

Prediction of Viscosities of Aqueous Two Phase Systems Containing Protein by Artificial Neural Network

Selvaraj Raja*, Varadavenkatesan Thivaharan, Vinayagam Ramesh and Vytla Ramachandra Murty

Department of Biotechnology, Manipal Institute of Technology, Manipal, Karnataka, India

Abstract

The viscosities of aqueous two phase system containing bovine serum albumin (BSA) were predicted by artificial neural network (ANN) as a function of concentration of poly-ethylene-glycol (PEG), concentration of BSA and temperature. A three layer feed forward neural network based on Levenberg-Marquardt (LM) algorithm which consisted of three input neurons, 10 hidden neurons and one output neuron (3:10:1) was developed. The performance parameters were calculated and compared with the conventional Grunberg-Nissan empirical model. The satisfactory values suggest that the proposed ANN model has the capability of predicting viscosity in a better way than the conventional empirical model.

Keywords: Aqueous two phase system; Artificial neural network; Levenberg-Marquardt (LM) algorithm; Grunberg-Nissan empirical model

Introduction

Aqueous two phase system (ATPS) is a separation method which is largely used to purify various biomolecules in a single step [1]. It is a liquid-liquid extraction method which can be prepared by mixing aqueous solution of water-soluble polymer and salt or two water-soluble polymers with water [2]. ATPS can also be formulated by mixing a water-soluble polymer and protein with water. One of such systems has been demonstrated by Johansson [3] for the partitioning of lactate dehydrogenase and some mitochondrial enzymes.

Viscosity is a transport property and the magnitude determines the mass transfer between phases. The various inter-related properties such as size, shape, structure, degree of polymerization and polymer-solvent interactions can be obtained from the viscosity data [4]. Studying and predicting the physical property, viscosity is very much essential for the design and optimization of an ATPS in large scale system. These data are helpful in designing the extraction process in large scale (since large volumes of phase components have to be handled and separated) and for the development of various mathematical models which can predict the partitioning behavior of biomolecules in aqueous two phase system [5].

To a large extent, viscosity of an ATPS depends on the concentration of phase component compositions and temperature. Substantial cost is involved in performing experiments in the entire range of phase components to determine the viscosity of solutions. In this context, mathematical equations should be available in predicting the physical properties. There are few polynomial equations and empirical equations are available in literature as a function of solute concentration for different temperatures [6]. However, empirical equations suffer because of their validity in a narrow range of dependent variables. Moreover the relative errors and deviations of these empirical models between experimental values and calculated values will be very high.

One of such empirical equations for viscosity is Grunberg-Nissan empirical model [7] which was successfully used by Gunduz [8] for an ATPS, consisting of poly-ethylene-glycol (PEG)- Bovine Serum Albumin (BSA)-Water system. The model (Eqn. 1) can be written as follows:

$$\ln(\eta_{sys}) = a_1 \ln(\eta_{PEG}) + a_2 \ln(\eta_{BSA}) + w_1 w_2 A \quad (1)$$

Where, A is an adjustable parameter which is the characteristic of the intermolecular interactions between PEG and BSA. a_1 and a_2 are weight fractions of PEG and BSA respectively and η_{PEG} and η_{BSA} are viscosities of aqueous PEG and BSA solutions respectively.

In recent years, an artificial intelligence technique, ANN has been widely used to model and predict linear and non-linear systems. It is a simplified artificial model of the biological neuron system which can provide the relationship between the input and output variables from the given data set. The artificial neuron simulates the basic functions of biological neurons [9].

The major advantage of ANN is its high learning capacity and predictive ability from the limited information fed to the system. Moreover, in contrast to conventional empirical model, ANN can accommodate multiple input and output variables and it has been proven by researchers that a well-trained ANN can predict various processes in a better way [10-13].

Recently, ANN has been used to predict viscosity of PEG solutions [14], density of ionic liquids [15] and vapor liquid equilibrium data of ionic liquids [16].

To the best of our knowledge, there are no reports available for predicting viscosities of aqueous two phase systems containing protein using artificial neural network. Therefore the objective of the current study is to model the experimental viscosity of the aforementioned ATPS by using ANN and compare with the Grunberg-Nissan empirical model.

Methodology

Collection of viscosity data from literature

For the current study, the dynamic viscosity data set was chosen

*Corresponding author: Selvaraj Raja, Department of Biotechnology, Manipal Institute of Technology, Manipal, Karnataka-576104, India, Tel: +9108202924322; Fax: +9108202571071; E-mail: rajaselvaraj@gmail.com

Received March 21, 2014; Accepted April 26, 2014; Published April 30, 2014

Citation: Raja S, Thivaharan V, Ramesh V, Murty VR (2014) Prediction of Viscosities of Aqueous Two Phase Systems Containing Protein by Artificial Neural Network. J Chem Eng Process Technol 5: 192. doi: [10.4172/2157-7048.1000192](https://doi.org/10.4172/2157-7048.1000192)

Copyright: © 2014 Raja S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

from the literature [8] which included the viscosity for the given concentration of PEG (4-11 % w/w), concentration of BSA (1-8 % w/w) and the process temperature (15-35 °C). A total of 35 experimental data were taken for the present study.

Neural network topology

Artificial Neural Network modeling codes were developed by using Matlab R2013a software. Normally, multilayers are used in ANN models and the simplest form of them is the three-layer models which contain input layer (I), one hidden layer (H) and output layer (O). Each layer consists of many nodes which are connected to each other by two coefficients known as weights (w) and biases (b). The commonly used feed forward multilayer neural network (FFMNN) was chosen for the present study. Choosing the number of neurons in the hidden layer and the transfer functions for hidden and output layer is crucial in developing an ANN model. For the present system, the number of neurons in hidden layer was chosen as 10 which was optimum.

Tangent and linear transfer functions were chosen as hidden and output layers respectively. The output of neuron is calculated by

$$O_j = F(I_j) \quad (2)$$

Where I_j , O_j and F are input of i^{th} neuron, output of i^{th} neuron and transfer function respectively.

In the present study, hyperbolic tangent transfer function and linear transfer functions have been employed for neurons located in hidden layer and output layers respectively. The input of each neuron (I_j) is calculated by the following equation with respect to outputs of previous layers (O_j), weights connecting i^{th} neuron to the j^{th} neuron (w_{ij}) and bias of the j^{th} neuron (b_j)

$$I_j = \sum_i w_{ij} y_i + b_j \quad (3)$$

In order to optimize the weights and biases, Levenberg-Marquardt (LM) back propagation algorithm has been used, which is one of the most popular tools [17].

The data set was divided into three sets, namely, training set (70%), validation set (15%) and testing set (15%). Training set is used to train the network and the network is adjusted according to its error. Validation set is used to measure the network generalization and to stop the training process when generalization stops improving. The testing set is used to evaluate the performance of trained model against new "unseen" data and hence gives an independent measure of network performance during and after training.

For the ANN model developed, the prediction accuracy and degree of fitness were evaluated by using the following factors namely, Mean Square Error (MSE), Root Mean Square Error (RMSE), Standard Error of Prediction (SEP), Average Absolute Relative Deviation (AARD) with Bias factor (B_p) and Accuracy factor (A_p) [18].

$$MSE = \frac{\sum_{i=1}^n [\mu_{i,exp} - \mu_{i,pred}]^2}{n} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\mu_{i,exp} - \mu_{i,pred}]^2}{n}} \quad (5)$$

$$SEP(\%) = \frac{RMSE}{\bar{\mu}_{i,exp}} \times 100 \quad (6)$$

$$AARD(\%) = 100 \times \frac{\sum_{i=1}^n \left| \frac{\mu_{i,exp} - \mu_{i,pred}}{\mu_{i,exp}} \right|}{n} \quad (7)$$

$$B_p = 10^{\left[\frac{\sum_{i=1}^n \log(\mu_{i,pred} / \mu_{i,exp})}{n} \right]} \quad (8)$$

$$A_p = 10^{\left[\frac{\sum_{i=1}^n \log(\mu_{i,pred} / \mu_{i,exp})}{n} \right]} \quad (9)$$

where $\mu_{i,exp}$ and $\mu_{i,pred}$ are the experimental and predicted viscosities and 'n' is the number of experiments.

Results and Discussions

Neural network modeling of viscosity

Coding for the proposed ANN model was written by using Matlab R2013a software with the aim to minimize mean squared error (MSE) and maximize the correlation coefficient value (R^2). A low value of MSE and high value of R^2 is preferred for the best model.

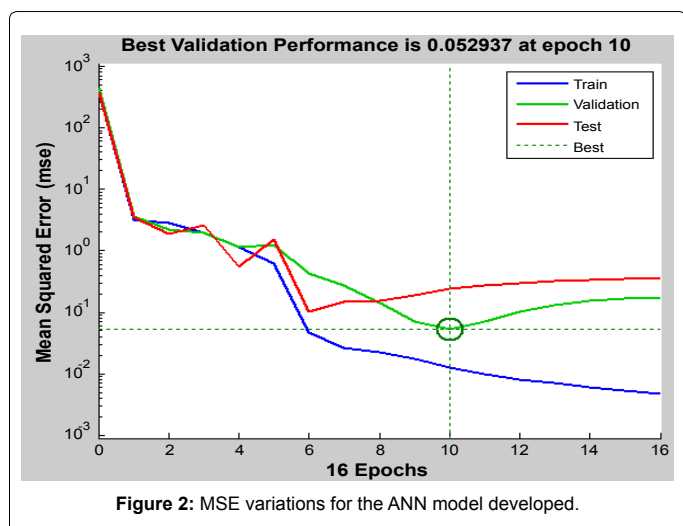
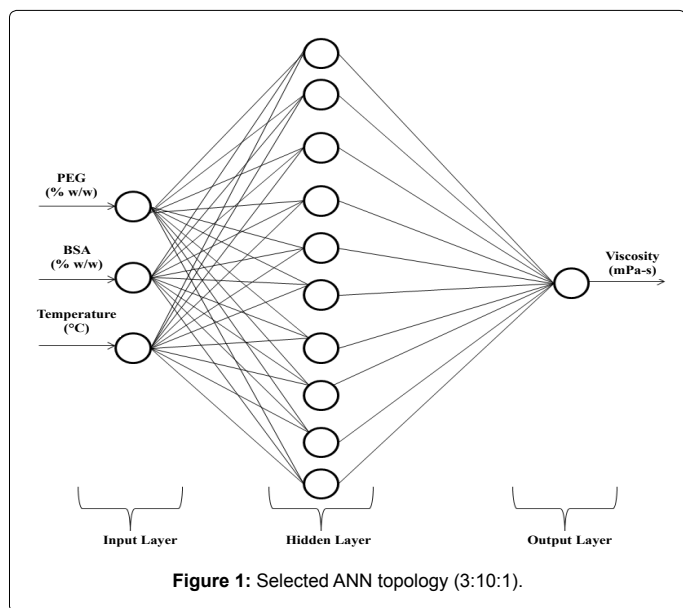
Various numbers of hidden layers (5, 10 and 15) were checked for the performance of the neural network. The optimum number of hidden layers was found to be 10 which are depicted in the Table 1. Lowest value of MSE (0.0511) and highest value of R^2 (0.9979) was obtained for 10 number of hidden layers and therefore considered as optimum.

As discussed earlier, the experimental points from the literature [8] were taken and trained by ANN topology (3:10:1). The selected network consists of three input neurons namely PEG (% w/w), BSA (% w/w) and Temperature (°C), 10 hidden neurons and one output neuron (Viscosity). The connection between the various layers for the developed ANN is shown in Figure 1.

"trainlm" training function was used in this work which updates weight and bias values according to Levenberg-Marquardt (LM) optimization. It is considered to be one of the fastest back propagation algorithms available. The training set data was trained by adjusting the strength and connections between neurons with an objective to fit the output of the entire network to be closer to the desired target and to minimize the performance function. The performance function used was MSE which measures the performance of the network according to the mean squared errors. The training process was halted once the goal was met (i.e., reaching a smallest value of MSE). In the present study,

Data Set	No. of hidden layers=5		No. of hidden layers=10		No. of hidden layers=15	
	MSE	R ²	MSE	R ²	MSE	R ²
Training	0.307	0.9857	0.013	0.9995	0.016	0.9802
Validation	0.149	0.9873	0.053	0.9965	0.243	0.6454
Testing	0.785	0.9251	0.24	0.997	0.202	0.8056
Overall	0.353	0.9839	0.051	0.9979	0.075	0.9481

Table 1: Optimum number of hidden layers based on MSE and R2 values.



the training was stopped after 16 iterations (epochs). Figure 2 shows the trend of MSE during the process of training, testing and validation. It can be noticed from the figure that MSE value reaches a minimum value of 0.0129 for training. Literature suggests that lower the value of MSE, higher the accuracy of the model. The best value of MSE (0.052937) was obtained after 10 epochs which was depicted (dotted lines) in the figure.

The optimal values of weights and biases of the ANN topology (3:10:1) based on LM algorithm is shown in Table 2.

In the Table 2, w_1 and w_2 are the input and hidden layer weight matrix, b_1 and b_2 are the biases of input and output layer respectively. The viscosity output values were predicted by using Eqn. 2 (data not shown) and the accuracy parameters were calculated.

The results obtained during training, validation and testing are shown in the Figure 3. The solid 45° line on the predicted versus experimental viscosities reveals that there is a perfect superpose of predicted and experimental data on each other. This confirms that the proposed model is valid. For all the data set, the value of R^2 was in acceptable range of greater than 0.99 (training: 0.9945, validation:

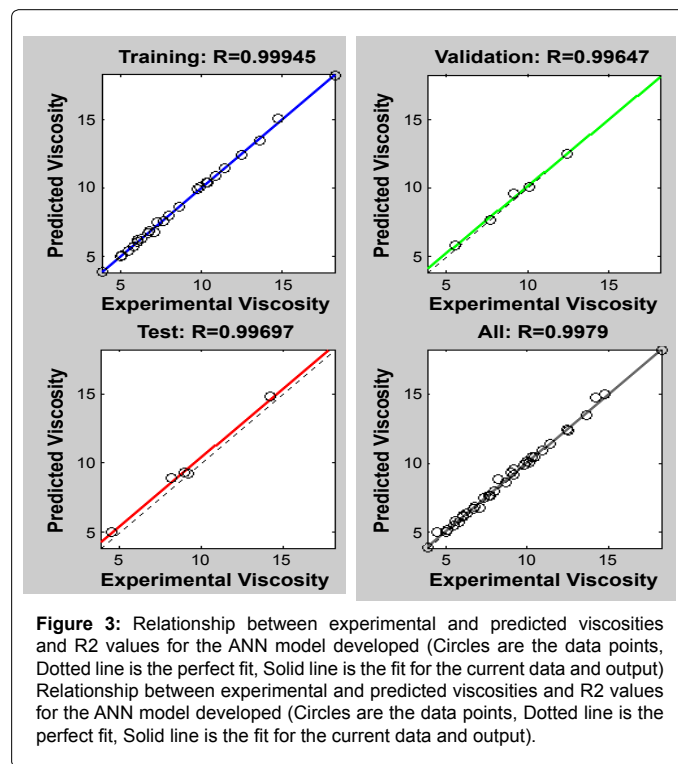
0.99647 Testing: 0.99697 and overall: 0.99790) which confirms a good agreement with the experimental data.

In addition to this, the prediction accuracy and degree of fitness were calculated (Equations 4 to 9) and shown in the Table 3.

The accuracy parameters of the ANN model are lesser than the Grunberg-Nissan model which suggests the best fitting. The B_f factor (1.01) and A_f factor (1.02) are closer to unity for the ANN model, which indicates that the model is “fail-safe” and shows a good concordance between the predicted and experimental values [19,20]. Because of these higher fitness and accuracy, the ANN models can be used to predict and model any linear or non-linear system in a better way than compared to polynomial or empirical equations.

Training Function	Weights and Biases of the neural networks					
	w_1		w_2		b_1	b_2
Trainm	-0.9447	-0.9069	1.9125	-2.2014	4.3511	1.996
	-2.7955	-1.191	0.3436	-1.8846	2.5251	
	-2.7081	2.9248	2.3544	-0.0795	0.1959	
	3.3501	-0.7734	-0.1839	0.1821	-0.2177	
	-0.3902	0.3684	-2.4848	0.1775	-1.2122	
	2.0133	2.4695	-0.3276	0.1491	-0.3456	
	-1.6439	2.0489	-2.1602	-0.2154	-3.6291	
	1.1543	0.1692	-1.0892	0.4619	2.4491	
	1.992	1.6889	-0.0257	0.9454	3.4215	
	0.361	-2.2842	1.743	-0.2438	-3.343	

Table 2: Weights and biases of the proposed ANN model.



Parameter	MSE	RMSE	SEP (%)	B_f	A_f	AARD (%)
Eqn. 1 [6]	0.32418	0.56937	6.46607	1.00596	1.06587	6.39
ANN	0.05106	0.22596	2.56611	1.01061	1.01689	1.71261

Table 3: Accuracy parameters for the viscosities from the literature [6] (Eqn.1) and present ANN model.

Temperature, °C	AARD (%)	
	Eqn. (1) [8]	ANN model
15	5.15	0.84
20	5.97	1.97
25	6.23	0.83
30	7.88	4.32
35	6.74	0.6
Overall AARD (%)	6.39	1.71

Table 4: AARD% values for the viscosities from the literature [8] (Eqn.1) and present ANN model.

PEG %	BSA %	Temperature °C	Value of parameter, A	Viscosity, mPa S			Absolute Relative Error, %	
				Experimental [8]	Eqn. 1 [8]	ANN model	Eqn. 1 [8]	ANN model
11	1	15	3.97	18.29	17.78	18.22	2.77	0.36
11	1	20	3.82	14.25	15.06	14.79	5.69	3.82
11	1	25	3.6	12.51	12.74	12.41	1.87	0.8
11	1	30	3.61	9.8	10.16	9.92	3.65	1.21
11	1	35	3.57	7.99	8.44	7.99	5.66	0.01

Table 5: Absolute relative error (%) values for the viscosities from the literature [8] (Eqn.1) and present ANN model.

Table 4 shows the comparison of AARD values of the literature [8] and the proposed ANN model. It is evident from the table that $AARD_{ANN} > AARD_{Eqn}$ for all the temperatures studied. Moreover, the overall AARD based on Grunberg-Nissan model and ANN model was 6.39% and 1.71% respectively which confirms that ANN model has lesser error for viscosity prediction.

In addition to these, the applicability of the ANN model for the generalized condition can be demonstrated by analyzing the absolute relative error (%) values for the proposed ANN model and Grunberg-Nissan model. A sample data values were shown in the Table 5. The interesting thing to be noted here is that the adjustable parameter A was not taken into consideration during the ANN topology. Even then, the absolute relative error (%) values are very low when compared to Grunberg-Nissan model. This significantly corroborates the validity of the ANN model.

Thence, the acceptable values of these parameters confirm that the developed ANN model defines the actual behavior of the system and the diversity of the model developed [18]. Therefore, the ANN model developed in the present study can be employed to predict the viscosity of PEG-BSA-Water ATPS with varying compositions and different temperatures.

Conclusions

An ANN model for the prediction of viscosity of ATPS which consisted of PEG-BSA-water was developed based on LM algorithm. The topology of ANN model developed was 3:10:1. Various accuracy parameters were calculated and compared with the existing empirical equation (Grunberg-Nissan model). The overall AARD (%) of the developed model was 1.71 which assures the validity of the model. A very high R^2 value (>.99) and low MSE justifies the selected ANN model for the prediction of viscosities. Therefore, it is concluded that ANN can be used as a predictive tool to evaluate the viscosities of aqueous two phase systems containing protein.

Acknowledgment

The authors gratefully acknowledge the Department of Biotechnology, MIT, Manipal University for providing the facilities to carry out the research work.

References

- Albertsson PA (1986) Partitioning of Cell Particles and Macromolecules: separation and purification of biomolecules, cell organelles, membranes, and cells in aqueous polymer two-phase systems and their use in biochemical analysis and biotechnology, 3rd edn. John Wiley and Sons, New York.
- Hatti-Kaul R (2000) Aqueous Two Phase Systems: Methods and Protocols. Humana Press: Totowa, NJ.
- Johansson G (1996) Aqueous two-phase systems with a liquid protein (bovine serum albumin) phase for partitioning of enzymes. J Chromatogr B Biomed Appl 680: 123-130.
- Modarress H, Mohsen-Nia M, Mahdavi MA (2004) Experimental and theoretical studies of polymer/solvent viscosity mixtures. Journal of Applied polymer science 91: 1724-1729.
- Gündüz U (2000) Viscosity prediction of polyethylene glycol-dextran-water solutions used in aqueous two-phase systems. J Chromatogr B Biomed Sci Appl 743: 181-185.
- Kalaivani S, Srikanth CK, Regupathi I (2012) Densities and Viscosities of Binary and Ternary Mixtures and Aqueous Two-Phase System of Poly(ethylene glycol) 2000 + Diammonium Hydrogen Citrate + Water at Different Temperatures. Journal of Chemical engineering data 57: 2528-2534.
- GRUNBERG L, NISSAN AH (1949) Mixture law for viscosity. Nature 164: 799.
- Gündüz U (2004) Evaluation of viscosities of aqueous two-phase systems containing protein. J Chromatogr B Analyt Technol Biomed Life Sci 807: 157-161.
- Haykin S (1998) Neural Networks: A Comprehensive Foundation. Prentice Hall, NJ, USA.
- Hezave AZ, Lashkarbolooki M, Raeissi S (2012) Using artificial neural network to predict the ternary electrical conductivity of ionic liquid systems. Fluid Phase Equilibria 314: 128-133.
- Khaye M, Cojocar C (2012) Artificial neural network modeling and optimization of desalination by air gap membrane distillation. Separation and Purification Technology 86: 171-182.
- Sharon Mano Pappu J1, Vijayakumar GK, Ramamurthy V (2013) Artificial neural network model for predicting production of Spirulina platensis in outdoor culture. Bioresour Technol 130: 224-230.
- Kumar S, Pai PS, Rao BRS (2012) Radial Basis Function Network Based Prediction of Performance and Emission Characteristics in a Bio Diesel Engine Run on WCO Ester. Advances in Artificial Intelligence.
- Mohagheghian M, Ghaedi AM, Vafaei A (2011) Prediction of Kinematic Viscosity of Binary Mixture of Poly (Ethylene Glycol) in Water using Artificial Neural Networks. World Academy of Science, Engineering and Technology 49: 158-161.
- Lashkarbolooki M, Hezave AZ, Babapoor A (2013) Correlation of density for binary mixtures of methanol + ionic liquids using back propagation artificial neural network. Korean Journal of chemical Engineering 30: 213-220.
- Fazlali A, Koranian P, Beigzadeh R, Rahimi M (2013) Application of artificial neural network for vapor liquid equilibrium calculation of ternary system including ionic liquid: Water, ethanol and 1-butyl-3-methylimidazolium acetate. Korean Journal of chemical Engineering 30: 1681-1686.

17. Demuth H, Beale M (2002) Neural Network Toolbox for Use with Matlab®, User's Guide, Version 4. Mathworks Inc.
18. Khaouane L, Si-Moussa C, Hanini S, Benkortbi O (2012) Optimization of Culture Conditions for the Production of Pleuromutilin from Pleurotus Mutilus Using a Hybrid Method Based on Central Composite Design, Neural Network, and Particle Swarm Optimization. Biotechnology and Bioprocess Engineering 17: 1048-1054.
19. Lu ZM, Lei JY, Xu HY, Shi JS, Xu ZH (2011) Optimization of fermentation medium for triterpenoid production from *Antrrodia camphorata* ATCC 200183 using artificial intelligence-based techniques. Applied Microbiology and Biotechnology 92: 371-379.
20. Khoramnia A, Ebrahimipour A, Beh BK, Lai OM (2011) Production of a Solvent, Detergent, and Thermotolerant Lipase by a Newly Isolated *Acinetobacter* sp. in Submerged and Solid-State Fermentations. Journal of Biomedicine and Biotechnology 102.