Biochimica et Biophysica Acta, 614 (1980) 613—624

BBA 69060

THE EVALUATION OF NON-HYPERBOLIC LIGAND BINDING AND SUBSTRATE SATURATION DATA USING VARIABLE HILL COEFFICIENTS

INGO KNACK and KLAUS-HEINRICH RÖHM

Institut für Physiologische Chemie der Universität, Auf den Lahnbergen, D-3550 Marburg/Lahn (F.R.G.)

(Received December 7th, 1979)

Key words: Ligand binding; Curve fitting; Hill coefficient; Substrate saturation

Summary

A fitting method for saturation curves is described that may be applied to any kind of ligand binding process, regardless of whether positive, negative or mixed cooperativity is involved. The procedure yields 'best values' of maximal saturation (or maximal velocity with enzymes); in addition, an empirical function is calculated that represents the concentration dependence of the Hill slope (i.e. the slope of the corresponding Hill plot).

Since the method does not require the maximal saturation to be known in advance, it can be successfully applied to systems which, due to experimental limitations, may not be saturated with ligand. It is shown that in these cases values of maximal saturation are obtained that are considerably more reliable than those estimated by customary methods. Moreover, it is demonstrated that the concentration dependence of the Hill slope, the other main result of the proposed procedure, may offer useful evidence as to the molecular events leading to non-hyperbolicity and thus provide a rational basis for selecting more specific models to describe the properties of the system studied.

Introduction

In many cases the simple Law of Mass Action will suffice to describe adequately the interactions of a protein or another biopolymer with ligands. When applied to a protein containing n identical and independent binding sites for a ligand L, it states that the degree of saturation, S (as measured by some quantity linearly related to the number of ligand molecules bound), is given by

$$S = \frac{S_{max}[L]}{K_d + [L]} \quad (1)$$

where S_{max} is the maximal saturation attained at infinite ligand concentration, and K_d is a thermodynamic dissociation constant.

With enzymes, the overall velocity in the steady state is usually proportional to the degree of saturation of the active site(s) with substrate, L_s, and thus

$$v = k \cdot [EL_s] = \frac{V \cdot [L_s]}{K_m + [L_s]} \tag{2}$$

The physical meaning of K_m, the so-called Michaelis constant, depends on the actual mechanism of catalysis; usually it differs from the substrate dissociation constant K_s.

Whenever the binding sites in the protein are intrinsically different or interact with each other, the saturation, S, is no longer hyperbolically dependent on [L]. For the analysis of such binding or velocity curves several methods of differing sophistication are in use. In favourable cases graphical procedures may furnish satisfactory results.

A Hill plot [1] is constructed by plotting log $\{S/(S_{max} - S)\}$ vs. log [L]. The graph is based on a linearization of the Hill equation

$$S = \frac{S_{max} \cdot [L]^{n_H}}{K + [L]^{n_H}} \tag{3}$$

Only if Eqn. 3 is strictly valid (i.e. in the improbable case of infinite cooperativity) the Hill plot will be a straight line, then the Hill coefficient n_H equals the number of binding sites. With real systems, n_H hardly ever reaches this maximal value and, in addition, is not a constant but depends on [L]. In consequence, Hill plots are usually curved (see Fig. 2b). It was shown [2,3] that the shape of such a plot reflects the underlying mechanism of ligand binding; the Hill slope, therefore, is useful evidence as to the kind and strength of cooperative interactions.

It is a common situation with ligand binding studies that the system cannot be completely saturated. In this case reliable estimates of S_{max} are unavailable, which excludes the application of Hill plots. To overcome such a problem Kurganov and his coworkers [4] and Endrenyi et al. [5] have proposed evaluation methods that, in principle, allow the estimation of Hill slopes when S_{max} is not known. However, both procedures are based on the validity of Eqn. 3 and thus on somewhat unrealistic assumptions. Moreover, cases of negative cooperativity may not be treated in this way. Finally, the application of both methods to actual data is rather cumbersome, since for the calculation of each point of such graphs equally spaced pairs or triplets of observations are required.

A Scatchard plot [6] is obtained if $S/[L]$ is plotted vs. S. It will be linear whenever Eqn. 1 holds, and curved in all other cases. S_{max} is not required for constructing Scatchard plots but may be estimated by extrapolation to $S/[L] = 0$. However, as illustrated by Fig. 2a, with strongly cooperative systems reliable estimates of S_{max} are not obtained unless the experimental readings closely approach maximal saturation.

The shape of a Scatchard plot may be of similar diagnostic value to that of the Hill plot. The kind of information that can be drawn from such a graph has been extensively discussed [3,7].

An alternative, although more demanding, approach to the analysis of non-hyperbolic saturation data is the use of curve-fitting procedures. The numerical calculations involved are extensive and thus a digital computer for data processing is indispensable.

A linear regression on the Hill plot would be mathematically very simple, however, such an approach suffers from the same uncertainty with respect to S_{max} as a purely graphical evaluation and is not expected to yield more reliable results. The direct fit of the Hill Eqn. 3 was proposed by several authors [8—11]. However, as mentioned above, Eqn. 3 is rarely valid for real systems. Therefore, as shown in more detail below, the outcome of such a fit when applied to strongly non-hyperbolic curves may be unsatisfactory or meaningless.

A fit of the general Adair equation (see Eqn. 5) to saturation data imposes much less bias. This approach has been succesfully applied to the binding of ligands to protein tetramers [12]. It is doubtful, however, whether meaningful results can be obtained when the number of binding sites exceeds 6—8 since, as a rule, the probability that a well defined optimal fit is achieved rapidly decreases as the number of parameters to be independently adjusted is increased.

The method proposed in this paper was designed to allow the analysis of incomplete saturation curves where a good estimate of S_{max} (and thus of the number of binding sites) is unavailable. As shown below, it will yield reliable values of S_{max} with systems exhibiting any type of cooperativity. Moreover, it provides data that describe the concentration dependence of the Hill slope. As already mentioned, this function may serve as a useful guide to the interpretation of cooperative phenomena in mechanistic terms.

The method

In brief, our method is based on a least-squares fit of the generalized Hill Eqn. 4 to saturation curves of S vs. [L].

$$S = \frac{S_{max} \cdot [L]^{n(L)}}{K + [L]^{n(L)}} \tag{4}$$

In this equation the Hill coefficient is no longer treated as a constant but allowed to vary with [L]. In the course of the calculations $n(L)$ is represented by a polynomial of the 3rd or 4th order *. The best fit of Eqn. 4 to the experimental points is obtained by an iterative search proceeding as follows (a more detailed description of the mathematical background is given in the Appendix).

(a) Choice of starting values of S_{max}

Initially the range has to be defined where the 'best value' of S_{max} is supposed to be found. This interval (AB in Fig. 1) is then subdivided into three equal parts. The sums of squares, SY (measuring the sum of residuals and thus

* As shown in the Appendix, $n(L)$ is not identical with the Hill slope $n_H(L)$; however, the latter is easily calculated from $n(L)$.

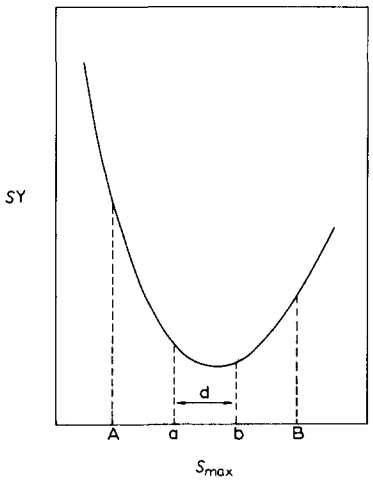


Fig. 1. Schematic representation of the method used to derive provisional estimates of S_{max} (see text).

the goodness of fit), are calculated as described below for values of S_{max} corresponding to points *a* and *b*. The interval with the higher *SY* at its inner border (*Aa* in Fig. 1) is cancelled, the remaining range (*aB*) is again divided into equal thirds, and the procedure is repeated until interval *d* becomes lower than a previously defined limit.

(b) Calculation of n(L)

Once a provisional value of S_{max} is defined the corresponding Hill plot may be constructed. Its slope as a function of [L] is calculated by numerical differentiation as described in the Appendix; then a polynomial that approximates $n_H(L)$ is determined by standard regression methods.

(c) Determination of K and improvement of S_{max}

When $n(L)$ is defined, the parameter K and an improved estimate of S_{max} are obtained by nonlinear regression.

(d) Calculation of SY

With the resulting set of parameters a theoretical curve is generated according to Eqn. 4. The sum of the squares, *SY*, is obtained by comparing this curve with the set of experimental readings

$$SY = \sum_{i=1}^{N} (S_i(\text{calcd.}) - S_i(\text{exp.}))^2$$

Finally, a new range of S_{max} is defined as described above and the minimization proceeds with a next cycle.

A computer program carrying out the calculations outlined above was written in FORTRAN. It was run on a CGK TR 440 system as the Rechenzentrum der Universität Marburg. Copies of the program are available from the authors on request.

Results

The performance of our method was tested by applying it to a variety of experimental and simulated data. The results obtained using data of Winslow et al. [13] on the saturation of whole human blood with oxygen will be discussed in some detail.

The saturation curve of S vs. [L] is shown in Fig. 4a, corresponding Hill and Scatchard plots are depicted in Fig. 2. In each case, the fit performed according to the method described here is represented by a solid line. The dotted line included in Fig. 2b corresponds to the best fit of Eqn. 3, i.e. of the Hill equation with constant Hill coefficient. Clearly, our procedure leads to a considerably improved fit, especially at low ligand concentrations. This impression is quantitatively supported by Table I. The concentration dependence of the Hill slope $n_H(L)$ is shown in Fig. 4b. In close agreement with the known properties of hemoglobin, a maximal Hill slope of 2.67 was found at about 60% saturation.

The efficiency of the method outlined here in the analysis of incomplete saturation curves was examined by fitting partial data sets, that were created from the original oxygen saturation data by increasingly omitting points from the right-hand part of the curve. The sets B—E obtained in this way (see Fig. 2) were also fitted using Eqn. 3; the results are summarized in Table I and in Fig. 3. It was found that the method described above yielded S_{max} with reasonable precision even when the curve was cut off above 30% saturation. In

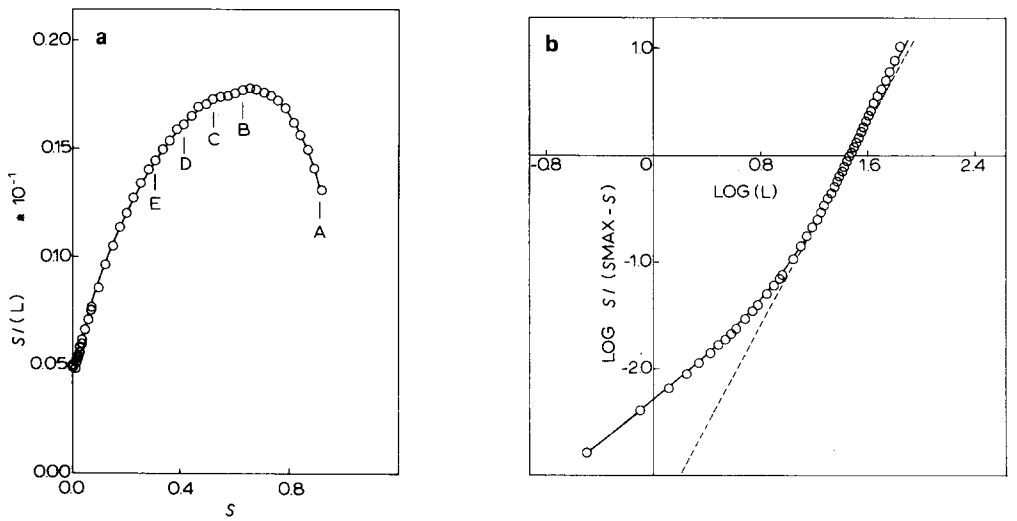


Fig. 2. Comparison of results obtained by curve fitting with customary graphical procedures. The data shown (○) are for the oxygen saturation of whole human blood (taken from Ref. 13, data set RW 2). S is the saturation of the system (expected S_{max} = 1.0); ligand concentrations [L], in this case, are partial pressures of O_2 (mmHg). In each plate the solid line represents the result of a fit performed as described in this paper; the dotted line in Fig. 2b is to illustrate the outcome of fitting the Hill equation with the Hill coefficient treated as a constant. Fig. 2a shows a Scatchard plot of the data. Capital letters indicate where the original data set (A) was cut off to generate the incomplete sets B—E (cf. Table I). In Fig. 2b the corresponding Hill plot is depicted. A more familiar representation of the same data may be found in Fig. 4a.

TABLE I

Comparison of fitting methods applied to incomplete saturation curves. Data set A contains the oxygen saturation data from Ref. 13, also shown in Figs. 2 and 4a, sets B—E were generated from A by successively omitting points from the right-hand part of the curve. The residual data were evaluated by fitting (a) the Hill equation with variable Hill coefficient (Eqn. 4) and (b) the Hill equation with constant Hill coefficient (Eqn. 3). The starting interval for S_{max}, introduced in the calculations is given for each set of data. The resulting parameters and the residual sums of squares are compared. n.m., no minimum.

Data set	Data pairs	Saturation	Starting interval	Resulting parameters after fitting the Hill equation with				
				variable n_H			constant n_H	
				S_{max}	K	$SY \cdot 10^3$	S_{max}	$SY \cdot 10^3$
A	49	0—0.91	0.95—3.0	1.01	195.9	2.9	1.06	8.7
B	38	0—0.62	0.60—3.0	0.95	198.1	2.0	1.48	5.4
C	34	0—0.51	0.50—3.0	0.93	192.0	1.6	1.49	5.5
D	30	0—0.41	0.40—3.0	0.97	194.2	1.1	n.m.	4—6
E	26	0—0.30	0.30—3.0	1.09	217.5	1.0	n.m.	4—5

contrast, fits using Eqn. 3 (performed as described in Ref. 11) were already poor when the upper third of the curve was missing, and no definite minima of SY were found when the residual curve did not exceed half-saturation.

The goodness of fit achieved with data sets A—E may be judged by the data given in Fig. 3. Distinct minima of SY as a function of S_{max} were found in all cases. In Fig. 3b the residuals are plotted vs. ligand concentrations. With set A there is a certain non-random fluctuation of the residuals, while with sets B—E this effect is much less conspicuous, if present at all. Finally, by inspection of Fig. 2 one can ascertain that the application of graphical methods to sets B—E would have been of little advantage. For instance, an attempt to estimate S_{max} by extrapolation of Scatchard plots would be a matter of mere intuition.

Diagnostic value of $n_H(L)$

In order to illustrate the kind of information that can be gathered from the concentration dependence of Hill slopes we applied our fitting method to a series of data sets calculated from the 4th degree Adair equation

$$\frac{S}{S_{max}} = \frac{K_1[L] + 2K_1K_2[L]^2 + 3K_1K_2K_3[L]^3 + 4K_1K_2K_3K_4[L]^4}{4(1 + K_1[L] + K_1K_2[L]^2 + K_1K_2K_3[L]^3 + K_1K_2K_3K_4[L]^4)} \quad (5)$$

An adequate choice of K_1—K_4 yields curves that represent the behaviour of various types of cooperative system encountered experimentally. According to Wong et al. [14] the kind and strength of cooperative interactions among binding sites may be quantitatively characterized by the 'cooperativity coefficient', γ. When for two successive binding steps $\gamma = 0$, cooperative interactions are not involved and the Hill coefficient, n_H, equals unity. Positive cooperativity is indicated by $\gamma > 0$ and $n_H > 1$, while negatively cooperative processes are characterized by $\gamma < 0$ and $n_H < 1$, respectively.

Some examples are shown in Fig. 4. Again the solid lines represent the result of fits performed as suggested here; in addition the concentration dependences of n_H are shown. The oxygen saturation curve of blood (see above) is brought about by two steps with substantial positive cooperativity followed by an event with slightly negative γ [13]. This is reflected by the shape of $n_H(L)$ which rises

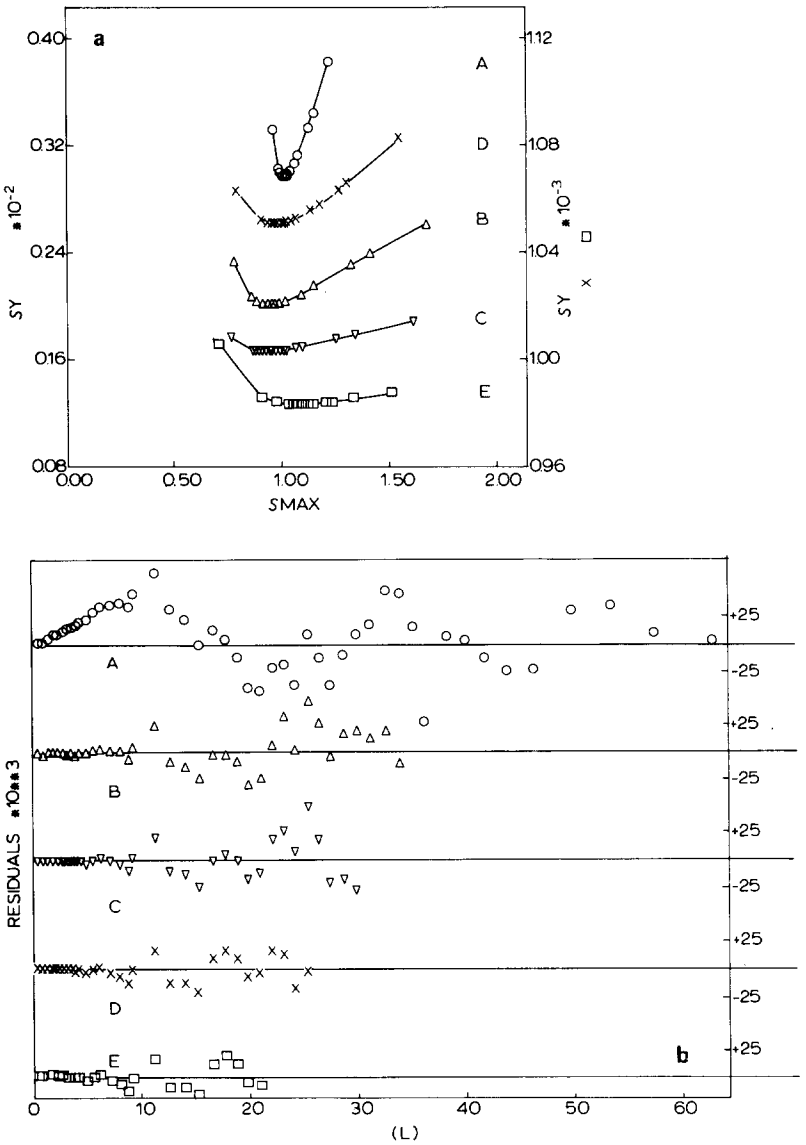
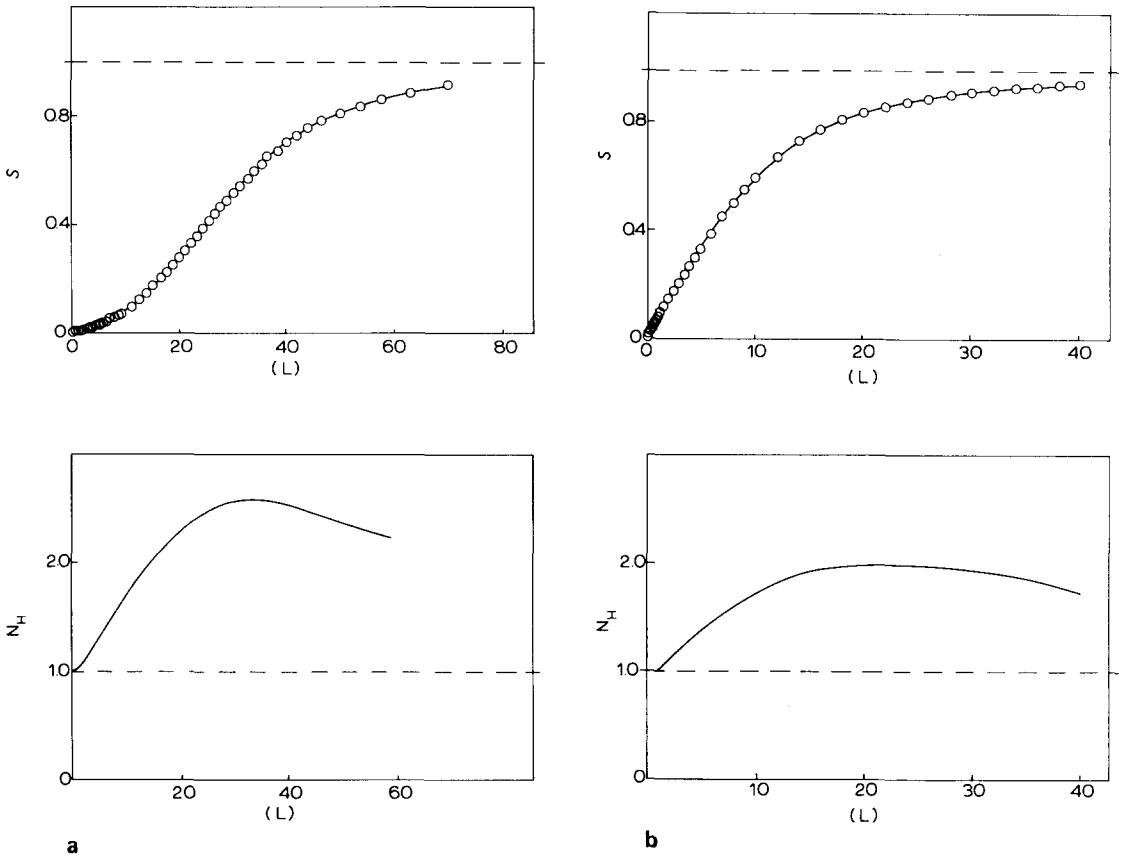


Fig. 3. Performance of the method on data sets A—E. Fig. 3a illustrates the actual course taken by the fitting procedure as applied to data sets A—E (cf. Table I). The sums of squares, SY, computed for each provisional value of S_{max} during the iteration process are shown. Note that well-defined minima of SY were attained in all cases. In Fig. 3b the residual errors are plotted as a function of ligand concentration. The solid lines correspond to zero deviation in each case.

above 2.5 and decreases again at high [L].

If two non-cooperative steps are followed by a positively cooperative one (see Fig. 4b), n_H(L) rises more slowly, reaching its maximal value (about 2.0) not before 80% of the sites are occupied with ligand. Fig. 4c shows a simulated case of strongly negative cooperativity. At first sight, one is tempted to interpret the graph as hyperbola with S_{max} around $S = 0.5$. However, the



decrease of n_H with increasing [L] unequivocally reveals the negative interactions are involved.

The last curve (Fig. 4d) would result from negative cooperativity between the sites occupied first, followed by two positively cooperative steps. Processes

TABLE II

Cooperativity parameters and results of fits to saturation curves 4a—d. Adair constants for data set 4a (same data as set A in Table I) were taken from Ref. 13. Sets 4b—4d were calculated from the 4th degree Adair equation using the constants K_1—K_4 indicated in each case. The corresponding cooperativity coefficients (see text) are also given. The results of fitting the Hill equation with variable Hill coefficient are shown by the final value of S_{max} (expected S_{max} was 1.0 in each case) and the residual sum of squares, SY.

Data set	Adair constants				Cooperativity coefficients			Starting interval	S_{max}	$SY \cdot 10^3$
	K_1	K_2	K_3	K_4	γ_1	γ_2	γ_3			
4a	0.004	0.043	0.262	0.039	1.45	1.14	—0.4	0.95—3.0	1.01	2.9
4b	0.326	0.122	0.054	0.203	0	0	1.0	0.5 —1.5	0.98	2.0
4c	0.326	0.122	$5.4 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	0	—1.0	—1.0	0.5 —3.0	1.04	0.05
4d	0.326	0.012	0.054	0.204	—1.0	1.0	1.0	0.5 —1.5	0.99	0.36

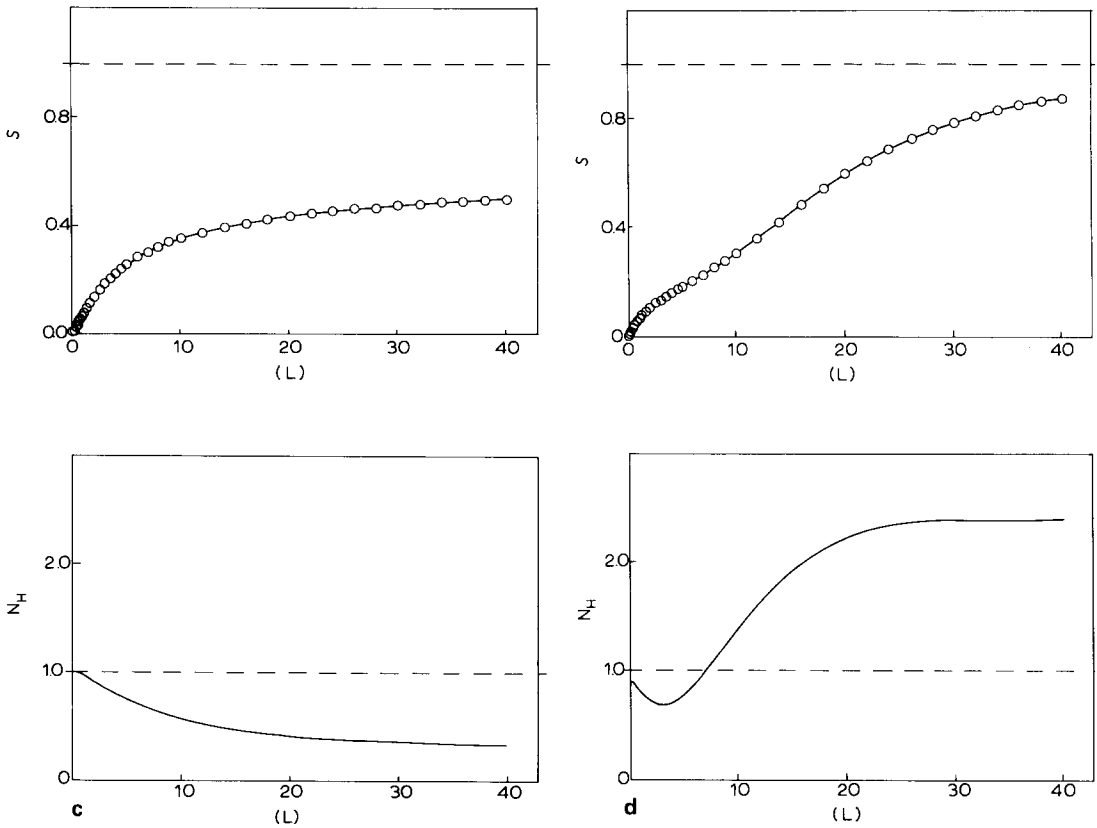


Fig. 4. Concentration dependence of the Hill slope for some types of cooperative behaviour. Fig. 4a contains the oxygen saturation data already shown in Fig. 2; the other data sets were simulated (cf. Table II).

of this type have actually been observed with oligomeric enzymes; they may be identified by an inflection in the corresponding saturation curve. The Hill slope clearly points to the underlying process being lower than unity at low [L] than steeply increasing to values above 2.

Discussion

The evaluation method described in the present paper was devised for the analysis of incomplete saturation curves which strongly deviate from hyperbolicity. In fact, none of the methods available at present yields satisfactory results in cases of strong cooperativity, unless S_{max} is already known or the experimental readings extent to at least 80—90% of this value. This is rarely attainable whenever systems with marked negative cooperativity are studied, and frequently impossible in consequence of experimental difficulties. Moreover, it may occur that one saturation process is superimposed by another, so that the curve can be evaluated only in part. In these instances our procedure yields reasonable values of S_{max} which are required to establish the total number of sites. In addition, the concentration dependence of the Hill slopes

becomes available, a piece of evidence that may lead to more detailed ideas as to the molecular basis of the observed non-hyperbolic behaviour. Of course, the method proposed here cannot replace the elaboration of more specific models which may follow in a later stage of the investigation. On the other hand, the approach adopted results in a wide applicability of the method to virtually any set of binding or kinetic data. If the Law of Mass Action is obeyed, the Hill slope will be unity at any ligand concentration. Other results should prompt further investigations of the kind which will be suggested by the observed shape of $n_H(L)$.

Since $n(L)$ is represented by a polynomial, one could argue that Eqn. 4 contains a total of at least five independent parameters; this would render the fitting of the equation rather problematic. However, in view of the algorithm used during iteration, the coefficients of $n(L)$ must not be regarded as truly independent variables. As shown in more detail in the Appendix, the coefficients are not individually varied but computed as a set by linear regression on a previously defined curve. In fact, numerous calculations with simulated data showed that local minima different from the correct one were very rarely selected, provided the standard deviation of the data did not substantially exceed 0.05.

On the other hand, it has to be born in mind that a polynomial is not ideally suited as a model equation for Hill slopes over the entire range of ligand concentrations. Therefore, our method yields a good approximation of the Hill slope only in the range of [L] where experimental points are available. Beyond this range considerable deviations may occur.

As is the case with all curve-fitting procedures, the reliability and significance of the parameters estimated as described in this paper depend crucially on the number and quality of the experimental readings. Since our method uses non-linear regression, a simple and straightforward measure of the precision of the resulting constants does not exist. However, procedures are available that allow the estimation of standard errors of non-linear parameters [15]. The question of statistical weighting of the data was not systematically investigated in the present study. If regarded necessary, weights may be introduced in the program.

Appendix

(a) Calculation of $n_H(L)$

From Eqn. 4, see text, one gets

$$\ln \frac{S}{S_{max} - S} = n(L) \cdot \ln[L] - \ln K \tag{6}$$

Therefore the Hill slope $n_H(L)$ is given by

$$\frac{d(\ln S/S_{max} - S)}{d(\ln[L])} = n_H(L) = n'(L) \cdot [L] \cdot \ln[L] + n(L) \tag{7}$$

$n(L)$ is represented by the polynomial

$$n(L) = \sum_{k=0}^{m} a_k \cdot [L]^k \tag{8}$$

and thus

$$n'(\mathrm{L}) = \sum_{k=1}^{m} k \cdot a_k [\mathrm{L}]^{k-1} \tag{9}$$

From Eqns. 7—9 it follows that

$$n_{\mathrm{H}}(\mathrm{L}) = a_0 + \sum_{k=1}^{m} a_k([\mathrm{L}]^k + k[\mathrm{L}]^k \cdot \ln[\mathrm{L}]) = a_0 + \sum_{k=1}^{m} a_k \cdot u_k \tag{10}$$

For a given value of [L], u_k is known; $n_{\mathrm{H}}(\mathrm{L})$ is obtained as follows: From the experimental data a continuous curve is determined by 'spline' interpolation. The resulting 'spline' function serves to compute data pairs $(\ln\{S/(S_{\max} - S)\}, \ln[\mathrm{L}])$ for each [L] using the presently available value of $S_{\max}$. Subsequently, these data are fitted by a polynomial, the first derivative of which yields $n_{\mathrm{H}}(\mathrm{L})$. Finally, the coefficients a_k are determined by the method of least-squares. The corresponding sum of squares, SQ, is given by

$$SQ = \sum_{i=1}^{\substack{Ua \\ N}} \left(a_0 + \sum_{k=1}^{m} a_k \cdot u_{k_i} - \left\{ n_{\mathrm{H}}(\mathrm{L}) \right\}_i \right)^2 \tag{11}$$

With the condition $\partial SQ/\partial a_k = 0$ one gets a set of linear equations that may be solved by standard procedures to yield the coefficients of Eqn. 8.

The smoothing of the experimentally determined saturation curve used in this step, was introduced to avoid numerical problems in the calculation of $n(\mathrm{L})$ which may arise if the scatter of the data is considerable. 'Spline' interpolation is not used in other steps; thus, even if an undue distortion of the saturation curve should occur here (which may be avoided by checking the resulting 'spline' fit) this would merely result in a small modification of $n(\mathrm{L})$, but hardly exert influence on the final value of $S_{\max}$.

(b) Determination of K

With the substitution $T = [\mathrm{L}]^{n(\mathrm{L})}$ Eqn. 4 becomes

$$S = S_{\max} \cdot T/(K + T) \tag{12}$$

K and $S_{\max}$ are computed according to the method of least-squares, i.e. by minimizing the sum of squares

$$SR = \sum_{i=1}^{N} (S(S_{\max}, K, T_i) - S_i)^2 \tag{13}$$

For $S_{\max}$ one obtains with the condition $\partial SR/\partial S_{\max} = 0$

$$S_{\max} = \frac{\sum_{i=1}^{N} S_i T_i/(K + T_i)}{\sum_{i=1}^{N} T_i/(K + T_i)^2} \tag{14}$$

The calculation of K is less straightforward. We use the Newtonian approximation

$$S(K) = S(K_0) + \frac{\partial S}{\partial K} \cdot \Delta K \quad \text{with} \quad K = K_0 + \Delta K \tag{15}$$

An estimate of K_0 is available from Eqn. 12. ΔK is calculated from Eqns. 13 and 15 as follows

$$\Delta K = \frac{\sum_{i=1}^{N} T_i^2/(K_0 + T_i)^3 - 1/S_{max} \sum_{i=1}^{N} S_i \cdot T_i/(K_0 + T_i)^2}{\sum_{i=1}^{N} T_i^2/(K_0 + T_i)^4} \tag{16}$$

In this way an improved value of K is obtained from Eqns. 15 and 16. The adjustment of K is repeated in a series of iterations until SR reaches a minimum.

References

1 Wyman, J.R. (1964) Adv. Protein Chem. 19, 233—286
2 Cornish-Bowden, A. and Koshland, D.E. (1975) J. Mol. Biol. 95, 201—212
3 Dahlquist, F.W. (1977) Methods Enzymol. 48, 270—299
4 Kurganov, B.I., Kagan, Z.S., Dorozhko, A.I. and Yakolev, V.A. (1974) J. Theor. Biol. 47, 1—41
5 Endrenyi, L., Fajszi, C. and Kwong, F.H.F. (1975) Eur. J. Biochem. 51, 317—328
6 Scatchard, G. (1949) Ann. N.Y. Acad. Sci. 51, 660—672
7 Schwarz, G. (1976) Biophys. Struct. Mech. 2, 1—12
8 Wieker, H.J., Johannes, K.J. and Hess, B. (1970) FEBS Lett. 8, 178—185
9 Atkins, G.L. (1973) Eur. J. Biochem. 33, 175—180
10 Nimmo, I.A. and Bauermeister, A. (1977) Anal. Biochem. 82, 468—472
11 Herrmann, V., Knack, I. and Röhm, K.-H. (1978) Hoppe-Seyler's Z. Physiol. Chem. 359, 47—61
12 Cornish-Bowden, A. and Koshland, D.E. (1970) Biochemistry 9, 3325—3336
13 Winslow, R.M., Swenberg, M.L., Berger, R.L., Shrager, R.I., Luzzana, M., Samaja, M. and Rossi-Bernardi, L. (1977) J. Biol. Chem. 252, 2331—2337
14 Endrenyi, L., Chan, M.-S. and Wong, J.T.-F. (1971) Can. J. Biochem. 49, 581—598
15 Chandler, J.P. (1977) Direct Search Optimization; Solution of Least Squares Problems. Program 307, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN, U.S.A.