Biophysical Chemistry 42 (1992) 271–281
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01637

Weighting functions and parameter resolvability for oxygenation data subject to error in the independent variable

Michael L. Doyle and Gary K. Ackers

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110 (USA)

(Received 25 June 1991; accepted in revised form 6 August 1991)

Abstract

Parameter resolvability and bias has been investigated for weighted nonlinear regression of data where the independent variable is subject to instrumental uncertainty. The specific example of cooperative oxygenation of hemoglobin was studied, where fractional saturation is determined spectrophotometrically and the oxygen activity is measured with a Clark polarographic electrode. For this system the instrumental uncertainty in the oxygen electrode was measured directly and the influence of the uncertainties on resolution of oxygen binding parameters was determined by Monte Carlo simulations. Four weighting functions were tested for their ability to minimize parameter uncertainty and bias: (1) uniform weighting; (2) "propagated weighting" whereby uncertainties in the independent variable are propagated into and added to uncertainties of the dependent variable; (3) Hill plot transform, or "end weighting"; and (4) maximum likelihood analysis, where deviations between fitting function and data are minimized as weighted horizontal and vertical distance vectors. Results of the Monte Carlo simulations favor the use of either uniform weighting, propagated weighting, or maximum likelihood weighting methods. Use of the Hill transform as a weighting function produced poorer parameter resolvability and inaccurate representation of the data in general. Bias error was negligible for all weighting functions.

Keywords: Weighted nonlinear regression; Parameter resolvability; Hemoglobin oxygenation; Oxygen activity measurements

1. Introduction

A widely-used general method for measuring oxygenation isotherms involves determining fractional saturation spectrophotometrically and oxygen activity with a Clark polarographic electrode [1,2]. The resulting data are then analyzed by nonlinear least-squares regression based on a fractional saturation fitting function, $\bar{Y}(X)$, which contains the physical parameters of the system being studied. Regression methods serve to adjust the values of the physical parameters in order to minimize the distances between data points and fitting function according to the familiar chi-squared function

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{\bar{Y}_i - \bar{Y}(X_i)}{\sigma_i} \right)^2 \tag{1}$$

Correspondence to: Dr. G.K. Ackers, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110 (USA).

Here $\bar{Y}_i$ is the ith observed value of the dependent variable and σ_i is the instrumental uncertainty for measuring $\bar{Y}_i$.

The weight assigned to each data point in eq. (1) is equal to $w_i = 1/\sigma_i^2$ and its functional dependence on oxygen activity is a subject of current debate [3–5]. Two divergent weighting functions are commonly used in the analysis of oxygenation isotherms obtained by these experimental techniques. Uniform weighting of all data points [2,6] where the instrumental error is regarded as constant for all $\bar{Y}_i$ values (i.e., as would arise solely from spectrophotometric random error). In contrast, and "end weighting" procedure has been employed [1] which treats the data points in the central region of the isotherm with 10,000 and 100,000 fold less weight than points at the extremes. The end weighting method is used to accommodate the view that the magnitude of σ_i in eq. (1) is dominated by uncertainty in measuring oxygen activity [4].

Which of these two weighting functions may be applicable for a given set of experimental results depends on the relationship between the random noise of the spectrophotometer and the random noise in the Clark electrode as it is propagated into the dependent variable. The purpose of this study was to measure directly the instrumental error in the spectrophotometric/Clark electrode isotherm system, and to assess the impact that various weighting procedures have upon parameter resolution and bias using data derived from these sources.

2. Methods

2.1 Oxygen electrode system

The activity of oxygen in solution, X_i, was monitored by a Beckman 39065 polarographic electrode [7] based on the following relationship.

$$X_i = X_\mathrm{T}\left[\frac{mv_i - mv_0}{mv_T - mv_0}\right] \tag{2}$$

The millivolt measurements mv_i, mv_0, and mv_T are made at X_i, zero oxygen activity, and the total oxygen activity, X_T, corresponding to the gas mixture with the highest partial pressure of oxygen. The value of X_T is calculated from the barometric pressure, water vapor pressure, percentage of oxygen in the gas mix, and solubility data [8].

The voltage signal was amplified with a Keithley model 150B microvolt ammeter using three orders of magnitude of input sensitivity. The ±1 volt output of the 150B was then digitized with a Nicolet model 3091 oscilloscope. A series *RC* circuit with a half-time of 0.15 second was connected to the output of the Keithley 150B as a high frequency filter. Temperature in the sample cell was regulated with a Lauda K-2/R water bath to a precision of ±0.02°C. The precision and accuracy of the temperature in the sample cell was measured with a National Bureau of Standards calibrated thermistor probe/Cole-Parmer 8502-20 digital thermometer. Solution conditions were chosen as those used in oxygenation experiments [2]: 0.1 *M* Tris, 0.1 *M* NaCl (0.18 *M* total chloride), 1m*M* Na_2EDTA, pH 7.40 and 21.50°C.

The noise properties of the oxygen electrode were characterized over a range of oxygen partial pressures. Partial pressures were obtained from Matheson certified standard gas mixtures of oxygen in nitrogen (certified to an accuracy of ±2% of the oxygen partial pressure).

2.2 Monte Carlo data synthesis

Noise-free data sets were generated for oxygen binding to cooperative tetrameric human hemoglobin according to the idealized fractional saturation function eq. (3) which assumes the absence of dissociation into dimers and that the fully oxygenated and deoxygenated asymptotic parameters are known exactly [1].

$$\bar{Y}(X) = \frac{\sum_{j=1}^{4} j\, \mathrm{e}^{-\Delta G_j/RT} X^j}{4\left(1 + \sum_{j=1}^{4} \mathrm{e}^{-\Delta G_j/RT} X^j\right)} \tag{3}$$

Here $\bar{Y}(X)$ is the fractional oxygen saturation at oxygen activity X, R is the gas constant, T is temperature and the Adair-constant free energies, which represent equilibria between the deoxygenated hemoglobin and the jth ligation state, are denoted by ΔG_j. The true values of the four Adair-constant free energies were taken from Chu et al. [2]. Noise-free data were generated as 150 points equally spaced in the logarithm of oxygen molarity from approximately -3.5 to -6.5.

Pseudo-random noise was superimposed on the perfect data by means of the Box–Muller transformation of uniform deviates [9]. Uniform deviates were obtained from Hewlett-Packard 9000/835 system calls and further randomized with the shuffling algorithm RANO as described by Press et al. [10]. The sequence of deviates was identical for all four weighting functions studied, i.e., they all started with the same seed. We assume that artifacts which might have arisen from less than perfect randomness in the sequence of deviates do not alter the general conclusions of this study. The pseudo-random noise was placed on each data point first on the dependent variable according to eq. (4)

$$\bar{Y}_i = \bar{Y}(X_i) + \phi_1\sigma_{\bar{Y}_i} \tag{4}$$

and then on the independent variable by eq. (5).

$$X_i = X_\mathrm{T}\left[\frac{(mv_i + \phi_2\sigma_{mv_i}) - mv_0}{mv_\mathrm{T} - mv_0}\right] \tag{5}$$

Here the ϕ terms are independent, Gaussian noise generators with mean values of zero and standard deviations of unity. The magnitude of the standard deviation term σ_{mv_i} was obtained from direct measurements of the instrumental uncertainty below. The value of $\sigma_{\bar{Y}_i}$ was taken from Myers et al. [3] as ± 0.0004, corresponding to the random noise in the dependent variable of an isotherm having a total absorbance change of 0.6 a.u. and measured with a high precision spectrophotometer in optimal condition.

The quantities X_T, mv_0 and mv_T in eq. (5) are measured only once per isotherm and therefore the errors in these measurements contribute systematically to the oxygenation isotherm (i.e., every data point in the isotherm is perturbed by the same amount). In general, systematic error factors can not be corrected by means of weighting functions. Nevertheless, the influence that these systematic error sources may have on parameter resolution and bias was also explored by extending eq. (5) to eq. (6).

$$X_i = X_\mathrm{T}\left[\frac{(mv_i + \phi_2\sigma_{mv_i}) - (mv_0 + \phi_3\sigma_{mv_0})}{(mv_\mathrm{T} + \phi_4\sigma_{mv_\mathrm{T}}) - (mv_0 + \phi_3\sigma_{mv_0})}\right] \tag{6}$$

Here individual calls to the noise generator ϕ_2 were made for every data point, while the noise generators ϕ_3 and ϕ_4 were called only once per isotherm. The random error in measuring X_T originates solely from reading a barometer and is therefore regarded as negligible in the present study (i.e., it contributes approximately 0.02% relative error in X_i, see Fig. 1 for comparison).

2.3 Monte Carlo nonlinear regression analysis

The synthetic data sets were fitted to eq. (3) by nonlinear regression for the four Adair-constant free energies. The Simplex algorithm [11] was used to minimize the generalized chi-squared function (eq. 1) until the convergence criterion of a change of less than 10^{-6} in chi-squared value was obtained. For each isotherm fitted the best-fit parameter values were stored and the resulting distributions of best-fit parameters evaluated after analysis of 1000 isotherms.

Four weighting procedures were compared in the present study. In the first case each data point was weighted uniformly so that all σ_i values in eq. (1) were equal. Uniform weighting is appropriate for experiments where the dominant source of random error originates uniformly from measurements of the dependent variable.

---

[1] We note that $\bar{Y}(X)$ values are calculated from the following three observables: the absorbance at oxygen activity X, and the fully oxygenated and deoxygenated asymptotic absorbancies. By standard error propagation theory and constant spectral noise it follows that $\bar{Y}(X)$ values in the asymptotic regions will have slightly more uncertainty than $\bar{Y}(X)$ values in the transition region.

In the second case an end weighting procedure was employed whereby the variance of each point is given by

$$\sigma_i^2 = \left[\bar{Y}_i(1-\bar{Y}_i)\right]^2 \tag{7}$$

The end weighting procedure is identical to transforming fractional saturations into Hill plot data ($\log[\bar{Y}_i/(1-\bar{Y}_i)]$) and then weighting uniformly. It is also similar in form to the derivative of eq. (3) with respect to ligand activity and is therefore a good approximation to a weighting function which would be applicable to experimental data where the dominant source of random error lies in measuring oxygen activity.

The third weighting procedure assumes that the variance at each point is comprised of a linear combination of instrumental uncertainties in both the dependent and independent variables as

$$\sigma_i^2 = \left(\frac{\partial \bar{Y}}{\partial X}\right)^2 (\sigma_{X_i})^2 + (\sigma_{\bar{Y}_i})^2 \tag{8}$$

Here it is assumed that uncertainty in the oxygen activity may be corrected by propagating it into the dependent variable and minimizing distances in the vertical direction only. It should be pointed out that there is no exact theory for this assumption in nonlinear cases and the results found in the present study with this "propagated weighting" procedure may not hold in general.

In a fourth data analysis method, "maximum likelihood" [12,13], instrumental error in both the dependent and independent variables is accounted for in a more direct manner. The maximum likelihood procedure minimizes weighted distance vector, D_i, between the data points and the fitting function. The maximum likelihood chi-squared function minimized is

$$\chi^2_{\mathrm{ml}} \equiv \sum_{i=1}^{N} \left(\frac{\bar{Y}_i - \bar{Y}(\bar{X}_i)}{\sigma_{\bar{Y}_i}}\right)^2 + \sum_{i=1}^{N} \left(\frac{X_i - \bar{X}_i}{\sigma_{X_i}}\right)^2$$

$$= \sum_{i=1}^{N} D_i^2 \tag{9}$$

where $\sigma_{\bar{Y}_i}$ and σ_{X_i} are the standard deviations in measuring the dependent and independent variables, and $\bar{X}_i$ is the optimal value of the independent variable which must be calculated at each data point and iteration. In the present study each $\bar{X}_i$ was estimated by searching the oxygen activity axis at a precision of 1% of the standard deviation in X_i (measured below, from eqs. 10 or 11) for the closest weighted vector distance indicated by eq. (9). (Changing the estimation precision of $\bar{X}_i$ to 2% of σ_{X_i} did not significantly change the results.)

The conclusions drawn from the present study are independent of initial parameter estimates, seed of the random number generator, magnitude of the convergence criterion (change in chi-squared), and number of simulations (several hundred trials sufficiently defines the distributions for the present purposes).

3. Results

3.1 Measurement of the random error in the oxygen activity of an isotherm

The standard deviation for measuring oxygen activity was measured by equilibrating a buffer sample in the oxygenation apparatus with known partial pressures of oxygen. The standard devia-

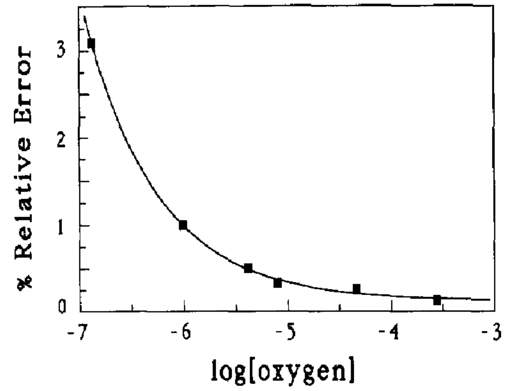

Fig. 1. Percent relative random error of the Beckman 39065 polarographic oxygen electrode versus the logarithm of the oxygen molarity. Solid curve is eq. (10).

tion was calculated as the square root of the variance for fitting the millivolt versus time data to an exponential response function. Figure 1 shows the relative random error (percent) in the oxygen electrode measured over a range of oxygen activities.

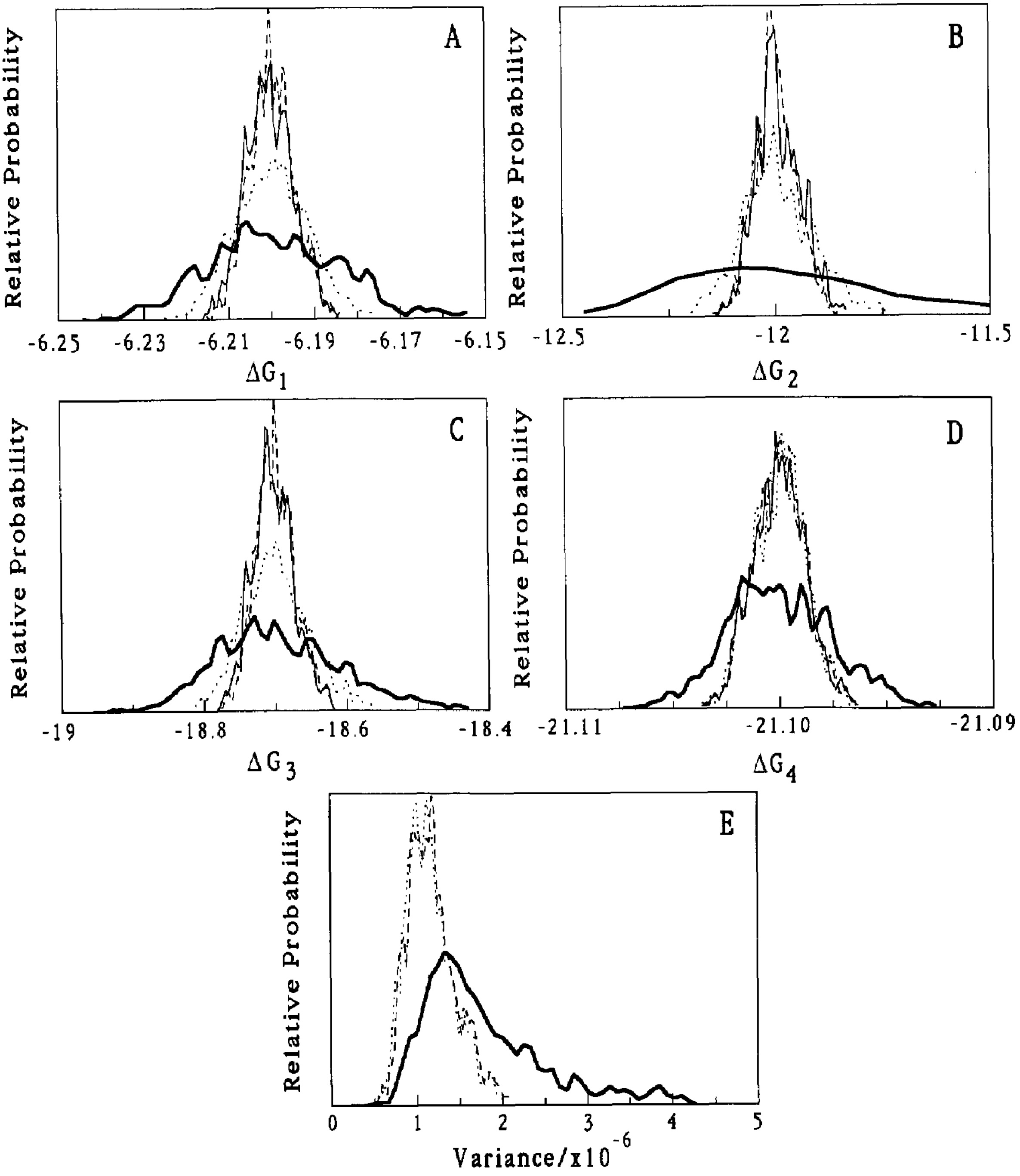

Fig. 2. Probability distributions for Monte Carlo estimation of parameters in eq. (3) for various weighting functions: maximum likelihood (thin solid curve), propagated (dashed curve), uniform (dotted curve), and Hill (thick solid curve). Random error was placed on both dependent and independent variables according to eqs. (4) and (5). True values of the Adair-constant free energies were taken from Chu et al. [2] as: -6.2 (A), -12 (B), -18.7 (C), and -27.1 (D) in units of kcal. Distributions of the variances of fit are shown in panel (E). The variances were calculated after convergence as the sum-of-squares (numerator eq. 1) divided by the degrees of freedom.

In order to model the random noise in the oxygen electrode as a function of oxygen molarity, the data in Fig. 1 were found to be adequately described by the following expression.

$$\%\sigma_{mv_i} = a\ e^{(b \log[O_2])} + c \qquad (10)$$

The constant values in eq. (10) obtained by nonlinear least-squares analysis of the data in Fig. 1 were: $a = 1.8\ 10^{-4}$, $b = -1.41$, and $c = 0.129$. (The absolute value of the electrode noise was therefore $\sigma_{mv_i} = mv_i \times \%\sigma_{mv_i}/100$.)

3.2 Resolvability of parameters when the oxygen activity is subject to random error

Synthetic data sets were generated according to eqs. (3), (4) and (5) using the standard error σ_{mv_i} as described by eq. (10). Analysis of 1000 isotherm data sets yielded the distributions of best-fit parameters shown in Fig. 2. The maximum likelihood and propagated weighting methods are found to be equally proficient in resolving parameters indicating that either method may be regarded as an optimal weighting function. The uniform weighting function is found to resolve parameters nearly as well as the optimal methods, while the end weighting function is approximately twice as poor as the uniform function, particularly for the intermediate species reactions (Table 1). The differences between the distribution medians and the true parameter values were insignificant in all cases, i.e., on the order of one percent of the standard deviation of the distributions or less.

Another measure of the appropriateness of a weighting function is derived from comparison of the distribution of variances of fit (Fig. 2E). The

Table 1

Resolvability and bias for parameter estimation using various weighting functions. Bias is listed at the top of each row. Lower and upper 68% confidence limits are in parenthesis. The two uncertainty models are described in the text. Values are in kilocalories

True values	Uniform	Hill transform	Propagated	Maximum likelihood
Random uncertainties				
−6.2	$-3\ 10^{-4}$ [a]	$1.2\ 10^{-4}$	$-5.5\ 10^{-5}$	$5.1\ 10^{-4}$
	(0.0093, 0.0081) [b]	(0.0143, 0.0167)	(0.0054, 0.0052)	(0.0056, 0.0057)
−12.0	-10^{-4}	-10^{-2}	$-2\ 10^{-4}$	$-3\ 10^{-3}$
	(0.076, 0.094)	(0.23, 0.34)	(0.048, 0.0050)	(0.047, 0.061)
−18.7	$9\ 10^{-4}$	$1\ 10^{-3}$	$1.8\ 10^{-4}$	$3.4\ 10^{-3}$
	(0.045, 0.049)	(0.080, 0.101)	(0.026, 0.026)	(0.030, 0.028)
−27.1	$2.6\ 10^{-5}$	$-3\ 10^{-5}$	$-3\ 10^{-5}$	$-2\ 10^{-5}$
	(0.0013, 0.0011)	(0.0022, 0.0026)	(0.0013, 0.0012)	(0.0013, 0.0012)
σ^2 [c]	$3.9\ 10^{-7}$	$6.1\ 10^{-7}$	$4.0\ 10^{-7}$	NA[c]
Systematic and random uncertainties				
−6.2	$-5.0\ 10^{-5}$	$-6.2\ 10^{-4}$	$-1.2\ 10^{-3}$	$2.5\ 10^{-3}$
	(0.039, 0.037)	(0.074, 0.074)	(0.048, 0.048)	(0.046, 0.047)
−12.0	$-6.0\ 10^{-3}$	$-1.7\ 10^{-2}$	$-4.8\ 10^{-4}$	$-4.4\ 10^{-2}$
	(0.096, 0.098)	(0.23, 0.37)	(0.062, 0.072)	(0.080, 0.038)
−18.7	$1.7\ 10^{-3}$	$1.7\ 10^{-3}$	$1.5\ 10^{-4}$	$1.1\ 10^{-2}$
	(0.049, 0.049)	(0.082, 0.095)	(0.047, 0.049)	(0.039, 0.040)
−27.1	$5.4\ 10^{-4}$	$-2.0\ 10^{-4}$	$5.1\ 10^{-4}$	$-1.5\ 10^{-5}$
	(0.024, 0.024)	(0.022, 0.022)	(0.023, 0.024)	(0.023, 0.023)
σ^2 [c]	$1.2\ 10^{-6}$	$2.6\ 10^{-6}$	$1.3\ 10^{-6}$	NA[c]

[a] Bias = median of distribution − true value.

[b] Confidence limits were determined by integration of the distributions in Figs. 2 and 4.

[c] Variances of fit are the mean values representing sum-of-squared deviations (numerator of eq. 1) divided by the degrees of freedom (146) for analysis of 1000 isotherms. The variance is not applicable for maximum likelihood for reasons given in the text.

similarity in the variance distributions of the propagated weight function and the uniform weight function demonstrates the equivalence of these functions in their description of the isotherm data. The significantly larger variances of fit and broader distribution seen with the end weighting function reflects the deemphasis on data points in the middle region of the isotherm and indicates the inappropriateness of this function. The distribution of variances is not shown for the maximum likelihood method since they represent entirely different distances (compare eqs. 1 and 9).

3.3 Random versus systematic error in the oxygen activity of an isotherm

A distinction must be made between random and systematic sources of uncertainty in measuring oxygen activity. The errors in measuring oxygen activity with an electrode will be independent only in situations where all four observables in eq. (2) are measured independently for every determination of X_i. In which case the resultant distribution of X_i values is expected to obey the following error program formula (neglecting cross terms).

$$\sigma_{Xi}^2 = \left(\frac{\partial X_i}{\partial mv_i}\right)^2 \sigma_{mv_i}^2 + \left(\frac{\partial X_i}{\partial mv_0}\right)^2 \sigma_{mv_0}^2 + \left(\frac{\partial X_i}{\partial mv_\mathrm{T}}\right)^2 \sigma_{mv_\mathrm{T}}^2 + \left(\frac{\partial X_i}{\partial X_\mathrm{T}}\right)^2 \sigma_{X_\mathrm{T}}^2 \qquad (11)$$

Figure 3 depicts the standard error in measuring X_i as a function of the logarithm of the oxygen molarity (or activity). Here all four observables in eq. (6) were measured independently for data points located at a specific oxygen molarity, while some correlation exists between sets of data points determined at different oxygen molarities as a result of the use of common measurements of mv_0 and mv_T. The dashed lines portray the random error in measuring oxygen activity according to eq. (11) using $\sigma_{mv_0} = 0.003$ mV and $\sigma_{mv_\mathrm{T}} = 0.00129\ mv_\mathrm{T}$. The magnitude of σ_{mv_0} is estimated from many measurements of mv_0 made in the course of oxygenation studies in our laboratory over a period of 15 years and σ_{mv_T} is determined from the data in Fig. 1 at high concentrations of oxygen molarity. The uncertainty in measuring X_T is much smaller than the other uncertainties (approximately 0.02% constant relative error on X_i, compare to Fig. 1) and σ_{X_T} is therefore set to zero in the present study.

The data in Fig. 3 also provide a test of the linearity of the oxygen electrode which is seen to be at least as good as the certified precision in the gas mixtures at the higher oxygen concentrations. Linearity at the lowest oxygen concentrations was difficult to assess due to the large amount of random error in the measurements (Fig. 3B), but is well within the expected error limits based on eq. (11).

In contrast to the random error of the measurements in Fig. 3, when measuring the oxygen activity values of an isotherm the only measure-

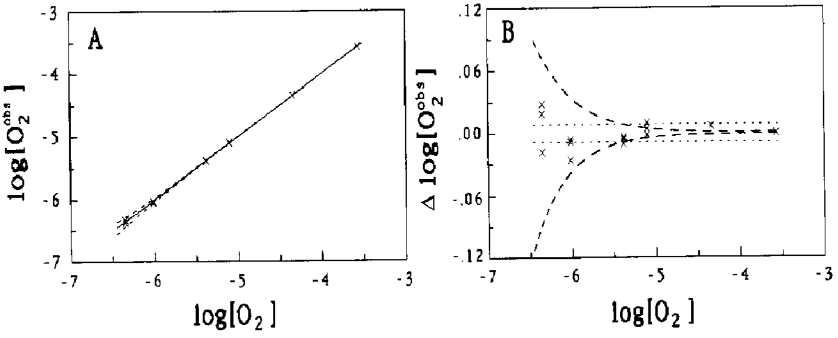

Fig. 3. (A) Logarithm of observed oxygen activity versus logarithm of expected oxygen molarity. Dashed curves delimit one standard deviation error limits in measuring oxygen activity (eq. 11). The straight solid line represents expected oxygen activity. Measured data points are indicated by ×. (B) Difference between observed and expected oxygen activities. Dashed curves delimit one standard (eq. 11) and dotted lines delimit uncertainty in the calibrating gas mixtures (±2% relative error in the certified gas mixtures). Measured values are indicated by ×.

ment in eq. (2) which is independent for all data points in mv_i. The other three experimental measurements (X_T, mv_T and mv_0) are made only once per isotherm and therefore represent systematic error sources. In general, systematic error sources in nonlinear least-squares data analysis

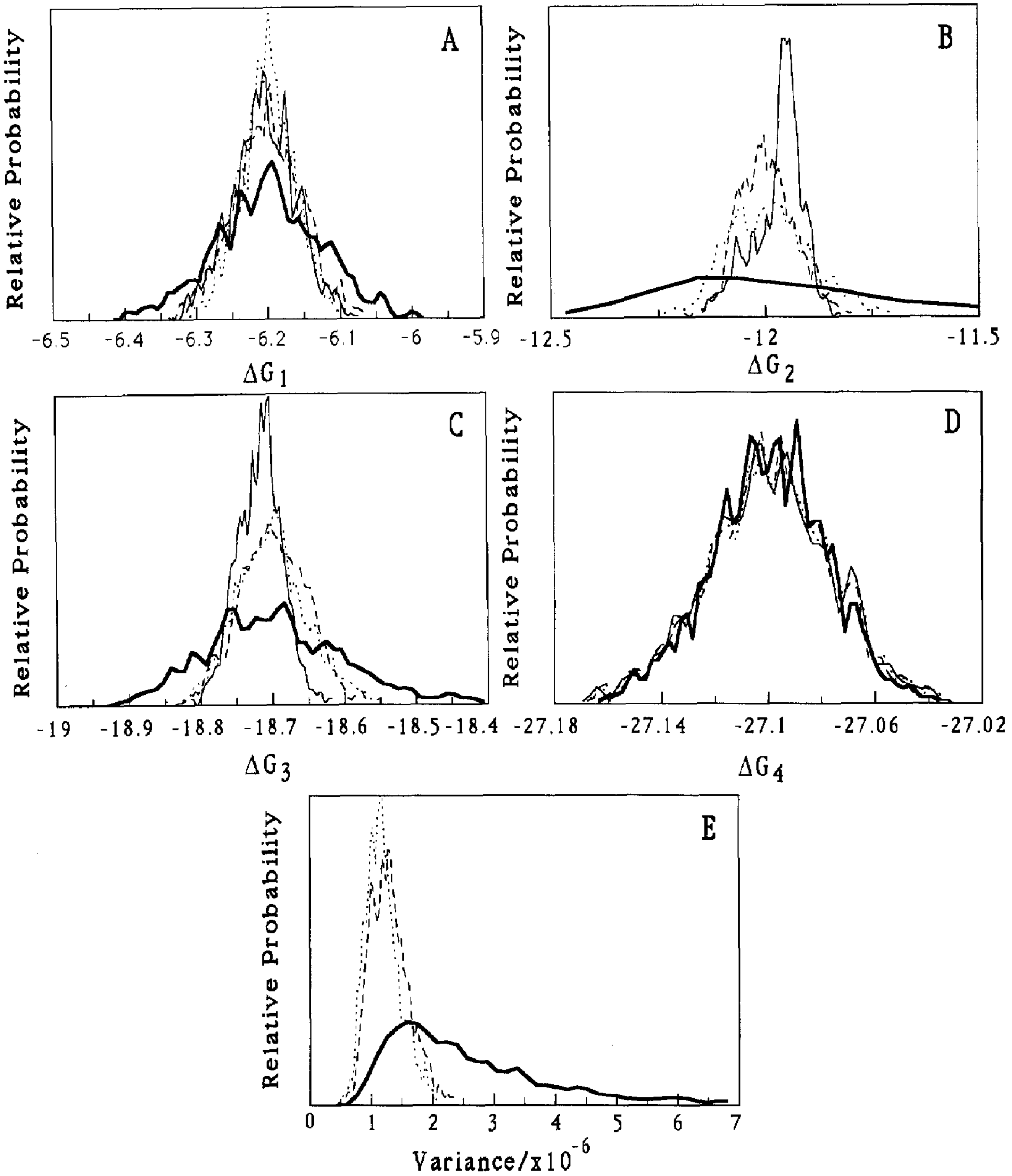

Fig. 4. Relative probability distributions for Monte Carlo estimation of parameters in eq. (3) for weighting functions described in the text: maximum likelihood (thin solid curve), propagated (dashed curve), uniform (dotted curve), and Hill (thick solid curve). Systematic and random error were placed on the data according to eqs. (4) and (6). True values of Adair-constant free energies were taken from Chu et al. [2] as: -6.2 (A), -12 (B), -18.7 (C), and -27.1 (D) in units of kcal. Distributions of the variances of fit are shown in panel (E).

cannot be corrected by the use of weighting functions.

3.4 Resolvability and bias in parameters when the oxygen activity contains random and systematic errors

To investigate the question of parameter resolution and bias due to both the random and systematic error sources described above, Monte Carlo studies were performed whereby the error on the data was governed by eqs. (4) and (6). The resulting distributions of best-fit parameters are shown in Fig. 4 for four different weighting functions. Due to the systematic nature of the simulated noise it is not possible to draw general conclusions. However, several pertinent observations can be made in the context of the specific true parameters and noise model studied. First, the impact of systematic error in the oxygen activity is largely to decrease parameter resolvability and not to generate significant bias error (Table 1). Second, the dependence of parameter resolvability on the weighting function used is unpredictable. For example, the uniform weighting function was found to resolve ΔG_1 better than the propagated weighting function (which weighted inversely to the actual systematic error). This lack of correspondence between known random and systematic error, weighting function, and parameter resolvability exemplifies the hazards in attempting to correct systematic uncertainties with weighting functions in nonlinear regression.

4. Discussion

4.1 Effect of weighting functions on parameter resolvability and bias

The major benefit of employing a weighting function to correct for differences in precision of the measured dependent variable is improved parameter resolvability. Bias error in parameter estimation by nonlinear least-squares regression is not a significant concern *per se* when the data are subject to random error only. After performing 1000 Monte Carlo "experiments" bias for all weighting functions studied becomes insignificant. However, for real experimental situations, where only one or a few isotherms are measured, the apparent bias (i.e., the difference between the true parameter and the best-fit parameter) is related to the width of the distributions in Fig. 2. The broader the distribution, the larger the apparent bias (although, correct assessment of confidence intervals should reflect larger variance in the best-fit parameter).

A more serious bias error situation occurs when the experimental data are perturbed by systematic error. For these situations one anticipates bias error in the best-fit parameters, even after many Monte Carlo simulations. The amount of bias can be assessed from the median of the best-fit parameter distribution. The amount of bias error found on any of the median values of the distributions in Fig. 4 was insignificant compared to the uncertainty (Table 1) indicating that the magnitude of the systematic uncertainties in eq. (6) are not problematic.

4.2 Implications for hemoglobin studies

The motivation for pursuing a Monte Carlo study with *simulated data* comes from *a priori* knowledge of the true parameters and true fitting function (model). The impact of various weighting functions can then be rigorously evaluated without complications regarding the physical meaning of the derived best-fit parameters and without the need for subjective criteria in assessing the virtues of weighting functions. The present study, using simulated data where the true model and parameters are known, demonstrates that different weighting procedures yield the same best-fit parameters but with different variances (Table 1). In contrast, recent comparison of weighting functions applied to real experimentally derived data showed that different weighting procedures produced different best-fit parameters [3]. The apparent discrepancy between the studies with real data and simulated data points to the existence of systematic error in the real experimental data, and strongly suggests that the problem of weighting function-dependent best-fit parameters in real data can be solved only by incorporation of a more accurate model.

A major source of systematic error which is frequently ignored but can be corrected by current methods [14] is the oxygen-linked dissociation of hemoglobin tetramers into dimers which occurs in typical experimental solution conditions. The substantial bias error which results when these effects are not corrected for has been reported [15,16]. Other systematic error sources such as the oxidation of ferrous heme sites to met hemoglobin (typically 1% or more [4]), or possible differences in optical properties of intermediate oxygenated states [17] may also be more consequential than the form of the weighting function used during regression analysis.

Nevertheless, certain weighting functions are more correct than others, and the benefits in terms of parameter resolvability indicate the need for using the most correct function. The uniform and end weighting functions represent model weighting functions corresponding to uncertainties which originate solely in the dependent and independent variables respectively. For the instrumental uncertainties reported here, the uniform weighting function is more proficient in resolving the intermediate oxygenation free energies than the end weighting function, indicating that the relative contribution of the uncertainty in the oxygen activity measurement to the total uncertainty is minimal. The similarity between the distributions of variances for the uniform and propagated weighting function fits and the large difference seen for the end weighting method also strongly implicate the use of a uniform weighting function.

An indirect approach for deducing a weighting function in the analysis of hemoglobin oxygenation isotherms has been reported whereby multiple isotherms were measured over the course of a year and the scatter in fractional saturation values evaluated [18]. Unfortunately this strategy is subject not only to random uncertainties in the spectrophotometer and oxygen electrode but also to several sources of systematic error (differential amounts of sample decomposition, dissociation into dimers, instrumental drift, inability of reproducing buffer pH to better than the certified precision of calibration buffers (±0.01 pH unit [2], etc.). Given the hypersensitive allosteric nature of hemoglobin to its environment and the general problems of sample decomposition, it is in practice more accurate and straightforward to measure instrumental uncertainties directly.

5. Conclusions

A general method has been presented for deducing the appropriate weighting function for nonlinear least-squares regression, and also for assessing the influence of instrumental uncertainties in both dependent and independent variables on parameter resolution and bias. Improved resolution is obtained when a weighting function is used which most accurately describes the random instrumental uncertainties, whereas bias error is insignificant and independent of the weighting function used. The dependence of parameter resolution on weighting function is unpredictable when the data are subject to systematic error.

For the case of hemoglobin oxygenation isotherms measured with spectrophotometric/Clark electrode methods, the noise in the oxygen activity measurement is found to be insignificant compared to the uncertainty in measuring fractional saturation with a high precision spectrophotometer. Consequently, a uniform weighting function is found to offer nearly optimal minimization of parameter uncertainties.

Acknowledgements

We would like to thank Enrico Di Cera for helpful suggestions and discussions. This work was supported by NIH grant R37-GM24486 and NSF grant DMB 9107244.

---

[2] For example, as a consequence of the Bohr effect [2], at pH 7.4 an uncertainty in pH as small as 0.01 of a pH unit would lead to an apparent uncertainty in relative oxygen activity of 1.3%.

References

1 K. Imai, Allosteric effects in haemoglobin (Cambridge University Press, Cambridge, 1982).

2 A.H. Chu, B.W. Turner and G.K. Ackers, Biochemistry 23 (1984) 604–617.

3 D. Myers, K. Imai and T. Yonetani, Biophys. Chem. 37 (1990) 323–340.

4 K. Imai, Biophys. Chem. 37 (1990) 197–210.

5 M.L. Johnson and A.E. Lassiter, Biophys. Chem. 38 (1990) 159–165.

6 F.C. Mills, M.L. Johnson and G.K. Ackers, Biochemistry 15 (1976) 5350–5362.

7 I. Fatt, Polarographic oxygen sensor: Its theory of operation and its application in biology, medicine, and technology (CRC Press, Cleveland, OH, 1976).

8 E. Wilhelm, R. Battino and R.J. Wilcock, Chem. Rev. 77 (1977) 219–259.

9 G.E.P. Box and M.E. Muller, Ann. Math. Stat. 29 (1958) 610–611.

10 W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical recipes: The art of scientific computing (Cambridge University Press, Cambridge, 1986).

11 J.A. Nelder and R. Mead, Computer J. 7 (1965) 308.

12 M.L. Johnson and S.G. Frasier, Methods Enzymol. 117 (1985) 301–342.

13 M.L. Johnson, Analytical Biochem. 148 (1985) 471–478.

14 B.W. Turner, D.W. Pettigrew and G.K. Ackers, Methods Enzymol. 76 (1981) 596–628.

15 M.L. Johnson and G.K. Ackers, Biophys. Chem. 7 (1977) 77.

16 M.L. Johnson and A.E. Lassiter, Biophys. Chem. 37 (1990) 231–238.

17 M.L. Doyle, E. Di Cera and S.J. Gill, Biochemistry 27 (1988) 820–824.

18 K. Imai, H. Morimoto, M. Kontani, H. Watari, W. Hirata and M. Kuroda, Biochim. Biophys. Acta 200 (1970) 820–824.