ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Analytical Biochemistry 321 (2003) 79–88

ANALYTICAL BIOCHEMISTRY

www.elsevier.com/locate/yabio

A study of statistical error in isothermal titration calorimetry

Joel Tellinghuisen*

Department of Chemistry, Vanderbilt University, Nashville, TN 37235, USA

Received 25 March 2003

Abstract

In isothermal titration calorimetry (ITC), the two main sources of random (statistical) error are associated with the extraction of the heat q from the measured temperature changes and with the delivery of metered volumes of titrant. The former leads to uncertainty that is approximately constant and the latter to uncertainty that is proportional to q. The role of these errors in the analysis of ITC data by nonlinear least squares is examined for the case of 1:1 binding, $M + X \rightleftarrows MX$. The standard errors in the key parameters—the equilibrium constant K^o and the enthalpy ΔH^o—are assessed from the variance–covariance matrix computed for exactly fitting data. Monte Carlo calculations confirm that these "exact" estimates will normally suffice and show further that neglect of weights in the nonlinear fitting can result in significant loss of efficiency. The effects of the titrant volume error are strongly dependent on assumptions about the nature of this error: If it is random in the integral volume instead of the differential volume, correlated least-squares is required for proper analysis, and the parameter standard errors decrease with increasing number of titration steps rather than increase.

Keywords: ITC; Data analysis; Nonlinear least squares; Monte Carlo; Statistical errors

The method of isothermal titration calorimetry (ITC)[1] is now widely used to obtain thermodynamics information about biochemical binding processes. In a typical application of this technique, one of the reactants (M), typically at a concentration of 1 mM, is contained in a reaction vessel of small volume (0.2–2.0 mL), and the second reactant (the titrant X) is added stepwise to beyond the endpoint of the reaction. The temperature changes that occur after each injection of titrant are analyzed to obtain the heat q associated with the chemical changes from that injection, and the experiment thereby produces a titration curve of q vs extent of reaction. Analysis of such titration curves yields the enthalpy change ΔH^o and the equilibrium constant K^o for the reaction [1–3].

Despite the burgeoning use of the ITC technique,[2] there has been surprisingly little effort directed toward understanding the role of statistical errors in ITC data and their effect on the estimates of ΔH^o and K^o, which are obtained by nonlinear least-squares (LS) analysis of the data [4]. Such information can be used to optimize the parameters chosen for a particular experiment with respect to the precision of determination of either ΔH^o or K^o or to find some judicious compromise between these two precisions. It is also of relevance to a better resolution of a simmering controversy over the extent to which the ΔH^o values estimated from the temperature dependence of K^o (the van't Hoff ΔH^os) are consistent with those obtained directly from the analysis of the ITC titration curves [5–7]. The latter issue was in fact the impetus for the present work.

In what follows, I investigate the role of statistical errors in the analysis of ITC data using the "exact" variance–covariance matrix **V** for the nonlinear fit model. This "exact" **V** is obtained for exactly fitting data of known statistical error structure and in fact is not exact in its prediction of the parameter errors for nonlinear fit models. However, Monte Carlo (MC) calculations on typically 10^5 equivalent simulated titration curves confirm that analysis of such data is in accord with a previously derived "10% rule of thumb" [8]: If the

* Fax: 1-615-343-1234/322-4936.

E-mail address: joel.tellinghuisen@vanderbilt.edu.

[1] *Abbreviations used:* ITC, isothermal titration calorimetry; LS, least-squares; MC, Monte Carlo.

[2] Ref. [1] has been cited over 800 times, with about 1/4 of these coming in the last 2 years alone.

0003-2697/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0003-2697(03)00406-8

relative standard error for a nonlinear fit parameter is less than 10%, the parameter's distribution is close enough to Gaussian to permit its confidence limits to be estimated within 10% from the "exact" **V**. The fit model avoids the differential approximation used in some software, permitting one to estimate the standard errors reliably for a small number of titrant injections, down to the number of adjustable fit parameters. Application to the simple case of 1:1 complexation ($X + M \rightleftarrows MX$) shows that parameter precision can either increase or decrease with the number of titrant injections, depending on the assumptions about the nature of the statistical error. Further, if the error in the titrant volume is random in the accumulated (integral) volume rather than in the incremental (differential) volume, a correlated LS model is required for proper analysis. From these dependences, it is clear that the conditions for extracting ΔH^o and K^o with optimal precision depend strongly on the model assumed for the structure of the data error.

Methods

Fit model

Following the ith injection of titrant X, the heat q_i determined from the temperature excursion represents the result of changes in the amount of the complex MX in the active volume V_0 of the reactant vessel. The cells in the most widely used instruments are of the perfusion type, in which a volume v of solution is expelled each time the same volume of titrant is injected. It is assumed that prior to each injection the system is uniform and at equilibrium, and solution of this composition is expelled on injection, after which the injected titrant mixes and reacts to achieve the new equilibrium. After the ith injection the total concentrations (free and complexed) of X and M are given by [3]

$$[X]_{0,i} = [X]_0(1 - d^i) \quad \text{and} \quad [M]_{0,i} = [M]_0 d^i, \tag{1}$$

where $[X]_0$ is the concentration of titrant in the syringe, and $[M]_0$ is the starting concentration of M in the reaction vessel. The dilution factor $d = 1 - v/V_0$.

At equilibrium the concentrations of reactants and product satisfy the equilibrium expression,

$$\frac{[MX]_i}{([X]_{0,i} - [MX]_i)([M]_{0,i} - [MX]_i)} = K \equiv K^o \times (\mathrm{L\,mol^{-1}}). \tag{2}$$

The number of moles of complex produced by the ith injection is thus

$$\begin{aligned}\Delta n_i &= V_0[MX]_i - (V_0 - v)[MX]_{i-1} \\ &= V_0\{[MX]_i - d[MX]_{i-1}\},\end{aligned} \tag{3}$$

and the associated heat is

$$q_i = \Delta H^o\, \Delta n_i. \tag{4}$$

For simplicity, I ignore such experimental complications as the need to estimate heats of dilution for the titrant and the related concentration dependence of q_i (though the latter effect is expected to be negligibly small for many systems at the very dilute concentrations typically employed in such studies). I also use the dimensionless thermodynamic K^o instead of K below, which is equivalent to assuming all activity coefficients are unity at all times. Within the framework of these assumptions, this model is exact for any injection volume v, including, with modification of Eqs. (1) and (3), variable v within a given run. By contrast, the differential model described by Wiseman et al. [1] and used in some software is an approximation that may deteriorate for a small number of titrant injections and is not well suited to variable v.

The model just described has two adjustable parameters, ΔH^o and K^o, and as many data points as injections. The software in general use for analyzing ITC data includes a third parameter, the "site number" n. For "1:1" complexes this parameter is typically within $\sim$0.05 of 1.0 and should usually be viewed as a concentration correction factor, needed to put the concentrations of X and M on a common footing. Since inclusion of this factor is important for achieving a good fit of typical ITC data, I have also included it in the present model, where I have taken it as a correction factor to $[M]_0$. (The matter of how this factor should be applied is discussed further below.)

For most of the calculations discussed below, $[M]_0$ was fixed at 1.00 mM, V_0 was 1.4 mL (as in the instrument of [1]), and ΔH^o was 10.0 kcal/mol. The total titrant volume for m injections was typically 0.1 mL (though this was altered in some tests), and the precisions in K^o, n, and ΔH^o were investigated as functions of K^o (more properly $[M]_0K$), the number of injections m, and the stoichiometry range of the experiment, $R_m = [X]_{0,m}/[M]_{0,m}$.

Computational approach

The nonlinear least-squares and MC computations were carried out using techniques like those that have been employed previously to examine bias and non-Gaussian parameter distributions in nonlinear LS [8] and in linear LS analysis of data that are "biased" (non-Gaussian, as a result of transformation) [9]. For present purposes, the key results of these studies may be summarized as follows.

(A) When the error structure of the data is known a priori, the estimated variance for the jth nonlinear fit parameter β_j is the corresponding diagonal element V_{jj} of the **V** matrix, given by

$$\mathbf{V} = \mathbf{A}^{-1} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}. \tag{5}$$

In ordinary weighted fitting, the weight matrix **W** is diagonal, with dimensionality equal to the number of data points and elements that are the inverse data variances, $W_{ii} = \sigma_i^{-2}$. In the fitting of correlated data, **W** contains off-diagonal elements and is defined as $\mathbf{W} = \mathbf{V}_d^{-1}$, where $\mathbf{V}_d$ is the variance–covariance matrix of the data [10,11]. The elements of the matrix **X** are $X_{ij} = (\partial F_i/\partial \beta_j)$, where F expresses the fit function in terms of the fit variables (x and y) and the adjustable parameters **β**; the partial derivatives are evaluated at each value x_i of the independent variable using the final (iterated) values of the fit parameters.

(B) **V** as defined by Eq. (5) may be called the a priori variance–covariance matrix, since it presumes prior knowledge about the random error structure of the data; it is computed using exactly fitting data for the model in question.[3]

(C) Although parameters estimated from nonlinear LS are inherently non-Gaussian in their distribution and may not even have a formally defined variance, the Gaussian assumption will likely suffice within ~10% to estimate their dispersion, provided that the standard error estimated from the **V** matrix is less than 1/10 the value of the parameter.

Knowledge of the error structure of the data is key to estimating the parameter standard errors, as it is also for realistic MC calculations. There are two clear sources of random error in ITC: (1) the extraction of q_i values from the recorded temperature changes and (2) the delivery of the metered volume v of titrant from the syringe. The first of these is essentially a sensitivity of measurement limitation and is expected to be roughly constant, independent of q_i. The effects of the second depend strongly on the specific assumptions about the volume error. If the incremental volume v is assumed to possess random error, simple error propagation yields a proportional error, $\sigma_{qi} = q_i(\sigma_v/v)$. On the other hand, if it is the total accumulated volume after i steps that is assumed to possess random error, the incremental volume, being the difference between two such independent quantities, possesses correlated error. These three kinds of error affect the precisions differently and are examined individually and in combination in the calculations described below. When simultaneous contributions from the measurement and volume errors are considered, the variances are assumed to be additive, e.g., $\sigma_i^2 = \sigma_q^2 + q_i^2(\sigma_v/v)^2$ for the random volume error. Note that random volume error leads to data that are inherently heteroscedastic, requiring a weighted fit for proper analysis.[4] Similarly, correlated volume error requires the use of weighted, correlated LS.

[3] For linear LS fits and the usual assumptions of unbiased data with normally distributed random error, the LS parameters are unbiased and normal, with standard errors given exactly by the a priori **V** [8–10]. Thus such cases can be used to check Monte Carlo LS codes.

[4] The assumption that the uncertainty in q_i is independent of q_i has been used widely and is the basis for the customary neglect of weights in the analysis of ITC data; it merits closer scrutiny.

For reference, the two uncertainties are estimated as $\sigma_q = 0.28\,\mu\text{cal}$ and $\sigma_v = 0.015\,\mu\text{L}$ in [1]; for the benchmark reaction of 2′CMP with RNase studied there, the volume error dominates over most of the titration curve in the random volume error model (see below).

A useful point to bear in mind in assessing least-squares parameter errors is that these scale with the error in the data. For example, if the computations for a given model are repeated after simply scaling σ_q and σ_v by a factor f, the parameter standard errors will scale by the same factor f. Since sufficiently small data errors yield adequately Gaussian parameter distributions, this means that the error structure for the model can always be evaluated using Eq. (5). The only question then is the extent to which this structure applies to the actual situation; in previously examined cases, the 10% rule of thumb has proved a reliable guideline for applicability.

The computational codes were written by me in FORTRAN and employed methods similar to those that are widely available [12,13]. In the MC computations the random deviates were taken from the random number generator of the computer (a DEC AlphaServer 2100A 4/275). These were converted to Gaussian using the Box–Muller algorithm [13]. In the fitting code the independent variable was taken as the titration index i, which is rigorously error free. The errors in q and v were taken to be normal at all times, and the solution concentrations were treated as exact. (Although uncertainty in the concentrations is significant in most actual experiments, this uncertainty is not manifested as point-to-point random error in a given experiment, and it is partly compensated through the parameter n, as was already noted and is discussed further below.)

Results and discussion

Check on the 10% rule of thumb

I address first the circumstances under which Eq. (5) can be trusted to yield reliable estimates of the parameter errors in the analysis of ITC curves. For reference, Fig. 1 illustrates typical titration curves spanning the approximate extremes of analyzable values of the product $K[\text{M}]_0$. These curves represent an extension of the example, $K[\text{M}]_0 = 1000$, explored in Fig. 3A of [4]. Since the latter MC analysis also included an assumed 2% error in the concentrations, the present results cannot be compared quantitatively; however, the results for this case are commensurate with those obtained there. It is interesting that even though K° is uncertain by 20% for $K[\text{M}]_0 = 1000$, ΔH° and n are actually quite well defined in this case. On the other hand, all three parameters are much less precise at the other extreme.

Fig. 2 illustrates the results of MC calculations for the case $K^\circ = 5 \times 10^6$ in Fig. 1. The results for K° are clearly

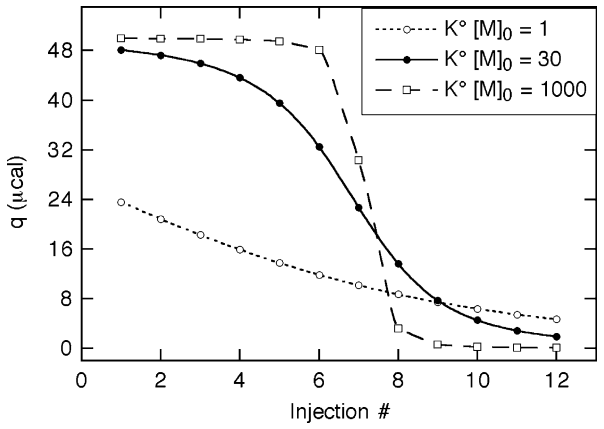

Fig. 1. Computed ITC curves for $K^o = 5 \times 10^3$, 1.5×10^5, and 5×10^6 and $[M]_0 = 0.200$ mM. Other conditions: $V_0 = 0.20$ mL, $R_m = 2.08$, $v = 10$ μL, $\Delta H^o = 10$ kcal/mol. Neglecting the error in v and taking $\sigma_q = 0.6$ μcal, as in [4], the calculated (Eq. (5)) standard errors in K^o are 1.845×10^3, 1.107×10^4, and 1.008×10^6, respectively, while the corresponding errors in ΔH^o are 2.486, 0.0932, and 0.0533 kcal/mol. The standard errors in the concentration correction parameter n (taken to be 1.00) are 0.105, 0.0074, and 0.0030 (same order).

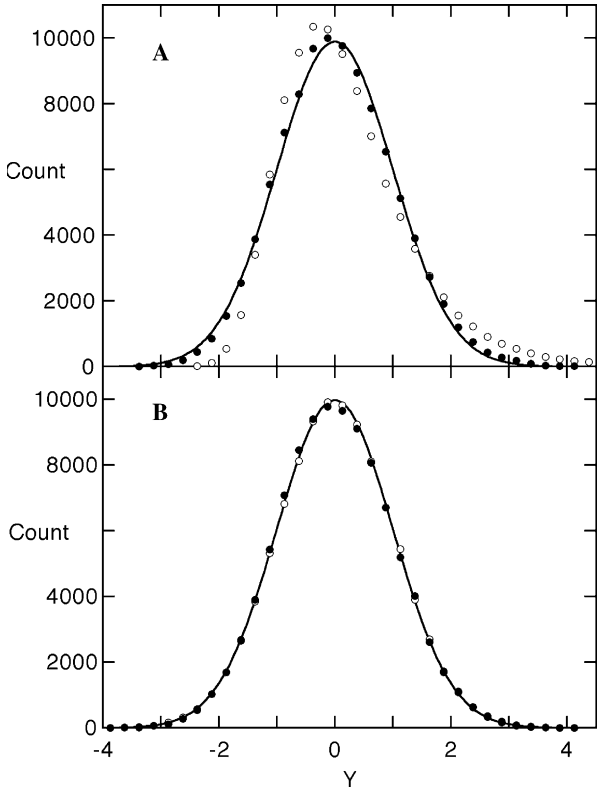

Fig. 2. Histogrammed results for (A) K^o (○) and K^{o-1} (●) and (B) ΔH^o (○) and n (●), from 10^5 MC fits of 12-point data sets like that shown for $K^o = 5 \times 10^6$ in Fig. 1, with superimposed random error having $\sigma_q = 0.6$ μcal. In each case the histogrammed quantity is $Y = (\beta - \beta_{true})/\sigma_\beta$, with the values for these quantities given in the caption to Fig. 1. The smooth curves are standard Gaussians, scaled into optimal agreement with the data for K^{o-1} in A and with ΔH^o in B.

non-Gaussian, with a bias of 3.4% (even though the peak in the distribution is shifted negatively). However, the MC statistical error in K^o exceeds the predicted value by less than 6%, showing that even for this case of 20% relative error, the predicted value from Eq. (5) would be adequate for many applications. The histograms for ΔH^o and n appear to be Gaussian, as expected from their smaller-percentage standard errors (0.5 and 0.3%, respectively). However, only the former actually fits the Gaussian with a sufficiently small χ^2—21.6 for 33 points.[5] It is interesting that the reciprocal of K^o (or the dissociation constant) is actually much closer to Gaussian than K^o itself; similar results were obtained in the large-K^o regime in the study of complexation equilibria in [8].

Monte Carlo calculations have been carried out for a number of other choices of the ITC parameters, for both constant and proportional error, random and correlated. None of the results has shown any problem with the 10% rule for the analysis of typical ITC data. The validity of Eq. (5) extends even to the low limit of three titration increments ($m = 3$), where there are no degrees of freedom and the LS equations yield exact fits at all times. Of course, Eq. (5) says nothing about the extent of nonnormality or the bias, so if these are at issue, MC calculations must be employed for the specific cases in question.

Dependence on stoichiometry range and titration steps

I turn now to the dependence of the parameter standard errors on the other experimental quantities, as predicted from the **V** matrix via Eq. (5), using exactly fitting data. Fig. 3 illustrates the computed standard errors in K^o and ΔH^o as functions of the range of titration and the number of titration increments, for the mid-range $K[M]_0$ value of 36 and constant error. Other parameters were chosen to resemble those for the instrument described in [1] (except the error, which was artificially large). The most interesting result of these calculations is the observation that better precision is generally achieved with fewer points and a larger titration range than is customarily employed in such work. For small m, the standard errors exhibit structure, showing that they are sensitive to just where the points fall on the titration curve.

The loss of precision with increasing number of points seems at odds with expectations, but it can be understood as follows. The total heat q is limited by the fixed amount of M in the reaction vessel, so increasing m decreases each q_i. If K^o and n are held constant, one obtains m estimates of ΔH^o from $\Delta H^o = q_i/\Delta n_i$ (Eq. (4)). The relative error in each such estimate of ΔH^o is approximately σ_q/q_i, and since q_i decreases with increasing m, the error in ΔH^o

[5] The average value of χ^2 is the number of degrees of freedom ν (the number of data points minus the number of adjustable parameters), so a proper fit should approximate this, consistent with its variance of 2ν.

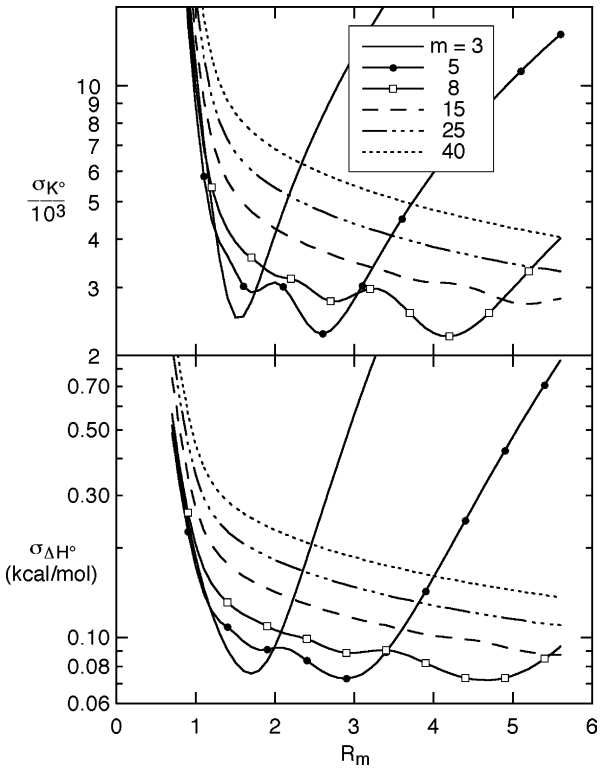

Fig. 3. Calculated standard errors in K° and ΔH° for $K^\circ = 36,000$ and $[M]_0 = 1.0$ mM, as functions of the stoichiometry range and the number of titration steps m, for constant absolute error $\sigma_q = 0.04$ mcal. Other parameters: $V_0 = 1.4$ mL, $\Delta H^\circ = 10$ kcal/mol, and $vm = 0.10$ mL. For reference, the total q for complete reaction is 14 mcal, so for $m = 15$ and $R_m = 5$, the error on q_1 is ~1%.

increases concomitantly. This is partially offset by the statistical averaging, which goes as $m^{-1/2}$. The net result is a standard error that increases roughly as $m^{1/2}$. Similar observations were made some time ago by Doyle et al. [14] in a study of a differential absorption technique for characterizing binding isotherms. Fig. 4 shows that the observed power is slightly higher than 1/2 for the cases examined in Fig. 3.

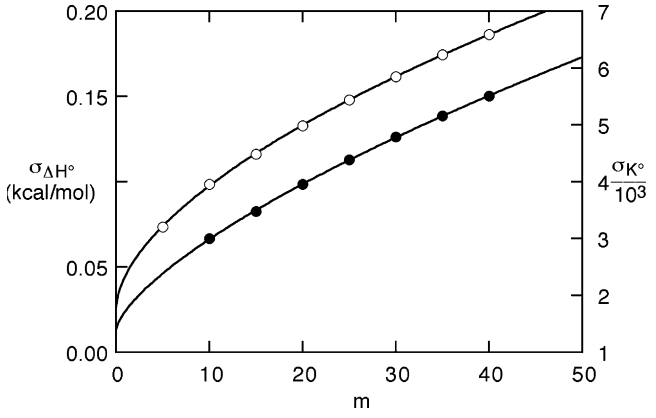

Fig. 4. Calculated standard errors in ΔH° (○) and K° (● scale to right) at $R_m = 3$, as functions of the number of titration increments m. Other conditions as in Fig. 3. The smooth curves represent fits to $a + bm^p$ and yield $p = 0.58$ for ΔH° and 0.68 for K°.

The problem of correlated volume error will be considered in detail below. For random error in the incremental titrant volume v, there are several cases to consider: (1) the concentration $[X]_0$ and v are fixed, and the operator decides upon m, and hence R_m; (2) $[X]_0$ and the stoichiometry range R_m are set in advance, and the operator chooses m; and (3) R_m and v are set, and the operator varies m. In case (1), an increase in m means simply adding additional points at the end of the titration range, which must decrease all of the parameter standard errors. In case (2) the total titrant volume is fixed, so increasing m decreases v; this leads to an increase in the relative uncertainty σ_v/v and hence an m-dependence similar to that shown for constant error σ_q in Fig. 3. Case (3) is the least feasible from an experimental standpoint, since it means altering the titrant concentration when either m is changed for fixed R_m or R_m is changed for fixed m. However, this case does lead to increasing precision with increasing m, since the relative error $\Delta v/v$ is now fixed. It should also be borne in mind that with increasing dilution of titrant, all q_i will decrease as m^{-1}, so that eventually the constant error σ_q in q will dominate and lead again to increasing standard errors with increasing m.

Fig. 5 illustrates the dependence of the relative errors in K° and ΔH° on K° and R_m for $m = 7$ and constant

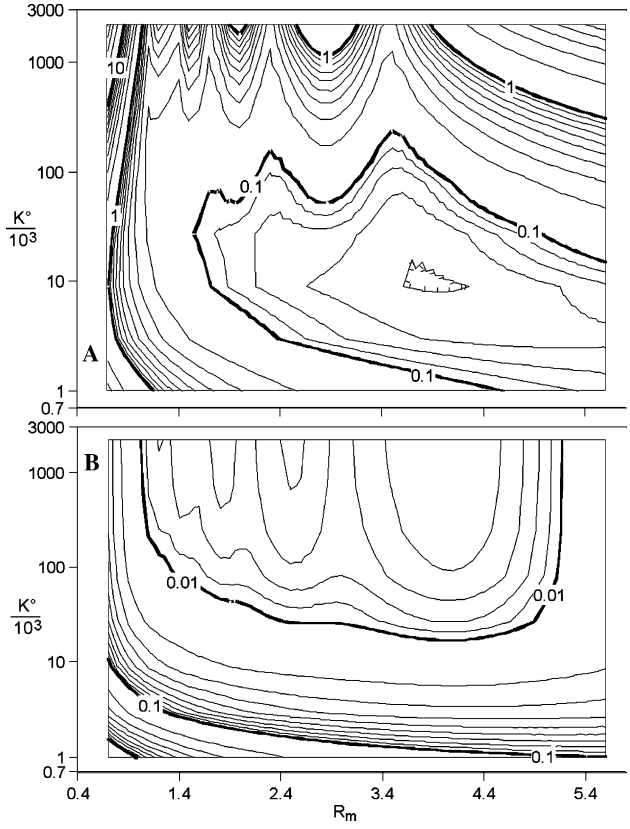

Fig. 5. Contour plots of the relative standard errors σ_β/β as functions of K° and R_m, for $m = 7$ and constant error, $\sigma_q = 0.04$ mcal: A, for K°; B, for ΔH° Other parameters are as in Fig. 3.

absolute error. There is a fairly flat minimum in the error surface for K^o centered near $K[M]_0 = 10$ and $R_m = 4$. ΔH^o is an order of magnitude more precise, with a large region at large K^o where the relative error is less than 1%. For both quantities there is considerable structure in the error surface for the relatively small m value of 7. Note again that the structure of these contour diagrams is unaffected by the actual value used for σ_q, so, for example, reducing σ_q by a factor of 10 would simply result in a relabeling of the contours by the same factor. Also, in regions where the relative errors in K^o and ΔH^o exceed ~0.2, the actual statistical distributions may be far from Gaussian, as already noted in the previous section.

Neglecting weights

I now examine more closely the prototype case from [1] (Figs. 4A and 5) and ask what happens when such data are analyzed by unweighted least-squares. In general, when heteroscedastic data are fitted with neglect of weights, there are two consequences: (1) the parameter distributions are broader, because minimum-variance estimates are obtained only when the weights are taken as the inverse variances, $W_{ii} = \sigma_i^{-2}$, and (2) the error estimates from the **V** matrix are not reliable and can be either pessimistic or optimistic [15]. The magnitude of these effects can be determined only through MC computations, in which it is also necessary to replace the definition of **V** in Eq. (5) by the usual a posteriori expression for analysis of data of unknown precision [8,15],

$$\mathbf{V} = \frac{S}{v}\mathbf{A}^{-1}, \tag{6}$$

where S is the sum of squared residuals from the fit, and v is the number of degrees of freedom.

For the aforementioned experiment in [1], the following parameter values apply: $v = 4\,\mu L$, $\Delta H^o = -13.7$ kcal/mol, $[M]_0 = 0.651$ mM, $K^o = 4.88 \times 10^4$, $R_m = 2.06$. As was noted earlier, σ_v and σ_q were estimated to be 0.015 μL and 0.28 μcal, respectively. The calculated q_i values range from 1.18 mcal for $i = 1$ to 0.04 mcal for $i = 20$. Since the relative error in v is 0.0038, the relative error will exceed the absolute error (σ_q) until $i = 17$, where $q_i = 0.07$ mcal. With data properly weighted for the combined relative and absolute errors, Eq. (5) yields $\sigma_{K^o} = 200$ and $\sigma_{\Delta H^o} = 21.0$ cal/mol.

For comparison, Monte Carlo calculations employing unweighted fits yield the results illustrated in Fig. 6. The neglect of weights has led to increases in both σ values, as anticipated. The corresponding loss of efficiency for the determination of K^o is $2.334^2 \approx 5$, meaning one would need to run five equivalent experiments with unweighted analysis to match the precision achievable through proper weighting. On the other hand, the biases

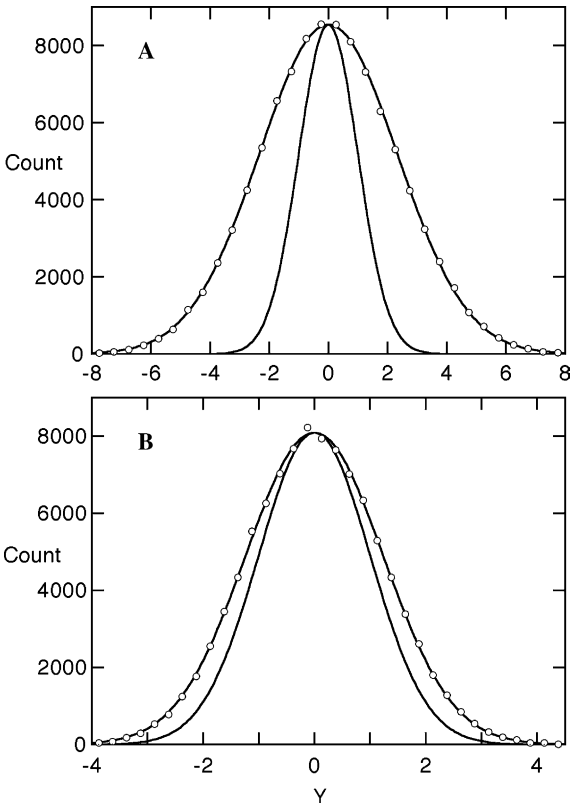

Fig. 6. Histogrammed results from 10^5 unweighted MC fits of 20-point data sets resembling that illustrated in Figs. 4A and 5 of [1], for K^o (A) and ΔH^o (B). In each case the histogrammed quantity is $Y = (\beta - \beta_{true})/\sigma_\beta$, using σ_β values as given in the text for a properly weighted fit. The inner curve in each case is the normal curve ($\sigma = 1$) scaled to the peak of the data, while the other curves are fitted Gaussians having $\sigma = 2.334$ (A) and 1.232 (B).

in both K^o and ΔH^o are not statistically significant. Interestingly, the rms estimates obtained from the MC results using Eq. (6) are surprisingly close to correct for σ_{K^o} (500 vs 472 observed) and only a bit worse for $\sigma_{\Delta H^o}$ (17.8 vs 25.7 cal/mol). These results also illustrate how Eq. (6), when applied naively to heteroscedastic data, can both over- and undershoot the actual values for the fit model in use.

Even the less efficient unweighted fits have yielded parameter standard errors that are smaller than those suggested from the repeated runs in [1]. As a check on this result, I have fitted the data depicted in Fig. 5 of [1] using the current fit model, obtaining apparent errors in K^o from 900 to 1400, and in ΔH^o from 30 to 120 cal/mol, for weighted and unweighted fits, with and without inclusion of an additional parameter for a constant background. In the weighted fits, the value of S/v in Eq. (6) should approximate the reduced χ^2, for which the expected value is unity. Observed values were 35 without and 16 with a background parameter. These and similar results obtained from analysis of ITC data for the binding of $BaCl_2$ by 18-crown-6 ether [16] imply that either the experimental random errors in v and q are

much greater than those indicated in [1] or other effects are limiting the precision of the determinations. The weighted fits were particularly sensitive to inclusion of the background parameter, suggesting that the manner of correcting for heat of dilution of titrant could be an underappreciated source of error in the determinations.

Correlated volume error

The motorized syringes used to inject the titrant in ITC apparatuses are programmed to travel a certain distance with each injection. It may be more appropriate to consider the end points of travel of the plunger as the quantities subject to random error, rather than the incremental volume v. Then the latter, being the difference between two independent quantities, possesses correlated error. The effects of this change in assumption about the error can be seen dramatically in Monte Carlo computations: With increasing m, the case (2) model above (fixed $[X]_0$ and R_m) now gives *decreasing* rather than increasing parameter error. This effect, too, was observed previously by Doyle et al. [14] in fitting differences extracted from an absorbance titration curve.

To treat this case, let us represent the total delivered titrant volume after i injections as u_i. Then the ith incremental volume is $v_i = u_i - u_{i-1}$. The v_i and u_i are related via a linear transformation,

$$\mathbf{v} = \begin{pmatrix} 1 & 0 & 0 & & 0 & 0 \\ -1 & 1 & 0 & & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ & & \dots & & & \\ 0 & 0 & 0 & & -1 & 1 \end{pmatrix} \mathbf{u} \equiv \mathbf{L}\mathbf{u}, \qquad (7)$$

and thus $\mathbf{V_v}$ and $\mathbf{V_u}$ are related by [10,17]

$$\mathbf{V_v} = \mathbf{L}\mathbf{V_u}\mathbf{L}^{\mathrm{T}}. \qquad (8)$$

Because the u_i are independent, $\mathbf{V_u}$ is diagonal, with elements σ_u^2 (constant by assumption). Since it is q_i that is fitted here, these σ_u^2 values must be converted to σ_{qi}^2 by error propagation, using numerical differentiation to assess $\mathrm{d}q_i/\mathrm{d}u_i$. Calling the resulting matrix $\mathbf{V'_u}$, $\mathbf{V'_v} = \mathbf{L}\mathbf{V'_u}\mathbf{L}^{\mathrm{T}}$ and $\mathbf{W} = \mathbf{V'^{-1}_v}$. It is noteworthy that $\mathbf{V'_v}$ is tridiagonal, with elements $(i,i) = (\sigma_{qi-1}^2 + \sigma_{qi}^2)$, $(i,i+1) = -\sigma_{qi}^2$, and $(i,i-1) = -\sigma_{qi-1}^2$ (with all indices limited to the range $1 \leqslant i \leqslant m$ of the data). The diagonal terms are thus seen to be the expected results for subtraction of two random quantities.

Use of this $\mathbf{W}$ with the model described and used above yielded $\mathbf{V}$-based parameter error estimates in good agreement with the results from the MC computations and well within the framework of the 10% rule of thumb. For example, in computations for $R_m = 3$ on the model described in Fig. 3 but having $\sigma_q = 0$ and $\sigma_v = 0.0015$ mL, the choice $m = 5$ yielded 13% relative standard error in K° and 6.4% in ΔH° as compared with estimates higher by 5.0 and 4.6%, respectively, from MC computations on 10^5 data sets. However, the reduced χ^2 from the MC computations was 1.034, which deviates more from the expected value of 1.00 than was found in the earlier MC results.

The volume error was set artificially high for the purpose of the MC test computations. Reducing it to 0.0001 mL makes the error for $m = 10$ about 1%, which is comparable to that for early points in the constant-error model used to produce the results in Fig. 3. Repetition of those calculations for the present correlated-v model yields the results shown in Fig. 7. As already noted, now the errors in both parameters decrease with increasing m. It is also noteworthy that the scales of both errors are about an order of magnitude smaller than those in Fig. 3, showing that a comparable nominal error for each point (1%) leads to much smaller actual errors in the parameters, due to the weighting and the compensating effects of the correlation. Interestingly, $R_m < 2$ yields optimal precision for ΔH° but the K° precision continues to improve with increasing R_m for all but $m = 3$.

Fig. 8 shows what happens when such data are analyzed with neglect of weights and correlation. The ordinary LS analysis correctly tracks the m-dependence predicted by the correlated model **V,** but gives standard errors too large by about a factor of 10 for K° and 2 for ΔH° On the other hand, the a posteriori $\mathbf{V}$ (Eq. (6)) yields estimates that are roughly correct for this

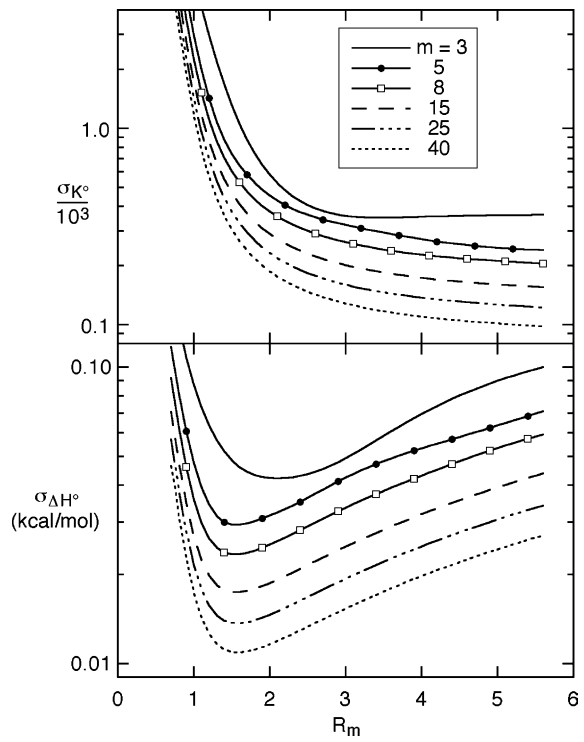

Fig. 7. Calculated standard errors in K° and ΔH° for the correlated volume model, with σ_v for the integral titrant volume u_i taken to be 0.0001 mL and $\sigma_q = 0$. Other parameters are as in Fig. 3.

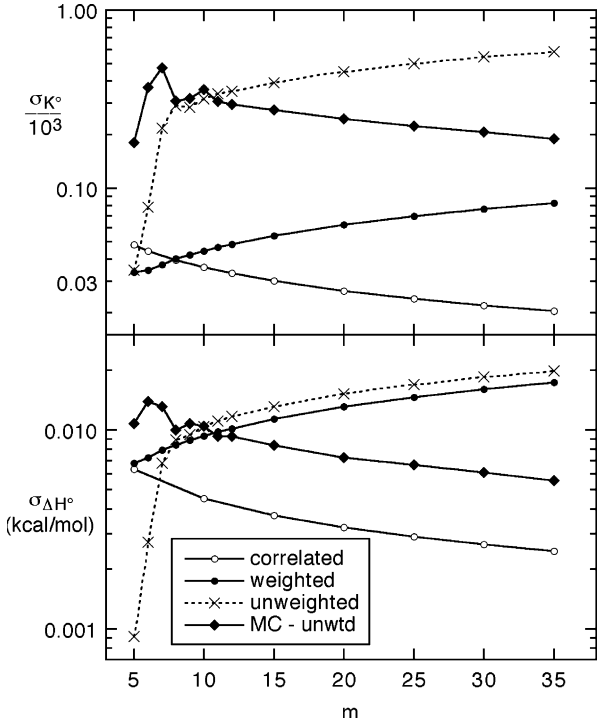

Fig. 8. Statistical errors in K° and ΔH° as functions of m, for model of Fig. 3, but with $\sigma_q = 0$ and random error $\sigma_v = 0.015\,\mu L$ (as in [1]) in the integral titrant volume u_i ("correlated"). The results marked "weighted" are obtained when this same random error is in v. The "MC" results are from the statistics of 10^4 data sets analyzed using unweighted LS, while those marked "unweighted" represent the apparent values (rms) returned by the a posteriori **V** of Eq. (6), from the same MC computations. The structure at small m in the "unweighted" and "MC" results is real.

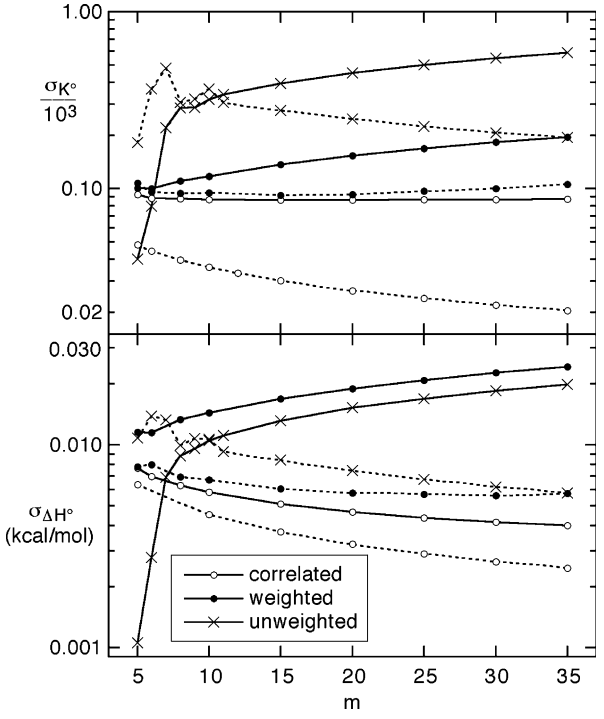

Fig. 9. Results for same model as in Fig. 8, but with addition of random measurement error $\sigma_q = 0.28\,\mu cal$ (as in [1]). The dashed "correlated" curves are from Fig. 8 ($\sigma_q = 0$). The other results are obtained by analyzing the same MC data sets with neglect of correlation, using weighted and unweighted LS. In both cases, the solid curves represent the apparent standard errors, from the a posteriori **V**, while the dashed curves represent the actual MC statistics from 10^4 data sets for each point.

unweighted model only for the range $m \approx 8$–12, deviating sharply to overly optimistic errors for smaller m and pessimistic for larger.

Since the error in measuring q_i is presumed to be random, and since this error leads to decreasing precision with increasing m (Fig. 3), we might expect that addition of random error in q_i to this correlated v model should neutralize the increased precision at large m in Fig. 8. That is in fact observed. To generate the data needed for the MC computations, the effects of the volume error are computed first, as before: The u_i are given random error, v_i is calculated from $v_i = u_i - u_{i-1}$, and the heat q_i is calculated using a variable-v version of Eqs. (1) and (3). Then random error is added to each q_i value. This error is correctly accommodated in the correlated fit model by adding σ_q^2 to the diagonal elements of $\mathbf{V}'_{\mathbf{v}}$, as was confirmed in the MC computations. Fig. 9 shows that the addition of the random measurement error has rendered K° less precise by a factor of $\sim$3 and made σ_{K° almost independent of m. A smaller reduction of precision occurs for ΔH° and its standard error still decreases with increasing m.

Fig. 9 also includes results of analyzing the same data with neglect of the correlation, using either unweighted LS or the weighted model that is correct for random error in both v and q_i. The MC statistics show that the weighted model does a good job of extracting the K° values but performs less well for ΔH° However, without repeating their experiments enough times to obtain ensemble statistics, users of the weighted model would normally rely upon the a posteriori **V** for error estimates. Accordingly, they would report errors for ΔH° that are significantly too large over the full range of m in Fig. 9 and somewhat less pessimistic errors for K° for most m. Not surprisingly, unweighted LS performs worse almost across the board; the one exception is in the assessment of ΔH° where the unweighted model actually betters the weighted model for $m > 35$.

Conclusion

An examination of the statistical error in the nonlinear least-squares analysis of ITC data for 1:1 complexation reactions has yielded the following results. (1) When the data uncertainty is dominated by a constant measurement uncertainty in the heat q, better precision is achieved in both K° and ΔH° for fewer titrant injections than are customarily used; as few as five may be

optimal in many cases. (2) Under usual experimental circumstances, the same result holds for data uncertainty dominated by the relative error in the differential titrant volume v, when this is assumed to be random. (3) On the other hand, for random error in the integral titrant volume, the error in v is correlated, and the parameter precisions increase with increasing titration steps. (4) For constant error, the relative precision for K° is an order of magnitude poorer than that for ΔH°; when the volume error dominates, the precision gap is smaller. (5) Actual ITC data are typically dominated by the relative error in v for large q (early titrant injections i) and by the constant absolute uncertainty in q for small q (large i); for optimal extraction of K° and ΔH° such data require analysis by weighted LS or, in the case of the correlated volume error, by weighted correlated LS. (6) Higher precision in K° is generally favored by larger stoichiometry ranges than are customarily used; the same holds for ΔH° in the constant error model, but $R_m < 2$ is optimal for ΔH° when the volume error dominates.

Probably the most important result of this study is the observation that the parameter precisions depend so strongly on assumptions about the random error in the data, especially the titrant volume error. Clearly more experimental work is needed to clarify the extent to which the two limiting models—random in v_i vs random in u_i—apply to real ITC data. It is also desirable to determine just how constant σ_q is for measurement of q_i values ranging over typically two orders of magnitude. The latter question is of particular interest for binding characterized by small ΔH° and limited to small $[M]_0$, where the measurement error may exceed the volume error over most of the titration curve.

The present study has been confined to just 1:1 binding. Even in situations where the random errors in q and v dominate, making smaller m preferable, it may still be necessary to use $m = 10$ or more to demonstrate that the stoichiometry actually is 1:1. Also, in practice it is wise to work with at least a few degrees of statistical freedom, to obtain some indication of the goodness of fit. Other cases, such as multiple binding [18], will have to be examined specifically; however, it is likely that there, too, better precision can be achieved using fewer injections than are commonly employed (30–60 in [18]).

The large reduced χ^2 values (>10) obtained from weighted analysis of typical experimental data suggest that factors other than the stated uncertainties in q and v may be limiting the precisions. Very precise data place severe demands on nonlinear fit models [15], so subtle limitations in current fit models may be at least partly responsible. In this connection it is worth noting that weighted analysis of similar data having correlated v error leads to only a factor of 2 increase in χ^2 when the correlation is neglected, so such neglect cannot fully explain the observed large χ^2 values.

With regard to limitations in fit models, there is the matter of the "site number" parameter n. As was noted earlier, in simple 1:1 cases this quantity is typically within 5% of unity and is mainly correcting for concentration errors. It is usually defined in the stoichiometry sense MX_n, which means that it is serving as a correction factor for M (i.e., $n = 1.05$ means that the true $[M]_0$ is 5% larger than stated). Since M is often a macromolecule and is typically harder to prepare to known concentrations, this is probably appropriate in most cases. However, for cases where the titrant concentration is less certain, it is proper to redefine the correction. The effect of such a redefinition is a correction factor of $1/n$ to $[X]_0$, and it results in a drop in both K° and ΔH° by the factor n. In cases where the site parameter is covering for a mix of macromolecules (e.g., 2.5% having two sites to yield $n = 1.05$), the 1:1 fit model is not really correct. Either way, the deviation of n from the "chemical" stoichiometry for the process under investigation is an indication of systematic error, and a conservative assessment of the parameter errors in such cases should include its consideration.

It is unfortunately a common practice to analyze inherently heteroscedastic data by unweighted LS, using Eq. (6) to estimate the parameter errors. It should be borne in mind that Eq. (6) always "lies" in such cases, and the extent of the lie can be determined only through Monte Carlo calculations. In the modeling of the key test case from [1] in the present study, the use of unweighted LS resulted in a 5-fold loss in efficiency. However, if correlated volume error is assumed to dominate over the entire titration curve, as it may for today's much more sensitive instruments, the loss of efficiency in estimating K° rises to a factor of 100 (10-fold increase in σ_{K°, Fig. 8)! On the other hand, the use of unweighted LS has not shown any significant biases in either K° or ΔH°, so the neglect of weights and correlation has probably not been a major factor in the reported inconsistencies between directly measured and van't Hoff estimates of ΔH°.

Throughout this work I have assumed that equal volume aliquots of titrant are added sequentially to generate the titration curve. This appears to be the standard mode used by workers in the field. However, in principle there is nothing wrong with using variable volumes, and from the structure evident in Figs. 3 and 7, it seems likely that for small m and a chosen R_m, some sequence of volumes that vary from step to step can yield smaller parameter errors than the default constant-v approach. This possibility is currently being investigated, with preliminary results indicating that σ_{K° can be lowered as much as 40% using variable instead of fixed v_i for $m < 10$.

Finally, it is worth noting that most of the parameter error calculations described here are not hard to carry out. Although I employed a programming language for

the Monte Carlo calculations and to expedite the examination of the dependence of the errors on m, R_m, and K^o, results for specific random data structures can be obtained easily using some simple and inexpensive desktop data presentation and analysis packages. For example, in the course of this work I have used the KaleidaGraph program [19] to double-check some of the results obtained from the FORTRAN programs.[6]

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] M. El Harrous, S.J. Gill, A. Parody-Morreale, Description of a new Gill titration calorimeter for the study of biochemical reactions. I: assembly and basic response of the instrument, Meas. Sci. Technol. 5 (1994) 1065–1070.
[3] M. El Harrous, O.L. Mayorga, A. Parody-Morreale, Description of a new Gill titration calorimeter for the study of biochemical reactions. II: operational characterization of the instrument, Meas. Sci. Technol. 5 (1994) 1071–1077.
[4] M. El Harrous, A. Parody-Morreale, Measurement of biochemical affinities with a gill titration calorimeter, Anal. Biochem. 254 (1997) 96–108.
[5] Y. Liu, J.J. Sturtevant, Significant discrepancies between van't Hoff and calorimetric enthalpies III, Biophys. Chem. 64 (1997) 121–126.
[6] J.B. Chaires, Possible origin of differences between van't Hoff and calorimetric enthalpy estimates, Biophys. Chem. 64 (1997) 15–23.
[7] J.R. Horn, D. Russell, E.A. Lewis, K.P. Murphy, van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: Are there significant discrepancies?, Biochemistry 40 (2001) 1774–1778.
[8] J. Tellinghuisen, A Monte Carlo study of precision, bias, inconsistency, and non-Gaussian distributions in nonlinear least squares, J. Phys. Chem. A 104 (2000) 2834–2844.
[9] J. Tellinghuisen, Bias and inconsistency in linear regression, J. Phys. Chem. A 104 (2000) 11829–11835.
[10] W.C. Hamilton, Statistics in Physical Science: Estimation, Hypothesis Testing, and Least Squares, The Ronald Press Co., New York, 1964.
[11] D.L. Albritton, A.L. Schmeltekopf, R.N. Zare, An introduction to the least-squares fitting of spectroscopic data, in: K. Narahari Rao (Ed.), Molecular Spectroscopy: Modern Research II, Academic Press, New York, 1976, pp. 1–67.
[12] P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1969.
[13] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge, UK, 1986.
[14] M.L. Doyle, J.H Simmons, S.J. Gill, Analysis of parameter resolution from derivatives of binding isotherms, Biopolymers 29 (1990) 1129–1135.
[15] J. Tellinghuisen, On the role of statistical weighting in the least-squares analysis of UV–visible spectrophotometric data, Appl. Spectrosc. 54 (2000) 1208–1213.
[16] Y. Liu, J.J. Sturtevant, Significant discrepancies between van't Hoff and calorimetric enthalpies II, Protein Sci. 4 (1995) 2559–2561.
[17] D.L. Albritton, A.L. Schmeltekopf, J. Tellinghuisen, R.N. Zare, Least-squares equivalence of different representations of rotational constants, J. Mol. Spectrosc. 53 (1974) 311–314.
[18] C.A. DiTusa, T. Christensen, K.A. McCall, C.A. Fierke, E.J. Toone, Thermodynamics of metal ion binding. 1. Metal ion binding by wild-type carbonic anhydrase, Biochemistry 40 (2001) 5338–5344.
[19] J. Tellinghuisen, Nonlinear least-squares using microcomputer data analysis programs: KaleidaGraph™ in the physical chemistry teaching laboratory, J. Chem. Educ. 77 (2000) 1233–1239.

[6] This was accomplished by defining several of the key quantities as library functions in KaleidaGraph; an example is available on request. In general such programs are not designed to handle the correlated error situation.