

Introdução 85-TP

## Estimation of the error implicit in modelling multisubstrate biochemical reactions as lumped-substrate systems

by F. XAVIER MALCATA

Escola Superior de Biotecnologia, Universidade Católica Portuguesa  
Rua Dr. António Bernardino de Almeida, 4200 Porto, Portugal

(Received 15 February 1992)

The necessary and sufficient condition associated with the maximum absolute error arising from the simulation of a biochemical reaction occurring in a batch multi-substrate system as if it were equivalent to a system with a pseudo single substrate is presented. The kinetics of the reaction catalyzed by the soluble enzyme is assumed to be described by a Michaelis-Menten rate equation in the absence of enzyme deactivation, whereas the batch system is assumed to be perfectly mixed. The relevance of the analysis is due to the fact that monitoring the concentration of multiple species in a reaction system (as, in principle, required for controlling purposes) is often technically and/or economically unfeasible especially on the industrial scale. Hence, the strategy reported is useful during the pre-design steps of enzymatic batch reactors as an aid for the decision of whether or not to use complex analytical equipment. According to the analysis described, the appropriateness of the lumped substrate approach needs evaluation at a single point using an algorithm which is of straightforward implementation.

### 1. Introduction

Most biological catalysts possess a very high degree of specificity due to their role in the biochemical pathways of living organisms. Several enzymes possess, however, affinity to a variety of substrates provided that their molecules share a common type of labile covalent bond [1]. In this situation, if a multi-substrate feedstock is provided, the various reactants therein compete for the active site of the enzyme, each one playing the role of a competitive inhibitor of each other. A typical example of this behaviour is found in the hydrolytic action of lipases on fats and oils of natural origin, which are formed by complex mixtures of glycerides.

Frequently in experiments and process operation, it is neither possible nor practical to measure and monitor the concentrations of all species present in the reaction mixture. One way to circumvent this difficulty is by lumping together all substrate species into a single imaginary species. For instance, measuring the concentration of each type of fatty acid residue in a mixture of glycerides requires expensive and time-consuming GC or HPLC techniques, whereas measuring the overall content of fatty acid residues can be easily and quickly achieved by ethanolic titration or reading of the overall absorbance of coloured derivatives thereof [2].

Lumping of this type is very widespread in all branches of biochemical kinetics. Although not often explicitly recognized in the literature of the life sciences and biotechnology, lumping is a pervasive practice when investigating, modelling, or designing biochemical reactors [1]. Possible pitfalls of lumping can be easily made apparent by varying the initial composition of the multi-substrate feedstock while keeping the total substrate concentration constant; in this case the observed values of

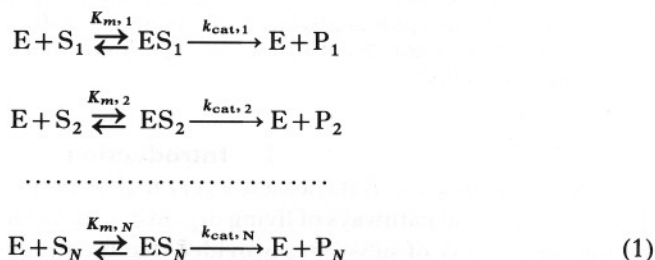
$v_{\max}$  and  $K_m$  will change even though the overall substrate concentration remains unchanged [1].

It is the purpose of this paper to show that there is an upper finite boundary for the absolute error arising from the assumption that a multi-substrate system undergoing an enzyme-catalyzed reaction behaves as a lumped substrate system, and that such value can be obtained using a straightforward numerical procedure. If the aforementioned maximum error is acceptable, then no other tests are in principle required, and the lumped substrate approach may be safely used.

## 2. Mathematical analysis

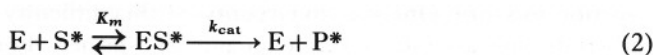
The biochemical reaction of interest here is assumed to be carried out in a batch reactor under isothermal conditions of operation, perfect micromixing, and absence of enzyme deactivation. This type of closed reactor is characterized by ease of operation, low construction costs, and absence of concentration gradients [3]; hence it has a widespread range of applications within the biochemical field on both bench and commercial scales. The enzyme kinetics is assumed to be accurately described by a multi-substrate Michaëlis-Menten rate equation [1].

The multi-substrate stoichiometry and mechanism using the rapid equilibrium approach can be represented as follows



where E denotes the enzyme,  $S_i$  a generic substrate, and  $ES_i$  and  $P_i$  the corresponding enzyme-substrate complex and product, respectively. The Michaëlis-Menten constants are denoted by  $K_{m,i}$ , whereas the maximum velocity is obtained from  $v_{\max,i} \equiv k_{cat,i}[E]_{\text{tot}}$  ( $[E]_{\text{tot}}$  is the total concentration of enzyme active sites).

Similarly, the stoichiometry and mechanism of the lumped substrate approximation can be represented as follows



where  $S^*$  and  $P^*$  are the lumped substrate and product, respectively. The kinetic parameters are now represented by  $K_m$  and  $v_{\max} \equiv k_{cat}[E]_{\text{tot}}$ .

The transient mass balance to each substrate undergoing the enzyme-catalyzed reaction can be written as

$$-\frac{dC_i}{dt} = \frac{\frac{v_{\max,i} C_i}{K_{m,i}}}{1 + \sum_{j=1}^N \frac{C_j}{K_{m,j}}} \quad i=1, 2, \dots, N \tag{3}$$

where  $C_i$  represents the molar concentration of the  $i$ th substrate out of  $N$  different substrates, and  $t$  stands for time. Equation (3) can be rearranged to give

$$\frac{dC_i}{\frac{v_{\max,i}}{K_{m,i}} C_i} = \frac{-dt}{1 + \sum_{j=1}^N \frac{C_j}{K_{m,j}}} = \frac{dC_n}{\frac{v_{\max,n}}{K_{m,n}} C_n} \quad i=1, 2, \dots, n-1, n+1, \dots, N \quad (4)$$

where subscript  $n$  denotes an arbitrary reference substrate. Integration of the left-hand side of the above first equality and the right-hand side of the above second equality subject to the initial condition

$$@C_n = C_{n,0}, C_1 = C_{1,0}, C_2 = C_{2,0}, \dots, C_{n-1} = C_{n-1,0}, C_{n+1} = C_{n+1,0}, \dots, C_N = C_{N,0} \quad (5)$$

leads to the following result

$$C_i = C_{i,0} \left( \frac{C_n}{C_{n,0}} \right)^{\frac{v_{\max,i} K_{m,n}}{v_{\max,n} K_{m,i}}} \quad i=1, 2, \dots, n-1, n, n+1, \dots, N \quad (6)$$

Using equation (6) in equation (4) and integrating both left and right-hand sides of the first equality subject to the initial condition

$$@t=0, C_n = C_{n,0} \quad (7)$$

one finally gets

$$t = \frac{K_{m,n}}{v_{\max,n}} \ln \left\{ \frac{C_{n,0}}{C_n} \right\} + \sum_{j=1}^N \frac{C_{j,0}}{v_{\max,j}} \left( 1 - \left( \frac{C_n}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}} \right) \quad (8)$$

If a lumped-substrate model were considered, then the mass balance to the substrate would read

$$-\frac{dC}{dt} = \frac{\frac{v_{\max}}{K_m} C}{1 + \frac{C}{K_m}} \quad (9)$$

which, upon integration from the initial condition

$$@t=0, C = C_0 \equiv \sum_{j=1}^N C_{j,0} \quad (10)$$

yields

$$t = \frac{K_m}{v_{\max}} \ln \left\{ \frac{C_0}{C} \right\} + \frac{1}{v_{\max}} (C_0 - C) \quad (11)$$

Here  $C$  represents the molar concentration of the hypothetical lumped substrate.

The necessary condition for a local maximum can be mathematically expressed as

$$\frac{\partial}{\partial t} \left\{ \left( \sum_{j=1}^N C_j \right) - C \right\} = 0 \quad (12)$$

Recalling the rules of differentiation and equation (6), equation (12) becomes

$$\frac{\partial C_n}{\partial t} \sum_{j=1}^N \frac{\partial}{\partial C_n} \left\{ C_{j,0} \left( \frac{C_n}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}} \right\} = \frac{\partial C}{\partial t} \quad (13)$$

Using equation (8) and substituting equation (9) in equation (13), one obtains

$$\frac{\sum_{j=1}^N v_{\max,j} \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}}{1 + \sum_{j=1}^N \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}} = \frac{v_{\max} C_{\text{opt}}}{K_m} \frac{1}{1 + \frac{C_{\text{opt}}}{K_m}} \quad (14)$$

where subscript opt refers to the conditions prevailing at the locus of the maximum absolute error. Some algebraic manipulation allows equation (14) to be rewritten as

$$C_{\text{opt}} = \frac{K_m \sum_{j=1}^N v_{\max,j} \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}}{1 + \sum_{j=1}^N \left( 1 - \frac{v_{\max,j}}{v_{\max}} \right) \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}} \quad (15)$$

Use of equations (6) and (15) in equation (12) leads to the following result for the maximum deviation between the multisubstrate and the lumped approaches ( $Er_{\text{opt}}$ ):

$$Er_{\text{opt}} \equiv \left\{ \left( \sum_{j=1}^N C_{j,0} \right) - C \right\}_{C_n = C_{n,\text{opt}}} \\ = \left( \sum_{j=1}^N C_{j,0} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}} \right) - \frac{K_m \sum_{j=1}^N v_{\max,j} \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}}{1 + \sum_{j=1}^N \left( 1 - \frac{v_{\max,j}}{v_{\max}} \right) \frac{C_{j,0}}{K_{m,j}} \left( \frac{C_{n,\text{opt}}}{C_{n,0}} \right)^{\frac{v_{\max,j} K_{m,n}}{v_{\max,n} K_{m,j}}}} \quad (16)$$

The algorithm to be used in obtaining the value of  $Er_{\text{opt}}$  is based on a direct interval having technique and may be outlined as follows:

- Specify the accuracy of the method by setting  $\varepsilon$  equal to a given value (the value of  $C_n$  yielding the maximum absolute error will then be known with an accuracy not worse than  $\pm \varepsilon/2$ ).
- Set  $C_n = C_{n,0}$  (denoted hereafter as  $\alpha$ ) and calculate  $t$  from equation (8) (denoted hereafter as  $t_{n\alpha}$ ), and  $C$  from equation (15); use  $C$  to calculate  $t$  from equation (11) (denoted hereafter as  $t_\alpha$ ); and compute  $\Lambda_\alpha \equiv t_{n\alpha} - t_\alpha$ .
- Set  $C_n = \delta$  ( $\delta$  is a very small positive value) (denoted hereafter as  $\beta$ ) and calculate  $t$  from equation (8) (denoted hereafter as  $t_{n\beta}$ ), and  $C$  from equation (15); use  $C$  to calculate  $t$  from equation (11) (denoted hereafter as  $t_\beta$ ); and compute  $\Lambda_\beta \equiv t_{n\beta} - t_\beta$  (if  $\Lambda_\beta$  has the same sign of  $\Lambda_\alpha$ , use a smaller value  $\delta$ ).
- Set  $C_n = (\alpha + \beta)/2 \equiv \gamma$  and calculate  $t$  from equation (8) (denoted hereafter as  $t_{n\gamma}$ ), and  $C$  from equation (15); use  $C$  to calculate  $t$  from equation (11) (denoted hereafter as  $t_\gamma$ ); and compute  $\Lambda_\gamma \equiv t_{n\gamma} - t_\gamma$ .
- Multiply  $\Lambda_\alpha$  by  $\Lambda_\gamma$ ; if the product is positive, set  $\alpha = \gamma$  and  $\Lambda_\alpha = \Lambda_\gamma$ , and iterate from (d); if the product is negative, set  $\beta = \gamma$  and  $\Lambda_\beta = \Lambda_\gamma$ , and return to step (d).
- Iterate step (e) a number of times equal to the first integer which exceeds  $\ln \{(C_{n,0} - \delta)/\varepsilon\} / \ln \{2\}$ .
- Compute  $Er_{\text{opt}}$  using the latest value of  $C_n$  in equation (16).

The applicability of the analysis developed above is emphasized in the following practical situation.

### 3. Numerical example

Consider the hydrolysis of melted butterfat at 40°C and pH 7.0 brought about in a batch reactor by a lipase from *Aspergillus niger*. This lipase is sn-1, 3 specific, i.e. it acts on either the sn-1 or the sn-3 ester bonds, one at a time. The substrate is a mixture of triglycerides with ten different fatty acid residues independently esterified in the sn-1, sn-2 and sn-3 positions of the glycerol backbone. The rate of release of each type of free fatty acid depends on the type of said fatty acid, but does not depend on whether the fatty acid is esterified at the sn-1 or sn-3 positions, whether the substrate is a triglyceride, diglyceride, or monoglyceride, or on the type of fatty acid residue esterified on any other position not currently under lipolytic attack. Therefore, one can consider each labile ester bond as an independent substrate pseudo-molecule. This is similar to saying that each triglyceride or sn-1, 3 diglyceride corresponds to two independent glyceride molecules, and that each sn-1, 2 or sn = 2, 3 diglyceride or each sn-1 or sn-3 monoglyceride corresponds to one glyceride molecule. The equivalent concentration of glycerides labile to lipase action is then: 0.48 M of butyryl, 0.14 M of caproyl, 0.03 M of caprylyl, 0.07 M of capryl, 0.06 M of lauryl, 0.18 M of myristyl, 0.46 M of palmityl, 0.19 M of stearyl, 0.36 M of oleyl, and 0.015 M of linolenyl [4]. Hence,  $C_0 = 1.985$  M. The values of  $v_{max,i}$  for each of the above substrates are: 0.0585, 0.1019, 0.1165, 0.1057, 0.0829, 0.0588, 0.0388, 0.0241, 0.0270, and 0.0301 M hr<sup>-1</sup>, respectively [2], whereas  $v_{max}$  is equal to 0.0524 M hr<sup>-2</sup>. The values of  $K_{m,i}$  are approximately all the same, and equal to  $K_m$ , which in turn is given by 0.141 M [2]. For convenience, butyryl will be handled as the reference substrate.

Substituting the above values in equations (8), (11), (15), and (16), one gets

$$t_{opt} = 46.0858 - 2.4103 \ln \{C_{1,opt}\} - 17.0937C_{1,opt} - 4.9339C_{1,opt}^{1.7419} - 1.1107C_{1,opt}^{1.9915} - 2.4944C_{1,opt}^{1.8068} - 2.0479C_{1,opt}^{1.4171} - 6.4016C_{1,opt}^{1.0051} - 19.2904C_{1,opt}^{0.6632} - 10.6676C_{1,opt}^{0.4120} - 18.7093C_{1,opt}^{0.4615} - 0.7270C_{1,opt}^{0.5145} \quad (17)$$

$$t_{opt} = 39.7266 - 2.6908 \ln \{C_{opt}\} - 19.0840C_{opt} \quad (18)$$

$$C_{opt} = \frac{1.1164C_{1,opt} + 0.9777C_{1,opt}^{1.7419} + 0.2877C_{1,opt}^{1.9915} + 0.5318C_{1,opt}^{1.8068} + 0.2686C_{1,opt}^{1.4171} + \dots}{1 - 0.8256C_{1,opt} - 3.3684C_{1,opt}^{1.7419} - 1.1226C_{1,opt}^{1.9915} - 1.9020C_{1,opt}^{1.8068} - 0.7008C_{1,opt}^{1.4171} + \dots} \dots \dots \dots \frac{\dots + 0.4224C_{1,opt}^{1.0051} + 0.5542C_{1,opt}^{0.6632} + 0.1182C_{1,opt}^{0.4120} + 0.2603C_{1,opt}^{0.4615} + 0.0126C_{1,opt}^{0.5145}}{\dots - 0.3261C_{1,opt}^{1.0051} + 1.3777C_{1,opt}^{0.6632} + 0.9847C_{1,opt}^{0.4120} + 1.7366C_{1,opt}^{0.4615} + 0.0660C_{1,opt}^{0.5145}} \quad (19)$$

and

$$Er_{opt} = C_{1,opt} + 0.5028C_{1,opt}^{1.7419} + 0.1294C_{1,opt}^{1.9915} + 0.2637C_{1,opt}^{1.8068} + 0.1698C_{1,opt}^{1.4171} + \dots \dots + 0.3764C_{1,opt}^{1.0051} + 0.7485C_{1,opt}^{0.6632} + 0.2571C_{1,opt}^{0.4120} + 0.5052C_{1,opt}^{0.4615} + 0.0219C_{1,opt}^{0.5145} - C_{opt} \quad (20)$$

respectively. Implementing eight iterations of a searching procedure similar to the one described before (with  $\delta = 0.001$ ), one finds that  $C_{1,opt} = 0.04404 \pm 0.0093$  M. Replacing this value in equations (20) and (18) (or (17)), one finally obtains  $Er_{opt} = 0.2185$  M and  $t_{opt} = 42.53$  hr, respectively. This means that the maximum error associated with the determination of the total concentration of substrate using

the unisubstrate approximation for the given multi-substrate system normalized by the total initial concentration of substrates is equal to  $0.2185/1.985 = 11\%$ . This error level is acceptable for practical purposes, so the extent of the reaction can be monitored via use of a lumped substrate analytical methodology.

#### 4. Discussion

Equation (12) is the necessary condition for existence of a critical point (which can be a minimum, maximum, or simply an inflexion point). Equation (10) states that at  $t=0$ ,  $C=C_0$ , which by definition is equal to the sum of the initial concentrations of all species labile to the enzyme catalytic action. Hence, the error of the lumped approach relative to the multi-substrate approach is nil at zero time. On the other hand, as time tends to infinity the concentrations of all substrates tend to zero, and so does the concentration of the hypothetical lumped substrate. Therefore, the error of the lumped approximation approaches zero as time grows without limit. Since equation (15) has a single solution for  $C_{n, \text{opt}}$ , the critical value of  $C_n$  (or, equivalently, the associated critical value for  $t$ ) must necessarily correspond to a maximum of  $|(\sum C_j) - C|$ . The accurateness of the analysis is, thus, established.

It should be noted that the algorithm developed is applicable only to situations where all products of the enzyme-catalyzed reaction cannot in turn function as substrates for the enzyme. If this constraint is not met, then the general situation thereby arising leads to a much more involved analysis, an asymptotic case of which is discussed elsewhere for continuous stirred tank reactors [5].

Although  $C_{n, \text{opt}}$  cannot be explicitly obtained, the numerical algorithm outlined above is fast and converges always to the true solution. Furthermore, it works on the conservative side by providing an estimate of the maximum deviation ever attainable between the results of the uni- and multi-substrate analyses. An alternative algorithm that did not take advantage of the locus of the maximum as given by equation (15) would require an exhaustive search; the steps of such trial-and-error method would consist on setting a value for  $C_n$ , obtain  $t$  from equation (8), obtain  $C_i$  ( $i=1, 2, \dots, n-1, n+1, \dots, N$ ) from equation (6), sum all values of  $C_i$  ( $i=1, 2, \dots, n-1, n, n+1, \dots, N$ ), solve equation (11) for  $C$  using an iterative method, compute  $(\sum C_j) - C$ , and redo the computation process with a new estimate of  $C_n$  until a maximum of  $|(\sum C_j) - C|$  is obtained. This alternative algorithm is much more cumbersome to implement and requires a far larger number of iterations for reasonable final estimates.

The algorithm reported in this paper is useful for the predesign steps of batch stirred reactors aimed at the performance of enzyme-catalyzed reactions. This situation is due to the fact that, by providing an upper boundary of the absolute error to be expected when a lumped substrate approach is employed in a multi-substrate system, the algorithm enables one to ascertain whether or not expensive analytical equipment (usually based on chromatographic techniques) is required to accurately monitor the extent of the reaction.

#### Nomenclature

- $C$  = concentration ( $\text{mol} \cdot \text{dm}^{-3}$ )
- $E$  = enzyme
- $Er$  = error involved in using the lumped substrate approximation to a multisubstrate system
- $ES$  = enzyme/substrate complex



$[E]_{\text{tot}}$  = total concentration of enzyme active sites ( $\text{mol.dm}^{-3}$ )

$i$  = integer variable

$j$  = integer dummy variable

$k_{\text{cat}}$  = first order kinetic constant ( $\text{hr}^{-1}$ )

$K_m$  = Michaëlis-Menten constant ( $\text{mol.dm}^{-3}$ )

$N$  = number of substrates

$P$  = product

$P^*$  = hypothetical lumped product

$S$  = substrate

$S^*$  = hypothetical lumped substrate

$t$  = time elapsed (hr)

$v_{\text{max}}$  = rate of reaction under saturation conditions of substrate  
( $\text{mol.dm}^{-3}.\text{hr}^{-1}$ )

### Subscripts

0 = at the initial conditions

$i$  = referring to the  $i$ th substrate and/or product

$n$  = reference substrate

opt = referring to the conditions leading to the maximum absolute value of  $Er$

### Greek letters

$\alpha$  = upper boundary of searching interval

$\beta$  = lower boundary of searching interval

$\gamma$  = half point of searching interval

$\delta$  = arbitrarily small positive value

$\epsilon$  = tolerance of interval halving technique

### References

- [1] BAILEY, J. E., and OLLIS, D. F., 1986, *Biochemical Engineering Fundamentals* (New York: McGraw-Hill).
- [2] MALCATA, F. X., HILL, C. G., and AMUNDSON, C.H., 1990, Paper presented at the Annual Meeting of the Institute of Food Technologists, Anaheim, CA, USA.
- [3] HILL, C. G., 1977, *An Introduction to Chemical Engineering Kinetics and Reactor Design* (New York: John Wiley and Sons).
- [4] NAWAR, W. W., 1985, in *Food Chemistry* (New York: Marcel Dekker), p. 139.
- [5] MALCATA, F. X., 1991, *J. Chem. Ed.*, **68**, 288-290.