Published in final edited form as:

Int J Biochem Cell Biol. 2015 July ; 64: 58–74. doi:10.1016/j.biocel.2015.03.011.

Decoding Mechanisms by which Silent Codon Changes Influence Protein Biogenesis and Function

Vedrana Bali and Zsuzsanna Bebok*

Department of Cell, Developmental and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL, USA

Abstract

Scope—Synonymous codon usage has been a focus of investigation since the discovery of the genetic code and its redundancy. The occurrences of synonymous codons vary between species and within genes of the same genome, known as codon usage bias. Today, bioinformatics and experimental data allow us to compose a global view of the mechanisms by which the redundancy of the genetic code contributes to the complexity of biological systems from affecting survival in prokaryotes, to fine tuning the structure and function of proteins in higher eukaryotes. Studies analyzing the consequences of synonymous codon changes in different organisms have revealed that they impact nucleic acid stability, protein levels, structure and function without altering amino acid sequence. As such, synonymous mutations inevitably contribute to the pathogenesis of complex human diseases. Yet, fundamental questions remain unresolved regarding the impact of silent mutations in human disorders. In the present review we describe developments in this area concentrating on mechanisms by which synonymous mutations may affect protein function and human health.

Purpose—This synopsis illustrates the significance of synonymous mutations in disease pathogenesis. We review the different steps of gene expression affected by silent mutations, and assess the benefits and possible harmful effects of codon optimization applied in the development of therapeutic biologics.

Physiological and medical relevance—Understanding mechanisms by which synonymous mutations contribute to complex diseases such as cancer, neurodegeneration and genetic disorders, including the limitations of codon-optimized biologics, provides insight concerning interpretation of silent variants and future molecular therapies.

Keywords

codon usage bias (CUB); sSNP; mRNA structure; translation dynamics; protein folding

© 2015 Published by Elsevier Ltd.

Address correspondence to: Zsuzsanna Bebok, M.D. Associate Professor, The Department of Cell, Developmental and Integrative Biology, University of Alabama at Birmingham, 1918 University Boulevard, MCLM 350A, Birmingham, AL 35294. Phone: 205-975-5449; bebok@uab.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

Half a century ago, following the discovery of the triplet nature of the genetic code and codon redundancy (Crick, 1955, Crick et al., 1961), in an article entitled “*Molecules and Documents of Evolutionary History*”, Emile Zuckerkandl and Linus Pauling predicted: “*Due to isosemantic (synonymous) substitutions, there probably is more evolutionary history inscribed in the base sequence of nucleic acids than in the amino-acid sequence of the corresponding polypeptide chain*” (Zuckerkandl and Pauling, 1965). Today, in light of the sequencing and analysis of multiple prokaryotic, eukaryotic (http://www.genome.jp/kegg/catalog/org_list.html), and human genomes (Venter et al., 2001), our knowledge of the origins and consequences of codon redundancy is exponentially increasing. Sequence analysis has identified a discrepancy between the low number of protein-coding genes and the structural and functional complexity of the organisms (Claverie, 2001). This observation directed attention to the contribution of codon redundancy and non-coding segments of the DNA to the diversity and complexity of species (Shabalina et al., 2004, Shabalina and Spiridonov, 2004, Levine et al., 2014). Understanding the molecular mechanisms by which individual nucleotides govern the function of gene products is now a viable experimental objective. Synonymous mutations leading to an altered phenotype or disease provide excellent models for these types of studies. Although reviews on similar topics have been published elsewhere (Hunt et al., 2014, Shabalina et al., 2013, Sauna and Kimchi-Sarfaty, 2011), here we concentrate on the mechanisms by which synonymous mutations alter gene expression from pre-mRNA processing to cotranslational protein folding.

HISTORICAL PERSPECTIVE AND TERMINOLOGY

From DNA structure to codons and the degeneracy of the genetic code

Elucidating the structure of DNA (Watson and Crick, 1953) led to identification of the nucleotide triplets that comprise the genetic code (Crick, 1955, Crick et al., 1961), its universality (Woese, 1965, Woese, 1964), discovery of how codons are transcribed into mRNA (Martin et al., 1962), and the pathways by which mRNA is translated into a protein (Nirenberg, 1965, Nirenberg, 1963). Codons are nucleotide triplets, comprised of four bases adenine (A), cytosine (C) guanine (G) and thymine (T). These permit 64 possible codon variations, three of which represent translation termination signals. The remaining 61 codons encode 20 possible amino acids, resulting in codon redundancy, indicating that multiple codons can encode the same amino acid. Indeed, 18 of the 20 amino acids are encoded by multiple codons, and 10 of the amino acids can be charged onto multiple tRNAs, further increasing the degenerate state of the system (Subramaniam et al., 2013).

Genomic distribution of synonymous codons, codon usage bias (CUB)

Beginning with the analysis of *E. coli* (Ikemura, 1981), and extending to the genomes of a variety of prokaryotes, eukaryotes, and multicellular organisms (Ikemura, 1985, Chen et al., 2004, Duret, 2002b, Duret et al., 2002, Hense et al., 2010, Lucks et al., 2008, Plotkin et al., 2004), it has become increasingly apparent that synonymous codon usage is not random and that specific organisms have a propensity for certain codons, known as codon usage bias (CUB). CUB is universal, but with variable preference between organisms, tissues, and

genes in the same genome (Sharp et al., 1988, Sharp et al., 1993, Shields et al., 1988, Andersson and Kurland, 1990, Plotkin et al., 2004). Studies of molecular evolution attributed CUB to both genetic drift (neutralist) (Kimura, 1968b, Kimura, 1968a) and natural selection (selectionist) mechanisms (Mayr, 1961, Mayr, 1967). Although the selectionist or neutralist theories are not exclusive of each other (Bilgin et al., 2013, Bulmer, 1991, Duret, 2002a, Duret et al., 2002, Lercher et al., 2002, Wagner, 2008), we lack a complete understanding of the relative contribution of these pathways with regard to the evolution of CUB (reviewed in: (Nei and Nozawa, 2011, Nei et al., 2010, Serohijos and Shakhnovich, 2014, Hershberg and Petrov, 2008).

Important CUB differences exist between prokaryotes, multicellular organisms, and mammals (Wain-Hobson et al., 1981, Archetti, 2004, Jabbari and Bernardi, 2000, Andersson and Kurland, 1990, dos Reis et al., 2004, Long and Gillespie, 1991, Shao et al., 2012, Shields et al., 1988). In bacteria, genes encoding abundant proteins have the greatest CUB (Kanaya et al., 1996, Duret and Mouchiroud, 1999, Ikemura, 1981). As an example, *E. coli* prefers codons that enhance both the rate and accuracy of translation based on the availability of the tRNAs (Stoletzki and Eyre-Walker, 2007). In mammals, the intragenomic pattern of codons is different than in other taxa, yet CUB impacts multiple steps of mammalian gene expression (Shabalina et al., 2013, Chen et al., 2004, Du et al., 2014, Sharp et al., 1988, Chamary and Hurst, 2009, Chamary et al., 2006).

It should be noted that based on differences in gene regulation in prokaryotes and eukaryotes (Struhl, 1999), the mechanisms by which synonymous codons influence gene expression in different phylogenetic kingdoms may exhibit commonalities, but also some significant differences. Here we concentrate on eukaryotic gene expression regulation by synonymous mutations in relation to disease.

Single Nucleotide Polymorphism (SNP) and Synonymous SNP (sSNP)—When a single nucleotide (A, T, C, or G) in the genome differs between individuals, or between paired chromosomes, it is termed single nucleotide polymorphism (SNP) (Antonarakis, 1998). SNPs are considered the most common type of genetic variation among humans and may be present throughout the genome (Castle, 2011).

Here we focus on SNPs in the protein coding regions of the genome. SNPs that change the amino acid sequence of the protein are termed nonsynonymous SNP (nsSNP). However, as the result of code degeneracy, SNPs may occur without altering the amino acid sequence of the encoded protein and are referred to as synonymous or silent SNP (sSNP) (Antonarakis, 1998). sSNPs have long been considered inconsequential. However, as the number of genome-wide association studies (GWAS) has increased, numerous sSNPs have been linked to disease. Although most sSNPs do not alter the function of proteins or affect human disorders, several have been identified as contributors to complex human diseases including heart disease, diabetes, cancer, and neurological disorders (reviewed in: (Sauna and Kimchi-Sarfaty, 2011, Hunt et al., 2014)). We refrain from use of the term silent polymorphism and refer instead to sSNPs as synonymous mutations.

With reference to the focus of this review, it is notable that codon conservation has been observed in numerous regions of the mammalian genome, including sequences involved in protein, DNA-, and or RNA-binding. Specific DNA sequences such as exon-intron junctions, functionally conserved domains of proteins, and micro-RNA (miRNA) target sites, for example, are highly conserved. Intriguingly, these sequences are greatly depleted for sSNPs but not for nsSNPs, suggesting that codon conservation was driven by negative selection (Schattner and Diekhans, 2006, Bustamante et al., 2002, Hellmann et al., 2003). Molecular mechanisms by which synonymous mutations in these regions alter disease outcome will be discussed herein.

MECHANISMS BY WHICH SYNONYMOUS MUTATIONS CONTRIBUTE TO GENE EXPRESSION AND DISEASES PHENOTYPE

1. *Synonymous mutations and the formation of messenger ribonucleoprotein particles (mRNPs)*

Eukaryotic mRNAs exist as messenger ribonucleotide particles (mRNPs) that are composed of proteins (Muller et al., 2010, Mitchell and Parker, 2014), long- (Yoon et al., 2013) and small noncoding RNAs (ncRNAs) often with multiple functions (reviewed in: (Maute et al., 2014)). The assembly of an mRNP affects nearly every step of transcription and translation (e.g.; to prevent RNA:DNA hybrid formation, to stimulate elongation (Huertas and Aguilera, 2003), to facilitate 5' capping, aid in splicing (Reed, 2003), promote 3' cleavage and polyadenylation (Gott and Emeson, 2000), to govern mRNP export from the nucleus (Bhatia et al., 2014, Bjork and Wieslander, 2014), direct delivery to specific cellular compartments (Eom et al., 2003, Song et al., 2015), and aid cellular quality control and degradation (Eberle and Visa, 2014, Santos-Pereira et al., 2014)). Some components of an mRNP complex may work by repressing translation until the appropriate cellular location is reached (Mitchell and Parker, 2014), while others may operate as cell- and tissue-specific enhancers (Andersson et al., 2014). Importantly, not all mRNP constituents associate with pre-mRNAs or mRNAs in a sequence-specific manner, yet the binding of these “general machineries” is coordinated in a canonical fashion. Thus, the association of mRNP components with an mRNA is ultimately governed by sequence specificity (Mitchell and Parker, 2014).

The primary nucleotide sequence determines the secondary structure of mRNAs (Mortimer et al., 2014). The secondary structure is formed by base pairing of nucleotides creating stems, loops, and combinations of these- including stem-loops, pseudoknots, and tetraloops. The final shape of an mRNA is recognized by mRNP components and other ligands (reviewed in: (Mauger et al., 2013). Considering that the formation of mRNA secondary structures is based on nucleotide pairing, a single nucleotide change can alter this organization considerably (Figure 1). Moreover, because the primary sequence and structure of an mRNA (local context) affects the affinity of mRNA-binding factors (Mitchell and Parker, 2014), it is apparent that synonymous codons can greatly influence the formation of mRNA secondary structure and the nature of the subsequent mRNP complexes.

The evaluation of mRNP formation and mRNA interactions presents a challenge in identifying the underlying molecular mechanisms by which an sSNP may lead to a disease phenotype. However there is an increasing body of bioinformatics data linking synonymous mutations to diseases, and growing number of reports showing that defects in mRNP formation can lead to human illness (Table 1). Below we describe synonymous mutations that influence gene expression through changes in mRNP composition and/or binding affinity at different steps of gene expression.

1.1. sSNPs associated with changes in mRNA and protein binding—Amyotrophic lateral sclerosis (ALS) is a progressive, lethal, neurodegenerative disorder associated with mutations in the *SOD1* gene (Ge et al., 2006). Numerous missense mutations and a number of synonymous mutations such as Gly10-GGC>GGT (Jackson et al., 1997), Ser59-AGT>AGC and Ala140-GCT>GCA (Hosler et al., 1996), Thr116-ACA>ACG (Calder et al., 1995), Asn139-AAC>AAT and Gln153-CAA>CAG (Andersen et al., 1997), have also been reported in *SOD1*. Ge and colleagues analyzed these synonymous variants using *in vitro* transcribed mRNAs as templates in a UV-crosslinking study using different tissue homogenates, each containing a unique tissue specific assortment of RNA binding proteins. They observed reduced mRNP formation leading to decreased stability of the mRNAs carrying the synonymous mutations in extracts specifically from the spinal cord. Although *SOD1* is ubiquitously expressed, these results suggest the contribution of reduced *SOD1* activity to neuronal tissue vulnerability in ALS (Ge et al., 2006). In their study, Ge *et al.*, did not identify a specific *SOD1* mRNA binding factor, but in a subsequent report describing both sporadic- and familial ALS cases, it has been proposed that ALS is caused by aberrant mRNA metabolism (Strong, 2010). Further studies will be necessary to determine specific mechanisms by which synonymous mutations contribute to this process, and to identify mRNA binding proteins.

1.2. Synonymous mutations and splicing—Mammalian genes consist of coding (exonic) and non-coding (intronic) domains. In mature mRNAs, exons are ligated following removal of introns by the dynamic function of the spliceosome (components of the mRNP machinery), containing small nuclear RNAs (snRNAs) and proteins (reviewed in: (Matera and Wang, 2014)). Synonymous nucleotide changes can alter the composition, affinity, and function of spliceosomes. Consequently, improper identification of exon–intron boundaries or the failure to remove an intron generates defective mRNAs, dysfunctional proteins, and can result in human disorders (Faustino and Cooper, 2003) (Figure 2).

In higher eukaryotes, exon–intron junctions are defined by multiple sequence motifs. Among these are intronic and exonic cis-elements important for proper splice-site recognition. Intronic cis-elements include the nearly invariant 5' GU and 3' AG splice sites and a consensus branch point site localized deep within the intron (Cartegni et al., 2002). Exonic cis-elements called splice enhancers (ESE) (Liu et al., 1998, Schaal and Maniatis, 1999b, Schaal and Maniatis, 1999a) or silencers (ESS) have also been identified (Wang et al., 2004) that stimulate or repress mRNA splicing, respectively (reviewed in: (Kornblihtt et al., 2004, Pascale and Govoni, 2012, Rieder and Reenan, 2012, Matera and Wang, 2014)).

In human genes, a strong CUB exists in splicing regulatory and modulatory elements (Fairbrother et al., 2004, Majewski and Ott, 2002). The frequency of sSNP is low at exon boundaries where ESE density is high, suggesting selection against sSNPs in these regions (Parmley et al., 2006). In support of this idea, Willie and coworkers observed that in human genes, codon preference varied with the distance from splice sites and ESEs (Willie and Majewski, 2004). Molecular evolutionary analysis of the breast cancer-1, early onset (BRCA1) gene has also provided evidence for negative selection against synonymous mutations in ESEs (Hurst and Pal, 2001).

It has been estimated that up to 15% of all point mutations causing human genetic disorders disrupt mRNA processing (Krawczak et al., 1992). Indeed, the Human Gene Mutation Database identifies >16,000 point mutations targeting splicing signals (<http://www.hgmd.cf.ac.uk>) (Stenson et al., 2014). The contribution of synonymous mutations to this number is not clear, since they might be under-reported if they are assumed neutral.

1.2.1. Examples of sSNP-associated splicing defects (Table 1): Charcot-Marie-Tooth disease type 1B (CMT1B) is caused by mutations in the gene encoding myelin protein zero. In a familial case of late onset CMT1B a synonymous mutation in exon three, Val102-GTG>GTA, renders the binding site for a small nuclear RNA (snRNA U1) enhanced, causing truncation of exon three (Crehalet et al., 2010). As a component of the spliceosome, snRNA U1 binds to the 5' splice site of an intron during canonical mRNA splicing (Matera and Wang, 2014). Strong U1 binding likely disturbs the dynamic assembly/disassembly of the spliceosome resulting in defective splicing.

Another report concerns the X-linked metabolic disorder (Leigh's encephalomyelopathy), caused by pyruvate dehydrogenase (PDH)-E1 α deficiency. In a severe case of familial E1 α deficiency, a synonymous mutation Gly185-GGA>GGG caused aberrant splicing in a subset of cases, resulting in the skipping of exon six. Exon six contains a thiamine pyrophosphate-binding site necessary for proper enzyme function. Therefore, skipping this exon results in a dysfunctional protein (De Meirleir et al., 1994).

Pagani and colleagues analyzed the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, mutations of which cause cystic fibrosis. Multiple synonymous mutations in exon 12 of the human *CFTR* gene were shown to induce exon skipping, leading to decreased *CFTR* mRNA. The underlying mechanism was interpreted as disruption of a composite exonic regulatory element of splicing (CERES). Therefore, CERES sequence variations may represent disease-causing mechanisms and could also be responsible for phenotypic variability. Utilization of functional splicing assays for distinguishing polymorphisms from pathogenic mutations was suggested in cystic fibrosis (Pagani et al., 2003). Using *CFTR* exon 12 as model, the same group established that ~25% of all synonymous variations in exon 12 result in exon skipping and inactive *CFTR* protein (Pagani et al., 2005).

Cutting and colleagues reported a *de novo* synonymous mutation Ser1204-TCA>TCC contributing to Treacher Collins syndrome (TCS), an autosomal-dominant disorder associated with craniofacial morphogenesis defects through mutations in the *TCOF1* gene (Macaya et al., 2009). Family members carrying a benign *TCOF1* mutation Ala41Val have

no phenotype. However, analysis of *TCOF1* mRNA in a patient with a *de novo* Ser1204-TCA>TCC silent mutation revealed a transcript without exon 22. These results suggest that TCS was caused by the haploinsufficiency of *TCOF1* as a result of the synonymous mutation that resulted in exon skipping.

In other studies, Nielsen *et al.*, used patient cells, mini-gene analysis, and *in vitro* assays to demonstrate that a sSNP in exon five (Thr92-ACA>ACC) of the medium-chain acyl-CoA dehydrogenase gene, *MCAD*, caused exon skipping through inactivation of an ESE. Interestingly, this ESE inhibits an adjacent ESS and is necessary to define a suboptimal 3' splice site. A nearby mutation, Thr96Ile is also disease-associated, and thought to inactivate the ESS. Surprisingly, the Thr96Ile mutation itself does not affect splicing; however when both variants are present, the non-synonymous polymorphism compensated for the silent Thr92-ACA>ACC mutation and reverted the deleterious effect. These findings emphasize context-dependent effects of single-nucleotide polymorphisms on gene-expression regulation (Nielsen *et al.*, 2007).

Mutations in *PKD1* and *PKD2* cause polycystic kidney disease, the most common monogenic human disorder. Most disease-causing mutations affect *PKD1*, and many are synonymous. Using bioinformatic tools, mutations with the potential to affect pre-mRNA splicing were tested in a minigene analysis. The authors evaluated 33 mutations and found two synonymous variants (Gly109-GGA>GGT and Arg3753-CGC>CGA) that caused splicing defects by introducing strong donor splice sites in exons three and 39 respectively, resulting in aberrant mRNA (Gonzalez-Paredes *et al.*, 2014).

Another example is the Phe199-TTT>TTC synonymous mutation in exon eight of the human X-linked hypoxanthine-guanine phosphoryl transferase (*HPRT*) gene. This sSNP caused extensive skipping of exon eight, leading to Lesch–Nyhan syndrome (LNS), also known as Kelley–Seegmiller syndrome or juvenile gout. The authors proposed that the mutation causes changes to the secondary structure of the mRNA leading to inefficient exposure of proper splicing sites (Steingrimsdottir *et al.*, 1992).

2. The Significance of mRNA Secondary Structure

As alluded to above, changes in mRNA secondary structure can alter the binding of RNP components at any step during the life of an mRNA and ultimately influence gene expression. The mechanism(s) by which mRNA structure regulates gene expression comprise virtually any step of the gene expression process (Mortimer *et al.*, 2014) including transcription (Frieda and Block, 2012), splicing (Warf and Berglund, 2010), stability (Ferre-D'Amare, 2011), and translation (Mauger *et al.*, 2013, Mandal and Breaker, 2004). mRNA structure is also required for proper miRNA-mediated regulation (Meijer *et al.*, 2013, Gu *et al.*, 2014b), and mRNA cellular localization (Jambor *et al.*, 2014, Martin and Ephrussi, 2009). Most significantly, it has been demonstrated that mRNA conformation governs the expression of their encoded proteins both *in vitro* (Wan *et al.*, 2011, Kertesz *et al.*, 2010) and *in vivo*, as indicated by genome-wide mRNA structural profiling (Ding *et al.*, 2014).

2.1. Evolutionary selection of mRNA structures—In view of the widespread influence of mRNA conformation with regard to gene expression, it is reasonable to imagine

that codon selection was particularly important for the formation of functional mRNAs during evolution. According to Chamary and Hurst, synonymous mutations are likely to be under evolutionary selection because they affect the thermodynamic stability of an mRNA (Chamary and Hurst, 2005). In this context, it has been observed that functional RNA secondary structure is conserved across species (Pedersen et al., 2006, Smith et al., 2013, Washietl et al., 2007). A recent study investigating relationships between nucleotide conservation and mRNA structure in *E. coli*, yeast, *D. melanogaster*, and mouse models concluded that there was significant conservation in the secondary structure of RNA between species, and predicted that this was the result of evolutionary selection (Gu et al., 2014a).

Another example of this conservation derives from the hypothesis that mRNA structures are less stable near the 5' region of a gene in order to allow for more efficient translation (Griswold et al., 2003). Gu and colleagues examined the genomes of 340 species, including bacteria, archaea, fungi, plants, insects, fishes, birds, and mammals for signs of reduced mRNA secondary structure. They concluded that reduced mRNA structural stability near the start codon is a universal feature of all cellular life (Gu et al., 2010).

In similar fashion, a comprehensive analysis of synonymous codons was used to evaluate gene expression in *E. coli*. From a library of 154 variants of the gene encoding the green fluorescent protein (GFP), random synonymous changes led a 250-fold range of expression in *E. coli*. Although the levels of the GFP mRNA, rates of degradation, and growth rates also varied, changes in mRNA folding near the ribosomal binding site was responsible for the majority of observed variations (Kudla et al., 2009). In a subsequent study, Goodman *et al.*, analyzed synthetic reporter gene expression in *E. coli* and quantified the effects of individual N-terminal codons on gene expression. The authors concluded that the RNA structure, not rare codon usage (discussed later), correlated best with levels of reporter gene expression (Goodman et al., 2013). These observations are consistent with the hypothesis that genes that are not highly expressed favor the formation of resilient mRNA structures and, conversely, high protein expression is facilitated by weak mRNA structural organization around the 5' mRNA region (Griswold et al., 2003).

The examples reviewed above indicate that synonymous changes can alter base pairing and secondary structure of mRNA, with significant effects on gene expression and function. In this context, it should be noted that although numerous computational methods exist to predict the impact of synonymous changes with regard to the secondary structure of mRNA (reviewed in: (Hunt et al., 2014)), the dynamic nature of mRNAs observed both *in vitro* and *in vivo* (Ding et al., 2014, Rouskin et al., 2014) imply that computational prediction methods should be considered only in combination with experimental validation. Both are necessary to best understand the relationships that exist between synonymous nucleotide alteration, mRNA folding, and protein expression. That being said, the lack of experimental methods to analyze large RNA structures has limited studies to confirm changes in mRNA secondary structure. Chemical assays such as selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (Rice et al., 2014, McGinnis et al., 2012, Wilkinson et al., 2006, Merino et al., 2005), structure specific RNase digestion, and circular dichroism (CD) (Bartoszewski et al., 2010) are examples of approaches to confirm predictions regarding to local mRNA

folding. As discussed below, a number of studies have used computational methods in combination with biochemical, cellular and functional assays to establish that changes in the secondary structure of mRNA are an underlying mechanism of altered gene expression following synonymous mutations. Although several downstream effects are described in the following sections, it is important to note that changes in mRNA structural stability do not necessarily alter mRNA decay rates, but may cause either enhanced or reduced mRNA half-life.

2.1.1. Examples of sSNP-associated mRNA folding defects: A study analyzing six different synonymous mutations in a human dopamine receptor gene, *DRD2*, determined that the Pro319-CCC>CCT variant decreased mRNA stability and protein levels and led to reduced dopamine response (Duan et al., 2003). The authors employed the mRNA structure prediction tool “Mfold” (Mathews et al., 2007, Zuker, 2003), and proposed that alterations in the secondary structure of the mRNA caused early decay of the Pro319-CCC>CCT transcript. Interestingly, a separate synonymous mutation, Lys367-AAG>AAA mutation rescued the effects of the Pro319-CCC>CCT change and stabilized the transcript.

More recently Duan *et al.*, analyzed the extent of genome-wide changes in RNA stability in seven human lymphoblastoid cell lines obtained from the HapMap database (hapmap.ncbi.nlm.nih.gov). The group found higher G/C ratios at the third-codon positions (GC3) in genes with altered mRNA structural stability. In these studies, higher GC3 contents corresponded to increased RNA half-life. Their findings led to the hypothesis that the more stable GC base pairing was responsible for increased thermodynamic mRNA stability (Duan et al., 2013).

An excellent study by Nackley and colleagues interrogated the consequences of two synonymous polymorphisms in the human Catechol-O-methyltransferase (*COMT*) gene with regard to pain sensitivity (Nackley et al., 2006). *COMT* is a key regulator of pain, cognitive function, and mood. Haplotypes with synonymous codon variants His62-CAC>CAT and Leu136-CTC>CTG exhibited a marked difference in *COMT* activity, due to reduced protein levels. The authors demonstrated that the most stable mRNA structures were associated with the lower protein levels and enzymatic activity. Diminishing the stability of the mRNA following site-directed mutagenesis led to an increase in protein expression. This study however, did not extend its analysis to any changes in mRNA half-life.

Bartoszewski *et al.* analyzed the synonymous change that is an element of the most frequent disease-causing mutation in the human *CFTR* gene (Bartoszewski et al., 2010). Deletion of three nucleotides led to loss of phenylalanine at the 508 position of the human *CFTR* protein (F508 *CFTR*) and resulted in a synonymous codon change at isoleucine 507 (Ile507-ATC>ATT F508 *CFTR*). Their studies provided biophysical (circular dichroism) and biochemical (structure specific RNase digestion) evidence for alterations in the secondary structure of the mRNA in the vicinity of the mutation. Interestingly, the loss of phenylalanine-508 without the synonymous nucleotide change (Ile507-ATC>ATT) had no measured effect on mRNA folding and produced a structurally and functionally more stable protein. The mRNA secondary structural changes in the Ile507-ATT F508 *CFTR* mRNA were shown to cause altered translational dynamics and reduced protein stability without

altering mRNA half-life. Analysis of the CFTR mRNA near the mutation revealed codon conservation across 43 species. The authors also analyzed three other synonymous mutations that were introduced to optimize plasmid propagation in the most widely used wild type and F508 CFTR expression vectors (Rommens et al., 1991) and found no changes in predicted structure of the mRNAs (Lazrak et al., 2013). These results are consistent with the idea that not every synonymous codon change alters mRNA secondary structure.

These founding studies have drawn our attention to the fact that changes in the secondary structure of mRNA are a likely mechanism by which synonymous mutations contribute to disease. Recent reports analyzing the possible contribution of sSNPs to disease phenotype in different genes appear to support this hypothesis. Examples include studies on sSNPs affecting *PARK7*, a gene defective in Parkinson's disease (Anvret et al., 2011), and the human estrogen receptor alpha (Fernandez-Calero et al., 2014). However, it should be noted, that these studies do not experimentally confirm sSNP-associated changes in mRNA secondary structure.

2.1.2. Example of a sSNP-associated enhanced mRNA stability: Individuals carrying mutations in the gene encoding the alpha subunit of the Gs protein (*GNAS1*) are predisposed to vasovagal syndrome (Lelonek et al., 2008), hypertension, and altered sensitivity to beta-blockers (Jia et al., 1999). Lelonek and colleagues identified a synonymous mutation (Ile131-ATT>ATC) that modifies the target site of endonuclease FokI. The investigators proposed that the sSNP increases the stability of the mRNA, which presents translation increases, production of cAMP, and further enhances PKA-mediated reaction (Lelonek et al., 2008).

3. Synonymous mutations and micro-RNA binding

MicroRNAs (miRNAs) are ~22 nucleotide-long regulatory RNAs expressed in most organisms. Function of miRNAs classically has been attributed to the downregulation of gene expression through the formation of an miRNA-induced silencing complex with Argonaute (miRISCs) that affects the 3' UTR of target mRNAs leading to their decay (for review: (Bartel, 2004, Bartel, 2009)). Subsequently, miRNAs have been shown to repress translation through interference with the initiation step that required the collective function of the 5' cap and the 3' poly(A) tail. The findings pointed to the cap-binding, eukaryotic initiation factor 4E as a molecular target (Humphreys et al., 2005). In contrast, using ribosome profiling, Guo and colleagues demonstrated that in mammalian cells miRNAs act predominantly (84%) through the destabilization of mRNAs (Guo et al., 2010).

Notably, plant miRNAs have been shown to predominantly target coding sequences (Jones-Rhoades et al., 2006). In similar fashion, miRNA target sites have been described in conserved coding regions of mammalian mRNAs (Forman and Collier, 2010, Forman et al., 2008, Ott et al., 2011). For excellent reviews on miRNA function refer to: (Filipowicz et al., 2008, Chekulaeva and Filipowicz, 2009, Fabian et al., 2010). In summary, the function of miRNAs is complex and includes miRNA-mRNA, miRNA-long-noncoding RNA (lncRNA) (Yoon et al., 2014), and miRNA-protein interactions. miRNAs also serve as components of

mRNPs, which participate in miRNA-mediated gene regulation. The multiple, regulatory mechanisms attributed to miRNAs elicit both translational repression and mRNA decay. These processes may be reversible or depend on the cellular localization of miRNA-mediated events (reviewed in: (Fabian et al., 2010, Filipowicz et al., 2008)).

Interestingly, computational analysis has revealed low synonymous substitution rates at several miRNA target sites in human protein coding regions (Hurst, 2006). High throughput approaches using Argonaute-bound targets indicate that the number of miRNA target sites in the coding sequence (CDC) is similar to those in the 3' UTR, but at a lower density (Hafner et al., 2010b, Hafner et al., 2010a). Interestingly, the inhibitory function of miRNAs that bind CDC sequences is less effective (Forman and Collier, 2010, Hafner et al., 2010b). A possible reason for this observation is that coding sequence miRNA binding sites are not accessible if covered by the translational complex or by alterations in the mRNA structure (Bartel, 2009). Thus, miRNAs might have affected synonymous codon choices evolutionarily, since coding sequence-associated miRNA sites appear to be in highly conserved regions of the DNA (Brest et al., 2011, Hurst, 2006, Tay et al., 2008b). In this context it has been proposed that the conservation of miRNA binding sites may be sustained by selection that eliminates deleterious alleles (Guo et al., 2008).

3.1. Examples of sSNPs associated with miRNA defects—To emphasize the significance of CDC targeting miRNAs, it has been experimentally validated that miR-134, miR-296, and miR-470 interact with the Nanog, Oct4, and Sox2 CDC regions, respectively, to modulate mouse embryonic stem cell differentiation (Tay et al., 2008a). Synonymous mutations in these target sites abolish miRNA activity, prevent downregulation of the corresponding genes. This results in a delay in development (Tay et al., 2008a, Tay et al., 2008b).

The biological significance of CDC targeting miRNAs is further supported by studies of Crohn's disease. GWAS analysis identified robust linkage disequilibrium between Crohn's disease and sequence variants of the autophagy-associated immunity-related GTPase family M protein (IRGM) (Parkes et al., 2007). Brest and colleagues have demonstrated that a linked susceptibility locus (Leu105-CTG>TTG) is a binding site for miR-196. When miR-196 is unable to bind, IRGM expression is enhanced leading to misprocessing of pathogens and contribute to Crohn's phenotype (Brest et al., 2011).

In another study, the analysis of 285 melanoma samples through exon sequencing identified a sSNP (Phe17-TTC>TTT) in the *BCL2L12* gene. The synonymous codon change inhibits miR-671-p binding and leads to increased BCL2L12 mRNA and protein levels, as well as increased affinity for p53. Enhanced binding to p53 sequesters BCL2L12, reduces UV-induced apoptosis and p53 target gene transcription (Gartner et al., 2013). As mentioned previously, it is likely that many deleterious sSNPs in miRNA binding regions have been evolutionally eliminated since their interactions are crucial for miRNA function. In addition, nucleotide sequences near miRNA target sites are important for miRNA function (Long et al., 2007) as exemplified in the *IRGM* for miR-196 reviewed above (Brest et al., 2011).

4. The impact of synonymous mutations on translation and protein folding

Mature mRNAs are translated as the ribosome processes along a message (Rodnina and Wintermeyer, 2011). To function, proteins must fold into their native, minimum free energy state (Anfinsen, 1973). The factors that determine protein structure and dynamics of protein folding have been increasingly well characterized, although important questions remain unanswered (reviewed in: (Dill and MacCallum, 2012)). Here we emphasize aspects of protein folding that are impacted by synonymous codon usage.

Analyzing small polypeptides by *in vitro* folding assays, Anfinsen proposed that amino acid sequence was the determinant of three-dimensional structure (Anfinsen, 1973). Although *in vitro* studies have been critical in understanding the fundamental features of protein folding and link experimental, computational, and theoretical approaches, it is also clear that most proteins are unable to fold when removed from the cellular environment (reviewed in: (Braselmann et al., 2013)). Inside cells, efficient protein folding requires the assistance of molecular chaperones to prevent off-pathway configurations and aggregation (Hartl, 1996, Frydman and Hartl, 1996, Ellis and Hartl, 1996). Studies analyzing the folding of polypeptides using *in vitro* translation systems (Komar et al., 1997) and with prokaryotes *in vivo* (Fedorov and Baldwin, 1999), lower eukaryotes (Frydman et al., 1999, Frydman and Hartl, 1996), and mammalian cells (Du and Lukacs, 2009, Kleizen et al., 2005) have determined that proteins reach their final conformation through cotranslational folding and posttranslational domain assembly events. Hence, more recent theories define protein folding as a self-assembly process that is determined not only by thermodynamic, but also by kinetic laws (Whitesides and Grzybowski, 2002).

As discussed below, synonymous codons may impact the kinetics of protein folding. Early reports suggested that mRNA itself provided meaningful information used during protein folding in prokaryotes (Thanaraj and Argos, 1996b) and mammals (Gu et al., 2003, Oresic and Shalloway, 1998, Xie and Ding, 1998). Today, whole-genome sequencing, crystallography, molecular simulations, and structural proteomics, together with theoretical and computational tools of chemical kinetics provide a more global view of protein sequence and structure relationships. These methods provide new means to test the influence of a particular codon on cotranslational protein folding, and evaluate the potential contribution of synonymous mutations to protein biogenesis and conformation (areviewed in: (O'Brien et al., 2014a).

Resources such as the Coding Sequence and Structure database (www.stats.ox.ac.uk/bioinfo/resources/) offer detailed mapping of thousands of protein structures along with their corresponding mRNAs (Saunders and Deane, 2010). Analyzing the sequence and structural relationships in this database led Saunders and colleagues conclude that cotranslational folding leads to a more compact amino-terminus in proteins with combined alpha and beta structures. Studies of this sort have been performed to validate the importance of the primary mRNA sequence with regard to cotranslational folding (Deane and Saunders, 2011).

For example, kinetic modeling was employed by O'Brien and colleagues to investigate relationships between translation speed and cotranslational protein assembly. In their studies, rapid translation through difficult regions was shown to increase domain folding and

reduce misfolding. The authors concluded that processivity, determined by specific codons, can strongly impact cotranslational folding (O'Brien et al., 2014b).

Pechmann and colleagues have demonstrated evolutionary conservation of codon usage in association with specific protein secondary structures. The authors argued that codon usage evolved to govern the rhythm of translation and to facilitate co-translational folding (Pechmann and Frydman, 2013). The finding has significance since proper function of multi-domain proteins is determined through both co-translational folding and post-translational domain assembly (reviewed in: (O'Brien et al., 2014a, Schlebach and Sanders, 2014)). As indicated by these examples, it is clear that redundancy of the genetic code may have enabled evolution to optimize gene expression not only at the level of the amino acid, but also at the level of the mRNA sequence (for review: (Tuller, 2014)).

4.1. Mechanisms by which codons govern translation and protein folding—A

number of features including codon usage (rare vs. frequent codons), tRNA availability, mRNA shape, and ribosome structure each contribute to translational dynamics and influence gene expression. In addition, determining the rhythm of translation appears to be critically important in determining final protein conformation. Studies described in the following sections provide a framework by which we better understand effects that synonymous codon substitutions may have on various steps of gene expression.

4.1.1. The role of ribosomes: Ribosomes play a vital role in translational processes including mRNA unwinding, advancement, selection of cognate tRNAs, and formation of the peptide bond (Rodnina and Wintermeyer, 2009, Rodnina et al., 2007, Wintermeyer et al., 2004). As protein domains are synthesized, they may form secondary structures cotranslationally within the ribosomal exit tunnel of the large subunit and tertiary structures after exit, (Clark, 2004, Kaiser et al., 2011, Thanaraj and Argos, 1996b, Tsai et al., 2008) (Figure 3).

Synonymous codon usage may affect ribosome traffic (Komar et al., 1999), cause translational pausing (reviewed in: (D'Onofrio and Abel, 2014)), or affect the rate at which a peptide emerges from the ribosome tunnel (Sander et al., 2014). It has also been proposed that at protein domain boundaries, codons that are “slow”, encode hydrophobic amino acids that adhere to the ribosome exit channel. In addition, the same portions of the mRNA/peptide requiring slow translation may have complex mRNA secondary structure that modulates translation dynamics (discussed below) (Thanaraj and Argos, 1996a, Thanaraj and Argos, 1996b). Consistent with these ideas, genomic analysis of mitochondrial genes revealed that rare codons (translated more slowly) were preferred for directing hydrophobic amino acids at several positions (Bhardwaj, 2014). Therefore, it is believed that translation speed may be determined in part by rare codons, the biochemical characteristics of encoded amino acids, and the structure of the mRNA. Additionally, frequent codons correspond to more abundant tRNA pools (Ikemura, 1981, Ikemura, 1985, Ikemura and Ozeki, 1983) and it is likely that this also impacts the rate of translation and protein folding (Tuller et al., 2011, Tuller et al., 2010b). Any of these features could be impacted by a synonymous mutation.

Taken together, these data suggest multiple layers of information utilized by the ribosome (carried within in the mRNA sequence), which impact the structure and biochemical features of a protein. Although experimental data linking codon usage and protein folding is rapidly accumulating, multidisciplinary approaches will be necessary to fully understand the underlying mechanism (Tuller, 2014). We discuss reports emphasizing specific aspects by which code degeneracy regulates translation and cotranslational protein folding, but note that these mechanisms likely work in concert and highlight additional robust complexity in the translational machinery.

4.1.2. The role of mRNA secondary structures in protein folding: The nucleotide sequence of mRNAs determines secondary structure, and in turn impacts protein expression (reviewed in: (Mauger et al., 2013). Focusing on translation, Tuller and colleagues have found reduced thermodynamic stability of mRNA secondary-structures near the start codon of both *E. coli* and *S. cerevisiae* (Tuller et al., 2010b). Moreover, in addition to mRNA folding energy, the available tRNA pool and amino acid charging also impact translation rate. These features may independently reduce ribosomal speed at the beginning of an open reading frame (ORF) in a manner that would improve ribosomal allocation (Tuller et al., 2011). New techniques such as genome wide *in vivo* mRNA secondary structure profiling are being used to experimentally validate the significance of RNA structure during gene expression (Ding et al., 2014).

Using elegant experimental systems that monitor translation on single ribosomes, Tinoco and colleagues demonstrated that translation proceeds through translocation-pause cycles (Wen et al., 2008). While the lengths of the pauses were variable, the translocation cycles were generally fast (<0.1 s). Such pauses were suggested to govern proper protein folding *in vivo* (Tinoco and Wen, 2009), and it was noted that mRNA structure also contributes to the length of translational pausing (Wen et al., 2008). Consistent with this finding, Bartoszewski and colleagues demonstrated that synonymous mutations altering mRNA structure also change translation dynamics, (Bartoszewski et al., 2010), leading to altered protein folding and function (Bartoszewski et al., 2010, Lazrak et al., 2013).

4.1.3 The function of rare versus frequent codons during translation and protein folding: In most organisms, highly expressed genes contain a greater percentage of codons corresponding to abundant tRNAs (Ikemura, 1981, Ikemura, 1985, Ikemura and Ozeki, 1983). This improves translational efficiency and contributes to fitness (Berg and Kurland, 1997, Dong et al., 1996). Conversely, low-level gene expression is typically associated with an increased number of rare codons (Plotkin and Kudla, 2011). Both the “rare codon effect” and tRNA availability have been proposed to help explain the mechanisms by which silent mutations alter protein folding and function in *E. coli* (Cortazzo et al., 2002) and human cells (Kimchi-Sarfaty et al., 2007). The mechanisms by which codon usage influence translation and protein folding can be quite variable, as described below.

Codons are translated at various rates (Dana and Tuller, 2014, Spencer et al., 2012, Xu et al., 2013), but rare codons usually translate less rapidly (Marin, 2008, Wu et al., 2004). Consistent with this observation, the “ramp theory” of mRNA translation predicts that rare codons are preferred within the first fifty amino acids of proteins because they function to

slow down elongation and prevent ribosomal crowding (Fredrick and Ibba, 2010, Tuller et al., 2010a). This provides an additional layer of protection for thermodynamically less stable mRNAs (Tuller, 2014, Tuller et al., 2010a, Tuller et al., 2011, Tuller et al., 2010b). In contrast, other studies have implied that positively charged residues, not rare codons, are more critical determinants of translational rate (Charneski and Hurst, 2014).

An alternative mechanism by which rare codons influence cotranslational events was suggested in a recent study investigating translation dynamics of secretory and membrane proteins in yeast (Pechmann et al., 2014). Pechmann and colleagues found that non-optimal codon usage at various sites along the molecule promote interaction with the signal recognition particle (SRP) that targets proteins into the ER membrane (for review on SRP see: (Akopian et al., 2013)). The authors demonstrated that non-optimal codons 35–40 amino acids downstream of an SRP-binding site were important for targeting the peptide to the ER. They emphasized the slow translation of these codons and termed them mRNA-encoded translation ramps (Pechmann et al., 2014). Please note that these “ramps” are not identical to the “ramp theory” that emphasizes slow translation at the beginning of protein synthesis to prevent ribosomal crowding (Fredrick and Ibba, 2010, Tuller et al., 2010a).

Consistent with Pechmann’s idea, studies in *E. coli* that lack SRPs - but contain Shine-Dalgarno sequences to help recruit the ribosome to the mRNA-, have demonstrated that Shine-Dalgarno-like elements (SDLE) are located immediately before, or just downstream of the sequences encoding transmembrane helices of membrane proteins. They suggested that SDLEs serve as SRP-like structures to slow translation. Most significantly, experimental evidence indicates that slow translation improves fidelity and folding after introduction of an experimental SDLE into a transmembrane protein sequence which reduced aggregation upon overexpression (Fluman et al., 2014).

Furthermore, since rare codons often define the boundaries of protein domains (Makhoul and Trifonov, 2002), it has been proposed that this arrangement contributes to the accuracy of co-translational protein folding (Komar and Jaenicke, 1995, Zhang et al., 2005, Adzhubei et al., 1996). Another interesting observation is that alpha helices are usually encoded by rapidly translating, and conversely, beta sheets are preferentially coded by slowly translating codons. Early reports suggested that protein folding requires discontinuous elongation rates and that an appropriate balance between translational delays and acceleration is important (Komar and Jaenicke, 1995, Sanchez et al., 2004, Komar, 2009). Sander and colleagues, recently showed that the rate at which a protein exits the ribosome (governed by the varied translation rate of codons), determine the final folded protein structure (Sander et al., 2014). Importantly, all these data underscore the importance of the primary mRNA sequence in translation and protein folding regulation.

Development of new techniques such as ribosome profiling allows genome wide analysis of translation in cells (Ingolia et al., 2009, Ingolia et al., 2012). Using a modified, pulse-chase approach applied to this technique, Ingolia and colleagues analyzed translation dynamics in mouse embryonic stem cells. They discovered numerous strong translational pause sites and concluded that ribosome elongation rates do not provide a strong correlation with codon frequency when analyzed across the genome (Ingolia et al., 2011). These results challenge

previously summarized work suggesting that neither rare codons, nor charge of the residues influences translation kinetics. However, more recently, Gardin *et al.* modified the ribosome profiling method and developed a novel analytic tool with the predisposition that codon-decoding rates do not strongly account for the variation in ribosome footprints across an mRNA. Their results revealed that frequent and high AT content codons, were decoded more rapidly than rare codons and that proline formed peptide bonds slowly.

These studies demonstrate that specific experimental situations may lead to varying conclusions regarding the impact of codon usage on ribosomal pausing and illustrate the difficulty to analyze translational dynamics at the genome level. Nonetheless, it is clear that understanding the contribution of codon usage to protein structure and function has advanced substantially since Anfinsen's original hypothesis that amino acid sequence itself determines native protein structure (Anfinsen, 1973). There appears to be a measure of agreement that translation begins slowly, and that the rate of translation varies to aid cotranslational folding. However, the precise mechanisms by which these steps occur are still being determined. It is clear that use of rare vs. frequent codons, tRNA availability, mRNA shape, and ribosome structure, may each contribute to translational dynamics and influence gene expression.

4.2. Synonymous mutations that alter protein structure and function and contribute to human disorders—It has been proposed that nsSNPs and sSNPs are equally as likely to be associated with human disorders (Chen *et al.*, 2010). While changes in protein conformation are often suggested as the underlying mechanism of synonymous mutation-associated disorders, studies able to confirm such changes, or protein function differences, are rare (Table 1).

Kimchi-Sarfaty and colleagues analyzed mutations in the multidrug resistance one gene (*MDR1*) and identified a synonymous mutation (Ile1145-ATC>ATT) that led to synthesis of a P-glycoprotein (P-gp) variant with altered drug and inhibitor binding properties. The authors found similar mRNA and protein levels, but the mutant P-gp was misfolded compared to the wild-type P-gp. They proposed that introduction of a rare codon (Ile-ATC vs. Ile-ATT) influenced cotranslational folding thereby altering the structure of both the substrate and inhibitor interaction sites (Kimchi-Sarfaty *et al.*, 2007). In a subsequent study, functional differences between the polymorphic (Ile1145-ATC>ATT) and wild type P-gps were evaluated in polarized epithelial monolayers. Importantly, the mutant P-gp exhibited prolonged recycling times and demonstrated altered response to P-gp inhibitors (Fung *et al.*, 2014).

Bartoszewski and colleagues analyzing a synonymous nucleotide change the human F508 CFTR (Ile507-ATC>ATT) and demonstrated that in addition to altered secondary structure of the mRNA near the mutation (discussed previously), the synonymous change influenced both translation dynamics and protein stability (Bartoszewski *et al.*, 2010). In a subsequent study, same authors compared trafficking and function of the Ile507-ATT and Ile507-ATC

F508 CFTR variants. Using a protein fragment analysis assay they confirmed structural differences between the synonymous variants and found that these protein-folding differences were accompanied by altered CFTR ion channel properties (Lazrak *et al.*, 2013).

Notably, P-gp and CFTR belong to the ABC transporter superfamily of proteins and exhibit complex tertiary structures, attained through co- and posttranslational folding events. The cellular environment also significantly contributes to both final protein conformation and function (for review see: (ter Beek et al., 2014)). Therefore, analyzing the consequences of synonymous mutations in natural biological context, such as performed recently for P-gp (Fung et al., 2014), will be necessary to draw accurate conclusions in future studies.

5. Codon optimization for the production of recombinant proteins

Protein defects are associated with a variety of human disorders. Recombinant peptide and protein therapeutics have become a common approach to treat complex diseases such as diabetes, cancer, viral infections, or hemophilia, among many others (reviewed in: (Carter, 2011, Leader et al., 2008, Spiess et al., 2015, Kimchi-Sarfaty et al., 2013) (El-Baky and Redwan, 2015). Historically these “biologics” were purified from human or animal tissues, but recombinant DNA technology has largely replaced such practices (Kamionka, 2011). To improve recombinant protein yield in heterologous systems, a gene engineering approach that applies synonymous codon variation to increase production, termed codon optimization, has become a common practice (Carter, 2011). Optimized genetic regions include: promoters, mRNA secondary structures, initiation codon environment, internal ribosome entry positions, or splice donor and acceptor sites (Gao et al., 2004, Gaspar et al., 2012, Fath et al., 2011). In addition to the production of recombinant peptides and proteins (Kane, 1995), codon optimization has been applied in nucleic acid therapies, such as gene therapy (Nathwani et al., 2011), mRNA therapy, and DNA/RNA vaccines (reviewed in: (Kane, 1995, Elena et al., 2014, Mauro and Chappell, 2014)).

The first codon optimized peptide, somatostatin, was synthesized in 1977. With the knowledge of only the amino acid sequence, Boyer and colleagues assembled a nucleic acid composed of the most frequently utilized codons by in *E. coli* to produce the 14 amino acid somatostatin fused to beta-galactosidase (Itakura et al., 1977). Based on easy manipulation and well understood genetic and biochemical features of microbial systems, they are widely used for recombinant protein production (reviewed in: (Elena et al., 2014)). Eukaryotic proteins often misfold and aggregate when produced in prokaryotic cells (Makrides, 1996). Therefore, to produce eukaryotic glycoproteins, other expression systems have been developed (Wurm and Bernard, 1999, Wurm, 2004, Meissner et al., 2001). Recombinant proteins intended for human therapeutics, vaccination, or diagnostics are often produced in mammalian cell lines such as the Chinese hamster ovary cell line (CHO) or the human embryonic kidney 293 cell line (HEK293) (Matasci et al., 2008, Alattia et al., 2013). Although codon optimization remains commonly used in prokaryotes (Burgess-Brown et al., 2008, Maertens et al., 2010), it has also been applied to eukaryotic (Kotula and Curtis, 1991, Wiedemann and Boles, 2008), and cell free systems (Kane, 1995, Swartz, 2001).

A growing demand for recombinant protein expression has facilitated development of codon-optimization programs, but with significant differences in the variables that are being emphasized. In addition to the original approach of applying the most frequently used codons by the host (Richardson et al., 2006, Villalobos et al., 2006), other programs incorporate criteria such as natural codon distribution (Gao et al., 2004, Gaspar et al., 2012,

Jayaraj et al., 2005, Raab et al., 2010, Roth et al., 2011, Wu et al., 2006), host tRNA concentrations (Fuglsang, 2003, Qian et al., 2012), or slow translating regions thought to be important for protein folding (Angov et al., 2008). A recent approach termed the “preferred human codon-optimized method” integrates the preferred human codons, with >60% GC content (Inouye et al., 2015). However, as we discussed above, certain protein coding regions such as translation initiation sites, domain boundaries, or catalytic sites of enzymes are extremely sensitive to synonymous codon changes, imposing a challenge to optimize these regions.

The importance of codon optimization as practice is emphasized by the growing demand of biologics (reviewed in: (Katsnelson, 2011, Kimchi-Sarfaty et al., 2013)). One-third of the new drugs approved by the USA Food and Drug Administration (FDA) are therapeutic proteins and most of these are produced from codon optimized genes (Kling, 2011). Notably, it has been observed that codon optimization does not automatically result in improved gene expression (Alexeyev and Winkler, 1999, Curran et al., 2013, Agashe et al., 2013). For example, analysis of 44 synthetic gene products at Blue Heron Biotechnology revealed that only 57% of them were expressed at higher, and 32% of them were expressed at lower levels when “optimized” (<http://www.blueheronbio.com/assets/documents/BlueHeronBioExpressionSurvey.pdf>). Nevertheless, development of codon optimization strategies with multiple variables that incorporate information from studies analyzing translation and cotranslational protein folding are of considerable interest (Kimchi-Sarfaty et al., 2013).

5.1. Examples of codon optimization with beneficial effects—A multi-parameter RNA and codon optimization approach was tested as a means to enhance expression of 50 candidate genes in a mammalian system (Fath et al., 2011). The parameters optimized included: codon choice, increase in GC-content, removal of destabilizing RNA elements and cryptic splice-sites, avoidance of intragenic poly(A)-sites and RNA secondary structure, and deletion of internal ribosomal entry sites. The study evaluated transcription factors, ribosomal and polymerase subunits, protein kinases, membrane proteins, and immune modulators. Interestingly, 86% of the tested genes showed unchanged or elevated expression following manipulation. Importantly, optimization did not affect solubility or functionality of the proteins analyzed. Molecular analysis of the sequence-optimized genes also showed positive transcriptional, translational, and mRNA stability effects. These findings illustrate ways in which multiple elements should be considered when optimizing gene expression *in vitro*.

A recent study introduced condition-specific codon optimization as an approach to recombinant protein synthesis in yeast (Lanza et al., 2014). The authors utilized codon for their codon optimization strategy. Using a green fluorescent protein originally tested *E. coli*, variants were generated that expressed protein constitutively, or at high levels during stationary growth. These studies reveal that it is possible to produce multiple transcripts that encode the same peptide yet have different temporal expression profiles. These examples further emphasize the importance of multi-parameter codon optimization based on the numerous elements of a transcript that may impact translation and ultimately the function of recombinant proteins.

5.2. Examples of harmful or functionally defective codon-optimized biologics

—In 2013, the functional analog of erythropoietin was recalled due to hypersensitivity reactions, including lethal anaphylaxis. Moreover, the development of two recombinant coagulation factor analogs was discontinued in Phase III clinical trials due to safety issues (Kimchi-Sarfaty et al., 2013). Additionally, interferon-alpha has been used in anti-viral and cancer therapies for decades, yet both acute and chronic side effects and immunogenicity remain limiting factors for use of this therapeutic (El-Baky and Redwan, 2015). These examples point to potential limitations of recombinant proteins synthesized following codon modifications for use as human therapeutics.

FUTURE PROSPECTS

As discussed above, codon optimization strategies may produce biologics with reduced function and lead to altered immunogenicity or unknown, harmful biological activity. These findings also illustrate the possible impact of synonymous mutations on protein biosynthesis. Since the expression of unmodified human gene sequences is often impossible using the presently available systems, testing the consequences of codon optimization is crucial prior to clinical application. With rapidly improving analytic tools such as hydroxyl radical labeling and tandem-ion mass spectrometry, that identify conformational differences between protein-variants (Kaur et al., 2015, Scott et al., 2015, Zhang et al., 2011), valuable information in this regard can often be obtained prior to clinical application. Since not all synonymous codon changes impact protein structure or function, identifying biologics with harmful effects and analyzing the underlying molecular mechanism will facilitate production of recombinant proteins with reduced risk to patients in the future.

Additionally, when mutant versions of recombinant genes are constructed for experimental purposes, it is advised to report which synonymous codons have been employed. Unfortunately, this information is often not provided, but in the light of our current understanding of protein biogenesis, it seems vital for proper interpretation of results in publications. Heterologous protein over-expression has provided invaluable information concerning protein biogenesis and function for decades. Therefore, understanding the mechanisms by which synonymous changes may alter protein structure and function should facilitate a more critical evaluation of expression vectors that are employed in these studies.

Development of *in silico* and experimental tools

Because of the difficulty identifying sSNPs with detrimental potential, it is possible that many sSNPs contributing to human disorders are filtered out during data analysis. To address this limitation, the American College of Medical Genetics (ACMG) has developed recommendations for sequence variation interpretation. These guidelines emphasize the limitations of sequence-based techniques and the importance of standardized terminology (2014, Richards et al., 2008). To aid the discovery of potentially significant sSNPs, prediction tools, such as the Silent Variant Analyzer (SilVA), have been developed (Buske et al., 2013). SilVA identifies disease-associated sSNPs by analyzing different genetic and genomic features of single nucleotide variants. These include the relative position in the transcript, CUB in the region, and mRNA folding status. Based on these attributes, SilVA determines a relative harmfulness score of sSNPs.

It is expected that the number and user friendliness of sSNP analysis tools available to assist with identification of potentially harmful sSNPs will increase. Furthermore, collaborative projects such as the human and mouse Consensus Coding Sequence (CCDS) (<http://www.ncbi.nlm.nih.gov/CCDS/>) (Farrell et al., 2014), the mutation position imaging toolbox (MuPIT) (<http://mupit.icm.jhu.edu>), and the NCBI database of SNPs (<http://www.ncbi.nlm.nih.gov/snp>), will each provide important resources to better characterize and study the consequences of synonymous changes.

Salari and colleagues developed remuRNA software that analyzes differences between native and mutant RNA structure (Salari et al., 2013). Using this software, it is possible to interrogate the impact of SNPs on mRNA folding and compare the findings to the effect of random point mutations. The authors observed that common SNP-induced mRNA structural changes are usually less significant than random point mutations resulting silent nucleotide changes (Salari et al., 2013).

Madanecki and colleagues developed a web server termed “RNA Digest” (Madanecki et al., 2014) (http://www.biology.pl/rna_digest). RNA Digest uses RNA structural models obtained through mfold (Zuker, 2003) to predict results of structure specific RNase digestion. The program is suitable for the identification of unique RNA fragments with a maximum length of 200 nucleotides and assists in the design of experiments to confirm changes in mRNA secondary structure.

It is obvious that a growing body of experimental data has facilitated the development of valuable prediction and experimental tools for analysis of synonymous mutations. Application of these tools for specific mutations will be crucial in determining accuracy and promote development of improved methods in this important area.

CONCLUDING REMARKS

Studies during the past 15 years have demonstrated that codon usage plays a crucial role regulating cellular function, and have led to a general understanding of the mechanisms by which these events occur (Tuller, 2014). It is clear that consequences of codon usage have impact at multiple steps in gene expression - including mRNP formation, mRNA structure, translational rate, protein folding, and function. It is also evident that the combination of codon usage, availability of cognate tRNAs, the ribosome unwinding mRNA, the number and length of pauses, and the overall processivity rate during translation all contribute to protein folding. Consequently, both cotranslational and posttranslational protein conformation appear to be strongly influenced by the dynamics of translation as key determinants of final folding state. At a fundamental level, these tasks are governed by mRNAs and gene-specific distribution of synonymous codons.

Acknowledgments

The growing number of publications on synonymous mutation-associated gene expression confirms significance of codon choice. The disciplines that interrogate this question include, but are not limited to: theoretical biology, molecular evolution, computational biology, molecular genetics, systems biology, biochemistry and cell biology, and physiology. We focused here on studies with relevance to synonymous mutation-associated human disorders. Although we made an effort to discuss conflicting ideas and work from various fields, we sincerely apologize to those whose studies are not cited in this review.

The authors thank Dr. Eric Sorscher for review of the manuscript and Amber K. O'Connor PhD of akoWriting LLC for editorial services. Supported by: NIH HL076587 and CFF BEBOK14P0.

ABBREVIATIONS

A	adenine
ACMG	American College of Medical Genetics
Ala	alanine
Asp	aspartic acid
BCL2L12	B-cell lymphoma-2-like protein 12
BRCA1	breast cancer-1 gene
C	cytosine
CCDS	Consensus Coding Sequence
CD	circular dichroism
CDC	coding sequence
cDNA	complementary DNA
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CERES	composite exonic regulatory element of splicing
CFTR	cystic fibrosis transmembrane conductance regulator
CHO	Chinese hamster ovary cell line
COMT	Catechol-O-methyltransferase
CRISPR-Cas9	clustered regularly interspaced short palindromic repeats-caspase 9
CUB	codon usage bias
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
F508	deletion of phenylalanine at position 508
DNA	deoxyribose nucleic acid
DRD2	dopamine receptor D2
<i>E. coli</i>	<i>Escherichia coli</i>
ER	endoplasmic reticulum
ESE	Exonic splicing enhancers
ESS	Exonic splicing silencers
G	guanine
GC3	guanine and cytosine at the third-codon position
GFP	green fluorescent protein
Gly	glycine

GWAS	genome wide association studies
HapMap	haplotype map database
HEK293	human embryonic kidney 293 cell line
HEK293T	derived from HEK293 with stable expression of simian virus-40 large T antigen
HGPRT	hypoxanthine-guanine phosphoryl transferase
His	histidine
Ile-507	isoleucine 507
Ile	isoleucine
IRGM	immunity-related guanosine triphosphatase (GTPase) family M protein
Leu	leucine
LNS	Lesch–Nyhan syndrome
MCAD	medium-chain acyl-CoA dehydrogenase
Mfold	mRNA structure prediction software
miRISCs	miRNA-induced silencing complexes with Argonaute
miRNA	micro-RNA
mRNA	messenger RNA
mRNP	mRNA and protein complex
MuPIT	mutation position imaging toolbox
ncRNA	small non-coding RNA
Nanog	transcription factor named fo the Irish mythology Tír na nÓg
NBD1	nucleotide binding domain 1
NCBI	National Center for Biotechnology Information
Oct4	octamer-binding transcription factor 4
ORF	open reading frame
PARK7	Parkinson protein 7
PDH-E1α	pyruvate dehydrogenase E 1 alpha subunit
Phe	phenylalanine
PKD1	polycystin 1
PKD2	polycystin 2
pre-mRNA	precursor mRNA
Pro	proline

remuRNA	bioinformatics software for measurement of single-nucleotide polymorphism-induced changes of RNA conformation
Arg	arginine
REST	mRNA-encoded translation ramp
RNA	ribonucleic acid
RNase	ribonuclease
Ser	serine
SHAPE assay	Selective 2'-hydroxyl acylation analyzed by primer extension assay
SiVA	Silent Variant Analyzer
SCC	synonymous codon change
SDLE	Shine-Dalgarno-like element
SNP	single nucleotide polymorphism
sSNP	synonymous single nucleotide polymorphism
snRNA	small nuclear RNA
Sox2	sex determining region Y-box 2; transcription factor
SRP	signal recognition particle
T	thymine
TCOF1	TCS-associated gene
TCS	Treacher Collins syndrome
Thr	threonine
tRNA	transfer RNA
U	uracil
3'UTR	three prime untranslated region
5'UTR	five prime untranslated region
Val	valine

REFERENCES CITED

1. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genet Med*.
2. ADZHUBEI AA, ADZHUBEI IA, KRASHENINNIKOV IA, NEIDLE S. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. *FEBS Lett*. 1996; 399:78–82. [PubMed: 8980124]
3. AGASHE D, MARTINEZ-GOMEZ NC, DRUMMOND DA, MARX CJ. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol Biol Evol*. 2013; 30:549–60. [PubMed: 23223712]
4. AKOPIAN D, SHEN K, ZHANG X, SHAN SO. Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem*. 2013; 82:693–721. [PubMed: 23414305]

5. ALATTIA JR, MATASCI M, DIMITROV M, AESCHBACH L, BALASUBRAMANIAN S, HACKER DL, WURM FM, FRAERING PC. Highly efficient production of the Alzheimer's gamma-secretase integral membrane protease complex by a multi-gene stable integration approach. *Biotechnol Bioeng*. 2013; 110:1995–2005. [PubMed: 23359429]
6. ALEXEYEV MF, WINKLER HH. Gene synthesis, bacterial expression and purification of the *Rickettsia prowazekii* ATP/ADP translocase. *Biochim Biophys Acta*. 1999; 1419:299–306. [PubMed: 10407080]
7. ANDERSEN PM, NILSSON P, KERANEN ML, FORSGREN L, HAGGLUND J, KARLSBORG M, RONNEVI LO, GREDAL O, MARKLUND SL. Phenotypic heterogeneity in motor neuron disease patients with CuZn-superoxide dismutase mutations in Scandinavia. *Brain*. 1997; 120 (Pt 10):1723–37. [PubMed: 9365366]
8. ANDERSSON R, GEBHARD C, MIGUEL-ESCALADA I, HOOF I, BORNHOLDT J, BOYD M, CHEN Y, ZHAO X, SCHMIDL C, SUZUKI T, NTINI E, ARNER E, VALEN E, LI K, SCHWARZFISCHER L, GLATZ D, RAITHEL J, LILJE B, RAPIN N, BAGGER FO, JORGENSEN M, ANDERSEN PR, BERTIN N, RACKHAM O, BURROUGHS AM, BAILLIE JK, ISHIZU Y, SHIMIZU Y, FURUHATA E, MAEDA S, NEGISHI Y, MUNGALL CJ, MEEHAN TF, LASSMANN T, ITOH M, KAWAJI H, KONDO N, KAWAI J, LENNARTSSON A, DAUB CO, HEUTINK P, HUME DA, JENSEN TH, SUZUKI H, HAYASHIZAKI Y, MULLER F, CONSORTIUM F, FORREST AR, CARNINCI P, REHLI M, SANDELIN A. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014; 507:455–61. [PubMed: 24670763]
9. ANDERSSON SG, KURLAND CG. Codon preferences in free-living microorganisms. *Microbiol Rev*. 1990; 54:198–210. [PubMed: 2194095]
10. ANFINSEN CB. Principles that govern the folding of protein chains. *Science*. 1973; 181:223–30. [PubMed: 4124164]
11. 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:e2189. [PubMed: 18478103]
12. ANTONARAKIS SE. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat*. 1998; 11:1–3. [PubMed: 9450896]
13. ANVRET A, BLACKINTON JG, WESTERLUND M, RAN C, SYDOW O, WILLOWS T, HAKANSSON A, NISSBRANDT H, BELIN AC. DJ-1 Mutations are Rare in a Swedish Parkinson Cohort. *Open Neurol J*. 2011; 5:8–11. [PubMed: 21532868]
14. ARCHETTI M. Selection on codon usage for error minimization at the protein level. *J Mol Evol*. 2004; 59:400–15. [PubMed: 15553093]
15. BARTEL DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116:281–97. [PubMed: 14744438]
16. BARTEL DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136:215–33. [PubMed: 19167326]
17. BARTOSZEWSKI RA, JABLONSKY M, BARTOSZEWSKA S, STEVENSON L, DAI Q, KAPPES J, COLLAWN JF, BEBOK Z. A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem*. 2010; 285:28741–8. [PubMed: 20628052]
18. BERG OG, KURLAND CG. Growth rate-optimised tRNA abundance and codon usage. *J Mol Biol*. 1997; 270:544–50. [PubMed: 9245585]
19. BHARDWAJ A. Investigating the role of site specific synonymous variation in disease association studies. *Mitochondrion*. 2014; 16:83–8. [PubMed: 24434286]
20. BHATIA V, BARROSO SI, GARCIA-RUBIO ML, TUMINI E, HERRERA-MOYANO E, AGUILERA A. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*. 2014; 511:362–5. [PubMed: 24896180]
21. BILGIN T, KURNAZ IA, WAGNER A. Selection shapes the robustness of ligand-binding amino acids. *J Mol Evol*. 2013; 76:343–9. [PubMed: 23689513]
22. BJORK P, WIESLANDER L. Mechanisms of mRNA export. *Semin Cell Dev Biol*. 2014; 32:47–54. [PubMed: 24813364]

23. BRASELMANN E, CHANEY JL, CLARK PL. Folding the proteome. *Trends Biochem Sci.* 2013; 38:337–44. [PubMed: 23764454]
24. BREST P, LAPAQUETTE P, SOUIDI M, LEBRIGAND K, CESARO A, VOURET-CRAVIARI V, MARI B, BARBRY P, MOSNIER JF, HEBUTERNE X, HAREL-BELLAN A, MOGRABI B, DARFEUILLE-MICHAUD A, HOFMAN P. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet.* 2011; 43:242–5. [PubMed: 21278745]
25. BULMER M. The selection-mutation-drift theory of synonymous codon usage. *Genetics.* 1991; 129:897–907. [PubMed: 1752426]
26. BURGESS-BROWN NA, SHARMA S, SOBOTT F, LOENARZ C, OPPERMAN U, GILEADI O. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr Purif.* 2008; 59:94–102. [PubMed: 18289875]
27. BUSKE OJ, MANICKARAJ A, MITAL S, RAY PN, BRUDNO M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics.* 2013; 29:1843–50. [PubMed: 23736532]
28. BUSTAMANTE CD, NIELSEN R, HARTL DL. A maximum likelihood method for analyzing pseudogene evolution: implications for silent site evolution in humans and rodents. *Mol Biol Evol.* 2002; 19:110–7. [PubMed: 11752196]
29. CALDER VL, DOMIGAN NM, GEORGE PM, DONALDSON IM, WINTERBOURN CC. Superoxide dismutase (glu100-->gly) in a family with inherited motor neuron disease: detection of mutant superoxide dismutase activity and the presence of heterodimers. *Neurosci Lett.* 1995; 189:143–6. [PubMed: 7624031]
30. CARTEGNI L, CHEW SL, KRAINER AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002; 3:285–98. [PubMed: 11967553]
31. CARTER PJ. Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res.* 2011; 317:1261–9. [PubMed: 21371474]
32. CASTLE JC. SNPs occur in regions with less genomic sequence conservation. *PLoS One.* 2011; 6:e20660. [PubMed: 21674007]
33. CHAMARY JV, HURST LD. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* 2005; 6:R75. [PubMed: 16168082]
34. CHAMARY JV, HURST LD. The price of silent mutations. *Sci Am.* 2009; 300:46–53. [PubMed: 19485088]
35. CHAMARY JV, PARMLEY JL, HURST LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet.* 2006; 7:98–108. [PubMed: 16418745]
36. CHARNESKI CA, HURST LD. Positive charge loading at protein termini is due to membrane protein topology, not a translational ramp. *Mol Biol Evol.* 2014; 31:70–84. [PubMed: 24077849]
37. CHEKULAEVA M, FILIPOWICZ W. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol.* 2009; 21:452–60. [PubMed: 19450959]
38. CHEN R, DAVYDOV EV, SIROTA M, BUTTE AJ. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One.* 2010; 5:e13574. [PubMed: 21042586]
39. CHEN SL, LEE W, HOTTES AK, SHAPIRO L, MCADAMS HH. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci U S A.* 2004; 101:3480–5. [PubMed: 14990797]
40. CLARK PL. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci.* 2004; 29:527–34. [PubMed: 15450607]
41. CLAVERIE JM. Gene number. What if there are only 30,000 human genes? *Science.* 2001; 291:1255–7. [PubMed: 11233450]
42. CORTAZZO P, CERVENANSKY C, MARIN M, REISS C, EHRLICH R, DEANA A. Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem Biophys Res Commun.* 2002; 293:537–41. [PubMed: 12054634]
43. CREHALET H, LATOUR P, BONNET V, ATTARIAN S, LABAUGE P, BONELLO N, BERNARD R, MILLAT G, ROUSSON R, BOZON D. U1 snRNA mis-binding: a new cause of CMT1B. *Neurogenetics.* 2010; 11:13–9. [PubMed: 19475438]

44. CRICK, FH. On Degenerate templates and the Adaptor Hypothesis: A Note for the RNA Tie Club. Wellcome Library for the History and Understanding of Medicine; London: 1955.
45. CRICK FH, BARNETT L, BRENNER S, WATTS-TOBIN RJ. General nature of the genetic code for proteins. *Nature*. 1961; 192:1227–32. [PubMed: 13882203]
46. CURRAN KA, LEAVITT JM, KARIM AS, ALPER HS. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab Eng*. 2013; 15:55–66. [PubMed: 23164574]
47. D'ONOFRIO DJ, ABEL DL. Redundancy of the genetic code enables translational pausing. *Front Genet*. 2014; 5:140. [PubMed: 24904640]
48. DANA A, TULLER T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res*. 2014; 42:9171–81. [PubMed: 25056313]
49. DE MEIRLEIR L, LISSENS W, BENELLI C, PONSOT G, DESGUERRE I, MARSAC C, RODRIGUEZ D, SAUDUBRAY JM, POGGI F, LIEBAERS I. Aberrant splicing of exon 6 in the pyruvate dehydrogenase-E1 alpha mRNA linked to a silent mutation in a large family with Leigh's encephalomyelopathy. *Pediatr Res*. 1994; 36:707–12. [PubMed: 7898978]
50. DEANE CM, SAUNDERS R. The imprint of codons on protein structure. *Biotechnol J*. 2011; 6:641–9. [PubMed: 21567957]
51. DILL KA, MACCALLUM JL. The protein-folding problem, 50 years on. *Science*. 2012; 338:1042–6. [PubMed: 23180855]
52. DING Y, TANG Y, KWOK CK, ZHANG Y, BEVILACQUA PC, ASSMANN SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 2014; 505:696–700. [PubMed: 24270811]
53. DONG H, NILSSON L, KURLAND CG. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol*. 1996; 260:649–63. [PubMed: 8709146]
54. DOS REIS M, SAVVA R, WERNISCH L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. 2004; 32:5036–44. [PubMed: 15448185]
55. DU J, DUNGAN SZ, SABOUHANIAN A, CHANG BS. Selection on synonymous codons in mammalian rhodopsins: a possible role in optimizing translational processes. *BMC Evol Biol*. 2014; 14:96. [PubMed: 24884412]
56. DU K, LUKACS GL. Cooperative assembly and misfolding of CFTR domains in vivo. *Mol Biol Cell*. 2009; 20:1903–15. [PubMed: 19176754]
57. DUAN J, SHI J, GE X, DOLKEN L, MOY W, HE D, SHI S, SANDERS AR, ROSS J, GEJMAN PV. Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci Rep*. 2013; 3:1318. [PubMed: 23422947]
58. DUAN J, WAINWRIGHT MS, COMERON JM, SAITOU N, SANDERS AR, GELERNTER J, GEJMAN PV. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet*. 2003; 12:205–16. [PubMed: 12554675]
59. DURET L. Detecting genomic features under weak selective pressure: the example of codon usage in animals and plants. *Bioinformatics*. 2002a; 18(Suppl 2):S91. [PubMed: 12385989]
60. DURET L. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev*. 2002b; 12:640–9. [PubMed: 12433576]
61. DURET L, MOUCHIROUD D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A*. 1999; 96:4482–7. [PubMed: 10200288]
62. DURET L, SEMON M, PIGANEAU G, MOUCHIROUD D, GALTIER N. Vanishing GC-rich isochores in mammalian genomes. *Genetics*. 2002; 162:1837–47. [PubMed: 12524353]
63. EBERLE AB, VISA N. Quality control of mRNP biogenesis: networking at the transcription site. *Semin Cell Dev Biol*. 2014; 32:37–46. [PubMed: 24713468]
64. EL-BAKY NA, REDWAN EM. Therapeutic alpha-interferons protein: structure, production, and biosimilar. *Prep Biochem Biotechnol*. 2015; 45:109–27. [PubMed: 24785737]
65. ELENA C, RAVASI P, CASTELLI ME, PEIRU S, MENZELLA HG. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Front Microbiol*. 2014; 5:21. [PubMed: 24550894]

66. ELLIS RJ, HARTL FU. Protein folding in the cell: competing models of chaperonin function. *FASEB J*. 1996; 10:20–6. [PubMed: 8566542]
67. EOM T, ANTAR LN, SINGER RH, BASSELL GJ. Localization of a beta-actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses. *J Neurosci*. 2003; 23:10433–44. [PubMed: 14614102]
68. FABIAN MR, SONENBERG N, FILIPOWICZ W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem*. 2010; 79:351–79. [PubMed: 20533884]
69. FAIRBROTHER WG, HOLSTE D, BURGE CB, SHARP PA. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol*. 2004; 2:E268. [PubMed: 15340491]
70. FARRELL CM, O'LEARY NA, HARTE RA, LOVELAND JE, WILMING LG, WALLIN C, DIEKHANS M, BARRELL D, SEARLE SM, AKEN B, HIATT SM, FRANKISH A, SUNER MM, RAJPUT B, STEWARD CA, BROWN GR, BENNETT R, MURPHY M, WU W, KAY MP, HART J, RAJAN J, WEBER J, SNOW C, RIDDICK LD, HUNT T, WEBB D, THOMAS M, TAMEZ P, RANGWALA SH, MCGARVEY KM, PUJAR S, SHKEDA A, MUDGE JM, GONZALEZ JM, GILBERT JG, TREVANION SJ, BAERTSCH R, HARROW JL, HUBBARD T, OSTELL JM, HAUSSLER D, PRUITT KD. Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res*. 2014; 42:D865–72. [PubMed: 24217909]
71. FATH S, BAUER AP, LISS M, SPRIESTERSBACH A, MAERTENS B, HAHN P, LUDWIG C, SCHAFER F, GRAF M, WAGNER R. Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLoS One*. 2011; 6:e17596. [PubMed: 21408612]
72. FAUSTINO NA, COOPER TA. Pre-mRNA splicing and human disease. *Genes Dev*. 2003; 17:419–37. [PubMed: 12600935]
73. FEDOROV AN, BALDWIN TO. Process of biosynthetic protein folding determines the rapid formation of native structure. *J Mol Biol*. 1999; 294:579–86. [PubMed: 10610781]
74. FERNANDEZ-CALERO T, ASTRADA S, ALBERTI A, HORJALES S, ARNAL JF, ROVIRA C, BOLLATI-FOGOLIN M, FLOURIOT G, MARIN M. The transcriptional activities and cellular localization of the human estrogen receptor alpha are affected by the synonymous Ala87 mutation. *J Steroid Biochem Mol Biol*. 2014; 143:99–104. [PubMed: 24607813]
75. FERRE-D'AMARE AR. Use of a coenzyme by the glmS ribozyme-riboswitch suggests primordial expansion of RNA chemistry by small molecules. *Philos Trans R Soc Lond B Biol Sci*. 2011; 366:2942–8. [PubMed: 21930586]
76. FILIPOWICZ W, BHATTACHARYYA SN, SONENBERG N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*. 2008; 9:102–14. [PubMed: 18197166]
77. FLUMAN N, NAVON S, BIBI E, PILPEL Y. mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. *Elife*. 2014; 3
78. FORMAN JJ, COLLIER HA. The code within the code: microRNAs target coding regions. *Cell Cycle*. 2010; 9:1533–41. [PubMed: 20372064]
79. FORMAN JJ, LEGESSE-MILLER A, COLLIER HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A*. 2008; 105:14879–84. [PubMed: 18812516]
80. FREDRICK K, IBBA M. How the sequence of a gene can tune its translation. *Cell*. 2010; 141:227–9. [PubMed: 20403320]
81. FRIEDA KL, BLOCK SM. Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*. 2012; 338:397–400. [PubMed: 23087247]
82. FRYDMAN J, ERDJUMENT-BROMAGE H, TEMPST P, HARTL FU. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*. 1999; 6:697–705. [PubMed: 10404229]
83. FRYDMAN J, HARTL FU. Principles of chaperone-assisted protein folding: differences between in vitro and in vivo mechanisms. *Science*. 1996; 272:1497–502. [PubMed: 8633246]
84. FUGLSANG A. Codon optimizer: a freeware tool for codon optimization. *Protein Expr Purif*. 2003; 31:247–9. [PubMed: 14550643]

85. FUNG KL, PAN J, OHNUMA S, LUND PE, PIXLEY JN, KIMCHI-SARFATY C, AMBUDKAR SV, GOTTESMAN MM. MDR1 synonymous polymorphisms alter transporter specificity and protein stability in a stable epithelial monolayer. *Cancer Res.* 2014; 74:598–608. [PubMed: 24305879]
86. GAO W, RZEWSKI A, SUN H, ROBBINS PD, GAMBOTTO A. UpGene: Application of a web-based DNA codon optimization algorithm. *Biotechnol Prog.* 2004; 20:443–8. [PubMed: 15058988]
87. GARTNER JJ, PARKER SC, PRICKETT TD, DUTTON-REGESESTER K, STITZEL ML, LIN JC, DAVIS S, SIMHADRI VL, JHA S, KATAGIRI N, GOTEV V, TEER JK, WEI X, MORKEN MA, BHANOT UK, PROGRAM NCS, CHEN G, ELNITSKI LL, DAVIES MA, GERSHENWALD JE, CARTER H, KARCHIN R, ROBINSON W, ROBINSON S, ROSENBERG SA, COLLINS FS, PARMIGIANI G, KOMAR AA, KIMCHI-SARFATY C, HAYWARD NK, MARGULIES EH, SAMUELS Y. Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc Natl Acad Sci U S A.* 2013; 110:13481–6. [PubMed: 23901115]
88. GASPAR P, OLIVEIRA JL, FROMMLET J, SANTOS MA, MOURA G. EuGene: maximizing synthetic gene design for heterologous expression. *Bioinformatics.* 2012; 28:2683–4. [PubMed: 22847936]
89. GE WW, LEYSTRA-LANTZ C, SANELLI TR, MCLEAN J, WEN W, STRONG W, STRONG MJ. Neuronal tissue-specific ribonucleoprotein complex formation on SOD1 mRNA: alterations by ALS SOD1 mutations. *Neurobiol Dis.* 2006; 23:342–50. [PubMed: 16730180]
90. GONZALEZ-PAREDES FJ, RAMOS-TRUJILLO E, CLAVERIE-MARTIN F. Defective pre-mRNA splicing in PKD1 due to presumed missense and synonymous mutations causing autosomal dominant polycystic disease. *Gene.* 2014; 546:243–9. [PubMed: 24907393]
91. GOODMAN DB, CHURCH GM, KOSURI S. Causes and effects of N-terminal codon bias in bacterial genes. *Science.* 2013; 342:475–9. [PubMed: 24072823]
92. GOTT JM, EMESON RB. Functions and mechanisms of RNA editing. *Annu Rev Genet.* 2000; 34:499–531. [PubMed: 11092837]
93. GRISWOLD KE, MAHMOOD NA IV, ERSON BL, GEORGIOU G. Effects of codon usage versus putative 5'-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein Expr Purif.* 2003; 27:134–42. [PubMed: 12509995]
94. GU W, LI M, XU Y, WANG T, KO JH, ZHOU T. The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes. *BMC Evol Biol.* 2014a; 14:87. [PubMed: 24758737]
95. GU W, XU Y, XIE X, WANG T, KO JH, ZHOU T. The role of RNA structure at 5' untranslated region in microRNA-mediated gene regulation. *RNA.* 2014b; 20:1369–75. [PubMed: 25002673]
96. GU W, ZHOU T, MA J, SUN X, LU Z. Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans Nanobioscience.* 2003; 2:150–7. [PubMed: 15376949]
97. GU W, ZHOU T, WILKE CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol.* 2010; 6:e1000664. [PubMed: 20140241]
98. GUO H, INGOLIA NT, WEISSMAN JS, BARTEL DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 2010; 466:835–40. [PubMed: 20703300]
99. GUO X, GUI Y, WANG Y, ZHU QH, HELLIWELL C, FAN L. Selection and mutation on microRNA target sequences during rice evolution. *BMC Genomics.* 2008; 9:454. [PubMed: 18831738]
100. HAFNER M, LANDTHALER M, BURGER L, KHORSHID M, HAUSSEER J, BERNINGER P, ROTHBALLER A, ASCANO M JR, JUNGKAMP AC, MUNSCHAUER M, ULRICH A, WARDLE GS, DEWELL S, ZAVOLAN M, TUSCHL T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell.* 2010a; 141:129–41. [PubMed: 20371350]
101. HAFNER M, LANDTHALER M, BURGER L, KHORSHID M, HAUSSEER J, BERNINGER P, ROTHBALLER A, ASCANO M, JUNGKAMP AC, MUNSCHAUER M, ULRICH A,

- WARDLE GS, DEWELL S, ZAVOLAN M, TUSCHL T. PAR-CLiP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp*. 2010b
102. HARTL FU. Molecular chaperones in cellular protein folding. *Nature*. 1996; 381:571–9. [PubMed: 8637592]
103. HELLMANN I, ZOLLNER S, ENARD W, EBERSBERGER I, NICKEL B, PAABO S. Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res*. 2003; 13:831–7. [PubMed: 12727903]
104. HENSE W, ANDERSON N, HUTTER S, STEPHAN W, PARSCH J, CARLINI DB. Experimentally increased codon bias in the *Drosophila* Adh gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. *Genetics*. 2010; 184:547–55. [PubMed: 19966063]
105. HERSHBERG R, PETROV DA. Selection on codon bias. *Annu Rev Genet*. 2008; 42:287–99. [PubMed: 18983258]
106. HOSLER BA, NICHOLSON GA, SAPP PC, CHIN W, ORRELL RW, DE BELLEROCHE JS, ESTEBAN J, HAYWARD LJ, MCKENNA-YASEK D, YEUNG L, CHERRYSON AK, DENCH JE, WILTON SD, LAING NG, HORVITZ HR, BROWN RH JR. Three novel mutations and two variants in the gene for Cu/Zn superoxide dismutase in familial amyotrophic lateral sclerosis. *Neuromuscul Disord*. 1996; 6:361–6. [PubMed: 8938700]
107. HUERTAS P, AGUILERA A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell*. 2003; 12:711–21. [PubMed: 14527416]
108. HUMPHREYS DT, WESTMAN BJ, MARTIN DI, PREISS T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A*. 2005; 102:16961–6. [PubMed: 16287976]
109. HUNT RC, SIMHADRI VL, IANDOLI M, SAUNA ZE, KIMCHI-SARFATY C. Exposing synonymous mutations. *Trends Genet*. 2014; 30:308–21. [PubMed: 24954581]
110. HURST LD. Preliminary assessment of the impact of microRNA-mediated regulation on coding sequence evolution in mammals. *J Mol Evol*. 2006; 63:174–82. [PubMed: 16786435]
111. HURST LD, PAL C. Evidence for purifying selection acting on silent sites in BRCA1. *Trends Genet*. 2001; 17:62–5. [PubMed: 11173101]
112. IKEMURA T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*. 1981; 151:389–409. [PubMed: 6175758]
113. IKEMURA T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985; 2:13–34. [PubMed: 3916708]
114. IKEMURA T, OZEKI H. Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring Harb Symp Quant Biol*. 1983; 47(Pt 2):1087–97. [PubMed: 6345068]
115. INGOLIA NT, BRAR GA, ROUSKIN S, MCGEACHY AM, WEISSMAN JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc*. 2012; 7:1534–50. [PubMed: 22836135]
116. INGOLIA NT, GHAEMMAGHAMI S, NEWMAN JR, WEISSMAN JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324:218–23. [PubMed: 19213877]
117. INGOLIA NT, LAREAU LF, WEISSMAN JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147:789–802. [PubMed: 22056041]
118. INOUE S, SAHARA-MIURA Y, SATO JI, SUZUKI T. Codon optimization of genes for efficient protein expression in mammalian cells by selection of only preferred human codons. *Protein Expr Purif*. 2015; 109C:47–54. [PubMed: 25665506]
119. ITAKURA K, HIROSE T, CREA R, RIGGS AD, HEYNEKER HL, BOLIVAR F, BOYER HW. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*. 1977; 198:1056–63. [PubMed: 412251]

120. JABBARI K, BERNARDI G. The distribution of genes in the *Drosophila* genome. *Gene*. 2000; 247:287–92. [PubMed: 10773468]
121. JACKSON M, AL-CHALABI A, ENAYAT ZE, CHIOZA B, LEIGH PN, MORRISON KE. Copper/zinc superoxide dismutase 1 and sporadic amyotrophic lateral sclerosis: analysis of 155 cases and identification of a novel insertion mutation. *Ann Neurol*. 1997; 42:803–7. [PubMed: 9392581]
122. JAMBOR H, MUELLER S, BULLOCK SL, EPHRUSSI A. A stem-loop structure directs oskar mRNA to microtubule minus ends. *RNA*. 2014; 20:429–39. [PubMed: 24572808]
123. JAYARAJ S, REID R, SANTI DV. GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res*. 2005; 33:3011–6. [PubMed: 15911632]
124. JIA H, HINGORANI AD, SHARMA P, HOPPER R, DICKERSON C, TRUTWEIN D, LLOYD DD, BROWN MJ. Association of the G(s)alpha gene with essential hypertension and response to beta-blockade. *Hypertension*. 1999; 34:8–14. [PubMed: 10406816]
125. JONES-RHOADES MW, BARTEL DP, BARTEL B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol*. 2006; 57:19–53. [PubMed: 16669754]
126. KAISER CM, GOLDMAN DH, CHODERA JD, TINOCO I JR, BUSTAMANTE C. The ribosome modulates nascent protein folding. *Science*. 2011; 334:1723–7. [PubMed: 22194581]
127. KAMIONKA M. Engineering of therapeutic proteins production in *Escherichia coli*. *Curr Pharm Biotechnol*. 2011; 12:268–74. [PubMed: 21050165]
128. KANAYA S, KUDO Y, NAKAMURA Y, IKEMURA T. Detection of genes in *Escherichia coli* sequences determined by genome projects and prediction of protein production levels, based on multivariate diversity in codon usage. *Comput Appl Biosci*. 1996; 12:213–25. [PubMed: 8872390]
129. KANE JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol*. 1995; 6:494–500. [PubMed: 7579660]
130. KATSNELSON A. Breaking the silence. *Nat Med*. 2011; 17:1536–8. [PubMed: 22146444]
131. 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. *Mol Cell Proteomics*. 2015
132. KERTESZ M, WAN Y, MAZOR E, RINN JL, NUTTER RC, CHANG HY, SEGAL E. Genome-wide measurement of RNA secondary structure in yeast. *Nature*. 2010; 467:103–7. [PubMed: 20811459]
133. KIMCHI-SARFATY C, OH JM, KIM IW, SAUNA ZE, CALCAGNO AM, AMBUDKAR SV, GOTTESMAN MM. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science*. 2007; 315:525–8. [PubMed: 17185560]
134. 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; 34:534–48. [PubMed: 24060103]
135. KIMURA M. Evolutionary rate at the molecular level. *Nature*. 1968a; 217:624–6. [PubMed: 5637732]
136. KIMURA M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet Res*. 1968b; 11:247–69. [PubMed: 5713805]
137. KLEIZEN B, VAN VLIJMEN T, DE JONGE HR, BRAAKMAN I. Folding of CFTR is predominantly cotranslational. *Mol Cell*. 2005; 20:277–87. [PubMed: 16246729]
138. KLING J. Fresh from the biologic pipeline--2010. *Nat Biotechnol*. 2011; 29:197–200. [PubMed: 21390016]
139. KOMAR AA. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci*. 2009; 34:16–24. [PubMed: 18996013]
140. KOMAR AA, JAENICKE R. Kinetics of translation of gamma B crystallin and its circularly permuted variant in an in vitro cell-free system: possible relations to codon distribution and protein folding. *FEBS Lett*. 1995; 376:195–8. [PubMed: 7498540]
141. KOMAR AA, KOMMER A, KRASHENINNIKOV IA, SPIRIN AS. Cotranslational folding of globin. *J Biol Chem*. 1997; 272:10646–51. [PubMed: 9099713]

142. KOMAR AA, LESNIK T, REISS C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* 1999; 462:387–91. [PubMed: 10622731]
143. KORNBLIHTT AR, DE LA MATA M, FEDEDA JP, MUNOZ MJ, NOGUES G. Multiple links between transcription and splicing. *RNA.* 2004; 10:1489–98. [PubMed: 15383674]
144. KOTULA L, CURTIS PJ. Evaluation of foreign gene codon optimization in yeast: expression of a mouse IG kappa chain. *Biotechnology (N Y).* 1991; 9:1386–9. [PubMed: 1367771]
145. KRAWCZAK M, REISS J, COOPER DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet.* 1992; 90:41–54. [PubMed: 1427786]
146. KUDLA G, MURRAY AW, TOLLERVEY D, PLOTKIN JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science.* 2009; 324:255–8. [PubMed: 19359587]
147. LANZA AM, CURRAN KA, REY LG, ALPER HS. A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. *BMC Syst Biol.* 2014; 8:33. [PubMed: 24636000]
148. LAZRAC A, FU L, BALI V, BARTOSZEWSKI R, RAB A, HAVASI V, KEILES S, KAPPES J, KUMAR R, LEFKOWITZ E, SORSCHER EJ, MATALON S, COLLAWN JF, BEBOK Z. The silent codon change I507-ATC->ATT contributes to the severity of the DeltaF508 CFTR channel dysfunction. *FASEB J.* 2013; 27:4630–45. [PubMed: 23907436]
149. LEADER B, BACA QJ, GOLAN DE. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov.* 2008; 7:21–39. [PubMed: 18097458]
150. LELONEK M, PIETRUCHA T, MATYJASZCZYK M, GOCH JH. Mutation T/C,Ile 131 of the gene encoding the alfa subunit of the human Gs protein and predisposition to vasovagal syncope. *Circ J.* 2008; 72:558–62. [PubMed: 18362425]
151. LERCHER MJ, SMITH NG, EYRE-WALKER A, HURST LD. The evolution of isochores: evidence from SNP frequency distributions. *Genetics.* 2002; 162:1805–10. [PubMed: 12524350]
152. LEVINE M, CATTOGLIO C, TJIAN R. Looping back to leap forward: transcription enters a new era. *Cell.* 2014; 157:13–25. [PubMed: 24679523]
153. LIU HX, ZHANG M, KRAINER AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* 1998; 12:1998–2012. [PubMed: 9649504]
154. LONG D, LEE R, WILLIAMS P, CHAN CY, AMBROS V, DING Y. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol.* 2007; 14:287–94. [PubMed: 17401373]
155. LONG M, GILLESPIE JH. Codon usage divergence of homologous vertebrate genes and codon usage clock. *J Mol Evol.* 1991; 32:6–15. [PubMed: 1901369]
156. LUCKS JB, NELSON DR, KUDLA GR, PLOTKIN JB. Genome landscapes and bacteriophage codon usage. *PLoS Comput Biol.* 2008; 4:e1000001. [PubMed: 18463708]
157. MACAYA D, KATSANIS SH, HEFFERON TW, AUDLIN S, MENDELSON NJ, ROGGENBUCK J, CUTTING GR. A synonymous mutation in TCOF1 causes Treacher Collins syndrome due to mis-splicing of a constitutive exon. *Am J Med Genet A.* 2009; 149A:1624–7. [PubMed: 19572402]
158. MADANECKI P, NOZELL S, OCHOCKA R, COLLAWN JF, BARTOSZEWSKI R. RNAdigest: a web-based tool for the analysis and prediction of structure-specific RNase digestion results. *PLoS One.* 2014; 9:e96759. [PubMed: 24801507]
159. MAERTENS B, SPRIESTERSBACH A, VON GROLL U, ROTH U, KUBICEK J, GERRITS M, GRAF M, LISS M, DAUBERT D, WAGNER R, SCHAFFER F. Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. *Protein Sci.* 2010; 19:1312–26. [PubMed: 20506237]
160. MAJEWSKI J, OTT J. Distribution and characterization of regulatory elements in the human genome. *Genome Res.* 2002; 12:1827–36. [PubMed: 12466286]
161. MAKHOUL CH, TRIFONOV EN. Distribution of rare triplets along mRNA and their relation to protein folding. *J Biomol Struct Dyn.* 2002; 20:413–20. [PubMed: 12437379]
162. MAKRIDES SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev.* 1996; 60:512–38. [PubMed: 8840785]

163. MANDAL M, BREAKER RR. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol.* 2004; 5:451–63. [PubMed: 15173824]
164. MARIN M. Folding at the rhythm of the rare codon beat. *Biotechnol J.* 2008; 3:1047–57. [PubMed: 18624343]
165. MARTIN KC, EPHRUSSI A. mRNA localization: gene expression in the spatial dimension. *Cell.* 2009; 136:719–30. [PubMed: 19239891]
166. MARTIN RG, MATTHAEI JH, JONES OW, NIRENBERG MW. Ribonucleotide composition of the genetic code. *Biochem Biophys Res Commun.* 1962; 6:410–4. [PubMed: 14470439]
167. MATASCI M, HACKER DL, BALDI L, WURM FM. Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discov Today Technol.* 2008; 5:e37–42. [PubMed: 24981089]
168. MATERA AG, WANG Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol.* 2014; 15:108–21. [PubMed: 24452469]
169. MATHEWS DH, TURNER DH, ZUKER M. RNA secondary structure prediction. *Curr Protoc Nucleic Acid Chem.* 2007; Chapter 11(Unit 11):2. [PubMed: 18428968]
170. MAUGER DM, SIEGFRIED NA, WEEKS KM. The genetic code as expressed through relationships between mRNA structure and protein function. *FEBS Lett.* 2013; 587:1180–8. [PubMed: 23499436]
171. MAURO VP, CHAPPELL SA. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med.* 2014; 20:604–13. [PubMed: 25263172]
172. MAUTE RL, DALLA-FAVERA R, BASSO K. RNAs with multiple personalities. *Wiley Interdiscip Rev RNA.* 2014; 5:1–13. [PubMed: 24039180]
173. MAYR E. Cause and effect in biology. *Science.* 1961; 134:1501–6. [PubMed: 14471768]
174. MAYR E. Evolutionary challenges to the mathematical interpretation of evolution. *Wistar Inst Symp Monogr.* 1967; 5:47–58. [PubMed: 6051130]
175. MCGINNIS JL, DUNKLE JA, CATE JH, WEEKS KM. The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc.* 2012; 134:6617–24. [PubMed: 22475022]
176. MEIJER HA, KONG YW, LU WT, WILCZYNSKA A, SPRIGGS RV, ROBINSON SW, GODFREY JD, WILLIS AE, BUSHELL M. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science.* 2013; 340:82–5. [PubMed: 23559250]
177. MEISSNER P, PICK H, KULANGARA A, CHATELLARD P, FRIEDRICH K, WURM FM. Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnol Bioeng.* 2001; 75:197–203. [PubMed: 11536142]
178. MERINO EJ, WILKINSON KA, COUGHLAN JL, WEEKS KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc.* 2005; 127:4223–31. [PubMed: 15783204]
179. MITCHELL SF, PARKER R. Principles and properties of eukaryotic mRNPs. *Mol Cell.* 2014; 54:547–58. [PubMed: 24856220]
180. MORTIMER SA, KIDWELL MA, DOUDNA JA. Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet.* 2014; 15:469–79. [PubMed: 24821474]
181. MULLER M, NEUGEBAUER KM, ECKMANN C. mRNA: a complex(ed) life. *Genome Biol.* 2010; 11:304. [PubMed: 20500909]
182. MULLER-MCNICOLL M, NEUGEBAUER KM. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet.* 2013; 14:275–87. [PubMed: 23478349]
183. NACKLEY AG, SHABALINA SA, TCHIVILEVA IE, SATTERFIELD K, KORCHYNSKYI O, MAKAROV SS, MAIXNER W, DIATCHENKO L. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science.* 2006; 314:1930–3. [PubMed: 17185601]
184. NATHWANI AC, TUDDENHAM EG, RANGARAJAN S, ROSALES C, MCINTOSH J, LINC DC, CHOWDARY P, RIDDELL A, PIE AJ, HARRINGTON C, O'BEIRNE J, SMITH K, PASI J, GLADER B, RUSTAGI P, NG CY, KAY MA, ZHOU J, SPENCE Y, MORTON CL, ALLAY J, COLEMAN J, SLEEP S, CUNNINGHAM JM, SRIVASTAVA D, BASNER-TSCHAKARJAN E, MINGOZZI F, HIGH KA, GRAY JT, REISS UM, NIENHUIS AW,

DAVIDOFF AM. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med*. 2011; 365:2357–65. [PubMed: 22149959]

185. NEI M, NOZAWA M. Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. *Genome Biol Evol*. 2011; 3:812–29. [PubMed: 21903731]
186. NEI M, SUZUKI Y, NOZAWA M. The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet*. 2010; 11:265–89. [PubMed: 20565254]
187. NIELSEN KB, SORENSEN S, CARTEGNI L, CORYDON TJ, DOKTOR TK, SCHROEDER LD, REINERT LS, ELPELEG O, KRAINER AR, GREGERSEN N, KJEMS J, ANDRESEN BS. Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am J Hum Genet*. 2007; 80:416–32. [PubMed: 17273963]
188. NIRENBERG M. Protein synthesis and the RNA code. *Harvey Lect*. 1965; 59:155–85. [PubMed: 5338236]
189. NIRENBERG MW. The genetic code. II. *Sci Am*. 1963; 208:80–94. [PubMed: 13938751]
190. O'BRIEN EP, CIRYAM P, VENDRUSCOLO M, DOBSON CM. Understanding the influence of codon translation rates on cotranslational protein folding. *Acc Chem Res*. 2014a; 47:1536–44. [PubMed: 24784899]
191. O'BRIEN EP, VENDRUSCOLO M, DOBSON CM. Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat Commun*. 2014b; 5:2988. [PubMed: 24394622]
192. ORESIC M, SHALLOWAY D. Specific correlations between relative synonymous codon usage and protein secondary structure. *J Mol Biol*. 1998; 281:31–48. [PubMed: 9680473]
193. OTT CE, GRUNHAGEN J, JAGER M, HORBELT D, SCHWILL S, KALLENBACH K, GUO G, MANKE T, KNAUS P, MUNDLOS S, ROBINSON PN. MicroRNAs differentially expressed in postnatal aortic development downregulate elastin via 3' UTR and coding-sequence binding sites. *PLoS One*. 2011; 6:e16250. [PubMed: 21305018]
194. PAGANI F, RAPONI M, BARALLE FE. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci U S A*. 2005; 102:6368–72. [PubMed: 15840711]
195. PAGANI F, STUANI C, TZETIS M, KANAVAKIS E, EFTHYMIADOU A, DOUDOUNAKIS S, CASALS T, BARALLE FE. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet*. 2003; 12:1111–20. [PubMed: 12719375]
196. PARKES M, BARRETT JC, PRESCOTT NJ, TREMELLING M, ANDERSON CA, FISHER SA, ROBERTS RG, NIMMO ER, CUMMINGS FR, SOARS D, DRUMMOND H, LEES CW, KHAWAJA SA, BAGNALL R, BURKE DA, TODHUNTER CE, AHMAD T, ONNIE CM, MCARDLE W, STRACHAN D, BETHEL G, BRYAN C, LEWIS CM, DELOUKAS P, FORBES A, SANDERSON J, JEWELL DP, SATSANGI J, MANSFIELD JC, CARDON L, MATHEW CG. WELLCOME TRUST CASE CONTROL C. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*. 2007; 39:830–2. [PubMed: 17554261]
197. PARMELEY JL, CHAMARY JV, HURST LD. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol Biol Evol*. 2006; 23:301–9. [PubMed: 16221894]
198. PASCALE A, GOVONI S. The complex world of post-transcriptional mechanisms: is their deregulation a common link for diseases? Focus on ELAV-like RNA-binding proteins. *Cell Mol Life Sci*. 2012; 69:501–17. [PubMed: 21909784]
199. PECHMANN S, CHARTRON JW, FRYDMAN J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol*. 2014; 21:1100–5. [PubMed: 25420103]
200. PECHMANN S, FRYDMAN J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. 2013; 20:237–43. [PubMed: 23262490]

201. PEDERSEN JS, BEJERANO G, SIEPEL A, ROSENBLOOM K, LINDBLAD-TOH K, LANDER ES, KENT J, MILLER W, HAUSSLER D. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol*. 2006; 2:e33. [PubMed: 16628248]
202. PLOTKIN JB, KUDLA G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011; 12:32–42. [PubMed: 21102527]
203. PLOTKIN JB, ROBINS H, LEVINE AJ. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci U S A*. 2004; 101:12588–91. [PubMed: 15314228]
204. QIAN W, YANG JR, PEARSON NM, MACLEAN C, ZHANG J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet*. 2012; 8:e1002603. [PubMed: 22479199]
205. RAAB D, GRAF M, NOTKA F, SCHODL T, WAGNER R. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst Synth Biol*. 2010; 4:215–25. [PubMed: 21189842]
206. REED R. Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*. 2003; 15:326–31. [PubMed: 12787775]
207. RICE GM, BUSAN S, KARABIBER F, FAVOROV OV, WEEKS KM. SHAPE Analysis of Small RNAs and Riboswitches. *Methods Enzymol*. 2014; 549:165–87. [PubMed: 25432749]
208. RICHARDS CS, BALE S, BELLISSIMO DB, DAS S, GRODY WW, HEGDE MR, LYON E, WARD BE. MOLECULAR SUBCOMMITTEE OF THE ALQAC. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008; 10:294–300. [PubMed: 18414213]
209. RICHARDSON SM, WHEELAN SJ, YARRINGTON RM, BOEKE JD. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res*. 2006; 16:550–6. [PubMed: 16481661]
210. RIEDER LE, REENAN RA. The intricate relationship between RNA structure, editing, and splicing. *Semin Cell Dev Biol*. 2012; 23:281–8. [PubMed: 22178616]
211. RODNINA MV, BERINGER M, WINTERMEYER W. How ribosomes make peptide bonds. *Trends Biochem Sci*. 2007; 32:20–6. [PubMed: 17157507]
212. RODNINA MV, WINTERMEYER W. Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol*. 2009; 21:435–43. [PubMed: 19243929]
213. RODNINA MV, WINTERMEYER W. The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation. *Biochem Soc Trans*. 2011; 39:658–62. [PubMed: 21428957]
214. ROMMENS JM, DHO S, BEAR CE, KARTNER N, KENNEDY D, RIORDAN JR, TSUI LC, FOSKETT JK. cAMP-inducible chloride conductance in mouse fibroblast lines stably expressing the human cystic fibrosis transmembrane conductance regulator. *Proc Natl Acad Sci U S A*. 1991; 88:7500–4. [PubMed: 1715567]
215. ROTH A, BOESS F, LANDES C, STEINER G, FREICHEL C, PLANCHER JM, RAAB S, DE VERA MUDRY C, WEISER T, SUTER L. Gene expression-based in vivo and in vitro prediction of liver toxicity allows compound selection at an early stage of drug development. *J Biochem Mol Toxicol*. 2011; 25:183–94. [PubMed: 21671310]
216. ROUSKIN S, ZUBRADT M, WASHIETL S, KELLIS M, WEISSMAN JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 2014; 505:701–5. [PubMed: 24336214]
217. SALARI R, KIMCHI-SARFATY C, GOTTESMAN MM, PRZYTYCKA TM. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic Acids Res*. 2013; 41:44–53. [PubMed: 23125360]
218. SANCHEZ IE, MORILLAS M, ZOBLEY E, KIEFHABER T, GLOCKSHUBER R. Fast folding of the two-domain semliki forest virus capsid protein explains co-translational proteolytic activity. *J Mol Biol*. 2004; 338:159–67. [PubMed: 15050831]
219. SANDER IM, CHANEY JL, CLARK PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014; 136:858–61. [PubMed: 24392935]

220. SANTOS-PEREIRA JM, GARCIA-RUBIO ML, GONZALEZ-AGUILERA C, LUNA R, AGUILERA A. A genome-wide function of THSC/TREX-2 at active genes prevents transcription-replication collisions. *Nucleic Acids Res.* 2014; 42:12000–14. [PubMed: 25294824]
221. SAUNA ZE, KIMCHI-SARFATY C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet.* 2011; 12:683–91. [PubMed: 21878961]
222. SAUNDERS R, DEANE CM. Protein structure prediction begins well but ends badly. *Proteins.* 2010; 78:1282–90. [PubMed: 20014025]
223. SCHAAL TD, MANIATIS T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol.* 1999a; 19:261–73. [PubMed: 9858550]
224. SCHAAL TD, MANIATIS T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol.* 1999b; 19:1705–19. [PubMed: 10022858]
225. SCHATTNER P, DIEKHANS M. Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res.* 2006; 34:1700–10. [PubMed: 16556911]
226. SCHLEBACH JP, SANDERS CR. The safety dance: biophysics of membrane protein folding and misfolding in a cellular context. *Q Rev Biophys.* 2014; 1–35. [PubMed: 25420508]
227. SCOTT D, LAYFIELD R, OLDHAM NJ. Ion mobility-mass spectrometry reveals conformational flexibility in the deubiquitinating enzyme USP5. *Proteomics.* 2015
228. SEROHIJOS AW, SHAKHNOVICH EI. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol.* 2014; 26C:84–91. [PubMed: 24952216]
229. SHABALINA SA, OGURTSOV AY, ROGOZIN IB, KOONIN EV, LIPMAN DJ. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Res.* 2004; 32:1774–82. [PubMed: 15031317]
230. SHABALINA SA, SPIRIDONOV NA. The mammalian transcriptome and the function of non-coding DNA sequences. *Genome Biol.* 2004; 5:105. [PubMed: 15059247]
231. SHABALINA SA, SPIRIDONOV NA, KASHINA A. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.* 2013; 41:2073–94. [PubMed: 23293005]
232. SHAO ZQ, ZHANG YM, FENG XY, WANG B, CHEN JQ. Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS One.* 2012; 7:e33547. [PubMed: 22432034]
233. SHARP PM, COWE E, HIGGINS DG, SHIELDS DC, WOLFE KH, WRIGHT F. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.* 1988; 16:8207–11. [PubMed: 3138659]
234. SHARP PM, STENICO M, PEDEN JF, LLOYD AT. Codon usage: mutational bias, translational selection, or both? *Biochem Soc Trans.* 1993; 21:835–41. [PubMed: 8132077]
235. SHIELDS DC, SHARP PM, HIGGINS DG, WRIGHT F. “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol.* 1988; 5:704–16. [PubMed: 3146682]
236. SMITH MA, GESELL T, STADLER PF, MATTICK JS. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res.* 2013; 41:8220–36. [PubMed: 23847102]
237. SONG T, ZHENG Y, WANG Y, KATZ Z, LIU X, CHEN S, SINGER RH, GU W. Specific interaction of KIF11 with ZBP1 regulates the transport of beta-actin mRNA and cell motility. *J Cell Sci.* 2015
238. SPENCER PS, SILLER E, ANDERSON JF, BARRAL JM. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol.* 2012; 422:328–35. [PubMed: 22705285]
239. SPIESS C, ZHAI Q, CARTER PJ. Alternative molecular formats and therapeutic applications for bispecific antibodies. *Mol Immunol.* 2015

240. STEINGRIMSDOTTIR H, ROWLEY G, DORADO G, COLE J, LEHMANN AR. Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene. *Nucleic Acids Res.* 1992; 20:1201–8. [PubMed: 1373235]
241. STENSON PD, MORT M, BALL EV, SHAW K, PHILLIPS A, COOPER DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet.* 2014; 133:1–9. [PubMed: 24077912]
242. STOLETZKI N, EYRE-WALKER A. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol.* 2007; 24:374–81. [PubMed: 17101719]
243. STRONG MJ. The evidence for altered RNA metabolism in amyotrophic lateral sclerosis (ALS). *J Neurol Sci.* 2010; 288:1–12. [PubMed: 19840884]
244. STRUHL K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell.* 1999; 98:1–4. [PubMed: 10412974]
245. SUBRAMANIAM AR, PAN T, CLUZEL P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc Natl Acad Sci U S A.* 2013; 110:2419–24. [PubMed: 23277573]
246. SWARTZ JR. Advances in *Escherichia coli* production of therapeutic proteins. *Curr Opin Biotechnol.* 2001; 12:195–201. [PubMed: 11287237]
247. TAY Y, ZHANG J, THOMSON AM, LIM B, RIGOUTSOS I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature.* 2008a; 455:1124–8. [PubMed: 18806776]
248. TAY YM, TAM WL, ANG YS, GAUGHWIN PM, YANG H, WANG W, LIU R, GEORGE J, NG HH, PERERA RJ, LUFKIN T, RIGOUTSOS I, THOMSON AM, LIM B. MicroRNA-134 modulates the differentiation of mouse embryonic stem cells, where it causes post-transcriptional attenuation of Nanog and LRH1. *Stem Cells.* 2008b; 26:17–29. [PubMed: 17916804]
249. TER BEEK J, GUSKOV A, SLOTBOOM DJ. Structural diversity of ABC transporters. *J Gen Physiol.* 2014; 143:419–35. [PubMed: 24638992]
250. THANARAJ TA, ARGOS P. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci.* 1996a; 5:1973–83. [PubMed: 8897597]
251. THANARAJ TA, ARGOS P. Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* 1996b; 5:1594–612. [PubMed: 8844849]
252. TINOCO I JR, WEN JD. Simulation and analysis of single-ribosome translation. *Physical biology.* 2009; 6:025006. [PubMed: 19571367]
253. TSAI CJ, SAUNA ZE, KIMCHI-SARFATY C, AMBUDKAR SV, GOTTESMAN MM, NUSSINOV R. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol.* 2008; 383:281–91. [PubMed: 18722384]
254. TULLER T. Challenges and obstacles related to solving the codon bias riddles. *Biochem Soc Trans.* 2014; 42:155–9. [PubMed: 24450644]
255. TULLER T, CARMI A, VESTSIGIAN K, NAVON S, DORFAN Y, ZABORSKE J, PAN T, DAHAN O, FURMAN I, PILPEL Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell.* 2010a; 141:344–54. [PubMed: 20403328]
256. TULLER T, VEKSLER-LUBLINSKY I, GAZIT N, KUPIEC M, RUPPIN E, ZIV-UKELSON M. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.* 2011; 12:R110. [PubMed: 22050731]
257. TULLER T, WALDMAN YY, KUPIEC M, RUPPIN E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A.* 2010b; 107:3645–50. [PubMed: 20133581]
258. VENTER JC, ADAMS MD, MYERS EW, LI PW, MURAL RJ, SUTTON GG, SMITH HO, YANDELL M, EVANS CA, HOLT RA, GOCAYNE JD, AMANATIDES P, BALLEW RM, HUSON DH, WORTMAN JR, ZHANG Q, KODIRA CD, ZHENG XH, CHEN L, SKUPSKI M, SUBRAMANIAN G, THOMAS PD, ZHANG J, GABOR MIKLOS GL, NELSON C, BRODER S, CLARK AG, NADEAU J, MCKUSICK VA, ZINDER N, LEVINE AJ, ROBERTS RJ, SIMON M, SLAYMAN C, HUNKAPILLER M, BOLANOS R, DELCHER A, DEW I, FASULO D, FLANIGAN M, FLOREA L, HALPERN A, HANNENHALLI S, KRAVITZ S,

LEVY S, MOBARRY C, REINERT K, REMINGTON K, ABU-THREIDEH J, BEASLEY E, BIDDICK K, BONAZZI V, BRANDON R, CARGILL M, CHANDRAMOULISWARAN I, CHARLAB R, CHATURVEDI K, DENG Z, DI FRANCESCO V, DUNN P, EILBECK K, EVANGELISTA C, GABRIELIAN AE, GAN W, GE W, GONG F, GU Z, GUAN P, HEIMAN TJ, HIGGINS ME, JI RR, KE Z, KETCHUM KA, LAI Z, LEI Y, LI Z, LI J, LIANG Y, LIN X, LU F, MERKULOV GV, MILSHINA N, MOORE HM, NAIK AK, NARAYAN VA, NEELAM B, NUSSKERN D, RUSCH DB, SALZBERG S, SHAO W, SHUE B, SUN J, WANG Z, WANG A, WANG X, WANG J, WEI M, WIDES R, XIAO C, YAN C, et al. The sequence of the human genome. *Science*. 2001; 291:1304–51. [PubMed: 11181995]

259. VILLALOBOS A, NESS JE, GUSTAFSSON C, MINSHULL J, GOVINDARAJAN S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics*. 2006; 7:285. [PubMed: 16756672]
260. WAGNER A. Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet*. 2008; 9:965–74. [PubMed: 18957969]
261. WAIN-HOBSON S, NUSSINOV R, BROWN RJ, SUSSMAN JL. Preferential codon usage in genes. *Gene*. 1981; 13:355–64. [PubMed: 7262559]
262. WAN Y, KERTESZ M, SPITALE RC, SEGAL E, CHANG HY. Understanding the transcriptome through RNA structure. *Nat Rev Genet*. 2011; 12:641–55. [PubMed: 21850044]
263. WANG Z, ROLISH ME, YEO G, TUNG V, MAWSON M, BURGE CB. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004; 119:831–45. [PubMed: 15607979]
264. WARF MB, BERGLUND JA. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci*. 2010; 35:169–78. [PubMed: 19959365]
265. WASHIETL S, PEDERSEN JS, KORBEL JO, STOCSTITS C, GRUBER AR, HACKERMULLER J, HERTEL J, LINDEMAYER M, REICHE K, TANZER A, UCLA C, WYSS C, ANTONARAKIS SE, DENOEUDE F, LAGARDE J, DRENKOW J, KAPRANOV P, GINGERAS TR, GUIGO R, SNYDER M, GERSTEIN MB, REYMOND A, HOFACKER IL, STADLER PF. Structured RNAs in the ENCODE selected regions of the human genome. *Genome Res*. 2007; 17:852–64. [PubMed: 17568003]
266. WEN JD, LANCASTER L, HODGES C, ZERI AC, YOSHIMURA SH, NOLLER HF, BUSTAMANTE C, TINOCO I. Following translation by single ribosomes one codon at a time. *Nature*. 2008; 452:598–603. [PubMed: 18327250]
267. WHITESIDES GM, GRZYBOWSKI B. Self-assembly at all scales. *Science*. 2002; 295:2418–21. [PubMed: 11923529]
268. WIEDEMANN B, BOLES E. Codon-optimized bacterial genes improve L-Arabinose fermentation in recombinant *Saccharomyces cerevisiae*. *Appl Environ Microbiol*. 2008; 74:2043–50. [PubMed: 18263741]
269. WILKINSON KA, MERINO EJ, WEEKS KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc*. 2006; 1:1610–6. [PubMed: 17406453]
270. WILLIE E, MAJEWSKI J. Evidence for codon bias selection at the pre-mRNA level in eukaryotes. *Trends Genet*. 2004; 20:534–8. [PubMed: 15475111]
271. WINTERMEYER W, PESKE F, BERINGER M, GROMADSKI KB, SAVELSBERGH A, RODNINA MV. Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine. *Biochem Soc Trans*. 2004; 32:733–7. [PubMed: 15494001]
272. WOESE CR. Universality in the Genetic Code. *Science*. 1964; 144:1030–1. [PubMed: 14137944]
273. WOESE CR. Order in the genetic code. *Proc Natl Acad Sci U S A*. 1965; 54:71–5. [PubMed: 5216368]
274. WU G, BASHIR-BELLO N, FREELAND SJ. The Synthetic Gene Designer: a flexible web platform to explore sequence manipulation for heterologous expression. *Protein Expr Purif*. 2006; 47:441–5. [PubMed: 16376569]
275. WU X, JORNVALL H, BERNDT KD, OPPERMANN U. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and

- secondary structure but not tRNA abundance. *Biochem Biophys Res Commun*. 2004; 313:89–96. [PubMed: 14672702]
276. WURM F, BERNARD A. Large-scale transient expression in mammalian cells for recombinant protein production. *Curr Opin Biotechnol*. 1999; 10:156–9. [PubMed: 10209142]
277. WURM FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. 2004; 22:1393–8. [PubMed: 15529164]
278. XIE T, DING D. The relationship between synonymous codon usage and protein structure. *FEBS Lett*. 1998; 434:93–6. [PubMed: 9738458]
279. XU Y, MA P, SHAH P, ROKAS A, LIU Y, JOHNSON CH. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*. 2013; 495:116–20. [PubMed: 23417065]
280. YOON JH, ABDELMOHSEN K, GOROSPE M. Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol*. 2013; 425:3723–30. [PubMed: 23178169]
281. YOON JH, ABDELMOHSEN K, GOROSPE M. Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol*. 2014; 34:9–14. [PubMed: 24965208]
282. ZHANG H, GAU BC, JONES LM, VIDAUSKY I, GROSS ML. Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system. *Anal Chem*. 2011; 83:311–8. [PubMed: 21142124]
283. ZHANG J, LONG M, LI L. Translational effects of differential codon usage among intragenic domains of new genes in *Drosophila*. *Biochim Biophys Acta*. 2005; 1728:135–42. [PubMed: 15833448]
284. ZUCKERKANDL E, PAULING L. Molecules as documents of evolutionary history. *J Theor Biol*. 1965; 8:357–66. [PubMed: 5876245]
285. ZUKER M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003; 31:3406–15. [PubMed: 12824337]

Figure 1. The consequences of a synonymous single nucleotide change on the predicted structure of the mRNA (mfold)

The predicted (mfold) structures of the Ile507-ATC and Ile507-ATT F508 *CFTR* mRNAs. The sequences represent human *CFTR* mRNA fragments encoding the region of NBD1 near the F508 mutation. The locations of the altered nucleotides (C and U) are highlighted in red.

Figure 2. The consequence of synonymous mutations in exonic splice regulatory sites
 sSNPs may disrupt critical elements necessary for splicing. In the example shown, this results in exon skipping. ESE: exonic splicing enhancer; ESS: exonic splicing suppressor. (For a review concerning pre-mRNA splicing refer to: (Muller-McNicoll and Neugebauer, 2013)).

Figure 3. Different steps of translation and cotranslational protein folding affected by synonymous mutation

sSNPs may alter translation initiation efficiency, translation elongation rate, ribosomal pause rhythm, cotranslational folding or the overall fate of the protein.

Table 1

Diseases associated with synonymous mutations.

DISEASE	GENE	CHANGE	AA	LOCATION	CONSEQUENCE	Reference
Phenylketonuria	PAH	C→G	Gly	exon 1	skipping of exon 1	Dobrowolski, S.F. et al, 2010
von Hippel-Lindau disease	VHL	A→C	Pro	exon 2	skipping of exon 2	Martella, M. et al, 2006
Acute intermittent porphyria	PBG	C→G	Arg	exon 3	skipping of exon 3	Llewellyn, D.H. et al, 1996
Autosomal dominant polycystic kidney disease	PKD1	A→T	Gly	exon 3	skipping of exon 3	Gonzalez-Paredes, 2014
Coeliac disease	CD44	C→T	His	exon 3	skipping of exon 3	Vidal, C. et al, 2009
Myasthenic syndrome	DPAGT1	G→C	Leu	exon 3	skipping of exon	Selcen, D. et al, 2014
Crouzon syndrome	FGFR2	G→A	Ala	exon 3c	skipping of exon 3	Delgatto, F. & Breathnach, 1995
X-linked parkinsonism with spasticity	ATP6AP2	C→T	Ser	exon 4	skipping of exon 4	Korvatska, O. et al, 2013
X-linked mental retardation Hedera type	ATP6AP2	C→T	Asp	exon 4	skipping of exon 4	Ramser, J. et al, 2005
PDH deficiency (range of symptoms)	PDHA1	C→T	Tyr161	c.483 in exon 5	skipping of exon 5	Biochard, A. et al, 2008
PDH deficiency (range of symptoms)	PDHA2	C→T	Ile 166	c.498 in exon 5	skipping of exon 5	Biochard, A. et al, 2008
MCD4 deficiency	MCAD	A→C	Thr	exon 5	skipping of exon 5	Nielsen, K.B. et al, 2007
Leigh's encephalopathy	PDH-E1α	A→G	Gly	exon 6	skipping of exon 6	De Meirleir et al, 1994
Congenital muscular dystrophies	POMGNT1	C→T	Phe	exon 7	skipping of exon 7	Oliveira, J. et al, 2008
Spinal muscular atrophy	SMN2	C→T	Phe	exon 7	skipping of exon 7	Cartegni and Krainer, 2002
Lesch-Nyhan syndrome	HGPRT	C→T	Phe	exon 8	skipping of exon 8 and hypothesized alteration of mRNA structure	Steingrimsdottir et al, 1992
Netherton syndrome	SPINK5	C→T	Cys	exon 11	skipping of exon 11	Lacroix, M. et al, 2012
Phenylketonuria	HPHG	A→T	Val	exon 11	skipping of exon 11	Chao et al, 2001
Isolated GH deficiency (IGHD) type IB	GHRHR	G→A	Glu	exon 12	skipping of exon 12	Inoue, H. et al, 2011
Cystic fibrosis	CFTR	T→C	Ile	exon 12	skipping of exon 12	Pagani F. et al, 2003
Kell phenotypes (blood group)	KEL	C→T	Gly	exon 16	skipping of exon 16	Karamatic Crew, V. et al, 2014
DFNB4 (congenital hearing loss and dilated vestibular aqueducts)	PDS or SLC26A4	G→A	Lys	exon 16	skipping of exon 16	Kim, Y. et al, 2013
Treacher Collins syndrome	TTCOF1	A→C	Ser	exon 22	skipping of exon 22	Macaya et al, 2009
von Willebrand disease type 2A/IIIE	VWF	C→T	Gly	exon 26	skipping of exon 26	Pagliari, M.T. et al, 2013

DISEASE	GENE	CHANGE	AA	LOCATION	CONSEQUENCE	Reference
Autosomal dominant polycystic kidney disease	PKD1	C→A	Arg	exon 39	skipping of exon 39	Gonzalez-Paredes, F. J. et al. 2014
Usher syndrome type 1	CDH23	G→A	Thr	exon 42	skipping of exon 42	von Brederlow, B. et al. 2002
Dystrophic epidermolysis bullosa pruriginosa	COL7A1	G→C	Leu	exon 87	skipping of exon 87	Covaciu, C. et al. 2011
Cystic fibrosis	CFTR	G→T	Gly	exon 15	partial splicing of exon 15	Faa, V. et al. 2010
Osteoporosis	CD44	G→A	Glu	exon 9	abolished ESE	Vidal, C. et al. 2009
Drug resistance	MDR1	C→T	Ile	exon 26	a rare codon alters translation rhythm due to changes in tRNA availability resulting in altered protein folding, conformation, protein activity, and substrate specificity	Kimchi-Sarfaty et al. 2007
Drug resistance	MDR1	C→T	Ile	exon 26	confirmed altered protein conformation, stability, and transporter specificity in epithelial monolayers	Fung, K.L. et al. 2014
Increased pain sensitivity	COMT	G→A	Ala	exon 6	↑ mRNA structural stability	Nackley et al. 2006
Cystic fibrosis	F508 CFTR	C→T	Ile	exon 12	↓ mRNA structural stability alters protein folding	Bartoszewski et al. 2010
Cystic fibrosis	F508 CFTR	C→T	Ile	exon 12	confirmed altered protein stability and function	Lazrak and Fu et al. 2013
Schizophrenia	DRD2	C→T	Pro	exon 7	↓ mRNA structural stability	Duan et al. 2002
Chron's disease	IRGM	C→T	Leu	exon 2	↑ mRNA half-life	Brest et al. 2011
Melanoma	BCL2L12	C→T	Phe	exon 1	↑ mRNA half-life	Gartner et al. 2013
Charcot-Marie-Tooth disease type 1B	CMT1B	G→A	Val	exon 3	increases the strength of a preexisting cryptic donor site which then overpowers the normal donor site	Crehalet et al. 2010
Vasovagal syndrome	GNAS1	T→C	Ile	exon 5	↓ mRNA half-life	Lelonek et al. 2008
Essential hypertension and altered sensitivity to beta-blockers	GNAS1	T→C	Ile	exon 5	absence of a restriction site for FokI endonuclease	Jia et al. 1999
Amyotrophic lateral sclerosis	SOD1	C→T	Gly	exon 1	reduced RNP complex formation → ↓ mRNA stability	Ge et al. 2006
Amyotrophic lateral sclerosis	SOD1	T→C	Ser	exon 3	reduced RNP complex formation → ↓ mRNA stability	Ge et al. 2006
Amyotrophic lateral sclerosis	SOD1	A→G	Thr	exon 4	reduced RNP complex formation → ↓ mRNA stability	Ge et al. 2006
Amyotrophic lateral sclerosis	SOD1	C→T	Ala	exon 5	reduced RNP complex formation → ↓ mRNA stability	Ge et al. 2006
Mild hemophilia B	F9	G→A	Val	exon 5	no exon skipping or intron inclusions	Knobe, K.E. et al. 2008

DISEASE	GENE	CHANGE	AA	LOCATION	CONSEQUENCE	Reference
Congenital heart disease	NKX2-5	G→A A→G	Gln Glu	c.543 c.63	in cis with Ala119Glu mutation further diminishes function of the protein	Reamon-Buettner et al, 2013
Amyotrophic lateral sclerosis	SOD1	C→T	Asn	exon 5	unknown	Andersen et al, 1997
Amyotrophic lateral sclerosis	SOD1	A→G	Gln	exon 5	unknown	Andersen et al, 1997
Phenylketonuria	PAH	C→T	Leu	exon 9	unknown	Ho, P.Y. et al, 2008
Phenylketonuria	PAH	A→G	Thr	exon 9	unknown	Ho, P.Y. et al, 2008
Autism spectrum disorder	NLGN3	G→A	Lys	unknown	unknown	Yanagi et al, 2012