Monte Carlo Simulation of Error Propagation in the Determination of Binding Constants from Rectangular Hyperbolae. 1. Ligand Concentration Range and Binding Constant

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · OCTOBER 1998

Impact Factor: 2.69 · DOI: 10.1021/jp9819947

2 AUTHORS, INCLUDING:

David Chen
University of British Columbia - Vancouver
112 PUBLICATIONS **3,578** CITATIONS

SEE PROFILE

Monte Carlo Simulation of Error Propagation in the Determination of Binding Constants from Rectangular Hyperbolae. 1. Ligand Concentration Range and Binding Constant

Michael T. Bowser and David D. Y. Chen*

Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

Received: April 24, 1998

Rectangular hyperbolae have been used both to estimate equilibrium constants and to describe chemical processes dictated by equilibria. The propagation of error from the experimental measurements to the estimated constants, however, has not been well understood. In this paper, simulated experiments are used in a Monte Carlo analysis to compare the distributions of binding constants estimated by various calculation methods under different experimental conditions. The necessity of matching the range of additive (ligand) concentrations to the binding constant of the chemical interaction is demonstrated. It is shown that the relative error in the binding constant estimate is lower when the additive concentrations cover the central to upper portion of the binding isotherm (i.e., where the fraction of analyte complexed is above 0.5). The difference in the slope of the binding isotherm at the lowest and highest additive concentration used for the measurements is a good indicator of the reliability of the binding constant estimated under a specific set of conditions.

Introduction

Over the past 85 years, rectangular hyperbolae have been used to describe a variety of chemical properties influenced by molecular equilibria[1] including spectrophotometry,[2−7] NMR chemical shifts,[8−13] Michaelis−Menten kinetics,[14−21] ion transport across membranes,[22] pharmacokinetics,[23−26] and even algal growth rates.[27−29] All of these studies are based on a certain chemical response dictated by a 1:1 interaction between a substrate and a ligand. A rectangular hyperbola generally can be expressed as[1]

$$y = \frac{dx}{f + ex} \tag{1}$$

where y is the response, x is the ligand concentration, and d, e, and f are constants. One of the constants is redundant since eq 1 can be rearranged, allowing the response to be described using two constants. One constant describes the equilibrium (the binding constant or dissociation constant) while the other describes the maximum response range (i.e., the difference between the maximum and minimum response). Equation 1 can be expressed using either association constants (common in spectroscopic studies) or dissociation constants (common in kinetic and biochemical studies).

Equations taking the general form of eq 1 have been developed independently in many research areas including spectrophotometry, NMR, and Michaelis−Menten kinetics. When spectrophotometric methods are used, eq 1 is usually expressed as

$$\frac{(A - A_S)}{b} = \frac{[S_0](\epsilon_{SL} - \epsilon_S)K[L]}{1 + K[L]} \tag{2}$$

where A is the measured absorbance, A_S is the absorbance of the substrate in the absence of ligand, b is the optical path length, $[S_0]$ is the initial concentration of the substrate, ϵ_{SL} and ϵ_S are the molar absorptivities of the substrate/ligand complex and the substrate respectively, K is the binding constant, and [L] is the ligand concentration.

If NMR is used to study chemical binding, eq 1 is usually written as

$$(\delta - \delta_S) = \frac{(\delta_{SL} - \delta_S)K[L]}{1 + K[L]} \tag{3}$$

where δ is the measured chemical shift, δ_S and δ_{SL} are the chemical shifts of the substrate and the substrate/ligand complex, respectively, K is the binding constant, and [L] is the ligand concentration.

Michaelis−Menten kinetics present eq 1 as

$$v = \frac{V_m[S]}{K_m + [S]} \tag{4}$$

where v is the observed rate of the reaction, V_m is the maximum rate of the reaction achieved when the catalyst is fully complexed, [S] is the concentration of the substrate, and K_m is the dissociation constant. Equation 4 takes a slightly different form than eqs 2 and 3 because it is expressed according to the dissociation constant ($K_m = 1/K$).

Recently, it has been shown that, in capillary electrophoresis (CE), analyte mobility in the presence of analyte-additive interactions can be described according to[30−40]

$$(\nu\mu_{ep}^{A} - \mu_{ep,A}) = \frac{(\mu_{ep,AC} - \mu_{ep,A})K_{AC}[C]}{1 + K_{AC}[C]} \tag{5}$$

where μ_{ep}^{A} is the net electrophoretic mobility of the analyte, ν is a correction factor which normalizes μ_{ep}^{A} to conditions where [C] approaches zero, [C] is the concentration of the complexation additive (analogous to [L]), K_{AC} is the formation constant of the complex AC, and $\mu_{ep,AC}$ and $\mu_{ep,A}$ are the electrophoretic

* To whom correspondence should be addressed: Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada V6T 1Z1. Tel: (604) 822-0878. Fax: (604) 822-2847. E-mail: chen@chem.ubc.ca.

S1089-5639(98)01994-X CCC: $15.00
Published on Web 09/23/1998

TABLE 1: Equations Used in CE and Variances in the Transformed *y* for the Different Calculation Methods

calculation method	equation	$\sigma_{y'}^{2\,a}$
nonlinear regression	$(\nu\mu_{ep}^{A} - \mu_{ep,A}) = \dfrac{(\mu_{ep,AC} - \mu_{ep,A})K_{AC}[C]}{1 + K_{AC}[C]}$	σ_y^2
double reciprocal	$\dfrac{1}{(\nu\mu_{ep}^{A} - \mu_{ep,A})} = \dfrac{1}{(\mu_{ep,AC} - \mu_{ep,A})K_{AC}}\dfrac{1}{[C]} + \dfrac{1}{(\mu_{ep,AC} - \mu_{ep,A})}$	$\dfrac{\sigma_y^2}{(\nu\mu_{ep}^{A} - \mu_{ep,A})^4}$
y-reciprocal	$\dfrac{[C]}{(\nu\mu_{ep}^{A} - \mu_{ep,A})} = \dfrac{[C]}{(\mu_{ep,AC} - \mu_{ep,A})K_{AC}} + \dfrac{1}{(\mu_{ep,AC} - \mu_{ep,A})K_{AC}}$	$\dfrac{[C]^2\sigma_y^2}{(\nu\mu_{ep}^{A} - \mu_{ep,A})^4}$
x-reciprocal	$\dfrac{(\nu\mu_{ep}^{A} - \mu_{ep,A})}{[C]} = -K_{AC}(\nu\mu_{ep}^{A} - \mu_{ep,A}) + K_{AC}(\mu_{ep,AC} - \mu_{ep,A})$	$\left(K_{AC} + \dfrac{1}{[C]}\right)^2\sigma_y^2$

[a] $\sigma_{y'}^2$ is the variance of the transformed *y*; σ_y^2 is the variance in $(\nu\mu_{ep}^{A} - \mu_{ep,A})$; the weight for each point is equal to $1/\sigma_y^2$.

mobilities of the analyte−additive complex AC and the uncomplexed analyte, respectively. Clearly, eq 5 is also analogous to eq 1.

The constants in eqs 2 to 5 are usually estimated by measuring the response at several ligand concentrations followed by one of several regression procedures. Although a nonlinear regression can be used to solve the constants directly, eq 1 is often linearized by one of the following equations:

$$\frac{1}{y} = \frac{f}{d}\frac{1}{x} + \frac{e}{d} \tag{6}$$

$$\frac{x}{y} = \frac{e}{d}x + \frac{f}{d} \tag{7}$$

$$\frac{y}{x} = -\frac{e}{f}y + \frac{d}{f} \tag{8}$$

The constants are estimated from the slopes and the intercepts of the straight lines.[1] The use of eqs 6−8 have acquired different names in different research fields but can be referred to most generally as double reciprocal (eq 6; also referred to as Lineweaver−Burk[15] or Benesi−Hildebrand[2] plots), *y*-reciprocal (eq 7), and *x*-reciprocal (eq 8; also referred to as Eadie[16] or Scatchard[41] plots) methods. The linearizations of eq 5 are shown in Table 1.

It has long been realized that although the nonlinear regression and each of the three linearizations are based on the same equation, they often give different estimates and confidence intervals for the constants when applied to the same data set.[42−45] Linearizing eq 1 to eqs 6 to 8 invalidates some of the assumptions made in performing the least squares regression analyses, including introducing error into the independent variable and transforming the error in the data to a non-Gaussian distribution. The data spacing is also changed when eq 1 is rearranged to eqs 6 and 8, which alters the weight on certain measurements. These problems can often be overcome if the data is weighted according to the functions listed in Table 1.[1]

Because of the complexity of the regression calculations, it is difficult to show how error is propagated through the different methods analytically. Dowd and Riggs[43] first used Monte Carlo analyses to compare the different calculation methods and their estimates of the constants. Since then a number of researchers have used Monte Carlo analyses to simulate binding experiments.[46−53] It has been shown that the nonlinear regression method minimizes both the error and the bias in the estimates of the constants.

Another concern is the effect of the experimental design on the constants estimated using the different methods. The range of ligand concentrations over which responses are measured has been identified as one of the most important aspects of the experimental design.[1,44] Intuitively, to minimize the error in the estimate of *K*, [L] should cover a significant portion of the possible response range with a number of points near the ligand concentration where half of the substrate is complexed. Theories do suggest that the error in the estimate of the equilibrium constant will be minimized when responses are measured over the central portion of the binding isotherm (where the fraction of analyte complexed ranges from 0.2 to 0.8).[44] Several Monte Carlo analyses have been made to determine the effect of the ligand concentration range on the estimate of the equilibrium constant.[46,47,50] Unfortunately, these studies only tested a few conditions and only qualitative interpretations could be extracted. Not enough data were collected to truly assess the theories proposed to describe the effect of ligand concentration on the error in binding constant estimates.

In this paper, the effect of the ligand concentration on the error in the estimate of *K* is studied more thoroughly. The nonlinear regression is compared with the three linearizations of the binding isotherm. The error and bias in the binding constant estimates is compared over a wide range of conditions, including cases where the fraction of substrate complexed is below 0.2 for all [L] or above 0.8 for all [L]. The terminology developed for CE (eq 5) will be used since this is the primary research interest of our group (i.e., analyte and additive are analogous to substrate and ligand respectively). The constants and errors in the data are typical for CE. It should be emphasized that although CE is used as an example in this paper, the equations are analogous to those used in many other research areas, allowing the conclusions presented here to be applied to complexation chemistry in general.

Methods

Monte Carlo simulations of a dynamic complexation CE experiment were performed using a Visual Basic macro in Microsoft Excel 5.0 on a Pentium PC. The simulations were made assuming the following: $\mu_{ep,A} = 2.5 \times 10^{-4}$ cm^2 V^{-1} s^{-1}; $\mu_{ep,AC} = 1.0 \times 10^{-4}$ cm^2 V^{-1} s^{-1}, separation potential = 30 kV, total capillary length = 57 cm, length to the detector = 50 cm. Simulations were made for binding constants ranging from 1 to 4000 M^{-1} with a set range of additive concentrations. The additive concentrations used were 5, 23.75, 42.5, 61.25, and 80 mM. The concentration of the analyte is assumed to be much smaller than that of the additive, allowing the initial additive concentration ($[C]_0$) to be substituted for [C] in eq 5.[40] This is consistent with the other methods described by eqs 2−4. Equation 5 was used to calculate the true net analyte mobility

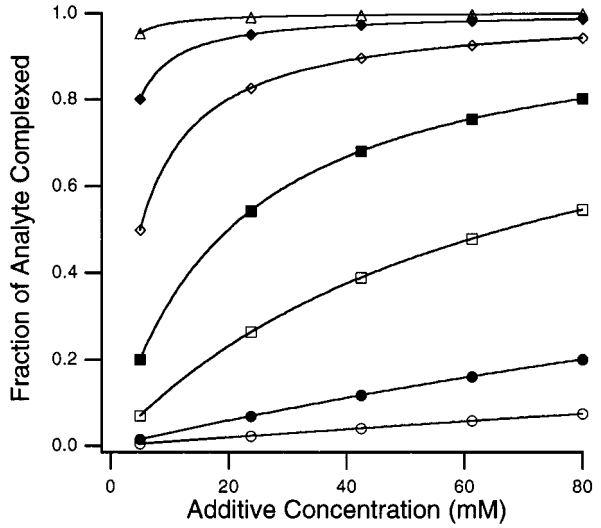

Figure 1. Binding isotherms for a set range of additive (ligand) concentrations at K = 1 M^{-1} (○), 3.125 M^{-1} (●), 15 M^{-1} (□), 50 M^{-1} (■), 200 M^{-1} (◇), 800 M^{-1} (◆), and 4000 M^{-1} (△).

at each additive concentration. The random number generator in Excel 5.0 was used to produce an experimental mobility according to a normal distribution which had a mean equal to the true mobility and a standard deviation of 8.75×10^{-7} cm^2 V^{-1} s^{-1}. Two experimental mobilities were generated for each additive concentration. Four experimental mobilities were generated for an additive concentration of 0 mM, again with a standard deviation of 8.75×10^{-7} cm^2 V^{-1} s^{-1}, to calculate $\mu_{ep,A}$. The experimental mobilities were then used to estimate the equilibrium constant according to one of the calculation methods. All regressions were made according to the least-squares variance−covariance method. This procedure was repeated 1000 times for each calculation method at each *K*. Thirty seven values for *K* were tested, giving rise to over 3.6 million simulated measurements, emphasizing the necessity of the computational approach used.

The effect of varying *K* for a set range of additive concentrations is shown in Figure 1. As *K* changes, so does the portion of the binding isotherm covered by the additive concentrations. At low *K*, the additive concentrations only cover the lower portion of the binding isotherm (i.e., the portion where the fraction of analyte complexed is low). Conversely, at high *K*, the additive concentrations only cover the upper portion of the binding isotherm (i.e., the portion where the fraction of analyte complexed is high). Therefore, varying *K* for a set group of additive concentrations is analogous to scaling the additive concentration range for a set *K*. The absolute values of *K* and the additive concentrations are not as important as the portion of the binding isotherm that is covered when both are considered. At $K = 50$ M^{-1}, the fraction of analyte complexed ranges from 0.2 to 0.8, which has been previously suggested as the ideal range of additive concentrations.[44] At $K = 3.125$ M^{-1}, the fraction of analyte complexed is equal to or below 0.2 for all of the additive concentrations used in this paper. At $K = 800$ M^{-1}, the fraction of analyte complexed is equal to or above 0.8 for all of the additive concentrations.

Results and Discussion

Error and Bias at Low Equilibrium Constants. Figure 2 depicts the distributions of the binding constants estimated using the different calculation methods for $K = 1-50$ M^{-1}. The markers represent the medians of the distributions. Medians were used because of their robustness and to minimize the influence of the grossly incorrect estimates that can occur when the additive concentrations do not cover the optimal range of the binding isotherm. The solid line indicates the true equilibrium constant. Therefore, deviation of the median from the solid line indicates bias in the calculation method. The dashed lines depict the 95% range of the distributions. The 95% range gives an indication of the reliability of the calculation method. Obviously a narrow distribution of binding constant estimates is preferable.

Comparing the different calculation methods shows that the 95% range of the nonlinear regression method is generally smaller than that of the unweighted linear methods. In all three linearizations the distributions using weighted data are narrower than the unweighted for $K = 1-50$ M^{-1}. When ideally weighted, all of the calculation methods studied gave nearly identical 95% ranges of the binding constant estimates. The weighted and unweighted nonlinear regressions are identical since no weighting is required when the errors in the data points are equal. The elimination of the weighting procedure is a major advantage of the nonlinear regression method. It should be emphasized that because the errors in this experiment were simulated, it was possible to calculate the exact weighting functions for each method. In practical applications, the error at each point is usually estimated, which may give rise to inaccurate weighting functions. The 95% ranges presented for the weighted methods therefore represent the minimum range possible with ideal weighting functions. The nonlinear regression method is the least dependent on the choice of the weighting function and should provide the most reliable binding constant estimate in practical applications.

As shown in Figure 1, at $K = 50$ M^{-1}, the fraction of analyte complexed ranges from 0.2 to 0.8. It has been previously suggested that this is the range of additive concentrations that minimizes the error in the estimate of the binding constant.[44] As the binding constant decreases from 50 M^{-1}, the fraction of analyte complexed at some of the additive concentrations drops below the 0.2 threshold. When *K* is below 3.125 M^{-1}, the fraction of analyte complexed is less than 0.2 for all of the additive concentrations used.

At low binding constants, where the fraction of analyte complexed is below 0.2, there is a significant increase in the 95% range of the *K* estimates, especially when the three linear methods are used. When the additive concentrations only cover the lower portion of the binding isotherm, it is impossible to accurately determine the value of *K*. In some cases, a substantial portion of the estimates was negative. At low *K* values, the distributions of binding constant estimates are no longer symmetrical around the median. This may cause problems when calculating the confidence intervals for the estimated binding constants.

Bias becomes significant in the three linearized methods when the binding constant is low. In some cases, the medians of the distributions were negative, indicating that the method gave more negative estimates for *K* than positive estimates. Weighting did not remove the bias from the binding constant estimates. Bias is usually a serious problem since it cannot be eliminated through replicate experiments. In binding experiments though, the amount of bias present at low *K* is small in relation to the 95% ranges of the distributions, suggesting that bias is not the most critical problem in this case. No bias was detected when the nonlinear method was used to estimate the equilibrium constant, further supporting the premise that the

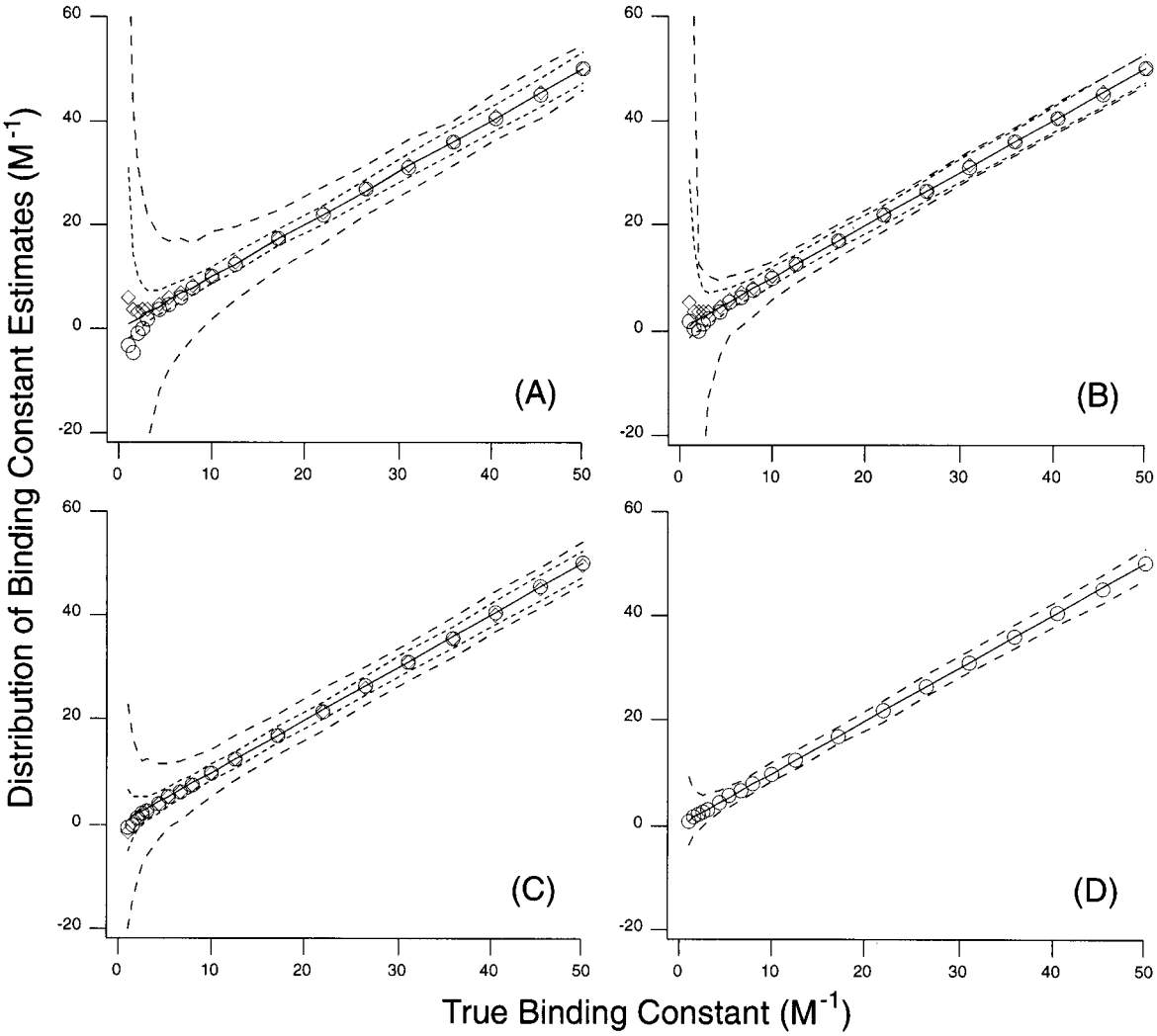

Figure 2. The distributions of the binding constant estimates calculated using (A) double reciprocal, (B) *y*-reciprocal, (C) *x*-reciprocal, and (D) nonlinear regression methods over $K = 1-50$ M^{-1}. The markers are the medians of the distributions for the weighted (○) and unweighted (◇) methods. The dashed lines define the 95% ranges of the distributions for the weighted (- -) and unweighted (− −) methods.

nonlinear regression method is more reliable than the three linearizations.

Error and Bias at High Equilibrium Constants. Figure 3 depicts the distributions of the binding constant estimates for $K = 50-4000$ M^{-1}. At $K = 50$ M^{-1} the fraction of analyte complexed ranges from 0.2 to 0.8. When *K* is higher than 800 M^{-1}, the fraction of analyte complexed is equal to or greater than 0.8 at all of the additive concentrations (e.g., the fraction of analyte complexed ranges from 0.952 to 0.997 for the additive concentrations used in this paper when $K = 4000$ M^{-1}). At higher *K* values, when the fraction of analyte complexed is above 0.8, the 95% ranges of the binding constant estimates increase. With the exception of the *y*-reciprocal plots, the ranges of the binding constant estimates for the nonlinear, double reciprocal, and *x*-reciprocal methods are nearly identical, and weighting does not seem to have any effect. The unweighted *y*-reciprocal method is significantly poorer than the other methods at higher *K* values. When ideally weighted though, the 95% ranges of the *y*-reciprocal binding constant estimates are comparable to the other methods. Again, it was possible to ideally weight the data in the simulated experiments, but would be difficult to do so in practical cases. Therefore, when the additive concentrations only cover the upper portion of the binding isotherm, it would be more reasonable to use one of the methods that are less dependent on choosing the proper weighting function (i.e., nonlinear, double reciprocal, or *x*-reciprocal).

Bias is significant in the *x*-reciprocal plot at higher *K* values, and weighting the data does not alleviate the problem. Bias is troublesome since it cannot be eliminated through replicate experiments. Therefore, the *x*-reciprocal plot is also unsuitable for estimating the binding constant in this case.

Comparing the Relative 95% Ranges at Low *K* and High *K*. Comparing the distributions reveals that the relative error in the binding constant estimates does not increase as quickly at high *K* as it does for low *K*. At low *K* the 95% range increases quickly, making it impossible to obtain accurate binding constant estimates. This does not occur at high *K*, even when the binding constant is increased to 4000 M^{-1}. Figure 4 shows that the relative 95% ranges of the *K* estimates at low binding constants are much higher that those at high *K*. The relative 95% range using the nonlinear regression method at $K = 3.125$ M^{-1} (fraction of analyte complexed ranging from 0.015 to 0.2) is approximately 14 times of that when $K = 800$ M^{-1} (fraction of analyte complexed ranging from 0.8 to 0.985), even though they both cover similar amounts of the binding isotherm (i.e., the difference in the fraction of analyte complexed at the lowest and highest additive concentration is the same for both). Even at $K = 4000$ M^{-1}, the relative 95% range is approximately one-quarter of that when $K = 3.125$ M^{-1} even though the

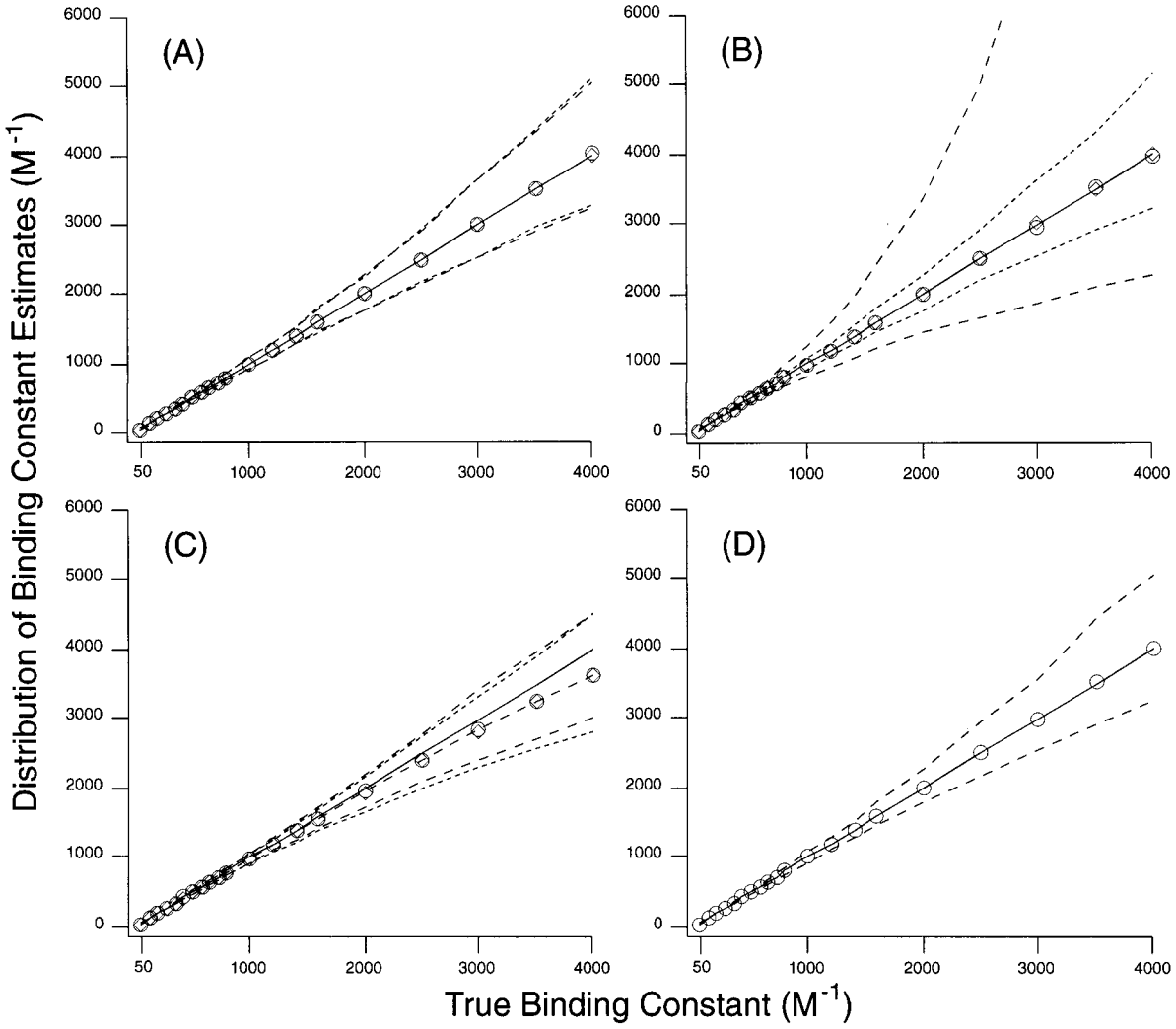

Figure 3. The distributions of the binding constant estimates calculated using (A) double reciprocal, (B) *y*-reciprocal, (C) *x*-reciprocal, and (D) nonlinear regression methods over K = 50−4000 M^{-1}. The symbols are the same as in Figure 2.

fraction of analyte complexed only ranges from 0.952 to 0.997. This contrasts what is suggested by some theories that the error contributed by the upper and lower portion of the binding isotherm should be symmetrical around the point where half of the analyte is complexed.[44]

Semiempirical Prediction of the 95% Range. An important goal of Monte Carlo simulations should be to determine the optimum conditions for actual experiments. It would be useful to predict the effect of the additive concentration range on the reliability of the estimated binding constant.

Several researchers have proposed methods to assess the reliability of binding constants estimated using a certain additive concentration range. Deranleau[44] suggested that propagation of error could be used to estimate the relative error in the calculated binding constant:

$$\frac{\Delta K}{K} \geq \Delta s \sqrt{\frac{1}{s^2} + \frac{1}{(1-s)^2}} \qquad (9)$$

where ΔK is the error in the binding constant K, s is the fraction of analyte complexed, and Δs is the error in the fraction of analyte complexed. According to eq 9, the error contributed by the lower and upper portion of the binding isotherm should be symmetrical around the point where half of the analyte is complexed, suggesting that the relative error in the estimated binding constant will be minimized when data are collected between $s = 0.2$ and 0.8. Unfortunately, this method only considers the error contributed by a single point, not a range of additive concentrations. It does not account for differences in the shape of the lower and upper portions of the binding isotherms.

Weber and Anderson[54] suggested that the information content of the measured responses should be maximized to ensure the reliability of the estimated equilibrium constant. The information content of a single measurement ($I(p)$) is given by

$$I(p) = -p \log_2 p - (1-p)\log_2(1-p) \qquad (10)$$

where p is the fraction of analyte complexed. The amount of information provided by a number of measurements can be found by summing the information content of the individual measurements. Although this approach accounts for the range of additive concentrations, the data are still considered individually, and the shape of the curve is not accounted for.

Carta et al.[55,56] proposed that reliable estimates for the equilibrium constant can be obtained when additive concentrations with a wide range of G values are tested. G represents the fraction of uncomplexed analyte and is given by

$$G = \frac{1}{K\sqrt{(a+b+K^{-1})^2 - 4ab}} \qquad (11)$$

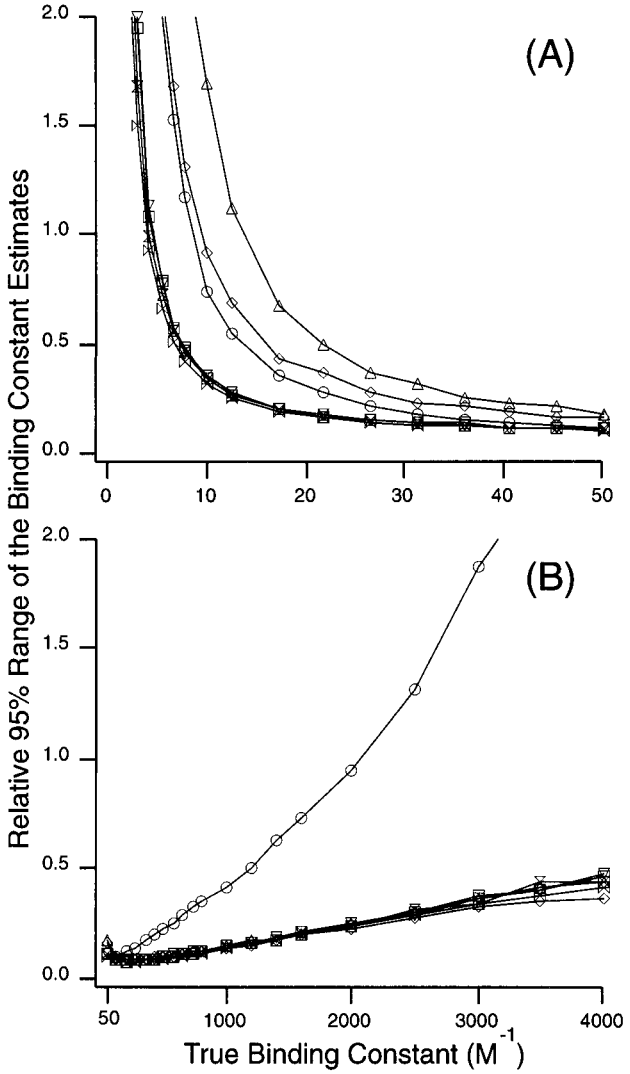

Figure 4. The relative 95% ranges of the binding constants estimated using the double reciprocal (△), weighted double reciprocal (□), *y*-reciprocal (○), weighted *y*-reciprocal (▽), *x*-reciprocal (◇), weighted *x*-reciprocal (⋈), and nonlinear regression methods (☒) for (A) $K = 1-50$ M^{-1} and (B) $K = 50-4000$ M^{-1}.

where a and b are the concentrations of the analyte (substrate) and the additive (ligand). When the concentration of the additive is much greater than that of the analyte (common in most CE binding studies), eq 11 simplifies to

$$G = \frac{1}{1 + Kb} \tag{12}$$

which is equal to the fraction of analyte uncomplexed at additive concentration b. The error in K should be minimized by maximizing

$$\sum_{i=1}^{n} (G_i - G_{\text{mean}})^2 \tag{13}$$

where G_{mean} is the average value of G and n is the number of measurements. This gives a measure of the range of the binding isotherm covered by the additive concentrations and is approximately symmetrical around the point where half of the analyte is complexed.

Unfortunately, none of the theories proposed above accurately explain the trends observed in this experiment. Figure 5 shows the dispersions of G values and the information contents of the

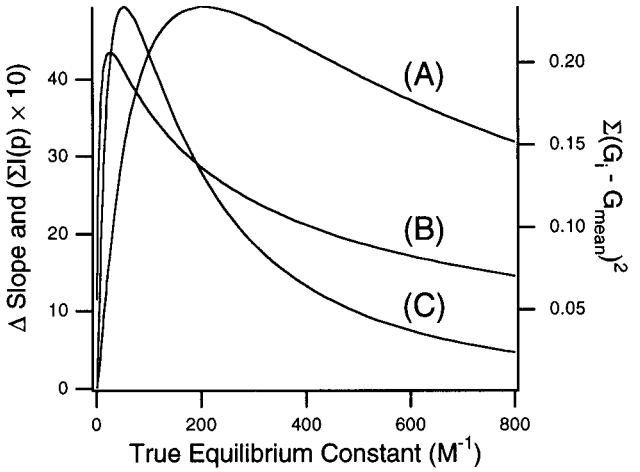

Figure 5. The (A) change in slope, (B) information content, and (C) dispersion of G for binding isotherms with $K = 1-800$ M^{-1} and a set range of additive (ligand) concentrations (5, 23.75, 42.5, 61.25, and 80 mM).

experiments simulated in this paper. The error in the estimate of K should be minimized at maximum values of $\Sigma(G_i - G_{\text{mean}})^2$ and $\Sigma I(p)$. The dispersion of G is greatest when $K = 50$ M^{-1}, where data is collected symmetrically around the point where half of the analyte is complexed. Also, the method proposed by Carta et al. suggests that there should be similar errors in the binding constant estimates at $K = 3.125$ M^{-1} and $K = 800$ M^{-1}. This does not agree with Figure 4 in which the minimum relative 95% range of the binding constants is at $K = 200$ M^{-1} and substantially narrower ranges are achieved when the additive concentrations cover the upper portion of the binding isotherm (i.e., where the fraction of analyte complexed is greater than 0.5). As shown in Figure 5B, information theory does not explain the trends in Figure 4 either, since it predicts that the error in the estimate of K will be minimized at $K = 25$ M^{-1} and that the error in K will generally be lower when the additive concentrations cover the lower portion of the binding isotherm.

Previous explanations of the effect of the additive concentration range focused on the amount of error contributed by the individual data points, but did not account for the geometric properties of the binding isotherm. In eqs 2 to 5, the equilibrium constant dictates the amount of curvature in the isotherm. Therefore, it is impossible to make an accurate estimate of K when there is no observable curvature in the isotherm. When this approach is used, it is clear that better estimates for K can be achieved when the additive concentrations cover the upper portion of the binding isotherm. Comparing the isotherms for $K = 3.125$ M^{-1} and $K = 800$ M^{-1} in Figure 1 reveals that even though both curves cover a similar amount of the binding isotherm (analyte complexed ranging from 0.015 to 0.2 for $K = 3.125$ M^{-1} and 0.8 to 0.985 for $K = 800$ M^{-1}), there is obviously more curvature in the plot for $K = 800$ M^{-1} than in that for $K = 3.125$ M^{-1}.

The amount of curvature in a binding isotherm can be estimated by finding the difference between the slope at the lowest additive concentration and that at the highest concentration. The slope at any point on the curve is equal to the first derivative of the response with respect to the additive concentration. For example, the first derivative of eq 5 is

$$\frac{\partial(\nu\mu_{\text{ep}}^{\text{A}})}{\partial[\text{C}]} = \frac{(\mu_{\text{ep,AC}} - \mu_{\text{ep,A}})K_{\text{AC}}}{(1 + K_{\text{AC}}[\text{C}])^2} \tag{14}$$

Figure 5A shows the difference in the slope of the lowest and

highest additive concentration over $K = 0-800\ M^{-1}$. The change in slope explains the trends of the relative 95% range presented in this paper remarkably well. The maximum difference in the slope occurs at approximately $K = 200\ M^{-1}$, in agreement with the minimum relative 95% range observed in Figure 4. Also, the difference in the slope is significantly larger when the data covers the upper portion of the isotherm, again in agreement with Figure 4.

Figure 6 shows the correlation between the relative 95% range of binding constant estimates vs the reciprocal of the difference in slopes, the information content, and the dispersion in *G*. There appears to be an excellent correlation between the relative 95% range and the reciprocal of the change in slope. However, closer inspection reveals two lines: one when the additive concentrations cover the lower portion of the isotherm and one when the additive concentrations cover the upper portion. The two distinct correlations indicate that the error introduced by the upper and lower portions of the binding isotherm propagate to the error in the binding constant estimates differently. The difference may be caused by the data spacing, which is not accounted for by the change in slope. When the data is evenly spaced, and the additive concentrations only cover the upper portion of the isotherm, the curvature in the plot is concentrated over several of the points, putting undue emphasis on these points. This may account for the larger relative 95% ranges of binding constant estimates than predicted by the inverse of the difference in slope. When the additive concentrations cover the lower portion of the isotherm, the curvature is more evenly spaced over the data set giving rise to a narrower range of binding constant estimates than would be predicted by the change in slope.

Regardless of the two distinct correlations, the inverse of the change in slope is a useful indicator of the significance of the estimated equilibrium constant. It correctly determines the conditions which minimize the relative 95% range of *K* estimates. Also the inverse change in slope is linearly correlated with the relative 95% range of equilibrium constant estimates over a wide range of conditions.

$1/\Sigma I(p)$ (Figure 6B) and $1/\Sigma(G_i - G_{mean})^2$ (Figure 6 C) do not correlate well with the relative 95% range of binding constant estimates. There are two distinct correlations for both the information content and dispersion in *G*. The differences in the slopes of the two correlations are much greater for the information content and the dispersion in G than that of the change in slope. Also, the correlations are not all linear. There is no well defined intersection in the correlations for the information content and the dispersion in *G*. The rounded intersections of the correlations indicate that the information content and the dispersion in G do not accurately predict the minimum relative 95% range (i.e., the minimum relative 95% range and the minimum $1/\Sigma I(p)$ and $1/\Sigma(G_i - G_{mean})^2$ do not occur at the same points), and are not suitable for predicting the optimum experimental conditions.

Correlation Coefficient

The correlation coefficient (R^2) is probably the most commonly quoted indicator of how well a data set fits a particular mathematical model, even though high R^2 values are sometimes obtained when the data does not fit the model. For this reason, visual inspections of the data plots should always be used to confirm the applicability of a model to a certain data set. Nonetheless, because it is impractical to visually inspect the many thousands of experiments simulated in this paper, the correlation coefficient was used to compare how well the quality of the data was reflected by the various calculation methods.

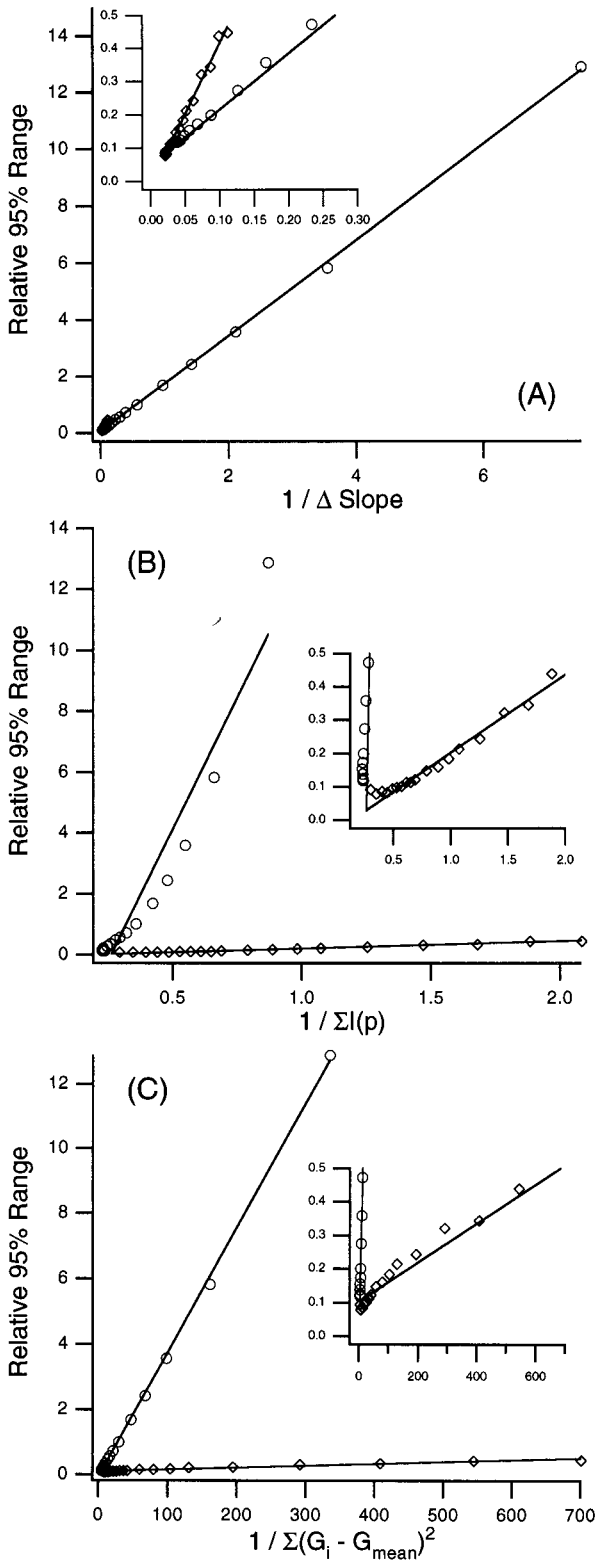

Figure 6. Correlations between the relative 95% ranges of binding constants estimated using the nonlinear regression method and (A) $(\Delta \text{ slope})^{-1}$, (B) $(\Sigma I(p))^{-1}$, and (C) $(\Sigma(G_i - G_{mean})^2)^{-1}$ for $K = 1-50\ M^{-1}$ (○) and $K = 50-4000\ M^{-1}$ (◇).

Figure 7 shows the median correlation coefficients for the different calculation methods over the range of binding constants tested. At low *K*, where the additive concentrations only cover the lower portion of the binding isotherm, the R^2 for the *x*-reciprocal and *y*-reciprocal plots decrease significantly, reflecting the unreliable nature of the *K* estimates under these conditions. Weighting the data increases R^2 for the *x*-reciprocal

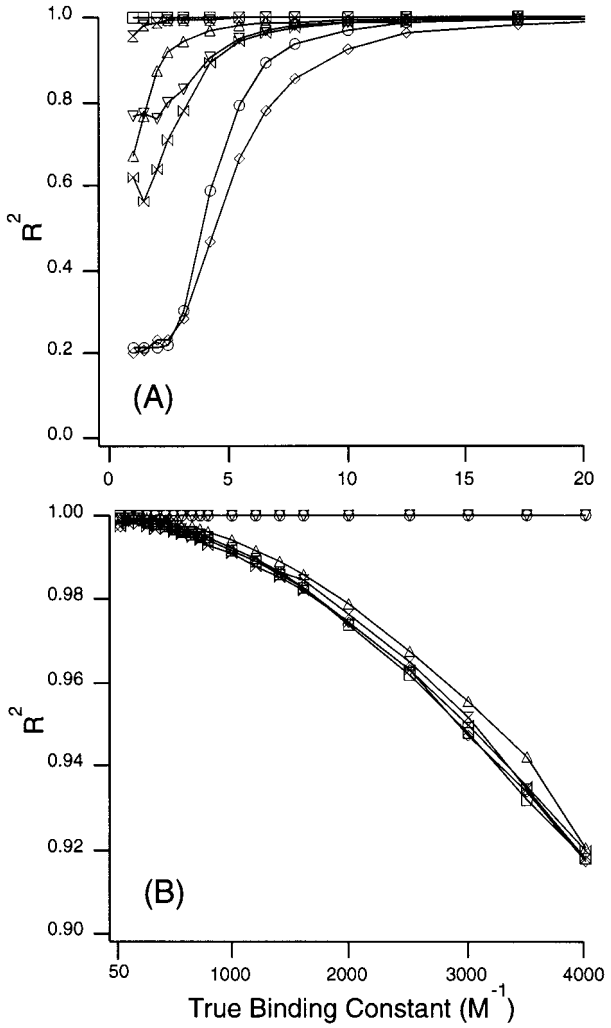

Figure 7. The median correlation coefficients (R^2) for the double reciprocal (△), weighted double reciprocal (□), *y*-reciprocal (○), weighted *y*-reciprocal (▽), *x*-reciprocal (◇), weighted *x*-reciprocal (⋈), and nonlinear regression methods (X̄) for (A) $K = 1-20$ M^{-1} and (B) $K = 50-4000$ M^{-1}.

and *y*-reciprocal plots, but this is in line with the narrower range of binding constant estimates obtained through weighting. The correlation coefficient is a very poor measure of the accuracy when double reciprocal plots are used to estimate binding constants when the additive concentrations only cover the lower portion of the binding isotherm. Very high R^2 values are obtained even when the relative 95% range of *K* estimates is high. This is misleading since it suggests that the data fits the model very well even though it is impossible to make valid estimates for *K*. The problem is more severe with weighted double reciprocal plots where the median R^2 is almost one even when $K = 1$ M^{-1}. The nonlinear fitting method also gives high R^2 values at low *K*.

The decrease in R^2 at high *K* is less pronounced than that observed at low *K*. The smaller decrease in R^2 is expected since the calculation methods generally provide better binding constant estimates when the additive concentrations cover the upper portion of the binding isotherm. With the exception of the *y*-reciprocal plots, the calculation methods all give similar R^2 at high *K*. The R^2 for the *y*-reciprocal plots is a poor indicator of the reliability of the estimated binding constant. Even at $K = 4000$ M^{-1}, the median R^2 for the *y*-reciprocal plots is almost one. Because *x*-reciprocal plots give the most realistic indication of the ability of a data set to provide reliable estimates of the binding constant at both low *K* and high *K*, this method should be used to assess the quality of the experimental data.

Conclusions

All of the calculation methods gave similar ranges of binding constant estimates when the ideal weighting functions were used. However, because it is difficult to obtain the weighting function accurately in practical applications, we recommend using the nonlinear regression method because the weighted and unweighted regression give similar results, making the estimated binding constant less dependent on the choice of the weighting function.

This paper demonstrated the importance of choosing an appropriate range of additive (ligand) concentrations when measuring a binding constant. The binding isotherm is determined by three parameters: the responses of the free and complexed analyte establish the minimum and maximum values of the isotherm, and the binding constant determines the curvature of the isotherm. In most cases the properties of the uncomplexed analyte can be accurately determined. Therefore, the accuracy of *K* is often dictated by the accuracy of the measurement of the maximum response and the curvature of the isotherm. As can be seen in Figure 1, at low *K* values, neither the maximum response nor the curvature of the isotherm can be accurately determined, leading to high uncertainty in the binding constant estimates. At high *K* values, the complex mobility can be accurately determined and there is more curvature observed in the isotherm, giving rise to more accurate estimates of the binding constant.

The optimal additive concentration range is related to the strength of the interaction to be studied. Although for this paper *K* was adjusted for a set range of additive concentrations, in practical experiments it is necessary to adjust the additive concentration range to match the binding constant. It was shown that the additive concentrations should cover the central to upper portion of the binding isotherm (i.e., where the fraction of analyte complexed is above 0.5). At higher *K*, the additive concentrations should be low so the fraction of analyte complexed falls within the ideal range. Similarly, at higher *K*, lower additive concentrations need to be used. In practical applications, a preliminary experiment should be performed to provide a rough estimate of the binding constant. The preliminary estimate of *K* can then be used to determine the range of additive concentrations which maximizes the difference in the slope of the isotherm, thereby minimizing the error in the final estimate of the binding constant.

References and Notes

(1) Connors, K. A. *Binding Constants—The Measurement of Molecular Complex Stability*; John Wiley & Sons: Toronto, 1987.

(2) Benesi, H.; Hildebrand, J. H. *J. Am. Chem. Soc.* **1949**, *71*, 2703−2707.

(3) Ramette, R. W. *J. Chem. Educ.* **1967**, *44*, 647−654.

(4) McBryde, W. A. E. *Talanta* **1974**, *21*, 979−1004.

(5) Slama-Schwok, A.; Teulade-Fichou, M. P.; Vigneron, J. P.; Taillandier, E.; Lehn, J. M. *J. Am. Chem. Soc.* **1995**, *117*, 6822−6830.

(6) Lahiri, J.; Fate, G. D.; Ungashe, S. B.; Groves, J. T. *J. Am. Chem. Soc.* **1996**, *118*, 2347−2358.

(7) Grabner, G.; Monti, S.; Marconi, G.; Mayer, B.; Klein, C.; Köhler, G. *J. Phys. Chem.* **1996**, *100*, 20068−20075.

(8) Mathur, R.; Becker, E. D.; Bradley, R. B.; Li, N. C. *J. Phys. Chem.* **1963**, *67*, 2190−2194.

(9) Hanna, M. W.; Ashbaugh, A. L. *J. Phys. Chem.* **1964**, *68*, 811−816.

(10) Carper, W. B.; Buess, C. M.; Hipp, G. R. *J. Phys. Chem.* **1970**, *74*, 4229−4234.

(11) Seal, B. K.; Mukherjee, A. K.; Mukherjee, D. C.; Farrel, P. G.; Westwood, J. V. *J. Magn. Reson.* **1983**, *51*, 318−322.

(12) Blixt, J.; Detellier, C. *J. Am. Chem. Soc.* **1995**, *117*, 8536−8540.
(13) Godínez, L. A.; Patel, S.; Criss, C. M.; Kaifer, A. E. *J. Phys. Chem.* **1995**, *99*, 17449−17455.
(14) Michaelis, L.; Menten, M. L. *Biochem. Z.* **1913**, *49*, 333−369.
(15) Lineweaver, H.; Burk, D. *J. Am. Chem. Soc.* **1934**, *56*, 658−666.
(16) Eadie, G. S. *J. Biol. Chem.* **1942**, *146*, 85−93.
(17) Coolen, H. K. A.; Meeuwis, J. A. M.; van Leeuwen, P. W. N. M.; Nolte, R. J. M. *J. Am. Chem. Soc.* **1995**, *117*, 11906−11913.
(18) Sun, S.; Duggleby, R. G.; Schowen, R. L. *J. Am. Chem. Soc.* **1995**, *117*, 7317−7322.
(19) Corey, E. J.; Noe, M. C. *J. Am. Chem. Soc.* **1996**, *118*, 319−329.
(20) Lamm, G.; Wong, L.; Pack, G. R. *J. Am. Chem. Soc.* **1996**, *118*, 3325−3331.
(21) Guedes da Silva, M. F. C.; Silva, J. A. L.; Fraústo da Silva, J. J. R.; Pombeiro, A. J. L.; Amatore, C.; Verpeaux, J.-N. *J. Am. Chem. Soc.* **1996**, *118*, 7568−7573.
(22) Aronson, J. K. *Clin. Sci.* **1990**, *78*, 247−254.
(23) Hoang, K.-C. T. *Toxicol. Lett.* **1995**, *79*, 99−106.
(24) Chen, T.-L.; Kennedy, M. J.; Anderson, L. W.; Kiraly, S. B.; Black, K. C.; Colvin, O. M.; Grochow, L. B. *Drug Metab. Dispos.* **1997**, *25*, 544−551.
(25) Kwon, Y.; Morris, M. E. *Pharm. Res.* **1997**, *14*, 774−779.
(26) Kwon, Y.; Morris, M. E. *Pharm. Res.* **1997**, *14*, 780−785.
(27) MacIsaac, J. J.; Dugdale, R. C. *Deep Sea Res.* **1969**, *16*, 45−57.
(28) Mullin, M. M.; Stewart, E. F.; Fuglister, F. J. *Limnol. Oceanogr.* **1975**, *20*, 259−262.
(29) Verity, P. G. *Limnol. Oceanogr.* **1991**, *36*, 729−750.
(30) Wren, S. A. C.; Rowe, R. C. *J. Chromatogr.* **1992**, *603*, 235−241.
(31) Wren, S. A. C.; Rowe, R. C. *J. Chromatogr.* **1992**, *609*, 363−367.
(32) Wren, S. A. C. *J. Chromatogr.* **1993**, *636*, 57−62.
(33) Shibukawa, A.; Lloyd, D. K.; Wainer, I. W. *Chromatographia* **1993**, *35*, 419−429.
(34) Penn, S. G.; Goodall, D. M.; Loran, J. S. *J. Chromatogr.* **1993**, *636*, 149−52.
(35) Penn, S. G.; Bergstrom, E. T.; Goodall, D. M.; Loran, J. S. *Anal. Chem.* **1994**, *66*, 2866−73.
(36) Penn, S. G.; Bergstrom, E. T.; Knights, I.; Liu, G.; Ruddick, A.; Goodall, D. M. *J. Phys. Chem.* **1995**, *99*, 3875−3880.
(37) Rundlett, K. L.; Armstrong, D. W. *J. Chromatogr.* **1996**, *721*, 173−186.
(38) Bowser, M. T.; Sternberg, E. D.; Chen, D. D. Y. *Electrophoresis* **1997**, *18*, 82−91.
(39) Peng, X.; Bowser, M. T.; Britz-McKibbin, P.; Bebault, G. M.; Morris, J.; Chen, D. D. Y. *Electrophoresis* **1997**, *18*, 706−716.
(40) Peng, X.; Bebault, G. M.; Sacks, S. L.; Chen, D. D. Y. *Can. J. Chem.* **1997**, *75*, 507−517.
(41) Scatchard, G. *Ann. N. Y. Acad. Sci.* **1949**, *51*, 660−672.
(42) Dixon, M. *Nature* **1959**, *184*, 296−298.
(43) Dowd, J. E.; Riggs, D. S. *J. Biol. Chem.* **1965**, *240*, 863−869.
(44) Deranleau, D. A. *J. Am. Chem. Soc.* **1969**, *91*, 4044−4049.
(45) Atkins, G. L.; Nimmo, I. A. *Anal. Biochem.* **1980**, *104*, 1−9.
(46) Ritchie, R. J.; Prvan, T. *J. Theor. Biol.* **1996**, *178*, 239−254.
(47) Berges, J. A.; Montagnes, D. J. S.; Hurd, C. L.; Harrison, P. L. *Mar. Ecol. Prog. Ser.* **1994**, *114*, 175−183.
(48) Sagnella, G. A. *Clin. Sci.* **1994**, *87*, 371−381.
(49) Tseng, S.; Hsu, J.-P. *J. Theor. Biol.* **1990**, *145*, 457−464.
(50) Currie, D. J. *Biometrics* **1982**, *38*, 907−919.
(51) Atkins, G. L.; Nimmo, I. A. *Biochem. J.* **1975**, *149*, 775−777.
(52) Madsen, B. W.; Robertson, J. S. *J. Pharm. Pharmac.* **1974**, *26*, 807−813.
(53) Cornish-Bowden, A.; Eisenthal, R. *Biochem. J.* **1974**, *139*, 721−730.
(54) Weber, G.; Anderson, S. R. *Biochem.* **1965**, *4*, 1942−1947.
(55) Carta, G.; Crisponi, G.; Nurchi, V. *Tetrahedron* **1981**, *37*, 2115−2119.
(56) Carta, G.; Crisponi, G. *J. C. S. Perkin II* **1982**, 53−58.