Structural identifiability of equilibrium ligand-binding parameters

Thomas R. Middendorf[1,2] and Richard W. Aldrich[1,2]

[1]Center for Learning and Memory and [2]Department of Neuroscience, University of Texas at Austin, Austin, TX 78712

Understanding the interactions of proteins with their ligands requires knowledge of molecular properties, such as binding site affinities and the effects that binding at one site exerts on binding at other sites (cooperativity). These properties cannot be measured directly and are usually estimated by fitting binding data with models that contain these quantities as parameters. In this study, we present a general method for answering the critical question of whether these parameters are identifiable (i.e., whether their estimates are accurate and unique). In cases in which parameter estimates are not unique, our analysis provides insight into the fundamental causes of nonidentifiability. This approach can thus serve as a guide for the proper design and analysis of protein–ligand binding experiments. We show that the equilibrium total binding relation can be reduced to a conserved mathematical form for all models composed solely of bimolecular association reactions and to a related, conserved form for all models composed of arbitrary combinations of binding and conformational equilibria. This canonical mathematical structure implies a universal parameterization of the binding relation that is consistent with virtually any physically reasonable binding model, for proteins with any number of binding sites. Matrix algebraic methods are used to prove that these universal parameter sets are structurally identifiable (SI; i.e., identifiable under conditions of noiseless data). A general approach for assessing and understanding the factors governing practical identifiability (i.e., the identifiability under conditions of real, noisy data) of these SI parameter sets is presented in the companion paper by Middendorf and Aldrich (2017. *J. Gen. Physiol.* https://doi.org/10.1085/jgp.201611703).

INTRODUCTION

One of the major functions of proteins is to bind other molecules. These binding reactions serve a variety of cellular functions, including buffering, transport, and signal transduction. Protein-binding ligands include a wide variety of chemical species, such as metal ions, peptides, other proteins, nucleotides, neurotransmitters, hormones, and nonbiological targets such as pharmaceuticals.

In some situations, such as high-throughput screening studies of protein–drug interactions, it may be sufficient to characterize protein–ligand binding using an empirical factor obtained directly from the binding curve, such as $K_{1/2}$, the half-saturating ligand concentration. However, for the large and important class of proteins containing multiple ligand-binding sites, the binding mechanism may be complex, and its elucidation may require quantitation of factors such as differences in the intrinsic ligand affinities of the sites, dependence of the site affinities on the conformation of the macromolecule, and cooperative interactions between the sites. These mechanistic parameters cannot be measured directly, but rather must be estimated by fitting a quantitative model to binding and/or conformation data.

For systems composed of multiple coupled equilibria, parameter estimation may be compromised by correlations between the parameters. For some combinations of model and data, parameter compensations during fitting may be sufficiently effective that multiple parameter sets provide equally good fits to the experimental data. If the range of parameter values spanned in these sets is large enough, then little or no knowledge is gained about the system under study. In such situations, progress requires either improvements in the data quality or else the adoption of alternative experimental approaches that provide stronger constraints on the parameter values.

An example of failed parameter estimation is shown in Fig. 1. The simple cooperative binding model (Fig. 1 A) represents a receptor that occupies a single conformational state and contains two (possibly nonidentical) binding sites. The three model parameters are the microscopic association equilibrium constants K_I and K_{II} for binding to sites I and II and an interaction (cooperativity) factor f that quantifies the fold change in a site-binding constant when the adjacent site is occupied by ligand. Detailed balance requires that $K_I\, f\, K_{II} = K_{II}\, f\, K_I$, so there is only one cooperativity factor for this model. (Cooperative interactions caused by unequal ligand affinities of a site in different conformations of a

Correspondence to Richard W. Aldrich: raldrich@austin.utexas.edu

Abbreviations used: PI, practically identifiable; SI, structurally identifiable.

The Rockefeller University Press $30.00
J. Gen. Physiol. 2017 Vol. 149 No. 1 105–119
https://doi.org/10.1085/jgp.201611702

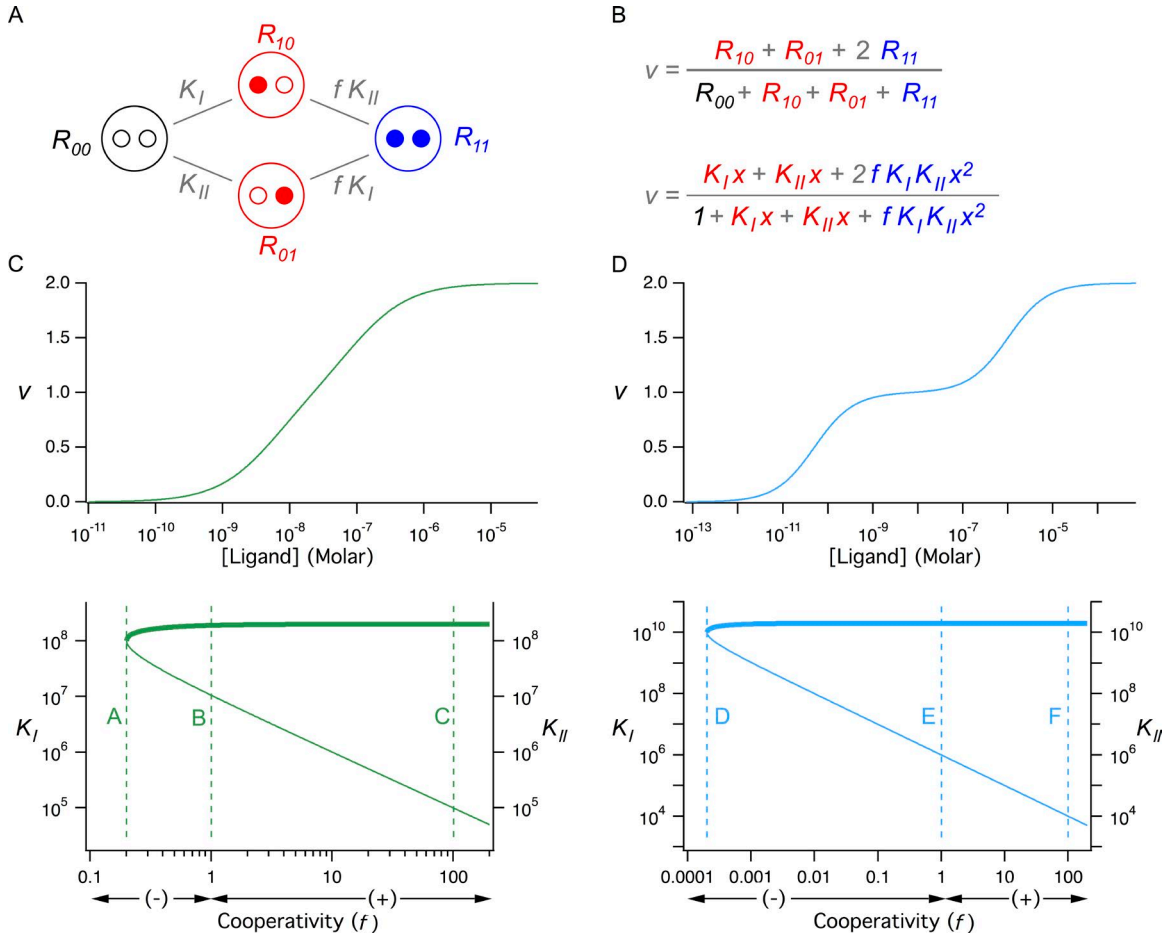

Figure 1. **Parameters of two-site allosteric model are not SI when constrained by total binding data.** (A) Diagram of two-site allosteric model. Large circles represent ligation states of the system. Small circles represent binding sites I (left) and II (right). States are designated by symbols R_{ij}, where i (j) is equal to 1 or 0 depending on whether site I (site II) is bound or not bound by ligand, respectively. States with zero, one, or two bound ligands are color coded black, red, and blue, respectively. Closed and open circles represent bound and unbound sites, respectively. Model parameters are the microscopic association equilibrium constants K_I and K_{II} for sites I and II, respectively, and cooperativity factor f. (B, top) Equation relating mean number of bound ligands, v, to concentrations of the ligated states. (bottom) Equation relating v to free ligand concentration, x, and model parameters. In the top and bottom, terms arising from states with zero, one, and two bound ligands are color coded black, red, and blue, respectively. (C, top) Simulated binding curve computed from the top equation in B using parameter values $\{f, K_I, K_{II}\} = \{10.017, 1.0034 \times 10^6\ M^{-1}, 1.99 \times 10^8\ M^{-1}\}$. Parameters were chosen to produce a relatively "unstructured" binding curve. (bottom) Locus of all parameter triples $\{f, K_I, K_{II}\}$ that yield total binding curve identical to the curve shown above. The curve was computed using Eqs. 12a and 12b. Parameter triples are determined by taking vertical lines, which determine the value of f, and their intersections with the bold and light green curves, which determine parameters K_I and K_{II}. Because of the symmetric appearance of K_I and K_{II} in the binding equation (B, bottom), the bold and light curves may correspond to either K_I and K_{II} or K_{II} and K_I, respectively. Dashed lines marked "A," "B," and "C" correspond to parameter triples $\{0.2, 10^8\ M^{-1}, 10^8\ M^{-1}\}$, $\{1.0086, 1.0462 \times 10^7\ M^{-1}, 1.8954 \times 10^8\ M^{-1}\}$, and $\{100.86, 99{,}193\ M^{-1}, 2.0 \times 10^8\ M^{-1}\}$, respectively. Arrows on abscissa delineate regions of negative ($f < 1$), zero ($f = 1$), and positive ($f > 1$) cooperativity. (D, top) Simulated total binding curve computed from the top equation in B using parameter values $\{f, K_I, K_{II}\} = \{0.10086, 1.999 \times 10^{10}\ M^{-1}, 9.9193 \times 10^6\ M^{-1}\}$. Parameters were chosen to produce a more "structured" binding curve than the one in the top of C. (bottom) Locus of all parameter triples that yield total binding curves identical to curve shown above. The curve was computed using Eqs. 12a and 12b. Dashed lines marked "D," "E," and "F" correspond to parameter triples $\{0.0002, 10^{10}\ M^{-1}, 10^{10}\ M^{-1}\}$, $\{1.0017, 2.0 \times 10^{10}\ M^{-1}, 9.9837 \times 10^5\ M^{-1}\}$, and $\{100.17, 2 \times 10^{10}\ M^{-1}, 9{,}983.2\ M^{-1}\}$, respectively.

macromolecule are also possible but are not considered in the model in Fig. 1 A.)

Many commonly used techniques, such as those based on uptake of radioactive ligands and calorimetric methods, do not provide distinguishable signals when ligands bind to distinct sites in a protein, but rather yield the total binding relation: the mean number of ligands bound to the protein as a function of ligand concentration. The total binding relation is the ratio of the concentration of bound ligands to the concentration of protein (Fig. 1 B, top). For the model in Fig. 1 A, this relation can be reexpressed in terms of the model parameters (Fig. 1 B, bottom). The binding curve in Fig. 1 C (top) was calculated from this relation using

specific values of the model parameters. However, an infinite set of parameter triples {f, K_I, K_{II}} (Fig. 1 C, bottom) yield total binding curves identical to this curve. (Note that the x axis of Fig. 1 C [bottom] is truncated; the locus of parameter triples continues to infinitely large values of f.) The dashed lines A, B, and C identify three parameter triples from this set that correspond to very different binding mechanisms. For parameter set A, sites I and II have identical affinities and interact with moderate negative cooperativity ($f < 1$); for parameter set B, the affinities differ by ∼20-fold, with no interaction between the sites ($f \sim 1$); for parameter set C, the site affinities differ by a factor of 2,000 and interact with strong positive cooperativity ($f = 100$). Because K_I and K_{II} appear symmetrically in the equation for total binding (Fig. 1 B, bottom), there is a further ambiguity in the relative magnitudes of the site affinities: the simulated curves are unaffected if the values of K_I and K_{II} are interchanged.

It might be argued that the curve in Fig. 1 C (top) is a "pathological" case that places anomalously poor constraints on the parameters. However, there is also an infinite locus of parameter values (Fig. 1 D, bottom) corresponding to the more highly structured synthetic binding curve in Fig. 1 D (top). The curves in Fig. 1 (C and D, bottom) indicate that, at best, fits to binding data can place a lower limit on f and an upper limit on the larger of K_I and K_{II}. The infinite range of parameter values yielding identical binding curves means that little or no mechanistic insight can be gained from total binding data analyzed using the two-site allosteric model (Fig. 1 A).

The simulations in Fig. 1 show that obtaining a good fit of a model to binding data provides no assurance that the estimated parameters are accurate or unique, even for simple models with a small number of parameters. In these situations, the parameters are deemed "not identifiable" (Ljung, 1987; Walter and Pronzato, 1997). It is useful to distinguish two aspects of parameter identifiability (Raue et al., 2009). Structural identifiability is an intrinsic mathematical property of a given model and the data to be fitted (Bellman and Åström, 1970; Audoly et al., 2001; Hengl et al., 2007; Chis et al., 2011). Parameters are structurally identifiable (SI) if they can be estimated accurately and uniquely from noiseless, bias-free data of a specified type. For example, the parameters of the model in Fig. 1 are not SI when constrained by total binding data because the synthetic data curves are fit exactly by an infinite number of parameter triples, for which the model parameters span infinite ranges. The unavoidable presence of noise in real experimental data adds to the uncertainty in fit-derived parameter values; parameters are practically identifiable (PI) if this added uncertainty is of an acceptable magnitude (Raue et al., 2009).

Several methods can be used to rigorously quantify the uncertainty in parameters estimated by fitting models to data. These include simulations, as in Fig. 1 (see also Colquhoun [1969]) and calculations of likelihood intervals (Colquhoun and Sigworth, 1983; Colquhoun and Ogden, 1988; Edwards, 1992) or Bayesian posterior distributions (Hines et al., 2014; Epstein et al., 2016). However, these approaches have important limitations. First, the "black box" nature of numerical methods tends to obscure the underlying features of the model and data that determine whether parameters are identifiable, particularly when there are multiple, correlated parameters. Second, the brute-force approach of mapping the entire error surface becomes computationally unreasonable for models with more than a few parameters. Finally, the estimation of parameters and their uncertainties must be repeated for each of the (possibly large number of) models under consideration. To overcome these limitations, we develop in this study and in a companion paper (see Middendorf and Aldrich in this issue) a simple, systematic, and essentially model-independent approach to assessing and understanding parameter identifiability for macromolecule–ligand binding systems at equilibrium.

Our goal in this paper is to understand the factors that determine the structural identifiability of binding parameters. The method presented in this study generates as output a set of fit parameters that are SI by construction, given two simple inputs: the number (n) of ligand-binding sites on the protein and the number of protein conformational states. The approach is general, as the SI parameter set is not derived from a specific model but directly from the conserved mathematical structure of the binding relation itself. The method can be applied to macromolecules with any number of binding sites.

Our analytical approach focuses on the question of the solvability of the system of equations that yields the parameter values, without actually requiring the solutions to be computed (as is done when numerical methods are used). By approaching the question in this way, we derive simple, general rules that govern binding parameter SI: (a) the parameters of any model consisting of any combination of bimolecular ligand–protein association reactions can be converted to a set of n SI fit parameters, {p_1, p_2, …, p_n}, because the form of the binding relation is conserved for all such models; and (b) models that also include protein conformational change are treated similarly, except that the SI parameter set comprises $n + 1$ fit parameters {p_0, p_1, p_2, …, p_n}.

The conserved form of the total binding relation for all models satisfying these criteria is

$$v = \frac{p_1 x^1 + 2 p_2 x^2 + \ldots + n p_n x^n}{1 + p_0 x^0 + p_1 x^1 + p_2 x^2 + \ldots + p_n x^n} \quad (1)$$

(Klotz, 1997), where v is the mean number of occupied sites at free ligand concentration x, n is the number of ligand-binding sites in the protein, and the parameters $\{p_0, p_1, \ldots, p_n\}$ are model-independent parameters. (Parameter p_0 in Eq. 1 is zero if the protein is assumed to occupy a single conformation.) Using matrix algebra methods, we show that each fit parameter p_b in Eq. 1 is obtained from the model parameters corresponding to all states with b bound ligands.

It is important to first establish, as we do here, that a set of parameters is SI before assessing whether they are PI because structural identifiability of parameters is a necessary (but not sufficient) condition that they are PI. Only when the parameters of a model are both SI and PI does one have confidence that analysis of binding data will yield meaningful estimates of molecular properties. A general approach to understanding the factors underlying the practical identifiability of the SI parameter sets described in this study is developed in the companion paper (Middendorf and Aldrich, 2017).

MATERIALS AND METHODS

Numerical calculations of binding isotherms (Fig. 1, C and D, top; and Fig. 2 C), perfect fit loci (Fig. 1, C and D, bottom; and Fig. 3), and design matrix determinants were performed using Igor Pro version 6.37 (WaveMetrics). Analytical derivations were performed by hand and checked using the symbolic mathematics software Maple 18 (Maplesoft).

RESULTS

In the next section, the mathematical structure of the total binding relation for the two-site, one-conformation model (Fig. 1 A) is derived, and the structural identifiability of its parameters is assessed and analyzed. In the following sections, it is shown that the results obtained for this simple system are readily generalized to all binding models comprising any combination of bimolecular association reactions and conformational equilibria, for proteins with any number of binding sites.

Structural nonidentifiability of model parameters for two-site allosteric model

By definition, the mean number of occupied binding sites, v, is the ratio of the concentration of ligand-bound sites to the total concentration of protein. For the model in Fig. 1 A, this ratio is given by the equation in Fig. 1 B (top). The symbols R_{ij} refer to ligation states with i ligands bound to site I and j ligands bound to site II ($i, j = 0$ or 1; Fig. 1 A) or to the concentration of those states (Fig. 1 B, top). (Which definition applies will be clear from the context.) The states and the terms derived from them are color coded black, red, and blue for zero, one, or two bound ligands, respectively. The reason for the color coding will be made clear shortly.

All of the elementary transitions in this model are ligand–protein association or dissociation processes, which are modeled as bimolecular reactions, such as

$$R_{00} + \text{Ligand} \overset{K_\text{I}}{\leftrightarrow} R_{10}, \tag{2}$$

which quantifies ligand binding to site I when site II is unoccupied. The equilibrium microscopic association constant K_I is given by the standard relation (Wyman and Gill, 1990; Winzor and Sawyer, 1995):

$$K_\text{I} = \frac{R_{10}}{R_{00}\, x}. \tag{3}$$

The quantities R_{10}, R_{00}, and x in Eq. 3 denote concentrations (which are used to approximate the activities of the corresponding species). The concentration of state R_{10} relative to the reference state (here the unliganded state R_{00}) is obtained by rearranging Eq. 3: $R_{10} = R_{00}\, K_\text{I}\, x$. The concentrations of states R_{01} and R_{11} are obtained similarly. Substituting these expressions into the equation in Fig. 1 B (top) yields the total binding relation (Fig. 2 A), which is not linear with respect to the model parameters. For real data, these parameters are properly estimated using nonlinear regression fitting (Seber and Wild, 2003; Jaqaman and Danuser, 2006). However, structural identifiability of parameters is assessed assuming noiseless data, allowing an important simplification: in the absence of a noise term, the binding relation (Fig. 2 A) can be cross-multiplied, yielding the equation in Fig. 2 B. This equation is linear with respect to the parameter set $\{K_\text{I}, K_\text{II}, f\, K_\text{I}\, K_\text{II}\}$. Here the quantity $f\, K_\text{I}\, K_\text{II}$ can be considered a "compound" parameter. The three equations required to determine the three unknown parameter values are obtained by evaluating the mean number of bound ligands (v_1, v_2, and v_3) at three ligand concentrations (x_1, x_2, and x_3; Fig. 2 C), yielding a system of linear equations (Fig. 2 D). The matrix representation of this system (Fig. 2 E) has the form

$$\boldsymbol{M}\boldsymbol{p} = \boldsymbol{v}, \tag{4}$$

where M is the design matrix, p is the parameter vector, and v is the vector of predicted values.

The power of the matrix algebra approach becomes apparent at this point, as the difficult question of whether the parameters of the two-site allosteric model are SI is replaced by the equivalent, but much simpler, question of whether there is a unique solution to Eq. 4. Left-multiplying each side of this equation by the inverse of the design matrix, M^{-1}, yields $p = M^{-1}\, v$. This system has a unique solution only if M is invertible or, equivalently, if the determinant of M is nonzero, which requires that the columns of M be linearly independent (Strang, 2003). This determination can be made by in-

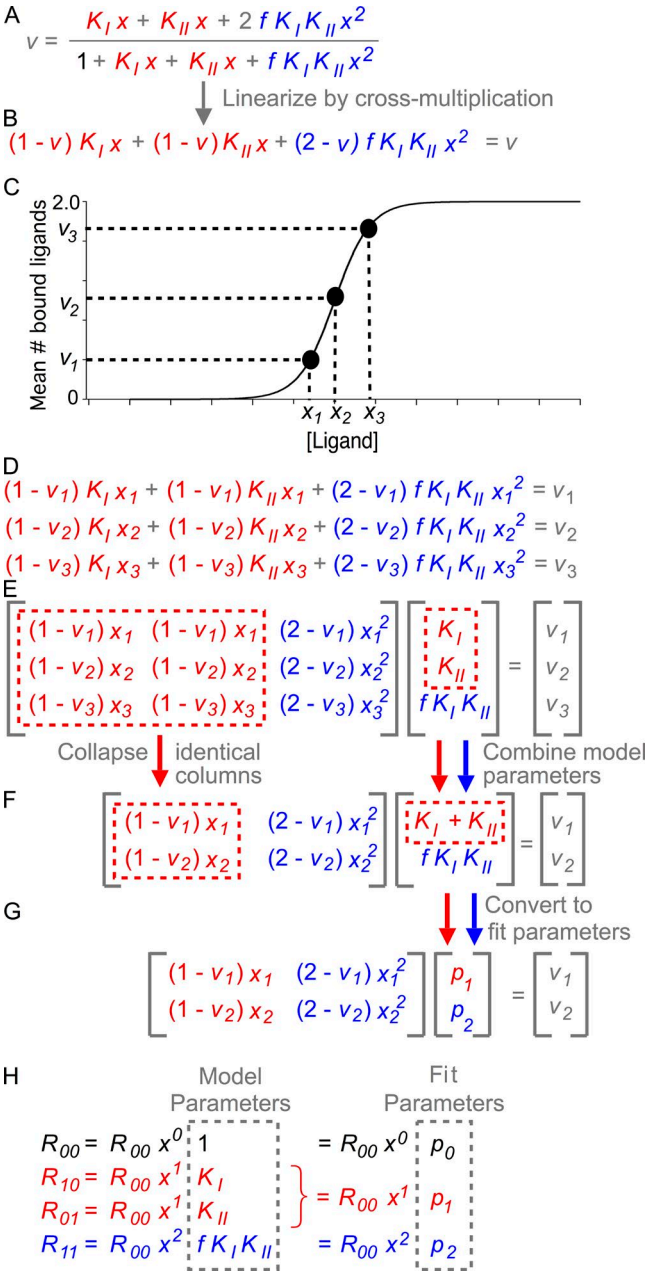

Figure 2. **Structural identifiability analysis of parameters for two-site binding models.** (A) Relation between mean number of bound ligands (v) and free ligand concentration (x) for two-site allosteric model from Fig. 1. Terms are color coded as in Fig. 1 B (bottom). (B) Linearized form of binding equation from A. (C) Synthetic total binding curve. The symbols v_1, v_2, and v_3 represent the mean number of bound ligands at free ligand concentrations x_1, x_2, and x_3. (D) System of equations obtained by evaluating equation in B at three ligand concentrations, as illustrated in C. (E) Matrix form of system of equations in D. Dashed red boxes indicate two identical columns in design matrix and corresponding parameters in parameter vector. (F) Reduced matrix equation obtained by combining parameters K_I and K_{II} into a single parameter: $K_I + K_{II}$. This procedure collapses the two identical red columns in the design matrix into a single column. (G) General form of reduced matrix equation in F, in which model-specific parameters $\{K_I + K_{II}, f K_I K_{II}\}$ are replaced by model-independent fit parameters $\{p_1, p_2\}$. (H) Expressions for state populations for model in Fig. 1 A and relation between model parameters and fit parameters: $p_1 = K_I + K_{II}$ and $p_2 = f K_I K_{II}$.

spection, without performing any computations: M is not invertible because the two red columns enclosed in the dashed box in Fig. 2 E are identical (and thus linearly dependent). This analysis indicates that the system does not have a unique solution, and hence the parameters of the two-site allosteric model are not SI, consistent with the simulation results in Fig. 1.

Structural identifiability of fit parameters for the two-site allosteric model

An advantage of the analytical approach used in this study is that the mathematical structure of matrix M reveals the exact cause of the identifiability failure and suggests a way to "repair" the parameter set (in the sense of restoring it to a condition of structural identifiability). Because the columns of the design matrix multiply the rows of the parameter vector in Fig. 2 E, combining K_I and K_{II} into a single parameter, $K_I + K_{II}$, merges the two identical red columns of M into a single column (Fig. 2 F). This reparameterization condenses the three-row by three-column (3×3), noninvertible design matrix (Fig. 2 E) into a "reduced" 2×2 design matrix (Fig. 2 F). The invertibility of the reduced design matrix is proved in the next section (Invertibility of reduced design matrix for the two-site allosteric model).

It is significant that the parameters in the first and second rows of the modified parameter vector ($K_I + K_{II}$ and $f K_I K_{II}$, respectively) are the parameters derived from the states with one and two bound ligands in Fig. 1 A. A final relabeling of the elements of this SI parameter set with subscripts denoting the number of bound ligands yields the set of SI fit parameters $\{p_1, p_2\}$ (Fig. 2 G):

$$p_1 = K_I + K_{II} \tag{5a}$$

and

$$p_2 = f K_I K_{II}. \tag{5b}$$

With the introduction of fit parameters, the binding relation in Fig. 2 A assumes the form of the general binding relation (Eq. 1) for the case of two binding sites:

$$v = \frac{p_1 x + 2 p_2 x^2}{1 + p_1 x + p_2 x^2}. \tag{6}$$

Thus, there are as many SI parameters as there are unique powers of ligand concentration in the total binding relation, which, in turn, is just the number of binding sites (n) in the protein. Fit parameters become particularly useful when analyzing proteins with a larger

number of binding sites because of the rapid increase in the number of model parameters.

An important consequence of Eq. 2 is that the expressions for state populations in binding models have a simple, conserved mathematical form. For each ligand-binding step, an additional power of ligand concentration (x) is accumulated in the expression. Thus, the concentration of any state with b bound ligands is proportional to the product of the reference state concentration (here the unliganded state R_{00}) and the free ligand concentration raised to the power b (Fig. 2 H). When considering the sum of the concentrations of all states with b bound ligands, the proportionality constant is just the fit parameter p_b. Thus, another advantage of introducing fit parameters is that identifiability can be treated without reference to a particular model. The form of Eq. 3 guarantees that SI analysis of any other two-site binding model consisting solely of bimolecular association reactions will yield the equation in Fig. 2 G after the appropriate combinations of model parameters have been converted into fit parameters.

Invertibility of reduced design matrix for the two-site allosteric model

We now show that the 2 × 2 "reduced" design matrix (Fig. 2 G), given by

$$\boldsymbol{M} = \begin{bmatrix} (1-v_1)\,x_1 & (2-v_1)\,{x_1}^2 \\ (1-v_2)\,x_2 & (2-v_2)\,{x_2}^2 \end{bmatrix}, \tag{7}$$

is invertible. The matrix elements in Eq. 7 can be simplified using Eq. 6, yielding

$$\boldsymbol{M} = \begin{bmatrix} \dfrac{(1-p_2{x_1}^2)x_1}{1+p_1x_1+p_2{x_1}^2} & \dfrac{(2+p_1x_1){x_1}^2}{1+p_1x_1+p_2{x_1}^2} \\ \dfrac{(1-p_2\theta^2{x_1}^2)\theta x_1}{1+p_1x_2+p_2{x_2}^2} & \dfrac{(2+p_1\theta x_1)\theta^2{x_1}^2}{1+p_1x_2+p_2{x_2}^2} \end{bmatrix}. \tag{8}$$

In Eq. 8, we have assumed, without loss of generality, that $x_2 = \theta x_1$, where $\theta > 1$ (Fig. 2 C). The determinant of this matrix can be expanded in the usual way, which, after simplification, yields

$$\mathrm{Det}(\boldsymbol{M}) = \frac{{x_1}^3\theta(\theta-1)\left[2+(\theta+1)p_1x_1+2\theta p_2{x_1}^2\right]}{\left(1+p_1x_1+p_2{x_1}^2\right)\left(1+p_1x_2+p_2{x_2}^2\right)}. \tag{9}$$

The right-hand side of Eq. 9 is greater than zero for all physically allowed (i.e., positive) values of p_1, p_2, and x. Thus, matrix M is invertible, the system in Fig. 2 G has a unique solution, and the fit parameters p_1 and p_2 are SI.

It is important to note that there is a cost to the process of "repairing" the original parameter vector (Fig. 2 E): by transforming the parameter set to achieve a condition of structural identifiability, the number of estimable parameters is reduced from three to two. None of the fundamental mechanistic parameters of the model in Fig. 1 A—the site affinities or the magnitude of the cooperative interaction between the sites—can be extracted from the values of the two fit parameters. The SI analysis indicates that less knowledge can be gained from the binding measurement than was anticipated by the model.

Infinite locus of model parameters yielding identical two-site binding curves

Having established that the fit parameters p_1 and p_2 completely specify any two-site binding curve (Eq. 6), we can now derive a general expression for calculating zero-error parameter contours (as in Fig. 1, C and D, bottom) for such curves.

Let $\{f, K_I, K_{II}\}$ and $\{p_1, p_2\}$ represent arbitrary sets of model and fit parameters related by Eqs. 5a and 5b. Similarly, let the sets $\{f^*, K_I^*, K_{II}^*\}$ and $\{p_1^*, p_2^*\}$ designate the correct values of the model and fit parameters for a given protein. Eq. 6 indicates that all parameter sets for which

$$K_I + K_{II} = p_1 = p_1^* \tag{10a}$$

and

$$fK_I K_{II} = p_2 = p_2^* \tag{10b}$$

will yield binding curves identical to the true binding curve for this molecule. Solving Eq. 10a for K_{II} and substituting this expression into Eq. 10b yields a quadratic equation in parameter K_I:

$$f{K_I}^2 - \left(fp_1^*\right)K_I + p_2^* = 0. \tag{11}$$

The two solutions of Eq. 11, given by the quadratic formula, are

$$K_I = \frac{p_1^*}{2} \pm \left[\frac{(p_1^*)^2}{4} - \frac{p_2^*}{f}\right]^{1/2}. \tag{12a}$$

Combining Eqs. 10a and 12a yields a similar expression for K_{II}:

$$K_{II} = \frac{p_1^*}{2} \mp \left[\frac{(p_1^*)^2}{4} - \frac{p_2^*}{f}\right]^{1/2}. \tag{12b}$$

Eqs. 12a and 12b can be used to compute the locus of all triples $\{f, K_I, K_{II}\}$ yielding two-site binding curves that are identical to the "true" curve with fit parameters $\{p_1^*, p_2^*\}$ (Fig. 3). Because K_I and K_{II} appear symmetrically in Eqs. 10a and 10b, there is an ambiguity in the values of K_I and K_{II} in Fig. 3, which is indicated by the signs in front of the square root terms in Eqs. 12a and 12b. If a triple $\{f, K_I, K_{II}\}$ is on the curve in Fig. 3, then the triple $\{f, K_{II}, K_I\}$, in which the values of the two association constants are switched, is also on the curve.

The salient features of the "perfect-fit" loci can be characterized by analyzing the mathematical properties of Eqs. 12a and 12b. Because K_I and K_{II} must be real-val-

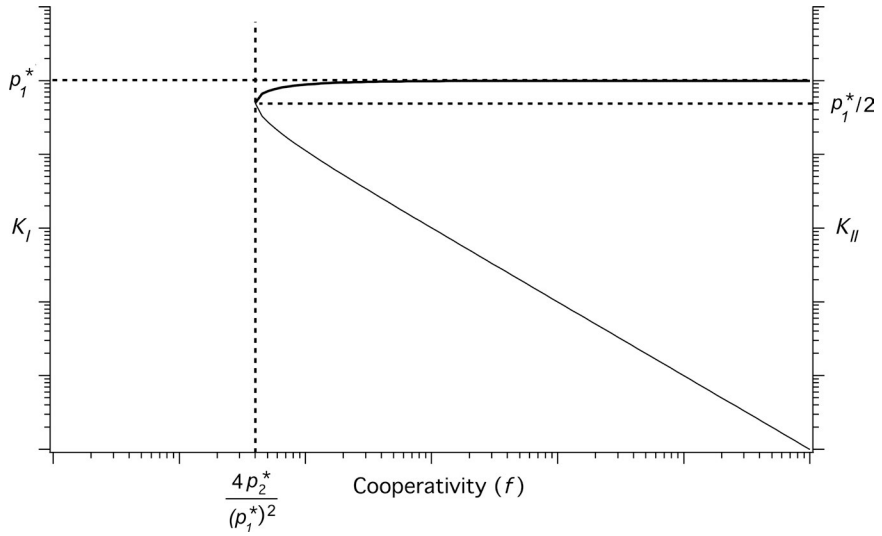

Figure 3. **Locus of points {f, K_I, K_{II}} yielding perfect fit to a two-site binding curve calculated using correct values of model parameters {f^*, K_I^*, K_{II}^*}.** Correct values of fit parameters are then $p_1^* = K_I^* + K_{II}^*$ and $p_2^* = f^* K_I^* K_{II}^*$. Curve was computed using Eqs. 12a and 12b. Upper (bold) and lower (normal) arms of curve correspond to values of either K_I and K_{II} or K_{II} and K_I. Minimum value of cooperativity parameter, f_{min}, and maximum value of the larger of K_I and K_{II} are also indicated.

ued, the quantity under the square root in Eqs. 12a and 12b must be greater than or equal to zero, which leads to the inequality

$$f \geq \frac{4p_2^*}{(p_1^*)^2}. \quad (13)$$

Thus, there is a minimum possible value of the cooperativity factor f for specified values of p_1^* and p_2^*. Substituting Eq. 13 into Eqs. 12a and 12b indicates that, at this minimum value of f, association constants K_I and K_{II} are equal. Finally, from Eq. 5a, the maximum possible value for the larger of K_I and K_{II} is p_1^*. These features are indicated on the generic curve in Fig. 3.

Curves like the one in Fig. 3 provide insight into how it is possible for many different sets of model parameters to yield the identical binding curve line shape. The curve in Fig. 1 D (bottom) illustrates the "push/pull" effect of cooperativity that underlies parameter nonidentifiability. For two sites with the same affinity (dashed line "D"), two distinct binding phases are apparent in the binding curve (Fig. 1 D, top) if there is strong negative cooperativity (i.e., binding to one site "pushes" the affinity of the adjacent site toward a lower value). For two sites with very different affinities (dashed line "F"), two binding phases with this same separation are obtained if there is strong positive cooperativity (i.e., binding to the high-affinity site "pulls" the very low-affinity site to higher affinity). The structure in the binding curve is the net result of two factors: (1) the ratio of the intrinsic affinities of the sites and (2) the magnitude of the cooperative interaction between the sites. An infinite number of combinations of values for these two factors yield the identical binding curve line shape (Fig. 1 D, top).

Structural nonidentifiability of parameters for three-site binding curves

The SI assessment strategy described in this study is readily applied to models of proteins with more than two binding sites. For example, the one-conformation model in Fig. 4 A comprises all possible ligated states for a receptor containing three binding sites. The model includes distinct microscopic association constants for all sites and distinct cooperative interactions between all pairs of sites. The nomenclature for the cooperativity factors highlights the conditional probabilistic nature of such models (Ben-Naim, 2001). For example, the equilibrium constant for the reaction in which state R_{001} binds ligand to form state R_{011} is $f_{II(III)}$ K_{II}. The symbol $f_{II(III)}$ represents the fold effect on binding to site II caused by occupancy of site III. Similarly, the symbol $f_{I(II,III)}$ represents the fold effect on binding to site I, given that sites II and III are occupied. After removing parameters that are redundant based on detailed balance considerations, the seven independent model parameters indicated in Fig. 4 A remain.

The procedure for estimating these unknown parameters is analogous to that for the two-site case (Fig. 2). The binding relation (Fig. 5 A) is obtained from the expressions for the concentrations of all ligated states of the system (Fig. 4 B). Seven constraint equations are obtained by evaluating the linearized form of the total binding relation (Fig. 5 B) at each of seven ligand concentrations, x_1 through x_7. The matrix representation of this system of equations (Fig. 5 C) has the same form as for the two-site case (Fig. 2 E): design matrix * parameter vector = vector of predicted values. The design matrix in Fig. 5 C is not invertible because it contains two sets of identical columns. Therefore, the set of seven model parameters is not SI when constrained by total binding data.

The identical red columns and the identical blue columns in Fig. 5 C arise from the three states with one bound ligand and the three states with two bound ligands, respectively. A clear pattern emerges when the matrix equations for three sites (Fig. 5 C) and two sites (Fig. 2 E) are compared: parameter nonidentifiability

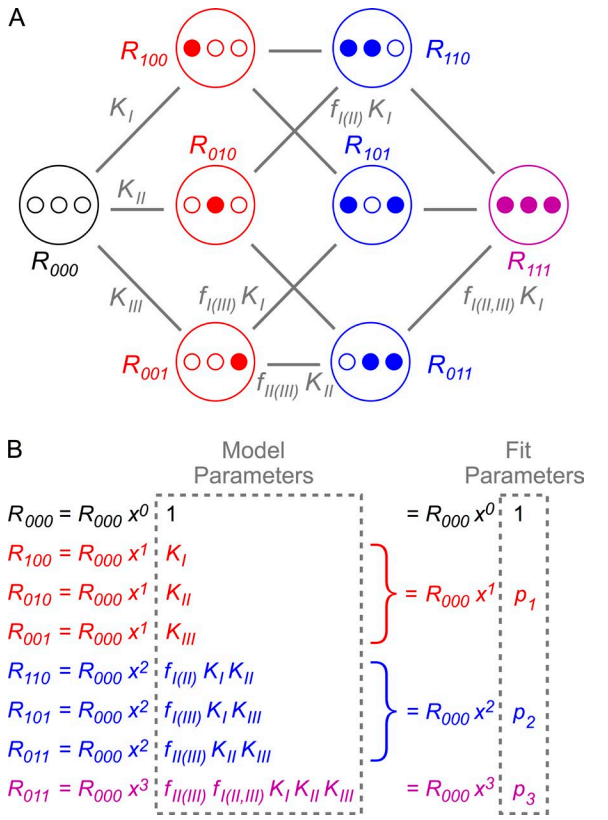

Figure 4. **Three-site allosteric binding model.** (A) State diagram in which symbols R_{ijk} designate states with i, j, and k ligands bound to sites I, II, and III (i, j, k = 0 or 1). K_I, K_{II}, and K_{III} represent microscopic equilibrium association constants for sites I–III. Conditional cooperativity factors $f_{J(K,L)}$ represent fold change in binding to site J given that sites K and L are occupied. (B) Expressions for state populations for model in A and relation between model parameters and fit parameters: $p_1 = K_I + K_{II} + K_{III}$, $p_2 = f_{I(II)}\,K_I\,K_{II} + f_{I(III)}\,K_I\,K_{III} + f_{II(III)}\,K_{II}\,K_{III}$, and $p_3 = f_{II(III)}\,f_{I(II,III)}\,K_I\,K_{II}\,K_{III}$. States and expressions are color coded black, red, blue, and purple for zero, one, two, and three bound ligands, respectively.

caused by identical columns in the design matrix is encountered whenever there is more than one state with a given number of bound ligands. The number of identical columns in the design matrix for a given value of b is equal to the number of ligated states with b bound ligands. This latter quantity is the number of ways of arranging b ligands on n nonidentical sites and is given by the binomial coefficient

$$\binom{n}{b} = \frac{n!}{b!\,(n-b)!}. \qquad (14)$$

The non-SI parameter set in Fig. 5 C can be transformed into one that is SI using the same strategy as in the two-site case (Fig. 2). The three model parameters for $b = 1$ (see Eq. 14) are condensed into a single fit parameter, $p_1 = K_I + K_{II} + K_{III}$, and the three (compound) parameters for $b = 2$ (see Eq. 14) are condensed into a single fit parameter, $p_2 = f_{I(II)}\,K_I\,K_{II} + f_{I(III)}\,K_I\,K_{III} + f_{II(III)}\,K_{II}\,K_{III}$ (Fig. 5, C–E). This process merges the three identical red columns into a single column and the three identical blue columns into a single column, reducing the original 7 × 7 design matrix to a 3 × 3 matrix. Again, the cost of achieving structural identifiability is a loss of knowledge regarding the molecular system because, at best, only three parameters, rather than the full seven of the model, can be estimated from fitting three-site binding curves. Also, it is not possible to evaluate any of the seven model parameters from knowledge of the three fit parameters.

Invertibility of reduced design matrix for three-site allosteric model

The "reduced" design matrix of Fig. 5 E is invertible, but obtaining a general proof of this issue is cumbersome for $n > 2$. We have adopted the alternative strategy of numerically evaluating the determinants of design matrices for a wide variety of three- and four-site binding curve line shapes. Although the determinants are always greater than zero, and therefore invertible, their magnitudes are often very small. Such matrices are "ill-conditioned" (Watkins, 1991), and inference for the linear system $M\,p = v$ (Fig. 5 E) is problematic: small changes in one of the elements in M or v may cause large changes in the solution for p. This property of design matrices for total binding indicates that, for real experimental conditions, small uncertainties in the measured binding site occupancy or in the ligand concentration (the v_i and x_i terms in Fig. 5 E) may cause large errors in the estimated values of one or more of the fit parameters (the p_i terms in Fig. 5 E). Thus, although the fit parameters $\{p_1, p_2, p_3\}$ are SI, we anticipate that there may be many cases for which they are not PI. This latter issue is the subject of the companion paper (Middendorf and Aldrich, 2017).

Structural identifiability of parameters for models of macromolecules with any number of binding sites

We have described a simple procedure for generating SI parameter sets for models of proteins containing two or three ligand-binding sites. The process consists of setting up a system of equations based on the linearized total binding relation, transforming the system into matrix form, and eliminating any identical columns in the design matrix by combining parameters derived from states with the same number of bound ligands. This process yields a matrix equation containing a "reduced" design matrix and a vector of fit parameters; some of the fit parameters are functions of multiple model parameters. Because of the conserved form of the total binding relation (Eq. 1), this approach can be extended to proteins with any number of ligand-binding sites.

Examples of the reduced design matrices for single-conformation models of proteins with two, three, four, and n binding sites are shown in Fig. 6. For n bind-

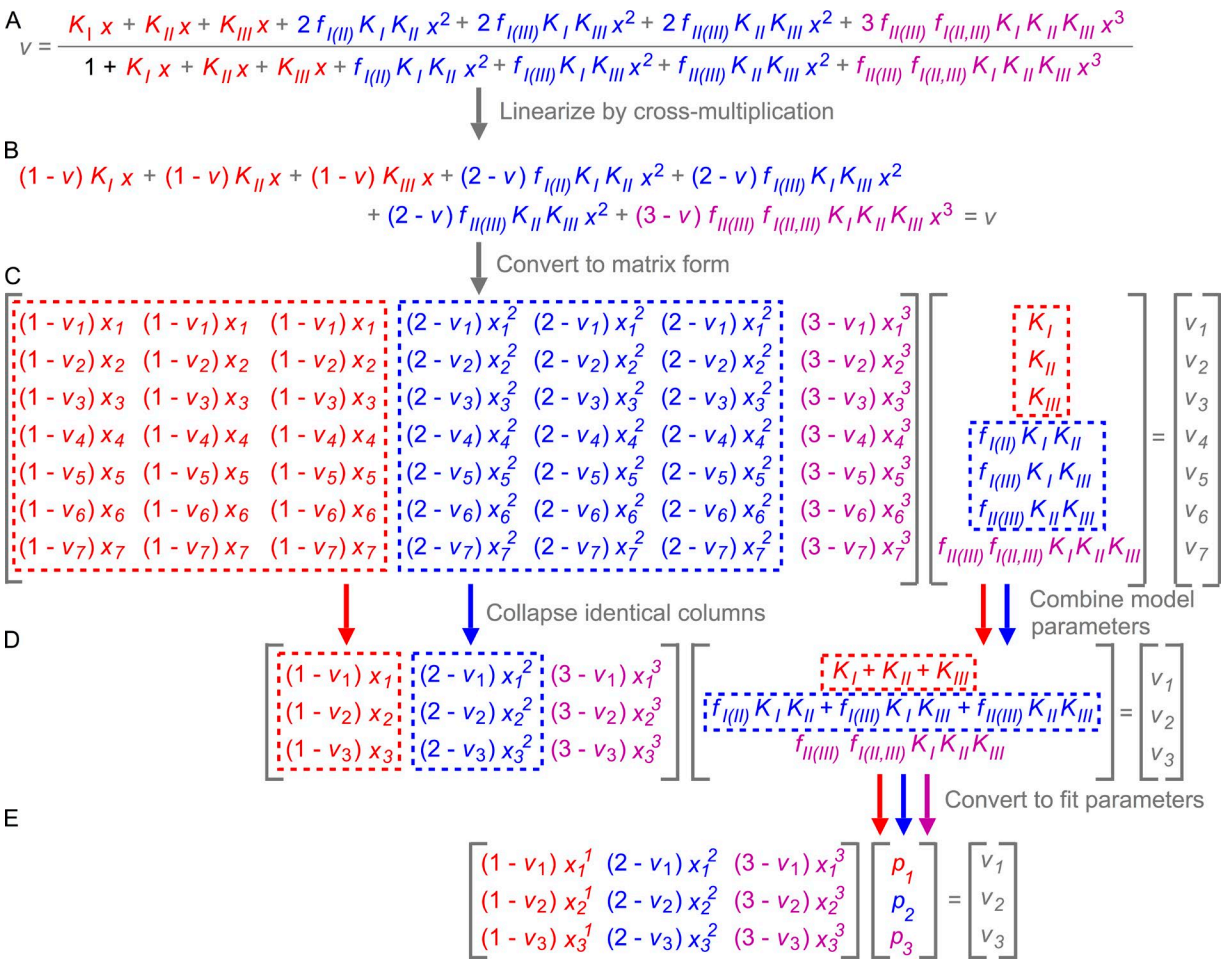

Figure 5. **Structural identifiability analysis of parameters for three-site allosteric binding model.** (A) Relation between mean number of bound ligands (v) and free ligand concentration (x) for the model in Fig. 4 A. Terms are color coded as in Fig. 4. (B) Linearized form of binding equation from A. (C) Matrix representation of system of equations obtained by evaluating the equation in B at seven ligand concentrations x_1 through x_7. Dashed red and blue boxes indicate identical columns in design matrix and corresponding parameters in parameter vector derived from states with one and two bound ligands, respectively. (D) Reduced matrix equation obtained by summing parameters K_I, K_{II}, and K_{III} from C into a single parameter and by summing parameters $f_{I(II)}\,K_I\,K_{II}$, $f_{I(III)}\,K_I\,K_{III}$, and $f_{II(III)}\,K_{II}\,K_{III}$ into a single parameter. This operation causes the identical red and blue columns to collapse into a single red and single blue column in the reduced design matrix. (E) General form of the reduced matrix equation in D, in which model-specific parameters are replaced by model-independent fit parameters $\{p_1, p_2, p_3\}$.

ing sites, the reduced design matrix contains n rows and n columns, and the element in row j, column k of this matrix, M_{jk}, is given by

$$M_{jk} = (k - v_j)\,x_j{}^k, \tag{15}$$

where the indices j and k are in the range $1 \le j, k \le n$. Because the reduced matrix equations have a canonical form for all values of n, all of the algebraic steps can be bypassed, and the reduced matrix equation can be written down once the value of n is specified. Analysis of the reduced matrix equation for macromolecules with n binding sites indicates that noiseless total binding data will constrain a set of n SI fit parameters $\{p_1, p_2, \ldots, p_n\}$ if it is assumed that the protein occupies a single conformational state.

Structural identifiability of parameters for models including conformational change

The central function of many ligand-binding proteins is to convert the free energy of ligand binding into conformational change to perform important cellular functions such as signal transduction. Our treatment of parameter SI is readily expanded to include models comprising any combination of binding and conformational equilibria. For example, for a two-site receptor, conformational change is modeled using equilibria of the form

$$M_{RS} = \frac{[S_{00}]}{[R_{00}]}, \tag{16}$$

which quantifies the distribution of unoccupied receptors between two conformations denoted R and S. Be-

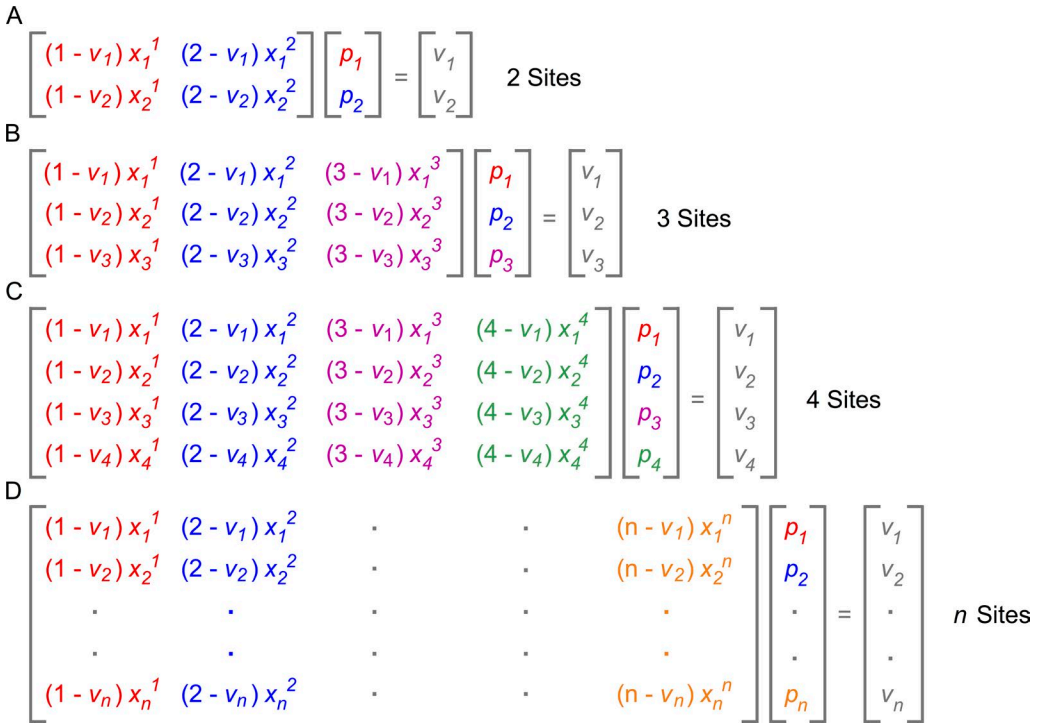

Figure 6. **Canonical form of reduced design matrices for binding models that include a single protein conformation.** (A–D) Reduced design matrices for proteins containing two (A), three (B), four (C), or *n* (D) binding sites. Proteins are assumed to occupy a single conformation. Matrix elements derived from states with one, two, three, four, or *n* bound ligands are color coded red, blue, purple, green, and orange, respectively. The general form of the matrix elements in all cases is given by Eq. 15.

cause the elementary conformational rearrangements are assumed to occur with no change in the ligation state of the receptor, this phenomenon is easily incorporated into our approach.

For example, the two-site model in Fig. 1 A can be expanded so that the macromolecule may occupy two (Fig. 7 B, left) or three (Fig. 7 C, left) global conformational states, denoted *R*, *S*, and *T*. In these models, protein conformation may influence ligand binding indirectly through binding at the other site (because of the conformation-specific cooperativity factors f_R, f_S, and f_T) and directly by explicit state dependence of the ligand affinities. The direct effect is quantified by the distinct association constants K_{IR}, K_{IS}, and K_{IT} for binding to site I and K_{IIR}, K_{IIS}, and K_{IIT} for binding to site II in the *R*, *S*, and *T* conformations.

Expressions for the state populations for the multiconformation models (Fig. 7, B and C, right) have the same conserved form as for the single-conformation model (Fig. 7 A, right). One new feature that emerges for models that include conformational change is the appearance of multiple unliganded states, but these adhere to the familiar pattern that their populations are given by the product of a reference state population (R_{00}), a factor that is a function of the model parameters, and ligand concentration raised to the power *b* (where $b = 0$; Fig. 7, B and C, right).

To solve for the 11 unknown, independent parameters of the model in Fig. 7 C (left), a system of equations is generated by evaluating the linearized binding relation at 11 (x_i, v_i) pairs. The parameters of this model are not SI because the design matrix is not invertible, resulting from the presence of multiple sets of identical columns (Fig. 8 A). The two identical black, six identical red, and three identical blue columns derive from states with zero, one, and two bound ligands, respectively. The now-familiar remedy of combining model parameters derived from states with the same number of bound ligands removes the linear dependencies in the design matrix by merging each set of identical columns (Fig. 8 B) and yields the set of SI fit parameters $\{p_0, p_1, p_2\}$ (Fig. 8 C). The unliganded states S_{00} and T_{00} are accounted for by the fit parameter p_0.

From the arguments made earlier, it is clear that these results generalize in a predictable way for models that include conformational change for proteins with any number of binding sites. Examples of the conserved form of the design matrices obtained for multiple-conformation models of proteins are shown in Fig. 9. As in Fig. 6, the matrix element M_{jk} is given by Eq. 15, except now the row and column numbers are in the range $0 \leq j, k \leq n$. For a protein with *n* binding sites and multiple conformational states, noiseless total binding data will constrain a set of $n + 1$ SI fit parameters $\{p_0, p_1, p_2, \ldots, p_n\}$.

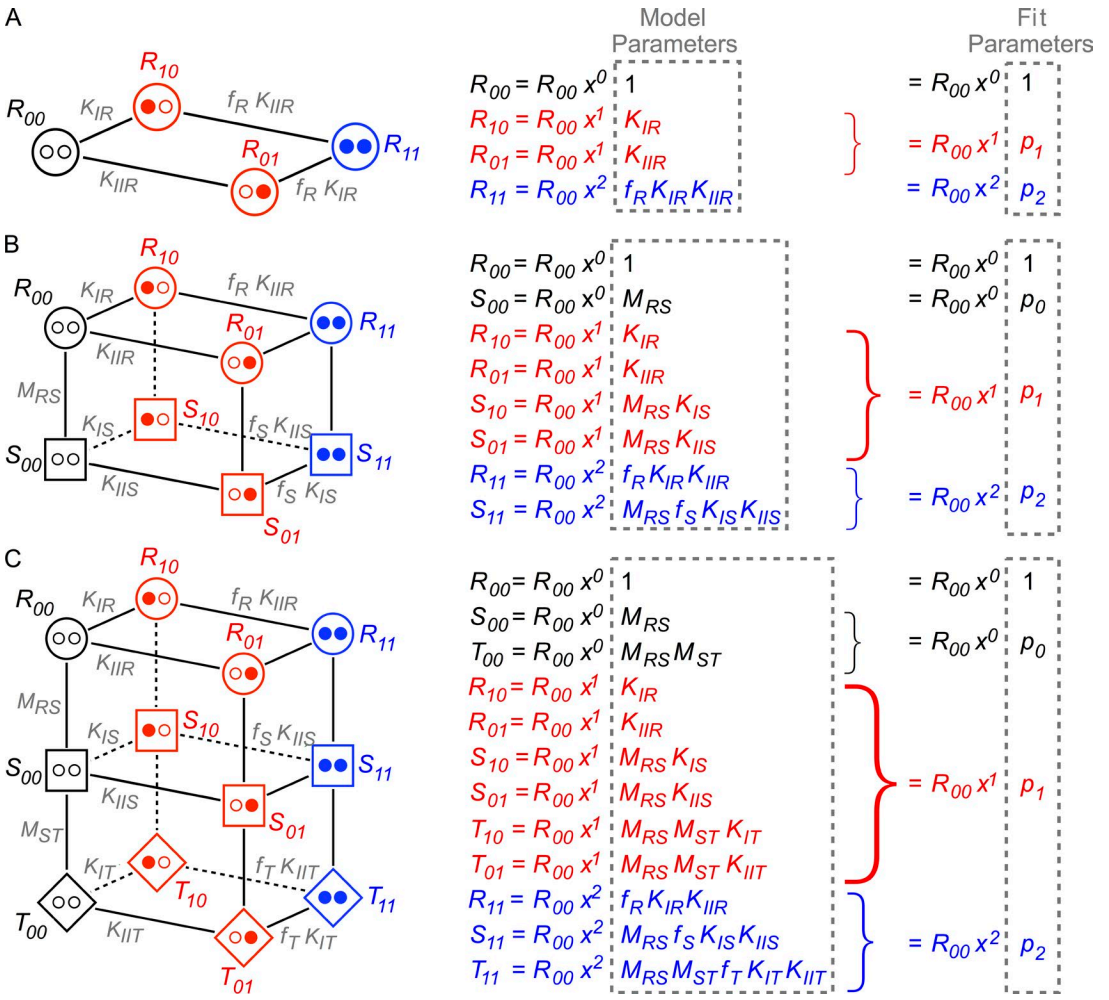

Figure 7. **State diagrams and state populations for two-site allosteric binding models that include multiple protein conformations.** (A–C, left) Models for proteins occupying one (A), two (B), or three (C) conformations. States R_{ij}, S_{ij}, and T_{ij} designate molecules in conformations *R*, *S*, and *T* with *i* and *j* ligands bound to sites I and II, respectively (*i*, *j* = 0 or 1). K_{IR}, K_{IS}, and K_{IT} represent microscopic equilibrium association constants for site I when the macromolecule is in the *R*, *S*, and *T* conformations, respectively. K_{IIR}, K_{IIS}, and K_{IIT} are the corresponding constants for site II. Cooperativity factors f_R, f_S, and f_T represent the fold change in binding to a site when the adjacent site is occupied and the protein is in the indicated conformation. M_{RS} and M_{ST} are the conformational equilibrium constants for the equilibria between states R_{00} and S_{00} and between states S_{00} and T_{00}, respectively. (A–C, right) Expressions for state populations for models shown on the left and relation between model parameters and fit parameters {p_0, p_1, p_2}. States and expressions are color coded black, red, and blue for zero, one and two bound ligands, respectively.

DISCUSSION

Our approach to assessing structural identifiability of total binding parameters

Much of our knowledge of the large and important class of macromolecular receptors that bind multiple ligands comes from estimates of binding parameters obtained by fitting total binding data. However, the question of the uniqueness and accuracy of these estimates has been largely ignored, likely because there is no general method for assessing binding parameter identifiability. We present in this study a method for determining the maximum number of SI binding parameters for a protein with *n* binding sites. The practical identifiability of these SI parameter sets is addressed in the companion paper (Middendorf and Aldrich, 2017).

Our approach to assessing binding parameter SI was guided by several considerations. It is important that the method be simple to apply so that the SI assessment can be made during the design phase of a proposed binding study. If the parameters of a candidate model are not SI, then the parameterization scheme is invalid; an ideal method would provide guidance on whether it is possible to "repair" non-SI parameter sets and, if so, how to modify those sets to achieve SI. Because it is very inefficient to reassess parameter SI for every candidate model under consideration, it would be preferable that the method generate a universal (i.e., model independent) pa-

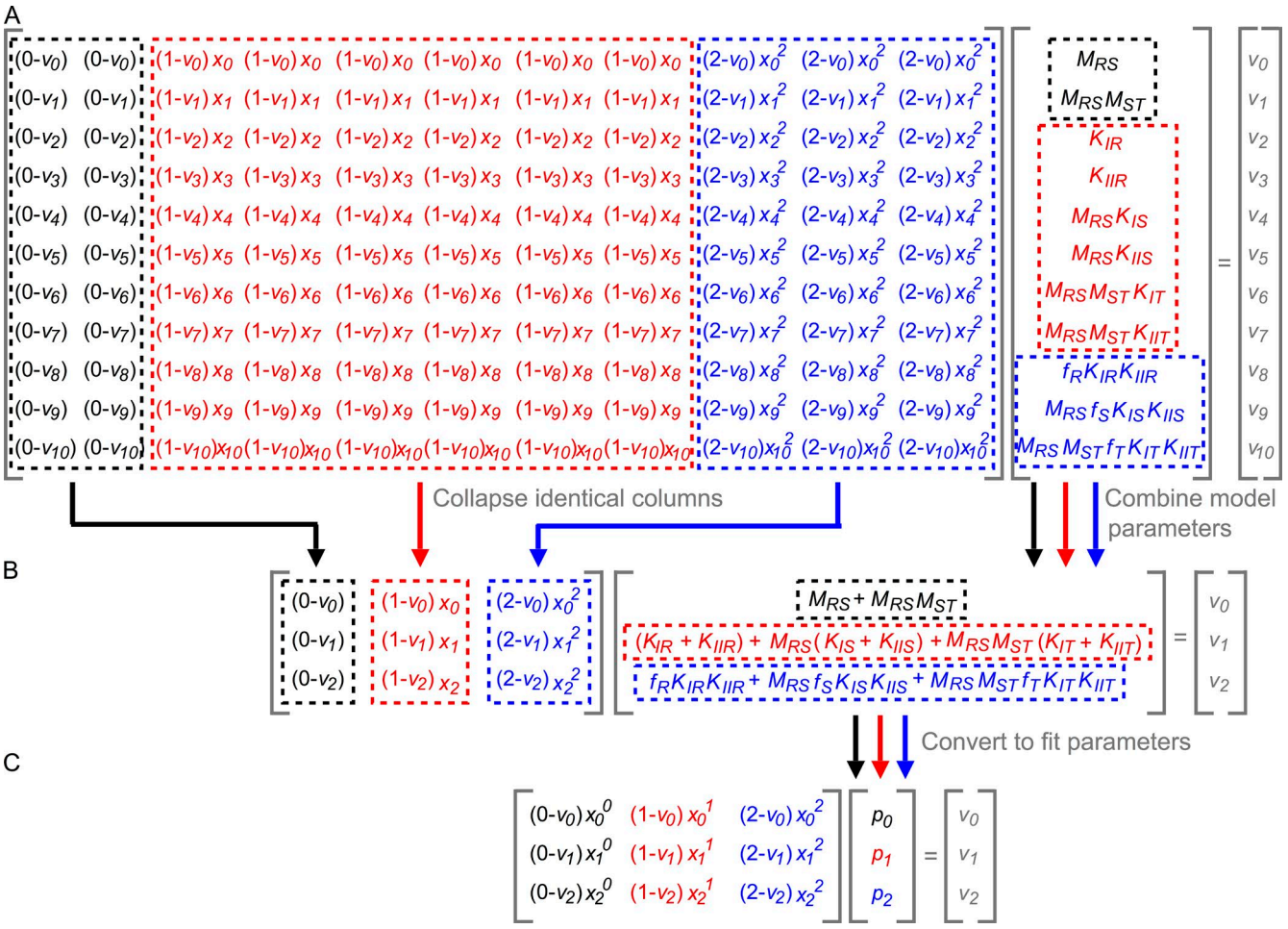

Figure 8. **Structural identifiability analysis of parameters for two-site, three-conformation binding model (Fig. 7 C, left).** (A) Matrix representation of system of equations obtained by evaluating linearized binding relation at 11 ligand concentrations x_0 through x_{10}. Dashed black, red, and blue boxes indicate identical columns in design matrix and corresponding parameters in parameter vector derived from states with zero, one, and two bound ligands, respectively. (B) Reduced matrix equation obtained by summing parameters in dashed boxes in A. This operation causes the identical black, red, and blue columns to collapse into a single black, single red, and single blue column in the reduced design matrix. (C) General form of reduced matrix equation in B, in which model-specific parameters are replaced by model-independent fit parameters $\{p_0, p_1, p_2\}$.

rameter set that is SI by design. Our method fulfills all of these criteria.

Three elements form the basis of our method. First, the state populations derived from the equations for binding (Eq. 3) and conformational equilibria (Eq. 16) have a conserved form for receptors with any number of binding sites: state population = reference state population * model parameter(s) * (free ligand concentration)^b, where b is the number of ligands bound to the state in question. Therefore, the total binding relation, which is the ratio of sums of these state populations, also has a conserved form (Eq. 1) and allows binding parameter SI to be treated in a general and model-independent fashion. Second, because the criteria for structural identifiability assume noiseless data, an intrinsically nonlinear problem can be linearized (Fig. 2, A and B). Third, matrix algebraic methods can be used to assess the solvability of linear systems of equations without performing calculations such as computing matrix inverses and are readily adapted to questions of parameter SI.

In our method for assessing SI, a system of equations is derived from the linearized form of the total binding relation. For any model composed solely of binding and conformational equilibria and for receptors with any number of ligand-binding sites, the matrix representation of this system has the invariant form: design matrix * parameter vector = vector of predicted values (Eq. 4). The question of whether the parameters are SI is equivalent to the question of whether the design matrix is invertible. The existence of multiple ligation states with the same total number of bound ligands (b) produces identical columns in the design matrix (Figs. 2 E, 5 C, and 8 A), which renders this matrix singular (i.e., noninvertible). An important advantage of our analytical approach over numerical methods is that the cause of

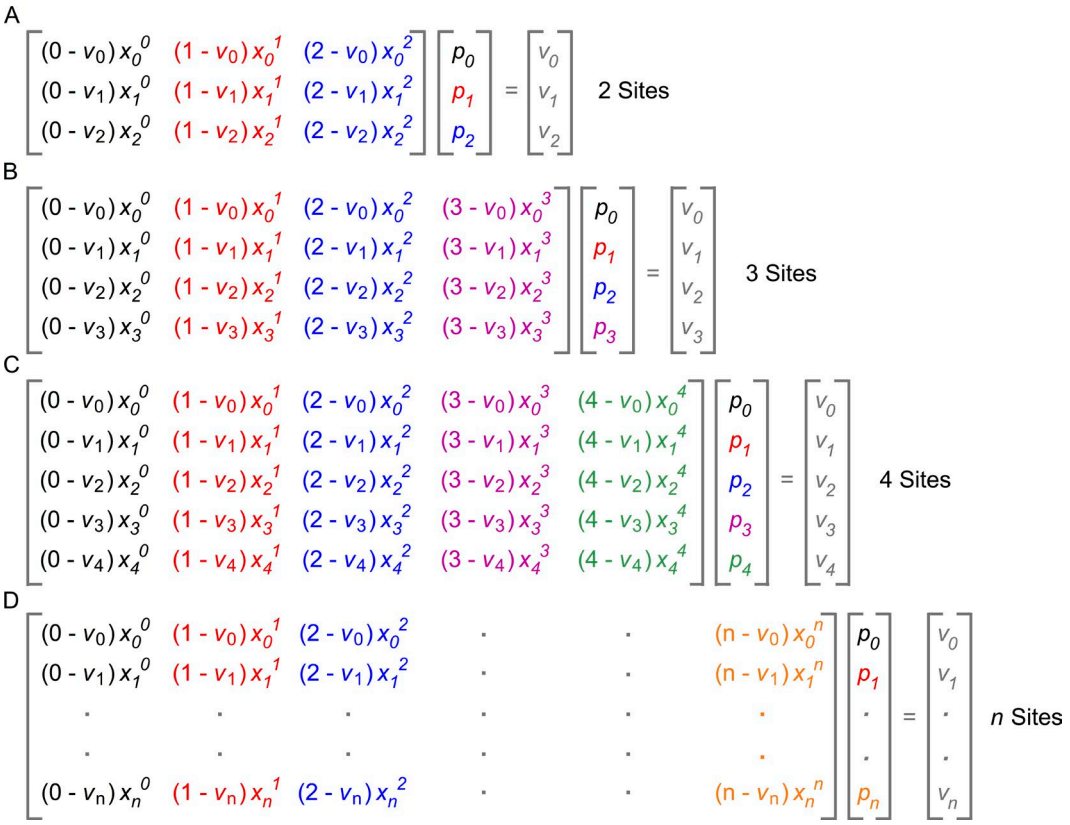

Figure 9. **Canonical form of reduced design matrices for binding models that include multiple protein conformations.** (A–D) Reduced design matrices for proteins containing two (A), three (B), four (C), or *n* (D) binding sites.Matrix elements derived from states with zero, one, two, three, four, and *n* bound ligands are color coded black, red, blue, purple, green, and orange, respectively. The general form of the matrix elements in all cases is given by Eq. 15, with the row and column numbering ranging from 0 to *n*.

the identifiability failure and its solution are revealed by the form of the design matrix. In all cases, the parameter set can be made SI by combining the model parameters for each group of states with the same value of b into a compound fit parameter, p_b. This transformation merges each group of identical columns in the design matrix into a single column, yielding a reduced, invertible design matrix and a set of SI fit parameters (Figs. 2 G, 5 E, and 8 C). Because the final matrix equation produced by this procedure has a canonical form (Figs. 6 and 9), the SI parameter set can be written down by inspection, with no calculations, and with no information other than the number of binding sites and whether the model includes conformational change.

In summary, we have derived a general strategy for generating the largest set of SI parameters for receptors with any number of binding sites, without reference to a specific physical binding model. The set of SI fit parameters $\{p_0, p_1, \ldots, p_n\}$ are the coefficients of powers of ligand concentration in the total binding relation (Eq. 1). The parameters of all models of protein–ligand interaction that consist of any combination of unitary steps comprising binding equilibria (Eq. 2) or conformational equilibria (Eq. 16) reduce to this canonical form. These very nonrestrictive criteria include virtually all physically reasonable binding models. These results also provide insight into why total binding data have relatively low power for constraining model parameters: the measurement acts as a coarse filter that sorts the states of the system into groups according to the number of bound ligands but does not distinguish between the states with a given value of b. Thus, many parameters that relate to the population of specific ligated intermediates cannot be estimated individually, but rather are folded into fit parameters that contain multiple terms. For example, the SI fit parameters for the model in Fig. 1 A are equal to the sum of the microscopic site binding constants ($p_1 = K_I + K_{II}$), and the product of the three model parameters ($p_2 = f K_I K_{II}$). None of the individual model parameters can be determined from fitting total binding curves, even in the absence of noise.

Assumptions of our approach to parameter identifiability

Our treatment of binding parameter SI incorporates several simplifying assumptions. General assumptions include the following: (a) Data are from total binding measurements performed at equilibrium. (b) Only

models that consist of binding equilibria and conformational equilibria of the form specified by Eqs. 3 and 16 are considered. (c) Binding between ligand and receptors containing n distinct (and generally nonidentical) ligand-binding sites occurs within a single, aqueous reaction phase.

Specific assumptions about the protein include the following: (d) All receptors are identical except for differences in site occupancy and conformation caused by ligand binding (i.e., there are no variations in stoichiometry or posttranslational modifications between receptors). (e) There are no interactions between functional receptors, which eliminates the possibility of dimerization of receptors or higher aggregate formation. (The functional receptors may be oligomeric; our assumption is that these oligomers do not interact.) (f) Protein is present at sufficiently low concentrations that ligand depletion effects (Goldstein and Barrett, 1987) are not significant. (g) Protein is present at sufficiently low concentrations that complications caused by molecular crowding (Zimmerman and Minton, 1993) can be ignored. (This assumption applies to the ligand as well.)

Specific assumptions about the ligand include the following: (h) There is a single ligand species present, and all ligand molecules are identical. (i) There are no interactions between ligands that are not bound to protein. (Interactions between multiple ligands bound to a single receptor are allowed.) (j) Ligands bind only at the specified sites on the protein: there is no nonspecific binding. (k) Ligands that are asymmetric bind in only one orientation in the protein-binding site. (l) Ligands bind to only one site at a time (i.e., ligand multivalence is not considered).

An important future direction of this research is to explore whether some of these assumptions may be relaxed. By properly modeling the effects, our general approach to parameter identifiability may be expanded to include an even wider range of phenomena. For example, we are extending the theory to account for ligand depletion (assumption f) and the presence of multiple, competing ligand species (assumption h). In addition, the single-phase approximation (assumption c) may be relaxed by incorporating the formalism developed by Wells (Hulme, 1992) to treat cases in which protein and ligand occupy multiple phases, such as aqueous and membrane compartments. Dimerization of functional receptors (assumption e) has been treated in the hemoglobin literature (Riggs, 1998) and may also be incorporated into our approach.

Limitations to inferring mechanism from analysis of total binding curves

Quantifying the microscopic site affinities, the magnitudes of cooperative interactions between binding sites, and possible conformational effects on these parameters are important goals of mechanistic binding studies. To specify these molecular properties, a total of $n * c$ parameters are required if the affinities of all n sites are assumed distinct in each of c protein conformations. The number of additional parameters required to specify the magnitudes of all possible site–site interactions increases rapidly as n increases. Thus, models that allow for distinct site affinities and cooperative interactions between the various sites require large numbers of parameters. In contrast, our analysis shows that the maximum number of SI parameters supported by equilibrium total binding data are smaller: n if the model is composed solely of binding equilibria (Eq. 3) and $n + 1$ if the model also includes conformational equilibria (Eq. 16). The discrepancy between the number of parameters required by detailed mechanistic models and the number that can be estimated reliably from experimental data indicates that a good fit to total binding data provides almost no information about the physical properties of binding sites in proteins. This observation may explain the popularity of much simpler models such as the Klotz–Adair model (Klotz, 1997), which, by distinguishing states based only on the total number of bound ligands, requires a total of n parameters. The inevitable trade-off required with this model is that the parameters are macroscopic association constants that do not distinguish between site affinity and cooperativity. Identifiability analysis underscores the need for other experimental measurements that provide stronger parameter constraints, such as equilibrium site-specific binding (Di Cera, 1995), binding kinetics, and conformation measurements.

Are both SI and PI assessments needed?

Structural identifiability is a necessary but not sufficient condition for ensuring that parameters obtained from fitting a model to data are accurate and unique (Bellman and Åström, 1970; Němcová, 2010). SI is assessed assuming ideal conditions (noiseless data) that are never achieved in real-world situations. When fitting experimental data containing noise, it is possible that the number of PI parameters may be even smaller than the number of SI parameters. Thus, it is natural to question whether it is worthwhile assessing parameter SI if parameter PI (which is the sufficiency condition) is to be determined separately.

For the case of total binding parameters, we find that the SI assessment is essential; the conclusions reached in the PI assessment phase vary depending on whether the parameter set is SI. For example, Fig. 1 (C and D, bottom) shows that the three parameters of the two-site allosteric model (Fig. 1 A) are not SI (and therefore not PI), regardless of the degree of resolved structure in the total binding curve (Fig. 1, C and D, top). In contrast, the practical identifiability of the two SI fit parameters for two-site binding curves (Fig. 2, G and H) is shown in the companion paper (Middendorf and Aldrich, 2017) to depend strongly on the amount of resolved structure in the binding curve.

ACKNOWLEDGMENTS

The authors wish to thank Keegan Hines for assistance in the early phases of this work and D. Brent Halling, Suzanna Bennett, and Ben Liebeskind for helpful discussions and critical reading of the manuscript.

This work was supported by National Institutes of Health grant R01-NS077821 to R.W. Aldrich.

The authors declare no competing financial interests.

Christopher Miller served as guest editor.

Submitted: 29 September 2016
Accepted: 23 November 2016

REFERENCES

Audoly, S., G. Bellu, L. D'Angiò, M.P. Saccomani, and C. Cobelli. 2001. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.* 48:55–65. http://dx.doi.org/10.1109/10.900248

Bellman, R., and K. Åström. 1970. On structural identifiability. *Math. Biosci.* 7:329–339. http://dx.doi.org/10.1016/0025-5564(70)90132-X

Ben-Naim, A. 2001. Cooperativity and Regulation in Biochemical Processes. Springer-Verlag US, New York, NY. 349 pp. http://dx.doi.org/10.1007/978-1-4757-3302-0

Chis, O.-T., J.R. Banga, and E. Balsa-Canto. 2011. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS One.* 6:e27755. http://dx.doi.org/10.1371/journal.pone.0027755

Colquhoun, D. 1969. A comparison of estimators for a two-parameter hyperbola. *J. R. Stat. Soc. Ser. C Appl. Stat.* 18:130–140.

Colquhoun, D., and D.C. Ogden. 1988. Activation of ion channels in the frog end-plate by high concentrations of acetylcholine. *J. Physiol.* 395:131–159. http://dx.doi.org/10.1113/jphysiol.1988.sp016912

Colquhoun, D., and F.J. Sigworth. 1983. Fitting and statistical analysis of single-channel records. *In* Single-Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York. 191–263. http://dx.doi.org/10.1007/978-1-4615-7858-1_11

Di Cera, E. 1995. Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules. Cambridge University Press, Cambridge, UK. 296 pp.

Edwards, A.W.F. 1992. Likelihood. Johns Hopkins University Press, Baltimore, MD. 275 pp.

Epstein, M., B. Calderhead, M.A. Girolami, and L.G. Sivilotti. 2016. Bayesian statistical inference in ion-channel models with exact missed event correction. *Biophys. J.* 111:333–348. http://dx.doi.org/10.1016/j.bpj.2016.04.053

Goldstein, A., and R.W. Barrett. 1987. Ligand dissociation constants from competition binding assays: Errors associated with ligand depletion. *Mol. Pharmacol.* 31:603–609.

Hengl, S., C. Kreutz, J. Timmer, and T. Maiwald. 2007. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics.* 23:2612–2618. http://dx.doi.org/10.1093/bioinformatics/btm382

Hines, K.E., T.R. Middendorf, and R.W. Aldrich. 2014. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *J. Gen. Physiol.* 143:401–416. http://dx.doi.org/10.1085/jgp.201311116

Hulme, E.C., editor. 1992. Receptor–Ligand Interactions: A Practical Approach. Oxford University Press, Oxford, UK. 458 pp.

Jaqaman, K., and G. Danuser. 2006. Linking data to models: Data regression. *Nat. Rev. Mol. Cell Biol.* 7:813–819. http://dx.doi.org/10.1038/nrm2030

Klotz, I.M. 1997. Ligand-Receptor Energetics: A Guide for the Perplexed. Wiley Interscience, New York, NY. 192 pp.

Ljung, L. 1987. System Identification: Theory for the User. Prentice Hall, Englewood Cliffs, NJ. 519 pp.

Middendorf, T.R., and R.W. Aldrich. 2017. The structure of binding curves and practical identifiability of equilibrium ligand-binding parameters. *J. Gen. Physiol.* 149. http://dx.doi.org/10.1085/jgp.201611703

Němcová, J. 2010. Structural identifiability of polynomial and rational systems. *Math. Biosci.* 223:83–96. http://dx.doi.org/10.1016/j.mbs.2009.11.002

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics.* 25:1923–1929. http://dx.doi.org/10.1093/bioinformatics/btp358

Riggs, A.F. 1998. Self-association, cooperativity and supercooperativity of oxygen binding by hemoglobins. *J. Exp. Biol.* 201:1073–1084.

Seber, G.A.F., and C.J. Wild. 2003. Nonlinear Regression. John Wiley & Sons, Hoboken, NJ. 768 pp.

Strang, G. 2003. Introduction to Linear Algebra. Third edition. Wellesley-Cambridge Press, Wellesley, MA. 568 pp.

Walter, E., and L. Pronzato. 1997. Identification of Parametric Models from Experimental Data. Springer-Verlag, Berlin. 413 pp.

Watkins, D.S. 1991. Fundamentals of Matrix Computations. John Wiley & Sons, New York. 449 pp.

Winzor, D.J., and W.H. Sawyer. 1995. Quantitative Characterization of Ligand Binding. Second edition. Wiley-Liss, New York, NY. 176 pp.

Wyman, J., and S.J. Gill. 1990. Binding and Linkage: Functional Chemistry of Biological Macromolecules. University Science Books, Mill Valley, CA. 330 pp.

Zimmerman, S.B., and A.P. Minton. 1993. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* 22:27–65. http://dx.doi.org/10.1146/annurev.bb.22.060193.000331