Analytical Biochemistry 414 (2011) 297–299

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Notes & Tips

Systematic errors in isothermal titration calorimetry: Concentrations and baselines

Joel Tellinghuisen [a,*], John D. Chodera [b]

[a] Department of Chemistry, Vanderbilt University, Nashville, TN 37235, USA
[b] California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form 21 March 2011
Accepted 22 March 2011
Available online 26 March 2011

ABSTRACT

In the study of 1:1 binding by isothermal titration calorimetry, reagent concentration errors are fully absorbed in the data analysis, giving incorrect values for the key parameters—K, ΔH, and n—with no effect on the least-squares statistics. Reanalysis of results from an interlaboratory study of a selected biochemical process demonstrates that concentration errors are likely responsible for most of the overall statistical error in these parameters. The concentration errors are approximately 10%, greatly exceeding expected levels. Furthermore, examination of selected data sets reveals a surprising sensitivity to the baseline, suggesting a need for great care in treating dilution heats.

Isothermal titration calorimetry (ITC)[1] is widely used to study complexation and binding, with one of its strongest assets being the unique ability to yield estimates of all key thermodynamic quantities—$\Delta H°$, $\Delta G°$, and $\Delta S°$—from an experiment at a single temperature [1]. For 1:1 binding, M + X $\rightleftarrows$ MX, commercially available instruments can produce results with approximately 1% relative standard error for the three key parameters: the binding enthalpy ΔH, the equilibrium binding constant K, and the stoichiometry parameter n [2]. However, systematic errors can lead to inaccuracies that exceed the least-squares estimated precisions. Among these are the cell volume V_0 (~5%) [3], the heat calibration (~1%) [4], and other problems of varying significance (e.g., baseline errors, gas bubbles, PV work, sorption, corrosion) [5]. In an interlaboratory study of the enzyme–inhibitor process, bovine carbonic anhydrase II (CA II) + 4-carboxybenzenesulfonamide (CBS), Myszka and coworkers found overall relative standard deviations for the three parameters of approximately 20%, exceeding the individual precision estimates by a factor of 10 [6]. They suggested that concentration errors were the likely cause and presented spectrophotometric results that supported this claim. In the current work, we have conducted a new analysis of their results that strongly supports this interpretation, indicating concentration errors in both reagents on the order of 10% across the 14 laboratories where the experiments were done. Because such errors should be significantly less than this if care is used in solution preparation, we believe that this situation warrants attention.

In an ITC experiment, the reagent that is more limited with respect to some combination of availability, cost, solubility, and quantitative characterization is usually placed in the sample cell (titrand M), and the other reagent is loaded in the syringe (titrant X). In standard analysis algorithms, the parameter n is associated with the cell reagent, where it serves to correct the stated concentration $[M]_0$ and absorb errors in the cell volume V_0, both in accord with the assumed stoichiometry of the reaction [3]. This means that the accuracy of the estimates of K and ΔH is totally reliant on an accurate concentration $[X]_0$ for the titrant. If the data from an experiment are reanalyzed with the stated concentrations $[X]_0$ and $[M]_0$ scaled by factors f and g, respectively, K and ΔH change by the factor $1/f$, whereas n changes by the factor f/g—with identical least-squares fit statistics and residuals [7].

With this dependence in mind, we ask whether the results reported in Table 3 of Ref. [6] might yield single overall estimates of K and ΔH through application of titrant concentration scale factors f for the individual experiments. This could be done if all experiments yielded the same ratio of $K/\Delta H$, but a quick check shows that this is not the case. On the other hand, there is enough similarity in these ratios to suggest that this approach might explain much of the interlaboratory variability. So, instead we ask, how much can we improve the global precisions for K and ΔH by adjusting the stated $[X]_0$ for each experiment?

To address this issue, we first examined the displayed data in Fig. 4 of Ref. [6] and concluded that results 2, 8, and 14 exhibited pathologies severe enough to justify exclusion. For the other 11 data sets, we then used information in Tables 2 and 3 in Ref. [6] to generate synthetic enthalpograms as heat q_i for each injection

* Corresponding author. Fax: +1 615 343 1234.
E-mail address: joel.tellinghuisen@vanderbilt.edu (J. Tellinghuisen).

[1] *Abbreviations used:* ITC, isothermal titration calorimetry; CA II, carbonic anhydrase II; CBS, 4-carboxybenzenesulfonamide; WSQ, sum of weighted squared residuals.

0003-2697/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2011.03.024

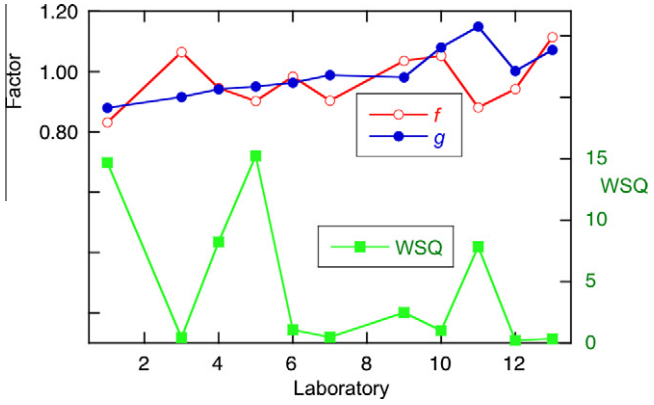

Fig.1. Concentration correction factors f and g and sum of weighted square residuals WSQ from the current global analysis displayed as functions of the participant number in Ref. [6]. Connecting lines are included as a visual aid and have no other significance.

i—a form analogous to that shown in the lower half of each frame in Fig. 4 of Ref. [6].[2] The data in the different sets were assigned weights based on the reported parameter standard errors (Table 3 of Ref. [6]).[3] Then all 11 data sets were refitted to a four-parameter model—K, ΔH, n, and f—with K and ΔH fixed globally but n and f adjusted for each data set. This computation was repeated with adjustment of K and ΔH until the cumulative sum of weighted squared residuals (WSQ) achieved a minimum. This minimum yielded the best estimates of K and ΔH and the 11 pairs of n and f values. Under the assumption of a single "true" $n = 0.94$, which was the average found in Ref. [6] and also from extensive calibration experiments on a MicroCal model VP-ITC [4], these n values were then converted to concentration scale factors g for the titrand.

This exercise yielded $K = 1.057 \times 10^6$ L/mol, $\Delta H = -11.17$ kcal/mol, and a cumulative WSQ of 50.8. The latter represents a sort of residual variance not explained by the model. (It would be zero if all data sets could be brought into exact agreement.) By comparison, when this computation was run with the same K and ΔH but with all f values frozen at unity, the cumulative WSQ was 2912, larger by a factor of 57. In short, the apparent discrepancies among the results in Table 3 of Ref. [6] are largely removed by application of concentration correction factors to the reagents.

Results for f, g, and WSQ are illustrated in Fig. 1. The averages for f and g are 0.97 and 0.99 with standard deviations 0.09 and 0.08, respectively. A slightly different perspective is obtained by applying these f and g values to the original K, ΔH, and n from Ref. [6], yielding the results shown in Fig. 2. From the caption, the adjustment massively decreases WSQ (here χ^2) for ΔH, but much less so for K. This is because the K values are determined with lower relative precision, so the adjustment is much less sensitive to them than to ΔH.

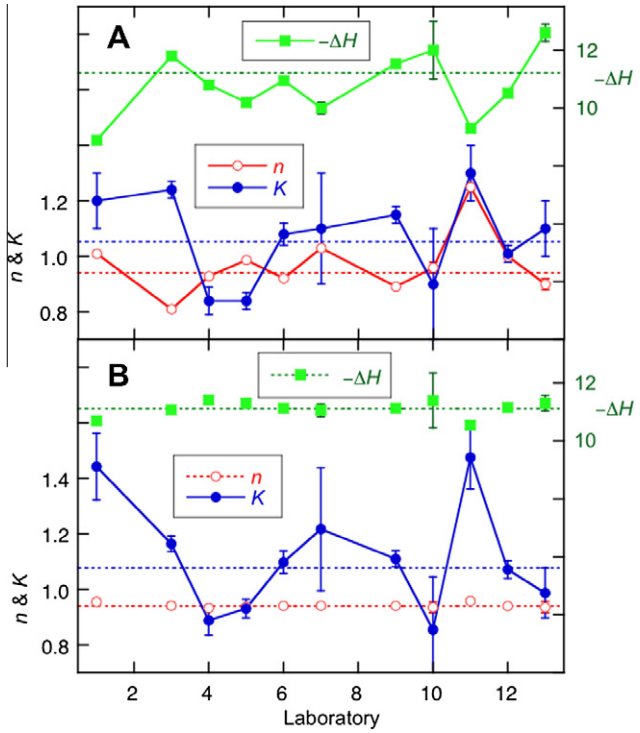

Fig.2. ITC parameters from Ref. [6] before (A) and after (B) adjustment with the concentration scale factors (units kcal/mol for ΔH and 10^6 L/mol for K). The sums of weighted squared residuals (χ^2) dropped from 2879 to 67 for ΔH and from 129 to 68 for K. From the adjusted values, the best estimates are −11.11(4) kcal/mol for ΔH and $1.08(4) \times 10^6$ L mol^{-1} for K. *Note*: 1.08(4) = 1.08 ± 0.04.

Myszka and coworkers identified concentration errors from the strong correlation between the derived ΔH values and the estimated molar absorptivities ε for the titrant (CBS) from the different laboratories at 272 nm (Fig. 6 in Ref. [6]). Assuming that these ε estimates were obtained by measuring absorbance A for a 1-cm cuvette, we have

$$\varepsilon_{\text{app}} = A/C_{\text{app}} = Af/C_{\text{true}} = \varepsilon_{\text{true}} f, \tag{1}$$

where the subscript "app" signifies apparent and C is concentration. Regression of our f values on ε_{app} supports this relationship: The intercept is statistically insignificant, and the fit to $b\varepsilon_{\text{app}}$ yields $\varepsilon_{272,\text{true}}$ (= $1/b$) = 1342(20) L mol^{-1} cm^{-1}.

If concentration errors accounted for all of the variability in K and ΔH, the χ^2 values given in Fig. 2B would be approximately 10, a factor of 7 smaller than that found. This means that there are additional sources of systematic error. The large WSQ values for data sets 1, 4, 5, and 11 in Fig. 1 prompted us to examine some of these in more detail. On fitting data for sets 1 and 11, obtained by digitizing the plots in Fig. 4 of Ref. [6], we needed to include a fourth adjustable parameter Δq for a baseline offset to reproduce the reported values. Furthermore, the results for ΔH and K were remarkably sensitive to this parameter. For set 1, the offset amounted to only 1.4% of the maximum heat in the first few injections, yet fixing it at zero changed ΔH by 2.5% and K by 11%. For set 11, the offset was more significant at approximately 11% of maximum; concomitantly, neglecting it altered ΔH by 20% and K by a whopping 50%.

In summary, researchers working with ITC must be mindful of systematic errors from poorly defined titrant concentrations and baselines. It should normally be possible to prepare titrant

[2] We were informed by the first author of Ref. [6] that the original data are not available; however, this lack is only an inconvenience for the current analysis.

[3] The exact data were first fitted to the standard three-parameter model assuming that $\sigma = 1$ for each q_i, yielding parameter standard errors from the a priori V matrix, which does not vanish for exactly fitting data [8]. Because the parameter standard errors scale with the data error σ, we could then adjust σ to match the reported parameter errors. For example, if $\sigma = 1$ yields $\sigma_n = 0.014$, whereas the reported value in Ref. [6] is 0.007, the corrected data error is $\sigma = 0.5$ and the weight is $1/(0.5)^2 = 4$. In fact, we used just the stated errors on n for this purpose and then checked whether the other two were reasonably close. The other assumption here is that all analyses behind the results in Ref. [6] employed unweighted least squares, which is the default choice for the standard algorithms. (Weights *are* needed here to treat the data collectively.)

solutions with concentration uncertainties below 1% by paying attention to the precision-limiting aspects of all operations.[4,5] If large concentration uncertainties are unavoidable, researchers should include these in their reported errors, which can be done by adding the percentage uncertainties in quadrature and taking the square root.[6] (For example, 3% titrant concentration uncertainty with 1% and 3% uncertainties in ΔH and K, respectively, will lead to overall uncertainties of 3.2% and 4.2%). Similarly, baselines must be treated with care. Dilution heats often are both significant and variable, invalidating the use of a compensating constant in the fit model. The variability is especially obvious when ionic titrants are injected into cells holding solution of different ionic strength [2,4]. Subtraction of the blank injection heats—obtained by running the experiment without M in the cell—is essential in such cases and can correctly account for most of the dilution effect [9]. In experiments with buffers, dilution heats should be minimal when the syringe and cell solutions are adjusted to common ionic strength. Cases that do not fit these patterns may require modification of the fit model beyond adding a constant [2,9]. At all times, dilution effects should be examined and their treatment should be stated clearly.

ITC experiments take time to set up and run. For most of the 14 experiments in Ref. [6], the run time alone was approximately 3 h, an amount that can become significant when many similar experiments are required. One of us has shown previously that, contrary to intuition, better precision is achieved with *fewer* injections rather than more, which clearly also improves throughput [7–10]. A protocol involving just 10 injections has been recommended for 1:1 processes, with adjustment of the titration range using

$$R_m = (6.4/c^{0.2}) + (13/c) \quad \text{or} \quad R_m = 1.1 \text{ minimally}, \tag{2}$$

where $c = K[\mathrm{M}]_0$ and R_m is the total X/M ratio after the last (mth) injection. Although the experiments in Ref. [6] were done before these results were known, most researchers today still follow standard manufacturers' recommendations—25–30 injections to $R_m \sim 2$. These procedures are often far from optimal and needlessly limit precision and productivity.

Acknowledgments

JDC was supported through a distinguished postdoctoral fellowship from the California Institute for Quantitative Biosciences (QB3) at the University of California, Berkeley.

References

[1] T. Wiseman, S. Williston, J.F. Brandts, L.-N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter, Anal. Biochem. 179 (1989) 131–137.
[2] L.S. Mizoue, J. Tellinghuisen, Calorimetric vs. van't Hoff binding enthalpies from isothermal titration calorimetry: Ba^{2+}-crown ether complexation, Biophys. Chem. 110 (2004) 15–24.
[3] J. Tellinghuisen, Volume errors in isothermal titration calorimetry, Anal. Biochem. 333 (2004) 405–406.
[4] J. Tellinghuisen, Calibration in isothermal titration calorimetry: Heat and cell volume from heat of dilution of NaCl(*aq*), Anal. Biochem. 360 (2007) 47–55.
[5] I. Wadsö, L. Wadsö, Systematic errors in isothermal micro- and nanocalorimetry, J. Therm. Anal. Cal. 82 (2005) 553–558.
[6] D.G. Myszka, Y.N. Abdiche, F. Arisaka, O. Byron, E. Eisenstein, P. Hensley, J.A. Thomson, C.R. Lombardo, F. Schwarz, W. Stafford, M.L. Doyle, The ABRF–MIRG '02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction, J. Biomol. Tech. 14 (2003) 247–269.
[7] J. Tellinghuisen, Optimizing experimental parameters in isothermal titration calorimetry, J. Phys. Chem. B 109 (2005) 20027–20035.
[8] J. Tellinghuisen, A study of statistical error in isothermal titration calorimetry, Anal. Biochem. 321 (2003) 79–88.
[9] J. Tellinghuisen, Optimizing experimental parameters in isothermal titration calorimetry: variable volume procedures, J. Phys. Chem. B 111 (2007) 11531–11537.
[10] J. Tellinghuisen, Isothermal titration calorimetry at very low c, Anal. Biochem. 373 (2008) 395–397.

[4] Replicate experiments run with a single set of stock solutions will not manifest concentration errors in their statistics, nor will experiments with individually prepared solutions reveal the bias that results from a repeated procedural error in their preparation.

[5] If for some reason the reagent that is preferred as titrant M is also easier to prepare to a known concentration, the analysis algorithm can be altered to incorporate the stoichiometry parameter n with the titrant concentration $[\mathrm{X}]_0$ instead of with $[\mathrm{M}]_0$. However, in that case the cell volume V_0 must be known reliably, whereas in the standard algorithms errors in V_0 are compensated by n [3].

[6] The least-squares parameter uncertainties and titrant concentration uncertainties are essentially independent, so their variances add to give this result.