Comparison of the extracellular full-length and truncated recombinant protein A production in *Escherichia coli* BL21 (DE3)

Garshasb Rigi[1], Parisa Beyranvand[2], Samira Ghaedmohammadi[2], Sara Heidarpanah[2], Kambiz Akbari Noghabi[2], Gholamreza Ahmadian[2,*]

[1]Department of Biology, College of Basic Science, Behbahan Khatam Alanbia University of Technology, Khuzestan, Iran
[2]Department of Industrial and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB) , Tehran, Iran

*Corresponding Author: email address: ahmadian@nigeb.ac.ir (G. Ahmadian)

ABSTRACT

Protein A is a commercially important protein in biotechnological and medicinal applications. The great value of this protein and its applications in genetic and protein engineering and microbial researches as well as the growing use in biochemical industries, biotechnology, medicine and pharmacology, highlight the importance of the present study. In this survey the encoding genes of full-length and truncated forms of protein A were expressed in *E. coli* under an optimized expression condition. Optimization of the culture conditions resulted in an increase in expression and secretion of both forms of the protein, the pattern of expression and secretion levels for two forms was completely different. A minimum of 10-fold higher expression was observed for the truncated protein in comparison to that of the full-length recombinant form. Hydropathy plot of both forms of proteins showed that the missing domains in the truncated form contain groups of amino acids with high hydrophobicity score. Deletion of the terminal region could led to a higher expression level of the recombinant protein in *E. coli*. The function of these two proteins was studied using ELISA, which showed a higher activity for the truncated form for binding to IgG, compared to the full-length protein.

Keywords: Expression; Protein A; Optimization; Medicinal application

INTRODUCTION

Protein A (SpA) is a major surface protein of *Staphylococcus aureus* that is covalently anchored in the staphylococcal cell wall through its carboxyl terminal end. It is comprised of 4 or 5 homologous repeat domains (E, D, A, B, C) of 56-61 residues followed by a polymorphic variable repeat region, Xr, and a conserved region, Xc, which includes a cell-wall attachment sequence [1-3]. The variability in the number and sequences of these repeated regions has been exploited as an epidemiological marker, although their significance is not established [4, 5]. Structural analysis of a single SpA domain revealed that it is composed of a 3 helical bundle. The solved structures of SpA domains in complex with IgG Fc and with a VH3-derived IgM Fab demonstrate how different parts of the SpA repeat are involved in the two interactions. In fact, a single SpA domain may bind each ligand simultaneously [6-9]. This protein has been widely used for quantitative and qualitative immunological techniques [10-12]. However, protein A is a commercially important protein in biotechnological and medicinal applications. Box and Wilson were the first describing Response surface methodology (RSM) for optimizing the conditions in a multivariable system. This method is a concise way of describing and predicting responses of a system of variables [13].

Furthermore, RSM defines the effect of the independent variables, alone or in combinations, on the process and generates a mathematical model that precisely describes the overall process.

A successful experimental factorial design together with RSM was already used in other fields of research and is well-suited with the study of the main and interaction of the factor in bioconversion yield [14].

Recent studies have indicated the use of RSM for analyzing effects of different factors on proteolytic activity and optimization of xylanase production. The purpose of this study is to investigate the effects of several factors on the production of a commercially important protein. In previous studies, the truncated form of *spa* gene was identified, cloned and expressed in *E. coli* BL21 (DE3). Expression was optimized by changing various parameters [15].

In this study, attempts were made to grow recombinant *E. coli* BL21 carrying the full-length and truncated forms of *spa* genes under flask and culture conditions, and to compare cell bulk, protein expression levels and extracellular production efficiency. This is the first study comparing the truncated and intact form of protein A in an appropriate condition and investigating their secretion in E. coli.

The great value of this protein and its applications in genetic and protein engineering and microbial researches as well as the growing use in biochemical industries, biotechnology, medicine and pharmacology, highlight the importance of the present study, because knowing the similarities and differences between these two forms of protein helps selecting the right one for each application.

MATERIALS AND METHODS

Bacterial strains and media used in this study have been presented in Table 1. Two strains of *E. coli* BL21 (DE3) (Novagen, Darmstadt, Germany) harbouring pET-26b (+), one of them entitled full-length (SpA1) consisted of a native signal peptide and it's full amino acids against another one, truncated (SpA2) form, that consisted of a native signal peptide and 5 immunoglobulin binding sites (residues 1 to 327) of *spa* genes, were used. The *spa* genes were PCR-amplified from the genomic DNA of *Staphylococcus aureus* (ATCC 6538) as the template, using the *Pfu* DNA polymerase (Fermentas, Germany) together with the following primers in Table 1. For designing the growth curves and optimization tests, cultures were grown in Luria-Bertani (LB) broth or on LB agar.

Table 1. Primers, plasmids and strains were used in this study.

Name	Sequence (5'→3')	Reference
Primers		
SPAF6	GGGGCAT**ATGAAAAAGAAAAACATTTATTC**	Rigi et al. 2013
SPAR8	GGGGCTCGAG**TTATTTTGGTGCTTGAGCATCG**	Rigi et al. 2013
SPAR6	GGGGCTCGAG**TTATAGTTCGCGACGACG**	This study
Plasmids		
pET-26b (+)		Novagen Cat. No.69862-3
Strains		
E. coli BL21 (DE3)		Novagen, Darmstadt, Germany
Staphylococcus aureus (ATCC 6538)		NIGEB collection

Primers matched the coding sequence of the gene (given in bold) and contained restriction sites (underlined).

Bioinformatics analysis

Numerous localization predictors have been developed for predicting the final destination of proteins and predicting signal peptides and corresponding cleavage sites in protein constructs based on their AA sequence [16, 17]. In prior comparative studies, the online program SignalP was identified to be the method of choice [16, 18]. Thus, we applied version 4.1 of this online tool for cleavage site prediction.

Regions of high hydrophobicity in full length and truncated protein A were determined using ProtScale Tool [19] in which hydropathic (hydrophobic) regions achieve a positive value.

RNA structures were analyzed using the algorithm of favorite thermodynamic by Vienna RNA

Secondary Structure Prediction program on the web at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi [20-22].

Bacterial growth and protein expression conditions

In this study, at first the growth conditions of two recombinant *E. coli* BL21 strains were constructed as follows. The cultures of both strains were inoculated into LB broth containing kanamycin (50 mg/ml). The cultures were then grown aerobically at 37 ˚C with shaking at 200 rpm on a rotary shaker (Jeio Tech, Seoul, South Korea). Sampling was carried out immediately after culturing and then at each hour. Optical density (OD) of the samples was measured at 600 nm (OD600) and sampling was continued until OD600 no longer increased. In order to achieve optimum growth conditions for the two recombinant *E. coli* strains, 26 experiments were designed by the response surface methodology (RSM) which was developed based on the full factorial central composite design (CCD) using the (Design – Expert v. 8.0.6, Stat – Ease, Inc., Minneapolis, MN, USA) software.

The experimental design was carried out under conditions applied in the previous research [23], and each test was performed in triplicate for every strain separately.

Samples were collected as before and analysed by 12% (w⁄v) sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) [24]. For western blot analysis, bacterial lysate from full-length and truncated clones were subjected to SDS-PAGE, transferred to nitrocellulose membrane (Roche, Germany).

The membrane was then incubated with IgG anti-goat HRP-conjugated antibody (Sigma, USA), and the blot was developed with 4-Chloro-1-naphthol substrate (Sigma, USA). It should be noted that in all the experiments regarding expressed proteins, equal amount of proteins were used. Also, the optimum conditions were then obtained from the CCD experiments. Subsequently, the production efficiency of the optimum conditions (which was repeated thrice) was calculated by multiplying the expression levels. The mean values of the three repeated efficiency assessments were compared with the independent t-test (two-tailed).

The ELISA test for Detection of two forms of SpA for comparison of their functional differences

Two forms of SpA were determined by ELISA test [25]: For polyclonal antibody production, purified proteins were injected in rabbits four times at 10 day intervals. In the first injection, protein concentration was 300-500 µg/ml with the same volume of complete adjuvant. Subsequent injections contained protein concentration of 200-300 µg/ml with the same volume of incomplete adjuvant and the third and fourth subcutaneous injections were also made, subsequently.

On day 44, total blood was collected and sera were tested against purified protein and optimal dilution of anti-amylase. Sera were established as 1:4000 for Sandwich ELISA for Detection of two forms of SpA. 96-well plates (Costar, USA) were coated with 100 ml of rabbit IgG (10 mg/ml) for 2 h at 37°C in 50mM $NaHCO_3$_$NaCO_3$ (pH 9.6). Then the plates were washed twice with 0.1% Tween 20 in PBS (PBST) and the wells were blocked overnight at 4°C with 200 ml 1% BSA in PBS. After two washes with PBST, the plates were incubated with 100 ml per well of SpA standard dilution samples (0.001–10 mg/ml) or serially diluted test samples for 30 min at 37°C.

After six washes with PBST, the plates were finally incubated with goat IgG conjugated horseradish peroxidase (HRP) at 1: 2000 dilution in PBS at room temperature for 30 min. The plates were finally washed again and 100 ml OPD substrate solution was added to each well. After stopping the reaction with 50 ml 1M H_2SO_4, the absorbance at 490 nm was determined using a microplate reader (ELX800, Biotech, USA). Then the functions of two forms were compared with the independent t-test**.**

RESULTS

Bioinformatics studies:

The predictions of SignalP correspond to proteins predicted to have the exactly the same cleavage site in two forms of protein A (Figure 1).

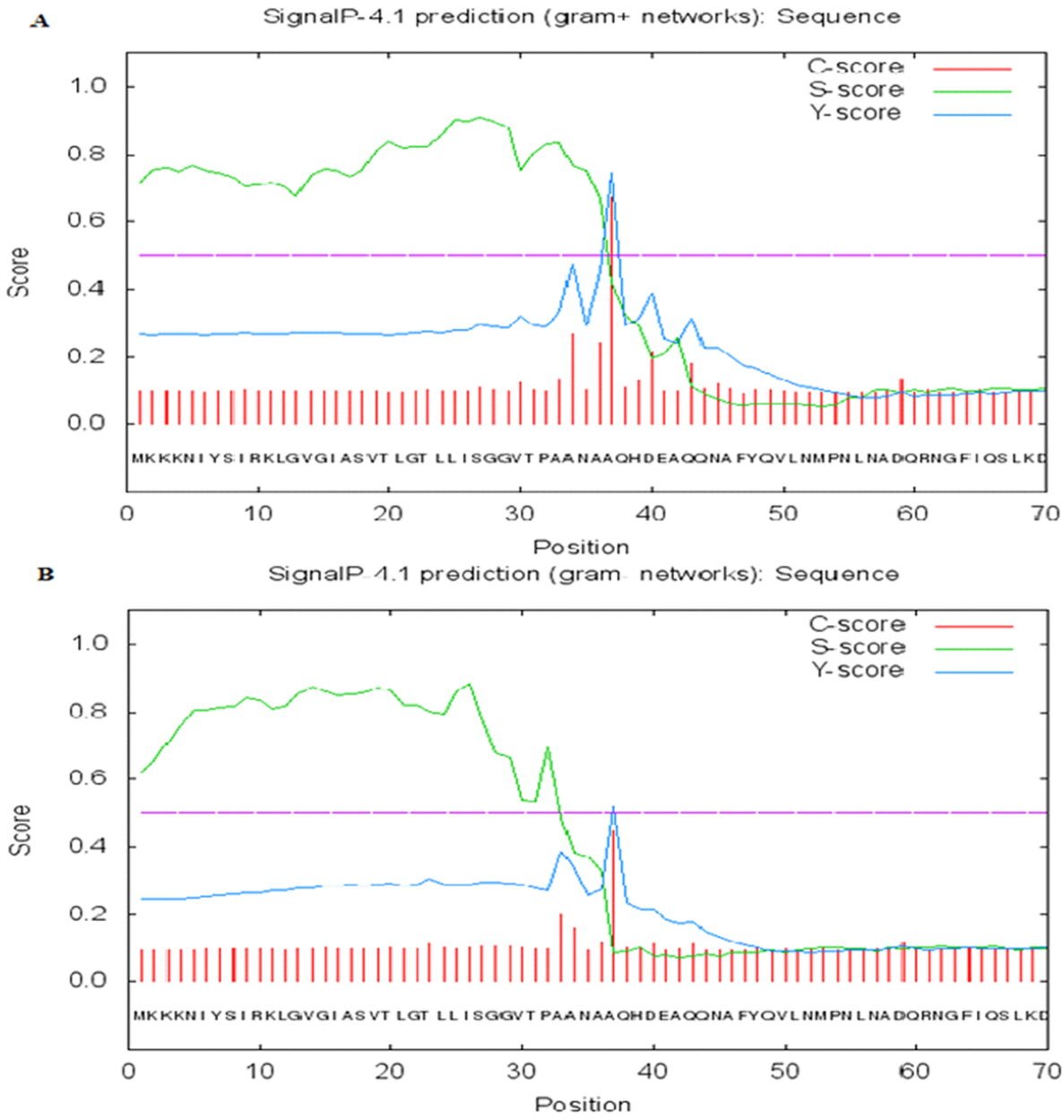

Figure 1. SignalP 4.1 analysis of amino acid sequence of protein A. "(A), Staphylococcal protein A for expression in gram positive bacteria. (B), Staphylococcal protein A for expression in gram negative bacteria.

There are various tools to analyze and predict the hydrophobicity pattern of a protein, among which the hydrophobicity criteria of Kyte-Doolittle is a widely applied scale for delineating hydrophobic character of a protein. Each amino acid is given a hydrophobicity score between -4.5 and 4.5. A score of 4.5 is the most hydrophobic and a score of - 4.5 is the most hydrophilic. Regions values above zero are hydrophobic in character. Using these criteria and ProtScale tool with a window size of 7 amino acids, one region has been scored with high hydrophobic values (Figure 2).

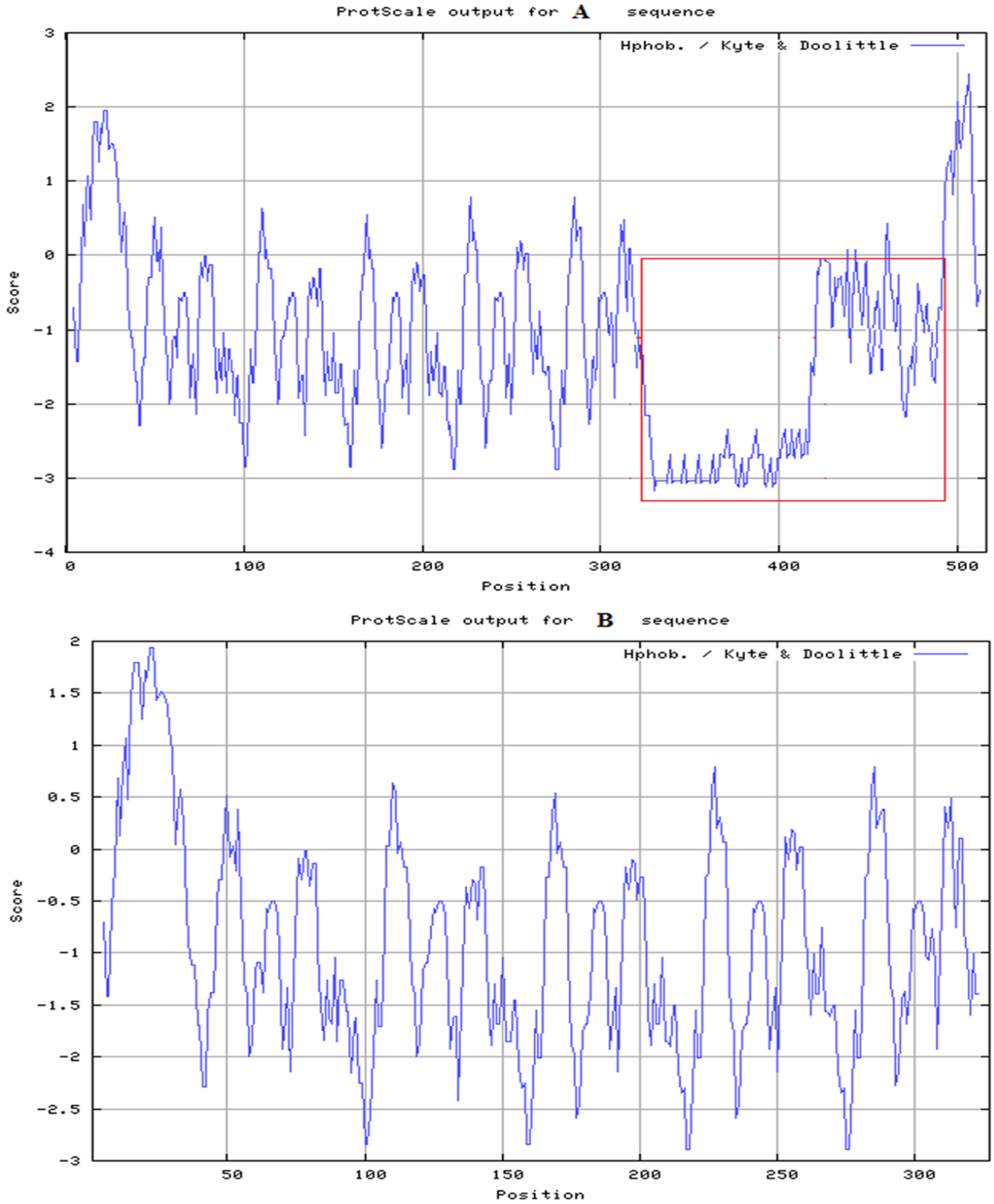

Figure 2. Hydropathy plot of the full-length (a), and truncated (b) protein A.
Main effects of different factors at different levels on the expression of the complete and truncated *spa* genes. For factors and levels applied to these experiments please refer to Rigi et al. (2013). Panel A, Complete SpA. Panel B, Shortened SpA.

Comparison of these graphs shows some changes in the hydrophobicity maxima in the full-length form. RNA Secondary Structure Prediction showed that the free energy of the thermodynamic ensemble for truncated *spa* gene is -208.24 kcal/mol and it is -337.36 kcal/mol for full-length *spa* gene. The number of loops in truncated mRNA were more than full-length mRNA and loops in structure of mRNA related to truncated *spa* gene also were bigger than the loops in structure of mRNA related to full-length *spa* gene (Figure 3).

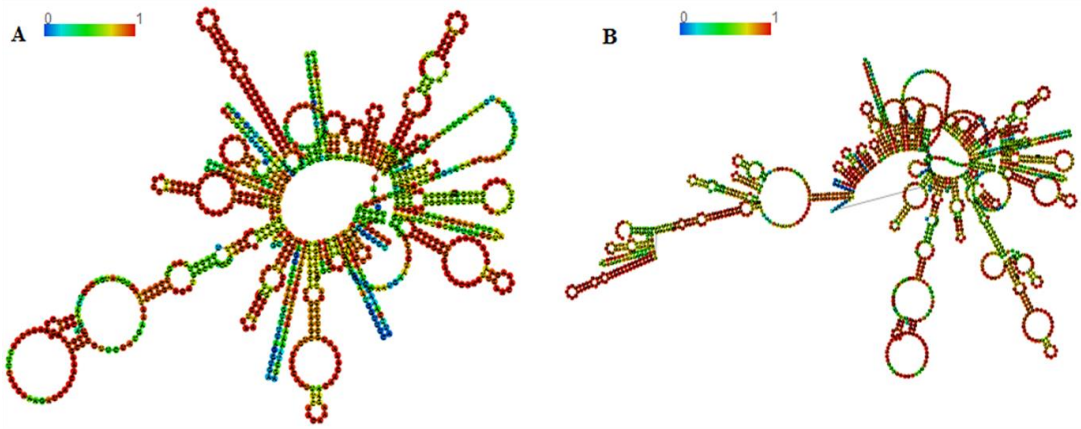

Figure 3. Secondary structure of *spa* RNA. (a) for truncated *spa* gene (ΔG = -208.24 kcal/mol); (b) for full-length *spa* gene (ΔG = -337.36 kcal/mol). The structure above is colored by base-pairing probabilities.

Protein expression and optimization of its conditions

The expression of full-length and truncated forms of protein A under control of the T7 promoter was performed in *E. coli*. Protein assay [26] was performed on samples and equal amount of proteins were loaded onto the wells. The result of SDS-PAGE is shown in figure 4 (Figure 4).

It should be noted that the expression of the full-length protein A is very low. We observed a higher expression of a specific protein band with expected size for truncated protein A in comparison with full-length protein A. The immunoreactivity of these bands was shown by Western blot analysis using IgG anti-goat HRP-conjugated antibody (Figure 5). The growth conditions of the two recombinant *E. coli* BL21 strains were optimized by application of the RSM method. After determination of the percentage production of the complete and shortened proteins, the results were analysed using Design – Expert software and the effect of different levels of each factor on the response were characterised in both constructs (Figure 6).

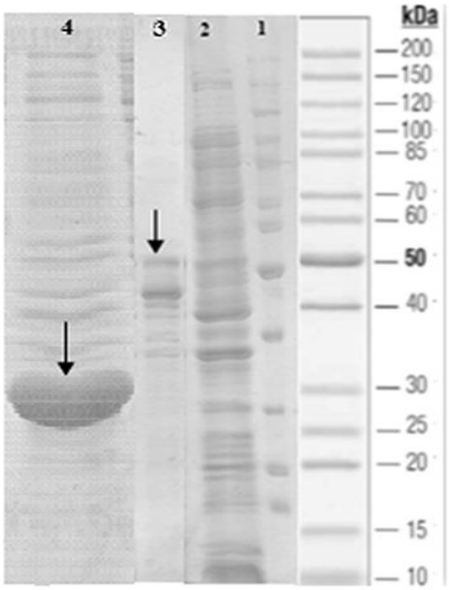

Figure 4. Comparison of extracellular complete and shortened protein A production under optimum conditions in the flask by SDS–PAGE analysis. Lane 1. Low molecular weight size marker. Lane 2: Total protein at T_0 (no induction). Lanes 3 and 4. Extracellular protein profiles related to the two constructs of *E. coli* BL21 harbouring pET-*spa1* and pET- *spa2* cultured in the flask, respectively. The related bands are shown by arrows.

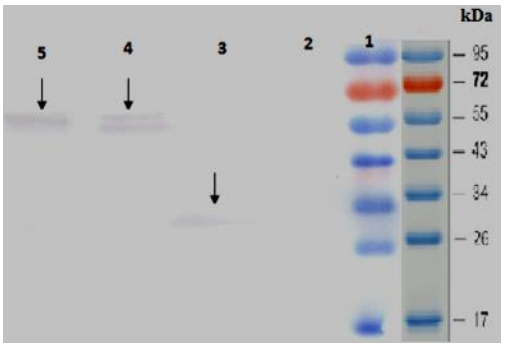

Figure 5. Confirmation of protein A (SpA2) by Western blot analysis using IgG anti goat HRP-conjugated antibody; Lane 1, Prestained protein molecular weight marker; Lane 2, The negative control that is total protein at T0 (no induction). Lane 3, Total supernatant of truncated protein A; Lane 4, Total pellet of full-length protein A; Lane 5, Total supernatant of full-length protein A.

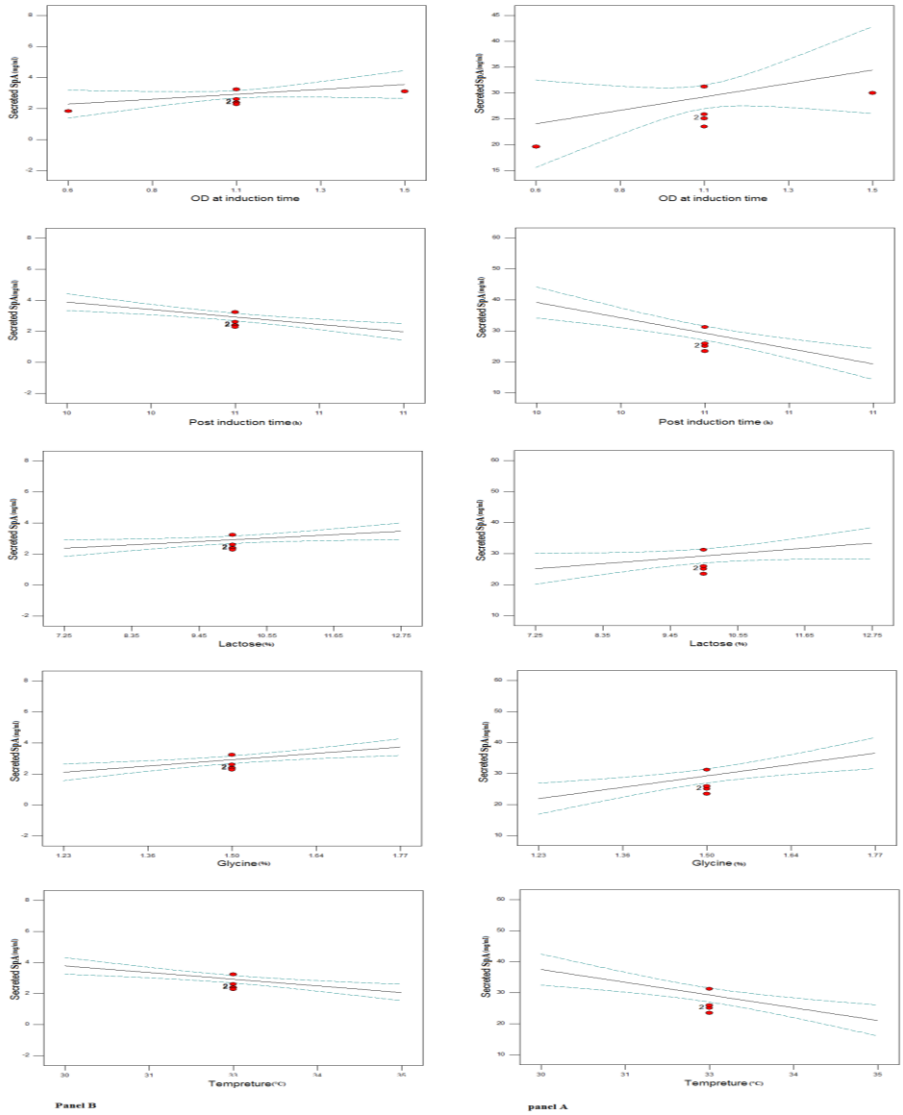

Figure 6. Main effects of different factors at different levels on the expression of the complete and truncated *spa* genes. For factors and levels applied to these experiments please refer to Rigi et al. (2013). Panel A, Complete SpA. Panel B, Shortened SpA

In accordance Figure 6, the level of factors as, time after induction 11, lactose concentration = 10 %, glycine concentration = 1.5% temperature = 30 and cell density (OD_{600}) at induction time = 1. For protein production from the expression of the complete *spa* gene, and level of factors as, time after induction = 11, temperature = 33, lactose concentration = 12% and glycine concentration = 1.77% and cell density (OD_{600}) for induction = 1.1 for protein production factor have the highest effect on the amount of the truncated (*spa2*) gene expression. The obtained results represent optimum conditions for maximum production of the two forms of SpA. Also, a central composite design (Design – Expert v. 8.0.6, Stat – Ease, Inc., Minneapolis, MN, USA) anticipated that under these conditions, interaction between cell density (OD_{600}) at induction time and lactose concentration are significant for full-length protein (Figure 7) and the interaction between cell density (OD_{600}) and lactose concentration, post-induction time and temperature, lactose concentration and glycine concentration are significant model terms for truncated form [23].

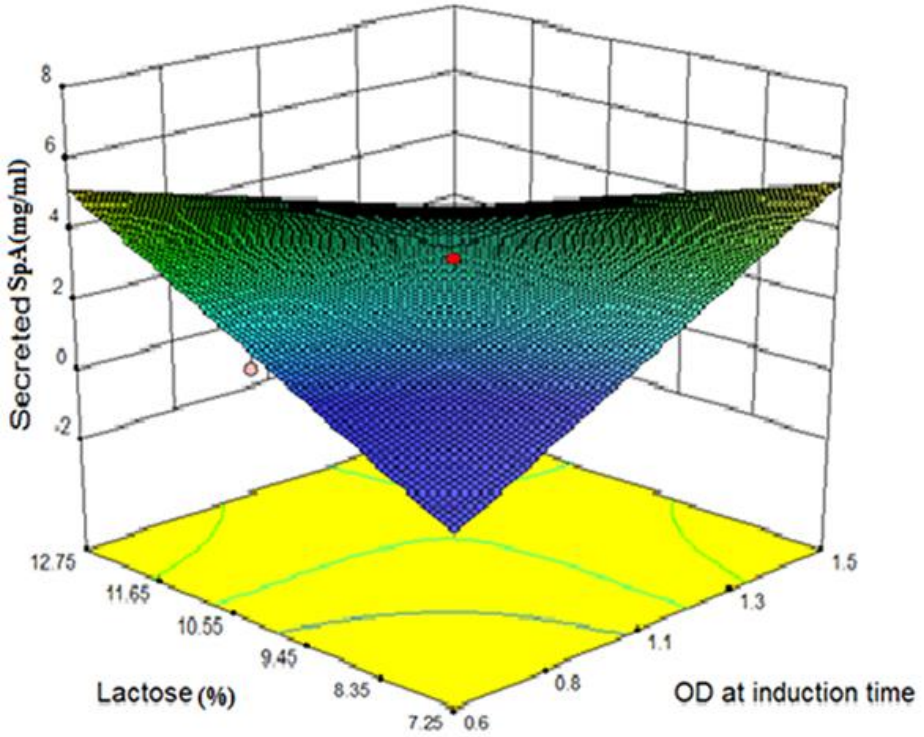

Figure 7. Response surface and contour curves of secreted full-length protein A from the recombinant *E. coli* showing interactions between lactose concentration and cell density ($OD._{600}$), at 33 ˚C , 100 rpm , post-induction time of 11 h and glycine concentration of 1.5.

Performing experimental tests under these conditions Analysis of these results (Table 2) using the SPSS software (version 16) showed that they were significant with 95% confidence interval (P value <0.05).

This, firstly confirms the advantage of the optimum conditions with a view to production efficiency and secondly indicates the advantage of the *E. coli* BL21 harbouring pET-*spa2* over *E. coli* BL21 harbouring pET-*spa1*.

Table 2. Variables showing observed values of secreted protein A in the supernatant.

Experiment run	Extracellular truncated protein A (mg/ml)	Extracellular full-length protein A (mg/ml)
1	25.342	2.321
2	13.351	1.521
3	19.173	1.834
4	28.597	2.871
5	28.647	2.831
6	25.902	2.611
7	34.270	3.331
8	50.846	5.129
9	25.081	2.359
10	16.314	1.711
11	61.418	6.215
12	52.339	5.179
13	38.137	3.901
14	36.274	3.998
15	19.006	1.897
16	42.117	4.389
17	30.008	3.103
18	21.225	2.012
19	23.503	2.264
20	36.948	3.509
21	31.217	3.226
22	20.989	2.003
23	15.439	1.439
24	25.080	2.406
25	19.641	1.839
26	30.829	3.117

For factors and levels applied to these experiments please refer to Rigi et al. 2013 [15].

Sandwich ELISA for Detection of two forms of SpA for a comparative study of amount of IgG captured at different amount of rabbit standard IgG and serum from 100 mg to 6 mg was investigated. The independent t-test showed that with 100 ml two forms of SpA, the adsorption amount of IgG rapidly increased for the truncated SpA.

DISCUSSION

Although *E. coli* is Gram-negative, the native signal peptide of SpA, from a Gram-positive bacterium was used in this study. Interestingly, the results revealed that the signal peptide from *Staphylococcus aureus* was recognized in *E. coli* and the resulting SpA was expressed and secreted into the medium. Expression and secretion level of the full-length form of protein was remarkably lower (about 10 fold) compared to the truncated form. Although, optimization of the condition showed an increase in expression and secretion for both forms of the protein, the difference between expression and secretion levels for the truncated form was completely evident in this condition.

Optimized conditions for the truncated form revealed that elevating concentrations of glycine in the medium of bacteria, increased protein expression as well as its secretion level.

Whereas, for the full-length protein, this factor only along with other factors led to an increase in protein expression which was not high as that in truncated form. Regarding to protein secretion level, the effect of glycine on secretion was not considerable compared to the truncated form.

We used hydrophobicity criteria of Kyte-Doolittle to analyze and predict the hydrophobicity pattern of these two forms of protein A. Using this program, one region in the full-length has been scored with high hydrophobic values (Figure 2). The comparison of the two profiles showed some shifts in the hydrophobicity profile from full-length to truncated form. One of the important factors in determining the expression levels of proteins in bacterial hosts is hydrophobic nature of polypeptide. There are reports of improved expression in bacteria by exchanging the hydrophobic residues into neutral or hydrophilic residues [27-29]. These amino acid replacements are especially important for surface residues where exposed hydrophobic residues may increase aggregations and misfolding inside host expression system. Total free energy (ΔG) and number and size of loops in the secondary structure and folding of mRNAs transcripted from these two forms of *spa* gene show that translation and expression in truncated form can be better than full-length. The function of these two proteins was studied using a molecular technique, sandwich ELISA, which showed a higher activity for the truncated form of protein A, interacting with IgG, compared to the full-length protein.

This augmented activity could be explained by lack of a hydrophobic extra segment and also lack of aggregations and misfolding which may cause more exposure of binding sites to Igg, therefore they show a higher activity. However, further studies are required to clarify the reason of elevated expression, secretion and activity for the truncated form.

ACKNOWLEDGMENTS

The authors would like to thank the National Institute of Genetic Engineering and Biotechnology (NIGEB) of Iran for providing the necessary equipments.
Funding for the project was provided by the Iran National Science foundation (Project no. 90004656).

REFERENCES

1.Uhlen, M., et al., Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. J Biol Chem, 1984. 259(3): p. 1695-702.
2.Moks, T., et al., Staphylococcal protein A consists of five IgG-binding domains. Eur J Biochem, 1986. 156(3): p. 637-43.
3.Langone, J.J., Protein A of *Staphylococcus aureus* and related immunoglobulin receptors produced by *streptococci* and *pneumonococci*. Adv Immunol, 1982. 32: p. 157-252.
4.Shopsin, B., et al., Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J Clin Microbiol, 1999. 37(11): p. 3556-63.
5.Jansson, B., M. Uhlen, and P.A. Nygren, All individual domains of staphylococcal protein A show Fab binding. FEMS Immunol Med Microbiol, 1998. 20(1): p. 69-78.
6.Sasso, E.H., G.J. Silverman, and M. Mannik, Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. J Immunol, 1989. 142(8): p. 2778-83.
7.Roben, P.W., A.N. Salem, and G.J. Silverman, VH3 family antibodies bind domain D of staphylococcal protein A. J Immunol, 1995. 154(12): p. 6437-45.
8.Graille, M., et al., Crystal structure of a *Staphylococcus aureus* protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A, 2000. 97(10): p. 5399-404.
9.Starovasnik, M.A., et al., Antibody variable region binding by Staphylococcal protein A: thermodynamic analysis and location of the Fv binding site on E-domain. Protein Sci, 1999. 8(7): p. 1423-31.
10.Lofdahl, S., et al., Gene for staphylococcal protein A. Proc Natl Acad Sci U S A, 1983. 80(3): p. 697-701.
11.Van Loghem, E., et al., Staphylococcal protein A and human IgG subclasses and allotypes. Scand J Immunol, 1982. 15(3): p. 275-8.
12.Tashiro, M. and G.T. Montelione, Structures of bacterial immunoglobulin-binding domains and their complexes with immunoglobulins. Curr Opin Struct Biol, 1995. 5(4): p. 471-81.
13.Murphy, T.D., Design and Analysis of Industrial Experiments. J Chem Eng, 1977: p. 168-182.
14.Gohel, V., Maisuria, V. and Chhatpar, H.S. , Utilization of various chitinous sources for production of mycolytic enzymes by *Pantoea dispersa* in benchtop fermenter. Enzyme and Microbial Technology, 2007. 40: p. 1608-1614.
15.Rigi, G., et al., Optimization of extracellular truncated Staphylococcal Protein A (SpA) expression in *Escherichia coli* BL21 (DE3). Biotechnol Appl Biochem, 2013. 13(10).
16.Petersen, T.N., et al., SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011 Sep 29;8(10):785-6. doi: 10.1038/nmeth.1701.
17.Emanuelsson, O., et al., Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc, 2007. 2(4): p. 953-71.
18.Choo, K.H., T.W. Tan, and S. Ranganathan, A comprehensive assessment of N-terminal signal peptides prediction methods. BMC Bioinformatics, 2009. 3(10): p. 1471-2105.

19.Kyte, J. and R.F. Doolittle, A simple method for displaying the hydropathic character of a protein. J Mol Biol, 1982. 157(1): p. 105-32.
20.Mathews, D.H., et al., Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(19): p. 7287-7292.
21.Gruber, A.R., et al., The Vienna RNA websuite. Nucleic Acids Res, 2008. 36(Web Server issue): p. 19.
22.Lorenz, R., et al., ViennaRNA Package 2.0. Algorithms Mol Biol, 2011. 6(26): p. 1748-7188.
23.Vahed M Fau - Motalebi, E., et al., - Improving the Chitinolytic Activity of *Bacillus pumilus* SG2 by Random Mutagenesis. J Microbiol Biotechnol, 2013. 23(11): p. 1519-28.
24.Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 1970. 227(5259): p. 680-5.
25.Nielsen, U.B. and B.H. Geierstanger, Multiplexed sandwich assays in microarray format. J Immunol Methods, 2004. 290(1-2): p. 107-20.
26.Bradford, M.M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem, 1976. 72: p. 248-54.
27.Ahmadian, G., et al., Bacterial expression and functional characterization of a naturally occurring exon6-less preprochymosin cDNA. 2005. 2005.
28.Linder, J.U., A. Hammer, and J.E. Schultz, The effect of HAMP domains on class IIIb adenylyl cyclases from *Mycobacterium tuberculosis*. Eur J Biochem, 2004. 271(12): p. 2446-51.
29.Strub, C., et al., Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. BMC Biochem, 2004. 5: p. 9.