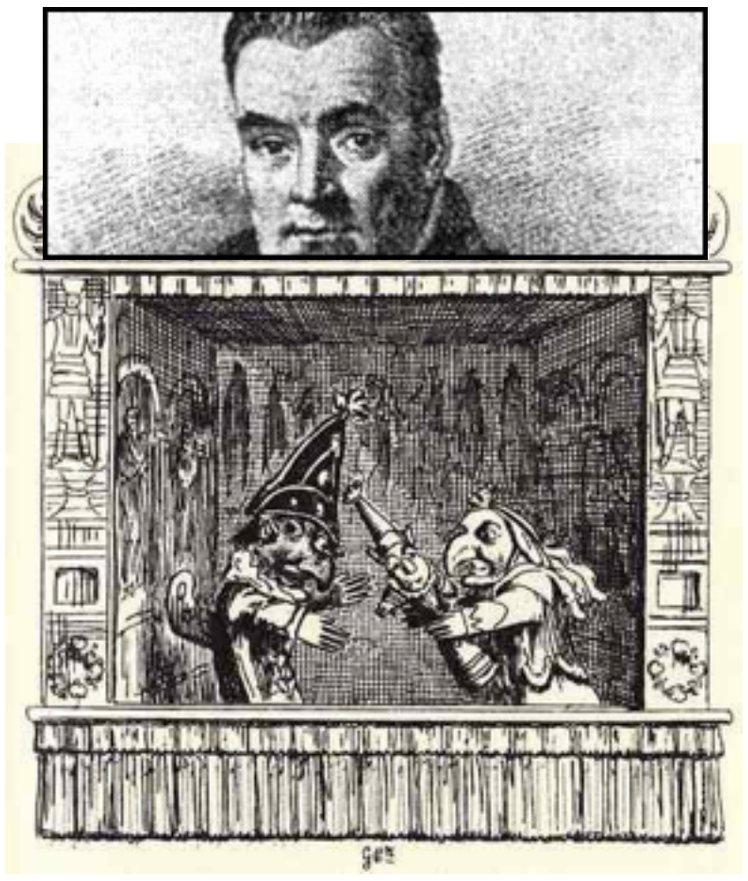
Story time and puppet show with Mr. Bayes

Data, Models, and Decisions



John D. Chodera
Computational Biology Program, Memorial Sloan-Kettering Cancer Center
http://www.choderalab.org

Tweet your questions to @jchodera

Slides, references, and IPython tutorials available at http://choderalab.org/cup-xiv

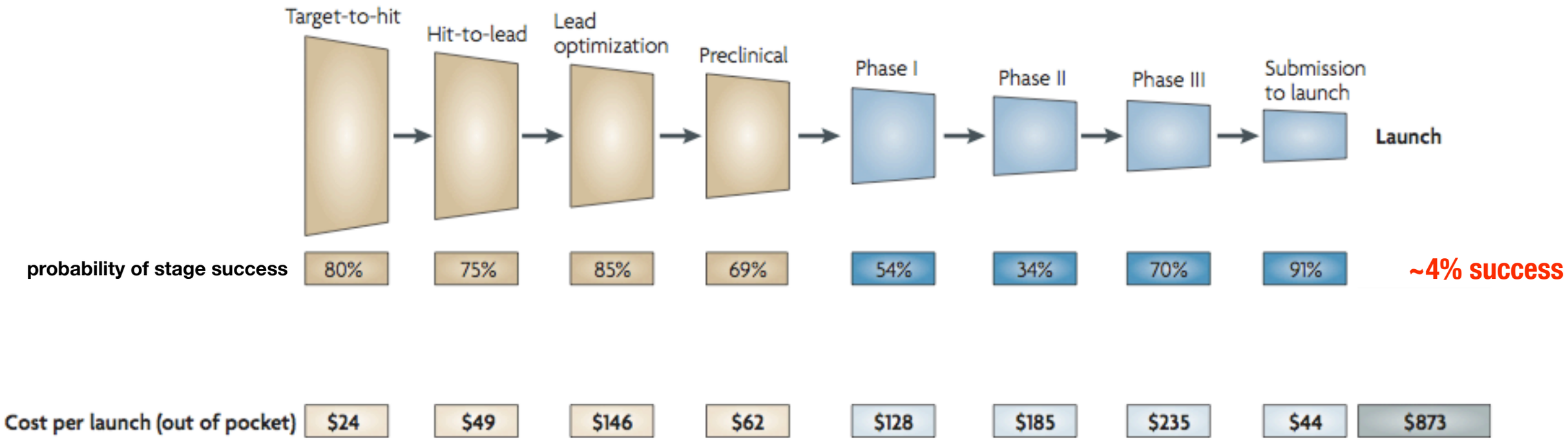
We are all here because we desperately hope modeling can improve the dismal success rates in drug discovery

---

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

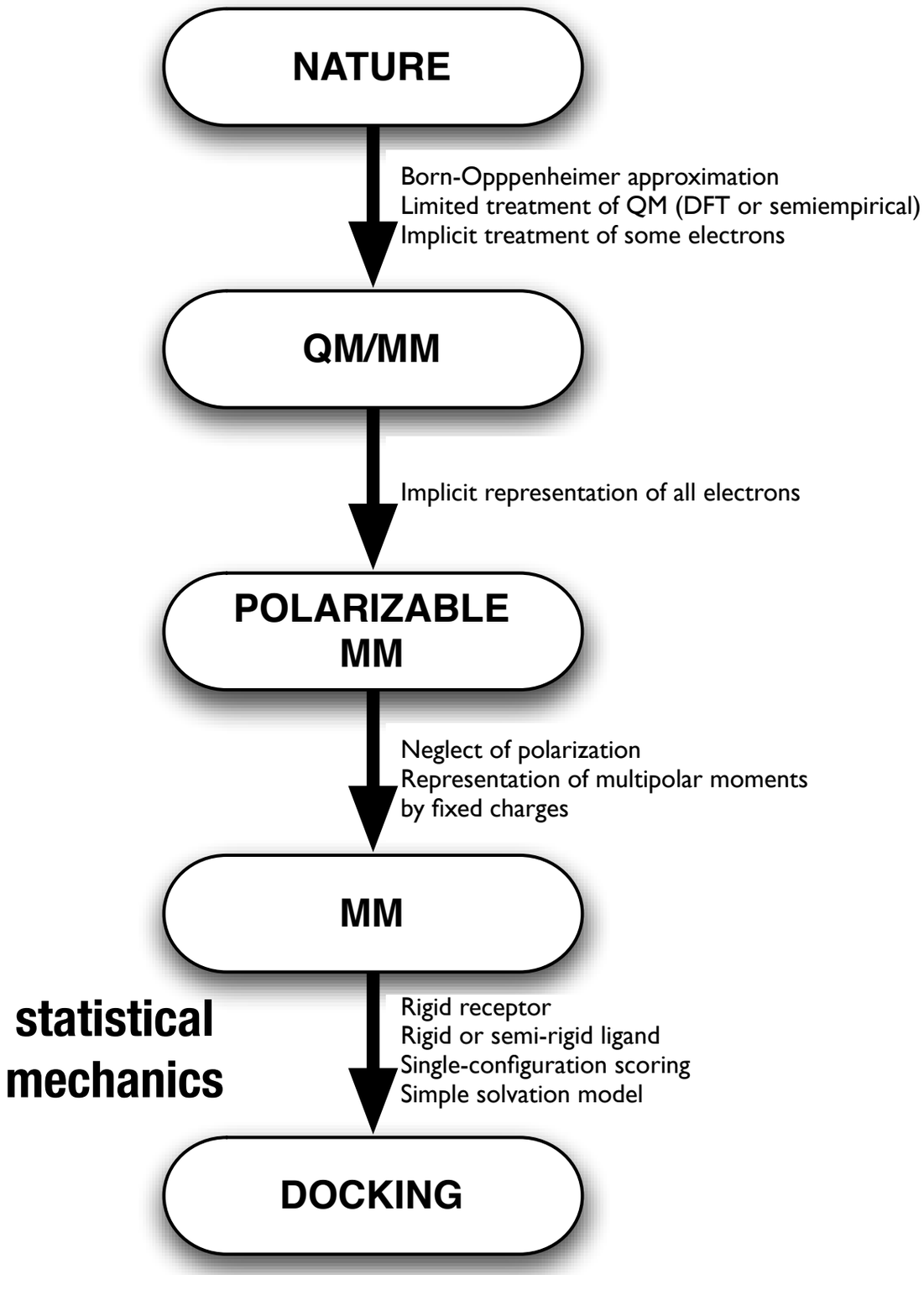
Number of new molecular entities approved by FDA 2005-2009 is half that from previous five years

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

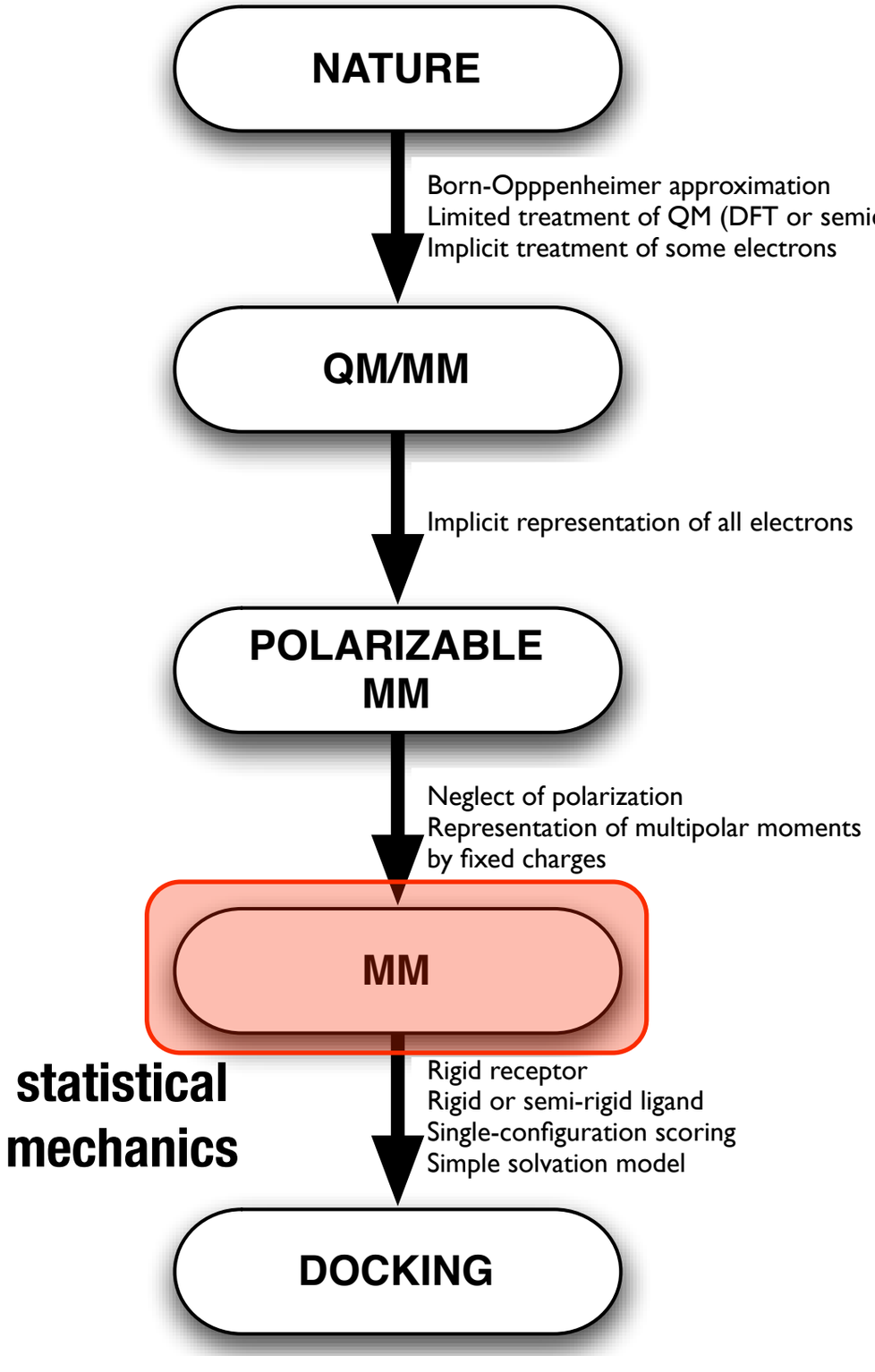


Paul et al. Nat. Rev. Drug Discover. 9:203, 2010.
Chodera et al. Curr. Opin. Struct. Biol., 21:150, 2011.

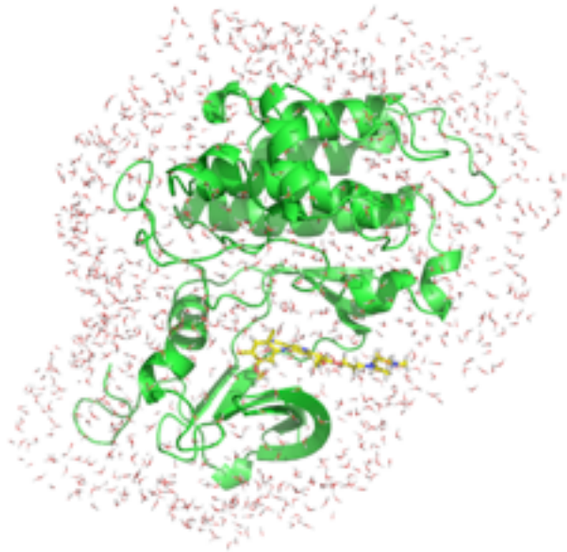
What details are crucial for useful accuracy in rational design? Our lab is trying to find out!



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



if insufficiently accurate, add detail in perturbative manner



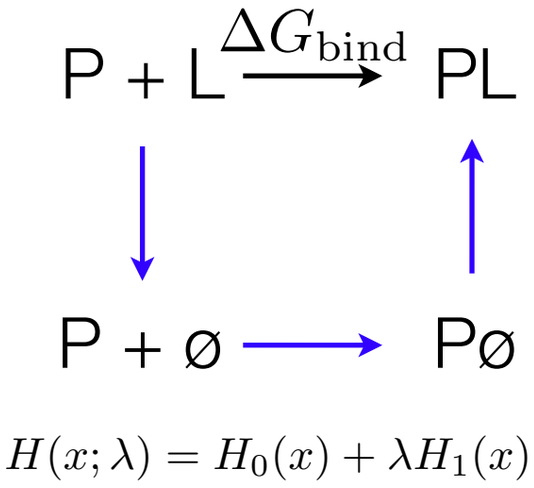
molecular mechanics potential energy function (e.g. AMBER)

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

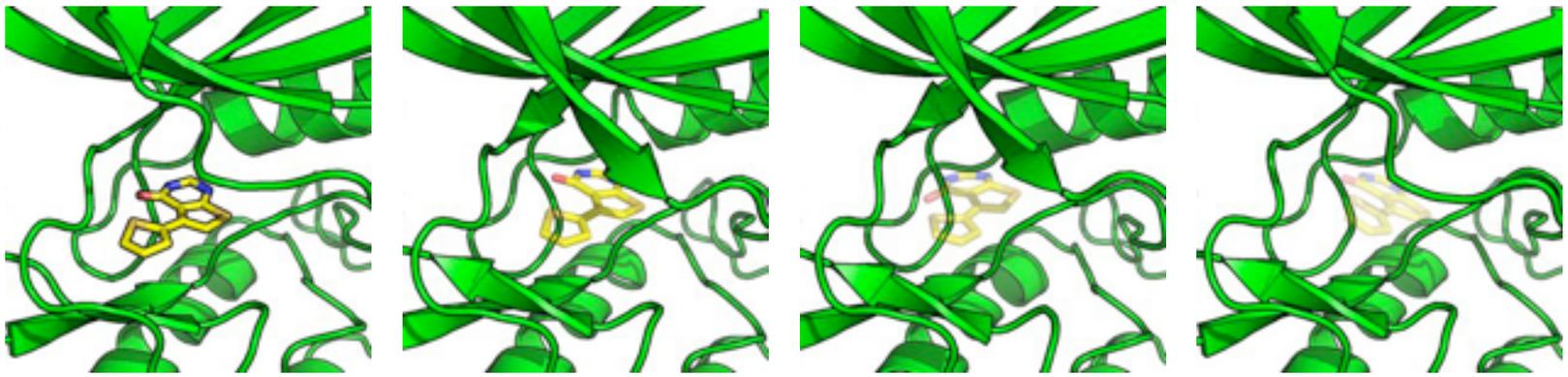
if accurate enough, systematically remove detail

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



$$H(x;\lambda) = H_0(x) + \lambda H_1(x)$$

multiple simulations of "alchemical intermediates" are run



imposition of restraints | discharging | vdW/steric decoupling | noninteracting

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).

Recent review: Chodera, Mobley, Shirts, Dixon, Branson, Pande. Curr Opin Struct Biol 21:150, 2011.

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

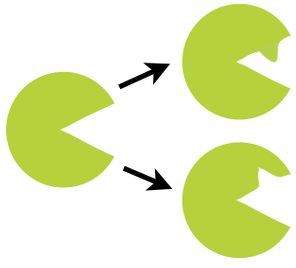
GPU acceleration

$$\pi(x)K(x,y) = \pi(y)K(y,x)$$
$$AH \leftrightharpoons A^- + H^+$$

MCMC framework

enhanced algorithms

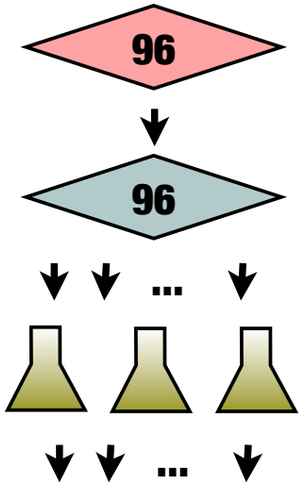
Testing predictions by mutating the target protein



mutate proteins instead of ligands

¢

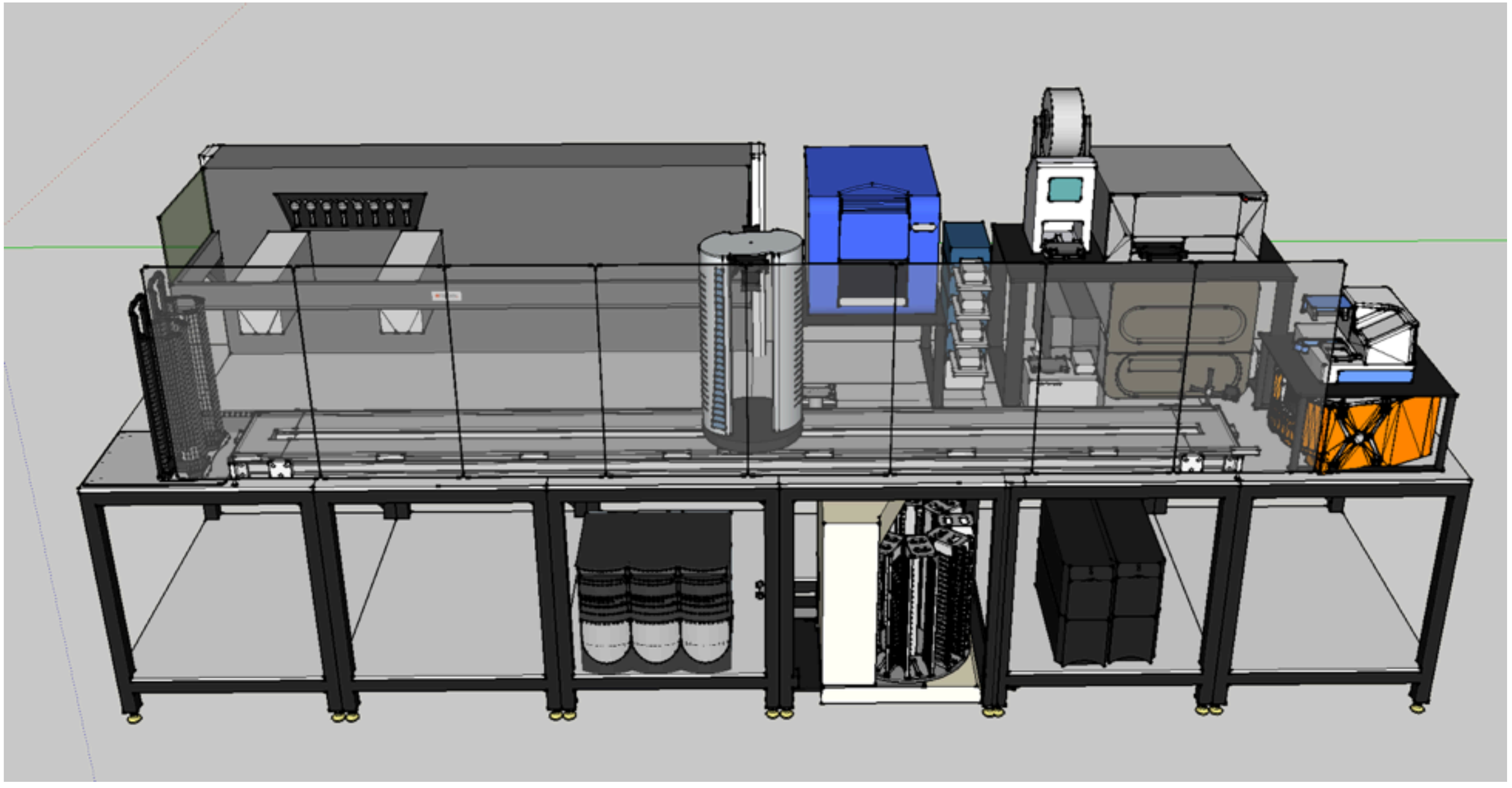
buy inexpensive ligands



high-throughput experiments

The last, best hope for data: RUG-1

(robotic undergraduate no. 1)

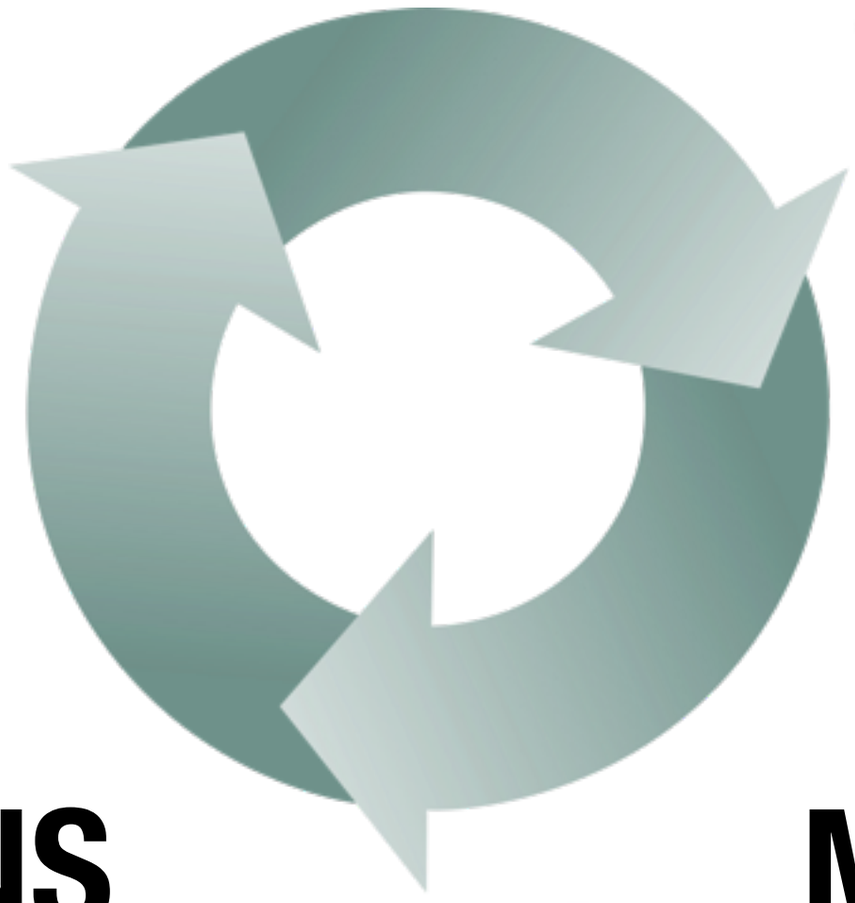


Automated platform for bacterial cloning, mutagenesis, expression, purification, and binding affinity measurement with 24/7 wallet-draining capability

The universal cycle of progress

The universal cycle of progress

DATA

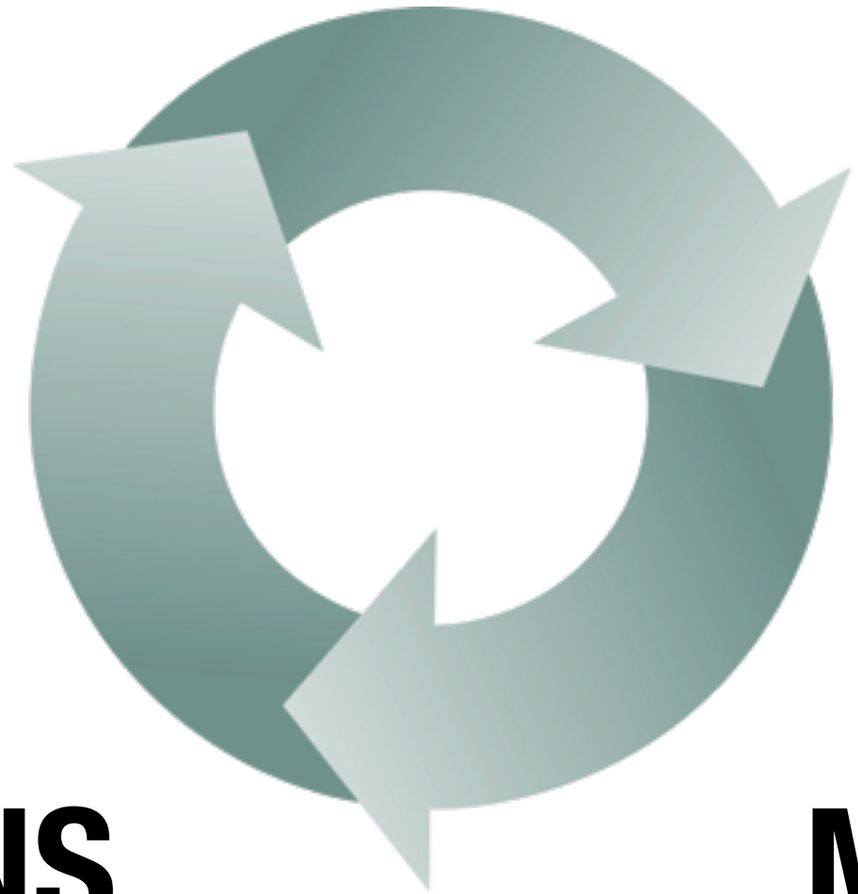


DECISIONS

MODELS

The universal cycle of progress

DATA



DECISIONS **MODELS**

How can we do all this without losing our minds?

Bayesian inference can drive progress

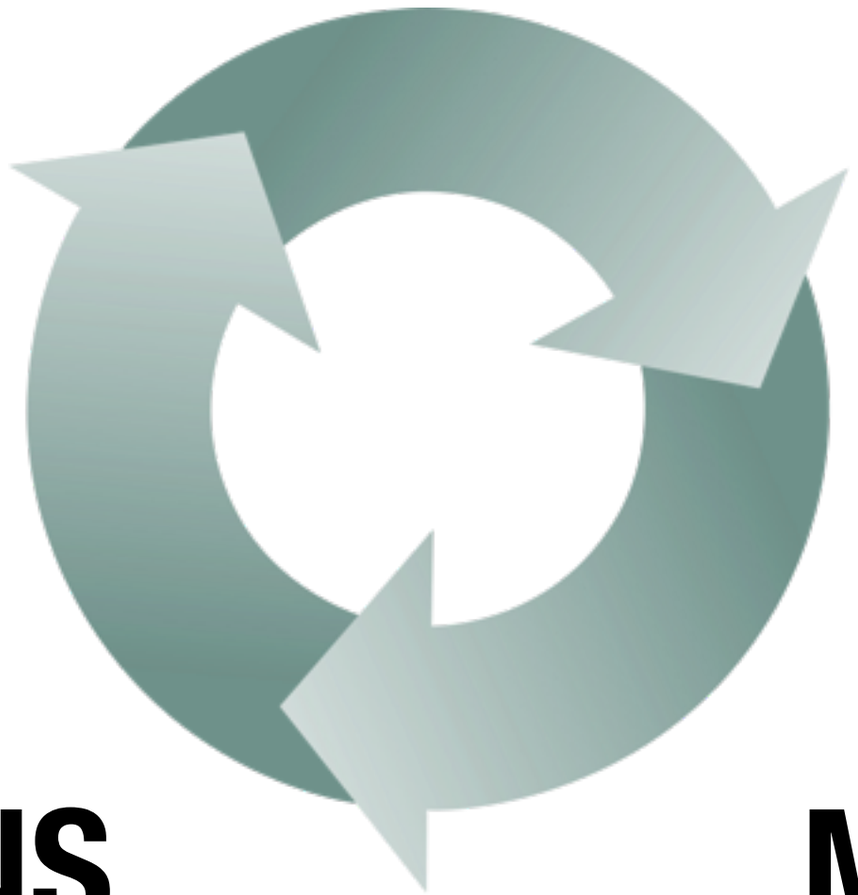
DATA

update knowledge conditioned on data

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

update model conditioned on data

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



DECISIONS

MODELS

make informed decisions based on model predictions

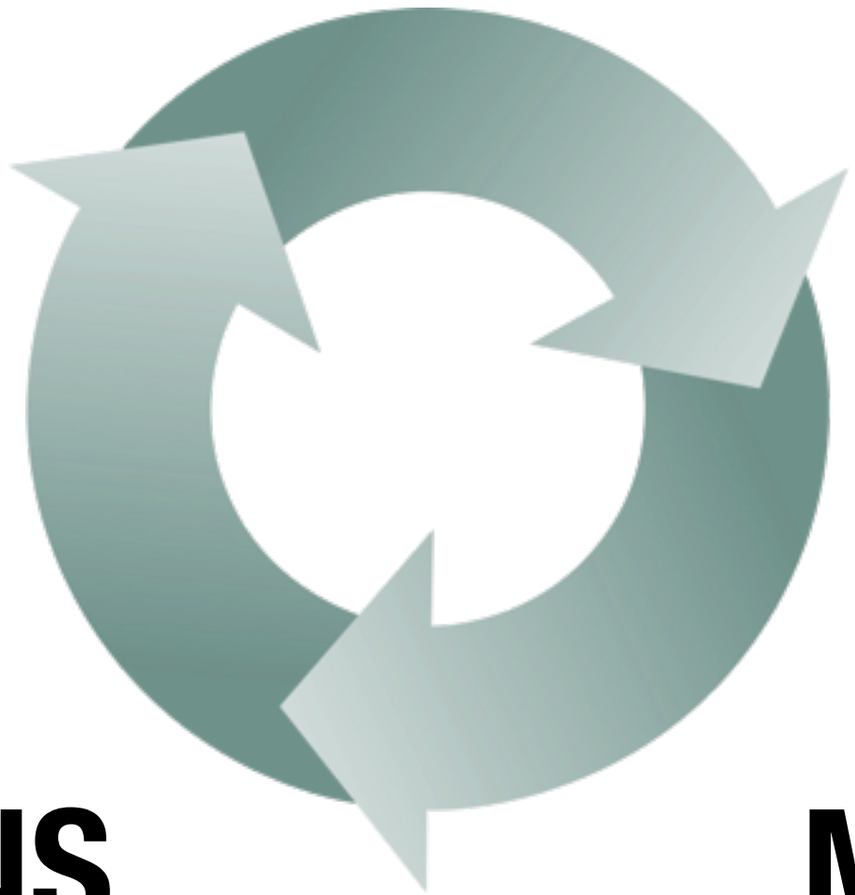
$$\text{argmax}_a \text{E}_a[f(\text{outcome})]$$

Bayesian inference can drive progress

DATA

update knowledge conditioned on data

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



update model conditioned on data

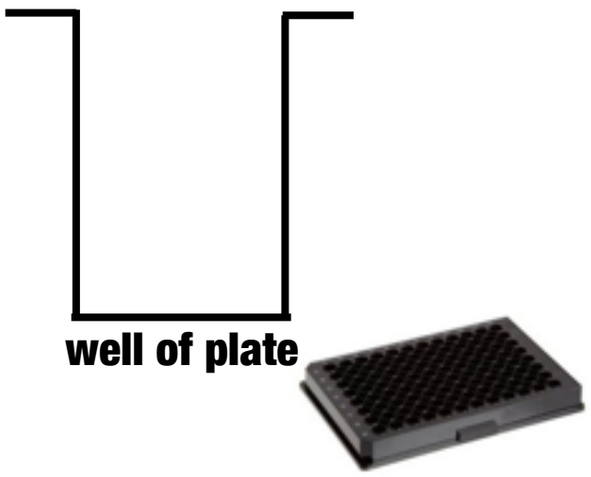
$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

DECISIONS **MODELS**

make informed decisions based on model predictions

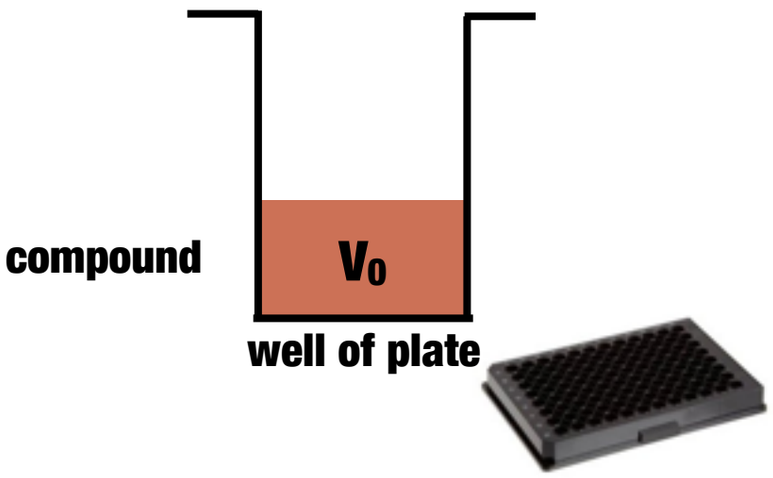
$$\text{argmax}_a \text{E}_a[f(\text{outcome})]$$

What is experimental error and where does it come from?



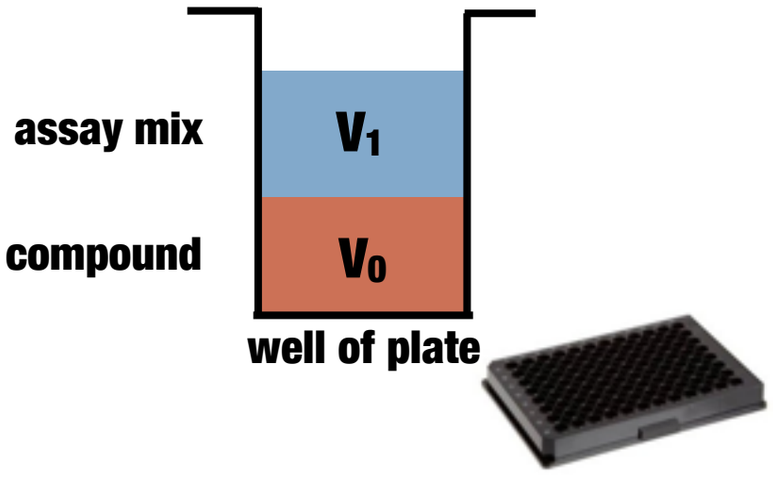
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?



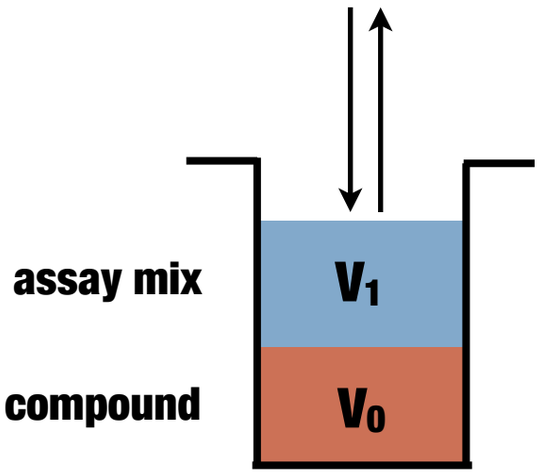
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?



IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

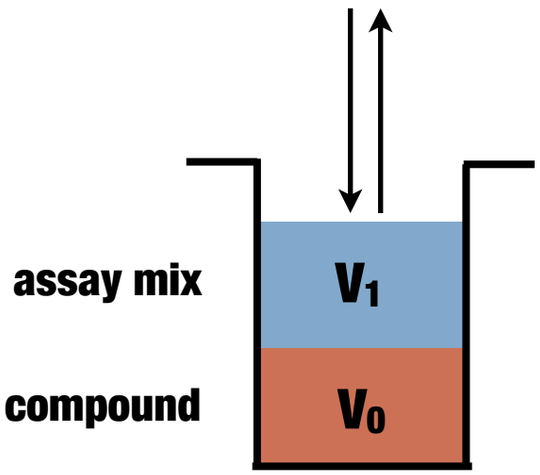
What is experimental error and where does it come from?



IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

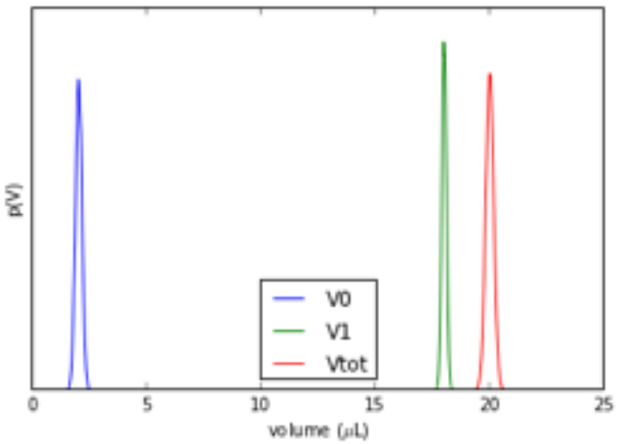
What is experimental error and where does it come from?

fluorescence measurement



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

finite pipetting precision



Biomek FX® Pipetting Performance Specifications

Span-8 Systems

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
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

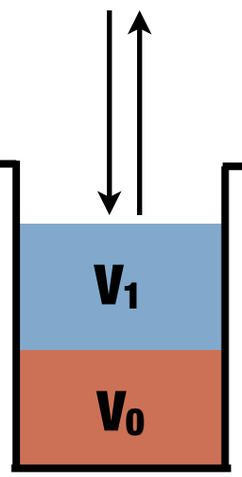
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?

fluorescence measurement

assay mix

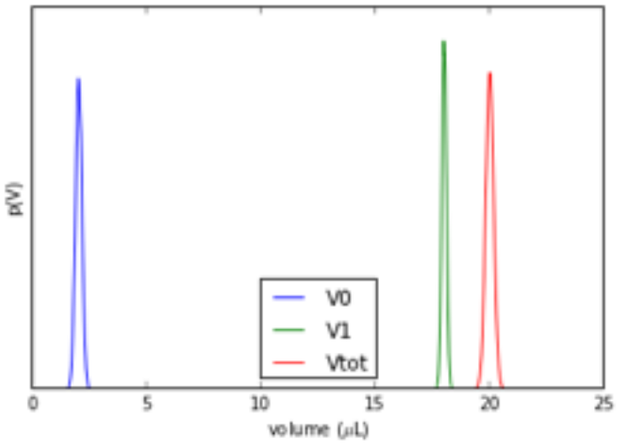
compound



well of plate

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

finite pipetting precision

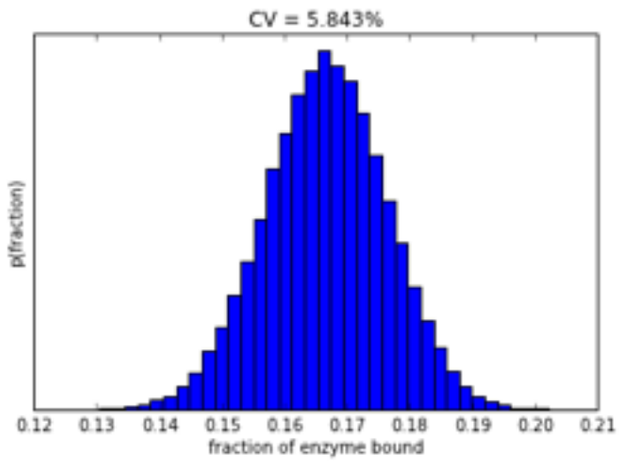


BIOMEK FX® PIPETTING PERFORMANCE SPECIFICATIONS

SPAN-8 SYSTEMS

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
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

variation in protein:ligand complex concentration



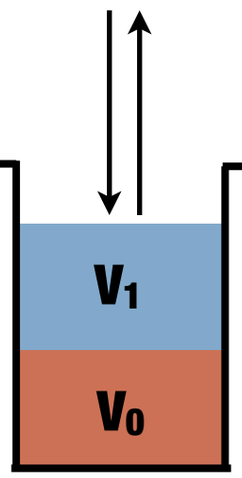
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?

fluorescence measurement

assay mix

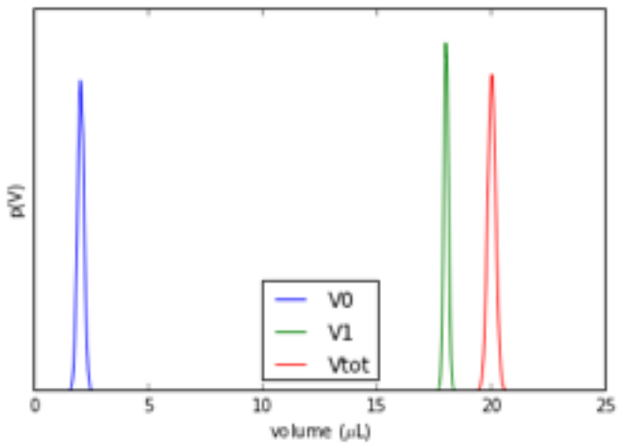
compound



well of plate

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

finite pipetting precision

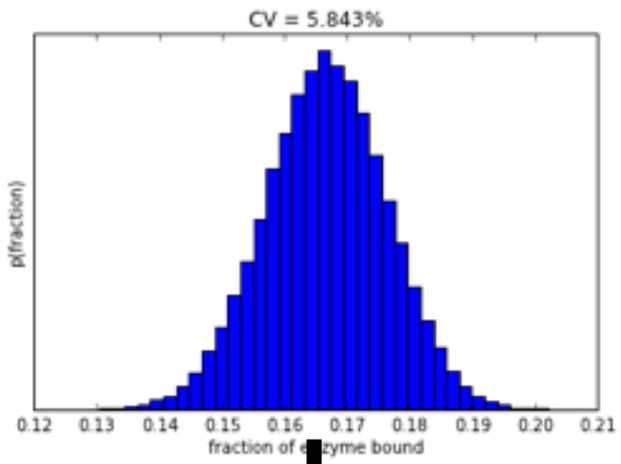


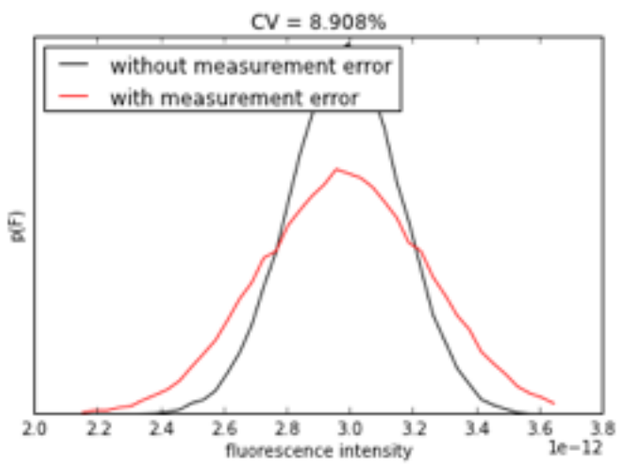
Biomek FX® Pipetting Performance Specifications

Span-8 Systems

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
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

variation in protein:ligand complex concentration





signal broadening due to measurement error

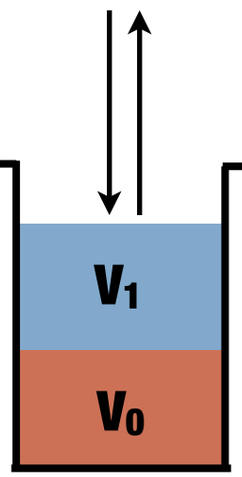
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?

fluorescence measurement

assay mix

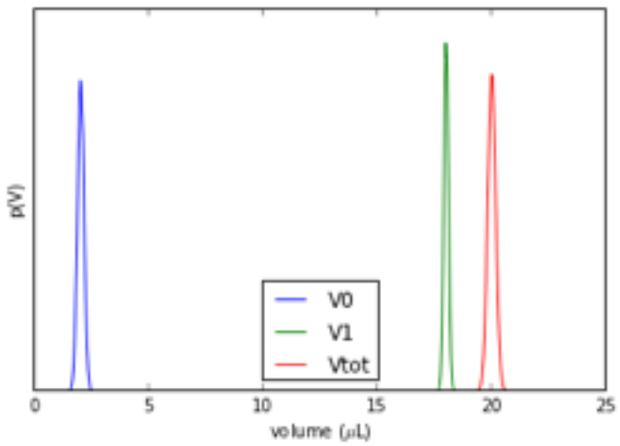
compound



well of plate

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

finite pipetting precision

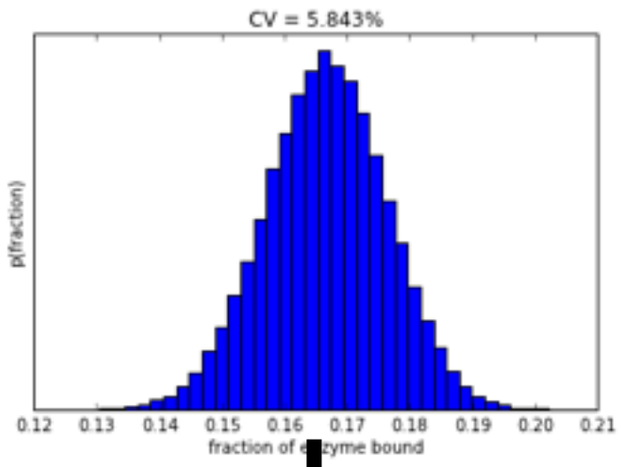


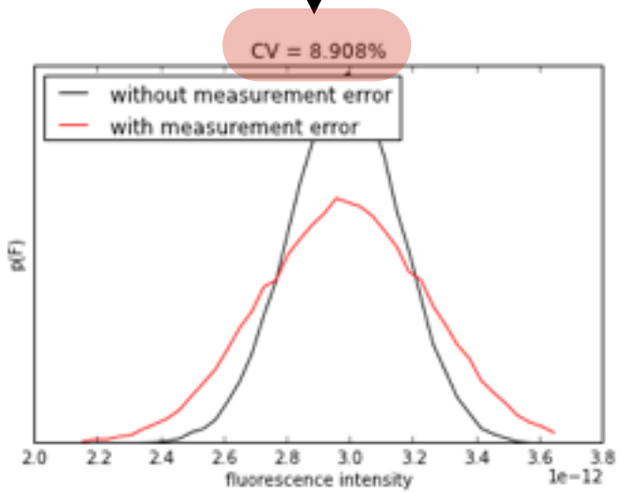
BIOMEK FX® PIPETTING PERFORMANCE SPECIFICATIONS

SPAN-8 SYSTEMS

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
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

variation in protein:ligand complex concentration





signal broadening due to measurement error

Is the expected CV small enough to be useful?

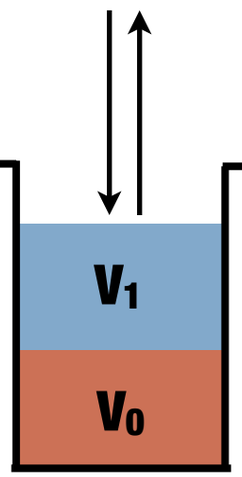
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

What is experimental error and where does it come from?

fluorescence measurement

assay mix

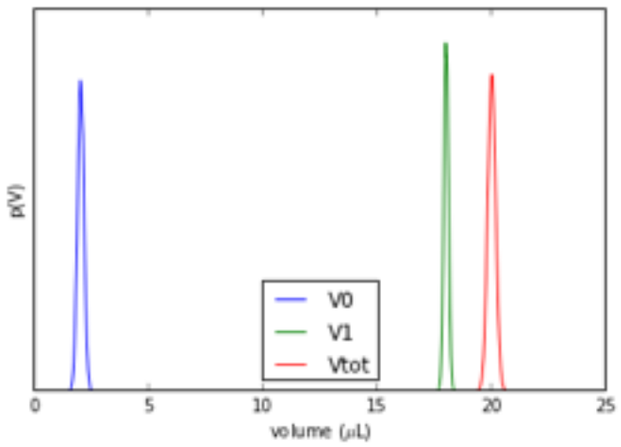
compound



well of plate

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

finite pipetting precision

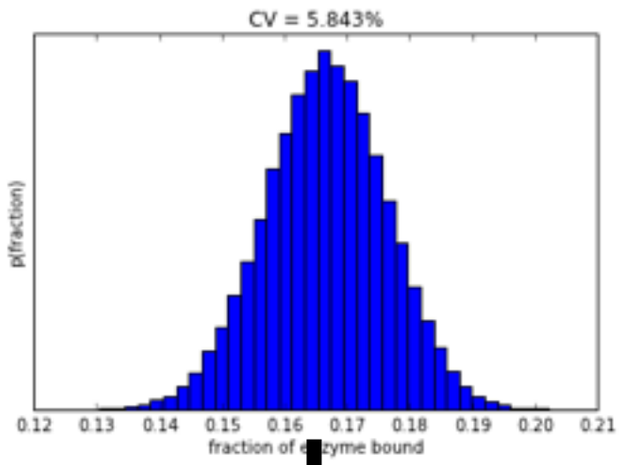


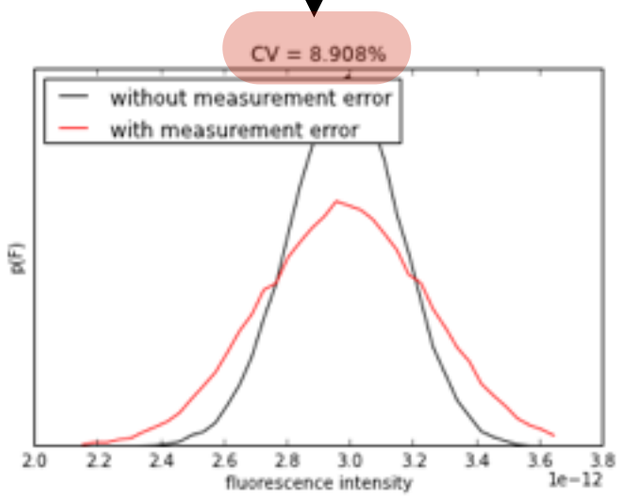
BIOMEK FX® PIPETTING PERFORMANCE SPECIFICATIONS

SPAN-8 SYSTEMS

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
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

variation in protein:ligand complex concentration

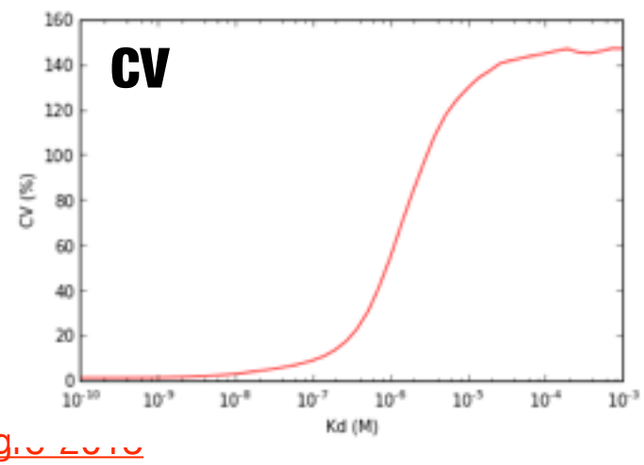




signal broadening due to measurement error

Is the expected CV small enough to be useful?

Modeling an experiment can ensure it will yield useful data or uncover unexpected issues



IPython tutorial available at https://github.com/choderalab/cadd-g...

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 »

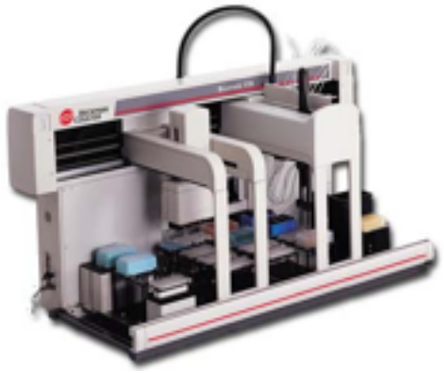
May 3, 2013

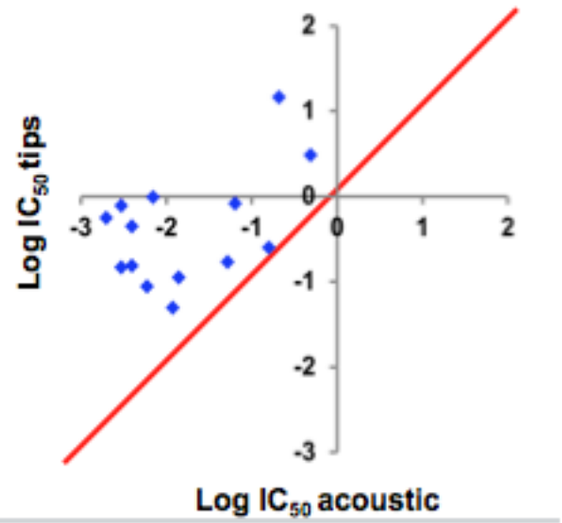
Drug Assay Numbers, All Over the Place

Posted by **Derek**

There's a truly disturbing paper out in *PLoSONE* 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

PLOS ONE

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

Sean Ekins[1]*, Joe Olechno[2], Antony J. Williams[3]

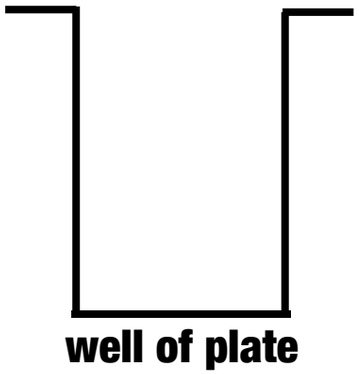
1 Collaborations in Chemistry, Fuquay-Varina, North Carolina, United States of America, **2** Labcyte Inc., Sunnyvale, California, United States of America, **3** Royal Society of Chemistry, Wake Forest, North Carolina, United States of America

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

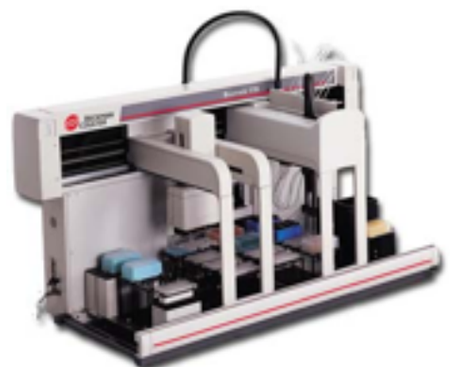
Is this just a result of pipetting accuracy differences?
Let's model the experiment to find out.

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

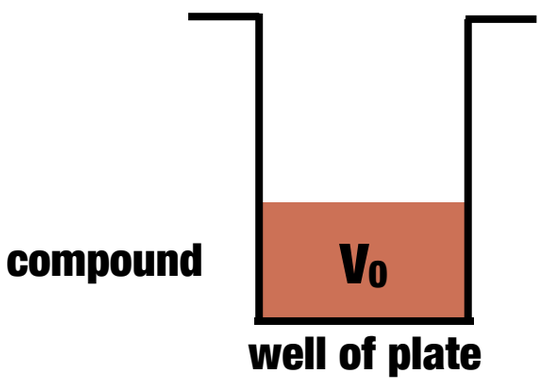
Modeling experimental error can be a valuable exercise: Preparing a dilution series via liquid-handling robot



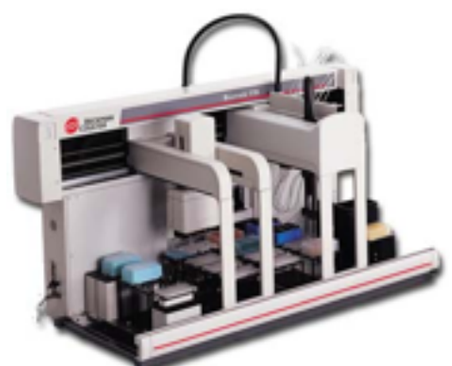
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013


Biomek NX/FX

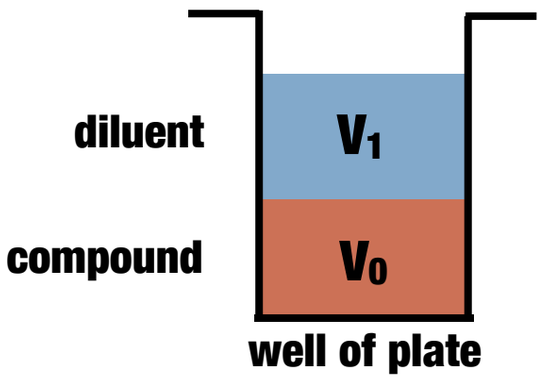
Modeling experimental error can be a valuable exercise: Preparing a dilution series via liquid-handling robot



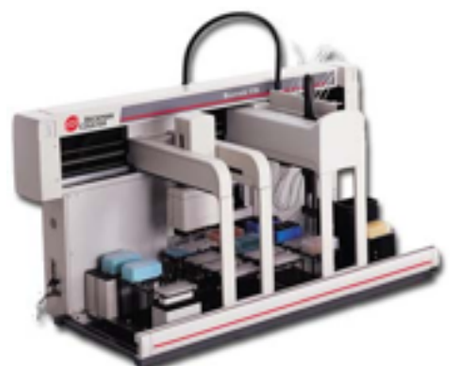
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013


Biomek NX/FX

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

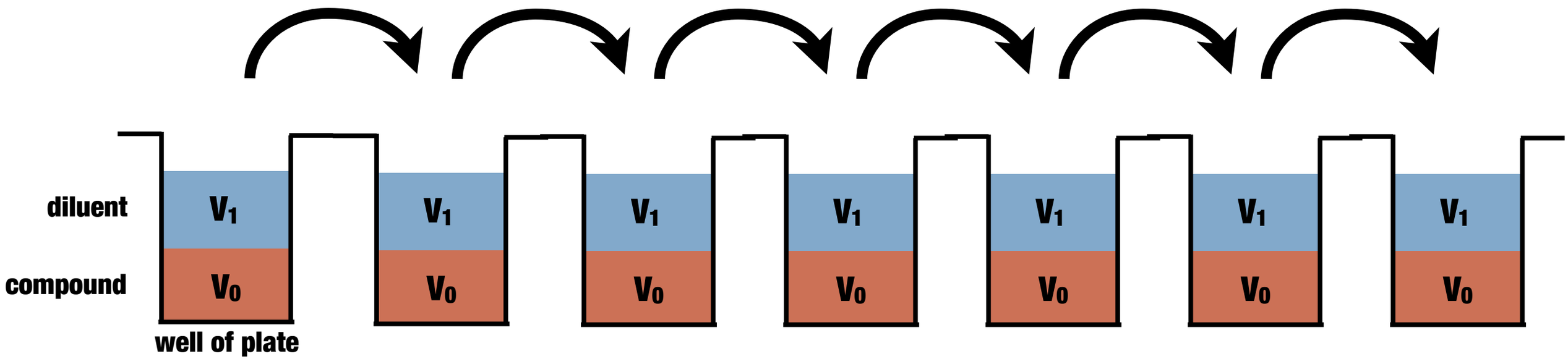


IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

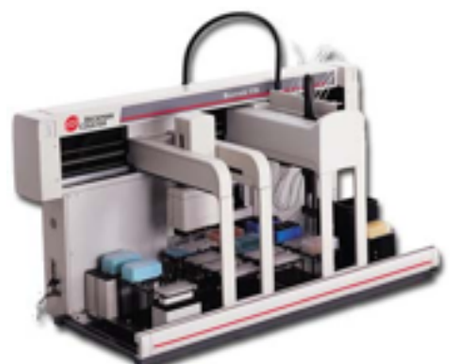


Biomek NX/FX

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

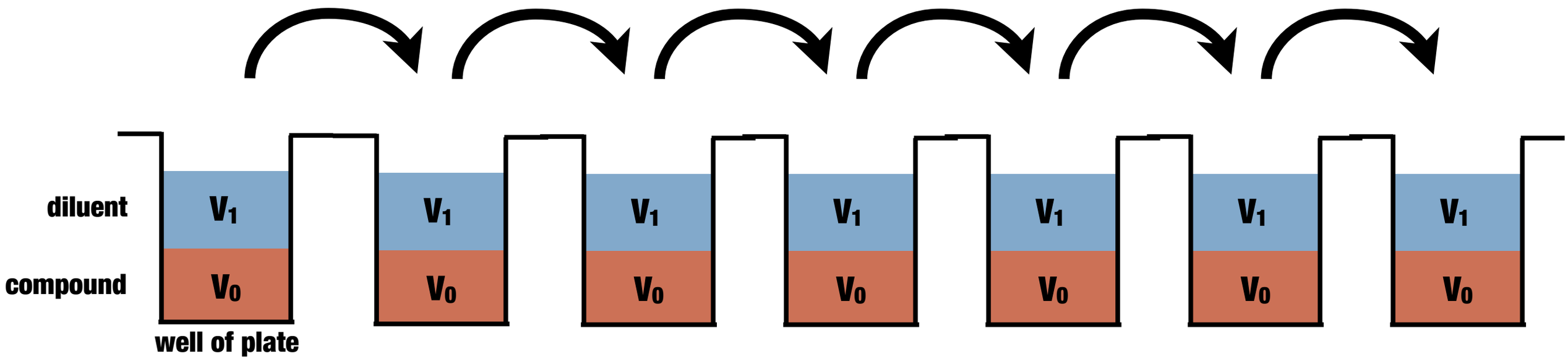


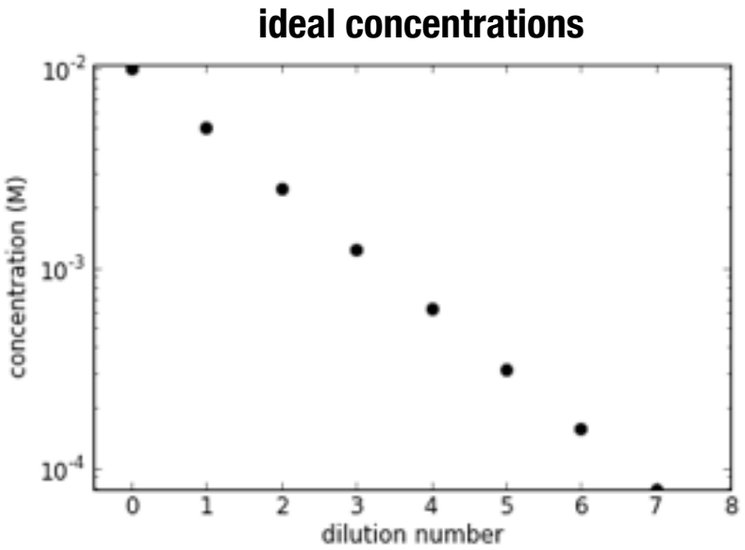
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013



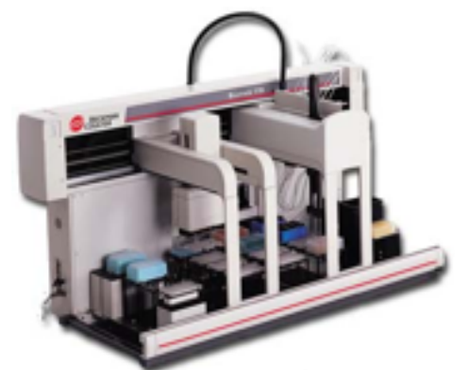
Biomek NX/FX

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



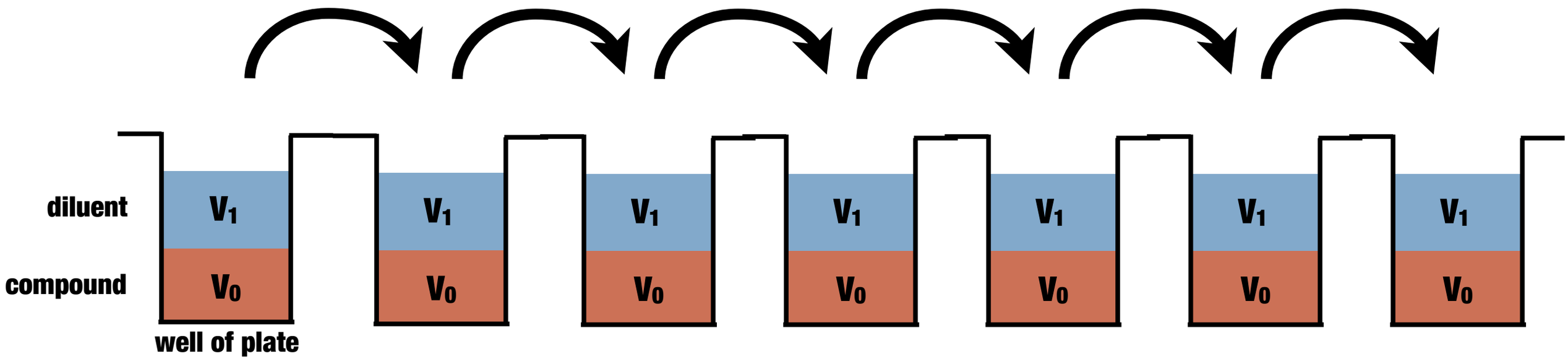


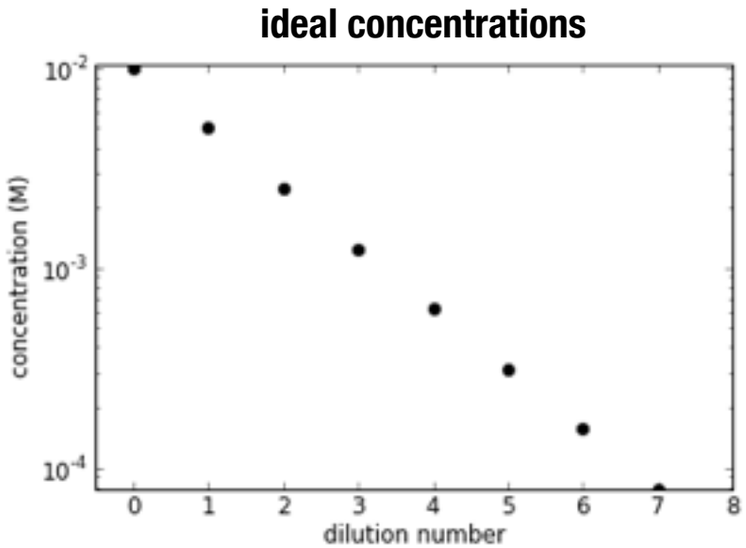
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

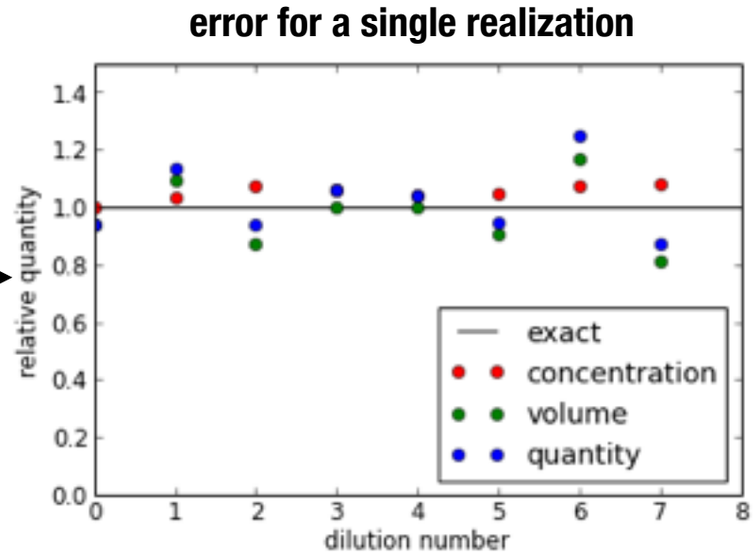


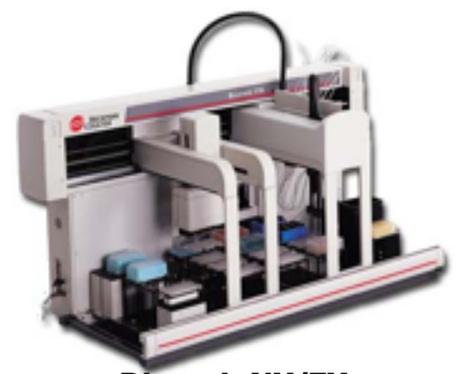
Biomek NX/FX

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



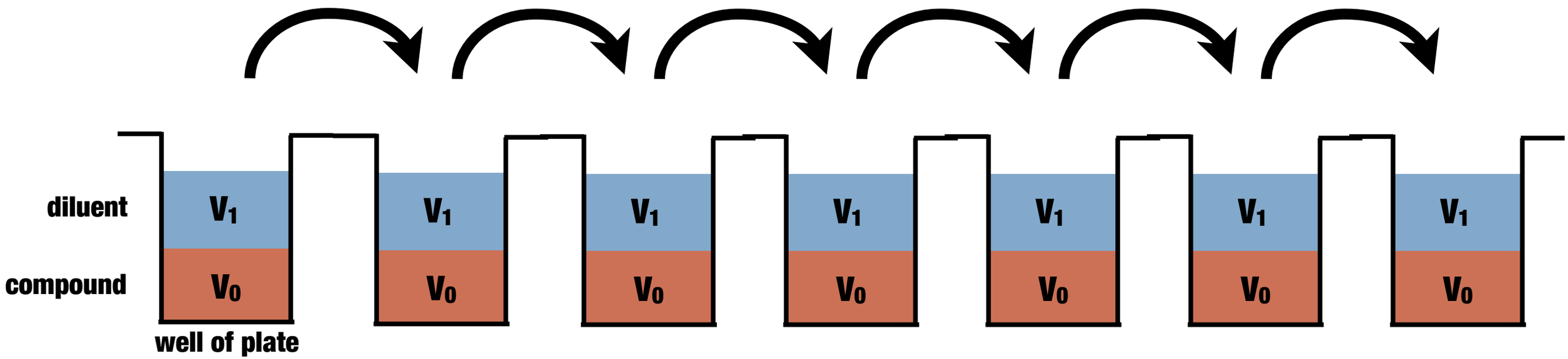




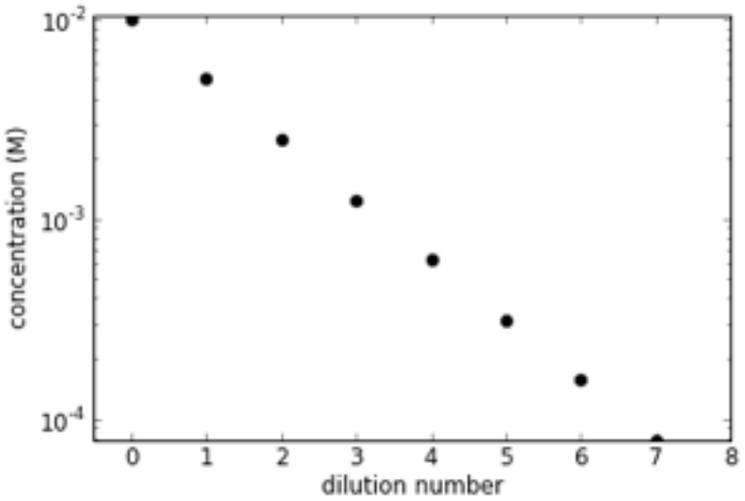

Biomek NX/FX

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

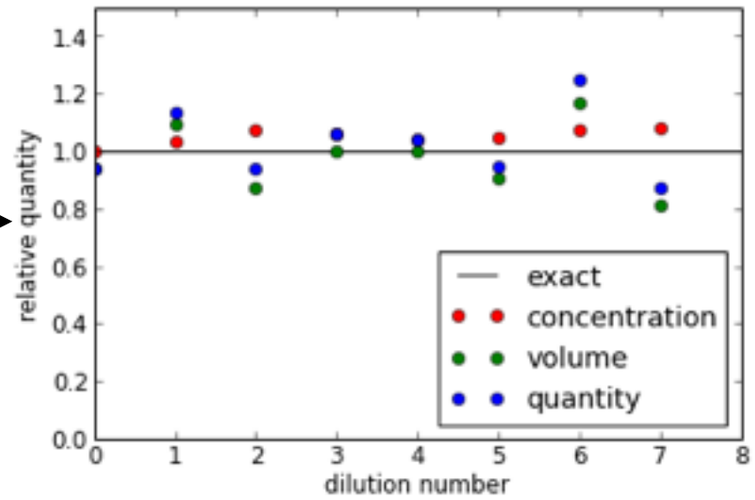
Modeling experimental error can be a valuable exercise: Preparing a dilution series via liquid-handling robot



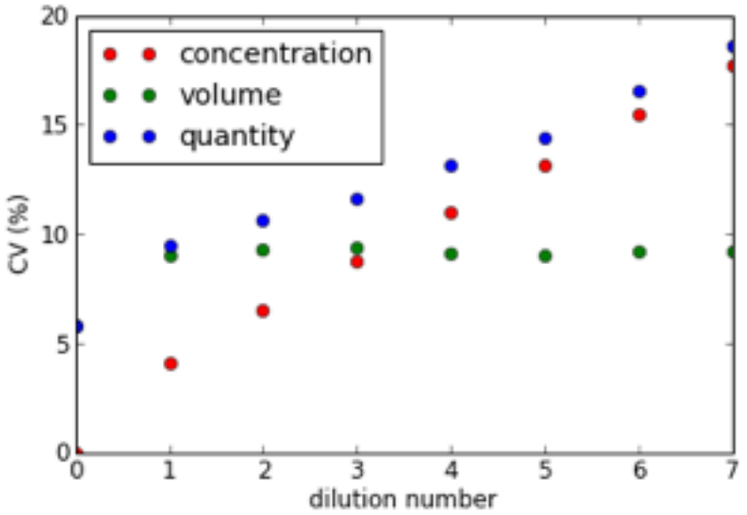
ideal concentrations



error for a single realization



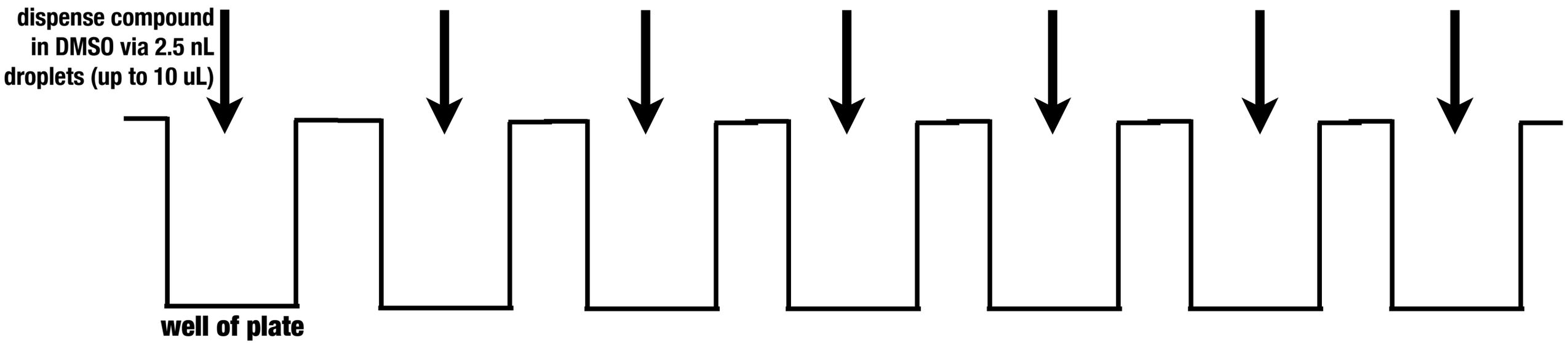
coefficients of variation (CV)



Biomek NX/FX

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

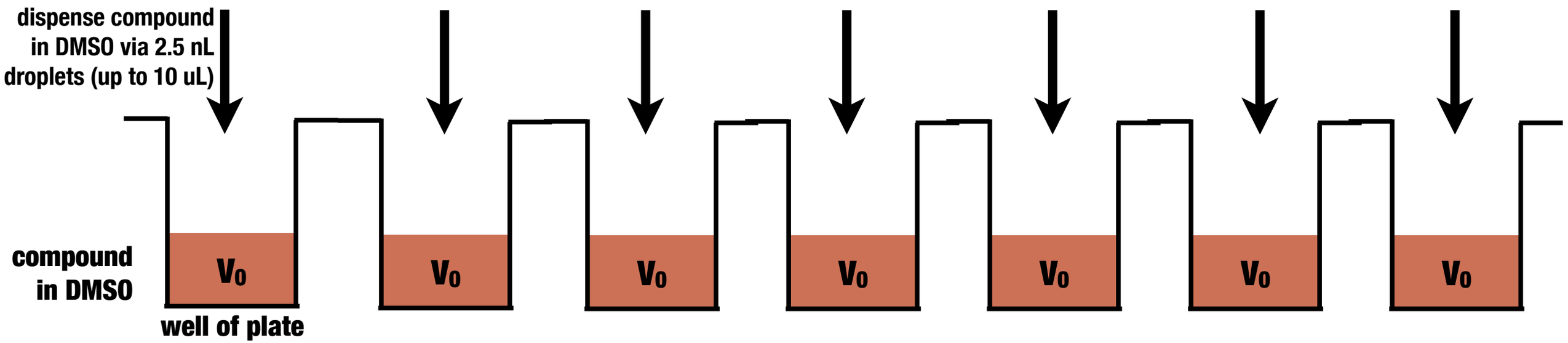
Modeling experimental error can be a valuable exercise: Preparing a dilution series via acoustic dispensing



IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

LabCyte Echo

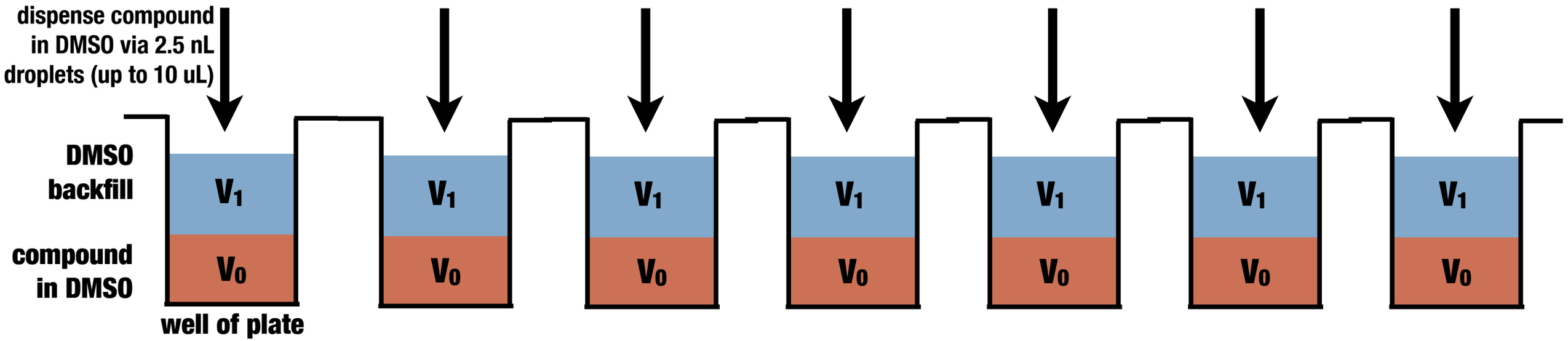
Modeling experimental error can be a valuable exercise: Preparing a dilution series via acoustic dispensing



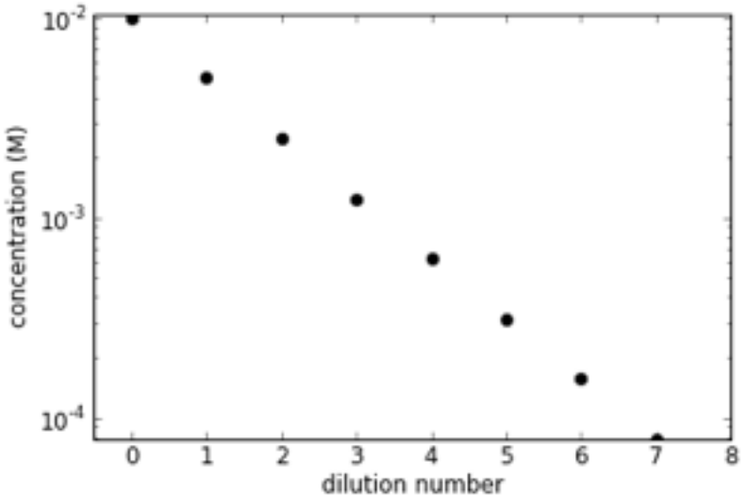
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

LabCyte Echo

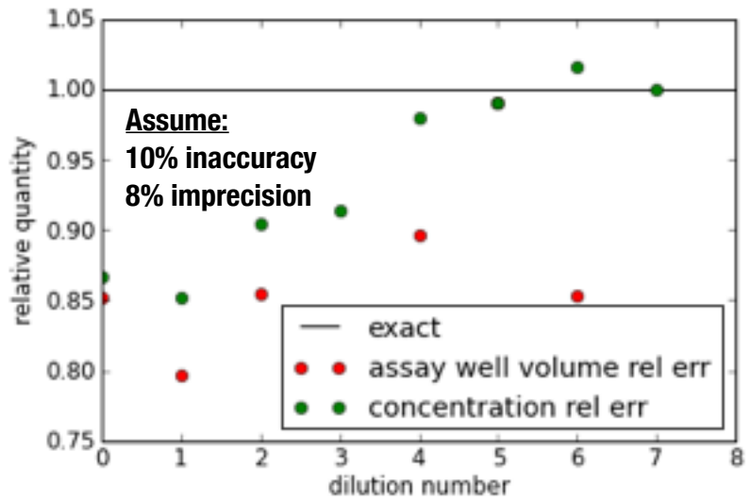
Modeling experimental error can be a valuable exercise: Preparing a dilution series via acoustic dispensing



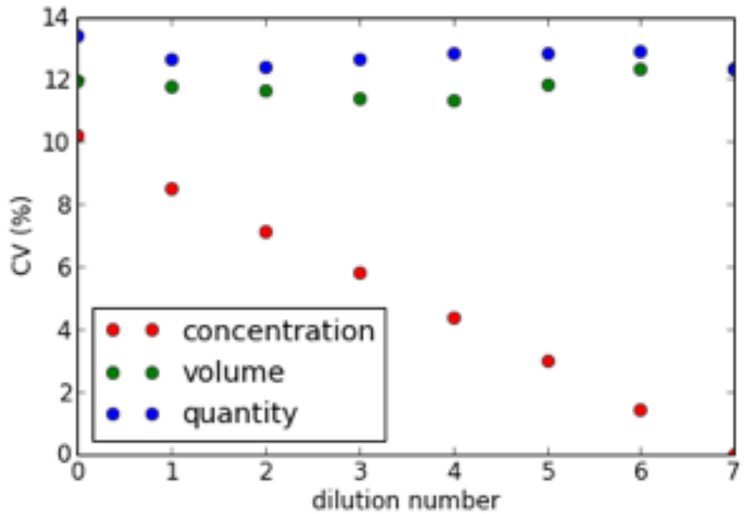
ideal concentrations



error for a single realization



coefficients of variation (CV)

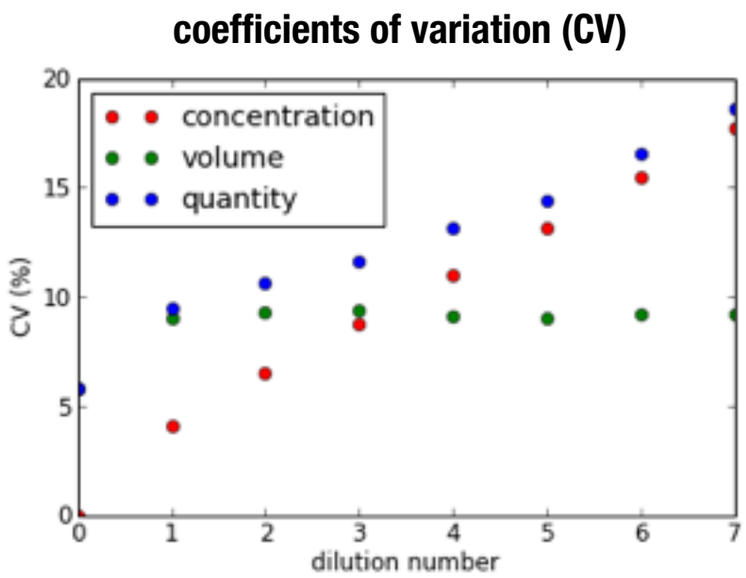


http://www.labcyte.com/sites/default/files/support_docs/Echo%205XX%20Specifications.pdf

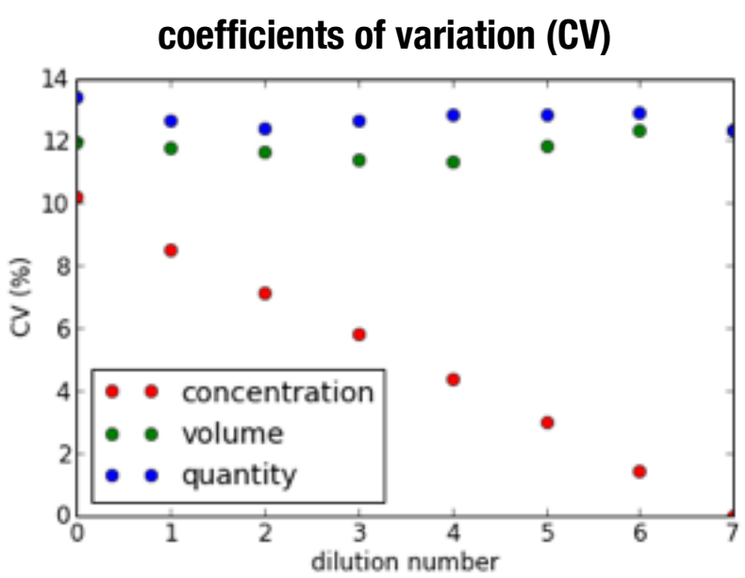
LabCyte Echo

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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



tip-based



acoustic

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

Modeling experimental error can be a valuable exercise: Assay model

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

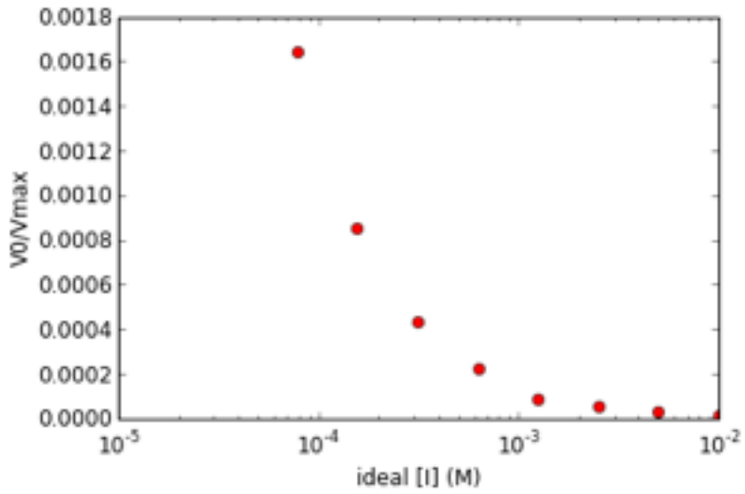
Assume for EphB4 assay:
Michaelis-Menten kinetics
competitive inhibition
[ATP] = 4 uM
[EphB4] = 6 uM
Km(ATP) = 1.71 uM

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

Modeling experimental error can be a valuable exercise: Assay model

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

measure initial reaction velocity



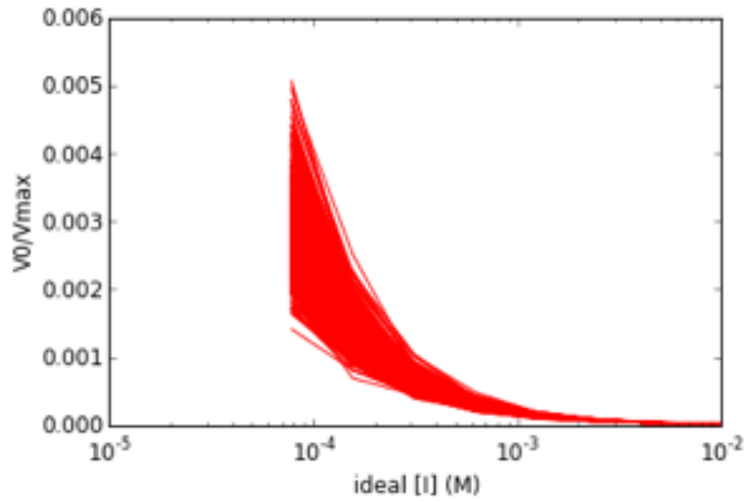
Assume for EphB4 assay:
Michaelis-Menten kinetics
competitive inhibition
[ATP] = 4 uM
[EphB4] = 6 uM
Km(ATP) = 1.71 uM

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

Modeling experimental error can be a valuable exercise: Assay model

<u>Dispense into assay plate:</u>
2 uL compound dilution
10 uL enzyme assay mix

measure initial reaction velocity



fit it to get IC50

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

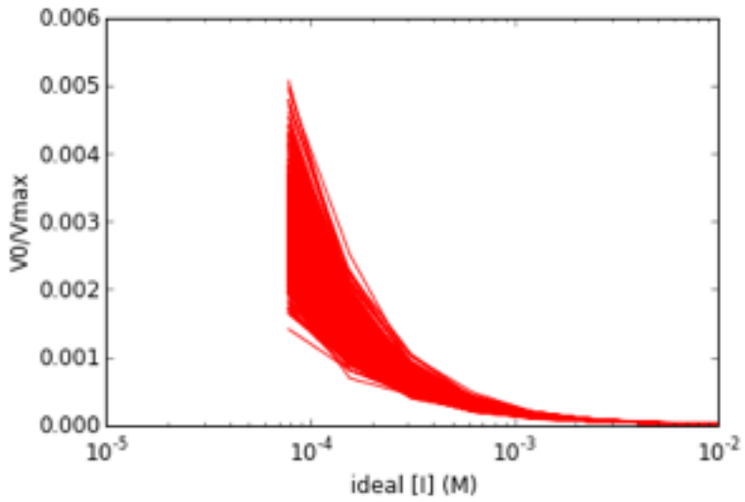
<u>IPython tutorial available at</u> https://github.com/choderalab/cadd-grc-2013

Modeling experimental error can be a valuable exercise: Assay model

<u>Dispense into assay plate:</u>
2 uL compound dilution
10 uL enzyme assay mix

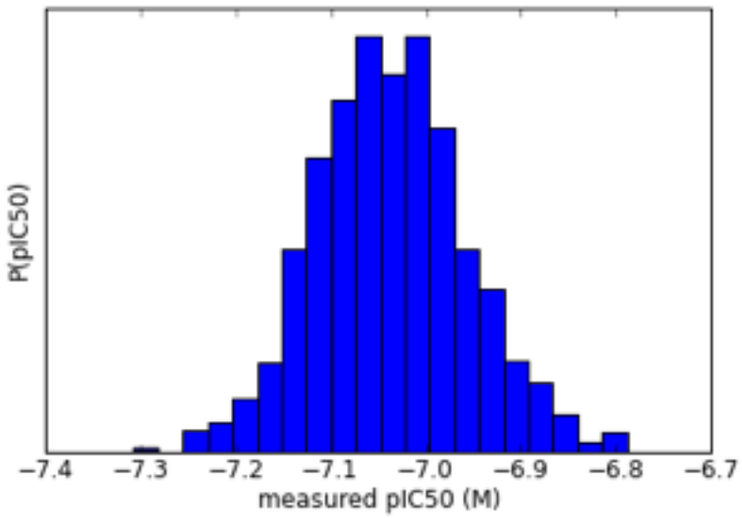
<u>Assume for EphB4 assay:</u>
Michaelis-Menten kinetics
competitive inhibition
[ATP] = 4 uM
[EphB4] = 6 uM
Km(ATP) = 1.71 uM

measure initial reaction velocity



fit it to get IC50

distribution of fit pIC50s



<u>IPython tutorial available at</u> https://github.com/choderalab/cadd-grc-2013

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 »

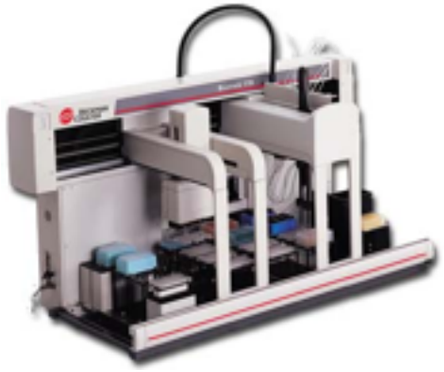
May 3, 2013

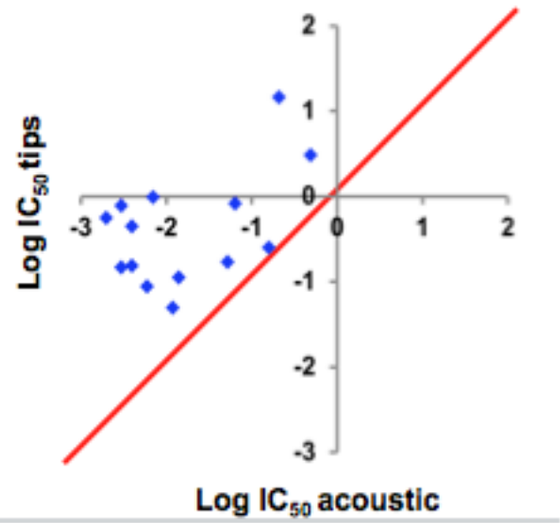
Drug Assay Numbers, All Over the Place

Posted by **Derek**

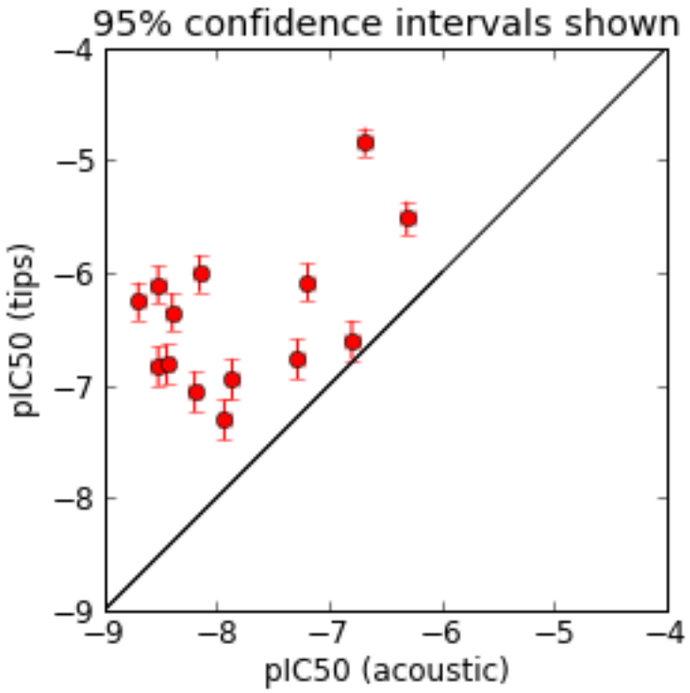
There's a truly disturbing paper out in *PLoSONE* with potential implications for a lot of assay data out there in the literature. The authors are looking at the results of biochemical assays as a function of how the compounds are dispensed in them, pipet tip versus acoustic, which is the sort of idea 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?

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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

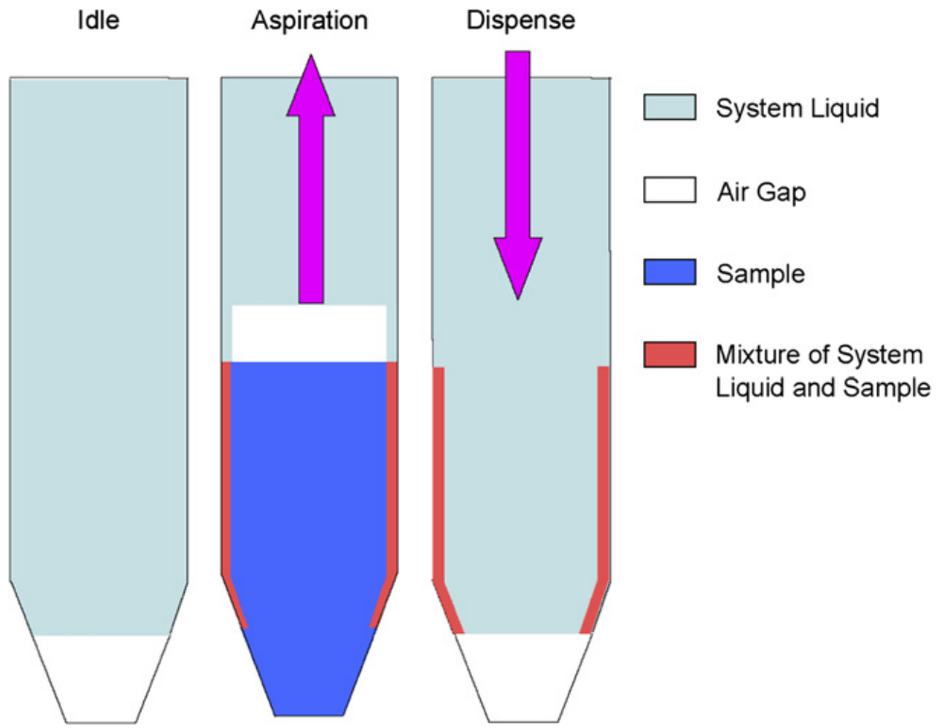


Figure 1. Schematic diagram of the dilution effect.

Dong, Ouyang, Liu, and Jemal. JALA 11:60, 2006. (BMS)

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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

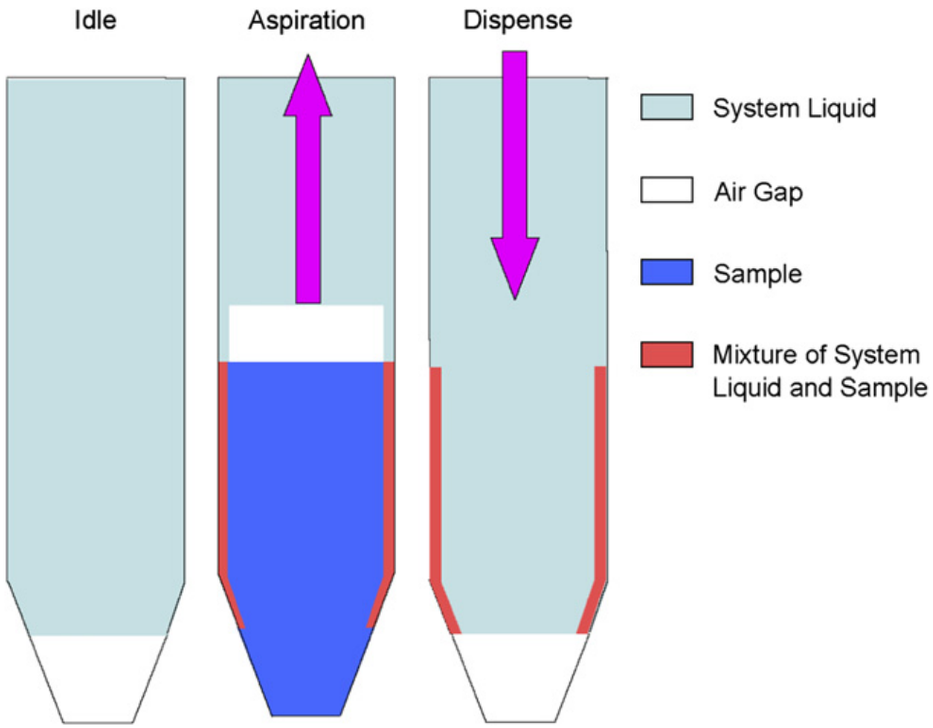


Figure 1. Schematic diagram of the dilution effect.

Table 1. Comparison between the MVS method and the gravimetry method—Tecan ALH using water liquid class

	MVS (μL)		Gravimetry (μL)	
Target volume	20	200	20	200
Mean volume (μL)	18.74	190.08	20.15	199.53
Inaccuracy (%)	−6.30	−4.96	0.75	−0.23
StDev	0.22	1.74	0.6	0.55
CV ($n = 96$) (%)	1.17	0.92	1.94	0.59

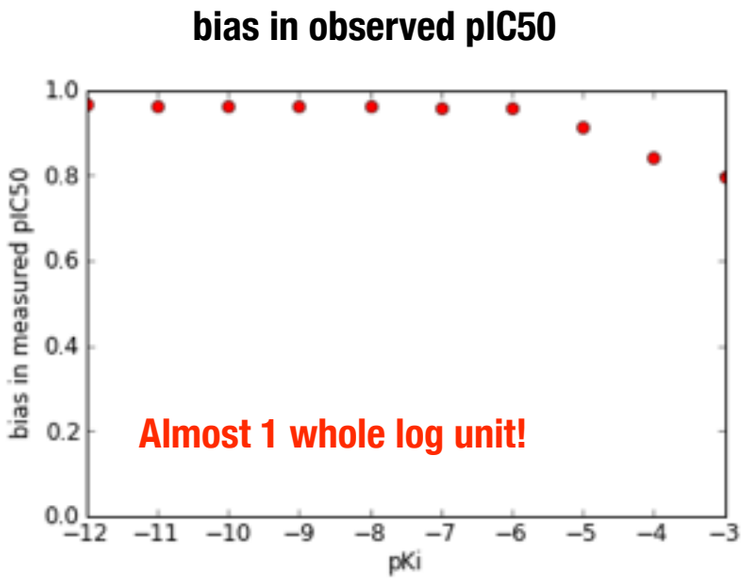
These solutions were delivered by the Tecan ALH using an aspirate/dispense protocol based on a water liquid class. Inaccuracy corresponds to 100 × (Observed volume − Target volume)/Target volume. Coefficient of variation (CV) corresponds to 100 × StDev/mean, where StDev is the standard deviation.

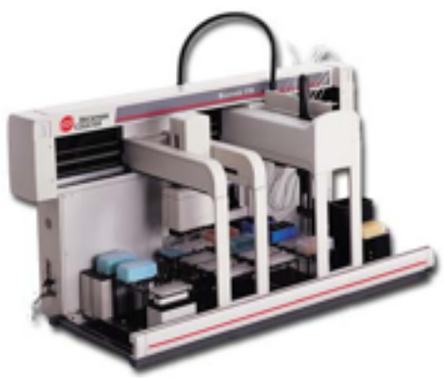
mixture of dyes **weight**

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

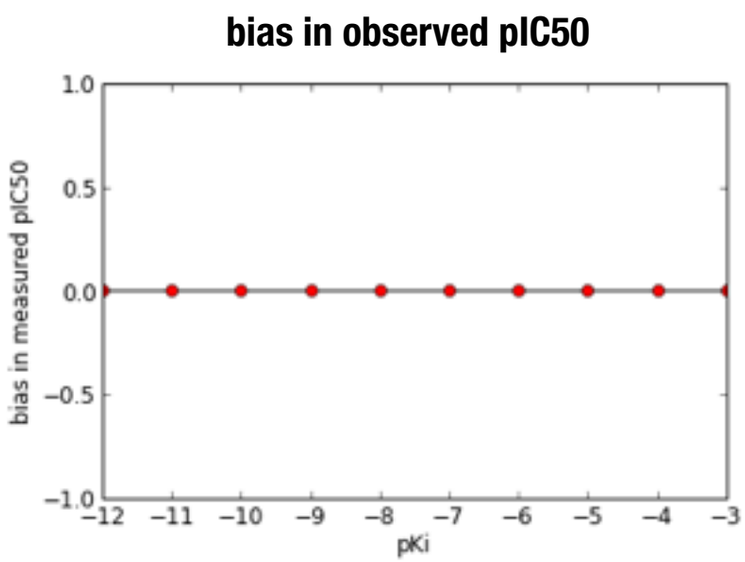
IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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





tip-based



acoustic

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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 »

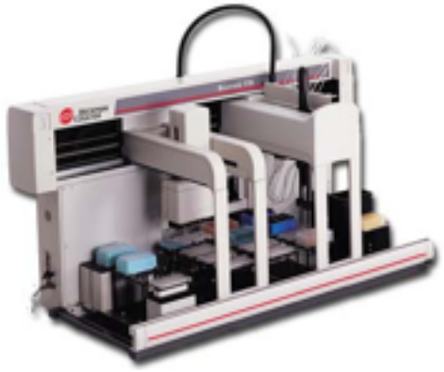
May 3, 2013

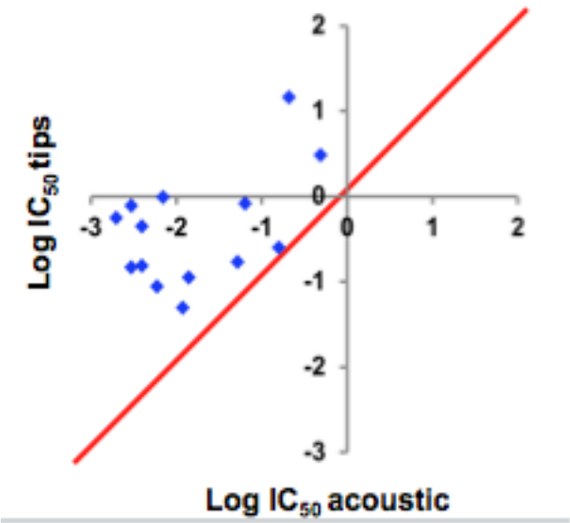
Drug Assay Numbers, All Over the Place

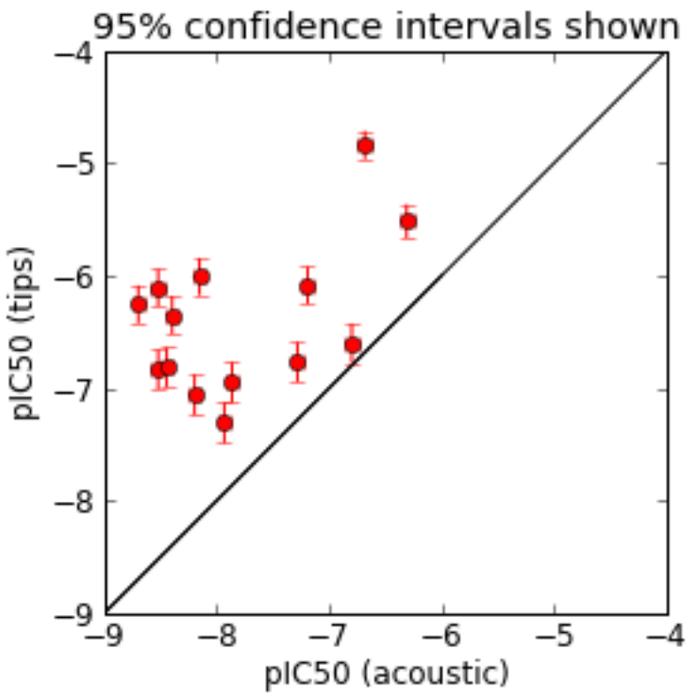
Posted by **Derek**

There's a truly disturbing paper out in *PLoSONE* 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.

IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

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 »

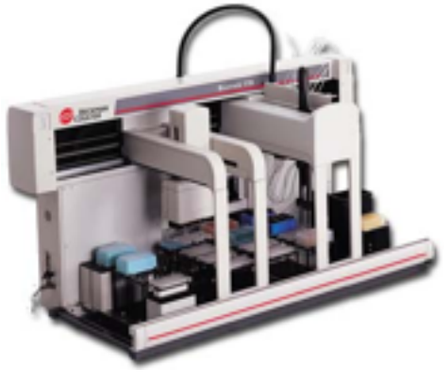
May 3, 2013

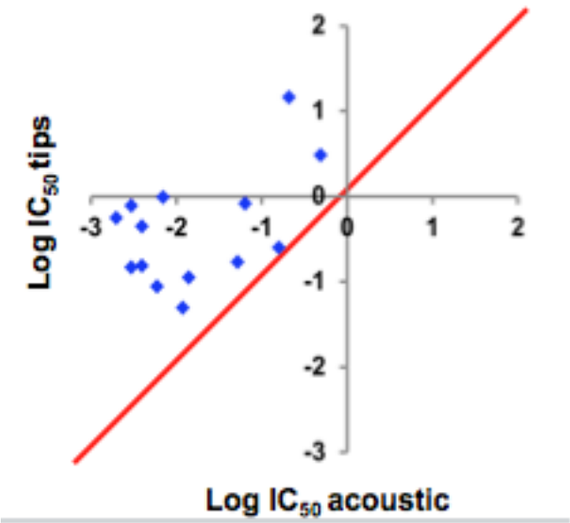
Drug Assay Numbers, All Over the Place

Posted by **Derek**

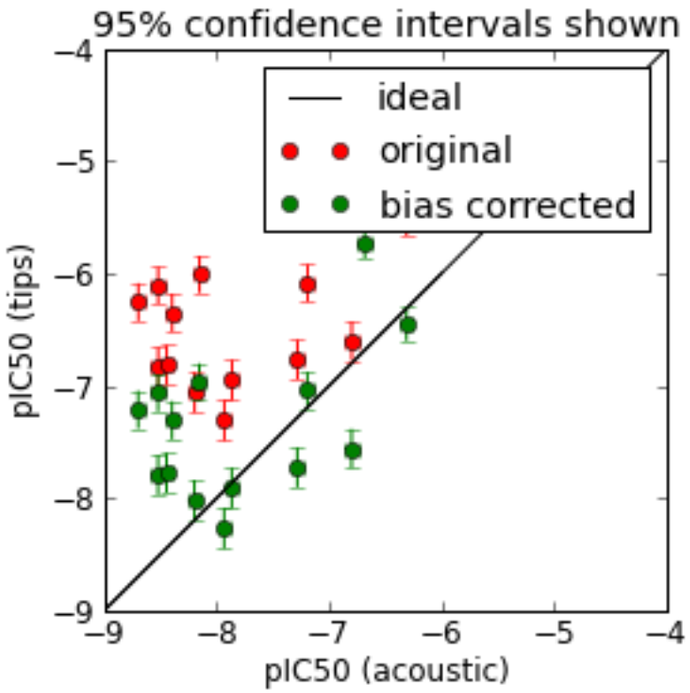
There's a truly disturbing paper out in *PLoSONE* with potential implications for a lot of assay data out there in the literature. The authors are looking at the results of biochemical assays as a function of how the compounds are dispensed in them, pipet tip versus acoustic, which is the sort of idea 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



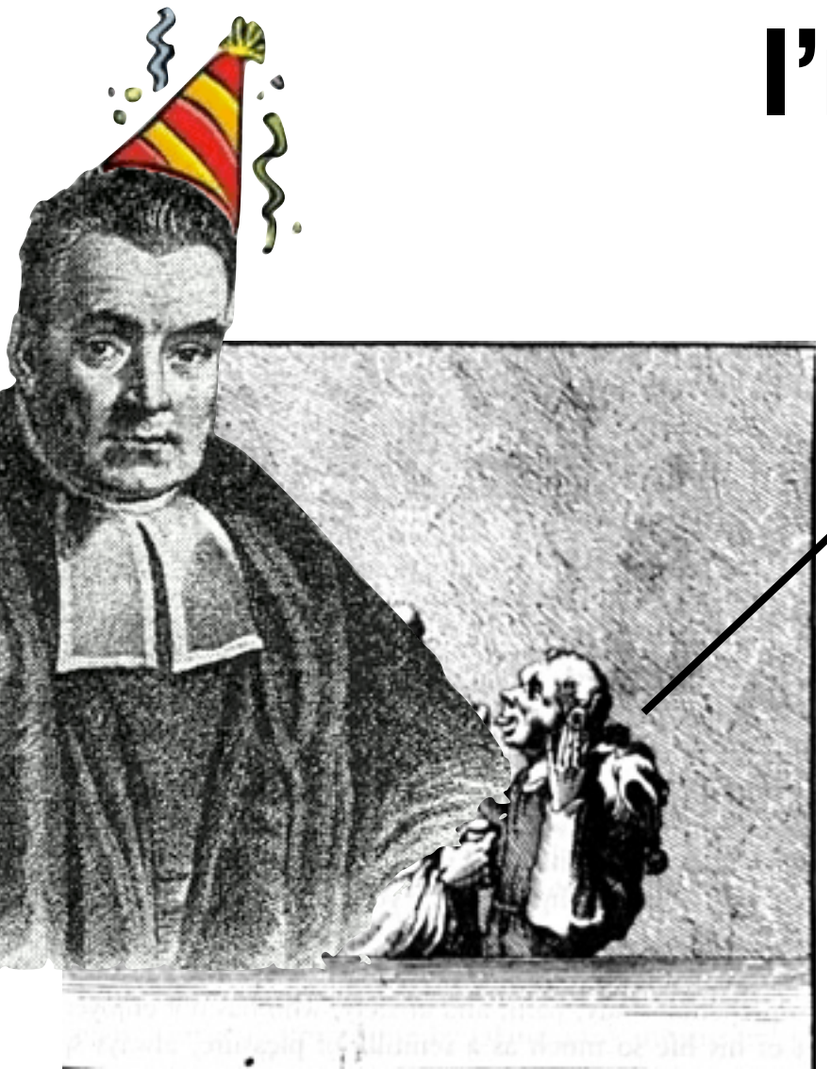


Elkins et al. PLoS One 8:e62325, 2013.



IPython tutorial available at https://github.com/choderalab/cadd-grc-2013

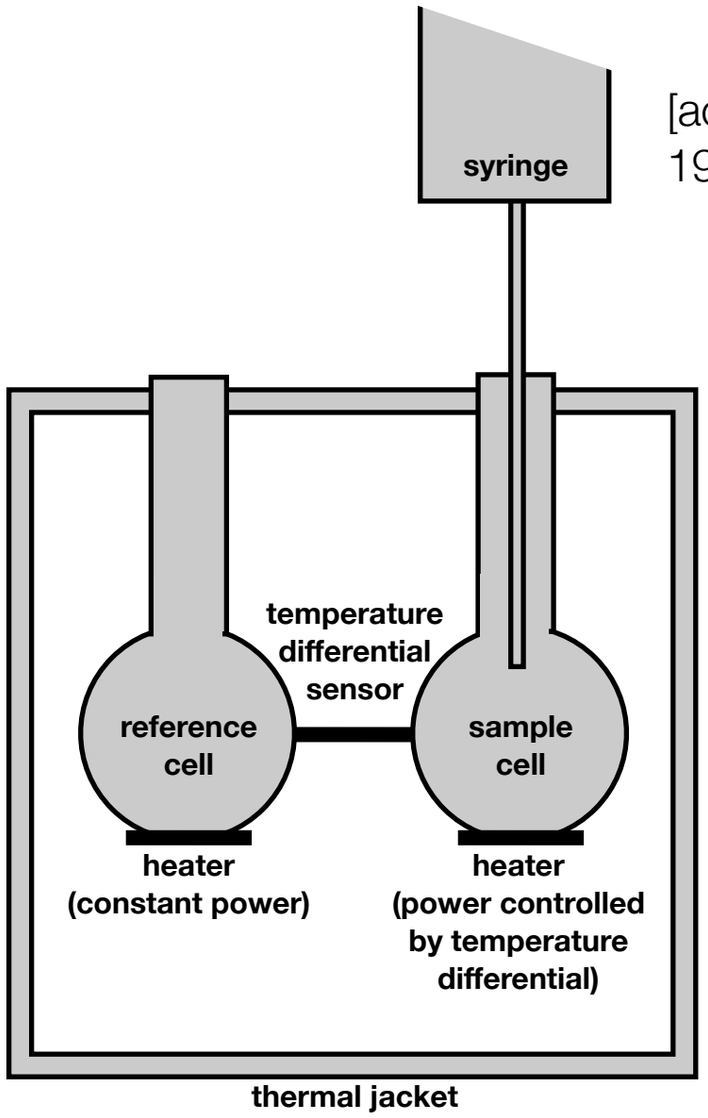
OK, that's all well and good, but how can we use Bayesian techniques to estimate the uncertainty in experimental data?



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

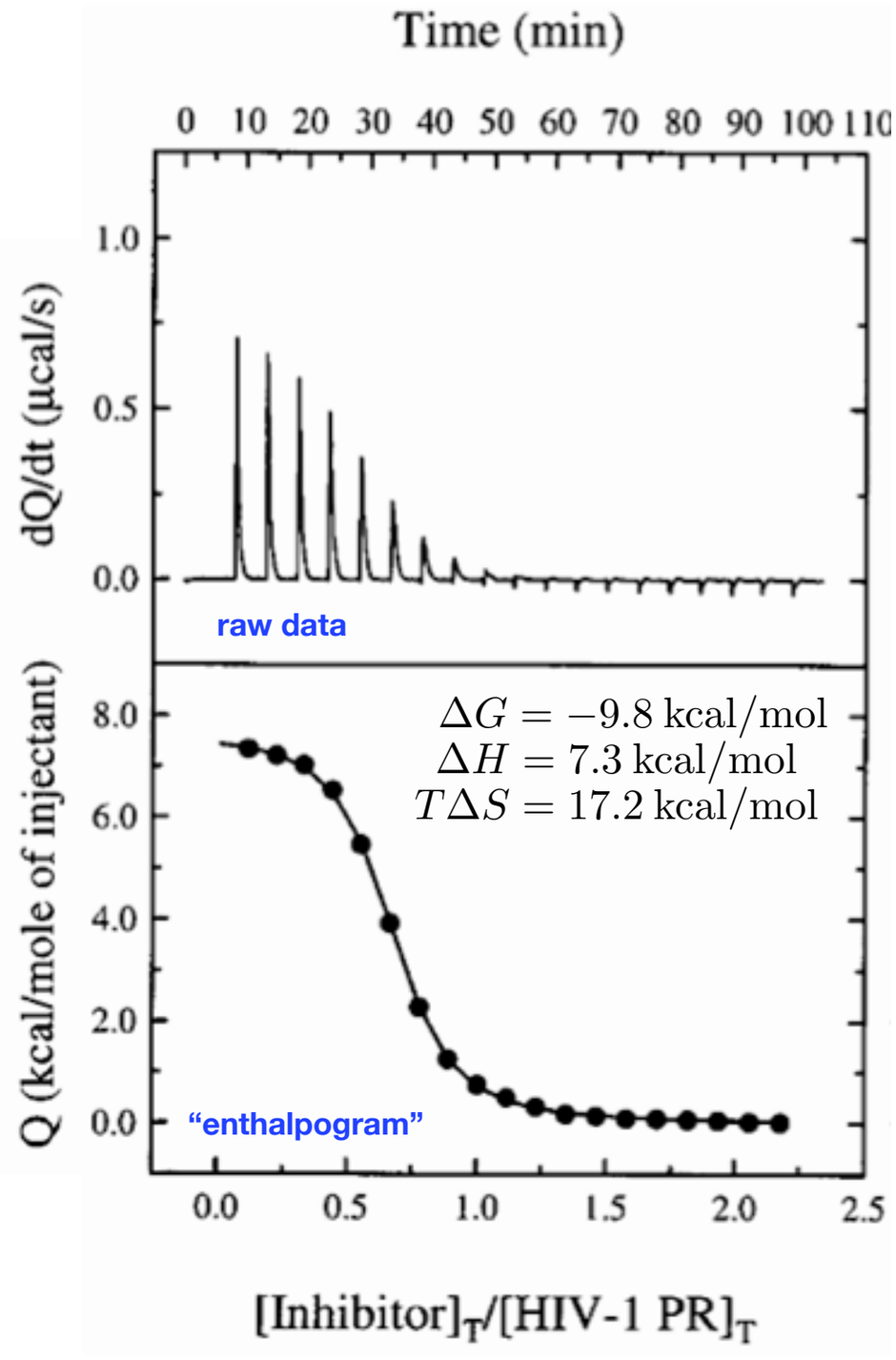
Isothermal titration calorimetry (ITC) can simultaneously interrogate free energies and enthalpies of binding

[illustrative experiment]



[acetyl pepstatin] = 200 uM
19 injections @ 10 uL/injection

[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.

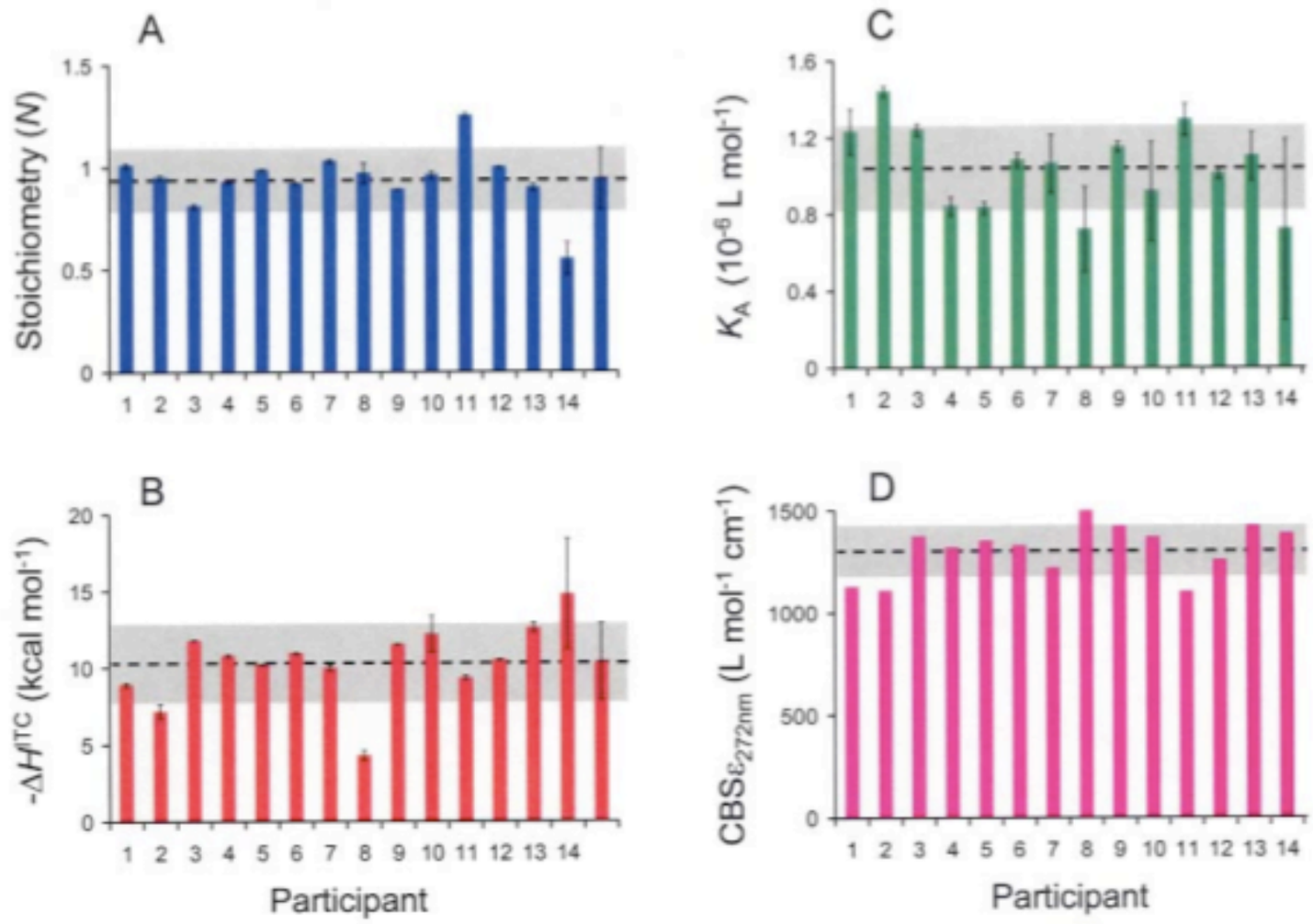


FIGURE 5

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 ± 2.5 kcal mol^{-1}). **C:** Affinity [(1.00 ± 0.22) × 10^6 L mol^{-1}]. **D:** molar extinction coefficient for CBS at 272 nm (1307 ± 126 L mol^{-1} 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.

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

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

TABLE 3

Summary of Reported Isothermal Titration Calorimetry Results[a]

Participant	Molar Binding Ratio (N)	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.

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

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

TABLE 3

Summary of Reported Isothermal Titration Calorimetry Results[a]

Participant	Molar Binding Ratio (*N*)	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.

ACTUAL STANDARD DEVIATION **± 0.2 µM** **± 2 kcal/mol**

± 0.1 kcal/mol

The reported error bars cannot be trusted.

They're often an order of magnitude (or more!) too small.

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

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

TABLE 3

Summary of Reported Isothermal Titration Calorimetry Results[a]

Participant	Molar Binding Ratio (*N*)	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.

ACTUAL STANDARD DEVIATION **± 0.2 µM** **± 2 kcal/mol**

± 0.1 kcal/mol

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.

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

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

TABLE 3

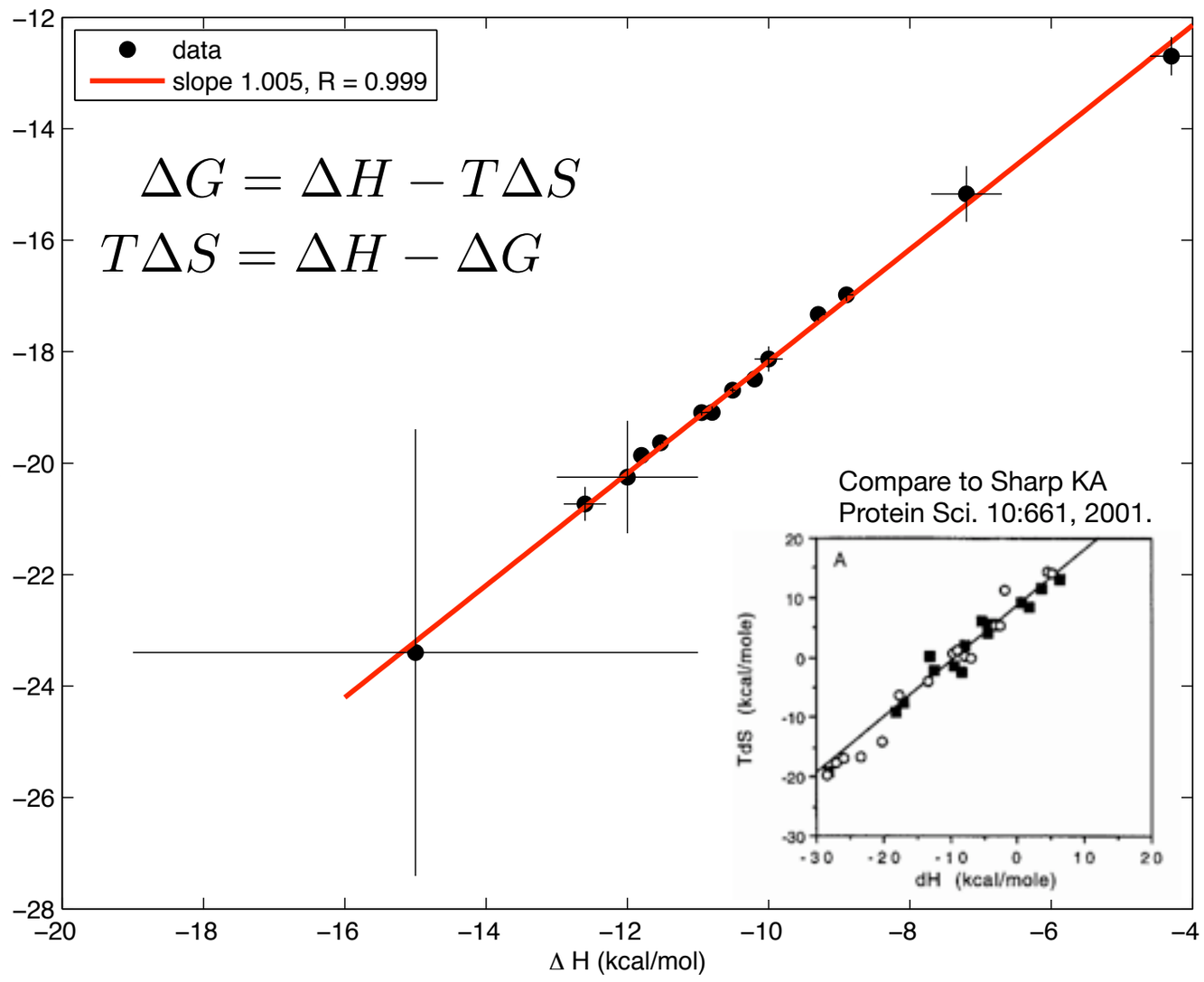
Summary of Reported Isothermal Titration Calorimetry Results[a]

Participant	Molar Binding Ratio (N)	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 analysi
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 analysi
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 analysi
6	0.921 ± 0.003	1.08 ± 0.04	−10.95 ± 0.05	39	CBS int lysi
7	1.03 ± 0.01	1.1 ± 0.2	−10.0 ± 0.2	55	CBS int
8	0.97 ± 0.05	0.7 ± 0.2	−4.3 ± 0.3	7.0	CBS int Both subtracted as part of lysis.
9	0.891 ± 0.002	1.15 ± 0.03	−11.53 ± 0.04	59	CBS into buffer, average val btracted 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.

plot

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

CBS/CA II measurements from different laboratories



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

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

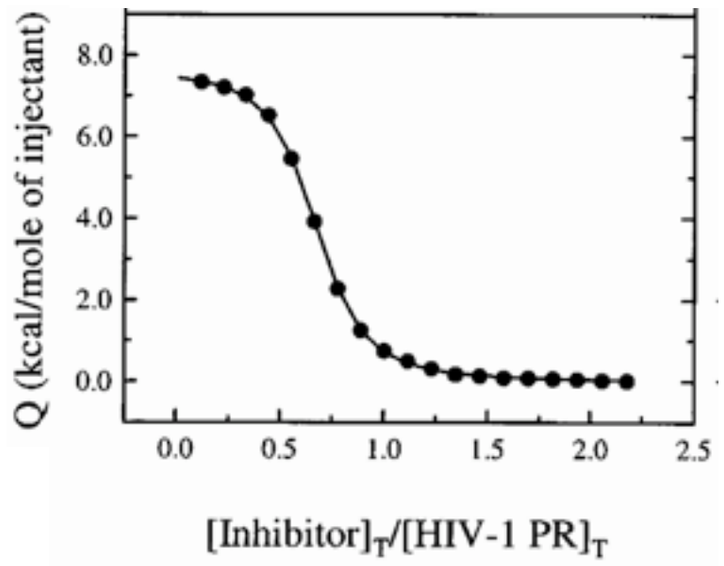
$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

Analysis of ITC experiments: The Bayesian way

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

$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

$\mathcal{D} = \{q_1, q_2, \ldots, q_N\}$ measurements of evolved heat



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

$\theta = \{\Delta G, \Delta H, T\Delta S, \Delta H_0\}$ thermodynamic parameters

$\Delta G = -kT \ln K_a$	free energy of binding
ΔH	enthalpy of binding
$T\Delta S$	entropic contribution to binding
ΔH_0	heat of dilution

... any additional parameters ...

Analysis of ITC experiments: The Bayesian way

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

$$p(\mathcal{D}|\theta) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(q_n - q_n^*)^2}{2\sigma^2}\right]$$ Gaussian error model

$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

q_n	measured heat of injection *n*
q_n^*	true heat of injection *n*
σ	std dev of error in measured heat (nuisance parameter)

$$P + L \overset{\Delta H}{\rightarrow} PL$$

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

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

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

$$p(\Delta G, \Delta H, \Delta H_0, \sigma) \propto \sigma^{-1}$$ prior on measurement noise

$$[M]_0^* \sim \mathcal{N}([M]_0, \sigma_M^2)$$
$$[L]_s^* \sim \mathcal{N}([L]_s, \sigma_L^2)$$
prior on cell and syringe concentrations

$\Delta G, \Delta H, \Delta H_0$ can be of any sign and value

$\sigma > 0$ scale parameter; can be of any magnitude

(Later, could build in some *a priori* 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

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

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

$\mathcal{D}$ data

θ model parameters

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

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

$p(\theta)$ prior

q_n^* nonlinear function of thermodynamic parameters

How can we sample from this awful function?

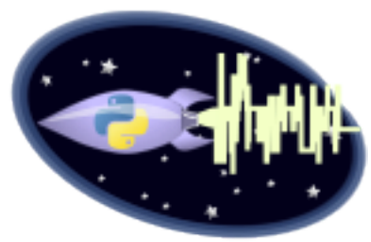
$$p(\theta|\mathcal{D}) = (2\pi)^{-N/2}\sigma^{-(N+1)}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(q_n - q_n^*)^2\right]$$ posterior

How can we sample from this awful function?

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

pymc is like ‘Bayes in a box’!

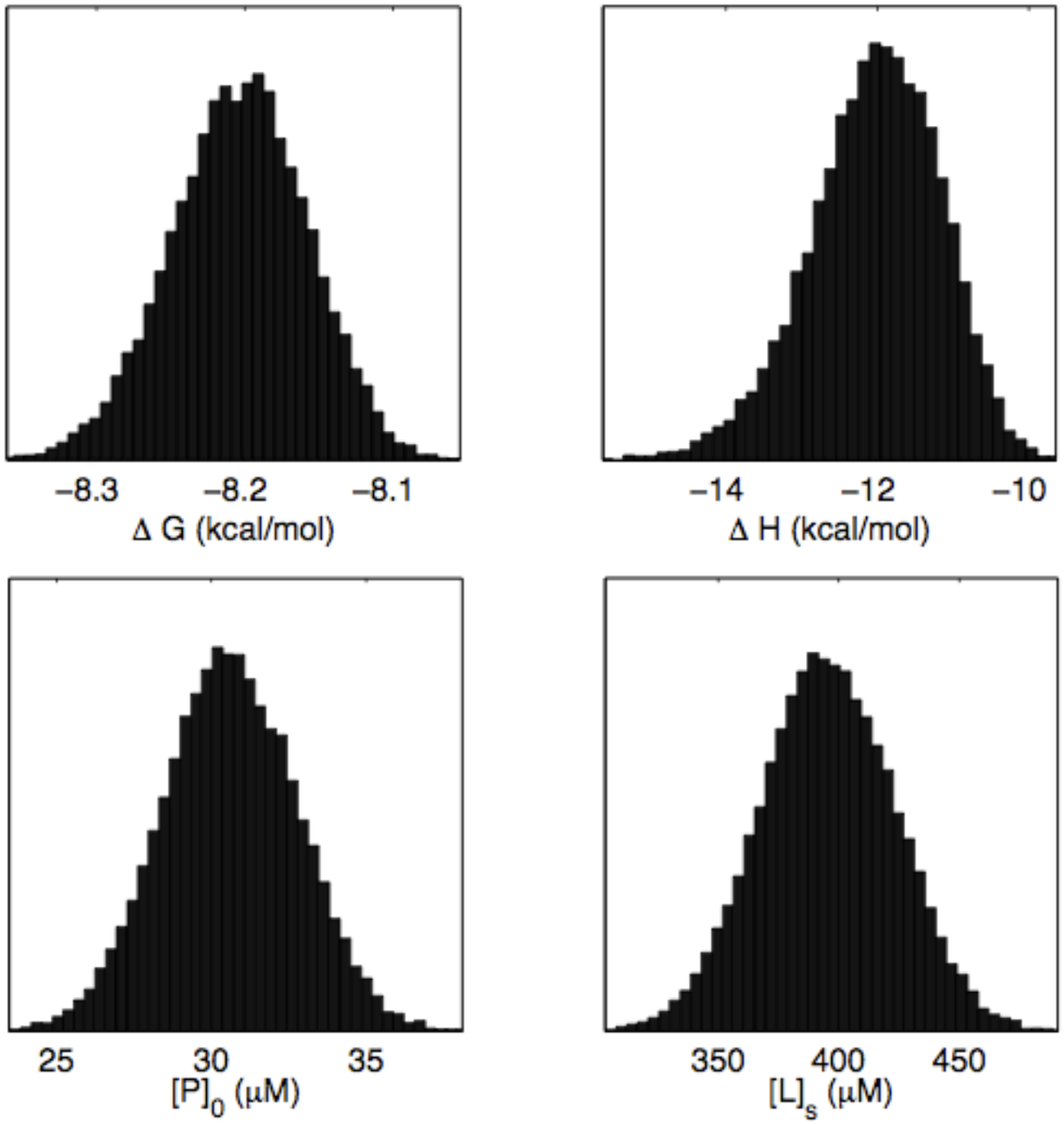
```
import pymc

# 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)
```

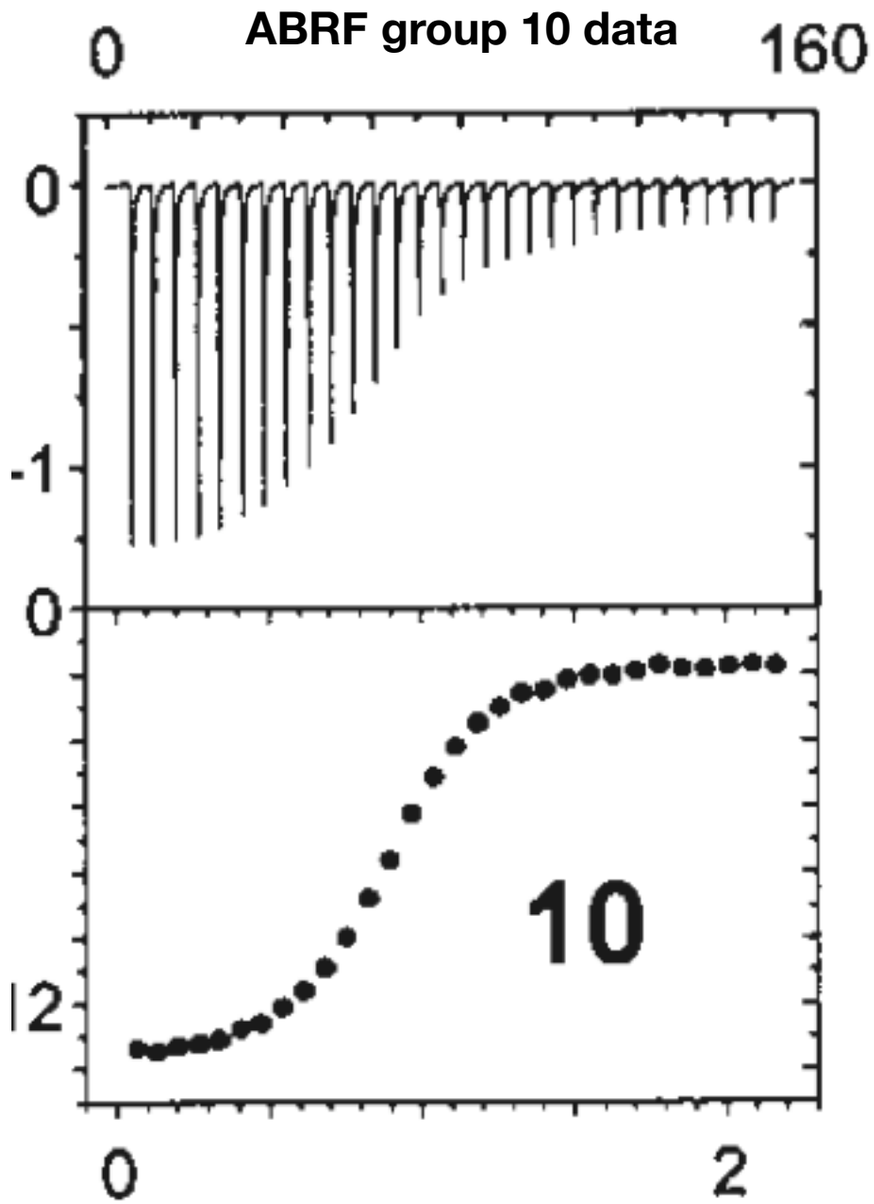


https://pypi.python.org/pypi/pymc

Analysis of ITC experiments: The Bayesian way

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

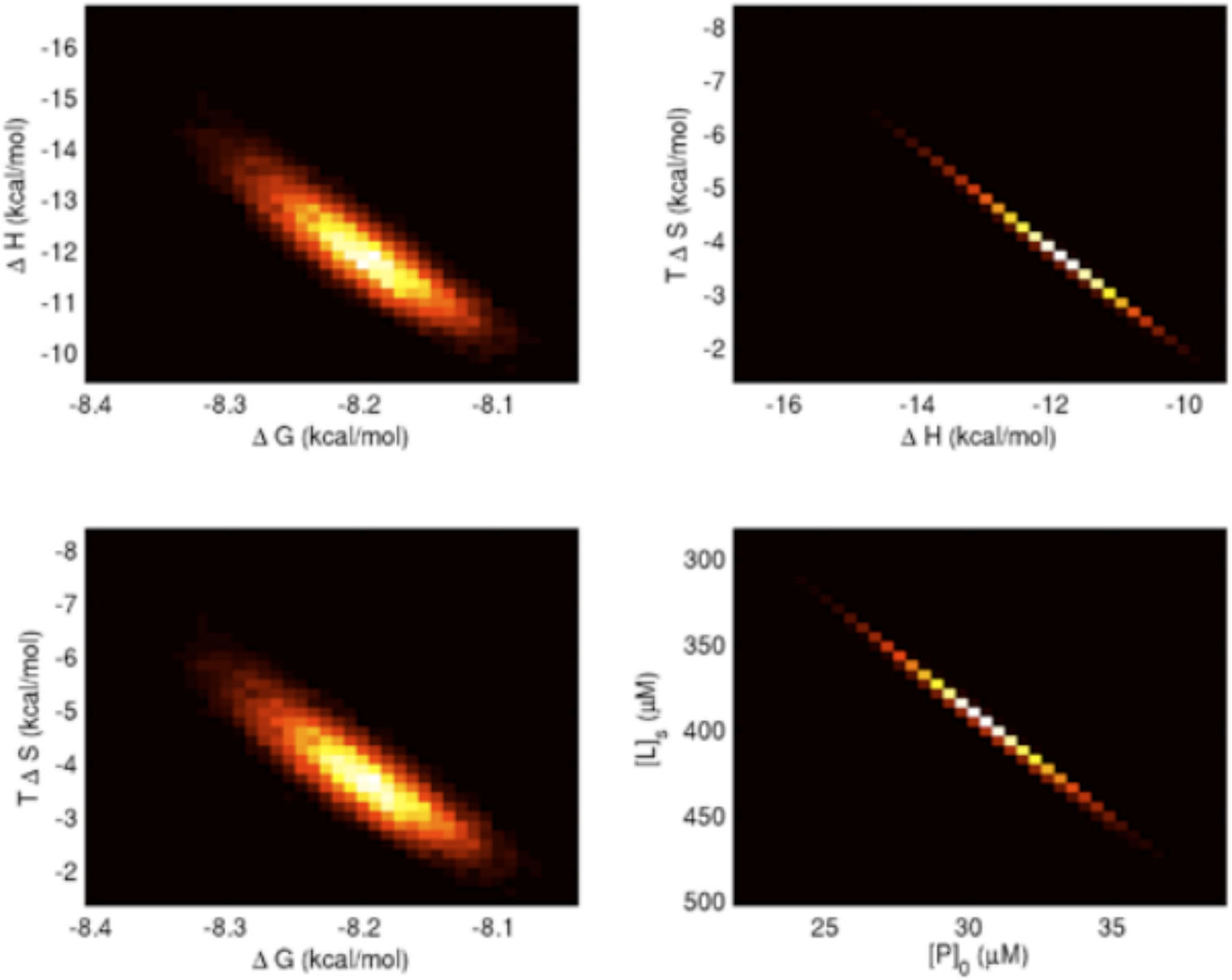




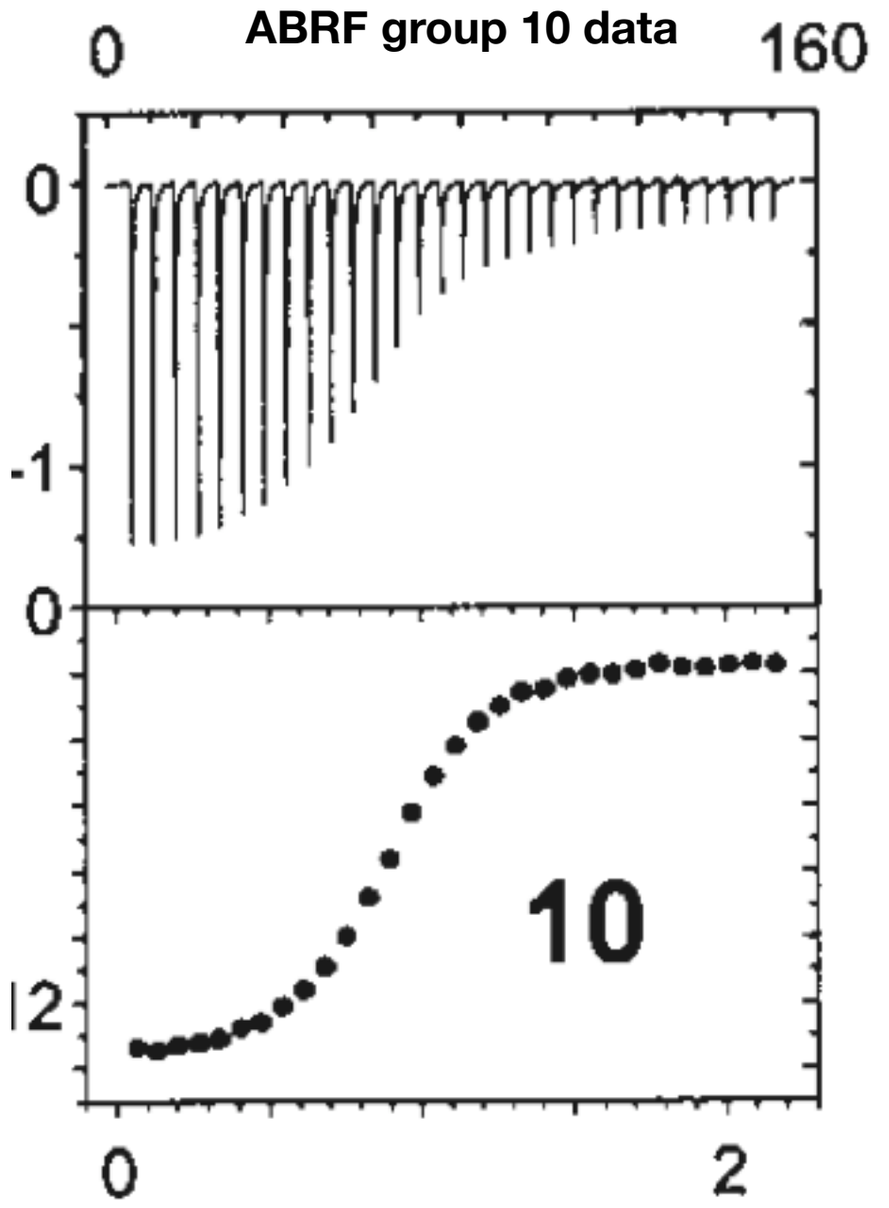
Myszka et al. J. Biomol. Tech. 14:247, 2003.

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 TΔS 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?"**

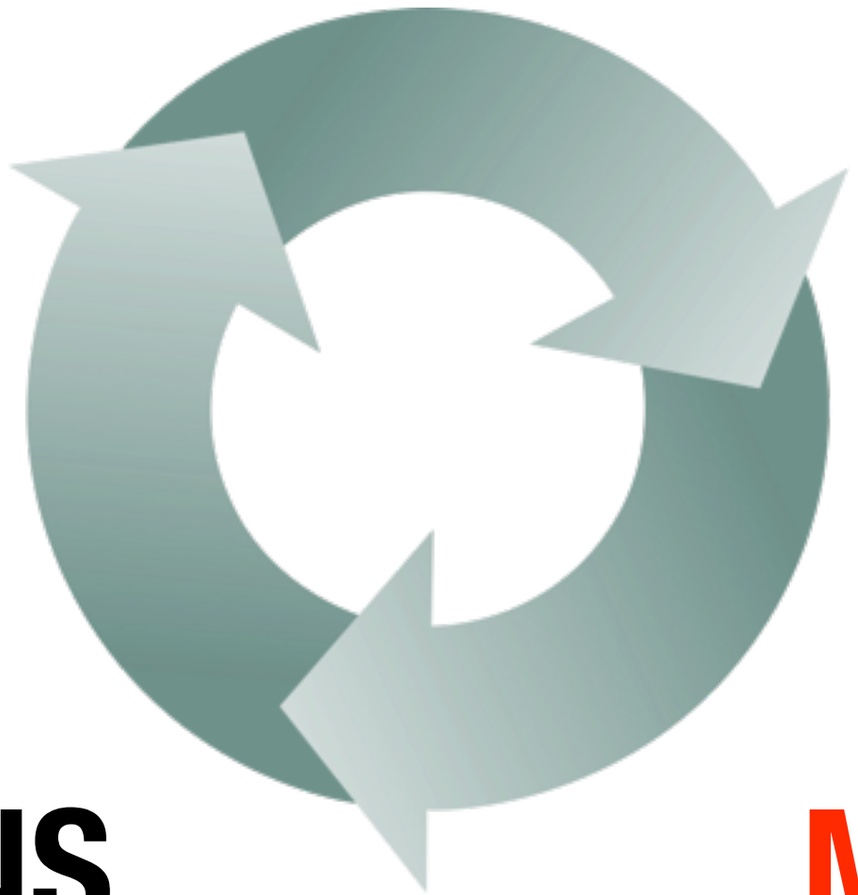
Code will be available at http://github.com/choderalab/bayesian-itc

The universal cycle of progress
Bayesian inference can drive progress

DATA

update knowledge conditioned on data

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



update model conditioned on data

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

DECISIONS

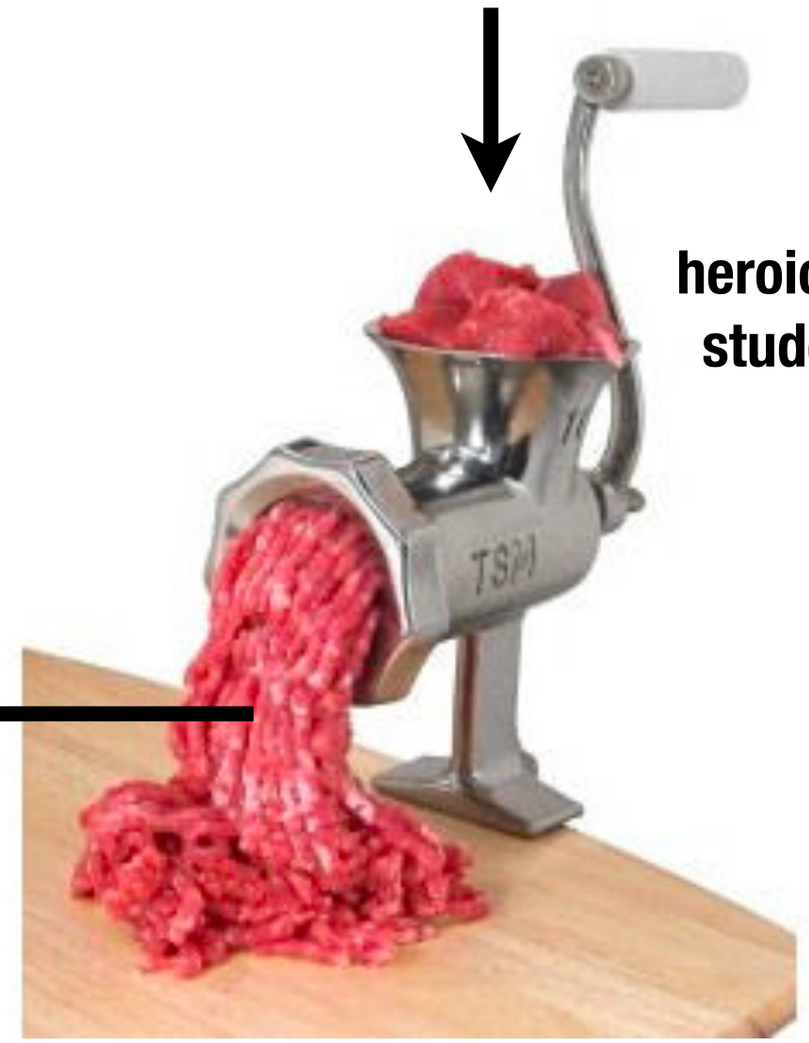
MODELS

make informed decisions based on model predictions

$$\text{argmax}_a \text{E}_a[f(\text{outcome})]$$

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

Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. JCTC 5:350, 2009.

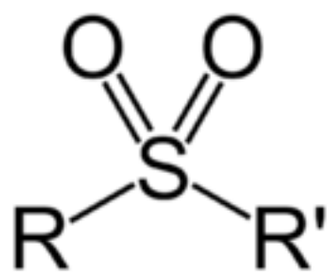
Poorly represented functional groups are problematic

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

Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. JCTC 5:350, 2009.

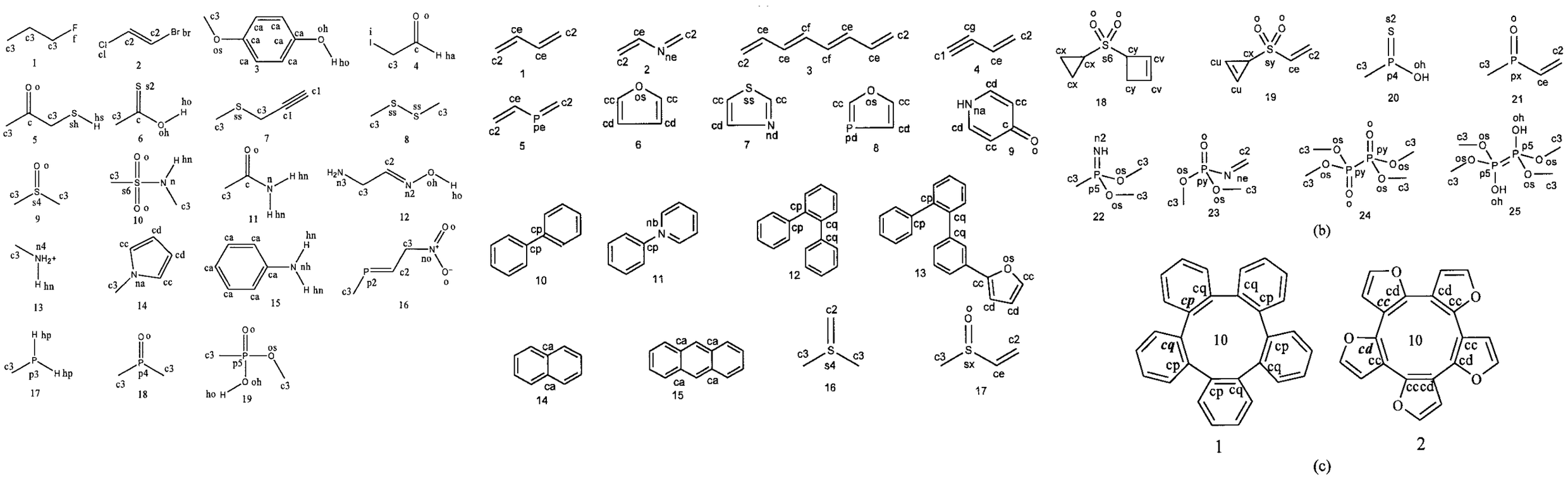
Poorly represented functionalities in parameterization set lead to pathologies in use.



sulfonyl moiety

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)

Wang J, Wolf RM, Caldwell JW, Kollman PA, and Case DA. J Comput Chem 25:1157, 2004.

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)

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

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

---

1. experimental/QM data
2. uncertainties/errors
3. spectroscopic model

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

MCMC

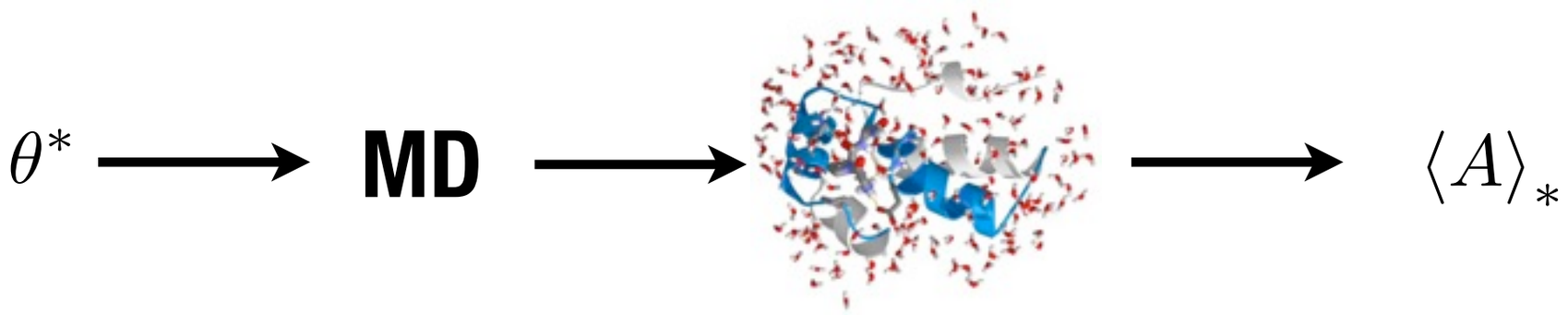
multiple parameter sets consistent with data

$$U_{\text{eff}}(\theta) \equiv -kT \log p(\theta|\mathcal{D})$$

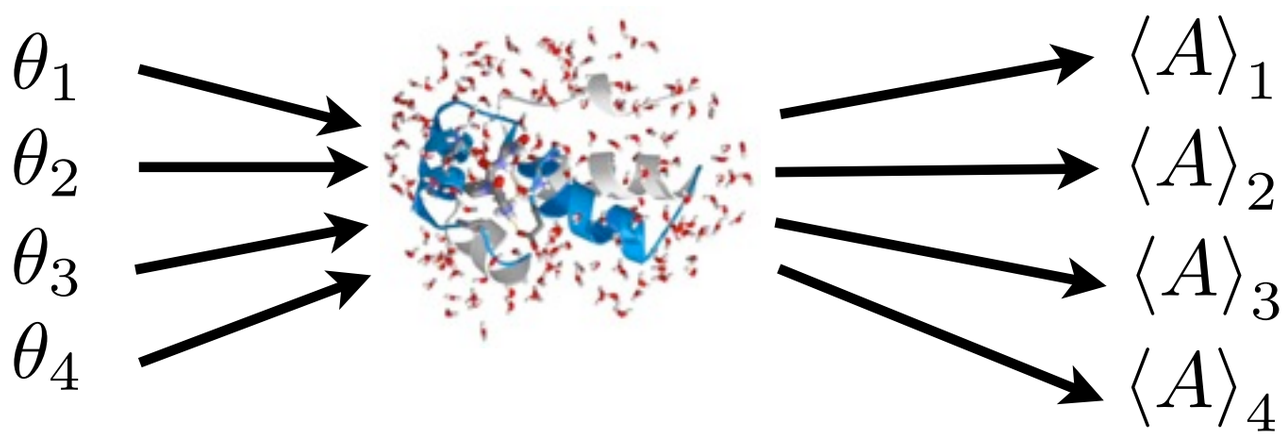
effective energy function

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



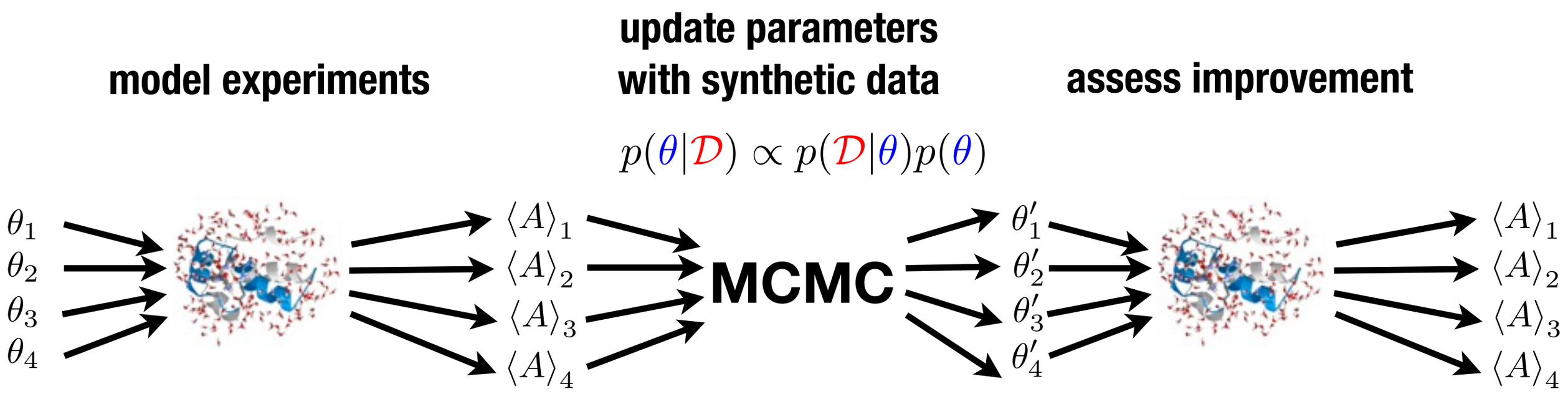
Estimate computed properties for other parameter sets from collection by reweighting (e.g. MBAR)



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

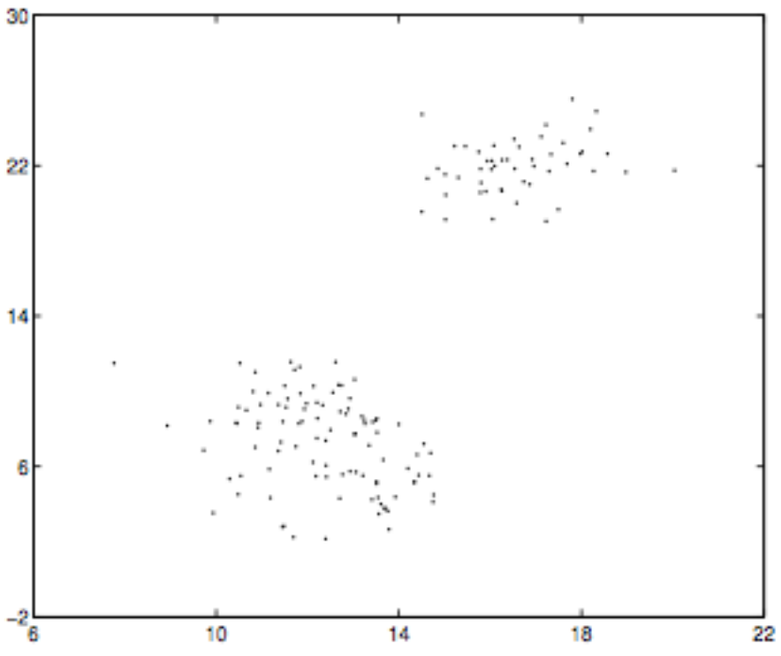
---

Classic Gaussian mixture model case:

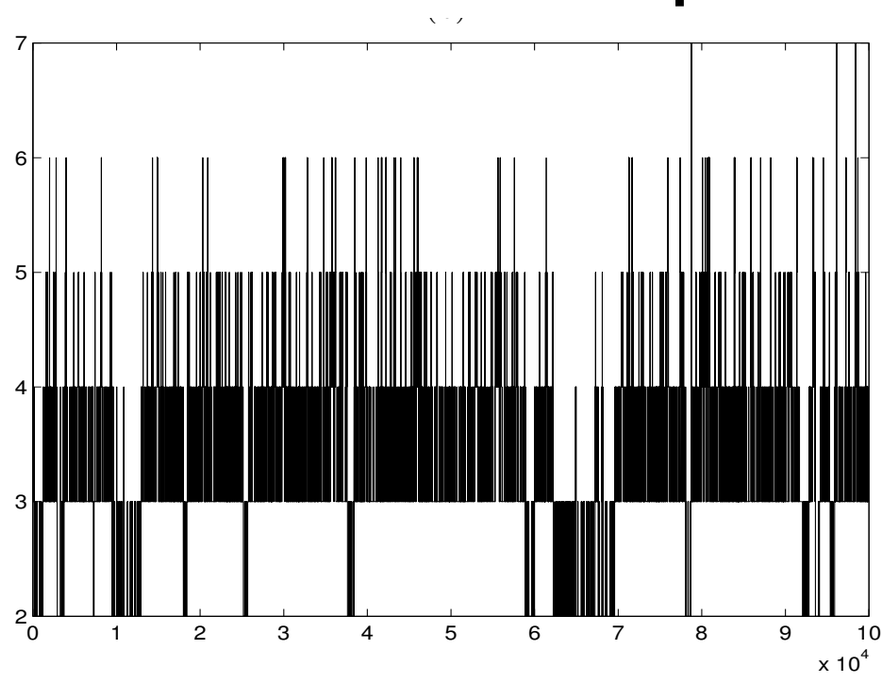
- How many components do the data support?

- What are their parameters?

data



number of Gaussian components



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	Lorentz-Berthelot
exponential-6	Waldman-Hagler
Halgren buffer 14-7	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?

experimental data
quantum chemistry
uncertainties

ensemble of
parameter sets

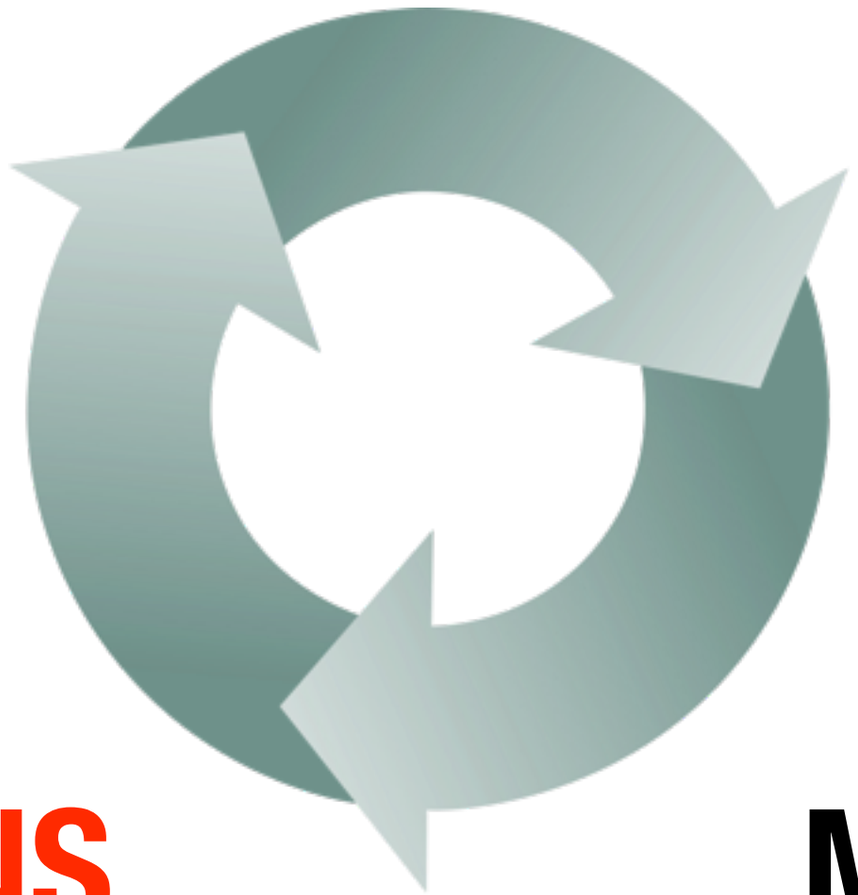
pull the trigger and go!

The universal cycle of progress
Bayesian inference can drive progress

DATA

update knowledge conditioned on data

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



update model conditioned on data

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

DECISIONS

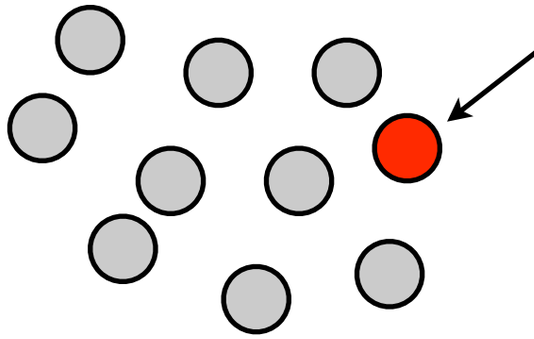
MODELS

make informed decisions based on model predictions

$$\text{argmax}_a \text{E}_a[f(\text{outcome})]$$

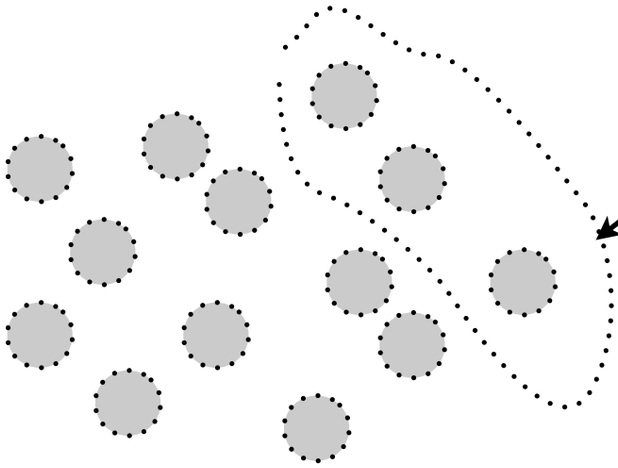
“Pick a winner”

Suppose we make predictions for 20 compounds the chemists can make. They don’t want to make all 20.



best one so far

some compounds we’ve made



how many the chemists want to make

a bunch of compounds we could make

What should they make?

Code at https://github.com/choderalab/cup-xiv

Let's pick a simple illustrative model

Simple model for affinity/potency/whatever:

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

feature/fingerprint vector $\mathbf{x} = [x_1 \; x_2 \; x_3]^{\mathrm{T}}$

model parameters $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

true model $\mathbf{A}^*$

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

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

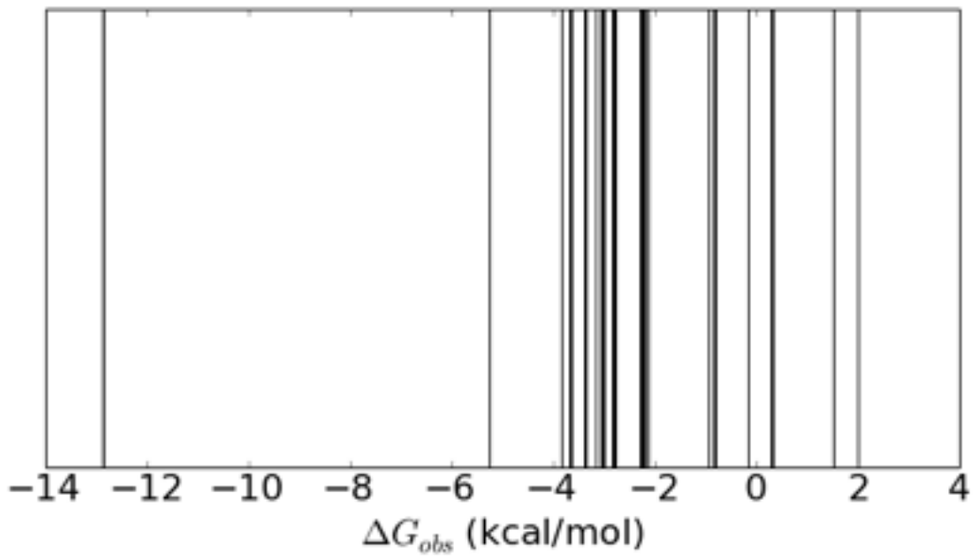
σ **measurement error standard deviation**

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

Code at https://github.com/choderalab/cup-xiv

Suppose we have already made some compounds

Measured affinities for 20 compounds

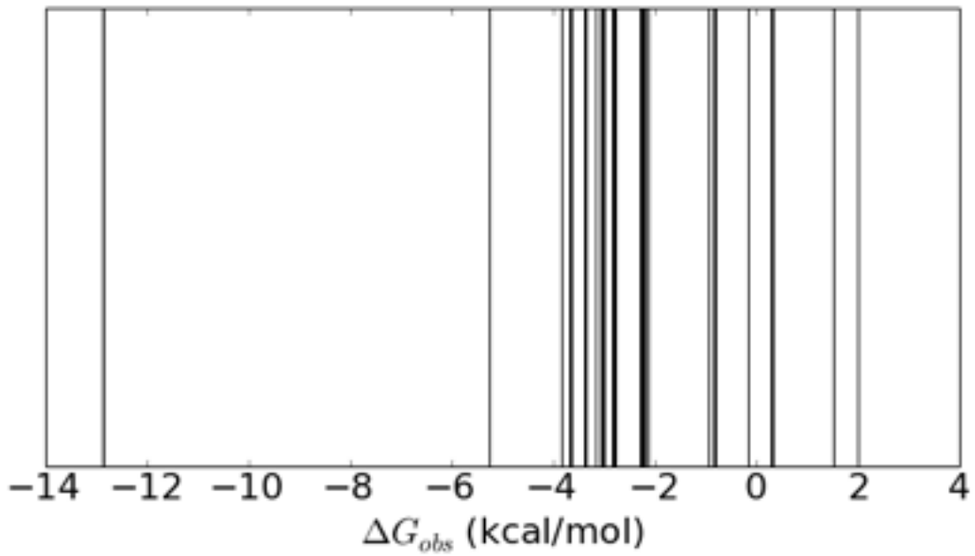


Which ones should we make?

Code at https://github.com/choderalab/cup-xiv

Suppose we have already made some compounds

Measured affinities for 20 compounds



Some pymc magic to sample models:

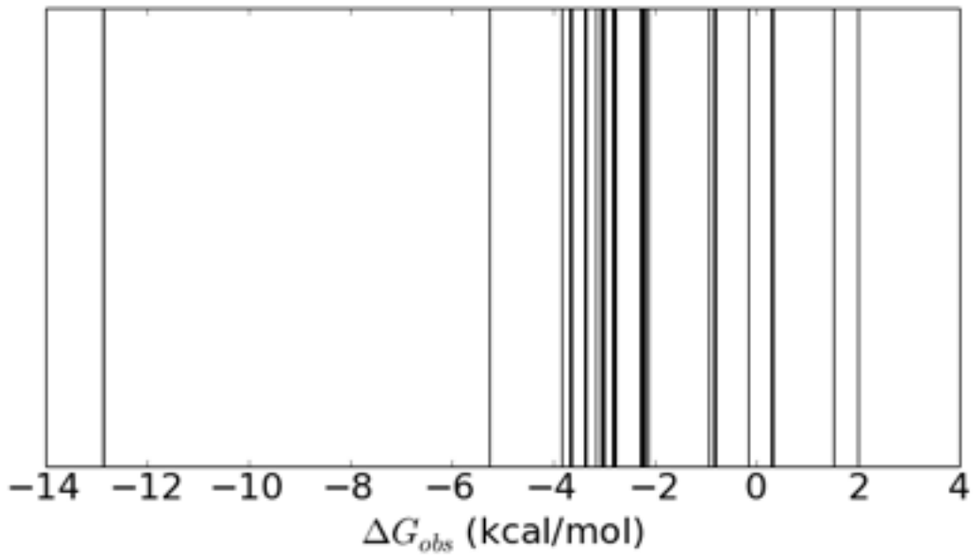
```
import pymc

# Create a pymc model
def make_model(x, DGobs_training):
    ntrain = len(x)

    A = pymc.Normal('A', mu=0.0, tau=1.0, size=[M,M])

    log_sigma = pymc.Uniform('log_sigma', lower=-3, upper=0, value=0.0)
    @pymc.deterministic
    def precision(log_sigma=log_sigma):
        return 1.0 / numpy.exp(log_sigma)

    @pymc.deterministic
    def DGmodel(A=A):
        DG = numpy.zeros([ntrain])
        for i in range(ntrain):
            DG[i] = model(x[i], A)
        return DG

    DGobs = pymc.Normal('DGobs', mu=DGmodel, tau=precision, size=[ntrain], observed=True, value=DGobs_training) # observed data

    # Construct dictionary of model variables.
    pymc_model = { 'A' : A, 'log_sigma' : log_sigma, 'precision' : precision, 'DGmodel' : DGmodel, 'DGobs' : DGobs }
    return pymc_model

pymc_model = pymc.Model(make_model(x_training, DGobs_training))

# Sample with MCMC
mcmc = pymc.MCMC(pymc_model, db='ram', name='Sampler', verbose=True)
mcmc.assign_step_methods()
mcmc.sample(iter=10000, burn=5000, thin=10, progress_bar=False)
```

Which ones should we make?

Code at https://github.com/choderalab/cup-xiv

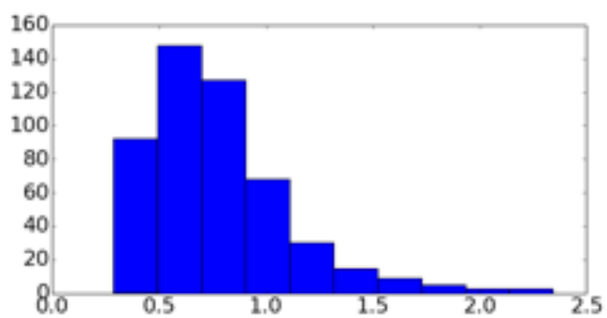
Suppose we have already made some compounds

Measured affinities for 20 compounds



Some pymc magic to sample models:

```
import pymc

# Create a pymc model
def make_model(x, DGobs_training):
    ntrain = len(x)

    A = pymc.Normal('A', mu=0.0, tau=1.0, size=[M,M])

    log_sigma = pymc.Uniform('log_sigma', lower=-3, upper=0, value=0.0)
    @pymc.deterministic
    def precision(log_sigma=log_sigma):
        return 1.0 / numpy.exp(log_sigma)

    @pymc.deterministic
    def DGmodel(A=A):
        DG = numpy.zeros([ntrain])
        for i in range(ntrain):
            DG[i] = model(x[i], A)
        return DG

    DGobs = pymc.Normal('DGobs', mu=DGmodel, tau=precision, size=[ntrain], observed=True, value=DGobs_training) # observed data

    # Construct dictionary of model variables.
    pymc_model = { 'A' : A, 'log_sigma' : log_sigma, 'precision' : precision, 'DGmodel' : DGmodel, 'DGobs' : DGobs }
    return pymc_model

pymc_model = pymc.Model(make_model(x_training, DGobs_training))

# Sample with MCMC
mcmc = pymc.MCMC(pymc_model, db='ram', name='Sampler', verbose=True)
mcmc.assign_step_methods()
mcmc.sample(iter=10000, burn=5000, thin=10, progress_bar=False)
```



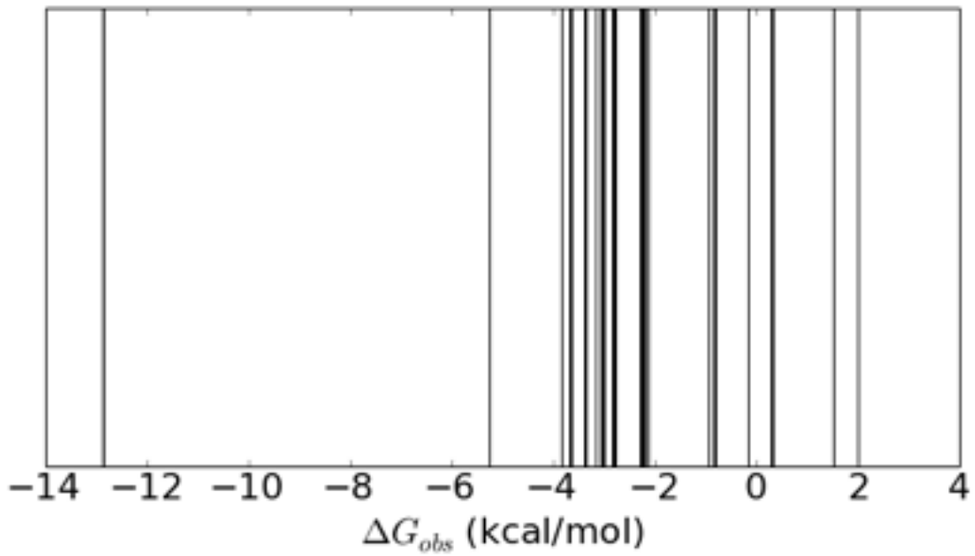
predicted experimental error (actual: 1)

Which ones should we make?

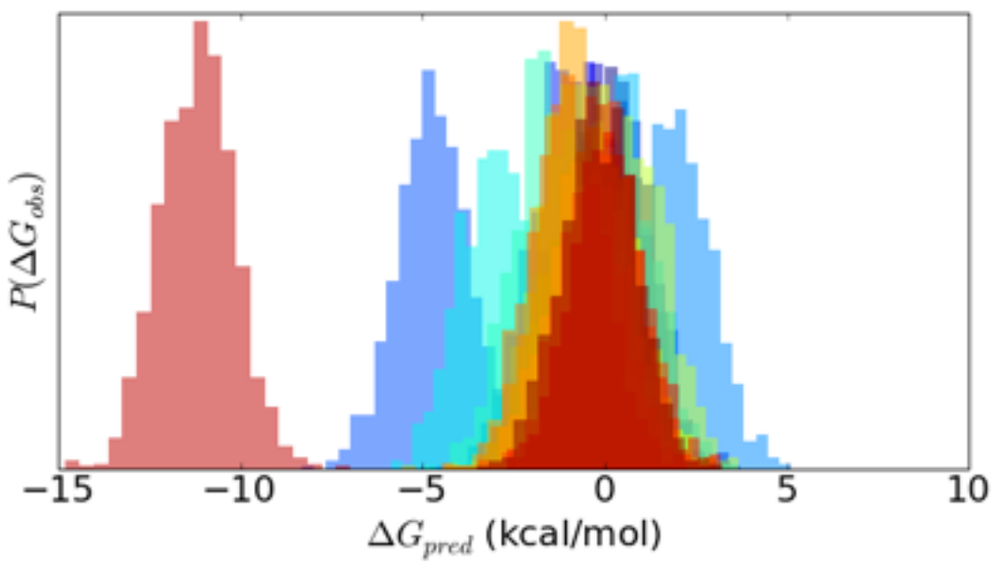
Code at https://github.com/choderalab/cup-xiv

Suppose we have already made some compounds

Measured affinities for 20 compounds

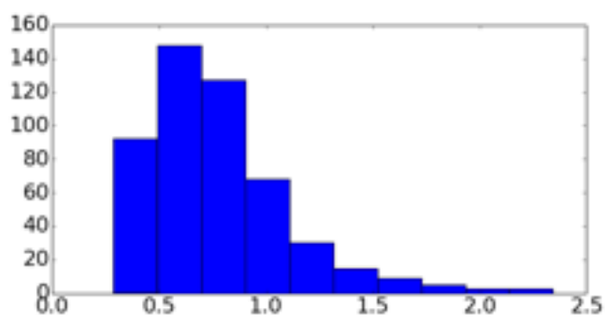


Predictions for next 20 compounds



Some pymc magic to sample models:

```
import pymc

# Create a pymc model
def make_model(x, DGobs_training):
    ntrain = len(x)

    A = pymc.Normal('A', mu=0.0, tau=1.0, size=(M,M))

    log_sigma = pymc.Uniform('log_sigma', lower=-3, upper=0, value=0.0)
    @pymc.deterministic
    def precision(log_sigma=log_sigma):
        return 1.0 / numpy.exp(log_sigma)

    @pymc.deterministic
    def DGmodel(A=A):
        DG = numpy.zeros([ntrain])
        for i in range(ntrain):
            DG[i] = model(x[i], A)
        return DG

    DGobs = pymc.Normal('DGobs', mu=DGmodel, tau=precision, size=[ntrain], observed=True, value=DGobs_training) # observed data

    # Construct dictionary of model variables.
    pymc_model = { 'A' : A, 'log_sigma' : log_sigma, 'precision' : precision, 'DGmodel' : DGmodel, 'DGobs' : DGobs }
    return pymc_model

pymc_model = pymc.Model(make_model(x_training, DGobs_training))

# Sample with MCMC
mcmc = pymc.MCMC(pymc_model, db='ram', name='Sampler', verbose=True)
mcmc.assign_step_methods()
mcmc.sample(iter=10000, burn=5000, thin=10, progress_bar=False)
```



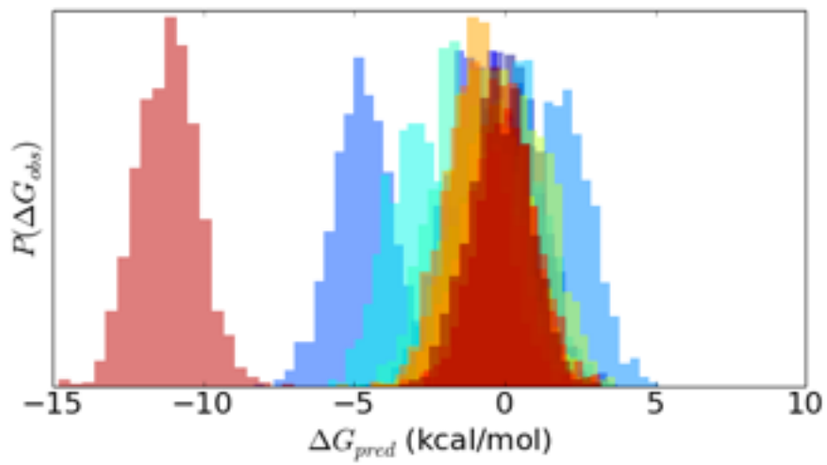
predicted experimental error (actual: 1)

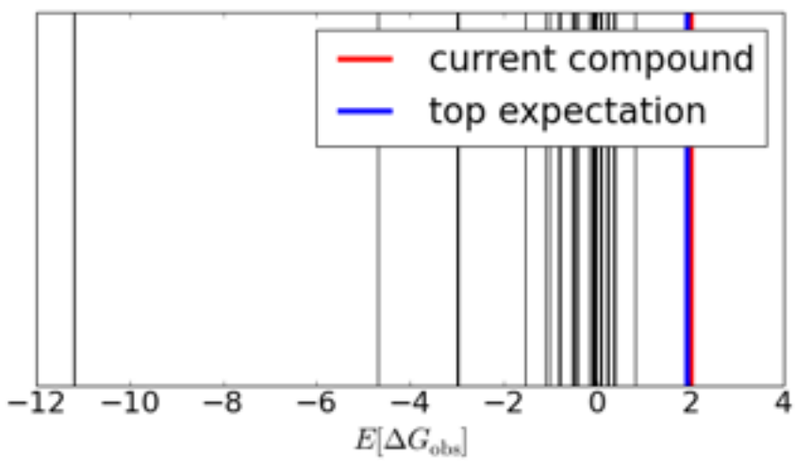
Which ones should we make?

Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

We could select a compound that maximizes our expected affinity gain (just like Bayesian Bandits)



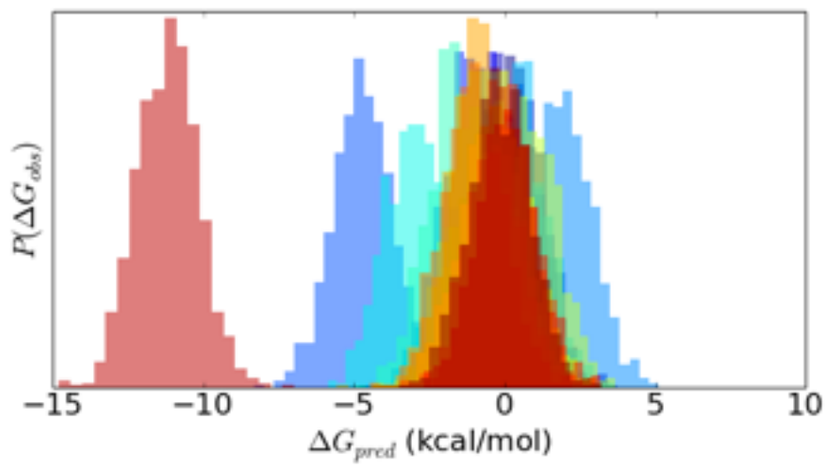


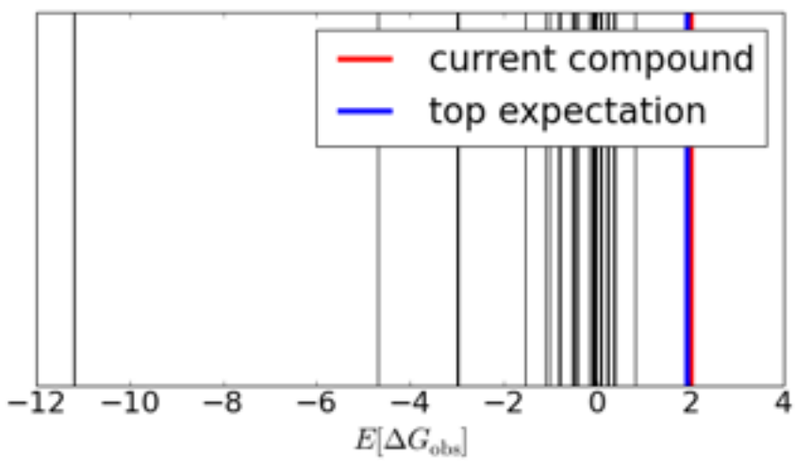
Uh oh. None of the 20 we were given have an expectation higher than the current best compound.

Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

We could select a compound that maximizes our expected affinity gain (just like Bayesian Bandits)





P(one of 20 compounds has affinity gain) = 0.538
P(top predicted expectation has affinity gain) = 0.361

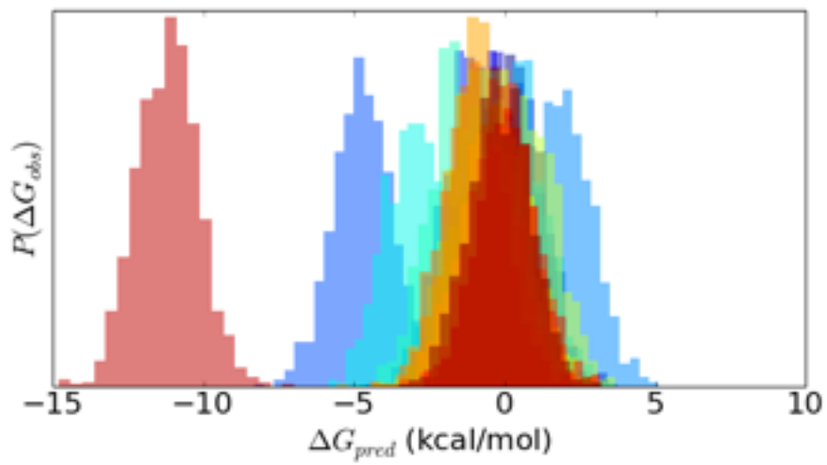
[Not unexpected for a model with 9 parameters fit to data for 10 compounds!]

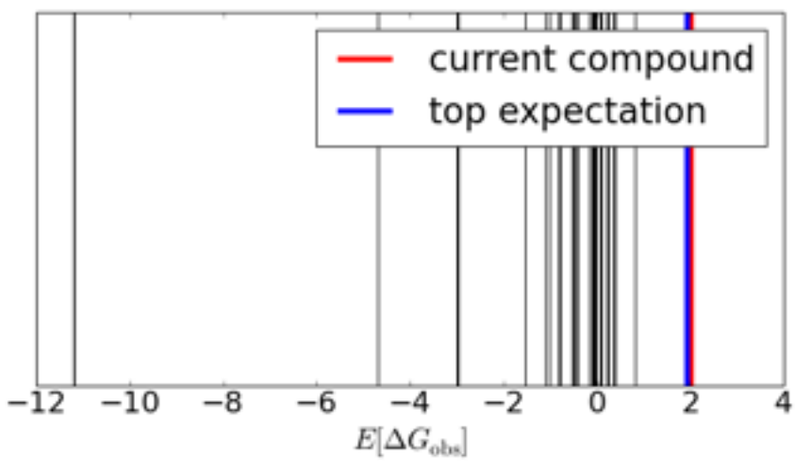
Uh oh. None of the 20 we were given have an expectation higher than the current best compound.

Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

What if we selected the five compounds that maximize expected affinity gain?

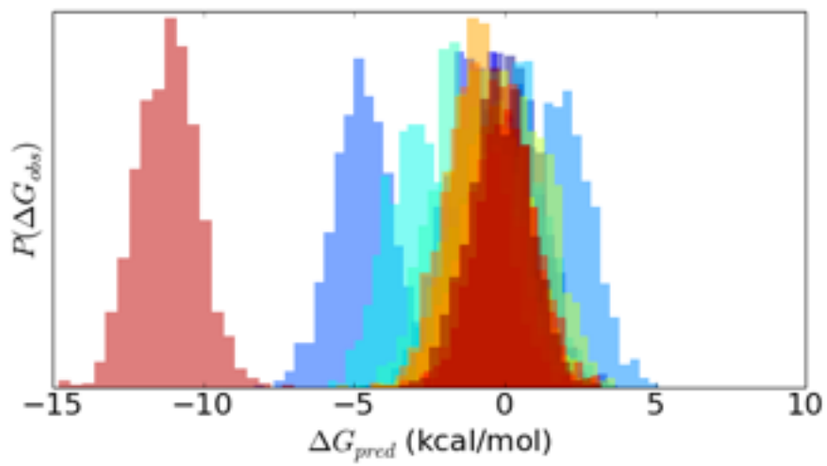


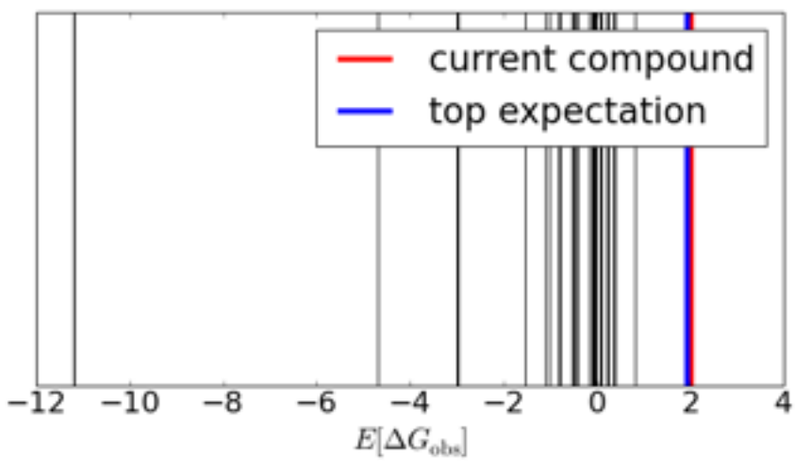


Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

What if we selected the five compounds that maximize expected affinity gain?





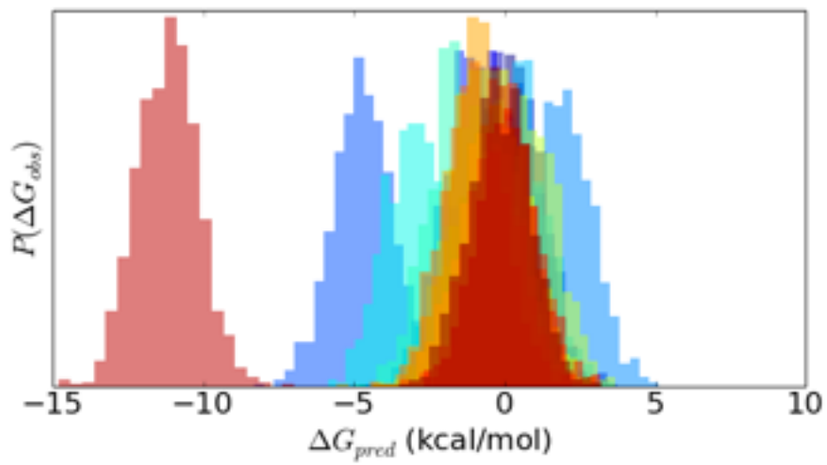
P(one of 20 compounds has affinity gain) = 0.538
P(top predicted expectation has affinity gain) = 0.361
P(affinity gain in top 5) = 0.502

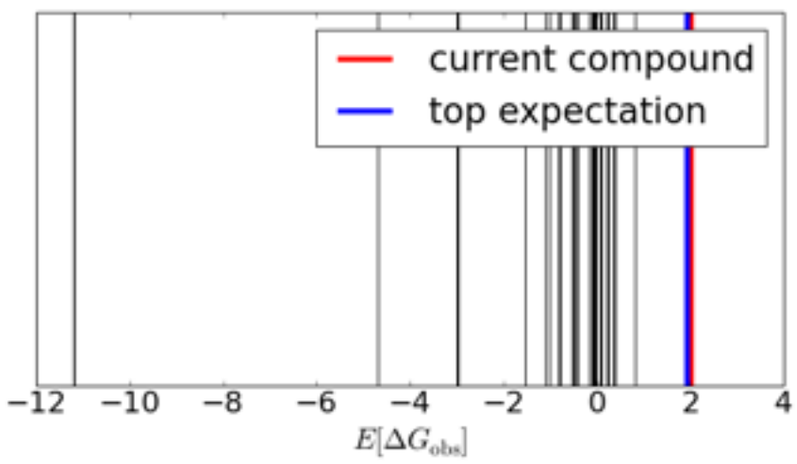
[Doing better!]

Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

What if we selected the five compounds that minimize probability of failure?

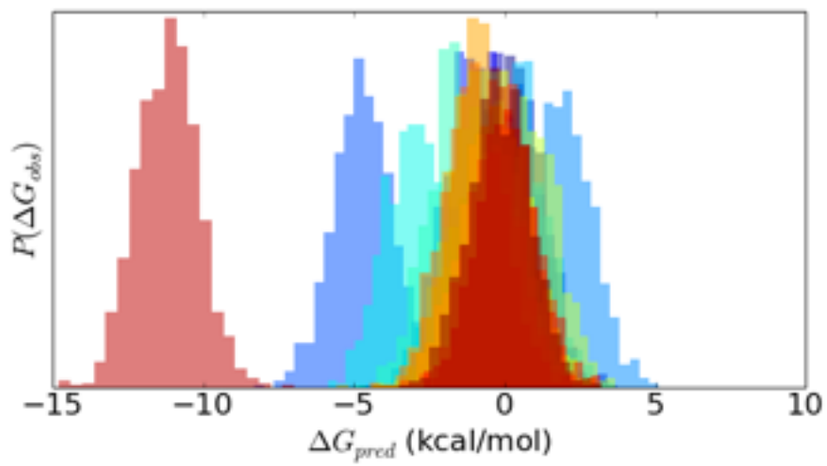


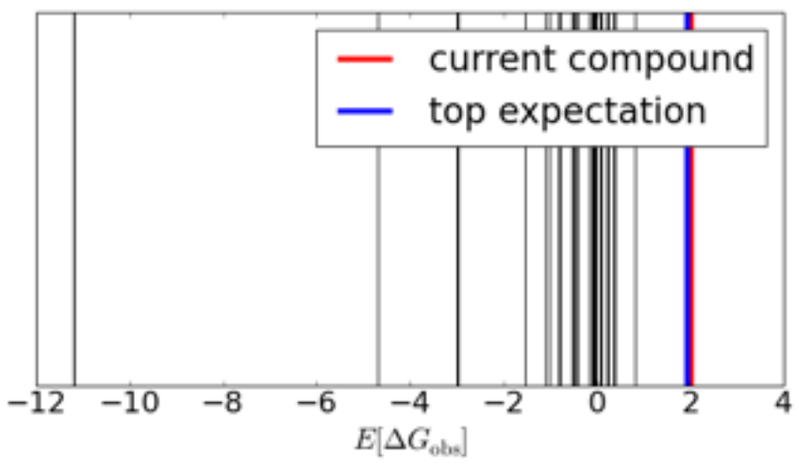


Code at https://github.com/choderalab/cup-xiv

Maximizing the expectation (as in Bayesian bandits)

What if we selected the five compounds that minimize probability of failure?





P(one of 20 compounds has affinity gain) = 0.538
P(top predicted expectation has affinity gain) = 0.361
P(affinity gain in top 5) = 0.502
P(affinity gain in "least worst" 5) = 0.539
[A bit better!]

Code at https://github.com/choderalab/cup-xiv

Where we need to be:
Decision-making 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 decision-making?

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)

Slides, references, and iPython tutorials available at http://choderalab.org/cup-xiv