



Application of the Artificial Neural Network and Response Surface Methodology Models for Predicting the Enzymatic Synthesis of Betulinic Acid Ester: A Comparative Study

M. GHAFFARI MOGHADDAM¹, F.B.H. AHMAD^{2,*}, M. BASRI² and M.B. ABDUL RAHMAN²

¹Department of Chemistry, Faculty of Science, University of Zabol, Zabol, Iran

²Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

*Corresponding author: E-mail: faujan@science.upm.edu.my; faujanahmad@gmail.com

(Received: 3 November 2011;

Accepted: 23 July 2012)

AJC-11867

In this paper, the estimation capabilities of the response surface methodology and artificial neural network, in an enzymatic reaction catalyzed by *Candida antarctica* lipase (Novozym 435), were investigated. The experiments were conducted based on a five-level, four-variable central composite rotatable design. The yield of ester in the enzymatic reaction was considered as a function of four independent variables, namely reaction time, reaction temperature, enzyme amount and substrate molar ratio. After predicting the model using response surface methodology and artificial neural network, two methodologies were then compared for their modeling. The results showed that the artificial neural network model is much more accurate in prediction as compared to the response surface methodology.

Key Words: Response surface methodology, Artificial neural network, Predictive capability, Enzymatic esterification, Betulinic acid.

INTRODUCTION

Betulinic acid (1) is a natural product which can be found in the outer bark of various *Betula* species such as the outer bark of yellow (*Betula alleghaniensis*) and white (*Betula papyrifera*) birches^{1,2}. Betulinic acid possesses several pharmacological activities including inhibition of human immunodeficiency virus (HIV), antibacterial, antimalarial, anti-inflammatory, anthelmintic, antioxidant and anticancer properties³. This compound is regarded by the scientific community as an accessible and valuable bioactive natural product⁴. The introduction of the polar groups, such as phthalates at C-3 position of betulinic acid, is an interesting way to increase the hydrosolubility and anticancer activity of betulinic acid⁵.

Response surface methodology (RSM) is a collection of mathematical and statistical techniques which are useful for the modeling and analysis of problems in which a response is influenced by several variables⁶. In more specific, the response surface methodology is a useful technique to be used in optimizing and evaluating the interactive effects of independent factors in numerous chemical and biochemical processes such as ester production by lipase⁷. Moreover, it is also superior to the traditional approach, in which optimization studies are carried out by varying one parameter at a time, while keeping others constant⁸. The methodology involves three steps: (a) experimental design in which the independent variables and

their experimental levels are set using well-established statistical experimental designs such as the central composite design; (b) response surface modeling through regression analysis; and (c) process optimization using the response surface model⁹.

Artificial neural network (ANN) is a highly simplified model of the structure of a biological network¹⁰. The fundamental processing element of artificial neural network is an artificial neuron (or simply a neuron). A biological neuron receives inputs from other sources, combines them, generally performs a non-linear operation on the result and then outputs the final result¹¹. The ability of the artificial neural networks, to recognize and reproduce the cause-effect relationships through training for the multiple input-output systems makes them efficient to represent even the most complex systems¹². The main advantages of artificial neural network as compared to response surface methodology include: (i) artificial neural network does not require any prior specification of suitable fitting function and (ii) artificial neural network also has a universal approximation capability to approximate almost all kinds of non-linear functions including quadratic functions, whereas response surface methodology is useful only for quadratic approximations¹³.

There are some studies in the literature where models were developed based on response surface methodology and artificial neural network using the same experimental design¹³⁻¹⁷. For example, Basri *et al.*¹⁴ reported the comparison of artificial

TABLE-1
CODED AND ACTUAL LEVELS OF VARIABLES CONSIDERED FOR THE DESIGN

Variable	Unit	Coded level of variable				
		-2	-1	0	1	2
X ₁ : Reaction time	h	8	12	16	20	24
X ₂ : Temperature	°C	40	45	50	55	60
X ₃ : Amount of enzyme	mg	50	100	150	200	250
X ₄ : Substrate molar ratio (betulinic acid/anhydride)	-	0.2	0.4	0.6	0.8	1.0

neural network and response surface methodology in the lipase-catalyzed synthesis of palm-based wax ester and they suggested the superiority of artificial neural network over response surface methodology. Both the artificial neural network and response surface methodology techniques were recently compared for their predictive and generalization capabilities, sensitivity analysis and optimization efficiency in fermentation media optimization¹³. It was found that the artificial neural network predicted model has higher accuracy and better generalization capability than response surface methodology, even with the limited number of experiments.

In this study, the response surface methodology and artificial neural network methodologies were applied for predicting the yield of ester in the enzymatic synthesis of 3-*O*-phthalyl-betulinic acid (Fig. 1). The results which were obtained through response surface methodology were then compared with those through artificial neural network.

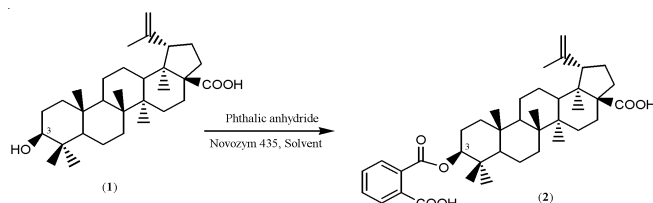


Fig.1. Enzymatic esterification of betulinic acid (1) with phthalic anhydride using lipase as a biocatalyst

EXPERIMENTAL

Immobilized enzyme (triacylglycerol hydrolase, EC 3.1.1.3; Novozym 435, 10000 PLU/g), *Candida antarctica* lipase, supported on a macroporous acrylic resin with a water content of 3 % (w/w), was purchased from Novo Nordisk A/S (Bagsvaerd, Denmark). Chloroform and *n*-hexane (Fisher Chemical, Loughborough, UK) were used as the organic solvents. Betulinic acid was isolated from the Malaysian *Callistemon speciosus* according to the procedure described by Ahmad *et al.*¹⁸. Phthalic anhydride was purchased from Acros Organics (Geel, Belgium). Ethyl acetate, Celite[®]545, Na₂SO₄, K₂CO₃ and HCl were purchased from Merck (Darmstadt, Germany). All the chemicals were of the analytical reagent grade.

Enzymatic esterification: To a magnetically stirred solution of betulinic acid (25 mg, 0.0547 mmol), K₂CO₃ (6 mg), Celite[®]545 (170 mg), different amounts of enzyme (50-250 mg), chloroform (10 mL) and hexane (10 mL) was added phthalic anhydride with difference molar ratio (betulinic acid/phthalic anhydride; 0.2-1). The reaction mixture was magnetically stirred (150 rpm) at different reaction temperatures (40-60 °C) and reaction times (8-24 h) as shown in Table-1. Each reaction was repeated in triplicate and the results represented the mean values of three independent experiments.

The control experiments were performed in the absence of enzyme. As a result, no chemical acyl transfer reaction was detected. Qualitative analysis of the reaction mixtures was made by thin layer chromatography (TLC) on silica gel plates eluted with system *n*-hexane/ethyl acetate (9:1, v/v). The plates were visualized under UV lamp and/or iodine vapour. Under these conditions, 3-*O*-phthalyl-betulinic acid (2) had an R_f of 0.9. The quantitative analysis of samples was carried out according to the procedure described by Kvasnica *et al.*,⁵. At the pre-determined time intervals, the flasks were taken and the enzyme was removed by filtration and washed with chloroform twice. The filtrate was evaporated to dryness and ethyl acetate was then added and washed with aqueous solution of HCl and twice with water. The organic layer was dried over Na₂SO₄ and concentrated under reduced pressure. The residue was chromatographed with gradient on silica gel 60 (*n*-hexane/ethyl acetate, 9:1-5:1, v/v). The ester fractions were combined and weighed after the evaporation of the solvents. The percentage of the isolated yield of ester (Yield %) is defined as:

$$\text{Yield}(\%) = \frac{\text{mmol isolated betulinic acid ester}}{\text{mmol initial betulinic acid}} \times 100 \quad (1)$$

The characterization of the product was made by recording the ¹H and ¹³C NMR spectra of the compound on a Varian Unity Inova 500 NMR spectrometer operating at 26 °C and this matched the data available in the literature⁵.

Experimental design for response surface methodology:

A 5-level-4-factor central composite rotatable design (CCRD) was employed in this study¹⁹. The fractional factorial design consisted of 16 factorial points, 8 axial points (two axial points on the axis of each design variable at a distance of 2 from the design center) and 6 center points. Table-1 presents the variables and their levels, which were selected for the study. Meanwhile, the actual experiments (in triplicate) which had been carried out to develop the model are shown in Table-2. The experimental data obtained based on the design were fitted to a second-order polynomial equation of the form:

$$Y = \beta_0 + \sum_{i=1}^4 \beta_i X_i + \sum_{i=1}^4 \beta_{ii} X_i^2 + \sum_{i=1}^3 \sum_{j=i+1}^4 \beta_{ij} X_i X_j \quad (2)$$

where, Y is the predicted response (percentage isolated yield of ester) used as a dependent variable; X_i and X_j represent the independent variables; β₀, β_i, β_{ii} and β_{ij} are constant and regression coefficients of the model. The subsequent regression analyses and analyses of variance (ANOVA) were generated using the Design Expert Software (version 6.0.6) from Stat-Ease Inc. (Minneapolis, MN, USA).

Experimental design for artificial neural network: The same experimental data, which had been used for the response surface methodology design, were also employed in designing

TABLE-2
CENTRAL COMPOSITE DESIGN MATRIX OF FOUR
VARIABLES AND THE EXPERIMENTALLY DETERMINED,
RSM MODEL PREDICTED AND ANN MODEL PREDICTED
VALUES OF ISOLATED YIELD OF ESTER

No.	X ₁	X ₂	X ₃	X ₄	Isolated yield of ester (%)		
					Actual	RSM predicted	ANN predicted
1	20	55	100	0.4	47.4	47.10	47.39
2	12	45	200	0.4	27.6	26.58	27.57
3	20	55	200	0.8	60.4	60.75	60.12
4	12	45	100	0.8	35.6	35.11	35.58
5	20	45	100	0.8	49.1	49.56	49.11
6	12	55	200	0.8	52.5	51.75	52.44
7	20	45	100	0.4	36.5	36.20	36.49
8	24	50	150	0.6	58.8	59.21	58.85
9	16	60	150	0.6	53.3	53.46	53.30
10	16	50	250	0.6	43.1	44.34	43.15
11	16	50	150	0.6	54.7	54.55	54.57
12	16	50	150	0.6	53.8	54.55	54.57
13	16	50	150	0.6	54.9	54.55	54.57
14	20	55	200	0.4	46.4	45.83	46.69
15	16	50	150	0.6	55.5	54.55	54.57
16	20	45	200	0.8	58.6	56.55	58.54
17	16	50	50	0.6	39.8	39.58	39.78
18	16	50	150	1.0	58.9	59.78	58.94
19	8	50	150	0.6	33.3	33.91	33.28
20	16	40	150	0.6	31.1	31.96	31.11
21	12	45	200	0.8	40.8	41.15	40.81
22	16	50	150	0.6	53.9	54.55	54.57
23	12	45	100	0.4	20.2	19.90	20.24
24	12	55	100	0.8	55.2	53.86	55.22
25	16	50	150	0.6	54.5	54.55	54.57
26	12	55	100	0.4	36.2	37.20	35.32
27	12	55	200	0.4	35.4	34.98	36.20
28	20	45	200	0.4	43.2	43.83	43.15
29	20	55	100	0.8	62.7	62.66	62.46
30	16	50	150	0.2	29.5	29.64	29.51

X₁= time (h); X₂= temperature (°C); X₃= enzyme amount (mg); X₄= substrate molar ratio (betulinic acid/ anhydride)

the artificial neural network. A multi-layer perceptron (MLP) based feed-forward artificial neural network, which makes use of the back-propagation learning algorithm, was applied for modeling the enzymatic reaction. The network consists of an input layer, one hidden layer and an output layer. The inputs for the network include reaction time, reaction temperature, enzyme amount and substrate molar ratio; output is the percentage of the isolated yield of ester. The performance of the artificial neural networks was statistically measured by the root mean squared error, the coefficient of determination (R²) and the absolute average deviation (AAD) obtained as follows:

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^n (y_i - y_{di})^2 \right)^{\frac{1}{2}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_{di})^2}{\sum_{i=1}^n (y_{di} - y_m)^2} \quad (4)$$

$$ADD = \left\{ \left[\sum_{i=1}^n (|y_i - y_{di}| / y_{di}) \right] / n \right\} \times 100 \quad (5)$$

where, n is the number of points, y_i is the predicted value, y_{di} is the actual value and y_m is the average of the actual values. A commercial artificial neural network software, known as NeuralPower version 2.5 was applied throughout the present study²⁰.

Model validation: The generalization ability of artificial neural network and response surface methodology can be determined using the unseen data which are not used for modeling¹³. Therefore, six additional experiments were performed in the range of values given in Table-1. Table-3 illustrates the actual values that were used to validate the models.

TABLE-3
ACTUAL AND PREDICTED VALUES FOR UNSEEN DATA

No.	X ₁	X ₂	X ₃	X ₄	Isolated yield of ester (%)		
					Actual	RSM predicted	ANN predicted
31	24	45	176	1.0	57.5	57.35	58.11
32	24	50	176	1.0	60.5	62.43	60.23
33	24	55	176	1.0	61.8	61.58	61.34
34	24	60	176	1.0	57.3	54.82	58.12
35	20	53	148	0.8	64.3	65.01	64.91
36	20	54	145	0.9	64.7	65.77	65.65

X₁= time (h); X₂= temperature (°C); X₃= enzyme amount (mg); X₄= substrate molar ratio (betulinic acid/ phthalic anhydride)

RESULTS AND DISCUSSION

Response surface methodology modeling: Response surface methodology was employed to model the enzymatic reaction between betulinic acid and phthalic anhydride, with 4 reaction parameters, namely reaction time, temperature, amount of enzyme and substrate molar ratio. Details of the response surface methodology modeling have been published elsewhere²¹. Fitting of the data to various models (linear, two factorial, quadratic and cubic) and their subsequent ANOVA showed that the reaction of betulinic acid and phthalic anhydride could most suitably be described using the quadratic polynomial model, as follows:

$$Y = 54.55 + 6.32 X_1 + 5.38 X_2 + 1.19 X_3 + 7.53 X_4 - 2X_1^2 - 2.96 X_2^2 - 3.15 X_3^2 - 2.46 X_4^2 - 1.6 X_1 X_2 + 0.24 X_1 X_3 - 0.46 X_1 X_4 - 2.22 X_2 X_3 + 0.55 X_2 X_4 - 0.16 X_3 X_4 \quad (6)$$

where Y is the percentage of the isolated yield of ester, X₁ is the reaction time, X₂ is the temperature, X₃ represents the amount of enzyme and X₄ is the substrate molar ratio.

Artificial neural network modeling: Four different algorithms, belonging to two different classes, namely gradient descent (in three versions; incremental back propagation, batch back propagation and quick propagation) and Levenberg-Marquardt were used to train the neural networks. Details of the artificial neural network modeling have been also published elsewhere²². The results showed that the quick propagation (QP) algorithm had a better performance relative to the incremental back propagation (IBP), batch back propagation (BBP) and Levenberg-Marquardt (LM) back propagation algorithms.

Comparison of response surface methodology and artificial neural network models: The estimation capabilities of the techniques, namely response surface methodology and

artificial neural network, were also examined in this study. For this purpose, the techniques were used to estimate the responses (*i.e.* the percentage of the isolated yield) at 36 experimental points (including CCRD matrix and the unseen data). The estimated responses, obtained from response surface methodology and artificial neural network, were then compared with the observed responses (*i.e.* the actual values), while the root mean squared error, coefficient of determination (R^2) and absolute average deviation were also determined. These values were used together to compare the response surface methodology and artificial neural network. The experimental and predicted isolated yields for the central composite rotatable design (CCRD) matrix are presented in Table-2.

The comparative values of root mean squared error, R^2 and absolute average deviation are given in Table-4. The root mean squared error for the central composite design matrix by response surface methodology and artificial neural network is 0.701 and 0.348, the coefficient of determination (R^2) is 0.996 and 0.999 and the absolute average deviation is 1.313 and 0.414. Fig. 2 shows the comparative parity plot for the two models predictions for the central composite design matrix. The error against observation order of both the models is compared in Fig.3. These results indicate that the response surface methodology model prediction has a greater deviation than the prediction made using the artificial neural network model. This also means that the experimental data had been fitted with a high accuracy using the artificial neural network model.

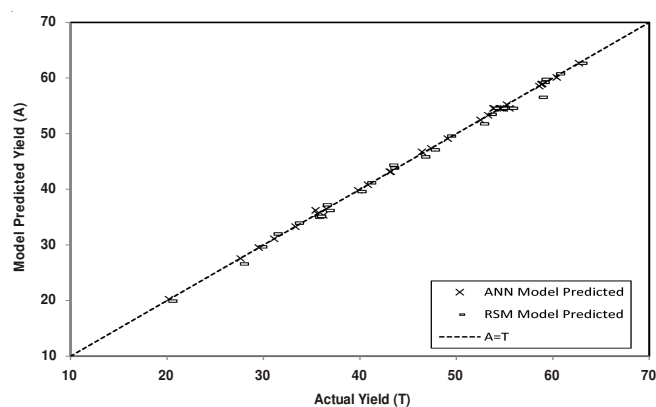


Fig. 2. Plot of RSM and ANN model predicted yield versus actual yield for central rotatable composite design (CCRD)

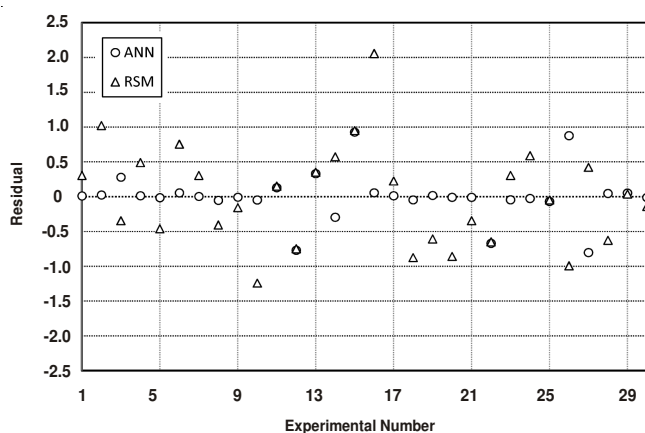


Fig. 3. Comparison of observation order with residuals for CCRD matrix

The generalization ability of artificial neural network and response surface methodology was determined using the unseen data which had not been used for modeling. The actual and predicted values, which were calculated using the response surface methodology and artificial neural network for the unseen data are presented in Table-3. The comparative values root mean squared error, R^2 and absolute average deviation for the unseen data are also given in Table-4. The root mean squared error for the unseen data by the response surface methodology and artificial neural network is 1.390 and 0.663, respectively; whereas the coefficient of determination (R^2) is 0.774 and 0.949, respectively; and the absolute average deviation is 1.816 and 1.021, respectively. Fig. 4 shows the comparative parity plot for the two models predictions for the unseen data. These results show that the artificial neural network model has higher generalization ability than the response surface methodology model. The higher predictive accuracy of the artificial neural network can be explained based on its universal ability to make approximation (regression) non-linearity of the system, whereas the response surface methodology is only limited to the second-order polynomial¹³.

TABLE-4
COMPARISON OF RSM AND ANN

Parameters	Design data ¹		Unseen data ²	
	RSM	ANN	RSM	ANN
RMSE	0.701	0.348	1.390	0.663
R^2	0.996	0.999	0.774	0.949
AAD (%)	1.313	0.414	1.816	1.021

¹Central composite rotatable design (CCRD) which used for modeling RSM and ANN; ²Unseen data or validation data which not used for modeling

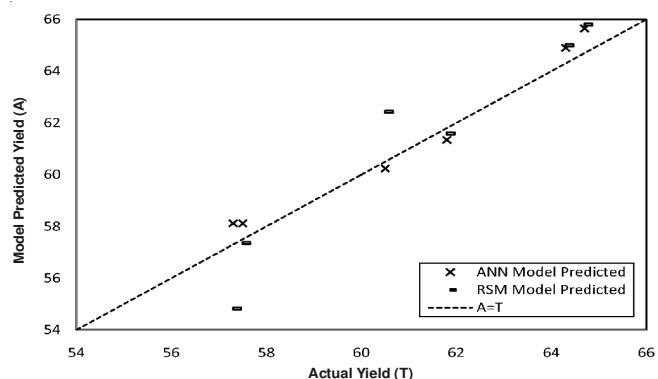


Fig. 4. Plot of RSM and ANN model predicted yield versus actual yield for unseen data

Conclusion

In the present work, two models (RSM and ANN) were used to predict the enzymatic synthesis of betulinic acid ester. The first step, *i.e.* response surface methodology, was applied for predicting the yield of ester in the enzymatic reaction. The next step was to train an artificial neural network model of the input-output data using the first step. The final step was to compare the two methodologies for their predictive and generalization capabilities. The present study indicates that the artificial neural network is much more robust and accurate to be used in estimating the values of yield of ester in comparison to the response surface methodology.

ACKNOWLEDGEMENTS

This research work was financially supported by a grant from Universiti Putra Malaysia (RUGS 9135500).

REFERENCES

1. M.M. O'Connel, M.D. Bently, C.S. Campbell and B.J.W. Cole, *Photochemistry*, **27**, 2175 (1988).
2. I. Habiyaremye, T. Stevanovic-Janezic, B. Riedl, F.-X. Garneau and F.-I. Jean, *J. Wood Chem. Technol.*, **22**, 83 (2001).
3. P. Yogeewari and D. Sriram, *Curr. Med. Chem.*, **12**, 657 (2005).
4. P.A. Krasutsky, *Nat. Prod. Rep.*, **23**, 919 (2006).
5. M. Kvasnica, J. Sarek, E. Klinotova, P. Dzubak and M. Hajduch, *Bioorg. Med. Chem.*, **13**, 3447 (2005).
6. D.C. Montgomery, *Design and Analysis of Experiments*, Wiley, New Jersey (2004).
7. W.-D. Chiang, S.-W. Chang and C.-J. Shieh, *Proc. Biochem.*, **38**, 1193 (2003).
8. H.S. Krishna, A.P. Sattur and N.G. Karanth, *Proc. Biochem.*, **37**, 9 (2001).
9. D. Bas and I.H. Boyaci, *J. Food Eng.*, **78**, 836 (2007).
10. D. Bas and I.H. Boyaci, *J. Food Eng.*, **78**, 846 (2007).
11. B. Manohar and S. Divakar, *Proc. Biochem.*, **40**, 3372 (2005).
12. V.K. Pareek, M.P. Brungs, A.A. Adesina and R. Sharma, *J. Photochem. Photobiol. A. Chem.*, **149**, 139 (2002).
13. K.M. Desai, S.A. Survase, P.S. Saudagar, S.S. Lele and R.S. Singhal, *Biochem. Eng. J.*, **41**, 266 (2008).
14. M. Basri, R.R. Zaliha, A. Ebrahimpour, A.B. Salleh, E.R. Gunawan and A. Rahman, *BMC Biotechnol.*, **7**, 53 (2007).
15. W. Lou and S. Nakai, *Food Res. Int.*, **34**, 573 (2001).
16. J. Bourquin, H. Schmidli, P.V. Hoogevest and A. Leuenberger, *Eur. J. Pharm. Sci.*, **7**, 5 (1998).
17. S. Agatonovic-Kustrin, M. Zecevic, L.G. Zivanovic and I.G. Tucker, *J. Pharm. Biomed. Anal.*, **17**, 69 (1998).
18. F.B.H. Ahmad, J. Omar and A.M. Ali, *Ultra Sci.*, **11**, 357 (1999).
19. W.G. Cochran and G.M. Cox, *Experimental Designs*, Wiley, New York (1992).
20. CPC-X Software. NeuralPower User Guide. Regsoft Inc. (2004). Available via DIALOG: <http://www.geocities.com/neuralpower> or <http://www.brothersoft.com/neuralpower-download-21356.html>
21. F.B.H. Ahmad, M.G. Moghaddam, M. Basri and M.B. Abdul Rahman, *Biocatal. Biotransform.*, **28**, 192 (2010).
22. M.G. Moghaddam, F.B.H. Ahmad, M. Basri and M.B.A. Rahman, *Elect. J. Biotechnol.*, **13**, 3 (2010).