Biophysical Chemistry, 34 (1989) 239–258
Elsevier

BIOCHE 01392

Point perturbation analysis of experimental data

Enrico Di Cera, Francesco Andreasi Bassi and Giuseppe Arcovito

Istituto di Fisica, Università Cattolica, Largo F. Vito 1, 00168 Roma, Italy

Received 24 April 1989
Revised manuscript received 26 July 1989
Accepted 28 July 1989

Data analysis; Critical phenomenon; Hemoglobin

A new method of data analysis is proposed. The method is based on discrete perturbation of experimental data points, which is used to probe the metric of the parameter hyperspace. Perturbation-induced fluctuations in the residual values are analysed by discrete Fourier transform to yield the autocorrelation function and a relaxation length for each experimental point. This parameter provides a quantitative measure of correlation and hence nonrandomness of residuals. The method is applied to the analysis of measurements of the shear viscosity of a 2,6-lutidine/water mixture near the critical point, and to the oxygen and carbon monoxide binding reactions to human hemoglobin. Relaxation profiles are constructed for several experimental data sets. Departure from random behavior in the residuals is discussed in connection with the theoretical interpretations of the phenomenon under consideration.

1. Introduction

Data analysis is a central issue in the description of any phenomenon in quantitative terms. Three basic steps in the approach to critical analysis of experimental data are in order: (i) linear or nonlinear regression to yield best-fit estimates of the parameters involved in the fitting function; (ii) determination of parameter error intervals; (iii) test of randomness of the residuals [1]. Phenomenological or mechanistic interpretations of experimental data rely on precise determination of the parameters involved in the model being proposed or tested. Best-fit parameters are estimated by minimizing a suitable function [1] which can be a linear or, more often, a nonlinear expression of the parameters depending upon the particular form of the fitting function. Errors on these estimates are extrapolated from the metric of the parameter space and, in the nonlinear case, by F-testing [2–4]. These first two steps usually receive most of the attention, since they yield the quantities of interest to the experimentalist and are supported by solid theoretical bases [1,4]. Tests on residuals, on the other hand, are seldom considered in standard data analysis, mostly because they stem from somewhat empirical grounds [1] and consequently lack a quantitative theoretical background, unlike the first two steps. It should also be noted that, although methods of data analysis are in principle 'equally partitioned' between fitting functions and experimental data, the fitting equation nevertheless is typically the focus of much attention, interested as we often are in the model and its reliability.

Quantitative evaluation of a given phenomenon cannot rule out, or even underestimate, tests for systematic errors. It is here that the strict interplay between fitting functions and experimental determinations becomes more evident and therefore deserves consideration. Systematic errors may be due either to the particular experimental technique used and/or to the functional form of the fitting

Correspondence address: E. Di Cera, Istituto di Fisica, Università Cattolica, Largo F. Vito 1, 00168 Roma, Italy.

function. When experimental data points are taken to be 'exact', systematic errors are indicative of failures in the model being considered. In other cases, one is faced with the opposite situation. As a matter of fact, the presence of systematic errors is assessed by analysis of residuals, the way they are distributed and whether or not they follow a Gaussian distribution [1–4]. It is well known, however, that such a procedure is only revealing of absolutely evident, or trivial, systematic errors and lacks the sensitivity required for detection of less evident errors [1]. Empirical methods aimed at checking the Gaussian nature of the residual distribution have also been proposed where the residuals are analysed in order to demonstrate possible departure from randomness. However, they are not very accurate and lack a sound theoretical basis [1]. Furthermore, none of these methods yields a quantitative measure of the degree of nonrandomness in the distribution of residuals, which could be used to characterize further the phenomenon under consideration.

In this paper, we wish to draw attention to a possible method of testing for systematic errors based on 'point perturbation analysis' (PPA) of experimental data. Perturbation of residuals is accomplished by dropping one experimental data point at a time, in order to yield a perturbation matrix which can be studied by standard techniques of Fourier analysis [5–7]. The method effectively probes the metric of the parameter hyperspace through the perturbation and provides a quantitative test for nonrandomness of residuals.

2. Theory

Before we begin, it is worth focussing on the matrix and vector notation used throughout this paper. The notation is similar to that used by Bard [1].

A matrix is denoted by a boldface capital, e.g., $\boldsymbol{A}$, and its elements by A_{ij}, while a boldface lower-case letter, e.g., $\boldsymbol{b}$, represents a column vector whose elements are b_i. $\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{-1}$ are the transpose and inverse of $\boldsymbol{A}$, respectively, while $\boldsymbol{b}^{\mathrm{T}}$ is the row vector with the same elements as $\boldsymbol{b}$. The summation Σ_i is assumed to span from 1 to n, unless otherwise specified.

We start from the basic function, Φ, which takes into account the n-dimensional vector of experimental data $\boldsymbol{y}$ and the fitting function $F(\boldsymbol{x},\boldsymbol{p})$. The fitting function interpolates the measurements $\boldsymbol{y}$ with the n-dimensional vector $\boldsymbol{x}$ of independent variables through the p-dimensional vector of parameters $\boldsymbol{p}$. The function Φ is equal to

$$\Phi = [\,\boldsymbol{y} - F(\boldsymbol{x},\boldsymbol{p})]^{\mathrm{T}}[\,\boldsymbol{y} - F(\boldsymbol{x},\boldsymbol{p})] = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = \Sigma_j r_j^2 \tag{1}$$

where summation is taken over the n experimental data points, and $\boldsymbol{r}$ is the n-dimensional vector of residuals. The function in eq. 1 is equal to the sum of squares of the residuals and is the form to be minimized when dealing with unweighted least squares. Weighted least squares, on the other hand, demands introduction of a diagonal matrix $\boldsymbol{W}$ such that

$$\Phi' = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{r} = \Sigma_j W_{jj} r_j^2 \tag{2}$$

which reduces to eq. 1 when all data points have the same weight. The vector $\boldsymbol{g} = \partial\Phi/\partial\boldsymbol{p}$ is the gradient of Φ, and the matrix of second-order partial derivatives $\boldsymbol{H} = \partial^2\Phi/\partial\boldsymbol{p}\partial\boldsymbol{p}$ is the Hessian of Φ. The existence of a true minimum for the function Φ implies that at the best-fit point in the parameter hyperspace, where $\boldsymbol{p} = \boldsymbol{p}^*$, one has $\boldsymbol{g}^* = 0$ and $\boldsymbol{H}^* > 0$. The Hessian is thus a positive definite form and the conditions above underlie the stability of point $\boldsymbol{p}^*$ in the parameter hyperspace with respect to small perturbations. The Hessian plays the role of the Gram matrix in the parameter hyperspace, so that information can be gained by probing its functional form. This can be accomplished by discrete point perturbations, which is the basis of the analysis described below.

Once the absolute minimum Φ^* of the function Φ has been determined we consider a particular data point, say y_k, and construct a diagonal matrix $\boldsymbol{W}_k$ such that $W_{kk} = 0$ and $W_{jj} = 1$ for any $j \neq k$. This is equivalent to dropping the particular data point y_k and hence to minimizing the function

$$\Phi_k' = \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{W}_k\boldsymbol{r}_k = \Sigma_{j\neq k} r_{kj}^2 \tag{3}$$

where the subscript k indicates that vectors and matrices are evaluated while dropping the data point k. The best-fit estimate $\Phi_k'^*$ contains all residuals except r_{kk}^*, i.e., the distance of point y_k relative to the fitting function F when the experimental point y_k is weighted zero. This residual can be added to $\Phi_k'^*$ so as to yield

$$\Phi_k^* = \Phi_k'^* + r_{kk}^{*2} = \Sigma_j r_{kj}^{*2} \tag{4}$$

and comparing eq. 4 with the best-fit value of Φ, Φ^*, obtained when all data points are included in the minimization, one necessarily has

$$\Phi^* \leqslant \Phi_k^* \tag{5}$$

because of the definition of Φ^* as a true minimum in the parameter space reached when fitting all data points. The inequality above is independent of k and expresses an important stability principle for the minimization procedure in the parameter hyperspace. It also implies that the residual of the point being dropped, r_{kk}^*, is always greater in absolute value than r_k^*, i.e., the residual obtained when the point is included in the minimization. One sees that the relevance of these stability conditions stems from the connection drawn between the metric of the parameter hyperspace and the discrete point perturbation introduced here.

Point perturbation provides a means of probing the metric of the parameter hyperspace. The change in the function Φ^* induced by point perturbation can be formulated in terms of a Taylor series as follows

$$\Phi_k^* = \Phi^* + \boldsymbol{g}^{*\mathrm{T}}(\boldsymbol{p}_k^* - \boldsymbol{p}^*) + (1/2)(\boldsymbol{p}_k^* - \boldsymbol{p}^*)^{\mathrm{T}}\boldsymbol{H}^*(\boldsymbol{p}_k^* - \boldsymbol{p}^*) + \ldots \tag{6}$$

and is a function of the 'unperturbed' gradient and Hessian obtained when fitting all data points. Here, $\boldsymbol{p}_k^*$ is the vector of best-fit parameter values obtained by dropping the data point k. Since the gradient vanishes at the minimum of the Φ hypersurface where $\boldsymbol{p} = \boldsymbol{p}^*$, then eq. 6 can conveniently be rewritten as

$$\delta\Phi_k^* \approx (1/2)\delta\boldsymbol{p}_k^{*\mathrm{T}}\boldsymbol{H}^*\delta\boldsymbol{p}_k^* \tag{7}$$

which shows that the perturbation of Φ^* can be expressed in terms of the perturbation of the best-fit parameter values through the Hessian. From eqs. 5 and 7 the positive definite nature of the matrix $\boldsymbol{H}^*$ follows immediately. The question of whether the fluctuations in the parameter values provide information on the metric of the parameter hyperspace can now be answered. From eq. 7 one has

$$2\delta\Phi_k^*\boldsymbol{V}^* \approx \delta\boldsymbol{p}_k^*\delta\boldsymbol{p}_k^{*\mathrm{T}} \tag{8}$$

where $\boldsymbol{V}$ denotes the covariance matrix obtained by inversion of $\boldsymbol{H}$ [1,4]. Consequently, for any couple of parameters i and j the following condition holds

$$2\delta\Phi_k^* V_{ij}^* \approx \delta p_{ki}^* \delta p_{kj}^* \tag{9}$$

where δp_{ki}^* is the change in the best-fit value of the parameter p_i when the data point k is dropped. The correlation between fluctuations of the best-fit parameter values is derived from eq. 9 by summing over all points and normalizing, i.e.,

$$r_{ij} = \Sigma_k \delta p_{ki}^* \delta p_{kj}^* / \sqrt{\Sigma_k \delta p_{ki}^{*2} \Sigma_k \delta p_{kj}^{*2}} \tag{10}$$

Hence,

$$r_{ij} \approx V_{ij}^* / \sqrt{V_{ii}^* V_{jj}^*} = R_{ij} \tag{11}$$

which shows that the correlation matrix obtained by point perturbation is, to a first approximation, identical with the correlation matrix derived from the covariance matrix $\boldsymbol{V}^* \approx \boldsymbol{H}^{*-1}$. The parameter changes can also be correlated to the fluctuations in the values of Φ. From eq. 7 one has

$$\partial\delta\Phi_k^*/\partial\delta\boldsymbol{p}_k^* \approx \boldsymbol{H}^*\delta\boldsymbol{p}_k^* \tag{12}$$

and assuming a parabolic shape for Φ near the minimum the partial derivatives on the left-hand side of eq. 12 can be approximated by linear expressions in $\delta\boldsymbol{p}_k^*$ for all k such as

$$\lambda\delta\boldsymbol{p}_k^* \approx \boldsymbol{H}^*\delta\boldsymbol{p}_k^* \tag{13}$$

and thus the λs are, to a first approximation, the eigenvalues of the Hessian from which linear estimates of the parameter errors can be derived [1–4]. Such an assumption, however, need not be made in general. In fact, the partial derivatives in

eq. 12 obtained by point perturbation can be used to construct cuts of the Φ hypersurface along various coordinates in the parameter hyperspace, in order to reveal eventually asymmetry of the parabolic branches of Φ. This is the situation often encountered in nonlinear parameter estimation [1–4]. However, for the purpose of the present paper it is important to focus on the theoretical bases of PPA.

The foregoing analysis has shown that discrete perturbations can effectively be used to probe the shape of the Φ hypersurface, at least to a first approximation, which guarantees the consistency of the method. With this result in mind, we proceed to address the major point of this paper, i.e., the problem of testing for systematic errors. The quantity of interest is the vector of residuals $\boldsymbol{r}$ constructed from the difference $\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x},\boldsymbol{p})$. Randomness of residuals at the best-fit point of the parameter hyperspace implies that $\boldsymbol{r}$ is a vector of random deviates. This may be difficult to assess, especially when dealing with a relatively small ($\approx$ 20–30) number of experimental determinations. Generally speaking, such a criterion seldom represents a strong enough condition for ruling out systematic errors and, more importantly, does not provide us with a quantitative measure of nonrandomness [1]. Such a measure would be most useful in data analysis, since it would represent an intrinsic property of the phenomenon being considered, just as other parameters like Φ^*. Consider the vector of best-fit residuals $\boldsymbol{r}^*$ and its change to $\boldsymbol{r}_k^*$ when dropping point k, which yields a perturbation vector $\delta\boldsymbol{r}_k^* = \boldsymbol{r}_k^* - \boldsymbol{r}^*$. The perturbation vector can be regarded as a displacement vector in the residual hyperspace around the 'equilibrium' position $\boldsymbol{r}^*$, as induced by point perturbation. Each element of the vector, say δr_{ki}^*, gives the projection of the perturbation due to dropping point k along the i-th axis of the manifold, i.e., the change in the distance of the i-th point from the fitting function when point k has been dropped. Consider now a given residual i and the n changes that it undergoes due to the n possible point perturbations. If the residuals are truly independent, as one expects in the absence of systematic errors, the n changes δr_{1i}^*, δr_{2i}^*, ... δr_{ni}^* are expected to be random. In other words, if the i-th equilibrium position r_i^* is perturbed by a 'stochastic force', such as a Gaussian white noise, there should be no correlation among displacements along the i-th axis of the residual hyperspace induced by point perturbation. This, of course, applies to all n axes. Therefore, one expects the n displacements along the i-th axis to be successive steps of a random walk in the absence of residual correlations, and hence systematic errors. Departure from random walk behavior along the i-th axis indicates that the i-th residual is responding nonrandomly to the n perturbations, and that a systematic error is present at the level of this residual. Point perturbations can thus be taken as successive steps of a walk in the residual hyperspace driven by an external force. This parallel allows one to analyse the perturbations within the framework of basic principles of nonequilibrium thermodynamics [8].

The displacement vector in the residual hyperspace can quantitatively be studied by Fourier transforming the displacements along each of the n axes separately. For each point k, one has

$$f_k(\tau) = \Sigma_j \delta r_{jk}^* e^{-i[2\pi(j-1)\tau]} \tag{14}$$

where f_k represents the discrete Fourier transform of the n changes in the k-th residual obtained by point perturbation, and the corresponding frequencies are $\tau = 0,\ 1/n,\ 2/n,\ \ldots\ (n-1)/n$ [5]. Euler's identity allows for an alternative expression of eq. 14 in terms of discrete cosine/sine coefficients, $A(\tau)$ and $B(\tau)$, so as to yield a representation of the displacements δrs as a sum of sinusoids with frequencies $\tau = 0,\ 1/n,\ 2/n,\ \ldots\ m/n$, where $m = n/2$ [5,6]. The cosine/sine coefficients are

$$A_k(0) = (1/n)\Sigma_j \delta r_{jk}^* \tag{15a}$$

$$A_k(i/n) = (2/n)\Sigma_j \delta r_{jk}^* \cos[2\pi i(j-1)/n]$$

$$i = 1,\ldots m-1 \tag{15b}$$

$$A_k(m/n) = (1/n)\Sigma_j (-1)^{j-1} \delta r_{jk}^* \tag{15c}$$

$$B_k(i/n) = (2/n)\Sigma_j \delta r_{jk}^* \sin[2\pi i(j-1)/n]$$

$$i = 0,\ldots m \tag{15d}$$

and it should be borne in mind that when n is odd

then $m=(n+1)/2$ and there is no frequency m/n [5,6]. The above coefficients are related to the fs as follows [5]

$$f_k(0)=A_k(0) \tag{16a}$$

$$f_k(j/n)=A_k(j/n)-iB_k(j/n)$$

$$j=1,\ldots m-1 \tag{16b}$$

$$f_k(m/n)=A_k(m/n) \tag{16c}$$

since $B_k(0)=B_k(m/n)=0$. The treatment given above includes only frequencies of the form j/n and hence the harmonics of the fundamental frequency $1/n$. Periodic components of these frequencies can easily be decoupled from noise and show up in the frequency spectrum as increased intensities. However, it is unlikely that a sine component with frequency f would match exactly any of the frequencies j/n, but in this case the frequency spectrum would show increased intensities of the fundamental frequencies around f. It is clear that for $n\to\infty$ the entire frequency spectrum can be obtained. For finite n, as in the cases encountered in practice, the entire spectrum is discretely sampled and empirical construction of the quantities of interest, namely, the power spectrum and the autocorrelation function, is obtained [5,6]. In particular, the Wiener-Khinchin theorem [8] allows derivation of the autocorrelation function directly from the power spectrum, once the Fourier transforms, fs, are known. The empirical power spectrum, or periodogram, for each point k is equal to [5]

$$I_k(\tau)=nf_k(\tau)f_k^{\circ}(\tau)/2$$

$$=n\left[A_k(\tau)^2+B_k(\tau)^2\right]/2 \tag{17}$$

where f° is the complex conjugate of f. Hence, the autocorrelation function $C(\alpha)$ for each point is given by the inverse Fourier transform of $I_k(\tau)$. Since the periodogram is an even function in the interval $\pm m/n$ (for n even) or $\pm(m-1)/n$ (for n odd), the autocorrelation function can be calculated as a cosine expansion in the same frequency range as follows [5]

$$C_k(\alpha)=\Sigma_j I_k(j/n)\cos(2\pi j\alpha/n)/\Sigma_j I_k(j/n) \tag{18}$$

where it should be noted that the summation now goes from 0 to m (n even), or from 0 to $m-1$ (n odd). Accordingly, the autocorrelation function $C_k(\alpha)$ is an even function computed in the lag range $\pm m$ or $\pm m-1$ and yields a quantitative measure of the coupling between different point perturbations acting on the k-th residual. Significant correlation values suggest presence of systematic errors. One also sees that the autocorrelation function is not defined outside the lag range $\pm m$ or $\pm(m-1)$ due to construction of the periodogram up to the frequency m/n or $(m-1)/n$, which implies that autocorrelation values at lags $>m$ (n even) or $>m-1$ (n odd) are arbitrarily assigned a value of zero [5]. This procedure is widely used in the estimation of autocorrelation functions [7] and fast Fourier transform theory [5]. It leads to no loss of generality, since values of the autocorrelation function for n points calculated outside the interval $\pm n/2$, or even $\pm n/4$ in some cases, are subject to severe fluctuations and are statistically irrelevant [5,7].

From the autocorrelation function one can derive a useful parameter characterizing the degree of nonrandomness of each residual. This parameter is defined here as the relaxation length for each point, $L(k)$, obtained from the integral equation

$$L(k)=\int_{-q}^{q} C_k(\alpha)\,\mathrm{d}\alpha \tag{19}$$

provided the integral converges in the interval of interest $\pm q$ ($q=m$ for n even; $q=m-1$ for n odd). This is always true when $n\to\infty$ and can readily be accomplished for arbitrary n by tapering the function $C_k(\alpha)$ with the Bartlett weights $W(\alpha)=1-\alpha/q$ [5,6]. The use of Bartlett's window is a standard procedure in discrete Fourier analysis aimed at smoothing the limiting regions of the autocorrelation function where fluctuations may be severe. The relaxation length defined above parallels the relaxation time of perturbation-induced fluctuations in a thermodynamic system [8], and hence is necessarily a positive quantity [9]. The relaxation length for a point subject to Gaussian white noise is zero, as a consequence of the fluctuation-dissipation theorem [8]. In general $L(k)$ gives the degree of coupling, and hence nonrandomness, of the point

perturbations relative to the k-th residual. The relaxation profile obtained by calculating L for each point encapsulates the results of PPA of the phenomenon under consideration and yields the 'distance of correlation' in point units for each residual. Therefore, in the absence of systematic errors the relaxation profile is expected to be as close as possible to the zero baseline, or at least should lie well below the boundary $L = 1$, i.e., the unit distance of correlation. *

The generality of the approach proposed here is evident, based as it is on fundamental principles of nonequilibrium thermodynamics. Details on the construction of a program for PPA are given in the appendix, along with a summary of the basic operational steps involved. Application of PPA is possible whenever the analysis of experimental data can be cast in terms of functionals such as eq. 1. In sections 3 and 4, we demonstrate application to two particular cases, namely, the viscosity of binary mixtures near the critical point and the oxygen and carbon monoxide binding reactions to human hemoglobin. Both examples are prototypic as they comply with intrinsic difficulties arising from the experimental approach and/or the theoretical interpretation. We trust these applications will provide sufficient details regarding the basic aspects of PPA.

3. PPA of the viscosity of binary mixtures near the critical point

Measurements of the viscosity of binary mixtures near the critical point have been given much consideration from both the theoretical [10–12] and experimental [13–17] points of view. Interest in this subject arises from the difficulty in interpreting the weak divergence of viscosity near the critical point. The temperature dependence of the viscosity, η, far from the critical point is given by the Arrhenius equation

$$\eta = Ae^{Bt} \tag{20}$$

where A and B denote constants, t the reduced temperature ($=|T - T_c|/T_c$), and T_c the critical temperature. The weak divergence approaching the critical point can be expressed by the empirical function of Debye [18]

$$\eta = Ae^{Bt}t^{-\phi} \tag{21}$$

where ϕ is the critical exponent. The divergence is accounted for by the term $t^{-\phi}$, which becomes very sharp when approaching the critical temperature. Other relations have been devised to correct for the nonideal character of the viscosity near the critical point, almost all of which make use of the Debye kernel $t^{-\phi}$ [12]. Yet, the phenomenon is still far from being understood from a theoretical standpoint and is further complicated by obvious experimental difficulties. This situation seems to be most appropriate for PPA. In fact, nonrandomness of residuals may stem from systematic errors due either to the experimental technique and/or to the functional form of the empirical expression, eq. 21.

Viscosity measurements for the system 2,6-lutidine/water were carried out using a capillary viscosimeter as described elsewhere [13] and analysed according to eq. 21 by nonlinear least-squares minimization. The results are shown in fig. 1, and the experimental determinations are given in table 1, along with best-fit parameter estimates and residuals. The data were then processed by PPA and the results of parameter perturbation are listed in table 2. It should be noted that the perturbations $\delta\Phi_k^*$ are all positive, in agreement with the stability principle given by eq. 5. Furthermore, the correlation matrix derived from PPA is in striking agreement with that obtained from the covariance matrix. The displacement vectors δr^* obtained by point perturbation are altogether shown in Fig. 2a in the form of a perturbation surface normalized by $\sqrt{\Phi^*}$. The surface readily reveals points or regions of high and low variability, along with structured patterns suggestive of nonrandom behavior. Perturbation surfaces were also calculated for three other sets of viscosity measurements for the 2,6-lutidine/water mixture, two of which were obtained by a capillary viscosimeter in other laboratories [15,16] (fig. 2b,c), the third being obtained

* The exact distribution of relaxation length values for pseudo-random deviates, i.e., in the absence of systematic errors, will be dealt with in a forthcoming paper.

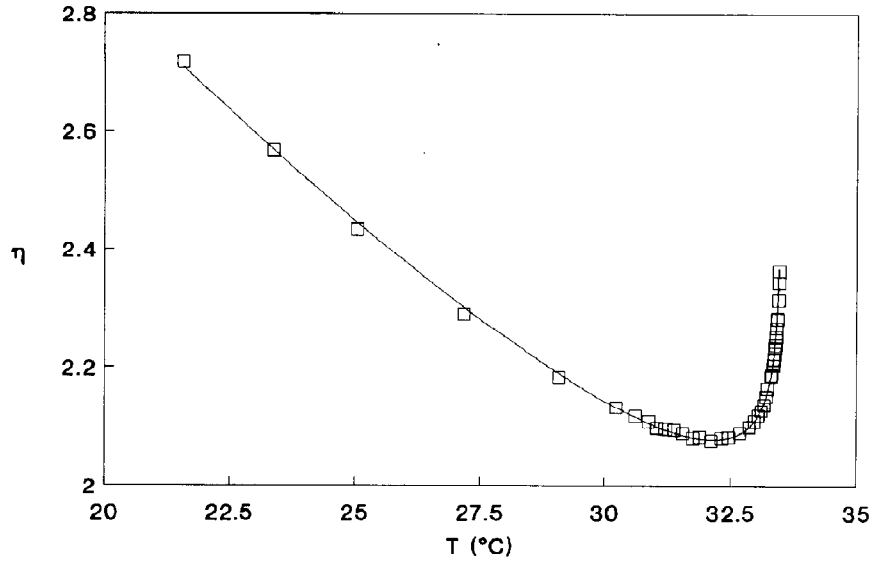

Fig. 1. Experimental determination of the shear viscosity of the binary mixture 2,6-lutidine/water plotted as a function of temperature. Measurements were made on a critical composition of 29.75 wt% of 2,6-lutidine by a standard Ubbelhode capillary viscosimeter modified as described elsewhere [13]. Temperature was controlled to $\pm 0.01°$ C by a water bath and measured by a Rosemount platinum-resistance thermometer (model E104-J4-4048-400) calibrated with a Leeds and Northrup platinum-resistance thermometer. The continuous line was drawn according to eq. 21 with the parameter values listed in table 1. The quality of the data ($\sigma = 0.00435$ cP) is comparable to that of other determinations [13–17]. The best-fit estimates of the relevant physical parameters, and especially the value of the critical exponent, are in agreement with those reported by others [15,16].

by the oscillating disk method [19] in our laboratory (fig. 2d). For each data set, autocorrelation functions were obtained by Fourier analysis through the empirical power spectra as discussed in section 2. Relaxation lengths were then computed by integration of eq. 19 and the relaxation profiles for the four data sets analysed in this study are altogether shown in fig. 3. For the sake of comparison, the profiles are plotted as a function of the logarithm of reduced temperature. In all cases, long relaxations ($L > 1$) suggest the presence of systematic errors. Interestingly, the three data sets (a–c) obtained with a capillary viscosimeter show a similar relaxation profile with a peak in the range of reduced temperature from 10^{-2} to 10^{-3}. This is the range where viscosity begins to depart from Arrhenius behavior. The presence of such a common trend in data sets obtained in different laboratories, as revealed by PPA, indicates that the theoretical treatment based on eq. 21 may not provide a correct interpretation of the divergence from Arrhenius behavior. This conclusion seems to be further supported by viscosity measurements made by the oscillating disk method, a different experimental approach, which also shows the presence of a relaxation peak in the same temperature range.

It is worth pointing out that PPA of the viscosity measurements discussed above not only provides an indication regarding the temperature range where nonrandomness of residuals is more evident, but also yields a quantitative measure of nonrandomness through the relaxation profile. More importantly, the conclusions reached via PPA would be difficult to predict from standard analysis of the residuals shown in table 1, that appear to be evenly distributed. The same holds true for the other three data sets analysed in this study, that have been tabulated in the original papers [15,16,19]. The relaxation peak observed in the case of 2,6-lutidine/water mixtures raises some questions as to the theoretical interpretation of the

Table 1

Experimental measurements of the shear viscosity (in cP) of the binary mixture 2,6-lutidine/water, along with the results of least-squares minimization using eq. 21

$A = 1.549 \pm 0.006$; $B = 10.61 \pm 0.065$; $\phi = 0.0455 \pm 0.0006$; $T_c = 33.488 \pm 0.002$; n (no. of points) = 39; p (no. of parameters) = 4; $\sigma = \sqrt{\Phi^*/(n-p)} = 0.00435$ cP.

No.	T (°C)	η_{exp}	η_{fit}	r/σ
1	21.58	2.718	2.710	1.76
2	23.37	2.568	2.567	0.33
3	25.05	2.433	2.442	−2.01
4	27.18	2.290	2.299	−1.96
5	29.07	2.183	2.188	−1.20
6	30.22	2.132	2.132	0.02
7	30.61	2.117	2.116	0.33
8	30.88	2.108	2.105	0.62
9	31.04	2.097	2.100	−0.63
10	31.21	2.096	2.094	0.40
11	31.37	2.094	2.090	1.01
12	31.55	2.088	2.085	0.67
13	31.75	2.080	2.081	−0.22
14	31.89	2.083	2.079	0.95
15	32.12	2.076	2.077	−0.23
16	32.32	2.080	2.078	0.55
17	32.46	2.082	2.080	0.55
18	32.68	2.089	2.087	0.56
19	32.87	2.099	2.098	0.15
20	32.98	2.109	2.109	−0.03
21	33.05	2.118	2.118	−0.06
22	33.11	2.127	2.128	−0.25
23	33.16	2.137	2.138	−0.28
24	33.21	2.151	2.151	0.08
25	33.24	2.165	2.160	1.24
26	33.31	2.185	2.187	−0.50
27	33.32	2.188	2.192	−0.96
28	33.35	2.204	2.210	−1.29
29	33.36	2.214	2.216	−0.56
30	33.37	2.224	2.224	0.02
31	33.38	2.233	2.232	0.20
32	33.39	2.238	2.241	−0.75
33	33.40	2.251	2.252	−0.12
34	33.41	2.267	2.263	0.88
35	33.42	2.282	2.277	1.24
36	33.43	2.283	2.292	−2.16
37	33.44	2.315	2.311	0.81
38	33.45	2.344	2.336	1.95
39	33.46	2.363	2.368	−1.05

initial departure from Arrhenius behavior in the viscosity curve, which will be considered in detail elsewhere. For the purpose of the present study it is sufficient to point out that PPA may effectively be used to identify possible discrepancies between experimental data and theoretical models.

4. PPA of oxygen and carbon monoxide binding to human hemoglobin

The study of the functional properties of human hemoglobin complies with several experimental difficulties that often complicate collection of high-precision measurements. Among the various techniques used in the determination of equilibrium binding constants spectroscopic methods have gained wide application in the past two decades due to their high reliability. Imai and co-workers were the first to recognize the importance of precise and reproducible spectroscopic measurements. Their technique [21] indeed set the stage for a quantitative analysis of the functional properties of human hemoglobin [22–24]. More recently, Dolman and Gill [25] have developed a different spectroscopic technique which makes use of a thin-layer cell. This technique is particularly suitable to control precisely the partial pressure of gaseous ligands and yields oxygen and carbon monoxide binding isotherms on highly concentrated (typically > 1 mM heme) hemoglobin samples [26,27]. The main features of the two techniques are discussed at length in the original papers [21,25]. Although both techniques are spectroscopic, they differ substantially in the strategy followed to achieve high-precision binding measurements. In the apparatus of Imai et al., the ligand saturation is derived from the expression

$$A(x) = A(0) + \delta A_T Y(x) \tag{22}$$

where $A(x)$ represents the absorbance measured when the ligand activity is x, δA_T the total absorbance change observed in taking the hemoglobin sample from the unligated to the fully ligated form, and $Y(x)$ the hemoglobin fractional saturation. In the method of Gill and co-workers, the underlying relation is

$$\begin{aligned}\delta A(\xi_i) &= A(x_i) - A(x_{i-1}) \\ &= \delta A_T [Y(x_i) - Y(x_{i-1})]\end{aligned} \tag{23}$$

where ξ_i is the geometric mean $\sqrt{x_i x_{i-1}}$. Here

Table 2

Results of PPA of the experimental data set reported in table 1

Perturbations are given as % values of the unperturbed parameter. $\Phi^* = 0.000662$; $p_1^* = 1.549$; $p_2^* = 10.61$; $p_3^* = 0.0455$; $p_4^* = 33.488$.

Correlation matrix from PPA

1.00 −0.88 0.99 −0.82
1.00 −0.83 0.55
1.00 −0.87
1.00

Correlation matrix from V

1.00 −0.89 0.99 −0.80
1.00 −0.85 0.61
1.00 −0.85
1.00

k	$\delta\Phi_k^*$	δp_{k1}^*	δp_{k2}^*	δp_{k3}^*	δp_{k4}^*
1	18.2973	0.3849	−1.1989	1.1986	−0.0029
2	0.1713	0.0228	0.0971	0.0687	−0.0001
3	2.6254	−0.0069	0.2579	−0.0011	−0.0002
4	1.2297	0.0942	0.0014	0.3158	−0.0010
5	0.3735	0.0923	−0.0917	0.2988	−0.0008
6	0.0001	−0.0015	0.0019	−0.0048	0.0000
7	0.0267	−0.0283	0.0378	−0.0895	0.0002
8	0.0914	−0.0522	0.0722	−0.1638	0.0004
9	0.0927	0.0519	−0.0734	0.1619	−0.0004
10	0.0363	−0.0326	0.0470	−0.1012	0.0002
11	0.2262	−0.0802	0.1182	−0.2473	0.0006
12	0.0982	−0.0519	0.0783	−0.1587	0.0003
13	0.0103	0.0161	−0.0251	0.0484	−0.0001
14	0.1784	−0.0658	0.1046	−0.1966	0.0004
15	0.0099	0.0142	−0.0236	0.0412	−0.0001
16	0.0511	−0.0296	0.0516	−0.0833	0.0001
17	0.0474	−0.0255	0.0466	−0.0688	0.0000
18	0.0436	−0.0177	0.0370	−0.0412	−0.0001
19	0.0029	−0.0023	0.0067	−0.0027	−0.0000
20	0.0001	−0.0001	−0.0006	−0.0014	0.0000
21	0.0005	−0.0006	−0.0007	−0.0045	0.0000
22	0.0096	−0.0044	−0.0000	−0.0246	0.0002
23	0.0120	−0.0072	0.0034	−0.0348	0.0002
24	0.0012	0.0027	−0.0019	0.0124	−0.0001
25	0.2816	0.0505	−0.0421	0.2180	−0.0011
26	0.0538	−0.0286	0.0295	−0.1153	0.0005
27	0.2034	−0.0570	0.0601	−0.2280	0.0009
28	0.3863	−0.0834	0.0931	−0.3263	0.0012
29	0.0740	−0.0370	0.0420	−0.1436	0.0005
30	0.0001	0.0011	−0.0013	0.0043	−0.0000
31	0.0092	0.0128	−0.0150	0.0491	−0.0002
32	0.1331	−0.0486	0.0583	−0.1850	0.0006
33	0.0034	−0.0076	0.0093	−0.0286	0.0001
34	0.1845	0.0495	−0.0622	0.1851	−0.0004
35	0.3856	0.0589	−0.0766	0.2173	−0.0003
36	1.3601	−0.0708	0.0989	−0.2544	−0.0004
37	0.2838	0.0045	−0.0119	0.0108	0.0007
38	4.0285	−0.1023	0.1035	−0.4038	0.0043
39	12.4404	0.3368	−0.3907	1.2762	−0.0089

absorbance changes are measured when the ligand activity changes stepwise from x_{i-1} to x_i, so that the δAs scale with the binding capacity of the macromolecule, i.e., with the derivative of the binding curve $dY(x)/d\ln x$ [28]. Both techniques tacitly assume linearity between absorbance and

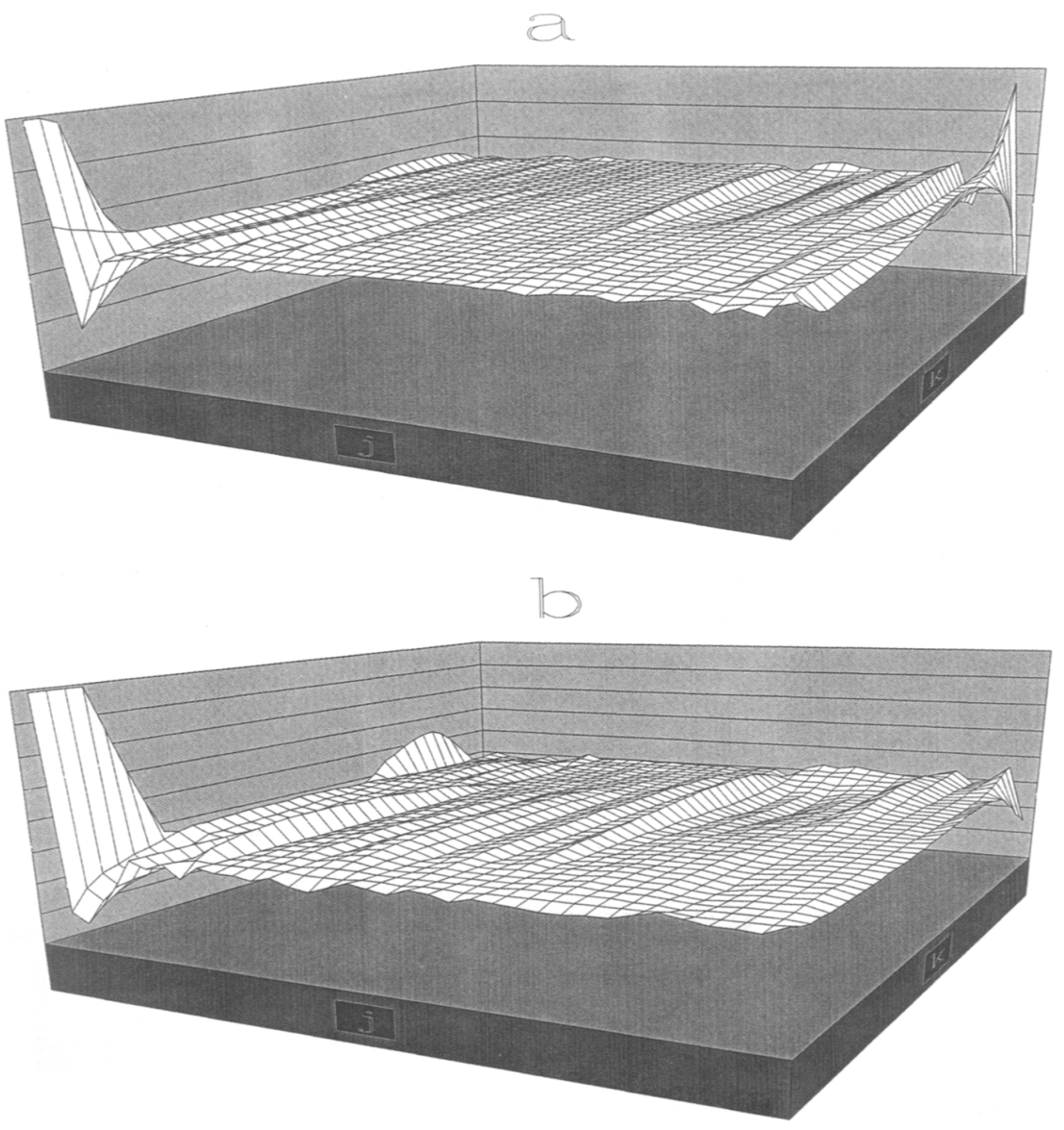

Fig. 2. Perturbation surfaces of the four determinations of shear viscosity of the binary mixture 2,6-lutidine/water analysed in this study: (a) data obtained in this study and reported in fig. 1; (b) data obtained by Stein et al. [15] using a capillary viscosimeter; (c) data obtained by Gulari et al. [16] using a capillary viscosimeter; (d) data obtained by Andreasi Bassi [19] using the oscillating disk method. Each surface is a perturbation (square) matrix whose columns correspond to experimental points and rows give the particular point being dropped. The element $\delta r_{j,k}^{*}$ thus gives the displacement of the k-th residual due to dropping the j-th point. The surfaces were normalized by $\sqrt{\Phi^{*}}$ for the sake of comparison among different data sets. The vertical axis of the surface goes from -0.5 to 0.5. The horizontal axes (labeled by j and k) run from 1 to n in the direction left → right. One should note the striking similarity of the four surfaces, which underlies a common relaxation profile as shown in fig. 3.

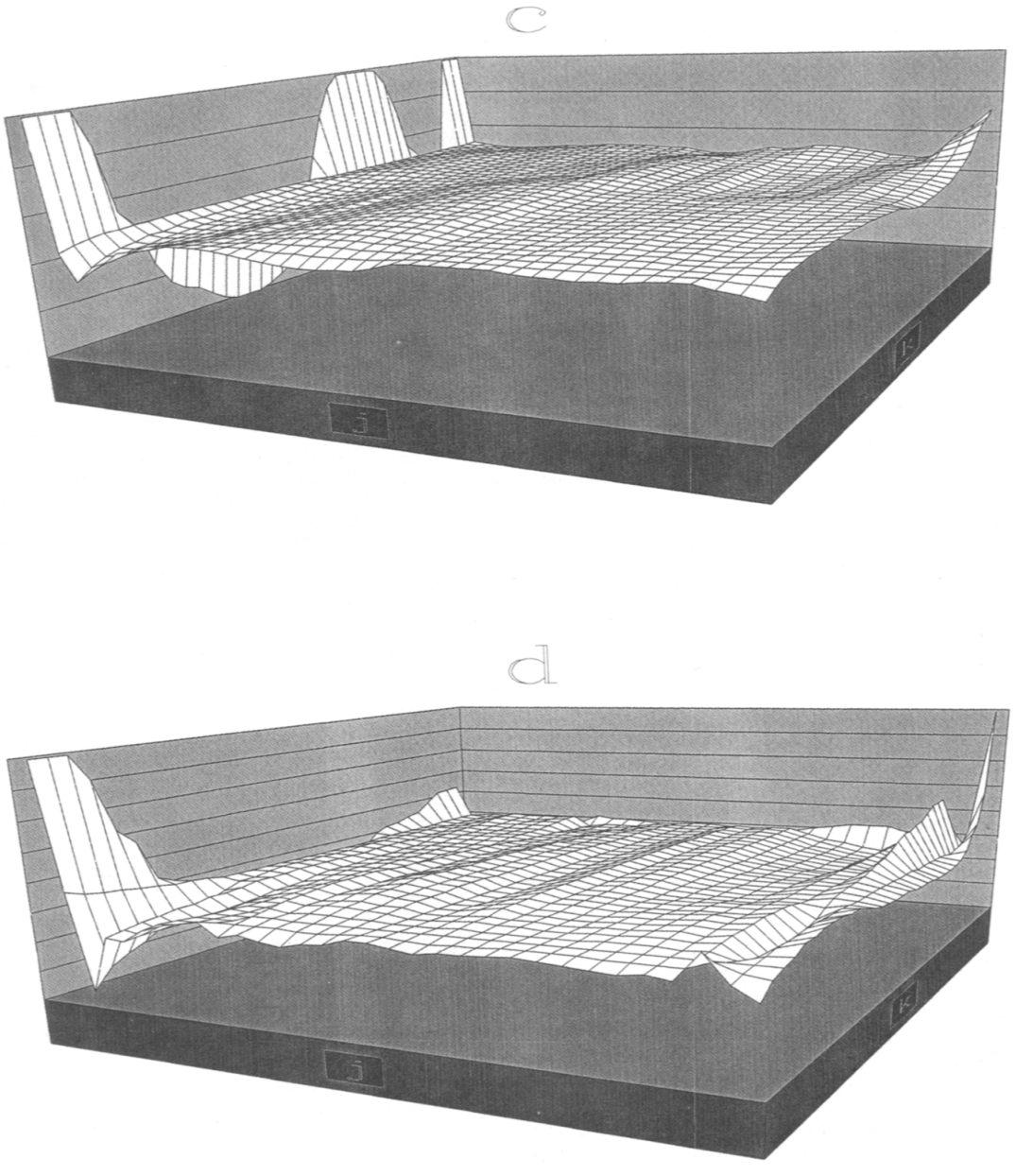

Fig. 2 (c,d).

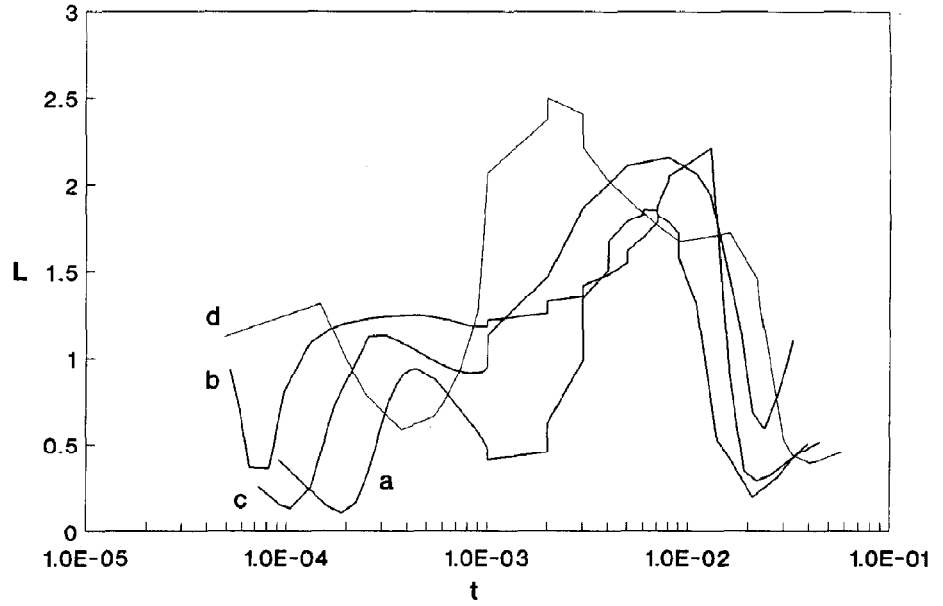

Fig. 3. Relaxation profiles for the four data sets analysed in this study and whose perturbation surfaces are shown in fig. 2. The relaxation length is plotted vs. the reduced temperature to allow quantitative comparison. Consistent with the similarity observed in the perturbation surfaces (see fig. 2), a common trend is also seen in the relaxation profiles. In particular, high relaxation values ($L > 1$) are observed in all cases in the reduced temperature range 10^{-2}–10^{-3}. Thick-line profiles correspond to measurements made by a capillary viscosimeter: (a) data of fig. 1; (b) data of ref. 15; (c) data of ref. 16. The thin-line profile (d) corresponds to measurements made by the oscillating disk method [19], which yields best-fit estimates of the physical parameters of eq. 21 consistent with those obtained by a capillary viscosimeter [19].

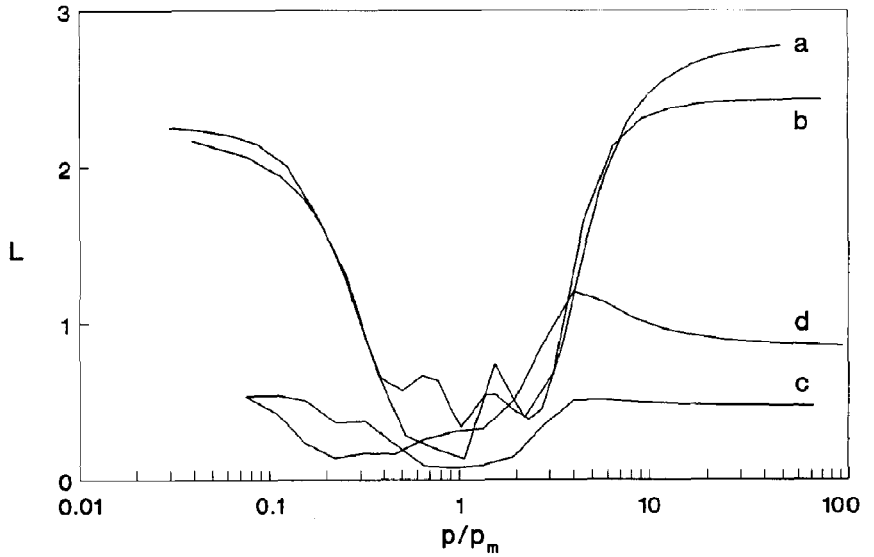

Fig. 4. Relaxation profiles for oxygen (a,b) and carbon monoxide (c,d) binding to human hemoglobin vs. the ligand partial pressure, p, as normalized by the median ligand partial pressure, p_m, to allow for quantitative comparison. Note the similarity between oxygen binding data obtained by the Imai apparatus (a) and the Gill method (b) that both yield high relaxations at relatively low and high partial pressures. Note also the difference between these two profiles (a,b) and the low relaxation profiles obtained for carbon monoxide determinations made by the Gill method (c,d). Experimental conditions: (a) 60 μM heme (HbA_0), 0.1 M phosphate, pH 7.5, 25 °C [22]; (b) 4 mM heme (HbA_0), 50 mM Hepes, 50 mM glycine, 50 mM Bicine, 0.1 M NaCl, 40 μM Na_2EDTA, pH 7.51, 25 °C [29]; (c) 2 mM heme (HbA_0), 0.1 M Hepes, 0.1 M NaCl, 15 mM diphosphoglycerate (DPG), pH 7.35, 37 °C [30]; (d) 1.2 mM heme (HbF_0), 0.1 M Hepes, 0.1 M NaCl, 15 mM DPG, pH 7.36, 37 °C [30].

saturation changes and hence eqs. 22 and 23 are taken to be 'exact'. This is because the function $Y(x)$ used in the fitting procedure can be cast exactly in terms of a partition function by application of the basic principles of statistical thermodynamics. Such a situation can conveniently be tested by PPA, just as in the case of the viscosity measurements analysed in section 3.

Typical data sets obtained with the apparatus of Imai et al. and the method of Gill and coworkers were first analyzed according to eqs. 22 or 23 and then processed by PPA. The relevant relaxation profiles for oxygen and carbon monoxide binding are shown in fig. 4 as a function of the logarithm of ligand partial pressure. One observes a striking difference between the oxygen and carbon monoxide determinations and similarity between profiles obtained in the case of oxygen measurements made by the two different experimental techniques. The presence of long relaxations at low and high oxygen partial pressures was confirmed by PPA of other binding determinations obtained by means of the two techniques, which naturally raises the question as to the origin of such a systematic error which is not seen in the case of carbon monoxide determinations. In order to explore further this point, we have also analysed oxygen binding curves of the isolated α and β chains of human hemoglobin, that represent a 'simpler' system than the macromolecule as a whole. The relaxation profiles are shown in fig. 5. The data of Mills [31], obtained by means of the apparatus of Imai et al., show a surprisingly high profile as opposed to the very low relaxations observed for data collected with the method of Di Cera and Gill [32]. This result is of interest as the binding data of Mills on the isolated chains have been reported to be 'inferior to that obtained for intact hemoglobin' (ref. 31, p. 143) for various technical reasons, including oxidation. It is important to note that such a discrepancy in the experimental data of Mills is readily revealed by PPA. Furthermore, the difference between figs. 4 and 5 is noteworthy. The relaxation profiles observed in the case of oxygen binding to human hemoglobin are not seen in the case of the iso-

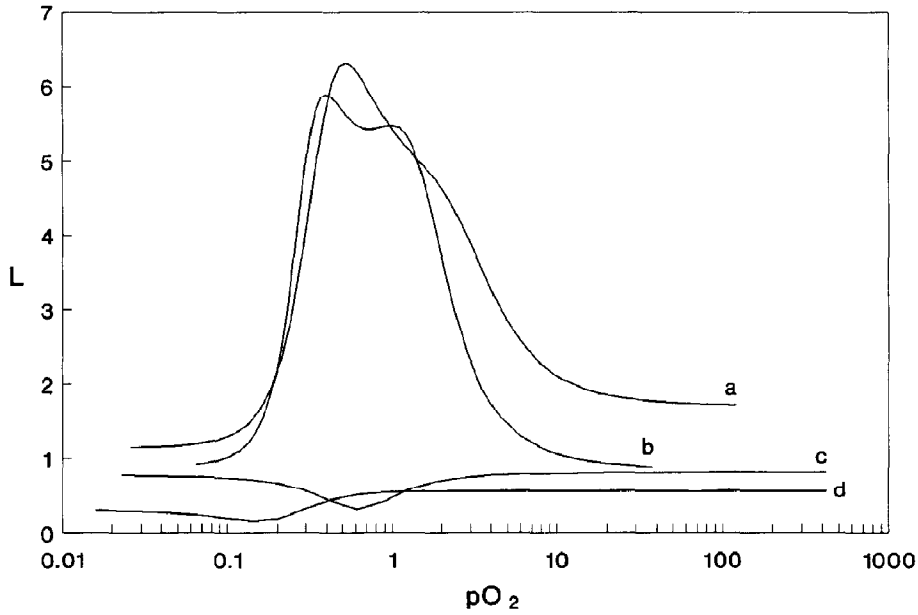

Fig. 5. Relaxation profiles for oxygen binding to the isolated α (a,c) and β (b,d) chains of human hemoglobin as a function of oxygen partial pressure. The data were obtained by the Imai apparatus (a,b) [31] and the Gill method (c,d) [32] under experimental conditions of 40–60 [31] or 200 [32] μM heme, at 25 °C.

lated chains, even in the presence of 'problems' such as oxidation.

5. Discussion

The foregoing analysis of experimental data has revealed the possibility of quantitatively testing for nonrandomness of residuals. The method is based on discrete perturbations of experimental data points that are used to probe the metric of the parameter hyperspace and to construct a perturbation surface of residuals. This fact considerably expands the degrees of freedom and hence the amount of information stored in the system under consideration. Analysis of the perturbation surface yields a quantitative measure of nonrandomness, i.e., the relaxation length for each experimental point subject to the perturbation. The physical parallel of the perturbation with a random walk provides a sound theoretical basis for the significance of the relaxation profiles constructed by PPA.

We have shown two applications of the method to widely different problems, such as the analysis of viscosity measurements near the critical point and the analysis of ligand binding to human hemoglobin. In both cases the method has pointed out nonrandomness of residuals, although of differing extents. Viscosity measurements of the 2,6-lutidine/water system share a relaxation peak in the temperature range characterizing the onset of 'nonideal' behavior, where the Arrhenius law no longer holds. The origin of nonrandom behavior of residuals in this particular temperature range, as pointed out by PPA, deserves consideration and should be analysed in connection with recent theoretical reevaluations of critical phenomena [33], which will be dealt with elsewhere. Other questions are opened by PPA in the case of human hemoglobin. First of all, a substantial difference is observed in the relaxation profiles of oxygen and carbon monoxide determinations, irrespective of the spectroscopic technique used. One possibility could be a difference in oxidation which is typically higher when measuring oxygen binding curves. However, this conclusion is not supported by relaxation profiles of the isolated chains obtained under experimental conditions of relatively high oxidation content. This draws attention to

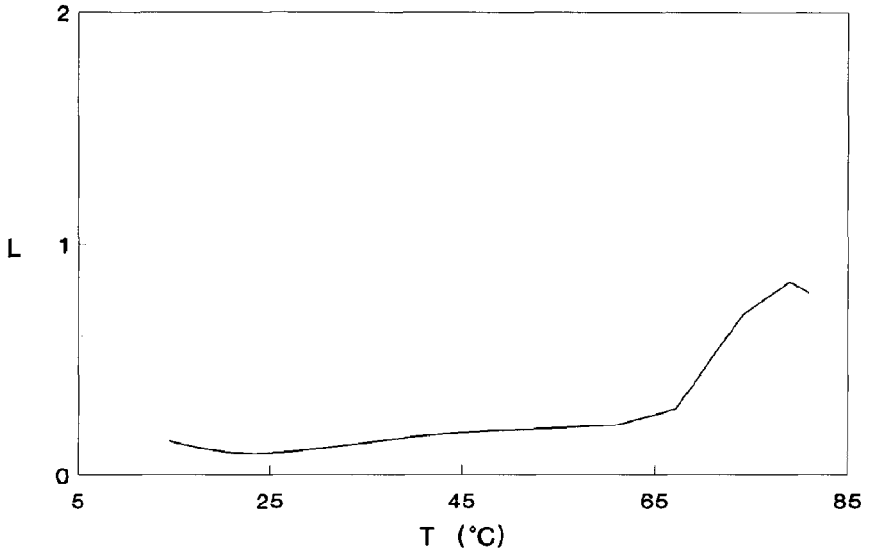

Fig. 6. Relaxation profile for the vapor pressure of benzene as a function of temperature [35]. The experimental data, listed in ref. 34, were fitted to the Antoine equation (see eq. 24) and then processed by PPA. The relaxation profile lies well below 1 in the temperature range explored experimentally, which indicates the absence of appreciable systematic errors.

the 'exactness' of eqs. 22 and 23, just as in the case of the viscosity measurements analysed in this study. The basic assumption that spectral changes are linearly related to saturation changes is probably a weak one and may be responsible for the observed nonrandom behavior of residuals. This point needs further investigation and can only be clarified by thorough analysis, or reanalysis, of experimental binding data obtained by spectroscopic and nonspectroscopic techniques. In all this PPA may provide key information.

It is worth pointing out that PPA correctly yields low relaxation profiles in the absence of systematic errors, as already pointed out in figs. 4 and 5. Extensive simulations of the experimental measurements discussed in this study were carried out by adding pseudorandom deviates to eqs. 21–23 and processing the synthetic data by PPA. In all cases, the relaxation profiles were found to be very low, with values of L well below unity. In addition, we have processed by PPA experimental determinations of the saturated vapor pressure of benzene as a function of temperature [34]. The Antoine equation

$$\log_{10} p = a - b/(T + c) \tag{24}$$

where p is the vapor pressure, T the absolute temperature, and a, b, and c are constants, was used as fitting equation. The relaxation profile is shown in fig. 6 for the data of Willingham et al. [35]. One sees that the relaxation lengths are very low and no appreciable systematic error is present.

The method we have proposed has turned out to be effective for detecting nonrandom behavior of residuals, which shows up in the relaxation profile in quantitative terms. As we have seen, departure from randomness can be detected by PPA even when the residuals are evenly distributed as in the case of the viscosity measurements listed in table 1. Conventional checks applied to residuals lack this sensitivity. For example, the residual autocovariances of a stretch of residuals $x_1, x_2, \ldots x_n$, are given by [5,7]

$$C(\alpha) = [n/(n-\alpha)]\Sigma_j x_j x_{j+\alpha}/\Sigma_j x_j^2 \tag{25}$$

with $\alpha = 0, 1, \ldots n-1$, and the summation in the

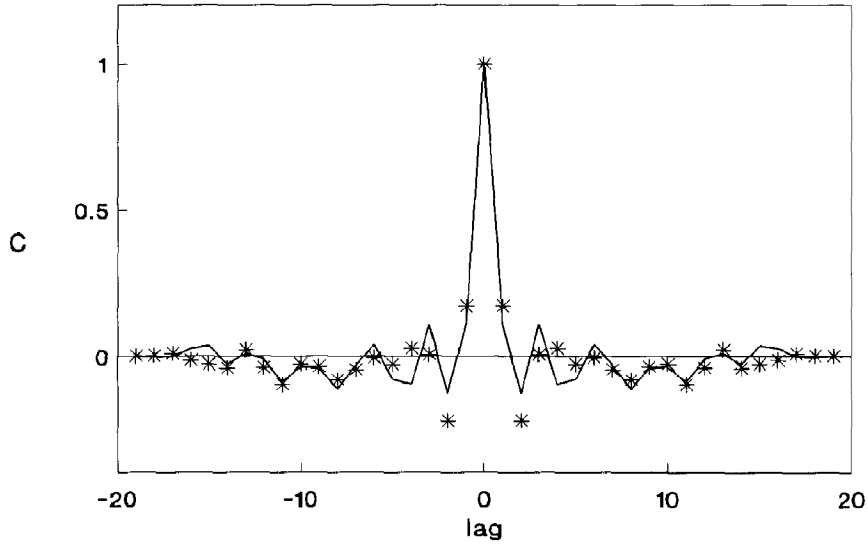

Fig. 7. Autocorrelation function for the residuals of the data listed in table 1. The function was calculated in the lag range $\pm q$ ($q = 19$) according to the Fourier analysis described in the text (points connected by a line) and the autocovariances (*) computed with eqs. 25 and 26. The area under the autocorrelation function was calculated using the extended trapezoidal rule [20] for both procedures, with a result of 0.34 for the values connected by a line and 0.12 for those depicted by stars.

numerator goes from 1 to $n-\alpha$. The autocovariances can conveniently be tapered in the interval $\pm q$ using the Bartlett window [5]

$$W(\alpha)=\begin{cases}1-\alpha/q & \text{for } \alpha \leqslant q\\ 0 & \text{for } \alpha > q\end{cases} \qquad (26)$$

with $q=n/2$ for n even, and $q=(n-1)/2$ for n odd. The empirical autocorrelation function for the data listed in table 1 was calculated from eqs. 25 and 26 and the results are shown in fig. 7. One observes that there is no significant correlation in the lag interval ± 19 ($n=39$ and hence $q=19$ for the data in table 1), so that from the autocovariance check of residuals one would conclude that the data in table 1 have no systematic error, contrary to what is observed by PPA. The same conclusion would be drawn from checking the residuals of other viscosity measurements analysed in this study. This shows that standard checks of residuals may not be appropriate, presumably because the amount of information stored in a stretch of residuals is seldom revealing of hidden periodicities. Construction of a perturbation matrix by PPA draws attention to the way in which the perturbation-induced fluctuations of each residual relax to the 'equilibrium' state. In practice, this method checks for hidden periodicities not just over n residuals, as in the case of standard periodogram tests [5], but over an $n \times n$ perturbation matrix which obviously contains much more information.

It is worth pointing out that the empirical autocovariances calculated according to eqs. 25 and 26 give a good approximation of the autocorrelation function derived by periodogram analysis of the residuals. In fig. 7 the autocovariances calculated for the residuals of the data listed in table 1 are depicted by stars and are nicely scattered around the line derived by Fourier analysis as described in section 2. The good agreement between the two methods may suggest construction of the relaxation profile from analysis of the perturbation matrix according to eqs. 25 and 26, rather than going through the complications of Fourier transforms. Although such a procedure would indeed be much simpler than the Fourier analysis proposed in this study, nevertheless calculation of the relaxation length from the empirical autocovariances leads to a severe underestimation of L, up to a factor of 3, in many cases. For this

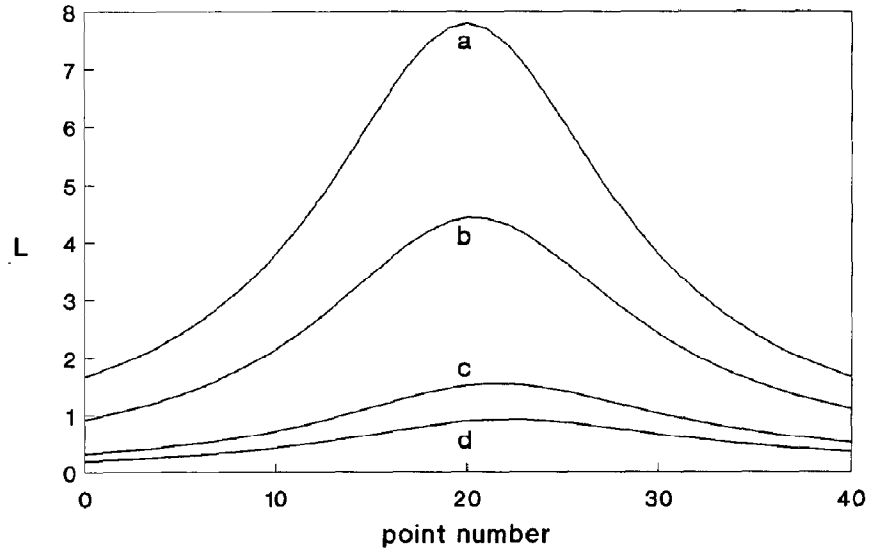

Fig. 8. Relaxation profiles for the simulation study mentioned in section 5. 40 data points were generated in the x range -0.5 to 0.5 according to the equation $x^2-0.125+k\sigma B$, where B denotes Box-Muller pseudorandom deviates [20], $k=0, 1, \ldots$, and $\sigma=0.076$. The simulated data points were then fitted to a straight line $a+bx$. (a) $k=0$; (b) $k=1$; (c) $k=3$; (d) $k=5$.

reason, we believe that such a procedure should not be used as an alternative to Fourier analysis, but at most in conjunction with it, in order to double check the overall consistency of the results. Our conclusion concurs with widely accepted 'recommendations' in Fourier analysis [5,7].

The sensitivity of PPA can best be checked by a simulation study. Data points were simulated according to the equation $x^2 - 0.125$, in the x range ± 0.5, and then fitted to the equation $a + bx$. A trivial systematic error in the residuals was recovered with a standard error $\sigma = 0.076$. The systematic error was then progressively hidden by superimposing Box-Muller pseudorandom deviates [20] to the original expression $x^2 - 0.125$ with a standard deviation $k\sigma$ ($k = 1, 2\ldots$), and the resulting data points were again fitted to a straight line. The relaxation profiles for four typical cases are shown in fig. 8. The profile begins to approach the zero baseline, and hence to become indistinguishable from truly random noise, for $k = 3$. On the other hand, standard analysis of residuals hardly reveals a significant departure from randomness even for $k = 1$.

A final remark may be appropriate regarding the fact that PPA per se cannot provide a rationale for nonrandom behavior. Relaxation profiles must always be evaluated in connection with the fundamental theoretical interpretations of the experimental facts. We trust such an interplay casts the analysis of experimental data in terms of a new, and perhaps more comprehensive, framework.

Appendix

Here, we wish to provide more details of the procedure introduced in this study in order to facilitate practical application of PPA. We trust this appendix is clear enough to allow the interested reader to construct a computer program for PPA. Alternatively, interested readers may request copies of our PPA program written in Quick-Basic for a Compaq 386/20e computer.

We want to process by PPA the data set reported in table 3, where x is the independent variable and y is a stretch of $n = 10$ simulated values using the equation

$$y_i = 1 + 0.1x_i + \Gamma_i \tag{A1}$$

with Γ being a Box-Muller pseudorandom error [20]. The data set is first analysed according to the equation

$$F(x_i) = a + bx_i \tag{A2}$$

to determine the best-fit values of a and b, and the residuals $r_{0i}^* = y_i - F(x_i)$. The subscript $0i$ indicates that the i-th residual is calculated using the same weight for all n points. Next the data set is analysed by dropping the weight of point 1 to zero, while that of the remaining points is unchanged, to obtain new residuals of the form $r_{1i}^* = y_i - F_1(x_i)$. The subscript $1i$ now indicates that the i-th residual is computed while dropping point 1, and F_1 is the new best-fit value of F. The procedure is repeated for all n points to yield a matrix of residuals as shown in table 4. Note that the value of the i-th residual obtained when dropping the i-th point must be included in the matrix. The r_{ii}^* values are most important in checking the stability condition expressed by eq. 5. These values must always be greater, in absolute value, than the r_{0i}^* values obtained without dropping any point. If this is not the case, then the minimization procedure has converged to a false minimum. In practice, it is useful to store all residuals in an $(n + 1) \times n$ matrix at the end of the $n + 1$ minimizations. The rows of this matrix indicate which point has been dropped (0 indicates that no point has been dropped), while columns refer to the

Table 3

Simulated data set obtained from eq. A1

i	x_i	y_i
1	1	1.32
2	2	1.28
3	3	1.19
4	4	1.18
5	5	1.55
6	6	1.70
7	7	1.64
8	8	1.82
9	9	1.85
10	10	1.98

Table 4

Matrix of residuals r_{ik}^{*} obtained by PPA of the data set reported in table 3

Rows indicate the particular data point being dropped, while columns refer to the residual being considered.

	1	2	3	4	5	6	7	8	9	10
0	0.17	0.04	−0.14	−0.24	0.04	0.10	−0.04	0.05	−0.01	0.03
1	0.26	0.11	−0.08	−0.19	0.08	0.12	−0.04	0.04	−0.03	−0.01
2	0.18	0.05	−0.13	−0.23	0.05	0.11	−0.04	0.05	−0.01	0.03
3	0.13	0.00	−0.17	−0.26	0.02	0.09	−0.05	0.04	−0.01	0.04
4	0.12	−0.01	−0.18	−0.27	0.01	0.08	−0.06	0.03	−0.02	0.03
5	0.17	0.04	−0.13	−0.23	0.05	0.11	−0.04	0.05	−0.01	0.03
6	0.18	0.05	−0.13	−0.23	0.05	0.12	−0.03	0.06	0.00	0.05
7	0.17	0.04	−0.14	−0.24	0.04	0.10	−0.05	0.04	−0.02	0.02
8	0.17	0.04	−0.14	−0.24	0.05	0.11	−0.04	0.06	0.00	0.04
9	0.17	0.04	−0.14	−0.24	0.04	0.10	−0.05	0.04	−0.01	0.03
10	0.16	0.03	−0.14	−0.24	0.05	0.11	−0.04	0.06	0.00	0.05

Table 5

Fourier coefficients and periodogram values obtained by PPA of the data set reported in table 3

k	$A_k(0)$	$A_k(1/10)$	$A_k(2/10)$	$A_k(3/10)$	$A_k(4/10)$	$A_k(5/10)$
1	0.001	0.017	0.035	0.014	0.014	0.009
2	−0.001	0.013	0.031	0.010	0.010	0.007
3	0.002	0.011	0.027	0.008	0.006	0.006
4	0.003	0.009	0.021	0.006	0.006	0.005
5	0.004	0.008	0.018	0.005	0.003	0.002
6	0.005	0.002	0.013	−0.001	0.002	−0.001
7	−0.004	0.000	0.008	−0.003	0.001	−0.002
8	−0.003	−0.001	0.005	−0.004	−0.002	−0.005
9	−0.001	−0.003	−0.002	−0.007	−0.002	−0.005
10	0.001	−0.007	−0.007	−0.014	−0.004	−0.009
k	$B_k(0)$	$B_k(1/10)$	$B_k(2/10)$	$B_k(3/10)$	$B_k(4/10)$	$B_k(5/10)$
1	0.000	−0.015	0.005	0.014	0.000	0.000
2	0.000	−0.015	0.005	0.014	0.000	0.000
3	0.000	−0.011	0.001	0.012	−0.002	0.000
4	0.000	−0.007	0.001	0.010	−0.002	0.000
5	0.000	−0.010	0.000	0.009	−0.005	0.000
6	0.000	−0.006	0.000	0.007	−0.005	0.000
7	0.000	−0.003	0.000	0.004	−0.005	0.000
8	0.000	−0.006	0.000	0.004	−0.008	0.000
9	0.000	−0.004	−0.001	0.002	−0.006	0.000
10	0.000	−0.001	−0.003	−0.002	−0.007	0.000
k	$I_k(0)$	$I_k(1/10)$	$I_k(2/10)$	$I_k(3/10)$	$I_k(4/10)$	$I_k(5/10)$
1	0.000005	0.002471	0.006097	0.002069	0.001043	0.000405
2	0.000005	0.001887	0.004794	0.001573	0.000546	0.000245
3	0.000020	0.001165	0.003535	0.001075	0.000225	0.000180
4	0.000045	0.000628	0.002281	0.000672	0.000179	0.000125
5	0.000080	0.000814	0.001680	0.000546	0.000160	0.000020
6	0.000125	0.000220	0.000858	0.000220	0.000142	0.000005
7	0.000080	0.000035	0.000309	0.000125	0.000131	0.000020
8	0.000045	0.000172	0.000118	0.000128	0.000342	0.000125
9	0.000005	0.000111	0.000031	0.000289	0.000209	0.000125
10	0.000005	0.000262	0.000275	0.000978	0.000365	0.000405

Table 6

Autocorrelation values and relaxation lengths obtained by PPA for the data set reported in table 3

k	$C_k(0)$	$C_k(1)$	$C_k(2)$	$C_k(3)$	$C_k(4)$	$C_k(5)$
1	1.00	0.17	−0.42	−0.34	0.01	0.18
2	1.00	0.20	−0.46	−0.36	0.03	0.18
3	1.00	0.22	−0.50	−0.39	0.08	0.22
4	1.00	0.20	−0.50	−0.39	0.11	0.27
5	1.00	0.28	−0.42	−0.32	0.00	0.16
6	1.00	0.24	−0.40	−0.27	0.11	0.43
7	1.00	0.06	−0.29	−0.09	0.14	0.49
8	1.00	−0.24	0.14	−0.02	−0.18	0.09
9	1.00	−0.36	−0.04	0.16	−0.04	−0.36
10	1.00	−0.31	−0.18	0.09	0.13	−0.44
k	$\langle C_k(0)\rangle$	$\langle C_k(1)\rangle$	$\langle C_k(2)\rangle$	$\langle C_k(3)\rangle$	$\langle C_k(4)\rangle$	$\langle C_k(5)\rangle$
1	1.00	0.13	−0.25	−0.14	0.00	0.00
2	1.00	0.16	−0.28	−0.14	0.01	0.00
3	1.00	0.18	−0.30	−0.16	0.02	0.00
4	1.00	0.16	−0.30	−0.15	0.02	0.00
5	1.00	0.23	−0.25	−0.13	0.00	0.00
6	1.00	0.19	−0.24	−0.11	0.02	0.00
7	1.00	0.05	−0.17	−0.03	0.03	0.00
8	1.00	−0.19	0.08	−0.01	0.04	0.00
9	1.00	−0.29	−0.02	0.06	−0.01	0.00
10	1.00	−0.24	−0.11	0.03	0.03	0.00

k	$L(k)$
1	0.48
2	0.50
3	0.48
4	0.46
5	0.70
6	0.72
7	0.76
8	0.68
9	0.48
10	0.42

particular residual being considered. The best-fit parameter values obtained in the $n+1$ minimizations should be stored in a different $(n+1)\times p$ matrix, where p is the number of parameters, whose rows and columns have the same significance as in the matrix of residuals. However, for the purpose of this paper only the matrix of residuals is of interest here.

Once the matrix of residuals has been constructed, the elements δrs of the $n\times n$ perturbation matrix can be obtained as follows

$$\delta r^*_{jk} = r^*_{jk} - r^*_{0k} \qquad \text{(A3)}$$

Here both j and k go from 1 to n. The value of δr^*_{jk} is simply the displacement of the k-th residual from its 'equilibrium' position r^*_{0k} due to the perturbation generated by dropping the j-th point. In practice, the perturbation matrix is the one to be used in the successive steps of PPA, that should be handled in different subroutines.

The columns of the perturbation matrix are in turn processed by Fourier analysis. In the first subroutine the periodogram is constructed by calculating the coefficients As and Bs according to eq. 15. The values of these coefficients are listed in table 5. The values of $I_k(\tau)$ $(\tau = 0, 1/n, \ldots m/n;$

$m = n/2$) are then readily obtained from eqs. 16 and 17, and are also listed in table 5. In the second subroutine, the autocorrelation function is constructed according to eq. 18. The resulting values of the Cs are listed in table 6. A third subroutine takes care of tapering the Cs according to the Bartlett window (see eq. 26) and then obtaining the relaxation length for each point using the extended trapezoidal rule [20]. The tapered autocorrelation function is calculated as

$$\langle C_k(\alpha)\rangle = C_k(\alpha)[1 - \alpha/q] \qquad \text{(A4)}$$

with $\alpha = 0, 1, \ldots q$, and $q = 5$ in this case. The values of the $\langle C\rangle$s are listed in table 6. Finally, the integral of eq. 19 is calculated by the trapezoidal rule. Since the autocorrelation function is even, one simply has

$$L(k) \approx 2[\langle C_k(0)\rangle/2 + \langle C_k(1)\rangle + \ldots + \langle C_k(q-1)\rangle + \langle C_k(q)\rangle/2] \qquad \text{(A5)}$$

The values of the Ls are listed in table 6.

Acknowledgements

E.D.C. is grateful to Dr. Gary K. Ackers for kindly providing a copy of the thesis of F.C. Mills and to Dr. Stanley J. Gill for helpful discussions. This work was supported by MPI and CNR.

References

1 Y. Bard, Nonlinear parameter estimation (Academic Press, New York, 1974).
2 P.R. Bevington, Data reduction and error analysis for the physical sciences (McGraw-Hill, New York, 1969).
3 M.E. Magar, Data analysis in biochemistry and biophysics (Academic Press, New York, 1972).
4 D.A. Ratkowski, Non-linear regression modeling (Dekker, New York, 1983).
5 M.B. Priestley, Spectral analysis and time series (Academic Press, New York, 1981).
6 P. Bloomfield, Fourier analysis of time series: an introduction (Wiley, New York, 1976).
7 G.E.P. Box and G.M. Jenkins, Time series analysis (Holden-Day, Oakland, CA, 1976).
8 S.R. de Groot and P. Mazur, Non-equilibrium thermodynamics (Dover, New York, 1984).
9 P. Grigolini and F. Marchesoni, Adv. Chem. Phys. 62 (1985) 29.
10 L.P. Kadanoff and J. Swift, Phys. Rev. 166 (1968) 89.
11 K. Kawasaki, Ann. Phys. 61 (1970) 1.
12 J.V. Sengers, Ber. Bunseges. Phys. Chem. 76 (1972) 234.
13 G. Arcovito, C. Faloci, M. Roberti and L. Mistura, Phys. Rev. Lett. 22 (1969) 1040.
14 H.M. Leister, J.C. Allegra and G.F. Allen, J. Chem. Phys. 51 (1969) 3701.
15 A. Stein, S.J. Davidson, J.C. Allegra and G.F. Allen, J. Chem. Phys. 56 (1972) 6164.
16 E. Gulari, A.F. Collings, R.L. Schmidt and C.J. Pings, J. Chem. Phys. 56 (1972) 6169.
17 R.F. Berg and M.R. Moldover, J. Chem. Phys. 89 (1988) 3694.
18 P. Debye, B. Chu and D. Woermann, J. Polym. Sci. A1 (1963) 249.
19 F. Andreasi Bassi, Thesis, Università Cattolica, Rome (1972).
20 W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical recipes (Cambridge University Press, New York, 1986).
21 K. Imai, H. Morimoto, M. Kotani, H. Watari, H. Waka and M. Kuroda, Biochim. Biophys. Acta 534 (1970) 189.
22 K. Imai, Allosteric effects in hemoglobin (Cambridge University Press, Cambridge, U.K., 1982).
23 F.C. Mills, M.L. Johnson and G.K. Ackers, Biochemistry 15 (1976) 5350.
24 A.H. Chu, B.W. Turner and G.K. Ackers, Biochemistry 23 (1984) 604.
25 S. Dolman and S.J. Gill, Anal. Biochem. 87 (1978) 127.
26 S.J. Gill, E. Di Cera, M.L. Doyle, G.A. Bishop and C.H. Robert, Biochemistry 26 (1987) 3995.
27 E. Di Cera, M.L. Doyle, P.R. Connelly and S.J. Gill, Biochemistry 26 (1987) 6494.
28 E. Di Cera and S.J. Gill, Biophys. Chem. 29 (1988) 351.
29 E. Di Cera, M.L. Doyle and S.J. Gill, J. Mol. Biol. 200 (1988) 593.
30 E. Di Cera, M.L. Doyle, M.S. Morgan, R. De Cristofaro, R. Landolfi, B. Bizzi, M. Castagnola and S.J. Gill, Biochemistry 28 (1989) 2631.
31 F.C. Mills, Thesis, The Johns Hopkins University, Baltimore, MD (1978).
32 E. Di Cera and S.J. Gill, Biophys. Chem. 32 (1988) 149.
33 J.C. Nieuwoudt and J.V. Sengers, J. Chem. Phys. 90 (1989) 457.
34 J. Timmermans, Physico-chemical constants of pure organic compounds (Elsevier, New York, 1950).
35 C.B. Willingham, W.J. Taylor, J.M. Pignocco and F.D. Rossini, J. Res. Nat. Bur. Standards 35 (1945) 219.