

An Overview of the Parameters for Recombinant Protein Expression in *Escherichia coli*

Bilgimol C Joseph¹, Suthakaran Pichaimuthu^{1,2}, Sankaranarayanan Srimeenakshi³, Musti Murthy¹, Kalimuthu Selvakumar², Ganesan M^{1,2} and Sadananda Rao Manjunath^{1,2*}

¹Micro Therapeutic Research Labs Pvt Ltd., Chennai, Tamilnadu, India

²KMS Health Centre Pvt Ltd., Chennai, Tamilnadu, India

³National Institute of Technology Rourkela, Rourkela, Odisha, India

Abstract

Escherichia coli, is one of the most widely preferred organism for the production of recombinant protein. Most of the FDA approved therapeutic proteins are produced in *E. coli*. The well-established cell factory of *E. coli* makes it a perfect heterologous system of choice for the production of recombinant proteins. In recent years, several advances have been taken place for modifying these cell factories for easy production of therapeutic proteins with utmost precision. Several molecular tools and protocols are available in hand for the high level protein production in heterologous expression system. Adapting the best strategy for producing recombinant proteins in *E. coli* can be obtained using several approaches. Combination of strategies will work well to enhance the expression and stability of produced protein. In this review we try to collate different strategies and approaches that can enhance the production as well as the stability of proteins expressed in *E. coli*.

Keywords: Recombinant protein expression; *Escherichia Coli*; Promoters

Introduction

Selection of suitable heterologous system for the production of recombinant protein plays a major role in producing therapeutic proteins. Proteins which are easy to be expressed are produced in *E. coli*. For the complex protein production, mammalian cell lines are used. The smallest to largest proteins can be produced using various heterologous systems. The most difficult protein to be expressed in heterologous system was recombinant factor VIII. But still, this protein has been successfully produced in heterologous system from 1984.

Bacterial expression system always remains as the preferred choice for the production of many recombinant proteins. Various researches conducted on foreign gene expression in *E. coli* promises to broaden the usefulness of it as a tool for gene expression. There are several aspects and factors which are discussed while using *E. coli* expression system. Tightly regulated prokaryotic promoters are highly preferred for obtaining high-level gene expression. Usage of appropriate terminators will also favor the production yield and stability. Modified host strains of *E. coli* are now available which favors the formation of disulfide bonds in the reducing environment of the cytoplasm. This will also enhance the protein yield and reduce the proteolytic degradation. Insights into the process of protein translocation across the bacterial membranes will eventually pave way for producing the proteins in desired bacterial compartments. Recombinant protein co-expression along with molecular chaperons is also explored and it is shown that in certain cases, chaperon can be very effect in obtaining improved protein folding, solubility and membrane transport. Currently, codon optimized services provided by different industry helps in improving the protein production in desired heterologous system. Finally, successful combination by standardization can also give insight into protein production, stability and yield.

Selection of Organisms

Choice of right heterologous expression system remains as the corner stone of production of recombinant proteins. The machinery

of each heterologous system has to be carefully understood for the production of any recombinant proteins. Various heterologous expression systems available for the production of recombinant proteins are *E. coli*, yeast, insect cell lines, mammalian cell lines and cell free system. All these systems have its own strengths and weakness but the choice of the system depends on the protein of interest [1,2]. If the protein requires post translational modifications, prokaryotic expression system is never the right choice. The eukaryotic expression with controlled post translational modification will be the right choice in this case [3].

The major advantage of choosing bacterial expression system is because of its cost effectiveness. However, *E. coli* stands in an exceptional position for the production of recombinant proteins. Usage of *E. coli* for the production of recombinant proteins stems from the experience of decades of research on its genetics, easy manipulation and easily available genetic engineering tools to modify the organism. Apart from these, the rapid growth rate of *E. coli*, capability for continuous fermentation, decreased media costs and high expression levels makes this organism one of the wonderful host to express the recombinant proteins [4,5].

Selection of Appropriate Promoters

The most sought approaches to exploit the recombinant protein production involve in selecting or designing the appropriate promoter. The promoter for use in *E. coli* should have certain characteristics

*Corresponding author: Sadananda Rao Manjunath, Micro Therapeutic Research Labs Pvt Ltd., Chennai, Tamilnadu, India and KMS Health Centre Pvt Ltd., Chennai, Tamilnadu, India, E-mail: smanjmicro@gmail.com

Received: June 13, 2015; Accepted: August 26, 2015; Published: August 29, 2015

Citation: Joseph BC, Pichaimuthu S, Srimeenakshi S, Murthy M, Selvakumar K, et al. (2015) An Overview of the Parameters for Recombinant Protein Expression in *Escherichia Coli*. J Cell Sci Ther 6: 221. doi:10.4172/2157-7013.1000221

Copyright: © 2015 Joseph BC, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

which are suitable for high level protein synthesis [6,7]. First of all the promoter selected must be strong enough in accumulating the protein up to 10-30% or more of the total cellular protein. Secondly, the promoter should exhibit minimal level of basal transcriptional activity. The low transcriptional activity reduces pre-induction strain on the host from the metabolic burden of recombinant protein production which in turn results in expression of host-toxic proteins. Thirdly, stringent regulation of promoter is essential for the synthesis of proteins which are detrimental to the host cell. Large scale gene expression preferably employs high cell density with minimal promoter activity followed by induction of the promoter. Some of the examples of *E. coli* promoters are; *lac*, *trp*, *lpp*, *phoA*, *recA*, *tetA*, *cspA*, T7, T7 *lac* operator, T3-*lac* operator, T5-*lac* operator etc. However, the toxicity to the host is not restricted to foreign genes alone. It also depends on certain native genes like *traT* gene, which encodes an outer membrane lipoprotein, the *EcoRI* restriction endonuclease in the absence of consistent protecting *EcoRI* modification methylase and the *lon* gene [8,9]. Additionally, promoters are engineered for improved recombinant expression. There has been an accepted relevance to the construction of a library of synthetic stationary-phase and stress promoters which are generated by randomized engineering [10]. The newly synthesized promoters exhibited three to four fold greater activities than the natural promoters. Recently, a mutant promoter library was constructed with randomization of *E. coli* consensus promoter sequences and the resultant promoter had 27.5 fold higher activities than *lac* promoter [11].

Current advances in research has paved way for engineering systems that utilize dual promoters which will help in expressing two desired recombinant proteins simultaneously. Certain promoters like *Ara* and *lac*-based promoters can be used simultaneously. It is known that IPTG is an inhibitor of *AraC* promoter. Thus *AraC* promoter was mutated and this promoter was used along with *lac* based promoter [12]. The combination of promoters might enhance the expression of proteins efficiently.

Selection of Suitable Terminators

Usage of stop codons plays a vital role in regulation of protein expression. Almost all the organisms use TAA, TAG and TGA as stop codons. Stop codon much preferred by *E. coli* is TAA than the other TAG and TGA. Conventional usage of multiple stop codons might increase the efficiency of transcription termination. Efficient transcription termination minimizes the cellular energy drain and reduces the metabolic burden for the host. Another important aspect is that, transcription terminator forms secondary structure at 3' end of the mRNA, which in turn will improve the stability of mRNA and substantially increase the protein production [13-15].

Nevertheless the major drawback of using *E. coli* is its lack of machinery for secreting the proteins to the growth medium and inability to enable disulfide bond formation. It also lacks machinery for other post translational modifications. Also the *E. coli* lacks specific molecular chaperons which help in proper protein folding. *E. coli* expression system is also in dearth of having an inefficient cleavage system for cleaving the amino terminal methionine which will result in lowering protein stability and increasing immunogenicity [4,5].

In prokaryotic expression system, transcription termination is affected by two different mechanisms mainly, Rho-dependent transcription termination and Rho-independent transcription termination. In rho-dependent transcription termination, the termination depends on hexameric protein rho. Rho will help in release of nascent RNA transcript from the template. In case of Rho-

independent transcription termination, the signals encoded in the template are responsible for transcription termination [16-19]. Efficient transcription terminators are crucial while selecting elements of expression vectors. These transcription terminators thus have several important roles. Promoter occlusion is a process whereby the promoter inhibits its function [20]. The promoter occlusion can be prevented by placing the promoter in appropriate position. Insertion of transcription terminator downstream of the coding sequence will prevent continuous transcription through another promoter. Likewise, a transcription terminator placed upstream of the promoter which drives the expression of specific gene will minimize background transcription [21]. In *E. coli*, there are two tandem transcription terminators T1 and T2 which are derived from *rrnB* rRNA operon of *E. coli* [22]. There are many other sequences which are also quite effective.

The sequence analysis for several expressed genes in *E. coli* reveals that TAA is certainly the major stop codon used [23]. Having TAA as the stop codon in *E. coli* expression system has several advantages over TAG or TGA stop codon usage [24-26]. TAA can be read by both release factors and the efficiency is comparable to the release factor specific codons [27]. Therefore, TAA as stop codon will not only secure termination by either of the two release factors but also ensures the termination takes place with high speed and accuracy. However, judicious positioning of strong transcription terminators will result in expressing the toxic genes in *E. coli*. The strategy could be adapted to several other related proteins which are difficult to be expressed [28].

The literature prove that mycoplasma lack the gene for release factor 2 *prfB* [29]. Thus these bacteria lack TGA codon and interestingly, it was found that TGA codes for tryptophan in these bacteria instead of stop codon [30]. It is noticed that in all occasion, the TGA codon is found to be followed by an immediate downstream TAA or TAG stop codon. This concept is true in 42% of *E. coli* genome. In case of TAG codon only 27% will have the double stop codons. The tendency of double stop signals is also noticed in yeasts and ciliates [31,32].

Selection of Proper Expression Host

There are many expression hosts available for *E. coli*. The efficiency of protein expression results from appropriate selection of expression host. All of these hosts have the advantages and disadvantages. Invariably, the initial expression of the protein expression is analyzed by BL21(DE3) or derivatives of K-12 lineage strains. A couple of major characteristics of BL21 cells are; they are deprived of Lon protease which degrades many foreign proteins [33]. K-12 lineages are also used for checking the basic protein expression. The AD494 and Origami strains are *trxB* mutants which are capable enhancing the disulfide bond formation in the cytoplasm [34]. Another strain of K-12 lineage is HMS174, a *recA* mutant version which has a positive effect on plasmid stability [35].

Large number of bacterial hosts have been selected and tested for efficient expression of proteins. Some of the strains are modified to improve recombinant protein expression. These strains are generally defective in protease such as Lon or other outer membrane protease, *OmpT*. The preferred choice of expression hosts are BL21 and its derivatives. Some of the hosts' strains are discussed in detail [36,37].

BL21(DE3): This strain is one of the widely used strain to check the basic protein expression in *E. coli*. Chromosomal DE3 prophage expresses T7 RNA polymerase under control of *lacUV5* promoter. It is noticed that BL21 derivatives lack Lon and *OmpT* proteases which will stabilize expression of some recombinant proteins. While using this strain it is estimated that there is a chance of significant basal T7

expression or leaky expression. Addition of 1% glucose to the growth medium will reduce the leaky expression. Even though, addition of 1% glucose is recommended for toxic clones it can be applied to several other clones too which have leaky expressions.

BL21(DE3) pLysS: This strain has similar characteristics of BL21(DE3). However, pLysS produces wild-type T7 lysozyme to reduce basal T7 expression of the gene of interest. BL21(DE3)pLysS strain is compatible with plasmids containing the ColE1 or pMB1 origin. Culturing of BL21(DE3)pLysS requires Chloramphenicol.

Lemo21(DE3): pLemo plasmid produces amidase negative T7 lysozyme (lysY) from a tunable promoter (Prha). This strain is compatible with plasmids containing the ColE1 or pMB1 origin. Most of the pET vectors are compatible with this strain. Chloramphenicol is required to maintain this strain.

BL21-AI: T7 RNA polymerase gene is controlled by the *ParaBAD* promoter. While using this strain, if pET vectors are used IPTG is also required for induction to titrate *LacI* repressor away from the T7-lac promoter on the vector. This strain is recommended while using pDEST vectors

TOP10: These are K - 12 strains that do not metabolize L-arabinose but may provide slight improvement when expressing membrane protein directly from *ParaBAD*. The strain is suitable for pBAD vectors.

C41(DE3): These strains are derivatives of BL21(DE3) which has lower levels of T7 RNA polymerase under non-inducing and inducing conditions. In this strain, glucose addition is not necessary as with BL21(DE3).

Tuner(DE3): This strain is a *lacZY* derivative of BL21. The lac permease mutation (*lacY1*) allows uniform entry of IPTG into all cells

in the population, which produces a concentration - dependent level of induction. The modified version of this strain is Tuner (DE3) pLysS. This is to express T7 lysozyme to control T7 expression in addition to the lac permease mutation.

Rosetta2 and Rosetta pLysS: Rosetta host strains are BL21 *lacZY*(Tuner) derivatives designed to enhance the expression of proteins which contains rare codons used in *E.coli*. These strains express tRNAs for rare codons on a compatible CamR plasmid. In case of pLysS strain, the rare tRNA genes and T7 lysozyme gene are carried by the same plasmid.

BL21 CodonPlus RIL and CodonPlus(DE3)-RIL/RIPL: BL21-Codon Plus strains are engineered to contain extra copies of genes that encode tRNAs which frequently limit the translation of heterologous proteins in *E. coli*. BL21-CodonPlus-RIL and BL21-CodonPlus (DE3)-RIL cells contain extra copies of the argU, ileY, and leuW tRNA genes. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively. The CodonPlusRIL strains have available the tRNAs that most frequently restrict translation of heterologous proteins from organisms that have AT-rich genomes. BL21-CodonPlus-RP and BL21-CodonPlus(DE3)-RP cells contain extra copies of the argU and proL genes. These genes encode tRNAs that recognize the arginine codons AGA and AGG and the proline codon CCC, respectively. The CodonPlus-RP strains have available the tRNAs that most frequently restrict translation of heterologous proteins of organisms that have GC-rich genomes. The BL21-CodonPlus (DE3)-RIPL cells contain extra copies of the argU, ileY, and leuW as well as the proL tRNA genes. This strain rescues expression of heterologous proteins from organisms that have either AT- or GC-rich genomes. Some of the features of different expression host are given in detail in Table 1.

Expression strain	Induction method	Advantages	Disadvantages
BL21	Infection/induction with Lambda bacteriophage CE6	Tightest control of Un-induced expression	The process of induction is tedious and the induction is not as efficient as DE3 derivatives
BL21(DE3)	sopropyl-1-thio-β-Dgalactopyranoside (IPTG) induction of T7 polymerase from lacUV5 promoter	High level of protein expression	Leaky expression of T7 polymerase can lead to uninduced expression of potentially toxic proteins
BL21(DE3)pLysS	IPTG induction of T7 polymerase	Ease of induction	Slight inhibition of induced expression when compared with BL21(DE3)
Lemo21(DE3)	IPTG induction of T7 polymerase	Optimizes overexpression of any given protein using only a single strain. Outperforms other systems in its ability to maximize the production of both routine and difficult-to-overexpress proteins.	The exact insight in the mechanism by which optimized expression yields are achieved in Lemo21(DE3) is lacking. It is not sure, whether the over expressed proteins are suitable for functional and structural studies.
BL21-AI	Induction of T7 polymerase with IPTG and arabinose	Promotes tight regulation and high yields, especially used for high level expression of toxic protein	Testing against a wider variety of proteins is necessary to demonstrate broad utility
C41(DE3) and C43 (DE3)	IPTG induction of T7 polymerase	The mutant strains C41(DE3) and C43(DE3) can minimize the phenomenon of plasmid instability for toxic proteins	Testing against a wider variety of proteins is required to demonstrate the broad utility
Tuner(DE3), and Tuner(DE3) pLysS	IPTG inducible T7 polymerase	The lac permease (<i>lacY</i>) mutation allows uniform entry of IPTG into all cells in the population. Expression can be regulated from very low expression levels up to the robust. Tuner(DE3)pLysS helps in tighter control of over expression	Different studies are to be carried out to understand the wider usage of the strain
Rosetta2 and Rosetta pLysS	IPTG inducible T7 polymerase	These strains are designed to enhance the expression of eukaryotic proteins that contain codons rarely used in <i>E. coli</i> .	The codon specificity might cause different problems when the proteins are expressed in high levels.
BL21 CodonPlus RIL and CodonPlus(DE3)-RIL/RIPL	IPTG inducible T7 polymerase		

Table 1: Different expression host for producing recombinant proteins and their main features.

Stability Enhancement of Recombinant Proteins

Protein stability is one of the important aspects in recombinant protein production. Stability of the protein is very important in case of its purification, formulation and storage. It is noted that, the properly folded proteins are stable during expression and purification. Some of the proteins appear to be unstable and insufficient amount of protein is produced. Various aspects affecting protein instability are like amino acid sequences of the protein, protein construction, host cell strain, expression and purification conditions will all affect the stability of the protein.

There are instances where the amino acid sequence of a protein itself is prone to degradation. Certain amino acids like Arg, Lys, Leu, Phe, Tyr and Trp residues at the N-terminus region can lead to protein degradation. Replacing these amino acids with compatible amino acids can greatly enhance the protein stability [38]. Many recombinant proteins are expressed with tags or fusion partners to prevent proteolytic degradation and increase the stability.

Some of the parameters which can improve the stability of recombinant proteins are like addition of special media for protein production. Customized media containing trace metals, minerals and vitamins can be supplied for enhancing the stability of the protein. Even though these chemicals may not be needed for host cell growth, they may still serve as co-factors, prosthetic groups or ligands for recombinant proteins. But these will be crucial in obtaining correct protein folding and stability. Medium pH should also be neutral to improve the stability of the protein. Induction of protein at lower temperature and for shorter duration will enhance the protein stability. Sometimes, changing the expression host also can result in getting the stable protein. Protein expression localization can also lead to the production of stable protein. For example, if streptokinase is expressed as soluble protein, it is tend to be unstable. But when it was directed to be expressed as inclusion bodies, it was found to be stable [39]. Lipid modifications of recombinant proteins can enhance stability of the protein. Lipid modification will effectively stabilize the molecule without perturbing its structure and function [39]. Several studies had been carried out to exploit the lipid modification strategies to enhance the stability of the protein. Generally, the non-lipoprotein is genetically engineered and converted into lipoprotein in *E. coli* and was proved to be successful [40]. Finally, molecular chaperon co-expression will also result in enhanced stability of the protein.

Factors Enhancing Recombinant Protein Yield

It is important to increase the yield while producing recombinant proteins. It is mainly controlled at the transcriptional level. DNA replication and post translational modifications also play an important role in protein yield. Some of the factors are discussed in detail.

Appropriate usage of vector for expressing the protein plays a major role in protein yield. The expression vector must contain structural units that allow protein expression. The structural units which have impact on protein yield are mainly promoter, ribosome binding site, start codon, stop codon and a terminator. Additionally, the expression vector should also contain a selection marker and origin of replication. These structural units determine the protein yield and expression level.

The strength of promoter determines mRNA expression level of recombinant protein. Usage of stronger promoter will result in getting a higher yield. In case of toxic protein, weaker promoters are used [41]. Another important factor which results in higher protein yield is ribosome binding site which is also known as Shine-Dalgarno sequence. The consensus rbs sequence is UAAGGAGG. It is reported

that secondary structure of rbs is important for ribosome binding or translation initiation [42]. Changes in rbs sequence can change expression levels over several magnitudes. The affinity between the rbs and ribosome is a critical factor influencing the efficiency of protein expression [43,44]. The distance between the rbs and the start codon also plays an important role in protein yield [45]. It is also reported that diverse protein production may be obtained from a different rbs for a protein. For a chosen recombinant protein, different rbs may give different expression level. Usage of stop codon also has an effect on protein expression. UAA gives the maximum protein expression level compared to other stop codons. The transcription terminator forms secondary structure at 3' end of the mRNA and gives stability to the protein produced. The origin of replication determines the copy number of the gene expressed. More number of genes are present higher the efficiency of protein expression. Selection marker also plays an important role in protein yield.

Generation of Successful Combination

There are number of options while designing the experiments for producing recombinant proteins in *E. coli*. Choosing the perfect combination is not possible prior to performing the experiments. Thus several trial and error methods have to be optimized to adapt the appropriate strategy for the production desired proteins. Currently, Bioinformatics tools are available to check the feasibility and yield of protein production prior to conducting the experiment. However, the successful combination can be obtained only through continuous experimental methods

Trouble Shooting in Recombinant DNA Expression in *E. coli*

It is common that the protein of interest is expressed so poorly in the heterologous system after all the preliminary checking. Possible reasons for poor protein expression are due to toxicity in the host cell, insolubility, or mRNA secondary structure preventing interactions with cellular machinery. Rarely, the gene of interest is rich in codons that are not inconsistent with the host strain's available supply of tRNAs. Unrestrained basal expression of desired protein can affect host cell growth resulting in decreased protein yield. However, the excessive robustness in induction will result in the formation of inclusion bodies. Exporting the protein to periplasm or to the inner membrane introduce more complications for targets that must be folded with disulfide bonds or incorporated into a membrane.

Effect of Composition of Medium

To optimize the level of expression, it is necessary to fine tune the culture conditions and culture medium because it is much cheaper and easier to manipulate the media compositions required for culture growth [46,47]. Concentration of some salts, peptone and yeast extract can increase the concentration of desired recombinant protein concentration [48,49]. Various media like LB, TB and 2YT can be used to optimize the protein concentration. Alternatively, addition of prosthetic groups or cofactors which are essential for proper folding or protein stability in the culture medium will prevent the formation of inclusion bodies. This will also enhance the protein solubility. While producing recombinant protein, the aggregation of protein secreted into the periplasmic space can be suppressed by allowing the cells to grow in comparatively high concentration of polyols like sorbitol or sucrose. The increase in osmotic pressure by these cofactors results in accumulation of osmo-protectants in the cell which stabilize the native protein structure. Other parameters or growth additives in the media

which can enhance the protein expression are ethanol, which will help in the expression of heat shock protein, low molecular weight thiols and disulfides which affect the redox state of the periplasmic space, thus influencing the disulfide bond formation and NaCl.

Choice of Expression Hosts

Choice of expression host for the production of recombinant protein also plays a major role in increasing the desired protein concentration. BL21 and its derivatives are routinely used for the recombinant protein production in *E. coli*. These host strains are deficient in Ion and OmpT proteases, which is responsible for increased protein stability. Different *E. coli* strains facilitate the expression of membrane proteins, proteins with rare codons, proteins with disulfide bonds, and proteins that are otherwise toxic to the cell. Most of the points about the choice of expression host are discussed in the previous section on selection of proper expression host.

Solutions for the Problems of Plasmid Instability

The most common method of retaining the stability of recombinant plasmid is the addition of selection antibiotics to culture media. This might not be feasible when the culture is taken to the large scale. An alternative strategy adapted to this is that usage of runaway-replication plasmid vectors. In this method, the plasmid copy number is relatively low at lower temperatures and increased when the temperature is elevated. Plasmid copy number is controlled by plasmid and host genetics and also by cultivation condition such as growth rates, media and temperature.

Solution to the Problem of mRNA Instability

Apart from the plasmid instability, the mRNA instability also plays a major role in controlling the recombinant protein expression. One of the existing used solutions to the mRNA instability is the addition of short specific DNA sequence to the distal end of the cloned gene. This will enhance the stability of mRNA transcribed thereby increasing the gene expression. It is also studied the rho-independent terminator of the mRNA can also stabilize the mRNA by protecting it from degrading exonucleases.

Addition of Fusion Tags

Addition of fusion tags in the protein sequences also can enhance the yield of protein, increase the solubility, and even promote proper folding of the protein.

There are number of other parameters which can be taken into account while trouble shooting the recombinant protein production in *E. coli*. Optical density during induction, inducer concentration, post induction time, usage of effective terminator codons is some of the strategies which can be implemented to enhance the desired protein production.

Challenges in Therapeutic Protein Production in *E. coli*

Though there are many advances and advantages in using *E. coli* expression system for the production of recombinant proteins, there are many challenges and hurdles in front to actualize the protein expression in *E. coli*. Some of the challenges are discussed in detail.

Most frequently utilized termination codon in bacterial genome is UAA followed by UGA and UAG. During translation, error in reading the termination codon lead to extended protein synthesis until another termination codon is encountered in the mRNA. Extended

reading results in producing larger peptide with several additional C-terminal amino acids. During the production of IFN- α 2b the termination codon UAA was replaced by UGA and this resulted in a 2 fold increase in protein expression level [50]. The reason behind this is that the transcription terminators stabilize the mRNA by creating a stem loop structure at the 3' end of the mRNA [51]. It is also found that the efficiency of translation termination can be improved by adding consecutive stop codons or by using a prolonged UAAU stop codon [52].

Future Perspectives of Recombinant Protein Production in *E. coli*

Even though there are different heterologous expression systems to produce recombinant proteins, the all-time favored heterologous expression host is *E. coli*. Thus novel technological advancements are unceasingly being prepared to advance the *E. coli* expression system. Major reasons for preferring *E. coli* expression system is due to the ease of genetic manipulations, well-characterized genome, availability of versatile plasmid vector, accessibility of different host strains, cost-effectiveness, and high expression levels of desired protein. However, there are certain limitations in efficiently and widely using *E. coli* system for the production of recombinant proteins. Biased codon usage, protein solubility, mRNA stability, and lack of post-translational modifications are some of them. Due to the presence of rare codons translational errors occur which result in mutation and production of undesired products. Thus, while expressing recombinant proteins in *E. coli*, paramount importance to be given for the usage of appropriate codons to express the protein in *E. coli*. The enhancement of expression level is achieved by replacing the rare codons with more favorable major codons. Correspondingly, the co-expression of genes encoding for tRNAs for rare codons could increase the expression level of therapeutic proteins. Additionally, directing the protein secretion to the periplasm offers several advantages. It helps in proper folding, solubility, ease in purification, and higher yield of specific protein. Recent advancements in therapeutic protein production in *E. coli* have shown that *E. coli* strains can be modified especially for each therapeutic protein to achieve high product yield as well as high quality products.

Conclusion

The ideal expression system for *E. coli* should be composed of DNA elements which are efficient in transcription, and powerful translation. This can lead to producing an authentic recombinant protein which have no truncation or extended version. It should not be also toxic to the organism. Such an ideal expression system should have a consensus promoter. Efficient transcription terminator will minimize the drain of cellular energy and will reduce the metabolic burden for the host. The transcription terminator should be able to form secondary structure at 3' end to improve the stability and protein yield. Appropriate host selection will favor the protein yield and enhance the stability of the protein. There is no ideal expression system working with all recombinant proteins. Every protein poses a new problem, high level synthesis and stability has to be optimized in each single case by empirical variation of the different parameters.

References

1. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 27: 297-306.
2. Adrio JL, Demain AL (2010) Recombinant organisms for production of industrial products. *Bioeng Bugs* 1: 116-131.
3. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* 307: 249-264.

4. Yin J, Li G, Ren X, Herrler G (2007) Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. J Biotechnol 127: 335-347.
5. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev 60: 512-538.
6. Goldstein MA, Doi RH (1995) Prokaryotic promoters in biotechnology. Biotechnol Annu Rev 1: 105-128.
7. Yansura DG, Henner DJ (1990) Use of *Escherichia coli* trp promoter for direct expression of proteins. Methods Enzymol 185: 54-60.
8. O'Connor CD, Timmis KN (1987) Highly repressible expression system for cloning genes that specify potentially toxic proteins. J Bacteriol 169: 4457-4462.
9. Thomas CD, Modha J, Razzaq TM, Cullis PM, Rivett AJ (1993) Controlled high-level expression of the lon gene of *Escherichia coli* allows overproduction of Lon protease. Gene 136: 237-242.
10. Miksch G, Bettenworth F, Friehs K, Flaschel E, Saalbach A, et al. (2005) Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in *Escherichia coli*. J Biotechnol 120: 25-37.
11. De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. BMC Biotechnol 7: 34.
12. Lee SK, Chou HH, Pfleger BF, Newman JD, Yoshikuni Y, et al. (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. Appl Environ Microbiol 73: 5711-5715.
13. Hayashi MN, Hayashi M (1985) Cloned DNA sequences that determine mRNA stability of bacteriophage phi X174 in vivo are functional. Nucleic Acids Res 13: 5937-5948.
14. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297-310.
15. Vasquez JR, Evin LB, Higaki JN, Craik CS (1989) An expression system for trypsin. J Cell Biochem 39: 265-276.
16. Chamberlin MJ (1992) New models for the mechanism of transcription elongation and its regulation. Harvey Lect 88: 1-21.
17. Richardson JP (1993) Transcription termination. Crit Rev Biochem Mol Biol 28: 1-30.
18. Richardson JP, Greenblatt J (1996) Control of RNA chain elongation and termination, *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D.C.
19. Yang MT, Scott HB 2nd, Gardner JF (1995) Transcription termination at the thr attenuator. Evidence that the adenine residues upstream of the stem and loop structure are not required for termination. J Biol Chem 270: 23330-23336.
20. Adhya S, Gottesman M (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. Cell 29: 939-944.
21. Nishihara T, Iwabuchi T, Nohno T (1994) A T7 promoter vector with a transcriptional terminator for stringent expression of foreign genes. Gene 145: 145-146.
22. Brosius J, Ullrich A, Raker MA, Gray A, Dull TJ, et al. (1981) Construction and fine mapping of recombinant plasmids containing the rrnB ribosomal RNA operon of *E. coli*. Plasmid 6: 112-118.
23. Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol 146: 1-21
24. Sharp PM, Bulmer M (1988) Selective differences among translation termination codons. Gene 63: 141-145.
25. Brown CM, Dalphin ME, Stockwell PA, Tate WP (1993) The translational termination signal database. Nucleic Acids Res 21: 3119-3123.
26. Bonetti B, Fu L, Moon J, Bedwell DM (1995) The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. J Mol Biol 251: 334-345.
27. Pavlov MY, Freistoffer DV, Dincbas V, MacDougall J, Buckingham RH, et al. (1998) A direct estimation of the context effect on the efficiency of termination. J Mol Biol 284: 579-590.
28. Sañda F, Uzan M, Odaert B, Bontems F (2006) Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. Curr Protein Pept Sci 7: 47-56.
29. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270: 397-403.
30. Inamine JM, Ho KC, Loechel S, Hu PC (1990) Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. J Bacteriol 172: 504-506
31. Liang H, Cavalcanti AR, Landweber LF (2005) Conservation of tandem stop codons in yeasts. Genome Biol 6: R31.
32. Adachi M, Cavalcanti AR (2009) Tandem stop codons in ciliates that reassign stop codons. J Mol Evol 68: 424-431.
33. Gottesman S (1996) Proteases and their targets in *Escherichia coli*. Annu Rev Genet 30: 465-506.
34. Derman AI, Prinz WA, Belin D, Beckwith J (1993) Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. Science 262: 1744-1747.
35. Marisch K, Bayer K, Cserjan-Puschmann M, Luchner M, Striedner G (2013) Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. Microb Cell Fact 12: 12-58
36. Tabor S, Richardson CC (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc Natl Acad Sci U S A 82: 1074-1078.
37. Bass SH, Yansura DG (2000) Application of the *E. coli* trp promoter. Mol Biotechnol 16: 253-260.
38. Hosur V, Johnson KR, Burzenski LM, Stearns TM, Maser RS, et al. (2014) Rhbdf2 mutations increase its protein stability and drive EGFR hyperactivation through enhanced secretion of amphiregulin. Proc Natl Acad Sci U S A 111: E2200-2209.
39. Suthakaran P, Balasubramanian J, Ravichandran M, Murugan V, Ramya LN, et al. (2014) Studies on lipidification of streptokinase: a novel strategy to enhance the stability and activity. Am J Ther 21: 343-351.
40. Banerjee S, Bhat MA (2007) Neuron-glia interactions in blood-brain barrier formation. Annu Rev Neurosci 30: 235-258.
41. Rhodius VA, Mutalik VK (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE. Proc Natl Acad Sci U S A 107: 2854-2859.
42. de Smit MH, van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A 87: 7668-7672.
43. Komarova AV, Tchufistova LS, Dreyfus M, Boni IV (2005) AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. J Bacteriol 187: 1344-1349.
44. Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. BMC Mol Biol 8: 100.
45. Chen H, Bjerknes M, Kumar R, Jay E (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. Nucleic Acids Res 22: 4953-4957.
46. Vincentelli R, Bignon C, Gruez A, Cavanaugh S, Sulzenbacher G, et al. (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. Acc Chem Res 36: 165-172.
47. Chambers SP, Swalley SE (2009) Designing experiments for high-throughput protein expression. Methods Mol Biol 498: 19-29.
48. El-Helow ER, Abdel-Fattah YR, Ghanem KM, Mohamad EA (2000) Application of the response surface methodology for optimizing the activity of an aprE-driven gene expression system in *Bacillus subtilis*. Appl Microbiol Biotechnol 54: 515-520.
49. Larentis AL, Argondizzo APC, Esteves GD, Jessouron E, Galler R, Medeiros MA (2011) Cloning and optimization of induction conditions for mature PsaA (pneumococcal surface adhesin A) expression in *Escherichia coli* and recombinant protein stability during long-term storage. Protein Expr Purif 78: 38-47
50. C sar S nchez J, Padr n G, Santana H, Herrera L (1998) Elimination

-
- of an HufN alpha 2b readthrough species, produced in *Escherichia coli*, by replacing its natural translational stop signal. J Biotechnol 63: 179-186.
51. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297-310.
52. Poole ES, Brown CM, Tate WP (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. EMBO J 14: 151-158.