ARTICLE IN PRESS

Mathematical and Computer Modelling ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Mathematical and Computer Modelling

journal homepage: www.elsevier.com/locate/mcm

Optimizing ethanol production selectivity

Raman Lall [a], Timothy J. Donohue [a,b], Simeone Marino [c], Julie C. Mitchell [a,d,*]

[a] BACTER Institute, University of Wisconsin, Madison, WI 53706, United States

[b] Department of Bacteriology and Great Lakes Bioenergy Center, University of Wisconsin, Madison, WI 53706, United States

[c] Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, United States

[d] Departments of Mathematics and Biochemistry, University of Wisconsin, Madison, WI 53706, United States

ARTICLE INFO

Article history:
Received 8 October 2009
Received in revised form 7 January 2010
Accepted 21 January 2010

Keywords:
Glycolysis pathway
Ethanol production
Selectivity maximization
Michaelis–Menten
S-systems

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 CO_2. A mathematical model of the pyruvate metabolism pathway that enhances ethanol production was developed from *in vivo* Nuclear Magnetic Resonance (NMR) time-series measurements that describe the dynamics of the metabolites in *L. lactis*. Both Michaelis–Menten and S-system models capture the observed *in vivo* dynamics of the glycolysis pathway in *L. lactis*, while prior models describe only the *in vitro* dynamics. The models provide insight into the maximization of selectivity of ethanol with respect to acetate and CO_2 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.

1. 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 CO_2, thus maximizing the production of ethanol and possibly limiting the accumulation of CO_2 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

* Corresponding author at: Departments of Mathematics and Biochemistry, University of Wisconsin, Madison, WI 53706, United States. Tel.: +1 011 608 432 5853.
E-mail address: jcmitchell@wisc.edu (J.C. Mitchell).

0895-7177/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcm.2010.01.016

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, Gylceraldehyde-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 [9–11] 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 [12–14]. Using this specific property, Torres et al. applied a linear optimization procedure to the production of ethanol by *Saccharomyces cervisiae* and suggested alternative enzyme profiles that predict higher production rates [13,15].

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 [16–18]. A direct optimization approach based on a multi-start stochastic algorithm was previously applied to the optimization of the Michaelis–Menten and Generalized Mass Action representations of ethanol, glycerol and carbohydrate production by *S. cervisiae* [19]. It was noted that when the goal was to maximize ethanol production, six enzymes involved in the ethanol synthesis needed to be modulated. In contrast, when the aim was to maximize the glycerol and carbohydrate production rate, the modification of only one enzyme that diverts flux from the main downstream to the final product leads to a significant increase in final metabolite production [15]. 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 CO_2 production are maximized in the LDH^- strain of *L. lactis*.

2. Methods

Parameter estimation and curve fitting are traditionally solved by minimizing a generalized distance between experimental data and model predictions. The Euclidean distance is the most commonly used, and the inverse problem is often referred to as a least squares (LS) formulation, or more generally, a non-linear least squares (NLLS) formulation. Many non-linear programming algorithms can numerically solve NLLS problems: Gauss–Newton and Levenberg–Marquardt are the most popular and frequently implemented [20–22]. Evolutionary algorithms and simulated annealing techniques are also used [23]. We use the approach described in [24] to fit the simultaneous ODE system. We briefly describe this procedure in Section 2.2.

2.1. Biochemical model

Models from Biochemical System Theory [25,26] 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

$$\dot{X}_i = \gamma_i \prod_{j=1}^{n+m} X_j^{f_{i,j}}, \tag{1}$$

where γ_i is the rate constant that describes the turnover rate of the process, and the exponent f_{ij} is a kinetic order that quantifies the 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 fluxes from the pool with a second power law term. The generic *S*-system structure is

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n} X_j^{h_{i,j}}, \quad i = 1, \ldots, n \tag{2}$$

where α and β 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 [27]. 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

$$A \longrightarrow D$$
$$A \longrightarrow U$$

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

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

$$r_D = k_D[A]^{\alpha_1} \tag{3}$$

$$r_U = k_U[A]^{\alpha_2} \tag{4}$$

where α_1 and α_2 are positive kinetic orders. The ratio of these rates is the selectivity parameter

$$S_{DU} = \frac{r_D}{r_U} = \frac{k_D}{k_U} \cdot [A]^{\alpha_1 - \alpha_2}. \tag{5}$$

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.

2.2. Computational fitting of parameters

In Section 3.1 and Appendix C, we give the functional forms of the Biochemical Systems Theory and Michaelis–Menten models that were fitted using the procedure outlined in this section. In principle, a system of ODEs like (2) could be fitted directly to data, but many computational issues arise. The dynamical system can become ill-conditioned or stiff due to non-optimal choices of parameter values during the minimization procedure. As a consequence, the computation time for numerically integrating the differential equations may increase to 95% or more of the total CPU time needed by the optimization algorithm [28]. Secondly, choosing a feasible and structurally stable initial guess for the parameter vector is sometimes is as difficult as solving the problem itself. Finally, experimental noise increases the likelihood of the existence of multiple local minima, which is amplified by an increasing size of the system. A possible strategy for reducing the complexity of the problem is to decouple the ODE system.

Decoupling may be accomplished by estimating slopes of all (n) time courses at many (N) time points, which reduces the system of n coupled differential equations to $n \times N$ algebraic equations [28]. Another means of decoupling the equations without disrupting the system of n coupled differential equations is by retaining the initial differential form and solving/fitting one equation at a time, using the data (stepwise linear interpolations) as input functions for the equation [29,24]. In this work, we use the decoupling approach and non-linear least squares (NLLS) fitting procedures described in [24] to fit parameters to the S-system and of the Michaelis–Menten models based on the time-series data.

The NLLS algorithm involves two steps, as described in detail in [24]. In Step 1 we fit each differential equation of each model independently (uni-dimensional ODE). Then, in Step 2, we plug the solutions from Step 1 into the simultaneous systems of ODEs, and fit the whole system of coupled differential equations. The idea is that Step 1 will refine the initial condition for the fitting of the larger, more complex and coupled non-linear models, facilitating convergence of the Levenberg–Marquardt or Gauss–Newton search direction scheme. Also, Step 1 avoids ill-conditioning and stiffness issues that might occur during the numerical solution step. In both Step 1 and Step 2, current values of the parameters are used to numerically solve the ODE system at each iteration of the minimization algorithm (NLLS). Optimality conditions are tested after each iteration, and, if not satisfied, a new iteration is performed using the direction of search is defined by the minimization algorithm (for example, Levenberg–Marquardt or Gauss–Newton). See [24] for a detailed description of the algorithm.

The model fitting module is implemented in MATLAB. We use the function *lsqcurvefit* to solve general non-linear least squares problems (in the present case, using the Levenberg–Marquardt algorithm) and *ode45* (an explicit, variable-step, Runge–Kutta method of fourth order) to numerically solve the differential equations for each of the metabolites ethanol, pyruvate, acetate and CO_2. Our code for applying this method to the system described in this paper is available by request to the authors.

2.3. 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 [30]. 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. 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 min 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. [1–13 °C] 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*

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

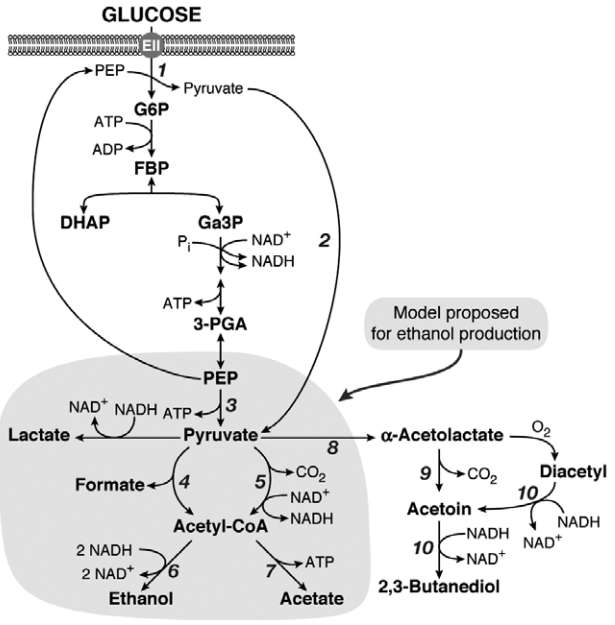

Fig. 1. The glycolyis pathway in *L. lactis*. The shaded region of the pathway describes the conversion of Phosphoenolpyruvate to ethanol.

under anaerobic conditions. The end products (lactate, pyruvate, ethanol and acetate) were quantified in the NMR-sample extract by H-NMR. Details of these experimental procedures can be found in Neves et al. [31–33].

Using the procedure described in the previous sections, parameters were derived to model the time course data for pyruvate, CO_2, ethanol, and acetate. Parameters for other metabolites, such as NADH and NAD^+, were not calculated, but the time-series data for all metabolites was used during the fitting procedure. Finally, because a time course of CO_2 concentrations had not been experimentally determined, it was calculated based on the other concentrations. In particular, the amount of CO_2 released can be calculated based on the concentrations of pyruvate, NAD^+ and Acetyl-CoA.

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.

3.1. 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

$$\dot{X_2} = \alpha_2 X_1^{g_{2,1}} - \beta_2 X_2^{h_{2,2}} X_4^{h_{2,4}} X_5^{h_{2,5}}. \tag{6}$$

The first term captures the production flux from Phosphyenolpyruvate (PEP), and the degradation term is driven by the concentration of pyruvate, NAD^+ and NADH. The formation of the end product CO_2 is driven by the concentration of pyruvate and NAD^+, that of ethanol is governed by the concentration of NADH and acetyl-CoA, and that of acetate by the concentration of acetyl-CoA. The dynamics of the production of CO_2, ethanol and acetate are given respectively by

$$\dot{X_3} = \alpha_3 X_2^{g_{3,2}} X_4^{g_{3,4}} \tag{7}$$

$$\dot{X_7} = \alpha_7 X_5^{g_{7,5}} X_6^{g_{7,6}} \tag{8}$$

$$\dot{X_8} = \alpha_8 X_6^{g_{8,6}} \tag{9}$$

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

X_1 = PEP concentration (mM)
X_2 = pyruvate concentration (mM)
X_3 = CO_2 concentration (mM)
X_4 = NAD^+ concentration (mM)
X_5 = NADH concentration (mM)
X_6 = acetyl-CoA concentration (mM)
X_7 = ethanol concentration (mM)
X_8 = acetate concentration (mM)

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

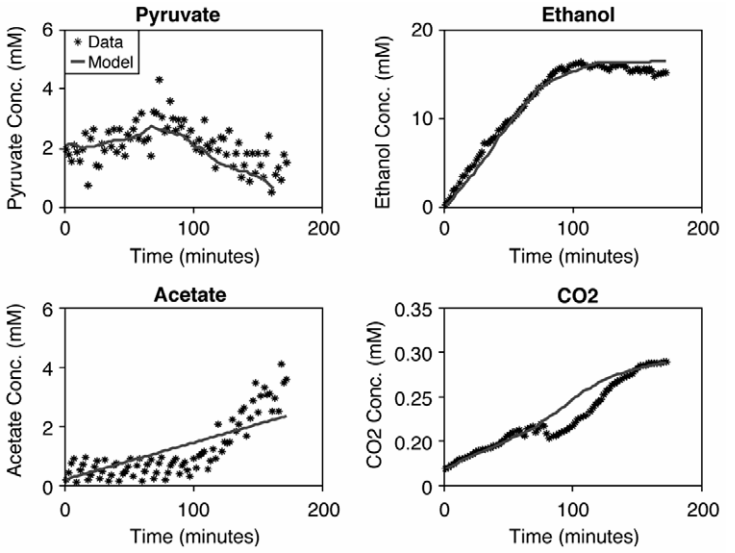

Fig. 2. The images shows the degree of fit between the experimental data and the biochemical systems theory model. Data and fitted curves are shown for the concentrations of pyruvate, ethanol, acetate, and CO_2.

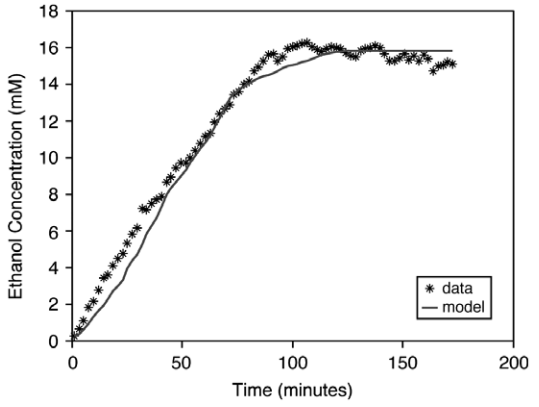

Fig. 3. The figure shows the degree of fit between the Michaelis–Menten model to the experimental data, where the parameters are as in the rightmost column of Table 1.

In the model for the glycolysis pathway that is proposed for ethanol production, the metabolites pyruvate, ethanol, acetate and CO_2 were fitted to the *in vivo* time-series data. The appropriateness of the functional form of the power law was first 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 $\dot{X}_i$), the functional form of the equation would be 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 non-linear optimization of the entire model. The global fits the non-linear regression (using the Levenberg–Marquardt 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.

3.2. Michaelis–Menten model for in vivo glycolysis

Good fits were obtained for describing *in vivo* data using a Michaelis–Menten model previously proposed for *in vitro* data for glycolysis in *L. lactis* [5,34]. In this case, not all reactions shown in Fig. 1 were fit to parameters, only those needed to perform the selectivity analysis described in the next section. The fitting procedure for the Michaelis–Menten model is similar to that described for the *S*-Systems representation, and the fit of the model to the data was comparably good. The fit for the ethanol concentration is shown in Fig. 3.

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

Table 1
In vitro and estimated *in vivo* parameters for the Michaelis–Menten model.

Param	*in vitro*	*in vivo* est.
$V_{\max_{6b}}$	3.690	0.071
$K_{eq_{6b}}$	5.675	1.403
$K_{m_{4,6b}}$	0.194	0.266
$K_{m_{5,6b}}$	1.404	0.037
$K_{m_{7,6b}}$	1.785	5.675
$K_{m_{9,6b}}$	0.266	3.370

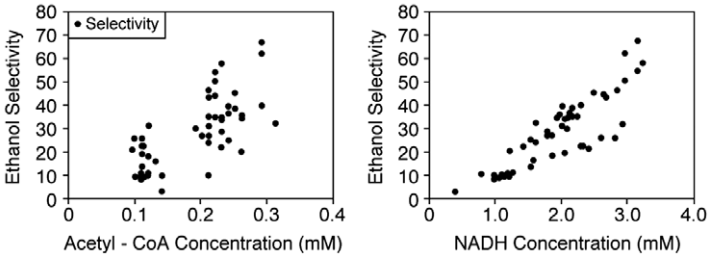

Fig. 4. The figure shows the selectivity analysis with ethanol as the desired product and acetate as the undesired product. The ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH.

Appendix C gives the functional form of the equations used in the Michaelis–Menten model. These rate equations were appropriate to fit the observed data. The optimized parameter values were in some cases quite different from those obtained from literature, but this is not surprising as the published values are based on *in vitro* enzyme assays. Table 1 compares the corresponding parameter values with those in the literature. The *in vitro* parameter values were taken from the JWS Online Model Database [35], which provide an updated version of the model presented in [5]. All but one of the K_m values was higher for the *in vivo* model, which is sensible given that reactions are apt to happen more slowly in a real biological environment as compared with a well-mixed solution. The K_m for NAD^+ was fairly comparable between the *in vitro* calculations and the *in vivo* model. For the hexokinase reaction within the glycolysis pathway in *Lactococcus lactis*, a comparisons of the K_m's for the *in vivo* and *in vitro* models is given in [36].

3.3. *Optimizing* ethanol/CO_2 *selectivity*

We can predict conditions for which ethanol production selectivity is maximized with respect to the undesired products acetate and CO_2. Ethanol selectivity was predicted from the ratio of the rate of production of ethanol with respect to the rate of production of acetate or CO_2. Using (7) for rate of CO_2 production, (8) for rate of ethanol production, and (9) for rate of acetate production, the selectivities for ethanol production versus CO_2 and acetate can be expressed respectively as

$$S_{\mathrm{ETH/ACT}} = \frac{\alpha_7 X_6^{g_{7,6}} X_5^{g_{7,5}}}{\alpha_8 X_6^{g_{8,6}}} \tag{10}$$

$$S_{\mathrm{ETH/CO_2}} = \frac{\alpha_7 X_6^{g_{7,6}} X_5^{g_{7,5}}}{\alpha_3 X_2^{g_{3,2}} X_4^{g_{3,4}}} \tag{11}$$

where the kinetic parameters $g_{7,6}, g_{8,6}, g_{7,5}, g_{3,2}$ and $g_{3,4}$ are those obtained from the model fits (see Appendix B for parameter values). Since $g_{7,6} > g_{8,6} > 0$ and $g_{7,5} > 0$ the Eq. (10) predicts that an increase in the concentration of acetyl-CoA and/or NADH increases ethanol selectivity. As $g_{7,6} > 0$, $g_{7,5} > 0$, $g_{3,2} > 0$ and $g_{3,4} > 0$, it is predicted from (11) that an increase in the concentration of acetyl-CoA and/or NADH increases ethanol selectivity, as does a decrease in concentration of pyruvate and/or NAD^+.

These trends are demonstrated graphically in Figs. 4 and 5, using concentration data for Acetyl-CoA, NADH, pyruvate and NAD^+ obtained from the experimental time course [33] along with the parameters from our fitted models. Based on the values of the kinetic parameters $g_{7,5}$, $g_{7,6}$, $g_{8,6}$, α_7 and α_8, ethanol production selectivity relative to Acetate was calculated using (10) and plotted against concentrations of Acetyl-CoA and NADH in Fig. 4. From this, it is predicted that an increase in the concentration of Acetyl-CoA or NADH will increase ethanol selectivity, which is consistent with the observations made above. Similarly, the kinetic parameters $g_{7,6}$, $g_{7,5}$, $g_{3,2}$, $g_{3,4}$, α_3 and α_7 along with (11) were used to create the data in Fig. 5. Here, increases in the concentrations of Acetyl-CoA and NADH or decreases in the concentrations of Pyruvate and NAD^+ are predicted to increase ethanol selectivity relative to production of CO_2.

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

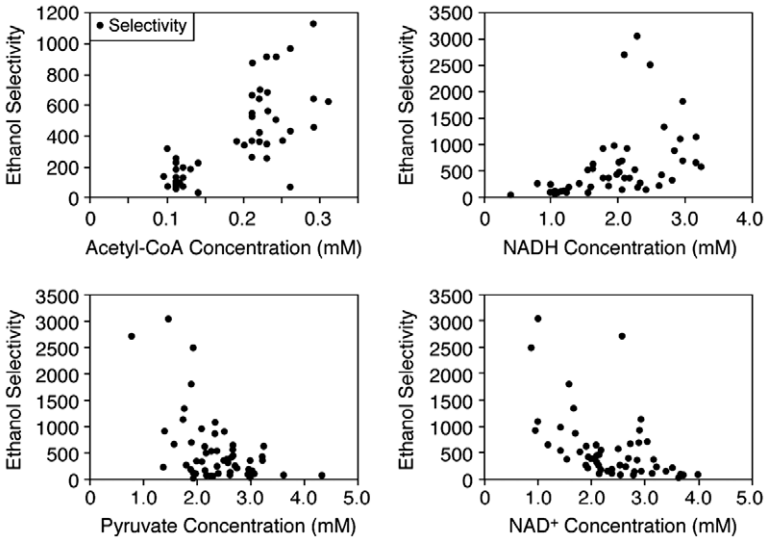

Fig. 5. The figure shows the selectivity analysis with ethanol as the desired product and CO_2 as the undesired product. The general trends indicate that ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH and decreases as a function of pyruvate or NAD^+.

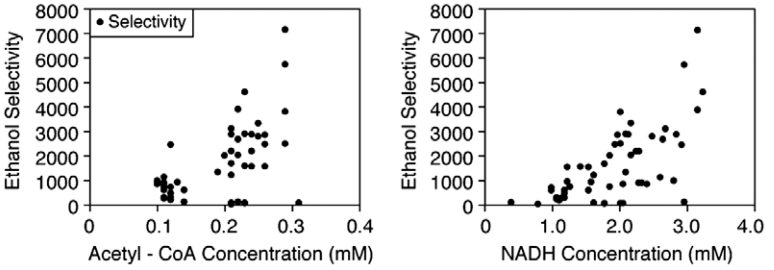

Fig. 6. The figure shows the selectivity analysis with ethanol as the desired product and acetate as the undesired product, calculated using the Michaelis–Menten model. The ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH.

The ethanol selectivity trends indicated by the Michaelis–Menten model are consistent with those derived from the BST system, as can be seen in Figs. 6 and 7. For the Michaelis–Menten model, ethanol selectivity was obtained from the ratios of the fluxes of producing ethanol to acetate (v_{6b}/v_7 in Reactions 6b and 7 in Appendix C) and plotted against concentrations of Acetyl-CoA and NADH in Fig. 6. As in Fig. 4, an increase in the concentration of Acetyl-CoA and NADH is predicted to increase ethanol selectivity relative to Acetate. In Fig. 7, ethanol selectivity relative to CO_2 was obtained using the ratio of the fluxes for producing ethanol and CO_2 (v_{6b}/v_5 in Reactions 5 and 6b from Appendix C) and plotted against concentrations of Acetyl-CoA, NADH, Pyruvate and NAD^+. The predictions are in agreement with those of Fig. 5, indicating that increases in the concentrations of Acetyl-CoA and NADH or decreases in the concentrations of Pyruvate and NAD^+ increase ethanol selectivity. The consistency between predictions made using Michaelis–Menten and Biochemical Systems Theory models is also seen in previous studies of the glycolysis pathway [19].

4. Discussion

Previous models for pyruvate distribution in *Lactococcus lactis* have been based on Michaelis–Menten kinetics for *in vitro* kinetic data [5,11]. The present work uses both Michaelis–Menten and 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.

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

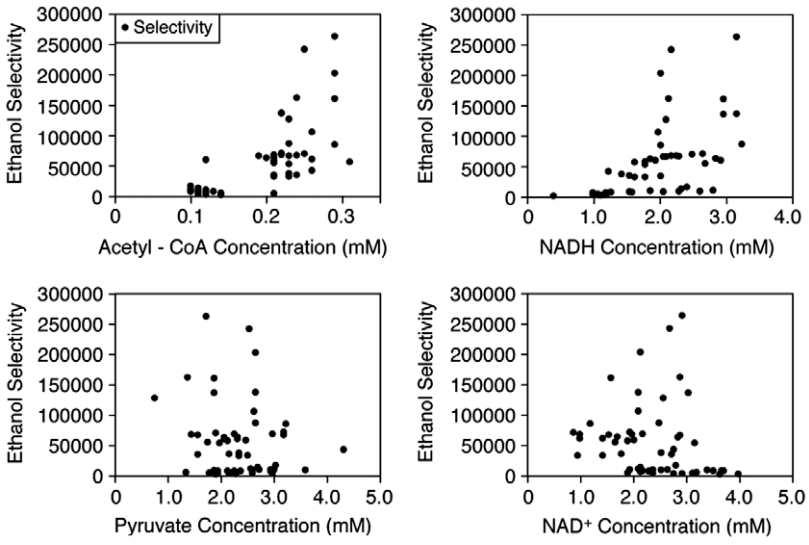

Fig. 7. The figure shows the selectivity analysis with ethanol as the desired product and CO_2 as the undesired product, calculated using the Michaelis–Menten model. The general trends indicate that ethanol selectivity increases as a function of the concentration of either acetyl-CoA or NADH and decreases as a function of pyruvate or NAD^+.

The estimation of parameters is typically solved through non-linear regression and is hampered by technical obstacles. Non-linear 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 [28].

Non-linear metabolic models for ethanol production were reduced so that methods of linear programming could be used [13]. These optimizations lead to profiles of enzyme activities that yield higher rates of the desired end products than original systems. Optimization was conducted with the *S*-system as the optimization problem was linear which was due to the fact that the steady-state equations are linear when represented in logarithmic coordinates. However, it was noted that in the case of linear optimization, the aggregation of fluxes at branch points is an indispensable feature of *S*-systems and leads to inaccuracies in the stoichiometry of reactions and subsequently in the optimum solution of the linear problem. On the other hand, the solution obtained with the non-linear optimization would depend on initial values and settings (mutation rate, iterations, etc.). The choice of using the linear method over the non-linear method depended on the topology of the pathway to be modeled. It was observed that in several biochemical models the occurrence of kinetic features such as bimolecularity associated with coupled cofactors, feedbacks, non-linear kinetics of enzymes constitute a limitation for some optimization procedures [7] and in most cases force us to use non-linear optimization.

The present work uses the concept of maximization of the selectivity of desired products to that of undesired products in multiple reactions [27]. Our models suggest that high concentrations of acetyl-CoA and NADH and low concentrations of NAD and pyruvate maximize ethanol selectivity. 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 these metabolites. The experimental data available for this work was at an initial concentration of 40 mM glucose under anaerobic conditions, and further experiments at varying initial concentrations of glucose may offer added insights on optimizing ethanol production selectivity. In particular, while the conclusions for selectivity relative to NADH and NAD^+ concentrations are obvious from the metabolites required in Reactions 5 and 6a–b, the prediction that low pyruvate concentrations enhance ethanol production selectivity is suggestive of the need for additional experiments. In particular, these results lead to a conjecture that fermentation in low glucose conditions, resulting in low pyruvate concentrations, will enhance ethanol production selectivity in a way that minimizes the production rate of CO_2 relative to the production rate of ethanol.

Acknowledgements

This work was supported by the BACTER Institute through a grant from the US 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.

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

Appendix A. Variable definitions

X_1 = PEP concentration (mM)
X_2 = pyruvate concentration (mM)
X_3 = CO_2 concentration (mM)
X_4 = NAD^+concentration (mM)
X_5 = NADH concentration (mM)
X_6 = acetyl-CoA concentration (mM)
X_7 = ethanol concentration (mM)
X_8 = acetate concentration (mM)
X_9 = ACAL concentration (mM)
X_{10} = AC concentration (mM)
X_{11} = ACP concentration (mM)
X_{12} = ADP concentration (mM)
X_{13} = ATP concentration (mM)
X_{14} = CoA concentration (mM)
X_{16} = FBP concentration (mM)

Appendix 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.

$$\dot{X_2} = \alpha_2 X_1^{g_{2,1}} - \beta_2 X_2^{h_{2,2}} X_4^{h_{2,4}} X_5^{h_{2,5}}$$
$$\dot{X_3} = \alpha_3 X_2^{g_{3,2}} X_4^{g_{3,4}}$$
$$\dot{X_7} = \alpha_7 X_5^{g_{7,5}} X_6^{g_{7,6}}$$
$$\dot{X_8} = \alpha_8 X_6^{g_{8,6}}$$

$\alpha_2 = 0.07357218341808$
$g_{2,1} = 0.03171344758464$
$\beta_2 = 0.02295975256347$
$h_{2,2} = 0.43556107753767$
$h_{2,4} = 1.12329782487184$
$h_{2,5} = 0.00221701611913$
$\alpha_3 = 0.56213505169305$
$g_{3,2} = 0.75322391540504$
$g_{3,4} = 1.16263523354082$
$\alpha_7 = 0.01248858503143$
$g_{7,5} = 0.00009141505280$
$g_{7,6} = 1.11262023999438$
$\alpha_8 = 1.66300224304295$
$g_{8,6} = 0.00000000000004.$

Appendix C. Michaelis–Menten model

The following equations were used to describe the rates of various reactions in the glycolysis pathway of *L. lactis*, using the numbering shown in Fig. 1. In the description of the reactions, number assigned to metabolite concentrations in Appendix A is shown as a subscript. Each K_m value is given two subscripts, the first of which denotes the metabolite number and the second the equation number. The values for $V_{\max}$, K_{eq}, K_i are given subscripts denoting the equation number. Similarly, the reaction rates, v_j, have subscripts indicating the reaction number.

Reaction 3

$$PEP_1 + ADP_{12} \longleftrightarrow PYR_2 + ATP_{13} + FBP_{15}$$

$$v_3 = V_{\max_3} \cdot \frac{\frac{X_1 \cdot X_{12}}{K_{m_{1,3}} \cdot K_{m_{12,3}}} - \frac{V_{\max_3} \cdot X_2 \cdot X_{13} \cdot X_{15}}{\left(K_{eq_3} \cdot K_{m_{1,3}} \cdot K_{m_{12,3}}\right)}}{\left(1 + \frac{X_1}{K_{m_{1,3}}} + \frac{X_2}{K_{m_{2,3}}}\right) \cdot \left(1 + \frac{X_{12}}{K_{m_{12,3}}} + \frac{X_{13}}{K_{m_{13,3}}}\right) \cdot \left(1 + \frac{X_{15}}{K_{m_{15,3}}}\right)}$$

$V_{\max_3} = 1.0355$
$K_{m_{1,3}} = 0.2671$
$K_{m_{2,3}} = 1.1921$
$K_{m_{12,3}} = 0.3487$
$K_{m_{13,3}} = 2.1467$
$K_{m_{15,3}} = 0.1089.$

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

Reaction 5

$$PYR_2 + NAD_4^+ + COA_{14} \longleftrightarrow NADH_5 + ACCOA_6$$

$$v_5 = V_{\max_5} \cdot \frac{\frac{X_2}{Km_{2,5}} \cdot \frac{X_4}{Km_{4,5}} \cdot \frac{X_{14}}{Km_{14,5}}}{\left(1 + \frac{X_2}{Km_{2,5}}\right) \cdot \left(1 + \frac{X_4}{Km_{4,5}} + \frac{X_5}{Km_{5,5}}\right) \cdot \left(1 + \frac{X_{14}}{Km_{14,5}} + \frac{X_6}{Km_{6,5}}\right) \cdot \left(\frac{1}{1 + Ki_5 \cdot \frac{X_5}{X_4}}\right)}$$

$V_{\max_5} = 1.6255$
$Km_{2,5} = 2.934$
$Km_{4,5} = 0.0706$
$Km_{5,5} = 1.4030$
$Km_{6,5} = 0.0353$
$Km_{14,5} = 0.9051.$

Reaction 6a–b

$$NADH_5 + ACCOA_6 \longleftrightarrow NAD_4^+ + COA_{14} + ACAL_9$$
$$NADH_5 + ACAL_9 \longleftrightarrow NAD_4^+ + ETOH_7$$

$$v_{6a} = V_{\max_{6a}} \cdot \frac{\frac{X_5 \cdot X_6}{Km_{5,6a} \cdot Km_{6,6a}} - \frac{V_{\max_{6a}} \cdot X_4 \cdot X_{14} \cdot X_9}{\left(Keq_{6a} \cdot Km_{5,6a} \cdot Km_{6,6a}\right)}}{\left(1 + \frac{X_4}{Km_{4,6a}} + \frac{X_5}{Km_{5,6a}}\right) \cdot \left(1 + \frac{X_6}{Km_{6,6a}} + \frac{X_{14}}{Km_{14,6a}}\right) \cdot \left(1 + \frac{X_9}{Km_{9,6a}}\right)}$$

$$v_{6b} = V_{\max_{6b}} \cdot \frac{\frac{X_5}{Km_{5,6b}} \cdot \frac{X_9}{Km_{9,6b}} \cdot \left(1 - \frac{X_4 \cdot X_7}{Keq_{6b} \cdot X_5 \cdot X_9}\right)}{\left(1 + \frac{X_7}{Km_{7,6b}} + \frac{X_9}{Km_{9,6b}}\right) \cdot \left(1 + \frac{X_4}{Km_{4,6b}} + \frac{X_5}{Km_{5,6b}}\right)}$$

$V_{\max_{6a}} = 1.6255$
$Km_{4,6a} = 0.0706$
$Km_{5,6a} = 1.4030$
$Km_{6,6a} = 0.0353$
$Km_{14,6a} = 0.9051$

$V_{\max_{6b}} = 3.36986$
$Keq_{6b} = 5.674845$
$Km_{4,6b} = 0.1989$
$Km_{5,6b} = 1.4038$
$Km_{7,6b} = 1.7854$
$Km_{9,6b} = 0.2656.$

Reaction 7

$$ACP_{11} + ADP_{12} \longleftrightarrow AC_{10} + ATP_{13}$$

$$v_7 = V_{\max_7} \cdot \frac{\frac{X_{11}}{Km_{11,7}} \cdot \frac{X_{12}}{Km_{12,7}} \cdot \left(1 - \frac{X_{10} \cdot X_{13}}{Keq_7 \cdot X_{11} \cdot X_{12}}\right)}{\left(1 + \frac{X_{10}}{Km_{10,7}} + \frac{X_{11}}{Km_{11,7}}\right) \cdot \left(1 + \frac{X_{12}}{Km_{12,7}} + \frac{X_{13}}{Km_{13,7}}\right)}$$

$V_{\max_7} = 0.5932$
$Keq_7 = 4.7477$
$Km_{10,7} = 1.771925$
$Km_{11,7} = 0.00001677$
$Km_{12,7} = 0.059508$
$Km_{13,7} = 0.16776.$

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016

References

[1] A. Sjoberg, I. Persson, M. Quednau, B. Hahn-Hagerdal, The influence of limiting and non-limiting growth conditions on glucose and maltose metabolism in *Lactococcus lactis* ssp. *lactis* strain, Appl. Microbiol. Biotechnol. 42 (1995) 931–938.
[2] T.D. Thomas, K.W. Turner, V.L. Crow, Galactose fermentation by *Streptococcus lactis* and *Streptococcus cremoris*: Pathways, products and regulation, J. Bacteriol. 144 (1980) 672–682.
[3] A.R. Neves, Metabolic strategies to reroute carbon fluxes in *Lactococcus lactis*: Kinetics of intracellular metabolite pools by *in vivo* nuclear magnetic resonance, Doctoral Dissertation, Universidade Nova de Lisboa, Instituto de Tecnologia Quimica e Biologica, 2001.
[4] S. Takahashi, K. Abbe, T. Yamada, Purification of pyruvate formate-Lyase from *Streptococcus* mutants and its regulatory properties, J. Bacteriol. 149 (1982) 1034–1040.
[5] H.N.M. Hoefnagel, C.J.M. Starrenburg, E.D. Martens, J. Hugenholtz, M. Kleerebezem, Van Swam II, R. Bongers, V.H. Westerhoff, L.J. Snoep, Metabolic engineering of lactic acid bacteria, the combined approach: Kinetic modeling, metabolic control and experimental analysis, Microbiology 148 (2002) 1003–1013.
[6] M.A. Savageau, Biochemical systems theory: Operational differences among variant representations and their significance, J. Theoret. Biol. 151 (1991) 509–530.
[7] H. Kacser, L. Acerenza, A universal method for achieving increases in metabolite production, Eur. J. Biochem. 216 (1993) 361–367.
[8] B.N. Kholodenko, M. Cascante, J.B. Hoek, H.V. Westerhoff, J. Schwaber, Metabolic design: How to engineer a living cell to desired metabolite concentrations and fluxes, Biotechnol. Bioeng. 59 (1998) 239–247.
[9] M.A. Savageau, Biochemical systems analysis. 1. Some mathematical properties of the rate law for the component enzymatic reactions, J. Theoret. Biol. 25 (1969) 365–369.
[10] M.A. Savageau, Biochemical systems analysis. 2. The steady-state solutions for an *n*-pool system using a power law approximation, J. Theoret. Biol. 25 (1969) 370–379.
[11] E.O. Voit, M.A. Savageau, Accuracy of alternative representations for integrated biochemical systems, Biochemistry 26 (1987) 6869–6880.
[12] L. Regan, I.D.L. Bogle, P. Dunnill, Simulation and optimization of metabolic pathways, Comput. Chem. Eng. 17 (5–6) (1993) 627–637.
[13] N.V. Torres, E.O. Voit, C. Gonzalez-Alcon, Optimization of nonlinear biotechnological processes with linear programming: Application to citric acid production by *Aspergillus niger*, Biotechnol. Bioeng. 49 (1996) 247–258.
[14] E.O. Voit, Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity, Van Nostrand Reinhold, New York, 1991, p. xii–365.
[15] N.V. Torres, E.O. Voit, C. Gonzalez-Alcon, F. Rodriguez-Acosta, An integrated optimization method for biochemical systems. Description of method and application to ethanol, glycerol and carbohydrate production in *Saccharomyces cervisiae*, Biotechnol. Bioeng. 55 (5) (1997) 758–772.
[16] S. Shuster, R. Heinrich, Minimization of intermediates concentrations as a suggested optimality principle for biochemical networks. I. Theoretical analysis, J. Math. Biol. 29 (1991) 425–442.
[17] S. Shuster, R. Schuster, R. Heinrich, Minimization of intermediates concentrations as a suggested optimality principle for biochemical networks, II. Time hierarchy, enzymatic rate laws, and erytrocyte metabolism, J. Math. Biol. 29 (1991) 443–455.
[18] G. Pettersson, Evolutionary optimization of the catalytic efficiency of enzymes, Eur. J. Biochem. 206 (1992) 289–295.
[19] F. Rodriguez-Acosta, R.M. Carlos, T.V. Nestor, Non-linear optimization of biotechnological processes by stochastic algorithms: Application to the maximization of the production rate of ethanol, glycerol and carbohydrates by *Saccharomyces cerevisiae*, J. Biotechnol. 68 (1999) 15–28.
[20] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.
[21] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1987.
[22] A. Bijork, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, PA, 1996.
[23] C.R. Reeves, J.E. Rowe, Genetic Algorithms: Principles and Perspectives: A Guide to GA Theory, Kluwer Academic Publishers, Boston, 2003.
[24] S. Marino, E.O. Voit, An automated procedure for the extraction of metabolic network information from time series data, J. Bioinform. Comput. Biol. 4 (3) (2006) 665–691.
[25] E.O. Voit, Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists, Cambridge University Press, New York, 2000.
[26] M.A. Savageau, Biochemical systems analysis: A study of function and design in molecular biology, in: Advanced Book Program, Addison Wesley Pub. Co., Reading, MA, USA, 1976, p. xvii-379.
[27] S.H. Fogler, Elements of Chemical Reaction Engineering, Prentice-Hall, New Jersey, 1992.
[28] E.O. Voit, J. Almeida, Decoupling dynamical systems for pathway identification from metabolic profiles, Bioinformatics 11 (2004) 1670–1681.
[29] Y. Maki, T. Ueda, M. Okamoto, N. Uematsu, Y. Inamura, Y. Eguchi, Inference of genetic network using the expression profile time course data of mouse P19 cells, Genome Informatics 13 (2002) 382–383.
[30] E.O. Voit, J. Almeida, S. Marino, R. Lall, G. Goel, A.R. Neves, H. Santos, Regulation of glycolysis in *Lactococcus lactis*: An unfinished systems biological case study, IEEE Proc. Syst. Biol. 153 (4) (2006).
[31] A.R. Neves, A. Ramos, M.C. Nunes, M. Kleerebezem, W.M. Hugenholtz J de Vos, J. Almeida, H. Santos, *in vivo* Nuclear Magnetic Resonance studies of glycolytic kinetics in *Lactococcus lactis*, Biotechnol. Bioeng. 64 (2) (1999) 200–212.
[32] A.R. Neves, A. Ramos, C. Shearman, M.J. Gasson, J.S. Almeida, H. Santos, Metabolic characterization of *Lactococcus lactis* deficient in lactate dehydrogenase using *in vivo* 13C NMR, Eur. J. Biochem. 267 (12) (2000) 3859–3868.
[33] A.R. Neves, R. Ventura, N. Mansour, C. Shearman, M.J. Gasson, C. Maycock, A. Ramos, H. Santos, Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? Kinetics of NAD^+ and NADH pools determined *in vivo* by 13C NMR, J. Biol. Chem. 277 (31) (2002) 28088–28098.
[34] D.M.W. Vos, J. Hugenholtz, Engineering metabolic highways in *Lactococci* and other lactic acid bacteria, Trends Biotechnol. 22 (2) (2004) 72–79.
[35] B.G. Olivier, J.L. Snoep, Web-based kinetic modelling using JWS online, Bioinformatics 20 (2004) 2143–2144.
[36] R. Lall, E.O. Voit, Parameter estimation in modulated, unbranched reaction chains within biochemical systems, Comput. Biol. Chem. 29 (2005) 309–318.

Please cite this article in press as: R. Lall, et al., Optimizing ethanol production selectivity, Mathematical and Computer Modelling (2010), doi:10.1016/j.mcm.2010.01.016