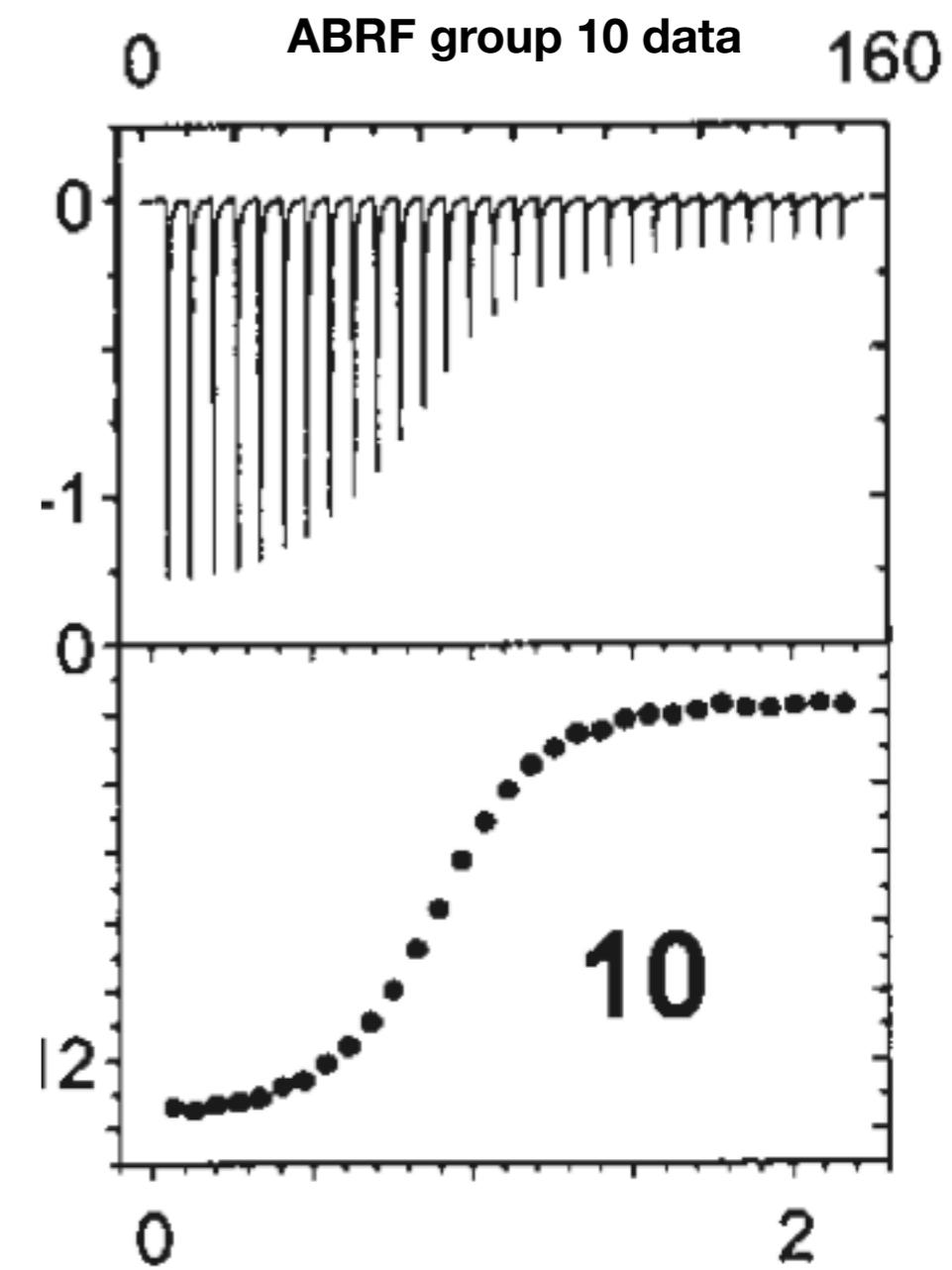
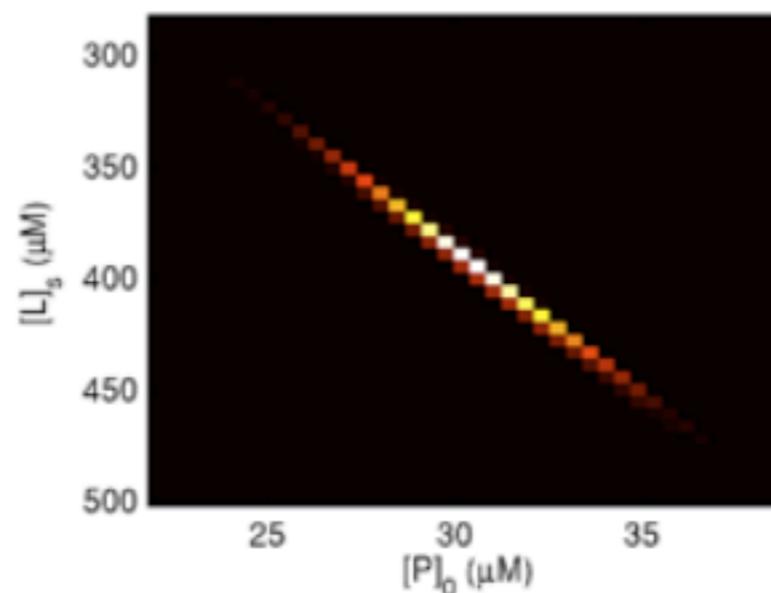
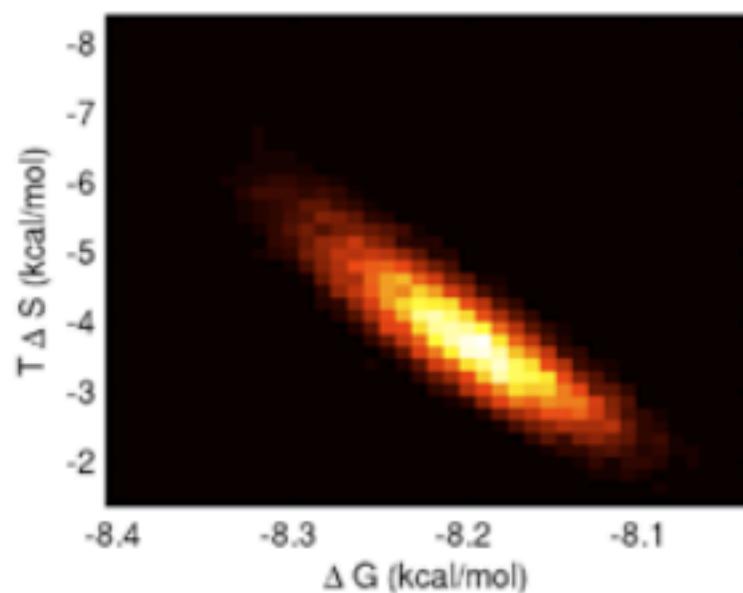
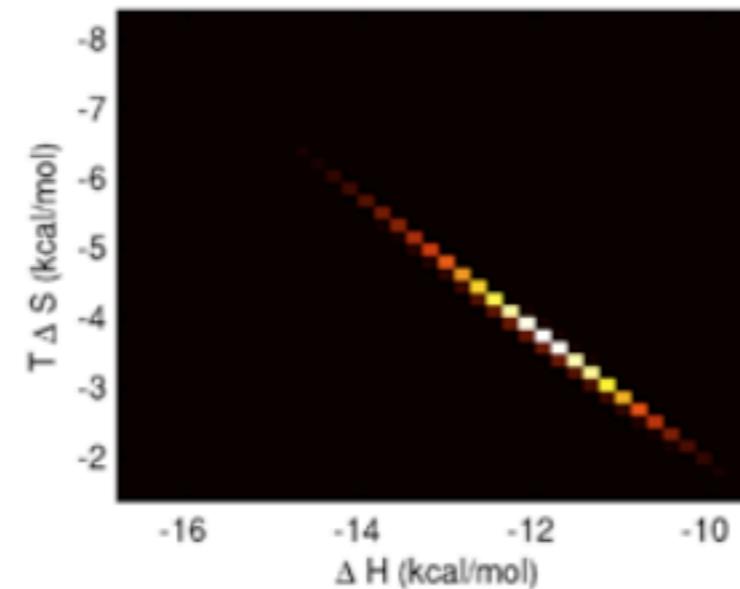
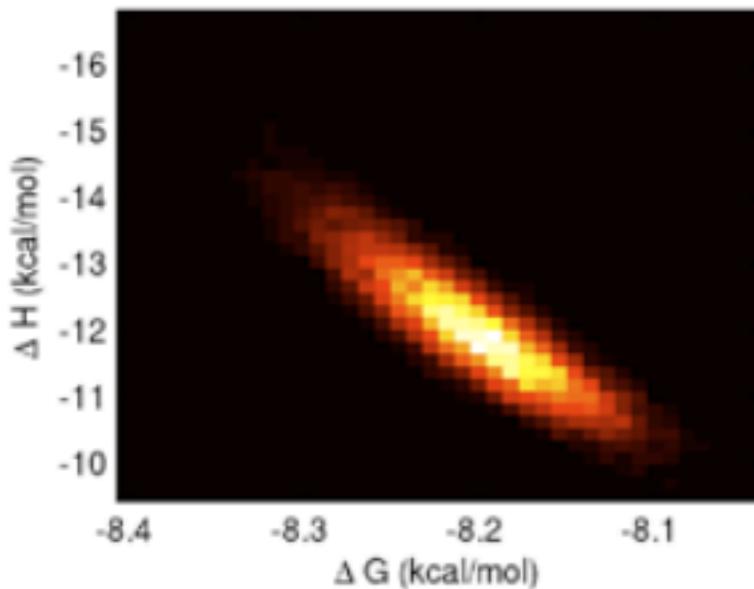
Analysis of ITC experiments: The Bayesian way

Marginal posterior probability distributions describe the uncertainty in each thermodynamic parameter inferred from the data.



Analysis of ITC experiments: The Bayesian way

Joint distribution functions describe the correlated uncertainty between any set of parameters.



Correlation in error in ΔH and $T\Delta S$ explains much of apparent entropy-enthalpy compensation behavior.

Myszka et al. J. Biomol. Tech. 14:247, 2003.

Bayesian approach has numerous advantages

Provides **true posterior joint distribution** of all thermodynamic parameters

Asymmetric confidence intervals and non-normal marginal distributions

Easy to “**plug in**” new binding models.

Make joint inferences from **multiple experiments**

Instrument parameters can be **conditioned on calibration data** (e.g. standard titrations)

Expected information content of new experiments can be estimated for protocol design

Experimental design:

“Will experiment X give me enough information to make it worthwhile?”

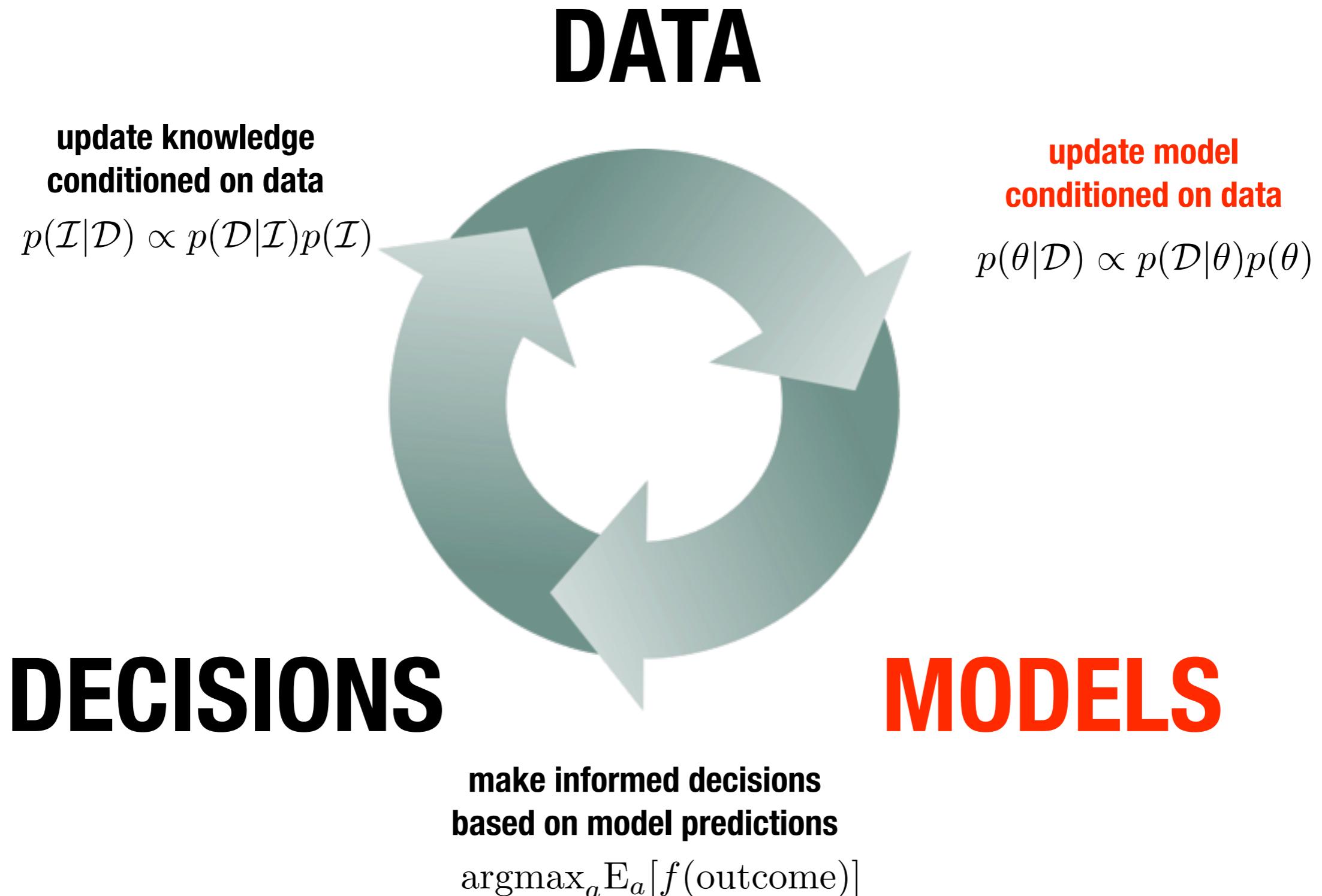
“What is the best experimental design to reduce the uncertainty in Z?”

“Do I have to run a baseline for sample X?”

Code will be available at <http://github.com/choderlab/bayesian-itc>

The universal cycle of progress

Bayesian inference can drive progress



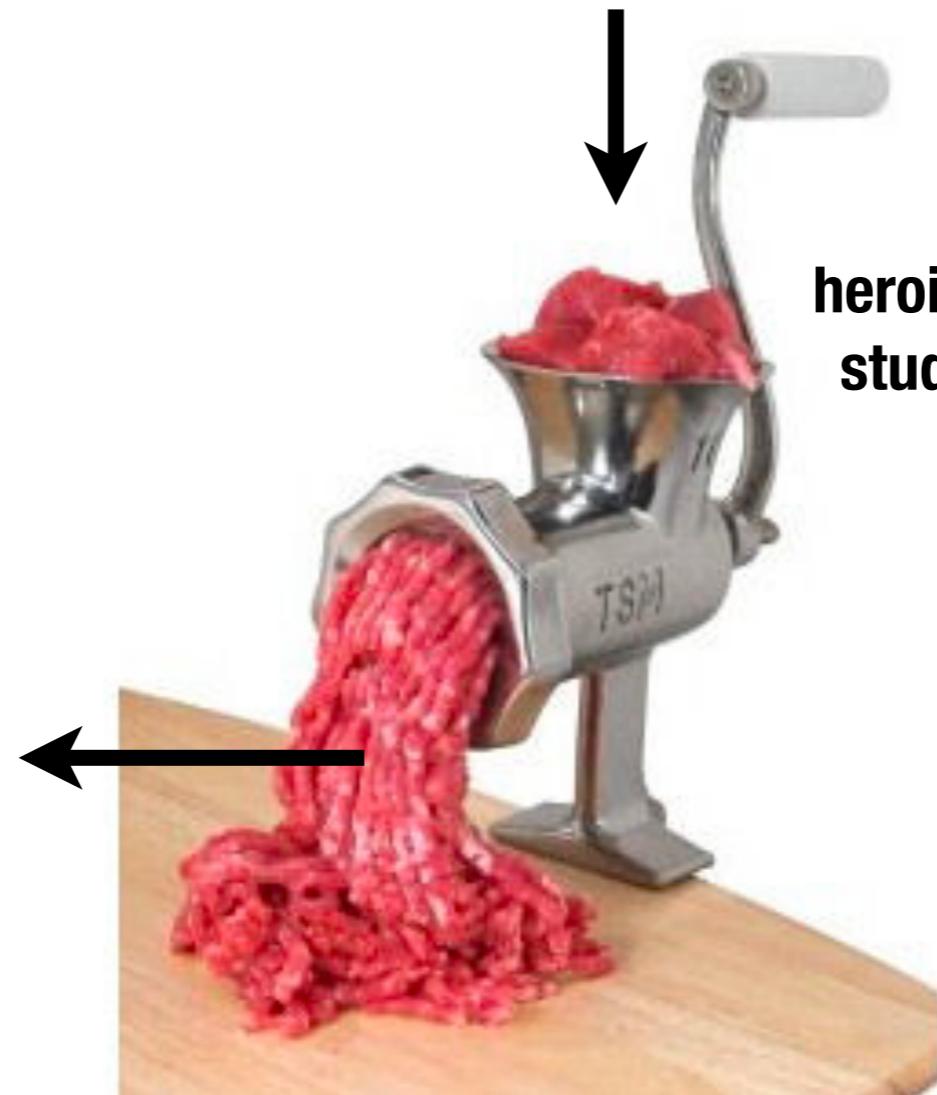
How are forcefields made?

**experimental data
quantum chemistry
keen chemical intuition**



**heroic effort by graduate
students and postdocs**

**a parameter set we
desperately hope someone
actually uses**

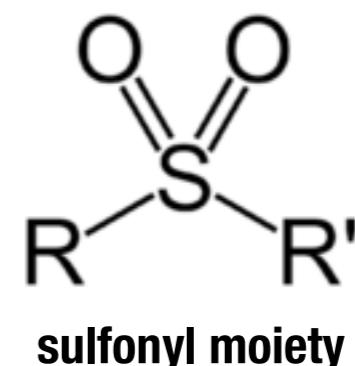


Poorly represented functional groups are problematic

Hydration free energy errors by functional group

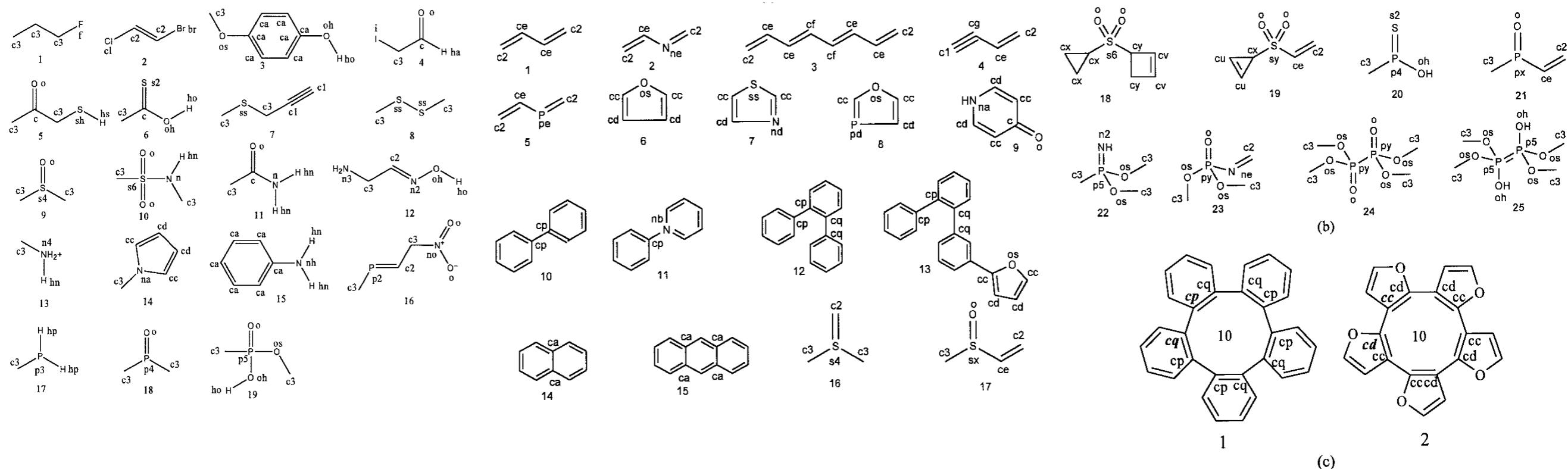
functional group	number	t-value	significance	mean error
acid	73	-7.43	4e-13	-0.34
alcohol	38	3.62	0.0003	1.29
aldehyde	20	-3.04	0.003	-0.07
alkanes	28	-1.69	0.09	0.31
alkene	35	2.34	0.02	1.07
alkyl bromide	17	3.31	0.001	1.50
alkyl chloride	31	2.31	0.02	1.09
alkyl iodide	9	0.59	0.6	0.86
alkyne	6	-0.38	0.7	0.49
amine	44	-0.65	0.5	0.55
aromatic compound	170	-1.05	0.3	0.55
aryl chloride	20	1.65	0.1	1.04
carbonitrile	12	3.22	0.001	1.63
cyclic hydrocarbon	8	-1.18	0.2	0.21
ester	8	-1.69	0.09	0.02
ether	42	2.18	0.03	1.01
halogen derivative	22	0.32	0.8	0.73
heterocyclic compound	48	2.38	0.02	1.02
hypervalent S	5	-4.55	7e-06	-1.50
ketone	25	-2.77	0.006	0.05
nitro compound	17	1.86	0.06	1.13
other	29	-0.48	0.6	0.55
phenol or hydroxyhetarene	33	2.72	0.007	1.16
thiol	5	0.51	0.6	0.89

Poorly represented functionalities in parameterization set lead to pathologies in use.



As drug discovery explores new parts of chemical space, how can forcefields keep up?

GAFF was parameterized for a limited diversity of compounds (and hence atom types):



Extension of GAFF by others “nontrivial” (effectively impossible for anyone but Junmei)

Our approach to parameterization has evolved over time, but it's still not completely automated by any measure

year	forcefield	parameter fitting	atom types
1990s	AMBER parm96	lots of “hand tweaking”	hand-picked
early 2000s	GAFF	genetic algorithm	hand-picked
mid 2000s	TIP4P-Ew	least-squares optimization	hand-picked

(an AMBER-centric view, but meant to be illustrative)

Torsion barrier for peptide bond from `parm96.dat`

```
X -C -N -X    4   10.00      180.0      2.      AA | check Wendy? & NMA
```

How can we move to entirely automated schemes that are easy to grow and refine?

What would we want out of a forcefield parameterization scheme (while we're dreaming)?

Everything is **automatic**; don't need to tweak things by hand.

There is only one **Christopher Bayly**, and he's pretty busy, so we shouldn't need to consult him too frequently for deep feats of chemical insight.

Automatically chooses optimal functional forms.

If we find ourselves in uncharted territory in chemical space, **can add more data**. (Maybe it could even tell us which new data would be useful!)

I am lazy and don't want to learn new algorithms.

Would give us an idea of **how reliable** it new predictions are expected to be.

Is there really a procedure that could fit these criteria?

Bayesian inference can help!

Bayes rule provides a **probability measure over unknown parameters given data**:

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

\mathcal{D} data

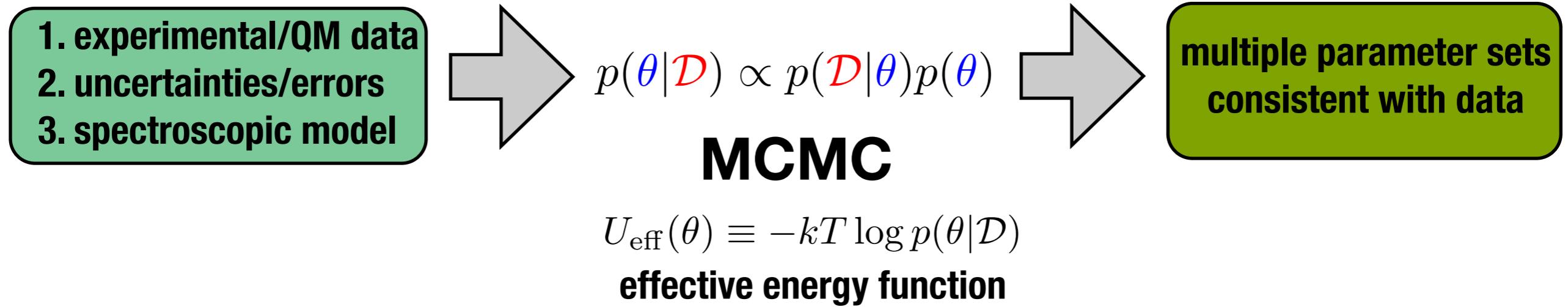
θ forcefield parameters

$p(\theta|\mathcal{D})$ posterior

$p(\mathcal{D}|\theta)$ data model

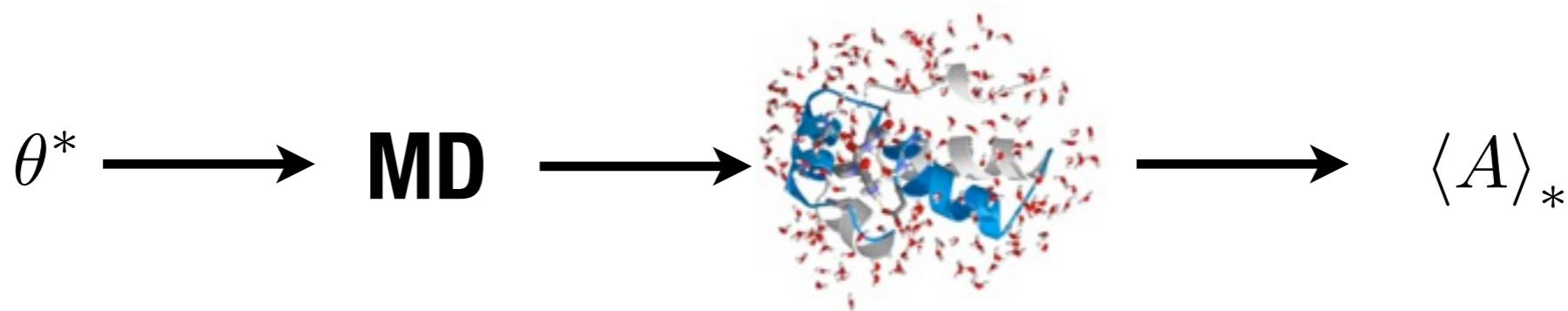
$p(\theta)$ prior on forcefield parameters

Bayesian parameterization can exploit the one tool we already know how to use well, Markov chain Monte Carlo

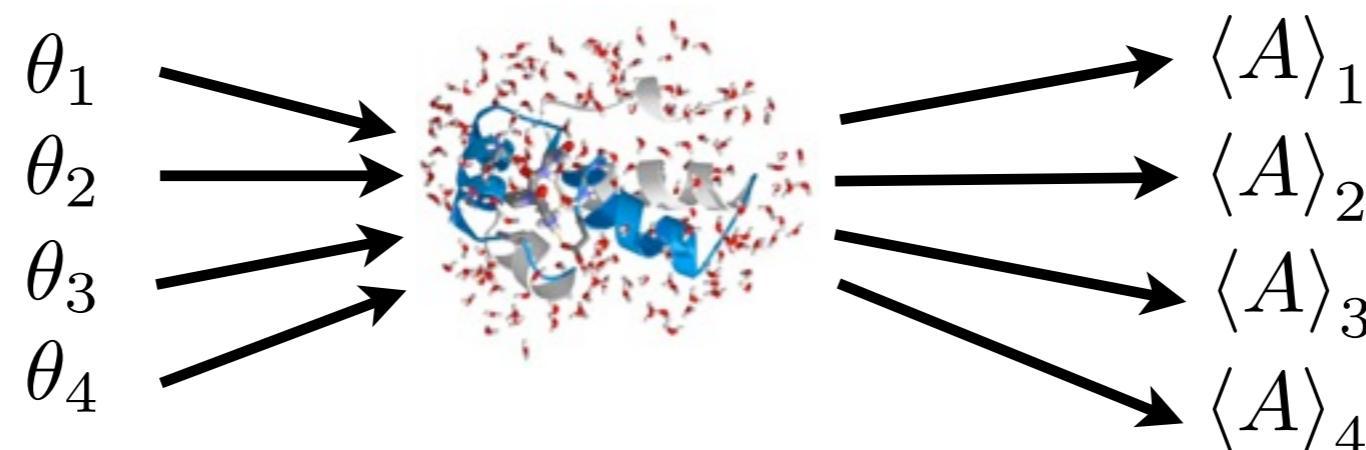


Evaluation of computed properties with multiple parameter sets gives an **estimate of systematic error** in forcefield

Compute properties with one representative parameter set from collection

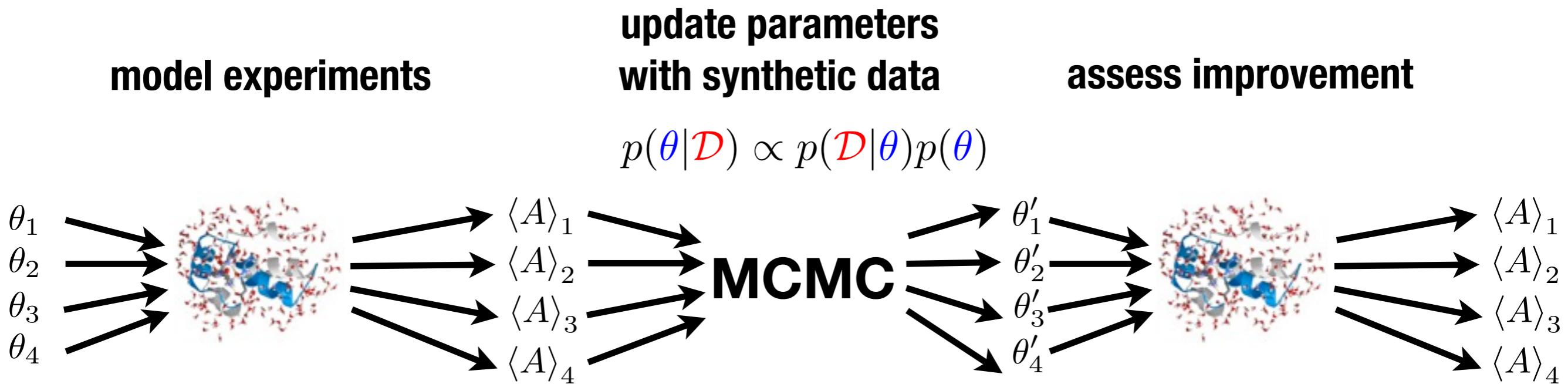


Estimate computed properties for other parameter sets from collection by reweighting (e.g. MBAR)



Can estimate both **statistical** and **systematic** components of computed property!

Bayesian experimental design can tell us which kinds of data would best help reduce systematic error

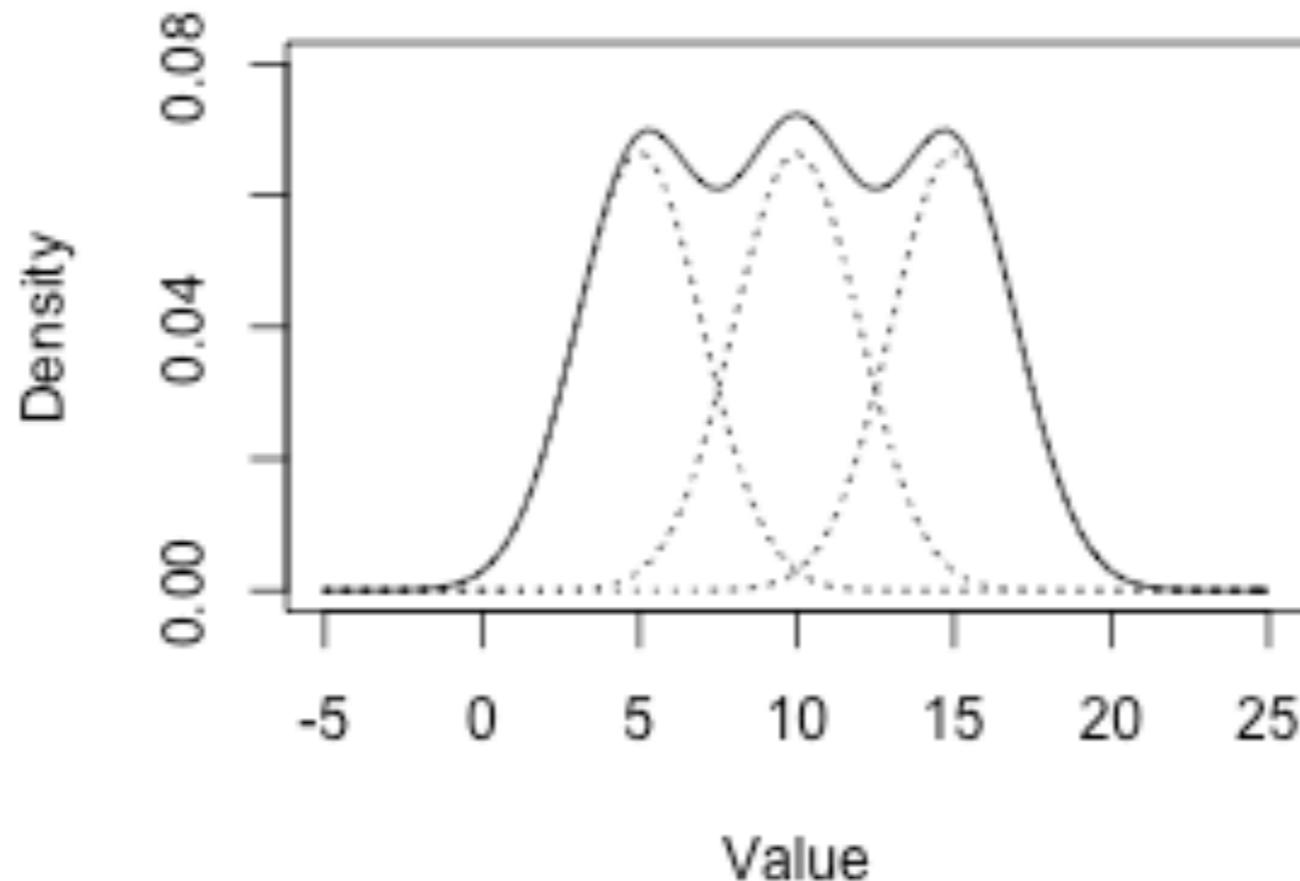


Modeling the outcome of new experiments can suggest which will be **most informative** in reducing systematic error

Bayesian model selection can be a powerful tool in forcefield development and parameterization

Classic Gaussian mixture model case:

- How many components do the data support?
- What are their parameters?

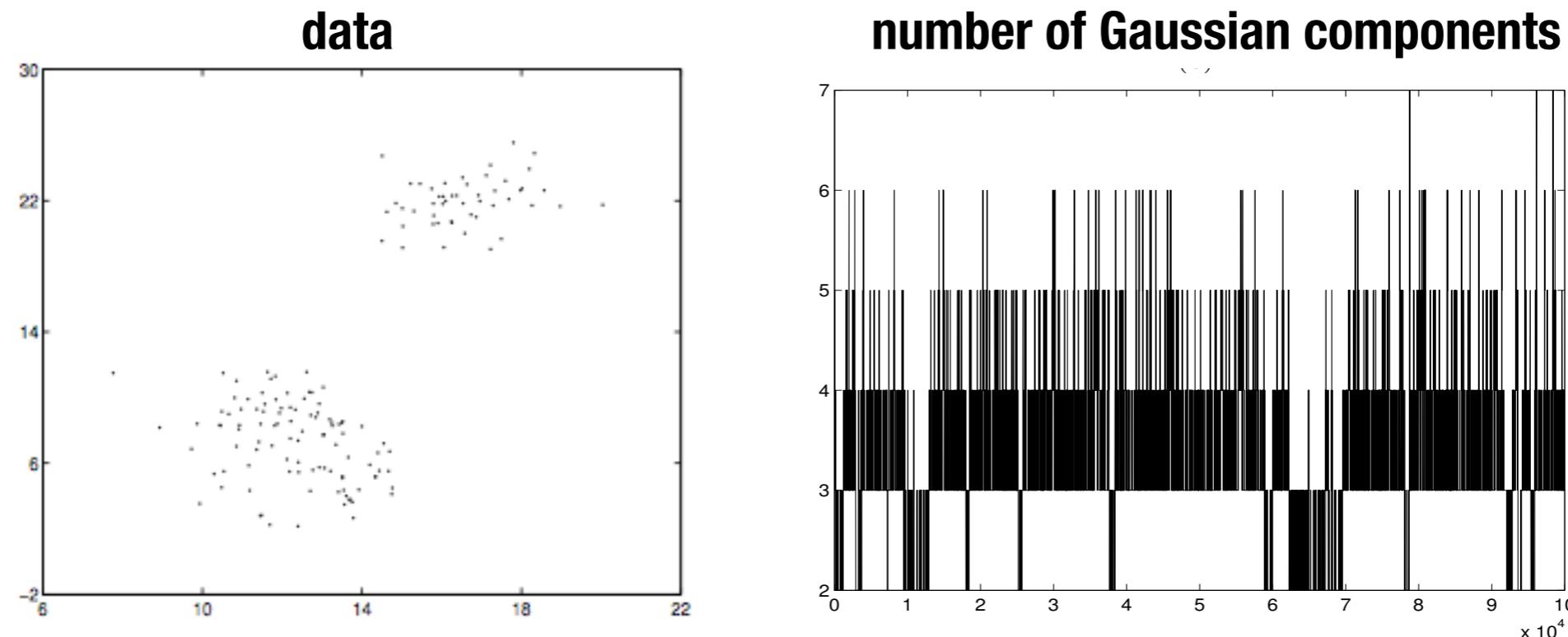


Bayesian inference penalizes complexity

Bayesian model selection can be a powerful tool in forcefield development and parameterization

Classic Gaussian mixture model case:

- How many components do the data support?
- What are their parameters?



Bayesian inference penalizes complexity

Bayesian model selection can be a powerful tool in forcefield development and parameterization

Reversible-jump Monte Carlo (RJMC) allows jumps among models; models best supported by data are preferentially sampled. (Calculation is analogous to grand canonical MC.)

RJMC moves simply have to satisfy detailed balance in model space (e.g. via Metropolis-Hastings):

$$P_{\text{accept}} = \min \left\{ 1, \frac{P(\text{new})}{P(\text{old})} \cdot \frac{P(\text{old}|\text{new})}{P(\text{new}|\text{old})} \right\}$$

RJMC can automate selection of **most appropriate functional form** and/or **mixing rules** for vdW:

Lennard-Jones

exponential-6

Halgren buffer 14-7

Lorentz-Berthelot

Waldman-Hagler

Halgren HHG

RJMC can automate selection of **atom types**: only as many types as the data supports!

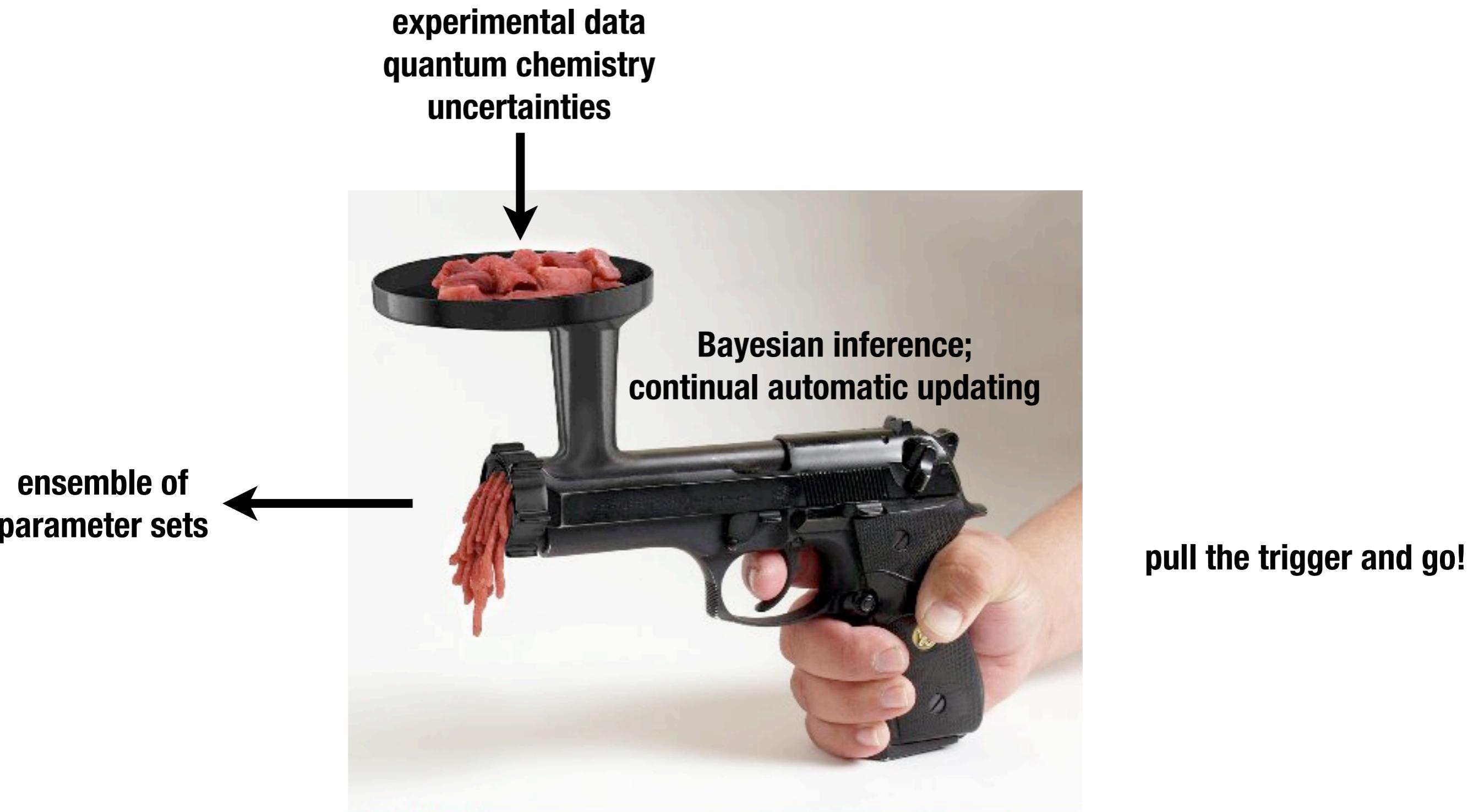
splitting moves that split off a subtype (e.g. sp3 carbon split from generic carbon type)

subtype deletion moves that remove subtypes

What would we want out of a forcefield parameterization scheme?

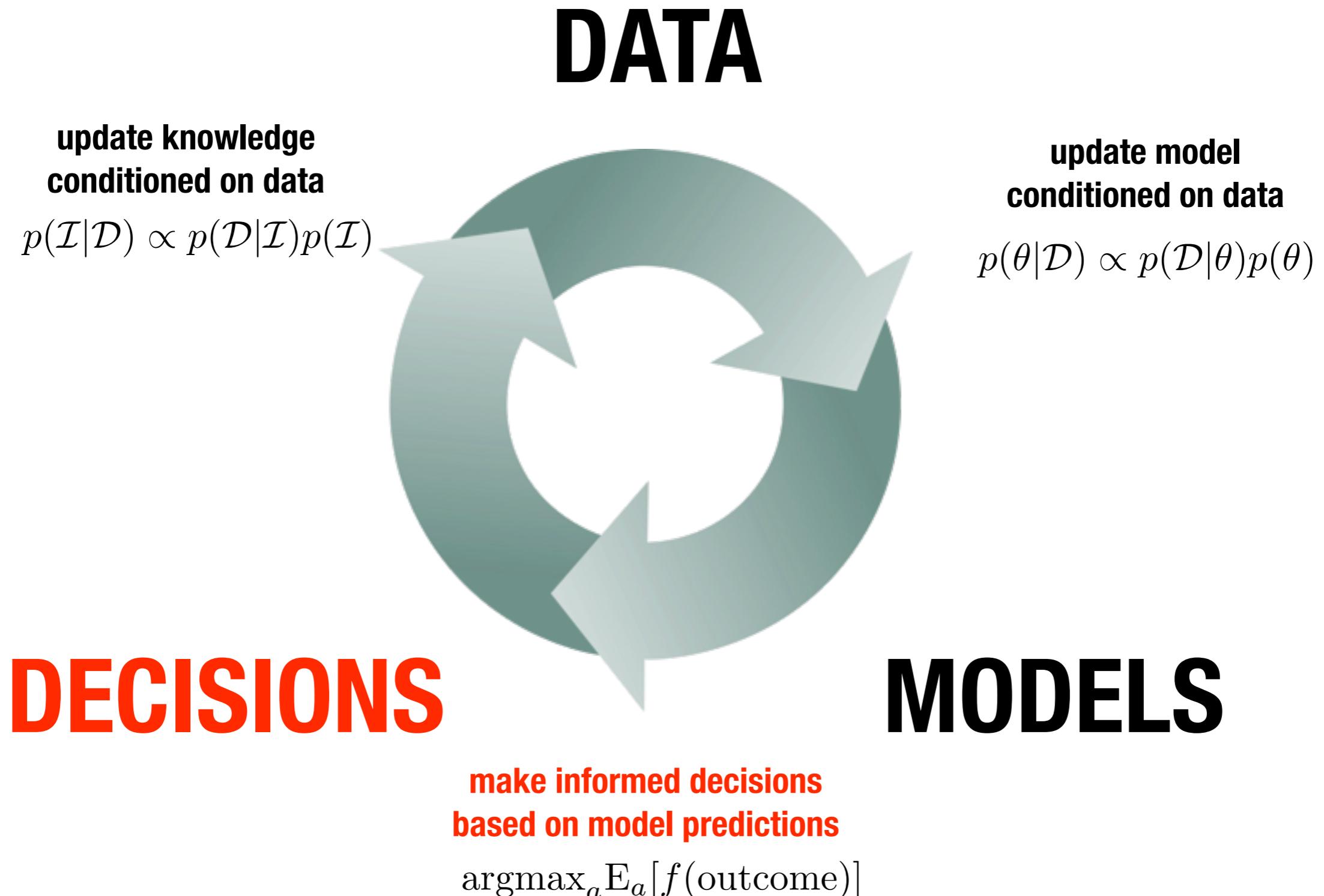
- ✓ Everything is automatic; don't need to tweak things by hand.
- ✓ Data goes in, parameters come out!
- ✓ Never need to choose atom types or perform immense feats of chemical intuition.
Bayesian model selection with RJMC!
- ✓ If we're uncertain about which functional forms are best, automatically chooses.
Bayesian model selection with RJMC!
- ✓ If we find ourselves in uncharted territory in chemical space, throw in more data.
(Maybe it could give us hints about which data would be useful to obtain?)
Bayesian updating and experimental design!
- ✓ Would be nice if parameterization procedure didn't require me to learn crazy new mathematics or algorithms (mainly because I'm lazy)
Isomorphic with statistical mechanics!
- ✓ Would give us an idea of how reliable it new predictions are expected to be.
Multiple parameter sets from Bayesian posterior

The future of forcefield parameterization?



The universal cycle of progress

Bayesian inference can drive progress



“Pick a winner”

**Suppose we make predictions for 20 compounds the chemists can make.
What should they make?**

Let's pick a simple illustrative model

Simple model for affinity/potency/whatever:

$$\Delta G = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$\mathbf{x} \in \mathbb{R}^3$ **feature/fingerprint vector**

$\mathbf{A} \in \mathbb{R}^{3 \times 3}$ **model parameters**

true model \mathbf{A}^*

error in measured affinities $\Delta G_{obs} = \Delta G + \xi$

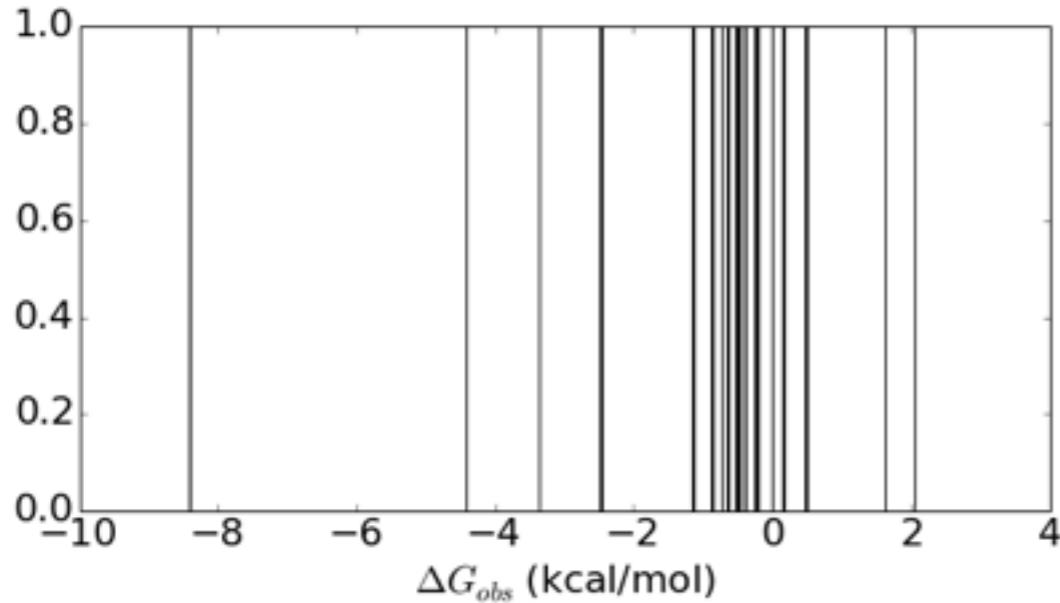
measurement error is Gaussian with zero mean $\xi \in \mathcal{N}(0, \sigma^2)$

σ **measurement error standard deviation**

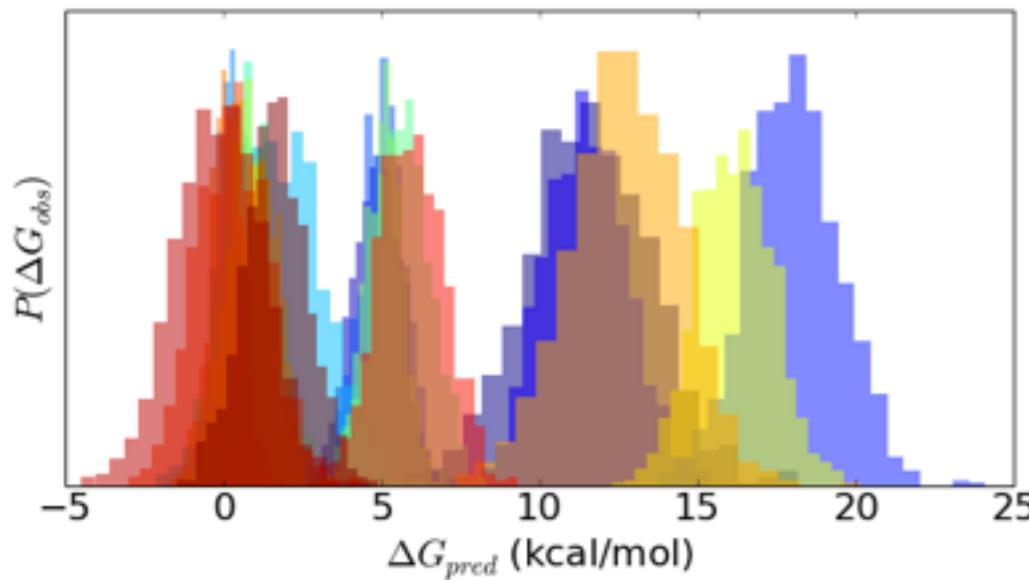
This model is simple, but these principles apply to whatever you want

Suppose we have already made some compounds

Measured affinities of the molecules that we've already made:



Now, we make predictions for 20 new molecules we **could make:**



Which ones **should we make?**

Maximizing the expectation (as in Bayesian bandits)

We could select a compound that maximizes our expected affinity gain

Decisionmaking for the risk-averse

Societal issues often penalize failure much more than rewarding success.

“You only get to be wrong once.”

How can we incorporate risk-aversion into decisionmaking?

A CONCRETE EXAMPLE:

Suppose we have model-derived predictions for 20 compounds, and the chemists want to synthesize 5.

How can we maximize the probability that at least ONE compound will be observed to improve potency?



Acknowledgments

Statistical fisticuffs

Ant Nicholls

Kim Branson

ITC and entropy-enthalpy compensation

**Kim Branson, Sarah Boyce, Paul Novick, Vijay Pande,
David Minh, David Mobley**

Telepresents

Johnny 5 and his support staff

(Brian Cole, Craig Bruce, Ben Ellingson)

