CellPress

Review

Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness

Miguel Angel Martínez,[1,*] Ana Jordan-Paiz,[1] Sandra Franco,[1] and Maria Nevot[1]

Synthetic genome recoding is a novel method of generating viruses with altered phenotypes, whereby many synonymous mutations are introduced into the protein coding region of the virus genome without altering the encoded proteins. Virus genome recoding with large numbers of slightly deleterious mutations has produced attenuated forms of several RNA viruses. Virus genome recoding can also aid in investigating virus interactions with innate immune responses, identifying functional virus genome structures, strategically ameliorating cis-inhibitory signaling sequences related to complex viral functions, to unravel the relevance of codon usage for the temporal regulation of viral gene expression and improving our knowledge of virus mutational robustness and adaptability. The present review discusses the impacts of synonymous genome recoding with regard to expanding our comprehension of virus biology, and the development of new and better therapeutic strategies.

Trends

Chemical synthesis of viral genomes without a natural template is a useful tool for genetic modification on a large scale.

Attenuated viruses can be obtained by synonymously redesigning virus genomes based on changes in codon usage, codon pair usage, and/or dinucleotide content.

Synonymously recoding of viruses may lead to the generation of new classes of live attenuated vaccines.

In vaccines generated by synonymous recoding, all viral proteins are encoded with wild-type amino acid sequences that are exactly the same.

Synonymous virus genome recoding involves hundreds or thousands of nucleotide substitutions, possibly reducing the risk of virus phenotypic reversion via point mutations.

Large-scale synonymous recoding of virus genomes is a new tool that can assist in exploring different aspects of virus biology.

Synonymous But Not Neutral

Synonymous mutations are alterations in a DNA or mRNA sequence that do not change the protein amino acid sequence. Although synonymous mutations do not affect the resulting protein, they still have important consequences for cellular processes in all organisms [1,2]. The genetic code that is common to all organisms is extensively redundant, with four bases creating 61 coding codons for 20 amino acids, as well as three stop codons. A given amino acid might be encoded by one, two, three, four, or six different codons (Figure 1), and the observed ratios of synonymous codons are highly nonrandom [3]. In *Escherichia coli*, foreign protein expression is substantially affected by the presence or absence of rare codons, and heterologous protein expression often requires coding sequence manipulation [4].

A proposed codon bias theory hypothesizes that preferred codons correlate with abundances of isoaccepting tRNAs, and the degree of this correlation relates to the protein production levels of individual genes [5]. Recent experiments reveal that silent mutations impact many processes, including chromatin organization [6], enhancer functions [7,8], mRNA structure and folding [9,10], mRNA splicing [6] and stability [11], microRNA targeting [8,12], cotranslational folding [13], and transcription-factor binding [14] (Figure 2). Each factor constraining codon choice has impacted protein evolution [15].

Advances in large-scale low-cost construction of desired DNA sequences are rapidly influencing both fundamental and applied biological research [16,17]. Synthetic techniques for designing and manufacturing DNA enable *de novo* synthesis of complete viral genomes [18,19], offering the promise of completely controlling an organism's genetic information. Most synthetic

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

*Correspondence: mmartinez@irsicaixa.es (M.A. Martínez).

 http://dx.doi.org/10.1016/j.tim.2015.11.002

UUU Phe 17.6	UCU Ser 15.2	UAU Tyr 12.2	UGU Cys 10.6
UUC Phe 20.3	UCC Ser 17.7	UAC Tyr 15.3	UGC Cys 12.6
UUA Leu 7.7	UCA Ser 12.2	UAA Stop 1.00	UGA Stop 1.6
UUG Leu 12.9	UCG Ser 4.4	UAG Stop 0.8	UGG Trp 13.2
CUU Leu 13.2	CCU Pro 17.5	CAU His 10.9	CGU Arg 4.5
CUC Leu 19.6	CCC Pro 19.8	CAC His 15.1	CGC Arg 10.4
CUA Leu 7.2	CCA Pro 16.9	CAA Gln 12.3	CGA Arg6.2
CUG Leu 39.6	CCG Pro 6.9	CAG Gln 34.2	CGG Arg 11.4
AUU Ile 16.0	ACU Thr 13.1	AAU Asn 17.0	AGU Ser 12.1
AUC Ile 20.8	ACC Thr 18.9	AAC Asn 19.1	AGC Ser 19.5
AUA Ile 7.5	ACA Thr 15.1	AAA Lys 24.4	AGA Arg 12.2
AUG Met 22.0	ACG Thr 6.1	AAG Lys 31.9	AGG Arg 12.0
GUU Val 11.0	GCU Ala 18.4	GAU Asp 21.8	GGU Gly 10.8
GUC Val 14.5	GCC Ala 27.7	GAC Asp 25.1	GGC Gly 22.2
GUA Val 7.1	GCA Ala 15.8	GAA Glu 29.0	GGA Gly 16.5
GUG Val 28.1	GCG Ala 7.4	GAG Glu 39.6	GGG Gly 16.5

Met	Ala	Glu	Gly	Pro	Gly
AUG	GCA	GAG	GGC	CCA	GGA
	GCC	GAG	GGU	CCU	GGU
	GCU		GGA	CCC	GGC
	GCG		GGG	CCG	GGG

Trends in Microbiology

Figure 1. The Standard Genetic Code and Codon Bias. Except methionine and tryptophan, each amino acid is encoded by multiple codons. For each codon, the human frequencies per thousand are noted to the right. The Kazusa database (http://www.kazusa.or.jp/codon/) was used to compile human codon usage. Codons carrying CpG and TpA dinucleotides are underrepresented. Hexapeptide degenerate encodings are also shown (see lower box). This hexapeptide has a total of 512 ($1 \times 4 \times 2 \times 4 \times 4 \times 4$) possible encodings.

reconstructions have been of RNA viruses, due to their smaller genomes (~10 kb) and importance in health and biotechnology. The approximate cost of synthesizing a single extensively recoded genome of 10 kb in length is 2500 $. A recently developed strategy for virus attenuation is synthetic genome recoding by introducing large-scale underrepresented and slightly deleterious synonymous substitutions into the virus genome protein-coding region without altering the encoded proteins (Figure 3, Key Figure). Approaches have included codon deoptimization [20–23], random codon recoding [24,25], codon-pair deoptimization [26–33], and increasing the CpG/UpA dinucleotide frequency [34–36] (Box 1). All of these methods are based on the hypothesis that selection can act on gene sequences without amino acid changes – that is, that global modification of the viral genome induces attenuation because synonymous sites were shaped by genome-wide mutational processes during virus evolution [37,38]. It is well established that synonymous codon usage is neither random nor neutral [2,39,40]. Different organisms and different genes from the same organism include specific codons at different frequencies [1].

Synonymous virus genome recoding has allowed attenuation of the following RNA viruses: poliovirus [20,21,26,34], influenza virus [27,29], human immunodeficiency virus type 1 (HIV-1) [28], simian immunodeficiency virus (SIV) [41], Chikungunya virus (CHIKV) [24], human respiratory syncytial virus (RSV) [23,30], porcine reproductive and respiratory syndrome virus, echovirus 7 [31], tick-borne encephalitis virus (TBEV) [25], vesicular stomatitis virus [32], and dengue virus (DENV) [33] (Table 1). Genome recoding has also been used to modify the replication capacity and pathogenicity of two DNA viruses: adeno-associated virus and papillomavirus [42,43]. Attenuation by synonymous virus genome recoding represents an exciting new strategy for generating novel live-attenuated vaccine candidates because the amino acid coding is completely unaffected, thereby avoiding the generation of novel and undesirable biological properties. Moreover, synonymous virus genome recoding involves hundreds or thousands of nucleotide substitutions, greatly reducing the risk of phenotypic reversion via point mutations or through recombination with homologous sequences in circulating strains. This is particularly