Modelling and optimization of fed-batch culture for 3-hydroxypropionic acid production

Christian de Oliveira Martins[a], Adilson José da Silva[a], Felipe Fernando Furlan[a*]

[a]Graduate Program in Chemical Engineering (PPGEQ), Federal University of São Carlos (UFSCar), São Carlos-SP, Brazil
**furlan@ufscar.br*

ABSTRACT

Sustainable bioprocess transition requires the efficient production of chemical alternatives for petrochemical products, like 3-hydroxypropionic acid (3-HP). However, using microorganisms involves operational challenges since these processes are not fully developed. This study developed a computational tool in Python to model and optimize the fed-batch production of 3-HP by *Escherichia coli*. The parameter identification step required multi-start optimization due to local minimum. Cultivation optimization was based on orthogonal-collocation transcription of the dynamic model and solution through interior-point optimizer. Results showed that Monod-Levenspiel model best captured product inhibition. Eight finite elements under three collocation points were necessary to achieve a 0.1 g/L precision for system dynamics simulation. Maximizing titer yielded an infeasible 511-hour process (product concentration: 135 g/L). Whereas maximizing productivity prioritizes the initial growth phase leading to a 20h process (product concentration: 58.1 g/L), with a productivity 1.68x greater than the experimental study.

Keywords: *Escherichia coli*, optimization, orthogonal collocation, 3-hydroxypropionic acid

1. Background

As the world searches for ways to establish a transition to eco-friendly energy, several traditional petroleum-based products have been substituted or had their production process altered. Examples of such trends are found in the fuel industry, with the substitution of petroleum-based fuel for alternatives like ethanol (Szklo et al., 2007), and the production of 1,3-propanediol using engineered organisms (Nielsen et al., 2022). In this context, 3-hydroxypropionic acid (3-HP) is a prominent platform molecule due to its properties as an acrylic acid precursor and a raw material for bioplastics (Werpy & Petersen, 2004). However, 3-HP large-scale production is absent, showing this product is still not economically feasible. While traditional chemical routes are highly polluting, biochemical pathways are still under development and not yet suitable for continuous industrial production (Nielsen et al., 2022; Wang et al., 2023).

The literature describes numerous efforts to overcome the industrial limitation of biochemical 3-HP production by combining bioreactor engineering and specialized genetically modified organisms. Despite these efforts, industrial-scale implementation remains unsuccessful. Further optimization is required to make this pathway economically and environmentally feasible. As with any microbial process, progress depends heavily on improving **TRY** (**T**iter, **R**ate and **Y**ield) metrics. Enhancing performance in these aspects can significantly reduce overall costs, thereby promoting the industrial viability of new biochemical processes (Konzock & Nielsen, 2024).

The *Escherichia coli* PSO932 strain was designed for the production of 3-HP at higher titers due to genetic modifications that grant the bacteria resistance to high product concentrations. This leads to increases in titer, yield and productivity (Batista et al., 2024). Moreover, fed-batch tests indicate the high potential to improve TRY aspects through feeding control. To achieve this goal, computer-aided tools, such as dynamic optimization algorithms, can be used. The dynamic nature of the system adds complexity, requiring the use of differential equations to describe its behavior. Consequently, solving such problems usually relies on numerical integration methods to discretize the system and in some cases solving the entire problem using an algebraic system approximation.

Among the strategies for solving dynamic optimization problems, the simultaneous approach stands out as a robust and computationally efficient alternative. The method relies on a full discretization of the differential algebraic system of equations through finite elements in time, with piecewise polynomial representations of the

state and control variables in each element. This approach makes it easier to work with path restrictions on variables (Biegler, 2010). In this sense, Pyomo is a collection of software packages that turns Python programming into a modelling environment for optimization. In this tool, it is possible to construct dynamic optimization problems to be solved using orthogonal collocation techniques like the use of Lagrange polynomials with Radau quadrature (Sandia National Laboratories, s. d.).

Therefore, this work aims to implement a discretization protocol to define the simplest, yet sufficiently accurate, description of the system. This approach was evaluated through two optimization problems: the maximization of titer and productivity. This study was conducted by modeling the biochemical production of 3-HP from glycerol using the *Escherichia coli* PSO932 strain in Pyomo, employing a simultaneous dynamic optimization approach.

2. Methods

The work was performed in three steps: first, fitting the kinetic model, second, identifying the lowest number of finite elements (NFE) required to maintain accuracy using three collocation points; and finally, determining the lowest NFE needed to reach the desired proximity to the objective function's optimal point.

2.1. Model selection and parameter estimation

The first step consisted in identifying the best model fit to experimental data for the *E. coli* PSO932 in fed-batch cultures after induction using exponential and linear feed profiles (Batista et al., 2024). The fed-batches mass balances were implemented in Python, assuming perfect pH control for base feeding estimation. This assumption was necessary since no data for base feeding profile were available. However, data retrieved directly from the authors indicate that the overall pH was approximately constant at 6.42 throughout the experiments, with a relative standard deviation of 2%.

The systems of differential equations were solved using the solve_ivp function from the Scipy library. Fitting was performed using LMFIT library through a least-squares method, employing the Levenberg-Marquardt algorithm for minimization. Four models for specific growth rate were tested: classical Monod, Contois, Monod with Levenspiel product inhibition and Monod with Exponential product inhibition, described by the equations 1 to 4.

$$\mu_{monod} = \mu_{max} \cdot \frac{S}{K_s+S} \quad (1)$$

$$\mu_{contois} = \mu_{max} \cdot \frac{S}{K_s \cdot X+S} \quad (2)$$

$$\mu_{levenspiel} = \mu_{monod} \cdot (1 - \frac{P}{P_m})^n \quad (3)$$

$$\mu_{exponential} = \mu_{monod} \cdot exp(-\frac{P}{P_m}) \quad (4)$$

Where μ_{model} represents the specific growth rate for each model, μ_{max} the maximum specific growth rate, S the substrate concentration, K_s represents the saturation constant, X is the cell concentration, P is the product concentration, P_m is the critical product concentration and n is an empirical constant.

The best fit was defined by the least sum of squares of residuals. Once the best model was identified, a multi-start optimization was applied to improve the adjustment. To guarantee the distribution of initial guesses in the optimization, 100 points were sampled from the parameters range using Latin Hypercube Sampling (LHS) modified to minimize the maximum correlation coefficient. The parameter set with the lowest sum of squared residuals was then selected.

2.2. Analysis of minimum number of finite elements

The dynamic optimization model included dissociation equilibria for 3-HP (pK_a = 4.5, Chemarin et al., 2019), the base used, ammonium hydroxide (pK_b = 4.75, Rumble, 2021), and water, as well as the electroneutrality of the mixture. This dissociation equilibrium information is necessary to obtain the optimal base-feeding protocol.

The appropriate number of finite elements (NFE) was determined by comparing the concentration values from the solve_ivp simulation with those from the discretized model. The simulation consisted of an initial batch phase for cell growth without 3-HP production and assuming constant specific growth rate of 0.18 h^{-1}, followed by a constant-feed fed-batch stage after induction. This operation follows the same initial flow rate from the experiments reported by Batista et al. (2024), however the feed flow applied in fed-batch phase was constant from

the exactly induction time. This change was applied to avoid numerical issues with discontinuities in the discretization method.

The results of the discretized model were compared with solve_ivp using the Radau method, with an absolute error tolerance of 10^{-11} and a relative error tolerance of 10^{-9}. The initial number of finite elements was set to 2; at each iteration, this value was doubled until the difference between the concentration of both models was under 10^{-2}.

2.3. Refining number of finite elements for optimization

Once a standard discretization was defined, the model was employed for optimization. The decision variables included the substrate feed profile, the feed concentration, the initial substrate concentration, and the base feed rule. Two different objective functions were evaluated: maximize the titer (final product concentration) and maximize the productivity (titer divided by the total culture time). The number of finite elements was increased using the same logic previously described until the difference between two consecutive optimization results was less than 10^{-2}.

3. Results

3.1. Model selection and parameter estimation

Parameter estimation for all four models was performed by minimize of the sum of the squared residuals. The sum of squared residuals value for each model is shown in Table 1. As can be seen, Monod with Levenspiel inhibition showed the best fit, indicating a possible product inhibition in 3-HP production by *E. coli,* the Contois model explained the data better than the pure Monod and Monod with exponential inhibition. Nevertheless, there is no information in literature on *E. coli* cell inhibition at the biomass concentrations observed in the experiments. Therefore, the performance of Contois model may be solely explained by the extra parameter.

Table 1. Least sum of squares of residuals of the fitting for each specific growth model tested.

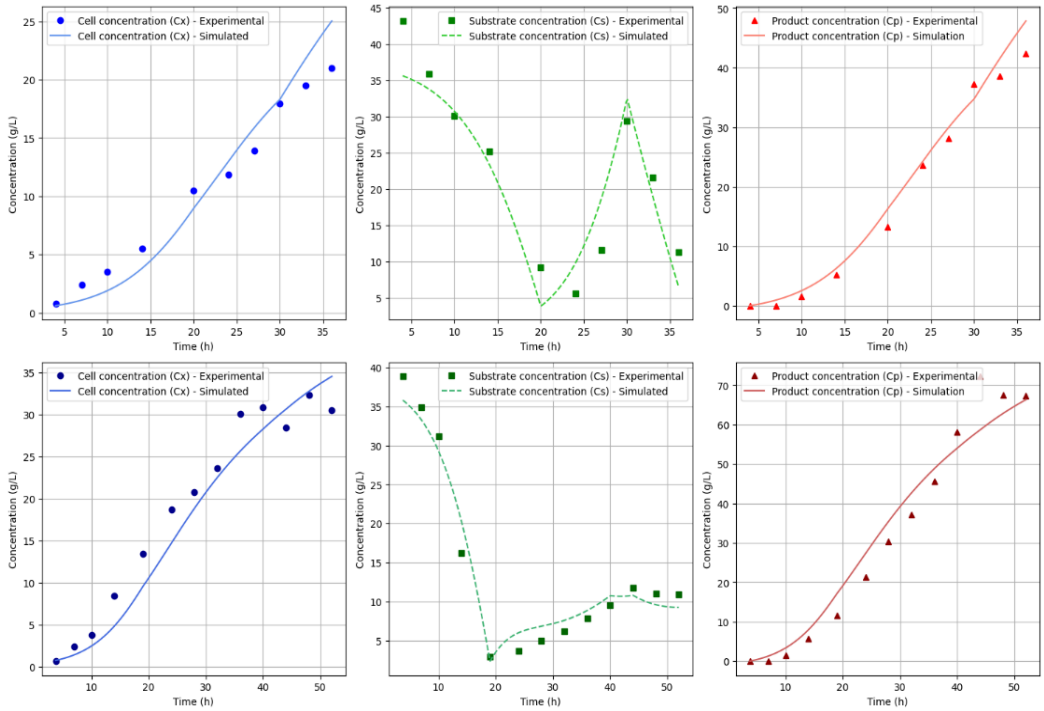
Model	Monod	Contois	Monod with Levenspiel inhibition	Monod with exponential inhibition
Sum of squared residuals	1987.31	1467.19	950.08	1864.08717

Once the model was chosen, a multi-start optimization was performed to avoid local optima using Latin Hypercube Sampling. The kinetic parameters are shown in Table 2, and the best fitting result is illustrated in Figure 1. It is possible to observe that the model accurately tracks the variation in concentrations even during the feeding stage.

Table 2. Kinetic parameters estimated using both experiments of fed-batch with *E. coli* PSO932 strain.

μ_{max}	K_S	Y_{PS}	Y_{XS}	P_m	n
0.19718258	0.11281338	0.51966683	0.26668491	175.598296	4.83099063

Figure 1. Experimental data and model representation for both experiments of fed-batch with *E. coli* PSO932 strain. The first line of graphics refers to the data and simulation for the experiment with exponential feed data and the second line refers to the experiment with linear feed data.



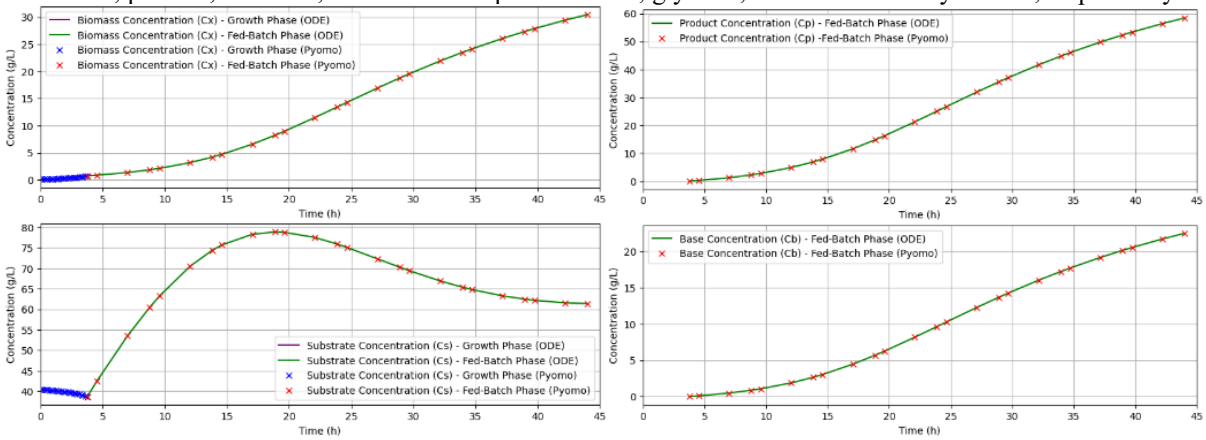
3.2. *Analysis of minimum number of finite elements*

The number of finite elements was chosen, as described in the methodology, in order to reach a maximum absolute error below 10^{-2}. Model refinement stopped with 8 finite elements points, the maximum absolute error for each concentration in the system under this configuration is shown in Table 3 and Figure 2 presents a comparison between the two models.

Table 3. Absolute error for each concentration in the definition of the number of finite elements step.

Biomass conc. (Growth Phase)	Substrate conc. (Growth Phase)	Product conc. (Batch Phase)	Biomass conc. (Batch Phase)	Substrate conc. (Batch Phase)
g/L				
$4x10^{-6}$	$1x10^{-5}$	$2.9x10^{-3}$	$1.5x10^{-3}$	$6.2x10^{-3}$

Figure 2. Comparison between results from Pyomo and solve_ivp simulations. Biomass represents E. coli PSO932; product, substrate, and base correspond to 3-HP, glycerol, and ammonium hydroxide, respectively.

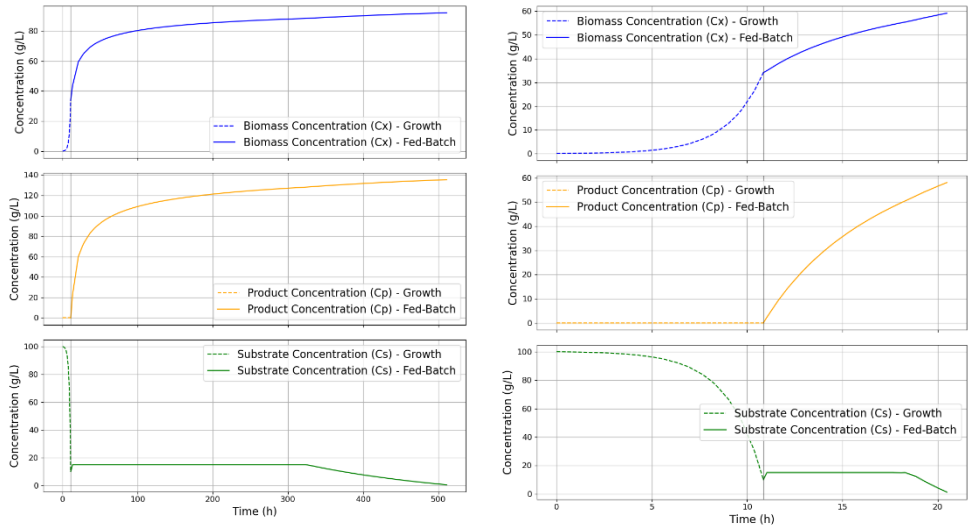


3.3. *Refining number of finite elements for optimization*

Finally, the number of finite elements for optimization was refined. Elements were iteratively doubled until the objective function variation dropped below 10^{-2}, resulting in 32 points for titer optimization and 16 for productivity. Figures 3 and 4 show the feeding and concentration profiles for both optimization cases. In these

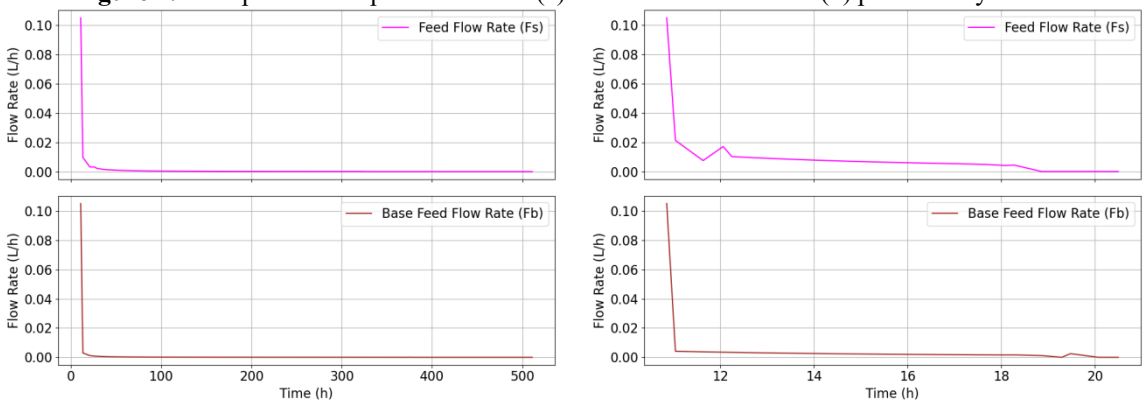
figures, the start of the solid line marks the induction point, separating the growth phase (simple batch without 3-HP production) from the fed-batch phase (with 3-HP production).

Figure 3. Concentrations profiles in optimization. (a) maximization of titer. (b) maximization of productivity.



(a) (b)

Figure 4. Feed profiles in optimization for (a) titer maximization and (b) productivity maximization.



(a) (b)

In both processes, the growth phase exhibited the same profile: a duration of approximately 11 h, reaching cell concentration of 34.12 g/L. This represents a longer phase (2.8-2.9x greater) and a higher cell concentration (9.8-10.6x greater) compared to the results reported by Batista et al. (2024). Additionally, the initial substrate concentration was set by the optimization to its upper limit. During the fed-batch phase, both protocols also showed similarities by maintaining the substrate concentration at the upper bound of 15 g/L to avoid metabolic pathways deviations, which typically occur above 10 g/L. While it is intuitive that higher cell concentrations improve productivity, the tool suggests that this strategy was also beneficial for titer optimization; these results will be further investigated in future studies. Consequently, the tool indicates that the primary difference between optimizing for productivity versus titer lies in the process duration itself.

A key observation is that the titer optimization results in a process that is impractical for laboratory testing. The lack of a time restriction leads to a 511-hour process with 11-hour growth phase, reaching a concentration of 135.39 g/L and a productivity of 0.27 g/L/h. Conversely, productivity optimization resulted in a

20-hour process, which also included an 11-hour growth phase. The final concentration was 58.09 g/L with a productivity of 2.77 g/L/h, which is 1.68 times higher than the productivity achieved by Batista et al. (2024).

In both scenarios, the substrate concentration reached its upper bound during the fed-batch phase. This constraint was introduced due to the observed decrease in cells' capacity to produce 3-HP at higher substrate concentrations; this phenomenon will be investigated in future experiments. Furthermore, although the optimizations led to rapid increases and decreases in feed flow, this profile could be filtered and smoothed for future experiments.

4. Conclusion

This study successfully developed a Python-based computational framework to model and optimize the fed-batch production of 3-HP by *Escherichia coli*. Best kinetic description was obtained by Monod-Levenspiel model indicating a possible product inhibition. Dynamic optimization highlighted a fundamental operational trade-off: prioritizing maximum titer demands a significantly extended batch time (511 hours) resulting in a titer of 135 g/L, whereas maximizing productivity necessitates a much shorter process (20 hours) with a productivity 1.68 times higher than the experimental results. These insights demonstrate the framework's capability and emphasize the need to incorporate metabolic limitations before experimental implementation.

Acknowledgements: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 through process 88887.948726/2024-00.

References

Batista, R. S., Chaves, G. L., Oliveira, D. B., Pantaleão, V. L., Neves, J. D. D. S., & Da Silva, A. J. (2024). Glycerol as substrate and NADP+-dependent glyceraldehyde-3-phosphate dehydrogenase enable higher production of 3-hydroxypropionic acid through the β-alanine pathway in E. coli. Bioresource Technology, 393, 130142. https://doi.org/10.1016/j.biortech.2023.130142

Biegler, L. T. (com Society for Industrial and Applied Mathematics). (2010). Nonlinear programming: Concepts, algorithms, and applications to chemical processes. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Chemarin, F., Moussa, M., Allais, F., Trelea, I. C., & Athès, V. (2019). Recovery of 3-hydroxypropionic acid from organic phases after reactive extraction with amines in an alcohol-type solvent. Separation and Purification Technology, 219, 260–267. https://doi.org/10.1016/j.seppur.2019.02.026

Konzock, O., & Nielsen, J. (2024). TRYing to evaluate production costs in microbial biotechnology. Trends in Biotechnology, 42(11), 1339–1347. https://doi.org/10.1016/j.tibtech.2024.04.007

Nielsen, J., Tillegreen, C. B., & Petranovic, D. (2022). Innovation trends in industrial biotechnology. Trends in Biotechnology, 40(10), Artigo 10. https://doi.org/10.1016/j.tibtech.2022.03.007

Rumble, J. R. (Org.). (2021). CRC handbook of chemistry and physics (102nd edition 2021-2022). CRC Press.

Sandia National Laboratories. (s. d.). Dynamic Optimization with pyomo.DAE — Pyomo 6.8.0 documentation. Recuperado 24 de março de 2026, de https://pyomo.readthedocs.io/en/6.8.0/modeling_extensions/dae.html#applying-multiple-discretization-transformations

Szklo, A., Schaeffer, R., & Delgado, F. (2007). Can one say ethanol is a real threat to gasoline? Energy Policy, 35(11), 5411–5421. https://doi.org/10.1016/j.enpol.2007.07.017

The SciPy community. (s. d.). solve_ivp—SciPy v1.17.0 Manual. Recuperado 24 de março de 2026, de https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html

Wang, X., Cui, Z., Sun, X., Wang, Z., & Chen, T. (2023). Production of 3-Hydroxypropionic Acid from Renewable Substrates by Metabolically Engineered Microorganisms: A Review. Molecules, 28(4), 1888. https://doi.org/10.3390/molecules28041888

Werpy, T., & Petersen, G. (2004). Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas (DOE/GO-102004-1992, 15008859; Número DOE/GO-102004-1992, 15008859, p. DOE/GO-102004-1992, 15008859). https://doi.org/10.2172/15008859