Received: 2 August 2018 | Revised: 17 December 2018 | Accepted: 7 February 2019

DOI: 10.1002/cem.3119

RESEARCH ARTICLE

WILEY Journal of CHEMOMETRICS

Sensitivity limits for determining 1:1 binding constants from spectrophotometric titrations via global analysis

Nathanael P. Kazmierczak | Joyce A. Chew | Anna R. Michmerhuizen | Seong Eun Kim | Zachary D. Drees | Andrew Rylaarsdam | Tasha Thong | Luke Van Laar | Douglas A. Vander Griend

Department of Chemistry & Biochemistry, Calvin College, Grand Rapids, Michigan, USA

Correspondence
Douglas A. Vander Griend, Department of Chemistry & Biochemistry, Calvin College, Grand Rapids, MI 49546.
Email: dvg@calvin.edu

Funding information
Camille and Henry Dreyfus Foundation, Henry Dreyfus Teacher Scholar Award; Arnold and Mabel Beckman Foundation, Beckman Scholars Program; National Science Foundation, Division of Chemistry, Grant/Award Number: 1310402

Abstract

The simultaneous nonlinear regression modeling of multiple wavelengths of spectrophotometric data allows binding constants to be determined with much higher precision than in previous single-wavelength methods; however, this method of global analysis has intrinsic limitations as well. Through Monte Carlo simulations on UV-vis titration data using various types of experimental errors, we demonstrate how the precision of binding constant calculation deteriorates under very strong binding regimes, as quantified by the product $K[H]_o$. We show that for a 1:1 binding model, global analysis can be reliably performed when $K[H]_o < 1000$, representing a significant improvement over previous recommendations. The relative impacts of different sources of error as well as the degree of overlap in molar absorptivity curves are quantified. Even under optimal conditions, errors in initial concentrations of the titration solutions are found to have the most impact on error in the calculated binding constant, while instrumental noise is largely weeded out by the global analysis technique. We propose experimental diagnostics indicating when the model has lost sensitivity to the binding constant and derive a novel experimental design formula for maximizing the precision of the binding constant calculation. The results imply the need to develop robust and accessible uncertainty estimation techniques competent to deal with concentration errors and asymmetric confidence intervals.

KEYWORDS
binding constant, hard modeling, Monte Carlo, titration, UV-vis

1 | INTRODUCTION

Binding constants in solution play a central role in a wide variety of chemical research areas. Supramolecular structures tuned for selective binding interactions can function as ionic sensors and separation agents.[1,2] Metal binding properties of polypeptides can yield insights into the protein-folding problem.[3] The binding strength of metal-ligand coordination reactions allows construction of a wide variety of geometric structures fabricated on the nanoscale.[4,5] Especially when the isolation of individual chemical components is not feasible, characterizing the thermodynamic properties of a solution constitutes a powerful route to controlling self-assembling molecular structures.[6,7] Consequently, precise, accurate,

efficient, and cheap determination of solution-phase binding constants is an important methodological goal in chemometric analysis.[8]

UV-vis spectrometry constitutes one popular method for ascertaining binding constants from simple titration data whenever most of the equilibrium species are sufficiently strong chromophores.[9,10] Traditionally, a spectrophotometric titration has involved monitoring the absorbance of the solution at one wavelength. The resulting data constitute a vector of absorbance readings, which could be linearized via a variety of approximate methods to obtain the binding constant between ligand and metal ion or supramolecular guest and host.[8] A single wavelength can also be used to estimate complexation stoichiometry via a Job's plot. However, the mathematical and statistical soundness of both methods has been called into serious question. Simultaneous use of all of the wavelengths, known as global analysis, is decidedly superior, as is rigorous nonlinear regression modeling for obtaining the best-fit binding constant. Stoichiometry is best determined by fitting and comparing multiple equilibrium models. If the compounds of interest do not absorb strongly, and pathlengths and solution concentration cannot be adjusted to compensate, then UV-vis may not be an appropriate instrumental choice for the given system.

In physicochemical global analysis, the data are treated as a matrix, and the mathematics of linear algebra and nonlinear regression are applied to obtain a simultaneous fit to all wavelengths.[11-14] This technique achieves superior resolution over averaging results from single-wavelength vectors[10] by making the global minimum more sharply defined on the error surface.[15] The basic implementation of global analysis in physicochemical modeling (also known as hard modeling) has been explained in detail elsewhere.[11-17] Briefly, the idea is to decompose the absorbance matrix, **D**, into the matrix product of smaller matrices of molar absorptivities, **R**, and equilibrium concentrations, **C**, with an additive residual error matrix: $\mathbf{D} = \mathbf{RC} + \mathbf{E}$. Nonlinear regression algorithms iteratively vary the trial binding constants in an attempt to find the values minimizing the root mean square of the residual errors matrix (RMSE). At each iteration, the equilibrium concentration matrix, **C**, is calculated from the initial concentrations. Linear regression is then used to obtain the optimal molar absorptivity profiles, **R**, for the given trial binding constants, thereby eliminating the linear parameters from the iterative fitting process. Multiple software packages have been developed,[18-23] of which two modern versions are freely available on the Web.[22,23] Physicochemical global modeling remains the method of choice for extracting precise binding constants from spectrophotometric titrations.[11,19]

When applying physicochemical global analysis to a UV-vis spectrophotometric titration, the chemometrician supplies a computer program with the equilibrium binding stoichiometry, the absorbance dataset, and initial guesses for the binding constant(s). In seconds, the analysis program returns the optimized value(s) for the binding constant(s) along with an estimate of the associated error bars. The estimated uncertainties often vary between these different methods,[24] and popular techniques such as linearized standard errors have limited reliability.[11] Five questions naturally arise:

1. Under what conditions will the model lose sensitivity to the binding constant; that is, fit equally well regardless of the precise value?
2. What sorts of experimental errors will have the biggest detrimental impact on the sensitivity of the results?
3. How do differing molar absorptivity profiles affect the propagation of this experimental error to the calculated binding constants?
4. What diagnostic metrics can the experimentalist apply to ascertain when the mathematical model has lost sensitivity?
5. How should a titration be designed so that the mathematical model retains maximum sensitivity?

As evidenced by the literature, these questions are answered most effectively through simulated data studies. Using visual inspection of simulated binding isotherms, Hirose has proposed that the product $K[\mathrm{H}]_o$, which effectively quantifies the strength of a 1:1 binding regime, should be less than 1 to ensure accurate results.[8] (Here, K refers to the association constant, K_a.) This work, however, did not use global analysis. Thordarson applied both visual inspection of binding isotherms and Monte Carlo methods on a simulated two-wavelength UV-vis titration to obtain a revised global analysis value of $K[\mathrm{H}]_o < 100$.[10] This work only considered experimental errors in the instrumental response (absorbance) data. Subsequent work[24] by Hibbert and Thordarson considered concentration and instrumental response error in Monte Carlo simulations using simulated NMR datasets with one resonance, but global analysis was not employed. Results from this study also suggested a $K[\mathrm{H}]_o < 100$ guideline. The same work includes simulations on 1:2 binding models and a discussion of uncertainty estimation techniques applied to published data.[25] Monte Carlo confidence

intervals were found to be superior to model comparison confidence intervals using the F statistic[26] and far superior to the symmetric linearized (asymptotic) uncertainty estimation.

These $K[H]_o$ guidelines give upper bounds on the strength of the binding regime that can be successfully analyzed. Consequently, these guidelines are particularly pertinent for UV-vis experiments, which generally probe stronger binding regimes than NMR titrations.[8,10] Calculation precision can also suffer in weak binding regimes.[8] To address this, Weber[27] and later Wilcox[28] have defined a P value metric that quantifies the "probability of binding," where P is defined as the equilibrium concentration of complex divided by either the initial guest or host concentration, whichever is smaller. They propose that weak binding regime titrations should range from at least 0.2 to 0.8 P. However, this metric is not helpful for strong-binding UV-vis systems because all P values approach one as complexation strength increases.[28] Indeed, under strong binding regimes, Thordarson observes that simulated isotherms only display substantial differences near the 1:1 equivalence point.[10] It is therefore important to identify which ratios of guest to host will maximize the sensitivity of the binding constant to the data, regardless of the binding regime strength, thus allowing for more rigorous determination of the intrinsic limitations for ascertaining a binding constant using global analysis.

This study focuses on the issue of sensitivity, seeking to quantify the conditions under which global analysis can quantify a binding constant for a model 1:1 equilibrium system. We rigorously address the five questions above using mathematical theorems and Monte Carlo simulations, taking care to recognize the inherent asymmetry involved. By refining Hirose's $K[H]_o$ metric and deriving a novel approach for maximizing model sensitivity that is applicable regardless of the strength of the binding regime, this work will empower researchers to better design titrations to minimize the uncertainty of the resulting binding constants. The results should give researchers confidence in using UV-vis global analysis not as a black box but instead as a substantial improvement upon single-wavelength methodology with well-understood limitations.

2 | METHODS

The Monte Carlo simulation,[29,30] a well-established computational statistics method, is particularly useful for this task. A large number of identical datasets are constructed from template materials (molar absorptivity profiles, initial concentrations, and $\Delta G°$ values), assuming that Beer's law and the law of mass action precisely hold. A unique pattern of random error with a common standard deviation is added to each dataset. Each dataset is then analyzed using the normal physicochemical global analysis technique, yielding a calculated result for the binding constant that can be compared with the "true" binding constant used to build the data. The difference represents the error introduced into the calculated binding constant because of the experimental error added to the data, and therefore a distribution can be obtained showing the spread in the calculated binding constants for the given error level. This procedure is repeated to find the limit at which the hard modeling process fails to realistically ascertain the binding constant.

We employ three types of normally distributed errors in our simulations: absorbance error, where error with a mean of 0 is added directly to the raw data; transmittance error, where error with a mean of 0 is added to the transmittance before converting back to absorbance for modeling; and composition error, where error with a mean of 1 is proportionally multiplied into the initial concentrations. The levels of transmittance and absorbance error were chosen to span typical instrument capabilities, given that most research-grade spectrometers have a transmittance error rating around 0.0002 T.[31] Composition error levels were chosen to cover an experimentally reasonable range from 0.01% to 1%.

Both experimental and idealized absorptivity data are used for the Monte Carlo simulations. Of the former nature, we include a dataset (hereafter referred to as "buckyball") modeled after a 1:1 supramolecular host-guest system involving fullerene nesting with macrocycle,[32] with the ligand absorptivity removed for ease of simulation. Of the latter nature, we include datasets where the molar absorptivity profiles are represented alternatively by Gaussians or downward facing parabolas with negative values truncated at 0, enabling a precise description of how absorptivity curve overlap influences calculation results without competing effects from curve shape. A3 parabolas refer to the case where host, guest, and the 1:1 host:guest complex all absorb, and each molar absorptivity profile has a width of 200 nm and is nonoverlapping with the others. B3 parabolas are identical to A3 parabolas, except the peaks of each parabola are offset by only 50 nm, resulting in overlap between the absorptivity profiles. C3 Gaussians refer to the case when all species absorb, but the peaks of each Gaussian are offset by only 17 nm. In C2 Gaussians, only the host and 1:1 host-guest complex absorb. All Gaussians have a half width of 200 nm.

The overlap between the molar absorptivity profiles is an important feature of the model sensitivity and can be conveniently summarized by treating the curves as vectors. The following standard formula gives the angle θ between any two absorptivity vectors:

$$\theta = \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{\left\|\vec{u}\right\|\left\|\vec{v}\right\|}\right).$$

As θ approaches 0°, the curves approach complete overlap; as it approaches 90°, the curves approach complete orthogonality. A3 parabolas have an angle of 90°; adjacent B3 parabolas, 40.1°; adjacent C2/3 Gaussians, 8.1°; buckyball curves, 3.8°. No systematic study is needed for the absolute values of the molar absorptivity coefficients because, while the absorbance in UV-vis spectrophotometry should be held less than 1, pathlength adjustments and dilutions can and should be tuned for each unique experimental system under consideration to place the signal on the correct scale.

Two main schemes were used for the initial solution compositions. "Dilution" profiles assume 50 equal aliquots of guest solution are added to a host solution until the final ratio of $[G]_o$ to $[H]_o$ reaches 2.5 equivalents. "Spiked" profiles assume the simulated titrant solution contains the same concentration of host as the analyte solution so that $[H]_o$ remains constant throughout the experiment. In these simulations, spiked profiles are run to 1.5 equivalents to ensure that the 1:1 equivalence is reached near the same point during the titration. Spiked profiles should be employed when the guest species does not absorb so that the dataset may be safely shifted upwards, eliminating negative values and any associated bias without introducing new errors.[17] These profiles are additionally employed in the existing experimental literature to reduce the possibility of aggregation side reactions,[10] to avoid unnecessary dilution factors in the data analysis,[10] and to permit easier visual identification of the titration endpoint. We therefore employ dilution profiles for simulations involving three absorbing compounds (A3 and B3 parabolas), but use spiked profiles for each simulation involving only two absorbing compounds (C2 Gaussians and buckyballs), as well as for C3 Gaussian simulations for direct comparison.

The choice of initial host concentration, $[H]_o$, influences the strength of the binding regime. $[H]_o$ is constant for spiked titrations; for dilution profiles, we use the value of $[H]_o$ in the first solution. We initially employ $[H]_o = 0.1$ M for the parabola and Gaussian curves and $[H]_o = 1 \times 10^{-4}$ M for the buckyball curves, approximating the concentration regimes used for common spectrophotometric titrations involving transition metal coordination and supramolecular host-guest binding, respectively. We run the initial A3 parabola simulations to 2.5 equivalents, those with C2 Gaussians out to 1.5 and the buckyball simulations to 1000 equivalents.[8] For the studies varying $[H]_o$ values, parabola curves are run to 2.5 equivalents, while Gaussian and buckyball curves are run to 1.5. In each case, these settings ensure each simulation contains datasets reaching at least 80% host complexation before the end of the titration.

3 | RESULTS AND DISCUSSION

To ascertain the relative impacts of different types of experimental error, Monte Carlo simulations using the A3 molar absorptivity profiles ($[H]_o = 0.1$ M) were run with absorbance error, transmittance error, composition error, and finally both composition and transmittance error together, which most closely simulates real experimental conditions.

As seen in Figure 1, increasing the exergonicity of the reaction increases the spread of the Monte Carlo distribution significantly and asymmetrically. This indicates that, as expected, stronger binding regimes admit to less precise determination of the equilibrium constants, especially in the exergonic direction. However, the quantities of transmittance error produced by a research-grade spectrometer (~0.0002 T) inflict relatively small errors on the calculated binding constant as long as $\log(K[H]_o) < 7$. A very similar result holds for a comparable absorbance error simulation (Figure S-5). In both cases, the calculations are slightly biased towards larger outliers on the exergonic side for ΔG° values, although the medians of the distributions stay close to 0. This asymmetry appears to arise from the nonlinear nature of the equilibrium concentration calculations.

Another asymmetric feature of the Monte Carlo data is the presence of extreme exergonic outliers. This effect arises in our simulations for a wide variety of absorptivity profiles, concentration regimes, and error types; it has also arisen in our experience modeling experimental datasets. We suspect that, in the exergonic limit, certain error patterns make it appear as if more product formed than was possible under mass balance constraints. This causes the model to send ΔG° towards negative infinity to best account for the data. Obviously, these strongly exergonic outliers indicate that the

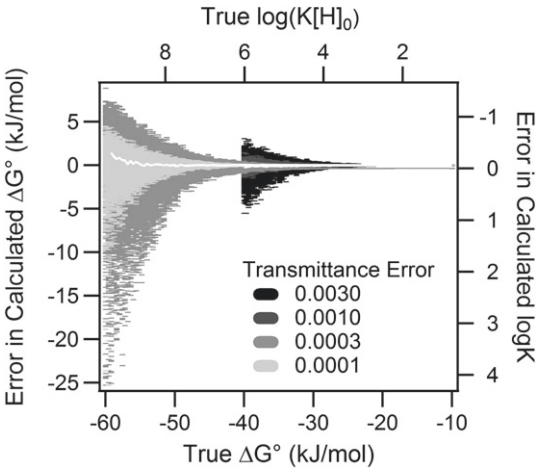

FIGURE 1 Monte Carlo (500 iterations × 101 true ΔG° settings (at 298 K) × 2 error settings + 500 × 151 × 2 = 252 000 datasets) using A3 parabolas with transmittance error. Shading corresponds to standard deviation of error. Four white lines represent medians. Seven hundred twenty-three exergonic outliers not shown. Each dot represents one optimized dataset

particular combination of error conditions and ΔG° values are not tenable for experimental work. Thordarson reports a Monte Carlo simulation resulting in an upper 95% confidence limit for K of +10%,[14] suggesting a similar outlier phenomenon in his analysis.[10]

Moving to composition error, the simulations tell a less optimistic story. For realistic 0.1% error, the calculated ΔG° value begins to diverge sharply after about ΔG° = −35 kJ/mol. The median error in calculated ΔG° becomes more positive as the true ΔG° becomes more exergonic, while simultaneously exergonic outliers begin to appear in abundance. Adding transmittance error in addition to the composition error has little additional effect (Figure 2; Figure S-6 shows result with composition error alone), showing that the distribution spread arising from composition and transmittance error is not additive. Clearly, composition error proves to be the more deleterious form of error at realistic experimental values. The asymmetries inherent in these distributions can be visualized through a histogram plot (Figure 3), which represents a vertical cross-section through Figure 2 at ΔG° = −40 kJ/mol.

To probe the effects of different absorptivity profiles on the calculation results, the same set of simulations was performed using the C2 Gaussians (θ = 8.1°) with $[H]_o$ = 0.1 M and buckyball curves, (θ = 3.8°) with $[H]_o$ = 0.0001 M. The former display far greater overlap than the A3 parabolas, and as a result, the Monte Carlo distributions (Figure 4) show both more error for any given true ΔG° value and an increased number of outliers. While the distributions of error remain relatively symmetric when the level of error is small, dramatic asymmetry occurs when the binding regime strengthens. Clearly, there can be no hope of reliably obtaining the binding constant at a true value

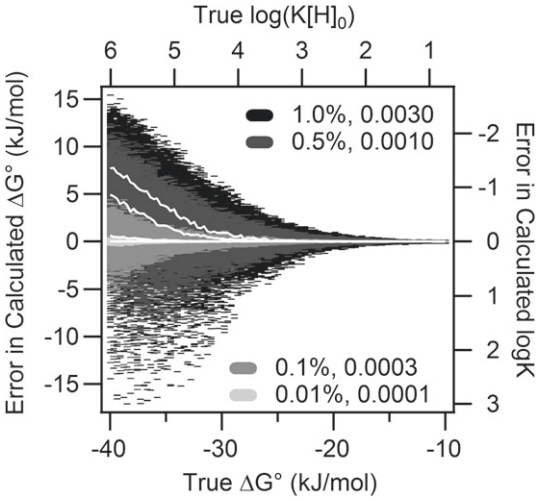

FIGURE 2 Monte Carlo (500 × 101 × 4) results using A3 parabolas with composition and transmittance error. Shading corresponds to error levels. Four white lines represent medians. One thousand four hundred exergonic outliers not shown

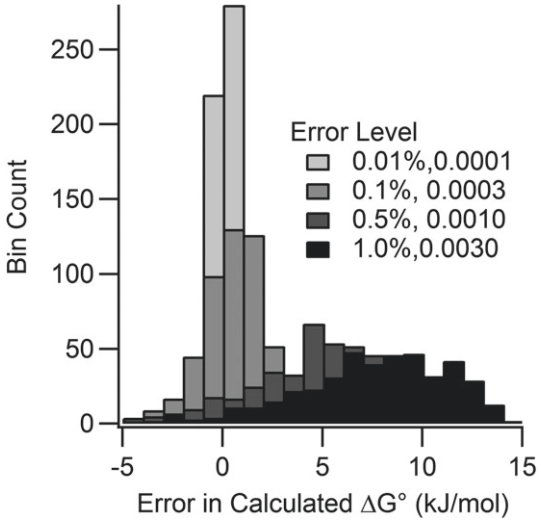

FIGURE 3 Histogram of Monte Carlo results for true ΔG° = −40 kJ/Mol simulation using A3 parabolas with composition and transmittance error. Ninety-eight exergonic outliers not shown

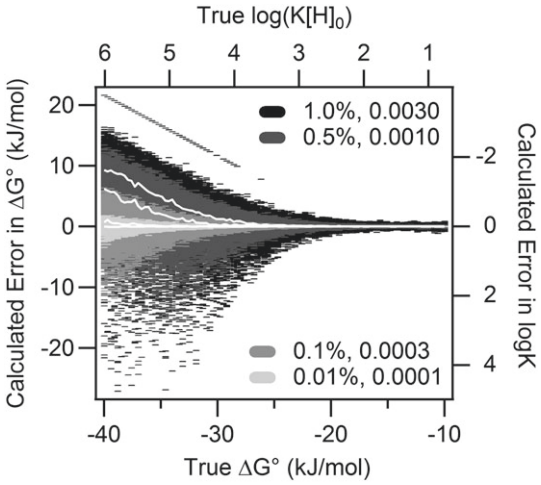

FIGURE 4 Monte Carlo (500 × 101 × 4) results using C2 Gaussians with composition and transmittance error. $[H]_o$ = 0.1 M. Shading corresponds to error levels. Four white lines represent medians. Four thousand three hundred thirty-one exergonic outliers not shown

of ΔG° of −40 kJ/mol for realistic values of composition (≥0.1%) and transmittance (≥0.0002) error. The buckyball curves display the greatest amount of overlap among the molar absorptivity profiles studied, and consequently, the simulations are extremely sensitive to added error (Figure 5). Even with composition error levels reduced by a factor of 10, the Monte Carlo distributions exhibit prohibitively asymmetric errors that increase in magnitude as the reactions become more exergonic. Under mild realistic error conditions (composition error of 0.1%), the binding constant cannot be ascertained with confidence beyond a true value of ΔG° of −25 kJ/mol.

Having observed the relative effects of different types of error and molar absorptivity curves, we now show numerically that the $K[H]_o$ metric serves as a useful invariant for binding strength across different concentration regimes. As exemplified by Figure 6 and also Figure S-4, Monte Carlo simulations demonstrate that the cutoff point at which the standard deviation of the calculated ΔG° value reaches 1.0 kJ/mol occurs very regularly, with log$K$ increasing by 1 as $[H]_o$ decreases by a factor of 10. Distributions are generally symmetric up until the $\sigma \geq 1$ kJ/mol cutoff and the density of each distribution tapers off smoothly with a slight clustering towards endergonic errors (Figure S-9).

While $K[H]_o$ remains quite unaffected by the value of $[H]_o$ chosen, the absorptivity profiles chosen have a substantial impact, varying the cutoff $K[H]_o$ by more than two orders of magnitude between the A3 parabolas and the buckyball absorptivity curves (Table 1). Predictably, absorptivity curves with a smaller amount of overlap have larger cutoff values

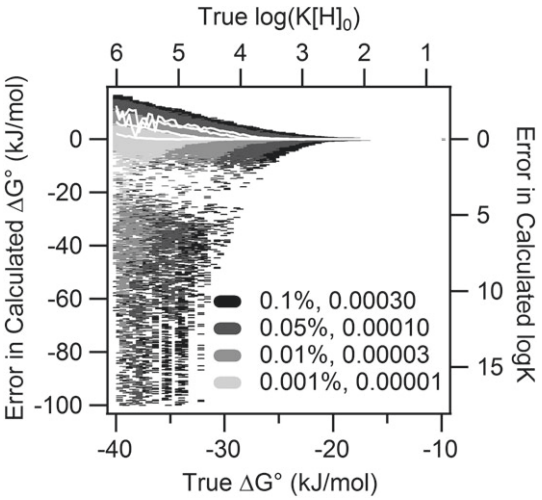

FIGURE 5 Monte Carlo (500 × 101 × 4) results using buckyball curves with composition and transmittance error. $[H]_o = 1 \times 10^4$ M. Shading corresponds to error levels. Four white lines represent medians. Twenty-two thousand nine hundred twenty-three exergonic outliers not shown

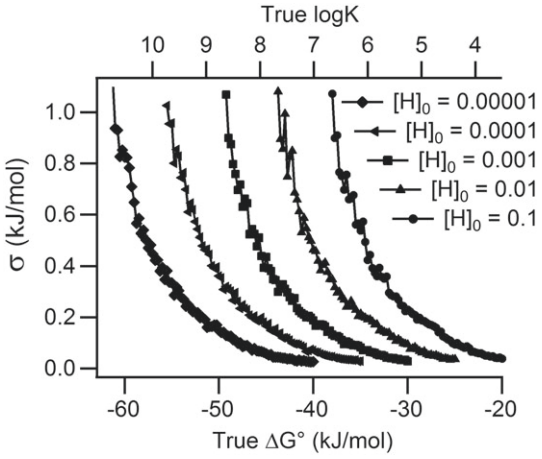

FIGURE 6 Graphical representation of the width of the Monte Carlo distribution for different concentration regimes, using A3 parabolas. ΔG° was decreased in increments of 0.25 kJ/Mol until the point where the sample standard deviation exceeded 1 kJ/Mol. Two hundred datasets were computed at each true value of ΔG° to form the sample distribution. Composition error level = 0.1%, transmittance error level = 0.0003

for $K[H]_o$. Interestingly, C2 Gaussians have higher cutoff values than those for C3 Gaussians, suggesting that titrations with a nonabsorbing guest may be slightly more resilient to error.

Table 1 captures how error level and the overlap of molar absorptivity curves impact sensitivity when determining a binding constant from UV-vis data. It should serve as a point of reference when designing a spectrophotometric titration. (Complete results of all combinations of the four error types and three molar absorptivity profiles are presented in Table S-I.) Given practical experimental constraints of low magnitudes of molar absorptivity coefficients, however, it may not be possible to construct chemical solutions satisfying the cutoff $K[H]_o$. Two situations are considered below, one amenable to a spectrophotometric titration with appropriate adjustment of the experimental conditions and one not.

In the first place, suppose that the molar absorptivity values of the equilibrium species are large, such that an initial host stock solution leads to too much absorbance for UV-vis measurement. In this case, the host stock solution can always be diluted, as dilution can only improve the sensitivity according to the $K[H]_o$ value. If a larger host concentration is desired given the concentration of the guest stock solution, then the pathlength can be decreased instead. Thus, strongly absorbing host compounds offer flexibility in titration design.

TABLE 1 Critical values of ΔG° (kJ/mol) and $K[H]_o$ at which the standard deviation of the Monte Carlo distribution becomes larger than 1 kJ/mol. "Mild error" refers to a transmittance error of 0.0003 and a composition error of 0.1%. "Harsh error" refers to a transmittance error of 0.001 and a composition error of 0.5%. "Δ isotherm RMS" is obtained by calculating the binding isotherm both at $\Delta G°_{cutoff}$ and $\Delta G°_{cutoff} - 1$ kJ/mol, then taking the root mean square of the difference between the two mole fraction isotherm vectors

	Error Level		0.1 M	0.01 M	0.001 M	0.0001 M	0.00001 M	Average $K[H]_o$	Δ Isotherm RMS ($\times 10^3$)
A3 parabolas (dilution)	Mild	ΔG°	−38	−43.75	−49.25	−55.5	−61.25	486 000 ± 42 000	0.14
		$K[H]_o$	460 000	470 000	430 000	530 000	540 000		
	Harsh	ΔG°	−29.25	−34.5	−39.75	−45.75	−51.75	11 200 ± 1400	1.69
		$K[H]_o$	13 000	11 000	9000	10 000	12 000		
B3 parabolas (dilution)	Mild	ΔG°	−33.75	−39.5	−45.5	−51.25	−57	91 000 ± 6000	0.39
		$K[H]_o$	82 000	84 000	94 000	96 000	97 000		
	Harsh	ΔG°	−25.5	−31.25	−37	−42.25	−48	2830 ± 230	4.51
		$K[H]_o$	2900	3000	3100	2500	2600		
C2 Gaussians (spiked)	Mild	ΔG°	−34.25	−40	−45.75	−51.5	−57	102 000 ± 4000	0.19
		$K[H]_o$	100 000	100 000	100 000	110 000	100 000		
	Harsh	ΔG°	−27.5	−32.75	−39	−44.75	−50.25	6470 ± 530	1.74
		$K[H]_o$	6600	5500	6900	7000	6400		
C3 Gaussians (spiked)	Mild	ΔG°	−32.5	−38.25	−44	−49.5	−55.25	49 500 ± 1900	0.36
		$K[H]_o$	50 000	51 000	52 000	47 000	48 000		
	Harsh	ΔG°	−24.5	−30.5	−36	−41.5	−47.5	2050 ± 100	3.97
		$K[H]_o$	2000	2200	2000	1900	2100		
Buckyballs (spiked)	Mild	ΔG°	−24	−29.25	−35.25	−41	−47	1550 ± 130	8.61
		$K[H]_o$	1600	1300	1500	1500	1700		
	Harsh	ΔG°	−15.75	−22.25	−27.5	−33.5	−39	70 ± 7	59.90
		$K[H]_o$	58	79	66	74	69		

In the second place, suppose that the molar absorptivity values are small, such that an initial stock solution does not give enough absorbance for a spectrometer, but the binding constant is large. Such a situation may be unsuitable for UV-vis analysis. To obtain enough absorbance for an acceptable signal-to-noise ratio, one would like to increase the concentrations. However, if K is large to begin with, this easily leads to a violation of the $K[H]_o$ metric, which would render any titration unreliable in accuracy and reproducibility as demonstrated above. The alternative solution would be to increase the pathlength, but this is often difficult or impossible. In these situations, if the researcher cannot find an alternate spectral region with larger molar absorptivity values, then an instrumental technique other than UV-vis should be used.

The $K[H]_o$ figure functions as a guideline to determine when a titration can be expected to yield reliable results or not. In this way, the theoretical Monte Carlo simulations directly inform experimental practice. However, it is also important that for any given experiment, the sensitivity to the data be ascertained after the fact. We therefore present diagnostics to quantify the sensitivity followed by insights regarding how to update the experimental design to achieve greater sensitivity.

While the impact of the molar absorptivity profiles and the precise magnitude of the residuals vary from system to system, the computed binding isotherm (mole fraction of complex, relative to $[H]_o$, plotted against equivalents of guest) provides a point of comparison. For 1:1 systems, the change in the concentration matrix caused by a change in ΔG° is entirely captured by the change in the single isothermal curve. This "difference isotherm" changes shape with the strength of the binding regime, and not all segments exhibit equal sensitivity to ΔG°. Starting at the cutoff value for ΔG°, we compute the root mean square of difference isotherm for a −1 kJ/mol perturbation (Table 1). This single number represents the sensitivity of the concentration profiles to the binding constant used, and this metric can be applied to any equilibrium model.

These observations can be formalized mathematically by calculating the number of guest equivalents needed to obtain maximum concentration sensitivity to the binding constant. As shown in the Supporting Information, the 1:1 equilibrium constant expression can be reparameterized in terms of just three variables: $X_{HG} = [HG]/[H]_o$, $E = [G]_o/[H]_o$, and $B = K[H]_o$, rather than the typical four variables K, $[H]_o$, $[G]_o$, and [HG]. Under this scheme, the analytical solution to the quadratic equation can be written in terms of the mole fraction of complex:

$$X_{HG} = \frac{1}{2}\left[1 + E + \frac{1}{B} - \sqrt{\left(1 + E + \frac{1}{B}\right)^2 - 4E}\right]. \quad (1)$$

Equation 1 completely describes the binding isotherm. This demonstrates that the mole fraction of complex is completely determined by just two variables, B, the $K[H]_o$ binding strength metric and E, the equivalents of guest added. This observation may additionally prove useful for experimentalists using NMR titrations, where X_{HG} is the quantity directly related to the change in the resonant frequency over the course of the titration.[9,10]

The derivative of this isotherm with respect to B (the strength of the binding regime) represents its sensitivity to K:

$$\frac{dX_{HG}}{dB} = \frac{1}{2B^2}\left[\frac{1 + E + \frac{1}{B}}{\sqrt{\left(1 + E + \frac{1}{B}\right)^2 - 4E}} - 1\right]. \quad (2)$$

Figure 7 displays how Equation 2 changes shape as a titration shifts between weak and strong binding regimes, which has significant implications for designing a suitable useful titration experiment. Exergonic systems are not merely challenging because the magnitude of the sensitivity decreases as $K[H]_o$ increases, as can be inferred from the dependence on B^{-2}, but additionally because the derivative shape effectively concentrates all of the chemical information about the binding constant into a very narrow region close to the 1:1 equivalence point. Thus, unless the titration is constructed carefully, only a few chemical solutions (perhaps even none) will exhibit sensitivity to the binding constant. The success or failure of the experiment will then ride entirely on the instrumental noise or concentration error in these few solutions, leading to a potentially precipitous breakdown in the calculated binding constant precision.

Equation 2 additionally allows analytical calculation of the regions of the titration that are most sensitive to a change in the binding constant. The maximum in this sensitivity function occurs at the number of equivalents of guest obtained by setting the derivative of Equation 2 with respect to $E = [G]_o/[H]_o$ equal to 0 and solving for E (Supporting Information).

$$E = \frac{[G]_0}{[H]_0} = 1 + \frac{1}{K[H]_0} = 1 + \frac{1}{B}. \quad (3)$$

Equation 3 identifies the number of equivalents in the single best chemical solution in a titration for determining the binding constant. Notice that in the exergonic limit as K goes to infinity, the most sensitive chemical solution approaches the 1:1 stoichiometry (Figure 7) in accord with previous observations[10] but is not precisely the 1:1 solution. Drawing on Weber's notion of the P value, we see that in the endergonic limit as K goes to 0, the best solution

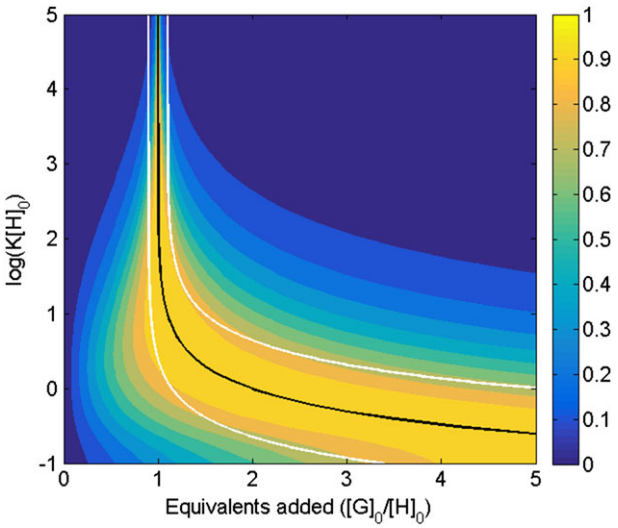

FIGURE 7 Contour map of Equation 2, normalized for each $K[H]_o$ value to a height of one. The central black line marks the most sensitive solution as expressed by Equation 3. Adjacent white lines mark the sensitivity envelope for $C = 0.1$ as defined by Equation 4

approaches a P value of 0.5, which is of course the center of the recommended 0.2-0.8 range. This can be confirmed by observing that for $P = 0.5$, $[G]_o/[H]_o = ½ + 1/(K[H]_o)$, where the second term dominates for small $K[H]_o$.

Equation 3 therefore provides a novel basis for identifying the most sensitive region of the titration because it applies to both strong and weak binding systems. This is especially useful for not only evaluating titrations but also designing retitrations. Hirose refined boundaries when he suggested running a titration between $1/5 + 1/(4K)$ and $4/5 + 4/K$ equivalents of guest. Equation 4 defines an equivalents envelope for a titration that is centered on the most sensitive solution while maintaining a minimal width of $2C$, where C is an arbitrary parameter.

$$(1 - C) + \frac{1}{4K[H]_0} < \frac{[G]_0}{[H]_0} < (1 + C) + \frac{4}{K[H]_0}. \tag{4}$$

This pair of equations is constructed to approach Hirose's boundaries as K decreases and yet converge around 1:1 as K increases.

If titrations are designed to focus on the envelope defined in Equation 4, could the sensitivity limits be extended, and if so, what is the optimal value of C? Monte Carlo simulations from $K[H]_o = 1 \times 10^{-6}$ to 1×10^{6} were performed to find out. "Control" titrations of 51 linearly spaced solutions were compared with "Envelope" titrations, in which the 50 linearly spaced solutions were compressed into the range defined by Equation 4 plus one solution of free host. As shown in Table 2, focusing titrations within the Equation 4 envelope improves $K[H]_o$ sensitivity as long as $C \geq 0.1$, but the sensitivity deteriorates dramatically for smaller values of C. This is because all of the chemical solutions become compressed in a very narrow region in which the total absorbance change is relatively small. Effectively, this situation results in a precipitous worsening of the signal-to-noise ratio since any instrumental errors will have more impact on the analysis results when the data are relatively featureless to begin with.

A second type of targeting strategy was employed in which 40 of the 51 solutions were evenly spaced within the envelope defined by Equation 4, with six more covering the range from free host to the lower bound and five more covering the range from the upper bound to 120% of the upper bound or two equivalents, whichever is larger. (The control titrations cover an identical range.) These "Telescoping" titrations are designed to balance the focus on the most sensitive region of a titration with the need to cover the full dynamic range of the signal. Graphical examples of Control, Envelope, and Telescoping titrations are provided in the Supporting Information (Figures S-12-S-13). Table 2 shows that telescoping profiles perform as well or better than either the Control or the Envelope strategy for all values of C tested. Therefore, after an initial titration, an experimenter may choose to titrate again using this telescoping strategy in order to maximize the sensitivity of the titration to the binding constant.

TABLE 2 $K[H]_o$ cutoff values for varying strategies of sensitivity targeting under two different levels of error. Control simulations consist of 51 evenly spaced solutions with no targeting. For envelope simulations, the solutions are still evenly spaced but compressed into the range defined by Equation 4 plus one solution of free host. Telescoping simulations consist of 40 solutions within the envelope of Equation 4 with six more covering the range from free host to the lower bound and five more covering the range from the upper bound to 120% of the upper bound or two equivalents, whichever is larger; 500 datasets were computed for each setting of K, using spiked C2 Gaussians with $[H]_o = 0.1$ M. Δ isotherm root mean square (RMS) is defined as in Table 1

		Mild Error (Comp. = 0.1%, Trans. = 0.0003)			Harsh Error (Comp. = 0.5%, Trans. = 0.001)		
C	**Diagnostic**	**Control**	**Envelope**	**Telescoping**	**Control**	**Envelope**	**Telescoping**
0.5	$K[H]_o$	3.6×10^4	4.8×10^4	4.8×10^4	5.3×10^3	6.9×10^3	6.9×10^3
	Δ isotherm RMS ($\times 10^3$)	0.59	0.67	0.63	2.39	2.75	2.56
0.25	$K[H]_o$	2.8×10^4	8.3×10^4	8.3×10^4	5.3×10^3	1.2×10^4	9.1×10^3
	Δ isotherm RMS ($\times 10^3$)	0.71	0.63	0.60	2.39	2.56	2.90
0.1	$K[H]_o$	3.6×10^4	1.5×10^5	1.5×10^5	5.3×10^3	1.2×10^4	1.6×10^4
	Δ isotherm RMS ($\times 10^3$)	0.59	0.64	0.60	2.39	3.69	2.83
0.01	$K[H]_o$	3.6×10^4	4.4×10^2	3.3×10^5	5.3×10^3	83	2.1×10^4
	Δ isotherm RMS ($\times 10^3$)	0.59	23.38	0.79	2.39	43.16	3.46
0	$K[H]_o$	3.6×10^4	1.9×10^2	5.8×10^5	5.3×10^3	63	2.8×10^4
	Δ isotherm RMS ($\times 10^3$)	0.59	32.69	0.68	2.39	46.83	3.05

Abbreviation: RMS: root mean square

It is instructive to note that the concentration profile sensitivity, as quantified by the difference (Δ) isotherm root mean square (RMS) cutoff values, remains relatively constant for both envelope and telescoping profiles for large to moderate C values regardless of targeting strategy even as the $K[H]_o$ cutoff value increases. This implies that for a given level of error and set of molar absorptivity profiles, the change in the binding isotherm constitutes a second metric for elucidating the exergonic limitations of the modeling. Consequently, the analytical intuition developed above is confirmed: envelope and telescoping targeting strategies effectively extend the $K[H]_o$ cutoff value because the sensitivity of the concentration profiles to the binding constant is maximized. The Δ isotherm RMS concept provides a useful diagnostic to complement the $K[H]_o$ metric, as it incorporates initial concentrations from throughout the titration. As crucial as it is, looking only at $K[H]_o$ fails to account for the necessity of structuring the initial concentrations to optimize sensitivity, which otherwise could allow the lack of sensitivity in a poorly structured titration to go undetected. The Δ isotherm RMS helps confirm the design of a particular experiment.

The preceding discussion explains the primary importance of experimental initial concentration errors in determining the accuracy of the hard-modeling equilibrium calculation under exergonic binding regimes. If stock solution concentrations are off even by a small amount, the narrowing sensitivity envelope near 1:1 may not be sufficiently probed, leading to potentially catastrophic effects. This empirical effect of concentration error has been noted previously. Thordarson observed that experimental errors in the $[G]_o/[H]_o$ ratio may have the largest impact.[9,24] In the related area of spectrophotometric kinetics, Billeter et al. reported that the variance in the initial reagent concentrations dominate the variance in the calculated rate constant, with little to no contribution from the variance of the residuals.[30] This work provides a rigorous numerical foundation for such intuition in the context of hard modeling for UV-vis equilibrium titrations.

Two key implications arise from pinpointing the dominant source of experimental error. First, analytical procedures should prioritize the minimization of uncertainty in the preequilibration concentrations. Second and most important, error estimation techniques must be designed to deal with initial concentration errors. Linearized standard errors, which have become the method implemented in most software for hard modeling of equilibrium spectrophotometric titrations, assume that all of the experimental errors are normally distributed in the response variables (in this case, the absorbance data).[33] Our study has shown that the most impactful errors lie in the predictor variables (initial concentrations) instead. This implies that linearized standard error methods will systematically underestimate the uncertainty in the calculated binding constants because these methods will only respond robustly to spectrometer noise. One indication of this comes from the asymmetry of the errors on the calculated ΔG° values seen in the Monte Carlo simulations. Linearized methods always yield symmetric error bars, a result clearly inappropriate for uncertainty estimation in this situation.

Following the outputs of programs such as SpecFit,[18] it has become common practice to publish binding constants with linearized uncertainty estimates, although this practice has been challenged recently.[24] The implication is that the literature likely contains uncertainty estimates that are systematically too small. It is therefore imperative to develop methods and accessible software that enable experimentalists to perform more robust uncertainty estimations than currently available. Norman and Maeder suggest that bootstrapping approaches may furnish more reliable uncertainty estimates;[11] additionally, Thordarson has implemented a related Monte Carlo technique capable of handling errors in initial concentrations.[24] This study lays a theoretical foundation for the importance of making such methodologies easily accessible in UV-vis hard modeling.

4 | CONCLUSIONS

The foregoing Monte Carlo simulations speak to both the power and the potential pitfalls of nonlinear regression modeling for equilibrium systems. On the one hand, the combination of global analysis and hard modeling enables accurate binding constants to be obtained in binding regimes far stronger than the previous guideline of $K[H]_0 < 1$. While the precise cutoff depends upon the overlap in the molar absorptivity profiles used, many experimental systems will have overlaps comparable with those employed in Table 1. The comparison (Table 1) of C2 and C3 systems suggests that titrations with nonabsorbing guests may be somewhat more robust, although further studies with additional molar absorptivity profiles would be needed to conclusively demonstrate a general trend. Ultimately, we suggest a revised guideline of $K[H]_0 < 1000$ for a conservative estimate of the magnitude of experimental errors, although well-executed titrations may admit to accurate determination even when $K[H]_0 = 10\,000$ or greater (Table 2). Nevertheless, hard modeling can still lead to large asymmetric errors in excessively strong binding regimes because the model loses

sensitivity to the precise value of K employed. This feature is fundamental to the nonlinear nature of chemical equilibrium and cannot be eradicated. As a corollary, systems with small absorptivity coefficients and large equilibrium constants may not be amenable to analysis by UV-vis methods. However, structuring the titration's initial concentrations to target the optimal chemical solutions can raise the $K[H]_0$ cutoff by up to an order of magnitude. When implementing hard modeling, it is imperative that sensitivity be ascertained if a binding constant is ultimately reported. Researchers and reviewers should expect diagnostic evaluations involving signal error estimation, concentration uncertainty, isotherm sensitivity, molar absorptivity curve overlap, and $K[H]_0$. Emphasis on these criteria will increase the reliability of binding constant data.

ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation Division of Chemistry (1310402) and a Teacher-Scholar Award from the Henry and Camille Dreyfus Foundation. We are also grateful to the Arnold and Mabel Beckman Foundation for supporting Calvin College and Nathanael Kazmierczak as a Beckman Scholar.

ORCID

Nathanael P. Kazmierczak https://orcid.org/0000-0002-7822-6769
Douglas A. Vander Griend https://orcid.org/0000-0002-8828-1112

REFERENCES

1. Zahran EM, Fatila EM, Chen C-H, Flood AH, Bachas LG. Cyanostar: C–H hydrogen bonding neutral carrier scaffold for anion-selective sensors. *Anal Chem*. 2018;90(3):1925-1933.
2. Wilson AM, Bailey PJ, Tasker PA, Turkington JR, Grant RA, Love JB. Solvent extraction: the coordination chemistry behind extractive metallurgy. *Chem Soc Rev*. 2014;43(1):123-134.
3. Pozdnyakova I, Wittung-Stafshede P. Biological relevance of metal binding before protein folding. *J Am Chem Soc*. 2001;123(41):10135-10136.
4. Hall BR, Manck LE, Tidmarsh IS, et al. Structures, host–guest chemistry and mechanism of stepwise self-assembly of M4L6 tetrahedral cage complexes. *Dalton Trans*. 2011;40(45):12132-12145.
5. Manck LE, Benson CR, Share AI, Park H, Vander Griend DA, Flood AH. Self-assembly snapshots of a 2 × 2 copper(I) grid. *Supramol Chem*. 2014;26(3-4):267-279.
6. Grzelczak M, Vermant J, Furst EM, Liz-Marzán LM. Directed self-assembly of nanoparticles. *ACS Nano*. 2010;4(7):3591-3605.
7. Jacobs WM, Frenkel D. Self-assembly of structures with addressable complexity. *J Am Chem Soc*. 2016;138(8):2457-2467.
8. Hirose K. Determination of Binding Constants. In: Schalley CA, ed. *Analytical Methods in Supramolecular Chemistry*. 1st ed. Weinheim: Wiley-VCH; 2007:17-54.
9. Thordarson P. Binding Constants and Their Measurement. In: *Supramolecular Chemistry*. Chichester, UK: John Wiley & Sons, Ltd; 2012:239-274.
10. Thordarson P. Determining association constants from titration experiments in supramolecular chemistry. *Chem Soc Rev*. 2011;40(3):1305-1323.
11. Norman S, Maeder M. Model-based analysis for kinetic and equilibrium investigations. *Crit Rev Anal Chem*. 2006;36(3-4):199-209.
12. Maeder M, Zuberbuehler AD. Nonlinear least-squares fitting of multivariate absorption data. *Anal Chem*. 1990;62(20):2220-2224.
13. Maeder M, Neuhold Y-M. *Practical Data Analysis in Chemistry*. Amsterdam, Netherlands: Elsevier; 2007.
14. Gampp H, Maeder M, Meyer CJ, Zuberbühler AD. Calculation of equilibrium constants from multiwavelength spectroscopic data—I: mathematical considerations. *Talanta*. 1985;32(2):95-101.
15. Beechem JM. Global Analysis of Biochemical and Biophysical Data. In: Brand L, Johnson M, Abelson J, Simon M, eds. Methods in Enzymology 210 *Numerical Computer Methods*. San Diego, California: Academic Press; 1992:37-54.
16. Sabatini A, Vacca A, Gans P. Mathematical algorithms and computer programs for the determination of equilibrium constants from potentiometric and spectrophotometric measurements. *Coord Chem Rev*. 1992;120:389-405.
17. Kazmierczak NP, Vander Griend DA. Properly Handling Negative Values in the Calculation of Binding Constants by Physicochemical Modeling of Spectroscopic Titration Data. Submitted for publication, 2018.
18. Gampp H, Maeder M, Meyer CJ, Zuberbühler AD. Calculation of equilibrium constants from multiwavelength spectroscopic data—II: Specfit: two user-friendly programs in basic and standard Fortran 77. *Talanta*. 1985;32:251-264.
19. Gans P, Sabatini A, Vacca A. Investigation of equilibria in solution. Determination of equilibrium constants with the HYPERQUAD suite of programs. *Talanta*. 1996;43(10):1739-1753.

20. ReactLab Equilibria|Jplus ReactLab http://jplusconsulting.com/products/reactlab-equilibria/ (accessed Jun 6, 2018).
21. Brugger J. BeerOz, a set of Matlab routines for the quantitative interpretation of spectrophotometric measurements of metal speciation in solution. *Comput Geosci*. 2007;33(2):248-261.
22. Thordarson P. Supramolecular.org http://supramolecular.org/ (accessed Jun 7, 2018).
23. Vander Griend DA. SIVVU http://sivvu.org/ (accessed Jun 7, 2018).
24. Hibbert DB, Thordarson P. The death of the job plot, transparency, open science and online tools, uncertainty estimation methods and other developments in supramolecular chemistry data analysis. *Chem Commun*. 2016;52(87):12792-12805.
25. Howe ENW, Bhadbhade M, Thordarson P. Cooperativity and complexity in the binding of anions and cations to a tetratopic ion-pair host. *J Am Chem Soc*. 2014;136(20):7505-7516.
26. Exner O. Calculating equilibrium constants from spectral data: reliability of the Benesi-Hildebrand method and its modifications. *Chemom Intell Lab Syst*. 1997;39(1):85-93.
27. Weber G. The Binding of Small Molecules to Proteins. In: Pullman B, Weissbluth M, eds. *Molecular Biophysics*. New York: Academic Press; 1965.
28. Wilcox CS. Design, Synthesis, and Evaluation of an Efficacious Functional Group Dyad. Methods and Limitations in the Use of NMR for Measuring Host-Guest Interactions. In: Schneider H-J, Dürr H, eds. *Frontiers in Supramolecular Organic Chemistry and Photochemistry*. New York, NY: VCH; 1991:123-143.
29. Vasquez VR, Whiting WB. Accounting for both random errors and systematic errors in uncertainty propagation analysis of computer models involving experimental measurements with Monte Carlo methods. *Risk Anal*. 2005;25(6):1669-1681.
30. Billeter J, Neuhold Y-M, Simon L, Puxty G, Hungerbühler K. Uncertainties and error propagation in kinetic hard-modelling of spectroscopic data. *Chemom Intell Lab Syst*. 2008;93(2):120-131.
31. Sooväli L, Rõõm E-I, Kütt A, Kaljurand I, Leito I. Uncertainty sources in UV-vis spectrophotometric measurement. *Accred Qual Assur*. 2006;11(5):246-255.
32. Caricato M, Coluccini C, Dondi D, Vander Griend DA, Pasini D. Nesting complexation of C60 with large, rigid D2 symmetrical macrocycles. *Org Biomol Chem*. 2010;8(14):3272-3280.
33. Motulsky H, Christopoulos A. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. USA: Oxford University Press; 2004.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kazmierczak NP, Chew JA, Michmerhuizen AR, et al. Sensitivity limits for determining 1:1 binding constants from spectrophotometric titrations via global analysis. *Journal of Chemometrics*. 2019;e3119. https://doi.org/10.1002/cem.3119