

# The Codon Usage Code for Cotranslational Folding of Viral Capsids

Rosa M. Pintó\* and Albert Bosch

Enteric Virus Laboratory, Section of Microbiology, Virology and Biotechnology, Department of Genetics, Microbiology and Statistics, School of Biology, and Institute of Nutrition and Safety, University of Barcelona, Barcelona, Spain

\*Corresponding author: E-mail: rpinto@ub.edu

Accepted: 22 April 2021

## Abstract

Codon bias is common to all organisms and is the result of mutation, drift, and selection. Selection for the efficiency and accuracy of translation is well recognized as a factor shaping the codon usage. In contrast, fewer studies report the control of the rate of translation as an additional selective pressure influencing the codon usage of an organism. Experimental molecular evolution using RNA virus populations is a powerful tool for the identification of mechanisms underlying the codon bias. Indeed, the role of deoptimized codons on the cotranslational folding has been proven in the capsids of two fecal-orally transmitted picornaviruses, poliovirus, and the hepatitis A virus, emphasizing the role of the frequency of codons in determining the phenotype. However, most studies on virus codon usage rely only on computational analyses, and experimental studies should be encouraged to clearly define the role of selection on codon evolution.

**Key words:** codon usage, translation selection, translation kinetics selection, protein folding, cotranslational folding, virus.

## Significance

The role of synonymous codons in controlling the rate of translation is a matter of debate in Evolutionary Biology. Experimental evolution using two picornaviruses has proven that deoptimized codons indeed influence the rate of mRNA translation and in turn in the folding of their capsids. This important finding needs to be confirmed in other viruses with different biological cycles and in other proteins too.

Codon bias, or the nonrandom usage of synonymous codons, is a common feature of most species in all domains of life. The particular codon usage of an organism reflects the evolutionary forces that have acted on its genome including mutation, genetic drift, and selection (Grantham et al. 1980; Hershberg and Petrov 2008; Sharp et al. 2010; Plotkin and Kudla 2011). It has been proposed that selection could favor preferred codons over rare codons to improve the efficiency and accuracy of the translation (i.e., translation selection), while mutational pressure and genetic drift would let minor codons to persist (Bulmer 1991; Duret 2002; Hershberg and Petrov 2008). However, there is also evidence of selection on rare or non-optimal codons to control the ribosome traffic on the translating mRNA (i.e., translation kinetics selection). The underlying

hypothesis is that tRNAs translating preferred codons would be abundant in the tRNA pool whereas tRNAs translating rare codons would be scarce. In this latter case, the longer time required for incorporating the scarce tRNAs into the ribosome A site would induce ribosome stalls slowing down translation elongation (Yang and Nielsen 2008; Komar 2009; Tuller et al. 2010; Plotkin and Kudla 2011; Wohlgemuth et al. 2013; Chaney and Clark 2015; Yu et al. 2015; Weinberg et al. 2016; Zhao et al. 2017). These ribosome stalls would play an essential role in the cotranslational folding of the nascent polypeptide by temporally separating the folding events, boosting the “favorable” and preventing the “unfavorable” interactions within the growing peptide (Yang and Nielsen 2008). An elegantly designed experiment demonstrated that the rate at

which a nascent protein exits from the ribosome can specify the folded structure of a protein (Sander et al. 2014). This theory has been proposed as the codon usage “code” for protein folding (Yu et al. 2015; Liu 2020; Liu et al. 2021).

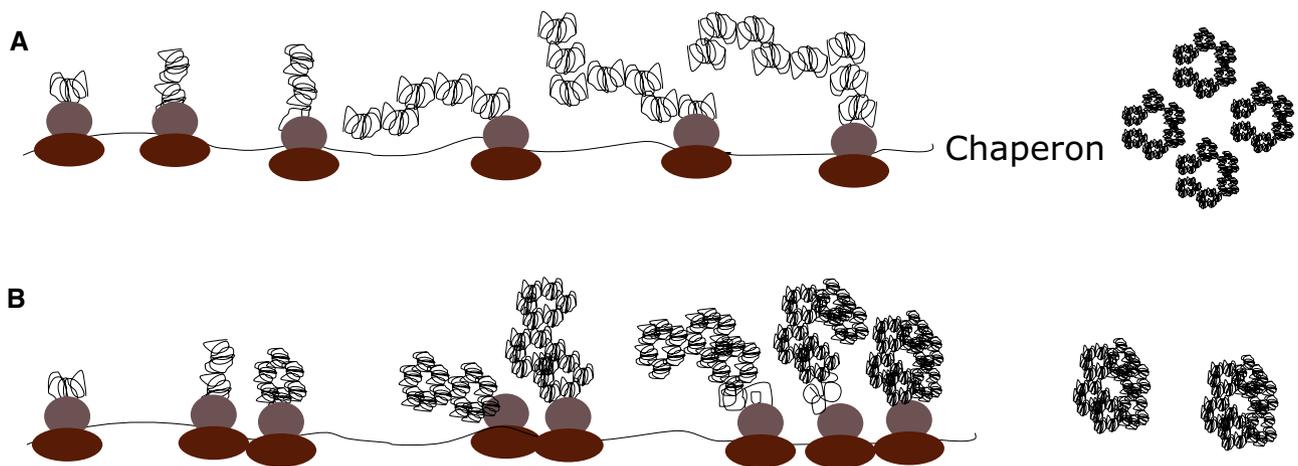
Several works have proven the role of nonoptimal codons in protein folding, including the *Neurospora* FREQUENCY protein (Zhou et al. 2013) and the *Drosophila* PERIOD protein (Fu et al. 2016), both essential for the circadian clock function. In humans, one of the most paradigmatic examples of how a silent mutation may influence the cotranslational folding and function of a protein is the synonymous single-nucleotide polymorphisms (SNPs) of the Multidrug Resistance 1 (MDR1) gene, which alters the substrate and inhibitor binding sites of its P-glycoprotein product (Kimchi-Sarfaty et al. 2007). Remarkably, synonymous SNPs have been correlated with human diseases in genome-wide association studies, but the underlying mechanisms have only been identified in a few of them (Chamary and Hurst 2009; Chen et al. 2010; Sauna and Kimchi-Sarfaty 2011). Genome projects generate massive amounts of sequences which may be comparatively analyzed using population genetics statistical models to detect footprints of selection (Kreitman 2000; Excoffier and Heckel 2006; Booker et al. 2017). However, these models do not usually consider selection for the efficiency, accuracy, and kinetics of translation. Hopefully, an alternative could arise from experimental evolution studies.

RNA viruses are a powerful tool for experimental evolution owing to their great evolutionary potential, which results from their very short generation time, their large population size, and their error-prone replication. RNA viruses exist as complex mutant swarms which are the target of selection (Domingo et al. 2012). The same molecular evolutionary rules apply to all organisms, from viruses to cells, but the relative weight of each driving process may vary between them. Different selective pressures may shape the codon usage of RNA viruses. For instance, viruses have evolved to very low CpG and UpA dinucleotide contents to evade antiviral cell responses, which in turn could influence their codon composition (Atkinson et al. 2014). Secondary structures in the genomic RNA, with functional roles through the interaction with viral or cellular proteins, could also constraint the codon usage (Gumpper et al. 2018), and finally yet importantly selection for an efficient and regulated translation may shape codon usage. An interesting example of a combined action of selection for translation efficiency and virus overcome of the cellular antiviral action is the codon usage of the avian influenza H3N2 virus PB1 mRNA, which is skewed towards interferon-altered human tRNA pools contributing to its codon bias (Smith et al. 2018).

Viruses are not able to synthesize tRNAs, hence they depend on the cellular tRNA pools for their own translation. Excessive codon usage similarity between virus and hosts may impede host translation, inducing a deleterious effect, and consequently, viruses have evolved to an optimal range of codon usage bias (Chen et al. 2020). Additionally, viruses

may be able or not to inhibit the cellular protein synthesis, that is, the host shut off, and this will have critical implications on tRNA availability. Assuming translation efficiency as the main selective pressure shaping the codon composition, a codon usage in the optimal range with the host should be expected in viruses able to shut down the protein synthesis, like poliovirus (PV), while a more deviated codon usage should apply in viruses unable to do so, like hepatitis A virus (HAV) (Pintó et al. 2018). In this latter example, selection for a slow translation elongation has been proven to be the main selection for its deviated codon usage. In experimental evolution studies, HAV was adapted to grow in the presence of actinomycin D, which specifically inhibits the cellular DNA-dependent RNA polymerase with no effects on the viral RNA-dependent RNA polymerase, and evolved with an initial increase of nonoptimal codons in the capsid coding region and thereafter with an increase of optimal codons (Aragonès et al. 2010; Costafreda et al. 2014). These codon adjustments induced significant changes in the rate of translation (D’Andrea et al. 2019), and in turn in capsid features such as the antigenic structure, the physical stability, and the uncoating efficiency (Costafreda et al. 2014; D’Andrea et al. 2019). Similarly, adaptation of PV to conditions of suppression of chaperon activity promoted the selection of clusters of codon-deoptimized mutations, which may assist the cotranslational folding (Geller et al. 2018). In both cases, there were significant increases in the diversity of the mutant swarms, and, although the selected genotypes mostly harbored few or even single mutations, they gave rise to new translation phenotypes that in turn increased the diversity of capsid phenotypes. PV uses chaperon Hsp90 to support capsid folding (Geller et al. 2012), while HAV does not require this chaperon and instead uses deoptimized codons for the control of the cotranslational folding (Aragonès et al. 2010; Costafreda et al. 2014; D’Andrea et al. 2019). It can thus be concluded that an efficient translation rate requires a tradeoff with an efficient cotranslational folding, which may be achieved by chaperon action or by selecting codon deoptimization at specific sites. The noncoding function of codons has been defined as the code within the genomic code (fig. 1).

There are still many questions to be resolved. Has the essential role of deoptimized codons in the control of cotranslational folding a relationship with the capsid symmetry and hence may be extended to other icosahedral capsids, or is it rather more related with aspects of the virus biology such as the mode of transmission? The fecal–oral route of transmission, used by PV and HAV, involves long periods outside the body host requiring highly stable capsids. In contrast, the arthropod-borne viruses, which have a broad spectrum of hosts and are transmitted by blood-feeding bites with no extra body phases, have evolved to codon usages resembling either their natural invertebrate hosts, when the vertebrates act as accidental dead-end hosts (bunyaviruses, rhabdoviruses, and reoviruses), or both invertebrate and vertebrate



**FIG. 1.**—General mechanism of the potential impact of codon deoptimization on protein folding. (A) The use of abundant codons ensures a rapid and efficient translation with a high production of protein at the cost of absence of cotranslational folding, instead chaperons assist the protein folding. This is the model for the poliovirus capsid folding. (B) The occurrence of clusters of rare codons and the associated ribosome stalls locally slows down the kinetics of translation promoting the cotranslational folding at the cost of a low production of proteins. This is the model for the hepatitis A virus capsid folding.

hosts (alphaviruses, flaviviruses, and orthomyxoviruses) when virus replication in vertebrates results in high-titer viremia (Velázquez-Salinas et al. 2016). In this latter scenario, virus transmission occurs through mosquito or tick bites, and to ensure transmission selection for an efficient translation would have contributed to shape their codon usage which would be in accordance with their capacity to induce the cellular shutoff (Fros and Pijlman 2016; Roth et al. 2017). Similarly, it has been described that viruses with a broad range of hosts have a worse codon usage match with their hosts than viruses with a narrow range of hosts (Tian et al. 2018) and that vector-borne viruses show lower codon bias than aerosol-transmitted viruses (Jenkins and Holmes 2003). Viral fitness is usually measured through the level of virus replication, but the efficient transmission is also a major factor with a critical impact in the overall fitness (Wargo and Kurath 2012). Codon usage may play a role in the mode of transmission of viruses, either facilitating the translation efficiency or controlling the capsid cotranslational folding.

It is important to emphasize the need to go one step further in identifying the underlying mechanisms of the codon usage bias in the different viruses, rather than simply describing their codon composition and comparing it with the host codon usage. Although there are outstanding studies showing selection for translation efficiency, that is, through the use of optimal codons and codon pairs, selection for the regulation of the translation rate, that is, through the right combination of optimal and nonoptimal codons, and selection for escaping the antiviral cell response, that is, through the low CpG and UpA contents, as mechanisms contributing to codon bias (Walsh et al. 2020; Burns et al. 2006; Mueller et al. 2006; Coleman et al. 2008; Mueller et al. 2010; Lauring et al. 2012; Tulloch et al. 2014; Martinez et al. 2016; Pintó et al. 2018; D'Andrea et al. 2019; Mueller et al. 2020), many others

are merely descriptive. For instance, the total number of publications found in the PubMed archive in the period 2000–2020 using the terms RNA+Virus+Codon+Usage was of 469, adding the term translation was of 153, and adding the term folding was of only 18, and corresponding most of these references to genomic comparative analysis. It is essential to determine the type of selection pressures contributing to the codon bias of each virus under study. The topics of selection for the translation efficiency and selection for the codon-driven translation kinetics may take advantage of experimental evolution in the framework of the new advances in the analysis of tRNA abundance and heterogeneity, which is host and cell specific. Finally, in the case of the codon-driven cotranslational folding, it is necessary to know how universal it is, and to what extent the codon frequency dictates the phenotypic landscapes of the protein folding of the mutant swarms in tRNA changing environments (hosts and cell types, shut off, chaperons, interferon, temperature, etc.). Robustness in the different conditions could anticipate the evolvability of virus populations. The potential constraints or benefits that the codon composition of a virus population may impose for the adaptation to new hosts could contribute to identify potential zoonotic spillovers between very distant species. A deep understanding of the codes within the genomic code would contribute to advancements in the field of virus evolution.

## Acknowledgments

Our Research is supported by Spanish Ministry of Economy grant BIO2017-83191-R cofinanced by the European Regional Development Fund, and by Grant 20182830 from La Marató de TV3 Foundation.

## Literature cited

- Aragonès L, Guix S, Ribes E, Bosch A, Pintó RM. 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* 6:e1000797.
- Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. 2014. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.* 42:4527–4545.
- Booker TR, Jackson BC, Keightley PD. 2017. Detecting positive selection in the genome. *BMC Biol.* 15:98.
- Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129:897–907.
- Burns CC, et al. 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J Virol.* 80:3259–3272.
- Coleman JR, et al. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784–1787.
- Costafreda MI, et al. 2014. Hepatitis A virus adaptation to cellular shutoff is driven by dynamic adjustments of codon usage and results in the selection of populations with altered capsids. *J Virol.* 88:5029–5041.
- Chamary JV, Hurst LD. 2009. The price of silent mutations. *Sci Am.* 300(6):46–53.
- Chaney JL, Clark PL. 2015. Roles for synonymous codon usage in protein biogenesis. *Annu Rev Biophys.* 44:143–166.
- Chen F, et al. 2020. Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection. *Nature Ecol Evol.* 4:589–600.
- Chen R, Davydov EV, Sirota M, Butte AJ. 2010. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One* 5:e13574.
- D'Andrea L, et al. 2019. The critical role of codon composition on the translation efficiency robustness of the hepatitis A virus capsid. *Genome Biol Evol.* 11:2439–2456.
- Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev.* 76:159–216.
- Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev.* 12:640–649.
- Excoffier L, Heckel G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet.* 7:745–758.
- Fros JJ, Pijlman GP. 2016. Alphavirus infection: host cell shut-off and inhibition of antiviral responses. *Viruses* 8(6):166.
- Fu J, et al. 2016. Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev.* 30:1761–1775.
- Geller R, Pechmann S, Acevedo A, Andino R, Frydman J. 2018. Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat Commun.* 9:1781.
- Geller R, Taguwa S, Frydman J. 2012. Broad action of Hsp90 as a host chaperone required for viral replication. *Biochim Biophys Acta.* 1823:698–706.
- Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8:r49–r62.
- Gumpper RH, Li W, Luo M. 2018. Constraints of viral RNA synthesis on codon usage of negative-strand RNA virus. *J Virol.* 93(5):e01775–18.
- Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet.* 42:287–299.
- Jenkins GM, Holmes EC. 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92:1–7.
- Kimchi-Sarfaty C, et al. 2007. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528.
- Komar AA. 2009. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci.* 34:16–24.
- Kreitman M. 2000. Methods to detect selection in populations with applications to the human. *Ann Rev Genom Hum Genet.* 1:539–559.
- Lauring AS, Acevedo A, Cooper SB, Andino R. 2012. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe.* 12(5):623–632.
- Liu Y. 2020. A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun Signal.* 18:145.
- Liu Y, Yang Q, Zhao F. 2021. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Ann Rev Biochem.* doi: 10.1146/annurev-biochem-071320-112701.
- Martinez MA, Jordan-Paiz A, Franco S, Nevot M. 2016. Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol.* 24:134–147.
- Mueller S, et al. 2010. Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotech.* 28:723–726.
- Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol.* 80:9687–9696.
- Mueller S, et al. 2020. A codon-pair deoptimized live-attenuated vaccine against respiratory syncytial virus is immunogenic and efficacious in non-human primates. *Vaccine* 38:2943–2948.
- Pintó RM, et al. 2018. Hepatitis A virus codon usage: implications for translation kinetics and capsid folding. *Cold Spring Harb Perspect Med.* 8(10):a031781.
- Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* 12:32–42.
- Roth H, et al. 2017. Flavivirus infection uncouples translation suppression from cellular stress responses. *MBio* 8(2):e02150–16.
- Sander IM, Chaney JL, Clark PL. 2014. Expanding Anfinsen’s principle: contributions of synonymous codon selection to rational protein design. *J Am Chem. Soc.* 136:858–861.
- Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet.* 12:683–691.
- Sharp PM, Emery LR, Zeng K. 2010. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.* 365:1203–1212.
- Smith BL, Chen G, Wilke CO, Krug RM. 2018. Avian influenza virus PB1 gene in H3N2 viruses evolved in humans to reduce interferon inhibition by skewing codon usage toward interferon-altered tRNA pools. *MBio* 9(4):e01222–18.
- Tian L, Shen X, Murphy RW, Shen Y. 2018. The adaptation of codon usage of +ssRNA viruses to their hosts. *Infect Genet Evol.* 63:175–179.
- Tuller T, et al. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141:344–354.
- Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. 2014. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife* 3:e04531.
- Velázquez-Salinas L, et al. 2016. Selective factors associated with the evolution of codon usage in natural populations of arboviruses. *PLoS One* 11:e0159943.
- Walsh IM, Bowman MA, Soto Santarriaga IF, Rodríguez A, Clark PL. 2020. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci U S A.* 117(7):3528–3534.
- Wargo AR, Kurath G. 2012. Viral fitness: definitions, measurement, and current insights. *Curr Opin Virol.* 2:538–545.
- Weinberg DE, et al. 2016. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* 14:1787–1799.
- Wohlgemuth SE, Gorochowski TE, Roubos JA. 2013. Translational sensitivity of the *Escherichia coli* genome to fluctuating tRNA availability. *Nucleic Acids Res.* 41:8021–8033.
- Yang Z, Nielsen R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol.* 25:568–579.

Yu CH, et al. 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell*. 59:744–754.

Zhao F, Yu CH, Liu Y. 2017. Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells. *Nucleic Acids Res*. 45:8484–8492.

Zhou M, et al. 2013. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495(7439): 111–115.

**Associate editor:** Ignacio Bravo