Original Article

Estimation of the amount of recombinant protein A secretion using Fuzzy regression

***Garshasb Rigi**[1] Tahereh Bahrami[2] Raham Armand[1] Zahra Piruzeh[2]

1- Assistant Professor, Department of Biology, Faculty of Science, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran
2-Department of Statistics, Faculty of Science, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran

***garshasbbiotech@gmail.com**

(Received: 6 May 2016; Revised: 27 August 2016; Accepted: 10 September 2016)

Abstract

***Background and purpose*:** Since protein A is considered an important protein from medical, medicinal, genetic engineering, and biotechnology point of view, the present study attempted to investigate and determine to what extent protein A is produced through regression, in addition to the production conditions of the protein.. Thus, a figure was introduced as for the estimation of the amount of protein A.

Methods: With the introduction of fuzzy mathematics and its combination with statistical methods, the kinds of regression models for estimating the amount of unknown variables were introduced. The utilization of fuzzy regression was developed from 1982 through the introduction of regression models, and the fuzzy data was based on a kind of linear plan. One of these regression models is fuzzy regression which considers the features of fuzzy numbers and the estimation obtained through them, and it has a higher level of reliability.

Results: In the present study, fuzzy regression method was introduced, and the number usage of this model in estimating the amount of secretion of protein A was investigated. It was then confirmed that this estimation method had a higher level of reliability. The type of regression used in this article was fuzzy regression that had a higher confidence level than the point classic regression. At the same time, the number of triangular fuzzy number was used in the current research in terms of computational handling, and it was found that triangular fuzzy number was much easier to use in comparison with the other species.

Conclusion: Secretion and extracellular production of recombinant protein is a wide production method which is currently developing. In the present study, it was observed that the statistical methods for improving the process in medical biotechnology are ideal methods. It was also documented that for laboratory designing of this important protein and achieving the best and most improved conditions for production and secretion, its amount of production must first be calculated through statistical methods.

***Keywords**:* Protein A; Extracellular; Fuzzy groups; Triangular Fuzzy Numbers; Fuzzy Regression

1. Introduction

Staphylococcal protein A (SPA) is a part of the cellular level of *Staphylococcus aureus* bacteria (1, 2). Protein A interacts with antibodies through two discrete connecting events : "classical" connecting place on FC (Fragment crystallizable) part of Immunoglobulines of IgG1 , IgG2 , and IgG4 of humans and "sub connection" place which is found in fab (Fragment antigen-binding) parts or connecting parts to antibody in IgG , IgM , IgA and IgE of humans and contains the heavy chains of VH3 sub-group. Protein A has long been proposed as a valuable immunologic tool, because of its features (3-8). It is a very important substance in biotechnology, medicine, and genetic engineering. Protein A is one of 30 chemicals – biological orders that scientists utilize in chemical materials and zymotic food stuff and the places for cellular culture. Protein A can be obtained through the culture of wild type *S. aureus* bacteria in a vast scale and then through making slippery the cellular suspensions or from the supernatant culture of *S. aureus* mutant bacteria (Secretory types). In both ways, the deliverances are usually done through utilizing chromatography of combining tendency on the stabilized IgG. About 7% (weight) of Cowan I type *S .aureus* which is a suitable productive type formed by protein A, is almost equivalent to 1/7 % of the whole protein. Since working with a high volume of *S.aureus* bacterial culture is not suitable due to producing illness, for generating big amounts of protein A in safe bacterial hosts, the molecular method must be utilized (9-11). In the previous studies, recombinant protein A was produced in laboratory conditions in *E.coli* bacteria as a host and a number of laboratory conditions were provided in order to increase the amount of production and secretion (12-14). For laboratory design of this important protein and achieving the best and the most improved conditions for production and secretion, first the amount of its production must be calculated through a statistical method. The statistical test utilized in the current study was fuzzy regression (15, 16). Fuzzy or practical data have many usages in different fields such as reliability, marketing, quality control, identification of pictures, and so on. One problem of fuzzy or vague data is the subject of personal deduction or variable identification language (like words such as sufficiently, good, enough, from which different deductions are obtained). In fact, many cases are seen that their meaning cannot be rightly identified, so only a personal deduction from them could be made, and a series of approximate descriptions or a group of concepts from them could be found out. Another problem of fuzzy or vague data is regarding their way of calculation. It means that in fact doing the calculations and algebraic works on these data is not in the domain of normal mathematics. Through the emergence of fuzzy mathematics and the features of fuzzy numbers, the problem of calculation related to these series of figures was also resolved (17). In analyzing fuzzy data, an investigation shows a relationship between fuzzy data and the equivalent (non-fuzzy) related variables and one or more independent fuzzy. Also, an investigation can, for example, be through fuzzy linear regression technique with these semi-widths. In this paper, in addition to introducing the fuzzy regression model, its

application in estimating the amount of protein A is investigated. The estimation obtained through this regression method is distance estimation which has a very higher level of reliability in comparison with spot estimation.

The procedure of the present study was as follows: First, the fuzzy groups and fuzzy numbers utilized in the current research were described (triangular fuzzy number). Then, there was a description of fuzzy regression, and finally the amount of protein A obtained from a laboratory sample was estimate (14).

2. Methods

2.1.Fuzzy Regression

In cases in which we deal with the two-value world, classic mathematics is a suitable tool for explaining different concepts. In the two-value world, the value of each statement is shown by two true or untrue values and through the mathematics language with {0,1}. But following the scientific and technological advancements and human mind growth, it is now essential for humans to find more suitable scientific tools to explain the complicated concepts of life and human atmosphere. Hence, the concepts which are based on two-value criteria shown through normal mathematics are no longer so much suitable or possible (18).

It is then observed that fuzzy mathematics is suitable for the type of needs which are for expressing multi-value concepts instead of two-value concepts, as well as the needs for expressing the world facts the way they are, instead of expressing them in the frames in which they are not included .

2.2.Membership function and degree of function

Consider the set of X. Suppose that Ã is a sub-group of X-Fuzzy if

Ã: X→ [0, 1] (1)

Be a function.

The amount which is related to X in this relation is called X membership degree. Ã fuzzy set can also be shown as Ã = {(x, Ã (x) | x ϵ X)}.

2.3. Definition of triangular fuzzy number

Triangular fuzzy number is displayed as triple for (a, r_a, l_a) , so that the first component is called the center of the triangular fuzzy number and the components of the second and third panels are half right and left triangular fuzzy number.

2.4.Membership Function

Membership function of triangular fuzzy number is specified as the following (16)

$$\mu \tilde{A}(x) \_ \begin{cases} \frac{x-a+l\,a}{la} \\ \frac{x-a+ra}{ra} \\ 0 \qquad else \end{cases} \qquad (2)$$

In which l_a is called a left semi-width and r_a is called the right semi-width. As it is deduced from the name of triangular fuzzy number, the diagram of these figures is in the shape of a triangle. For example, triangular fuzzy number for $l_a = 3$, a = 2 and $r_a = 1$ will be like the following:

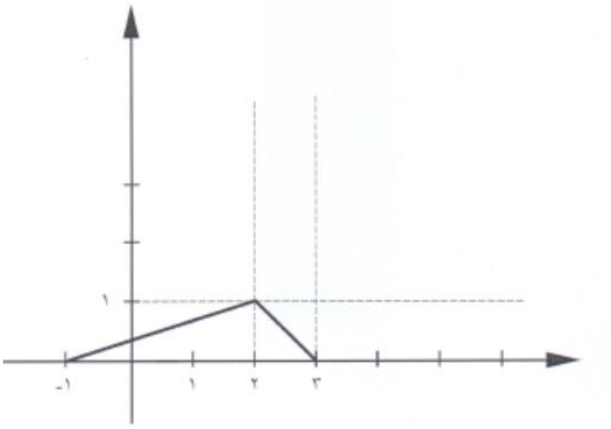

Figure 1: Diagram of triangular fuzzy number of example 3 – 2.

In regression analysis, the task is to address the dependence on one or more variables, this means that one or more independent variables exist in the regression model according to which, the dependent variable is predicted.

In general, the **classic regression** equation between more than two variables can be written as follows:

$$Y = A_0 + A_1x_1 + A_2x_2 + \cdots + A_kx_k$$

In this equation, $x_1, x_2, \dots, x_k$ are Independent variables and $A_0, A_1, \dots, A_K$ are multiple regression coefficients, and are the fixed numbers which are appointed using observed data and independent data. Y is the dependent variable that is anticipated using observed data and independent variables.

Utilization of fuzzy regression model was developed from 1982 with the introduction of regression models and fuzzy data based on a kind of linear planning (20, 21). Like normal regression, in fuzzy regression, researcher also deals with the estimation of one related variable according to one or several independent variables (20, 21).

A fuzzy regression model can be shown as the following (22):

$$Y_i = A_0 + A_1x_{i1} + A_2x_{i2} + \cdots + A_kx_{ip} \quad i = 1,2,..,n \qquad (1)$$

In this formula, x_{ij} are real numbers , $Y_i = [C_i - S_i\ , C_i + S_i]$ is the related variable of triangular fuzzy number , C_i is the center of the triangular fuzzy number , S_i is the spread , and $A_m = [a_m - r_m, a_m + r_m]$ are the fuzzy regression parameters which are all triangular fuzzy numbers (21). It must also be noted that the triangular fuzzy number used in this study is the symmetrical triangular fuzzy number, which means that left and right semi-width of the fuzzy number are equal.

We may treat $y_{Li} = C_i - S_i$ and $y_{Ri} = C_i + S_i$ as the left and right end point of the sample data, respectively. For the left end data points $\{(\ y_{Li}, x_{i1}, x_{i2}, \dots, x_{ip}) | i = 1,2,..,n\}$ and similarly to the right end data points $\{(\ y_{Ri}, x_{i1}, x_{i2}, \dots, x_{ip}) | i = 1,2,\dots,n\}$ we may use the linear regression model $y = \beta_0 + \beta_1x_1 + \cdots + \beta_px_p$ to obtain the following estimates, respectively;

$$\hat{y}_{Li} = \hat{L}_0 + \hat{L}_1x_{i1} + \hat{L}_2x_{i2} + \cdots + \hat{L}_px_{ip}\ , i = 1,2,..,n$$

$$\hat{y}_{Ri} = \hat{R}_0 + \hat{R}_1x_{i1} + \hat{R}_2x_{i2} + \cdots + \hat{R}_px_{ip}\ , i = 1,2,..,n$$

Then, $\hat{A}_m = [\hat{a}_m - \hat{r}_m, \hat{a}_m + \hat{r}_m]$ where

$\hat{a}_m = \frac{\hat{L}_m + \hat{R}_m}{2}$, $\hat{r}_m = \frac{|\hat{R}_m + \hat{L}_m|}{2}$

Based on the definition (1), we can use ordinary least-squares method to estimate the fuzzy

parameters in the general fuzzy regression model. Assuming that $y_i = (C_i, S_i)$ and $A_m = (a_m, r_m)$ have the same membership function, after appropriate translation, we can make all of $x_{ij} > 0$. Then (1) can be expressed as

$$(C_i, S_i) = (a_0, r_0) + (a_1, r_1)x_{i1} + (a_2, r_2)x_{i2} + \cdots + (a_p, r_p)x_{ip}$$

According to the Euclidean distance formula of (2)

$$D = \sqrt{(m_a - m_b)^2 + (\alpha_a - \alpha_b)^2} \quad \text{where } A = (m_a, \alpha_a), B = (m_b, \alpha_b) \quad (2)$$

The least-square estimates of a_i and r_i are the values of a_i, r_i which minimize the value of D^2 Where

$$D^2 = \sum_{i=1}^{n}[(C_i - (a_0 + a_1x_{i1} + \cdots + a_px_{ip}))^2 + (s_i - (r_0 + r_1x_{i1} + \cdots + r_px_{ip}))^2]$$

Let $\|\vec{v}\|$ denote the length of vector, that by using vectors and matrix expression, D^2 can be rewritten as $D^2 = ||Xa - C||' + ||Xa - S||'$ where X is a n× $(p + 1)$ design matrix,

$$a = (a_0, a_1, \dots, a_p)', r = (r_0, r_1, \dots, r_p)', C = (C_0, C_1, \dots, C_p)', S = (S_0, S_1, \dots, S_p)'$$

Let $\frac{\partial D^2}{\partial a} = 0$ and $\frac{\partial D^2}{\partial r} = 0$ then the solution of a and r which minimize D^2 are as follows:

$$\hat{a} = (X'X)^{-1}X'C \quad , \quad \hat{r} = (X'X)^{-1}X'S \quad 3)$$

The above method used regression with respect to center and spread.

2.5. Utilization of fuzzy regression number in the estimation of the amount of protein A secretion

In this part of the study, the effects of five variables (lactose concentration, glycine concentration, and bacterial OD at the time of induction, temperature, and the time after induction) on the expression performance and exploiting the form of the recombinant protein A and extracellular maturity of bacteria were all documented. These bacteria maturity consists of all mixtures from five variables through utilizing Central Composite Design (CCD) statistical plan and the software of Design-Expert V. USA, MN, Minneapolis, Inc., Stat-Ease, 6, 0, 8 in the laboratory conditions (14). The amounts of independent variables and the levels of their utilization in CCD as well as the amounts of protein secretion obtained in laboratory are explained in table 1.

Table 1. Variables showing the observed amounts of secreted recombinant protein A

Run	Actual response (mg/ml)	Predicted response (mg/ml)
1	25.342	25.090
2	13.351	13.099
3	19.173	18.921
4	28.597	28.345
5	28.647	28.395
6	25.902	29.902
7	34.270	34.018
8	50.846	44.195
9	25.081	29.266
10	16.314	11.254
11	61.418	61.169
12	52.339	47.279
13	38.137	37.885
14	36.274	36.791
15	19.006	18.502
16	42.117	41.605
17	30.008	34.450
18	21.225	21.742
19	23.503	29.266
20	36.948	36.698
21	31.217	29.266
22	20.989	14.338
23	15.439	15.927
24	25.080	29.266
25	19.641	24.083
26	30.829	30.577

For factors and levels applied to these experiments please refer to Rigi et al. 2013 (14). After reviewing the obtained results, the secretion amount of this protein using fuzzy regression model was estimated.

Hence, the estimation of fuzzy regression coefficients and the dependent using the described fuzzy regression led to the following conclusions:

$\tilde{A}_0$: [0, 0]
$\tilde{A}_1$: [0, 0]
$\tilde{A}_2$: [0.866, 0]
$\tilde{A}_3$: [15.178, 0]
$\tilde{A}_4$: [9.721, 16.375]
$\tilde{A}_5$: [0, 0]

Fuzzy data analysis work can be carried out. In other words, to deal with the uncertainty of the relationship between variables, the ordinary least square regression is done using fuzzy number not by random variables. Secondly, the incorrect data is also analyzed (through confidence interval or fuzzy number). The resulting estimates is an interval estimates with a confidence level of 0/90, which can be a reasonable estimate (Table 2).

Table 2 .Protein secretion estimation, utilizing fuzzy regression.

protein secretion amount	Right Spread	Left Spread
y_1	32.14	21.28
y_2	31.93	21.28
y_3	27.54	13.1
y_4	32.14	21.28
y_5	35.74	13.1
y_6	34.32	18.01
y_7	40.32	18.02
y_8	34.32	18.01
y_9	34.32	18.01
y_{10}	34.32	16.01
y_{11}	40.13	21.28
y_{12}	34.32	18.01
y_{13}	40.13	21.28
y_{14}	36.75	18.02
y_{15}	27.07	13.1
y_{16}	26.73	18.01
y_{17}	33.35	24.56
y_{18}	33.59	18.01
y_{19}	34.32	18.01
y_{20}	35.26	13.8
y_{21}	34.32	18.01
y_{22}	34.32	18.01
y_{23}	27.72	18.01
y_{24}	34.32	18.01
y_{25}	29.42	9.8
y_{26}	35.74	13.1

3. Discussion and conclusion

The production of recombinant protein is a wide production method which is currently developing. The secretion of extracellular protein is very desirable so that it can reduce the complexity of lower part processes and can help the improvement of the product (23-25). Statistical methods for improving the process in industrial biotechnology are ideal methods. In the improvement phase of the process, we improve some of the specified parameters from a group without any defect of some limitations. Of its most important goals are minimizing the cost, maximizing the power and efficiency for the next phase of preliminary studies and in a big scale. This is one of the important quantitative tools in industrial decision making (26, 27). For laboratory designing of this important protein and achieving the best and most improved conditions for production and secretion, its amount of production must first be calculated through statistical methods.

In the present study, the fuzzy regression method was utilized for estimating the amount of protein A.

The statistical technique that is regression analysis can be used for data which is itself dependent on other conditions. In these

instances, fuzzy data is used, and fuzzy regression analysis for fuzzy data is also applied to achieve at accurate divination (28-30). There has been a full definition and explanation of fuzzy regression in 'method' section of the article.

In some previous studies, the fuzzy regression method has been used for estimating the value of the proteins and recombinant proteins such our protein. Stefan Gnoth et al., said that quantitative knowledge can be activated for regulating the culture condition in recombinant protein production projects (31). In another recent study in this field, fuzzy regression model was used to optimizing Fed-Batch fermentation system to produce the recombinant γ-interferon protein in *E. coli* (32). Carl-Fredrik Mandenius also introduced some methods for monitoring, modeling, and controlling biological production systems, so it can be shown that statistical methods like fuzzy regression can help us (33). Some other investigations entitled quantitative analyses and evaluations of statistical software are also found which use fuzzy system in order to estimate culture conditions in an *Escherichia coli* cultivation expressing a recombinant protein (34), and even in Rat mammary gland such as estrogen receptor proteins (35). Thus far, there has been found no record of research estimating the amount of extracellular recombinant protein A using fuzzy regression, which is performed in the present study.

Considering the features of fuzzy mathematics and the figure utilized in this study, i.e. triangular fuzzy number, the resulting estimation is a distance estimation which has a very much higher reliability level than the spot estimation, and also the obtained estimations have the reliability level of 90 percent which can be a suitable reliability level for this estimation.

Through utilizing normal regression considering the mentioned conditions, an estimation can be made for the production of the mentioned protein. But our reason for estimating the amount of protein A through utilizing fuzzy regression has been, first, to increase the level of reliability of the mentioned estimation, and then, the fuzzy estimation of the production amount. This protein has less variance and in fact less distribution than normal regression method, which can be a special advantage for fuzzy regression in comparison with normal regression.

Conflict of Interest

The Authors have no conflict of interest.

Acknowledgments

Hereby, the authors would like to thank the Behbahan Khatam-al-Anbia University of Technology. This article is derived from the research project no: 94-10-00.

References

1. Jansson B, Uhlen M, Nygren PA. All individual domains of staphylococcal protein A show Fab binding. FEMS Immunol Med Microbiol. 1998;20(1):69-78.
2. Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, et al. Evaluation of protein A gene polymorphic region DNA sequencing for typing of Staphylococcus aureus strains. J Clin Microbiol. 1999;37(11):3556-63.
3. Benito Y, Kolb FA, Romby P, Lina G, Etienne J, Vandenesch F. Probing the structure of RNAIII, the Staphylococcus aureus agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. Rna. 2000;6(5):668-79.

4. Dunman PM, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, Wu S, et al. Transcription profiling-based identification of Staphylococcus aureus genes regulated by the agr and/or sarA loci. J Bacteriol. 2001;183(24):7341-53.
5. Novick RP. Autoinduction and signal transduction in the regulation of staphylococcal virulence. Mol Microbiol. 2003;48(6):1429-49.
6. Roben PW, Salem AN, Silverman GJ. VH3 family antibodies bind domain D of staphylococcal protein A. J Immunol. 1995; 154(12):6437-45.
7. Sasso EH, Silverman GJ, Mannik M. Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. J Immunol. 1989;142(8):2778-83.
8. Uhlen M, Guss B, Nilsson B, Gatenbeck S, Philipson L, Lindberg M. Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. J Biol Chem. 1984;259(3):1695-702.
9. Dossett JH, Kronvall G, Williams RC, Jr., Quie PG. Antiphagocytic effects of staphylococfcal protein A. J Immunol. 1969;103(6):1405-10.
10. Fournier B, Klier A, Rapoport G. The two-component system ArlS-ArlR is a regulator of virulence gene expression in Staphylococcus aureus. Mol Microbiol. 2001;41(1):247-61.
11. Lawman MJ, Thurmond MC, Reis KJ, Gauntlett DR, Boyle MD. Solid-phase radioimmunoassay for the detection of immunoglobulins against bovine Brucella abortus. Veterinary immunology and immunopathology. 1984;6(3-4):291-305.
12. Ghaedmohammadi S, Rigi G, Zadmard R, Ricca E, Ahmadian G. Immobilization of Bioactive Protein A from Staphylococcus aureus (SpA) on the Surface of Bacillus subtilis Spores. Molecular biotechnology. 2015; 57(8) : 756-66.
13. Rigi G, Beyranvand P, Ghaedmohammadi S, Heidarpanah S, Akbari Noghabi K, Ahmadian G. Comparison of the extracellular full-length and truncated recombinant protein A production in Escherichia coli BL21 (DE3). Journal of Paramedical Sciences. 2015;6(3):35-45.
14. Rigi G, Mohammadi SG, Arjomand MR, Ahmadian G, Noghabi KA. Optimization of extracellular truncated Staphylococcal Protein A (SpA) expression in Escherichia coli BL21 (DE3). Biotechnol Appl Biochem. 2013;13(10).
15. Chang P-T, Lee ES. A generalized fuzzy weighted least-squares regression. Fuzzy Sets and Systems. 1996;82(3):289-98.
16. Dubois DJ. Fuzzy sets and systems: theory and applications: Academic press; 1980.
17. BAHRAMI T, CHASHIANI S. Estimating daily peak load using symmetrical doubly linear adaptive-fuzzy regression model. Cumhuriyet Science Journal. 2015; 36(3) : 715-20.
18. BAHRAMI T, BAHRAMI M. Claim reserving with fuzzy regression. Cumhuriyet Science Journal. 2015; 36(3): 704-9.
19. Chang P-T. Fuzzy seasonality forecasting. Fuzzy Sets and Systems. 1997;90(1):1-10.
20. Peters G. Fuzzy linear regression with fuzzy intervals. Fuzzy Sets and Systems. 1994;63(1):45-55.
21. Tanaka H, Watada J. Possibilistic linear systems and their application to the linear regression model. Fuzzy sets and systems. 1988;27(3):275-89.
22. Zimmermann H-J. Fuzzy set theory—and its applications: Springer Science & Business Media; 2011.
23. Cornelis P. Expressing genes in different Escherichia coli compartments. Curr Opin Biotechnol. 2000;11(5):450-4.
24. Makrides SC. Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev. 1996;60(3):512-38.
25. Mergulha OF, Monteiro, G.A. . Secretion capacity limitations of the Sec pathway in Escherichia coli. . J Microb Biotechnol. 2004;14:128-33.
26. Oskouie SFG, Tabandeh F, Yakhchali B, Eftekhar F. Response surface optimization of medium composition for alkaline protease production by Bacillus clausii. Biochemical Engineering Journal. 2008;39(1):37-42.
27. Singh AK, Mehta G, Chhatpar HS. Optimization of medium constituents for improved chitinase production by Paenibacillus sp. D1 using statistical approach. Lett Appl Microbiol. 2009;49(6):708-14.

28. Diamond P, Tanaka H. Fuzzy Regression Analysis. In: Słowiński R, editor. Fuzzy Sets in Decision Analysis, Operations Research and Statistics. Boston, MA: Springer US; 1998. p. 349-87.
29. Hong T, Wang P. Fuzzy interaction regression for short term load forecasting. Fuzzy optimization and decision making. 2014;13(1):91-103.
30. Kovac P, Rodic D, Pucovsky V, Savkovic B, Gostimirovic M. Application of fuzzy logic and regression analysis for modeling surface roughness in face milliing. Journal of Intelligent manufacturing. 2013;24(4): 755-62.
31. Gnoth S, Jenzsch M, Simutis R, Lübbert A. Control of cultivation processes for recombinant protein production: a review. Bioprocess and Biosystems Engineering. 2008;31(1):21-39.
32. Feyzdar M, Vali Ahmad R, Babaeipour V. Identification and Optimization of Recombinant E. coli Fed-Batch Fermentation Producing γ-Interferon Protein. International Journal of Chemical Reactor Engineering2013. p. 123.
33. Mandenius C-F. Recent developments in the monitoring, modeling and control of biological production systems. Bioprocess and Biosystems Engineering. 2004; 26(6): 347-51.
34. Warth B, Rajkai G, Mandenius C-F. Evaluation of software sensors for on-line estimation of culture conditions in an Escherichia coli cultivation expressing a recombinant protein. Journal of Biotechnology. 2010;147(1):37-45.
35. Saji S, Sakaguchi H, Andersson S, Warner M, Gustafsson J-Å. Quantitative Analysis of Estrogen Receptor Proteins in Rat Mammary Gland. Endocrinology. 2001; 142(7):3177-86.