ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Volume 506
November 15, 1987

BIOCHEMICAL ENGINEERING V[a]

Editors and Conference Chairmen
MICHAEL L. SHULER and WILLIAM A. WEIGAND

Intelligent Sensors in Biotechnology

Applications for the Monitoring of Fermentations and Cellular Metabolism

JOSEPH J. VALLINO AND
GREGORY N. STEPHANOPOULOS
Department of Chemical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

INTRODUCTION

In order to control a fermentation, it is necessary to be able to estimate the state of the fermentor from available measurements because the inherent complexity of biological processes precludes the use of solely theoretical identification schemes. However, many of the parameters required for the accurate description of a fermentor are difficult—or virtually impossible—to measure directly due to the diversity of the fermentation environment, the low concentrations of many of the chemical species, or simply the inability to detect the compound of interest. Consequently, what is needed is a sensor that can extract detailed state information from the available measurements by relying on acquired intelligence to dictate the processing of the data; hence the term "intelligent sensors".

An intelligent sensor is any algorithm that couples measurements with some knowledge about the system to produce estimates of states that are not directly measurable. For instance, in a case studied previously, the measured quantities were the rates of oxygen consumption and carbon dioxide production; however, the states of interest were the specific growth rate and the biomass concentration. By using methods similar to the ones presented in this paper, these states were accurately estimated, along with several others. The algorithm used to extract this information is the intelligent sensor.

The objective of this paper is twofold. First, to show how the concentrations of substrate, product, biomass, and oxygen in the gas and liquid phases, as well as the yields, specific growth rate, and the oxygen-mass-transfer coefficient can be estimated from the on-line measurements of oxygen and carbon dioxide in the off-gas and the dissolved oxygen concentration. The robustness and accuracy of the algorithm is demonstrated with an emulsan producing fermentation of *Acinetobacter calcoaceticus*. Second, to describe an algorithm for determining the distribution of carbon among primary metabolic pathways. This algorithm is somewhat unrelated to the first, but it also incorporates the idea of intelligent sensors by amplifying the knowledge obtained by measuring the rates of production of all extracellular compounds and invoking a few pseudo-steady-state assumptions. This algorithm is still in the developing stages, so no quantitative results are yet available. The amino acid fermentation of *Corynebacterium glutamicum* will be the first application of the algorithm.

IDENTIFICATION OF FERMENTATION PARAMETERS

Fermentation

The algorithm described in this section is applied to a fermentation of *Acinetobacter calcoaceticus* carried out in a 14-L fed-batch fermentation at 30 °C with an airflow rate of 4 L/min. In this fermentation, two products are produced: an emulsan that is a polysaccharide with three amino sugars, and an extracellular protein. The carbon substrate used is ethanol and the nitrogen source is ammonium hydroxide; both are added in discrete amounts throughout the fermentation. The available on-line measurements are the dissolved oxygen and the concentrations of carbon dioxide and oxygen in the off-gas, which were sampled every six minutes. Off-line data were taken for the concentrations of substrate, products, and biomass. In addition, for the fermentation considered, acetate was not produced in measurable amounts. The data for the fermentation were obtained from E. J. Schaefer's Ph.D. thesis,[1] which should be consulted for further details on the fermentation. This fermentation was chosen in order to show the applicability of the algorithm to batch fermentations and to show its robustness in diverse applications. It was also chosen because of the more involved oxygen-transfer process prevalent in this non-Newtonian medium of time-varying rheological properties. Hence, the fermentation is merely used as a demonstration and is not important in itself.

The algorithm to be described is an extension of the one originally presented by K. Y. San, G. N. Stephanopoulos, and R. Grosz,[2–5] and it is a modified version of the one presented by J. J. Vallino.[6] Consequently, only the major changes will be elaborated on in this paper. The desired states of interest for this particular fermentation are the concentrations of biomass B, ethanol S, emulsan E, extracellular protein P, dissolved oxygen $[O_2]_d$, and oxygen in the off-gas $[O_2^*]$, along with the specific growth rate μ, the oxygen-mass-transfer coefficient k_la, the rate of ethanol uptake R_S, and the rate of oxygen uptake R_{O_2}. To obtain estimates of these states, a stoichiometric equation for cell growth is incorporated to generate input variables from the on-line measurements. These variables are then fed to an extended Kalman filter, which produces the desired estimates. The algorithm is described below in the order that it is executed on the computer.

The stoichiometric equation for cell growth is a technique devised by C. L. Cooney *et al.*[7] and further extended by K. Y. San[5] to determine the rate of substrate uptake and the rate of biomass and product synthesis from on-line measurements of the oxygen uptake rate (OUR), carbon dioxide evolution rate (CER), and the rate of nitrogen uptake rate (NUR) if a product is being formed. This technique relies on the observation that the cell composition does not change with time or conditions. With this in mind, the following stoichiometric equation can be written for cell growth for the emulsan fermentation:

$$\underset{\text{Ethanol}}{aC_2H_6O} + bO_2 + cNH_3 \rightarrow \underset{\text{Biomass}}{dC_{4.6}H_{8.8}O_2N} + eCO_2 + fH_2O + \underset{\text{Emulsan}}{gC_{10.4}H_{17.1}O_{4.6}N} + \underset{\text{Protein}}{hC_{2.8}H_8O_2N}, \quad (1)$$

where the coefficients a–h are time-varying unknowns. To determine these unknowns, four equations are generated from the elemental balances of carbon, hydrogen, oxygen, and nitrogen. For example, the carbon balance produces $2a = 4.6d + e + 10.4g + 2.8h$. Two more equations are also obtained from the on-line measurements of oxygen uptake rate and carbon dioxide evolution rate shown here:

$$R_{O_2} = 32b, \quad (2)$$

$$R_{CO_2} = 44e. \quad (3)$$

TABLE 1. Initial Conditions and Covariance Matrices for the Kalman Filter Algorithm

State Space Representation

$\mathbf{x}(t) = [B, S, E, P, [O_2]_d, [O_2^*], \mu, k_la, R_S, R_{O_2}]^T$

Initial Conditions

State:

$\bar{\mathbf{x}}(0) = [0.5, 15, 0, 0, 0.007, 0.007, 0.1, 100, 0.1, 0.1]^T$

State-error covariance matrix:

$\mathbf{P}(0) = \text{diag}\{0.01, 0.25, 0.1, 0.1, 2.5 \times 10^{-5}, 2.5 \times 10^{-5}, 0.25, 1 \times 10^5, 1, 3\}$

Noise Covariance Matrices

State:

$\mathbf{Q}(k) = \text{diag}\{1 \times 10^{-4}, 0, 0, 0, 2.5 \times 10^{-7}, 2.5 \times 10^{-7}, 4 \times 10^{-4}, 900, 0.01, 0.01\}$

Measurement:

$$\mathbf{R}(k) = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & 0 & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_{2,2} & 0 & \sigma_{2,4} \\ 0 & 0 & \sigma_{3,3} & 0 \\ \sigma_{1,4} & \sigma_{2,4} & 0 & \sigma_{4,4} \end{bmatrix}$$

where

$\sigma_{1,1} = [0.35Y_1(k)]^2$

$\sigma_{2,2} = [0.40Y_2(k)]^2 \qquad \sigma_{1,1} = \sigma_{1,2}$

$\sigma_{3,3} = [0.05Y_3(k)]^2 \qquad \sigma_{4,4} = \sigma_{1,4} = \sigma_{2,4}$

$\sigma_{4,4} = [0.05Y_4(k)]^2$

However, this still leaves the set of equations underdetermined by two dimensions. Because there were no more useful on-line measurements available to generate independent equations from, the following approximations were made. The off-line data were used to determine the emulsan yield, Y_E, and the extracellular protein yield, Y_P (proportional to g and h, respectively), thus providing the two independent equations needed for closure. Because the off-line data were taken infrequently, the

approximate yields lagged behind their true values from 3 to 10 hours (we will discuss the consequences of this later in the paper).

With the set of equations closed, the following input variables are generated from the measurements:

$$Y_1 = 110d = \mu B + \eta_1, \tag{4}$$

$$Y_2 = 46a = R_S + \eta_2, \tag{5}$$

$$Y_3 = [O_2]_d^m = [O_2]_d + \eta_3, \tag{6}$$

$$Y_4 = [O_2^*]^m = [O_2^*] + \eta_4, \tag{7}$$

where Y_i are the input variables, which are equal to the variables in the center columns of the above equations. The terms on the far right-hand sides of these equations relate the input variables to the state variables of interest. The η_i represent added noises to account for measurement uncertainties in the apparatus and inconsistencies in stoichiometric equation 1. Their standard deviations are set *a priori* based on the accuracy of the equipment, and they are assumed to be Gaussian in nature. The measurement-noise covariance matrix, $\mathbf{R}(k)$, which is listed in TABLE 1, shows the actual values used in the algorithm. Because both input variables Y_1 and Y_2 are functions of the concentration of oxygen in the off-gas, there exists some correlation between η_1, η_2, and η_4. Consequently, off-diagonal terms exist in the measurement-noise covariance matrix, although they do not seem to alter the estimates significantly. Once the input variables are generated, an extended Kalman filter (EKF) is used to extract estimates of the desired state variables. The dynamics of the latter is described by the corresponding balances presented in the following section.

Process Model

A process model for a Kalman filter consists of a set of differential equations describing the dynamics of each state variable as functions of the state space, the disturbances, and time. The following equations are used to model the fed-batch fermentation:

$$\frac{dB}{dt} = \left(\mu - \frac{F_0}{V} - \frac{\dot{V}}{V}\right)B + w_1, \tag{8}$$

$$\frac{dS}{dt} = \frac{F_f S_f}{V} - \left(\frac{F_0}{V} + \frac{\dot{V}}{V} + \frac{1}{100}\right)S - R_S + w_2, \tag{9}$$

$$\frac{dE}{dt} = \mu B Y_E - \left(\frac{F_0}{V} + \frac{\dot{V}}{V}\right)E + w_3, \tag{10}$$

$$\frac{dP}{dt} = \mu B Y_P - \left(\frac{F_0}{V} + \frac{\dot{V}}{V}\right)P + w_4, \tag{11}$$

$$\frac{d[O_2]_d}{dt} = k_l a([O_2^*] - [O_2]_d) - R_{O_2} + w_5, \tag{12}$$

$$\frac{d[O_2^*]}{dt} = \frac{F_A}{V}([O_2^*]_f - [O_2^*]) - k_l a([O_2^*] - [O_2]_d) + w_6, \tag{13}$$

$$\frac{d\mu}{dt} = 0 + w_7, \tag{14}$$

$$\frac{dk_l a}{dt} = 0 + w_8, \tag{15}$$

$$\frac{dR_S}{dt} = 0 + w_9, \tag{16}$$

$$\frac{dR_{O_2}}{dt} = 0 + w_{10}. \tag{17}$$

All but the last four equations represent mass balances around the fermentor. The F_0/V and $\dot{V}/V$ terms in equations 8–11 account for losses due to sampling and dilution effects, respectively. The $1/100$ term in the ethanol balance equation accounts for losses due to evaporation and was determined theoretically by assuming that the liquid-phase ethanol was in equilibrium with the gas phase (i.e., no mass-transfer resistance). The variables, μ, $k_l a$, R_S, and R_{O_2}, occur as parameters in the first six differential equations (equations 8–13); however, because these parameters are unknown and time-varying, they are treated as state variables that the Kalman filter algorithm adjusts to minimize the squared residual error. Also, because their dynamics are unknown, their differentials with respect to time are set to zero. This parameter identification technique differs slightly from that presented previously[2] in that no additional states are incorporated into the right-hand sides of state equations 14–17. It has been found that these simplified equations work well provided that the state-noise covariance matrix, $\mathbf{Q}(k)$, is set correctly.

The w_i in the above model equations account for stochastic disturbances that are not easily represented (e.g., uncertainties in the measured parameters like V, or modeling errors in the equations themselves arising from imperfect mixing, short instrument excursions, and other unaccounted effects). The magnitudes of the w_i are set in the state-noise covariance matrix, which is calculated differently than previously reported. In the previous papers,[2,3] the Riccati equation for the propagation of the state errors was solved in the continuous time domain; this, though, required the numerical integration of a matrix of differential equations. To increase the computational speed and to reduce the likelihood of integration problems, the Riccati equation is placed in discrete form so that only the state transition matrix requires calculation. Although no generally agreed upon method exists for the calculation of the state transition matrix, the algorithm given by R. Ward[8] seems to work quite well. By discretizing the Riccati equation, the state-noise covariance matrix (and *not* the state-noise spectral density matrix) needs to be determined. Previously, the state-noise spectral density matrix was determined adaptively from a modified version of an algorithm presented by Jazwinski[9,10] that relied on the size of the residuals to set the variance of w. Although this technique performs generally well, it was found not to be robust enough to handle large measurement noises or model errors; in fact, the state estimates can diverge under these conditions. To overcome this problem, a technique is employed that is typically

used in parameter estimation.[11] In this scheme, the diagonal terms of the state-noise covariance matrix are set as follows. For each state variable, a magnitude is determined *a priori* by which that variable could possibly be expected to change over the sample interval. For instance, if k_la is not expected to change more than 50 L/h over a six-minute sample period, then the state-noise covariance element that corresponds to k_la is set to 2500.[2] Note that this is only done for those state variables with unknown dynamics, that is, equations 14–17; for equations 8–13, w_i are set to account for reasonable process noise (often zero or very small).

Good knowledge of the system's dynamics is instrumental in determining the above matrix, but once the state-noise covariance matrix is set for a particular process, it need not be changed for different runs that have similar conditions. The value of the state-noise covariance matrix, as well the initial conditions used for the algorithm, are given in TABLE 1.

Further modifications include the addition of two differential equations (equations 12 and 13) to describe the dynamics of oxygen in the gas and liquid phases. Although these equations are usually approximated at steady state, a better (i.e., smoother) estimate of k_la is obtained by using the full differential expressions. Details of the particular Kalman filter algorithm used are explained below.

Extended Kalman Filter Algorithm

The equations and their derivation for the extended Kalman filter can be found in Jazwinski[9] or other references,[11,12] so they are not given here. To minimize problems with numerical integration, a discrete-continuous filter is implemented; that is, discrete equations are used to represent the measurements, while the state equations are solved in the continuous time domain (except for the Riccati equation, which was discretized as explained above). For the actual equations used for the algorithm, see Jazwinski (reference 9, pp. 272–278). The only modification is that the state-noise covariance matrix is set as mentioned above instead of calculated from the state-noise spectral density matrix, and that the measurement-noise covariance matrix, $\mathbf{R}(k)$, is averaged over 10 sample intervals to smooth out spikes.

Advantages of Kalman Filters

Kalman filter algorithms have several advantages over classical least squares (LS) methods. First, LS is for parameters that are time-invariant; this means that the estimates will approach some constant value even though the actual parameters may be (and often are) time varying. By window averaging, this problem can be alleviated; however, the variance of the estimates is often quite large. Secondly, the KF acts as an observer. LS methods can only produce estimates of directly measured parameters, while KF algorithms can estimate indirectly measured states. For instance, in this particular fermentation, only four input variables are generated, but ten states are estimated from them. Thirdly, because the KF can balance the reliability of the model against that of the measurements, very noisy measurements do not necessarily corrupt the state estimates provided that the model accurately describes the process. This can be accomplished by weighting state equations more than the measurements. The only

way to obtain a smooth estimate from LS methods under these conditions is to increase the sampling window; this, though, introduces large filter delays. Finally, the KF can produce estimates of nonobservable states. In this case, the differential equations governing the nonobservable states are simply integrated for all time without updating from the measurements; however, the parameters and states that comprise those differential equations are updated. It should be noted that if the equation governing the nonobservable state is inaccurate, then the estimate can diverge. In this particular implementation, the concentrations of substrate, emulsan, and extracellular protein are all nonobservable states; yet, because the state equations are reasonably accurate descriptions of the true process, fairly good estimates are obtained for these states that would be completely unidentifiable by most other methods.

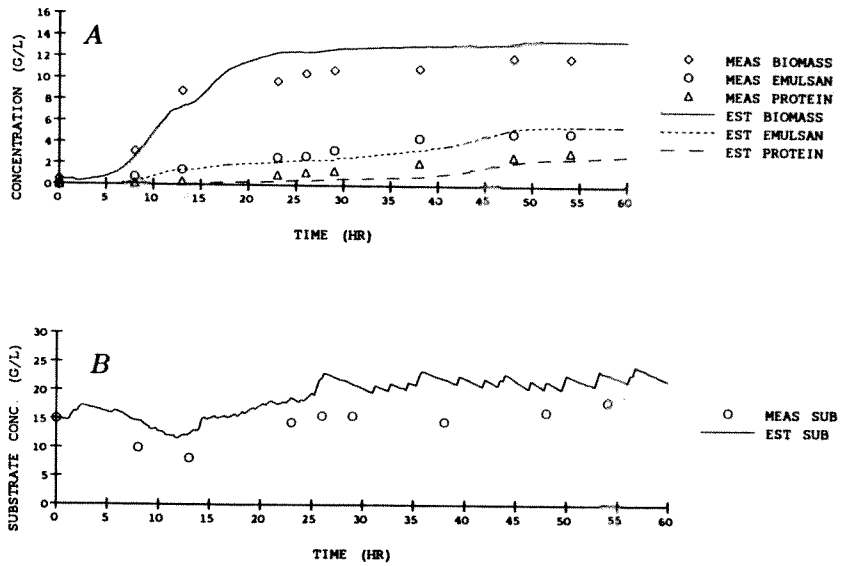

FIGURE 1. Estimated and measured concentrations for the fed-batch fermentation of *Acinetobacter calcoaceticus* of (A) biomass, emulsan, and extracellular protein, and (B) ethanol.

Results

FIGURES 1 and 3 show plots of the state estimates, as well as their comparison to off-line data (when applicable). In FIGURE 1, it can be seen that the estimates of the biomass, emulsan, protein, and ethanol concentrations overshoot or undershoot their measured values by ca. 10–20%. The main reason for this deviation is the fact that the carbon balance used for stoichiometric equation 1 did not close, thus possibly indicating the formation of an unaccounted-for product. In addition, the yields used in the calculations of the input variables lagged 3 to 10 hours, and the ethanol loss term was calculated theoretically for lack of a better estimate. When all these errors are taken into consideration, they produce enough uncertainty to account for the discrepancies between the estimates and the off-line measurements. It should be mentioned

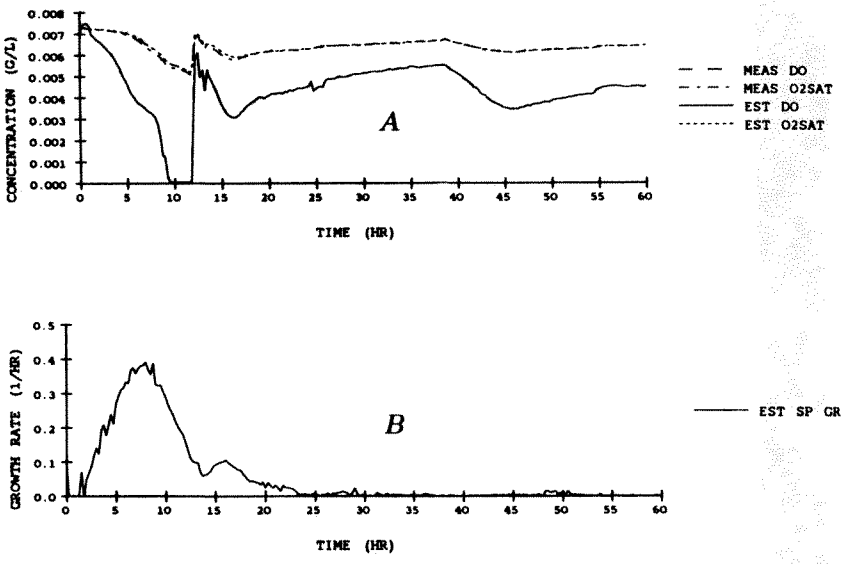

FIGURE 2. (A) Estimated and measured oxygen concentrations in the liquid and gas phases. (B) Estimated specific growth rate.

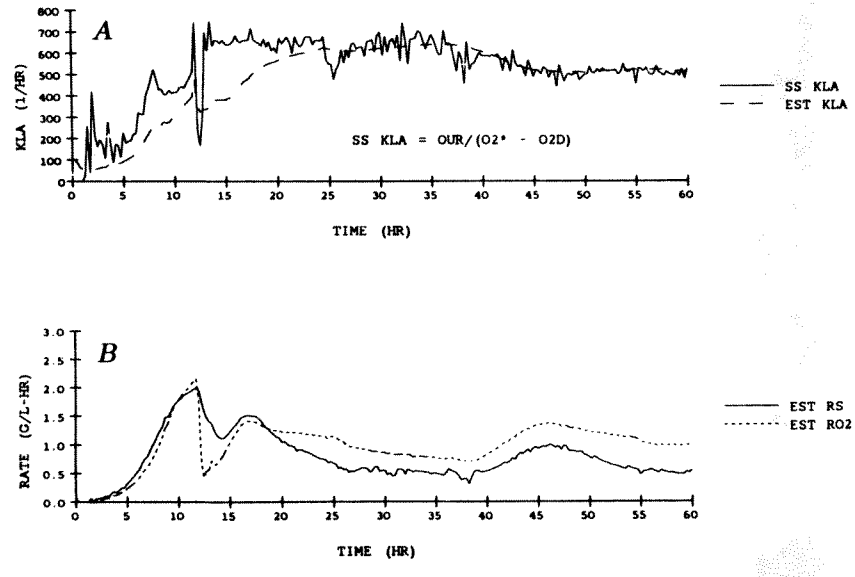

FIGURE 3. (A) Estimated oxygen-mass-transfer coefficient from the EKF algorithm and steady-state equation 18. (B) Estimated rates of substrate and oxygen uptake.

that three of these states are unobservable and could not be estimated at all by most other techniques. Also, the spikes in the ethanol concentration estimate are due to the discrete addition of the feed and not to process or measurement noise.

FIGURE 2A is a plot of the estimated dissolved and gas-phase oxygen concentrations, and includes their corresponding on-line measurements for comparison. The estimates and measurements coincide closely (following the assumption of fairly accurate measurements). TABLE 1 shows the actual uncertainties used. FIGURE 2B shows a plot of the estimated specific growth rate. After about 25 hours, it can be seen that no appreciable growth is occurring. This state could be quite useful in control because it indicates quite vividly where growth stops.

FIGURE 3A shows a plot of the estimated oxygen-mass-transfer coefficient, as well as the value of k_la calculated using the steady-state approximation:

$$(k_la)_{ss} = \frac{\text{OUR}}{([O_2^*] - [O_2]_d)}. \tag{18}$$

Although the steady-state method and the KF estimate produce approximately the same mean values, the steady-state estimate is very noisy and it would be extremely difficult to control a process from such an estimate.

Finally, FIGURE 3B is a plot of the rate of oxygen and substrate uptake. The use of these states is for calculating the yields from the specific growth rate and the biomass concentration:

$$Y_S = \frac{R_S}{\mu B}, \tag{19}$$

$$Y_{O_2} = \frac{R_{O_2}}{\mu B}. \tag{20}$$

Conclusions

These studies provide another example of the application of Kalman filter algorithms for the estimation of unmonitored states for biological processes. The estimation methodology applies equally well to transient and steady-state operations and does not require that the dynamic characteristics of all parameters be known. For example, the specific growth rate was not assumed to follow any particular growth model, such as a Monod model, yet an accurate estimate of this parameter is still produced. This property is valuable because it allows KF algorithms to be applied to a large class of fermentations. Of course, if the specific growth rate was known to follow some type of model, this could be incorporated into the process equations, thereby enhancing the performance of the filter.

Although the KF algorithm can be applied to many fermentations, there are limitations. The method is based on a stoichiometric equation for cell growth to generate input variables; therefore, more measurements are needed to estimate fermentation parameters where several products or substrates are involved. The use of specific probes (such as for glucose, ethanol, organic acids, etc.) in a fermentation is one way to increase the number of measurements. However, the majority of such probes have characteristic times similar to chromatographic measurements, thus

raising the possibility of using HPLC methods for essentially on-line fermentation monitoring. In this case, it may be possible to monitor more variables than needed for state estimation. The next part of this paper will then describe an algorithm that takes advantage of fermentations that are well monitored, along with the concept of intelligent sensors for identification of cellular metabolism.

MONITORING OF CELLULAR METABOLISM

In this section, a new approach is proposed for determining the cellular metabolic activity during product synthesis. According to this scheme, the known primary metabolic pathway stoichiometry is coupled with rate measurements for all extracellular compounds. Mass conservation constraints are then used to determine the extent of reaction for each pathway in the metabolic network. With this information, it may be possible to determine where metabolic bottlenecks exist in product synthesis or to examine how energy gets partitioned between product formation and biomass production. Detected bottlenecks can possibly be alleviated through either proper fermentor control or genetic engineering techniques to increase the amount of those enzymes that display poor activity. The validity of proposed pathways for microbes with poorly understood biochemistry might also be possibly tested by this approach.

Methodology

The derivation of the bioreaction network equation follows the work of E. T. Papoutsakis *et al.*[13,14] on solvent and mixed-acid fermentations. However, the scheme proposed here encompasses the majority of the metabolic pathways, and the emphasis is on the determination of metabolic activity and not on the prediction of rates of product synthesis. To start, all the primary metabolic pathways leading to the synthesis of appreciable amounts of products must be included for the particular microbe of interest. In addition, pathways indirectly involved (such as the respiratory chain, synthesis of intermediates, etc.) must be included in the network. It is only necessary to include compounds that occur at a branch point or are the extracellular terminus of a pathway, so all intracellular species considered are involved in more than two reactions. For example, the synthesis of serine and cysteine from 3-phosphoglycerate is written:

$$3PG + Glut \rightarrow NADH_2 + \alpha KG + Ser \quad (21)$$

$$Ser + AcCoA + H_2S \rightarrow Ac + Cys \quad (22)$$

$$SO_4^{2-} + 3\,ATP + 4\,NADH_2 \rightarrow H_2S \quad (23)$$

$$2\,ATP + Ac \rightleftharpoons AcCoA. \quad (24)$$

In this example, notice that all of the intermediates are omitted and that the synthesis of H_2S and the regeneration of acetyl-Coenzyme A are included. It is further assumed that $NADH_2$ is equivalent to $NADPH_2$ and that ATP → AMP is equivalent to 2 ATP → 2 ADP. Also, for the respiratory chain, the number of ADPs phosphorylated per atom of oxygen, the P/O ratio, is three, although this can be changed if necessary. Once completed, a simplified biosynthetic reaction network emerges like the one

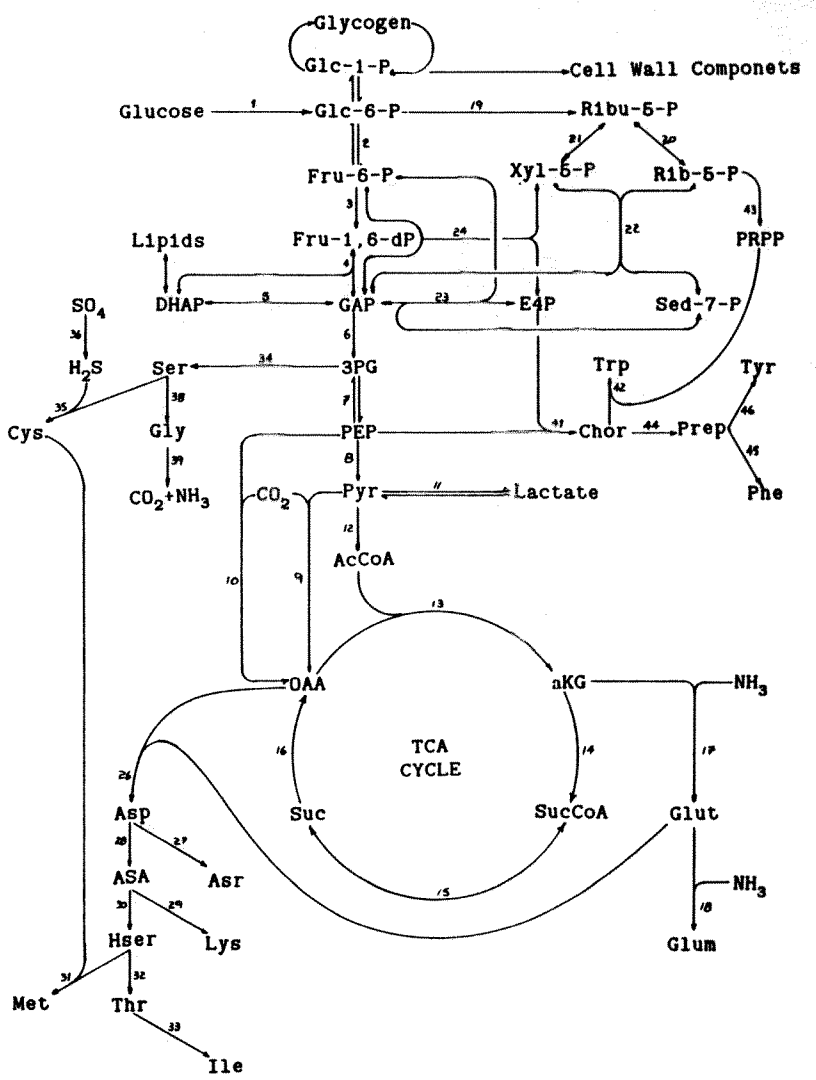

FIGURE 4. Biosynthetic reaction network considered for the production of selected amino acids from *Corynebacterium glutamicum*.

depicted in FIGURE 4, which shows the pathways considered for the synthesis of a few selected amino acids by *Corynebacterium glutamicum*.

Biomass Equation

Biosynthetic pathways need not be considered because biomass is treated as a product; therefore, all pathways needed for biomass synthesis are lumped into a single

equation. This equation is derived as follows. Assuming all carbon and nitrogen that comprise the biomass are obtained from glucose and ammonia, respectively, and if $C_4H_{4p}O_{4n}N_{4q}$ represents the composition of the biomass, then 2 moles of glucose plus 12q moles of NH_3 are required to produce 3 moles of biomass. The amount of reducing power, in the form of $NADH_2$, is determined by a balance of the degrees of reductance. For the biomass as written above, the degree of reductance is given by

$$r_b = 16 + 4p - 8n - 12q. \quad (25)$$

Therefore, $\frac{1}{2}(3r_b - 48)$ moles of $NADH_2$ are required to produce 3 moles of biomass. From the ATP yield, the number of moles of ATP required is $3(MW)_b/Y_{ATP}$, where $(MW)_b = (48 + 4p + 64n + 56q)$ and it is assumed that Y_{ATP} equals 10.5 grams biomass produced per mole ATP consumed (as proposed by W. J. Payne[15]). The overall

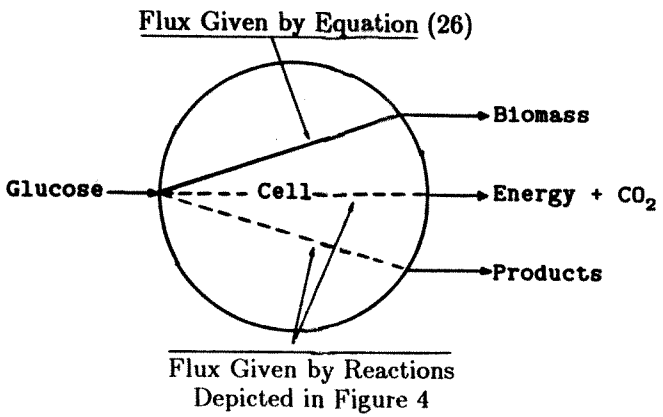

FIGURE 5. Graphical representation of the decoupling of biomass synthesis from product synthesis and energy production.

biomass equation is given by

$$2C_6H_{12}O_6 + 12qNH_3 + \tfrac{1}{2}(3r_b - 48)NADH_2 + 3(MW)_b/Y_{ATP}ATP \rightarrow 3C_4H_{4p}O_{4n}N_{4q}. \quad (26)$$

This equation, in effect, decouples the product formation from biomass production, as illustrated in FIGURE 5.

PROBLEM SETUP

The equations describing the bioreaction network can now be placed in a solvable form as follows. Each compound in the network represents a node; therefore, the rate of production at each node is given by the summation of all flows entering that node *minus* the summation of all flows leaving that node. For example, the rate of

production of ammonia as shown in FIGURE 4 is given by

$$R_{NH_3} = -(17) - (18) + (31) + (33) + (39) - 12q(47), \quad (27)$$

where the numbers in the parentheses indicate the fluxes through the corresponding reactions and flow (47) represents the extent of biomass synthesis (equation 26). These equations are written for each compound and the resulting set of equations are placed in matrix form as follows:

$$\mathbf{Ax} = \mathbf{r} \quad \text{subject to} \quad \mathbf{Cx} \geq \mathbf{b}, \quad (28)$$

where **A** = bioreaction network matrix, **x** = flux vector (unknown), **r** = production rate vector, **C** = constraint matrix, and **b** = contraints. This bioreaction network equation is an extension of the fermentation equation presented by E. T. Papoutsakis.[13] The **r** vector is the measured rate of production of all compounds. For those compounds that are not measurable (i.e., intracellular), a pseudo-steady-state approximation can be invoked, which implies, in effect, that the rate of change of the corresponding compound is small compared to its turnover rate. Hence, all rates that correspond to intracellular species are set to zero. Another approximation is also used if the bioreaction network matrix is not of full rank. In that case, those flows that are undeterminable must be either set or removed from the network through lumping with another reaction, or they must be determined by another independent equation. In the case considered in FIGURE 4, there are 47 unknown flows involving 51 species (i.e., nodes). The dimension of the corresponding bioreaction network matrix is found to be one less of full rank. In this case, flow nos. 8, 9, and 10 (as depicted in FIGURE 4) are indeterminate, and the approximation used to solve this problem is to set flow no. 10 equal to zero because it is believed to be insignificant.

Applicable constraints usually arise from reaction irreversibility. If the solution is not constrained, infinite loops may result where ATP is produced without energy consumption. Consequently, constraints are necessary to insure that thermodynamic laws are not violated. The presence of constraints changes the problem from a matrix inversion to a least squares one, which is usually not amenable to analytical solutions.

A numerical iteration–type scheme based on the procedure proposed by C. K. Liew[16] has been developed for the solution of inequality constrained problems of this type. The algorithm has been tested successfully with the bioreaction network of FIGURE 4 and more tests are planned for the purpose of developing an algorithm of general utility. Experimental application of this approach is presently under way with the amino acid fermentation system of *C. glutamicum*. In this system, the production rates of biomass, glutamate, lactate, CO_2, and amino acids of interest, as well as the consumption rates of glucose, NH_3, and O_2, are being measured. This thus results in an overdetermination of the problem. Additional equations resulting from the extra measurements will be used to test the consistency of the overall biochemical topology, the assumptions invoked, and the measurements obtained. If inconsistencies are detected, attempts will be made to identify the source of errors through a systematic elimination of assumptions and measurements, along with an evaluation of a properly defined consistency index. This method has been applied before to other fermentations.[17] Its application in the context of the studies described in this section can lead to an overall self-consistent structure that can be used for the essentially on-line monitoring of essential features of cellular metabolism. In this sense, the obtained

extra intelligence allows for the amplification of the fermentation measurements for more efficient investigations and better fermentation control.

REFERENCES

1. SCHAEFER, E. J. 1985 (June). Synthesis and release of emulsan by *Acinetobacter calcoaceticus*. Ph.D. dissertation. Massachusetts Institute of Technology.
2. STEPHANOPOULOS, G. & K. Y. SAN. 1984. Studies on on-line bioreactor identification. I. Theory. Biotechnol. Bioeng. **26:** 1176–1188.
3. SAN, K. Y. & G. STEPHANOPOULOS. 1984. Studies on on-line bioreactor identification. II. Numerical and experimental results. Biotechnol. Bioeng. **26:** 1189–1197.
4. GROSZ, R., G. STEPHANOPOULOS & K. Y. SAN. 1984. Studies on on-line bioreactor identification. III. Sensitivity problems with respiratory and heat evolution measurements. Biotechnol. Bioeng. **26:** 1198–1208.
5. SAN, K. Y. & G. STEPHANOPOULOS. 1984. Studies on on-line bioreactor identification. IV. Utilization of pH measurements for product estimation. Biotechnol. Bioeng. **26:** 1209–1218.
6. VALLINO, J. J. 1985 (May). On-line estimation of the oxygen-mass-transfer coefficient and other state variables in a chemostat. Master's thesis. California Institute of Technology. Pasadena, California.
7. COONEY, C. L., H. Y. WANG & D. I. C. WANG. 1977. Computer-aided material balancing for prediction of fermentation parameters. Biotechnol. Bioeng. **19:** 55–67.
8. WARD, R. C. 1977. Numerical computation of the matrix exponential with accuracy estimate. SIAM J. **14:** 600–610.
9. JAZWINSKI, A. H. 1971. Stochastic Process and Filtering Theory. Academic Press. New York.
10. JAZWINSKI, A. H. 1967. Adaptive filtering. Interm Report No. 67-6. Analytical Mechanics Assoc. Lanham, Maryland.
11. GELB, A., Ed. 1974. Applied Optimal Estimation. MIT Press. Cambridge, Massachusetts.
12. BROWN, R. G. 1983. Introduction to Random Signal Analysis and Kalman Filtering. Wiley. New York.
13. PAPOUTSAKIS, E. T. 1984. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol. Bioeng. **26:** 174–187.
14. PAPOUTSAKIS, E. T. 1985. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. Biotechnol. Bioeng. **27:** 50–66.
15. PAYNE, W. J. 1970. Energy yields and growth of heterotrophs. Ann. Rev. Microbiol. **24:** 17–52.
16. LIEW, C. K. 1976. Inequality constrained least-squares estimation. J. Am. Stat. Assoc. **71:** 746–751.
17. WANG, N. S. & G. STEPHANOPOULOS. 1983. Application of macroscopic balances to the identification of gross measurement errors. Biotechnol. Bioeng. **25:** 2177–2208.

APPENDIX

Nomenclature

- **A** Bioreaction network matrix (equation 28)
- a Stoichiometric coefficient (moles/L-h)
- B Biomass concentration (g/L)
- **b** Constraints (equation 28)
- b Stoichiometric coefficient (moles/L-h)

- **C** Constraint matrix (equation 28)
- c Stoichiometric coefficient (moles/L-h)
- d Stoichiometric coefficient (moles/L-h)
- E Emulsan concentration (g/L)
- e Stoichiometric coefficient (moles/L-h)
- F_A Airflow rate (converted to L/h from Henry's law constant)
- F_f Substrate feed rate (L/h)
- F_0 Sampling rate (L/h)
- f Stoichiometric coefficient (moles/L-h)
- g Stoichiometric coefficient (moles/L-h)
- h Stoichiometric coefficient (moles/L-h)
- $k_l a$ Oxygen-mass-transfer coefficient (1/h)
- n Biomass composition subscript
- $[O_2]_d$ Dissolved oxygen concentration (g/L)
- $[O_2^*]$ Effective concentration of oxygen in off-gas (g/L)
- $[O_2^*]_f$ Effective concentration of oxygen in feed (g/L)
- **P** State-error covariance matrix
- P Extracellular protein concentration (g/L)
- p Biomass composition subscript
- $\mathbf{Q}(k)$ State-noise covariance matrix
- q Biomass composition subscript
- $\mathbf{R}(k)$ Measurement-noise covariance matrix
- R_{CO_2} Rate of carbon dioxide evolution (g/L-h)
- R_{O_2} Rate of oxygen consumption (g/L-h)
- R_S Rate of substrate consumption (g/L-h)
- **r** Production rate vector
- r_b Degree of reduction of the biomass
- S Substrate concentration (g/L)
- S_f Substrate feed concentration (g/L)
- t Time (h)
- V Fermentor liquid volume (L)
- w State-noise vector
- **x** Flux vector (moles/L-h)
- Y_{ATP} ATP yield (grams biomass/mole ATP)
- Y_i Input variables (i = 1–4)
- Y_E Emulsan yield (grams emulsan/grams biomass)
- Y_{O_2} Oxygen yield (grams O_2/grams biomass)
- Y_P Extracellular protein yield (grams protein/grams biomass)
- Y_S Substrate yield (grams substrate/grams biomass)

Greek Variables

- η_i Measurement-noise vector
- μ Specific growth rate (1/h)
- $\sigma_{i,j}$ Element of $\mathbf{R}(k)$

Superscripts

m	Measured quantity
·	Derivative with respect to time
T	Transpose operator
−1	Inverse operator