CrossMark

Attenuation of Viruses by Large-Scale Recoding of their Genomes: the Selection Is Always Biased

Nikolaus Osterrieder [1] · Dusan Kunec [1]

Abstract

Purpose of Review This review summarizes the current understanding of virus attenuation by large-scale recoding of viral genomes and discusses what would ultimately be necessary for construction of better and safer live modified virus vaccines.
Recent Findings It has been shown that codon and codon pair deoptimization are rapid and robust methods that can be used for the development of attenuated vaccine candidates. The viruses attenuated by large-scale recoding have the added value that they are extremely genetically stable. However, the exact mechanisms that lead to viral attenuation by recoding are yet to be determined.
Summary While the advantages of large-scale recoding (speed, simplicity, potency, and universal applicability) have been known for more than a decade, this approach has been only inadequately explored and the attention was focused on a limited number of RNA viruses. Attenuation of viruses by large-scale recoding should be explored to combat known and future viral threats.

Keywords Codon bias · Codon pair bias · CpG dinucleotides · Live modified virus vaccines · Synthetic attenuation virus engineering · Large-scale recoding

Introduction

Viruses account for the majority of newly emerging human pathogens. Over the past few years, many different viruses such as SARS coronavirus, MERS coronavirus, avian influenza viruses, hantaviruses, Zaire ebolavirus, or Zika virus have (re)emerged as human pathogens [1–7]. Vaccines are the most efficient and cost-effective tools to fight infectious diseases, particularly virus infections. Millions of people and domestic animals worldwide still suffer from many devastating infectious diseases for which no (efficient) vaccines exist. We lack a rapid, universal, and reliable strategy that could be used for attenuation of viruses and production of vaccines.

From the three basic types of viral vaccines, modified live virus vaccines are the most efficacious and preferred vaccines for healthy individuals, because they evoke broad, strong, and durable immune responses and generally outperform inactivated and subunit vaccines [8–10].

Traditionally, modified live virus vaccines have been prepared empirically by serial passage of virulent viruses in cell culture and/or laboratory animals [8, 9]. However, attenuation by this procedure is costly, time consuming, and highly unpredictable [9]. While serial passage results in accumulation of a large number of mutations, often only a handful of them contribute to attenuation [11, 12]. Consequently, some vaccines prepared by serial passage are prone to reversion to virulence [13–15]. This safety concern is the biggest limiting factor for use of such vaccines [8, 9, 11].

Recent advances in the de novo synthesis of DNA ushered in the era of synthetic biology and the nascent field of modified live virus vaccines that are prepared by large-scale recoding of pathogen genomes [16–19]. In contrast to viral attenuation by serial passage, the attenuating mutations are introduced into viral genomes deliberately, according to rationally designed recoding principles.

The overall concept of attenuation by large-scale recoding is simple and effective: viruses with recoded genomes

✉ Dusan Kunec
dusan.kunec@fu-berlin.de

[1] Institut für Virologie, Zentrum für Infektionsmedizin, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany

replicate efficiently in culture systems, which is favorable for viral vaccine production, but their replication capacity and virulence in vivo is severely reduced or absent [17, 20]. The reduction of the reproductive fitness enables the host to gain the upper hand in controlling virus replication by innate and adaptive immune responses.

Typically, the goal of the recoding is to change dinucleotide, codon, or codon pair composition of the recoded viral genomes, because it was shown that all three types of (interrelated) modifications could lead to replication-competent, but severely attenuated viruses. Importantly, while recoding introduces hundreds of point mutations into viral genomes, the amino acid composition of the encoded proteins remains preserved. Consequently, the recoded viruses are antigenically identical with their pathogenic parents. The antigenic identity and replicative potential enable attenuated viruses to induce immune responses that are similar to those of virulent strains. Recoded viruses represent very promising vaccine candidates, because it might be possible to achieve the desired level of attenuation by adjusting the level of recoding [17]. In addition, viruses attenuated by large-scale recoding are extremely genetically stable, which is explained by the sheer number of introduced mutations [16, 17, 20–22].

Virus Attenuation by Codon Deoptimization

Amino acids, except for methionine and tryptophan, can be encoded by two or more synonymous codons, but these are used at unequal frequencies, a phenomenon known as *codon bias*. Synonymous codons also differ in translational accuracy [23], propensity to mutate to non-synonymous and non-conservative codons [24, 25], abundance of tRNA that decode them [26], and capacity to allow non-standard (wobble) base pairing between the third base of the codon and the first base of the anticodon [27].

Codon choice affects translation efficiency [28], protein folding [29], and mRNA stability [30], but the significance of codon bias despite of decades of investigation remains unclear. The prevailing hypothesis predicts that frequently used codons are translated more rapidly than rare codons, because frequent codons are often decoded by abundant tRNAs [26, 31, 32]. Consequently, utilization of rare codons reduces translation rates and protein yield because these are decoded slowly by rare tRNAs [30, 33]. Yet, only little direct in vivo evidence supports this hypothesis [34, 35]. In addition, it was shown that codon-optimized genes are often not translated as efficiently as expected [33].

The first attenuated virus that was prepared by large-scale recoding was a poliovirus (Enterovirus C) and the recoding modified codon usage of the capsid coding region [16, 18]. The rationale for recoding was the opposite of the codon optimization strategies. The goal of the recoding was to modify viral genomes to contain more codons that are infrequently used by the virus [18] or the virus host [16], because it was assumed that these might reduce speed of translation elongation and thus also protein yield. The codon deoptimization resulted in severely attenuated viruses in vitro [16, 18] and in vivo [16]. As expected, maximization of codons that are underrepresented in the virus host decreased translation capacity and protein yields of the recoded viruses. Surprisingly, viruses that contained the increased number of its own infrequent codons showed unaltered protein production, but diminished viral RNA yields and specific infectivity of purified virions [18].

Since then, others followed suit and many other viruses including rabies virus [36], influenza A virus, human respiratory syncytial virus [22, 37, 38], lymphocytic choriomeningitis virus [39, 40], and foot-and-mouth disease virus [41] were recoded using the codon deoptimization principles. In most cases, codon deoptimization resulted in production of highly attenuated viruses in vitro and in vivo [39–42]. However, some codon-deoptimized viruses remained pathogenic [36], or became only moderately attenuated [22].

Interestingly, experiments with human respiratory syncytial virus showed that viruses that were codon deoptimized according to the viral host codon usage had decreased protein production and were attenuated, whereas viruses that were deoptimized according to the virus codon usage were not [22].

Virus Attenuation by Codon Pair Deoptimization

The fact that codon usage alone could not explain observed differences in protein production implicated that other sequence features, such as neighboring nucleotides, or codons (codon context) must influence translation elongation. Recent studies accumulated compelling evidence that different mRNA context cues modulate eukaryotic translation (reviewed in [43]).

Similar to codon bias, *codon pair bias*, that is *codon pair combinations*, in protein coding genes is not random either [44, 45]. Some codon pairs are found in open reading frames (ORFs) significantly more or less frequently than would be expected based on the overall frequencies of two codons that form a particular codon pair [16, 21, 44]. The level of under- and overrepresentation of each codon pair can be measured with the codon pair score (CPS) statistic [21].

Codon pair bias was found in every species studied [46] and can be radically dissimilar between different species [20], but closely related species have essentially the same codon pair bias [45, 47]. Its existence has been known for many years, but it was on the periphery of scientific inquiry, and thus its biological significance and the forces shaping it are only poorly understood [46].

The attenuation by *codon pair (bias) deoptimization*, also known as "synthetic attenuated virus engineering" (SAVE), was pioneered in 2008 by the group of Eckard Wimmer at Stony Brook University when the effects resulting from the alternation of the codon pair bias was explored by recoding of poliovirus [21]. The attenuation of viruses by codon pair deoptimization involves reshuffling of existing codons in a protein coding sequence without changing the codon bias or amino acid composition of the encoded protein [20, 21]. The goal of reshuffling is to maximize the number of codon pairs that are underrepresented in the protein coding sequences of the virus host.

In the seminal study, the recoding by codon pair deoptimization involved the P1 region of the virus, which encodes the viral capsid [21]. Remarkably, a poliovirus with a fully codon pair-deoptimized P1 region, "PV-Min," could not be rescued in cell culture, despite the fact that no new rare codons were introduced into recoded viral segment. On the other hand, "PV-Max" virus with codon pair-optimized P1 segment had biological properties of the wild-type parent.

Since its initial description, codon pair deoptimization enabled rapid and highly efficient attenuation of a wide variety of viruses, including influenza A virus [20, 48, 49•], human immunodeficiency virus [50], human respiratory syncytial virus [51], vesicular stomatitis Indiana virus [52], and dengue virus [53•]. Some of the recoded viruses have shown 100,000-fold attenuation in comparison to pathogenic parents and have been successfully used as highly protective experimental vaccines with a wide margin of safety [48, 49•].

There are two main competing hypotheses that propose different molecular mechanisms that lead to attenuation of viruses by codon pair deoptimization. One hypothesis suggests that the increased numbers of underrepresented, or "non-preferred" codon pairs in recoded sequences, are themselves the reason for attenuation, because they create conditions that are not conducive for efficient protein production or processing [21, 53•]. It is speculated that physical properties of some tRNA molecules hamper their efficient interaction at adjacent A-site and P-site in the translating ribosome. As a consequence, codon pair-deoptimized sequences do not support efficient protein translation, and these are prone to increased mistranslation, stalled translation, or premature termination [49•]. The alternate hypothesis suggests that not the codon pairs themselves, but the increased number of CpG (and TpA) dinucleotides that are present in codon pair-deoptimized sequences (see below for explanation) is responsible for decrease of mRNA levels and thus also protein yields and virus attenuation [54–56].

While recoding by codon pair deoptimization has always led to decrease of protein production, it is unknown whether this decrease is caused primarily by suboptimal protein translation, or could be also caused by the reduced mRNA levels, because it was shown that codon pair deoptimization can be responsible for extensive reduction of mRNA levels [48]. However, the reduction of mRNA levels does not occur universally, and often the reduction of RNA levels is disproportional to the magnitude of reduction of protein levels [48, 49].

Typically, codon pair deoptimization introduces several hundred nucleotide changes into recoded genes. It is not known which genetic changes that arise through codon pair deoptimization are responsible for reduced protein production and ensuing virus attenuation. There are three possible options: (1) reduction of protein production is caused by a large number of underrepresented codon pairs that exert small negative effects on protein production, (2) reduction of protein production is caused primarily by a small number of codon pairs that exert strong negative effects on protein production, and (3) other yet unknown sequence features are responsible for decrease of mRNA stability, or faster turnover of mRNA transcript.

In 2016, in an elegant study, Gamble et al. provided compelling evidence—through experimentation with 35,000 GFP variants in yeast species *Saccharomyces cerevisiae*—that codon pairs rather than individual codons can exert a potent effect on translation elongation [57••]. The study identified 17 inhibitory codon pairs that were implicated in low protein production of the superfolding GFP. The inhibitory effect could not be assigned to individual codons, or six-base sequence, or encoded dipeptide, since reduced protein production was observed only when both codons of the inhibitory pair were present, in-frame, and adjacent in a proper order. The correct ordering suggested that tRNA interactions with mRNA on the ribosome mediated the inhibitory effect.

Codon and Codon Pair Deoptimization Increases the Number of CpG Dinucleotides in Recoded Genes

It was discovered that codon and codon pair deoptimization of vertebrate viruses not only increases the number of codons, or codon pairs that are underrepresented in coding sequences of the host, but also increases the frequency of CpG and, to lesser degree, TpA (UpA) dinucleotides in recoded sequences [45, 54, 55, 58]. The increase of CpG and TpA dinucleotides by codon pair deoptimization is inadvertent, as codon pairs that contain CpG and TpA dinucleotides at the codon pair boundary (NNC-p-GNN) are among the most underrepresented codon pairs in vertebrates [45, 55]. For example, 97 of the 100 most underrepresented codon pairs contain CpG at the codon pair boundary [45]. As a result, recoding by codon pair deoptimization does not increase the number of CpG dinucleotides that are already present in the shuffled codons, because recoding preserves codon bias, but creates new CpG and TpA dinucleotides at the boundary between the new codon pairs.

Similarly, because CpG and TpA dinucleotides are significantly suppressed in the genomes of higher eukaryotes [59], synonymous codons that contain CpG and TpA dinucleotides, for example, alanine's GCG, or leucine's CTA and TTA codons, are also infrequently used in protein coding sequences. Thus, codon deoptimization of vertebrate viruses also results in elevated number of CpG and TpA dinucleotides in recoded sequences. It is, therefore, unclear whether the increase of underrepresented codons, codon pairs, or less-favored dinucleotides in recoded sequences is primarily responsible for virus attenuation.

Consequently, an alternative hypothesis suggests that the cause of attenuation is to be found in the increased number of CpG (and TpA) dinucleotides, which are recognized by an as yet uncharacterized self/non-self-recognition system that stimulates enhanced innate immune responses to such recoded viruses [55, 56]. Since codon pair preferences and dinucleotide frequencies are intimately related (the most underrepresented codon pairs contain CpG and TpA dinucleotides at the codon pair boundary), dissecting the effects of the two phenomena is exquisitely difficult [56, 60].

Virus Attenuation by Increase of CpG/TpA Dinucleotides Frequencies

Vertebrate genomes have low CpG levels and the CpG suppression can be plausibly explained by the methylation-deamination hypothesis [61]. This hypothesis suggests that abundant methylation of cytosine in CpG dinucleotides is responsible for CpG suppression, because methylated cytosine often mutates to thymine by spontaneous deamination [62]. As a result, methylated CpG dinucleotides decay into TpG (and CpA) dinucleotides over time.

It remains enigmatic why CpG and TpA (UpA) dinucleotides occur at lower frequency also in the genomes of most RNA and small DNA viruses that infect vertebrates [45, 63]. For example, human papillomaviruses exhibit a frequency of CpG dinucleotides in their genomes that is only ∼ 50% of the expected number. Even more striking, human immunodeficiency viruses contain reduced CpG to only ∼ 25% and human polyomaviruses to less than 10% of the expected numbers [45]. Because CpG methylation does not occur on RNA, the methylation-deamination hypothesis, nor viral sequence constraints, can explain underrepresentation of CpG dinucleotides in genomes of vertebrate RNA viruses [64].

An alternative explanation for suppression of CpG dinucleotides in the genomes of small viruses suggests that CpG dinucleotides act as immunostimulatory motifs that trigger antiviral immune responses [45, 54, 64]. However, the identity of the hypothetical receptors recognizing CpG-rich RNA molecules remains elusive.

A recent study by Takata et al. showed that the host zinc-finger antiviral protein (ZAP) is such a long-suspected ssRNA CpG receptor, which specifically binds to CpG-rich RNA and targets them for degradation by the RNA exosome [65••]. These results suggest that the selective pressure mounted by ZAP drives vertebrate RNA viruses to reduce the levels of CpG dinucleotides in their genomes. Thus, increasing the number of CpG dinucleotides in viral genomes could be responsible for viral attenuation, because viral RNA with high CpG content is better recognized and then removed from the cytoplasm. Since ZAP expression is induced by interferon, viruses that can block interferon responses or counter the action of ZAP should be resistant to selection pressure exerted by ZAP. It remains to be determined whether ZAP is the only host factor that can recognize CpG-rich sequences, and what immune evasion strategies viruses employ to avoid the action of ZAP.

The early experiments with recoded poliovirus that had artificially elevated CpG and UpA dinucleotides in its capsid coding region showed that recoding only minimally affected protein production, protein processing, or the overall production of viral particles, but had significant negative effect on virus fitness, especially on specific infectivity of viral particles [66]. The fitness of the recoded virus was reduced to the threshold of viability when CpG and UpA dinucleotides were maximized within the recoded genome segment [66].

Since both codon pair bias and CpG dinucleotides appear to influence virus replication, a study with echovirus 7 attempted to separate the effect of two phenomena by creating mutants in which the two parameters were independently varied. The authors increased either the frequency of CpG and UpA dinucleotides in the viral genome and left codon pair bias constant, or vice versa [54]. Phenotypic characterization of the resulting mutants showed that only alternation of the CpG and UpA frequencies, but not codon pair bias, had a negative effect on viral fitness [54]. Interestingly, a complementary study showed that virus mutants that lacked CpG an UpA dinucleotides in their genomes had enhanced replication, produced larger plaques, and readily outcompeted wild-type parents in competition assays [55].

A subsequent study from the same group demonstrated that elevation of CpG frequencies in influenza A virus can also result in moderate attenuation of the virus in vitro and in vivo [58]. However, since recoded viruses with increased CpG or UpA frequencies did not have the same codon bias as the control—unmodified and permuted—viruses, changes in the codon bias could confound observed viral properties. In addition, because influenza A virus antagonizes ZAP activity [67, 68], it remains to be determined if ZAP alone can inhibit replication of recoded influenza A viruses with increased CpG levels.

Conclusion

Although the capacity to attenuate viruses by the three alternative attenuation methods is yet to be directly compared, viruses that are designed by codon pair deoptimization show consistently high levels of attenuation, to the extent that some codon pair-deoptimized viruses are nonviable in permissive cells [21, 52].

The major drawback of the attenuation by alternation of codon, codon pair, or CpG dinucleotide frequencies is that the molecular mechanisms responsible for attenuation remain largely unknown. Until this problem is solved, it will not be possible to improve this attenuation method further and design rationally better and safer vaccines. In addition, it will not be possible to assess reversion of attenuation based on observed genetic changes in attenuated viruses. Also, it is yet to be determined if DNA viruses can be attenuated by the same attenuating principles as small RNA viruses.

The continuously decreasing cost of synthetic DNA might soon allow us to characterize the phenotype of thousands of differently recoded viruses. Once phenotype is connected with genotype, unbiased and agnostic approaches might be able to precisely identify sequence features that are essential and sufficient for development of highly effective and safer viral vaccines.

In contrast to existent attenuation methods, recoded vaccine candidates can be designed within minutes and produced synthetically within days. The potential applications that might originate from these approaches are immense and could be universally applicable for attenuation of many known viruses and bacteria, but also to yet unknown viral threats as they emerge [17, 20].

Compliance With Ethical Standards

Conflict of Interest   The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent   This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A. 2005;102(39):14040–5. https://doi.org/10.1073/pnas.0506735102.

2. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676–9. https://doi.org/10.1126/science.1118391.

3. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814–20. https://doi.org/10.1056/NEJMoa1211721.

4. Li KS, Guan Y, Wang J, Smith GJ, Xu KM, Duan L, et al. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature. 2004;430(6996):209–13. https://doi.org/10.1038/nature02746.

5. Kruger DH, Tkachenko EA, Morozov VG, Yunicheva YV, Pilikova OM, Malkin G, et al. Life-threatening Sochi virus infections, Russia. Emerg Infect Dis. 2015;21(12):2204–8. https://doi.org/10.3201/eid2112.150891.

6. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, et al. Emergence of Zaire Ebola virus disease in Guinea. N Engl J Med. 2014;371(15):1418–25. https://doi.org/10.1056/NEJMoa1404505.

7. Zanluca C, Melo VC, Mosimann AL, Santos GI, Santos CN, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Mem Inst Oswaldo Cruz. 2015;110(4):569–72. https://doi.org/10.1590/0074-02760150192.

8. Kusters I, Almond JW. Vaccine Strategies. In: BWJM, Regenmortel MHVv, editors. Encyclopedia of virology (Third Edition). Oxford: Academic Press; 2008. p. 235–243.https://doi.org/10.1016/B978-012374410-4.00597-5.

9. Lauring AS, Jones JO, Andino R. Rationalizing the development of live attenuated virus vaccines. Nat Biotechnol. 2010;28(6):573–9. https://doi.org/10.1038/nbt.1635.

10. Bazin H. A brief history of the prevention of infectious diseases by immunisations. Comp Immunol Microbiol Infect Dis. 2003;26(5–6):293–308. https://doi.org/10.1016/S0147-9571(03)00016-X.

11. Spatz SJ, Volkening JD, Gimeno IM, Heidari M, Witter RL. Dynamic equilibrium of Marek's disease genomes during in vitro serial passage. Virus Genes. 2012. https://doi.org/10.1007/s11262-012-0792-z.

12. Spatz SJ. Accumulation of attenuating mutations in varying proportions within a high passage very virulent plus strain of Gallid herpesvirus type 2. Virus Res. 2010;149(2):135–42. https://doi.org/10.1016/j.virusres.2010.01.007.

13. Witter RL. Attenuated revertant serotype 1 Marek's disease viruses: safety and protective efficacy. Avian Dis. 1991;35(4):877–91. https://doi.org/10.2307/1591624.

14. Abraham R, Minor P, Dunn G, Modlin JF, Ogra PL. Shedding of virulent poliovirus revertants during immunization with oral poliovirus vaccine after prior immunization with inactivated polio vaccine. J Infect Dis. 1993;168(5):1105–9. https://doi.org/10.1093/infdis/168.5.1105.

15. Kew O, Morris-Glasgow V, Landaverde M, Burns C, Shaw J, Garib Z, André J, Blackman E, Freeman CJ, Jorba J, Sutter R, Tambini G, Venczel L, Pedreira C, Laender F, Shimizu H, Yoneyama T, Miyamura T, van der Avoort H, Oberste MS, Kilpatrick D, Cochi S, Pallansch M, de Quadros C Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. Science 2002;296(5566):356–359.https://doi.org/10.1126/science.1068284.

16. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol. 2006;80(19):9687–96. https://doi.org/10.1128/JVI.00738-06.

17. Wimmer E, Mueller S, Tumpey TM, Taubenberger JK. Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat Biotechnol. 2009;27(12):1163–72. https://doi.org/10.1038/nbt.1593.

18. Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A, Quay J, et al. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol. 2006;80(7):3259–72. https://doi.org/10.1128/JVI.80.7.3259-3272.2006.

19. Coleman JR, Papamichail D, Yano M, Garcia-Suarez Mdel M, Pirofski LA. Designed reduction of Streptococcus pneumoniae pathogenicity via synthetic changes in virulence factor codon-pair bias. J Infect Dis. 2011;203(9):1264–73. https://doi.org/10.1093/infdis/jir010.

20. Mueller S, Coleman JR, Papamichail D, Ward CB, Nimnual A, Futcher B, et al. Live attenuated influenza virus vaccines by computer-aided rational design. Nat Biotechnol. 2010;28(7):723–6. https://doi.org/10.1038/nbt.1636.

21. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. Virus attenuation by genome-scale changes in codon pair bias. Science. 2008;320(5884):1784–7. https://doi.org/10.1126/science.1155761.

22. Meng J, Lee S, Hotard AL, Moore ML. Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. MBio. 2014;5(5):e01704–14. https://doi.org/10.1128/mBio.01704-14.

23. Dix DB, Thompson RC. Codon choice and gene expression: synonymous codons differ in translational accuracy. Proc Natl Acad Sci U S A. 1989;86(18):6888–92. https://doi.org/10.1073/pnas.86.18.6888.

24. Lauring AS, Acevedo A, Cooper SB, Andino R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe. 2012;12(5):623–32. https://doi.org/10.1016/j.chom.2012.10.008.

25. Archetti M. Genetic robustness and selection at the protein level for synonymous codons. J Evol Biol. 2006;19(2):353–65. https://doi.org/10.1111/j.1420-9101.2005.01029.x.

26. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125–43. https://doi.org/10.1093/nar/14.13.5125.

27. Crick FH. Codon–anticodon pairing: the wobble hypothesis. J Mol Biol. 1966;19(2):548–55. https://doi.org/10.1016/S0022-2836(66)80022-0.

28. Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. Mol Syst Biol. 2011;7:481.

29. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol. 2009;16(3):274–80. https://doi.org/10.1038/nsmb.1554.

30. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015;160(6):1111–24. https://doi.org/10.1016/j.cell.2015.02.029.

31. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 1985;2(1):13–34. https://doi.org/10.1093/oxfordjournals.molbev.a040335.

32. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011;12(1):32–42. https://doi.org/10.1038/nrg2899.

33. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004;22(7):346–53. https://doi.org/10.1016/j.tibtech.2004.04.006.

34. Gardin J, Yeasmin R, Yurovsky A, Cai Y, Skiena S, Futcher B. Measurement of average decoding rates of the 61 sense codons in vivo. eLife. 2014;3 https://doi.org/10.7554/eLife.03735.

35. Lareau LF, Hite DH, Hogan GJ, Brown PO. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. eLife. 2014;3:e01257. https://doi.org/10.7554/eLife.01257.

36. Wirblich C, Schnell MJ. Rabies virus (RV) glycoprotein expression levels are not critical for pathogenicity of RV. J Virol. 2011;85(2):697–704. https://doi.org/10.1128/JVI.01309-10.

37. Rostad CA, Stobart CC, Gilbert BE, Pickles RJ, Hotard AL, Meng J, et al. A recombinant respiratory syncytial virus vaccine candidate attenuated by a low-fusion F protein is immunogenic and protective against challenge in cotton rats. J Virol. 2016;90(16):7508–18. https://doi.org/10.1128/JVI.00012-16.

38. Stobart CC, Rostad CA, Ke Z, Dillard RS, Hampton CM, Strauss JD, et al. A live RSV vaccine with engineered thermostability is immunogenic in cotton rats despite high attenuation. Nat Commun. 2016;7:13916. https://doi.org/10.1038/ncomms13916.

39. Cheng BY, Nogales A, de la Torre JC, Martinez-Sobrido L. Development of live-attenuated arenavirus vaccines based on codon deoptimization of the viral glycoprotein. Virology. 2017;501:35–46. https://doi.org/10.1016/j.virol.2016.11.001.

40. Cheng BY, Ortiz-Riano E, Nogales A, de la Torre JC, Martinez-Sobrido L. Development of live-attenuated arenavirus vaccines based on codon deoptimization. J Virol. 2015;89(7):3523–33. https://doi.org/10.1128/JVI.03401-14.

41. Diaz-San Segundo F, Medina GN, Ramirez-Medina E, Velazquez-Salinas L, Koster M, Grubman MJ, et al. Synonymous deoptimization of foot-and-mouth disease virus causes attenuation in vivo while inducing a strong neutralizing antibody response. J Virol. 2015;90(3):1298–310. https://doi.org/10.1128/JVI.02167-15.

42. Nogales A, Baker SF, Ortiz-Riano E, Dewhurst S, Topham DJ, Martinez-Sobrido L. Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. J Virol. 2014;88(18):10525–40. https://doi.org/10.1128/JVI.01565-14.

43. Brar GA. Beyond the triplet code: context cues transform translation. Cell. 2016;167(7):1681–92. https://doi.org/10.1016/j.cell.2016.09.022.

44. Gutman GA, Hatfield GW. Nonrandom utilization of codon pairs in Escherichia coli. Proc Natl Acad Sci U S A. 1989;86(10):3699–703. https://doi.org/10.1073/pnas.86.10.3699.

45. Kunec D, Osterrieder N. Codon pair bias is a direct consequence of dinucleotide bias. Cell Rep. 2016;14(1):55–67. https://doi.org/10.1016/j.celrep.2015.12.011.

46. Moura G, Pinheiro M, Silva R, Miranda I, Afreixo V, Dias G, et al. Comparative context analysis of codon pairs on an ORFeome scale. Genome Biol. 2005;6(3):R28. https://doi.org/10.1186/gb-2005-6-3-r28.

47. Moura GR, Pinheiro M, Freitas A, Oliveira JL, Frommlet JC, Carreto L, et al. Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. PLoS One. 2011;6(10):e26817. https://doi.org/10.1371/journal.pone.0026817.

48. Yang C, Skiena S, Futcher B, Mueller S, Wimmer E. Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. Proc Natl Acad Sci U S A. 2013;110(23):9481–6. https://doi.org/10.1073/pnas.1307473110.

49.• Broadbent AJ, Santos CP, Anafu A, Wimmer E, Mueller S, Subbarao K. Evaluation of the attenuation, immunogenicity, and efficacy of a live virus vaccine generated by codon-pair bias deoptimization of the 2009 pandemic H1N1 influenza virus, in ferrets. Vaccine. 2016;34(4):563–570. **This study showed that codon pair deoptimization of a pandemic H1N1 influenza A virus (2009 pH1N1) rendered the virus attenuated for outbred ferrets. It also showed that attenuated virus unexpectedly replicated to high titers in upper respiratory tract of infected animals.** https://doi.org/10.1016/j.vaccine.2015.11.054.

50. Martrus G, Nevot M, Andres C, Clotet B, Martinez MA. Changes in codon-pair bias of human immunodeficiency virus type 1 have

profound effects on virus replication in cell culture. Retrovirology. 2013;10(1):78. https://doi.org/10.1186/1742-4690-10-78.

51. Le Nouen C, Brock LG, Luongo C, McCarty T, Yang L, Mehedi M, et al. Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. Proc Natl Acad Sci U S A. 2014;111(36):13169–74. https://doi.org/10.1073/pnas.1411290111.

52. Wang B, Yang C, Tekes G, Mueller S, Paul A, Whelan SP, et al. Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. mBio. 2015;6(2). https://doi.org/10.1128/mBio.00237-15.

53.• Shen SH, Stauft CB, Gorbatsevych O, Song Y, Ward CB, Yurovsky A, et al. Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc Natl Acad Sci U S A. 2015. **Arboviruses need to replicate in two completely different environments (primate and insect cells). This study showed that recoding of virus genome can diminish replication potential of the virus in one environment (human), while preserving replication fitness in the other (insect)**. https://doi.org/10.1073/pnas.1502864112.

54. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. eLife. 2014;4. https://doi.org/10.7554/eLife.04531.

55. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. 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. 2014;42(7):4527–45. https://doi.org/10.1093/nar/gku075.

56. Simmonds P, Tulloch F, Evans DJ, Ryan MD. Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. Proc Natl Acad Sci U S A. 2015;112(28):E3633–4. https://doi.org/10.1073/pnas.1507339112.

57.•• Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent codons act in concert to modulate translation efficiency in yeast. Cell. 2016;166(3):679–690. **This study showed that codon context can dramatically affect protein synthesis. The study identified 17 codon pairs that had negative effect on GFP protein synthesis in *S. cerevisiae*.** https://doi.org/10.1016/j.cell.2016.05.070.

58. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, Nicol MQ, et al. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. eLife. 2016;5:e12735. https://doi.org/10.7554/eLife.12735.

59. Karlin S, Burge C. Dinucleotide relative abundance extremes: a genomic signature. Trends in genetics : TIG. 1995;11(7):283–90. https://doi.org/10.1016/S0168-9525(00)89076-9.

60. Futcher B, Gorbatsevych O, Shen SH, Stauft CB, Song Y, Wang B, et al. Reply to Simmonds et al.: Codon pair and dinucleotide bias have not been functionally distinguished. Proc Natl Acad Sci U S A. 2015;112(28):E3635–6. https://doi.org/10.1073/pnas.1507710112.

61. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. Molecular basis of base substitution hotspots in Escherichia coli. Nature. 1978;274(5673):775–80. https://doi.org/10.1038/274775a0.

62. Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 1980;8(7):1499–504. https://doi.org/10.1093/nar/8.7.1499.

63. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? J Virol. 1994;68(5):2889–97.

64. Greenbaum BD, Rabadan R, Levine AJ. Patterns of oligonucleotide sequences in viral and host cell RNA identify mediators of the host innate immune system. PLoS One. 2009;4(6):e5969. https://doi.org/10.1371/journal.pone.0005969.

65.•• Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, et al. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature. 2017;550(7674):124–127. **This study showed that CpG suppression is essential for replication of HIV-1 and that the host zinc-finger antiviral protein (ZAP) is a long-suspected ssRNA CpG receptor, which binds to viral CpG-rich RNA and targets them for degradation by the RNA exosome.** https://doi.org/10.1038/nature24039.

66. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J Virol. 2009;83(19):9957–69. https://doi.org/10.1128/JVI.00508-09.

67. Liu CH, Zhou L, Chen G, Krug RM. Battle between influenza A virus and a newly identified antiviral activity of the PARP-containing ZAPL protein. Proc Natl Acad Sci U S A. 2015;112(45):14048–53. https://doi.org/10.1073/pnas.1509745112.

68. Tang Q, Wang X, Gao G. The short form of the zinc finger antiviral protein inhibits influenza A virus protein expression and is antagonized by the virus-encoded NS1. J Virol. 2017;91(2). https://doi.org/10.1128/JVI.01909-16.