A General Method for the Quantitative Determination of Saturation Curves for Multisubunit Proteins*

Athel Cornish-Bowden and D. E. Koshland, Jr.

ABSTRACT: A curve-fitting procedure has been developed for determining the binding constants of a multiple subunit protein showing cooperative effects. The procedure is mathematically simple and provides a method of estimating the precision of the assigned parameters. Testing the procedure with a wide range of simulated data gave satisfactory results in all cases. Application to a tetrameric protein with simulated data approximating typical equilibrium dialysis results indicates that 15–20 observations spread over the middle portion of a saturation curve are sufficient to give accurate assessments of intrinsic binding constants K_1', K_2', etc. The conformational model which can explain the cooperative effects can usually be identified from the relative values of these intrinsic constants.

The quantitative expression of a saturation curve for a regulatory protein is needed for two main reasons. In the first place, most regulatory proteins are influenced by many substrates and effectors. Some of the effectors are important at physiological concentrations while others may be evident only in artificial situations. In order to ascertain the role of an effector molecule its quantitative influence on the enzyme it regulates must be evaluated in the presence of other ligands at physiological concentrations. Only with such quantitative evaluation can the roles of substrate and effector concentrations in the cellular economy be understood (*cf.* review by Atkinson (1966) for the importance of relating effector levels to physiological conditions).

This need at the metabolic level is matched by a second one at the molecular level. Current attempts to explain the behavior of regulatory proteins by molecular models (Monod *et al.*, 1965; Koshland *et al.*, 1966) lead to predictions of saturation curves which are often significantly different from each other. In Table I, for example, are shown the intrinsic constants predicted by some current models for the binding of a ligand to a tetrameric protein. Since the relationships between the constants are correlated with differences in protein interactions, an analysis which provided a unique identification of these constants would reveal important properties of the regulatory protein. In the case of glyceraldehyde 3-phosphate dehydrogenase from rabbit muscle, an extreme example of negative cooperativity was uncovered which allowed the individual determination of the successive binding constants (Conway and Koshland, 1968). With many proteins, particularly those exhibiting positive cooperativity, the binding of the second molecule of substrate to a given species occurs before the binding of the first molecule is complete and so forth. In that case only a detailed analysis of the saturation curve can disentangle the individual binding parameters. A curve-fitting procedure which would evaluate these constants could provide this information.

The classical saturation curve for binding of a ligand to a multisubunit protein can be expressed in the form of eq 1.

$$N_X = \text{number of moles, X, per mole of protein} = \frac{K_1(X) + 2K_1K_2(X)^2 + 3K_1K_2K_3(X)^3 + 4K_1K_2K_3K_4(X)^4}{1 + K_1(X) + K_1K_2(X)^2 + K_1K_2K_3(X)^3 + K_1K_2K_3K_4(X)^4} \quad (1)$$

In this equation the constants K_1, K_2, K_3, and K_4 represent the binding of the first, second, third, and fourth molecules of ligand as shown in eq 2. In general, equations which

$$K_1 = \frac{(EX)}{(E)(X)} \quad K_2 = \frac{(EX)_2}{(EX)(X)} \quad K_3 = \frac{(EX_3)}{(EX_2)(X)} \quad K_4 = \frac{(EX_4)}{(EX_3)(X)} \quad (2)$$

are nonlinear cannot be analyzed by the standard statistical methods of obtaining a least-squares fit to a linear equation. Since eq 1 is identical with the equation for the formation of a complex ion, it might be expected that the methods which have been developed for analyzing such data might be relevant to the present problem. Unfortunately this is not the case. The graphical methods of Leden (1941) and Fronaeus (1950) require an independent measure of the unliganded metal ion (or free enzyme) concentration which is not generally available in enzyme systems. The method of Schroder (1966) is applicable in principle to enzyme chemistry but is unsuitable because it requires a small number of highly accurate measurements of the ligand concentration and the fraction saturation. In most enzyme systems highly accurate values cannot be obtained and it is usually necessary to use a larger number of less accurate measurements.

General methods, including computer programs, are available for analyzing nonlinear systems (see, for example, Swann, 1969). However, the most widely used methods require good starting guesses for the parameters, which are generally not available in protein binding studies. Furthermore, it is desirable for protein studies to have a method which is simple and can be followed in detail in case anomalous results must be interpreted. The development and analysis

* From the Department of Biochemistry, University of California, Berkeley, California 94720. *Received March 2, 1970.* This work was supported in part by research grants from the U. S. Public Health Services (AM-GM-09765) and the National Science Foundation (GB-7057).

TABLE I: Relationship of Intrinsic Constants Describing the Saturation Curve of a Tetrameric Protein for Various Models.

Models	Relationship of Intrinsic Binding Constants[a]
Identical binding sites, no subunit interactions (*i.e.*, Michaelis–Menten)	$K_1' = K_2' = K_3' = K_4'$
Nonidentical, independent binding sites	$K_1' > K_2' > K_3' > K_4'$
Simplest sequential model:	
(a) Square	$\frac{1}{3} \leqslant \frac{K_2'}{K_1'} = \frac{K_4}{K_3'} \geqslant \frac{K_3'}{K_2'}$
(b) Tetrahedral	$\frac{K_2'}{K_1'} = \frac{K_3'}{K_2'} = \frac{K_4'}{K_3'}$
Symmetry model:	
(a) Binding to one conformation only	$K_1' < K_2' = K_3' = K_4'$
(b) Binding to both conformations	$K_1' < K_2' \leq K_3' \leq K_4'$
Simple positive cooperativity	$K_1' < K_2' < K_3' < K_4'$
Simple negative cooperativity	$K_1' > K_2' > K_3' > K_4'$
Mixed positive and negative cooperativity	No restrictions on K_i' values
Mixture of isozymes	Does not fit Adair equation, thus there are no K_i' terms

[a] In some cases the restrictions are stronger than those listed; *e.g.*, the symmetry model with binding to both conformations is more restricted than simple positive cooperativity, because of additional restrictions on the values of the K' terms which are too complicated to list in this table. Intrinsic constants, K_i', are association constants, K_i, corrected for statistical factors.

of such a procedure, which allows the quantitative fitting of saturation curves and the determination of standard errors of the assigned parameters, is the subject of this paper.

It was found during this analysis that the curve-fitting procedures could be applied more readily with higher efficiency if the standard binding equation were rewritten in the form of eq 3 (Adair, 1925). The evaluation of ψ values can

$$N_X = \frac{\psi_1(X) + 2\psi_2(X)^2 + 3\psi_3(X)^3 + 4\psi_4(X)^4}{1 + \psi_1(X) + \psi_2(X)^2 + \psi_3(X)^3 + \psi_4(X)^4} \quad (3)$$

then lead directly to the calculation of the individual binding constants (which are true thermodynamic parameters) as shown in eq 4.

$$\psi_1 = K_1 \qquad \frac{\psi_2}{\psi_1} = K_2 \qquad \frac{\psi_3}{\psi_2} = K_3 \qquad \frac{\psi_4}{\psi_3} = K_4 \quad (4)$$

Theoretical Consideration of Curve-Fitting Procedure

The general binding equation can be expressed in a number of ways of which two are shown in eq 1 and 3. Although eq 1 is attractive in the sense that the ultimate quantities to be determined, K_1, K_2, K_3, etc., appear in the equation, their appearance as product terms makes curve-fitting procedures more difficult to apply. On the other hand, it was found that the use of eq 3 allowed a convenient evaluation of the ψ terms. Since the individual K values can then be determined from eq 4, we shall confine our discussion to an analysis of the curve-fitting procedures for eq 3.

The most widely recognized criterion for estimating goodness of fit between a set of experimental points and a theoretical curve is the value of the sum of the squares of deviations from the fitted equation, as shown in eq 5, where the sum of squares is designated by ϵ.

$$\epsilon = \Sigma(N_{X_i(obsd)} - N_{X_i(calcd)})^2 \quad (5)$$

The procedure will be applied for illustrative purposes in this paper to the equation for a tetrameric protein, because this is the most complex case commonly encountered, but the same procedure can be applied to a similar binding equation with smaller or larger numbers of interacting sites.

Although in theory it might be possible to minimize the sum of squares by trial and error with respect to one parameter, keeping the others constant, repeat with each parameter in turn, and repeat the whole sequence until a minimum is reached, this would be extremely laborious in practice. Instead it was found possible to estimate the position of the minimum with a small number of measurements of the error function if certain assumptions were made about the shape of the error curve. The true shape of a plot of the sum of squares, ϵ, against ψ_i, although not a simple equation, has two general features. Firstly, there is a wide region around the minimum in which the second derivative of ϵ with respect to ψ_i is positive. Secondly, in all of the examples we have examined the curve shows only one minimum. Consequently it is not unreasonable to use the quadratic equation of a parabola as a first approximation for the ϵ function as shown in eq 6. The first and second derivatives of this approximation are given in eq 7 and 8. If we

$$\epsilon = a + b\psi_i + c\psi_i^2 \quad (6)$$

$$\frac{\partial \epsilon}{\partial \psi_i} = b + 2c\psi_i \quad (7)$$

$$\frac{\partial^2 \epsilon}{\partial \psi_i^2} = 2c \quad (8)$$

use the notation ψ_{i_k} to represent the value of ψ_i at the kth iteration, we can in principle measure the value of ϵ, its first and second partial derivatives with respect to ψ and have three equations which can be solved for three unknowns, a, b, and c of eq 6–8. It may readily be shown that the relationship between ψ_{i_k} and $\psi_{i_{k+1}}$ is given by eq 9 and this equation

$$\psi_{i_{k+1}} = \psi_{i_k} - \frac{\partial \epsilon / \partial \psi_i}{\partial^2 \epsilon / \partial \psi_i^2} \quad (9)$$

can be used repeatedly to obtain better approximations until a minimum in the sum of squares is reached.

There are several practical problems in this approach. In the first place negative values of ψ_i are physically meaningless but the use of eq 9 does not prevent negative values from being assigned. Secondly, in the vicinity of the minimum eq 9 is not an accurate representation of the plot of ϵ *vs.* ψ_i and consequently the iteration formula is not perfectly accurate. In some cases, it may call for changes in ψ_i which are so large that the series diverges instead of converges. Thirdly, if the value of ψ_{i_k} is very different from the value of the true minimum, the second derivative may be negative in which case eq 9 is grossly in error and will result in successively worse approximations. Fourthly, it has not been demonstrated that the plot of ϵ *vs.* ψ_i is a simple curve with a single minimum. It there were more than one minimum, the iteration formula would produce convergence toward the minimum closest to the starting point, which might not necessarily be the deepest minimum.

Fortunately it was found that all of these objections could be avoided fairly simply. To meet the first objection it was found that calculations could be carried out in terms of ln ψ_i instead of ψ_i, representing ϵ by the approximation shown in eq 10, thereby excluding negative values of ψ_i.

$$\epsilon = a + b \ln \psi_i + c(\ln \psi_i)^2 \qquad (10)$$

In this case the iteration formula of eq 9 is replaced by eq 11.

$$\ln \psi_{i_{k+1}} = \ln \psi_{i_k} - \frac{\partial\epsilon/\partial \ln \psi_i}{\partial^2\epsilon/\partial(\ln \psi_i)^2} \qquad (11)$$

This change had the further fortuitous but very satisfactory advantage of disposing of most of the second objection since a plot of ϵ *vs.* ln ψ_i is in general described very closely by a quadratic over a range of about one order of magnitude. This is shown in Figure 1 where a plot of the true ϵ *vs.* ln ψ_i is compared to a parabolic curve. It is seen here that the assumed equation for the sum of squares is a reasonable approximation for the true one.

The danger of making too large a change in ψ_i in one iteration was avoided by (a) setting a maximum limit to the change which is permitted, and (b) checking that the new value of ϵ is in fact smaller than the previous value. If the new value is greater than the previous value, the procedure requires that a smaller change in the value be made, *e.g.*, half the amount indicated by the iteration formula. In practice it has proved to be very rare that it is necessary to reduce the magnitude of the correction in this way.

A simple criterion was developed to determine whether or not the third objection pertained: if $\partial^2\epsilon/\partial(\ln \psi_i)^2 \lesseqgtr 0$, then eq 10 must not be used. Instead an improvement in ϵ can be obtained by making a change of fixed magnitude in a direction indicated by the sign of $\partial\epsilon/\partial \ln \psi_i$. Thus, eq 11 is replaced by eq 12, where the sign in front of α is opposite

$$\ln \psi_{i_{k+1}} = \ln \psi_{i_k} \pm \alpha \qquad (12)$$

to that of $\partial\epsilon/\partial \ln \psi_i$. Experience has shown that for equations of the type used to describe regulatory proteins a value of 2.3 (*i.e.*, ln 10) is appropriate for α, being small enough to ensure convergence in almost every case, and large enough to minimize the number of iterations required.

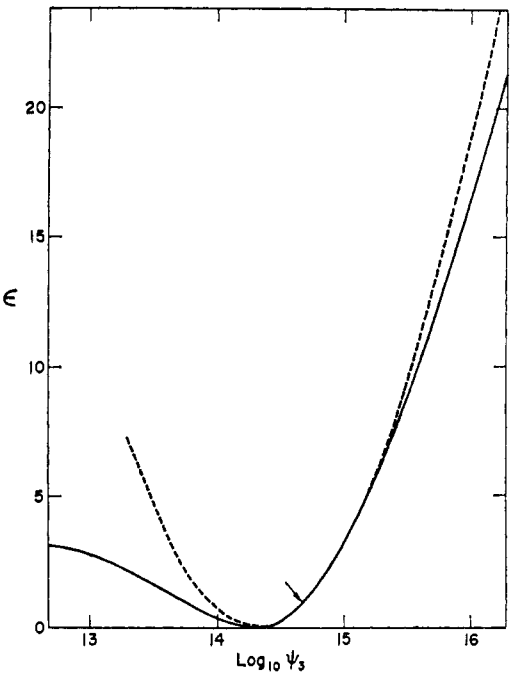

FIGURE 1: Comparison of the true sum of squares of the errors for a binding equation with a quadratic curve. The values of the sum of squares, ϵ (solid curve), were obtained for curves of the binding of NAD to yeast glyceraldehyde 3-phosphate dehydrogenase (see Results), by determining ϵ at different values of ψ_3, with the other three parameters held constant at their best fit values. The dashed line represents a quadratic function obtained as described in the text by measuring the derivatives of the real curve at the point marked with an arrow.

Although no theoretical test has been devised to show whether there is only one minimum value of ϵ, a convenient practical test has been developed, *i.e.*, to follow the minimization of ϵ with an initial input of a number of different starting values for the ψ_i terms. Some of these starting values can be highly improbable and far from the finally found minimum. If several checks of this sort produce a unique final set of ψ_i values, it seems highly implausible that an erroneous local minimum has been obtained.

We must now consider the more useful, but much more complex case of minimization with respect to all four parameters. The naive procedure of minimizing the sum of squares with respect to each parameter in a cyclic sequence until an overall minimum is reached results in inefficient minimization. This is because the best value of each parameter is dependent on the values of the other parameters. In order to permit minimization with respect to all parameters simultaneously, ϵ is approximated by an expression of the form shown in eq 13, which is a generalization of eq 10, the simple

$$\epsilon = a + \sum_{i=1}^{p} b_i \ln \psi_i + \sum_{i=1}^{p} \sum_{j=i}^{p} c_{ij} \ln \psi_i \ln \psi_j \qquad (13)$$

quadratic used for the one parameter case. In eq 13 a, b_i, and c_{ij} are a set of constants, and p is the number of parameters, which is 4 in the case of the Adair equation for four binding sites (eq 3). By taking derivatives twice with respect to each parameter, and rearranging the equations, the con-

TABLE II: Application of the Curve-Fitting Procedure to Data[a] for the Binding of NAD to Yeast Glyceraldehyde 3-Phosphate Dehydrogenase at 4°.

Iteration		(1)	(2)	(3)	(4)	ϵ
0	Parameters[b]	1.75×10^5	1.14×10^{10}	3.33×10^{14}	3.64×10^{18}	2.352
	First derivatives[c]	0.0714	0.2559	−0.3895	1.494	
	Matrix of second derivatives[d]	0.729	−0.087	−0.604	−0.417	
		−0.087	0.726	0.314	−0.308	
		−0.604	0.314	0.771	0.084	
		−0.417	−0.308	0.084	0.830	
1	Parameters	1.75×10^4 [f]	1.26×10^9 [e]	1.17×10^{14} [e]	3.64×10^{17} [f]	1.000
	First derivatives	−0.0493	−0.5262	−1.697	0.6886	
	Matrix of second derivatives	0.052	0.102	−0.086	−0.068	
		0.102	−0.210[g]	0.677	−0.092	
		−0.086	0.677	1.787	−0.306	
		−0.068	−0.092	−0.306	0.743	
2	Parameters	2.82×10^4 [h]	3.98×10^9 [i]	1.88×10^{14} [h]	2.29×10^{17} [h]	0.0406
	First derivatives	0.0395	0.0067	−0.1412	0.0927	
	Matrix of second derivatives	0.153	0.072	−0.302	−0.036	
		0.072	0.591	0.339	−0.156	
		−0.302	0.339	1.778	−0.146	
		−0.036	−0.156	−0.146	0.415	
3	Parameters[j,k]	2.28×10^4	3.70×10^9	1.88×10^{14}	1.77×10^{17}	0.0199
4	Parameters[j,k]	1.91×10^4	3.81×10^9	1.90×10^{14}	1.67×10^{17}	0.01893
5	Parameters[j]	1.83×10^4	3.83×10^9	1.89×10^{14}	1.66×10^{17}	1.01890
	First derivatives[l]	0.0001	0.0000	−0.0002	−0.0000	
	Matrix of second derivatives[m]	0.060	0.065	−0.214	−0.019	
		0.065	0.611	0.283	−0.127	
		−0.214	0.283	1.825	−0.060	
		−0.019	−0.127	−0.060	0.268	
	Matrix of variances and covariances	0.1561	−0.0261	0.0224	0.0036	
		−0.0261	0.0096	−0.0045	0.0017	
		0.0224	−0.0045	0.0048	0.0005	
		0.0036	0.0017	0.0005	0.0113	

[a] Data for this run consisted of 18 observed pairs of values for log (X) and N_X, an estimate of −4.64 for $\log(X_{0.5})$, and criteria for termination as follows: (i) every first partial derivative has absolute value less than 0.001, or (ii) no parameter changed by more than 0.1% in the last iteration, or (iii) iteration limit reached (of ten iterations). [b] Values of ψ_i (for eq 3). The starting values were calculated by the computer on the basis of a Michaelis–Menten curve having $\log(X_{0.5}) = -4.64$. [c] That is, $\partial\epsilon/\partial \ln \psi_i$ terms. [d] That is, $\partial^2\epsilon/\partial \ln \psi_i d \ln \psi_i$ terms. The pure second derivatives, $\partial^2\epsilon/\partial(\ln \psi_i)^2$, appear on the diagonal. Each of these four terms must be positive for the model implied by eq 13 to be a useful approximation. Since this is the situation here the following set of linear simultaneous equations may be set up, where c_1, c_2, etc., are the increases required in the natural logs of ψ_1, ψ_2, etc. [e] Corrections applied as calculated from above equations. [f] Corrections calculated for ψ_1 and ψ_4 were −3.73 and −4.38, respectively; *i.e.*, both were greater than the limit of one order of magnitude. So corrections of −2.3 were applied to these parameters. [g] $\partial\epsilon/\partial(\ln \psi_2)$ is negative, so the set of linear simultaneous equations cannot be used. [h] ψ_1, ψ_3, and ψ_4 were corrected using the formula $c_i = -(1/2)(\partial\epsilon/\partial \ln \psi_i)/(\partial^2\epsilon/\partial(\ln \psi_i)^2)$. [i] ψ_2 is not corrected using the formula given in note *h*, because $\partial^2\epsilon/\partial(\ln \psi_2)^2$ is negative. Instead the correction $c_2 = -(1/2) \ln (10) = -1.15$, is used, where the correction has the sign opposite that of the first derivative. In this formula and in the formula given in *h* the scale factor of one-half is included in order to make the correction more conservative. [j] Corrections to the parameters were obtained by solving the set of simultaneous equations analogous to those in note *d*, without complications. [k] The derivatives were calculated for these iterations, but are omitted from the table for simplicity. [l] Every first derivative has an absolute value less than 0.001, so that criterion (i) for termination (footnote *a*) is satisfied. [m] The matrix of second partial derivatives is converted into the matrix of variances and covariances as follows: (i) the matrix is inverted, using a standard computer routine (Wiberg, 1964); (ii) each element is multiplied by twice the residual sum of squares, 2ϵ, and divided by the number of degrees of freedom, which is the number of observations less the number of parameters, *i.e.*, 18 − 4 = 14, in this example. The resulting matrix contains the variances and covariances of the natural logarithms of the parameters. To obtain the variances and covariances of the parameters themselves, multiply each element $\text{cov}(\ln \psi_i, \ln \psi_j)$ by $\psi_i\psi_j$, giving $\text{cov}(\psi_i,\psi_j)$.

$$
\begin{aligned}
0.0714 &+ 0.729c_1 - 0.087c_2 - 0.604c_3 - 0.417c_4 = 0\\
0.2559 &- 0.087c_1 + 0.726c_2 + 0.314c_3 - 0.308c_4 = 0\\
-0.3895 &- 0.604c_1 + 0.314c_2 + 0.771c_3 + 0.084c_4 = 0\\
1.494 &- 0.417c_1 - 0.308c_2 + 0.084c_3 + 0.830c_4 = 0
\end{aligned}
$$

stants a, b_i, and c_{ij} can be eliminated, to yield, instead of a simple iteration formula, a set of simultaneous linear equations in $\ln(\psi_{jk+1}/\psi_{jk})$ of the type shown in eq 14. These

$$\frac{\partial \epsilon}{\partial \ln \psi_i} + \sum_{j=1}^{p} (\psi_{jk+1}/\psi_{jk}) \frac{\partial^2 \epsilon}{\partial \ln \psi_i \partial \ln \psi_j} = 0 \qquad (14)$$

may readily be solved for the improved values of the parameters ψ_{jk+1}, using any standard method for solving simultaneous linear equations. With some modifications, this is the method which we have used. Just as in the one parameter case, the validity of eq 13 is dependent on how close the currect values of the parameters are to the true minimum, and if any pure second partial derivative, $\partial^2\epsilon/\partial(\ln \psi_i)^2$, is zero or negative, eq 13 is sufficiently inaccurate to be rejected as a model. (Note that this consideration does not apply to *mixed* second derivatives, such as $\partial^2\epsilon/\partial \ln \psi_i \partial \ln \psi_2$, which may be either positive or negative without affecting the validity of eq 13.) In the event that any pure second derivative is zero or negative, the parameters are changed simultaneously using eq 11 for parameters for which the second derivative is positive, and eq 12 for parameters for which the second derivative is zero or negative. In order to make the changes more conservative, however, a scale factor of 0.5 is introduced into eq 11, 12, *i.e.*, only half the calculated correction is made. Regardless of the method of computing the corrections, a limit of ± 2.3 is placed on the correction permitted in any parameter, and if any correction is calculated which is outside these limits, it is reduced accordingly.

A computer program based on this method has been written in FORTRAN IV and computations were carried out on a Control Data Corporation 6400 digital computer. Copies of the curve-fitting program are available on request.

A sample calculation is shown in Table II.

Precision of Parameters. In the general nonlinear case there is no accepted method of estimating the precision of parameters. However, the model is only moderately nonlinear when eq 13 gives an accurate measure of the sum of squares in the vicinity of the minimum. In that case the matrix of the variances and covariances of the natural logarithms of the parameters is approximately represented by the matrix $\mathbf{D}^{-1}2\epsilon/\nu$, where $\mathbf{D}$ is the matrix of second partial derivatives, $\partial^2\epsilon/\partial \ln \psi_i \partial \ln \psi_j$, ϵ is the residual sum of squares, and ν is the number of degrees of freedom, *i.e.*, the number of observations − the number of parameters. Multiplication of each element cov($\ln \psi_i$, $\ln \psi_j$) of this matrix by $\psi_i\psi_j$ gives the matrix of variances and covariances of the parameters themselves, cov(ψ_i,ψ_j). The diagonal elements of this matrix are the variances of the parameters, and the square roots of these elements yield the standard errors. A full discussion of the validity of estimates of precision in nonlinear regression has been given by Beale (1960).

Results

The curve-fitting procedure described has been applied to published experimental data and two examples of this will be shown below. In fits to actual data there is no independent measure of the correctness of the results. In order to assess the method more completely therefore we have used simulated data generated by eq 3 where the true values of the parameters are known and the magnitudes of the error in the data points are also known.

TABLE III: Comparison of the Errors in Assigned Values of Parameters with Errors Expected for a Normal Distribution.[a]

Error in Parameter (s = Std Error of Parameter)	Frequency of Occurrence of Stated Error (%)	
	In Test	Expected for a Normal Distribution of Errors
Less than $-3s$	0.75	0.4
$-3s$ to $-2s$	2.00	2.8
$-2s$ to $-s$	14.00	13.0
$-s$ to 0	37.25	33.8
0 to s	31.75	33.8
s to $2s$	12.00	13.0
$2s$ to $3s$	1.75	2.8
Greater than $3s$	0.5	0.4

[a] Theoretical saturation curves were calculated from eq 3 using values of ψ_1, ψ_2, and ψ_3 taken from a random number generator, with ψ_1 and ψ_3 values in the range 0.4–40, and ψ_2 values in the range 0.6–60. ψ_4 was equal to 1.0 in all cases. Errors were introduced into the calculated N_X values by the random number generator with a standard deviation of 0.025. The errors in the parameters were obtained by subtracting the values used to generate the data from the values assigned by the curve-fitting procedure, and these true errors were compared with the calculated standard errors. Since there were 20 points/plot, the expectations of errors in the right-hand column are based on the values of Student's t with 16 (*i.e.*, 20–4) degrees of freedom.

In the most general experiment 100 sets of four parameters each were selected ranging from 0.4 to 40 for values of ψ_1 and ψ_3 and from 0.6 to 60 for ψ_2. Since only three parameters are needed to determine the shape of the saturation curve for the tetrameric case, the fourth parameter, ψ_4, was set equal to 1 without loss of generality. A random number generator was used for the selection of the various ψ_1, ψ_2, and ψ_3 and the values in the selected range were evenly distributed on a logarithmic scale. (If a linear scale had been used, there would have been a strong bias toward negatively cooperative systems which was not desired in this test.) Using the sets of parameters selected in this way 20 substrate concentrations were chosen such that the N_X values corresponding to these substrate concentrations varied between 5 and 95% saturation. Random normal errors were introduced into the N_X values such that the true standard deviation of the observations from the true curve was 0.025. For each set of data then the curve-fitting procedure was used to determine the values of the parameters, and their standard errors, as described in the previous section, and the calculated values were compared to the true values.

From the summary of the results in Table III it may be seen that the standard errors calculated as described agree

TABLE IV: Investigation of the Possibility that the Curve-Fitting Procedure Might Converge to a Local Minimum.[a]

No.	Starting Values				Final Values			
	ψ_1	ψ_2	ψ_3	ψ_4	ψ_1	ψ_2	ψ_3	ψ_4
1	100	10,000	1	0.01	0.28	34.7	191	1.17
2	0.01	1	10,000	100	0.25	34.8	190	1.19
3	100	1	1	100	0.28	34.7	191	1.17
4	0.01	10,000	10,000	0.01	0.25	34.8	190	1.19
5	100	1	10,000	0.01	0.29	34.7	191	1.17
6	0.01	10,000	1	100	0.25	34.8	190	1.19
7	1000	100	100	1	0.28	34.7	191	1.17
8	1	100,000	100	1	0.28	34.7	191	1.17
9	1	100	100,000	1	0.28	34.7	191	1.17
10	1	100	100	1000	0.28	34.7	191	1.17

[a] Starting values are the values given the computer to begin the curve-fitting procedure. Final values are the values obtained by the computer at the termination point. Essentially the same final values are obtained in every case, even though the starting values of the parameters were in error by factors as high as 3000. The conclusion is that in this case at least there is only one combination of parameters which corresponds to a minimum in the sum of squares. The "data" used in this experiment were simulated to resemble the shape of the curve for the 4° data for yeast glyceraldehyde 3-phosphate dehydrogenase but with much larger errors, *i.e.*, a standard deviation of 0.125 from the "true" values, in order to provide the method with a more severe test than the real data offered. Similar experiments were carried out with the real data, with essentially identical results.

very well with the expectation for a normal distribution of errors. This implies that within the range of this test, the method does give a valid measure of the parameters, and of the standard errors. Since the standard deviation of the observations in this experiment was rather low, a second test was carried out in exactly the same way using a standard deviation of 0.05 in the data, which is more typical of the precision actually observed in experiments. For this test the results were very similar to those shown in Table III, except that about 3% of the estimated standard errors were very much less than the actual errors. In all of these cases the true error was more than six times the calculated error, a discrepancy well outside statistical expectations. This discrepancy arises from the constraint requiring that each parameter be positive. If the true minimum is a point at which one or more parameters are negative (which is mathematically possible), then the termination point of the program will be a different point, and thus the calculated standard errors will be incorrect. Fortunately this situation is very easy to recognize in practice, since in all the cases where it was observed, the calculated standard error of the parameter in question was much larger (more than ten times) than the value assigned to the parameter itself. In the usual case, the validity of the standard errors is not affected by this consideration.

Possibility of a Local Minimum. To test the possibility that the curve-fitting procedure might converge to a local minimum which would not correspond to the true values of the parameters, various calculated theoretical equations were tested with various starting values. A typical result is shown in Table IV for a case in which the data were simulated using parameters from an actual case observed in the enzyme yeast glyceraldehyde 3-phosphate dehydrogenase (Cook *et al.*, 1970). Widely different starting values were given the computer which was then asked to seek the correct values for the parameters by the iterative procedure described in the theoretical section. Essentially the same final values were obtained in every case even though the starting values of the parameters were in error by factors as high as 3000. In no case was a local minimum found. Similar results were obtained in other examples. It appears therefore that the curve-fitting procedure has no inherent flaw which would lead to an incorrect minimum, and in any experimental case one can have confidence in the ψ values obtained. Moreover, a check with several widely different starting values can be used to insure the reliability of the best fit values.

If the starting guesses for some or all of the parameters differ from the best fit values by several orders of magnitude, the values for $\partial\epsilon/\partial \ln \psi_i$ calculated by the approximate method described may be exactly zero when the true values are very small or negative numbers. This can arise even if the computer carries many significant figures (for example, 15 on the CDC 6400 computer used in this work). When a value of $\partial\epsilon/\partial \ln \psi_i$ is found to be exactly zero, the sign as well as the magnitude is lost and consequently the computer is unable to determine whether the parameter should be increased or decreased. Since the errors in the starting guesses have to be very large for this to happen, this problem can usually be avoided by ensuring that the starting guesses are reasonable. In practice this has not presented any difficulties if the starting guesses of the ψ_i values are selected on the basis of an assumed Michaelis–Menten curve calculated from the half-saturation value. However, this difficulty should be kept in mind if wildly improbable starting guesses are used as a means of testing for local minima.

Effect of Number of Points on Precision of ψ Values. In practice it is often difficult to obtain accurate binding data for a protein covering most of the range from 0 to 100% saturation, and even when this is possible it is often incon-

venient to make a very large number of measurements to get a well-defined saturation curve. In the literature it is found that very few saturation curves have been published for multisubunit enzymes where there are more than 15 points per curve (the data of Anderson and Weber (1965) for the binding of NADH to lactic dehydrogenase are an exception to this), and many curves have been published with as few as five or six points. As a comparison, it may be pointed out that to fit a four-parameter model to data containing only five observations is analogous to fitting a straight-line plot with three points. It is therefore pertinent to ask how many observations are needed to define all four parameters of the Adair equation accurately, how much of the range need be covered, and what information can be obtained from data which fall short of these two requirements. We have carried out further experiments with simulated data in order to answer these questions.

In order to determine how many observations are needed to define the parameters, a set of experiments was simulated in which the "true" parameters were those of a Michaelis–Menten curve, with $\psi_1 = 4.0$, $\psi_2 = 6.0$, $\psi_3 = 4.0$, $\psi_4 = 1.0$, the number of data points in the various experiments ranged from 5 to 40, and the standard error of an individual determination of N_X was chosen as 0.05. The curve-fitting procedure was used to obtain the best-fit parameters, and the standard errors of the values obtained were calculated as described previously. It is not strictly valid to apply the values of Student's t to a nonlinear model, but this may be done as a tentative approximation, and was used in the present case to determine 95% confidence limits for the assigned parameters, using values of t with $n - 4$ degrees of freedom, where n is the number of observations, and the values of t were obtained from standard tables. The results of this set of experiments are shown in Figure 2, where the plot is drawn through the calculated values for the 95% confidence limits. The actual errors are also shown, and it may be seen that in 15 cases out of 16 the actual error was less than that given by the 95% confidence limit, which is a very satisfactory result. Examination of the curve suggests that 15–20 points are optimal, since the precision decreases sharply if fewer than 15 points are used, and is not greatly improved by using more than 20 points. It must be emphasized that this conclusion was reached with data evenly distributed over the range 5–95% saturation, and that in real experiments a somewhat higher number of observations may be needed to compensate for an uneven distribution. Moreover, the required numbers of points may be somewhat different for proteins which deviate significantly from Michaelis–Menten behavior.

Effect of the Range of N_X Values on Precision of ψ Values. In a practical situation the range of N_X values covered will be important as well as the number of points. The limits at 1% saturation and 99% saturation are difficult to determine experimentally. The middle points near 50% saturation may be determined most accurately but may not provide much discrimination between ψ values.

To determine how much of the range in saturation should be covered, a set of experiments very similar to those described in the section were carried out, but in this case the range of N_X was varied, instead of the number of data points, which was held constant at 15. The results of two typical evaluations are shown in Figure 3 for a protein which follows

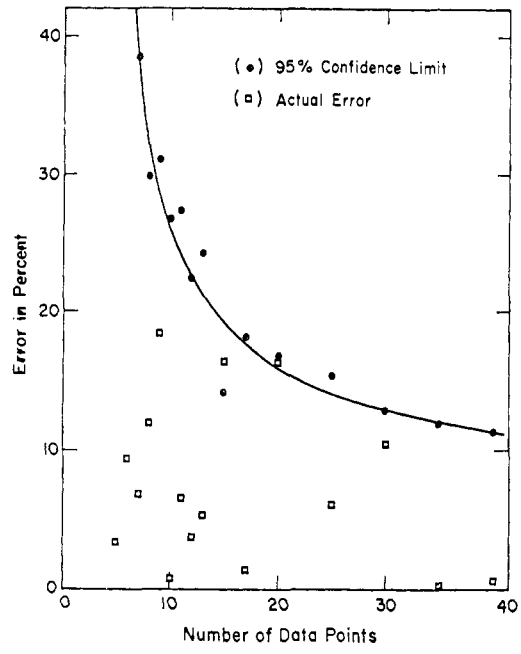

FIGURE 2: Determination of the number of observations needed to define the values of the parameters of the Adair equation to a given degree of precision. The data points were simulated for a Michaelis–Menten-type protein with four binding sites and a standard error in N_X of 0.05. The data points were evenly spaced over the range 5–95% saturation. The solid circles and the curve represent 95% confidence limited calculated from the standard errors with $(n - 4)$ degree of freedom. The open circles represent the actual differences (omitting signs) between the "true" and found values of the parameter. These results were for ψ_4, but the results for the other three parameters were very similar, and in 62 out of 64 cases (97%) the calculated values of the parameters were within the confidence limits.

Michaelis–Menten binding ($\psi_1 = 4.0$, $\psi_2 = 6.0$, $\psi_3 = 4.0$ and $\psi_4 = 1.0$), and in Figure 4 for a protein with positive cooperativity approximating the cooperativity of hemoglobin ($\psi_1 = \psi_2 = \psi_3 = 0.4$ and $\psi_4 = 1.0$). Tests were also made for a negatively cooperative protein ($\psi_1 = 40$, $\psi_2 = 100$, $\psi_3 = 40$, and $\psi_4 = 1.0$), with results which were similar in general appearance to the Michaelis–Menten case and hence are not reproduced here.

Examination of the figures leads to the following conclusions. (1) The accuracy of the individual ψ_i values increases as a greater fraction of the total N_X range (0–4) is covered. (2) The accuracy of the individual ψ_i values will depend on the range of N_X covered if the range is not large enough. Thus in Figure 3 there is little dependence of ψ_1 on the midpoint of the range covered when the N_X values cover more than 50% of the range, but the accuracy varies strongly with the midpoint when only 20% of the range is covered. (3) There is a tendency for a maximum in the error for a given ψ_i parameter in the region where the midpoint of the N_X range equals i. (4) In general the accuracy of the ψ_1, ψ_2, and ψ_3 values is very good for the Michaelis–Menten case when 50% of the range is covered and quite good for ψ_4. The same is true for the cooperative case only the errors are in general larger (20% instead of 5%) and ψ_4 is determined more accurately than ψ_1, ψ_2, and ψ_3. (5) The results emphasize the interrelations of the constants.

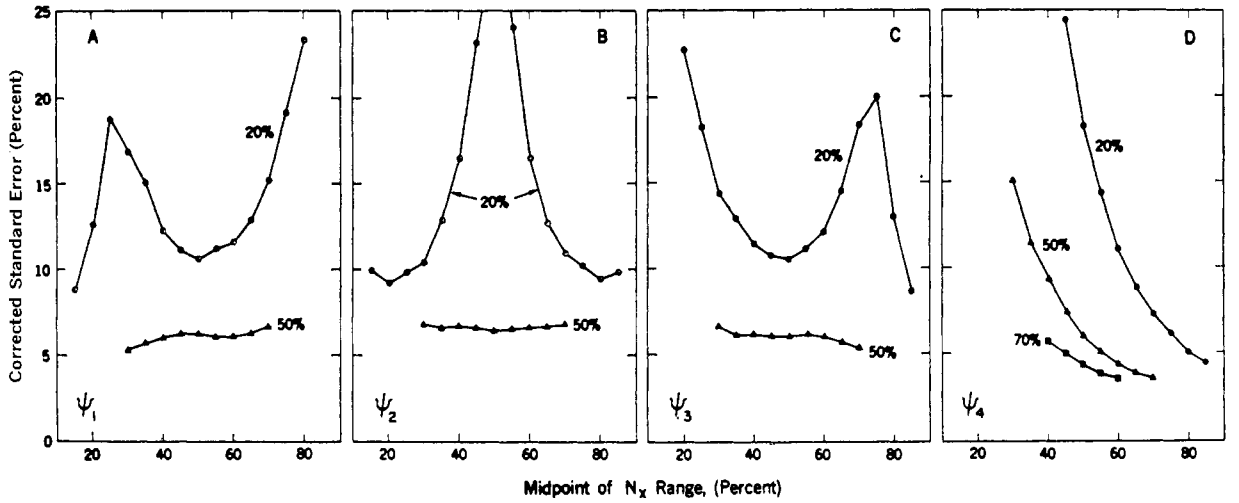

FIGURE 3: Standard errors of the ψ values as a function of the range in N_X covered by the data points for a protein with true parameters $\psi_1 = 4.0$, $\psi_2 = 6.0$, $\psi_3 = 4.0$, and $\psi_4 = 1.0$ (Michaelis–Menten). The standard error of the N_X values used was 0.02, and there were 15 points in each plot. The root-mean-square deviation between the fitted curve and the points varied statistically in the range 0.009–0.025, and this variation was strongly reflected in the calculated standard errors of the parameters, obscuring the trends in these errors. To compensate for this, the standard errors were corrected for a root-mean-square fit to the curve of 0.02 (by multiplying each standard error by 0.02/found root mean square). For each parameter, the corrected standard errors are plotted against the midpoint of the range of N_X represented in the data, and the plots are labeled with the extent of this range. Plots for 30 and 40% of the range were intermediate between those for 20 and 50% shown, and, except in the case of ψ_4, plots for a range of greater than 50% were very similar to the 50% plots.

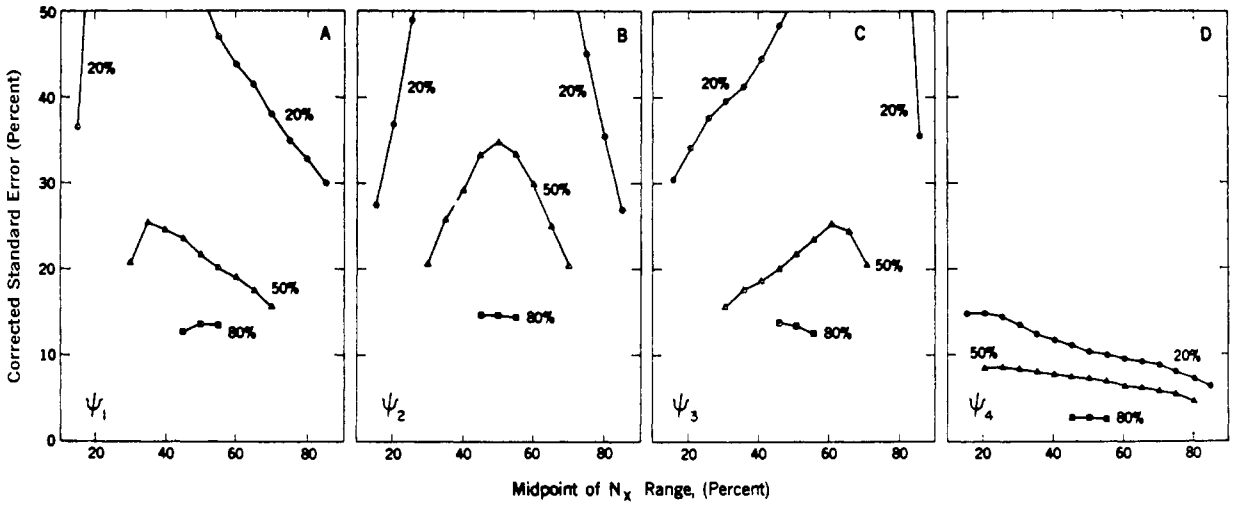

FIGURE 4: Standard errors of the ψ values as a function of the range in N_X covered by the data points for a protein with true parameters $\psi_1 = 0.4$, $\psi_2 = 0.4$, $\psi_3 = 0.4$, and $\psi_4 = 1.0$ (positively cooperative, to about the same degree as hemoglobin). Conditions and corrections as described under Figure 3.

For this reason measurement over a narrow range of N_X will rarely lead to highly accurate ψ values. The gain in accuracy for ψ_2, for example, would then be offset by increased inaccuracy for ψ_3 or ψ_4. As a rule of thumb, therefore a range of at least 50% of the saturation curve and a midpoint for this range at 50% saturation seems desirable for a routine saturation curve, but there is considerable flexibility in the range and the midpoint needed to obtain quite reasonable accuracy. (6) The results with a negatively cooperative system only accentuate the trend on proceeding from positive cooperativity to Michaelis–Menten. The errors in the ψ values are somewhat more dependent on the midpoint of the range covered in a negatively cooperative case. In an extreme case in which three sites are saturated before the fourth begins to be filled, collecting data in the first 75% of the curve would give no information about ψ_4.

While the three cases considered do cover a wide range of possibilities, it must be emphasized that all three of them represent a limited class, in that the approximate equality shown in eq 15 holds, to within a factor of 2.5. We have found that if these three quantities (which together define the shape of the saturation curve completely) are not of the same magnitude, the smallest of them becomes very ill defined, and hence the ψ_i term in the numerator becomes

$$\frac{\psi_1}{\psi_4^{1/4}} \approx \frac{\psi_2}{\psi_4^{1/2}} \approx \frac{\psi_3}{\psi_4^{3/4}} \qquad (15)$$

ill defined. The situation is observed in the real case of yeast glyceraldehyde 3-phosphate dehydrogenase, where $\psi_1/\psi_4^{1/4}$ is much smaller than the other terms, and ψ_1 is accordingly rather ill defined. The conclusions drawn above must, therefore, relate to illustrative cases in a moderate range. As the ψ_i values vary more widely, greater errors are found.

A second, and quite different, conclusion may be drawn from the results for the cooperative case, namely, that regardless of the data, ψ_4 is much better defined than the other parameters. For example, the worst situation shown for ψ_4, with data covering only the 5–25% range, produces a standard error about 14% of the value, which is almost as good as the best cases for any of the other parameters, *e.g.*, the standard error of 12% found for ψ_1 with data covering the range 5–85%. The implication of this is that for cooperative enzymes it is likely to be very difficult to obtain highly accurate unique values of all of the binding parameters, even though the experimental data can be fitted precisely with a theoretical curve (*cf.* Koshland *et al.*, 1966).

Distinction between Models. The curve-fitting procedure will provide a best-fit set of ψ_i values from which the true experimental association constants can be calculated from eq 4. For comparison to theoretical models the intrinsic constants, K_i', are preferable and these can be obtained from the K_i values by the correction for statistical factors as shown in eq 16 where n is the total number of sites. The

$$K_i' = \frac{i}{n - i + 1} K_i \qquad (16)$$

K_i' values can be compared to see if any of the relationships of Table I are satisfied. If so, a *prima facie* case for the theoretical relation implied in the model has been obtained.

Once the best fit has been obtained it is still necessary to test alternative models to determine whether they may also fit the data. In some cases it may not be possible to distinguish between models because they give the same intrinsic constant relationships. Thus, the square and tetrahedral models both give the $K_1'K_4' = K_2'K_3'$ relationship and hence other methods, *e.g.*, $\bar{B}$ *vs.* $\bar{Y}$ curves or the added relations of Table I, are needed to distinguish between them. In other cases there is a clear distinction between the theoretical relationships but the experimental data are not sufficiently accurate to allow a decision between the theories.

For illustrative purposes one may take the hemoglobin data of Rossi-Fanelli *et al.* (1961) and apply the present curve-fitting methods. In each the graphical relationship is illustrated in Figure 5. Clearly the Michaelis–Menten curve does not fit the data at all. The symmetry model with exclusive binding fits the data poorly but the other models (symmetry with nonexclusive binding or ligand-induced model in the tetrahedral or square arrangement) all fit the data equally well. The calculated sum of squares for each model is shown in Table V and it is seen that the models which "fit" all give very low sums of squares. Clearly one cannot choose between these on the basis of the saturation curve alone even though with absolutely accurate data

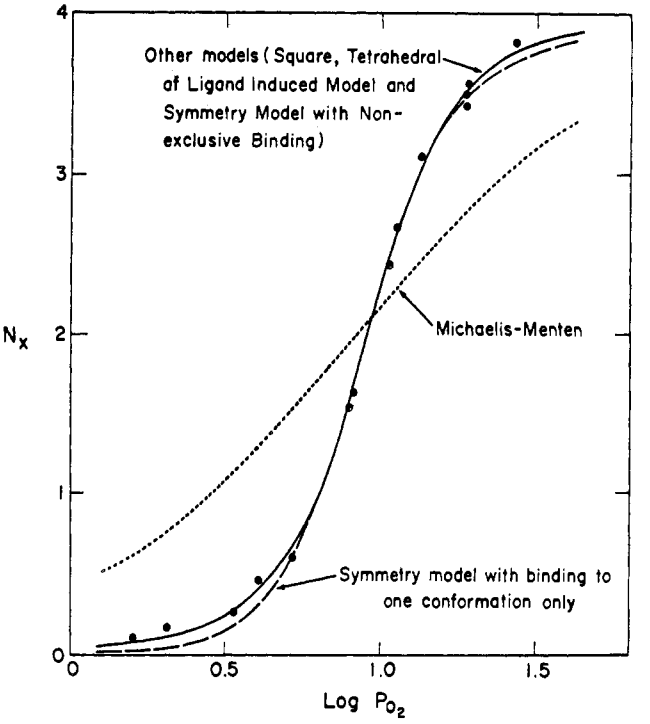

FIGURE 5: Data for the binding of oxygen to hemoglobin at high ionic strength (Rossi-Fanelli *et al.*, 1961) fitted by (a) Michaelis–Menten equation, (b) "square" version of simplest sequential model, (c) symmetry model with binding to one conformation only, (d) symmetry model with binding to both conformations, and (e) unrestricted Adair equation. The curves of b, d, and e were not distinguishable by eye, and therefore are not drawn separately.

one can distinguish between the concerted and sequential models. The data are probably good enough to exclude the symmetry model with exclusive binding. It is, or course, known from additional experiments (Ogawa and McConnell, 1967; Shulman *et al.*, 1969; Haber and Koshland, 1970; Antonini and Brunori, 1969; Guidotti, 1967) that the sequential model is required for hemoglobin but in this case we are merely analyzing what can be learned from the saturation curve alone.

In Table V there is also a comparison of the values selected by the nomogram method of Koshland *et al.* (1966) and the sum of squares in the curves calculated from their parameters is compared to the best-fit computer values. It can be seen that the nomogram procedure gives very good values and certainly could be used for cooperative curves as a first approximation if one does not wish to resort to a computer program.

In some cases, of course, only one relationship may satisfy the saturation equation and all other alternatives are excluded. The data for rabbit muscle glyceraldehyde 3-phosphate dehydrogenase (Conway and Koshland, 1968) and the accompanying paper on the yeast enzyme fall in these categories. An illustrative example with simulated data can be seen in Figure 6 and Table VI. In that example, the illustrated data were calculated for (a) the symmetry model and (b) the square example of the simplest sequential model. The data in each case were then fitted by (a) the symmetry model equation, (b) the square model equation, and (c) the

TABLE V: Values of Parameters Assigned in Fitting Various Theoretical Models to Data from the Binding of Oxygen to Hemoglobin.[a]

Model	Parameter	Values from Nomogram Fit	Values from Best Fit by Computer Procedure	Sum of Squares of Errors
Michaelis–Menten	K_S		0.117	6.07
"Square"	$K_sK_tK_{BB}$	0.109	0.109	0.040
	$K_{AB}/K_{BB}^{1/2}$	0.272	0.269	
Symmetry model with				
(a) Binding to one conformation only	K_R	0.63	0.60	0.084
	L	1197	1,197	
(b) Binding to both conformations	K_R		1.23	0.036
	L		15,800	
	c		0.0097	
General ligand-induced model	ψ_1		0.058	0.035
	ψ_2		0	
	ψ_3		0.000587	
	ψ_4		0.000144	

[a] Date of Rossi-Fanelli *et al.* (1961) for the binding of oxygen to human hemoglobin at 20°, pH 7.0, in 0.25 M phosphate. Parameters in the column labeled "nomogram fit" were calculated from those given by Koshland *et al.* (1966), which were obtained using nomograms described by those authors. Values in the column labeled "best fit" were calculated by the procedure described in this paper.

general equation of this article. The original equation fit the data perfectly in these cases since no random error was introduced. It is seen in Figure 6 that the square model cannot be made to fit the concerted data nor can the concerted model fit the square data. This is reflected in the larger errors of Table VI. This illustrates that a unique distinction between these models is possible if the data are good enough.

It may trouble some readers to note in Table VIb that the fit of eq 3 gives an even lower sum of squares than applying the concerted equation to data generated from that equation itself. This is an effect of the approximations in the curve-fitting procedure which proceeds in "jumps" defined by the first and second derivatives as described above. If the last jump reduces the error below a predetermined value, the program terminates. The error at this point is therefore not the least error possible but will depend fortuitously on how close the last jump came to the true minimum. Since all the errors of 10^{-4} are very small, it is not important whether the values are 3×10^{-4} or 1×10^{-4} or even 10^{-7}. All of these curves fit the data very closely. One may select the termination point at any reasonable level but it is incorrect to conclude one model is better if both fit within this predetermined value.

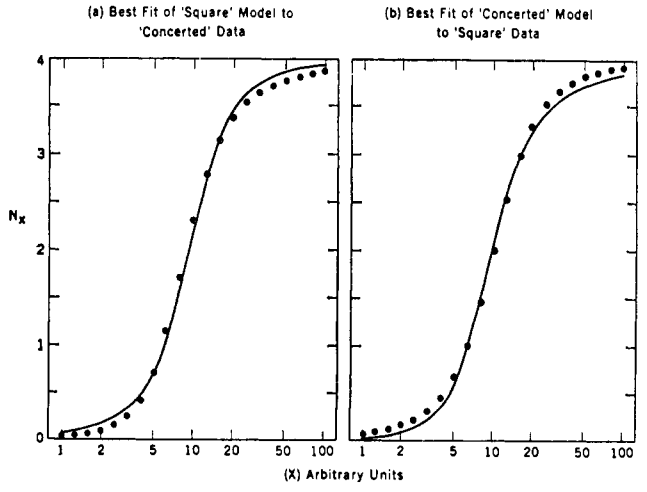

FIGURE 6: Plots showing that with perfect data the "square" and "concerted" models are distinguishable on the basis of saturation curves alone.

Another example is illustrated in Figure 7 in which some data for yeast glyceraldehyde 3-phosphate dehydrogenase are fitted by (a) the general binding eq 3, (b) the symmetry model with exclusive binding, (c) the square model, (d) a model assuming only simple positive cooperativity, and (e) a model assuming only a simple negative cooperativity. All of these equations except (a) imply some restrictions on the values of the intrinsic association constants, K_i', as shown in Table VII, which also shows the values of the sum of squares at the best fit which are obtained with each equation. The best fit curves for the various equations are shown in Figure 7, and it is readily apparent that the general equation fits the data better than any of the other possibilities. The sum of squares for this equation is much the lowest, and the scatter of points about the fitted curve appears to be random, whereas all of the other curves exhibit significant lack of fit. This is an example in which a saturation curve alone can distinguish between models. As in kinetics, isotope techniques, spin labels, or any other scientific tool, "consistent with" does not establish a mechanism. If another model is devised which fits the same data, then further experiments

TABLE VI: Performance of Curve-Fitting Procedure When Applied to Synthetic Data without Experimental Error.

(a) Parameters used to generate the data, compared to parameters found by curve-fitting procedure for the general eq 3.

I "Concerted" Data		II "Square" Data	
Theoretical Parameters Used to Generate Data	Parameters Found by Curve-Fitting Procedure	Theoretical Parameters Used to Generate Data	Parameters Found by Curve-Fitting Procedure
$\psi_1 = 8.74 \times 10^{-3}$	1.39×10^{-2}	$\psi_1 = 6.00 \times 10^{-2}$	6.51×10^{-2}
$\psi_2 = 4.72 \times 10^{-3}$	4.07×10^{-3}	$\psi_2 = 6.48 \times 10^{-3}$	5.97×10^{-3}
$\psi_3 = 1.13 \times 10^{-3}$	1.18×10^{-3}	$\psi_3 = 6.07 \times 10^{-4}$	6.35×10^{-4}
$\psi_4 = 1.02 \times 10^{-4}$	1.03×10^{-4}	$\psi_4 = 1.02 \times 10^{-4}$	1.03×10^{-4}

(b) Residual sums of squares after fitting theoretical equations to data generated in a above.

Equation Used to Fit Data	I "Concerted" Data	II "Square" Data
"Concerted" equation	3.77×10^{-4}	1.15×10^{-1}
"Square" equation	1.03×10^{-1}	4.55×10^{-7}
General eq 3	1.64×10^{-4}	1.37×10^{-4}

may be needed. However, in many cases like this, saturation curves can exclude mechanisms, and may exclude all but one of the reasonable mechanisms. An example of an alternative mechanism which fits the same data quite well is a mixture of two species with different properties.

Discussion

A general procedure for fitting theoretical saturation curves to experimental data has been obtained. This procedure was derived by assuming that the error function can be approximated by a quadratic. The error at any one value of the parameter and the first and second derivatives of the error function allow one to devise a computer program which gives a best fit of the general binding equation to experimental data. The analysis of the results shows the assumption of a quadratic is justified. Moreover, the procedure is not limited to this specific equation and can thus in principle be used for obtaining "best fits" for any theoretical equation of this nonlinear type. Safeguards against local minima can be obtained by using widely different starting values.

When this procedure is applied for illustrative purposes to the binding curve for a tetrameric protein, a number of interesting conclusions and rules of thumb result. Using a normally achievable error of 0.05 in the N_X value it is found that 15–20 points are sufficient to obtain quite accurate values of the four ψ_i values which characterize the binding equation. It is desirable that the N_X values which are measured at various X_t concentrations be distributed as evenly and over as large a range as possible. However, an unreasonable spread, for example, 1–99% saturation is not required. In fact in many cases with reasonable accuracy for the individual points, 20 points ranging from $N_X = 0.5$ to $N_X = 3.5$ will give a very good evaluation of the respective ψ_i values. The accuracy of the individual binding constants calculated from these ψ_i values will, of course, depend on the cooperativity of the particular protein but they are seen to be quite good in both positively and negatively cooperative systems. An error of 20% in a value for ψ_i may at first glance seem quite high based on the usual analysis of single association constants. Yet such an error can only be judged in the light of the need for more accurate ψ_i values. In most cases the accuracy of the ψ_i values even with 20–50% error is more than sufficient to obtain an excellent fit to the data and to obtain K_i values which can distinguish between different theoretical models.

Once the individual intrinsic constants are calculated they can be compared to the intrinsic constants to be

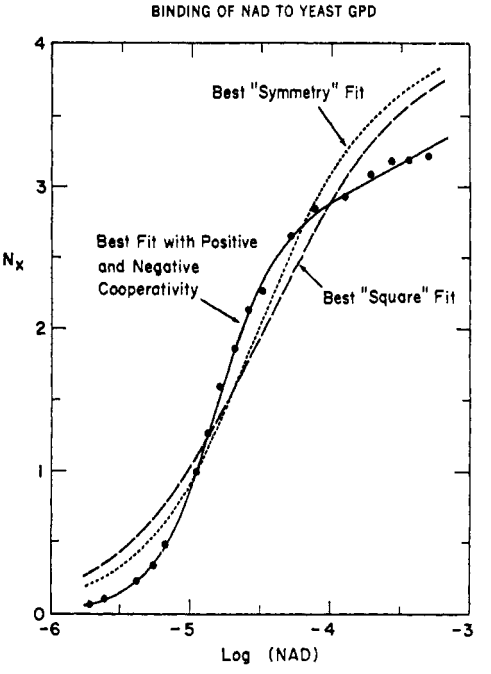

FIGURE 7: Data for the binding of NAD to yeast glyceraldehyde 3-phosphate dehydrogenase at 4° (Cook *et al.*, 1970), fitted by (a) general sequential model, *i.e.*, unrestricted Adair equation, (b) simplest sequential model ("square" or "tetrahedral"), and (c) symmetry model (with or without exclusive binding). Two other restrictions, which are omitted from this figure to avoid obscuring it, are listed in the tabulation of the same results in Table VII.

TABLE VII: Fit of Various Theoretical Models to Data for NAD Binding to Yeast Glyceraldehyde 3-Phosphate Dehydrogenase.[a]

Model	Sum of Squares of Errors $\Sigma(N_{X(obsd)} - N_{X(calcd)})^2$
General ligand-induced model	0.019
Symmetry model:	
(a) Binding to one conformation only	1.47
(b) Binding to both conformations	1.47
Simplest sequential model:	
(a) "Square"	1.39
(b) "Tetrahedral"	1.39
Simple positive cooperativity	1.47
Simple negative cooperativity	0.81
Mixture of two isozymes, each obeying the "square" model[b]	0.10

[a] Date of Cook and Koshland (1970) for the binding of NAD to yeast glyceraldehyde 3-phosphate dehydrogenase at 4°, pH 8.5 in 0.05 M pyrophosphate. Best fits were obtained by the procedure described in this paper. [b] In this model the enzyme was considered to be a mixture of two species, each of which separately obeyed the "square" model, with two independent sets of parameters. The molar ratio of the two species was also allowed to vary.

expected from various theoretical models. For example, a relationship $K_1'K_4' = K_2'K_3'$ suggests a simplest sequential model whereas a relationship $K_1' < K_2' = K_3' = K_4'$ suggests a concerted model with exclusive binding. A complex relationship of the intrinsic constants or negative cooperativity requires a general ligand-induced model. An example of such a determination was given above. Contrary to general lore of the moment therefore, saturation curves can by themselves in many cases distinguish between models.

Even when there are clearly different predictions for the relationships between intrinsic constants for different models, however, it is sometimes difficult to attain sufficient experimental accuracy to establish a distinction. In these cases two or more models may give saturation curves which are experimental indistinguishable even though they are different theoretically. Such was the case for the hemoglobin example. In such cases other diagnostic aids may be needed. In other cases, however, the current analysis suggests that additional points and/or additional accuracy in the points obtained may be sufficient to allow an identification of a model with the saturation curve data only.

An advantageous feature of this procedure is that it carries a continual analysis of the goodness of the fit. It may not be desirable in all cases to record the error and the first and second derivatives of the error at intermediate stages in the fitting. However, this is possible in any case in which the course of the analysis needs to be followed. Moreover, it allows an objective analysis of the final answer. Finally the procedure can be applied to any type of equation, though in some cases it may be desirable to rewrite the equation (*cf.* the rewriting of eq 1 as eq 3), in order to reduce correlation between the parameters. Thus, the best fit of the general binding equation can be compared to the best fit of any detailed model in order to determine which theory most closely represents the data. Even when one does not wish to analyze subunit interactions or the details of protein design, it may be desirable to obtain a quantitative expression for the influence of ligand concentration on the binding or velocity curve of an enzyme, repressor, receptor, binding protein, etc. The empirical curve-fitting procedures outlined here can therefore be used for any curve for which the binding equation gives a good representation of the data.

References

Adair, G. S. (1925), *J. Biol. Chem. 63*, 529.

Anderson, S. R., and Weber, G. (1965), *Biochemistry 4*, 1948.

Antonini, E., and Brunori, M. (1969), *J. Biol. Chem. 244*, 3909.

Atkinson, D. E. (1966), *Annu. Rev. Biochem. 35*, 85.

Beale, E. M. L. (1960), *J. Roy. Stat. Soc. B 22*, 41.

Conway, A., and Koshland, D. E., Jr. (1968), *Biochemistry 7*, 4011.

Cook, R. A., and Koshland, D. E., Jr. (1970), *Biochemistry 9*, 3337.

Fronaeus, S. (1950), *Acta Chem. Scand. 4*, 72.

Guidotti, G. (1967), *J. Biol. Chem. 242*, 3673.

Haber, J. E., and Koshland, D. E., Jr. (1970), *Biochim. Biophys. Acta* (in press).

Koshland, D. E., Jr. (1970), Enzymes, 3rd ed, New York, N. Y., Academic (in press).

Koshland, D. E., Jr., Nemethy, G., and Filmer, D. (1966), *Biochemistry 5*, 365.

Leden, I. (1941), *Z. Phys. Chem. (Leipzig) 188*, 160.

Monod, J., Wyman, J., and Changeux, J.-P. (1965), *J. Mol. Biol. 12*, 88.

Ogawa, S., and McConnell, H. (1967), *Proc. Nat. Acad. Sci. U. S. 58*, 19.

Rossi-Fanelli, A., Antonini, E., and Caputo, A. (1961), *J. Biol. Chem. 236*, 397.

Schroder, K. H. (1966), *Acta. Chem. Scand. 20*, 1401.

Shulman, R. G., Ogawa, S., Wuthrigh, K., Yamane, T., Peisach, J., and Blumberg, W. E. (1969), *Science 165*, 251.

Swann, W. H. (1969), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 2 (Suppl.)*, S39.

Wiberg, K. B. (1964), Physical Organic Chemistry, New York, N. Y., Wiley, p 541.