

---

---

# Modeling Data From Titration, Amide H/D Exchange, and Mass Spectrometry to Obtain Protein-Ligand Binding Constants

Mei M. Zhu, Don L. Rempel, and Michael L. Gross

Department of Chemistry, Washington University, St. Louis, Missouri, USA

---

We recently reported a new method for quantification of protein–ligand interaction by mass spectrometry, titration and H/D exchange (PLIMSTEX) for determining the binding stoichiometry and affinity of a wide range of protein–ligand interactions. Here we describe the method for analyzing the PLIMSTEX titration curves and evaluate the effect of various models on the precision and accuracy for determining binding constants using H/D exchange and a titration. The titration data were fitted using a 1:*n* protein:ligand sequential binding model, where *n* is the number of binding sites for the same ligand. An ordinary differential equation was used for the first time in calculating the free ligand concentration from the total ligand concentration. A nonlinear least squares regression method was applied to minimize the error between the calculated and the experimentally measured deuterium shift by varying the unknown parameters. A resampling method and second-order statistics were used to evaluate the uncertainties of the fitting parameters. The interaction of intestinal fatty-acid-binding protein (IFABP) with a fatty-acid carboxylate and that of calmodulin with Ca<sup>2+</sup> are used as two tests. The modeling process described here not only is a new tool for analyzing H/D exchange data acquired by ESI-MS, but also possesses novel aspects in modeling experimental titration data to determine the affinity of ligand binding. (J Am Soc Mass Spectrom 2004, 15, 388–397) © 2004 American Society for Mass Spectrometry

---

**H**ydrogen/deuterium exchange and mass spectrometry are a promising combination to probe protein–ligand interactions [1]. In most of the research on H/D exchange of protein–ligand complexes, the goal is to measure the differences in H/D exchange kinetics for the apo protein and protein–ligand complexes. The information level may be increased to the peptide or even amino-acid level by digesting the protein before analysis and using MS/MS [2, 3]. Nonlinear least squares regression [2, 4, 5], maximum entropy [6], and graphical methods [7] are often used for extracting average rate constants and estimating the number of hydrogens in each rate group.

There are only a few examples for quantitative analysis of thermodynamic properties of protein–ligand binding in solution using H/D exchange and MS. Fitzgerald and coworkers [8] demonstrated a MALDI/MS-based stability determination method (SUPREX) that yields the  $K_d$  of a protein–ligand complex. We described a complementary means [9] to quantify protein–ligand interactions in solution by mass spectrometry, titration and H/D exchange (PLIMSTEX). We demonstrated in that article that PLIMSTEX can be

applied to determine the conformational change, binding stoichiometry, and affinity of a wide range of protein–ligand interactions including those that involve small molecules, metal ions, and peptides.

Here we describe the method of modeling PLIMSTEX titration curves and examine the effect of model modifications on the precision and accuracy that can be achieved when determining binding constants. A preliminary description of the modeling accompanied the first description of PLIMSTEX [9] as supplemental material, but since that time, we extended the model with more parameters, improved it by adding resampling, and tested it in a more complete manner. The interaction of intestinal fatty-acid-binding protein (IFABP) with a fatty-acid carboxylate and that of calmodulin with Ca<sup>2+</sup> are used as two test systems to demonstrate the H/D exchange titration method and the protein–ligand fractional species model. The modeling process described here not only is a new tool for analyzing H/D exchange data acquired by ESI-MS, but also possesses some novel aspects in modeling experimental titration data to determine the affinity of ligand binding.

The basis for the modeling is nonlinear least squares (NLLS). Among all the curve fitting methods used in the biochemical literature, NLLS regression is probably the most common [10, 11]. The “least squares” method was introduced by Legendre in 1805, but the first

---

Published online February 7, 2004

Address reprint requests to Dr. M. L. Gross, Department of Chemistry, Washington University, No. 1 Brookings Drive, Box 1134, St. Louis, MO 63130, USA. E-mail: mgross@wuchem.wustl.edu

general-purpose program for using nonlinear least squares was written in 1958 by Booth and Peterson, as described on page 7 of [10]. NLLS has been widely used for the analysis of titration experiments that probe ligand-binding, and one of the first chemical applications was to a potentiometric titration in 1968 [12]. Leggett and coworkers [13] demonstrated a computational approach to the spectrophotometric determination of stability constants using the Gauss-Newton nonlinear least squares method. The reader who is interested in the various issues on the proper use of NLLS in biochemical research are referred to a review by Johnson and coworkers [14–18].

## Experimental

### *H/D Exchange Experiment*

The experiment was begun by allowing the protein to equilibrate with different concentrations of ligand in aqueous buffer solutions. D<sub>2</sub>O containing the same concentrations of buffer and salts as in the starting solution was added to initiate H/D exchange. The protocol utilized a high D/H ratio in the forward and a high H/D ratio in the back-exchange, and carried the added advantage of in-situ desalting. When the system reached a near steady state (1 to 3 h of exchange) where the fast exchangeable hydrogens had reached equilibrium while the slow exchangers had not, as had been determined by a kinetic study conducted previously, the exchange was quenched by adding cold 1-M HCl to decrease the pH to 2.5. The solution was then loaded on a small C18 column, which is described below, cooled to 0 °C, and the labile, non-amide sites of the immobilized protein were back-exchanged to the H form. The solution was desalted by washing with ice-cold, aqueous formic acid (pH 2.5). The protein, which now bears an isotopic exchange “signature” in its amide linkages reflecting its state in the initial solution, was then introduced into a mass spectrometer, and its molecular weight was determined. Rapid elution (by an isocratic flow of solvent at 30–35  $\mu$ L/min with high organic composition or with a fast, pH 2.5 gradient) delivered the protein to an electrospray ionization (ESI) source.

### *LC/ESI-MS with an Ion-Trap Mass Spectrometer*

ESI mass spectra of IFABP were obtained with a Finnigan Classic LCQ ion-trap mass spectrometer (San Jose, CA) equipped with an ESI source. The sample was loaded on an Opti-Guard C18 column (10 mm  $\times$  1 mm i.d.; Cobert Associates) using a six-port Rheodyne 7725 manual sample injector (P.J. Cobert Associates, St. Louis, MO). After desalting by syringe injection of the wash solution in the “load” mode, the injector was switched to the “inject” mode, and the protein was rapidly eluted to the mass spectrometer by using a fast gradient: Mobile phase B increased from 30 to 100% in 1 min, and then held at 100% for 5 min. Solvent A

consisted of 3.0:96.7:0.3 (vol:vol:vol) CH<sub>3</sub>CN:H<sub>2</sub>O:HCOOH; Solvent B consisted of 80:19:1 (vol:vol:vol) CH<sub>3</sub>CN:H<sub>2</sub>O:HCOOH. Separation by HPLC was carried out using a Waters 600MS pump (Milford, MA). The flow, divided with a splitter (LC-Packings, San Francisco, CA) at 10:1, was introduced to the column and then the mass spectrometer at a flow rate of 33  $\mu$ L/min. The spray voltage was 5.0 kV, and the capillary temperature was 200 °C. All data were acquired in the positive-ion mode at unit mass resolving power for ions of  $m/z$  between 600 and 2000. Normally ten scans were averaged and processed using Finnigan Xcalibur 1.1 software. The raw data were transformed by using Bioworks software supplied by Finnigan to give mass centroids of the “deconvoluted” peaks, which were used for calculating the deuterium shift.

### *LC-ESI/MS with a Q-TOF Mass Spectrometer*

To obtain better mass resolving power and more accurate mass assignment, a Micromass Q-ToF Ultima GLOBAL mass spectrometer (Manchester, UK) with a Z-spray ESI source was used for the Ca-titrations of calmodulin. A solution of 80:19:1 (vol:vol:vol) of CH<sub>3</sub>CN:H<sub>2</sub>O:HCOOH was introduced into an Opti-Guard C18 column (10 mm  $\times$  1 mm i.d.; Cobert Associates), and then into the mass spectrometer at a flow rate of 32  $\mu$ L/min using a Waters Cap-LC system (Milford, MA). The capillary voltage was 3.4 kV, and the source and desolvation temperatures were 80 °C and 150 °C, respectively. The cone-gas flow was 40 L/h, and the nitrogen desolvation gas flow was 400 L/h. All data were acquired in the positive-ion mode, using the TOF at 10,000 mass resolving power between  $m/z$  800 and 2000. Normally ten scans were summed and processed by using MassLynx 3.5 software that was supplied with the mass spectrometer system by Micromass. The molecular mass was determined using the maximum entropy algorithm (MaxEnt1 in MassLynx 3.5).

## Modeling Procedure

### *Overview of the Modeling Procedure*

PLIMSTEX is a new method developed in our laboratory for the quantitative analysis of protein–ligand binding [9]. A plot of the mass difference between the deuterated and non-deuterated protein (deuterium uptake) versus the total ligand concentration (or the ratio of total ligand concentration to the total protein concentration) gives the PLIMSTEX curve (Figure 1).

To quantify affinity, PLIMSTEX requires that a change occur in the extent of H/D exchange during a titration. This extent is given by  $\Delta D$ , the overall deuterium shift. The change may be a conformational change and/or stability difference between the apo- and ligand-bound protein. The titration data are fitted using a 1: $n$  protein:ligand sequential binding model, where  $n$  is the number of binding sites for the same ligand.

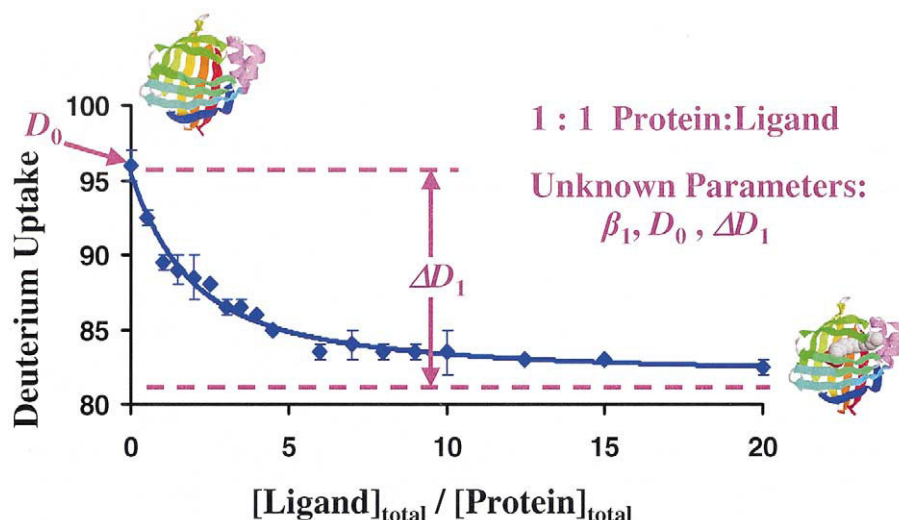


Figure 1. A typical PLIMSTEX curve for 1 to 1 protein:ligand binding.

There are two assumptions for the modeling: (1) The ligand binds to the protein in a stepwise fashion; (2) H/D exchange of each amide hydrogen is independent (i.e., does not depend on exchange at any other site of the protein). A nonlinear least squares (NLLS) regression is involved, and all modeling procedures were implemented with Mathcad 2001 Professional (MathSoft, Inc., Cambridge, MA).

For modeling the titration data, the  $\Delta D$  is calculated as a function of the total ligand concentration ( $[\text{Lig}_T]$ ), the overall binding constants ( $\beta_i$ , which is the product of the stepwise macroscopic binding constants  $K_i$ ,  $i = 1$  to  $n$ ) and the species deuterium shifts ( $D_0$  and  $\Delta D_i$ ,  $i = 1$  to  $n$ ).  $D_0$  is the shift in the molecular weight of the apo protein caused by H/D exchange (deuterium uptake). To minimize experimental errors, we do not accept the experiment value of  $D_0$  (the deuterium uptake of the apo-protein) but rather take it as a variable or unknown parameter.  $\Delta D_i$  is the difference between the average deuterium level of the complex and that of the apo-form (see Figure 1). A positive  $\Delta D_i$  indicates that binding of  $i$  ligand(s) to the protein leads to more protection and less  $D$  uptake as compared to the apo-form. A negative  $\Delta D_i$  points to the formation of a more open structure relative to its apo form. When  $\Delta D_i \approx 0$ , little conformational change occurs upon binding. In the latter case, PLIMSTEX may not be appropriate for determining the corresponding equilibrium constant ( $\beta_i$ ).

The best fit is obtained by a search, which iterates through a sequence of trials to minimize the error between the calculated overall deuterium shift,  $\Delta D$ , and the experimentally measured shift by varying the unknown parameters ( $\beta_i$ ,  $D_0$ , and  $\Delta D_i$ ). Normally at least two runs under similar conditions are performed for each set of PLIMSTEX data. The average data are used for the curve fitting to give mean values for the unknown parameters ( $\beta_i$ ,  $\Delta D_i$ , and  $D_0$ ). The macroscopic  $K_i$ 's are calculated from  $\beta_i$ 's. Finally, a resampling

statistical analysis is used to evaluate the precision for each parameter in the search.

#### Modeling the Titration Data

The goal for the modeling is to obtain the best-fit curve for the experimental titration data and to extract the values of the unknown parameters ( $\beta_i$ ,  $\Delta D_i$ , and  $D_0$ ) from the best fit. To fit the data, the model must build a relationship between the overall deuterium shift and the total ligand concentration. It then calculates the expected overall deuterium shift,  $\Delta D$ , as a function of the total ligand concentration,  $[\text{Lig}_T]$ , from a sum of the deuterium shifts for each species (eq 1):

$$\Delta D(\beta_1, \dots, \beta_n, D_0, \Delta D_1, \dots, \Delta D_n, [\text{Lig}_T]) = D_0 - \sum_{i=1}^n \Delta D_i \frac{[\text{Prot-Lig}_i]}{[\text{Prot}_T]} \quad (1)$$

In eq 1,  $[\text{Prot}_T]$  is the total experimental concentration of the protein, a known value for a specific titration, and  $D_0$ ,  $\Delta D_i$ , and  $\beta_i$  are the unknown parameters. We expect that each form of the protein/ligand complex would show a deuterium shift  $\Delta D_i$  ( $i = 1$  to  $n$ ), which is the difference between the average deuterium level of the complex and that of the apo-form.  $\Delta D_i$  is weighted by its binding fraction  $[\text{Prot-Lig}_i]/[\text{Prot}_T]$ , which is a function of total ligand concentration  $[\text{Lig}_T]$  and the overall binding constant  $\beta_i$  ( $i = 1$  to  $n$ ), the latter of which is sometimes referred to as an Adair constant [19].  $\beta_i$  is the product of all the stepwise equilibrium binding constants, and  $K_i$  is the equilibrium constant for the  $i$ th step forming a protein–ligand species (eq 2).

$$\beta_i = K_1 K_2 \cdots K_i \quad (2)$$

To predict the overall deuterium shift ( $\Delta D$ ), the

binding fractions must be calculated. The binding fraction (eq 3) for each protein–ligand species ( $[\text{Prot-Lig}_i]/[\text{Prot}_T]$ ) is calculated for each sequential reaction as governed by the binding constants:

$$\frac{[\text{Prot-Lig}_i]}{[\text{Prot}_T]} = \frac{\beta_i[\text{Lig}]^i}{P} \quad (3)$$

Each fractional solution concentration ( $[\text{Prot-Lig}_i]$ ) is a function of free  $[\text{Lig}]$ ,  $[\text{Prot}_T]$ , and  $\beta_i$  ( $i = 1$  to  $n$ ). The binding polynomial is given by eq 4:

$$P = 1 + \sum_{i=1}^n \beta_i[\text{Lig}]^i \quad (4)$$

The eqs 3 and 4 have been commonly used in modeling ligand-binding reactions (see Chapter 3 of reference [19]).

The calculation of the binding fractions requires that the free ligand concentration,  $[\text{Lig}]$ , in solution be calculated first. Solving for the free ligand concentration  $[\text{Lig}]$  as a function of the total ligand concentration  $[\text{Lig}_T]$  is the most challenging part of the modeling procedure because the solution is difficult to achieve by direct algebraic methods. The problem is made more complex because the relation between the equilibrium concentration of the ligand,  $[\text{Lig}]$ , in solution and the mass spectrometric signal corresponding to a gas-phase ion is not understood, and the relation between MS signals and solution concentrations are likely to be different for ligand, protein, and complex. Nevertheless,  $[\text{Lig}]$  can be inferred from the total ligand concentration by numerically solving the ordinary differential equation [20, 21] formed by the analytical derivative of  $[\text{Lig}]$  with respect to  $[\text{Lig}_T]$  (eq 5):

$$\begin{aligned} \frac{d[\text{Lig}]}{d[\text{Lig}_T]}(\beta_1, \dots, \beta_n, [\text{Prot}_T], [\text{Lig}_T]) \\ = f([\text{Lig}], \beta_1, \dots, \beta_n, [\text{Prot}_T], [\text{Lig}_T]) \end{aligned} \quad (5)$$

The right hand side of eq 5 is not possible to write directly; as a result, it is evaluated as the reciprocal of  $d[\text{Lig}_T]/d[\text{Lig}]$  (i.e.,  $1/f$ ). The expression  $d[\text{Lig}_T]/d[\text{Lig}]$  was obtained from the closed form expression of  $[\text{Lig}_T]$ , which is a function of  $[\text{Lig}]$ ,  $[\text{Prot}_T]$ , and  $\beta_i$ . Evaluation of the reciprocal of  $d[\text{Lig}_T]/d[\text{Lig}]$  leads to eq 6.

$$\begin{aligned} f([\text{Lig}], \beta_1, \dots, \beta_n, [\text{Prot}_T], [\text{Lig}_T]) \\ = \frac{\sum_{i=0}^n \beta_i[\text{Lig}]^i}{\sum_{i=0}^n (i+1)\beta_i[\text{Lig}]^i + \text{Prot}_T \left( \sum_{i=1}^n i^2\beta_i[\text{Lig}]^{i-1} \right) - \text{Lig}_T \left( \sum_{i=1}^n i\beta_i[\text{Lig}]^{i-1} \right)} \end{aligned} \quad (6)$$

For 1:1 protein–ligand binding,  $\beta_1$  is the same as  $K$ . The two equations (eqs 5 and 6) can be simplified to give eq 7 and 8:

$$\frac{d[\text{Lig}]}{d[\text{Lig}_T]}(K, [\text{Prot}_T], [\text{Lig}_T]) = f([\text{Lig}], K, [\text{Prot}_T], [\text{Lig}_T]) \quad (7)$$

$$\begin{aligned} f([\text{Lig}], K, [\text{Prot}_T], [\text{Lig}_T]) \\ = \frac{1 + K[\text{Lig}]}{1 + 2K[\text{Lig}] + K[\text{Prot}_T] - K[\text{Lig}_T]} \end{aligned} \quad (8)$$

The differential eq 5 was solved by integration using the “Rkadapt” function in Mathcad. For the initial condition,  $[\text{Lig}]$  was taken to be zero when  $[\text{Lig}_T]$  is zero. The integration produces a lookup table for  $[\text{Lig}]$  as a function of  $[\text{Lig}_T]$ ; we used a table with a constant step size of 1000 intervals over the ligand range in the titration. Each  $[\text{Lig}_T]$  has a matching value in the lookup table. One integration was performed for each trial. This approach appears to be the first reported use of an ordinary differential equation to calculate  $[\text{Lig}]$  from  $[\text{Lig}_T]$ , and may be of use to those who model titrations by using other methods.

After a relationship between the overall deuterium shift and the total ligand concentration is built by means of the procedures above, the selected unknown species deuterium shifts ( $D_0$  and  $\Delta D_i$ ) and  $\beta_i$  are solved in a sequence of trials. In each trial, the search algorithm postulates a set of binding constants ( $\beta_1, \dots, \beta_n$ ), deuterium mass shifts ( $\Delta D_1, \dots, \Delta D_n$ ) for each protein/ligand complex, and a deuterium uptake for the apo protein ( $D_0$ ), and then computes the expected overall deuterium shift ( $\Delta D$ ) as a function of  $[\text{Lig}_T]$ . The residuals of the trial are the differences between the experimentally measured overall mass shifts ( $A_{j,i}$ ) and the expected overall mass shifts ( $\Delta D$ ) at each  $[\text{Lig}_T]$ . If more than one set of titration data is available, then the measured mass shifts are averaged at each experimental  $[\text{Lig}_T]$  ( $A_{j,0}$ ), and the averaged mass shifts are used for the measured mass shifts in the calculation of the residuals. The error,  $\varepsilon$ , of the trial was measured as the square root of the mean of the squares (RMS) of the residuals (eq 9), where  $M$  is the number of data points in each titration:

$$\varepsilon = \sqrt{\frac{\sum_{j=0}^{M-1} (A_{j,1} - \Delta D(\beta_1, \dots, \beta_n, D_0, \Delta D_1, \dots, \Delta D_n, A_{j,0}))^2}{M}} \quad (9)$$

The search, which minimizes the trial error, is a standard nonlinear, least squares regression procedure, to which the normal assumptions required for a NLLS regression [14] apply. For example, the experimental

uncertainty for determining the difference between the mass centroids of the peaks corresponding to the deuterated and non-deuterated protein is assumed to follow a Gaussian distribution with a mean of zero. The regression is conducted with the “Minimize” function of Mathcad. The starting values for the search are guesses often informed by inspection of the titration curve, prior experience, or values derived from the literature. The lower the RMS, the better is the fit.

### *Evaluating the Precision for the Unknown Parameters*

Once a best-fit PLIMSTEX curve is obtained from the modeling, the values of the unknown parameters ( $D_0$ ,  $\Delta D_i$ , and  $\beta_i$ ) and the deviation between the experimental data and the fitted curve (RMS) are reported. In most circumstances, multiple independent runs under similar experimental conditions are performed for each protein–ligand titration. A resampling statistical method is used to evaluate the precision of the unknown parameters. The titration data are resampled by randomly selecting data points from the multiple trials at each [Lig]<sub>T</sub>. For example, if a *duplicate* titration is performed with 15 different [Lig]<sub>T</sub> values in each trial,  $2^{15}$  possible combinations of titration can be re-sampled. Second-order statistics [10] of the unknown parameters are computed as described in the following.

Each replicate of the titration experiment, if separately fitted as described in the titration modeling, would produce slightly different answers for the unknown parameters because there are measurement uncertainties. Rather than producing a large number of experiment replicates, the residuals from a few replicates are resampled, and their effect on the fitted unknown parameters are estimated to first order.

The evaluation procedure has three parts. In the **first part**, the changes in free ligand, [Lig], as a function of changes in  $\beta_i$  are computed. Because the free ligand concentration as a function of [Lig]<sub>T</sub> was computed as a solution to an ordinary differential equation, one should expect that the first and second derivatives of [Lig] with respect to  $\beta_i$  be also obtained as solutions to ordinary differential equations. These are the “sensitivity equations” and were described in Section 8-4b of reference [10]. In the present development, the sensitivity equations make use of the closed-form expressions of the first and second derivatives of  $d[\text{Lig}]/d[\text{Lig}]_T$ , which is  $f$  in eqs 5 and 6, with respect to [Lig] and  $\beta_i$ . The sensitivity equations are combined with the differential equation described in the modeling to form a system of ordinary differential equations that are solved simultaneously by integration to give a lookup table as before. The values for the unknown parameters obtained from the error minimization step in the modeling are used during the integration.

In the **second part of the evaluation**, we obtain a matrix that gives the derivative of the fitted parameters

with respect to the measured mass shifts at the various titration points. The calculation is facilitated by differentiating the “normal equations” (see Section 24.1 of reference [22] with respect to the unknown parameters and the experimentally measured mass shifts at the various titration points. The normal equations make use of first derivatives of the trial error function (eq 9) from the modeling. Differentiation of the normal equations, then, uses the second derivatives of the trial error function. This is where the first and second derivatives computed in part one of the evaluation are utilized.

In the **third part**, residuals produced by the replicate experiments are resampled to obtain second-order statistics for the unknown parameter estimates. Each experiment replicate produces a vector of residuals. Each vector coordinate value is a residual between the expected mass shift for the fitted unknown parameter values found in the modeling and the replicate measured mass shift at the corresponding total-ligand concentration of the titration. A random vector generator produces a uniformly distributed list of indices that match the titration points in number. Each index in the list corresponds to a titration point and specifies from which replicate residual vector to extract a coordinate value for the corresponding titration point. Each random vector of indices specifies another mix of residuals to form a new vector of residuals. The first-order estimate of the changes in the unknown parameter is given by product of the matrix computed in the second part of this component and the new vector of residuals. For a titration performed as  $X$  replicates with  $Y$  points in each run, this process is repeated approximately  $X^Y$  times, from which the second-order statistics are tabulated.

## Results and Discussion

### *Method for Data Analysis*

To obtain PLIMSTEX data, we sample various points in a titration of a protein with a ligand by measuring the extent of H/D exchange of the protein. We quench the H/D exchange, desalt the solution, and determine the molecular weight of the protein by admitting it to the mass spectrometer. With quenching and desalting, the ligands usually dissociate from the protein, allowing us to determine for the newly freed protein the number of deuteriums introduced on solvent-accessible amide nitrogens. This is simply the mass difference between the deuterated and non-deuterated protein (deuterium uptake), which is plotted vs. the ratio of total ligand to total protein concentration. We then fit the PLIMSTEX curve by using the model described above.

There are several new aspects of the model. It is the first model of the relationship between mass shift in H/D exchange and the binding constant of protein–ligand interaction. The most challenging step in the model is the extraction of a free ligand concentration in solution from the total ligand concentration. This step

**Table 1.** Binding properties obtained from PLIMSTEX analysis for IFABP–Oleate interaction

Data Parameter	Run 1 <sup>a</sup>	Run 2 <sup>b</sup>	Ave & Dev. for Run 1 and Run 2	Set Average <sup>c</sup> with statistics	Literature values
$K$ ( $\times 10^6 \text{ M}^{-1}$ )	2.4	2.8	$2.6 \pm 0.2$	$2.6 \pm 0.6$	$3.0^d$ ; $4.8^e$ or $25^f$
$\Delta D_1$	13.1	14.5	$13.8 \pm 0.7$	$13.8 \pm 0.7$	N/A
$D_0$	95.0	95.9	$95.4 \pm 0.5$	$95.4 \pm 0.7$	N/A
RMS	0.86	0.81	0.84	0.49	

<sup>a</sup>For the blue curve shown in Figure 2.

<sup>b</sup>For the magenta curve shown in Figure 2.

<sup>c</sup>For the curve fit to the average data of Run 1 and Run 2.

<sup>d</sup>From Ref. [28].

<sup>e</sup>From kinetic analysis in Ref. [27].

<sup>f</sup>From equilibrium analysis in Ref. [27].

has usually been avoided because it is difficult to solve for free ligand concentration,  $[\text{Lig}]$ , as a function of  $[\text{Lig}_T]$  by a direct algebraic method. To our knowledge, this is also the first time that an ordinary differential equation has been employed to calculate  $[\text{Lig}]$  from  $[\text{Lig}_T]$ . We believe the model employs a computationally efficient method and will be of general interest to chemists and biochemists who model reaction equilibrium.

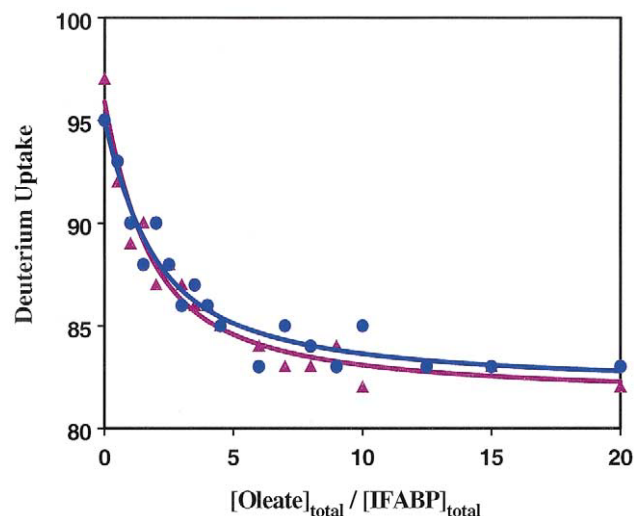
Furthermore, in most non-mass spectrometric methods, the free  $[\text{Lig}]$  is directly measured in solution, but this is not possible when using MS (i.e., the complex must be introduced into the gas phase). When using MS, one must infer the corresponding solution concentrations from the gas-phase measurements, a difficult proposition owing to the discriminatory effects of ESI and MALDI.

In addition, we have incorporated a resampling method to evaluate the uncertainties of the fitting parameters. With this step, one is not required to acquire large data sets and fit them separately to obtain an average and a standard deviation for each fitting parameter, an onerous requirement and sometimes impossible when the protein is difficult to obtain. The current model, which is accompanied by statistical analysis and resampling, evaluates the reliability of the affinity determination. Further, we can use the model to obtain a deeper understanding of the titration curve and the selection of parameters. However, if multiple data sets were acquired under considerably different experimental conditions, the resampling statistical analysis should not be used. Each data set should be analyzed separately, using the first two components in the modeling procedure, to give an average and a standard deviation for the binding constant(s).

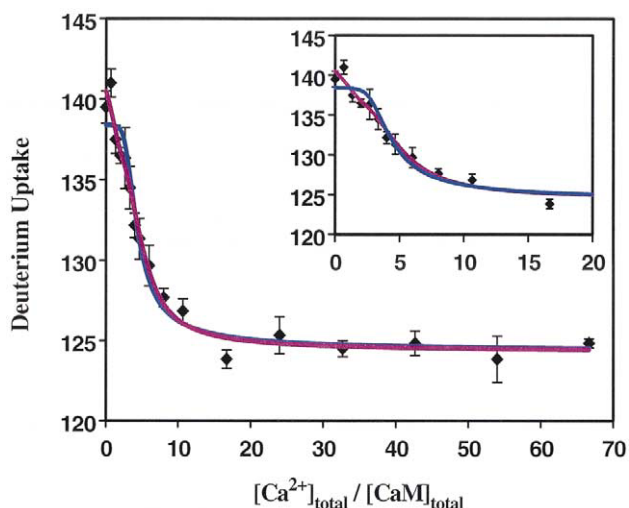
We chose not to use the standard resampling statistical methods such as Bootstrap or Jackknife [23–26] although our method resembles Bootstrap. The reason for our choice is that various data points in the PLIMSTEX curve contribute differently to the extracted binding constants. Therefore, the residuals of the titration data were resampled at every  $[\text{Lig}_T]$ . Further, the titration curve has a nonlinear nature and is comprised of a relatively small number of experimental data in practice.

### One-to-One Protein–Ligand Binding

The interaction between intestinal fatty acid binding protein and potassium oleate involves one-to-one protein–ligand binding as shown by others [27, 28] and recently by us using PLIMSTEX [9]. Although we previously published one titration curve and the results of other titrations [9], we chose this system to demonstrate the modeling of a H/D exchange titration and the use of the newly implemented statistical analysis of the uncertainties. We acquired two independent sets of PLIMSTEX data for titrating wild-type rat I-FABP with  $\text{K}^+$ -oleate under similar experimental conditions, and fit them by using a 3-parameter, 1:1 binding model without and with statistical analysis (see Table 1 for the fit results). Fitting separately the two sets of titration data (Figure 2) gave average values and deviations for the search parameters ( $\beta_1$ , same as  $K_1$ ;  $\Delta D_1$ ; and  $D_0$ ) (Table 1). The RMS values for the two runs are 0.86 and 0.81 mass units, respectively. The best-fit obtained using the 1:1 binding model for average values from the two data sets can be seen by consulting [9], Figure 1.



**Figure 2.** Two sets of independent PLIMSTEX data individually fit with a 3-parameter 1:1 binding model for 0.3- $\mu\text{M}$  wild type I-FABP titrated with  $\text{K}^+$ -oleate in 95%  $\text{D}_2\text{O}$ , 20-mM pyrophosphate buffer containing 135-mM KCl, and 10-mM NaCl (apparent pH = 9.0), after 3 h of exchange.



**Figure 3.** PLIMSTEX data averaged from three independent runs for titration of 15- $\mu\text{M}$  porcine CaM with  $\text{Ca}^{2+}$  in 99%  $\text{D}_2\text{O}$ , 50-mM HEPES buffer (apparent pH = 7.4), H/D exchange time = 1 h. The error bars for each data point were the standard deviations from the three runs. The solid blue line is the fitted curve for the average data using a 1:4 binding model and a 4-parameter ( $\beta_3$ ,  $\beta_4$ ,  $D_0$ ,  $\Delta D_4$ ) fit. The magenta line is the fitted curve that was obtained by using a 7-parameter ( $\beta_3$ ,  $\beta_4$ ,  $D_0$ ,  $\Delta D_1$ ,  $\Delta D_2$ ,  $\Delta D_3$ ,  $\Delta D_4$ ) fit.

Here we invoke for the first time in our efforts the resampling statistical analysis to evaluate the uncertainties for the search parameters  $K_1$ ,  $\Delta D_1$ , and  $D_0$  for I-FABP (see Table 1).

The RMS for averaged data decreases to 0.49 u, which indicates a better fit than can be obtained for the individual data sets. Despite the better fit, the mean values of  $K_1$ ,  $\Delta D_1$ , and  $D_0$  from fitting the average data are similar to the average results from the separate fits. We have found similar trends for other protein–ligand systems [9] when comparing results with and without resampling. The uncertainty estimated from limited data sets without statistical analysis may be biased because only results from two or three titrations were used, and the error distribution may not represent that of a larger population of data. Resampling the IFABP-oleate titration gave  $2^{15}$  non-replicated combinations of titration data when resampling was used for at least two independent titrations acquired under similar experimental conditions. Thus, the calculated error distribution for each search parameter should be more similar to the real error distribution. Moreover, less sample quantity and less time for data acquisition are required when one employs resampling.

### One to Four Protein–Ligand Binding

To explore a more complex binding situation, we applied PLIMSTEX to the  $\text{Ca}^{2+}$  binding to calmodulin (CaM). Figure 3 shows the average titration data for 15  $\mu\text{M}$  of porcine calmodulin titrated with  $\text{Ca}^{2+}$ . We used 17 data points to define the titration and three trials for each point. The H/D exchange time was 1 h for each

titration point; at this time the extent of exchange had reached steady state and there was a significant difference in the extent of exchange between apo and holo CaM. With the first additions of  $\text{Ca}^{2+}$ , the extent of H/D exchange either stayed constant or increased slightly, indicating no gain, and even a small loss, of protection for CaM when binding to one  $\text{Ca}^{2+}$ . As more  $\text{Ca}^{2+}$  was added, the extent of H/D exchange dropped significantly and leveled off as the protein became saturated with  $\text{Ca}^{2+}$ . Use of a 4-parameter ( $\beta_3$ ,  $\beta_4$ ,  $\Delta D_4$  and  $D_0$ ) fit gives the blue curve (Figure 3), a fit obtained for a 1:4 protein: $\text{Ca}^{2+}$  sequential binding model in which we assume that the change in H/D exchange behavior is due principally to an increase in protection afforded by the CaM-4Ca form. Because the binding of the fourth, and possibly the third,  $\text{Ca}^{2+}$  are important in changing the conformation of CaM, only  $\beta_3$  and  $\beta_4$  were searched in the model (for the calculation of  $K_3$  and  $K_4$ , we took  $\beta_1$  and  $\beta_2$  from the literature [29]). The averages and standard deviations for the various search parameters are based on three independent titrations. The stepwise macroscopic binding constants ( $K_i$ ), which are readily calculated from  $\beta_i$ , compare favorably with the literature values (Table 2).

One shortcoming of the 4-parameter model is that it does not fit well the initial part of the titration curve, probably because we ignored contributions to the extent of exchange by Ca-bound species other than CaM-4Ca. To correlate better the data, we implemented a more sophisticated 7-parameter model ( $\beta_3$ ,  $\beta_4$ ,  $D_0$ ,  $\Delta D_1$ ,  $\Delta D_2$ ,  $\Delta D_3$  and  $\Delta D_4$ ) to take into account any changes in exchange due to all five CaM- $x$ Ca ( $x = 0$ –4) species. The resulting fit (magenta curve in Figure 3) not only accommodates the early data in the titration but also has a reduced RMS (0.7). The outcome of the search with seven parameters (summarized in Table 2) gives information for all the binding species, but the relative deviation for each search parameter is larger when seven parameters were used instead of four. The poorer precision suggests that the number of data for defining the titration curve is insufficient and the outcome too sensitive to experimental error. After all, one expects that the introduction of more parameters should be compensated for by using more data points in the titration.

As a test of the outcome, a better defined (i.e., by using more points) titration is the Ca-titration of CaM in 50 mM of HEPES buffer in which 59 data points were used, and each point was obtained in duplicate. The percentage of  $\text{D}_2\text{O}$  was decreased from 99 to 90% in the H/D exchange media to minimize dilution of the protein and to conserve it. The new data were fit by using both the 4-parameter and the 7-parameter models; the results are in Table 3 (the curves were published in our recent article [30]). The relative standard deviation for each binding constant was 40–70% less than those estimated when using the smaller number of data in the titration. This comparison clearly shows that more

**Table 2.** Binding properties obtained from PLIMSTEX analysis for CaM–Ca interaction (17 data points)<sup>a</sup>

Models Parameters	4-Parameter fit ( $\beta_3, \beta_4, D_0, \Delta D_4$ )		7-Parameter fit ( $\beta_3, \beta_4, D_0, \Delta D_1,$ $\Delta D_2, \Delta D_3, \Delta D_4$ )		Literature values <sup>b</sup>
	Values	RSTD	Values	RSTD	
$\beta_3$ ( $\times 10^{16}\text{M}^{-3}$ )	$18 \pm 11$	66%	$2.0 \pm 1.9$	92%	5.0
$\beta_4$ ( $\times 10^{21}\text{M}^{-4}$ )	$13 \pm 9$	68%	$2.2 \pm 1.9$	84%	16
$K_3$ ( $\times 10^4\text{M}^{-1}$ )	$14 \pm 9$	66%	$1.6 \pm 1.5$	92%	4.0
$K_4$ ( $\times 10^5\text{M}^{-1}$ )	$0.7 \pm 0.7$	94%	$1.1 \pm 1.3$	124%	3.2
$\sqrt{K_3 K_4}$ ( $\times 10^4\text{M}^{-1}$ )	$10 \pm 3$	34%	$5.0 \pm 0.8$	42%	11
$D_0$	$138.4 \pm 0.4$	0.4%	$140.5 \pm 0.8$	0.6%	
$\Delta D_1$	0		$-0.3 \pm 9.9$	3973%	
$\Delta D_2$	0		$4.4 \pm 1.1$	26%	N/A
$\Delta D_3$	0		$0.3 \pm 7.2$	2526%	
$\Delta D_4$	$14.1 \pm 0.5$	3.6%	$16.2 \pm 0.9$	5.2%	
RMS		1.0		0.7	N/A

<sup>a</sup>For curves in Figure 3: [CaM] = 15  $\mu\text{M}$ ; in 50 mM HEPES, 99% D<sub>2</sub>O, apparent pH = 7.4; T = 21.5 °C.

<sup>b</sup>From Ref. [29].

reliable results for the binding constants and  $\Delta D$  values can be obtained when more data points are taken in the titration. Another advantage of the larger amount of data is that the precision for  $K_3$ ,  $K_4$ ,  $\Delta D_4$ , and  $D_0$  with the 7-parameter fit is now comparable to that obtained with the 4-parameter model.

Although the precision of the binding constants is better when more data are used, two other features of the results do not show improvement. First, the values for  $\Delta D_1$  and  $\Delta D_3$  still do not have high confidence. This is due to the nature of binding. There is a strong positive cooperativity in the binding of the first two Ca<sup>2+</sup> to CaM, as well as that of the last two. CaM-1Ca and CaM-3Ca never became the predominant species in the titration. Second, the fits (RMS values) that were obtained with the 4- and 7-parameter models applied to the titrations defined by 59 data points are similar to those achieved for the titrations shown in Figure 3.

The shape of the titration curve is determined principally by the complexation of the fourth Ca<sup>2+</sup>, as discussed earlier. Thus, PLIMSTEX is not a reliable way to determine  $K_1$  and  $K_2$  in this situation, and these values must be taken from the literature or, in the case of an unknown protein, estimated until they can be measured by another method. Thus, an important question is: What is the sensitivity of  $K_3$  and  $K_4$  to the values of the values of  $K_1$  and  $K_2$  that are input into the 4-parameter model? We also posed this question and provided an answer in another article [30]. When the values of  $K_1$  and  $K_2$  were artificially increased or decreased by a factor of 8, both  $\beta_3$  and  $\beta_4$  obtained with the 4-parameter fit changed, but  $K_4$  remained nearly constant (within a factor of 1.2), and  $K_3$  only varied within a factor of 2. Thus, the relative insensitivity of the values of  $K_3$  and  $K_4$  to the choices of  $K_1$  and  $K_2$  speaks well for the method given that  $K_1$  and  $K_2$  are not

**Table 3.** Binding properties obtained from PLIMSTEX analysis for CaM–Ca interaction (59 data points)<sup>a</sup>

Models Parameters	4-Parameter fit ( $\beta_3, \beta_4, D_0, \Delta D_4$ )		7-Parameter fit ( $\beta_3, \beta_4, D_0, \Delta D_1,$ $\Delta D_2, \Delta D_3, \Delta D_4$ )		Literature values <sup>b</sup>
	Values	RSTD	Values	RSTD	
$\beta_3$ ( $\times 10^{16}\text{M}^{-3}$ )	$9 \pm 2$	27%	$2.5 \pm 0.6$	25%	5.0
$\beta_4$ ( $\times 10^{21}\text{M}^{-4}$ )	$10 \pm 3$	27%	$3 \pm 1$	34%	16
$K_3$ ( $\times 10^4\text{M}^{-1}$ )	$7 \pm 2$	27%	$2.0 \pm 0.5$	25%	4.0
$K_4$ ( $\times 10^5\text{M}^{-1}$ )	$1.1 \pm 0.4$	39%	$1.1 \pm 0.5$	42%	3.2
$\sqrt{K_3 K_4}$ ( $\times 10^4\text{M}^{-1}$ )	$9 \pm 1$	14%	$5.0 \pm 0.8$	17%	11
$D_0$	$126.7 \pm 0.3$	0.2%	$128.4 \pm 0.8$	0.59%	
$\Delta D$	0		$-1 \pm 9$	1571%	
$\Delta D_2$	0		$3.5 \pm 0.7$	19%	N/A
$\Delta D_3$	0		$0.5 \pm 1.3$	272%	
$\Delta D_4$	$12.6 \pm 0.3$	2.6%	$14.8 \pm 0.7$	5%	
RMS		1.0		0.8	N/A

<sup>a</sup>[CaM] = 15  $\mu\text{M}$ ; in 50 mM HEPES, 90% D<sub>2</sub>O, apparent pH = 7.4; T = 21.5 °C.

<sup>b</sup>From Ref. [29].



reliably determined by PLIMSTEX.  $K_3$  and  $K_4$  calculated from the 7-parameter model, however, are more sensitive to changes in the input values of  $K_1$  and  $K_2$  than from the 4-parameter model.

The positive cooperativity in binding of the last two  $\text{Ca}^{2+}$  ions to CaM, as shown in other studies [29], suggests that the product of  $K_3K_4$  would be better determined than the individual  $K$  values. This is also a favorable outcome because either the product of  $K_3K_4$  or the geometric average binding affinity for N-domain ( $\sqrt{K_3K_4}$ ) is used to compare results from various studies made under similar experimental conditions. The geometric average binding affinity ( $\sqrt{K_3K_4}$ ) and the individual  $K$ 's for CaM-Ca binding using either the 4- or 7-parameter fitting of PLIMSTEX data (Table 3) agree with previously reported values within a factor of 3.

## Conclusion

The modeling process described here not only is a new tool in analyzing H/D exchange data acquired by ESI-MS, but also contains novel aspects for quantification of ligand binding by titrations. It is the first time that H/D exchange accompanying a titration of a protein with a ligand has been modeled to give the binding constant for the interaction. An ordinary differential equation was used, also for the first time, to calculate the free ligand concentration from the total ligand concentration. A nonlinear least squares regression method minimizes the error between the calculation-produced and the experimentally-measured overall deuterium shift by varying the underlying parameters. A resampling method and second order statistics were used to evaluate the uncertainties of the fit parameters. The statistical analysis also can give a better understanding of results from H/D exchange titrations and be a valuable asset in designing a sequel round of experiments.

We demonstrated the model by applying it to two 1:n sequential binding in various protein-ligand systems. Although three parameters ( $\beta_1$ ,  $D_0$ , and  $\Delta D_1$ ) are sufficient for modeling 1:1 protein:ligand binding, more parameters are required when more than one ligand binds. To choose the appropriate search parameters for the model, one should consider the number of titration points that are needed and the ability of H/D exchange to respond to the various binding steps. A statistical analysis of the uncertainty of each search parameter ( $\beta_i$ ,  $D_0$ , and  $\Delta D_i$ ) also provides a basis for designing an appropriate titration.

## Acknowledgments

The authors thank Professors Amy Walker and Lev Gelb of Washington University for helpful discussions in the modeling process, and Professor David Cistola and Dr. Benhur Ogbay for providing I-FABP. The National Centers for Research Resources of the NIH provided financial support (grant no. 2P41RR00954).

## References

- Smith, D. L.; Dharmasiri, K. Protein-Ligand Binding Studied by Amide Hydrogen Exchange and Mass Spectrometry. *NATO ASI Ser. C* **1998**, *510*, 45–58.
- Zhang, Z.; Smith, D. L. Determination of Amide Hydrogen Exchange by Mass Spectrometry: A New Tool for Protein Structure Elucidation. *Protein Sci.* **1993**, *2*, 522–531.
- Resing, K. A.; Hoofnagle, A. N.; Ahn, N. G. Modeling Deuterium Exchange Behavior of Erk2 Using Pepsin Mapping to Probe Secondary Structure. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 685–702.
- Zhang, Z.; Post, C. B.; Smith, D. L. Amide Hydrogen Exchange Determined by Mass Spectrometry: Application to Rabbit Muscle Aldolase. *Biochemistry* **1996**, *35*, 779–791.
- Demmers, J. A.; van Duijn, E.; Haverkamp, J.; Greathouse, D. V.; Koeppe, R. E., II; Heck, A. J.; Killian, J. A. Interfacial Positioning and Stability of Transmembrane Peptides in Lipid Bilayers Studied by Combining Hydrogen/Deuterium Exchange and Mass Spectrometry. *J. Biol. Chem.* **2001**, *276*, 34501–34508.
- Zhang, Z.; Li, W.; Logan, I. M.; Li, M.; Marshall, A. G. Human Recombinant [C22a] Fk506-Binding Protein Amide Hydrogen Exchange Rates from Mass Spectrometry Match and Extend Those from NMR. *Protein Sci.* **1997**, *6*, 2203–2217.
- Demmers, J. A. A.; Haverkamp, J.; Heck, A. J. R.; Koeppe, R. E. II; Killian, J. A. Electrospray Ionization Mass Spectrometry as a Tool to Analyze Hydrogen/Deuterium Exchange Kinetics of Transmembrane Peptides in Lipid Bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3189–3194.
- Powell, K. D.; Ghaemmaghani, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. A General Mass Spectrometry-Based Assay for the Quantitation of Protein-Ligand Binding Interactions in Solution. *J. Am. Chem. Soc.* **2002**, *124*, 10256–10257.
- Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. Quantification of Protein-Ligand Interactions by Mass Spectrometry, Titration, and H/D Exchange: PLIMSTEX. *J. Am. Chem. Soc.* **2003**, *125*, 5252–5253.
- Bard, Y. *Nonlinear Parameter Estimation*; Academic Press: New York, NY, 1974.
- Bates, D. M.; Watts, D. G. *Nonlinear Regression Analysis and Its Applications*; Wiley: New York, NY, 1988.
- Young, S. L.; Matijevic, E.; Meites, L. Titrimetric Applications of Multiparameter Curve Fitting. Iii. Evaluation of the Parameters Characterizing Acidimetric Potentiometric Titrations of Laurate Ion. *J. Phys. Chem.* **1974**, *28*, 2626–2631.
- Leggett, D. J.; Kelly, S. L.; Shiu, L. R.; Wu, Y. T.; Chang, D.; Kadish, K. M. A Computational Approach to the Spectrophotometric Determination of Stability Constants II. Application to Metalloporphyrin-Axial Ligand Interactions in Non-Aqueous Solvents. *Talanta* **1983**, *30*, 579–586.
- Johnson, M. L. Why, When, and How Biochemists Should Use Least Squares. *Anal. Biochem.* **1992**, *206*, 215–225.
- Johnson, M. L.; Frasier, S. G. Nonlinear Least Squares Analysis. *Methods Enzymol* **1985**, *117* (Enzyme Struct. Pt. J), 301–342.
- Johnson, M. L. Parameter Correlations While Curve Fitting. *Methods Enzymol.* **2000**, *321*, 424–446.
- Johnson, M. L. Outliers and Robust Parameter Estimation. *Methods Enzymol.* **2000**, *321*, 417–424.
- Johnson, M. L.; Faunt, L. M. Parameter Estimation by Least Squares Methods. *Methods Enzymol.* **1992**, *210*, 1–37.
- Wyman, J.; Gill, S. J. *Binding and Linkage: Functional Chemistry of Biological Macromolecules*; University Science Books: Mill Valley, CA, 1990.
- Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed; Cambridge University Press: New York, NY, 1992.

21. Gear, C. W. Numerical Initial Value Problems in Ordinary Differential Equations; Prentice-Hall: Englewood Cliffs, NJ, 1971, Section 3.2.
22. Draper, N. R.; Smith, H. Applied Regression Analysis, 3rd ed; John Wiley and Sons, Inc: New York, 1998, Section 24.1.
23. Efron, B.; Tibshirani, R. J. An Introduction to the Bootstrap; Chapman and Hall/CRC: Boca Raton, FL, 1998.
24. Shao, J.; Tu, D. The Jackknife and Bootstrap; Springer Verlag: New York, NY, 1995.
25. Almansa López, E.; Bosque-Sendra, J. M.; Cuadros Rodríguez, L.; García Campaña, A. M.; Aaron, J. J. Applying Non-Parametric Statistical Methods to the Classical Measurements of Inclusion Complex Binding Constants. *Anal. Bioanal. Chem.* **2003**, 375, 414–423.
26. Merle, M.; Graves, P. V.; Labouesse, B. Substrate Depletion Analysis as an Approach to the Pre-Steady-State Anticooperative Kinetics of Aminoacyl Adenylate Formation by Tryptophanyl-Trna Synthetase from Beef Pancreas. *Biochemistry* **1984**, 23, 1716–1723.
27. Cistola, D. P.; Kim, K.; Rogl, H.; Frieden, C. Fatty Acid Interactions with a Helix-Less Variant of Intestinal Fatty Acid-Binding Protein. *Biochemistry* **1996**, 35, 7559–7565.
28. Kurian, E.; Kirk, W. R.; Prendergast, F. G. Affinity of Fatty Acid for Rat Intestinal Fatty Acid Binding Protein: Further Examination. *Biochemistry* **1996**, 35, 3865–3874.
29. Linse, S.; Helmersson, A.; Forsén, S. Calcium Binding to Calmodulin and Its Globular Domains. *J. Biol. Chem.* **1991**, 266, 8050–8054.
30. Zhu, M. M.; Rempel, D. L.; Zhao, J.; Giblin, D. E.; Gross, M. L. Probing Ca<sup>2+</sup>-Induced Conformational Changes in Porcine Calmodulin by H/D Exchange and ESI-MS: Effect of Cations and Ionic Strength. *Biochemistry* **2003**, 42, 15388–15397.