Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Exploring Design Space and Optimization of nutrient factors for maximizing lipid production in *Metchnikowia pulcherrima* with Design of Experiments

Nichakorn Fungprasertkul[a]*, James Winterburn[a], and Peter Martin[a]

[a] School of Chemical Engineering, University of Manchester, Manchester, M13 9PL, UK
* Main Author: nichakorn.fungprasertkul@postgrad.manchester.ac.uk.

ABSTRACT

Due to the importance of unsaturated fatty acids for human health and the increasing global demand in the food and food crop area, oleaginous yeasts are promising alternative microorganisms for commercial lipid production due to the high volumetric productivity, with *Metchnikowia pulcherrima* being an underexplored oleaginous yeast with potential as a lipid producer. Critical to achieving high productivity lipid production are nutrient factors. A sensitivity test identified carbon and nitrogen sources as important factors in nitrogen limited broth (NLB) for lipid production in *M. pulcherrima* i.e. glucose, yeast extract and Ammonium sulphate. Response Surface Methodology (RSM) involving sets of 15 experimental runs of three-factor three-level Box-Behnken Design (BBD) was implemented for exploring the design of the carbon and nitrogen source in the growth media composition. Quadratic surfaces were least-square fitted and used to identify regions of optimal lipid yield. Multiple sets of runs were conducted with the parameter range progressively adapted and repeated until a clear optimum was identified. The highest total lipid production was in the low carbon concentration range (2.27-21.5 g/L) which suggests the process is more productive with the optimised media compared to NLB media (30.4 g/L of carbon). The optimal carbon concentration was 14.8 g/L whereas the dependence on nitrogen was not found to be significant. The yield of optimal point ($Y_{P/S}$) was 2.39 times higher than NLB media after validation because the depletion of glucose at the end of fermentation (72-104 h) contributes the high increase of lipid accumulation (mg Lipid/ 100 ml culture) of the optimal point (94.5%) which was higher than NLB media (34.5%) for 60%.

Keywords: Fermentation, Food & Agricultural Processes, Microbial Oil, Plackett-Burman design, Box-Behnken design

1. INTRODUCTION

In 2023 the World Health Organization (WHO) suggested that unsaturated fatty acids should be a primary source of fat consumption for humans in its guidance as they can decrease the risk of heart disease by lowering cholesterol levels [1]. The global population is predicted to reach 9 billion people by 2050, resulting in greater demand for food and food crops [2]. In the same way, lipid consumption has increased globally from 83.5 to over 223 million tonnes in past 30 years to serve several purposes including; biofuels, food, supplements, oleochemicals and animal feed. Oleaginous yeasts are considered as viable microorganisms for commercial lipid production because of the high specific growth rate and lipid accumulation. *Metchnikowia pulcherrima* is an under-explored oleaginous yeast with potential as a lipid producer [3]. Critical to achieving high productivity lipid production are nutrient factors [3-4].

This study focuses on understanding the effect of factors toward the growth and lipid production of *M. pulcherrima* and optimizing the important factors to maximize lipid production. A sensitivity test was used to identify the most influential factors on the responses [5]. Plackett-Burman design with a first degree polynomial model was used within this sensitivity test because it focuses

on main effects and properties to minimise the variances in independent estimates with a small number of experimental runs [6]. Box-Behnken design and a quadratic model were used for exploring the design space and optimization. Box-Behnken design can meet a requirement of quadratic system with fewer runs than central composite design [7]. To maximize the lipid production in *M. pulcherrima* by optimizing the important media compositions that have a significant impact on Total Lipid from the sensitivity test.

2. METHODOLOGY

2.1 Sample preparation

2.1.1 Microorganism and Cultivation Techniques

The yeast strain *Metchnikowia Pulcherrima* NCYC2580 (National Collection of Yeast Cultures, Norfolk, UK) was stored in a 23% (v/v) glycerol stock at – 80 °C. For preculture, one cryovial (1.6 ml) of working stock was incubated in 20 ml MEA media (malt extract 30 g/L; mycological peptone 5 g/L) in a 125 ml Erlenmeyer flask at 200 rpm at 20 °C in Infors HT Multitron 2 Incu-bator Shaker with Cooling for 24 h. Fermentation for the control flasks was prepared using nitrogen limited broth (NLB) (NLB: KH_2PO_4 7 g/L; $(NH4)_2SO_4$ 2 g/L; Na_2HPO_4 1 g/L; $MgSO_4{\cdot}7H_2O$ 1.5 g/L; yeast extract 1 g/L; glucose 80 g/L; pH 5). The fermentation was carried out by adding 1 ml of preculture in 100 ml of broth in 500 ml Erlenmeyer flask. The flasks were incubated at 200 rpm at 15 °C for 104 h [4].

2.1.2 Residue biomass, Lipid Content and Total Lipid measurement

After the fermentations had reached their end point at 104 h, biomass was measured by centrifuging 1 ml of culture broth, followed by removal of the supernatant and drying of the obtained cell pellets in a vacuum drier at 50 °C until constant weight. Biomass was recorded in units of g/g (dried cell weight/ culture weight). Lipid content was determined as the percentage of lipid weight per dry cell weight after lipid extraction process as the details in supplementary material (Section S1.1). Residual biomass was calculated by subtraction of the amount of the lipid in 1 g of biomass. Total Lipid (mg Lipid/100 ml culture) was recorded as the total amount of lipid obtained from one flask.

2.1.3 Additional Chemical Analysis

Total Organic Carbon (TOC) and Total Nitrogen (TN) were measured for understanding carbon and nitrogen source consumption. The liquid samples were diluted for 2 and 10 times for TOC and TN analysis respectively and filtered (0.22 µm, Millipore). The TOC and TN values were measured on Shimadzu TOC-V cpn Dissolved Carbon and TNM-1 Nitrogen Analyser at Manchester Analytical Geochemistry Unit (MAGU), The University of Manchester, UK. The measurement was done in duplicates. Glucose concentration in culture was measured daily to understand carbon source consumption. Glucose concentration was quantified as the details in supplementary material (Section S1.2).

2.2 Sensitivity Test

The experimental design for a sensitivity test was constructed to study the influences of seven nutrient factors on *M. pulcherrima* Residual Biomass and Lipid Content. The seven nutrient factors are six factors in NLB and glycerol supplementation. The levels in this experiment were considered from the concentration of the factors in NLB and the possible concentration range that can be developed for optimization as shown in Table 1. The experimental design was constructed on seven-factor Plackett-Burman design as the details in supplementary material (Section S2).

Table 1: Levels of factors for Sensitivity Test

Factors (g/L)	Symbol	Low Level (-)	High Level (+)
Glucose	x_1	40	80*
Glycerol	x_2	0	9
Yeast Extract	x_3	1*	4
$(NH_4)_2SO_4$	x_4	2*	8
Na_2HPO_4	x_5	1*	4
$MgSO_4{\cdot}7H_2O$	x_6	1.5*	6
KH_2PO_4	x_7	7*	15.75

*Concentration in Nitrogen limited broth (NLB)

Following the running of the Plackett-Burman design experiment (eight runs for three replications per the levels of factor in Table 1), a multiple linear regression model was fitted, equation 1. Where Y is the predicted response, a_0 is the intercept of mean and $a_1 \ldots a_7$ are the respective coefficients of factor x_1 until x_7 respectively [6]. Statistical analysis of the multiple linear regression model and response surface plots were carried out in Python 3.11.3 by Statsmodels and Plotly Library.

$$Y = a_0 + a_1x_1 + \cdots + a_7x_7 \quad (1)$$

2.3 Exploring the design space and Optimization

After running the sensitivity test, the important factors for residual biomass and lipid content were identified. The level of the important factor was assigned to the same range with the sensitivity test to visualize the potential range of the optimal point. A three-factor three-level Box-Behnken design was constructed for exploring the design space as the details in supplementary material (Section S3).

Following running the Box-Behnken design experiment (15 runs for two replications as per the levels of

factor in Table S2), a quadratic model was fitted as the equation 2. Where Y is the predicted response, $\beta_o, \beta_i, \beta_{ii}, \beta_{ij}$ are regression coefficients of the model for the intercept, linear, square, interaction effects respectively, and x_i, x_j are the factors [8]. RSM Library in R programming was used for fitting the quadratic model and analysing Analysis of variance (ANOVA) of the models [9], and response surface plots were carried out in Python 3.11.3 using the Plotly Library.

$$Y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum \beta_{ij} x_i x_j \quad (2)$$

Subsequent to exploring the design space, the parameter range was progressively adapted and repeated until a clear optimum was identified. It can be identified when the peak of the important factor of a quadratic model is shown on a response surface clearly. The optimal point was identified after running Box-Behnken design as the levels of factor as shown in Table 2 in two replications. A three-factor three-level Box-Behnken design was constructed for optimization as the details in supplementary material (Section S4). The statistical analysis and response surface fitting were done in the same manner as for the previous exploring the design space step.

Table 2: Levels of factors for optimization

Factors (g/L)	Symbol	Level (-1)	Level (0)	Level (+1)
Glucose	x_1	5	27.5	50
Yeast Extract	x_2	1	4	7
$(NH_4)_2SO_4$	x_3	2	8	14

After the optimal point was identified, it was validated by running an experiment at the optimal point in triplicate and comparing with the control sample (NLB). The yield of the optimal sample was calculated using equation 3 and compared with the control sample.

$$Y_{P/S} = \frac{mass\ product\ produced\ (g)}{mass\ substrate\ consumed\ (g)} \quad (3)$$

2.4 Repeatability of Exploring the design space and Optimization Experiments

The experiments were repeated by using the different preculture for two and five replications for the exploring the design space and optimization respectively to observe the percentage change as per equation 4. Where x_1, x_2 are the old and new values accordingly.

$$Percentage\ change = \frac{x_2 - x_1}{x_1} \times 100 \quad (4)$$

3. RESULTS AND DISCUSSION

3.1 Sensitivity Test

Subsequent to running the experiment of the sensitivity test, the multiple linear regressions of Residual Biomass and Lipid Content were fitted. The linear regression coefficient (R^2 = 0.912, Adjusted R-squared 0.873, F-statistic: 23.57) of multiple regression linear model of Residual Biomass indicates a goodness of the model to explain total variation among responses. The significant factors for Residual Biomass are yeast extract (1-4 g/L), glucose (40-80 g/L) and $(NH_4)_2SO_4$ (2-8 g/L) which are positive impacts. The response surface in figure 1 shows the increase of the Residual Biomass with the glucose and $(NH_4)_2SO_4$ concentrations. These important factors for Residual Biomass are carbon and nitrogen sources in the media. In the initial phase of fermentation, the carbon and nitrogen sources are important for increasing the cell division and protein synthesis to increase the cell density, resulting in the increase of biomass [10]. In the case of Lipid Content, the linear regression coefficient (R^2 = 0.921, Adjusted R-squared 0.887, F-statistic: 26.81) of the model of Lipid Content indicates goodness of the model to explain total variation among responses. Although yeast extract (1-4 g/L), glucose (40-80 g/L) and $(NH_4)_2SO_4$ (2-8 g/L) are significant factors for Lipid Content as well, they have negative effects for lipid content. Figure 1 shows that Lipid Content decreased with the increase of yeast extract and glucose in the response surface from the multiple linear regression model because of the limitation of the nutrient contributes the lipid accumulation inside of an oleaginous cell by activating citrate formation, thus ATP Citrate lyase start to change citrate to a substrate to a triglyceride formation or lipid droplet in the end [10].

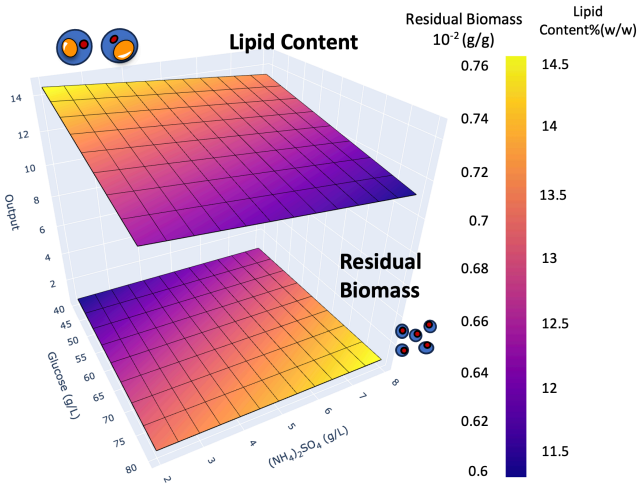

Figure 1: Response surface of multiple regression linear model of Lipid Content and Residual Biomass model from plotting Glucose (g/L) and $(NH_4)_2SO_4$ (g/L).

3.2 Exploring the design space and Optimization

To maximize the lipid production in the culture volume (100 ml), Total Lipid (mg Lipid/100 ml culture) was defined in order to optimize the important factors to

maximize the lipid production. Since the important factors for Residual Biomass and Lipid Content are yeast extract, glucose and $(NH_4)_2SO_4$ which are carbon and nitrogen sources, the important factors were converted to the concentrations of carbon and nitrogen from TOC and TN analysis. From running the experiments as the Box-Behnken design for exploring the design space, the data were fitted as a quadratic model. From the statistical analysis, the model has a moderate correlation (Multiple R-squared: 0.5948, Adjusted R-squared: 0.3696) with non-significant of Lack of fit (Pr(>F) = 0.12870, exceed 0.05) as shown in Table 3, thus the model is valid for the work. The statistical result in Table 4 shows that carbon is the important factor in both linear (x_1) and pure quadratic term (x_1^2), and nitrogen was not found to be significant. This could be because of the insignificant nitrogen concentration in this design space, the model has the low adjusted R-squared because the model cannot fully represent the data. The carbon sources in this process are glucose and yeast extract. They can contribute the high cell density of biomass in fermentation [11]. It is the first stage of lipid production. During the excess of the nutrients condition, the cells actively grow, resulting in the cell mass increase [10]

Table 3: Analysis of variance (ANOVA) for the quadratic regression model of Total Lipid (TL) after running the exploring the design space experiments (1).

	Df	Sum Sq	Mean Sq	F value	Pr(>|t|)
Residuals	9	3108.39	345.38		
Lack of fit	7	2988.41	426.92	7.1163	0.12870[b]

[b] Not significant under 95% level of confidence.

Table 4: Analysis of variance (ANOVA) for the quadratic regression model of Total Lipid (TL) after running the exploring the design space experiments when x_1 is carbon concentration (g/L) and x_2 is nitrogen concentration (g/L) (2).

	Estimate	Std. Error	Pr(>|t|)
Intercept	303.87271	112.7535	0.02459[a]
x_1	-19.69963	8.2744	0.04117[a]
x_2	30.38043	71.3737	0.68036
x_1x_2	-1.46881	1.9027	0.45993
x_1^2	0.41444	0.1689	0.03651[a]
x_2^2	-1.26624	21.5253	0.95438

[a] Significant under 95% level of confidence.

The contour plot from the response surface of the model in Figure 2 shows the highest Total Lipid in the low carbon range. For this reason, the parameter range in the experimental design was adapted progressively towards the to lower carbon concentration for optimization. Since nitrogen was found to be insignificant, the range of nitrogen concentration was expanded to include higher concentrations (2-14 g/L) for optimization.

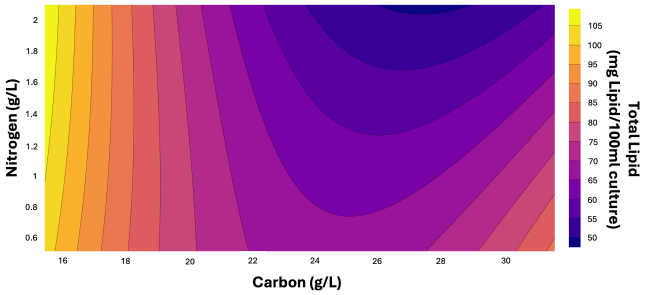

Figure 2: Contour plot of the quadratic model of Total Lipid from plotting Carbon (g/L) and Nitrogen (g/L) from Exploring the design space experiment.

After running the low carbon concentration experiments as per Table 2 for optimising the carbon and nitrogen concentration to maximise the Total Lipid, a quadratic model was fitted. From statistical analysis, the quadratic model has the moderate correlation (Multiple R^2: 0.5672 and Adjusted R^2: 0.3268). Lack of fit of the model is nonsignificant (Pr(>F) = 0.2359, exceed 0.05) as shown in Table 5, so the model is valid for the work. The statistical result in Table 6 shows that carbon (2.27-21.5 g/L) is an important factor for Total Lipid production in both linear (x_1) and pure quadratic term (x_1^2), and nitrogen was not found to be significant. In the same way with the Total Lipid model in exploring the design space, the irrelevant nitrogen concentration in this optmization design space could cause the low adjusted R-squared because the model cannot fully represent the data.

The optimal carbon concentration was 14.8 g/L, and it can be observed from the response surface as shown Figure 3. The optimal nitrogen concentration was identified from the highest Total Lipid that can be achieved at the optimal carbon concentration within the design space. Thus, the model was experimental validated at 14.8 g/L of carbon and 0.52 g/L of nitrogen which is comprised of glucose 38.5 g/L, yeast extract 1 g/L and $(NH_4)_2SO_4$ 2 g/L. After the model validation, the different percentage between the model and the experimental validation point (54.1 ± 2.78 (mg Lipid/ 100 ml culture)) was 6.93%. Carbon to nitrogen ratio of this optimal point in this process was 28.46 (g/g) which is a parameter for a fermentation process for lipid production by an oleaginous yeast [12]. The yield of the optimal point sample and NLB sample which was considered from the Total Lipid (mg lipid/100 ml culture) and the consumed carbon

concentration (mg consumed carbon/100 ml culture). The yield of optimal point ($Y_{P/S}$) was 2.39 times higher than NLB media after validation. The reasons of the high optimal yield were studied further from the glucose consumption and lipid accumulation daily examination.

Table 5: Analysis of variance (ANOVA) for the quadratic regression model of Total Lipid (TL) after running optimization experiments (1).

	Df	Sum Sq	Mean Sq	F value	Pr(>|t|)
Residuals	9	1703.91	189.32		
Lack of fit	8	1683.63	210.45	10.3731	0.2359[b]

[b] Not significant under 95% level of confidence.

Table 6: Analysis of variance (ANOVA) for the quadratic regression model of Total Lipid (TL) after running the optimization experiments when x_1 is carbon concentration (g/L) and x_2 is nitrogen concentration (g/L) (2).

	Estimate	Std. Error	Pr(>|t|)
Intercept	-9.47305	21.1502	0.66482
x_1	7.54068	2.7051	0.0211[a]
x_2	0.79361	8.6870	0.92921
x_1x_2	-0.11518	0.6229	0.85738
x_1^2	-0.25198	0.1003	0.0331[a]
x_2^2	-0.00802	0.2029	0.96935

[a] Significant under 95% level of confidence.

Since glucose is comprised 97.47% (w/w) of the total carbon the change of glucose with Total Lipid of the optimal point was examined every day to understand the metabolism of lipid accumulation and carbon source consumption by *M. pulcherrima* during fermentation. The study found that the optimal sample gave a high yield because of the depletion of glucose at the end of fermentation (72-104 h), contributing to the increase of lipid accumulation (mg Lipid/ 100 ml culture) at the optimal point (94.5%) which was higher than NLB media (34.5%) for 60%. This is because there are two stages of lipid accumulation for the carbon limitation media of the optimal point. In the first stage, the cells grow and accumulate gradually at the same time. However, at the end of the fermentation (72-104 h), the glucose was completely consumed, and the Lipid Content was increased dramatically to 7.67% (1.86 times of NLB samples) [13]. In contrast, the lipid accumulation in the NLB fermentations increased gradually with biomass. At the end of the process, the Lipid Content at 104 h was only 4.14%, and there was residual glucose for 15.17 g/L.

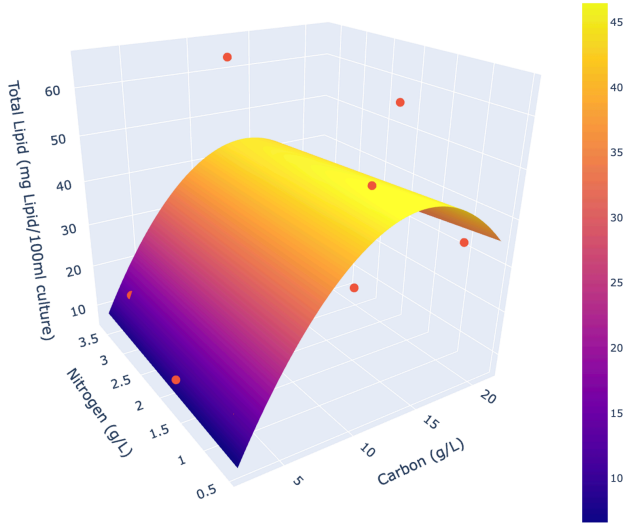

Figure 3: Response surface of quadratic model of Total Lipid from plotting Carbon (g/L) and Nitrogen (g/L) from Low Carbon Concentration from Box-Behnken Design.

3.3 Repeatability of Exploring the design space and Optimization Experiments

From the study of repeatability of the exploring design space experiment, the highest percentage change, 55.78% of the total lipid. This value indicates the likelihood of repeating the experiment by running the different preculture batch. In the same way, the repeatability of the optimization experiment was studied to observe the different change from running the different preculture batch and error during the experimental run as shown in Figure 4. The higher than expected total lipid of the third replication was observed. This is because of contamination of biomass during lipid separation, thus the weight of lipid in this data set are higher than the other replication. The different percentage the Total Lipid data was decreased for 4.97% from 92.1% to 87.2% by excluding the third replication in the data set. Removing an error data set before plotting a quadratic model was operated because the value of the third replication cannot represent the likelihood of the data set. Thus, the model could be representing the behavior of *M. pulcherrima* precisely by including only the data set that was experimented correctly.

4. CONCLUSION

The sensitivity test, the exploring the design space and optimization successfully maximized lipid production in the limited volume (100 ml) from *M. pulcherrima.* The sensitivity test from Plackett-Burman design and multiple linear regression model identified carbon and nitrogen sources as the important factors as expected in nitrogen limited broth (NLB) for lipid production by *M. pulcherrima* which are glucose, yeast extract and Ammonium sulphate. The important factors have a positive impact on

the Residual Biomass but a negative impact on the Lipid Content because they are the nutrients for the cell growth, but the starvation of nutrients contributes the lipid accumulation. Box-Behnken Design with quadratic model fitting suggests the maximal lipid production point on the response surface at the low carbon concentration range (2.27-21.5 g/L) for optimization of the Total Lipid. The optimal carbon concentration was 14.8 g/L, thus the optimal carbon to nitrogen ration was 28.46 (g/g). The yield of optimal point was higher than NLB media at 2.39 times because the depletion of glucose at the end of fermentation (72-104 h) contributes the high increase of lipid accumulation of the optimal point. The repeatability of the exploring the design space and optimization experiments were tested to understand the behavior of *M. pulcherrima*. and potential experimental error during running the experiments.

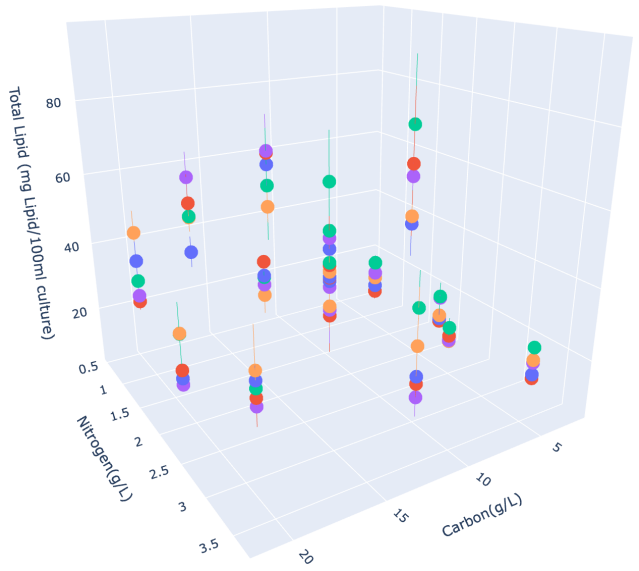

Figure 4: Scatter from plotting the experimental data of Total Lipid model on carbon and nitrogen concentrations axes from running the optimization experiment (blue-first replication, red-second replication, green-third replication, fourth-purple replication and orange-fifth replication).

DIGITAL SUPPLEMENTARY MATERIAL

Digital supplementary material can be found at https://psecommunity.org/LAPSE:2025.0028

REFERENCES

1. World Health Organization. https://www.who.int/publications/i/item/9789240061668
2. United Nations. https://www.un.org/youthenvoy/2013/09/fao-food-and-agriculture-organization-of-the-united-nations/
3. Abeln, F., & Chuck, C. J. (2021). The history, state of the art and future prospects for oleaginous yeast research. *Microb Cell Fact*, 20(1), 221. https://doi.org/10.1186/s12934-021-01712-1
4. Abeln, F., & Chuck, C. J. (2020). The role of temperature, pH and nutrition in process development of the unique oleaginous yeast Metschnikowia pulcherrima. *Journal of Chemical Technology & Biotechnology*. https://doi.org/10.1002/jctb.6301
5. Lin, S., Xie, M., Wu, M., & Zhou, W. (2018). Global Sensitivity Analysis of Large Reaction Mechanisms Using Fourier Amplitude Sensitivity Test. Journal of Chemistry, 2018(1), 5127393. https://doi.org/https://doi.org/10.1155/2018/5127393
6. Vanaja, K., & Shobha Rani, R. H. (2007). Design of Experiments: Concept and Applications of Plackett Burman Design. *Clinical Research and Regulatory Affairs*, 24(1), 1–23. https://doi.org/10.1080/10601330701220520
7. Liu, S. (2020). Design of experiment. In Bioprocess Engineering (pp. 885-933). https://doi.org/10.1016/b978-0-12-821012-3.00020-8
8. Ghorbel-Bellaaj, O., Hmidet, N., Jellouli, K., Younes, I., Maâlej, H., Hachicha, R., & Nasri, M. (2011). Shrimp waste fermentation with Pseudomonas aeruginosa A2: Optimization of chitin extraction conditions through Plackett–Burman and response surface methodology approaches. *International Journal of Biological Macromolecules*, 48(4), 596-602. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2011.01.024
9. Lenth, R. V. (2009). Response-Surface Methods in R, Using rsm. *Journal of Statistical Software*, 32(7), 1 - 17. https://doi.org/10.18637/jss.v032.i07
10. Chopra, J., Rangarajan, V., & Sen, R. (2022). Recent developments in oleaginous yeast feedstock based biorefinery for production and life cycle assessment of biofuels and value-added products. *Sustainable Energy Technologies and Assessments*, 53. https://doi.org/10.1016/j.seta.2022.102621
11. Abeln, F., & Chuck, C. J. (2019). Achieving a high-density oleaginous yeast culture: Comparison of four processing strategies using *Metschnikowia pulcherrima*. *Biotechnol Bioeng*, 116(12), 3200-3214. https://doi.org/10.1002/bit.27141
12. Lei, Y., Wang, X., Sun, S., He, B., Sun, W., Wang, K., Chen, Z., Guo, Z., & Li, Z. (2024). A review of lipid accumulation by oleaginous yeasts: Culture mode.

Sci Total Environ, 919, 170385. https://doi.org/10.1016/j.scitotenv.2024.170385

13. Amaretti, A., Raimondi, S., Sala, M., Roncaglia, L., De Lucia, M., Leonardi, A., & Rossi, M. (2010). Single cell oils of the cold-adapted oleaginous yeast Rhodotorula glacialis DBVPG 4785. *Microbial cell factories*, 9, 73. https://doi.org/10.1186/1475-2859-9-73