Biophysical Chemistry, 30 (1988) 133–141
Elsevier

BPC 01252

Analysis and parameter resolution in highly cooperative systems

Stanley J. Gill, Patrick R. Connelly, Enrico Di Cera and Charles H. Robert

Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309-0215, U.S.A.

Received 21 October 1987
Accepted 20 January 1988

Data analysis; Cooperativity; Intermediate; Hemoglobin

We have examined common methods of analysis of highly cooperative systems such as oxygen binding by hemoglobin and thermal denaturation. Through extensive simulation of ligand-binding data for a tetrameric macromolecule we show that careful attention must be paid to the formulation of the fitting function and to proper assessment of the number of parameters involved. We conclude that the partition function should be formulated in terms of overall reaction parameters as opposed to stepwise reaction parameters and that bias is introduced by fixing physical parameters such as extrapolated end points.

1. Introduction

Highly cooperative reactions such as the thermal denaturation of proteins [1] and oxygen binding to respiratory proteins [2] place special demands upon the formulation of the relevant descriptive equations and the evaluation of reaction parameters. Analysis of the temperature dependence of the excess heat capacity of the thermal denaturation of ribonuclease A shows that two states, native and fully denatured, dominate the denaturation process, combined with the presence of a small fraction ($< 5\%$) of an intermediate state [3]. The analysis of oxygen binding to hemoglobin also indicates, as would be expected from the Hill slope of 3, the low population of intermediate oxygenated species [4].

A critical step in the analysis of thermal and ligand-binding experiments is the determination of equilibrium constants which characterize the cooperative properties of the system under study. Nonlinear least-squares methods are routinely used in determining the relevant parameters from heat capacity and binding curve data. Recently we have found that precise binding capacity * data for high concentrations of human hemoglobin show that the triply oxygenated species is virtually negligible [6]. This finding stands in contrast to previous determinations [4,7]. The observation has drawn our attention to the particular need for proper formulation of the fitting problem [8], especially when intermediates make small contributions to the binding process. In comparison to thermal binding processes [9], the chemical binding problem is defined in terms which are met more rigorously under careful experimental investigation. For these reasons we have chosen to treat this situation in detail.

In order to examine the influence of parameterization and fixing values of selected parameters we have carried out a detailed simulation study of three cases of different cooperativity for a tetrameric macromolecule. From a different stand-

Correspondence address: S.J. Gill, Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309-0215, U.S.A.

* The binding capacity is defined in parallel with the heat capacity and represents the change in the number of moles of ligand X per macromolecule ($\bar{X}$) due to a change in the chemical potential of the ligand $\partial\bar{X}/\partial\mu_X$ [5].

point, Johnson et al. [10] have investigated the dimer-tetramer linkage scheme for oxygen binding to hemoglobin. This scheme requires knowledge of ten parameters: the dimer-tetramer equilibrium constant, six oxygen-binding constants, the heme concentration, and two optical parameters. They conclude that with available optical absorbance precision this linkage scheme could not be resolved without fixing some parameters. Our focus here is to assess (1) the effect of fixing certain parameters, such as the optical parameters, which are commonly fixed to extrapolated values, and (2) the influence of formulating the fitting problem in terms of either stepwise or overall reactions.

2. Analytical form of the fitting function

For a ligand-binding situation in which the optical absorbance is linearly related to the degree of saturation, as is true to a good approximation with hemoglobin [11], the measured optical absorbance at a specific ligand X activity x is given by

$$A(x) = A(0) + [A(\infty) - A(0)]\theta(x) \qquad (1)$$

where $\theta(x)$ is the fractional ligand saturation, $A(0)$ and $A(\infty)$ denoting the optical absorbances at zero ($\theta(0) = 0$) and complete ($\theta(\infty) = 1$) saturation, respectively.

The most general approach to the formulation of $\theta(x)$ is the phenomenological one where the different stoichiometric species of ligation are considered [12,13]. In the case of human hemoglobin we can denote the ligated species by M, MX, MX_2, MX_3 and MX_4, where X represents oxygen. The two principal ways of describing the various equilibrium reactions among these species are:

(1) The overall reactions, in which the unligated species reacts with a given number of ligand molecules leading to the equilibria $M + jX \rightarrow MX_j$.

(2) The stepwise reactions, in which one ligand molecule reacts with a given ligated species leading to the equilibria $MX_{j-1} + X \rightarrow MX_j$.

From a thermodynamic point of view the overall and stepwise reactions are entirely equivalent for describing the binding process. The equilibrium constant for the j-th overall reaction is denoted by β_j and the equilibrium constant for the j-th stepwise reaction is denoted by K_j. The relation between the equilibrium constants for the two formulations is readily found as $\beta_j = K_1 K_2 \ldots K_j$ and $K_j = \beta_j / \beta_{j-1}$ with $\beta_0 = 1$. One sees that the concentration of any species MX_j is given by $[MX_j] = [M]\beta_j x^j = [M]K_1 K_2 \ldots K_j x^j$.

For this system, Ξ, the binding polynomial [12] or binding partition function [13], gives the function $\theta(x)$ by a simple differentiation. In terms of the overall constants βs, Ξ can be written

$$\Xi(x) = 1 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \qquad (2)$$

or in terms of the stepwise constants Ks, as

$$\Xi(x) = 1 + K_1 x + K_1 K_2 x^2 + K_1 K_2 K_3 x^3 + K_1 K_2 K_3 K_4 x^4 \qquad (3)$$

The fractional saturation $\theta(x)$ for a four-site macromolecule is obtained from the binding partition function $\Xi(x)$ by

$$\theta(x) = \frac{1}{4} \frac{\mathrm{d} \ln \Xi(x)}{\mathrm{d} \ln x} \qquad (4)$$

Either formulation of $\Xi(x)$ can be used in eq. 1 to express $\theta(x)$. It can be noted that analogous expressions determine the excess enthalpy of a multistate macromolecule from the derivative of ln Ξ with respect to $1/T$, and the excess heat capacity is then obtained by differentiation with respect to T [9].

With the expression for $\theta(x)$ and consequently $A(x)$, one can define the fitting function. Assume a set of N experimental optical readings A_i ($i = 1, 2, \ldots N$) obtained at different ligand activities x_i. In the fitting procedure one then minimizes the chi-squared statistic [14], a function of the squares of the deviations of the fitting function $A(x)$ from the data points. From eq. 1 we see that in the case of a tetramer there are four Adair constants (overall or stepwise) to be determined in the minimization, along with the asymptotic absorbance values $A(0)$ and $A(\infty)$.

Linear fitting functions when applied to the analysis of experimental data give rise to normally distributed parameters. A nonlinear fitting function like $A(x)$ generally gives rise to parameter

estimates that are not normally distributed [15]. A useful measure of the nonlinearity of a given parameter in a fitting equation is the bias [15], the deviation of a parameter determination with respect to the true value. From a large number of data sets the bias of the mean parameter values is considered to be a better indicator of nonlinearity than the cross-correlation matrix [15]. Ideally, either formulation of $\theta(x)$, i.e., in terms of the overall or stepwise reactions, would provide equivalent answers to the binding problem. However, the degree of nonlinearity of the fitting function depends upon the formulation used, and, as we shall see, in practice this feature can bias the parameter estimates. The resulting bias is increased with the extent of cooperativity observed in the system.

3. Influence of cooperativity on resolution of binding parameters

The cooperative nature of a ligand-binding process influences one's ability to resolve the relevant equilibrium constants of the system. If only the unligated and fully ligated species exist, i.e., all the intermediate overall binding constants are zero, then the system exhibits complete cooperativity. On the other hand, if all the binding sites are equivalent and independent then the situation is noncooperative, and the intermediate species are significantly populated. This applies equally well to thermal transition phenomena, where the amounts of intermediates determine the cooperative nature of the process [1,3]. As we have seen, the overall reaction constants determine the relative concentrations of the reaction species. For ligand binding the fractional population of the j-th ligated species is defined as $\alpha_j(x) = \beta_j x^j / \Xi(x)$. Since the degree of ligation can be defined in terms of the fractional populations of the various ligated species as $\theta(x) = (1/4)\Sigma j\alpha_j(x)$, clearly the most difficult situation arises when the intermediate species fractions make very low contributions to $\theta(x)$, and thus to the experimental measurements.

4. Data simulation

In order to investigate the effects of both nonlinearity and cooperativity on the resolvability of the equilibrium constants, we have fitted 'data' generated by computer simulation in which random experimental error of a given magnitude is added to each point of a binding curve. Three sets of 100 binding curves each were generated in the form of optical absorbances A_i assuming noncooperative (case A), cooperative (case B) and highly cooperative (case C) binding. A random normal error was added to each of the 25 optical readings of each binding curve, corresponding to the error usually found in accurate optical measurements ($\sigma = 0.001$ absorbance units). All data sets were then fitted to eq. 1, with $\theta(x)$ expressed in terms of the βs or the Ks, using a nonlinear least-squares Gauss-Newton algorithm as modified by Marquardt and others [14]. The true values of the βs and the Ks used in generating the data were employed as starting guesses in the fitting procedure. Starting guesses were such that both formulations converged in the same region in every case. The 300 total fitted determinations of all the parameters were collected and the mean and the standard deviation of the distribution of each parameter are reported in table 1.

5. Results

Typical simulated binding curves are shown in fig. 1 for the three different cooperativities beneath the corresponding distributions of the stoichiometric species.

5.1. Effect of the analytical form of the fitting function

The overall constants (β) and the stepwise constants (K) are essentially normally distributed in cases A and B. This is not so for case C, where the cooperativity is high, as shown in fig. 2. In this case, the βs retain a Gaussian distribution whereas K_3 and K_4 are clearly not normally distributed. This is also shown by the lower bias of the overall Adair constants as given in table 1. The stepwise

Table 1

Values of the parameters obtained in fitting the simulated cases as compared to the true values used in generating the data

The distribution of each parameter is characterized by the mean 'm' and the standard deviation 'σ'. The % error is given by $100(\sigma/m)$. The value of $\beta_4 = 1$ indicates that the unligated and fully ligated species are equally populated when the ligand activity has a value of unity.

	True	Fitted		% bias	% error
		m	σ		
Case A					
β_1	4.00	3.99	0.0823	0.2	2.1
β_2	6.00	5.99	0.0676	0.2	1.1
β_3	4.00	3.99	0.0784	0.2	2.0
β_4	1.00	0.997	0.0155	0.3	1.5
K_1	4.00	3.99	0.0823	0.2	2.1
K_2	1.50	1.51	0.0359	−0.7	2.4
K_3	0.667	0.666	0.0160	0.1	2.4
K_4	0.250	0.249	0.00501	0.4	2.0
$A(0)$	0.100000	0.100028	0.000698	0.0	0.7
$A(\infty)$	1.100000	1.100118	0.000745	0.0	0.1
Case B					
β_1	0.535	0.536	0.0300	−0.2	5.6
β_2	0.353	0.357	0.0389	−1.1	10.9
β_3	0.168	0.166	0.0303	1.2	18.2
β_4	1.00	1.00	0.0112	0.0	1.1
K_1	0.535	0.536	0.0300	−0.2	5.6
K_2	0.660	0.648	0.111	1.8	17.1
K_3	0.476	0.504	0.143	−5.9	28.4
K_4	5.95	6.24	1.26	−4.9	20.2
$A(0)$	0.100000	0.099939	0.000548	0.1	0.5
$A(\infty)$	1.100000	1.099948	0.000602	0.0	0.0
Case C					
β_1	0.535	0.534	0.0291	0.2	5.4
β_2	0.0353	0.0375	0.0382	−6.2	101.9
β_3	0.0168	0.0160	0.0273	4.8	170.6
β_4	1.00	1.00	0.00875	0.0	0.9
K_1	0.535	0.534	0.0291	0.2	5.4
K_2	0.0660	0.0737	0.0761	−11.7	103.2
K_3	0.476	2.22	48.7	−366.4	2193.7
K_4	59.5	41.0	235	31.1	573.2
$A(0)$	0.100000	0.099985	0.000525	0.0	0.5
$A(\infty)$	1.100000	1.100011	0.000541	0.0	0.0

Adair constants show high percentage errors and high bias. Thus, as demonstrated by this measure of nonlinearity, the Ks are nonlinear parameters and the extent of nonlinearity increases with cooperativity. On the other hand, the overall Adair constants show lower percentage errors and can be considered quasi-linear parameters, with only a slight increase of nonlinearity with cooperativity. Consequently, in the cases of moderate or low cooperativity there is no difference between the stepwise or overall reaction formulations, but in the highly cooperative case the binding system is

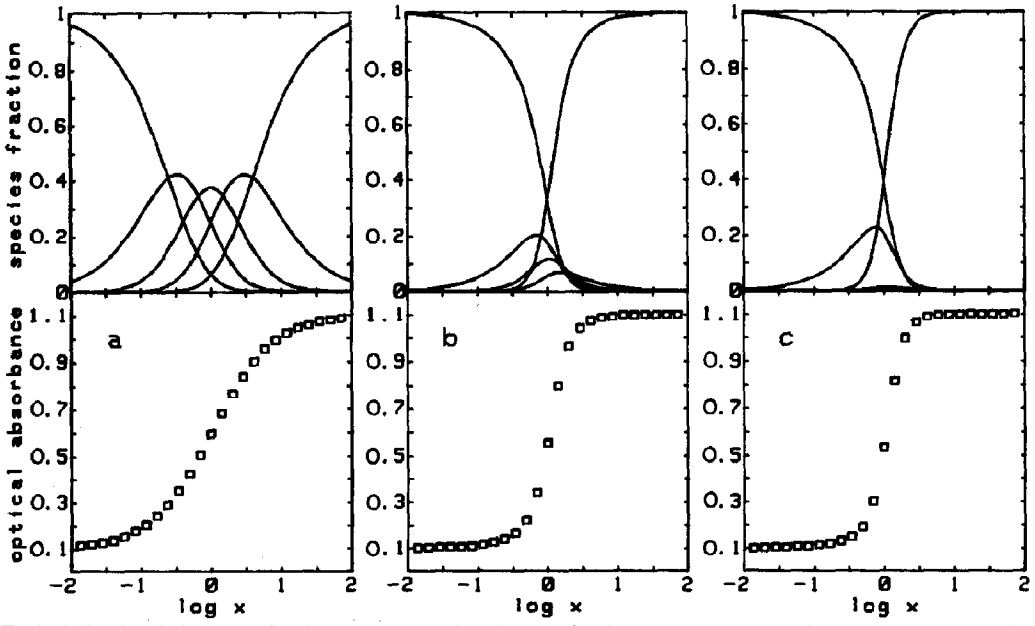

Fig. 1. Typical simulated data sets for (a) non-cooperative (case A in the text), (b) cooperative (case B, text), and (c) highly cooperative (case C, text) tetramer plotted as optical absorbances vs. the logarithm of the ligand activity. At the top of the binding curve is shown the distribution of the stoichiometric species vs. the logarithm of the ligand activity.

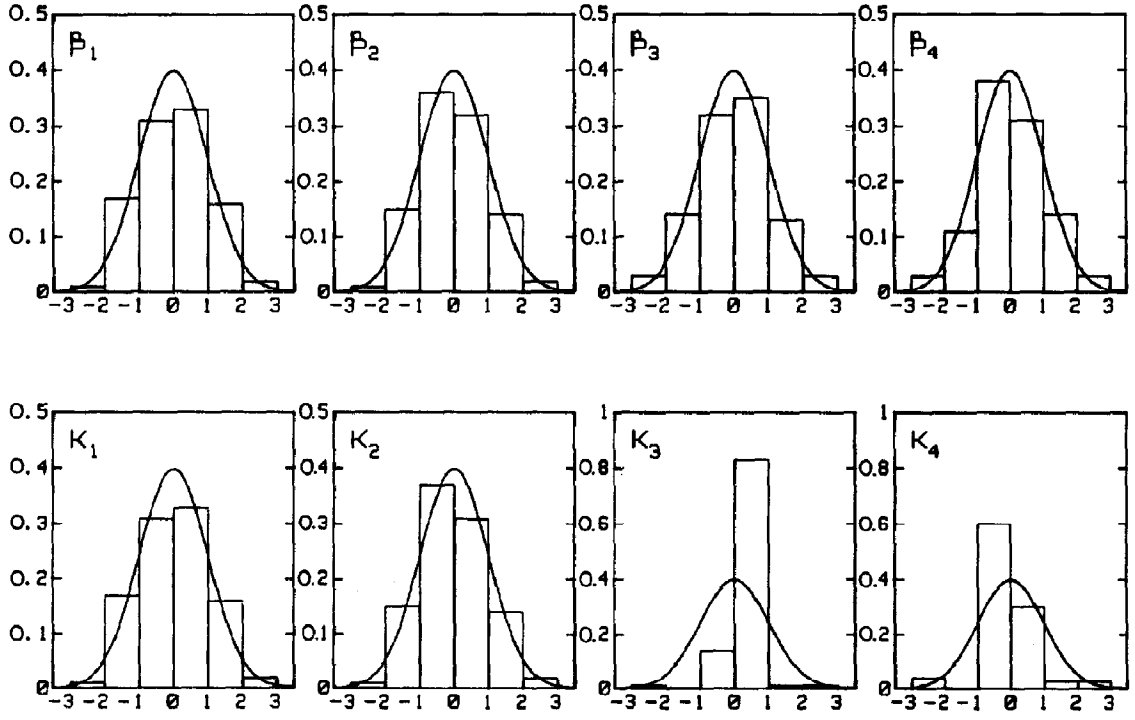

Fig. 2. Histograms of fitted values obtained for case C discussed in the text. The mean of the distribution has been placed at zero and the resulting values of the parameters, normalized by the standard deviation, have been collected in the intervals shown (± 1, 2, 3 standard deviations) to give the frequency profile in its standard form [10]. The continuous line depicts the expected normal distribution for a linear parameter. Note that the frequency scale is different for K_3 and K_4.

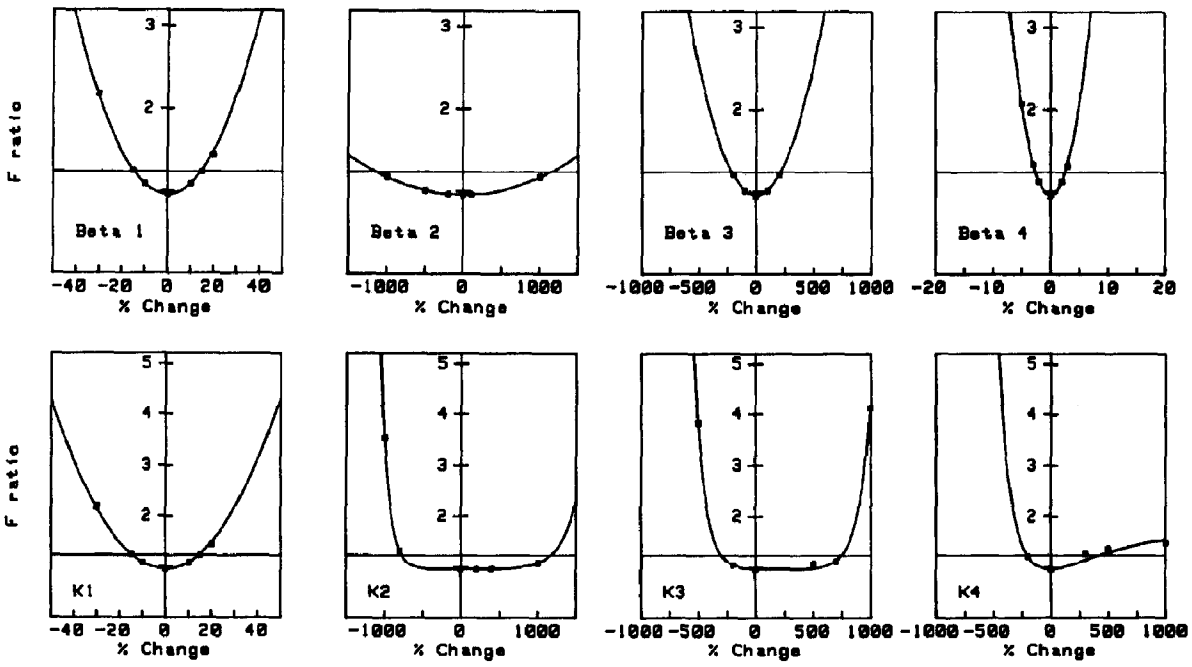

Fig. 3. Confidence intervals for the overall and stepwise Adair constants obtained by F-testing a typical binding curve from case C discussed in the text.

Table 2

Effect of fixing the end parameters $A(0)$ and $A(\infty)$ on the determination of the overall and stepwise Adair constants for a cooperative binding system

	True	Fitted		% bias	% error
		m	σ		
β_1	0.535	0.517	0.0385	3.4	7.4
β_2	0.353	0.385	0.0484	−9.1	12.6
β_3	0.168	0.141	0.0368	19.1	26.1
β_4	1.00	1.01	0.0155	−1.0	1.5
K_1	0.535	0.517	0.0385	3.4	7.4
K_2	0.660	0.755	0.146	−14.4	19.3
K_3	0.476	0.382	0.149	19.7	39.0
K_4	5.95	7.79	2.64	−30.9	33.4
Case C					
β_1	0.535	0.513	0.0353	4.1	6.9
β_2	0.0353	0.0693	0.0431	−96.3	62.2
β_3	0.0168	0.00834	0.0330	50.3	395.7
β_4	1.00	1.01	0.0155	−1.0	1.5
K_1	0.535	0.513	0.0353	4.1	6.9
K_2	0.0660	0.140	0.0933	−112.1	66.6
K_3	0.476	0.761	6.24	−59.9	820.0
K_4	59.5	3.60	470	93.9	13055.5

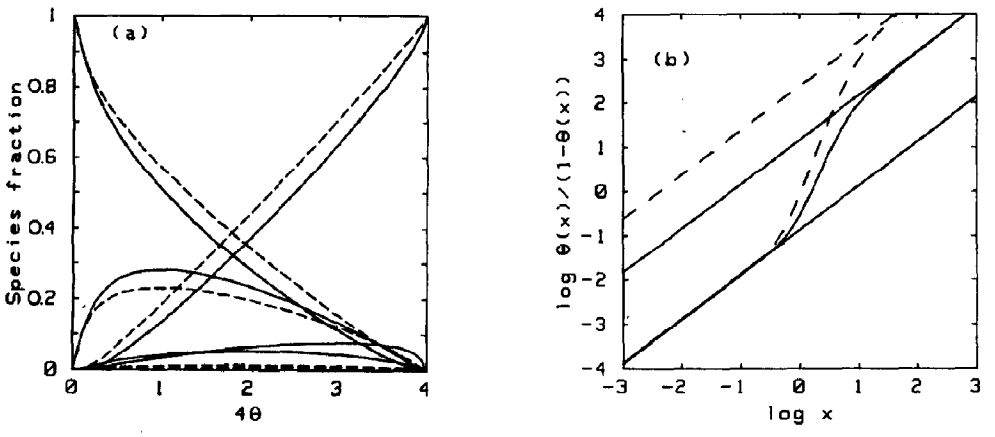

Fig. 4. Effect of fixing the end parameters $A(0)$ and $A(\infty)$ on the resolvability of the Adair constants for case C discussed in the text, according to the mean values of the stepwise constants Ks reported in table 2. (a) Distribution of the fractional population of stoichiometric species of ligation: broken line shows true distribution. (b) Hill plot: broken line shows the true curve.

better resolved when the problem is cast in terms of the overall Adair constants. This conclusion is also supported by the fact that the Ks calculated from the mean values of the fitted βs are less biased than the fitted Ks.

5.2. *Significance of the F-test for evaluation of confidence intervals*

A major concern of nonlinear least-squares analysis is the determination of confidence intervals on fitted parameters. The standard deviation of a parameter distribution can be used as a confidence interval for the parameter only if the distribution is symmetric. In the case of nonlinear problems the F-test is useful for evaluation of intervals at approximate confidence [6,16]. An F-test yields asymmetric confidence intervals if the parameter is highly nonlinear [14,16]. Fig. 3 shows in case C that the confidence intervals for the Ks are far more asymmetric than those for the βs. Indeed, the reliability of the F-test is reduced when applied to increasingly nonlinear fitting functions [6,14]. Thus, in the analysis of real experimental data the confidence intervals of the overall constants will be more reliable than those determined for the stepwise constants, because of the higher nonlinearity of the stepwise constants.

5.3. *Relevance of physical parameters*

The asymptotic parameters $A(0)$ and $A(\infty)$ are practically linear parameters, as seen from table 1 for all cases. Consequently, their resolvability does not present a problem. Yet these parameters are often arbitrarily fixed at values obtained from extrapolation of the absorbances at low and high ligand activities, and are not optimized in the fitting procedure [4,17]. This procedure introduces a bias into the binding constant determinations.

For emphasis of this point we have fitted the data sets of cases B and C by exactly such a procedure: in all cases the first ten points ($\theta(x) < 0.04$) have been extrapolated by a quadratic in the ligand activity x: $A(0) + c_1x + c_2x^2$, with c_1 and c_2 being the fitting constants; the last ten points ($\theta(x) > 0.96$) were extrapolated with a quadratic in the inverse activity $1/x$: $A(\infty) + c_3/x + c_4/x^2$, where c_3 and c_4 are the fitting constants. Such a procedure is widely used [4,17]. With $A(0)$ and $A(\infty)$ held fixed at the values obtained by these extrapolations, we obtained the results listed in table 2. One sees that all the Adair constants, overall and stepwise, are biased when this procedure is used. The effect is especially severe in the highly cooperative case, as is encountered with human hemoglobin.

6. Discussion

There have been many determinations of Adair constants for human hemoglobin in the past three decades which have inspired a variety of mechanistic models of ligand binding and a continued effort towards the improvement of experimental techniques. Nevertheless, it still seems that inadequate attention has been given to critical statistical analysis of experimental data. The results presented here suggest that the procedures often employed in the analysis of highly cooperative systems such as human hemoglobin may be flawed. Use of the stepwise equilibrium constants has been historically preferred in the fitting procedure, and in many cases these constants have been obtained by fixing the end parameters $A(0)$ and $A(\infty)$ [4,17], and then used in selection or exclusion of particular mechanisms for cooperativity in ligand binding [18,19].

The distortion of intermediate populations that can be caused by fitting with an incorrect parameterization is best seen in fig. 4a, where we depict the species fractions for the high-cooperativity case as calculated by the mean values of the Ks with end parameters $A(0)$ and $A(\infty)$ held fixed. In this figure the true distribution of the ligated species is shown by broken lines. Note that fitting with the Ks and with the ends fixed greatly overestimates the true contribution of both the doubly and triply ligated species to the ligation process. The bias in the binding parameters determined with such a procedure invalidates even qualitative mechanistic interpretation of the underlying processes. The popular Hill plot exaggerates the discrepancy, as shown in fig. 4b.

The expectation that fixing thermodynamic parameters would lead to bias has been noted in a previous study [10]. However, the effect of fixing asymptotic absorbance values on the determination of oxygen-binding parameters for human hemoglobin was not recognized [17]. In subsequent experiments these end parameters were freed for optimization and resulting best-fit estimates for intermediate species populations were greatly decreased [7], in agreement with the results of the present simulation study. It is thus not surprising that the conclusion of recent studies [6,11,20–22] showing a negligible contribution of the triply ligated species to the oxygenation process of human hemoglobin has perhaps been obscured previously by inadequate analysis of experimental data.

We have examined the analysis of cooperative ligand binding in some detail. The conclusions reached extend beyond this situation to generalized binding [9] which includes for example thermal and pressure denaturation phenomena. Particularly in the study of highly cooperative processes the partition function Ξ should be formulated in terms of overall reaction parameters as opposed to stepwise reaction parameters. Additionally, care must be taken to avoid bias introduced by fixing physical parameters that describe particular experiments to arbitrary values such as interpolated baselines and extrapolated end points.

Acknowledgement

Financial support from NIH grant HL22325 is gratefully acknowledged.

References

1 P. Privalov, Adv. Protein Chem. 33 (1979) 167.
2 J. Wyman, Adv. Protein Chem. 4 (1948) 407.
3 E. Freire and R.L. Biltonen, Biopolymers 17 (1978) 167.
4 K. Imai, Allosteric effects in haemoglobin (Cambridge University Press, Cambridge, 1982).
5 E. Di Cera, S.J. Gill and J. Wyman, Proc. Natl. Acad. Sci. USA 85 (1988) 449.
6 S.J. Gill, E. Di Cera, M.L. Doyle, G.A. Bishop and C.H. Robert, Biochemistry 26 (1987) 3995.
7 A.H. Chu, B.W. Turner and G.K. Ackers, Biochemistry 23 (1984) 604.
8 W.G. Bardsley and P.B. McGinlay, J. Theor. Biol. 126 (1987) 183.
9 S.J. Gill, B. Richey, G.A. Bishop and J. Wyman, Biophys. Chem 21 (1985) 1.
10 M.L. Johnson, H.R. Halvorson and G.K. Ackers, Biochemistry 15 (1976) 5363.
11 M.L. Doyle, E. Di Cera and S.J. Gill, Biochemistry 27 (1988) 820.
12 J. Wyman, Adv. Protein Chem. 19 (1964) 315.
13 T.L. Hill, Cooperativity theory in biochemistry (Springer-Verlag, Berlin, 1985).
14 M.E. Magar, Data analysis in biochemistry and biophysics (Academic Press, New York, 1972).

15 D.A. Ratkowski, Non-linear regression modeling (Marcel Dekker, New York, 1983).
16 M.L. Johnson, Biophys. J. 44 (1983) 101.
17 F.C. Mills, M.L. Johnson and G.K. Ackers, Biochemistry 15 (1976) 5350.
18 G.K. Ackers and M.L. Johnson, J. Mol. Biol. 147 (1981) 559.
19 M.F. Perutz, G. Fermi, B. Luisi, B. Shaanan and R.C. Liddington, Acc. Chem. Res. 20 (1987) 309.
20 M.L. Doyle, E. Di Cera, C.H. Robert and S.J. Gill, J. Mol. Biol. 196 (1987) 927.
21 A. Parody-Morreale, C.H. Robert, G.A. Bishop and S.J. Gill, J. Biol. Chem. 262 (1987) 10994.
22 E. Di Cera, M.L. Doyle, P.R. Connelly and S.J. Gill, Biochemistry 26 (1987) 6494.