



Tehran University of Medical
Sciences Publication
<http://tums.ac.ir>

Iran J Parasitol

Open access Journal at
<http://ijpa.tums.ac.ir>



Iranian Society of Parasitology
<http://isp.tums.ac.ir>

Review Article

Polycistronic Expression of Multi-Subunit Complexes in the Eukaryotic Environment: A Narrative Review

Afshin Abdi Ghavidel¹, Vahid Jajarmi², Mojgan Bandehpour², *Bahram Kazemi^{2,3}

1. Student Research Committee, Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Cellular and Molecular Biology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received 21 Jan 2022
Accepted 15 Mar 2022

Keywords:

Leishmania;
Multi-subunit protein;
Polycistronic expression
system;
Trans-splicing;
Peptide

*Correspondence Email:

Bahram_14@yahoo.com

Abstract

Protein complexes are involved in many vital biological processes. Therefore, researchers need these protein complexes for biochemical and biophysical studies. Several methods exist for expressing multi-subunit proteins in eukaryotic cells, such as 2A sequences, IRES, or intein. Nevertheless, each of these elements has several disadvantages that limit their usage. In this article, we suggest a new system for expressing multi-subunit proteins, which have several advantages over existing methods meanwhile it, lacks most of their disadvantages. *Leishmania* is a unicellular eukaryote and member of the *Trypanosomatidae* family. In the expression system of *Leishmania*, pre-long RNAs that contain several protein sequences transcribe. Then these long RNAs separate into mature mRNAs in the process named trans splicing. For producing multi-subunit protein, *Leishmania* transformed with a vector containing the sequences of all subunits. Therefore, those subunits translate and form the complex under eukaryotic cell conditions. The sequence of each protein must separate by the spatial sequence needed for trans splicing. Based on a *Leishmania* expression pattern, not only is it possible to produce the complexes with the correct structures and post-translational modifications, but also it is possible to overcome previous method problems.

Introduction

A broader range of proteins needs for biochemical and biophysical studies. It is difficult to get these proteins from

natural sources. Because even if a large amount of animal and plant tissues or large volumes of biological fluids provided, just



Copyright © 2022 Abdi Ghavidel et al. Published by Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.
(<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited

small quantities of proteins would obtain. In addition, while scientists need to study variation and mutations in the main protein, a recombinant source is needed. Recombinant protein expression requires obtaining the gene sequence of the interested protein, inserting the sequence into an expression vector, and transforming the host with that vector (1). To express the desired recombinant protein in a large quantity, the host cell must be chosen wisely.

Some proteins act individually, therefore, they can be studied uniquely. However, some others, such as those acting in gene regulation, cell-cycle control, and metabolism pathway, are multi-subunit proteins, so for a detailed review, they must study together and in complex form. It is challenging to simultaneous and balance the co-expression of multiple proteins and isolate them for many critical applications such as the production of multimeric pharmaceutical proteins (2), synthesis of protein complexes for biochemical, biophysical, and structural studies (3), the study of metabolic pathways and metabolic networks regulation (4, 5). One method is to produce each complex component individually, isolate them from recombinant hosts, and form the complex under in vitro conditions. The yeast TFIIA/TBP1/DNA complex and the nucleosome are two examples of complex proteins produced by this procedure (6, 7).

Although such examples succeed with the mentioned method, complex formation yields in defined in vitro conditions are meager. Moreover, when each subunit of a complex is expressed individually, they might not fold into native structures of the complex protein. To overcome this problem, the co-expression of all subunits and folding of the complex in a cellular environment seems a better method (8). Myosin and TAFs are two examples of complex proteins produced by this method (9, 10).

Controlling the expression amount of each gene, in the case of using multiple

monocistronic vectors in the same host, is too tricky. Besides, transforming one cell by several plasmid vectors is challenging. Moreover, the same amount of all subunits cannot be reached by using the same type of promoter for all multiple monocistronic vectors (11). Using just one plasmid with sequences of all subunits seems a solution. However, if each gene has resided on the vector with the individual promoter, terminator, upstream, and downstream sequences, a large structure is created. This colossal structure tends to break during purification, is difficult to manipulate, and decreases transformation efficiency (12).

Polycistronic Expression System in prokaryote

In *Escherichia coli*, it is difficult to maintain more than two plasmids since each plasmid must have different antibiotics for selection and a unique compatible origin of replication. *E. coli* has a polycistronic expression system that can co-express several genes with just one promoter. Thus, there is no need for several antibiotic resistance genes, multiple plasmid transformations, controllable promoters for each gene, and compatible replication origins. In the polycistronic expression system method, one promoter regulates all downstream genes, and mRNA with the sequences of all those genes was produced. Therefore, all of them can be regulated together. Besides, the yield and quality of the desired complex in this method are higher than those in the noncellular situation. Translation of each subunit needs special elements named ribosome-binding site (RBS). The RBS sequence must be placed upstream of each gene (Fig.1). RBS is responsible for recruiting a ribosome during the initiation of protein translation. Therefore, downstream sequences of RBS translate into proteins. In an mRNA, there can be only one RBS sequence, which is called monocistronic mRNA, and only one type of protein translates by that mRNA (Fig.1.a), or it can have several RBS sequences, which is called

polycistronic mRNA, and they translate into several different proteins (Fig.1.b). NuA4 his-

tone acetyltransferase is one of the successful complexes obtained with this method (13).

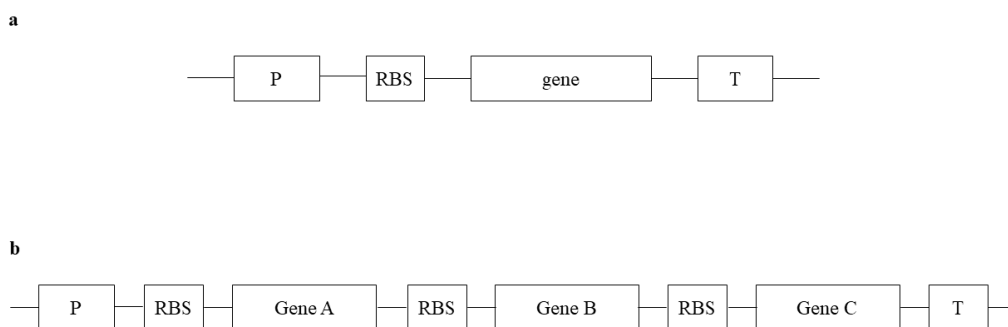


Fig.1: **a.** Monocistronic structure. **b.** Polycistronic structure. (P: promoter. RBS: ribosome binding site. T: transcriptional terminator.)

Post-translational modifications (PTMs) are essential for some proteins to fold correctly and have a function. Therefore, they must be produced in a eukaryotic host to obtain those proteins with the correct structure and function. In eukaryotic, unlike prokaryotes, a polycistronic expression system does not exist. Several elements detected in many studies can act as RBS in eukaryotic hosts. We will review them below and add some more arguments about them.

Internal Ribosomal Entry Sites (IRES)

Although most eukaryotes use the cap-dependent mechanism for translation initiation, some of them utilize cap-independent

translation mechanisms (14). In the cap-independent method, translation initiates with secondary structural RNA at the 5' untranslated region (5' UTR). A special element in the 5' UTR of those mRNAs, named IRESs, mediate internal translation initiation. This element can use for polycistronic expression in eukaryotic cells (Fig.2.a). A long mRNA is transcribed from those genes. The first gene has the cap-dependent translation, and the rest of the gene(s), which has (have) an IRES sequence in its (their) 5' UTR, initiates translation with the cap-independent mechanism (Fig.2.b).

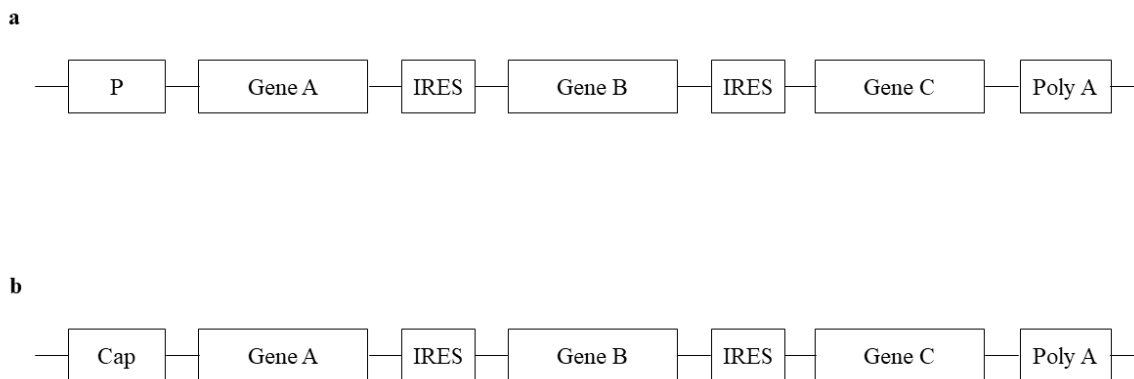


Fig.2: **a.** DNA structure. **b.** Mature RNA structure. Gene A has the cap and initiates cap-dependent translation. Gene B and C have the IRES and initiate translation with the cap-independent mechanism (P: promoter. IRES: Internal Ribosome Entry Site.)

The main problem of this method is the lower level of the translation in cap-independent initiation compared to the 5'-cap mediated initiation (11). In addition, some of these elements have tissue specificity, and each IRES in a different type of cell has a different transcription rate (15). The co-expression of interleukin 12 subunits is one of the best examples of this method (16).

2A peptides

2A is the name of a viral peptide that can use for polycistronic gene expression in eukaryotic cells. The 2A peptide is a self-cleavage peptide with cleavage activity between the last glycine and proline in its sequences (Fig.3.a). For the co-expression of multi-subunit protein, the 2A sequence must place between sequences of genes encoding complex subunits (Fig.3.b). Between the C terminal of the first protein and the N terminal of the second protein, a typical peptide bond forms because the termination codon is deleted from the C terminal of the first protein sequences. Therefore, ribosomes do not detect any stop codon at the end of the first protein sequences, and the first protein does not release from ribosomes (Fig.3.c). After completing the translation process, each protein was separated from the other(s) by the self-cleavage mechanism of the 2A peptide. A few extra amino acids remain attached to both proteins (the C terminal of the first protein and the N terminal of the second protein), and these amino acids can affect protein folding and functions (Fig.3.d). Sometimes the protein after the 2A sequence might not be translated (Fig.3.e) (17), or the polyprotein might not separate into its constituent proteins (Fig.3.f) (18). β -carotene and penicillin gene clusters are two successful examples of using the 2A peptide for polycistronic gene expression (19, 20).

Intein

The intein is a protein splicing element that autocatalytically and without any cofactors undergo self-excision from a protein precursor. The sequence of intein can use to separate polyprotein into its constituent proteins. Intein mediates a protein-splicing reaction of each polyprotein that has the intein sequence. This system, unlike previous methods, can proceed efficiently in both prokaryote and eukaryote cells. Nevertheless, in this method, several extra amino acids remain attached to each side of the proteins, and sometimes, these extra amino acids can change protein properties (11).

Leishmania expression system

Since recombinant proteins were studied, many organisms from prokaryote to eukaryote have been analyzed as expression hosts to produce different recombinant proteins with sufficient biological activity. Each host has a different codon usage pattern, variable post-translational modification process, and folding mechanism. *Leishmania* is a flagellated unicellular eukaryotic parasite from the *Trypanosomatidae* family. One of the non-pathogenic species of the genus *Leishmania*, named *L. tarentolae*, could be used as the available expression host for recombinant protein production (21). Production of recombinant proteins in *L. tarentolae* has several advantages, such as a higher growth rate of *L. tarentolae* compared to mammalian cells with a doubling time of 4–5 h, the same codon usage as human, and mammalian-type N-glycosylation pattern (22). However, the pathway of mRNA synthesis and regulation of transcription in *Leishmania* and other *Kinetoplastida* are different from most other eukaryotes (23).

Regulation of gene expression in *Leishmania* due to consensus promoter is different from other eukaryotes.

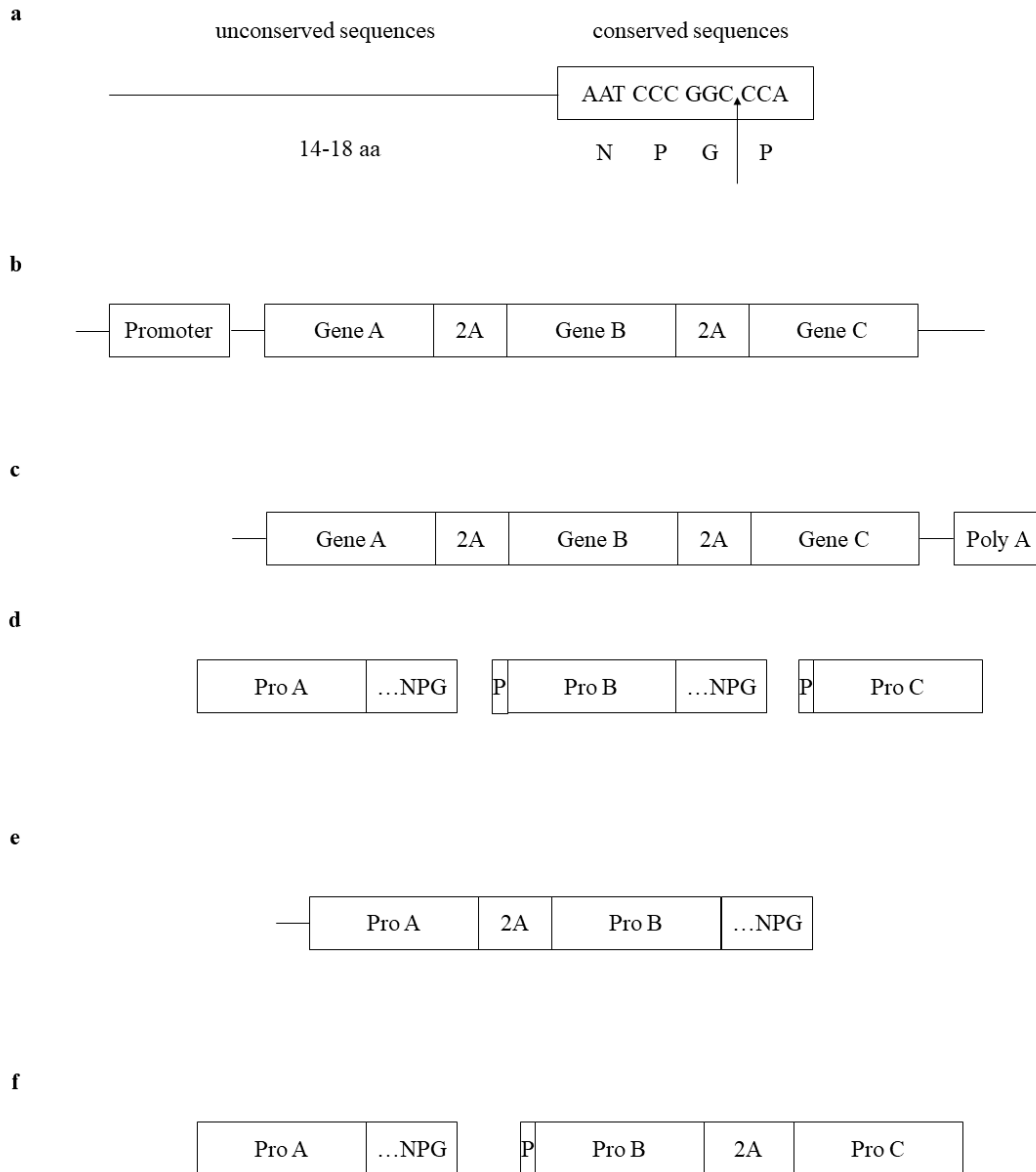


Fig.3: **a:** 2A peptides structure. Unconserved sequences contain 14-18 amino acids, conserved sequences have four conserved amino acids (asparagine, proline, glycine, proline). **b.** DNA structure for the co-expression of multi-subunit protein. **c.** mRNA structure contains sequences of multi-subunit protein. **d:** Translated proteins. Several amino acids from the 2A sequence remain in the C terminal of upstream protein, and proline from the 2A sequence remains in the N terminal of downstream protein. **e.** Uncompleted mRNA translation. By increasing the length of the translating protein, the possibility of releasing polypeptide from the ribosome before completing the translation is increased. **f.** The incomplete protein separation by the incomplete activity of 2A peptides. (aa: amino acid, N: asparagine, P: proline, G: glycine, ↑: self-cleavage site, Pro: protein, ...NPG: several amino acids from 2A sequence.)

RNA polymerases transcribe genes into polycistronic precursor RNA molecules. Then these pre-long RNAs are separated into con-

stituent mRNAs in a process named trans-splicing. Trans-splicing needs two independently transcribed RNA molecules, spliced

leader (SL) RNA and pre-long mRNA. Trans-splicing has begun with the addition of a 39-nucleotides named mini-exon (SL) sequence from the SL RNA to the 5' ends of each constituent mRNAs in pre-long RNAs. Then the mRNA capping processes start in the 5' ends of each mRNA. Finally, the poly (A) sequences will be added to the 3' ends of the upstream mRNA. The spliced leader (SL) was discovered in *Trypanosoma Brucei* when it had found that all the mature mRNAs in *Trypanosoma brucei* carry a typical 39-nucleotide sequence. The cap structure of *Leishmania* mRNA is named cap four because the first four transcribed nucleotides in mature mRNAs are modified (24).

Usually, the splice sites of each constituent mRNAs have AG dinucleotide (Fig.4.a), and the SL RNAs splice sites have a GU dinucleotide (Fig.4.b) (25). The site of trans-splicing has identical UUUCAG sequences as the site of cis-splicing (26). It is highly believed, polyadenylation couples with trans-splicing, however, the conventional polyadenylation signal (AAUAAA), do not cooperate in *Leishmania* polyadenylation processes (27). After cleavage at the splice site, the SL sequence, which has derived from SL RNA, joins to the 5' end of each mRNA, and at the same time, the upstream genes were polyadenylated (Fig.4.c). Finally, the rest of the SL RNA with the extra part of pre-long RNA will release (Fig.4.d)(28).

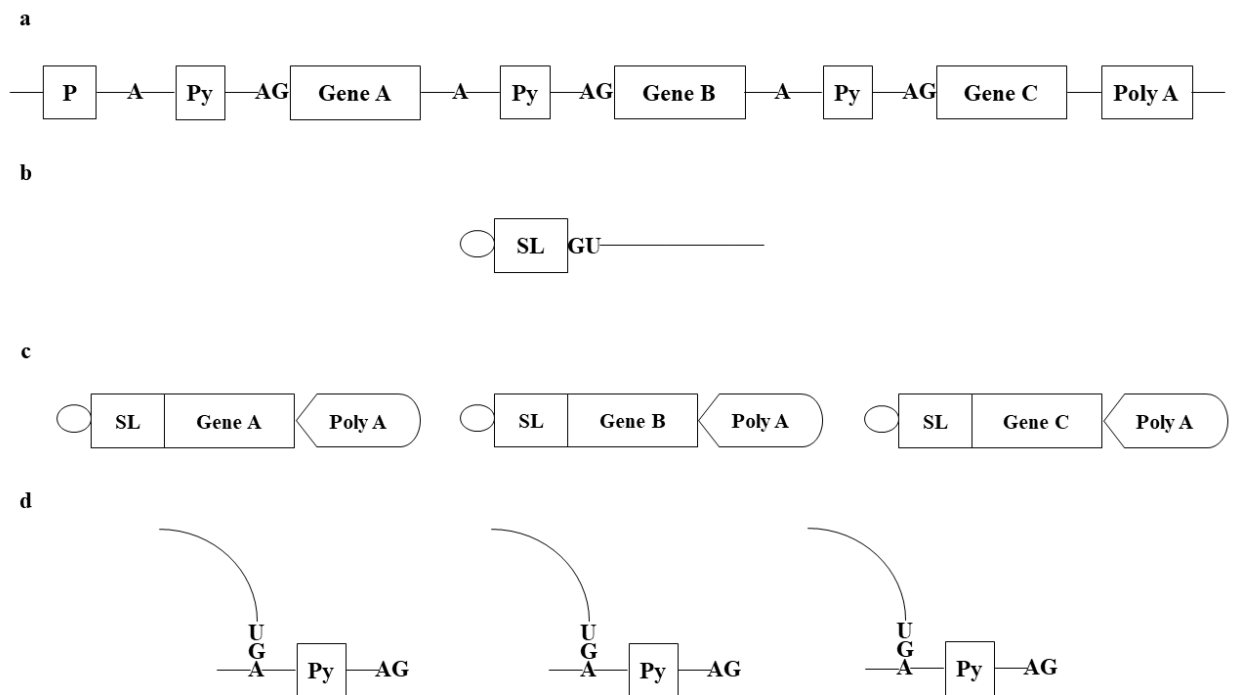


Fig.4: a. Gene's structure in *Leishmania*. b. SL RNA structure. c. Mature mRNA structure. In the process named trans-splicing, guanosine in SL RNA performs a nucleophilic attack to adenosine in 5' of each gene, and SL sequences from SL RNAs join to 5' end of each gene. d. Part of mRNA and SL RNA, which separate from the central part of pre-long RNA and SL RNA, are released after trans-splicing. (SL: Spliced Leader. P: promoter. G: Guanosine. U: Urasil. A: Adenosine. Py: polypyrimidine nucleotides. ○: Cap. 5'. ◻: Poly A tail.)

In *Leishmania*, each gene does not have separate promoters for itself. Therefore, there is

no regulation of gene expression at the transcription level, and all mRNAs transcript at

the same levels. It seems the translation levels of each mRNA depend on the degradation of the mRNAs. However, many mRNAs are much less abundant than expected from their half-lives (29). The regulatory activities were detected in sequences that present upstream of each mRNA. These sequences can regulate the efficiency of trans-splicing for different genes, and in different situations, enhance or repress the splicing rate of that mRNA (30).

Moreover, *Leishmania* must have alternative mechanisms for regulating protein concentrations in cells, including mRNA stability and protein half-life (31). mRNA stability for each gene can depend on several elements such as the type of SL sequences attached to each mRNA (32), elements which need for transporting mRNAs to the cytoplasm (33), regulatory motifs primarily exist in the 5' and 3' UTRs of the mature mRNAs (34), and (or) regulatory proteins bind to mRNAs (35). For example, if mRNAs have a binding site for RBP6 protein, the stability of those mRNAs will increase (36) or U-rich sequence in the 3' UTRs of mRNA, decreasing mRNA stability (37). *Leishmania* rarely has introns; therefore, *Leishmania* does not control gene expression by alternative cis-splicing (38).

The Hypothesis

This paper presents a new hypothesis on the host for the polycistronic expression of multi-subunit complexes. Based on the information mentioned above, it is possible to produce several proteins from a polycistronic vector in *Leishmania*. The vectors' structure must design like the *Leishmania* genome (Fig.4.A), which means the up and downstream of each gene must have special sequences that need for the trans-splicing process. If identical sequences are used for each gene, the same level of all subunits will produce. The recombinant genes transcribe to the pre-long mRNA. Trans-splicing separates each mRNA, then capping, and polyadenylation occurs. Then, mature mRNAs can lead to translation. It is possible to use different sequences upstream of each

mRNA to produce different subunit levels. After translation, all the complex subunits have almost the identical PTM as the mammalian proteins and fold in the eukaryotic situation. Another advantage of using *Leishmania* as a host cell is that the intron's presence in recombinant genes is not necessary to improve the expression level of those genes.

Heat shock protein 70 (HSP70) is one of the highly expressed proteins in *Leishmania*, which accounted for more than 2.1% of the total protein in *Leishmania* promastigotes (39). Based on what has been mentioned so far and higher expression of HSP70, the 3'UTR and 5'-UTR sequences of the HSP 70 are highly recommended for use in both sides of recombinant genes to achieve higher stable transcripts from recombinant genes. However, if the purpose is producing one subunit more than another, the 3'UTR of the L₂₃ transcripts can be used. LmjF.35. T3800 and LmjF.35. T3790 are two transcripts of L₂₃ protein. They have the same 5'-UTR and CDS, but their difference is in the 3'UTR sequence. The transcript level of LmjF.35. T3790 was 3-fold lower than the transcript level of LmjF.35. T3800. Therefore, these 3'UTRs can be used to express unequal levels of subunits.

Discussion

Various multi-subunit protein expression systems have been developed in bacteria, yeast, mammalian cells, and plants. The optimal expression system must be selected based on the advantages and disadvantages of the expression systems and the purpose of the desired protein.

The polycistronic expression system successfully expresses multi-subunit proteins in prokaryotes. However, PTMs in prokaryotes are different from eukaryotic which are essential for the folding and function of some proteins. Therefore, they must be produced in a eukaryotic host. One of the expression systems used to express the multi-subunit protein

in a eukaryotic environment is IRES. Meanwhile, several proteins are successfully expressed by this method in eukaryotes, the expression rate of multi-subunit protein was reduced because of the lower level of the cap-independent expression and tissue or cell specificity of each IRES. The 2A method is also used to express the multi-subunit protein in eukaryotic hosts. However, the 2A sequence can be used to produce several multi-subunit proteins, the self-cleavage activity of the 2A can be skipped, which affects the amount and structure of produced proteins. Although the intein system can produce the multi-subunit protein, extra amino acids that remain attached to each side of the proteins in this method may change protein properties.

The *Leishmania* expression system is approximately similar to the RBS method in prokaryotes, meanwhile, the PTM of producing proteins through this system is nearly the same as the mammalian proteins. In contrast to the 2A method, not only the amount of each subunit can be strictly determined by using the *Leishmania* expression system but also increasing the number of genes will not affect the expression rate of downstream genes. In comparison with the intein method, *Leishmania*-produced proteins do not have any extra amino acids.

Previously Sugino et al. expressed a large heterotrimeric glycoprotein in the *L. tarentolae*; however, they are used several different vectors with different antibiotics (40) instead of one vector and trans-splicing method for each subunit. In addition, they did not use different UTR to reach the different amounts of each subunit, which are essential in many multi-subunit proteins. Sangani et al. used the 2A method to produce two proteins, which was placed on one vector (41) in the *Leishmania*. Although the goal of this study was not to produce multi-subunit proteins, all the mentioned problems about the 2A method remain and none of the *Leishmania* expressing benefits

including polycistronic expression system were used.

In the case of the ability to produce different amounts of each subunit regardless of their correspondence gene position, *Leishmania* can be considered a novel effective host for the expression of multi-subunit proteins. These proteins cannot only fold into their unique 3D structures in eukaryotic cells, express the almost amount of each subunit, but also have nearly similar PTMs to mammalian proteins. Producing multi-subunit proteins in the *Leishmania* in addition to possessing those proteins in the eukaryotic environment, they are produced in the inexpensive host, which does not require extensive biosafety facilities, has high growth rates compared to that mammalian cells, can be cultured in simple, inexpensive nutrient media at 26 °C, and does not require a CO₂ incubator.

Conclusion

Based on mentioned information about the *Leishmania* expression system, it is possible to use the trans-splicing method and special UTR for producing multi-subunit proteins in this host similar to their natural counterparts with the same structure and function.

Acknowledgements

This study is related to the project NO 1399/655969 From the Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran. We appreciate the “Student Research Committee” and “Research & Technology Chancellor” at Shahid Beheshti University of Medical Sciences for their financial support of this study.

Conflicts of interest

The authors declare that they have no known competing financial interests or per-

sonal relationships that could have influenced the work reported in this paper

References

- Langlais C, Korn B. Recombinant Protein Expression in Bacteria. Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 1609-16.
- Hamorsky KT, Grooms-Williams TW, Husk AS, Bennett LJ, Palmer KE, Matoba N. Efficient single tobamoviral vector-based bioproduction of broadly neutralizing anti-HIV-1 monoclonal antibody VRC01 in *Nicotiana benthamiana* plants and utility of VRC01 in combination microbicides. *Antimicrob Agents Chemother.* 2013;57(5):2076-86.
- Tan S. A Modular Polycistronic Expression System for Overexpressing Protein Complexes in *Escherichia coli*. *Protein Expr Purif.* 2001;21(1):224-34.
- Brazier-Hicks M, Edwards R. Metabolic engineering of the flavone-C-glycoside pathway using polyprotein technology. *Metab Eng.* 2013;16:11-20.
- Vranová E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol.* 2013;64:665-700.
- Luger K, Rechsteiner TJ, Flaus AJ, Wayne MM, Richmond TJ. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol.* 1997;272(3):301-11.
- Tan S, Hunziker Y, Sargent DF, Richmond TJ. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature.* 1996;381(6578):127-51.
- Jaenicke R. Protein self-organization in vitro and in vivo: partitioning between physical biochemistry and cell biology. *Biol Chem.* 1998;379(3):237-43.
- Gangloff YG, Werten S, Romier C, et al. The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs. *Mol Cell Biol.* 2000;20(1):340-51.
- McNally EM, Goodwin EB, Spudich JA, Leinwand LA. Coexpression and assembly of myosin heavy chain and myosin light chain in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1988;85(19):7270-3.
- Zhang B, Rapolu M, Liang Z, Han Z, Williams PG, Su WW. A dual-intein autoprocessing domain that directs synchronized protein co-expression in both prokaryotes and eukaryotes. *Sci Rep.* 2015;5:8541.
- Ohse M, Takahashi K, Kadowaki Y, Kusaoke H. Effects of plasmid DNA sizes and several other factors on transformation of *Bacillus subtilis* ISW1214 with plasmid DNA by electroporation. *Biosci Biotechnol Biochem.* 1995;59(8):1433-7.
- Selleck W, Tan S. Recombinant protein complex expression in E. coli. *Curr Protoc Protein Sci.* 2008;Chapter 5:Unit-5.21.
- Kozak M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol Cell Biol.* 1987;7(10):3438-45.
- Créancier L, Mercier P, Prats AC, Morello D. c-myc Internal ribosome entry site activity is developmentally controlled and subjected to a strong translational repression in adult transgenic mice. *Mol Cell Biol.* 2001;21(5):1833-40.
- Zitvogel L, Tahara H, Cai Q, et al. Construction and characterization of retroviral vectors expressing biologically active human interleukin-12. *Hum Gene Ther.* 1994;5(12):1493-506.
- Liu Z, Chen O, Wall JBJ, et al. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports.* 2017;7(1):2193.
- de Felipe P, Luke GA, Brown JD, Ryan MD. Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnol J.* 2010;5(2):213-23.
- Beekwilder J, van Houwelingen A, Cankar K, et al. Valencene synthase from the heartwood of Nootka cypress (*Callitropsis nootkatensis*) for biotechnological production of valencene. *Plant Biotechnol J.* 2014;12(2):174-82.
- Unkles SE, Valiante V, Mattern DJ, Brakhage AA. Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem Biol.* 2014;21(4):502-8.
- Azizi H, Hassani K, Taslimi Y, Najafabadi HS, Papadopoulou B, Rafati S. Searching for virulence factors in the non-pathogenic parasite

- to humans *Leishmania tarentolae*. Parasitology. 2009;136(7):723-35.
22. Basile G, Peticca M. Recombinant protein expression in *Leishmania tarentolae*. Mol Biotechnol. 2009;43(3):273-8.
 23. Cuyper B, Domagalska MA, Meysman P, et al. Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method for RNA-seq in Trypanosomatids. Scientific Reports. 2017;7(1):3725.
 24. Freistadt M, Cross G, Robertson H. Discontinuously synthesized mRNA from *Trypanosoma brucei* contains the highly methylated 5'cap structure, m7GpppA* A* C (2'-O) mU* A. J Biol Chem. 1988; 263(29):15071-5.
 25. Mayer MG, Floeter-Winter LM. Pre-mRNA trans-splicing: from kinetoplastids to mammals, an easy language for life diversity. Mem Inst Oswaldo Cruz. 2005;100(5):501-13.
 26. Yague-Sanz C, Hermand D. SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data. Gigascience. 2018;7(7):giy084.
 27. Dillon LAL, Okrah K, Hughitt VK, et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. Nucleic Acids Res. 2015;43(14):6799-813.
 28. Kazemi B. Genomic organization of *Leishmania* species. Iran J Parasitol. 2011;6(3):1-18.
 29. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. Open Biol. 2019;9(6):190072.
 30. Liang X-h, Haritan A, Uliel S, Michaeli S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell. 2003;2(5):830-40.
 31. Requena JM. Lights and shadows on gene organization and regulation of gene expression in *Leishmania*. Front Biosci (Landmark Ed) [Internet]; 2011 2011/06//; 16:[2069-85 pp.]. <http://europepmc.org/abstract/MED/21622163>
 32. Lamontagne J, Papadopoulou B. Developmental regulation of spliced leader RNA gene in *Leishmania donovani* amastigotes is mediated by specific polyadenylation. J Biol Chem. 1999;274(10):6602-9.
 33. Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol. 2007;156(2):93-101.
 34. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet. 2007;8(7):533-43.
 35. Cuervo P, Domont GB, De Jesus JB. Proteomics of trypanosomatids of human medical importance. J Proteomics. 2010;73(5):845-67.
 36. Karamysheva ZN, Gutierrez Guarnizo SA, Karamyshev AL. Regulation of Translation in the Protozoan Parasite *Leishmania*. Int J Mol Sci. 2020;21(8):2981.
 37. De Gaudenzi JG, Noé G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. Essays Biochem. 2011;51:31-46.
 38. Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. Science. 2005;309(5733):436-42.
 39. Rastrojo A, Carrasco-Ramiro F, Martín D, et al. The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. BMC Genomics. 2013;14(1):223.
 40. Sugino M, Niimi T. Expression of multisubunit proteins in *Leishmania tarentolae*. Methods Mol Biol. 2012;824:317-25.
 41. Salehi Sangani G, Jajarmi V, Khamesipour A, Mahmoudi M, Fata A, Mohebbi M. Generation of a CRISPR/Cas9-Based Vector Specific for Gene Manipulation in *Leishmania major*. Iran J Parasitol. 2019;14(1):78-88.