Gene Reports 9 (2017) 46–53

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems

Adnan B. Al-Hawash[a,b], Xiaoyu Zhang[a], Fuying Ma[a,*]

[a] Key Laboratory of Molecular Biophysics of MOE, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Ministry of Education, Directorate of Education, Basra, Iraq

A R T I C L E I N F O

Keywords:
Codon optimization
Synthetic biology
Gene engineering
Microbial systems

A B S T R A C T

Codon optimization includes strategies in gene design engineering, which uses synonymous codon changes to improve levels of protein production. Codon usage bias indicates that specific codons are used more often from other synonymous codons by gene translation. In the past decade, a variety of codon optimization methods have been used to design genes for optimal expression, which requires selection from a vast number of possible DNA sequences. Specific species differences in codon usage often indicate primary causes that affect levels of protein expression. The use of synthetically designed genes provides means for researchers to exercise more control on heterologous protein expression. Current studies show that codon optimization can affect protein conformation and function and increase expression levels. Over the past several years, considerable achievement in speed and cost of gene synthesis has facilitated complete redesign of entire gene sequences, significantly improving likelihood of high-level protein expression. This methodology significantly affects economic feasibility of microbial-based biotechnological processes for instance, by increasing volumetric productivities of recombinant proteins or facilitating the redesign of new biosynthetic routes for production of metabolites. Expression of proteins in heterologous hosts has become a cornerstone of modern biotechnology. This review discusses various codon optimization approaches that lead to high levels of protein expression and also a clarification to current applications of this technology.

1. Introduction

Microorganisms are crucial to the production of industrial enzymes, pharmaceuticals, and fine chemicals. Meeting commercial-level demands of target proteins and/or metabolites in numerous cases entails utilizing heterologous expression of genes. In the past decade, contribution of synthetic biology has markedly decreased cost of many products manufactured in microbial systems, where over-expression of just one gene is required. This process has driven replacement of native sequences, which are optimally produced in most cases, with the possible double-fold increase of target proteins. The dissimilarity in codon usage is one of the main factors that affect protein expression levels. In consequence, rare codons can decrease the rate of translation and also incite translation errors that significantly influence economy of recombinant by a bacterium production processes (Ikemura, 1981). This simple adjustment bears importance. Therefore, most products are considered traded commodities; consequently, a continuous persistence with reduced manufacturing cost is required to remain competitive in the global market (Menzella, 2011). The succeeding step to reach goals of synthetic biology is to further reduce the cost and time needed to develop recombinant organisms through the use of pre-assembled parts that provide stable, predictable protein expressions (Dellomonaco et al., 2010). The first human protein (somatostatin) was produced in a bacterium by Genentech scientists and their academic collaborators in 1977. Heterologous hosts for protein expression played crucial roles and exerted marked effects on the launch of biotechnology industry (Itakura et al., 1977). During this period, only the amino acid sequence of somatostatin was familiarly known; therefore, the Genentech group synthesized the 14-codon-long somatostatin gene by using oligonucleotides without cloning it from the human genome. Itakura et al. (1977) designed these oligonucleotides by relying on three standards. First, despite the limited knowledge on *Escherichia coli* genome DNA sequence at that period, codons favored by phage MS2 were used preferentially, and MS2 phage had been sequenced at the time to supposedly provide a better guide for codons being used in immensely expressed *E. coli* genes. Second, the researchers ensured elimination of

Abbreviations: TRNAs, transfer RNAs; mRNA, messenger RNA; miRNAs, microRNAs; GFP, green fluorescent protein; RBSs, ribosomal binding sites; bGH, bovine growth hormone; TnT, troponin T; hCG-β, human chorionic gonadotropin β gene; xynB, xylanase gene; CALB, *Candida antarctica* lipase B; LIP, lipase gene; GLA, gamma-linolenic acid

* Corresponding author.
E-mail address: mafuying@hust.edu.cn (F. Ma).

http://dx.doi.org/10.1016/j.genrep.2017.08.006
Received 7 February 2017; Received in revised form 24 July 2017; Accepted 24 August 2017
Available online 30 August 2017
2452-0144/ © 2017 Published by Elsevier Inc.

undesirable inter- and intra-molecular pairing of overlapping oligonucleotides, which affect gene synthesis. Third, GC-rich and AT-rich sequences were avoided because such sequences can terminate transcription. This finding considered the first functional polypeptide production in the synthetic gene (Ikemura, 1981). *E. coli* is highly important in synthetic biology because most progress has been achieved using this bacterium owing to its several advantages, including rapid growth, well-understood genetics, and low-cost fermentation media. Therefore, *E. coli* is the preferred host for heterologous protein production (Burgess-Brown et al., 2008; Welch et al., 2009; Menzella, 2011). To expand the applications of synthetic biology, combined efforts are currently focused in finding other hosts, including *Corynebacterium glutamicum* (Becker and Wittmann, 2012), *Streptomyces* species (Medema et al., 2011), yeast (Krivoruchko et al., 2011), and algae (Wang et al., 2012; Gimpel et al., 2013). The goal of this expanded search is to take advantage of the natural potential to synthesize precursors and cofactors required to produce certain targets. In this review, we summarize current codon optimization strategies for expression of heterologous proteins.

2. Codon optimization

Codon optimization is defined as gene design engineering, and it functions without altering amino acid sequences to improve recombinant genes based on different standards (Gaspar et al., 2012). Each protein sequence comprises only 20 standard amino acids; however, 64 different codons consisting of three stop codons exist, and another 61 codons encode the amino that code these 20 amino acids (Crick et al., 1961) (Table 1). A three-nucleotide codon in a nucleic acid sequence identifies a single amino acid. Only a few amino acids are encoded by a single codon. Most amino acids are encoded by two to six different codons, which complicate decoding by allowing more than one codon to encode single-amino-acid residues in proteins. This side of genetic code is referred to as “degeneracy” of codons. Synonymous codons are different codons that encode the same amino acid (Hershberg and Petrov, 2008; Sharp et al., 2010), as shown in Table 1. Numerous studies have demonstrated that the use of synonymous codons is a non-random process (Hershberg and Petrov, 2008; Sharp et al., 2010; Plotkin and Kudla, 2011). For example, alanine can be encoded by four codons, namely, GCC, GCG, GCU, and GCA, whereas phenylalanine can be encoded by two codons, namely, UUU and UUC, and leucine by six codons CUA, CUC, CUG, CUU, UUA, and UUG. Codon usage signifies the non-random use of codons in mRNAs. Codon usage in numerous organisms has been quantified by using various calculations, including the frequency of relative synonymous codon usage (Sharp et al., 1986), codon bias index (Bennetzen and Hall, 1982), use of optimal codons (Ikemura, 1981), effective number of codons (Wright, 1990), and codon adaptation index (Sharp and Li, 1987), which indicate differences in frequency of synonymous codons in coding DNA. For highly express genes in *E. coli* and yeast, the usage of non-random synonymous codons was found to be related with tRNA abundance. Most amino acids are encoded by more than one codon; thus, the majority of codon optimization programs aim to avoid the use of rare codons (Raab et al., 2010). Optimization of codons for the custom design of genes is a good approach for increasing efficiency of translation system and/or accuracy of protein synthesis. Numerous scientists have recently used the codon optimization technique of increasing efficiency of translation of the desired gene to improve protein expression (Gaspar et al., 2012). To maximize expression of proteins in living organisms, codon optimization is used to improve translational efficiency of target genes by converting DNA sequences of nucleotides of one species into DNA sequences of nucleotides of another species, for example, human sequence to bacterial or yeast sequences, plant sequence to human sequence, and fungal sequence to yeast sequence. Protein expression and codon bias are linked. Therefore, expression may be improved by mimicking patterns of codon bias of highly expressed mRNAs, thereby enhancing the possibility of developing numerous codon-optimization programs and services for commercial endeavors. These methods vary in how codon biases are measured, the number of variables considered potential applications, and execution. These programs should avoid the use of rare codons, which are believed to decrease the rate of translation elongation. An increasing number of programs are characterized to facilitate cloning, gene modification, and gene synthesis, whereas particular care regarding rare codons should be practiced to prevent features that may decrease protein expression. Most codon optimization strategies are not obliged by natural codon usage of the gene but absolutely require an amino acid sequence as input. Different strategies of codon optimization have led to production of codon-optimized mRNA sequences, which can vary markedly according to how they quantify codon usage and implement changes in codons.

Several strategies ideally use (frequently used) codons for all cases of amino acids or variations of this strategy (Richardson et al., 2006; Villalobos et al., 2006). Other strategies adjust codon usage to be commensurate with natural distribution of host organisms (Richardson et al., 2006). These strategies include codon coordination, which aims

Table 1
Codon table.

First base of codon	Second base of codon: U		C		A		G		Third base of codon
U	UUU	Phenylalanine phe	UCU	Serine ser	UAU	Tyrosine tyr	UGU	Cysteine cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine leu	UCA		UAA	STOP codon	UGA	STOP codon	A
	UUG		UCG		UAG		UGG	Tryptophan trp	G
C	CUU	Leucine leu	CCU	Proline pro	CAU	Histidine his	CGU	Arginine arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine ile	ACU	Threonine thr	AAU	Asparagine asn	AGU	Serine ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine lys	AGA	Arginine arg	A
	AUG	Methionine met (start codon)	ACG		AAG		AGG		G
G	GUU	Valine val	GCU	Alanine ala	GAU	Aspartic acid asp	GGU	Glycine gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid glu	GGA		A
	GUG		GCG		GAG		GGG		G

Download English Version:

https://daneshyari.com/en/article/5589960

Download Persian Version:

https://daneshyari.com/article/5589960

Daneshyari.com