JVI Accepts, published online ahead of print on 29 January 2014
J. Virol. doi:10.1128/JVI.03223-13

Large-scale nucleotide optimization of simian immunodeficiency virus (SIV) reduces its capacity to stimulate type-I IFN *in vitro*

Nicolas Vabret[1,4], Marc Bailly-Bechet[2], Alice Lepelley[3], Valérie Najburg[1], Olivier Schwartz[3], Bernard Verrier[4], Frédéric Tangy[1] *

[1]Unité de Génomique Virale et Vaccination, CNRS UMR-3569, Institut Pasteur, Paris, France

[2]Laboratoire de Biométrie et Biologie Evolutive, CNRS UMR-5558, Université Claude Bernard Lyon 1, Villeurbanne, France

[3]Unité Virus et Immunité, Institut Pasteur, CNRS UMR-3569, Paris, France

[4]Laboratoire de biologie Tissulaire et Ingénierie thérapeutique, UMR-5305, University of Lyon 1, Lyon, France.

*To whom correspondence should be addressed.
Frédéric Tangy
Unité de Génomique Virale et Vaccination, Institut Pasteur,
25-28 rue du Dr. Roux, 75724 PARIS Cedex 15, France
Tel: +33 (0)1 45 68 87 70
Fax: +33 (0)1 40 61 31 67
Email: ftangy@pasteur.fr

ABSTRACT

Lentiviral RNA genomes present a strong bias in their nucleotide composition with extremely high frequencies of A-nucleotide in HIV-1 and simian immunodeficiency virus (SIV). Based on the observation that human optimization of RNA virus gene fragments may abolish their ability to stimulate type-I interferon (IFN-I) response, we identified the most biased sequences along SIV genome and showed that they are the most potent IFN-I stimulators. With the aim of designing an attenuated SIV genome based on a reduced capacity to activate IFN-I response, we synthesized artificial SIV genomes whose biased sequences were optimized towards macaque average nucleotide composition without altering their regulatory elements or amino acid sequences. A synthetic SIV optimized with 169 synonymous mutations in *gag* and *pol* genes showed a 100-fold decrease in replicative capacity. Interestingly, a synthetic SIV optimized with 70 synonymous mutations in *pol* had a normal replicative capacity. Its ability to stimulate IFN-I was reduced when infected cells were cocultured with reporter cells. IRF3 transcription factor was required for IFN-I stimulation, implicating cytosolic sensors in the detection of SIV biased RNA in infected cells. No reversion of introduced mutations was observed for both optimized viruses after 10 serial passages. In conclusion, we have designed large-scale nucleotide-modified SIVs that may display attenuated pathogenic potential.

IMPORTANCE

In this study, we synthesized artificial SIV genomes in which the most hyper biased sequences were optimized to bring them closer to the nucleotide composition of the macaque SIV host. Interestingly, we generated a stable synthetic SIV optimized with 70 synonymous mutations in *pol* gene, which had a normal replicative capacity but a reduced ability to stimulate type-I IFN. This demonstrates the possibility to rationally change viral nucleotide composition to design replicative and genetically stable lentiviruses with attenuated pathogenic potentials.

INTRODUCTION

RNA viruses carry their whole genetic information on a single or double stranded genomic RNA molecule. They have relatively small genomes, probably due to their high mutation rate (26) and to capsid size constraints (50). Their genomic RNA encodes for few proteins, sometimes with overlapping segments, and contains many non-coding sequences (41). These regulatory sequences play critical roles in controlling transcription, translation, sub-cellular trafficking or packaging, but also virulence functions such as immune evasion (7, 19, 27). Recently, viral RNA properties relying on global features such as nucleotide composition or codon usage have also been revealed crucial in virus biology (48, 49).

It was proposed 30 years ago that each species was subjected to specific genomic pressures on nucleotide composition resulting in a distinctive bias in synonymous codon usage (20). This is true for viruses, which have species-specific nucleotide compositions with most RNA viruses displaying A-rich and C-poor codons in their coding strand (5). For example, the genomes of HIV-1 and other lentiviruses are strongly biased in their nucleotide composition as compared to their primate hosts, with as much as 35 % adenosines (67, 68). As a consequence, both their average amino acid composition and their synonymous codon usage are different from those of their hosts (8, 9). Lentiviral biased nucleotide composition has been explained by dNTP pool imbalance during reverse transcription (17) and by antiviral activity of the cellular Apobec 3G (A3G) cytidine deaminase, which mutates G to A in HIV provirus (10) and is counteracted by the viral protein Vif (17, 61). The specific nucleotide composition of lentiviruses may impact genome structure and stability (38, 70), nuclease sensitivity (43) and viral RNA recognition by innate immunity receptors (39, 57).

To study the impact of global nucleotide composition or codon usage on virus biology, it is necessary to rely on chemical synthesis of large genomic fragments and on reverse genetics to generate modified viruses. In the last years, several groups have synthesized large artificial viral sequences to proceed to genome-scale modifications. For example, the effect of shifting poliovirus codon usage on its replication has been investigated (13, 49). Poliovirus RNA genome was deoptimized without altering the amino-acid sequence by changing synonymous codons from

frequently to rarely used codons. This resulted in imbalanced synthesis of viral proteins and generated attenuated viruses. Attenuated polio and influenza viruses were further developed by deoptimizing the codon pair bias (22) (15, 48). Both deoptimized viruses protected mice from infection after subsequent challenge with a wild-type virus, highlighting their potential as vaccine. In the case of HIV-1, it was observed that systematic replacement of wild-type codons by synonymous GC-rich codons in *gag* and *pol* genes led to a profound delay in replication kinetics with at least 5-fold loss of infectivity (33). Suppression of viral infectivity was caused by enhanced dimer stability of viral RNA genome and subsequent reduction of viral cDNA synthesis (33). These examples demonstrate that genome-scale changes in viral sequence allow the design of attenuated vaccines that are genetically stable because of the large number of mutations involved.

We recently reported that HIV-1 biased nucleotide composition triggers over-stimulation of type I interferon (IFN-I) response after RNA transfection in human cells, indicating that RNA sequences are discriminated according to their nucleotide composition (66). Type I IFN is a major antiviral cytokine thought to contribute to chronic activation of the immune system and progression to AIDS during HIV infection (24, 30). We also proposed a putative link between pathogenicity and divergent HIV-1 nucleotide composition compared to host (66). These results suggested a new determinant for the pathogenicity of lentiviral infections, and raised the possibility of altering virus-host interactions by artificially changing the nucleotide frequency of viral genome.

In the present work, we observed that codon optimization of viral genes, a technique commonly used to increase antigen expression in vaccine candidates (1, 6, 40), abolishes the capacity of viral RNA to induce IFN-I in human cells. Based on the demonstration that distinct regions of SIV genomic RNA trigger different levels of IFN-I according to their nucleotide composition, we designed an attenuated artificial SIV genome by sequence optimization. Viral sequences were made closer to SIV host macaque average nucleotide composition without altering the regulatory elements and the amino acid composition. The corresponding synthetic viruses were produced by reverse genetics and were analyzed for their ability to replicate and to stimulate IFN-I production *in vitro* in a human T-cell line and in human and macaque PBMC.

MATERIALS AND METHODS

In silico design of nucleotide optimized SIV genomes. To design new SIVopt genomes we selected three modifiable regions within *gag*, *pol* and *env* genes of SIVmac239 genome where nucleotide optimization was possible while preserving amino acid sequences and regulatory regions. For each region, the actual profile of chi-square divergence of nucleotide composition between virus and *rhesus* macaque host was computed with a sliding window of 101 nucleotides, as previously described (66). Then, a threshold of maximal acceptable chi-square divergence was selected, according to similar profiles computed with non-pathogenic lentiviruses-hosts couples (SIVagm / African green monkey *Sabeus*, SIVsm / Sooty mangabey) or couples with reduced lentiviral pathogenicity (HIV-2/human, SIVcpz/chimpanzee). The selected thresholds were 0.04 in the *gag* region, 0.14 in the *pol* region, and 0.05 in the *env* region. To identify the nucleotide modifications that would reduce local divergence, each region was divided into peaks of divergence separated from non-divergent sections (Fig. 2B). Inside each region, all genome positions with a chi-square above the given threshold and distant less than 101 nucleotides from another were included in the same peak of divergence. The distance of 101 nucleotides corresponds to the window size for chi-square computation, thus local chi-square at a given position depends only on the nucleotide composition of the 50 surrounding nucleotides and two positions distant of 101 nucleotides are computed independently. This division of genomic regions into independent peaks of divergent and non-divergent sections allowed optimizing the nucleotide content of each peak separately and quickened most computations. Inside each peak, all possible one-nucleotide changes synonymous for all CDS coded by the position were ranked based on their impact on the local chi-square divergence. To minimize the number of mutations necessary to reduce the divergence to the pre-defined threshold, the single nucleotide modification causing the higher diminution of divergence was selected. After each modification, the chi-square profile of the peak was recomputed and new modifications were made iteratively, until either no more genome position in the peak had a chi-square above threshold, or all possible mutations in the peak had been realized, meaning that the selected threshold was unreachable. The next peak was then selected and solved iteratively in the same way until resolution of all peaks inside each region.

Cells and reagents. Human PBMCs were isolated from the blood of healthy donors by Ficoll centrifugation. The blood was provided by the EFS (Etablissement Français du Sang, the French Official Blood Bank). Macaque PBMCs were isolated from Two Chinese RMs (Macaca mulatta) housed in single cages within the "Commissariat à l'Energie Atomique" (Fontenay-aux-Roses, France) facilities according to national guidelines. Whole blood was collected on sodium heparin. Human and Macaque PBMCs and Cemx174 were grown in RPMI medium with 10% heat-inactivated fetal bovine serum (FBS). Hela P4C5 (45), HEK293T and derivatives were grown in DMEM supplemented with 10% FBS.

Virus production. Viruses were produced by transfection of infectious plasmids into HEK293T cells and co-cultivation with Cemx174 cells. Mutated *gag* and *pol*-optimized DNA sequences were chemically synthesized (GeneScript) then cloned into plasmid p239SpSp5. Mutated *env*-optimized sequence was cloned in p239SpE3' nef Open. Plasmids p239Sp5 and p239SpE3' nef Open contain respectively the 5' and 3' halves of SIVmac239 infectious clone (54). Each plasmid (10 µg) containing opt or wt version of *gag*, *pol* or *env* genes were digested with SphI (p239Sp5) or SphI + AatII (p239SpE3' nef Open), purified on agarose gel and ligated together with T4 DNA ligase for 48h at 4°C. The total ligation product (40µl) was transfected with 6µl Lipofectamine (Invitrogen) in 8.10^5 HEK293T cells plated the day before on 6-well plates. Two days after transfection, DMEM was replaced by 3ml RPMI 10% FBS containing 5.10^5/ml Cemx174. Syncytia formation in the coculture indicated virus production. Two weeks after transfection, culture supernatant was harvested, centrifuged at low speed and filtrated trough 0.45µm pores. Viral stock were kept at -80°C.

RNA preparation. Primers (Table S1) were designed to amplify 40 overlapping fragments of approximately 500pb long by PCR reaction (Enzyme Phusion, 35 cycles, Ta=60°C, 30 sec. elongation). Primers were also designed to amplify sequences (1min. elongation) from wt or opt sequence of viral genes. A 5' tail containing T7 promoter sequence was included in every forward primer to allow the subsequent *in vitro* transcription reaction. PCR products were purified and used as template for T7 RNA synthesis according to manufacturer's instructions (T7 RiboMA Express, Promega). Resulting RNA were purified using RNeasy mini kit (Qiagen) and concentration was determined by Nanodrop measurement.

IFN-luciferase reporter assays. Expression of IFN-α/β was determined by transient transfection of HEK293 cells with either reporter plasmid pISRE-Luc containing five ISRE enhancer elements upstream of the firefly luciferase gene (Stratagene) or reporter plasmid pIFNβ-Luc containing the firefly luciferase gene under the control of IFNβ promoter (provided by Drs. R. Weil and J. Hiscott). For RNA transfections, Hela P4C5 or HEK293T cells were plated in 24-well plates (2×10^5 per well). After 24 hours, cells were transfected using 1µl Lipofectamine 2000 (Invitrogen) with pISRE-Luc reporter plasmid (250 ng/well) (HEK293T) or pIFNβ-Luc (Hela P4C5), a plasmid harboring a thymidine kinase (Tk) promoter upstream of the renilla luciferase gene (25 ng/well), and 12ng of each RNA fragment (500pb) or 20 ng of wt/opt viral genes. After 20 h, cells were lysed, and the firefly and renilla luciferase activities were measured in cell lysates using the Dual-luciferase Reporter Assay System (Promega) according to manufacturer's instructions. Reporter activity was calculated as a triplicate of the ratio of firefly luciferase activity to reference renilla luciferase activity.

For virus activation analysis during coculture, Hela P4C5 cells were plated in 24-well plates (10^5 cells per well). One day later, cells were transfected using 1µl Fugen (Roche) with pIFNβ-Luc reporter plasmid (200 ng/well) and a plasmid harboring a thymidine kinase (Tk) promoter upstream of the renilla luciferase gene (20 ng/well). One day later infected Cemx174 cells were added to the culture at a concentration of 10^6 cells per ml (Cemx174 / Hela P4C5 ratio = 1:1). The percentage of Cemx174 infection was assayed by SIV p27 staining and flow cytometry.

IFN-I detection. IFN-I secretion was quantified using the reporter cell line HL116, that carries the luciferase gene under the control of the IFN-inducible 6-16 promoter (65) (a kind gift from Sandra Pellegrini, Institut Pasteur, France). HL116 were grown in DMEM supplemented with 10% FBS and HAT (H: 20 µg/mL, T: 20 µg/mL, A: 0.2 µg/mL). HL116 cells (2×10^4) plated 16 h prior the assay in 96-well plate, were incubated for 7 h with the desired culture supernatants or standards containing a titration of human IFNα2a (Immunotools). Cells were then lysed (Luciferase Cell Culture Lysis, 5X Reagent, Promega) and luciferase activity measured using the Luciferase Assay Reagent (Promega). Samples were analyzed using Perkin Elmer Wallac 1420. IFN levels are expressed as equivalent of IFNα2a concentration in

Units/ml.

Flow cytometry staining. Cells were intracellularly stained with anti-SIV Gag p27 (Clone 55-2F12, NIH AIDS Research & Reference Reagent Program, from Dr Niels Pedersen) and anti-human MxA (Dr. O. Haller). Briefly, cells were fixed for 10 min with PBS 4 % paraformaldehyde, washed, permeabilized and stained for 45 min in PBS containing 1% BSA and 0.05% saponin. Isotype-matched mAbs were used as negative controls. Samples were analyzed by flow cytometry using a FacsCalibur (Becton Dickinson) or a FacsCanto II (Becton Dickinson) with FlowJo or FacsDIVA softwares.

Lentiviral transduction. Hela P4C5 cells were transduced with a lentiviral vector (LV) expressing BVDV-Pro and previously described (25). The LV also expresses the puroR gene. Two days after transduction, Hela P4C5 cells were selected in the presence of 1 µg/mL puromycin. Resistant populations grew in few days and were used without further cloning.

WB analysis. One week after transduction and selection, $2x10^6$ Hela P4C5 cells were lysed in PBS-1% Triton X-100 (Sigma-Aldrich) supplemented with protease inhibitors (Roche). Cell lysates were analyzed by SDS-gel electrophoresis using 4-12% NuPAGE gels (Invitrogen). IRF-3 western blot was performed using rabbit anti-IRF3 (Clone FL-425, Santa Cruz). As control, actin specific primary antibody was used.

RESULTS

Codon-optimization of viral RNAs abolishes their ability to induce IFN-I. Optimization of codon usage by introducing host cell synonymous codons is a widely used mean to improve recombinant protein expression for DNA-based (1, 6) or viral vector-based vaccines (40) produced in bacteria, yeast or plants (21). Genes with a codon usage matching the specific cellular tRNAs abundance are the most highly expressed (2, 28). Interestingly, in mammalian cells, only a weak positive correlation is found between optimal codon usage and gene expression levels (16, 52, 53, 64). Since we previously observed that the nucleotide composition of HIV-1 RNA can modulate IFN-I stimulation in human cells (66) we wondered whether codon optimization of various viral sequences would alter their potential immunostimulatory capacity.

We chose several genes of different sizes and functions from different viral species (*gag, pol, env* from HIV-1, *gag* from SIVmac, *hemmagglutinin* and *neuraminidase* from Influenza H5N1 and H1N1, *nucleoprotein* and *spike* from SARS-Coronavirus, *core, E1 and E2* from Hepatitis C virus). These sequences were obtained from both wild type viral cDNAs and commercial synthetic DNAs that were optimized for human codon usage (GeneScript). For each viral sequence, wild type (wt) and human codon optimized (opt) versions were used as template for *in vitro* transcription with T7 RNA polymerase into uncapped and unpolyadenylated RNA fragments. To ensure that no protein was expressed from these RNA, the first nucleotide of the ATG codon was mutated in each construct. The capacity of these RNA fragments to stimulate IFN-I was determined by using a very sensitive method (66). RNA fragments were co-transfected into HEK-293T cells together with a reporter plasmid expressing the luciferase gene under the control of five interferon-stimulated response elements (ISRE-luciferase) (14). IFN-I response was determined by measuring luciferase activity 24 h after transfection (Fig. 1A). In this system, all wt genes (Fig 1A, black bars) significantly induced IFN-I. However, differences in intensity were observed: RNA derived from HIV *pol* gene was the most potent stimulator of IFN-I and RNAs from HCV were the less stimulatory. According to our previous results, human codon optimization of all viral genes reduced their capacity to stimulate IFN-I (Fig. 1A, grey bars), with the exception of HCV core RNA whose

optimization increased the luciferase signal. These data show that, regardless of their origin, RNAs derived from viral genes stimulate IFN-I, while codon optimization decreases this property (Wilcoxon paired unilateral test, $p=10^{-3}$). We then measured IFN-I production by human PBMC upon stimulation with the same RNAs complexed to DOTAP. IFN-I concentration in PBMC supernatants was measured 20h after stimulation using a reporter cell line (65). Human codon optimization also reduced the ability of all genes tested to induce IFN-I in PBMC, albeit to a lesser extent than observed in HEK-293T (Fig. S1).

Codon optimization modifies both codon usage and global nucleotide composition. To evaluate the influence of nucleotide composition in our observations, we performed a Principal Component Factorial Correspondence Analysis on the nucleotide composition of all wt and opt RNA fragments (Fig. 1B). This analysis highlights that wt viral genes are either A or U-rich, with the exception of HCV genes, while most optimized versions are G/C-rich. Comparison of nucleotide composition with luciferase activity showed that A richness of RNAs correlated with IFN-I stimulation (Fig 1B, $R=0.856$, $p=1.4.10^{-7}$).

The most biased regions of SIV genome are the most potent IFN-I stimulators. IFN-I is the principal mediator of antiviral innate immunity and its sustained expression a major difference observed in host responses between pathogenic and non-pathogenic lentiviral infections (24, 30). Viral RNAs with high content of A/U nucleotides are strong stimulators of IFN-I and lentiviral genomes have a particularly A-rich composition. With the aim of designing an attenuated SIV genome based on a reduced capacity to activate IFN-I response, we investigated whether local A-rich regions of SIV genes were stronger stimulators of IFN-I. To analyze the repartition of IFN-stimulating sequences along SIVmac239 genome, we measured the ability of a set of 40 overlapping RNA fragments of approximately 500 bp long, covering the entire genome of SIVmac239, to induce IFN-I *in vitro* in HEK-293T cells (Fig 2A). Activating fragments were mostly found clustered in the *pol* region, the 5' region of *Gag* and 3' region of *env*. We then looked for the repartition of local nucleotide bias on the whole genome. We computed the Chi-square distance between the A/C/G/U frequencies of a sliding window 500 nt wide along SIVmac239 genome and the corresponding frequencies of the entire coding sequences of the macaque genome (Fig. 2B). This analysis shows that the most divergent regions

locate within the three large viral genes (*gag, pol, env*), while overlapping coding regions and cis-active regulatory sequences are not biased. A-richness contributes to the majority of the observed divergence, with the exception of a short central portion of *env* where C-poverty is the major contributor. The ability of each fragment to induce IFN-I (Fig. 2A) and the divergence to host in nucleotide composition (Fig. 2B) correlated significantly (R=0.36, p=0.02), as expected from our previous work on HIV-1 genome (66).

Design of SIVmac239 optimized sequences. Based on this observation, we hypothesized that a synthetic SIV with a nucleotide composition optimized to be closer to its host would have a reduced capacity to stimulate IFN-I, and might thus represent a model of attenuated macaque lentiviral infection. SIVmac239 is highly pathogenic in *rhesus* macaques (62). To attenuate its virulence, we optimized SIVmac239 genomic sequence by changing its nucleotide content towards macaque average composition. We used the original SIVMM239 sequence available at GenBank (accession N° M33262). Our strategy was to reduce as much as possible the major peaks of nucleotide divergence, which are also the most potent stimulators of IFN-I, while preserving the amino acid sequence. Based on previous work showing that codon optimization of HIV-1 genome may lead to defective virus (33), we excluded all known or putative regulatory sequences from optimized regions. From the literature available on SIVmac239 genome, and by homology with known regulatory sequences of HIV-1 genome, we excluded from the optimized regions the sequences corresponding to the genomic features described in Table 1. Overlapping ORFs that were impossible to optimize while maintaining amino acid sequences were unmodified. This led us to select three separate regions on SIVmac239 genome to be optimized to the macaque average nucleotide composition. These regions are located within *gag*, *pol* and *env* genes (Fig 2A, white areas). Each region was divided in short domains of nucleotide divergence and synonymous mutations were locally computed to reduce divergence. This method was applied iteratively on all divergent peaks until the divergence in each region was below a pre-selected threshold. We tested different thresholds inside each region and selected for synthesis and further experiments those leading to a total of 348 changes divided in 99 synonymous mutations in *gag* region, 70 in *pol* region and 189 in *env* region. Optimized sequences are available on GenBank (SIVopt1 accession number KJ152770 and

SIVopt2 accession number KJ152771) and the statistics of mutations are presented in Table S1. Mutated fragments were chemically synthesized (GeneScript) and sub-cloned in replacement of wt counterparts into p239Sp5 and p239SpE3' nef Open plasmids that contain respectively the 5' and 3' halves of SIVmac239 infectious clone (54).

Optimization of *pol* or *gag/pol* regions leads to replicating viruses. Wt and optimized SIV were produced after transfection of reconstituted proviral DNA into HEK293 cells and co-cultivation with Cemx174 cells. All combinations of wt and optimized *gag*, *pol* and *env* fragments were tested. Replicating viruses could be obtained with genomes containing an optimized *pol* (SIVopt1) or optimized *gag* and *pol* (SIVopt2) (Fig. 3A). No replicating virus was obtained with optimized *env* or optimized *gag* alone. Viral stocks of SIVopt1, SIVopt2 and SIVwt were produced and their replication kinetics compared by infecting Cemx174 cells and macaque PBMC with different viral doses. Viral replication was assessed by p27 immunostaining and flow cytometry (Fig 3BC). SIVwt and SIVopt1 displayed the same replicative kinetics in both Cemx174 and PBMC, while SIVopt2 was greatly reduced in Cemx174 and did not replicate in PBMC. In Cemx174 cells, a 100-fold higher viral input was needed for SIVopt2 to reach viral peak on day 8 after infection, compared to SIVwt and SIVopt1. Because SIVopt2 had a reduced infectivity, it was not further evaluated. Conversely, the 70 synonymous mutations introduced in *pol* gene did not alter the replicative capacity of SIVopt1.

In the perspective of vaccine design, the stability of attenuated synthetic viruses is of uttermost importance to prevent reversion. We therefore evaluated the stability of optimized virus genomes. SIVopt1, SIVopt2 and SIVwt were passaged 10 times on Cemx174 cells (cultured for 10 weeks, which represents approximately 30 replication cycles). The genomes of resulting viruses were fully sequenced and their consensus sequences were compared to their original sequences before passaging. A few mutations occurred in the three passaged viruses (listed in table S2). One mutation common to the three passaged viruses was found in the 5'-LTR. Other mutations were found in different regions of the genome (Table S2). However and interestingly, no reversion was observed in the optimized regions of optimized viruses, indicating that these artificially introduced mutations are not less stable than wild type

SIVmac239 sequence. The lack of reversion remains to be determined in primary cells such as macaque PBMC.

SIVopt1 has a reduced capacity to stimulate IFN-I response in human PBMC. To compare the capacity of SIVopt1 and SIVwt to stimulate IFN-I response in human PBMC, we measured the induction of MxA, an interferon stimulated gene (23). As expected from the previous experiment, SIVwt and SIVopt1 grew at similar rates in human PBMC. Percentage of SIV Gag p27 positive infected cells and cellular MxA induction were assessed by immunostaining and flow cytometry analysis at different time points. To standardize with viral growth, MxA induction was calculated relative to the percentage of Gag p27 positive cells for both viruses (Fig 4A). When comparing 7 human PBMC donors, we found that SIVopt1 had a slightly reduced capacity to stimulate MxA expression as compared to SIVwt, (Fig 4B, paired Wilcoxon test, $p= 3.10^{-5}$). This effect is not due to a difference in replication between both viruses (paired Wilcoxon test p=0.07 calculated for 7 donors at each time point on more than 150 pairs), and these results are not affected by the post-replication peak cytotoxic effect of viruses ($p=4.10^{-4}$ for the reduction of MxA expression, and p=0.12 for the difference in replicativity when removing all data points beyond 12 days, same tests). Viral replication and MxA induction profiles for individual human and macaques PBMC are available in Supp. Mat. (Supp. Fig. S2 and S3). This effect indicates a reduced ability of SIVopt1 to induce IFN-I, and suggests that a molecular sensor involved in innate sensing and IFN-I response is able to detect the difference between SIVwt and SIVopt1, which consists in 70 synonymous mutations within the *pol* region in genomic RNA.

SIVopt1-infected cells induce a reduced IRF3-dependent IFNβ response. To investigate the role of cellular sensors in differential detection of SIVwt and SIVopt1, we used a co-culture system consisting of infected Cemx174 lymphocytes and Hela-derived reporter cells (P4C5 cells expressing SIV receptors CD4 and CCR5 (3)) transfected with a luciferase gene under the control of IFNβ promoter (IFNβ-luc). This system allows investigating the cytoplasmic sensing of viral RNA after cell-to-cell contact, which is a major mode of virus spreading in lentiviral infections (58). SIVwt and SIVopt1 differ only by their genomic RNA. Viral RNAs are detected by endosomal TLRs (TLR3/7/8) and cytosolic receptors (RIG-I, MDA5) (11). TLR7 pathway is not functional in Hela cells, as confirmed by the absence of IFNβ

stimulation by CL097, a TLR7 ligand (not shown). Therefore, this experimental system allows addressing whether TLR7-independent pathways (37) can detect SIV-infected lymphocytes and if this detection is sensitive to nucleotide composition of the viral genome.

We first controlled whether our optimization algorithm decreased the capacity of SIV RNA to induce IFN-I in this cell line, as did commercial human codon optimization of most viral RNA in HEK293 cells (Fig 1A). We produced *in vitro* transcribed RNA from the *Pol* region of SIVwt and SIVopt1 and evaluated their ability to trigger IFN-I response after transfection in Hela P4C5 target cells (Fig. 5A, P4C5ctrl). As expected, the optimized version of *pol* RNA induced a reduced IFNβ response as compared to wt RNA.

We then evaluated the induction of IFNβ-luciferase when Cemx174 cells were cocultured with P4C5 reporter cells. Cemx174 cells infected with SIVwt or SIVopt1 activated the IFNβ promoter. The intensity of luciferase induction was directly related to the percentage of SIV-infected cells in the coculture (Fig. 5B, SIVwt line R^2=0,87, SIVopt1 line R^2=0,86). Interestingly, SIVopt1 displayed a lower rate of IFNβ stimulation than SIVwt, as illustrated by the difference in the slopes of the regression lines (Fig. 5B, t test, p = 0.001). Thus, SIV triggers IFNβ when transmitted from cell-to-cell and SIVopt1 is attenuated in this capacity.

During infection, binding of certain viral RNAs to cytosolic receptors RIG-I or MDA5 leads to conformational changes that expose their CARD-like domains to MAVS, inducing down-stream signaling for IFNβ transcription through IRF3 (31, 32, 35, 42, 47, 60, 72, 73). To characterize the pathway that mediates recognition of SIV-infected cells, we stably expressed in Hela P4C5 reporter cells BVDV-Pro, a molecule that degrades IRF-3 (25). Western blot analysis confirmed that IRF3 levels were reduced in Hela P4C5 cells expressing BVDV-Pro protease (Fig 5A). As expected (60), the absence of IRF3 abrogated the activation of IFNβ upon RNA transfection (Fig 5A P4C5 BVDV-Pro). IRF3 was also involved in cellular signaling induced by SIV-infected Cemx174 cells (Fig 5C, reduction of approx. 50% for coculture with SIVwt infected cells in IRF3-silenced cells). This observation indicates that cytosolic sensors play a role in the detection of SIV-infected cells.

DISCUSSION

The redundancy of genetic code leads to a specific codon usage bias for each species that refers to differences in occurrence frequency of synonymous codons. The origin of nucleotide and codon bias has been extensively debated and can be explained by mutational and selective mechanisms (52). Consequently, RNA viruses, which have co-evolved with their hosts, can still express their proteins in host cells despite a very biased RNA nucleotide composition imposed by their genomic constraints. However, such biased nucleotide composition is likely detected as non-self in infected host cells. Innate antiviral immune response is initiated by the recognition of non-self Pathogen-Associated Molecular Patterns (PAMPs) by Pattern Recognition Receptors (PRRs). Viral nucleotide sequences are recognized by toll-like membrane receptors (TLRs) and RIG-I-like cytosolic receptors (RLRs) (29), which upon activation, trigger signaling cascades that converge to induce IFN-α/β secretion and amplification of the antiviral response (29). In this work, we show that nucleotide optimization of viral RNA sequences affects their recognition by PRRs. Indeed, once optimized to human codon usage, viral RNAs virtually lose their capacity to induce IFN-α/β expression in human cells. The A/U richness of viral RNAs correlated with their capacity to stimulate IFN-α/β whereas optimized G/C-rich sequences were attenuated in this capacity. As a practical consequence of this observation, the use of optimized viral genes in genetic vaccines strategies could be re-evaluated considering the potential immunostimulatory effect of viral sequences.

We further identified local A-rich regions of SIV genes as strong stimulators of IFN-I, as we previously showed for the HIV-1 genome (66). With the aim of designing an attenuated SIV, we generated 70 synonymous mutations that unbalanced the *pol* gene towards macaque average nucleotide composition. The resulting SIVopt1 was growing as efficiently as SIVwt in a T-cell line and PBMC, showing that it is possible to proceed to large-scale modification of a lentivirus genome without interfering with its *in vitro* replicative capacity. This also suggests the absence of any unknown regulatory sequences in the optimized *pol* region. Conversely, the delay in replication kinetics observed with SIVopt2, which contains 99 additional mutations in *gag*, and the failure to obtain a replicative virus with 189 mutations in *env* may be due to the existence of regulatory sequences or secondary structures necessary for viral

463 replication. This could be experimentally investigated by designing less optimized
464 versions of these genomic regions. More intriguingly, we were unable to obtain a
465 *gag*-only optimized virus, whereas a virus optimized both in *gag* and *pol* was
466 replicative. We cannot rule out the possibility that a *gag*-optimized replicative virus
467 may be obtained after additional attempts. However, since nucleotide optimization
468 alter the flexibility of RNA molecules (33), it is likely that high local structural
469 constraints induced by optimizing the *gag* region would inhibit certain viral replication
470 functions, such as packaging or reverse transcription. Changing additionally the
471 nucleotide composition of the *pol* region might have induced compensatory effect in
472 the global flexibility of the genomic RNA molecule, which would restore viral
473 replicative function, although with a reduced efficacy. In a previous work (33), codon
474 optimization of A-rich sequences within HIV-1 *pol* led to a 100-fold decrease of
475 infectivity that was explained by enhanced dimer stability of the viral RNA genome
476 and reduction of viral cDNA synthesis. A report published while we were preparing
477 this manuscript (46), describes the use of synthetic attenuated virus engineering
478 (SAVE) approach (15) to generate optimized and deoptimized HIV-1. Although very
479 few synonymous mutations (less than 40) were introduced, deoptimized viruses had
480 significantly lower viral replication capacity and reverted to wild-type virulence after
481 serial passages. On the contrary, optimized viruses remained stable but were not
482 attenuated (46). Due to the low number of mutations introduced, the global A-
483 richness of these genomes was probably minimally affected.

484 We previously highlighted a correlation between the nucleotide composition of
485 different HIV-1 subtypes and their pathogenicity (66). We proposed that biased RNA
486 would drive a high IFN-I level in persistently infected cells and might be involved in
487 the over-activation of the immune system during chronic pathogenic lentiviral
488 infection. Here, we studied the consequences of optimizing the nucleotide
489 composition of SIVopt1 genome on IFN-I stimulation. We observed that both
490 transfection of optimized RNA and coculture with SIVopt1-infected cells triggered a
491 lower level of cellular IFN-I response than SIVwt. In non-immune cells, this suggests
492 the existence of a cellular sensor able to recognize RNAs according to their
493 nucleotide composition. IRF3 is the transcription factor leading to IFN-I upregulation
494 after triggering of cytosolic viral sensors (18). We demonstrated that IRF3 is involved
495 in IFN-I induction upon detection and differentiation of opt and wt SIV RNAs after

transfection or coculture with infected cells. Upstream of IRF3, the cytosolic RLR (RIG-I-like receptors) helicases RIG-I and MDA5 directly recognize multiple and distinct forms of intracellular dsRNA. For instance, putative RIG-I ligands include short 5'-triphosphate RNAs with double stranded structures (31, 35, 42, 55, 59). The RNAs used in our experiment were *in vitro* transcribed by T7 RNA polymerase, and, in consequence were not capped nor- polyadenylated. Thus, RIG-I and MDA5 could be involved in their recognition. Addressing precisely their role will require further investigation.

Differences observed between the recognition of SIVopt1 and SIVwt also raise the question of the nature of the viral ligand involved. Our results strongly suggest a role for nucleotide bias recognition in our system. However, other motifs can change with nucleotide modification. HIV-1 genomic RNA is highly structured (70) and this is likely the case for SIV genome. Changes in RNA nucleotide composition may be either directly detected by RLRs or through the modification of viral RNA structure. Codon optimization of HIV-1 *pol* in a previous study (33) led to destabilization of RNA dimerization and impairment of reverse transcriptase processing. We cannot exclude that nucleotide modifications in SIVopt1 affected recognition by innate sensors without impacting viral replication. This could be the case by increased production of defective particles or accumulation of nucleic acid replication intermediates.

In vivo studies will be required to analyze the fitness of SIVopt1 and SIVopt2 in macaque, and the influence of their nucleotide composition on AIDS induction. Live attenuated SIV vaccines are highly protective in the macaque AIDS model (34), but their transposition to human application is not considered for safety concerns. Indeed, attenuated SIV persists and reversion to virulence eventually occurs (71). In this context, original vaccine strategies are desired to reproduce the protection efficacy of attenuated viruses without retaining their associated risks. The mutations artificially introduced in SIVopt1 (70 mutations) and SIVopt2 (169 mutations) did not revert after 10 weeks of culture on Cemx174 cells. Although genetic stability has to be further documented in primary cells and in vivo, this first promising observation indicates that our optimization strategy did not alter viral fitness, at least in CEM cells. Using our optimization system, it is thus possible to increase the number of synonymous mutations with the aim to finely tuning viral replication and/or IFN-I response, while preventing reversion of attenuated virus. SIVopt1 has a reduced

ability to induce IFN-I, but shows *wt* replicative kinetics *in vitro*. Although type I IFN response is obviously necessary for a vaccine to shape adaptive immune responses and memory, an excessive response might be deleterious for the pathogenicity and the stability of a live attenuated SIV/HIV vaccine. A considerable amount of work will be required to analyse if these properties are conserved *in vivo* and whether they impact the pathogenicity of infection. In conclusion, our work demonstrates the possibility to rationally change viral nucleotide composition to design replicative and genetically stable lentiviruses with attenuated pathogenic potentials.

ACKNOWLEDGMENTS

This work was supported by the Centre National de la Recherche Scientifique (CNRS-PEPS), the Agence Nationale de Recherche contre le Sida (ANRS-2010), and the Institut Pasteur. NV was a fellow from Ecole Normale Supérieure (ENS-Lyon, France). The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: p239SpE3' nef Open and p239SpSp5' from Dr. Ronald Desrosiers; SIVmac p27 mAb (55-2F12) from Dr. Niels Pedersen.

REFERENCES

1. **Abdulhaqq, S. A., and D. B. Weiner.** 2008. DNA vaccines: developing new strategies to enhance immune responses. Immunol Res **42:**219-232.
2. **Akashi, H., and A. Eyre-Walker.** 1998. Translational selection and molecular evolution. Curr Opin Genet Dev **8:**688-693.
3. **Amara, A., S. L. Gall, O. Schwartz, J. Salamero, M. Montes, P. Loetscher, M. Baggiolini, J. L. Virelizier, and F. Arenzana-Seisdedos.** 1997. HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. J Exp Med **186:**139-146.
4. **Arya, S. K., and R. C. Gallo.** 1988. Human immunodeficiency virus type 2 long terminal repeat: analysis of regulatory elements. Proc Natl Acad Sci U S A **85:**9753-9757.
5. **Auewarakul, P.** 2005. Composition bias and genome polarity of RNA viruses. Virus Res **109:**33-37.
6. **Barouch, D. H.** 2006. Rational design of gene-based vaccines. J Pathol **208:**283-289.
7. **Baum, A., and A. Garcia-Sastre.** 2011. Differential recognition of viral RNA by RIG-I. Virulence **2:**166-169.
8. **Berkhout, B., A. Grigoriev, M. Bakker, and V. V. Lukashov.** 2002. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS Res Hum Retroviruses **18:**133-141.
9. **Berkhout, B., and F. J. van Hemert.** 1994. The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. Nucleic Acids Res **22:**1705-1711.
10. **Bishop, K. N., R. K. Holmes, A. M. Sheehy, and M. H. Malim.** 2004. APOBEC-mediated editing of viral RNA. Science **305:**645.
11. **Brennan, K., and A. G. Bowie.** 2010. Activation of host pattern recognition receptors by viruses. Curr Opin Microbiol **13:**503-507.
12. **Browning, M. T., F. Mustafa, R. D. Schmidt, K. A. Lew, and T. A. Rizvi.** 2003. Sequences within the gag gene of feline immunodeficiency virus (FIV) are important for efficient RNA encapsidation. Virus Res **93:**199-209.
13. **Burns, C. C., J. Shaw, R. Campagnoli, J. Jorba, A. Vincent, J. Quay, and O. Kew.** 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol **80:**3259-3272.
14. **Caignard, G., A. V. Komarova, M. Bourai, T. Mourez, Y. Jacob, L. M. Jones, F. Rozenberg, A. Vabret, F. Freymuth, F. Tangy, and P. O. Vidalain.** 2009. Differential regulation of type I interferon and epidermal growth factor pathways by a human Respirovirus virulence factor. PLoS Pathog **5:**e1000587.
15. **Coleman, J. R., D. Papamichail, S. Skiena, B. Futcher, E. Wimmer, and S. Mueller.** 2008. Virus attenuation by genome-scale changes in codon pair bias. Science **320:**1784-1787.
16. **Comeron, J. M.** 2004. Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. Genetics **167:**1293-1304.
17. **Deforche, K., R. Camacho, K. V. Laethem, B. Shapiro, Y. Moreau, A. Rambaut, A. M. Vandamme, and P. Lemey.** 2007. Estimating the relative contribution of

dNTP pool imbalance and APOBEC3G/3F editing to HIV evolution in vivo. J Comput Biol **14:**1105-1114.

18. **Dixit, E., and J. C. Kagan.** 2013. Intracellular pathogen detection by RIG-I-like receptors. Advances in immunology **117:**99-125.
19. **Gale, M., Jr., and M. G. Katze.** 1998. Molecular mechanisms of interferon resistance mediated by viral-directed inhibition of PKR, the interferon-induced protein kinase. Pharmacol Ther **78:**29-46.
20. **Grantham, R., C. Gautier, M. Gouy, R. Mercier, and A. Pave.** 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res **8:**r49-r62.
21. **Gustafsson, C., S. Govindarajan, and J. Minshull.** 2004. Codon bias and heterologous protein expression. Trends Biotechnol **22:**346-353.
22. **Gutman, G. A., and G. W. Hatfield.** 1989. Nonrandom utilization of codon pairs in Escherichia coli. Proc Natl Acad Sci U S A **86:**3699-3703.
23. **Haller, O., S. Stertz, and G. Kochs.** 2007. The Mx GTPase family of interferon-induced antiviral proteins. Microbes Infect **9:**1636-1643.
24. **Harris, L. D., B. Tabb, D. L. Sodora, M. Paiardini, N. R. Klatt, D. C. Douek, G. Silvestri, M. Muller-Trutwin, I. Vasile-Pandrea, C. Apetrei, V. Hirsch, J. Lifson, J. M. Brenchley, and J. D. Estes.** 2010. Downregulation of robust acute type I interferon responses distinguishes nonpathogenic simian immunodeficiency virus (SIV) infection of natural hosts from pathogenic SIV infection of rhesus macaques. J Virol **84:**7886-7891.
25. **Hilton, L., K. Moganeradj, G. Zhang, Y. H. Chen, R. E. Randall, J. W. McCauley, and S. Goodbourn.** 2006. The NPro product of bovine viral diarrhea virus inhibits DNA binding by interferon regulatory factor 3 and targets it for proteasomal degradation. J Virol **80:**11723-11732.
26. **Holmes, E. C.** 2011. What Does Virus Evolution Tell Us about Virus Origins? J Virol **85:**5247-5251.
27. **Hu, C., D. T. Saenz, H. J. Fadel, W. Walker, M. Peretz, and E. M. Poeschla.** 2010. The HIV-1 central polypurine tract functions as a second line of defense against APOBEC3G/F. J Virol **84:**11981-11993.
28. **Ikemura, T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol **2:**13-34.
29. **Iwasaki, A.** 2012. A virological view of innate immune recognition. Annual review of microbiology **66:**177-196.
30. **Jacquelin, B., V. Mayau, B. Targat, A. S. Liovat, D. Kunkel, G. Petitjean, M. A. Dillies, P. Roques, C. Butor, G. Silvestri, L. D. Giavedoni, P. Lebon, F. Barre-Sinoussi, A. Benecke, and M. C. Muller-Trutwin.** 2009. Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. J Clin Invest **119:**3544-3555.
31. **Jiang, F., A. Ramanathan, M. T. Miller, G. Q. Tang, M. Gale, Jr., S. S. Patel, and J. Marcotrigiano.** 2011. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature **479:**423-427.
32. **Kawai, T., K. Takahashi, S. Sato, C. Coban, H. Kumar, H. Kato, K. J. Ishii, O. Takeuchi, and S. Akira.** 2005. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. Nat Immunol **6:**981-988.
33. **Keating, C. P., M. K. Hill, D. J. Hawkes, R. P. Smyth, C. Isel, S. Y. Le, A. C. Palmenberg, J. A. Marshall, R. Marquet, G. J. Nabel, and J. Mak.** 2009. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. Nucleic Acids Res **37:**945-956.

34. **Koff, W. C., P. R. Johnson, D. I. Watkins, D. R. Burton, J. D. Lifson, K. J. Hasenkrug, A. B. McDermott, A. Schultz, T. J. Zamb, R. Boyle, and R. C. Desrosiers.** 2006. HIV vaccine design: insights from live attenuated SIV vaccines. Nat Immunol **7:**19-23.
35. **Kowalinski, E., T. Lunardi, A. A. McCarthy, J. Louber, J. Brunel, B. Grigorov, D. Gerlier, and S. Cusack.** 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. Cell **147:**423-435.
36. **Le, S. Y., M. H. Malim, B. R. Cullen, and J. V. Maizel.** 1990. A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. Nucleic Acids Res **18:**1613-1623.
37. **Lepelley, A., S. Louis, M. Sourisseau, H. K. Law, J. Pothlichet, C. Schilte, L. Chaperot, J. Plumas, R. E. Randall, M. Si-Tahar, F. Mammano, M. L. Albert, and O. Schwartz.** 2011. Innate sensing of HIV-infected cells. PLoS Pathog **7:**e1001284.
38. **Lesnik, E. A., and S. M. Freier.** 1995. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. Biochemistry **34:**10807-10815.
39. **Li, M., E. Kao, X. Gao, H. Sandig, K. Limmer, M. Pavon-Eternod, T. E. Jones, S. Landry, T. Pan, M. D. Weitzman, and M. David.** 2012. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature **491:**125-128.
40. **Li, S., E. Locke, J. Bruder, D. Clarke, D. L. Doolan, M. J. Havenga, A. V. Hill, P. Liljestrom, T. P. Monath, H. Y. Naim, C. Ockenhouse, D. C. Tang, K. R. Van Kampen, J. F. Viret, F. Zavala, and F. Dubovsky.** 2007. Viral vectors for malaria vaccine development. Vaccine **25:**2567-2574.
41. **Liu, Y., E. Wimmer, and A. V. Paul.** 2009. Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochim Biophys Acta **1789:**495-517.
42. **Luo, D., S. C. Ding, A. Vela, A. Kohlway, B. D. Lindenbach, and A. M. Pyle.** 2011. Structural insights into RNA recognition by RIG-I. Cell **147:**409-422.
43. **Malathi, K., B. Dong, M. Gale, Jr., and R. H. Silverman.** 2007. Small self-RNA generated by RNase L amplifies antiviral innate immunity. Nature **448:**816-819.
44. **Mangeot, P. E., D. Negre, B. Dubois, A. J. Winter, P. Leissner, M. Mehtali, D. Kaiserlian, F. L. Cosset, and J. L. Darlix.** 2000. Development of minimal lentivirus vectors derived from simian immunodeficiency virus (SIVmac251) and their use for gene transfer into human dendritic cells. J Virol **74:**8307-8315.
45. **Marcheschi, R. J., D. W. Staple, and S. E. Butcher.** 2007. Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem-loop. J Mol Biol **373:**652-663.
46. **Martrus, G., M. Nevot, C. Andres, B. Clotet, and M. A. Martinez.** 2013. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology **10:**78.
47. **Meylan, E., J. Curran, K. Hofmann, D. Moradpour, M. Binder, R. Bartenschlager, and J. Tschopp.** 2005. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. Nature **437:**1167-1172.
48. **Mueller, S., J. R. Coleman, D. Papamichail, C. B. Ward, A. Nimnual, B. Futcher, S. Skiena, and E. Wimmer.** 2010. Live attenuated influenza virus vaccines by computer-aided rational design. Nat Biotechnol **28:**723-726.
49. **Mueller, S., D. Papamichail, J. R. Coleman, S. Skiena, and E. Wimmer.** 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon

689 deoptimization causes attenuation of viral virulence by lowering specific
690 infectivity. J Virol **80:**9687-9696.
691 50. **Nurmemmedov, E., M. Castelnovo, C. E. Catalano, and A. Evilevitch.** 2007.
692 Biophysics of viral infectivity: matching genome length with capsid size. Q Rev
693 Biophys **40:**327-356.
694 51. **Olsen, H. S., S. Beidas, P. Dillon, C. A. Rosen, and A. W. Cochrane.** 1991.
695 Mutational analysis of the HIV-1 Rev protein and its target sequence, the Rev
696 responsive element. J Acquir Immune Defic Syndr **4:**558-567.
697 52. **Plotkin, J. B., and G. Kudla.** 2011. Synonymous but not the same: the causes and
698 consequences of codon bias. Nat Rev Genet **12:**32-42.
699 53. **Plotkin, J. B., H. Robins, and A. J. Levine.** 2004. Tissue-specific codon usage and
700 the expression of human genes. Proc Natl Acad Sci U S A **101:**12588-12591.
701 54. **Regier, D. A., and R. C. Desrosiers.** 1990. The complete nucleotide sequence of a
702 pathogenic molecular clone of simian immunodeficiency virus. AIDS Res Hum
703 Retroviruses **6:**1221-1231.
704 55. **Rehwinkel, J., and C. Reis e Sousa.** 2010. RIGorous detection: exposing virus
705 through RNA sensing. Science **327:**284-286.
706 56. **Reinhart, T. A., M. J. Rogan, and A. T. Haase.** 1996. RNA splice site utilization by
707 simian immunodeficiency viruses derived from sooty mangabey monkeys.
708 Virology **224:**338-344.
709 57. **Saito, T., D. M. Owen, F. Jiang, J. Marcotrigiano, and M. Gale, Jr.** 2008. Innate
710 immunity induced by composition-dependent RIG-I recognition of hepatitis C
711 virus RNA. Nature **454:**523-527.
712 58. **Sattentau, Q.** 2008. Avoiding the void: cell-to-cell spread of human viruses.
713 Nature reviews. Microbiology **6:**815-826.
714 59. **Schlee, M., E. Hartmann, C. Coch, V. Wimmenauer, M. Janke, W. Barchet, and**
715 **G. Hartmann.** 2009. Approaching the RNA ligand for RIG-I? Immunol Rev
716 **227:**66-74.
717 60. **Seth, R. B., L. Sun, C. K. Ea, and Z. J. Chen.** 2005. Identification and
718 characterization of MAVS, a mitochondrial antiviral signaling protein that
719 activates NF-kappaB and IRF 3. Cell **122:**669-682.
720 61. **Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim.** 2002. Isolation of a
721 human gene that inhibits HIV-1 infection and is suppressed by the viral Vif
722 protein. Nature **418:**646-650.
723 62. **Silvestri, G.** 2008. AIDS pathogenesis: a tale of two monkeys. Journal of medical
724 primatology **37 Suppl 2:**6-12.
725 63. **Unger, R. E., M. W. Stout, and P. A. Luciw.** 1991. Simian immunodeficiency virus
726 (SIVmac) exhibits complex splicing for tat, rev, and env mRNA. Virology **182:**177-
727 185.
728 64. **Urrutia, A. O., and L. D. Hurst.** 2003. The signature of selection mediated by
729 expression on human genes. Genome Res **13:**2260-2264.
730 65. **Uze, G., S. Di Marco, E. Mouchel-Vielh, D. Monneron, M. T. Bandu, M. A.**
731 **Horisberger, A. Dorques, G. Lutfalla, and K. E. Mogensen.** 1994. Domains of
732 interaction between alpha interferon and its receptor components. J Mol Biol
733 **243:**245-257.
734 66. **Vabret, N., M. Bailly-Bechet, V. Najburg, M. Muller-Trutwin, B. Verrier, and F.**
735 **Tangy.** 2012. The biased nucleotide composition of HIV-1 triggers type I
736 interferon response and correlates with subtype D increased pathogenicity. PLoS
737 One **7:**e33502.

67. **van der Kuyl, A. C., and B. Berkhout.** 2012. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. Retrovirology **9:**92.
68. **van Hemert, F. J., A. C. van der Kuyl, and B. Berkhout.** 2013. The A-nucleotide preference of HIV-1 in the context of its structured RNA genome. RNA biology **10:**211-215.
69. **Victoria, J. G., and W. E. Robinson, Jr.** 2005. Disruption of the putative splice acceptor site for SIV(mac239)Vif reveals tight control of SIV splicing and impaired replication in Vif non-permissive cells. Virology **338:**281-291.
70. **Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, Jr., R. Swanstrom, C. L. Burch, and K. M. Weeks.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-716.
71. **Whitney, J. B., and M. A. Wainberg.** 2007. Recovery of fitness of a live attenuated simian immunodeficiency virus through compensation in both the coding and non-coding regions of the viral genome. Retrovirology **4:**44.
72. **Wu, B., A. Peisley, C. Richards, H. Yao, X. Zeng, C. Lin, F. Chu, T. Walz, and S. Hur.** 2013. Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. Cell **152:**276-289.
73. **Xu, L. G., Y. Y. Wang, K. J. Han, L. Y. Li, Z. Zhai, and H. B. Shu.** 2005. VISA is an adapter protein required for virus-triggered IFN-beta signaling. Mol Cell **19:**727-740.

Table 1. Genomic regulatory sequences excluded from algorithm optimization in SIVmac239

Sequences excluded	Function during viral replication	Reference
Long Terminal Repeats	Regulation of transcription, translation and genome packaging	(4)
300 first nucleotides of Gag	Genome packaging in FIV or SIV-derived vector	(12) (44)
Gag-Pol frameshift	Translation of Pol gene	(45)
5' A-rich sequence of Pol	Important for cDNA synthesis during HIV-1 reverse transcription	(33)
Central PolyPurine Tract and Central terminal Sequence	Regulation of Reverse transcription	(44)
Rev Responsive Element	RNA subcellular trafficking	(36) (51)
Sequences containing donor and acceptor splicing sites	RNA maturation	(63) (69) (56)

FIGURE LEGENDS

Figure 1. Codon optimization of viral RNAs modulates their ability to stimulate IFN-I

A. Measure of Luciferase expression reflecting IFN-I stimulation after transfection of RNA fragments transcribed in vitro from wild type (wt) or optimized (opt) viral sequences from different origins. RLU: Relative Luciferase Units. In capital letter: virus names, in small letters: viral gene names. **B.** Plot of the two principal components (96.2% of the total variance) of the correspondence analysis of the nucleotide composition of viral sequences. The position of viral sequences depends on their composition (black dots correspond to wt sequences and grey dots to opt sequences). Projection of the A, C, G, U components indicate which nucleotide is enriched in each direction on the graph. Projection of the Luciferase expression (as in Fig 1A) is also shown (Luc), and is highly correlated with A nucleotide frequency in the strains (Pearson correlation test, R=0.85, p = 1.4 10-7), as can visually be assessed by the small angle between the A and Luc arrows.

Figure 2. The nucleotide divergence of SIVmac239 RNA with *Rhesus* macaque correlates with the ability to stimulate IFN-I

A. Luciferase reporter activity determined in HEK-293T cells co-transfected with SIVmac239 RNA fragments and pISRE-Luc reporter plasmid. **B.** Repartition of nucleotide divergence along SIVmac239 genome. The black line shows the nucleotide divergence of SIVmac239 compared to macaque rhesus genome, as measured by a Chi-square distance in a 500 bp sliding window along the SIVmac239 sequence. Individual contributions of A and C nucleotides to this divergence are shown respectively in black and grey dotted lines. Nucleotide divergence correlates with luciferase activity (R=0.36, p<0.02). White area: regions modified by our optimization algorithm – the limits of the region are reported along the x-axis. SIVmac schematic genome map is represented below the figure.

Figure 3. Replicative capacity of SIVopt1 and SIVopt2

A. Schematic representation of the SIVmac239-based constructs used in this study. The codon-modified regions are indicated in black. **B.** Replicative kinetics of SIVopt1 and SIVopt2 as compared to SIVwt in Cemx174. Nucleotide modification does not impair viral replication kinetics of SIVopt1. However SIVopt2 display a 2-log decrease in viral infectivity. **C.** Replication kinetics of SIVwt and SIVopt1 in macaque PBMC. The amount of viral input used to infect cells is indicated (p27 ng).

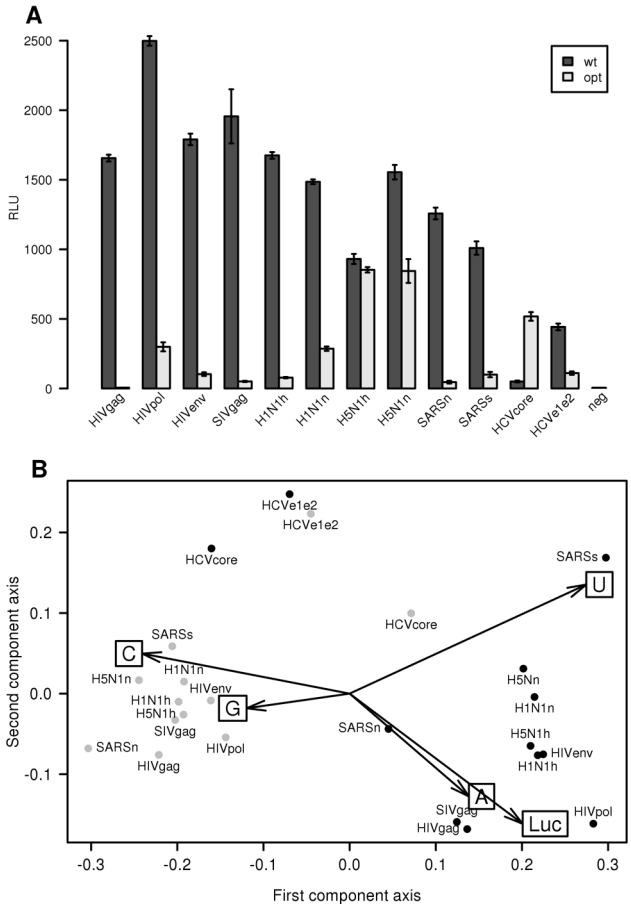
Figure 4. Differential sensing of SIV replication by PBMC

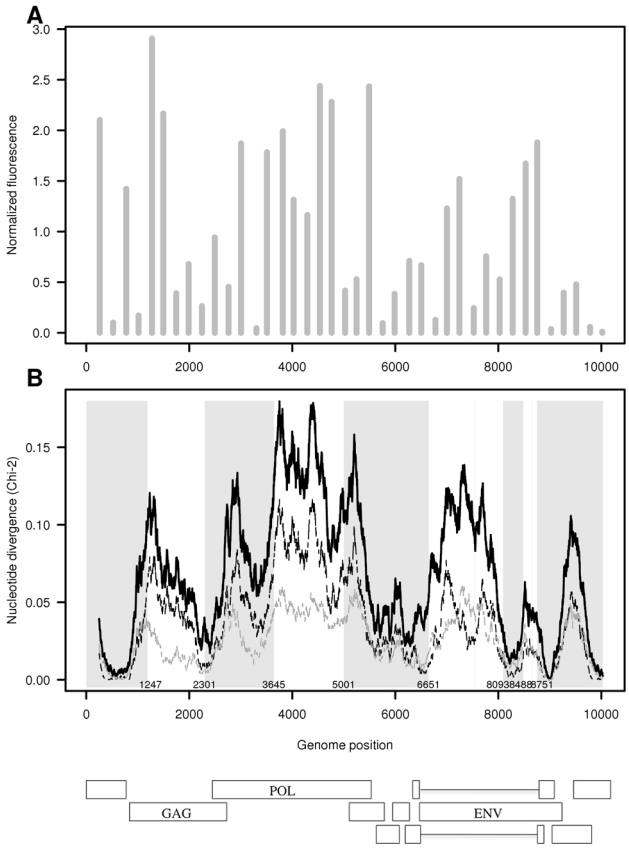
A. Example for one experiment. Human PBMC were infected by SIVwt or SIVopt1. IFN-I Induction was measured with the profile of intracellular MxA expression, an interferon stimulated gene. SIVopt1 shows a reduced ability to induce MxA expression as compared to SIVwt. Rebound at day 14 is likely caused by cell mortality. **B.** Boxplot of SIVwt and SIVopt1 for every measured MxA expression point from 7 different human PBMC donors infected by SIVopt1 and SIVwt. Each PBMC donor sample was divided in several cultures infected by different viral input concentrations (always paired between SIVwt and SIVopt1). MxA expression was measured at regular time points following infection (fig. S3). SIVopt1 globally stimulated less MxA synthesis than SIVwt (paired Wilcoxon test, $p= 3. 10^{-5}$ on more than 150 pairs).

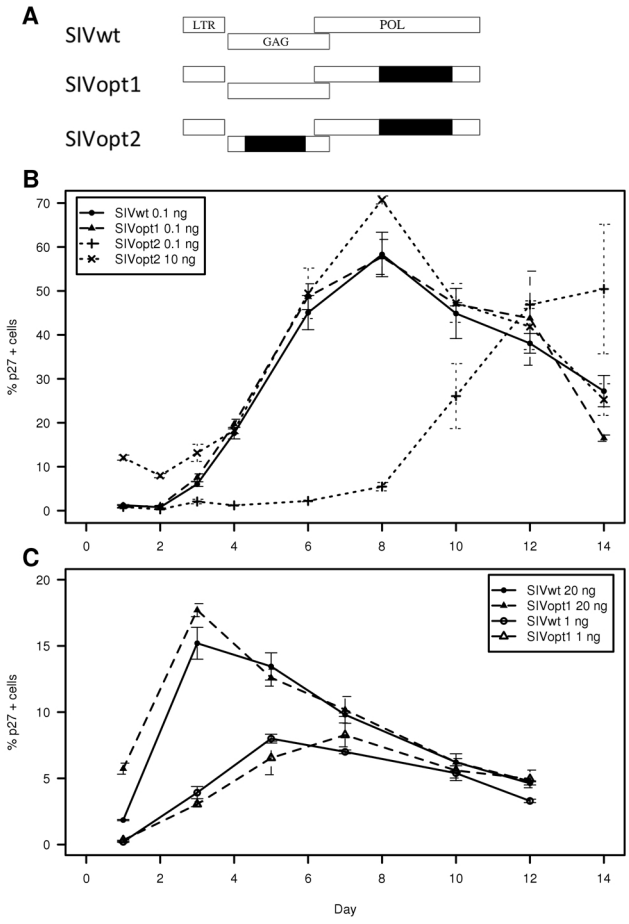
Figure 5. Differential sensing of SIV-infected lymphocytes by Hela-derived epithelial cells

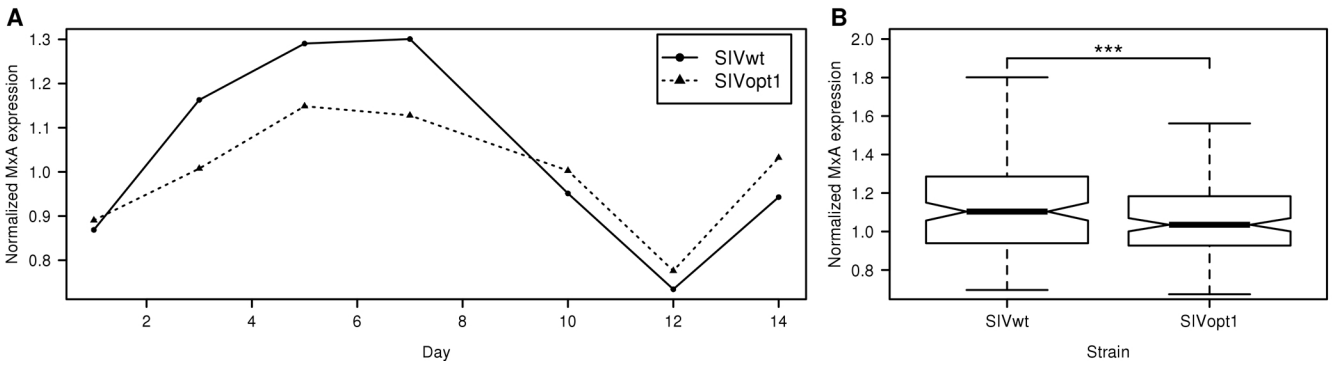
A. SIV Pol derived RNA transfection. Hela P4C5 cells were transduced with a lentiviral vector (LV) expressing BVDV-Pro or an irrelevant (CTRL) siRNA. IFN-I stimulation was measured in cells co-transfected with IFNβ-luciferase reporter plasmid and SIVwt or SIVopt1 Pol RNA sequence. The fold induction of luciferase activity, compared to a negative control with no RNA transfection is shown. Wt Pol RNA was able to stimulate a higher IFN-I response than nucleotide optimized version of the RNA. IFN-I induction was lost in BVDV-Pro (which cleaves IRF-3) expressing cells. Right Panel: IRF3 and actin levels, assessed by western blot, in control (CTRL)

and silenced Hela P4C5 cells. **B. Sensing of SIV-infected lymphocytes.** Induction of IFN-I in Hela P4C5 cells transfected with IFNβ-luciferase reporter plasmid and cocultured for 16 h with CEMx174 cells infected by SIVwt or SIVopt1 at various level of infection (X-axis). The fold induction of luciferase activity, compared to non-stimulated cells is shown (Y-axis). SIVwt-infected cells show a higher IFNβ promoter induction in Hela P4C5 than SIVopt1 infected cells ($p=0.001$). **C. Role of IRF-3.** Control Hela P4C5 cells or expressing BVDV-Pro were cocultured with approximately 25% SIVwt or SIVopt1-infected Cemx174 cells and assayed for IFN-β promoter activity. The paramyxovirus Sendai virus (SeV) was used as a positive control.









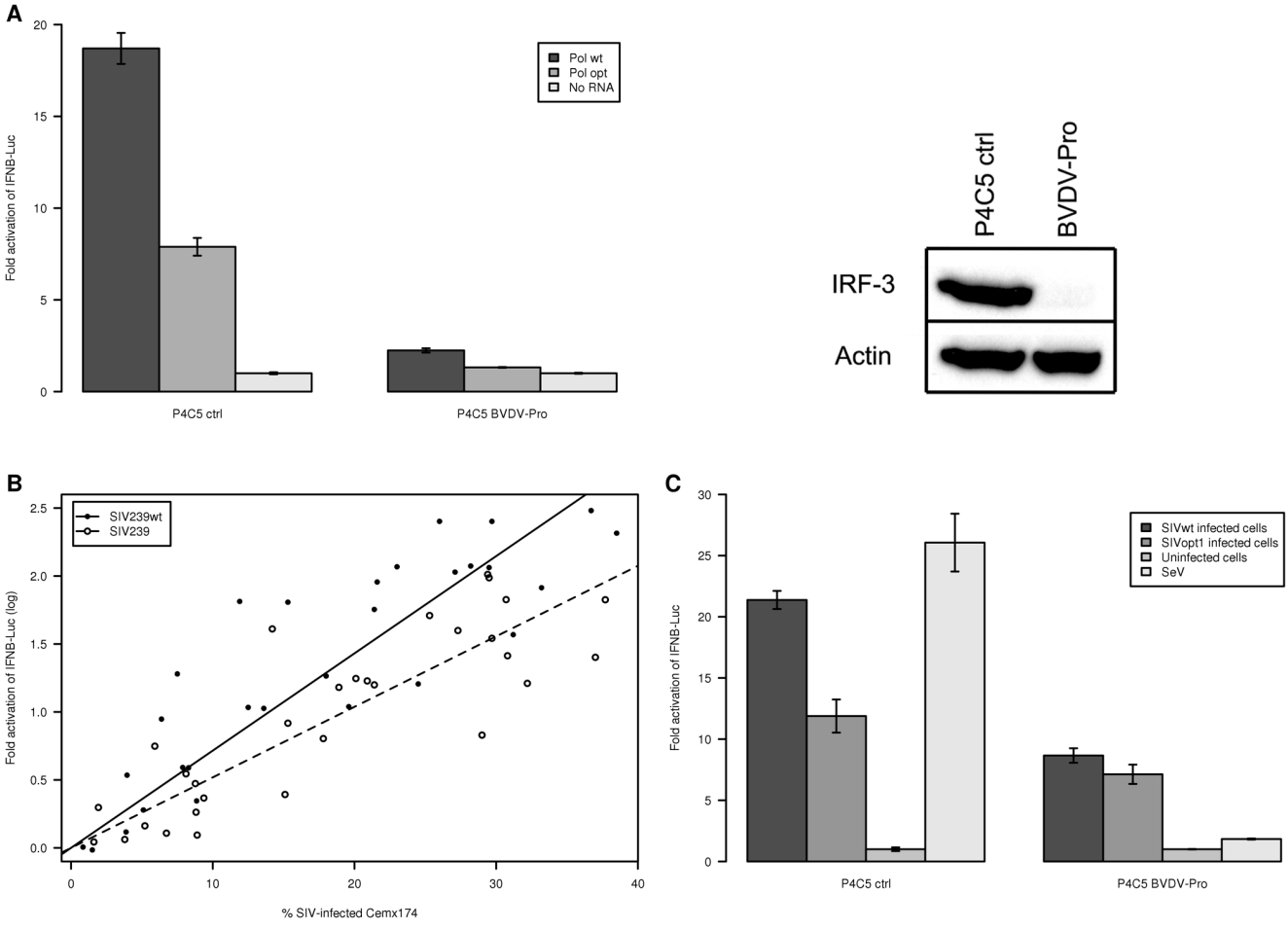


Table 1. Genomic regulatory sequences excluded from algorithm optimization in SIVmac239

Sequences excluded	Function during viral replication	Reference
Long Terminal Repeats	Regulation of transcription, translation and genome packaging	(4)
300 first nucleotides of Gag	Genome packaging in FIV or SIV-derived vector	(12) (44)
Gag-Pol frameshift	Translation of Pol gene	(45)
5' A-rich sequence of Pol	Important for cDNA synthesis during HIV-1 reverse transcription	(33)
Central PolyPurine Tract and Central terminal Sequence	Regulation of Reverse transcription	(44)
Rev Responsive Element	RNA subcellular trafficking	(36) (51)
Sequences containing donor and acceptor splicing sites	RNA maturation	(63) (69) (56)