Growth rate data fitting of *Yarrowia lipolytica* NCIM 3589 using logistic equation and artificial neural networks

Sarat Babu Imandi[1,*], Sita Kumari Karanam[2], Surekha Darsipudi[3], Hanumantha Rao Garapati[1,4]

[1]Department of Biotechnology, ANITS, Bheemunipatnam, India;
[2]M. R. College of Pharmacy, Phool Baugh, India;
[3]Koringa College of Pharmacy, Korangi, India;
[4]Department of Chemical Engineering, Andhra University, Andhra Pradesh, India.
Email: saratbabu_imandi@yahoo.com; saratbabuimandi@gmail.com; sita_karanam@yahoo.com; sitakaranam@gmail.com

ABSTRACT

Growth rate of *Yarrowia lipolytica* NCIM 3589 was observed in a fermentation medium consisting of peptone, yeast extract, sodium chloride. Logistic equation was fitted to the growth data (time vs. biomass concentration) and compared with the prediction given by Artificial Neural Networks (ANN). ANN was found to be superior in describing growth characteristics. A single MATLAB programme is developed to fit the growth data by logistic equation and ANN.

Keywords: *Yarrowia lipolytica*; Logistic Equation; Artificial Neural Networks

1. INTRODUCTION

Yarrowia lipolytica is one of the most extensively studied ''non-conventional'' yeasts which is currently used as a model for the study of protein secretion, peroxisome biogenesis, dimorphism, degradation of hydrophobic substrates, and several new fields. Recently, the entire sequence of the six *Y. lipolytica* chromosomes has been determined [1], allowing its admission into the ''omic'' disciplines such as genomics, transcriptomics and proteomics. Several reviews have already been published on its physiology and genetics [2-4], secretion [5-7], dimorphism [8], peroxisome biogenesis [9,10] and mitochondrial complex I [11].

Yarrowia lipolytica is ascomycetous yeast which has been assigned to the family *Dipodascaceae* [12]. This organism is non-pathogenic and oil degrading yeast. It is suitable for the aerobic biodegradation and detoxification of oil mill waste water. It is good for the waste water purification and reduction of pollution due to its ability to reduce chemical oxygen demand and biological oxygen demand. It is also suggested for a deliberate use in cheese ripening due to its extra cellular enzyme activities. It is therefore important to know the growth behavior of this microorganism. The growth capacity of the organism can be evaluated by using growth curves which can be plotted by taking log number of cells versus incubation time. Logistic equation is conveniently used to evaluate the growth capacity of the organism.

Recently, a number of new models have been introduced which involve the application of artificial neural networks to describe the complex growth of yeasts. ANN is highly interconnected network consisting of many simple processing elements capable of performing a massively parallel computation for data processing inspired by the elementary principles of the nervous system. ANN tends to correlate the input data with the output data. In this paper, the growth data were fitted by using both logistic equation and ANN.

2. MATERIALS AND METHODS

2.1. Microorganism

Y. lipolytica NCIM3589 obtained from National Chemical Laboratory, Pune, India, was used throughout the study.

2.2. Growth Conditions

The culture was maintained on MGYP slants having the composition (%): malt extract 0.3, glucose 1.0, yeast extract 0.3, peptone 0.5 and agar agar 2.0. The pH of the medium was adjusted to 6.4–6.8 and culture was incubated at 30°C for 48 h. Sub-culturing was carried out once in every 2 weeks and the culture was stored at 4°C.

2.3. Growth of *Y. lipolytica*

The yeast *Yarrowia* strain was cultivated in a medium containing peptone 5 g, yeast extract 3 g and sodium chloride 3 g/l of distilled water. The cells were culti-

Table 1. Experimental and predicted values of growth data.

Sl. No	Time(hrs)	Biomass concentration(µg/ml)		
		Experimental data	Predicted by logistic equation	Predicted by GRNN
1	0	0.214	0.214	0.214
2	1	1.262	0.433	1.262
3	2	1.291	0.806	1.291
4	3	1.303	1.319	1.303
5	4	1.578	1.851	1.578
6	6	2.29	2.518	2.29
7	7	2.57	2.652	2.57
8	8	3	2.717	3
9	9	3	2.748	3
10	10	3	2.762	3
11	11	3	2.769	3
12	12	3	2.771	3
13	13	3	2.773	3
14	14	3	2.773	3
15	15	3	2.774	3
16	17	3	2.774	3
17	18.3	1.886	2.774	1.886
18	19.3	1.886	2.774	1.886

vated in this medium at 30°C on a shaker at 200 rpm for 24 h [13]. The growth data was taken using double beam spectrophotometer with which optical density was taken at 570 nm. The observations were noted each hour until the organism has reached the stationary stage of growth and the data were reported in **Table 1**.

2.4. Logistic Growth Model

Using the logistic model, the growth curve assumes a sigmoidal shape by plotting biomass vs. time. This shape can be predicted by combining the Monod equation with the growth equation and an equation for the yield of cell mass based on substrate consumption is given by

$$\frac{dX}{dt} = \frac{\mu_m S}{k_s + S} X \tag{1}$$

The specific growth rate is related to the amount of unused carrying capacity as

$$\mu_g = k\left(1 - \frac{X}{X_\infty}\right) \tag{2}$$

Thus,

$$\frac{dX}{dt} = kX\left(1 - \frac{X}{X_\infty}\right) \tag{3}$$

The integration of the equation yields [14],

$$X = \frac{X_0 e^{kt}}{1 - \frac{X_0}{X_\infty}(1 - e^{kt})} \tag{4}$$

where X_0 = initial concentration of biomass, g/l

X_∞ = concentration of biomass at infinite time, g/l

k = rate constant, h^{-1}.

This is the logistic function relating biomass "X" with time "t" with unknown coefficients k and X_∞ which are to be estimated using non-linear least squares. However, non-linear least squares routine requires the initial guess values of k and X_∞. The guess value for X_∞ might be taken as 1.01 times the end value of biomass, while a guess value for k can be calculated by approximating the **Eq.3** for a pair of data points as follows:

$$k = \frac{\frac{1}{\overline{X}}\frac{\Delta X}{\Delta t}}{\left(1 - \frac{\overline{X}}{X_\infty}\right)}$$

where $\overline{X}$ is the average of two data points, ΔX is the difference between the two data points, Δt is the corresponding difference in time. Non-linear regression routine, 'nlinfit' of MATLAB 7, was used to estimate the values of k and X_∞. The predicted biomass values were reported in **Table 1** as the fourth column.

2.5. Artificial Neural Networks

A neural network is a mathematical representation of the neurological functioning of the brain. It stimulates the brain learning process by mathematical modeling of the network structure of interconnected nerve cells. The essential requirement of neural network modeling was sufficient number of data as it operates directly on input-output data. Thus, ANN is purely data driven model made up of inter connected processing elements called neurons

that are arranged in layers. The most important of this modeling methodology is its ability to reduce the complexity of the network. A good introduction to the subject with respect to MATLAB usage is given by Demuth *et al.* [15].

The type of ANN used is the Generalized Regression Neural Networks (GRNN) which is mainly based on non-linear regression theory. ANN approximates any arbitrary function between input and output vectors, drawing function estimate directly from the training data. GRNN has four layers: input, a layer of radial centers, a layer of regression units and output. The radial layer units represent the centers of clusters of known training data. This layer must be trained by a clustering algorithm such as sub sampling. Thus, GRNN is a universal approximator for smooth function, so it should be able to solve any smooth function approximation problem given enough data.

Using MATLAB 7, the biomass values and time data were fitted by GRNN and predicted values were reported in **Table 1** as the fifth column.

3. RESULTS AND DISCUSSION

The experimental growth data was fitted by using both logistic equation and ANN. In logistic growth model, the parameter values (k and $1/X_\infty$) are obtained as 0.7946 and 0.3605 respectively using 'nlinfit' routine and using the parameter estimates the biomass values are predicted at different periods of times with neural networks, GRNN was used to predict the biomass values. Several iterations were conducted with different spread values ranging from 0.1 to 1.0. A larger spread leads to a large area around the input vector where first layer neurons will respond with significant outputs. If the spread is small the function is very steep, so that the neuron with the weight vector
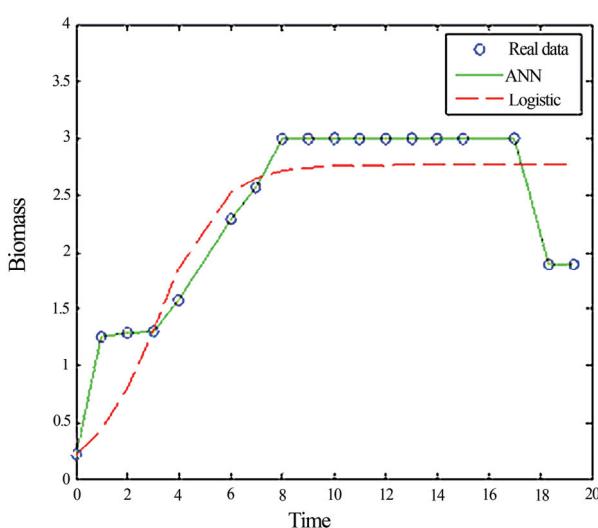

Figure 1. Growth data fitting with ANN & logistic equation.

closes to input will have a much larger output than other neurons. Finally, with a spread value of 0.2, we could achieve the predicted values which are almost identical to the experimental values (**Figure 1**).

Figure 1 represents the fitting of growth rate data using ANN and logistic equation. The solid line, dotted line, the circles represent the ANN curve, logistic curve and experimental data respectively. It is evident from **Figure 1** that ANN curve best fits with the experimental data curve when compared to logistic curve. Thus logistic equation is not adequate in fitting the data.

4. CONCLUSIONS

A comparative study has been made for the fitting of growth rate data of *Yarrowia lipolytica* NCIM 3589 using logistic equation and GRNN respectively. It is found that logistic equation is poor in fitting the growth data when compared to ANN. Thus ANN was found to be superior in describing the growth characteristics of the organism.

REFERENCES

[1] Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuve´glise, C., Talla, E., Goard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, R., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P. and Souciet, J.L. (2004) Genome evolution in yeasts. *Nature*, **430**, 35-44.

[2] Barth, G. and Gaillardin, C. (1996) *Yarrowia lipolytica* in: nonconventional yeasts in biotechnology (Wolf, K., Ed.), *A Handbook*, Springer, Berlin, Heidelberg, New York, pp. 313-388.

[3] Barth, G. and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiology Reviews*, **19**, 219-237.

[4] Barth, G., Beckerich, J.M., Dominguez, A., Kerscher, S., Ogrydziak, D., Titorenko, V. and Gaillardin, C. (2003) Functional genetics of *Yarrowia lipolytica* In: (de Winde, J.H., Ed.), *Functional Genetics of Industrial Yeasts*, Springer, Berlin, Heidelberg, New York, 227-271.

[5] Dominguez, A., Ferminan, E., Sanchez, M., Gonzalez, F.J., Perez-Campo, F.M., Garcia, S., Herrero, A.B., San Vicente, A., Cabello, J., Prado, M., Iglesias, F.J., Choupina, A., Burguillo, F.J., Fernandez-Lago, L. and

    ABB

Lopez, M.C. (1998) Non-conventional yeasts as hosts for heterologous protein production. *International Microbiology*, **1**, 131-142.

[6] Nicaud, J.M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P. and Gaillardin, C. (2002) Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Research*, **2**, 371-379.

[7] Madzak, C., Gaillardin, C. and Beckerich, J.M. (2004) Heterologous protein expression and secretion in the nonconventional yeast *Yarrowia lipolytica*: a review. *Journal of Biotechnology*, **109**, 63-81.

[8] Dominguez, A., Ferminan, E. and Gaillardin, C. (2000) *Yarrowia lipolytica*: an organism amenable to genetic manipulation as a model for analyzing dimorphism in fungi. *Contributions to Microbiology*, **5**, 151-172.

[9] Titorenko, V.I., Smith, J.J., Szilard, R.K. and Rachubinski, R.A. (2000) Peroxisome biogenesis in the yeast *Yarrowia lipolytica*. *Cell Biochemistry and Biophysics*, **32**, 21-26.

[10] Titorenko, V.I. and Rachubinski, R.A. (2001) Dynamics of peroxisome assembly and function. *Trends in Cell Biology*, **11**, 22-29.

[11] Kerscher, S., Drose, S., Zwicker, K., Zickermann, V. and Brandt, U. (2002) *Yarrowia lipolytica*, a yeast genetic system to study mitochondrial complex I. *Biochimica et Biophysica Acta*, **1555**, 83-91.

[12] Kurtzman, C. and Fell, J., Eds., (1998). The yeasts: a taxonomy study, Elsevier, Amsterdam.

[13] Imandi, S.B., Bandaru, V.V.R., Somalanka, S.R. and Garapati, H.R. (2007). Optimization of medium constituents for the production of citric acid from byproduct glycerol using doehlert experimental design. *Enzyme and Microbial Technology*, **40**, 1367-1372.

[14] Shuler M.L. and Kargi, F. (2002). Bioprocess engineering-basic concepts, 2nd Edition, Prentice-Hall of India pvt ltd, New Delhi.

[15] Demuth H., Beale M. and Hagan M. (2006). Neural network toolbox for use with MATLAB. *Neural Network Toolbox*, IEE Savoy Place, London.