Bayesian analysis of isothermal titration calorimetry data

John D. Chodera,[1,*] David D. L. Minh,[2] Sarah E. Boyce,[1] Joel Tellinghuisen,[3,†] Paul A. Novick,[4,‡] Kim Branson,[4,§] and Vijay S. Pande[4,¶]

[1]*California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720*
[2]*Biosciences Division, Argonne National Laboratory, 9700 S. Cass Av., Argonne, Illinois 60439, USA*[**]
[3]*Department of Chemistry, Vanderbilt University, Nashville, TN 37235*
[4]*Department of Chemistry, Stanford University, Stanford, CA 94305*

(Dated: June 30, 2015)

[JDC: Author order has not been finalized yet, but I wanted to make sure I listed everybody who made substantial contributions to this work.]

Isothermal titration calorimetry (ITC) is the only experimental technique able to reliably measure both the free energies of macromolecule-ligand association as well as its decomposition into enthalpic and entropic contributions in a single experiment. Due to the way in which the thermodynamic parameters are extracted from the data, errors in the free energy, enthalpy, and entropy of binding can differ in magnitude, and are generally highly correlated. Additionally, the use of multiple measurements to improve statistics, measure affinities of stronger ligands by competing off weaker ones, and separate proton uptake effects from binding by titrating in multiple buffers can complicate the propagation of these uncertainties to the physical quantities of interest. Here, we present a simple Bayesian framework for computing the full posterior distribution of all thermodynamic parameters and other quantities of interest from one or more ITC experiments, allowing their uncertainties and correlations to be quantitatively assessed. Use of this Bayesian approach leads to uncertainties that can be orders of magnitude larger than those typically reported when data are re-analyzed, but which more accurately represent the true variability in experiments from laboratory to laboratory. The framework is general and flexible, and further allows the modeling of new experiments in a way that aids the experimenter in selecting experimental parameters that will maximize the expected information gain. A Python implementation suitable for use with most popular calorimeter data formats is freely available online at `http://www.simtk.org/home/bayesian-itc`.

I. INTRODUCTION

Isothermal titration calorimetry (ITC) [**?**] has proven to be a powerful technique for measuring the free energy of association of two soluble species, most notably finding application in providing a quantitative measure of the binding affinity of small molecule ligands to biological macromolecules such as proteins and RNA [**? ?**]. [JDC: Add some references to reviews and applications to biomoleculecular interactions, such as those by Ernesto Friere.] From a single experiment, estimates of the free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) of binding can be simultaneously extracted, providing both a direct assessment of binding affinity as well as insight into the nature of the thermodynamics of binding. Beyond this, ITC also allows the study of more complex reactions, such as competitive binding reactions [**? ?**], binding events in the presence of changes in protonation [**? ?**] or tautomeric [**?**] states, and in certain cases, even kinetics of binding [**?**]. Recently, several groups have argued that ITC could play a more central role in lead optimization efforts in drug discovery [**? ?**].

In a typical ITC experiment, small quantities of a *titrant* (such as a ligand dissolved in buffer) contained in a syringe are injected into a sample cell containing the *titrate* (often a macromolecule in identical buffer), and the quantity of heat liberated or absorbed as a result of each injection is measured. Given a model of the heat liberated from each injection due to the thermodynamics of binding, the thermodynamic parameters of interest are then extracted from a fit of the evolved heat per injection to the binding model [**?**]. Only in the case that the reaction is *isenthalpic* (in which no heat is produced or consumed) can no useful measurement can be made with ITC.

Because various effects contribute to variation in the experimental operation—e.g. error in titrant or titrate concentrations, unintended variation in injection volume, noise in the measured heat signal—the reported thermodynamic quantities of interest will be determined only up to some degree of uncertainty or error. While some properties of this noise can be assessed by careful calibration runs, how the measurement error propagates into the thermodynamic quantities will depend on the actual binding characteristics of the system under study and the experimental protocol used to make the measurements. Further complicating this is the problem that some quantities, such as the degree to which the quantity of titrant actually injected in the first injection is diminished (the so-called "first injection

* Corresponding author; jchodera@stanford.edu
† joel.tellinghuisen@vanderbilt.edu
‡ pnovick@stanford.edu
§ kim.branson@stanford.edu
¶ pande@stanford.edu
** daveminh@anl.gov

anomaly" [**?** **?**]), are simply unknown. While good protocols attempt to *minimize* the sources of these errors [**?** **?** **?** **?**], they cannot *eliminate* them. With the Bayesian approach we describe below, however, we can at least attempt to *account* for how these various sources of error affect the determined parameters.

[JDC: Do we need to include a paragraph describing traditional nonlinear fitting procedures and their limitations?]

Here, we present a simple Bayesian formalism for inferring the full posterior distribution of the thermodynamic quantities of interest (free energy, entropy, and enthalpy of binding) as well as any unknown instrument parameters from the collected calorimetry data. The Bayesian approach provides numerous advantages over more traditional approaches based on nonlinear least-squares error fitting [**?** **?** **?**]. Unlike least-squares methods, the Bayesian approach provides the full posterior distribution of the inferred parameters, from which joint distributions of the parameters of interest, asymmetric confidence intervals, and measures of covariance of estimates that do not depend on linearity assumptions and asymptotically normal error estimates can be extracted.

Additionally, Bayesian methods allow for true binding model selection; instead of jointly inferring a parameter that selects among models—such as the binding stoichiometry n—the Bayesian method can assess the weight of evidence for each model and, given that model, what the unknown thermodynamic parameters are. Multiple experiments on identical or multiple titrants and macromolecules under different experimental conditions can be analyzed *simultaneously* to provide the best estimate of complex quantities of interest, like differences in binding affinities or entropies between lead candidates. Design of additional experiments can be aided by prior experimental data, knowledge of binding affinities, or even probable ranges, and the experimental protocol expected to yield the largest information gain selected. Finally, instead of using *ad hoc* subtraction techniques to deal with reference spectra or use of instrument noise reference values, calibration data or runs with blanks can be included directly in the inference procedure.

This paper is organized as follows: Section **??** describes the isothermal titration calorimetry experiment in detail. In Section **??**, we describe the Bayesian formulation for the posterior distribution of unknown thermodynamic parameters given the experimental data. An efficient scheme for sampling from the posterior and computing useful quantities from this sample is described in Section **??**. We then illustrate and validate the Bayesian analysis scheme by applying it to fully independent replications of an experiment in which $MgCl_2$ is titrated into a cell containing the chelator EDTA to study the thermodynamics of the Mg^{2+}:EDTA reaction in Section **??**. Use of the Bayesian method to inform experimental design is described in Section **??**, and various common binding models (which can be "plugged-into" the Bayesian formalism) are provided in Appendix **??**. A free Python implementation of the Bayesian methodology described here, along with all datasets used in Section **??**, is available from `http://www.simtk.org/home/bayesian-itc`.

II. ISOTHERMAL TITRATION CALORIMETRY

[JDC: Should we describe ITC generally, or focus on the MicroCal instrument(s) in particular?] In performing an isothermal titration calorimetry (ITC) experiment with a modern titration microcalorimeter, two identical cells are isolated from the environment by a thermal jacket, as depicted in Figure **??**. [JDC: How is the VP-ITC thermal jacket thermostatted to the sample and reference cell temperatures? Is it to minimize heat flow between cells and jacket?] One cell, the *sample cell*, is filled with a solution containing one or more molecular species dissolved in buffer (the *titrate*). This usually consists of a macromolecule to which the binding of some ligand is to be assessed, but may also include a weak ligand to be displaced by a stronger ligand, as in the case of competitive binding experiments [**?**]. The other cell, the *reference cell*, is filled with a solution of identical heat capacity to the titrate. A syringe with automatic injection control, mounted above the sample cell, contains the *titrant*, which usually consists of a ligand in buffer. In order to minimize heats of mixing upon injection that will obscure the binding signal, the titrant and titrate buffers must be identical. Because of this, common protocols recommend dialysis of both against a common buffer during sample preparation [**?**].

Initially, both cells and the syringe are thermostatted to the desired experimental temperature. Throughout the experiment, the reference cell is slowly heated by application of constant known power to a resistive heating element. This applied *reference power* is small—generally in the range of 2–30 μcal/s for a $\sim$ 1.4 ml sample cell [**?**]—to ensure the total change in sample temperature over the duration of the experiment (generally tens of minutes to hours) is minimal (much less than one degree Celsius). To maintain identical temperatures in both sample and reference cells, the sample cell is heated by a separate resistive heating element slaved to a highly accurate sensor that measures the thermal differential between sample and reference cells, applying heat to the sample cell to minimize the thermal difference and making accurate measurements of the power required to do so.

During the course of the experiment, a series of N injections are performed in which a known volume of titrant is injected into the sample cell. The volume of the injection may vary with injection number, and the volume of injection n is denoted here by ΔV_n. Many protocols, for example, recommend a reduced volume on the first injection due to the belief (now invalidated [**?**])

that some titrant is inevitably lost from the syringe needle prior to the first injection [**?**]. The majority of instruments (such as the MicroCal VP-ITC) utilize perfusion-type configurations, in which a quantity of liquid equal to the injection volume flows out of the sample cell into an inactive tube, where it no longer contributes to the detected temperature differential [**? ?**]. The interval of time τ_n between the beginning of injection n and the beginning of the next injection (or the end of the recording period) is chosen to be large enough to allow adequate mixing and exceed all relevant binding (and, if present, reorganization) kinetics, as well as allow the feedback heating mechanism sufficient time to equalize the temperature between the sample and reference cells prior to the next injection. If these criteria are not met, it becomes difficult (if not impossible) to deconvolute the heat signals from sequential injections.

While the experiment is running, the temperature difference between the reference and sample cells is sampled at discrete subsecond intervals, converted into power units, and used to drive the heater of the sample cell. This *differential* applied power between the reference and sample cells (which should average to zero in the absence of titrant injections) is averaged over fixed time intervals τ_{filter} and stored to disk, along with readings of the cell temperatures and performance of the instrument.

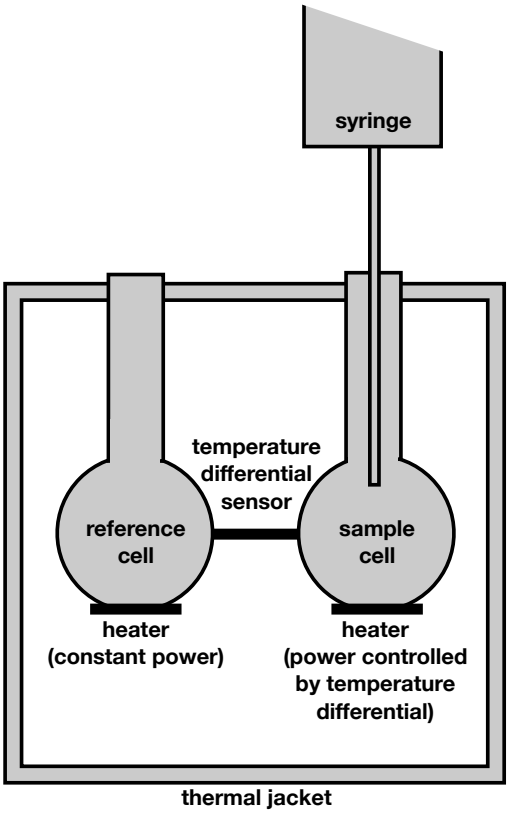
If the binding reaction is either *endothermic* (consumes heat) or *exothermic* (produces heat), the temperature of the sample cell will deviate from that of the reference cell, and the integrated difference in power applied to bring the sample cell to the same temperature as the reference cell will provide a measure of the quantity of heat absorbed during the binding reaction. The various thermodynamic quantities of interest are then extracted from a fit of the evolved heat per injection to the binding model. Should the reaction be isenthalpic (in which no heat is produced or consumed), no useful measurement can be made with ITC. The power applied to the reference cell is chosen to minimize the duration of the experiment while ensuring that the temperature rise over the course of the experiment is minimal.

[JDC: This section contains old material that will be moved and reworked.]

We concern ourselves with M molecular species denoted X_m. Most commonly, $M = 2$, where one species is a macromolecule and the other species is a ligand or binding partner whose affinity for the macromolecule is to be determined.

A sample containing one or more of the species X_m is prepared in buffer and loaded into a *sample cell* of volume V_0, where the concentration of each species X_m is initially $[X_m]_0$. Most often, the sample cell solution contains only a macromolecule, but a weak ligand of known affinity might also present in order to allow the affinity of very strong ligands to be determined. Another solution containing one or more species X_m (usually, the ligand of interest) is prepared in identical buffer at concentrations $[X_m]_s$, and loaded into the injector syringe.

FIG. 1. **Schematic diagram of an isothermal titration calorimeter.** [JDC: The top of this figure needs to be clipped off.]



A. Evolved heat due to binding

For a simple association reaction of two components P and L without cooperativity,

$$P + L \stackrel{\Delta G_a}{\rightarrow} PL$$

where ΔG denotes the free energy of binding, the *dissociation constant*, K_d, is given by

$$K_d \equiv \exp[-\beta \Delta G_a]\,(1\ \mathrm{M}) = \frac{[P][L]}{[PL]}\,(1\ \mathrm{M}). \qquad (1)$$

where $[X]$ denotes the concentration of species X. We note that K_d is expressed in units of *molarity* (mol/L).

With each injection, the volume of buffer in the sample cell increases, so that we can write the volume V_n

after injection n as

$$V_n = V_0 + n\Delta V. \tag{2}$$

The total quantity (number of moles) of protein P and ligand L_n in the cell after injection n is easily seen to be

$$\begin{aligned} P &= V_0\,[P_T]_0 \\ L_n &= n\,\Delta V\,[L_T]_s \end{aligned} \tag{3}$$

Conservation of mass gives us the constraints

$$\begin{aligned} P &= V_n\left([P]_n + [PL]_n\right) \\ L_n &= V_n\left([L]_n + [PL]_n\right) \end{aligned} \tag{4}$$

Combining Eqs.

With each injection, three effects will contribute to the true quantity of heat q_n^* liberated by injection n: (1) the association of P with L, the (2) the dilution of ligand and buffer into the protein solution (as most solutions are nonideal), and (3) the mechanical heat produced due by the injection and stirring. We subsume the latter two components into a single term ΔH_0, and write

$$q_n^* = V\Delta H\left([PL]_n - [PL]_{n-1}\right) + \Delta H_0 \tag{5}$$

where ΔH is the enthalpy change for the reaction $P + L \rightarrow PL$, and ΔH_0 is the heat of dilution and stirring.

Given K_d, we can determine the concentration of bound ligand after n injections, $[PL]_n$, from V_n, L_n, and P by (see Appendix **??**)

$$\begin{aligned} V_n\,[PL]_n = \frac{1}{2}\{&(V_nK_d + L_n + P) \\ &- \left[(V_nK_d + L_n + P)^2 - 4L_nP\right]\}^{1/2} \end{aligned} \tag{6}$$

The *measured* heat q_n is assumed to differ from the *actual* heat q_n^* by a normally-distributed error:

$$p(q_n|q_n^*,\sigma) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(q_n - q_n^*)}{2\sigma^2}\right] \tag{7}$$

where we have introduced the nuisance parameter σ to quantify the magnitude of this error. Because each heat measurement is assumed to be independent, the likelihood of observing the data $\mathbf{q}$ given the actual evolved heats $\mathbf{q}^*$ is simply

$$p(\mathbf{q}|\mathbf{q}^*,\sigma) = \prod_{n=1}^{N}\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(q_n - q_n^*)}{2\sigma^2}\right] \tag{8}$$

In the absence of any calibration information, we can assign the noise parameter σ a Jeffreys prior

$$p(\sigma) \propto \sigma^{-1} \tag{9}$$

Putting this all together, we construct the complete posterior probability density function

$$\begin{aligned} &p(\Delta G, \Delta H, \Delta H_{\mathrm{dil}}, \sigma|\{q_n\}_{n=1}^N) \\ &\propto \left\{\prod_{n=1}^{N}\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(q_n - q_n^*)^2}{2\sigma^2}\right]\right\}\sigma^{-1} \\ &\propto \sigma^{-(N+1)}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(q_n - q_n^*)^2\right] \end{aligned} \tag{10}$$

where the true heats per injection q_n^* are a function of ΔG, ΔH, and ΔH_0, as given by Eq. **??**.

III. BAYESIAN FORMULATION

In Bayesian inference, we wish to infer the *posterior* distribution of some unknown parameters Θ given observed data $\mathcal{D}$ and any prior information $\mathcal{I}$. The posterior quantifies our complete state of knowledge about how well various choices for the *true values* of the unknown parameters Θ are supported by the available data and prior information. By Bayes' theorem, this posterior probability can be written (up to an irrelevant normalization constant independent of Θ) as

$$p(\Theta|\mathcal{D},\mathcal{I}) \propto p(\mathcal{D}|\Theta)\,p(\Theta|\mathcal{I}) \tag{11}$$

The conditional probability $p(\mathcal{D}|\Theta)$, termed the *likelihood*, describes the probability of observing data $\mathcal{D}$ given underlying model parameters $\boldsymbol{\theta}$. Due to various sources of random error, each realization of the same protocol of the experiment will generate a different set of observed data $\mathcal{D}$, sampled from this distribution $p(\mathcal{D}|\boldsymbol{\theta})$. In practice, while we may not know $p(\mathcal{D}|\boldsymbol{\theta})$ exactly, we can often come up with an extremely good model for this distribution if we know in detail what the experiment is measuring.

The quantity $p(\boldsymbol{\theta}|\mathcal{I})$, termed the *prior*, expresses our state of knowledge of $\boldsymbol{\theta}$ before accounting for the data $\mathcal{D}$. This distribution characterizes whatever prior information $\mathcal{I}$ we may have, be it physical considerations or prior experimental data.

In an ITC experiment, the data $\mathcal{D}$ consists of a series of measurements of the differential power applied to the sample cell, given a protocol describing the concentrations of various species in the sample cell and syringe, the injection volumes and times, and the time between injections. The unknown parameters $\boldsymbol{\theta}$ consists of the thermodynamic parameters of interest such as the free energy of binding ΔG, and its decomposition into enthalpic ΔH, and entropic $-T\Delta S$ components. Also included in $\boldsymbol{\theta}$ are any unknown quantities that play a role in determining the distribution of observed data $\mathcal{D}$, such as the heat of dilution, heat due to stirring, magnitude of measurement errors, and so on. The prior information $p(\boldsymbol{\theta}|\mathcal{I})$ includes any prior information we might have from calibration runs or prior experimental data.

The posterior represents the joint distribution of *all* parameters—not just the ones we might care about at any moment. As a result, we can *marginalize* the posterior by integrating out parameters we don't care about, expressing the posterior distribution of only those of interest. For example, we may only be interested in the distribution of the free energy of binding $p(\Delta G|\mathcal{D},\mathcal{I})$, or perhaps the joint distribution of the enthalpic and entropic contributions $p(\Delta H, -T\Delta S|\mathcal{D},\mathcal{I})$. Useful statistical quantities like means, (co)variances, and confidence intervals then be extracted from these distributions. The

remaining parameters, which have been integrated out, are referred to as *nuisance variables*.

Below, we describe the various components necessary to model the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{I})$ in a way that accounts for many aspects of experimental error.

A. Prior for thermodynamic parameters

Given our prior information $\mathcal{I}$, we must first assign a prior $p(\boldsymbol{\theta}|\mathcal{I})$ to the unknown thermodynamic parameters $\boldsymbol{\theta}$.

In the absence of any previous knowledge about the parameters, we choose to assign flat priors to binding affinities ΔG and enthalpies of association ΔH:

$$p(\Delta G, \Delta H|\mathcal{I}) \propto 1 \tag{12}$$

A uniform prior is appropriate because the quantities ΔG and ΔH can be of either sign, and of any value. (The entropic contribution to the binding affinity, $-T\Delta S$, is implicitly determined in terms of ΔG and ΔH by the definition of the Gibbs free energy $\Delta G = \Delta H - T\Delta S$.)

If we do have prior information about how, say, the binding free energy must be within a given range $[\Delta G_{\mathrm{low}}, \Delta G_{\mathrm{high}}]$ (e.g. as determined by another experimental technique), we can write

$$p(\Delta G, \Delta H|\mathcal{I}) \propto \begin{cases} 1 & \text{if } \Delta G \in [\Delta G_{\mathrm{low}}, \Delta G_{\mathrm{high}}] \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Alternatively, if another technique has determined the binding affinity to be $\Delta G^* \pm \delta\Delta G$, where the error bars denote the standard error, we can use a normal prior

$$p(\Delta G, \Delta H|\mathcal{I}) \propto \delta\Delta G^{-1} \exp\left[-\frac{1}{2}\frac{(\Delta G - \Delta G^*)^2}{(\delta\Delta G)^2}\right] \tag{14}$$

Finally, if we already have collected data $\mathcal{D}^*$ from a separate ITC experiment involving one or more of the thermodynamic parameters under study, we can simply use the posterior $p(\boldsymbol{\theta}|\mathcal{D}^*,\mathcal{I})$ as a suitable prior for $\boldsymbol{\theta}$. This can be particularly useful when combining data from multiple experiments on a set of ligands, for example, to determine relative binding affinities or in the analysis of competition experiments where a weaker ligand is competed of by multiple stronger ligands in separate experiments.

B. Measurement error in evolved heat

The calorimeter makes numerous measurements each second of the temperature difference between the reference and sample cell. This temperature difference is translated into power units and used to control the resistive element heating the sample cell. The resulting quantity is termed the *differential power*, by convention negative when an exothermic reaction in the sample cell results in less power applied to the sample cell than the reference cell, and positive when an exothermic reaction requires more heat to be applied to the sample cell than the reference cell to minimize the temperature differential.

Many sequential differential power measurements are averaged over a larger time window—termed the *filter period* τ_{filter}—to produce an estimate of the true average differential power applied to the sample cell over this time. Because the true temperature differentials are very small, both the measured temperature differential (and hence reported differential power) and the true applied power will contain random errors. (We presume the instrument has been correctly calibrated so that *bias* is minimal compared to the magnitude of these random errors.) Because of the filtering stage, the central limit theorem demands the filtered differential power measurements will be normally distributed about the true average differential power regardless of the actual distribution of random errors in the individual measurements. Additionally, the filtering windows do not overlap, and so the random errors in different filtered power measurements can further be assumed to be independent.

The kinetics of injection, mixing, and binding are not relevant in determining the binding thermodynamics, provided the time between injections is long enough for the sample cell to reach the same temperature as the reference cell before the beginning of the next injection (or the end of the measurement period). This is generally satisfied provided the timescale for binding/unbinding kinetics and any associated conformational changes is much shorter than the time between injections. The total quantity of heat absorbed or emitted from the sample cell as a result of each injection is therefore a sufficient statistic for determination of the thermodynamic parameters of binding. We denote the measurement of the heat liberated (or absorbed) as a result of injection n by

$$q_n = \sum_{t=1}^{T_n} \Delta t \, P_{nt} \tag{15}$$

where the sum runs over all of the T_n differential power measurements P_{nt} taken starting with the beginning of injection n and ending just prior to injection $n+1$ (or when the recording period terminates).

We characterize the measurement noise ϵ in the reported differential power (resulting from both errors in the measured temperature differential and errors in the true applied power averaged over many samples during τ_{filter}) by a normal distribution of width σ:

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{16}$$

The use of the normal distribution for noise ϵ_t is, again, justified by the central limit theorem. The noise in the measured heat q_n will therefore also be normal, but with

a variance that depends on the number of filtering periods T_n between injections n and $n+1$:

$$q_n \sim \mathcal{N}(q_n^*, T_n \sigma^2) \tag{17}$$

In the absence of any calibration information, we can assign the noise parameter σ a Jeffreys prior [**?**]

$$p(\sigma) \propto \sigma^{-1} \tag{18}$$

If we do have calibration data on the magnitude of σ, we can condition the prior on this data. For example, from a calibration run $\mathcal{D}_{\mathrm{cal}}$ consisting of measuring the reported differential power $P_{\mathrm{cal},n}$ over N filter periods where no ligand is injected, we can update the prior on σ using the same assumption of normality for the P_n as above to obtain:

$$p(\sigma^2|\mathcal{D}_{\mathrm{cal}}) \propto \sigma^{-(N+1)} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N} P_{\mathrm{cal},n}^2\right] \tag{19}$$

[JDC: We should allow more complex error models to be included and, ideally, automatically selected. Look at Joel's paper on this topic [**?**]. Ideally we can extend his models to include the injection duration.]

C. Errors in sample preparation

Because the preparation of the sample and titrant will inevitably errors in the concentrations of the solutions, we can account for these errors as well during our analysis to ensure that the reported binding affinities reflect these experimental uncertainties as well. In particular, errors in quantities like the reported initial concentrations of macromolecule or ligand will have a direct effect on binding affinities. [JDC: How much will a 1% error in initial concentrations affect the the reported binding affinities – is this directly proportional?] We can therefore include an error model to describe the uncertainties in the true initial concentrations given the reported initial concentrations.

Suppose, for example, we determine that the solution loaded into the sample cell had a concentration of macromolecule $[M]_0 \pm \delta[M]_0$, where the standard deviation δM was estimated by careful propagation of error during the sample preparation process (from estimates of known pipetting error magnitudes, known analytical balance accuracies, and reported compound purities). We can model the true initial concentration of macromolecule $[M]_0^*$ by a normal distribution:

$$[M]_0^* \sim \mathcal{N}([M]_0, (\delta[M]_0)^2) \tag{20}$$

D. Volume displacement by shaft and stirring paddle of syringe assembly

Insertion of the shaft and stirring paddle of the injector syringe assembly into the sample cell displaces a small but significant volume which must be accounted for in order to avoid subsequent errors in the measured binding affinity [**?**]. For the MicroCal VP-ITC instrument, this volume has been estimated at 0.044 ± 0.005 mL[1] though Calorimetry Sciences Corp. recommends performing a separate calibration procedure for their instruments [**?**].

To account for this reduction in sample cell volume, we provide two options:

1. If calibration data is available (of the form $\Delta V_{disp}^{\mathrm{cal}} \pm \delta\Delta V_{disp}^{\mathrm{cal}}$), a normal prior can be assigned to the displacement volume ΔV_{disp} of the assembly:

$$\Delta V_{disp} \sim N(\Delta V_{disp}^{\mathrm{cal}}, (\delta\Delta V^{\mathrm{cal}})^2) \tag{21}$$

2. If no calibration data is available, or the calibration data is distrusted, the displacement volume can be inferred completely by assigning a broad uniform prior:

$$\Delta V_{disp} \sim U(0, V_{cell}) \tag{22}$$

Note that, in traditional ITC data analysis, the site parameter n is partially able to absorb errors in the stated cell volume, allowing correct binding affinities and enthalpies to be obtained at the expense of deviations from the expected $n = 1$ for 1:1 complexation reactions [**?**].

E. The "first injection anomaly"

A commonly encountered problem in ITC experiments performed according to manufacturer-recommended protocols (such as the VP-ITC User Guide [**?**]) is the observation that the first injection will yield an integrated heat of magnitude smaller than expected, a phenomenon termed the "first injection anomaly" [**?**]. Widespread practice has been to attempt to minimize the effect of this phenomenon on the fit to the binding equations by performing a small initial injection that is discarded before analysis, attributing the "first injection anomaly" to a loss of titrant from the syringe needle prior to the first injection.

This phenomenon is now understood to be the result of the "purge-and-refill" syringe preparation procedure suggested by instrument manufacturers [**?**]. As a result of this procedure, the worm screw in the syringe that converts rotary stepper motion to vertical motion to expel the syringe contents must change direction from "fill" to "purge" during the first injection, resulting in

[1] The statement "With allowance for uncertainties ... this value is likely within 0.01 mL of the true value" [**?**] is taken to mean that the 95% confidence interval is ± 0.01 mL, suggesting a standard deviation of ± 0.005 mL as referenced in the text.

an expulsion of titrant that is smaller than expected. The anomaly can be eliminated by simply issuing the syringe a short "plunger down" command prior to loading the syringe into the apparatus [**?**]. This modified procedure is highly recommended over the common practice of discarding the initial injection, especially because such a procedure will change the observed binding behavior of subsequent titrant injected into the sample cell because some titrate is already complexed with titrant.

However, it is recognized that many datasets have already been collected using a protocol which results in a first-injection anomaly. In order to permit analysis of these datasets within our framework, we provide the option of inferring the quantity of first injection volume shortfall through the following procedure. We introduce two nuisance parameters: v_{out}, denoting volume of titrant lost external to the sample chamber or a shortfall of injection volume due to work gear reversal, and v_{in}, denoting volume of titrant lost within the sample chamber prior to the start of the first injection. We apply the reasonable constraint that the total volume of this loss must be less than the total volume of the first injection, ΔV_1. Other than this, we apply no additional knowledge of what the magnitude of this loss may be, assigning a uniform prior to v_{out} and v_{in} in this range. Their actual values will be inferred—and marginalized out, if they are not of interest—during the analysis from the magnitudes of heat liberated or consumed from all the injections.

$$p(v_{\mathrm{out}}, v_{\mathrm{in}}) \propto \begin{cases} 1 & \text{if } (v_{\mathrm{out}} > 0) \cap (v_{\mathrm{in}} > 0) \\ & \quad \cap (v_{\mathrm{out}} + v_{\mathrm{in}} \leq \Delta V_1) \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

F. Other effects

JDC: Can we include other things into our model to make analysis more robust?

- Baseline correction/drift
- Better models for heat of mixing/dilution
- Bubbles?
- Multiple experiments and including titrations with blanks

IV. SAMPLING FROM THE POSTERIOR

Because the posterior distribution (Eq. **??**) is not amenable to direct sampling, we employ a Markov chain Monte Carlo (MCMC) procedure to sample from it for the purposes of computing means, variances, and approximations to the joint and marginal distributions [**?**].

JDC: Things to discuss here include

- Gibbs sampling for parameters like σ
- David's normally-distributed MCMC moves
- Automated step size adjustment during burn-in
- Correlated moves among parameters for increased acceptance probability
- Gibbs sampling to allow convenient introduction of additional nuisance parameters?
- Assessing sampling burn-in and convergence

V. ANALYSIS

[JDC: This section is still under construction.]

In most practical applications, we will want to integrate out parameters that are not of direct interest, producing marginal distributions for the free energy of binding

$$p(\Delta G|\mathcal{D},\mathcal{I}) = \int d\Delta H \int d\Delta H_0 \cdots \times p(\Delta G, \Delta H, \Delta H_0, \ldots|\mathcal{D},\mathcal{I})$$

or the joint distribution of the enthalpic (ΔH) and entropic ($T\Delta S$) contributions to binding

$$p(\Delta H, T\Delta S|\mathcal{D}) = \int d\Delta G \int d\Delta H_0 \times \int d\sigma\, p(\Delta H - T\Delta S, \Delta H, \Delta H_0, \sigma|\mathcal{D}) \tag{24}$$

A. Confidence intervals

B. Joint distributions

VI. EXPERIMENTAL DESIGN

[JDC: This section is still under construction.]

The Bayesian framework presented here can also be used to aid in the design of new experiments using previously-collected experimental data or prior information. A natural way to judge the utility of a given choice of experimental design parameters $\boldsymbol{y}$ is to maximize the *expected information content* (EIC) of the experiment $\boldsymbol{y}$ given prior information $\mathcal{I}$:

$$\mathrm{E}\left[I(\boldsymbol{y})\right] \equiv \int d\boldsymbol{\theta}\, I(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathcal{I}) \tag{25}$$

This prior knowledge may simply be the bare prior assumptions outlined above $p(\boldsymbol{\theta})$, or it may be conditioned

on additional expectations about the potential range of binding affinity, prior experimental data $\mathcal{D}$, or any information about the behavior of the apparatus (such as control experiments). Though prior experimental data on the system of interest will likely be the most informative, *any* information (provided it is not wrong) can be useful in making the estimation of EIC more precise.

Instead of computing only the mean EIC, our strategy computed the posterior distribution of the EIC based on the given assumptions and data by a double sampling approach, allowing the operator to determine whether one experiment will clearly deliver more information than another experiment. (In fact, the probability that this is the case case be estimated.)

To estimate the posterior $p(\mathrm{EIC})$, the following strategy is used:

1. Given the prior information and any data or assumptions to condition on, a proposed true model $\boldsymbol{\theta}^*$ is sampled from $p(\boldsymbol{\theta}|\boldsymbol{D}_1)$.

2. A random outcome $\mathcal{D}_y$ is chosen for one or more experiments $\boldsymbol{y}$ by sampling from $p(\mathcal{D}_y|\boldsymbol{\theta}^*)$.

3. The information content of the new data $\mathcal{D}_y$ given prior data $\mathcal{D}$ and prior information $p(\boldsymbol{\theta})$ is computed and tallied, and the process repeated.

$$\begin{aligned} I(\boldsymbol{y}|\boldsymbol{\theta}) &= \int d\mathcal{D}_y \, p(\mathcal{D}_y|\mathcal{I}) \, I(\mathcal{D}_y|\mathcal{I}) \\ &= \int d\mathcal{D}_y \, p(\mathcal{D}_y|\mathcal{I}) \; (H[p(\boldsymbol{\theta}|\mathcal{I})] - H[p(\boldsymbol{\theta}|\mathcal{D}_y, \mathcal{I})]) \\ &= \end{aligned} \tag{26}$$

$$\begin{aligned} \mathrm{E}\left[I(\boldsymbol{y})\right] &\equiv \int d\boldsymbol{\theta} \, p(\boldsymbol{\theta}) \, I(\boldsymbol{y}|\boldsymbol{\theta}) \\ &= \int d\boldsymbol{\theta} \, p(\boldsymbol{\theta}|\mathcal{D}) \end{aligned} \tag{27}$$

VII. ILLUSTRATIVE APPLICATIONS

A. Simple 1:1 complexation of Mg^{2+} to EDTA

We illustrate the effectiveness of the Bayesian approach in describing the *true* uncertainty in the experimental measurements by studying a simple complexation reaction—the 1:1 binding of Mg^{2+} to the chelator EDTA. In order to assess the true variation among measurements, including errors in the concentrations of titrant and titrate, the experiment was repeated *from scratch* ten times. For each trial, the titrant, $MgCl_2$, titrate, EDTA, and buffer, 50 mM Tris-HCl at pH 8.0, were weighed and dissolved to prepare solutions at the planned concentrations; of course, the *actual* concentration of the solutions in the syringe and sample cell differed due to various measurement, preparation, and solution transfer errors.

For each replicate of the experiment, a 0.5 M solution of $MgCl_2$ was prepared to act as the titrant and 0.05 M solution of EDTA to act as as the titrate. Magnesium Chloride Hexahydrate ($MgCl_2$) was purchased from Fisher Scientific (Catalog No. BP214-500, Lot No. 006533) and Ethylenediaminetetraacetic acid, anhydrous (EDTA) was purchased from Sigma-Aldrich (Catalog No. E6758-500G, Batch No. 034K0034). Tris Base was purchased from Fisher Scientific (Catalog No. BP154-1, Lot No. 082483). Buffer was prepared by weighing Tris base, adding MilliQ water, and adjusting the final pH to 8.0. Solutions were prepared by weighing powder (0.1 g for $MgCl_2$ and 0.01 g for EDTA) and adding the appropriate amount of buffer, neglecting the volume occupied by buffer, to make a concentrated solution (15 mM for $MgCl_2$ and 1.0 mM for EDTA). The solutions were then further diluted with buffer to prepare the titrant and titrate.

The ITC experiment consisted of a total of 24 injections, with the first programmed to deliver 2 μL of titrant ($MgCl_2$) and the remaining 23 injections programmed to deliver 12 μL. Data was collected for 60 s prior to the first injection and 300 s for each injection. The injection rate for all injections was 0.5 μL/s. All experiments were conducted at 298.1 K, and the reference power was fixed at 5 μcal/s.

[JDC: More here.]

VIII. DISCUSSION

We have described a simple Bayesian framework for the analysis of isothermal titration calorimetry experiments that significantly extends what was possible with earlier analysis techniques. First, the approach more accurately captures the uncertainties in estimated thermodynamic parameters by allowing rigorous confidence intervals, rather than just standard deviations of the mean, to be estimated from the data. Second, by capturing the full posterior distribution of the thermodynamic parameters, complex (often nonlinear) covariations in the estimated parameters can be accurately assessed. This may, for example, have great utility in assessing whether entropy-enthalpy compensation is present in a system under study in a way that is statistically meaningful. Thirdly, the ability to combine data from multiple experiments conducted under different conditions allows binding affinities and other thermodynamic parameters to be determined much more precisely than before, and their errors to be assessed yet more accurately. Fourth, the ability to use existing information or data to aid in the design of new experiments using a rigorous information theoretic criterion gives the experimenter a powerful new tool that goes far beyond the 'rules of

thumb' that have been the mainstay of the field for many years.

[JDC: Add a paragraph summarizing observations on the applications appearing in the paper.]

With automated, high-throughput microcalorimeters, such as the MicroCal Auto-iTC$_{200}$ (capable of running 75 samples/day, and 384 samples unattended) and the Vivactis MiDiCal II (projected to be capable of running up to 1000 samples/day) on the horizon, it becomes increasingly important that robust, reliable, and automated methods be available for analyzing the large quantities of ITC data that are becoming available. There is obvious potential for software which automatically determines the optimal experimental design parameters (e.g. concentrations, competition assays) for follow-up experiments (within specified resource restrictions) that become too tedious or complex for a human laboratory worker. The Bayesian framework described here provides a natural choice for these high-throughput environments, where all experimental data can be simultaneously analyzed and used to decide which follow-up experiments are necessary to confirm or reject hypotheses with statistical certainty.

A Python implementation of this method is freely available online at http://www.simtk.org/home/bayesian-itc.

IX. ACKNOWLEDGMENTS

The authors are indebted to Sarah E. Boyce (UCSF), Alan P. Graves (GSK), Veena L. Thomas (UCSF), Brian K. Shoichet (UCSF), and Phillip L. Geissler (UCB) for insightful discussions. JDC acknowledges support from a QB3-Berkeley Distinguished Postdoctoral Fellowship. VSP acknowledges support through an NSF grant for Cyberinfrastructure (NSF CHE-0535616). JDC and SEB thank Brian K. Shoichet (UCSF) for generous use of his laboratory to conduct some of the experiments presented here. [JDC: Add appropriate acknowledgments for other authors.]

Appendix A: Binding models

Any reversible association model in which the equilibrium concentrations of the various species $[X_n]$ can be computed (potentially numerically) in terms of the total quantity of each species in the sample cell can be used in the inference procedure. The inference procedure can be even used to test multiple models simultaneously by along transitions *between* models: The fraction of samples generated from each model reflects the degree by which that model is preferred (see Bayesian model selection [**?**]).

In developing a model of binding, we fundamentally must arrive at a way to compute enthalpy potential $H(\boldsymbol{X})$ as a function of the total quantities of the various species X_n that are present in the sample cell. Usually, this is obtained by first constructing a physical binding model and identifying both the relevant association constants and other parameters and the enthalpy change parameters associated with these events. Using this information, the enthalpy potential can be expressed in terms of the equilibrium quantities of all species in the sample cell. For a given set of total quantities of each dissociable species, we can solve (potentially numerically) for the equilibrium concentrations of all species and compute the new enthalpy potential.

To illustrate this process, we have worked out the solution of the equilibrium concentrations and enthalpy potential for various binding models in common use below.

1. Simple two-component binding

In a two-component 1:1 complexation reaction, we have reversible association between a macromolecule M and ligand L

$$M + L \stackrel{K_a}{\rightleftharpoons} ML \tag{A1}$$

where the equilibrium bound and free concentrations are characterized by an *association constant* K_a

$$K_a = \frac{[ML]}{[M][L]} \tag{A2}$$

where the total number of moles of macromolecule M and ligand L are given by the constraints

$$\begin{aligned} M &= V([\mathrm{M}] + [\mathrm{ML}]) \\ L &= V([\mathrm{L}] + [\mathrm{ML}]) \end{aligned} \tag{A3}$$

Combining these relations, we obtain a quadratic equation for the complex concentration $[ML]$

$$ML/V^2 - (M/V + L/V + 1/K_a)\,[\mathrm{ML}] + [\mathrm{ML}]^2 = 0 \tag{A4}$$

where the only solution that satisfies $0 \leq [\mathrm{ML}] \leq \min\{[\mathrm{M}]\,, [\mathrm{L}]\}$ is

$$[\mathrm{ML}] = \frac{1}{2V}\left\{(M + L + V/K_a) - \left[(M + L + V/K_a)^2 - 4ML\right]^{1/2}\right\} \tag{A5}$$

The change in molar enthalpy ΔH of the association reaction is denoted by

$$M + L \overset{\Delta H}{\rightarrow} ML \tag{A6}$$

If we assign the $[ML] = 0$ state a reference heat potential of 0, we can write the heat potential Q for a solution of M moles of macromolecule and L moles of ligand in a volume V as

$$Q(M, L, V) = \Delta H \cdot V \cdot [ML] \tag{A7}$$

where the concentration $[ML]$ is a function of M, L, V, and K_a determined from Eq. **??**.

2. Multiple independent binding sites of equal affinity

A binding model traditionally used in the analysis of macromolecule-ligand association in ITC experiments [**?**] assumes there are N independent binding sites on the macromolecule M for ligand L, all with equal association constants K_a:

$$\begin{aligned} M + L &\overset{K_a}{\rightleftharpoons} ML \\ ML + L &\overset{K_a}{\rightleftharpoons} ML^2 \\ &\vdots \\ ML^{N-1} + L &\overset{K_a}{\rightleftharpoons} ML^N \end{aligned} \tag{A8}$$

Note that, even though traditional analysis allows n to be fractional, N *must*, by physical necessity, be a whole number. (Fractional deviations from whole numbers in more traditional analysis can be considered as the fitting procedure attempting to correct for errors in the sample cell volume [**?**].)

All of the equilibrium constants are assumed to be equal:

$$K_a = \frac{[ML^n]}{[ML^{n-1}][L]} \;\;,\;\; n = 1, \ldots, N \tag{A9}$$

Again, we also have the conservation of mass equations

$$\begin{aligned} M &= V([M] + \sum_{n=1}^{N}[ML^n]) \\ L &= V([L] + \sum_{n=1}^{N}[ML^n]) \end{aligned} \tag{A10}$$

To determine the heat potential Q, we further assume that the enthalpy change of each association event is identical

$$ML^{n-1} + L \overset{\Delta H}{\rightarrow} ML^n \;\;,\;\; n = 1, \ldots, N \tag{A11}$$

Assigning the fully dissociated macromolecule a reference heat potential of zero, the heat potential can be written:

$$Q = \Delta H(\sum_{n=1}^{N} n\,[ML^n]) \tag{A12}$$

If N is unknown, it can be determined by assigning a prior to it (such as $p(N|\mathcal{I}) \propto N^{-1}$) and sampling over a collection of models to determine which model is best supported by the evidence. Once determined, N can be fixed, and the thermodynamic parameters of binding determined given the fixed-N model.

3. N-component competitive binding

In an N-component competitive binding scenario, the macromolecule M can associate with any ligand $L_1, \ldots, L_N$

$$\begin{aligned} M + L_1 &\overset{K_{a1}}{\rightleftharpoons} ML_1 \\ &\vdots \\ M + L_N &\overset{K_{aN}}{\rightleftharpoons} ML_N \end{aligned} \tag{A13}$$

The various equilibrium constants are defined as

$$K_{an} = \frac{[ML_n]}{[M][L_n]} \tag{A14}$$

For $N > 1$, there is no longer a simple analytical solution for $[ML_n]$ given total quantities of macromolecule M and ligands L_n, so we must solve for the $[ML_n]$ numerically.

Given the association constants $\{K_{an}\}_{n=1}^{N}$ and total quantities of macromolecule M and ligand $\{L_n\}_{n=1}^{N}$, we solve for the equilibrium concentrations $[M]$, $[ML_n]$, and $[L_n]$ given the set of equations:

$$\begin{aligned} M &= V\left([M] + \sum_{n=1}^{N}[ML_n]\right) \\ L_n &= V([L_n] + [ML_n]) \\ [ML_n] &= K_{an}[M][L_n] \end{aligned} \tag{A15}$$

subject to the constraints that $[M] > 0$, $[L_n] > 0$, and $[ML_n] > 0$. This can be easily done with any number of nonlinear root-finding methods.

We presume that the binding of each species L_n is governed by a separate enthalpy of association ΔH_n:

$$M + L_n \overset{\Delta H_n}{\rightarrow} ML_n \tag{A16}$$

Assigning the fully dissociated state a reference heat potential of zero, the heat potential can be written:

$$Q = \sum_{n=1}^{N} \Delta H_n \, [ML_n] \tag{A17}$$

4. Cooperative binding with a Hill coefficient

For the case where ligand L binds cooperatively to macromolecule M with Hill coefficient n, association is described by the process

$$M + nL \overset{K_a, n}{\rightleftharpoons} ML^n \tag{A18}$$

where the association constant is defined by

$$K_a = \frac{[ML^n]}{[M][L]^n} \tag{A19}$$

Together with the conservation equations

$$\begin{aligned} M &= V([M] + [ML^n]) \\ L &= V([L] + n[ML^n]) \end{aligned} \tag{A20}$$

we can again numerically solve for $[M]$, $[L]$, and $[ML^n]$.

If we again define the fully dissociated species as having a heat potential of zero, the heat potential Q can be written

$$Q = \Delta H \, [ML^n] \tag{A21}$$

where ΔH refers to the enthalpy of association for n ligands. The per-ligand association enthalpy can be obtained by dividing ΔH by n.