

Modelling Bioprocesses Using Component Profiles in Natural Media, Named “Substratome”, Obtained by Performing Comprehensive Non-Targeted Analysis

Masaaki KONISHI^{1*}, Kazuki WATANABE², Seiga TACHIBANA²

¹ Biotechnology and Food Chemistry Course Program, School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, 165 Koen-cho, Kitami, Hokkaido, 090-8507, Japan.

² Department of Biotechnology and Environmental Chemistry, Graduate School of Engineering, Kitami Institute of Technology, 165 Koen-cho, Kitami, Hokkaido, 090-8507, Japan.

Abstract Natural media are often used for various commercial bioprocesses by manufacturers to cut raw material cost. However, the components of the raw materials varies between lot-to-lots and brand-to-brands. The varieties of raw materials influence to the cell growths and materials productivities, and results in unstable production across batches in manufacturing processes. To ensure the quality of raw materials among batches, it is necessary to perform a laboratory screening to purchasing the optimal one, and ensure a desirable performance in industrial process. To solve the serious problems in bioprocesses, it is developing that a modelling methodology using composition of raw materials, named us “substratome”, obtained by non-targeted metabolomics-like methods can estimate the cell growth and bio-productions. Here, we will present that two model studies: [1] *Escherichia coli* growths have been estimated from hydrophilic components in yeast extract obtained by gas chromatography-mass spectrometry (GC-MS), and [2] bioethanol production have been estimated by the volatile components in corncob and corn stover hydrolysates obtained by GC-MS; by partial least square regression (PLS-R). Additionally, we will present preliminary results to solve the same issues by using artificial intelligence.

1 Problems Using Raw Materials in Bioprocesses

Raw materials such as yeast extracts, meat extract, malt extract, peptone, molasses, corn steep liquor, etc. are often used for various commercial bioprocesses, including microbial and mammalian cell processes, for cutting cost. Lignocellulosic biomasses are remarkable raw materials for second generation bioethanol production and biorefinery-conceptual bioproductions. The raw materials composes of various chemicals, and their compositions varies as lot-to-lots and brand-to-brands differences by the cultivation conditions of source materials and manufacturing conditions. The variety of composition often causes the unstable yields of the bioproducts in commercial processes. Indeed, the lot-to-lots difference in yeast extract can lead to up to a 50% variation in the levels of biomass and growth rate of *Escherichia coli* (Portvin *et al.*, 1997). It have been described that different peptones obtained visceral and fish muscle residues promoted growth of lactic acid bacteria when applied individually or in combination (Vázquez *et al.*, 2006). Different quality types of sugar-cane molasses have been effects on the fermentation by *Zymomonas mobilis* (Doelle *et al.*, 1989). Seasonal and geographical would influence on the compositions of almost of biomass

feedstocks. Therefore, quality of raw materials affect the cell growth and productivity. In industrial bioprocesses, the instability of growth and production is a serious issue. Therefore, to assure stable raw material quality across batches, it is necessary to carried out a laboratory screening to prior to use the optimal one. Efforts for modelling the medium compositions to cultivating results have been reported that correlation of recombinant fermentation yield wit yeast extract composition was characterized by near-infrared (NIR) spectroscopy, with models for predicting cell mass in different scale cultivations (Kasprow *et al.*, 1998). However it have been difficult to estimate the culture results from only the NIR spectra.

2 “Substratome” Concept for Bioprocess Engineering

In biology, English neologism omics are often used such as genome, proteome, and metabolome. The omics means a multivalent analysis of genes, proteins, and metabolites, respectively. The development and spread of next generation sequencers and analytical techniques using mass spectrometer accelerated to apply in various fields. Metabolomics approach was applied to food engineering.

* Corresponding author: konishim@kitami-it.ac.jp

For example, the approach have been applied to predict geographical origin of hazelnuts (Klockmann *et al.*, 2016), Japan green tea (Jumtee *et al.*, 2011) and sake (Mimura *et al.*, 2014). The numerous data in omics analysed using statistical analysis between food compositions and evaluation functions. In the issues for the difference of raw material compositions, the composition data of raw materials, named “substratome”, can estimate the enhancing and inhibiting the cultivations, by metabolomics like approaches (Tachibana *et al.*, 2019). We defined the approach as “substratemics”, which can estimate the cultivating results from “substratome” of initial media components (Figure 1).

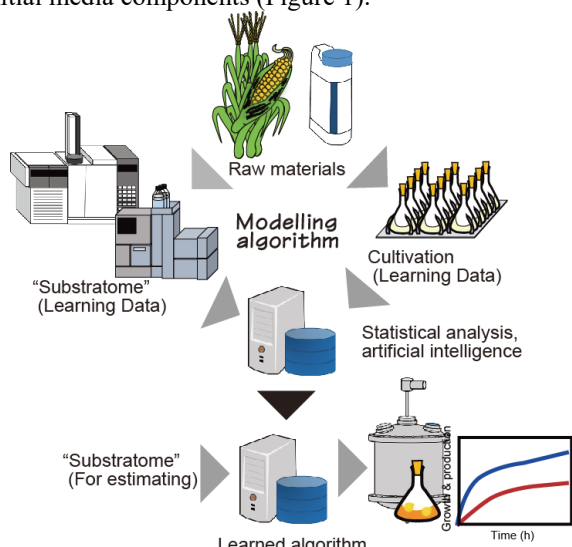


Figure 1. Schematic illustrate to estimate culture results by “substratome”

3 Model Studies for “Substratome” Concept

3.1 Influences of component variation in yeast extract on *Escherichia coli* growth

To confirm the concept, Tachibana *et al.* (2019) demonstrated that the component of different brands and lots of yeast extract (YE-a, YE-b, YE-c, YE-d, YE-e, and YE-f) estimated *E. coli* growth in batch culture in flasks by using gas chromatograph-mass spectrometry (GC-MS) and partial least square regression. *E. coli* NBRC 3301, synonym K12, was used for cultivation. To evaluate the effect of yeast extract in simulating bacterial growth, different concentrations of YE (0.5, 1.0, 2.0, 5.0 g/l) were supplied to M9 broth. (All patterns of media components were 16 patterns) The cultivations were performed at 30°C at 200 rpm. Optical density at 600 nm as index of growth were measured by using a spectrophotometer. Not only the amount of added yeast extracts, but also brands and lots were influence on the growths. In case of control experiment without yeast extract, bacterial growth gradually increased to 2.45 in 24 h. The growth increased with the amount of YE added. All concentrations of yeast extract, except for that in YE-d, resulted in increased cell

growth to a turbidity of more than 9.0. However, the turbidity, in the case of 5.0 g/l YE-d addition, reached only 6.95±0.08 at 24 h. At 18 h, turbidity in the media YE-a, YE-b, and YE-c increased to 10.80±0.28, 10.50±0.21, and 10.86±0.32, respectively. Turbidity in YE-e and YE-f increased to 9.60±0.50 and 9.13±0.16, respectively, whereas that in YE-d was only 6.83±0.09 at 18 h.

To confirm the difference in low-molecular-weight hydrophilic components, GC-MS analysis was performed, where Agilent GC-MS system, 7980B and 5977A MSD, was used for the hydrophilic component analysis. HP-5ms UI (30 m× φ 0.25 mm × i.d. 0.25 μm) column was used. Of the total 165 peaks detected from all samples, 109 peaks were assigned. Of those, 6 sugars, 26 amino acids, 20 glycosides, and 3 fatty acids were estimated, including different degrees of TMS derivatives. The remaining peaks were not classified.

To construct models between the GC-MS results and cultivation results, partial least squares regression were carried out. Good correlations between observed growth and predicted growth were observed in modeling data in all cases. The degree of agreement between the derived model and the data, R^2 , were indicated in the range between 0.986 and 0.997. The best of Q^2 was recorded 0.994 (Figure 2). The values indicate the model were construct with good accuracy.

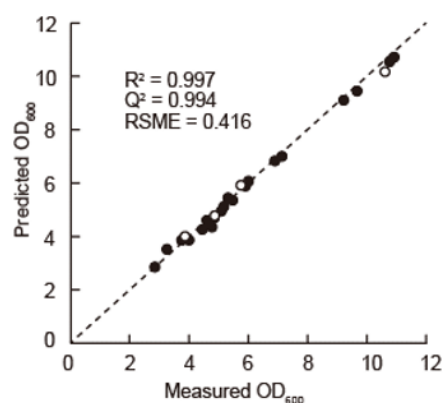


Figure 2. Measured and predicted OD₆₀₀ plots. R^2 and Q^2 indicate validation metrics for learning data and for predicting data, respectively. RMSE: root mean square error. Symbols: closed circles, estimated by learning data; open circles, estimated by test data (Tachibana *et al.*, 2019, Figure 3 modified)

Filtering by correlation coefficient of each compound, important variables that contribute substantially to model construction were calculated. In amino acids, Phe, Gly, Ile, Leu, Val, Asn, Ser, Glu, Asp, Lys, Tyr, 5-oxoproline, and ornithine, were also predicted as growth activators, but Met gave inhibitory effects. In fact, L-Asp, L-Val, L-Glu and L-Try showed tendency of growth stimulation in the range of 1.3 to 8.0%. L-Met obviously inhibited the growth; a 12% decrease in growth was observed in a validating cultivation.

The results demonstrates that substratome obtained TMS-GC-MS can estimate *E. coli* growth, and the significant compounds be estimated from the models.

3.2 Modelling of bioethanol production from inhibitory materials in hydrolysates of lignocellulosic biomass

In order to apply the modeling using “substratome” for bioethanol production. Watanabe *et al.* (2019) demonstrated that volatile compounds in corncobs and corn stover hydrolysates had influences on the yeast growth and ethanol fermentations. Three different corn cobs derived from different region, and corn stover were model raw materials. The materials were hydrolysed by acid-hydrolyzation with different concentration of sulphuric acid and enzyme-assisted saccharification. Further activated carbon treatment performed for removing growth inhibitors. Sixteen different quality of hydrolysate were prepared. Bioethanol production were performed by *Saccharomyces cerevisiae* S288c in the different qualities of hydrolysates. The cell growth and ethanol productions were inhibited in hydrolysates prepared by higher concentrations of sulphuric acid. In hydrolysates prepared 3%, w/v of sulphuric acid, cell growths and fermentations were no longer observed. On the other hand, cell growth and fermentation were enhanced in hydrolysates with activated carbon treatment in spite of the difference of raw materials.

Here, it was considered that inhibitory materials caused with pyrolysis in the hydrolysate preparation, and inhibit the cell growth and fermentation. Fran aldehydes, phenolic compounds, and acids have been known well as growth and fermentation inhibitory materials. Since almost of those are volatile chemicals, volatile compounds profiles in hydrolysates were measured by GC-MS and acetate, and glucose and xylose were measured by a HPLC. 208 chemicals were observed in total ion chromatograms of all hydrolysates. After the varieties of the composition were checked by principal component analysis (PCA), PLS-R analyses were performed using the inhibitory material component profiles as explanatory variables and the turbidity and the ethanol concentration at 72h after inoculation as response variables in model for predicting growth and ethanol fermentation. In cross validation, R^2 values were more than 0.99 in all cases, the training data were well fitted the algorithm. When compounds profiles of corn-cobs hydrolysates were used as training data, the cultivating results in corn stover hydrolysates could not be estimated. The data was over-fitted the algorithm with negative Q^2 value. The other cases well fitted. The results demonstrated that volatile component of hydrolysates includes enough information for estimating the results of ethanol fermentation. The remarkable compounds to fit the algorithm were considered as significant compounds for bioethanol production. The significant compounds were estimated by calculating variable importance in projection (VIP) values from correlation coefficients each compounds, and Jack-knife test. Acetate, furfural, 5-hydroxymethylfurfural (5-HMF), and vanillin, known as growth and fermentation inhibitors, indicated high VIP scores and negative correlation coefficient. Further, apocynin, furyl hydroxymethyl ketone, *m*-methoxyacetophenone, and syringaldehyde were

estimated as novel inhibitory materials as strong inhibitors of minor compounds.

4 Insights into future

Here two different cases of modelling bioprocess from “sbstratome” were describes. The modelling strategy can be applied to the versatile processes using raw materials. The fields would never limited in microbial bioprocesses, and can be spread to pharmaceutical, agricultural, and food engineering. In present studies, linear regression models were used. The algorithm would limiting the estimating the cultivation results in case of more complex issues. Artificial intelligence (AI) technology can promote the modelling using “substratome”. We will present also the preliminary results of “substratome” modelling using AI technologies.

References

- Doelle, M. B., and H. W. Doelle ; "Sugar-cane Molasses Fermentation by *Zymomonas mobilis*," *Appl. Microbiol. Biotechnol.* **33**, 31–35 (1990)
- Jumtee, K., H. Komura, T. Bamba, and E. Fukusaki; "Prediction of Japanese Green Tea (Sen-cha) Ranking by Volatile Profiling Using Gas Chromatography Mas Spectrometry and Multivariate Analysis," *J. Biosci. Bioeng.* **112**, 252–255 (2011)
- Kaspro, P. R., A. J. Lange, and D. J. Kirwan; "Correlation of Fermentation Yield with Yeast Extract Composition as Characterized by Near-Infrared Spectroscopy," *Biotechnol. Prog.* **14**, 318–325 (1998)
- Klokmann, S., E. Reiner, R. Bachmann, T. T. Hackl, and M. Fisher; "Food Fingerprinting: Metabolomics Approaches for Geographical Origin Discrimination of Hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS," *J. Agric. Food Chem.*, **64**, 9253–9262 (2016)
- Mimura, N., A. Isogai, K. Iwashita, T. Bamba, and E. Fukusaki; "Gas Chromatography/Mass Spectrometry-Based Component Profiling and Quality Prediction for Japanese Sake," *J. Biosci. Bioeng.*, **118**, 406–414 (2014)
- Potvin, J., E. Fonchy, J. Conway, and C. P. Champagne; "An Automatic Turbidimetric Method to Screen Yeast Extracts as Fermentation Nutrient Ingredients," *J. Microbiol. Methods*, **29**, 153–160 (1997)
- Tachibana, S., K. Watanabe, and M. Konishi; "Estimating Effects of Yeast Extract Compositions on *Escherichia coli* Growth by a Metabolomics Approach," *J. Biosci. Bioeng.*, **128**, 468–474 (2019)
<https://doi.org/10.1016/j.jbiosc.2019.03.012>
- Watanabe, K., S. Tachibana, and M. Konishi: "Modeling Growth and Fermentation Inhibition during Bioethanol Production Using Component Profiles Obtained by

Performing Comprehensive Targeted and Non-Targeted Analyses," *Bioresouc. Technol.*, **281**, 260–268 (2019)

Vázquez, J. A., M. P. González, and M. A. Murado; "Preliminary Tests on Nisin and Pediocin Production Using Waste Protein Sources: Factorial and Kinetic Studies," *Bioresouc Technol.*, **97**, 605–613 (2006)