Korean J. Chem. Eng., 34(4), 1109-1121 (2017)
DOI: 10.1007/s11814-016-0358-1

pISSN: 0256-1115
eISSN: 1975-7220

Medium optimization for high yield production of extracellular human interferon-γ from *Pichia pastoris*: A statistical optimization and neural network-based approach

Ashish Anand Prabhu, Bapi Mandal, and Veeranki Venkata Dasu†

Biochemical Engineering Laboratory, Department of Bioscience and Bioengineering,
Indian Institute of Technology, Guwahati 781039, Assam, India
(*Received 5 May 2016 • accepted 23 December 2016*)

Abstract−Medium development for high level expression of human interferon gamma (hIFN-γ) from *Pichia pastoris* (GS115) was performed with the aid of statistical and nonlinear modeling techniques. In the initial screening, gluconate and glycine were found to be key carbon and nitrogen sources, showing significant effect on production of hIFN-γ. Plackett-Burman screening revealed that medium components., gluconate, glycine, KH_2PO_4 and histidine, have a considerable impact on hIFN-γ production. Optimization was further proceeded with Box-Behnken design followed by artificial neural network linked genetic algorithm (ANN-GA). The maximum production of hIFN-γ was found to be 28.48 mg/L using Box-Behnken optimization (R^2=0.98), whereas the ANN-GA based optimization had displayed a better production rate of 30.99 mg/L (R^2=0.98), with optimal concentration of gluconate=50 g/L, glycine=10.185 g/L, KH_2PO_4=35.912 g/L and histidine 0.264 g/L. The validation was carried out in batch bioreactor and unstructured kinetic models were adapted. The Luedeking-Piret (L-P) model showed production of hIFN-γ was mixed growth associated with the maximum production rate of 40 mg/L of hIFN-γ production.

Keywords: *Pichia pastoris*, Human Interferon Gamma, Plackett-Burman, Box-Behnken, Artificial Neural Network-Genetic Algorithm (ANN-GA)

INTRODUCTION

With rapid outbreak of epidemics worldwide, there is a huge demand for producing an extensive range of therapeutic proteins to remediate chronic diseases. Encompassed by several therapeutic proteins, Interferon Gamma (IFN-γ) is found to be one such protein that has a broad range of biological activity, such as antigen presenting, lysosome activity of macrophages, stimulation of antiviral and antiparasitic activity, promoting adhesion and binding of leukocytes and displaying a cell proliferation and apoptosis effect [1]. IFN-γ is associated with class II type cytokine, majorly secreted by natural killer (NK), natural killer T (NKT) cells as well as by the CD4 and CD8 cytotoxic T1 lymphocytes. The mature human interferon gamma (hIFN-γ) consists of 143 amino acids (17 kDa) protein majorly comprised of lysine and arginine residues. It has been approved for the treatment of chronic granulomatous disease and severe malignant osteopetrosis by FDA [2].

There are several prominent hosts used for heterologous protein production. Among them, *Pichia pastoris*, a methylotrophic yeast, is one of the most extensively studied eukaryotic system apart from *Saccharomyces cerevisiae* and has become most convenient and versatile host system for the production foreign proteins under tightly regulated methanol inducible alcohol oxidase 1 promoter (pAOX1) [3,4]. It has the advantage of post translational modification, which includes glycosylation and disulfide bond formation [5]. The genome of *P. pastoris* is relatively simple to manipulate, and foreign proteins can be directed towards secretory pathways by signal peptide, which eases the downstream processing. More than 500 proteins are reported to be successfully expressed in *Pichia pastoris*; with all its convenience, *Pichia pastoris* is now drawing attention from several industries and academicians [6,7].

To date, only defined basal salts medium (BSM) or modified BSM media known as FM22 have been extensively used for the high cell density cultivation of *Pichia pastoris*; as per the guidelines proposed by Invitrogen (USA), these media result in about gram per liter levels of a recombinant protein yield. The major issue related with this media is the formation of precipitates at higher ionic strength [4]. These precipitates have adverse effects on fermentation, leading to unbalanced nutrient supply, contamination of secreted products with intracellular materials due to abrasive nature, cell disruption, clog gas-sparger of fermentor, interfering with measurement of cell density and cause tedious downstream processing [8]. Hence, it is necessary to develop a physiologically rational and suitable medium for recombinant protein production in *P. pastoris*. Regarding this context, only few studies have been carried out till date. Zhang et al. [9] used glycerophosphate as a phosphorus source and compared the production of Interferon-τ with the organism grown in FM22 medium. Ghosalkar et al. [10] used minimal medium designed for the growth of *S. cerevisiae* and optimized it for production of biomass in *P. pastoris*. Since, central carbon metabolism for both the yeasts is similar and metabolic flux ratio profiles for amino acid biosynthesis are also similar [11].

†To whom correspondence should be addressed.
E-mail: veeranki@iitg.ernet.in

Moreover, a successful expression of any heterologous protein relies on the carbon source used. In *Pichia pastoris* fermentation, glycerol is the most extensively used carbon source added together with methanol during induction phase. With this combination the volumetric productivity will be enhanced, but the specific productivity of heterologous protein may get decreased as excess glycerol may tend to repress the AOX promoter, thus limiting the expression level [12]. Hence there is a requirement of an alternative carbon source that supports growth but does not repress the AOX I promoter activity. Just as glycerol, sorbitol is another widely used nonrepressive carbon source that has shown a similar level of expression for foreign protein [13]. Alternatively other carbon sources such as Mannitol, alanine, trehalose and lactic acid are also reported to be employed as nonrepressible carbon sources in various studies of *Pichia pastoris* [14]. But there is no report on the use of sodium gluconate as a carbon source for heterologous protein production in *Pichia pastoris.* Furthermore, it is very well known that the transcription of the carbohydrate metabolism genes are affected by the quality of nitrogen source and inorganic phosphate source used in the media [15]. *Pichia pastoris* can be able to utilize vast array of nitrogenous compounds and among them ammonium sulfate and glutamine are the most preferred ones. For the metabolism of nitrogen compounds, they have developed special regulation mechanisms that provide preemptive absorption of these compounds. Under this condition, the preferential nitrogen sources are not available in the medium; yeast cells try to switch their metabolic pathways and begin to utilize poor nitrogen sources like urea and proline. Since, the genes which control proline and other amino acids catabolism and related permeases are regulated by nitrogen catabolite repression, which leads to low expression of foreign gene [16].

Media optimization is one of the crucial methodologies applied for increased yield of fermentative products at the industrial level. Classical optimization with one factor at a time method is time consuming, which tends to overlook the effects of interaction among the factors and might lead to misinterpretation of results. In contrast, statistical approach has been considered the most effective method for media optimization and there is ample amount of literature available on various statistical approaches [17]. Plackett-Burman, response surface methodology (RSM) and factorial designs are examples of such means. However, other mathematical methodologies such as artificial neuron network (ANN) coupled with GA have also become remarkably successful in the last few years, as these system persist high predicting capabilities of nonlinear functions [18]. Plackett-Burman is a two level factorial design system used for rapid and efficient screening of numerous significant factors by using least number of experiments [19]. In addition, RSM is used as an adequate experimental tool for the determination of optimal conditions in a multivariable system [20]. To optimize nonlinear based systems, more advanced techniques like ANN have been used in recent years, as this system mimics the structure of biological networks called as a neuron. A neuron receives a signal from a source; these signals are operated through nonlinear functions to receive an appropriate output. The network was created for defining a function approximation using back-propagation algorithm, which utilizes the experimental data for underlying a training framework. Genetic Algorithm (GA), which is a heuristic method for determining a global solution and is often coupled with ANN for achieving a precise output. These methods incorporate a stochastic search algorithm, which generates a new population from old ones. To figure out a new population, it uses operators like selection, crossover and mutation on a primarily random population [21].

In our previous study, we showed the effect of chaperon and codon optimization on expression level of hIFN-γ in *Pichia pastoris*; also, we emphasized on optimizing various parameters that affected the expression level of hIFN-γ [22]. In this context, the rationale behind the present study was to develop an appropriate medium composition for enhanced production of hIFN-γ, with the aid of statistical and artificial intelligence methodology.

MATERIALS AND METHODS

1. Strain and Media

Recombinant *P. pastoris* strain GS115/Mut$^+$/*hIFN-*γ^{opt}, expressing human interferon gamma under the control of alcohol oxidase promoter, was used for optimization studies [22]. Stock culture was maintained on YPD agar plates (yeast extract 10 g/L, peptone 20 g/L, dextrose 20 g/L and agar 20 g/L).The production of recombinant human interferon gamma (rhIFN-γ) was studied in the modified FM22 media containing (gram per liter): 40 carbon source; 42.9 KH_2PO_4, 5 nitrogen source, 1.0 $CaSO_4{\cdot}2H_2O$, 14.3 K_2SO_4, 11.7 $MgSO_4{\cdot}7H_2O$ 1 ml/L vitamins solution and 4 ml/L trace elements solution (PTM 4) composition of PMT4 (gram per liter): 2.0 $CuSO_4{\cdot}5H_2O$, 0.08 NaI, 3.0 $MnSO_4{\cdot}H_2O$, 0.2 $Na_2MoO_4{\cdot}2H_2O$, 0.02 H_3BO_3, 0.5 $CaSO_4{\cdot}2H_2O$, 0.5 $CoCl_2$, 7 $ZnCl_2$, 22 $FeSO_4{\cdot}7H_2O$, 0.2 biotin, 1 mL conc. H_2SO_4. Composition of vitamins solution (gram per liter) was 0.05 D-biotin, 1.00 Ca D-panthothenate, 1.00 nicotinic acid, 25.0 myo-inositol, 1.00 thiamin hydrochloride, 1.00 pyridoxol hydrochloride and 0.20 p-amino benzoic acid. Vitamins and trace metal solutions were filter sterilized separately and then the whole medium was aseptically reconstituted. Finally, the pH was set at 5 using 1 N KOH prior to inoculation. The inoculum was prepared by inoculating a single clone of GS115/pPICZαA-hIFN-γ^{opt} in 25 mL BMGY (having 1% yeast extract, 2% peptone, 1% glycerol, 1.34% YNB (w/o amino acids), 4×10^{-5}% biotin, and 100 mM potassium phosphate, pH 6.0) and incubated at 30 °C at 250 rpm for 24 h. The cells were harvested by centrifugation at 3,000 ×g for 10 min at room temperature and two percent of the inoculum from the above-said culture was added to 50 mL of the medium in 250 mL baffled flask. The flask was incubated in a shaking incubator at 25 °C, 250 rpm, which was optimized in our previous study [22]. Samples were withdrawn at regular time intervals and measured for hIFN-γ production. Experiments were conducted in triplicate, and ELISA for hIFN-γ was performed in triplicate for each sample.

2. Primary Screening of Carbon and Nitrogen Sources and Their Effect on hIFN-γ Production

Eight carbon sources (sorbitol, mannitol, gluconate, lactose, glycerol, whey, galactose, and maltose) and six nitrogen sources (ammonia, urea, glutamate, glycine, and ammonium sulfate and sodium nitrite) were screened based on the hIFN-γ production in the abovementioned media. The effects of carbon and nitrogen sources are the values stated in reference to a base yield from unmodified medium.

3. Effect of Aeration on Protein Expression

Influence of aeration on the protein expression was studied by cultivating cells with modified FM22 media with above screened carbon (gluconate) and nitrogen (glycine) source in baffled and non-baffled flasks.

4. Effect of Casamino Acid on Protein Expression

Effect of casamino acid supplementation in the medium was studied by comparing the medium comprising 1% w/v casamino acid and the media without casamino acid.

5. Screening of Significant Medium Components by the Plackett-Burman Experimental Design Technique

To screen the significant medium components that influence hIFN-γ production, Plackett-Burman experimental design was adapted [23]. A total of nine parameters, gluconate, glycine, KH_2PO_4, $MgSO_4 \cdot 7H_2O$, histidine, trace elements, vitamins, EDTA and triton X-100, were considered for the screening experiments. High and low levels of each variable were denoted by (+1) and (−1), respectively (Table 1). According to Plackett-Burman design, 12 experiments were performed. The levels of variables and design matrix in the coded levels and real values are shown in Table 2. The Plackett-Burman experimental design is based on a first-order polynomial model:

Table 1. Coded values of independent variables for Plackett-Burman screening

Variable	Symbol code	Levels	
		Low (−1)	High (+1)
Gluconate (g/L)	X_1	20	80
Glycine (g/L)	X_2	2	20
KH_2PO_4 (g/L)	X_3	20	60
$MgSO_4 \cdot 7H_2O$ (g/L)	X_4	2	20
Trace elements (ml/L)	X_5	2	10
Vitamins (ml/L)	X_6	1	5
Histidine (mg/L)	X_7	0.2	1
EDTA (mM)	X_8	2	10
Tritonx-100 (%)	X_9	0.01	0.1

Table 2. Plackett-Burman design matrix in coded units and real values (in parenthesis) along with the observed and predicted for hIFN-γ production

Run order	Variables and their levels									hIFN-γ (mg/L)	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	[a]Observed	Predicted
1	80(+1)	2(−1)	60(+1)	2(−1)	2(−1)	1(−1)	1(+1)	10(+1)	0.1(+1)	11.44±0.09	10.98
2	80(+1)	20(+1)	20(−1)	20(+1)	2(−1)	1(−1)	0.2(−1)	10(+1)	0.1(+1)	6.58±0.01	8.35
3	20(−1)	20(+1)	60(+1)	2(−1)	10(+1)	1(−1)	0.2(−1)	2(−1)	0.1(+1)	4.91±0.12	4.28
4	80(−1)	2(−1)	60(+1)	20(+1)	2(−1)	5(+1)	0.2(−1)	2(−1)	0.01(−1)	3.74±0.03	5.01
5	80(+1)	20(+1)	20(−1)	20(+1)	10(+1)	1(−1)	1(+1)	2(−1)	0.01(−1)	9.03±0.06	9.3
6	80(+1)	20(+1)	60(−1)	2(−1)	10(+1)	5(+1)	0.2(−1)	10(+1)	0.01(−1)	2.30±0.68	1.76
7	20(−1)	20(+1)	60(+1)	20(+1)	2(−1)	5(+1)	1(+1)	2(−1)	0.1(+1)	8.63±0.17	9.99
8	20(−1)	2(−1)	60(+1)	20(+1)	10(+1)	1(−1)	1(+1)	10(+1)	0.01(−1)	9.21±0.23	10.39
9	20(−1)	2(−1)	20(−1)	20(+1)	10(+1)	5(+1)	0.2(−1)	10(+1)	0.1(+1)	14.38±0.9	14.65
10	80(+1)	2(−1)	20(−1)	2(−1)	10(+1)	5(+1)	1(+1)	2(−1)	0.1(+1)	16.31±0.1	16.19
11	20(−1)	20(+1)	20(−1)	2(−1)	2(−1)	5(+1)	1(+1)	10(+1)	0.01(−1)	14.66±0.8	14.62
12	20(−1)	2(−1)	20(−1)	2(−1)	2(−1)	1(−1)	0.2(−1)	2(−1)	0.01(−1)	12.12±0.1	12.16

[a]The observed values of hIFN-γ concentration, were the mean values of triplicates with standard deviation (mean±SD)

Table 3. Statistical analysis of the Plackett-Burman design showing coefficient, t, and P values for each variable

Term	Symbol code	Effect	Coef	T	P
Constant			9.447	0.2317	0.001[a]
Gluconate (g/L)	X_1	−2.418	−1.209	−5.22	0.035[a]
Glycine (g/L)	X_2	−3.513	−1.757	−7.58	0.017[a]
KH_2PO_4 (g/L)	X_3	−5.475	−2.738	−11.82	0.007[a]
$MgSO_4 \cdot 7H_2O$ (g/L)	X_4	−1.696	−0.848	−3.66	0.067[b]
Trace elements (ml/L)	X_5	−0.171	0.2317	−0.37	0.747[b]
Vitamins (ml/L)	X_6	1.125	0.562	2.43	0.136[b]
Histidine (mg/L)	X_7	4.21	2.105	9.08	0.012[a]
EDTA (mM)	X_8	0.638	0.319	1.38	0.302[b]
Tritonx-100 (%)	X_9	1.865	0.932	4.02	0.057[b]

R^2=99.42% R^2(pred)=79.21% R^2(adj)=96.82%

[a]Significant

[b]Nonsignificant at P>0.05

$$Y=\beta_0+\Sigma\beta_i X_i \quad (1)$$

where, Y is the response (hIFN-γ production), β_0 is the model intercepts, β_i is the linear coefficient, and X_i is the level of the independent variable. The standard error (S.E.) of the concentration effect is the square root of the variance of an effect; and the significance level (p-value) of each concentration effect is determined by using Student's t-test shown in Eq. (2):

$$t_{xi}=\frac{E(X_i)}{S.E} \quad (2)$$

where, E (X_i) is the effect of variable X_i.

All experiments were performed in duplicates and the data was represented as mean±SD. The variables, with confidence levels greater than 95% were considered to be significantly influencing hIFN-γ production and were further optimized using Box-Behnken Design (Table 3).

6. Optimization of Screened Components by Box-Behnken Design

To maximize the production of rhIFN-γ we employed a Box-Behnken factorial design consisting of four factors and three levels [24]. The model included three replicated center points and the set of points lying at the midpoints of each edge of the multidimensional cube that defined the region of interest and were used for fitting a second order response surface. The effect of the screened medium constituents, gluconate, glycine, KH_2PO_4, and histidine, on the expression level of hIFN-γ were evaluated using this experimental design. The four screened variables were designated as X_1, X_2, X_3, X_7, and hIFN-γ production was designated as Y, which is a response. The levels of each variable are shown in Table 4. Twenty-seven experiments were performed with three replications at the center points to evaluate the pure error. The total number of experi-

Table 4. Coded values of independent variables for box benkhen design

Indepenent variable	Symbol code	Coded value		
		−1	0	1
Gluconate (g/L)	X_1	20	50	80
Glycine (g/L)	X_2	2	11	20
KH_2PO_4 (g/L)	X_3	20	40	60
Histidne (g/L)	X_7	0.2	0.6	1

Table 5. Box-Behnken design matrix with un-coded and coded values along with observed and Predicted response for hIFN-γ production

Run order	Variables and their level				hIFN-γ (mg/L) (Y)		ANN predicted (mg/L)
	X_1	X_2	X_3	X_7	[a]Observed	Predicted	
1	20(−1)	2(−1)	40(0)	0.6(0)	10.50±3.41	11.39	10.51
2	80(+1)	2(−1)	40(0)	0.6(0)	8.47±0.05	8.91	8.47
3	20(−1)	20(+1)	40(0)	0.6(0)	8.16±0.79	8.68	8.16
4	80(+1)	20(+1)	40(0)	0.6(0)	3.55±0.39	3.62	3.55
5	50(0)	11(0)	20(−1)	0.2(−1)	11.75±2.36	12.01	11.02
6	50(0)	11(0)	60(+1)	0.2(−1)	9.18±0.05	9.88	9.18
7	50(0)	11(0)	20(−1)	1(+1)	17.55±1.00	17.82	19.01
8	50(0)	11(0)	60(+1)	1(+1)	8.42±1.34	9.12	8.42
9	20(−1)	11(0)	40(0)	0.2(−1)	9.14±1.61	9.86	9.14
10	80(+1)	11(0)	40(0)	0.2(−1)	5.16±0.05	5.15	5.16
11	20(−1)	11(0)	40(0)	1(+1)	10.85±0.48	11.45	10.85
12	80(+1)	11(0)	40(0)	1(+1)	8.73±0.36	8.61	8.73
13	50(0)	2(−1)	20(−1)	0.6(0)	16.65±0.16	17.34	16.65
14	50(0)	20(+1)	20(−1)	0.6(0)	11.19±3.13	11.25	11.19
15	50(0)	2(−1)	60(+1)	0.6(0)	9.29±0.28	9.84	9.3
16	50(0)	20(+1)	60(+1)	0.6(0)	8.02±0.28	7.92	8.02
17	20(−1)	11(0)	20(−1)	0.6(0)	16.54±0.17	15.31	14.45
18	80(+1)	11(0)	20(−1)	0.6(0)	3.50±0.43	3.45	3.51
19	20(−1)	11(0)	60(+1)	0.6(0)	3.33±0.68	1.81	1.53
20	80(+1)	11(0)	60(+1)	0.6(0)	6.46±0.28	6.12	6.46
21	50(0)	2(−1)	40(0)	0.2(−1)	9.94±1.9	8.6	9.94
22	50(0)	20(+1)	40(0)	0.2(−1)	16.58±2.13	16.24	16.58
23	50(0)	2(−1)	40(0)	1(+1)	24.00±1.23	22.77	24.01
24	50(0)	20(+1)	40(0)	1(+1)	7.35±0.09	7.12	6.82
25	50(0)	11(0)	40(0)	0.6(0)	27.18±2.40	26.86	26.71
26	50(0)	11(0)	40(0)	0.6(0)	24.94±2.74	26.86	26.71
27	50(0)	11(0)	40(0)	0.6(0)	28.48±3.39	26.86	26.71

[a]The observed values of hIFN-γ concentration, were the mean values of triplicates with standard deviation (mean±SD)

ments was calculated from Eq. (3) [25].

$$N=2K(K-1)+C_0 \quad (3)$$

where K is the number of factors and C_0 is the number of central point.

The minimum and maximum ranges of the variables and the full experimental plan with regards to their values in actual and coded form are provided in Table 5. This methodology allowed the modelling of a second-order equation that described the interaction of the process variables on the response (objective function). rhIFN-γ production was analyzed by multiple regression through the least squares method to fit Eq. (4):

$$Y=\beta_0+\Sigma\beta_i X_i+\Sigma\beta_{ij}x_i x_j+\Sigma\beta_{ii}x_i^2 \quad (4)$$

where, Y is the measured response variable; β_0, β_i, β_{ij}, β_{ii} are constant and regression coefficients of the model, and x_i, x_j represent the independent variables in coded values. The coding was done by Eq. (5)

$$x_i=(X_i-\overline{X}_i)/\Delta X_i \quad (5)$$

where, x_i is the coded value of an independent variable, X_i is the real value of an independent variable, $\overline{X}_i$ is the real value of an independent variable at the centre point, and ΔX_i is the step change value. The optimum conditions were verified by conducting validation experiments. Responses were monitored and results were compared with model predicted [21,26]. The fitted polynomial equation was expressed as response plots using the MINITAB (version 16) software to visualize the relation between the response and experimental levels of each factor and to deduce the optimum condition.

7. Artificial Neural Network Linked with Genetic Algorithm (ANN-GA) as a Modelling and Optimization Tool

We used the input and output data of Box-Behnken to train ANN (Table 5). The total experimental data was divided into three different sets 21, 3 and 3 and these were used for training, validating and testing, respectively. The data used for training would compute the network parameters, robustness of the parameters was checked by the validation data. While running a data, if a network trains too well then training data rules might fit for the overall data. To avoid this overfitting of data, testing data was used to control error, which stopped when the error was increased. The testing data was used to assess the predictive ability of the generated model [21,26].

We adapted a multilayer perceptron feed forward neural network, which consisted of three layers: input, hidden and output. The process variables, gluconate, glycine, KH_2PO_4 and histidine were considered as input layer, while the concentration of hIFN-γ was considered as output layer. These layers were interconnected via weights (w) (Real number quantity associated with the connection between two neurons) and Biases (b) that were considered to be parameters of the neural network (NN). The neurons in the input layer simply introduced the scaled input data via w to the hidden layer. In the network architecture proposed, the data flowed in forward direction that was from input to output via hidden layer. The neurons in the hidden layer carried out two major functions [27]. First, they summed up all the weighted input to the neurons including biases; Eq. (6) is given as

$$\Sigma_{i=1}^{n} x_i w_i + b \quad (6)$$

where, w_i (i=1 to n) were the connection weights, b is called bias and x_i is the input parameter.

The summation of weighted output was passed through the transfer function. In the present study, tansig was used as transfer function between input and hidden layer, and the output produced by the hidden layer became the input to the output layer where purelin transferred functions and produced output same as hidden layer [28]. Eqs. (7) and (8) for purelin and tansig, respectively, are given below:

$$\text{purelin (sum)}=\text{sum} \quad (7)$$

$$\text{tansig}=\frac{1-\exp(-\text{sum})}{1+\exp(-\text{sum})} \quad (8)$$

The error function was calculated based on the difference between actual output and predicted output. ANN is a iterative method which is pre-specified to minimize error function and adjust weight appropriately [27]. The commonly used error function was the mean squared error (MSE), which was used in the present study and is given by Eq. (9):

$$MSE=\frac{1}{N}\Sigma_{i=1}^{N}(Y_a-Y_p)^2 \quad (9)$$

where, Y_a is the actual output, Y_p is the predicted output and N is the number of data points. In ANN, there are several algorithms used, but the most commonly used algorithm in feed forward neural network is back propagation method [28]. Back propagation is an iterative optimization method where the MSE is minimized by

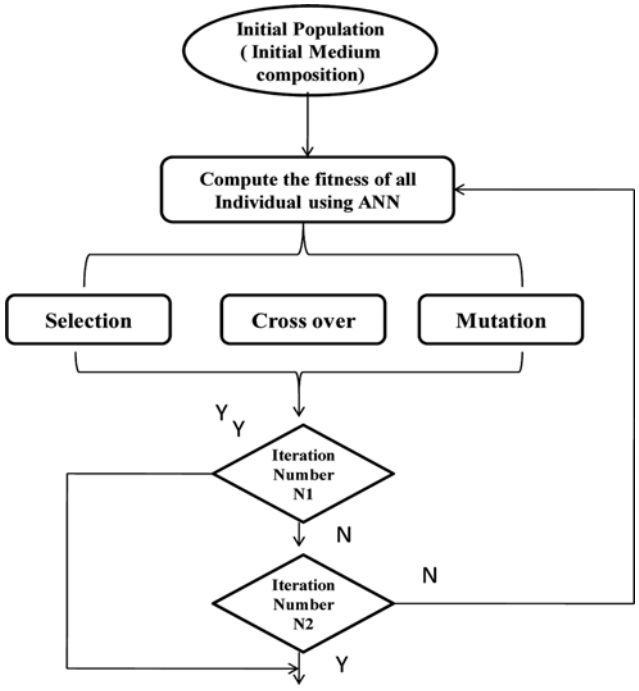

Fig. 1. Flowchart of procedure describing ANN and Genetic algorithm models adopted for medium optimization.

adjusting the weights and biases appropriately. During training step the weight and biases are iterated by Levenberg-Marquardt algorithm until the convergence to the certain value is achieved [21]. We employed a neural network toolbox of MATLAB (2010a) mathematical software to predict the hIFN-γ concentration.

8. Genetic Algorithm (GA)

Once the ANN model was developed, its input space was further optimized by using GA as demonstrated in Fig. 1. The input variable of ANN would become the decision variable for GA. The optimization in GA followed four stages: during the first stage, initialization of the solution for the population known as chromosome takes place, followed by the fitness computation, which in turn is dependent on the objective function; thereby selecting best chromosome, the selected chromosome undergoes a genetic propagation using genetic operators like crossover and mutation, which leads to the creation of a new set of chromosomes. This process is repeated until a suitable result is achieved. GA can be described as a global optimization procedure with the advantage of not being dependent on the initial value to achieve the convergence [21]. The objective function of GA can be given by:

$$\text{Maximize } Y = f(x, w),\ x_i^L \le x_i \le x_i^u,\ i = 1, 2, 3 \cdots P$$

where f is objective function (ANN model): x denotes input vector: w denotes corresponding weight vector; Y refers to the hIFN-γ experimental yield. X denotes operating conditions. P denotes number of input variables, and x_i^L & x_i^U lower and upper bounds of x_i fitness of each candidate solution were evaluated based on following fitness function:

$$\text{error}_j = 1 - \frac{1}{Y_{pred}^j};\ j = 1, 2, 3, \ldots, n$$

where error j denotes the fitness value of the candidate solution, and Y_{pred} denotes the MLP model predicted hIFN-γ yield of given candidate solution.

9. Validation of the Optimized Conditions

To validate the optimized conditions, triplicate experiments were performed based on the results obtained by Box-Behnken and ANN-GA experiments. The average value of the experiments was compared with the predicted values of the optimized conditions and the accuracy and suitability of the optimized conditions were determined.

10. Stirred Tank Bioreactor Cultivation

Inoculum for the bioreactor was prepared using the above-specified BMGY medium. The sterile medium (100 mL) contained in a 250-mL Erlenmeyer flask was inoculated with loop of recombinant *Pichia* culture as explained above. The flask was incubated at 30 °C, 250 rpm for 16 h. The entire content of the flask with an average O.D (optical density) of 5-6 was used for inoculating the bioreactor. Batch fermentations were performed using a 2 L Biostat B plus (Sartorius, Germany) stirred tank bioreactor. At initiation, the bioreactor contained 1 L of the liquid medium. The incubation temperature, agitation speed and aeration rate were regulated at 25 °C, 600 rpm and 1.5 L min^{-1}, respectively. The dissolved oxygen level was maintained at 30% of the air saturation value. The pH was controlled at 4.5 by automatic addition of alkali (1 M KOH) and acid (1 M HCl), as required.

11. Kinetic Modeling

The measured batch fermentation profiles of biomass concentration (X) and hIFN-γ(P) were simulated using unstructured kinetic models. The fermentation kinetic parameters were estimated using nonlinear regression to fit the models to the measured data. Levenberg-Marquardt (LM) algorithm based on iterative solution method was used in obtaining the solutions to the model equations. MATLAB 7.1 was used for nonlinear regression.

11-1. Model of Recombinant *Pichia pastoris* Growth

The dry biomass concentration was modeled by using the logistic equation, which describes as follows:

$$\frac{dX}{dt} = \mu_{max} X\left(1 - \frac{X}{X_{max}}\right) \tag{10}$$

where, dX/dt is the rate of biomass production (g $L^{-1}h^{-1}$), μ_{max} is the maximum specific growth rate (h^{-1}), X is the biomass concentration (g L^{-1}) and X_{max} is the model predicted maximum biomass concentration for the fermentation (g L^{-1}). The integrated form of Eq. (10) is the following:

$$X = \frac{X_0 \exp(\mu_{max} t)}{1 - \left(\frac{X_0}{X_{max}}\right)(1 - \exp(\mu_{max} t))} \tag{11}$$

where, X_0 is the initial biomass concentration (g L^{-1}) and t is time (hour).

11-2. Model of Human Interferon Gamma (hIFN-γ) Production

The production of hIFN-γ was modeled using the Luedeking-Piret equation:

$$\frac{dP}{dt} = \alpha\left(\frac{dX}{dt}\right) + \beta X \tag{12}$$

where, dP/dt is the rate of hIFN-γ production (mg $L^{-1}h^{-1}$), dX/dt is the rate of biomass production (g $L^{-1}h^{-1}$), X is the biomass concentration (g L^{-1}), α is a growth associated constant (mg g^{-1}) and β is a non-growth associated constant (mg $g^{-1}h^{-1}$). The values of α and β depend on the fermentation conditions. By substituting Eqs. (10) And (11) in (12) results in the following relationship:

$$\frac{dP}{dt} = \alpha\left[\mu_{max} X\left(1 - \frac{X}{X_{max}}\right)\right] + \beta\left[\frac{X_0 \exp(\mu_{max} t)}{1 - \left(\frac{X_0}{X_{max}}\right)(1 - \exp(\mu_{max} t))}\right] \tag{13}$$

Eq. (13) can be integrated using the initial condition t=0, $X=X_0$ and $P=P_0$, to produce the following equation:

$$P - P_0 = \alpha\left[\frac{X_0 \exp(\mu_{max} t)}{1 - \left(\frac{X_0}{X_{max}}\right)} - X_0 + \beta\left[\frac{X_{max}}{\mu_{max}} \ln\left(1 - \left(\frac{X_0}{X_{max}}\right)(1 - \exp(\mu_{max} t))\right)\right]\right] \tag{14}$$

where, P is the hIFN-γ concentration (mg L^{-1}), P_0 is the initial hIFN-γ concentration (mg L^{-1}) and t is time (hours).

12. Analytical Methods

ELISA for the quantification of hIFN-γ was performed using

the Biolegend ELISA MAX™ Deluxe set. The dry cell weight measurement was carried out by harvesting 5 ml of culture broth and centrifugation at 10,000 ×g for 10 min followed by drying at 80 °C in vacuum oven; the dry biomass was then weighed. Dry cell weight (DCW) was plotted against OD @ 600 of the samples in the range of linearity (0-1) OD.1 unit of OD corresponded to 0.272 g DCW. The specific growth rate was calculated in the exponential phase. The specific growth rate (μ) in the exponential phase was calculated as the slope of plot drawn between ln (γ_x) vs time; where, γ_x is the dry cell mass obtained at a particular time.

RESULTS AND DISCUSSION

To date, the hIFN-γ is cloned and expressed in various popular host systems such as CHO cell lines, baculovirus and *E. coli*, but the major bottleneck in expressing this protein is that it easily tends to form inclusion bodies due to lack of glycosylation. Even though the glycosylation mechanism is present in mammalian cells, the production level is very low and the medium cost is expensive [29]. *Pichia pastoris* has the added advantage of glycosylating and secreting the foreign protein extracellularly into the defined media, which eases the downstream processing. There have been only a few reports available on the expression of hIFN-γ in *Pichia pastoris*.

1. Effect of Carbon Source on hIFN-γ Production

In BSM and FM22 medium, glycerol is the most commonly used carbon source, but it is found that usage of glycerol in the medium perpetuates a distinct repressible effect on the AOX gene, which ultimately hinders the heterologous protein production in *Pichia pastoris*. So, it is necessary to screen an appropriate carbon and nitrogen source that enhances the hIFN-γ production without exhibiting a repression effect. Addition of non-repressible carbon substrate with methanol is reported to have a substantial effect on the recombinant protein production in *P. pastoris*. Recently, some reports suggested that a carbon source apart from glycerol has a positive effect with recombinant *P. pastoris* strains [30]. In the fermentation, methanol serves as a carbon source/inducer during the induction phase. Apart from glycerol, other non-repressible carbon sources such as sorbitol, mannitol, trehalose, alanine, and lactic acid with methanol have been used and an elevated expression of recombinant proteins has been reported [15]. With the view of understanding the effect of carbon source on the hIFN-γ production, we selected eight non-repressible carbon sources: sorbitol, mannitol, lactose, galactose, gluconate, maltose, whey, glycerol at 40 g/L with 1% methanol as an inducer. Among the carbon sources screened, gluconate exhibited a prominent effect on the expression of hIFN-γ, which resulted in 6.2 mg/L of protein yield, followed by galactose with 5.8 mg/L of hIFN-γ. In contrast, sorbitol resulted in about 4.2 mg/L of hIFN-γ production. Maltose, galactose and glycerol showed higher specific growth rate of 0.036 h^{-1}, while gluconate showed specific growth rate of 0.022 h^{-1}; whey, a by-product of dairy industry, showed low specific growth rate and also resulted in lower product yield. Product yield and specific growth rate of different carbon sources are illustrated in Fig. 2(a). In this study, gluconate emerged as a prominent carbon source with substantial effect on the production of hIFN-γ. In bacteria, gluconate is usually metabolized through the Entner-Doudoroff pathway, but also

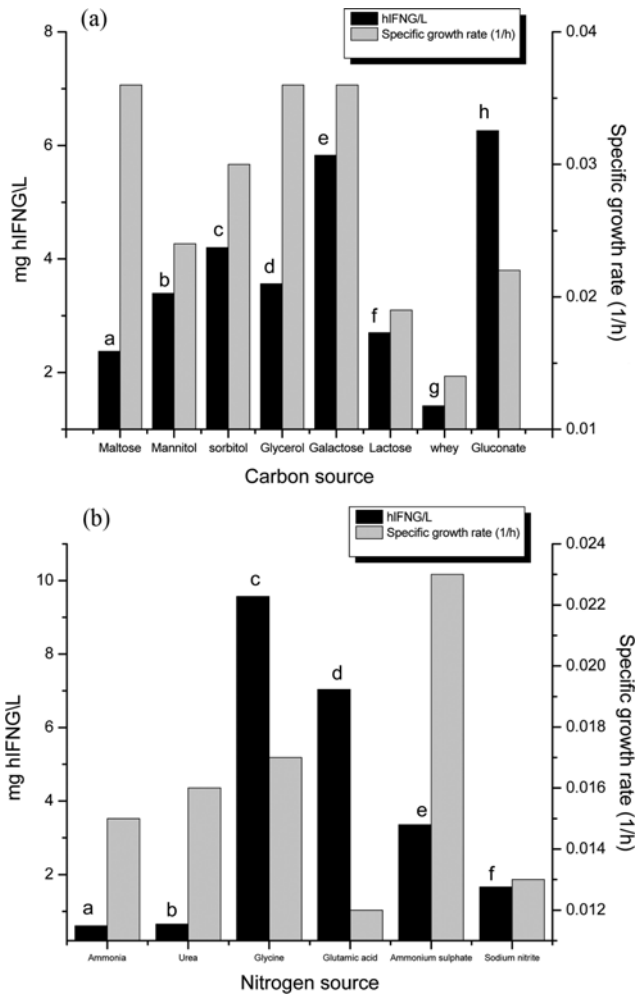

Fig. 2. (a) Effect of carbon sources (sorbitol, mannitol, gluconate, lactose, glycerol, whey, galactose, and maltose) on the specific growth rate and production of hIFN-γ with 1% methanol induction, (b) effect of nitrogen sources (ammonia, urea, glutamate, glycine, ammonium sulfate and sodium nitrite) on the specific growth rate and production of hIFN-γ with 1% methanol induction.

uptake can be through the pentose phosphate pathway. Industrially, gluconate salts are used as food additives or as secondary carbon sources for bio-related products formation [31]. In *Pichia pastoris*, the uptake of gluconate is executed through an enzyme gluconokinase, which converts gluconate to gluconate-6-phosphate with the consumption of 1 mole of ATP, thereby entering into pentose phosphate pathway. Further, gluconate-6-phosphate is converted to D-Ribose-5-Phosphate with the generation of 1 mole of NADPH, which will be utilized in biosynthetic pathway for the generation of biomass and for the generation of important amino acids like lysine, which comprises major proportion of amino acid in hIFN-γ. The D-Ribose-5p is branched to PRPP pathway for generation of nucleic acid and other branch leads to glyceraldehydes-3P which enters glycolysis (Fig. S1). Bianchi et al., 2001 reported about 14 g/L of lysine production in *Corynebacterium glutamicum* using gluconate as sole carbon source. Recently Wu et al., 2013, reported higher uptake rate of gluconate compared to glucose as well as stable pH

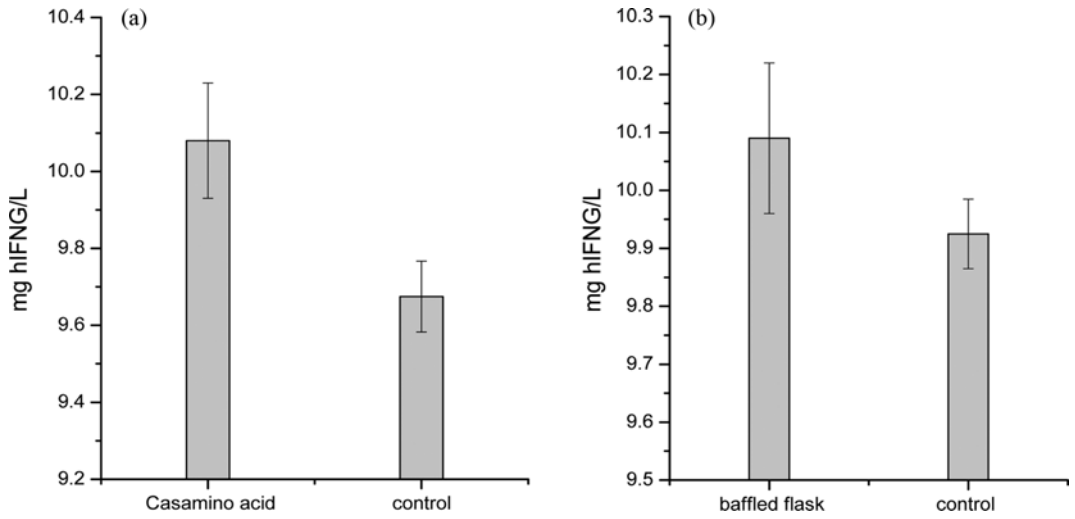

Fig. 3. (a) Effect of Casamino acid (1%) supplementation on the production of hIFN-γ, (b) Effect of baffled flask on the production of hIFN-γ.

throughout in the ethanol fermentation by *E. coli*.

2. Effect of Nitrogen Source on hIFN-γ Production

In yeast, nitrogen source plays a major role in exhibiting higher protein expression level, as the transcription of carbon metabolizing gene depends on the source of nitrogen used. In this investigation, the effect of eight nitrogen sources, glycine, glutamic acid, ammonium sulfate, urea, ammonia and sodium nitrate on hIFN-γ production, was studied. The production of hIFN-γ was enhanced by 9.5 mg/L when glycine and gluconate were used as nitrogen and carbon source, respectively. Similarly, significant effect was observed when glutamic acid was used as nitrogen source, where around 7 mg/L of hIFN-γ production was achieved. Other nitrogen sources such as urea and ammonia had a negligible effect on the expression of hIFN-γ. Meanwhile, ammonium sulfate and glycine had the maximum growth rate of 0.017 and 0.023 h^{-1}, while sodium nitrate showed a low growth rate compared to other nitrogen sources used (Fig. 2(b)). We found that glycine as prominent nitrogen source resulted in the significant enhancement of hIFN-γ production when compared to other sources used. In *Pichia pastoris*, glycine metabolism branches into two pathways; in one pathway, glycine is metabolized to 5, 10-methylene-THF and 1 mole of ammonia is liberated, thereby used as a nitrogen source, and further 5, 10-methylene-THF is used in the formation of lipoylprotein, whereas in other pathways glycine is converted into serine (bi-directional) followed by pyruvate formation (Fig. 2S). Glycine has been found to induce morphological changes in *E.coli* by enhanced translocation of protein. Supplementation of glycine in the fermentation medium may result in slight disruption on peptidoglycan cross-linkages and cell membrane integrity [32]. Yang et al. [33] reported that adding 2% (w/v) glycine and 1% triton X-100 dramatically increased extracellular production of sFV/TNF-α and β-glucosidase in bacterial system. Similar results were observed in this study with 1% glycine and 0.01% triton X-100 addition.

3. Effect of Aeration and Casamino Acid on hIFN-γ Production

Aeration is one of the most crucial parameters that affect protein production in *Pichia pastoris*. Reports have shown that the usage of baffled flasks can enhance the oxygenation efficiency as compared to the non-baffled ones [34]. To understand the effect of aeration on hIFN-γ expression, we carried out expression studies in baffled and non-baffled flasks with modified FM22 with gluconate as carbon and glycine as nitrogen source, but no significant improvement was achieved in hIFN-γ concentration using baffled flasks (Fig. 3(a)). Also in the current study, we observed no significant changes in hIFN-γ production with the supplementation of 1% casamino acid (Fig. 3(b)). Similar results were observed by Batra et al. [35], where they witnessed reduction in the production of β-glucosidase, which was expressed in *Pichia pastoris* using casamino acid as a nitrogen source. Addition of casamino acid is reported to have a stabilizing effect on protein production by reducing the proteolytic activity as it serves as a nitrogen source in nutrient-starved condition [36]. Hu et al. [37] observed higher growth rate and increase in product yield by supplementing casamino acid in the medium. whereas Batra et al. [35] and Xie et al. [14] observed no effect on protein production with addition of 1% casamino acid.

To evaluate the higher production of hIFN-γ production, it is very important to study the significance of each medium component and interaction among them. Hence, experiments were carried out to screen the significant medium components and optimize their levels using the Plackett-Burman and Box-Behnken experimental design techniques, respectively. Furthermore, precise optimization of nonlinear system was carried out using ANN-GA based optimization.

4. Screening of Essential Medium Components

Plackett-Burman experiments were performed according to the design matrix given in Table 2. The observed and predicted response of hIFN-γ concentration is specified in Table 2. The concentration of hIFN-γ varied from 2.3 mg/L to 16.3 mg/L. This wide variation in hIFN-γ concentration reflected the importance of the optimization of medium constituents on protein production. The first-order polynomial for hIFN-γ production is given in Eq. (10)

$$Y_{hIFN\gamma}=9.447-2.418X_1-3.513X_2-5.475X_3+4.21X_7 \quad (10)$$

where, X_1=Gluconate (g/L), X_2=Glycine (g/L), X_3=KH_2PO_4 and X_7=Histidine (g/L)

First-order regression model was fitted to check the adequacy of the model. Analysis of Variance (ANOVA) was carried out for screening the important components, and the significance of the variables was judged by using Student's *t* test. Table 3 represents the effects, values of coefficients, t and P values of each component from the responses. Generally, a higher t-value and low p-value indicated high significance in the model term. The main effect of each variable was estimated as the difference between both averages of measurements made at the high level (+1) and at the low level (−1) of that variable. In this study, ANOVA showed that the variables such as gluconate, glycine, KH_2PO_4, $MgSO_4{\cdot}7H_2O$, had negative effect, whereas vitamins, histidine, EDTA and Triton X-100 had positive effect. The absolute values of the variable effects on the response are shown in Table 3. These results indicated the relative contribution of the variable on the responses; the positive sign indicated that the higher level of variable resulted in higher response, whereas the negative sign specified the lower level of variable resulting in higher response. The significant components were selected based on the P<0.05. In the present study, gluconate, glycine, KH_2PO_4 and histidine showed significance (P<0.05) and were considered as prominent variables for further optimization studies.

The effect of the variables on the response was illustrated in a Pareto chart (Fig. 4), which is a pictorial way to view the results of a Plackett-Burman experimental design. The ranking was made according to the absolute values of standardizing effects, which is important in designing further optimization processes. The reference line (4.30) indicated that effects were significant with α value of 0.05. The variable crossing the line was considered as significant at that particular α value. Fig. 4 shows the standardized effect of the t-test, which is calculated by dividing each variable coefficient with its standard error. Fig. 4 represents that the variables such as gluconate, glycine, KH_2PO_4 and histidine have a significant role in enhancing hIFN-γ concentration. All other insignificant variables were not included in the further optimization experiment, but instead used at their middle level (center point) and Triton X-100 was used at the low level, as higher triton concentration had shown reduced growth of *Pichia*.

5. Optimization of Screened Variables for Maximization of hIFN-γ

We adopted a three-level Box-Behnken design to optimize and investigate the effect of screened variables on response hIFN-γ. The design matrix and the corresponding results of observed and predicted responses (production of hIFN-γ) are given in Table 5. The concentration of hIFN-γ varied from 3.33 mg/L to 28.48 mg/L. To check the model adequacy, multiple regression analysis was carried out and second-order polynomial model was fitted in Eq. (11).

$$Y_{hIFN\gamma} = 26.86 - 1.88X_1 - 2.00X_2 - 2.70X_3 + 1.26X_7 - 11.81X_1^2 - 6.89X_2^2 - 8.37X_3^2 - 6.28X_7^2 - 0.64X_1X_2 + 4.04X_1X_3 + 0.46X_1X_7 + 1.04X_2X_3 - 5.82X_2X_7 - 1.64X_3X_7 \quad (11)$$

where, X_1=Gluconate (g/L), X_2=Glycine (g/L), X_3=KH_2PO_4 and X_7=Histidine (g/L)

The data were analyzed by ANOVA; the results are shown in Table 6(a). According to ANOVA the model Fisher *F* test (mean square regression: mean square residual is 62.56) was highly significant, with P<0.05; the result demonstrated that the interaction between the variables had a significant effect in enhancing the production of hIFN-γ. The model goodness of fit was determined by coefficient of determination (R^2); the model R^2 was found to be 0.986, which implied that 98.6% of variation in the model could be explained. R^2 gives the proportion of the total variation in the responses predicted by the models; the predicted R^2 was found to be 0.939, which is the measure of how good the model predicts a response value. The model lack of fit had shown insignificant (p> 0.05) with F value of 0.36, The lack of fit measures the failure of the model to represent data in the experimental domain at points

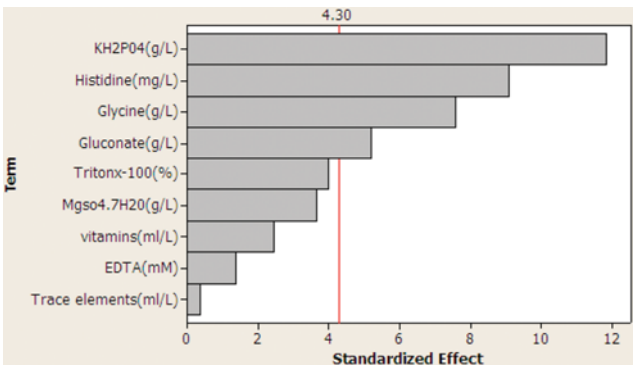

Fig. 4. Pareto chart of the standardized effects of the factors on the hIFN-γ production, α=0.05.

Table 6(a). ANOVA for quadratic model

Source	DF	SS	MS	F	P
Model	14	1306.01	93.286	62.56	0
Residual (error)	12	17.9	1.491		
Lack-of-fit	10	11.51	1.151	0.36	0.89
Pure error	2	6.39	3.195		
Total	26	1323.9			

R^2=98.65%, R^2 (pred)=93.91%, R^2 (adj)=97.07%
DF=degrees of freedom, SS=sum of squares, MS=mean square

Table 6(b). Model coefficient estimated by multiple linear regressions

Term	Coef	SE Coef	T	P
Constant	26.8687	0.705	38.109	0
X_1	−1.8869	0.3525	−5.352	0
X_2	−2.0015	0.3525	−5.678	0
X_3	−2.708	0.3525	−7.682	0
X_7	1.2633	0.3525	3.584	0.004
X_1*X_1	−11.8139	0.5288	−22.342	0
X_2*X_2	−6.8994	0.5288	−13.048	0
X_3*X_3	−8.3766	0.5288	−15.841	0
X_7*X_7	−6.2817	0.5288	−11.88	0
X_1*X_2	−0.6445	0.6106	−1.056	0.312
X_1*X_3	4.0424	0.6106	6.621	0
X_1*X_7	0.4655	0.6106	0.762	0.461
X_2*X_3	1.046	0.6106	1.713	0.112
X_2*X_7	−5.821	0.6106	−9.533	0
X_3*X_7	−1.6405	0.6106	−2.687	0.02

that are not included in the model. The student *t* test and the corresponding p-value, along with the parameter estimation, are shown in Table 6(b). The probability value for all linear and square termsswas found to be significant, but the interaction between glycine with gluconate and KH_2PO_4, gluconate with histidine was found to be insignificant.

With the help of the 3D response plots constructed, production of hIFN-γ was predicted for different values of the tested variables. The plot was built in a way where the response (hIFN-γ concen-

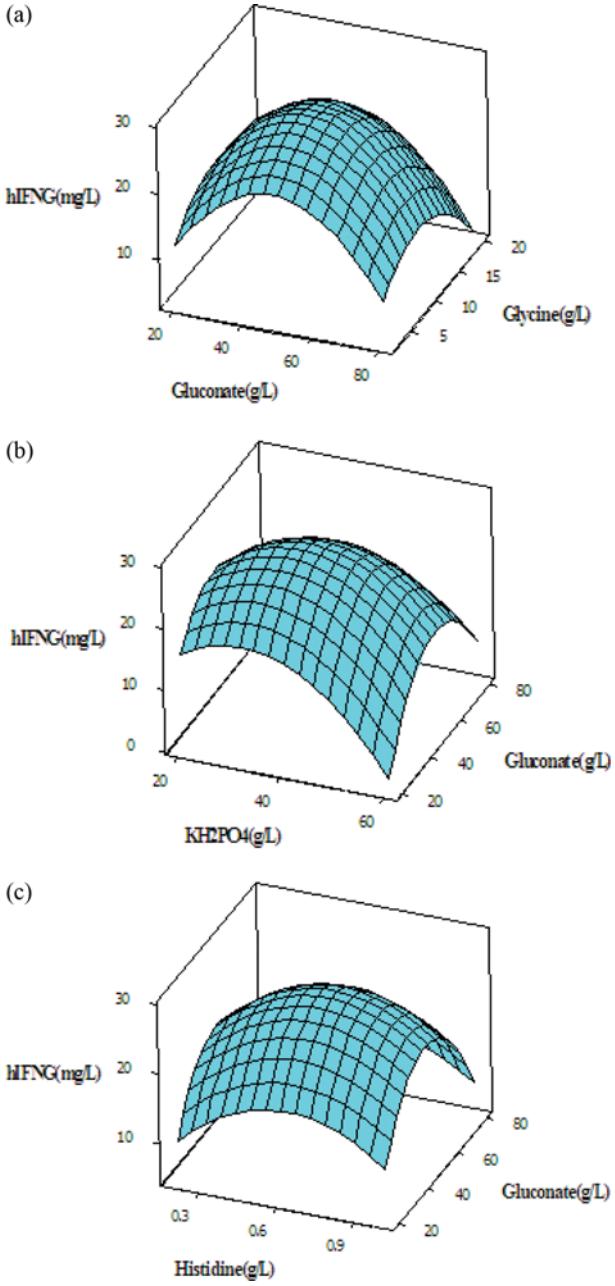

Fig. 5. Three-dimensional response surface plot for hIFN-γ production showing the interactive effects of (a) gluconate and glycine (b) gluconate and KH_2PO_4 (c) gluconate and histidine with the remaining factors kept constant at the middle level of the Box Behnken experimental design.

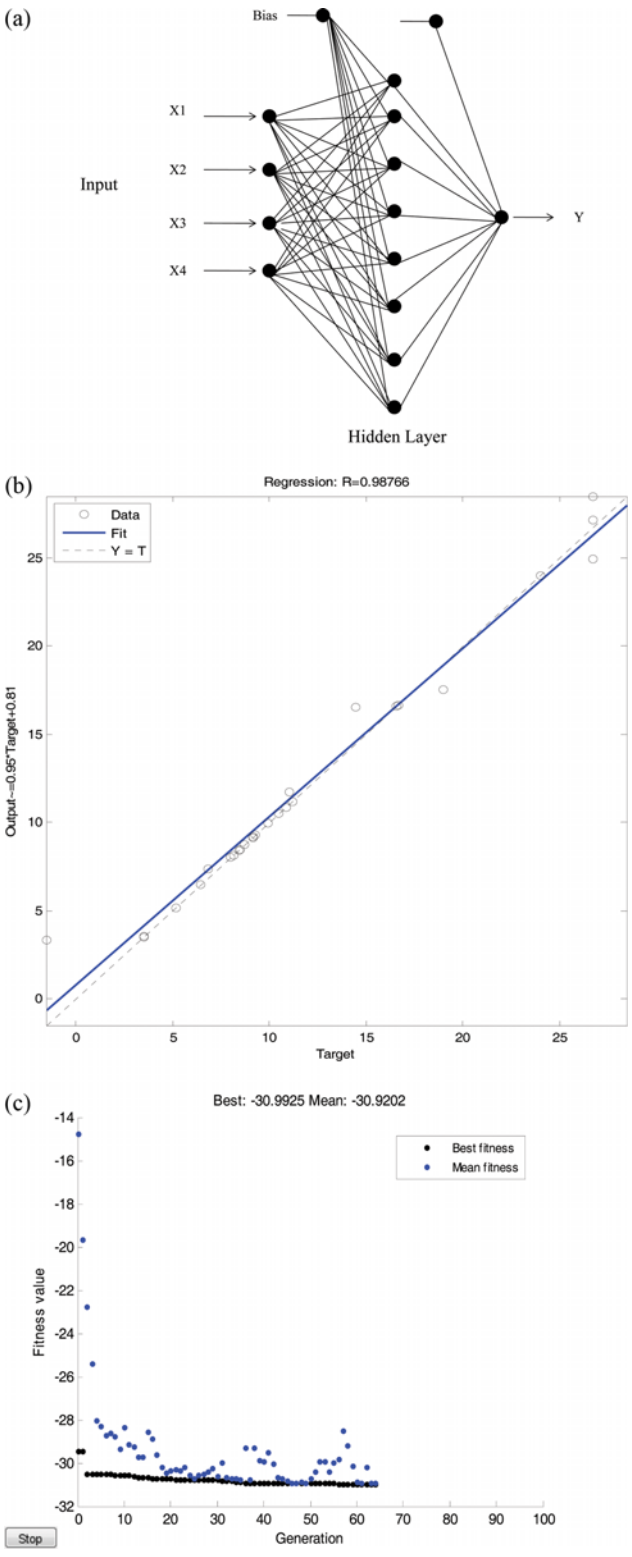

Fig. 6. (a) Schematic representation of a (4-8-1) neural network (having three neurons in the input layer, eight neurons in the hidden layer and one in the output layer). (b) The prediction performance of ANN models for the hIFN-γ production. (c) Representative plots generated from the optimization by GA using MATLAB (2010 a) Best and average fitness values with successive generations showed gradual convergence to the optimum value for hIFN-γ production.

tration) was plotted on z axis against any two independent variables while maintaining other variables at their optimal levels given in Fig. 5(a)-(c). There was a direct correlation between hIFN-γ production and concentration of glycine. There was a steep increase of hIFN-γ production with increasing amount of glycine, which led to the maximum hIFN-γ production between 10-15 g/*l* (Fig. 5(a)). Likewise, the same pattern was followed by gluconate and KH_2PO_4 (Fig. 5(b)), which with increasing concentration showed an improved maximum production between 55-60 g/*l* and 40 g/*l*, respectively. Past this level the hIFN-γ expression was hindered and a sheer decrease in production was detected. There was a very prominent interaction witnessed from the figure between KH_2PO_4 with gluconate ($P<0.05$) and histidine with gluconate (Fig. 5(c)). There was an observed gradual decrease in the hIFN-γ production on increasing the concentration of Gluconate and Glycine above their mid value.

6. Hybrid ANN-GA Modelling

6-1. Predictive Modelling by Artificial Neural Network

We employed the most commonly used feed forward back propagation ANN namely, multi-layer perception (MLP), to train the data. The input node represents the screened variables, gluconate, glycine, KH_2PO_4 and histidine, while the output node represents the concentration of hIFN-γ. Network topology plays an important role in predicting results. The number of hidden layers was determined by training ANN topology several times until least MSE (mean square error) was achieved. In this algorithm, the ANN was trained using Levenberg-Marquardt (LM) method, which approximates Newton's method and it is the most suitable method for training ANN. This algorithm uses a second-order equation for better convergence of MSE between desired output and actual output. The training was done for 1000 epochs and the optimum value was reached by 9 epochs. The optimal result was obtained with network topology of four inputs, eight hidden layers and one output layer, which is illustrated in Fig. 6(a). The MSE and determination coefficient (R^2) for training validation and test are shown in Table S1. Model output versus prediction output is shown in Fig. 6(b). The R^2 of the model was found to be 0.9876 and only 0.127 of the total variations was not explained by the model. The predicted value of ANN is given in Table 5. The error and learning curve of the training, validation and test is shown in Fig. S3.

6-2. Optimization Based on Genetic Algorithm

Once the ANN architecture was trained, the GA technique was used to optimize input space with the aim of maximizing hIFN-γ yield. The values of GA specific parameters used in the optimization technique were as follows: population size=20, cross over probability=0.8, mutation probability=0.01, No. of generation=100. The optimal solution of GA would be restricted between the levels specified in Box-Behnken. The GA was repeated several times with different initial parameter conditions until a global optimum was obtained. It showed that after 51 iterations, the optimized value of 30.99 mg/L of hIFN-γ was achieved by maintaining the variables: Gluconate=50 g/L, Glycine=10.185 g/L, KH_2PO_4=35.912 g/L, Histidine= 0.264 g/L, the optimum solution was found heuristically Fig. 6(c) [38,39].

7. Validation of Box-Behnken and ANN-GA

To verify the validity of the model, experiments were carried out in triplicate at optimal levels of significantly influenced medium components and at middle level of other medium components; the values were compared with that of predicted value and also keeping BSM medium as a control. The observed value (27.14 mg hIFN-γ/L) was in good agreement with the predicted value of (28.48 mg hIFN-γ/L). The ANN-GA based optimization was validated by carrying out the fermentation at GA-specified optimum conditions. The hIFN-γ yield obtained in the verification experiment was 29.72 mg/L, which is in close agreement with the hybrid ANN-GA solution of 30.99 mg/L, while the BSM medium resulted in about 5.8 mg/L of hIFN-γ yield. The optimization with Box-Behnken and ANN-GA enhanced hIFN-γ yield by 4.6- and 5.1-fold compared to BSM medium. In our experiments, we observed precipitate formation in BSM medium, while no traces of precipitate were found in modified FM22 medium.

Comparative studies between RSM and ANN results showed the usage of the neural network as an empirical model for predicting a nonlinear system. ANN model was found to have excellent prediction accuracy and generalizability. The network models not only fit the training data very well, but also closely predicted the validation data. With optimum concentration of Gluconate=50 g/L, Glycine=10.185 g/L, KH_2PO_4=35.912 g/L, Histidine=0.264 g/L we found maximum of 30 mg/L hIFN-γ yield with ANN linked GA optimization, whereas Box-Behnken design yielded 28 mg/L of product yield. The coefficient of determination of ANN-GA model (R^2= 0.9876) was higher than that of Box-Behnken. In this current investigation, we achieved a maximum of 30 mg/L of hIFN-γ production from *Pichia pastoris*. This study indicated that modifying the existing FM22 medium with optimizing the medium components achieved a high yield of extracellular heterologous protein production.

8. Unstructured Model Prediction in Batch Reactor

The profile growth and hIFN-γ production in bioreactor under controlled conditions are illustrated in Fig. 7. There was a direct correlation between hIFN-γ production and biomass formation until 80 h and declination profile was observed thereafter. Kinetic

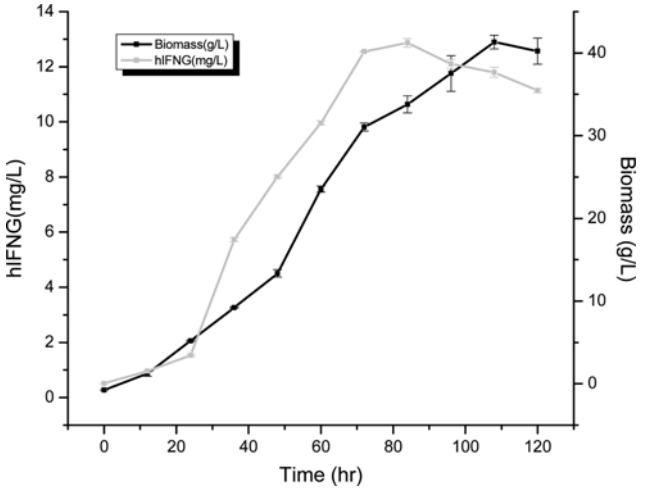

Fig. 7. Comparison of time course profiles for biomass and hIFN-γ production, respectively, in batch fermentation using modified FM22 medium.

Table 7. Parameters estimated by logistic and Leudeking-Pirate model equation

Model parameters	Experimental values	R^2
X_0 (g/L)	0.653	
X_{max} (g/L)	12.46	
μ_m (h^{-1})	0.211	0.94
α (mg/g)	1.36	
β (mg/g·h)	0.015	

parameters involved in the process were estimated using the models mentioned in Eqs. (11) and (14). These models describe the kinetics of growth and product formation. The estimated kinetic parameters values obtained from these models are given in Table 7. The coefficients of determination (R^2) values obtained by fitting the various models to the experimental data were found to be very significant (R^2>0.94). Using Leudeking-Piret model, α and β value values were predicted. The L-P equation suggested that the production of hIFN-γ was mixed growth associated. The maximum biomass and hIFN-γ was found to be 12.89 g DCW/L and 40.18 mg/L, respectively. The growth and product formation are illustrated in Fig. 7.

CONCLUSIONS

Pichia pastoris is a well-known host organism for the production of many heterologous proteins. For high yield, the expression of any protein defining proper medium components is essential. In industries, the most commonly used defined media are BSM and FM22, which leads to problems such as precipitation, higher ionic strength and also glycerol as a carbon source has more repressible effect on the AOX gene, hence modification of the existing medium with additive ingredient for enhanced growth and product formation is required. In the present study, we attempted to modify existing FM22 medium with addition of components such as vitamins, triton X-100 and EDTA for enhanced growth and extracellular secretion of human interferon gamma protein in medium, which eased the cost of downstream processing. Apart from this, various carbon and nitrogen sources were screened for increased product yield; gluconate and glycine appeared as eminent carbon and nitrogen sources for hIFN-γ production. Usage of casamino acid and baffled flask showed less effect on protein production. Screening of main components which influenced product formation was done using Placket-Burman screening, and the screened components were further subjected to Box-Behnken design and ANN-GA for precise optimization. We found maximum of 30 mg/L hIFN-γ yield with ANN linked GA optimization, whereas the Box-Behnken design yielded in 28 mg/L of product yield. It was seen that the coefficient of determination of ANN-GA model (R^2= 0.9876) was higher than that of Box-Behnken. Since ANN-GA is a more accurate and more generalized model than quadratic RSM, it is better equipped to reach the global optimum. Moreover, no precipitate formation was observed in the medium. Also, the batch reactor kinetics was fitted for the optimized medium; maximum biomass of 12.89 g DCW/L was found with the production 40.18 mg/L. Also, the L-P equation showed that the production of hIFN-γ was mixed growth associated.

ACKNOWLEDGEMENT

The authors would like to thank Nitin Kumar, Rajat Pandey, Sushma C, and Yachna G for their support to carry out experiments. This study was financially supported by Indian Institute of Technology Guwahati, Guwahati, India.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals.

SUPPORTING INFORMATION

Additional information as noted in the text. This information is available via the Internet at http://www.springer.com/chemistry/journal/11814.

REFERENCES

1. P. W. Gray and D. V. Goeddel, *Nature*, **298**, 859 (1982).
2. H. M. Younes and B. G. Amsden, *J. Pharm. Sci.*, **91**, 2 (2002).
3. S. Macauley-Patrick, M. L. Fazenda, B. McNeil and L. M. Harvey, *Yeast*, **22**, 249 (2005).
4. O. Cos, R. Ramón, J. L. Montesinos and F. Valero, *Microb. Cell Factories*, **5**, 17 (2006).
5. A. Idiris, H. Tohda, H. Kumagai and K. Takegawa, *Appl. Microbiol. Biotechnol.*, **86**, 403 (2010).
6. B. A. Plantz, J. Sinha, L. Villarete, K. W. Nickerson and V. L. Schlegel, *Appl. Microbiol. Biotechnol.*, **72**, 297 (2006).
7. G. Potvin, A. Ahmad and Z. Zhang, *Biochem. Eng. J.*, **64**, 91 (2012).
8. W. Zhang, H. Liu and J. Chen, *Biochem. Eng. J.*, **12**, 1 (2002).
9. W. Zhang, J. Sinha and M. M. Meagher, *Appl. Microbiol. Biotechnol.*, **72**, 139 (2006).
10. A. Ghosalkar, V. Sahai and A. Srivastava, *Bioresour. Technol.*, **99**, 7906 (2008).
11. A. Solà, H. Maaheimo, K. Ylönen, P. Ferrer and T. Szyperski, *Eur. J. Biochem.*, **271**, 2462 (2004).
12. R. A. Brierley, C. Bussineau, R. Kosson, A. Melton and R. S. Siegel, *Ann. N. Y. Acad. Sci.*, **589**, 350 (1990).
13. E. D. Thorpe, M. C. d'Anjou and A. J. Daugulis, *Biotechnol. Lett.*, **21**, 669 (1999).
14. J. Xie, Q. Zhou, P. Du, R. Gan and Q. Ye, *Enzyme Microb. Technol.*, **36**, 210 (2005).
15. "Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale | Science." Available: http://science.sciencemag.org/content/278/5338/680.long (2016).
16. B. Magasanik, *Cold Spring Harb. Monogr. Arch.*, **21**, 283 (1992).
17. V. V. Dasu and T. Panda, *Bioprocess Eng.*, **22**, 45 (2000).
18. C. Sivapathasekaran, S. Mukherjee, A. Ray, A. Gupta and R. Sen,

Bioresour. Technol., **101**, 2884 (2010).
19. Y. Yu, X. Zhou, S. Wu, T. Wei and L. Yu, *Electron. J. Biotechnol.*, **17**, 311 (2014).
20. S. Kumar, K. Pakshirajan and V. V. Dasu, *Appl. Microbiol. Biotechnol.*, **84**, 477 (2009).
21. Y. Yasin, F. B. H. Ahmad, M. Ghaffari-Moghaddam and M. Khajeh, *Environ. Nanotechnol. Monit. Manag.*, **1**, 2 (2014).
22. A. A. Prabhu, V. D. Veeranki and S. J. Dsilva, *Process Biochem.*, **51**, 709 (2016).
23. R. L. Plackett and J. P. Burman, *Biometrika.*, **33**, 305 (1946).
24. G. E. P. Box and D. W. Behnken, *Technometrics.*, **2**, 455 (1960).
25. J. P. Maran, S. Manikandan, B. Priya and P. Gurumoorthi, *J. Food Sci. Technol.*, **52**, 92 (2013).
26. X. Song, A. Mitnitski, C. MacKnight and K. Rockwood, *J. Am. Geriatr. Soc.*, **52**, 1180 (2004).
27. K. M. Desai, S. A. Survase, P. S. Saudagar, S. S. Lele and R. S. Singhal, *Biochem. Eng. J.*, **41**, 266 (2008).
28. M. Khayet and C. Cojocaru, *Sep. Purif. Technol.*, **86**, 171 (2012).
29. P. Christova, K. Todorova, I. Timtcheva, G. Nacheva, A. Karshikoff and P. Nikolov, *Zeitschrift Für Naturforschung.*, **58**, 288 (2003).
30. C. Jungo, J. Schenk, M. Pasquier, I. W. Marison and U. von Stockar, *J. Biotechnol.*, **131**, 57 (2007).
31. D. Bianchi, O. Bertrand, K. Haupt and N. Coello, *Enzyme Microb. Technol.*, **28**, 754 (2001).
32. J. H. Choi and S. Y. Lee, *Appl. Microbiol. Biotechnol.*, **64**, 625 (2004).
33. J. Yang, T. Moyana, S. MacKenzie, Q. Xia and J. Xiang, *Appl. Environ. Microbiol.*, **64**, 2869 (1998).
34. W. Klöckner and J. Büchs, *Trends Biotechnol.*, **30**, 307 (2012).
35. J. Batra, D. Beri and S. Mishra, *Appl. Biochem. Biotechnol.*, **172**, 380 (2013).
36. K. Sreekrishna, R. G. Brankamp, K. E. Kropp, D. T. Blankenship, J.-T. Tsay, P. L. Smith, J. D. Wierschke, A. Subramaniam and L. A. Birkenberger, *Gene*, **190**, 55 (1997).
37. S. Hu, L. Li, J. Qiao, Y. Guo, L. Cheng and J. Liu, *Protein Expr. Purif.*, **47**, 249 (2006).
38. A. A. Prabhu, S. Chityala, Y. Garg and V. V. Dasu, *Prep. Biochem. Biotechnol.*, **1**, 1 (2016).
39. A. A. Prabhu and A. Jayadeep, *Prep. Biochem. Biotechnol.*, **1** (2016).