Optimization of an Artificial Neural Network for Modeling Protein Solubility

Amith D. Naik and Sunil S. Bhagwat*

Chemical Engineering Division, Institute of Chemical Technology, University of Mumbai, Matunga, Mumbai-400019, India

Solubility models for four protein−salt systems have been developed with the aid of an artificial neural network technique. The solubility of proteins in salt solutions is a complex phenomenon dependent on the type of protein, pH, temperature, concentration, and type of salt. In these models, the solubility has been correlated as a function of temperature, pH, and salt concentrations. The four systems are carboxyhaemoglobin in potassium phosphate solutions, ovalbumin in ammonium sulfate solutions, glucose isomerase in ammonium sulfate solutions, and concanavalin A in ammonium sulfate solutions. The models predicted the solubilities with an average quadratic error ranging from 0.00025 to 0.002. The model predictions were then analyzed to study the effect of pH, temperature, and salt concentrations. The predictions were found to be qualitatively in agreement with reports from the literature.

Introduction

Downstream processing of proteins accounts for a major portion of the cost in the bioprocess industry. Research has been ongoing to find cost-effective separation techniques to lower production costs. Crystallization, which is used industrially for the recovery and purification of many inorganic and organic materials, can be used for the recovery and purification of proteins. The crystallization technique is scalable to any production requirement, provided that solution and kinetic data for the protein are known.[1] Solubility data is an important tool for process design and control in crystallization. Solubility is the first information needed for designing a crystallizing system.[2] It is also a key to understanding the protein crystal growth process and gives insight into a protein's behavior and function in liquid and crystalline states.[3]

Although they are important, solubility data are available for only a few proteins, and an accurate model for the solubility of a protein has not been defined. Efforts to model protein solubility date back more than a half century. The first well-known relation between protein solubility and salts was proposed by Cohn.[4] Melander and Horvath[5] related the hydrophobic effect to protein solubility. Przybycien and Bailey[6] showed that the empirical correlation proposed by Melander and Horvath[5] was valid for only conformally inert proteins. Three simple empirical equations were proposed by Jenkins,[7] in which protein solubility was fitted with salt concentration in terms of either the salt molarity, the salt activity, or the water activity. A UNIQUAC model with temperature-dependent parameters to model protein solubility was proposed by Agena et al.[3] The two systems applied to the model were lysozymes in sodium chloride solution and concanavalin A in ammonium sulfate solution. The models developed were limited to the systems and conditions studied. Moreover, the model did not include the pH variation term. A relation between the second virial coefficient of the solution and the solubility of proteins was suggested by Wilson and co-workers.[8] However, the model was analyzed only for lysozymes in sodium chloride solutions. A correlation between the osmotic second virial coefficient and the solubility of proteins from classical thermodynamics was developed by Ruppert et al.[9] The model was fit to two systems, lysozymes in sodium chloride solution and ovalbumin in ammonium sulfate solution, but agreement between the model and the experimental data was good only for protein solubilities up to 30 mg/mL.

Artificial neural networks have been successfully applied to a number of disciplines such as chemistry,[10,11] medicine,[12] molecular biology[13] and chemical engineering.[14−16] In biotechnology, ANN has been utilized for many applications such as the prediction of variables, optimization and modeling, and process control.[17−23] Neural networks provide a simple, straightforward approach to ill-defined biological problems owing to their ability to handle the highly nonlinear, complex, and dynamic problems in bioprocesses.

Protein solubility is a complex function of a number of factors such as the physical and chemical nature of proteins themselves and environmental parameters such as pH, temperature, and nature of the salt or organic solvent and its concentration.[24] In this study, an artificial neural network technique is used to develop a model for predicting the solubility of proteins. The systems examined here are

1. carboxyhaemoglobin in aqueous potassium phosphate solutions;
2. ovalbumin in aqueous ammonium sulfate solutions;
3. glucose isomerase in aqueous ammonium sulfate solutions; and
4. concanavalin A in aqueous ammonium sulfate solutions.

To the best of our knowledge, a thermodynamic-based model has not been developed for glucose isomerase−ammonium sulfate systems and carboxyhaemoglobin−potassium phosphate systems.

Experimental Data

Ovalbumin is a major protein in egg whites. It can be recovered and purified from solution by bulk crystallization

* Corresponding author. E-mail: ssb@udct.org. Fax: +91-22-24145614.

10.1021/je049713d CCC: $30.25
Published on Web 02/18/2005

using high concentrations of ammonium sulfate.[1] In the present study, solubility data of ovalbumin in ammonium sulfate solutions were taken from the published literature.[25] Ovalbumin solubility in aqueous ammonium sulfate solutions ranging from (23.7 to 30.1) kg of ammonium sulfate·(100 kg of water)$^{-1}$ at 273 K, 285 K, 291 K, 293 K, 297 K, and 302 K over the pH range from 3.92 to 5.3 was measured by Sorensen and Hoyrup[26] and Sorensen.[27] A majority of these measurements were made at 291 K. The solubility of ovalbumin in aqueous ammonium sulfate solutions ranging from (18.1 to 28) kg of ammonium sulfate·(100 kg of water)$^{-1}$ at 303 K over the pH range from 4.57 to 5.42 was measured by Judge et al.[28] The solubility data of glucose isomerase from *S. rubiginosis* was obtained from Dalziel.[29] Glucose isomerase solubility in aqueous ammonium sulfate solutions was from (2.5 to 25) kg of ammonium sulfate·(100 kg of solution)$^{-1}$, temperatures from 275 K to 325 K, and pH from 5.5 to 8.5. Solubility data of horse carboxyhaemoglobin was taken from Green.[30,31] It covered potassium phosphate concentrations from 0.005 K·mol·m^{-3} to 1.908 K·mol·m^{-3} at temperature of 298 K and 273 K. The effect of pH from 5.89 to 7.53 was also considered. Concanavalin A solubility data from Jack Bean in aqueous ammonium sulfate was taken from Mickol and Giege.[32] It covered ammonium sulfate concentrations from 0.4 K·mol·m^{-3} to 1.6 K·mol·m^{-3} and temperature from 277 K and 313 K. The effect of pH from 5 to 7 was also included.

Artificial Neural Network

An artificial neural network is an information processing paradigm that is inspired by the way that biological nervous systems, such as the brain, process information. They offer an attractive approach to black-box modeling of highly complex, nonlinear systems having a large number of inputs and outputs. They require relatively little time to construct and do not require any prior knowledge of the relationships between the process variables in question.[33] They consist of simple units called neurons or processing elements that are working in parallel and are connected via directed links. Among the variety of neural network architectures that have been proposed, the feedforward artificial neural network with an error back-propagation learning algorithm appears to be the most commonly used network paradigm for approximating the nonlinear functional relationships between input−output variables of complex systems.[34] The error back-propagation (EBP) structure is mainly composed of three layers: an input layer, an intermediate layer called a hidden layer, and an output layer. Each layer contains a number of neurons. Each neuron may be connected to all of the neurons in the next layer by connection weight. The problem of neural network modeling is to find a set of weights such that the error in prediction is minimal. The weights are randomly chosen initially and then adjusted according to an error minimization technique until the prediction error falls to an acceptable level. Thus, the network acquires knowledge through a learning process that involves the modification of connection weights in a systematic manner. This model can then be used to predict output that was not included in the training set. Thus, ANNs have the ability to generalize beyond the training data, which is one of the important advantages. Another advantage of ANNs is the inherent fault tolerance; the overall performance is not affected significantly even if a few data are abnormal because of experimental errors.

The input layer consists of n neurons that serve as distribution points. Each hidden layer of neurons computes the weighted sum of all of the inputs according to eq 1 and then transforms it using a nonlinear activation function[35]

$$\text{net} = \sum_{i=1}^{n} x_i w_i - \theta \tag{1}$$

where w_i ($i = 1, n$) represents the connection weights and θ is called the bias. The output y is evaluated according to

$$y = f(\text{net}) \tag{2}$$

where y is the final output and f is an activation function.

Activation Function. The activation function transforms processing elements in a linear or nonlinear manner. The nonlinear function approximation capability of multilayer feedforward networks is attributed to the use of nonlinear transfer functions for computing nodal outputs. In this study, the logistic function given by eq 3 was used as the activation function.

$$f(\text{net}) = \frac{1}{1 + e^{(-\text{net})}} \tag{3}$$

Learning Function: Standard Back-Propagation. The learning rule defines the steps needed to arrive at the right weight values so that the network learns individual patterns of the training data. The back-propagation algorithm, which is the most famous learning algorithm, was used in the present study.

The Stuttgart Neural Network Simulator (version 4.2)[36] was used for correlating the solubility.

Results and Discussion

Data Preprocessing and Analysis. In this study, the temperature, pH, and ammonium sulfate concentration were chosen as input parameters for the glucose isomerase−ammonium sulfate system and the concanavalin A−ammonium sulfate system. In the case of ovalbumin and carboxyhaemoglobin, the solubility decreases as the temperature increases at high salt concentrations.[27,30] The inverse temperature was hence used as an input parameter for these two systems. The solubility falls off exponentially as the salt concentration is increased at a given pH and temperature.[28,30] Also, the chemical potential of protein is a logarithmic function of protein concentration (solubility), and hence the logarithm of solubility was used as the dependent parameter (i.e., output). The inputs and outputs were then scaled to avoid numerical overflows during training. Because a sigmoidal form of the activation function was used, the input and output data for all of the systems was scaled between 0.05 and 0.95 and not between 0 and 1. This was done so that the network can extrapolate results with respect to the new inputs lying just outside the region of the training input vectors and also because to generate the normalized desired output of 0 and 1 using the logistic sigmoid function the output-layer weights would have to be very large, leading to an increase in training time.[35] The scaled data were then randomized.

Training of Neural Network. For neural network training, the data set for all of the systems was divided into a training set, a validation set, and a testing set.

- The training set is used to train a neural network. The error in this data set is minimized during training.
- The validation set is used to determine the performance of a neural network on patterns that are not trained during learning.

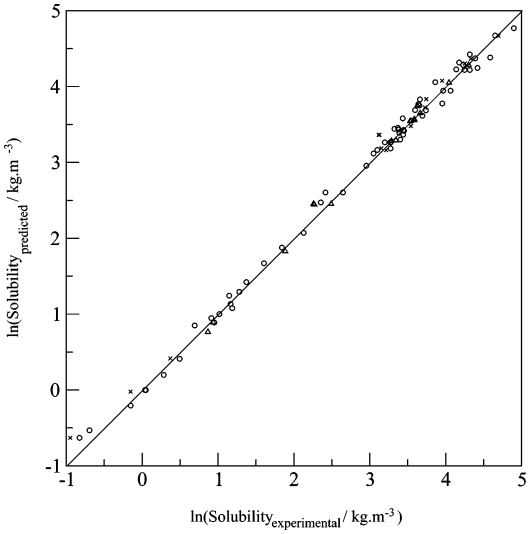

Figure 1. Parity plot of experimental and predicted ANN (3-4-1) solubility of carboxyhaemoglobin in potassium phosphate solutions of varying pH, temperature, and potassium phosphate concentration: ○, training; ×, testing; △, validation.

• A test set for finally checking the overall performance of a neural network.

The carboxyhaemoglobin−potassium phosphate system consisting of 89 patterns was divided into 60 patterns for training, 15 patterns for validation, and 14 patterns for testing. The ovalbumin−ammonium sulfate system consisting of 96 patterns was divided into 62 patterns for training, 20 patterns for validation, and 14 patterns for testing. The glucose isomerase−ammonium sulfate system consisting of 69 patterns was divided into 49 patterns for training, 10 patterns for validation, and 10 patterns for testing. The concanavalin A−ammonium sulfate system consisting of 161 patterns was divided into 110 patterns for training, 26 patterns for validation, and 25 patterns for testing.

To decide the optimum number of hidden-layer neurons, neural networks with 1−10 hidden-layer neurons were trained for all of the systems. The training of the network was continued until the validation error reached a minimum value. At this point, the net has the best generalization ability. Overtraining deteriorates the performance of the net, despite the fact that the error in the training data may decrease. The trained net was finally checked with the third set, the test set. The average quadratic error (AQE) for training, validation, testing, and the total was calculated as shown in eq 4 to quantify the network performance.

$$\mathrm{AQE} = \frac{(\sum_{P}^{i=1}(\mathrm{predicted} - \mathrm{experimental})^2)}{P} \qquad (4)$$

where P is the number of data points.

Carboxyhaemoglobin−Potassium Phosphate System. The optimum number of hidden-layer neurons for the carboxyhaemoglobin−potassium phosphate system was found to be four. The total AQE for this configuration was 0.000254, and the predicted AQE was 0.000454. The parity plot of the trained network is shown in Figure 1. To test the robustness of the trained network further, data sets that span a wide range of pH, temperature, and phosphate concentration were constructed, and the corresponding solubility (output) was predicted by the trained model. As shown in Figure 2 , solubility passes through a minimum, which is in agreement with theory. The point of minimum solubility is called the isoelectric point. Moreover, it can be seen from Figure 2 that the point of minimum solubility of carboxyhaemoglobin lies between pH 6.55 and pH 6.65, which is close to the isoelectric point reported in the literature (pI 6.6).[30] The effects of phosphate concentration and temperature on solubility were also examined. Figure 3 shows the "salting-in" and "salting-out" effects of salt. In the salting-in region, increasing temperature increases the solubility, whereas in the salting-out region, increasing temperature decreases solubility. This behavior is in agreement with theory.[30] The effect of temperature on solubility is shown in Figure 4. Thus, the trained neural network

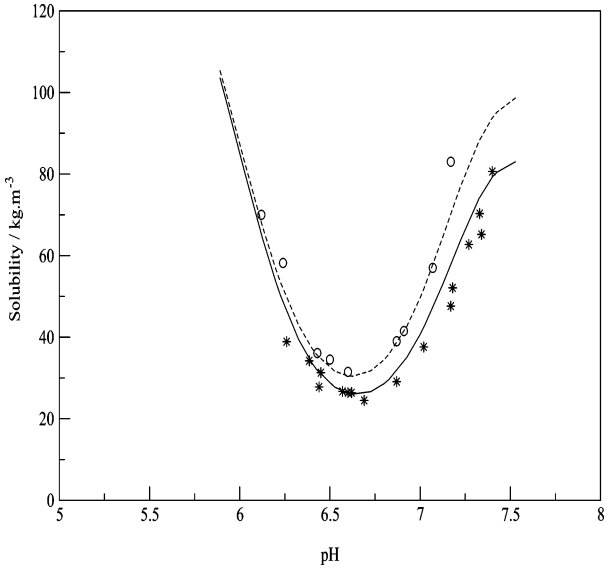

Figure 2. Solubilities of carboxyhaemoglobin estimated by neural network (3-4-1) at a fixed temperature 298 K and different potassium phosphate concentrations as a function of pH: *, phosphate concentration 0.02 K·mol·m^{-3}; ○, phosphate concentration 0.04 K·mol·m^{-3}. Lines represent ANN model calculations, and symbols represent experimental data.

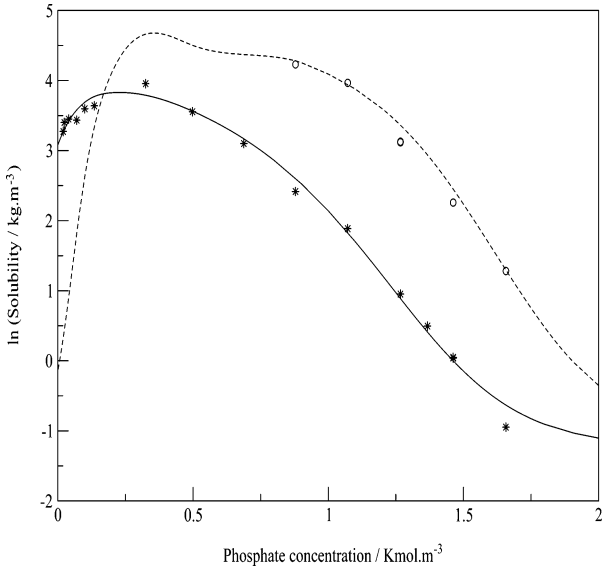

Figure 3. Solubilities of carboxyhaemoglobin estimated by neural network (3-4-1) at pH 6.6 and varying temperature as a function of potassium phosphate concentration: *, temperature 298 K; ○, temperature 273 K. Lines represent ANN model calculations, and symbols represent experimental data.

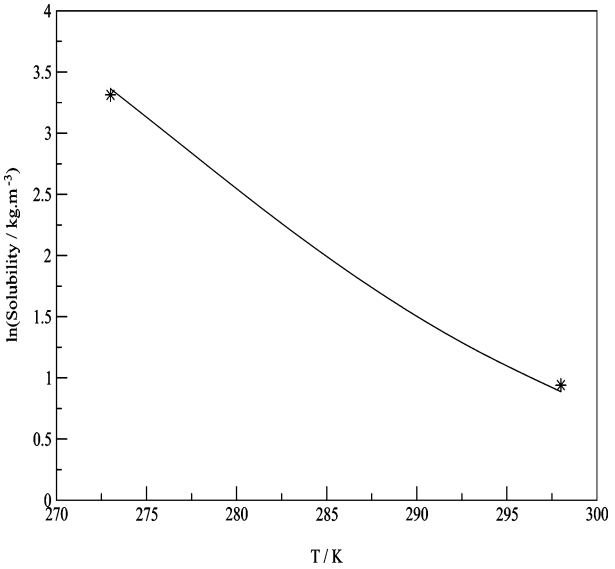

Figure 4. Solubilities of carboxyhaemoglobin estimated by neural network (3-4-1) at a fixed potassium phosphate concentration (1.267 K·mol·m^{-3}) and pH 6.6 as a function of temperature. The line represents ANN model calculations, and the symbols represent experimental data.

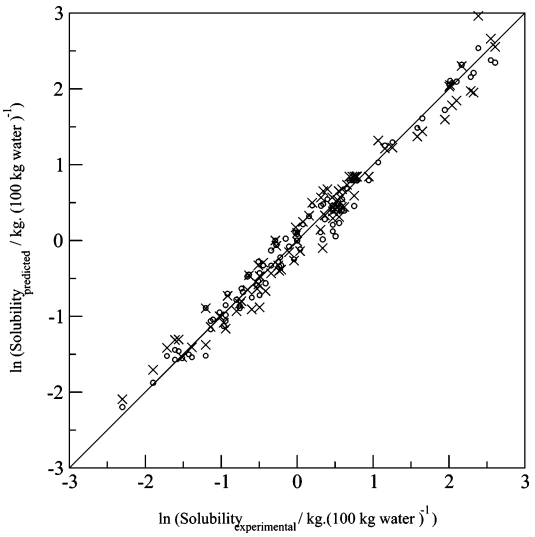

Figure 5. Parity plot of experimental and predicted ANN (3-3-1) solubility of ovalbumin in ammonium sulfate solutions of varying pH, temperature, and ammonium sulfate concentrations: ○, ANN model (3-3-1); ×, Judge et al. (1996).

model prediction of carboxyhaemoglobin solubility is qualitatively and quantitatively correct.

Ovalbumin−Ammonium Sulfate System. For the ovalbumin−ammonium sulfate system, the network with the three hidden-layer neurons predicted the solubility of ovalbumin with the lowest total AQE of 0.000831 and prediction AQE of 0.001103. The parity plot of the trained network is shown in Figure 5. Similar to the above system, the solubility was calculated for various input patterns using the trained model, and the effects of pH, ammonium sulfate concentration, and temperature were investigated. The effect of pH on solubility is shown in Figure 6. The solubility behavior is in agreement with theory. This model predicts the minimum solubility to be in the pH range of 4.2 to 4.6, which also closely follows the isoelectric point of ovalbumin (pI 4.58).[28] The effect of ammonium sulfate concentration and temperature on ovalbumin solubility was examined. As seen from Figure 7, the logarithm of the

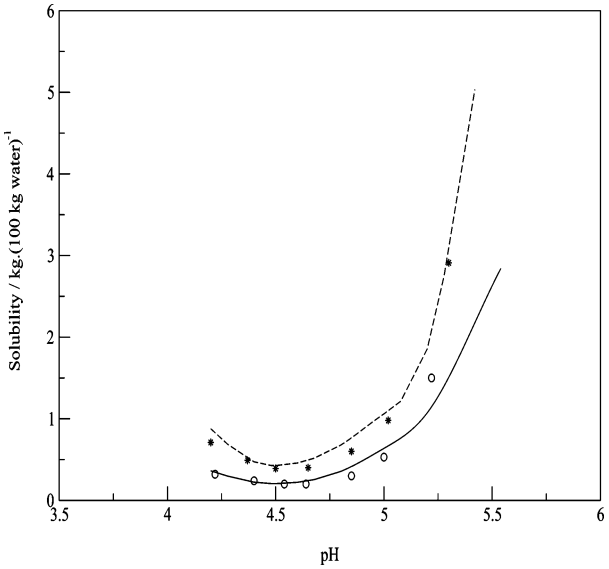

Figure 6. Solubilities of ovalbumin estimated by neural network (3-3-1) at fixed temperature of 291 K and varying ammonium sulfate concentrations as a function of pH: *, ammonium sulfate concentration 25.9 kg·(100 kg of water)$^{-1}$; ○, ammonium sulfate concentration 27.1 kg·(100 kg of water)$^{-1}$. Lines represent ANN model calculations, and symbols represent experimental data.

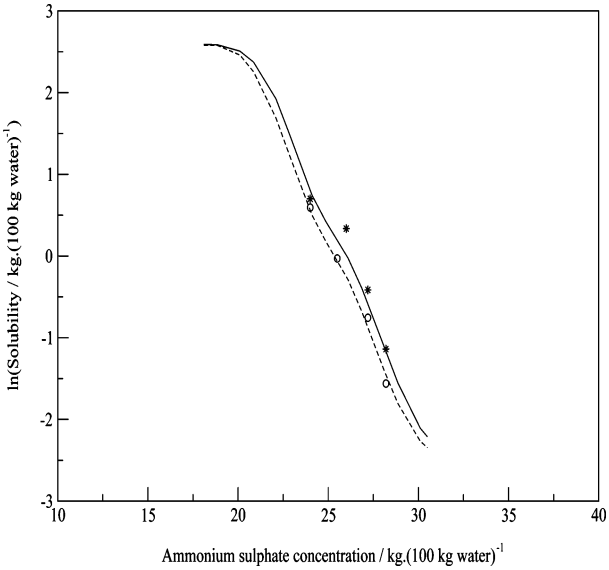

Figure 7. Solubilities of ovalbumin estimated by neural network (3-3-1) at pH 4.86 and different temperature as a function of ammonium sulfate concentration: *, temperature 273 K; ○, temperature 285 K. Lines represent ANN model calculations, and symbols represent experimental data.

solubility of ovalbumin decreases linearly with increasing ammonium sulfate concentration and decreases as temperature increases (Figure 8). These findings are also in agreement with theory.[28,30] The ANN model was then compared with the correlation developed by Judge et al.[28] The correlation is as given in the following expression

$$\log_{10}(C_0) = 5.06 - 0.006t - 0.205C_A + 0.5(\mathrm{pH} - 4.58) + 1.1(\mathrm{pH} - 4.58)^2 \quad (5)$$

where C_0 is the ovalbumin solubility concentration in (kg·(100 kg of water)$^{-1}$), t is the temperature in °C, and C_A is the ammonium sulfate concentration in (kg·(100 kg of water)$^{-1}$). This correlation is valid for ammonium sulfate concentrations of ($18 \leq C_A$/(kg·(100 kg of water)$^{-1}$) ≤ 30),

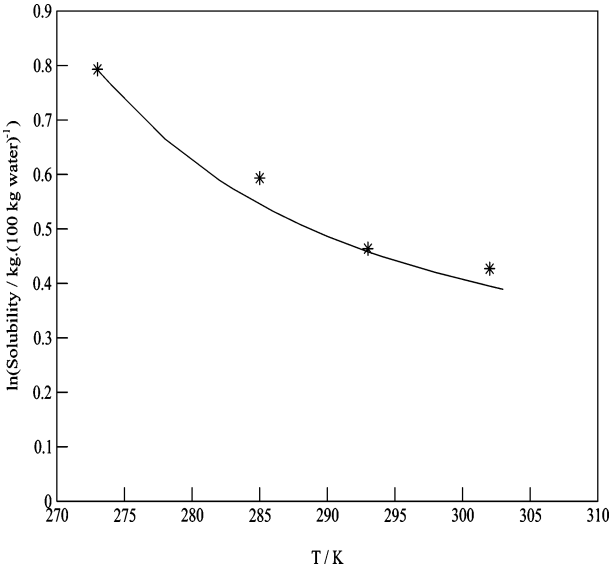

Figure 8. Solubilities of ovalbumin estimated by neural network (3-3-1) at a fixed ammonium sulfate concentration (24 kg·(100 kg of water)$^{-1}$) and pH 4.86 as a function of temperature. The line represents ANN model calculations, and the symbols represent experimental data.

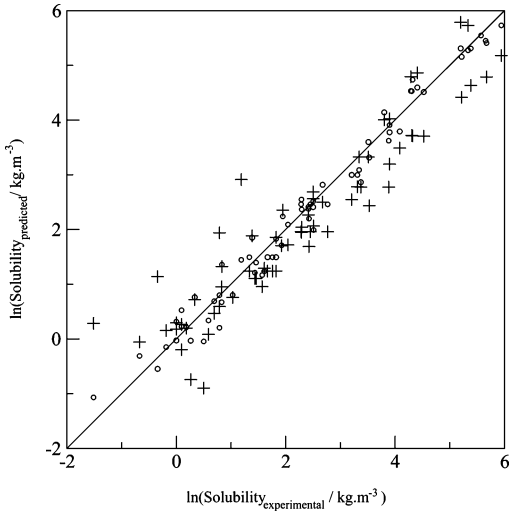

Figure 9. Parity plot of experimental and predicted ANN (3-5-1) solubility of glucose isomerase in ammonium sulfate solutions of varying pH, temperature, and ammonium sulfate concentrations: ○, ANN model (3-5-1); +, Dalziel (2000).

pH (4.58 ≤ pH ≤ 5.4), and temperature (0 ≤ *t*/°C ≤ 30). This correlation gave an AQE of 0.00158, which was comparable to the AQE of the ANN model. The parity plot for the correlation is given in Figure 5.

Glucose Isomerase−Ammonium Sulfate System. Neural network topology 3-5-1 was found to be optimum for the glucose isomerase−ammonium sulfate system. The total AQE of this model was 0.001026, and the prediction AQE was 0.000656. The parity plot of the trained network is shown in Figure 9. As shown in Figure 10, the trained model predicts the minimum solubility to lie between pH 6.5 and pH 7.5. This is reasonably correct with the isoelectric point of glucose isomerase (pI 7).[29] The effect of the ammonium sulfate concentration and temperature on solubility was also examined. From Figure 11, it can be seen that the solubility decreases exponentially with the increasing ammonium sulfate concentration and increases with increasing temperature (Figure 12), as suggested by Visuri.[29] The ANN model was then compared with the

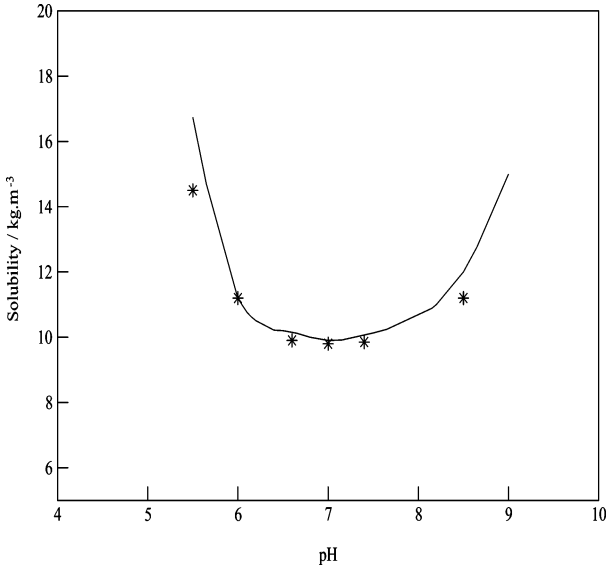

Figure 10. Solubilities of glucose isomerase estimated by neural network (3-5-1) at a fixed temperature 303 K and ammonium sulfate concentration (10 kg·(100 kg of solution)$^{-1}$) as a function of pH. The line represents ANN model calculations, and the symbols represent experimental data.

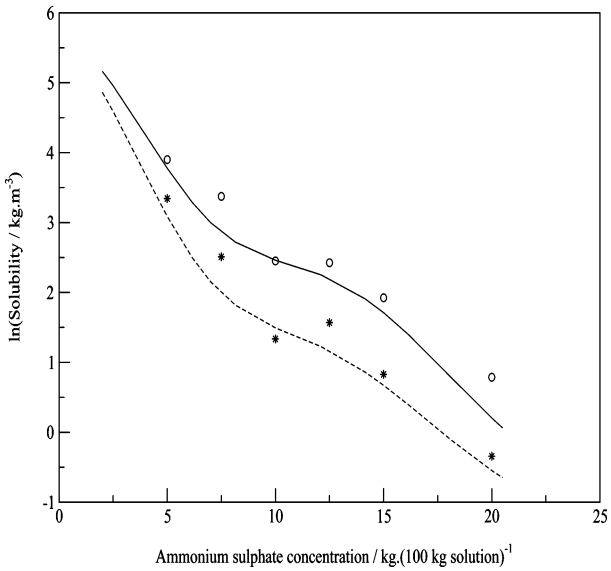

Figure 11. Solubilities of glucose isomerase estimated by neural network (3-5-1) at pH 7 and different temperatures as a function of ammonium sulfate concentration: *, temperature 298 K; ○, temperature 303 K. Lines represent ANN model calculations, and symbols represent experimental data.

correlation developed by Dalziel.[29] The correlation is as given in the following expression

$$C^* = 1.7(\mathrm{pH} - 7.25)^2 + [0.05 + 18e^{(-AS/1.6)}] \times e^{[T/(7.5 - 0.07AS)]} \quad (6)$$

where C^* is the glucose isomerase solubility in (kg·m^{-3}), T is the temperature in °C, and AS is the ammonium sulfate concentration in (kg·(100 kg of solution)$^{-1}$). This correlation gave an AQE of 0.00597, which was greater than the value from the ANN model. The parity plot for the correlation is given in Figure 9.

Concanavalin A−Ammonium Sulfate System. For this system, the optimum network consisted of seven hidden-layer neurons. The total AQE for this network

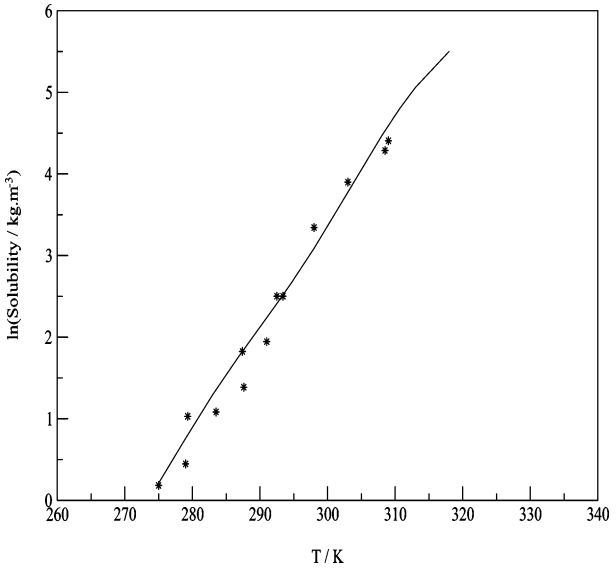

Figure 12. Solubilities of glucose isomerase estimated by neural network (3-5-1) at a fixed ammonium sulfate concentration (5 kg·(100 kg of solution)$^{-1}$) and pH 7 as a function of temperature. The line represents ANN model calculations, and the symbols represent experimental data.

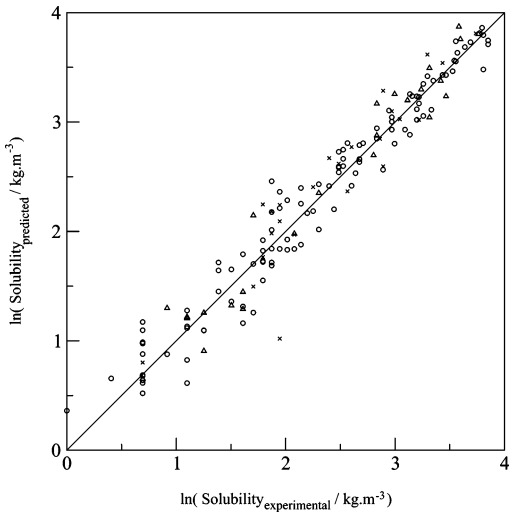

Figure 13. Parity plot of experimental and predicted ANN (3-7-1) solubility of concanavalin A in ammonium sulfate solutions of varying pH, temperature, and ammonium sulfate concentration: ○, training; ×, testing; △, validation.

topology was 0.002013, and the prediction AQE was 0.00307. The parity plot is shown in Figure 13. The effects of pH, temperature, and ammonium sulfate concentration were also analyzed. As shown in Figure 14, solubility decreases as pH increases as the isoelectric point of concanavalin A (pI 8.1[32]), which is in agreement with trends from theory. The effects of the ammonium sulfate concentration and temperature are shown in Figure 15 and Figure 16. The behavior is consistent with results from the literature.[32] The ANN model was then compared with the UNIQUAC model proposed by Agena et al.[3] The performance of the UNIQUAC model was evaluated by means of the root-mean-square deviation (rmsd). For the concanvalin A−ammonium sulfate system, the deviation of the experimental and modeled solubility was 4.5% rmsd. The ANN-based model gave similar results, but the ANN-based model was able to extend to include pH variation, with a deviation of 7% rmsd.

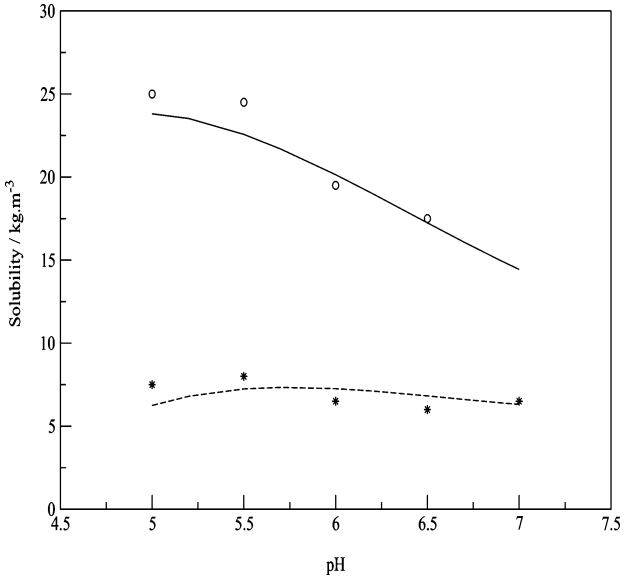

Figure 14. Solubilities of concanavalin A estimated by neural network (3-7-1) at a fixed temperature of 293 K and varying ammonium sulfate concentrations as a function of pH: *, ammonium sulfate concentration 1.2 K·mol·m^{-3}; ○, ammonium sulfate concentration 0.8 K·mol·m^{-3}. Lines represent ANN model calculations, and symbols represent experimental data.

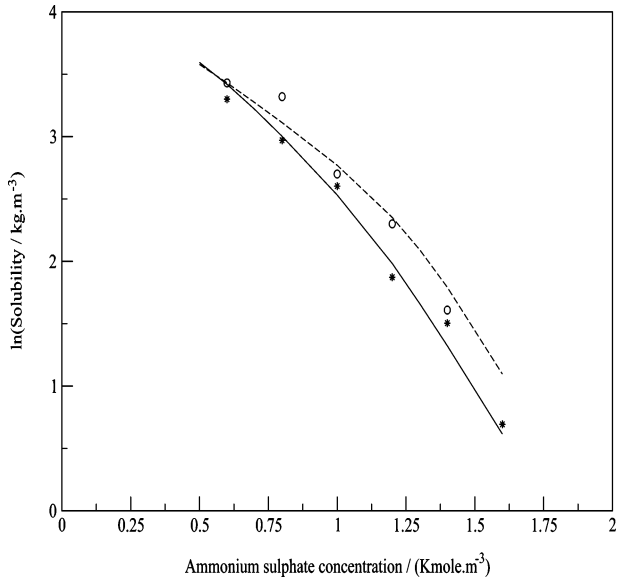

Figure 15. Solubilities of concanavalin A estimated by neural network (3-7-1) at pH 6 and varying temperature as a function of ammonium sulfate concentration: *, temperature 293 K; ○, temperature 303 K. Lines represent ANN model calculations, and symbols represent experimental data.

Conclusions

Protein solubility in salt solutions is a complex phenomenon dependent on many factors such as the type of protein, pH, temperature, and concentration and type of salt. As a result, a fundamental model has not been developed. Artificial neural networks, because of their ability to model nonlinear, multivariable systems and because they do not require a complete knowledge of the process to be modeled, offer a novel solution to the problem of protein solubility modeling.

In this work, solubility models for four different proteins—carboxyhaemoglobin, ovalbumin, glucose isomerase, and concanavalin A—have been developed with the artificial neural network technique. Solubility has been modeled as

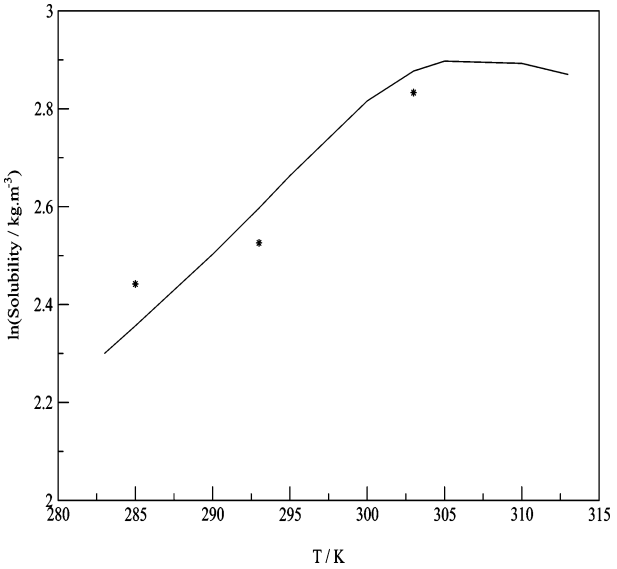

Figure 16. Solubilities of concanavalin A estimated by neural network (3-7-1) at fixed ammonium sulfate concentration (1 K·mol·m^{-3}) and pH 5.5 as a function of temperature. The line represents ANN model calculations, and the symbols represent experimental data.

function of pH, temperature, and salt concentration. The average quadratic errors (AQEs) for carboxyhaemoglobin, ovalbumin, glucose isomerase, and concanavalin A were found to be 0.00025, 0.00083, 0.001, and 0.002, respectively. A performance evaluation of these models was made by studying the effects of pH, temperature, and salt concentration. The carboxyhaemoglobin solubility model was able to predict the salting-in and salting-out behavior of salt because the network was trained for low and high salt concentrations. This shows that the ANN model can be enhanced by the availability of a wide range of data. The models were also able to predict the minimum solubility of the proteins near their isoelectric point. Thus, the models can be said to be qualitatively and quantitatively correct. Though the models can effectively predict the solubilities within the range of conditions for which the network was trained, extrapolation to predict solubilities cannot be guaranteed.

Supporting Information Available:

Procedure for developing the artificial neural network (ANN) technique for the carboxyhaemoglobin−potassium phosphate system and training of the neural network. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Judge, R. A.; Johns, M. R.; White, E. T. Protein purification by bulk crystallization: The recovery of ovalbumin. *Biotechnol. Bioeng.* **1995**, *48*, 316−323.
(2) Mulin, J. W. *Crystallization*, 2nd ed.; Butterworths: London, 1972.
(3) Agena, S. M.; Pusey, M. L.; Bogle, D. L. Protein solubility modeling. *Biotechnol. Bioeng*. **1999**, *64*, 144−150.
(4) Cohn, E. J.; Edsall, J. T. *Proteins, Amino Acids and Peptides*; Reinhold: New York, 1943.
(5) Melander, W.; Horvath, C. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of lyotropic series. *Arch. Biochem. Biophys*. **1977**, *183*, 200−215.
(6) Przybycien, T. M.; Bailey, J. E. Structure-function relationships in the inorganic salt induced precipitaion of a chymotrypsin. *Enzyme Microb. Technol*. **1989**, *11*, 264−276.
(7) Jenkins, W. T. Three solutions of the protein solubility problem. *Protein Sci.* **1998**, *7*, 376−382.
(8) Hass, C.; Drenth, J.; Wilson, W. W. Relations between Solubility of Proteins in Aqueous Solutions and the Second Virial Coefficient of the Solutions. *J. Phys. Chem. B* **1999**, *103*, 2808−2811.
(9) Ruppert, S.; Sandler, S. I.; Lenhoff, A. M. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnol. Prog*. **2001**, *42*, 376−380.
(10) Burns, J. A.; Whitesides, G. M. Feed-Forward Neural Networks in Chemistry: Mathematical Systems for Classification and Pattern Recognition. *Chem. Rev.* **1993**, *93*, 2583−2601.
(11) Cartwright, H. M. *Applications of Artificial Intelligence in Chemistry*; Oxford University Press: Oxford, U.K., 1993.
(12) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural Networks Applied to Structure−Activity Relationships. *J. Med. Chem*. **1990a**, *33*, 905−908.
(13) Quis, N.; Sejnowski, T. Predicting the secondary structure of globular proteins using neural networks models. *J. Mol. Biol*. **1998**, *202*, 865−884.
(14) Baratti, R.; Vacca, G.; Servida, A. Neural networks modeling of distillation column. *Hydrocarbon Process.* **1995**, *June*, 35−38.
(15) Thompson, M. L.; Kramer, M. A. Modeling chemical process using prior knowledge and neural networks. *AIChE J*. **1994**, *40*, 1328−1340.
(16) Prasad, J. W.; Bhagwat, S. S. Simple neural network models for prediction of physical properties of organic compounds. *Chem. Eng. Technol*. **2002**, *25*, 1041−1046.
(17) Thibault, J.; Van Breusegen, V.; Cheruy, A. On-line prediction of fermentation variables using neural networks. *Biotechnol. Bioeng*. **1990**, *36*, 1041−1048.
(18) Massimo, C.; Montague, G. A.; Willis, M. J.; Tham, M. T.; Morris, A. J. Towards improved penicillin fermentation via artificial neural networks. *Comput. Chem. Eng*. **1992**, *16*, 283−291.
(19) Ochoa, G. F.; Castro, E. G. Estimation of oxygen mass transfer coefficient in stirred tank reactors using tank reactors using artificial neural networks. *Enzyme Microb. Tech*. **2001**, *28*, 560−569.
(20) Yang, H.; Fang, B. S.; Reuss, M. k_La correlation established on the basis of a neural network model. *Can. J. Chem. Eng*. **1999**, *77*, 838−843.
(21) Thoulder, A.; Ramirez, W. F. Optimization of fed-batch bioreactors using neural network parameter function models. *Biotechnol. Prog*. **1996**, *12*, 302−309.
(22) Glassey, J. A.; Montague, G.; Ward, A. C.; Kara, B. V. Artificial neural networks based experimental design procedures for enhancing fermentation development. *Biotechnol. Bioeng*. **1994**, *44*, 397−405.
(23) Syu, M. J.; Tsaw, G. T. Neural network modeling of batch cell growth pattern. *Biotechnol. Bioeng*. **1993**, *42*, 376−380.
(24) Arakawa, T.; Timasheff, S. N. Theory of Protein solubility. *Methods Enzymol.* **1985**, *114*, 49.
(25) Judge, R. A. Investigating the Bulk Crystallization of Proteins. Ph.D. Thesis, Department of Chemical Engineering, The University of Queensland, Australia, 1995.
(26) Sorensen, S. P. L.; Hoyrup, M. Studies of proteins. 4. On the state of equilibrium between crystallized egg-albumin and surrounding mother liquor and on the applicability of Gibbs' phase rule to such systems. *Compt. Rend. Lab. Carlsberg* **1915−1917**, *12*, 213−261.
(27) Sorensen. S. P. L. The constitution of soluble proteins. (Reversibly dissociable component systems). *Compt. Rend. Lab. Carlsberg* **1930**, *18*, 1−124.
(28) Judge, R. A.; Johns, M. R.; White, E. T. Solubility of Ovalbumin in Ammonium Sulphate Solutions. *J. Chem. Eng. Data* **1996**, *41*, 422−424.
(29) Dalziel, S. M. Crystallization Studies on Glucose Isomerase. Ph.D. Thesis, The University of Queensland, Brisbane, Australia 4072, 2000.
(30) Green, A. A. Studies in the physical chemistry of the proteins VIII. The solubility of hemoglobin in concentrated salt solutions. A study of the salting out of proteins. *J. Biol. Chem*. **1931a**, *93*, 495−515.
(31) Green, A. A. Studies in the physical chemistry of the proteins IX. The effect of electrolytes on the solubility of haemoglobin in solutions of varying hydrogen ion activity with a note on the comparable behaviour of casein. *J. Biol. Chem*. **1931b**, *93*, 517−542.
(32) Mickol, V.; Geige, R. Phase diagram of a crystalline protein: Determination of the solubility of concanavalin A by a microquantitaion assay. *J. Cryst. Growth* **1989**, *97*, 324−332.
(33) Linko, S.; Zhu Y. H.; Linko P. Applying neural networks as software sensors for enzyme engineering. *Tibtech* **1999**, *17*, 155−162.
(34) Jha, B. K.; Tambe, S. S.; Kulkarni, B. D. Estimating diffusion coefficients of a micellar system using a artificial neural network. *J. Colloid Interface Sci*. **1995**, *170*, 392−398.
(35) Tambe, S. S.; Kulkarni, B. D.; Deshpande, P. B. *Elements of Artificial Neural Networks with Selected Applications in Chemical Engineering, Chemical and Biological Sciences*. Simulation and Advanced Controls Inc.: Louisville, KY, 1996.
(36) Zell, A.; Mamier, G.; Vogt, M.; Mache, N.; Sommer, T.; Hubner, R.; Schmalzl, M.; Soyez, T.; Doring, S.; Posselt, D.; Herrmann,

K.; Hazigeorgiou, A.; Kett, B.; Clemente, G.; Wieland, J.; Schreiner, T. *Stuttgart Neural Network Simulator* (SNNS V 4.2) *IPVR*; University of Stuttgart: Stuttgart, Germany, 1990−1995 and WSI, University of Tübingen: Tübingen, Germany, 1990−1998.

Received for review August 3, 2004. Accepted January 12, 2005.

JE049713D