



Impact of Synonymous Genome Recoding on the HIV Life Cycle

Ana Jordan-Paiz[†], Sandra Franco and Miguel Angel Martínez*

IrsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona (UAB), Badalona, Spain

OPEN ACCESS

Edited by:

Cara Carthel Burns,
Centers for Disease Control
and Prevention (CDC), United States

Reviewed by:

Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands
Zuben E. Sauna,
United States Food and Drug
Administration, United States

*Correspondence:

Miguel Angel Martínez
mmartinez@irsicaixa.es

† Present address:

Ana Jordan-Paiz,
Heinrich Pette Institute, Leibniz
Institute for Experimental Virology,
Hamburg, Germany

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 14 September 2020

Accepted: 25 February 2021

Published: 16 March 2021

Citation:

Jordan-Paiz A, Franco S and
Martínez MA (2021) Impact
of Synonymous Genome Recoding on
the HIV Life Cycle.
Front. Microbiol. 12:606087.
doi: 10.3389/fmicb.2021.606087

Synonymous mutations within protein coding regions introduce changes in DNA or messenger (m) RNA, without mutating the encoded proteins. Synonymous recoding of virus genomes has facilitated the identification of previously unknown virus biological features. Moreover, large-scale synonymous recoding of the genome of human immunodeficiency virus type 1 (HIV-1) has elucidated new antiviral mechanisms within the innate immune response, and has improved our knowledge of new functional virus genome structures, the relevance of codon usage for the temporal regulation of viral gene expression, and HIV-1 mutational robustness and adaptability. Continuous improvements in our understanding of the impacts of synonymous substitutions on virus phenotype – coupled with the decreased cost of chemically synthesizing DNA and improved methods for assembling DNA fragments – have enhanced our ability to identify potential HIV-1 and host factors and other aspects involved in the infection process. In this review, we address how silent mutagenesis impacts HIV-1 phenotype and replication capacity. We also discuss the general potential of synonymous recoding of the HIV-1 genome to elucidate unknown aspects of the virus life cycle, and to identify new therapeutic targets.

Keywords: HIV-1, synonymous, mutations, virus, phenotype

INTRODUCTION

Despite the relatively recent introduction of HIV-1 within the human population, this virus has already exhibited enormous diversification. This high genetic diversity results from its fast replication cycle, with the generation of about 10^{10} virions daily in an infected individual, coupled with its high mutation rate of approximately 3×10^{-5} per nucleotide base per replication cycle, and the recombinogenic properties of its reverse transcriptase (RT) (Coffin and Swanstrom, 2013). The HIV-1 RNA genome comprises an above-average proportion of adenine (A) nucleotides, while being extremely poor in cytosine (C) (van der Kuyl and Berkhout, 2012). Interestingly, despite the high variability of the HIV-1 genome, its base composition is surprisingly stable over time, varying by <1% per base per isolate regardless of whether it originates from the early or later years of the epidemic (van der Kuyl and Berkhout, 2012). This stability of the peculiar base composition of the HIV-1 genome strongly impacts its synonymous codon and codon pair usage, as well as dinucleotide frequencies.

All organisms share the same genetic code, in which four different nucleotides generate codon triplets – such that 4³, or 64, different codons are possible. These 64 codons encode 20 different amino acids and 3 translation stop codons. Of these 20 amino acids, 18 are encoded by more than one synonymous codon, and only methionine and tryptophan are encoded by a unique triplet codon. The ratios of synonymous codons are highly non-random, i.e., some synonymous codons appear more frequent than others (Grantham et al., 1980), a phenomenon termed codon usage bias. Codon usage differs among different species, strongly suggesting that codon usage is an adaptive trait affected by selective pressure and random drift. Another bias that can be observed in organism genomes is that codon pair frequencies are not random, which is termed codon-pair bias. The frequencies of codon-pairs can be different from what would be expected based on the individual codon usage bias of a given genome, as reviewed by Alexaki et al. (2019). Indeed, codon-pair bias have been also described in other organisms, including HIV-1 (Martrus et al., 2013).

In addition to the translation and abundance of isoaccepting tRNAs, synonymous codon mutations can impact many other molecular phenotypes, including transcription modifications (Zhou et al., 2016; Findlay et al., 2018), translation initiation (Kudla et al., 2009; Goodman et al., 2013; Stergachis et al., 2013), translation elongation (Sorensen et al., 1989; Boel et al., 2016), translation accuracy (Akashi, 1994; Drummond and Wilke, 2008), RNA stability (Presnyak et al., 2015), RNA structure and folding (Shabalina et al., 2006; Kudla et al., 2009), RNA splicing (Pagani et al., 2005; Takata et al., 2018), RNA toxicity (Mittal et al., 2018), cotranslational folding (Pechmann and Frydman, 2013), chromatin organization (Warnecke et al., 2009), enhancer functions (Lin et al., 2011; Birnbaum et al., 2012), and microRNA targeting (Brest et al., 2011; Birnbaum et al., 2012). These impacts of synonymous mutations on cell phenotype further indicate that the distribution of synonymous substitutions throughout genes and genomes is neither random nor neutral, and is thus subjected to selective forces. As mentioned above, HIV-1 is a good example since, although the genome is highly variable, the genomic base composition has been tremendously stable over time.

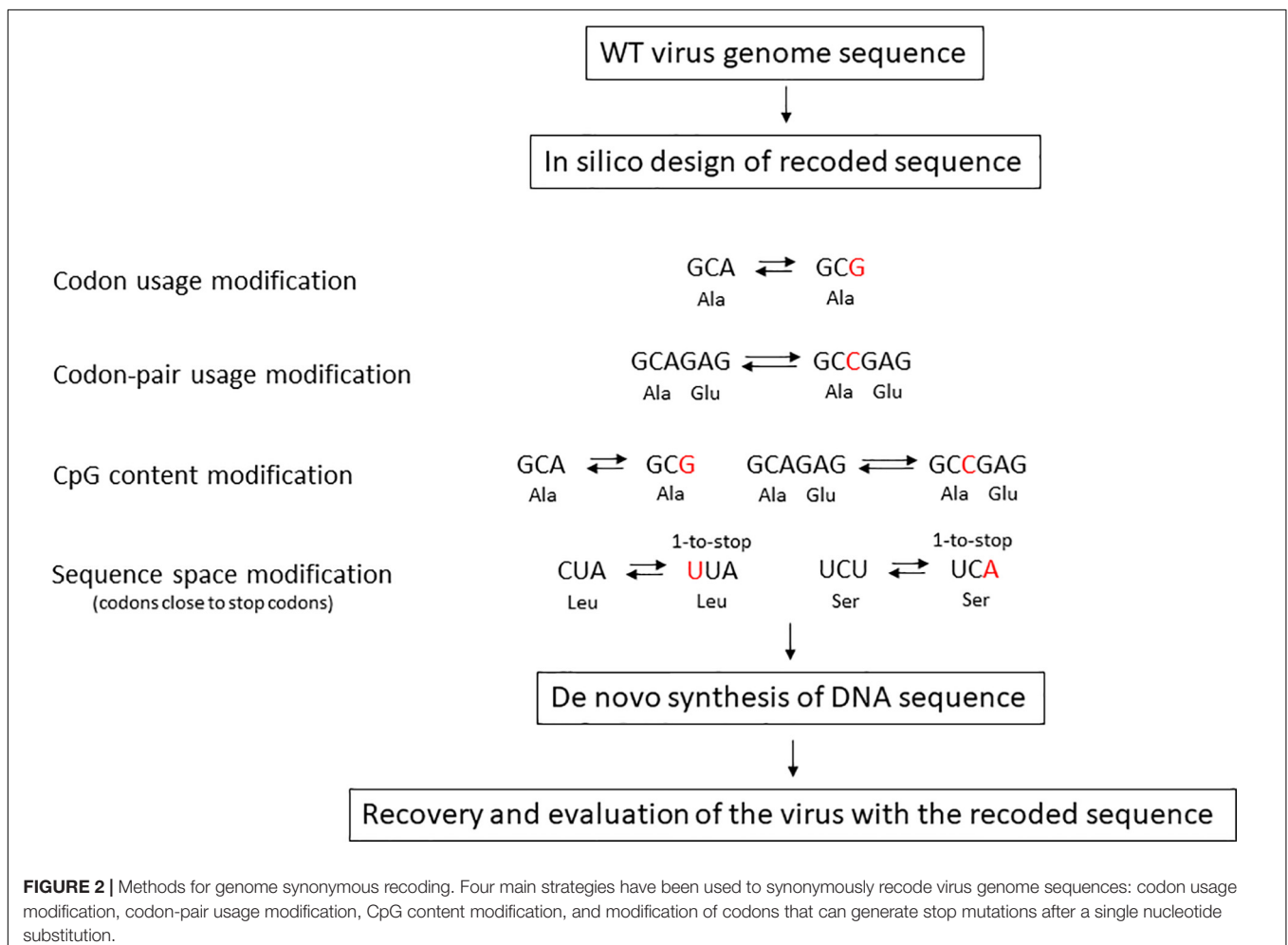
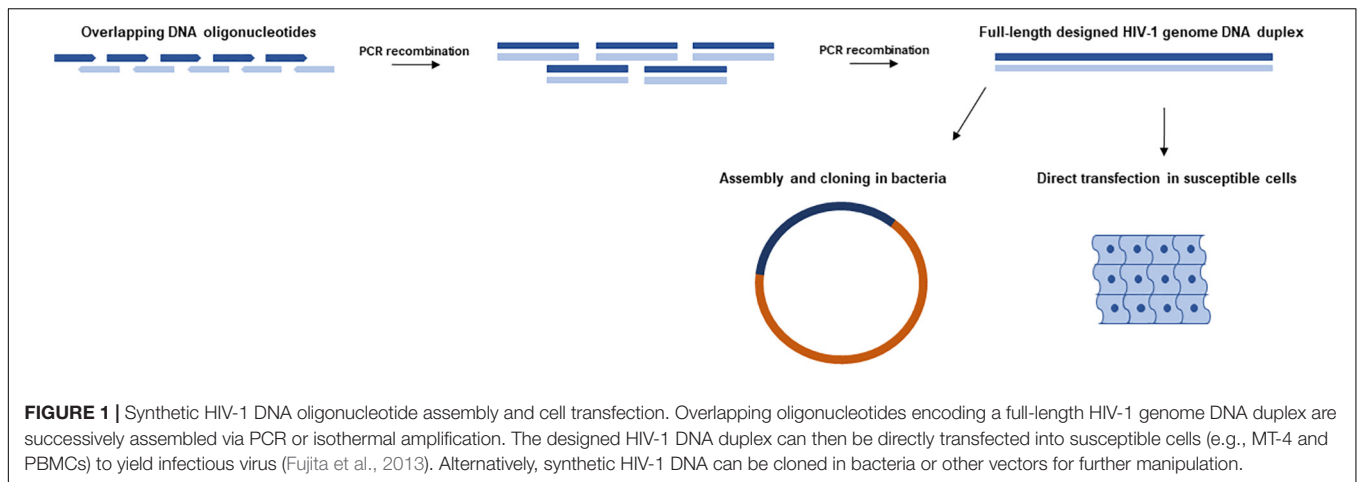
Forty years ago, the invention of PCR and the chemical synthesis of DNA oligonucleotides opened the door to synthetic genomics. Nowadays, DNA chemical synthesis enables the synthesis of 200-nucleotide-long oligonucleotides (Hughes and Ellington, 2017). Overlapping single-stranded DNA oligonucleotides of 100–200 nucleotides in length can easily be assembled by PCR or isothermal amplification to generate DNA fragments of 1,000–2,000 base pairs (bps) (Figure 1). Conventional cloning methods can be applied to clone these synthetic DNA fragments into a bacterial plasmid vector, and individual clones can be isolated and sequenced by Sanger sequencing. Virus genome recoding is a recent tool that is enabling us to elucidate fundamental aspects of virus biology (Figure 2). Synthetic recoding can also help us to develop better therapeutic tools, such as new synthetic vaccines and virus-based gene therapy vectors. Next, we will discuss how synthetic HIV-1 synonymous genome recoding (Table 1) is uncovering HIV-1 biology, and opening the door to new therapeutic opportunities.

Remarkably relevant is how HIV-1 synonymous genome recoding has allowed the description of previous unknown functions of cell factors involved in the innate immune response (Li et al., 2012; Takata et al., 2017).

SYNONYMOUS SUBSTITUTIONS AND HIV-1 REPLICATION CAPACITY

A well-known and common application of synonymous nucleotide recoding is synonymous codon optimization to increase protein expression in various systems (Mauro and Chappell, 2014). On the other hand, an interesting and less known application of synonymous nucleotide recoding is to synonymously deoptimize codon usage, codon-pair usage, or dinucleotide frequencies to reduce protein expression and attenuate virus replication capacity, which has been described for several RNA viruses (Figure 3; Martínez et al., 2016, 2019). Recoding viral genomes through numerous synonymous but suboptimal substitutions represents a new source of live attenuated vaccine candidates. In pioneer research with poliovirus, the introduction of 542 synonymous substitutions among the 2,555 nucleotides of the virus capsid region reduced the virus replication capacity by up to 98% in HeLa cells (Burns et al., 2006). A similar approach involving synonymous deoptimization of the poliovirus capsid coding region generated a virus that exhibited a neuro-attenuated phenotype in transgenic mice (Mueller et al., 2006). Large-scale synonymous codon usage recoding has been used to generate prototypes of live attenuated vaccines for several RNA viruses, including poliovirus (Burns et al., 2006; Mueller et al., 2006), influenza virus (Mueller et al., 2010; Yang et al., 2013), respiratory syncytial virus (RSV) (Nouën et al., 2014), vesicular stomatitis virus (Wang et al., 2015), porcine reproductive and respiratory syndrome virus (Ni et al., 2014), dengue virus (Shen et al., 2015), zika virus (Li et al., 2018), echovirus 7 (Fros et al., 2017), foot and mouth disease virus (Diaz-San Segundo et al., 2021), and the plant cucumber mosaic virus (Mochizuki et al., 2018); arboviruses, such as Chikungunya virus (Nougairede et al., 2013) and tick-borne encephalitis virus (de Fabritus et al., 2015); and DNA viruses, such as Marek's disease herpesvirus (Conrad et al., 2018; Eschke et al., 2018). Clinical trials have been performed using codon-deoptimized type 2 poliovirus. These trials found the vaccine candidate to be safe and immunogenic in infants and toddlers (Van Damme et al., 2019; Konopka-Anstadt et al., 2020; De Coster et al., 2021; Sáez-Llorens et al., 2021).

The HIV-1 genome has been also synonymously deoptimized (Martrus et al., 2013; Klaver et al., 2017). In one study, up to 118 substitutions were introduced in the 1,508-nucleotide structural Gag-coding region, generating viruses with lower replicative capacities in an established MT-4 cell line and in peripheral blood mononuclear cells (PBMCs) from uninfected donors (Martrus et al., 2013). The replication capacity of these recoded variants was reduced up to 39 and 85% in MT-4 cells and PBMCs, respectively (Martrus et al., 2013). Similarly, the introduction of 41 substitutions in protease coding region also generated a virus



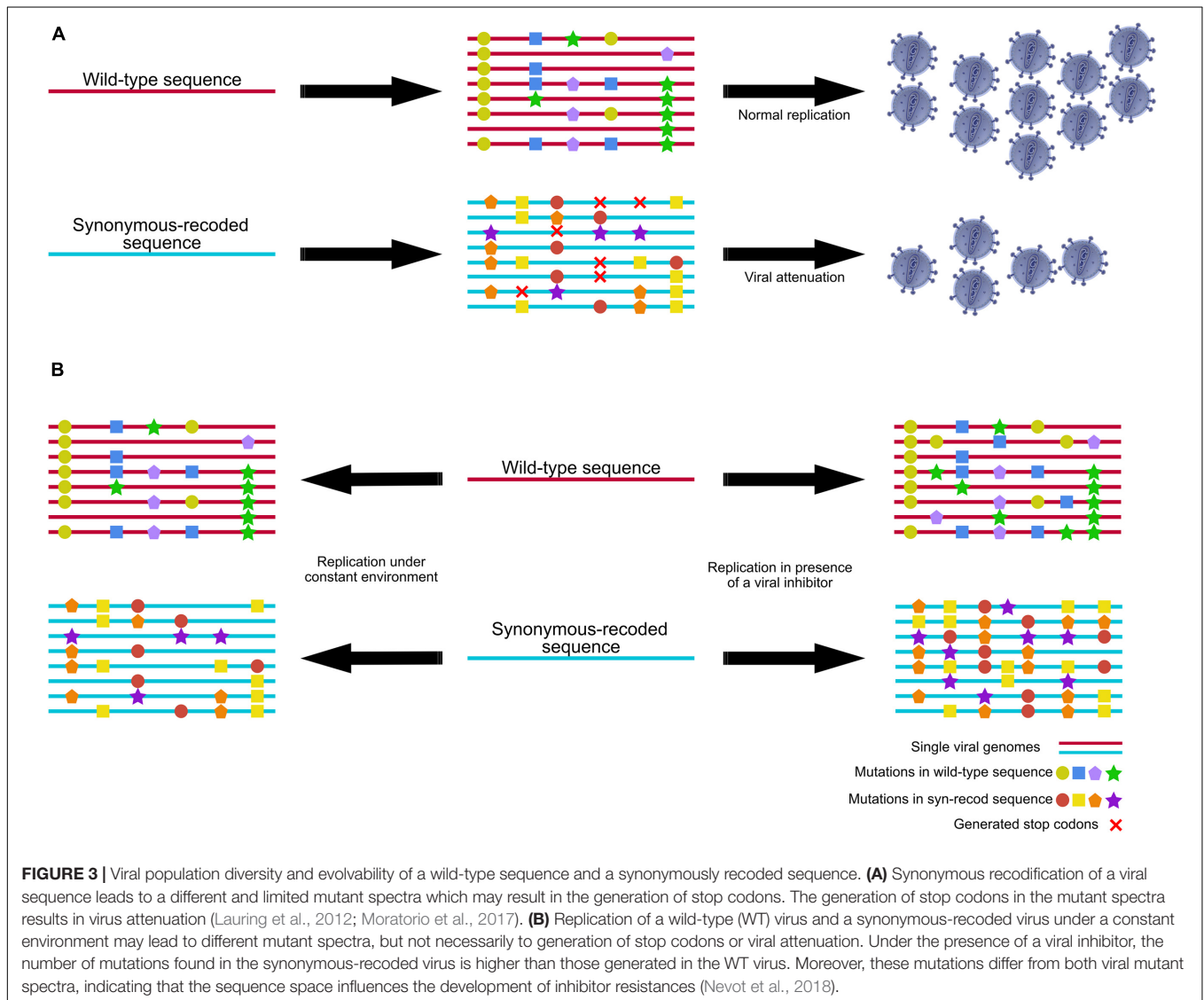
with reduced replication. To test the phenotypic stability of these protease and *gag* variants, the viruses were serially propagated in MT-4 cells, revealing that all deoptimized viruses recovered wild-type (WT) replication capacity after 60 days of cell culture propagation (Martrus et al., 2013). Individual virus clones were obtained after cell culture passages, and sequencing revealed

that several deoptimizing synonymous substitutions had reverted to WT; many additional synonymous and non-synonymous mutations were also detected. The clinical development of an attenuated HIV-1 vaccine is thus improbable. However, these experiments confirm the rapid evolution of an RNA virus, and the necessity of rigorous experiments to evaluate the stability of

TABLE 1 | Examples of HIV-1/SIV phenotype modification by genome synonymous substitutions.

Targeted Gene	Recoding method	Implicated host factor	Resulting phenotype	References
<i>gag</i>	Codon usage	Innate response (SLFN11)	Translation inhibition (lethality)	Li et al., 2012
<i>gag, pol</i>	Codon-pair usage	ND	Translation inhibition (attenuation, lethality)	Martrus et al., 2013
<i>env</i>	Codon usage	Innate response (Type 1 interferon)	Transcription inhibition (attenuation)	Vabret et al., 2014
<i>gag, pol, env</i>	CpG content	Innate response (ZAP)	Transcription inhibition (attenuation, lethality)	Takata et al., 2017
<i>pol</i>	Codon-pair usage	None	Sequence space modification (none)	Nevot et al., 2018
<i>gag, pol, env</i>	CpG content	mRNA splicing	Suppression of splice sites (attenuation, lethality)	Takata et al., 2018
<i>env</i>	Codon usage	mRNA splicing	Suppression of essential mRNA structures (attenuation, lethality)	Jordan-Paiz et al., 2020

ND, not determined.



all candidates for new attenuated virus vaccines based on genome synonymous recoding.

In another lentivirus, simian immunodeficiency virus (SIV), the introduction of 169 synonymous nucleotide optimizing mutations in *gag* and *pol* yielded a virus with a 100-fold decrease

of its replication capacity (Vabret et al., 2014). Interestingly, the recoded virus exhibited a reduced ability to stimulate type I interferon, which may have attenuated its pathogenic potential. Analogously, synonymous deoptimization of the *Streptococcus pneumoniae* pneumolysin gene with underrepresented codon

pairs resulted in an attenuated phenotype in a mouse pulmonary infection model, which was associated with a markedly reduced inflammatory response in the lungs (Coleman et al., 2011).

Reports have also described *cis*-acting functions in the HIV-1 coding sequence for viral gene expression. Synonymous adaptive mutations in the HIV-1 3' *pol* gene result in parallel increases or decreases in the expression levels of late viral proteins and in viral replication capacity (Nomaguchi et al., 2014). These findings suggest that viral fitness is altered through nucleotide-dependent modulation of the expression pattern of viral mRNAs. A global silent mutagenesis experiment was performed to identify new *cis*-acting RNA elements in the HIV-1 genome that are important for virus replication (Takata et al., 2018). Sixteen mutant proviruses were designed and synthesized, which contained clusters of ~50 to ~200 synonymous mutations spanning nearly the entire HIV-1 protein coding sequence. These mutant viruses were analyzed and categorized into three phenotypic groups: (1) mutants exhibiting near WT replication, (2) mutants exhibiting replication defects accompanied by perturbed RNA splicing, and (3) mutants exhibiting replication defects without obvious splicing perturbation (Takata et al., 2017). Mutants of the second group generally contained point mutations that reduced proximal splice site utilization. Mapping the changes responsible for splicing perturbations in these viruses revealed several RNA sequences that apparently suppressed the use of cryptic or canonical splice sites. These findings indicated complex negative regulation of HIV-1 splicing via RNA elements in various regions of the HIV-1 genome, which maintained a balance between splicing and viral replication (Takata et al., 2018). Overall, these experiments demonstrated that synonymous HIV-1 recoding may provide insights into uncharacterized elements in the HIV-1 genome that determine the fate and splicing of HIV-1 RNA and thus the ability of HIV-1 to replicate.

SYNONYMOUS SEQUENCE SPACE

One unknown aspect of the genetic architecture of RNA viruses, including HIV-1, is how codon choice influences population diversity and evolvability. Early comparisons of the nucleotide sequences of homologous genes revealed higher numbers of synonymous substitutions compared to non-synonymous mutations, promoting an initial assumption that synonymous mutations were selectively neutral. This postulation contributed to the foundations of the neutral theory of molecular evolution, in which organisms evolve mainly through the random drift of genomes carrying neutral or quasi-neutral mutations (King and Jukes, 1969; Kimura, 1983). However, although genome random drift may play an important role in molecular evolution, the currently available evidence indicates that synonymous mutations are not neutral (Martínez et al., 2019; Domingo, 2020).

Synonymous substitutions can determine the evolutionary trajectory of a genome. Different codons that encode the same amino acid can have different evolutionary potential in terms of the amino acids that they can access through a point mutation, which can determine their likelihood of reaching beneficial mutations that facilitate adaptation (Lauring et al.,

2013). Research with polio, coxsackie B3, and influenza A viruses has revealed that synonymous recoding of the virus genome can change its starting position in sequence space and limit its access to mutational neighborhoods (Lauring et al., 2012; Moratorio et al., 2017), potentially resulting in virus attenuation. Mutational neighborhoods refer to a network of variants organized in sequence space around a single master sequence (Domingo, 2020). In the cases of coxsackie B3 and influenza A viruses, leucine and serine codons were recoded to favor the possibility of nonsense mutations resulting in stop codons. The virus variants were attenuated *in vivo* and exhibited increased numbers of stop codons. These findings suggested the possibility of changing a virus' starting position in sequence space, and redirecting it toward detrimental mutational neighborhoods to generate self-limiting vaccine strains (Moratorio et al., 2017). Similarly, a synonymously recoded poliovirus exhibited unique mutant spectra, showing significantly different distributions of polymorphic amino acid substitutions in the capsid (Lauring et al., 2012). This recoded virus exhibited normal replication capacity in tissue culture, but displayed an attenuated phenotype in an animal model of infection, demonstrating the importance of mutant neighborhoods in determining viral pathogenesis.

To explore whether the synonymous sequence space influences the development of HIV-1 protease inhibitor (PI) resistance, WT HIV-1 was compared to a variant carrying a protease gene with 38 synonymous mutations (13% of the protease sequence) (Nevot et al., 2018). The 38 synonymous substitutions were scattered throughout the protease coding region, and were selected to improve protease gene codon-pair bias without modifying the codon bias or folding free energy (Martrus et al., 2013). Importantly, replication in MT-4 cells or PBMCs was indistinguishable between the recoded variant and the WT virus. Similar to the studies performed with poliovirus and coxsackie virus, this investigation was designed to explore how HIV-1 evolvability was influenced by the natural position in sequence space. In contrast to previous work, this study explored how synonymous substitutions affected the specific selection pressure targeting a precise HIV-1 gene. Interestingly, upon PI treatment, the recoded virus displayed different patterns of resistance mutations, demonstrating that sequence space position affects evolvability. However, although the WT and recoded proteases occupied different sequence spaces, they showed similar levels of development of phenotypic resistance to PIs – i.e., the recoded and WT virus exhibited the same robustness to overcome a specific selective pressure (Nevot et al., 2018). Interestingly, the recoded virus showed significantly higher population diversity in the recoded and targeted gene following propagation in both the absence and presence of PIs. It is tempting to speculate that the recoded virus was subjected to greater pressure to change or revert to a WT-synonymous background. As discussed above, positioning a virus in a detrimental sequence space to reduce its capacity to produce fit progeny may be a new strategy for attenuated poliovirus vaccine development (Lauring et al., 2012). However, the research involving the HIV-1 protease (Nevot et al., 2018) suggests that this approach must be cautiously developed, with careful testing

of the long-term stability of synonymously recoded individual candidate viruses.

SYNONYMOUS SUBSTITUTIONS AND HIV-1 RNA SECONDARY STRUCTURE

RNA virus genomes contain multiple functional RNA elements that are required for translation or RNA replication. Synonymous genome recoding has been exploited to identify specific RNA structures required for virus replication. Recoding does not affect virus growth unless it destroys the sequence/structure of a functional RNA element (Song et al., 2012). SHAPE experiments have revealed that in HIV-1 RNA, individual nucleotides exhibit widely divergent tendencies to be base-paired (Wilkinson et al., 2008; Watts et al., 2009). HIV-1 RNA secondary structures are conserved between strains, and thus might have a function in HIV-1 replication. As previously discussed, synonymous recoding of the HIV-1 genome has elucidated the presence of unknown RNA *cis*-regulatory elements that influence balanced splicing and viral replication (Takata et al., 2018). Using dimethyl sulfate mutational profiling and sequencing (DMS-MaPseq) to investigate the HIV-1 RNA structure in cells has revealed that the same RNA sequence can assume alternative conformations (Tomezsko et al., 2020). These findings have revealed heterogeneous regions of RNA structure across the entire HIV-1 genome, as well as alternative conformations at critical splice sites that influence the ratio of transcript isoforms. Overall, these results strongly suggest that HIV-1 RNA conformation regulates splice-site use and viral gene expression (Tomezsko et al., 2020).

In SIV and HIV-1, the expression of Env depends on the nature of the codons used (Shin et al., 2015). One interesting hypothesis is that in different families of persistent viruses, codon usage is skewed in a distinctive manner to enable temporal regulation of late-expressed structural gene products, which is the case for HIV-1 Env. Temporal regulation of the lentiviral Env protein ensures its production late in the lytic cycle of these persisting viruses. Notably, expression of these late gene products is typically induced by viral transducers produced earlier in the viral replication cycle. One example of such a transducer is the HIV-1/SIV protein Rev. A nuclear localization signal encoded in the *rev* gene enables localization of the Rev protein to the nucleus, where it participates in the export of unspliced and incompletely spliced mRNAs. In the absence of Rev, mRNAs of the late (structural) HIV-1/SIV genes (e.g., Env protein) are retained in the nucleus, preventing their translation (Karn and Stoltzfus, 2012).

Env exhibits unusual codon usage that differs from that of the host cell. It was recently demonstrated that Rev induction of Env protein expression is dependent on this biased codon usage (Shin et al., 2015). To determine whether codon usage affected HIV-1 Env protein expression and virus viability, the codons AGG, GAG, CCU, ACU, CUC, and GGG of the HIV-1 *env* gene were substituted with the synonymous codons CGU, GAA, CCG, ACG, UUA, and GGA, respectively (Jordan-Paiz et al., 2020). This approach revealed that synonymous recoding

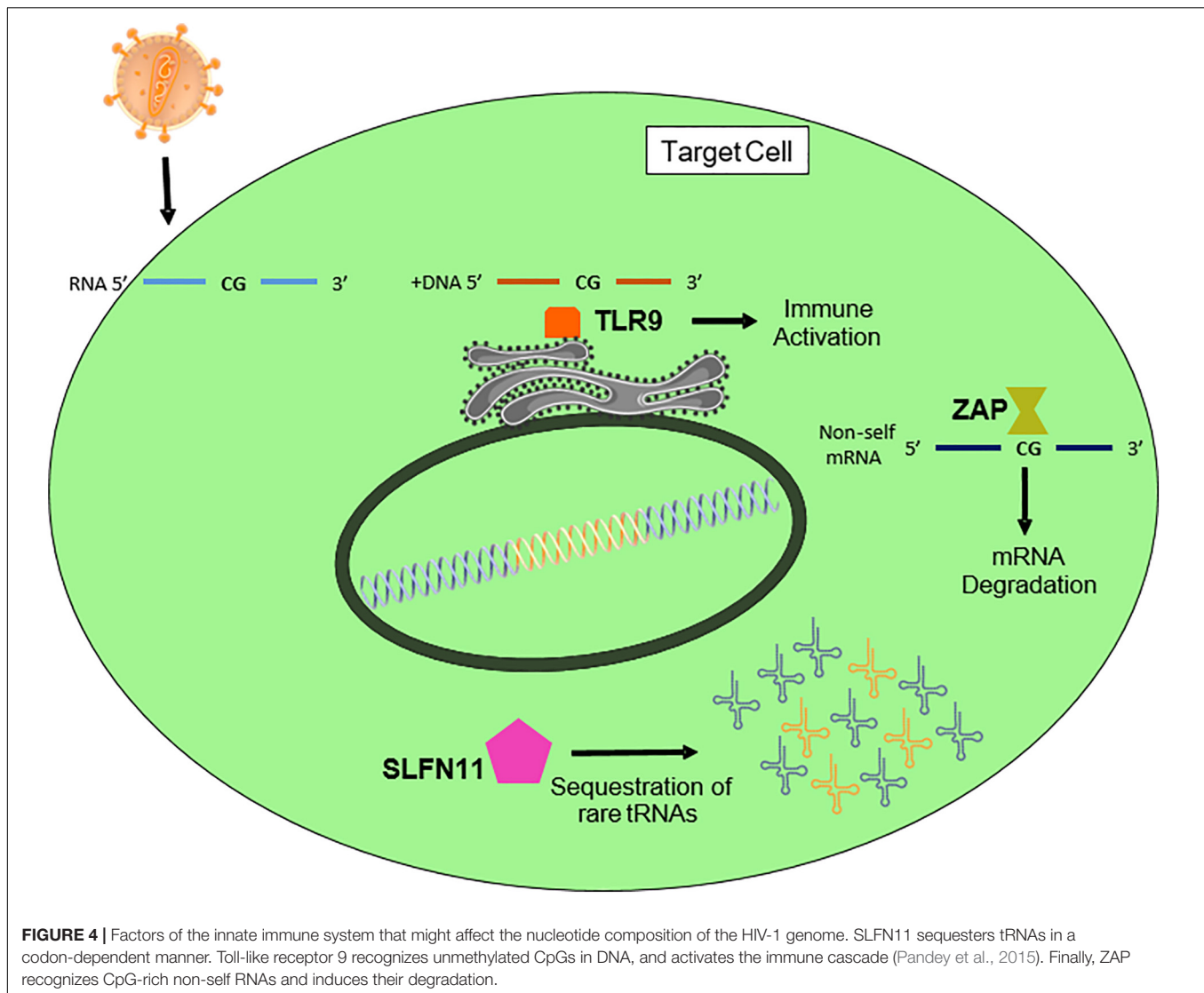
of the Env protein gp120 coding region did not significantly affect virus replication capacity, despite the introduction of 15 new CpG dinucleotides (see in the next section the relevance of CpGs). In contrast, changing a single codon (AGG to CGU) within the gp41 coding region (HXB2 *env* position 2,125–2,127), which is located in the intronic splicing silencer (ISS), completely abolished virus replication and Env expression (Jordan-Paiz et al., 2020). Computational analyses of this mutant revealed severe disruption of the ISS RNA secondary structure. Moreover, a variant that restored the ISS secondary RNA structure also re-established Env production and virus viability. These findings indicate that external ISS loop disruption strongly affected HIV-1 replication and Env translation – again highlighting the relevance of synonymous recoding in maintaining biologically relevant RNA structures.

HIV-1 DINUCLEOTIDE FREQUENCIES AND INNATE IMMUNE RESPONSE

As previously discussed, lentiviral RNA genomes (e.g., HIV-1 and SIV) exhibit a strong bias in their nucleotide composition, with high frequencies of A and low content of C (van der Kuyl and Berkhout, 2012). In accordance with the nucleotide composition, a biased dinucleotide frequency is also observed in HIV-1, with the most frequent occurrence of the dinucleotide ApA and a lower-than-expected proportion of CpG. Reduced CpG frequencies are also observed in many other RNA viruses, and in most vertebrate genomes. The low CpG abundance in vertebrate genomes can be explained by the methylation/deamination process that promotes the mutation of CpG to TpG/CpA (Karlin and Mrázek, 1997). In contrast to HIV-1, although many RNA viruses mimic the CpG suppression of their vertebrate hosts (Simmonds et al., 2013), this phenomenon cannot be explained by methylation since they lack a DNA intermediate. Thus, the explanation for the low CpG abundance in RNA viruses remains controversial.

Studies of several RNA viruses, including HIV-1, have revealed that increases of CpG in the virus genome negatively impact their replication capacity (Burns et al., 2009; Atkinson et al., 2014; Tulloch et al., 2014; Gaunt et al., 2016; Antzin-Anduetza et al., 2017; Takata et al., 2017). In HIV-1-infected individuals, the *in vivo* generation of *de novo* CpG sites carries twice the fitness cost of mutations that do not generate CpG sites (Theys et al., 2018). These findings raise questions regarding why CpG sites are suppressed in the HIV-1 genome, and why an increase of this dinucleotide carries a fitness cost.

Toll-like receptor 9 (TLR9) recognizes bacterial and viral DNA that is rich in unmethylated CpG DNA (Figure 4; Pandey et al., 2015). TLR9 is highly expressed in plasmacytoid dendritic cells. Since the replication of HIV-1 in plasmacytoid dendritic cells is not fully demonstrated, the role of TLR9 in HIV-1 CpG suppression is also unclear. Recently, Takata and colleagues generated several CpG-rich HIV-1 variants through the synthetic random synonymous mutagenesis of different HIV-1 regions (Takata et al., 2017). As previously described (Atkinson et al., 2014; Gaunt et al., 2016; Antzin-Anduetza et al., 2017), they



found that variants containing larger numbers of CpG sites had significantly lower replication capacity. By targeting different cellular proteins implicated in mRNA degradation, they revealed that knockdown of a zinc-finger antiviral protein (ZAP) restored the normal replication capacity of CpG-rich HIV-1 variants (Figure 4). ZAP is a factor involved in the innate immune response, which was first described as an inhibitor of viral RNA production (Gao et al., 2002). ZAP specifically binds viral mRNAs (Guo et al., 2004) and prevents their cytoplasmic accumulation by recruiting and promoting degradation of the RNA processing exosome (Guo et al., 2007) – thereby promoting specific loss of cytoplasmic viral mRNA without affecting nuclear mRNA. While previous studies could not determine how ZAP targets viral mRNA (Guo et al., 2004), Takata and colleagues demonstrated that ZAP showed specificity for CpG-rich HIV-1 mRNA exonic regions. These authors suggested that RNA virus evolution may have favored low CpG content to avoid selective inhibition by ZAP. Examination of another RNA virus has corroborated the

relationship between CpG frequency, ZAP, and virus replication capacity (Odon et al., 2019).

Since the demonstration that ZAP induces the degradation of viral CpG-rich mRNA, several studies have focused on the mechanism underlying this binding and inhibition. As previously described, ZAP recruits the RNA processing exosome (Guo et al., 2007) and this complex degrades CpG-rich viral mRNAs. However, the regulation of ZAP's antiviral activity is not completely understood, as other cellular cofactors might be involved in its activation. The E3 ubiquitin ligase TRIM25 reportedly enhances ZAP and is required for its activity (Li et al., 2017; Zheng et al., 2017). ZAP interacts with several components of the RNA exosome complex, suggesting that TRIM25 may not be the only cofactor that mediates ZAP activity. Indeed, it was recently demonstrated that KHNYN, a newly described cytoplasmic human protein, is also essential for ZAP activity against foreign CpG-rich mRNAs (Ficarelli et al., 2019b). This study revealed how KHNYN interacted with ZAP, and that

its overexpression dramatically reduced the infectivity capacity of a synonymous recoded CpG-rich HIV-1 variant. Moreover, KHNYN inhibition enabled this CpG-rich variant to reach WT replication levels.

The N-terminal domain of ZAP includes four CCCH-type zinc-finger motifs that together form the ZAP RNA-binding domain (RBD), and are responsible for targeting viral mRNA (Chen et al., 2012). A recent paper describes the RBD structure, and how it selectively binds CpG-rich mRNA sequences (Meagher et al., 2019). Structural analysis of ZAP revealed a pocket on the protein surface, which only binds CpG dinucleotides (Meagher et al., 2019). This ZAP structure may explain how ZAP avoids low-CpG host mRNA and recognizes foreign mRNAs. Importantly, another recent publication describes the molecular mechanism through which ZAP only detects CpG sites in a single-stranded form (Luo et al., 2020). The four zinc-finger motifs of ZAP form a specific architecture that enables extensive interactions with RNA. They further revealed that an RNA containing several ZAP-binding sites can be recognized by multiple ZAP molecules, which likely enhances the activation of the exosome complex. The authors hypothesize that a single CpG dinucleotide cannot bind ZAP and activate the exosome complex, but rather must be single-stranded and surrounded by other CpG sequences that can bind several ZAP molecules, potentially exerting a synergistic effect.

Despite recent descriptions of the ZAP structure and molecular mechanism, there remains some controversy surrounding how CpG-rich mRNAs are targeted. Investigations of the genomes of different primate lentiviruses do not reveal any correlation between the number of CpG sites and the sensitivity to ZAP. As previously noted, the HIV-1 genome contains a very low proportion of CpG sites. In contrast, HIV-2 shows less CpG suppression. However, although the CpG content is higher in HIV-2 than in HIV-1, no ZAP inhibition is observed in HIV-2 (Kmiec et al., 2020). This phenomenon has also been observed for another lentivirus, SIVmac, which exhibits a higher CpG content but significantly lower ZAP inhibition compared to HIV-1. This raises the question of why ZAP can inhibit HIV-1, but not other related lentiviruses.

It is hypothesized that a region in the genome, termed ZAPsen, may confer ZAP sensitivity. In HIV-1, this ZAPsen region is described as located in the 5' region of the *env* gene (from position 6,239–6,947 of the HIV-1 HXB2 reference genome) (Kmiec et al., 2020). And increase of synonymous CpG sites in ZAPsen confers higher ZAP sensitivity and lower virus replication capacity (Ficarelli et al., 2019a; Kmiec et al., 2020). However, this is not the only factor that affects replication. ZAP sensitivity is also increased by the introduction of other ZAPsen mutations that alter the RNA secondary structure, without modifying the number of CpGs. Thus, ZAP may detect the ZAPsen region based on either the number of CpG sites, or changes in the RNA secondary structure. ZAP shows greater affinity to ZAPsen when higher amounts of CpGs are introduced, but other factors may also explain why CpG suppression is observed along the whole HIV-1 genome. Accordingly, synonymous mutations that increase the number of CpGs in other HIV-1 regions also lead to reduced replication capacity (Ficarelli et al., 2019a).

These variants were not inhibited by ZAP, but rather through other ZAP-independent mechanisms, mainly related to pre-mRNA splicing. As described in a previous section, synonymous mutations can also alter splicing (Takata et al., 2018). In this case, synonymous CpGs were responsible for aberrant mRNA splicing and reduced viral fitness.

As discussed above, CpG suppression in HIV-1 might be partly explained by the actions of the innate response factor ZAP. However, other mechanisms may also contribute, including mRNA splicing and mRNA secondary structure. The exact mechanism of ZAP targeting and binding remains to be elucidated. This phenomenon may be further studied through the generation of different HIV-1 variants with increasing amounts of CpG sites in different regions of the genome, which could also be useful for identifying new factors that might be involved in the inhibition of HIV-1 replication.

HIV-1 CODON USAGE AND INNATE IMMUNE RESPONSE

HIV-1-biased nucleotide composition can produce overstimulation of the type I interferon, suggesting that RNA sequences may be discriminated based on their nucleotide composition (Vabret et al., 2012). Type I interferon is a major antiviral cytokine that contributes to chronic immune system activation and progression to AIDS during HIV infection (Jacquelin et al., 2009). Pathogenicity is reportedly correlated with divergent nucleotide composition of HIV-1 compared to host (Vabret et al., 2012), suggesting that virus-host interactions might be altered by artificially changing the nucleotide frequency of the HIV-1 genome. With this aim, SIV codon usage was optimized to be closer to the average nucleotide composition of the SIV macaque host (Vabret et al., 2014). A synthetic SIV optimized with 169 synonymous mutations in *gag* and *pol* exhibited a 100-fold decrease of replicative capacity. Interestingly, this optimized variant also exhibited reduced ability to stimulate type I interferon in infected human and macaque PBMCs (Vabret et al., 2014). No reversion of the introduced mutations was observed after ten serial cell passages, suggesting that this variant may be a safe candidate for an attenuated vaccine. Still, further experiments should be performed to confirm the stability of this attenuated variant. Live attenuated SIV vaccines are highly protective in the macaque AIDS model (Koff et al., 2006). However, in addition to safety concerns, this optimized SIV variant raises intriguing questions related to the fact that a type I interferon response is necessary for a vaccine to shape adaptive immune responses and memory.

The Schlafen (SLFN) gene family was first discovered in 1998, as regulators of T-cell maturation. The name Schlafen means “to sleep” in German, and was chosen based on the observation that enhanced SLFN1 expression resulted in G₀/G₁ cell cycle arrest (Schwarz et al., 1998). SLFN genes are categorized as interferon-stimulated genes (ISGs), as their expression is induced by type I interferon. Some SLFN proteins possess RNA cleavage activity, and exhibit antiviral activity against RNA and DNA viruses. In particular, SLFN11 potently and selectively inhibits

HIV-1 protein translation and virus production (Li et al., 2012), and is considered an HIV-1 restriction factor. SLFN11 also inhibits other retroviruses, including murine leukemia virus and feline immunodeficiency virus (Stabell et al., 2016), such that its antiretroviral effect has been classified as host-specific but virus-independent.

Interestingly, the inhibitory effect of SLFN11 on viral protein expression is intrinsic to the transcripts. SLFN11 acts at the point of virus protein synthesis by exploiting the unique viral codon bias toward A/U nucleotides, sequestering tRNAs in a codon-dependent manner (Figure 4). Accordingly, SLFN11 only affects WT HIV-1, and does not recognize synonymously recoded viruses in which the HIV-1 structural *gag* sequences are optimized for human codon usage (i.e., without A/U in the third position) (Li et al., 2012). Again, and similar to the findings obtained with SIV, HIV-1 replication was attenuated by codon optimization for host translation. This model is in line with findings that SLFN11's antiviral activity extends to other viruses with rare codon bias (e.g., influenza, which also has a high A content) but not to adeno-associated virus or herpes simplex virus (Li et al., 2012). Overall, investigations in SIV and HIV-1 have shown that innate immunity restricts sequence landscapes by targeting specific sequences and sequence patterns that are primarily found in pathogens (Vabret et al., 2017).

CONCLUSION

Synonymous rewriting of the HIV-1 genome is helping to elucidate essential genome functions. However, mammalian codon optimization is not straightforward, since synonymous mutations are often not neutral. On the other hand, intentional deoptimization of codon, codon-pair usage, or dinucleotide

frequencies has been applied in several RNA virus genomes to generate new attenuated vaccines. Nevertheless, the safety and stability of these attenuated vaccines remain to be elucidated. Due to safety concerns, it is very difficult to envision the successful development of an attenuated HIV-1 vaccine. However, recoded HIV-1 variants can be used in gene therapy, as a vaccine vector for immunization against diverse microorganisms, or in immunotherapy to elicit specific innate immune responses to treat particular conditions. Importantly, artificial HIV-1 synonymous recoding has greatly increased our knowledge regarding its interaction with the host. Specifically, HIV-1 synonymous recoding impacts virus RNA nucleotide composition and RNA secondary structure which regulate splice-site use and viral gene expression. Together with structural RNA features, changes in nucleotide composition also affect HIV-1 susceptibility to endogenous cell innate responses and correlated with differences in clinical progression rates, suggesting a potential role of virus RNA nucleotide composition in HIV-1 *in vivo* pathogenicity.

AUTHOR CONTRIBUTIONS

AJ-P, SE, and MM discussed and wrote this review. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Spanish Ministry of Science and Innovation (PID2019-103955RB-I00). AJ-P was supported by a contract from the Spanish Ministry of Science and Innovation (grant BES-2014-069931).

REFERENCES

- Akashi, H. (1994). Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136, 927–935. doi: 10.1093/genetics/136.3.927
- Alexaki, A., Kames, J., Holcomb, D. D., Athey, J., Santana-Quintero, L. V., Lam, P. V. N., et al. (2019). Codon and Codon-Pair Usage Tables (CoCoPUTs): Facilitating Genetic Variation Analyses and Recombinant Gene Design. *J. Mol. Biol.* 431, 2434–2441. doi: 10.1016/j.jmb.2019.04.021
- Antzin-Anduetza, I., Mahiet, C., Granger, L. A., Odendall, C., and Swanson, C. M. (2017). Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* 14:49.
- Atkinson, N. J., Witteveldt, J., Evans, D. J., and 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. doi: 10.1093/nar/gku075
- Birnbaum, R. Y., Clowney, E. J., Agamy, O., Kim, M. J., Zhao, J., Yamanaka, T., et al. (2012). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* 22, 1059–1068. doi: 10.1101/gr.133546.111
- Boel, G., Letso, R., Neely, H., Price, W. N., Wong, K. H., Su, M., et al. (2016). Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529, 358–363. doi: 10.1038/nature16509
- Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., et al. (2011). 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.* 43, 242–245. doi: 10.1038/ng.762
- Burns, C. C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J., and Kew, O. (2009). Genetic Inactivation of Poliovirus Infectivity by Increasing the Frequencies of CpG and UpA Dinucleotides within and across Synonymous Capsid Region Codons. *J. Virol.* 83, 9957–9969. doi: 10.1128/jvi.00508-09
- Burns, C. C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J., et al. (2006). Modulation of Poliovirus Replicative Fitness by Increasing the Frequencies of Synonymous Codon Usage in the Capsid Region. *J. Virol.* 80, 3259–3272. doi: 10.1128/jvi.80.7.3259-3272.2006
- Chen, S., Xu, Y., Zhang, K., Wang, X., Sun, J., Gao, G., et al. (2012). Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. *Nat. Struct. Mol. Biol.* 19, 430–435. doi: 10.1038/nsmb.2243
- Coffin, J., and Swanstrom, R. (2013). HIV pathogenesis: Dynamics and genetics of viral populations and infected cells. *Cold Spring Harb. Perspect. Med.* 3:a012526. doi: 10.1101/cshperspect.a012526
- Coleman, J. R., Papamichail, D., Yano, M., Del Mar, García-Suárez, M., and Pirofski, L. A. (2011). Designed reduction of *Streptococcus pneumoniae* pathogenicity via synthetic changes in virulence factor codon-pair bias. *J. Infect. Dis.* 203, 1264–1273. doi: 10.1093/infdis/jir010
- Conrad, S. J., Silva, R. F., Hearn, C. J., Climans, M., and Dunn, J. R. (2018). Attenuation of Marek's disease virus by codon pair deoptimization of a core gene. *Virology* 516, 219–226. doi: 10.1016/j.virol.2018.01.020

- De Coster, I., Leroux-Roels, I., Bandyopadhyay, A. S., Gast, C., Withanage, K., Steenackers, K., et al. (2021). Safety and immunogenicity of two novel type 2 oral poliovirus vaccine candidates compared with a monovalent type 2 oral poliovirus vaccine in healthy adults: two clinical trials. *Lancet* 397, 39–50. doi: 10.1016/s0140-6736(20)32541-1
- de Fabritus, L., Nougairède, A., Aubry, F., Gould, E. A., and de Lamballerie, X. (2015). Attenuation of Tick-Borne Encephalitis Virus Using Large-Scale Random Codon Re-encoding. *PLoS Pathog.* 11:1004738. doi: 10.1371/journal.ppat.1004738
- Diaz-San Segundo, F. L., Medina, G. N., Spinard, E., Kloc, A., Ramirez-Medina, E., Azzinaro, P., et al. (2021). Use of synonymous deoptimization to derive modified live attenuated strains of foot and mouth disease virus. *Front. Microbiol.* 11:610286. doi: 10.3389/fmicb.2020.610286
- Domingo, E. (2020). *Virus as Populations: Composition, Complexity, Dynamics, and Biological Implications*. Amsterdam: Elsevier Inc.
- Drummond, D. A., and Wilke, C. O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352. doi: 10.1016/j.cell.2008.05.042
- Eschke, K., Trimpert, J., Osterrieder, N., and Kunec, D. (2018). Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. *PLoS Pathog.* 14:e1006857. doi: 10.1371/journal.ppat.1006857
- Fiorelli, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A. E., Sertkaya, H., Wilson, H., et al. (2019a). CpG Dinucleotides Inhibit HIV-1 Replication through Zinc Finger Antiviral Protein (ZAP)-Dependent and -Independent Mechanisms. *J. Virol.* 94:e01337.
- Fiorelli, M., Wilson, H., Galão, R. P., Mazzon, M., Antzin-Anduetza, I., Marsh, M., et al. (2019b). KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife* 8:e46767.
- Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., et al. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature* 562, 217–222. doi: 10.1038/s41586-018-0461-z
- Fros, J. J., Dietrich, I., Alshaiikhahmed, K., Passchier, T. C., Evans, D. J., and Simmonds, P. (2017). CpG and upA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *Elife* 6:e29112.
- Fujita, Y., Otsuki, H., Watanabe, Y., Yasui, M., Kobayashi, T., Miura, T., et al. (2013). Generation of a replication-competent chimeric simian-human immunodeficiency virus carrying env from subtype C clinical isolate through intracellular homologous recombination. *Virology* 436, 100–111. doi: 10.1016/j.virol.2012.10.036
- Gao, G., Guo, X., and Goff, S. P. (2002). Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* 297, 1703–1706. doi: 10.1126/science.1074276
- Gaunt, E., Wise, H. M., Zhang, H., Lee, L. N., Atkinson, N. J., Nicol, M. Q., et al. (2016). Elevation of CpG frequencies in influenza a genome attenuates pathogenicity but enhances host response to infection. *Elife* 5:e12735.
- Goodman, D. B., Church, G. M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479. doi: 10.1126/science.1241934
- Grantham, R., Gautier, C., and Gouy, M. (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* 8, 1893–1912. doi: 10.1093/nar/8.9.1893
- Guo, X., Carroll, J.-W. N., MacDonald, M. R., Goff, S. P., and Gao, G. (2004). The Zinc Finger Antiviral Protein Directly Binds to Specific Viral mRNAs through the CCCH Zinc Finger Motifs. *J. Virol.* 78, 12781–12787. doi: 10.1128/jvi.78.23.12781-12787.2004
- Guo, X., Ma, J., Sun, J., and Gao, G. (2007). The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proc. Natl. Acad. Sci. U S A.* 104, 151–156. doi: 10.1073/pnas.0607063104
- Hughes, R. A., and Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* 9:a023812. doi: 10.1101/cshperspect.a023812
- Jacquelin, B., Mayau, V., Targat, B., Liovat, A. S., Kunkel, D., Petitjean, G., et al. (2009). Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. *J. Clin. Invest.* 119, 3544–3555.
- Jordan-Paiz, A., Nevot, M., Lamkiewicz, K., Lataretu, M., Franco, S., Marz, M., et al. (2020). HIV-1 lethality and loss of Env protein expression induced by single synonymous substitutions in the virus genome intronic splicing silencer. *J. Virol.* 94, 1108–1120e.
- Karlin, S., and Mrázek, J. (1997). Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. U S A.* 94, 10227–10232. doi: 10.1073/pnas.94.19.10227
- Karn, J., and Stoltzfus, C. M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med.* 2:a006916. doi: 10.1101/cshperspect.a006916
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- King, J. L., and Jukes, T. H. (1969). Non-Darwinian evolution. *Science* 164, 788–798. doi: 10.1126/science.164.3881.788
- Klaver, B., van der Velden, Y., van Hemert, F., van der Kuyl, A. C., and Berkhout, B. (2017). HIV-1 tolerates changes in A-count in a small segment of the pol gene. *Retrovirology* 14:43.
- Kmieć, D., Nchioua, R., Sherrill-Mix, S., Stürzel, C. M., Heusinger, E., Braun, E., et al. (2020). CpG frequency in the 5' third of the env gene determines sensitivity of primary HIV-1 strains to the zinc-finger antiviral protein. *MBio* 11:e02903.
- Koff, W. C., Johnson, P. R., Watkins, D. I., Burton, D. R., Lifson, J. D., Hasenkamp, K. J., et al. (2006). HIV vaccine design: Insights from live attenuated SIV vaccines. *Nat. Immunol.* 7, 19–23. doi: 10.1038/nri1296
- Konopka-Anstadt, J. L., Campagnoli, R., Vincent, A., Shaw, J., Wei, L., Wynn, N. T., et al. (2020). Development of a new oral poliovirus vaccine for the eradication end game using codon deoptimization. *NPJ Vaccines* 5:26.
- Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258. doi: 10.1126/science.1170160
- Lauring, A. S., Acevedo, A., Cooper, S. B., and Andino, R. (2012). Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe* 12, 623–632. doi: 10.1016/j.chom.2012.10.008
- Lauring, A. S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* 11, 327–336. doi: 10.1038/nrmicro3003
- Li, M. M. H., Lau, Z., Cheung, P., Aguilar, E. G., Schneider, W. M., Bozzacco, L., et al. (2017). TRIM25 Enhances the Antiviral Action of Zinc-Finger Antiviral Protein (ZAP). *PLoS Pathog.* 13:e1006145. doi: 10.1371/journal.ppat.1006145
- Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., et al. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 491, 125–128. doi: 10.1038/nature11433
- Li, P., Ke, X., Wang, T., Tan, Z., Luo, D., Miao, Y., et al. (2018). Zika Virus Attenuation by Codon Pair Deoptimization Induces Sterilizing Immunity in Mouse Models. *J. Virol.* 92, 701–718e.
- Lin, M. F., Kheradpour, P., Washietl, S., Parker, B. J., Pedersen, J. S., and Kellis, M. (2011). Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.* 21, 1916–1928. doi: 10.1101/gr.108753.110
- Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G., et al. (2020). Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. *Cell Rep.* 30, 46–52. doi: 10.1016/j.celrep.2019.11.116
- Martínez, M. A., Jordan-Paiz, A., Franco, S., and Nevot, M. (2016). Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness. *Trends Microbiol.* 24, 134–147. doi: 10.1016/j.tim.2015.11.002
- Martínez, M. A., Jordan-Paiz, A., Franco, S., and Nevot, M. (2019). Synonymous genome recoding: a tool to explore microbial biology and new therapeutic strategies. *Nucleic Acids Res.* 47, 10506–10519. doi: 10.1093/nar/gkz831
- Martus, G., Nevot, M., Andres, C., Clotet, B., and Martínez, M. A. (2013). Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. *Retrovirology* 10:78. doi: 10.1186/1742-4690-10-78
- Mauro, V. P., and Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends Mol. Med.* 20, 604–613. doi: 10.1016/j.molmed.2014.09.003
- Meagher, J. L., Takata, M., Gonçalves-Carneiro, D., Keane, S. C., Rebendene, A., Ong, H., et al. (2019). Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc. Natl. Acad. Sci. U S A.* 116, 24303–24309. doi: 10.1073/pnas.1913232116

- Mittal, P., Brindle, J., Stephen, J., Plotkin, J. B., and Kudla, G. (2018). Codon usage influences fitness through RNA toxicity. *Proc. Natl. Acad. Sci. U S A.* 115, 8639–8644. doi: 10.1073/pnas.1810022115
- Mochizuki, T., Ohara, R., and Roossinck, M. J. (2018). Large-Scale Synonymous Substitutions in Cucumber Mosaic Virus RNA 3 Facilitate Amino Acid Mutations in the Coat Protein. *J. Virol.* 92, 1007–1018e.
- Moratorio, G., Henningson, R., Barbezange, C., Carrau, L., Bordería, A. V., Blanc, H., et al. (2017). Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.* 2:17088.
- Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., et al. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.* 28, 723–726. doi: 10.1038/nbt.1636
- Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S., and 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. doi: 10.1128/jvi.00738-06
- Nevot, M., Jordan-Paiz, A., Martrus, G., Andrés, C., García-Cehic, D., Gregori, J., et al. (2018). HIV-1 Protease Evolvability Is Affected by Synonymous Nucleotide Recoding. *J. Virol.* 92, 777–718e.
- Ni, Y. Y., Zhao, Z., Opiessnig, T., Subramaniam, S., Zhou, L., Cao, D., et al. (2014). Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology* 450–451, 132–139. doi: 10.1016/j.virol.2013.12.009
- Nomaguchi, M., Miyake, A., Doi, N., Fujiwara, S., Miyazaki, Y., Tsunetsugu-Yokota, Y., et al. (2014). Natural Single-Nucleotide Polymorphisms in the 3' Region of the HIV-1 pol Gene Modulate Viral Replication Ability. *J. Virol.* 88, 4145–4160. doi: 10.1128/jvi.01859-13
- Nouën, C., Le, Brock, L. G., Luongo, C., McCarty, T., Yang, L., et al. (2014). Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc. Natl. Acad. Sci. U S A.* 111, 13169–13174. doi: 10.1073/pnas.1411290111
- Nougairde, A., de Fabritus, L., Aubry, F., Gould, E. A., Holmes, E. C., and de Lamballerie, X. (2013). Random Codon Re-encoding Induces Stable Reduction of Replicative Fitness of Chikungunya Virus in Primate and Mosquito Cells. *PLoS Pathog.* 9:e1003172. doi: 10.1371/journal.ppat.1003172
- Odon, V., Fros, J. J., Goonawardane, N., Dietrich, I., Ibrahim, A., Alshaikahmed, K., et al. (2019). The role of ZAP and OAS3/RNaseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Res.* 47, 8061–8083. doi: 10.1093/nar/gkz581
- Pagani, F., Raponi, M., and Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. U S A.* 102, 6368–6372. doi: 10.1073/pnas.0502288102
- Pandey, S., Kawai, T., and Akira, S. (2015). Microbial sensing by toll-like receptors and intracellular nucleic acid sensors. *Cold Spring Harb. Perspect. Biol.* 7:a016246. doi: 10.1101/cshperspect.a016246
- Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* 20, 237–243. doi: 10.1038/nsmb.2466
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124. doi: 10.1016/j.cell.2015.02.029
- Sáez-Llorens, X., Bandyopadhyay, A. S., Gast, C., Leon, T., De, DeAntonio, R., et al. (2021). Safety and immunogenicity of two novel type 2 oral poliovirus vaccine candidates compared with a monovalent type 2 oral poliovirus vaccine in children and infants: two clinical trials. *Lancet* 397, 27–38. doi: 10.1016/s0140-6736(20)32540-x
- Schwarz, D. A., Katayama, C. D., and Hedrick, S. M. (1998). Schlafen, a new family of growth regulatory genes that affect thymocyte development. *Immunity* 9, 657–668. doi: 10.1016/s1074-7613(00)80663-9
- Shabalina, S. A., Ogurtsov, A. Y., and Spiridonov, N. A. (2006). A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 34, 2428–2437. doi: 10.1093/nar/gkl287
- Shen, S. H., Stauff, C. B., Gorbatshevych, O., Song, Y., Ward, C. B., Yurovsky, A., et al. (2015). Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. *Proc. Natl. Acad. Sci. U S A.* 112, 4749–4754. doi: 10.1073/pnas.1502864112
- Shin, Y. C., Bischof, G. F., Lauer, W. A., and Desrosiers, R. C. (2015). Importance of codon usage for the temporal regulation of viral gene expression. *Proc. Natl. Acad. Sci. U S A.* 112, 14030–14035. doi: 10.1073/pnas.1515387112
- Simmonds, P., Xia, W., Baillie, J. K., and McKinnon, K. (2013). Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla - selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics* 14:610. doi: 10.1186/1471-2164-14-610
- Song, Y., Liu, Y., Ward, C. B., Mueller, S., Futcher, B., Skiena, S., et al. (2012). Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. *Proc. Natl. Acad. Sci. U S A.* 109, 14301–14307. doi: 10.1073/pnas.1211484109
- Sorensen, M. A., Kurland, C. G., and Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* 207, 365–377. doi: 10.1016/0022-2836(89)90260-x
- Stabell, A. C., Hawkins, J., Li, M., Gao, X., David, M., Press, W. H., et al. (2016). Non-human Primate Schlafen11 Inhibits Production of Both Host and Viral Proteins. *PLoS Pathog.* 12:e1006066. doi: 10.1371/journal.ppat.1006066
- Stergachis, A. B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., et al. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342, 1367–1372. doi: 10.1126/science.1243490
- Takata, M. A., Gonçalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., et al. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* 550, 124–127. doi: 10.1038/nature24039
- Takata, M. A., Soll, S. J., Emery, A., Blanco-Melo, D., Swanstrom, R., and Bieniasz, P. D. (2018). Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. *PLoS Pathog.* 14:e1006824. doi: 10.1371/journal.ppat.1006824
- Theys, K., Feder, A. F., Gelbart, M., Hartl, M., Stern, A., and Pennings, P. S. (2018). Within-patient mutation frequencies reveal fitness costs of CpG dinucleotides and drastic amino acid changes in HIV. *PLoS Genet.* 14:e1007420. doi: 10.1371/journal.pgen.1007420
- Tomezsko, P. J., Corbin, V. D. A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., et al. (2020). Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* 582, 438–442. doi: 10.1038/s41586-020-2253-5
- Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D., and Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife* 3:e04531.
- Vabret, N., Bailly-Bechet, M., Lepelley, A., Najburg, V., Schwartz, O., Verrier, B., et al. (2014). Large-Scale Nucleotide Optimization of Simian Immunodeficiency Virus Reduces Its Capacity To Stimulate Type I Interferon In Vitro. *J. Virol.* 88, 4161–4172. doi: 10.1128/jvi.03223-13
- Vabret, N., Bailly-Bechet, M., Najburg, V., Müller-Trutwin, M., Verrier, B., and Tangy, F. (2012). The biased nucleotide composition of HIV-1 triggers type I interferon response and correlates with subtype D increased pathogenicity. *PLoS One* 7:e33502. doi: 10.1371/journal.pone.0033502
- Vabret, N., Bhardwaj, N., and Greenbaum, B. D. (2017). Sequence-Specific Sensing of Nucleic Acids. *Trends Immunol.* 38, 53–65. doi: 10.1016/j.it.2016.10.006
- Van Damme, P., De Coster, I., Bandyopadhyay, A. S., Revets, H., Withanage, K., De Smedt, P., et al. (2019). The safety and immunogenicity of two novel live attenuated monovalent (serotype 2) oral poliovirus vaccines in healthy adults: a double-blind, single-centre phase 1 study. *Lancet* 394, 148–158. doi: 10.1016/s0140-6736(19)31279-6
- van der Kuyl, A. C., and Berkhout, B. (2012). The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* 9:92.
- Wang, B., Yang, C., Tekes, G., Mueller, S., Paul, A., Whelan, S. P. J., et al. (2015). Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. *MBio* 6:e00237.
- Warnecke, T., Weber, C. C., and Hurst, L. D. (2009). Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence. *Biochem. Soc. Trans.* 37, 756–761. doi: 10.1042/bst0370756
- Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Swanstrom, R., et al. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716. doi: 10.1038/nature08237
- Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., et al. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 6, 883–899. doi: 10.1371/journal.pbio.0060096

- Yang, C., Skiena, S., Futcher, B., Mueller, S., and Wimmer, E. (2013). Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraproductive live vaccine in mice. *Proc. Natl. Acad. Sci. U S A.* 110, 9481–9486. doi: 10.1073/pnas.1307473110
- Zheng, X., Wang, X., Tu, F., Wang, Q., Fan, Z., and Gao, G. (2017). TRIM25 Is Required for the Antiviral Activity of Zinc Finger Antiviral Protein. *J. Virol.* 91, 88–17e.
- Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C. H., Fu, J., et al. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. U S A.* 113, E6117–E6125.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Jordan-Paiz, Franco and Martínez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.