European Biophysics Journal
© Springer-Verlag 1991

Original papers

Variance reduction by simultaneous multi-exponential analysis of data sets from different experiments

K.-H. Müller and Th. Plesser

Max-Planck-Institut für Ernährungsphysiologie, Rheinlanddamm 201, W-4600 Dortmund 1, Federal Republic of Germany

Received June 27, 1990/Accepted in revised form December 18, 1990

Abstract. The analysis of experimental data from the photocycle of bacteriorhodopsin (bR) as sums of exponentials has accumulated a large amount of information on its kinetics which is still controversial. One reason for ambiguous results can be found in the inherent instabilities connected with the fitting of noisy data by sums of exponentials. Nevertheless, there are strategies to optimize the experiments and the data analysis by a proper combination of well known techniques. This paper describes an applicable approach based on the correct weighting of the data, a separation of the linear and the non-linear parameters in the process of the least squares approximation, and a statistical analysis applying the correlation matrix, the determinant of Fisher's information matrix, and the variance of the parameters as a measure of the reliability of the results. In addition, the confidence regions for the linear approximation of the non-linear model are compared with confidence regions for the true non-linear model. Evaluation techniques and rules for an optimum experimental design are mainly exemplified by the analysis of numerically generated model data with increasing complexity. The estimation of the number of exponentials significant for the interpretation of a given set of data is demonstrated by using records from eight absorption and photocurrent experiments on the photocycle of bacteriorhodopsin.

Key words: Multi-exponentials – Relaxation – Data analysis – Photocycle – Bacteriorhodopsin

Introduction

The decomposition of experimentally recorded time series by exponentials is a powerful technique for the analysis and evaluation of the underlaying relaxation processes which generate the observed signals. A principal drawback of the multi-exponential analysis is the non-orthogonality of exponential functions (Lanczos 1956) which is reflected in a high correlation of the time constants leading to an extreme sensitivity with respect to any systematic errors in the data generated, for example, by error statistics and by data truncation (Atkins 1974). McWhirter and Pike (1978) considered a measured time series as a convolution of a sum of delta functions with a resolution function. This approach attributed the decomposition of a signal to a deconvolution, i.e. an inverse Laplace transform in the case of exponentials. The authors found that the numerical inverse Laplace transform is totally unstable. From this, one must conclude that the numerical fitting of data by sums of exponentials has intrinsic difficulties. The problem may become particularly serious if data from different experiments and various experimental techniques are combined to form one data base for parameter estimation by a simultaneous multi-exponential analysis. The ambiguities arising from the inherent difficulties in the data approximation by sums of exponentials can be minimized if appropriate statistical tools are combined to optimize the data analysis and the design of the experiments.

In this paper we focus on experiments aiming to elucidate the mechanism of the photocycle of bacteriorhodopsin (bR) (Stoeckenius et al. 1979). The photocycle has been investigated very extensively by various experimental techniques, especially by light absorption measurements. These experiments covered a broad range of the visible spectrum (e.g. Lozier et al. 1975; Xie et al. 1987) and the infrared (Siebert et al. 1980; Siebert et al. 1982; Engelhard et al. 1985; Gerwert et al. 1990). Another set of data was produced by measurement of charge transport in oriented sheets of the purple membrane fixed in a polyacrylamide gel (Eisenbach et al. 1977; Dèr et al. 1985). The strategies of analysis derived for the photocycle are, however, also applicable to other systems of similar error structure.

This paper considers the problem from a methodological point of view and demonstrates the advantages of the developed method with examples of increasing complexity. A similar approach to that presented in this paper was

Offprint requests to: K.-H. Müller

used by Nagle et al. (1982) and Maurer et al. (1987) in their work on the analysis of the photocycle of bacteriorhodopsin. The advance in our work is made by the introduction of weighting factors and the use of the variance-covariance matrix for determination of the standard deviation of estimated parameters and their correlation. The calculation of true non-linear confidence regions emphazises the importance of a proper experimental design.

Techniques of optimum experimental design show that a simultaneous fit of all available data generates a significant gain of information compared to an analysis of one experiment at a time. The approach results in even better determined parameters if sophisticated methods are applied for the optimum design of each measurement in a set of experiments. For a general introduction into these techniques see Fedorov (1972). A biochemical example can be found in Markus and Plesser (1976; 1981) where the design of experiments for the determination of kinetic parameters of enzymes by progress curve analysis is demonstrated. However, there are no general statistical theories for this approach, especially for those models which include non-linear parameters.

The paper is organized into three parts. The next section explains the theoretical and methodological background of the approach and discusses the tools available in the literature which can be used to attack the problem of simultaneous non-linear analysis of time series measured in experiments using different apparatus and techniques. The second section is devoted to numerical model calculations which show step by step how the significant decrease of the parameter variance is achieved by accumulation of the appropriate data. In the last section the techniques learned from the model calculations are applied to measurements. Special attention is given to the number of exponentials significant for the interpretation of the data.

Methodological considerations

The most advanced concepts of parameter extraction from experimental data are available for models with linear relationships. The basic theorem (Himmelblau 1970; Fedorov 1972) for linear parameter estimation or regression analysis, formulates that the best linear unbiased estimator of the unknown parameter vector $\boldsymbol{\theta}$ is given by the dispersion matrix multiplied with the vector $\boldsymbol{Y}$ of the measured data. The dispersion matrix, the inverse of Fisher's information matrix, takes into account the error variance of each data item contained in $\boldsymbol{Y}$.

Parameters estimated from models with non-linear relationships $f(x, \boldsymbol{\theta})$ cannot be expressed by a closed formula as in the linear case. They must be calculated by sophisticated non-linear optimization techniques (Himmelblau 1970; Powell 1965, 1972). The standard optimization criterion for the parameter extraction from experimental data is the minimization of the sum of squares of the residuals

$$\sum_l w_l (y_l - f(x_l, \boldsymbol{\theta}))^2 = \text{Minimum}. \qquad (1)$$

In this paper we discuss only those functional relationships for which $x_l = t_l$ denotes an instant in time and $f(t_l, \boldsymbol{\theta})$ is composed of the sum of relaxing exponentials as given by the expression (2).

$$f(t_l) = \sum_i A_i \, e^{+k_i t_l} \qquad k_i \leq 0. \qquad (2)$$

The value $k_i = 0$ allows for the inclusion of a baseline. Special attention has to be given to the weights w_l of the measured data y_l in (1). They are crucial with respect to the reliability of the estimated parameter sets A_i and k_i. For independent data items, $w_l = 1/s_l^2$, where s_l^2 is the error variance of the particular measured datum l.

Since we are going to apply data sets from various experimental sources it is reasonable to group the data points according to the serial number m of the experiment. The estimation of parameters leads therefore to minimization of the expression

$$\sum_{m=1}^{M} \sum_{l=1}^{n_m} w_{l,m} \left(y_{l,m} - \sum_{i=1}^{N} A_{i,m} \, e^{+k_i t_{l,m}} \right)^2 = \text{Minimum} \qquad (3)$$

with respect to the linear parameters $A_{i,m}$ and the non-linear parameters k_i. The symbols are explained in the following list:

M	– number of recorded time series,
n_m	– number of data points in the time series m,
$y_{l,m}$	– data point l at time instant $t_{l,m}$ of the time series m,
$w_{l,m}$	– weight of data point l of the time series m,
$t_{l,m}$	– point l in time of the time series m,
N	– number of exponentials,
$A_{i,m}$	– amplitude related to the process i in time series m,
k_i	– rate constant of process i.

The number of the non-linear parameters $k_i (1 \leq i \leq N)$ is determined by the mechanism under investigation whereas the number of the linear parameters $A_{i,m} (0 \leq i \leq N, 1 \leq m \leq M)$ is $N+1$ times M and increases proportionally to the number of experiments included in the minimization procedure.

There are several strategies available to optimize the set of parameters. One approach is to apply a non-linear least-squares fitting algorithm to the total set of parameters. This is inefficient, takes a large amount of computer power and carries a high risk of ending in a local minimum. A much more efficient method is the separation of the linear and non-linear parameters (Golub and Pereyra 1973; Ruhe and Wedin 1980). It is applicable since the set of linear parameters is unique for a given set of rates and experimental data.

The variance-covariance and the correlation matrix of the non-linear parameters, the rates, are calculated when the optimization algorithm detects a minimum in the sum of squares. Normally the variances of the amplitudes are calculated from the linear minimization. The errors of these values, owing to the dispersion of the rates, are not significant for the applications in mind (see below). The matrices for the full set of parameters were only determined for some test cases, since matrix inversion by numerical methods is very time consuming for large matrices.

The weights $w_{l,m}$ in (3) are, as already mentioned, crucial for the reliability of the estimated parameters. They have to be determined from the error structure of each set of experiments. Special attention has to be given to error correlation as well as to truncation and digitization errors. In any case one has to look for a good estimate of the variance of each recorded data item. Examples can be found in the third section of this paper and in Müller et al. (1991).

The last unknown and very crucial quantity in (3) is the number N of exponentials required to fit the experimental data. The determination of the number of significant exponentials, – some-times called the recognition problem (Bergner et al. 1973) – has to proceed in an iterative manner starting with the lowest possible value of N which is still in accord with other information about the mechanism under investigation (Xie et al. 1987). Our discrimination process is based on three pieces of information, the residuals, the variances of the optimized parameters, and the correlation between the parameters. An exponential term is added as long as systematic deviations of the residuals from the zero mean are thereby reduced. The second criterion forces a significant reduction of the standard deviations of the parameters calculated from the residuals and the degree of freedom of the fit. Two other indicators are invoked if there is only a small reduction in standard deviation. Firstly, the value of at least one of the amplitudes introduced with the added exponential must differ with statistical significance from zero, i.e. its absolute value has to be at least three times as high as its estimated standard deviation (3 σ limit); and secondly, the new rate constant has to be appraised with respect to its correlation with any of the other rates. Further details are given in the examples of the section "Analysis of Experimental Data".

Estimators of the rates and the amplitudes, the baselines included, are obtained as point estimators from the overall non-linear least-squares fit. Approximate confidence regions for the parameters and other useful information can be derived from a linearization of the model in the vincinity of the minimum. The information is obtained from the variance-covariance and the correlation matrix in the same manner as for the linear models (Himmelblau 1970). The range of validity of the linearization is checked by comparison with levels of constant sums of squares (Beale 1960).

The routine VA05AD of the Harwell subroutine library (Harwell 1987) was applied for non-linear optimization and the routine MA44AD for the linear optimization which determines the amplitudes. The non-linear fitting was not performed with the rate constant k_i but with its logarithm $\ln(-k_i)$. The logarithmic transformation guarantees that the parameters to be optimized are of the same order of magnitude, a prerequisite for non-linear optimization by gradient methods. The standard deviation of a parameter calculated in this case at the optimum is a relative error (Clore and Chance 1978). Software packages for non-linear parameter optimization of arbitrary functions such as FACSIMILE/CHEKMAT (1987) in combination with HOWGOOD (1989) apply statistical tests at a similar level to those discussed here.

Numerical model calculations

The efficiency of the methods considered in the preceeding section will be demonstrated with artificial data sets generated from models of increasing complexity.

The two exponential model is the most simple example, except for the trivial case of one exponential. We consider the function

$$f(t)=A_1\,e^{+k_1 t}+A_2\,e^{+k_2 t} \tag{4}$$

with

$$A_1=-0.2 \quad A_2=+0.8 \quad k_1=1.0 \quad k_2=-3.0.$$

The calculations were performed with 251 data items equally distributed over a five second time interval and with an error of standard deviation 0.01. This example allows for instructive analysis of the confidence regions by formulas and graphical means.

The boundary of a confidence region with the significance level $1-\alpha$ is given by the sum of squares

$$\phi_{1-\alpha}=\phi_{\min}\left(1+\frac{p}{n-p}F_{1-\alpha}(p,n-p)\right) \tag{5}$$

$\phi_{\min}$ denotes the sum of squares in the minimum and $F_{1-\alpha}(p,\,n-p)$ is the value of the variance-ratio (Fisher-) distribution with p and $n-p$ degrees of freedom. As usual n denotes the number of data points, and p is the number of parameters.

Stange (1971) pointed out that the confidence regions form ellipses for the linearized model as expressed by (6).

$$\left(\frac{k_1-\hat{k}_1}{\hat{\sigma}_1}\right)^2-2\,c\hat{o}r(k_1,k_2)\left(\frac{k_1-\hat{k}_1}{\hat{\sigma}_1}\right)\left(\frac{k_2-\hat{k}_2}{\hat{\sigma}_2}\right)+\left(\frac{k_2-\hat{k}_2}{\hat{\sigma}_2}\right)^2 =2\ln(1/\alpha) \tag{6}$$

The label ^ denotes the estimated standard deviation of a parameter k_i and $c\hat{o}r(k_1, k_2)$ is the estimated correlation between k_1 and k_2. The estimated standard deviations $\hat{\sigma}_1$ and $\hat{\sigma}_2$ and the correlations are computed from the inverse Hessian matrix of the model, including the amplitudes as parameters. The elements of the Hessian are given by

$$h_{i,j}=\frac{1}{2}\frac{\partial^2\phi}{\partial\theta_i\,\partial\theta_j}=\sigma^2\sum_l\frac{\partial f(t_l)}{\partial\theta_i}\,\frac{\partial f(t_l)}{\partial\theta_j}. \tag{7}$$

The partial derivatives of the function $f(t)$ at the optimum, which is known in the model calculations, have been introduced in (7) as analytical expressions derived from (1) and (2), respectively.

Formula (5) is only an approximation of the sum of squares level $1-\alpha$ in the non-linear case and no explicit expressions for the calculation of the boundary of the confidence regions are available. The following algorithm was developed for the calculation of contour lines of constant sum of squares allowing for a comparison of the confidence regions of the linearized model with those from the full non-linear model. Let $\phi_1^{(0)}(k_1^{(1)}, k_2^{(1)})$ and $\phi_2^{(0)}(k_1^{(2)}, k_2^{(2)})$ be two points in the parameter space with the Euclidean distance

$$0<d(\phi_1^{(0)}, \phi_2^{(0)})\le h_{\max} \quad \text{and} \quad \phi_1^{(0)}=\phi_2^{(0)}=\phi_{1-\alpha}.$$

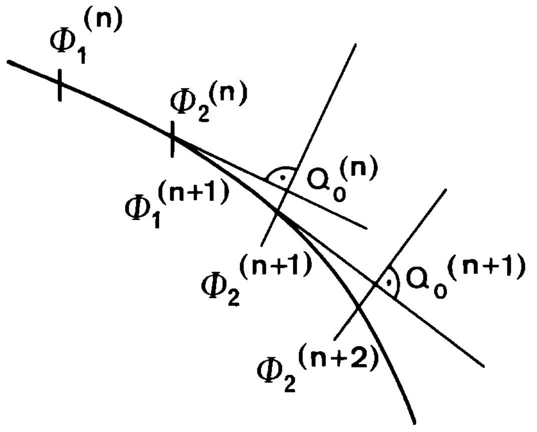

Fig. 1. Sketch of the geometry explaining the iterative algorithm for the determination of contour lines of constant sums of squares ϕ in the parameter space of two rate constantes. $\phi_1^{(n)}$ and $\phi_2^{(n)}$ determine a line along which the algorithm makes the next step of length h to find the point $Q_0^{(n)}$. A Newton algorithm starts at $Q_0^{(n)}$ to find the boundary $\phi=\phi_{1-\alpha}$ on a line perpendicular to the direction of the step

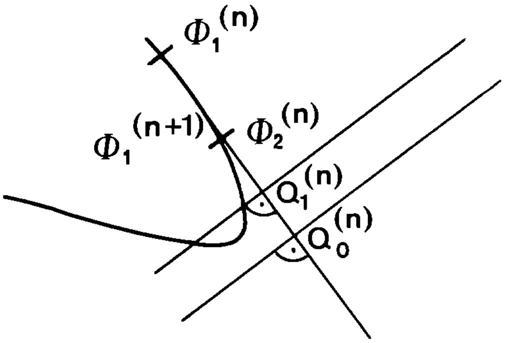

Fig. 2. Sketch of the geometry related to the algorithm for the determination of sum of squares contour lines $\phi=\phi_{1-\alpha}$ in the vincinity of a turning point. For further explanations see text and the legend of Fig. 1

The $h_{\max}$ value depends on the problem and must be chosen to compromise between computing speed and smoothness of the contour line. The first point of the contourline is found on the line from the point (k_1, k_2) to the origin by Newton's method, the second one is on a line parallel to the former one with a distance of $h_{\max}$. At the beginning, let $h=h_{\max}$. The algorithm then proceeds in the following manner (see Figs. 1 and 2):

1. Compute $Q_0^{(n)}$ on the line through the points $\phi_1^{(n)}$ and $\phi_2^{(n)}$ so that $d(Q_0^{(n)}, \phi_2^{(n)})=h$.

2. Find the point $\phi_2^{(n+1)}$ on the line through $Q_0^{(n)}$ orthogonal to the line through the points $\phi_1^{(n)}$ and $\phi_2^{(n)}$ so that $\phi_2^{(n+1)}=\phi_{1-\alpha}$ using Newton's method starting at $Q_0^{(n)}$.

3. If this point exists on the line, let $\phi_1^{(n+1)}=\phi_2^{(n)}$ and $h=\min(2h, h_{\max})$. Go to step 1 (Fig. 1).

4. If no point exists (Newton's algorithm diverges), set h to $h/2$ and goi to step 1 (Fig. 2).

5. The algorithm stops by the criterion $d(\phi_2^{(n)}, \phi_1^{(0)})\leq h_{\max}$.

The algorithm yields results, as long as the contour line has no edges.

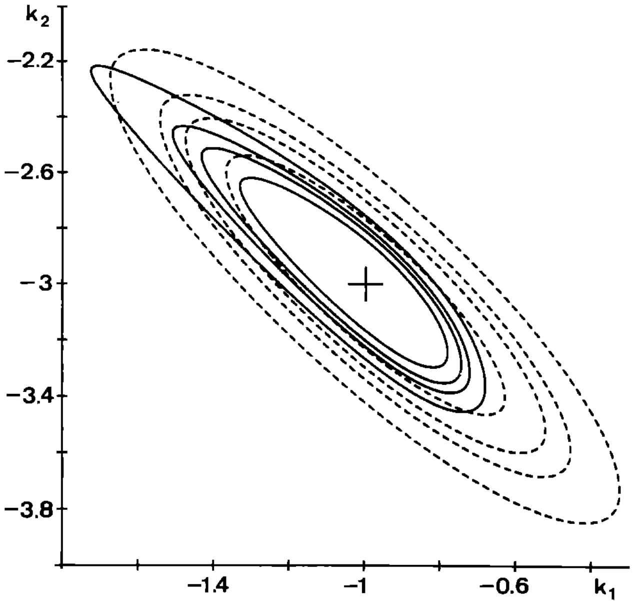

Fig. 3. Regions of confidence and contour lines for the function $f(t)=-0.2\,e^{-t}+0.8\,e^{-3t}$ calculated at 251 points evenly distributed over a 5 s time interval having a standard deviation of the error of 0.01. The dashed lines represent the confidence levels of 99%, 95%, 90%, and 75% (from the outside) for the linearized model. The solid lines display the contour lines of the non-linear model for the same (approximate) set of confidence levels. The cross marks the minimum sum of squares at $k_1=-1.0$ and $k_3=-3.0$

Figure 3 shows a comparison of the contour lines calculated with the covariance matrix (dashed lines) and with the non-linear model in the parameter space of (4). The confidence levels are from the outside 99%, 95%, 90%, and 75%, respectively. The linear approximation overestimates the errors for all shown confidence levels and gives by its very nature centrosymmetric confidence regions in contrast to the true regions.

In the next step we simulate two signals obtained from a mechanism characterized by two exponential terms

$$f_1(t)=A_{1,1}\,e^{+k_1t}+A_{2,1}\,e^{+k_2t} \qquad (8)$$

$$f_2(t)=A_{1,2}\,e^{+k_1t}+A_{2,2}\,e^{+k_2t}$$

$$A_{1,1}=-1.0 \quad A_{1,2}=1.0 \quad A_{2,1}=1.0 \quad A_{2,2}=1.0$$

$$k_1=-1.0 \quad k_2=-3.0.$$

The standard deviation of the data error is 0.01.

As expected from the previous results the separate evaluation of each equation shows a high correlation between the parameters k_1 and k_2 (Fig. 4). The ellipses are "orthogonal" to each other since the following relationships between amplitudes hold: $A_{1,1}=-A_{2,1}$ and $A_{1,2}=A_{2,2}$. The correlation is significantly reduced if the two data sets are evaluated simultaneously as shown by the inner circle (dashed lines: linear approximation; solid lines: non-linear model). In addition, the difference between the linear approximation and the non-linear model becomes negligible.

A critical case of (8) is considered in Fig. 5 A by choosing k_2 only 1.5 times larger than k_1. Even in this example, with very similar rate constants, a decoupling is possible by the evaluation of two "orthogonal" data sets. The

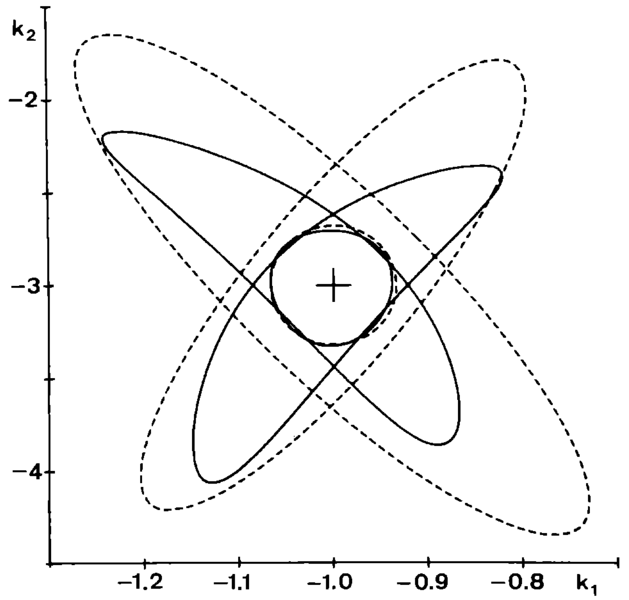

Fig. 4. Confidence regions (dashed) and contour lines (solid) for an "orthogonal" design of two experiments, see (8) in the text. The perpendicular ellipses display the results of a separate evaluation of each experiment. The inner circles are the boundaries for a simultaneous data evaluation of the two data sets. The cross marks the minimum sum of squares at $k_1 = -1.0$ and $k_2 = -3.0$

dashed line represents, as in Fig. 4, the 99% confidence level for the linear approximation and the solid line the same (approximate) level for the non-linear model.

The contours in Fig. 5B display the other extreme for the same rate constants as in Fig. 5A when the amplitude $A_{1,1}$ is changed from -1.0 to 1.0, i.e. (8) is reduced to the structure of (4). The size of the confidence regions increases by about one order of magnitude and their shape calculated by the linear approximation (dashed line) and the true non-linear model (solid line) overlap only slightly.

The rates and amplitudes for a data set with five "measurements" simulated by a two exponential model are compiled in Table 1. The number of data points is the same in each evaluation, i.e. one hundred per time series in an analysis of one "experiment" at a time and five times twenty for the simultaneous analysis. The constant number of evenly distributed data points retained in the various evaluations guarantees comparable sums of squares and discloses just that part of the additional information which is only related to the structure of the data and not to the size of the database. Figure 6 shows the corresponding 99% confidence regions. The lines marked with the digits 1 to 5 display the results of evaluations of one "measurement" at a time. It is obvious that, for example, for curves 3 or 5, respectively, one rate is estimated much better than the other. A significant reduction of the standard deviation of both rate constants is expected for a combined analysis of these two data sets. For quantitative figures see Table 4 below. The small ellipse around the point (k_1, k_2) marked with a cross, which is nearly the crossection of the individual confidence regions, is computed from the full data set. Its form shows that the two rate constants are still correlated.

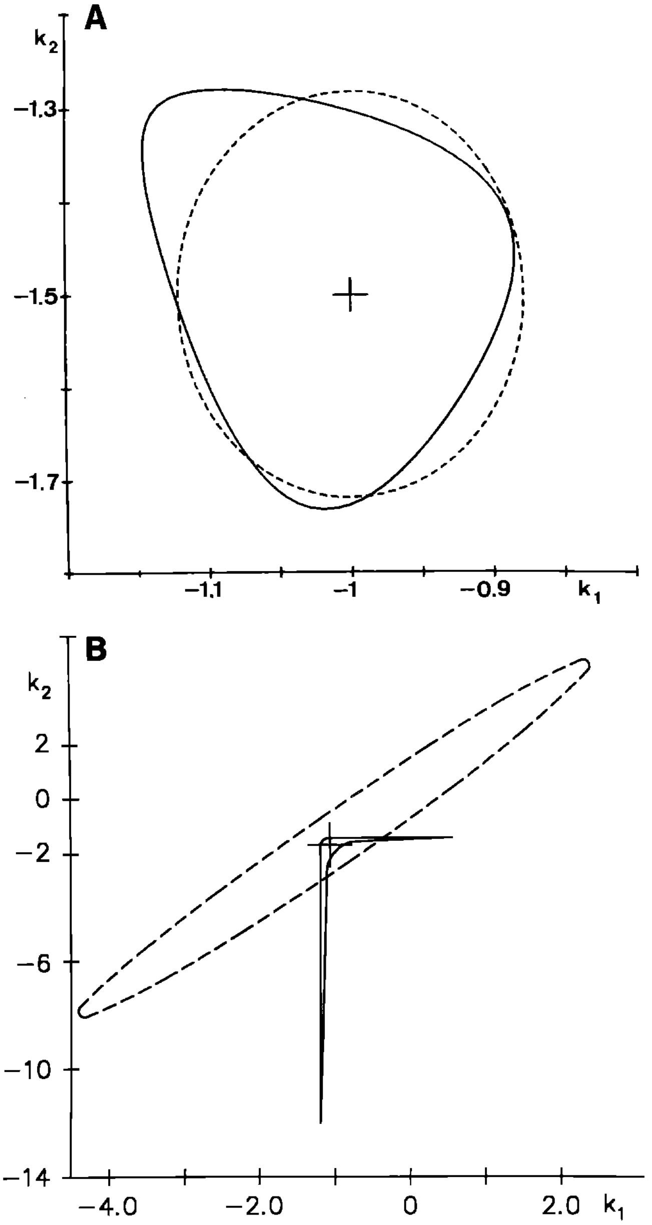

Fig. 5. **A** 99% confidence region (dashed) and contour line (solid) for $f_1(t) = -e^{-t} + e^{-1.5t}$ and $f_2(t) = +e^{-t} + e^{-1.5t}$, functions with a very small ratio of the rate constants. **B** 99% confidence region (dashed) and contour line (solid) only for $f_2(t)$. The cross marks the minimum sum of squares at $k_1 = -1.0$ and $k_2 = -1.5$ in both plots

Now we consider as the most complex model a description of the photocycle of bacteriorhodopsin in the purple membrane of halobacterium halobium (Stoeckenius et al. 1979). A set of six rate constants and their corresponding amplitudes, tabulated in Table 2, is used to model absorption experiments at five different wavelenths in the range of the K-bR transition.

The standard deviation of the data error was arbitrarily set to 1.0 per datum and the total number of 1280 data items was evenly distributed among the five simulated data traces. The logarithmic time scale in the simulations ranges from 0.01 µs to 90 ms and the weights w_i were chosen to simulate the linlog transient recorder as described in the next section and in Müller et al. (1991):

$$t_l = \begin{cases} 0 & l=1 \\ 0.01 \cdot 1.05^{(l-2)} & l>1 \end{cases} \tag{9}$$

and

$$w_l = w \cdot 2^{1+[(l-1)/20]}; \quad l=1, \dots, 256 \tag{10}$$

where

$$w = \begin{cases} 1 & \text{in Tables 2, 3, and 4 (column a)} \\ 1/n & \text{in Table 4 (column b)} \end{cases}$$

The square brackets [ ] (floor function) indicate the truncation to the largest integer contained in the argument and n is the number of different data sets. Table 3 displays for a separate non-linear fit of the kinetics of each wavelength the standard deviations as the percentage of the corresponding rate constants (Table 2) and the logarithm of the determinant of the Fisher information matrix. The variation in the information content is not significant as reflected by the data in the last column.

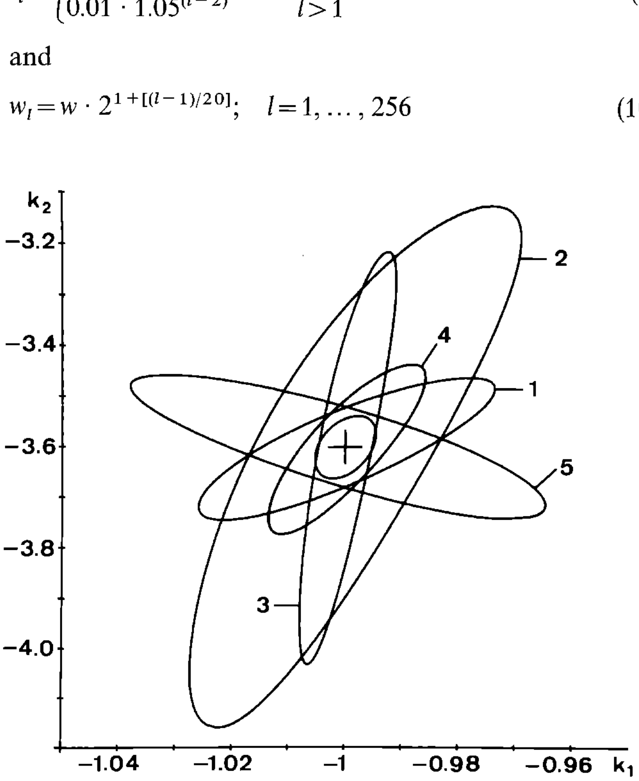

Fig. 6. The 99% non-linear contour lines shown for a set of simulated data which model the bacteriorhodopsin M-bR transition at five different wavelengths (1 to 5). The small ellipse, not marked with a number, is the 99% confidence region for the simultaneous evaluation of the "diluted" data set. The cross marks the true parameters $k_1 = -1.0$ and $k_2 = -3.6$

Table 1. Amplitudes and rate constants for a model with two exponentials. The corresponding graphs of the data analysis are shown in Fig. 6

i	A_{1i}	A_{2i}
1	2.81	3.35
2	−2.54	−0.92
3	−8.70	−1.16
4	−5.30	−2.80
5	2.01	−3.32
	$k_1 = -1.0$	$k_2 = -3.6$

It is impossible for this example to present the results of the variance analysis in easily understood pictorial form, so tables must be shown. Table 4 contains, in the columns labelled $i = 1$ to 6, the relative errors of the rate constants calculated for various combinations of data sets. The first row contains results obtained from a combined analysis of all data generated for the five wavelengths. Significant reductions of the standard deviations of the rates as compared to the standard deviations determined in the separate fits (Table 3) are obtained. An exception is the rate constant k_6. The largest decrease of the standard deviations of the rate constants in a combined analysis is generated by the step from one to two wavelengths. Each added wavelength, however, yields a smaller standard deviation. The last two columns show the increase of information by the successive accumulation of experiments. Column (a) displays the total information change due to the structure of the data set and the number of data items. In column (b) the information content is normalized per curve, i.e. the numbers reflect the information increase by the structure of the added data only.

Analysis of experimental data

In this section we apply the techniques learned in the model calculations to data sets obtainted from photocycle experiments on bR in the visible and infrared range and from recordings of the photocurrent in bR sheets.

The linlog transient recorder, used for all measurements described in Müller et al. (1991), records data in three segments in the log-mode, a linear pretrigger, a fast linear dwell time of 128 data points, and a logarithmic

Table 2. Amplitudes and rate constants for a model analysis with six exponentials. The amplitudes $A_i(\lambda)$ and the rates k_i are taken from an evaluation of photocycle data from bR. The labels in the first column under the heading λ/nm refer to the wavelength to which the set of amplitudes belongs

λ/nm	$A_1(\lambda)$	$A_2(\lambda)$	$A_3(\lambda)$	$A_4(\lambda)$	$A_5(\lambda)$	$A_6(\lambda)$
415	−1.50	−8.10	−2.50	3.50	6.00	0.40
570	−3.80	8.00	3.50	2.70	−17.20	−0.70
650	3.80	2.50	1.50	−7.70	5.00	0.25
600	6.40	4.80	−0.50	−4.30	−7.00	−0.40
500	−2.00	2.00	0.40	1.40	−4.00	−0.30
k_i	−0.5	-0.1×10^{-1}	-0.3×10^{-2}	-0.4×10^{-3}	-0.15×10^{-3}	-0.3×10^{-4}

Table 3. Standard deviations of the rate constants given as percentage of the rate. The last column contains the logarithm of the determinant of Fisher's information matrix

λ/nm	$i=1$ %	2 %	3 %	4 %	5 %	6 %	$\log_{10}(\det(\mathrm{I}))$
415	15	5	21	17	9	27	69.15
570	6	5	15	22	3	15	71.43
650	6	17	36	8	11	43	68.60
600	4	9	107	14	7	27	68.86
500	12	21	133	42	13	37	65.18

Table 4. Standard deviations of the rate constants given in percentage of the rate and the logarithm of the determinant of Fisher's information matrix. The first column with label λ/nm indicates which data sets have been evaluated together. The last two columns display the full information content (a) and the normalized information content (b). The latter reflects only the increase of information by the structure of the data

λ/nm	$i=1$ %	2 %	3 %	4 %	5 %	6 %	$\log_{10}(\det(\mathrm{I}))$	
							a)	b)
415, 570 650, 600 500	2.5	1.8	5	1.3	1.2	19	169.49	140.83
415, 570 650, 600	2.5	1.9	5	1.4	1.3	20	145.97	125.51
415, 570 650	3.5	2.6	8	1.7	1.5	23	121.58	108.70
415, 570	5	2.7	8	3.3	1.6	25	97.15	91.20
415	15	5	21	17	9	27	69.15	69.15

Table 5. Standard deviation of the residuals as a function of the number N of exponentials fitted to a set of light absorption data measured at 415, 500, 570, 600, and 650 nm. Column a: evaluation with theoretical weights; column b: evaluation with empirically corrected weights. The star indicates high correlations between some of the rate constants

N	$\hat{\sigma}^{\mathrm{vis}}$	
	a)	b)
2	12.05	10.60
3	2.05	1.49
4	1.68	1.37
5	1.34	1.06
6	1.26	0.99
7*	1.21	0.94
8*	1.24	0.96

segment. In the two linear segments the sweep time per point can be varied from 0.2 µs up to 2 s. This parameter was always set to 0.2 µs in the measurements described. The logarithmic part starts after the linear dwell time with a decade of 2 µs per point. In the second decade, two data points are averaged before being stored, in the third decade four data points are averaged, and so on. In the n^th^ decade $M=2^{n-1}$ data points are averaged so that the accuracy of the data gets better in each decade. The datum stored as an average of M points is only a true statistical representation of the physical signal if it changes linearly in time over the averaged time interval.

The apparatus makes no use of any sample statistics. Therefore some error statistics of the data in the logarithmic segment have to be done during the data analysis. Provided the errors are normally distributed, that is they are, for example, not a function of time and the error due to the piecewise linearization of the (unknown) physical signal is negligible compared to the measuring noise, then the following formula for the variances of the data in each decade of the logarithmic segment can be used:

$$\sigma_M^2=\frac{\sigma_1^2}{M} \quad \text{and} \quad w_M=\frac{M}{\sigma_1^2} \qquad (11)$$

where σ_1^2 is the variance of the data points before average, σ_M^2 is the variance of the mean of M successive averaged data values, and w_M is the corresponding weight of the mean.

An estimator $\hat{\sigma}_1^2$ for the variance of the data error σ_1^2, which is also used for the weights of the linear part of the record, can be calculated from the variance of the linear pretrigger baseline by standard methods.

Error structure of light absorption data

Light absorption changes were measured with the linlog recorder at 415, 500, 570, 600, and 650 nm, respectively. Data evalution was carried out as described in the following. The reliability of the estimated weights $\hat{w}_M$ by (11) for the logarithmic part was experimentally tested at those wavelengths which were used in the experiments. At each wavelength, multiple data traces were recorded without bleaching the purple membrane by a laser flash. The plateau value of each measurement was adjusted to zero by the least erroneous data values at times greater than ten seconds. The empirical weights $w_E(t)$ were calculated for each time point from the variance of the adjusted data sets. Comparison of the empirical weights with the w_M values calculated by (11) leads for all wavelength to the same approximate correction formula for the trend in the data error

$$w_E(t)\approx w_M(1-0.14\cdot\log_{10}(0.1\cdot t))^2. \qquad (12)$$

The time t is given in microseconds. The error in time was not taken into account because it can be neglected compared to the uncertainty of the correction formula. The correction formula was applied only to the logarithmic part of the recorded data.

The correction formula (12) indicates that either the pretrigger baseline is too short for a good estimate of the variance, or/and that the experimental setup consisting of measuring lamp, first monochromator, sample, second monochromator, photomultiplier, and preamplifier does not produce a normally distributed error. The error in the range of ten to one hundred milliseconds is underestimated only by a factor of about two and even less in earlier time regions as estimated from the ratio w_E/w_M.

Table 5 shows the behaviour of a multi-exponential fit with an increasing number of exponential terms. The left column (a) displays the standard deviation of the residuals when the weights are calculated according to (11). The right column (b) shows the same quantity after correction of the weights w_M with the empirically formula (12).

It can be seen that the numbers in Table 5 approach a nearly constant level which is close to unity in column (b) as expected from theoretical considerations and the definition of the weights w_i in (1).

For seven and eight exponential functions the standard deviations are smaller than in the 6-exponential fit, but the absolute value of the correlation coefficient between some rates in the millisecond range reaches more than 0.9, a reason to reject these solutions.

Error structure of IR data

A new stategy for getting estimators of the weights for data recorded in the IR has to be applied since the IR detector used produces random errors with an $1/f$ distribution, and is thus not normally distributed and in particular not time-independent as required for (11).

The error distribution over the full experimental time range can be estimated when the data of n repeated experiments are not averaged in the linlog recorder but stored separately, i.e. there are n data items available for each time instant t_i. These time series are first adjusted to zero by their least erroneous data values in the time range from 500 to 1 000 ms. In a second step the adjusted data are averaged, resulting in a mean value $\bar{y}_i$ for each time instant t_i. Now the original data traces are corrected by an individual factor C_k that takes into account long range laser energy fluctuations. The factors C_k are calculated by the least squares condition

$$\sum_{i=1}^{n} (\bar{y}_i - C_k\, y_{ik})^2 = \text{minimum} \tag{13}$$

leading to

$$C_k = \frac{\sum_{i=1}^{n} y_{ik}\, \bar{y}_i}{\sum_{i=1}^{n} \bar{y}_i^2}. \tag{14}$$

y_{ik} is the i^{th} data value from the k^{th} repeated measurement. Finally the empirical variances of the means are calculated, and the weights are estimated by

$$w_i = n(n-1) / \sum_{k=1}^{n} (\bar{y}_i - C_k\, y_{ik})^2. \tag{15}$$

The estimators for the weights are not unbiased since the original data were used twice, for the computation of the correction factors (as member of a whole data trace) and for the empirical data variances (as member of a set of repeated data). However, the error seems to be negligible. Table 6 displays the decrease of the standard deviation of the residuals with an increasing number of exponentials. $\hat{\sigma}$ levels off for $N=6$ similar to those for the optical absorption data. The 10% deviation of $\hat{\sigma}$ from the theoretical value 1.0 indicates that the correction procedure explained above cannot fully cope with the complicated error structure of the IR data.

Table 6. Standard deviation of the residuals as a function of the number N of exponentials fitted to a set of absorption data measured in the IR. The star indicates high correlations between some of the rate constants

N	2	3	4	5	6	7*	8*
$\hat{\sigma}^{\text{IR}}$	2.18	1.34	1.25	1.21	1.10	1.10	1.08

Table 7. Standard deviation of the residuals and relaxation times as a function of the number N of exponentials fitted to photocurrent data. The star indicates high correlation between τ_4 and τ_5. The relaxation times τ_i are given in microseconds

N	$\hat{\sigma}^{\text{el}}$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
2	20.2	47	1 225					
3	6.9	7	269	1 935				
4	3.7	3	58	296	2 130			
5	1.03	3	63	350	2 850	9 950		
6*	1.005	3	67	364	2 277	2 348	10 965	
7*	1.006	3	67	363	2 208	2 513	8 525	21 200

Error structure of photocurrent data

The error structure of data recorded from photocurrent experiments of the excited purple membrane differ from the former ones in some respects. Firstly, the signal varies over four orders of magnitude from the volt- to the millivolt range and is thus measured in two parts. Secondly, the (random) data error is small compared to the errors of optical absorption data and of unknown type. In addition, small systematic errors such as electrode drifts violate the presumption of normally distributed data errors, especially in the millivolt range. In the volt range limited 8-bit resolution of the fast A/D-converter introduces cut-off errors, which are not compensated by a baseline drift as in the millivolt range.

The same strategy as for the IR data was used for the estimation of the weights w_M for photocurrent data. It was observed, however, that the error is underestimated in the first part of the measurements (limitations of the A/D converter) and overestimated in the second part because the adjustment to zero contains errors due to non-repetitive electrode drifts.

Table 7 shows the standard deviation $\hat{\sigma}^{el}$ estimated from the residues of the fit for an increasing number of exponentials. A sharp step appears between $N=4$ and $N=5$. In the right part of the table the half-times of the rate constants belonging to the exponential fits are shown. In the case $N=6$ two rates, $\tau_4 = 2\,277$ µs and $\tau_5 = 2\,348$ µs, with a correlation coefficient $r_{4,5} = -1.00$ appear instead of $\tau_4 = 2\,850$ µs for $N=5$. Therefore five exponentials are sufficient to fit the data. The significant transition between $N=4$ and $N=5$ is also reflected in Fig. 7 where the residuals are plotted vs. time.

Now we are in the position to discuss the results of a combined analysis of data sets from light absorption and photocurrent experiments. Since for this purpose the total fit has to be compared with contributions from either of the two experiments, the goodness of fit is estimated by the

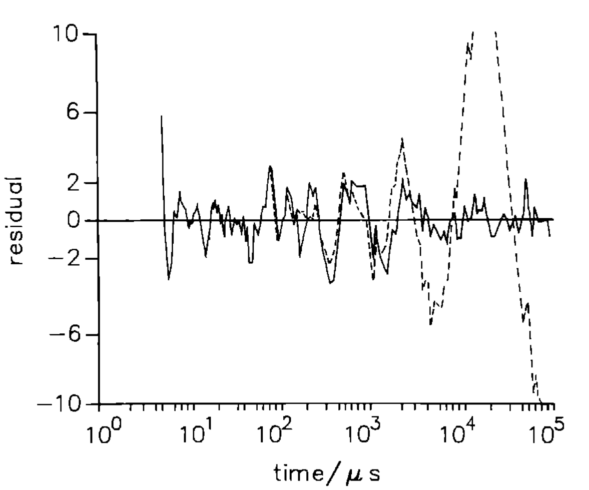

Fig. 7. Residuals from a fit of photocurrent data with four (dashed line) and five exponentials (solid line). The residuals for $N=4$ show significant fluctuations for times larger than a millisecond. The fifth exponential introduces a relaxation constant in the 10 ms range

Table 8. Sums of squared residuals (s.o.s.r.) as a function of the number N of exponentials fitted to photocurrent and light absorption data separately and simultaneously. Columns a: s.o.s.r. of the combined fit; columns b: s.o.s.r. of the separate fits

N	Combined	Photocurrent		Light absorption	
	s.o.s.r. · 10^3	s.o.s.r. · 10^3	s.o.s.r. · 10^3	s.o.s.r. · 10^3	s.o.s.r. · 10^3
		a)	b)	a)	b)
2	322	131	93	191	121
3	84	74	10.8	10.1	2.4
4	16	11.6	3.06	3.9	2.0
5	6.5	4.5	0.23	2.0	1.2
6	1.32	0.24	0.22	1.08	1.03
7*	1.22	0.215	0.215	1.005	0.92
8*	1.11	0.211	–	0.91	0.97

sum of squared residuals (s.o.s.r.) and not by its standard deviation since the latter depends on the degrees of freedom which are difficult to define for subsets of fitted data. Results are displayed in Table 8. The analysis of the combined data sets requires 6 exponentials for a good fit. For more than 6 exponentials a slight decrease of the s.o.r.s. is observed, coupled with high correlations among rate constants. The columns (a) under the headings photocurrent and light absorption show their contribution to the s.o.s.r. in the combined fit. The columns (b) display the same quantity as calculated by fits of the corresponding subsets of the data. It is obvious for the light absorption data that its contribution to the combined analysis, column (a), is comparable to the s.o.r.s. in the separate analysis, column (b), for $N=6$. The same consideration for the photocurrent data leads to the conclusion that the s.o.r.s. of the separate analysis, column (b), with $N=5$ compares to the s.o.r.s. for $N=6$ in the combined analysis.

Discussion

The paper explains in detail the considerations leading to an optimum design of those experiments which measure first order kinetics by various techniques. In addition, the evaluation of the minimum number of exponentials necessary for an interpretation of the data within the limits of the error statistics is demonstrated with experimental data obtained by three different techniques from bacteriorhodopsin and recorded with a linlog transient recorder. It turned out that the correct weighting of the data is of crucial importance. Sensitive criteria for a correct weighting of the data are given for the various experimental techniques. The applied statistical tools have to be slightly modified if modern linear transient recorders with large memories are used instead of linlog recorders.

Model calculations with data sets of increasing complexity demonstrate very clearly that the simultaneous evaluation of data recorded from various properly designed experiments significantly reduces the standard deviation of the estimated parameters and their correlation. This holds for both groups of parameters in a sum of exponentials, the non-linear ones, the relaxation constants, and the linear ones, the amplitudes. The amplitudes are of critical importance if the existence of kinetic processes are under discussion. In this case the non-zero expectation value of an amplitude must be proven. The detection of a zero amplitude may be caused by two reasons, a missing kinetic process in a certain time interval or a probe which is not sensitive to the specific process. In the latter case the simultaneous evaluation of data sets recorded with different experimental techniques is of special importance if some techniques report a process and others do not.

Another technical advantage of the software developed and presented in this paper is the option to calculate baselines. This property makes it feasible to adjust experiments performed independently at different intervals of time by a frame overlap. The technique allows to cover a very broad time range from nanoseconds to seconds in one parameter optimization run. The reliability of the estimated parameters can be inspected by the shape and the extension of contour lines calculated at constant levels of sum of squares.

The technique has been successfully applied to data sets recorded from bacteriorhodopsin by light absorption in the visible and in the infrared combined with data collected by photocurrent measurements (Müller et al. 1991). Other examples came from Fe^{3+} substituted bR (Engelhard et al. 1990), kinetic FTIR experiments with bR (Gerwert et al. 1991), and from studies of the sensory rhodopsin SR II (Scharf et al. 1990, unpublished work). The most critical step during the analysis of these examples was the precise determination of the statistical weighting factors of the data from multiple baseline experiments or from a large number of replicated experiments.

The combined analysis of well designed experiments with different probes led in each case to a reliable minimum set of amplitudes and rate constants. The decision, however, that these sets of numbers reflect the response of one and only one process in the measured sample must be based on other arguments. Furthermore, it must be emphasized that in a series of experiments the control parameters of the sample have to be regulated in such a manner that the process under investigation is always in the same state.

Appendix. The software package MEXFIT is written in standard FORTRAN 77 and is therefore fairly independent of the hardware. The package has about 4 000 lines of source code and needs 300 K bytes of main memory. It has been tested under the UNIX operating system on the workstations SUN 386i, SUN 4, Personal Iris 4D from Silicon Graphics, and on the minisuper-computer C2 series from CONVEX. Originally the software was developed on a minicomputer Perkin-Elmer 3230. The source code is available on request.

Acknowledgement. We thank Dr. M. Engelhard for clarifying discussions and Mrs. H. Kritzler for preparation of the manuscript.

References

Atkins GL (1974) Weighting functions and data truncation in the fitting of multi-exponential functions. Biochem J 175:125–127

Beale EML (1960) Confidence region in non-linear estimation. J R Stat Soc B22:41–75

Bergner PE Takeuchi K, Lui YY (1973) The recognition problem: application to sum of exponentials. Math Biosci 17:315–337

Clore GM, Chance EM (1978) The kinetics and thermodynamics of the reaction of solid-state fully reduced membrane-bound cytochrome oxidase with carbon monoxide as studied by dual-wavelength multichannel spectroscopy and flash photolysis. Biochem J 175:709–725

Dèr A, Hargittai P, Simon J (1985) Time-resolved photoelectric and absorption signals from oriented purple membrane imobilized in gel. J Biochem Biophys Methods 10:295–300

Eisenbach M, Weissmann C, Tanny G, Caplan SR (1977) Bacteriorhodopsin-loaded charged synthetic membranes. FEBS Lett 81:77–80

Engelhard M, Gerwert K, Hess B, Kreutz W, Siebert F (1985) Light-driven protonation changes of internal aspartic acids of bacteriorhodopsin: An investigation by static and time-resolved infrared difference spectroscopy using [4–13C] aspartic acid labeled purple membrane. Biochemistry 24:400–407

Engelhard M, Kohl KD, Müller KH, Hess B, Heidemeier J, Fischer M, Parak F (1990) The photocycle and the structure of iron containing bacteriorhodopsin – a kinetic and Mössbauer spectroscopy investigation. Eur Biophys J 19:11–18

FACSIMILE/CHEKMAT (1987) Harwell Laboratory, Oxfordshire AERE-Report 12805

Fedorov VV (1972) Theory of optimal experiments. Academic Press, New York London, pp 23–63

Gerwert K, Souvignier G, Hess B (1990) Light-induced protonation changes of protein-side-groups, chromophore-isomerization and backbone-motion of bacteriorhodopsin simultaneously monitored by time-resolved-FTIR-spectroscopy. Proc Nat Acad Sci, 89:9774–9778

Golub GH, Pereyra V (1973) The differentiation of pseudo inverses and nonlinear least-squares problems whose variable separate. SIAM J Numer Anal 10:412–432

Harwell Subroutine Library (1987) Harwell Laboratory Oxfordshire, England

Himmelblau DM (1970) Process analysis by statistical methods. Wiley, New York London Sidney, pp 176–207

HOWGOOD (1989) Harwell Laboratory, Oxfordshire AERE-Report 12864

Lanczos C (1956) Applied analysis. Prentice Hall, Englewood Cliffs, p 28

Lozier RH, Bogomolni RA, Stoeckenius W (1975) Bacteriorhodopsin: a light-driven proton pump in Halobacterium halobium. Biophys J 15:955–962

Markus M, Plesser Th (1976) Design and analysis of progress curves in enzyme kinetics. Biochem Soc Trans 4:361–364

Markus M, Plesser Th (1981) Progress curves in enzyme kinetics: Design and analysis of experiments. In: Endrenyi L (ed) Kinetic data analysis of enzyme and pharmacokinetic experiments. Plenum Press, New York pp 317–339

Maurer R, Vogel J, Schneider S (1987) Analysis of flash photolysis data by a global fit with multi exponentials. – I. determination of the minimal number of intermediates in the photocycle of bacteriorhodopsin by the 'stability criterion'. Photochem Photobiol 46:247–253

McWhirter JG, Pike ER (1978) On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kined. J Phys A 11:1729–1745

Müller KH, Butt HJ, Bamberg E, Fendler K, Hess B, Siebert F, Engelhard M (1991) The reaction cycle of bacteriorhodopsin: An analysis using visible absorption, photocurrent and infrared techniques. Eur Biophys J 19:241–251

Nagle JF, Parodi LA, Lozier RH (1982) Procedure for testing kinetic models of the photocycle of bacteriorhodopsin. Biophys J 38:161–174

Powell MJD (1965) A method for minimizing a sum of squares of non-linear functions without calculating derivatives. Comput J 7:303

Powell MJD (1972) Problems related to unconstrained optimization. In: Murray W (ed) Numerical methods of unconstrained optimization. Academic Press, New York London, p 29

Ruhe A, Wedin PA (1980) Algorithms for separable nonlinear least squares problems. SIAM Rev 22:318–337

Siebert F, Mäntele W, Kreutz W (1980) Flash-induced kinetic infrared spectroscopy applied to biochemical systems. Biophys Struct Mech 6:139–146

Siebert F, Mäntele W, Kreutz W (1982) Evidence for the protonation of two internal carboxylic groups during the photocycle of bacteriorhodopsin. FEBS Lett 141:82–87

Stange K (1971) Angewandte Statistik, vol II. Springer, Berlin Heidelberg New York

Stoeckenius W, Lozier RH, Bogomolni RA (1979) Bacteriorhodopsin and the purple membrane of Halobacteria. Biochim Biophys Acta 505:215–278

Xie AH, Nagle JF, Lozier RH (1987) Flash spectroscopy of purple membrane. Biophys J 51:627–635