Phytochemistry, 1976, Vol. 15, pp. 1367–1377. Pergamon Press. Printed in England.

DETERMINING THE STRUCTURE OF A BIOSYNTHETIC PATHWAY FROM INCORPORATION MEASUREMENTS

IAIN M. CAMPBELL
Department of Biochemistry, Faculty of Arts and Sciences, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, U.S.A.

(*Revised received* 3 *March* 1976)

Key Word Index—Detailed analysis; two component biosynthetic system; flow matrix; estimates of enzyme activity and metabolic load.

Abstract—Incorporation data, collected judiciously from a biosynthetic tracer experiment, can be processed to yield a considerable body of detailed information about a given biosynthetic pathway. In addition to determining what substances are *bona fide* intermediates in the pathway and establishing the order in which they occur, it is shown that estimates can be obtained of: (a) the activities of enzymes in the pathway, (b) the extent to which components in the pathway are common to other pathways of metabolism in the particular cell colony, and (c) the metabolic load being carried by the pathway under study. The various mathematical operations that need to be applied to experimentally derived incorporation data to produce this information, are described; examples are worked and step-by-step summaries of procedure are given.

INTRODUCTION

In a recent paper [1] it was shown that incorporation ($I\%$) and dilution (D) values could be predicted for a simple biosynthetic tracer experiment provided the structure of the biosynthetic pathway and the sampling time(s) were known. 'Knowing the structure' of a biosynthetic pathway implied a knowledge not only of the identity of the precursor(s) and of the various intermediate(s) in the pathway and their relative position, but also of the intracellular pool size of each component, the intrapool fluxes of mass, and the manner and extent to which components of the pathway were common to contiguous areas of intermediary metabolism. The analysis dealt with a simple three component system of a precursor, an intermediate and a product, and allowed some well-accepted, but emperically established, biosynthetic facts to be placed in theoretical perspective.

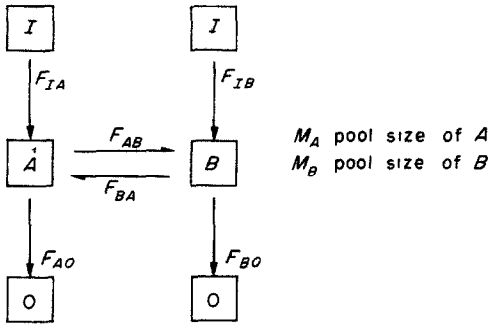
There emerged from the study several suggested changes in experimental design which, if adopted, should increase the yield of biological information emerging for isotope incorporation experiments. A particularly tantalizing possibility was that of using experimentally derived incorporation values to determine the structure of biosynthetic pathways. If this proved feasible, biogenetic hypotheses could be checked quickly and in detail, and there would be immediate access to the *in vivo* rate constant of each enzyme in the pathway. Moreover, each pathway so studied would be placed in proper quantitative context as far as the rest of the metabolism of the particular cell colony was concerned. The present report describes how this possibility can be realized in principle.

The paper has three main sections. In the first, the general relationship between a biosynthetic pathway of known structure and the flux of administered isotope through it, is developed. This differs somewhat from the former paper in that biosynthetic pathways are represented in a matrix format which not only defines them uniquely and succinctly but also permits facile calculation of the various $I\%$ versus time functions. These functions govern the discrete $I\%$ value in any component of the pathway at any time following isotope addition to the pathway. In the second section it is shown how the various constants in these $I\%$ vs time functions, together with intracellular pool size data, can be used to regenerate the pathway specific matrices and hence lay bare the structure of the biosynthetic pathway. The last section deals with the problem of constructing $I\%$ vs time functions from discrete, experimentally determined $I\%$ values. The simplest of all biosynthetic pathways—the two component system—is used in illustrations throughout to ensure that manipulative detail does not occlude basic principle. Worked examples are provided at each stage, as are step-by-step summaries of procedure. It is hoped that these devices will help break down the bioorganic chemist's understandable diffidence in considering an approach to biosynthetic investigation somewhat far removed from the classical one.

It must be emphasized at the outset that the methods used in this report are not new. Compartmental analysis has been used in processing isotope incorporation data since the mid-1930's and several excellent monographs are available in which the substantial literature is summarized [2–4]. However, the bulk of this work has been done in the physiological context and with emphasis on whole animals. The objective of this study was to bring compartmental analysis functionally into the context of plant and fungal secondary metabolism; we perceive such methods to be useful in processing data obtained from combined radio gas chromatography—mass spectrometric analysis [5] of liquid cultures of algae, fungi and cells derived from higher plants.

METHODS AND RESULTS

Attention will be focused on the two component system shown in Scheme 1



Scheme 1.

B is the biosynthetic product and *A* the precursor. The designations *I* and *O* represent those other areas of the organism's intermediary metabolism that supply mass to (*I*), and receive mass from (*O*) the two components of the system. It is assumed (a) that the intracellular pools of each component, M_A and M_B, do not alter significantly in size during the tracer experiment, (b) that the administered isotopically enriched substrate, whether *A* or *B*, is either singly labelled or its multiply labelled centres are preserved as a unit in all the reactions covered by Scheme 1*, and (c) that the labelled substrate is introduced to the system over a time interval that is short compared with the duration of the experiment. The various flows associated with each arrow in Scheme 1, F_{AB}, F_{BA}, etc. represent the molar (or mM, μM, etc.) flux per unit time *from* pool *A* *to* pool *B*. *from* pool *B* *to* pool *A*, etc. respectively. These flows are measures of the activity of the enzyme that catalyses that particular step in the pathway. When divided by the molar mass of the pool denoted by the first of its subscripts, flows yield '*k* values', e.g. $F_{AB}/M_A = k_{AB}$. These latter values can be considered fluxes measured, not in mol/unit time, but in units of pool size equivalents per unit time. The amount of isotope in any pool of the system is given by Q_A^0 and Q_B^0 initially, by Q_A and Q_B at any time subsequently†.

Defining a pathway of known structure in matrix format

As was shown previously [1], two easily constructed‡ equations govern movement of any administered isotope in the system of Scheme 1, viz.

$$\frac{dQ_A}{dt} = k_{BA}Q_B - (k_{AB} + k_{AO})Q_A$$
$$= k_{BA}Q_B - k_{AA}Q_A \quad (1)$$

$$\frac{dQ_B}{dt} = k_{AB}Q_A - (k_{BA} + k_{BO})Q_B$$
$$= k_{AB}Q_A - k_{BB}Q_B \quad (2)$$

where $k_{AB} + k_{AO} = k_{AA}$ and $k_{BA} + k_{BO} = k_{BB}$. These two linear simultaneous differential equations can be written in matrix format (3) or (4), where in (4), $\mathbf{Q}'$ is a two element vector (column) of dQ/dt values, $\mathbf{Q}$ is a two element vector of Q values, and $\mathbf{k}$ is a 2 × 2 matrix of k values.

$$\begin{vmatrix} \frac{dQ_A}{dt} \\ \frac{dQ_B}{dt} \end{vmatrix} = \begin{vmatrix} -k_{AA} + k_{BA} \\ k_{AB} - k_{BB} \end{vmatrix} \cdot \begin{vmatrix} Q_A \\ Q_B \end{vmatrix} \quad (3)$$

$$\mathbf{Q}' = \mathbf{k} \cdot \mathbf{Q} \quad (4)$$

Note particularly how the **k** matrix is set up. Consider the left hand column to be headed by $\mathscr{A}$, the right hand column by $\mathscr{B}$. Each column therefore contains respectively those *k* values in which *A* and *B* are the first of the two subscript letters (see 5 below), i.e. those *k* values which relate to flows *from* A and B respectively. In descending order consider the rows indexed $\mathscr{A}$, $\mathscr{B}$. Rows

	$\mathscr{A}$	$\mathscr{B}$
$\mathscr{A}$	$-k_{AA}$	k_{BA}
$\mathscr{B}$	k_{AB}	$-k_{BB}$

(5)

	$\mathscr{A}$	$\mathscr{B}$
$\mathscr{A}$	$-k_{AA}$	k_{BA}
$\mathscr{B}$	k_{AB}	$-k_{BB}$
$\mathscr{O}$	k_{AO}	k_{BO}

(6)

will contain *k* values associated with flows *to* the component specified by the second of the two subscripts. If an additional row is tacked onto the foot of the **k** matrix, is labelled $\mathscr{O}$ (*to* the outside), and is filled with the values k_{AO} and k_{BO}, the particularly valuable "augmented *k*" matrix ($\bar{\mathbf{k}}$) (6) is formed. Note firstly, that each column in the augmented **k** sums to zero, e.g. $-k_{AA} + k_{AB} + k_{AO} = -(k_{AB} + k_{AO}) + k_{AB} + k_{AO} = 0$; and secondly that if each column is multiplied by the mass of the pool designated by its heading, **k** is converted directly to the flow matrix (**F**) (7).

	$\mathscr{A}$	$\mathscr{B}$
$\mathscr{A}$	$-F_{AA}$	F_{BA}
$\mathscr{B}$	F_{AB}	$-F_{BB}$
$\mathscr{O}$	F_{AO}	F_{BO}

(7)

	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{I}$
$\mathscr{A}$	$-F_{AA}$	F_{BA}	F_{IA}
$\mathscr{B}$	F_{AB}	$-F_{BB}$	F_{IB}
$\mathscr{O}$	F_{AO}	F_{BO}	$-F_{IO}$

(8)

---

* Situations such as the biosynthesis of a polyketide where one precursor (e.g. acetate) is incorporated into the product several times, are not covered by this present analysis.

† Q_A^0 and Q_B^0 are the Q_A ($t = 0$) and Q_B ($t = 0$) of ref. [1] likewise Q_A replaces Q_A (t), etc.

‡ For any system of known structure with *n* components there are *n* differential equations, one defining the rate of change of isotope content (dQ/dt) in each pool of the system. For the component *M*, flanked in the system by components *L* and *N*, the appropriate equation is:

$$\frac{dQ_M}{dt} = k_{LM}Q_L + k_{NM}Q_N - Q_M(k_{ML} + k_{MN} + k_{MO}),$$

the sum $(k_{ML} + k_{MN} + k_{MO})$ often being represented as k_{MM}. In words, the procedure for obtaining the correct differential equation is: the rate of change of isotope in any pool is the sum of the "$k.Q$" products for all pools feeding the one in question minus the product of the Q for the pool in question and the sum of the k values for all the pools that the pool in question feeds.

Augmentation of $\mathbf{F}$ by inclusion of another column on its right-hand side, with heading $\mathscr{I}$ and filled with the terms F_{IA}, F_{IB} and F_{IO}, yields the all-important matrix $\bar{\mathbf{F}}$ (8) in which all the flows occurring in Scheme 1 appear together with a few that do not. As with $\mathbf{k}$, the columns of $\bar{\mathbf{F}}$ sum to zero—the element $-F_{IO}$ is defined as the negative sum $(F_{IA} + F_{IB})$ to keep it in harmony with the other columns. As will be discussed later, this element is a biologically significant one since it is a measure of the flux of material into (and therefore also out of) the system considered as a whole. Unlike $\mathbf{k}$, however, the rows of $\bar{\mathbf{F}}$ also sum to zero. This latter situation occurs because the pool size of each component in the system must be held constant, i.e. inflow must equal outflow; for component A, for instance $-F_{AA}$ must equal $(F_{BA} + F_{IA})$. These important properties of $\bar{\mathbf{F}}$ will be used later.

From the earlier understanding of what is implied by the phrase "knowing the structure of a biosynthetic pathway", it follows that $\bar{\mathbf{F}}$ can be written down directly for any known pathway; $\bar{\mathbf{F}}$ becomes therefore a symbolic representation of any known pathway. Together with pool size information it defines the pathway uniquely and completely.

Obtaining incorporation vs time functions and discrete I% values from structure defining matrices†*

For the purpose of argument, let label in the amount of Q_A^0 be added to pool A in Scheme 1 at time zero. The two incorporation vs time functions required are‡

$$I_{AA}\% = \frac{Q_A}{Q_A^0} \times 100 \quad \text{and} \quad I_{AB}\% = \frac{Q_B}{Q_A^0} \times 100 \tag{9, 10}$$

Expressions for Q_A and Q_B come from solutions of equations (3) or (4). In the former paper [1], this was achieved through standard transformation methods—a process which though adequate for the two and three component system is difficult to extend to larger systems, should such extension be needed. A simple, general solution to equation (4) exists which is quite compatible with the representation of biosynthetic pathways in matrix form.

If matrix $\mathbf{k}$ is available, and of course it is for any known pathway, the solutions for Q_A and Q_B are:

$$Q_A = Z_1 X_{A1} e^{g_1 t} + Z_2 X_{A2} e^{g_2 t} \tag{11}$$

$$Q_B = Z_1 X_{B1} e^{g_1 t} + Z_2 X_{B2} e^{g_2 t} \tag{12}$$

or more conveniently:

$$\begin{vmatrix} Q_A \\ Q_B \end{vmatrix} = Z_1 \begin{vmatrix} X_{A1} \\ X_{B1} \end{vmatrix} e^{g_1 t} + Z_2 \begin{vmatrix} X_{A2} \\ X_{B2} \end{vmatrix} e^{g_2 t} \tag{13}$$

where g_1 and g_2 are the eigenvalues of $\mathbf{k}$ and

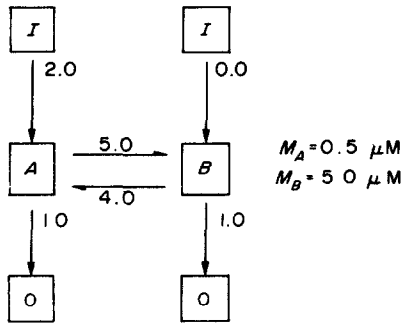
$$\begin{vmatrix} X_{A1} \\ X_{B1} \end{vmatrix} \text{ and } \begin{vmatrix} X_{A2} \\ X_{B2} \end{vmatrix}$$

are the corresponding eigenvectors§. The factors Z_1 and Z_2 are normalization constants whose role is to ensure that at time $t = 0$, all the isotope resides in the pool to which it was added, and none in any other.

From equations (11) and (12) it is an easy task to calculate the incorporation vs time functions, and therefore to obtain discrete incorporation values for particular sampling points in any biosynthetic experiment involving a pathway of known structure.

Example of the calculation of discrete I% values for a pathway of known (or hypothetical) structure

It is believed that the structure shown in Scheme 2



Scheme 2. (all flows, μM/hr).

defines a particular biosynthetic pathway. To check its validity, two experiments were set up. In the first, isotope was added at time zero to the A pool while in the second it was added initially to the B pool. In both experiments incorporation measurements were made in each pool every 30 min for 5 hr by standard dilution analysis. The values predicted by Scheme 2 were needed to compare with the experimentally determined ones and so check the proposal.

From Scheme 2, the augmented flow matrix $\bar{\mathbf{F}}$ can be constructed directly as (14). The diagonal elements, $-F_{AA}$, etc. are computed as the negative sum of

	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{I}$
$\mathscr{A}$	−6.0	4.0	2.0
$\mathscr{B}$	5.0	−5.0	0.0
$\mathscr{O}$	1.0	1.0	−2.0

(14)

	$\mathscr{A}$	$\mathscr{B}$
$\mathscr{A}$	−12.0	0.8
$\mathscr{B}$	10.0	−1.0

(15)

all the other elements in that particular column. From $\bar{\mathbf{F}}$, $\mathbf{k}$ (15), the matrix whose eigenparameters are needed, is obtained by dividing each column of $\bar{\mathbf{F}}$ by the appro-

* In the physiological field (e.g. Atkins [2], and Shipley and Clark [4], greater use of specific activity time courses is made than of incorporation value time courses. To demonstrate that both can yield the same information, Appendix 1 traces the relationship between $\bar{\mathbf{F}}$ and specific activity versus time functions.

† Dilution values *per se* will not be considered in this article. They are obtainable from incorporation values through the relationship

$$D_{xy} = \frac{100}{I_{xy}\%} \cdot \frac{M_y}{M_x}.$$

‡ In describing incorporation functions and values the first subscript indicates the pool to which label was added initially, the second the one in which it is sought eventually. When both subscripts are the same, e.g. $I_{AA}\%$, depletion from the fed pool itself is being examined.

§ A quick word of consolation may be in order here: most, if not all, computer centres have 'canned' programmes which, with little or no modification by the bioorganic chemist, will yield eigenparameters for any $\mathbf{k}$ matrix provided. The biosynthetic investigator interested in using this approach, need not become an expert computer programmer or mathematician.

priate pool size and eliminating the $\mathscr{C}$ row and $\mathscr{S}$ column. The eigenparameters of **k**, once obtained*, allow the general incorporation vs time functions for the system of Scheme 2 to be established as 16 and 17:

$$\begin{vmatrix} Q_A \\ Q_B \end{vmatrix} = Z_1 \begin{vmatrix} -0.760 \\ 0.650 \end{vmatrix} e^{-12.68t} + Z_2 \begin{vmatrix} -0.068 \\ -0.998 \end{vmatrix} e^{-0.32t} \quad (16)$$

$$(17)$$

In the first experiment isotope was to be added to the A pool at time zero; hence for it at time zero, Q_A^0 will be finite while Q_B^0 will be zero. Substituting these values and $t = 0$ in equations (16) and (17) and solving for Z_1 and Z_2 yields†:

$$Z_1 = -1.243, Z_2 = -0.810 \quad (18)$$

These values of the normalization coefficients, taken with equations (9), (10), (16) and (17), provide equations (19) and (20) which are the incorporation versus time functions governing the particular form of Scheme 2 in which isotope is fed initially to the A pool. From these functions the discrete $I\%$ values expected at every half hour sampling point in the first experiment, can be calculated. These values are shown in Table 1, part A.

$$\begin{vmatrix} I_{AA}\% \\ I_{AB}\% \end{vmatrix} = \begin{vmatrix} 94.5 \\ -80.8 \end{vmatrix} e^{-12.68t} + \begin{vmatrix} 5.5 \\ 80.8 \end{vmatrix} e^{-0.32t} \quad (19)$$

$$(20)$$

To obtain the corresponding values for experiment 2 in which isotope was fed initially to the B pool, equations (16) and (17) are solved for the Z factors such that at zero time Q_A^0 will be zero but Q_B^0 will be finite. The appropriate incorporation vs time functions for the B feeding, equations (21) and (22), yield the discrete data shown also in Table 1, part B.

$$\begin{vmatrix} I_{BA}\% \\ I_{BB}\% \end{vmatrix} = \begin{vmatrix} -6.5 \\ 5.5 \end{vmatrix} e^{-12.68t} + \begin{vmatrix} 6.5 \\ 94.5 \end{vmatrix} e^{-0.32t} \quad (21)$$

$$(22)$$

Note that the two exponential coefficients remain unchanged even though the point of label administration was changed from pool A to pool B. This arises because these coefficients are eiqenvalues of **k**, and **k** is invariant for any system of known structure irrespective of where label is added. This fact proves very useful in pathway mapping, as will be shown later.

Summary of the step-by-step procedure of obtaining discrete I% values for a pathway of known structure‡

The following protocol, although introduced here in the context of the two component system, is quite general. It will yield incorporation vs time functions for a pathway of any size and any structure that abides by the three constraints mentioned at the beginning of "Methods and Results."

(a) Set out the structure of the pathway diagrammatically as in Scheme 2, filling in all the flows and noting all pool sizes,

(b) establish the matrix $\bar{\mathbf{F}}$ (8) using the flows found in the structure diagram, and employing the fact that its rows and columns must sum to zero, to obtain values that do not appear in the structure diagram‡,

(c) derive the **k** matrix from $\bar{\mathbf{F}}$ and the pool size data‡,

(d) find the eigenparameters of **k** and so establish the general incorporation vs time functions (e.g. (16) and (17)) that govern isotope flow in the particular structure‡,

(e) specify which pool has to receive the isotope initially and solve for the specific set of Z factors‡.

---

*Most scientific subroutine packages have various programmes to solve eigenparameter problems. Our programme INCORP uses the subroutines ELMMES, ELTRAN and HQR2 of EISPCK available through Argonne National Laboratories (applied Mathematics Division), Argonne, Ill., 60439, U.S.A. For small matrices such as (15), the eigenparameters can be calculated by hand. While we recommend that readers make use of computer facilities whenever possible—it allows $I\%$ values for any size of system to be obtained with equal ease—we accept the advice of a reviewer and summarize below the procedure of solving the **k** matrix (15).

Let an eigenvalue of (15) be λ and the corresponding eigenvector be

$$\mathbf{X} = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix},$$

then by definition

$$\mathbf{k} \cdot \mathbf{X} = \lambda \mathbf{X}$$

i.e.

$$(\mathbf{k} - \lambda \mathbf{1})\mathbf{X} = 0$$

For the above equation to have a non-trivial solution, the determinant $|\mathbf{k} - \lambda \mathbf{1}|$ must be zero, i.e.

$$\begin{vmatrix} (-12 - \lambda) & 0.8 \\ 10.0 & (-1.0 - \lambda) \end{vmatrix} = 0$$

i.e.

$$(12 + \lambda)(1 + \lambda) - 8 = 0$$

whence

$$\lambda = -12.68 \text{ or } -0.32$$

Each of these eigenvalues must obey the condition $\mathbf{k} \cdot \mathbf{X} = \lambda \mathbf{X}$

i.e.

$$\begin{vmatrix} -12.0 & 0.8 \\ 10.0 & -1.0 \end{vmatrix} \cdot \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = -12.68 \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}.$$

i.e.

$$-12.0x_1 + 0.8x_2 = -12.68x_1$$

and

$$10.0x_1 - 1.0x_2 = -12.68x_2$$

whence

$$x_1 + 1.18x_2 = 0$$

For **X** to be normalized,

$$\sqrt{(x_1^2 + x_2^2)} = 1$$

since

$$x_1 = -1.18x_2$$

$$\sqrt{(1.39x_2^2 + x_2^2)} = 1$$

hence

$$x_2 = \pm 0.65 \text{ and } x_1 = \mp 0.76$$

Since **X** is still an eigenvector when multiplied by any constant, both

$$\begin{vmatrix} -0.76 \\ 0.65 \end{vmatrix} \text{ and } \begin{vmatrix} 0.76 \\ -0.65 \end{vmatrix}$$

are answers, choose any one; the Z factors in (16) and (17) will accommodate the choice to the specific system under study.

† It can be shown that Z_1 and Z_2 are the elements of that column in the inverse of **k** ($\mathbf{k}^{-1}$) that corresponds to the row in **k** whose index is the pool into which isotope was added initially, i.e. if isotope was added to the A pool, Z_1 and Z_2 will be the elements of the first column of $\mathbf{k}^{-1}$.

‡ The steps marked ‡ are accomplished directly by the FORTRAN IV programme INCORP (written for DEC PDP-10). A listing of INCORP and all other programmes mentioned in these summaries are available from the author at no charge.

Table 1. Discrete incorporation values obtained from the structure diagram shown in Scheme 2

Time (hr)	Part A Label added to A I_{AA}%	I_{AB}%	Part B Label added to B I_{BA}%	I_{BB}%
0.5	4.85	68.71	5.53	80.54
1.0	4.00	58.67	4.72	68.62
1 5	3.40	50.00	4.02	58.48
2.0	2.90	42.61	3.43	49.83
2.5	2 47	36.31	2.92	42.46
3.0	2.11	30.94	2.49	36.18
3.5	1 80	26.36	2.12	30.83
4.0	1.53	22.47	1.81	26.28
4 5	1.30	19.14	1.54	22 39
5.0	1 11	16.31	1.31	19.08

(f) use the Z factors to obtain the incorporation vs time functions [e.g. (19)–(22)] that apply to the particular feeding experiment*,

(g) calculate the discrete I% values for any desired sampling point(s) from the functions obtained in (*f*)‡.

Obtaining the augmented flow matrix $\bar{\mathbf{F}}$, *and the pathway structure diagram from the coefficients of incorporation vs time functions*

The way back to the structure of a pathway from incorporation data has two stages, the second of which is examined first. Let it therefore be assumed that sets of discrete, experimentally determined I% values, such as those appearing in Table 1, have been fitted to functions of the form $I = \Sigma X_i e^{g_i t}$*. It will be assumed further that we are again dealing with a two component system, and that isotope was fed initially to A. From the outcome of the section on page 1369, it follows that for the two pool system there will be two incorporation versus time functions, I_{AA} and I_{AB}, each of which will be the sum of two or less terms†. Each of these terms will feature as its exponential coefficient one of the two eigenvalues of the **k** matrix that defines the system. Moreover, when the incorporation versus time functions are arranged in consecutive rows in accord with the known metabolic sequence, and such that the terms with a common exponential coefficient fall in columns, the preexponential coefficients, X, form the eigenvectors of the **k** matrix‡. The eigenvectors are associated with the eigenvalue that is the common exponential coefficient of the column. Specifically, if equations (23) and (24) are the properly arranged incorporation versus time functions, g_1 and g_2 are the two eigenvalues of **k**, and

$$\begin{vmatrix} X_{A1} \\ X_{B1} \end{vmatrix} \text{ and } \begin{vmatrix} X_{A2} \\ X_{B2} \end{vmatrix}$$

are the corresponding eigenvectors.

$$I_{AA} = X_{A1}\, e^{g_1 t} + X_{A2}\, e^{g_2 t} \tag{23}$$

$$I_{AB} = X_{B1}\, e^{g_1 t} + X_{B2}\, e^{g_2 t} \tag{24}$$

To obtain the structure of the pathway that gave rise to the incorporation vs time function, the **k** matrix for the system has to be generated from its eigenparameters. Fortunately, this is an operation that can be performed readily and again the necessary programmes will be found in most computer centers. The procedure involves the following steps:

(a) make the matrix **X** (25) in which the preexponential coefficients of equations (23) and (24) appear as elements,

$$\begin{matrix} X_{A1} & X_{A2} \\ X_{B1} & X_{B2} \\ (25) \end{matrix} \quad \Big\| \quad \begin{matrix} g_1 & 0 \\ 0 & g_2 \\ (26) \end{matrix}$$

(b) make the matrix **g** (26) in which the diagonal elements are g_1 and g_2 and all other elements are zero,

(*c*) invert **X** to $\mathbf{X}^{-1}$ and perform the matrix multiplication $\mathbf{X}.\mathbf{g}.\mathbf{X}^{-1} = \mathbf{k}$.

(It is essential to recognize at this stage that, while it is important in setting up (25) and (26) to associate each eigenvector with the appropriate eigenvalue, i.e. if the first column in (25) is

$$\begin{vmatrix} X_{A1} \\ X_{B1} \end{vmatrix}$$

the first column in (26) must be

$$\begin{vmatrix} g_1 \\ 0 \end{vmatrix},$$

there is no natural or significant sequencing of the eigenvalues of a matrix. Hence,

$$\begin{vmatrix} X_{A2} \\ X_{B2} \end{vmatrix}$$

could just have easily been the first column of (25); if this had been done, however,

$$\begin{vmatrix} g_2 \\ 0 \end{vmatrix}$$

would have to have been the first column in (26). This fact simplifies pathway structure mapping considerably.)

With the **k** matrix in hand, $\bar{\mathbf{k}}$ can be constructed using the fact introduced in the section on page 1368, that the elements of any column in $\bar{\mathbf{k}}$ must sum to zero. To go to either of the flow matrices requires that pool size data be available. Division of each column in $\bar{\mathbf{k}}$ by the appropriate mass yields **F** which can be augmented to $\bar{\mathbf{F}}$ using the fact that the elements of the $\mathscr{S}$ column must be chosen such that the sum of the rows of $\bar{\mathbf{F}}$ be zero. From $\bar{\mathbf{F}}$ it is a simple task to construct the appropriate structure diagram of the biosynthetic pathway.

Example of the generation of the structure diagram of a biosynthetic pathway from incorporation vs time functions

In this section, we will work back to Scheme 2 from the incorporation vs time equations (19) and (20), and

* From this point on it is more convenient to switch from the I% designation of incorporation to I, with the understanding that $I\% = I \times 100$. Such a switch reduces inaccuracies associated with round-off errors in matrix inversion.

† If one, or more elements in the **X** matrix (see later) is zero, less than a full complement of terms will appear in one, or more, of the incorporation vs time functions. For more details see Appendix 2.

‡Recall that if an eigenvector **M** of a given matrix is multiplied by a constant, e.g. Z, the product is also an eigenvector of the matrix. Hence, normalizing the Q functions to obey the initial value conditions does not alter the fact that the preexponential coefficients of the incorporation vs time functions are eigenvectors of **k**.

(21) and (22). To convert the $I\%$ functions to I functions, divide all the preexponential coefficients by 100. It is readily seen that the appropriate form of the **X** and **g** matrices for the situation where isotope was added to pool A initially are (27) and (28) respectively.

$$\begin{array}{cc} 0.945 & 0.055 \\ -0.808 & 0.808 \end{array} \quad \begin{array}{cc} -12.68 & 0 \\ 0 & -0.32 \end{array}$$

(27) (28)

Inversion of (27) gives (29), and performance of the multiplication (27). (28). (29) yields (30) for **k**. Provided the information that $M_A = 0.5$ μM and $M_B = 5$ μM,

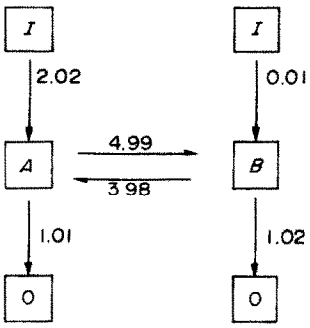
$$\begin{array}{cc} 1.00 & -0.068 \\ 1.00 & 1.70 \end{array} \quad \begin{array}{cc} & \mathscr{A} \quad\quad \mathscr{B} \\ \mathscr{A} & -12.000 \quad 0.795 \\ \mathscr{B} & 9.987 \quad -1.000 \end{array}$$

(29) (30)

the **F** matrix can be calculated from (30). From **F**, the structure defining $\bar{\mathbf{F}}$ matrix is obtained as (31); and from the latter the structure diagram (Scheme 3) can be constructed.

$$\begin{array}{cccc} & \mathscr{A} & \mathscr{B} & \mathscr{I} \\ \mathscr{A} & -6.00 & 3.98 & 2.02 \\ \mathscr{B} & 4.99 & -5.00 & 0.01 \\ \mathscr{O} & 1.01 & 1.02 & -2.03 \end{array}$$

(31)

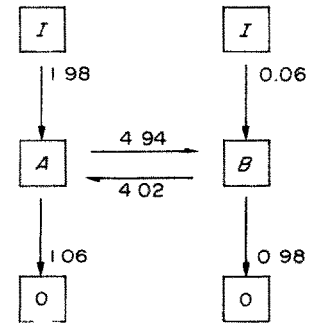


Scheme 3. (all flows, μM/hr)

Equations (21) and (22), which govern the case where isotope was fed initially to the B pool yields by similar procedures the $\bar{\mathbf{F}}$ matrix (32) and the structure diagram of Scheme 4. The slight inconsistencies between Schemes 2, 3, and 4 are

$$\begin{array}{cccc} & \mathscr{A} & \mathscr{B} & \mathscr{I} \\ \mathscr{A} & -6.00 & 4.02 & 1.98 \\ \mathscr{B} & 4.94 & -5.00 & 0.06 \\ \mathscr{O} & 1.06 & 0.98 & -2.04 \end{array}$$

(32)



Scheme 4. (all flows, μM/hr)

occasioned by round-off errors in making loops such as Scheme 2 → equations (19) and (20) → Scheme 3. Structure diagrams such as Schemes 3 and 4, with evidences of small leaks and reversals of flow, are typical of what real systems can be expected to yield.

*Summary of the step-by-step procedure of obtaining the structure diagram of a biosynthetic pathway from its incorporation vs time function**

As with the previous summary, this protocol is general and can be applied to a system of any size of structure:

(a) Convert $I\%$ function to I functions by dividing through by 100,

(b) arrange the functions in rows in accord with the known or proposed metabolic sequence, i.e. I_{AA} above I_{AB} or I_{BA} above I_{BB},

(c) position the terms in each row such that those with the same exponential coefficient form columns. The sequencing of the columns is not important,

(d) construct the **X** and **g** matrices from the preexponential and exponential coefficients, taking care to place the eigenvalue in the same column of **g** as its eigenvector occupies in **X**,

(e) invert **X** to $\mathbf{X}^{-1}$ and perform the multiplication $\mathbf{X}.\mathbf{g}.\mathbf{X}^{-1}$ to get **k**,*

(f) fill in the $\mathscr{O}$ row of **k** from the column elements of **k** and the fact that the sum of column elements in **k** must be zero,*

(g) input pool sizes and calculate **F**,*

(h) fill in the $\mathscr{I}$ column of $\bar{\mathbf{F}}$ from the row elements of **F** and the fact that the sum of the row elements in $\bar{\mathbf{F}}$ must be zero,*

(i) draw out the structure diagram of the pathway.*

Obtaining incorporation vs time functions from discrete experimental I% values

Up till this point, everything that has been done, although described in terms of the two component biosynthetic system, has been applicable generally. When it comes to fitting sets of discrete experimentally determined $I\%$ values to functions of the forms $\Sigma X_i e^{g_i t}$, however, generalization is not easy. Nonlinear regression methods do not handle the problem well; indeed there are some who believe it should not even be tackled [6]. In our hands, the most effective approach is the orthogonal matrix transformations of Jennings and Osborne [7]. Unfortunately, this method does require that quite a large set of $I\%$ values be available to solve three component systems and larger. Constraints, unique to the biosynthetic tracer experiment, may allow this number to be reduced significantly; that remains to be seen. There is, however, a simple technique for obtaining a fit of the data for a two component system, and for any larger system which can be reduced in part or *in toto* to that level. It is reviewed briefly here; for more details see Shipley and Clark [4], or Atkins [2].

Formally, the problem is to find the numerical value of six constants, the four preexponential (X_{A1}, X_{A2}, X_{B1}, and X_{B2}) and the two exponential (g_1 and g_2) coefficients in the incorporation vs time functions (23) and (24)†. Expressed alternatively, this amounts to finding the four unknowns in the **X** matrix (25) and the two unknowns in the **g** matrix (26). The number of unknowns in (25) can be reduced quickly from four to two because of the situation that is known to prevail initially, i.e. $I^0_{A4} = 1$ and $I^0_{AB} = 0$. Substituting these values in (23) and (24) with $t = 0$, yields $X_{A1} + X_{A2} = 1$ and $X_{B1} + X_{B2} = 0$.

* The steps marked * are accomplished directly by the FORTRAN IV programme STRUCT (written for DEC PDP-11).

† For the sake of argument, isotope will again be assumed to have been added at zero time to the A pool.

Since the ordering of columns in **X** is immaterial (see above), (32a) and c are equivalent, as are (32b) and d.

$$\begin{vmatrix} X_{A1} & (1-X_{A1}) \\ X_{B1} & -X_{B1} \end{vmatrix} \quad \text{(32a)} \qquad \begin{vmatrix} X_{A1} & (1-X_{A1}) \\ -X_{B2} & X_{B2} \end{vmatrix} \quad \text{(32b)}$$

$$\begin{vmatrix} (1-X_{A2}) & X_{A2} \\ -X_{B2} & X_{B2} \end{vmatrix} \quad \text{(32c)} \qquad \begin{vmatrix} (1-X_{A2}) & X_{A2} \\ X_{B1} & -X_{B1} \end{vmatrix} \quad \text{(32d)}$$

Provided the correct time frame has been chosen in which to collect the experimental I data, there is a very simple way of obtaining estimates of at least some of the remaining unknown constants. Equations such as (23) and (24), that are the sums of exponential terms in which the exponential coefficients are always negative*, tend to become dominated as the independent variable t increases, by the term that contains the exponential coefficient of smaller absolute value†. The existence of this dominance is best detected by the linearity of $\log_e I$ with time. Moreover, from this linear relationship, when it occurs, come estimates of the smaller exponential coefficient (gradient of the linear portion), and the preexponential coefficient (ordinate intercept of the linear portion). Hence, an effective way to start searching for the four unknowns needed to solve the two component system is to plot the $\log_e$ of the experimentally determined I_{AA} and I_{AB} values against time and check for linearity. If linearity is found the gradient and ordinate intercept should be measured. This can be done graphically or by standard methods of linear regression. If no linear zone‡ is found for I_{AA} and/or I_{AB}, the time frame of the experiment should be expanded to include some later sampling points.

Assuming a linear zone has been found in both the I_{AA} and I_{AB} data, there are two possible outcomes. If two distinct gradients are found, i.e. the $\log_e I_{AA}$ and $\log_e I_{AB}$ vs time plot are not parallel, this most commonly means X_{A2} is zero§. Hence the gradient obtained from tracking the I_{AA} data can be considered g_1 and the corresponding intercept will be X_{A1}; X_{A1} should have a value of unity (see Appendix 2). The gradient obtained from the I_{AB} data plot will be g_2 and the intercept will be X_{B2}. In accord with form (32b) of the **X** matrix, X_{B1} will be $-X_{B2}$. The system is therefore defined completely.

If both gradients obtained by the plotting procedure are equal, i.e. the plots are parallel, the value of the gradient obtained can be considered g_1. The intercept of the I_{AA} data plot will be X_{A1}, that of the I_{AB} plot will be X_{B1}. To solve the system, the other g value has to be obtained. This can be done easily provided incorporation data has been collected in the nonlinear zone of I_{AA} and/or I_{AB}. As is shown in the examples given below, use of such incorporation data points together with equations (23) and/or (24) which have been modified to reflect the fact that g_1 and all four X values are known, yields g_2. If the linear zone of neither I_{AA} nor I_{AB} has been penetrated in the experiment, the time frame of the latter would best be modified to include some earlier sampling points. If this cannot be done feasibly, only a maximum limiting value of the outstanding g value can be obtained. This is obtained from the fact that at the earliest sampling point in the experiment, the term $X_{A2}e^{g_2 t}$ in equation (23) or $X_{B2}e^{g_2 t}$ in equation (24) is no larger in magnitude than the experimental error in determining any I value. Since X_{A1} and X_{B2} can both be calculated ($X_{A1} + X_{A2} = 1$, $X_{B1} + X_{B2} = 0$)

$$g_2 \leqslant -\left[\log_e \frac{X_{A2}(\text{or } X_{B2})}{\text{experimental error}}\right] \Big/ \text{time of earliest sampling} \qquad (33)$$

Examples of obtaining structure diagrams for two component biosynthetic systems direct from discrete, experimentally determined incorporation data

Table 2, parts A and B contains sets of $I\%$ values recorded in two hypothetical biosynthetic experiments. The steady state pool sizes of the precursor and product are also indicated. The plots of $\log_e I$ vs time for experiment 1 are shown in Fig. 1. Linear regression analysis established that the four $I_{AA}\%$ points fall on the line $\log_e I_{AA}\% = 4.6\text{–}20.0\,t$. Thus, in the time frame $0.1 < t$ hr, $I_{AA} = 1.00\,e^{-20.0t}$; hence g_1 must be -20.0 and X_{A1} is 1.00. The unity intercept found in tracking I_{AA} indicates that in the system under study there is no reflux of isotope from the pool B back to A (see Appendix 2 for proof). From Fig. 1 it is seen that the linear zone of the I_{AB} data is $1.0 < t$ hr, wherein $I_{AB} = 1.02\,e^{-0.4t}$. Hence g_2 is -0.4 and X_{B2} is 1.02. The complete **X** matrix of the system is as shown in (34), in accord with (32b), and the **g** matrix is (35).

$$\begin{vmatrix} 1.00 & 0 \\ -1.02 & 1.02 \end{vmatrix} \quad \text{(34)} \qquad \begin{vmatrix} -20.0 & 0 \\ 0 & -0.4 \end{vmatrix} \quad \text{(35)}$$

Applying the procedure summarized on p. 1372 to (34)

* This results from the fact that **k** is diagonally dominant and all the diagonal elements are negative.

† This is easily seen. Consider the function $y = e^{-ax} + e^{-bx}$, with a and b both being positive and $a > b$. For all positive values of x, $ax > bx$ and $e^{-ax} < e^{-bx}$. The latter relation will become even more true as x increases. Eventually, e^{-ax} will become so small relative to e^{-bx}, that for all intents and purposes, $y = e^{-bx}$.

‡ The phrases linear/nonlinear zone will be used to signify those time frames wherein the natural logarithm of any incorporation function is linear/nonlinear with time. Since A (the precursor) is usually a fairly small, actively metabolized pool while B (the product) is most often quite large in size, the linear zone for I_{AA} normally occurs much sooner than that for I_{AB}. Indeed, if there is no reflux of isotope from B back to A, there is no nonlinear zone for I_{AA} (see Appendix 2).

§ Equations (23) and (24) contain both g_1 and g_2. Hence, both will be dominated by the same absolutely smaller exponent as t increases, and the plots of $\log_e I_{AA}$ and $\log_e I_{AB}$ data versus time will be parallel. If the latter is not the case, the preexponential coefficient of the term containing the absolutely smaller exponent must be zero in one case. As shown in the Appendix 2, this situation occurs when there is no reflux from B back to A, i.e. $X_{A2} = 0$.

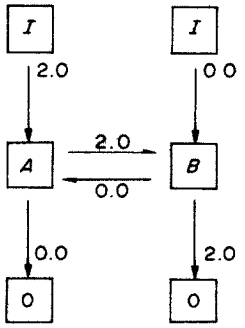
Table 2. Discrete incorporation data for use in the examples of the Section on p. 1373

	Part A		Part B		Part C	
M_A	0.1 μM		0.5 μM		—	
M_B	5.0 μM		5.0 μM		—	
Time (hr)	I_{AA}%	I_{AB}%	I_{AA}%	I_{AB}%	I_{AA}%	I_{AB}%
0.1	13.53	—	18.58	—	45.38	—
0.2	1 83	—	9 27	—	23.15	—
0.3	0 25	—	8.09	—	14.06	—
0 4	0.034	—	7.83	—	10.29	—
0 5	—	—	7.68	—	8.68	—
1	—	68.37	7.02	77.26	7.03	77.25
3	—	30 72	4.90	53.90	4.90	53.90
5	—	13.80	3.42	37.61	3.42	37.61
7	—	6.20	2.38	26.24	2.38	26.24
9	—	2 79	1.66	18.31	1 66	18.31

and (35) yields flow matrix (36) and the pathway structure diagram of Scheme 5.

	𝒜	ℬ	𝒥
𝒜	−2.0	0.0	2.0
ℬ	2.0	−2.0	0.0
𝒪	0.0	2.0	−2.0

(36)



Scheme 5. (all flows, μM/hr)

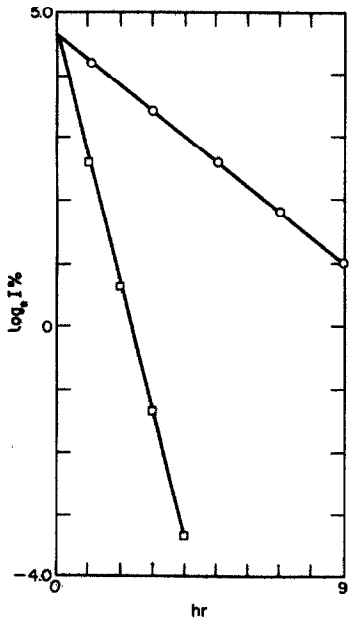


Fig. 1. Plots of the natural logarithm of the incorporation data shown in Table 2, part A, against time. The I% form, rather than the I form has been used. Key: I_{AA}%, □—□—□, I_{AB}%, O—O—O. The time axis for the I_{AA}% data was increased ten fold to allow I_{AA}% and I_{AB}% data to be plotted together. The I_{AA}% points fall on the line $\log_e I_{AA}\% = 4.6\text{–}20.0\,t$, while the I_{AB}% points fall on the line $\log_e I_{AB}\% = 4.63\text{–}0.4\,t$.

The plots of the $\log_e I$% vs time for the second experiment are shown in Fig. 2. Clearly, in the time zone $1.0 < t$ hr, the I_{AA} and I_{AB} functions are being controlled by a common term with exponential coefficient −0.18. This can be considered g_1. The ordinate intercepts are 0.084 (X_{A1}) and 0.925 (X_{B1}) respectively. The second g value can be obtained from any of the data points of Table 2 that are in the nonlinear zone of either I_{AA}% or I_{AB}% or both. The best procedure is to use all data points and find the average g value that is produced. In the present example, there are two points well in the nonlinear zone—the I_{AA}% samplings at 0.1 and 0.2 hr of 18.58 and 9.27%, respectively. These are converted to I values and inserted separately into equation (37) which is the appropriate form of equation (23). The constant 0.916 is obtained from the fact that $X_{A1} + X_{A2} = 1$, and $X_{A1} = 0.084$.

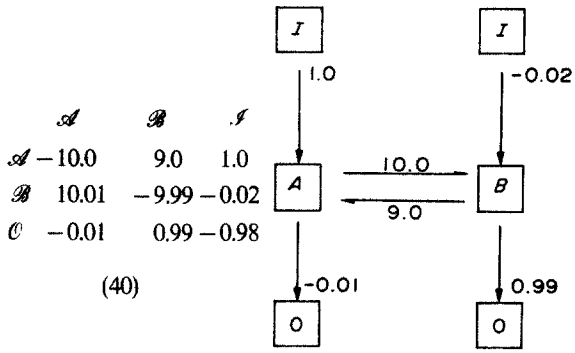
$$I_{AA} = 0.084\,e^{-0.18t} + 0.916\,e^{g_2 t} \qquad (37)$$

An average value of $g_2 = -21.8$ results.

The system can now be solved. The appropriate **X** and **g** matrices are (38) and (39), respectively; together these yield (40) as $\bar{\mathbf{F}}$ and Scheme 6 as the structure diagram.

$$\left|\begin{matrix} 0.084 & 0.916 \\ 0.925 & -0.925 \end{matrix}\right| \left\|\begin{matrix} -0.18 & 0 \\ 0 & -21.8 \end{matrix}\right|$$

(38) (39)



	𝒜	ℬ	𝒥
𝒜	−10.0	9.0	1.0
ℬ	10.01	−9.99	−0.02
𝒪	−0.01	0.99	−0.98

(40)

Scheme 6. (all flow, μM/hr).

In this example, should it have happened that the readings of I_{AA}% up to and including the 0.4th hr had not been made, the specific calculation of the second g value would not have been possible. Using equation (33), however, a maximum value for g_2 could have been obtained. Assuming that the experimental error (probably best estimated from the standard error of the least square line that establishes the linear zone of $\log_e I_{AA}$ vs t) is 0.01, i.e. a 1% error in the incorporation values, equation (33) at the earliest sampling point (0.5 hr) yields

$$g_2 \leqslant -\left[\log_e\left(\frac{0.916}{0.01}\right)\right]\Big/0.5 = -9.03$$

With this value for g_2, the flow matrix (41) and the structure diagram of Scheme 7 would be obtained. Clearly

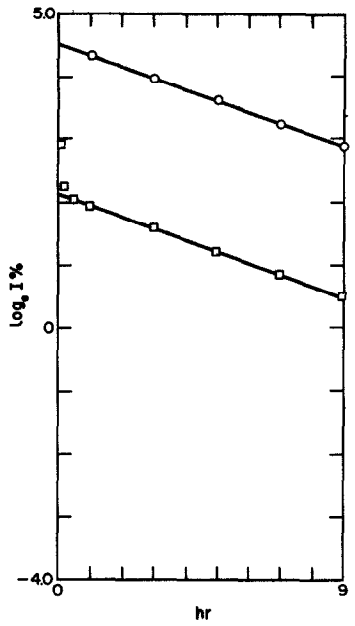
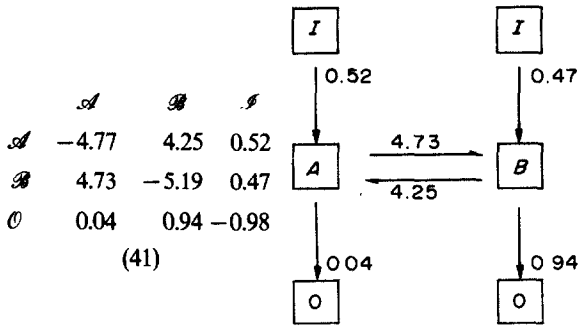


Fig. 2. Plots of natural logarithm of incorporation data shown on Table 2, part B, against time. Key: $I_{AA}\% = \square\text{—}\square\text{—}\square$, $I_{AB}\% = \text{O—O—O}$. The $I_{AB}\%$ points at 0.3 and 0.4 hr have been omitted. The $I_{AA}\%$ points fall on the line $\log_e I_{AA}\% = 2.13\text{–}0.18\,t$, while the $I_{AB}\%$ points fall on the line $\log_e I_{AB}\% = 4.53\text{–}0.18\,t$.

Scheme 7 is not a very good approximation



	𝒜	ℬ	𝒞
𝒜	−4.77	4.25	0.52
ℬ	4.73	−5.19	0.47
𝒪	0.04	0.94	−0.98

(41)

Scheme 7. (all flows, μM/hr).

to Scheme 6. Nevertheless, as is confirmed by the discrete incorporation values calculated from Scheme 7 and collected in Table 2, part c, the structure of Scheme 7 predicts perfectly the experimental data within the set limits of precision and the time frame wherein it was calculated, $0.5 < t < 9$ hr. This happening underscores the necessity of choosing judiciously the time frame in which incorporation data is collected.

Summary of the step-by-step procedure of obtaining the structure diagrams of two component systems from discrete I% *data*

(a) Either plot the logarithm of the I_{AA} and I_{AB} data against time and check for zones of linearity, or programme a search of the data to accomplish the same thing,

(b) if two linear zones are not found, reset the time frames for collecting I_{AA} and/or I_{AB} data,

(c) if two linear zones are found, obtain gradients and ordinate intercepts of both,

(d) if the two gradients differ, set the one obtained from I_{AA} equal to g_1 and the corresponding intercept to X_{A1}. The latter should be unity, or within experimental error thereof. The gradient from the I_{AB} data plot will be g_2 and the corresponding intercept will be X_{B2}. Use the relationship of (32b) to complete the **X** matrix,

(e) if only one gradient emerges from the plots set it equal to g_1 and the intercepts of the I_{AA} and I_{AB} plot to X_{A1} and the X_{B1}, respectively. Obtain g_2 from the data points in the nonlinear zone of I_{AA} and/or I_{AB} using equations (23) and/or (24). If no such data points have been taken, use equation (33) to estimate a maximum value of g_2. Use relation (32a) to construct the **X** matrix,

(f) fill up the **g** matrix in accord with the **X** matrix,

(g) follow the instruction of the section on p. 1372 to derive the structure diagram.

DISCUSSION

Incorporation and dilution values have been used extensively in biosynthetic investigations in plants and fungi since such investigations began in earnest in the early 1950s. They have figured in searches for primary precursors of a given product and in tests to determine if proposed intermediates actually occur in a specific pathway*. In the majority of cases, however, these parameters have been used in a semi-quantitative manner; interest has centered on their orders of magnitude as opposed to their precise value. Such formal lack of rigour was quite justified where the main concern of the studies was to establish the basic metabolic grids of plant and fungal secondary metabolism.

With a few notable exceptions, the outlines of these grids now stand revealed; what remains to be done is fill in the details. Intermediates need to be identified and positioned appropriately on the grids. Actual, intracellular pool sizes of precursors, intermediates and products, require measurement. Some account has also to be taken of the fact that the components in the grid are dynamically related to each other; a means of determining *in vivo*, the rate constant of the various reactions that appear in the grid, has to be evolved. In our opinion, it is only when this whole body of information is in hand, that it can be claimed justifiably that the structure of the grids, or any subsection of them, is 'known'. On the more molecular biological plane, an important set of 'details' relates to how the flux of mass into, through, and out of the grid is prescribed, monitored and regulated.

The concept of the incorporation value, applied with full quantitative rigour, has a role to play in many of these more sophisticated studies. Indeed, in some instances, incorporation value analysis may be the most effective experimental tool to use to obtain results. With regard to the problem of positioning intermediates appropriately in a grid, for instance, the fact that a common set of exponential coefficients (the eigenvalue of the **k** matrix that defines the system) characterizes the incorporation vs time functions of all the components in a

* The place of incorporation and dilution values in biosynthetic methodology has been reviewed by Brown [8].

given biosynthetic pathway, means that there is at least one formal criterion, apart from chemical intuition, for placing a given substance on any particular pathway. Moreover, the specific position that the substance occupies in the pathway is dictated by the ordering of rows in the **X** matrix. These are important outcomes as far as pathway structure mapping is concerned and ones that will be developed more fully hereafter. Their impact on the bioorganic chemist, however, tends to be lessened by the fact that in most cases they will only confirm, albeit rigorously, what has been suggested by other methods.

Such should not be the case when the significance of the augmented flow matrix, $\bar{\mathbf{F}}$, is reviewed. As was shown, this matrix can be constructed from discrete, precise incorporation values that have been collected from a given system as a function of time. The flow matrix contains measures not only of the flux of mass between each component in the system, but also *to* each component *from* the 'outside,' *from* each component to the 'outside,' and *through* the entire system. The flows between components in the system are nothing less than estimates of the actual *in vivo* rate constants of the enzymes that catalyse each specific step in the pathway. The flows between each component and the 'outside' indicate the degree to which intermediates in the pathway are common to other processes of intermediary metabolism. The flow through the system is a measure of the metabolic load the particular pathway is carrying. This kind of information is hard to obtain by other means; as shown above it is easily accessible through incorporation value analysis.

The augmented flow matrix has one other quite intriguing attribute—it provides a means of defining a biosynthetic pathway as a single, integrated operational unit. As attention focuses on the biological significance of the pathways of secondary metabolism and on their relationship to these of primary metabolism, this property of the flow matrix should prove of considerable utility.

Acknowledgements—The author acknowledges with grateful thanks the help, advice and encouragement of his colleagues, C. Bartman, L. Ernst, C. Nulton, J. Slayback, M. Venza and F. Wimberly in preparing this manuscript. The work was performed during tenure of grant number RR00273 and AI11819 of the U.S. Public Health Service and grant number P2B4120 of the National Science Foundation.

REFERENCES

1. Campbell, I. M. (1975) *Phytochemistry* **14**, 683.
2. Atkins, G. L. (1969) in *Multicompartment Models for Biological Systems*, Methuen, London.
3. Jacquez, J. A. (1972) in *Compartmental Analysis in Biology and Medicine*, Elsevier, Amsterdam.
4. Shipley, R. A. and Clark, R. E. (1972) in *Tracer Methods for In Vivo Kenetics* Academic Press, New York.
5. Nulton, C. P., Naworal J. D., Campbell I. M. and (in part) Grotzinger, E. W. *Analyt. Biochem*, in press.
6. Acton, F. S. (1970) in *Numerical Methods that Work*, p. 252. Harper & Row, New York.
7. Jennings, L. S. and Osborne, M. R. in *Applications of Orthogonal Matrix Transformations to the Solution of Systems of Linear and Nonlinear Equations*, Technical Report No. 37 of the Computer Centre, The Australian National University, Canberra, A.C.T. 2600.
8. Brown, S. A. (1972) in *Specialist Periodical Reports—Biosynthesis*, Vol. 1, p. 1. The Chemical Society, London.

APPENDIX 1

In this section it is established that time courses of specific activity versus time can yield $\bar{\mathbf{F}}$. Equations (1) and (2) (see text) that govern the movement of isotope in Scheme 1 can be converted to read in terms of specific activities by dividing through respectively by M_A and M_B; viz. for equation 1.

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{Q_A}{M_A} = k_{BA}\frac{Q_B}{M_A} - k_{AA}\frac{Q_A}{M_A}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}SA_A = k_{BA}\frac{Q_B}{M_A} - k_{AA}SA_A \qquad \text{(a)}$$

since

$$SA_A = Q_A/M_A$$

Recalling

$$k_{BA} = F_{BA}/M_B, \quad k_{AA} = F_{AA}/M_A$$

$$\frac{\mathrm{d}}{\mathrm{d}t}SA_A = \frac{F_{BA}}{M_A}SA_B - \frac{F_{AA}}{M_A}SA_A \qquad \text{(b)}$$

Similarly equation (2) can be restated as

$$\frac{\mathrm{d}}{\mathrm{d}t}SA_B = \frac{F_{AB}}{M_B}SA_A - \frac{F_{BB}}{M_B}SA_B \qquad \text{(c)}$$

If we now let $k^{\dagger}_{AB} = F_{AB}/M_B$, $k^{\dagger}_{BA} = F_{BA}/M_A$ etc.,
i.e. whereas k_{xy} is obtained by dividing F_{xy} by M_x, the pool size designated by the first of the subscripts, $k^{\dagger}_{xy}$ is obtained by dividing F_{xy} by M_y, the pool size designated by the second subscript, equations (b) and (c) can be represented as (d)

$$\mathbf{SA}' = \mathbf{k}^{\dagger}.\mathbf{SA} \qquad \text{(d)}$$

This equation can be solved as was (4) to give SA_A and SA_B as a function of the eigenparameters of $\mathbf{k}^{\dagger}$. Likewise the coefficients in the specific activity versus time functions can be processed to yield $\mathbf{k}^{\dagger}$. It is easily shown that $\mathbf{k}^{\dagger}$ (e) can be augmented $\bar{\mathbf{k}}^{\dagger}$ (f) by inclusion of an I column. From $\bar{\mathbf{k}}^{\dagger}$, $\bar{\mathbf{F}}$ is obtained by multiplying each row in $\bar{\mathbf{k}}^{\dagger}$ by the mass of its index (recall $-k_{AA}{}^{\dagger} = -F_{AA}/M_A$, $k_{BA}{}^{\dagger} = F_{BA}/M_A$, etc.)

	$\mathscr{A}$	$\mathscr{B}$
$\mathscr{A}$	$k^{\dagger}_{AA}$	$k^{\dagger}_{BA}$
$\mathscr{B}$	$k^{\dagger}_{AB}$	$-k^{\dagger}_{BB}$

(e)

	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{I}$
$\mathscr{A}$	$k^{\dagger}_{AA}$	$k^{\dagger}_{BA}$	$k^{\dagger}_{IA}$
$\mathscr{B}$	$k^{\dagger}_{AB}$	$-k^{\dagger}_{BB}$	$k^{\dagger}_{IB}$

(f)

and filling in the $\mathscr{O}$ row on the basis that the columns of $\bar{\mathbf{F}}$ must sum to zero.

APPENDIX 2

The purpose of this Appendix is to provide proof of some statements made in the text and establish for future use a method of using known characteristics of the **k** matrix to simplify the task of finding elements in the **X** matrix.

The **k** matrix of any system is obtained by multiplying together the **X** matrix, the **g** matrix and the inverse of **X**. For the two component system this amounts to the equation:

$$\underset{\text{(i)}}{\begin{vmatrix} -k_{AA} & k_{BA} \\ k_{AB} & -k_{BB} \end{vmatrix}} = \underset{\text{(ii)}}{\begin{vmatrix} X_{A1} & X_{A2} \\ X_{B1} & X_{B2} \end{vmatrix}} \cdot \underset{\text{(iii)}}{\begin{vmatrix} g_1 & 0 \\ 0 & g_2 \end{vmatrix}} \cdot \underset{\text{(iv)}}{\begin{vmatrix} X^{-1}_{A1} & X^{-1}_{A2} \\ X^{-1}_{B1} & X^{-1}_{B2} \end{vmatrix}}$$

Matrix (iv) can be expressed in terms of X_{A1}, X_{A2}, X_{B1} and X_{B2} as (v).

$$\begin{vmatrix} \frac{X_{B2}}{D} & \frac{-X_{A2}}{D} \\ \frac{-X_{B1}}{D} & \frac{X_{A1}}{D} \end{vmatrix} \qquad D = X_{A1}X_{B2} - X_{A2}X_{B1}.$$

(v)

For the case where isotope is added initially to pool A, $X_{A1} + X_{A2} = 1$ and $X_{B1} + X_{B2} = 0$. Under these conditions (v) reduces to (vi) or (vii), which are equivalent.

$$\begin{vmatrix} 1 & \frac{-X_{A2}}{X_{B2}} \\ 1 & \frac{X_{A1}}{X_{B2}} \end{vmatrix} \qquad \begin{vmatrix} 1 & \frac{X_{A2}}{X_{B1}} \\ 1 & \frac{-X_{A1}}{X_{B1}} \end{vmatrix}$$

(vi) (vii)

If either of these are used and the matrix multiplication above is performed, and elements in the product are equated with elements in (i), the outcome is

$$-k_{AA} = g_1X_{A1} + g_2X_{A2} \tag{viii}$$

$$k_{AB} = (g_1 - g_2)X_{B1} = (-g_1 + g_2)X_{B2} \tag{ix}$$

$$k_{BA} = (g_1 - g_2)\frac{X_{A1}X_{A2}}{X_{B1}} = (-g_1 + g_2)\frac{X_{A1}X_{A2}}{X_{B2}} \tag{x}$$

$$-k_{BB} = g_1X_{A2} + g_2X_{A1} \tag{xi}$$

From these equations several conclusions can be drawn:

(a) By virtue of what is known of the **k** matrix, k_{AB} and k_{BA} are both positive; hence the product $k_{AB} \cdot k_{BA}$ will also be positive. Using equations (ix) and (x), it follows that the product $(g_1 - g_2)^2 X_{A1} X_{A2}$ will also be positive, i.e. X_{A1} and X_{A2} must have the same sign. Since $X_{A1} + X_{A2} = 1$, this means both X_{A1} and X_{A2} must be positive real numbers between 0 and 1.

(b) If $k_{BA} = 0$, i.e. there is no flow from B back to A, either $(g_1 - g_2)$ or the product $X_{A1} X_{A2}$ has to be zero. The former case implies $g_1 = g_2$ and indicates that the two component system is behaving as a one component system, i.e. $-k_{AA} = -k_{BB} = g_1$ and $k_{AB} = k_{B0} = 0$. The latter implies that either X_{A1} or X_{A2} are zero. Since there is no significant order to eigenvalues or the corresponding vectors, let $X_{A2} = 0$, i.e. $X_{A1} = 1$. Thus if there is no reflux from B back to A, I_{AA} will be governed solely by the term $X_{A1}e^{g_1t}$. Resultantly, the plots of $\log_e I_{AA}$ and $\log_e I_{AB}$ will not be parallel and there will be no nonlinear zone for I_{AA}.