

The extraction of fibroin protein from *Bombyx mori* silk cocoon: Optimization of process parameters

Mahesh Kumar Sah^{1*}, Arvind Kumar², Pramanik K.³

^{1,3} Department of Biotechnology and Medical Engineering, National Institute of Technology Rourkela-769 008 Orissa (INDIA), pearlusbiotech@gmail.com

² Department of Chemical Engineering, National Institute of Technology, Rourkela-769 008 Orissa (INDIA)

Abstract- Optimization of protein extraction using silk cocoon was carried out by the design of experiment (DOE) to obtain the response surface methodology (*RSM*). Box-Behnken rotatable design was used to create an experimental program to provide data to model the effects of various factors on protein extraction efficiency. The variables chosen were sodium carbonate concentration (X_1), Lithum bromide concentration (X_2) and temperature (X_3). The mathematical relationship between protein extraction efficiency and three significant independent variables can be approximated by a second order quadratic model. *RSM* was used to describe the individual and interactive effects of three variables at three levels, combined according to a Box-Behnken Design. The coefficient of determination (R^2) for the model is 0.9761. Probability value ($P < 0.0001$) demonstrates a very high significance for the regression model indicating that Box-Behnken Design can be applied to the protein extraction from silk cocoon, and it is an economical way of obtaining the maximum amount of information with the fewest experiments.

Keywords: *Bombyx mori*; Silk Fibroin; Box-Behnken Design; Response Surface Methodology; Design of Experiments

Introduction

Silk fiber produced by the silkworm *Bombyx mori* (*B. mori*) possesses excellent mechanical properties.[1,2] Silk protein also attracts the attention because of their valuable properties utilizable in biotechnological and biomedical areas.[3] Silk must be regenerated into a desirable form to meet a specific biomedical application.[4] Silk fibers are composed of two different proteins: sericin and fibroin.[5] The sericin, a natural gum can be removed by degumming process in order to dye the silk evenly.[6] Sericine contains 18 kinds of amino acids, including all 8 kinds of human need ones. Because the hydrophilic amino acids amount measures up to about 80%, the sericine can be removed from the fabric to the aqueous solution using even small amount of weaker nucleophiles eg. sodium bicarbonate.[7] The fibroin has highly oriented crystalline domain and it is insoluble. The fibroin content of naturally spun silk fibers can be separated from sericin *in vitro*.[8] Degummed silk fibers can be dissolved in water using a variety of chaotropic salts (like CaCl_2 , LiBr), leading to a transparent solution.[9] The solution was extensively dialyzed against deionized water to yield the regenerated silk fibroin (RSF) solution for further study. The objective of the present study is to investigate the

maximum protein yield at the optimum condition of process variables used in the present work, by using the 10 g silk cocoon as a starting material. The study uses the Box–Behnken design in the optimization of experiments using RSM to understand the effect of important process variables and their interactions on the protein purification process. The process variables used are the Na_2CO_3 concentration, LiBr concentration and temperature .

Materials and methods

Bombyx mori silk cocoons were obtained from silkworms reared in controlled conditions at the mulberry farms in Chittoor district, Hyderabad. Na_2CO_3 (S. D. Fine chemicals) and LiBr (>99% purity, Sigma Aldrich, Germany) were used as such without further purification. Dried *Bombyx mori* silk cocoons were cut into small pieces and then treated with boiling aqueous solution of 0.02M Na_2CO_3 for 20 minutes with stirring. The whole mass was washed with distilled water (Milli-Q water) repeatedly to remove the glue-like sericine protein and then dried in hot air oven. The removal of Sericine from the silk fibroin surface was confirmed by SEM study. The extracted silk fibroin solution was prepared by dissolving 10 g of degummed silk in 9.30M LiBr solution at 70°C for 2.5 h. The fibroin solution was dialyzed in a cellulose membrane based dialysis cassette (molecular cutoff 12,400.)

against deionized water for 3 days changing water every 6 hr in order to remove LiBr. After dialysis, concentrated solution was obtained by centrifuging silk fibroin solution at 5-10°C and 9000 rpm for 20 min. The concentrated solution was stored at 4 °C for further study. The sequential experimental procedure for protein extraction from *Bombyx mori* silk cocoon has been shown in Fig. (1).

The fibroin protein concentration was measured by the Bradford protein assay method.[10] The fibroin solution was added to the Bradford reagent and incubated at 30°C for 5 min and the absorbance at 595nm was measured. Bovine serum albumin (Sigma-Aldrich) was used as a standard protein.

RSM is used in the empirical study of relationships between one or more measured response and a number of input variables and the objective is to optimize this response. It offers solutions to critical questions such as how particular response is affected by a given set of input variables over some specified region of interest, what settings of factors will give a product simultaneously satisfying desired specifications, and what values (in range) of the factors will yield a maximum for a specific response.[11]

The design procedure for RSM is as follows [12]:

- Performing a series of experiments for adequate and reliable measurement of the response of interest.
- Developing a mathematical model of the second-order response surface with the best fit.
- Determining the optimal set of experimental parameters that produce a maximum or minimum value of response.
- Selecting the best models by conducting sequential *F*-test, lack-of-fit test and other adequacy.
- Representing the direct and interactive effects of process parameters through two and three-dimensional (3-D) plots.

To analyze a process or system including a response *Y*, where *Y* depends on the input factors x_1, x_2, \dots, x_k , the relationship between the response and the input process parameters can be described as:

$$Y = f(x_1, x_2, \dots, x_k) + \varepsilon$$

(1)

Where, *f* is the real response function with unknown format, and ε is the residual error, which describes the differentiation that can be included by the function *f*. Since the relationship between the response and the input parameters can be described as a surface of the x_1, x_2, \dots, x_k ordinates in the graphical sense, the research

into these relationships is known as a response surface study [13]. Box-Behnken rotatable design (BBRD) for three independent variables each at three levels was adopted in this study. The process variables used are the Na₂CO₃ (*X*₁), LiBr (*X*₂) concentrations and temperature (*X*₃). A total of 17 experiments were necessary for the estimation of the various coefficients of the model. Starting from the matrix of the Table-1 a matrix of 17 experiments was obtained to evaluate the surface of optimization. These experiments were carried out in the laboratory to compare the experimental values and those obtained with the model. The quadratic response surface model was fitted to the following equation [14, 15]

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{j=1}^k \beta_{ij} X_i X_j + \varepsilon \quad (2)$$

Where, X_1, X_2, \dots, X_k are the input variables, which influence on the response *Y*;

$$\beta_i (i = 1, 2, \dots, k),$$

$$\beta_{ij} (i = 1, 2, \dots, k; j = 1, 2, \dots, k) \quad \text{the}$$

unknown parameters and ε is a random error.[16] Actual values of independent variables in Box-Behnken design for the protein extraction using silk cocoon is shown in Table 1.

It is important to include the second order model to provide good prediction throughout the region of interest. The second order response surface design is rotatable; this means that the variance of the predicted response is the same at all points. Rotatability is a reasonable basis for the selection of response surface design. Because the purpose of response surface methodology (*RSM*) is optimization and as the location of the optimum is unknown prior to running the experiment, it makes sense to use design that provides equal precisions of estimation in all directions [17]. A three-dimensional response surface and contour plots of the independent variables and their interactions were generated using the statistical software, Design Expert software version 6.0.6 (STAT-EASE Inc., Minneapolis, US). The optimization of protein extraction process was aimed to establish the maxima levels within the independent variables such as Na₂CO₃ concentration (*X*₁), LiBr concentration (*X*₂) and temperature (*X*₃).

Results and discussion

Response surface optimization is more advantageous than the traditional single parameter optimization in that it saves time, space and raw material. All 17 of the designed experiments were conducted for optimizing the three individual parameters in the current Box–Behnken design. The results were analyzed by multiple regression analysis. Table 2 shows the experimental conditions and the results of extracted protein yield according to the factorial design. Maximum yield of protein (2.10 mg/ml) was recorded under the experimental conditions of Na_2CO_3 concentration 0.01 M, $LiBr$

concentration 9.10 M, and Temperature $70^\circ C$. SEM performed for each combination of process conditions shows the extent of removal of sericine from the silk fibroin surface. Fig. (2) depicts how actually sericine removal proceeds during degumming. To decide about the adequacy of the model for protein yield by silk cocoon, three different tests such as the Sequential model Sum of Squares, Lack of Fit Tests and Model Summary Statistics were carried out and p values for all the regressions were lower than 0.01 (see Table 3) and the quadratic model. This means that at least one of the terms in the regression equation had a significant correlation with the response variable. The model summary statistic showed the regression coefficient is found to be highest ($R^2 = 0.9761$) for the quadratic model with minimum standard deviation (0.041). ANOVA analysis confirmed that the form of the model chosen to explain the relationship between the factors and the response is correct [18]. In order to determine whether or not the quadratic model is significant, it is necessary to conduct ANOVA analysis. The P -values were used as a tool to check the significance of each coefficient, which also indicated the interaction strength of each parameter. The smaller the P -values are, the bigger the significance of the corresponding coefficient. [11] Here, the P -value of the model was smaller than 0.0001, which indicated that the model was suitable for use in this experiment. The P -value of “lack of fit” was 0.64 ($P > 0.01$), which indicated that “lack of fit” was insignificant relative to the pure error. The coefficient of determination (R^2) and adjust coefficient of determination ($Adj.R^2$) were also shown in Table 4. The values indicated that the accuracy and general availability of the polynomial model were adequate [18]. The regression coefficients and the corresponding P -values were presented in Table 5. From the P -values of each model term, it could be

concluded that three linear coefficients (X_1, X_2, X_3), three quadratic coefficients (X_1^2, X_2^2, X_3^2), and three cross-product coefficients (X_1X_3) were significant. Using the designed experimental data (Table 2), the polynomial model for the protein extraction was regressed and shown as below (in term of actual factors):

$$Y = 1.37 + 0.0001X_1 - 0.0001X_2 + 0.0001X_3 + 0.0001X_1^2 - 0.0001X_2^2 + 0.0001X_3^2 + 0.0001X_1X_3 \quad (3)$$

Many parameters can influence the protein extraction concentration. Eq. (3) shows that the yield of protein extraction has a complex relationship with independent variables that encompass both first and second-order polynomials and may have more than one maximum point. Fig. (3) represents the effects of Na_2CO_3 concentration, temperature and their reciprocal interactions on protein yield. An increase of temperature increased the degumming protein yield at a constant Na_2CO_3 concentration. The appropriate maximal protein yield was determined at a temperature of $85^\circ C$. Fig. (4) shows the effects of $LiBr$ concentration, temperature and their reciprocal interactions on protein yield. At a high temperature, the yield of protein increased with temperature increasing. The fragments of fibroin macromolecules are nonpolar hydrophobic aliphatic hydrocarbon (alanine, leucine, isoleucine, valine, proline) and aromatic (phenylalanine) substituents, polar hydrophilic hydroxyl- containing residues of serine, threonine, and tyrosine, carboxy groups of aspartic and glutamic. The polar carboxy and amino groups in fibroin can also be involved in dipole interactions, and in the case of proton transfer, also in electrostatic interactions.[19] As the content of acid and base groups in fibroin is low, the electrostatic factor is not decisive in the formation of the secondary structure; it, however, may become decisive in the fibroin dissolution. Since fibroin consists to 3/4 of nonpolar hydrophobic amino acids, it is necessary to take into account also hydrophobic (dispersion) interactions making fibroin resistant to the majority of solvents. The particular feature associated with dissolution of fibroin macromolecule in various systems is its molecular structure. Dissolution in salt systems is due to interaction of solvent ions with functional groups of fibroin macromolecules. The rupture of inter and intramolecular hydrogen bonds results due to nucleophilic attack by the anion.[20] Thus the amino acid composition of fibroin precipitated

from the salt system differs from that of initial fibres.[21] The group reported the change in amino acid composition of LiBr system in dissolution of *B. mori*. Silk. They found that irrespective of dissolution temperature, the relative content of glycine and alanine residues characteristics of the hydrophobic areas increased while content of the residue of all in amorphous areas decreased after the precipitation of polymer.

Validation of the model

The optimal values of the selected variables were obtained by solving the regression equation (Eq. (3)) using Design-Expert 6 software. The optimal conditions for protein yield estimated by the model equation were as follows: $X_1 = 0.01$ M, $X_2 = 9.23$ M and $X_3 = 70^\circ$ C. The theoretical protein yield (Y) predicted under the above conditions was 2.09 mg/ml. In order to verify the prediction of the model, the optimal reaction conditions were applied to three independent replicates for protein yield. The average protein yield was 2.1 ± 0.01 mg/ml, a figure well within the estimated value of the model equation. This demonstrated the validation of the *RSM* model. The good correlation between these results confirmed that the response model was adequate for reflecting the expected optimization. The results also suggested that the models of Eq. (3) are satisfactory and accurate.

Conclusion

Optimization of the process parameters for maximal protein was carried out. RSM was used to optimize the extraction of maximum protein yield. The coefficient of determination (R^2) for the model is 97.61 %. Probability value ($p < .0001$) demonstrates a very high significance for the regression model. Maximum protein yield of 2.1 ± 0.01 mg/ml from silk cocoon was obtained at the set optimum conditions of process variables. Validation experiments verified the availability and the accuracy of the model. The predicted value was in agreement with the experimental value. The study proved the response surface method to be useful for optimization of process parameters for protein extraction using silk cocoon and statistical analysis is proved to be a useful and powerful tool in developing optimal production conditions.

References

- [1] Gosline J.M., Guerette P.A., Ortlepp C.S., Savage K.N. (1999) *Journal of Experimental Biology*, 202,3295–3303.
- [2] Vollrath F., Knight D.P. (2001) *Nature*, 410,541–548.
- [3] Zhu Z.H., Ohgo K., Asakura T. (2008) *eXPRESS Polymer Letter*, 2(12), 885–889.
- [4] Nagarkar S., Patil A., Lele A., Bhat S., Bellare J., Mashelkar R.A. (2009) *Doi: 10.1021/ie801723f*
- [5] Perez-Rigueiro J., Viney C., Llorca J., Elices M. (1998) *Journal of Applied Polymer Science*, 70, 2439–2447.
- [6] Gulrajini M.L., Gupta S. (1989) *Silk dyeing, printing and finishing*.
- [7] Kadhar K.M.A., Hariraj G., Roy S., Somashekar T.H. (2008) *Journal of Silk Science & Technology Japan* 17,43-49.
- [8] Rui H.G. (1998) *Silk Reeling, Oxford & IBH Publishing Co Pvt Ltd, New Delhi*.
- [9] Ayub Z.H., Arai M., Hirabayashi K. (1993) *Bioscience, Biotechnology and Biochemistry*, 57,1910–1912.
- [10] Bradford M. (1976) *Anal Biochemistry*, 72,248–254.
- [11] Zhao H., Wang J., Lu Z. (2009) *Carbohydrate Polymer*, 77,677–680.
- [12] Aslan N. (2008) *Powder Technology*, 185,80–86.
- [13] Huiping L., Guoqun Z., Shanting N., Yiguo L. (2007) *Composite Material Science* 38(3),561–570.
- [14] Evans M. (2003) *Optimization of Manufacturing Processes: A Response Surface Approach, Carlton House Terrace, London*.
- [15] Guaracho V.V., Kaminari N.M.S., Ponte M.J.J.S., Ponte H.A. (2009) *Doi:10.1016/j.jhazmat. 2009. 07. 100*.
- [16] Guvenc A., Kapucu N., Kapucu H., Aydogan O., Mehmetoglu U. (2007) *Enzyme and Microbial Technology* 40,778–785.
- [17] Kumar A., Prasad B., Mishra I.M. (2008) *Journal of Environmental protection Science* 2, 54 – 62.
- [18] Tarangini K., Kumar A., Satpathy G.R., Sangal V.K. (2009) *Clean* 37(4 – 5), 319 –327.
- [19] Dill K.A. (1990) *Biochemistry* 29(31), 7133 - 7155.
- [20] Dawsey T.R. , McCormick C.L. (1990) *Macromolecular Chemistry and Physics* 30(3),405- 440.
- [21] Tsukada M., Goto Y., Minoura S. (1990) *Journal of Sericulture Science Japan* 59(5), 325-330.

Table 1- Experimental range and levels of independent variables

X ₁	Range and levels		
	-1	0	+1
X ₁ : Na ₂ CO ₃ concentration (mg/l)	0.01	0.02	0.03
X ₂ : LiBr concentration (mg/l)	9.10	9.30	9.50
X ₃ : Temperature (°C)	55	70	85
Protein yield (mg/ml)			

Table 2- Selection of a satisfactory model for protein extraction

Sequential Model Sum of Squares						
Source	Sum of Squares	df	Mean Square	F	p	Remark
Mean	49.2	1	49.2	-	-	-
Linear	0.36	3	0.12	11.56	0.0006	-
2FI	0.019	3	6.22E-03	0.53	0.6735	-
Quadratic	0.11	3	0.035	20.6	0.0008	suggested
Cubic	3.88E-03	3	1.29E-03	0.64	0.6299	Aliased
Residual	8.12E-03	4	2.03E-03	-	-	-
Total	49.7	17	2.92	-	-	-
Lack of Fit Tests						
Source	Sum of Squares	df	Mean Square	F	p	Remark
Linear	0.13	9	0.014	7.03	0.038	-
2FI	0.11	6	0.018	9.01	0.0259	-
Quadratic	3.88E-03	3	1.29E-03	0.64	0.6299	suggested
Cubic	0	0	-	-	-	Aliased
Pure error	8.12E-03	4	2.03E-03	-	-	-
Model Summary Statistics						
Source	Std. Dev.	R2	Adj. R2	Pre. R2	PRESS	Remark
Linear	0.1	0.7274	0.6645	0.4501	0.28	-
2FI	0.11	0.7647	0.6235	-0.1233	0.56	-
Quadratic	0.041	0.9761	0.9453	0.8509	0.075	suggested
Cubic	0.045	0.9838	0.9352	-	-	Aliased

Table 3- Box–Behnken design matrix of the three variables in coded units along with the experimental and predicted values of protein yield.

Trial no	X ₁	X ₂	X ₃	Protein yield (mg/ml)	
				Experimental	Predicted
1	0.01	9.10	70	1.89	1.92
2	0.01	9.50	70	1.84	1.85
3	0.03	9.10	70	1.71	1.70
4	0.03	9.50	70	1.65	1.62
5	0.02	9.10	55	1.45	1.45
6	0.02	9.50	55	1.30	1.32
7	0.02	9.10	85	1.76	1.74
8	0.02	9.50	85	1.74	1.74
9	0.01	9.30	55	1.64	1.61
10	0.03	9.30	55	1.50	1.51
11	0.01	9.30	85	2.1	2.09
12	0.03	9.30	85	1.72	1.75
13	0.02	9.30	70	1.70	1.72
14	0.02	9.30	70	1.69	1.72
15	0.02	9.30	70	1.73	1.72
16	0.02	9.30	70	1.80	1.72
17	0.02	9.30	70	1.70	1.72

Table 4- Analysis of variance for the fitted quadratic polynomial model of protein yield

Source	Sum of squares	df	Mean square	F-value	Probability (P)>F
Model	0.49	9	0.054	31.71	<0.0001
Lack of fit	0.0039	3	0.0013	0.64	0.6299
Pure error	0.0082	4	0.002		
Correlation total	0.50	16	-	-	-

R²=0.9761; Adj R²=0.9453

Table 5- Regression coefficients and their significance of the quadratic model of protein yield

Model term	Coefficient estimate	df	Standard error	Probability (P)>F
Intercept	1.72	1	0.019	-
X ₁	-0.11	1	0.015	0.0001
X ₂	-0.35	1	0.015	0.0481
X ₃	0.18	1	0.015	<0.0001
X ²	0.11	1	0.020	0.0008
X ² ₂	-0.065	1	0.020	0.0151
X ² ₃	-0.097	1	0.020	0.0019
X ₁ X ₃	-0.060	1	0.021	0.0230

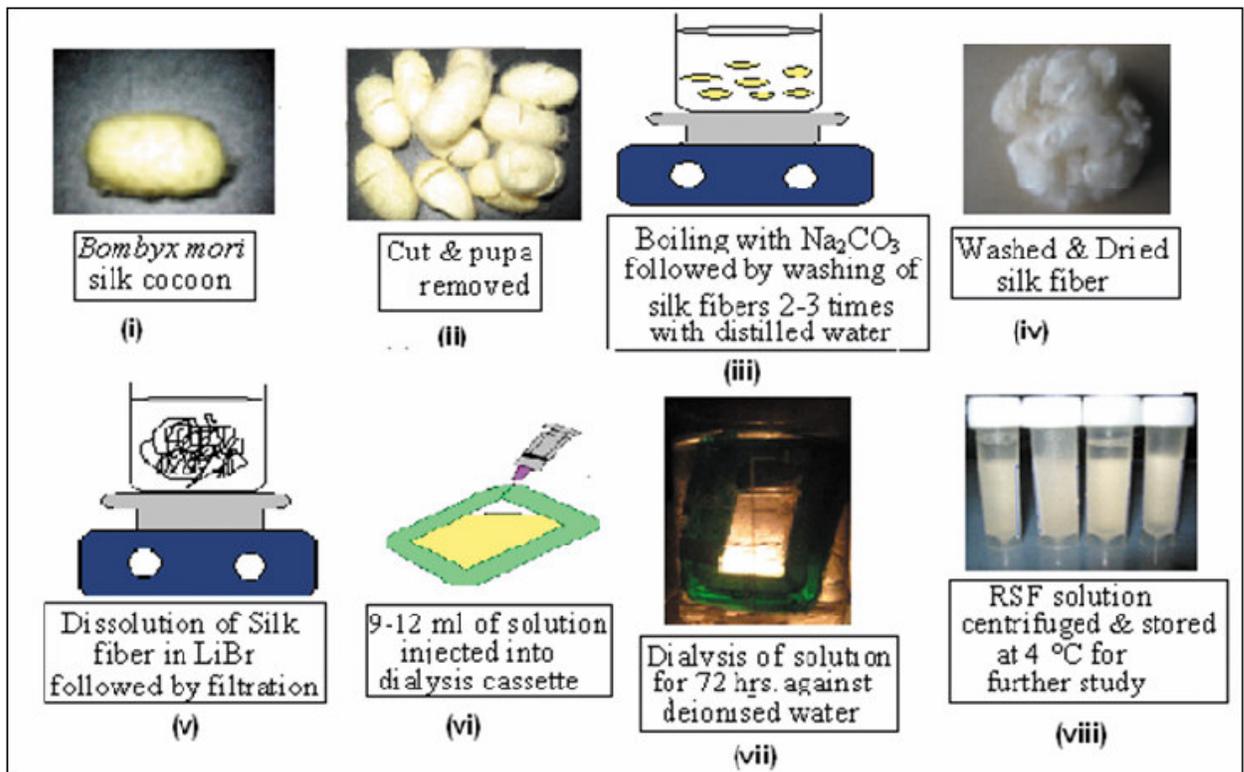


Fig. 1- Sequential experimental procedure for protein extraction from *Bombyx mori* silk cocoon.

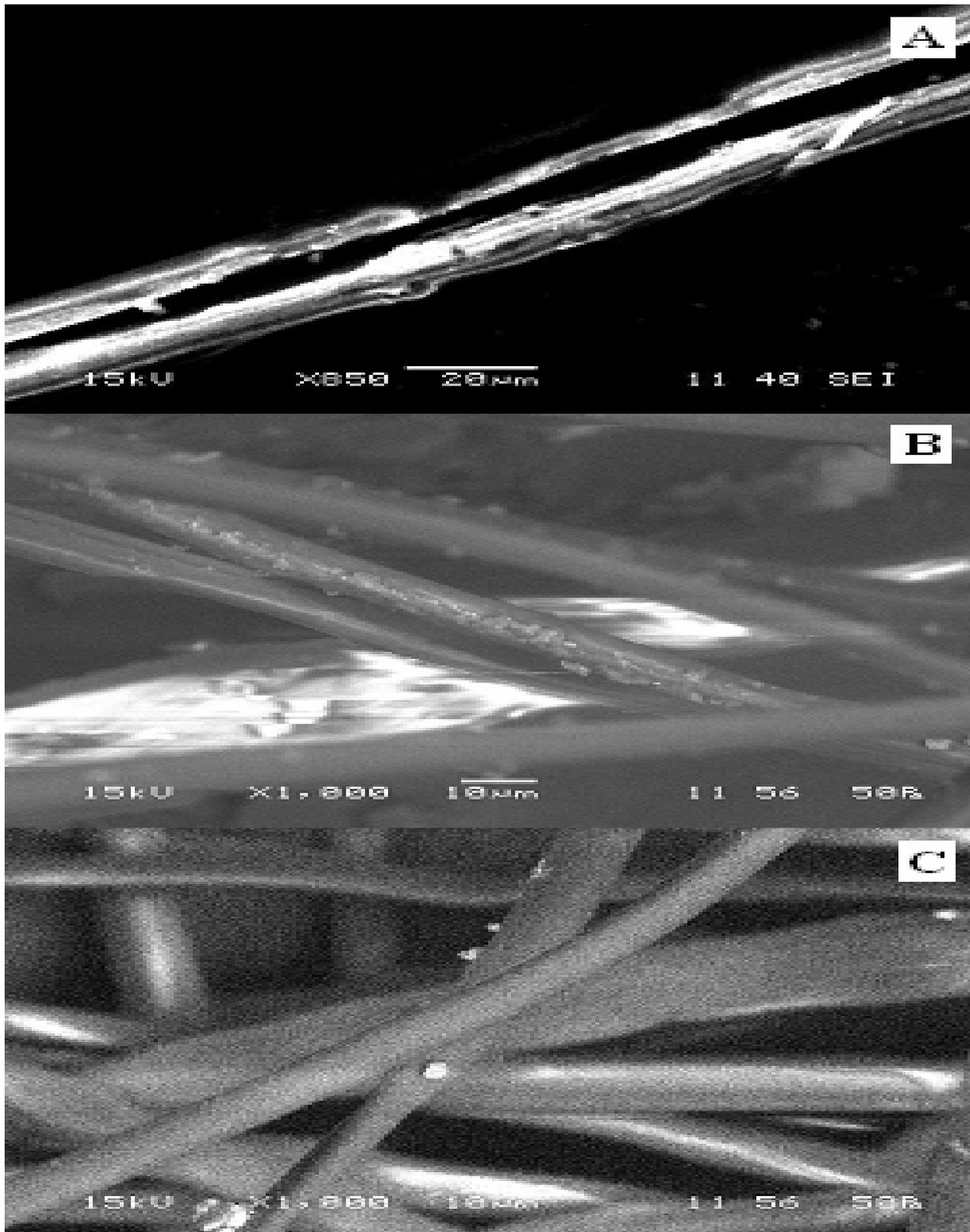


Fig. 2- Scanning electron micrographs of Native silk fibre (A) Incompletely degummed (B) and Completely degummed (C) silk fibre from *Bombyx mori* silk cocoon.

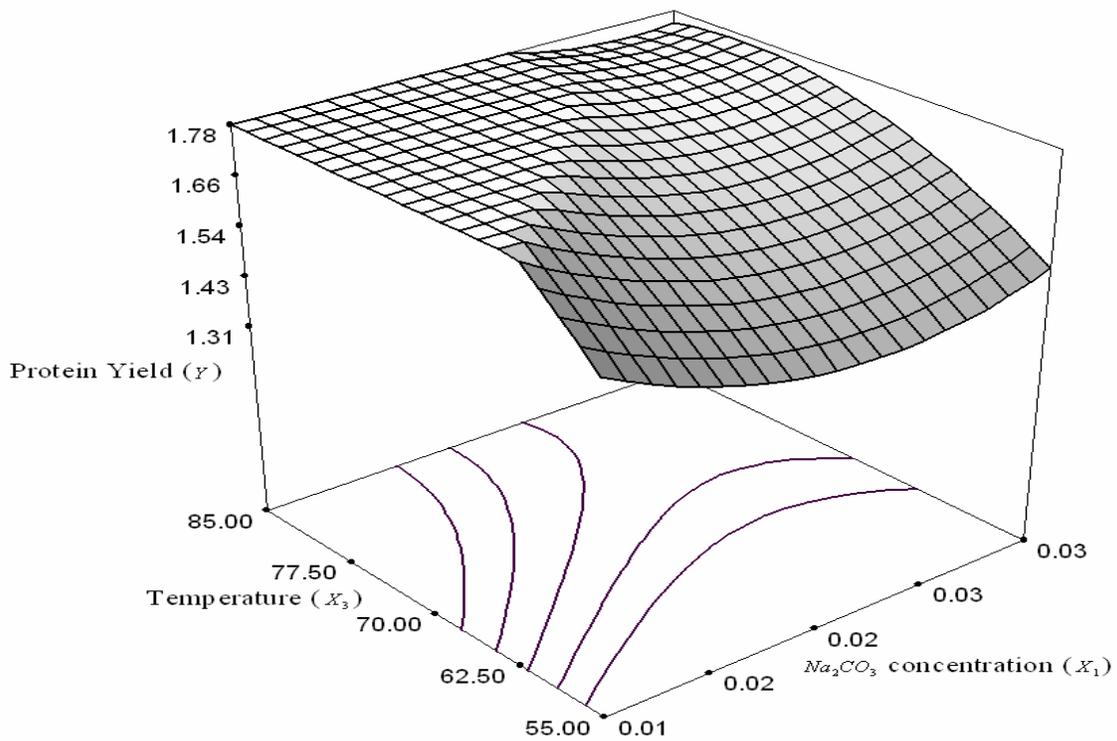


Fig. 3- Response surface plots representing the effect of Na_2CO_3 concentration and temperature, and their reciprocal interaction on protein yield.

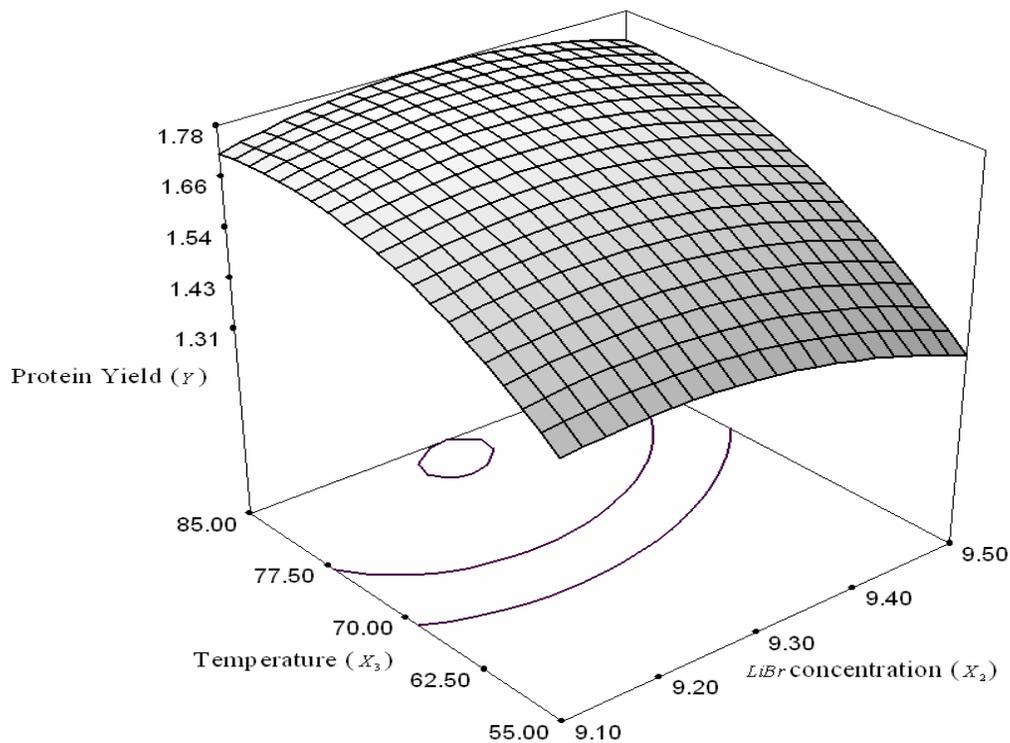


Fig. 4- Response surface plots representing the effect of LiBr concentration and temperature, and their reciprocal interaction on protein yield.