



# Optimum culture medium composition for lipopeptide production by *Bacillus subtilis* using response surface model-based ant colony optimization

J SATYA ESWARI<sup>1</sup>, M ANAND<sup>2,\*</sup> and C VENKATESWARLU<sup>1</sup>

<sup>1</sup>Chemical Engineering Sciences Division, Indian Institute of Chemical Technology, Hyderabad 500007, India

<sup>2</sup>Department of Chemical Engineering, Indian Institute of Technology Hyderabad, Hyderabad 502205, India

e-mail: eswari\_iit@yahoo.co.in; anandm@iith.ac.in; chvenkat.iict@gmail.com

MS received 21 May 2015; accepted 28 June 2015

**Abstract.** Central composite rotatable design (CCRD) of experiments was used to obtain data for Lipopeptide and Biomass concentrations from fermentation medium containing the following five components: glucose, monosodium glutamate, yeast extract,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , and  $\text{K}_2\text{HPO}_4$ . Data was used to develop a second order regression response surface model (RSM) which was coupled with ant colony optimization (ACO) to optimize the media compositions so as to enhance the productivity of lipopeptide. The optimized media by ACO was found to yield 1.501 g/L of lipopeptide concentration which was much higher compared to 1.387 g/L predicted by Nelder–Mead optimization (NMO). The optimum from ACO was validated experimentally. RSM-based ACO is thus shown to be an effective tool for medium optimization of biosurfactant production.

**Keywords.** Ant colony optimization; *Bacillus subtilis*; central composite rotatable design; lipopeptide; response surface methodology.

## 1. Introduction

Biosurfactants are microbial compounds that exhibit pronounced surface and emulsifying activities. There is a great deal of interest in biosurfactants since they are considered as “green” alternatives to synthetic surfactants. Biosurfactants comprise a wide range of chemical structures, such as glycolipids, lipopeptides, polysaccharide–protein complexes and phospholipids: see Desai & Banat [1] and Abdel-Mawgoud *et al* [2]. Among the various biosurfactants, lipopeptides are particularly interesting because of their high surface activities and antibiotic potential. The bioactive peptides such as surfactin, fengycin and Turing A, B and C, mycosubtilins and bacillomycins fall under the category of lipopeptide biosurfactants.

The lipopeptide surfactin produced by *Bacillus subtilis* is the most powerful biosurfactant [3] with potential biotechnological and biomedical applications. Surfactin lipopeptide belongs to a group of cyclic lipopeptide containing beta-hydroxyl fatty acids and D2/L-amino acid residues: Peypoux *et al* [4], Haddad *et al* [5] and Tang *et al* [6]. It possesses various biological activities; anti-microbial, antiviral, anti-tumor, blood anticoagulant and fibrinolytic activities [7]. The cell growth and the accumulation of metabolic products of a lipopeptide biosurfactant process are strongly

influenced by medium compositions such as carbon sources, nitrogen sources, phosphorous sources, growth factors, and inorganic salt concentrations. The potential applications considered for this biosurfactant depend on whether it can be produced economically. In this regard, experimental design provides greater insight in studying the impact of potential variables affecting the process.

Experimental design and response surface methodology (RSM), in general, have been studied for different biosurfactant processes in Abalos *et al* [8], Al-Araji *et al* [9], Rispoli *et al* [10] and Rikalovic *et al* [11]. The influence of media components such as carbon-, nitrogen-, and potassium-sources, and environmental factors for the growth and production of lipopeptides by *Bacillus subtilis* has been evaluated experimentally by various researchers (see, for instance, Suwansukho *et al* [12]). Gu *et al* [13] have employed central composite design (CCD) based RSM to optimize the levels of sucrose substrate, ammonium chloride, ferrous sulphate, and zinc sulphate for the production of a lipopeptide by *Bacillus subtilis* in shaker flask fermentation. Media optimization of biosurfactant production by *Bacillus subtilis* has also been studied by Abdel Mawgoud *et al* [14]. Liu *et al* [15] have applied central composite rotatable design (CCRD) and RSM to optimize the medium composition for enhanced productivity of C15-surfactin by a *Bacillus* genus. Mutalik *et al* [16] employed RSM to find the optimum medium composition for the biosurfactant

\*For correspondence

production by *Rhodococcus spp. MTCC 2574*. Seghal Kiran et al [17] employed a quadratic model – fitted to experimental data – to optimize the critical control factors involved in the production of lipopeptide biosurfactant by marine *Brevibacterium aureum MSA13*. In most of the above studies, media optimization is performed either by the analysis of experimental results or RSM results. Very few studies have considered the use of efficient optimization algorithms to explore the design search space for optimal solution of a biosurfactant process. De Lima et al [18] developed an empirical model using bioreactor data, and combined it with an optimization method developed in Maple VIII (release 4) software to optimize the conditions for biosurfactant production by *Pseudomonas aeruginosa*. Pal et al [19] applied artificial neural network (ANN) coupled with genetic algorithm (GA) for media optimization of biosurfactant production by *Rhodococcus erythropolis MTCC 2794*. Satya Eswari et al [20] employed artificial neural network-based response surface model (ANN RSM) coupled with non-dominated sorting differential evolution (DE) to optimize the medium composition for Rhamnolipid production by *Pseudomonas aeruginosa AT10*.

The use of optimization algorithms in conjunction with the response surface models that are formulated based on the data of designed experiments can better explore the design search space for optimal solution of a process. To overcome the limitations of classical optimization techniques, a host of new optimization algorithms that operate in a different way have been developed. In fact, evolutionary optimization techniques such as the GA [21–23], simulated annealing [24, 25], ant colony optimization (ACO) [26–28], particle swarm optimization [29], and DE [30, 31] are capable of yielding more flexible solutions than the classical optimization techniques. To the best of our knowledge, the RSM coupled with ACO has not been reported so far for biosurfactant processes. Hence, in this work, it is proposed to optimize the lipopeptide biosurfactant process by coupling the CCRD based response surface models with ACO. ACO could be a very powerful and flexible tool well suited for modeling the fermentation process due to an implicit corrective action arising from the training methodology and the associated estimation procedure.

RSM is the most preferred method for fermentation media optimization. The CCRD is the most popular method to obtain data for an RSM. ACO is the most efficient meta-heuristic search algorithm used to solve combinatorial optimization problems. ACO is based on the observation that ants can find the optimal path between a food source and their nest exploiting a mix of probabilistic behavior and pheromone depositing. In ACO, a set of artificial ants simulate the behavior of real ants; the artificial ants move on the graph representation of a combinatorial optimization problem and build solutions probabilistically. The probabilities are biased by artificial pheromones that ants deposit while building solution. In this work, CCRD was used to design

the experiments, and data was generated for lipopeptide production by *Bacillus subtilis*. The data was used to develop response surface models for lipopeptide and biomass productivities. These models were coupled with ACO algorithm to optimize the media composition for enhancing the lipopeptide activity. The optimized activity was validated experimentally. The results of ACO were compared with those of Nelder–Mead optimization method, and used to generate conclusions.

## 2. Materials and methods

### 2.1 Microorganism and culture conditions

*Bacillus subtilis*, a bacterial strain as lyophilized culture (2423) from IMTECH-MTCC, was activated in 5 ml of nutrient broth under laminar air flow. The nutrient broth used was beef extract (1 g/l), yeast extract (2 g/l), peptone (5 g/l) and sodium chloride (5 g/l). The broth was incubated in a rotary shaker run at 190 rpm at 37°C.

The inocula were prepared as follows: The pure microbial culture, i.e. *Bacillus subtilis MTCC 2423* from the broth was grown on nutrient agar slants for 24–48 h. The agar slants were sub-cultured for every two weeks. From the sub-culture, one isolated colony was dispensed in nutrient broth at room temperature ( $30\pm 2^\circ\text{C}$ ) and kept in rotary shaker at 200 rpm for 16 h. This was used as an inoculum at the concentration of 10% v/v. For biosurfactant production, *Bacillus subtilis* from the inocula was grown in 250 ml flask with 100 ml of Minimal medium at the same conditions for 16 h. The concentration of the Minimal medium was glucose (2.5 g/l), monosodium glutamate (1 g/l), yeast extract (0.3 g/l),  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$  (0.1 g/l),  $\text{K}_2\text{HPO}_4$  (0.1 g/l), and KCl (0.05 g/l); the medium was sterilized at 121°C for 20 min. Further 50 ml of culture from the inoculum was grown in a 1,000 ml conical flask with 300 ml of Minimal medium for 96 h on a rotary shaker at 160 rpm and 30°C. This experiment is performed to observe the productivity of biosurfactant by Minimal medium and further to design the experiments for response surface analysis.

### 2.2 Analytical methods

**Biomass.** The biomass was determined from the cells after centrifugation of the culture broth at 6,700g (10,000 rpm), 4°C for 10 min. The dry cell weight (DCW) was obtained from the cell pellets by washing twice with distilled water and drying in hot air oven at 105°C for 24 h.

**Surface activity measurement.** Culture samples were centrifuged at 6,700g (10,000 rpm) for 20 min for cell removal, and the supernatant was subjected to surface activity measurements. Surface tension (ST) and interfacial tension (IT) were determined with a Kruss Tensiometer: this was performed at room temperature using the ring method.

The instrument was calibrated by first measuring the surface tension of pure water. The measurement was repeated at least three times, and the average was reported as the surface tension of the sample.

**Extraction of crude biosurfactant.** The crude biosurfactant was isolated from the cell-free broth of culture kept for 96 h. The bacterial cells were removed from surfactant-containing culture broth by centrifugation at 6,700g (10,000 rpm), 4°C for 20 min. The supernatant was kept overnight at 4°C and precipitated by adding concentrated HCl to achieve a final pH of 2.0. Gray white pellets (of lipids and proteins) formed by precipitation were collected by centrifugation at 6,700g (10,000 rpm), 4°C for 20 min. The pellets were lyophilized and weighed for quantification. For the extraction of biosurfactant compounds, 50 ml of chloroform–methanol mixture (2:1 v/v) was added to 500 mg of the dry product and incubated in a rotatory shaker at 250 rpm, 30°C ( $\pm 0.5^\circ\text{C}$ ) for 15 min. The extract was evaporated to dryness

and weighed for quantification. Assays were carried out in triplicates.

### 3. Design of experiments and data generation

In this study, CCRD is used to design the experiments for lipopeptide biosurfactant production. The CCRD is the most popular class of designs used for fitting second-order models. The total number of tests required for CCRD is  $2^{k-1} + 2k + n_0$ , which includes  $2k$  factorial points with its origin at the center,  $2k$  points fixed axially at a distance  $\beta$  ( $\beta = 2k/4$ ) from the center to generate the quadratic terms, and replicate tests at the center ( $n_0$ ); where  $k$  is the number of independent variables. A design should include enough replications (often at the center point) to provide an independent estimate of the experimental error allowing it to be tested for lack of fit of the model. For five variables, the recommended number of tests at the center is six. Hence the total number of tests required for five independent variables is  $2^4 + (2 \times 5) +$

**Table 1.** Experimental design matrix.

0	Actual values					Coded values					Responses	
	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	g Biomass l <sup>-1</sup>	g Lipopeptide l <sup>-1</sup>
1	1	2	0.2	0.2	0.2	-1	-1	-1	-1	+1	3.0	0.90
2	2	2	0.2	0.2	0.4	+1	-1	-1	-1	-1	3.0	1.10
3	1	4	0.2	0.2	0.4	-1	+1	-1	-1	-1	2.6	0.70
4	2	4	0.2	0.2	0.2	+1	+1	-1	-1	+1	1.8	0.12
5	1	2	0.4	0.2	0.4	-1	-1	+1	-1	-1	1.8	0.74
6	2	2	0.4	0.2	0.2	+1	-1	+1	-1	+1	2.4	1.06
7	1	4	0.4	0.2	0.2	-1	+1	+1	-1	+1	3.0	0.86
8	2	4	0.4	0.2	0.4	+1	+1	+1	-1	-1	4.0	1.26
9	1	2	0.2	0.4	0.4	-1	-1	-1	+1	-1	2.2	1.02
10	2	2	0.2	0.4	0.2	+1	-1	-1	+1	+1	1.6	1.16
11	1	4	0.2	0.4	0.2	-1	+1	-1	+1	+1	4.0	0.76
12	2	4	0.2	0.4	0.4	+1	+1	-1	+1	-1	4.0	1.36
13	1	2	0.4	0.4	0.2	-1	-1	+1	+1	+1	4.0	1.32
14	2	2	0.4	0.4	0.4	+1	-1	+1	+1	-1	4.4	1.14
15	1	4	0.4	0.4	0.4	-1	+	+1	+1	-1	3.4	1.24
16	2	4	0.4	0.4	0.2	+1	+	+1	+1	+1	2.6	0.66
17	0.5	3	0.3	0.3	0.3	-2	0	0	0	0	2.6	0.64
18	2.5	3	0.3	0.3	0.3	2	0	0	0	0	2.2	0.42
19	1.5	1	0.3	0.3	0.3	0	-2	0	0	0	2.2	0.50
20	1.5	5	0.3	0.3	0.3	0	2	0	0	0	2.6	0.38
21	1.5	3	0.1	0.3	0.3	0	0	-2	0	0	2.0	0.24
22	1.5	3	0.5	0.3	0.3	0	0	2	0	0	12	1.12
23	1.5	3	0.3	0.1	0.3	0	0	0	-2	0	9.8	1.04
24	1.5	3	0.3	0.5	0.3	0	0	0	2	0	7.2	1.32
25	1.5	3	0.3	0.3	0.1	0	0	0	0	-2	13.2	1.24
26	1.5	3	0.3	0.3	0.5	0	0	0	0	2	12.4	1.30
27	1.5	3	0.3	0.3	0.3	0	0	0	0	0	12.8	0.92
28	1.5	3	0.3	0.3	0.3	0	0	0	0	0	13	0.90
29	1.5	3	0.3	0.3	0.3	0	0	0	0	0	12.6	0.90
30	1.5	3	0.3	0.3	0.3	0	0	0	0	0	12.2	0.92
31	1.5	3	0.3	0.3	0.3	0	0	0	0	0	12.8	0.92
32	1.5	3	0.3	0.3	0.3	0	0	0	0	0	12.4	0.90

6=32. For statistical calculation, the experimental variables  $x_i$  has been coded as  $X_i$  as per the transformation equation:

$$X_i = \frac{x_i - x_0}{\Delta x_i}. \quad (1)$$

Here  $X_i$  is the dimensionless coded value of the  $i^{\text{th}}$  independent variable,  $x_i$  is the uncoded value of the  $i^{\text{th}}$  independent variable,  $x_0$  is the value of  $x_i$  at the center point, and  $\Delta x_i$  is the step change value of the real variable  $x_i$ .

In this study, the five independent process variables chosen were Glucose ( $x_1$ ), Monosodium glutamate ( $x_2$ ), Yeast extract ( $x_3$ ),  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  ( $x_4$ ), and  $\text{K}_2\text{HPO}_4$  ( $x_5$ ). Five levels ( $\pm\beta, \pm 1, 0$ : where  $\beta = 2^{4/4} = 2$ ) and six replicates at the central points were used to design the experiments. The levels of lowest, low, center, high and highest for the design variables in g/l were specified as  $x_1$ : (0.5, 1, 1.5, 2, 2.5),  $x_2$ : (1, 2, 3, 4, 5),  $x_3$ : (0.1, 0.2, 0.3, 0.4, 0.5),  $x_4$ : (0.1, 0.2, 0.3, 0.4, 0.5), and  $x_5$ : (0.1, 0.2, 0.3, 0.4, 0.5), respectively.

The response variables measured were the biomass ( $Y_1$ ) and lipopeptide concentrations ( $Y_2$ ): note that  $Y_i$  refers to (dimensional) concentration in g/l of  $i^{\text{th}}$  response variable. Six replicates at the center of the design were used to estimate the sum of squares error. Experiments were randomized in order to maximize the effects of unexplained variability in the observed responses due to extraneous factors. The experiments were conducted according to the CCRD design given in table 1.

## 4. Modeling and optimization

### 4.1 Response surface methodology (RSM)

The first step in RSM is to find a suitable approximation for the true functional relationship between the response ( $Y$  in g/l) and the set of independent variables. An important assumption is that the independent variables are continuous and controllable by experiments with negligible errors. Curvature is present in our system, and hence the RSM we use is a polynomial of high degree (the second-order model):

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \beta_{ij} X_i X_j + \varepsilon, \quad (2)$$

where  $\beta_{ii}$  represents the quadratic effect of the  $i^{\text{th}}$  factor and  $\beta_{ij}$  represents the cross product effect, or interaction effect, between the  $i^{\text{th}}$  and  $j^{\text{th}}$  factors.

### 4.2 Ant colony optimization (ACO)

ACO introduced in Dorigo *et al* [32] is one of the most recent techniques to solve optimization problems. The ACO mimics the way real ants find the shortest route between a food source and their nest exploiting a mix of probabilistic behavior and pheromone depositing. The ants communicate with one another by means of pheromone trails and

exchange information about which path should be followed. This autocatalytic and collective behavior of ants results in the establishment of the shortest route from the nest to the food source and back. This pheromone-mediated intelligent foraging behavior of ants is exploited by means of ACO algorithm to solve a number of optimization problems [33].

In this work, the ACO algorithm with its global search features is used to optimize the media compositions (g/l): glucose ( $x_1$ ), monosodium glutamate ( $x_2$ ), yeast extract ( $x_3$ ),  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  ( $x_4$ ),  $\text{K}_2\text{HPO}_4$  ( $x_5$ ) involved in the production of lipopeptide surfactin by *Bacillus subtilis*. An objective function is defined based on the actual process measurements and the model predictions as given by

$$J = f(\theta) = \sum_{i=1}^l (Y_i - \hat{Y}_i)^2. \quad (3)$$

where  $J$  is the cost function,  $\theta$  is the vector of parameters,  $l$  is the number of observations,  $Y_i$  is the measured value of the  $i^{\text{th}}$  variable, and  $\hat{Y}_i$  is the corresponding predicted value. Iterative convergence of this equation leads to minimization of the value of  $J$  thus providing the optimum parameter values.

The ACO-IM problem can be stated as

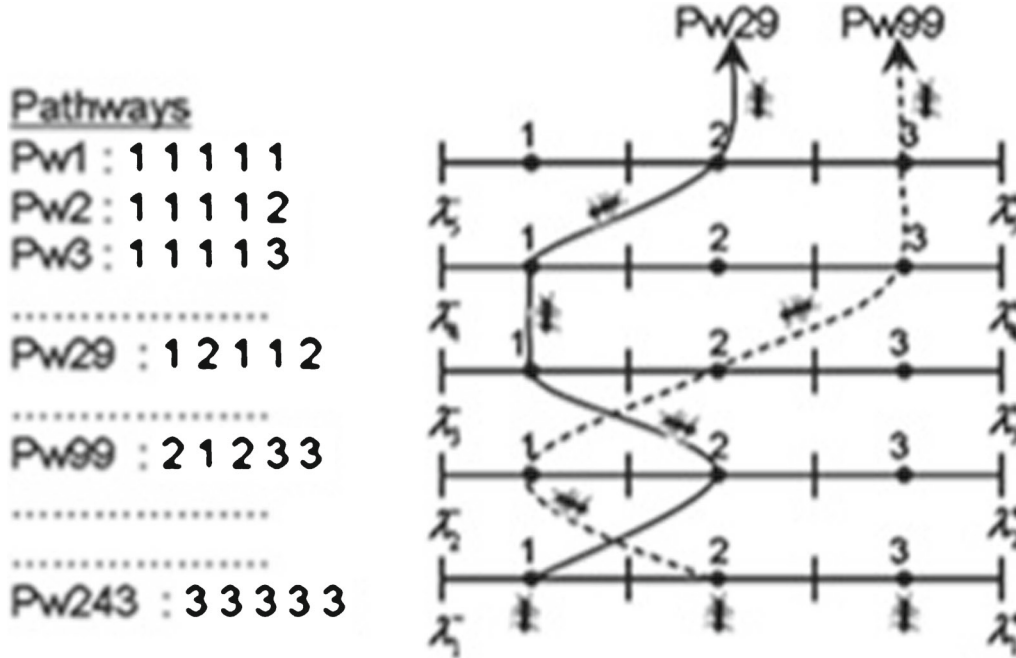
$$f(\theta) : D \rightarrow R, \quad (4)$$

where  $\theta$  is  $p$ -dimensional parameter vector,  $R$  is the space of real numbers and  $D$  is the search space of  $\theta$ . The objective is to find the parameter vector  $\theta$  that minimizes  $f(\theta)$ .  $D$  is considered to be a hyper parallelepiped

$$D = \{\theta_i \mid \theta_i^- \leq \theta_i \leq \theta_i^+ \}; \quad i = 1, \dots, p, \quad (5)$$

where  $\theta_i^-$  and  $\theta_i^+$  denote the lower and upper levels of parameter  $\theta_i$ . The search space  $D$  is first divided in the interval  $[\theta_i^-, \theta_i^+]$  for each parameter  $\theta_i$  into a number of strata,  $m_i$ . If each stratum be represented by the value at the middle of the stratum, then there will be  $M = m_1, \dots, m_p$  permutations or possible pathways through the search space of input parameters. Let  $\theta_{ij}$  ( $i = 1 \dots p$  and  $j = 1 \dots m_i$ ) be the stratum  $j$  of parameter  $\theta_i$  in the search space  $D$ . The parameters can be initially specified such that they uniformly cover the whole parameter space. Use of more strata can speed up the convergence of the optimization problem but the increase of strata would increase the computational burden.

Let there be  $M$  ants representing  $M$  strata. The possible number of pathways,  $N$ , that the ants can travel through the parameter space  $p$  is given by  $N = M^p$ . For instance, each ant follows one of the  $N$  pathways from the pathway structure list shown in the left side of figure 1. Accordingly, each ant will follow any one of the  $N$  pathways, e.g., pathway 29 (1,2,1,1,2), or pathway 99 (2,1,2,3,3), as shown in figure 1. The ants perform the tasks such as selection of pathways that they pass, remembering the parameter strata along the pathways, passing these parameter values to the model of



**Figure 1.** Graphical representation of discretized parameter space and the pathway structure.

the process, evaluating the value of objective function for each path way, and updating the pheromone based on the objective function values.

The computations involved in ACO implementation are as follows. The trail intensity also called the pheromone deposit,  $\tau_u$  on each pathway  $u$  ( $u = 1 \dots N$ ) is computed based on the cost function values of  $J$ . The cost function  $J$  of each pathway is represented as  $J_u$  and the minimum of the cost functions,  $J_{min}$  is found. The mean,  $\mu_J$  and the standard deviation,  $\sigma_J$  of the objective functions are evaluated. The ratio of standard deviation to mean represents the variability in the function values and the critical cost  $C$  which is evaluated as

$$C = J_{min} + C_c \frac{\sigma_J}{\mu_J}, \quad (6)$$

where  $C_c$  is a constant.

The critical cost,  $C$  is used to evaluate the trail intensity,  $\tau_u$  on each path according to the formula

$$\tau_u = \begin{cases} \exp \left[ 4.6 \left( \frac{J_u - C}{J_{min} - C} \right) \right], & J_u \leq C \\ 0, & J_u > C \end{cases} \quad (7)$$

According to this equation, pathways with the cost function larger than  $C$  receive no trail intensity, while those below  $C$  receive larger trail intensity values. It can be observed that any single stratum  $\phi_{ij}$  in  $D$  may be the crossroad of many ant pathways. The trail share  $\Gamma_u$  calculated of each stratum  $\theta_{ij}$  from each pathway can be summed to yield  $\phi_{ij}$  as  $\phi_{ij} = \sum_{u=1}^N \tau_u$   $u \in$  crossing pathway.

Here, the crossing pathways are those that cross the stratum  $\phi_{ij}$ . The mean,  $\mu_{ij}$  and the standard deviation,  $\sigma_{ij}$  of the

cost functions are used to compute the scores,  $S_{ij}$  according to the formula:

$$S_{ij} = \frac{(\phi_{ij})^A (\sigma_{ij})^B}{\sum_i \sum_j (\phi_{ij})^A (\sigma_{ij})^B}, \quad (8)$$

where

$$B = C_S \frac{\sigma_{ij}}{\mu_{ij}}, \quad (9)$$

where  $A$  and  $C_S$  are constants. Eqs. (8) and (9) in ACO structure represent trail intensity and transition probability [34]. The scores,  $S_{ij}$  are used to eliminate the strata with lowest scores while retaining the highest score. The interval of highest retained score results in narrowing the ranges for the parameters. This strategy is subsequently reinitialized with updated parameter ranges and the procedure is repeated until the convergence in objective defined by Eq. (3) is achieved. The constant  $C_c$  in Eq. (6) signifies the variability of the trail received by each stratum.

## 5. Results and discussion

### 5.1 Experimental observations

The lower and upper ranges considered for the media compositions in CCRD design were: glucose ( $x_1$ ): 0.5–2.5 g/l, monosodium glutamate ( $x_2$ ): 1–5 g/l, yeast extract ( $x_3$ ): 0.1–0.5 g/l,  $MgSO_4 \cdot 7H_2O$  ( $x_4$ ): 0.1–0.5 g/l, and  $K_2HPO_4$  ( $x_5$ ): 0.1–0.5 g/l. Response data was generated for biomass and lipopeptide productivities. The design matrix of 32 experiments with the actual and coded compositions, and the response data are shown in table 1. The

carbon source and nitrogen source were observed to influence the biosurfactant production. The experimental results have shown that the media composition with glucose ( $x_1$ ): 2 g/l, monosodium glutamate ( $x_2$ ): 4 g/l, yeast extract ( $x_3$ ): 0.2 g/l,  $MgSO_4 \cdot 7H_2O$  ( $x_4$ ): 0.4 g/l and  $K_2HPO_4$  ( $x_5$ ): 0.4 g/l yielded maximum lipopeptide productivity. Increase of glucose or monosodium glutamate is observed to increase the biosurfactant productivity. Studies by various researchers have shown that the presence of glucose in the production medium increased the biosurfactant production [35, 36]. The presence of nitrogen can play an important role in the regulation of biosurfactant synthesis, and it was found that nitrogen plays an important role in the production of surface-active compounds by microbes [1, 37]. Decreasing glucose or monosodium glutamate concentration did not show any improvement in biosurfactant production. Increase of yeast extract has no influence on the production of biosurfactant. Decrease of  $MgSO_4 \cdot 7H_2O$  and  $K_2HPO_4$  also did not show any improvement in biosurfactant production.

## 5.2 Response surface model

The aim of developing a response surface model based on the experimental design data is to employ these models to optimize the compositions of McKeen medium so as to maximize lipopeptide production. The data generated from the designed experiments for lipopeptide surfactin production by *Bacillus subtilis* was used to develop response surface models (RSM). A second order full regression model with factor interactions was considered to build relations between the factors and responses. The form of the model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{24} X_{24} + \beta_{34} X_{34} + \beta_{15} X_{15} + \beta_{25} X_{25} + \beta_{35} X_{35} + \beta_{45} X_{45} + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{44} X_4^2 + \beta_{55} X_5^2 \quad (10)$$

The experimental data in table 1 was used to determine the regression coefficients of the second-order multiple regression models. The coefficients of the models representing the biomass and lipopeptide concentrations were determined by using the method of least squares. The empirical

models identified for biomass and lipopeptide productivities are given as follows:

$$Y_{bio} = 12.997 + 0.260X_1 + 0.158X_2 + 0.975X_3 - 0.025X_4 - 0.192X_5 - 0.063X_{12} + 0.163X_{13} - 0.138X_{23} - 0.113X_{14} + 0.038X_{24} + 0.113X_{34} - 0.688X_{15} - 0.138X_{25} - 0.013X_{35} - 0.038X_{45} - 3.051X_1^2 - 2.92X_2^2 - 1.771X_3^2 - 1.396X_4^2 - 0.321X_5^2 \quad (11)$$

$$Y_{lipo} = 8.802 - 0.023X_1 - 0.717X_2 + 1.217X_3 + 1.033X_4 - 0.667X_5 - 0.4X_{12} - 0.25X_{13} + 0.625X_{23} - 0.65X_{24} - 1.25X_{34} - 1.625X_{15} - 0.475X_{25} + 0.00001X_{35} + 0.001X_{45} - 0.640X_1^2 - 0.877X_2^2 - 0.277X_3^2 + 0.973X_4^2 + 1.198X_5^2 \quad (12)$$

The validity of the second order regression models was studied by F test and the significance of the regressed model coefficients was evaluated by student t test as described below.

## 5.3 Analysis of variance and test of significance

The validity of the full regression models was studied using analysis of variance (ANOVA). Table 2(a) and 2(b) showed the ANOVA results of quadratic response-surface model fitting for both response  $Y_{bio}$  (Biomass concentration) and  $Y_{lipo}$  (Lipopeptide concentration), respectively. The higher values of calculated  $F$  than the table values and corresponding lower  $P$  values indicate less significance of the full order models. The  $R^2$  values of these models were found to be 0.901 and 0.883. The significance of each of the coefficients of the full order regression model was determined by using student  $t$ -test. The student  $t$  values and the corresponding  $P$  values are given in columns 4 and 5 of table 3 and table 4, respectively. A large magnitude for  $t$ -value, and small  $P$ -value, indicates the higher significance of the corresponding coefficient. The elimination of insignificant coefficients led to the following reduced models:

$$Y_{bio} = 12.997 + 0.260X_1 + 0.158X_2 + 0.975X_3 - 0.192X_5 - 0.063X_{12} + 0.163X_{13} - 0.138X_{23} - 0.113X_{14} + 0.038X_{24} + 0.113X_{34} - 0.688X_{15} - 0.138X_{25} - 0.038X_{45} - 3.051X_1^2 - 2.92X_2^2 - 1.771X_3^2 - 1.396X_4^2 - 0.321X_5^2 \quad (13)$$

**Table 2.** ANOVA results of full order regression models: (a) Biomass; (b) Lipopeptide.

Source	Sum of squares	Degree of freedom	Mean square	$F_{cal} < F_{21,10 \text{ table}}$	P
(a)					
Regression	1702.474	21	85.124	32.88 < 2.32	0.05
Residual	51.778	10	2.354		
Total	1754.248	31			
(b)					
Regression	1289.336	21	64.467	SSR/SSE=33.91 < 2.32	0.05
Residual	38.014	10	1.728		
Total	1327.344	31			

**Table 3.** Test of significance for the coefficients of Biomass regression model.

Intercept	Coefficient estimate	Standard error	$t_{table} < t_{calculated}; t_{0.05,31} = 1.695$	P-value
X <sub>0</sub>	12.997	0.050	260.491	< 0.001
X <sub>1</sub>	0.260	0.030	8.609	0.0005
X <sub>2</sub>	0.158	0.025	6.210	< 0.05
X <sub>3</sub>	0.975	0.025	38.243	< 0.0005
X <sub>4</sub>	-0.025	0.025	-0.981	< 0.2
X <sub>5</sub>	-0.192	0.025	-7.518	0.001
X <sub>1</sub> X <sub>2</sub>	-0.063	0.031	-2.002	0.05
X <sub>1</sub> X <sub>3</sub>	0.163	0.031	5.204	0.001
X <sub>2</sub> X <sub>3</sub>	-0.138	0.031	-4.404	0.0005
X <sub>1</sub> X <sub>4</sub>	-0.113	0.031	-3.603	0.003
X <sub>2</sub> X <sub>4</sub>	0.038	0.031	1.201	0.15
X <sub>3</sub> X <sub>4</sub>	0.113	0.031	3.603	0.001
X <sub>1</sub> X <sub>5</sub>	-0.688	0.031	-22.018	0.001
X <sub>2</sub> X <sub>5</sub>	-0.138	0.031	-4.404	0.0005
X <sub>3</sub> X <sub>5</sub>	-0.013	0.031	-0.400	< 0.25
X <sub>4</sub> X <sub>5</sub>	-0.038	0.031	-1.201	0.15
X <sub>1</sub> <sup>2</sup>	-3.052	0.026	-116.642	< 0.025
X <sub>2</sub> <sup>2</sup>	-2.922	0.023	-125.809	0.003
X <sub>3</sub> <sup>2</sup>	-1.772	0.023	-76.292	< 0.003
X <sub>4</sub> <sup>2</sup>	-1.397	0.023	-60.145	< 0.003
X <sub>5</sub> <sup>2</sup>	-0.322	0.023	-13.857	< 0.02

**Table 4.** Test of significance for the coefficients of lipopeptide regression model.

Intercept	Coefficient estimate	Standard error	$t_{table} < \text{absolute } t_{calculated}; t_{0.05,31} = 1.695$	P-value
X <sub>0</sub>	8.802	0.334	26.370	< 0.01
X <sub>1</sub>	-0.023	0.202	-0.116	0.25
X <sub>2</sub>	-0.717	0.171	-4.202	<0.025
X <sub>3</sub>	1.217	0.171	7.133	0.001
X <sub>4</sub>	1.033	0.171	6.058	< 0.05
X <sub>5</sub>	-0.667	0.171	-3.909	< 0.001
X <sub>1</sub> X <sub>2</sub>	-0.40	0.209	-1.915	0.15
X <sub>1</sub> X <sub>3</sub>	-0.25	0.209	-1.197	0.15
X <sub>2</sub> X <sub>3</sub>	0.625	0.209	2.992	0.02
X <sub>1</sub> X <sub>4</sub>	-0.225	0.209	-1.077	< 0.2
X <sub>2</sub> X <sub>4</sub>	0.15	0.209	0.718	0.25
X <sub>3</sub> X <sub>4</sub>	-0.65	0.209	-3.112	0.01
X <sub>1</sub> X <sub>5</sub>	-1.25	0.209	-5.984	0.0005
X <sub>2</sub> X <sub>5</sub>	-1.625	0.209	-7.779	< 0.003
X <sub>3</sub> X <sub>5</sub>	0.475	0.209	2.274	0.025
X <sub>4</sub> X <sub>5</sub>	0.001	0.209	0.0001	< 0.25
X <sub>1</sub> <sup>2</sup>	-0.640	0.175	-3.659	0.003
X <sub>2</sub> <sup>2</sup>	-0.877	0.155	-5.645	< 0.0005
X <sub>3</sub> <sup>2</sup>	-0.277	0.155	-1.784	0.0005
X <sub>4</sub> <sup>2</sup>	0.973	0.155	6.262	< 0.05
X <sub>5</sub> <sup>2</sup>	1.198	0.155	7.710	< 0.001

$$\begin{aligned}
 Y_{lipo} = & 8.802 - 0.717X_2 + 1.217X_3 + 1.033X_4 - 0.667X_5 \\
 & - 0.4X_{12} - 0.25X_{13} + 0.625X_{23} - 1.25X_{34} \\
 & - 1.625X_{15} - 0.475X_{25} + 0.00001X_{35} + 0.640X_1^2 \\
 & - 0.877X_2^2 - 0.277X_3^2 + 0.973X_4^2 + 1.198X_5^2
 \end{aligned}
 \tag{14}$$

The resulting reduced models were statistically validated and tested for their predictive ability by comparing with the

experimental results. The ANOVA results of the reduced models are shown in table 5(a) and 5(b), respectively. The lower values of calculated *F* compared to the table values and the corresponding higher *P* values indicate the adequacy of the reduced models. The *R*<sup>2</sup> values of reduced models representing the biomass and lipopeptide were calculated as 0.912 and 0.904. The experimental and model prediction results of biomass and lipopeptide concentrations were

**Table 5.** ANOVA results of reduced order regression models: (a) Biomass; (b) Lipopeptide.

Source	Sum of squares	Degree of freedom	Mean square	F <sub>cal</sub> < F <sub>18,13table</sub>	P
(a)					
Regression	16.089	18	0.88	12.15 < 12.16	0.05
Residual	1.9	13	0.344		
Total	17.999	31			
(b)					
Regression	14.089	18	0.725	10.1 < 12.1	0.05
Residual	2.9	13	0.144		
Total	16.999	31			

shown in figures 2 and 3, respectively. These results showed the predictive ability of the fitted regression models for both the biomass and lipopeptide concentrations.

Sensitivity analysis was performed to find the media components that significantly influence the lipopeptide productivity. The variables that most influence the lipopeptide productivity were found by using the normalized sensitivity studies. The normalized sensitivities are evaluated by using the relation:  $\left(\frac{\partial Y}{\partial X_i}\right) / \left(\frac{Y}{X_i}\right)$ . Here  $Y$  refers to the lipopeptide concentration and  $X_i$  refers to the respective independent variable, i.e., the (non-dimensional) media composition. The absolute normalized sensitivity values of lipopeptide productivity with respect to glucose ( $X_1$ ), monosodium glutamate ( $X_2$ ), yeast extract ( $X_3$ ),  $MgSO_4 \cdot 7H_2O$  ( $X_4$ ), and  $K_2HPO_4$  ( $X_5$ ) were computed as 2.026, 2.283, 0.659, 3.110, and 3.795, respectively. This sensitivity analysis shows that the variables  $X_4$  and  $X_5$  exhibit greater influence on the productivity of lipopeptide. The contour plots represent the effect of the significant variables and their interaction in the response variable. The response surface counter plot drawn for  $X_5$  vs.  $X_4$  on lipopeptide productivity is given in figure 4.

#### 5.4 Optimization of lipopeptide production

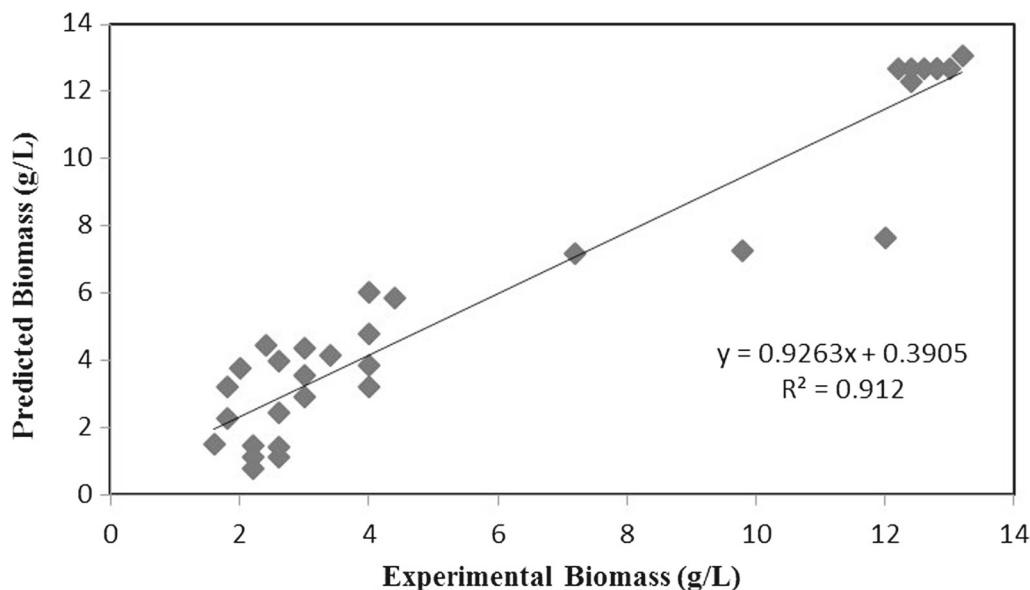
The reduced order response surface model detailed in the previous section was coupled with ACO to optimize the compositions of fermentation process medium components (Glucose, Monosodium glutamate, Yeast extract,  $MgSO_4 \cdot 7H_2O$ , and  $K_2HPO_4$ ) so as to maximize the lipopeptide productivity. The optimization problem is stated as

$$\text{Maximize : } Y_{lipo}(x_1, x_2, x_3, x_4, x_5) \quad (15)$$

within the ranges of medium composition (in g/l):

$$\begin{aligned} 0.5 &\leq x_1 \leq 2.5 \\ 1 &\leq x_2 \leq 5 \\ 0.1 &\leq x_3 \leq 0.5 \\ 0.1 &\leq x_4 \leq 0.5 \\ 0.1 &\leq x_5 \leq 0.5 \end{aligned} \quad (16)$$

The ACO was designed and coupled with the reduced order empirical models to solve the optimization problem. The cost function  $J$  in Eq. (3) was defined based on the experimental and model predicted lipopeptide concentrations.

**Figure 2.** Predicted and experimental values of biomass ( $Y_{bio}$ ) concentration.



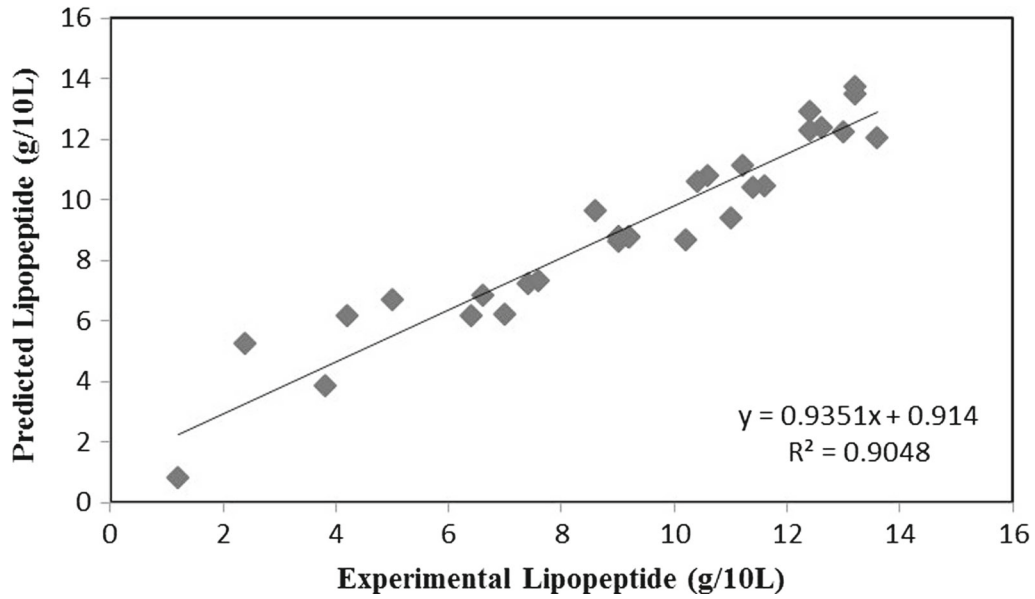


Figure 3. Predicted and experimental values of lipopeptide ( $Y_{lipo}$ ) concentration.

The dimension of the vector ( $\alpha$ ) representing the medium compositions was set as 5. The number of strata ( $M$ ) was assigned as 3. The number of ant sets chosen is 20. The constants involved in Eqs. (6), (8) and (9) were appropriately

tuned as  $C_c = 0.5$ ,  $A = 1.0$  and  $C_s = 0.3$ . The ACO algorithm was implemented iteratively to provide the updated values of compositions until convergence in the cost function was achieved. The number of pathways generated for

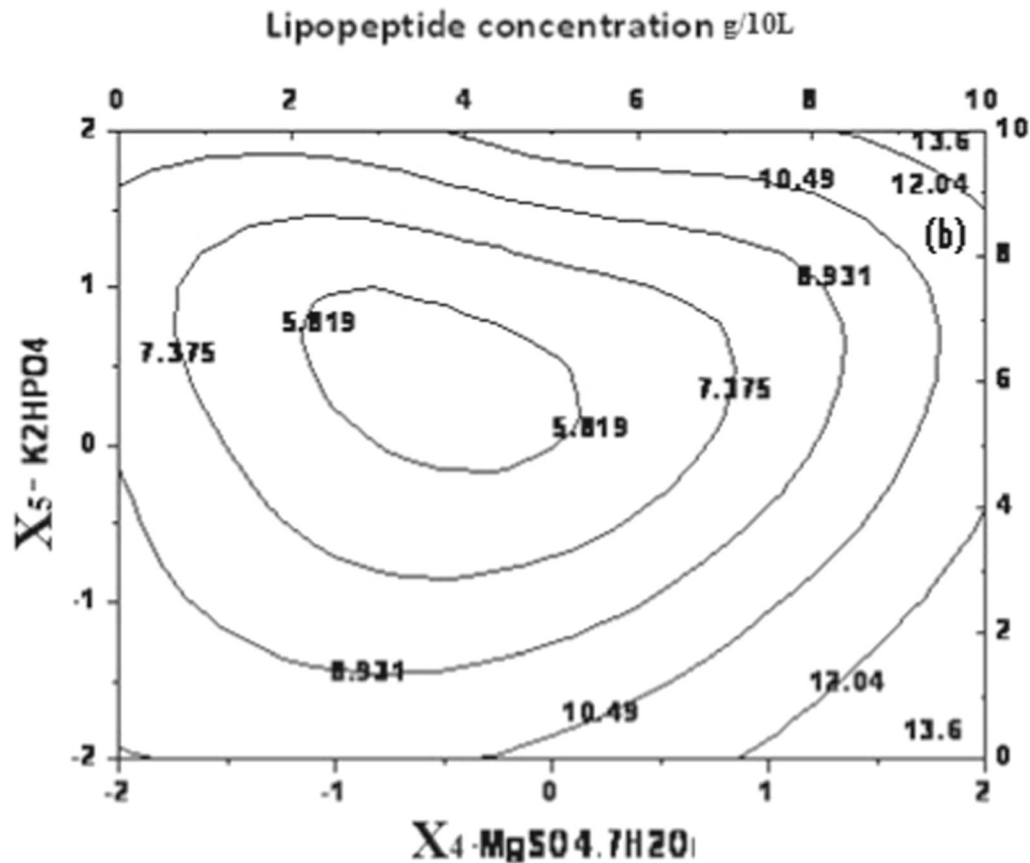
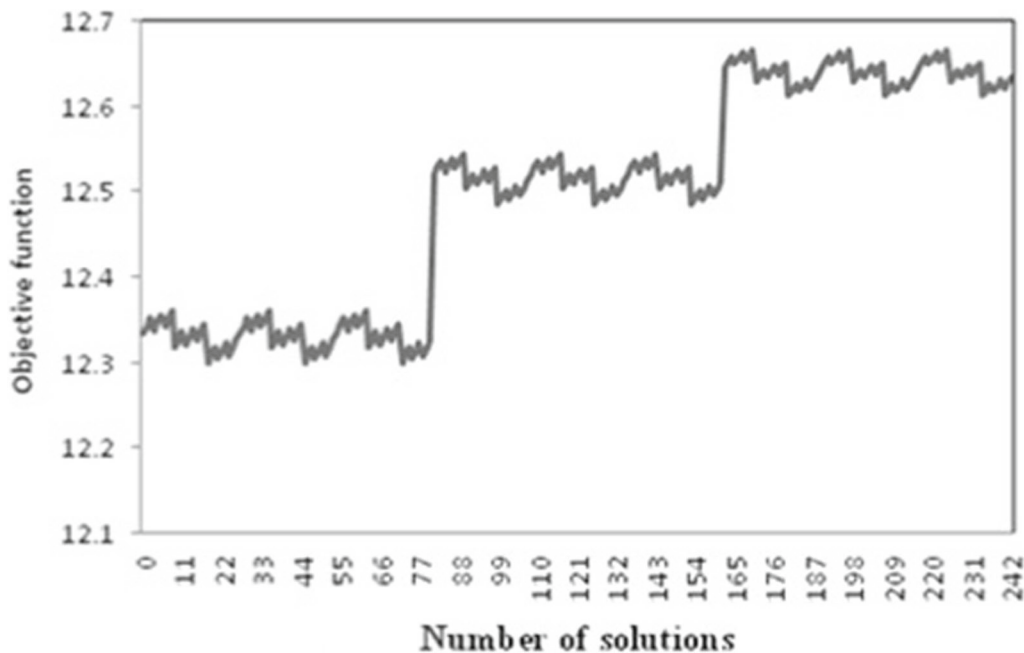


Figure 4. Response surface counter plot of  $X_5$  vs.  $X_4$  on lipopeptide concentration.



**Figure 5.** Convergence in objective function of ACO with ants=20, strata=3, and estimated parameters=5.

ants' travel was 243. The convergence in optimal solution defining the lipopeptide productivity evolved for each ant set by ACO was shown in figure 5. The ACO was executed by writing the program in C language. The optimized medium composition by ACO was found as  $x_1 = 1.098$  g/l,  $x_2 = 4.01$  g/l,  $x_3 = 0.426$  g/l,  $x_4 = 0.431$  g/l, and  $x_5 = 0.219$  g/l. The maximum lipopeptide concentration obtained due to the optimized medium composition was 1.501 g/l ( $y_{lipo}$ ) and the corresponding biomass concentration was obtained as 4.291 g/l ( $y_{bio}$ ).

A classical Nelder–Mead optimization (NMO) method [38] was also employed to compare with ACO. The NMO has been successfully applied for modeling and optimization of many chemical and biological problems [39, 40]. The tuning parameters involved in NMO were the reflection, contraction and expansion coefficients, which were set as 1.0, 2.0 and 0.5, respectively. The optimum medium components found by NMO are  $x_1 = 2$  g/l,  $x_2 = 3.8$  g/l,  $x_3 = 0.2$  g/L,  $x_4 = 0.4$  g/l and  $x_5 = 0.4$  g/l with the maximum lipopeptide concentration of 1.387 g/L and the corresponding biomass concentration was obtained as 3.99 g/l. The NMO predicted optimum lipopeptide concentration 1.387 g/l was found to be lower than the 1.498 g/l that was predicted by ACO. The results show that optimizing the medium composition by ACO can identify the maximum lipopeptide productivity more accurately. The experimentally validated lipopeptide concentration based on ACO optimized media composition was found to be 1.498 g/L. These results thus exhibit the effectiveness of ACO in optimizing the media composition for lipopeptide production.

## Acknowledgements

Financial assistance from DST through the grant SR/WOS-A/ET-20/2009 is gratefully acknowledged.

## References

- [1] Desai J D and Banat I M 1997 Microbial production of surfactants and their commercial potential. *Microbiol. Mol. Biol. R.* 61(1): 47–64
- [2] Abdel-Mawgoud A M, Lépine F and Déziel E 2010 Rhamnolipids: Diversity of structures, microbial origins, and roles. *Appl. Microbiol. Biot.* 86(1): 323–1336
- [3] Morikawa M, Hirata Y and Imanaka T 2000 A study on the structure–function relationship of lipopeptide biosurfactants. *BBA-Mol. Cell Biol. L.* 1488(3): 211–218
- [4] Peypoux F, Bonmatin J M and Wallach J 1999 Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biot.* 51(5): 553–563
- [5] Haddad N I, Liu X, Yang S and Mu B 2008 Surfactin isoforms from *bacillus subtilis* HSO121: Separation and characterization. *Protein. Peptide Lett.* 15(3): 265–269
- [6] Tang J, Gao H, Hong K, Yu Y, Jiang M and Lin H 2007 Complete assignments of  $^1H$  and  $^{13}C$  NMR spectral data of nine surfactin isomers. *Magn. Reson. Chem.* 45(9): 792–796
- [7] Cameotra S S and Makkar R S 2004 Recent applications of biosurfactants as biological and immunological molecules. *Curr. Opin. Microbiol.* 7(3): 262–66
- [8] Abalos A, Maximo F, Manresa M A and Bastida J 2002 Utilization of response surface methodology to optimize the culture media for the production of rhamnolipids by *Pseudomonas Aeruginosa* AT10. *J. Chem. Technol. Biot.* 77: 777–784

- [9] Al-Araji L I Y, Abd-Rahman R N Z R, Basri M and Salleh A B 2007 Optimization of rhamnolipids produced by *Pseudomonas aeruginosa* 181 using response surface modeling. *Ann. Microbiol.* 57(4): 571–575
- [10] Rispoli F J, Badia D and Shah V 2010 Optimization of the fermentation media for sophorolipid production from *Candida bombicola* ATCC 22214 using a simplex centroid design. *Biotechnol. Progr.* 26(4): 938–944
- [11] Rikalović M G, Gojgić-cvijović G, Vrvić M M and Karadzic I 2012 Production and characterization of rhamnolipids from *Pseudomonas Aeruginosa*-AI. *J. Serb. Chem. Soc.* 77: 27–42
- [12] Suwansukho P, Rukachisirikul V, Kawai F and H-Kittikun A 2008 Production and applications of biosurfactant from *Bacillus subtilis* MUV4, Songklanakarin. *J. Sci. Technol.* 30(Suppl 1): 87–93
- [13] Gu X, Zheng Z, Yu H, Wang J, Liang F and Liu R 2005 Optimization of medium constituents for a novel lipopeptide production by *Bacillus Subtilis* MO-01 by a response surface method. *Process Biochem.* 40(10): 3196–3201
- [14] Abdel-Mawgoud M, Aboulwafa M M and Hassouna N A 2008 Optimization of surfactin production by *bacillus subtilis* isolate BS5. *Appl. Biochem. Biotech.* 150: 305–325
- [15] Liu X, Ren B, Gao H, Liu M, Dai H, Song F, Yu Z, Wang S, Hu J, Kokare C R and Zhang L 2012 Optimization for the production of surfactin with a new synergistic antifungal activity. *PLoS ONE* 7(5): e34430
- [16] Mutalik S R, Vaidya B K, Joshi R M, Desai K M and Nene S N 2008 Use of response surface optimization for the production of biosurfactant from *Rhodococcus spp.* MTCC 2574. *Bioresource Technol.* 99(10): 7875–7880
- [17] Seghal Kiran S, Anto-Thomas T, Joseph S, Sabarathnam B and Lipton A P 2010 Optimization and characterization of a new biosurfactant produced by marine *Brevibacterium aureum* MSA 13 in solid state culture. *Bioresource Technol.* 101: 2389–2396
- [18] De Lima C J, Ribeiro E J, Servulo E F, Resende M M and Cardoso V L 2009 Biosurfactant production by *Pseudomonas aeruginosa* grown on residual soybean oil. *Appl. Biochem. Biotech.* 152(1): 156–168
- [19] Pal M P, Vaidya B K, Desai K M, Joshi R M, Nene S N and Kulkarni B D 2009 Media optimization for biosurfactant production by *Rhodococcuserythropolis* MTCC 2794: Artificial intelligence versus a statistical approach. *J. Ind. Microbiol. Biot.* 36: 747–756
- [20] Satya Eswari J, Anand M and Venkateswarlu C 2013 Optimum culture medium composition for Rhamnolipid production by *Pseudomonas Aeruginosa* AT10 using a novel multi-objective optimization method. *J. Chem. Technol. Biot.* 88(2): 271–279
- [21] Chen L Z, Nguang S K, Chen X D and Li X M 2004 Modeling and optimization of fed-batch fermentation processes using dynamical neural networks and genetic algorithms. *Biochem. Eng. J.* 22(1): 51–61
- [22] Deb K 2001 *Multi-objective optimization using evolutionary algorithms*. New York: John Wiley & Sons
- [23] Shopova E G and Vaklieva-Bancheva N G 2006 BASIC-A genetic algorithm for engineering problem solution. *Comput. Chem. Eng.* 30(8): 1293–1309
- [24] Faber R, Jockenhovet T and Tsatsaronis G 2005 Dynamic optimization with simulated annealing. *Comput. Chem. Eng.* 29: 273–290
- [25] Kirkpatrick S, Gelatt Jr. C D and Vecchi M P 1983 Optimization by simulated annealing. *Science* 220(4598): 671–680
- [26] Blum C 2005 Beam-ACO- hybridizing ant colony optimization with beam search: An application to open shop scheduling. *Comput. Oper. Res.* 32(6): 1565–1591
- [27] Dorigo M, Birattari M and Stutzle T 2006 Ant colony optimization. *IEEE Comput. Intell. Mag.* 1(4): 28–39
- [28] Socha K, Sampels M and Manfrin M 2003 Ant algorithms for the university course timetabling problem with regard to the state-of-the-art. *Proceedings of the Third European Workshop on Evolutionary Computation in Combinatorial Optimization*, Essex: UK, pp. 334–345
- [29] Parsopoulos K E and Vrahatis 2002 Recent approaches to global optimization problems through particle swarm optimization. *Nat. Comput.* 1(2–3): 235–306
- [30] Anand P, Bhagvanth Rao M and Venkateswarlu C 2013 Multi-stage dynamic optimization of a copolymerization reactor using differential evolution. *Asia Pac. J. Chem. Eng.* 8(5): 687–698
- [31] Satya Eswari J and Venkateswarlu C 2012 Optimization of culture conditions for Chinese Hamster Ovary (CHO) cells production using differential evolution. *Int. J. Pharm. Pharm. Sci.* 4(1): 465–470
- [32] Dorigo M, Maniezzo V and Colorni A 1996 Ant system: Optimization by a colony of cooperating agents. *IEEE. Trans. Syst. Man Cybern. B: Cybern.* 26(1): 29–41
- [33] Dorigo M and DiCaro G 1999 *The ant colony meta-heuristic in “New Ideas in Optimization”*. New York: McGraw-Hill
- [34] Dorigo M, Bonabeau E and Theraulaz G 2000 Ant algorithms and stigmergy. *Future Gener. Comp. Syst.* 16(8): 851–871
- [35] Guerra-Santos L, Kappeli O and Fiechter A 1984 *Pseudomonas aeruginosa* biosurfactant production in continuous culture with glucose as carbon source. *Appl. Environ. Microb.* 48(2): 301–305
- [36] Wei Q F, Mather R R and Fotheringham A F 2005 Oil removal from used sorbents using a biosurfactant. *Bioresource Technol.* 96(3): 331–334
- [37] Cameotra S S and Makkar R S 1998 Synthesis of biosurfactants in extreme conditions. *Appl. Microbiol. Biot.* 50: 520–529
- [38] Kuester J L and Mize J H 1973 *Optimization techniques with Fortran*. New York: McGraw-Hill
- [39] Satya Eswari J and Venkateswarlu C 2013 Evaluation of anaerobic biofilm reactor kinetic parameters using ant colony optimization. *Environ. Eng. Sci* 30(9): 527–535
- [40] Venkateswarlu C and Gangiah K 1992 Dynamic modeling and optimal state estimation using extended Kalman filter for a kraft pulping digester. *Ind. Eng. Chem. Res.* 31(3): 848–855