ROBERT KOCH INSTITUT

Originally published as:

**Hanke, K., Kramer, P., Seeher, S., Beimforde, N., Kurth, R., Bannert, N.
Reconstitution of the ancestral glycoprotein of human endogenous retrovirus K and modulation of its functional activity by truncation of the cytoplasmic domain
(2009) Journal of Virology, 83 (24), pp. 12790-12800.**

**DOI: 10.1128/JVI.01368-09**

This article is not available in the publisher's version.

# Reconstitution of the Ancestral Glycoprotein of a Human Endogenous Retrovirus-K and Modulation of its Functional Activity by Truncation of the Cytoplasmic Domain

Kirsten Hanke[1§], Philipp Kramer[1§], Sandra Seeher[1], Nadine Beimforde[1,2], Reinhard Kurth[1], Norbert Bannert[1*]

[1]Robert Koch Institute, Centre for Biological Safety 4, Nordufer 20, 13353 Berlin, Germany

[2]Charite-Universitätsmedizin Berlin, Hindenburgdamm 30, 12203 Berlin, Germany

Running Title: Reconstruction and characterization of HERV-K Env

§ K.H. and P.K. contributed equally to this work.

*Corresponding author: Robert Koch Institute
Centre for Biological Safety 4
Nordufer 20
13353 Berlin
Germany
Phone: +49 (0) 30-4547-2549
Fax: +49 (0) 30-4547-2334
Email: BannertN@rki.de

Abstract Word Count: 241

Text Word Count: 8721

**Abstract**

Endogenous retroviruses present in the human genome provide a rich record of ancient infections. All presently recognized elements, including the youngest and most intact proviruses of the Human Endogenous Retrovirus-K(HML-2) family, have suffered postinsertional mutations during their time of chromosomal residence and genes encoding the envelope glycoprotein (Env) have not been spared these mutations. In this study we have, for the first time, reconstituted an authentic Env of a HERV-K(HML-2) provirus by back-mutation of putative postinsertional amino acid changes of the protein encoded by HERV-K113. Supported by codon-optimized expression we demonstrate that the reconstituted Env has regained its ability to be incorporated into retroviral particles and to mediate entry. The original ancient HERV-K113 Env is synthesized as a moderately glycosylated gp95 precursor protein cleaved into surface (SU) and transmembrane (TM) subunits. Of the nine N-linked oligosaccharides, four are part of the TM subunit contributing 15 kDa to its apparent molecular mass of 41 kDa. The carbohydrates as well as the cytoplasmic tail are critical for efficient intracellular trafficking, processing, stability and particle incorporation. Whereas deletions of the carboxy-terminal six residues completely abrogate cleavage and virion association, more extensive truncations slightly enhance incorporation but dramatically increase the ability to mediate entry of pseudotyped lentiviruses. Although the first HERV-K(HML-2) elements infected human ancestors about 30 million years ago, our findings indicate that their glycoproteins are in most respects remarkably similar to classical contemporary retroviruses and can still mediate efficient entry into mammalian cells.

## Introduction

Retroviruses that infect germ line cells or their precursors can become vertically transmitted genetic elements and spread in a host population during subsequent generations (35). About 8% of the human genome consists of stably integrated endogenous retroviruses acquired during early and more recent evolution by our primate and hominid ancestors. These fossils are grouped into several distinct families (2, 42). Their sequence and genomic structure typically resemble one of the genera of current exogenous retroviruses providing an exceptional archive for the study of many aspects of viral and host co-evolution and its dynamics (25). In contrast to humans, several animals including mice and sheep contain replication competent present-day exogenous and endogenized forms of the same retrovirus (1, 10). The most striking example is the koala, a species with a profound ongoing endogenization burst with a highly oncogenic gammaretrovirus (52).

The most recently integrated human elements belong to the betaretrovirus-like HERV-K(HML-2) family (7, 25, 54). Infectious viruses of this family appear to have started invading the chromosomes before the evolutionary split of Old World monkeys and hominoids about 30 million years ago (41, 50). Several of its members are human specific, indicating a continuing active replication in our ancestors following the deviation of the chimpanzee lineage 5-6 million years ago (4, 7, 12). The recent acquisition of several HERV-K(HML-2) elements is further substantiated by the lack of fixation (not present in all individuals) (6, 54). However, during their residency in the host genome every one of the currently known proviruses has suffered from mutations, deletions or recombination events. In many cases homologous recombination has left only a single LTR at the integration site (31). None of the more complete proviruses appear to be replication competent, although some of them have retained the capacity to form particles (8, 11). Recently, infectious HERV-K(HML-2) viruses have been produced by generating consensus sequences based on human specific elements and, in an alternative approach, by assembling functional regions of three authentic proviruses

into a single element (21, 39). These studies clearly demonstrate that (i) HERV-K(HML-2) is able to form viral particles infecting human cells and (ii) certain recombination events such as template switching during reverse transcription or gene conversion might re-establish fully functional chimeric HERV-K(HML-2) elements by combining conserved sequences from partially crippled proviruses.

One of the best preserved full length HERV-K(HML-2) elements is HERV-K113 (5, 54). Due to its low prevalence of less than 20% it escaped recognition by the Human Genome Project and the provirus was identified in a human BAC library on chromosome 19p13.11 of an unknown DNA donor (54). As previously reported by us and by others, despite a functional LTR promoter and open reading frames for all proteins, a few substitutions in the reverse transcriptase and one critical substitution in the *gag* gene of the provirus contribute to its lack of replication (5, 29, 54). In addition, the envelope protein of the cloned HERV-K113 provirus is not incorporated into viral particles, suggesting further postinsertional damage by mutation(s) (5, 20). These changes are not necessarily present in all HERV-K113 variants in the human population since some degree of polymorphism between carriers from different ethnic groups can be expected (44). Thus functional HERV-K113 envelope genes might still exist in some humans, as has been demonstrated for the related HERV-K108 element (20).

Retroviral envelope proteins are synthesized as fusion inactive precursor glycoproteins which trimerize and get cleaved by furin-like endopeptidases in the Golgi compartment of the cell into SU and TM subunits (32, 38). The mature N-glycosylated trimeric proteins eventually reach the cell membrane where they become incorporated into particles (18). The cytoplasmic domain of the protein has been shown to be critical for transport, incorporation and fusogenicity of virions (9, 13, 15, 37). Some retroviruses release a short carboxy-terminal R-peptide, a step necessary for full fusogenic activity of the glycoprotein (27). With the exception of the syncytin-1 and -2 glycoproteins which are the fusogenic envelope proteins of a HERV-W and HERV-FRD element involved in the formation of the placental

syncytiotrophoblast layer, the envelope proteins of other HERVs are poorly characterized (14).

In the present report we have identified inactivating mutations in the envelope gene of a present day HERV-K113 provirus. All or almost all of the changes most likely occurred gradually following integration in the genome more than a million years ago. In this case the already endogenous virus was probably able to replicate for a certain period of time, although damaging pre-integrational mutations introduced by the RNA polymerase or during reverse transcription by errors of the enzyme and by the action of APOBEC proteins can not be completely ruled out. We reconstituted the original sequence presumed to exist at the time of integration. Facilitated by codon-optimized expression we investigated various aspects of the glycoprotein and carboxy-terminal truncation mutants in terms of cleavage, N-glycosylation, incorporation into viral particles and mediation of viral entry into target cells.

## Materials and Methods:

### Cell culture.

HEK 293T, Tera-1, HeLa and CrFK cells were cultured in complete Dulbecco's Modified Eagle Medium (DMEM) containing 10% FBS, penicillin (50 U/ml), streptomycin (50 µg/ml) and L-glutamine (2 mM).

### Codon optimization, mutagenesis and cloning of Env constructs.

A synthetic version of the HERV-K113 (GenBank AY037928) Env protein, optimized for expression in mammalian cells, was designed and synthesized by Geneart (Regensburg, Germany). The codon optimized sequence, coEnv (Fig. S1), was cloned into a C-terminal V5-tag-containing pcDNA3 expression vector (Invitrogen) using the EcoRI/ NotI cleavage sites. The construct for the expression of the putative ancestral (original) oricoEnv-V5 (pcDNAoricoEnv-V5) as well as all derived mutants were generated using the QuickChange Multi Site-Directed Mutagenesis Kit or the QuickChange Site-Directed Mutagenesis Kit (Stratagene) and primer pairs listed in Fig. S2. These mutants include the N-glycosylation deficient variant oricoEnvGly$^-$, the cleavage site mutants oricoEnvCS$^-$ and oricoEnvCS_MMTV, the C-terminal truncation mutants oricoEnvΔ659-699, oricoEnvΔ680-699, oricoEnvΔ693-699 and the envelope proteins without V5-tag. The last four were generated by the introduction of stop codons at the respective positions in pcDNAoricoEnv-V5. All mutations were confirmed by DNA sequencing.

### Envelope expression analysis.

For Env expression analysis, 4 x $10^6$ cells grown in 100 mm dishes were transfected with the appropriate expression plasmid using the calcium phosphate method. Cells were lysed in 500 µl lysis buffer (1% Triton X-100, 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA,

and protease inhibitor mix (Roche)) and cell debris was removed by centrifugation. For immunoprecipitation experiments 25 µl of αV5 sepharose (Sigma) were added to the lysate and incubated on an overhead shaker for 4 h. The sepharose beads were washed 5 times with washing buffer (1% Triton X-100, 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.02% SDS and protease inhibitor mix), mixed with Laemmli sample buffer (BioRad), briefly boiled and pelleted. For simple immunoblotting, the cleared lysates were directly mixed with sample buffer and cleared by centrifugation before being subjected to SDS-PAGE. The separated proteins were transferred to a polyvinylidene diflouride membrane (Roth) using the semi dry transfer method. The membranes were blocked in blocking buffer (PBS, 5% skimmed milk powder, 0.1% Tween) and finally incubated with specific antibodies. An αV5-HRP antibody (Invitrogen) was used for quantification. It was detected by a secondary Alexa Fluor 680 donkey anti-mouse IgG before measurement of band fluorescence intensity using an Odyssey Scanning and Analysis (LI-COR) unit. In experiments without quantification, αV5-HRP or TM specific (HERM-1811-5, Austral Biologicals) monoclonal antibodies were used in conjunction with an HRP-conjugated secondary anti-mouse antibody. Bands were visualized using Enhanced Chemiluminescence Reagents (Pierce).

**Immunofluorescence microscopy.**

Cells (1 x $10^5$) were grown in chamber slides (Nalge Nunc) and transfected with 0.2 µg of plasmid DNA using Effectene Reagent (Qiagen). After 48 h, cells were fixed with 2% paraformaldehyde in PBS for 30 min, rinsed briefly with PBS, permeabilized with PBS-buffered 0.5% Triton-X 100 and again washed 3 times with PBS. After blocking with Marvel (1% non fat milk in PBS) for 20 min, cells were incubated with primary antibodies diluted in blocking buffer for 60 min at 37 °C. HERM-1811-5 (Austral Biologicals) or αV5 (Serotec) were used as primary antibodies. The slides were washed 3 times with PBS and a Cy3-conjugated mouse specific secondary antibody (Sigma) diluted 1:200 was added for 30 min.

After extensive washing the cells were mounted in Mowiol and examined using a Zeiss LSM 510 confocal laser scanning microscope.

**Deglycosylation experiments.**

PNGase F deglycosylation experiments were performed with precipitated V5-tagged envelope proteins immunoprecipitated from lysates of transfected cells grown in 6-well plates. Cells were lysed 48 h post transfection in 200 µl lysis buffer. The precipitates were briefly boiled in 1x denaturation buffer at 98 °C for 10 min and 3 µl 10x G7- buffer (New England Biolabs), 3 µl 10% NP-40 and 1500 units of N-glycosidase F (PNGase F, New England Biolabs) were added for digestion. The tubes were incubated for various times at 37 °C and the deglycosylation reaction was stopped by adding Laemmli sample buffer (BioRad). The proteins were immunoblotted using the αV5- HRP specific antibody (Invitrogen).

To inhibit intracellular N-glycosylation, cells were preincubated for 2.5 h with DMEM containing 30 or 50 µg/ml of tunicamycin. After the pretreatment, cells were transfected using the Polyfect Transfection Reagent Kit (Qiagen) and incubated for an additional 24 h in medium containing tunicamycin before lysis and immunoblotting.

**Production of pseudotyped SHIV reporter viruses.**

The SIVmac based retroviral vectors pSIvec1ΔenvLuc and pSIvec1ΔenvGFP, the HIV-1-Rev expressing plasmid pCMV-Rev and the plasmid pSVIII-ΔKS (expressing a non-functional HIV-1 envelope to allow non-specific virus uptake to be determined) are described elsewhere (3). To generate infectious SHIV particles capable of a single round of infection, 4 x $10^6$ HEK 293T cells were transfected with 15 µg of pSIvec1ΔenvLuci, 5 µg pCMV-Rev and 0.5 µg of an envelope expression vector in 10 cm dishes using the calcium phosphate method. At 2 days post-transfection, supernatants were harvested, centrifuged at 6000 rpm for 10 min and

filtered through 0.45 µm pore-sized membranes to remove residual cells. Gag-specific ELISA and RT-activity (GE Healthcare) were used to assess and normalize virus preparations.

**Ultracentrifugation of cell supernatants.**

Viral particles were concentrated by ultracentrifugation of cell supernatants through a 20% sucrose cushion for 2 h, at 4 °C and 175,000g. The viral pellet was resuspended in 50 µl of 0.05 M Hepes, pH 7.2.

**Infectivity assays.**

For entry experiments, CrFK target cells were sown in 6-well plates (600,000 cells/well). The next day, cells were rinsed with PBS and incubated with 400 µl normalized virus supernatant and 200 µl DMEM for 2 h before addition of 1.5 ml DMEM medium and culturing for 2 days. Infection was determined by measuring the luciferase activity in cell lysates using the Luciferase Assay System (Promega) according to the manufacture's instructions.

## Results

### Enhanced expression of the HERV-K113 envelope protein.

The Env of HERV-K(HML-2) is translated from a singly spliced transcript. A second splicing event removing large parts of the coding region results in the mRNA for the nucleocytoplasmic shuttle protein Rec at the expense of Env (Fig. 1). This internal splicing, the low GC content of 42% and the use of rare codons are the most likely causes for the poor expression of the HERV-K(HML-2) envelope protein from CMV promoter-driven constructs (5). To overcome this limitation for the Env of HERV-K113, which significantly hampers a functional characterization of the protein, we have generated a synthetic version optimized for expression in mammalian cells by synonymous exchange of 71% of the codons (no change in amino acid sequence) and raising the GC content of the sequence to 63%. The codon optimization and the concomitant removal of the internal splice sites resulted in a drastic increase in expression (over 50-fold) as demonstrated by immunoprecipitation (Fig. 2A) of the glycoproteins from transiently transfected HEK 293T cells and by immunofluorescence microscopy (Fig. 2B). For immunoprecipitation and detection purposes, a V5-tag that increases the molecular size by 1.7 kDa has been fused to the C-terminus of wtEnv and coEnv. On immunoblots, specific bands with apparent molecular masses in the range of about 75-95 kDa and 32-38 kDa are visible (Fig. 2A), sizes that match those expected for the glycosylated precursor protein and the transmembrane (TM) subunit (5).

### Reconstitution of the ancestral sequence of the HERV-K113 envelope protein.

We and others have previously reported that the HERV-K113 envelope protein is not incorporated into viral particles (5, 20) and it was assumed that the cause lay in postinsertional mutations acquired during its time in the human genome. In an approach similar to that recently used to generate two related HERV-K(HML-2) consensus sequences (21, 39), we aligned the amino acid sequence of HERV-K113 Env with the corresponding

sequences of 10 well-preserved human specific elements belonging to this family to identify and repair postinsertional amino acid substitutions in the HERV-K113 coEnv protein (Fig. 3). To differentiate between authentic polymorphisms in the integrating ancient viruses and postinsertional changes, we assumed a nonsynonymous postinsertional mutation if none or only one of the aligned proviruses matched the amino acid of HERV-K113 and assumed a shared polymorphism if two or more of the elements had the same amino acid as HERV-K113, even if different from the consensus sequence. Three such variable positions (T137, I369 and V676) were identified and left unchanged (Fig. 3). In contrast, 8 putative protein-relevant postinsertional substitutions were identified, 4 in the SU and 4 in the TM region (Fig. 3). Using site-directed mutagenesis we reversed these 8 amino acids in the pcDNAcoEnv-V5 plasmid, taking the consensus amino acids as the originals. The resulting construct, pcDNAoricoEnv-V5, encodes a V5-tagged version of the putatively ancestral HERV-K113 envelope protein existing at the time of integration into a human ancestor.

**The original HERV-K113 envelope protein is moderately N-glycosylated.**

We first compared the expression of coEnv (present day) and oricoEnv (original) by immunoblotting of lysates from transfected HEK 293T, Tera-1 and HeLa cells. Although both glycoproteins had similar levels of expression, considerable differences in the processing and migration patterns of the presumed TM subunits were seen. Although the ratio of precursor to TM varied somewhat from lysate to lysate, oricoEnv was more efficiently cleaved than coEnv in all experiments. The presumed TM subunit of oricoEnv-V5 migrates as a 43 kDa band in all cells tested. In contrast, coEnv-V5 is very inefficiently cleaved into a 35 kDa protein in Tera-1 cells, and in HEK 293T and HeLa cells a fuzzy doublet of about 33 and 35 kDa can be seen (Fig. 4A).

Differences in the carbohydrate composition of the two proteins were considered to be a likely reason for the discrepancy in size. CoEnv contains 10 consensus NXS/T motifs for

N-linked glycosylation (23, 55), with two of the sites (N506 and N507) being adjacent (Fig. 3). The N506K back mutation altered this motif giving 9 potential sites for N-linked carbohydrates in oricoEnv: 5 in the SU subunit and four in the ectodomain of TM (Fig. 4B). Treatment of precipitated glycoproteins from transfected HeLa cells with peptide-N-glycosidase F (PNGase F), which removes all N-linked glycans, decreased the apparent molecular weight of the presumed oricoEnv-V5 TM subunit to 28 kDa (Fig. 4C), the size calculated for the non-glycosylated TM-V5 subunits. However, the supposed TM doublet of coEnv-V5 was also shifted by PNGase F treatment although the deglycosylated proteins clearly migrated above 28 kDa, indicating further N-terminal processing sites and a lower carbohydrate content of the TM subunit compared to that of oricoEnv. The precursor sizes of approximately 95 and 70 kDa are consistent with the molecular masses of the glycosylated and non-glycosylated proteins (the theoretical size of the unmodified protein without signal peptide is 68 kDa).

To ascertain the nature of the detected protein bands we incubated pcDNAoricoEnv-V5 transfected HEK 293T cells with tunicamycin, an agent that prevents the transfer of saccharidic units to the nascent peptide chain in the ER. Immunoblot analysis show that tunicamycin treatment results in the disappearance of the glycosylated precursor and TM bands and instead a protein pattern similar to that observed after PNGase F treatment (70 kDa precursor and 28 kDa TM) is seen (Fig. 4D). The TM protein in tunicamycin treated cells is barely visible, indicating an inefficient furine protease-mediated cleavage of the non-glycosylated protein. This assumption is further strengthened by the expression and immunoblot analysis of a mutant oricoEnv protein with all nine potential N-glycosylation sites mutated by changing the serine or threonine residue of the NXS/T motif into alanine. The oricoEnvGly$^-$ protein shows a band pattern similar to that already described for cells treated with moderate levels of tunicamycin (Fig. 4E). Together, these results illustrate that prevention of N-linked glycosylation results in a 70 kDa precursor protein whose processing

is extremely inefficient. Furthermore, the strength of a specific 64 kDa band increases with treatment. This protein appears to represent an N-terminal truncation product of the unglycosylated precursor, suggesting a protective role for the N-terminal glycans against cellular proteases and degradation.

Interestingly, the position of one of the potentially glycosylated asparagines (N461) is only 5 amino acids upstream of the predicted furin cleavage site (Fig. 4B). An N-glycan at this location might influence the processing efficiency of the precursor protein. To test this by preventing glycosylation at this position we substituted the serine in the $^{461}$NRSKR sequence to alanine ($^{461}$NRAKR). This mutation obliterates the glycosylation signal and changes the supposed cleavage site motif of oricoEnv into a sequence matching that of the closely related betaretrovirus mouse mammary tumor virus (MMTV) (Fig. 4B). As shown in Fig. 4F, this substitution had no significant effect on processing efficiency. In contrast, replacing the positively charged amino acids in the cleavage site by alanine ($^{461}$NASAA mutant) completely abolished the release of the 43 kDa TM subunit. This result therefore implies that either N461 is not glycosylated or that the carbohydrate at this location neither inhibits nor enhances processing. Furthermore, the $^{461}$NASAA mutant provides further evidence that the 43 kDa TM protein is generated by processing at the assumed SU/TM cleavage site.

**Incorporation of the reconstituted HERV-K113 Env into lentiviral particles is modulated by C-terminal sequences.**

We next investigated whether or not the HERV-K113 oricoEnv protein is incorporated into lentiviral particles. Together with vectors for the expression of the lentiviral proteins and the packaged RNA containing a luciferase gene, 0.5 µg of an Env plasmid of choice were included to produce pseudotyped SHIV-reporter particles in transfected HEK 293T cells, as described previously (3). Furthermore, for several retroviruses a C-terminal truncation of the cytoplasmic tail has been shown to influence particle incorporation and/or fusogenic

properties of the envelope protein (13-15). We therefore created three mutants with C-terminal truncations, namely oricoEnvΔ659-699, oricoEnvΔ680-699 and oricoEnvΔ693-699 (Fig. 5A), and included them with the cleavage and glycosylation minus mutants oricoEnvGly$^-$ and oricoEnvCS$^-$ in the experiments. All proteins were expressed at approximately equivalent levels in HEK 293T cells (Fig. 5B). The Env proteins were detected using a commercially available monoclonal antibody that recognizes an epitope in the ectodomain of the TM subunit and the changes in size of the truncated TM subunits were noticeable for oricoEnvΔ659-699 and oricoEnvΔ680-699 (Fig. 5). Interestingly, no processing occurred with the oricoEnvΔ693-699 mutant that lacks just seven amino acids (Fig. 5B, left panel). To examine incorporation into virions, the supernatants were harvested, carefully clarified and concentrated by ultracentrifugation through a 20% sucrose cushion. Immunoblot analysis of the pellets using the TM specific antibody revealed varying degrees of incorporation into SHIV particles (Fig 5B, right panel). Whereas no incorporation was detectable for coEnv, oricoEnvΔ693-699 and the glycosylation and cleavage deficient oricoEnv mutants, substantial incorporation of oricoEnvΔ680-699 and oricoEnvΔ659-699 was observed at levels approximately 2 and 5 times higher, respectively, than the parental oricoEnv. It is of particular interest that the mutant almost completely lacking the cytoplasmic tail (oricoEnvΔ659-699) is exclusively incorporated as the cleaved glycoprotein.

Together, these results demonstrate that progressive truncation of the HERV-K Env cytoplasmic tail influences processing and particle association.

**The ancestral HERV-K113 envelope protein mediates entry of retroviral particles and its infectivity can be augmented by an extensive C-terminal truncation.**

The infectivity of the SHIV-luciferase reporter particles carrying various envelope proteins was tested next. The single round entry assays were performed using Crandell feline kidney

(CrFK) cells or dog Cf2Th cells as targets and virus suspensions normalized for reverse transcriptase activity. CrFK cells have been previously reported to be susceptible to infection with HERV-K(HML-2) (39). Although specific entry of virions carrying coEnv or oricoEnvΔ693-699 could not be detected (Fig. 6), significant entry was achieved by pseudotyping with the ancestral oricoEnv, giving levels of luciferase activity two orders of magnitude higher than the background determined with a non-functional HIV envelope protein (ΔKS) (3). Similar results were obtained in experiments with Cf2Th cells (data not shown). Unexpectedly, the efficiencies of entry with oricoEnvΔ680-699 and oricoEnvΔ659-699 pseudotypes were even 10 or 100-fold higher, respectively (Fig. 6). In parallel experiments using pseudotypes expressing GFP instead of luciferase, flow cytometry revealed that up to 30% of the target cells were infected with oricoEnvΔ659-699 (data not shown). Therefore, in contrast to the modern proviral HERV-K113 envelope protein, the putative ancestral envelope is efficiently incorporated into retroviral particles and mediates entry. Removal of the last 7 amino acids of the glycoprotein completely blocked cleavage and particle incorporation and although truncation of the cytoplasmic tail by 20 or 42 C-terminal amino acids resulted in only a slightly higher incorporation there was a much stronger enhancement of virus entry.

**Two arginine-to-cysteine mutations are the predominant cause for the loss of coEnv function.**

To further analyze the impact of the postinsertional mutations in the *env* gene of HERV-K113 we generated coEnvΔ659-699 constructs containing the SU or TM sequences of oricoEnv. These two envelope proteins (coEnvΔ659-699 SU-orico and coEnvΔ659-699 TM-orico) did not facilitate entry of pseudotyped reporter viruses, indicating that mutations in both the SU and the TM subunit of coEnv contribute to the lack of entry (Fig. 7A). Eight oricoEnvΔ659-

699 mutants, each carrying a single reversion, were then generated in order to identify the extent with each of the eight postinsertional mutations contributes to the loss of coEnv function. Entry experiments using these pseudotyped reporter viruses showed that only the R140C mutation in the SU subunit and the R577C mutation in the TM essentially completely abolish entry whereas the K506N and P626T mutations each reduce entry by a factor of approximately 10 (Fig. 7A). Similar results were obtained with mutants in the context of the full length oricoEnv protein (Fig. 7B).

## Discussion

Cell entry of retroviruses is triggered by a specific interaction of the viral envelope glycoprotein with one or more receptor proteins in the cellular membrane. The receptor expression profile largely defines the virus's cellular tropism. Despite being present in many healthy and tumorigenic human tissues and being capable of pseudotyping other retroviruses, the envelope protein of HERV-K(HML-2) has been only poorly characterized (17, 20). Two major obstacles have long hampered the study of the protein's fundamental characteristics: the inefficient protein expression in mammalian cells, even when under the control of strong promoters, and the uncertainty concerning the impact of postinsertional mutations (53). Recently however, recognition that the HERV-K108 glycoprotein has residual functional activity plus the *in silico* generation of consensus prototypic HERV-K(HML-2) sequences have added considerably to our basic understanding of the virus and its proteins (20, 21, 39).

The data presented here describe a strategy to overcome the poor expression and document efforts to reconstitute a HERV-K(HML-2) Env protein coded for by the HERV-K113 element that originally entered the human germ line as an infectious retrovirus more than 800,000 years ago (33, 54). This provirus is one of the most studied elements. It has preserved open reading frames for all viral proteins, but has been inactivated by postinsertional mutations effecting, amongst others, the expression and function of Env (5, 29).

The use of a synthetic DNA sequence optimized for mammalian cells elevated the expression of HERV-K113 Env by a factor of over 50 compared with that of the proviral sequence. Codon optimization has previously been successfully used to boost the expression of several other retroviral proteins, including envelope glycoproteins (19, 45). Full-length retroviral genomes require a balanced expression of their proteins for optimal replication and their DNA, RNA and protein sequences are under a multitude of constrains and selective pressures. The consequential use of rare codons, low GC content, unstable RNA and internal splicing, (e.g. the generation of Rec transcripts from HERV-K *env* mRNA) results in submaximal

levels of expression. A limited expression of toxic viral proteins can as well minimize adverse effects on cell viability. Viral RNA is also a target for antiviral responses that result in specific transcript degradation and low protein expression. Recently a large number of microRNAs has been identified that target exogenous and endogenous retroviruses, including HERV-K (28). The role played by such silencing small RNAs in the inefficient expression of HERV-K proteins deserves further investigation.

The efficient expression of the Env protein prompted us to identify and correct mutations that have altered the original protein sequence of the primordial infectious retrovirus and to perform a comparative study with the contemporary Env. Such sequence changes may have been already introduced before integration, e.g. by the RNA polymerase in the producer cell, during reverse transcription, as a result of the inherent low fidelity of the enzyme or by the action of APOBEC deaminases. Alternatively, the mutations have occurred gradually over evolutionary time since integration. Because it is unlikely that there are more than one or two RT errors in a single round and as the action of APOBEC deaminases has not been shown to significantly effect HERV-K113 in a recent report (40), it is likely that most, if not all, of the mutations occurred after integration.

Although the method used to identify postinsertional changes was based, like those previously used to generate consensus sequences (21, 39), on sequence alignments, it also allows likely mutations and shared polymorphisms in a minority of elements to be discriminated (29). The young age of the 11 proviruses used for alignment makes it very unlikely that the shared polymorphisms are the result of matching postinsertional mutations in more than two individual proviruses or of ectopic recombination events (30) between the *env* genes of HERV-K113 and two or more of the other aligned elements, although these cannot be completely ruled out.

The assumed shared polymorphisms account for three amino acid differences between the reconstituted HERV-K113 Env and the deduced consensus sequence (Fig. 3). In contrast,

amongst the aligned sequences eight amino acids are unique to the contemporary HERV-K113 Env or match with only one element. The corresponding non-synonymous mutations were therefore assumed to be postinsertional changes and were inverted to restore the original protein sequence. As shown by the immunoblot TM band, the reversions yielding oricoEnv altered the protein's N-glycan pattern. Although one consensus N-glycosylation motif in the TM subunit was lost by the N506K back mutation, an increase in the apparent molecular mass was seen. Aberrant folding and impaired transport through the ER-Golgi compartment are the likely reasons for the peculiar glycosylation, migration characteristic in the Western blot, and the reduced stability of the modern HERV-K113 Env. This is consistent with the very low level of surface expression and lack of incorporation into viral particles (20). In contrast, the TM subunit of the reconstituted original protein (oricoEnv) displayed a more regular glycosylation pattern, reached the cell surface (data not shown) and was efficiently incorporated into lentiviral particles. In agreement with the known consensus sequences for HERV-K(HML-2), oricoEnv contains nine potential N-glycosylation sites, 5 in the SU and 4 in the TM subunit. The number and SU/TM ratio of 5/4 are similar to that of Jaagsiekte sheep retrovirus (JSRV; ratio 6/3). The mouse mammary tumor virus (MMTV; ratio 3/2) or syncytin 1 (HERV-W; ratio 6/1) are less glycosylated (14). Deglycosylation experiments shifted the V5-tagged oricoEnv TM by 15 kDa from 43 kDa to 28 kDa. Assuming a reduction of at least 2.8 kDa per oligosaccharide (14), it is likely that all four N-glycosylations sites predicted for TM are modified in HEK 293T cells and probably in most other human cells. The same seems also to be true for the five sites in the oricoEnv SU subunit. Focusing on the HERV-K108 glycoprotein, which contains one additional N-glycosylation site in its SU subunit, Dewannieux and colleagues demonstrated a 20 kDa shift following PNGase F digestion (20). This result and the size difference between the glycosylated and non-glycosylated precursor proteins indicate that most, if not all, theoretical sites carry an N-linked carbohydrate. Of particular interest in this respect is N461 that lays only 5 amino acids upstream of the furin

cleavage site ($^{461}$NRSKR). Since cleavage occurs after initial glycosylation, an oligosaccharide at this position might influence processing efficiency (55) and indeed, an impaired cleavage has previously been shown for HERV-K(HML-2) Env proteins (8). To test this, we substituted the serine with alanine, eliminating the glycosylation motif and immediately creating the cleavage site of MMTV (26). The mutated protein was cleaved with the same efficiency as the oricoEnv protein, indicating that in HEK 293T cells, at least, the presence of an N-glycosylation site close to the cleavage site does not impede the furin protease or asparagine 461 is not glycosylated. Alternatively, a cleavage enhancement by the N-glycan prevention in the $^{461}$NRAKR mutant might be exactly balanced by a negative effect of the serine to alanine mutation. As tunicamycin treatment or expression of non-glycosylated oricoEnv resulted in poor processing and a susceptibility to intracellular proteolysis, it would appear that efficient cleavage of Env depends on proper glycosylation. Several studies have already described the importance of extensive SU and TM glycosylation for folding, transport, processing, fusogenicity and immunological shielding of numerous retroviral envelope proteins (36, 47, 49). For HIV-1, which contains between 18 and 33 N-linked carbohydrates (4 in TM), mutation of glycosylation sites in gp41 has been shown to lead to a considerable inhibition of Env transport, cleavage and fusion activity (16, 23, 55). This impediment results from a substantial arrest of the protein in the Golgi (23). Furthermore, non-glycosylated HERV-W envelope proteins also remain trapped in the secretory compartment as misfolded proteins (14).

In agreement with these observations and with previously published data, it was not possible to demonstrate incorporation of the contemporary HERV-K Env (coEnv) into lentiviral particles or the production of infectious virus (5, 20). It is likely that one or both of the arginine to cysteine mutations which render the protein non-functional prevent the correct folding of the Env protein, which in turn results in aberrant glycosylation and a lack of particle incorporation. In contrast to coEnv, pseudotyping of SHIV particles by the ancient

oricoEnv was very efficient. The protein was able to efficiently mediate entry into target cells, demonstrating a functional reconstitution. In contrast, abrogating N-glycosylation or preventing processing into SU and TM subunits by mutation of the cleavage site dramatically reduced oricoEnv particle incorporation.

HERV-K(HML-2) Env has a relatively short putative cytoplasmic tail of only 44 amino acids (20) and truncation of the cytoplasmic tail is known to modulate particle incorporation and the fusogenicity of retroviral glycoproteins (13, 15, 37). In the case of MLV and related gammaretroviruses, the R-peptide, a short C-terminal peptide, is cleaved off by the viral protease during maturation, a step necessary for full fusogenicity. Although there is no evidence for the presence of an R-peptide in HERV-K(HML-2), we addressed whether arbitrary progressive C-terminal truncations of oricoEnv by insertion of stop codons would effect Env incorporation and associated viral functions. Surprisingly, the mutant with the shortest deletion (seven amino acids) had a late maturation defect with no cleavage or particle incorporation. The intermediate truncation of 18 amino acids did not significantly affect processing and incorporation efficiency but particle infectivity was increased by a factor of 10. A further truncation of 41 amino acids, leaving only 3 amino acids of the putative cytoplasmic domain, yielded particles bearing approximately 5-fold more Env but having a capacity for fusion almost 100-fold higher than with the full-length oricoEnv. Although the underlying mechanism for this difference remains unclear, it is possible that the higher glycoprotein content with the short-tailed Env helps the budding lentiviral particle by reducing the interference between the matrix protein of the nascent Gag shell and the C-terminal tail (13, 24, 43). Indeed, envelopes with short cytoplasmic tails are known to be better incorporated into SIV particles (34, 48). Alternatively, the mutant might have improved kinetics of transport and maturation or a lower rate of endocytosis resulting in a higher glycoprotein concentration at the budding site.

Truncation of the Env cytoplasmic tail has been reported to augment fusogenicity for a number of retroviruses (13, 46, 51). For the HERV-K Env it is tempting to speculate that the predominant or almost exclusive incorporation of completely processed glycoproteins is one reason for the considerably higher infectivity using the short-tailed truncation mutants. Other possible explanations include an increased mobility in the viral membrane (which could promote Env clustering during fusion) or an inherent structural difference in the TM ectodomain that favours fusogenicity (22, 56). Further studies are needed to elucidate the exact cause of this enhancement and to determine its relevance for integration into HERV-K particles.

Given the abundance and polymorphism of endogenous retrovirus-K elements in humans and other primates it is likely that many more functional HERV-K(HML-2) glycoproteins are to be found in several genomes in addition to the known functional HERV-K108 Env. The reconstitution and efficient expression of the original ancient HERV-K113 Env protein facilitated rational studies of the N-glycosylation, processing, incorporation and mediation of entry of this fossil retrovirus. The results show that the protein can pseudotype retroviruses and mediate efficient entry into cells. Its use as a functional model protein should assist in the search for the cellular receptor and in studies focusing on the pathophysiologic as well as physiologic implications of HERV-K(HML-2) Env expression.

**Acknowledgments**

We thank Sandra Klein for her excellent technical support and Dr. Kazimierz Madela for providing his expertise with the confocal microscope. We are grateful to Steve Norley for his critical reading of the manuscript and helpful discussions. The work was supported in part by a donation from the Heinz Kuthe de Mouson Legacy to R.K.

## Figure Legends

**Fig. 1.** Genomic organization of the HERV-K113 provirus and its envelope protein. The Env precursor protein is translated from a singly spliced RNA and is composed of the N-terminal signalpeptide (SP), the surface glycoprotein (SU) and the transmembrane subunit (TM) comprising the hydrophobic membrane spanning domain (MSD). The cleavage site (CS) for cellular furine proteases is indicated. The theoretical molecular weights of each subunit (SP, TM, SU) as well as the last amino acid of relevant domains are shown.

**Fig. 2.** Expression of wtEnv and coEnv. HEK 293T cells were transfected with pcDNAwtEnv-V5 and pcDNAcoEnv-V5. (A) Western blot analyses of immunoprecipitates using an αV5-HRP antiserum (1:5000). Assumed precursor proteins at 75-90 kDa and TM subunits at 32-38 kDa are visible. Quantification of the band intensities indicates an increase in expression of 50-fold or more resulting from Env codon optimization. (B) Immune fluorescence analysis of wtEnv and coEnv using an αV5-Cy3 antibody. The signal intensity and the number of cells expressing Env is significantly higher in cells transfected with coEnv.

**Fig. 3.** Alignment of the amino acid sequences of 11 HERV-K(HML-2) envelope proteins. Sequences of HERV-K101 (AF164609), K102 (AF164610), K104 (AC116309), K107 (AF164613), K108 (AC072054), K109 (AC055116), K115 (AY037929), Y178333, AP000776 and AC025420 are compared to the K113 (AY037928) wtEnv sequence. The consensus sequence and the putative original sequence (oricoEnv) obtained by applying an algorithm to identify shared polymorphisms (see results for details) are included. The consensus sequence and oricoEnv differ at positions 137, 369 and 676 due to shared polymorphisms (highlighted by rectangles). The eight sites in HERV-K113 wtEnv changed to the consensus sequence in order to reconstitute the ancestral protein sequence are in grey.

**Fig. 4.** Reconstitution of the ancient K113 oricoEnv sequence alters TM subunit processing. (A) Western blot analysis of coEnv-V5 and oricoEnv-V5 expressed in HEK 293T cells, Tera-1 and HeLa cells. (B) Position of the potential glycosylation sites (NXS/T) within the Env molecule. The localization of the cleavage site (CS) and of amino acid changes resulting in oricoEnv are indicated. (C) PNGase F treatment of cell lysates from HeLa cells expressing coEnv-V5 and oricoEnv-V5. Lysates were incubated with and without PNGase F at 37 °C for 1 h. Treated samples show a shift in the precursor and TM bands. (D) Incubation of HEK 293T cells expressing oricoEnv-V5 with tunicamycin (30 and 50 µg/ml). The generation of a 64 kDa N-terminal Env precursor degradation product (marked by an asterisk) is enhanced by tunicamycin treatment. (E) Elimination of the predicted glycosylation sites by mutagenesis results in a decrease in the molecular weights of the precursor and TM subunit and inhibits processing. (F) Analysis of cleavage site (CS) mutants in HEK 293T cells. In oricoEnv_CS$^{-}$-V5 the cleavage site sequence $^{461}$NRSKR was changed to $^{461}$NASAA. OricoEnv_CS/MMTV-V5 carries a mutation ($^{461}$NRSKR to $^{461}$NRAKR) that prevents potential glycosylation at N461 and matches the MMTV cleavage site.

**Fig. 5.** Incorporation of oricoEnv and C-terminal truncation mutants into lentiviral particles. (A) Amino acid sequences of the membrane spanning domain (boxed) and the cytoplasmic region are shown. The Env truncation mutants are named according to the number of amino acids deleted by the introduction of a stop codon (*). (B) The left hand panel shows an immunoblot of HEK 293T cells transfected with vectors used to produce pseudotyped SHIV reporter viruses plus the respective Env expression plasmids, visualized by a TM-ectodomain specific monoclonal antibody. Note the size shifts of the mutated TM subunits. The corresponding supernatants were collected, concentrated by ultracentrifugation through a sucrose cushion and also subjected to immunoblotting with the TM specific antibody (right

hand panel). Finally, the blot was stripped and incubated with an SIV Gag specific antibody to determine particle load.

**Fig. 6.** Infectivity assay with pseudotyped SHIV reporter particles carrying a luciferase gene. The lentiviral particles were produced in HEK 293T cells, normalized and used to infect CrFK cells. 48 h later, cells were lysed and assayed for luciferase activity. The background level was determined by infection with a truncated non-functional HIV Env (ΔKS). Mock treated cells were not infected. The means and standard deviations of six replicates are shown. Similar results were achieved in four independent experiments.

**Fig. 7.** Impact of postinsertional mutations in the envelope protein of HERV-K113 on the entry efficiency of pseudotyped SHIV reporter viruses. (A) Mutations in the context of the C-terminal truncation mutant with the highest entry level. (B) The substitutions have similar effects on the function of the full length oricoEnv. The means and standard deviations of six replicates are shown. Similar results were obtained in two independent experiments.

**Supplementary figure legends**

**Fig. S1.** Codon optimized sequence of the HERV-K113 *env* gene.

**Fig. S2.** Table of mutagenesis primers used for generating oricoEnv, oricoEnvCS$^{-}$, oricoEnvGly$^{-}$, oricoEnvΔ659-699, oricoEnvΔ680-699, oricoEnvΔ693-699 and coEnv mutants.

## References


1. **Arnaud, F., M. Varela, T. E. Spencer, and M. Palmarini.** 2008. Coevolution of endogenous betaretroviruses of sheep and their host. Cell Mol Life Sci **65:**3422-32.
2. **Bannert, N., and R. Kurth.** 2006. The evolutionary dynamics of human endogenous retroviral families. Annu Rev Genomics Hum Genet **7:**149-73.
3. **Bannert, N., D. Schenten, S. Craig, and J. Sodroski.** 2000. The level of CD4 expression limits infection of primary rhesus monkey macrophages by a T-tropic simian immunodeficiency virus and macrophagetropic human immunodeficiency viruses. J Virol **74:**10984-93.
4. **Barbulescu, M., G. Turner, M. I. Seaman, A. S. Deinard, K. K. Kidd, and J. Lenz.** 1999. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol **9:**861-8.
5. **Beimforde, N., K. Hanke, I. Ammar, R. Kurth, and N. Bannert.** 2008. Molecular cloning and functional characterization of the human endogenous retrovirus K113. Virology **371:**216-25.
6. **Belshaw, R., A. L. Dawson, J. Woolven-Allen, J. Redding, A. Burt, and M. Tristem.** 2005. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol **79:**12507-14.
7. **Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem.** 2004. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci U S A **101:**4894-9.
8. **Bieda, K., A. Hoffmann, and K. Boller.** 2001. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. J Gen Virol **82:**591-6.

9. **Blot, V., S. Lopez-Verges, M. Breton, C. Pique, C. Berlioz-Torrent, and M. P. Grange.** 2006. The conserved dileucine- and tyrosine-based motifs in MLV and MPMV envelope glycoproteins are both important to regulate a common Env intracellular trafficking. Retrovirology **3:**62.

10. **Boeke, M., and J.P. Stoye.** 1997. Retrotransposons, endogenous retroviruses and the evolution of retroelements., p. 343-436. *In* S. H. H. J.M. Coffin, and H.E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

11. **Boller, K., K. Schonfeld, S. Lischer, N. Fischer, A. Hoffmann, R. Kurth, and R. R. Tonjes.** 2008. Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. J Gen Virol **89:**567-72.

12. **Buzdin, A., S. Ustyugova, K. Khodosevich, I. Mamedov, Y. Lebedev, G. Hunsmann, and E. Sverdlov.** 2003. Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. Genomics **81:**149-56.

13. **Celma, C. C., M. G. Paladino, S. A. Gonzalez, and J. L. Affranchino.** 2007. Importance of the short cytoplasmic domain of the feline immunodeficiency virus transmembrane glycoprotein for fusion activity and envelope glycoprotein incorporation into virions. Virology **366:**405-14.

14. **Cheynet, V., A. Ruggieri, G. Oriol, J. L. Blond, B. Boson, L. Vachot, B. Verrier, F. L. Cosset, and F. Mallet.** 2005. Synthesis, assembly, and processing of the Env ERVWE1/syncytin human endogenous retroviral envelope. J Virol **79:**5585-93.

15. **Cote, M., Y. M. Zheng, L. M. Albritton, and S. L. Liu.** 2008. Fusogenicity of Jaagsiekte sheep retrovirus envelope protein is dependent on low pH and is enhanced by cytoplasmic tail truncations. J Virol **82:**2543-54.

16. **Dash, B., A. McIntosh, W. Barrett, and R. Daniels.** 1994. Deletion of a single N-linked glycosylation site from the transmembrane envelope protein of human

immunodeficiency virus type 1 stops cleavage and transport of gp160 preventing env-mediated fusion. J Gen Virol **75 ( Pt 6):**1389-97.

17. **de Parseval, N., V. Lazar, J. F. Casella, L. Benit, and T. Heidmann.** 2003. Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. J Virol **77:**10414-22.

18. **Demirov, D. G., and E. O. Freed.** 2004. Retrovirus budding. Virus Res **106:**87-102.

19. **Deml, L., A. Bojak, S. Steck, M. Graf, J. Wild, R. Schirmbeck, H. Wolf, and R. Wagner.** 2001. Multiple effects of codon usage optimization on expression and immunogenicity of DNA candidate vaccines encoding the human immunodeficiency virus type 1 Gag protein. J Virol **75:**10991-1001.

20. **Dewannieux, M., S. Blaise, and T. Heidmann.** 2005. Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. J Virol **79:**15573-7.

21. **Dewannieux, M., F. Harper, A. Richaud, C. Letzelter, D. Ribet, G. Pierron, and T. Heidmann.** 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res **16:**1548-56.

22. **Edwards, T. G., S. Wyss, J. D. Reeves, S. Zolla-Pazner, J. A. Hoxie, R. W. Doms, and F. Baribaud.** 2002. Truncation of the cytoplasmic domain induces exposure of conserved regions in the ectodomain of human immunodeficiency virus type 1 envelope protein. J Virol **76:**2683-91.

23. **Fenouillet, E., J. C. Gluckman, and I. M. Jones.** 1994. Functions of HIV envelope glycans. Trends Biochem Sci **19:**65-70.

24. **Freed, E. O., and M. A. Martin.** 1995. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. J Virol **69:**1984-9.

25. **Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem.** 2005. Evolution and distribution of class II-related endogenous retroviruses. J Virol **79:**6478-86.

26. **Goodman, L. J., S. R. Kain, and G. L. Firestone.** 1993. Trafficking of wild-type and an endoproteolytic-site mutant of the mouse mammary tumor virus glycoprotein. J Biol Chem **268:**2329-36.

27. **Green, N., T. M. Shinnick, O. Witte, A. Ponticelli, J. G. Sutcliffe, and R. A. Lerner.** 1981. Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide. Proc Natl Acad Sci U S A **78:**6023-7.

28. **Hakim, S. T., M. Alsayari, D. C. McLean, S. Saleem, K. C. Addanki, M. Aggarwal, K. Mahalingam, and O. Bagasra.** 2008. A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. Biochem Biophys Res Commun **369:**357-62.

29. **Heslin, D. J., P. Murcia, F. Arnaud, K. Van Doorslaer, M. Palmarini, and J. Lenz.** 2009. A single amino acid substitution in a segment of the CA protein within Gag that has similarity to human immunodeficiency virus type 1 blocks infectivity of a human endogenous retrovirus K provirus in the human genome. J Virol **83:**1105-14.

30. **Hughes, J. F., and J. M. Coffin.** 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. Genetics **171:**1183-94.

31. **Hughes, J. F., and J. M. Coffin.** 2004. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. Proc Natl Acad Sci U S A **101:**1668-72.

32. **Hunter, E., and R. Swanstrom.** 1990. Retrovirus envelope glycoproteins. Curr Top Microbiol Immunol **157:**187-253.

33. **Jha, A. R., S. K. Pillai, V. A. York, E. R. Sharp, E. C. Storm, D. J. Wachter, J. N. Martin, S. G. Deeks, M. G. Rosenberg, D. F. Nixon, and K. E. Garrison.** 2009. 371 Cross sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before Homo sapiens. Mol Biol Evol. doi:10.1093/molbev/msp180

34. **Johnston, P. B., J. W. Dubay, and E. Hunter.** 1993. Truncations of the simian immunodeficiency virus transmembrane protein confer expanded virus host range by removing a block to virus entry into cells. J Virol **67:**3077-86.

35. **Katzourakis, A., A. Rambaut, and O. G. Pybus.** 2005. The evolutionary dynamics of endogenous retroviruses. Trends Microbiol **13:**463-8.

36. **Kayman, S. C., R. Kopelman, S. Projan, D. M. Kinney, and A. Pinter.** 1991. Mutational analysis of N-linked glycosylation sites of Friend murine leukemia virus envelope protein. J Virol **65:**5323-32.

37. **Kim, F. J., N. Manel, Y. Boublik, J. L. Battini, and M. Sitbon.** 2003. Human T-cell leukemia virus type 1 envelope-mediated syncytium formation can be activated in resistant Mammalian cell lines by a carboxy-terminal truncation of the envelope cytoplasmic domain. J Virol **77:**963-9.

38. **Kowalski, M., J. Potz, L. Basiripour, T. Dorfman, W. C. Goh, E. Terwilliger, A. Dayton, C. Rosen, W. Haseltine, and J. Sodroski.** 1987. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. Science **237:**1351-5.

39. **Lee, Y. N., and P. D. Bieniasz.** 2007. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog **3:**e10.

40. **Lee, Y. N., M. H. Malim, and P. D. Bieniasz.** 2008. Hypermutation of an ancient human retrovirus by APOBEC3G. J Virol **82:**8762-70.

41. **Macfarlane, C., and P. Simmonds.** 2004. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. J Mol Evol **59:**642-56.

42. **Mager, D. L., and M. Patrik.** 2003. Retroviral Repeat Sequences. *In* D. N. Cooper (ed.), Nature encyclopedia of the human genome. Nature Pub. Group, London ; New York.

43. **Manrique, J. M., C. C. Celma, E. Hunter, J. L. Affranchino, and S. A. Gonzalez.** 2003. Positive and negative modulation of virus infectivity and envelope glycoprotein incorporation into virions by amino acid substitutions at the N terminus of the simian immunodeficiency virus matrix protein. J Virol **77:**10881-8.

44. **Mayer, J., T. Stuhr, K. Reus, E. Maldener, M. Kitova, F. Asmus, and E. Meese.** 2005. Haplotype analysis of the human endogenous retrovirus locus HERV-K(HML-2.HOM) and its evolutionary implications. J Mol Evol **61:**706-15.

45. **Megati, S., D. Garcia-Hand, S. Cappello, V. Roopchand, A. Masood, R. Xu, A. Luckay, S. Y. Chong, M. Rosati, S. Sackitey, D. B. Weiner, B. K. Felber, G. N. Pavlakis, Z. R. Israel, L. R. Smith, J. H. Eldridge, M. K. Sidhu, and M. A. Egan.** 2008. Modifying the HIV-1 env gp160 gene to improve pDNA vaccine-elicited cell-mediated immune responses. Vaccine **26:**5083-94.

46. **Mulligan, M. J., G. V. Yamshchikov, G. D. Ritter, Jr., F. Gao, M. J. Jin, C. D. Nail, C. P. Spies, B. H. Hahn, and R. W. Compans.** 1992. Cytoplasmic domain truncation enhances fusion activity by the exterior glycoprotein complex of human immunodeficiency virus type 2 in selected cell types. J Virol **66:**3971-5.

47. **Pinter, A., W. J. Honnen, and J. S. Li.** 1984. Studies with inhibitors of oligosaccharide processing indicate a functional role for complex sugars in the transport and proteolysis of Friend mink cell focus-inducing murine leukemia virus envelope proteins. Virology **136:**196-210.

48. **Puffer, B. A., S. Pohlmann, A. L. Edinger, D. Carlin, M. D. Sanchez, J. Reitter, D. D. Watry, H. S. Fox, R. C. Desrosiers, and R. W. Doms.** 2002. CD4 independence of

simian immunodeficiency virus Envs is associated with macrophage tropism, neutralization sensitivity, and attenuated pathogenicity. J Virol **76:**2595-605.

49. **Quinones-Kochs, M. I., L. Buonocore, and J. K. Rose.** 2002. Role of N-linked glycans in a human immunodeficiency virus envelope glycoprotein: effects on protein function and the neutralizing antibody response. J Virol **76:**4199-211.

50. **Reus, K., J. Mayer, M. Sauter, H. Zischler, N. Muller-Lantzsch, and E. Meese.** 2001. HERV-K(OLD): ancestor sequences of the human endogenous retrovirus family HERV-K(HML-2). J Virol **75:**8917-26.

51. **Ritter, G. D., Jr., M. J. Mulligan, S. L. Lydy, and R. W. Compans.** 1993. Cell fusion activity of the simian immunodeficiency virus envelope protein is modulated by the intracytoplasmic domain. Virology **197:**255-64.

52. **Tarlinton, R. E., J. Meers, and P. R. Young.** 2006. Retroviral invasion of the koala genome. Nature **442:**79-81.

53. **Tonjes, R. R., C. Limbach, R. Lower, and R. Kurth.** 1997. Expression of human endogenous retrovirus type K envelope glycoprotein in insect and mammalian cells. J Virol **71:**2747-56.

54. **Turner, G., M. Barbulescu, M. Su, M. I. Jensen-Seaman, K. K. Kidd, and J. Lenz.** 2001. Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr Biol **11:**1531-5.

55. **Vigerust, D. J., and V. L. Shepherd.** 2007. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol **15:**211-8.

56. **Wyss, S., A. S. Dimitrov, F. Baribaud, T. G. Edwards, R. Blumenthal, and J. A. Hoxie.** 2005. Regulation of human immunodeficiency virus type 1 envelope glycoprotein fusion by a membrane-interactive domain in the gp41 cytoplasmic tail. J Virol **79:**12231-41.

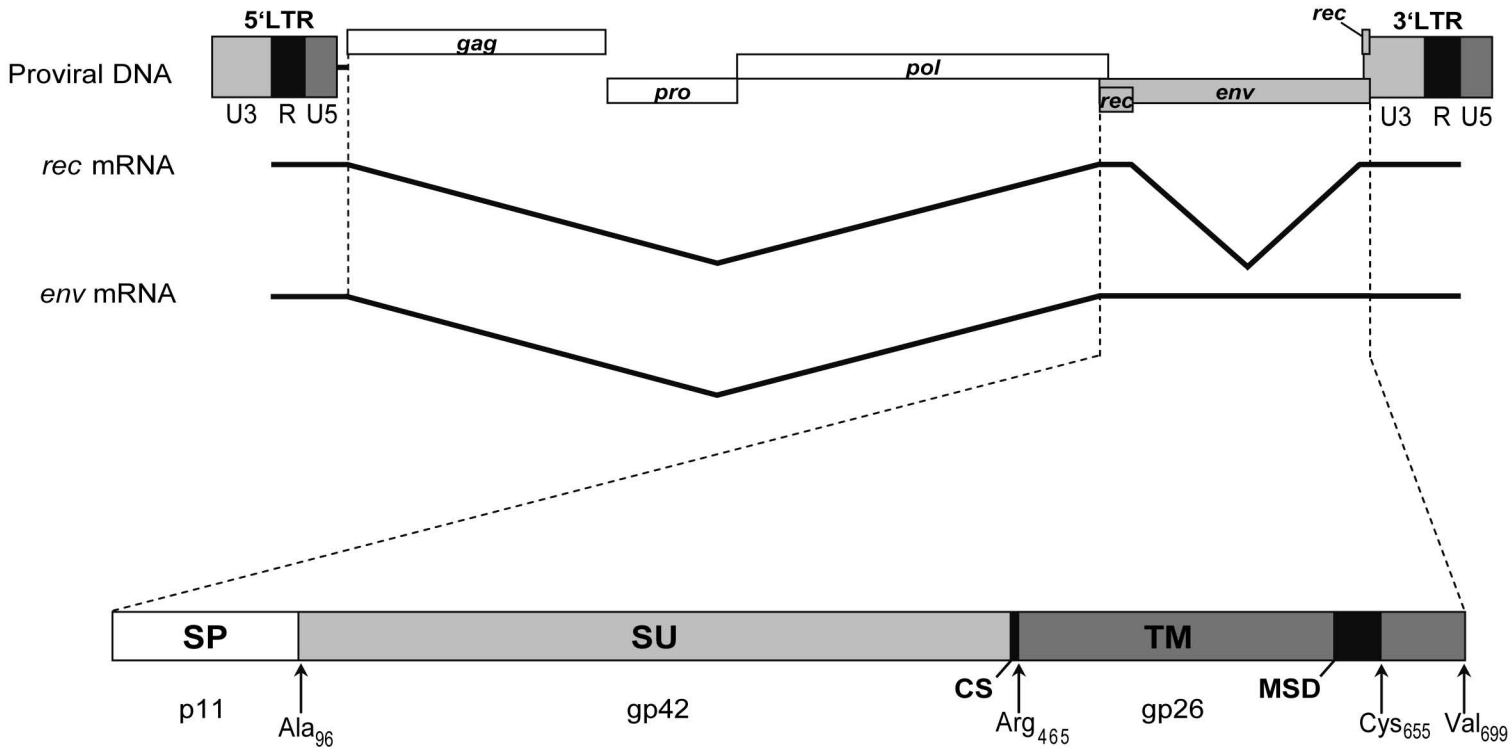
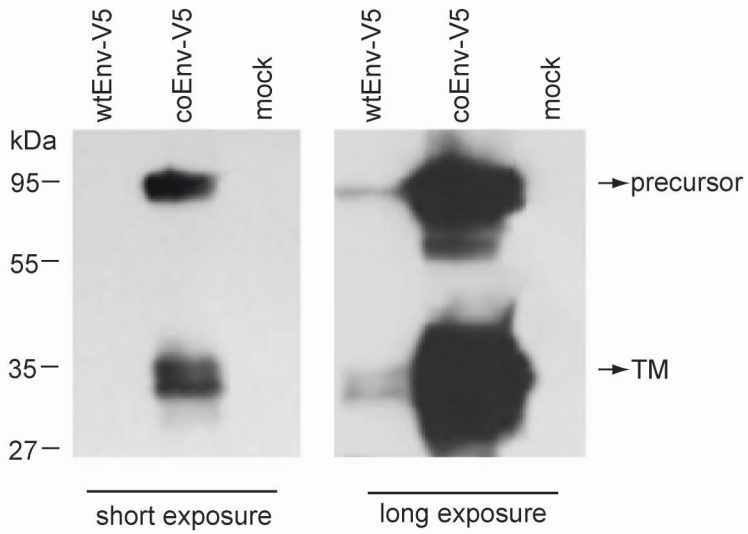

5'LTR
gag
pro
pol
rec
env
rec
3'LTR
Proviral DNA
U3 R U5
U3 R U5
rec mRNA
env mRNA
SP
SU
TM
p11
Ala96
gp42
CS
Arg465
gp26
MSD
Cys655
Val699

Figure 1

A


wtEnv-V5
coEnv-V5
mock
wtEnv-V5
coEnv-V5
mock
kDa
95
55
35
27
precursor
TM
short exposure
long exposure

B

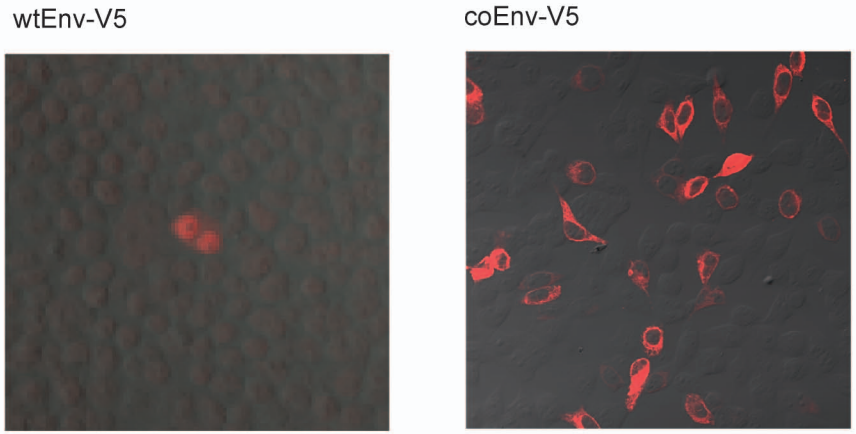

wtEnv-V5
coEnv-V5

Figure 2

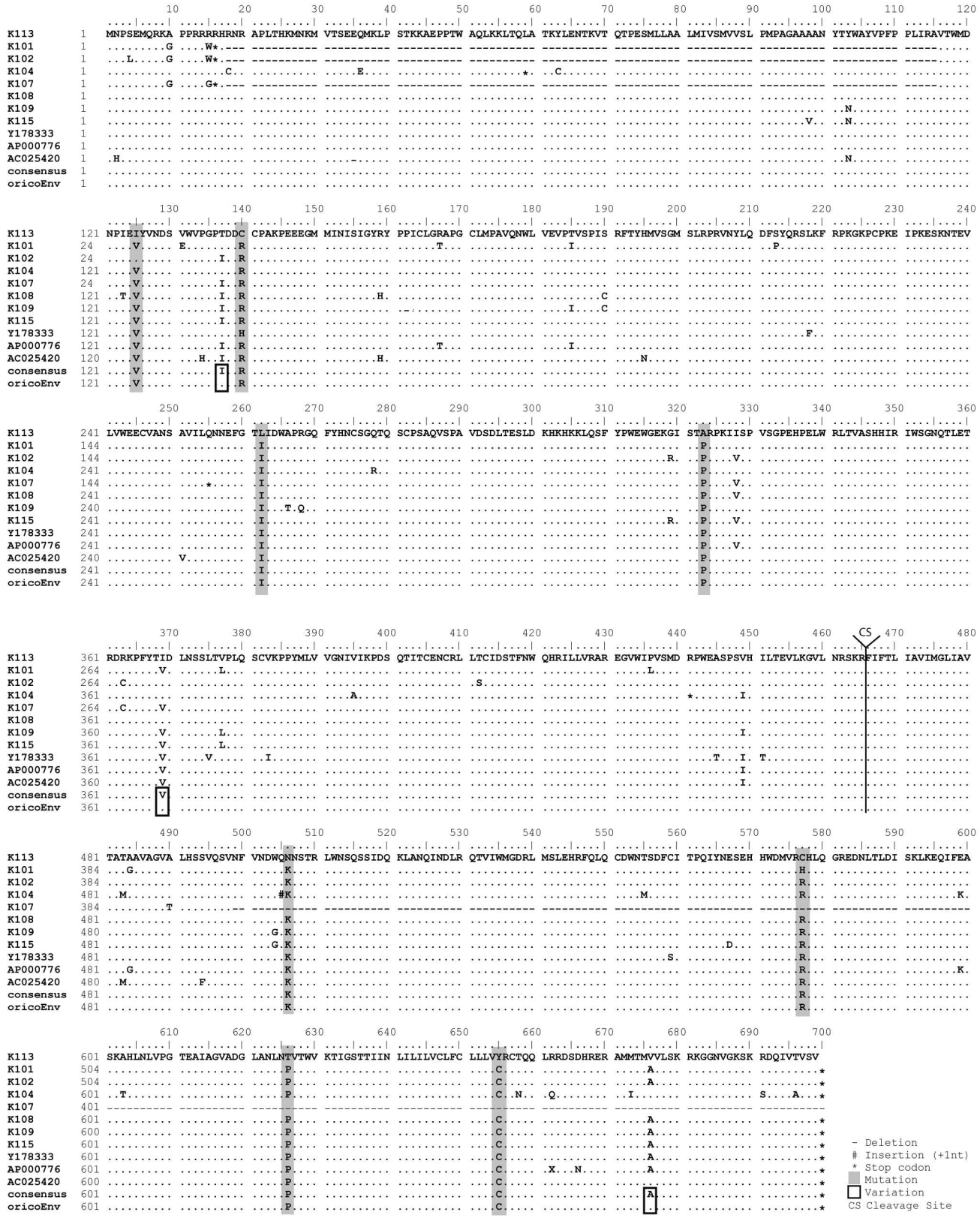

CS
- Deletion
# Insertion (+1nt)
* Stop codon
Mutation
Variation
CS Cleavage Site

Figure 3

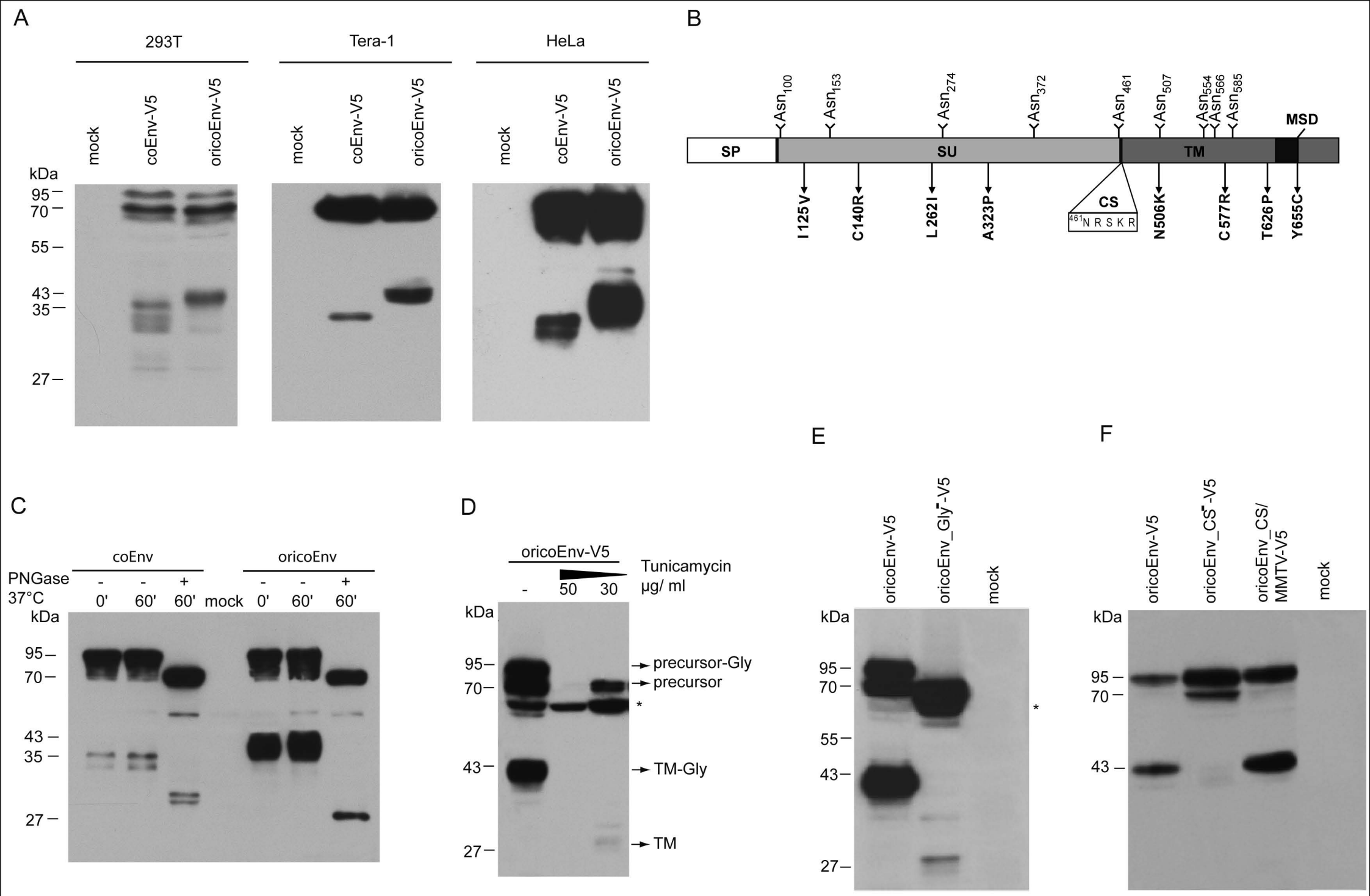
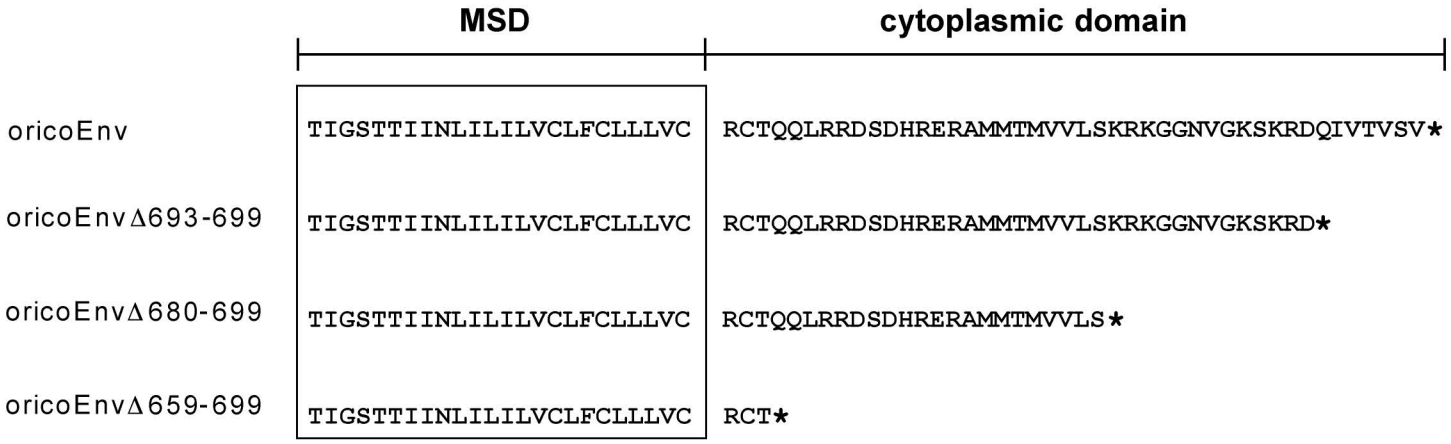

A
293T
Tera-1
HeLa
mock
coEnv-V5
oricoEnv-V5
kDa
95
70
55
43
35
27
B
Asn100
Asn153
Asn274
Asn372
Asn461
Asn507
Asn554
Asn566
Asn585
MSD
SP
SU
TM
CS
461N R S K R
I 125V
C140R
L 262I
A323P
N506K
C 577R
T626 P
Y655C
C
coEnv
oricoEnv
PNGase
37°C
0'
60'
mock
D
oricoEnv-V5
Tunicamycin
µg/ ml
precursor-Gly
precursor
TM-Gly
TM
E
oricoEnv_Gly-V5
F
oricoEnv_CS-V5
oricoEnv_CS/
MMTV-V5

Figure 4

A


MSD
cytoplasmic domain
oricoEnv
TIGSTTIINLILILVCLFCLLLVC
RCTQQLRRDSDHRERAMMTMVVLSKRKGGNVGKSKRDQIVTVSV*
oricoEnvΔ693-699
TIGSTTIINLILILVCLFCLLLVC
RCTQQLRRDSDHRERAMMTMVVLSKRKGGNVGKSKRD*
oricoEnvΔ680-699
TIGSTTIINLILILVCLFCLLLVC
RCTQQLRRDSDHRERAMMTMVVLS*
oricoEnvΔ659-699
TIGSTTIINLILILVCLFCLLLVC
RCT*

B

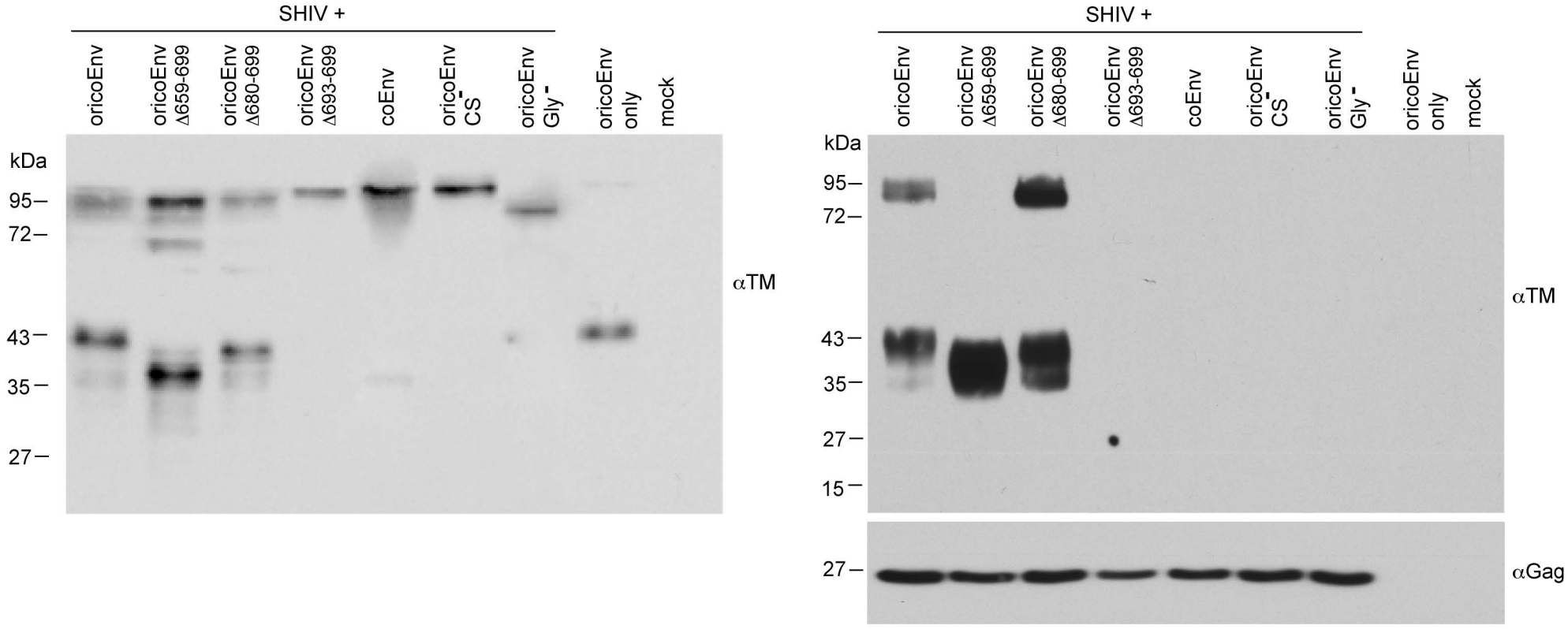

SHIV +
oricoEnv
oricoEnv Δ659-699
oricoEnv Δ680-699
oricoEnv Δ693-699
coEnv
oricoEnv CS
oricoEnv Gly-
oricoEnv only
mock
kDa
95
72
43
35
27
αTM
SHIV +
oricoEnv
oricoEnv Δ659-699
oricoEnv Δ680-699
oricoEnv Δ693-699
coEnv
oricoEnv CS
oricoEnv Gly-
oricoEnv only
mock
kDa
95
72
43
35
27
15
αTM
27
αGag

Figure 5

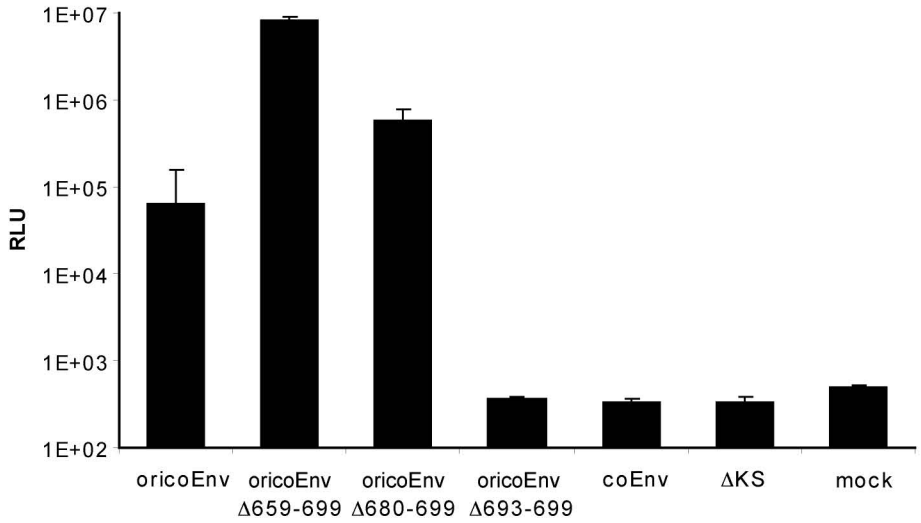
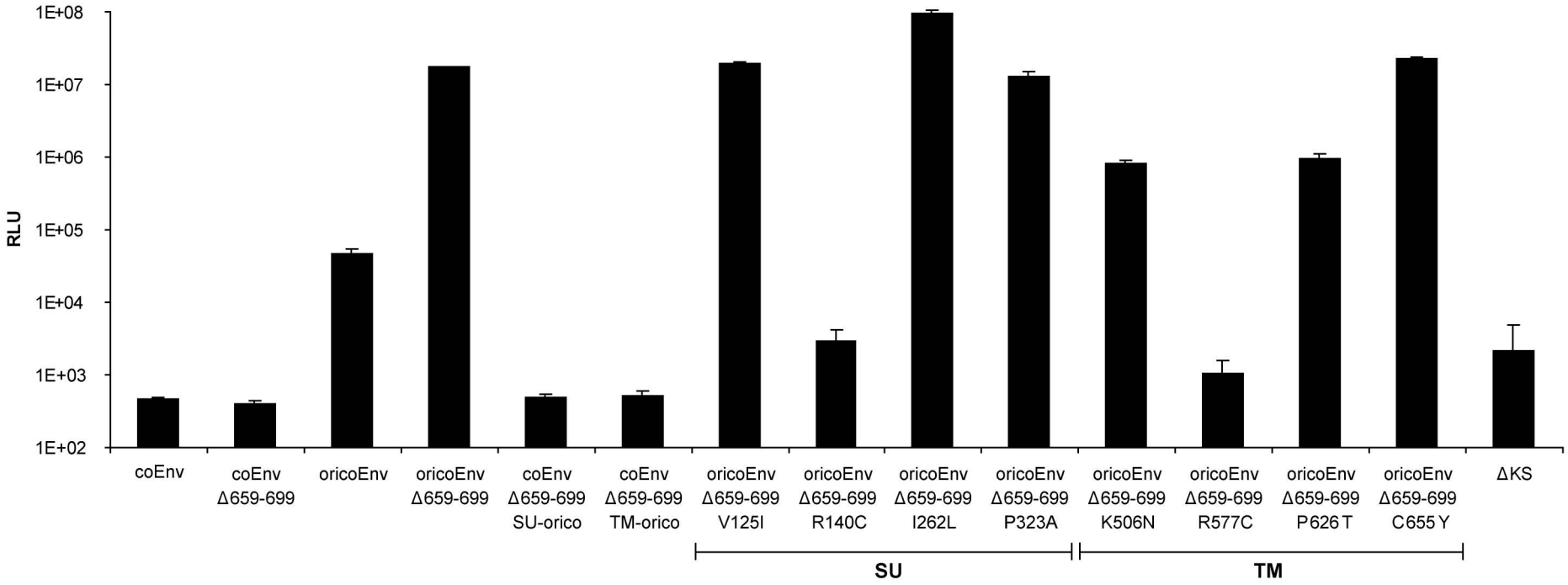

1E+07
1E+06
1E+05
1E+04
1E+03
1E+02
RLU
oricoEnv
oricoEnv Δ659-699
oricoEnv Δ680-699
oricoEnv Δ693-699
coEnv
ΔKS
mock

Figure 6

A


1E+08
1E+07
1E+06
1E+05
1E+04
1E+03
1E+02
RLU
coEnv
coEnv Δ659-699
oricoEnv
oricoEnv Δ659-699
coEnv Δ659-699 SU-orico
coEnv Δ659-699 TM-orico
oricoEnv Δ659-699 V125I
oricoEnv Δ659-699 R140C
oricoEnv Δ659-699 I262L
oricoEnv Δ659-699 P323A
oricoEnv Δ659-699 K506N
oricoEnv Δ659-699 R577C
oricoEnv Δ659-699 P626T
oricoEnv Δ659-699 C655Y
ΔKS
SU
TM

B

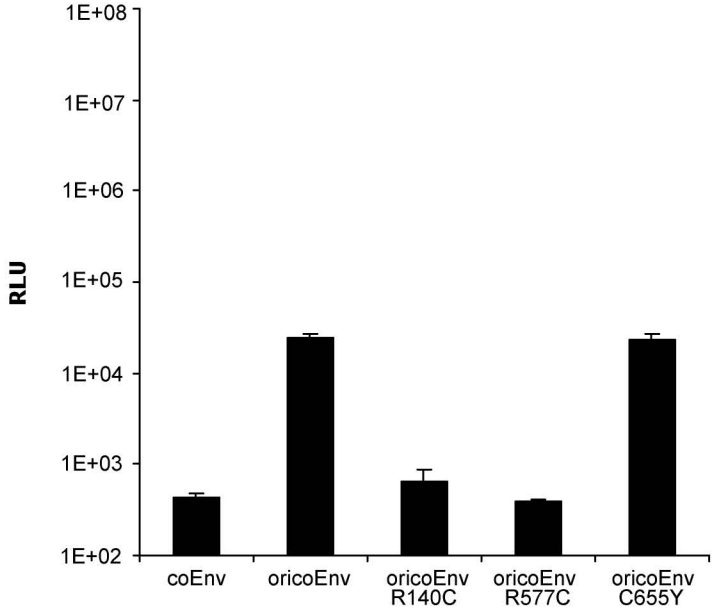

1E+08
1E+07
1E+06
1E+05
1E+04
1E+03
1E+02
RLU
coEnv
oricoEnv
oricoEnv R140C
oricoEnv R577C
oricoEnv C655Y

Figure 7

ATGAACCCCAGCGAGATGCAGCGGAAGGCCCCTCCCAGGCGGCGGAGACACCGGAACAGAGCCCCCCTGACCCAC
AAGATGAACAAGATGGTGACCAGCGAGGAACAGATGAAGCTGCCCAGCACCAAGAAGGCCGAGCCCCCCACCTGG
GCCCAGCTGAAGAAGCTGACCCAGCTGGCCACCAAGTACCTGGAAAACACCAAGGTGACCCAGACCCCCGAGAGC
ATGCTGCTGGCCGCCCTGATGATCGTGAGCATGGTGGTGTCCCTGCCCATGCCTGCCGGCGCCGCGGCTGCCAAC
TACACCTACTGGGCCTACGTGCCCTTCCCCCCCCTGATCCGGGCCGTGACCTGGATGGACAACCCCATCGAGATC
TACGTGAACGACAGCGTGTGGGTGCCAGGCCCCACCGACGACTGCTGCCCCGCCAAGCCCGAGGAAGAGGGCATG
ATGATCAACATCAGCATCGGCTACAGATACCCCCCCATCTGCCTGGGCAGGGCCCCTGGCTGCCTGATGCCCGCC
GTGCAGAACTGGCTGGTGGAGGTGCCCACCGTGAGCCCCATCAGCCGGTTCACCTACCACATGGTGTCCGGCATG
AGCCTGCGGCCCAGGGTTAACTACCTGCAGGACTTCAGCTACCAGCGGAGCCTGAAGTTCCGGCCCAAGGGCAAG
CCCTGCCCCAAAGAGATCCCCAAAGAGTCCAAGAACACCGAGGTGCTGGTGTGGGAGGAATGCGTGGCCAACAGC
GCCGTGATCCTGCAGAACAACGAGTTCGGCACCCTGATCGACTGGGCCCCCAGGGGCCAGTTCTACCACAACTGC
AGCGGCCAGACCCAGAGCTGCCCCAGCGCCCAGGTGTCCCCTGCCGTGGACAGCGACCTGACCGAGAGCCTGGAC
AAGCACAAGCACAAAAAGCTGCAGAGCTTCTACCCCTGGGAGTGGGGCGAGAAGGGGATATCCACCGCCAGGCCC
AAGATCATCAGCCCCGTGAGCGGCCCTGAGCACCCCGAGCTGTGGCGGCTGACCGTGGCCAGCCACCACATCCGG
ATCTGGTCCGGCAACCAGACCCTGGAAACCCGGGACCGGAAGCCCTTCTACACCATCGACCTGAACAGCAGCCTG
ACCGTGCCCCTGCAGAGCTGCGTGAAGCCCCCCTACATGCTGGTGGTGGGCAACATCGTGATCAAGCCCGACAGC
CAGACCATCACCTGCGAGAACTGCCGGCTGCTGACCTGCATCGACAGCACCTTCAACTGGCAGCACCGGATCCTG
CTCGTACGGGCCAGGGAAGGCGTGTGGATCCCCGTGTCCATGGACCGGCCCTGGGAGGCCAGCCCCAGCGTGCAC
ATCCTGACAGAGGTGCTGAAGGGCGTGCTGAACAGAAGCAAGCGGTTCATCTTTACCCTGATCGCCGTGATCATG
GGCCTGATTGCCGTGACCGCCACAGCCGCCGTGGCCGGCGTGGCCCTGCACAGCTCTGTACAGAGCGTGAACTTC
GTGAACGACTGGCAGAACAACAGCACCCGGCTGTGGAACAGCCAGAGCAGCATCGACCAGAAGCTGGCCAACCAG
ATCAACGACCTGCGGCAGACCGTGATTTGGATGGGCGACCGGCTGATGAGCCTGGAACACCGGTTCCAGCTGCAG
TGCGACTGGAACACCAGCGATTTCTGCATCACCCCCCAGATCTACAACGAGAGCGAGCACCACTGGGACATGGTG
CGGTGCCACCTGCAGGGCCGGGAGGACAATCTGACCCTGGACATCAGCAAGCTGAAAGAGCAGATCTTCGAGGCC
AGCAAGGCCCATTTAAATCTGGTGCCCGGCACCGAGGCCATCGCCGGGGTGGCCGACGGCCTGGCCAACCTGAAC
ACCGTGACATGGGTGAAAACCATCGGCAGCACCACCATCATCAACCTGATCCTGATCCTGGTGTGCCTGTTCTGC
CTGCTGCTGGTGTATCGGTGCACCCAGCAGCTGCGGAGAGACAGCGACCACCGGGAGCGGGCCATGATGACCATG
GTGGTGCTGTCCAAGCGGAAGGGCGGCAACGTGGGCAAGAGCAAGCGGGACCAGATCGTGACCGTGAGCGTGTGA

Figure S1

| Primer name | sequence (5`→3`) |
|---|---|
| **oricoEnv- mutagenesis** | |
| coEnv_I125V | AAC CCC ATC GAG GTC TAC GTG AAC GAC |
| coEnv_C140R | GTG CCA GGC CCC ATC GAC GAC CGC TGC CCC GCC AAG CCC |
| coEnv_L262I | ACA ACG AGT TCG GCA CCA TTA TCG ACT GGG CCC CCA G |
| coEnv_A323P | GGG ATA TCC ACC CCC AGG CCC AAG ATC GTC AGC CCC GTG AGC GGC |
| coEnv_N506K | GTG AAC GAC TGG CAG AAA AAC AGC ACC CGG |
| coEnv_C577R | GAC ATG GTG CGG CGC CAC CTG CAG GGC |
| coEnv_T626P | GCC AAC CTG AAC CCC GTG ACA TGG GTG |
| coEnv_Y655C | CTG CTG CTG GTG TGT CGG TGC ACC CAG |
| **coEnv- mutagenesis** | |
| coEnvΔ659-699 | GTG TAT CGG TGC ACC TAG CAG CTG CGG AGA G |
| oricoEnv_V125I | AC AAC CCC ATC GAG ATC TAC GTG AAC GAC |
| oricoEnv_R140C | GGC CCC ACC GAC GAC TGC TGC CCC GCC AAG C |
| oricoEnv_I262L | AAC GAG TTC GGC ACC CTT ATC GAC TGG GCC C |
| oricoEnv_P323A | AG GGG ATA TCC ACC GCC AGG CCC AAG ATC |
| oricoEnv_K506N | G AAC GAC TGG CAG AAC AAC AGC ACC CGG CTG |
| oricoEnv_R577C | TGG GAC ATG GTG CGG TGC CAC CTG CAG GGC C |
| oricoEnv_P626T | TG GCC AAC CTG AAC ACC GTG ACA TGG GTG |
| oricoEnv_C655Y | CTGCCTGCTGCTGGTGTATCGGTGCACCCAGCAGC |
| **oricoEnvCS$^-$ - mutagenesis** | |
| oricoEnv_ΔCS1384 | GTG CTG AAG GGC GTG CTG AAC GCA AGC GCG GCG TTC ATC TTT ACC CTG ATC GCC |
| oricoEnv_MMTV1387 | GTG CTG AAG GGC GTG CTG AAC AGA GCC AAG CGG TTC ATC TTT ACC CTG ATC |
| **oricoEnvGly$^-$ - mutagenesis** | |
| ΔGly1_oricoEnv_SU304 | GCG GCT GCC AAC TAC GCC TAC TGG GCC TAC G |
| ΔGly2_oricoEnv_SU463 | ATG ATG ATC AAC ATC GCC ATC GGC TAC AGA TA |
| ΔGly3_oricoEnv_SU826 | TTC TAC CAC AAC TGC GCC GGC CAG ACC CAG AG |
| ΔGly4_oricoEnv_SU1120 | ATC GAC CTG AAC AGC GCC CTG ACC GTG CCC CT |
| ΔGly5_oricoEnv_TM1525 | TGG CAG AAA AAC AGC GCC CGG CTG TGG AAC A |
| ΔGly6_oricoEnv_TM1666 | TGC GAC TGG AAC ACC GCC GAT TTC TGC ATC AC |
| ΔGly7_oricoEnv_TM1702 | CAG ATC TAC AAC GAG GCC GAG CAC CAC TGG GA |
| ΔGly8_oricoEnv_TM1759 | CGG GAG GAC AAT CTG GCC CTG GAC ATC AGC A |
| **oricoEnvΔCT- mutagenesis** | |
| oricoEnvΔ659-699_1975 | GTG TGT CGG TGC ACC TAG CAG CTG CGG AGA G |
| oricoEnvΔ680-699_2038 | ATG GTG GTG CTG TCC TAG CGG AAG GGC GGC A |
| oricoEnvΔ693-699_2077 | AAG AGC AAG CGG GAC TAG ATC GTG ACC GTG A |

Figure S2