Hidden Silent Codes in Viral Genomes

Eli Goz, Hadas Zur and Tamir Tuller

Abstract Viruses are small infectious agents that replicate only inside the living cells of other organisms and comprise approximately 94% of the nucleic acid-containing particles in the oceans. They are believed to play a central role in evolution, are responsible for various human diseases, and have important contributions to biotechnology and nanotechnology. Viruses undergo evolutionary selection for efficient transmission from host to host by exploiting the host’s gene expression machinery (e.g., ribosomes) for the expression of the genes encoded in their genomes. As a result, viral genes tend to be expressed via non-canonical mechanisms that are very rare in living organisms. Many of the gene expression stages and other aspects of the viral life cycle are encoded in the viral transcripts via ‘silent codes’, and are induced by mutations that are synonymous to the viral amino acid content. In a series of studies that included the analyses of dozens of organisms from the three domains of life, it was shown that there are overlapping ‘silent codes’ in the genetic code that are related to all stages of gene expression regulation. The aim of this chapter is to summarize the current knowledge related to the silent codes in viral genomes and the open questions in the field.
## 1 Hidden Information Related to Gene Expression Regulation and Other Aspects is Encoded in the Transcripts and Affects Organismal/Viral Fitness

Proteins are the principal actors in all intracellular activities. Gene expression is the process by which the information encoded in a gene is used to synthesize the corresponding protein. The major cellular biophysical stages of gene expression are transcription, splicing (in eukaryotes), mRNA degradation, translation, and protein degradation; each of these stages has several sub-stages (e.g., initiation, elongation, and termination of translation). For many years, researchers referred to the promoter (which mainly determines the transcription initiation rates) as the ‘module’ that includes almost all the information related to gene expression regulation, while the information related to protein structure is contained in the coding sequence via the genetic code.

However, in recent years, it was shown that such a modularity is only a raw approximation of the reality (Quax et al. 2015; Supek 2016; Sauna and Kimchi-Sarfaty 2013; Fredrick and Ibba 2010; Cannarozzi et al. 2010; Bahir et al. 2009; Gorgoni, et al. 2014; Gu et al. 2010; Zafrir and Tuller 2017; Tuller and Zur 2015; Ben-Yehezkel et al. 2015; Zafrir and Tuller 2015a, Yofe et al. 2014; Diament et al. (in press); Dana and Tuller 2014a; Zur and Tuller 2012, 2013, 2016; Tuller et al. 2010a, b, 2011a; Zafrir et al. 2016; Goz et al. 2017). Various signals (codes) related to all the stages of gene expression regulation, including its dynamics and amplitude, appear also in the coding sequence (ORF) itself and in the untranslated regions (UTRs), and are involved in biophysical interactions with the other segments of the transcript, and various macromolecules involved in gene expression regulation (Quax et al. 2015; Supek 2016; Sauna and Kimchi-Sarfaty 2013; Fredrick and Ibba 2010; Cannarozzi et al. 2010; Bahir et al. 2009; Gorgoni et al. 2014; Gu et al. 2010; Zafrir and Tuller 2015b; 2017; Tuller and Zur 2015; Ben-Yehezkel et al. 2015; Yofe et al. 2014; Diament et al. (in press); Dana and Tuller 2014b; Zur and Tuller 2012; 2013; 2016; Tuller et al. 2010a, b, 2011a; Zafrir et al. 2016; Goz et al. 2017) (see Figs. 1 and 2). Transcripts tend to also include information related to/affecting additional phenomena such as co-translational protein folding (Thommen et al. 2016; Chaney and Clark 2015) and regulation by the bacterial immune system (Terns and Terns 2011).

Specifically, it is interesting to emphasize that many of these ‘silent’ codes are encoded in the coding regions via the redundancy of the genetic code. A certain protein can be encoded by an exponential number of codon combinations; replacing a codon with a synonymous one can significantly affect the expression of the transcript (Quax et al. 2015; Supek 2016; Sauna and Kimchi-Sarfaty 2013; Fredrick and Ibba 2010; Cannarozzi et al. 2010; Bahir et al. 2009; Gorgoni et al. 2014; Gu et al. 2010; Tuller and Zur 2015, 2017; Ben-Yehezkel et al. 2015; Zafrir and Tuller 2015a; Yofe et al. 2014; Diament et al. (in press); Dana and Tuller 2014b; Zur and Tuller 2012; 2013; 2016; Tuller et al. 2010a, b, 2011a; Zafrir et al. 2016; Goz et al. 2017; Bazzini et al. 2016; Morgunov et al. 2014; Sin et al. 2016).
The information related to these codes is considered ‘hidden’ as it is partially encoded in synonymous/‘silent’ aspects of the transcript, and is much harder to model than the genetic code. Regulation of gene expression is clearly at the heart of

Fig. 1 Some of the interactions of the mRNA molecule with the gene expression machinery. The affinities of these interactions are encoded in the UTRs and ORFs of the genes

The information related to these codes is considered ‘hidden’ as it is partially encoded in synonymous/‘silent’ aspects of the transcript, and is much harder to model than the genetic code. Regulation of gene expression is clearly at the heart of
every biological system. Thus, understanding how aspects of this process are encoded in the transcript should have important ramifications to every biomedical discipline (e.g., human health, synthetic biology, molecular evolution, genetics, systems biology, etc.).

It is important to emphasize that while there are studies that suggest codon usage bias is related to mutation drift (Bulmer 1991), various lines of evidence have recently demonstrated that codon usage bias directly affects the translation elongation speed. Specifically, based on direct experimental measurements of ribosome densities (which are related to elongation rates) over the entire transcriptome at a single codon resolution, it was shown that different codons have different elongation rates, which correlate with corresponding tRNA levels (Dana and Tuller 2014b; Gardin et al. 2014). It was also experimentally shown that changing the codon content of a protein directly affects protein levels (Ben-Yehezkel et al. 2015; Gustafsson et al. 2004), and thus the organism’s fitness.

2 The Importance of Understanding the ‘Silent’ Information Encoded in Viral Genomes

Viruses are small infectious agents that replicate only inside the living cells of other organisms. They are comprised of genetic material (RNA or DNA molecule(s)) and often additional enzymes that are enclosed within a protective coat of lipids and proteins. The viral genome contains all necessary information to initiate and complete a replication cycle within a cell. Viruses can infect all living organisms (fungi, plants, bacteria, mammals, etc.); we eat and breathe billions of virus particles regularly and carry viral genomes as part of our own genetic material.

Viruses are by far the most abundant biological entities in the oceans, comprising approximately 94% of the nucleic acid-containing particles (Zimmer 2011). They are believed to play a central role in evolution as they are important natural
means of transferring genes between different species (Zimmer 2011). In addition, viruses are responsible for various human diseases: Some of them are common (e.g., common cold, influenza, chickenpox, cold sores), others are severe and fatal (e.g., ebola, AIDS, avian influenza, and SARS). Viruses also have important implications to biotechnology (e.g., they are often used as vectors to introduce genes into cells), and even to materials science and nanotechnology (e.g., they can be used as organic nanoparticles) (Fischlechner and Donath 2007).

The viral genomes undergo evolutionary selection for efficient transmission from host to host, and for exploiting the gene expression machinery of the host (e.g., ribosomes, various transcription/translation factors, etc.) for efficient synthesis of the encoded proteins and the efficient expression of various types of genes (e.g., see Gale et al. 2000). As a result, viral genes tend to be expressed via non-canonical mechanisms that are either specific only to viruses, or very rare in living organisms (e.g., see Gale et al. 2000; Firth and Brierley 2012; Rohde et al. 1994; Brierley 1995; Lopez-Lastra et al. 2010; Fig. 3). For example, viruses tend to include

---

**Fig. 3** Examples of non-canonical viral mRNA translation. **a** Canonical mRNA translation, which translates a single ORF from the start codon (AUG) to the stop codon (UAG, or UAA or UGA). **b, c, d, e** and **f** Non-canonical translation mechanisms: **b** Ribosomal re-initiation. **c** Diversity of start codon and near-cognate start codons. **d** Cap-independent translation via IRES (internal ribosome entry site). **e** Upstream-ORF (uORF). **f** Multiple/internal ORFs (either in-frame or out-of-frame)
overlapping ORFs and they often include a long ORF translated into a single polyprotein that is cleaved posttranslationally into a set of mature proteins. Viruses also tend to initiate translation from internal ribosome entry sites (IRES), and not via canonical scanning from the 5’ end of the transcript. Furthermore, frequently viral genes contain functional mRNA structures related to all stages of their expression regulation. Furthermore, frequently they include strong mRNA structure related to all stages of their gene expression regulation. Regularly, the viral genetic material is RNA and not DNA and can undergo a series of transformations before translation into proteins. Finally, most of the viral genomes are very compact and include all their gene expression information in a very short genome (typically a few thousand nucleotides), etc.

One gene expression aspect common to all viruses is the fact that all types of viruses must use the ribosomes (and other expression machinery) of their host.

It is important to emphasize that viruses evolve to include non-canonical gene expression mechanisms since these non-canonical regulatory mechanisms contribute to their fitness. Specifically, often during viral development, the canonical gene expression mechanisms in the cell are ‘shut down’ (e.g., due to down-regulation of relevant initiation factors); since viruses bypass these canonical mechanisms (via non-canonical mechanisms, e.g. IRES), they can successfully exploit the intracellular gene expression resources (e.g., ribosomes and tRNAs). Some of these non-canonical mechanisms (e.g., overlapping ORFs) enable a more efficient (in terms of energy) production of viruses, and decreasing the probability of deleterious mutations (Holmes 2009). Among others, this means that it is less trivial to understand the viral silent gene expression codes as they are relatively rare and unique (Gale et al. 2000; Firth and Brierley 2012; Rohde et al. 1994; Brierley 1995; Lopez-Lastra et al. 2010; Adrian et al. 2005; Holland 2012).

3 Previous Relevant Studies Concerning ‘Silent’ Information Related to Gene Expression in Viruses

Various studies in recent years have provided statistical evidence that silent aspects in the viral genomes are related to their fitness.

Specifically, among others, it was suggested that very basic features, such as mRNA folding, codon decoding times, codon or nucleotide pairs distributions (or other low order statistics of genomic sequences), may be induced by synonymous mutations and play an important role in controlling the viral life cycle (Bahir et al. 2009; Cuevas et al. 2012; Lobo et al. 2009; Jenkins et al. 2001; Greenbaum et al. 2008; van Hemert et al. 2007; Pride et al. 2006; Cardinale and Duffy 2011; Shackelton et al. 2006; Carbone 2008; Gu et al. 2004; Sau et al. 2005a, 2007; Zhao et al. 2008; Cheng et al. 2012; Lucks et al. 2008; Mueller et al. 2006). In this subsection, detail the different silent aspects of viral gene expression that have been reported thus far.
3.1 Selection for Codon Preference in Viral Genomes

The most basic property of the viral coding sequences is the frequencies of the different codons. The tendency to choose specific codons has been shown to affect/regulate intracellular mechanisms (Supek 2016; Sauna and Kimchi-Sarfaty 2013; Tuller and Zur 2015; Novoa et al. 2012): for example, it may affect translation elongation (Dana and Tuller 2014b; Gardin et al. 2014; Ben-Yehezkel et al. 2015), translation initiation (Tuller and Zur 2015; Zur and Tuller 2013; Kozak 1986), splicing (Zafir and Tuller 2015b; Chamary and Hurst 2005), mRNA folding (Gu et al. 2010; Zur and Tuller 2012; Tuller et al. 2010), protein folding (Pechmann and Frydman 2013; Kramer et al. 2009), and more; thus, we expect that viral codon bias will be under selection pressure.

Indeed, many studies have suggested that viral codons may be under selection to improve the viral fitness, for example, via adaptation to the host tRNA pool (or other translation resources) (Bahir et al. 2009; Burns et al. 2006; Tao et al. 2009; Jia et al. 2009; Zhou et al. 2010; Liu et al. 2010, 2011; Das et al. 2006; Cai et al. 2009; Sau et al. 2005b; Wong et al. 2010; Zhong et al. 2007; Zhang et al. 2013; Novella et al. 2004; Michely et al. 2013; Roychoudhury and Mukherjee 2010; Ma et al. 2011; Aragones et al. 2010; Tsai et al. 2007; Su et al. 2009; Bull et al. 2012; Zhao et al. 2005). The adaptation of the viral codon usage bias to the tRNA pool is expected to improve translation efficiency via better allocation of the limited translation resources (e.g., ribosomes and tRNA molecules) (Dana and Tuller 2014b; Rocha 2004; Sharp et al. 2005).

In order to study the effects and extents of codon usage bias many measures have been developed (Sharp and Li 1987; dos Reis et al. 2004; Sabi et al. 2016; Wright 1990).

For example, Bahir et al. (Bahir et al. 2009) analyzed a large data set of viruses that infect hosts ranging from bacteria to humans. They show that bacteria-infecting viruses are strongly adapted to their specific hosts in terms of codon usage bias but that they differ from other unrelated bacterial hosts. Viruses that infect humans, but not those that infect other mammals or aves, show a strong resemblance to most mammalian and avian hosts, in terms of codon preferences. This observation can be partially explained by the following points: (1) There is similarity in the codon usages among most mammals (Bahir et al. 2009). (2) The codon usage bias among bacteria is very high (Bahir et al. 2009). (3) Bacteria (and thus probably also their viruses) usually undergo stronger selection for codon usage bias and for various aspects of translation optimality (among others due to their larger population size) relatively to most eukaryotes (dos Reis et al. 2004; dos Reis and Wernisch 2009). (4) Additional explanations may be related to the recent expansion of humans and the coevolution of their viruses, or to the hypothesis that large portions of the human genome are actually of viral origin (Bahir et al. 2009; Kazazian 2004).

Pavesi et al. suggested that the fact viruses undergo selection to include specific codons can help detecting new and ancestral viral coding regions (Pavesi et al. 2013). Aragonès et al. suggested that the Hepatitis A virus undergoes various types of adaptations to fine-tune the translation kinetics, among others, via selection on
codon usage bias (Aragones et al. 2010). A study by Bull et al. (2012) has shown that when reengineering the major capsid gene of the bacteriophage T7 with varying levels of suboptimal synonymous codons, the fitness of the constructs declines linearly with the number of suboptimal changes. These experiments/analyses suggest a direct relation between codon usage bias and fitness/fitness-recovery. Similarly, a related study by Lauring et al. (2012) compared the wild-type poliovirus to synthetic viruses carrying reengineered capsid sequences with hundreds of synonymous mutations. They found that such mutations are related to the rewiring of the population’s mutant network which reduced its robustness to mutations and attenuated the virus in an animal model of infection.

It is important to mention that some of these codon usage bias patterns may be associated with regulatory signals not necessarily directly related to tRNA levels (Gog et al. 2007); alternative or partial explanations to viral codon usage bias are mutational bias, asymmetrical mutational bias in two DNA strands, temperature, viral replication mechanisms, protein folding, dinucleotide distribution, mRNA folding, and more (Das et al. 2006; Zhang et al. 2011, 2013; Sau and Deb 2009; Adams and Antoniw 2004; Cardinale et al. 2013; Berkhout et al. 2002; Pinto et al. 2007; Cladel et al. 2008; Choi et al. 2005; Zhou et al. 2013; Burns et al. 2009; Liu et al. 2012). The effect on chromatin structure and nucleosome positioning is another potential constraint on the viral codon frequency distribution, as viruses are exposed to histones produced by the host (Eslami-Mossallam et al. 2016; Cohanim and Haran 2009; Babbitt and Schulze 2012).

Interestingly, a recent line of studies suggested that codon pairs’ distribution is an important feature under selection in various viruses, which may be used for their attenuation for developing new vaccines (Coleman et al. 2008; Mueller et al. 2008; Martrus et al. 2013). However, there is a debate regarding this feature, while some researchers believe that it is related directly to the distribution of codon pairs (Coleman et al. 2008; Mueller et al. 2008; Martrus et al. 2013), others have suggested that it is related to the distribution of dinucleotides (Tulloch et al. 2014) which affect RNA folding (see, e.g., Babak et al. 2007), or may be related to the enhanced innate immune responses to viruses with elevated CpG/UpA dinucleotide frequencies rather than the viruses themselves being intrinsically defective (Tulloch et al. 2014; Belalov and Lukashev 2013). These possible explanations still connect the viral fitness to silent features of its genome, demonstrating their importance and influence on viral fitness and evolution. Finally, it is important to emphasize the fact that many silent viral codes are localized to specific regions within the genome (Dumans et al. 2004).

### 3.2 Evidence for Condition Specific Adaptation to Codon Bias

Intriguingly, a recent study has provided evidence of selection for distinct compositions of synonymous codons in viral genes that are expressed at different stages of the viral life cycle (e.g., early and late viral genes): It was shown that in the
bacteriophage lambda, evolution of viral coding regions is driven, among others, by
codon ‘selection’ which is specific to the expression time of the gene during the
viral development (e.g., early expressed genes versus late expressed genes).
Specifically, during the initial/progressive stages of infection, the decoding rates in
early/late genes were found to be superior to those in late/early genes, respectively
(Goz et al. 2017) (Fig. 4). This study is important since it is the first to show that the
selection for codon usage in the virus is directly related to translation elongation
rates. In addition, it was shown for the first time that codon elongation rates change
during viral evolution; thus, this is expected to affect the codons ‘selected’ for each
viral gene based on its expression time during the viral development cycle.
Currently, due to the absence of experimental measurements, this result has been
demonstrated only in one virus (bacteriophage lambda), since to perform such an
analysis one needs to infer the codon decoding rates in different time points of the
viral development. This can be achieved only via relevant experiments (Ingolia
et al. 2009) and data filtering (Dana and Tuller 2014b), and the viral genome alone
is not enough.

Specifically, one type of relevant experiment is Ribo-Seq which provides
large-scale information (the entire transcriptome) related to the probability for
seeing a ribosome over each codon in the transcriptome in vivo (Ingolia et al.
2009). These experiments, when performed in different viral development stages,
can be used for estimating the decoding rates of different codons in different viral
conditions (Goz et al. 2017; Liu et al. 2013). Since Ribo-Seq data includes various
sources of bias and noise, the data should be analyzed with tools tailored speci-

cally for parameter estimation and bias filtering in the Ribo-Seq experiments (Dana
and Tuller 2014b; Diament and Tuller 2016).

As can be seen in (Fig. 4a), to estimate codon decoding rates we do not compute
a simple average of the normalized Ribo-Seq footprint count (NFC). The main
reasons that a simple average does not work are related to: (1) The fact that
Ribo-Seq includes various types of non-trivial biases (e.g., very extreme values in
certain positions due to the biochemistry of the protocol) (Dana and Tuller 2012,
2014b; Diament and Tuller 2016; Gerashchenko and Gladyshev 2017). (2) Codons
upstream of slower codons will have more reads due to traffic jams (Dana and
Tuller 2014b). (3) Codons downstream of slower codons will have more reads due
to incomplete halting of the ribosomes movement during the Ribo-Seq experiment
(Hussmann et al. 2015).

Consequently, the NFC always has a very thick right tail. It was shown via
simulations of the Ribo-Seq procedure (Dana and Tuller 2014a, b) that without the
aforementioned problems, the NFC distribution is close to normal (resembles a
Gaussian without the thick right tail). Thus, to estimate the nominal decoding rate,
we must filter the right tail. It was shown via Ribo-Seq simulations that the sug-
gested filtering estimates the correct decoding times, but due to the reasons
explained above merely taking the mean of the entire NFC distribution does not
correlate with the actual decoding times (Dana and Tuller 2014b).

We believe that in the future, similar results will be reported for additional
viruses.
(a) Arrest translation

Purify fragments protected by ribosomes

Digest unprotected mRNA

Map fragments to genome and create a normalized footprints count (NFC) profile

Typical codon decoding rate (TDR) estimation

(b) Relative ribo-seq viral read counts

(c) Relative translation elongation efficiency coefficient (RTEC)

(d) Probability

Mean MTDR
Various previous studies have suggested that the UTRs of many viruses include important functional structures (Watts et al. 2009; Firth et al. 2011; Brown et al. 1992; Hyde et al. 2014; Abbink and Berkhout 2003). For example, it was demonstrated that extensive structural elements that modulate RNA replication via different conformations appear in the 5′ and 3′ UTRs of Dengue and other flaviviruses. The promoter for Dengue virus RNA synthesis is a large stem-loop structure located at the 5′ end of the genome. This structure specifically interacts with the viral polymerase NS5 and promotes RNA synthesis at the 3′ end of a circularized genome. The circular conformation of the viral genome is mediated by long-range RNA–RNA interactions that span thousands of nucleotides (Fig. 5).

As another example, the genomes of human hepatitis C virus (HCV), and the animal pestiviruses responsible for bovine viral diarrhea (BVDV) and hog cholera (HChV), have a conserved (and probably functional) stem-loop structure in the 3′ 200 bases of the 5′UTR (Brown et al. 1992). A different study (Hyde et al. 2014) suggested that the pathogenic alphaviruses use secondary structural motifs within the 5′UTR as part of an evasion mechanism by which viruses avoid immune restriction.
Interestingly, Firth et al. (2011) analyzed the \( C_{24} \) nt 3′-adjacent to the stop codon (UGA) in Sindbis, Venezuelan equine encephalitis related alphaviruses, and in the plant virus genera (Furovirus, Pomo virus, Tobravirus, Pecluvirus and Benyvirus); they found a phylogenetically conserved stem-loop structure. Mutational analysis of the predicted structure demonstrated that the stem-loop increases read-through by up to ten-fold. Thus, this structure has an important function: increasing read-through probability.

An interesting question is related to the possibility that such important functional structures appear inside the coding regions of viruses. To check this possibility, the strength of the structures within the coding regions of viral genomes can be compared to the ones we ‘expect to obtain’ under a ‘null evolutionary model’ that generates viral genomes with similar properties to the original genome (such as, encoded proteins, GC content, codon frequencies, identical distances/alignment-scores between the viral strains of the same virus). Two recent studies have performed such analyses (Goz and Tuller 2015, 2016).

In these papers (Goz and Tuller 2015, 2016), 1666 genomes of the four Dengue serotypes and the HIV genome were analyzed, using statistical/computational analyses to detect dozens of positions suspected to undergo selection for weak/strong local mRNA folding (probably many of them are related to viral fitness), while controlling for the false discovery rate.

An extensive position-specific selection for global and local mRNA structures in these viruses was demonstrated (Goz and Tuller 2015, 2016) (see also Goz et al. 2017). In addition, since robustness to mutations is an important factor that influences viral evolution (expressly in the case of RNA viruses) (Lauring et al. 2013), it was specifically interesting to provide evidence related to the robustness of some of these structures to mutations/errors (Goz and Tuller 2016) (Fig. 6).

Inference of the HIV RNA structure (Watts et al. 2009) suggested that there is correlation between high levels of RNA structure and sequences that encode inter-domain loops in HIV proteins.

It was shown that RNA structure can effect translation elongation rates (Tuller et al. 2011a; Dana and Tuller 2012); it was also shown that the elongation rates can effect co-translational folding (Zur and Tuller 2016; Yang et al. 2014; Faure et al. 2017).
Hidden Silent Codes in Viral Genomes

(a) AGCGGGGAGACAUAUAUCAGCCU...

(b) 150 nt sliding window for MFE prediction in position i

(c) WT

\[
\begin{align*}
&\text{CGTCTGG} \\
&\text{CGTCTGG} \\
&\text{CGTCTGG} \\
&\text{CGTCTGG} \\
&\text{CGTCTGG}
\end{align*}
\]

\[
\text{MT}
\]

\[
\begin{align*}
&\text{ACGGTCGTC} \\
&\text{ACGGTCGTC} \\
&\text{ACGGTCGTC} \\
&\text{ACGGTCGTC} \\
&\text{ACGGTCGTC}
\end{align*}
\]

\[
\text{SMR} = \frac{1}{|\text{Mutations}|} \sum_{\text{Mutations}} \frac{L - d[S(\text{WT}), S(\text{MT})]}{L}
\]

(d) MINIMUM FREE FOLDING ENERGY (Kcal/mol)

(e) Structure preserving randomization model

p-value = 0.028, z-score = 1.2
Thus, it is possible that, among others, the RNA structure modulates ribosome elongation to promote native protein folding. It was shown that unstructured RNA regions tend to include splice site acceptors and hypervariable regions. The HIV genome also includes a functional ribosomal gag-pol frameshift stem-loop.

These results suggest that the coding regions, and not only the UTRs, of various viruses are populated with local RNA structures that are important for the viral life cycle and fitness.

### 3.4 Longer Hidden Codes in the Viral Coding Sequence

As we mentioned in the previous section, we expect the viral coding region to include many codes/patterns that are important for the viral fitness and are longer and more complex than the single codon distribution. Thus, to show this we recently performed large-scale analyses of most all the viruses with available genomes and their hosts with a novel method for detecting hidden silent codes (that cannot be explained by codon bias) (Zur and Tuller 2015) in the viral genetic material. The new statistical measure compares that mean repetitive patterns in the
viral and host genome and identify signals that are not expected to appear in these genomes based only on the distribution of single codons (Goz et al. 2017; Zur and Tuller 2015; Goz and Tuller 2017) (see Fig. 7).

Based on this analysis, we were able to detect significant patterns of such codes (repetitive sequences) in a high percentage of the analyzed viruses (33–90% for different groups of viruses classified according to their host) and in 90% of the bacteriophages (Goz et al. 2017; Goz and Tuller 2017).

Fig. 7 a The statistical approach for evaluating the tendency of a viral coding region to include long subsequences that tend to appear in the host. At each position in the coding region, the length of the longest subsequence starting in this position that also appears in the host is computed. The average longest host repetitive score (AHRS) is the average of all these lengths. b To evaluate the statistical significance of the AHRS in the viral coding sequences, the score was compared to the ones obtained for randomized versions of the viral genomes maintaining the proteins, codon frequencies, dinucleotide frequencies, and GC content. The figure includes the analysis for the bacteriophage lambda (Goz et al. 2017).
4 Discussion

Engineering viruses

It is important to mention that there are some preliminary studies regarding gene expression engineering/modeling and other related aspects (see, e.g., Gorgoni et al. 2014; Tuller et al. 2011a; Sin et al. 2016; Konur et al. 2016; Sanassy et al. 2015; Wu et al. 2016; Schoech and Zabet 2014; Cheng et al. 2016; Pan et al. 2016; Haldane et al. 2014; Raveh et al. 2016; Reuveni et al. 2011), but none dealing with complete viruses. Thus, one open question is related to the development of practical strategies for engineering viruses based on the hidden/silent information. Developing approaches for controlling these codes should enable us to manipulate (e.g., increase or decrease) the expression levels of viral genes, and thus to modulate various viral phenotypes such as replication rates.

Therefore, based on such an approach, it will be possible to efficiently engineer viruses (Wimmer et al. 2009) for various objectives related to human health such as the design of live attenuated and killed vaccines (Lauring et al. 2010). Today, almost all the approaches for designing vaccines are based on non-synonymous alterations of the viral genomes, ignoring the largest fraction of the information (i.e., the silent information) encoded in the viral genome. Indeed, some preliminary studies have suggested that modulating simple features, such as codon and codon-pair usage, and local mRNA folding, can be used for the development of live attenuated vaccines (Coleman et al. 2008; Goz and Tuller 2015; Nogales et al. 2014).

Such an approach can also be generalized to engineer bacteriophages for various objectives such as ‘fighting’ pathogenic bacteria resistance to antibiotics, and engineering the human microbiome. It may also be used to design better oncolytic viruses with improved replication/fitness in cancerous cells but not in healthy ones.

Is it possible that some of the silent codes are related to the immune system?

In this book chapter, we emphasized the relation between sequence patterns in the viral coding sequences and transcripts, and viral fitness, via their effect on gene expression. However, it is possible that some of these patterns are related not only to gene expression, but also to the evolution of the virus for escaping the host immune system. It is important to emphasize that in most of the analyses, we and others reported (some are mentioned above), the amino acid content of the viral genes was controlled for. Thus, the reported signals cannot trivially be attributed only to the classical mechanisms, such as viral recognition by the host (e.g., antibodies), as these mechanisms are traditionally believed to be based on interactions between proteins. However, it is very plausible that they are related to alternative known and/or unknown mechanisms.

One very relevant such mechanism in bacteria is clustered regularly interspaced short palindromic repeats (CRISPR; see Fig. 8) (Marraffini 2015; Horvath and Barrangou 2010). This mechanism is based on creating fragments in the viral genome that are transcribed to short RNA molecules (crRNAs); these short RNA molecules match a certain region in the viral genome and ‘guide’ a protein complex
(CAS-crRNA complex) that cuts the viral DNA in this region and inactivates the virus. Since this mechanism is based on the recognition of short DNA subsequences that should appear in the virus/phage but not in the host, this may trigger evolution of the nucleotide composition of the virus/phage to be similar to the host. This may result in similar patterns of codons, and longer sequences that appear in the phage and the host, explaining some of the results reported here (Goz and Tuller 2017).

Relation to horizontal gene transfer (HGT)

Finally, it is important to emphasize that similarly to viral adaptation to the host, silent features of the coding regions are expected to affect related phenomena such as horizontal gene transfer (HGT). In this case, a transferred gene is expected to be successfully expressed in a new host if its silent features are compatible (Tuller et al. 2011b; Tuller 2011, 2012; Roller et al. 2013; Medrano-Soto et al. 2004). Thus, many of the results reported here may be generalized to the case of HGT.

It is important to emphasize that a central HGT mechanism is transduction, the process in which bacterial DNA is moved from one bacterium to another by a virus/bacteriophage (Soucy et al. 2015). Thus, the reported relations between (1) the

Fig. 8 The short palindromic repeat (CRISPR)-Cas system provides adaptive immunity against foreign elements in prokaryotes: Upon viral injection, a small sequence of the viral genome, known as a spacer, is integrated into the CRISPR locus to immunize the host cell. Spacers are transcribed into small RNA guides that direct the cleavage of the viral DNA by Cas nucleases (Horvath and Barrangou 2010)
host silent codes and (2) the transferred gene silent codes have much overlap:
The fact that viral fitness is related to the similarity of its silent aspects/codes to the
host should directly improve its ability to transfer genes; it is also directly related to
the fact that the silent aspects/codes in the transferred genes are more adapted to the
new host since the virus undergoes evolution to be better adapted to the host.

Some preliminary studies of heterologous gene expression have suggested that
introducing a foreign gene with a distinct codon distribution to the host results in a
decrease in the host’s fitness and the gene’s protein levels (Gustafsson et al. 2004;
Ben-Yehezkel et al. 2015; Tuller et al. 2011; Welch et al. 2009). Computational
models have suggested that this is partially due to the fact that such genes recruit
more ribosomes (slower codons result in ribosomes spending more time on the
mRNA), the number of available ribosomes decreases, the global initiation rates of
the host genes decreases, and thus the host fitness decreases (Raveh et al. 2016;
Tuller et al. 2011b; Tuller 2011) (though many additional explanations exist (Tuller
and Zur 2015; Tuller 2012; Welch et al. 2009; Angov 2011)). However, additional
experimental studies should be performed to better understand the effect of the
codon bias of a transferred gene on the transferred gene expression and the host
fitness.

Acknowledgements This study was supported in part by a fellowship from the Edmond J. Safra
Center for Bioinformatics at Tel-Aviv University. TT is partially supported by the
Minerva ARCHES award.

References

Abbink TE, Berkhout B (2003) A novel long distance base-pairing interaction in human
immunodeficiency virus type 1 RNA occludes the gag start codon. J Biol Chem 278:11601–11611
Adams MJ, Antoniw JF (2004) Codon usage bias amongst plant viruses. Arch Virol 149:113–135
Adrian JG et al (2005) Molecular basis of virus evolution. University Press, Cambridge
Angov E (2011) Codon usage: nature’s roadmap to expression and folding of proteins.
Biotechnol J 6:650–659
Aragones L et al (2010) Fine-tuning translation kinetics selection as the driving force of codon
usage bias in the hepatitis A virus capsid. PLoS Pathog 6:e1000797
Babak T et al (2007) Considerations in the identification of functional RNA structural elements in
genomic alignments. BMC Bioinform 8:33
Babbitt GA, Schulze KV (2012) Codons support the maintenance of intrinsic DNA polymer
flexibility over evolutionary timescales. Genome Biol Evol 4:954–965
Bahir I et al (2009) Viral adaptation to host: a proteome-based analysis of codon usage and amino
cid preferences. Mol Syst Biol 5:1–14
Bazzini AA et al (2016) Codon identity regulates mRNA stability and translation efficiency during
the maternal-to-zygotic transition. EMBO J 35:2087–2103
Ben-Yehezkel T et al (2015) Rationally designed, heterologous S. cerevisiae transcripts expose
novel expression determinants. RNA Biol 12:972–984
Belalov IS, Lukashev AN (2013) Causes and implications of codon usage bias in RNA viruses.
PLoS One 8:e56642
Ben-Yehezkel T et al (2015) Rationally designed, heterologous S. cerevisiae transcripts expose
novel expression determinants. RNA Biol 12:972–984
Berkhout B et al (2002) Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS Res Hum Retroviruses 18:133–141
Brierley I (1995) Ribosomal frameshifting viral RNAs. J Gen Virol 76(Pt 8):1885–1892
Brown EA et al (1992) Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. Nucleic Acid Res 20:5041–5045
Bull JJ et al (2012) Slow fitness recovery in a codon-modified viral genome. Mol Biol Evol 29:2997–3004
Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907
BURNS CC et al (2006) Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol 80:3259–3272
Burns CC et al (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
Cai MS et al (2009) Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. Intervirology 52:266–278
Cannarozzi G et al (2010) A role for codon order in translation dynamics. Cell 141:355–367
Carbone A (2008) Codon bias is a major factor explaining phage evolution in translationally biased hosts. J Mol Evol 66:210–223
Cardinale DJ, Duffy S (2011) Single-stranded genomic architecture constrains optimal codon usage. Bacteriophage 1:219–224
Cardinale DJ et al (2013) Base composition and translational selection are insufficient to explain codon usage bias in plant viruses. Viruses 5:162–181
Chamary JV, Hurst LD (2005) Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? Trends Genet 21:256–259
Chaney JL, Clark PL (2015) Roles for synonymous codon usage in protein biogenesis. Annu Rev Biophysics 44:143–166
Cheng Z et al (2016) Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. Mol Syst Biol 12:855
Cheng XF et al (2012) High codon adaptation in citrus tristeza virus to its citrus host. Virol J 9:113
Choi IR et al (2005) An internal RNA element in the P3 cistron of wheat streak mosaic virus revealed by synonymous mutations that affect both movement and replication. J Gen Virol 86:2605–2614
Cladel NM et al (2008) CRPV genomes with synonymous codon optimizations in the CRPV E7 gene show phenotypic differences in growth and altered immunity upon E7 vaccination. PLoS One 3:e2947
Cohenim AB, Haran TE (2009) The coexistence of the nucleosome positioning code with the genetic code on eukaryotic genomes. Nucleic Acid Res 37:6466–6476
Coleman JR et al (2008) Virus attenuation by genome-scale changes in codon pair bias. Science 320:1784–1787
Cuevas JM et al (2012) The fitness effects of synonymous mutations in DNA and RNA viruses. Mol Biol Evol 29:17–20
Dana A, Tuller T (2012) Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. PLoS Comput Biol 8:e1002755
Dana A, Tuller T (2014a) Properties and determinants of codon decoding time distributions. BMC Genomics 15(Suppl 6):S13
Dana A, Tuller T (2014b) The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acid Res 42:9171–9181
Das S et al (2006) Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res 117:227–236
Diament A, Tuller T (2016) Estimation of ribosome profiling performance and reproducibility at various levels of resolution. Biol Direct 11:24
Diament A et al (2014) Three dimensional genomic organization of eukaryotic genes is correlated with their expression and function. Nat Commun (in press)
dos Reis M, Wernisch L (2009) Estimating translational selection in eukaryotic genomes. Mol Biol Evol 26:451–461

dos Reis M et al (2004) Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acid Res 32:5036–5044

Dumans AT et al (2004) Synonymous genetic polymorphisms within Brazilian human immunodeficiency virus Type 1 subtypes may influence mutational routes to drug resistance. J Infect Dis 189:1232–1238

Eslami-Mossallam B et al (2016) Multiplexing genetic and nucleosome positioning codes: a computational approach. PLoS One 11:e0156905

Faure G et al (2016) Role of mRNA structure in the control of protein folding. Nucleic Acid Res 44:10898–10911

Firth AE, Brierley I (2012) Non-canonical translation in RNA viruses. J Gen Virol 93:1385–1409

Firth AE et al (2011) Stimulation of stop codon readthrough: frequent presence of an extended 3’ RNA structural element. Nucleic Acid Res 39:6679–6691

Fischlechner M, Donath E (2007) Viruses as building blocks for materials and devices. Angew Chem Int Ed Engl 46:3184–3193

Fredrick K, Ibba M (2010) How the sequence of a gene can tune its translation. Cell 141:227–229

Gale MJr et al (2000) Translational control of viral gene expression in eukaryotes. Microbiol Mol Biol Rev 64:239–280

Gardin J et al. (2014) Measurement of average decoding rates of the 61 sense codons in vivo. Elife 3:10.7554/eLife.03735

Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res 45(2):e6

Gorgoni B et al (2014) Controlling translation elongation efficiency: tRNA regulation of ribosome flux on the mRNA. Biochem Soc Trans 42:160–165. doi:10.1042/BST20130132

Goz E et al (2017) Evidence of translation efficiency adaptation of the coding regions of the bacteriophage lambda, DNA Res

Greenbaum BD et al (2008) Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog 4:e1000079

Gu W et al (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput Biol 2010(6):1–8

Gog JR et al (2007) Codon conservation in the influenza A virus genome defines RNA packaging signals. Nucleic Acids Res 35:1897–1907

Goz E, Tuller T (2015) Widespread signatures of local mRNA folding structure selection in four dengue virus serotypes. BMC Genom 16(Suppl 10):S4

Goz E, Tuller T (2016) Evidence of a direct evolutionary selection for strong folding and mutational robustness within HIV coding regions. J Comput Biol 23:641–650

Goz E, Tuller T (2017) Widespread selection for complex patterns of synonymous information in viral coding regions. In Review

Gu W et al (2004) Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales, Virus Res 101:155–161

Gustafsson C et al (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346–353

Haldane A et al (2014) Biophysical fitness landscapes for transcription factor binding sites. PLoS Comput Biol 10:e1003683

Holland JJ (2012) Genetic diversity of RNA viruses. Springer Science & Business Media

Holmes EC (2009) The evolution and emergence of RNA viruses. Oxford University Press Inc, New York

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167–170

Husssmann JA et al (2015) Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. PLoS Genet 11:e1005732

Hyde JL et al (2014) A viral RNA structural element alters host recognition of nonself RNA. Science 343:783–787
Ingolia NT et al (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
Jenkins GM et al (2001) Evolution of base composition and codon usage bias in the genus Flavivirus. J Mol Evol 52:383–390
Jia R et al (2009) Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. Virus Genes 38:96–103
Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632
Konur S et al (2016) An integrated in silico simulation and biomatter compilation approach to cellular computation. In: Adamatzky A. (ed) Advances in unconventional computing, vol 23. pp. 655–676
Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292
Kramer G et al (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol 16:589–597
Lauring AS et al (2010) Rationalizing the development of live attenuated virus vaccines. Nat Biotechnol 28:573–579
Lauring AS et al (2012) Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe 12:623–632
Lauring AS et al (2013) The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol 11:327–336
Liu YS et al (2010) Analysis of synonymous codon usage in porcine reproductive and respiratory syndrome virus. Infect Genet Evol 10:797–803
Liu YS et al (2011) The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. Infect Genet Evol 11:1168–1173
Liu XS et al (2012) Patterns and influencing factor of synonymous codon usage in porcine circovirus. Virol J 9:68
Liu X et al (2013) High-resolution view of bacteriophage lambda gene expression by ribosome profiling. Proc Natl Acad Sci USA 110:11928–11933
Lopez-Lastra M et al (2010) Translation initiation of viral mRNAs. Rev Med Virol 20:177–195
Lobo FP et al (2009) Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. PLoS One 4:e6282
Lucks JB et al (2008) Genome landscapes and bacteriophage codon usage. PLoS Comput Biol 4: e1000001
Marraffini LA (2015) CRISPR-Cas immunity in prokaryotes. Nature 526:55–61
Martrus G et al (2013) Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology 10:78
Ma MR et al (2011) The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern. Virol J 8:544
Medrano-Soto A et al (2004) Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. Mol Biol Evol 21:1884–1894
Michely S et al (2013) Evolution of codon usage in the smallest photosynthetic eukaryotes and their giant viruses. Genome Biol 5:848–859
Morgunov AS, Babu MM (2014) Optimizing membrane-protein biogenesis through nonoptimal-codon usage. Nat Struct Mol Biol 21:1023–1025. doi:10.1038/nsmb.2926
Mueller S et al (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol 80:9687–9696
Mueller S et al (2008) Live attenuated influenza virus vaccines by computer-aided rational design virus attenuation by genome-scale changes in codon pair bias. Nat Biotechnol 28:723–726
Nogales A et al (2014) Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. J Virol 88:10525–10540
Novella IS et al (2004) Positive selection of synonymous mutations in vesicular stomatitis virus. J Mol Biol 342:1415–1421
Novoa EM, Ribas de Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet 28:574–81. doi:10.1016/j.tig.2012.07.006. Epub 23 Aug 2012
Pan W et al (2016) Online model selection for synthetic gene networks. CDC 776–782
Pavesi A et al (2013) Viral proteins originated de novo by overprinting can be identified by codon usage: application to the “gene nursery” of Deltaretroviruses. PLoS Comput Biol 9:e1003162
Pechmann S, 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. Epub 23 Dec 2012
Pinto RM et al (2007) Codon usage and replicative strategies of hepatitis A virus. Virus Res 127:158–163
Pride DT et al (2006) Evidence of host-virus co-evolution in tetrancleotide usage patterns of bacteriophages and eukaryotic viruses. BMC Genomics 7:8
Quax TE et al (2015) Codon bias as a means to fine-tune gene expression. Mol Cell 59:149–161
Raveh A et al (2016) A model for competition for ribosomes in the cell. J R Soc Interface 13:20151062
Reuveni S et al (2011) Genome-scale analysis of translation elongation with a ribosome flow model. PLoS Comput Biol 1–18
Rocha EP (2004) Codon usage bias from tRNA’s point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res 14: 2279–2286. Epub 12 Oct 2004
Rohde W et al (1994) Plant viruses as model systems for the study of non-canonical translation mechanisms in higher plants. J Gen Virol 75(Pt 9):2141–2149
Roller M et al (2013) Environmental shaping of codon usage and functional adaptation across microbial communities. Nucleic Acid Res 41:8842–8852
Roychoudhury S, Mukherjee D (2010) A detailed comparative analysis on the overall codon usage pattern in herpesviruses. Virus Res 148:31–43
Sabi R et al (2016) stAIcalc: tRNA adaptation index calculator based on species-specific weights.
Bioinformatics
Sanassy D et al (2015) Meta-stochastic simulation of biochemical models for systems and synthetic biology. ACS Synth Biol 4:39–47
Sau K, Deb A (2009) Temperature influences synonymous codon and amino acid usage biases in the phages infecting extremely thermophilic prokaryotes. Silico Biol 9:1–9
Sau K et al (2005a) Factors influencing the synonymous codon and amino acid usage bias in AT-rich Pseudomonas aeruginosa phage PhiKZ. Acta Biochim Biophys Sin (Shanghai) 37:625–633
Sau K et al (2005b) Synonymous codon usage bias in 16 Staphylococcus aureus phages: implication in phage therapy. Virus Res 113:123–131
Sau K et al (2007) Studies on synonymous codon and amino acid usage biases in the broad-host range bacteriophage KVP40. J Microbiol 45:58–63
Sauna ZE, Kimchi-Sarfaty C (2013) Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 12:683–691
Schoech AP, Zabet NR (2014) Facilitated diffusion buffers noise in gene expression. Phys Rev E Stat Nonlin Soft Matter Phys 90:032701
Shackelton LA et al (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol 62:551–563
Sharp PM, Li WH (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acid Res 15:1281–1295
Sharp PM et al (2005) Variation in the strength of selected codon usage bias among bacteria. Nucleic Acid Res 33:1141–1153. Print 2005
Sin C et al (2016) Quantitative assessment of ribosome drop-off in E. coli. Nucleic Acid Res 44:2528–2537
Soucy SM et al (2015) Horizontal gene transfer: building the web of life. Nat Rev Genet 16:472–482
Su MW et al (2009) Categorizing host-dependent RNA viruses by principal component analysis of their codon usage preferences. J Comput Biol 16:1539–1547
Supek F (2016) The code of silence: widespread associations between synonymous codon biases and gene function. J Mol Evol 82:65–73
Tao P et al (2009) Analysis of synonymous codon usage in classical swine fever virus. Virus Genes 38:104–112
Terns MP, Terns RM (2011) CRISPR-based adaptive immune systems. Curr Opin Microbiol 14:321–327
Thommen M et al (2016) Co-translational protein folding: progress and methods. Curr Opin Struct Biol 42:83–89
Tsai CT et al (2007) Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus Res 126:196–206
Tuller T et al (2010a) Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci USA 107:3645–3650
Tuller T et al (2010b) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354
Tuller T (2011) Codon bias, tRNA pools, and horizontal gene transfer. Mob Genet Elem
Tuller T (2012) The effect of codon usage on the success of horizontal gene transfer. In: In lateral gene transfer in evolution
Tuller T et al (2011a) Composite effects of gene determinants on the translation speed and density of ribosomes. Genome Biol 12:R110
Tuller T et al (2011b) Association between translation efficiency and horizontal gene transfer within microbial communities. Nucleic Acid Res 22
Tuller T, Zur H (2015) Multiple roles of the coding sequence 5’ end in gene expression regulation. Nucleic Acid Res 43:13–28
Tulloch F et al (2014) RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife 3:e04531
van Hemert FJ et al (2007) Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. Virology 361:447–454
Watts JM et al (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711–716
Welch M et al (2009) Design parameters to control synthetic gene expression in Escherichia coli. PLoS ONE 4:1–10
Wimmer E et al (2009) Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat Biotechnol 27:1163–1172
Wong EH et al (2010) Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. BMC Evol Biol 10:253
Wright F (1990) The ‘effective number of codons’ used in a gene. Gene 87:23–29
Wu H et al (2016) Multiensemble Markov models of molecular thermodynamics and kinetics. Proc Natl Acad Sci USA 113:E3221–E3230
Yang JR et al (2014) Codon-by-codon modulation of translational speed and accuracy via mRNA folding. PLoS Biol 12:e1001910. doi: 10.1371/journal.pbio.1001910. eCollection Jul 2014
Yofe I et al (2014) An intronic code for gene expression regulation in S.cerevisiae. PLoS Genet 10: e1004407
Zafrir Z, Tuller T (2015a) Selection for nucleotide composition adjacent to intronic splice sites improves splicing efficiency via its effect on pre-mRNA local folding in fungi. RNA 21: 1704–1718
Zafrir Z, Tuller T (2015b) Nucleotide sequence composition adjacent to intronic splice sites improves splicing efficiency via its effect on pre-mRNA local folding in fungi. RNA 21: 1704–1718
Zafrir Z, Tuller T (2017) Unsupervised detection of regulatory gene expression information in different genomic regions enables gene expression ranking. BMC Bioinform 18:77
Zafrir Z et al (2016) Selection for reduced translation costs at the intronic 5’ end in fungi. DNA Res 23:377–394
Zhang Y et al (2011) Analysis of synonymous codon usage in hepatitis A virus. Virol J 8:174
Zhang Z et al (2013) Analysis of synonymous codon usage patterns in torque tenosus virus 1 (TTSuV1). Arch Virol 158:145–154
Zhao KN et al (2005) Gene codon composition determines differentiation-dependent expression of a viral capsid gene in keratinocytes in vitro and in vivo. Mol Cell Biol 25:8643–8655
Zhao S et al (2008) Analysis of synonymous codon usage in 11 human bocavirus isolates. Biosystems 92:207–214
Zhong J et al (2007) Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes 35:767–776
Zhou JH et al (2010) Analysis of synonymous codon usage in foot-and-mouth disease virus. Vet Res Commun 34:393–404
Zhou JH et al (2013) The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus. Infect Genet Evol 16:270–274
Zimmer C (2011) A Planet of Viruses. University Of Chicago Press, Chicago
Zur H, Tuller T (2012) Strong association between mRNA folding strength and protein abundance in S. cerevisiae. EMBO Rep
Zur H, Tuller T (2013) New universal rules of Eukaryotic translation initiation fidelity. PLoS Comput Biol 9:e1003136
Zur H, Tuller T (2015) Exploiting hidden information interleaved in the redundancy of the genetic code without prior knowledge. Bioinformatics 31:1161–1168
Zur H, Tuller T (2016) Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. Nucleic Acid Res 44:9031–9049