



Gene Multimerization in Expression Vector: A Potential Strategy for Enhanced Protein Expression in *E. coli*

Aadil Sultan, Mohsin Ahmad Khan*, Nadeem Ahmed, Muhammad Hassan, Rashid Bhatti, Hafsa Naeem, Muhammad Islam Khan, Saad Tahir, Samia Afzal* and Ahmad Ali Shahid

Center of Excellence in Molecular Biology, 87-West Canal Bank Road, Thokar Niaz Baig, University of the Punjab, Lahore, Pakistan

ABSTRACT

Protein production in any expression system can be enhanced either by optimizing the culturing parameters or targeting the genetic factors that enhance protein production. One of those genetic factors is gene dosage, which can be increased by increasing the vector copy number. However, increased number of expression vector poses some metabolic burdens on bacterial cells. The current study provides gene multimerization strategy as an alternate way to amplify the gene dosage and explores its effect on the expression of human epidermal growth factor (hEGF) by constructing three types of expression vectors i.e., pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3, each containing single, double and triple expression cassettes, respectively. These vectors were transformed in *E. coli* strain Rosetta-gami 2 (DE3) to develop three different bacterial populations. Expression of human epidermal growth factor was analyzed by SDS-PAGE and subsequent densitometric analysis was done using ImageJ software. The difference in protein expression of all three populations was analyzed by using one-way ANOVA. Clones with double and triple copies of genes showed an increased expression up to 3.8 and 3.1 folds as compare to the clones having single copy of gene. Gene multimerization strategy can be used to enhance production of recombinant proteins in bacterial expression system.

Article Information

Received 12 November 2021
Revised 28 December 2021
Accepted 19 January 2022
Available online 11 May 2022
(early access)
Published 22 February 2023

Authors' Contribution

MAK conceived the idea of study. AS, MH, HN, MIK and ST helped in lab work. NA and RB wrote the manuscript. SA and AAS proof read and finalized the article.

Key words

Expression system, Gene multimerization, Recombinant proteins, Bioprocessing, Gene dosage

INTRODUCTION

Bacterial expression system has been used for research and commercial based production of both prokaryotic and eukaryotic proteins. Bacterial strain, *E. coli* has been used to produce biopharmaceutical products such as insulin, interferon, human growth factors and other recombinant proteins (Tripathi, 2016). Bacterial strain *E. coli* offers different advantages, such as easy to grow, high growth rate and cheap media demands. Moreover, its cellular components and genetics are well studied, which make it a host, that can bear fruitful genetic manipulations for expression of recombinant proteins (Joseph *et al.*, 2015; Rosano and Ceccarelli, 2014; Sivashanmugam *et al.*, 2009; Tripathi, 2016). However, in spite of all these advantages,

protein production in bacterial expression system, has never been a straight forward method. Certain disadvantages such as lack of post-translational modifications, formation of inclusion bodies, protein toxicity etc. may result in production of proteins either with improper structure and function or with no or very low yield (Rosano and Ceccarelli, 2014). Different methodologies have been developed to target those factors that might be responsible for the problem of low yield in bacterial expression system. Briefly, these methodologies can be broadly classified as (i) those dealing with optimization of bioprocessing and culture conditions and (ii) those that deal at molecular level for enhancement of protein production, such as strain engineering, increased plasmid number, codon optimization and genetic manipulations of certain genetic elements.

Optimization of bioprocessing and culturing conditions include several strategies. For enhanced production of recombinant proteins, usually cultures with high cell densities are preferred in batch and fed-batch fermentation processes. Cultures can either be induced or grown in auto-induction media for enhanced production of recombinant proteins in a regulated way (Sivashanmugam *et al.*, 2009). Different media compositions, have been studied and shown to affect the protein expression level.

* Corresponding author: samiaraza@live.com, mohsin@cemb.edu.pk
0030-9923/2023/0003-1051 \$ 9.00/0



Copyright 2023 by the authors. Licensee Zoological Society of Pakistan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Specific media formulations can be screened for improved protein expression (Broedel *et al.*, 2001). Besides that, different recombinant proteins such as hIGF (Zhang *et al.*, 2010), neutrophil-acting protein (Lu *et al.*, 2015), Fab antibody fragments (Ukkonen *et al.*, 2013), Phenylalanine Ammonia lyase (Jaliani *et al.*, 2014) and many others require optimization of different culturing conditions such as temperature, pH, aeration, concentration of inducer and post-induction period can also be optimized for enhanced protein expression. Selection of appropriate host for recombinant protein expression is also one of the crucial parameters that might depend upon the type of protein being produced. Strain engineering has been extensively reviewed. Targeted and whole genome based mutagenesis have been used for development of host strains, such as Origami, Rosetta and many others for enhanced protein expression. Specific strains have been developed to carry out post-translational modifications, such as disulfide-bridge formation, acetylation and glycosylation, hence leading to proper folding of recombinant protein and ultimately improved protein expression (Makino *et al.*, 2011). Though the process of optimization improves the protein yield and acts as bridge between lab and large-scale production, however, the process is time taking, tedious and needs a systematic study for every culturing parameter.

Apart from above mentioned techniques, different techniques of molecular biology have been used to enhance the heterologous protein expression. The transcriptional rate for any heterologous gene expression can be related to the strength of the promoter. Different promoters, such as T7, *lac*, *trp*, *lacUV5*, *recA*, *tetA*, etc. have been used to produce different recombinant proteins in *E. coli* (Joseph *et al.*, 2015). The strength of promoters need to be experimentally determined via comparative studies and can be considered as a function of its nucleotide sequences (Li and Zhang, 2014).

The translational rate can also be a considered as a critical factor that might affect the protein expression. The translational rate of any recombinant protein might be affected by its mRNA secondary structures and the codon biasness of host. The 5' end of mRNA contains specific genetic elements such as initiation codon, Shine-Dalgarno sequences and enhancers that affect the rate of translation (Vimberg *et al.*, 2007). Codon biasness can be overcome using codon optimization. In this modern era of synthetic biology, the codons that cannot be easily translated by the translation machinery of the host, are replaced by the ones that favor the efficient protein translation, hence overcoming the problem of codon biasness (Burgess-Brown *et al.*, 2008; Elena *et al.*, 2014). Besides that, vectors containing genes for expression of rare tRNAs can also be co-expressed in host cells to avoid the problem of codon biasness. This has been practiced using the

Rosetta strains of *E. coli* (Novy *et al.*, 2001). Use of some universal translation initiation tags might also improve the translation initiation rate and hence improved protein expression (Parret *et al.*, 2016; Vimberg *et al.*, 2007).

Gene dosage is another critical factor that might contribute to increase protein expression. The effect of gene dosage has been well studied in *Pichia pastoris* expression system, which provides opportunities for inserting more than one gene copy number in its genome, thus generating multicopy clones (Aw and Polizzi, 2013). The effect of gene dosage on expression of several recombinant proteins such as IFN α 2b, insulin precursor, etc. has been studied in *Pichia pastoris* expression system (Khan *et al.*, 2014; Mansur *et al.*, 2005). The gene dosage in bacterial expression system ultimately depends upon the copies of expression vector, as each vector contains single gene. For such studies, different plasmids having high, medium and low copy numbers are used (Chakravarty and Cronan, 2015; Friehs, 2004). The effect of increased gene dosage to enhance the protein expression, is usually studied by increasing the plasmid copy number as has been done for production of recombinant proteins in *Leuconostoc citreum* (Son *et al.*, 2016). However, the apparently simple relationship between plasmid copy number and protein expression does not always correlate with each other. A high plasmid copy number poses additional difficulties on the strain for its maintenance by saturating the replication machinery. Besides that, increased copy number has been known not only to affect the metabolism of strain but also affects its growth rate and cell physiology. The effects of high copy number have been extensively reviewed (Silva *et al.*, 2012). In current study, we have reported for the first time an alternate way for studying the effect of increased gene dosage by increasing gene copies per plasmid keeping the plasmid copy number unaltered for enhanced protein production in bacterial expression system. The strategy has been named as gene multimerization in expression vector of bacterial expression system. An expression cassette containing T7 promoter, 6X His-tag and gene for human epidermal growth factor has been prepared in the backbone of pET28 a (+) expression vector. The expression cassette has been multimerized to prepare the expression vectors containing double and triple copies of expression cassettes. Three different populations of clones of *E. coli* strain Rosetta-gami 2 (DE3) containing expression vector with single, double and triple copies of expression cassettes have been generated and comparative protein expression studies have been carried out to analyze the effect of gene multimerization on protein expression level.

MATERIALS AND METHODS

Strains and plasmids

E. coli Top10 F' cells were used for routine cloning

experiments. *E. coli*. Rosetta *gami*^{TM2} (DE3) (Novagen) was used as expression strain for protein expression studies. pCR^{TM2}.1 Vector was used for gene cloning and pET28a (+) (Novagen) was used as expression vector.

Enzymes and culture conditions

T4 DNA ligase of Rapid DNA ligation Kit (Thermo Scientific, Catalog # K1422) was used at every ligation step. *Bam*H I, *Bgl* II and *Nde* I were used for restriction digestion. All these enzymes were used from Thermo Scientific. Calf Intestine Alkaline Phosphatase (CIAP), (Merck) was used to prevent self-ligation where necessary. Culturing of cloning and expression strains was carried out in Lauria-Bertani (LB) medium at 37°C and 200rpm, containing appropriate drugs as described in a section given below.

Multimerization of hEGF and hIGF in expression vectors

The genes *hegf* derived from Huh cell line, was PCR amplified using gene specific primer pairs from constructs prepared previously. Gene specific primers used in this study were *Nde*I-EGF-F and *Bam*H1-EGF-R with sequences of 5'CATATGAATAGTGACTCTGAATGTCCCCTG T3' and 5'GGATCCTTAGCGCAGTTCCCACCACTTC3' respectively. The forward and reverse primers contained restriction sites for *Nde*I and *Bam*H1, respectively. The amplified PCR product was ligated to PCR 2.1 vector and sequenced using M13 Forward and M13 Reverse primer pairs. The insert (*hegf*) was released using enzymes *Nde*I and *Bam*H1 and ligated to expression vector pET28 a (+) for preparing single copy constructs i.e., pET28-EGF-C1 and pET28-IGF-C1. For preparing double copy construct, i.e., pET28-EGF-C2, having two copies of genes, the inserts from single copy constructs were released using enzymes *Bam*H1 and *Bgl* II and ligated back to already prepared single copy constructs, linearized with *Bam*H1 only. The prepared double copy constructs were analyzed for successful preparation via restriction digestion with *Bam*H1 and *Bgl* II. Moreover, it also served as a strategy for analyzing whether both the copies of genes were in desired orientation i.e. in head to tail fashion having promoter at 5' end and terminator at 3' end. The double copy constructs of both genes were further linearized with *Bam*H1, to which the inserts released from single copy constructs using enzymes *Bam*H1 and *Bgl* II, were ligated, to prepare triple copy construct i.e., pET28-EGF-C3. The triple copy constructs were analyzed by digestion with enzymes *Bam*H1 and *Bgl* II. The constructs linearized with *Bam*H1 only, were given CIAP treatment to prevent self-ligation (Fig. 4).

Expression studies of prepared constructs in Rosetta-gami^{TM2} (DE3)

To study the effect of gene multimerization on protein expression, the prepared constructs for both the genes i.e., *hegf* and *hegf*, were transformed into *E. coli*. Rosetta-*gami*^{TM2}(DE3), to obtain three different populations of clones containing single, double and triple copy constructs, respectively. Expression studies for single, double and triple copy constructs were carried out in three *iNde*p*Nde*nt experiments. Five clones from among each population were selected for protein expression studies. Each clone was cultured in 5 ml LB media containing tetracycline (12.5 µg/ml), streptomycin (50 µg/ml), kanamycin (50 µg/ml) and chloramphenicol (34.5 µg/ml), for overnight growth at 37°C. OD₆₀₀ of the overnight grown cultures was measured and volume was calculated to dilute the cultures in 10 ml fresh LB media containing already mentioned drugs, such that their OD₆₀₀ was maintained to be 0.1. OD₆₀₀ was measured at regular intervals of two h. As its values reached above 0.6, cultures were induced with 1mM isopropyl-β-d-thiogalactopyranoside (IPTG). Cultures were grown for 4 h of post-induction period and harvested by centrifugation at 6000 rpm for 5 min. 1 ml culture was separated before induction to be used as negative control for subsequent studies.

Protein expression analysis

The cells, harvested, were suspe*Nde*d in 500 µl of 1X PBS, pH 7.3 (Oxoid), to which equal volume of 2X protein sample buffer (Tris-Cl, pH 6.8, 120mM, Glycerol 10%, SDS 2%, Bromophenol blue 0.05% and β-mercaptoethanol 5%) was added. Heat shock was given at 100°C for 10 min which was followed by centrifugation at 13,000 rpm, 4°C for 10 min. Supernatant was analyzed on 15% polyacrylamide gel via sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein bands were visualized by staining the gels with Coomassie Brilliant R-250. Bands, at appropriate size, having reduced intensity or absent in negative control were considered to be that of interest. Image J software was used for densitometric analysis. The band appeared with the unaltered intensity in all the gels, was used for normalization of our bands of interest. Western blotting was performed for validating the expression of recombinant hEGF and hIGF. For this purpose, the protein samples of single, double and triple copy clones, showing maximum expression were resolved in 15% SDS-polyacrylamide gel and bands were transferred to nitrocellulose blotting membrane (AmershamTMProtanTM 0.45µm NC). Followed by overnight blocking in 5% skim milk (5% w/v in PBS), membrane was incubated with mouse anti-histidine antibody diluted as 1: 1000, for three h. After thorough washing with PBST (0.05% Tween 20 in PBS),

membrane was incubated for one hour with AP conjugated goat-anti mouse IgG (Santa Cruz Biotechnology, Inc.) diluted as 1: 7500. Bands were visualized by applying NBT/BCIP (Sigma-Aldrich) solution.

Plasmids structural stability

The prepared constructs for both genes, were analyzed for structural stability as well. Plasmid DNA was isolated by using Plasmid DNA Extraction Mini Kit (Favorgen Biotech Corp.) from clones containing single, double and triple copy constructs, after overnight growth of post-induction period. The isolated plasmid DNA was subjected to restriction digestion analysis by enzymes *Bam*HI and *Bgl* II.

Statistical analysis

The peak areas obtained by densitometric analysis were further statistically analyzed for each population. The mean expression level in selected five clones for each population of single, double and triple constructs, were analyzed by ANOVA to determine whether the difference in protein expression in these populations was significant. One sample Kolmogorov-Smirnov test and test of homogeneity of Variances was performed to check if the data was fulfilling the criteria for ANOVA.

RESULTS

Multicopy expression vectors and molecular confirmation gene multimerization.

The gene *hegf* was successfully amplified by PCR and ligated to pET28a (+) to prepare single copy construct i.e pET28-EGF-C1. An expression cassette of about ~340bp having T7 promoter, 6 His Tag and *hegf* with its stop codon at 3' end was released by its digestion with enzymes *Bam*H1 and *Bgl* II and was used to prepare double copy pET28-EGF-C2 and triple copy pET28-EGF-C3 constructs. Each prepared expression vector was analyzed by restriction digestion with *Bam*H1 and *Bgl*III. Release of inserts of expected size i.e., ~ 700 bp and ~1000 bp clearly demonstrates the successful preparation of expression vectors containing double and triple copy constructs with desired orientation of expression cassettes i.e. in head to tail fashion having promoter at 5' end and terminator at 3' end. (See supplementary material for results of restriction digestion analysis). These results indicated successful construction of all three recombinant constructs required to study the effect of gene multimerization in *E. coli*.

Protein expression studies

The single, double and triple copy constructs for *hegf* were transformed and expressed in *E. coli* strain Rosetta-

gamiTM2 (DE3). Five clones for each of the constructs were selected for protein expression analysis by PAGE. To maintain the similar cell densities for comparative expression studies, cells were induced as their OD₆₀₀ reached just above 0.6 and were allowed to grow for a post-induction period of 4 h. The total cell lysate was analyzed on 15% Tris-Glycine polyacrylamide gel (Fig. 1). The un-induced sample for each construct was used as a control. The protein band appearing at appropriate size i.e., ~7.3 kDa was considered to be the protein of interest which was absent or had leaky expression in un-induced sample. The intensity of the bands obtained was analyzed by densitometric analysis using image J software. The ratio of intensity of bands of *hEGF* to that of bands of protein appearing at size above 20kDa, in all gels, showing uniform intensity in all clones for each construct was considered as normalized protein expression. The clones having double copy constructs i.e., pET28-EGF-C2 showed an overall highest expression in all three populations (Fig. 2). The average protein expression in clones having pET28-EGF-C2 and pET28-EGF-C3 was increased up to almost 4 and 3 folds respectively as compared to the protein expression in clones having single constructs i.e., pET28-EGF-C1 (Table I).

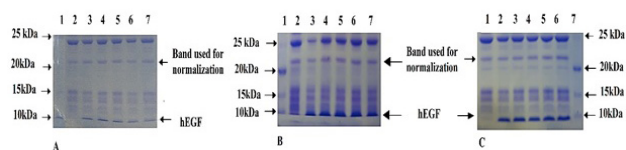


Fig. 1. Expression analysis of hEGF on 15% polyacrylamide gel. **A**, Expression of hEGF for single copy construct. Lane 1, PAGE Ruler TM unstained protein ladder. Lane 2, Un-induced sample. Lanes 3 to 7 show induced samples having bands of hEGF slightly below than 10kDa are clearly visible. **B**, Expression of hEGF for double copy construct. Lane 1, PAGE Ruler TM unstained protein ladder. Lane 2: Un-induced sample. Lanes 3 to 7: Induced samples. Bands of hEGF with much bright intensity are clearly visible. **C**, Expression of hEGF for triple copy construct. Lane 7: PAGE Ruler TM unstained protein ladder. Lane 1, Un-induced sample. Lane 2 to 6, Induced samples. Bands of hEGF are visible. Intensity of bands were analyzed by densitometric analysis.

Expression confirmation by western blotting

The clones showing highest protein expression from each of the population were selected for confirmation of expression of hEGF by Western blotting. Each expression cassette in clones having pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3 has a sequence of His Tag between T7 promoter and *hegf* gene sequence. Therefore, the protein expression was validated by using mouse anti-histidine antibodies followed by their detection with AP conjugated

goat anti-mouse antibodies. Reddish brown bands appeared, that demonstrated the successful expression of human epidermal growth factor in clones having single, double and triple copy constructs (Fig. 3). Western blotting results further confirms that all three types of clones carrying single and multimerized genes are actively expressing recombinant hEGF. One way Analysis of Variance (ANOVA) was performed using SPSS (Statistical Package for Social Sciences) to consider whether the variations in expression level of hEGF across the three types of populations, i.e., clones having pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3 was significant. Before applying ANOVA it was ensured by One-Sample Kolmogorov-Smirnov Test that data was following normal distribution. Moreover, tests of Homogeneity Variance taking p value of 0.05 also confirmed that variations within each population were not significant. For ANOVA the level of confidence was taken as 0.01. The significance obtained by ANOVA was 0.000 which was below the selected level of confidence, hence indicating that variation in protein expression across the populations i.e., single, double and triple copy clones was significant (Table II).

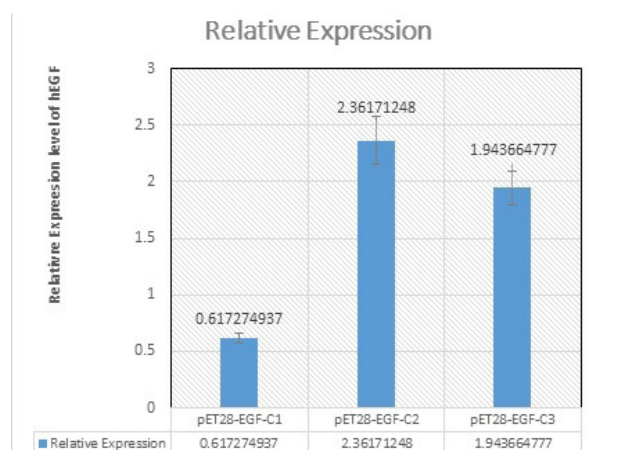


Fig. 2. Showing relative expression level of human epidermal growth factor in 3 different populations of Rosetta-gami™ 2 (DE3) strain having single, double and triple copies of expression cassette. The mean value for expression level in clones containing pET28-EGF-C1 was found to be 0.61 ± 0.04 . The mean value for expression level in clones containing pET28-EGF-C2 was found to be 2.36 ± 0.21 which is 3.8 folds higher as compare to their counterparts having single expression cassette. The clones containing expression vector with three expression cassettes i.e. pET28-EGF-C3 showed the mean expression level of 1.94 ± 0.154 which is 3.1 folds higher as compare to their single copy counterparts. The error bars indicate the standard error for each population.

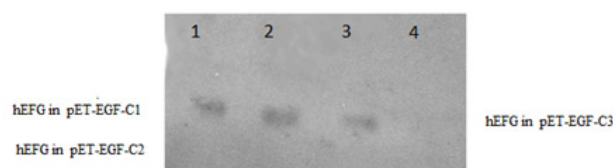


Fig. 3. Western blotting for confirmation of expression of human epidermal growth factor in clones having expression cassettes. **Lanes 1-3**, Bands appeared on nitrocellulose membrane after NBT/BCIP treatment showing successful expression of hEGF in clones containing pET-EGF-C1, pET-EGF-C2 and pET-EGF-C3 respectively. **Lane 4**, Un-induced sample used as negative control. No band appeared for negative control.

Table I. Normalized protein expression for clones having pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3 constructs. The average values are shown along with standard errors.

Clones	Normalized single copy expression	Normalized double copy expression	Normalized triple copy expression
1	0.61	2.94	2.14
2	0.62	1.68	2.02
3	0.71	2.58	2.35
4	0.65	2.14	1.68
5	0.46	2.43	1.50
Average	0.61 ± 0.04	2.36 ± 0.21	1.94 ± 0.15

Table II. Statistics of one way ANOVA. The significant value of 0.000 shows that these three data sets (expression level of hEGF in clones having pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3) have a significant variation between their means.

	Sum of squares	df	Mean square	F	Sig.
Between groups	8.29	2	4.14	35.43	0.000
Within groups	1.40	12	0.11		
Total	9.70	14			

Plasmid stability

Structural stability of prepared constructs i.e., pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3 was very important to be analyzed. The clones of *E. coli* strain Rosetta-gami™ 2(DE3) having single, double and triple copy constructs were allowed to go through the whole process of culturing and post-induction period of 16 h. The plasmid DNA extracted from the clones was analyzed

by digestion with *Bam*HI and *Bgl*III on 2% agarose gel. Detection of inserts of ~340bp, ~700bp and ~1000bp from pET28-EGF-C1, pET28-EGF-C2 and pET28-EGF-C3 (supplementary material), confirmed that the prepared constructs were structurally stable during the expression studies.

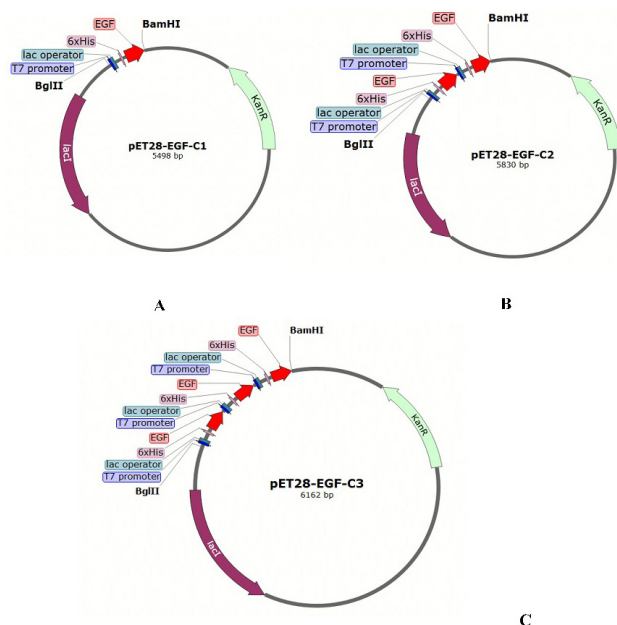


Fig. 4. Maps of single, double and triple copy constructs prepared for the expression of human epidermal growth factor by multimerization of *hegf* gene. The expression cassette of about ~340bp was released by digestion with *Bgl*II and *Bam*HI from single copy construct Figure 1A and inserted back to its *Bam*HI linearized form for preparation of double copy construct i.e., pET28-EGF-C2 as shown in Figure 1B. The same expression cassette was inserted to the backbone of double copy construct, linearized with *Bam*HI, for preparation of triple copy constructs i.e. pET28-EGF-C3 as shown in Figure 1C. Sequential steps repeated in same pattern can generate constructs with multiple copies of genes. Sequence maps were generated using SnapGene software version for Windows.

DISCUSSION

Bacterial expression system has been used for production of several recombinant proteins. Two major approaches can be used for enhancing the protein production. The first approach includes such methodologies that deal with optimization of culturing parameters such as pH, temperature, aeration, media formulations, concentration of inducer, use of appropriate strain etc. (Jaliani *et al.*, 2014; Ranjbari *et al.*, 2015; Sivashanmugam *et al.*, 2009). On the other hand, second approach utilizes

such methodologies that deal at molecular level for improved protein expression. Molecular techniques manipulate certain genetic elements which may increase the protein expression level either at transcriptional or translational level (Makino *et al.*, 2011). Synthetic biology enables us to create and use strong promoters such as T7, *trc*, etc. that allow us to achieve high expression level by increased transcriptional rate (Tegel *et al.*, 2011). Besides, optimization of translational initiation region and codon optimization has also been utilized to overcome translational difficulties and codon biasness for enhanced protein production in a particular host (Tegel *et al.*, 2011). In addition to that the effect of gene dosage has also been a crucial factor for protein expression level. Gene dosage has been studied in yeast as well as in bacterial expression system. Multiple copies of genes can be inserted in genome of yeast to study the effect of gene dosage for enhanced protein expression (Liu *et al.*, 2020). However, gene dosage in bacterial expression system directly relies on expression vector copy number. Usually, plasmid copy number is increased for increasing the gene dosage in bacteria. This poses additional difficulties in growth and physiology of bacteria. The current study provides an alternate strategy for increasing the gene dosage without altering the copy number of expression vector (Kittleson *et al.*, 2011).

In this study, we prepared an expression cassette having T7 promoter, 6X His-tag and gene for human epidermal growth factor in the backbone of pET28a (+) expression vector. The expression cassette was successfully multimerized to prepare the expression vectors with two and three copies of expression cassettes (see Fig. 1). The prepared constructs were transformed into *E. coli* strain Rosetta-gami 2 (DE3) and expression level of human epidermal growth factor was checked by SDS-PAGE for all three types of bacterial populations having single, double and triple copy constructs as shown in Figure 2. The intensities of protein bands were analyzed using ImageJ and the normalized protein expression of clones having double and triple copies of expression cassette was compared with that of the clones having single copy of it. One way ANOVA clearly demonstrated that a significant difference in the expression level of human epidermal growth factor was present across the three types of populations (see Table II). As shown in graph 01, the expression level was increased up to 3.8 and 3.1 folds in clones having double and triple copies of expression cassettes respectively. The three expression cassettes might have led to the saturation of host metabolic machinery due to which decrease in expression level has been observed. The current study is an initiative to explore the effect of gene multimerization in bacterial expression system, which can be a useful strategy

for enhancing the production of several other recombinant proteins in bacterial expression system. However, a more systematic study at genomic and transcriptomic level can be carried out to analyze the extent of linear relationship between gene dosage and protein expression level which can increase its potential of enhanced protein production not only at lab scale but can facilitate industrial processes in an economical way.

CONCLUSION

In conclusion, we have successfully used gene multimerization strategy for preparation of structurally stable expression constructs containing single, double and triple copies of expression cassettes and studied the effect of gene dosage on protein expression without altering the plasmid copy numbers. We consider it worth considerable strategy for enhanced protein expression in bacterial expression system.

ACKNOWLEDGEMENTS

We are thankful to Centre of Excellence in Molecular Biology, University of The Punjab, Lahore, Pakistan for providing facilities for this study.

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

- Aw, R., and Polizzi, K.M., 2013. Can too many copies spoil the broth? *Microb. Cell Factor.*, **12**: 128. <https://doi.org/10.1186/1475-2859-12-128>
- Broedel, S., Papiak, S.M., and Jones, W.R., 2001. *The selection of optimum media formulations for improved expression of recombinant proteins in E. coli*. Athena Enzyme Systems Technical Bulletin.
- Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O., 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr. Purif.*, **59**: 94-102. <https://doi.org/10.1016/j.pep.2008.01.008>
- Chakravarty, V., and Cronan, J.E., 2015. A series of medium and high copy number arabinose-inducible *Escherichia coli* expression vectors compatible with pbr322 and pacyc184. *Plasmid*, **81**: 21-26. <https://doi.org/10.1016/j.plasmid.2015.03.001>
- Elena, C., Ravasi, P., Castelli, M.E., Peirú, S., and Menzella, H.G., 2014. Expression of codon optimized genes in microbial systems: Current industrial applications and perspectives. Recombinant protein expression in microbial systems. *Front. Microbiol.*, **5**: 21. <https://doi.org/10.3389/fmicb.2014.00021>
- Friebs, K., 2004. *Plasmid copy number and plasmid stability. New trends and developments in biochemical engineering*. Springer. pp. 47-82. <https://doi.org/10.1007/b12440>
- Jalilani, H.Z., Farajnia, S., Safdari, Y., Mohammadi, S.A., Barzegar, A., and Talebi, S., 2014. Optimized condition for enhanced soluble-expression of recombinant mutant *Anabaena variabilis* phenylalanine ammonia lyase. *Adv. Pharma. Bull.*, **4**: 261.
- Joseph, B.C., Pichaimuthu, S., Srimeenakshi, S., Murthy, M., Selvakumar, K., Ganesan, M., and Manjunath, S.R., 2015. An overview of the parameters for recombinant protein expression in *Escherichia coli*. *J. Cell Sci. Ther.*, **6**: 1. <https://doi.org/10.4172/2157-7013.1000221>
- Khan, M.A., Hassan, N., Ahmad, N., Khan, M.I., Zafar, A.U., Khan, F., and Husnain, T., 2014. Studies to analyse the relationship between *ifna2b* gene dosage and its expression, using a *Pichia pastoris* based expression system. *Yeast*, **31**: 13-28. <https://doi.org/10.1002/yea.2990>
- Kittleson, J.T., Cheung, S., and Anderson, J.C., 2011. Rapid optimization of gene dosage in *E. coli* using dial strains. *J. Biol. Eng.*, **5**: 10. <https://doi.org/10.1186/1754-1611-5-10>
- Li, J., and Zhang, Y., 2014. *Relationship between promoter sequence and its strength in gene expression*. arXiv Preprint arXiv, pp. 14052726. <https://doi.org/10.1140/epje/i2014-14086-1>
- Liu, J., Han, Q., Cheng, Q., Chen, Y., Wang, R., Li, X., Liu, Y., and Yan, D., 2020. Efficient expression of human lysozyme through the increased gene dosage and co-expression of transcription factor *haclp* in *Pichia pastoris*. *Curr. Microbiol.*, **77**: 1-9. <https://doi.org/10.1007/s00284-019-01872-9>
- Lu, J., Song, Q., Ji, Z., Liu, X., Wang, T., and Kang, Q., 2015. Fermentation optimization of maltose-binding protein fused to neutrophil-activating protein from *Escherichia coli* t_{b1}. *Electron. J. Biotechnol.*, **18**: 281-285. <https://doi.org/10.1016/j.ejbt.2015.05.002>
- Makino, T., Skretas, G., and Georgiou, G., 2011. Strain engineering for improved expression of recombinant proteins in bacteria. *Microb. Cell Factories*, **10**: 32. <https://doi.org/10.1186/1475-2859-10-32>
- Mansur, M., Cabello, C., Hernández, L., País, J., Varas,

- L., Valdés, J., Terrero, Y., Hidalgo, A., Plana, L., and Besada, V., 2005. Multiple gene copy number enhances insulin precursor secretion in the yeast *pichia pastoris*. *Biotechnol. Lett.*, **27**: 339-345. <https://doi.org/10.1007/s10529-005-1007-7>
- Novy, R., Drott, D., Yaeger, K., and Mierendorf, R., 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations*, **12**: 1-3.
- Parret, A.H., Besir, H., and Meijers, R., 2016. Critical reflections on synthetic gene design for recombinant protein expression. *Curr. Opin. Struct. Biol.*, **38**: 155-162. <https://doi.org/10.1016/j.sbi.2016.07.004>
- Ranjbari, J., Babaeipour, V., Vahidi, H., Moghimi, H., Mofid, M.R., Namvaran, M.M., and Jafari, S., 2015. Enhanced production of insulin-like growth factor i protein in *Escherichia coli* by optimization of five key factors. *Iran. J. Pharm. Res.*, **14**: 907.
- Rosano, G.L., and Ceccarelli, E.A., 2014. *Recombinant protein expression in Escherichia coli: Advances and challenges. Recombinant protein expression in microbial systems.* pp. 7. <https://doi.org/10.3389/978-2-88919-294-6>
- Silva, F., Queiroz, J.A., and Domingues, F.C., 2012. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnol. Adv.*, **30**: 691-708. <https://doi.org/10.1016/j.biotechadv.2011.12.005>
- Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q., 2009. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.*, **18**: 936-948. <https://doi.org/10.1002/pro.102>
- Son, Y.J., Ryu, A.J., Li, L., Han, N.S., and Jeong, K.J., 2016. Development of a high-copy plasmid for enhanced production of recombinant proteins in *leuconostoc citreum*. *Microb. Cell Factor.*, **15**: 12. <https://doi.org/10.1186/s12934-015-0400-8>
- Tegel, H., Ottosson, J., and Hober, S., 2011. Enhancing the protein production levels in *Escherichia coli* with a strong promoter. *FEBS J.*, **278**: 729-739. <https://doi.org/10.1111/j.1742-4658.2010.07991.x>
- Tripathi, N.K., 2016. Production and purification of recombinant proteins from *Escherichia coli*. *Chem. BioEng. Rev.*, **3**: 116-133. <https://doi.org/10.1002/cben.201600002>
- Ukkonen, K., Veijola, J., Vasala, A., and Neubauer, P., 2013. Effect of culture medium, host strain and oxygen transfer on recombinant fab antibody fragment yield and leakage to medium in shaken *E. coli* cultures. *Microb. Cell Factor.*, **12**: 73. <https://doi.org/10.1186/1475-2859-12-73>
- Vimberg, V., Tats, A., Remm, M., and Tenson, T., 2007. Translation initiation region sequence preferences in *Escherichia coli*. *BMC Mol. Biol.*, **8**: 100. <https://doi.org/10.1186/1471-2199-8-100>
- Zhang, D., Wei, P., Fan, L., Lian, J., Huang, L., Cai, J., and Xu, Z., 2010. High-level soluble expression of hgf-1 fusion protein in recombinant *Escherichia coli*. *Process. Biochem.*, **45**: 1401-1405. <https://doi.org/10.1016/j.procbio.2010.05.014>