Codon usage limitation in the expression of HIV-1 envelope glycoprotein

Jürgen Haas*‡, Eun-Chung Park† and Brian Seed*†

Background: The expression of both the *env* and *gag* gene products of human immunodeficiency virus type 1 (HIV-1) is known to be limited by *cis* elements in the viral RNA that impede egress from the nucleus and reduce the efficiency of translation. Identifying these elements has proven difficult, as they appear to be disseminated throughout the viral genome.

Results: Here, we report that selective codon usage appears to account for a substantial fraction of the inefficiency of viral protein synthesis, independent of any effect on improved nuclear export. The codon usage effect is not specific to transcripts of HIV-1 origin. Re-engineering the coding sequence of a model protein (Thy-1) with the most prevalent HIV-1 codons significantly impairs Thy-1 expression, whereas altering the coding sequence of the jellyfish green fluorescent protein gene to conform to the favored codons of highly expressed human proteins results in a substantial increase in expression efficiency.

Conclusions: Codon-usage effects are a major impediment to the efficient expression of HIV-1 genes. Although mammalian genes do not show as profound a bias as do *Escherichia coli* genes, other proteins that are poorly expressed in mammalian cells can benefit from codon re-engineering.

Addresses: *Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. †Melvin and Barbara Nessel Gene Therapy Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

‡Present address: Institut für Medizische Virologie, Universität Heidelberg, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany.

Correspondence to: Brian Seed
E-mail address: seed@molbio.mgh.harvard.edu

Received: 21 December 1995
Revised: 31 January 1996
Accepted: 31 January 1996

Current Biology 1996, Vol 6 No 3:315–324

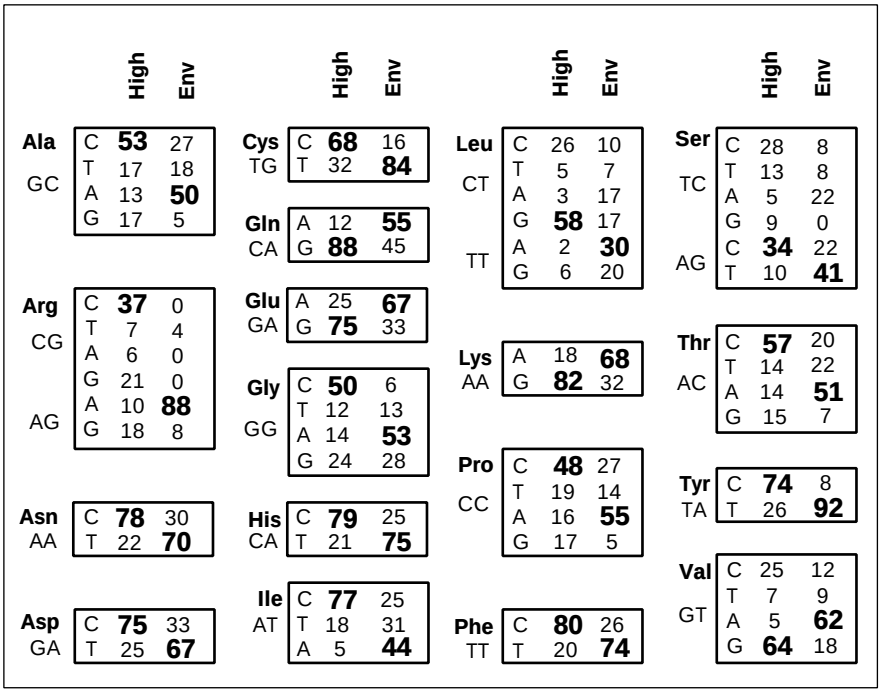
Background

The mature envelope glycoproteins of human immunodeficiency virus type 1 (HIV-1), gp120 and gp41, form a heterodimeric membrane complex embedded in the lipid bilayer that circumscribes the viral inner capsid. Expression of envelope proteins, like those of the gag polyprotein and reverse transcriptase, is facilitated by a *cis*-acting element in the viral mRNA known as the rev-responsive element or RRE [1–7], and by the action of rev, a small RNA-binding polypeptide encoded by a subgenomic RNA overlapping the coding region of the envelope glycoproteins [8,9].

Rev has been reported to exert its effect at two or more stages in the expression of a viral gene product: first by expediting the export to the cytoplasm of viral RNAs bearing the RRE [1,4,5,10–16], and subsequently by improving the translational potency of RRE-containing transcripts [17–19]. What causes viral mRNAs to be inadequately expressed is not clear. Several studies point to the existence of poorly localized sequences within the *gag* and *env* coding regions that attenuate expression [10,20–24]. One of these inhibitory sequences has been shown to overlap the RRE itself [25–27], but others have proven more difficult to localize. For example, the concerted change of four coding segments, of approximately 30 nucleotides each, within the 5′ end of the *gag* coding sequence produced a transcript that was substantially *rev*-independent; further dissection of the inhibitory sequences proved difficult, however, as mutation of all four regions was required to sustain the increased protein synthesis [28]. Cellular factors, including heterogenous ribonucleoprotein C (hnRNP C), or a serologically related protein, have been found to form complexes with a 270 residue inhibitory sequence, suggesting that intranuclear retention may be mediated by sequence-specific interactions [29]. Nonetheless, an explanation for the lower translational efficiency of the mRNA template remains to be advanced.

The envelope proteins of HIV-1 are a natural target for vaccines and for post-infection treatments to limit viral spread. Despite a great deal of effort aimed at developing and studying the viral envelope proteins, little is known about the factors that limit their expression, the role of non-CD4 cofactors in viral entry, or the mechanisms of tissue tropism. The poor expression of envelope proteins limits the titer of retroviral pseudotypes, making it difficult to develop analytical tools that restrict the success or failure of infection to membrane-proximal events. In attempting to address this problem, we considered the profound bias in the codon usage of the *env* proteins, a bias that extends to the *gag* and *pol* proteins [30–33]. Figure 1 illustrates the divergence between the codon prevalence found in the coding region for the envelope protein of the HIV-1 LAV isolate and that found in a compendium of highly expressed human genes. Here, we report on the consequences of codon bias, and on the favorable outcome

Figure 1



Codon usage of HIV-1 envelope (env) and highly expressed human (high) genes. The frequencies (×100) of the individual codons are shown for each of the degenerately encoded amino acids, and the most prevalent codon is shown in bold.

of the systematic replacement of the native codons of gp120 with codons chosen to reflect more closely the codon preference of highly expressed human genes.

The surprising efficacy of codon replacement for gp120 led us to explore the consequences of codon re-engineering for the expression of two small test proteins: Thy-1, an abundant cell-surface protein of rodent thymocytes, and green fluorescent protein (GFP) [34–36], a jellyfish (*Aequorea victoria*) protein that is relatively poorly expressed in mammalian cells. Conversion of the Thy-1 coding sequence to that of a synthetic gene with the codon preference of HIV-1 envelope protein gave severely attenuated expression, whereas replacement of the endogenous codons of GFP with those of highly expressed mammalian proteins resulted in a substantial increase in synthetic efficiency.

Results

Construction of a synthetic gp120 gene based on optimal codon usage

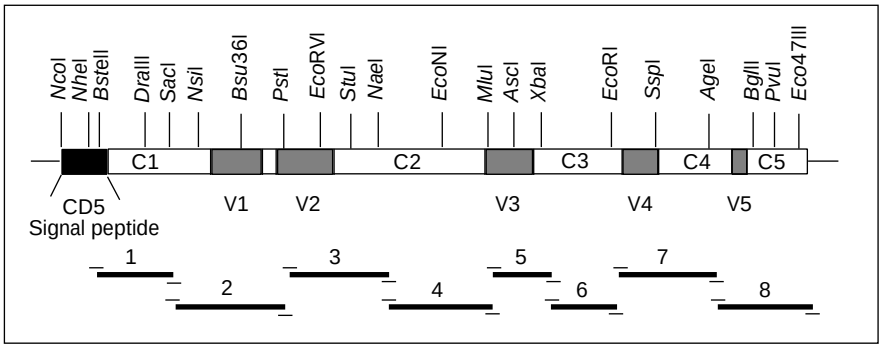
To explore whether codon bias accounted for the poor expression of envelope glycoproteins we constructed a synthetic gene encoding the gp120 segment of HIV-1, based on the sequence of the prototype virus of the most common North American subtype, HIV-1MN. The synthetic gp120 was assembled from chemically synthesized long oligonucleotides that were subsequently amplified in the crude state by polymerase chain reaction (PCR). Some deviations from strict adherence to optimized codon usage were made to accommodate the introduction of unique restriction sites into the resulting gene at approximately 100 base-pair intervals (Fig. 2). In addition, because the endogenous secretory peptide of envelope is known to function inefficiently [37], it was replaced by the leader peptide of the human CD5 antigen, which efficiently directs the synthesis and export of secreted and membrane-bound proteins [38]. The chimeric gp120 precursor gene was inserted in a mammalian cDNA expression vector under the control of the human cytomegalovirus (CMV) immediate early promoter.

Expression of wild-type and synthetic gp120

To evaluate the relative potency of the wild-type and synthetic gp120 coding sequences, we compared the outcomes of transient transfection of the synthetic and native genes, as well as the result of replacement of the endogenous gp120 secretory leader sequence with that of the human CD5 antigen. As shown in Figure 3, the synthetic gene product was expressed at a very high level compared with that of the wild-type gp120, whether expressed with the native or CD5 leader, and as assessed either by immunoprecipitation with a CD4–immunoglobulin fusion protein or by enzyme-linked immunosorbent assay; in the latter assay, expression of synthetic gp120 exceeded that of the native protein with CD5 leader by 41-fold to several hundred-fold, depending on the experiment. Because optimal expression of native envelope sequences requires both the RRE in *cis* and *rev* in *trans*, a similar experiment was conducted in which both the synthetic and endogenous genes were endowed with the RRE, and expression

Figure 2

Synthesis of a gp120 (HIV-1MN) gene with the codon preference of highly expressed human genes. The shaded portions marked V1 to V5 indicate segments of the envelope that show high variability between different natural isolates, and C1 through C5 denote regions that remain relatively constant by the same criteria. Unique restriction sites engineered into the sequence are shown above, and a bar diagram of the chemically synthesized DNA fragments which served as PCR templates is shown below the sequence.



was measured by immunoprecipitation following cotransfection of the envelope expression plasmids and a plasmid expressing *rev*. The results show that *rev* had no effect on the expression of the synthetic gene product, but significantly enhanced expression of the native gene product (Fig. 4). Thus, the action of *rev* is not apparent on a substrate which lacks the coding sequence of the endogenous viral envelope sequences. Because *rev* appears to exert its effect at two steps in the expression of a viral transcript, we sought to clarify the possible role of export in the improved expression of the synthetic gene; we therefore turned to expression from recombinant vaccinia viruses.

Cytoplasmically transcribed synthetic gp120 is more efficiently expressed

Native gp120 contains a transcriptional termination sequence for vaccinia early transcripts, so we used for comparison a variant virus from which this sequence had been removed for higher production [39]. As shown in Figure 5, increased expression of the synthetic gene was demonstrable when the endogenous gene product and the synthetic gene product were expressed from vaccinia virus recombinants under the control of the strong mixed early and late 7.5 k promoter. Densitometry showed the improvement to be from 8–50 fold, depending on the cell type. As vaccinia virus transcripts are created and translated in the cytoplasm of infected cells, the increased expression of the synthetic gp120 gene in this circumstance cannot be attributed to improved export from the nucleus. RNA blot analysis showed that the vaccinia transcripts encoding the synthetic gp120 were less abundant than the transcripts encoding native gp120, indicating that the increased efficiency of expression was not attributable to increased mRNA stability (Fig. 6).

Figure 3

Expression of synthetic gp120 in transient transfection assays. (a) Gel electrophoresis of immunoprecipitated supernatants of 293T cells transfected with plasmids expressing gp120 encoded by the IIIB isolate of HIV-1, (gp120IIIb), the MN isolate (gp120mn), the MN isolate modified by substitution of the endogenous leader peptide with that of the CD5 antigen (gp120mnCD5L), or the chemically synthesized gene encoding the MN variant with the human CD5 leader (syngp120mn). Supernatants were harvested following a 12 h labeling period, 40 h post-transfection, and immunoprecipitated with CD4–IgG1 fusion protein and protein A agarose. (b) ELISA of supernatants of transiently transfected cells. Supernatants of 293T cells transfected by calcium phosphate with the constructs described above were harvested after 4 days and tested in a gp120/CD4 ELISA.

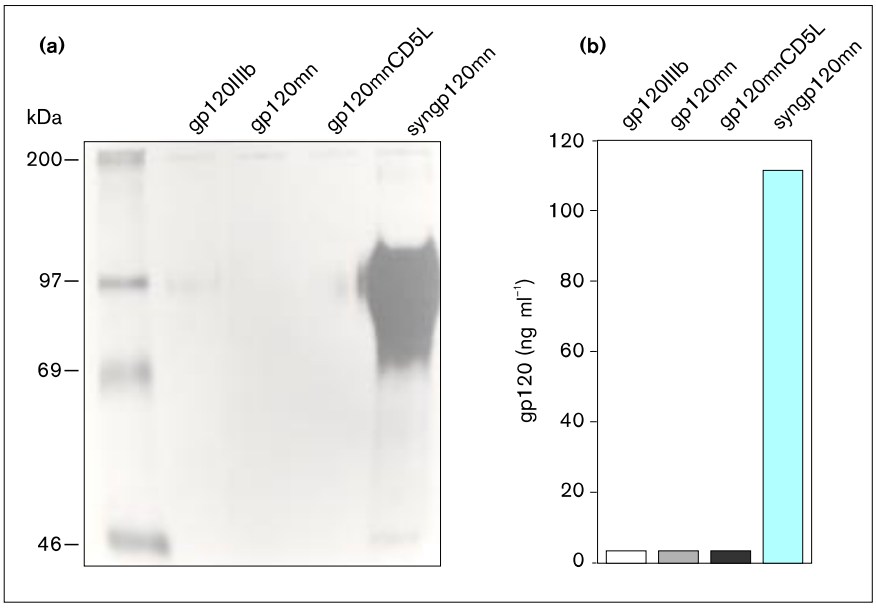
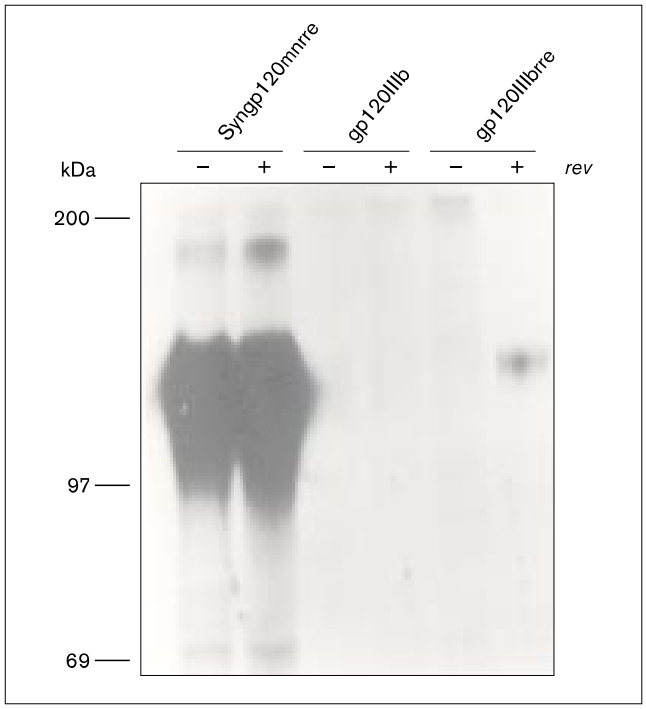
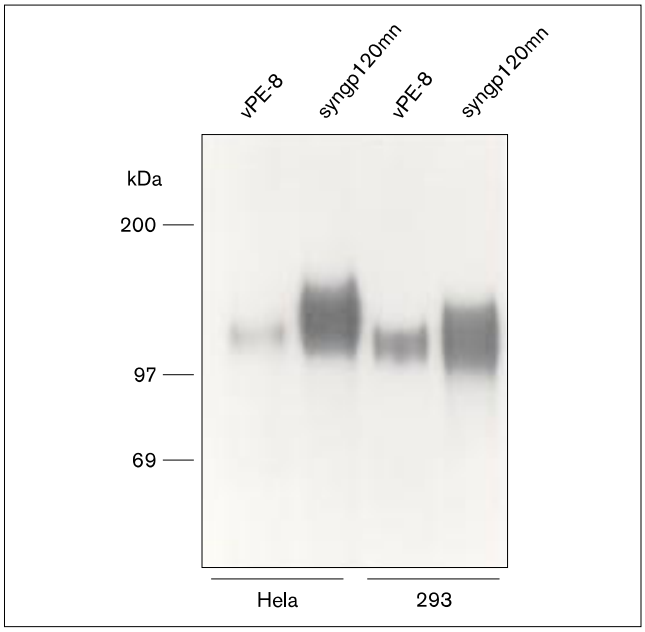


Figure 4



Expression of synthetic and wild-type envelope sequences in the presence of *rev* in *trans* and the RRE in *cis.* 293T cells were transiently transfected by calcium phosphate coprecipitation of 10 μg plasmid expressing: the synthetic gp120MN sequence and RRE in *cis* (syngp120mnrre), the gp120 portion of HIV-1 IIIB (gp120IIIb), and the same sequence with the RRE in *cis* (gp120IIIbrre); each gp120 expression plasmid was cotransfected with 10 μg of either pCMVrev or CDM7 plasmid DNA. Supernatants were harvested 60 h post-transfection and immunoprecipitated with CD4–IgG fusion protein and protein A agarose. The gel exposure time was extended to allow the induction of gp120IIIBrre by *rev* to be demonstrated. Shorter exposures showed no significant dependence of syngp120mnrre expression upon *rev.*

Figure 5



Cytoplasmic expression of syngp120mn by vaccinia virus. Immunoprecipitation of supernatants of human 293 or Hela cells infected with vaccinia virus expressing wild-type gp120IIIb (vPE-8) or synthetic gp120mn (syngp120mn). Cells were infected at a multiplicity of infection of at least 10. Supernatants were harvested after 24 h of infection and immunoprecipitated with CD4–IgG fusion protein and protein A agarose.

HIV codon patterns undermine the efficiency of Thy-1 expression

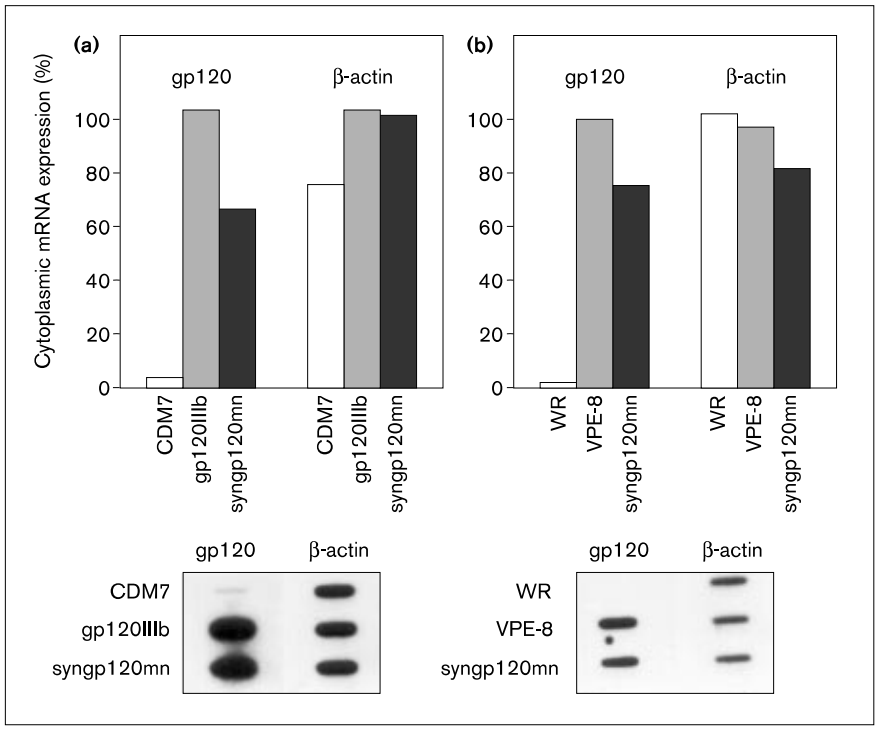
To explore further the importance of codon usage in envelope protein expression, we replaced the codons of a small, typically highly expressed cell-surface protein, the rat Thy-1 antigen, with the codons most frequently used by the HIV-1 envelope protein. The resulting sequence was then edited to remove any message destabilization motifs of the form AUUUA that had been created, and to introduce two restriction sites for ease of creation and manipulation of the resulting sequence. The synthesis strategy is shown in Figure 7. As shown in Figure 8, the synthetic rat Thy-1 was expressed approximately 100-fold less well than the native gene.

As codon usage for highly expressed proteins can vary from organism to organism, sequence re-engineering may prove to be a generally useful way to improve the expression of genes poorly expressed in mammalian cells. To explore the generality of this approach we selected a second small model protein.

Codon replacement enhances green fluorescent protein expression

The GFP of the jellyfish *Aequorea victoria* [34–36] has attracted attention recently for its possible use as a marker or reporter for transfection and lineage studies [40], but has been found to be expressed poorly in mammalian cells. Examination of a codon usage table constructed from the native coding sequence showed that the GFP codons favored either A or U in the third position. The bias in this case favors A less than does the bias of gp120, but is substantial. A synthetic gene was created in which the GFP sequence was recreated in the same manner as for gp120, and the initiation consensus was replaced with sequences corresponding to the translational initiation consensus. The expression of the resulting protein was contrasted with that of the wild-type sequence, similarly engineered to bear an optimized translational initiation consensus (Fig. 9). In addition, the effect of replacing the serine at position 65 with a threonine (S65T), reported to improve excitation efficiency at 490 nm and hence preferred for fluorescence microscopy [41], was examined (Fig. 9).

Figure 6



Increased mRNA concentration does not account for the higher output of synthetic gp120. Cytoplasmic RNA was prepared from (a) COS cells or (b) 293 cells transfected with expression plasmids (a) or infected with recombinant vaccinia viruses (b). RNA was quantitated by slot blot and the relative concentrations are shown above images of the slot blot autoradiogram. Actin concentrations are shown as a control. In all cases the mRNA concentration of the synthetic gene was less than that of the native gene. CDM7, expression vector only. WR, wild-type vaccinia. vPE-8, recombinant vaccinia expressing native gp120IIIb.

Codon engineering conferred a significant increase in expression efficiency (and concomitant percentage of cells apparently positive for transfection), and the combination of the S65T mutation and codon optimization resulted in a DNA segment encoding a highly visible mammalian marker protein (Fig. 9). There was a net improvement in fluorescence per cell of between 40–120 fold, depending on detection conditions.

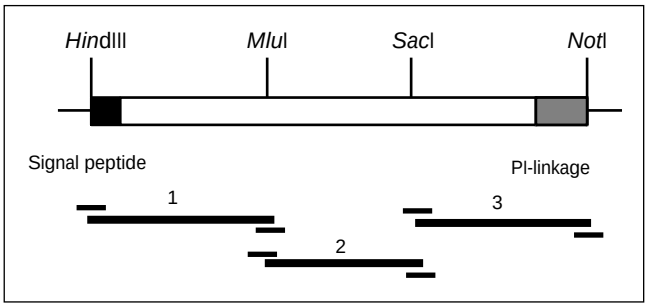
Discussion

Phenomenological rules for codon patterns

Examination of Figure 1 shows that three simple rules describe the pattern of envelope codon usage: first, preferred codons maximize the number of adenine residues in the viral RNA; second, T is preferred over C whenever the third position degeneracy is a pyrimidine; and third, the dinucleotide CG is highly under-represented. In all cases but one, the first rule implies that, if the third position can be A, that codon will be most frequently used [31]. In the special case of serine, three codons equally contribute one A residue to the mRNA; together these three comprise 85 % of the codons found in envelope transcripts. A particularly striking example of the combined effects of first and third rules is found in the codon choice for arginine, in which the AGA triplet comprises 88 % of all codons. Finally, the third rule implies that the third position is much less likely to be G whenever the second position is C, as in the codons for alanine, proline, serine and threonine (Fig. 1).

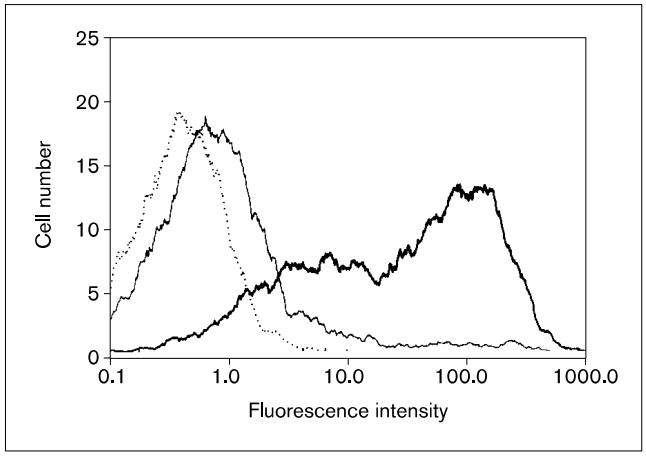
Several studies have pointed to a tendency for HIV reverse transcriptase to exhibit a high frequency of G to A transition, a phenomenon called G to A hypermutation [42–44], and it is known that other lentiviruses undergo similar hypermutation [45]. Although this helps to account for the general prevalence of A and T, additional hypotheses are needed to account for the bias of A over T in the third position, a bias that appears to reflect a more general excess of purines in the viral RNA strand. One suggestion is that the nucleotide bias should be restated as high A, low C, and is driven by two factors: the low intracellular

Figure 7



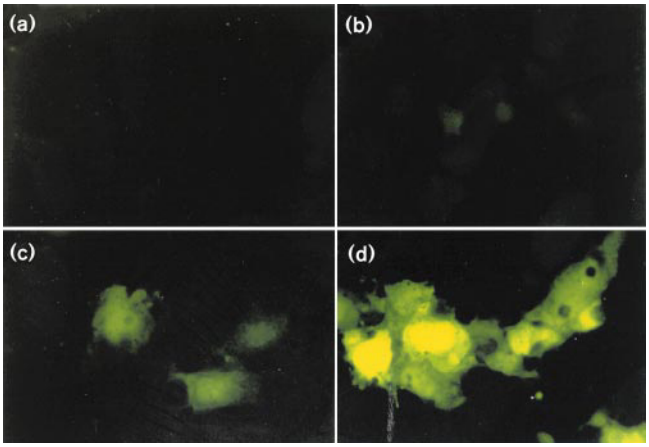
Schematic diagram of a rat Thy-1 gene with HIV-1 envelope codon usage (rTHY-1env). The darkened boxes in the rTHY-1env construct denote the leader peptide and the sequences in the precursor that direct the attachment of a phosphatidylinositol glycan anchor. Unique restriction sites used for assembly of the construct are shown above.

Figure 8



Surface expression of rat Thy-1 with HIV-1 envelope codon usage. Shown are flow cytometry histograms of 293T cells transiently transfected with either wild-type rat Thy-1 (thick line), rat Thy-1 with envelope codons (thin line) or vector only (dotted line). 293T cells were transfected with the different expression plasmids by calcium phosphate coprecipitation and stained with anti-rat Thy-1 monoclonal antibody OX7 followed by a polyclonal FITC-conjugated anti-mouse IgG antibody 3 days after transfection.

Figure 9



Expression of GFP. (a) COS cells transfected with vector only. (b) COS cells transfected with a CDM7 expression plasmid encoding native GFP engineered to include a consensus translational initiation sequence. (c) COS cells transfected with an expression plasmid having the same flanking sequences and initiation consensus as in (b), but bearing a codon-optimized gene sequence. (d) COS cells transfected with an expression plasmid as in (c), but encoding GFP(S65T).

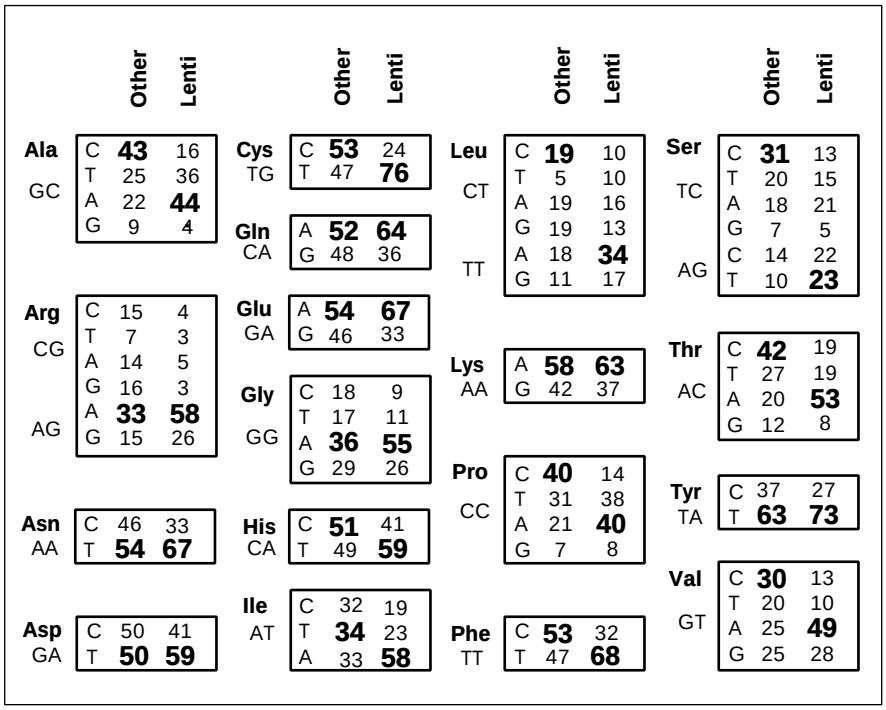
concentration of dCTP relative to TTP; and the higher error rate of HIV reverse transcriptase on RNA templates than on DNA, which may promote transversions from pyrimidine residues in the RNA ([44,46]; S. Wain-Hobson, personal communication).

Why should inefficient codons be prevalent in HIV-1?

Although it is often thought that viruses occupy an evolutionary niche which places a premium on rapid replication, persistent infection of a fraction of susceptible individuals is a common feature of several viral strains and may be important for the maintenance of a reservoir of viral genomes that can be transmitted to new hosts. This appears to be especially true of HIV-1 and HIV-2. Codon bias may allow the expression of such viruses to be suppressed in order to minimize the antigenic profile. An opposing hypothesis, that the presence of adenine residues in genomic RNA is generally favored in the retroviral life cycle — for example, by facilitating the process of replication — is not consistent with earlier reports that Visna virus and HIV-1 share similar codon usage, whereas human T-cell leukemia virus type 1 (HTLV-1) and HIV-1 do not [30], or that murine leukemia virus and HIV-1 are similarly discordant [31]. Figure 10 illustrates this point with a compilation of codon frequencies chosen from envelope glycoproteins of a variety of other retroviruses — but excluding the lentiviruses — compared with those drawn from the envelope sequences of five lentiviruses that are not closely related to HIV-1: bovine immunodeficiency virus, caprine arthritis encephalitis virus, equine infectious anemia virus, feline immunodeficiency virus, and Visna virus. The codon usage patterns for the lentiviruses is strikingly similar to that of HIV-1 — in all cases the preferred codon for HIV-1 is the same as the preferred codon for the other lentiviruses. In contrast, the non-lentiviral envelope codons do not show a similar predominance of A residues, and are also not as skewed toward third position C and G residues as are the highly expressed human genes. In general the non-lentiviral retroviruses appear to exploit the different codons more equally, a pattern they share with less highly expressed human genes (data not shown). In addition to the prevalence of codons including A, the lentiviral codons also show the HIV-1 pattern of strong CpG under-representation, such that the third position for alanine, proline, serine and threonine triplets is rarely G. The retroviral envelope triplets of the comparison group show a similar, but less pronounced under-representation of CpG. The most obvious difference between lentiviruses and other retroviruses with respect to CpG prevalence lies in the usage of the CGX variant of arginine triplets, which is relatively frequently represented among the retroviral envelope coding sequences, but is almost never present among lentiviral sequences. The under-representation of CpG-bearing triplets in lentiviruses could be a consequence of genetic selection to avoid transcriptional silencing by methylation of CpG cytosines [47], consistent with the finding that CpG dinucleotides are also under-represented among untranslated segments of the HIV-1 genome (data not shown).

Figure 10

Codon usage of lentiviral (lenti) and other retroviral envelope genes. The frequencies (×100) of the individual codons are shown for each of the degenerately encoded amino acids, and the most prevalent codon is shown in bold.



Representatives of two unusual types of retrovirus were not included in the group compared with the lentiviruses in Figure 10: the spumaretroviruses, exemplified by bovine syncytial virus and the human and simian spumaviruses; and Walleye dermal sarcoma virus (WDSV), which appears to constitute a new genus [48]. Viruses of these two classes show envelope codon-usage patterns similar to those of the lentiviruses and, like the lentiviruses, have large genomes that encode several short open reading frames in addition to the *gag*, *pol* and *env* constituents [48]. Intriguingly, the spumaviruses, WDSV, and at least some of the lentiviruses share the presence of single-stranded gaps in their unintegrated linear DNA [48–52]. This small constellation of similar attributes suggests they may have arisen from a common progenitor.

Why is there codon preference in any form?

Codon bias has been observed in many species [53–56], and a correlation has been noted between high expression and the use of a stereotyped pattern of codons [54,57]. Although this correlation is not universal ([58]; see also discussion in [59]), in several cases it has been found that expression of exogenous gene products in *Escherichia coli* can be enhanced by systematic substitution of the endogenous codons with triplets over-represented in highly expressed *E. coli* genes [59,60]. Although highly expressed mammalian genes show nonrandom codon-usage patterns, the degree of over-representation of favored codons is not as pronounced as for highly expressed bacterial genes. Nonetheless, the data presented here clearly suggest that codon usage can play an important role in determining translational efficiency in a mammalian cell context.

It has been widely assumed that translational efficiencies of mammalian gene products are governed by initiation; if they were not, mRNAs would, in general, be maximally loaded with ribosomes. This study suggests that poorly translated mRNAs may indeed be maximally loaded. However, the finding that inferior codons limit translational efficiency is not necessarily inconsistent with the view that translational efficiency is governed by initiation. First, the limitation imposed by inferior codons need not be kinetic. Instead, an abundance of less favored codons could incur a significant cumulative probability of failure to complete the nascent polypeptide. Recent evidence, however, suggests that premature termination does not occur with a high enough efficiency to account for the observed stimulation (E.C.P. and B.S., unpublished observations). Another possibility is that the inferior codons reflect a selection for RNA structures that influence the rate of initiation — for example, if access to ribosomes is controlled by cues distributed throughout the RNA. In such a case, lentiviral codons could predispose the RNA to accumulate in a pool of poorly initiated RNAs. The sequestered RNA might be given an improved rate of initiation by the action of *rev* or other regulatory proteins. No stimulation by *rev* would be seen in the event that the

RNA did not bear the unfavored sequences, consistent with these results. Some precedent for the existence of discrete pools of mRNA can be found in studies of maternal transcript distribution and use in early embryogenesis [61,62]. RNA adenine methylation, a candidate RNA modification that might dictate transcript use, seems unlikely to participate in translational suppression because the known methylation consensus [63] is not substantially over-represented among HIV sequences relative to other retroviruses (data not shown; there is a profound strand bias in the frequency of the RNA methylation consensus, however, which is consistent with the purine excess in the sense strand).

Conclusions

The creation of a synthetic coding sequence based on codons over-represented in highly expressed human genes overcomes a major limitation to the translational efficiency of HIV-1 envelope glycoprotein. Although codon optimization has been used in the past to improve expression in *E. coli* of genes from other organisms, similar studies have not been conducted in mammalian cells, perhaps because the codon bias of highly expressed mammalian genes is not as striking as that of *E. coli* genes. However, the results obtained here with three unrelated proteins — HIV-1 gp120, rat cell-surface antigen Thy-1 and GFP from *Aequorea victoria* — suggest that codon optimization may prove to be a fruitful strategy for improving the expression in mammalian cells of genes that show limited translational efficiency in their native form.

Materials and methods

Plasmid constructions

The synthetic gp120 gene was generated from eight 160–200 base oligonucleotides, produced on a Milligen 8750 synthesizer. After elution with 30 % ammonium hydroxide, the oligonucleotides were deblocked at 55 °C for 12 h, precipitated with n-butanol and resuspended in H_2O. 15–25-mer oligonucleotides complementary to the ends were used to amplify the long oligonucleotides by PCR. Typically, PCR was carried out using 35 cycles with 55 °C annealing temperature and 0.2 min extension time. The products were gel purified, phenol extracted, and used in a subsequent overlap PCR to generate longer fragments consisting of two adjacent small fragments. These fragments were cloned into a CDM7-derived plasmid containing a leader sequence of the CD5 surface molecule followed by a *Nhe*1/*Pst*1/*Mlu*1/*Eco*R1/*Bam*H1 polylinker. The correct sequence was confirmed by DNA sequencing. The gp120IIIb construct was generated by PCR using a *Sal*1/*Xho*1 HIV-1 HXB2 envelope fragment as template, followed by an exchange of a *Kpn*1/*Ear*1 fragment from a proviral clone. The wild-type gp120 mn constructs used as controls were cloned by PCR from HIV-1MN (NIH AIDS Repository) infected C8166 cells and contained either the native *env* leader or a CD5 leader sequence. Two clones of each construct were tested to avoid PCR-induced artefacts. The rat Thy-1 gene with the HIV envelope codon usage (rTHY-1env) was generated using three 150–170-mer oligonucleotides. In contrast to the syngp120mn, the PCR products were directly cloned and assembled in pUC12, and subsequently cloned into CDM7.

A GFP coding sequence was assembled in a similar manner from six fragments of approximately 120 bp each, using a strategy for assembly that relied on the ability of the restriction enzymes *Bsa*I and *Bbs*I to cleave outside their recognition sequence. Long oligonucleotides were synthesized which contained portions of the coding sequence of GFP embedded in flanking sequences encoding *Eco*RI and *Bsa*I sites at one end, and *Bam*HI and *Bbs*I sites at the other end, in the configuration *Eco*RI–*Bsa*I–GFP–*Bbs*I–*Bam*HI. The restriction site ends generated by the *Bsa*I and *Bbs*I sites were designed to yield compatible ends between adjacent GFP fragments, each of which was unique and non-selfcomplementary. The crude synthetic DNA segments were amplified by PCR, inserted between *Eco*RI and *Bam*HI in pUC9, and sequenced. Subsequently, the intact coding sequence was assembled in a six fragment ligation, using insert fragments prepared with *Bsa*I and *Bbs*I. Two of six plasmids resulting from the ligation bore an insert of correct size, and one contained the desired full length sequence. Mutation of Ser65 to Thr was accomplished by PCR, using a primer that overlapped a unique *Bss*SI site in the synthetic GFP.

Immunoprecipitation

293T cells were transfected by calcium phosphate using small dishes of 50–70 % confluent cells and 10 μg plasmid DNA. In cotransfection experiments with *rev*, cells were transfected with 10 μg gp120IIIb, gp120IIIbrre, syngp120mnrre or rTHY-1enveg1rre expression plasmid DNAs and 10 μg pCMVrev (AIDS Repository) or CDM7 plasmid DNA. After 48–60 h medium was exchanged and cells were incubated for additional 12 h in Cys/Met-free medium (Gibco) containing 200 μCi [^{35}S]Cys+Met (ICN). Supernatants were harvested and spun for 15 min at 3000 rpm to remove debris. After addition of the protease inhibitors leupeptin, aprotinin and PMSF (Sigma) to 2.5 μg ml^{-1}, 50 μg ml^{-1} and 100 μg ml^{-1}, respectively, 1 ml of supernatant was incubated with either 10 ml of packed protein A sepharose (Sigma) alone (rTHY-1enveg1rre) or with protein A sepharose and 3 mg of a purified CD4–immunoglobulin fusion protein (kindly provided by Behring) at 4 °C for 12 h. The protein A beads were washed 5 times with a buffer containing 100 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$, 1 % NP-40. After the final wash, 10 μl of loading buffer containing 10 % glycerol, 4 % SDS, 4 % mercaptoethanol and 0.002 % bromphenol blue was added, samples were boiled for 3 min and applied on 7 % (all gp120 constructs) or 10 % (rTHY-1enveg1rre) SDS polyacrylamide gels. Gels were fixed in 10 % acetic acid and 10 % methanol, incubated with Amplify (Amersham) for 20 min, dried and exposed for 12 h.

ELISA

The concentration of gp120 in culture supernatants was determined using CD4-coated ELISA plates and goat anti-gp120 antisera in the soluble phase. Supernatants of 293T cells transfected by calcium phosphate were harvested after 4 days, spun at 3000 rpm for 10 min to remove debris and incubated for 12 h at 4° C on the plates (NEN Dupont). After 6 washes with PBS, 100 μl goat anti-gp120 antisera diluted 1:200 (AIDS Repository) were added for 2 h. The plates were washed again and incubated for 2 h with a peroxidase-conjugated rabbit anti-goat IgG antiserum 1:1000 (Cappel). Subsequently, the plates were washed and incubated for 30 min with 100 μl substrate solution containing 2 mg ml^{-1} o-phenylendiamine (Sigma) in sodium citrate buffer. The reaction was finally stopped with 100 μl 4 M sulfuric acid. Plates were read at 490 nm with a Coulter microplate reader. As a control, purified recombinant gp120IIIb (Repligen) was used.

Vaccinia virus recombinants

Recombinant vaccinia virus expressing syngp120mn under the control of the 7.5 promoter was generated by cloning a *Hin*d3/*Not*1 fragment containing the synthetic gp120mn gene into pTKG. Vaccinia recombination, plaque purification and generation of high-titer virus stocks was done in CV-1 cells essentially as described [64]. vPE-8 expresses HIV-1 IIIB gp120 under the 7.5 promoter [39]. Hela or 293 cells were infected at a multiplicity of infection of at least 10. Supernatants of ^{35}S-labelled cells were harvested 24 h post infection and immunoprecipitated as described above.

Immunofluorescence

293T cells were transfected by calcium phosphate and analyzed for surface Thy-1 expression after 3 days. Cells were detached with 1 mM EDTA in PBS and stained with the monoclonal antibody OX-7 (Accurate) and a FITC-conjugated goat anti-mouse immunoglobulin antiserum (Cappel). The analysis was done on a EPICS XL cytofluorometer (Coulter).

Compilation of codon frequency tables

Codon usage by retroviruses was compiled from the envelope precursor sequences of avian leukosis virus, bovine leukemia virus, feline leukemia virus, gibbon ape leukemia virus, human T-cell leukemia virus type I, Mason–Pfizer monkey virus, mouse mammary tumor virus, the 10A1 isolate of murine leukemia virus, the 4070A amphotropic isolate of MLV, rat leukemia virus, and simian sarcoma virus. The codon frequency tables for the non-HIV, non-SIV lentiviruses were compiled from the envelope precursor sequences for bovine immunodeficiency virus, caprine arthritis encephalitis virus, equine infectious anemia virus, feline immunodeficiency virus, and Visna virus.

Acknowledgements

We thank Simon Wain-Hobson for a stimulating discussion and many suggestions, and Patricia Earl, Robert Gallo, Marie-Louise Hammarskjöld, Bernard Moss and David Rekosh for provision of reagents under the auspices of the NIH AIDS Research and Reference Reagent Program. J.H. was supported by a fellowship from the DKFZ. Portions of this work were supported by NIH grant HL53694 and an award to the Massachusetts General Hospital from Hoechst AG.

References

1. Daly TJ, Cook KS, Gray GS, Maione TE, Rusche JR: Specific binding of HIV-1 recombinant rev protein to the Rev-responsive element ***in vitro***. *Nature* 1989, 342:816–819.
2. Hadzopoulou-Cladaras M, Felber BK, Cladaras C, Athanassopoulos A, Tse A, Pavlakis GN: The ***rev*** (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a ***cis***-acting sequence in the env region. *J Virol* 1989, 63:1265–1274.
3. Zapp ML, Green MR: Sequence-specific RNA binding by the HIV-1 ***rev*** protein. *Nature* 1989, 342:714–716.
4. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR: The HIV-1 ***rev*** transactivator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 1989, 338:254–257.
5. Olsen HS, Nelbock P, Cochrane AW, Rosen CA: Secondary structure is the major determinant for interaction of HIV ***rev*** protein with RNA. *Science* 1990, 247:845–848.
6. Huang XJ, Hope TJ, Bond BL, McDonald D, Grahl K, Parslow TG: Minimal Rev-response element for type 1 human immunodeficiency virus. *J Virol* 1991, 65:2131–2134.
7. Kjems J, Brown M, Chang DD, Sharp PA: Structural analysis of the interaction between the human immunodeficiency virus ***rev*** protein and the rev response element *Proc Natl Acad Sci USA* 1991, 88:683–687.
8. Feinberg MB, Jarrett RF, Aldovini A, Gallo RC, Wong- Staal F: HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* 1986, 46:807–817.
9. Sodroski J, Goh WC, Rosen C, Dayton A, Terwilliger E, Haseltine W: A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature* 1986, 321:412–417.
10. Emerman M, Vazeux R, Peden K: The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. *Cell* 1989, 57:1155–1165.
11. Felber BK, Derse D, Athanassopoulos A, Campbell M, Pavlakis GN: Cross-activation of the Rex proteins of HTLV- I and BLV and of the rev protein of HIV-1 and nonreciprocal interactions with their RNA responsive elements. *New Biol* 1989, 1:318–328.
12. Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green,SM, Karn J, *et al*: HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the rev response element region. *Cell* 1990, 60:685–693.
13. Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR: HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell* 1990, 60:675–683.
14. Malim MH, Cullen BR: ***rev*** and the fate of pre- mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes *Mol Cell Biol* 1993, 13:6180–6189.
15. Fischer U, Meyer S, Teufel M, Heckel C, Luhrmann R, Rautmann G: Evidence that HIV-1 rev directly promotes the nuclear export of unspliced RNA. *EMBO J* 1994, 13:4105–4112.
16. Bogerd HP, Fridell RA, Madore S, Cullen BR: Identification of a novel cellular cofactor for the rev/rex class of retroviral regulatory proteins. *Cell* 1995, 82:485–494.
17. Arrigo SJ, Chen IS: ***rev*** is necessary for translation but not cytoplasmic accumulation of HIV-1 vif, vpr, and env/vpu 2 RNAs. *Genes Dev* 1991, 5:808–819.
18. Lawrence JB, Cochrane AW, Johnson CV, Perkins A, Rosen CA: The HIV-1 rev protein: a model system for coupled RNA transport and translation. *New Biol* 1991, 3:1220–1232.
19. D'Agostino DM, Felber BK, Harrison JE, Pavlakis GN: The ***rev*** protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. *Mol Cell Biol* 1992, 12:1375–1386.
20. Rosen CA: HIV regulatory proteins: potential targets for therapeutic intervention. *AIDS Res Hum Retroviruses* 1992, 8:175–181.
21. Dayton AI, Terwilliger EF, Potz J, Kowalski M, Sodroski JG, Haseltine WA: Cis-acting sequences responsive to the ***rev*** gene product of the human immunodeficiency virus J. *AIDS* 1988, 1:441–452.
22. Cochrane AW, Jones KS, Beidas S, Dillon PJ, Skalka AM, Rosen CA: Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression. *J Virol* 1991, 65:5305–5313.
23. Maldarelli F, Martin MA, Strebel K: Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation. *J Virol* 1991, 65:5732–5743.
24. Schwartz S, Felber BK, Pavlakis GN: Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of ***rev*** protein. *J Virol* 1992, 66:150–159.
25. Brighty DW, Rosenberg M: A ***cis***-acting repressive sequence that overlaps the Rev-responsive element of human immunodeficiency virus type 1 regulates nuclear retention of env mRNAs independently of known splice signals. *Proc Natl Acad Sci USA* 1994, 91:8314–8318.
26. Zolotukhin AS, Valentin A, Pavlakis GN, Felber BK: Continuous propagation of RRE(-) and Rev(-)RRE(-) human immunodeficiency virus type 1 molecular clones containing a cis- acting element of simian retrovirus type 1 in human peripheral blood lymphocytes. *J Virol* 1994, 68:7944–7952.
27. Barksdale SK, Baker CC: The human immunodeficiency virus type 1 rev protein and the Rev-responsive element counteract the effect of an inhibitory 5' splice site in a 3' untranslated region. *Mol Cell Biol* 1995, 15:2962–2971.
28. Schwartz S, Campbell M, Nasioulas G, Harrison J, Felber BK, Pavlakis GN: Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J Virol* 1992, 66:7176–7182.
29. Olsen HS, Cochrane AW, Rosen C: Interaction of cellular factors with intragenic cis-acting repressive sequences within the HIV genome. *Virology* 1992, 191:709–715.
30. Grantham R, Perrin P: AIDS virus and HTLV-I differ in codon choices. *Nature* 1986, 319:727–728.
31. Kypr J, Mrazek J: Unusual codon usage of HIV. *Nature* 1987, 327:20–20.
32. Sharp PM: What can AIDS virus codon usage tell us? *Nature* 1986, 324:114–114.
33. Chou KC, Zhang CT: Diagrammatization of codon usage in 339 human immunodeficiency virus proteins and its biological implication. *AIDS Res Hum Retrovir* 1992, 8:1967–976.
34. Ward WW: Energy transfer processes in bioluminescence. *Photochem Photobiol Rev* 1979, 4:1–57.
35. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ: Primary structure of the ***Aequorea victoria*** green-fluorescent protein. *Gene* 1992, 111:229–233.
36. Cody CW, Prasher DC, Westler WM, Prendergast FG, Ward WW: Chemical structure of the hexapeptide chromophore of the ***Aequorea*** green-fluorescent protein. *Biochemistry* 1993, 32:1212–1218

37. Lasky LA, Groopman JE, Fennie CW, Benz PM, Capon DJ, Dowbenko DJ, *et al*.: Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. *Science* 1986, 233:209–212.
38. Aruffo A, Stamenkovic I, Melnick M, Underhill CB, Seed B: CD44 is the principal cell surface receptor for hyaluronate. *Cell* 1990, 61:1303–1313.
39. Earl PL, Koenig S, Moss B: Biological and immunological properties of human immunodeficiency virus type 1 envelope glycoprotein: analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. *J Virol* 1991, 65:31–41
40. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC: Green fluorescent protein as a marker for gene expression. *Science* 1994, 263:802–805.
41. Heim R, Cubitt AB, Tsien RY: Improved green fluorescence. *Nature* 1995, 373:663–664.
42. Vartanian JP, Meyerhans A, Asjo B, Wain-Hobson S: Selection, recombination, and G-A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol* 1991, 65:1779–1788.
43. Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, *et al*.: Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* 1992, 358:495–499.
44. Vartanian JP, Meyerhans A, Sala M, Wain-Hobson,S: G→A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc Natl Acad Sci USA* 1994, 91:3092–3096.
45. Wain-Hobson S, Sonigo P, Guyade M, Gazit A, Henry M: Erratic G→A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* 1995, 209:297–303.
46. Sala M, Wain-Hobson,S, Schaeffer F: Human immunodeficiency virus type 1 reverse transcriptase tG:T mispair formation on RNA and DNA templates with mismatched primers: a kinetic and thermodynamic study. *EMBO J* 1995, 14:4622–4627.
47. Eden S, Cedar H: Role of DNA methylation in the regulation of transcription. *Curr Opin Genet Dev* 1994, 4:255–259.
48. Holzschu DL, Martineau D, Fodor SK, Vogt VM, Bowser PR, Casey JW: Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. *J Virology* 1995, 69:5320–5331.
49. Blum HE, Harris JD, Ventura P, Walker D, Staskus K, Retzel E, Haase AT: Synthesis in cell culture of the gapped linear duplex DNA of the slow virus visna. *Virology* 1985, 142:270–277.
50. Kupiec JJ, Tobaly-Tapiero J, Canivet M, Santillana-Hayat M, Flugel RM, Peries J, Emanoil-Ravier R: Evidence for a gapped linear duplex DNA intermediate in the replicative cycle of human and simian spumaviruses. *Nucleic Acids Res* 1988, 16:9557–9565.
51. Charneau P, Clavel F: A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J Virol* 1991, 65:2415–2421.
52. Renshaw RW, Gonda MA, Casey JW: Structure and transcriptional status of bovine syncytial virus in cytopathic infections. *Gene* 1991, 105:179–184.
53. Ikemura T: Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and ***Escherichia coli*** with reference to the abundance of isoaccepting transfer RNAs. *J Mol Biol* 1982, 158:573–597.
54. Grantham R, Gautier C, Gouy M: Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res* 1980, 8:1893–1912.
55. Wain-Hobson S, Nussinov R, Brown RJ, Sussman, JL: Preferential codon usage in genes. *Gene* 1981, 13:355–364.
56. Grosjean H, Fiers W: Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 1982, 18:199–209.
57. Ikemura T: Correlation between the abundance of ***Escherichia coli*** transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the ***E coli*** translational system. *J Mol Biol* 1981, 151:389–409.
58. Holm L: Codon usage and gene expression. *Nucleic Acids Res* 1986, 14:3075–3087.
59. Makoff AJ, Oxer MD, Romanos MA, Fairweather NF, Ballantine S: Expression of tetanus toxin fragment C in ***E coli***: high level expression by removing rare codons. *Nucleic Acids Res* 1989, 17:10191–10202.
60. Williams DP, Regier D, Akiyoshi D, Genbauffe F, Murphy JR: Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed ***Escherichia coli*** genes. *Nucleic Acids Res* 1988, 16:10453–10467.
61. Ding D, Lipshitz HD: Localized RNAs and their functions. *Bioessays* 1993, 15:651–658.
62. Kislauskis EH, Singer RH: Determinants of mRNA localization. *Curr Opin Cell Biol* 1992, 4:975–978.
63. Narayan P, Ludwiczak RL, Goodwin EC, Rottman FM: Context effects on N6-adenosine methylation sites in prolactin mRNA. *Nucleic Acids Res* 1994, 22:419–426.
64. Romeo C, Seed B: Cellular immunity to HIV activated by CD4 fused to T cell or Fc receptor polypeptides. *Cell* 1991, 64:1037–1046.