

RESEARCH ARTICLE

# Bayesian analysis of isothermal titration calorimetry for binding thermodynamics

Trung Hai Nguyen<sup>1</sup>\*, Ariën S. Rustenburg<sup>2,3</sup>\*, Stefan G. Krimmer<sup>4</sup>, Hexi Zhang<sup>1</sup>, John D. Clark<sup>1</sup>, Paul A. Novick<sup>5</sup>\*, Kim Branson<sup>5</sup>\*, Vijay S. Pande<sup>5</sup>, John D. Chodera<sup>2\*</sup>, David D. L. Minh<sup>1\*</sup>

**1** Department of Chemistry, Illinois Institute of Technology, Chicago, IL, United States of America, **2** Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America, **3** Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY, United States of America, **4** Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, Marburg, Germany, **5** Department of Chemistry, Stanford University, Stanford, CA, United States of America

\* These authors contributed equally to this work.

\*a Current address: Nurix, San Francisco, CA, United States of America

\*b Current address: Genentech, South San Francisco, CA, United States of America

\* john.chodera@choderalab.org (JDC); dminh@iit.edu (DDLML)



**OPEN ACCESS**

**Citation:** Nguyen TH, Rustenburg AS, Krimmer SG, Zhang H, Clark JD, Novick PA, et al. (2018) Bayesian analysis of isothermal titration calorimetry for binding thermodynamics. *PLoS ONE* 13(9): e0203224. <https://doi.org/10.1371/journal.pone.0203224>

**Editor:** Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

**Received:** May 23, 2018

**Accepted:** August 16, 2018

**Published:** September 13, 2018

**Copyright:** © 2018 Nguyen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All computer code and data files are available from the github repository at the following link: <https://github.com/nguyentrunghai/bayesian-itc/tree/d8cbf43240862e85d72d7d0c327ae2c6f750e600>.

**Funding:** THN, DDLML, and JDC acknowledge support from the National Institutes of Health, <http://www.nih.gov/> (Grants P30 CA008748 and R01 GM121505 to JDC and Grant R15 GM114781 to DDLML). The funders had no role in study design,

## Abstract

Isothermal titration calorimetry (ITC) is the only technique able to determine both the enthalpy and entropy of noncovalent association in a single experiment. The standard data analysis method based on nonlinear regression, however, provides unrealistically small uncertainty estimates due to its neglect of dominant sources of error. Here, we present a Bayesian framework for sampling from the posterior distribution of all thermodynamic parameters and other quantities of interest from one or more ITC experiments, allowing uncertainties and correlations to be quantitatively assessed. For a series of ITC measurements on metal:chelator and protein:ligand systems, the Bayesian approach yields uncertainties which represent the variability from experiment to experiment more accurately than the standard data analysis. In some datasets, the median enthalpy of binding is shifted by as much as 1.5 kcal/mol. A Python implementation suitable for analysis of data generated by MicroCal instruments (and adaptable to other calorimeters) is freely available online.

## Introduction

Isothermal titration calorimetry (ITC) is a widely used biophysical technique for measuring the binding affinity between small molecules and biological macromolecules (such as proteins and RNA [1–4]), as well as between proteins [5]. In addition to simple two-component (one-to-one) binding processes, ITC may also be used to study more complex processes such as competitive binding [1, 6], binding cooperativity [7], and binding events coupled to changes in the protonation state [8, 9] or tautomeric state [10] of one or more components. Provided reaction rates are slower than cell mixing times, ITC can even be used to study the kinetics of association [11–15].

data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** JDC is a member of the Scientific Advisory Board for Schrödinger LLC. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Here, we focus on the thermodynamics of simple two-component association (one-to-one binding). A unique and powerful property of ITC is that it can not only determine the free energy of binding ( $\Delta G$ ), but also decompose it into enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) without having to resort to multiple experiments at different temperatures to determine these quantities via the van't Hoff equation. This decomposition has been used to draw conclusions into, for example, how entropy is related to antibody flexibility [16] and ordering of disordered loops [17] during antibody affinity maturation. It has also been used to suggest that iterative improvements in generations of drugs result in their interactions being increasingly driven by enthalpy [18]. Furthermore, it has been used to suggest how force fields might be improved [19].

It is possible to perform enthalpy-entropy decomposition with ITC because the instrument not only detects a binding process, but can determine the heat of binding. The raw data from an ITC instrument is the differential power required to maintain the *titrand* in a sample cell (usually a macromolecule dissolved in buffer) at the same temperature as a reference cell as a *titrant* (usually a small molecule ligand) is injected into the former. The experimental data  $\mathcal{D}$  can be summarized as the measured heats of injection,  $\mathcal{D} \equiv \{q_1, q_2, \dots, q_N\}$  obtained by integrating the differential power over the duration of each injection. Thermodynamic parameters are then determined by fitting binding heat models (expressions for the heat in terms of unknown thermodynamic and experimental parameters) to the integrated heat [20]. The standard protocol for parameter estimation, implemented in the Origin software package [21] distributed with the popular MicroCal VP-ITC instrument [22], uses a nonlinear least squares fit to estimate the association constant  $K_a$ , enthalpy  $\Delta H$ , and stoichiometry  $n$  (number of binding sites per mole of receptor), along with their estimated uncertainties.

Unfortunately, this established procedure for analyzing ITC data does not accurately determine uncertainties for enthalpy-entropy decomposition because it fails to account for all relevant sources of error. In a large-scale interlaboratory study (ABRF-MIRG'02) of a model protein: small molecule binding reaction—the binding of carboxybenzenesulfonamide (CBS) to bovine carbonic anhydrase II (CAII)—the variation among the reported ITC binding constant and enthalpy from 14 participants was more than an order of magnitude larger (and up to *three* orders of magnitude larger) than standard errors reported by the individual least squares analyses [23].

Spectrophotometric results suggested that titrant concentration errors were likely a major cause of this unexpectedly large variation. The standard analysis method accounts for error in the titrand concentration by treating the stoichiometry  $n$  as a free parameter that can take any real and positive value. On the other hand, the titrant concentration, likely an important source of discrepancies among laboratories [24], is often treated as exactly known. While precise titrant concentrations are systematically achievable [25], strong evidence suggests that large (10–20%) errors in titrant concentration are widespread even amongst laboratories skilled in biomolecular calorimetry [23]. It is possible to explicitly treat titrant concentration error in nonlinear least squares fitting [25], but this is not typically performed.

In addition to concentration error, another important source of error that is frequently neglected is the so-called *first injection anomaly*, in which the heat of injection from the first injection is smaller than expected. The anomaly often emerges due to backlash in the motorized screw mechanism used to drive the syringe plunger [26]; if the last operation of the plunger prior to the first injection is upwards, then less titrant will be injected via a subsequent downward movement of the plunger. This issue may be overcome by executing a short downward movement of the plunger prior to insertion into the sample cell. Another contributing factor to the first injection anomaly is leakage of titrant out of the syringe during instrument equilibration. Because the initial injection generally carries the largest magnitude of heat per

mole of titrant injected, the first injection anomaly (or the inability to account for it) can lead to significant errors in reported measurements.

Here we introduce a new data analysis protocol that accounts for these sources of error and, as we shall show, more accurately estimates the uncertainty in derived thermodynamic parameters—especially entropy and enthalpy. The approach is modular; additional sources of uncertainty or variability can be modeled through simple extensions of the model. Importantly, this analysis procedure also allows the joint uncertainties in entropy and enthalpy to be resolved, an essential requirement to evaluating hypotheses regarding entropy-enthalpy compensation. Our approach is based on Bayesian statistics, which uses the *posterior* probability distribution,

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (1)$$

where  $p(\mathcal{D}|\boldsymbol{\theta})$  is the *likelihood*, a conditional probability of observing data  $\mathcal{D}$  (in our case, the injection heats  $\{q_1, \dots, q_N\}$ ) given unknown thermodynamic parameters  $\boldsymbol{\theta}$ .  $p(\boldsymbol{\theta})$  is the *prior* probability, a function describing foreknowledge of the parameters  $\boldsymbol{\theta}$  before conditioning this distribution on the observed data  $\mathcal{D}$  from this experiment.

A Bayesian analysis has several significant potential advantages over the standard analysis protocol, including:

1. **Multimodal posteriors:** Bayesian analysis makes no assumptions about the shape of the posterior. Therefore, it can treat multimodal posteriors in which two or more distinct sets of parameters describe the data. On the other hand, the standard analysis assumes a multivariate Gaussian, which is based on a single mode.
2. **Nonlinear parameter correlation:** It is feasible to determine whether parameters are correlated, even if correlations are nonlinear.
3. **Modularity:** Additional sources of uncertainty can be incorporated in a modular fashion simply by adding more random variables (nuisance parameters) with associated priors.
4. **Integration of multiple experiments:** It is possible to incorporate information from multiple measurements and even from multiple experimental techniques. The posterior probability of a parameter is simply the product of posteriors for each measurement. Information from control experiments, such as a blank titration or prior standard measurements, can be incorporated into the prior.
5. **Optimal experimental design:** New experiments that maximize the gain of new information can be automatically identified. By using techniques from Bayesian experimental design [27], one can choose among many potential experiments those that would maximize the gain of *new* information, either sequentially or in batches.

To clarify, it is possible for analyses based on nonlinear regression to integrate some of these features. In nonlinear regression, parameter distributions are inherently non-Gaussian and two-dimensional contour plots of different parameters may have non-ellipsoid shapes, indicating nonlinear correlations [28, 29]. It is also possible to integrate multiple experiments with a global fit [30]. However, these features are not available in the standard protocol.

Recently, Duvvuri et al. [31] described a new python package for the Bayesian analysis of ITC experiments. For the analysis of single experiments, their results were consistent with Origin. They were also able to integrate data from multiple buffers, titrant/titrant ratios, and temperatures. However, they did not perform substantial error analysis.

Our present work is based on a different new python package and we more carefully consider the uncertainty of different analysis protocols. The primary criterion we use to evaluate and compare analysis protocols is based on interval estimates. Interval estimates have

somewhat different meanings in Bayesian and frequentist statistics. In frequentist statistics, the  $\alpha\%$  confidence interval is expected to contain the true value  $\alpha\%$  of the time. A confidence interval is inaccurate if the percentage of estimated intervals that contain the true value deviates from  $\alpha\%$ . In Bayesian statistics, the credible interval is not necessarily intended to contain the true value a specific percentage of the time; it is simply a region that contains  $\alpha\%$  of the posterior probability. Nonetheless, for the purposes of comparing uncertainties, we evaluate whether the Bayesian credible interval (BCI) obtained from our model serves as an accurate confidence interval compared to the confidence interval from the standard nonlinear regression protocol (NIRCI). Previously, BCIs have been shown to work well as confidence intervals for binding thermodynamics and reference scattering patterns in analyses of X-ray scattering experiments of protein:ligand binding [32].

## Methods

### Simulated isothermal titration calorimetry data

To assess the accuracy of BCIs in a context where all sources of error are known, we simulated ITC data in which the integrated heat is given by,

$$q_i = q_i^*(\Delta G, \Delta H, \Delta H_0, [R]_0, [L]_s) + \epsilon_i, \quad (2)$$

where  $q_i^*$  is the model integrated heat at injection  $i$  for an ITC without instrument noise, given by Eq S3 in [S1 Appendix](#). It is a function of several parameters,

- $\Delta G$ : the free energy of binding,
- $\Delta H$ : the enthalpy of binding,
- $\Delta H_0$ : the enthalpy of dilution and stirring per injection, and
- $[L]_s$  and  $[R]_0$ : the concentrations of titrant in the syringe and of the titrand in the cell, respectively.

In the simulated curves,  $\Delta G = -10$  kcal/mol,  $\Delta H = -5$  kcal/mol, and  $\Delta H_0 = 0.5$   $\mu$ cal. The concentrations were drawn from lognormal distributions with the stated values of 0.1 mM for  $[R]_0$  and 1 mM for  $[L]_s$  and with uncertainty 10%.  $\epsilon_i$  represents error in the measurement of  $q_i$  and was modeled as normal variable with zero mean and standard deviation of 1  $\mu$ cal. The number of injections was 24 ( $i \in [1, 24]$ ). A total of 50 simulated heat curves were generated.

### Titration of Mg(II) into EDTA

In order to assess the effectiveness of the Bayesian approach in describing the true uncertainty in the experimental measurements, we studied a simple complexation reaction—the 1:1 binding of Mg(II) to the chelator EDTA—for which multiple experimental replicates can be easily collected. The entire ITC experiment was repeated *from scratch*—with all solutions prepared completely independently so that any concentration errors would be fully independent—a total of 14 times. This is critical, as simply repeating the experimental measurement with the same stock solutions would not capture the true experimental variability. For each trial, the titrant (MgCl<sub>2</sub>), titrand (EDTA), and buffer (50 mM Tris-HCl pH 8.0) were weighed and dissolved to prepare solutions at the two planned concentrations for the titrant MgCl<sub>2</sub> and the titrand EDTA. In the first five trials, we prepared the titrant and titrand concentrations as 1.0 mM and 0.1 mM, respectively. In the other nine trials, the titrant and titrand concentrations were prepared as 0.5 mM and 0.05 mM, respectively.

Magnesium chloride hexahydrate [ $\text{MgCl}_2 \cdot (\text{H}_2\text{O})_6$ ] was purchased from Fisher Scientific (Catalog No. BP214-500, Lot No. 006533) and anhydrous ethylenediaminetetraacetic acid (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 by dropwise titration with HCl or NaOH. Solutions were prepared by weighing powder and adding the appropriate amount of buffer, neglecting the volume occupied by powder, to make a concentrated solution (15 mM for  $\text{MgCl}_2$  and 1.0 mM for EDTA). To maximize the number of significant figures, at least 0.1 g of  $\text{MgCl}_2$  and 0.01 g of EDTA were weighted out. The solutions were then further diluted with buffer to prepare the titrant and titrand. For example, to prepare a 0.1 mM solution of EDTA, a pipetman was used to measure 9 parts buffer to 1 part of 1.0 mM EDTA.

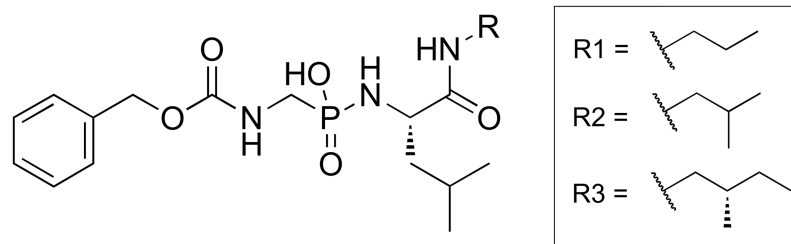
ITC measurements were performed on a MicroCal VP-ITC calorimeter. The experiments consisted of a total of 24 injections, with the first injection programmed to deliver 2  $\mu\text{L}$  of titrant ( $\text{MgCl}_2$ ) into the sample cell, and the remaining 23 injections programmed to deliver 12  $\mu\text{L}$ . 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  $\mu\text{L}/\text{s}$ . All experiments were conducted at 298.1 K, and the reference power was fixed at 5  $\mu\text{cal}/\text{s}$ .

The baseline was corrected and injection heats integrated using NITPIC [33].

### Titration of phosphoramidate-type inhibitors into thermolysin

To demonstrate our approach on protein:ligand systems, we also analyzed titrations of phosphoramidate-type inhibitors into thermolysin initially described in Krimmer et al. [34]. For each individual measurement, lyophilized thermolysin powder was freshly weighed (1.5–2 mg) and dissolved in an appropriate volume of buffer to achieve a concentration of 30  $\mu\text{M}$ . The concentration was confirmed by ultraviolet absorption at 280 nm. Prior to measurement, the thermolysin solution was centrifuged for 8 min at 8150 g. In contrast, one solution was prepared for all measurements with each ligand by dissolving the pure powder (0.3–0.4 mg) in buffer without the addition of DMSO. A MX5 microbalance from Mettler Toledo (Switzerland) with a readability of 1  $\mu\text{g}$  and a repeatability of 0.8  $\mu\text{g}$  was used for the sample weighting. Measurements were repeated in this fashion at least nine times. Results reported in [34] were based on three repetitions with a fresh batch of thermolysin and after optimizing ITC parameters. In contrast, our present analysis was based on all available data for each system except for a small subset with a large baseline shift in the middle of an injection.

Lyophilized thermolysin (EC number 3.4.24.2) from *Bacillus thermoproteolyticus* was purchased from Calbiochem (EMD Biosciences). The inhibitors (Fig 1) P-(((benzyloxy)carbonyl)amino)methyl-N-((S)-4-methyl-1-oxo-1-(propylamino)pentan-2-yl)phosphoramidic acid (ligand 1), P-(((benzyloxy)carbonyl)amino)methyl-N-((S)-1-(isobutylamino)-4-methyl-1-oxopentan-2-yl)phosphoramidic acid (ligand 2), and P-(((benzyloxy)carbonyl)amino)methyl-N-((S)-4-methyl-1-(((S)-2-methylbutyl)amino)-1-oxopentan-2-yl)phosphoramidic acid (ligand 3), were generously provided by Nader Nasief and David G. Hangauer (University of Buffalo, Buffalo, New York, USA), who synthesized and purified them as previously described [35]. In the paper [35], they claimed that the compounds were at least 95% pure. (Crystal structures of ligands 1 (PDB ID 4MXJ), 2 (PDB ID 4MTW), and 3 (PDB ID 4MZN) in complex with thermolysin have been previously reported [34]). All measurements were performed with a buffer composed of 20 mM HEPES (pH 7.5), 200 mM NaSCN, and 2 mM  $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$ . HEPES was purchased from Carl Roth (Catalog No. 9105.3, Batch no. 192184596), NaSCN was purchased from Fluka Analytical (Catalog No. 71938-1KG, Lot no.



**Fig 1. Chemical structures of thermolysin ligands used in this study.**

<https://doi.org/10.1371/journal.pone.0203224.g001>

BCBC9384V), and  $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$  was purchased from Carl Roth (Catalog No. T886.2, Lot no. 433205269). Prior to measurement, the buffer was filtered through a  $0.22 \mu\text{m}$  filter and degassed under reduced pressure.

ITC measurements with thermolysin were performed on an MicroCal ITC<sub>200</sub> calorimeter from GE Healthcare (Piscataway, New Jersey). After an initial delay 170 or 180 sec, the initial injection (0.3 to 0.5  $\mu\text{L}$ ) was followed by 19 to 26 main injections (1.2 to 1.5  $\mu\text{L}$ ). The duration of the injection (in sec) was twice the value of the volume (in  $\mu\text{L}$ ). All measurements were performed at a temperature of the measurement cell of 298.15 K, a stirring speed of 1000 rpm, titrand (thermolysin) concentration of 30  $\mu\text{M}$ , and a titrant (ligands 1-3) concentration of 400  $\mu\text{M}$ . For details on each protocol, see [S1 Table](#).

As with the  $\text{Mg(II):EDTA}$  data, the baseline was corrected and injection heats integrated using NITPIC [33]. Representative differential power and integrated heat curves for  $\text{Mg(II):EDTA}$  binding and thermolysin:ligand binding are shown in [S1 Fig](#). Some data were excluded due to large baseline shifts that could not be readily corrected.

### CBS:CAII dataset from the ABRF-MIRG'02 study

Finally, we considered a protein:ligand ITC dataset from a previously published study which demonstrated large interlaboratory variation far in excess of reported error estimates [23]. As data were unavailable from the authors, injection heat data were digitized from Fig 4 in [23], the ABRF-MIRG'02 paper, which includes 14 ITC datasets measured fully independently on identical source material (aliquots of CAII and dry powder stocks of CBS) by independent laboratories. Dataset 2 was generated by an instrument called the CSC 4200 ITC (see Table 2 in [23]) for which we could not find the user's manual to obtain information such as the cell volume. Therefore, we excluded this dataset. We also excluded dataset 4 because we were not able to reliably digitize the large number of injections. For other datasets, the experimental design parameters were taken from Table 2 in [23], while the reported thermodynamic parameters and standard errors were taken from Table 3 in [23]. In the ABRF-MIRG'02 study [23], most experiments obtained standard errors were using a nonlinear least squares fit. The exceptions were datasets 10 and 14, in which the standard deviation was obtained by repeating the same experiment 3 and 5 times, respectively. In these datasets, it was not clearly specified whether the entire experiment or just the titration was repeated in each replicate.

Digitization leads to some error. As a quantitative estimate of this error, consider the experiment from Group 1 in Fig 4 in [23], the ABRF-MIRG'02 paper. When zoomed in, the center of the axes and the markers can be identified to within a single pixel. The markers are 8 pixels high, and the plot is 182 pixels high. This translates to a maximum error of 1 pixel (marker center) on an axis conversion of 11 kcal/mol /  $182 \pm 2$  pixels  $0.060 \pm 0.001$  kcal/mol. Hence, the error (0.06 kcal/mol/injection) is rather small, even if done by eye.

### Frequentist confidence intervals

Origin software was used to perform nonlinear least squares fit of the heat data to obtain the binding constant  $K_a$ , enthalpy  $\Delta H$ , and the stoichiometry number  $n$ , and their corresponding standard errors. Each parameter was assumed to be normally distributed and the standard error was used as a standard deviation. The lower and upper bounds of the  $\alpha\%$  confidence interval were the  $1 - \alpha/2$  and  $1 + \alpha/2$  percentile, respectively, of the normal distribution with a mean as the point estimate and standard deviation as the reported uncertainty.

### Sampling from the Bayesian posterior

Our Bayesian model is constructed to infer the unknown true parameters,

$$\theta \equiv (\Delta G, \Delta H, \Delta H_0, [L]_s, [R]_0, \sigma), \tag{3}$$

In addition to the parameters described above in “Simulated ITC data”, this model includes  $\sigma$ , the standard error of heat measurement per injection.  $\sigma$  is a nuisance parameter in that it is not the objective of a measurement but is necessary to calculate the likelihood. The assumption of a constant  $\sigma$  is most reasonable if all injections include the same number of power measurements.

**Likelihood.** The data  $\mathcal{D} \equiv \{q_1, q_2, \dots, q_N\}$  consists of the observed heats per injection determined by integrating the differential power over the injection time. The corresponding data likelihood function was based on the assumption that, because the *observed* injection heat  $q_n$  is the sum of many power measurements, the measurement error added to the *true* (unknown) heat  $q_n^*$  will be normally distributed due to the central limit theorem,

$$q_n \sim \mathcal{N}(q_n^*(\theta), \sigma^2). \tag{4}$$

The total data likelihood for  $\mathcal{D} \equiv \{q_1, q_2, \dots, q_N\}$  is therefore given by

$$p(\mathcal{D}|\theta) = \frac{1}{(2\pi)^{N/2} \sigma^N} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=1}^N (q_n - q_n^*(\theta))^2 \right]. \tag{5}$$

The model heats  $q_n^*(\theta)$  are a function of the parameters  $\theta$ . See [S1 Appendix](#) for details of the binding model relating  $\theta$  to the true heats  $q_n^*$ .

**Priors.** The prior  $p(\theta)$  was a product of priors for each parameter,  $p(\theta) = \prod_j p(\theta_j)$ . Uniform priors were chosen for  $\Delta G$ ,  $\Delta H$ , and  $\Delta H_0$ :

$$\Delta G \sim \text{Uniform}(-40 \text{ kcal/mol}, 40 \text{ kcal/mol}), \tag{6}$$

$$\Delta H \sim \text{Uniform}(-100 \text{ kcal/mol}, 100 \text{ kcal/mol}), \tag{7}$$

$$\Delta H_0 \sim \text{Uniform}(q_{min} - \Delta q, q_{max} + \Delta q), \tag{8}$$

where  $q_{min} = \min\{q_1, q_2, \dots, q_N\}$ ,  $q_{max} = \max\{q_1, q_2, \dots, q_N\}$  and  $\Delta q = q_{max} - q_{min}$ , usually reported in units of cal.

We used three different sets of priors for the true concentrations of titrant in the syringe,  $[L]_s$ , and receptor in the cell,  $[R]_0$  ([Table 1](#)): General, Flat  $[R]_0$ , and Comparison. All of the concentration models make use of the fact that concentrations must be positive. In the General model, both concentrations are assigned lognormal priors with the mean and standard deviation given by their stated experimental values and corresponding experimental uncertainties

**Table 1. Summary of concentration priors used in this manuscript.**

| Name         | $[R]_0$   | $[L]_s$   | Purpose                           |
|--------------|---|---|-----------------------------------|
| General      | Lognormal,<br>$\delta[R]_0 = 0.1[R]_0^{stated}$ | Lognormal,<br>$\delta[L]_s = 0.1[L]_s^{stated}$   | Most future analyses              |
| Flat $[R]_0$ | Uniform   | Lognormal,<br>$\delta[L]_s = 0.1[L]_s^{stated}$   | When the true $[R]_0$ is unknown  |
| Comparison   | Uniform   | Lognormal,<br>$\delta[L]_s = 0.001[L]_s^{stated}$ | Comparison with standard analyses |

<https://doi.org/10.1371/journal.pone.0203224.t001>

due to preparation steps,

$$\ln [X]_0 \sim \mathcal{LN}([X]_0, (\delta[X]_0)^2). \tag{9}$$

The lognormal prior prevents negative values and the lognormal distribution is the maximum entropy distribution when the mean and variance of the logarithm of the parameter is specified. In the absence of specific quantification of the titrant concentration uncertainty, we assumed a value of  $\delta[X]_0$  equal to 10% of the provided  $[X]_0$ . This specified uncertainty is in line with quantification of typical laboratory titrant concentration errors observed by Myszka et. al. [23]. In cases where the practitioner uses an orthogonal method to quantify titrant concentration or carefully tracks the uncertainty during preparation steps, as described in Boyce et. al. [25], this more precise concentration uncertainty could be used instead. Alternatively,  $\delta[X]_0$  could be treated as a free nuisance parameter. Although the parameter may not be precisely determined from a single ITC experiment, it could potentially be elucidated by sampling from a Bayesian posterior based on multiple measurements with the same titrand solutions, analogous to a global fitting procedure possible with nonlinear least squares [36].

In the Flat  $[R]_0$  model, a uniform prior was used for  $[R]_0$  such that,

$$[R]_0 \sim \text{Uniform}\left(\frac{[R]_0^{stated}}{10}, 10[R]_0^{stated}\right). \tag{10}$$

This model is useful in cases where the receptor concentration is not clearly known, such as when the sample is impure or partially degraded. Due to potential degradation of protein used in some ITC measurements, we used this model in our analysis of data for thermolysin. Finally, in the Comparison model, we used a uniform prior for  $[R]_0$  and a sharply peaked prior for  $[L]_s$  such that  $\delta[L]_s = 0.001[L]_s^{stated}$ .

The Comparison model mimics the treatment of concentrations in standard nonlinear least squares fitting. In the standard procedure,  $[L]_s$  is assumed to be precisely the stated value while  $[R]_0$  can take any positive value that minimizes the total residual sum of squares. There is no penalty for changing  $[R]_0$  from its stated value. This is consistent with flat prior for  $[R]_0$  a sharp prior for  $[L]_s$ . On the other hand, the General model allows for but penalizes deviations from the stated values. In the absence of further information, we believe that the General model is the most justified of the three models because concentrations are likely to be close to their stated values. Our main reason for performing calculations with the Comparison model was to isolate the effects of concentration models from other aspects of the Bayesian analysis.

Finally, since even its order of magnitude may be unknown, an uninformative Jeffreys prior [37] was assigned to the noise parameter  $\sigma$ ,

$$p(\sigma) \propto (\sigma/\sigma_0)^{-1}. \tag{11}$$



where  $\sigma_0 \equiv 1 \text{ cal}$  is a reference quantity that simply renders the ratio  $\sigma/\sigma_0$  dimensionless. This model assumes that the injection heat measurement uncertainty  $\sigma$  is constant for all injections. This may be a good approximation when the same number of power measurements are integrated for each injection (i.e., when injections are of identical duration), but when experiments contain injections of different durations, the noise variance  $\sigma^2$  should be proportional to the number of power measurements summed to give the injection heat (with all other things being held constant). More complex noise variance models (such as those considered in [38]) could also be considered. The noise model could also be improved using calibration experiments based on the same protocol (such as blank titrations), or even other data collected on the instrument for other systems; in these cases, likelihoods from independent experiments are simply multiplied.

While we used uninformative priors (except for our concentrations) in this study, alternative priors for other parameters can be used. If some knowledge of thermodynamic parameters or concentrations is available from another type of experiment, e.g., spectrophotometric measurements, then these can be incorporated into their respective priors. In such cases, the prior could be normally-distributed with the sample mean and standard deviation as parameters. Another way to parameterize the prior for concentrations is 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). Alternatively, the posterior from a previous (e.g., pilot) ITC experiment can be used as the prior to integrate the information from a second ITC experiment with different experimental parameters.

**Sampling from the posterior.** In principle, Bayesian statistical inference could be performed based on a direct analysis of the posterior distribution (Eq 1) to obtain properties such as the mean, median, mode, credible intervals, and marginal distributions. However, most Bayesian posterior distributions are complex and multidimensional. They are not amenable to exact mathematical solutions. An alternative approach is to generate random samples from a distribution and analyze these samples to obtain the desired properties. Ideally, it would be feasible to use acceptance-rejection or another method that generates independent and identically distributed variates. Unfortunately, the complexity of Bayesian posteriors typically precludes these types of algorithms. In the last few decades, computing advances have made it possible to sample from Bayesian posterior distributions using Markov chain Monte Carlo (MCMC) [39] methods. In MCMC simulation, samples from a distribution are not entirely independent. Rather, a new sample is generated based on a perturbing the previous sample. After a sufficiently number of MCMC iterations, samples are no longer correlated with each other and can be regarded as independent. With a sufficient number of independent samples from the posterior, summary statistics of interest can be calculated.

In our MCMC simulations, initial values were chosen as follows:

- for  $[L]_s$  and  $[R]_0$ , the stated (intended) concentration was used
- for  $\Delta H$ ,  $\Delta G$ , and  $\Delta H_0$ , initial values of zero (in their appropriate energy units) were used
- for  $\sigma$ , the standard deviation of the last four injection heats was used as an initial guess

Parameters were updated by sequential Gibbs sampling. In sequential Gibbs sampling, one parameter is updated at a time using the Metropolis-Hastings algorithm [40, 41]:

1. For each single parameter, a proposal is drawn from a normal distribution centered at the current value, and a scale of unity for  $\Delta H$ ,  $\Delta G$ , and  $\Delta H_0$ , or the initial guess value for  $\sigma$ ,  $[L]_s$ , and  $[R]_0$ ;

- The trial move is accepted or rejected according to the Metropolis criterion. If it is accepted, the next value in the Markov chain is the trial move. If it is rejected, the next value in the Markov chain is the original value.

MCMC was performed using a python library that we wrote, `bayesian-itc` (<https://github.com/choderalab/bayesian-itc>). `bayesian-itc` uses the Metropolis-Hastings implementation in the PyMC [42] library to perform MCMC sampling. For each experiment, sampled parameters were stored after every 2000 MCMC trial moves for a total of 5000 samples. These calculations take about 11 hours on a single modern CPU. Each sample from the Bayesian posterior is a set of six values, as described in Eq 3. For each parameter, the  $\alpha\%$  BCI is estimated based on the shortest interval that contains  $\alpha\%$  of the MCMC samples.

The precise version of the library used in this manuscript was committed to github on May 2, 2018. It is freely available at <https://github.com/nguyentrunghai/bayesian-itc/tree/d8cbf43240862e85d72d7d0c327ae2c6f750e600>. The directory entitled `analysis_of_Mg2EDTA_ABRF-MIRG02_Thermolysin` contains all the data needed to reproduce the figures in this manuscript.

### The Kullback-Leibler divergence quantifies differences between thermodynamic parameter distributions obtained from Bayesian and nonlinear least-squares approaches

To compare posterior marginal distributions, we computed the Kullback-Leibler divergence (KL-divergence), between the posterior marginal densities in the two most important thermodynamic quantities of interest,  $(\Delta G, \Delta H)$ ,

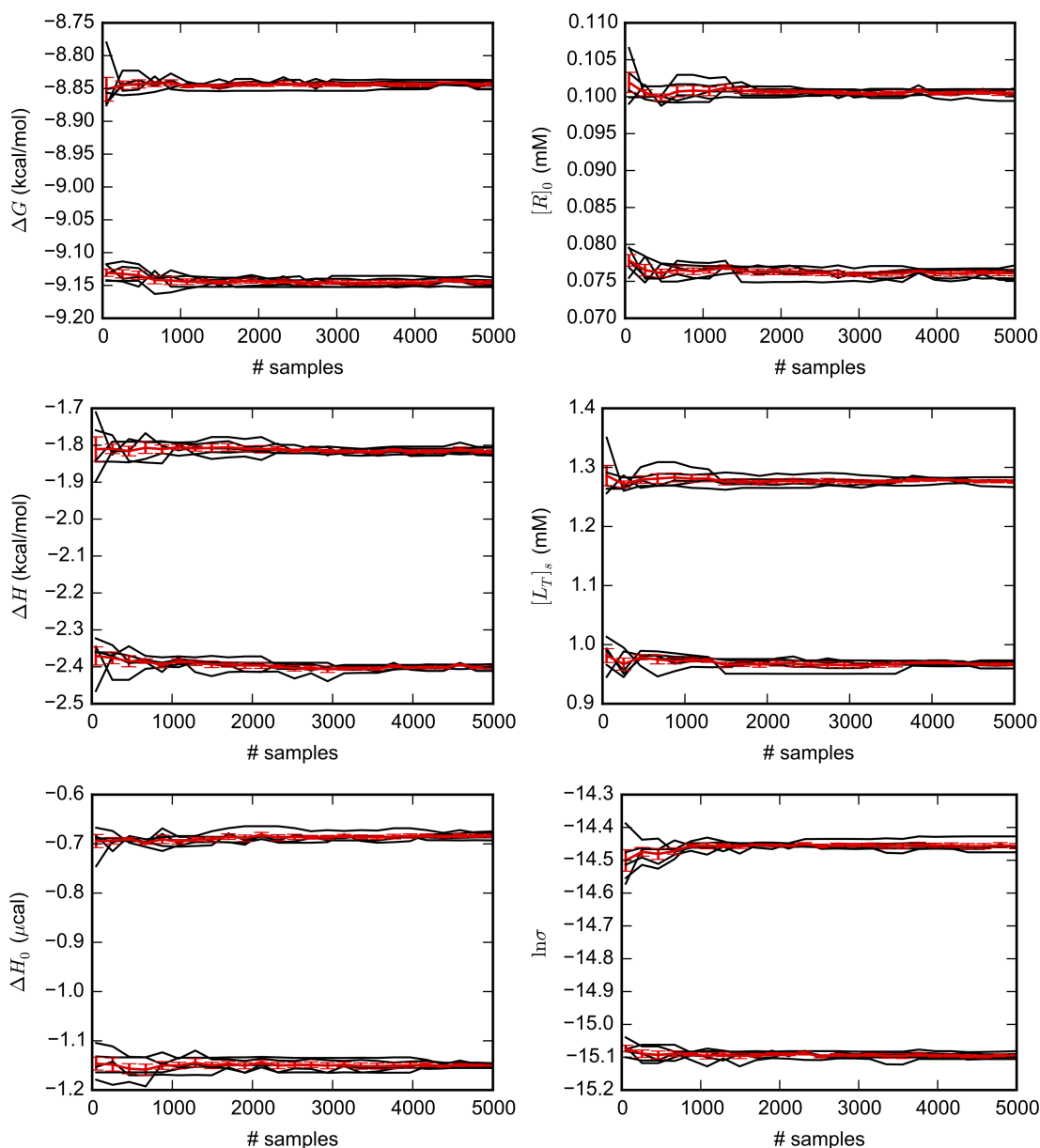
$$D_{KL}[p(\Delta G, \Delta H|\mathcal{D}_1)||p(\Delta G, \Delta H|\mathcal{D}_2)] = \int d\Delta G \int d\Delta H p(\Delta G, \Delta H|\mathcal{D}_1) \ln \frac{p(\Delta G, \Delta H|\mathcal{D}_1)}{q(\Delta G, \Delta H|\mathcal{D}_2)}. \quad (12)$$

$p(\Delta G, \Delta H|\mathcal{D}_1)$  and  $p(\Delta G, \Delta H|\mathcal{D}_2)$  are the posterior marginal densities specified by two different experiments with associated datasets  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . This metric, commonly used as a measure of deviation between two probability densities, can be interpreted as the amount of information lost when  $p(\Delta G, \Delta H|\mathcal{D}_2)$  is used to approximate  $p(\Delta G, \Delta H|\mathcal{D}_1)$ . The marginal posterior density for each experiment was estimated by using a Gaussian kernel density estimate (KDE) based on MCMC samples for  $\Delta G$  and  $\Delta H$  (ignoring other parameters). We used the `Kernel-Density` package implemented in `scikit-learn` [43] to estimate the density  $p(\Delta G, \Delta H)$ . The bandwidth for Gaussian kernel was set to 0.03 kcal/mol. Although the Kullback-Leibler divergence can be analytically computed for Gaussian densities, we also used the same KDE method to estimate probability densities  $p(\Delta G, \Delta H)$  for nonlinear regression. Samples for  $\Delta G$  and  $\Delta H$  were drawn from Gaussian distributions with the mean and standard deviation based on nonlinear regression point estimates and errors, respectively.

## Results and discussion

### Markov chain Monte Carlo sampling leads to precise estimates of Bayesian credible intervals

Our MCMC sampling protocol appears to yield precise estimates of 95% BCIs (Fig 2 and S2–S5 Figs). In all of the selected systems, the estimated 95% BCIs do not substantially change after considering about 2000 samples. The standard deviation of estimated upper and lower bounds over the five independent simulations in each system was less than 5% of the length of the average interval. Therefore, we are confident that the number of MCMC samples and



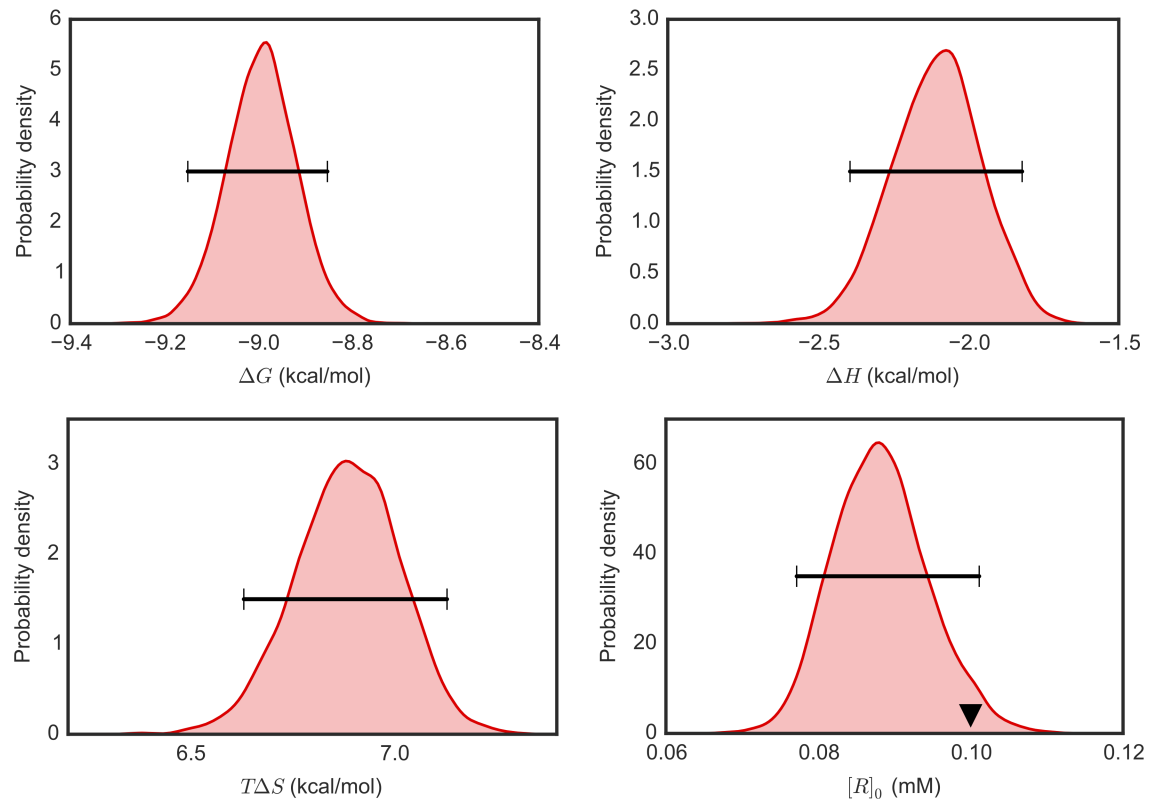
**Fig 2. Convergence of 95% Bayesian credible intervals (BCIs) with MCMC sampling.** 5000 MCMC samples were generated from the Bayesian posterior (General model) for several variables based on one ITC experiment measuring Mg(II):EDTA binding. For five independent repetitions of the MC simulations, the black lines are running estimates, as the number of samples is increased, of the upper and lower limits of 95% BCIs. The red line and error bars are the average and standard deviation across the five independent simulations. Similar plots for ligands 1-3 binding to thermolysin and CBS:CAII are available as S2–S5 Figs.

<https://doi.org/10.1371/journal.pone.0203224.g002>

mixing of the MCMC chain is sufficient to yield consistent estimates of the BCIs and other statistics of interest.

### Bayesian analysis yields unimodal distributions of linearly correlated parameters

Bayesian analysis permits multimodal posteriors and nonlinear parameter correlations to be investigated. Qualitative trends in the posterior density may be visualized by generating



**Fig 3. Representative 1D marginal distributions of thermodynamic parameters from Bayesian ITC analysis.** 1D marginal probability densities for thermodynamic parameters of interest were estimated based on 5000 MCMC samples generated from the Bayesian posterior (General model) for one ITC experiment measuring Mg(II):EDTA binding. Horizontal bars show 95% Bayesian credible intervals. The triangle in density plot of  $[R]_0$  indicates the stated value.

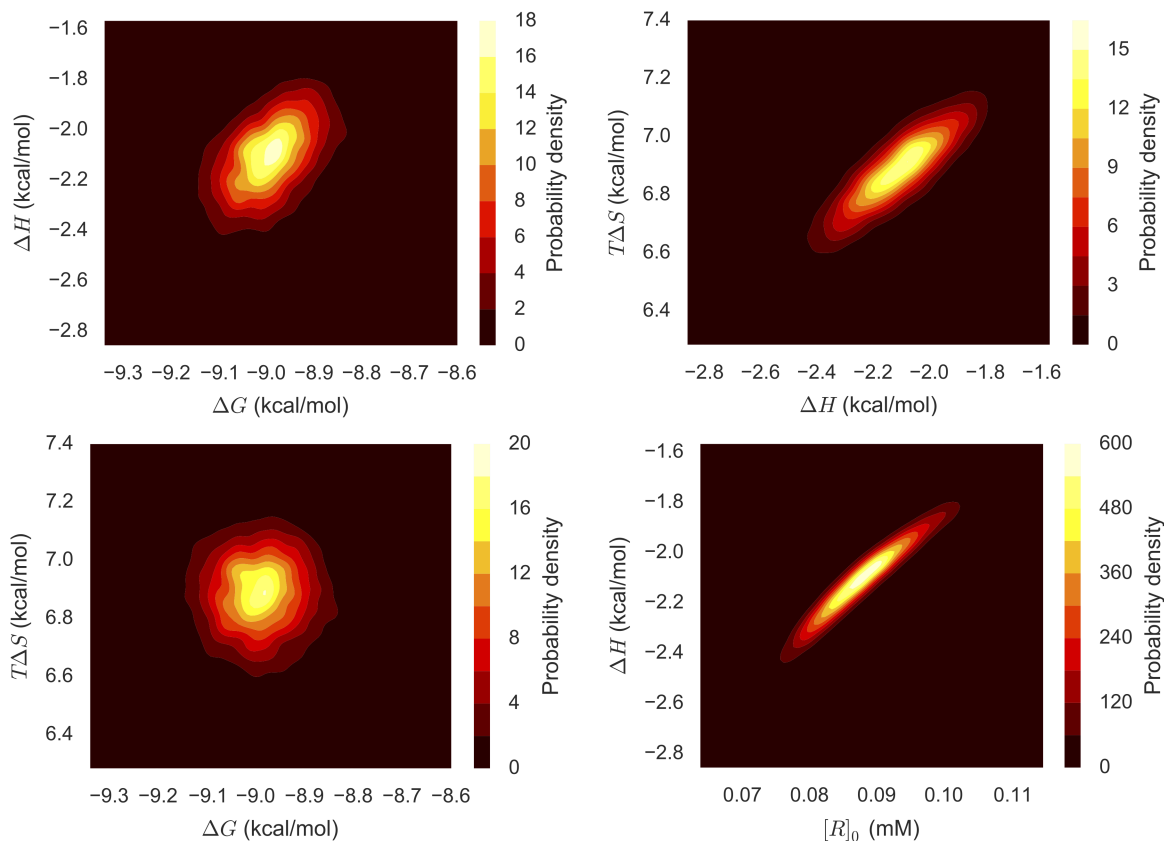
<https://doi.org/10.1371/journal.pone.0203224.g003>

histograms of MCMC samples drawn from the posterior. For our systems, representative 1D marginal distributions of key parameters (Fig 3) are unimodal. Although some skew is evident in  $\Delta H$ , the Gaussian distribution could be considered a reasonable approximation for most of these parameters. Our observation is consistent with previous analyses of nonlinear regression which showed that a Gaussian assumption is appropriate when the magnitude of statistical error is less than 10% of a parameter [28, 29].

Representative 2D marginal distributions (Fig 4) show that some pairs of parameters are nearly independent and others are highly correlated, with varying degrees of correlation in between. Of particular interest is the fact that while the free energy  $\Delta G$  and enthalpy  $\Delta H$  are mostly uncorrelated (top left of Fig 4), there is high correlation between the enthalpic ( $\Delta H$ ) and entropic ( $T\Delta S$ ) contributions to binding (top right of Fig 4) and between  $\Delta H$  and the receptor concentration  $[R]_0$  (bottom right of Fig 4). These correlations are not considered in the standard nonlinear regression analysis.

Given that the correlations appear to be linear, they can be succinctly summarized via the correlation coefficient. The estimated correlation matrix shown in Table 2 indicates that the titrant  $[L]_s$  and titrand concentrations  $[R]_0$  are highly correlated with each other and with the enthalpy  $\Delta H$  but only weakly with  $\Delta G$ . This result is consistent with Tellinghuisen [44], who evaluated the sensitivity of the binding constant and enthalpy to changes in concentration.

Estimates of concentrations and  $\Delta H$  are correlated because the effect of changing one of the parameters can be largely counteracted by changing another. When samples from a Bayesian



**Fig 4. Representative 2D marginal distributions of pairs of thermodynamic properties from Bayesian ITC analysis.** 2D joint marginal probability densities were estimated based on 5000 MCMC samples generated from the Bayesian posterior (General model) for one ITC experiment measuring Mg(II):EDTA binding.  $T\Delta S$  was derived from the sampled parameters  $\Delta G$  and  $\Delta H$  to aid our discussion of enthalpy-entropy compensation.

<https://doi.org/10.1371/journal.pone.0203224.g004>

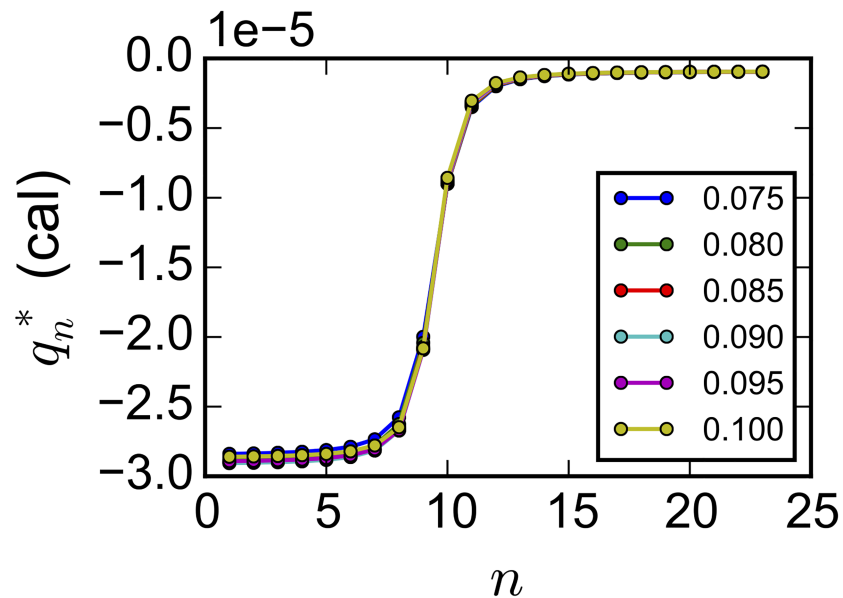
posterior for Mg(II):EDTA binding were used to parameterize a simple linear model for  $[L]_s$  and  $\Delta H$  as a function of  $[R]_0$ , different parameter values led to essentially the same integrated heat curve (Fig 5). An important implication of this enthalpy-concentration compensation is that given a measured integrated heat curve, the precise values of the three parameters are underdetermined; by itself, ITC cannot simultaneously determine the titrant or titrand concentration and the enthalpy of binding.

Due to the observed correlations, *apparent* enthalpy-entropy compensation [45] is a possible consequence of simple concentration error. If multiple measurements of the *same*

**Table 2. Correlation matrix estimated from the Bayesian posterior (General model) for an Mg(II):EDTA binding dataset.** Numbers in parentheses denote the uncertainty in the last digit.

|              | $\Delta G$ | $\Delta H$ | $\Delta H_0$ | $[L]_s$    | $[R]_0$    | $\ln \sigma$ |
|--------------|------------|------------|--------------|------------|------------|--------------|
| $\Delta G$   | 1          | 0.50(1)    | 0.280(8)     | 0.541(9)   | 0.553(9)   | 0.00(2)      |
| $\Delta H$   | 0.50(1)    | 1          | -0.07(1)     | 0.9893(2)  | 0.9884(2)  | 0.01(1)      |
| $\Delta H_0$ | 0.280(8)   | -0.07(1)   | 1            | 0.00(1)    | 0.01(1)    | 0.016(9)     |
| $[L]_s$      | 0.551(9)   | 0.9893(2)  | 0.00(1)      | 1          | 0.99887(3) | 0.01(1)      |
| $[R]_0$      | 0.553(9)   | 0.9884(2)  | 0.01(1)      | 0.99887(3) | 1          | 0.01(1)      |
| $\ln \sigma$ | 0.00(2)    | 0.01(1)    | 0.016(9)     | 0.01(1)    | 0.01(1)    | 1            |

<https://doi.org/10.1371/journal.pone.0203224.t002>



**Fig 5. Integrated heat for different values of titrand concentration  $[R]_0$  for Mg(II):EDTA binding.** Corresponding values of  $[L]_s$  and  $\Delta H$  were based on a simple linear regression of  $[L]_s$  and of  $\Delta H$  versus  $[R]_0$ . The other parameters ( $\Delta G$ ,  $\Delta H_0$ ) took the last value from the MCMC time series (General model). The legend shows the titrand concentration  $[R]_0$ , in mM, corresponding to each line.

<https://doi.org/10.1371/journal.pone.0203224.g005>

receptor-ligand pair were performed with different disturbances to the receptor concentration (e.g. small dilutions),  $\Delta H$  would be perturbed more significantly than  $\Delta G$ , leading to an apparent trend between  $\Delta G$  and  $T\Delta S$ .

### Median enthalpy estimates are sensitive to the titrand concentration model

Even though nonlinear least squares fitting and Bayesian analysis are based on the same binding model, other variations in the analysis procedure may lead to different estimates of  $\Delta G$  and  $\Delta H$ . We compare different analysis methods by considering how the median (which is less sensitive to outliers than the mean) of each quantity within a dataset. For all datasets, the median estimate of  $\Delta G$  is largely consistent across the different analysis methods. In contrast, with the thermolysin datasets,  $\Delta H$  estimates are consistent between all models except for the General model, which differ by as much as 1.5 kcal/mol (Table 3).

The consistency between all models except for the General model indicates that the major reason for discrepancy is the prior on the receptor concentration. In all but the General model, the titrand concentration freely changes (subject to the constraint  $[R]_0 > 0$ ) from stated

**Table 3. Median estimates of  $\Delta H$  (kcal/mol).** For nonlinear least squares, the value is the median of the different point estimates across different measurements. For Bayesian analysis, it is the median of the median sample from each Bayesian posterior. The numbers in parentheses are standard deviations estimated by bootstrapping: resampling the datasets (for nonlinear least squares) or the MCMC samples (for Bayesian analysis) with replacement using 1000 replicates.

| Dataset              | Nonlinear least squares | Comparison    | Flat $[R]_0$ | General       |
|----------------------|-------------------------|---------------|--------------|---------------|
| Mg(II):EDTA          | -2.25 (0.06)            | -2.24 (0.05)  |              | -2.19 (0.04)  |
| ligand 1:thermolysin | -5.84 (0.27)            | -5.89 (0.22)  | -5.92 (0.21) | -4.41 (0.31)  |
| ligand 2:thermolysin | -5.08 (0.07)            | -4.92 (0.06)  | -4.96 (0.07) | -4.23 (0.09)  |
| ligand 3:thermolysin | -5.28 (0.08)            | -5.10 (0.09)  | -5.14 (0.1)  | -4.46 (0.29)  |
| CBS:CAII             | -10.36 (0.6)            | -10.22 (0.53) |              | -10.05 (0.76) |

<https://doi.org/10.1371/journal.pone.0203224.t003>

concentration without penalty. In the General model, the prior penalizes deviations from the stated value of  $[R]_0$ . Estimates of the concentration affect  $\Delta H$  but not  $\Delta G$  because concentrations are highly correlated with  $\Delta H$  but not with  $\Delta G$ .

It is also evident that the titrand concentration is the determining factor for the shift in median  $\Delta H$  because the titrant concentration is lognormal in all the Bayesian priors. Modifying the standard deviation in the lognormal distribution affects credible intervals but does not change the median. By elimination, the factor that leads to the shift in the median is usage of a lognormal instead of uniform prior for the titrand concentration.

One result that is insensitive to the choice of concentration model is the one-to-one stoichiometry of these binding processes. Even when the prior on  $[R]_0$  is completely flat, concentrations sampled from the posterior are fairly close to the stated concentration. Sampled  $[R]_0$  that are a factor of two (or more) greater than the stated concentrations could indicate multiple ligands binding to each receptor molecule. Since sampled concentrations are similar to stated concentrations, the selected one-to-one binding model is suitable for these systems.

### Bayesian confidence intervals are more consistent with each other than those from the standard analysis

In addition to the median enthalpy, the width and consistency of intervals is also dependent on the concentration model (See Table 4, Fig 6, and S8–S19 Figs). For  $\Delta H$  and  $[R]_0$  in particular, NIRCIs and BCIs based on the Comparison model are narrower and correspondingly less consistent with one another than BCIs based on other concentration models. BCIs based on the General model are substantially broader and those based on the flat  $[R]_0$  model are broader still. However, all the BCIs and NIRCIs for  $\Delta G$  are of comparable magnitude.

In contrast with the dependence of the shift in the median enthalpy on the titrand concentration model, the change in interval size is primarily driven by the titrant concentration model. The Comparison and Flat  $[R]_0$  model have the same uniform prior for the titrand concentration. However, the size of the  $\Delta H$  and  $[R]_0$  intervals for the Flat  $[R]_0$  model is much larger because the standard deviation in the lognormal model for  $[L]_s$  is larger. (Data are in figures noted in Table 4).

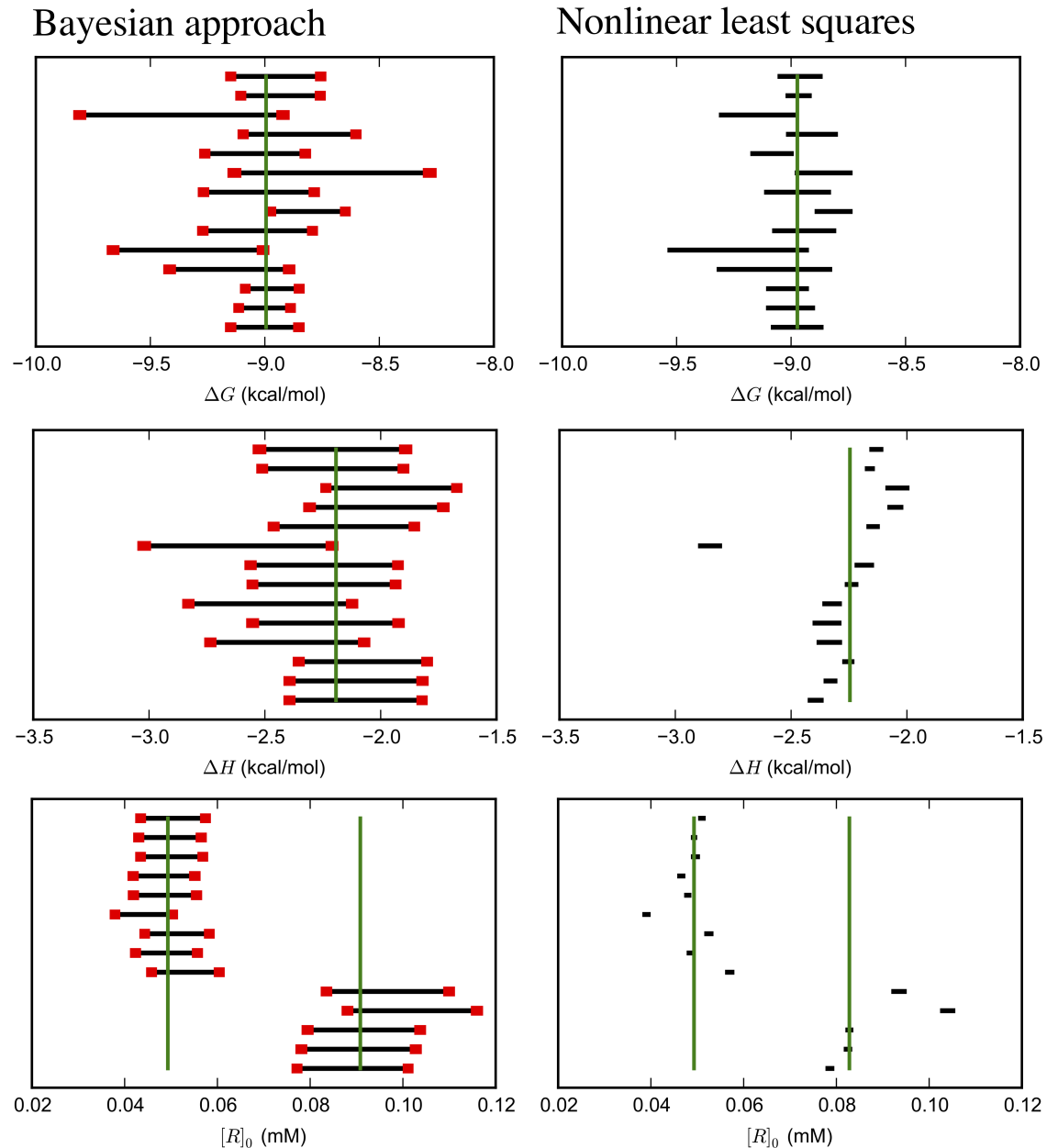
On a note related to the width and consistency between confidence intervals, nearly every pair of 95% BCIs for  $\Delta G$  and  $\Delta H$  from the General and Flat  $[R]_0$  model have at least some overlap with one another. (The 95% BCIs for  $[R]_0$  do not overlap when the stated concentrations differ, as in the Mg(II):EDTA and CBS:CAII datasets.) As with other statistics, BCIs based on the Comparison model are very similar to NIRCIs (S7, S10, S13, S16 and S18 Figs).

One complication with assessing confidence interval estimates is that we do not know the “true” value. Because we do not know the “true” value, we used the median value from repeated experiments as an approximation. The mean value is also a suitable choice, but the median is less sensitive to outliers.

**Table 4. Figure numbers for confidence interval plots in this manuscript.**

| System               | General | Flat $[R]_0$ | Comparison |
|----------------------|---------|--------------|------------|
| Mg(II):EDTA          | Fig 6   |              | S5 Fig     |
| ligand 1:thermolysin | S8 Fig  | S9 Fig       | S10 Fig    |
| ligand 2:thermolysin | S11 Fig | S12 Fig      | S13 Fig    |
| ligand 3:thermolysin | S14 Fig | S15 Fig      | S16 Fig    |
| CBS:CAII             | S17 Fig |              | S18 Fig    |

<https://doi.org/10.1371/journal.pone.0203224.t004>

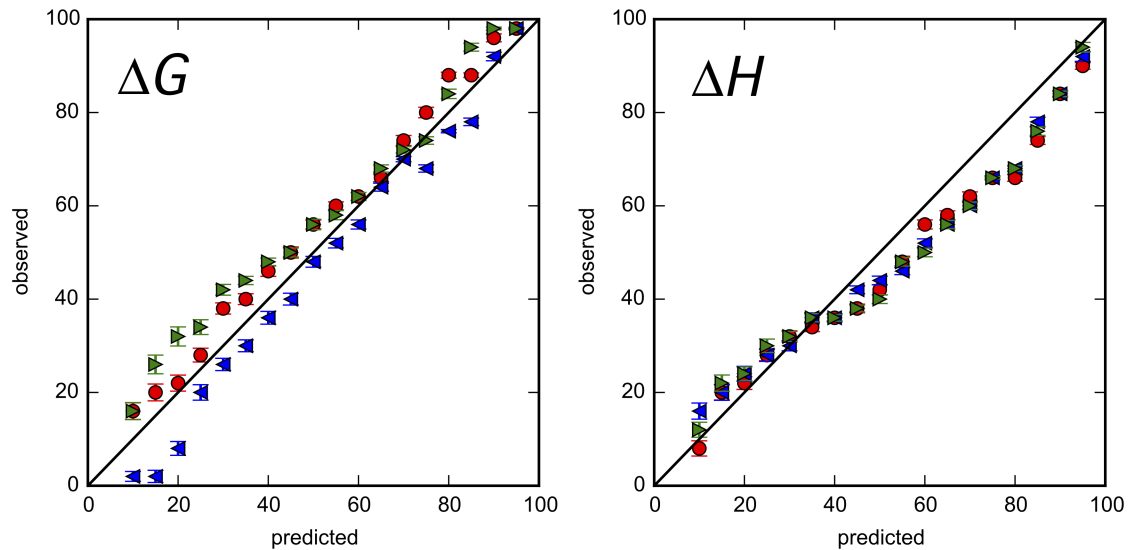


**Fig 6. Uncertainty estimates from Bayesian and nonlinear least squares analyses of Mg(II):EDTA ITC replicates.** 95% credible intervals estimated from the Bayesian posterior based on the General model (left) and confidence intervals from nonlinear least squares (right) for parameters specifying magnesium binding to EDTA. The vertical green lines are the median of the median MCMC samples. There are two median estimates for  $R$  because the experiments were done at two different concentrations. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

<https://doi.org/10.1371/journal.pone.0203224.g006>

Most of the 95% BCIs for  $\Delta G$ ,  $\Delta H$ , and  $[R]_0$  from the General and Flat  $[R]_0$  models contain the median. One exception is for the CBS:CAII dataset, in which BCIs for  $\Delta G$  capture the median less consistently. In contrast, while most 95% NIRCIs for  $\Delta G$  contain the median (except in the CBS:CAII dataset), the 95% NIRCIs for  $\Delta H$  and  $[R]_0$  generally do not. BCIs from the Comparison model behave similarly to NIRCIs (S5, S8, S11, S14 and S16 Figs). The size of these intervals appear to be significantly underestimated in all of our systems.

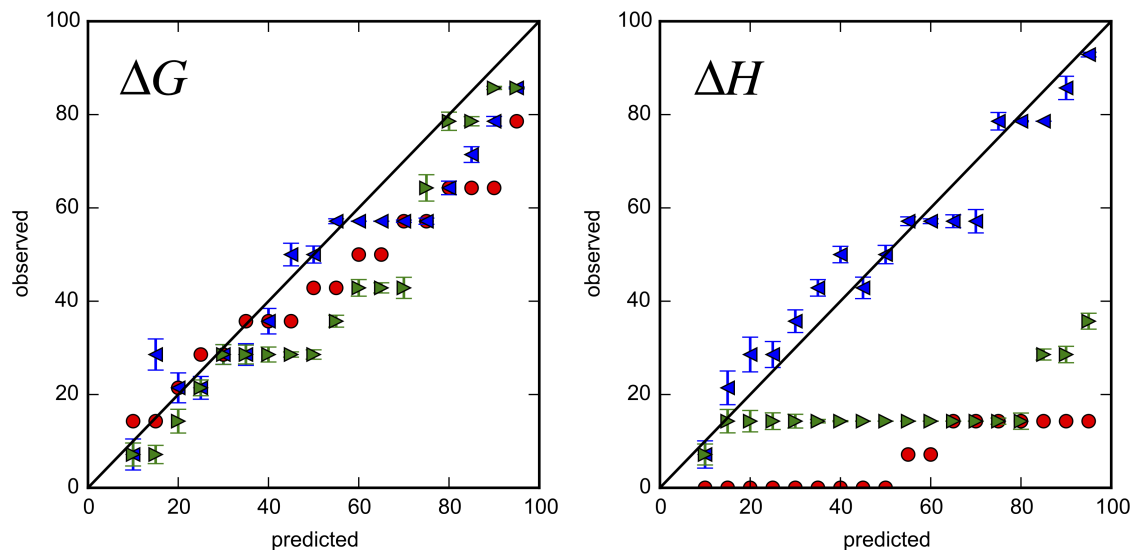




**Fig 7. Uncertainty validation for Bayesian analysis of simulated data.** The predicted versus observed rate (%) in which BCIs contain the true value (red circles), the mean (blue leftward triangles) or the median (green rightward triangles) for binding parameters are shown. Error bars are standard deviations based on bootstrapping.

<https://doi.org/10.1371/journal.pone.0203224.g007>

A better way to visualize the performance of confidence intervals is to compare the fraction of intervals that contain the true value with the stated confidence level. If stated levels are accurate, they should reflect the probability that the interval contains the true value. In this type of plot, therefore, data points should lie along the diagonal, the solid line of Figs 7 and 8, and S19–S22 Figs Points below the diagonal indicate that stated confidence intervals are too small. Conversely, points above the diagonal indicate that they are too large.



**Fig 8. Uncertainty validation for Bayesian and nonlinear least squares analyses of Mg(II):EDTA data.** For the Mg<sub>2</sub>:EDTA binding experiments, the predicted versus observed rate (%) in which intervals contain the median value for binding parameters is shown. Intervals were BCIs based on the General model (blue leftward triangles), Comparison model (green rightward triangles), or nonlinear least squares confidence intervals (red circles). Error bars are standard deviations based on bootstrapping.

<https://doi.org/10.1371/journal.pone.0203224.g008>

Simulated data are useful for performing this type of assessment in a context where true parameter values and all sources of error are known. We generated 50 simulated heat curves (S6 Fig) and estimated BCIs for each. As an assessment of the BCIs, we then plotted the fraction of BCIs that contain the true value against their stated levels. For the simulated data, the observed fraction of BCIs containing the true value (red dots) is very close to the stated levels (Fig 7), indicating that the Bayesian approach produces accurate confidence intervals.

The simulated data are also valuable for assessing our choice of medians as approximations to true parameter values. For the simulated data, the mean and median give almost the same estimate for the true value, but the median is slightly closer (Table 5). The blue and green dots in Fig 7 correspond to the observed fraction of BCIs containing the mean and the median, respectively. Except for the fraction containing the mean  $\Delta G$  at low confidence (blue dots on left panel of Fig 7), they are very close to the observed fraction of BCIs containing the true value. Because the performance of the median and mean is similar and the median is more robust with respect to outliers, the median is a reasonable approximation to the true value.

Based on uncertainty validation, BCIs based on the General model perform nearly ideally for Mg(II):EDTA and less reliably for the other experimental datasets. In the cases of Mg(II):EDTA and ligand 2:thermolysin binding, the observed fraction of BCIs (General model) for  $\Delta G$  and  $\Delta H$  that contain the median is very close to the ideal line (Fig 8 and S20 Fig). For the other datasets, BCIs based on the General model are less consistent with observed rates. In the cases of ligand 1:thermolysin and ligand 3:thermolysin, the median-containing frequency of  $\Delta G$  BCIs is also very close the ideal line whereas that  $\Delta H$  BCIs deviates from ideality, especially for larger confidence intervals (S19 and S21 Figs). In the CBS:CAII dataset, however, BCIs for  $\Delta H$  are more consistent with observed rates than for  $\Delta G$ .

Intervals from other models had variable performance. NIRCIs of  $\Delta G$  have similar performance to BCIs but the observed rate at which NIRCIs for  $\Delta H$  contain the median is significantly less than ideal. This deviation from ideality is consistent with the poorly overlapping 95% confidence intervals for  $\Delta H$ . BCIs from the Comparison model behave similarly to NIRCIs. In contrast with intervals from other models, BCIs based on the flat  $[R]_0$  model generally overestimate the width of intervals for the thermolysin model. The overestimation of intervals suggests that the uniform prior employed in this analysis is too uninformative.

Overall, our Bayesian method (with the General model) led to reasonable BCIs for multiple measurements performed by a single individual within a single laboratory. The performance of BCIs in accounting for laboratory-to-laboratory variability in the CBS:CAII datasets digitized from the ABRF-MIRG'02 paper [23] was weaker. In this dataset, there must be one or more significant sources of error that the present approach fails to account for.

The strong correlation between concentrations and  $\Delta H$  explains the dramatic improvement of the credible intervals of  $\Delta H$  (e.g. Fig 8) when the uncertainty in  $[L]_s$  is included in the Bayesian analysis. In the same vein, the weak correlation between concentrations and  $\Delta G$  explains why NIRCIs for  $\Delta G$  are reasonable (Fig 8) even if the titrant concentration was treated as exactly known in the fit. Trends in the accuracy of confidence intervals are consistent with previous analyses based on error propagation [24, 25, 44, 45], which showed that titrant

**Table 5. Comparison of Bayesian estimates of mean and median with the true values for simulated data.** The numbers in parentheses denote uncertainty in the last digit, which are standard deviations estimated by bootstrapping by resampling the MCMC samples and the datasets with replacement 1000 times.

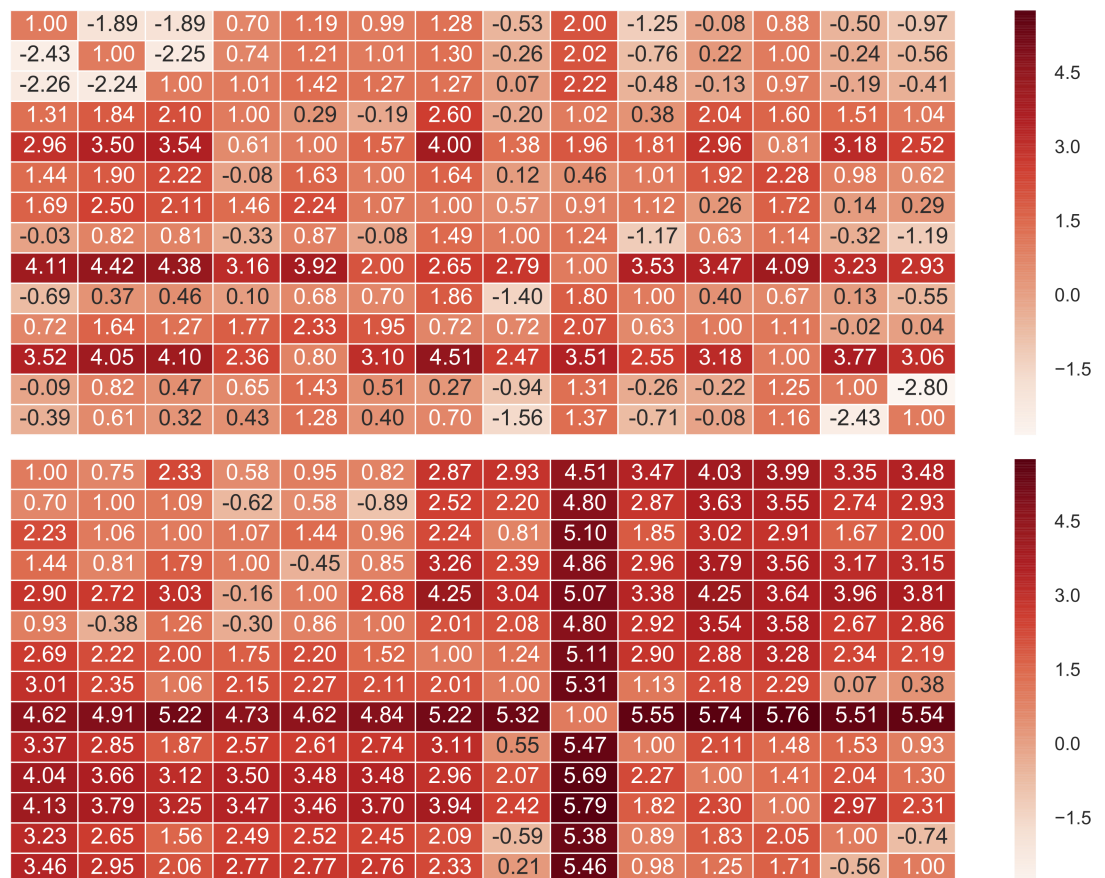
| Parameter             | True value | mean      | median    |
|-----------------------|------------|-----------|-----------|
| $\Delta G$ (kcal/mol) | -10        | -10.04(2) | -10.03(1) |
| $\Delta H$ (kcal/mol) | -5         | -5.04(5)  | -5.03(5)  |

<https://doi.org/10.1371/journal.pone.0203224.t005>

concentration errors propagate to small relative errors in  $\Delta G$  but large relative errors in  $\Delta H$ . If the error in titrant concentration is correctly propagated, it may be possible to make NIRCIs more accurate [25], but testing this is beyond the scope of the present work. In subsequent analysis, we will only consider the General model.

### Binding parameter distributions are more consistent with Bayesian analysis than nonlinear regression

In most datasets, the estimated Kullback-Leibler divergence between pairs of Bayesian posteriors is smaller than those estimated for nonlinear regression (Fig 9 and S23–S26 Figs). For the thermolysin datasets where the flat  $[R]_0$  model was tested, the Kullback-Leibler divergence for the flat  $[R]_0$  model was even smaller than for the general model. Therefore, marginals of the Bayesian posteriors are more consistent with one another than the Gaussian distributions from nonlinear regression. This finding agrees with above analyses that the Bayesian posterior captures the variance among experiments better than nonlinear least squares. The one exception is with the CBS:CAII dataset, in which the Kullback-Leibler divergence matrix based on the Bayesian method is comparable to the one from nonlinear regression (S26 Fig).



**Fig 9. The natural logarithm of Kullback-Leibler divergence between Bayesian approach (General model) and nonlinear least-squares.** The natural logarithm of the KL-divergence between posterior marginal distributions (top) and between Gaussian distributions of nonlinear least squares errors (bottom) is shown. Each column and row corresponds to one of the 14 datasets of Mg(II):EDTA binding. The diagonal elements should be  $\ln 0 = -\infty$  but were set to 1 for visualization.

<https://doi.org/10.1371/journal.pone.0203224.g009>

## Conclusion

In this study we have applied Bayesian statistics to analyze ITC data for the first time. We were able to account for various sources of error including, most importantly, uncertainties in the titrand and titrant concentrations. Due to the inclusion of concentration uncertainties, BCIs more accurately capture the variance between independent experiments than NIRCIs. In some datasets, the concentration error model led to differences in binding enthalpy estimates. Correlation between different parameters computed from the Bayesian posterior helps rationalize the effects of concentration uncertainty on the accuracy of  $\Delta G$  and  $\Delta H$ . Our analysis methods are freely accessible and extensible to more complex binding models, including the consideration of complex stoichiometry and cooperativity.

## Supporting information

**S1 Appendix. Description of simple two-component (1:1) association binding model.**  
(PDF)

**S1 Table. Experimental parameters of thermolysin ITC measurements.**  
(PDF)

**S1 Fig. Representative differential power and integrated heat.** From top to bottom: Mg(II): EDTA, ligand 1:thermolysin, ligand 2:thermolysin and ligand 3:thermolysin.  
(PDF)

**S2 Fig. Convergence of 95% credible intervals for ligand 1:thermolysin.** 5000 MCMC samples were generated from the Bayesian posterior (General model) for several variables based on one ITC dataset. For five independent repetitions of the MC simulations, the black lines are running estimates, as the number of samples is increased, of the upper and lower limits of 95% BCIs. The red line and error bars are the average and standard deviation across the five independent simulations.  
(PDF)

**S3 Fig. Convergence of 95% credible intervals for ligand 2:thermolysin.** 5000 MCMC samples were generated from the Bayesian posterior (General model) for several variables based on one ITC dataset. For five independent repetitions of the MC simulations, the black lines are running estimates, as the number of samples is increased, of the upper and lower limits of 95% BCIs. The red line and error bars are the average and standard deviation across the five independent simulations.  
(PDF)

**S4 Fig. Convergence of 95% credible intervals for ligand 3:thermolysin.** 5000 MCMC samples were generated from the Bayesian posterior (General model) for several variables based on one ITC dataset. For five independent repetitions of the MC simulations, the black lines are running estimates, as the number of samples is increased, of the upper and lower limits of 95% BCIs. The red line and error bars are the average and standard deviation across the five independent simulations.  
(PDF)

**S5 Fig. Convergence of 95% credible intervals for CBS:CAII.** 5000 MCMC samples were generated from the Bayesian posterior (General model) for several variables based on one ITC dataset for binding of CBS to CAII digitized from the ABRF MIRG'02 paper [23]. For five independent repetitions of the MC simulations, the black lines are running estimates, as the number of samples is increased, of the upper and lower limits of 95% BCIs. The red

line and error bars are the average and standard deviation across the five independent simulations.

(PDF)

**S6 Fig. Fifty simulated heat curves.** Parameters for the curves are in the Experimental section of the main text.

(PDF)

**S7 Fig. Uncertainty estimates from Bayesian (Comparison model) and nonlinear least squares analyses of Mg(II):EDTA ITC replicates.** 95% credible intervals estimated from Bayesian analysis (left) and confidence intervals from nonlinear least squares (right) for parameters specifying magnesium binding to EDTA. The vertical green lines are the median. There are two median estimates for  $R$  because the experiments were done at two different concentrations. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S8 Fig. Uncertainty estimates from Bayesian (General model) and nonlinear least squares analyses of ligand 1:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 1 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S9 Fig. Uncertainty estimates from Bayesian (Flat  $[R]_0$  model) and nonlinear least squares analyses of ligand 1:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 1 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S10 Fig. Uncertainty estimates from Bayesian (Comparison model) and nonlinear least squares analyses of ligand 1:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 1 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S11 Fig. Uncertainty estimates from Bayesian (General model) and nonlinear least squares analyses of ligand 2:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 2 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S12 Fig. Uncertainty estimates from Bayesian (Flat  $[R]_0$ ) and nonlinear least squares analyses of ligand 2:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for

parameters specifying ligand 2 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S13 Fig. Uncertainty estimates from Bayesian (Comparison model) and nonlinear least squares analyses of ligand 2:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 2 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S14 Fig. Uncertainty estimates from Bayesian (General model) and nonlinear least squares analyses of ligand 3:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 3 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S15 Fig. Uncertainty estimates from Bayesian (Flat  $[R]_0$  model) and nonlinear least squares analyses of ligand 3:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 3 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S16 Fig. Uncertainty estimates from Bayesian (Comparison model) and nonlinear least squares analyses of ligand 3:thermolysin ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying ligand 3 binding to thermolysin. The vertical green lines are the median. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S17 Fig. Uncertainty estimates from Bayesian (General model) and nonlinear least squares analyses of CBS:CAII ITC replicates.** 95% credible intervals estimated from the Bayesian posterior (left) and confidence intervals from nonlinear least squares (right) for parameters specifying CBS binding to CAII. The vertical green lines are the median. Note that each experiment was done at different concentration. Red bars denote the standard deviations of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S18 Fig. Uncertainty estimates from Bayesian (Comparison model) and nonlinear least squares analyses of CBS:CAII ITC replicates.** 95% credible intervals estimated from the Bayesian analysis (left) and confidence intervals from nonlinear least squares (right) for parameters specifying CBS binding to CAII. The vertical green lines are the median. Note that each experiment was done at different concentration. Red bars denote the standard deviations

of the lower and upper bounds, estimated by bootstrapping, and are a total of two standard deviations wide.

(PDF)

**S19 Fig. Uncertainty validation for Bayesian and nonlinear least squares analyses of ligand 1:thermolysin data.** For the ligand 1:thermolysin experiments, the predicted versus observed rate (%) in which intervals contain the median value for binding parameters is shown. Intervals were BCIs based on the General (blue leftward triangles), Flat  $[R]_0$  (black squares), and Comparison (green rightward triangles) models or nonlinear least squares confidence intervals (red circles). Error bars are standard deviations based on bootstrapping.

(PDF)

**S20 Fig. Uncertainty validation for Bayesian and nonlinear least squares analyses of ligand 2:thermolysin data.** For the ligand 2:thermolysin experiments, the predicted versus observed rate (%) in which intervals contain the median value for binding parameters is shown. Intervals were BCIs based on the General (blue leftward triangles), Flat  $[R]_0$  (black squares), and Comparison (green rightward triangles) models or nonlinear least squares confidence intervals (red circles). Error bars are standard deviations based on bootstrapping.

(PDF)

**S21 Fig. Uncertainty validation for Bayesian and nonlinear least squares analyses of ligand 3:thermolysin data.** For the ligand 3:thermolysin experiments, the predicted versus observed rate (%) in which intervals contain the median value for binding parameters is shown. Intervals were BCIs based on the General (blue leftward triangles), Flat  $[R]_0$  (black squares), and Comparison (green rightward triangles) models or nonlinear least squares confidence intervals (red circles). Error bars are standard deviations based on bootstrapping.

(PDF)

**S22 Fig. Uncertainty validation for Bayesian and nonlinear least squares analyses of CBS:CAII data.** For the CBS:CAII experiments, the predicted versus observed rate (%) in which intervals contain the median value for binding parameters is shown. Intervals were BCIs based on the General (blue leftward triangles), Flat  $[R]_0$  (black squares), and Comparison (green rightward triangles) models or nonlinear least squares confidence intervals (red circles). Error bars are standard deviations based on bootstrapping.

(PDF)

**S23 Fig. Logarithm of Kullback-Leibler divergence between posterior marginal distributions based on the General model (top) and flat  $[R]_0$  model (middle), and between Gaussian distributions of nonlinear least squares errors (bottom).** Each column and row corresponds to one of the 10 datasets of ligand 1:thermolysin binding. The diagonal elements should be  $\ln 0 = -\infty$  but were set to 1 for visualization.

(PDF)

**S24 Fig. Logarithm of Kullback-Leibler divergence between posterior marginal distributions based on the General model (top) and flat  $[R]_0$  model (middle), and between Gaussian distributions of nonlinear least squares errors (bottom).** Each column and row corresponds to one of the 11 datasets of ligand 2:thermolysin binding. The diagonal elements should be  $\ln 0 = -\infty$  but were set to 1 for visualization.

(PDF)

**S25 Fig. Logarithm of Kullback-Leibler divergence between posterior marginal distributions based on the General model (top) and flat  $[R]_0$  model (middle), and between**

**Gaussian distributions of nonlinear least squares errors (bottom).** Each column and row corresponds to one of the 11 datasets of ligand 3:thermolysin binding. The diagonal elements should be  $\ln 0 = -\infty$  but were set to 1 for visualization.

(PDF)

**S26 Fig. Logarithm of Kullback-Leibler divergence between posterior marginal distributions based on the General model (top) and flat  $[R]_0$  model (middle), and between Gaussian distributions of nonlinear least squares errors (bottom).** Each column and row corresponds to one of the 10 datasets of CBS:CAII binding. The diagonal elements should be  $\ln 0 = -\infty$  but were set to 1 for visualization.

(PDF)

## Acknowledgments

We thank Gerhard Klebe for facilitating the sharing of ITC data collected by his former student Stefan Krimmer. We thank Nader Nasief and David G. Hangauer for providing the ligands for Krimmer's ITC measurements. We also thank Joel Tellinghuisen for helpful discussions and comments on the manuscript. We thank an anonymous reviewer for suggesting the analysis of simulated data. JDC thanks Sarah Boyce for training on ITC instruments. Finally, we thank our summer interns at Illinois Tech: Mateus Pires Schneider (funded by the Capes Foundation within the Brazilian Ministry of Education) for performing NITPIC integrations on the thermolysin systems; and Erica Cusnariov for assistance with KL divergence figures.

## Author Contributions

**Conceptualization:** John D. Chodera, David D. L. Minh.

**Data curation:** Trung Hai Nguyen, John D. Chodera.

**Formal analysis:** Trung Hai Nguyen.

**Funding acquisition:** John D. Chodera, David D. L. Minh.

**Investigation:** Trung Hai Nguyen, Stefan G. Krimmer, Hexi Zhang, John D. Clark, Paul A. Novick, Kim Branson, David D. L. Minh.

**Methodology:** Ariën S. Rustenburg, John D. Chodera, David D. L. Minh.

**Project administration:** John D. Chodera, David D. L. Minh.

**Software:** Trung Hai Nguyen, Ariën S. Rustenburg, John D. Clark, John D. Chodera.

**Supervision:** Vijay S. Pande, John D. Chodera, David D. L. Minh.

**Visualization:** Trung Hai Nguyen, Hexi Zhang.

**Writing – original draft:** Trung Hai Nguyen, Stefan G. Krimmer.

**Writing – review & editing:** Ariën S. Rustenburg, John D. Chodera, David D. L. Minh.

## References

1. Leavitt S, Freire E. Direct Measurement of Protein Binding Energetics by Isothermal Titration Calorimetry. *Curr Opin Struct Biol.* 2001; 11:560–566. [https://doi.org/10.1016/S0959-440X\(00\)00248-7](https://doi.org/10.1016/S0959-440X(00)00248-7) PMID: 11785756
2. Rajarathnam K, Rösgen J. Isothermal Titration Calorimetry of Membrane Proteins—Progress and Challenges. *Biochim Biophys Acta.* 2014; 1838:69–77. <https://doi.org/10.1016/j.bbamem.2013.05.023> PMID: 23747362



3. Feig AL. Applications of Isothermal Titration Calorimetry in RNA Biochemistry and Biophysics. *Biopolymers*. 2007; 87:293–301. <https://doi.org/10.1002/bip.20816> PMID: 17671974
4. Salim NN, Feig AL. Isothermal Titration Calorimetry of RNA. *Methods*. 2009; 47:198–205. <https://doi.org/10.1016/j.ymeth.2008.09.003> PMID: 18835447
5. Velazquez-Campoy A, Leavitt SA, Freire E. Characterization of Protein-Protein Interactions by Isothermal Titration Calorimetry. In: Fu H, editor. *Protein-Protein Interactions: Methods and Applications*. Totowa, NJ: Humana Press; 2004. p. 35–54.
6. Velazquez-Campoy A, Kiso Y, Freire E. The Binding Energetics of First- and Second-Generation HIV-1 Protease Inhibitors: Implications for Drug Design. *Arch Biochem Biophys*. 2001; 390:169–175. <https://doi.org/10.1006/abbi.2001.2333> PMID: 11396919
7. Brown A. Analysis of Cooperativity by Isothermal Titration Calorimetry. *Int J Mol Sci*. 2009; 10: 3457–3477. <https://doi.org/10.3390/ijms10083457> PMID: 20111687
8. Czodrowski P, Sotriffer Ca, Klebe G. Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments. *J Mol Biol*. 2007; 367:1347–1356. <https://doi.org/10.1016/j.jmb.2007.01.022> PMID: 17316681
9. Steuber H, Czodrowski P, Sotriffer Ca, Klebe G. Tracing Changes in Protonation: A Prerequisite to Factorize Thermodynamic Data of Inhibitor Binding to Aldose Reductase. *J Mol Biol*. 2007; 373:1305–1320. <https://doi.org/10.1016/j.jmb.2007.08.063> PMID: 17905306
10. Jin L, Amaya-Mazo X, Apel ME, Sankisa SS, Johnson E, Zbyszynska MA, et al. Ca<sup>2+</sup> and Mg<sup>2+</sup> Bind Tetracycline with Distinct Stoichiometries and Linked Deprotonation. *Biophys Chem*. 2007; 128: 185–196. <https://doi.org/10.1016/j.bpc.2007.04.005> PMID: 17540497
11. Egawa T, Tsuneshige A, Suematsu M, Yonetani T. Method for Determination of Association and Dissociation Rate Constants of Reversible Bimolecular Reactions by Isothermal Titration Calorimeters. *Anal Chem*. 2007; 79:2972–2978. <https://doi.org/10.1021/ac062183z> PMID: 17311466
12. Nilsson M, Valente AJM, Olofsson G, Söderman O, Bonini M. Thermodynamic and Kinetic Characterization of Host-Guest Association between Bolaform Surfactants and Alpha- and Beta-Cyclodextrins. *J Phys Chem B*. 2008; 112:11310–11316. <https://doi.org/10.1021/jp802963x> PMID: 18702539
13. Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, et al. KinITC: A New Method for Obtaining Joint Thermodynamic and Kinetic Data by Isothermal Titration Calorimetry. *J Am Chem Soc*. 2012; 134:559–565. <https://doi.org/10.1021/ja209057d> PMID: 22126339
14. Vander Meulen KA, Butcher SE. Characterization of the Kinetic and Thermodynamic Landscape of RNA Folding Using a Novel Application of Isothermal Titration Calorimetry. *Nucleic Acids Res*. 2012; 40:2140–2151. <https://doi.org/10.1093/nar/gkr894> PMID: 22058128
15. Di Trani JM, De Cesco S, O’Leary R, Plescia J, do Nascimento CJ, Moitessier N, et al. Rapid Measurement of Inhibitor Binding Kinetics by Isothermal Titration Calorimetry. *Nat Commun*. 2018; 9:893. <https://doi.org/10.1038/s41467-018-03263-3> PMID: 29497037
16. Thielges MC, Zimmermann J, Yu W, Oda M, Romesberg FE. Exploring the Energy Landscape of Antibody-Antigen Complexes: Protein Dynamics, Flexibility, and Molecular Recognition. *Biochemistry*. 2008; 47:7237–7247. PMID: 18549243
17. Cho S, Swaminathan CP, Bonsor DA, Kerzic MC, Guan R, Yang J, et al. Assessing Energetic Contributions to Binding from a Disordered Region in a Protein-Protein Interaction. *Biochemistry*. 2010; 49:9256–9268. <https://doi.org/10.1021/bi1008968> PMID: 20836565
18. Ladbury JE, Klebe G, Freire E. Adding Calorimetric Data to Decision Making in Lead Discovery: A Hot Tip. *Nat Rev Drug Discov*. 2010; 9:23–27. <https://doi.org/10.1038/nrd3054> PMID: 19960014
19. Henriksen NM, Fenley AT, Gilson MK. Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. *J Chem Theory Comput*. 2015; 11:4377–4394. <https://doi.org/10.1021/acs.jctc.5b00405> PMID: 26523125
20. Wiseman T, Williston S, Brandts JF, Lin LN. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Anal Biochem*. 1989; 179:131–137. [https://doi.org/10.1016/0003-2697\(89\)90213-3](https://doi.org/10.1016/0003-2697(89)90213-3) PMID: 2757186
21. MicroCal. *Data Analysis in Origin*. 1998.
22. MicroCal. *VP-ITC Users Manual*. 2001.
23. Myszkowski DG, Abdiche YN, Arisaka F, Byron O, Eisenstein E, Hensley P, et al. The ABRF-MIRG’02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction. *J Biomol Tech*. 2003; 14:247–269. PMID: 14715884
24. Tellinghuisen J, Chodera JD. Systematic Errors in Isothermal Titration Calorimetry: Concentrations and Baselines. *Anal Biochem*. 2011; 414:297–299. <https://doi.org/10.1016/j.ab.2011.03.024> PMID: 21443854

25. Boyce SE, Tellinghuisen J, Chodera JD. Avoiding Accuracy-Limiting Pitfalls in the Study of Protein-Ligand Interactions with Isothermal Titration Calorimetry. *bioRxiv*. 2015; 023796.
26. Mizoue LS, Tellinghuisen J. The Role of Backlash in the “First Injection Anomaly” in Isothermal Titration Calorimetry. *Anal Biochem*. 2004; 326:125–127. <https://doi.org/10.1016/j.ab.2003.10.048> PMID: 14769346
27. Chaloner K, Verdinelli I. Bayesian Experimental Design: A Review. *Stat Sci*. 1995; 10:273–304. <https://doi.org/10.1214/ss/1177009939>
28. Tellinghuisen J. A Study of Statistical Error in Isothermal Titration Calorimetry. *Anal Biochem*. 2003; 321:79–88. [https://doi.org/10.1016/S0003-2697\(03\)00406-8](https://doi.org/10.1016/S0003-2697(03)00406-8) PMID: 12963058
29. Tellinghuisen J. Can You Trust the Parametric Standard Errors in Nonlinear Least Squares? Yes, with Provisos. *Biochim Biophys Acta*. 2018; 1862:886–894. <https://doi.org/10.1016/j.bbagen.2017.12.016> PMID: 29289616
30. Zhao H, Piszczek G, Schuck P. SEDPHAT—A Platform for Global ITC Analysis and Global Multi-Method Analysis of Molecular Interactions. *Methods*. 2015; 76:137–148. <https://doi.org/10.1016/j.ymeth.2014.11.012> PMID: 25477226
31. Duvvuri H, Wheeler LC, Harms MJ. pytc: A Python Package for Analysis of Isothermal Titration Calorimetry Experiments. *bioRxiv*. 2017; 234682.
32. Minh DDL, Makowski L. Wide-Angle X-Ray Solution Scattering for Protein-Ligand Binding: Multivariate Curve Resolution with Bayesian Confidence Intervals. *Biophys J*. 2013; 104:873–83. <https://doi.org/10.1016/j.bpj.2012.12.019> PMID: 23442966
33. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P. High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Anal Chem*. 2012; 84:5066–5073. <https://doi.org/10.1021/ac3007522> PMID: 22530732
34. Krimmer SG, Betz M, Heine A, Klebe G. Methyl, Ethyl, Propyl, Butyl: Futile but Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors. *ChemMedChem*. 2014; 9:833–846. <https://doi.org/10.1002/cmcd.201400013> PMID: 24623396
35. Nasief NN, Hangauer D. Influence of Neighboring Groups on the Thermodynamics of Hydrophobic Binding: An Added Complex Facet to the Hydrophobic Effect. *J Med Chem*. 2014; 57:2315–2333. <https://doi.org/10.1021/jm401609a> PMID: 24479949
36. Tellinghuisen J. Calibration in Isothermal Titration Calorimetry: Heat and Cell Volume from Heat of Dilution of NaCl(aq). *Anal Biochem*. 2007; 360:47–55. <https://doi.org/10.1016/j.ab.2006.10.015> PMID: 17107650
37. Jeffreys H. An Invariant Form for the Prior Probability in Estimation Problems. *Proc Math Phys Eng Sci*. 1946; 186:453–461. <https://doi.org/10.1098/rspa.1946.0056>
38. Tellinghuisen J. Statistical Error in Isothermal Titration Calorimetry: Variance Function Estimation from Generalized Least Squares. *Anal Biochem*. 2005; 343:106–115. <https://doi.org/10.1016/j.ab.2005.04.026> PMID: 15936713
39. Liu JS. Monte Carlo Strategies in Scientific Computing. 2nd ed. New York: Springer; 2001.
40. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller A, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys*. 1953; 21:1087–1092. <https://doi.org/10.1063/1.1699114>
41. Minh DDL, Minh DL. Understanding the Hastings Algorithm. *Commun Stat Simul Comput*. 2015; 44:332–349. <https://doi.org/10.1080/03610918.2013.777455>
42. Patil A, Huard D, Fonnesbeck C. PyMC: Bayesian Stochastic Modelling in Python. *J Stat Softw*. 2010; 35:1–81. <https://doi.org/10.18637/jss.v035.i04> PMID: 21603108
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in {P}ython. *J Mach Learn Res*. 2011; 12:2825–2830.
44. Tellinghuisen J. Optimizing Experimental Parameters in Isothermal Titration Calorimetry. *J Phys Chem B*. 2005; 109:20027–20035. <https://doi.org/10.1021/jp053550y> PMID: 16853587
45. Chodera JD, Mobley DL. Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognitions and Design. *Annu Rev Biophys*. 2013; 42:121–142. <https://doi.org/10.1146/annurev-biophys-083012-130318> PMID: 23654303