

# Neural Modeling of Bromelain Extraction by Reversed Micelles

Ana Maria Frattini Fileti<sup>\*</sup>, Gilvan Anderson Fischer and Elias Basile Tambourgi

Faculdade de Engenharia Química; Universidade Estadual de Campinas; C. P.: 6066; 13083-970; Campinas - SP - Brasil

## ABSTRACT

*A pulsed-cap microcolumn was used for bromelain extraction from pineapple juice by reversed micelles. The cationic micellar solution used BDBAC as the surfactant, isooctane as the solvent and hexanol as the co-solvent. In order to capture the dynamic behavior and the nonlinearities of the column, the operating conditions were modified in accordance with the central composite design for the experiment, using the ratio between the light phase flow rate and the total flow rate, and the time interval between pulses. The effects on the purification factor and on total protein yield were modeled via neural networks. The best topology was defined as 16-9-2, and the input layer was a moving window of the independent variables. The neural model successfully predicted both the purification factor and the total protein yield from historical data. At the optimal operating point, a purification factor of 4.96 and a productivity of 1.29 mL/min were obtained.*

**Key words:** bromelain, reversed micelles, extraction, neural networks, pineapple

## INTRODUCTION

In many biotechnological industries, including the food and pharmaceutical sectors, the selective separation of proteins from the fermentation broths or vegetable sources is of prime interest for downstream processing operations. However, it can be difficult as well as expensive to recover a targeted protein from a broth due to low concentration of the protein and similarity between its physical properties and those of other proteins present in the same solution.

Liquid-liquid extraction involves the transferring a substance from one liquid mixture to another (either immiscible or partially miscible) by placing them in contact. This process is widely used in the chemical and pharmaceutical industries, such as in

the recovery of antibiotics or organic acids from fermentation broths. Nevertheless, its application is still limited for protein purification due to the risk of denaturation when proteins come into contact with organic solvents, thus yielding useless products (Aires-Barros et al., 1994).

In order to avoid the protein denaturation, liquid-liquid extraction by reversed micelles was used in the present study. Reversed micelles are aggregates of surfactant molecules containing an inner core of water molecules dispersed in a continuous organic solvent. They are capable of selectively solubilizing polar compounds in apolar solvents and are useful for recovering specific biomolecules from aqueous solutions, such as fermentation broths or cell culture media. Kilikian et al. (2000) suggested reversed micelle extraction

<sup>\*</sup>Author for correspondence: frattini@feq.unicamp.br

a versatile and useful process for purifying proteins.

Bromelain is the name of a group of powerful protein-digesting, or proteolytic, enzymes found in the pineapple plant (*Ananas comosus*). Discovered in 1957, and widely studied since then, bromelain is particularly useful for reducing the muscle and tissue inflammation, as well as for aiding the digestion. It is employed not only for its pharmacological effects, but also in food industry activities such as brewing and meat processing.

Although liquid-liquid extraction is usually modeled according to first principles, this is not trivial in the case of reversed micelles. On the one hand, calculating equilibria, hydrodynamics, mass transfers and material balances requires excessive amounts of time for computer simulations. On the other, artificial neural networks have the ability to “learn” the non-linear behavior and interactions of the complex system studied in the present research. Moreover, since prior knowledge of the process is not prerequisite to finding a suitable model, neural modeling was the empirical approach chosen for the purpose of this study.

Within the field of chemical processes, there is increasingly more literature describing the use of artificial neural networks for a diverse range of engineering applications, such as fault detection and signal processing, in addition to process modeling and control (Himmelblau, 2000). In a study by Pinto et al. (2005) on the partial enzyme hydrolysis of cheese whey proteins, an off-line smoothing algorithm, based on penalized least squares, was implemented in the monitoring system. For filtering on-line signals, various algorithms were compared: artificial neural networks, a moving average and a smoothing algorithm. In the on-line pH control system, the filters based on neural networks had smoother control action and no effect on the inference system.

The aim of the current study was to use an extremely simple approach to accurately represent a continuous bromelain extraction process. The operating conditions were modified according to the central composite design. Given a representative set of experimental observations concerning the inputs and their corresponding outputs, the neural network integrated these data into its structure. By means of a single neural model, two variables could easily be predicted: total protein yield and the purification factor. The

model developed should enable a human operator to adjust the operating variables – such as pulse frequency and the ratio between the light flow rate and the total flow rate – during experimental runs in order to achieve target outputs.

### Neural Networks

Artificial neural networks (ANNs) are mathematical models composed of various neurons arranged in interconnected layers (input, hidden and output). According to Equation (1), the neuron in one layer is responsible for the sum of the signals from the neurons in the previous layers,  $y_j$  (amplified or weakened by weighted values,  $w_{k,j}$ ) and a value called bias,  $b_k$ . A transfer function,  $f$  – such as a hyperbolic tangent, sigmoid or linear function – is used to activate the neuron output,  $y_k$ .

$$y_k = f \left( \sum_{j=1}^N (w_{k,j} y_j) + b_k \right) \quad (1)$$

A multi-layered feedforward network, the most suitable topology for empirical modeling and engineering applications, was used in this study.

The training procedure requires a set of the process inputs and outputs. During the procedure, the weights and biases are iteratively adjusted to minimize an objective function. The conventional training algorithm – backpropagation – moves the network parameters in the direction of the negative gradient (Demuth and Beale, 2002). Levenberg-Marquardt optimization is also a valuable training method. However, both methods can lead to data overfit if carelessly implemented.

One of the best methods for improving generalization and avoiding overfit is called regularization. The method involves modifying the objective function, which generally computes only the sum of the squared errors (SSE) of the training set. When the regularization method is used, a term consisting of the mean of the sum of the squared weights (SSW) is added to the SSE calculation in the objective function (Equation [2]).

$$F = \beta \cdot SSE + \alpha \cdot SSW \quad (2)$$

in which  $\beta$  and  $\alpha$  are fitted parameters (Demuth and Beale, 2002).

According to Hagan and Foresse (1997), when this objective function is coupled with a Levenberg-

Marquardt algorithm, the network will have smaller weights, which, in turn, will cause the network response to be smoother and less subject to overfit. Another important feature of this algorithm is that it provides a measure as to how many parameters (weights and biases) are effectively used by the network. This effective number of parameters is termed  $\gamma$  and differs from the total number of parameters in the network,  $N$ . When the effective number of parameters remains the same, the ideal number of neurons for the hidden layer has been attained.

## MATERIALS AND METHODS

### Bromelain sample preparation

Fruit bromelain (EC 3.4.22.5) was obtained from the fruit extract of the Perola pineapple species. The pineapple pulp was triturated. Distilled water was used at a dilution rate of 1:1. Solids were filtered from the mixture by means of paper filters. The filtrate, called pineapple juice, contained the bromelain enzyme. The samples were frozen at  $-5^{\circ}\text{C}$  (Cesar et al., 1999) and stored (for a maximum of three months) until used in the experiments.

### Micellar solution

The cationic micellar solution contained BDBAC as the surfactant, isooctane as the solvent and hexanol as the co-solvent (Kilikian et al., 2000).

### Backward extraction solution

The bromelain-rich (raffinated) phase was treated

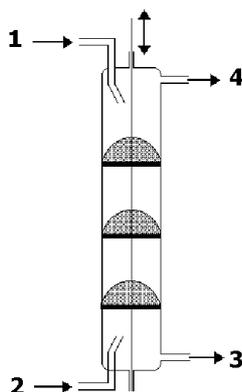
with a buffered phosphate solution (citric acid/sodium phosphate) and sodium chloride.

### Previously determined conditions

In a previous study (Fileti et al., 2008), it was found that the best concentrations of the surfactant agent, co-solvent, salt and pH for back and forward batch extraction were 100 mM, 10% v/v, 1 M, 3.5 and 8, respectively. These same conditions were maintained for continuous extraction. On comparing the purification factor obtained under these conditions – 3.29 – to that obtained via two-phase aqueous extraction with PEO-PPO-PEO block copolymers (Rabelo et al., 2004) – 1.25 – it became evident that the reversed micelle process was well worth studying.

### Extractor description

A diagram of the pulsed-cap extraction microcolumn used in the present study is shown in Figure 1. The glass column had a height of 19 cm and an internal diameter of 2.54 cm. Three perforated caps were mounted on a central stainless steel stem at 4-cm distances. The caps were made of MESH-24 sieves, with a 38% free-flow area. A pulse frequency controller drove the 2.8-cm vertical movements of the stem. Thus, part of the light phase was pulverized and uniform dispersion was attained. Another portion of the light phase was retained beneath the cap, increasing the contact between the phases. The inlet flow rates of the dispersed and continuous phases were maintained constant by previously calibrated peristaltic pumps.



**Figure 1** - Extraction microcolumn with pulsed caps: (1) Pineapple juice inlet (continuous phase); (2) Micellar solution inlet (dispersed); (3) Heavy phase outlet (extracted); (4) Light phase outlet (raffinated).

## Experimental procedures

### Forward extraction

The column was filled with pineapple juice. The feeding pumps were activated with the total flow rate (pineapple juice plus micellar solution streams) set at 8mL/min (constant) and the ratio between the light phase flow rate and the total flow rate set at 0.5. The pulse frequency (i.e., the time interval between the pulses) was then adjusted to 4 seconds.

In order to capture the dynamic behavior and the nonlinearities of the column, samples were

collected from the raffinated phase (micellar phase outlet) every three minutes. Once the steady state was reached (approximately 21 minutes), the operating conditions were modified according to the central composite design for the experiment (Tables 1 and 2). The independent variables were: the ratio between the light phase flow rate and the total flow rate; and the time interval between the pulses. The effects on the purification factor (PF) and on total protein ( $P_i$ ) were measured and then modeled via neural networks.

**Table 1** - Levels and values of the central composite design.

Independent variable	Level		
	(-1)	(0)	(+1)
Light phase flow rate/ Total flow rate	0.3	0.5	0.7
Time interval between pulses (s)	6	4	2

### Backward Extraction

The protein-rich (micellar) phase obtained from the extraction was mixed with an equal volume of the backward extraction solution. The contents of the tube were stirred in a vortex for three minutes.

In order to split the phases, the mixture was centrifuged at 8000 rpm for 5 min. The light phase (micellar) was rejected. Assays of enzyme activity and total protein were conducted on samples from the heavy aqueous phase.

**Table 2** - Matrix of the central composite design.

Run no.	Light/total flow rate	Time interval
1	-1	-1
2	-1	+1
3	+1	-1
4	+1	+1
5	-1.41421	0
6	+1.41421	0
7	0	-1.41421
8	0	+1.41421
9 (C)	0	0
10 (C)	0	0
11 (C)	0	0
12 (C)	0	0

### Enzyme activity assay

Enzyme activity was determined by enzymatic hydrolysis of 2% casein (w/v) at pH 7.5 and 37°C for 10 min. Trichloroacetic acid (TCA) was used in the precipitation of the non-hydrolyzed product. The amount of soluble peptides in the TCA was determined by measuring the absorbance at 280 nm. The method defines a unit of enzyme activity as the amount of enzyme required to modify the absorbance at 280 nm by 1.0.

### Total protein assay

Total protein was determined using the Lowry method (1951) and a BSA standard.

### Extraction performance indexes

a) Total protein yield (TP):

$$TP(\%) = \frac{P_2}{P_1} \cdot 100 \quad (3)$$

where  $P_1$  is the total protein concentration (g/L) of the pineapple juice and  $P_2$  is the total protein in the backward extraction solution.

b) Enzyme activity yield (EA):

$$EA(\%) = \frac{A_2}{A_1} \cdot 100 \quad (4)$$

where  $A_1$  is the enzyme activity measurement (in units per liter obtained from the pineapple juice) and  $A_2$  is the enzyme activity of the backward extraction solution.

c) Purification factor (PF), which represents the increase in the purity of the bromelain enzyme:

$$PF = \frac{A_2/P_2}{A_1/P_1} \quad (5)$$

### Neural modeling

The modeling process employed a feedforward network and a Levenberg-Marquardt optimization algorithm in conjunction with Bayesian regularization. In the Neural Networks Toolbox of the MATLAB® Software, this algorithm is denoted "trainbr". Because neural network training would be more efficient when preprocessing normalization was performed for the input and target output variables, the data set was normalized in the [0.1, 0.9] range. A sigmoid function was applied to the neurons in the hidden

and output layers. The input layer was comprised of a moving window of the operating variables, i.e. the ratio between the light phase flow rate and the total flow rate; and the time interval between the pulses. Total protein yield and the purification factor were the target variables, i.e. the output layer neurons. For the purpose of determining the best number of backward steps for the moving window and number of neurons for the hidden layer, the mean square error (MSE) and the effective number of parameters ( $\gamma$ ) were monitored for test and training sets.

In all the experimental runs, the purification factor (PF) and the total protein (TP) were measured at the micellar phase outlet every three minutes. An experimental data set of 272 input/output vectors was obtained, of which 204 vectors constituted the training set and 68 vectors the test set.

## RESULTS AND DISCUSSION

### Experimental design

In accordance with the results obtained during the experiment, the range of each variable is summarized in Table 3. In order to determine the best operating point, the objective function "desirability" of the Statistica® software was used. This function assumes values in the 0 to 1 range, and maximization of the function yields the optimal global operating point, i.e. the highest productivity rate (total protein multiplied by raffinated phase flow rate) capable of guaranteeing an elevated purification factor.

**Table 3** - Operating ranges of the input and output variables in the neural model.

	Input		Output	
	Light phase/total flow rate	Time between pulses (s)	Purif. factor	Total prot. yield (%)
Upper bound	0.782	1	7.64	12.08
Lower bound	0.217	7	0.89	4.79

Figure 2 shows the desirability response surface as a function of the light phase/total flow rate versus the time interval between pulses. The optimal conditions obtained were: light phase/total flow rate equal to 0.67 and time interval between pulses equal to 1 second. At this point, productivity was 1.29 mL/min and the purification factor was 4.96. As previously cited, the batch extraction by reversed micelles yielded a purification factor of

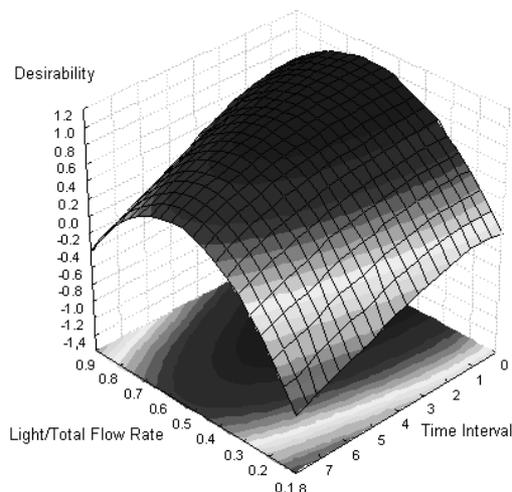
3.29 (Fileti et al., 2008). In the present work, a purification factor of 4.96 and productivity of 1.29 mL/min were obtained via continuous extraction. These results also confirmed that the extraction by reversed micelles was superior to two-phase aqueous extraction with PEO-PPO-PEO block copolymers, which yielded a purification factor of only 1.25 (Rabelo et al., 2004).

### Neural modeling

An increase in the number of backward steps of the moving window means changing the number of neurons in the network input layer (see Table 4). To do so, five different neural models were tested.

The mean square errors (MSE) of the target variables – the purification factor and total protein – were monitored for the test and training sets (Fig. 3 and 4). For this assessment, the number of

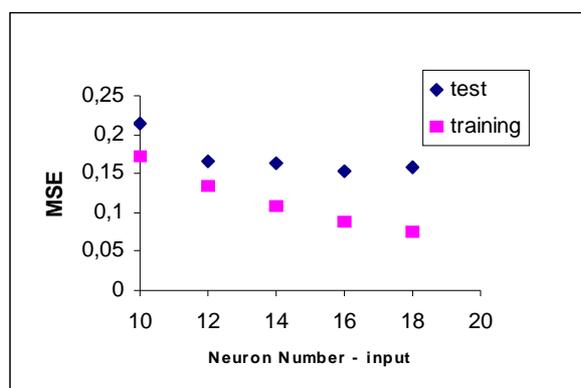
neurons in the hidden layer was initially set at a very high level since any excess connections would be eliminated by the training method adopted (Levenberg-Marquardt algorithm with Bayesian regularization). The test set curves depicted in Figures 3 and 4 indicated that the mean error (MSE) did not decrease when more than 16 input neurons were used. Hence, the backward time adopted was 21 minutes (16 neurons), as shown in Figure 5.



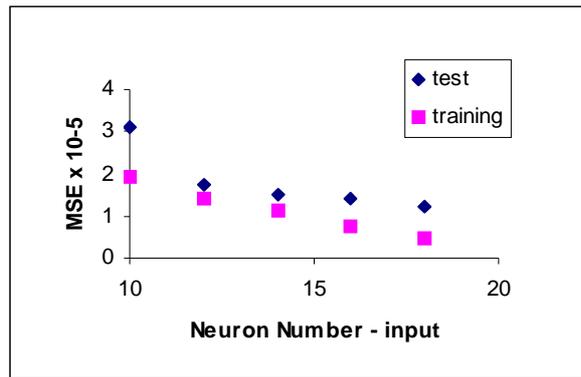
**Figure 2** - Desirability response surface.

**Table 4** - Input moving window and corresponding number of neurons in input layer.

Neuron number input layer	Backward time (min)
10	12
12	15
14	18
16	21
18	24



**Figure 3** - Mean square errors of the purification factor (PF), estimated via neural model, for test and training sets.



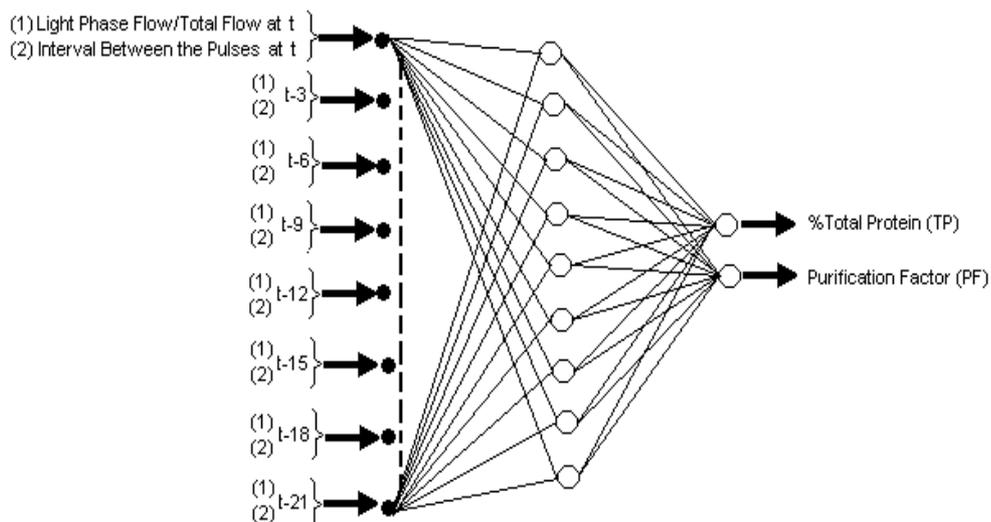
**Figure 4** - Mean square errors of the total protein yield (TP), estimated via neural model, for test and training sets.

Using the inputs and outputs presented in Figure 5, the number of neurons in the hidden layer changed as shown in Table 5. The mean square error (MSE) and the effective number of parameters ( $\gamma$ ) were used as indexes for evaluating the performance of the model.

According to the Hagan and Fosse criterion described above and the data presented in Table 5, the best number of neurons for the hidden layer is 9, given that the effective number of parameters did not change as the number increased. The analysis of the MSE for the test set (last two

columns in Table 5) confirmed this choice. Therefore, the best network topology was defined as 16-9-2, in exact accordance with the scheme portrayed in Figure 5.

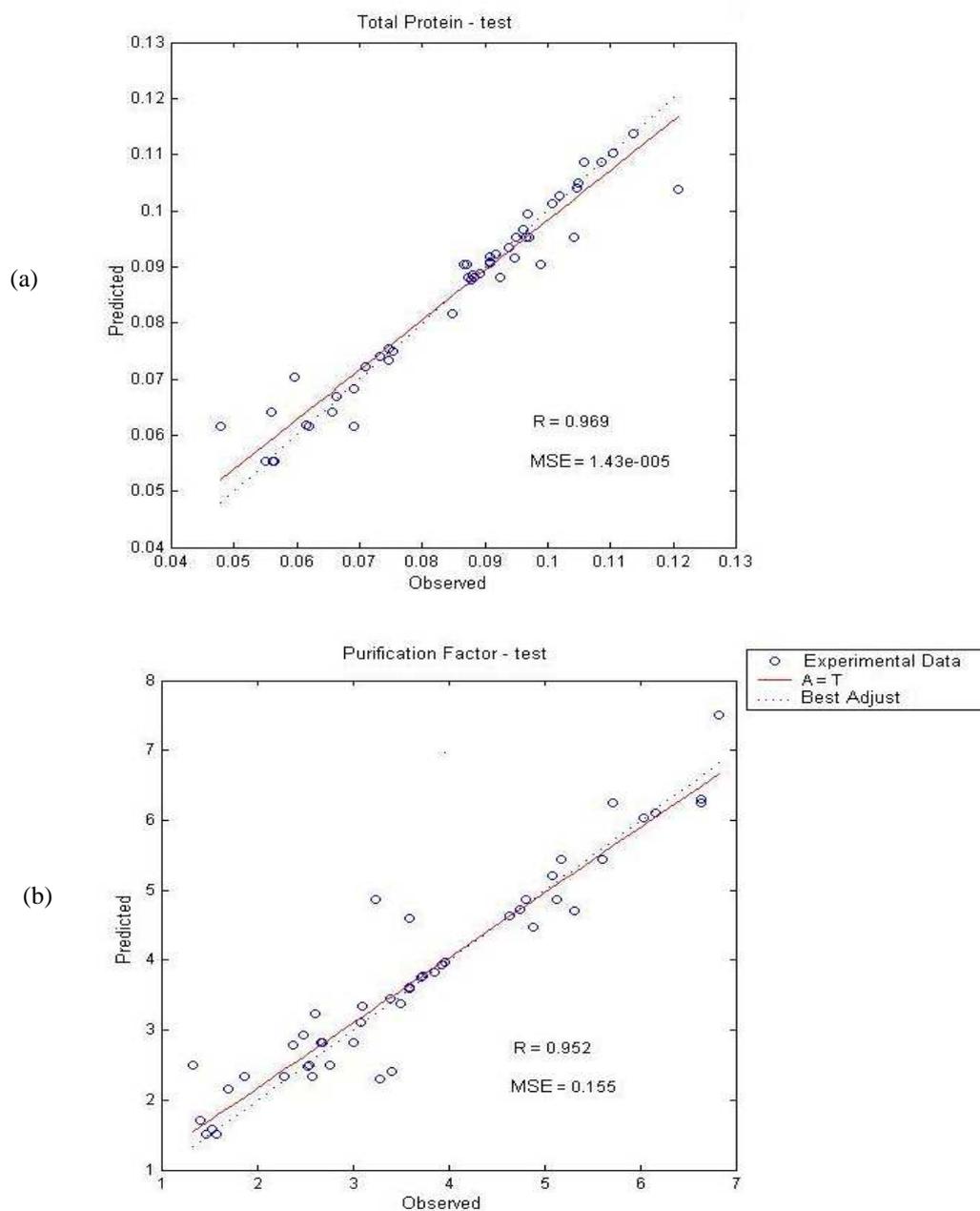
A comparison of observed (T) and predicted (A) points for the test set is presented in Figure 6. The closer the points to the diagonal line (A=T), the more accurate the neural model. The dispersion plots showed that the neural model successfully and simultaneously predicted both the purification factor and the total protein yield from the historical data employed.



**Figure 5** - Neural model architecture of the pulsed-cap microcolumn.

**Table 5** - Mean square errors (MSE) and effective number of parameters ( $\gamma$ ): indexes used to drive the search for the best number of neurons in the hidden layer.

Hidden layer neurons	Effective number of parameters( $\gamma$ )	MSE (training set)		MSE (test set)	
		TP( $10^{-6}$ )	PF	TP ( $10^{-5}$ )	PF
6	104	8.31	0.0973	1.55	0.176
7	114	7.93	0.0868	1.55	0.161
8	119	7.56	0.0873	1.44	0.157
9	122	7.56	0.0861	1.43	0.155
10	122	7.80	0.0859	1.49	0.152



**Figure 6** - Dispersion plots of: (a) total protein yield (%) and (b) purification factor.

## CONCLUSIONS

The developed neural model successfully and simultaneously predicted the purification factor and the total protein yield from the historical data used. This kind of model could be particularly important due to its capacity to help a human operator adjust variables during experimental runs in order to achieve target performance indexes.

The findings indicated that aligning the central composite designs and neural models based on moving windows provided concise information for constructing the suitable empirical patterns for nonlinear and transient processes.

## ACKNOWLEDGEMENTS

The authors would like to thank CNPq for the financial support extended to this project.

## RESUMO

Uma micro-coluna com campânulas pulsantes foi utilizada para a extração de bromelina a partir de suco de abacaxi, usando micelas reversas. A solução catiônica micelar foi composta do surfactante BDBAC, do solvente iso-octano e do co-solvente hexanol. Seguindo um planejamento experimental, perturbações foram impostas à coluna de extração com o objetivo de capturar seu comportamento dinâmico e suas não-linearidades, usando a razão entre a vazão da fase leve e vazão total, e o intervalo de tempo entre os pulsos. Os efeitos das variáveis independentes sobre o fator de purificação e sobre o rendimento em proteínas totais foram modelados via redes neurais artificiais. A melhor topologia de rede obtida foi definida como 16-9-2, usando um esquema de janela móvel no tempo das variáveis independentes. O modelo neural obtido do histórico do processo se mostrou adequado para prever simultaneamente o fator de purificação e o rendimento do processo em proteínas totais.

No ponto ótimo de operação, foi encontrado um fator de purificação de 4.96, com produtividade de 1.29 mL/min.

## REFERENCES

- Aires-Barros, M. R., Taipa, M. A. and Cabral, J. M. S. (1994), Isolation and purification of lipases. In *Lipases: their structure, biochemistry and application*, ed. Wooley, P. and Petersen, S. B. Cambridge University Press, Great Britain, pp. 234-270.
- Cesar, A. C., Silva, R. and Lucarini, A. C. (1999), Pineapple stem and rind proteolytic enzyme recovery (In Portuguese). *Revista de Iniciação Científica*, 1, 47-53.
- Fileti, A. M. F., Fischer, G. A., Santana, J. C. C. and Tambourgi, E. B. (2008), Batch and continuous extraction of bromelain enzyme by reversed micelles. Submitted to *Brazilian Archives of Biology and Technology*.
- Hagan, M. (2002), Backpropagation. In *Neural Network Toolbox User's Guide for Use with MATLAB®, Version 4*, ed. Demuth, H. and Beale, M. The Mathworks, Inc., Natick, MA, chap. 5. URL: <http://www.mathworks.com/support/product/NN>
- Hagan, M. T. and Fousse, F. D. (1997), Gauss-Newton approximation to Bayesian learning. In *Proceedings of the 1997 International Joint Conference on Neural Networks*. URL: <http://elec-engr.okstate.edu/mhagan/>.
- Himmelblau, D.M. (2000), Applications of artificial networks in chemical engineering. *Korean J. Chem. Eng.*, **17** (4), 373-392.
- Kilikian, B. V., Bastazin, M. R., Ninami, N. M., Gonçalves, E. M. R. and Pessoa Jr., A. (2000), Liquid-liquid extraction by reversed micelles in biotechnological processes. *Brazilian Journal of Chemical Engineering*, **17** (1), 29-38.
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951), Protein measurement with the folin phenol reagent. *J. Biol. Chem.*, **193** (1), 265-275.
- Pinto, G. A., Sousa Jr., R. and Giordano, R. C. (2005), Comparison of performance of different algorithms in noisy signals filtering of process in enzymatic hydrolysis of cheese whey. *Brazilian Archives of Biology and Technology*, **48**, 151-159.
- Rabelo, A. P. B., Tambourgi, E. B. and Pessoa Jr., A. (2004), Bromelain partitioning in two-phase aqueous systems containing PEO-PPO-PEO block copolymers. *Journal of Chromatography B*, **807**, 61-68.

Received: December 07, 2007;

Revised: August 21, 2008;

Accepted: August 20, 2009.