

## **Optimal Selection of Metabolic Fluxes for *in vivo* Measurement. I. Development of Mathematical Methods**

JOANNE M. SAVINELL\* AND BERNHARD O. PALSSON†

*Cellular Biotechnology Laboratory, Department of Chemical Engineering,  
The University of Michigan, Ann Arbor, MI 48109-2136, U.S.A.*

*(Received on 16 May 1991, Accepted on 10 October 1991)*

The measurement of uptake and secretion rates is often not sufficient to allow the calculation of all internal metabolic fluxes. Measurements of internal fluxes are needed and these additional measurements are used in conjunction with mass-balance equations to calculate the complete metabolic flux map. A method is presented that identifies the fluxes that should be selected for experimental measurement, and the fluxes that can be computed using the mass-balance equations. The criterion for selecting internal metabolic fluxes for measurement is that the values of the computed fluxes should have low sensitivity to experimental error in the measured fluxes. A condition number indicating the upper bound on this sensitivity, is calculated based on stoichiometry alone. The actual sensitivity is dependent on both the flux measurements and the error in flux measurements, as well as the stoichiometry. If approximate physiologic ranges of fluxes are known a realistic sensitivity can be computed. The exact sensitivity cannot be calculated since the experimental error is usually unknown. The most probable value of the actual sensitivity for a given selection of measured fluxes is estimated by selecting a large number of representative error vectors and calculating the actual sensitivity for each of these. A frequency distribution of actual sensitivities is thus obtained giving a representative range of actual sensitivities for a particular experimental situation.

### **Introduction**

The study of steady-state fluxes within the cell's metabolic network and the changes in these fluxes in response to environmental stimuli provides information about the regulation of intermediary metabolism and is important for a number of clinical and biotechnological applications. Metabolic networks can be divided into two general classifications. The first type of network has a large ratio of end products to branch points, so that measurements of product secretion and nutrient uptake rates coupled with mass balances on the intermediates allow the fluxes throughout the network to be estimated. Networks which include only major metabolic and product formation pathways are examples of this type of network. This type of network is commonly used when analyzing bacterial metabolism and product synthesis.

\* Present address: Department of Chemical Engineering, Cleveland State University, Cleveland, OH 44115, U.S.A.

† Author to whom correspondence should be addressed.

The second type of metabolic network has a large number of branch points and relatively few secreted products. The steady-state fluxes through these networks cannot usually be obtained solely from measurements of the fluxes of components entering and leaving the cell. Additional measurements of fluxes within the cell must normally be made, typically by the use of  $^{14}\text{C}$  labeling or *in vivo* NMR techniques. Networks describing mammalian cell metabolism and detailed networks of bacterial metabolism are examples of this type of network.

Examples of the first type of metabolic network are found in butanediol and mixed-acid bacteria. The fluxes through these networks have been computed using measurements of the production rates of the fermentation products (Blackwood *et al.*, 1956) and mass-balance equations (Papoutsakis & Meyer, 1985). The uptake and production rates involved with lysine fermentation of *Brevibacterium* have also been measured (Erickson *et al.*, 1978), and the fluxes within the network calculated (Vallino & Stephanopoulos, 1989). The set of mass balances and data result in an overdetermined system of equations, in which linear regression is used to estimate the unknown fluxes.

Several algorithms have been formulated for studying this type of network. The number of flux measurements that are needed to verify the stoichiometry of the biochemical network was derived (Tsai & Lee, 1988; Niranjan & San, 1989). The consistency of a set of flux measurements can also be checked by means of the mass-balance equations (Papoutsakis & Meyer, 1985). A complete procedure for analyzing this type of network is discussed in (Vallino & Stephanopoulos, 1989). This procedure provides a means for identifying errors in measurements and verifying the appropriateness of the steady-state approximation. The condition number of the system is used to determine whether the system is well-posed, and thus whether the fluxes within the network can be calculated based on measurements of uptake and production rates. A high condition number indicates that the network, as formulated, is ill-posed. The sensitivity of each calculated flux to each measurement can be estimated in order to identify the reactions that should be measured with the greatest accuracy.

In order to estimate the fluxes in the second type of network, i.e. those with a large number of branch points, some of the reaction fluxes within the network must be measured. The common technique is to expose cells to nutrients labeled with  $^{14}\text{C}$ , measure the concentrations of intermediates and products with the  $^{14}\text{C}$  label, and then calculate the steady-state fluxes using the mass-balance equations. The carbon fluxes in the metabolic network of the protozoa *Tetrahymena* have been measured in this manner for low glucose, high glucose, and low oxygen conditions (Stein & Blum, 1979, 1980, 1981). The results from this work included the quantification of futile cycling in the cell and the effect of glucose on the glyoxylate bypass. The fluxes through liver cells removed from rats on different diets have also been measured (Sauer *et al.*, 1970; Stucki & Walter, 1972; Crawford & Blum, 1983; Rabkin & Blum, 1985), as well as fluxes in cardiac tissue (Safer & Williamson, 1973). The effects of hormones on the fluxes in the liver were determined and the control characteristics of a series of fluxes were observed.

The minimum number of fluxes that must be measured using labeling techniques in order to uniquely calculate all the fluxes is equal to the number of independent

compound balances subtracted from the number of reactions. The mass-balance equations are used in conjunction with the data to calculate the values of the remaining fluxes. If the number of measurements is greater than the minimum number, linear regression is used to determine the unknown fluxes.

The procedures mentioned previously for analyzing metabolic networks are not appropriate for this second type of network, since they are based on an over-determined system of equations in which measurements of all secreted products and nutrients allow the direct calculation of fluxes. The choice in experimental design arises primarily from the amount of precision required for each measurement. In metabolic networks in which fluxes within the network must be measured, in addition to product secretion and nutrient uptake rates, one must determine which fluxes should be measured. There are no clear rules presently available for selecting the nutrients to be labeled and the intermediates to be measured. It has been suggested (Blum & Stein, 1982) that compounds with labeled carbon atoms that participate in a large number of reactions should be selected for measurement. However, even with these restrictions, there may be some sets of measurements which yield good calculated results and others which give poor results. In this report, we present a mathematical algorithm for identifying the reactions which would be the best choices for measurements, based on the assumption that the computed fluxes should have low sensitivity to errors in the experimental measurements. In part II of this series (Savinell & Palsson, 1992) we apply this algorithm to two important biological systems: a subset of metabolism in *Escherichia coli*, and intermediary metabolism in the hybridoma cell. The usefulness of this method of analysis as a guide for experimental design is discussed.

### Mathematical Methods

An example of the type of reaction network to be considered here is shown in Fig. 1. The fluxes that represent transport cross the cell membrane are termed *external fluxes* and the reactions which occur entirely within the cell are termed *internal fluxes*. The primary characteristic of the systems considered here is that the number of reactions is greater than the number of compounds. The dynamic mass-balance for each compound in the network can be represented by the system of equations:

$$\frac{d\mathbf{X}}{dt} = \mathbf{S} \cdot \mathbf{v} - \mathbf{b} \quad (1)$$

where  $\mathbf{X}$  is the vector of metabolite amounts per cell,  $\mathbf{S}$  is the  $n \times m$  stoichiometric matrix,  $n$  is the number of metabolites,  $\mathbf{v}$  is the vector of  $m$  metabolic fluxes, and  $\mathbf{b}$  is the vector of known substrate consumption rates, waste production rates, and biosynthetic fluxes. The element  $s_{ij}$  is the stoichiometric coefficient of the  $i$ th compound in the  $j$ th reaction. Since the rate of volume expansion due to growth is slow compared to metabolic transients, we can assume that the following condition

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \quad (2)$$

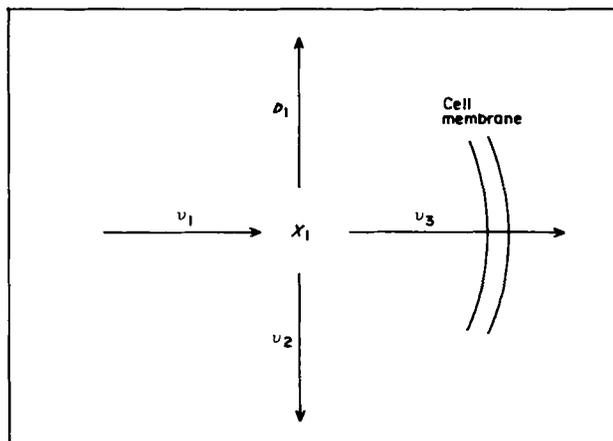


FIG. 1. Example of a metabolic reaction network. Reactions  $v_1$  and  $v_2$  are metabolic generation and degradation, respectively, and are termed *internal fluxes*. Reaction  $v_3$  involves transport across the cell membrane and is termed an *external flux*. The biosynthetic demand for  $X_1$  is given by  $b_1$ , which is either known from the composition of the cell and the growth rate, or can be measured directly. The stoichiometric matrix for this system has one mass balance and three fluxes, so that two of the three fluxes must be measured.

holds. Since the matrix  $S$  is known from the structure of the metabolic pathways and  $b$  is known from experimental data, this equation can be used to calculate the steady-state metabolic fluxes  $v$ .

#### PARTITIONING OF FLUXES INTO MEASURED AND COMPUTED FLUXES

In order to calculate the fluxes through the network, a minimum of  $(m-n)$  fluxes must be measured, and the remaining  $n$  fluxes calculated. The vector  $v$  and matrix  $S$  can be partitioned as follows:

$$v = \begin{bmatrix} v_c \\ - \\ v_e \end{bmatrix}, \quad S = [S_c | S_e]$$

where  $v_e$  are the  $(m-n)$  fluxes that are experimentally measured, and  $v_c$  are the  $n$  fluxes to be calculated. Equation (2) now becomes:

$$S_c v_c + S_e v_e = b \quad (3)$$

and  $v_c$  can be calculated by rearranging eqn (3):

$$v_c = S_c^{-1} \cdot (b - S_e v_e). \quad (4)$$

Two preconditions exist for partitioning  $v$  and  $S$ . First,  $S_c$  must be non-singular, and second, the fluxes in  $v_e$  must be measurable. Herein we put forth an additional criterion: the fluxes that are calculated,  $v_c$ , should have low sensitivity to experimental

error in the measured flux values,  $v_e$ . This criterion can be justified as follows. If a small error in a flux measurement is magnified greatly in the calculated fluxes, the calculated solution may be very inaccurate, and therefore meaningless. On the other hand, if the calculated solution depends on the data, and yet is relatively insensitive to error in the data, then one can have confidence in the accuracy of the calculations.

Let the matrix  $S$  be appended with a  $p \times m$  matrix  $I_p$ , where  $p = m - n$ , and  $b$  is appended with a  $p \times 1$  vector  $b_p$ , where  $b_p = v_e$ :

$$b_r = \begin{bmatrix} b \\ \hline b_p \end{bmatrix}, \quad S_r = \begin{bmatrix} S \\ \hline I_p \end{bmatrix}$$

such that:

$$S_r \cdot v = b_r \quad (5)$$

is true.

The matrix  $I_p$  is constructed such that each row in  $I_p$  will have one and only one element with value 1, and the column in which the 1 is located will be designated as the  $k$ th reaction. All other elements in the row will be zero, and all other elements in the  $k$ th column will be zero. The corresponding row of  $b_p$  will contain the experimental measurement of the  $k$ th flux. For example, for the stoichiometric matrix  $S$  corresponding to the network in Fig. 1,  $S_r$  may be formed such that:

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ \hline 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ \hline b_2 \\ b_3 \end{bmatrix} \quad (6)$$

where  $b_2$  is the measurement of flux  $v_1$ ,  $b_3$  is the measurement of flux  $v_3$ , and  $v_e = (v_1, v_3)^T$ .

#### MAXIMUM SENSITIVITY TO EXPERIMENTAL ERROR

The error in the data, represented by  $\Delta b_r$ , will lead to an error in the calculations, represented by  $\Delta v$ . The relationship between the two, obtained from eqn (5), is:

$$S_r \cdot \Delta v = \Delta b_r \quad (7)$$

The following equations are used to estimate the maximum error that can occur in the calculated fluxes.

From the property of the matrix norm (Stewart, 1973) and from eqn (5) one obtains:

$$\|S_r\| \cdot \|v\| \geq \|b_r\| \quad (8)$$

Similarly, after rearranging eqn (7), one obtains:

$$\|S_r^{-1}\| \cdot \|\Delta b_r\| \geq \|\Delta v\| \quad (9)$$

Multiplying the terms of eqn (8) by the terms of eqn (9) and rearranging yields:

$$\frac{\|\Delta v\|}{\|v\|} \leq C \cdot \frac{\|\Delta b_r\|}{\|b_r\|} \quad (10)$$

where

$$C = \|S_r\| \cdot \|S_r^{-1}\|. \quad (11)$$

The fractional error in the calculated solution is given by  $\|\Delta v\|/\|v\|$ , and the fractional error in the data is given by  $\|\Delta b_r\|/\|b_r\|$ . The quantity  $C$ , termed the *condition number* of matrix  $S_r$ , is a measure of the maximum sensitivity of the solution  $v$  to errors in the data  $b_r$ . Note that  $C$  is a function only of  $S_r$ . The role of the condition number in error analysis is derived and discussed thoroughly in (Wilkinson, 1963), and in many other texts on linear algebra, (e.g. Stewart, 1973).

The condition number  $C$  is thus a measure of the upper bound of the sensitivity of the calculated fluxes to error in experimental measurements in  $b_r$ . In other words,  $C$  is a measure of the possible error amplification. By changing the location of the 1s and 0s in  $I_p$ , and calculating the condition number for each new  $S_r$ , one can determine the possible error amplification for each set of fluxes to be measured. In the example given above, there are three different combinations of fluxes possible, given by three different  $I_p$ :

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ \hline 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where the first matrix is for  $v_1$  and  $v_2$  measured, the second matrix is for  $v_1$  and  $v_3$  measured, and the third matrix is for  $v_2$  and  $v_3$  measured. The best combination of fluxes to be measured is that which gives  $S_r$  with the lowest condition number, and thus the lowest sensitivity to error in data. The worst choice of fluxes is given by the matrix with the highest condition number. Infinite-valued condition numbers represent flux combinations for which a computed solution is not possible.

The value of the condition number depends on the choice of norm that is used. The norm is chosen according to the type of error description that is desired. Three of the most commonly used norms are the 1-norm, the 2-norm, and the  $\infty$ -norm (Stewart, 1973). When the  $\infty$ -norm is used,  $\|\Delta v\|/\|v\|$  represents the largest error in a single flux, normalized to the flux with the largest value (Stewart, 1973). This means of error description is not useful here and a better description is given by the 2-norm. When the 2-norm is used,  $\|\Delta v\|/\|v\|$  represents the sum of all the error normalized to the sum of all the fluxes. The error description provided by the 1-norm is similar to the 2-norm. The 1- and 2-norms thus are measures of the average error overall in the calculations. In this report, the condition number will be calculated using the 2-norm, due to the significance of the error interpretation and the widespread use of this norm. Results using other norms may differ from those shown here, due to the difference in error interpretation.

The 2-norms of  $S_r$  and  $S_r^{-1}$  are given by (Strang, 1980):

$$\|S_r\| = \bar{\sigma}$$

$$\|S_r^{-1}\| = \frac{1}{\sigma^0}$$

where  $\bar{\sigma}$  and  $\sigma^0$  are the maximum and minimum singular values of  $S_r$ , respectively. The condition number  $C$  can then be calculated by:

$$C = \frac{\bar{\sigma}}{\sigma^0} \quad (12)$$

#### ACTUAL SENSITIVITY TO EXPERIMENTAL ERROR

The value of  $C$  obtained from eqn (12) is only an upper bound on the sensitivity. Because this is an upper bound, a set of fluxes with a high condition number has the *potential* to result in a more sensitive solution than the low-condition number set, although the high-condition number set will not necessarily be more sensitive than the low-condition number set. The actual sensitivity,  $R$ , of the calculated fluxes to the experimental data is defined by:

$$\frac{\|\Delta v\|}{\|v\|} = R \cdot \frac{\|\Delta b_r\|}{\|b_r\|} \quad (13)$$

where  $R$  must always be less than or equal to  $C$ . The actual sensitivity,  $R$ , is a function not only of  $S_r$ , but also of the values of the experimental measurements and the error, given by  $b_r$  and  $\Delta b_r$ , respectively.

The condition for  $R = C$  occurs when the inequalities in eqns (8) and (9) become equalities. The vectors  $b_{r,max}$ ,  $\Delta b_{r,max}$ ,  $v_{max}$ , and  $\Delta v_{max}$ , which satisfy these equalities are defined by:

$$\|S_r\| \cdot \|v_{max}\| = \|b_{r,max}\| \quad (14)$$

$$\|S_r^{-1}\| \cdot \|\Delta b_{r,max}\| = \|\Delta v_{max}\|. \quad (15)$$

The values of  $b_r$  and  $\Delta b_r$  are obtained by the singular value decomposition of  $S_r$  and  $S_r^{-1}$ . By the singular value decomposition theorem,  $S_r$  can be decomposed as (Strang, 1980):

$$S_r = U\Sigma W^T$$

where  $\Sigma$  is a diagonal matrix, with each element along the diagonal called a singular value, and  $U$  and  $W$  are orthonormal matrices.

It is shown in the Appendix that a sufficient condition for eqn (14) to be true is when:

$$b_{r,max} = \alpha S_r \bar{w} \quad (16)$$

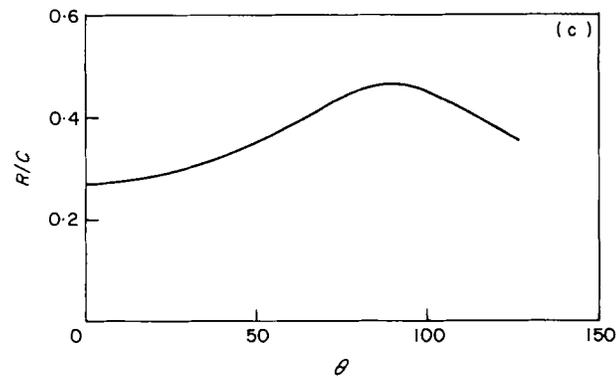
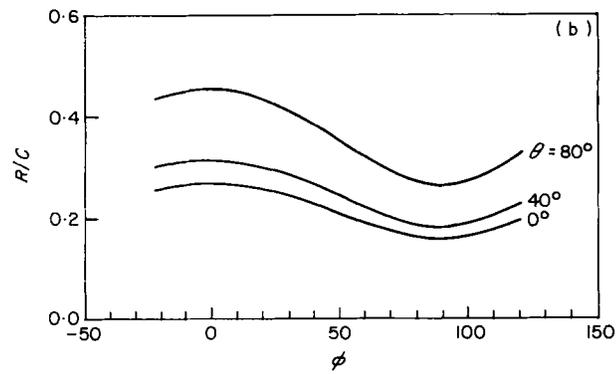
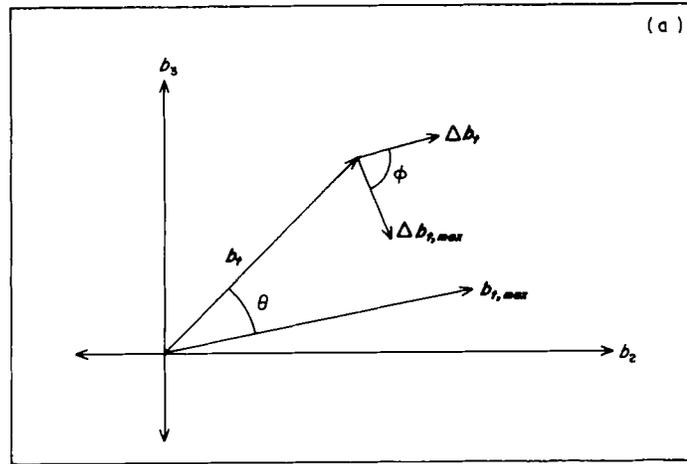


FIG. 2. (a) Illustration of the angles  $\theta$  and  $\phi$  for the example system in Fig. 1. The biosynthetic demands and flux measurements are represented by  $b_i$  and  $\Delta b_i$ , represents the error in the measurements. The vectors  $b_{i,max}$  and  $\Delta b_{i,max}$ , given by eqns (16) and (17), result in  $R=C$ . The value of the biosynthetic demand,  $b_i$ , is set to  $\gamma$  and the value of the error in biosynthetic demand,  $\Delta b_i$ , is set to  $\delta$ . The projection of  $b_i$  and  $\Delta b_i$  onto the plane  $[b_i(1)=\gamma]$  is shown and  $\theta$  is the angle between these projections. Similarly, the projection of  $\Delta b_i$  and  $\Delta b_{i,max}$  onto the plane  $[\Delta b_i(1)=\delta]$  is shown and  $\phi$  is the angle between these projections. (b) The value of sensitivity/condition number as a function of the angle  $\phi$  between the projections of  $\Delta b_i$  and  $\Delta b_{i,max}$  onto the plane  $[b_i(1)=0]$  for the system given by eqn (18). The angle between the projections of  $b_i$  and  $b_{i,max}$  onto the plane  $[b_i(1)=0]$  is given by  $\theta$ . (c) The value of sensitivity/condition number as a function of  $\theta$  when  $\phi=0$ .

The parameter  $\alpha$  is any scalar value, and  $\bar{w}$  is the column of  $\mathbf{W}$  that corresponds to the maximum singular value of  $\mathbf{S}_r$ . Similarly, eqn (15) is true when:

$$\Delta \mathbf{b}_{t,max} = \beta \mathbf{u}^0 \quad (17)$$

where  $\mathbf{u}^0$  is the column of  $\mathbf{U}$  that corresponds to the minimum value in  $\Sigma$  and  $\beta$  is any scalar value. Thus, the directions of  $\mathbf{b}_t$  and  $\Delta \mathbf{b}_t$  for which  $R$  is exactly equal to  $C$ , are given by eqns (16) and (17).

The influence of  $\mathbf{b}_t$  and  $\Delta \mathbf{b}_t$  on  $R$  is illustrated by means of the example system illustrated in Fig. 1, where  $\mathbf{S}_r$  is given by:

$$\mathbf{S}_r = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (18)$$

and  $\mathbf{b}_t$  is a  $3 \times 1$  vector. The first element of  $\mathbf{b}_t$  is the biosynthetic demand, with the value  $b_1$ , and the error in this demand,  $\Delta \mathbf{b}_t$ , has the value  $\Delta b_1$ . All possible directions of  $\mathbf{b}_t$  and  $\Delta \mathbf{b}_t$ , with the restriction that  $b_t(1) = b_1$  and  $\Delta b_t(1) = \Delta b_1$ , can be considered by independently varying the value of the second element of each vector and calculating the third element, keeping the magnitudes constant. The actual sensitivity,  $R$ , is calculated for each  $\mathbf{b}_t$  and  $\Delta \mathbf{b}_t$  using eqn (13), and  $R$  is plotted as a function of two angles  $\phi$  and  $\theta$ . As illustrated in Fig. 2(a),  $\phi$  is the angle between the projections of  $\Delta \mathbf{b}_t$  and  $\Delta \mathbf{b}_{t,max}$  in the plane of  $\Delta b_t(1) = \Delta b_1$ , and  $\theta$  is the angle between the projections of  $\mathbf{b}_t$  and  $\mathbf{b}_{t,max}$  in the plane of  $b_t(1) = b_1$ .

Figure 2(b) and (c) illustrates that both  $\phi$  and  $\theta$  influence the sensitivity. Although the maximum theoretical value of  $R/C$  is 1, this value will rarely occur in a biological system. The reason that  $R/C$  is small is that the values of  $\Delta \mathbf{b}_{t,max}$  and  $\mathbf{b}_{t,max}$  are given by the singular value decomposition of  $\mathbf{S}_r$ , and thus most elements of these vectors are non-zero. However, many of the biosynthetic fluxes in  $\mathbf{b}_t$  have zero values, since not all metabolites are involved in biosynthesis. Hence, the physiological  $\Delta \mathbf{b}_t$  and  $\mathbf{b}_t$  will always be displaced from  $\Delta \mathbf{b}_{t,max}$  and  $\mathbf{b}_{t,max}$ , causing  $R$  to be less than  $C$ . Figure 2(b) and (c) demonstrates, for this example, that the condition number will always overestimate the actual sensitivity by at least a factor of 2, regardless of the values of the flux measurements and the measurement error. When systems with larger dimension are considered, the maximum value of the ratio  $R/C$  tends to decrease. However, as will be shown later, there is a correlation between the sensitivity  $R$  obtained using an estimate of  $\mathbf{b}_t$  and the condition number. Consequently, the condition number can be used to estimate relative sensitivities between flux arrangements, even though the actual sensitivity may be much less than that given by the condition number.

#### EFFECT OF ERROR DIRECTION

The values of  $\mathbf{b}_t$  are characteristic of the system and the environment, while the values of  $\Delta \mathbf{b}_t$  are determined by the precision of the instruments and techniques used in measurement. Consequently, the values of  $\Delta \mathbf{b}_t$  are essentially unknown. Since the

elements of  $\Delta b_i$ , represent errors in measurement that are mostly independent, the vector  $\Delta b_i$  can theoretically have any direction in  $(m-n)$  dimensional space. The magnitude of  $\Delta b_i$  can be considered to be constant, since the magnitudes of  $\Delta b_i$  and  $\Delta v$  cancel in eqn (13). Thus, all possible directions of  $\Delta b_i$  can be thought to be represented by all points on the surface of a sphere in  $(m-n)$ -dimensional space.

Figure 3(a) illustrates the surface traced by  $\Delta b_i$ , for the example system given by eqn (6). In this example the surface is a circle because only two measurements ( $m-n=2$ ) are needed. An even distribution of points on this surface is needed in order to determine the effect of the direction of the error vector, represented by each point, on the sensitivity. Each element of  $\Delta b_i$  is assigned a random value, and the vector magnitude is calculated. All vectors with magnitude greater than 1 are discarded, and those with magnitude less than or equal to 1 are normalized to 1, thus ensuring an even distribution of points. The sensitivity of the computed solution to the measured solution, for a particular experimental design given by  $S_i$ , can be calculated for each of these remaining error vectors. Since the objective is to determine the most probable sensitivity when the measurement error is unknown, a frequency distribution is obtained from the calculated sensitivities, as shown in Fig.

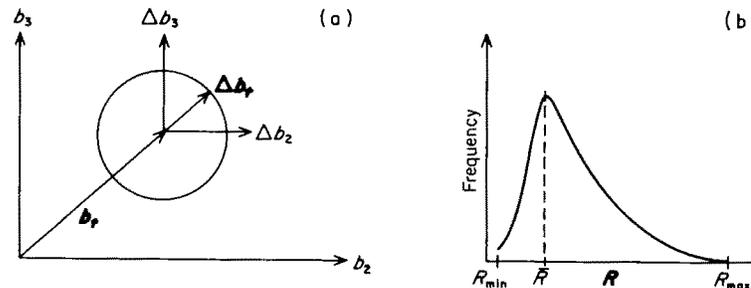


FIG. 3. (a) Illustration of the surface traced by all possible directions of  $\Delta b_i$ , for the example system given by eqn (6). Since only two fluxes must be measured for this example, the surface in this example is a circle in two-dimensional space. (b) A hypothetical frequency distribution of the actual sensitivities,  $R$ , calculated for each  $\Delta b_i$ , where  $\bar{R}$  represents the most probable sensitivity.

3(b). The maximum value of this curve represents the most probable value of  $R$ . This procedure will be illustrated in detail using two biological models in part II of this series (Savinell & Palsson, 1992).

#### EFFECT OF ADDITIONAL MEASUREMENTS

The analysis presented above has assumed that only the minimum number of fluxes would be measured. In many instances, a greater number of fluxes are measured, in order to check for the consistency of the data and the accuracy of the metabolic network. If the number of fluxes measured is greater than  $m-n$ , the appended stoichiometric matrix is no longer square and is designated by  $S_i^\eta$ , where  $\eta$  represents the number of additional measurements. The resulting set of mass balances is thus

an overdetermined set of possibly inconsistent equations, where the best estimate of the fluxes,  $\bar{v}$ , are calculated using linear regression:

$$\mathbf{S}_i^{\eta T} \mathbf{S}_i^{\eta} \bar{v} = \mathbf{S}_i^{\eta T} \mathbf{b}_i. \quad (19)$$

This method of solution is discussed in most linear algebra texts (e.g. Strang, 1980). From eqn (19) one can derive:

$$\frac{\|\Delta \bar{v}\|}{\|\bar{v}\|} \leq C_{\eta} \cdot \frac{\|\mathbf{S}_i^{\eta T} \Delta \mathbf{b}_i\|}{\|\mathbf{S}_i^{\eta T} \mathbf{b}_i\|} \quad (20)$$

where

$$C_{\eta} = \|\mathbf{S}_i^{\eta T} \mathbf{S}_i^{\eta}\| \cdot \|(\mathbf{S}_i^{\eta T} \mathbf{S}_i^{\eta})^{-1}\|. \quad (21)$$

In this derivation of sensitivity, the error vector and the measurement vector are multiplied by  $\mathbf{S}_i^{\eta T}$ , so that the interpretation of  $C_{\eta}$  is somewhat different from the interpretation of  $C$ . To compare the maximum possible sensitivity using excess measurements with the minimum number of measurements, we note that when  $\eta=0$ , i.e.  $\mathbf{S}_i^{\eta} = \mathbf{S}_i$ , then

$$C_0 = \|\mathbf{S}_i^T \mathbf{S}_i\| \cdot \|(\mathbf{S}_i^T \mathbf{S}_i)^{-1}\| = C^2$$

is true. The value of  $C_{\eta}$  is calculated for  $\eta=0, 1, 2, \dots, m-(m-n)$  for the example systems in part II (Savinell & Palsson, 1992).

The actual sensitivity of the computed solution when excess flux measurements are made,  $R_{\eta}$ , is defined analogous to eqn (20):

$$\frac{\|\Delta \bar{v}\|}{\|\bar{v}\|} = R_{\eta} \cdot \frac{\|\mathbf{S}_i^{\eta T} \Delta \mathbf{b}_i\|}{\|\mathbf{S}_i^{\eta T} \mathbf{b}_i\|}. \quad (22)$$

In order to calculate the upper bound on  $R_{\eta}$ , the following expression is used, analogous to eqn (17):

$$\mathbf{S}_i^{\eta T} \Delta \mathbf{b}_{i\max} = \beta \mathbf{u}^0 \quad (23)$$

where  $\mathbf{u}^0$  is the column of  $\mathbf{U}$  corresponding to the smallest singular value of the matrix  $\mathbf{S}_i^{\eta T} \mathbf{S}_i^{\eta}$ .

### Discussion

The estimation of steady-state fluxes in a biochemical network is usually obtained by measuring some of the fluxes, and using the mass-balance equations to compute the other fluxes. An algorithm is derived for estimating the sensitivity of the computed fluxes to experimental error in the measured fluxes. The upper bound on this sensitivity is given by the condition number, which is a function only of stoichiometry. If estimates of the fluxes are available before the experiment is performed, then a more realistic sensitivity can be calculated. The experiment is then designed such that the combination of fluxes which yield low sensitivity is selected for measurement. This

algorithm will be illustrated with two model systems, *E. coli* and a hybridoma cell line, in part II of this series.

This research was supported by National Science Foundation grant BCS-9009389. The authors thank James S. Freudenberg for his help with the derivation in the Appendix.

#### REFERENCES

- BLACKWOOD, A., NEISH, A. & LEDINGHAM, G. (1956). Dissimilation of glucose at controlled pH values by pigmented and non-pigmented strains of *Escherichia coli*. *J. Bact.* **72**, 497.
- BLUM, J. J. & STEIN, R. B. (1982). On the analysis of metabolic networks. In *Biological Regulation and Development* Vol. 3A (Goldberger, R. & Yamamoto, K., eds) pp. 99-125. New York: Plenum Press.
- CRAWFORD, J. M. & BLUM, J. J. (1983). Quantitative analysis of flux along the gluconeogenic, glycolytic and pentose phosphate pathways under reducing conditions in hepatocytes isolated from fed rats. *Biochem. J.* **212**, 595-598.
- ERICKSON, L., SELGA, S. & VIESTURS, U. (1978). Application of mass and energy balances regularities to product formation. *Biotechnol. Bioengng.* **20**, 1623.
- NIRANJAN, S. & SAN, K.-Y. (1989). Analysis of a framework using material balances in metabolic pathways to elucidate cellular metabolism. *Biotechnol. Bioengng.* **34**, 496-501.
- PAPOUTSAKIS, E. T. & MEYER, C. L. (1985). Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotechnol. Bioengng.* **27**, 50-66.
- RABKIN, M. & BLUM, J. (1985). Quantitative analysis of intermediary metabolism in hepatocytes incubated in the presence and absence of glucagon with a substrate mixture containing glucose, ribose, fructose, alanine and acetate. *Biochem. J.* **225**, 761-786.
- SAFER, B. & WILLIAMSON, J. (1973). Mitochondrial and cytosolic interactions in perfused rat heart. Role of coupled transamination in repletion of citric acid cycle intermediates. *J. biol. Chem.* **248**(7), 2570-2579.
- SAUER, F., ERFLF, J. & BINNS, M. (1970). Turnover rates and intracellular pool size distribution of citrate cycle intermediates in normal, diabetic and fat-fed rats estimated by computer analysis from specific activity decay data of  $^{14}\text{C}$ -labeled citrate cycle acids. *Eur. J. Biochem.*, **17**, 350-363.
- SAVINELL, J. M. & PALSSON, B. O. (1992). Optimal selection of metabolic fluxes for *in vivo* measurement. II. Application to *Escherichia Coli* and hybridoma cell metabolism. *J. theor. Biol.* **155**, 215-241.
- STEIN, R. B. & BLUM, J. (1979). Quantitative analysis of intermediary metabolism in *tetrahymena pyriformis*; cells grown in proteose-peptone and resuspended in a defined nutrient-rich medium. *J. biol. Chem.* **254**(20), 10385-10395.
- STEIN, R. B. & BLUM, J. (1980). Quantitative analysis of intermediary metabolism in *tetrahymena pyriformis*; cells grown in glucose-supplemented medium. *J. biol. Chem.* **255**(9), 4198-4205.
- STEIN, R. B. & BLUM, J. (1981). Quantitative analysis of intermediary metabolism in *tetrahymena pyriformis*; cells kept under static conditions for four hours after growth in glucose-supplemented medium. *J. biol. Chem.* **256**(6), 2752-2760.
- STEWART, G. (1973). *Introduction to Matrix Computations*. London: Academic Press.
- STRANG, G. (1980). *Linear Algebra and its Applications*. London: Academic Press.
- STUCKI, J. W. & WALTER, P. (1972). Pyruvate metabolism in mitochondria from rat liver: measured and computer-simulated fluxes. *Eur. J. Biochem.* **30**, 60-72.
- TSAI, S. & LEE, Y. (1988). Application of Gibbs' rule and a simple pathway method to microbial stoichiometry. *Biotechnol. Prog.* **4**(2), 82-88.
- VALLINO, J. J. & STEPHANOPOULOS, G. (1989). Flux determination in cellular bioreaction networks: applications to lysine fermentations. In: *Frontiers in Bioprocessing* (Sikdar, S. K., Bier, M. & Todd, P., eds) pp. 205-219. Boca Raton, FL: CRC Press.
- WILKINSON, J. (1963). *Rounding Errors in Algebraic Processes*. Englewood Cliffs, NJ: Prentice-Hall.

#### APPENDIX

##### Derivation of Conditions for $R = C$

The objective of this proof is to show that the expression for  $\Delta b_{r,max}$  given by eqn (16) is a sufficient condition for eqn (14) to be true. Before presenting the proof we

note that the rearrangement of eqns (5) and (14) yields:

$$\|\mathbf{S}_t \mathbf{v}_{max}\| = \|\mathbf{S}_t\| \cdot \|\mathbf{v}_{max}\| = \|\mathbf{b}_{t,max}\|. \quad (\text{A.1})$$

The singular value decomposition (SVD) of  $\mathbf{S}_t$  is given by:

$$\mathbf{S}_t = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^T = \mathbf{\Sigma}\sigma_i \mathbf{u}_i \mathbf{w}_i^T. \quad (\text{A.2})$$

Since  $\mathbf{U}$  and  $\mathbf{W}^T$  are orthonormal matrices (by the SVD theorem), the following can be obtained from eqn (A.2):

$$\mathbf{S}_t \mathbf{W} = \mathbf{U}\mathbf{\Sigma} \quad (\text{A.3})$$

$$\mathbf{S}_t \mathbf{w}_i = \sigma_i \mathbf{u}_i. \quad (\text{A.4})$$

Let  $\bar{\sigma}$  be the largest singular value of  $\mathbf{S}_t$ , and  $\bar{\mathbf{u}}$  and  $\bar{\mathbf{w}}$  be the corresponding columns of  $\mathbf{U}$  and  $\mathbf{W}$ , respectively. From the definition of the 2-norm, and using the SVD theorem, one can derive the following (Strang, 1980):

$$\|\mathbf{S}_t\| = \bar{\sigma}. \quad (\text{A.5})$$

Further, since  $\mathbf{U}$  and  $\mathbf{W}^T$  are orthonormal matrices, the following is true:

$$\|\mathbf{u}_i\| = \|\mathbf{w}_i\| = 1. \quad (\text{A.6})$$

We begin the proof by assuming that  $\mathbf{v}_{max}$  is given by:

$$\mathbf{v}_{max} = \alpha \bar{\mathbf{w}} \quad (\text{A.7})$$

and thus

$$\mathbf{b}_{t,max} = \mathbf{S}_t \mathbf{v}_{max} = \alpha \mathbf{S}_t \bar{\mathbf{w}}. \quad (\text{A.8})$$

Combining eqns (A.4) and (A.8), we obtain:

$$\mathbf{S}_t \mathbf{v}_{max} = \alpha \bar{\sigma} \bar{\mathbf{u}}. \quad (\text{A.9})$$

Using eqn (29), we obtain from eqn (A.9):

$$\|\mathbf{S}_t \mathbf{v}_{max}\| = \alpha \bar{\sigma}.$$

Inserting eqn (28) into the previous equation gives:

$$\|\mathbf{S}_t \mathbf{v}_{max}\| = \alpha \|\mathbf{S}_t\|. \quad (\text{A.10})$$

From eqns (A.6) and (A.7):

$$\|\mathbf{v}_{max}\| = \alpha \quad (\text{A.11})$$

and inserting this expression for  $\alpha$  into eqn (A.10) gives

$$\|\mathbf{S}_t \mathbf{v}_{max}\| = \|\mathbf{S}_t\| \cdot \|\mathbf{v}_{max}\| = \|\mathbf{b}_{t,max}\| \quad (\text{A.12})$$

which is eqn (A.1). Therefore,  $\mathbf{b}_{t,max} = \alpha \mathbf{S}_t \bar{\mathbf{w}}$  is a sufficient condition for eqn (14) to be true.

The derivation of the expression for  $\Delta \mathbf{b}_{t,max}$  given by eqn (17), which must satisfy eqn (15), proceeds along similar lines. The singular value decomposition of  $\mathbf{S}_t^{-1}$  is given by:

$$\mathbf{S}_t^{-1} = \mathbf{W}\mathbf{\Sigma}^{-1}\mathbf{U}^T. \quad (\text{A.13})$$

It can be shown, by analogy with the above derivation, that

$$\Delta \mathbf{b}_{t,max} = \beta \mathbf{u}^0 \quad (\text{A.14})$$

where  $\mathbf{u}^0$  is the column of  $\mathbf{U}$  corresponding to the maximum value of  $\mathbf{\Sigma}^{-1}$ . The maximum value of  $\mathbf{\Sigma}^{-1}$  corresponds to the minimum singular value of  $\mathbf{\Sigma}$ . Thus,  $\mathbf{u}^0$  corresponds to the minimum singular value of  $\mathbf{S}_t$ .

It can also be proven that  $\Delta \mathbf{b}_{t,max}$  is orthogonal to  $\mathbf{b}_{t,max}$ , as illustrated in Fig. 2.