Oxygenation-Linked Subunit Interactions in Human Hemoglobin: Analysis of Linkage Functions for Constituent Energy Terms†

Michael L. Johnson,‡ Herbert R. Halvorson,§ and Gary K. Ackers*

ABSTRACT: Resolution of the linkage functions between oxygenation and subunit association–dissociation equilibria in human hemoglobin into the constituent microscopic terms has been explored by numerical simulation and least-squares analysis. The correlation properties between parameters has been studied using several choices of parameter sets in order to optimize resolution. It is found that, with currently available levels of experimental precision and ranges of variables, neither linkage function can provide sufficient resolution of all the desired energy terms. The most difficult quantities to resolve always include the dimer–tetramer association constant for unliganded hemoglobin and the oxygen binding constants to $\alpha\beta$ dimers. A feasible experimental strategy for overcoming these difficulties lies in independent determination of the dimer–tetramer association constants for unliganded and fully oxygenated hemoglobin. These constants, in combination with the median ligand concentration, provide an estimate of the energy for total oxygenation of tetramers which is essentially independent of the other constituent energies. It is shown that if these separately determinable parameters are fixed, the remaining terms may be estimated to good accuracy using data which represents either linkage function. In general it is desirable to combine information from both types of experimental quantities. A previous paper (Mills, F. C., Johnson, M. L., and Ackers, G. K. (1976), *Biochemistry, 15,* the preceding paper in this issue) describes the experimental implementation of this strategy.

It has been recognized for some time that subunit dissociation studies may provide an effective means of probing intersubunit contact energy changes which accompany cooperative ligand binding in human hemoglobins and a number of experimental studies have been aimed toward this goal (Anderson et al., 1970; Kellett, 1971; Gilbert et al., 1975; Thomas and Edelstein, 1972, 1973; Ip et al., 1976; Williams and Kim, 1975; Ackers et al., 1976). The dissociation properties of hemoglobin tetramers into dimers are linked to the ligand binding reactions of those species so that an important manifestation of subunit dissociation is the concentration dependence of oxygenation curves. Recent technical developments (Imai et al., 1970; Imai, 1973) have greatly increased the feasibility of experimentally studying this important linkage function. Such a study is described in the previous paper (Mills et al., 1976). Progress is also being made in the development of more sensitive techniques capable of directly studying subunit dissociation of partially liganded hemoglobins in dilute solution (Crepeau et al., 1974; Williams, 1976; Ackers et al., 1976). Although subunit dissociation is not of physiological interest per se, studies of dissociation are of great value in relating changes in the intersubunit contact energies to functional properties of hemoglobin tetramers (Noble, 1969; Thomas and Edelstein, 1972; Weber, 1972; Ackers and Halvorson, 1974; Ip et al., 1976). They provide a means to explore the ways in which cooperative interactions between remote binding sites are expressed as energetic constraints across subunit contact planes.

Recently a model-independent linkage analysis has been formulated, delineating the thermodynamic relationships between oxygenation and subunit dissociation equilibria in hemoglobin (Ackers and Halvorson, 1974). In that study some preliminary numerical analyses were carried out to explore the sensitivity of the linkage functions to various ways of distributing the oxygenation-linked subunit interaction energy with respect to the sequential binding steps.

In the present study we have investigated, by numerical analysis, the problem of actually resolving all the relevant energy terms from the experimentally determinable linkage functions $\bar{Y}$ vs. $[P_t]$ (Figure 1) and ${}^{x}K_2$ vs. [X] (Figure 2). Numerical analysis provides considerable insight into the requirements of experimental design which must be implemented to extract the desired information. Since the parameters to be estimated are not statistically independent, it is of critical importance to establish the conditions under which a particular set of derived energy terms will be reliable. In the previous paper (Mills et al., 1976) we presented an experimental study which provides a complete resolution of the energy distribution problem for human hemoglobin under one set of conditions. The analyses described in this paper provide the background necessary to the interpretation of that study.

Since the relationships between energy terms obtained under many sets of conditions and for different hemoglobins (e.g., mutants and chemically modified species) may ultimately be more important than their actual values, we are mainly concerned at this stage with generating the capability (both experimentally and analytically) of studying a variety of hemoglobin systems rather than with the determination of a particular set of energy terms pertaining to any one set of conditions. Accordingly, we present results of detailed analyses on an illustrative case which is relevant to real hemoglobin systems, and which embodies the general features and conclusion of a much larger class of possibilities. It is recognized that saturation curves, $\bar{Y}$, as a function of protein concentration, $[P_t]$, and macroscopic dimer–tetramer association constants,

† From the Department of Biochemistry, University of Virginia, Charlottesville, Virginia 22901. *Received April 26, 1976.* Supported by Grants BMS74-24507 from the National Science Foundation, and GM-14493 from the National Institutes of Health.

‡ Recipient of United States Public Health Service Postdoctoral Fellowship AM03201.

§ Present address: Edsel B. Ford Institute of Medical Research, Detroit, Michigan 48202.

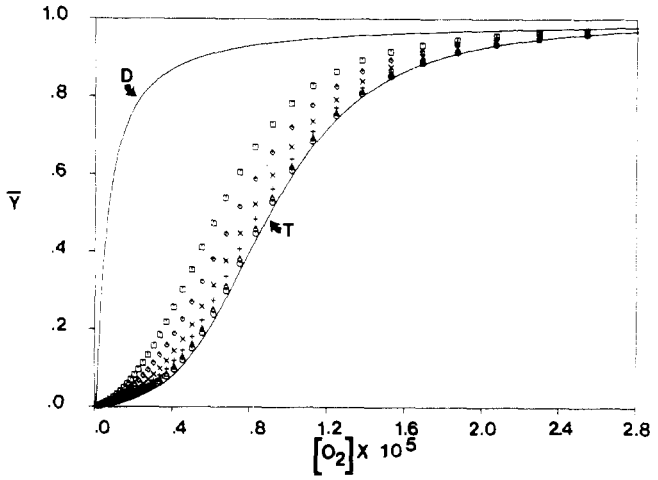

FIGURE 1: Synthetic data illustrating the effect of protein concentration $[P_t]$ on ligand binding curves. Data were calculated from experimentally determined parameters corresponding to human hemoglobin A in 0.1 M Tris,[1] 0.1 M NaCl, 1 mM Na_2EDTA and titrated to pH 7.4 with HCl, at 21.5 °C (Mills et al., 1976). $K_{21} = 3.454 \times 10^6$ mol^{-1}, $K_{22} = 2.980 \times 10^{12}$ mol^{-2}, $K_{41} = 4.583 \times 10^4$ mol^{-1}, $K_{42} = 5.800 \times 10^9$ mol^{-2}, $K_{43} = 2.462 \times 10^{14}$ mol^{-3}, $K_{44} = 1.708 \times 10^{20}$ mol^{-4}, and ${}^0K_2 = 5.108 \times 10^{10}$ mol^{-1}. The outside lines represent the binding isotherms for dimer and tetramer and the data points represent (left to right) values of $[P_t]$ to 1.0, 5.0, 10.0, 50.0, 100, and 500 μM heme. Each curve consists of 50 data points at different oxygen concentratons between 0.1 and 60 mmHg.

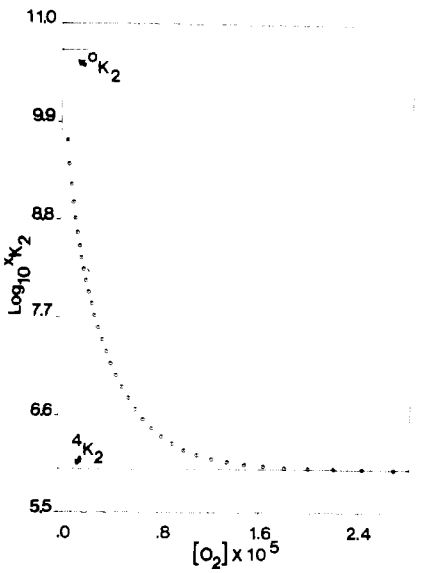

FIGURE 2: Synthetic data illustrating the effect of ligand concentration on the dimer to tetramer association constant, ${}^xK_2$. Data were simulated for the same hemoglobin system as described in Figure 1. In this case, the oxygen concentration was varied (0.263 to 60.0 mmHg) such that ${}^xK_2$ varied over the range (5×10^9 to 9.8×10^5).

${}^xK_2$ vs. pO_2, may be obtained by a number of experimental techniques. Although the analyses presented here are independent of particular techniques used to obtain the linkage functions, the error levels and ranges for the variables have been chosen with currently feasible methods in mind.

Definition of Terms

The specific phenomena with which we will be concerned are the dissociation equilibria involving cleavage of tetramers along the plane which forms dimers of the $\alpha^1\beta^1$ type. It is this dissociation process which is known to occur in dilute solutions of hemoglobin near neutral pH (Rosemeyer and Huehns, 1967; Soni, 1973). In Figure 1 of the preceding paper in this issue the various liganded states of the hemoglobin tetramer are shown on the right and those of the dimer on the left. We assume no appreciable dissociation beyond the dimeric stage into separate α and β chains (Ackers and Thompson, 1965; Chiancone et al., 1968; Kellett and Schachman, 1971; Elbaum and Herskovits, 1974). A state designated $(\alpha_2\beta_2)X_i$ ($i = 1, 2, 3, 4$) represents all tetrameric species having i ligands bound, with no partitioning of these into microscopic forms in which the bound ligands are distributed differently. A state designated $(\alpha\beta)X_i$ ($i = 1, 2$) represents all dimeric species with i ligands bound, regardless of microscopic distribution. The species designations and the corresponding energy relationships are well-defined averages over many possible microscopic states. We are concerned with well-defined average energy changes at the dimer–dimer contact region (Ackers and Halvorson, 1974).

In Figure 1 of the previous paper (Mills et al., 1976) the subunit association equilibria are depicted. Equilibrium constants to be used in this study are defined below. The formation constants for various liganded states of tetramer are:

$$ {}^iK_2 = \frac{[(\alpha_2\beta_2)X_i]}{[(\alpha\beta)X_j][(\alpha\beta)X_m]}, \qquad j + m = i \tag{1}$$

[1] Abbreviations used: Tris, tris(hydroxymethyl)aminomethane; EDTA, ehtylenediaminetetraacetic acid.

It should be noted that ${}^2K_2$ is ambiguous since it can be defined in two ways ($j = m = 1$ or $j \neq m$). The stepwise binding constants are defined as

$$k_{2j} = \frac{[(\alpha\beta)X_j]}{[(\alpha\beta)X_{j-1}][X]} \tag{2}$$

$$k_{4j} = \frac{[(\alpha_2\beta_2)X_j]}{[(\alpha_2\beta_2)X_{j-1}][X]} \tag{3}$$

The product binding constants (Adair constants) are defined as

$$K_{mi} = \prod_{j=1}^{i} k_{mj} \tag{4}$$

To define the equilibria depicted in the linkage scheme seven independent parameters are required. They may be individual equilibrium constants, or combinations of the constants. Several choices will be explored in this paper.

At this point it is worth noting that, in addition to the tetramer and dimer binding constants, primary interest lies in ultimately obtaining estimates of the ratios ${}^0K_2/{}^1K_2$, ${}^1K_2/{}^3K_2$, and ${}^3K_2/{}^4K_2$. These ratios provide highly useful information on the way in which the oxygenation-linked subunit interaction energy is partitioned with respect to the sequential binding steps.

The functional forms of the saturation function, $\bar{Y}$, and subunit association constant, ${}^xK_2$, have been derived previously (Ackers and Halvorson, 1974) and are given by eq 5 and 6, respectively:

$$\bar{Y} = \frac{Z_2' + Z_4'(\sqrt{Z_2^2 + 4\,{}^0K_2Z_4[P_t]} - Z_2)/4Z_4}{Z_2 + \sqrt{Z_2^2 + 4\,{}^0K_2Z_4[P_t]}}$$

$$ {}^xK_2 = {}^0K_2\frac{Z_4}{(Z_2)^2} \tag{5}$$

where

$$Z_2 = 1 + K_{21}[X] + K_{22}[X]^2$$

$$Z_2' = K_{21}[X] + 2K_{22}[X]^2$$

$$Z_4 = 1 + K_{41}[X] + K_{42}[X]^2 + K_{43}[X]^3 + K_{44}[X]^4$$

$$Z_4' = K_{41}[X] + 2K_{42}[X]^2 + 3K_{43}[X]^3 + 4K_{44}[X]^4 \tag{6}$$

and $[P_t]$ is the protein concentration expressed as total moles of heme per liter.

A third linkage function of considerable use is the median ligand concentration $[\overline{X}]$ which is defined (Wyman, 1964) for any saturation function as:

$$\int_0^{[\overline{X}]} \overline{Y}\, d \ln [X] = \int_{[\overline{X}]}^{\infty} (1 - \overline{Y})\, d \ln [\overline{X}] \quad (7)$$

For binding by a single macromolecular species onto n sites, $[\overline{X}]$ is related to the binding constant K_{nn} for total ligation of the macromolecule by the relationship: $[\overline{X}]^n = K_{nn}^{-1}$ (Wyman, 1964). For a system of dimers and tetramers in equilibrium, the value of $[\overline{X}]$ is a function of the subunit association constants ${}^0K_2$, and ${}^4K_2$, as well as the tetrameric Adair constant K_{44}. The total saturation function for such a system may be written:

$$\overline{Y} = f_2\overline{Y}_2 + f_4\overline{Y}_4 = \overline{Y}_4 + f_2(\overline{Y}_2 - \overline{Y}_4) \quad (8)$$

where $\overline{Y}_2$ and $\overline{Y}_4$ are saturation functions for dimer and tetramer, respectively; f_2 and f_4 are the weight fractions of hemoglobin present as dimer and tetramer, respectively (these fractions depend upon state of ligation as well as total hemoglobin concentration, but always sum to unity).

Since

$$\frac{d \ln {}^xK_2}{d \ln [X]} = 4(\overline{Y}_4 - \overline{Y}_2) \quad (9)$$

(Wyman, 1964; Ackers and Halvorson, 1974), we can write eq 8 as:

$$\overline{Y} = \overline{Y}_4 - \frac{1}{4} f_2 \frac{d \ln {}^xK_2}{d \ln [X]} \quad (10)$$

Substituting this result and the relation $\overline{Y}_4 = \frac{1}{4} d \ln Z_4 / d \ln [X]$ into eq 7 yields:

$$\int_0^{[\overline{X}]} d \ln Z_4 - \int_{[\overline{X}]}^{\infty} d \ln \left\{\frac{[X]^4}{Z_4}\right\} = \int_0^{\infty} f_2 d \ln {}^xK_2 \quad (11)$$

Substituting

$$f_2 = \frac{\sqrt{1 + 8[P_t]{}^xK_2} - 1}{4[P_t]{}^xK_2} \quad (12)$$

into eq 11 the integrals may be evaluated to yield the final result:

$$K_{44} = [X]^{-4} \left\{\frac{1 - {}^4f_2}{1 - {}^0f_2}\right\} \exp({}^0f_2 - {}^4f_2) \quad (13)$$

where:

$${}^4f_2 = \frac{\sqrt{1 + 8[P_t]{}^4K_2} - 1}{4[P_t]{}^4K_2},$$

$${}^0f_2 = \frac{\sqrt{1 + 8[P_t]{}^0K_2} - 1}{4[P_t]{}^0K_2}$$

It can be seen from eq 13 that the median ligand concentration $[\overline{X}]$ provides a determination of the binding constant K_{44} (and hence the total energy of saturation) for tetramer, provided the subunit association constants ${}^0K_2$, ${}^4K_2$, and the hemoglobin concentration $[P_t]$ are known.

Numerical Methods

Minimization Procedure. The minimization procedure used was based upon the Gauss–Newton method (Magar, 1972, p 149; Box, 1960). We modified this procedure to allow for weighting factors and to improve the convergence properties. These modifications are described below.

The basic Gauss–Newton procedure for finding several parameters, α_j, by fitting experimental data to some function, ϕ, is based on a series expansion about the independent variables, X_i, and the current best approximation of the parameters, α^k, with a weighting factor, W_i

$$Y_i \simeq W_i\phi\{\alpha^f, X_i\} \simeq W_i\phi\{\alpha^k, X_i\} + \sum_j W_i \frac{\partial\phi\{\alpha^k, X_i\}}{\partial\alpha_j{}^k}(\alpha_j{}^f - \alpha_j{}^k) \quad (14)$$

where the f superscripts refer to the final values of the parameters, k superscripts refer to some intermediate iteration, the unsubscripted α is a vector of all the parameters, the j subscripts refer to a particular parameter, and the i subscripts refer to a particular data point. This can be expressed in matrix notation where ϵ is a vector approximating the difference between α^f and α^k, $\mathbf{Y}^*$ is a vector which is the weighted difference between the data points and the function evaluated at the current best approximation of the parameters, and P is a nonsquare matrix of partial derivatives. The correction vector, ϵ, can then be evaluated by using the transpose, P', of the matrix P:

$$\epsilon = (P'P)^{-1}P'\mathbf{Y}^* \quad (15)$$

This value of the correction vector can then be used to find a new set of parameters, α^{k+1}, and the process repeated until the correction vector is arbitratily close to zero. The modification of the method which we used to improve the convergence properties was to introduce a constant, t, which is varied at each iteration to obtain a minimum variance:

$$\alpha^{k+1} = \alpha^k + t\epsilon \quad (16)$$

For the analyses of simulated data, the values of parameters used to synthesize the data were used as initial guesses. The fitting is thus biased in favor of the "correct" answers since the probability of convergence to these is maximized whenever they lie near a minimum in residual space. For analysis of real data, described in the previous paper (Mills et al., 1976), a wide range of initial guesses was employed.

Correlation between Parameters. The cross-correlation coefficient, CC_{ij}, between two fitted parameters (i and j) is defined in terms of the ij element of the inverse matrix $(P'P)^{-1}$ as:

$$CC_{ij} = \frac{\{(P'P)_{ij}{}^{-1}\}}{\sqrt{\{(P'P)_{ii}{}^{-1}\}\{(P'P)_{jj}{}^{-1}\}}} \quad (17)$$

This coefficient always has an absolute value less than or equal to unity and is a measure of the degree to which uncertainty in the estimated value of one parameter may be linked to uncertainty in the estimated value of the second parameter. The degree of correlation between parameters is greater if their cross-correlation coefficient is closer to unity and only if the coefficient is zero are they uncorrelated altogether. The coefficients depend upon the functional form of the equation, the values of parameters, and the range of the independent variable, but are essentially independent of random experimental error. It is obvious that highly correlated parameters will be difficult to resolve and thus the correlation coefficient is a useful measure of the difficulty to be expected in a particular minimization problem. The correlation coefficients also reflect the sensitivity to small systematic errors which are likely to occur in experimental measurements.

TABLE I: Cross-Correlation Coefficients for Parameter Set 1.[a]

A. Oxygenation Data

K_{21}	1.000						
K_{22}	0.99996	1.000					
K_{41}	−0.510	−0.507	1.000				
K_{42}	0.576	0.574	−0.917	1.000			
K_{43}	−0.464	−0.461	0.814	−0.936	1.000		
K_{44}	0.346	0.345	−0.004	0.292	−0.421	1.000	
$^0K_2$	−0.99996	1.00000	−0.507	0.575	−0.462	0.346	1.000
	K_{21}	K_{22}	K_{41}	K_{42}	K_{43}	K_{44}	$^0K_2$

B. Subunit Association Data

K_{21}	1.000						
K_{22}	0.103	1.000					
K_{41}	−0.843	0.425	1.000				
K_{42}	0.718	−0.525	−0.967	1.000			
K_{43}	−0.539	0.574	0.852	−0.953	1.000		
K_{44}	−0.921	0.288	0.977	−0.893	0.732	1.000	
$^0K_2$	0.996	0.175	−0.809	0.685	−0.509	−0.891	1.000
	K_{21}	K_{22}	K_{41}	K_{42}	K_{43}	K_{44}	$^0K_2$

[a] Data were simulated as in Figures 1 and 2.

Estimation of Confidence Limits. The confidence intervals for simultaneously determined parameters are found by searching two sets of vectors, each in both directions from the minimum of residual space, until the desired variance ratio (F statistic, Zelen and Severo, 1964) is achieved. One set of vectors is obtained by varying each of the parameters independently and the other set of vectors is the axes of the "ellipsoidally shaped" region in n-dimensional space given by the equation:

$$(\alpha - \alpha^j)PP'(\alpha - \alpha^j) \leq ns^2F \qquad (18)$$

where s^2 is the variance of the minimum, F is the variance ratio from statistical tables, and α is a set of vectors which determines the axes of this region of space (cf. Magar, 1972, p 243). The largest deviations (in each direction) of each parameter are taken to be the positive and negative confidence intervals. It should be noted that these regions are not symmetrical, reflecting the nonlinearity of the problem. The confidence limits for parameter estimates presented in this paper correspond to one standard deviation.

Generation of Simulated Data. A wide range of cases relevant to real hemoglobin systems were explored using the approaches to be described. We present here only one typical case (i.e., one set of assumed energy terms) which corresponds to the real system experimentally studied in the previous paper (the case C analysis of Mills et al., 1976).

For the analyses to be presented, synthetic data were simulated as shown in Figures 1 and 2. Normally distributed random errors corresponding to a specified variance were added to the ligand binding data. Since the variance of these errors was constant throughout a given series of calculations, no weighting factors were used in the fitting to these data. Median ligand concentration values were obtained by numerical integration of the simulated data; again no weighting factors were used. Subunit association constants, $^xK_2$, were calculated from eq 6 (see Figure 2) and random error was added as a linear function of log $^xK_2$. An appropriate weighting factor was used such that the weighted variance of the errors was not a function of [X]. In general a weighting factor should be used if some of the experimental data sets are more precise than others.

Choice of Functional Form. Although only seven equilibrium constants are needed to describe the linkage between subunit association and ligand binding in Figure 2, the optimum choice of parameters is not obvious. We have explored the use of three such sets of independent parameters with varying degrees of success. The choice of which set to use may be dictated by the parameters one wishes most to determine (thus avoiding propagation of the confidence intervals), or by elimination of the parameters that can be evaluated independently.

(1) The first set of parameters studied consisted of the deoxy subunit association constant, $^0K_2$, the product Adair constants for dimer, K_{2i}, and those for tetramer, K_{4i}. The cross-correlation coefficients for these parameters are shown in Table I. In our experience correlation coefficients with absolute values less than 0.95 nearly always permit tractable resolution of parameters. Frequently more highly correlated parameters (up to 0.98–0.99) can be resolved, whereas values greater than 0.99 are nearly always intractable cases. For a given set of data the resolvability of a pair of parameters having a particular correlation coefficient increases if the total number of parameters to be estimated is lower. Thus, if some of the parameters can be estimated independently a higher correlation between the remaining ones can be tolerated. It can be seen that for oxygenation data the correlations between the subunit association constant, $^0K_2$, and the dimer Adair constants, K_{2i}, are very high and consequently they should be very difficult to resolve using this parameter set and range of data. Comparing Table IA with Table IB it is evident that data representing the subunit association constant, $^xK_2$, as a function of ligand concentration would provide better resolution since the parameters are not as highly correlated in that linkage function. Unfortunately, neither function alone is adequate. We have tried analyzing the simulated noise perturbed data using this set of seven parameters, and found the results to be so sensitive to the effects of small errors that no meaningful resolution of parameters could be achieved. This conclusion pertains to cases where the correct values are used as initial guesses. These essentially negative results will not be described further in this paper.

(2) The second set of parameters explored included the

TABLE II: Cross-Correlation Coefficients for Parameter Set 2.[a]

A. Oxygenation Data

k_{21}	1.000						
k_{22}	−0.909	1.000					
k_{41}	−0.745	0.943	1.000				
k_{42}	0.625	−0.853	−0.971	1.000			
k_{43}	−0.595	0.828	0.957	−0.998	1.000		
k_{44}	0.366	−0.606	−0.793	0.911	−0.932	1.000	
$^0K_2$	0.978	−0.834	−0.671	0.563	−0.536	0.329	1.000
	k_{21}	k_{22}	k_{41}	k_{42}	k_{43}	k_{44}	$^0K_2$

B. Subunit Association Data

k_{21}	1.000						
k_{22}	−0.917	1.000					
k_{41}	−0.506	0.572	1.000				
k_{42}	0.572	−0.580	−0.927	1.000			
k_{43}	−0.563	0.573	0.909	−0.998	1.000		
k_{44}	0.472	−0.524	−0.794	0.926	−0.947	1.000	
$^0K_2$	0.99996	−0.913	−0.504	0.571	−0.561	0.470	1.000
	k_{21}	k_{22}	k_{41}	k_{42}	k_{43}	k_{44}	$^0K_2$

[a] Data were simulated as in Figures 1 and 2.

TABLE III: Cross-Correlation Coefficients for Parameter Set 3.[a]

A. Oxygenation Data

$^0K_2$	1.000						
$^4K_2$	0.978	1.000					
K_{44}	0.344	0.479	1.000				
$\sqrt{k_{43}}$	−0.562	−0.598	−0.312	1.000			
K_{coop2}	0.9998	0.978	0.345	−0.564	1.000		
$^0K_2/^1K_2$	0.99990	0.978	0.343	−0.569	0.99995	1.000	
$^3K_2/^4K_2$	−0.782	−0.816	−0.462	0.879	−0.789	0.789	1.000
	$^0K_2$	$^4K_2$	K_{44}	$\sqrt{k_{43}}$	K_{coop2}	$^0K_2/^1K_2$	$^3K_2/^4K_2$

B. Subunit Association Data

$^0K_2$	1.000						
$^4K_2$	0.346	1.000					
K_{44}	−0.774	−0.566	1.000				
$\sqrt{k_{43}}$	−0.536	−0.854	0.852	1.000			
K_{coop2}	0.907	0.515	−0.960	−0.757	1.000		
$^0K_2/^1K_2$	0.701	0.715	−0.973	−0.948	0.907	1.000	
$^3K_2/^4K_2$	−0.407	−0.944	0.696	0.960	−0.610	−0.836	1.000
	$^0K_2$	$^4K_2$	K_{44}	$\sqrt{k_{43}}$	K_{coop2}	$^0K_2/^1K_2$	$^3K_2/^4K_2$

[a] Data were simulated as in Figures 1 and 2.

stepwise binding constants for dimer, k_{2i}, and tetramer, k_{4i}, and the unliganded subunit association constant, $^0K_2$. As can be seen in Table II, there exists a high correlation between $^0K_2$ and k_{2i} and also between the tetramer binding constants, k_{4i}. Consequently with this parameter set it is difficult to resolve the dimer binding constants k_{2i} from the subunit association constant $^0K_2$ and it is also difficult to resolve the tetramer stepwise binding constants, k_{4i}. These predictions, based upon the correlation properties, were borne out in practice. It was found that no reliable estimate of the seven parameters could be achieved using this parameter set with the specified ranges and accuracies of data.

Two points should be emphasized in relation to the attempts at resolution of parameters from simulated data described in 1 and 2 above. (a) The conclusions drawn embody a wide range of simulated cases (including, among others, all those described by Ackers and Halvorson, 1974) and are not restricted to the particular energy distribution of our model case. (b) Inability to estimate simultaneously a set of seven independent constants is not an intrinsic property of the linkage functions. Using perfect simulated data (i.e., with no random errors added) and a much wider range of values, we are able to estimate all the constants quite successfully using these methods.

Since neither $\bar{Y}$ vs. $[P_t]$ data nor $^xK_2$ vs. [X] data were found capable of yielding the quantities necessary to define completely the linkage scheme, at currently feasible levels of experimental accuracy, an obvious strategy is to estimate some of the parameters independently, and to introduce these values as fixed constants in estimating the remaining parameters from linkage data. This strategy is particularly promising if some of the most highly correlated parameters can be estimated independently. In addition, it places less severe demands upon the linkage data in that fewer parameters are estimated simultaneously. This is the strategy adopted in choosing parameter set 3, described below: first, identify those parameters which can be determined independently; second, choose those

TABLE IV: Confidence Profiles for Random Error Perturbed Synthetic Data Analyzed in Terms of Parameter Set 3.[a]

	$\bar{Y}$ Data	$^{x}K_2$ Data
$^{0}K_2$	−315.5 + 498.3	−53.5 + 107.0
$^{4}K_2$	−15.1 + 16.0	−4.7 + 4.1
K_{44}	−4.4 + 4.9	−38.9 + 21.7
$\sqrt{k_{43}}$	−50.9 + 83.6	−565.8 + 311.4
K_{coop2}	−70.4 + 129.3	−37.4 + 50.3
$^{0}K_2/^{1}K_2$	−145.0 + 254.7	−53.3 + 228.2
$^{3}K_2/^{4}K_2$	−29.6 + 63.6	−105.5 + 89.5

[a] Typical Confidence profiles, expressed as a percent. In this example the ligand binding data, $\bar{Y}$, were generated as in Figure 1. Pseudo-random errors were than added so that the variance was 1.0×10^{-4}. The subunit association data were calculated as in Figure 2. Pseudo-random errors were then added such that the fractional error in the particular data points is given by the formula $-0.55 + 0.1 \log {}^{x}K_2$. The $^{x}K_2$'s were then weighted with the inverse of values generated by this formula so that the weighted deviations would be constant.

TABLE V: Cross-Correlation Coefficients for Median Ligand Data as a Function of Hemoglobin Concentration.[a]

Hemoglobin Concn	Cross-Correlation Coefficients		
$[P_t]$	$^{4}K_2 - K_{22}$	$^{0}K_2 - K_{22}$	$^{4}K_2 - {}^{0}K_2$
1.0×10^{-9}	0.596	0.9997	0.614
1.0×10^{-8}	0.780	0.99995	0.785
1.0×10^{-7}	0.926	0.99999	0.927
1.0×10^{-6}	0.989	1.00000	0.989
5.0×10^{-6}	0.998	1.00000	0.998

[a] Calculated as a function of the lowest protein concentration at which the ligand binding curve was calculated.

TABLE VI: Correlation Matrices for $\bar{Y}$ Data Assuming Values for $^{0}K_2$, $^{4}K_2$, K_{22}, and K_{44}.[a]

K_{21}	1.000			
K_{41}	−0.248	1.000		
K_{42}	0.079	−0.942	1.000	
K_{43}	−0.231	0.913	−0.946	1.000
	K_{21}	K_{41}	K_{42}	K_{43}
$\sqrt{k_{43}}$	1.000			
K_{coop2}	−0.102	1.000		
$^{0}K_2/^{1}K_2$	−0.632	0.816	1.000	
$^{3}K_2/^{4}K_2$	0.852	−0.567	−0.895	1.000
	$\sqrt{k_{43}}$	K_{coop2}	$^{0}K_2/^{1}K_2$	$^{3}K_2/^{4}K_2$

[a] Data were simulated in Figure 1.

TABLE VII: Correlation Matrices of $^{x}K_2$ Data Assuming Values for $^{0}K_2$, $^{4}K_2$, K_{22}, and K_{44}.[a]

K_{21}	1.000			
K_{41}	0.759	1.000		
K_{42}	−0.620	−0.931	1.000	
K_{43}	0.508	0.778	−0.924	1.000
	K_{21}	K_{41}	K_{42}	K_{43}
$\sqrt{k_{43}}$	1.000			
K_{coop2}	0.613	1.000		
$^{0}K_2/^{1}K_2$	−0.928	−0.725	1.000	
$^{3}K_2/^{4}K_2$	0.932	0.477	−0.770	1.000
	$\sqrt{k_{43}}$	K_{coop2}	$^{0}K_2/^{1}K_2$	$^{3}K_2/^{4}K_2$

[a] Data were simulated as in Table I.

which are of greatest interest to a particular problem; and finally, add the ones needed to complete the energy scheme.

(3) We have recently developed a method for independent determination of the equilibrium constant, $^{0}K_2$, based upon kinetic measurements of the forward and reverse rates (Ip et al., 1976). Using this method a value of 5.11×10^{10} l. mol^{-1} dimer was determined for the particular hemoglobin system which corresponds to the case being presented here. The association constant $^{4}K_2$ for fully liganded hemoglobin in this system was also determined by analytical gel chromatography (cf. Ackers et al., 1975), yielding a value of 9.81×10^{5} l. mol^{-1} of dimer. The strategy chosen was to use these dissociation constants and estimate K_{44} from the median ligand concentration of oxygen saturation data (eq 13). In our system the oxygenation-linked subunit dissociation energy was 6.34 kcal (Ip et al., 1976; Mills et al., 1976). We are interested in determining how this energy is partitioned with binding of the first and last ligands to the tetramer. These values also define the sum of energy changes associated with the middle two ligand binding steps combined. The intersubunit contact energies of the middle steps are not individually resolvable from data of these kinds, even in principle (Ackers and Halvorson, 1974). The quantities of interest for the energetic partitioning can be represented by the ratios $^{0}K_2/^{1}K_2$ and $^{3}K_2/^{4}K_2$.

Another useful combination of equilibrium constants is formulated by noting that for a noncooperative dimer K_{22} is the square of the intrinsic binding constant and K_{21} is twice that intrinsic binding constant (because of the statistical factor for the number of binding sites). Consequently, a parameter representing cooperativity of the dimer can be defined by the equation:

$$K_{21} = 2K_{coop2}\sqrt{K_{22}} \tag{19}$$

One more parameter is needed to complete the energy balance. We arbitrarily chose $\sqrt{k_{43}}$. In recapitulation, the parameter set chosen consisted of: $^{0}K_2$, $^{4}K_2$, K_{44}, $\sqrt{k_{43}}$, K_{coop2}, $^{0}K_2/^{1}K_2$, $^{3}K_2/^{4}K_2$. Our ultimate strategy was to separately estimate the first three parameters. However, we first explored the correlation properties and analyses of simulated data using this third parameter set to estimate all seven parameters simulaneously. We found that the correlation properties of this parameter set were generally no more favorable than the other two sets for resolution of all seven parameters simultaneously. Table III shows the correlation for the third set of parameters. In this case the dimer cooperativity term K_{coop2} is highly correlated with the unliganded subunit association constant. It was again found that in the subunit association function, $^{x}K_2$, the parameters are generally not as highly correlated as in the ligand saturation function.

Table IV illustrates the precision with which we can expect to analyze experimental data for all of the parameters using the third parameter set. It is obvious that the confidence profiles leave much to be desired, and that accurate determination of the desired parameters is again not achievable using currently feasible experimental procedures. The high correlation between the dimer binding constants and the subunit association constants in each of the three parameter sets is a mani-

TABLE VIII: Typical Percentage Confidence Profiles for Estimating Parameters from Data, $\bar{Y}$, at Various Error Levels and Assuming the Values of $^4K_2$, $^0K_2$, and K_{44} Are Known.[a]

σ^2	9.00×10^{-7}	9.75×10^{-6}	9.49×10^{-5}	1.07×10^{-3}
Set 1				
K_{21}	−7.3 + 6.6	−20.3 + 26.0	−81.0 + 63.2	−157.5 + 358.3
K_{41}	−4.7 + 8.3	−21.7 + 19.8	−85.3 + 48.5	−246.3 + 192.2
K_{42}	−118.2 + 67.3	−281.8 + 308.6	−690.0 + 1214.1	−2735.7 + 3507.2
K_{43}	−12.1 + 21.2	−55.4 + 50.5	−217.8 + 123.9	−629.1 + 490.7
Set 3				
$\sqrt{k_{43}}$	−15.4 + 46.0	−43.2 + 137.1	−159.3 + ∞	− ∞ + ∞
K_{coop2}	−4.1 + 2.9	−6.4 + 5.8	−22.8 + 18.1	−64.7 + 62.7
$^0K_2/^1K_2$	−3.1 + 6.1	−12.1 + 10.9	−30.5 + 26.5	−63.3 + 99.9
$^3K_2/^4K_2$	−7.1 + 6.7	−16.3 + 21.9	−39.7 + 55.2	−111.1 + 164.8

[a] Data were simulated as in Figure 1. K_{22} is directly calculable from $^0K_2$, $^4K_2$, K_{44}. Confidence profiles correspond to one standard deviation.

TABLE IX: Typical Percentage Confidence Profile for Estimating Parameters from Subunit Association Data, $^xK_2$, at Various Error Levels Assuming the Values of $^4K_2$, $^0K_2$, and K_{44} Are Known.[a]

Relative Pseudo-Random Noise	0.05 Log $^xK_2 - 0.275$	0.1 Log $^xK_2 - 0.55$	0.2 Log $^xK_2 - 1.1$
Set 1			
K_{21}	−20.3 + 29.1	−35.9 + 63.2	−46.0 + 100.8
K_{41}	−165.8 + 237.9	−293.6 + 517.3	−376.6 + 824.9
K_{42}	−2593.3 + 1807.4	−5637.6 + 3199.7	−8990.1 + 4104.2
K_{43}	−163.2 + 234.1	−288.9 + 509.0	−379.6 + 811.7
Set 3			
$\sqrt{k_{43}}$	−∞ + ∞	−∞ + ∞	−∞ + ∞
K_{coop2}	−18.8 + 29.5	−33.8 + 61.7	−43.3 + 79.1
$^0K_2/^1K_2$	−38.1 + 45.4	−58.1 + 252.3	−51.9 + 88.6
$^3K_2/^4K_2$	−37.5 + 45.4	−83.0 + 75.8	−125.0 + 85.6

[a] Data were simulated as in Figure 2. K_{22} is directly calculable from the known values of $^0K_2$, $^4K_2$, and K_{44}. Confidence profiles correspond to one standard deviation.

festation of the fact that, within the concentration range where data can be collected, the concentration of unliganded dimer is insignificant. The alternative, as mentioned above, is to determine as many parameters as possible separately and then fix their values in the determination of the remaining ones.

The Median Ligand Concentration, $[\bar{X}]$. The problem can be simplified by evaluating the median ligand concentration $[\bar{X}]$ of each of the oxygen binding curves. This experimental quantity is a function of $^0K_2$, $^4K_2$, K_{22}, and K_{44}, only three of which are independent. For present purposes we chose K_{44} to be the dependent parameter.

We first explored the correlation properties of the $[\bar{X}]$ function (eq 13) in terms of the four constituent parameters. Table V gives the cross-correlation coefficients as a function of the lowest protein concentration for which ligand binding curves were simulated. It can be seen that near the lowest experimentally feasible concentrations, e.g., 10^{-7} M heme, the correlations between $^0K_2$ and $^4K_2$ or $^4K_2$ and K_{22} are quite tractable. It should be noted that the cross-correlation between $^0K_2$ and K_{22} would remain very high even if we could collect data at a hemoglobin concentration of 1.0×10^{-9} M (heme). It is again indicated that either $^0K_2$ or K_{22} must be measured independently.

As was previously noted, values of $^0K_2$ and $^4K_2$ are experimentally accessible (Ip et al., 1976). Using the independently determined values of these constants for our hemoglobin system, the determination of K_{22}, and consequently K_{44}, is found to be an easy single parameter fit to the function $[\bar{X}]$ vs. $[P_t]$. Consequently three of the seven parameters can be accurately determined and the seven parameter fitting problem is reduced to a four-parameter problem.

The Four-Parameter Problem is Tractable. Tables VI and VII show that, by determining the values K_{44}, K_{22}, $^0K_2$, and $^4K_2$ separately, the problem of the high correlation between the variables is essentially solved, with either parameter sets 1 or 3. The same is true of set 2 (not shown). Not only are the correlation coefficients lower, but their effect is greatly reduced by there being fewer parameters to be resolved from a particular data set.

The effects of random experimental error upon the feasibility of obtaining valid estimates of the four parameters is illustrated in Tables VIII and IX. For oxygenation data a variance of 10^{-5} is approximately what is currently achievable in the best experimental measurements. Table VIII illustrates that all of the quantities of interest can be found at this variance with reasonable precision. We have found that, if the value of K_{coop_2} is fixed at unity (noncooperative dimer case), the confidence profiles are considerably narrower. These results with simulated cases establish the validity of interpretations drawn from real experimental data in the accompanying paper (Mills et al., 1976).

Finally, it should be noted that in principle there is no need to vary hemoglobin concentration in order to estimate all other parameters of eq 5. Oxygenation data of exceedingly high

TABLE X: Effects of Random Experimental Errors Upon Estimation of Parameters from Oxygenation Data at a Single Hemoglobin Concentration (60 μM Heme) and Assuming the Values of ${}^0K_2$, K_{22}, and K_{44} to Be Known.[a]

	Estimated Value	% Confidence Limits	Correct Values
σ^2	9.86×10^{-6}		
K_{44}	1.707×10^{20}	$-2.8 + 2.9$	1.708×10^{20}
$\sqrt{K_{43}}$	366.56	$-32.8 + 48.4$	651.52
${}^0K_2/{}^1K_2$	84.24	$-11.1 + 9.9$	75.35
${}^3K_2/{}^4K_2$	0.908	$-22.8 + 33.7$	1.245
σ^2	1.08×10^{-4}		
K_{44}	1.718×10^{20}	$-9.1 + 8.6$	1.708×10^{20}
$\sqrt{K_{43}}$	286.42	$-49.6 + 90.6$	651.52
${}^0K_2/{}^1K_2$	83.56	$-31.5 + 33.4$	75.35
${}^3K_2/{}^4K_2$	0.785	$-67.9 + 84.9$	1.245
σ^2	9.70×10^{-4}		
K_{44}	1.736×10^{20}	$-22.9 + 27.6$	1.708×10^{20}
$\sqrt{K_{43}}$	2.173×10^{11}	$-\infty + \infty$	651.52
${}^0K_2/{}^1K_2$	58.33	$-40.6 + 103.9$	75.35
${}^3K_2/{}^4K_2$	1.160	$-42.9 + 202.4$	1.245

[a] Data were simulated for the same energy distribution as in Figures 1 and 2, but for oxygenation data at a single hemoglobin concentration.

precision at a single hemoglobin concentration (reflecting appreciable contributions from dimeric species) will suffice. Since many such data have been reported, it is of interest to explore the effects of random errors upon the four parameters derivable from these curves assuming that ${}^0K_2$, ${}^4K_2$, and K_{44} were independently determined. Results shown in Table X illustrate the results of analyses in terms of parameter set 3 for oxygenation curves simulated at a hemoglobin concentration of 60 μM (heme) as a function of the experimental precision. The variances shown correspond to random experimental errors assigned separately to three such simulated curves at the same hemoglobin concentration. The curves were then analyzed together for the four parameters. It can be seen that good estimation of the four parameters is quite feasible when the variance lies in the range between 10^{-5} and 10^{-4}, corresponding to the experimentally achievable range.

Discussion

In this study we have explored the problem of resolving experimentally derivable linkage functions into the constituent equilibrium constants necessary to define the intersubunit contact energy changes associated with cooperative ligand binding. The objectives were twofold: (a) to determine conditions and strategies whereby currently obtainable data on the two linkage functions can provide meaningful resolution; (b) to establish background information on the analyses necessary to interpretation of the experimental study described in the previous paper (Mills et al., 1976). Although we have used hemoglobin parameters of that experimental case for detailed presentation here, the results are typical of a much wider range of energy distributions that we have explored.

The first conclusion established by these studies is that the precision of currently available techniques for determining either oxygen binding data, $\bar{Y}$, or subunit association data, ${}^xK_2$, is not sufficient to resolve all seven independent parameters necessary to define the linkage scheme using either one of the linkage relationships. Data of either kind may nevertheless be useful in eliminating certain models of the energy distribution, and, with increased precision, greater success at resolution may be expected in the future. A much more promising and currently useful strategy, based upon these studies, is to combine experimental information both from direct dissociation studies and from ligand binding studies.

Secondly, it is important to estimate some of the parameters in an independent fashion, particularly those which are most highly correlated with the other parameters. It should be noted, however, that the separately estimated parameters must be determined to high accuracy. Small errors in their estimated values will be reflected as errors in the determination of the remaining parameters. Similarly, the confidence regions for the remaining parameters will not include the effect of variations in the fixed parameters. Because of this, the choice of parameters for least-squares minimization should be dictated as much as possible by the information which the experimenter finally wants to determine as well as by the above considerations. This eliminates the problem of propagating the confidence intervals of the determined parameters.

Results of this study by numerical analysis provide the necessary basis for interpretation of the experimental results described in the previous paper (Mills et al., 1976) on the concentration dependence of oxygenation curves, $\bar{Y}$ vs. $[P_t]$. The key to resolution of the seven requisite free energies of the linkage system in that study was obtaining an accurate independent estimate of the dimer–tetramer association constant ${}^0K_2$ for unliganded hemoglobin (Ip et al., 1976). This quantity is intrinsically the most difficult to resolve in either $\bar{Y}$ or ${}^xK_2$ data because of its high value (i.e., 10^{10} M^{-1}) and because of its high correlation with the dimer binding constants. Similarly, the association constant ${}^4K_2$ for fully oxygenated hemoglobin was determined independently by analytical gel chromatography. Then the measurements of accurate oxygenation curves permitted valid estimation of the remaining parameters using the procedures developed in this study. For data exhibiting only random deviations from the best fit, the variance was found to lie well within the limits for valid estimation of physically meaningful parameters.

Finally, it should be noted that variations in hemoglobin concentration are not required in order to resolve the four remaining parameters (assuming ${}^0K_2$, ${}^4K_2$, K_{44}) from currently practicable oxygenation data. As illustrated in Table X, a valid resolution is possible from data at a single protein concentra-

tion which reflects appreciable contributions from both dimeric and tetrameric species.

Acknowledgments

We thank the University of Virginia Computing Center for enormous amounts of time on the CDC 6400 computer.

References

Ackers, G. K., Brumbaugh, E. E., Ip, S. H. C., and Halvorson, H. R. (1976), *Biophys. Chem. 4,* 171-179.

Ackers, G. K., and Halvorson, H. R. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4312-4316.

Ackers, G. K., Johnson, M. L., Mills, F. C., Halvorson, H. R., and Shapiro, S. (1975), *Biochemistry 14,* 5128-5134.

Ackers, G. K., Johnson, M. L., Mills, F. C., and Ip, S. H. C. (1976), *Biochem. Biophys. Res. Commun. 69,* 135-142.

Ackers, G. K., and Thompson, T. E. (1965), *Proc. Natl. Acad. Sci. U.S.A. 53,* 342-349.

Anderson, N. M., Antonini, E., Burnori, E., and Wyman, J. (1970), *J. Mol. Biol. 47,* 205-213.

Box, G. E. P. (1960), *Ann. N.Y. Acad. Sci. 86,* 792-816.

Chiancone, E., Gilbert, L. M., Gilbert, G. A., and Kellett, G. L. (1968), *J. Biol. Chem. 243,* 1212-1219.

Crepeau, R. H., Hensley, C. P., Jr., and Edelstein, S. J. (1974), *Biochemistry 13,* 4860-4865.

Elbaum, D., and Herskovits, T. T. (1974), *Biochemistry 13,* 1268-1278.

Gilbert, G. A., Gilbert, L. M., and Shawky, N. A. F. (1975), *Biochem. Biophys. Acta 379,* 488-495.

Imai, K. (1973), *Biochemistry 12,* 789-808.

Imai, K., Morimoto, H., Kotani, M., Watari, H., Waka, H., and Juroda, M. (1970), *Biochim. Biophys. Acta 200,* 189-196.

Ip, S. H. C., Johnson, M. L., and Ackers, G. K. (1976), *Biochemistry 15,* 654-660.

Kellett, G. L. (1971), *Nature (London) 234,* 189-191.

Kellett, G. L., and Schachman, H. K. (1971), *J. Mol. Biol. 59,* 387-399.

Magar, M. (1972), Data Analysis in Biochemistry and Biophysics, New York, N.Y., Academic Press.

Mills, F. C., Johnson, M. L., and Ackers, G. K. (1976), *Biochemistry,* the preceding paper in this issue.

Noble, R. W. (1969), *J. Mol. Biol. 39,* 479-491.

Rosemeyer, M. A., and Huehns, E. R. (1967), *J. Mol. Biol. 25,* 253-273.

Soni, S. K. (1973), Ph.D. Thesis, University of Tennessee.

Thomas, J. O., and Edelstein, S. J. (1972), *J. Biol. Chem. 247,* 7870-7874.

Thomas, J. O., and Edelstein, S. J. (1973), *J. Biol. Chem. 249,* 2901-2905.

Weber, G. (1972), *Biochemistry 11,* 864-878.

Williams, R. C., Jr. (1976), *Biophys. Chem.* (in press).

Williams, R. C., Jr., and Kim, H. (1975), *Arch. Biochem. Biophys. 170,* 368-374.

Wyman, J. (1964), *Adv. Protein Chem. 19,* 223-289.

Zelen, M., and Severo, N. (1964), in NBS Handbook of Mathematical Functions, Abrumowitz, M., and Stegun, I., Ed., p 946.