The International Journal of Biochemistry & Cell Biology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

The International Journal of Biochemistry & Cell Biology

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

Review

Decoding mechanisms by which silent codon changes influence protein biogenesis and function

Vedrana Bali, Zsuzsanna Bebok*

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

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form 2 March 2015
Accepted 14 March 2015
Available online xxx

Keywords:
Codon usage bias (CUB)
sSNP
mRNA structure
Translation dynamics
Protein folding

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.

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; *C. elegans*, *Caenorhabditis elegans*; 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; *D. melanogaster*, *Drosophila melanogaster*; ΔF508, deletion of phenylalanine at position 508; DNA, deoxyribose nucleic acid; DRD2, dopamine receptor D2; *E. coli*, *Escherichia coli*; 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 triphospatase (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, polycistin 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; SilVA, 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.

* Corresponding author at: The Department of Cell, Developmental and Integrative Biology, University of Alabama at Birmingham, 1918 University Boulevard, MCLM 350A, Birmingham, AL 35294, USA. Tel.: +1 205 975 5449.
E-mail address: bebok@uab.edu (Z. Bebok).

http://dx.doi.org/10.1016/j.biocel.2015.03.011
1357-2725/

Please cite this article in press as: Bali V, Bebok Z. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. Int J Biochem Cell Biol (2015), http://dx.doi.org/10.1016/j.biocel.2015.03.011

Contents

1. Introduction

Q2 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.

2. Historical perspective and terminology

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

Elucidating the structure of DNA (Watson and Crick, 1953) led Q3 to identification of the nucleotide triplets that comprise the genetic code (Crick, 1955; Crick et al., 1961), its universality (Woese, 1965, 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, 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).

2.2. 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

Please cite this article in press as: Bali V, Bebok Z. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. Int J Biochem Cell Biol (2015), http://dx.doi.org/10.1016/j.biocel.2015.03.011