Vol. 23, No. 3(2024) Sim24312
Revista Mexicana de Ingeniería Química

Modelling and analysis on the ethanol production by the *Torulaspora delbrueckii* yeast

Modelado y análisis sobre la producción de etanol por la levadura *Torulaspora delbrueckii*

P.A. Valle[1]*, Y. Salazar, N. O. Soto-Cruz[3], J. B. Páez-Lerma[3,*], L. N. Coria[1], M. E. Núñez-Guerrero[4], R. Rodriguez-Herrera[5], L. D. Herrera[1]

[1] Tecnológico Nacional de México / IT Tijuana, Departamento de Ingeniería Eléctrica y Electrónica, Grupo de Investigación BioMath, Tijuana, Baja California, paul.valle@tectijuana.edu.mx, luis.coria@tectijuana.edu.mx, luis.herrera18@tectijuana.edu.mx

[2]Tecnológico Nacional de México / IT Durango, Departamento de Ingeniería Eléctrica y Electrónica, Durango, Durango. ysalazar@itdurango.edu.mx

[3] Tecnológico Nacional de México / IT Durango, Departamento de Ingenierías Química y Bioquímica, Durango, Durango. nsoto@itdurango.edu.mx, jpaez@itdurango.edu.mx.

[4]Hospital General de Zona 1 del Instituto Mexicano del Seguro Social, Durango, Durango. marthang83@hotmail.com

[5]Facultad de Ciencias Químicas, Ciencia y Tecnología de Alimentos, Universidad Autónoma de Coahuila, C. José Cárdenas Valdez and Av. Venustiano Carranza, Saltillo 25280, Mexico; raul.rodriguez@uadec.edu.mx

Received: April 17, 2024; Accepted: August 8, 2024

Abstract

This work presents two mechanistic models composed of first-order nonlinear Ordinary Differential Equations (ODEs). These systems describe the evolution of alcoholic fermentation (ethanol production) by *Torulaspora delbrueckii* yeast when glucose and fructose are used as substrates. The set of ODEs was fitted to three sets of experimental data for the concentration (in g/L per hour) of glucose [$x(t)$], fructose [$y(t)$], and ethanol [$z(t)$]. The parameter values for each equation were estimated using a computational algorithm for nonlinear regression through least squares with a 95% confidence interval to determine if the obtained values were statistically significant. The fitting ability of the models to reproduce experimental data was also assessed using the coefficient of determination (R^2) and the Akaike Information Criterion (AIC). Additionally, numerical simulations were employed to illustrate the analysis results and predict ethanol concentration after the 72 hours during which the experimental data were recorded. Furthermore, the asymptotic stability of each formulated system was analyzed using the Lyapunov stability theory, and its global dynamics were explored through the method of Localization of Compact Invariant Sets.

Keywords: Alcoholic fermentation; Biostatistics; Dynamical systems; Experimental data; *In- silico* experimentation.

Resumen

En este trabajo se presentan dos modelos mecanicistas compuestos por Ecuaciones Diferenciales Ordinarias (EDOs) no lineales de primer orden; estos sistemas describen la evolución de una fermentación alcohólica (producción de etanol) por la levadura *Torulaspora delbrueckii* cuando se utilizan como sustratos glucosa y fructosa. El conjunto de EDOs se ajustaron a tres muestras de datos experimentales de la concentración (en g/L por hora) de glucosa [$x(t)$], fructosa [$y(t)$] y etanol [$z(t)$], los valores de los parámetros de cada ecuación se estimaron mediante un algoritmo computacional de regresión no lineal por mínimos cuadrados con un intervalo de confianza del 95 % para determinar si los valores obtenidos son estadísticamente significativos, también se evaluó la capacidad de ajuste de los modelos para reproducir los datos experimentales mediante el coeficiente de determinación (R^2) y el Criterio de Información de Akaike (AIC). Adicionalmente, las simulaciones numéricas permitieron ilustrar los resultados del análisis y se utilizaron para predecir la concentración de etanol después de las 72 horas en que se registraron los datos experimentales. Por otro lado, se analizó la estabilidad asintótica de cada sistema formulado mediante la teoría de estabilidad en el sentido de Lyapunov y su dinámica global por el método de Localización de Conjuntos Compactos Invariantes.

Palabras clave: Bioestadística, Datos experimentales, Fermentación alcohólica, Experimentación *in-silico*, Sistemas dinámicos.

*Corresponding author. *E-mail*: paul.valle@tectijuana.edu.mx; jpaez@itdurango.edu.mx.;
https://doi.org/10.24275/rmiq/Sim24312
ISSN:1665-2738, *issn-e*: 2395-8472

1 Introduction

Torulaspora delbrueckii is a non-*Saccharomyces* yeast that has been studied in biotechnological industries for its role as a biocontrol agent against spoilage organisms, reducing the need for chemical preservatives. It has emerged as an alternative to sulfur dioxide, which can have negative health effects, especially for asthmatic individuals, and for its potential in ethanol production from glucose and fructose. This yeast stands out due to its ability to introduce diversity into the established standards of the wine and bakery markets, which have been dominated by the *Saccharomyces cerevisiae* yeast (Fernandes *et al*., 2021; Escribano *et al*., 2022). Dutraive *et al*. (2019) reported that *T. delbrueckii* was inoculated to start the grape fermentation, and *S. cerevisiae* was introduced after 24-48 h. It enhanced the wine aroma, although the *T. delbrueckii* population decayed 2-4 days after S. cerevisiae inoculation. When co-cultured, *T. delbrueckii* downregulates glycolysis and overexpresses catalases to control reactive oxygen species and maintain homeostasis (Tondini *et al*., 2020). It could allow better competition against *S. cerevisiae* during the first hours of fermentation when a high osmotic stress occurs.

Despite the well-identified phases of a batch fermentation process (Yang & Sha, 2019), accurately predicting the evolution of *T. delbrueckii* and its ethanol production is challenging. This difficulty arises because it involves the interaction of multiple variables and nonlinear processes. Therefore, there are tools to aid in understanding and predicting its behavior under specific conditions, one of which is the development of mechanistic mathematical models through Systems Biology. The aim is to integrate experimental data into the modelling process to provide a more precise and accurate description of the system (Madni *et al*., 2019).

Mathematical models in predictive microbiology can be classified according to different functionalities based on the type and number of variables. For instance, they can be categorized as primary or secondary models. They can also be differentiated based on their mathematical basis as mechanistic or empirical models, and further categorized as structured or unstructured depending on the complexity of biomass chemical compounds and the relationship between their variables (García *et al*., 2021; Salazar *et al*., 2023). In this work, two mechanistic models are formulated and analyzed. These kinds of models are formulated based on a set of assumptions about the scientific principles underlying the phenomenon being modeled (Ledder, 2022). This approach was chosen because it allows for a more comprehensive understanding of the factors that influence the fermentation process.

The mechanistic models formulated in this work allow for *in-silico* experiments to predict the evolution of *T. delbrueckii* yeast in various scenarios with different initial conditions of substrates (glucose and fructose). This has the potential to reduce costs and accelerate the development of more efficient production strategies. Additionally, it is important to highlight that the parameters of each model were estimated using nonlinear regression algorithms by fitting each model to three sets of samples, each one with their corresponding data presented as concentrations in g/L over a 72-hour interval. For each estimated value, biostatistics such as the standard error (SE), margin of error (MoE), 95% confidence intervals (CI95%), and p-values were calculated. This is done with the purpose of discussing the statistical significance of the results (Motulsky, 2014). Characteristics concerning the global dynamics of the systems are studied by applying nonlinear systems theories such as stability in the sense of Lyapunov (Khalil, 2002) and the Localization of Compact Invariant Sets (LCIS) method (Krishchenko & Starkov, 2006; Valle *et al*., 2021).

Hence, the fundamental objective of this research is to provide both a mathematical and computational tool that is able to estimate substrate consumption (glucose and fructose) for ethanol production without measuring biomass concentration throughout the process.

2 Materials and methods

2.1 Experimental data

The *T. delbrueckii* strain ITD-00014a was obtained from the yeast collection of the Microbial Biotechnology Laboratory at the Durango Institute of Technology by Nuñez *et al*., 2016. Fermentation kinetics were conducted in triplicate using *Agave duranguensis* juice with an initial sugar concentration of 129±4.7 g/L (12% glucose and 88% fructose) and a C/N ratio of 73 g C/g N as substrate. Fermentation kinetics were carried out in glass tubes (20×150 mm), each containing 15 mL of agave juice inoculated with 1×10^8 cells/mL. These tubes were then incubated at 28°C for 72 hours. Samples were analyzed every 8 hours using high-performance liquid chromatography (HPLC) (De los Ríos *et al*., 2015) to determine the concentrations of glucose, fructose, and ethanol (Guerrero *et al*., 2019).

2.2 *Modelling, nonlinear regression, and biostatistics*

This section presents the mathematical models formulated by means of first-order nonlinear ODEs to describe the dynamics of the three variables involved in the fermentation process, where $x_i(t)$ represents glucose, $y_i(t)$ represents fructose, and $z_i(t)$ represents ethanol, with $i = A, B, C$ as each letter identifies each sample. Variables are measured as concentrations in g/L and the time units are given in hours (h). **Variant 1** is formulated by Equations (1)-(3) as indicated below:

$$\dot{x}_1 = -\rho_4 x_1 z_1 - \rho_1 x_1, \tag{1}$$

$$\dot{y}_1 = -\rho_5 y_1 z_1 - \rho_2 y_1, \tag{2}$$

$$\dot{z}_1 = +\rho_6 (x_1 + y_1) z_1 - \rho_3 z_1. \tag{3}$$

Meanwhile, **Variant 2** is formulated by Equations (4)-(6):

$$\dot{x}_2 = -\frac{\rho_4 x_2 z_2}{\rho_7 + x_2} - \rho_1 x_2, \tag{4}$$

$$\dot{y}_2 = -\frac{\rho_5 y_2 z_2}{\rho_7 + y_2} - \rho_2 y_2, \tag{5}$$

$$\dot{z}_2 = +\rho_6 (x_2 + y_2) z_2 - \rho_3 z_2. \tag{6}$$

The equations corresponding to the consumption dynamics of glucose [(1) and (4)] and fructose [(2) and (5)] are modeled by different growth laws. In Variant 1, the law of mass action is applied; meanwhile, in Variant 2, the process is modelled by the classic Monod like-kinetics. However, ethanol production is described by the law of mass action in both variants. These differences will be illustrated in the process of both approximation and prediction through *in-silico* experimentation.

Now, let us describe the dynamics of each equation in our proposed mechanistic models. As mentioned before, the consumption of glucose and fructose by *T. delbrueckii* for ethanol production is modeled by means of the law of mass action in Equations (1) and (2) as indicated, respectively, by terms $-\rho_4 x_1 z_1$ for glucose, and $-\rho_5 y_1 z_1$ for fructose. Meanwhile, the latter is described in Equations (4) and (5) by terms $-\rho_4 x_2 z_2 / (\rho_7 + x_2)$ and $-\rho_5 y_2 z_2 / (\rho_7 + y_2)$, respectively. Further, spontaneous non-biological decomposition terms are considered for each substrate, $-\rho_1 x_i$ for glucose and $-\rho_2 y_i$ for fructose, as well as an ethanol degradation term $-\rho_3 z_i$ in both mathematical models. Ethanol production is modeled by a combination of variables following the law of mass action as indicated by the term $+\rho_6 (x_i + y_i) z_i$ in Equations (3) and (6). Here, it should be noted that parameter ρ_6 represents the combined conversion rate of both glucose and fructose into ethanol (the complete substrate) by the *T. delbrueckii* yeast. Table 1 presents descriptions and units of each variable and parameter of Variants 1 and 2.

Additionally, by considering the positivity property in nonlinear systems established by De Leenheer & Aeyels in 2001, the following remark can be established:

Remark I: Nonnegative solutions. *Solutions of Equations (1)-(6) will be positive for non-negative initial conditions, and the semi-trajectories will be positively forward invariant. Therefore, the dynamics of each system is located in the nonnegative octant defined by the following domain:*

$$\mathbf{R}^3_{+,0} = \{x_i(t), y_i(t), z_i(t) \geq 0 \mid i = 1, 2\}.$$

Degradation rates of glucose (ρ_1) and fructose (ρ_2) are calculated based on the estimates of the half-life ($t_{1/2}$) of each substrate, which were determined by Wolfenden & Yuan in 2008 with the following results:

Table 1. Description and units of variables and parameters.

Variables / Parameters	Description	Units Variant 1	Units Variant 2
$x_i(t)$	Glucose concentration.	g/L	g/L
$y_i(t)$	Fructose concentration.	g/L	g/L
$z_i(t)$	Ethanol concentration.	g/L	g/L
ρ_1	Spontaneous non-biological decomposition rate of glucose.	h^{-1}	h^{-1}
ρ_2	Spontaneous non-biological decomposition rate of fructose.	h^{-1}	h^{-1}
ρ_3	Degradation rate of ethanol.	h^{-1}	h^{-1}
ρ_4	Glucose consumption rate for ethanol production.	L/(g×h)	h^{-1}
ρ_5	Fructose consumption rate for ethanol production.	L/(g×h)	h^{-1}
ρ_6	Combined conversion rate of glucose and fructose into ethanol by the *T. delbrueckii* yeast.	L/(g×h)	L/(g×h)
ρ_7	Affinity constant of substrates for ethanol production.	g/L	g/L

$$t_{1/2}[glucose] = 96\ years = 840{,}960\ hours,$$
$$t_{1/2}[fructose] = 70\ days = 1{,}680\ hours.$$

Therefore, by considering first-order kinetics (Byers & Sarver, 2009), degradation rates are calculated as follows:

$$\rho_1 = \frac{\ln(2)}{t_{1/2}[glucose]} = \frac{\ln(2)}{840{,}960\ h} = 824.233 \times 10^{-9}\ h^{-1},$$
$$\rho_2 = \frac{\ln(2)}{t_{1/2}[fructose]} = \frac{\ln(2)}{1{,}680\ h} = 412.588 \times 10^{-6}\ h^{-1}.$$

Regarding the value of parameter ρ_3, it was set to $1.1\times \max\{\rho_1,\rho_2\}$ in order to fulfill the bounding conditions derived in Section 2.3 with further discussion provided in Section 3. Hence, the ethanol degradation rate is set as follows:

$$\rho_3 = 453.846 \times 10^{-6}\ h^{-1}.$$

Now, based on the above, the nonlinear regression algorithm is designed to estimate the values of the following parameters when fitting each system to the experimental data sets of samples A, B, and C:

$$Variant\ 1\ [Equations\ (1)-(3)]:\ \rho_4,\ \rho_5,\ \rho_6,$$
$$Variant\ 2\ [Equations\ (4)-(6)]:\ \rho_4,\ \rho_5,\ \rho_6,\ \rho_7.$$

The *fitnlm* function is the core of the algorithm developed in MATLAB (MathWorks, n.d.) with the following initial conditions for the parameter estimation:

$$Variant\ 1\ :\ \rho_0 = [0.1, 0.1, 0.1],$$
$$Variant\ 2\ :\ \rho_0 = [0.1, 0.1, 0.1, 0.1].$$

As stated before, the nonlinear regression algorithm was developed in MATLAB, particularly in release 2024a. Equations (1)-(6) were numerically solved by means of Heun's method (also known as the improved Euler's method) with an integration step (Δ_t) equal to 1×10^{-3}. Further, all *in-silico* experiments were carried out on a computer with the following specifications: CPU Ryzen 9 7950X, 192 GB of RAM DDR5 CL40, and a 1 TB Samsung 990 Pro Gen 4 NVMe M.2 drive. The algorithm is provided as supplementary material under the INDAUTOR registration number 03-2024-021609004700-01.

2.3 Localizing domain

The Localization of Compact Invariant Sets (LCIS) method was established by Krishchenko & Starkov in 2006 to study the dynamics in both short- and long-term of nonlinear systems of first-order ODEs. The General Theorem is written as follows: *Each compact invariant set* Γ *of an autonomous nonlinear system of the form* $\dot{x} = f(x)$ *is located either inside or at the boundaries of the* ***localizing domain****:*

$$K(h) = \left\{h_{\inf} \le h(x) \le h_{\sup}\right\},$$

where $f(x)$ is a C^∞-differentiable vector function and $x \in \mathbf{R}^n$ is the state vector. Examples of compact invariant sets are the following: equilibrium points; periodic, homoclinic, and heteroclinic orbits; limit cycles, and chaotic attractors. In order to apply the LCIS method, first one should propose the so-called **localizing function** $h(x) : \mathbf{R}^n \to \mathbf{R}$, a C^∞-differentiable function that is not the first integral of $f(x)$, and from which one can determine the set $S(h) = \left\{x \in \mathbf{R}^n | L_f h(x) = 0\right\}$, where $L_f h(x)$ is the Lie derivative of $\dot{x} = f(x)$ and is given by $L_f h = (\partial h/\partial x) f(x)$. From the latter, by basic algebraic calculations, one can define the lower and upper bounds as $h_{\inf} = inf\{h(x)|x \in S(h)\}$, and $h_{\sup} = sup\{h(x)|x \in S(h)\}$. Further, $h|_S$ denotes the restriction of $h(x)$ on a set $S(h) \subset \mathbf{R}^n$. Now, if all compact invariant sets are located in the $K(h_i)$ and $K\left(h_j\right)$, with $K(h_i), K\left(h_j\right) \subset \mathbf{R}^n$, then they are located in the set $K(h_i) \cap K\left(h_j\right)$. Hence, if the location of all compact invariant sets is inside the domain $\Lambda \subset \mathbf{R^n}$, then the set $K(h) \cap \Lambda$ may be formulated as well. However, if $K(h) \cap \Lambda = \varnothing$, then system $\dot{x} = f(x)$ has no compact invariant sets located in Λ (Valle *et al*., 2021).

However, before applying the LCIS method, it is evident by Remark I that all components in the vector fields of Equations (1)-(2) and (4)-(5) are negative (see also Section 1.5 of Garfinkel *et al*., 2017). Hence, all solutions of Equations (1)-(2) $[x_i(t)]$ and (4)-(5) $[y_i(t)]$; $i = 1,2$, will always be decreasing from a positive initial condition, i.e., $x_i(0), y_i(0) > 0$, and we can establish their lower and upper bounds as follows:

$$K_{x_i} = \{0 \le x_i(t) \le x(0)\},$$
$$K_{y_i} = \{0 \le y_i(t) \le y(0)\}.$$

Now, with the aim of defining the localizing domain of system (1)-(3), the following function is proposed:

$$h_1 = \beta_1 x_1 + \beta_2 y_1 + \beta_3 z_1,$$

whose Lie derivative is given by:

$$L_f h_1 = -\beta_1\rho_1 x_1 - \beta_2\rho_2 y_1 - \beta_3\rho_3 z_1 - (\beta_1\rho_4 - \beta_3\rho_6) x_1 z_1 - (\beta_2\rho_5 - \beta_3\rho_6) y_1 z_1,$$

therefore, set $S(h_1) = \left\{L_f h_1 = 0\right\}$ may be written as follows when considering that $\beta_3 z_1 = h_1 - \beta_1 x_1 - \beta_2 y_1$:

$$S(h_1) = \left\{h_1 = \frac{\beta_1}{\rho_3}(\rho_3 - \rho_1) x_1 + \frac{\beta_2}{\rho_3}(\rho_3 - \rho_2) y_1 - \frac{\beta_1\rho_4 - \beta_3\rho_6}{\rho_3} x_1 z_1 - \frac{\beta_2\rho_5 - \beta_3\rho_6}{\rho_3} y_1 z_1\right\}$$

hence, if the next set of conditions hold:

$$\rho_3 > \max\{\rho_1, \rho_2\},$$
$$\beta_3 < \frac{1}{\rho_6}\, min\{\beta_1\rho_4, \beta_2\rho_5\},$$

then, the **Iterative Theorem** (see Krishchenko & Starkov, 2006; and Valle *et al*., 2021) is applied by intersecting sets $S(h_1) \cap K_{x_1} \cap K_{y_1}$ with the next result:

$$K_{z_1} = \left\{ z_1(t) \leq \frac{\beta_1(\rho_3 - \rho_1)}{\beta_3\rho_3} x_1(0) + \frac{\beta_2(\rho_3 - \rho_2)}{\beta_3\rho_3} y_1(0) \right\},$$

and we can conclude the following:

Result I: Localizing domain of Variant 1. *All compact invariant sets of Variant 1, defined by the system of Equations (1)-(3), are located within or at the boundaries of the localizing domain:*

$$K_{\Gamma_1} = K_{x_1} \cap K_{y_1} \cap K_{z_1}.$$

Now, let us exploit the next function in order to determine the localizing domain of system (4)-(6):

$$h_2 = \gamma_1 x_2 + \gamma_2 y_2 + \gamma_3 z_2,$$

and compute its Lie derivative as follows:

$$L_f h_2 = -\gamma_1\rho_1 x_2 - \gamma_2\rho_2 y_2 - \gamma_3\rho_3 z_2 + \frac{\gamma_3\rho_6 x_2 + \gamma_3\rho_6\rho_7 - \gamma_1\rho_4}{\rho_7 + x_2} x_2 z_2 + \frac{\gamma_3\rho_6 y_2 + \gamma_3\rho_6\rho_7 - \gamma_2\rho_5}{\rho_7 + y_2} y_2 z_2,$$

therefore, set $S(h_2) = \left\{ L_f h_2 = 0 \right\}$ may be written as shown below when considering that $\gamma_3 z_2 = h_2 - \gamma_1 x_2 - \gamma_2 y_2$:

$$S(h_2) = \left\{ h_2 = \frac{\gamma_1}{\rho_3}(\rho_3 - \rho_1) x_2 + \frac{\gamma_2}{\rho_3}(\rho_3 - \rho_2) y_2 + \frac{\gamma_3\rho_6 x_2 + \gamma_3\rho_6\rho_7 - \gamma_1\rho_4}{\rho_3(\rho_7 + x_2)} x_2 z_2 + \left(\frac{\gamma_3\rho_6 y_2 + \gamma_3\rho_6\rho_7 - \gamma_2\rho_5}{\rho_3(\rho_7 + y_2)} \right) y_2 z_2 \right\},$$

and, if the next set of conditions hold:

$$\rho_3 > max\{\rho_1, \rho_2\},$$
$$\gamma_3 < \frac{1}{\rho_6} \min\left\{ \frac{\gamma_1\rho_4}{x_2(0) + \rho_7}, \frac{\gamma_2\rho_5}{y_2(0) + \rho_7} \right\},$$

one should apply the **Iterative Theorem** by intersecting sets $S(h_2) \cap K_{x_2} \cap K_{y_2}$ with the following result:

$$K_{z_2} = \left\{ z_2(t) \leq \frac{\gamma_1(\rho_3 - \rho_1)}{\gamma_3\rho_3} x_2(0) + \frac{\gamma_2(\rho_3 - \rho_2)}{\gamma_3\rho_3} y_2(0) \right\}.$$

Result II: Localizing domain of Variant 2. *All compact invariant sets of Variant 2, defined by the system of Equations (4)-(6), are located within or at the boundaries of the localizing domain:*

$$K_{\Gamma_2} = K_{x_2} \cap K_{y_2} \cap K_{z_2}.$$

2.4 Asymptotic stability

Lyapunov demonstrated that certain types of functions, called Lyapunov candidate functions, can be used to establish the stability of the equilibrium point $x^* = 0$ (the origin of the state space) of a nonlinear, continuous, and time-invariant dynamical system of the form $\dot{x} = f(x)$ (Khalil, 2002).

A Lyapunov candidate function is usually denoted by $V(x) : \mathbf{R}^n \to \mathbf{R}$, where x represents the state vector of the dynamical system, it is continuously differentiable positive definite function, i.e., $V(0) = 0$, and $V(x) > 0$ for $x \neq 0$; and radially unbounded, which implies that $V(x) \to \infty$ when $\|x\| \to \infty$. Its temporal derivative is given by $\dot{V}(x) = (\partial V/\partial x) f(x)$. From the latter, the Theorem of Global Asymptotic Stability in the Sense of Lyapunov is formulated as follows: *The equilibrium point x^* is globally asymptotically stable if there exists a positive definite, radially unbounded, and decreasing function $V(x)$ such that its temporal derivative $\dot{V}(x)$ is negative definite*. A function $V(x)$ that satisfies the properties of this theorem is called a **Lyapunov function**.

First, the asymptotic stability of Variant 1, defined by Equations (1)-(3), is analyzed by proposing the following candidate Lyapunov function:

$$V_1 = x_1 + y_1 + \alpha_1 z_1,$$

Whose temporal derivative is given by:

$$\dot{V}_1 = -\rho_1 x_1 - \rho_2 y_1 - \alpha_1\rho_3 z_1 - (\rho_4 - \alpha_1\rho_6) x_1 z_1 - (\rho_5 - \alpha_1\rho_6) y_1 z_1,$$

hence, if the next condition holds:

$$\alpha_1 < \frac{1}{\rho_6} \min\{\rho_4, \rho_5\},$$

then, the following result is established:

Result III: Asymptotic stability of Variant 1. *System (1)-(3) is asymptotically stable in the nonnegative octant $\mathbf{R}^3_{+,0}$, and all trajectories of Variant 1 will converge to the unique biologically feasible equilibrium point defined by $\left(x_1^*, y_1^*, z_1^*\right) = (0,0,0)$.*

Now, in order to investigate the asymptotic stability of Variant 2, defined by Equations (4)-(6), the following candidate Lyapunov function is proposed:

$$V_2 = x_2 + y_2 + \alpha_2 z_2,$$

then, the temporal derivative is computed as follows:

$$\dot{V}_2 = \frac{\alpha_2\rho_6 x_2 + \alpha_2\rho_6\rho_7 - \rho_4}{\rho_7 + x_2} x_2 z_2 + \frac{\alpha_2\rho_6 y_2 + \alpha_2\rho_6\rho_7 - \rho_5}{\rho_7 + y_2} y_2 z_2 - \rho_1 x_2 - \rho_2 y_2 - \alpha_2\rho_3 z_2.$$

It is evident that $\dot{V}_2(0) = 0$, therefore, when considering localizing sets K_{x_i} and K_{y_i}, one can constrain both rational terms as follows:

$$\alpha_2\rho_6 x_2 + \alpha_2\rho_6\rho_7 - \rho_4 \leq \alpha_2\rho_6 x_2(0) + \alpha_2\rho_6\rho_7 - \rho_4 < 0$$
$$\alpha_2\rho_6 y_2 + \alpha_2\rho_6\rho_7 - \rho_5 \leq \alpha_2\rho_6 y_2(0) + \alpha_2\rho_6\rho_7 - \rho_5 < 0$$

and, if the next condition holds:

$$\alpha_2 < \frac{1}{\rho_6}\min\left\{\frac{\rho_4}{x_2(0)+\rho_7}, \frac{\rho_5}{y_2(0)+\rho_7}\right\},$$

then, following result is established:

Result IV: Asymptotic stability of Variant 2. *System (4)-(6) is asymptotically stable in the nonnegative octant* $\mathbf{R}^3_{+,0}$*, and all trajectories of Variant 2 will converge to the unique biologically feasible equilibrium point defined by* $\left(x_2^*, y_2^*, z_2^*\right) = (0,0,0)$.

3 Results and discussion

Now, as stated before, an algorithm was designed in MATLAB by applying the nonlinear regression function *fitnlm*. Hence, in order to better fit Equations (1)-(6) to the experimental data illustrated by the × marker in Figure 1 (green for sample A, red for sample B, and blue for sample C) we decided to normalize the data[1]. Results obtained for Variant 1 defined by Equations (1)-(3) are presented in Table 2, while Table 3 shows results for Variant 2 defined by Equations (4)-(6). The margin of error (MoE) can be calculated by multiplying the standard error (SE) by the two-tailed Student's t-value corresponding to each model, where the latter depends on the degrees of freedom (dof), i.e., the number of experimental data points minus the parameters to be estimated, and the significance level (α), which is selected as 0.05 to determine the 95% confidence intervals (CI95%) for the estimated value of each parameter, and as a threshold to determine the statistical significance of the results by the computed p-values (Motulsky, 2014).

Regarding the goodness of fit of the approximated solutions provided by Variants 1 and 2, Table 4 presents the R-squared (R^2) results for each sample, and Table 5 shows the Akaike Information Criterion (AIC) results for the fitting capabilities of each variant to the complete set of data.

The *in-silico* experimentation performed in Figure 1 allows us to illustrate and compare results from each mathematical model, where solutions $x_1(t)$, $y_1(t)$, and $z_1(t)$ correspond to Variant 1 [Equations (1)-(3)] and solutions $x_2(t)$, $y_2(t)$, and $z_2(t)$ correspond to Variant 2 [Equations (4)-(6)]. The data illustrates the consumption of both substrates, which is why both concentrations tend to zero, as one can see in the upper and middle panels. The increase in ethanol concentration as a product of this process is observed in the three lower panels of the figure. One can see that both systems are able to approximate the experimental data for the concentrations of glucose [$x_i(t)$], fructose [$y_i(t)$] and ethanol [$z_i(t)$] in the three corresponding samples identified by the × marker in each panel. Further, both variants are able to predict a reasonable ethanol concentration for two times (144 hours) the original 72-hour interval on which of the experimental data was measured (Padmanabhan *et al*., 2021).

Table 2. Estimation results for parameters of Variant 1 defined by Equations (1)-(3). Biostatistical calculations are performed considering 30 experimental data points and 3 parameters to be estimated, resulting in 27 dof and a Student's t-value of 2.0518.

Parameter	Estimate	SE	CI95%	p-value
		Sample A results		
ρ_4	199.706×10^{-3}	18.219×10^{-3}	$(162.324\times10^{-3}, 237.089\times10^{-3})$	19.123×10^{-12}
ρ_5	141.169×10^{-3}	10.079×10^{-3}	$(120.489\times10^{-3}, 161.849\times10^{-3})$	66.795×10^{-15}
ρ_6	79.458×10^{-3}	2.365×10^{-3}	$(74.605\times10^{-3}, 84.311\times10^{-3})$	14.529×10^{-24}
		Sample B results		
ρ_4	200.121×10^{-3}	20.078×10^{-3}	$(158.924\times10^{-3}, 241.318\times10^{-3})$	152.379×10^{-12}
ρ_5	138.764×10^{-3}	10.933×10^{-3}	$(116.331\times10^{-3}, 161.197\times10^{-3})$	680.489×10^{-15}
ρ_6	75.727×10^{-3}	2.425×10^{-3}	$(70.7518\times10^{-3}, 80.701\times10^{-3})$	99.167×10^{-24}
		Sample C results		
ρ_4	171.738×10^{-3}	20.144×10^{-3}	$(130.405\times10^{-3}, 213.071\times10^{-3})$	3.868×10^{-9}
ρ_5	133.451×10^{-3}	13.615×10^{-3}	$(105.514\times10^{-3}, 161.387\times10^{-3})$	217.922×10^{-12}
ρ_6	74.066×10^{-3}	2.912×10^{-3}	$(68.092\times10^{-3}, 80.041\times10^{-3})$	21.241×10^{-21}

[1]The original data supporting the findings of this research is available through authors Nicolás O. Soto-Cruz and Jesús B. Páez-Lerma upon reasonable request.

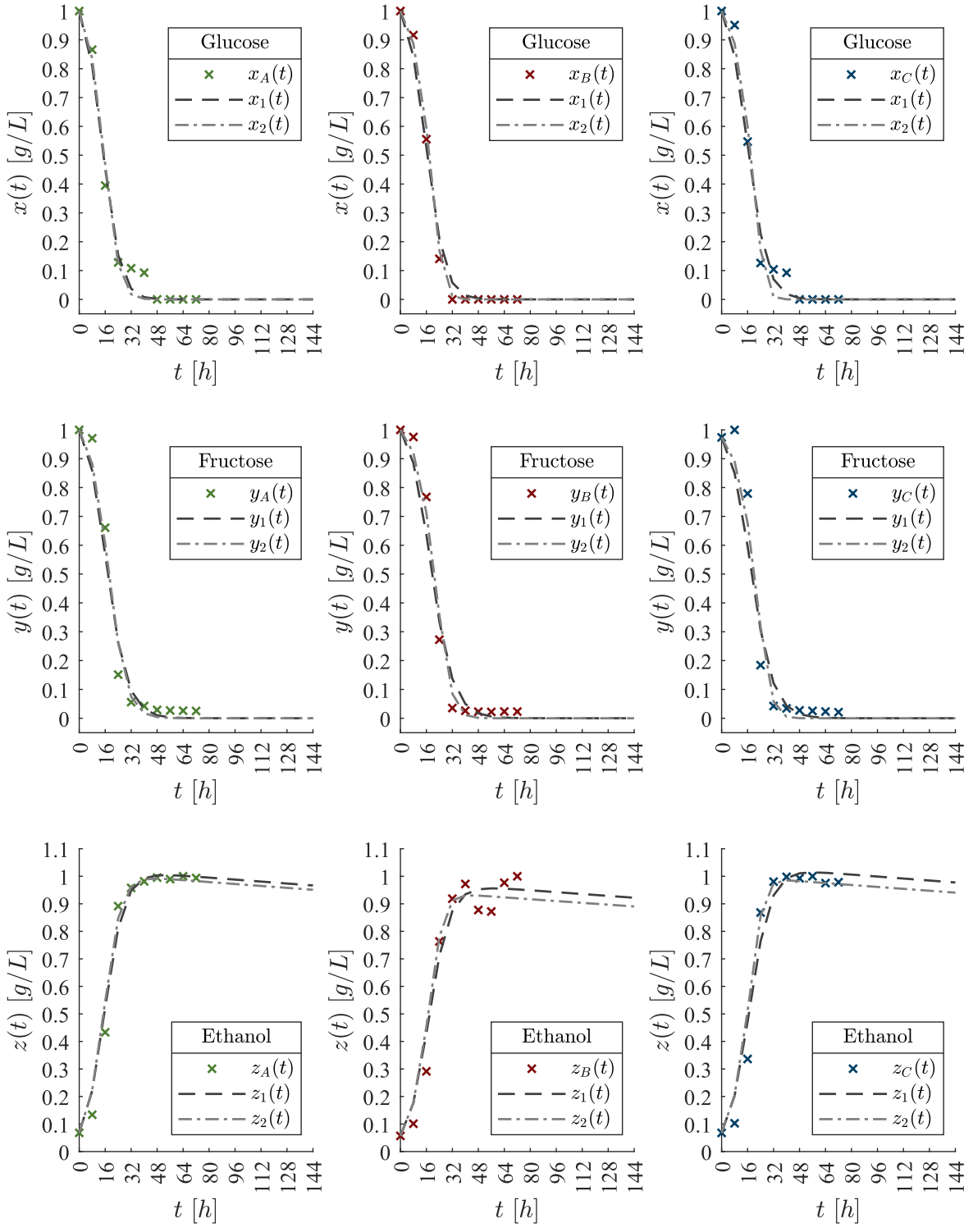

Figure 1. The *in-silico* experimentation illustrates both the goodness of fit and the prediction capabilities of Variant 1 [$x_1(t)$, $y_1(t)$, $z_1(t)$], and Variant 2 [$x_2(t)$, $y_2(t)$, $z_2(t)$] over an interval of 0 to 144 hours. Panels in the left column correspond to the experimental data of sample A [$x_A(t), y_A(t), z_A(t)$], those in the middle column to sample B [$x_B(t), y_B(t), z_B(t)$], and those in the right column to sample C [$x_C(t), y_C(t), z_C(t)$].

Table 3. Estimation results for parameters of Variant 2 defined by Equations (4)-(6). Biostatistical calculations are performed considering 30 experimental data points and 4 parameters to be estimated, resulting in 26 dof and a Student's t-value of 2.0555.

Parameter	Estimate	SE	CI95%	p-value
		Sample A results		
ρ_4	409.916×10^{-3}	252.745×10^{-3}	$(-109.610\times10^{-3},929.442\times10^{-3})$	116.900×10^{-3}
ρ_5	278.678×10^{-3}	177.591×10^{-3}	$(-86.366\times10^{-3},643.722\times10^{-3})$	128.691×10^{-3}
ρ_6	78.960×10^{-3}	2.245×10^{-3}	$(74.347\times10^{-3},83.574\times10^{-3})$	18.543×10^{-24}
ρ_7	1.486	1.242	$(-1.067\times10^{-3},4.039)$	242.419×10^{-3}
		Sample B results		
ρ_4	194.933×10^{-3}	58.645×10^{-3}	$(74.386\times10^{-3},315.480\times10^{-3})$	2.645×10^{-3}
ρ_5	131.174×10^{-3}	41.070×10^{-3}	$(46.754\times10^{-3},215.594\times10^{-3})$	3.657×10^{-3}
ρ_6	74.295×10^{-3}	1.838×10^{-3}	$(70.518\times10^{-3},78.072\times10^{-3})$	530.225×10^{-27}
ρ_7	442.204×10^{-3}	284.415×10^{-3}	$(-142.420\times10^{-3},1.027)$	132.088×10^{-3}
		Sample C results		
ρ_4	150.689×10^{-3}	50.405×10^{-3}	$(47.081\times10^{-3},254.298\times10^{-3})$	6.036×10^{-3}
ρ_5	108.300×10^{-3}	38.025×10^{-3}	$(30.138\times10^{-3},186.461\times10^{-3})$	8.482×10^{-3}
ρ_6	73.333×10^{-3}	2.280×10^{-3}	$(68.645\times10^{-3},78.020\times10^{-3})$	181.999×10^{-24}
ρ_7	323.536×10^{-3}	272.032×10^{-3}	$(-235.634\times10^{-3},882.707\times10^{-3})$	245.059×10^{-3}

Table 4. Results of the goodness of fit by means of the R^2 for system (1)-(3) [Variant 1] and system (4)-(6) [Variant 2] for the three samples (A, B, and C) of experimental data concerning the concentrations of glucose [$x_i(t)$], fructose [$y_i(t)$], and ethanol [$z_i(t)$].

	Variant 1			Variant 2		
Variable	A	B	C	A	B	C
$x_i(t)$	0.985	0.990	0.980	0.982	0.998	0.982
$y_i(t)$	0.975	0.975	0.952	0.983	0.989	0.973
$z_i(t)$	0.982	0.964	0.968	0.983	0.962	0.970

Table 5. Results of goodness of fit by means of the AIC for system (1)-(3) [Variant 1] and system (4)-(6) [Variant 2] for the three samples (A, B, and C) of experimental data.

Mathematical model	Samples		
	A	B	C
Variant 1	-85.836	-78.400	-66.513
Variant 2	-86.114	-86.842	-73.024

As mentioned before, the main difference between Variants lies in the ODEs that describe the dynamics of glucose [Equations (1) and (4)] and fructose [Equations (2) and (5)] concentrations. Variant 1 describes the substrates consumption by the law of mass action, whilst Variant 2 describes this process by following a Monod type-kinetics. Furthermore, the latter model also introduces an additional parameter known as the affinity constant of substrates for ethanol production, ρ_7. However, the biostatistical analysis shown in Table 3 forces us to conclude that fitting results for Variant 2 are not statistically significant, i.e., some CI95% have a negative lower bound, and their corresponding p-values are greater than the predetermined value of the significance level, α. In nonlinear regression models, the latter implies that we cannot reject the null hypothesis (H_0) regarding the influence of parameter ρ_7 in the dynamics of the observed data, i.e., the parameter value should be zero. Nonetheless, the goodness of fit indicates an overall better fitting for Variant 2 to the observed data, as shown in Tables 4 and 5 by the R-squared and the AIC, respectively.

Now, when considering all results, one can conclude that the two Variants adequately describe the variance of the experimental data in the interval from 0 to 72 hours, both quantitatively and qualitatively. Although differences are minor, one can make the following remarks regarding the results determined with each variant: The *in-silico* experimentation illustrates that Variant 1 estimates a higher concentration of ethanol as well as higher

final value for the prediction phase, whilst Variant 2 better approximates the consumption dynamics of both substrates (glucose and fructose).

Concerning the nonlinear dynamics of each system, we were able to compute all lower and upper bounds for the two variants, i.e., a localizing domain where all compact invariant sets are located, as stated in Results I and II. Furthermore, sufficient conditions for asymptotic stability were derived for both systems, as stated in Results III and IV. The latter implies that all solutions describing the concentrations of glucose $[x_i(t)]$, fructose $[y_i(t)]$, and ethanol $[z_i(t)]$ will be bounded in the nonnegative octant and will go the unique biologically feasible equilibrium point, i.e., $\left(x_i^*, y_i^*, z_i^*\right) = (0,0,0)$. Hence, as time tends to infinity, all concentrations will completely degrade.

Following the latter, degradation rates for glucose (ρ_1) and fructose (ρ_2) were calculated from the half-life values reported by Byers & Sarver, 2009; while ethanol degradation rate (ρ_3) was set to fulfill conditions for the localizing domains of each system, as their existence depends on the following constraint: $\rho_3 > max\{\rho_1, \rho_2\}$. It also worth noting that we could not find a reported value for ethanol degradation in the medium where the fermentation kinetics were conducted, nor could we estimate a biologically feasible value by the nonlinear regression process. However, we expect that a long-term measurement of ethanol concentration in the original medium will allow us to better estimate this rate.

Conclusion

In this work, we formulated, analyzed, and validated two nonlinear systems of first-order ODEs which were identified as Variant 1 [Equations (1)-(3)], and Variant 2 [Equations (4)-(6)]. These mechanistic models were proposed to describe an alcoholic fermentation process conducted by the *T. delbrueckii* yeast where the changes in concentrations (g/L) of glucose $[x_i(t)]$, fructose $[y_{i(t)}]$, and ethanol $[z_i(t)]$ were measured every 8 hours in a 72-hour interval. As illustrated in Figure 1, the experimental data was divided into three samples (A, B, and C) with 30 observed data points for each sample (10 for every variable). The main difference between our two models is observed in Figure 2 at the prediction stage. The *in-silico* experimentation for Variant 1 illustrates a higher ethanol estimation, whereas the numerical simulations of Variant 2 better approximate the substrate consumption. Concerning the results of the localizing domain and stability, boundedness ensures that the concentrations of substrates and the product will stay within biologically acceptable ranges. In contrast, dissipativity and asymptotic stability are linked to the system's energy balance, which is often associated with the metabolic and thermodynamic aspects of biological processes.

Acknowledgements

This research was fulfilled within the TecNM projects "Estrategias *in silico* integradas con biomatemáticas y sistemas dinámicos no lineales para el modelizado, análisis y control de sistemas biológicos [19377.24-P]", "Modelizado de sistemas no lineales para procesos de fermentación basados en la dinámica de crecimiento de microorganismos: parte 2 [19805.24-P]", the academic research group Project "Sistemas dinámicos no lineales" ITIJ-CA-10, and the RICCA "Red Internacional de Control y Cómputo Aplicados".

References

Byers, J. P., & Sarver, J. G. (2009). *Pharmacokinetic modeling*. In Pharmacology (pp. 201-277). Academic Press. https://doi.org/10.1016/B978-0-12-369521-5.00010-5

De Leenheer, P., & Aeyels, D. (2001). Stability properties of equilibria of classes of cooperative systems. *IEEE Transactions on Automatic Control*, 46(12), 1996-2001. https://doi.org/10.1109/9.975508

De los Rios-Deras, G.C., Rutiaga-Quiñones, O. M., López-Miranda, J., Páez-Lerma, J. B., López, M., & Soto-Cruz, N. O. (2015). Improving *Agave duranguensis* must for enhanced fermentation: C/N ratio effects on mezcal composition and sensory properties. *Revista Mexicana de Ingeniería Química*, 14(2), 363-371.

Dutraive, O., Benito, S., Fritsch, S., Beisert, B., Patz, C. D., & Rauhut, D. (2019). Effect of sequential inoculation with non-*Saccharomyces* and *Saccharomyces* yeasts on riesling wine chemical composition. *Fermentation*, *5*(3), 79. https://doi.org/10.3390/fermentation5030079

Escribano-Viana, R., González-Arenzana, L., Garijo, P., Fernández, L., López, R., Santamaría, P., & Gutiérrez, A. R. (2022). Bioprotective effect of a Torulaspora delbrueckii/Lachancea thermotolerans-mixed inoculum in red winemaking. *Fermentation*, 8(7), 337. https://doi.org/10.3390/fermentation8070337

Fernandes, T., Silva-Sousa, F., Pereira, F., Rito, T., Soares, P., Franco-Duarte, R., & Sousa,

M. J. (2021). Biotechnological importance of Torulaspora delbrueckii: From the obscurity to the spotlight. *Journal of Fungi*, 7(9), 712. https://doi.org/10.3390/jof7090712

Garcia, B. E., Rodriguez, E., Salazar, Y., Valle, P. A., Flores-Gallegos, A. C., Rutiaga-Quiñones, O. M., & Rodriguez-Herrera, R. (2021). Primary model for biomass growth prediction in batch fermentation. *Symmetry*, 13(08), 1468. https://doi.org/10.3390/sym13081468

Garfinkel, A., Shevtsov, J., & Guo, Y. (2017). *Modeling life: the mathematics of biological systems*. Springer International Publishing AG.

Guerrero, M. E. N., Vázquez, E. S., Páez-Lerma, J. B., Herrera, R. R., & Soto-Cruz, N. O. (2019). Physiological characterization of two native yeasts in pure and mixed culture using fermentations of agave juice. *Ciencia e investigación agraria: revista latinoamericana de ciencias de la agricultura*, 46(1), 1-11.

Khalil, H. K. (2002). *Nonlinear systems*. Prentice-Hall, USA.

Krishchenko, A. P., & Starkov, K. E. (2006). Localization of compact invariant sets of the Lorenz system. Physics Letters A, 353(5), 383-388. https://doi.org/10.1016/j.physleta.2005.12.104

Ledder, G. (2022). Mentoring Undergraduate Research in Mathematical Modeling. *Bulletin of Mathematical Biology*, 84(8), 77. https://doi.org/10.1007/s11538-022-01040-4

Madni, A. M., Madni, C. C., & Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. Systems, 7(1), 7.

MathWorks. (n.d.) fitnlm: Fit nonlinear regression model. https://www.mathworks.com/help/stats/fitnlm.html. Acceso: Febrero 4, 2024.

Motulsky, H. (2014). *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. Oxford University Press, USA.

Nuñez-Guerrero, M. E., Páez-Lerma, J. B., Rutiaga-Quiñones, O. M., González-Herrera, S. M., & Soto-Cruz, N. O. (2016). Performance of mixtures of Saccharomyces and non-Saccharomyces native yeasts during alcoholic fermentation of Agave duranguensis juice. *Food Microbiology*, 54, 91-97. https://doi.org/10.1016/j.fm.2015.10.011

Padmanabhan, R., Meskin, N., & Al Moustafa, A. E. (2021). *Mathematical models of cancer and different therapies*. Springer, Singapore.

Salazar, Y., Valle, P. A., Rodríguez, E., Soto-Cruz, N. O., Páez-Lerma, J. B., & Reyes-Sánchez, F. J. (2023). Mechanistic Modelling of Biomass Growth, Glucose Consumption and Ethanol Production by Kluyveromyces marxianus in Batch Fermentation. *Entropy*, 25(3), 497. https://doi.org/10.3390/e25030497

Tondini, F., Onetto, C. A., & Jiranek, V. (2020). Early adaptation strategies of *Saccharomyces cerevisiae* and *Torulaspora delbrueckii* to co-inoculation in high sugar grape must-like media. *Food Microbiology*, 90, 103463. https://doi.org/10.1016/j.fm.2020.103463

Valle, P. A., Coria, L. N., Plata, C., & Salazar, Y. (2021). CAR-T cell therapy for the treatment of ALL: Eradication conditions and in silico experimentation. *Hemato*, 2(3), 441-462. https://doi.org/10.3390/hemato2030028

Wolfenden, R., & Yuan, Y. (2008). Rates of spontaneous cleavage of glucose, fructose, sucrose, and trehalose in water, and the catalytic proficiencies of invertase and trehalas. *Journal of the American Chemical Society*, 130(24), 7548-7549. https://doi.org/10.1021/ja802206s

Yang, Y., & Sha, M. (2019). A beginner's guide to bioprocess modes–batch, fed-batch, and continuous fermentation. Enfield, CT: Eppendorf Inc, 408, 1-16.