Food Chemistry 111 (2008) 683–686

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Extraction of pepsin-soluble collagen from grass carp (*Ctenopharyngodon idella*) skin using an artificial neural network

Lingzhao Wang [a], Bao Yang [b,*], Rui Wang [b], Xiuqiao Du [a]

[a] School of Marine Science and Technology, Huaihai Institute of Technology, Lianyungang 222005, China
[b] South China Botanical Garden, Chinese Academy of Sciences, Guangzhou Leyiju 510650, China

ARTICLE INFO

Article history:
Received 5 December 2007
Received in revised form 4 January 2008
Accepted 15 April 2008

Keywords:
Grass carp skin
Collagen
Pepsin
Artificial neural network
Ctenopharyngodon idella

ABSTRACT

A multilayer feed-forward neural network trained with an error back-propagation algorithm was used to evaluate the effects of pepsin amount, reaction time and pH on the yield of pepsin-soluble collagen. A positive correlation was observed between the yield and the amount of pepsin and also the reaction time. The yield increased with an increase of pH to nearly 3, thereafter yield decreased. The trained network gave a regression coefficient (r^2) of 0.97 and a mean squared error (MSE) of 0.21, which implied a good generalisation of the network. Based on the genetic algorithm, the optimal extraction conditions to obtain the highest yield were determined to be pH 3.4, 53.3 unit/mg of pepsin and 35.2 h. The predicted yield value was 30.3%. As the estimated optimal extraction conditions were used in the actual preparation of the pepsin-soluble collagen, the yield was measured experimentally to be 29.3 ± 0.8%, which was not significantly different ($p > 0.05$) from the predicted value. The response surface plots showed the yield of pepsin-soluble collagen as a function of two factors under various extraction conditions.

1. Introduction

Grass carp is a stomachless freshwater fish species widely dispersed in lakes and rivers of Asia. It is one of the "four major cultured fish species" with black carp (*Mylopharyngodon piceus*), silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) in China (Li, Lu, & Zhou, 1995). Due to its special taste and abundant nutrients (e.g., unsaturated fatty acids and proteins), it is well accepted by consumers around the world. Grass carp skin has significant amounts of high-quality proteins, which can be used by the food, medicine and cosmetic industries (Wasswa, Tang, Gu, & Yuan, 2007).

Collagen has been employed in medical applications for decades with good biocompatibility and safety. It can form strong and stable fibres by self-aggregation and cross-linking, which makes it useful in drug delivery systems (Lee, Singla, & Lee, 2001). The structural characteristics of collagen have been established by earlier researchers, through amino acid analysis and physicochemical evaluation (Ramachandran & Sasisekharan, 1965). The collagen molecule consists of three polypeptide chains twisted around one another into a triple helix. The strands are bound together primarily by hydrogen bonds between carbonyl and amino groups. Covalent bonds are also involved in some cases (Harkness, 1966). The $(Gly\text{-}X\text{–}Y)_n$ repeating structure is found in the "collagenous" domains of all collagens. The glycine residue is the structural prerequisite for triple helix formation. The X and Y positions are often occupied by proline and hydroxyproline (Gelse, Pöschl, & Aigner, 2003).

Since pepsin was crystallised by Northrop in 1929, its activity has been extensively investigated (Campos & Sancho, 2003). It belongs to the family of aspartic proteases. Pepsin is a monomeric, two-domain, β-sheet-rich protein, with abundant acidic residues (43 out of 327). The catalytic site is formed by two aspartate residues (Antonov et al., 1978).

In this work pepsin was used to prepare collagen from grass carp skin. Pepsin amount, reaction time and pH are important variables affecting the extractability of collagen from grass carp skin. So far, the available literature on this topic is limited. A multilayer feed-forward neural network trained with an error back-propagation algorithm was employed to further evaluate the effects of these variables. Genetic algorithms were also used to optimise the conditions needed to obtain the highest yield of pepsin-soluble collagen.

2. Materials and methods

2.1. Materials

Several grass carp (*Ctenopharyngodon idella*) with average weight of 3.0–3.2 kg were purchased alive from a local market in Lianyungang, China. After being killed, the skins were removed, descaled and washed with distilled water. The cleaned skins were

* Corresponding author. Tel./fax: +86 20 37252960.
E-mail address: yangbao@china.com.cn (B. Yang).

0308-8146/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2008.04.037

minced into very small pieces (less than 4 mm) by a cutting mill with a size of 780 × 600 × 850 mm (ZB20L, Ruiheng Food Machine Co., Zhucheng, China). Three knives in the mill were rotated at 2000 rpm. Then the skin pieces were frozen at –18 °C for less than two days, prior to collagen extraction.

2.2. Chemicals

Lyophilised pepsin powder from porcine gastric mucosa with an activity of 3520 units/mg protein was purchased from Sigma–Aldrich (Guangzhou, China). Sodium chloride, acetic acid, sodium hydroxide were purchased from Guangzhou Chemical Company (Guangzhou, China). All other chemicals were of analytical grade.

2.3. Extraction of pepsin-soluble collagen

The pepsin-soluble collagen was extracted from grass carp skin according to the method of Kittiphattanabawon, Benjakul, Visessanguan, Nagai, and Tanaka (2005), with slight modification. Five grams of skin were precisely weighed. The following experiments were done in a room at 4 °C. To remove non-collagenous proteins, the skin was mixed with 100 ml of 0.1 M NaOH. The mixture was stirred using a magnetic stirrer (78-2, Kexing Instrument Co, Jintan, China) for 6 h, while the NaOH solution was changed every 2 h by filtration. After removal of the supernatant, the sample was washed with distilled water until the pH was neutral. The deproteinised skins were defatted with 100 ml of diethyl ether for 24 h. The diethyl ether was changed every 8 h. Defatted skins were washed with distilled water three times.

Collagen was extracted with pepsin in 0.5 M acetic acid. Pepsin amount (20–60 unit/mg defatted skin), reaction time (12–36 h), and pH (2.0–4.0) were chosen as the variables. The pepsin amount was calculated on the basis of its activity. Sodium acetate–acetic acid buffer was used to regulate the pH of the bulk solution. The extract was filtered with Whatman No. 1 paper (Whatman, Maidstone, UK). The collagen in the filtrate was precipitated by adding NaCl to a final concentration of 2.6 M in the presence of 0.05 M tris(hydroxymethyl)aminomethane (pH 7.0). The resulting sediment was collected by centrifugation at 20,000*g* for 40 min. Two centrifuge tubes of 250 ml and a GL-21M centrifuge (SAITE Xiangyi Centrifuge Instrment Co., Changsha, China) were used. Fifty millilitres of 0.5 M acetic acid were used to dissolve the sediment, then the solution was dialysed against 0.1 M acetic acid and then distilled water. The collagen was obtained by freeze-drying. The yield of pepsin-soluble collagen was calculated on the basis of skin weight after cleaning and expressed as a weight percentage.

2.4. Modelling of neural network

The multilayer feed-forward neural network has proven to be an excellent universal approximator of non-linear functions (Dreyfus & Dreyfus, 2003). In this work, a feed-forward neural network trained with an error back-propagation algorithm was employed using MATLAB (Version 6.5, Mathworks, Natick, MA) Neural Network Toolbox was used to model the yield of pepsin-soluble collagen, as a function of the independent variables used during the extraction. The input parameters chosen in this study were pepsin amount, reaction time and pH. Supervised learning was used to train this network. The predicted output and desired output were compared with one another while the errors were calculated between the predicted output and actual output. An error back-propagation algorithm was used for adjusting the network weights. It used a gradient descent approach, in which weights were changed in proportion to the negative value of the error gradient. The training iterations were stopped when the validation error reached a set minimum (Desai, Vaidya, Singhal, & Bhagwat, 2005).

2.5. Optimisation of extraction conditions

Genetic algorithms can solve well linear and non-linear problems, by exploring all regions of the state spaces and exploiting promising areas through selection, crossover and mutation operations, applied to individuals in the population (Shen, Wang, & Li, 2007). Once the feed-forward neural network model was completely trained, genetic algorithms could be used to determine the optimisation of extraction conditions for the highest yield.

3. Results and discussion

3.1. Effects of pepsin amount, reaction time and pH

The effects of pepsin amount, reaction time and pH on the yield of pepsin-soluble collagen as well as their interactions are shown in Figs. 1–3. Pepsin amount showed a similar effect to reaction time (Fig. 2). A positive correlation was found between the yield and the pepsin amount, and the yield and the reaction time. The yield of pepsin-soluble collagen increased very obviously with the increase of pepsin amount in the range of 20–40 unit/mg; a slight increase was observed when the pepsin amount increased in the range of 40–60 unit/mg. The pH value exhibited a different effect. A saddle-shaped profile is seen in Fig. 1 for pH and pepsin amount. The yield of pepsin-soluble collagen increased with pH increase to 3.0 and then decreased thereafter. The turning point of pH value was close to 3.0.

Collagen can be isolated by direct extraction with organic acids, like acetic, citric and lactic acids (Sadowska, Kołodziejska, & Niecikowska, 2003). Amongst these, acetic acid is the most popular, due to the high extractability. The solubility of collagen in the bulk solution plays a key role in the extractability from the grass carp skin. The covalent cross-linking through condensation of aldehyde groups at the telopeptide regions of collagen chains and the intermolecular cross-linked collagen was not readily solubilised by acetic acid (Jongjareonrak, Benjakul, Visessanguan, & Tanaka, 2005). The cross-linked telopeptide regions could be cleaved by pepsin without damaging the integrity of the triple helix of collagen (Nalinanon, Benjakul, Visessanguan, & Kishimura, 2007).

pH is an important parameter that affects the conformation and activity of pepsin. When the pH is above 7, pepsin is in an expanded conformation, devoid of well-defined tertiary interactions, with a reduced content of secondary structure. But it is still compact enough to have a significant buried tryptophan (Campos &

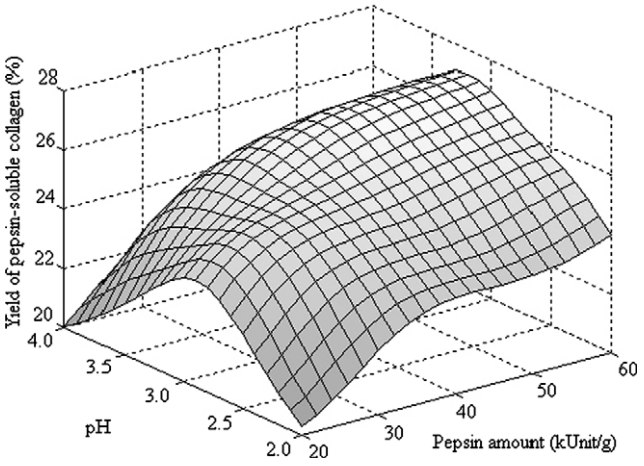

Fig. 1. Response surface plot showing the effects of pepsin amount and pH value, as well as their interactions, on the yield of pepsin-soluble collagen. The time was set at 24 h.

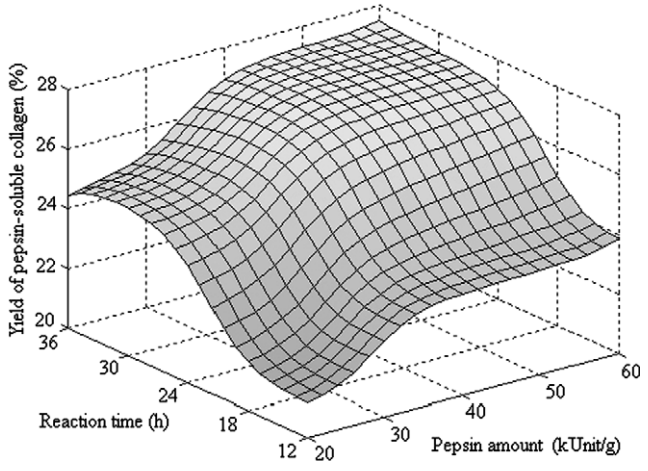

Fig. 2. Response surface plot showing the effects of pepsin amount and reaction time, as well as their interactions, on the yield of pepsin-soluble collagen. The pH was set at 3.0.

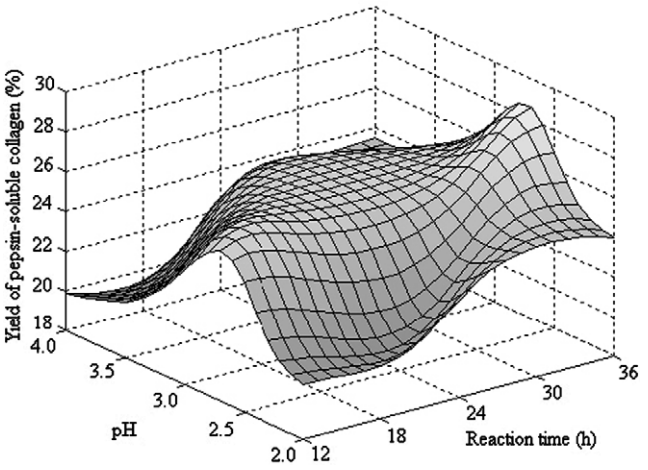

Fig. 3. Response surface plot showing the effects of pH and reaction time, as well as their interactions, on the yield of pepsin-soluble collagen. The pepsin amount was set at 40 kUnit/g.

Sancho, 2003). This conformation is enzymatically inactive. At the lower pH range of 2–4, pepsin exhibits an active conformation. However, from the results obtained, a significant difference in the yield of pepsin-soluble collagen was shown in this range. It indicated that there was an optimum pH for pepsin preparation of collagen. Higher or lower than this value would lead to a lower efficiency of extraction.

A positive correlation was found between reaction time and yield of collagen. Sadowska et al. (2003) have reported a similar result for isolation of collagen from Baltic cod skin. Mass transfer rates of analyte from the matrix plays a key role in the efficiency of extraction (Bartle, Boddington, Clifford, Cotton, & Dowle, 1991). In classical extraction, the mass transfer rate is controlled by the diffusion process, which is time-associated. In the diffusion-controlled process, the recovery of analyte will keep increasing with the extension of reaction time. This may explain the effect of reaction time in this work.

Collagen is a thermally instable protein that can be denatured at room temperature. This sensitivity to temperature is associated with its chemical structure. Collagens with lower levels of hydroxyproline show lower thermal stability than those with higher levels of hydroxyproline (Muyonga, Cole, & Duodu, 2004). This is due to the role of hydroxyproline in forming inter-chain linkages by hydrogen bond, that stabilise the triple-helix structure of collagen. Therefore, a low extraction temperature (e.g., 4 °C) was chosen to prepare collagen in this work, to avoid protein denaturation.

3.2. Evaluation of model predictability

The ratio of the explained variation to the total variation, r^2, reflects the degree of fit for the mathematical model (Nath & Chattopadhyay, 2007). The closer the r^2 value is to 1, the better the model fits to the actual data (Sin, Yusof, Hamid, & Rahman, 2006)

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

where n is the number of points, y_i is the predicted value obtained from the neural network model, y_{di} is the actual value, and y_m is the average of the actual values.

MSE is another important index to show the degree of fit of the model. It is calculated by Eq. (2). The network having minimum MSE and maximum r^2 is considered as the best neural network model (Izadifar & Jahromi, 2007)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{di})^2 \quad (2)$$

In this work the range of three independent variables for building the neural network was set. The input matrix and the yield of pepsin-soluble collagen are listed in Table 1. To understand the generalisation capacity of the network, 27 input values were divided into three sets, 21 values for the training set and three values each for the validation set and the testing set. Fig. 4 shows the plot

Table 1
Predicted values of the neural network and the actual values for the yield of pepsin-soluble collagen

Experiments	Pepsin amount (unit/mg)	Reaction time (h)	pH	Pepsin-soluble collagen yield (%)	
				Predicted values	Actual values
Training set					
1	20	12	2	18.4	18.5
2	60	12	2	21.3	21.6
3	20	12	3	21.4	21.0
4	40	12	3	23.1	24.1
5	60	12	3	24.9	24.6
6	40	12	4	20.4	20.4
7	60	12	4	21.6	21.1
8	40	24	2	22.9	23.5
9	60	24	2	24.2	23.8
10	20	24	3	23.5	23.4
11	60	24	3	26.7	26.9
12	20	24	4	20.8	20.1
13	60	24	4	23.1	23.7
14	20	36	2	20.8	20.9
15	40	36	2	24.6	24.1
16	60	36	2	24.9	24.7
17	20	36	3	23.5	23.8
18	40	36	3	27.4	27.0
19	60	36	3	27.6	27.8
20	20	36	4	20.1	20.3
21	60	36	4	24.6	24.2
Validation set					
22	40	12	2	21.6	21.2
23	20	12	4	18.2	17.7
24	40	24	4	22.5	23.0
Testing set					
25	40	36	4	22.9	23.4
26	40	24	3	27.1	26.3
27	20	24	2	19.6	20.2

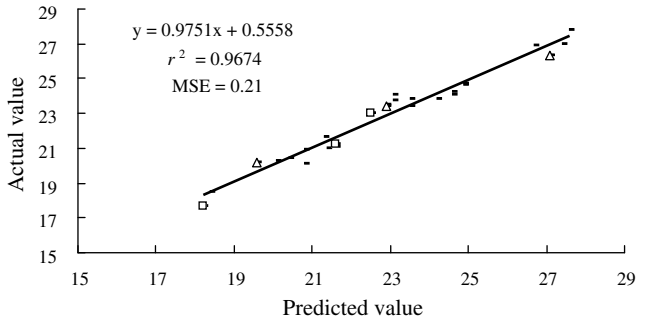

Fig. 4. Correlation between the predicted values of the neural network and the actual values for the yield of pepsin-soluble collagen. —, training set; □, validation set and Δ, testing set.

of the predicted values and actual values of the yield of pepsin-soluble collagen as well as r^2 and MSE. The trained network gave an r^2 of 0.97 and an MSE of 0.21. The r^2 values of the training set, the validation set and the testing set were 0.97, 0.95 and 0.93, while the MSE values of the three sets were 0.18, 0.22 and 0.42, respectively. The network could predict the yield of pepsin-soluble collagen within a range of ±4.1% of the actual value. The r^2, MSE and prediction range indicated a good agreement between the predicted value of the neural network model and the actual value, which also confirmed a good generalisation of the network.

By application of the artificial neural network-genetic algorithm, the optimum extraction conditions for the highest yield of pepsin-soluble collagen were determined to be pH 3.4, 53.3 unit/mg and 35.2 h. The predicted value of the yield was 30.3%. As the estimated optimal extraction conditions were used in the actual preparation of the pepsin-soluble collagen, the yield was measured experimentally to be 29.3 ± 0.8%, which was not significantly different ($p > 0.05$) to the predicted value.

4. Conclusions

Based on the results obtained, abundant collagen was present in grass carp skin, and pepsin was shown to be an efficient tool for its extraction. Significant effects on the yield of pepsin-soluble collagen were determined for pepsin amount, reaction time and pH. The artificial neural network provided a model to search for the non-linear nature between extraction conditions and yield in an efficient manner. The trained network gave an r^2 value of 0.97 and an MSE value of 0.21, which implied a good agreement between the predicted values and the actual values for the yield of pepsin-soluble collagen, and confirmed a good generalisation of the network. Based on the artificial neural network-genetic algorithm, the optimal extraction conditions to obtain the highest yield were determined to be pH 3.4, 53.3 unit/mg of pepsin and 35.2 h. The predicted value was 30.3%, not significantly different from the actual value (29.3 ± 0.8%). Furthermore, more specific chemical effects of pepsin on the structure of collagen need to be investigated in future work.

References

Antonov, V. K., Ginodman, L. M., Kapitannikov, Y. V., Barshevskaya, T. N., Gurova, A. G., & Rumsh, L. D. (1978). Mechanism of pepsin catalysis: General base catalysis by the active-site carboxylate ion. *FEBS Letters, 88*, 87–90.

Bartle, K. D., Boddington, T., Clifford, A. A., Cotton, N. J., & Dowle, C. J. (1991). Supercritical fluid extraction and chromatography for the determination of oligomers in poly(ethylene terephthalate) films. *Analytical Chemistry, 63*, 2371–2377.

Campos, L. A., & Sancho, J. (2003). The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH. *FEBS Letters, 538*, 89–95.

Desai, K. M., Vaidya, B. K., Singhal, R. S., & Bhagwat, S. S. (2005). Use of an artificial neural network in modeling yeast biomass and yield of β-glucan. *Process Biochemistry, 40*, 1617–1626.

Dreyfus, C., & Dreyfus, G. (2003). A machine learning approach to the estimation of the liquidus temperature of glass-forming oxide blends. *Journal of Non-crystalline Solids, 318*, 63–78.

Gelse, K., Pöschl, E., & Aigner, T. (2003). Collagens – Structure, function, and biosynthesis. *Advanced Drug Delivery Reviews, 55*, 1531–1546.

Harkness, R. D. (1966). *Collagen. Science Progress (Oxford), 54*, 257–274.

Izadifar, M., & Jahromi, M. Z. (2007). Application of genetic algorithm for optimization of vegetable oil hydrogenation process. *Journal of Food Engineering, 78*, 1–8.

Jongjareonrak, A., Benjakul, S., Visessanguan, W., & Tanaka, M. (2005). Isolation and characterization of collagen from bigeye snapper (*Priacanthus marcracanthus*) skin. *Journal of the Science of Food and Agriculture, 85*, 1203–1210.

Kittiphattanabawon, P., Benjakul, S., Visessanguan, W., Nagai, T., & Tanaka, M. (2005). Characterisation of acid-soluble collagen from skin and bone of bigeye snapper (*Priacanthus tayenus*). *Food Chemistry, 89*, 363–372.

Lee, C. H., Singla, A., & Lee, Y. (2001). Biomedical applications of collagen. *International Journal of Pharmaceuticals, 221*, 1–22.

Li, S. F., Lu, Q. Q., & Zhou, B. Y. (1995). Evaluation on the potential capacity of the swan oxbow for the conservation of the major Chinese carps. *Aquaculture, 137*, 46–47.

Muyonga, J. H., Cole, C. G. B., & Duodu, K. G. (2004). Characterisation of acid soluble collagen from skins of young and adult Nile perch (*Lates niloticus*). *Food Chemistry, 85*, 81–89.

Nalinanon, S., Benjakul, S., Visessanguan, W., & Kishimura, H. (2007). Use of pepsin for collagen extraction from the skin of bigeye snapper (*Priacanthus tayenus*). *Food Chemistry, 104*, 593–601.

Nath, A., & Chattopadhyay, P. K. (2007). Optimization of oven toasting for improving crispness and other quality attributes of ready to eat potato-soy snack using response surface methodology. *Journal of Food Engineering, 80*, 1282–1292.

Ramachandran, G. N., & Sasisekharan, V. (1965). Refinement of the structure of collagen. *Biochimica et Biophysica Acta – Protein Structure and Molecular Enzymology, 109*, 314–316.

Sadowska, M., Kołodziejska, I., & Niecikowska, C. (2003). Isolation of collagen from the skins of Baltic cod (*Gadus morhua*). *Food Chemistry, 81*, 257–262.

Shen, C., Wang, L., & Li, Q. (2007). Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method. *Journal of Materials Processing Technology, 183*, 412–418.

Sin, H. N., Yusof, S., Hamid, N. S. A., & Rahman, R. A. (2006). Optimization of enzymatic clarification of sapodilla juice using response surface methodology. *Journal of Food Engineering, 73*, 313–319.

Wasswa, J., Tang, J., Gu, X. H., & Yuan, X. Q. (2007). Influence of the extract of enzymatic hydrolysis on the functional properties of protein hydrolysate from grass carp (*Ctenopharyngodon idella*) skin. *Food Chemistry, 104*, 1698–1704.