International Journal of Natural and Engineering Sciences 2 (3): 51-56, 2008
ISSN: 1307-1149 www.nobel.gen.tr

Optimization Of The Fermentation Media Using Statistical Approach and Artifical Neural Networks for the Production of An Alkaline Protease from *Bacillus subtilis*.

Velaga PRASANTHI Murali Yugandhar NIKKU* Sandhya Poornima VUDDARAJU Kiran Kumar NALLA
Chaduvula Asha Immanuel RAJU Sri Rami Reddy DONTHIREDDY

Center for Biotechnology, Department of Chemical Engineering,

Andhra University College of Engineering (A) , Visakhapatnam-530 003, INDIA.

*Corresponding Author
E-mail: ynikku@gmail.com

Received: April 28, 2008
Accepted: June 05, 2008

ABSTRACT

Optimization of the fermentation medium for maximum alkaline protease production was carried out by *Bacillus subtilis* using Artificial Neural Networks (ANN) and Response Surface Methodology (RSM) and a predictive model was built, for the combined effects of independent variables (moisture content, concentration of carbon and nitrogen supplementation). Maximum alkaline protease produced was 701.9 U/ml with the application of RSM and the protease production further increased to 753.453 U/ml when ANN was used. The results demonstrated a higher prediction accuracy of ANN when compared to RSM. The domination of ANN over other multi factorial approaches would make this estimation technique a very helpful tool for fermentation monitoring and control.

Key Words: Alkaline protease, *Bacillus subtilis,* optimization, Response Surface Methodology, Artificial Neural Networks, back propagation.

INTRODUCTION

The extra cellular enzymes from microorganisms have attracted the researchers all over the world due to their board biochemical diversity, feasibility of mass culture and ease of genetic manipulation [8, 5]. Among all the other groups of enzymes proteolytic enzymes was the most important group which is produced commercially [1]. Among various proteases, bacterial proteases were the most significant enzymes when compared with animal and fungal proteases [3, 7]. Proteases occupy 60% of the world market of industrial enzymes, owing to their use in a wide range of different applications such as detergent, food, pharmaceutical, leather, silk, waste processing industries, protein recovery or solubilization and for recovery of silver from used x-ray films [1, 6, 9, 15]. Moreover alkaline proteases specifically account for nearly 25% of the world's enzyme market with a predominant share (35%) taken by detergents [2]. Recently solid state fermentation (SSF) has generated much interest due to its simplicity [7]; lower manufacturing costs by using unprocessed or moderately processed raw materials and the use of agro industrial wastes as substrates. When compared to submerged fermentation it requires less energy, low initial capital costs, low waste water output and improved product recovery [1, 4].

The "one at a time strategy" for improving the fermentation conditions are most frequently used in biotechnology to obtain high yields of the desired product. This approach is time consuming and also ignores the combined interactions between physico- chemical parameters. But response surface methodology (RSM) includes factorial design and regression analysis which helps in evaluating the effective factors, selection of the optimum conditions of variables for a desirable response and building models to study interactions [10].

An artificial neural network (ANN) is a superior and more accurate modeling technique when compared to the RSM method, as it represents the non linearities in a much better way [10]. However, surface and contour plots provide a good way to visualize the interactions between the independent and dependent variables. Therefore, both techniques are often used in unison for predicting optimum conditions for the production of various microbial products.

It is well known that extra cellular protease production in microorganism is greatly influenced by physical factors and the medium constituents like pH, moisture content, carbon and nitrogen supplementation. The present investigation is aimed at optimization of three variables (moisture content, concentration of carbon and nitrogen supplementation) which have been predicted to play a significant role in enhancing the production of alkaline protease from *Bacillus subtilis* using RSM and ANN. Extensive studies were performed on the effect of the initial values of connection weights on the accuracy of the back propagation learning method which was employed in the training of the artificial neural network. The effectiveness of the neural network with the proposed RSM technique was demonstrated and found that the neural network provided a more accurate prediction of the response.

MATERIALS AND METHODS

Microorganisms and Maintainance

The protease producing *Bacillus subtilis* (NCIM 2724), obtained from National Collection of Industrial Microorganisms, Pune, was used in the present study. The strain was grown and maintained on Nutrient Agar Medium (high medium) after cultivating for 24 h at 30 ^{0}C. Cultures were preserved at 4 ^{0}C for short term storage and were sub cultured every 4 weeks.

Substrate

The agro waste (green gram husk) material was collected from the local market and dried in hot air oven for 10 min at 100 ^{0}C. Then the husk was sieved using standard sieve set of nos. 14, 18, 25, 36, 44 and 52 to obtain mean particle sizes of 0.85, 0.6, 0.425, 0.355 and 0.3 mm and the material is stored for further use.

Inoculum preparation

The production of protease enzyme requires the preparation of inoculum. Culture was scraped off and washed from the slant culture with 10 ml sterile water and 2 ml of this inoculum was added to each of 250 ml flasks containing the production medium.

Solid State Fermentation

10 grams of substrate was taken in 250 ml Erlenmeyer conical flasks and to this a 4ml of water was added, mixed thoroughly and autoclaved at 121 ^{0}C for 15 min after cooling the flasks to room temperature, the flasks were inoculated with 2 ml of 24 h grown culture strain under sterile conditions. The inoculum was prepared by adding sterile distilled water to a 24 h old slant. The contents were mixed thoroughly and incubated in a slanting position to provide maximum surface area at 30 °C.

Enzyme extraction

After the incubation period, enzyme was extracted from the fermentation medium by mixing thoroughly with 50 mM glycine–NaOH buffer, pH 11 for 30 min and the extract was separated by squeezing through a cloth. This process was repeated three times and the extracts were pooled together and then centrifuged. The supernatant was used as an enzyme source for protease assay [12, 16].

Enzyme assay

Alkaline protease activity was estimated by the Anson–Hagihara method [11]. The enzyme (0.5 ml) was added to 3.0 ml casein (0.6 %w/v in 20 mM borax– NaOH buffer, pH 10) and the reaction mixture was incubated at 37 °C for 10 min before the addition of 3.2 ml of TCA mixture (0.11 M trichloroacetic acid, 0.22 M sodium acetate, 0.33 M acetic acid). The terminated reaction mixture was incubated for 30 min at room temperature. The precipitates were removed by filtration through Whatman No. 1 filter paper and the absorbance of the filtrate was measured at 280 nm. One unit of alkaline protease activity was defined as the amount of enzyme liberating 1 µg of tyrosine per minute under assay conditions. Enzyme units were measured using tyrosine (0–100 µg) as a standard.

EXPERIMENTAL DESIGN AND PROTEASE PRODUCTION

RSM consists of a group of empirical technology devoted to the evaluation of relations existing between a cluster of controlled experimental factors and the measured responses, according to one or more selected criteria. Prior knowledge and understanding of the process variables is necessary for achieving a realistic model. Based on the preliminary experiments results moisture content, concentration of carbon and nitrogen supplementation were found to be the major variables in protease production. Both RSM and ANN were used to study the interactive effects of the three variables i.e. moisture content, concentration of carbon and nitrogen supplementation for improving the total protease production. Experiments were conducted in triplicate and results were the average of these three independent trials.

Response Surface Methodology

Using RSM, the relationship among the variables, i.e., moisture content, concentration of carbon and nitrogen supplementation were expressed mathematically in the form of a quadratic polynomial model which gave the response as a function of relevant variables. The present work was based on Box-Behnken design which was utilized to obtain the experimental data, which would fit an empirical, full second order polynomial model representing the response surfaces over a relatively broad range of parameters. RSM has not only been used for optimization of culture parameters in the fermentation process but also for studying the combined effects of medium components.

The range and levels of experimental variables investigated in this study were presented in Table 1. The central values (zero level) chosen for experimental design were: 40% (v/w) moisture content (X_1), 2.5% (w/w) corn flour (X_2), 2.5% (w/w) yeast extract (X_3). The production of protease was optimized using Box-Behnken design [17], when protease production is related to independent variables by a response equation

$$Y = f(x_1, x_2, x_3, \ldots, x_k) \quad (1)$$

The true relationship between Y and x_k may be complicated and, in most cases, it is unknown; however a second-degree quadratic polynomial can be used to represent the function in the range interest

$$Y = R_0 + \sum_{i=1}^{k} R_i X_i + \sum_{i=1}^{k} R_i X_i^2 + \sum_{i=1,i<j}^{k-1} \sum_{j=2}^{k} R_j X_i X_j + \epsilon \quad (2)$$

where $X_1, X_2, X_3 \ldots X_k$ are the independent variables which affect the response Y, R_0, R_i, R_{ii} and R_{ij} (i=1-k, j=1-k) are the known parameters, € is the random error. A second order model is designed such that variance of Y is constant for all points equidistant from the center of the design.

The Box-Behnken design helps in investigating linear, quadratic and cross-product effects of these factors each varied at these levels and also includes three center points for replication. The design is performed because relations for experimental combination of the variables are adequate to estimate potentially complex response functions. The 'STATISTICA' software was used for regression and graphical analysis of the data obtained. The optimum values of the selected variables were obtained by solving the regression equation and also by analyzing the response surface plots.

Table 1. Independent Variables in the experimental plan

Variables	Coded levels		
	-1	0	+1
Moisture Content (% v/w) X_1	20	40	60
Corn Flour (%w/w) X_2	1.5	2.5	3.5
Yeast Extract (% w/w) X_3	1.5	2.5	3.5

Table 2. The Box-Behnken design matrix employed for three independent variables in coded units along with observed values and predicted values of both RSM and ANNs

Run No.	X_1	X_2	X_3	Protease Activity (μ/ml)		
				Observed values	Predicted by RSM	Predicted by ANN
1	20	1.5	2.5	230.2	267.1712	437.54
2	60	1.5	2.5	402.5	434.1437	402.5
3	20	3.5	2.5	436.5	404.8562	436.5
4	60	3.5	2.5	543.5	506.5487	543.5
5	20	2.5	1.5	290.6	276.8275	290.6
6	60	2.5	1.5	442.4	433.9350	442.4
7	20	2.5	3.5	389.3	397.6650	389.3
8	60	2.5	3.5	495.5	509.2225	641.79
9	40	1.5	1.5	411.5	388.2713	411.5
10	40	3.5	1.5	412.5	457.8663	535.42
11	40	1.5	3.5	496.3	450.8838	496.3
12	40	3.5	3.5	568.2	591.3788	568.2
13	40	2.5	2.5	684.5	683.4167	683.43
14	40	2.5	2.5	689.3	683.4167	683.43
15	40	2.5	2.5	676.5	683.4167	683.43

Artificial neural networks

An artificial neural network is a biologically inspired computational model formed from hundreds of single units, artificial neurons, connected with coefficients (weights) which constitute the neural structure. They are also known as processing elements (PE) as they process information. Each PE has weighed inputs, transfer function and one output.

ANNs analysis is quite flexible as regards to the amount and form of the training (experimental) data which makes it possible to use more informal experimental designs than with statistical approaches. Also, neural network models might generalize better than regression models since regression analyses are dependent on predetermined statistically significant levels. This means less significant terms are not included in the model. ANNs uses all the data potentially, making the models more accurate. ANNs along with RSM has been used to optimize the culture parameters for protease production from a newly isolated *Pseudomonas* sp. [10].

Usually a neural network in its basic form is composed of several layers of neurons, there being one input layer, one output layer and at least one hidden layer (Fig. 1). The use of at least one hidden layer enables the ANNs to describe nonlinear systems. A problem in constructing ANNs is to find the optimal number of hidden neurons.

W_{ij} is the weight-connection to neuron j from neuron i, x_i denotes the input values and bias_j is the bias of neuron j. The activation of the j^{th} neuron (Net_j) is defined as the sum of the weighted input signal to that neuron:

$$\text{Net}_j = \Sigma\, w_{ij}x + \text{bias}_j \quad (3)$$

This activation is transformed to the neuron output by a transform function. Different ANN classes use different definitions of the activation function. The most common transform function in back-propagation neural networks (BNNs) is a sigmoidal function:

$$y_j = \frac{2}{1 + e^{-\text{Net}_j}} - 1 \quad (4)$$

Each neuron in the input layer is connected to each neuron in the hidden layer and each neuron in the hidden layer is connected to each neuron in the output layer to produce the output vector. Information in a BNN is stored as weights, which are connections between neurons in successive layers and as bias values (neuron activation threshold). The neural network used in this work is the feed-forward, back-propagation neural network type, most often used in analytical applications. Information from various sets of inputs was fed forward through the BNN to optimize the weight between neurons, or to 'train' it. The error, or bias, in prediction is then propagated through the system and the inter-unit connections were changed to minimize the error in the prediction. This process is continued with multiple training sets until the error is minimized across many sets.

During training, neural techniques need to have some way of evaluating their own performance. Since they are learning to associate the inputs with outputs, evaluating the performance of the network on the training data may not produce the best results. If a network was left to train for too long, it will over train and will lose the ability to generalize. Thus, two types of data sets were used - training data: used to train network and test data: used to monitor the neural network performance during training. The MATLAB version 7.0 was used for neural network program.

RESULTS AND DISCUSSION

Production of protease enzyme from *Bacillus subtilis* (NCIM 2724) was conducted in solid state fermentation. Preliminary experiments on protease production from the above strain indicated that the most important factors were the moisture content, concentrations of corn flour and yeast extract. Hence these factors were considered as the independent variables and their effects on protease production were studied using a Box-Behnken design of RSM and back propagation of ANNs.

The results of Box- Behnken design experiments and ANNs for studying the effects of three independent variables, viz., moisture content, concentrations of corn flour and yeast extract, on protease production are presented in Table 2 along with the predicted and observed responses. The application of RSM [13] yielded the following regression equation, which is an empirical relation ship between the enzyme yield and test variables in coded units

$$Y = 426.6 + 67.2X_1 + 52.5X_2 + 11.4X_1X_3 + 17.7\ X_2X_3 + 87.0X_{11} + 53.1X_{22} + 52.5X_{33} \quad (5)$$

Where Y= enzyme yield

X_1, X_2, X_3 are the coded values of the test variables moisture content, concentrations of corn flour and yeast extract.

Statistical testing of the model was done by the Fischer's statistical test for analysis of variance (ANOVA) and the results are shown in Table 3. The calculation of regression analysis gives the value of the determination coefficient (R^2 = 0.96) indicates that only 4.0 % of the total variations are not explained by the model and the F-value of 65.77 indicates that protease production by *Bacillus subtilis* has a good model fit due to the high values of R^2 and F. The p-values are used

as a tool to check the significance of each coefficient, which also indicate the interaction strength between each independent variable. The smaller the p-values, the bigger the significant of the corresponding coefficient [14].

The model equation (5) indicated that moisture content (X_1) had a significant effect ($p<0.01$) on Y and it had the largest coefficient followed by corn flour (X_2) and yeast extract (X_3). The statistical analysis of the design shows a high precision of the polynomial model that reflects high degree of fitting between the predicted and the experimental data. This great similarity between the predicted and the observed results reflects the accuracy and applicability of the Box- Behnken model in the optimization processes.

The relationship between coded variables and responses can be better understood by examining the series of surface plots (Figs. 2-4). These response surfaces display the variation of two factors while the third is kept at the optimum level.

The analysis revealed a maximum protease yield of 701.9 U/ml which was 1.83 % more than the value (683.93 U/ml) obtained with the initial experiments, at the points where moisture content, concentrations of corn flour and yeast extract were 43.46 %v/w, 2.75 %w/v and 2.84 %w/v respectively.

Further, ANN methodology was applied to provide a non-linear mapping between input variables (moisture content, concentrations of corn flour and yeast extract) and the out put variable (protease yield) for the runs were reported in Table 2. The type of ANNs chosen was the back propagation network having a feed forward structure. The simulated values of the response (protease yield) were listed in the last column of Table 2.

Figure 1: The Neural network topology with Single hidden layer

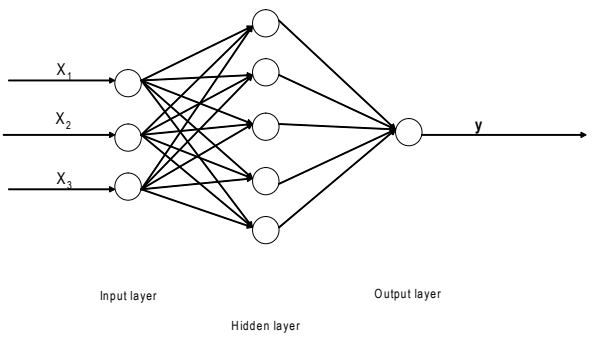

Table 3: Analysis of Variance (ANOVA) for the quadratic model

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-value	P-value
$X_1+X_1X_1$	147830	2	73915.02	35.09	0.0011
$X_2+X_2X_2$	63770.9	2	31885.44	15.14	0.0075
$X_3+X_3X_3$	59972.8	2	29986.42	14.23	0.0086
X_1X_2	1065.7	1	1065.7	0.5	0.508
X_1X_3	518.7	1	518.7	0.24	0.640
X_2X_3	1256.7	1	1256.7	0.57	0.474
Error	10530.1	5	2106.01		
Total SS	261542.5	14			

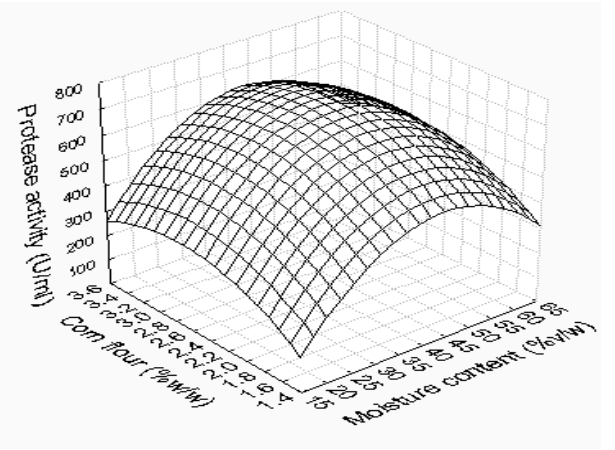

Figure 2. Effect of moisture content, corn flour on protease production by *Bacillus subtilis*.

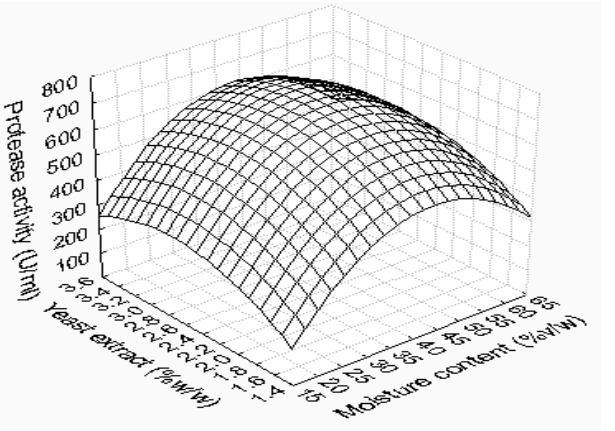

Figure 3. Effect of moisture content, yeast extract on protease production by *Bacillus subtilis*.

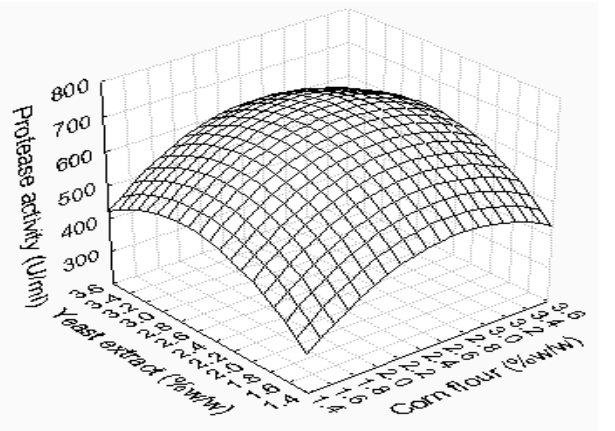

Figure 4. Effect on corn flour, yeast extract on protease production by *Bacillus subtilis*.

The configuration of the neural network developed in this work (a 3-14-1 structure: three input neurons- fourteen neurons in hidden layer- one out put neuron) was determined by trail and error and the topology of network was shown in Fig 5.

The transfer functions used in the neural networks were 'transig' and 'purelin' at the hidden layer and out put layer respectively. 'Newff' function was used for the training of the neural networks. The training function 'trainlm' was used in this work. The following equation was the outcome of the

neural network training, relating the input variables (x_1, x_2 and x_3) to the out put variable, **y**, in terms of weights and biases.

$$\mathbf{y} = \mathbf{w_2}^* (2. / (1.0+\exp (-2*(\mathbf{w_1}*\mathbf{xt}^1+\mathbf{b_1}))) -1) +b_2 \text{ -------- (6)}$$

Where $\mathbf{w_1}$ and $\mathbf{w_2}$ were the weights, $\mathbf{b_1}$ and $\mathbf{b_2}$ were the biases whose values are reported in Table 14. '**y**' is the predicted value from the neural network and **xt** was the row vector of 3 independent variables (x_1, x_2 and x_3) , while $\mathbf{xt}^1$ represent the transpose of the vector with a dimension of (3x1); x_1, x_2 and x_3 represent moisture content, concentrations of corn flour and yeast extract respectively. For any given set of x_1, x_2 and x_3, protease yield (**y** value) can be predicted using the above equation.

The equation (6) represent the out put, **y** (protease yield), for the given set of independent variables represented in '**xt**' when 'transig' was used as the transfer function in the hidden layer and 'purelin' was used as the transfer function in the out put layer. The input data of the independent variables were transformed between -1 and +1 using the built-in function 'premmnf' prior to neural network training while 'postmnmx' was used to transform back the optimized set of independent variables into the original scale, after the global optimization method was applied. However, the output data was used without any transformation. The simulated values of d yield as predicted by equation (6) were in the close agreement with those of the experimental values as evident from the values in the Table 2.

The optimum protease yield (753.453 U/ml) predicted by ANNs is higher than the value predicted by Box-Behnken design; however, it is 10.16 % higher than the value obtained from preliminary runs (683.93 U/ml). At these points where moisture content, concentrations of corn flour and yeast extract were 52.34 %v/w, 2.27 %w/w and 2.54 %w/w, respectively were found to be optimum. This demonstrates the superiority of ANNs represented by equation (6) over the RSM represented by equation (5) in the fermentation medium.

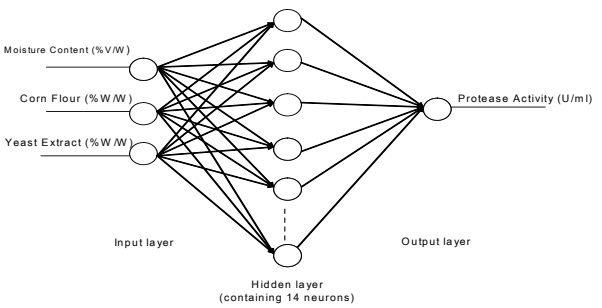

Figure 5: The neural network topology with single hidden layer having fourteen neurons.

CONCLUSIONS

This work has demonstrated the performance of the Box-Behnken design of RSM and back propagation of ANN in determining the conditions leading to the maximum yield of protease production. Both RSM and ANN estimated the fermentation performance parameters (moisture content, concentrations of corn flour and yeast extract) and provided quality predictions for the three independent variables in terms of protease production. ANN models performed and generalized better than the RSM models in giving the critical concentrations of the parameters which are optimum, resulting in maximum protease production. The optimum protease yield (753.453 U/ml) predicted by ANNs is higher than the value (701.9 U/ml) predicted by Box-Behnken design; however, it was 10.16 % higher than the value obtained by the conventional method at a time optimization (683.93 U/ml). Thus, ANN could be a very powerful and a flexible tool which is well suited for modeling the fermentation process.

REFERENCES

[1] Fikret Uyar, Zübeyde Baysal. 2004. Production and optimization of process parameters for alkaline protease production by a newly isolated *Bacillus* sp. under solid state fermentation. Process Biochemistry. 39:1893–1898.

[2] Ganesh Kumar C, Han-Seung Joo, Yoon-Mo Koo, Seung Paik R, Chung-Soon Chang. 2004. Thermostable alkaline protease from a novel marine haloalkalophilic *Bacillus clausii* isolate. World Journal of Microbiology & Biotechnology. 20:351–357.

[3] Ellaiah P, Adinarayana K, Siva Prasad D. 2003. Purification and partial characterization of thermostable serine alkaline protease from a newly isolated *Bacillus subtilis* PE-11. J. AAPS Pharm Sci Tech. 5(3):272-278.

[4] Murat Elibol, Antonio R Moreira. 2005. Optimizing some factors affecting alkaline protease production by a marine bacterium *Teredinobacter turnirae* under solid substrate fermentation. Process Biochemistry. 40:1951–1956.

[5] Rajesh Patel, Mital Dodia, Satya Singh P. 2005. Extracellular alkaline protease from a newly isolated haloalkaliphilic *Bacillus* sp.: Production and optimization. Process Biochemistry. 40:3569–3575.

[6] Folasade Olajuyigbe M, Joshua Ajele O. 2005. Production dynamics of extracellular protease from *Bacillus* species. African Journal of Biotechnology. 4(8):776-779.

[7] Valeria Soares F, Leda Castilho R, Elba Bon P S, Denise Freire M G. 2005. High-Yield *Bacillus subtilis* protease production by solid-state fermentation. Applied Biochemistry and Biotechnology. 121–124.

[8] Shikha, Adhyayan Sharan, Nandan Darmwal S. 2007. Improved production of alkaline protease from a mutant of alkalophilic *Bacillus pantotheneticus* using molasses as a substrate. Bioresource Technology. 98:881-885.

[9] Ellaiah, P, Adinarayana, K. 2002. Response surface optimization of the critical medium components for the production of alkaline protease by a newly isolated *Bacillus* sp., J. Pharm Pharmaceut Sci. 5(3):272-278.

[10] Jayati R D, Pranab K D, Rintu B. 2004. Optimization of culture parameters for extra cellular protease production from a newly isolated *Pseudomonas* Sp. under response surface and artificial neural network models. Process Biochemistry. 39:2193-2198.

[11] Patel R K, Dodia, M S, Joshi R H , Singh S P. 2006. Production of extracellular halo-alkaline protease from a

newly isolated haloalkaliphilic *Bacillus* sp. isolated from seawater in Western India. World Journal of Microbiology & Biotechnology. 22: 375–382.

[12] Box G E P, Hunter W G, Hunter J S. 1978. Statistics for Experiments. Wiley 291-334. New York.

[13] Cui F J, Li F, Xu Z H, Xu H Y, Sun K, Tao W Y. 2006. Optimization of the medium composition for the production of mycelial biomass and exo-polymer by *Grifola frondosa* GF 9801 using response surface methodology. Bioresour Technol. 97:1209-1216.

[14] Khuri A L, Cornell J A. 1987. Response surfaces: Design and Analysis, Marcel Dekker. New York.

[15] Prakasham R S, Subba Rao Ch, Sarma P N. 2006. Green gram husk—an inexpensive substrate for alkaline protease production by *Bacillus* sp. in solid-state fermentation. Bioresource Technology. 97:1449–1454.

[16] Box G E P, Behnken D W. 1960. Three level design for the study of quantitative variables. Technometrics. 2:455-475.

[17] Annadurai G, Sheeja R Y. 1998. Use of Box-Behnken design experiments for the adsorption of Verofix red using biopolymer. Bioprocess Engg. 18:463-466.