Importance of codon usage for the temporal regulation of viral gene expression

Young C. Shin^{a,1}, Georg F. Bischof^{a,b,1}, William A. Lauer^a, and Ronald C. Desrosiers^{a,2}

^aDepartment of Pathology, Miller School of Medicine, University of Miami, Miami, FL 33136; and ^bInstitute of Clinical and Molecular Virology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

Edited by Stephen P. Goff, Columbia University College of Physicians and Surgeons, New York, NY, and approved September 10, 2015 (received for review August 6, 2015)

The glycoproteins of herpesviruses and of HIV/SIV are made late in the replication cycle and are derived from transcripts that use an unusual codon usage that is quite different from that of the host cell. Here we show that the actions of natural transinducers from these two different families of persistent viruses (Rev of SIV and ORF57 of the rhesus monkey rhadinovirus) are dependent on the nature of the skewed codon usage. In fact, the transinducibility of expression of these glycoproteins by Rev and by ORF57 can be flipped simply by changing the nature of the codon usage. Even expression of a luciferase reporter could be made Rev dependent or ORF57 dependent by distinctive changes to its codon usage. Our findings point to a new general principle in which different families of persisting viruses use a poor codon usage that is skewed in a distinctive way to temporally regulate late expression of structural gene products.

codon usage | SIV Rev | RRV ORF57 | glycoprotein induction | temporal regulation

Some viruses are able to maintain lifelong infections in their natural host despite strong host immune surveillance measures. Such persisting viruses include members of the herpes, lenti-retro, polyoma, papilloma, adeno, and parvo families of viruses. It is perhaps no accident that these six families of persisting viruses strictly regulate their gene expression in a temporal fashion. Structural proteins are made late in the virus replication cycle and expression of these late gene products is typically induced by viral transinducers made earlier in the viral replication cycle. A variety of strategies are used by different persisting viruses to achieve this temporal regulation of structural gene expression; these include both transcriptional and posttranscriptional mechanisms. The γ herpesviruses encode a transcription factor called ORF50/RTA that is expressed immediately early and that can act on the promoters of a variety of herpesviral genes expressed during the lytic cycle of replication (1–3); the herpesviral-encoded ORF57 family of proteins act posttranscriptionally to facilitate expression of the late structural genes (4–7).

Billelo et al. noticed an uncanny set of similarities when comparing the expression of the viral-encoded envelope glycoproteins of lenti-retroviruses (the gp160 of the human immunodeficiency virus, HIV, and the simian immunodeficiency virus, SIV) and of the viral-encoded envelope glycoproteins of the herpesviruses (8). Glycoprotein H of the γ -2 herpesvirus of rhesus monkeys called RRV (rhesus monkey rhadinovirus) has been used as the example for some of these comparisons (8). Expression of both gp160 and RRV gH is temporally regulated such that each is made late in the lytic cycle of these persisting viruses. Both gp160 and gH have an unusual, skewed codon usage. If one puts a strong promoter (such as the cytomegalovirus (CMV) immediate early promoter/enhancer) in front of each reading frame and transfects the expression cassette into human embryonic kidney (HEK) 293T cells, essentially no protein expression can be detected. There are two ways in which protein expression can be rescued in these assays: by changing the codon usage to one more compatible with expression or by providing the natural viral-encoded transinducer *in trans*. For HIV and SIV, the natural transinducer is the early viral gene product called regulator of virion expression (Rev). For RRV, the natural

transinducer is the early viral gene product called ORF57. The levels of viral glycoprotein expression from a sequence whose codon usage has been ideally altered far exceed those from the natural coding sequence even when optimal levels of transinducer are provided (7, 8). These observations prompted us to investigate further the influence of codon usage on the regulated expression of viral genes.

Results

Induction of SIV gp160 and RRV gH Expression from Codon-Modified Sequences. Comparison of the codon usage of SIV gp160 with that of RRV gH revealed several codons that were overrepresented in one sequence but not the other, i.e., the nature of the unusual skewed codon usage was different (Fig. 1A and Table S1). Six codons were selected based on their bias of usage. To make SIV gp160 more gH-like in its codon usage, codons were changed in a way that reduced the usage of codons that have relatively high frequencies in SIV gp160 and/or increased the usage of codons that have relatively high frequencies in RRV gH. The opposite codon changes were performed to make RRV gH more gp160-like in its codon usage. More specifically, six codons (AGG, GAG, CCT, ACT, CTC, and GGG) in SIV gp160 were changed to CGT, GAA, CCG, ACG, TTA, and GGA, respectively. The number of altered codons was 93, which comprised 10.5% of the total codons in SIV gp160 (Table S1). The codon changes in SIV gp160 and RRV gH were distributed rather evenly throughout the entire coding region (Fig. 1B and Fig. S1A and B). The Rev response element (RRE) region in SIV gp160 was not codon modified; the 1,520–1,876 nucleotide stretch that contains the RRE (9, 10) was left intact. A splicing donor (SD) sequence was also present in front of the ATG start codon of each coding sequence because it is required for efficient Rev-mediated induction as has been previously

Significance

We describe here that different families of persisting viruses consistently use highly unusual codon usage for synthesis of their structural gene products; that the specific nature of the skewed codon usage differs fundamentally from one virus family to another; and that viral-encoded regulatory proteins (specifically Rev of HIV/SIV and ORF57 of herpesviruses) recognize the specific nature of the skewed codon usage to allow expression of their structural gene products. The data include a stunning demonstration of the ability to flip Rev inducibility and ORF57 inducibility simply by changing the nature of the codon usage in the target gene.

Author contributions: Y.C.S., G.F.B., and R.C.D. designed research; Y.C.S., G.F.B., and W.A.L. performed research; Y.C.S., G.F.B., W.A.L., and R.C.D. analyzed data; and Y.C.S. and R.C.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹Y.C.S. and G.F.B. contributed equally to this work.

²To whom correspondence should be addressed. Email: r.desrosiers@med.miami.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515387112/-DCSupplemental.

Fig. 1. Influence of codon usage on SIV gp160 and RRV gH induction. (A) Comparison of codon usage in SIV gp160 and RRV gH. The coding sequences of SIV239 gp160 and RRV gH were analyzed using the GCUA software (gcu.schoedl.de/) (40). The use of each indicated codon in RRV gH as percent of all codons for that particular amino acid is plotted vs. that for gp160 using a XY scatter plot with a linear regression line and a coefficient of determination (R^2). Codons with the most skewed usage between the two sequences are those most extensively deviated from $y = x$ line. To generate a codon-modified (gH-like) SIV gp160, some codons of SIV gp160 (empty colored circles) were changed into corresponding synonymous codons (filled colored circles) to reflect the codon usage of RRV gH. To generate a codon-modified (gp160-like) RRV gH, codon changes were performed in the exact opposite direction. (B) Schematic representation of expression cassettes with the codon-modified (gH-like) SIV gp160 and codon-modified (gp160-like) RRV gH sequence. Splicing donor (SD) sequence (indicated by gray arrow, aaacaagtaagt), and Rev-response element (RRE, red colored box) are shown. Each codon change is represented by a vertical line in the gp160 and gH box. Note that there were no codon changes in the RRE region. (C) Expression cassettes of SD-gp160 with the unmodified (U) (lanes 1, 2, and 3) or codon-modified (M, gH-like) SIV gp160 sequence (lanes 4, 5, and 6), and gp160 (no SD) with the unmodified (U) (lanes 7, 8, and 9) or codon-modified (M, gp160-like) SIV gp160 sequence (lanes 10, 11, and 12) were transfected into HEK 293T cells together with either control, Rev, or ORF57 expression plasmid. The cells were harvested at 46 h after transfection and proteins were subjected to SDS/PAGE and immunoblotting. The expression of gp160 was detected using the 3.11H antibody. The expression of ORF57 and Rev was detected using a v5 antibody. (D) The total levels of gp160 RNA (copies per microgram) were measured by qPCR. (E) Expression cassettes of gH with the unmodified (U) or codon-modified (M, gp160-like) RRV gH sequence with or without flanking SD and RRE sequences was transfected into HEK 293T cells together with either control, Rev, or ORF57 expression plasmid. The cells were harvested at 46 h after transfection and proteins were subjected to SDS/PAGE and immunoblotting. The expressions of RRV gH, ORF57, and Rev were detected using a v5 antibody. The two bands for RRV gH appear to correspond to two bands previously noted for the gH of the closely related Kaposi sarcoma herpesvirus (46). (F) The total levels of gH RNA (copies per microgram) were measured by qPCR.

noted (11, 12). The frequencies of A, G, C, and T in the codon-modified gp160 construct differed only very modestly from the parental reading frame from which it was derived (Table S2).

Expression of SIV gp160 from the expression cassette with natural codon usage and the strong CMV immediate early promoter/enhancer was not detectable in the absence of any inducer (Fig. 1C, lane 1), but was readily detectable in the presence of its natural inducer, Rev (Fig. 1C, lane 2), as has been repeatedly observed in many previous studies (12, 13). However, little if any induction of SIV gp160 expression from this cassette was observed in the presence of ORF57 (Fig. 1C, lane 3), indicating a specific induction of SIV gp160 by Rev but not by ORF57. The codon-modified version of the SIV gp160 expression cassette showed completely different results. The codon-modified (gH-like) SIV gp160 also showed little or no expression in the absence of trans-inducer (Fig. 1C, lane 4). Surprisingly, the codon-modified gp160 cassette yielded very little expression in the presence of the natural inducer, Rev, even though it still contained the *cis*-acting sequences RRE and SD that are known to be required for Rev-mediated induction (Fig. 1C, lane 5). However, expression of this cassette was now very strongly induced by the heterologous inducer, ORF57 (Fig. 1C, lane 6). In the absence of SD, induction of unmodified gp160 by Rev was lost (Fig. 1C, lane 8) while that of codon-modified (gH-like) gp160 by ORF57 was not affected (Fig. 1C, lane 12). Thus, we were amazingly able to switch the induction pattern of SIV gp160 simply by changing the codon use.

We next sought to do similar experiments by modifying the RRV gH coding sequence to make it more resemble the codon use of SIV gp160 (Fig. 1E). Again, six codons were changed but this time in the exact opposite direction of what was used for the codon changes performed with SIV gp160. CGT, GAA, CCG, ACG, TTA, and GGA in RRV gH were changed to AGG, GAG, CCT, ACT, CTC, and GGG, respectively. The number of altered codons in RRV gH was 104, amounting to 14.3% of its total codons (Table S1). Again, the frequencies of A, G, C, and T in the codon-modified construct differed only very modestly from the parental gH reading frame from which it was derived (Table S2). The RRV gH expression cassette with the natural codon usage was inducible by adding its homologous inducer, ORF57, but not by Rev (Fig. 1D, lane 2 and 3). However, as observed with codon-modified gp160 above, we could completely switch the induction pattern of RRV gH simply by changing its codon usage. The codon-modified (gp160-like) version of gH was inducible by Rev but not by ORF57 (Fig. 1E, lane 5 and 6). No Rev-mediated induction was observed in the absence of SD and RRE (Fig. 1E, lane 11), showing the necessity of these *cis*-acting elements for Rev activity in addition to the nature of the codon usage. Inductions by ORF57 (the natural gH sequence and the codon-modified gp160 sequence) occurred whether or not the splice donor was present (Fig. 1E, lane 3 and 9).

In parallel with the protein expression measurements, we also measured the corresponding RNA level in each condition by quantitative real-time PCR (qPCR) using a gp160- or gH-specific probe set. Overall, the RNA levels generally reflected the levels

of protein expression (Fig. 1 *D* and *F*). Compared with the control, the level of unmodified gp160 RNA was increased by approximately sixfold in the presence of Rev (Fig. 1*D*, lane 2) and that of codon-modified (gH-like) gp160 RNA was increased by approximately threefold in the presence of ORF57 (Fig. 1*D*, lane 6). In the absence of SD, the enhancement of RNA level in the former but not the latter was disrupted (Fig. 1*D*, lane 8 and 12).

To examine the influence of codon usage in a completely different system other than RRV gH and SIV gp160, we changed the coding sequence of firefly luciferase to make it either SIV gp160-like or RRV gH-like in its codon usage. The resulting codon-modified luciferase sequences showed very skewed patterns of codon usage compared with the parental luciferase sequence (Fig. S2 *A* and *B*). Both codon-modified cassettes showed drastic reductions in the basal level of firefly luciferase expression as expected: parental (862) vs. gp160-like (1) vs. gH-like (14) in relative light units (RLU). Expression of firefly luciferase from the expression cassette with gp160-like codon usage was efficiently induced by Rev but much less so by ORF57, whereas firefly luciferase expression from the cassette with gH-like codon usage was preferentially induced by ORF57 but much less so by Rev (Fig. 2*A*). When there were no SD and RRE, Rev-mediated induction of gp160-like luciferase disappeared, whereas ORF57-mediated induction of gH-like luciferase did not (Fig. 2*B*), consistent with the gp160 and gH induction results. Again, the luciferase RNA levels reflected to a reasonable extent the luciferase activities in each sample (Fig. 2, *Lower*).

Our creation of a Rev-inducible luciferase reporter appears analogous to the creation of a Rev-inducible GFP reporter by Graf et al. (14). However, in contrast to Graf et al. (14) who used massive G→A and C→A substitutions and an increase in A content from 24.2% to 45.7%, our Rev-inducible luciferase reporter is only marginally different in AGCT content compared with the parental sequence from which it was derived (Table S2).

Contributory Sequences Are Distributed. We next sought to address whether the sequence changes that impart ORF57 inducibility and

loss of Rev inducibility to gp160 could be localized to discrete segments of the gp160 reading frame (Fig. 3 and Fig. S3). Constructs with the Arg + Leu changes alone, the Glu + Gly changes alone, or the Pro + Thr changes alone were able to impart only a low level of ORF57 inducibility, far short of the full effect that was seen with the full set of six codons changed. Also, they retained most of Rev-mediated inducibility. We next exchanged all six codons in specific, confined regions of the gp160 sequence. Codon changes in the N-terminal 1/3 region had more significant impact than those in the middle or C-terminal 1/3 region. Induction of gp160 by Rev was more diminished in the 1/3N than in the 1/3M or 1/3C construct. Induction of gp160 by ORF57 was quite prominent in the 1/3N but not in the 1/3M or 1/3C construct. Codon changes in only a small portion (1/12N) of the gp160 sequence caused an impressive loss of Rev-mediated induction but only a slight induction by ORF57. Impressive induction of codon-modified gp160 by ORF57 required codon changes in 1/4 or more at the N terminus of gp160, and the degree of induction was proportional to the length of gp160 N-terminal sequence that had codon changes. Based on these results, we conclude that the N-terminal region of gp160 is critical to impart the Rev- or ORF57-mediated induction in a codon usage-dependent manner. The data on which this mapping summary (Fig. 3) are based can be found in Fig. S3.

Elicitation of Anti-env Antibodies by Recombinant RRV with Codon-Modified gp160 Sequences. In previous studies with recombinant RRV-SIVenv that used the CMV immediate early promoter and a codon-altered version of gp160 gene sequences that had been optimized for expression in an unregulated fashion, anti-gp160 antibodies were not elicited upon monkey infection (15). We thus constructed a recombinant RRV using the codon-modified version of gp160 sequences whose expression should now be temporally regulated as a late gene product. The CMV immediate early promoter was again used for this new recombinant RRV. When early passage rhesus fibroblasts were infected with recombinant RRV at low multiplicity of infection, RRV-SIVenv with

Fig. 2. Induction of luciferase from the expression cassettes with unmodified and codon-modified luciferase sequences. The coding sequence of parental firefly luciferase was modified to make it gp160-like or gH-like. In addition to the six codon changes shown in Table S1, the following additional synonymous codon changes were made. For gp160-like, GCC→GCA, CGC→AGG, CGG→AGG, AAC→AAT, TGC→TGT, ATC→ATA, CTG→TTG, CCG→CCA, TAC→TAT, and GTT→GTA. For gH-like, GCC→GCA, GAC→GAT, TGC→TGT, ATC→ATA, CTC→TTG, CTG→TTA, AAG→AAA, TTC→TTT, TAC→TAT, and GTC→GTA. Each luciferase sequence was flanked by splicing donor (SD) sequence and the Rev-responsive element (RRE) in *A*. No SD and RRE were added in *B*. The levels of firefly luciferase expression from each expression cassette in the presence of empty vector, ORF57, or Rev were measured after transfection of corresponding plasmid DNAs into HEK 293T cells. The cells were harvested at 46 h after transfection. The levels of renilla luciferase expression from a pRL-SV plasmid (Promega) were used to calibrate the transfection efficiency among the samples. Shown are representative results for three independent experiments done in triplicate. Error bars indicate SD. RLU, relative light units. The calibrated level of firefly luciferase activity from the transfection of gp160-like luciferase sequence (with SD and RRE) in the presence of empty vector was set as 1 RLU. The RLU value of each transfection was indicated above the corresponding bars. Shown at the bottom are the corresponding luciferase RNA levels measured by qPCR.

SIV239 gp160 constructs			w/ rev	w/ orf57
U		RRE	+++++	-
M		RRE	+	++++
R/L only		RRE	+++++	+
E/G only		RRE	+++++	+
P/T only		RRE	+++++	+
1/3C		RRE	+++	-
2/3C		RRE	++	-
1/3M		RRE	+++++	-
1/3N		RRE	++	+++
2/3N		RRE	++	++++
1/12N		RRE	++	+
1/6N		RRE	++	+
1/4N		RRE	++	++

wt sequence (U)
 gH-like sequence (M)

Fig. 3. Summary of induction of gp160 by Rev and ORF57 using altered coding sequences. HEK 293T cells in each well of a six-well plate were transfected with 1.5 μ g plasmid DNA of the gp160 constructs and 0.5 μ g each inducer (Rev or ORF57) and cultured for 42–48 h before harvest. Protein input was normalized using a Bradford assay. The level of SIV239 gp160 in the harvested cell pellets was detected by immunoblotting. Very little or no expression is denoted with a minus (–) sign, and the relative degree of induction of each construct is represented with varying numbers of a plus (+) sign.

gH-like codon usage displayed replication kinetics and peak levels of virus production that were indistinguishable from those of RRV-SIVenv with optimized codon usage for the env cassette (Fig. S44). However, levels of SIVenv expression in the infected cells appeared to peak later with the RRV-SIVenv with gH-like codon usage, consistent with temporally regulated expression (Fig. S4B). Upon experimental infection of rhesus monkeys, anti-gp160 antibody responses were now readily measured (Fig. 44). These antibodies have persisted for more than 1 y and were capable of neutralizing SIV316 (Fig. 4A and B) in vitro. It should be noted that anti-gp160 antibody responses were also not achieved with a recombinant CMV that used a codon-optimized version of gp160 sequences (16, 17).

Codon Usage for Structural Gene Products of Additional Virus Families.

We analyzed codon usage for structural gene products of representative members of six families of classically persisting viruses. All have temporal regulation of viral gene expression such that structural gene products are made late in the replication cycle. These six are RRV of the γ -2 subgroup of herpesviruses; HIV/SIV of the lentivirus subgroup of retroviruses; HPV16 of the papillomavirus family; SV40 of the polyomavirus family; human adenovirus 2 of the adenovirus family; and B19 in the parvovirus family. Compared with codon usage of the human exome, all six structural gene products exhibited marked skewing of codon usage, with R^2 values ranging from 0.25 to 0.50 (average, 0.35; Fig. 5A). We then compared these values to values for structural gene products of representative members of five families of acute, nonpersisting viruses. These nonpersisting virus families were selected on the basis of not exhibiting temporal regulation of viral gene expression. These five are poliovirus in the picornavirus family; measles virus in the paramyxovirus family; Ebola virus in the filovirus family; rabies virus in the rhabdovirus family; and yellow fever virus in the flavivirus family. Codon usage for the structural gene products of these five nonpersisting viruses was much more aligned with the codon usage in the human exome (Fig. 5B). R^2 values ranged from 0.67 to 0.87 (average, 0.74). These differences are statistically significant ($P < 0.0001$ by an unpaired t test). Some of the other families of acute, nonpersisting

viruses exhibit temporal regulation of viral gene expression, and at least some of these, for example, coronaviruses and arenaviruses, possess a skewed codon usage for their structural gene products.

Finally, we asked whether the expression limitations imposed by the skewed codon usage for RRV and SIV reported here could be extended to other persisting virus families. Expression of SV40 VP1 and rhesus papillomavirus L1 from an expression cassette with natural codon usage was below the limit of detection, whereas codon optimization of VP1 and L1 coding sequence led to readily detectable levels of expression (Fig. S5).

Discussion

Our data show that Rev induction of Env protein expression depends on the nature of the skewed codon usage. Rev inducibility was lost when 10.5% of the codons were replaced with synonymous codons reflecting the codon usage of RRV gH. While it had been previously known that Rev induction depended on the Rev response element (the RRE) and the presence of a splice donor upstream of the coding sequence (11, 18), the dependence on the nature of the skewed codon usage had not been previously known, although publications by Graf et al. (14) and Shuck-Lee et al. (19) are consistent with it. Similarly, our data show that ORF57 induction of RRV gH protein expression also depends on the nature of the skewed codon use. ORF57 inducibility was lost when 14.3% of the codons were replaced with synonymous codons, reflecting the skewed codon usage of SIV gp160. Again, this dependence of ORF57-inducing activity on the nature of the skewed codon usage had not been previously known. Remarkably, Rev dependence of gp160 expression can be flipped to ORF57 dependence simply by changing the nature of the codon usage; similarly, ORF57 dependence of RRV gH expression can be flipped to Rev dependence simply by changing the nature of the codon usage (Fig. 1). Furthermore, it was possible to make firefly luciferase expression heavily dependent on ORF57 or Rev by imparting distinctive, synonymous changes to codon usage (Fig. 2). It is important to note that no complicated algorithms have been used, at least to date, to select the codon biases that have resulted in this remarkable flip in dependencies.

Although our data unambiguously make the case for the importance of the nature of aberrantly skewed codon usage for the specificity of transinduction across these two families of persisting viruses, it should by no means be assumed that this is the only mechanism for regulating late structural gene expression or that the specific mechanisms of codon usage dependence is necessarily the same across multiple families. Although the early literature emphasized the effects of Rev in regulating the egress of unspliced and minimally spliced RNAs out of the nucleus (18, 20), a large body of literature has also emphasized the effects of Rev on expression of RNA in the cytoplasm (21–24) and on other regulatory effects (23,

Fig. 4. Elicitation of anti-env antibodies by recombinant RRV with codon-modified gp160 sequence. (A) Recombinant RRV (rRRV) harboring a codon-modified (gH-like) gp160 sequence was inoculated into two rhesus macaques (1×10^9 genome copies per animal), and levels of anti-SIV envelope antibodies were measured by an ELISA. (B) Neutralization of SIV316 by sera taken at weeks 17 and 33 after inoculation with rRRV with codon-modified gp160 sequence. SPF, negative control sera from specific pathogen-free animals; RLU, relative light units.

Fig. 5. Codon usage of persistent and nonpersistent viruses. (A) The coding sequences of representative late gene products from six persistent viruses were analyzed using the GCUA software (gcu.schoedl.de/) (40). The usage (%) of each codon in RRV gH, SIV239 gp160, human papilloma virus 16 (HPV16) L1, simian virus 40 (SV40) VP1, human adenovirus 2 (hAV2) fiber, and parvovirus B19 VP1 was plotted on the *x* axis vs. the codon usage of human exome sequences on the *y* axis. Also shown is the coefficient of determination (R^2). (B) The coding sequences of representative structural gene products from five acute, nonpersistent viruses were analyzed as in A. Listed are poliovirus P1, measles virus HA, Ebola virus gp, rabies virus gp, and yellow fever virus glycoprotein sequences. The GeneBank accession number for each sequence is as follows: SIV gp160 (HM800148), RRV gH (AF210726), SV40 VP1 (EF579804), HPV16 L1 (K02718), hAV2 fiber (AY224420), parvovirus B19 vp1 (AY768535), poliovirus P1 (V01149), measles virus HA (AY445032), Ebola virus gp (L11365), rabies virus gp (S72465), and yellow fever virus glycoprotein (KM388818).

25–27). The literature is similarly diverse for the posttranscriptional effects of ORF57 of γ -2 herpesviruses with respect to RNA egress, RNA stability, and translation (7, 28–31). It is important to note that the efficiency of translation can have dramatic effects on RNA stability and consequently the levels of RNA and that such effects on stability will vary markedly with the individual RNA (32).

An unusual, skewed codon usage for structural gene products is not confined to the lenti-retro and herpes groups of persisting viruses. Structural gene products of the polyoma, papilloma, adeno- and parvo families of persisting viruses also have unusual, skewed codon usage patterns compared with that of the human exome. All six have early $\rightarrow$ late transitions for expression of structural gene products. We put a strong promoter in front of four of these, and this yielded essentially no detectable protein expression under transfection conditions that yielded very strong levels of expression of their codon-optimized versions (Fig. S5). These results alone provide suggestive evidence that multiple families of persisting viruses use abnormal codon usage to achieve regulated expression of their late structural gene products. Five families of nonpersisting viruses lacking temporal regulation of viral gene expression do not have such a skewed pattern of codon usage for their structural genes; structural genes with natural codon usage from them can

be expressed just fine without any codon changes, without any transinducer. It needs to be noted, however, that some families of nonpersisting viruses also use temporal regulation of viral gene expression, and at least some of these exhibit skewed codon usage for their structural gene products.

What exactly is being recognized to achieve the codon usage-dependent transinduction is a fascinating question. Synonymous codon changes can sometimes lead to unexpected outcomes by changing the structure of mRNA (33, 34) or protein itself (35, 36). Using the secondary structure prediction software RNAfold, the overall structure of RNA was predicted to be changed significantly by the synonymous codon changes, whereas the RRE region was not (Fig. S6). Perhaps even more attractive is the possible contribution of negative regulatory sequences, so-called inhibitory and instability elements, to the poor expression of the natural coding sequences for gp160 and gH (37–39). By this scenario, Rev and ORF57 would overcome the negative effects of distinct inhibitory/instability elements that are imparted by the codon usage of their target RNAs. In other words, codon usage per se may not be the critical factor but rather, distinct negative regulatory elements imparted by the nature of the skewed codon usage. With this scenario, the characteristic codon changes that we used would be postulated to have imparted distinctive inhibitory/instability elements that are distinctly overcome by Rev or by ORF57.

What advantage might temporal regulation of expression of structural proteins provide to persisting viruses? Perhaps in the face of persisting immune responses, highly restrictive temporal regulation of the expression of structural proteins may allow release of at least some progeny virions before immune-mediated destruction. It is fascinating that at least some persisting viruses have independently evolved a skewed, inhibitory codon usage as a means for achieving this temporal regulation.

Our observations on the influence of codon usage have practical utility for the use of persisting viruses as vaccine vectors or vectors for gene therapy. The altered codon usage in our recombinant RRV-SIVenv has allowed us for the first time, to our knowledge, to achieve anti-gp160 antibody responses with this herpesvirus vector system.

Materials and Methods

Codon Analysis, Modification, and RNA Secondary Structure. Codon usage was analyzed using the graphical codon usage analyzer (GCUA) software (gcua.schoedl.de/) (40) and Gribkov codon preference plot (41). Sequence alignment was performed using a MacVector program. Codon adaptation index (CAI) was calculated using JCat (www.jcat.de) (42). Codon-modified SIVgp160, RRV gH, firefly luciferase, and codon-optimized SV40 VP1 sequences were synthesized in vitro (GenScript). Prediction of RNA secondary structure was obtained using a RNAfold software (rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (43).

Plasmids. The cDNA sequences encoding unmodified, codon-modified (gH-like), and expression-optimized SIV gp160 (obtained from George N. Pavlakis, National Cancer Institute, Bethesda, MD) were cloned in pcDNA6-v5-hisA plasmid vector (Life Technologies) with the splicing donor sequence (aaacaagaatg) in front of the translational start codon of gp160 coding sequence. The cDNA sequences expressing unmodified and codon-modified (gp160-like) RRV gH were cloned in pcDNA6-v5-hisA plasmid vector with or without the splicing donor sequence and the RRE as indicated in Fig. 1B. The cDNA sequences expressing unmodified and codon-modified (gp160-like and gH-like) firefly luciferase were cloned in pcDNA6-v5-hisA plasmid vector with the splicing donor sequence in front of the translational start codon and the RRE after the stop codon. The RRV ORF57 and HIV-1 Rev cDNA sequences (from pCMV-HIV-1, National Institutes of Health AIDS reagent program) were cloned in pEF1-v5-hisA (Life Technologies) and pcDNA6-v5-hisA plasmid vector, respectively. The WT and codon-optimized sequence of SIV gp160, RRV gH, rHPV L1, and SV40 VP1 were cloned in pcDNA6-v5-hisA plasmid.

Transfection. HEK293T cells that had been seeded into six-well plates 1 d earlier were transfected, unless otherwise indicated, with 1.5 μ g of gp160 or gH, 0.5 μ g of Rev, or ORF57 DNAs using a jetPRIME transfection reagent (Polyplus transfection) and cultured for 42–48 h before harvest.

Antibodies and Reagents. The antibodies and reagents used in this study were obtained from the following sources: mouse and rabbit anti-v5 (Life Technologies) and mouse and rabbit anti-FLAG (Sigma). Anti-gp120 antibody (3.11H) was purified from hybridoma cell line acquired from J. E. Robinson, Tulane University, New Orleans.

Immunoblotting and SIVgp120 ELISA. Immunoblotting and SIVgp120 ELISA were performed as described previously (7, 15).

qPCR. For quantitative real-time PCR analysis, HEK293T cells in six-well plates were transfected with plasmids expressing unmodified or codon-modified sequences of interest. For the coexpression of inducers, empty vector (pEF1-FLAG), Rev (pCMV-HIV-1), or ORF57 (pIRES-ORF57-FLAG) was used. Total RNA from cells was purified using an RNeasy plus miniprep kit (Qiagen) according to the manufacturer's instructions and quantified by measuring absorbance at 260 nm (NanoDrop). qPCR was performed using gp160-N probe set (forward; GGTGATTATTCAGAAAGTGGCCCTTA, reverse; TCTATTGCCTGTTCTGTGACTGT, reporter; FAM-CCAGGCATCAAAGCTT-NFQ), gH-N probe set (forward; GGTA-ATTACATTGTCTGTAATT, reverse; TCCACAGTTGATATATCGCTT, reporter; FAM-ATAGTTCTCTTACCCTGCC-NFQ), or Luc-N probe set (forward; TGCTTTTACAGATGCACATAT, reverse; GGCTGCGAAATGCCACT, reporter; FAM-GCTAT-GAAACGATATGG-NFQ). Each purified RNA sample was mixed with 4x Taqman Fast Virus 1-step Master mix (Life Technologies), and the real-time signal

was detected by AB 7500 real-time PCR system. The amplification process consisted of reverse transcription (50 °C, 10 min) and inactivation (95 °C, 20 s) followed by 40 cycles of amplification (95 °C, 3 s; 60 °C, 30 s). The number of RNA copies per microgram of total purified RNA was calculated using the corresponding plasmid DNA as standards.

Generation of Recombinant RRV. Recombinant RRV expressing SIV env from the codon-modified (gH-like) gp160 sequence was made by overlapping cosmid transfection as described earlier (44), and the virus titers in the culture supernatants of rhesus fibroblast cells were measured by qPCR using a primer set specific to the sequence of latency-associated nuclear antigen (LANA) of RRV (Forward; ACCGCTGTGCGTGTTA, reverse; CAATCGCCAACGCTCAA, reporter; FAM-CAGGCCCCATCCCC).

Neutralization Assay. Neutralization assays using TZM-bl cells were performed using monkey sera taken at weeks 17 and 33 after inoculation with rRRV according to methods described previously (45).

ACKNOWLEDGMENTS. We thank L. Zhou for her help with Fig. 4B; J. Jung, M. Farzan, D. Evans, D. Watkins, M. Stevenson, A. Hahn, J. Martinez-Navio, S. P. Fuchs, and J. Termini for discussion; and S. Pantry and S. Pedreño for reading the manuscript. This work was supported by National Institutes of Health Grant 063928 (to R.C.D.).

- Damania B, et al. (2004) Comparison of the Rta/Orf50 transactivator proteins of gamma-2-herpesviruses. *J Virol* 78(10):5491–5499.
- Zaloni S, Holley-Guthrie E, Kenney S (1996) Epstein-Barr viral latency is disrupted by the immediate-early BRLF1 protein through a cell-specific mechanism. *Proc Natl Acad Sci USA* 93(17):9194–9199.
- Sun R, et al. (1998) A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. *Proc Natl Acad Sci USA* 95(18):10866–10871.
- Kirshner JR, Lukac DM, Chang J, Ganem D (2000) Kaposi's sarcoma-associated herpesvirus open reading frame 57 encodes a posttranscriptional regulator with multiple distinct activities. *J Virol* 74(8):3586–3597.
- Sandri-Goldin RM, Mendoza GE (1992) A herpesvirus regulatory protein appears to act post-transcriptionally by affecting mRNA processing. *Genes Dev* 6(5):848–863.
- Ruvolo V, Wang E, Boyle S, Swaminathan S (1998) The Epstein-Barr virus nuclear protein SM is both a post-transcriptional inhibitor and activator of gene expression. *Proc Natl Acad Sci USA* 95(15):8852–8857.
- Shin YC, Desrosiers RC (2011) Rhesus monkey rhadinovirus ORF57 induces gH and gL glycoprotein expression through posttranscriptional accumulation of target mRNAs. *J Virol* 85(15):7810–7817.
- Bilello JP, Morgan JS, Desrosiers RC (2008) Extreme dependence of gH and gL expression on ORF57 and association with highly unusual codon usage in rhesus monkey rhadinovirus. *J Virol* 82(14):7231–7237.
- Cochrane AW, Chen CH, Rosen CA (1990) Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA. *Proc Natl Acad Sci USA* 87(3):1198–1202.
- Malim MH, et al. (1990) HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* 60(4):675–683.
- Chang DD, Sharp PA (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell* 59(5):789–795.
- Lu XB, Heimer J, Rekosh D, Hammarikjöld ML (1990) U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced. *Proc Natl Acad Sci USA* 87(19):7598–7602.
- Hammarikjöld ML, et al. (1989) Regulation of human immunodeficiency virus env expression by the rev gene product. *J Virol* 63(5):1959–1966.
- Graf M, Ludwig C, Kehlenbeck S, Jungert K, Wagner R (2006) A quasi-lentiviral green fluorescent protein reporter exhibits nuclear export features of late human immunodeficiency virus type 1 transcripts. *Virology* 352(2):295–305.
- Bilello JP, et al. (2011) Vaccine protection against simian immunodeficiency virus in monkeys using recombinant gamma-2 herpesvirus. *J Virol* 85(23):12708–12720.
- Hansen SG, et al. (2009) Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. *Nat Med* 15(3):293–299.
- Hansen SG, et al. (2011) Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. *Nature* 473(7348):523–527.
- Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 338(6212):254–257.
- Shuck-Lee D, Chang H, Sloan EA, Hammarikjöld ML, Rekosh D (2011) Single-nucleotide changes in the HIV Rev-response element mediate resistance to compounds that inhibit Rev function. *J Virol* 85(8):3940–3949.
- Felber BK, Hadzopoulou-Cladaras M, Cladaras C, Copeland T, Pavlakis GN (1989) rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci USA* 86(5):1495–1499.
- D'Agostino DM, Felber BK, Harrison JE, Pavlakis GN (1992) The Rev protein of human immunodeficiency virus type 1 promotes polyomal association and translation of gag/pol and vpu/env mRNAs. *Mol Cell Biol* 12(3):1375–1386.
- Perales C, Carrasco L, González ME (2005) Regulation of HIV-1 env mRNA translation by Rev protein. *Biochim Biophys Acta* 1743(1–2):169–175.
- Blissenbach M, Grewe B, Hoffmann B, Brandt S, Ueberl K (2010) Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *J Virol* 84(13):6598–6604.
- Arrigo SJ, Chen IS (1991) Rev is necessary for translation but not cytoplasmic accumulation of HIV-1 vif, vpr, and env/vpu 2 RNAs. *Genes Dev* 5(5):808–819.
- Brandt S, et al. (2007) Rev proteins of human and simian immunodeficiency virus enhance RNA encapsidation. *PLoS Pathog* 3(4):e54.
- Groom HC, Anderson EC, Lever AM (2009) Rev: Beyond nuclear export. *J Gen Virol* 90(Pt 6):1303–1318.
- Jeang KT (2012) Multi-faceted post-transcriptional functions of HIV-1 Rev. *Biology (Basel)* 1(2):165–174.
- Nekorchuk M, Han Z, Hsieh TT, Swaminathan S (2007) Kaposi's sarcoma-associated herpesvirus ORF57 protein enhances mRNA accumulation independently of effects on nuclear RNA export. *J Virol* 81(18):9990–9998.
- Pilkington GR, et al. (2012) Kaposi's sarcoma-associated herpesvirus ORF57 is not a bona fide export factor. *J Virol* 86(23):13089–13094.
- Malik P, Blackburn DJ, Clements JB (2004) The evolutionarily conserved Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with REF protein and acts as an RNA export factor. *J Biol Chem* 279(31):33001–33011.
- Williams BJ, et al. (2005) The prototype gamma-2 herpesvirus nucleocytoplasmic shuttling protein, ORF 57, transports viral RNA through the cellular mRNA export pathway. *Biochem J* 387(Pt 2):295–308.
- Presnyak V, et al. (2015) Codon optimality is a major determinant of mRNA stability. *Cell* 160(6):1111–1124.
- Nackley AG, et al. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314(5807):1930–1933.
- Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324(5924):255–258.
- Kimchi-Sarfaty C, et al. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315(5811):525–528.
- Zhou M, et al. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495(7439):111–115.
- Schneider R, Campbell M, Nasioulas G, Felber BK, Pavlakis GN (1997) Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. *J Virol* 71(7):4892–4903.
- Barreau C, Paillard L, Osborne HB (2005) AU-rich elements and associated factors: Are there unifying principles? *Nucleic Acids Res* 33(22):7138–7150.
- Megati S, et al. (2008) Modifying the HIV-1 env gp160 gene to improve pDNA vaccine-elicited cell-mediated immune responses. *Vaccine* 26(40):5083–5094.
- Fuhrmann M, et al. (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol Biol* 55(6):869–881.
- Gribkov M, Devereux J, Burgess RR (1984) The codon preference plot: Graphic analysis of protein coding sequences and prediction of gene expression. *Nucleic Acids Res* 12(1 Pt 2):539–549.
- Grote A, et al. (2005) JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33(Web Server issue):W526–W531.
- Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31(13):3429–3431.
- Bilello JP, Morgan JS, Damania B, Lang SM, Desrosiers RC (2006) A genetic system for rhesus monkey rhadinovirus: Use of recombinant virus to quantitate antibody-mediated neutralization. *J Virol* 80(3):1549–1562.
- Montefiori DC (2009) Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol Biol* 485:395–405.
- Naranatt PP, Akula SM, Chandran B (2002) Characterization of gamma2-human herpesvirus-8 glycoproteins gH and gL. *Arch Virol* 147(7):1349–1370.
- Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9(1):133–148.