Modeling and Optimization of L-asparaginase Production by Enterobacter Aerogenes Using Artificial Neural Network Linked Genetic Algorithm

G.Baskar, V.Rajasekar, and S.Renganathan

Abstract— **In the present work the artificial neural network linked genetic algorithm was applied for the optimization of fermentation media components like carbon and nitrogen sources for L-asparaginase production by Enterobacter aerogenes MTCC 2823 in submerged fermentation. Artificial neural network (ANN) based back propagation algorithm was used to train and test the neural network using the experimental activity obtained by central composite design. Higher value of coefficient of determination (R2=0.984) of artificial neural network justified an excellent correlation between the media components and L-asparaginase activity, the artificial neural network model fitted well with high statistical reliability and significance than RSM model (R2=0.871) developed by central composite design. The predicted optimum concentration of the media components using artificial neural network linked genetic algorithm was sodium citrate 2.09%, DAHP 0.25% and L-asparagine 0.92% with the maximum predicted L-asparaginase activity of 18.59 IU/mL which was close to the experimental L-asparaginase activity of 18.72 IU/mL at simulated optimum conditions.**

Index Terms—**Fermentation; Optimization; Polynomial model; Artificial Neural Network; Response Surface Methodology.**

I. Introduction

The fermentative production of any product is influenced by operating conditions such as temperature, pH, and agitation rate and the various media components to fulfill the nutritional requirements of the microbial growth and synthesis of fermentation product. Hence the optimization of nutritional requirements and operating conditions is an important step in any bioprocess development.

Statistical experimental designs have been used in several steps of optimization strategy and it is better acknowledged than traditional one variable at a time method [1].

The central composite design (CCD) using response surface methodology (RSM) is an efficient statistical design for optimization of multiple variables in order to predict the best performance conditions with the minimum number of experiments. RSM is suited for studying the main and interaction effects of factors on growth or metabolite formation during microbial fermentation [2,3]. The development of accurate models for a biological reaction system on a chemical and physical basis is still a critical challenge, mainly due to the non-linear nature of the biochemical network interactions. It has been shown that machine learning techniques such as artificial neural networks (ANN) and genetic algorithms (GA) mimic different aspects of biological information processing for data modeling and could prove to be useful in media optimization [4-6].

L-asparaginase (L-asparagine amidohydrolase; EC.3.5.1.1), catalyzes the deamidation of L-asparagine to L-aspartic acid and ammonia, is used as a chemotherapeutic agent for acute lumphocytic leukeamia and less frequently for acute myeloblastic leukeamia, chronic lumphocytic leukeamia, Hodgkin's disease, melonosarcoma and non-Hodgkin's lymphoma. Although Clementi in 1922 had reported its presence in guinea-pig serum, the anti-tumour properties of this enzyme were only recognized some time later [7,8]. Tsuji first reported deamidation of L-asparagine by extracts of E. coli [9]. Broome in 1961 discovered that the regression of lymphosarcoma transplants in mice treated with guinea-pig serum was due to the nutritional dependence of the malignant cells on exogenous L-asparagine [10]. The production of L-asparaginase by bacterial sources is mainly regulated by different degree of carbon catabolite and oxygen repression [11,12].

In the present work various effects of sodium citrate, di-ammonium hydrogen phosphate and L-asparagine (selected based on literature survey) were studied and optimized using work ANN linked GA for enhanced L-asparaginase production by Enterobacter aerogenes MTCC 2823 under shake flask fermentation conditions.

II. Materials And Methods

A. Central Composite Design

The important media components namely sodium citrate(X_1), DAHP (X_2) and L-asparagine (X_3) for L-asparaginase production by *Enterobacter aerogenes* was optimized using Central Composite Design (CCD). The variables were prescribed into three levels, -1, 0, + 1 for low, middle and high and the central composite experimental design was developed using Minitab15 *software* in coded units. Table 1 shows the experimental CCD design in actual units (%, w/v). The response variable was fitted into quadratic model to correlate the effect of the variables on L-asparaginase activity.

$$Y = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_{11}X_1^2+\beta_{22}X_2^2+\beta_{33}X_3^2+ \beta_{12}X_1X_2+\beta_{13}X_1X_3+\beta_{23}X_2X_3 \quad (1)$$

Manuscript received January 30, 2009.

G.Baskar is with the Department of Biotechnology, St.Joseph's College of Engineeirng, Chennai-600119, India. (corresponding author to provide phone: +91 94436 78571;email: basg2004@gmail.com)

V.Rajasekar is with the Department of Chemical Engineering, Manipal Institute of Technology, Manipal-576 104, India. (corresponding author to provide +91 9916824207S;email: rajashekardv@gmail.com)

.Renganathan is with the Department of Chemical Engineering, A.C.Technology, AnUniversity Chennai, Chennai-600 025. India.(corresponding author to provide +91 9941613532;email: rengsah@rediffmail..com,)

where Y is the predicted response, β_0 model constant; X_1, X_2 and X_3 are independent variables; β_1, β_2 and β_3 are linear coefficients; β_{12}, β_{13} and β_{23} are cross product coefficients; β_{11} β_{22} and β_{33} are the quadratic coefficients [3]. The optimum levels of the selected variables were obtained by solving this regression model.

B. Artificial Neural Network Linked Genetic Algorithm

The development of accurate model is still a critical challenge, mainly due to the non-linear nature of the biochemical network interactions. It has been shown that machine learning techniques such as ANN mimic different aspects of biological information processing for data modeling and could prove to be useful in media optimization. ANN is capable of predicting the output when any input similar to the pattern that it has learnt is fed. A multilayer feed forward ANN trains and evaluates the system performance using adaptive gradient learning rule. The learning rate of the network was set to value, that resulted in an optimal coefficient of correlation (R^2) for the neural network. Regression based response surface models require the order of the model to be stated, while ANN tends to implicitly match the input vector to the output vector [4, 13]. In the present work three neurons in the input layer, three in the hidden layer and one in the output layer of the network was used to model the dependence of the L-asparaginase production on three media components. GA on the other hand, is the commonly used global optimization technique that optimizes a given function over a particular range, and is based on the evolutionary methods of natural selection of the best individuals in a population. The genetic algorithm explores all regions of the solution space using a population of individuals. Each individual represents a set of independent variables. Initially, a population of individuals is formed randomly. The fitness of each individual is evaluated using an objective function [5,14]. Once the ANN model was developed, GA was used to determine the maximum L-asparaginase production and optimum concentration of the media components.

C. Production of L-Asparaginase

The bacterial culture *Enterobacter aerogenes* MTCC 2823 was obtained from Institute of Microbial Technology, Chandigarh, India. It was grown on nutrient agar slants at 35°C for 24 h, stored at 4°C and subcultured periodically. Culture suspension of 5% inoculum size was transferred to Erlenmeyer flasks with 100 mL of liquid Czapek-Dox medium prepared using carbon and nitrogen sources based on experimental design (Table 2) at pH 6.7 with fixed concentration of other nutrients such as glucose 0.5%; potassium chloride 0.05%; $MgSO_4.7H_2O$ 0.05%; $FeSO_4.7H_2O$ 0.001% and K_2HPO_4 0.1%. The culture was kept in orbital shaker in 186 rpm and at 35°C. A culture sample of 2 mL was collected at maximum L-asparaginase production time of 6 h [3].

A. Assay of L-asparaginase Activity

The cells were separated from fermentation broth using refrigerated centrifuge (5°C) at 10,000 rpm, cell mass was suspended and shaken vigorously with 2 mL phosphate buffer (pH 7.0) containing triton X-100 (0.01%) for 5 min to and centrifuged. The cell mass was suspended in 1.5 mL sodium-borate buffer pH 8.65, and assayed for intracellular L-asparaginase activity by Nesslarization [10].

TABLE I: EXPERIMENTAL, CCD AND ANN PREDICTED L-ASPARAGINASE ACTIVITY

Std. Run	X_1	X_2	X_3	L-asparaginase activity (IU/mL)		
				Experimental	CCD Predicted	ANN Predicted
1	1	0.25	0.25	14.53	15.80	14.53
2	3	0.25	0.25	14.89	14.98	14.89
3	1	0.75	0.25	15.44	16.35	15.44
4	3	0.75	0.25	14.63	15.28	14.63
5	1	0.25	0.75	18.42	18.66	18.71
6	3	0.25	0.75	17.11	17.10	17.11
7	1	0.75	0.75	14.01	14.82	14.84
8	3	0.75	0.75	13.38	13.01	13.39
9	0.32	0.5	0.5	18.49	17.01	18.47
10	3.68	0.5	0.5	14.58	14.80	14.58
11	2	0.33	0.5	18.17	17.66	18.16
12	2	1.17	0.5	15.44	14.68	14.47
13	2	0.5	0.33	16.73	15.43	16.73
14	2	0.5	1.17	15.89	15.92	15.89
15	2	0.5	0.5	19.22	18.67	18.63
16	2	0.5	0.5	18.76	18.67	18.48
17	2	0.5	0.5	18.57	18.67	18.64
18	2	0.5	0.5	18.57	18.67	18.63
19	2	0.5	0.5	18.22	18.67	18.58
20	2	0.5	0.5	18.46	18.67	18.62

III. RESULTS AND DISCUSSION

A. Modeling and Optimization by Response Surface Methodology

The effect of sodium citrate, DAHP and L-asparagine on L-asparaginase production was studied using the experimental L-asparaginase activity given in table 1 and was subjected to multiple linear regression analysis [3]. The second order polynomial model given in equation (2) was fitted in coded unit of the variables. A higher value of coefficient of correlation (R2 = 0.871) obtained for this model indicates that the model justified a good correlation between dependent variables L-asparaginase activity and the independent variables namely sodium citrate, DAHP and L-asparagine. The optimal concentration of independent variables was reported as sodium citrate 1.87% (w/v), DAHP 0.57% (w/v) and L-asparagine 0.85% (w/v).

B. Modeling and Optimization by Artificial Neural Network Linked Genetic Algorithm

Back propagation algorithm is a multilayer feed forward ANN with three neurons in input layer, three in the hidden layer and one in the output layer using 'Tanh' transfer function was used to model the dependence of L-asparaginase production on independent variables such as sodium citrate, DAHP and L-asparagine (figure 1). The ANN predicated L-asparaginase activity after 787950 iterations is given in table 1. Although both RSM regression model and ANN model provided accurate predictions, ANN model (R^2=0.984) showed better correlation with the experimental L-asparaginase activity than RSM regression model (R^2=0.871).

GA with population size of 30, mutation rate of 0.1 and uniform cross overrate of 0.8 was used to optimize the ANN model and determine the maximum L-asparaginase production and optimum concentration of the media components. The importance of the media components on L-asparaginase production was found in the order of DAHP (38.38%), sodium citrate (37.47%) and L-asparagine (24.15%) as shown in figure 2. The predicted optimum concentration of the media components was sodium citrate 2.09%, DAHP 0.25% and L-asparagine 0.92% with the maximum predicted L-asparaginase production of 18.59 IU/mL.

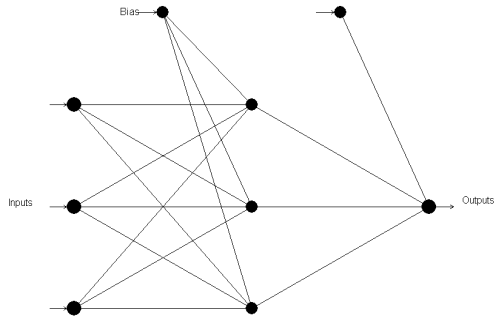

Figure 1. Multilayer normal feed forward batch back propagation neural network

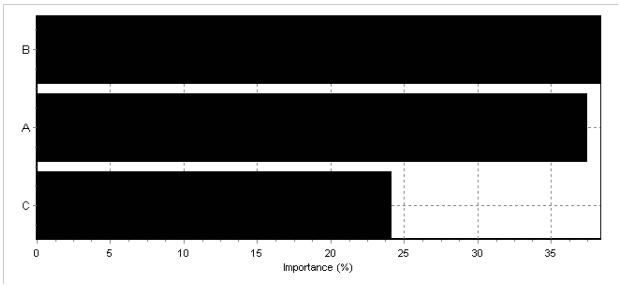

Figure 2. Importance of media components on L-asparaginase production (A-Sodium citrate, B-DAHP and C=L-asparagine)

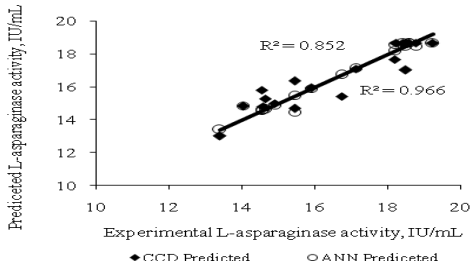

Figure 3. Distribution coefficients for CCD and ANN predicted L-asparaginase activity

The ANN predicted condition was experimentally verified. The verification experiment in triplicate was conducted at the predicted optimum values media components by ANN model for validation. All the other fermentation conditions were fixed as CCD experiments. The experimental L-asparaginase activity of 18.72 IU/mL was obtained at predicted optimal conditions of ANN model. Figure 3 shows the predicted distribution coefficient of the CCD (Pred. R2=0.852) and RSM (Pred. R2=0.967) predicted L-asparaginase activity and are closer to the distribution coefficient (R2). The high value of predicted distribution coefficient indicates that the ANN is highly accurate in successful prediction of L-asparaginase activity.

Hence ANN can be effectively used to adequately represent the relationship between the input variables and L-asparaginase production. The optimization using ANN model linked with GA is found to be the more effective for L-asparaginase production with a high degree of accuracy.

IV. Conclusion

ANN based BPN algorithm provided more accurate predictions than CCD based second order polynomial model on dependence of media components for L-asparaginase production on media using Enterobacter aerogenes MTCC 2823. The predicted optimum concentration of the media components ANN based BPN algorithm was found to be 2.09% sodium citrate, 0.25% DAHP and 0.92% L-asparagine with the maximum predicted L-asparaginase production of 18.59 IU/mL. The experimental L-asparaginase activity of 18.72 IU/mL was obtained at the predicted optimal conditions.

References

[1] S. K. Rai, A. K. Mukherjee. (2010) Statistical optimization of production, purification and industrial application of a laundry detergent and organic solvent-stable subtilisin-like serine protease (Alzwiprase) from Bacillus subtilis DM-04. Biochem. En J. 48, pp. 173–180.

[2] M. Hymavathi, T. Sathish, Ch. Subba Rao, R. S. Prakasham (2009) Enhancement of L-asparaginase production by isolated Bacillus circulans (MTCC 8574) using response surface methodology. Appl. Biochem. Biotechnol. 159, pp. 191–198.

[3] G. Baskar, M. Dharmendira Kumar, A. Anand Prabu, S. Renganathan, C. K. Yoo (2009) Optimization of carbon and nitrogen sources for L-asparaginase production by Enterobacter aerogenes using response surface methodology. Chem. Biochem. Eng. Q. 23, pp. 393–397.

[4] C. Sivapathasekaran, S. Mukherjee, A. Ray, A. Gupta, R. Sen (2010) Artificial neural network modeling and genetic algorithm based medium optimization for the improved production of marine biosurfactant. Bioresource Technol. 101, pp. 2884–2887.

[5] R. Ebrahimpour, D. Rahman, M. Ean Chng, A. Basri, A. Salleh (2008) A modeling study by response surface methodology and artificial neural network on culture parameters optimization for thermostable lipase production from a newly isolated thermophilic Geobacillus sp. strain ARM. BMC Biotechnol. 8, 96.

[6] M. P. Pal, B. K. Vaidya, K. M. Desai, R. M. Joshi, S. N. Nene, B. D. Kulkarni (2009) Medium optimization for biosurfactant production by Rhodococcus erythropolis MTCC 2794: Artificial intelligence verses a statistical approach. J. Ind. Microbiol. Biotechnol. 36, pp. 747–756.

[7] J. C. Wriston, T. O. Yellin (1973) L-asparaginase. A review. Adv. Enzymol. Relat. Areas Mol. Biol. 39, pp. 185–248.

[8] C. Howard, J. H. Schwartz (1968) Production of L-asparaginase II by Escherichia coli. J. Bacteriol. 96. pp. 2043–2048.

[9] J. D. Broome (1963) Evidence that the L-asparaginase of guinea pig serum is responsible for its antilymphoma effects: I. properties of the L-asparaginase of guinea pig serum in relation to those of the antilymphoma substance. J Exp Med. 118, pp. 99–120.

[10] J. Mukherjee, S. Majumdar, T. Scheper (2000) Studies on nutritional and oxygen requirements for production of L-asparaginase by Enterobacter aerogene. Appl. Microbiol. Biotechnol. 53, pp. 180–120.

[11] H. Geckil, S. Gencer, M. Uckun (2004) Vitreoscilla hemoglobin expressing Enterobacter aerogenes and Pseudomonas aeruginosa respond differently to carbon catabolite and oxygen repression for production of L-asparaginase, an enzyme used in cancer therapy. Enzyme Microb. Tech. 35, pp. 182–189.

[12] H. Geckil, S. Gencer (2004) Production of L-asparaginase in Enterobacter aerogenes expressing Vitreoscilla hemoglobin for efficient oxygen uptake. Appl. Microbiol. Biotechnol. 63, pp. 691–697.

[13] J. Holland (1975) Adaptation in natural and artificial systems, Ann Arbor: University of Michigan Press.

[14] D. Goldberg (1989) Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley.