Published in final edited form as:

Anal Biochem. 2018 January 01; 540-541: 64–75. doi:10.1016/j.ab.2017.10.013.

Using two-site binding models to analyze microscale thermophoresis data

Shih-Chia Tso^a, Qiuyan Chen^{b,†}, Sergey A. Vishnivetskiy^b, Vsevolod V. Gurevich^b, T. M. Iverson^{b,c}, and Chad A. Brautigam^{a,d,*}

^aDepartment of Biophysics, The University of Texas Southwestern Medical Center, Dallas, TX, USA

^bDepartment of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA

^cDepartment of Biochemistry, Vanderbilt University Medical Center, Nashville, TN, USA

^dDepartment of Microbiology, The University of Texas Southwestern Medical Center, Dallas, TX, USA

Abstract

The emergence of microscale thermophoresis (MST) as a technique for determining the dissociation constants for bimolecular interactions has enabled these quantities to be measured in systems that were previously difficult or impracticable. However, most models for analyses of these data featured the assumption of a simple 1:1 binding interaction. The only model widely used for multiple binding sites was the Hill equation. Here, we describe two new MST analytic models that assume a 1:2 binding scheme: the first features two microscopic binding constants ($K_D(1)$ and $K_D(2)$), while the other assumes symmetry in the bivalent molecule, culminating in a model with a single macroscopic dissociation constant ($K_{D,M}$) and a single factor (α) that accounts for apparent cooperativity in the binding. We also discuss the general applicability of the Hill equation for MST data. The performances of the algorithms on both real and simulated data are assessed, and implementation of the algorithms in the MST analysis program PALMIST is discussed.

Keywords

Microscale thermophoresis; Protein-ligand interactions; Protein-protein interactions; Arrestin-3; DNA aptamer

*To whom correspondence should be addressed: 5323 Harry Hines Blvd., Dallas, TX 75390-8816, Phone: +1-214-645-6384, chad.brautigam@utsouthwestern.edu.

[†]Current address: Department of Biological Sciences, Purdue University, West Lafayette, IN, USA.

Conflicts: The authors declare no conflicting interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

A full understanding of the functioning of biomolecules usually involves quantitative characterization of their interactions with other biomolecules. Such interactions can entail protein:protein, protein:small molecule, protein:nucleic acid, nucleic acid:small molecule, and other binding events. In quantifying the energetics of these interactions, a convenient and intuitive measure is the equilibrium dissociation constant (K_D). In fact, $K_D = 1/K_A$, and K_A is directly dependent on the free energy of the interaction: $G^0 = -RT \ln K_A$, where R is the universal gas constant, and T is temperature in Kelvins. A number of biophysical methods can be exploited to measure K_D , including analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), surface plasmon resonance, fluorescence anisotropy, and many others [1–7].

One of the newest biophysical methods available to the researcher seeking to examine the affinities of biomolecular interactions is microscale thermophoresis (MST) [8,9]. In its most popular format, this method requires that one of the molecules under study be fluorescent in the visual spectrum; this is usually achieved via the modification of the molecule with an extrinsic dye, although fluorescent-protein tags or intrinsic fluorophores (when available) can also be used. The experiments are carried out in transparent capillary tubes in which the spatial migration of the labeled molecules can be monitored in a small volume of the capillary using a confocal microscope. “Thermophoresis”, i.e. the net movement of molecules along a temperature gradient, is initiated by illuminating the monitored volume of the capillary with an infrared (IR) laser. The change in fluorescence is recorded as a function of time, and then the IR laser is extinguished. A comparison of the tracked fluorescence before and some time after the actuation of the IR laser is quantified as “thermophoresis.” The thermophoresis of the labeled molecules usually changes as a function of whether they are liganded or not, and a binding curve can be obtained from plotting thermophoresis as a function of ligand concentration.

An important methodological aspect of most MST experiments to determine K_D is that these use low-concentration, “tracer” methods. That is, only one of the molecules under study is detected (labeled), and it is supplied at a very low concentration, usually well below K_D . Similar experimental designs are widely used to characterize binding interactions via techniques such as fluorescence quenching and fluorescence anisotropy. These contrast with other technologies, such as NMR, EPR, and ITC, where molecular concentrations are often near to or even exceeding K_D . This character of the MST assay has advantages and disadvantages. One advantage is that it can consume very small quantities of the labeled molecule and still yield interpretable data. However, this particular strength has a drawback: information on stoichiometry is not readily available from such a low-concentration experiment. Consequently, most users plan MST experiments under the assumption that the ligand and the labeled receptor interact with a 1:1 stoichiometry. If the user suspects non-unitary stoichiometry, additional experiments can be performed to yield this datum in an alternatively configured MST experiment [10].

Currently, only two analytic models for analyzing MST data are in wide use: a 1:1 interaction model that yields K_D , and the so-called Hill model, which uses a variant of the Hill equation [11] to allow the fitting of the data. However, binding events that are known to have a 1:2 molar ratio of interaction are usually not well served by these models. This is because such data can have two inflection points, but both of the current models can only account for one. One proposed strategy is to split a biphasic binding curve into two sub-curves; two subsequent separate analyses using the 1:1 model can then be performed [8]. While this approach can yield reasonable results, it unnecessarily excludes information and provides for sparse sampling of both binding events.

In the current work, we describe two 1:2 fitting algorithms that have been derived for use with MST data. The first is a microscopic model, in which the two binding sites on the labeled molecule are assumed to be asymmetric (i.e. independent) and non-cooperative. Two K_D values, $K_D(1)$ and $K_D(2)$, are obtained in such an analysis. The second model is a macroscopic one, implying that the two binding sites are symmetric (i.e. identical). In this case, a single K_D , $K_{D,M}$, is obtained, and a factor α accounts for any observed cooperativity. Additionally, these methods have been adapted to provide quick, rigorous, automated estimates of confidence intervals for the fitted parameters. The performance of these algorithms was assessed in simulations and on empirical data, and the implications of the findings on experimental planning and performance were considered. Also, the Hill model was assessed for use in MST experiments, with important implications for the interpretation of results from such an analysis. Finally, we integrated these algorithms into the freeware program PALMIST.

Materials & Methods

Materials

DNA modifying enzymes were from New England Biolabs (Ipswich, MA). DNA purification kits were from Zymo Research (Irvine, CA). Heparin Sepharose, Q- and SP-Sepharose were from GE Healthcare (Marlborough, MA). All other reagents were from Amresco (Solon, OH) or Sigma-Aldrich (St. Louis, MO).

Protein Methods

The gene encoding bovine arrestin-3 was cloned into the pTrcHisB vector (Invitrogen; Carlsbad, CA) with the codon for L394 mutagenized to a stop codon (TGA). This created a fully functional version of arrestin-3 truncated immediately after the C-tail attachment point (encoding residues 1-393 and deleting residues 394-408) [12]. The N_{IP6} arrestin-3 (R8Q, K11A, K12A, K108Q, K161Q, K295Q) mutant was generated on this background of truncated arrestin-3 (1-393) as described [13]. The pTrcHisB plasmid containing N_{IP6} arrestin-3 was transformed into *E. coli* BL21 Gold and grown in 1-L cultures in Luria-Bertani medium supplemented with 100 mg/L ampicillin. Cells were grown at 30 °C overnight with shaking at 250 rpm, then protein expression was induced with the addition of 35 μ M isopropyl β -D-1-thiogalactopyranoside for 4 h. Cells were harvested by centrifugation and the pellet was stored at -80 °C.

N_{IP6} arrestin-3 was purified using a modification of a previously described protocol [14]. Briefly, cell pellets were resuspended in buffer containing 10 mM Tris, pH 8.0, 5 mM ethylene-bis(oxyethylenenitrilo)tetraacetic acid (EGTA), 2 mM dithiothreitol (DTT), 10 mM glucose, phenylmethylsulfonyl fluoride, and protease-inhibitor cocktail (Sigma). The suspension was treated with 0.1 mg/ml lysozyme for 40 min at 4 °C. The suspension was subsequently frozen and thawed to facilitate cell lysis. The cells were then sonicated three times with ~ 1 min intervals at 4 °C (15 sec each time, 90% amplitude; Fisher Dismembrator, Model 500), and treated with RNase and DNase on ice for 40 min, as described [14]. The lysate was clarified by centrifugation at $20,800 \times g$ (SLA 3000 rotor, RC-5B Plus centrifuge, Beckman, Palo Alto, CA) for 90 min, and arrestin was precipitated by the addition of $(NH_4)_2SO_4$ to a final concentration 0.32 mg/mL. Precipitated N_{IP6} arrestin-3 was collected by centrifugation at $20,800 \times g$ for 90 min. The supernatant including floating lipids was removed, and the pellet was dissolved in buffer containing 10 mM MOPS pH 7.2, 2 mM EGTA, 2 mM ethylenediaminetetraacetic acid (EDTA) and 2 mM DTT, then centrifuged again at $20,800 \times g$ for 90 min to remove particulates. The supernatant containing soluble N_{IP6} arrestin-3 was applied to a Heparin Sepharose column and eluted with a linear NaCl gradient. Fractions containing N_{IP6} arrestin-3 were identified by SDS-PAGE and Western blot, then combined. The NaCl concentration of the pooled fractions was adjusted to 80 mM, and the solution was loaded onto linked Q-Sepharose and SP-Sepharose columns. At a NaCl concentration of 80 mM, N_{IP6} arrestin-3 flows through the Q-Sepharose (while most contaminants bind), but binds to the SP-Sepharose. The columns were uncoupled and a linear NaCl gradient was used to elute N_{IP6} arrestin-3 from the SP-Sepharose. The fractions containing N_{IP6} arrestin-3 were identified by SDS-PAGE and combined, concentrated with a 30 kDa cutoff concentrator, then further purified using a Superdex 200 increase 10/300 GL column (GE Healthcare) equilibrated with 20 mM MOPS pH 7.2, 150 mM NaCl, and 2 mM tris(2-carboxyethyl)phosphine (TCEP). This final step was considered to serve as quality-control step that assessed the correct folding and monodispersity of the protein preparation.

DNA Aptamer

The DNA aptamer described by Entzian and Schubert [15] was also employed in this study. The sequence was 5' Cy5-CCTGGGGGAGTATTGCGGAGGAAGG-3', where "Cy5" represents the Cy5 fluorescent dye. The oligonucleotide was purchased from Integrated DNA Technologies (Coralville, IA). The lyophilized DNA was dissolved in a volume of H₂O sufficient to bring the concentration to 95 μ M. It was then frozen at -20 °C.

Microscale thermophoresis

Arrestin N_{IP6} —MST experiments were carried out on a NT.115 (NanoTemper GmbH, Munich, Germany) instrument; the arrestin-3 N_{IP6} studies were conducted in an instrument equipped with a blue/green filter set, and only the "blue" channel was used. All N_{IP6} experiments were carried out in MST Buffer (20 mM 3-morpholinopropane-1-sulfonic acid pH 7.5, 150 mM NaCl, and 1 mM TCEP). The N_{IP6} protein contained eight cysteine residues and was labeled with AlexaFluor-488 C5 maleimide ester by incubating the 200 μ M of the dye with a 20 μ M solution of the protein for 30 min in the dark. The labeling mixture was subsequently applied to a G-25 column (GE Healthcare, Inc.) that had been equilibrated

with MST Buffer. After elution from this column, the concentration of the labeled protein was quantified spectrophotometrically. In general, the experiments consisted of 15-point titration series in which the ligand concentrations were generated as a 1:1 dilution of the previous sample. At this point, all buffers were supplemented to 1% (w/v) maltose neopentyl glycol-3 (MNG-3). To this, a 100 nM solution of labeled N_{IP6} (also supplemented to 1% MNG-3) was added to the ligand-containing mixtures, bringing the final concentration of the protein to 50 nM. Usually, a 16th sample that contained no ligand was characterized. During analysis, this sample was assigned a very low (nonzero) concentration of ligand to facilitate the logarithmic representation of the data. All samples were incubated in the dark for ca. 45 min prior to being loaded into premium-coated capillary tubes (NanoTemper). A typical MST experiment featured LED illumination at 50% power and IR illumination at 80% power. The pre-IR phase was 5 s, followed by an IR-on phase of 10 s and an IR-off phase of 5 s. The short IR-on phase did not allow the traces to approach a new equilibrium value, but this was deemed unnecessary because only the regions just before and just after the actuation of the IR laser (“T-Jump”; [9,10]) were to be analyzed. All data were analyzed using PALMIST [9] using the modifications described in the “Theory” section below.

DNA Aptamer—For DNA aptamer/AMP experiments, the 1 μ L of thawed aptamer was diluted to a final concentration of 190 nM in Binding Buffer (20 mM Tris pH 7.6, 300 mM NaCl, 5 mM $MgCl_2$, 0.01% (v/v) Tween-20). An AMP stock at 10 mM was prepared in Binding Buffer. A 16-point titration was prepared using the 10 mM as the top concentration of the ligand and using a 1:1 serial dilution strategy in 10 μ L sample volumes. Next, 10 μ L of the aptamer solution was added to each sample, yielding a final concentration of aptamer of 95 nM, and the top concentration of AMP of 5 mM. All samples were incubated in the dark for ca. 30 min. before being loaded into standard capillary tubes (NanoTemper). These studies were conducted in an NT.115 instrument equipped with blue and red filters; only the red filter set was used. The LED power was set to 5%, and the IR illumination power to 40%. The pre-IR illumination period was 5 s, followed by an IR-on phase of 30 s and an IR-off phase of 5 s. The experiment was conducted three times, and the data were averaged and fitted in PALMIST. For both the N_{IP6} and aptamer data, the data-fit-residuals plots from MST were made using GUSSI [16].

Theory

Thermophoresis

First, in all cases, we define “thermophoresis” as established by convention [9,17]. The user or analysis program establishes two regions of the fluorescence time-trace curve; commonly these are before the actuation of the IR laser (the “cold” region) and some time afterwards (the “hot” region). Thermophoresis, F_n , can then be defined as:

$$F_n = \frac{F_h}{F_c} \times 1,000 \quad (1)$$

where F_h is the mean fluorescence in the “hot” region and F_c is the mean fluorescence in the “cold” region.

Microscopic binding model

Here, we employ a mathematical strategy that is capable of calculating the concentrations of all possible species given the assumed K_D values and the inputted (“tot”) concentrations of the ligand (“A” hereafter) and the labeled, bivalent receptor (“B*”). This approach is necessary because the signals measured by MST are not strictly proportional to the occupancy of the sites on B*. For example, the complex AB* may exhibit positive thermophoresis (manifesting as normalized thermophoresis value, F_n , of less than 1,000) and the AB*A complex could have a negative thermophoresis ($F_n > 1,000$). Such a scenario demands that the analysis take into account the concentrations of all labeled species. For reasons discussed below, labeling the bivalent receptor is strongly preferred and assumed throughout the Theory section.

The microscopic model used in this paper makes the assumption that B* has two non-identical sites for the ligand, A, termed Site 1 and Site 2 herein. Throughout, we consider Site 1 to have the higher affinity of the two. Thus, there will be two distinguishable species that feature a single bound molecule of A: these are termed AB* and B*A here. There are four species or complexes whose concentrations must be accounted for in the model: B*, AB*, B*A, and AB*A, with the two microscopic binding constants, $K_D(1)$ and $K_D(2)$ governing the binding of A to Sites 1 and 2, respectively, on B*. The algorithm shown here is based on that propounded by Wang & Jiang [18]. First, they state that:

$$A_b = \frac{A \cdot B_{\text{tot}}^*}{K_D(1) + A} + \frac{A \cdot B_{\text{tot}}^*}{K_D(2) + A}, \quad (2)$$

where we use the shorthand of depicting the concentrations of species as italicized respective letters. A_b is the total concentration of A bound on B*. It follows that the first term on the right-hand side of Eq. (2) is the concentration of A bound at Site 1, and the second term is the concentration of A bound at Site 2. Thus

$$AB^* + AB^*A = \frac{A \cdot B_{\text{tot}}^*}{K_D(1) + A} \quad (3)$$

and

$$B^*A + AB^*A = \frac{A \cdot B_{\text{tot}}^*}{K_D(2) + A}. \quad (4)$$

Also, there are two mass-conservation equations that must hold:

$$A_{\text{tot}} = A + AB^* + B^*A + 2AB^*A \quad (5)$$

and

$$B_{\text{tot}}^* = B^* + AB^* + B^*A + AB^*A. \quad (6)$$

By inserting the Eq. (3) and Eq. (4) into Eq. (5), the result is the equation

$$A^3 + pA^2 + qA + r = 0 \quad (7)$$

where

$$\begin{aligned} p &= K_D(1) + K_D(2) + 2B_{\text{tot}}^* - A_{\text{tot}} \\ q &= (K_D(1) \cdot K_D(2)) + \{(B_{\text{tot}}^* - A_{\text{tot}})(K_D(1) + K_D(2))\} \\ r &= -K_D(1) \cdot K_D(2) \cdot A_{\text{tot}} \end{aligned} \quad (8)$$

Eq. (7) has only one physically realistic root [18,19]:

$$A = -\frac{p}{3} + \frac{2}{3}\sqrt{p^2 - 3q} \cos\left(\frac{\theta}{3}\right) \quad (9)$$

where

$$\theta = \arccos\left(\frac{-2p^3 + 9pq - 27r}{2\sqrt{(p^2 - 3q)^3}}\right). \quad (10)$$

Thus, A is easily obtained from knowledge of A_{tot} and B_{tot}^* , along with the assumed knowledge of $K_D(1)$ and $K_D(2)$. To obtain B^* , it is convenient to recognize that there are numerous alternative definitions of $K_D(1)$ and $K_D(2)$ that are equally valid. Among them are

$$K_D(1) = \frac{A \cdot B^*A}{AB^*A}; A + B^*A \rightleftharpoons AB^*A$$

$$K_D(2) = \frac{A \cdot B^*}{B^*A}; A + B^* \rightleftharpoons B^*A$$
(11)

(see Supplemental Theory; mass action equations are given to the right of each K_D definition). Multiplying the dissociation constants in Eq. (11) and rearranging, we find that

$$AB^*A = \frac{A^2 \cdot B^*}{K_D(1)K_D(2)}. \quad (12)$$

Eq. (6) can then be subtracted from Eq. (5), and, after substituting in the r.h.s. of Eq. (12) for AB^*A , we obtain

$$A_{\text{tot}} - B_{\text{tot}}^* - A = \frac{A^2 \cdot B^*}{K_D(1)K_D(2)} - B^*. \quad (13)$$

which, following further rearrangement, yields an equation for B^* :

$$B^* = \frac{A_{\text{tot}} - B_{\text{tot}}^* - A}{\left(\frac{A^2}{K_D(1)K_D(2)} - 1 \right)}. \quad (14)$$

With B^* known, its value can be inserted into Eq. (12) to yield AB^*A . AB^* and B^*A are obtained by inputting known values into the appropriate parts of Eq. (11). With the concentrations of all pertinent species known, the quantity “thermophoresis” (F_n) is assumed to be a linear combination of the species’ individual thermophoresis values (depicted with a respective subscript):

$$F_n = \left(\frac{B^*}{B_{\text{tot}}^*} \cdot F_{B^*} \right) + \left(\frac{AB^*}{B_{\text{tot}}^*} \cdot F_{AB^*} \right) + \left(\frac{B^*A}{B_{\text{tot}}^*} \cdot F_{B^*A} \right) + \left(\frac{AB^*A}{B_{\text{tot}}^*} \cdot F_{AB^*A} \right) \quad (15)$$

Simulations (see Results) demonstrate that $K_D(1)$ and $K_D(2)$ must be significantly (ca. 100×) different to allow their unique determination from MST data; thus, there will be a very low B^*A at all points in the titration in most MST experiments. We thus make the simplification that B^*A will have the same thermophoresis as AB^* , as follows.

$$F_n = \left(\frac{B^*}{B_{\text{tot}}^*} \cdot F_{B^*} \right) + \left(\frac{AB^* + B^*A}{B_{\text{tot}}^*} \cdot F_{AB^*} \right) + \left(\frac{AB^*A}{B_{\text{tot}}^*} \cdot F_{AB^*A} \right). \quad (16)$$

Importantly, this model cannot account for cooperativity between the two sites; the K_D values and any putative cooperativity factor would be too strongly correlated in such experiments to yield rational results.

Macroscopic binding model

When a bivalent molecule B^* has two identical sites (i.e. it is “symmetric”, a situation easily arising if B^* is dimeric), two strategies are available for the analysis of binding. One would be the microscopic model (see above). This, however, discards information that is known about B^* , i.e. its symmetry. The second strategy would be to treat both sites as though they had identical microscopic dissociation constants, and thus account for any observed differences between the two as being attributable to cooperativity. If B^* is truly symmetric (e.g. it is a two-fold symmetric homodimer), this is the most accurate model for this situation, and its theoretical description follows.

For this “macroscopic” model, we follow the approach of Velazquez-Campoy and coworkers [20] to define two “overall” association constants:

$$\begin{aligned} \beta_{M1} &= \frac{AB^*}{A \cdot B^*} \\ \beta_{M2} &= \frac{AB^*A}{A^2 \cdot B^*} \end{aligned} \quad (17)$$

Here, in contrast to the microscopic model, there can be no distinction between AB^* and B^*A ; in the macroscopic view, these species are necessarily identical (and hence below will have identical thermophoresis values). Whether AB^* and B^*A will truly be identical depends on the symmetry of B^* ; for example, if B is a homodimer with two-fold rotational symmetry, the identity of AB^* and B^*A is very likely. Vega et al. [21] have shown that

$$\beta_{M2}A^3 + (\beta_{M1} + \beta_{M2}(2B_{\text{tot}}^* - A_{\text{tot}}))A^2 + (1 + \beta_{M1}(B_{\text{tot}}^* - A_{\text{tot}}))A - A_{\text{tot}} = 0 \quad (18)$$

This cubic equation is analogous to Eq. (7). We may therefore find the value of A by the use of Eq. (9) and Eq. (10), with the following new definitions for p , q , and r :

$$\begin{aligned}
 p &= \frac{\beta_{M1} + (\beta_{M2}(2B_{\text{tot}}^* - A_{\text{tot}}))}{\beta_{M2}} \\
 q &= \frac{1 + (\beta_{M1}(B_{\text{tot}}^* - A_{\text{tot}}))}{\beta_{M2}} \quad . \quad (19) \\
 r &= \frac{-A_{\text{tot}}}{\beta_{M2}}
 \end{aligned}$$

This strategy works well in cases of negative cooperativity (see simulations and arrestin-3 experiments in the Results), but for strong, positive cooperativity, an imaginary quantity can result from the application of Eq. (10). Our algorithm checks for this situation and dynamically switches to an alternative root-finding formula if it occurs (see Supplemental Theory).

Vega et al. [21] also demonstrate that (in a rearrangement of their Eq. 54):

$$B^* = \frac{A_{\text{tot}} - A}{\beta_1 \cdot A + 2\beta_2 \cdot A^2} \quad . \quad (20)$$

From the knowledge of A and B^* , AB^* and AB^*A can be calculated using Eq. (17). As for the microscopic case, the quantity thermophoresis, F_n , is then calculated as the linear combination of the thermophoreses of all labeled species F_{B^*} , F_{AB^*} , and F_{AB^*A} :

$$F_n = \left(\frac{B^*}{B_{\text{tot}}^*} \right) F_{B^*} + \left(\frac{AB^*}{B_{\text{tot}}^*} \right) F_{AB^*} + \left(\frac{AB^*A}{B_{\text{tot}}^*} \right) F_{AB^*A} \quad . \quad (21)$$

In our macroscopic algorithm, the microscopic dissociation constants are considered to be identical, placing a constraint on the relative values of the macroscopic values. Specifically, $K_{D,M2} = 4K_{D,M1}$. Any deviations from this rule must be due to cooperativity; we accommodate that concept by utilizing the cooperativity factor α , as previously introduced by Schuck and colleagues [22]: $K_{D,M2} = 4\alpha K_{D,M1}$. In terms of the β values described above,

$$\begin{aligned}
 K_{D,M1} &= K_{D,M} = \frac{1}{\beta_{M1}} \\
 \alpha &= \frac{\beta_{M1}^2}{4\beta_{M2}} \quad , \quad (22)
 \end{aligned}$$

Operationally, β_{M1} and β_{M2} are refined by the algorithm, but $K_{D,M}$ and α are reported to the user.

Hill model

The Hill equation [11], in terms of two-site binding and thermophoresis, can be written as

$$F_n = F_{B^*} + \frac{F_{AB^*} A^n}{K_{50}^n + A^n}, \quad (23)$$

where K_{50} is the ligand concentration at which 50% of the receptors are filled. It is common to recast the “ K ” term as an “effective concentration at which a 50% signal is seen” (EC_{50}). Notably, there is no term for F_{AB^*} in Eq. (23). This is due to the implicit assumption in the Hill model of strong positive cooperativity such that the AB^* complex is infinitesimally present under all circumstances [23]. The validity of this assumption will be examined below.

General fitting procedures

The above algorithms were incorporated into the analysis program PALMIST [9] for the purpose of using them to fit MST data. Importantly, two key features of that program were retained for the present study: (1) the strategy of alternating least-squares optimization procedures (Levenberg—Marquardt [24,25] vs. Nelder—Mead simplex [26]), and (2) the automated, adaptive confidence-interval search for all refined parameters. The overall performance of these features in these contexts was good.

Results

Simulation of two-site binding models

As described in the Theory section, we have chosen to analyze the MST data from two-site binding systems using two models. The first is a microscopic model (termed “1:2 Micro.” hereafter; see “Microscopic binding model” in “Theory”). In this model, molecule B has two non-identical binding sites (Sites 1 and 2) for molecule A, and thus two K_D 's, $K_D(1)$ and $K_D(2)$, may be refined. The second model is macroscopic (“1:2 Macro.” below), and assumes identical Sites 1 and 2 (e.g., B could be a two-fold symmetric homodimer); hence, only a single K_D , $K_{D,M}$, is refined, with a cooperativity factor α quantifying the degree to which the binding of A affects the binding of a second A. As modeled (see “Macroscopic binding model” in “Theory”), values of α above 1.0 indicate negative cooperativity and those below 1.0 would reflect positive cooperativity.

Operational Symmetry

The first goal of simulation of two-site binding systems was to determine whether there is an operational symmetry in typical MST titrations, where the concentration of the labeled entity is well below the highest K_D studied. That is, under these circumstances, can the bivalent

molecule be held constant and the univalent one titrated and *vice versa*? To examine this issue, we scrutinized species plots of simulated data using the 1:2 Micro. model described above, plotting the fractional populations of all species along with the resulting binding curve for a favorable binding system (Figs. 1 & 2), i.e. with $K_D(1)$ at 10 nM, $K_D(2)$ at 10 μ M, thermophoresis of the free labeled species at 700, that of all singly bound species at 800, and that of the doubly bound species at 750. The labeled species is held constant at 50 nM, a realistic value for MST experiments, and the titrant was varied from 50 pM to 100 μ M in a 22-point 1:1 serial dilution. Although these data are fully computer-generated, the underlying parameters resemble those obtained from empirical data (see arrestin-3 data below). We plotted two Scenarios. In Scenario I (Fig. 1), we have labeled the bivalent B to yield B*, and have titrated in univalent, unlabeled A. Two important findings emerge from this simulation. First, it is only at high A that the AB*A complex becomes significantly occupied. Second, the B*A complex, i.e. the singly bound species with A occupying Site 2 only, is essentially not observed. Stated in another way, the concentrations necessary to start occupying Site 2 are sufficiently high to ensure the saturation of Site 1. We examined other concentration domains that would retain the low concentration of B* but allow significant population of B*A at equilibrium. This can occur when $K_D(2)$ is lowered to values near to $K_D(1)$ (not shown). However, such a scenario does not allow for the unique determination of both K_D 's (see below). For these reasons, B*A can be safely ignored in the calculation of thermophoresis in an analyzable MST experiment performed under typical conditions (see Eq. (16)).

In Scenario II (Fig. 2), we imagine that univalent A is the labeled molecule, yielding A*. The bivalent B is thus the titrant, and, in this microscopic simulation, possible complexes are A*B, BA*, and A*BA*. Under these conditions, however, the only complex displaying significant concentrations is A*B. The large molar excesses of B never favor occupancy of its Site 2. The result of this titration configuration is that only one site is evident in the binding curve. We analyzed data generated by this simulation using a standard 1:1 model [9] and found that the reported K_D was very close to the simulated $K_D(1)$. Hence, this Scenario is likely capable of accurately reflecting the K_D of the tighter binding site if naively analyzed using a 1:1 binding model. Maintaining the constraint that the concentration be low relative to $K_D(2)$, we systematically explored different conditions to determine if one could be found in which there was a significant A*BA*; this could allow a higher probability to derive both K_D 's. While such conditions can be found (see Supplemental Fig. S1 for an example), they occurred only when the K_D 's were close together, disallowing a unique determination of these parameters from the data. Thus, for the best chance at capturing information from both binding sites in a typical MST setting, unlabeled A (the univalent molecule) should be titrated in labeled B*.

We also examined both of these scenarios with the 1:2 Macro. analytic model (Supplemental Figs. S2 and S3). In these cases, $K_{D,M}$ was set to 10 nM, and α to 250, with all other parameters remaining the same as simulated above. We found a similar pattern (with one exception discussed below in the context of the Hill equation): with A as the titrant, two phases could be discerned in the data (i.e. both the AB* and AB*A complexes were observable), whereas with a titrated B, only the A*B complex could be detected. Given these results, in all further studies in this paper, B is the labeled species.

Similarity of $K_D(1)$ and $K_D(2)$

Under optimal circumstances, binding curves characterizing two-site interactions display two inflection points that are widely separated in concentration space (e.g. the synthetic data shown in Fig. 3A, thin line). This represents a favorable circumstance for the successful refinement of both $K_D(1)$ and $K_D(2)$. However, the K_D 's in such curves are very different ($1,000\times$ different in Fig. 3A), raising the question of how different they must be to ensure their reliable estimation from the data. To address this point, we performed Monte Carlo simulations on synthetic data sets, holding $K_D(1)$ in the 1:2 Micro. model constant at 10 nM, but varying $K_D(2)$, i.e. systematically lowering it (e.g. Fig. 3B for a monotonic curve). We performed analogous simulations with the 1:2 Macro. model, holding $K_{D,M}$ at 10 nM and varying the cooperativity factor α , starting with a high value (i.e. a large negative cooperativity) and systematically reducing it. In all of these simulations, the concentration of B^* was set to 50 nM, and the titrant, A, was added in a 22-point, 1:1 dilution series in which the concentration of A varied from 480 pM to 1 mM. Each titration experiment was simulated in triplicate with a root-mean-square noise of 5 thermophoresis units (randomly sampled from a Gaussian distribution). These triplicates were averaged and fitted using the approaches described in the Theory section. Ten thousand Monte Carlo simulation trials were performed for each set of input parameters; the results are shown in Tables 1 and 2 for 1:2 Micro. model and 1:2 Macro. model, respectively. Both non-monotonic and monotonic binding curves were simulated.

Unsurprisingly, as the K_D 's approach equality (or as α becomes smaller), the ability of the analytic algorithms to estimate the simulated, underlying parameters degrades. With $F_{AB^*A} < F_{AB^*}$ (non-monotonic binding curves; see Fig. 3A, thick line and top panels of Tables 1 and 2), K_D or cooperativity values can be accurately estimated when the factor relating the two K_D 's is 100 or greater ($K_D(2) = 1000$ nM in the 1:2 Micro. model or $\alpha = 25$ in the 1:2 Macro. model). Overall, the 1:2 Macro. model performed worse than the 1:2 Micro. model. At low α values, it failed to reach convergence more often (cf. "Successful trials" in the respective tables). The mean α values are also poorly determined compared to the $K_D(2)$ values calculated by the 1:2 Micro. model.

For monotonic binding curves ($F_{AB^*A} > F_{AB^*}$; Fig. 3A, thin line), both models in general performed poorer than with non-monotonic curves (Tables 1 & 2, lower panels). Neither model achieved 100% convergence over 10,000 trials except 1:2 Micro. with the highest simulated $K_D(2)$ (Table 1). With $K_D(2) < 10$ μ M or $\alpha < 250$, the ability of the models to arrive at the simulated values degraded significantly. It therefore appears that the non-monotonic binding curves perform much better in these analyses than the monotonic ones, and that, in general, the two binding constants must differ substantially (Fig. 3B, darkest line) to allow the accurate determination of the underlying thermodynamic parameters.

One-Step versus Two-Step Fitting

In the literature, it has been suggested that an effective means of analyzing MST data with two inflection points (Fig. 3) is to separate the binding curve into two segments, treating the resulting binding curves as separate 1:1 binding events [8]. Implicit in this approach is that the two K_D 's are very different. To compare the efficacy of this stratagem (a "two-step"

procedure) to that of a single, two-site fitting model (the “one-step” procedure), we simulated fifty 22-point MST binding curves using the 1:2 Micro. paradigm (each curve was the result of averaging three curves that had independent, random noise, as above). We followed the two-step separation procedure and used PALMIST's 1:1 binding model to determine separate K_D 's. The separation of the data sets was achieved “by eye” without regard to the underlying simulated values. We compared these values to those obtained with the one-step 1:2 Micro. analytic model (Table 3).

Two scenarios again were tested; we first examine the results with simulated values of $F_{AB^*} = 800$ and $F_{AB^*A} = 750$ (Fig. 3A, thick line). When viewed in a categorical scatter plot as in Fig. 4A, a significant trend is evident in the two-step analyses: the average refined values of K_D systematically diverge from the simulated values, with $K_D(1)$ skewing lower and $K_D(2)$ higher. Conversely, the refined K_D values for the one-step 1:2 Micro. model stay closer to the simulated values on average. As expected from our simulations described above, the accuracy of both strategies suffered at $K_D(2)$ values of 30 and below.

An important finding of these simulations is that both methods have similar precision, as quantified by the standard deviations of the mean K_D values (Table 3, Fig. 4). This observation only breaks down at low $K_D(2)$ values, where the precision of the K_D 's obtained from the two-step analyses is perhaps slightly better for $K_D(1)$. However, two facts still undermine the adoption of the two-step approach: (1) the overall accuracy of the one-step 1:2 Micro. model is better, and (2) the 1:1 model in the two-step procedure is artificially constrained. Expanding on point (2), the implicit assumption of widely divergent K_D 's is untrue in these cases, and thus parameter correlations are masked. We conclude that the one-step 1:2 Micro. model outperforms the two-step strategy for this set of simulations.

The second scenario that was tested features $F_{AB^*} = 750$ and $F_{AB^*A} = 800$ (Fig. 3, thin line). In this configuration, when the factor between the K_D 's is less than 100, the binding curves appear essentially monophasic (Fig. 3B), and a naïve analyst would not choose a 2-site binding model. Therefore, no analyses on simulations with $K_D(2)$ less than 1000 nM were conducted. Both the one-step 1:2 Micro. model and the two-step strategy arrive at similarly accurate and precise estimates for the $K_D(2)$ values (Fig. 4, Table 3). However, the ability of the two-step approach to converge on an accurate $K_D(1)$ values appears to be poorer than that of the one-step 1:2 Micro. model. Thus, the latter model appears to be superior even in this monotonic case.

The Hill model

With the 1:2 Macro model, as the value of α goes below 1 (it is always positive and nonzero theoretically), by definition the cooperativity in the system is positive. In MST data, this phenomenon sometimes manifests as only a single observable inflection point. A favorable example is shown in Fig. 5 (circles), where $F_{B^*} = 700$, $F_{AB^*} = 725$, $F_{AB^*A} = 750$, $K_{D,M} = 10$ μ M, $B^* = 50$ nM, and $\alpha = 0.01$ (100-fold cooperativity). A naïve observer would likely assume a single binding event and therefore would attempt to use a 1:1 binding model. After the fitting session with the latter model, however, it is clear that there would be substantial systematicity in the residuals (Fig. 5, dashed line). Fitting such data with a Hill model significantly improves the quality of the fit (Fig. 5, solid line). We simulated this two-site-

binding scenario 10,000 times (3 replicates with 5 thermophoresis units of noise) and analyzed each of them with the Hill equation (Eq. (23)). The results indicate that, on average, an analyst using the Hill equation would arrive at an n and EC_{50} of 1.9 and 2.1 μM , respectively. Conservatively interpreted, the analyst would (accurately) conclude that there are at least two binding sites and that the concentration at which 50% of the effect is observed is 2.1 μM .

But what if the analyst were to interpret EC_{50} as the concentration at which 50% of all B^* 's were occupied with two A's (K_{50})? This assumption is demonstrably incorrect; only approximately 46% of B^* 's meet this criterion, with ca. 44% of the B^* 's free. The remaining ~10% of B^* 's have a single A bound. Nonetheless, 46% is not far from 50%, and perhaps this level of error could be tolerated. However, it must be realized that in the above simulation, with $F_{AB^*} - F_{B^*} = 0.5(F_{AB^*A} - F_{B^*})$, the 10% of B^* in AB^* complexes does not contribute adversely to the overall thermophoresis value. However, there is no physical law requiring this proportionality to be true of MST data. Indeed, it seems much more likely that this assumption will be unrealistic. To illustrate this point, we re-performed the above simulation using the same K_D values, but with $F_{AB^*} = 900$. Although the resulting noiseless data would not be perfectly monotonic, this fact might be difficult to discern in noisy data (Fig. 6). On average (using a Monte Carlo simulation), Hill fits to these data yield n and EC_{50} values of 2.0 and 1.0 μM , respectively. But here, only ~18% of B^* 's are doubly occupied with A, and the analyst would be grossly incorrect in equating EC_{50} with K_{50} . Thus, the value of EC_{50} is not fully governed by the underlying K_D values; it is significantly influenced by the relative thermophoresis values of the complexed species. A more extreme example can be simulated in which F_{AB^*} has been set to 1000 (Fig. 6, diamonds). Now, the binding curve has an obviously biphasic appearance. The Hill model cannot accommodate this well; it by necessity ignores the non-monotonicity of the curve (Fig. 6, dashed line). Although the 1:2 Macro. model can arrive at an excellent fit for these data (Fig. 6, dotted line), we found that the refined parameters only loosely resembled the simulated ones, and the 68.3% confidence intervals were large; the refined values were $K_{D,M} = 1.6 [1.0, 4.5] \mu\text{M}$, $\alpha = 0.7 [0.1, 4.8]$, $F_{AB^*} = 782 [772, 851]$.

Even in systems having very high cooperativity ($\alpha = 0.001$), the influence of F_{AB^*} on the result is felt. This may seem counterintuitive, as the population of AB^* is very low at all concentrations of A under these circumstances, and thus a single inflection point would almost always be present in the data. Nonetheless, our Monte Carlo simulations of highly cooperative two-site binding scenarios (Table 4) demonstrate that F_{AB^*} wields an unequivocal influence on the value of EC_{50} , even when the underlying dissociation constants are the same in all simulations. Modest deviations from the “ideal” value of 725 caused significant and reproducible changes in EC_{50} . These aspects cast doubt on the utility of the Hill model in quantitatively evaluating MST data.

In the course of simulating systems with strong positive cooperativity, we found that, contrary to our general recommendation, these can benefit from the univalent species being the labeled one (resulting in A^*). In Supplemental Fig. S4, we show a simulation in which $[A^*] = 50 \text{ nM}$, $F_{A^*B} = 750$, $F_{A^*BA^*} = 800$, $F_{A^*} = 700$, $K_{D,M} = 100 \text{ nM}$, and $\alpha = 0.01$. With these parameters, titrating the bivalent, unlabeled B results in a biphasic, non-monotonic

curve. We performed 10,000 Monte Carlo simulations on this scenario, fitting with the 1:2 Macro. model. The resulting refined binding parameters were well defined by the data (the mean fitted values were $K_{D,M} = 100$ [80,120] nM, $\alpha = 0.012$ [0.006, 0.017]). This result reveals an opportunity for the experimenter: if the data are best fitted by a Hill model when B is the labeled species, it may be beneficial to redo the experiment, labeling A in the second trial (if possible).

The interaction of arrestin-3 N_{IP_6} and IP_6

To examine the performance of the algorithms described above on experimental data, we chose the arrestin-3 protein as one model system. Arrestin-3 was shown to bind the small molecule inositolhexakis phosphate (IP_6) via two structurally distinct sites localized in the N- and C-domain of the molecule [13]. In the N_{IP_6} variant, the N-domain binding site was altered by the following set of mutations: R8Q, K11A, K12A, K108Q, K161Q, K295Q.

The characteristics of the $N_{IP_6}:IP_6$ binding interaction in solution were explored using an MST experiment (Fig. 7). In this case, N_{IP_6} was labeled with a fluorophore and held at a constant concentration of 50 nM, while the concentration of IP_6 was varied. We used a four-titration strategy to cover a large amount of IP_6 -concentration space (Fig. 8, gray-filled circles); in all, 22 IP_6 concentrations were studied, and replicates were averaged. We found that there were measurable, concentration-dependent changes in the “T-Jump” region of the fluorescence traces; this strategy compares the fluorescence prior to the activation of the IR laser to that immediately after activation (see Fig. 9 for an example with a different interaction pair). We used this signal to determine the K_D 's. Clearly, the resulting binding curve (Fig. 7) had two inflection points, indicative of (at least) two classes of binding sites.

Given that N_{IP_6} should be a monomer at the concentration used and that the two binding sites are different [13], we reasoned that the microscopic model was the most realistic two-site binding model to use for analysis. Using a strategy of alternating least-squares the Simplex and Levenberg—Marquardt minimization routines [9], the fitting algorithm quickly arrived at values of $K_D(1) = 20$ nM and $K_D(2) = 14$ μ M, and the fit to the data is excellent (Fig. 7). During the fitting session, an adaptive, automated error-surface projection method [9] was used to determine the 68.3% confidence intervals on these parameters, which were [14, 29] nM and [10, 21] μ M, respectively. Repeated efforts at fitting arrive at the same refined values and confidence intervals, demonstrating the consistency of the method.

Although the macroscopic model is clearly incorrect for the $N_{IP_6}:IP_6$ interaction (i.e. N_{IP_6} is not a multimer with identical IP_6 -binding sites), we decided to use these data to test the performance of the macroscopic binding model. In this model, only a single $K_{D,M}$ is refined, while differences in the two presumably identical binding sites are attributed to cooperativity. We take the approach of refining a “cooperativity factor”, α , which should be 1 in the case of no cooperativity; $\alpha > 1$ indicates negative cooperativity. For the current data, the fitting algorithm arrives at another excellent fit (not shown), with values of 20 [13, 29] nM for K_D and 180 [90, 340] for α . Again, the consistency of the method is demonstrated by identical parameter values being refined for multiple trials.

DNA Aptamer/AMP binding

A second experimental test of the multi-site algorithm used DNA aptamers as a model system. DNA aptamers are oligonucleotides usually designed to bind to a target small molecule. Often, they are the products of *in vitro* selection (or “SELEX”) [27]. The aptamer used in this study (labeled with Cy5 at its 5′ end) was designed to bind to ATP [28], but we use the more stable AMP, which also binds well. We followed the protocol of Entzian & Schubert [15] to monitor the binding of AMP to the aptamer, which was known to bind to two molecules of AMP [29]. We performed three replicates and averaged the resulting data points.

The resulting MST data (Fig. 9) were analyzed using the temperature-jump (T-Jump) region of the thermophoresis traces [9,10]. There was only one apparent inflection point in the data. A 1:1 model refining K_D in PALMIST [9] could be fitted to the data, but the residuals were significant and systematic (Fig. 9); the root-mean-square deviation (r.m.s.d.) of the data from the fit was 0.854 T-Jump units. However, a Hill model (Eq. (23)) implemented in PALMIST could fit the data with more precision (Fig. 9), resulting in an r.m.s.d. of 0.365. The residuals were also less systematic. The refined parameters of the Hill fit were $n = 1.8$ [1.5, 2.1] and $EC_{50} = 58$ [52, 66] μM . Notably, characterization of the interaction using a different set of analysis zones (i.e. using “Thermophoresis” instead of “T-Jump”; Fig. 9) leads to a different value for EC_{50} (132 [122, 142] μM); this comports with our earlier conclusion [9] that fitted parameters can vary depending on the analysis zones, and, when possible, T-Jump data should be analyzed to the exclusion of other potential MST analytic modes. The 1:2 Micro. model could also be fitted to the data (not shown) with an excellent r.m.s.d. (0.388 T-Jump units), but the refined K_D 's were identical and had large 68.3% confidence intervals ($K_D(1) = K_D(2) = 20$ [4, 40] μM), indicating that these values are likely not well defined by the data.

Thus, the Hill binding model can be used to fit data well in this context. However, as noted above, the EC_{50} value carries a tenuous physical relationship to the species populations in solution. In other words, the EC_{50} should be considered to reveal only qualitative information regarding the associated binding constants.

Discussion

Numerous recent applications have demonstrated that MST can be an effective means of measuring K_D values for the interactions of biological macromolecules [8,9,30,31]. In this work, we have extended the number of models that may be used to analyze MST data in which two binding events are evident. In the past, software designed to analyze MST data could only approach accurately fitting such data by the exclusion of approximately half of the binding curve [8]. In contrast, here we have adapted a microscopic binding model capable of simultaneous refinement of $K_D(1)$ and $K_D(2)$ for MST data. Further, we have also introduced a macroscopic model that assumes identical binding sites and refines the $K_{D,M}$ for the first binding event and the cooperativity (α) for the second. The choice of model should incorporate the experimenter's knowledge of the species under study: for example, if divalent B is a two-fold symmetric homodimer, then the binding sites are identical, and the macroscopic model is most appropriate. For a monomeric B with two non-identical binding sites, the microscopic model should be chosen.

Our simulations of the microscopic and macroscopic binding models demonstrated the performance and limits of these binding scenarios under typical MST assay conditions. These findings have a significant impact on experimental planning. First, they indicated which molecule should be labeled if one is known to be bivalent. The labeled molecule—which is to be held at a constant, low concentration—must (in most cases) be the bivalent one for a successful simultaneous assessment of both K_D 's or $K_{D,M}$ and α (Fig. 1). The opposite titration configuration (Fig. 2), i.e. labeling the univalent molecule and treating the bivalent partner as the titrant, could only report on the lower of the two K_D 's (or, in the macroscopic view, it can only report on the first binding event). It is therefore possible for non-unitary stoichiometry to remain unrecognized in a standard MST experiment, and analysts should bring outside information to bear on the problem before assuming that their interactions feature only a single binding site. Although some systems may be amenable to stoichiometry determination using MST under different experimental conditions [10], this task is usually better accomplished with other methods, e.g. AUC [32,33] or ITC [2].

The simulations suggest that two visible, distinct inflection points are necessary for a stable simultaneous quantification of two thermodynamic binding parameters (Tables 1 & 2). While our simulations do not represent an exhaustive exploration of concentration space, preliminary additional simulations in likely MST configurations (not shown) imply that the need for two inflection points generally holds. Operationally, this means that the two inflection points must be well separated in concentration space (e.g. Fig. 3). However, another characteristic of the binding curves was identified that had a significant impact on the reliability of the algorithms tested. This aspect was the nature of the signal from the doubly bound B^* (i.e. the AB^*A complex). If its thermophoresis (or T-Jump signal or initial, pre-IR fluorescence signal) causes a reversal (Fig. 3, thick line) in the binding-curve trend observed for the singly bound B^* (the AB^* complex), this trend reversal is inherently informative and adds to the algorithm's ability to discriminate between the two binding events, thus enhancing the experimenter's ability to determine reliable refined parameters. The obvious converse of this observation is that monotonic binding curves (Fig. 3A, thin line) present more difficulties in these models.

Our experience with the $N_{IP6}:IP_6$ interaction has brought other experimental design concepts to light. For example, a standard, commercial MST cassette offers only 16 positions for capillary tubes. In a typical 1:1 serial dilution series, a concentration range of more than four decades can therefore be explored. However, in a two-site system with the microscopic K_D 's widely separated (or exhibiting a strong degree of negative cooperativity), even this concentration range may not be broad enough to allow the accurate refinement of parameters. The experimenter is thus faced with the choice of covering the high end or low end of the binding curve. While different dilution ratios can offer more coverage within the 16-capillary limitation, these sample concentration space more sparsely, and are therefore unsatisfactory. Moreover, any dilution strategy must balance the desirability of experimental replicates with the expense of the consumable capillary tubes. For the $N_{IP6}:IP_6$ interaction, we devised a scheme (Fig. 8, left) that utilizes three overlapping, 16-capillary titrations, all having a 1:1 dilution strategy except for the last capillary in each replicate, which had no ligand. A fourth titration covered the ends of concentration space that were neglected in the first three. The result was a 22-point titration in which 16 concentrations were sampled in

triplicate and two in quadruplicate, with the remaining four points sampled twice. This scheme allows roughly six decades of concentration space to be sampled, with a no-ligand control sampled three times. An experimenter can, of course, modify this strategy to cover other areas of interest or alter the sampling. For example, a simple modification of this schema results in the elimination of the quadruplicates and only two concentrations being sampled twice (Fig. 8, right).

Another experimental aspect that should be addressed in two-site binding systems is the possibility that the second observed inflection point is the result of factors other than ligand binding. For example, in systems that are known to have only one binding site, a sharp reversal of the trend in the binding curve can manifest at high protein titrant concentrations [9]. We have speculated that this could be due to hydrodynamic nonideality exacerbated by the high protein concentrations encountered in such experiments. For suspected two-site systems, a control experiment can be engineered to effectively eliminate this possibility: the titration can be performed with a chemically similar, non-binding ligand. If no strong trends are observed, then both inflection points are probably consequent to ligand binding and not non-specific solution effects. In the case of the $N_{IP6}:IP_6$ interaction, this control is probably not necessary, as we have never observed aberrant trend reversals in the presence of a small-molecule ligand. Indeed, we found that sodium sulfate, an analog of IP_6 , did not elicit a strong trend in the T-Jump data with wild-type arrestin-3 (Supplemental Fig. S5).

It is worth noting that many of the principles discussed here in the context of MST assays hold for other binding assays as well. A subset of ligand-binding assays (“tracer” assays) utilize titrations in which one labeled species is held at a very low concentration while an assay-silent binding partner is titrated; these include (but are not limited to) fluorescence-quenching and fluorescence anisotropy experiments. Experienced practitioners of these methods are doubtless aware of the pitfalls described herein, but those new to these methods can benefit from the experimental-design principles enumerated above.

A generalized multi-site model that is widely used to analyze ligand-binding curves is the Hill model [23]. The Hill equation, restated in thermophoretic form in Eq. (23), was developed to study the interaction of molecular oxygen with hemoglobin in classic studies on cooperativity in ligand binding [11]. Indeed, the Hill equation assumes such extreme positive cooperativity that no intermediate forms are accounted for in the analysis [23]. The data that Hill was analyzing were presented in terms of the percent saturation of blood or hemoglobin with the ligand, O_2 . Hence, the ordinate was directly proportional to the amount of ligand bound, and obviously the trend in the binding curves was monotonic. Under these circumstances, it is likely that the contribution of intermediate forms to the overall signal does not severely impact the quantitative conclusions that may be drawn from a Hill analysis. The species in an MST experiment, however, are not constrained to have proportional signals (see Table 4 and Figs. 3 & 6), and thus the physical meaning of EC_{50} in thermophoretic experiments is suspect, and is almost certainly different from EC_{50} s measured using truly proportional methods. We recommend that good Hill fits to MST data be treated as phenomenological descriptions of the binding properties, but nothing further. Hill fits that are demonstrably better than 1:1 fits point to an opportunity for the experimenter, however. If it is feasible to label the univalent species and treat the bivalent

species as the titrant, it is possible to fit the resulting titration data with the 1:2 Macro. model (Supplemental Fig. S4).

As stated above, the 1:2 Micro. model cannot be used to analyze 1:2 binding systems in which the binding sites are asymmetric and exhibit cooperativity. In such a case, the binding constants and the cooperativity parameter relating them would be too correlated to allow convergence on a rational result. There are experimental strategies that could result in a cogent analysis for such a system. For example, individual mutagenesis of the two binding sites on B* would enable independent determinations of the microscopic K_D 's; these could be fixed in an analysis using data obtained from the wild-type B*. Although we have not implemented such a model, it would be straightforward to do so.

Could the principles described herein be extended to quantify the binding properties of molecules harboring three or more binding sites for a ligand? This possibility exists but seems very unlikely. If numeric K_D 's are desired, then the number of distinct phases in the curves would have to be equal to the number of binding sites. Taking the case of a three-site molecule, this would imply that all three binding sites have very different microscopic binding constants or significantly negative cooperativity. Moreover, to provide adequate information on all binding sites, we estimate that the necessary concentration range would have to span more than nine decades, which seems impracticable except in the case of readily soluble small molecules.

We have incorporated the algorithms and strategies described above into version 1.2 of the freeware program PALMIST. In the program's "Expert Mode", the 1:2 Micro., 1:2 Macro, and Hill models are available for analysis of MST data. Further, the non-Hill models can be used for Monte Carlo or bootstrap-type simulations to help in experimental planning. The program automatically finds rigorous confidence intervals for all fitted parameters where possible. Additionally, the software's experimental-planning and documentation features have been extended to the new models. It can be downloaded from <http://biophysics.swmed.edu/MBR/software.html>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Part of the research in this report was supported by grants to VVG (GM077561, GM109955, (these two R01 grants were replaced by R35 GM122491), and EY011500) and TMI (GM120569, GM061606, AI06987, and DA043680) from the National Institutes of Health, a Vanderbilt Discovery Grant (TMI/VVG), and the Vanderbilt Core Grant in Vision Research (P30EY008126).

References

1. Zhao H, Brautigam CA, Ghirlando R, Schuck P. Overview of Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation. *Curr Protoc Protein Sci.* 2013; 71:20.12.1–20.12.49.
2. Pierce MM, Raman CS, Nall BT. Isothermal titration calorimetry of protein-protein interactions. *Methods.* 1999; 19:213–21. [PubMed: 10527727]

3. Arac D, Murphy T, Rizo J. Facile detection of protein-protein interactions by one-dimensional NMR spectroscopy. *Biochemistry*. 2003; 42:2774–2780. [PubMed: 12627942]
4. Schuck P. Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu Rev Biophys Biomol Struct*. 1997; 26:541–566. [PubMed: 9241429]
5. Pollard TD. A guide to simple and informative binding assays. *Mol Biol Cell*. 2010; 21:4061–4067. [PubMed: 21115850]
6. Hulme EC, Trevethick MA. Ligand binding assays at equilibrium: Validation and interpretation. *Br J Pharmacol*. 2010; 161:1219–1237. [PubMed: 20132208]
7. Hanson SM, Francis DJ, Vishnivetskiy SA, Klug CS, Gurevich VV. Visual arrestin binding to microtubules involves a distinct conformational change. *J Biol Chem*. 2006; 281:9765–9772. [PubMed: 16461350]
8. Seidel SAI, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, Lazic A, et al. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*. 2013; 59:301–315. [PubMed: 23270813]
9. Scheuermann TH, Padrick SB, Gardner KH, Brautigam CA. On the acquisition and analysis of microscale thermophoresis data. *Anal Biochem*. 2016; 496:79–83. [PubMed: 26739938]
10. Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P, et al. MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct*. 2014; 1077:101–113.
11. Hill AV. The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. *Biochem J*. 1913; 7:471–480. [PubMed: 16742267]
12. Gurevich VV, Benovic JL. Visual arrestin binding to rhodopsin. *J Biol Chem*. 1995; 270:6010–6016. [PubMed: 7890732]
13. Chen Q, Perry NA, Vishnivetskiy SA, Gilbert NC, Zhuo Y, Berndt S, et al. Structural basis of arrestin-3 activation and signaling. *Nat Commun*. 2017 in press.
14. Vishnivetskiy, SA., Zhan, X., Chen, Q., Iverson, TM., Gurevich, VV. *Curr Protoc Pharmacol*. John Wiley & Sons, Inc; 2001. Arrestin Expression in *E. coli* and Purification.
15. Entzian C, Schubert T. Mapping the binding site of an aptamer on ATP using microscale thermophoresis. *J Vis Exp*. 2017:e55070.
16. Brautigam CA. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol*. 2015; 562:109–134. [PubMed: 26412649]
17. Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun*. 2010; 1:100. [PubMed: 20981028]
18. Wang ZX, Jiang RF. A novel two-site binding equation presented in terms of the total ligand concentration. *FEBS Lett*. 1996; 392:245–249. [PubMed: 8774854]
19. Burington, RS. *Handbook of Mathematical Tables and Formulas*. 5th. McGraw-Hill Book Company; New York: 1973.
20. Freire E, Schön A, Velazquez-Campoy A. Isothermal titration calorimetry: general formalism using binding polynomials. *Methods Enzymol*. 2009; 455:127–55. [PubMed: 19289205]
21. Vega S, Abian O, Velazquez-Campoy A. A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry. *Methods*. 2015; 76:99–115. [PubMed: 25305413]
22. Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, et al. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci*. 2007; 16:30–42. [PubMed: 17192587]
23. Weiss JN. The Hill equation revisited: uses and misuses. *FASEB J*. 1997; 11:835–841. [PubMed: 9285481]
24. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Q J Appl Math*. 1944; II:164–168.
25. Marquardt DW. An algorithm for the least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math*. 1963; 11:431–441.
26. Nelder J, Mead R. A simplex method for function minimization. *Comput J*. 1964; 7:308–313.

27. Breaker RR. DNA aptamers and DNA enzymes. *Curr Opin Chem Biol.* 1997; 1:26–31. [PubMed: 9667831]
28. Huizenga DE, Szostak JW. A DNA Aptamer That Binds Adenosine and ATP. *Biochemistry.* 1995; 34:656–665. [PubMed: 7819261]
29. Lin CH, Patel DJ. Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. *Chem Biol.* 1997; 4:817–832. [PubMed: 9384529]
30. Lippok S, Seidel SAI, Duhr S, Holthoff HP, Jenne D, Braun D. Direct detection of antibody concentration and affinity in human serum using microscale thermophoresis. *Anal Chem.* 2012; 84:3523–3530. [PubMed: 22397688]
31. Seidel SAI, Wienken CJ, Geissler S, Jerabek-Willemsen M, Duhr S, Reiter A, et al. Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angew Chemie - Int Ed.* 2012; 51:10656–10659.
32. Balbo A, Minor KH, Velikovsky CA, Mariuzza RA, Peterson CB, Schuck P. Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc Natl Acad Sci U S A.* 2005; 102:81–86. [PubMed: 15613487]
33. Padrick SB, Brautigam CA. Evaluating the stoichiometry of macromolecular complexes using multisignal sedimentation velocity. *Methods.* 2011; 54:39–55. [PubMed: 21256217]

Abbreviations

1:2 Macro	a two-site macroscopic fitting model
1:2 Micro	a two-site microscopic fitting model
AUC	analytical ultracentrifugation
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene-bis(oxyethylenitrilo)tetraacetic acid
EPR	electron paramagnetic resonance
IP₆	inositolhexakis phosphate
IR	infrared
ITC	isothermal titration calorimetry
NMR	nuclear magnetic resonance
MST	microscale thermophoresis

Figure 1. Data and population plot for a labeled bivalent species

In circles and the thick line are shown a simulation with $F_{B^*} = 700$, $F_{AB^*} = F_{B^*A} = 800$, $F_{AB^*A} = 750$, $K_D(1) = 10$ nM, and $K_D(2) = 10$ μ M. $[B^*]$ is held constant at 50 nM. The fractional populations of the species depicted in the inset are shown, with values described on the right-hand y-axis.

Figure 2. Data and population plot for a labeled univalent species

In circles and the thick line are shown a simulation with $F_{A^*} = 700$, $F_{A^*B} = F_{BA^*} = 800$, $F_{A^*BA^*} = 750$, $K_D(1) = 10$ nM, and $K_D(2) = 10$ μ M. $[A^*]$ is held constant at 50 nM. The fractional populations of the species depicted in the inset are shown, with values described on the right-hand y-axis.

Figure 3. Monotonic and non-monotonic MST binding curves

(A) A comparison of non-monotonic and monotonic binding curves. A 1:2 Micro. simulation is shown for both lines, with $K_D(1) = 10$ nM, and $K_D(2) = 10$ μ M. In the thick line and circles is shown a simulation with $F_{B^*} = 700$, $F_{AB^*} = 800$, and $F_{AB^*A} = 750$, while the thin line and squares depict a simulation featuring $F_{B^*} = 700$, $F_{AB^*} = F_{B^*A} = 750$, and $F_{AB^*A} = 800$. $[B^*]$ was held constant at 50 nM. (B) Simulations of monotonic binding curves with differing $K_D(2)$'s. The same monotonic simulation as shown in part A is shown, except progressively lighter shades of the lines represents the different $K_D(2)$'s (shown in nM units in the inset legend) that were simulated, holding $K_D(1) = 10$ nM.

Figure 4. Results for the one-step and two-step analyses for simulated MST data

(A) Results for the simulated non-monotonic binding curves. Shown are comparisons of K_D 's estimated by 1:2 Micro. model (one-step fitting) or by two iterations of the 1:1 model (two-step fitting). Each column represents one set of data simulations. The two K_D values used to simulate data sets are indicated with dashed lines in the plots; all the $K_D(1)$'s were set to 10 nM, and $K_D(2)$'s were set to 10,000, 3,000, 1,000, 300, or 100 nM. The estimated $K_D(1)$'s and $K_D(2)$'s from fitting the curves are shown as solid circles and empty circles, respectively, and the average values of the ten result points are represented by solid lines. (A) The scenario of $F_{B^*} = 700$, $F_{AB^*} = 800$, and $F_{AB^*A} = 750$ (a non-monotonic curve). (B) The scenario of $F_{B^*} = 700$, $F_{AB^*} = 750$, and $F_{AB^*A} = 800$ (a monotonic curve). Data sets for $K_D(2) = 300$ and 100 nM were not tested, and a naïve observer would likely assume single-site binding for these.

Figure 5. A comparison of 1:1 and Hill fits

A positively cooperative binding curve is simulated with noise (circles) in the upper plot. Shown also are two fits to these data as described in the inset. The lower plot shows the residuals for these two fits: the residuals for the Hill fit are shown as circles, while those for the 1:1 fit are stars. Note the strong systematicity near to 1,000 nM in the 1:1 fit.

Figure 6. Hill and 1:2 Macro. Fits to simulated data

Simulated data are depicted as markers in the upper plot, using the 1:2 Macro. simulation with $[B^*] = 50$ nM, $K_{D,M} = 10$ μ M, $\alpha = 0.01$, $F_{B^*} = 700$, $F_{AB^*A} = 750$, and F_{AB^*} as shown in the inset. Solid and dashed lines represent fits using the Hill model to simulations with $F_{AB^*} = 900$ and $F_{AB^*} = 1,000$, respectively. The dotted line is a 1:2 Macro. fit to the latter data set. The lower plot shows residuals to the fits, with markers as indicated in the inset, except the open diamonds represent the residuals between the $F_{AB^*} = 1,000$ data and the 1:2 Macro. fit.

Figure 7. N_{IP_6} binding to IP_6 monitored by fluorescence in an MST instrument

The upper plot shows squares with error bars depicting the means and standard deviations of the means, respectively, of two, three, or four independent determinations of T-Jump values as a function of $[IP_6]$ (see also Fig. 8, left). The line is the 1:2 Micro. fit to these data, resulting in $K_D(1) = 20$ nM and $K_D(2) = 14$ μ M. The lower plot shows the residuals between the data and the fit.

Figure 8. Four-experiment, 22-point titration strategies

On the left is shown the strategy used for the N_{IP6}/IP_6 interaction shown in Fig. 7. Each filled-in circle represents a capillary at the given concentration (far left). The first three replicates have staggered top concentrations, while the fourth has duplicates at under-represented concentrations. The strategy results in two quadruplicates (12 and 6 nM), three duplicates (100, 50, and 25 μM), and seventeen triplicates (all remaining concentrations). On the right, an alternative (“optimized”) strategy with open circles representing the capillaries is shown. The same number of capillaries is shown, but now there are no quadruplicates and only one duplicate (25 μM).

Figure 9. MST data for AMP binding to a DNA aptamer

The upper plot shows the thermophoresis traces (three experiments are overlaid), and the gray boxes represent the analysis zones that could be used for analysis. The left and middle boxes could be compared for T-Jump, and the middle and right boxes for Thermophoresis. T-Jump was used to derive the data points (open circles, middle plot). These were the average of the three replicates, and the error bars show the standard error of these means. These data were fitted with a Hill model or a 1:1 K_D model, as indicated in the inset legend. Residuals to the Hill model are also shown open circles in the lower plot, whereas the residuals to the 1:1 model are closed triangles.

Table 1

Monte Carlo results for the 1:2 Micro. simulations

Simulated values $K_D(1) = 10$ nM; $K_D(2)$ (nM)		100	300	1000	3000	10,000
$F_{AB}^* = 700, F_{AB}^{**} = 800, F_{AB}^{**A} = 750$; non-monotonic						
Successful trials		9,445	9,985	10,000	10,000	10,000
$K_D(1)$						
Mean value		80	200	11	10	10
C.I.		[10, 160] [*]	[0, 400]	[7, 14]	[8, 13]	[8, 12]
$K_D(2)$						
Mean Value		70	200	1,000	3,100	10,000
C.I.		[10, 140]	[0, 400]	[700, 1400]	[2400, 3800]	[8000, 12000]
$F_{AB}^* = 700, F_{AB}^{**} = 750, F_{AB}^{**A} = 800$; monotonic						
Successful trials		9399	8313	9756	9753	10,000
$K_D(1)$						
Mean value		89	218	5	5	11
C.I.		[73, 103]	[192, 256]	[0, 13]	[0, 12]	[6, 15]
$K_D(2)$						
Mean Value		0.286	3.3E9	900	2,500	10,000
C. I.		[0.0, 0.0]	[2.1E9, 4.1E9]	[600, 1100]	[1800, 3300]	[8000, 12,000]

^{*} Bracketed values are 68.3% confidence intervals from the Monte Carlo results.

Table 2

Monte Carlo results for the 1:2 Macro simulations

Simulated values $K_{D,M} = 10$ nM; α						
<div>2.57.52575250</div>						
$F_{B^*} = 700, F_{AB^*} = 800, F_{AB^*A} = 750$; non-monotonic						
Successful trials		8,629	9,874	10,000	10,000	10,000
$K_{D,M}$						
Mean value		10	12	11	10	10
C.I.		[4, 16]*	[6, 18]	[7, 14]	[8, 13]	[8, 12]
α						
Mean Value		12	12	30	80	270
C.I.		[1, 14]	[3, 19]	[10, 50]	[50, 120]	[170, 360]
$F_{B^*} = 700, F_{AB^*} = 750, F_{AB^*A} = 800$; monotonic						
Successful trials		7,359	7,812	9,607	9,974	9,998
$K_{D,M}$						
Mean value		94	144	12	11	10
C.I.		[87, 102]	[153, 161]	[4, 20]	[5, 16]	[6, 15]
α						
Mean Value		1.05E6	6.11E5	40	100	290
C. I.		[1.07E6, 1.08E6]	[6.00E5, 7.05E5]	[20, 50]	[50, 130]	[180, 390]

^{*} Bracketed values are 68.3% confidence intervals from the Monte Carlo results.

Table 3

Results of one- and two-step analyses

Simulated Values $K_D(1) = 10$ nM; $K_D(2)$ (nM)		100	300	1000	3000	10,000
$F_{AB}^* = 700, F_{AB}^{*A} = 800, F_{AB}^{*A} = 750$; non-monotonic						
One-step curve fit						
$K_D(1)$		$9 \pm 6^*$	12 ± 7	12 ± 3	9 ± 2	9 ± 2
$K_D(2)$		140 ± 70	357 ± 177	930 ± 210	3110 ± 570	$10,000 \pm 2000$
Two-step curve fit						
$K_D(1)$		4 ± 2	9 ± 6	7.2 ± 2.3	8 ± 2	9 ± 2
$K_D(2)$		170 ± 60	500 ± 200	1100 ± 300	3600 ± 800	$12,100 \pm 3000$
$F_{AB}^* = 700, F_{AB}^{*A} = 750, F_{AB}^{*A} = 800$; monotonic						
One-step curve fit						
$K_D(1)$		-	-	12 ± 5	11 ± 7	8 ± 3
$K_D(2)$		-	-	1200 ± 200	3200 ± 900	$11,400 \pm 1600$
Two-step curve fit						
$K_D(1)$		-	-	23 ± 8	19 ± 9	10 ± 5
$K_D(2)$		-	-	1080 ± 160	2800 ± 800	$10,500 \pm 1400$

* Values are presented as the mean plus or minus the standard deviation of the mean

Table 4
Results of highly cooperative binding analyzed with the Hill model

Simulated F_{AB^*}	Refined EC_{50} (nM)	Refined n
600 [‡]	810 [730, 880] [‡]	2.0 [1.7, 2.3]
725	680 [620, 750]	2.0 [1.7, 2.3]
900	540 [490, 590]	2.0 [1.7, 2.3]
1000	470 [420, 520]	2.0 [1.7, 2.4]

[‡]All simulations featured $F_{B^*} = 700$, $F_{AB^*A} = 750$, $K_{D,M} = 10 \mu\text{M}$, $B^* = 50 \text{ nM}$, and $\alpha = 0.001$.

[‡]Values in brackets represent the 68.3% confidence interval from the Monte Carlo simulation.