

CODON PAIR BIAS AND VIRAL VACCINE DESIGN

Garry A. Luke*, Uday S. Pathania, Fiona Tulloch, Martin D. Ryan

Biomedical Sciences Research Complex, Biomolecular Sciences Building, University of St Andrews,
North Haugh, St Andrews, Fife KY16 9ST, Scotland, U.K.

*Corresponding author gal@st-andrews.ac.uk

Live attenuated vaccines (LAVs) have prevented morbidity and mortality against a number of important viral diseases (such as smallpox and polio) *via* long-lived humoral and cell-mediated immunity. Unlike inactivated, subunit or recombinant protein vaccines which require multiple inoculations this cost-effective approach requires only one or two doses to generate a robust immune response (Graham and Crowe, 2007). Furthermore, live vaccines are able to elicit both mucosal and systemic protective responses. Unfortunately, conventional LAVs have two major drawbacks. Firstly, attenuation of the pathogenic phenotype by either random gene mutation or by passage in unnatural conditions depends on chance and cannot be universally applied to a variety of virus types. Secondly, since viruses are generally attenuated on the basis of only a few mutations the risk of reversion to virulence remains a key aspect of developing attenuated virus vaccines.

Novel methods to rationally design vaccines include control by microRNAs (miRNAs) or zinc finger nucleases (ZFNs). While each of these may be useful against persistent or latent viruses, miRNAs and ZFNs are limited to RNA and non-integrating DNA viruses respectively (Lauring et al., 2010). Two other approaches have been recently proposed. In one approach, increased replication fidelity of error-prone RNA polymerases through mutations, reduces the likelihood of reversion from attenuation to wild type (Vignuzzi et al., 2008). A second approach, dubbed synthetic attenuated virus engineering (SAVE), can be used to create non-virulent strains of viruses, which could provide a platform technology for vaccine development against a wide variety of viral diseases (Coleman et al., 2008; Wimmer et al., 2009; Mueller et al., 2010). The new vaccine strategy exploits the fact that the genetic code is degenerate in that most amino acids are encoded by more than one three-nucleotide codon (e.g. GCA, GCC, GCG, and GCT all code for the amino acid alanine). Many organisms, including viruses, tend to have biases toward certain synonymous codons (the “codon bias”) and codon pairs (the “codon pair bias”) in their genes (Gutman and Hatfield, 1989; Moura et al., 2007; Tats et al., 2008). Codon pair bias refers to the preferential pairing of certain codons over others (e.g. in human genes the codon pair GCC-GAA encodes the adjacent amino acids alanine-glutamic acid less often than GCA-GAG, even though GCC and GAA are the most used codons, Coleman et al., 2008). Unfavourable host codon pairs introduced into a viral genome may adversely affect protein translation and thereby provide a method for producing attenuated viruses.

As a proof of concept for SAVE, a team led by Eckard Wimmer at Stony Brook University used a computer-guided approach to design weakened poliovirus strains (Coleman et al., 2008). The research is an outgrowth of a technique they pioneered to assemble the virus from scratch (Cello et al., 2002). The single stranded positive-sense RNA genome of poliovirus contains a long open reading frame encoding a polyprotein that is cleaved into P1 (capsid precursor), P2 and P3 (precursors to non-structural proteins) flanked by 5' and 3' nontranslated regions. Briefly, they engineered the P1 coding region (encoding the viral capsid, 2643 nucleotides) with under- or over-represented codon pairs: the attenuated strains caused decreased rates of protein translation and protected mice challenged with a lethal dose of wild-type poliovirus (Figures 1 & 2, *vide infra*). The promising results of this work suggested that SAVE could play a role in creating new vaccines for influenza virus. The “flu” virus

genome is composed of eight segments of negative-sense RNA that encode up to 12 proteins. Segments targeted for attenuation encode proteins that play important roles in replication and assembly of influenza virus - *nucleoprotein (NP)*, *polymerase subunit B1(PB1)*, and *hemagglutinin (HA)*. Synthetic strains of influenza virus PR8 containing one (NP, HA, or PB1), two (NP/HA or HA/PB1) or three (PB1/NP/HA, short PR8^{3F}) codon pair deoptimized gene segments formed plaques and grew *in vitro*. The PR8^{3F} strain was highly attenuated *in vivo* and could protect mice challenged with wild-type influenza virus (Mueller et al., 2010).

As a general strategy for vaccine development, SAVE offers several advantages. First, because this approach involves the introduction of hundreds of “silent” mutations, unlike the handful of mutations found in current vaccine strains, the probability of reversion of codon-pair deoptimized vaccines to virulence is essentially zero. Second, given the changes do not alter the amino acid sequences, the proteins expressed are identical to wild type virus and possess identical immunogenic properties. Finally, as attenuation is systematic rather than empiric, it can be applied speedily to a wide range of viruses in various species.

References

Cello J, Paul AV, Wimmer E. (2002). Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template. *Science* 297, 1016-1018.

Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science* 320(5884): 1784-1787.

Graham BS, Crowe JE. (2007) Immunization against viral diseases. In: “Fields Virology”, Fifth Edition, David M. Knipe and Peter M. Howley, editors. Lippincott Williams & Wilkins. Philadelphia, PA, Chapter 15: 487-538.

Gutman GA, Hatfield GW. (1989) Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc Natl Acad Sci USA* 86, 3699-3703.

Lauring AS, Jones JO, Andino R. (2010) Rationalizing the development of live attenuated virus vaccines. *Nature biotechnology* 28(6): 573-579.

Moura G, Pinheiro M, Arrais J, Gomes AC, Carreto L, Freitas A, Oliveira JL, Santos MA. (2007) Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PLOS ONE* 2, e847.

Mueller S, Coleman JR, Papamichail D, Ward CB, Nimnual A, Futcher B, Skiena S, Wimmer E. (2010) Live attenuated influenza virus vaccines by computer-aided rational design. *Nature biotechnology* 28(7): 723-726.

Tats A, Tenson T, Remm M. (2008). Preferred and avoided codon pairs in three domains of life. *BMC Genomics* 9:463 doi:10.1186/1471-2164-9-463.

Wimmer E, Mueller S, Tumpey TM, and Taubenberger JK. (2009) Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nature biotechnology* 27(12): 1163-1172.

Vignuzzi M, Wendt E, Andino R. (2008) Engineering attenuated virus vaccines by controlling replication fidelity. *Nature medicine* 14(2): 154-161.

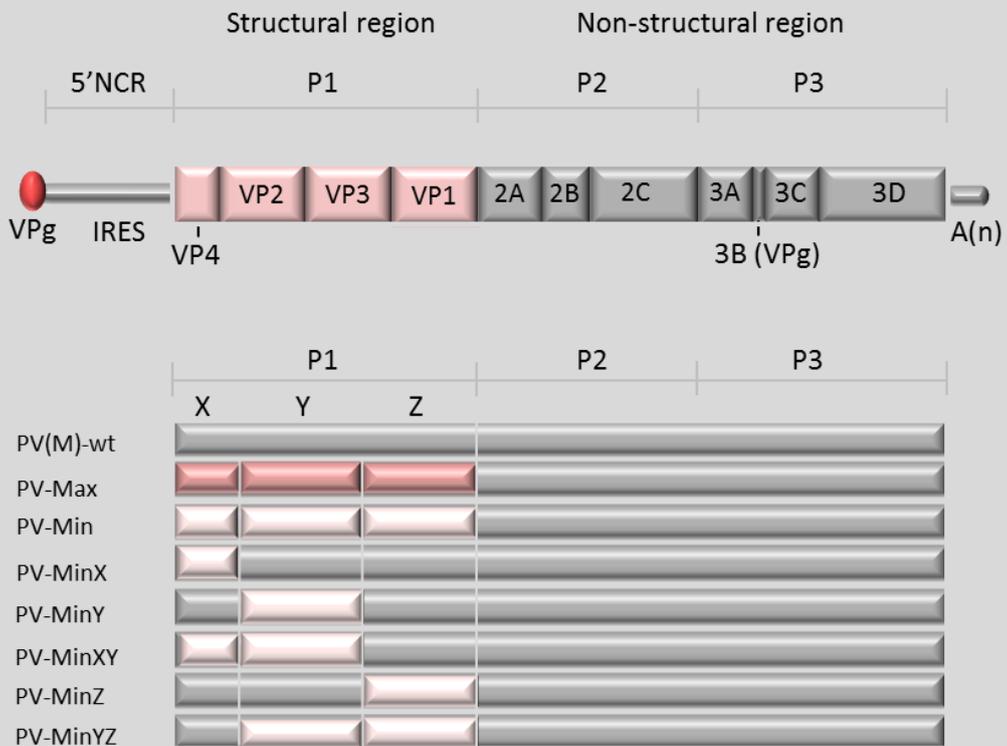


Figure 1. The genome of poliovirus aligned with partly synthetic poliovirus constructs. The single-stranded RNA genome is shown with the terminal protein VPg and the large internal ribosomal entry site (IRES) at the 5' end and the 3' NTR with the poly(A) tail at its 3' end. The polyprotein has been divided into 3 domains (P1, P2, P3) of which P1 is the precursor to four capsid proteins VP1-4, P2 and P3 specify replication proteins. The general organization of the various synthetic viruses is shown below. PV(M)-wt is the wild-type virus. PV-Min and PV-Max contain under and overrepresented codon pairs respectively (adapted from Coleman et al., 2008).

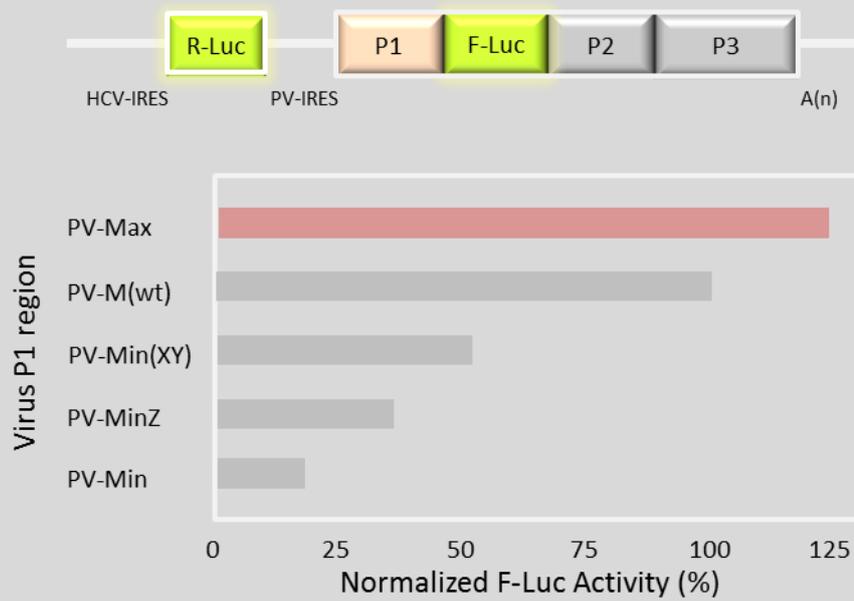


Figure 2. *In vivo* translational efficiency and codon pair bias. Structure of the dicistronic reporter used to measure the effect of codon pair bias on translation and below the amount of F-Luc produced by the various P1 encodings (adapted from Coleman et al., 2008).