Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

A mechanism-assisted data-driven model to improve the efficiency of sophorolipids by Candida bombicola

Yang Chen[1], Xiwei Tian[1], Xiaofeng Zhu[1], Xu Tang[1], Ying-Ping Zhuang[1], Song Zhang[2], Shiyuan Hu[3], and Ju Chu[4]

[1]East China University of Science and Technology
[2]Evonik (Shanghai) Investment Management Co., Ltd
[3]Evonik (Shanghai) Investment Management Co., Ltd.
[4]State Key Laboratory of Bioreactor Engineering

April 16, 2024

Abstract

In this study, we developed a mechanism-assisted data-driven model to regulate substrate feedback to improve the production efficiency of sophorolipids (SLs). First, we used a variety of on-line biosensors to establish a multi-scale parameter detection system. We found that the production of SLs by fed-batch fermentation could be divided into three stages: a stage that was limited by cell production capacity, a stage that was inhibited by high product concentration, and a stage that was limited by oxygen supply. Subsequently, we used process parameters to develop a data-driven model, and this was then combined with the analysis of cell metabolic mechanisms. The optimal production of SLs was achieved in the first and second stages by the precise feedback regulation of substrate feeding, which increased the titer of SLs by 4.9%. The control error of the substrate was reduced from more than 15% to less than 5%. The mechanism-assisted data-driven model was then applied for semi-continuous fermentation during the production of SLs. This effectively alleviated the oxygen limitation during the third stage, and further increased the productivity of SLs to 2.30 g/L/h, 40.2% higher than the fed-batch fermentation method.

Introduction

Sophorolipids (SLs) as renewable glycolipid biosurfactants are mainly produced by microorganisms of the *Saccharomyces* ssp. (Van Bogaert et al., 2007). Glucose and fatty acids are the main precursors for the synthesis of SLs. First, fatty acids are catalyzed by P450 monooxygenase into hydroxy fatty acids, and then UDP-glucoses are sequentially connected to form non-acetylated acidic SLs *via* the action of glucose transferase I and II (Lodens et al., 2020). Finally, the functional activities of acetyltransferase and extracellular lactonase lead to the acetylation and lactonization of SLs. Up to 20 different structural forms of SLs are known to exist, and this variation is the result of differences in acetylation, hydroxyl position, the length of the fatty acid chain, and the unsaturation degree of fatty acids (Hu & Ju, 2001). Thus, the production of SLs is a complex and multiphase fermentation process, involving a gas phase (air), a solid phase (cells, SLs crystals), a hydrophilic liquid phase (acidic SLs, glucose) and a hydrophobic liquid phase (lactonic SLs, oil), which pose a significant challenge to the efficient and stable production of SLs (Tian, Li, Chen, Mohsin & Chu, 2021).

At present, the optimization of microbial fermentation can be divided into three different aspects: (1) obtain high-performance producers *via* mutagenic breeding or genetic engineering, (2) develop and utilize cheap substrates to reduce the costs of fermentation , and (3) optimize the fermentation process to achieve the efficient synthesis of product (Dolman, Kaisermann, Martin & Winterburn, 2017; Li, Chen, Tian & Chu, 2020; Tian et al., 2021; Wang et al., 2020a). Of these processes, the rational and precise regulation of

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

fermentation remains the most significant factor in achieving high-efficiency production. The identification of key process parameters form the basis of process control and optimization. By regulating key process parameters, it is possible to achieve the regulation of cell metabolism in a flexible manner, this allowing high titer, productivity, and yield (Wang et al., 2020c). The continuous development of sensing detection and information processing technologies, along with the real-time detection of conventional environmental parameters by on-line sensors, cellular macro-physiological, and metabolic parameters, has led to the availability of key parameters on-line, including living cell amount, oxygen uptake rate (OUR), carbon dioxide evolution rate (CER), and respiratory quotient (RQ). This information creates a database for the fermentation process (Chen, Lin, Tian, Li & Chu, 2019; Feng et al., 2021).

However, the mining of sensitive process parameters still relies upon correlation analysis and manual experience and has yet to be studied from the perspective of big data analysis (Davila, Marchal & Vandecasteele, 1997). On the other hand, it is gratifying that various data processing and analytical methods are gradually being introduced into the fermentation process. For example, linear and non-linear algorithms, neural networks, support vector machines, and other mathematical models, can quickly process many on-line and off-line parameters, and can therefore be correlated with regulatory processes during the fermentation process (Safarian, Saryazdi, Unnthorsson & Richter, 2021; Zhang et al., 2020). Overall, the current options for regulating the fermentation process arise predominantly from the process control perspective and lack rationality (Kim, Yun & Kim, 2009). Other options include guidance provided by cellular metabolic characteristics, but this approach lacks universal applicability.

In terms of SLs fermentation, the supplementation of glucose and oil substrates is essential for the synthesis of SLs. During the late fermentation stage, and with the gradual accumulation of SLs, the rheological properties of the fermentation broth undergoes significant changes, thus increasing viscosity. These changes exert a key impact on mass transfer and mixing, thus resulting in a limited oxygen supply, consequently, the productivity of SLs synthesis decreases notably. The development of a semi-continuous fermentation process could significantly alleviate the influence of oxygen limitation on SLs synthesis (Zhang et al., 2018). It has been found that controlling the content of oil, and the ratio of oil to SLs, can exert influence on the morphology of SLs (crystalline or non-crystalline types), thereby significantly changing the sedimentation characteristics of SLs (Chen et al., 2021b). In turn, this affects the efficiency of semi-continuous fermentation. Therefore, it is vital that we are able to precisely control the process of oil supplementation so that we can achieve the efficient production of SLs.

In this study, we established a multi-scale parameter detection system for the SLs fermentation process by applying a range of on-line sensors, mainly including a near-infrared spectrometer and a process mass spectrometer. First, we studied the differences in macro-physiological and metabolic parameters under different rates of oil supplementation. Subsequently, we used a range of process parameters to construct a data-mechanism fusion model, which was accomplished by integrating data modeling with cellular metabolic mechanisms, for feedback feeding of oil and glucose. Finally, this model was applied to semi-continuous fermentation to achieve a highly efficient production system for SLs.

2 Materials and methods

2.1 Strain, media, and culture conditions

Candida bombicola ATCC 22214 was obtained from the Guangdong Culture Collection Center (China) and stored at -80 in 20% glycerol solution. Seed medium contained 50 g/L of glucose, 1 g/L of KH_2PO_4, 4 g/L of $(NH_4)_2SO_4$, 0.5 g/L of $MgSO_4{\cdot}7H_2O$, and 10 g/L of corn steep liquor (CSL). Seed was cultured in a 1 L baffled shake flask with 200 mL working volume at 200 rpm and 25 for 48 h.

The initial fermentation medium was placed in a 5 L bioreactor (Shanghai Guoqiang Bioengineering Equipment Co., Ltd., China) and consisted of 100 g/L of glucose, 1 g/L of KH_2PO_4, 4 g/L of $(NH_4)_2SO_4$, 0.5 g/L of $MgSO_4{*}7H_2O$, and 10 g/L of CSL. All culture media were sterilized at 115 for 30 min. An initial volume of 2.5 L was mixed with 3% of inoculum and cultured at 25 for 168 h. Aeration was provided at 0.5 vvm and dissolved oxygen (DO) was maintained above 30% of the saturation concentration by adjusting

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

agitation in a stepwise manner. A pH of 3.5 was maintained during the entire process by the addition of 4 M NaOH solution. The fed-batch fermentation cycle lasted 168 h. During the fed-batch fermentation process, the feeding rate for rapeseed oil was 1 g/L/h in the first 24 h, and was then set to 2.0 (Low), 2.6 (Medium), and 3.2 (High) g/L/h. During the late stage of fermentation, the residual oil concentration of the fermentation broth was controlled so that it did not exceed 30 g/L. During semi-continuous fermentation, the concentration of glucose and rapeseed oil was controlled at 40 g/L and 10 g/L, respectively, and the glucose concentration was maintained below 30 g/L by controlling the feeding rate prior to *in-situ* separation. *In-situ*separation was carried out when the concentration of SLs in the fermenter exceeded 140 g/L (non-normalized), as determined by the on-line sensor.

2.2 Determination of fermentation process parameters

During the fermentation process, the concentration of SLs, rapeseed oil, and glucose, were determined by using a real-time on-line detection platform involving a near-infrared spectrum (Chen et al., 2021a). The relationship between off-line data and near-infrared spectrum data was established during the early stages to achieve the real-time detection of substance concentration in the fermentation broth. OUR, CER, and RQ were calculated on-line through detecting the oxygen and carbon dioxide proportion in the inlet air and off-gas by a process mass spectrometer as described previously by Chen et al. (2019). For dry cell weight (DCW) determination, 2 mL of fermentation broth was sampled, washed three times with an equal volume of 70% ethanol (v/v) solution, and then dried in an oven at 80degC. Next, the sample was weighed. SLs structure was determined by LC-MS as described previously by Chen et al. (2020).

Extracellular organic acids were analyzed by the HPLC method. A 2 mL sample of fermentation broth was taken at 24 h, 48 h, 72 h, and 96 h. The samples were then centrifuged to isolate the supernatant. The supernatant was filtered into a liquid phase vial with a 0.22 μm filter and then analyzed by HPLC, including a VARIAN Metacarb-H plus chromatographic column, a RID detector, 0.01 mol/L dilute sulfuric acid as the mobile phase, a flow rate of 0.4 mL/min, an injection volume of 10 μL, a column temperature of 50°C, and a detection temperature of 35°C.

The composition of rapeseed oil was determined by the GC method, as described previously by Gao, Liu, Jin & Wang (2019).

2.3 The influence of SLs concentration on SLs synthesis

First, we added 200 mL of fermentation broth (the concentration of the mixture of glucose and rapeseed oil was 50 g/L) into a 1 L baffled shake flask. Next, we added SLs to the shake flask to prepare the following concentrations: 0, 15, 30, 45, 60, 75, 90, and 105 g/L. Then, we took 50 mL of the fermentation broth and cultured this sample in a 5 L fermenter for 30 h; this was followed by centrifugation to remove the supernatant. After washing, the cells were resuspended in 20 mL of sterile water and inoculated in a shake flask. This was then cultured for 24 h at 200 rpm and 25°C.

2.4 Feedback feeding model

The feedback feeding model achieved stable, real-time, and rational regulation of fermentation process feeding based on on-line parameters and fermentation control. This model was mainly divided into an on-line parameter acquisition module, a parameter analysis module, and a feedback feeding module. First, we established a real-time detection system for fermentation process parameters (independent variables: OUR, CER, RQ, pH, DO, temperature, agitation, aeration, and DCW). This was achieved by a variety of on-line sensors (near infrared spectroscopy, a process mass spectrometer, a dissolved oxygen electrode, temperture electrode and pH electrode). We also monitored a range of dependent variables, including SLs productivity, glucose consumption rate, and rapeseed oil consumption rate. Then, we used six algorithms to fit these key parameters, including multiple linear regression, partial least squares, a support vector machine, random forest, and gradient boosting regression. The correlations between parameters (R^2) were then used to construct a data correlation model. Next, we used the data correlation model to select the optimal feedback value for feed output in accordance with the on-line parameters and fermentation control. Finally, we used the

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

output correlation (R^2) for multiple algorithm models to achieve real-time control over the feed module. The model could read a range of detection data (SLs productivity, fermenter volume, DCW, residual oil concentration, residual glucose concentration, OUR, and other parameters) every 1 h. The control of oil and glucose concentrations enabled real-time feedback and regulation of oil and glucose supplementation. During the verification process, fed-batch fermentation was performed to control the residual oil concentration of the fermentation broth to 2 g/L and 10 g/L; the glucose concentration was controlled at 40 g/L. During semi-continuous fermentation, the residual oil concentration was controlled at 10 -15 g/L, and the glucose concentration was controlled at 30-40 g/L. We used gravity sedimentation separation and washing recovery to achieve *in-situ* separation and fermentation for 234 h. The fermentation control platform and feedback feeding model are shown in Fig. 1.

2.5 Data analysis

Due to the consideration of changes in working volume during feeding, all of the data presented in the figures and tables were normalized to the initial volume (except those marked 'non-normalized in part'), and represent the real-time detection of SLs concentration in the fermenter. All experiments were performed in triplicate. The data shown in the tables and figures represent the mean ± standard deviation of three experiments. One-way analysis of variance (ANOVA) and the T-test ($P < 0.05$) were used to identify significant differences between treatments (GraphPad Prism 8.3.0, GraphPad Software Inc., USA).

3 Results

3.1 The effect of oil feeding rate on SLs production during fed-batch fermentation

During 5 L fed-batch fermentations, SLs were synthesized with different rapeseed oil feeding rates (high, medium, and low modes) for 168 h. There were no significant differences between the three modes with respect to the titer of SLs, total glucose consumption, and total oil consumption (Fig. 2A, Table 1), however, the phased synthesis rate of SLs differed substantially during the fermentation process. The higher the oil feeding rate, the faster the accumulation of SLs in the early stages of fermentation, although the accumulation rate decreased during the middle and late stages. We also found that the phased productivity of SLs differed significantly between the three modes (Fig 2B). The highest productivity of SLs was 3.53 g/L/h under the high mode, while the highest productivity in the medium and low modes was 2.63 and 2.09 g/L/h, thus representing reductions of 25.5% and 40.8%, respectively. Notably, after 48 h, the productivity of SLs declined rapidly under the high mode until 110 h and was then maintained at 0.98 g/L/h to the end of the fermentation process. In contrast, the productivity of SLs under the low mode was maintained at an average of 2.04 g/L/h until 120 h and then began to decrease. As shown in Fig 2C, the difference between the high and low modes reached a maximum of 139.7 g at around 87 h, while the maximum difference between the high and medium modes was 85.7 g at 73 h. Otherwise, there were significant differences in the OUR and CER profiles when compared between the three modes during the fermentation process. In high mode, the maximum OUR reached up to 64 mmol/L/h, gradually decreased after 48 h, and was then maintained at 47 mmol/L/h after 110 h. In contrast, the OUR in low mode was maintained at approximately 54 mmol/L/h for 120 h. The CER showed a similar trend to OUR (Fig 2D).

3.2 Cellular metabolic characteristics in different stages of SLs production

The fed-batch fermentation process underlying the production of SLs predominantly consisted of two phases: the primary metabolism of cell growth and the secondary metabolism of SLs synthesis. We found that the maximum productivity of SLs was related to the oil feeding rate and that the highest value was 3.53 g/L/h. However, a further increase in the oil feeding rate did not enhance SLs productivity, meaning that the maximum specific productivity for SLs reached 0.15 $g/g_{DCW}/h$ due to limitations imposed by the production capacity of the cells. Moreover, we observed that the production of SLs in the shake flask began to decrease significantly (by 15.6%) when the concentration of SLs exceeded 75 g/L (non-normalized) (Fig. 3A); at this point, the productivity of SLs in the fermenter under the three modes also gradually decreased, thus inferring that product inhibition may have occurred. Furthermore, the degree of inhibition increased with an increase in the concentration of SLs. Interestingly, although the OUR and CER decreased during this phase,

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

the RQ remained unchanged, thus showing that product inhibition only limited the metabolic capacity of the cells, while the metabolic pathway remained unchanged. Notably, when the concentration of SLs in the fermentation broth exceeded 140 g/L (non-normalized), although the OUR and CER continued to decrease with the decline of SLs productivity, the RQ began to increase from approximately 0.70 to 0.76 (Fig. 3B). This result demonstrated that the metabolic pathways of cells underwent significant changes at this time. By further analyzing the viscosity of the fermentation broth, we found that the viscosity of the fermentation broth would rise sharply during the middle and late phases of fermentation, and when the concentration of SLs exceeded 140 g/L (non-normalized), the viscosity exceeded 20 cP (Fig. 3C).

The fermentation of SLs is a high oxygen consumption process. High viscosity can exert serious effects on oxygen supply, thus resulting in alterations in cell metabolism and the synthesis of SLs. Further analysis of the RQ profiles demonstrated that when the average RQ was approximately 0.70 (high mode: 24-96 h; medium mode: 24-110 h; low mode: 24-120 h), the average yields of SLs to glucose and rapeseed oil were 1.21 g/g and 1.39 g/g, respectively. However, when the concentration of SLs exceeded 140 g/L (non-normalized) (high mode: after 96 h; medium mode: after 110 h; low mode: after 120 h), the average yields of SLs to glucose and rapeseed oil changed to 0.82 g/g and 1.46 g/g, respectively, and the average RQ reached 0.76 (Table 2). This implied that once the oxygen supply became limited, the yield of SLs to glucose was reduced, while the yield to rapeseed oil was increased, thus leading to an increase in RQ value.

In summary, the synthesis of SLs during fed-batch fermentation could be divided into three phases (Fig. 4). During the initial phase, the synthesis of SLs was mainly limited by the production capacity of the cells as well as the oil supply. The maximum specific SLs productivity could reach up to 0.15 $g/g_{DCW}/h$ but was affected by the oil feeding rate. During the second phase, when the concentration of SLs in the fermentation broth exceeded 75 g/L (non-normalized), the production capacity of the cells began to be inhibited by high product concentration. The higher the concentration of SLs, the stronger the degree of inhibition. However, cellular metabolic pathways remained unchanged during this phase. During the last phase, the concentration of SLs reached 140 g/L (non-normalized) and the production capacity of the cells became significantly restricted by oxygen supply. On the basis of these analyses, the first and second phases were considered to represent the main stages of SLs accumulation during fed-batch fermentation. Increasing the productivity of SLs as much as possible during these phases *via* the fine control of substrate feeding strategy was of great significance to the improvement of production efficiency. In terms of the third phase, although it was difficult to improve production efficiency without increasing oxygen supply during the conventional fermentation mode, the semi-continuous fermentation mode, involving the*in-situ* separation of SLs, which was regulated by the concentration of oil in the fermentation broth (Chen et al., 2021b), could effectively alleviate product inhibition and significantly reduce the viscosity of the fermentation broth to eliminate oxygen limitation (Liu et al., 2019).

3.3 Construction of an intelligent feedback feeding model based on multi-parameter process data

Next, we constructed and optimized an efficient, stable, and rational fermentation process for SLs production by regulating feeding during the first and second phases. First, we established correlations between the process macroscopic parameters and the production capacity of cells. All the on-line parameters, including agitation, temperature, pH, DO, glucose concentration, oil concentration, SLs concentration, OUR, CER, and RQ were applied as independent variables. The oil consumption rate, glucose consumption rate, and SLs productivity were set as dependent variables. Then, six linear and non-linear mathematical equations were adopted to establish the correlation between independent and dependent variables, including unary linear regression equations (LR), multiple linear regression equations (MLR), support vector machines (SVMs), partial least squares (PLS), random forest algorithms (RFs) and gradient boosting regression algorithms (GBRs). Finally, the pros and cons of the mathematical equations were compared by the correlation coefficient (R^2) between the predicted values and the real values (Fig. 5). Approximately 110 sets of data were used as the training set, and 20 sets of data were used as the testing set. Three dependent variables relating to SLs (productivity, oil consumption rate, and glucose consumption rate) were all strongly correlated with the actual values when applying the six equations ($R^2 > 0.96$), thus showing that these six algorithms accurately reflected the

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

relationship between the consumption of substrates and the synthesis of product and fermentation process parameters (Table 3). Notably, the R^2 for the RF and GBR algorithms decreased slightly after the testing set was added, while the R^2 for the other four equations increased as the number of parameters increased. It is possible that with the continuous input of parameters, the database of equations also increased; consequently, the accuracy of the predicted value would also be improved. Non-linear equations might be associated with over-fitting problems; therefore, when there is less data, the predicted values of ULR, MLR, SVM, and PLS, were more reliable. In contrast, with a continuous increase in sample number, the prediction values for RF and GBR became more accurate.

When the model produced a predicted value (SLs productivity, glucose consumption rate, and oil consumption rate) with an optimal R^2, then the model would add data to the database, thus increasing the data sample. This feedback feeding model was combined with the data provided by on-line sensors. This meant that the model could read the on-line data, perform equation fitting, and output the optimal predicted value by the computer. By integrating the working volume, the required concentrations of glucose and oil, as well as the amounts of consumed glucose and oil, the model could automatically guide the precise feeding rate *via* a computer-controlled feedback module. Simultaneously, the model performed data overlay and self-learning based on the continuous accumulation of data, thereby further improving the accuracy of the model.

The feedback feeding model was validated in a 5 L fermenter. As shown in Fig. 6A and B, the concentration of residual oil was well controlled at 2 g/L and 10 g/L respectively, and the concentration of residual glucose was maintained at 40 g/L from 24 h until the RQ was increased to end the fermentation process. In addition, by applying the feedback feeding model, the SLs titer with a residual oil concentration at 2 g/L was improved by 4.9% during the first and second phases of fed-batch fermentation, when compared to the high oil feeding mode. More importantly, the feedback feeding model shortened the feeding interval from 6 to 12 h by conventional manual intervention to 1 h by automatic feeding. The model also reduced the error related to feeding control from more than 15% to less than 5%, thus providing significant savings in terms of labor and time costs.

3.4 An intelligent feedback feeding model facilitated semi-continuous fermentation for the efficient production of SLs

To alleviate oxygen limitation in the late phase of fed-batch fermentation, we used gravity sedimentation for the *in-situ* separation of SLs, thus permitting semi-continuous fermentation. When the concentration of SLs reached 140 g/L (non-normalized), we began the *in-situ* separation process and allowed this to proceed until the concentration of SLs fell below 60 g/L, thus alleviating the influence of a high SLs concentration on the rheological properties, as well as relieving product inhibition. As illustrated in Fig. 7A, the feedback feeding model could stably control the concentrations of residual glucose and residual oil at appropriate levels during the entire fermentation process, so as to ensure the efficient in-situ separation of SLs (Chen et al., 2021b). After two rounds of *in-situ* separation of product, the total production of SLs reached 1348.3 g within 234 h. Moreover, the overall SLs productivity and yield were 2.30 g/L/h and 0.57 g/g, respectively, thus representing increases of 40.2% and 18.7% in comparison to the fed-batch fermentation with a high oil feeding mode (Fig. 7B). Notably, the productivity of SLs did not exhibit a downwards trend at the end of semi-continuous fermentation (Fig. 7C). Therefore, it was reasonable to speculate that the production efficiency could be further improved by performing longer operations with the assistance of an intelligent feedback feeding model.

4. Discussion

Traditional fermentation process regulation is based on manual experience. The analysis of relevant parameters can identify connections between sensitive parameters and key indicators (Chen et al., 2013; Zhang et al., 2014). With the development of sensor technology, a variety of advanced on-line sensors have been applied in the fermentation process to achieve multi-dimensional monitoring of cell metabolism, thus making the process control more rational and reliable. These sensors have included process mass spectrometry, near-infrared spectroscopy, raman spectroscopy, and viable cell biosensors (Chen et al., 2021a; Iversen, Berg &

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Ahring, 2014). In the present study, we developed an on-line monitoring platform for the fermentation of SLs for the first time. This platform allowed the simultaneous real-time detection of substrates, products, and cellular metabolic states, thus creating a significant amount of process data. We found that the production of SLs under different oil feeding strategies during the fed-batch fermentation could be predominantly divided into three stages: the first stage was limited by cell production capacity; this stage was associated with a relatively higher specific productivity of SLs (up to 0.15 $g/g_{DCW}/h$). Then, the cells entered a stage that was inhibited by high product concentration when the concentration of SLs reached 75 g/L (non-normalized); at this point, the SLs productivity decreased, whereas the RQ value stabilized at 0.70. Finally, when the concentration of SLs exceeded 140 g/L (non-normalized), the cells entered a third stage which was limited by oxygen supply. During this stage, the SLs productivity continued to decline, and the RQ began to rise. Thus, RQ is a key parameter to characterize the difference of substrate metabolism in SLs fermentation and can be adopted as a potential control parameter to guide process optimization.

The method of mining key parameters from a huge amount of process data by manual analysis is both time-consuming and laborious. With recent advancements in data science technology, the regulation of fermentation based on mathematical algorithms is gradually being applied to realize the processing of many data samples, to identify the relationship between multiple variables and build association models (Zhu et al., 2021). Currently, mathematical algorithms predominantly include linear and non-linear algorithms, such as unary linear models, neural network models, and support vector machines. A data model combining a large number of parameters and mathematical algorithms could help us to predict and regulate the fermentation process, thereby significantly reducing labor costs (Lopez et al., 2013). In a previous study, Lu et al. (2016) established a continuous fermentation control system for feedback that adjusted the glucose supplementation rate during the fermentation process of sodium gluconate by correlating the changing trend between OUR and DO, thus significantly improving the fermentation efficiency of sodium gluconate. In another study, Wang et al. (2020b) proposed a non-linear predictive method for the real-time control of product concentration during L-lysine fermentation to improve the efficiency. Data showed that predictive control, as based on Grey-Wolf Optimization, led to better levels of prediction accuracy, adaptability, real-time tracking ability, overall error, and control accuracy. Herein, six common mathematical algorithms, including LR, MLR, SVMs, PLS, RFs and GBRs, have been used to establish a parameter correlation model for the first and second stages of SLs synthesis, as based on the productivity of SLs and using the consumption of glucose and rapeseed oil as dependent variables. The integrated application of multiple mathematical algorithms was able to achieve the real-time and rational control of process feeding based on fermentation process parameters. Furthermore, the model could accomplish self-optimization under continuous data iteration. However, it was not possible for the model to fully understand the physiological state of the cells.

On the other hand, an intracellular metabolic flux model was established for cells to interpret the mechanism underlying the data model. The processes involving the two substrates (glucose and rapeseed oil) included four main pathways: glucose oxidation, β-oxidation of fatty acids, the synthesis of SLs, and the synthesis of extracellular by-products. Through structure and component analyses, the main proportion of the lactone-form of SLs reached 85.0%; of these, the C18:1 lactone-form of SLs accounted for the highest proportion at 58.2%. This may have been because the fatty acid component in rapeseed oil was mainly composed of oleic acid and linoleic acid; the proportion of oleic acid reached 68.5% (Tables S1 and S2). The extracellular organic acids were mainly citric acid (CIT), pyruvic acid (PYR), malic acid (MAL), and succinic acid (SUC); these organic acids were present in only very low concentrations (Fig. S1). The carbon ratio of extracellular organic acids to the added substrates during the synthesis of SLs was only 7.7%.

During the fed-batch fermentation of 24-96 h, we calculated the consumptions of glucose and rapeseed oil, the synthesis of CO_2, SLs, and extracellular organic acids; this allowed us to calculate that the carbon balance at this stage was 95.8% (Table S3). Metabolic flux analysis demonstrated that when the RQ was 0.70, the ratio of glucose and rapeseed oil to SLs was fixed. The synthesis of 1 g of SLs required 0.824 g and 0.719 g of glucose and rapeseed oil, respectively. Furthermore, 66.5% of the carbon sources in the substrate were transferred to SLs, 21.7% to CO_2, and 7.7% to organic acids (Fig. S2). Therefore, there was a strong relationship between the substrate consumption rate, the productivity of SLs and respiration

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

metabolism (carbon dioxide production rate, oxygen consumption rate) during the synthesis stage of SLs. The difference in respiration metabolism is closely related to the change of culture conditions. The multi-parameter integrated control platform could realize precise and intelligent control according to the changes in the culture environment and cellular metabolism. The mechanism-assisted data model was applied for the rational and precise control of residual oil concentration in semi-continuous SLs fermentation. The oxygen limitation problem encountered during the third stage was effectively alleviated, thereby further increasing the SLs productivity to 2.30 g/L/h with an increase of 40.2%.

5. Conclusion

According to the physiological metabolic state of cells, the fed-batch fermentation process for SLs synthesis can be divided into three stages: a stage that is limited by cell production capacity, a stage that is inhibited by high product concentration, and a stage that is limited by oxygen supply. Based on the detection of process multi-parameters, we established a data-driven feedback feeding model. This was combined with analyses of cellular metabolic mechanisms such that an optimal production state was achieved in the first and second stages,in which there was a 4.9% increase in SLs production. Furthermore, the mechanism-assisted data model was applied to the rational and precise control of residual oil concentration during semi-continuous SLs fermentation. The problem associated with oxygen limitation that was encountered during the third stage was effectively alleviated, thereby further increasing the SLs productivity to 2.30 g/L/h. The mechanism-assisted data modeling method developed in this study could be easily extended to other fermentation processes. We also expect our new process strategy for the production of SLs to be applied on an industrial-scale.

Author Contribution Statement

Chen Y: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft & editing, Visualization.

Tian XW: Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Supervision,

Zhu XF, Tang X, Zhang S, Hu SY: Conceptualization, Methodology, Supervision.

Chu J, Zhuang YP: Conceptualization, Methodology, Validation, Writing - review & editing, Supervision, Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Supplementary data

E-supplementary data of this work can be found in online version of the paper.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgement and Funding

This work was financially supported by the National Key Research Development Program of China (2019YFA0904803), the Scientific Research Think Tank of Biological Manufacturing Industry in Qingdao (QDSWZK202004), the Fundamental Research Funds for the Central Universities (JKF01211116, JKVF1211036, JKF01211708). The authors thank Research Center of Analysis and Test of East China University of Science and Technology for the help on the characterization.

References

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Chen, Y., Chen L. L., Guo, M. J., Li, X., Liu, J. S., Liu, X. F., Chen, Z. B., Tian, X. J., Zheng, H. Y., Tian, X. W., Chu, J., & Zhuang, Y. P. (2021a). Application of near-infrared spectroscopy technology in the complex fermentation system to achieve high-efficiency production. Bioresources and Bioprocessing, 8, 96. https://doi.org/10.1186/s40643-021-00452-9

Chen, Y., Lin, Y. M., Tian, X. W., Li, Q. H., & Chu, J. (2019). Real-time dynamic analysis with low-field nuclear magnetic resonance of residual oil and sophorolipids concentrations in the fermentation process of *Starmerella bombicola* . Journal of Microbiological Methods, 157, 9-15. https://doi.org/10.1016/j.mimet.2018.12.007.

Chen, Y., Tang, X., Li, Y., Liu, C., Zhuang, Y. P., Tian, X. W., & Chu, J. (2021b). Ultrasound assisted in situ separation of sophorolipids in multi-phase fermentation system to achieve efficient production by *Candida bombicola* . Biotechnology journal, https://doi.org/10.1002/biot.202100478.

Chen, Y., Tian, X. W., Li, Q. H., Li, Y., Chu, J., Hang, H. F., & Zhuang, Y. P. (2020). Target-site directed rational high-throughput screening system for high sophorolipids production by *Candida bombicola* . Bioresource Technology, 315, 123586. https://doi.org/10.1016/j.biortech.2020.123856.

Chen, Y., Wang, Z. J., Chu, J., Zhuang, Y. P., Zhang, S. L., & Yu, X. G. (2013). Significant decrease of broth viscosity and glucose consumption in erythromycin fermentation by dynamic regulation of ammonium sulfate and phosphate. Bioresource Technology, 134, 173-179. https://doi.org/10.1016/j.biortech.2013.02.023.

Davila, A. M., Marchal, R., & Vandecasteele, J. P. (1997). Sophorose lipid fermentation with differentiated substrate supply for growth and production phases. Applied Microbiology and Biotechnology, 47, 496-501. https://doi.org/10.1007/s002530050962.

Dolman, B. M., Kaisermann, C., Martin, P. J., & Winterburn, J. B. (2017). Integrated sophorolipid production and gravity separation. Process Biochemistry, 54, 162-171. https://doi.org/10.1016/j.procbio.2016.12.021.

Feng, Y., Tian, X. W., Chen, Y., Wang, Z. Y., Xia, J. Y., Qian, J. C., Zhuang, Y. P., & Chu, J. (2021). Real-time and on-line monitoring of ethanol fermentation process by viable cell sensor and electronic nose. Bioresources and Bioprocessing, 8. https://doi.org/10.1186/s40643-021-00391-5.

Iversen, J., Berg, R., & Ahring, B. (2014). Quantitative monitoring of yeast fermentation using Raman spectroscopy. Analytical & Bioanalytical Chemistry, 406, 4911-4919. https://doi.org/10.1007/s00216-014-7897-2

Kim, Y. B., Yun, H. S., & Kim, E. K. (2009). Enhanced sophorolipid production by feeding-rate-controlled fed-batch culture. Bioresource Technology, 100, 6028-6032. https://doi.org/10.1016/j.biortech.2009.06.053.

Gao, P., Liu, R. J., Jin, Q. Z., & Wang, X. G. (2019). Comparative study of chemical compositions and antioxidant capacities of oils obtained from two species of walnut: Juglans regia and Juglans sigillata. Food Chemistry, 279, 279-287. https://doi.org/10.1016/j.foodchem.2018.12.016.

Hu, Y., & Ju, L. (2001) Purification of lactonic sophorolipids by crystallization. Journal of biotechnology, 87, 263-272. https://doi.org/10.1016/S0168-1656(01)00248-6.

Lodens, S., Roelants, S. L. K. W., Ciesielska, K., Geys, R., Derynck, E., Maes, K., Pattyn, F., Van Renterghem, L., Mottet, L., Dierickx, S., Vanhaecke, L., Devreese, B., De Maeaeneire, S. L., & Soetaert, W. (2020). Unraveling and resolving inefficient glucolipid biosurfactants production through quantitative multiomics analyses of *Starmerella bombicola* strains. Biotechnology and Bioengineering, 117, 453-465. https://doi.org/10.1002/bit.27191.

Lopez C, D. C., Barz, T., Peuela, M., Villegas, A., Ochoa, S., & Wozny, G. (2013). Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. Biotechnology Progress, 29, 1064-1082. https://doi.org/10.1002/btpr.1753.

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Li, Y., Chen, Y., Tian, X. W., & Chu, J. (2020). Advances in sophorolipid-producing strain performance improvement and fermentation optimization technology. Applied microbiology and biotechnology, 104, 10325-10337. https://doi.org/10.1007/s00253-020-10964-7.

Liu, Z. P., Tian, X. W., Chen, Y., Lin, Y. M., Mohsin, A., & Chu, J. (2019) Efficient sophorolipids production via a novel in situ separation technology by *Starmerella bombicola* . Process Biochemistry, 81, 1-10. https://doi.org/10.1016/j.procbio.2018.12.005.

Lu, F., Li, C., Wang, Z. J., Zhao, W., Chu, J., Zhuang, Y. P., & Zhang, S. L. (2016). High efficiency cell-recycle continuous sodium gluconate production by *Aspergillus niger* using on-line physiological parameters association analysis to regulate feed rate rationally. Bioresource Technology, 220, 433-441. https://doi.org/10.1016/j.biortech.2016.08.062.

Safarian, S., Saryazdi, S. M. E., Unnthorsson, R., & Richter, C. (2021). Artificial Neural Network Modeling of Bioethanol Production Via Syngas Fermentation. Biophysical Economics and Sustainability, https://doi.org/6. 10.1007/s41247-020-00083-2.

Tian, X. W., Li, Y., Chen, Y., Mohsin, A., & Chu, J. (2021). System optimization of an embedding protocol to immobilize cells from Candida bombicola to improve the efficiency of sophorolipids production. Bioresource Technology, 340, 125700. https://doi.org/10.1016/j.biortech.2021.125700.

Van Bogaert, I. N.A., Saerens, K., De Muynck, C., Develter, D., Soetaert, W., & Vandamme, E. J. (2007). Microbial production and application of sophorolipids. Applied microbiology and biotechnology, 76, 23-34. https://doi.org/10.1007/s00253-007-0988-7.

Wang, B., Shahzad, M., Zhu, X. L., Ur Rehman, K., Ashfaq, M., & Abubakar, M. (2020a). Soft-sensor modeling for l-lysine fermentation process based on hybrid ICS-MLSSVM. Scientific Reports, 10, 1-15. https://doi.org/10.1038/s41598-020-68081-4.

Wang, H. M., Kaur, G., To, M. H., Roelants, S. L. K. W., Patria, R. D., Soetaert, W., & Lin, C. S. K. (2020b). Efficient in-situ separation design for long-term sophorolipids fermentation with high productivity. Journal of Cleaner Production, 246, 118995. https://doi.org/10.1016/j.jclepro.2019.118995.

Wang, Z. J., Zhang, X. Z., Wang, P., Sui, Z. W., Guo, M. J., Zhang, S. L., & Zhuang, Y. P. (2020c). Oxygen Uptake Rate Controlling Strategy Balanced with Oxygen Supply for Improving Coenzyme Q10 Production by *Rhodobacter sphaeroides* . Biotechnology and Bioprocess Engineering, 25, 459-469. https://doi.org/10.1007/s12257-019-0461-3.

Zhang, A. H., Zhu, K. Y., Zhuang, X. Y., Liao, L. X., Huang, S. Y., Yao, C. Y., & Fang, B. S. (2020). A robust soft sensor to monitor 1,3-propanediol fermentation process by Clostridium butyricum based on artificial neural network. Biotechnology and bioengineering, 117, 3345-3355. https://doi.org/10.1002/bit.27507.

Zhang, Q., Chen, Y., Hong, M., Gao, Y., Chu, J., Zhuang, Y. P., & Zhang, S. L. (2014). The dynamic regulation of nitrogen and phosphorus in the early phase of fermentation improves the erythromycin production by recombinant Saccharopolyspora erythraea strain. Bioresources and Bioprocessing, 1, 15. https://doi.org/10.1186/s40643-014-0015-7.

Zhang, Y. G., Jia, D., Sun, W. Q., Yang, X., Zhang, C. B., Zhao, F. L., & Lu, W. Y. (2018). Semicontinuous sophorolipid fermentation using a novel bioreactor with dual ventilation pipes and dual sieve-plates coupled with a novel separation system. Microbial Biotechnology, 3, 455-464. https://doi.org/10.1111/1751-7915.13028.

Zhu, X. F., Mohisn, A., Zaman, W. Q., Liu, Z. B., Wang, Z. J., Yu, Z. H., Tian, X. W., Zhuang, Y. P., Guo, M. J., & Chu, J. (2021). Development of a novel non-invasive quantitative method to monitor *Siraitia grosvenorii* cell growth and browning degree using an integrated computer-aided vision technology and machine learning. Biotechnology and bioengineering, https://doi.org/10.1002/bit.27886.

Figure caption

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Fig. 1. Feedback feeding platform in SLs fermentation process. The solid lines represent material pipelines, dotted lines represent signal pipelines.

Fig. 2. Process parameters of SLs fermentation with different oil feeding modes. A: SLs titer; B: phased SLs productivity; C: difference in total SLs production; D: process CER and OUR profilings

Fig. 3. Metabolic characteristics of SLs at different production stages. A: The effect of the product on the synthesis of SLs; B: Changes in RQ during SLs synthesis; C: Relation between SLs concentration and viscosity.

Fig. 4. Analysis of SLs synthesis stage.

Fig. 5. Intelligent feedback feeding model. OUR for the oxygen uptake rate (mmol/L/h), CER for the CO_2 release rate (mmol/L/h), RQ for the respiratory quotient, RPM for the fermentation speed (rpm), DO for the dissolved oxygen (%), Temp for the fermentation temperature (), pH for the fermentation pH. SLs represents the productivity of SLs (g/L/h), Oil represents the consumption rate of rapeseed Oil (g/L/h), Glc represents the consumption rate of glucose (g/L/h), V represents the working volume (L), c1 and c3 represent the concentrations of rapeseed Oil and glucose in the actual fermentation broth (g/L), c2 and c4 represent the controlling concentrations of rapeseed oil and glucose in fermentation broth (g/L).

Fig. 6. Feedback feeding control. A: Glucose concentration; B: Rapeseed oil concentration.

Fig. 7. Semi-continuous fermentation of SLs production with the assistance of intelligent feedback feed model. A: Rapeseed oil and glucose concentration as well as DCW; B: SLs titer; C: SLs productivity and yield.

Hosted file

`Table.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

`Figure 1.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

`Figure 2.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

`Figure 3.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

`Figure 4.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

`Figure 5.docx` available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-

bombicola

Hosted file

Figure 6.docx available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Hosted file

Figure 7.docx available at https://authorea.com/users/373831/articles/713059-a-mechanism-assisted-data-driven-model-to-improve-the-efficiency-of-sophorolipids-by-candida-bombicola

Posted on 31 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170670053.32139291/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.