MOLECULAR PHARMACOLOGY, 13, 387–399

Alternative Models for the Analysis of Drug-Protein Binding

JOHN E. FLETCHER

Laboratory of Applied Studies, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20014

ARTHUR A. SPECTOR

Departments of Biochemistry and Medicine, University of Iowa, Iowa City, Iowa 52242

(Received July 23, 1976)

(Accepted December 14, 1976)

SUMMARY

FLETCHER, JOHN E. & SPECTOR, ARTHUR A. (1977) Alternative models for the analysis of drug-protein binding. *Mol. Pharmacol.*, 13, 387–399.

Mathematical models that are used to describe experimentally obtained drug-protein binding data often exhibit serious limitations. These limitations are demonstrated by a thorough examination of the binding of salicylate to human serum albumin. It is shown that simple models with nonintegral coefficients are inadequate for the interpretation of such binding data. A generalized mathematical formulation that includes both the Scatchard and stepwise equilibrium binding models is derived. This formulation is simpler to fit to data than the stepwise formulation, yet is not restricted to the Scatchard assumption of independent and noninteracting binding sites. This alternative formula is appropriate for the analysis of all equilibrium binding data. Even with this generalized mathematical formulation and sophisticated computerized fitting techniques, it is not possible to define precisely a single binding model for the salicylate binding data. The utility of model fitting in such a case is discussed, and procedures are demonstrated that characterize the information that can be obtained.

INTRODUCTION

Recent improvements in the measurement of drug binding to proteins have stimulated interest in the mathematical modeling of this interaction. The use of a mathematical model to interpret drug-protein binding usually has one or more of the following objectives: (a) determination of the number and class of binding sites that a carrier protein molecule contains, (b) partition of the drug-protein mixture into its constituent species, that is, fraction bound, fraction free, etc., and (c) inferences about the distribution of the drug between bound and free forms under physiological conditions.

A number of alternative mathematical models are available for the modeling of drug-protein interactions. These models vary in complexity and in the type of information that they can provide. Because of the increasing interest among investigators in applying mathematical models, it seems appropriate to describe the use and potential pitfalls of such modeling efforts. To illustrate the theoretical points, we have utilized the data on the binding of salicylate to human serum albumin previously reported by Mais *et al.* (1).

MATHEMATICAL FORMULATION OF MODELS

Drug-protein interactions are often characterized in terms of the total drug present and fraction bound to the protein

ISSN 0026-895x

carrier, for these values are readily obtained from the experimentally determined free and total drug concentrations. Most equilibrium binding models, however, formulate the fraction bound as a function of the free drug concentration. With such a formulation, the total drug concentration T and fraction bound, FB, are commonly given by the equations

$$T = P_t \bar{\nu} + A$$

and (1)

$$\text{FB} = \frac{P_t \bar{\nu}}{T}$$

where P_t is the total protein concentration in moles, $\bar{\nu}$ is the bound drug concentration in moles per mole, and A is the concentration of the free drug in moles. The most common models present the bound drug as a function of the free concentration A. For this reason, extensive statistical analyses of the fitted results will not be attempted in this paper. We concentrate instead on the question: What mathematical formulation is most appropriate for fitting to equilibrium binding data? A statistical analysis of the model parameters and the variances generated by fitting this model is then a separate issue.

Site-specific models. The simplest case of drug binding is that involving a single class of independent, noninteracting binding sites. The number of moles of drug bound per mole of protein is given by the Scatchard model (2),

$$\bar{\nu} = N \frac{kA}{1 + kA} \tag{2}$$

where N is the number of protein binding sites and k is the association constant for these sites. Equation 2 has the convenient property that the Scatchard variables $y = \bar{\nu}/A$ and $x = \bar{\nu}$ transform this simple model to the linear form

$$y = Nk - kx \tag{3}$$

In this form the binding parameters N and k can (in theory) be read directly from the intercepts (i.e., y–0, x=N; x=0, y=Nk) of a graph of the experimental data.

Unfortunately, few drug-protein interactions are described by this simple model. In general, binding data, when exhibited in the Scatchard variables, will exhibit a curvilinear form, suggesting the presence of a multiplicity of classes of binding sites or the occurrence of an allosteric effect during the binding process (3–5).

If there is more than one class of independent binding sites, but the sites are noninteracting, the appropriate model is a generalization of Eq. 2 (6),

$$\bar{\nu}(c) = N_1 \frac{k_1 A}{1 + k_1 A} + \cdots + N_m \frac{k_m A}{1 + k_m A} \tag{4}$$

where m is the number of classes of binding sites. The binding constant k_i represents the microscopic, or site-specific, binding affinity for the sites in each class.

In the Scatchard variables, $y = \bar{\nu}/A$ and $x = \bar{\nu}$, the multiclass model described by Eq. 4 can be expressed in the form

$$\frac{N_1 k_1}{y + k_1 x} + \frac{N_2 k_2}{y + k_2 x} + \cdots + \frac{N_m k_m}{y + k_m x} = 1 \tag{5}$$

From this equation it can be shown that the Scatchard plot slope of this model is given by

$$\frac{dy}{dx} = -\frac{\sum_{i=1}^{m} \frac{N_i k_i^2}{(y + k_i x)^2}}{\sum_{i=1}^{m} \frac{N_i k_i}{(y + k_i x)^2}} \tag{6}$$

The quantity on the right side of Eq. 6 is always negative. Therefore, if the binding model represented by Eq. 4 is to apply to a set of drug binding data, the Scatchard plot of the data must have an everywhere negative slope. The Scatchard plot is therefore a means for graphically identifying positive cooperativity in drug-protein binding studies. If the slope on this plot is non-negative over any binding range, positive cooperativity is present in the drug binding. However, as we shall demonstrate, an everywhere negative slope on this graph does not guarantee that the

Scatchard model will apply to a given set of binding results.

The intercepts on the Scatchard plot, as in the simple case, indicate the total number of drug binding sites; i.e., at $y = 0$,

$$x = N_1 + N_2 + \cdots + N_m$$

and at $x = 0$, (7)

$$y = N_1k_1 + N_2k_2 + \cdots + N_mk_m$$

However, these equations are generally not useful for determining the separate values for N_i and k_i in each class of binding sites. Many subjective graphical methods have been used to approximate the parameters in this Scatchard model. An excellent summary of the correct mathematical formulae in the respective graphical presentations has been given by Klotz and Hunston (7). We shall not discuss subjective graphical methods here, but concentrate instead on the more accurate computer-based fitting methods and results that can be obtained with them. Specifically, all model parameters will be determined by minimizing the sums of squares of deviations $\sum (\hat{\nu} - \bar{\nu})^2$, where $\hat{\nu}$ is the experimentally measured value and $\bar{\nu}$ is the computed value from a mathematical model. The independent variable is always taken as the free concentration A.

Site-specific data analysis. In order to fit Eq. 4 to experimental data, it is necessary to determine and interpret the unknown parameters in this model. The approach often taken is to fit Eq. 4 to experimental data by permitting the coefficients N_i to take on nonintegral values (8). An examination of the data on the binding of salicylate to human serum albumin, as determined by Mais *et al.* (1, 9, 10), will illustrate the difficulties associated with this practice. These data were selected because they appear relatively smooth in graphical presentations and the probable number of binding sites is small, thus permitting a detailed analysis of the binding. For many data sets, such as our own (11, 12), the uncertainty of the total number of binding sites and scatter in the experimental data prevent a thorough, detailed analysis. The site-specific analysis and methodology of Mais *et al.* have been used to predict the relative contributions to the bound fraction of the so-called secondary sites (1, 13). Since these data and detailed analyses are available in published form, they provide practical examples of the points we wish to amplify.

We have combined the Mais data at 0.3 g % protein concentration and fitted them to a Scatchard model using procedures which we have described previously (14–17). The results for a four-parameter, two-class Scatchard binding model are

$$N_1 = 1.27961,\ k_1 = 0.708556$$

$$N_2 = 3.80413,\ k_2 = 0.032721$$

with an unweighted sum of squares = 0.04506. These values are the same as those reported by Mais *et al.* (1). The Scatchard plots of these data and the fitted free-parameter curve are shown in Fig. 1.

The first difficulty in properly interpreting these results is the nonintegral values for N_1 and N_2. Theoretically, only integral values can be assumed for these parameters (2, 5, 16). Therefore we must ask what possible meaning these values can have. There are at least three possibilities:

1. The binding molecules are inhomogenous; that is, the protein molecules in a given mixture do not have the same or similar numbers of binding sites for the

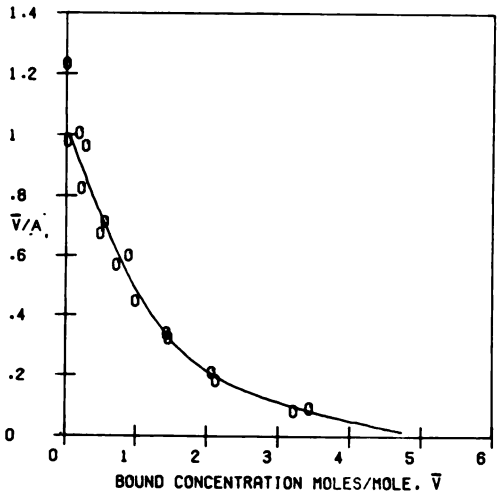

FIG. 1. *Scatchard plot, $\bar{\nu}/c$ vs. $\bar{\nu}$, for binding of salicylate to human serum albumin*

——, free-parameter fitted curve; O, experimental data as determined by Mais *et al.* (1).

drug. In such cases N_1 and N_2 are average values over a large number of different molecules and would not, in general, be integral.

2. The protein concentration is incorrect, or a second ligand is present and competes for some of the sites on the protein. For example, if the serum albumin is not defatted, some drug binding sites may be occupied by fatty acids (18). Both these defects would result in nonintegral coefficients.

3. The binding model is inappropriate even though the noninteger model appears to fit the data satisfactorily. In short, the data can be adequately fitted because the free-parameter model has sufficient degrees of freedom to conform to the variations of the data and hence "fits," but it does not represent the molecular mechanism.

Integer coefficient model fitting. To be theoretically valid, a Scatchard model should have integer values for N_i. Therefore our initial approach to reconcile theory and experimental results was to refit the salicylate-albumin data to a Scatchard model with integral coefficients. The Scatchard plot of the salicylate binding data in Fig. 1 suggests the presence of at least four salicylate binding sites on the human albumin molecule. The fitted free-parameter results suggest that integral coefficients for a two-class model could occur in the following ways: 1,3; 1,4; 1,5; 2,2; 2,3; 2,4; or 3,3. A summary of the trial results for these models is given in Table 1. The "better" fitting 1,4 and 2,4 models are compared in Fig. 2, where each is plotted along with the free-parameter model. Neither the 1,4 nor the 2,4 model provides an acceptable representation of these data. The 1,4 model overpredicts the data at low $\bar{\nu}$ values, while the 2,4 model underpredicts the data in this region. The free-parameter model fits the data because its nonintegral coefficients interpolate between the two integer coefficient models. The inability to obtain a satisfactory fit with a two-class integer coefficient model demonstrates how the free-parameter model may give

TABLE 1

Summary of trial results for two-class Scatchard models

N_1	k_1	N_2	k_2	Sum of squares
1.27961	0.70856	3.80413	0.03272	0.0451
1	0.71625	3	0.07462	0.1233
1[a]	1.03138	4	0.03943	0.0524
1	1.22344	5	0.02572	0.1196
2	0.30897	2	0.05292	0.1810
2	0.36022	3	0.02362	0.0854
2	0.38288	4	0.01485	0.0641
3	0.20567	3	0.00692	0.1752

[a] Best fitting integer coefficient model.

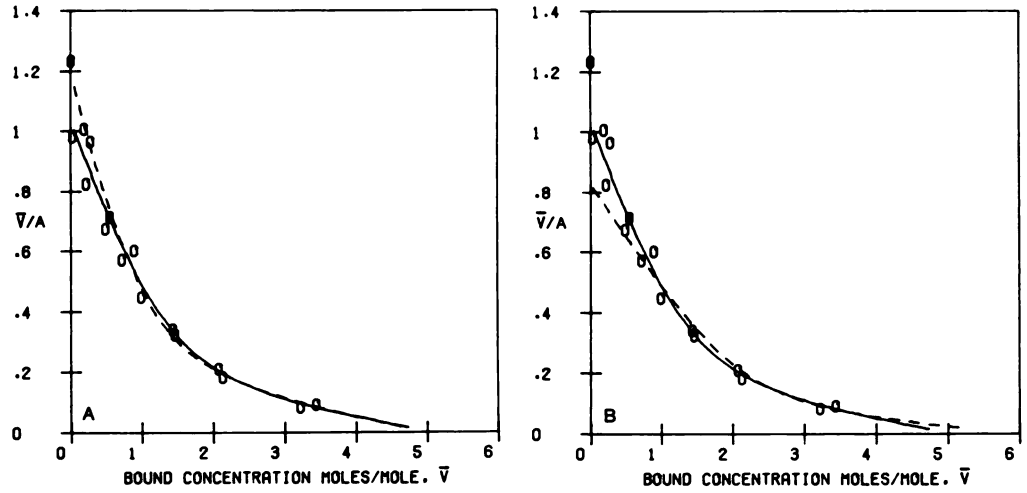

FIG. 2. *Comparison of 1,4 (A) and 2,4 (B) integer coefficient Scatchard models with fitted free-parameter model*

- - -, integer coefficient models.

misleading results. The free-parameter model contains more flexibility in data fitting than its assumptions permit. The finding that the "best" integer models here are not satisfactory indicates that further model refinement is necessary.

The results obtained by adding a third class of binding sites, giving a total of six free parameters in the model, are shown in the first line of Table 2. The sum of squares is reduced slightly compared with the two-class, free-parameter model given in the first line of Table 1. However, upon the addition of the third class of binding sites, the computer algorithm used for data fitting becomes poorly conditioned. This suggests that the model is inappropriate or that there may be too many free parameters to be determined accurately from this set of data. To relieve this ill conditioning, the number of free parameters in the model was reduced by assigning integer coefficients for each of the three classes of binding sites (14). A summary of these trials is given in the last eight lines of Table 2. The better fitting models reveal that the possible subdivisions of the three classes occur in the following ways: 1,1,4; 1,2,4; 1,2,3; and 1,1,3. These results suggest that human albumin may contain one strong site, one or two intermediate sites, and three or four weaker sites for binding salicylate. However, to distinguish the "best" among these models, based on the sum of squares, is quite subjective, with only a slight preference being given the 1,1,4 and 1,2,4 models. For purposes of discussion and comparison with a later analysis, the 1,1,4 model has been selected as the "best" fitting Scatchard model. The distinct classes of sites will be classified further in the subsequent analysis. Note that the sum of squares for this model with only three free parameters is very close to that of the free-parameter, two-class model, which has four free parameters (degrees of freedom).

Complex specific models. A second approach to the analysis of multiple binding is to classify the different liganded complexes rather than the specific sites on the protein to which the drug is bound. Using this approach, multiple binding is described as an ordered sequence of protein-drug equilibria represented by the equations

$$\begin{aligned} P \quad &+ A \rightleftharpoons PA_1 \\ PA_1 \quad &+ A \rightleftharpoons PA_2 \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ PA_{N-1} &+ A \rightleftharpoons PA_N \end{aligned} \tag{8}$$

where N represents the maximum number of binding reactions that can occur. This sequence represents multiple equilibria even when the binding induces allosteric changes in the protein binding sites (5, 19). The Scatchard model, i.e., Eq. 4, cannot describe such allosteric reactions. The stoichiometric association constants for the formation of each of these complexes are defined by

TABLE 2

Summary of trial results for three-class Scatchard models

N_1	k_1	N_2	k_2	N_3	k_3	Sum of squares
0.56462	1.16169	1.02576	0.32818	3.79857	0.02474	0.04488
1[a]	0.82449	1	0.15455	4	0.01615	0.04522
1	0.89220	2	0.08118	4	0.00721	0.04549
1	0.90510	2	0.07723	3	0.01088	0.04557
1	0.91598	1	0.09603	3	0.02928	0.04585
1	0.93633	2	0.06529	2	0.02333	0.04626
1	0.93945	3	0.05551	3	0.00328	0.04632
1	0.76964	1	0.18697	5	0.01133	0.04635
2	0.38289	2	0.01488	2	0.01482	0.06412
1	0.71627	1	0.07462	2	0.07462	0.12326

[a] Best fitting integer coefficient model.

$$K_1 = \frac{[PA_1]}{[P][A]}$$
$$K_2 = \frac{[PA_2]}{[PA_1][A]}$$
$$\vdots \qquad \vdots$$
$$K_N = \frac{[PA_N]}{[PA_{N-1}][A]} \tag{9}$$

The binding isotherm for the collective species as a function of free ligand concentration is represented by the rational function

$$\bar{\nu}(A) = \frac{K_1A + 2K_1K_2A^2 + \cdots + NK_1\cdots K_NA^N}{1 + K_1A + K_1K_2A^2 + \cdots + K_1\cdots K_NA^N} \tag{10}$$

where A is the unbound drug concentration and the K_i are the stoichiometric binding constants. This model is commonly called a stepwise equilibrium model (5, 19), or an Adair model (20).

In general, equilibrium measurements do not reveal the individual molecular pathways in a liganding process. Therefore the system of Eq. 8 cannot distinguish the specific site position influence of any cooperative phenomena. Instead this system identifies all constituent species having the same number of bound ligands with the same complex. (For a discussion of these points in depth see ref. 5.) The model Eq. 10 represents the same data analyzed in the previous section, but here the identification is with the drug-protein complexes rather than specific protein binding sites. This equation may be recast, for fitting purposes, in many equivalent forms. One approach is to define "macroscopic" constants,

$$B_1 = K_1$$
$$B_2 = K_1K_2$$
$$\vdots \qquad \vdots$$
$$B_N = K_1K_2\cdots K_N \tag{11}$$

With these definitions, Eq. 10 becomes

$$\bar{\nu}(A) = \frac{B_1A + 2B_2A^2 + \cdots + NE_NA^N}{1 + B_1A + B_2A^2 + \cdots + B_NA^N} \tag{12}$$

It has been shown that when the generalized Scatchard model (Eq. 4) is applicable, the macroscopic coefficients in Eq. 12 can be obtained from the constants in Eq. 4 (6, 16). If we number consecutively from 1 to N, repeating the constant for each class of sites according to its multiplicity (i.e., N_i sites in the class of affinity k_i means repeat k_i N_i times in the list), then

$$B_1 = k_1 + k_2 + \cdots + k_N = \sum_{i=1}^{N} k_i$$
$$B_2 = k_1k_2 + k_1k_3 + \cdots + k_{N-1}k_N = \sum_{i_1=1}^{N-1}\sum_{i_2=i_{1+1}}^{N} k_{i_1}k_{i_2}$$
$$B_3 = k_1k_2k_3 + \cdots + k_{N-2}k_{N-1}k_N = \sum_{i_1=1}^{N-2}\sum_{i_2=i_1+1}^{N-1}\sum_{i_3=i_2+1}^{N} k_{i_1}k_{i_2}k_{i_3} \tag{13}$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$B_p = \sum_{i_1=1}^{N-p+1}\cdots\sum_{i_p=i_{p-1}+1}^{p} k_{i_1}k_{i_2}\cdots k_{i_p},\ 1 \leqq p \leqq N$$

The stoichiometric constants can then be computed directly from Eqs. 13 according to

$$K_i = \frac{B_i}{B_{i-1}} \quad (14)$$

If we assume that all binding sites are available to bind drug and that all free sites are identical after each liganding reaction, these stoichiometric constants are related to the intrinsic site constants by the linear factor

$$k_i' = \frac{i}{N - (i-1)} K_i$$

This relationship is sometimes called a "statistical factor." In general, the actual pathways and site constants cannot be identified from the stoichiometry.

Data analysis with the stepwise equilibrium model. The salicylate binding data of Fig. 1 can be directly fitted using the stepwise equilibrium model. We begin the analysis by converting the site-specific constants for the 1,1,4 Scatchard model to the stoichiometric constants by using Eqs. 13 and 14. The results obtained are shown in the first line of Table 3. Since the stepwise model is not restricted to the assumption of independent, noninteracting sites, it is possible that values other than those given in the first line of Table 3 can provide a representation of the binding data. This is considered by using the values in the first line of Table 3 as starting estimates for the parameters in Eq. 10 and refining these values by fitting Eq. 10 directly to the salicylate binding data. Since six free parameters generate a poorly conditioned system for fitting these data, we initially have restricted the stepwise model to the case of six identical sites with negative cooperativity. Negative cooperativity is imposed by constraints

$$\frac{i}{i+1} K_i > \frac{N-i+1}{N-i} K_{i+1} \quad (15)$$

We consider a negative cooperativity model first, because it is known that negative cooperativity and independent classes of sites are not distinguishable by model fitting or by graphical analysis of equilibrium data (16, 4).

As a result of directly fitting the stepwise equilibrium model, a new, lower minimum was found for the parameter values shown in the third line of Table 3. Note that the direct fitting of this model resulted in the adjustment of K_6 to an extremely small value. If K_6 is set equal to zero, the remaining constants are adequate to describe the salicylate data at the same sum of squares (fourth line of Table 3). This reduction suggests that perhaps only five complexes are identifiable in these data. Therefore it is possible that cooperativity among only five binding sites instead of six identical interacting sites, or six independent sites of varying affinity, could have produced the binding data shown in Fig. 1.

Since these new values do not satisfy the constraints for negative cooperativity for five identical interacting sites, the constraints were removed and the five parameters were permitted to assume whatever values produce a lower sum of squares. The final parameter values are given in the fifth line of Table 3. These values suggest that the successive liganding reactions may induce both negative and positive cooperativity.

In an attempt to identify the minimal parameter model, we also set $K_5 = 0$ and refitted Eq. 10 to the data with only four

TABLE 3

Summary of trial results for stepwise equilibrium models

K_1	K_2	K_3	K_4	K_5	K_6	Sum of squares
1.04363	0.184190	0.050870	0.022080	0.010250	0.003920	0.04522
1.06001	0.177662	0.053073	0.022485	0.009721	0.003182	0.04510
1.04892	0.183798	0.048089	0.027050	0.009147	2.78×10^{-16}	0.04478
1.04892	0.183798	0.048089	0.027050	0.009147	0.0	0.04478
1.01497	0.206349	0.028508	0.070540	0.000601	0.0	0.04455
1.01111[a]	0.209073	0.026355	0.079368	0.0	0.0	0.04455

[a] Best fitting stepwise model.

adjustable parameters. The four adjusted parameters are given in the sixth line of Table 3. Note that reduction of the model did not increase the sum of squares. These parameter values should be compared with results in the first two lines of Table 2, and with the first, third, fourth, and fifth lines of Table 3.

In contrast to the Scatchard model results, the adjusted stepwise parameter values suggest mixed cooperativity among four salicylate binding sites. This contrast is further illustrated by the Klotz affinity plots [iK_i vs. i (5)] shown in Fig. 3. The stoichiometric constants generated by the five-site and four-site stepwise models suggest an initial affinity decrease, then a positive allosteric effect, possibly occurring after the third salicylate complex is formed. We emphasize again that the above results do not prove cooperativity or even identify it. They merely show that a mathematical model that is consistent with cooperativity can be fitted to the binding data. In fact, this model gives a slightly better fit of the experimental data with exactly the same number of adjustable parameters.

Generalized mathematical formulation. In order to resolve the difficulties encountered with the Scatchard and stepwise models, we have developed a new mathematical expression that includes the Scatchard model as a special case. This formulation is also directly equivalent to the stepwise model and hence is appropriate for fitting to all equilibrium data, including those exhibiting cooperativity. This formulation will aid in explaining the contrasting results obtained in the previous sections.

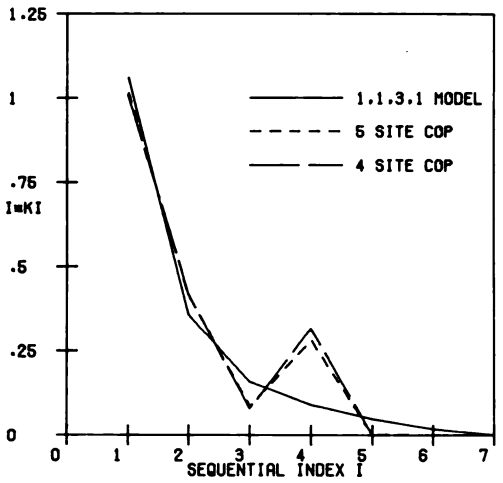

FIG. 3. *Klotz affinity plot (5) for stoichiometric binding constants from a five-site and a four-site stepwise equilibrium model*

COP = cooperative or allosteric model.

Equation 12 can be written in the alternative mathematical form

$$\bar{\nu}(A) = A \frac{d}{dA}[\log_e P_N(A)] \qquad (16)$$

where $P_N(A) = 1 + B_1A + B_2A^2 + \cdots + B_NA^N$. The polynomial $P_N(A)$ can be expressed in terms of the factors of its N roots (16, 21),

$$P_N(A) = (A - R_1)(A - R_2) \cdots (A - R_N) \cdot B_N \qquad (17)$$

It can be shown by direct substitution of Eq. 17 into Eq. 16 that Eq. 12 (and therefore Eq. 10) has the alternative mathematical expression

$$\bar{\nu}(A) = \frac{A}{A - R_1} + \frac{A}{A - R_2} + \cdots + \frac{A}{A - R_N} \qquad (18)$$

Since the macroscopic constants of Eqs. 11 are all positive, it follows that all real roots R_i in Eq. 17 must be negative; therefore, by setting $k_i = -1/R_i$ and rearranging, it is seen that Eq. 18 and Scatchard's model, Eq. 4, are identical. This artifact of mathematical equivalence means that one cannot confirm that the Scatchard type of simple binding has occurred by merely fitting Eq. 4 to data. That is, success in fitting Eq. 4 to data does not necessarily imply that the derived constants are site-specific or microscopic site binding parameters. One may, in fact, only have located the real roots R_i in the polynomial $P_N(A)$ of Eq. 16 or its equivalent, Eq. 17.

Equation 18 can itself be fitted directly to the data on salicylate–human serum albumin binding. The starting estimates for finding values for the R_i by directly fitting Eq. 18 are obtained from the 1,1,4 Scatchard model. That is, we set $R_1 = -1/k_1$;

$R_2 = -1/k_2$; and $R_3 = R_4 = R_5 = R_6 = -1/k_3$. These estimates are refined by fitting Eq. 18 directly to the data of Fig. 1. The resulting fitted parameters are given in the first line of Table 4. We are able to derive three distinct roots, and one multiple root of multiplicity 3. That is, the 1,1,4 Scatchard model has been refined to a 1,1,3,1 Scatchard model. The refined "Scatchard" constants are given in parentheses in the second line of Table 4, and the stoichiometric constants derived from them, using Eqs. 13 and 14, are given in the second line of Table 3. The affinity curve generated by these parameters is shown as the solid line in Fig. 3.

This refinement shows that the common practice of lumping weaker sites into a large class may not be an appropriate procedure. Theoretically this is clear, for the number of repetitions of a root in Eq. 17 has no relationship to the magnitude of the root. Therefore "site" groupings need not occur in any particular order in the equivalent "Scatchard" model. The analysis with Eq. 18 provides a refined five-site Scatchard model containing only four parameters and probably reveals the most that is extractable from these data with a Scatchard model.

However, it remains to explain or otherwise interpret the slightly better fitting stepwise models that require only four or five, possibly interacting, binding reactions to explain the data. We have obtained only Scatchard model parameters because we have implicitly forced all the R_i to be real numbers. This is equivalent to requiring that a Scatchard model fit the data. However, there is no requirement a priori that all roots in Eq. 17 be real numbers. Therefore the possible occurrence and meaning of pairs of complex conjugate roots must be considered. This can be done by combining pairs of terms in Eq. 18 to obtain alternative compound terms of the form

$$\bar{\nu}_j(A) = \frac{A(2A - R_j^a)}{A^2 - R_j^a A + R_j^m} \qquad (19)$$

where the real parameters $R_j^a/2$ and R_j^m will represent, respectively, the real parts and moduli of complex conjugate pairs of roots R_j and $\bar{R}_j$, and the subscript j will index the compound terms. Because Eq. 19 will also represent pairs of real roots, there is no loss in model generality by considering terms of this type. The most general form of the model can then be expressed as

$$\bar{\nu}(A) = \sum_{j=1}^{p} \frac{A(2A - R_j^a)}{A^2 - R_j^a A + R_j^m} + \sum_{i=1}^{m} N_i \frac{A}{A - R_i} \qquad (20)$$

where $N = 2p + \sum_1^m N_i$ is the total number of binding sites on the protein. The derived

TABLE 4

Summary of trial results for alternative mathematical model

Simple model						
R_1, k_1	R_2, k_2	R_3, k_3	R_4, k_4	R_5, k_5	R_6, k_6	Sum of squares
−1.17302[a]	−7.29157	−46.62602	−46.62602	−46.62602	−165.919	0.04510
(0.85250)[a]	(0.13715)	(0.02145)	(0.02145)	(0.02145)	(0.00603)	0.04510
Compound model						
$-R_1^a$	R_1^m	N_1	R_1	N_2	R_2	Sum of squares
−7.684	7.848	4.0	−61.929	0.0	0.0	0.04522
−8.405	8.507	2.0	−40.262	2.0	−97.338	0.04512
−8.460	8.549	3.0	−46.666	1.0	−165.312	0.04510
−11.505	11.368	3.0	−34.151	0.0	0.0	0.04585
−35.265	37.307	2.0	−34.138	1.0	−10.414	0.04585
−5.618	476.196	2.0	−2.176	0.0	0.0	0.04588
−7.425[a]	434.447	1.0	−1.367	1.0	−3.807	0.04455

[a] Best fitting alternative model.

parameters in this expression do not yield the site-specific binding constants directly, but they do represent the roots of the polynomial in Eq. 17. These roots can be converted to the stoichiometric binding constants by the following procedure. Formally define $k_i = -1/R_i$ for each root, real or complex, in Eq. 17. Then Eqs. 13 and 14 can be used to produce the desired stoichiometric constants.

The occurrence of pairs of complex roots in Eq. 17 probably reflects site interaction, i.e., allosteric phenomena in the binding. Intuitively this is consistent, since in such cases a site constant is not "constant" but depends on the state of neighboring sites. Thus one should not be able to extract independent parameters such as site-specific constants from the binding data. Rather, the site activities "interact," and this interaction is expressed by terms of the type given in Eq. 19 (16, 17). Therefore, a mathematical model that can account for allosteric site interaction, is a direct equivalent of the stepwise model, and contains the Scatchard model is expressed by Eq. 20.

The parameters in the compound terms of Eq. 20 can be examined for complex roots by using the relations $R_j^1 + R_j^2 = R_j^a$ and $R_j^1 \cdot R_j^2 = R_j^m$. These equations lead directly to the familiar quadratic formulae

$$R_j^1 = \frac{R_j^a + (R_j^{a2} - 4R_j^m)^{1/2}}{2}$$
$$R_j^2 = \frac{R_j^a - (R_j^{a2} - 4R_j^m)^{1/2}}{2} \qquad (21)$$

Complex roots are present, of course, whenever $R_j^{a2} - 4R_j^m < 0$.

The generalized formulation, as expressed by Eq. 20, has a relatively simple mathematical expression with a minimum number of independent parameters, and hence is simple to fit to data. Yet it is free of the undesirable assumption of independence of binding sites. Moreover, this formulation does not require one parameter for each liganding reaction as does the stepwise model. This latter characteristic often limits the use of the stepwise model when there is a large number of protein-ligand complexes. Finally, possible site interaction is detected by the determination of parameters which generate complex roots in Eqs. 21. We emphasize "possible site interactions," since independent validation of such phenomena is necessary before such interactions can be claimed.

This new formulation yields results that are equivalent to those obtained for the salicylate binding data using both the Scatchard and stepwise equilibrium models. Only the results corresponding to the four-parameter models will be presented, for these values correspond to the best results found with those models. If $p=1$ and $m=2$ in Eq. 20, N_1 and N_2 can be varied through values such that $N_1 + N_2 \leqq 4$. In this way, up to six possible binding sites can be examined with only four parameters. The results are listed under the heading "Compound model" in Table 4. The parameter values in the first five lines of this section of Table 4 correspond respectively to 1,1,4; 1,1,2,2; 1,1,3,1; 1,1,3; and 1,1,2,1 Scatchard models. No complex roots are derived with the generalized model until the total number of binding sites is reduced to 4. The results in the last two lines of Table 4 reveal that for $N_1 = 2$, $N_2 = 0$, and for $N_1 = 1$, $N_2 = 1$, complex roots are derived when these models are fitted to the data. This suggests that site interaction, which is indicated by the derived parameter values, could be an artifact of fitting the data with a model containing too few binding sites rather than true binding site interactions. The results in this table include the parameters of the best models in both Tables 2 and 3, and the results under "Simple model" in the first line of Table 4 are rederived as the third line under "Compound model." The generalized formulation thus demonstrates that it is an alternative to both the Scatchard model and the stepwise model.

The best results of fitting the generalized formulation verify that two quite different binding models are compatible with the binding data of Fig. 1. The first model is a 1,1,3,1 Scatchard model, which corresponds to all real roots in Eq. 17. The second model is a four-site stepwise model, which suggests mixed site interaction and gives rise to two real and two complex

conjugate roots in Eq. 17. The fitted curves corresponding to each of these models are shown in a Scatchard plot in Fig. 4.

INTERPRETATION OF MODELING RESULTS

Model fitting results are often used to interpret and predict the individual site or site class contributions to total drug-protein binding. In this context it is important that the model be as nearly correct as the data permit, and that possible alternative models be examined. For example, a secondary site analysis of salicylate binding to albumin gives a very different picture when the results using a 1,1,3,1 model are compared to Mais' findings with a two-class, nonintegral model (1). The contrasting results are shown Figs. 5 and 6, where we have classified as primary ($\bar{\nu}_p$) the 1,1 sites and the remaining 3,1 sites are labeled as secondary ($\bar{\nu}_s$).

A more serious concern is raised by the analytical results, which showed that the Scatchard model binding constants need not represent the affinity constants associated with any binding site. This was shown to be the case even when the bind-

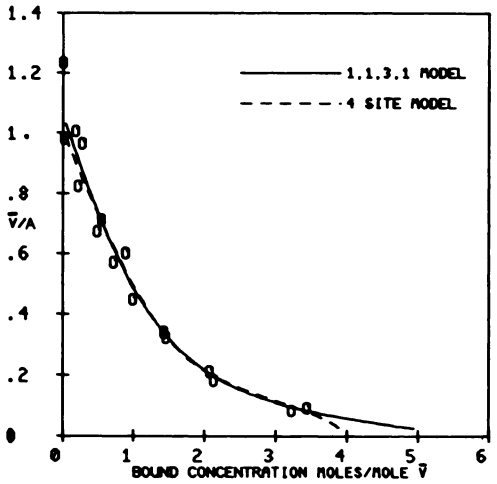

FIG. 4. *Comparison of fitted curves on a Scatchard plot for a 1,1,3,1 Scatchard model and a four-site stepwise equilibrium model*

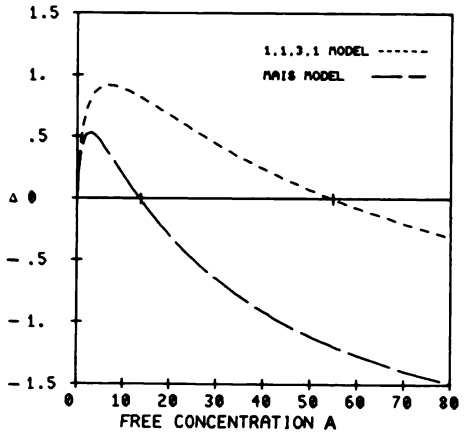

FIG. 6. *Comparison of net contribution of secondary sites to binding as measured by the class difference ($\Delta = \bar{\nu}_p - \bar{\nu}_s$) plotted against free drug concentration*

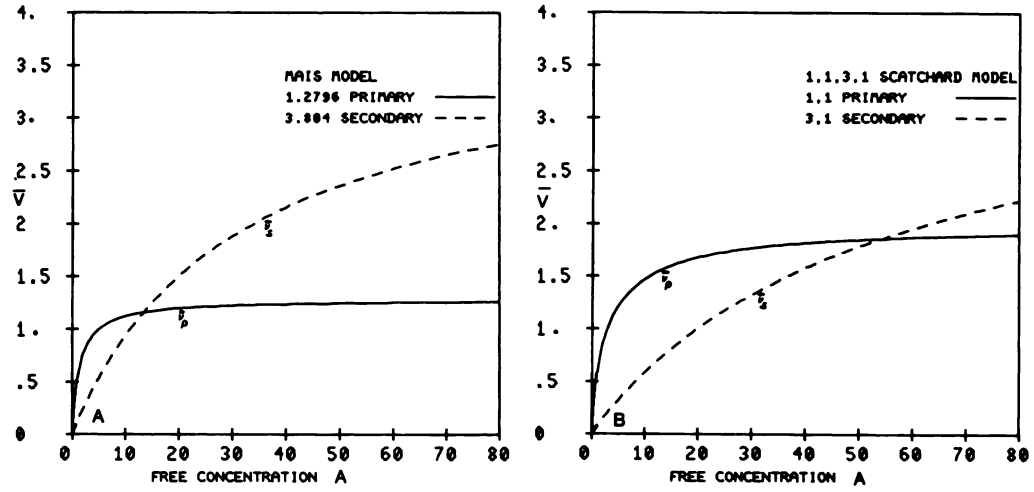

FIG. 5. *Comparison of primary and secondary binding (1), using a four-parameter free model (A) and a four-parameter integer coefficient model (B)*

ing parameters were obtained from a Scatchard model with integer coefficients. In general, the parameters only approximate the roots of Eq. 17. Such results confirm that it is impossible, by equilibrium binding measurements, to distinguish independent classes of noninteracting binding sites from sites that interact negatively. Therefore site-specific analyses may be without valid meaning in the absence of independent validation that this type of binding actually occurs. On the other hand, the stoichiometric constants (B_i) are always meaningful, and can be computed from the parameters derived by fitting Eq. 20, then using Eqs. 13 and 14 to obtain the stoichiometric constants. The resulting stoichiometric constants are valid even though the parameters derived from fitting Eq. 20 do not necessarily represent site-specific binding affinities.

It has also been clearly demonstrated that the parameters obtained from different models cannot be directly interchanged, as is sometimes assumed (13). This is clearly illustrated by the fact that the "best" Scatchard constants from the second line of Table 4, the equivalent stoichiometric constants in the second line of Table 3, and the parameters from the alternative model given in the fifth line of Table 4 are not identical.

The second four-parameter model that we have derived for the salicylate binding data is the four-parameter stepwise model or its equivalent, the generalized formulation, given in the last line of Table 4. This model suggests possible allosteric site interaction in the binding process and, on the basis of sum of squares, gives a better fit of the experimental data points than does the Scatchard model. However, this model implies that there may be only four interacting binding sites rather than six independent sites. The primary difficulty with this result is to determine whether it reflects the actual drug binding or is an artifact of the model fitting process.

To gain some insight into this crucial point, we have computed noise-free data by evaluating the best Scatchard model (1,1,3,1) at the free drug concentrations of the data shown in Fig. 1. These noise-free data are then fitted by the stepwise model with first six, then five, and finally four adjustable parameters. The results are given in Table 5. Note the similarity of the values in the third line of Table 5 to those in the last line of Table 3. It is clear from these results that truncating the stepwise model to fewer parameters than available binding sites (i.e., six sites) forces the parameters to adjust in such a way as to suggest cooperative binding when, in fact, no such binding need be present. Similar results are obtained when the generalized model is used with the coefficient values $N_1 = 1$ and $N_2 = 1$. The derived parameters are given in the last line of Table 5 and should be compared with the last line of Table 4. This truncation effect demonstrates the importance of having a good estimate of the actual number of drug binding sites on a protein before the stepwise equilibrium model can be fitted and expected to give reliable results. A part of the problem with the analysis of salicylate binding is that the data are sparse in the $\bar{\nu}$

TABLE 5

Summary of trial results for generated noise-free data

Stepwise model						
K_1	K_2	K_3	K_4	K_5	K_6	Sum of squares
1.06001[a]	0.177662	0.053073	0.022485	0.009721	0.003182	0.00000
1.06070	0.177157	0.053860	0.020910	0.013034	0.000	0.00000
1.01959	0.205092	0.0276861	0.076251	0.0	0.0	0.00033
Alternative model						
R_1[a]	R_1[m]	N_1	R_1	N_2	R_2	Sum of squares
−8.460[a]	8.549	3.0	−46.666	1.0	−165.312	0.00000
−7.810	427.044	1.0	−1.333	1.0	−3.981	0.00033

[a] Exact parameter values for generated data.

range 2.0–3.0, and the $\bar{\nu}$ range near site saturation is not well defined. Thus one cannot precisely identify site interaction or the maximum number of binding sites available. Such limitations must be considered when attempting to model any set of binding data. However, the modeling efforts have identified ranges where additional data are needed to gain more insight into the binding process – information that is essential for designing further binding experiments.

In summary, simple Scatchard models with nonintegral coefficients provide an inadequate model for interpretation of the binding of salicylate to human serum albumin. As an alternative, we have derived a generalized formulation that includes the best features of both the Scatchard model and the stepwise equilibrium formulation. This formulation is simpler to fit to data than the stepwise model, yet is not restricted to the assumption of independent and noninteracting binding sites. It would seem appropriate to adopt this mathematical formulation for the analysis of all equilibrium binding data, including the binding of drugs to proteins.

While the precise details of modeling brought out here are specific to the salicylate–human albumin binding system, the procedures and models apply to more general studies of macromolecule-ligand binding. Such alternative modeling formulations should be considered before a model is accepted as a basis for interpretation of experimental results. As was demonstrated by our analysis of the salicylate binding data, the proper alternative mathematical formulations and sophisticated computerized fitting techniques may still be inadequate to define precisely a single model for a given set of experimentally determined data.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the generous cooperation of Dr. Roland F. Mais of the Veterans Administration Hospital, Neuropharmacology Research Service, Hines, Illinois, in supplying the data used to illustrate the methodology in this paper. We also acknowledge the technical assistance of Mrs. Elizabeth Fleetwood in the preparation of the manuscript.

REFERENCES

1. Mais, R. F., Keresztes-Nagy, S., Zaroslinski, J. F. & Oester, Y. T. (1974) *J. Pharm. Sci.*, **63**, 1423–1427.
2. Scatchard, G. (1949) *Ann. N. Y. Acad. Sci.*, **51**, 660–6762.
3. Steinhardt, J. & Reynolds, J. (1969) in *Multiple Equilibria in Proteins*, pp. 1–9, Academic Press, New York.
4. Hammes, G. G. & Wu, C. W. (1974) *Annu. Rev. Biophys. Bioeng.*, **3**, 1–33.
5. Klotz, I. M. & Hunston, D. L. (1975) *J. Biol. Chem.*, **250**, 3001–3009.
6. Scatchard, G., Hughes, W. L., Jr., Gurd, F. R. N. & Wilcox, P. E. (1954) in *Chemical Specificity in Biological Interactions*, pp. 193–219, F. R. N. Gurd, Ed. Academic Press, New York.
7. Klotz, I. M. & Hunston, D. L. (1971) *Biochemistry*, **10**, 3065–3069.
8. Priore, R. L. & Rosenthal, H. E. (1976) *Anal. Biochem.*, **70**, 231–240.
9. Zaroslinski, J. F., Keresztes-Nagy, S., Mais, R. F. & Oester, Y. T. (1974) *Biochem. Pharmacol.*, **23**, 1767–1776.
10. Kresztes-Nagy, S., Mais, R. F., Oester, Y. T. & Zaroslinski, J. F. (1972) *Anal. Biochem.*, **48**, 80–89.
11. Ashbrook, J. D., Spector, A. A. & Fletcher, J. E. (1975) *J. Biol. Chem.*, **250**, 2333–2338.
12. Ashbrook, J. D., Spector, A. A. & Fletcher, J. E. (1972) *J. Biol. Chem.*, **247**, 7038–7042.
13. Wosilait, W. D. & Soler-Argilaga, C. (1975) *Life Sci.* **17**, 159–166.
14. Fletcher, J. E. & Spector, A. A. (1968) *Comput. Biomed. Res.*, **2**, 164–175.
15. Fletcher, J. E. & Shrager, R. I. (1968) *DCRT Technical Report No. 1*, National Institutes of Health, Bethesda, Md.
16. Fletcher, J. E., Spector, A. A. & Ashbrook, J. D. (1970) *Biochemistry*, **9**, 4580–4587.
17. Fletcher, J. E. & Ashbrook, J. D. (1973) *Ann. N. Y. Acad. Sci.*, **226**, 69–81.
18. Spector, A. A. (1975) *J. Lipid Res.*, **16**, 165–189.
19. Klotz, I. M. (1946) *Arch. Biochem.*, **9**, 109–117.
20. Adair, G. S. (1925) *J. Biol. Chem.*, **63**, 529–545.
21. MacDuffee, C. C. (1954) *Theory of Equations*, pp. 29–95, Wiley, New York.