Pak. J. Biochem. Mol. Biol., 2021, 54(3-4):21-32

Review

Translational Regulation of Recombinant Protein Expression in *E. coli* - A Review

M. Waheed Akhtar

School of Biological sciences, University of the Punjab, Lahore-54590, Pakistan.

Abstract

The progress made in the advancement and application of recombinant DNA technology has been largely associated with the development of expression systems based on *Escherichia coli*. Advancements have been made in optimization of *E. coli* expression system both at transcription and translation levels. Using this system, a large variety of heterologous proteins have been produced for applications in medical, industrial, agricultural and environmental fields. Here the recent developments mainly regarding the factors involved in the translation process are dealt with. The various aspects related to Shine-Dalgarno (SD) sequence, sequence and size of the oligonucleotide between SD sequence and the start codon, the initial 5'-end sequence of the mRNA encoding the protein of interest, codon bias between the target and the host, size of the protein to be expressed and the use of tags to achieve different goals are discussed in this review.

Key words: Translation, recombinant protein, expression, *E. coli*

Introduction

Escherichia coli remains the commonly employed organism for expression of heterologous proteins. The choice for using *E. coli* as the expression system is due to the several advantages, which most other systems lack. The ability of *E. coli* to grow fast in a relatively cheaper medium to achieve high cell density cultures with healthy expression levels offer great advantages for obtaining the product in large scale economically. Development of a variety of molecular tools and protocols likea variety of plasmids, a large number of the bacterial strains optimized for expression, the versatile protein-specific molecular strategies and the variety of culture methodologies have been instrumental in achieving the desired objectives. However, for the proteins requiring post-translational processing, alternate options need to be considered.

Rate of expression of a protein is regulated at both the transcription and the translation levels. Several factors are involved at both the levels in determining the rate of protein expression. Many reviews have discussed in detail the factors regulating expression at transcription and translation levels in *E. coli* and the related aspects in detail (1-4).

The elements involved in protein expression at transcription level are rather of more general nature and we have seen outstanding advancements in this respect. Availability of engineered *E. coli* strains combined with matching expression vectors have made it possible to achieve high expression levels under optimal conditions. However, at the translation level the specific nature of the protein, which is determined by its amino acid sequence, becomes highly relevant in determining the expression levels.

The translation mechanism involves the formation of initiation complex involving many factors. The translation rate has been known to be associated with the efficiency with which the initiation complex is formed. Formation of the initiation complex requires

attachment of the ribosomal binding sequence of the mRNA, which is rich in purines, to the corresponding sequence at the 3'-end of the 16S ribosomal RNA, which is rich in pyrimidines. Many factors affect formation of the initiation complex and thus the progress of the translation process. These include the distance between the start codon and the Shine-Dalgarno sequence, possibility of the formation of secondary structures between the bases around the start codon, stability of the mRNA structure, preference for the codons in the 5'-end of the coding sequence and the role of rare codons in general. Each of these factors are discussed in the light of the work done in the recent years.

Base Sequence Upstream of the Start Codon

Poor expression in *E. coli* can be affected by the nucleotide sequence at the linkage region between coding sequence and the vector. The vector sequence between the start codon and the Shine–Dalgarno sequence can be of much significance for expression. PCR based sequence variants were generated and the library was transformed in *E. coli*. Screening of the cells was done using a marker fluorescent dye linked to the sequence of interest (5). More recently the screening was done through translation by linking the coding sequence to the gene encoding the antibiotic (6). Using this strategy, the authors reported the results for the sequence variants modulating expression by hundreds of folds. A low GC content in this region and increased stability of the mRNA seemed to be important for increased expression.

The effect of certain codons in the sequence immediately before the mRNA encoding β-galactosidase was reported to have sound effect on expression (7). Introducing the triplets UAU and CUU at -1 position produced 20-fold higher expression as compared to when UUC, UCA, or AGG were used in the same position. A U residue at -1 and an A residue at -2 position enhanced the expression level. However, it was suggested that the degree of expression enhancement depends on the nature of the the neighboring bases. Another study showed that placing the SD sequence between the two initiation codons the initiation took place mainly at the downstream initiation codon (8). However, when the distance between upstream codon and the SD sequence was increased initiation at the upstream codon became possible indicating the importance of the initial codons downstream from the start codon in the translation process. Expression of bovine growth hormone was enhanced from 3% to 25% by introducing AT rich block just upstream of the SD sequence. Likewise enriching spacer region between Shine-Dalgarno sequence and start codon with A and T resulted in expression level from 0.8% to 20% (9). Enhanced expression was explained due to destabilization of secondary structure in these regions.

Eliminating the possibility of hairpin loop structure formation involving the start codon of G+C rich 3-isopropylmalate dehydrogenase of *Thermus thermophilus* through mutation the expression level was enhanced manifold (10). Also, by inserting short open reading frames between the promoter sequence and the start codon of this enzyme and that of tryptophan synthase of the same organism the expression could be enhanced. As a result of formation of a hairpin loop caused burying the initiation codon of murine μ heavy chain immunoglobulin, which resulted in its low-level expression (11). Removal of the secondary structure by alteration of nucleotide sequence by mutagenesis, resulted in over ninety-fold increase in protein expression.

Binding threonyl-tRNA synthetase of *E. coli* to its own mRNA located at the site upstream of the initiation codon was attributed to negative autoregulation of the gene (12). Promoting enhanced base-pairing and thus reduced binding between ribosomal binding site and the corresponding site on the ribosome by introducing mutations in the in the operator region had strong effect in repression. The enhanced repressor effect of the enzyme caused in the case of mutated mRNA was explained on the basis of the competitive advantage due to reduced binding of the mutated mRNA to the ribosome.

5'-End Coding Sequence

Secondary structure of mRNA and its structure stability particularly in the sequence around the initiation codoncan play an important role in translation efficiency. The major factors regarding the 5’-end of the coding sequence that govern regulation of expression include the GC content, secondary structure formation, stability of the mRNA structure, bias in codon usage as well as efficiency in the translation initiation step. Formation of strong secondary structure like a hairpin loop at a location involving the Shine-Dalgarno sequence and the initiation codon can lower expression levels markedly (13).

A study on the interplay between the sequences in the upstream region and the sequence immediately downstream of the initiation codon reported that the most efficient immediate downstream sequence along with a Shine-Dalgarno sequence which lacked complementarity with anti Shine-Dalgarno sequence of 16S rRNA showed as much expression as that in the case of an appropriate Shine-Dalgarno sequence but with the altered downstream sequence (14).

Expression of two different sequences of bovine growth hormone produced by modifications in the early coding sequence was studied by varying nucleotide sequence in the SD and the spacer region of nine bases between SD sequence and the initiation codon. Out of a series of constructs produced those having ΔG value of less than -26.3 kcal/mole, calculated on the basis of the sequence between transcription initiation site to +51 nucleotide of bGH cDNA using mfold web server software, showed expression levels in the range of 6.9% to 8.5%, as compared to nothing in the case of those having higher ΔG values (15).

Lowering stability of secondary structure formation by introducing silent mutations in the 5’-end of mRNA can have positive effect on kinetics of translation at both the level of formation of the initiation complex (16) as well as the chain elongation (17). Lowering the possibility of mRNA hairpin structure stabilityat its 5'-end by enhancing the relative AU codon content can lead to significantly greater expression levels (18). Removing the possibility of strong secondary structure formation in the the 5’-end of the coding sequences would promote attachment of ribosomes on mRNA and thus construction of translation initiation complex required for efficient translation and peptide elongation (19).

Position of a rare codon in the 5’-end region of the gene can dictate the expression level. Substitution of the codon at position +2 of the gene encoding strep to kinase with the rare codon AGG, the enzyme was over-expressed in *E. coli*, but shifting this codon to +3 or +5 positions resulted in nil expression (20). However, moving this codon downstream to +9 or +11 positions resulted in over-expression again. It was thus concluded that besides the choice of rare codons their precise position is important in determining expression level of the heterologous gene.

Expression of native cDNA encoding bovine somatotropin using a vector derived from pBR322 was not significant. However, inserting a stretch of 16 codons from *trp*LE in front of somatotropin gene cDNA resulted in the yield of more than 20% of total cell protein as fusion protein (21). Changing some nucleotides to destabilize stem loop structures in the early coding sequence increased expression but only up to 0.5-1.0 % of the total cell proteins. It was suggested that the nucleotide sequence or adjoining amino acids within the beginning of somatotropin sequence,along with the secondary structure, had a role in translation initiation.

In a study on the influence of RNA sequence on expression of β-galactosidase introducing silent mutations in the early downstream codons resulted in highly variable β-galactosidase expression levels (22). Certain silent mutations in the +2 to +6 codons showed ~9-fold increase in expression as compared to that in the case of the native codons. Introducing NGG sequence (where N is non-G) at +2 to +5 positions resulted in low expression of *lacZ* reporter gene, although growth was not affected (23). This was considered to be due to presence of NGG codons in early coding position trigger peptidyl-tRNA drop-of, which was explained on the basis of the evidence that growth of a mutant *E. coli* strain with a defective peptidyl-tRNA hydrolase could not grow as the peptidyl-tRNA after drop-of could not be recycled.

Modifying the sequences in the initial codons of the gene encoding bubaline somatotropin as silent mutations (24) or introducing codons encoding a hexa-histidine sequence (25) resulted in higher expression levels of the product. This seemed to be due to lowering the possibility of formation of secondary structure of mRNA. In a more extensive study introduction of non-silent and silent mutations in the early codons in +2 to +8 positions of bubaline somatotropin mRNA raised the expression levels in *E. coli* to varying extents, as compared to barely discernible level in the case of the native form (26). Deletion of the GCC codon at +2 position of the native sequence and addition at +2 position of repeat codons for hexa-histidine, commonly used for immuno-purification of the product, enhanced the expression up to 48% of the total cell proteins. Differences in expression related to the nature of early codons (especially +2) seemed to confirm therole of thesecondary structurer of mRNA in lowering expression level.

The impact of codon modulation in the 5′-end of the mRNA encoding human granulocyte colony stimulating factor, on its expression in *E. coli* was studied by introducing single or more synonymous substitutions at positions between +2 to +7 codons (27). Changing the codon for threonine from ACC toACA at +2 position, either alone or in combination with Pro codons CCC, CCT or CCA at +3 position resulted in expression enhancement to 45–50% of the total cell proteins, as compared to otherwise poorly expressed native GCSF sequence. The differences in the levels of GCSF expression could be attributed to the codon usage, stability of mRNA structure, and formation of mRNA secondary structure in the early region.

In another study introduction of silent mutations in the region between ribosomal binding site and +24 bases of mRNA coding for caprine growth hormone had strong influence on the expression levels (28). From almost negligible expression level in the case of the native mRNA, silent mutations in itsearly 5'-end region seem to have a positive impact on translation without altering the free-energy values considerably.

A β-glucosidase of *Bacillus halodurans* was expressed in a partially soluble form at levels >35 % of *E. coli* cell proteins (29). However, the variant in which serine codon TCG at +3 position was replaced with TCA resulted in total insolubility of the expressed enzyme. Examining the expression levels at different time intervals after induction IPTG showed bglA release by the cells immediately after induction was more rapid when transformed with the mutated variant. This faster expression rate is not likely to leave enough time for the enzyme to fold and assume native conformation. Lowering the cultivation temperature to 18 °C decreased the rate of synthesis resulting in the expression in a soluble.

Codon Usage Bias

Proteins are often difficult to express in a heterologous system due to difference in the levels of synonymous codons between the foreign coding DNA and that of the host *E. coli*. The genes encoding a protein from a particular organism may use specific codon or expression-limiting regulatory elements in the coding sequence for their specific needs. The rare codons scenario of *E. coli* with the limited availability of the corresponding tRNAs may cause limitations in expression of the genes carrying such codons frequently. Such a dissimilar situation may lead to limitations including misincorporation of amino acid and early termination of the polypeptide chain elongation. In order to overcome obstacles in heterologous expression in *E. coli* several strategies including sequence modifications in the initial translation initiation regions, alterations related to secondary structure of mRNA and use of codon biases have been reviewed (3). Several online apps are available for checking the existence and positions of rare codons in gene sequence of interest with respect to the host organism *E. coli.* The codons, which are used by *E. coli* at a frequency of less than one percent, as for example AGG, have been defined as the rare codons (30).

Two possible approaches can be adopted to overcome the issues related to codon usage bias. These include increasing the availability in the host cells of the tRNAs specific for carrying amino acid residues corresponding to the rare codons or optimization of the foreign coding sequence making it more acceptable to the host strain (31). Engineered strains of *E. coli* equipped with extra copies of the genes encoding the limiting tRNAs have been produced and these are available commercially.

The other strategy aims at introducing silent mutations, most often by changing wobble base of the codon, to modify the rare codons to match the codon usage of the host (32, 33). The approach to modify the codons in such a way should ordinarily not alter the amino acid sequence of the gene (34), unless required to introduce mutation at a particular site of the target protein for achieving a specific objective. Introduction of silent mutations can be achieved using different approaches like site-directed mutagenesis, or having the whole gene or a part of it resynthesized. If the codon optimization is required in the 5'- or the 3'-ends of the target gene the it can be achieved by designing the primers using the preferred codons and using these for PCR-amplification of the gene. In case the codon alterations are required in the sites located more centrally in the gene then the whole gene can be split two or more fragments. Suitable primers with the required codon alterations and having complimentary restriction sites in the successive fragments can be designed for amplification of each of these fragments. Amplified fragments using such primers can be restricted with the corresponding enzymes and ligated to generate the complete gene sequence with the desired codon alterations. The decision for introducing the codon alteration can be

based on the algorithms available for this purpose (35-37).

If the mutations are required to be introduced in the terminal 5'- or the 3'-ends of the coding sequence then it can be achieved relatively more easily by designing the primers with the sequence that shall allow introduction of the required mutations. Using this approach, we introduced a series of mutations in the 5'-end sequence of the genes encoding bubaline somatotropin (24, 26), human granulocyte colony stimulating factor (27) and a β-glucosidase to study their effects on their expression and solubility property (29). However, the procedure for introduction of mutations within the regions distant from 5'- or 3'-end of the sequence becomes more cumbersome. The approaches involving silent mutations based onsite-directed mutagenesis or synthesis of the whole gene or parts of it are not trivial options (3).

Designing strategies based on codon optimization may be tricky quite often. Efficiency in the formation of initiation complex is required as the first step in the translation process. In addition to other factors the sequence of the 5'-end of the mRNA may on its own or in combination with the pre-start codon sequence may give rise to an unstable secondary structure, which may seriously hinder efficiency in the formation of initiation complex. Rarity of the codon in the N-terminus sequence may be more favorable for expression if it results in reduction of RNA secondary structure (38). Another factor determining protein expression can be determined by the speed of binding and decoding of the codons in the initiation site (39). Clearance of the ribosome at the initiation site in a rapid manner shall allow their ready availability for binding to a free start codon thus engaging in translation. Another factor involved in the expression regulation is the presence of Shine-Dalgarno-like sequences in the target mRNA that may hybridize with 16S rRNA and may cause what is called as translational pausing (40). Slowing down translational process due to translational pausing may have a beneficial effect on folding of the expressed protein as it promotes folding at the intermediate stage (41, 42).

For a rational design of the genes for achieving the desired goals of expression the various factors controlling translation process pose a challenge. The factors like 5'-end RNA sequence allowing generation of secondary structure, location of Shine–Dalgarno-like motifs in a strategic position, clearance rates of ribosomes at the initiation site are relevant. Likewise, ribosome stalling caused by depletion of low abundance tRNA, maybe followed by its detachment from the RNA strand before completion of the full-length protein (43). To overcome this issue *E. coli* strains having plasmids with extra copies of the deficient tRNAs genes have been developed and these are available commercially. The BL21(DE3) CodonPlus strain (Stratagene) contains the pRIL plasmid, which provides the commonly used p15A replicon, and additional genes for the tRNAs for Arg, Ile, Leu and Pro. The Rosetta (DE3) strains from Novagen supply tRNAs for Gly in addition to all the above-mentioned codons. However, improvement in the speed and amount of protein production can lead to a decrease in solubility of the expressed protein (44).

Size of the Expressed Protein

Expression of a protein, with all the other factors like nature of 5'-end codons, mRNA secondary structure, codon bias, etc. common, the number of the completed molecules per unit time shall depend on the size of the molecule. Expression of a larger size protein shall require the ribosomes longer time to traverse a longer size mRNA

to complete the process. Other factors like the possible hiccups in covering a longer mRNA sequence and supply of tRNAs may be relevant in successful completion of the process. Expression of relatively smaller size proteins have been commonly reported to be expressed in larger molar quantities under similar conditions.

Regarding the effect of the gene length on expression it was found expression of shorter sequences isgreater in comparison to longer sequences (45). A computational analysis for the relationship between the average gene length and the expression rate showed that the expression was found almost inversely proportional to gene length. This study agrees with the previous reports of the researchers that the genes requiring lesser inputs shall be preferred by the metabolic systems (46).

In our study on the effect of reducing the size of the target molecule on expression we found that reducing the size results in significant increase in expression. PstS1, a 38-kDa antigen of *M. tuberculosis* was truncated by removing 96 residues from the N- and 14 from the C-terminal to produce tnPstS1 (29.5 kDa). Expression level of the truncated version increased to 25% of the total cell proteins as compared to 15% in the case the native molecule (47). In addition to the higher expression the truncated variant showed significantly higher sensitivity in detecting the antibody.

Xylanases XynC and XynZ from *Clostridium thermocellum* were expressed with and without the non-catalytic domains in *E. coli* (48). From the molecular sizes of XynC and XynZ with the respective carbohydrate binding modules of 58 and 60 kDa, respectively, their sizes were reduced to 38 kDa each. Whereas both these enzymes with the carbohydrate binding domains were expressed at about 30% of the total cell proteins, expression level for the enzymes without the CBM increased to 45% in both the cases. Moreover, specific activity of XynZ without CBD increased by ~5-fold. Enhanced expression levels of the catalytic domains only seemed to be due to their smaller sizes as well as a more favorable codon arrangement particularly in the 5'-end of the encoding sequence (49).

Cel5A, a 37 kDa cellulase of *Thermotoga maritima,* was expressed in *E. coli* in a partially soluble state. Removing ten residues from its C-terminal to produce tCel5A1, the expression was not only in a completely soluble state, but the truncated version also showed several-fold increase in activity (50). *In silico* analysis of tCel5A1 revealed that the catalytic cavity with the active site residues Glu253, Trp286, and Phe292 lying close to the C-terminus, were exposed with removal of over-hanging C-terminal peptide, making the active site residues more accessible to the substrate thus resulting in increased activity.

Tags for Expression andAffinity Purification

Although overexpression of the required protein is the primary objective, but several other aspects are important for obtaining the product in a final state for application. The expressed protein needs to be in a soluble state, easily detectable and purified to homogeneity following an efficient protocol. These issues have been addressed commonly by using tags and fusion partners, which facilitate efficient processes like affinity purification and other targets for downstream processing of the expressed proteins.

Several tags have been reported for use as solubility enhancers, but there is hardly any of these, which can be used for a wide variety of proteins. Some of the commonly used tags used for expression of proteins in a soluble form include, maltose binding protein, glutathione S-transferase, and small

ubiquitin-like family of proteins (1). However, as these tags may work for some proteins but not for the others, choosing the one for a particular protein requires additional efforts to make the right choice. In the recent years, many reports describing affinity tags and new fusion partners have been reviewed (51, 52). A tag encoding the sequence for the tetrapeptide serine-lysine-isoleucine-lysine when added after the initial methionine codon improved expression of recombinant proteins markedly (53). A tag derived from the first 11 amino acid residues of a carbonic anhydrase when used in combination with His-tag enhanced production of the tested proteins in a soluble state (54). Proteins bearing a 34-amino acid heparin-binding peptide tag was used to purify proteins using heparin attached agarose resin (55, 56). A simple NaCl gradient could be used to elute the tagged proteins. A tag constructed from a truncated maltotriose-binding protein of *Pyrococcus furiosus* in association with a modified His-tag, consisting of intercalating histidine and glutamate residues, showed high expression in a soluble form (57). A rather small metal-binding protein (<10 kDa), which facilitate soluble expression as well as affinity purification with themetal ion-bound resins, was used as fusion tags (58, 59).

We showed that HSPX, a heat shock protein of *Mycobacterium tuberculosis*, when attached to the n-terminal of other antigens of this organism allowed soluble expression of the protein, which was otherwise expressed as inclusion bodies (60-62). Soluble expression of the fusion molecules consisting of more than antigens in combination with HSPX was an additional advantage in designing the molecules with enhanced sensitivity in serodiagnosis of tuberculosis. Based on the pET32 plasmid a vector was designed for expressing histidine-tagged proteins with enhanced solubility (63, 64). This expression system was based on a restriction-free cloning protocol using only a pair of primers for all vectors.

Although strategies for soluble expression of heterologous protein have been extensively investigated, there has been increasing tilt in favor of expression as inclusion bodies (65, 66). Formation of inclusion bodies may in fact help in achieving high yields and also help in overcoming toxic effect of the foreign expressed protein. The support in favor of inclusion body formation also develops due to availability of refolding procedures, which can be adopted for proteins commonly with success. In some cases, expressing proteins as inclusion bodies may be the only choice. A signal sequence consisting of 39 amino acid residues of the protein TorA from *E. coli*, which promoted the formation of inclusion bodies of highly soluble protein like thioredoxin, may become tool of choice in particular cases (67).

The vast improvements in the protocols for the production of recombinant proteins using *E. coli* expression system have yielded enormous gains in the fields of health, industry, agriculture and environment. Further advancements shall result in emergence of strategies which should make the production of more difficult products easier.

References

1. Rosano GL and Ceccarelli EA. (2014) Recombinant protein expression in *E. coli:* advances and challenges. *Frontiers in Microbiology* 5, 172.

2. Angov E (2011) Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol. J.* 6, 650-659.

3. Gustafsson C, Govindarajan S and Minshull J (2004) Codon bias and hetereologous protein expression. *Trends in Biotechnology* 22, 346-353.

4. Rosano GL, Morales ES and Ceccarelli EA (2019) New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Science* 2, 1412–1422.

5. Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Norholm MH and Daley DO (2015) Enhanced protein production in Escherichia coli by optimization of cloning scars at the vector-coding sequence junction. *ACS Synth Biol* 4:959–965.

6. Rennig M, Martinez V, Mirzadeh K, Dunas F, Rojsater B, Daley DO, Norholm MHH (2018) TARSyn: Tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. *ACS Synth Biol* 7:432–442.

7. Hui A, Hayflick J, Dinkelspiel K and de Boer HA (1984) Mutagenesis of the three bases preceding the start codon of the β-galactosidase mRNA and its effect on translation in *Escherichia coli. EMBO J.*3(3), 623-629.

8. Jin H, Zhao Q, Ernesto I., de Valdivia G, Ardell DH, Stenström M and Isaksson LA (2006) *Mol. Microbiol.* 60(2), 480-492.

9. Choi JW, Park H, Yun JW,Song CH and Lee YS (1999) The artificial AT-rich block enhanced the production of bovine growth hormone in *Escherichia coli*. *Bioprocess Engineering* 21, 293-298.

10. Ishida M and Oshima T (1994) Overexpression of genes of an extreme thermophile, *Thermus thermophilus*, in *Escherichia coli* cells. *J. Bact.* 176, 2767-2770.

11. Wood CR, Boss MA, Patel TP and Emtage JS (1984) The influence of messenger RNA secondary structure on expression of an immunoglobulin heavy chain in *Escherichia coli*. *Nucleic Acid Res.* 12(9), 3937-3950.

12. Brunel C, Romby P, Sacerdot C, de Smit M, Graffe M, Dondon J, van Duin J, Ehresmann B, Ehresmann C and Springer M (1995) Stabilised secondary structure at a ribosomal binding site enhances translational repression in *E. coli. J. Mol. Biol*. 253, 277-290.

13. Kubo, M and Imanaka, T (1989) mRNA secondary structure in an open reading frame reduces translation efficiency in *Bacillus subtilis*. *J. Bacteriol. 171*, 4080–4082.

14. Stenström CM, Holmgren ME, and Isaksson LA (2001) Cooperative effects by the initiation codon and its flanking regions on translation initiation. *Gene* 273, 259-265.

15. Paik SY, Ra KS, Cho HS, Koo KB, Baik HS, Lee MC, Yun JW and Choi JW (2006) The influence of the nucleotide sequences of Random Shine-Dalgarno and spacer region on bovine growth hormone gene expression. *J. Microbiol*. 44, 64-71.

16. de Smit MH and van Duin J, Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proc. Natl. Acad. Sci. USA* 1990, *87*, 7668–7672.

17. Tuller T, Carmi A, Vestsigian K, Navon S, et al., An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 2010, *141*, 344–354.

18. Voges D, Watzele M, Nemetz C, Wizemann, S and Buch- berger B, Analyzing and enhancing mRNA translational efficiency in an *Escherichia coli* in vitro expression system. *Biochem. Biophys. Res. Commun.* 2004, *318*, 601–614.

19. Angov E, Legler PM and Mease RM (2011). Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol. Biol.* 705, 1–13. doi: 10.1007/978-1-61737-967-3_1

20. Thangadurai C, Suthakaran P, Barfal P, Anandraj B, Pardhan SN, Boneya HK, Ramalingan S and Murugan V (2008) Rare codon priority and its position specificity at the 5′ of the gene modulates heterologous expression in *E. coli*. *Biochemical and Biophysical Research communications* 376, 647-652.

21. Tomich CSC, Olson ER, Olson MK, Kaytes PS, Rockenbach SK and Hatzebuhler NT (1989) Effect of nucleotide sequences directly downstream from the AUG on the expression of bovine somatotropin in *E. coli*. *Nucleic Acids Res*. 17, 3179-3197.

22. Stenström CM and Isaksson LA (2002) Influences on translation initiation and early elongation by the messenger RNA region flanking the initiation codon at the 3' side. *Gene* 288, 1-8.

23. de Valdivia EIG and Isaksson LA (2005) Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding regions of mRNA. *FEBS Lett.* 272, 5306-5316.

24. Sadaf S, Khan MA, Wilson DB and Akhtar MW (2007) Molecular Cloning, Characterization, and Expression Studies of Water Buffalo (*Bubalus bubalis*) Somatotropin. *Biochemistry (Moscow)* 72, 162-169.

25. Sadaf S, Khan MA and Akhtar MW (2007) Production of bubaline somatotropin by auto-induction in *Escherichia coli*. *Biotechnol. Appl. Biochem*. 47, 21–26.

26. Sadaf S, Khan MA, and Akhtar MW (2008) Expression enhancement of bubaline somatotropin in *E. coli* through gene modifications in the 5'-end coding region. *J. Biotech*. 135, 134-139.

27. Bashir S, Iqbal M, Sadaf S, Akhtar MW (2017) Synonymous codon changes at the 5′-end of the gene strongly impact the heterologous protein expression in *Escherichia coli. Applied Biochemistry and Microbiology*. 53 (3), 296–303.

28. Khan MA, Sadaf S and Akhtar MW (2007) Role of silent gene mutations in the expression of caprine growth hormone in *Escherichia coli*. *Biotechnology Progress*. 23, 1049-1052.

29. Naz S, Ikram N, Sadaf Sand Akhtar MW (2010) 5′-end coding sequence of β-glucosidase gene of *Bacillus halodurans* determines the rate of expression and folding of the product in *E. coli. Pak. J. Biochem. Mol. Biol. 2010; 43(3):153-158.*

30. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 6, 494-500. (95)80082-4.

31. Sorensen H Pand Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115, 113-128.

32. Welch M, Govindarajan S, Ness J E, Villalobos A, Gurney A, Minshull J, et al. (2009). Design parameters to control synthetic gene expression in *Esherichia coli*. *PLoS ONE* 14;4(9):7002

33. Menzella HG (2011). Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli*. *Microb. Cell Fact.* 10 (15):1475-2859.

34. Burgess-Brown NA, 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.

35. Gao W, Rzewski A, Sun H, Robbins PD and Gambotto A (2004). UpGene: application of a web-based DNA codon optimization algorithm. *Biotechnol. Prog.* 20, 443–448.

36. Jayaraj S, Reid R, and Santi DV (2005). GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res.* 33, 3011-3016.

37. Puigbo P, Guzman E, Romeu A, and Garcia-Vallve S (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* 35, W126–W131.

38. Goodman D B, Church GM and Kosuri S (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479.

39. Chu D, Kazana E, Bellanger N, Singh T, Tuite MF and Von Der Haar T (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 33, 21–34.

40. Li GW, Oh E and Weissman JS (2012) The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature.* 484, 538–541.

41. Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, and Nussinov R (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.* 383, 281-291.

42. Yona AH, Bloom-Ackermann Z, Frumkin I, Hanson-Smith V, Charpak-Amikam Y, Feng Q, etal. (2013). tRNA genes rapidly change in evolution to meet novel translational demands. *Genome Research Vol.* 24 (11): 1797-1807.

43. Buchan JR and Stansfield I (2007). Halting a cellular production line: responses to ribosomal pausing during translation. *Biol. Cell* 99, 475-487.

44. Rosano GL and Ceccarelli EA (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb. Cell Fact.* 8 (41):1475-2859.

45. Gajendra PS Raghava and Joon H Han (2005) Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein. *BMC Bioinformatics.* 6(59): 1471-2105.

46. Akashi H (2003). Translational selection and yeast proteome evolution. *Genetics.*164 (4):1291-303.

47. Khurshid S, Khalid R, Afzal M and Akhtar MW (2013) Truncation of PstS1 antigen of *Mycobacterium tuberculosis* improves diagnostic efficiency. *Tuberculosis* 93, 654-659.

48. Sajjad M, Khan MIM, Akbar NS, Ahmad S, Ali I and Akhtar MW (2010) Enhanced expression and activity yields of *Clostridium thermocellum* xylanases without non-catalytic domains. *J. Biotechnol.* 145, 38-42.

49. Khan MIM, Sajjad M, Sadaf S, Zafar R, Niazi UHK and Akhtar MW (2013) Influence of transposition and insertion of additional binding domain on expression and characteristics of XylanaseZ of *Clostridium thermocellum*. *J. Biotechnology*. 168:403-408.

50. Basit A and Akhtar MW (2018) Truncation of the processive Cel5A of *Thermotoga maritima* results in soluble expression and several fold increase in activity. *J. Biotechnology & Bioengineering* 115:1675-1684.

51. Paraskevopoulou V and Falcone FH (2018) Polyionic tags as enhancers of protein solubility in recombinant protein expression. *Microorganisms* 6:47.

52. Vandemoortele G, Eyckerman S, Gevaert K (2019) Pick a tag and explore the functions of your pet protein. *Trends Biotechnol.* 37(10):1078-1090.

53. Ojima-Kato T, Nagai S and Nakano H (2017) N-terminal SKIK peptide tag markedly improves expression of difficult-to-express proteins in Escherichia coli and Saccharomyces cerevisiae. *J. Biosci Bioeng* 123:540-546.

54. Nguyen TKM, Ki MR, Son RG and Pack SP (2019) The NT11, a novel fusion tag for enhancing protein expres- sion in Escherichia coli. *Appl Microbiol Biotechnol*. 103: 2205-2216.

55. Jayanthi S, Gundampati RK and Kumar TKS (2017) Simple and efficient purification of recombinant proteins using the heparin-binding affinity tag. *Curr Protoc Prot Sci.* 90(6):11–16.

56. Morris J, Jayanthi S, Langston R, Daily A, Kight A, McNabb DS, Henry R and Kumar TKS (2016) Heparin- binding peptide as a novel affinity tag for purification of recombinant proteins. *Prot Express Purif* 126:93–103.

57. HanY,GuoW,SuB,GuoY,WangJ,ChuB and YangG (2018) High-level expression of soluble recombinant pro- teins in Escherichia coli using an HE-maltotriose-binding protein fusion tag. *Prot Express Purif*. 142:25-31.

58. Vargas-Cortez T, Morones-Ramirez JR, Balderas-Renteria I and Zarate X (2016) Expression and purification of recombinant proteins in Escherichia coli tagged with a small metal-binding protein from Nitrosomonas europaea. *Prot Express Purif.* 118:49–54.

59 Vargas-Cortez T, Morones-Ramirez JR, Balderas- Renteria I, Zarate X (2017) Production of recombinant proteins in Escherichia coli tagged with the fusion pro- tein CusF3H. *Prot Express Purif.* 132:44–49.

60. Khalid R, Afzal M, Khurshid S, Paracha RZ, Khan IH and Akhtar MW(2016) Fusion Molecules of Heat Shock Protein HSPX with Other Antigens of Mycobacterium tuberculosis Show High

Potential in Serodiagnosis of Tuberculosis. *PLOS ONE*. 11 (9): 163-349.

61.Akhter M, Arif S, Khaliq A, Zaib un Nisa, Khan IH and Akhtar MW (2020) Designing fusion molecules from antigens of *Mycobacterium tuberculosis* for detection of multiple antibodies in plasma of TB patients. *Tuberculosis*.124, 101981.

62. Shaista, Arif, Akhter, Mohsina, Khaliq, Aasia, Nisa, Zaib Un, Khan, Imran H.and Akhtar, Muhammad Waheed (2021) Serodiagnostic evaluation of fusion proteins from multiple antigens of Mycobacterium tuberculosis for active TB. *Tuberculosis.* 127, 1020153.

63. Correa A, Ortega C, Obal G, Alzari P, Vincentelli R and Oppezzo P (2014) Generation of a vector suite for protein solubility screening. *Front Microbiol*. 5:67.

64. Ortega C, Prieto D, Abreu C, Oppezzo P and Correa A (2018) Multi-compartment and multi-host vector suite for recombinant protein expression and purification. *Front Microbiol.* 9:1384.

65. Wurm DJ, Quehenberger J, Mildner J, Eggenreich B, Slouka C, Schwaighofer A, Wieland K, Lendl B, Rajamanickam V, Herwig C and Spadiut O (2018) Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in E. coli. *Appl Microbiol Biotechnol* 102:667–676.

66. Ramon A, Senorale-Pose M and Marin M (2014) Inclusion bodies: not that bad. Front Microbiol 5:56.

67. Jong WSP, Vikström D, Houben D, van den Berg van Saparoea HB, de Gier J-W and Luirink J (2017) Application of an *E. coli* signal sequence as a versatile inclusion body tag. *Microb Cell Fact* 16:50.