OPTIMIZING ETHANOL PRODUCTION SELECTIVITY

RAMAN LALL¹, TIMOTHY J. DONOHUE¹,², and JULIE C. MITCHELL¹,³,⁴

*Corresponding author
1 Department of Mathematical Sciences, University of Delaware, Newark, DE 19716
2 Department of Bacteriology, University of Wisconsin, Madison WI 53706
3 Department of Mathematics, University of Wisconsin, Madison WI 53706
4 Department of Biochemistry, University of Wisconsin, Madison WI 53706

Abstract

Lactococcus Lactis metabolizes glucose homofermentatively to lactate. However, after disruption of the gene coding for lactate dehydrogenase, LDH, a key enzyme in NAD+ regeneration, the glycolytic flux shifts from homolactic to mixed acid fermentation with the redirection of pyruvate towards production of formate, acetate, ethanol and carbon dioxide. A mathematical model of the pyruvate metabolism pathway that enhances our understanding of ethanol production was developed from in vivo nuclear magnetic resonance (NMR) time series measurements that describes the dynamics of the metabolites in L. lactis. An S-system model based on the power law representation was able to capture the observed dynamics of the pyruvate metabolism pathway in L. lactis in vivo. The model provides insights into the maximization of selectivity of ethanol with respect to acetate and carbon dioxide as undesired products in multiple reactions. High concentrations of NADH and acetyl-CoA and low concentrations of pyruvate and NAD appear to maximize ethanol selectivity. These conclusions help to design strategies for improvement of ethanol products from sugars or other biomass. The results bear relevance in the educational field due to its usefulness in metabolic engineering applications.

Keywords  Pathway, Modeling, Ethanol, Selectivity, S-system, time-series

Introduction

Extracting biological information from time-series data requires a modeling framework able to capture the dynamics of the data. As the models are non-linear coupled systems of differential equations, fitting parameters to the models is complex and presents convergence issues as the systems grow in size. In spite of such issues, the analysis of in vivo time series data is very useful in gauging the response of in vivo systems that have not undergone artificial isolation and purification. In particular, such data accurately reflect the activity of cells and organisms and how they respond to signals and stimuli. Living organisms must coordinate biological machinery across several levels of organization, from gene expression to dynamic changes in protein abundance to adaptive changes in metabolic profiles and physiological response. These changes cannot always be deduced from in vitro measurements of rate constants and other kinetic parameters. For this reason, we develop models based on in vivo metabolomics data, in the present case for the glycolysis pathway that converts glucose to ethanol. A surge of interest in biofuels makes the study and optimization of this pathway very desirable. Our specific goal is to predict conditions under which the pathway will maximize the ratio of ethanol to CO₂, thus
maximizing the production of ethanol and possibly limiting the accumulation of CO₂ during the fermentation process.

*L. lactis* is a gram positive, non-spore forming, typically homofermentative bacterium in which the major end product of fermentation is lactate. During anaerobic growth on sugars such as maltose [1] or Galactose [2] or Mannitol [3] the glycolytic flux shifts towards the mixed-acid fermentation products formate, acetate and ethanol. Traditionally, the shift from homolactic to mixed acid fermentation has been captured by models of the enzymes lactate Dehydrogenase (LDH) and pyruvate formate-lyase (PFL), which compete for pyruvate under anaerobic conditions. Glycolytic intermediates, Glyceraldehyde-3-Phosphate (GAP) and Dihydroxyacetone-Phosphate (DHAP) are strong inhibitors of PFL, whereas the glycolytic intermediate Fructose 1,6 Diphosphate (FDP) is an activator of LDH [4]. When grown on less favorable sugars such as Galactose, the levels of FDP, GAP and DHAP are lower, resulting in lower PFL inhibition, LDH activation and mixed-acid product formation [5]. In order to optimize fluxes toward the desired end products, various theoretical frameworks such as biological system theory [6], metabolic control analysis [7] and metabolic design [8] have been used to analyze multi-enzyme systems in a quantitative, predictive fashion.

Time-dependent models have traditionally been represented by Michaelis-Menten equations, which have proven useful for characterizing isolated reaction mechanisms in vitro. Power law formulations [9, 10, 11] offer an alternative having a simplified structure that retains some of the non-linear features of the original system. In the S-system formulation of the Biochemical System Theory, as a consequence of flux aggregation, the resulting steady state equations are linear after a logarithmic transformation [12, 13].

Michaelis-Menten representations do not allow linearization of the steady state equation. As a result, there have been relatively few attempts to optimize metabolic systems based on these representations [14, 15, 16]. The present work examines the use of optimization, based on Michaelis-Menten and S-Systems models of in vivo time series data, to predict conditions under which the ratios of ethanol to acetate and ethanol to CO₂ production are maximized in the LDH⁻ strain of *L. lactis*.

**Methods:**

A decoupling procedure is used for the solution of this model of nonlinear ordinary differential equations [17]. The procedure uses the slopes as estimates of the true differentials on the left-hand side of the ODE model. As a result, the nonlinear ODE is reformulated from one involving \( n \) non-linear ODE’s to a larger system of \( n \times N \) algebraic equations where each set may be treated as an independent regression task in \( N \) equations. The results from the estimates of the decoupled equations serve as good starting values for a subsequent nonlinear optimization of the entire pathway. The optimized parameter values were obtained using a non-linear least squares optimization method (Levenberg-Marquardt) which is implemented in Matlab. Models from Biochemical System Theory [18, 19] have been used to approximate time-series data. These
power law models can be regarded as a generalization of linear regression analysis. In this framework, processes are represented as products of power law functions that are mathematically derived from Taylor's theorem applied to variables in logarithmic space. As a result, the model for each differential \( \dot{X}_i \) involves at most \( n \) dependent (state) variables and \( m \) independent variables, and it has the following form

\[
\dot{X}_i = \gamma_i \prod_{j=1}^{n+m} X_j^{f_{ij}}
\]

where \( \gamma_i \) is the rate constant that describes the turnover rate of the process, and the exponent \( f_{ij} \) is the kinetic order which quantifies the direct effect of variable \( X_j \) on \( \dot{X}_i \). An activating effect is indicated by a positive kinetic order whereas a negative kinetic order represents inhibition.

The S-system is a useful representation of BST and represents a collection of inward fluxes into a given pool with a single power-law term and the collection of outward effluxes from the pool with a second power law term. The generic S-system structure is:

\[
\dot{X}_i = \alpha_i \prod_{j=1}^{n} X_j^{\gamma_{ji}} - \beta_i \prod_{j=1}^{n} X_j^{\gamma_{ij}} \quad i=1,\ldots,n.
\]

where \( \alpha \) and \( \beta \) are non-negative rate constants and the exponents \( g \) and \( h \) are real valued kinetic orders.

Selectivity analysis has been used to maximize the yield of desired product relative to undesired products [20]. Selectivity is defined as the rate of production of the desired product to that of the undesired product in multiple reactions. For example, in the competing reactions

\[
A \rightarrow D
\]

\[
A \rightarrow U
\]

one might wish to maximize the formation of the desired product \( D \) and minimize the formation of the undesired product \( U \). The rate of formation of \( D \) and \( U \) are given from the rate laws as:

\[
r_D = k_D[A]^{\alpha_1}
\]

\[
r_U = k_U[A]^{\alpha_2}
\]

where \( \alpha_1 \) and \( \alpha_2 \) are positive kinetic orders. The ratio of these rates is the rate selectivity parameter \( S \) which needs to be maximized

\[
S_{DU} = \frac{r_D}{r_U} \left( \frac{k_D}{k_U} \right) [A]^{\alpha_1 - \alpha_2}
\]
For the case where the reaction order of the desired product is greater than the reaction order of the undesired product ($\alpha_1 > \alpha_2$), the selectivity is maximized by keeping the concentration $[A]$ as high as possible. It is seen from (5) that a higher concentration of A results in a high ratio of the desired to the undesired product. Conversely, when the reaction order of the undesired product exceeds that of the desired product ($\alpha_1 < \alpha_2$), the selectivity is maximized when $[A]$ is minimized.

2.1. Modeling the Glycolysis Pathway for Enhanced ethanol Production

$L. lactis$ is a gram positive non sporulating bacterium that is used widely in the dairy industry for the production of cheese, buttermilk and yogurt. Here, we develop a model for the LDH strain of $L. lactis$. The deletion of the LDH gene shifts the glycolytic flux towards ethanol production. The simplified pathway underlying our model analysis is illustrated in Fig. 1. A previous kinetic model was developed for the conversion of glucose to pyruvate [21]. The focus of our model is on the flux through pyruvate formate-lyase as this is the step that converts pyruvate to acetyl-CoA, which is the precursor for ethanol. For the purpose of this model, data are taken from published time-series data [22]. In this experimental study, Nuclear Magnetic Resonance (NMR) measurements were used to monitor the pool of labeled intermediates and end products with a time resolution of 2.2 minutes in non-growing $L. lactis$ cell suspensions that followed a pulse of 13C-labeled glucose. In vivo NMR experiments were performed using 50 mL mini-bioreactors that were designed to maintain cell suspensions in defined conditions of gas atmosphere, pH and temperature. [1-13C] glucose (40 mM) was supplied to cell suspension and time courses for substrate consumption, product formation and dynamics of intracellular metabolites were monitored in vivo under anaerobic conditions. The end products (lactate, pyruvate, ethanol and acetate) were quantified in the NMR-sample extract by H-NMR. Further details of these experimental procedures can be found in Neves et. al [22, 23, 24].

3. Results

We present ethanol production selectivity results with respect to two models, one based on the Michaelis-Menten equations and the other based on a simplified S-System representation. Both models are able to capture the dynamics of the time-series data, and they lead to similar conclusions regarding the determinants of ethanol versus CO$_2$ selectivity.

Biochemical Systems Theory Model for in vivo Glycolysis

The BST equations for the pyruvate metabolism pathway are readily set up in symbolic form, according to well–documented guidelines [25]. The dynamics of pyruvate is given symbolically as:

$$\dot{X}_2 = \alpha_2 [X_1]^{\beta_{21}} - \beta_2 [X_2]^{\beta_{22}} [X_4]^{\beta_{24}} [X_5]^{\beta_{25}}$$  \[6\]
The first term captures the production flux from Phosphoenolpyruvate (PEP) and the sthe degradation term is driven by the concentration of pyruvate, NAD+ and NADH. The formation of the end product CO₂ is driven by the concentration of pyruvate and NAD⁺, that of ethanol by the concentration of NADH and Acetyl-CoA and that of acetate by the concentration of Acetyl-CoA. The dynamics of the production of carbon dioxide, ethanol and acetate are given by:

\[
\dot{X}_3 = \alpha_3 [X_2]^{g_{32}} [X_4]^{g_{34}} \\
\dot{X}_7 = \alpha_7 [X_5]^{g_{75}} [X_6]^{g_{76}} \\
\dot{X}_8 = \alpha_8 [X_6]^{g_{86}}
\]

where the following variables denote the metabolite concentrations as a function of time.

\[
X_1 = \text{PEP concentration (mM)} \quad X_5 = \text{NADH concentration (mM)} \\
X_2 = \text{Pyruvate concentration (mM)} \quad X_6 = \text{Acetyl-CoA concentration (mM)} \\
X_3 = \text{CO₂ concentration (mM)} \quad X_7 = \text{Ethanol concentration (mM)} \\
X_4 = \text{NAD⁺ concentration (mM)} \quad X_8 = \text{Acetate concentration (mM)}
\]

In the model for pyruvate metabolism pathway that was proposed for ethanol production, the metabolites pyruvate, ethanol, acetate and CO₂ were fitted to the in vivo time-series data. The appropriateness of the functional form of the power law was tested by fitting the equation for each metabolite separately. The rationale was that if a good fit could be obtained for each individual equation (for \(X_i\)), the functional form of the equation is adequate whereas if the model was unable to capture the data, the lack of fit could be attributed to the form of the S-system equation. The metabolite concentrations were approximated very well using this procedure, and the estimates from these individual fits served as good starting values for a subsequent nonlinear optimization of the entire model. The global fits the non-linear regression (using the Levenberg-Marquard algorithm in Matlab) found a good solution that was able to capture the global dynamics. The results of these global fits are shown in Fig. 2, and the parameters we obtained are given in Appendix B.

For the purpose of enhancing ethanol production, conditions were analyzed for maximizing ethanol selectivity with respect to the undesired products acetate and carbon dioxide. Ethanol selectivity was determined from the ratio of the rate of production of ethanol with respect to the rate of production of acetate or carbon dioxide. Using (7) for rate of carbon dioxide production and (8) for rate of ethanol production and (9) for the rate of acetate production, the selectivities for ethanol production versus CO₂ and acetate can be expressed as:

\[
S_{\text{ETH/CO₂}} = \frac{\alpha_7 / \alpha_3 [X_6]^{g_{76}} [X_5]^{g_{75}} /[X_2]^{g_{12}} [X_4]^{g_{14}}}{}
\]
The kinetic parameters $g_{76}$, $g_{86}$, $g_{75}$, $g_{32}$ and $g_{34}$ were obtained from the model fits (see Appendix A for parameter values). Since $g_{76} > g_{86} > 0$ and $g_{75} > 0$ it is Eq. 11 predicts that an increase in the concentration of acetyl-CoA and/or NADH increases ethanol selectivity. As $g_{76} > 0$, $g_{75} > 0$, $g_{32} > 0$ and $g_{34} > 0$, it is predicted from Eq. 10 that an increase in the concentration of acetyl-CoA and/or NAD+ increases ethanol selectivity as does a decrease in concentration of pyruvate and/or NAD+. These trends are demonstrated graphically in Figures 3-4.

Discussion:

Previous kinetic models for pyruvate distribution in Lactococcus lactis have been based on Michaelis-Menten kinetics for in vitro kinetic data (Hoefnagel et. al. 2002, Voit et. al. 1987). The present work uses both Michaelis-Menten and the the S-system representation of the Biochemical System Theory (BST) in conjunction with in vivo time-series data. While size and parameter values of the S-system depend entirely on the complexity of the investigated phenomenon, the structure of the model is the same. This fact has led to a powerful repertoire of algebraic and numerical methods for the analysis of S-system models and to interactive software that executes the methods with high efficiency. The canonical structure of the S-system is more tractable than the Michaelis-Menten dynamics, where a priori, it is unclear what exact type and rate law would best characterize a particular reaction.

The estimation of parameters is typically solved through non-linear regression and is hampered by technical obstacles. Nonlinear regressions are far more complex than linear regression, where a fast and unique solution is usually guaranteed. For example, it is known that gradient-based non-linear searches have a tendency to get stuck in local minima and may not converge to an optimal solution within the parameter space. Objective functions based on systems of differential equations require many numerical integrations that can use more than 95% of the entire search time. Search algorithms often select parameter combinations artificially, making the system stiff and thereby dramatically slowing down the solution [25].

The present work uses the concept of maximization of the selectivity of desired products to that of undesired products in multiple reactions. Our models suggest that high concentrations of acetyl-CoA and NADH and low concentrations of NAD and pyruvate maximize ethanol selectivity. These results suggest specific reactor schemes to selectively produce ethanol and minimize CO2 release. In order to keep the concentration of a specific metabolite as high as possible, batch or plug-flow reactors are recommended. In such reactors, the concentration of a metabolite starts at a high value and drops progressively during the course of the reaction. On the other hand, in order to keep the concentration of a specific metabolite as low as possible, a continuously stirred tank reactor (CSTR) is recommended. This maintains the concentration of
reactant at a low level. A recycle stream, in which the product stream acts as a diluent, can also be used to maintain the concentrations of the metabolite at low levels [20].

In our specific case, we wish to keep the concentration of certain metabolites high and others low, so a semibatch bioreactor is advised. In this reactor, scheme metabolites requiring low concentration are fed slowly into a large amount of the metabolite with the higher concentration. Other bioreactor designs include a tubular reactor with side streams of the metabolite with the low concentration continuously fed to the bioreactor, along with a series of small CSTR’s with the metabolite with the high concentration fed only to the first bioreactor and metabolite with the lower concentration fed to each bioreactor.

As the metabolites pyruvate, acetyl-CoA, NADH and NAD$^+$ are intracellular metabolites, it is not feasible to introduce them extracellularly. This suggests the need to study how glucose concentrations can be optimized to affect concentrations of pyruvate and acetyl-CoA. The experimental data available for this work was at an initial concentration of 40mM glucose under anaerobic conditions, and further experiments at varying initial concentrations of glucose may offer added insights on optimizing ethanol production selectivity. As a result, there is a need to investigate how glucose concentrations can be optimized to affect concentrations of these intracellular metabolites.

5. ACKNOWLEDGMENTS

This work was supported by the BACTER Institute through a grant from the Department of Energy as part of the Genomics:GTL program (DE-FG02-04ER25627). The authors also thank Daniel Noguera and Laura Vanderploeg of the University of Wisconsin for support during the project.

6. REFERENCES


Bioeng 64 (2): 200-212 (1999)


.VARIABLE DEFINITIONS

X1 = PEP concentration (mM) 
X9 = ACAL concentration (mM) 
X2 = pyruvate concentration (mM) 
X10 = AC concentration (mM) 
X3 = CO2 concentration (mM) 
X11 = ACP concentration (mM) 
X4 = NAD+ concentration (mM) 
X12 = ADP concentration (mM) 
X5 = NADH concentration (mM) 
X13 = ATP concentration (mM) 
X6 = acetyl-CoA concentration (mM) 
X14 = CoA concentration (mM) 
X7 = ethanol concentration (mM) 
X16 = FBP concentration (mM) 
X8 = acetate concentration (mM)

B. S-SYSTEM MODEL

Parameters of the S-system used to model the measured time courses of pyruvate metabolism in L. lactis (LDH□) were estimated by numerical integration using the ODE solver in MATLAB. The numerical implementation was used to estimate observed data in Fig. 2.

\[
\dot{X}_2 = \alpha_2 [X_1]^{g_{21}} - \beta_2 [X_2]^{g_{22}} [X_4]^{g_{24}} [X_5]^{g_{25}} \\
\dot{X}_3 = \alpha_3 [X_2]^{g_{32}} [X_4]^{g_{34}} \\
\dot{X}_7 = \alpha_7 [X_5]^{g_{76}} [X_6]^{g_{75}} \\
\dot{X}_8 = \alpha_8 [X_6]^{g_{86}} \\
\]

\[\alpha_2 = 0.07357218341808\]
\[g_{2,1} = 0.03171344758464\]
\( \beta_2 = 0.02295975256347 \)
\( h_{2,2} = 0.43556107753767 \)
\( h_{2,4} = 1.12329782487184 \)
\( h_{2,5} = 0.00221701611913 \)

\( \alpha_3 = 0.56213505169305 \)
\( g_{3,2} = 0.75322391540504 \)
\( g_{3,4} = 1.16263523354082 \)

\( \alpha_7 = 0.01248858503143 \)
\( g_{7,5} = 0.00009141505280 \)
\( g_{7,6} = 1.1126202399438 \)

\( \alpha_8 = 1.66300224304295 \)
\( g_{8,6} = 0.0000000000004 \)

FIG. 1. The glycolysis pathway in L. lactis. The shaded region of the pathway describes the conversion of Phosphoenolpyruvate to ethanol.
FIG. 2: The images show the degree of fit between the experimental data and the biochemical systems theory model. Data and fitted curves are shown for the concentrations of pyruvate, ethanol, acetate and CO₂.

FIG. 3. The figure shows the selectivity analysis with ethanol as the desired product and acetate as the undesired product. The ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH.
FIG. 4. The figure shows the selectivity analysis with ethanol as the desired product and CO$_2$ as the undesired product. The general trends indicate that ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH and decreases as a function of pyruvate or NAD$^+$. 