

Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations

Vincent P. Mauro¹ 

Published online: 1 February 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract Biotherapeutics are increasingly becoming the mainstay in the treatment of a variety of human conditions, particularly in oncology and hematology. The production of therapeutic antibodies, cytokines, and fusion proteins have markedly accelerated these fields over the past decade and are probably the major contributor to improved patient outcomes. Today, most protein therapeutics are expressed as recombinant proteins in mammalian cell lines. An expression technology commonly used to increase protein levels involves codon optimization. This approach is possible because degeneracy of the genetic code enables most amino acids to be encoded by more than one synonymous codon and because codon usage can have a pronounced influence on levels of protein expression. Indeed, codon optimization has been reported to increase protein expression by > 1000-fold. The primary tactic of codon optimization is to increase the rate of translation elongation by overcoming limitations associated with species-specific differences in codon usage and transfer RNA (tRNA) abundance. However, in mammalian cells, assumptions underlying codon optimization appear to be poorly supported or unfounded. Moreover, because not all synonymous codon mutations are neutral, codon optimization can lead to alterations in protein conformation and function. This review discusses codon optimization for therapeutic protein production in mammalian cells.

Key Points

Codon optimization is a method that is commonly used to increase the expression of biotherapeutic recombinant proteins through the use of synonymous codon mutations in messenger RNA (mRNA) coding regions.

A key assumption underlying codon optimization is that protein synthesis is restricted by rare codons; this assumption appears to be poorly supported in mammalian cells, which are frequently used to express recombinant proteins.

An unintended consequence of codon optimization is that it disrupts different types of information that overlap coding regions, which can affect local rates of translation elongation, lead to alterations in protein conformation, and increase immunogenicity.

1 Introduction

Various proteins, including hormones, monoclonal antibodies, enzymes, and blood factors, have great utility as drugs. In some cases, it has been possible to use material purified from natural sources for protein replacement therapy. For instance, diabetes mellitus has been treated using the peptide hormone insulin purified from cow and pig [1]; similarly, hemophilia A has been treated with clotting factor VIII purified from human blood plasma [2]. However, a major limitation with using proteins purified

✉ Vincent P. Mauro
vmauro@scripps.edu; vincent.p.mauro@gmail.com

¹ The Scripps Research Institute, La Jolla, CA 92037, USA

from animal or human sources is that many proteins with therapeutic potential are expressed at such low levels that it is not realistic to purify them. In these cases, protein therapy can become feasible when recombinant proteins are over-expressed in genetically engineered cells on an industrial scale [3]. Although a variety of cell types, including bacterial, yeast, insect, and mammalian, have proven useful for recombinant protein expression, most approved protein drugs are produced in mammalian cell lines and the most commonly used cell lines are derived from Chinese hamster ovary (CHO) [4, 5]. These cells have numerous features suitable for the production of therapeutic proteins, including the ability to grow in suspension, the ability to grow in chemically defined serum-free media, and the ability to be cultured on a large scale (> 10,000 L) [6]. In addition, CHO cells provide post-translational processes similar to those in human cells, which include co-translational folding, chaperone binding, and glycosylation. Moreover, CHO and other mammalian cell lines are less likely to yield undesired post-translational modifications that can lead to a protein being recognized as foreign by the patient.

1.1 Overview of Recombinant Protein Expression

In general, the process of recombinant protein expression in mammalian cells involves cloning a suitable complementary DNA (cDNA) sequence into an expression vector, such as a DNA plasmid, and then introducing the construct into a host cell line, which can be achieved by different methods including transfection, nucleofection, the use of virus vectors, as well as other methods [7–10]. Plasmids that enter the nucleus of transfected cells are transcribed, and messenger RNA (mRNA) encoding the target protein is translated. This process of transient gene expression is used to generate recombinant protein for several days and can produce milligram to gram amounts of recombinant protein, which is useful for academic studies and preclinical work. However, transient expression is not efficient for generating larger amounts of protein, which are required for clinical studies and commercialization.

The ability to express recombinant proteins on an industrial scale is possible because of tremendous advances in cell culture over the past century, including the development of antibiotics, sterile techniques, and chemically defined culture media [11–14]. Large-scale production is facilitated by generating stable transfected cell lines, which involves integrating an expression construct into the chromosomal DNA of the host cell line. Integration can occur at random chromosomal sites or in a site-directed manner which targets one or more chromosomal locations that may have been preselected for their abilities to facilitate high levels of expression and stability [15–19]. For

generating stable transfected pools and clonal cell lines, a marker gene on the expression construct is typically used to enable screening or selection of cells [20, 21]. For example, the enhanced green fluorescent protein gene can be used as a screening marker to identify cells that express this protein and separate them from non-expressing cells by using fluorescence activated cell sorting (FACS). Stable transfected cells can also be generated by using a selectable marker gene and a method to kill cells that do not express this gene. Selectable markers include antibiotic-resistance genes, the neomycin-resistance gene, the dihydrofolate reductase (*DHFR*) gene, and the glutamine synthetase (*GS*) gene. This process can be illustrated using the *GS* marker gene: cells are transfected with an expression construct containing the *GS* gene; following transfection, cells are cultured under conditions that enable stably transfected cells, which express *GS*, to grow but kill cells that do not express this protein. Selection involves growing cells without glutamine and in the presence of methionine sulfoximine which inhibits endogenous *GS* activity, or by using an auxotrophic cell line that lacks *GS* activity.

Following selection or screening, clonal cell lines can be obtained by limited dilution or other cloning method, including FACS or growth in semi-solid matrix. At this stage, individual stably transfected cells express recombinant protein at dramatically different levels. Expression is affected by numerous variables, including the chromosomal insertion site or sites, and the number of integrated plasmids. Even after high expressing clones are identified, their expression levels can still be affected by various factors, including genetic instability of the insertion site and methylation-induced transcriptional silencing [22]. Considerable heterogeneity can also occur within clonal cell lines [23]. In addition, endogenous genes are sometimes disrupted by insertion of the expression construct, which can have unanticipated effects. For these reasons, screening for cell lines generated by random integration is much more extensive than for targeted integration and can involve testing thousands of cell lines. For secreted proteins, a particularly powerful method for identifying high expressing clones involves culturing cells in a methylcellulose semi-solid matrix containing fluorescently labeled antibodies that recognize the recombinant protein [24, 25]. Single cells form colonies in the matrix and fluorescent halos develop around the colonies. The size and intensity of the halos correlates with the amount of secreted recombinant protein. High expressing colonies are picked using an automated picker, e.g., ClonePix (Molecular Devices, Sunnyvale, CA, USA), based on halo size/intensity and other parameters, including the size and shape of the colony, as well as its vicinity to other colonies.

Once clonal cell lines with suitable expression, stability, and growth properties are identified, expression can be optimized for maximal production by adjusting culture conditions (see Wurm [26]). The amount of protein produced in stable cell lines can vary dramatically for different proteins, but yields of 1–10 g/L can typically be reached in CHO-based fed-batch cultures [27].

1.2 Use of Recombinant Proteins as Therapeutic Drugs

Over the years, numerous therapeutic proteins have been approved by the US Food and Drug Administration (FDA) [28]. Tissue plasminogen activator (tPA) was the first recombinant protein produced in mammalian cells (CHO) that was approved for clinical use [29]. tPA illustrates the advantage of using genetically engineered cells to over-express a protein of interest as this protein can be expressed at high concentrations as a recombinant protein (50 pg/cell/day) but is only secreted naturally by mammalian cells at a low concentration [30].

A review of recent approvals of therapeutic recombinant proteins by the FDA for the period between 1 January 2011 and 31 August 2016 identified 62 proteins [5]. The majority are monoclonal antibodies (48%), which includes antibody–drug conjugates as well as antibody Fab fragments. Other major categories of proteins are coagulation factors (19%) and replacement enzymes (11%). The remaining therapeutics (22%) are fusion proteins, hormones, growth factors, and plasma proteins. The primary therapeutic indications for these approved proteins are in oncology (26%) and hematology (29%). Other indications are in cardiology/vascular disease, dermatology, endocrinology, gastroenterology, genetic disease, immunology, infectious diseases, musculoskeletal, nephrology, ophthalmology, pulmonary/respiratory disease, and rheumatology. Of these approved proteins, 50% were granted orphan designation.

More recently, there have been another 27 approvals by the FDA: 24 at the Center for Drug Evaluation and Research (CDER) between 1 September 2016 and 31 December 2017 and three at the Center for Biologic Evaluation and Research (CBER) between 1 September 2016 and 12 December 2017. These approvals included five biosimilars. Compared to the previous set of approvals discussed in Lagasse et al. [5], the percentage of approvals for new monoclonal antibodies was much higher (78 vs. 48%) and included one bispecific antibody and two antibody–drug conjugates. There were also two enzyme replacements, two vaccine antigens, and one Fc-fusion protein. In addition, a recombinant enzyme was approved in combination with a previously approved monoclonal antibody (mAb). The primary therapeutic indications for these proteins are in oncology (30%) and rheumatology

(19%). Other indications are in dermatology, infectious diseases, hematology, genetic disease, immunology, musculoskeletal, and pulmonary/respiratory disease.

1.3 Expression Challenges Associated with Recombinant Proteins

As discussed in Sect. 1.1, the process of recombinant protein expression involves numerous steps that can affect expression, protein quality, and cell physiology. For many recombinant proteins, expression levels can determine commercial viability and often present a bottleneck for further development. Fortunately, there are numerous variables that can be considered for enhancing productivity (e.g., see Ayyar et al. [31]). In some cases, protein expression can be improved by using an alternative promoter to drive transcription of the recombinant mRNA, as different natural and synthetic promoters vary in strength and stability [32–34]. In addition, improvements in expression and stability can be realized by minimizing negative effects associated with some chromosomal sites, e.g., by including a chromosomal insulator sequence on the expression plasmid [35]. Recombinant mRNA levels can also be increased by purposefully generating cell lines with multiple gene copies, for example by using an expression construct containing the *DHFR* gene as a selectable marker [36]. For this approach, the DHFR construct is introduced into CHO cells that are deficient for DHFR and stable transfected cells are selected by using increasing concentrations of methotrexate, a drug that inhibits DHFR activity. Cell lines with multiple gene copies can also be generated by using site-directed integration approaches [18]. It is anticipated that the development of new integration strategies will provide even greater control of expression levels.

Increasing recombinant mRNA levels can be useful up to a point, beyond which there is no obvious further benefit, or even a negative effect [26, 37]. However, it should be recognized that high levels of mRNA are not necessarily linked to high levels of protein [38, 39]. For example, cells selected using the DHFR selection method can contain up to thousands of genomic copies of the expression construct, but protein levels are maximally increased by only 10- to 20-fold (e.g., Wurm [26]). Problems associated with large numbers of genomic copies of an expression construct include reduced stability of the trans-genes and other effects, including position effects and disruption of endogenous genes [26, 40, 41]. In addition, it is likely that high levels of recombinant mRNAs limit protein production by non-specific effects, e.g., by titrating transcription factors or RNA binding factors. Indeed, our own studies have shown that protein expression from an mRNA optimized for translation efficiency can be dramatically higher

when transcription is driven by a weaker promoter than a stronger promoter (Mauro and Chappell, unpublished observations). In addition, some negative effects associated with overexpression are related to the biological activity or toxicity of the recombinant protein which may affect cell physiology.

Other features of recombinant genes can be modified to increase expression levels. For instance, protein production is often increased by including one or more introns in the recombinant gene [42]. In addition, the ability of an mRNA to compete with other mRNAs for the translation machinery and its efficiency of translation initiation can be enhanced by modifying the 5' leader sequence, or by replacing it completely [43, 44]. One approach involves inserting natural or synthetic translation enhancing elements into the 5' leader. Alternatively, initiation can be enhanced by completely replacing the 5' leader with the 5' leader of an efficiently translated mRNA, such as β -globin, or with a synthetic sequence optimized for ribosome recruitment and initiation [43]. Modification of 3' untranslated region sequences can also yield increased expression by enhancing ribosome recruitment and mRNA stability [45].

Some proteins are inherently difficult to express because of features in the coding regions of the genes. This situation is not unexpected as some proteins, such as enzymes and hormones, are typically required at very low levels, can be harmful at higher levels, and are necessarily expressed poorly in the body. For example, the blood clotting factor VIII is required at low levels in the body and increased levels of this protein are associated with increased risk of thrombosis and stroke [46]. In cultured cells, this protein is notoriously difficult to express, and in the body, the factor VIII gene has evolved numerous features which limit its expression [47]. Unfortunately, some of these same evolved features likely make it difficult to overexpress the recombinant protein in cultured cells.

2 Codon Optimization

Codon optimization refers to approaches used for maximizing protein expression by overcoming expression limitations associated with codon usage. It is routinely used for applications in bioproduction as well as for in vivo nucleic acid therapeutic applications [31, 48]. Codon optimization has been reported to increase protein expression by up to > 1000-fold [49], although most reports are much more modest. Interestingly, synonymous codon mutations have also been used to de-optimize expression in order to fine-tune the expression of one of two light chain genes of a bispecific antibody, which resulted in increased the expression of this antibody [92]. An overview of the

process of mRNA translation is included below to provide appropriate background and context for this approach.

2.1 Messenger RNA Translation

Translation is the process whereby an mRNA template is decoded into a polypeptide sequence. This process consists of three steps: initiation, elongation, and termination [50]. Initiation involves recruitment of the small 40S ribosomal subunit by the mRNA, either at the 5' m7G cap structure or at an internal site. The 40S subunit then moves to a start site, which is typically an AUG codon that is recognized by the initiator-methionine transfer RNA (tRNA) associated with the small subunit. The large 60S ribosomal subunit subsequently joins to form a ribosomal complex which is capable of peptide synthesis. During the elongation cycle, the ribosome facilitates base pairing interactions between codons in mRNAs and anti-codons in aminoacyl-tRNAs, which are tRNA molecules covalently linked to their cognate amino acids [51]. Figure 1a shows the codon-amino acid associations that comprise the genetic code. In the elongation cycle, the peptidyl transferase activity of the ribosome mediates the transfer of amino acids from tRNAs to a growing polypeptide chain. Polypeptide synthesis stops when the translating ribosome reaches a stop codon, which leads to dissociation of the ribosomal complex and release of the newly synthesized protein.

2.2 Altering Codon Usage

Codon optimization strategies attempt to increase protein expression by altering the codon usage of the gene. Altering codon usage is possible because 20 amino acids are encoded by 61 codons (Fig. 1a). Although methionine (Met) and tryptophan (Trp) are encoded by a single codon each, all other amino acids are specified by two, three, four, or six codons. Because of this degeneracy in the genetic code, it is possible for mRNA sequences with different synonymous codon compositions to encode the same polypeptide [52] (Fig. 1b). Synonymous codons therefore provide a great deal of flexibility. In fact, for recombinant protein expression, a gene can be synthesized without even knowing the mRNA sequence by reverse translating the amino acid sequence. This process of reverse translation was used to express the first recombinant peptide, somatostatin, without knowing the mRNA sequence [53]. As gene sequences became available and were analyzed, it became evident that synonymous codon usage in nature is not random. Bias in codon usage varies between different organisms, between different tissues of the same organism, and even between different parts of the same gene [54, 55]. Factors affecting codon bias in bacteria, yeast, and *Drosophila* include correlations between codon bias and

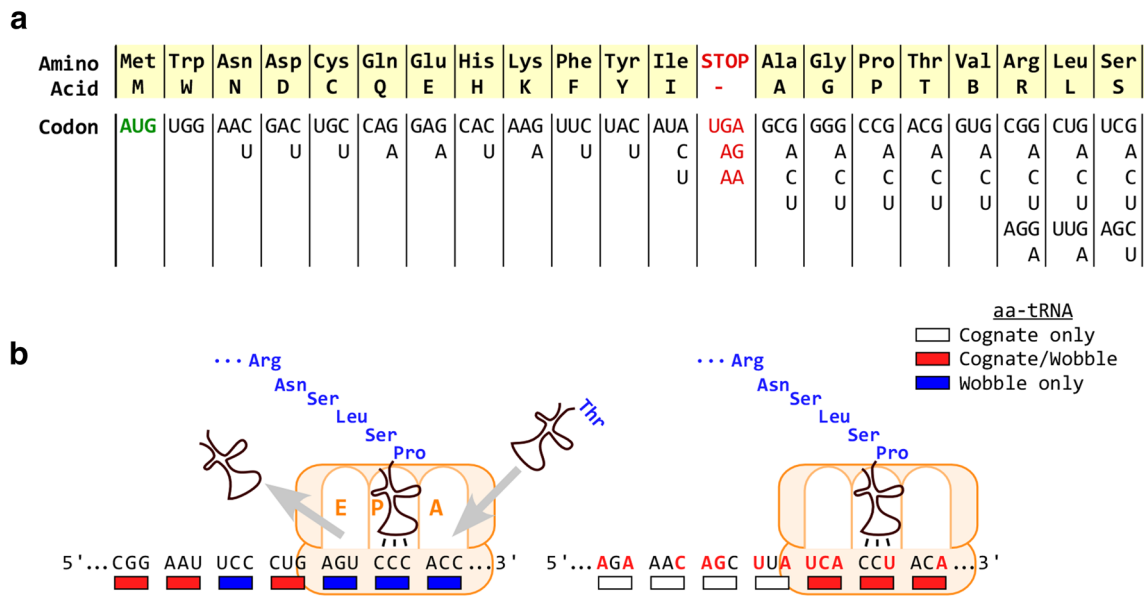


Fig. 1 Degeneracy of the genetic code. **a** Codon–amino acid associations. For each amino acid, both the three-letter and one-letter abbreviations are indicated. The AUG start codon, which encodes methionine, is indicated in green. This same codon is used to specify methionine residues within coding regions. Three stop codons are indicated in red; they do not specify amino acids but terminate translation. With exception of methionine and tryptophan, all amino acids are coded by two or more codons. **b** Degeneracy enables mRNAs containing different synonymous codons to encode the same polypeptide. This example shows how the same peptide sequence can be translated from mRNAs that differ significantly in their primary structure. In this example, the mRNA sequences in the left and right panels encode the same peptide but do not use any of the same codons and are only $\approx 43\%$ identical at the nucleotide level. The nucleotide differences are indicated in red bold type in the right panel. Based on human codon usage [65], codons underlined by white bars can only be translated by the corresponding (cognate) aa-tRNA; codons underlined by red bars can be translated by both cognate and wobble tRNAs, and those underlined by blue bars can only be translated by

wobble tRNAs because these codons lack a corresponding tRNA gene. In these illustrations, ribosomal subunits are indicated schematically as peach-colored structures; the smaller structure represents the 40S subunit, and the larger one represents the 60S subunit. The tRNA binding sites are labeled A, P, and E. For simplicity, each ribosome is shown with a tRNA molecule in the P site; the tRNA molecules are represented as cloverleaf structures. The tRNA in the P site is shown with the peptide chain encoded by the mRNA sequence shown. The next elongation cycle would involve recognition of the codon in the A site (ACC in the left panel; ACA in the right panel) by an aminoacyl (charged) Thr-tRNA. The peptidyl transferase activity of the ribosome would transfer the peptide chain from the tRNA in the P site to the threonine on the tRNA in the A site. A one-codon shift of the mRNA through the ribosome in the 3' direction would then leave an uncharged tRNA in the E site, the tRNA with the growing peptide chain in the P site, and an empty A site, ready for the next aminoacyl tRNA. A aminoacyl, aa-tRNA aminoacyl-tRNA, E exit, mRNA messenger RNA, P peptidyl, tRNA transfer RNA

translation efficiency [56–60]. Other variables affecting codon bias include the background nucleotide composition of the genome, which can vary significantly even within genomes [61]. In addition, codon bias can be affected by the expression levels of various tRNAs, which can vary between different tissues [62–64]. Moreover, even within individual genes, codon bias can be influenced by various constraints, which include splicing motifs, conserved mRNA secondary structures, amino-terminal coding sequences (codon ramp), as well as constraints affecting protein folding [55].

Different codon optimization strategies use synonymous codons to alter numerous features of mRNA coding sequences that can inhibit expression, including putative splice donor and acceptor sites. In addition, synonymous codons are used for convenience, e.g., to facilitate gene synthesis and cloning (reviewed in Mauro and Chappell [65]). However, the primary tactic for enhancing protein

expression involves increasing the rate of synthesis by eliminating or minimizing occurrences of rare codons. The assumption is that poor expression is caused by poor codon usage. Over the years, codon optimization approaches have ranged from relatively simple approaches that replace all codons with the most frequently used ones [66, 67], to seemingly more sophisticated approaches, such as codon harmonization, which try to maintain regions of slow translation that are thought to be important for protein folding [68]. This approach of maintaining regions of slow translation may be oversimplified as various lines of evidence suggest that protein folding can be affected both by codons that are typically thought to mediate a slow rate of translation—to increase folding—as well as by codons thought to mediate a fast rate, which may be important for reducing the possibility of misfolded intermediates [69].

Together with my colleague Stephen Chappell, we have previously discussed and critically analyzed various codon

optimization approaches for use in in vivo applications [65]. We identified three key assumptions that underlie various codon optimization strategies: (1) rare codons are rate-limiting for protein production; (2) synonymous codons are interchangeable without affecting protein structure and function; and (3) protein production can be increased by replacing rare codons with frequently used ones. A review of the literature indicates that these assumptions were either poorly supported or not generalizable. For example, the notion that rare codons are rate limiting for protein production is based on studies in *Escherichia coli* and lower eukaryotes and there is little evidence to support this idea in mammalian cells. In addition, there is abundant evidence demonstrating that synonymous codon changes, even individual codon changes, can significantly alter the formation of messenger ribonucleoprotein particles (mRNPs), mRNA secondary structure, mRNA stability, microRNA binding, translation, and protein folding [70–72].

2.2.1 Codon Usage in Mammals

One of the reasons codon usage is different in mammals is that a significant amount of variation in synonymous codon usage appears to be correlated with differences in the GC content of chromosomal regions known as isochores [61]. Isochores are large segments of DNA that have a uniform GC composition and encompass both coding and non-coding regions. An analysis of synonymous codon usage of different functional categories of human genes revealed that $\approx 70\%$ of the variation in synonymous codon usage between genes could be explained by the GC content of the chromosomal region, as well as meiotic recombination, which is more common in these regions. Notably, synonymous codon differences caused by large-scale variations in GC content were found to be independent of the functional category of the genes. This observation indicates that different highly expressed genes in the same cell have different patterns of synonymous codon usage. For many of these genes, codon usage does not match tRNA abundance [61, 63, 73].

In non-mammalian organisms, various studies have indicated that highly expressed genes contain more frequently used codons, which in many cases correlate with the expression levels of the corresponding tRNAs. Recent studies in *E. coli*, fungi, yeast, and *Drosophila* have demonstrated that frequently used synonymous codons have faster elongation rates than less frequently used codons [56–60]. Although there is not yet any comparable evidence in mammalian cells, analyses of mRNA and tRNA populations do not support this idea. Indeed, various studies have reported good correspondence between overall codon usage in cells and corresponding tRNA levels.

In one study, an analysis of different human cell types identified two distinct tRNA pools that are differentially expressed in proliferating or differentiated cells [74]. The authors found that codon usage in the transcriptome was coordinated with the expression of corresponding tRNAs such that there was a balance between the codon populations and the tRNA pools that were required for their translation. Similar results were found in a study that determined the frequency of usage for all codons as well as tRNA expression levels in mouse liver and brain tissues at eight different developmental stages [63]. The results showed that the codon pools from the expressed mRNAs and the anticodon pools were highly correlated in both tissues through development. In addition, it was noted that there did not appear to be differential codon usage between highly expressed and poorly expressed genes. In another study, tRNA pools and codon usage were analyzed in human and mouse liver cancer cell lines (in vitro) and quiescent liver cells (in vivo) [73]. The authors concluded that the tRNA pool of any of these cell types was capable of translating the mRNA transcriptomes of any other cell type with similar efficiency. In addition, no evidence was found to support the notion that highly expressed mRNAs in the different cell types were optimized for translation efficiency. The authors suggested that any variabilities in codon usage between different gene sets were best explained by variations in GC content.

In mammals, lack of evidence for slower elongation rates at rare codons also comes from ribosome profiling studies. Ribosome profiling is a technique that uses deep sequencing to identify segments of mRNAs that are protected by ribosomes in cells. In a study performed using mouse embryonic stem cells, cells were treated with the drug harringtonine to stall new initiation events at the start codon. By using ribosome profiling to monitor run-off elongation, it was possible to determine the kinetics of translation in these cells [75]. This study reported that translation speed was largely independent of codon usage and there was no evidence of ribosomal pausing at rare codons. Although the authors did not rule out the possibility of specific examples, they found no evidence for a large effect of codon usage on the overall rate of elongation.

Another study supporting the notion that rare codons are not limiting for expression comes from an analysis of protein coding sequences in the human genome [76]. This study found that rare codons for alanine, proline, serine, and threonine are used preferentially in the first 50 codons of the coding region. The effect on expression of the rare alanine codon was tested in constructs with multiple alanine codons in the first 50 codons of a synthetic fusion protein. The results showed that expression from constructs containing the rare alanine codon was much higher than

from those containing the more frequently used alanine codons.

2.2.2 Wobble Decoding

An important element that can affect the rate of elongation and is disrupted upon codon optimization is the type of tRNA interaction, i.e., whether a codon uses standard (Watson–Crick) or wobble tRNA base pairing interactions. A codon pairs to its cognate tRNA via three Watson–Crick interactions; by contrast, a codon can base pair to a non-cognate tRNA via a wobble interaction that uses standard base pairing for the first two nucleotides and less stringent base pairing for the third nucleotide, e.g., G:U base pairing. Ribosome profiling in *Caenorhabditis elegans* and a human cell line (HeLa) indicated that the rate of elongation is slower at codons decoded by wobble tRNA interactions than at codons decoded by Watson–Crick tRNA interactions [77]. In human cells, there was an ≈ 65 to 300% increase in ribosome occupancy at codon positions for which the third base interaction was a wobble G:U base pair compared to a standard G:C base pair, consistent with a slower rate of elongation at these codons. An in-depth analysis of ribosome profiling data in yeast also demonstrated that recognition of codons by wobble base pairing is slower than for codons translated by Watson–Crick base pairing [78].

In yeast, wobble appears to be associated with another finding, which is that specific pairs of adjacent codons significantly reduce the rate of elongation, independent of any dipeptide effects [79]. In this study, it was observed that for 16 of 17 inhibitory codon pairs, one or both codons were wobble codons. In addition, for 10 of 11 pairs, it was shown that codon order was important, suggesting that the slower translation at some codon pairs was caused by more than just the additive effects of each codon. Moreover, the inhibitory effects could be suppressed more effectively by overexpressing a non-native tRNA with an exact match to the anticodon, than with native (wobble decoding) tRNAs. In another study, these inhibitory codon pairs were shown to be associated with faster mRNA decay [80]. Additional evidence that specific di-codon pairs affect translation in mammalian cells comes from an analysis of 35 synonymous single nucleotide polymorphisms (sSNPs) in 27 different genes for 22 human genetic diseases or traits, which identified disruptions determined by pairs of consecutive codons rather than by individual codon bias [81].

Wobble decoding is associated with significant complexity, which is disrupted by codon optimization. This complexity is illustrated in Fig. 1 of Mauro and Chappell [65]. Additional complexity comes from the fact wobble itself can vary between organisms that express different subsets of the 61 possible aminoacyl tRNAs. Synonymous

codon changes can disrupt the pattern of cognate and wobble tRNA interactions because some codons are decoded by only one cognate tRNA, other codons are decoded by both cognate and wobble tRNAs, and still other codons lack a corresponding tRNA gene and are decoded by only non-cognate tRNAs. In Fig. 1b, notice how the pattern of cognate, cognate/wobble, and wobble codon usage is completely different for the two mRNAs. tRNA wobble is only one variable, but shows the complexity of trying to understand and recreate the elongation rhythm of an mRNA.

2.2.3 Additional Considerations

The goal of maintaining the natural folding pattern of a recombinant protein by preserving the elongation rhythm of the natural mRNA in the body is not trivial. There are numerous differences between the natural cell type in which a protein of interest is expressed, e.g., liver sinusoidal cells in the body, and a production cell line, such as CHO or human embryonic kidney 293 (HEK293), in a bioreactor under production conditions. Differences that could affect elongation include tRNA concentrations, levels of other mRNAs that determine whether translation conditions are competitive or non-competitive, and the codon composition of the transcriptome. tRNA concentrations are determined in part by which tRNA genes are present, the number of genes, and their expression levels. An additional potential consideration for production cell lines involves variations in codon usage that may be influenced by culture conditions, which are likely to affect both tRNA expression and the transcriptome. Moreover, overexpression of recombinant mRNA, either by transcription or translational enhancement, may itself disrupt the balance of codon demand and tRNA abundance, causing some tRNAs to become limiting and inadvertently altering elongation rates at specific codons. Even an unmodified natural mRNA coding sequence is likely to be translated differently in a production cell line than in the body. An important question is how do these differences affect protein folding?

Another significant consideration associated with the use of codon-optimized constructs for in vivo applications, including gene therapy, RNA therapeutics, and DNA/RNA vaccines, is translation from out-of-frame cryptic translation start sites in coding regions [65]. Many out-of-frame reading frames are altered by codon optimization and encode novel peptides that may have undesirable properties. An example of this type of cryptic initiation was reported by Lorenz et al. [82] who codon optimized a papillomavirus E7 oncoprotein mRNA to isolate E7-specific T cell receptors for T cell receptor gene therapy. The codon-optimized mRNA was expressed from

transfected dendritic cells that were incubated with T cells. The results revealed a T cell response with the codon-optimized but not wild-type sequence. This response was mapped to a cryptic peptide from the +3-alternative reading frame. Granted, expression of novel cryptic peptides from codon-optimized mRNAs is less serious when expressing therapeutic protein in a bioreactor because the therapeutic proteins are purified. However, it is still a consideration because the novel cryptic peptides may have unexpected biological effects which may negatively affect the physiology of the cells or the expression and processing of the therapeutic protein.

The various lines of evidence discussed here indicate that trends regarding codon usage and elongation rates in mammals are much weaker than in other organisms. These lines of evidence include the effects of chromosomal isochores on GC distribution patterns and codon usage, the observed balance in codon and tRNA pools, as well as the effects associated with wobble decoding. However, these findings do not rule out possible effects for some genes or under certain conditions. For example, a study in HEK293 cells suggested that non-optimal codons are critical for promoting the translation of selective mRNAs during amino acid starvation [83].

2.3 Mammalian Codon Optimization: What's the Harm?

Synonymous codon mutations are known to potentially affect protein expression at various levels and there is mounting evidence indicating that translation itself is affected and can lead to dramatic alterations in the conformation and processing of some proteins. Numerous examples in various reviews document this evidence (see McCarthy et al. [81], Gotea et al. [84], and Hunt et al. [85]).

A critical issue with codon optimization is that while it maintains the amino acid sequence of a protein, it can disrupt multiple other layers of information encoded in mRNA coding sequences [86, 87]. These overlapping functional elements are often difficult to identify. However, some of these elements can affect the rate of elongation locally, alter protein folding, and lead to changes in protein conformation and post-translational modifications. The non-neutral nature of synonymous codon mutations has been exploited in various studies which have screened synonymous mRNA variants to identify conformational variants of the encoded proteins with altered function (e.g., Cheong et al. [88]). The non-interchangeability of synonymous codons is also the basis for large-scale random recoding, which has been used successfully to attenuate more than a dozen viruses [89–91]. The approach of using synonymous codon mutations to alter protein function is

very useful for particular applications, including industrial enzyme optimization. However, the possible effects of synonymous codon mutations on protein conformation are much riskier in the production of therapeutic proteins as they may lead to problems in the patient, including production of anti-drug antibodies that reduce drug efficacy, as well as immunogenic complications [93–95].

Disruption of overlapping information defining mRNA secondary structures that affect the rate of elongation at specific sites in the coding region was suggested to explain results obtained following codon optimization of a feline endogenous retroviral RD114-TR envelope protein [96]. Although codon optimization resulted in increased protein yield, there were associated glycosylation defects that interfered with correct processing of the envelope protein which led to the production of an *inactive* protein.

Factors associated with production of recombinant therapeutic proteins in CHO, or other cell lines, can lead to differences with the natural protein that trigger production of anti-drug antibodies in patients. Differences may include glycosylation, factors affecting the integrity of the recombinant protein, and conformational alterations. Recombinant erythropoietin (EPO) illustrates the type of problem that might occur if anti-drug antibodies also recognize the endogenous protein. Some patients treated with recombinant EPO for anemia associated with chronic renal failure developed neutralizing antibodies against EPO [97, 98]. These antibodies inhibited the activities of both the recombinant and endogenous proteins, which stopped red blood cell production and caused patients to develop pure red cell aplasia. In one of these studies, recombinant EPO preparations from different manufacturers were compared and it was found that some formulations were more or less likely to result in the development of anti-EPO antibodies [98]. While it is not known if codon optimization of recombinant EPO constructs contributed to this problem, it illustrates the type of problem that might be expected if a codon-optimized mRNA gives rise to a recombinant protein with an altered conformation.

Synonymous codon mutations are worrying inasmuch as many diseases have been linked to single synonymous codon mutations. A codon-optimized mRNA can be altered by up to 80% from its native form [99]; consequently, the net result is the introduction of a large number of synonymous codon mutations into an mRNA. A recent example illustrating the effects of a single synonymous codon mutation in mammalian cells comes from the analysis of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene [64]. This study demonstrated that a synonymous mutation of a threonine codon in this gene (ACT to ACG) affected both the conformation and function of the CFTR protein. Analysis of ribosome-protected fragments in a cystic fibrosis bronchial epithelial cell line

revealed that ribosome occupancy of ACG codons was much higher than that of ACT codons; indeed, ACG was amongst the codons with the highest ribosome occupancy, suggesting that the mutated codon is one of the most slowly translated codons in these cells, and that the natural ACT codon is translated much more rapidly. These results were corroborated by data showing that the tRNA levels for these two codons were correlated with the predicted relative translation speeds of these codons. In addition, the authors showed that the structural and functional defects in the mutated CFTR protein could be rescued by increasing levels of the tRNA corresponding to the mutated ACG codon. These results strongly support the notion that a single synonymous codon mutation in the CFTR protein causes both structural and functional deficits because of slower translation at the mutated codon. Although the effects of a rare codon in this example seem contrary to those reported in many other studies in mammalian cells, it provides an example of the complexity of codon usage because the tRNA corresponding to the mutated ACG codon was not found to be rare in other human tissues, suggesting that the effects observed in the epithelial cell line are tissue specific.

Codon optimization should be considered one of various possible factors that may contribute to the immunogenicity of a recombinant protein. In addition, not all biologicals are equivalent in terms of potential safety issues that may arise. For example, recombinant monoclonal antibodies that function by targeting other molecules may be inherently safer than recombinant versions of natural proteins, which can have dramatic consequences if anti-drug antibodies against the recombinant protein recognize the endogenous protein. Nevertheless, in any case, an additional goal of codon optimization, beyond increased expression, is increased safety.

2.4 Lost Opportunities?

A potential problem associated with codon optimization is that it is routinely used to try to increase protein yields when a protein moves from academic and preclinical studies to clinical trials. However, it is likely that many academic and preclinical studies are performed using gene constructs based on natural mRNA sequences. Codon-optimized variants may behave differently and underlie instances in which a protein generated very promising data in preclinical studies but failed to perform as expected after being scaled up under GMP conditions. The concern is that highly effective protein drugs may be negatively affected or even fall by the wayside when a codon-optimized version of a protein is used.

For instance, molecules that are potentially very useful for vaccine development include broadly neutralizing

antibodies, e.g., from rare HIV-infected patients. The development of these antibodies in patients can take many years, often involving multiple rounds of extensive somatic mutation [100, 101]. Subtle changes such as those that arise from synonymous mutations associated with codon optimization may affect the binding activities of these broadly neutralizing antibodies and prevent them from functioning identically to those on which they were based, reducing or perhaps even eliminating their usefulness. This example is provided to illustrate the type of problems that may occur, and is not limited to broadly neutralizing antibodies from rare HIV-infected patients.

We do not know the extent to which codon optimization of recombinant proteins has resulted in reduced efficacy or increased immunogenicity. However, it is likely that in some cases these proteins represent lost opportunities and may be worth revisiting with non-codon-optimized mRNAs. This is particularly true for any proteins or antibodies that did not behave as expected after codon optimization, e.g., after being scaled up for clinical trials.

2.5 Why Does Codon Optimization Sometimes Increase Expression?

If codon optimization does indeed increase protein yields in mammalian cells because of enhanced codon usage and more efficient elongation, then this effect should be robust and reproducible. Although there is some evidence that the translation rates of some codon-optimized mRNAs are faster than those of non-optimized mRNAs (reviewed in Hanson and Coller [72]), increased expression does not seem to be a general finding, and numerous studies report little or no effect [65, 102]. In unpublished studies, we ordered codon optimized light and heavy chain genes for a mAb. Three light chain genes and three heavy chain genes were ordered from the same commercial provider. Comparison of the codon-optimized nucleotide sequences revealed that they were all different, i.e., different synonymous codon mutations were used for each gene. Strikingly, when combinations of these light and heavy chain genes ($t = 9$) were expressed in transiently transfected CHO cells and mAb expression levels were compared, the results showed that expression varied by > 5 -fold. The magnitude of the difference in mAb expression between different light and heavy chain combinations is difficult to reconcile with the proposed mechanism of increased elongation rates and does not inspire confidence regarding the expected expression properties of codon-optimized genes.

It seems unlikely that the increased expression of some codon-optimized mRNAs in mammalian cells is due to increased elongation rates, but rather the result of an inadvertent event. For example, increased expression may

result from elevated recombinant mRNA levels, which can occur by various mechanisms, including disruption of a miRNA seed sequence, decreased degradation of the mRNA, or increased transcription. In yeast, it was observed that codon usage bias was positively correlated with mRNA levels, which was at least partially due to effects on mRNA stability [103]. In addition, several recent studies have indicated that the effects of codon optimization occur at the level of transcription. In *Neurospora* it was shown that increased mRNA and protein levels obtained from codon-optimized mRNAs were not due to increased mRNA stability or translation but increased transcription [104]. This study suggested that some genes with non-optimal codons undergo transcriptional silencing at the chromatin level. A similar conclusion was reached in studies performed in mammalian cells which analyzed two Toll-like receptors (TLRs) [105]. This study showed that codon optimization of TLR7 increased its expression by 40-fold, whereas codon optimization of a closely related protein (TLR9) had no effect. Ribosome profiling studies indicated that the translation efficiency of codon-optimized TLR7 was only modestly increased and that the effect on expression was caused primarily by increased mRNA levels that resulted from increased transcription. The authors suggested that the effect on transcription was caused by an increase in GC content following codon optimization.

2.6 Suggested New Goals for Codon Optimization

In many cases, codon optimization enhances protein expression, and it is expected that these methods will continue to improve as algorithms incorporate empirical observations based on codon usage and patterns that are correlated with high protein expression [106]. This is acceptable if the goal is increased expression, and it is appropriate for some applications, e.g., for protein evolution and increasing the expression and/or activity of industrial enzymes. However, for recombinant expression of natural therapeutic proteins in targeted cells, an additional goal should be to maintain the conformation and processing of the natural protein sequences. As suggested earlier, the best approach for increasing protein production is to increase the rate of translation initiation, directly or through factors affecting this process, for example, by incorporating translation enhancer elements, increasing mRNA levels, or using introns. Because of the potential problems associated with synonymous codon mutations, it is suggested that they should be used sparingly if at all, and, if so, with scientific justification. For the production of therapeutic proteins, it seems difficult to justify the large number of synonymous mutations associated with codon optimization.

In light of the possibility that codon optimization can lead to alterations in protein conformation, it has been suggested that it is crucial to assess the consequences of codon optimization before using a recombinant protein drug in patients [70]. The increased use of high-resolution methods for comparing conformational differences between proteins derived from natural and codon-optimized mRNAs is useful in identifying protein variants that may be potentially harmful [107]. It is expected that the development of new methods for rapidly and more easily probing protein conformation will enable screening of large numbers of protein variants at an early stage of development. It seems that there is still a need for additional research.

3 Summary and Conclusions

Numerous studies indicate that the scientific bases for codon optimization in mammals are poorly supported; because of this, it is difficult to justify the use of codon optimization as a tool for bioproduction of therapeutic proteins. The question that therefore needs to be asked is why is codon optimization still commonly used? One possible reason is that in some cases, higher levels of protein expression are required for clinical trials and commercialization, and these expression levels can sometimes be obtained by using codon-optimized mRNAs—regardless of the underlying mechanism. Unfortunately, some of the potential problems associated with codon optimization, which can affect protein function and increase immunogenicity, may not be seen until the drug is in late stage clinical trials, or after the drug is on the market [99].

It is surprising that biotherapeutic approvals by the FDA do not yet require disclosure of gene sequences [5], as knowledge of gene structure—native or codon optimized—would be useful in determining whether particular problems affecting drug safety are associated with codon optimization. Gene sequence information should be an important component in the FDA's quality by design considerations. Thankfully, the effects of synonymous codon usage and potential problems associated with codon optimization have been recognized and are actively being studied by scientists at the FDA [5, 102]. Hopefully the FDA will soon take steps to address this situation. It should be noted that the absence of nucleic acid information also significantly impacts the generation of biosimilars, for which similarity is hard to achieve without knowing the gene sequence of the innovator drug. However, because biosimilars are replacing proteins developed using older technologies, it has been suggested that it may actually be better if biosimilars are not identical to the reference

protein [5]. Moving towards the use of more natural mRNA sequences would be a step in the right direction.

Acknowledgements I would like to thank Stephen Chappell and Daiki Matsuda for critical reading of the manuscript and valuable comments, Kathryn Crossin for helpful discussions on how codon optimization might lead to lost opportunities during therapeutic protein development, and Daniel Ivansson for a series of thoughtful discussions that sparked my interest in this topic.

Compliance with Ethical Standards

Funding No funding has been received for the conduct of this study and/or preparation of this manuscript.

Conflict of interest Vincent Mauro declares no conflict of interest.

References

- Ladisch MR, Kohlmann KL. Recombinant human insulin. *Biotechnol Prog*. 1992;8(6):469–78.
- Lieuw K. Many factor VIII products available in the treatment of hemophilia A: an embarrassment of riches? *J Blood Med*. 2017;8:67–73.
- Andersen DC, Krummen L. Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol*. 2002;13:117–23.
- Dumont J, Eewart D, Mei B, Estes S, Kshirsagar R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol*. 2016;36(6):1110–22.
- Lagasse HA, Alexaki A, Simhadri VL, Katagiri NH, Jankowski W, Sauna ZE, et al. Recent advances in (therapeutic protein) drug development. *F1000Res*. 2017;6:113.
- Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol*. 2012;93(3):917–30.
- Davami F, Eghbalpour F, Barkhordari F, Mahboudi F. Effect of peptone feeding on transient gene expression process in CHO DG44. *Avicenna J Med Biotechnol*. 2014;6(3):147–55.
- Delafosse L, Xu P, Durocher Y. Comparative study of polyethylenimines for transient gene expression in mammalian HEK293 and CHO cells. *J Biotechnol*. 2016;10(227):103–11.
- Lattenmayer C, Loeschel M, Schriegl K, Steinfeldner W, Sterovsky T, Trummer E, et al. Protein-free transfection of CHO host cells with an IgG-fusion protein: selection and characterization of stable high producers and comparison to conventionally transfected clones. *Biotechnol Bioeng*. 2007;96(6):1118–26.
- Kramer O, Klausung S, Noll T. Methods in mammalian cell line engineering: from random mutagenesis to sequence-specific approaches. *Appl Microbiol Biotechnol*. 2010;88(2):425–36.
- Harrison RG. Observations on the living developing nerve fiber. *Proc Soc Exptl Biol Med*. 1907;4:140–3.
- Chain E, Florey HW, Adelaide MB, Gardner AD, Oxford DM, Heatley NG, et al. Penicillin as a chemotherapeutic agent. *Lancet*. 1940;236:226–8.
- Schatz A, Bugie E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proc Soc Exp Biol Med*. 1944;55:66–9.
- Eagle H. Nutrition needs of mammalian cells in tissue culture. *Science*. 1955;122:501–14.
- Thyagarajan B, Calos MP. Site-specific integration for high-level protein production in mammalian cells. *Methods Mol Biol*. 2005;308:99–106.
- Wirth D, Gama-Norton L, Riemer P, Sandhu U, Schucht R, Hauser H. Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol*. 2007;18(5):411–9.
- Campbell M, Corisdeo S, McGee C, Kraichely D. Utilization of site-specific recombination for generating therapeutic protein producing cell lines. *Mol Biotechnol*. 2010;45(3):199–202.
- Suzuki T, Kazuki Y, Oshimura M, Hara T. A novel system for simultaneous or sequential integration of multiple gene-loading vectors into a defined site of a human artificial chromosome. *PLoS One*. 2014;9(10):e110404.
- Ahmadi M, Damavandi N, Akbari Eidgahi MR, Davami F. Utilization of site-specific recombination in biopharmaceutical production. *Iran Biomed J*. 2016;20(2):68–76.
- Nakamura T, Omasa T. Optimization of cell line development in the GS-CHO expression system using a high-throughput, single cell-based clone selection system. *J Biosci Bioeng*. 2015;120(3):323–9.
- Priola JJ, Calzadilla N, Baumann M, Borth N, Tate CG, Betenbaugh MJ. High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol J*. 2016;11(7):853–65.
- Kim M, O'Callaghan PM, Droms KA, James DC. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol Bioeng*. 2011;108(10):2434–46.
- Pilbrough W, Munro TP, Gray P. Intracolonial protein expression heterogeneity in recombinant CHO cells. *PLoS One*. 2009;4(12):e8432.
- Dharshanan S, Chong H, Hung CS, Zamrod Z, Kamal N. Rapid automated selection of mammalian cell line secreting high level of humanized monoclonal antibody using Clone Pix FL system and the correlation between exterior median intensity and antibody productivity. *Electron J Biotechnol*. 2011;14(2). <https://doi.org/10.2225/vol14-issue2-fulltext-7>.
- Tsuruta LR, Lopes Dos Santos M, Yeda FP, Okamoto OK, Moro AM. Genetic analyses of Per. C6 cell clones producing a therapeutic monoclonal antibody regarding productivity and long-term stability. *Appl Microbiol Biotechnol*. 2016;100(23):10031–41.
- Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. 2004;22:1393–8.
- Kunert R, Reinhart D. Advances in recombinant antibody manufacturing. *Appl Microbiol Biotechnol*. 2016;100(8):3451–61.
- Kinch MS. An overview of FDA-approved biologics medicines. *Drug Discov Today*. 2015;20(4):393–8.
- Jayapal KP, Wlaschin KF, Hu WS, Yap MG. Recombinant protein therapeutics from CHO cells—20 years and counting. *CHO Consortium SBE Special Section 2007*:40–7.
- Kretzmer G. Industrial processes with animal cells. *Appl Microbiol Biotechnol*. 2002;59:135–42.
- Ayyar BV, Arora S, Ravi SS. Optimizing antibody expression: the nuts and bolts. *Methods*. 2017;01(116):51–62.
- Brown AJ, James DC. Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol Adv*. 2016;34(5):492–503.
- Wang W, Jia YL, Li YC, Jing CQ, Guo X, Shang XF, et al. Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. *Sci Rep*. 2017;7(1):10416.
- Ebadat S, Ahmadi M, Nematpour F, Barkhordari F, Mahdian R, et al. Evaluating the efficiency of CHEF and CMV promoter with IRES and Furin/2A linker sequences for monoclonal antibody expression in CHO cells. *PLoS One*. 2017;12(10):e0185967.

35. Majocchi S, Aritonovska E, Mermod N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res.* 2014;42(1):193–204.
36. Kaufman RJ. Overview of vector design for mammalian gene expression. *Methods Mol Biol.* 1997;62:287–300.
37. Gu MB, Kern JA, Todd P, Kompala DS. Effect of amplification of dhfr and lac Z genes on growth and beta-galactosidase expression in suspension cultures of recombinant CHO cells. *Cytotechnology.* 1992;9:237–45.
38. Payne SH. The utility of protein and mRNA correlation. *Trends Biochem Sci.* 2015;40(1):1–3.
39. Vogel C. Evolution. Protein expression under pressure. *Science.* 2013;342(6162):1052–3.
40. Wurm FM, Pallavicini MG, Arathoon R. Integration and stability of CHO amplicons containing plasmid sequences. *Dev Biol Stand.* 1992;76:69–82.
41. Kim SJ, Lee GM. Cytogenetic analysis of chimeric antibody-producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng.* 1999;64:741–9.
42. Gallegos JE, Rose AB. The enduring mystery of intron-mediated enhancement. *Plant Sci.* 2015;237:8–15.
43. Chappell SA, Edelman GM, Mauro VP. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci USA.* 2000;97:1536–41.
44. Chappell SA, Edelman GM, Mauro VP. Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc Natl Acad Sci USA.* 2006;103(48):18077–82.
45. Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* 2012;9(5):563–76.
46. Gouse BM, Boehme AK, Monlezun DJ, Siegler JE, George AJ, Brag K, et al. New thrombotic events in ischemic stroke patients with elevated factor VIII. *Thrombosis.* 2014;2014:302861.
47. Kumar SR. Industrial production of clotting factors: challenges of expression, and choice of host cells. *Biotechnol J.* 2015;10(7):995–1004.
48. Williams JA. Improving DNA vaccine performance through vector design. *Curr Gene Ther.* 2014;14(3):170–89.
49. Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M. Engineering genes for predictable protein expression. *Protein Expr Purif.* 2012;83(1):37–46.
50. Van Der Kelen K, Beyaert R, Inze D, De Veylder L. Translational control of eukaryotic gene expression. *Crit Rev Biochem Mol Biol.* 2009;44(4):143–68.
51. Ling C, Ermolenko DN. Structural insights into ribosome translocation. *Wiley Interdiscip Rev RNA.* 2016;7(5):620–36.
52. Welch M, Villalobos A, Gustafsson C, Minshull J. You're one in a googol: optimizing genes for protein expression. *J R Soc Interface.* 2009;6(6 Suppl 4):S467–76.
53. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, et al. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science.* 1977;198(4321):1056–63.
54. Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, et al. A new and updated resource for codon usage tables. *BMC Bioinform.* 2017;18(1):391.
55. Supek F. The code of silence: widespread associations between synonymous codon biases and gene function. *J Mol Evol.* 2016;82(1):65–73.
56. Gardin J, Yeasmin R, Yurovsky A, Cai Y, Skiena S, Fletcher B. Measurement of average decoding rates of the 61 sense codons in vivo. *eLife.* 2014;3. <https://doi.org/10.7554/eLife.03735>.
57. Dana A, Tuller T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* 2014;42(14):9171–81.
58. Dana A, Tuller T. Mean of the typical decoding rates: a new translation efficiency index based on the analysis of ribosome profiling data. *G3.* 2014;5(1):73–80.
59. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, et al. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell.* 2015;59(5):744–54.
60. Paulet D, David A, Rivals E. Ribo-seq enlightens codon usage bias. *DNA Res Int J Rapid Publ Rep Genes Genom.* 2017;24(3):303–10.
61. Pouyet F, Mouchiroud D, Duret L, Semon M. Recombination, meiotic expression and human codon usage. *eLife.* 2017;6. <https://doi.org/10.7554/eLife.27344>.
62. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* 2006;2(12):e221.
63. Schmitt BM, Rudolph KL, Karagianni P, Fonseca NA, White RJ, Talianidis I, et al. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Res.* 2014;24(11):1797–807.
64. Kirchner S, Cai Z, Rauscher R, Kastelic N, Anding M, Czech A, et al. Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLoS Biol.* 2017;15(5):e2000779.
65. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med.* 2014;20(11):604–13.
66. Richardson SM, Wheelan SJ, Yarrington RM, Boeke JD. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* 2006;16(4):550–6.
67. Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S. Gene designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinf.* 2006;7:285.
68. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One.* 2008;3(5):e2189.
69. Wang E, Wang J, Chen C, Xiao Y. Computational evidence that fast translation speed can increase the probability of cotranslational protein folding. *Sci Rep.* 2015;21(5):15316.
70. Bali V, Bebek Z. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *Int J Biochem Cell Biol.* 2015;64:58–74.
71. Diederichs S, Bartsch L, Berkman JC, Froese K, Heitmann J, Hoppe C, et al. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Mol Med.* 2016;8(5):442–57.
72. Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol.* 2018;19(1):20–30.
73. Rudolph KL, Schmitt BM, Villar D, White RJ, Marioni JC, Kutter C, et al. Codon-driven translational efficiency is stable across diverse mammalian cell states. *PLoS Genet.* 2016;12(5):e1006024.
74. Gingold H, Tehler D, Christoffersen NR, Nielsen MM, Asmar F, Kooistra SM, et al. A dual program for translation regulation in cellular proliferation and differentiation. *Cell.* 2014;158(6):1281–92.

75. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011;147(4):789–802.
76. Park JH, Kwon M, Yamaguchi Y, Firestein BL, Park JY, Yun J, et al. Preferential use of minor codons in the translation initiation region of human genes. *Hum Genet*. 2017;136(1):67–74.
77. Stadler M, Fire A. Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*. 2011;17(12):2063–73.
78. Wang H, McManus J, Kingsford C. Accurate recovery of ribosome positions reveals slow translation of wobble-pairing codons in yeast. *J Comput Biol*. 2017;24(6):486–500.
79. Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*. 2016;166(3):679–90.
80. Harigaya Y, Parker R. The link between adjacent codon pairs and mRNA stability. *BMC Genom*. 2017;18(1):364.
81. McCarthy C, Carrea A, Diambra L. Bicodon bias can determine the role of synonymous SNPs in human diseases. *BMC Genom*. 2017;18(1):227.
82. Lorenz FK, Wilde S, Voigt K, Kieback E, Mosetter B, Schendel DJ, et al. Codon optimization of the human papillomavirus E7 oncogene induces a CD8 + T cell response to a cryptic epitope not harbored by wild-type E7. *PLoS One*. 2015;10(3):e0121633.
83. Saikia M, Wang X, Mao Y, Wan J, Pan T, Qian SB. Codon optimality controls differential mRNA translation during amino acid starvation. *RNA*. 2016;22(11):1719–27.
84. Gotea V, Gartner JJ, Qutob N, Elnitski L, Samuels Y. The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigm Cell Melanoma Res*. 2015;28(6):673–84.
85. Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. *Trends Genet*. 2014;30(7):308–21.
86. Firth AE. Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res*. 2014;42(20):12425–39.
87. Fahraeus R, Marin M, Olivares-Illana V. Whisper mutations: cryptic messages within the genetic code. *Oncogene*. 2016;35(29):3753–9.
88. Cheong DE, Ko KC, Han Y, Jeon HG, Sung BH, Kim GJ, et al. Enhancing functional expression of heterologous proteins through random substitution of genetic codes in the 5' coding region. *Biotechnol Bioeng*. 2015;112(4):822–6.
89. Martinez MA, Jordan-Paiz A, Franco S, Nevot M. Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol*. 2016;24(2):134–47.
90. de Fabritius L, Nougaiarde A, Aubry F, Gould EA, de Lamballerie X. Attenuation of tick-borne encephalitis virus using large-scale random codon re-encoding. *PLoS Pathog*. 2015;11(3):e1004738.
91. Wang B, Yang C, Tekes G, Mueller S, Paul A, Whelan SP, et al. Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. *MBio*. 2015;6(2):1–10.
92. Magistrelli G, Poitevin Y, Schlosser F, Pontini G, Malinge P, Jossierand S, et al. Optimizing assembly and production of native bispecific antibodies by codon de-optimization. *mAbs*. 2017;9(2):231–9.
93. Perez-De-Lis M, Retamozo S, Flores-Chavez A, Kostov B, Perez-Alvarez R, Brito-Zeron P, et al. Autoimmune diseases induced by biological agents. A review of 12,731 cases (BIO-GEAS Registry). *Expert Opin Drug Saf*. 2017;16(11):1255–71.
94. Strand V, Balsa A, Al-Saleh J, Barile-Fabris L, Horiuchi T, Takeuchi T, et al. Immunogenicity of biologics in chronic inflammatory diseases: a systematic review. *BioDrugs*. 2017;31(4):299–316.
95. Piga M, Chessa E, Ibba V, Mura V, Floris A, Cauli A, et al. Biologics-induced autoimmune renal disorders in chronic inflammatory rheumatic diseases: systematic literature review and analysis of a monocentric cohort. *Autoimmun Rev*. 2014;13(8):873–9.
96. Zucchelli E, Pema M, Stornaiuolo A, Piovan C, Scavullo C, Giuliani E, et al. Codon optimization leads to functional impairment of RD114-TR envelope glycoprotein. *Mol Ther Methods Clin Dev*. 2017;17(4):102–14.
97. Casadevall N, Nataf J, Viron B, Kolta A, Kiladjian JJ, Martin-Dupont P, et al. Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N Engl J Med*. 2002;346(7):469–75.
98. Cournoyer D, Toffelmire EB, Wells GA, Barber DL, Barrett BJ, Delage R, et al. Anti-erythropoietin antibody-mediated pure red cell aplasia after treatment with recombinant erythropoietin products: recommendations for minimization of risk. *J Am Soc Nephrol*. 2004;15(10):2728–34.
99. Katsnelson A. Breaking the silence. *Nat Med*. 2011;17(12):1536–8.
100. Derdeyn CA, Moore PL, Morris L. Development of broadly neutralizing antibodies from autologous neutralizing antibody responses in HIV infection. *Curr Opin HIV AIDS*. 2014;9(3):210–6.
101. McCoy LE, Burton DR. Identification and specificity of broadly neutralizing antibodies against HIV. *Immunol Rev*. 2017;275(1):11–20.
102. Kimchi-Sarfaty C, Schiller T, Hamasaki-Katagiri N, Khan MA, Yanover C, Sauna ZE. Building better drugs: developing and regulating engineered therapeutic proteins. *Trends Pharmacol Sci*. 2013;34(10):534–48.
103. Chen S, Li K, Cao W, Wang J, Zhao T, Huan Q, et al. Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level. *Mol Biol Evol*. 2017;34(11):2944–58.
104. Zhou Z, Dang Y, Zhou M, Li L, Yu CH, Fu J, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci USA*. 2016;113(41):E6117–25.
105. Newman ZR, Young JM, Ingolia NT, Barton GM. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci USA*. 2016;113(10):E1362–71.
106. Gustafsson C, Vallverdu J. The best model of a cat is several cats. *Trends Biotechnol*. 2016;34(3):207–13.
107. Kaur P, Kiselar J, Yang S, Chance MR. Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (MS). *Mol Cell Proteomics*. 2015;14(4):1159–68.