School of Pharmacy & Pharmaceutical Sciences
Isfahan University of Medical Sciences

Original Article

Improvement of solubility and yield of recombinant protein expression in *E. coli* using a two-step system

Tahereh Sadeghian-Rizi[1,2], Azade Ebrahimi[1,2], Fatemeh Moazzen[1], Hesam Yousefian[2], and Ali Jahanian-Najafabadi[1,*]

[1]*Department of Pharmaceutical Biotechnology, School of Pharmacy and Pharmaceutical Sciences,*
Isfahan University of Medical Sciences, Isfahan, I.R. Iran.
[2]*Student Research Committee, School of Pharmacy and Pharmaceutical Sciences,*
Isfahan University of Medical Sciences, Isfahan, I.R. Iran.

Abstract

Overexpression of recombinant proteins in *Escherichia coli* results in inclusion body formation, and consequently decreased production yield and increased production cost. Co-expression of chaperon systems accompanied by recombinant protein is a general method to increase the production yield. However, it has not been successful enough due to imposed intense stress to the host cells. The aim of this study was to balance the rate of protein production and the imposed cellular stresses using a two-step expression system. For this purpose, in the first step, green fluorescent protein (GFP) was expressed as a recombinant protein model under control of the T7-TetO artificial promoter-operator, accompanied by Dnak/J/GrpE chaperon system. Then, in the next step, TetR repressor was activated automatically under the control of the stress promoter ibpAB and suppressed the GFP production after accumulation of inclusion bodies. Thus in this step incorrect folded proteins and inclusion bodies are refolded causing increased yield and solubility of the recombinant protein and restarting GFP expression again. Total GFP, soluble and insoluble GFP fractions, were measured by Synergy H1 multiple reader. Results showed that expression yield and soluble/insoluble ratio of GFP have been increased 5 and 2.5 times using this system in comparison with the single step process, respectively. The efficiency of this system in increasing solubility and production yield of recombinant proteins was confirmed. The two-step system must be evaluated for expression of various proteins to further confirm its applicability in the field of recombinant protein production.

Keywords: Chaperons; *E. coli*; GFP; Inclusion bodies; Protein solubility; Recombinant proteins.

INTRODUCTION

Recombinant proteins have many applications in the course of molecular biology, food industries, and pharmaceutical companies (1). The universality of the genetic codons permits protein expression in heterologous organisms (2). *Escherichia coli* (*E. coli)* is one of the most frequently used production host due to the extensive information of its physiological and genetics processes, ease of cultivation, low cost of its growth, and achievement of high cell density fermentations (3,4). Overexpression of the recombinant protein in this organism often leads to incorrect folding, aggregation, and inclusion body formation. Limitation of the host chaperon capacity may be one of the causes of this problem (5). Chaperons help proper folding of the newly synthesized proteins and refolding aggregated and misfolded proteins to the native conformation as protein quality control system in cells (6). DnaK and GroEL/ES systems are two main chaperon systems in the *E. coli* cells (7-9). General approaches for increasing protein production yields are based on the expression of recombinant proteins accompanied by chaperons whose unbalanced expression cause severe stress in host cells lowering the success of the system (10,11).

*Corresponding author: A. Jahanian-Najafabadi
Tel: +98-3137927056, Fax: +98-3136680011
Email: jahanian@pharm.mui.ac.ir

Other researchers designed expression processes to equilibrate protein production and cellular stresses. Previously, a two-step procedure reported by Marco *et al.* showed efficient enhancement of solubility. In the first step of this procedure the four chaperon systems were coordinately co-overproduced with recombinant proteins to optimized folding. Then in the next step protein biosynthesis was inhibited by removal of isopropyl β- d-1-thiogalactopyranoside (IPTG) to permit chaperone-mediated refolding of the misfolded and aggregated proteins *in vivo*. This strategy increased the solubility of 70% of 64 different heterologous proteins tested up to 42-fold (12). Disadvantage of this method is the need of external intervention for protein expression inhibition which is impractical and time consuming with expensive process in the industrial scale. Dragosits *et al.* presented a novel self-regulatory protein production system that couples the control of recombinant protein production with a stress-induced, negative feedback mechanism. In their study the soluble/insoluble ratio of recombinant protein increased but total protein yield decreased (13). In the present research we investigated the latter system by some modifications in the expression vectors, expression conditions, and in the presence of DnaK/J/GrpE chaperon system for increasing protein solubility without decreasing total protein yield. In this manner, recombinant protein production was suppressed by the precise feedback system in the appropriate time by the cell itself and after refolding of the misfolded and aggregated proteins, expression of recombinant protein was restarted.

MATERIALS AND METHODS

Bacterial strains

E. coli strains used in this study were XL1-blue and BL21 [DE3]. All strains were grown on LB agar plates or in LB broth supplemented with 100 μg/mL ampicillin, 50 μg/mL kanamycin and/or 34 μg/mL chloramphenicol as necessary.

Plasmid construction
GFP and TetR plasmids

The schematic presentation of the produced recombinant vectors is represented in Fig. 1. Briefly, to produce the presented constructs, Tet operator (TetO) sequence was inserted downstream of the T7 promoter sequence and a ribosome binding site sequence of the pET28a vector (13). Then, the GFP coding sequence was codon optimized for expression in *E. coli BL21*(DE3) cells and cloned downstream of the ribosome binding site sequence. Finally, the corresponding DNA sequence was ordered to be synthesized by Biomatik Company (Canada) and cloned into the *Bgl*II and *Xho*I restriction sites of the pET28a plasmid.

The TetR repressor sequence was also codon optimized for optimal expression in *E. coli BL21* and was cloned under the control of stress-sensitive IbpAB promoter (P$_{IbpAB}$) and a mutant weak ribosome binding site sequence. Finally, the corresponding DNA sequence was ordered to be synthesized by Biomatik Company (Canada) and cloned to the *Bgl*I and *Xho*I restriction sites of the pET15-b plasmid. All of the above cloning procedures were authenticated by restriction endonuclease digestion and finally by DNA sequencing (14). The confirmed pET28a derived plasmid was designated as pET28-T7-TetO-GFP and pET15b as pET15b-IPAB-TetR plasmids.

Under this circumstance, GFP is expected to be expressed under the control of T7 promoter until formation of inclusion bodies enough for activation of P$_{IbpAB}$ promoter, which results in the production of TetR protein, and consequently inhibition of GFP production via activation of Tet-O operator on the pET28-GFP.

Chaperon plasmid

In order to verify the effect of simultaneous application of the chaperones on the production yield of recombinant GFP protein produced by the mentioned controllable system, commercial vector pKJE7 from Takara Company (Takara, Japan) was used for expression of DnaK/J/GrpE chaperon system under control of L-arabinose promoter.
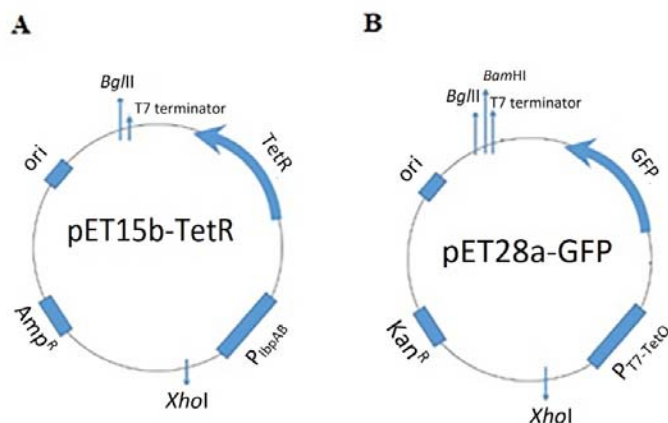
Fig. 1. The schematic presentation of the produced recombinant vectors. A, pET15b-TetR vector and B, pET28a-GFP vector

Protein expression

Co-transformation of the plasmids was performed by electroporation using Electro-competent *E. coli BL21 (DE3)* cells. Then, the cells were cultivated in the presence of suitable antibiotics in accordance with the plasmids each cell harbored. For the one-step procedure, a single colony from the cells transformed with the pET28a-GFP was used to inoculate 5 mL of LB medium supplemented with kanamycin (50 µg/mL) and incubated overnight at 37 °C and 180 rpm. The overnight culture (500 µL) was used to inoculate the 50 mL LB medium supplemented with kanamycin (50 µg/mL) and incubated in conditions mentioned above. At an optical density of 0.6 at 600 nm, GFP expression was induced by addition of 1 mM IPTG and cells were further incubated overnight at 20 °C. Expression of GFP was also optimized based on tuning of three variables, including temperature (25, 30, and 37 °C), IPTG concentration (0, 0.5, 1, and 1.5 mM), and post-induction cultivation time (2, 4, 6, 8, and 24 h) (15). For the two-step procedure, a single colony of the cells transformed with the pET28a-GFP, pET15b-TetR, and pKJE7 was cultured in 5 mL of LB medium supplemented with kanamycin (50 µg/mL), ampicillin (100 µg/mL), and chloramphenicol (25 µg/mL), and incubated overnight as mentioned above. L-arabinose (0.4 g/mL) was added first to induce chaperon system expression, and IPTG (1 mM) was added at

an optical density of 0.6 at 600 nm. Then, the cells were further incubated overnight at 20 °C. GFP expression was investigated in various combinations of pET28a-GFP with the two other plasmids (pET15b-TetR and pKJE7) (12,13). In fact, host cell was transformed with the plasmids in 3 combinations including combination 1, pET28a-GFP and pET15b-TetR, for evaluation of the effect of TetR on the GFP expression yield and soluble/insoluble ratio; combination 2, pET28a-GFP and pKJE7, for evaluation of the effect of DnaK/J/GrpE chaperon system on GFP expression yield and soluble/insoluble ratio; and combination 3, pET28a-GFP, pET15b-TetR, and pKJE7, for investigation of the two-step procedure effect on total GFP expression and soluble/insoluble ratio.

Protein purification and analysis of soluble and insoluble protein ratio

Preparation of soluble and insoluble GFP fractions were performed according to the method based on single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form (16). Briefly, the cells were harvested at 5000 *g* for 20 min at 4 °C, resuspended in 15 mL phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, $12H_2O$, 2 mM KH_2PO_4, pH 8.0), and then lysed by sonication on ice. The lysate was centrifuged at 12,000 *g* for

20 min at 4 °C. The clarified supernatant was collected as soluble fraction of GFP and retained for analysis by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and fluorescence measurement. The insoluble pellet was resuspended in 20 mL washing buffer (50 mM tris-Hcl pH 8.0, 100 mM NaCl, 1 mM EDTA, 1% Triton™ X-100, 1M urea) and then analyzed by SDS-PAGE. The pellet was washed with 20 mL washing buffer for three times, then washed with PBS buffer to remove Triton™ X-100 detergent and centrifuged at 12000 g for 20 min at 4 °C. For subsequent solubilization and refolding, the purified inclusion body pellet was resuspended in PBS buffer containing 2 M urea, frozen at -20 °C for overnight, thawed at room temperature, and centrifuged at 12,000 g for 15 min at 4 °C. Supernatant was collected as insoluble fraction of GFP and retained for analysis by SDS-PAGE and fluorescence measurement.

Fluorescence measurements

The total GFP, soluble, and insoluble GFP amounts were measured using Synergy H1 Multiple Reader (BioTek, USA) using the following settings: 37 °C, excitation wavelength 485 nm, emission wavelength 535 nm and gain 75. The GFP fluorescence values were reported as GFP fluorescence per optical density at 600 nm (13).

Statistical analyses

Data are presented as mean ± standard deviation (SD) of at least three experiments. Statistical significance was tested by two-tailed t-test and two-way ANOVA test. Analyses were performed using GraphPad Prism 8.0.2 software. $P < 0.05$ was considered significant.

RESULTS

Optimization of GFP expression in one-step procedure

Expression of the GFP protein by the transformed *E. coli* BL21 cells with pET28a-GFP plasmid was verified by fluorescent microscopy (Fig. 2A). Then, the optimized conditions for GFP expression was investigated by evaluating the effects of temperature, IPTG concentration, and post-induction incubation time. As shown in Fig. 2B, the maximal GFP expression was achieved at 25 °C, 1 mM IPTG and a post-induction incubation time of 16 h (results related to the effect of time are not presented).
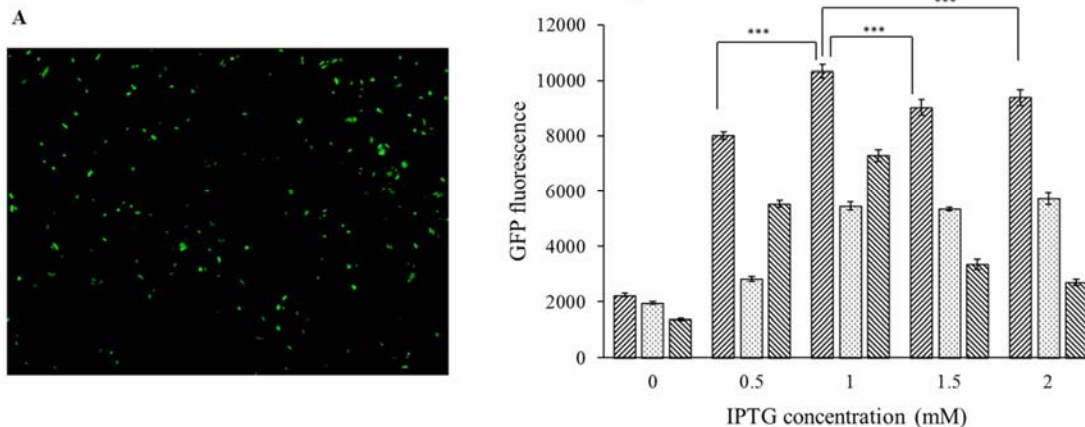
Fig. 2. GFP expression. A, Fluorescent microscopy of *Escherichia coli BL12* transformed with pET28a-GFP plasmid. B, Optimization of GFP expression in one-step procedure. The maximal GFP expression was achieved at 25 °C, 1 mM IPTG and a post-induction time of 16 h. *** Indicates significant differences between groups, $P \leq 0.001$. GFP, green fluorescent protein; IPTG, isopropyl β-d-1-thiogalactopyranoside.
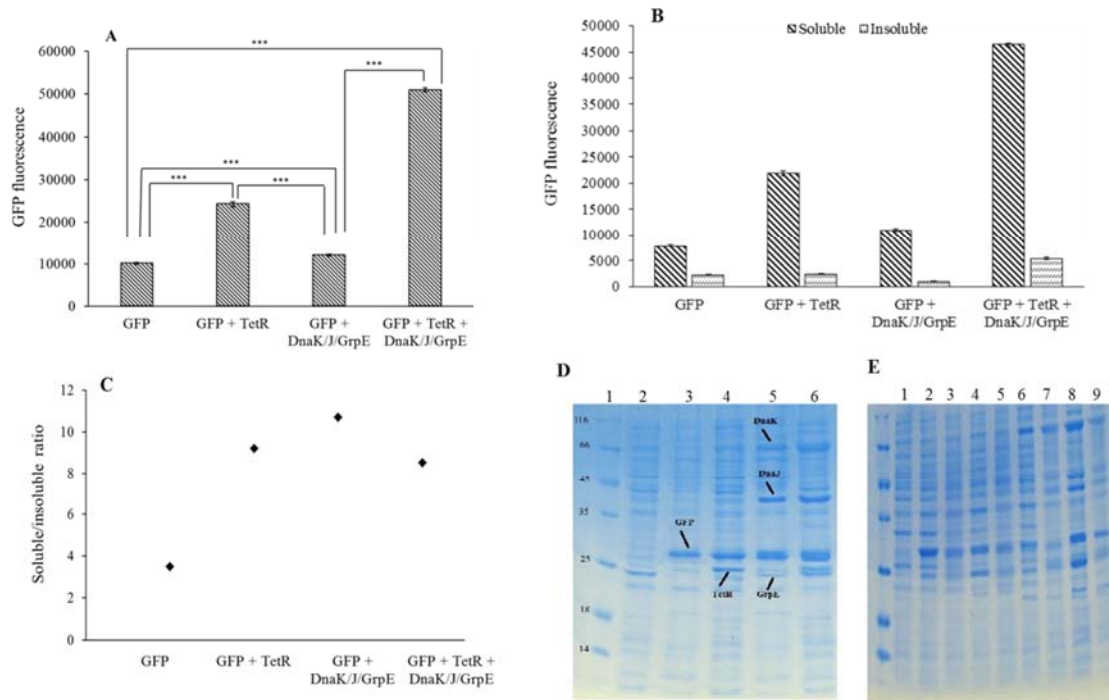
Fig. 3. Effect of two-step procedure on GFP expression. A, Effect of each plasmid combination (GFP, coexpression of GFP and TetR plasmids, coexpression of GFP and DnaK/J/GrpE plasmids, and coexpression of GFP, TetR and DnaK/J/GrpE plasmids) on total GFP expression. B, Effect of each plasmid combination on soluble and insoluble fraction of GFP. C, Soluble/insoluble ratio of GFP in the presence of each plasmid combination. D, SDS-PAGE pattern of total GFP in different combinations; M, protein marker (Thermo Scientific, #26610, USA); lane 1, *E. coli* BL21 (DE3) without any plasmid; lane 2, expression of GFP plamid; lane 3, coexpression of GFP and TetR plasmids; lane 4, coexpression of GFP and DnaK/J/GrpE plasmids; lane 5, coexpression of GFP, TetR, and DnaK/J/GrpE plasmids. E, SDS-PAGE pattern of soluble and insoluble fractions of GFP in different combination; M, protein marker (Thermo Scientific, #26610, USA); lane 1, *E. coli* BL21 (DE3) without any plasmid; lanes 2 and 3, soluble and insoluble fraction of GFP in the presence of GFP plamid; lanes 4 and 5, soluble and insoluble fraction of GFP in coexpression of GFP and TetR plasmids; lanes 6 and 7, soluble and insoluble fraction of GFP in coexpression of GFP and DnaK/J/GrpE plasmids; lanes 8 and 9, soluble and insoluble fraction of GFP in coexpression of GFP, TetR, and DnaK/J/GrpE plasmids. *** Indicates significant differences between groups, $P \leq 0.001$. GFP, green fluorescent protein; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; *E. coli*, *Escherichia coli*.

Effect of two-step procedure on GFP expression

Figure 3 shows the effect of TetR expression on GFP production. We observed GFP expression was increased 2.4 times where the feedback loop, encoded by pET15b-TetR, was induced (Fig. 3A). Expression of TetR under control of stress-sensitive IbpAB promoter translated to increased soluble/insoluble ratio of the produced GFP (2.6 times higher than the expression of GFP in the absence of TetR) (Fig. 3B,3C). As shown in Fig. 3A-3C, co-expression of DnaK/J/grpE with GFP did not exhibit notable effect on the yield of total GFP expression compared with

the two other groups but a significant increase of soluble/insoluble ratio was observed (3 times). Figure 3A-3C shows that total GFP expression was increased 5 times using the two-step expression procedure in comparison with GFP expression in the absence of TetR and DnaK/J/GrpE. On the other hand, the main fraction of total GFP was in the soluble form (soluble/insoluble ratio 8.5).

DISCUSSION

Bacterial inclusion bodies are insoluble protein aggregates formed in the bacterial cytoplasm and periplasm during the

404

overproduction of recombinant proteins, especially those from mammalian or viral origin (17,18). Factors including high copy number of target gene, strong promoter, as well as high temperature during protein expression, and high inducer concentrations favor high level expression of recombinant protein and increase the possibility of the inclusion body formation (19-22). Reducing environment of bacterial cytosol, amino acid sequence of protein with highly hydrophobic proteins, lack of post-translational modification, and eukaryotic chaperones also favor this event (23). Inclusion bodies formation imposes a great obstacle in the production and downstream processing of recombinant proteins (24). Solubilization and refolding of inclusion bodies into bioactive form is a challenging task. If an efficient and convenient way of decreasing formation of inclusion bodies without decreasing protein production yield could be developed, it would reduce the need of excess chromatographic purification steps and downstream processing costs (25). At the midstream level, co-expression of folding modulators such as chaperones, lowering the transcription rate, and reducing the growth temperature have been thoroughly explored strategies which have been successful for an important number of specific proteins but they do not offer generic solutions to this matter (26-28). Co-expression of molecular chaperones and foldases from accessory plasmids with no further control results in imbalanced expression of these proteins and has detrimental effects on the production process (29,30). Dragosits *et al.* developed a synthetic expression platform that enables the cell to shut down the recombinant protein production mechanism automatically by itself through stress-induced feedback to address challenges such as balancing protein production and cellular stress. This system limited stressful protein expression and increased the total soluble fraction but significantly decreased the protein yield. The authors suggested further investigation on promoter and repressor engineering, other recombinant proteins, and different inducer concentrations for avoiding such loss (13).

In another study, Marco *et al.* used a two-step procedure that shows the strongest enhancement of the solubility. In the first step of this procedure the four chaperon systems were coordinately co-overproduced with recombinant proteins to optimized folding and in the next step protein biosynthesis was inhibited by removal of IPTG to permit chaperone mediated refolding of the misfolded and aggregated proteins *in vivo*. This strategy increased the solubility of 64 different heterologous proteins up to 42 folds without decreasing total protein yield (12). Disadvantage of this method is requirement of external intervention to inhibit protein expression and an impractical and expensive process in industrial scale. We combined the two methods with some modifications. We substituted pNf-TetR with pET15b-TetR and pET23b-GFP with pET28a. In this study, both soluble/insoluble ratio and total GFP expression was increased (2.5 and 5 times, respectively), while in the study of Dragosits *et al.* soluble/insoluble ratio was increased up to 1.5 times but total GFP expression was decreased 1.4 times as explained earlier. This may be due to the use of pET15b vector for cloning of TetR, which retains similar copy number to pET28a as carrier of GFP encoding sequence in comparison with pNF. Second reason may be related to the use of DnaK/J/GrpE chaperon system and increasing of chaperon capacity of bacterial cells which resulted in folding of misfolded and aggregated GFP, suppression of stress-sensitive IbpAB promoter and excessive expression of TetR as suppressor of GFP expression and resuming of its expression. It would be interested to test the efficacy of this system with other recombinant proteins.

CONCLUSION

Altogether, in the present study, the use of an automatic negative feed-back system, combined with co-expression of major cytosolic chaperon system DnaK/J/GrpE, resulted in an increased solubility and total recombinant protein expression. Therefore, this system is very applicable,

beneficial, and cost-effective in recombinant protein expression field of biotechnology. Further studies on using this system for expression of various recombinant therapeutic proteins are undergoing in our laboratory.

ACKNOWLEDGMENTS

REFERENCES

1. Ben-Nissan G, Vimer S, Warszawski S, Katz A, Yona M, Unger T, *et al*. Rapid characterization of secreted recombinant proteins by native mass spectrometry. Commun Biol. 2018;1(1):213-224.
2. Kyratsous CA, Silverstein SJ, DeLong CR, Panagiotidis CA. Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*. Gene. 2009;440(1-2):9-15.
3. Chen R. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol Adv. 2012;30(5):1102-1107.
4. Panda AK. Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*. Adv Biochem Eng Biotechnol. 2003;85:43-93.
5. Yang Z, Zhang L, Zhang Y, Zhang T, Feng Y, Lu X, *et al*. Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. PLoS One. 2011;6(7):e22981.
6. Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol. 2013;14(10):630-642.
7. Calloni G, Chen T, Schermann SM, Chang HC, Genevaux P, Agostini F, *et al.* DnaK functions as a central hub in the *E. coli* chaperone network. Cell Rep. 2012;1(3):251-264.
8. Aguilar-Rodríguez J, Sabater-Muñoz B, Montagud-Martínez R, Berlanga V, Alvarez-Ponce D, Wagner A, *et al*. The molecular chaperone DnaK is a source of mutational robustness. Genome Biol Evol. 2016;8(9):2979-2991.
9. Kedzierska S, Staniszewska M, Wegrzyn A, Taylor A. The role of DnaK/DnaJ and GroEL/GroES systems in the removal of endogenous proteins aggregated by heat-shock from *Escherichia coli* cells. FEBS Lett. 1999;446 (2-3):331-337.
10. Gopal GJ, Kumar A. Strategies for the production of recombinant protein in *Escherichia coli*. Protein J. 2013;32(6):419-425.
11. Xiao S, Shiloach J, Betenbaugh MJ. Engineering cells to improve protein expression. Curr Opin Struct Biol. 2014;26:32-38.
12. de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnol. 2007;7:32-40.
13. Dragosits M, Nicklas D, Tagkopoulos I. A synthetic biology approach to self-regulatory recombinant protein production in *Escherichia coli*. J Biol Eng. 2012;6(1):2-11.
14. Jahanian-Najafabadi A, Soleimani M, Azadmanesh K, Mostafavi E, Majidzadeh AK. Molecualr cloning of the capsular antigen F1 of Yersinia pestis in pBAD/gIII plasmid. Res Pharm Sci. 2015;10(1): 84-89.
15. Long X, Gou Y, Luo M, Zhang S, Zhang H, Bai L, *et al*. Soluble expression, purification, and characterization of active recombinant human tissue plasminogen activator by auto-induction in *E. coli*. BMC Biotechnol. 2015;1(15):13-21.
16. Qi X, Sun Y, Xiong S. A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. Microb Cell Fact. 2015;14:24-35.
17. Garcia-Fruitos E. Inclusion bodies: a new concept. Microb Cell Fact. 2010;9:80-82.
18. Rodriguez-Carmona E, Cano-Garrido O, Seras-Franzoso J, Villaverde A, Garcia-Fruitos E. Isolation of cell-free bacterial inclusion bodies. Microb Cell Fact. 2010;9:71-79.
19. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. Microb Cell Fact. 2015;14:41-50.
20. Carrio MM, Villaverde A. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J Bacteriol. 2005;187:3599-3601.
21. Chang CC, Song J, Tey BT, Ramanan RN. Bioinformatics approaches for improved recombinant protein production in *Escherichia coli*: protein solubility prediction. Brief Bioinform. 2014;15(6):953-962.
22. Rosano GL, Morales ES, Ceccarelli EA. New tools for recombinant protein production in *Escherichia coli*: A 5-year update. Protein Sci. 2019;28(8):1412-1422
23. Carrio MM, Cubarsi R, Villaverde A. Fine architecture of bacterial inclusion bodies. FEBS Lett. 2000;471(1):7-11.
24. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:172-188.
25. Salehinia J, Sadeghi HMM, Abedi D, Akbari V. Improvement of solubility and refolding of an anti-human epidermal growth factor receptor 2

single-chain antibody fragment inclusion bodies. Res Pharm Sci. 2018;13(6):566-574.

26. Garcia-Fruitos E, Sabate R, de Groot NS, Villaverde A, Ventura S. Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J. 2011;278(14):2419-2427.

27. Liu L, Yang H, Shin HD, Chen RR, Li J, Du G, *et al*. How to achieve high-level expression of microbial enzymes: strategies and perspectives. Bioengineered. 2013;4(4):212-223.

28. Martínez-Alonso M, González-Montalbán N, García-Fruitós E, Villaverde A. Learning about protein solubility from bacterial inclusion bodies. Microb Cell Fact. 2009;8(1):4-8.

29. Martínez-Alonso M, García-Fruitós E, Ferrer-Miralles N, Rinas U, Villaverde A. Side effects of chaperone gene co-expression in recombinant protein production. Microb Cell Fact. 2010;9(1): 64-69.

30. Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, *et al*. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact. 2008;7(1):11-28.