CHEMICAL ENGINEERING TRANSACTIONS
VOL. 43, 2015
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Chief Editors: Sauro Pierucci, Jiří J. Klemeš

ISBN 978-88-95608-34-1; **ISSN** 2283-9216
DOI: 10.3303/CET1543237

Use of Artificial Neural Networks to predict Aqueous Two-Phases System Optimal Conditions on Bromelain's Purification

Diego de Freitas Coêlho[a], Camila Alves Silva[a], Camila Sacconi Machado[a], Edgar Silveira[b], Elias Basile Tambourgi[a]

[a]Department of Chemical Engineering Systems, School of Chemical Engineering, State University of Campinas - UNICAMP - Av. Albert Einstein, 500. P.O. 6066, Zip Code: 13083-970. Campinas-SP, Brazil.
[b]Institute of Genetics and Biochemistry, Federal University of Uberlandia – UFU - Campus Umuarama - Bloco 2E - Sala 246 - 2º Piso, Av. Pará, 1720. Zip Code: 38400-902, Uberlândia – MG, Brazil.
dfcoelho@feq.unicamp.br

Bromelain is the denomination chosen to the group of endoproteases obtained from pineapple and from most of plants belonging to *Bromeliaceae* family. These enzymes have being widely studied in researches across the world due its physiological activity and biotechnological potential. While Brazil still cultivating over 60,000 hectares of pineapple, there is a optimistic trend that aim bromelain's recovery from agriculture residues (stalk and leaves) and fruit processing residues (stem and bark) leading to a fully integrated process which aggregate value to vegetal residues. Our previous studies applied Aqueous Two-Phases Systems and Fractional Precipitation to purify bromelain and achieve purification factor and yield of 11.80 and 87.36, respectively. However, such studies were designed and analysed using Design Of Experiments (DOEs), which lead to an optimal condition but cannot predict with accuracy the complex phenomena of partitioning using ATPS. This work is part of an initiative that aims establish a protocol to calculate more accurate partitioning data through the use of Artificial Neural Networks (ANNs) over a dataset that has being improved continuously. The ANN will determine the relationship between five input parameters (temperature, PEG's molar mass, concentration of PEG, concentration of ammonium sulphate and dilution factor of sample) with three output parameters (protein partition coefficient, Activity partition coefficient and purification factor). The method applied a feed-forward neural network trained with Levenberg-Marquardt algorithm and the Bayesian regularization over the normalized experimental data. The network generated proved the reliability of the method which combined datasets from different DOEs and obtained regression coefficient (~.99) and error (MSE ~0.02) satisfactory for such amount of data used so far.

1. Introduction

The group of thiol-endopeptidases known as Bromelain can be extracted from any plant belonging to the *Bromeliaceae* family (Heinicke and Gortner, 1957) and originally was used as a folk medicine by the aboriginal inhabitants of Central and South America to treats several sicknesses (Taussig and Batkin, 1988). These enzymes had proven therapeutic applications as an anti-inflammatory drug (Salas et al., 2008), in the treatment of allergic disease (Secor Jr et al., 2005), carcinopreventive agent (Harrach et al., 1994), antithrombotic and fibrinolytic activities (Maurer, 2001).
Unlike most enzymes, bromelain is stable and highly active in both acid and alkaline solutions (which expand its range of possible applications) and holds its proteolytic activity even at 60°C, when most enzymes denatures (Bhattacharya and Bhattacharyya, 2009).
Their purification has being studied for extensively in the last decade: While Harrach et al. (1995) applied Fast Protein Liquid Chromatography as a way to Isolate and characterise the enzymes, Rabelo et al. (2004)

Please cite this article as: Coelho D.F., Silva C.A., Machado C.S., Silveira E.C., Tambourgi E.B., 2015, Use of artificial neural networks to predict aqueous two-phases system optimal conditions on bromelain's purification, Chemical Engineering Transactions, 43, 1417-1422
DOI: 10.3303/CET1543237

decided to study the use of Aqueous Two-Phase Systems, which has a higher throughput capacity than chromatography techniques, to purify the same enzymes. Over the years, researchers have investigated the use of alternatives bulk recovery techniques, such as expanded bed absorption (Silveira et al., 2009) and Fractional Precipitation (Martins et al., 2014) but the use of ATPS if by far the most employed: Reverse micelles (Umesh Hebbar et al., 2008) , High-speed counter-current chromatography (Yin et al., 2011), combining it with fractional precipitation (Coelho et al., 2012) and endless variety of salts and polymers, such as PEG/potassium phosphate (Ferreira et al., 2014) and PEO–PPO–PEO block polymers (Rabelo et al., 2004).

While the researchers seem to have tried to exhaust the possible combinations of components and modes when using ATPS, they all lack in use a fast and reliable method to determine the best operational conditions.

At this moment, there is no method to determine such characteristics with no use of a time-consuming and laborious experimental method. ATPS's characterisations rely on empirical determination of purification parameters for every single modification in the systems under study. One might use statistic methods (such as Design of Experiments) to reduce experimental work but it still lacks in handle trade-off problems as a purification process.

What if we could use a cluster of randomly distributed data obtained to optimize specific parameters in a much broader purpose? That is exactly the purpose of this initiative: to combine all data generated through decades of research in a database that can be constantly improved.

2. Materials and Methods

2.1 ATPS Data acquisition

All experimental data was acquired in projects realized previously, in which we used Design of Experiments and Response Surface Methodology to optimize the parameters or determine a specific operational condition. At this study we restricted the data to those a limited number of variables and in a specific range. The chosen input variables and the correspondents ranges are presented in the Table 1. Those variables were selected from studies in which their effects in the purification were evaluated. The ones presented here showed higher impact during experiments.

Table 1: Input variables used in the neural model and their range

Input variables	Description	Range
Temperature (°C)	Operational Temperature	5 - 25
MMPEG	PEG Molecular Mass	2,000 – 4,000 – 6,000
(m/m,%) $(NH_4)_2SO_4$	Concentration of Ammonium Sulphate	7 a 20
(m/m,%) PEG	PEG Concentration	9 a 30
Dilution (%)	Dilution Factor	25 - 50 - 75

2.2 Mathematical definition of Output Variables

As output variables, we chose the protein partition coefficient (K_P) and the enzymatic partition coefficient (K_A), as described in the table 2. Coelho et al. (2013) describes in details the equationing for the chosen output variables.

Table 2: Output variables used in the neural model and their range.

Output variables	Description	Minimum	Maximum
KP	Protein Partition Coefficient	0	100
KA	Enzymatic partition Coefficient	0	100
PF	Purification Factor	0	98.75

2.3 Results and Discussion

As no mathematical model can predict the complex nature of aqueous two-phase systems in enough accuracy, we decided to evaluate the application of Artificial Neural Networks (ANN's) in the modelling and prediction of partitioning parameters.
Basically, an artificial neural network is a system composed of hundreds of units; artificial neurons (AN) or processing elements (PE), which are connected with coefficients (weights) constituting the neural structure and is arranged in layers as can be seen in the figure 1 (Chrislb, 2005).

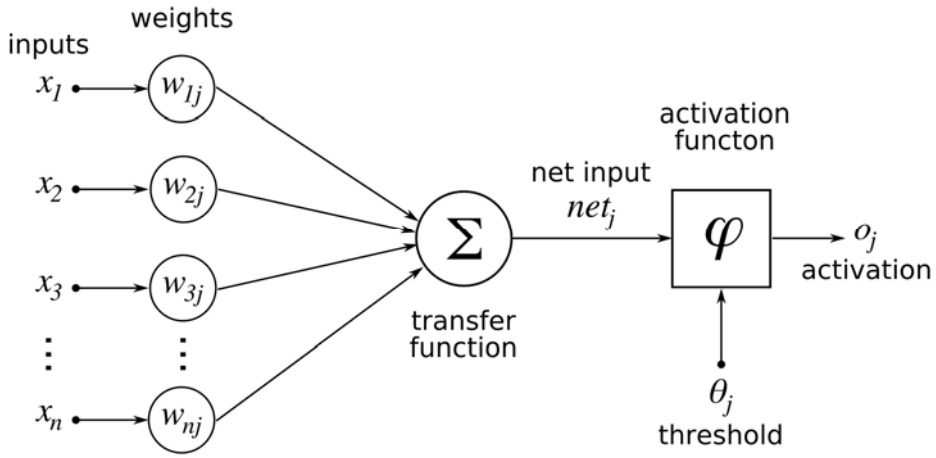

Figure 1: Diagram of an artificial neuron

When a set of input and output data is used to stimulated a "learning" network, such data is used to adjust each neuron's "weights" through successive changes in its values so that the network implement and execute the desired functions (Brumatti, 2005) and apply the "knowledge" gained from past experiences to new problems or conditions.
This study used the Levenberg-Marquardt optimization as the training algorithm but it was used Bayesian regularization in order to improve generalization and avoid overfit. This gain is a consequence of the smaller weights calculated by the algorithm, which make the network respond smoother (Foresee and Hagan, 1997).
As mentioned, the neural model used either a backpropagation network or a feedforward network coupled with Levenberg-Marquardt and Bayesian regularization optimization algorithms, all available in the Neural Network toolbox from MATLAB ® Software (The MathWorks Inc., 2013). All variables were normalized between 0 and 0.9.
As activation functions, were tested hyperbolic tangent function, sigmoid function and a linear function. The neural networks were set and trained combining the different neural models and activation functions (besides the number of neurons) in order to determine which topology converged faster. To estimate the deviation between the ANN's results and the experimental data, we used the Mean Squared Error (MSE) and the regression coefficient (R), which are the most common parameters used on its analysis (Beale et al.). The neural network would be considered fit to the experimental data when MSE tend to zero and R to 1.
Among the results (Table 3) obtained from the topology optimization for the neural network, the best configuration is the T3, which used 30 neurons and no intermediary layers. We compared values of R, MSE and also the convergence time, being the last one the main factor that made T3 better than T7. These results were obtained during the initial step of this research project and hence used a dataset with only 120 experiments and such variance obtained is expected. Although Aqueous Two-Phases Systems has been used for decades, there is no such thing as a model that can precisely predict any property from those systems with no experimental data. This creates an even harder task to find an appropriated approach to study it.

Table 3: Neural Networks training results using KP, KA and PF as output variables

T	(Nrn,Lyr)	R	MSE	KP	KA	PF	Inter. Function	Inter. Function 2	Output Function
T1	(10,1)	0.8862	0.14369	0.88394	0.91154	0.67951	Tansig	Purelin	Purelin
T2	(20,1)	0.9514	0.06371	0.95463	0.95836	0.87268	Tansig	Purelin	Purelin
T3	(30,1)	0.9846	0.02045	0.99418	0.98574	0.92548	Tansig	Purelin	Purelin
T4	(30,2)	0.6465	0.40178	0.63385	0.66119	0.49902	Tansig	Purelin	Purelin
T5	(30,2)	0.9807	0.02573	0.99149	0.98097	0.91795	Logsig	Purelin	Purelin
T6	(40,2)	0.9858	0.01901	0.99351	0.98713	0.93540	Tansig	Purelin	Purelin
T7	(50,2)	0.9854	0.01953	0.99307	0.98720	0.93223	Tansig	Purelin	Purelin
T8	(5,2)	0.7577	0.28486	0.75959	0.76642	0.54001	Tansig	Purelin	Purelin
T9	(10,2)	0.8814	0.14927	0.88072	0.90113	0.70156	Tansig	Purelin	Purelin
T10	(20,2)	0.9688	0.04160	0.98592	0.97862	0.80179	Tansig	Purelin	Purelin
T11	(30,3)	0.9840	0.02118	0.99243	0.98553	0.92806	Tansig	Purelin	Purelin
T12	(30,4)	No Convergence							

Tansig: Hyperbolic Tangent, Purelin: Linear, Logsig: Sigmoidal, Nrn: Neurons, Lyr: Layers

In this set of simulations we tried to test an even bigger number of combinations between the number of neurons, number of intermediary layers and even the activation functions but most of them couldn't even converge. Thus, the topology with 30 neurons and using a hyperbolic function as activation function provided the best results.

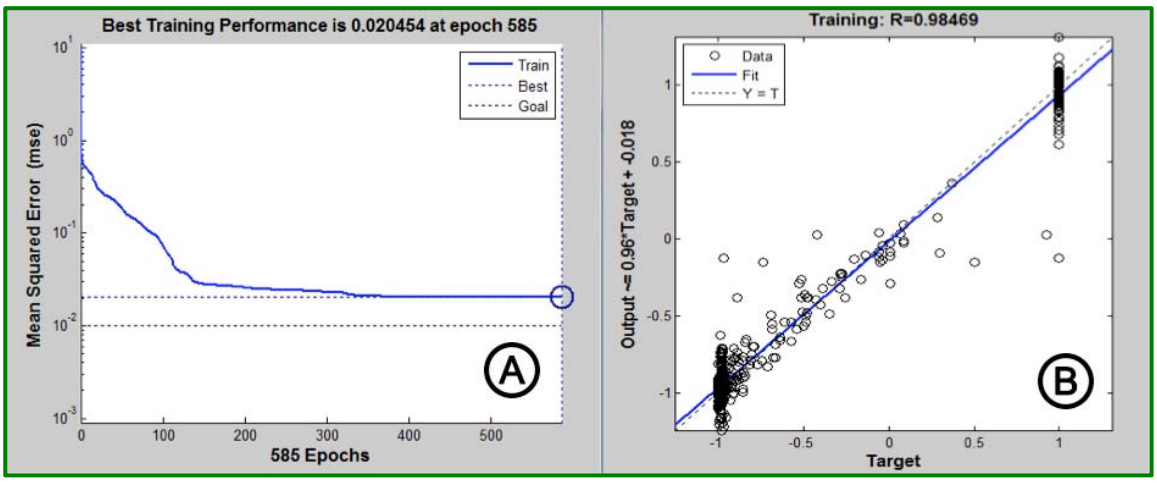

Figure 2: Convergence (A) and Regression (B) for the best topology obtained (T3, Table 3)

Figure 2 presents the fitting parameters results obtained for T3 topology, which was the one that returned the best results. Figure 3 presents the regression data using the topology T3 for the output variables (K_A, K_P and P_F). It is noticeable that data represents well the experimental data but we still expect to be able to improve the model in at least 5 %. Positive results are mainly due Bayesian regularization, which improve the generalization capability of the model even in a reasonable high operational range (Fileti et al., 2010).

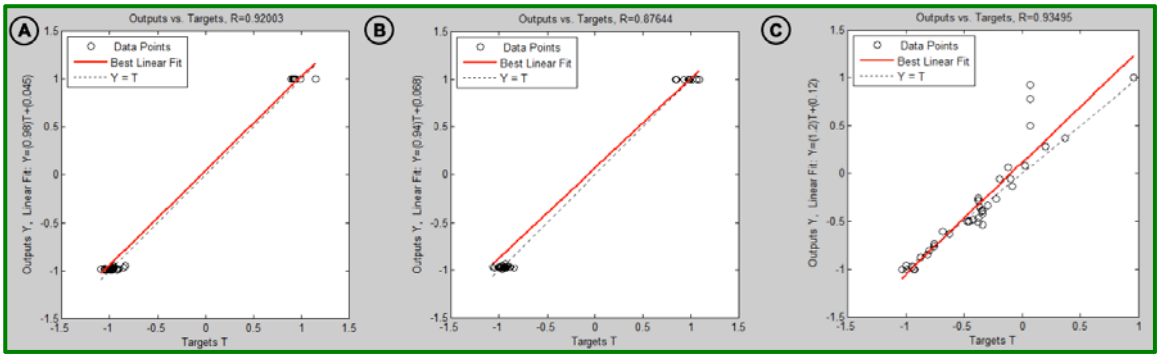

Figure 3: Output Test with T3 Topology output Neural Network for KA, KP and PF respectively

However, the study still lacks in explain why we couldn't decrease MSE and the variation observed in figure 2 and at this point the network just proved that was able to correlate data from several experiments and show we can improve the network and use it to obtain a better understanding of the process.

3. Conclusions

The network generated proved the reliability of the method by modelling combined data from different experimental designs and obtaining reasonable regression coefficient (~.99) and error (MSE ~0.02).
At this stage the neural network was able to model and predict with certain precision the data handled. However, it is necessary to improve the robustness by increasing the number of input in the database used to train the network. When complete, the neural model will be able to predict operational points, analyze influence of different factors and select conditions in which a trade-off is present.

Acknowledgements

The authors would like to acknowledge the financial support of FAPESP (São Paulo Research Foundation), PROPP-UFU (Dean of Research and Graduate Studies at the Federal University of Uberlândia) and CNPq (National Council for Scientific and Technological Development).

References

Beale, M. H., Hagan, M. T. & Demuth, H. B., Neural network toolbox 7.

Bhattacharya, R. & Bhattacharyya, D., 2009, Resistance of bromelain to sds binding, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1794**,** 698-708.

Brumatti, M., 2005, Redes neurais artificiais, *Vitória, Espírito Santo*.

Chrislb,2005, Diagram of an artificial neuron.,*In:* Artificialneuronmodel_English.Png (ed.) *(created by Chrislb) [GFDL (*http://www.gnu.org/copyleft/fdl.html*) or CC-BY-SA-3.0 (*http://creativecommons.org/licenses/by-sa/3.0/)%5D*, via Wikimedia Commons*, http://commons.wikimedia.org/wiki/File:ArtificialNeuronModel_english.png: Wikimedia Commons.

Coelho, D., Silveira, E., Pessoa Junior, A. & Tambourgi, E., 2012, Bromelain purification through unconventional aqueous two-phase system (peg/ammonium sulphate), *Bioprocess and Biosystems Engineering,* 35**,** 1-8.

Coelho, D. F., Silveira, E., Pessoa Junior, A. & Tambourgi, E. B., 2013, Bromelain purification through unconventional aqueous two-phase system (peg/ammonium sulphate), *Bioprocess and Biosystems Engineering,* 36**,** 185-192.

Ferreira, J. F., Sbruzzi, D., Barros, K. V. G., Ehrhardt, D. D. & Basile, E., 2014, Purification of bromelain enzyme from curauá (ananaserectifolius lb smith) white variety, by aqueous two-phase system peg 4000/potassium phosphate, *J. Chem,* 8**,** 395-399.

Fileti, A. M. F., Fischer, G. A. & Tambourgi, E. B., 2010, Neural modeling of bromelain extraction by reversed micelles, *Brazilian Archives of Biology and Technology,* 53**,** 455-463.

Foresee, F. D. & Hagan, M. T. Gauss-newton approximation to bayesian learning. Proceedings of the 1997 international joint conference on neural networks, 1997. Piscataway: IEEE, 1930-1935.

Harrach, T., Eckert, K., Schulze-Forster, K., Nuck, R., Grunow, D. & Maurer, H. R., 1995, Isolation and partial characterization of basic proteinases from stem bromelain, *Journal of Protein Chemistry,* 14**,** 41-52.

Harrach, T., Garbin, F., Munzig, E., Eckert, K. & Maurer, H. R., 1994, Bromelain: An immunomodulator with anticancer activity, *European Journal of Pharmaceutical Sciences,* 2**,** 164.

Heinicke, R. M. & Gortner, W. A., 1957, Stem bromelain—a new protease preparation from pineapple plants, *Economic Botany,* 11**,** 225-234.

Martins, B. C., Rescolino, R., Coelho, D. F., Zanchetta, B., Tambourgi, E. B. & Silveira, E., 2014, Characterization of bromelain from ananas comosus agroindustrial residues purified by ethanol factional precipitation, *Chemical Engineering Transactions,* 37**,** 781-786.

Maurer, H. R., 2001, Bromelain: Biochemistry, pharmacology and medical use, *Cellular and Molecular Life Sciences,* 58**,** 1234-1245.

Rabelo, A. P. B., Tambourgi, E. B. & Pessoa, A., 2004, Bromelain partitioning in two-phase aqueous systems containing peo-ppo-peo block copolymers, *Journal of Chromatography B,* 807**,** 61-68.

Salas, C. E., Gomes, M. T. R., Hernandez, M. & Lopes, M. T. P., 2008, Plant cysteine proteinases: Evaluation of the pharmacological activity, *Phytochemistry,* 69**,** 2263-2269.

Secor Jr, E. R., Carson Iv, W. F., Cloutier, M. M., Guernsey, L. A., Schramm, C. M., Wu, C. A. & Thrall, R. S., 2005, Bromelain exerts anti-inflammatory effects in an ovalbumin-induced murine model of allergic airway disease, *Cellular Immunology,* 237**,** 68-75.

Silveira, E., Souza-Jr, M. E., Santana, J. C. C., Chaves, A. C., Porto, A. L. F. & Tambourgi, E. B., 2009, Expanded bed adsorption of bromelain (e.C. 3.4.22.33) from *ananas comosus* crude extract, *Brazilian Journal of Chemical Engineering,* 26**,** 149-157.

Taussig, S. J. & Batkin, S., 1988, Bromelain, the enzyme complex of pineapple (ananas comosus) and its clinical application. An update, *Journal of Ethnopharmacology,* 22**,** 191-203.

The Mathworks Inc.,2013, Matlab ® software.*In:* 8.1.0.604 ed, Natick, Massachusetts: The MathWorks Inc.,.

Umesh Hebbar, H., Sumana, B. & Raghavarao, K. S. M. S., 2008, Use of reverse micellar systems for the extraction and purification of bromelain from pineapple wastes, *Bioresource Technology,* 99**,** 4896-4902.

Yin, L., Sun, C. K., Han, X., Xu, L., Xu, Y., Qi, Y. & Peng, J., 2011, Preparative purification of bromelain (ec 3.4.22.33) from pineapple fruit by high-speed counter-current chromatography using a reverse-micelle solvent system, *Food Chemistry,* 129**,** 925-932.