
Evolutionary Molecular Engineering to Efficiently Direct *in vitro* Protein Synthesis

Manish Biyani, Madhu Biyani, Naoto Nemoto and Yuzuru Husimi

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52311>

1. Introduction

With the completion of human genome project in 2003, the 50th anniversary year of the discovery of the structure of DNA, we entered in the post-genomic era that concentrates on harvesting the fruits hidden in the genomic text. Since then we have witnessed the generation of a tremendous volume of DNA information (*genetic information*). As of September 2011, the Genomes OnLine Database (GOLD, <http://www.genomesonline.org>) has documented 1914 complete genome projects which comprise 1644 bacterial, 117 archaeal and 153 eukaryal genomes [1]. However, only a fraction of these DNA data are associated with their encoded proteins, i.e., their phenotypes (*functional information*) [2]. Even when a phenotype is associated with the encoding gene, the function of a particular gene cannot be fully understood until it is possible to describe all of the phenotypes that result from the wild-type and mutant forms of that gene. Moreover, unlike a genome that contains a fixed number of genes, the levels of proteins within cells are likely an order of magnitude greater than the number of genes. Therefore, the focus of the scientific community has recently been shifted from gene sequencing to annotation of gene function and regulation through elucidation of protein abundance, expression, post-translational modifications, and protein-protein interactions. While the pre-genomic era which lasted less than 15 years, the post-genomic era can be expected to last much longer, probably extending over several generations, and thus there is an increasing need for high throughput expression of the genome encoded proteins to profile the entire proteome and get a deeper understanding of protein abundance and reveal novel protein functions. Protein synthesis is therefore a powerful tool for large-scale analysis of proteins for a large-variety of low- and high-throughput applications (see Fig.1) and an essential tool for bridging the gap between genomics and proteomics in the post-genomic era. Noteworthy, the ribosome that catalyses and provide the platform for protein synthesis was in the spotlight recently, as the Nobel

Prize in Chemistry 2009 was awarded to the work that unlocked the structure and function of the ribosomes.

2. Significance of protein synthesis enhancer sequences (5'UTR)

Cell-based (*in vivo*) and cell-free (*in vitro*) methods have been developed for production of protein synthesis [3]. Cell-based host systems such as bacteria, yeast, worms, mammals used for protein synthesis and protein expression analysis, however, have been unable to meet the requirement of producing large amounts of purified and functional proteins which is a prerequisite to facilitate structure-based functional analysis. For example, purified proteins are necessary to grow protein crystals whose X-ray diffraction patterns provide the most precise structural information. Other limitations in host organisms includes such as bacteria don't have the intracellular organelles found in eukaryotes; yeast lack a dimension of complexity in intracellular communication observed in metazoans; and even other mammalian system are different from human in important aspects of both normal physiology and disease pathogenesis [4]. In addition, many biochemical pathways are simply difficult to study in the larger context of other events happening at the same time within the cell. In contrast to the cell-based systems, cell-free protein expression systems are now becoming the favored alternative with far greater fidelity as it offers a simple and flexible system for the rapid synthesis of functional proteins. There is currently a wide range of cell-free translation systems due to the ready availability of cell extracts prepared from various cell sources, including *Escherichia coli*, yeast, wheat germ, rabbit reticulocytes, *Drosophila* embryos, hybridomas, and insect, mammalian, and human cells [5-11]. Although encouraging, there would be some major issues in the use of cell-free systems. First, a major

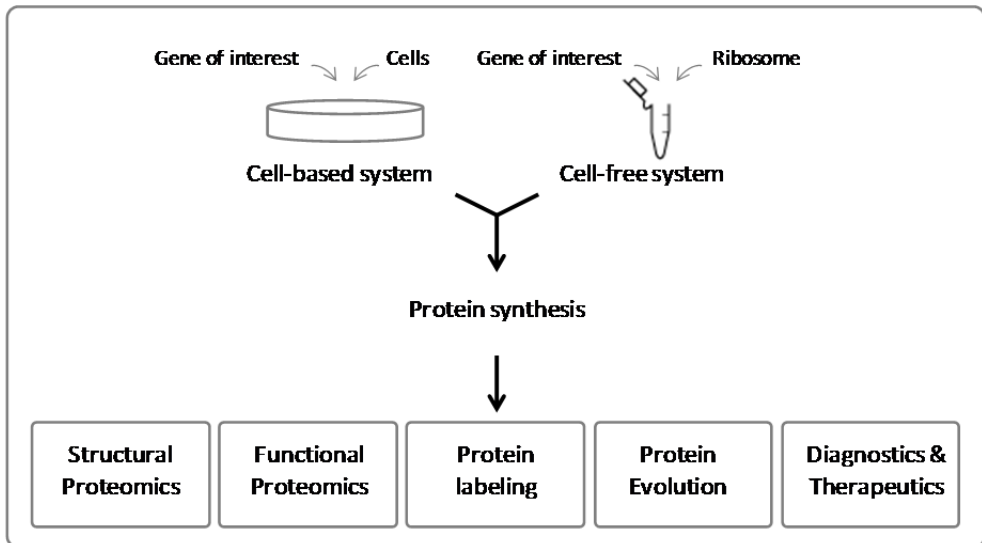


Figure 1. Application of Protein synthesis

drawback of synthesizing proteins in the lysate is that the lysate contains a large portion of the cellular proteins and nucleic acids that are not necessarily involved in the targeted protein synthesis and can lead to low protein yields through interfering with the subsequent purification reactions. In addition, the presence of proteases and nucleases in the lysates could be inhibitory to protein synthesis. In order to addressing this issue, cell-free protein synthesis system was reconstituted *in vitro* from purified components of the *E. coli* translation machinery. This system, termed the “protein synthesis using recombinant elements” (PURE) system, contains all necessary translation factors, purified with high specific activity, and allows efficient protein production [12]. Remarkably, this reconstituted system has been shown to catalyze efficient *in vitro* protein synthesis by providing a much cleaner background than a lysate-based system [13].

The second issue is that existing cell-free systems differ substantially from each other with respect to their efficiency and scalability to produce proteins and therefore these systems has to be programmed for given exogenous mRNA templates. Although different lysates may contain specific cellular factors that promote protein synthesis, a key factor in ensuring high protein production is the use of strong translational enhancer sequences (untranslated regions, UTRs) in the mRNA templates, which has long been known to enhance protein production up to several hundred-folds [14]. UTRs are known to play crucial role in the post-transcriptional regulation of gene expression, including modulation of the transport of mRNAs out of the nucleus and of translation efficiency [15]. The average length of UTRs motifs located at the 5′end of the exon, called 5′-UTR, ranges between 100 and 200 nucleotides and strikingly varies a lot within a species, e.g., in humans, the longest known 5′UTR is 2,803 while the smallest is just 18 nucleotides [16,17] (Fig.2).

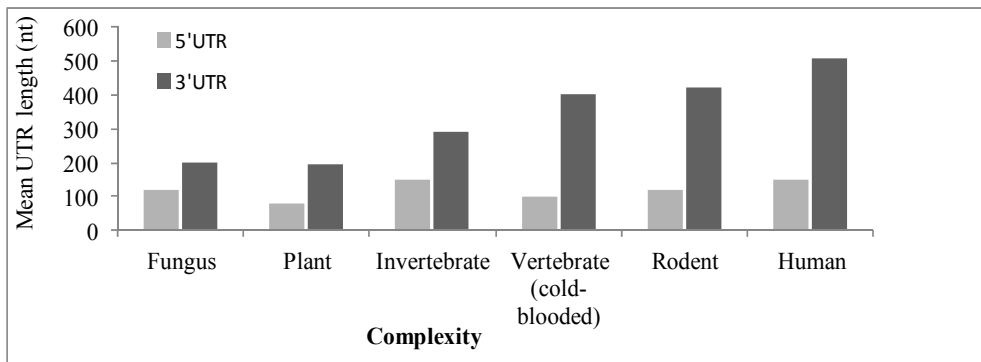


Figure 2. The average length of untranslated region sequences in the different taxonomic classes. Grey bar representing 5′-UTR and black bar is 3′-UTR.

The structural features of the 5′UTR have a major role in the control of protein synthesis. Those proteins which are involved in developmental processes, including growth factors, transcription factors or proto-oncogenes, often have longer 5′UTR than an average and thus

untranslated regions of mRNAs have crucial roles in protein regulation through protein synthesis. Structural elements of the eukaryotic mRNA, including the 5' cap and 3' poly(A) tail, and a series of protein-mRNA and protein-protein interactions, including several eIF (eukaryotic initiation factors), are important determinants of translation initiation (Fig.3). In eukaryotes, a multifactor complex of eukaryotic initiation factors are involved in the initiation phase of protein synthesis. But, in particular, 5'UTR plays a major role in the translation initiation, a critical step in protein synthesis which is determining qualitatively and quantitatively which proteins are made, when and where. 5'UTR is composed of several regulatory elements, including the Shine-Dalgarno (SD) and the AU-rich sequences which facilitates 16s rRNA-specific ribosome binding to initiate the protein synthesis [18,19]. In cell-based or in vivo systems, the translation of natural mRNAs is finely regulated by several mechanisms using 5'-capped and 3'-poly(A) containing long-untranslated regions (UTRs).

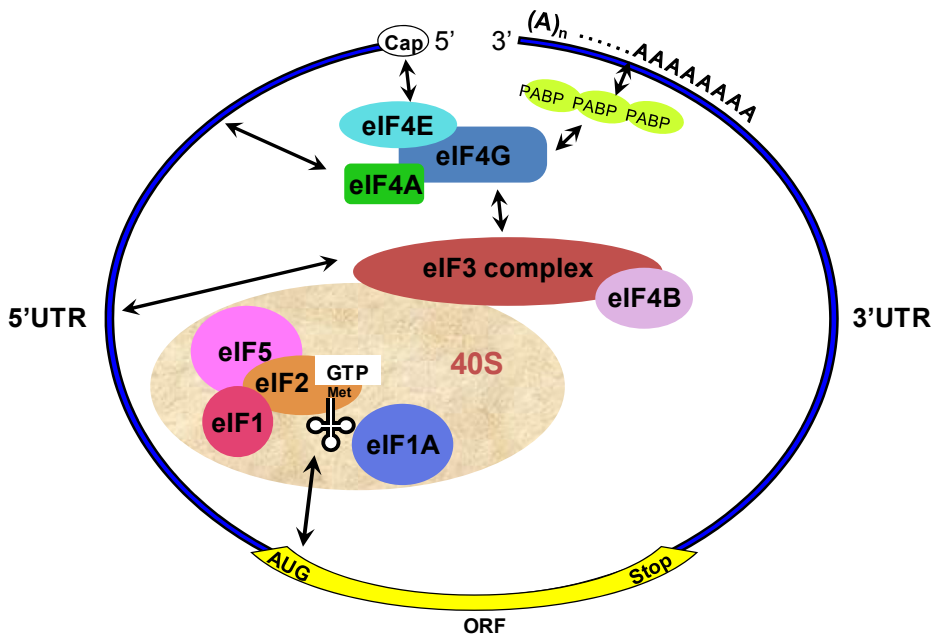


Figure 3. A schematic drawing of a eukaryotic mRNA, illustrating some post-transcriptional regulatory elements that affect initiation of protein synthesis.

Therefore, the efficiency of a cell-free translation system which is reconstituted using crude cell extract is restricted due to the problematic of maintaining long-natural UTRs in the in vitro construct. Even if so, the obvious question here is that “are the natural UTRs can meet the requirements of various translation factors in a cell-free system to carry similar mechanisms as in in vivo system?” Looking at this ‘black box’ may open a new window into the post-transcriptional regulation of gene expression using cell-free translation systems.

Therefore, it is prerequisite to find an alternate for natural UTRs dependency and optimization of translation initiation in cell-free system for next-generation *in vitro* high throughput protein synthesis systems. In a recent study using cell-free systems, the translation-enhancing activity of some commonly used natural enhancer sequences, such as omega from tobacco mosaic virus and the 5'UTR of β -globin mRNA from *Xenopus laevis*, was reported to vary from 1- to 10-fold, depending on the source of the cell-free extract used (e.g., wheat germ, rabbit reticulocyte lysate, insect) [20]. Therefore, optimization of enhancer sequences of an exogenous mRNA template with a given crude cell extract is desirable before using a cell-free protein synthesis system. A recent new development has been the remarkable generation of a universal cell-free translation system that mediates efficient translation in multiple prokaryotic and eukaryotic systems by bypassing the need for early translation initiation steps [21].

3. Co-evolutionary relationship between translational initiation and protein synthesis

In the course of evolution on the Earth, how the early life evolved beginning with a hypothetical RNA world-to-the world we know today (DNA world) is the persistent issue of debate for evolutionary biologist. In 1968, Francis Crick argued about the existence of the RNA world in the initial stage of evolution in which RNA molecules assembled from a nucleotide soup and supposed to carry both the genetic and catalytic information (Fig.4). In later stage, some special types of RNA molecules (now termed as Ribozymes) was considered to catalyzes its own self-replication and therefore to develop an entire range of enzymatic activities to form DNA world through an intermediate RNP (RNA/Protein) world. However, there are certain questions that cannot be answered with proposed RNP world. These include: 1. How did 'RNA-world (Ribozyme-type)' evolved to 'DNA-world (cell-type)' since there is no record exists of the intermediates between the RNA-world and organized complexity of cell? 2. What was the first Protein evolved out of an RNA world? 3. How could it have evolved and how the process of translation emerged? 4. If ribosome make protein then how the first ribosomal protein appeared? 5. Why is ribosome made half of protein and half of RNA ?

The recent advances in evolutionary molecular engineering have revealed the bonding strategy of the genotype to its phenotype as a unique and essential nature of a 'virus' and thus the role of virus-type strategy in the course of evolution on the Earth. In 1995, Nemoto and Husimi proposed a 'virus-early and cell-late model' that a virus-like molecule consist of genotype (mRNA) and phenotype (its coded protein) molecules emerged in the latter period of RNA world was the key molecule which enforced the transition from RNA-to-RNP world by co-evolving the translation system and a virus-like molecule coded a primitive protein of replicase [22]. In this theory, they also showed that such virus-like molecule could introduce Darwinian evolution into the Eigen's hypercycle members (RNA replicase of RNA, RNA translation members, RNA replicase of protein) resulting in carrying out co-evolution

between translation system and protein replicase. This was later reinforced by inventing and demonstrating a genotype-phenotype linked method (IVV, in vitro virus) for evolutionary molecular engineering [26] and this strongly suggest the potential of IVV method to understand the relation between ribosome-mRNA interaction.

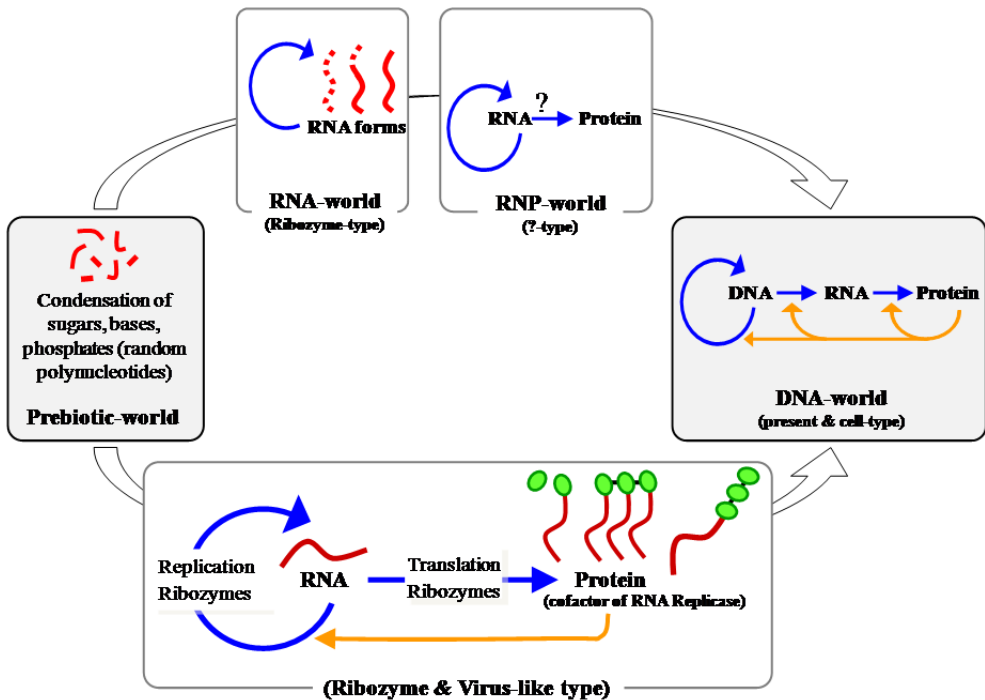


Figure 4. A schematic drawing to represent co-evolution of the translation initiation and protein synthesis system, prior to 'birth of first cellular life'.

4. Directed molecular evolution and screening of protein synthesis enhancer sequences

Directed molecular evolution mimics the natural Darwinian evolution process to evolve new functional molecules in the laboratory rather than in the jungle and in days rather than in millenniums and thus has emerged as a dominant approach for exploiting the sequence space to generate biomolecules with novel functions. Directed molecular evolution rely on the application of selection pressure to identify a bio-molecule with desirable properties from a diverse pools (or 'libraries') of bio-molecules with hundreds of millions of mutations and consist of four essential and repeating cycles: the creation of mutation and diversity at the DNA molecular level; the coupling of genetic information (DNA/mRNA) to functional information (Protein); the application of selection pressure; and the amplification of selected molecules (Fig.5).

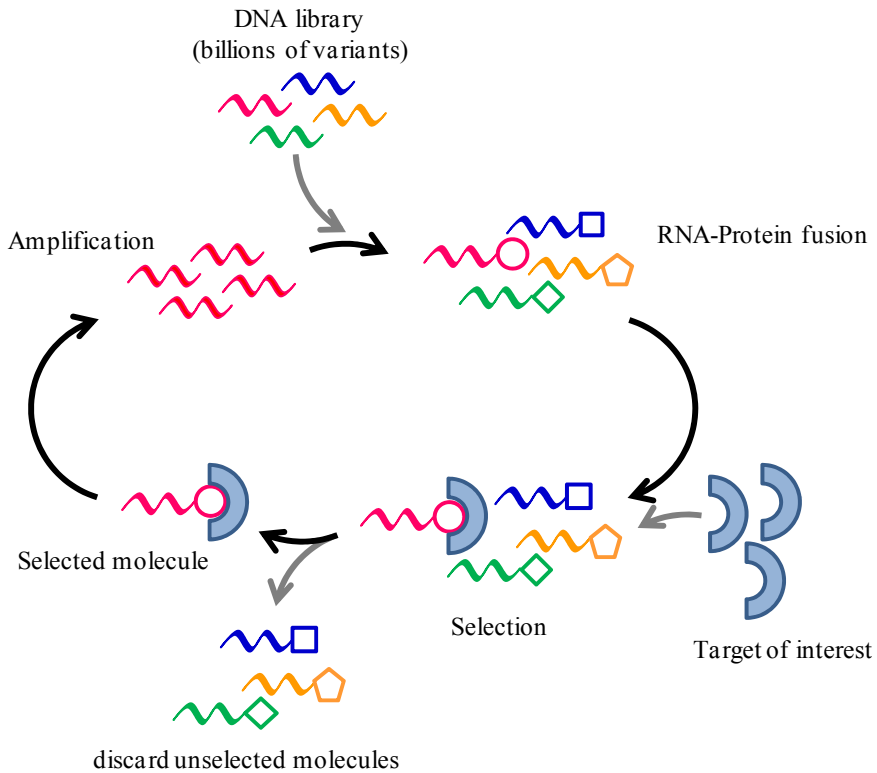


Figure 5. A schematic drawing of Directed Molecular Evolution

A number of well-established strategies, called display technologies, have been developed which use natural cell-based environment, such as yeast surface display, bacterial surface display, phage display or use a cell-free environment, such as ribosome display, mRNA display (in vitro virus), cDNA display, CIS display, IVC (in vitro compartmentalization) (Fig.6).

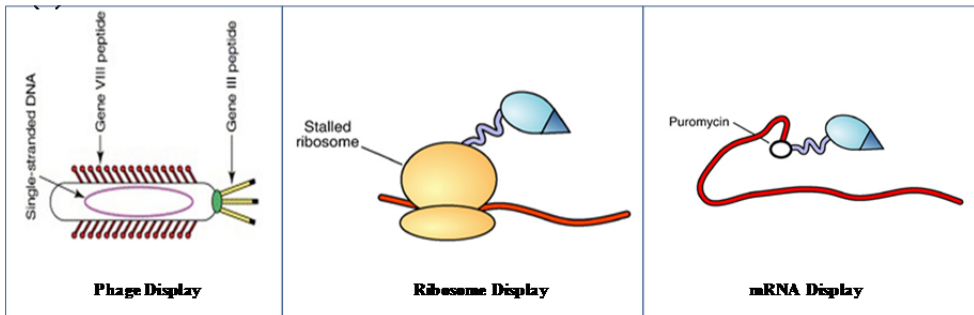


Figure 6. A schematic drawing of well-known strategies to perform Directed Molecular Evolution in the laboratory

Interestingly, a few groups have reported the application of directed molecular evolution to the screening of enhancer sequences with high translation efficiency in a cell-free translation system using ribosome display or polysome-mediated selection methods [23-25]. Recently, a novel strategy is also described for the *in vitro* selection of strong translation enhancer sequences for use in a cell-free translation system using an mRNA display method. The mRNA display method (originally called an “*in vitro* virus”) [26,27], which covalently links the mRNA molecule (genotype) to its encoded protein (phenotype), is a powerful evolutionary method for searching for functional protein molecules in a large-scale library. In this strategy, a simplified new gel shift assay system was developed to demonstrate that short but efficient translation enhancer sequences can be created for use in a given cell-free translation system (Fig.7). This method is based on an mRNA display method in which a covalent linkage is formed between the mRNA and the encoded protein through the antibiotic molecule puromycin. The steps involved in the synthesis of the covalently linked mRNA–protein fusion, and in the selection of 5'UTR sequences, are summarized below. First, a model gene construct is designed (Fig.7A) as a positive control (wt), which consists of a T7 promoter and a natural 5'UTR sequence (*X. laevis b-globin*) upstream of the PDO coding sequence. The stop codon is deleted to facilitate RNA–protein fusion, and a short DNA fragment complementary to a Puro-linker DNA sequence is ligated downstream of the coding sequence. Second, a random variable 5'UTR library is constructed by replacing the cognate secondary structure part of the *X. laevis b-globin* UTR sequence (36 nt) with a randomized 20-nt-long sequence with all possible combinations of the four nucleotides (N_{20}) (Fig.7B), resulting in an initial library size of approximately 10^{12} (4^{20}) molecules. Third, the cDNA library is then transcribed into an mRNA library using T7 RNA polymerase with/without the cap analogue (m7GpppG). Fourth, the 3'-terminal end of the mRNA library is ligated to a synthetic Puro-linker DNA. Fifth, the resulting mRNA–Puro-linker conjugate library is then used as a template in a given cell-free translation system and is converted into an mRNA–protein fusion library. Sixth, to select efficient 5'UTR candidates from inefficient ones, the resulting mRNA–protein fusion is analyzed using SDS–PAGE. As shown in Fig.7F, fusion products (translated products) of efficiently translated 5'UTRs will migrate with a decreased mobility compared with untranslated products from 5'UTR regions with no and/or slow translation efficiency. Thus, translated and nontranslated candidates can be distinguished, and translated candidates can be clearly identified, by a shift in the gel band pattern. Seventh, the fusion product of translated candidates is then carefully excised from the gel, and the associated mRNAs that represent selected 5'UTR candidates for efficient translation are directly reverse-transcribed and amplified using a single-step RT–PCR. This PCR step completes one round of selection. Finally, the selected 5'UTR candidates are then used as templates for a subsequent selection round for further enrichment of efficient 5'UTR sequences. Using this gel-shift assay, the translation of an mRNA template using a population of randomized 20-nt-long sequences upstream of a Pou-specific DNA-binding domain of Oct-1 (PDO) was screened with a rabbit reticulocyte extract and the time for translation was successively shortened. A total of five selection rounds were performed, starting with a translation time of 45 min and reducing the time by 10 min for each subsequent round. The final round used a translation time of only 5 min. The total yield of RNA–protein fusion constructs following translation after each round was evaluated using SDS–PAGE analysis and reported to gradually increased with each successive round of selection [28].

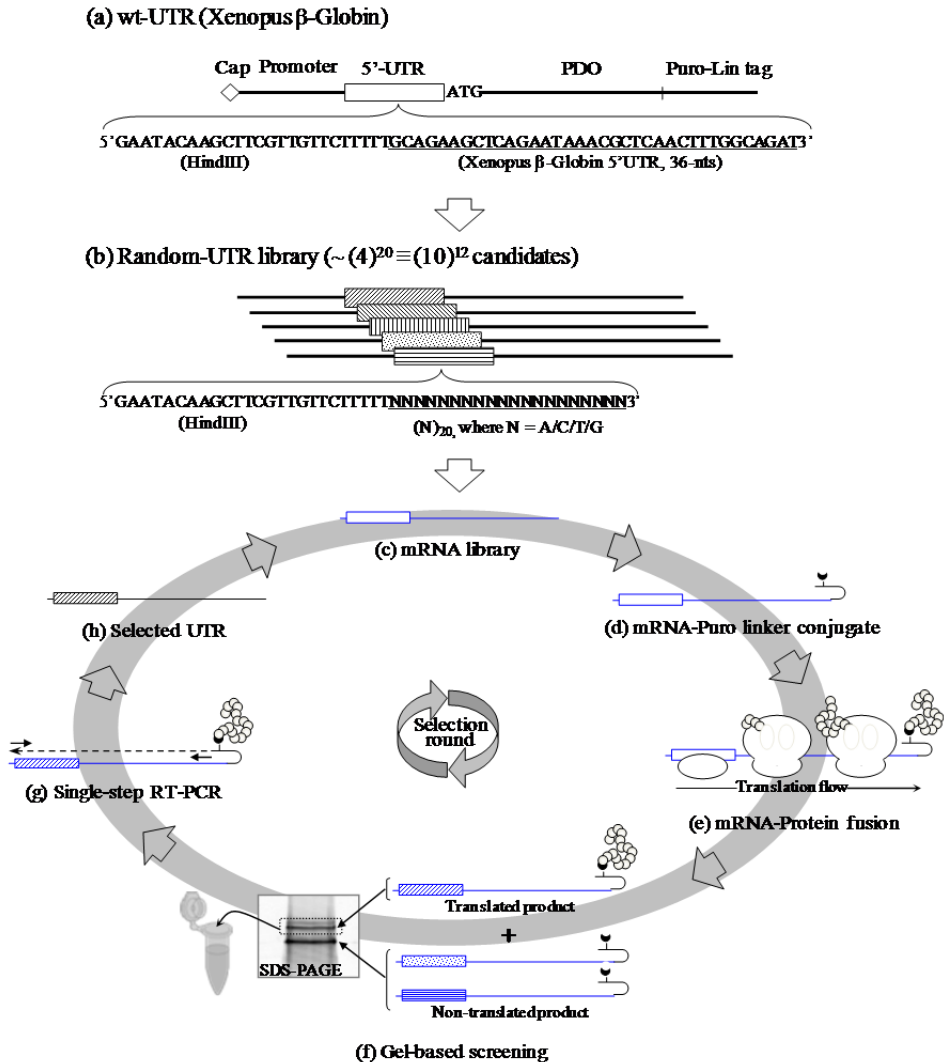


Figure 7. Schematic representation and flow diagram of a novel gel-shift selection method for searching a strong translation enhancer sequence against a given cell-free translation system using mRNA display. DNA template constructs used in the screening experiments including known 5'-untranslated region of *Xenopus*-Globin (wt-UTR) (a) and random-UTR library (b). The mRNA library which is lacking a stop codon (c) is ligated at the 3'-terminus end to the complementary portion of 5'-terminus end of the puromycin-linker DNA (d) and translated in a cell-free translation system (e). The ribosome stalls at the mRNA and linker-DNA junction during translation. This permits puromycin to enter the ribosomal A-site and to bind to the nascent polypeptide chain. Translated products are analyzed by SDS-PAGE analysis and carefully excised from gel to separate them from non-translated products (f). The associated mRNAs which represent the selected 5'UTR candidates were then directly reverse-transcribed and amplified using single-step RT-PCR (g) and used as template for next selection-round (h).

This increase confirmed that the selected library is successively enriched for strong translation enhancer sequences after each round of selection and thus the gel shift selection method using mRNA display is indeed a simple and effective method of screening for strong translation enhancer sequences. The analysis of selected sequences showed the richness of T and G bases with an average of 53% and 35%, respectively, indicating a significant role of U and G bases in the translation enhancer sequences. In addition, these selected sequences was confirmed to show higher translation efficiency in comparison with the natural and longer enhancer sequences. These results encouraged that the described gel-shift method could be applied to a rapid screening of novel 5'UTR which can facilitate cap-independent (IRES-mediated) protein synthesis in cell-free translation systems without the assistance of the full set of initiation factors. Very recently, a few interesting 5'UTRs have been proposed to accelerate the translation initiation reaction [29,30]. These findings of simple and effective 5'UTR suggest the possibility of improvement of 5'UTR under the conditions in various cell-free translation systems. Our approach can be applied to the further searching for 5'UTR by combining with these researches. In conclusion, gel-shift method demonstrated that shorter but strong translation enhancer sequences which should be easier to handle than long natural sequences can be selected rapidly by simple and robust mRNA display method. Searching for novel 5'UTR will contribute much toward the development of proteomics and evolutionary protein engineering research by improvements of cell-free translation methodologies.

5. Conclusion and future perspective

This chapter represents a simple, rapid, easy, and novel strategy, called 'Gel-shift selection', to obtain strong translation enhancer sequence variants for tunable protein synthesis using cell-free system. This method can further explore for (i) discovering of nuclease-resistant stable hairpin secondary structure to stabilize the 5'-terminus end of mRNA template with improved half-life instead of using synthetic 5'-cap analog; (ii) optimization of strong translational enhancer motifs which is free of 5'-cap dependency of translation initiation to improve the translational efficiency on given mRNAs under given translational conditions in cell-free system; (iii) optimization of enhancer motifs which is free of 3'-poly(A) dependency to eliminate the poly(A) leader effect which provide the abolition of the inhibition of translation at excess mRNA concentration.

Author details

Manish Biyani^{1,2,*}, Madhu Biyani^{1,3}, Naoto Nemoto³ and Yuzuru Husimi⁴

¹*Department of Biotechnology, Biyani Group of Colleges, R-4, Sector No 3, Jaipur, India*

²*Department of Bioengineering, The University of Tokyo, Tokyo, Japan*

³*Department of Functional Materials Science, Saitama University, Saitama, Japan*

⁴*Innovative Research Organization, Saitama University, Saitama, Japan*

* Corresponding Author

6. References

- [1] Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC., The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 2012, 40 (Database issue):D571-9.
- [2] Prasun, P., Pradhan, M., Agarwal, S., One gene, many phenotypes. *J. Postgrad. Med.*, 2007, 53, 257-261.
- [3] Yokoyama S., Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol.* 2003, 7, 39-43.
- [4] Winslow J, Insel T. The social deficits of the oxytocin knockout mouse. *Neuropeptides* 2002, 36, 221 – 229.
- [5] Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., Yokoyama, S., Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression, *J. Struct. Funct. Genomics* 2004, 5, 63–68.
- [6] Hartley A.D., Santos M.A., Colthurst D.R., Tuite M.F., Preparation and use of yeast cell-free translation lysate, *Methods Mol. Biol.* 1996, 53, 249–257.
- [7] Erickson A.H., Lobel G., Cell-free translation of messenger RNA in a wheat germ system, *Methods Enzymol.* 1983, 96, 38–50.
- [8] Jackson R.J., Hunt T., Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA, *Methods Enzymol.* 1983, 96, 50–74.
- [9] Mikami S., Kobayashi T., Yokoyama S., Imataka H., A hybridoma-based *in vitro* translation system that efficiently synthesizes glycoproteins, *J. Biotechnol.* 2006, 127, 65–78.
- [10] Ezure, T., Suzuki, T., Higashide, S., Shintani, E., Endo, K., Kobayashi, S., Shikata, M., Ito, M., Tanimizu, K., Nishimura, O., Cell-free protein synthesis system prepared from insect cells by freeze–thawing, *Biotechnol. Prog.* 2006, 22, 1570–1577.
- [11] S. Mikami, T. Kobayashi, M. Masutani, S. Yokoyama, H. Imataka, A human cell derived *in vitro* coupled transcription/translation system optimized for production of recombinant proteins, *Protein Exp. Purif.* 2008, 62, 190–198.
- [12] Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T., Cell-free translation reconstituted with purified components. *Nat Biotechnol.* 2001, 19, 751-755.
- [13] Shimizu Y, Kanamori T, Ueda T. Protein synthesis by pure translation systems. *Methods* 2005, 36, 299-304.
- [14] Falcone, D., Andrews, D.W., Both the 5'untranslated region and the sequences surrounding the start site contribute to efficient initiation of translation *in vitro*, *Mol. Cell. Biol.* 1991, 11, 2656–2664.
- [15] van der Velden AW, Thomas AA: The role of the 5'untranslated region of an mRNA in translation regulation during development. *Int J Biochem Cell Biol* 1999, 31, 87-106.
- [16] Mignone F, Gissi C, Liuni S, Pesole G: Untranslated regions of mRNAs. *Genome Biology* 2002, 3, 1-10.
- [17] Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S: Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 2001, 276, 73-81.

- [18] Grunberg-Manago, M., Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu. Rev. Genet.* 1999, 33, 193–227.
- [19] Salvador, M.L., Klein, U., Bogorad, L., 5' sequences are important positive and negative determinants of the longevity of *Chlamydomonas* chloroplast gene transcripts. *Proc. Natl. Acad. Sci. USA* 1993, 90, 1556–1560.
- [20] Suzuki, T., Ito, M., Ezure, T., Kobayashi, S., Shikata, M., Tanimizu, K., Nishimura, O., Performance of expression vector, pTD1, in insect cell-free translation system. *J Biosci Bioeng.* 2006, 102, 69-71.
- [21] Mureev, S., Kovtun, O., Nguyen, U.T. and Alexandrov, K. Species-independent translational leaders facilitate cell-free expression. *Nat Biotechnol.* 2009, 27, 747-752.
- [22] Nemoto, N. and Husimi, Y., A Model of the Virus-type Strategy in the Early Stage of Encoded Molecular Evolution., *J. Theor. Biol.* 1995, 176, 67-77.
- [23] Kamura, N. and Sawasaki, T. Selection of 5'UTR that enhance initiation of translation in a cell free protein synthesis system from wheat embryos. *Bioorg Med Chem Lett.* 2005, 15, 5402-5406.
- [24] Nagao, I. and Obokata, J. In vitro selection of translational regulatory elements. *Anal Biochem.* 2006, 354, 1-7.
- [25] Mie, M., Shimizu, S., Takahashi, F. and Kobatake, E. Selection of mRNA 5'-untranslated region sequence with high translation efficiency through ribosome display. *Biochem Biophys Res Commun.* 2008, 373, 48-52.
- [26] Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. In vitro virus: bonding of mRNA bearing puromycin at the 30-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* 1997, 414, 405–408.
- [27] Roberts, R. W., and Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl. Acad. Sci. USA* 1997, 94, 12297-12302.
- [28] Biyani, M., Biyani, M., Nemoto, N., Ichiki, T., Nishigaki K., Husimi, Y. Gel-shift selection of translation enhancer sequences using mRNA display. *Anal Biochem* 2011, 409, 105-11.
- [29] Shirokikh, N.E. and Spirin, A.S. Poly(A) leader of eukaryotic mRNA bypasses the dependence of translation on initiation factors. *Proc Natl Acad Sci U S A.* 2008, 105, 10738-10743.
- [30] Elfakess, R. and Dikstein, R., A translation initiation element specific to mRNAs with very short 5'UTR that also regulates transcription., *PLoS One* 2008, 3, e3094.