Influence of experimental errors on the determination of flux control coefficients from transient metabolite concentrations

Magnus EHLDE and Guido ZACCHI

Department of Chemical Engineering 1, University of Lund, P.O. Box 124, S-221 00 Lund, Sweden

The influence of experimental errors on the determination of flux control coefficients from transient metabolite concentrations with the method proposed by Delgado and Liao [(1992) *Biochem. J.* **282**, 919–927] has been investigated by using Monte Carlo simulations. The method requires least-squares fitting of the transient metabolite concentrations. Three different fitting methods have been evaluated. Simulated metabolite concentrations of a fictive metabolic pathway were scattered randomly, emulating experimental errors, before performing the fits. This

was repeated a large number of times; the mean values and standard deviations of the resulting control coefficients are reported. The results show that the proposed method for determining control coefficients is too sensitive to experimental errors to be practicable, with theoretically justified fitting methods. This is in particular due to the high degree of correlation between the concentrations. An alternative *ad hoc* fitting method produced biased mean values of the estimates of the control coefficients, but with remarkably low standard deviations.

INTRODUCTION

The control of metabolic pathways is a complex mechanism, which can only be analysed by a quantitative approach. Such an approach is Metabolic Control Analysis (MCA) [1–3], where flux and concentration control coefficients are defined. A flux control coefficient quantifies the degree of control exerted by an enzyme on a steady-state flux in the pathway. Although control coefficients are defined for infinitesimal changes around specific steady states, their values have a conceptual importance and may increase the understanding of the control of metabolic pathways.

A major obstacle in the application of MCA is associated with the experimental determination of control coefficients. If the stoichiometry of a metabolic pathway and the kinetics of the individual enzymic reactions, i.e. the elasticity coefficients, are known, the control coefficients can be calculated by using the theorems of MCA. However, the kinetics of all enzymes in a metabolic pathway are seldom available and the determination of individual enzyme kinetics, either *in vitro* or *in vivo*, is very time-consuming. Furthermore, if any kinetic interactions are overlooked, for instance an allosteric interaction, the results will be erroneous. For this reason, the results must be complemented with direct measurement of at least one control coefficient. The direct measurement can be done by altering the activity of one enzyme in the pathway, by genetic means, and measuring the effect on the steady-state flux. This also is a very time-consuming process. The results will be based on a few steady-state measurements, so a high degree of experimental accuracy is required. Various experimental approaches suggested for the determination of elasticity and control coefficients together with some applications of these approaches have been reviewed by Fell [3].

Although the control coefficients of a metabolic pathway are valid only at steady state, they are basically kinetic properties. Delgado and Liao [4–6] have shown that if certain assumptions are fulfilled, the control coefficients can be related to the metabolite concentrations of the pathway obtained in a transient state, i.e. after a perturbation of the steady state of the system. The use of such an approach to determine control coefficients is appealing, because transient data contain more information about the kinetics of a pathway than do a few steady-state measurements.

The method of Delgado and Liao is based on linearization of the individual kinetic rate equations in the pathway. They have derived a relationship that relates the metabolite concentrations during a transient in a linear manner, where the control coefficients are parameters. All control coefficients are determined simultaneously by fitting the transient metabolite concentrations to this relationship by least-squares fitting, yielding the best estimates of the control coefficients. The relationship is strictly valid only if the individual rate equations can be approximated as linear in the transient region. Because enzymic rate equations are known to be usually non-linear, this is a potential source of error. Experimental difficulties associated with the method include the determination of all metabolite concentrations in the pathway during the transient. Also, if the number of enzymes in the pathway is large, the fitting procedure will be very sensitive to experimental errors in these determinations.

In the present study we investigate the sensitivity of the determined control coefficients to experimental errors in the measured metabolite concentrations. This was done with simulated data from the same fictive pathway used by Delgado and Liao [5]. The simulated concentrations were scattered randomly before they were fitted to a slightly modified version of the linear relationship derived by Delgado and Liao. The whole procedure was repeated a large number of times; the mean values and standard deviations of the resulting control coefficients, using three different fitting methods, are reported.

THEORY

Metabolic control analysis

The control of the steady-state flux through a specific enzymic reaction in a metabolic pathway is shared between all enzymes in the pathway. The amount of control of the flux through enzyme k , J_k , by the activity of enzyme j , e_j , can be quantified by the flux control coefficient, defined as:

$$C_{e_j}^J = \left. \frac{dJ_k}{de_j} \frac{e_j}{J_k} \right|_{ss} \quad (1)$$

The subscript ss indicates that the expression should be evaluated at steady state. A control coefficient is specific to the steady state

at which it is defined, and is a function of all parameters affecting the steady state. The flux control coefficients for the flux J_k through enzyme k with respect to each enzyme activity in the pathway will add up to unity according to the summation theorem:

$$\sum_{j=1}^n C_{e_j}^J = 1 \quad (2)$$

where n is the number of enzymes in the pathway. Because in the following we shall always refer to the same steady-state flux, the flux control coefficients will, for simplicity, be written as C_j^J .

Transient method

Delgado and Liao devised a method for determining control coefficients of metabolic pathways, using measurements of metabolite concentrations in a transient state. The method is based on linearization of the enzymic rate equations of the pathway. They first derived an expression to relate the transient flux through each enzyme of the pathway to the flux control coefficients of the pathway, as follows [4]:

$$\sum_{j=1}^n C_j^J \left(\frac{v_j(t)}{J_j} \right) = 1 \quad (3)$$

where n is the number of enzymes in the pathway and $v_j(t)$ and J_j are the transient and steady-state fluxes through enzyme j , respectively.

In a later refinement of their method [5], they derived an integral form of eqn. (3) in which the transient fluxes are implicitly calculated and that relates the transient metabolite concentrations of the pathway by a linear constraint:

$$\sum_{j=1}^m \alpha_j (x_j(t) - x_j(0)) = t \quad (4)$$

where m is the number of metabolites, t is the time and $x_j(0)$ is the concentration of metabolite j at an initial time point. The flux control coefficients of the pathway are related to the coefficients α_j by:

$$[C_1^J \ C_2^J \ \dots \ C_n^J] = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_m] \mathbf{A} \mathbf{J} \quad (5)$$

where n is the number of enzymes in the pathway, $\mathbf{J}$ is an $n \times n$ diagonal matrix, containing the steady-state fluxes through each enzyme, and $\mathbf{A}$ is the full $m \times n$ stoichiometric matrix, containing stoichiometric coefficients for both internal and external metabolites. The element in the i th row and j th column of $\mathbf{A}$ is the stoichiometric coefficient of metabolite i in the reaction j .

Delgado and Liao have also derived a relation that relates metabolite concentrations during a transient to the concentration control coefficients of the pathway [6]. The present study deals only with the determination of flux control coefficients, which are generally of most interest.

The relations (4) and (5) are valid if the following assumptions are fulfilled: (i) the external metabolites do not affect the pathway kinetics, or these metabolites are buffered; (ii) the kinetic rate equations for the fluxes and enzymic reactions in the pathway can be approximated as linear in the transient region; (iii) it is theoretically possible to calculate each flux and reaction rate in the pathway from the transient metabolite concentrations; and (iv) the metabolites are homogeneously distributed in the cell.

The flux control coefficients are determined by inducing a step or a pulse transient in the pathway and measuring all metabolite concentrations in the pathway as a function of time, as the system relaxes towards steady state. The measured transient

metabolite concentrations are fitted to eqn. (4) by least-squares fitting. This yields the best estimates of the coefficients α_j , which are then used in eqn. (5) to calculate the flux control coefficients.

Least-squares fitting to eqn. (4) is not possible if linear constraints are present among the metabolite concentrations, i.e. if one concentration is a linear combination of other concentrations. Such constraints may be present owing to stoichiometric properties of the pathway. As an example, if the concentrations are expressed in mass units, they are constrained by the mass-balance equation:

$$\sum_{j=1}^m (x_j(t) - x_j(0)) = 0 \quad (6)$$

If the concentrations are expressed in molar units, a similar constraint can be derived. This means that each metabolite concentration is a linear combination of the other concentrations. Other stoichiometric constraints may be present, which can all be deduced from the stoichiometry matrix $\mathbf{A}$. The number of independent metabolites in a pathway is determined by the rank of the stoichiometry matrix. Before performing a least-squares fit, the dependent metabolite concentrations and the corresponding α values should be deleted from eqn. (4). Also, the corresponding rows in $\mathbf{A}$ should be deleted when calculating the control coefficients using eqn. (5). The concentrations chosen as dependent should be those determined with the greatest uncertainty. These values will be implicitly determined from other concentrations, determined with higher accuracy, during the calculations.

Even if linear constraints arising from stoichiometric properties are not present between the variables in eqn. (4) (as the dependent metabolites have been deleted), high degrees of linear correlation may still be present owing to the kinetic properties of the pathway. If, for instance, one reaction in the pathway is much faster than the others, this reaction will reach a quasi-equilibrium very early during the transient. This means that the concentrations of the substrate and the product of this reaction will be related by the equilibrium constant of the reaction. If this is known *a priori* these metabolites can be lumped together in a pool, as shown by Delgado and Liao [5]. In Appendix A it is also illustrated how linear dependencies can arise if the variation of one metabolite is much smaller than the variations of the other metabolites during the transient. The degree of linear correlation between the variables is a critical quantity that greatly influences the results of the fits, and it may be decreased by using data from several different transients.

The matrix resulting after the deletion of the dependent rows from $\mathbf{A}$ is called $\mathbf{A}_R$. This matrix is invertible, because all rows are independent. Using this fact, we have found it convenient to rewrite eqn. (4) by solving the vector of α values from eqn. (5) and inserting the result into eqn. (4), yielding:

$$\sum_{j=1}^n C_j^J y_j(t) = t \quad (7)$$

where the new variables $y_j(t)$ are calculated as:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \mathbf{J}^{-1} \cdot \mathbf{A}_R^{-1} \cdot \begin{pmatrix} x_1(t) - x_1(0) \\ x_2(t) - x_2(0) \\ \vdots \\ x_n(t) - x_n(0) \end{pmatrix} \quad (8)$$

The values of the new variables $y_j(t_i)$, which in the following are denoted y_{ij} , are constructed from the concentration measurements of the independent metabolites by using eqn. (8), and the estimates of the control coefficients can be determined as a direct result of least-squares fitting to eqn. (7).

FITTING METHODS

The least-squares fitting of the experimental data to eqn. (7) can be performed in several ways. Three different methods have been evaluated, referred to in the text as fitting methods 1, 2 and 3. In all equations, a hat (^) above a symbol denotes an estimate of the quantity represented by that symbol.

In traditional multi-dimensional least-squares fitting, as used by Delgado and Liao [5], the best estimates of the parameters C_j^j are determined by minimization of the merit function:

$$Q_1 = (\hat{\mathbf{t}} - \mathbf{t})^T \sigma_t^{-1} (\hat{\mathbf{t}} - \mathbf{t}) \quad (9)$$

where $\mathbf{t}$ is a $p \times 1$ vector containing the p experimental values t_i , and $\hat{\mathbf{t}}$ is a $p \times 1$ vector containing the estimated values $\hat{t}_i$. These latter values are calculated from the estimated parameters values $\hat{C}_j^j$ as:

$$\hat{t}_i = \sum_{j=1}^n \hat{C}_j^j y_{ij} \quad i = 1, \dots, p \quad (10)$$

σ_t in eqn. (9) is the $p \times p$ variance matrix, which contains the variances of the time (t_i) measurement errors in the diagonal elements and the covariances in the off-diagonal elements. The variables $y_{i1}, y_{i2}, \dots, y_{in}$ are calculated from eqn. (8) with the experimental concentration values at time t_i .

The errors in the time measurements are likely to be uncorrelated, i.e. the off-diagonal elements of the variance matrix are zero. If the standard deviations of the time measurements can be approximated to be the same for each experimental point, the variance matrix can be written as an identity matrix multiplied by a scalar. This means that it can be omitted from eqn. (9), which then reduces to:

$$Q'_1 = (\hat{\mathbf{t}} - \mathbf{t})^T (\hat{\mathbf{t}} - \mathbf{t}) = \sum_{i=1}^p (\hat{t}_i - t_i)^2 \quad (11)$$

This method, which is here called method 1, is almost always used whenever a least-squares problem is encountered. The merit function has only one minimum, which is easy to find because it can be analytically derived [7]. However, care must be taken because an underlying assumption of the method is that the independent variables (i.e. y_{ij}) are known exactly and that the dependent variable (i.e. t_i) is measured with a gaussian-distributed experimental error. This is not always perfectly true and in this case it is certainly not. The errors (which may be approximated as gaussian distributed) reside mainly in the concentration measurements, i.e. in the variables y_{ij} , whereas the time, t_i , can be measured with relative exactness. This implies that this method produces biased estimates of the control coefficients [8].

Lybanon [9] has presented a generalized least-squares fitting method, which is recommended when more than one variable contains errors. The method is here called method 2 and is based on the results of Jefferys [8]. The concept of the method is to determine the best estimates not only for the parameters C_j^j , but also for the error-containing variables y_{ij} . The least-squares problem is formulated as minimization of the merit function:

$$Q_2 = (\hat{\mathbf{y}} - \mathbf{y})^T \sigma_y^{-1} (\hat{\mathbf{y}} - \mathbf{y}) \quad (12)$$

subject to the p constraints:

$$\mathbf{f}(\hat{\mathbf{y}}, \hat{\mathbf{C}}) = \mathbf{0} \quad (p \times 1) \quad (13)$$

where p is the number of experimental points, $\mathbf{y}$ is a $p \cdot n \times 1$ vector containing the original variable values, $\hat{\mathbf{y}}$ is a $p \cdot n \times 1$ vector containing the best estimates of the variable values, and σ_y is the $p \cdot n \times p \cdot n$ variance matrix containing the variances and

covariances of the errors in $\mathbf{y}$. The variables in the vector $\mathbf{y}$ and the corresponding best estimates in the vector $\hat{\mathbf{y}}$ here represent the quantities calculated from eqn. (8), and the values in the vectors are organized as:

$$\mathbf{y} = (y_{1,1} y_{1,2} \dots y_{1,n} \quad y_{2,1} y_{2,2} \dots y_{2,n} \dots y_{p,1} y_{p,2} \dots y_{p,n})^T \quad (14)$$

$\hat{\mathbf{C}}$ is an $n \times 1$ vector, containing the best estimates, $\hat{C}_j^j$, of the parameters. In the transient method the constraints (13) are:

$$f_i(\hat{\mathbf{y}}, \hat{\mathbf{C}}) = \sum_{j=1}^n C_j^j \hat{y}_{ij} - t_i = 0 \quad i = 1, \dots, p \quad (15)$$

The constrained minimization can be solved by finding values of $\hat{\mathbf{y}}$, $\hat{\mathbf{C}}$ and $\hat{\boldsymbol{\mu}}$ that minimize:

$$Q'_2 = (\hat{\mathbf{y}} - \mathbf{y})^T \sigma_y^{-1} (\hat{\mathbf{y}} - \mathbf{y}) + \mathbf{f}^T(\hat{\mathbf{y}}, \hat{\mathbf{C}}) \hat{\boldsymbol{\mu}} \quad (16)$$

where $\hat{\boldsymbol{\mu}}$ is a $p \times 1$ vector containing Lagrange multipliers. The merit function (16) cannot be minimized analytically and an iterative approach must be used. Lybanon [9] has outlined how to apply Newton's method to find the minimum. It is crucial to assign good initial values to the parameters at the start of the minimization, because several local minima may exist, depending on the number of variables and the nature and size of the experimental errors. After the reformulation of eqn. (4) to eqn. (7) we know that, for an unbranched pathway, the values of the parameters, i.e. C_j^j , attain values between 0 and 1. This facilitates the choice of initial values, but the results of the fits do not change as a result of the reformulation. We can also restrict the search for the minimum to the domain where $C_j^j \geq 0$ and $C_j^j \leq 1$. The minimum found might be a local minimum, but it is probably the minimum of interest because other values of the control coefficients have no physical significance.

Method 1 and method 2 both rely on a sound theoretical basis and are derived from the concept of maximum likelihood [7], although the underlying assumption of method 1 of error-free independent variables is violated in this case. The third method investigated, method 3, is an *ad hoc* method, where the fact that the errors reside mainly in the dependent variables, y_{ij} , is taken into account. Contrary to method 2, only the best estimates of the parameters $\hat{C}_j^j$ are determined. This means, as in method 1, the estimates are based on erroneous values of the dependent variables. The fitting is performed by finding values of the parameters $\hat{C}_j^j$ that minimize the function:

$$Q_3 = (\hat{\mathbf{y}} - \mathbf{y})^T \sigma_y^{-1} (\hat{\mathbf{y}} - \mathbf{y}) \quad (17)$$

where σ_y has the same significance as in method 2, and the values of the elements in $\hat{\mathbf{y}}$ are calculated by rearranging eqn. (7) to:

$$\hat{y}_{ij} = \frac{1}{C_j^j} \left(t_i - \sum_{k=1}^{n, k \neq j} C_k^j y_{ik} \right) \quad i = 1, \dots, p \quad j = 1 \dots n \quad (18)$$

If all measurements are independent (zero covariances) and the standard deviations can be considered to be equal for all variables, y_{ij} , eqn. (17) reduces to:

$$Q'_3 = \sum_{i=1}^p \sum_{j=1}^n (\hat{y}_{ij} - y_{ij})^2 \quad (19)$$

This merit function may also contain several local minima, so an iterative approach to minimization is required.

RESULTS

The influence of errors in the simulated experimental data on the resulting best estimates of the control coefficients, determined by the three fitting methods described above, was investigated by

Table 1 Results of Monte Carlo simulation of data with zero correlation

Mean values and standard deviations of 10 000 fits. Data are scattered with a standard deviation of 0.15.

	Correct values	Method 1		Method 2		Method 3	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
C_1	0.20	0.194	0.093	0.204	0.099	0.305	0.061
C_2	0.40	0.392	0.067	0.401	0.071	0.415	0.063
C_3	0.50	0.488	0.067	0.500	0.071	0.501	0.066
C_4	0.60	0.587	0.066	0.601	0.070	0.594	0.066
C_5	0.80	0.761	0.084	0.797	0.090	0.765	0.085

using a Monte Carlo simulation method. Random gaussian-distributed errors were imposed on simulated metabolite concentrations before performing the fits.

To make a basic comparison between the three fitting methods, data were constructed that satisfied eqn. (7) with $n = 5$ and with zero linear correlation between the variables y_1 to y_5 . Zero correlation will represent the 'best case', which cannot be obtained from a real or simulated transient experiment, where the variables are more or less correlated. The number of points was 12 ($p = 12$) and all variable values (y_{ij}) ranged between 0 and 1. The values of the parameters C_j^j (which in this case do not represent control coefficients) used when constructing the data are shown in Table 1. Before each fit, random gaussian-distributed errors, with a standard deviation of 0.15, were imposed on the variable values. The errors in the variables were non-correlated and because the standard deviations of the errors were equal for each variable, the variance matrix could be omitted from the merit functions (see the Theory section). Here it must be emphasized that correlation between errors in variables is different from correlation between the variables themselves. The errors of two variables may be uncorrelated whereas the variables themselves are correlated, and vice versa.

As pointed out by Jefferys [8], method 1 yields a biased estimate of the parameters. The method underestimates the values of all coefficients whereas method 2 yields the correct mean values. Method 3 also produces a biased estimate of the parameters, but in this case the large coefficients are underestimated whereas the small coefficients are overestimated. For this system the standard deviations of the estimates are comparable for all three fitting methods. However, for method 1 and method 3 an increase in the standard deviation of the random errors in the variables implies not only an increase in the standard deviation of the estimated parameters, but also a shift in the mean values further away from the correct values.

To evaluate the control coefficients determined from transient metabolic concentrations, data were taken from simulations of a fictive metabolic pathway. The simulations were performed with the metabolic simulation program MIST, which is a tool for dynamic simulation of user-defined metabolic pathways [10]. When the control coefficients are determined from transient concentration measurements, two sources of error can corrupt the results: systematic errors and random errors. Systematic errors arise if any of the underlying assumptions of the transient method are not correct. The aim of this study was to investigate only the influence of random errors introduced by errors in the experimental concentration measurements. For this reason, the pathway simulated was chosen such that the assumptions of the transient method were applicable, including the assumption of

Figure 1 Stoichiometry of the simulated pathway

X_1 and X_6 represent external metabolites, X_2 to X_5 internal metabolites and E_1 to E_5 enzymes.

Table 2 Steady-state values of the simulated pathway

X_2	X_3	X_4	X_5	J
0.189	0.074	0.151	0.071	0.024

Figure 2 Progression of the metabolite concentrations during transient 1

Symbols: $\circ$, x_2 ; $\square$, x_3 ; $\triangle$, x_4 ; $\bullet$, x_5 ; $\blacksquare$, $x_6/10$.

linear kinetic rate equations. The pathway was identical to that used by Delgado and Liao [5] in their example 1 and has the stoichiometry shown in Figure 1.

The kinetic rate equations of the enzymes are given by:

$$v_1 = -0.4x_2 + 0.1 \quad (20a)$$

$$v_2 = 0.5x_2 - 0.8x_3 - 0.3x_5 + 0.01 \quad (20b)$$

$$v_3 = 0.6x_3 - 0.2x_4 + 0.01 \quad (20c)$$

$$v_4 = 0.15x_4 - 0.12x_5 + 0.01 \quad (20d)$$

$$v_5 = 0.2x_5 + 0.01 \quad (20e)$$

The concentrations of the external substrate, x_1 , and of the external product, x_6 , never reach a steady value but because these concentrations do not affect the rate of any enzyme in the pathway, the system is able to reach a quasi-steady state. The flux, J , and the metabolite concentrations, x_j , obtained at the quasi-steady state are shown in Table 2.

A transient was induced in the pathway by setting initial values of the concentration x_1 to 2 and the remaining concentrations to 0 at the start of the simulation. A simulation was started and the concentrations were recorded at 12 equally spaced times as the system relaxed towards the quasi-steady state. When the new variables, y_{ij} , were constructed (see eqn. (8)), the concentration

Figure 3 Progression of the new variables during transient 1

 Symbols: $\circ$, y_1 ; $\square$, y_2 ; $\triangle$, y_3 ; $\bullet$, y_4 ; $\blacksquare$, y_5 .

Table 3 Results of Monte Carlo simulation with data from transients 1

Mean values and standard deviations of 10000 fits. Concentration values are scattered with a relative standard deviation of 0.01%

	Correct values	Method 1		Method 2		Method 3	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
C_1	0.158	0.157	0.026	0.158	0.017	0.156	0.001
C_2	0.126	0.127	0.125	0.126	0.080	0.136	0.005
C_3	0.168	0.167	0.134	0.168	0.086	0.158	0.005
C_4	0.224	0.225	0.078	0.224	0.051	0.230	0.004
C_5	0.324	0.323	0.043	0.324	0.028	0.320	0.003

x_1 was chosen as dependent and deleted from the calculations. The progressions of the metabolite concentrations and of the new variables during the transient are shown in Figures 2 and 3 respectively.

The Monte Carlo simulation was performed by imposing relative random gaussian-distributed errors, with a standard deviation of 0.01%, on the metabolite concentration values obtained from the simulation (see Figure 2), before constructing the new variables y_{ij} (see Figure 3) and performing the fits. The errors in the concentrations are uncorrelated but owing to the linear combination of the concentrations in the construction of the new variables [see eqn. (8)], the errors in these are correlated and the variance matrix is not diagonal. In Appendix B it is shown how the covariances, i.e. the off-diagonal elements of the variance matrix, are calculated. The results of the Monte Carlo simulation are shown in Table 3, together with the correct values of the control coefficients. These were calculated by the simulation programme MIST, by using elasticity coefficients and the theorems of MCA.

Even for the very small relative errors imposed on the simulated experimental values, the standard deviations of the control coefficients determined by method 1 and method 2 are high. They are slightly lower for method 2, which is the correct method to use from a theoretical point of view. This method also yields the correct mean values. The biased estimations of method 1, demonstrated in the previous example, are very small in the example, owing to the small errors imposed. In method 3, however, the mean values are evidently biased, but the standard

Figure 4 Progression of the metabolite concentrations during transient 2

 Symbols: $\circ$, x_2 ; $\square$, x_3 ; $\triangle$, x_4 ; $\bullet$, x_5 ; $\blacksquare$, $x_6/10$.

Figure 5 Progression of the new variables during transient 2

 Symbols: $\circ$, y_1 ; $\square$, y_2 ; $\triangle$, y_3 ; $\bullet$, y_4 ; $\blacksquare$, y_5 .

Table 4 Results of Monte Carlo simulation with data from transients 1 and 2

Mean values and standard deviations of 10000 fits. Concentration values are scattered with a relative standard deviation of 0.5%

	Correct values	Method 1		Method 2		Method 3	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
C_1	0.158	0.155	0.041	0.158	0.015	0.150	0.004
C_2	0.126	0.133	0.100	0.125	0.047	0.145	0.004
C_3	0.168	0.170	0.037	0.168	0.024	0.151	0.004
C_4	0.224	0.211	0.177	0.225	0.064	0.246	0.005
C_5	0.324	0.331	0.098	0.323	0.041	0.311	0.009

deviations are remarkably small compared with the other two methods.

A high linear correlation between the variables increases the sensitivity of the fittings to the errors in the variables. This is so

Table 5 Results of Monte Carlo simulation with data from transient 1 and 2

Mean values and standard deviations of 10000 fits. Concentration values are scattered with a relative standard deviation of 5.0%.

	Original values	Method 3	
		Mean	S.D.
C_1	0.158	0.166	0.012
C_2	0.126	0.179	0.006
C_3	0.168	0.183	0.004
C_4	0.224	0.208	0.021
C_5	0.324	0.214	0.028

for the transient shown in Figure 3, where the mean value of the correlation coefficients between the variables is 0.990. This is due to the large difference between the variations in the concentrations during the transient. The concentration x_6 (the values are divided by 10 in Figure 2) do reach a steady value but grow continuously. The variation in this concentration is thus much larger than the variations in the other concentrations and because this concentration is present in all the new variables, this results in the high correlation (see Appendix A).

To decrease the correlation, a second transient was simulated, in which the concentrations were changing in the opposite direction. The transient was induced by setting initial values of all the concentrations to 0.3 at the start of the simulation, and the concentration values were recorded at 12 equally spaced times as the system relaxed towards the quasi-steady state indicated in Table 2. The progressions of the metabolite concentrations and of the new variables during the second transient are shown in Figures 4 and 5 respectively.

When the data from the two transients are combined, the correlation between the variables decreases drastically: the new mean value of the correlation coefficients was 0.709. A Monte Carlo simulation was performed with data from both transients simultaneously, which radically decreased the influence of random errors on the fitting results. The standard deviation of the errors imposed on the concentrations was increased to 0.5%. The results are compiled in Table 4.

Although the magnitude of the relative errors imposed on the concentration values was increased 50-fold over the previous example, the two results are comparable. The improvements are mainly due to the decreased correlation of the variables. The increased number of experimental points has itself a minor effect on the results. The use of data from even more transients lowers the correlation by only small amounts. More simulated experimental data, either from the same transients or from yet another transient, will thus not significantly improve the accuracy of the results. Again, method 1 yields coefficients that are very slightly biased but with large standard deviations, whereas method 2 yields the correct mean values with lower standard deviations. Method 3 yields coefficients that are even more biased than in the previous example, but still with very small standard deviations.

If the standard deviation of the imposed errors is increased to more than 1%, the standard deviations of the coefficients determined by methods 1 and 2 become very high (more than 1). Method 3 still yields biased estimates of the coefficients but with very low standard deviations. A Monte Carlo simulation was performed, in which the standard deviation of the imposed errors

was 5%. Fits were made by using fitting method 3. The results are compiled in Table 5.

DISCUSSION

The only measurable variables in a metabolic pathway that reveal information about the kinetic properties of the pathway are metabolite concentrations. The transient concentrations are complex functions of the kinetic properties of the pathway, as are the control coefficients. The transient method of Delgado and Liao differs from other methods suggested for the determination of control coefficients [3] in the sense that transient measurements are used instead of steady-state measurements. This is appealing, because transient data obviously contain more information about the kinetics of a pathway than do steady-state measurements.

However, measuring transient metabolite concentrations *in vivo* is not a trivial task and the concentration measurements are bound to have significant experimental errors. The results above indicate that even small errors drastically influence the results obtained.

If no errors are present in the simulated concentrations, all three fitting methods investigated yield the correct control coefficients. This is due to the perfectly linear kinetic rate equations used. If the rate equations were non-linear, they would still be approximated as linear in the range of the transient, which would introduce a systematic error. As always, when derivatives of non-linear functions are determined by linear approximations (e.g. other methods for determination of control coefficients), a compromise must be made to minimize the effect of both systematic and random errors. Increasing the size of the transient, i.e. the magnitudes of the variations in the metabolite concentrations, decreases the influence of experimental errors. This, however, also increases the influence of systematic errors, because the linear approximations might not be justified far away from the steady state.

It must thus be emphasized that the simulated data used in the fittings above represent a highly idealized situation. The two main reasons for this are that the rate equations are linear, which eliminates methodical errors, and that in a simulated system it is possible to make independent transients, yielding data with low correlations. In a real system, where the number of ways of including a transient is limited, it might be difficult to generate data with low correlations. Also, the number of enzymes in real systems is often more than five, which will also make the fits more sensitive to errors.

Fitting method 2 is shown to be superior to method 1, both from a theoretical point of view and from the results obtained from the fits. Method 2 always yields the correct mean values of the estimated parameters and with lower standard deviations than in method 1. However, despite the idealized situation, the effects of the random errors are severe. This would make the method useless in a practical situation, where the experimental errors are likely to be larger than in the Monte Carlo simulations performed.

For the *ad hoc* fitting method 3, even relatively large experimental errors have only a minor effect on the standard deviation of the estimated parameters. The mean values of the parameters are biased, but for an experimental error in the concentrations of 5%, the results are still acceptable. Again, it must be emphasized that the simulated variables are very slightly correlated, and that, if the rate equations were non-linear, the mean values would have been shifted further away from the correct values, owing to the systematic error.

In conclusion, determining control coefficients by the transient method proposed by Delgado and Liao would not be possible in

a real situation, with the use of theoretically sound fitting methods. Further investigations will reveal the reason for the biased estimates and the low standard deviations of fitting method 3, and whether this method could be used in a real situation.

We thank the Swedish National Board for Industrial and Technical Development (NUTEK) for financial support.

REFERENCES

- 1 Kacser, H. and Burns, J. A. (1973) Symp. Soc. Exp. Biol. **27**, 65–104
- 2 Heinrich, R. and Rapoport, T. A. (1974) Eur. J. Biochem. **42**, 89–95
- 3 Fell, D. A. (1992) Biochem. J. **286**, 313–330
- 4 Delgado, J. P. and Liao, J. C. (1991) Biotechnol. Prog. **7**, 15–20
- 5 Delgado, J. P. and Liao, J. C. (1992) Biochem. J. **282**, 919–927
- 6 Delgado, J. P. and Liao, J. C. (1992) Biochem. J. **285**, 965–972
- 7 Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) in Numerical Recipes in FORTRAN: the Art of Scientific Computing, 2nd edn., pp. 650–670, Cambridge University Press, Cambridge
- 8 Jefferys, W. H. (1980) Astron. J. **85**, 177–181
- 9 Lybanon, M. (1984) Am. J. Phys. **52**, 22–26
- 10 Ehde, M. and Zacchi, G. (1995) Comp. Appl. Biosci. **11**(2), 201–207

APPENDIX A

In this appendix eqn. (7) is applied to the stoichiometry of the pathway depicted in Figure 1. The full stoichiometric matrix, $\mathbf{A}$, of the pathways is given by:

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A1})$$

The rank of the matrix is 5 and any of the rows can be considered to be dependent and deleted. The result will, in principle, be the same because the deleted concentration will be implicitly calculated from the independent concentrations by the mass-balance equation (6). If the first row of $\mathbf{A}$ is deleted and the resulting matrix $\mathbf{A}_R$ is inverted, the result is:

$$\mathbf{A}_R^{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A2})$$

Because in a linear pathway all steady-state fluxes are equal, multiplication by the matrix $\mathbf{J}^{-1}$ in eqn. (7) is equivalent to division by the scalar J . The values of the new variables $y_1, y_2, \dots, y_5$ can be calculated as:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \frac{1}{J} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta x_5 \\ \Delta x_6 \end{pmatrix} =$$

$$\frac{1}{J} \cdot \begin{pmatrix} \Delta x_2 + \Delta x_3 + \Delta x_4 + \Delta x_5 + \Delta x_6 \\ \Delta x_3 + \Delta x_4 + \Delta x_5 + \Delta x_6 \\ \Delta x_4 + \Delta x_5 + \Delta x_6 \\ \Delta x_5 + \Delta x_6 \\ \Delta x_6 \end{pmatrix} \quad (\text{A3})$$

where $\Delta x_j = x_j(t) - x_j(0)$. Eqn. (7) can now be written as:

$$C_1^J \sum_{j=2}^6 \Delta x_j + C_2^J \sum_{j=3}^6 \Delta x_j + C_3^J \sum_{j=4}^6 \Delta x_j + C_4^J \sum_{j=5}^6 \Delta x_j + C_5^J \Delta x_6 = J \cdot t \quad (\text{A4})$$

From eqn. (A4) it follows that if the variations of the different concentrations Δx_j are not of the same magnitude, the new variables, i.e. the sums in eqn. (A4), will be highly correlated. If for instance the variation of Δx_5 is much smaller than the others, the third and fourth sum in eqn. (A4) will be almost equal and the corresponding variables y_4 and y_5 will be highly linearly correlated.

APPENDIX B

The covariance between two variables y_1 and y_2 can be calculated as:

$$\text{Cov}(y_m, y_k) = E(y_m \cdot y_k) - E(y_m) \cdot E(y_k) \quad (\text{B1})$$

where $E(y_k)$ is the expected value of variable y_k . Suppose that the n variables $y_1, \dots, y_n$ are constructed from the n variables $x_1, \dots, x_n$, which contain uncorrelated random errors, by the linear combination:

$$y_k = \sum_{j=k}^n x_j \quad k = 1, \dots, n \quad (\text{B2})$$

We now seek the covariances between the new variables. From eqn. (B2) the following relation between two variables, y_k and y_m , can be described:

$$y_m = y_k + \sum_{j=m}^{k-1} x_j \quad k > m \quad (\text{B3})$$

Insertion of this relation into eqn. (B1) yields:

$$\begin{aligned} \text{Cov}(y_m, y_k) &= E\left(y_k y_k + y_k \sum_{j=m}^{k-1} x_j\right) - E\left(y_k + \sum_{j=m}^{k-1} x_j\right) E(y_k) \\ &= E(y_k y_k) - E(y_k) E(y_k) + E\left(y_k \sum_{j=m}^{k-1} x_j\right) \\ &\quad - E(y_k) E\left(\sum_{j=m}^{k-1} x_j\right) \\ &= \text{Cov}(y_k, y_k) + \text{Cov}\left(y_k, \sum_{j=m}^{k-1} x_j\right) \quad k > m \end{aligned} \quad (\text{B4})$$

Because y_k and $\sum_{j=m}^{k-1} x_j$ do not share any common variables x_j , the errors in these two quantities are uncorrelated and the covariance between them is 0. The result is thus:

$$\begin{aligned} \text{Cov}(y_m, y_k) &= \text{Cov}(y_k, y_m) = \text{Cov}(y_k, y_k) = \text{Var}(y_k) \\ &= \sum_{j=k}^n \text{Var}(x_j) \quad k > m \end{aligned} \quad (\text{B5})$$

where $\text{Var}(x_j)$ is the variance of the error in variable x_j .