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

ARTICLE INFO

Keywords:
Codon optimization
Synthetic biology
Gene engineering
Microbial systems

ABSTRACT

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

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

to identify and maintain slow translation regions that are considered important for protein folding (Villalobos et al., 2006; Angov et al., 2008). Alternative strategies involving the use of codons are believed to correspond to abundant tRNAs by using codons according to tRNA concentrations, selectively replacing rare codons, or avoiding the selection of known and slowly translating codon pairs (Gaspar et al., 2012; Qian et al., 2012).

3. Codon optimization analysis

To improve protein expression by codon optimization, strategies depend on three essential assumptions. First, rate of protein synthesis is limited by rare codons. Second, without affecting protein structure and function, synonymous codons can be used interchangeably. Third, optimization can be improved through increased production of protein by replacing rare codons with frequently used ones.

3.1. First assumption: rate of protein synthesis is limited by rare codons

In bacteria, different codons are translated at different rates. For example, an *E. coli* analysis of 29 codons denoted that aminoacyl–tRNA selection rates differ by up to 25-fold (Curran and Yarus, 1989). Overexpression of several recombinant proteins can drain one or more tRNAs and limit their expression (Kurland and Gallant, 1996). However, no additional evidence proves that rare codons limit production of proteins in mammalian cells. Several studies on bacteria showed absence of correlation of translation rates of specific codons with either tRNA abundance or frequency of codon use (Bonekamp et al., 1989). In a study by Rosano and Ceccarelli (2009) rare codons increased rate of translation corresponding to and increasing tRNAs expression but observed protein misfolding and aggregation. The extent to which different amino acids are encoded varies up to 10-fold in humans, but the overall trend indicates that more frequently encoded amino acids possess higher synonymous codons than those encoded less frequently. Differences are reduced by threefold when amino acid frequency is normalized to the synonymous codon number per amino acid. Individual codon usage frequencies also differ by up to tenfold. However, when a codon is rate-limiting, probability will depend on other variables, including tRNA levels. Codon usage tables that do not consider other variables that can affect codon frequencies, including codon usage by out-of-frame or substitute in-frame initiation events, may significantly skew codon usage. No significant evidence indicates that protein synthesis is limited by rare codons; this concept is independently supported by reports referring to translation initiation, not elongation, as rate-limiting factor for protein synthesis (Hershey et al., 2012). In other words, codons designated as “rare” may be classified incorrectly and may not be rate-limiting for translation.

3.2. Second: without affecting protein structure and function, used synonymous codons become interchangeable

Although the elongation rhythm determined for codon usage can slow down or stop ribosomes at certain sites, this state may be required in specific cases to correct protein folding (Komar, 2009). Although rare codons have been implicated in slowing translation and modeling stop sites, several studies revealed that codons do not require diminished local translation rates (Komar, 2009). One study showed that translation by cognate and wobble tRNAs occurs at various rates, with wobble coupling occurring more slowly by up to threefold in HeLa cells (Komar, 2009). Researchers assumed that wobble-dependent deceleration of elongation may select a mechanism for protein folding and is thus largely independent of tRNA levels. Decoding of efficiency variations may allow a mechanism to fine-tune elongation temporal patterns and is possibly important for protein conformation. Ribosomal pausing in *E. coli* involves base pairing of rRNA to Shine–Dalgarno sequences, such as sequences in coding regions (Li et al., 2012). In codon-optimized mRNA, synonymous codon substitutions may disrupt the information encoded in primary gene sequences. For instance, this disruption may influence translation by base pairing to rRNA or other RNAs when deleting complementary matches, including noncoding RNAs, such as microRNAs. These kinds of interactions can affect initiation, shunting, stop-over, frame shifting, reinitiation, and mRNA stability (Li et al., 2012). In addition to disrupting such interactions, codon optimization may inadvertently introduce new RNA binding sites. Finally, this view of protein function and synonymous codons, which are not necessarily interchangeable, is the most prominent. In one study, 342 antibodies were verified with synonymous codon variants and identified by altered expression, solubility, and antibodies with binding affinities (Sander et al., 2014). To understand the effects of synonymous codon changes, as recently implemented by Sander et al. (2014) a fluorescent protein gene was engineered to exhibit various fluorescent properties based on folded structures. Sander's group proved that fluorescent characteristics of proteins with synonymous codon changes were altered because of modified protein folding.

3.3. Third: protein production is increased by replacing rare codons with frequently used codons

Numerous examples support the assumption that major codons are regularly replaced with minor codons in the gene, showing diminished expression of phosphoglycerate kinase (Hoekema et al., 1987). When half of the codons replaced with codons used regularly through abundant proteins in yeast are increased, an immunoglobulin kappa protein is expressed (Kotula and Curtis, 1991). These types of studies present certain constraints; they do not consider multiple variables that may unintentionally be affected in expression or indicate whether other codon-optimized alternatives are tested. Without additional data, we cannot determine whether reported effects are attributed to altered codon bias or other mechanisms. Studies using synonymous variants of formaldehyde-activating enzyme proved that enzyme expression and cell fitness are not correlated with the use of either rare codons or frequent codons (Agashe et al., 2013). Other mechanisms are indicated elsewhere. For instance, an analysis of ribosomal footprint data showed that translation rate is not slower when ribosomes translate rare codons or clusters of rare codons but is affected by amino acid charge (Charneski and Hurst, 2013). Several factors affecting elongation rates comprise mRNA secondary structure and adaptation of codons to the tRNA pool (Chen et al., 2013). Plotkin's group conducted a more extensive experimental investigation of contributions of different variables to protein expression. In their report, a library of 154 green fluorescent protein (GFP) genes was synthesized. These genes vary randomly in codon usage, but all genes encode the similar GFP protein. Fluorescence and GFP levels varied 250-fold across the library when expressed in *E. coli*; however, no relation was observed between level of expression (fluorescence) and codon bias (which was evaluated using two measures) or between fluorescence and the number of rare codons pairs. The research discovered a relation between codon usage and cell fitness, which was low for cells expressing mRNAs with large numbers of rare codons. Codon usage in highly expressed mRNAs influences the numbers of free ribosomes and global translation because of the influence of fitness. Bacteria and yeast grow fast and exhibit strong codon bias in highly expressed genes; on the contrary, higher eukaryotes exhibit more instead of less codon bias (Subramanian, 2008). In highly expressed genes, codon bias is conversely related to species' generation time, varying by more than four orders of magnitude and with the lowest level bias occurring in mammals. We can hardly predict that codon usage in optimizing mammalian genes should improve protein expression because codon usage in mammals has not been studied as systematically as that in *E. coli*. Frequently used codons do not cause to high level of expression despite highly expressed genes (at least in microbes) genes evolving an optimal codon bias, thereby suggesting that codon bias per se does not yield high expression but requires other

features (Andersson and Kurland, 1990; Klumpp et al., 2012).

4. Study codon bias, and what is codon usage bias

Coding of genome sequences is the blueprint of gene products; it provides valuable information on gene function and organism evolution. Consequently, genome-wide probe codon bias patterns, their causes and results, and identification of selective forces that shape their evolution are important in genome biology research. Analysis of codon usage bias presents significant applications in biotechnological processes, whereas heterologous gene expression is a powerful strategy for producing recombinant products, such as antibiotics insulin and vaccines. Codon bias of host organisms expressing foreign gene will probably be a major limiting factor in the yield of required products (Kane, 1995). Thus, foreign gene codon optimization is needed based on style of codon usage of host organisms to avoid affecting translation of recombinant genes (Fletcher et al., 2007). Research has also proven that optimal pairs can be replaced with synonymous codons in viral genomes to develop attenuated viral vaccines (Coleman et al., 2008). In other areas of studies and applications in modern biology, numerous codons (synonymous codons) are translated into the same amino acid. These concepts indicate the importance of accurate and significant analysis of codon bias in target organisms because of the degeneracy of the genetic code. Codon bias is a phenomenon synonymous to the use of codons at different frequencies. The defining characteristic in all genomes is a codon bias, which is maintained by a balance between selection, genetic drift, and mutation (Rocha, 2004; Hershberg and Petrov, 2008). Codon biases differ markedly between organisms despite the relative universality of the genetic code and preservation of translation machinery across species. Therefore, the rarest or most frequent codons in a gene differ both between and within species (Hershberg and Petrov, 2008). In general, the speed at which ribosomes decode a codon is based on cellular concentration of tRNA that recognizes it (Dong et al., 1996). However, this assumption remains controversial (Li et al., 2012), specifically regarding the concept of the most abundant codon pairs featuring more abundant tRNAs and vice versa. Several studies proved that synonymous triplet difference in all species is driven by codon usage adaptation to tRNA abundances or vice versa.(Ikemura, 1981; Kanaya et al., 1999). According to this finding, close relationships exist between gene codon bias and expression levels of a gene in organisms as diverse as *E. coli*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Arabidopsis thaliana* (Ikemura, 1981; Duret, 2002; Ghaemmaghami et al., 2003; Goetz and Fuglsang, 2005). The search for a relationship between tRNA abundance and codon usage has been successful in certain organisms (Ikemura, 1981; Dong et al., 1996). When in use, particular codons can improve expression levels of genes by > 1000-fold (Zhou et al., 1999).

5. Effect of codon bias on protein expression

Codon usage has been identified as the primary and most critical factor in prokaryotic gene expression (Lithwick and Margalit, 2003) because preferred codons bear relations with abundance of cognate tRNAs available within cells. This relationship improves translational level and balances codon concentration with isoacceptor tRNA concentration (Ikemura, 1981). Codon usage and isoacceptor tRNA concentrations have possibly co-evolved, and selection pressure for this co-evolution is clearer for high-level expressed genes compared with genes expressed at low levels (Bulmer, 1987). The co-evolution of isoacceptor tRNAs together with codon frequencies has resulted in several cases of departure from the canonical genetic code (Massey et al., 2003), whereas studies of comparative genomics shed new light on continuing evolution of the genetic code (Santos et al., 2004). In different organisms, considering the existence of slight variations in codes is a massive barrier to heterologous expression levels. Several organisms, especially ciliates, which played a critical role in elucidation of telomere biology, possess tRNAs that read the canonical stop codons TAA and TAG as Glu, which disable heterologous expression of genes.

6. Codon frequencies with tRNA abundances

Protein synthesis in tRNA translates codons into amino acids; in this process, tRNA isoacceptors comprise multiple tRNAs that read codons for the same amino acid, which is found in all organisms. Several studies proved that synonymous triplet differences across species drive adaptation of codon usage to tRNA abundances (Ikemura, 1981; Kanaya et al., 1999; Zhou et al., 2009; Novoa et al., 2012). The present study on the correlation between codon usage and tRNA abundance succeeded only in certain organisms (Ikemura, 1981; Dong et al., 1996). However, the present works was unsuccessful in bacteria, eukaryotes, and numerous species (Kanaya et al., 1999; dos Reis et al., 2004), suggesting that in lower organisms, translation efficiency may not be the main factor influencing codon usage (Duret and Mouchiroud, 1999; dos Reis et al., 2004). However, a new report present two distinct modifications at the wobble position of certain anticodons, which are at the center of this clear lack of relevance (Novoa et al., 2012). In bacteria and eukaryotes, these amendments “extend” wobble pairing ability of anticodons and affect codon usage bias, eventually affecting codon usage and genomic tRNA compositions. Inclusion of this amendment corrects formerly reported discrepancies between tRNA abundance and codon usage across all extant major phylogenetic groups. These findings not only indicate that codon usage and tRNA abundance coevolved but also that diversification of the genetic code usage in development is at least partially driven by emergence of certain tRNA modifications (Novoa et al., 2012). At present, speeds of translation for all sense codons from *Saccharomyces cerevisiae* are determined in vivo (Qian et al., 2012) by using genome-wide ribosome data profiling. Surprisingly, similar speeds of translation among synonymous codons were identified, indicating that preferentially used codons in expressions of high-level proteins are not translated faster than non-preferred codons. However, a relationship between codon usage bias and cognate tRNA abundances was actually observed. These results propose that codon usage bias found in highly expressed genes is a product of natural selection for an overall cellular efficiency rather than a stronger choice product for translation efficiency in highly expressed genes.

7. Technology of synthetic DNA for codon optimization

Several distinct steps that depend on codon optimization strategies for protein expression must be considered. Two methodologies have been used in codon optimization. The first method, designated as “one amino acid–one codon,” encodes all occurrences of a given amino acid in an optimized sequence through the majorly abundant codons of hosts (Villalobos et al., 2006; Feng et al., 2010; Marlatt et al., 2010; Wang et al., 2010). This approach, which was the most well known in early days of gene synthesis technology, exhibits a main defect; with this design, a powerful transcribed mRNA from a gene will contain high percentage of a subset of codons, yielding an imbalance in the tRNA pool and consequently reducing growth due to tRNA depletion (Gong et al., 2006; Villalobos et al., 2006). The second approach is called codon randomization. This method utilizes tables of translation according to frequency distribution of codons in an entire genome or a subset of genes of highly expressed codons. These tables attach weight to every codon by randomly assigning a likelihood given by respective weights (Kodumal et al., 2004; Welch et al., 2009; Wang et al., 2010). This last approach was speedily adopted by the community of synthetic biology and was considered superior. In addition to improving yields of desired products, codon randomization provides numerous advantages. For instance, flexibility in codon selection facilitates gene design by preventing the following: a. recurring elements that may result in gene deletions; b. secondary mRNA structures (Welch et al., 2009), and; c. internal ribosomal binding, transcriptional terminator sites, or

polyadenylation signals. By facilitating elimination of undesirable restriction sites, codon randomization assists assembly of large constructs (Villalobos et al., 2006). Several methodology reports conducted in the past years described variations in this approach and offered data on effects of sequence variables (Welch et al., 2009; Allert et al., 2010). In addition to codon optimization, other parameters must be considered in gene design for effective translation; these parameters include GC content, existence of mRNA sequence motifs (Pertzev and Nicholson, 2006) with the local context of a given codon (Villalobos et al., 2006), and sequence of the region containing the first 10 codons (Goodman et al., 2013) of several web-based free basic or advanced software, such as DNA Works (Hoover and Lubkowski, 2002), Optimizer (Puigbo et al., 2007), GeMS (Jayaraj et al., 2005), Gene Designer (Villalobos et al., 2006), and Gene Designer Synthetic (Wu et al., 2006). At present, GenScript, DNA2.0, Genewiz, and GeneArt are considered the top synthetic DNA suppliers that provide sequence optimization services using proprietary algorithms without additional cost. In engineered microorganisms for technology, synthetic DNA is not restricted to gene redesign. Synthetic biology can provide a group of promoters and ribosomal binding sites (RBSs), which can offer various levels of gene expression to achieve this purpose (synthetic gene design) with low cost (Boyle and Silver, 2012; Meng et al., 2013; Vogl et al., 2013). To date, most promoters that have been used in their natural sequences lead to satisfactory gene expression. Both bacteriophage T7 and AOX promoters are used on a wide scale to supply high levels of transcription in *E. coli* (Studier and Moffatt, 1986) and *Pichia pastoris* (Tschopp et al., 1987). Currently, libraries of synthetic promoters for tunable gene expression are obtainable from numerous microorganisms, such as *E. coli* (Wu et al., 2013), *Bacillus subtilis* (Hansen et al., 2009), *Pichia pastoris* (Vogl et al., 2013), and *Corynebacterium glutamicum* (Yim et al., 2013), with industrially relevant properties. Synthetic RBSs can also be used to regulate gene expression (Basu et al., 2005; Pfleger et al., 2006). Recently, a novel method for automatic artificial RBS design to control gene expression has been described to expand the use of artificial sequence toolbox to custom genetic engineering circuits (Salis et al., 2009). However, frequent efforts for accurate forecast remained unsuccessful in terms of responses to any given promoter or RBS. Anonymous interactions among isolated components probably significantly affect optimal level of gene expression required to achieve a particular flux through a reaction or pathway, and promoters play a crucial role in controlling biosynthetic pathways (Keasling, 2012). Kosuri et al. (2013) synthesized 12,563 combinations of common promoters and RBSs and simultaneously measured DNA, RNA, and protein levels from the entire library. These researchers discovered that RNA and protein expression were within twofold of projected levels (80% to 64% of the time, respectively), and that the poorest 5% of constructs deviated from prediction by 13-fold; this outcome can notably hinder genetic engineering projects. Given that inducible promoters are one of the most effective approaches to induce gene expression (Keasling, 2012), synthetic plasmid construction and assembly of parts that were created by Shetty et al. (2011) gained popularity among the synthetic biology community despite availability of several other designs. Shetty's group suggested the BioBrick standard, in which all parts are flanked by a common group of restriction sites, allowing combination, joining, and rapid assembly of genetic parts to create functional units of gene expression. To date, most experiments for creating synthetic vectors are carried out using *E. coli*. Constante et al. (2011) also used the BioBrick principle to engineer eukaryotic hosts by building several model constructs that encompass concatemers of protein binding site motifs and a variety of plasmids for stable eukaryotic cloning and chromosomal insertion.

8. Practical results of codon optimization strategies

The use of protein expression by codon optimization constantly increases in various applications, including pharmaceuticals, industrial enzymes, biofuels, and other metabolites (Table 2). However, the main obstacle to commercial production of these molecules is the development of robust yield, productivities, and processes. George et al. (1985) optimized mature bovine growth hormone (bGH) and highly expressed *E. coli* genes to increase accumulation of hormones by 10- to 15-fold; this experiment resulted in the replacement of 13 of the first 23 codons of mature bGH. Williams et al. (1988) reported a synthetic gene encoding human interleukin-2 (IL-2), in which the synthetic IL-2 gene produced 16- fold more IL-2 than its native counterpart. As an antibody product, troponin T (TnT) has been replaced by two pairs of consecutive AGG codons successfully (AGG165 and AGG166; and AGG 215 and AGG 216) in the cDNA with the main synonymous codon CGT by site-directed mutagenesis; in this approach, substitution of one pair of rare arginine codons increased TnT expression by 10-fold, whereas a 40-fold incremental increase was achieved when both pairs of rare codons were replaced in *E. coli* (Hu et al., 1996). Genes were replaced with a frequently used codon (GAG replacing GAA) in the human-type 1 neurofibromin, and expression of soluble protein improved by threefold (Hale and Thompson, 1998). Zolotukhin et al. (1996) constructed gfph, a synthetic version of the jellyfish *Aequorea victoria*, from gfb cDNA. A gfp10 cDNA was designed, synthesized, and cloned as a humanized mutant appropriate for high-level expression in mammalian cells. The researchers observed increased sensitivity of this reporter gene system by 22-fold for gfph1 and a 45-fold gfph2 increase compared with original jellyfish gene. Vervoort et al. (2000) investigated the crucial role of codon usage and of DNA sequence upstream of the ATG start codon. The human chorionic gonadotropin β gene possesses three clusters with infrequently used codons that replaced the codon preferred in *Dictyostelium*, thereby optimizing the first 17 codons and inducing a four- to fivefold increase in expression. Yadava and Ockenhouse (2003) increased F2 (F2 domain of EBA-175) clone protein expression by four- to ninefold in *E. coli* and *Pichia pastoris*, respectively. Bradel-Tretheway et al. (2003) investigated the effect of codon optimization on protein expression of a herpes virus gene. Codon optimization of the U51 gene (human herpes viruses (HHV)-6 and HHV-7) leveled up expression by approximately 10- to 100-fold in humans. Production of biofuels, such as ethanol and biodiesel, is considered one of the most advantageous applications for synthetic biology. De novo construction of pathways in model host organisms, such as *E. coli*, provides a platform for reconstituting and studying behaviors of complex enzyme systems in vivo. Bond-Watts et al. (2011) demonstrated construction of a chimeric pathway assembled from three different organisms for high-level production of n-butanol (4650 ± 720 mg/L). Recently, more fuels and chemicals have been produced from fatty acids of plant and animal oils

Table 2
Collection of publications of gene expression by strategy of codon optimized.

Protein name	Host	Optimization	References
IL-2	*Escherichia coli*	16-fold	(Williams et al., 1988)
bGH	*E. coli*	10- to 15-fold	(George et al., 1985)
TnT	*E. coli*	10 and 40-fold	(Hu et al., 1996)
Neurofibromin	*E. coli*	3 fold	(Hale and Thompson, 1998)
$gfph_{1\ and\ 2}$	*Aequorea victoria*	22 and 45 fold	(Zolotukhin et al., 1996)
hCG-β	*Dictyostelium*	4–5 fold	(Vervoort et al., 2000)
F1, F2 domain of EBA	*E. coli and Pichia pastoris*	4 and 9 fold	(Yadava and Ockenhouse, 2003)
HHV-6 and -7	Mammalian	10–100-fold	(Bradel-Tretheway et al., 2003)
xynB	*E. coli*	2.8-fold	(Jia et al., 2012)
LIP2	*P. pastoris*	2.82 g/L	(Zhou et al., 2015)
endochitinase	*P. pastoris*	1.35-fold	(Yu et al., 2013)
Irisin	*P. pastoris*	77.98 mg/L	(Duan et al., 2015)
CALB	*P. pastoris*	0.8-fold	(Yang et al., 2013)
Delta-6 desaturase	Mammalian	2.2–4.8 fold	(Chen et al., 2011)

than those in the past decades. Steen et al. (2010) engineered *E. coli* to produce fatty acid ethyl esters (C12–C18) directly from glucose and ethanol, resulting in 400 and 700 mg/L production, respectively. The use of mevalonate pathway expression was further enhanced in *E. coli* synthase screening, and codon optimization of bisabolene led to a 10-fold improvement in bisabolene titers; meanwhile, *Saccharomyces cerevisiae* screening of bisabolene synthases yielded bisabolene titers > 900 mg/L (Peralta-Yahya et al., 2011). Paddon et al. (2013) used synthetic biology to develop strains of *Saccharomyces cerevisiae* for high-level biological production of artemisinic acid, a promisor of artemisinin, through fermentation. This approach resulted in a high yield of 25 g/L of artemisinic acid. Xylanase presents wide-scale applications in paper and pulp; energy and industries; and foodstuffs, including animal feed. Expressed in *Pichia pastoris*, a xylanase gene (xynB) from the hyperthermophilic *Thermotoga maritima* was optimized based on codon usage of the former. In analysis of DNA sequence of the native xynB, a graphical codon usage analyzer detected that several amino acid residues were encoded by codons that are rarely used in *Pichia pastoris*. These codons, CTC (Leu), TCG (Ser), AGC (Ser), and GCG (Ala), share < 10% of usage percentage. Low level of expression was due to rare codons in xynB in *Pichia pastoris*; rare codons were replaced by frequently used codons, and G + C content was adjusted to an appropriate range of 42.7% to 43.1%. Using codon optimization, Jia et al. (2012) showed a 2.8-fold increase in xynB gene expression. Yang et al. (2013) optimized the codon of *Candida antarctica* lipase B (CALB) gene through the use of a de novo design and synthesis strategy. After optimization, expression level of the codon-optimized gene was approximately 0.8-fold higher than that of the native CALB gene; codon optimization on Alpha-factor can also effectively improve expression level of CALB gene in *Pichia pastoris*. Heterologous gene expression is a significant tool for synthetic biology; it allows metabolic engineering and production of non-natural biologics in a variety of host organisms. By using codon optimization strategy, two heterologous genes demonstrated improved translational efficiency, and they included fluorescent protein-encoding eGFP and catechol 1,2-dioxygenase gene CatA, which are expressed in *Saccharomyces cerevisiae*. The strategy led to a 2.9-fold improvement over commercial gene optimization algorithms (Lanza et al., 2014). Gamma-linolenic acid (GLA, 18:3 n6) belongs to the omega-6 family of essential poly-unsaturated fatty acids and is extremely important in prevention and/or treatment of various health problems. Chen et al. (2011) increased GLA content in mammalian cells by using the delta-6 gene from *Borago officinalis*. This gene was codon-optimized and introduced into HEK293 cells by lipofectin transfection. The highest conversion efficiency of LA into GLA was 6.9- times higher than that of the control group, and the study showed that expression of delta-6 desaturase yielded a 2.2–4.8 fold increase in GLA when compared with that in the control cells. A simple design method for codon optimization of genes was applied to express heterologous proteins in mammalian cells; by using this method, sole selection of codons preferentially used in humans with > 60% GC content can enhance protein expression in other mammalian cells. This method, called the "preferred human codon-optimized method," was used in a study by Inouye et al. (2015) to express six proteins, including photoproteins of aequorin and clytin II, coelenterazine-utilizing luciferases of *Gaussia luciferase*, *Luciola cruciata*, firefly luciferases from *Photinus pyralis*, and *Renilla luciferase*. Both preferred human codon-optimized genes showed higher luminescence activity compared with wild-type genes. Lipases (EC 3.1.1.3) from *Yarrowia lipolytica* are significant and promising biocatalysts for industrial applications. The codon-optimized lipase gene demonstrated high lipase production, with the protein content in fermentation reaching high estimated level of 2.82 g/L (Zhou et al., 2015). Several products of commercial interest were recently produced from strains with codon optimization; these products include L-amino acids in *Corynebacterium glutamicum* and *E. coli* (Becker and Wittmann, 2012), and polyhydroxybutyrate and methyl halides in *Saccharomyces cerevisiae* (Kocharin et al., 2013). Yu et al. (2013) worked based on codon bias of *Pichia pastoris* GS115 optimized by cDNA of endochitinase from *Trichoderma viride* sp., showed high expression of *Trichoderma viride* sp. endochitinase in *Pichia pastoris* GS115 by using codon optimization. Endochitinase increased by 1.35-fold compared with the native strain. Irisin, a novel hormone, was found to be important in metabolic diseases, insulin resistance, and body weight regulation. Duan et al. (2015) used a codon-optimized irisin gene, which was designed according to Pichia pastoris GS115 synonymous codon usage bias, and showed that codon optimization resulted in the highest heterologous irisin production, reaching approximately 77.98 mg/L. Selection of expression vectors and transcriptional promoters are crucial in expression of heterologous proteins although codon bias in genes plays a critical role in their expression.

9. Conclusion

Codon optimization can induce high levels of gene expression; any change in the genetic code alters the meaning of a codon. Expression of heterologous genes offers a powerful methodology in biotechnological processes that produce recombinant products, such as insulin, antibiotics, and other necessary industrial items. Numerous suggestions primarily consider codon optimization for generating high-expression constructs, thereby supporting approaches that can improve protein expression. However, mRNAs evidently contain numerous layers of information that overlap with the amino acid code, creating a complexity that implies likelihood of disruption by codon optimization. Efforts on frequent codon usage by researchers over the past decade focused on decreasing costs of technologies by assembling large fragments of DNA and enabling production of multiple simultaneous changes to wild-type genomes, which are now becoming available. Accordingly, we can expect designing of custom-made microorganisms for specific applications in the future. At present, low synthesis costs, along with automation, will allow broad applications to genomics of synthetic products. At a wide-scale level, current strategies for codon optimization in industrial biotechnology have demonstrated that this approach is being rapidly adopted by strain developers to match competitions in current worldwide markets. More studies, a deeper understanding of underlying molecular principles of living systems, and further development of bioinformatics tools are needed to provide assistance in modeling behavior of synthetic genomes to fully harness the potential of synthetic biology. With increased exposure and now high-profile operations, this field of technology will continue to achieve tremendous progress in the following years.

Acknowledgements

Not applicable.

Authors' contributions

Al-Hawash conceived the structure and wrote this manuscript. Al-Hawash and Zhang improved the structure and language of the manuscript. Al-Hawash and Ma designed the tables. Ma and Zhang revised the manuscript. All authors read and approved the final manuscript.

Conflict of interest

Authors have no potential conflict of interest to disclose.

References

Agashe, D., Martinez-Gomez, N.C., Drummond, D.A., Marx, C.J., 2013. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol. Biol. Evol. 30, 549–560.

Allert, M., Cox, J.C., Hellinga, H.W., 2010. Multifactorial determinants of protein

expression in prokaryotic open reading frames. J. Mol. Biol. 402, 905–918.
Andersson, S., Kurland, C., 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. 54, 198–210.
Angov, E., Hillier, C.J., Kincaid, R.L., Lyon, J.A., 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One 3, e2189.
Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H., Weiss, R., 2005. A synthetic multicellular system for programmed pattern formation. Nature 434, 1130–1134.
Becker, J., Wittmann, C., 2012. Bio-based production of chemicals, materials and fuels–*Corynebacterium glutamicum* as versatile cell factory. Curr. Opin. Biotechnol. 23, 631–640.
Bennetzen, J.L., Hall, B., 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026–3031.
Bond-Watts, B.B., Bellerose, R.J., Chang, M.C., 2011. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat. Chem. Biol. 7, 222–227.
Bonekamp, F., Dalbøge, H., Christensen, T., Jensen, K.F., 1989. Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in *Escherichia coli*. J. Bacteriol. 171, 5812–5816.
Boyle, P.M., Silver, P.A., 2012. Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab. Eng. 14, 223–232.
Bradel-Tretheway, B.G., Zhen, Z., Dewhurst, S., 2003. Effects of codon-optimization on protein expression by the human herpesvirus 6 and 7 U51 open reading frame. J. Virol. Methods 111, 145–156.
Bulmer, M., 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325, 728–730.
Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., Gileadi, O., 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. Protein Expr. Purif. 59, 94–102.
Charneski, C.A., Hurst, L.D., 2013. Positively Charged Residues are the Major Determinants of Ribosomal Velocity.
Chen, Q., Nimal, J., Li, W., Liu, X., Cao, W., 2011. Delta-6 desaturase from borage converts linoleic acid to gamma-linolenic acid in HEK293 cells. Biochem. Biophys. Res. Commun. 410, 484–488.
Chen, C., Zhang, H., Broitman, S.L., Reiche, M., Farrell, I., Cooperman, B.S., Goldman, Y.E., 2013. Dynamics of translation by single ribosomes through mRNA secondary structures. Nat. Struct. Mol. Biol. 20, 582–588.
Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., Mueller, S., 2008. Virus attenuation by genome-scale changes in codon pair bias. Science 320, 1784–1787.
Constante, M., Grunberg, R., Isalan, M., 2011. A biobrick library for cloning custom eukaryotic plasmids. PLoS One 6, e23685.
Crick, F., Barnett, L., Brenner, S., Watts-Tobin, R.J., 1961. General nature of the genetic code for proteins. Nature 192, 1227–1232.
Curran, J.F., Yarus, M., 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J. Mol. Biol. 209, 65–77.
Dellomonaco, C., Fava, F., Gonzalez, R., 2010. The path to next generation biofuels: successes and challenges in the era of synthetic biology. Microb. Cell Factories 9, 1–15.
Dong, H., Nilsson, L., Kurland, C.G., 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. 260, 649–663.
dos Reis, M., Savva, R., Wernisch, L., 2004. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 32, 5036–5044.
Duan, H., Wang, H., Ma, B., Jiang, P., Tu, P., Ni, Z., Li, X., Li, M., Ma, X., Wang, B., 2015. Codon optimization and expression of irisin in *Pichia pastoris* GS115. Int. J. Biol. Macromol. 79, 21–26.
Duret, L., 2002. Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. 12, 640–649.
Duret, L., Mouchiroud, D., 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. 96, 4482–4487.
Feng, Z., Zhang, L., Han, X., Zhang, Y., 2010. Codon optimization of the calf prochymosin gene and its expression in *Kluyveromyces lactis*. World J. Microbiol. Biotechnol. 26, 895–901.
Fletcher, S.P., Muto, M., Mayfield, S.P., 2007. Optimization of recombinant protein expression in the chloroplasts of green algae. Adv. Exp. Med. Biol. 616, 90–98.
Gaspar, P., Oliveira, J.L., Frommlet, J., Santos, M.A., Moura, G., 2012. EuGene: maximizing synthetic gene design for heterologous expression. Bioinformatics 28, 2683–2684.
George, H.J., L'ITALIEN, J.J., PILACINSKI, W.P., GLASSMAN, D.L., KRZYZEK, R.A., 1985. High-level expression in *Escherichia coli* of biologically active bovine growth hormone. DNA 4, 273–281.
Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.
Gimpel, J.A., Specht, E.A., Georgianna, D.R., Mayfield, S.P., 2013. Advances in microalgae engineering and synthetic biology applications for biofuel production. Curr. Opin. Chem. Biol. 17, 489–495.
Goetz, R.M., Fuglsang, A., 2005. Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. Biochem. Biophys. Res. Commun. 327, 4–7.
Gong, M., Gong, F., Yanofsky, C., 2006. Overexpression of tnaC of *Escherichia coli* inhibits growth by depleting tRNA2Pro availability. J. Bacteriol. 188, 1892–1898.
Goodman, D.B., Church, G.M., Kosuri, S., 2013. Causes and effects of N-terminal codon bias in bacterial genes. Science 342, 475–479.
Hale, R.S., Thompson, G., 1998. Codon optimization of the gene encoding a domain from human type 1 neurofibromin protein results in a threefold improvement in expression level in *Escherichia coli*. Protein Expr. Purif. 12, 185–188.
Hansen, M.E., Wangari, R., Hansen, E.B., Mijakovic, I., Jensen, P.R., 2009. Engineering of *Bacillus subtilis* 168 for increased nisin resistance. Appl. Environ. Microbiol. 75, 6688–6695.
Hershberg, R., Petrov, D.A., 2008. Selection on codon bias. Annu. Rev. Genet. 42, 287–299.
Hershey, J.W., Sonenberg, N., Mathews, M.B., 2012. Principles of translational control: an overview. Cold Spring Harb. Perspect. Biol. 4, a011528.
Hoekema, A., Kastelein, R., Vasser, M., De Boer, H., 1987. Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol. 7, 2914–2924.
Hoover, D.M., Lubkowski, J., 2002. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 30, 1–7 e43.
Hu, X., Shi, Q., Yang, T., Jackowski, G., 1996. Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in *Escherichia coli*. Protein Expr. Purif. 7, 289–293.
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.
Inouye, S., Sahara-Miura, Y., Sato, J.-I., Suzuki, T., 2015. Codon optimization of genes for efficient protein expression in mammalian cells by selection of only preferred human codons. Protein Expr. Purif. 109, 47–54.
Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., Boyer, H.W., 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198, 1056–1063.
Jayaraj, S., Reid, R., Santi, D.V., 2005. GeMS: an advanced software package for designing synthetic genes. Nucleic Acids Res. 33, 3011–3016.
Jia, H., Fan, G., Yan, Q., Liu, Y., Yan, Y., Jiang, Z., 2012. High-level expression of a hyperthermostable *Thermotoga maritima* xylanase in *Pichia pastoris* by codon optimization. J. Mol. Catal. B Enzym. 78, 72–77.
Kanaya, S., Yamada, Y., Kudo, Y., Ikemura, T., 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238, 143–155.
Kane, J.F., 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol. 6, 494–500.
Keasling, J.D., 2012. Synthetic biology and the development of tools for metabolic engineering. Metab. Eng. 14, 189–195.
Klumpp, S., Dong, J., Hwa, T., 2012. On ribosome load, codon bias and protein abundance. PLoS One 7, e48542.
Kocharin, K., Siewers, V., Nielsen, J., 2013. Improved polyhydroxybutyrate production by *Saccharomyces cerevisiae* through the use of the phosphoketolase pathway. Biotechnol. Bioeng. 110, 2216–2224.
Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M., Santi, D.V., 2004. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc. Natl. Acad. Sci. U. S. A. 101, 15573–15578.
Komar, A.A., 2009. A pause for thought along the co-translational folding pathway. Trends Biochem. Sci. 34, 16–24.
Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., Church, G.M., 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 110, 14024–14029.
Kotula, L., Curtis, P.J., 1991. Evaluation of foreign gene codon optimization in yeast: expression of a mouse IG kappa chain. Nat. Biotechnol. 9, 1386–1389.
Krivoruchko, A., Siewers, V., Nielsen, J., 2011. Opportunities for yeast metabolic engineering: lessons from synthetic biology. Biotechnol. J. 6, 262–276.
Kurland, C., Gallant, J., 1996. Errors of heterologous protein expression. Curr. Opin. Biotechnol. 7, 489–493.
Lanza, A.M., Curran, K.A., Rey, L.G., Alper, H.S., 2014. A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst. Biol. 8, 33.
Li, G.-W., Oh, E., Weissman, J.S., 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484, 538–541.
Lithwick, G., Margalit, H., 2003. Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Res. 13, 2665–2673.
Marlatt, N.M., Spratt, D.E., Shaw, G.S., 2010. Codon optimization for enhanced *Escherichia coli* expression of human S100A11 and S100A1 proteins. Protein Expr. Purif. 73, 58–64.
Massey, S.E., Moura, G., Beltrão, P., Almeida, R., Garey, J.R., Tuite, M.F., Santos, M.A., 2003. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. Genome Res. 13, 544–557.
Medema, M.H., Breitling, R., Takano, E., 2011. Synthetic biology in *Streptomyces* bacteria. Methods Enzymol. 497, 485–502.
Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G., Wang, Y., 2013. Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. PLoS One 8, e60288.
Menzella, H.G., 2011. Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli*. Microb. Cell Factories 10, 10.1186.
Novoa, E.M., Pavon-Eternod, M., Pan, T., de Pouplana, L.R., 2012. A role for tRNA modifications in genome structure and codon usage. Cell 149, 202–213.
Paddon, C.J., Westfall, P., Pitera, D., Benjamin, K., Fisher, K., McPhee, D., Leavell, M., Tai, A., Main, A., Eng, D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–532.
Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., Lee, T.S., 2011. Identification and microbial production of a terpene-based advanced biofuel. Nat. Commun. 2, 483.
Pertzev, A.V., Nicholson, A.W., 2006. Characterization of RNA sequence determinants

and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. Nucleic Acids Res. 34, 3708–3721.
Pfleger, B.F., Pitera, D.J., Smolke, C.D., Keasling, J.D., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat. Biotechnol. 24, 1027–1032.
Plotkin, J.B., Kudla, G., 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42.
Puigbo, P., Guzman, E., Romeu, A., Garcia-Vallve, S., 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 35, W126–W131.
Qian, W., Yang, J.-R., Pearson, N.M., Maclean, C., Zhang, J., 2012. Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. 8, 1–18 e1002603.
Raab, D., Graf, M., Notka, F., Schödl, T., Wagner, R., 2010. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Syst. Synth. Biol. 4, 215–225.
Richardson, S.M., Wheelan, S.J., Yarrington, R.M., Boeke, J.D., 2006. GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome Res. 16, 550–556.
Rocha, E.P., 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14, 2279–2286.
Rosano, G.L., Ceccarelli, E.A., 2009. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. Microb. Cell Factories 8, 41.
Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27, 946–950.
Sander, I.M., Chaney, J.L., Clark, P.L., 2014. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. J. Am. Chem. Soc. 136, 858–861.
Santos, M.A., Moura, G., Massey, S.E., Tuite, M.F., 2004. Driving change: the evolution of alternative genetic codes. Trends Genet. 20, 95–102.
Sharp, P.M., Li, W.-H., 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15, 1281–1295.
Sharp, P.M., Tuohy, T.M., Mosurski, K.R., 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14, 5125–5143.
Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that influence the evolution of codon bias. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 365, 1203–1212.
Shetty, R., Lizarazo, M., Rettberg, R., Knight, T.F., 2011. Assembly of BioBrick standard biological parts using three antibiotic assembly. Methods Enzymol. 498, 311–326.
Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B., Keasling, J.D., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463, 559–562.
Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113–130.
Subramanian, S., 2008. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. Genetics 178, 2429–2432.
Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., Gingeras, T.R., 1987. Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Res. 15, 3859–3876.
Vervoort, E.B., van Ravestein, A., van Peij, N.N., Heikoop, J.C., van Haastert, P.J., Verheijden, G.F., Linskens, M.H., 2000. Optimizing heterologous expression in *Dictyostelium*: importance of 5′ codon adaptation. Nucleic Acids Res. 28, 2069–2074.
Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., Govindarajan, S., 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinf. 7, 285.
Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., Glieder, A., 2013. Synthetic core promoters for *Pichia pastoris*. ACS Synth. Biol. 3, 188–191.
Wang, X., Li, X., Zhang, Z., Shen, X., Zhong, F., 2010. Codon optimization enhances secretory expression of *Pseudomonas aeruginosa* exotoxin A in *E. coli*. Protein Expr. Purif. 72, 101–106.
Wang, B., Wang, J., Zhang, W., Meldrum, D.R., 2012. Application of synthetic biology in cyanobacteria and algae. Front. Microbiol. 3, 1–15.
Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., Gustafsson, C., 2009. Design parameters to control synthetic gene expression in *Escherichia coli*. PLoS One 4, 1–10 e7002.
Williams, D., Regier, D., Akiyoshi, D., Genbauffe, F., Murphy, J., 1988. Design, synthesis and expression of a human interteukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes. Nucleic Acids Res. 16, 10453–10468.
Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.
Wu, G., Bashir-Bello, N., Freeland, S.J., 2006. The synthetic gene designer: a flexible web platform to explore sequence manipulation for heterologous expression. Protein Expr. Purif. 47, 441–445.
Wu, Y., Zhang, Y., Tu, R., Liu, H., Wang, Q., 2013. Construction of synthetic promoters for *Escherichia coli* and application in the biosynthesis of cis, cis-muconic acid. Chin. J. Biotechnol. 29, 760–771.
Yadava, A., Ockenhouse, C.F., 2003. Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. Infect. Immun. 71, 4961–4969.
Yang, J.-K., Liu, L.-Y., Dai, J.-H., Li, Q., 2013. De novo design and synthesis of *Candida antarctica* lipase B gene and α-factor leads to high-level expression in *Pichia pastoris*. PLoS One 8, e53939.
Yim, S.S., An, S.J., Kang, M., Lee, J., Jeong, K.J., 2013. Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. Biotechnol. Bioeng. 110, 2959–2969.
Yu, P., Yan, Y., Gu, Q., Wang, X., 2013. Codon optimisation improves the expression of *Trichoderma viride* sp. endochitinase in *Pichia pastoris*. Sci Rep 3, 1–6.
Zhou, J., Liu, W.J., Peng, S.W., Sun, X.Y., Frazer, I., 1999. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. J. Virol. 73, 4972–4982.
Zhou, T., Weems, M., Wilke, C.O., 2009. Translationally optimal codons associate with structurally sensitive sites in proteins. Mol. Biol. Evol. 26, 1571–1580.
Zhou, W.-J., Yang, J.-K., Mao, L., Miao, L.-H., 2015. Codon optimization, promoter and expression system selection that achieved high-level production of *Yarrowia lipolytica* lipase in *Pichia pastoris*. Enzym. Microb. Technol. 71, 66–72.
Zolotukhin, S., Potter, M., Hauswirth, W.W., Guy, J., Muzyczka, N., 1996. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J. Virol. 70, 4646–4654.