Identification and Analysis of Cleavage Sites and Subdomains in the Group Specific Antigen (Gag) of the Human Endogenous Retrovirus-K(HML-2)

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von

Maja Muguluvila George
aus Berlin

2014

List of Figures

1. Summary

Interestingly at molecular level humans and chimpanzees are very similar. The average DNA sequence difference between them is ~4% (Varki and Altheide, 2009). Contributing to the difference in humans are human endogenous retroviruses (HERV), which constitute about 8% of our genome. The young age and the similarity to other cancer causing betaretroviruses such as MMTV have led to intense research on the influence of some HERVs on the human genome as well as oncogenic aspects of HERV encoded proteins. Fundamental properties of these ancient retroviruses although have remained undiscovered, due to the low expression of the retroviral proteins. One example is the Gag polyprotein, the major structural protein of all retroviruses with the ability to form virus like particles in the absence of any other retroviral protein. The aim of this project was to decipher the processing sites of the Group specific antigen (Gag) protein of a specific member of the HERV family, HERV-K113, and to identify all cleavage products.

Although the provirus of HERV-K113 encodes open reading frames for genes of a complex retrovirus and produces mature virions on low expression levels, it has accumulated mutations which render it uninfectious. To elevate the expression non-synonymous post-integration mutations in the *gag-pro-pol* region of HERV-K113 had previously been identified by a specific algorithm and reconstituted. The amino acid sequence had then been codon optimized for mammalian cell. Gag was cloned together with the protease and polymerase sequence into the vector pcDNA3.1 to set it under the control of a CMV-promoter. Expression of this new construct, oricoHERV-K113_GagProPol, in mammalian cells led to an enhanced viral protein expression as seen on Western Blot and in Electron microscopic pictures of purified viral pellets. The protease defective construct oricoHERV-K113_GagPro$^-$Pol was established to investigate the maturation process. Comparison of protease defective viral proteins to protease active viral proteins on silver nitrate stained SDS gel analysis revealed the processing of Gag proteins and the maturation of VLPs after the release from cells. Alignments of HERV-K113 Gag with the sequences of two beta retroviruses, MPMV and MMTV, had already shown that the endogenous retroviral Gag harbours the typical domains matrix (MA), capsid (CA) and nucleocapsid (NC) as well as a phosphorylated protein with a late domain between MA and CA. The high expression rate of oricoHERV-K113 enabled the investigation of the exact Gag cleavage sites. Therefore viral proteins were separated by reverse phase high pressure liquid chromatography. Fractions were screened for Gag subdomains on western blot with

previously established antibodies and, in some cases, by mass spectrometry. Proteins suspected to contain a subdomain were N-terminally sequenced, to determine the N-terminal cleavage site amino acid sequence. Additional to the common retroviral Gag subdomains MA, CA and NC the subdomain p15 was confirmed between MA and CA. Typical for a late domain encoding protein p15 was shown to be phosphorylated by mass spectrometry. All cleavage sites except the one between MA and p15 belong to type 1 or type 2 cleavage sites. Mutation of the P1 position at the cleavage sites led to changes in cleavage patterns detected on Western Blot. Surprisingly two new short proteins were discovered following NC. Both are rich in glutamine and proline and hence were named QP1 and QP2.

In summary the reconstituted protein sequence of oricoHERV-K113_GagProPol facilitated the discovery that the HERV-K113 Gag is cleaved by the protease to release the subdomains MA, p15, CA, NC, QP1 and QP2. The functions of QP1 and QP2 have to be investigated in the future. Since processing is a prerequisite for replication, analysis of the subdomains and their processing could lead to the discovery why no original or improved HERV sequence until now has been shown to replicate.

2. Introduction

2.1 Retroviruses

The family of retroviruses (*Retroviridae*) has been named after their main characteristic enzyme, the reverse transcriptase (RT). They are enveloped viruses with a homodimer of linear single-stranded positive sense RNA as their genome. Their ability to integrate their genetic information into the genome of the host cell after being reverse transcribed into DNA by the RT sets them apart from any other virus family.

Retroviruses were first studied approximately 100 years ago, when V. Ellerman, O. Bang and P. Rous produced cell free extracts from chicken tumour cells and showed that injection of these extracts caused tumours in healthy chicken (Ellermann V., 1908; Vogt, 1997). John Bittner proved in 1936 for the first time that a virus, now known as the mouse mammary tumour virus, induced the growth of tumours in mice and was transmissible through the milk of the mother. The enzyme RT was characterized in 1970 when Howard Temin and David Baltimore described its mechanism independently (Baltimore, 1970; Temin and Mizutani,

1970). Hence the viruses, which encode the RT, were named reverse transcriptase oncoviruses or short, retroviruses. Since then numerous retroviruses have been discovered which are able to infect vertebrates. In humans the retrovirus Human T lymphotropic virus type 1 (HTLV-1) which causes adult T-cell leukaemia (ATL) is the best known tumour virus (Poiesz et al., 1980). But by far the most well-known human retrovirus in general is the Human Immunodeficiency Virus (HIV) which induces the Acquired Immunodeficiency Syndrome (AIDS) (Barré-Sinoussi et al., 1983).

The taxonomic classification of retroviruses has been formalized by the International Committee on Taxonomy of Viruses (ICTV) as stated in figure 2.1. They have divided retroviruses based on their genetic heritance and distinguish between *Orthoretroviridae*, which comprises the alpha-, beta-, gamma-, delta-, epsilon-Retroviruses as well as lentiviruses, and *Spumaretroviridae*, to which spuma-, better known as foamyviruses belong to. Additionally retroviruses are differentiated into simple and complex viruses according to the composition of their main structural proteins, the enzymes and the envelope glycoproteins. Complex viruses carry additional proteins and essential factors with regulatory function, e.g. some can modulate the host's reaction to the virus or counteract the antiviral defense mechanisms.

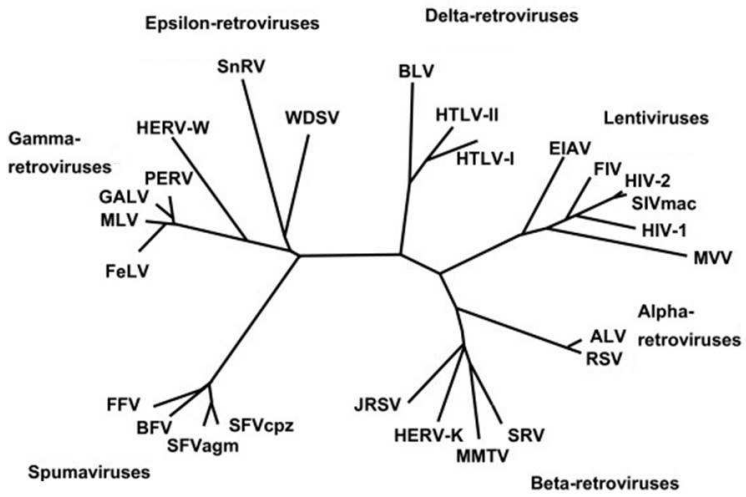


Fig. 2-1 Phylogeny of retroviruses. Phylogenetic tree according to the comparison of the amino acid sequence of the RT. ALV- avian leukosis virus; BFV- bovine foamy virus; BLV- bovine leukaemia virus; EIAV- equine infectious anaemia virus; FFV- feline foamy virus; FeLV- feline leukemia virus; FIV- feline immunodeficiency virus; HERV-K/-W- human endogenous retrovirus-K/-W; GALV- gibbon ape leukaemia virus; HIV-1- human immunodeficiency virus-1; HIV-2- human immunodeficiency virus-2; HTLV-1- human T-cell leukaemia virus type 1; HTLV-2- human t-cell leukaemia virus 2; JSRV- Jaagsiekte sheep retrovirus; MLV- murine leukaemia virus; MMTV- mouse mammary tumour virus; MPMV- Mason Pfizer monkey virus; MVV- Maedi-Visna virus; PERV- porcine endogenous retrovirus; RSV- rous sarcoma virus; SFVagm- simian foamy virus African green monkey; SFVcpz- simian foamy virus chimpanzee; SIVmac- simian immunodeficiency virus Macaque; SRV- simian retrovirus; SnRV- snakehead retrovirus; WDSV- walleye dermal sarcoma virus; Visna- visna virus; (adapted from (Weiss, 2006))

2.2 Morphology

The morphology of all mature retroviral virions resembles each other. They are roughly spherical and between 100-150 nm in size (Kurth and Bannert, 2010). The outside of the virion is studded with envelope (Env) proteins, which are embedded in the lipid bilayer of the virion envelope and in the shell consisting of matrix (MA) proteins. Inside the shell is the protein core composed by capsid (CA) proteins, which surrounds the highly condensed RNA genome. Associated with the genome are the nucleocapsid (NC) proteins. Additionally there are the enzymes RT, the integrase (IN) and the protease (PR) inside the capsid (de Marco et al., 2010).

The structural proteins MA, CA and NC are subdomains of a polyprotein, the Group specific antigen (Gag) (Willsand and Craven, 1991). Gag is processed by the protease after the virion

has been released from the cell. This process is called maturation. An immature virion features a characteristic electron lucent core. Upon cleavage of Gag by the protease the morphology of the virion changes dramatically (Demirov and Freed, 2004). The now mature virus has an electron dense core and a distinct outer shell as shown schematically in figure 2.2.

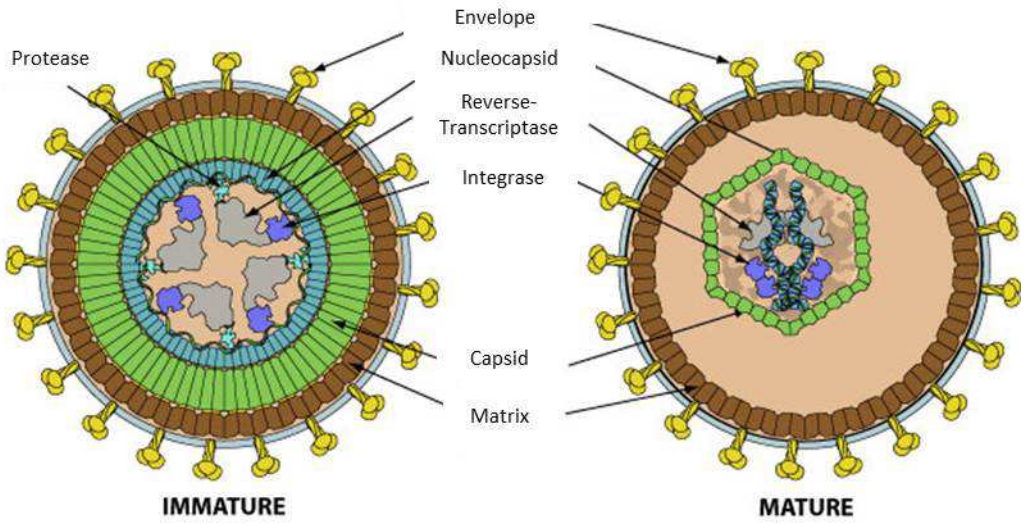


Fig. 2-2 Schematic picture of the morphology of immature and mature retroviruses. The morphology of an immature capsid is spherical with radially oriented Gag polyproteins. The N-terminal domain of Gag, MA, is anchored in the lipid bilayer. The RNA genome is in contact with the NC proteins. During or after release the Gag precursor is cleaved. This leads to the conversion of the particle into a virus with a distinct core consisting of CA proteins which contain several enzymes such as the Polymerase as well as the single stranded RNA genome associated with NC. The envelope is studded with trimeric Env proteins. (taken from http://viralzone.expasy.org/all_by_species/66.html)

2.3 The retroviral genome and proteins

The retroviral genome consists of a homodimer of two identical RNA sequences, which are capped at the 5' end (Zhou et al., 2003) and carry a polyadenylation (poly A) sequence at the 3' end. Every retrovirus encodes the *gag, pro, pol* and *env* genes. These genes are flanked by regulatory sequences which are important for the reverse transcription and the integration into the cellular host DNA. Following the 5'cap as well as again immediately before the poly A sequence is the R region. Downstream of the 5`R region is the so called unique region (U5), which encodes sequences important for the integration of the provirus into the host genome. Next to the U5 Region is the primer binding (pbs) site. This 18 nucleotides long sequence

hybridizes a host t-RNA to the genome, which will serve for the initiation of the transcription of the minus strand by the RT. The region between the pbs and the gag gene encodes often a splice donor site for the production of subgenomic mRNA. This spliced mRNA serves mostly for the production of the Env protein and, in case of complex viruses like lentiviruses, for additional regulatory proteins. Just upstream of *gag* is the Psi element, which is necessary for the encapsidation of the viral genome into the virion. Downstream of *env* lays the polypurine tract (ppt) with at least nine Adenosines and Guanines, which are important for initiation of the reverse transcription. Just before the 3' R region will be the U3 region. Similar to the U5 region it encodes sequences for the integration and is essential for the gene expression of the integrated Provirus (Goff, 2001). A simplified overview of the RNA genome is shown in figure 2.3.

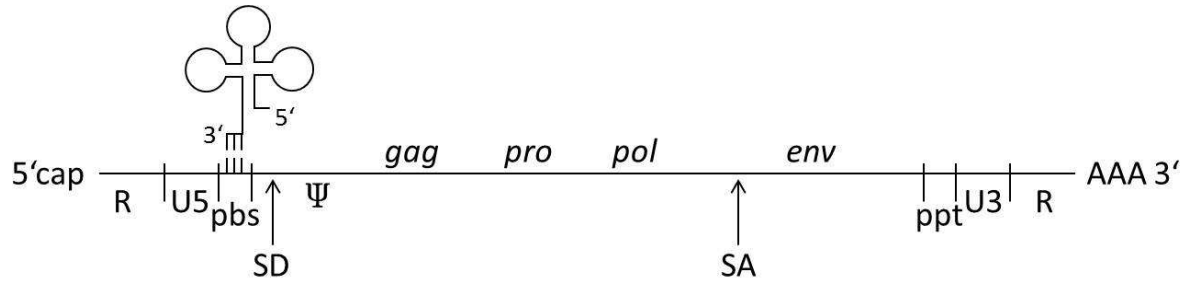


Fig. 2-3 Organization of the retroviral RNA genome. The single stranded RNA is depicted as a line with the genes *gag*, *pol* and *env* as well as the regulatory region encoded at the 5' and 3' end. R- repeated region; U- unique region; pbs- primer binding site; SD– splice donor; **Ψ**- Psi element; SA- splice acceptor; ppt– polypurine tract. (adapted from (Knipe and Howley, 2013))

The gag gene encodes the Gag polyprotein. This precursor protein is cleaved by the protease after the virion is released from the cell to give rise to the mature viral proteins MA, CA and NC, as well as additional proteins. The position for the gene of the Protease, *pro*, is in almost all retroviral genomes between the gag and pol genes. But it can be fused to the gag gene, as well as to the *pol* gene or simply possess its own reading frame. The *pol* gene gives rise to a precursor polyprotein which is also cleaved like Gag into several proteins during maturation. These proteins are only necessary in low quantities, such as the RT and the Integrase enzymes.

To keep the balance between the different proteins which are part of the virion, there are two mechanisms employed, translational read-through, and translational frameshifting, which both lead to the translation of less Pro and Pol proteins than Gag. For the translational read-through

the open readings frames (ORF) of genes are in the same reading frame but separated by a single UAG. This stop codon is suppressed about 5-10% of the time and leads thereby to the rise of the fusion proteins comprising Gag, Pro and Pol. The translational read-through is employed by the gamma and epsilon retroviruses, which have the Gag ORF and the Pro-Pol ORF separated by UAG (Balvay et al., 2007; *Fields Virology*, 2006).

For the translational frameshifting the ORFs lie in different reading frames. Depending on the virus there is a sequence of homopolymeric bases, consisting of oligo A or oligo U, which is called slippery site at the end of *gag* and or *pol*. At this sequence the ribosome slips back one nucleotide and thereby changes into a new reading frame. In the case of alpha- and lentivurses *gag* and *pol* are encoded in different reading frames. Betaretroviruses and deltaretroviruses have separate ORFs for *gag*, *pro* and *pol* and need two successive frameshifts for the Gag-Pro-Pol fusion protein (Jacks et al., 1988; Moore et al., 1987)

2.4 The main structural retroviral protein Gag

All retroviruses encode the gene for the viral polyprotein Gag at the 5‘end of the viral genome. Interestingly the Gag precursor is the sole protein necessary for the formation of virion-like particles (VLP) (Delchambre et al., 1989; Gheysen et al., 1989; Wills and Craven, 1991). It is synthesized from the unspliced genomic RNA, assembles by itself in the presence of nucleic acids and is released even if all other viral proteins are absent. The assembly takes either place at the plasma membrane or in the cytosol (Goff, 2001). Type C retroviruses which comprise alpharetroviruses, gammaretroviruses and lentiviruses assemble into electron-dense structures at the plasma membrane, whereas Type B and D retroviruses assemble in the cytosol. Retroviral Gag precursor proteins, except those of spumaretroviruses, are cleaved during or after the VLPs have been released from the cell. There are three main subdomains always present after cleavage: MA, CA and NC (Leis et al., 1988). The order of these proteins in the Gag protein is conserved across all retroviruses (Fig. 2.4) and correlates with the order in which the mature viral proteins MA, CA and NC are located in the virion (Fig. 2-2). The main structural Gag subdomains are sometimes separated by additional peptide sequences of often unknown function. The most noticeable is a protein located between MA and CA produced by some alpha-, beta- and gamma retroviruses. These often called phosphoproteins contain sometimes one or two late assembly domains, which mediate virus release (Demirov

and Freed, 2004). Additionally other proteins shorter than 5 kDa, often named spacer protein (SP), can be located between major subdomains.

2.4.1 Matrix (MA)

All retroviral MA proteins consist of four α-helices, which form a globular core (Murray et al., 2005). Although retroviral MA proteins do not share sequence homology all of them are able to target the Gag polyprotein to the inner leaflet of the plasma membrane. One important feature for this ability is for most retroviruses, such as HIV (Medizin et al., 2008; Pal et al., 1990), M-PMV (Schultz and Oroszlant, 1983; Schultz and Rein, 1989) or MLV (Henderson et al., 1983), the posttranslational N-terminal myristylation. Additionally all retroviruses share a patch of basic amino acids, which is also implicated in the anchorage of MA in the plasma membrane (Murray et al., 2005). Indeed exchanging the basic amino acids with differently charged amino acids resulted in no particle production in HTLV-1 MA (Le Blanc et al., 1999) and FIV MA (Manrique et al., 2001). Murray et al. 2005 showed in structural models that the mutation led to an impaired MA/membrane association (Murray et al., 2005). It is speculated that in retroviruses regularly lacking a basic patch in their MA, such as Rous sarcoma virus (RSV) or equine infectious anemia virus, the basic patch is sufficient for the association with the plasma membrane (Kurth and Bannert, 2010).

2.4.2 Capsid (CA)

The exterior shell of the retroviral core in all mature retroviruses, which contains the retroviral genome with associated NC proteins as well as the retroviral enzymes and additional proteins is formed by a conglomerate of CA proteins. The mature core morphology varies widely. It can be cone shaped as in HIV, roughly spherical as in betaretroviruses, polygonal or cylindrical. The CA domain is the largest Gag cleavage product and is encoded downstream of MA (Fig. 2-3). It is composed of two α-helical domains, called N-terminal domain (NTD) and C-terminal domain (CTD) (Gamble, 1997; Kingston et al., 2000). The NTD forms an N-terminal β-hairpin after the proteolytic processing of immature Gag, which triggers the assembly of the mature core (Gitti et al., 1996). The most highly conserved sequence among all retroviruses, the major homology region (MHR) is encoded in the CTD. The MHR consists of a sequence of approximately 20 amino acids, which is conserved among all replication-competent orthoretroviruses. The exact function of the MHR still needs to be elucidated (Kurth and Bannert, 2010).

2.4.3 Nucleocapsid (NC)

The NC protein is a small highly basic protein encoded in all orthoretroviruses downstream of CA. The main features of NC are one or two Cystidin-Histidine or CCHC motifs that coordinate a single zinc ion, called zinc knuckle domain. Through the zinc knuckle domains NC binds specifically to the packaging signal on genomic retroviral RNA (Amarasinghe et al., 2000; D'Souza and Summers, 2005; De Guzman et al., 1998) and thereby facilitate the encapsidation of the viral genome (Méric and Goff, 1989). But it can also bind non-specifically to nucleic acids in the host cell (Bowzard et al., 1998; Cimarelli et al., 2000; Ganser-Pornillos et al., 2008) and supports the use of RNA as a structural element in retroviral particles (Muriaux and Darlix, 2010; Muriaux et al., 2001). During reverse transcription NC contributes to tRNA positioning (Li et al., 1996) and the specific transfer of DNA strands (Allain et al., 1994; Rodríguez-Rodríguez et al., 1995).

2.4.4 Additional proteins encoded in Gag

Beside the major Gag subdomains, which all orthoretroviruses share, different retroviruses exhibit a variety of additional peptides after cleavage of the precursor Gag protein. They are either encoded between the MA, CA or NC or C-terminal of Gag. Members of the alpha-, beta and gammaretroviruses for example encode between Ma and CA the 10-20kDA phosphoproteins. These proteins, as in the case of RSV or M-PMV, sometimes contain late assembly domains important for viral release. Also very frequent are subdomains encoded at the C-terminus of Gag, as the p6 protein in HIV or the p4 protein of M-PMV (Hizi et al., 1989). In the case of HIV the p6 domain encodes the late assembly domain (Müller et al., 2002). Additionally there are spacer proteins, p1 and p2 released after cleavage by the protease between CA/NC and NC/p6. The function of many of the spacer proteins still needs to be elucidated (Kurth and Bannert, 2010).

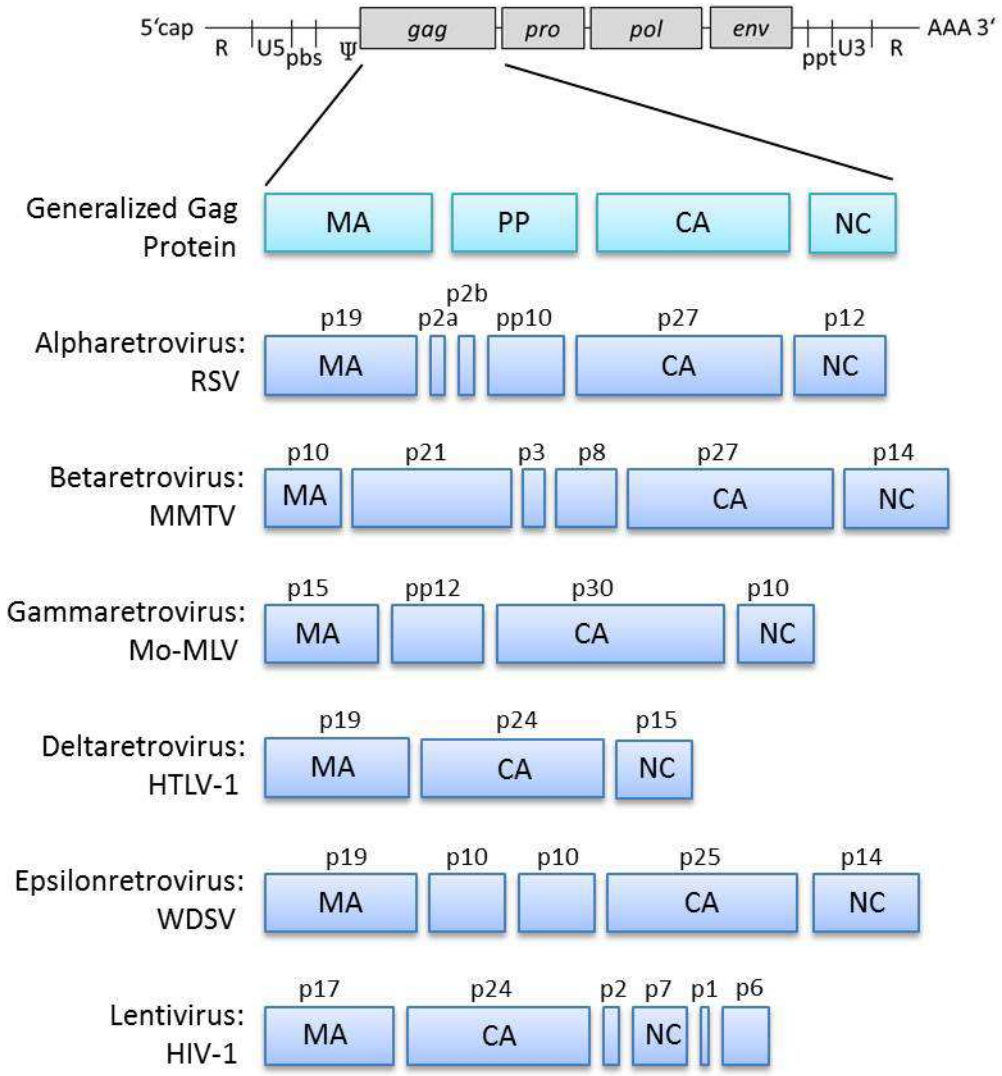


Fig. 2-4 Cleavage products of retroviral Gag proteins. In the viral RNA the gag gene is encoded at the 5´end. The generalized Gag scheme depicts the order the Gag domains of most retroviral Gag share: MA, PP, CA, NC. Gag Cleavage products of one member from each retroviral family are shown. Cleavage products have been assigned a name which reflects its apparent molecular weight, e.g. p2a of RSV, in some cases they are additionally used to the common subdomain name. (adapted from Wills and Craven 1991 and Kurth and Bannert 2010 (Kurth and Bannert, 2010; Wills and Craven, 1991)).

2.5 Cleavage site conformation

The nomenclature of proteolytic cleavage sites was introduced by Schechter and Berger in 1967. The substrate amino acid residues at the cleavage site are labelled as Pn, …, P2, P1, P1´, P2´, …, Pn`, where P1-P1´denotes the hydrolyzed bond (Fig. 2-5).

Fig. 2-5 Nomenclature of a cleavage site in a protease substrate. Amino acids N-terminal of the cleavage site are named P1 to Pn with P1 next to the cleavage site and amino acids C-terminal from the cleavage site are called P1´ to Pn` with P1` next to the cleavage site. Cleavage is indicated by the red arrow.

Till today the exact mechanism of processing site recognition by retroviral proteases has not been elucidated. Although the retroviral PRs are highly specific in site selection, the amino acid sequence recognized for cleavage is variable for each retroviral protein. Studies investigating the minimal number amino acids necessary for specific and efficient processing have shown that the enzyme requires at least four amino acids on the N-terminal site and 3 amino acids on the C-terminal site (Bagossi et al., 2005; Eizert et al., 2008). According to the nomenclature introduced by Schechter and Berger these are referred to as P4-P1 and P1´- P3´ (Cameron et al., 1993; Schechter and Berger, 2012). Pettite et al. investigated 46 known processing sites and discovered that about 80% of all known PR cleavage sites belong to one of two types. Type 1 cleavage sites, as often present at the N-terminus of CA, encode an aromatic residue at P1 and Pro at P1`. Type 2 sites, usually present at the C-terminus of CA, have a hydrophobic residue (excluding Ile and Val) at P1 and prefer Val, Leu or Ala at P1' (Pettit et al., 1991).

2.6 Lifecycle

The retroviral infection begins with the receptor mediated attachment by Env to the cell surface (Goff, 2007; Hernandez et al., 1996). This triggers the fusion of the viral and cellular membranes (Skehel and Wiley, 1998; Sommerfelt, 1999) and leads either directly, or indirectly after internalization, to the release of the viral core into the cytoplasm (Marsh and Helenius, 2006). During or after uncoating of the viral capsid the reverse transcriptase is activated (Goff, 2007; Narayan and Young, 2004) and copies the single stranded positive sense RNA genome into double stranded DNA. The full length DNA forms with other enzymes such as the integrase the preintegration complex (PIC). Depending on the retrovirus, PIC is actively transported into the nucleus (Bukrinsky et al., 1992) or depends on the nuclear membrane dissolution to access the chromosomes during the M-phase of the cell growth cycle (Roe et al., 1993). The integrase then facilitates the integration of the viral DNA into the host

cell chromosomes (Brown et al., 1987; Bushman et al., 2005). The integrated viral genome, called provirus, can be transmitted vertically during cell division from one cell to the progeny cells. The viral long terminal repeats (LTRs) flanking the provirus enable cellular RNA polymerase II to synthesise viral genomic RNA. The RNA is transported either differentially spliced or unspliced out of the nucleus into the cytoplasm where translation takes place by the host-cell translation machinery. Full length RNA is mainly used for the translation of Gag, Gag-Pro or Gag-Pro-Pol precursor protein. A fraction of the unspliced genome-length RNA is bound by the NC subdomain of the unprocessed Gag polyprotein and shuttled to the plasma membrane by the MA subdomain (Finzi et al., 2007; Jouvenet et al., 2006). The plasma membrane serves as a scaffold to concentrate Gag, Gag-Pro and Gag-Pro-Pol precursor molecules, which is visualized by electron microscopy as an electron dense patch. The Env precursor is translated at the ER from spliced RNA and cleaved by a cellular protease within the secretory pathway into the components SU (surface unit) and TM (transmembrane unit). These two proteins stay together as one complex and are anchored in the plasma membrane. Here TM interacts with the MA during the formation of the viral particle.

The immature particle assembles through lateral addition of Gag precursor proteins by binding to the viral genomic RNA underneath the plasma membrane. As the patch grows the plasma membrane protrudes while the virus buds off to be released from the cell (Meng and Lever, 2013). Some betaretroviruses, such as MMTV, assemble in the cytosol and are transported to the cell surface to bud through the plasma membrane. During or after release the immature particles mature by the activation of the retroviral protease and the subsequent cleavage of the precursor Gag proteins. The mature viral protein MA remains associated with the viral lipid envelope, CA assembles to the protein shell called capsid and NC assemples to a complex with the viral RNA genome inside the capsid. The viral maturation renders the retrovirus infectious (Ganser-Pornillos et al., 2008).

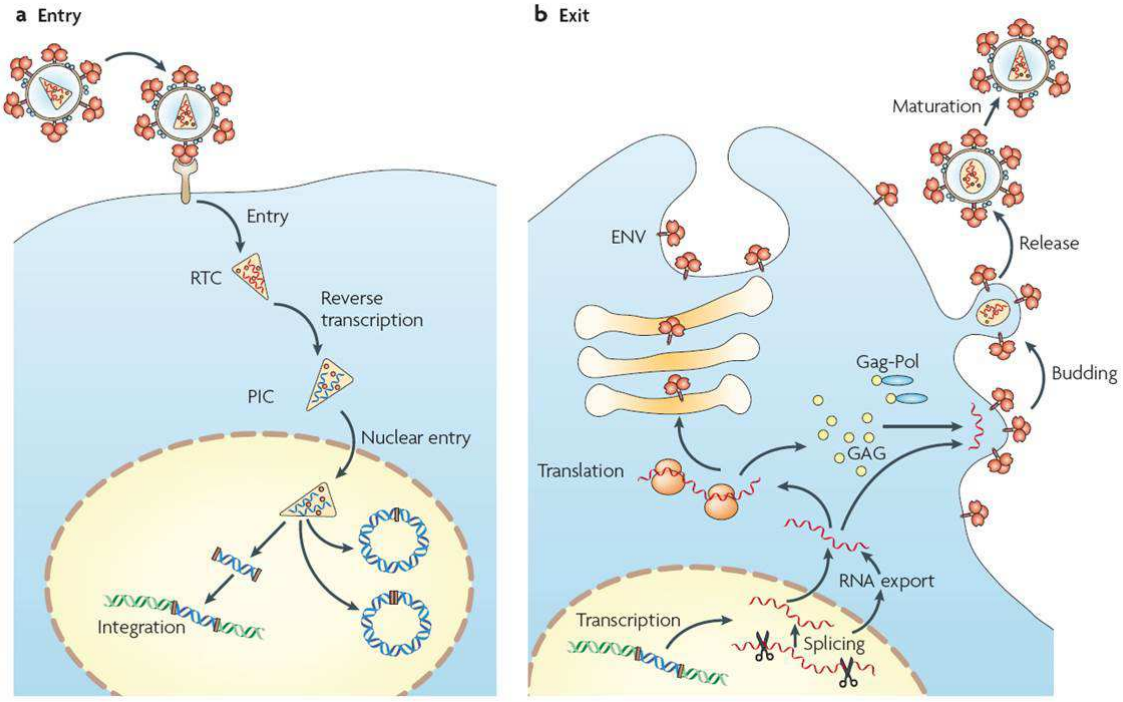


Fig. 2-5 Lifecycle of a orthoretrovirus. (A) Receptor mediated fusion of the viral and cell membrane leads to the release of the virus capsid into the cytosol. The reverse transcription complex evolves after reverse transcription of the viral RNA genome into the preinitiation complex which is shuffled into the nucleus. Inside the nucleus integration in the cellular genome takes place. (B) The viral provirus is transcribed by the RNA polymerase II. After transcription, spliced as well as unspliced RNA is transported into the cytosol where it is transcribed by the cellular transcription machinery. Gag precursor proteins assemble at the plasma membrane and encapsidate thereby an unspliced RNA dimer as the retroviral genome. Env interacts with the MA subdomain of Gag polyproteins and is thereby integrated in the immature particle which buds from the cell membrane. After release from the cell the virus matures through processing of Gag precursor as well as related Gag polyproteins. (Figure taken from Goff 2007(Goff, 2007))

2.7 Endogenous retroviruses

Exogenous retroviruses infect somatic cells leading to the transmission of the inserted provirus in progeny cells. When the cell or the host dies the retrovirus disappears as well (Bannert and Kurth, 2006). Occasionally retroviruses integrate into the genome of a mammalian germline cell and are passed vertically from the individual to its offspring following a mendelian pattern (Huebner and Todaro, 1969; Katzourakis and Gifford, 2010; Weiss, 2006). Sometimes the virus is then also able to re-infect other cells in the germline, leading to an increase in copy number in the offspring. The most recent model supporting this hypothesis has been the observation of a new type of endogenous retroviruses in Koalas. The

Koala Retrovirus (KoRV) causes leukemia and lymphomas in some freshly infected Koalas but it has also been shown to exist as an endogenous retrovirus in most natural populations (Tarlinton et al., 2006). The same, the parallel existence of endogenous and exogenous forms of one retrovirus, have been shown in mice with the mouse mammary tumour virus (MMTV), which is transmitted vertically to offspring as Mendelian traits but also by infection through breast milk (Robbins et al., 1986).

The endogenisation of retroviruses can have beneficial as well as detrimental effects on the host. Since the LTR of newly integrated retrovirus possess promoter and enhancer domains, it can up or down regulate the expression of downstream genes (Beyer et al., 2011; Jern and Coffin, 2008; Kettmann et al., 1982; Landry et al., 2002; Medstrand et al., 2001). Detrimental effects have been shown in oncogenic animal ERVs, such as MLV and MMTV, feline leukemia virus (FeLV), porcine endogenous virus (PERV) and the aforementioned KoERV (Kurth and Bannert, 2010). Beneficial effects can arise when retrotransposon-mediated sequence transduction and gene duplication lead to the creation of novel genes (Brandt et al., 2005). Endogenous retroviruses also can foster the diversity and stability of multigene families such as the major histocompability complex class II (Doxiadis et al., 2008). Another beneficial effect is the interference of proteins of endogenous retrovirus during the infection cycle of exogenous retroviruses. An example is the receptor blocking by Env protein expressed by endogenous proviruses, which can prevent the entry of an exogenous retrovirus, as it was shown for MLV (Jern and Coffin, 2008; Kozak et al., 1984). This was also shown for endogenous JSRV which inhibits the entry of exogenous JSRV. Another example for the interference of ERV proteins is the prevention of replication of exogenous JSRV by the expression of mutant enJSRV gag. If both gag are expressed in one cell this leads to hybrid capsids, which cannot exit the cell (Palmarini et al., 2004).

The taxonomy of endogenous retroviruses has been developed apart from the one of exogenous retroviruses. Currently ERVs are classified in three groups with class I retroviruses clustering with gamma- and epsilonretroviruses, class II ERVs with mostly beta– and very distantly with delta- and lenti retroviruses and class III ERVs with spumaretroviruses.

2.8 Human endogenous retroviruses (HERV)

Over the course of human evolution cells have been continually infected with retroviruses and some proviruses have become fixed in the germline as human endogenous retroviruses (HERVs). It is estimated that about 8 % of the human genome is of retroviral and retrotransposon origin (Lander et al., 2001) and approximately 3900 full length HERVs exist with both LTRs (Bannert and Kurth, 2006). Since there are several HERV present in Old and New World monkeys it is estimated that colonization has occurred more than 35 million years (Mya) ago. One of the oldest HERVs known probably integrated 60-70 Mya ago when the first primates begun to appear (Bannert and Kurth, 2006; Tristem, 2000). Most proviruses in humans exhibit very little polymorphisms if compared from one individual to the next and are even found at the same location in chimpanzees as well (Jern and Coffin, 2008).

The taxonomy of human endogenous retroviruses differs greatly from the one of exogenous retroviruses. The dominant classification scheme is to group HERVs into families according to the specific cellular tRNA that initiates the revers transcription process. This is indicated by the addition of the one-letter code of the amino acid specifity of the tRNA as a suffix following HERV. Examples therefor are the family HERV-W, with an $tRNA^{Trp}$, or HERV-K with an $tRNA^{Lys}$ most probable as a starter tRNA. Due to this historic classification sometimes unrelated elements were grouped together and had later assigned to new families or classes (Lavie et al., 2004).

Most HERVs have accumulated various mutations over time, which inactivated them, except when there was a beneficial effect for the host. The best investigated effect is the product of the env gene of the HERV-W provirus, syncetin-1. During embryogenesis the expression of syncetin-1 mediates cell fusion and syncytium formation in cells, called trophoblasts, which form the outer layer of the placenta (Mallet et al., 2004).

2.9 HERV-K 113

HERV-K elements have been first described in the early 1980s as human sequences similar to MMTV (Callahan et al., 1982) and were first called human MMTV like (HML) viruses. Until today there have been at least 550 HERV-K elements identified dispersed throughout the human genome, which marks HERV more of a class II superfamily (Bannert and Kurth,

2006). They have the organization of a reading frame for Gag, Pro and Pol in common and most of them encode a dUTPase domain N-terminal to Pro. The first full-length well conserved provirus described by Ono et al. belongs to the HERV-K family. There have been at least 73 human specific retroviral elements described and 15-20 are well preserved as well (Macfarlane and Simmonds, 2004).

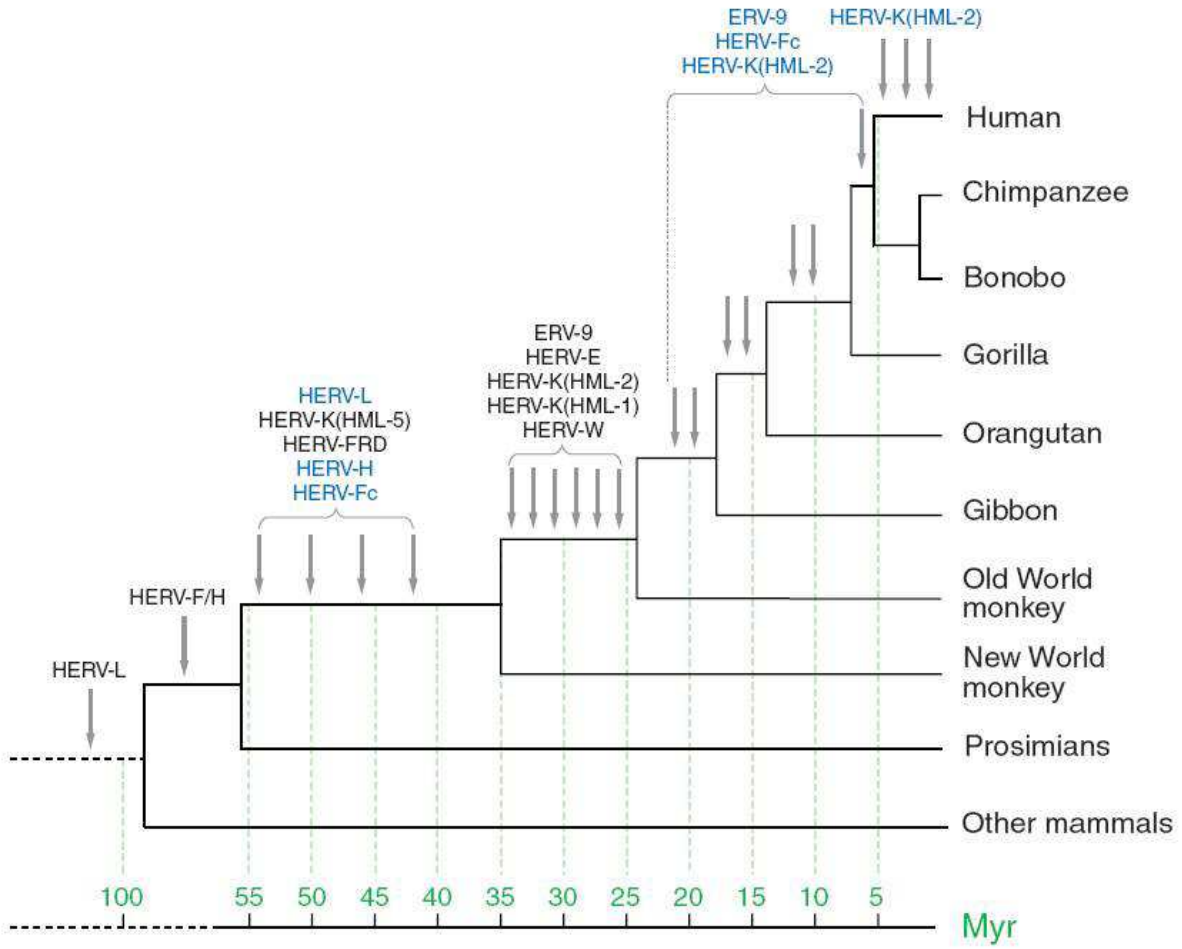


Fig. 2-6 Primary colonization times of selected HERV families. Over the last 100 Myr retroviruses continuously integrated into the genome of human ancestors. Some members of the HERV-K family integrated after the split of humans from the chimpanzee and bonobo predecessor. Depicted in blue are important amplification periods of some HERV families (Bannert and Kurth, 2006).

Turner et al. published in 2001 the first HERV element with almost identical LTR sequences and open reading frames for all viral proteins, called HERV-K113 (Turner et al., 2001). It lacks any substitutions in known functionally critical amino acids and has shown to be only present in a certain percentage in the human population, depending on ethnicity. In a recent study the geographic distribution of HERV-K113 among 156 HIV-1 positive subjects from

the United States, including African Americans and Caucasians was investigated. It was discovered that 21% of African Americans compared to 9% Caucasians carried HERV-K113 (Jha et al., 2009). This indicated a recent integration of the provirus and was in accordance with a previous publication which showed that the provirus integration was higher in individuals from Africa (21.8%) than in the United Kingdom (4.16%) (Moyes et al., 2005).

Because of their similarity to MMTV HERV-K elements have been implicated with tumours. First it was shown that teratocarcinoma cell lines express HERV-K particles (Boller et al., 2008; Büscher et al., 2005). Then it was revealed that breast cancer tissue exhibited elevated env RNA level (Ejthadi et al., 2005). Interestingly gag RNA was also detected in the peripheral blood cells of leukemia patients, as well as in breast, prostate and ovarian cancers (Depil et al., 2002; Ishida et al., 2008; Wang-Johanning et al., 2007). But it is still unknown if the activation of HERV-K elements is a side effect of the altered gene expression in tumour cells or if the virus actively induces tumour growth. Recent investigations have implicated that the HERV-K encoded Rec protein plays a role in tumour genesis. It was shown to interact with promyelocytic leukemia zinc-finger protein (PLZF), the related testicular zinc –finger protein (TZFP), the human small glutamin rich tetratricopeptide and with Staufen-1. All these proteins are proposed to be oncogenically active proteins (Hanke et al., 2013a, 2013b; Kaufmann et al., 2010).

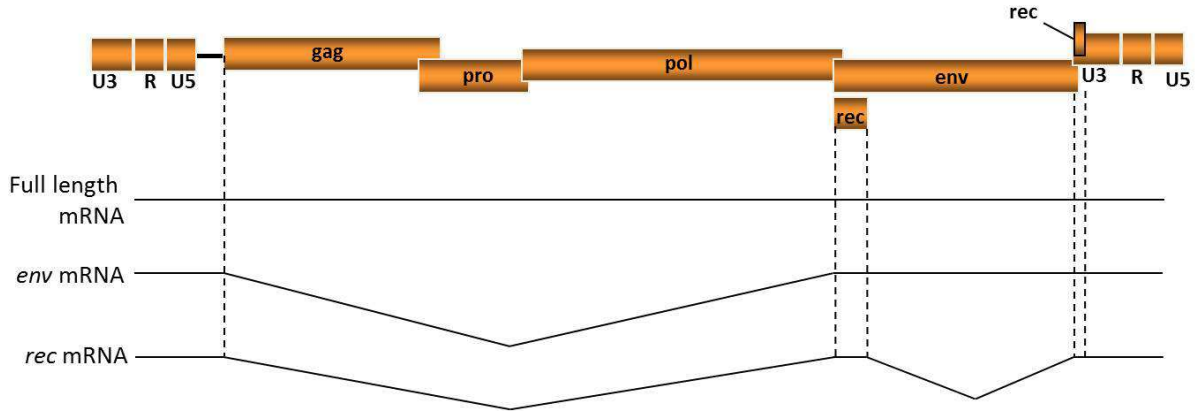


Fig. 2-7: Proviral sequence of HERV K 113. LTRs flank the proviral sequence of open reading frames for gag, pro, pol, env and rec. The rec mRNA is a product of a double splicing event.

The 9472bp HERV-K113 is localized on the human chromosome 19p13.11. The genome structure of HERV-K113 is similar to exogenous retroviruses. It is flanked by short duplicated host DNA and by the LTRs at the 5' and the 3' end. As visualized in Figure 2-7 it codes for the

proteins Gag, Pro, Pol, Env and Rec (Bannert and Kurth, 2004; Löwer et al., 1984). There are two ribosomal -1 frameshifts necessary for the expression of Pro and Pol, which are then expressed as the polyprotein Gag-Pro and Gag-Pro-Pol. The precursor proteins are cleaved by the virus encoded protease after viral release from the cell as indicated by the existence of mature proteins (Boller et al., 2008). Incomplete splicing leads to Env and a double splicing event leads to Rec mRNA. There is also a third 1.5 kb splice variant, called *hel*-transcript, which role has not yet been elucidated.

2.10 MALDI-TOF

The Matrix-assisted laser desorption/ionization technique is a two step process which allows the analysis of large organic molecules, such as proteins. The matrix, e.g. hydroxyl cinnamic acid (HCCA) and the analyte are cocristallized on a metal plate. Next the analyte is ionized by a nanosecond pulse, which leads to the ablation of the upper layer of the matrix material and a hot plume containing not only different species of matrix material but also ionized analyte molecules. The ionized analytes, e.g. biomolecules, are accelerated in an electric field and enter the flight tunnel, where they are separated according to their mass to charge ration and reach the detector at different times.

3. Materials and Methods

3.1 Materials

3.1.1 Laboratory devices

Device	**Company**
AlphaImager 2200 MultiImage Light Cabinet	Biozym Scientific GmbH, Oldendorf, Germany
1200 Series HPLC System	Agilent, Palo Alto, USA
384-spot polished steel tar- get plate	Bruker Daltonics, Bremen, Germany
4.6 × 150 mm 3.5 micron Zorbax 300SB-C8 reverse-phase column	Agilent, Palo Alto, USA
Autoflex I mass spectrometer	Bruker Daltonics, Bremen, Germany
Biological Safety Cabinet Thermo HeraSafe	Kendro, Langenselbold, Deutschland
CO_2-Incubator C 200 (eukaryotic cells)	Labotect, Göttingen, Germany
Cooling-heating thermomixer KTMR-133	HLC, Bovenden, Germany
Coulter-Counter Z2	Coulter Electronics Inc., Miami, USA
Counting chamber, Neubauer	Paul Marienfeld, Lauda-Königshofen, Germany

Cryo-Freezer	Carl Roth GmbH, Karlsruhe, Germany
Digital Grafic Printer UP-D895	Sony Corporation, USA
DNA Engine Thermocycler	BioRad, Hercules, USA
EM 902	(Zeiss)
Freezing container, Nalgene® Mr. Frosty	Sigma-Aldrich GmbH, Munich, Deutschland
Gene Pulser XCell + Shock Pod	BioRad, Hercules, USA
Gene Pulser XCell + Shock Pod	BioRad, Hercules, USA
HB-500 Minidizer Hybridization Oven	UVP, Upland, USA
Heating bath HB4 Basic	IKA Labortechnik GmbH, Staufen im Breisgau, Germany
Heraeus Sepatech Varifuge 3.0R	Heraeus Sepatech GmbH, Osterode, Germany
Horizontal Shaker GFL 3016	GFL Gesellschaft für Labortechnik GmbH, Burgwedel, Deutschland
Innova 4200 Incubator Shaker	New Brunswick Scientific, Edison, USA
Magnet stirrer IKA RTC	IKA Labortechnik GmbH, Staufen im Breisgau, Germany
Microcentrifuge 5415 D	Eppendorf, Wesseling-Berzdorf, Germany
Microplate Reader Sunrise	Tecan, Männedorf, Switzerland
Mini-Protean 3 Electrophoresis	BioRad, Hercules, USA
Multifuge 1S-R (Centrifuge)	Kendro, Langenselbold, Germany
NanoDrop Spectrophotometer ND-1000	Nanodrop, Wilmingon, USA
Optima L100K (Ultracentrifuge)	Beckman Coulter, Fullerton, USA
Overhead Lab Shaker Reax 2	Heidolph Instruments GmbH & Co. KG, Schwabach, Germany
pH Meter	Metrohm, Filderstadt, Germany
Photometer UV-1202	Shimadzu, Duisburg, Germany
Pipetus®	Hirschmann Laborgeräte, Herrenberg, Deutschland
Power Pac HC	BioRad, Hercules, USA
Precision Balance TE	Satorius Stedim Biotech GmbH, Göttingen, Germany
Rocking Platform Shakers	VWR, Darmstadt, Germany
Sub-Cell GT Agarose Gel Electrophoresis System	BioRad, Hercules, USA
SW 32Ti Rotor	Beckman Coulter, Fullerton, USA
Tabletop processor Agfa Curix 60	Agfa-Gevaert, Mortsel, Belgium
Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell	BioRad, Hercules, USA
Transelektronenmikroskop Tecnai 12; CCD-Kamera MegaView II	FEI Company, Hillsboro, USA; Olympus Soft Imaging Solutions GmbH, Hamburg, Germany
Trimmer EM Rapid	Leica Microsystems, Wetzlar, Germany
Ultracentrifuge-Tubes: 1. Thinwall Ultra-Clear™ Tubes 2. Thinwall Polyallomer konical Tubes	Beckman Coulter, Fullerton, USA
Ultraflex II MALDI-TOF/TOF	Bruker Daltonics, Bremen, Germany
Ultramikrotom Ultracut S	Leica Microsystems, Wetzlar, Germany
Ultramikrotom Ultracut UCT	Leica Microsystems, Wetzlar, Germany
UV-Transilluminator	BioView, Billerica, USA
Vortex Mixer	neoLab GmbH, Heidelberg, Germany

Table 1 The devices used are listed in this table

3.1.2 Laboratory equipment

Equipment	Company
100mm plates	TPP/Biochrom, Berlin, Germany
24 well plates	TPP/Biochrom, Berlin, Germany
6 well plates	TPP/Biochrom, Berlin, Germany
Blot Paper, extra thick	Biorad, Hercules, USA
Blottmembran Roti® PVDF-Membran, 0,45µm	Carl Roth GmbH, Karlsruhe, Germany
Chilling jar	Carl Roth GmbH, Karlsruhe, Germany
Cell culture flask, 150 cm2	TPP/Biochrom, Berlin, Germany
Cell culture flask, 75 cm2	TPP/Biochrom, Berlin, Germany
Cell scraper	TPP/Biochrom, Berlin, Germany
Cryo tubes, 2ml	TPP AG, Trasadingen, Schweiz
Electroporation Cuvette, 2mm	Biorad, Hercules, USA
Eppendorf tube 1,5/2ml	Eppendorf AG, Hamburg, Germany
Falkon Tubes, steril, 15ml	TPP AG, Trasadingen, Switzerland
Falkon Tubes, steril, 50ml	TPP AG, Trasadingen, Switzerland
Film 18 x 24cm, Kodak Medical X-ray l	Carestream Health Deutschland GmbH, Stuttgart, Germany
Filter paper, ø 90cm	Macherey-Nagel GmbH & Co.KG, Düren, Germany
Glass beads	Carl Roth GmbH, Karlsruhe, Germany
Kimwipes, Precision Wipes Tissue Wipers	Kimberly-Clark Professional, Koblenz, Germany
Parafilm, size S and M	Pechiney Plastic Packing, Chicago, USA
Pasteur Pipettes, single use	Carl Roth GmbH, Karlsruhe, Germany
PCR-Tubes 200µl, Softstrips	Biozym Scientific GmbH, Oldendorf, Germany
Serologicals Pipettes, single-use	TPP AG, Trasadingen, Switzerland
Single-use syringe, sterile, 20ml	B.Braun Melsungen AG, Melsungen, Germany
Syringe Filter, pore size 0,45µm	Satorius Stedim Biotech GmbH, Göttingen, Germany
Tissue Culture dish, sterile, 10cm	TPP AG, Trasadingen, Switzerland
Tube, Thinwall, konical™	Beckman Coulter, Inc., Fullerton, USA
Tube, Thinwall, Ultra-Clear™	Beckman Coulter, Inc., Fullerton, USA
X-ray cassette, Kodak X-Ray	Carestream Health Deutschland GmbH, Stuttgart, Germany
ZipTip C18 tips	Millipore, Bedford, MA, USA

Table 2 List of Laboratory equipment

3.1.3 Laboratory chemicals and enzymes

Chemicals and Enzymes	Company
Acetonitrile	Sigma-Aldrich Chemie GmbH, Munich, Germany
Agarose	Peqlab Biotechnologie GmbH, Erlangen, Germany
Alpha-Cyano-4-hydroxy-cinnamic acid (HCCA)	Bruker Daltonics, Bremen, Germany
Ammoniumpersulfate (APS)	Carl Roth GmbH, Karlsruhe, Germany
Ampicillin	Carl Roth GmbH, Karlsruhe, Deutschland
Calciumchlorid	Carl Roth GmbH, Karlsruhe, Germany
Calciumchlorid	Carl Roth GmbH, Karlsruhe, Germany
Complete Protease Inhibitor Tablet	Roche Diagnostics GmbH, Mannheim, Germany
Dimethylsulfoxid (DMSO)	Carl Roth GmbH, Karlsruhe, Germany
dNTP's	Fermentas International Inc., Burlington, Canada
Ethanol	Carl Roth GmbH, Karlsruhe, Germany
Ethidiumbromid	Carl Roth GmbH, Karlsruhe, Germany
Ethylendiamintetraacetat (EDTA)	Serva Electrophoresis GmbH, Heidelberg, Germany
Fetal calf serum	Biochrom, Berlin, Germany
Glutaraldehyd (25%)	TAAB Laboratories, Berkshire, England
Hydroxyethyl-piperazineethane-sulfonic acid buffer (HEPES)	Carl Roth GmbH, Karlsruhe, Germany
HPLC grade water	Mallinckrodt Baker B.V., Deventer, The Netherlands).
Isopropanol	Carl Roth GmbH, Karlsruhe, Germany
Laufpuffer, 10x	BioRad, Hercules, USA
Methanol	Carl Roth GmbH, Karlsruhe, Germany
Paraformaldehyd (PFA)	Carl Roth GmbH, Karlsruhe, Germany
Penicillin/Streptomycin 100x	Biochrom, Berlin, Deutschland
Pfu-DNA-Polymerase	Stratagene Europe, Amsterdam, Netherlands
PNGaseF (N-Glycosidase F)	New England Biolabs Inc., Beverly, USA
Polybrene®	Sigma-Aldrich Chemie GmbH, Munich, Germany
Ponceau S	Sigma-Aldrich Chemie GmbH, Munich, Germany
Restriction enzyme	New England Biolabs Inc., Beverly, USA
Rinderserumalbumin (BSA), Fraktion V, 96%	Sigma-Aldrich Chemie GmbH, Munich, Germany
Rotiphorese® Gel 30 (37,5:1)	Carl Roth GmbH, Karlsruhe, Germany
SDS – Lösung 10%	Biorad, München, Germany
ß-Mercaptoethanol	Carl Roth GmbH, Karlsruhe, Germany
Sucrose	Sigma-Aldrich Chemie GmbH, Munich, Germany
T4-DNA-Ligase	New England Biolabs Inc., Beverly, USA
Taq-DNA-Polymerase	Roche Diagnostics GmbH, Mannheim, Germany
Tetramethylethylendiamin, 1,2-Bis(dimethylamino)-ethan (TEMED)	Carl Roth GmbH, Karlsruhe, Germany
Transfer Buffer, 10x	BioRad, Hercules, USA
Trifluoroacetic acid (TFA)	Uvasol, Merck, Darmstadt, Deutschland
Tris(hydroxymethyl)-aminomethan	Carl Roth GmbH, Karlsruhe, Germany
Triton X-100	Carl Roth GmbH, Karlsruhe, Germany

Trypsin (0,25%)/EDTA (0,2%)	Biochrom AG, Berlin, Germany
Tween 20	Carl Roth GmbH, Karlsruhe, Germany
Urea	Carl Roth GmbH, Karlsruhe, Germany

Table 3 List of Chemicals and Enzymes

3.1.4 Standards

DNA and Protein standards	**Company**
Generuler 100 bp Ladder	Fermentas International Inc., Burlington, Canada
Generuler 1kb DNA Ladder Plus	Fermentas International Inc., Burlington, Canada
PageRuler Prestained Protein Ladder	Fermentas International Inc., Burlington, Canada
PageRuler Prestained Protein Ladder Plus	Fermentas International Inc., Burlington, Canada
Peptide calibration standard II	Bruker Daltonics, Bremen, Germany
Protein calibration standard I	Bruker Daltonics, Bremen, Germany

3.1.5 Buffer and Media

Buffer and Media	**Composition**
2x YT-Medium	1 % Bacto-Trypton; 1 % Bacto yeast extract; 0,5 NaCl
DNA loading buffer (6x)	10 mM Tris-Acetat, 50 mM EDTA, 10 % Ficoll-400 (w/v)(Serva), 0,4 % Orange-G (w/v) (Sigma) in H2O
LB Agar	LB-Medium with 20 g/ l Agar
LB Medium	1 % Bacto yeast extract; 1 % NaCl, pH 7,0
Ligation buffer, 1x	50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP, 10 mM Dithiothreitol, 25 µg/ ml BSA
SOB Medium	20 g Trypton, 5 g Bacto yeast extract, 0,186 g KCl or 2,5 ml 1 M KCl, 0,584 g NaCl or 10 ml 1 M NaCl-Lösung ad. 970 ml H20, pH 7,0
SOC Medium	9,7 Teile SOB, 0,1 Teil Mg-Mix, 0,2 Teile Glukose
TAE, 1x	40 mM Tris-Acetat, 1 mM EDTA, pH 7,2
TE, 1x	10 mM Tris-HCl, 1 mM EDTA, pH 8,0
APS, 10%	10% Ammoniumpersulfat in A.bidest., filtered sterile
Blocking buffer	5% skimmed milk in PBS/Tween20 (1%)
Blot Processing-Puffer	10% 10xTris/Glycerin/SDS Puffer + 90% A.bidest.
Blot transfer buffer	10% 10xTris/Glcerin Puffer + 20% Methanol + 70% A.bidest.r
Cell lysis buffer	1% Triton X-100, 20mM Tris (pH 7,7), 150mM Natriumchloride
Freeszing media for eukaryotic cells	90% FKS, 10% DMSO
PBS/Tween20 (0,1%)	0,1% Tween20 in PBS
Phosphate Buffer Saline (PBS)	123mM NaCl, 2, 7mM Kaliumchlorid (KCL), 10mM, Di-Natriumhydrogenphosphat, 2mM Kalium-Di-

	Hydrogenposphat; pH 7,0

Table 4 List of Buffers and Media

3.1.6 Kits

Kits	Company
ECL Plus Western Blot Detection System	GE Healthcare, Fairfield, USA
Endo-free Plasmid Maxi Kit	Qiagen GmbH, Hilden, Germany
Expand High Fidelity PCR System	Roche Diagnostics GmbH, Mannheim, Germany
HS-Mg RT Activity Kit	Cavidi, Uppsala, Schweden
PolyFect Transfection Reagent	Qiagen GmbH, Hilden, Germany
QIAprep Spin Miniprep Kit	Qiagen GmbH, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen GmbH, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden, Germany
QuikChange® Multi Site-Directed Mutagenesis Kit	Stratagene Europe, Amsterdam, Netherlands
Rodeo Sensitive Western Blot Detection	USB, Cleveland, USA
Silver Stain Kit	BioRad, Hercules, USA
Super Signal West Dura Extended Duration Substrate	Thermo Scientific, Logan, USA
Super Signal West Femto Max. Sensitivity Substrate	Thermo Scientific, Logan, USA

Table 5 List of Kits

3.1.7 Bacteria and cell lines

Equipment	Company/Origin
Escherischia coli (E.coli) One Shot Top 10	Invitrogen Corporation, Carlsbad, USA
HEK 293T	Human, kidney cells

Table 6 Bacteria and Cell lines used for the experiments

3.1.8 Antibodies

Equipment	Species	Origin
α-Rat IgG-HRP	Goat, monoclonal	Sigma-Aldrich Chemie GmbH, Munich, Germany
α-CA	rat, polyclonal	George M, 2011
α-MA	rat, polyclonal	George M, 2011
α-mCherry	rabbit, polyclonal	
α-p15	rat, polyclonal	George M, 2011

Table 7 List of antibodies used for detection

All antibodies were mixed with blocking buffer and incubated for at least 30 min with western blots.

3.1.9 Constructs

Equipment	Company/Origin
oricoHERV-K113	N. Beimforde (George et al., 2011)
oricoHERV-K113_GagProPol	(George et al., 2011)
oricoHERV-K113_GagPro¯Pol	(George et al., 2011)
oricoHERV-K113_GagProPol MA-SP1	(George et al., 2011)
oricoHERV-K113_GagPro-Pol p15-CA	(George et al., 2011)
oricoHERV-K113_GagPro-Pol CA-NC	(George et al., 2011)
oricoHERV-K113_NC_F624D	(George et al., 2011)
oricoGagCherry	Anja Zimmermann, n.p.

Table 8 List of constructs and their origin, if published

3.1.10 Software

Software	Company
Adobe® Illustrator CS5.1	Adobe Systems GmbH, München, Deutschland
BioEdit Sequence Alignment Editor v7.0.9.0	Tom Hall (1999)
Clone Manager 9 Professional Edition	Sci-Ed Software, Cary, USA
EndNote	Thomson Reuters, Philadelphia, USA
FlexControl 3.0	Bruker Daltonics, Bremen, Germany
iTEM	Olympus Soft Imaging Solutions GmbH, Hamburg, Deutschland
Microsoft Excel 2010	Microsoft Corporation, Redmond, USA
Microsoft Power Point 2010	Microsoft Corporation, Redmond, USA
Microsoft Word 2010	Microsoft Corporation, Redmond, USA
NanoDrop ND-1000 v.3.3	Nanodrop, Wilmingon, USA
Biotools 3.0	Bruker Daltonics, Bremen, Germany
FlexAnalysis	Bruker Daltonics, Bremen, Germany

Table 9 List of software used

3.2 Methods

3.2.1 Preparation of electrocompetent E. coli One shot Top 10

Electrocompetent E.coli Top 10 were plated on 2xTY Agar and incubated over night at 37°C. The next day single colonies were picked and used to grow a starter 10ml culture by incubation over night at 37°C and 220rpm. Of this culture 2.5ml were used to inoculate 250 ml 2xTY-Medium and grown at 37°C and while shaken again at 220rpm, for approximately 4 hours until an OD of approximately 0.6-0.8 measured by a Photometer was reached. 2x 250ml culture were united in a Glas flask and cooled for 15min on ice. Centrifuge bottles were disinfected with 0.1N HCl and washed with A. bidest. Then the culture was moved into the bottles and the cells spinned for 20 min at 256g and 4°C. The supernatant was discarded, the Pellet resuspended in ice cold 1mM HEPES and again centrifugated as before. Next the Pellet was was resuspended in 125ml and after another centrifugation step the pellet was resuspended in 25ml 1mM HEPES/10% Glycerin and pipetted into a 50ml Falcon tube. The suspension was centrifugated at 4000g, 4°C for 20 min. The supernatant was discarded, the pellet resuspended in 3ml ice cold 10%Glycerin and aliquoted á 55µl in 1.5ml Eppedorf tubes. These tubes were then directly thown into a liquid nitrogen and later kept at -80°C. Efficiency and contamination tests were done with the pUC19 vector as a control on the next day.

3.2.2 Electroporation

Competent bacteria are able to take up plasmid DNA and are used to multiply DNA. The take up can be initiated by e.g. electroporation. Therefore an aliquot of electrocompetent cells (see 3.1.1) was mixed with 1-50ng DNA in a cold electroporation cuvette and immediately after electroporation (2.5kV, 25µF, 200Ohm) moved to ice. 250-1000µl SOC-Medium were added, the cell mix transferred into 1.5 Eppendorf tubes and incubated at 37°C, 200rpm for 1 hour.

3.2.3 Mini and Maxi DNA preparation

Small and large quantities of plasmid DNA were prepared by transforming E. coli Top Tenn cells with the preferred plasmids and extracting plasmid DNA through Mini and Endo-free Plasmid Maxi Kit.

For the Mini preparation a single colony of transformed E.coli were used to inoculate 5ml LB-Amp-Medium and grown over night at 37 and 200rpm. Plasmid DNA was isolated then according to the instruction with the QIAprep Spin Miniprep Kit. For the Maxi preparation first a single colony was incubated for 8h at 200rpm and then this starter culture was used to inoculate 250ml in a baffled Erlenmeyer flask. This culture was grown over night at 37°C and 220rpm. The next day the plasmid DNA was isolated with the Endo-free Plasmid Maxi Kit according to the instructions. DNA was eluted in the last step with 200µl A.bidest.

3.2.4 Measuring DNA concentration

The concentration of isolated DNA was measured via the Spectralphtotmeters NanoDrop ND-1000. The concentration was measured by λ= 260nm or 280nm and the device was first calibrated with A. bidest.

3.2.5 Mutagenesis PCR

To introduce specific mutations in constructs the in vitro Multi Site-Directed Mutagenesis Kit was used. The primer for this reaction had to be specially designed. The mutation had to lie in the middle of a 31-46 bp long sequence. The melting temperature (T_m) had to be ≥75°C and was estimated according to Stratagene with this formula

$$Tm = 81.5 + 0.41(\%GC) - \frac{675}{N} - \%mismatch$$

The length of the primer was N and the GC content had to be calculated in % for this formula and should have been at about 40%. With this kit it was possible to use up to 5 potential DNA mutations at one. The PCR mix and the cycling procedure were conducted according to the kits manual.

3.2.6 Restriction digest

Bacterial Type-2-endonukleases are used to digest DNA at specific sites, most often recognized through a palindrom structure. The digest can be used to verify a mutation in sequence or identify a vector through the specific band pattern. It is also used to for molecular cloning. All restriction enzymes used and their buffers were from NEB. A digest was done by incubation over night at room temperature or 1-2h at 37 °C.

3.2.7 DNA-Agarosegelelectrophoresis

With this method mixes of different DNA fragments can be separated by size. The agarose forms a three dimensional structure through which the DNA molecules can pass through pores and channels depending on their size. Depending on the range of DNA sizes expected the Gel consisted of 0.75-2% Agarose and was generated by heating the appropriate amount of Agarose with 1x TAE buffer. When all agarose had melted with the buffer into a clear solution, 0.5µg/ml Ethidiumbromide was added after the solution had cooled off a little bit. The probes were mixed with 6x DNA loading buffer, leading to a single dilution and applied into the left out pockets in the gel. The DNA fragments were separated by 70-110 V, visualized under UV-Light and a picture was taken with the AlphaImager 2200.

3.2.8 Purification of DNA fragments

After a digest the different DNA fragments of the digest can be separated on an agarose gel. Depending on what fragment is of interest, e.g. Vector or insert, the specific band in the agarose gel is detected on the UV-Transilluminator and cut out with a scalpel. The agarose piece is transferred to a 2ml reaction tube and the DNA isolated according to the QIAquick Gel Extraction Kit manual.

3.2.9 Ligation

The T4-DNA-Ligase catalyzes the ligation of free 3'-hydroxy and the 5'-phosphate end of a Vector and Insert, which were cut with the same restriction enzyme before. In order to achieve a favorable balance of vector molecules to the number of insert molecules the amount of vector and insert was determined with this equation $Insert_{DNA}$

$$\text{Insert} = 3x20ng\ Vector\ x\ bpVector/bpInsert$$

bpVector- number of basepairs of the vector; bpInsert- number of basepairs of the Insert

The ligation mix was set up according to the New England Biolabs T4-DNA Ligase manual and kept for 1h at room temperature or over night at 16°C. After the incubation step the mix was used to transform electrompetent *E.coli*.

3.2.10 Sequencing

The exact nucleotide sequence of a DNA fragment can be determined via the chain termination method developed by F. Sanger. Modified nucleotide substrates lead to a termination of the DNA synthesis reaction and the analysis of the fragments reveals the DNA sequence. Sequencing was applied to check for mutations in the vector or after the introduction of mutations in the sequence. Sequencing primers are 18-25 nucleotides in length and should have a GC content of at least 40%. Sequencing was conducted at the institutes own sequencing laboratory and the results analyzed with the program Clone Manager. The reaction mixture was prepared according to the sequencing laboratory guidelines:

Components	**Volume/concentration**
DNA Template	1-4µl (150-300ng)
Primer (10µM)	0,5µl (10µM)
Abi BigDye 3.1® Mix	1µl
Abi BigDye 3.1 Puffer (5x)	1µl
A.bidest.	ad 10µl

Table 10 Components for sequencing mix

The DNA sequence of interest was then amplified in a PCR reaction also according to the recommended conditions by the sequencing laboratory:

Step	**Temperature [°C]**	**Time**	**Cycle**
Denature	96	2 min	1
Denature	96	10 sec	⎫
Annealing	53	10 sec	⎬ 25
Elongation	60	4 min	⎭
Store	4	∞	

Table 11 Steps of sequencing program

3.2.11 Cell culture

Adherent HEK 293T cells were cultured at 37°C, 5%CO2 and 98% humidity in Dulbecco's modified Eagle medium containing 10% fetal bovine serum, penicillin (50U/ml), streptomycin (50Uµg/ml) and L-Glutamine (2mM). Every 48-72 hours at a confluency of approximately 80% media was taken off, cells washed with PBS and separated from the bottom of the flask by incubation with 1ml trypsin/ECTA. The cell number was determined

via the Coulter cell counter Z2 and, depending on the size of the new container, a fraction of cells was seeded into new flasks or plates. All procedures were done inside the bio safety cabinet wearing gloves and a long sleeved lab coat.

3.2.12 Freezing and thawing cell aliquots

Cells were stored in a freeze medium, consisting of 90% FKS and 10% DMSO. Therefore at least $1x10^7$ cells were resuspended in 1 ml of freeze medium, transferred into a cryo-tubes and slowly frozen in a chilling-jar filled with Isopropanol overnight -80°C. The next day cells were transferred into liquid nitrogen. Frozen cells were thawed in 20ml Medium and cultivated in CO_2-Incubator.

3.2.13 Counting cells

The exact number of cells is necessary to seed the right amount of cells which can then be transfected and successfully grow virus until the harvest. Cells were therefore washed with PBS and separated from the bottom of the flask or plate with trypsin. After that they were pelleted at 1200 rpm for 4 min at room temperature and resuspended in 6ml of fresh medium. To count the cells in the Coulter Counter Z2 10µl of the cell suspension were transferred into a specific cell counter vial with 10ml Isoton buffer and measured according to the manual. For transfection on the next day the following number of cells were seeded.

Plate/Flask	Number of HEK293T cells	Medium added
6 Well- plate	$6x10^5$	3ml
10mm plate	$4x10^6$	8ml
Cell culture flask, $75cm^2$	$5.5x10^6$	15ml
Cell culture flask, $100cm^2$	$7.5x10^5$	25ml

Table 12

3.2.14 Transfection of cells

DNA can be introduced into eukaryotic cells by transfection. This is achieved by opening up transient pores in the cell membrane to allow the uptake of genetic material. Cells seeded in 6-well-plates were transfected with the PolyFect Transfection Kit according to the Kits manual. Cells grown in 10mm plates and cell culture flasks were transfected via the Calciumphosphate method. Plasmids were therefore diluted in a specific volume, Calciumphosphate added drop by drop and this mixed solution then again dropwise added to 2xHBS. This was then

incubated for 30 min and then dropwise added to cells with fresh medium. The cells were incubated for 8-12h in the CO2-Incubator and then the medium was exchanged against fresh media. Cells were then cultured until they reached the desired confluency.

Plate/Flask	**DNA [µg]**	**ad. H_2O [µl]**	**2.5M $CaCl_2$ [µl]**	**2xHBS [µl]**	**Medium [ml]**
6 Well- plate	3-4	67	8	1.2	3ml
10mm plate	25-30	450	50	8	8ml
Cell culture flask, 150cm^2	60-63	1125	125	20	25ml

3.2.15 Concentration of Virus like particles

HEK 293T cells were transfected with the plasmids expressing a variant of oricoHERV-K113. 48 hours post-transfection the supernatants were collected, centrifuged at 3345 × g for 8 min and filtered through 0.45 µm-pore-size membranes to remove cell debris. The viral particles were then concentrated by ultracentrifugation through a 20% sucrose cushion at 175,000 × g for 3h at 4°C. Viral pellets were resuspended in either 100 µl 5 M urea containing 1% glacial acidic acid for high pressure liquid chromatography or in 0.05 M Hepes buffer, pH 7.2 for Western blot analysis.

3.2.16 Cell lysis

Cells were lysed to analyse the proteins expressed by them on SDS Page. Therefore the cells were washed, the buffer taken off and the just covered with incubated with cell lysis buffer, to which a protease Inhibitor had been added. The lysed cell mix was centrifugated at 4°C and 15700g for 5 min and the supernatant was transferred into an 1.5ml Eppendorf tube. The proteins were kept at -20°C until further analysis.

3.2.17 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Similar to Agarose gel electrophoresis also during SDS-PAGE proteins are separated by size. Before the sample is added to the gel, it is mixed with Laemmli loading buffer, which contains β-Mercaptoethanol. This chemical leads to the breakage of disulfide during the cooking process of 10min at 95°C. In addition the anionic detergent SDS is added to the polyacrylamide gel, which denatures secondary and tertiary protein structures. It also applies negative charge to each protein. The SDS gel is arranged in two phases, at the top the stacking gel and at the bottom the separation gel. Depending on the sample size the gel is 1,0 or 1,5mm thick. After application of the sample to the top of the gel the proteins migrate first slowly

through the stacking gel at 80-100V. When they reach the separation gel the Voltage is increased to 120-160V.

Components	Stacking Gel (2x) 5%	Separation Gel (1x)	
		10%	15%
A.bidest.	5,7ml		4,1ml
Acryamid 30%	1,7ml	3,3ml	4,0ml
1,5M Tris pH 8,8	-	2,5ml	2,5ml
0,5M Tris pH 6,8	2,5ml	-	-
SDS 10%	100µl	100µl	100µl
APS 10%	50µl	50µl	50µl
TEMED	10µl	10µl	10µl

3.2.18 Silver nitrate staining

To visualize the proteins separated through SDS-PAGE in the acrylamide gel they can be stained with silver nitrate. Therefore the proteins in the SDS gels are first fixated with 10% silver nitrate. Silver ions are able to adhere to negatively charged sidechains of aminoacids.

3.2.19 Western Blot analysis

Western blot analysis enables the specific detection of proteins through specific antibodies. Proteins are therefore transferred in a process called blotting after the separation through SDS-PAGE from the acrylamide gel onto a PVDF membrane. The first step of blotting is to incubate the gel in the transfer buffer for 10min and then to lay it on top of a PVDF membrane which was activated with Methanol before. These two layers are sandwiched between two blotting paper drenched in transfer buffer. The proteins are then transferred at a voltage of 20V for 40 min from the gel onto the membrane. This membrane is then incubated in a blocking solution for one hour, containing skimmed milk, to inhibit unspecific binding of antibodies later on. After that the proteins are specifically detected with antibody solubilized in blocking solution either at room temperature for one hour or over night at 4°C. If the primary antibody is not coupled to a horse radish peroxidase (HRP), it has to be detected by a secondary antibody which is coupled to HRP. Before the incubation with the secondary antibody the membrane is washed 3x á 10 min with PBS/Tween 0.1% at room temperature. After the incubation with the secondary antibody with the same conditions as the first antibody the proteins are detected through a chemiluminescence reaction. The light emitted is able to darken an X-ray film which allows to picture the transferred proteins.

3.2.20 Reversed phase high pressure liquid chromatography (RP-HPLC)

Protein separation was performed on an Agilent (Palo Alto, CA) 1200 series binary HPLC fitted with a 4.6 × 150 mm 3.5 micron Zorbax 300SB-C8 reverse-phase column (Agilent, Palo Alto, CA). The column was maintained at 40°C. Solvent A consisted of 0.1% trifluoroacetic acid in water and solvent B of 0.08% trifluoroacetic acid in acetonitrile. Proteins were eluted at a flow rate of 0.5 ml/min from 8-58 min and 1 ml/min from 0-6min and 60-62 min employing the following gradient: solvent B, 0-6 min, 0%; 8 min, 15%; 49 min, 40%; 58 min, 95%; 58-60 min, 95%; 62 min, 0%. The eluate was monitored at 280 nm and 0.5 ml fractions were collected.

3.2.21 MALDI-TOF mass spectrometry

Following evaporation to dryness proteins of each HPLC fraction were dissolved in 20 µl of TA2 (2:1 (v/v) mixture of 100% acetonitrile and 0.3% TFA). 1 µl of each fraction was spotted onto a 384-spot polished steel target plate (Bruker Daltonics, Bremen, Germany) and mixed with 1 µl alpha-Cyano-4-hydroxy-cinnamic acid (HCCA) solution (6 mg/ml in TA2) and air dried.

Mass spectra were collected by an Autoflex I mass spectrometer (Bruker Daltonics). The instrument was controlled by Bruker's FlexControl 3.0 data collection software and was equipped with a UV-nitrogen laser (l = 337 nm). MS measurements were carried out in linear mode using an acceleration voltage of 20.00, or 18.45 kV (ion source 1 and 2), respectively. Lens voltage was 6.70 kV. Spectra were stored in mass range between 0.7-10 kDa and 2-20 kDa, depending on the expected size of the peptides. External calibration was performed employing protein calibration standard I and peptide calibration standard II, respectively (Bruker Daltonics). To achieve a high signal to noise ratio each spectrum represents the integration of at least 600 individual laser shots. In order to determine the exact positions of the N- and C-terminal ends of the processed Gag fragments the respective proteins were digested with trypsin. Tryptic peptides were purified using ZipTip C18 tips (Millpore, Bedford, MA, USA) and measured in the reflectron mode using an acceleration voltage of 19.40, or 16.90 kV (ion source 1 and 2), respectively. Lens voltage was 8 kV. Spectra were stored in the mass range between 0.7-4 kDa. Mass peaks that did not correspond to tryptic peptides predicted by theoretical in silico tryptic digests were further analyzed by MS/MS to generate sequence information.

3.2.22 De novo protein sequencing

Sequencing by MALDI-TOF MS was performed under the control of FlexControl software (Bruker Daltonics) using an Ultraflex II MALDI-TOF/TOF mass spectrometer (Bruker Daltonics). It was equipped with a near infrared solid state smartbeam™ laser Nd:YAG laser (l = 1064 nm) which operated at 100 Hz. Fractions containing target peptides were identified by recording spectra in linear positive mode with external calibration using a standard mixture of peptides. In order to assign a mass window for fragmentation and peptide sequencing in the 'LIFT' MS/MS mode, an exploratory scan from 2000 to 5000 Da was performed in the reflectron mode. Spectra were obtained by averaging up to 3000 laser shots acquired at a fixed laser power, which had been set to the minimum laser power necessary for ionization of selected samples before starting the analyses. The mass spectra were visualized and processed using FlexAnalysis software and sequence tag hints were obtained by analyzing tandem MS spectra employing the Biotools 3.0 software (Bruker Daltonics). For N-terminal sequencing by Edman degradation, proteins were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), blotted to a PVDF membrane and stained with Ponceau S. Protein bands of the expected size were cut out and sent for sequencing (Proteome Factory, Berlin, Germany).

4. Results

4.1 Alignment of retroviral Gag polyprotein sequences

The protein sequences of several Gag polyproteins have been studied. In all retroviral gag proteins the three subdomains MA (Matrix), CA (Capsid) and NC (Nucleocapsid) are found in the same order (Leis et al., 1988). Interspersed between them are various subdomains with often unknown functions.

Wills et al. showed that HIV Gag is cleaved into MA, CA, NC and p6, as well as two spacer proteins called SP1 and SP2. The mature viral proteins of MMTV have also been published, which are p10, pp21, p3, p8, p27 and p14 (Hizi et al., 1989). For MPMV Henderson et al. concluded that Gag is cleaved into p10, pp24, p12, p27, p14 and p4(Henderson et al., 1984).

The exact protease cleavage sites of the HERV Gag polyprotein have not yet been discovered. To investigate the gag subunits, the protein sequence of HERV-K113 was aligned first with two other beta retroviruses MMTV and MPMV. These alignments allowed the preliminary interpretation that the subunit is organized in the following order. First, there is a MA region analogous to MMTV and MPMV. Encoded in this region are the 18 amino acids responsible for type D-specific morphogenesis (Choi et al., 1999). This region is followed by an amino acid sequence which encodes a late domain with the amino acid sequence “PTAP”, similar to MMTV. Since proteins, which contain the late domain, are likely to be phosphorylated this sequence was hence named phosphoprotein (p15). The Capsid domain can be identified by the major homology region, which encodes the three absolutely conserved polar residues Gln, Glu, and Arg, as well as the conserved hydrophobic residues (Purdy et al., 2008). The Alignment also revealed two conserved zinc finger domains. In many retroviral NC at least one of these zinc finger domains is encoded (Darlix et al., 1995).

```
HERV*Gag MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDW
MMTV gag MGVSGSKGQKLFVSVL---QRLLSERGLHVKESSAIEFYQFLIKVSPWFPEEGGLNLQDW
MPMV gag MGQE----LSQHERYVEQLKQALKTRGVKVKYADLLKFFDFVKDTCPWFPQEGTIDIKRW

HERV*Gag KRIGKELKQ-AGRK--G-NIIPLTVWNDWAIIKAALEPFQTEE-DSVSVSDAPGSCIIDCN
MMTV gag KRVGREMKRYAAEH--GTDSIPKQAYPIWLQLREIL----TEQSDLVLLSAEAKSVTEEE
MPMV gag RRVGDCFQD-YYNTFGP-EKVPVTAFSYWNLIKELI-------DKKEVNPQVMAAVAQTE

HERV*Gag EKTRKKSQKETESLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLK---LEGKGPEL-VG
MMTV gag LEEGLTGLLSTSSQEKTYGTRGT-AYAEIDTEVDKLSEHIYDEPYE---EKEKADK----
MPMV gag EILKSNSQTD------------LTKTSQNPDLDLISLDSDDEGAKSSSLQDKGLS----

HERV*Gag PSESKPRGTSPLPAGQV---PVTLQPQKQVKENKT-------QPPV-AYQY--WPPAELQ
MMTV gag -NEEKDHVRKIKKVVQR---KENSEGKRKEKDSKAFLATDWNDDDL-SPED--WDDLEEQ
MPMV gag -STKKPKRFPVLLTAQTSKDPEDPNPSEVDWDGLE-------DEAA-KYHNPDWPP--FL

HERV*Gag ---YRPPPESQYGYPGMPPAPQGRAPY-PQPPTRRLNPTAPPSRQGSELHE-IIDKSRK-
MMTV gag AAHYHDDDEL------ILPVKR-KVVK-KKPQALRRKPLPPVGFAGA------MAEARE-
MPMV gag ---TRPPPYNK----ATPSAPTVMAVVNPKEELKEKIAQLEEQIKLEELHQALISKLQKL

HERV*Gag EGDTEAWQFPVTL-----EPMPPGEGAQEGE-----------------PPTVEAR----
MMTV gag KGDL-TFTFPVVF---------MGESDEDDT-----------------PV---------
MPMV gag KTGNETVTHPDTAGGLSRTPHWPGQHIPKGKCCASREKEEQIPKDIFPVTETVDGQGQAW

HERV*Gag --YKSFSIKMLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAKSSLSPSQFL
MMTV gag --WEPLPLKTLKELQSAVRTMGPSAPYTLQVVDMVA-SQWLTPSDWHQTARATLSPGDYV
MPMV gag RHHNGFDFAVIKELKTAASQYGATAPYTLAIVESVA-DNWLTPTDWNTLVRAVLSGGDHL

HERV*Gag QFKTWWIDGVQEQVRRN--RAANPPVNIDADQLLGIGQNWSTISQQALMQNEAIEQVRAI
MMTV gag LWRTEYEEKSKEMVQK---AAGKRKGKVSLDMLLGTGQFLSP-SSQIKLSKDVLKDVTTN
MPMV gag LWKSEFFENCRDTAKRN--QQAGN--GWDFDMLTGSG-NYSSTDAQMQYDPGLFAQIQAA

HERV*Gag CLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIA---DEKARKVIVELM
MMTV gag AVLAWRAIPPPGVKKTVLAGLKQGNEESYETFISRLEEAVYRMMP---RGEGSDILIKQL
MPMV gag ATKAWRKLPVKGDPGASLTGVKQGPDEPFADFVHRLITTAGRIFG---SAEAGVDYVKQL

HERV*Gag AYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMHKAMLMAQAITGVV----L
MMTV gag AWENANSLCQDLIRPIR-----KTGTIQDYIRACLDASPAVVQGMAYAAAMRGQK----Y
MPMV gag AYENANPACQAAIRPYRKKTD-----LTGYIRLCSDIGPSYQQGLAMAAAFSGQTVKDFL

HERV*Gag GGQVR-TFGG---K----------CYNCGQIGHLKKN---CPVLNKQNITIQATTTGREP
MMTV gag STFVKQTYGG---GKGGQGAEGPVCFSCGKTGHIRKD---CK---------DEKGSKRAP
MPMV gag NNKNK-EKGG---C----------CFKCGKKGHFAKN---CH---------EHAHNNAEP

HERV*Gag --PDLCPRCKKGKHWASQCRSKFDKNGQPL---SGNEQRGQPQAPQQTGAFPIQPFVPQG
MMTV gag --PGLCPRCKKGYHWKSECKSKFDKDGNPL----------------------------
MPMV gag KVPGLCPRCKRGKHWANECKSKTDNQGNPIPPHQGNGWRGQPQAPKQ--AYGAVSFVPAN

HERV*Gag FQGQQPPLSQVFQGISQLPQ----YNNCPPPQAAVQQ
MMTV gag -----PPLETNAENSKNL------------------
MPMV gag -------KNNPFQSLPEPPQEVQDWTSVPPPTQY---
```

Fig. 4-1: Alignment of oricoHERV-K113-Gag (HERV*Gag), MMTV Gag and MPMV Gag. The well-established Gag subdomains, which both, MMTV and MPMV, have in common are color coded: Matrix (MA), Capsid (CA) and Nucleocapsid (NC); in light green is the MPMV Cytoplasmic Targeting/Retention Signal (Choi et al., 1999) highlighted; in purple is the Late domain (Yasuda and Hunter, 1998) highlighted; the Major homology region is highlighted in light brown with the conserved polar residues Gln(Q), Glu (E), Phe (F) and Arg(R) highlighted in yellow (Mammano et al., 1994); the characteristic zinc finger domain motiv (CCHC) amino acids are highlighted in light blue:…-Cys-X2-Cys-X4-His-X4-Cys-…; underscored are amino acids oricoHerv-K113 Gag has in common in the before specified regions with MMTV and MPMV. The alignment was generated using BLOSUM 62 (Clone Manager) and manually refined.

4.1.1 Cloning of oricoHERV-K113_GagProPol

The proviral sequence of HERV-K integrated into the human genome about 80.000 years ago and during this time acquired several incapacitating mutations. To overcome this obstacle a method had been in developed in our group to reverse mutations in the genome of HERV-K 113, which were most probably introduced after the integration into the human genome.

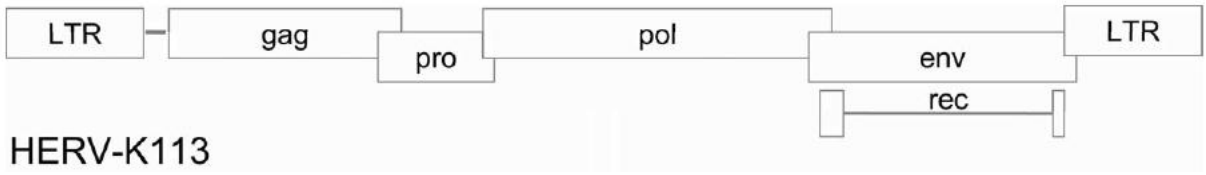


Fig. 4-2: Schematic picture of HERV-K113 genome (George et al., 2011)

Therefore the amino acid sequences of 10 well-preserved human specific HERV-K (HML-2) viruses were aligned with the sequence of HERV-K113. To identify amino acid changes, which had been accidentally introduced after the insertion of the viral sequence into the host's genome, each amino acid of HERV-K-113 differing from the other sequences was assessed under two aspects:

- If there was none or only one other sequence bearing the same differing amino acid as in the HERV-K113 sequence it was considered a mutation introduced after integration. Subsequently it was exchanged by side directed mutagenesis.
- If there were two or more sequences bearing the same nucleotide at this position it was left unchanged and named a variation.

According to this algorithm 16 amino acids in the protein sequence of HERV-K113 were identified as post-insertional mutations and changed, five of them encoded in Gag, three in Pro and eight in Pol by Nadine Beimforde. This sequence was then called oriHERV-K 113, since it resembles the original sequence at the time of integration. To further enhance the protein expression, the main sections of the nucleotide sequence of Gag, Pro and Pol were codon optimized for mammalian cells (GeneArt), except the overlapping gene regions between each of the genes (Fig.3-3). These regions were spared, to ensure that the ribosome was still able to frame shift during translation from the Gag sequence into the Pro sequence as well from Pro to Pol. The codon optimized construct was then termed oricoHERV-K 113. At last the nucleotide sequence of gag, pro and pol was cloned into the pcDNA3.1 plasmid by facilitating the restriction sites KpnI and XhoI. It was thereby set under the control of a CMV promoter to enhance protein expression. Improved virus expression was seen in electron microscopy, Western Blot and Cavidi Assays.

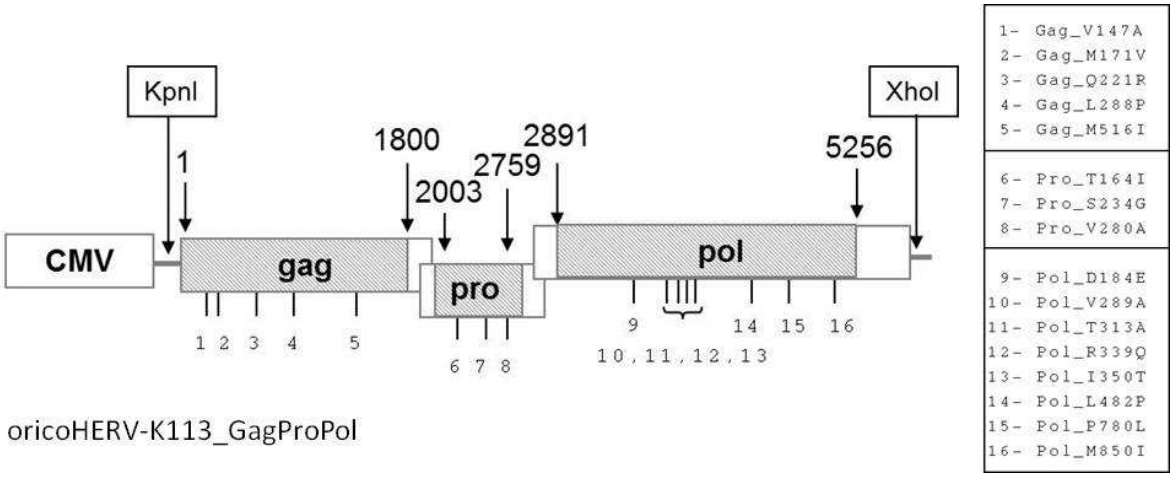


Fig. 4-3 Schematic presentation of oricoHERV-K113_GagProPol (oricoHERV-K113 GPP) in the vector pcDNA3. The sequence gag, pro and pol were partially codon optimized (gray areas). Numbers underneath the construct indicate the amino acid substitutions. Numbers above the construct indicate the amino acids, beginning with 1 at the start of Gag. Grey shaded boxes represent the codon optimized sequences. Overlapping sequences between gag-pro and pro-pol were not codon optimized; hence the boxes are left white. (adapted from (George et al., 2011))

4.1.2 Generating the protease mutant oricoHERV-K113_GagPro$^-$Pol

Maturation of retroviruses occurs when the Gag proteins in the immature virus are cleaved to release the mature viral protein e.g. MA, CA and NC. Hence the maturation process is

impaired when the proteolytic process is inhibited. To investigate the immature viral proteins mutations in the viral protease were introduced. The HERV-K113 protease is a member of the aspartyl protease family (Kuhelj et al., 2001). The active site consists of the conserved sequence: Asp - Thr - Gly. To inactivate the active site these conserved amino acids were replaced by site directed mutagenesis with Alanine (D204A, T205A and G206A) (Fig. 3-4) in the oricoHERV-K113_GagProPol construct.

```
HIV          --PQVTLWQR PLVTIKIGGQ LKEALLDTGA DDTLVEEMSL P----GRWKP
HERV         YWASQVSENR PVCKAIIQGK QFEGLVDTGA DVSIIALNQW PKNWPKQKAV
HERV_Pro-    YWASQVSENR PVCKAIIQGK QFEGLVAAAA DVSIIALNQW PKNWPKQKAV

HIV          KMIGGIGGFI KVRQYDQILI EICGHKAIGT VLVGPT--PV NIIGRNLLTQ
HERV         TGLVGIGTAS EVYQSTEILH CLGPDNQEST VQPMITSIPL NLWGRDLLQQ
HERV_Pro-    TGLVGIGTAS EVYQSTEILH CLGPDNQEST VQPMITSIPL NLWGRDLLQQ

HIV          IGATLNF--- ---------- ---------- ---------- ----------
HERV         WGAEITMPAP LYSPTSQKIM TKMGYIPGKG LGKNEDGIKV PVEAKINQKR
HERV_Pro-    WGAEITMPAP LYSPTSQKIM TKMGYIPGKG LGKNEDGIKV PVEAKINQKR

HIV          -------
HERV         EGIGYPF
HERV_Pro-    EGIGYPF
```

Fig. 4-4 Alignment of retroviral Protease Amino acid sequences of the protease of HIV and HERV-K113 (HERV) as well as of HERV-K113_Pro$^-$ (HERV_Pro$^-$) are shown. The box indicates the active site of the proteases DTG. HERV_Pro$^-$ carries three alanine instead of D-T-G rendering the protease a mutant inactive enzyme. (Adapted from (Kuhelj et al., 2001))

The resulting construct carrying the inactive protease was named oricoHERV-K113_GagPro$^-$Pol. Expression of the construct in HEK 293 cells led to the release of immature VLPs. This was investigated by electron microscopy, SDS gel analysis and mass spectrometry.

4.2 Expression of oricoHERV-113_GagProPol and oricoHERV-K113_GagPro$^-$Pol

To assess the functionality of oricoHERV-113_GagProPol and oricoHERV-K113_GagPro$^-$Pol, HEK-293T cells were transfected with these constructs and viral pellets collected by ultracentrifugation. Virus pellets and cell lysates of cells infected with oricoHERV-113_GagProPol and oricoHERV-K113_GagPro$^-$Pol were compared in i) electron microscopy, ii) silver nitrate stained SDS gels and iii) Western Blots.

4.2.1 Electron microscopic pictures of virus like particles

To investigate the functionality of the constructs HEK 293T cells were transfected with oricoHERV-K113_GagProPol and the protease mutant oricoHERV-K113_GagPro$^-$Pol. After two days the supernatants were taken off, viral particles collected by ultracentrifugation and prepared for Electron microscopy. Additionally the producing cells were scraped from cell plates and prepared for electron microscopy. Both, virus and cell pellets, exhibited virus production.

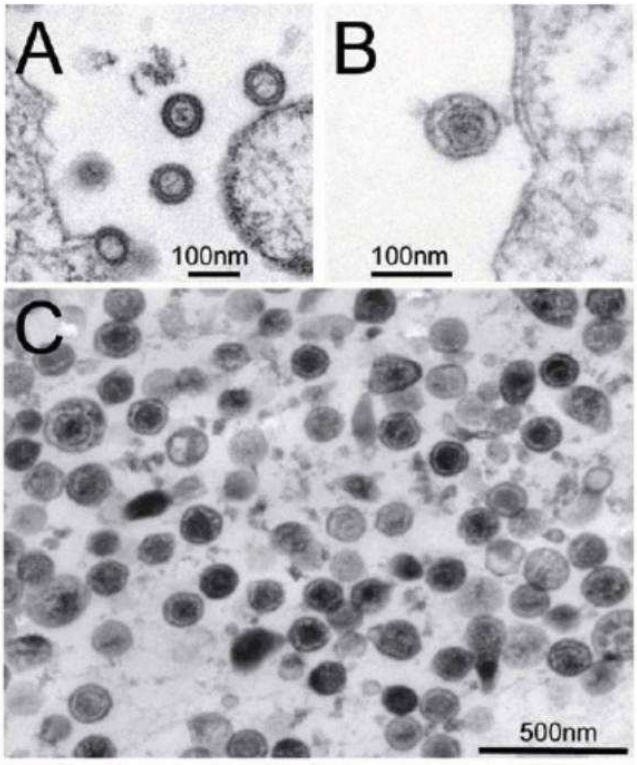


Fig. 4-5: Electron microscopic picture of virus like particles. (A) Cells transfected with oricoHERV-K113_GagPro$^-$Pol produce immature particles identified by the thick electron dense outer membrane and the electron lucent core. (B) Cells transfected with oricoHERV-K113_GagProPol produced mature viral proteins with a narrow electron dense outer membrane and an electron dense core. (C) Section of ultracentrifugated pellet of mature VLP. (George et al., 2011)

The cells transfected with oricoHERV-K113_GagProPol produced immature (Fig. 4-5 A) particles which matured after release from the cell (Fig. 3-5 B). Immature particles have an electron dense outer ring structure and electro lucent core, whereas mature particles carry an electron dense core. The core of mature particles is only formed after the mature viral proteins MA, CA and NC are released from the polyprotein Gag by the active viral protease. In virus

pellets of the construct orico-HERV-K113_GagProPol only mature virus particles were observed (Fig. 3-5 C).

4.2.2 Protein gel analysis of virus like particles

Maturation of the retroviral particle occurs when the polyprotein Gag is cleaved by the protease to release the mature subdomain proteins. Since there were mature virus particles visible in electron microscope pictures, it was of interest if the difference of the mature virus subdomain proteins and the immature virus Gag polyprotein would also be detectable on the protein level. Therefore cells were transfected with oricoHERV-K113_GagProPol and the oricoHERV-K113_GagPro$^-$Pol and virus pellets collected. The virus pellets were run on a 15% SDS gel and the gel then stained with silver nitrate.

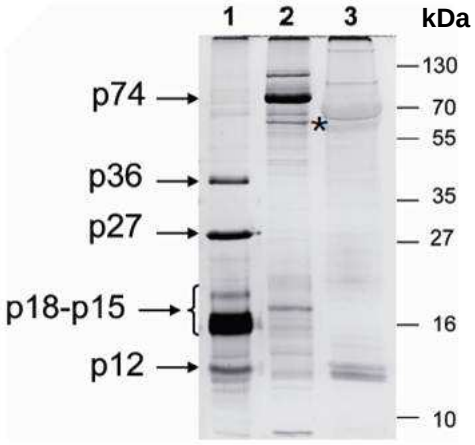


Fig. 4-6: Proteins of virus pellets of the constructs oricoHERV-K113_GagProPol and oricoHERV-K113_GagPro$^-$Pol on silver nitrate stained SDS gel. Lane 1: In protease active VLPs Gag is processed into the smaller subunits p36, p27, p18-p15 and p-12. Lane 2: Protease deficient VLPs Gag is not processed. Accordingly there is one prominent band p74 visible. Band marked with a star indicates unspecific N-terminal degradation. Lane3: Empty vector control. (George et al., 2011)

One prominent band was visible at approximately 70kDa in the lane loaded with the virus pellet of the protease defective mutant (Fig. 3-6 Lane 2). This protein migrated at a size, which correlated with the calculated size of Gag of 74kDa. In the VLP pellets of the active PR this band was only barely visible, compared to the prominent bands at 36 kDa, 27 kDa, 15-18 kDa and 12 kDa (Fig.3-6 lane 1). These bands represented the presumably processed Gag polypeptides.

4.2.3 Western Blot detection of MA, CA and NC

In the silver nitrate stained SDS gel of VLPs of oricoHERV-K113_GagProPol and oricoHERV-K113_GagPro$^-$Pol (Fig. 4-6) it was shown that processing led to several new smaller proteins. To identify the proteins spotted on silver nitrate stained gels of mature and immature VLPs they were investigated on Western Blots with antibodies raised against MA, PP and CA. Three specific rat sera had been raised before against different sections of the Pr74 (Gag) protein. The antisera were named αMA, αCA and αp15, which indicated the presumed Gag-domain they were generated against. Indeed antibodies in the sera identified certain protein bands in the PR-active VLP lysates on Western-Blot. The αMA serum bound to a protein about the size of 16 kDa, as well as to a 36 kDa protein (Fig. 4-7). The αp15 serum bound to a 36kDa big protein and to a triplet of bands around 16 kDa. The 36 kDa protein recognized by αMA and αp15 represents a processing intermediate comprising the matrix and the phosphoprotein.

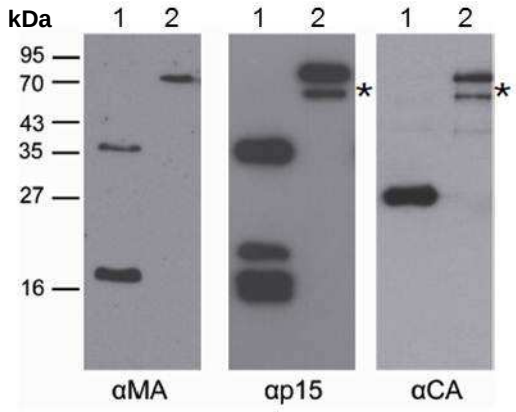


Fig. 4-7 Western Blot of oricoHERV-K113_GagProPol and oricoHERV-K113_GagPro$^-$Pol with antibodies against MA, CA and p15. Specific Gag subdomains were detected by Western Blot. Lane 1: UC virus pellet of oricoHERV-K113_GagProPol producing cells. Lane 2: UC virus pellet of oricoHERV-K113_GagPro$^-$Pol producing cells. The bands marked with an asterisk in the lane with protease mutant VLPs indicate unspecific degradation. (George et al., 2011)

4.3 Separation and analysis of viral subdomains

After the identification of the subdomains with specific antibodies the aim was to purify the subdomains, which could be used to identify the correct protein sequence of Gag cleavage product including its exact cleavage sites. The subdomains were analyzed as follows. Virus

pellets were generated and viral proteins separated by reverse-phase high pressure chromatography (RP-HPLC). Fractions of the RP-HPLC were then examined on SDS Page and/or Western Blots with the antibodies described in 3.2.1 to identify subdomains of Gag. If a subdomain was detected by Western Blot in a fraction, the fraction was either directly used for MALDI-TOF mass spectrometry or the protein first trypsin digested and then analyzed by Mass spectrometry. In case more than one protein eluted in the same fraction proteins were separated on a silver nitrate gel, the corresponding band cut out and trypsin digested to be measured by MALDI-TOF. Additionally the amino acid sequence of trypsin fragments of interest were determined by MALDI-TOF MS/MS experiments.

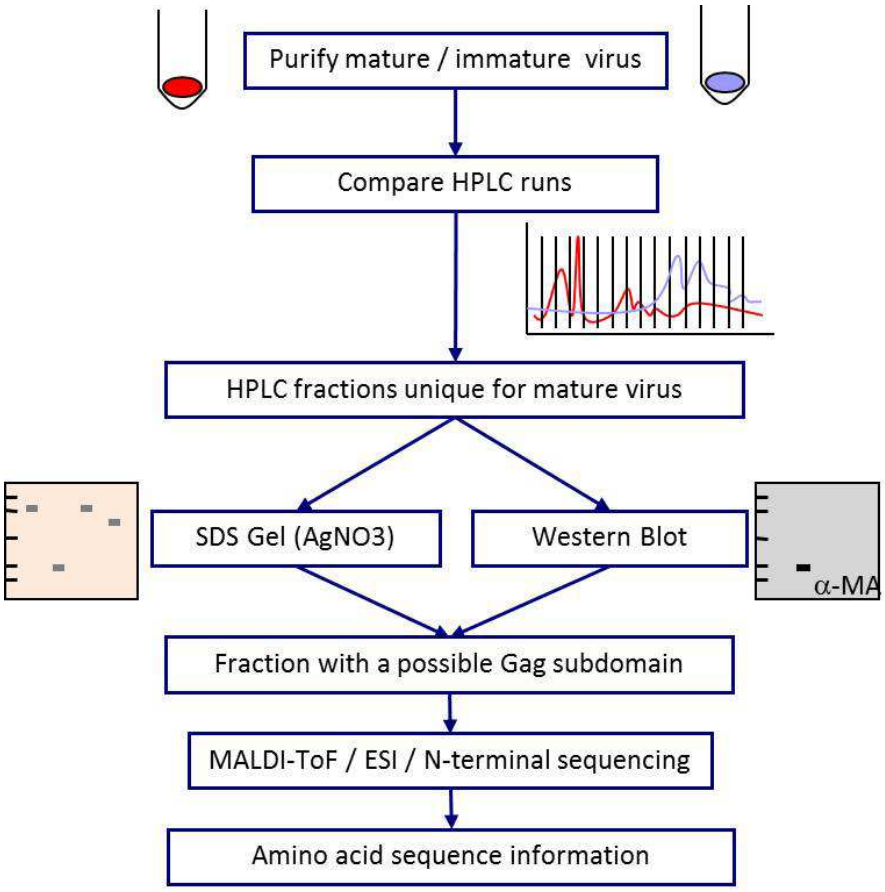


Fig. 4-8 Flowchart of the experimental design. Mature VLP pellet is depicted in red, immature VLP pellet is depicted in blue.

4.3.1 RP-HPLC of virus pellets

To assess the ability to purify Gag subdomains by RP-HPLC, first it was investigated if the RP-HPLC chromatograms of VLP pellets generated with oricoHERV-K113_GagProPol and oricoHERV-K113_GagPro^{-}Pol differed. Therefore, virus pellets were diluted in Urea buffer and filtrated through centrifugation tubes. The proteins dissolved in the supernatant were loaded on the reverse phase column and eluted for 60 min by a rising concentration of acetonitrile.

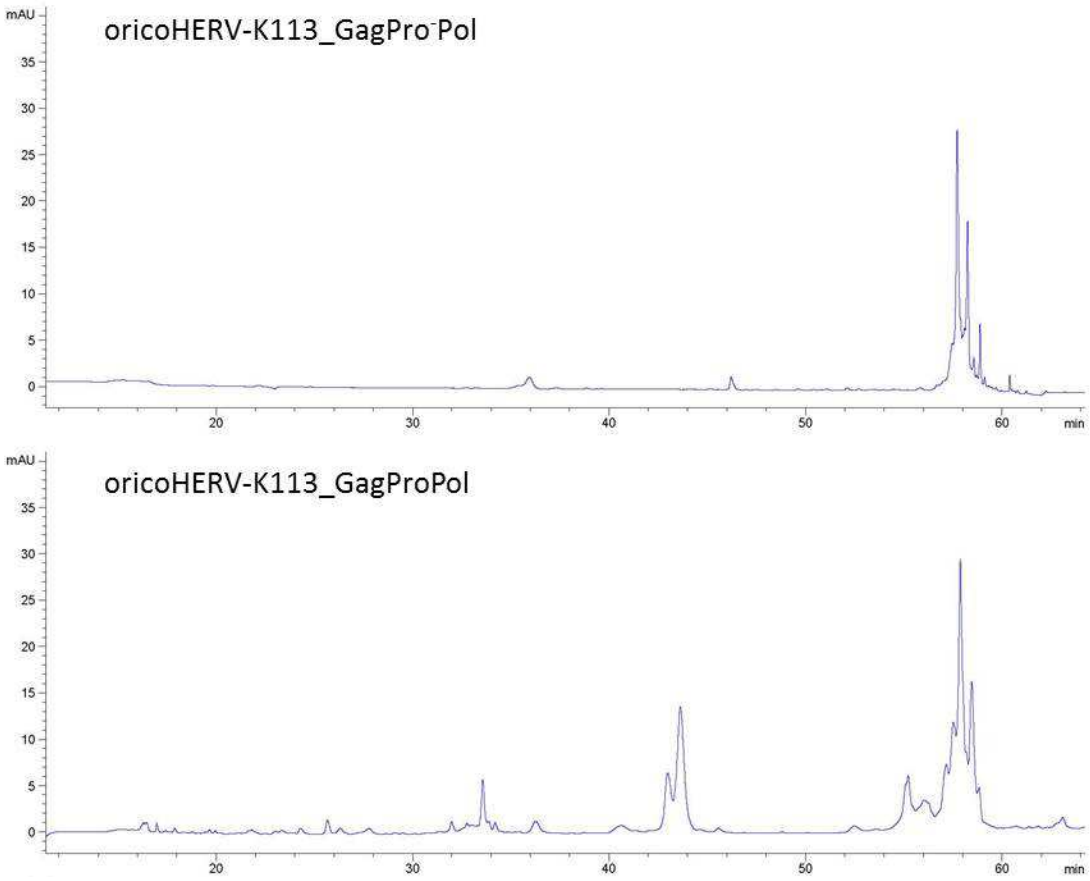


Fig. 4-9 RP-HPLC with virus pellets of protease active and protease mutant oricoHERV_GagProPol. The blue line depicts the elution of proteins measured at 280 nm. Peaks indicate the release of proteins from the reverse phase column. Virus proteins of oricoHERV-K113_GagProPol eluted earlier, resulting in several distinct peaks compared to oricoHERV-K113_GagPro^{-}Pol.

The elution of proteins was measured at 280nm. The chromatogram of virus proteins of oricoHERV-K113_GagProPol showed several peaks. The proteins started eluting after 10 min with major peaks at 33, 44 and 45 minutes. Several small peaks and the biggest peak eluted

after 54 min. Proteins from the virus pellets of oricoHERV-K113_GagPro-Pol eluted later and mainly in one major peak after 57 minutes slightly indicating a larger size which was expected from Gag protein (Fig. 4-9) although RP-HPLC does not generally .

In summary RP-HPLC chromatograms of virus pellets containing mature viral proteins differed from the ones containing immature viral proteins. Next the proteins eluting at peaks were collected and subjected to Western Blot analysis.

4.3.2 Isolation and identification of MA

Most retroviral Gag polyproteins encode N-terminally the Matrix (MA) protein. To identify MA in mature VLPs, VLP pellets of oricoHERV-K113 GagProPol were purified and used for RP-HPLC. All fractions were collected, dried and resolved in loading buffer. The samples were blotted and incubated with αMA. With a 60min run MA was detected in a fraction after 58 min.

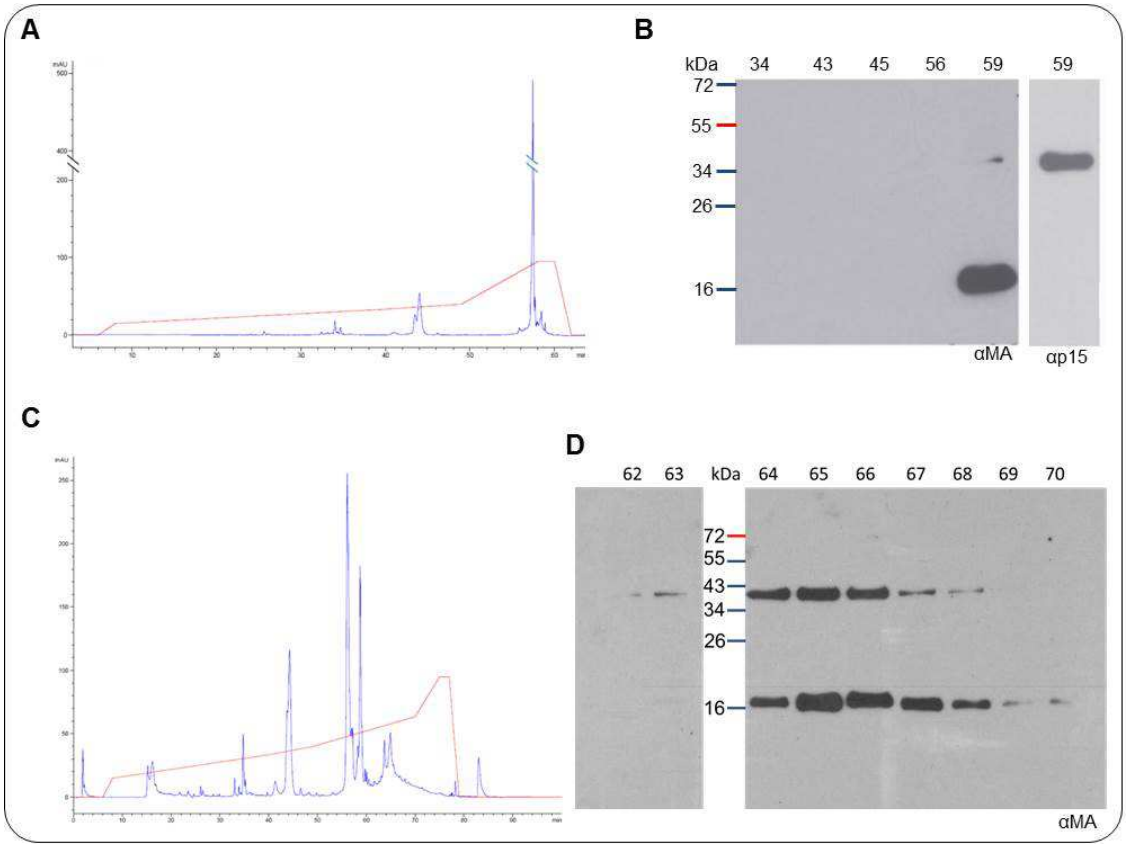


Fig. 4-10 Detection of MA and MA-p15 in VLP pellets by RP-HPLC and Western Blot. (A) RP-HPLC chromatogram of oricoHERV-K113_GagProPol viral proteins (60 min) (George et al. 2011); (B) Identification of MA and MA-p15 on Western Blot with fraction eluting at 34, 43, 45, 56, 59 minutes with αMA and αp15 (George et al. 2011); (C) Chromatogram of the extended RP-HPLC run for 90 minute of oricoHERV-K113_GagProPol viral proteins; (D) Detection of MA and MA-p15 with αMA in several fractions of an 90min RP-HPLC run.

To improve the resolution of the identified peaks the elution gradient was elongated to rise at 70 min from 50% acetonitrile to 100% at 80 minutes. This led to a refined separation of MA from CA and an elution over several fractions. Together with MA eluted the approximately 36 kDa processing intermediate spanning MA-p15, which was detectable with αMA and αp15 (Fig. 4-10 B).

4.3.3 Isolation and identification of the Phosphoproteins

Sequence comparison indicated an additional protein with an estimated size of 15 kDa between MA and CA, which was putatively named phosphoprotein (p15) similar to pp24 of MMTV. To detect this sequence a polyclonal antibody serum, αp15, had been raised against it in rats. Consecutive fractions of a RP-HPLC run of oricoHERV-K113_GagProPol were analysed on Western Blots. On these blots αp15 polyclonal antibody identified two obvious protein bands and one light band.

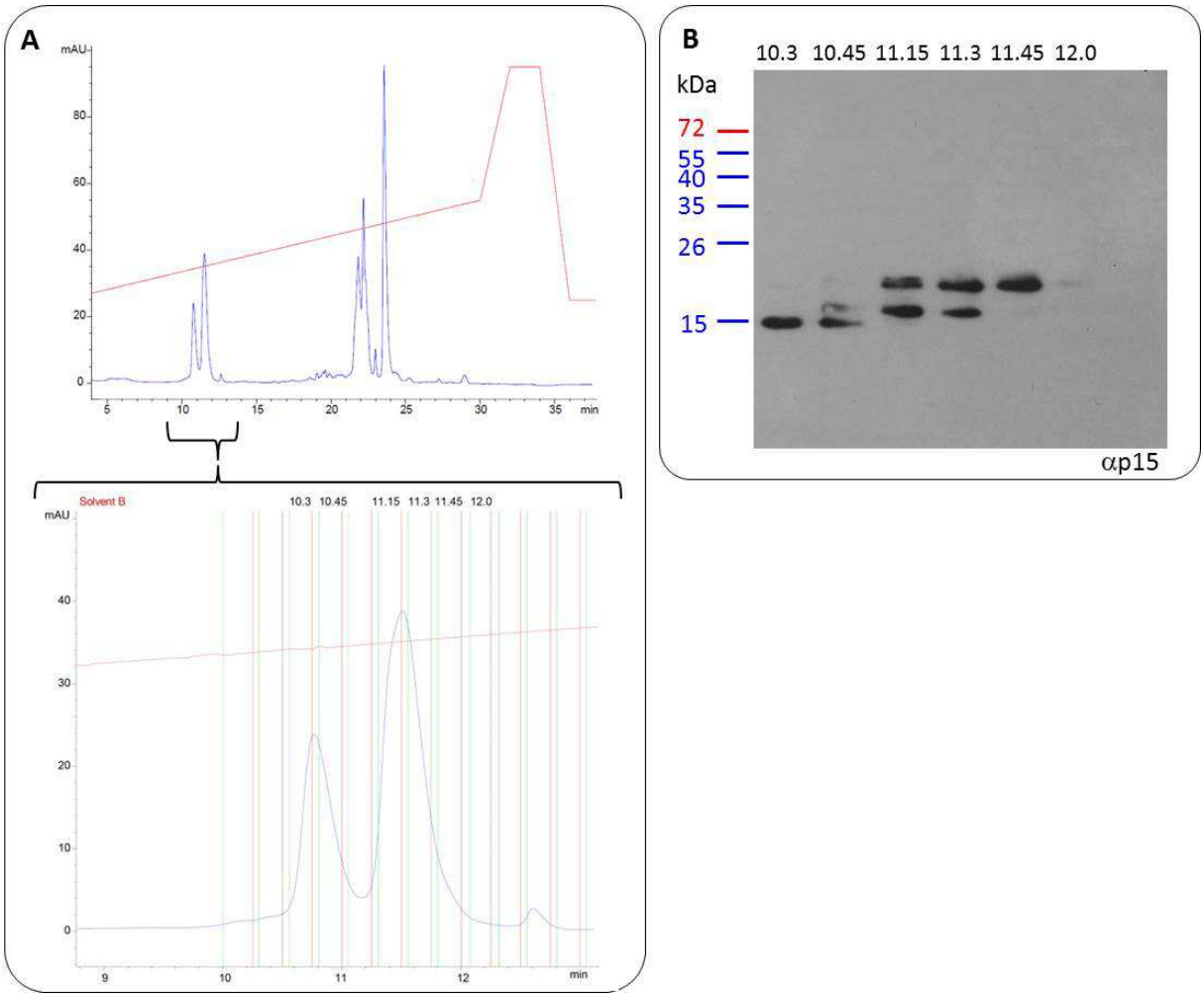


Abb. 4-1 Identification of RP-HPLC fractions containing phosphoproteins. (A) The HPLC run over 30min and zoom in on a detailed analysis of the peaks eluting minutes 10 to 12. Green lines indicate the beginning of the sample collection. Red lines indicate the end of sample collection. Rising red line indicates in both graphs the rising solvent gradient. Fraction were collected every 15 seconds (B) Western Blot analysis with αp15 of the fractions eluting from minute 10.3 till minute 12.

The fractions containing the phosphoproteins eluted at approximately 30% acetonitrile. Fractions containing phophoproteins were dried and proteins either measured directly without trypsin digestion by MALDI-MS or analyzed after the proteins had been separated on a silver nitrate stained SDS gel and trypsin digested.

MALDI MS analysis of undigested phosphoproteins fractions revealed 3 proteins, a 15 kDa protein and 2 proteins between 16kDa and 16.5kDa. These three proteins shared a common C-terminus, a 1210 Da fragment which had been detected in all three SDS gel purified and trypsinated phosphoproteins. It consisted of the sequence “KEGDTEAWQHPV” and was verified by ESI analysis. The N-terminus of the 15,008 Da phosphoprotein was also detected

in the trypsin digest and determined by the 1,738 Da fragment matching the sequence “YNQLQEVIYPETLK”. This fragment was unique for the 15,008 Da phosphoprotein and not found in any of the other phosphoproteins.

The phosphoproteins larger than 15 kDa were determined undigested by MALDI-MS (Fig.3-13 A) with a size of 16,489.527 Da and 16,187.920 Da. Since they had the same C-terminus in common, the N-terminus was initially resolved by deducing the amino acid sequence from the size of the whole protein (Fig. 3-13B). The calculated N-terminus sequence fitted very well with the 16,187 Da protein, corresponding accordingly to the starting aminoacids “PVMA…”. The measured size of 16, 489 Da differed by 10 Da from the assumed start of the last phosphoprotein at the amino acid sequence “VAE…” leading to the assumed size of 16,479 Da. Later the two most prominent phosphoproteins were additionally N-terminal sequenced and the results matched the expected sequences including the of the 16,489 Da protein. The 16,187 Da protein although visible on some silvernitrate stained gels and detectable by Maldi-MS was not detected in every virus pellet and less often detected on Western Blot. The sequence between p15 and the cleavage site producing the 16,489.527 Da protein was identified as an spacer protein and named SP1.

4.3.4 Isolation and identification of CA

HERV-K 113 contains like all retroviruses a central core particle that encapsidates the viral RNA as well as other Gag subdomains, such as nucleocapsid. As was shown by sequence alignment CA proteins are encoded after MA and carry the Major Homology Region (MHR) with conserved polar aminoacids (Fig. 4.1). To identify the exact amino acid sequence of this subdomain VLP Pellets of oricoHERV-K113 GagProPol were separated through RP chromatography and fractions blotted for Western Blot analysis. αCA detected proteins in fractions eluting consistently at approximately 40% acetonitrile. CA eluted in several large peaks during the HPLC run, depending on how fast the rise of the solvent was. MALDI-TOF analysis of the trypsin digested CA subunit revealed a 2,489.085Da fragment which matched the N-terminal sequence “PVTLEPMPPGEGAQEGEPPTVEAR”. This was later confirmed by N-terminal sequencing. The protease cleavage site between p15 and CA was confirmed as “…AWQF-PVTL…”. This cleavage site is in accordance with a typical type 1 retroviral cleavage site (Pettit et al., 1991) characterized by an aromatic amino acid at the P1 site and a proline at the P1’ site.

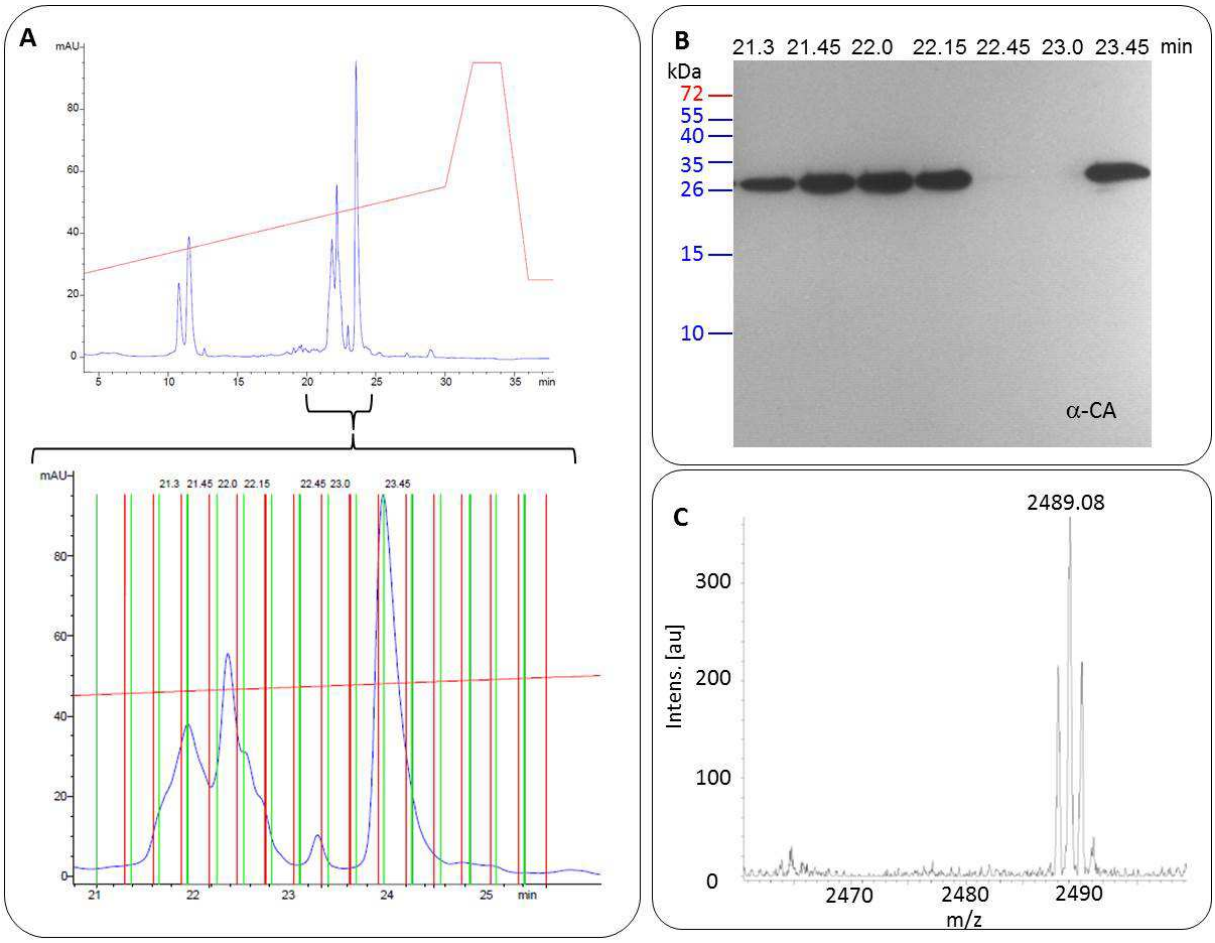
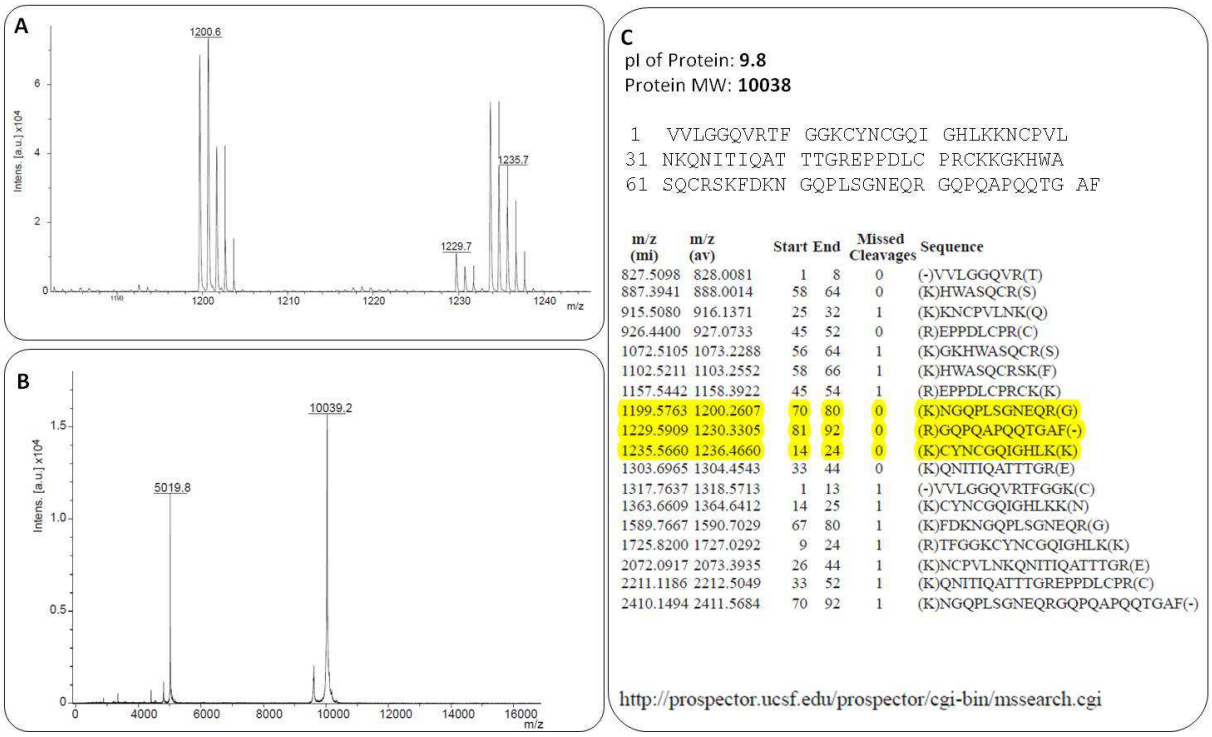


Fig. 4-11 RP-HPLC fraction of CA. (A) The HPLC run over 30min is shown on top. A zoom in of the peaks eluting at 21 to 23.5 minutes is depicted on the bottom. Green vertical lines indicate the start of the collection of the fraction, red lines indicate the end. Red rising horizontal line monitors the rise of solvent B of 0.08% trifluoroacetic acid in acetonitrile. (B) Selected fractions taken every 15 seconds from 21 to 22.3 minutes were analyzed by Western Blot and proteins were identified by αCA. (C) MALDI-TOF analysis of a trypsin digest of CA cut out of a silver nitrate stained gel reveals a fragment with the size of 2,489 Da, which corresponds to the sequence "PVTLEPMPPGEGAQEGEPPTVEAR".

4.3.5 Identification of Nucleocapsid (NC)

Beside MA and CA all retroviruses encode a nucleocapsid domain. This was also identified for HERV-K113 with an alignment of the Gag sequences of MPMV and MMTV. Zinc fingers with the characteristic sequence '$CX_2CX_4HX_4C$' flanked by small domains rich in basic residues encoded in this sequence confirmed the notion. To characterize the cleavage sites flanking NC VLP pellets of oricoHERV-K 113 GagProPol were produced and the viral proteins separated through RP-HPLC. Lacking an antibody to identify NC on Western Blot, each fraction was divided into two parts. One part was measured directly by Maldi-TOF and the other was first trypsin digested and then measured by Maldi-TOF analysis. Trypsin

fragments correlating with the expected in silica trypsin fragments in the NC sequence were measured in fractions eluting at approximately 20% acetonitrile. An especially unambiguous sign for NC were fragments with sizes correlating with in silica calculated fragments, which contained parts of the zinc fingers (Fig.4-12 A and C). The same fraction was also measured undigested and revealed surprisingly a NC smaller than expected with a size of 10,039 Da. Maldi MS/MS experiments confirmed the sequence of the first trypsin fragment of NC being "VVLG".



m/z (mi)	m/z (av)	Start	End	Missed Cleavages	Sequence
827.5098	828.0081	1	8	0	(-)VVLGGQVR(T)
887.3941	888.0014	58	64	0	(K)HWASQCR(S)
915.5080	916.1371	25	32	1	(K)KNCPVLNK(Q)
926.4400	927.0733	45	52	0	(R)EPPDLCPR(C)
1072.5105	1073.2288	56	64	1	(K)GKHWASQCR(S)
1102.5211	1103.2552	58	66	1	(K)HWASQCRSK(F)
1157.5442	1158.3922	45	54	1	(R)EPPDLCPRCK(K)
1199.5763	1200.2607	70	80	0	(K)NGQPLSGNEQR(G)
1229.5909	1230.3305	81	92	0	(R)GQPQAPQQTGAF(-)
1235.5660	1236.4660	14	24	0	(K)CYNCGQIGHLK(K)
1303.6965	1304.4543	33	44	0	(K)QNITIQATTTGR(E)
1317.7637	1318.5713	1	13	1	(-)VVLGGQVRTFGGK(C)
1363.6609	1364.6412	14	25	1	(K)CYNCGQIGHLKK(N)
1589.7667	1590.7029	67	80	1	(K)FDKNGQPLSGNEQR(G)
1725.8200	1727.0292	9	24	1	(R)TFGGKCYNCGQIGHLK(K)
2072.0917	2073.3935	26	44	1	(K)NCPVLNKQNITIQATTTGR(E)
2211.1186	2212.5049	33	52	1	(K)QNITIQATTTGREPPDLCPR(C)
2410.1494	2411.5684	70	92	1	(K)NGQPLSGNEQRGQPQAPQQTGAF(-)

Fig. 4-12 MALDI-TOF Analysis of RP-HPLC fractions containing NC. (A) MALDI-TOF spectra of trypsin digested fraction of oricoHERV-K113_GagProPol proteins. The fragment of 1,200.6 arbitrary units (a.u.) correlates with the average size of the fragment NGQPLSGNEQR. The fragment 1,229.7 a.u. correlates with the monoisotopic size of GQPQAPQQTGAF and the fragment 1,235.7 with the monoisotopic fragment CYCGQIGHLK; Y-axis: intensity in arbitrary units (a.u.); X-axis: mass to charge ratio (m/z); (B) MALDI-TOF analysis of the NC domain of oricoHERV-K113. The first major peak represents doubly charged NC (z = 2) and the second major peak NC with a single charge. (C) List of fragments by in silico digestion of 10,038 Da NC. Highlighted in yellow are the fragments correlating with the peaks in (A). pI-isoelectric point; MW-molecular weight; The in silico analysis was performed with an online tool by the university of California, San Franciso: http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest.

4.3.6 N-terminal sequencing

To confirm the identity of the so far discovered proteins, proteins of oricoHERV-K113 VLPs were fractioned in a 60min HPLC run, employing the following gradient: solvent B, 0-6 min, 0%; 8 min, 15%; 49 min, 40%; 58min, 95%; 58-60 min, 95%; 62 min, 0%. Fractions containing the proteins of NC, the Phosphoproteins p15 and p16.5 as well as CA were collected and dried. Then they were blotted on PVDF membrane and stained with the dye Pounceau red. The protein bands were cut out and sent to be sequenced by Proteome Factory, Berlin. Sequencing confirmed the cleavage sites (Table 1).

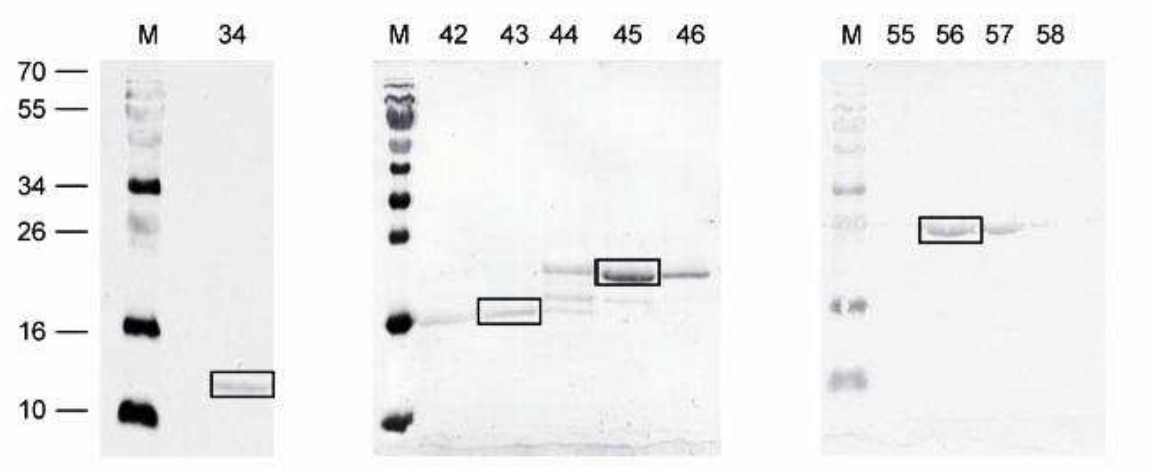


Fig. 4-13 Cleavage sites determined by N-terminal sequencing. Fractions containing the proteins NC, P15, P16.5 and CA were blotted on PVDF membrane and stained with Pounceau red. Framed protein bands were cut out for N-terminal sequencing. M-Marker, sizes are indicated on the left. Numbers on top of blot indicate the elution time in minutes. (George et al.)

Sequencing confirmed the identity and the N-terminus of the Phosphoproteins, Capsid and Nucleocapsid. It also substantiated that the N-terminus of p15 and p16 is separated by the 14 amino acid peptide “VAEPVMAQSTQNVD”, which had been named SP1.

Table 4-1 Cleavage sites identified by N-terminal sequencing of purified Gag domains

Subdomain	P4	P3	P2	P1	-	P1’	P2’	P3’	P4’	P5’
MA –SP1	His-	Cys-	Glu-	Tyr-	:	-Val	-Ala	-Glu	-Pro	-Val
Sp1-p15	Gln-	Asn-	Val-	Asp-	:	-Try	-Asn	-Gln	-Lys	-Gln

P15-CA	Ala-	Trp-	Gln-	Phe-	:	-Pro	-Val	-Thr	-Lys	-Glu
CA-NC	Ala-	Iso-	Thr-	Gly-	:	-Val	-Val	-Lys	-Gly	-Gly

Aminoacids identified by N-terminal sequencing are underlined (George et al.)

4.3.7 Mutational analysis of cleavage sites identified by N-terminal sequencing

Oshima et al. 2004 showed in their experiments that the exchange of just one amino acid at the P1 site of Gag cleavage sites in Moloney murine leukemia virus abrogated cleavage. To confirm the cleavage site for maturation, several P1 mutants were created by site directed mutagenesis. At the P1 site between MA and SP1 Tyrosine was exchanged with Alanine leading to a profound decrease of mature 16kDa Matrix and an increase of MA-SP1. The Exchange of Phenylalanine with Aspartic Acid at the P1 site of PP and CA severely hampered the maturation process. It led to a decrease in Matrix as well as reduced the release of CA to near elimination and completely abrogated the release of free p15. It increased also an intermediate protein containing Matrix, SP1, p15 and CA since this band was detected with αMA, αp15 and αCA (Fig. 4-14 B). Interestingly, the exchange of Glycine to Aspartic acid at the P1 site of CA-NC led to a slight decrease of mature Capsid protein but to a strong increase of other fusion proteins, such as CA-NC. This implicates that the viral protease of oricoHERV-K113_GagProPol tolerates an Aspartic acid at the P1 site and cleavage is reduced but not abolished.

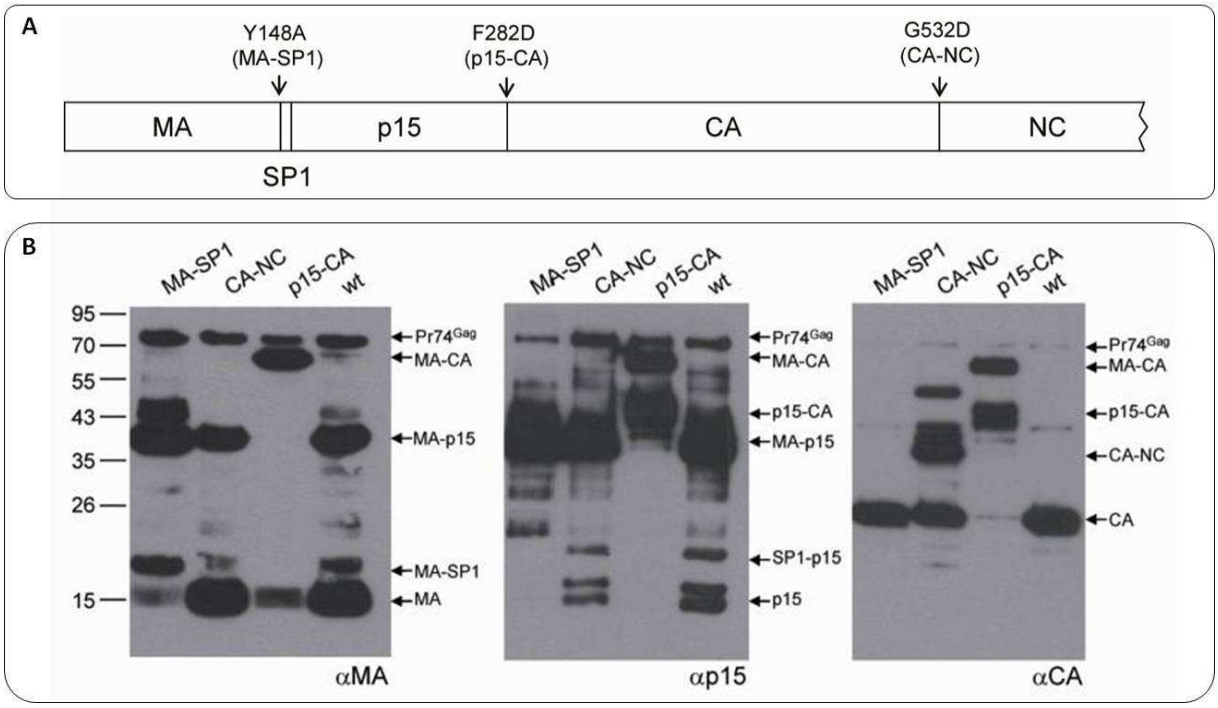


Fig. 4-14 Western Blot analysis of cleavage site mutants. (A) Schematic picture of Gag confirmed by N-terminal sequencing without the C-terminal sequence downstream of NC. Indicated are the introduced amino acid exchanges at the P1 site of MA-SP1, p15-CA and CA-NC. (B) Western Blot analysis of virus proteins of the oricoHERV-K113 GagProPol mutants MA-SP1 (lane 1), CA-NC (lane 2) and p15-CA (lane 3) (adapted from George et al. 2011)

4.3.8 Identification of QP1

Amino acid sequence analysis suggested that similar to HIV or MPMV the HERV-K113 Gag sequence encodes for a C-terminal protein with a size of approximately 5kDa. Confirming this hypothesis was the fact that detailed MALDI-TOF analysis of NC trypsin fragments had already revealed the C-terminal cleavage site of NC as “TGAF”, a typical type 1 cleavage site for a retroviral protease. Because of this it was clear that there had to exist an additional Gag subdomain. But detailed analysis of all fractions of RP-HPLC fractionated UC-Pellets did not reveal an approximately 5 kDa protein for the remaining Gag sequence “PIQPFVPQGFQGQQPPLSQVFQGISQLPQYNNCPPPQAAVQQ”. One obstacle was that there was no antibody available against this amino acid sequence. Additionally this Glutamine (Q) and Proline (P) rich sequence did not contain a Lysine (K) or an Arginine (R) and therefore could not be processed by trypsin for a detailed MALDI-TOF analysis.

To further analyze the proteins C-terminal to NC, the Phenylalanine at the P1 position of NC was exchanged in the construct oricoHERV-K113_GagProPol with an Aspartic Acid by site

directed mutagenesis (F624D mutation). Analysis of the UC-Pellets of this new mutant HERV-K113 NC F624D on silver nitrate stained SDS gels, revealed that the mutation decreased the amount of mature NC (Fig. 4-15 A). But another protein band about 2.5 kDa in size bigger than NC appeared. Together this confirmed that the cleavage site was severely impaired by the mutation and let to the rise of a new fusion protein. Next the proteins of UC-Pellets of the mutant oricoHERV-K113_GagProPol_F624D were separated on RP-HPLC and fractions containing the mutant NC F624D were trypsin digested. MALDI-TOF analysis of the trypsin fragments revealed a new fragment of the size of 3705Da. This fragment matched the sequence "RGQPQAPQQTGAD- PIQPFVPQGFQGQQPPLSQVFQG", the C-terminus of NC as well as the new identified protein. Since the fragment is not generated via trypsin it could have only be generated by the viral protease. Due to its Glutamine and Proline rich sequence this protein was named QP-1.

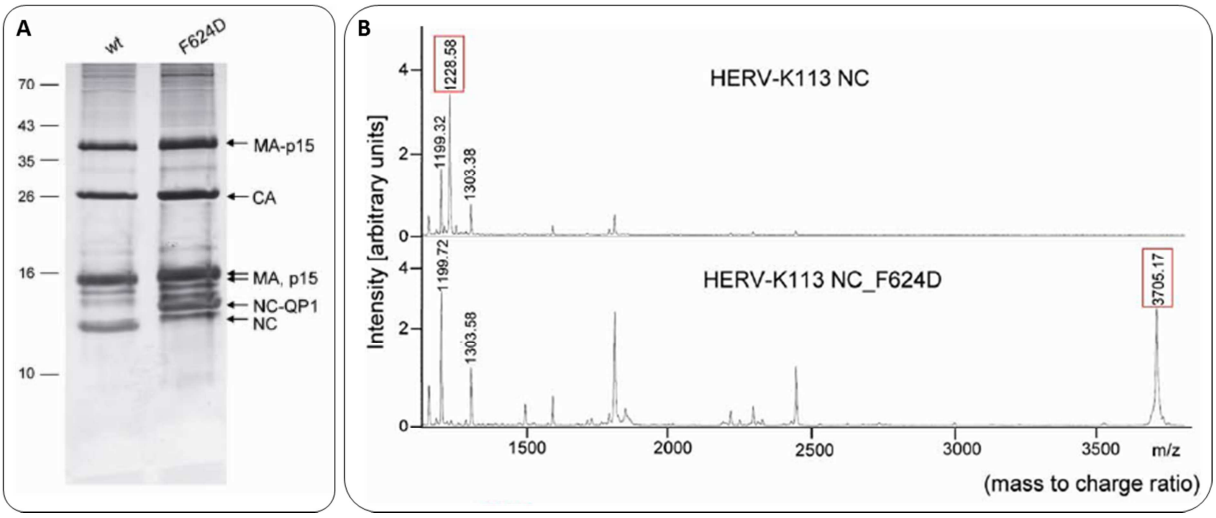
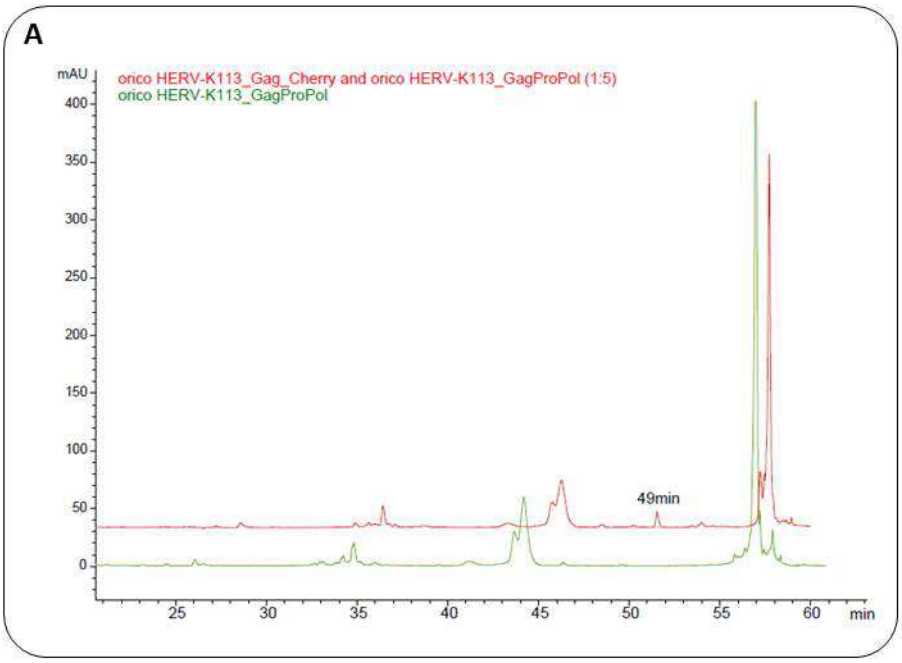


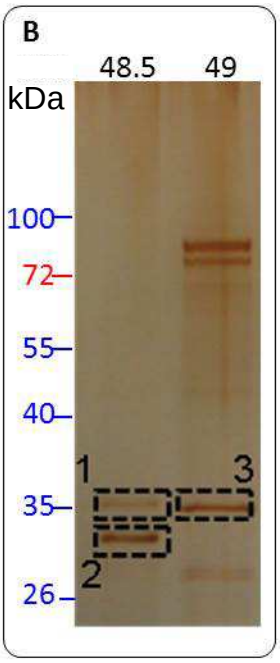
Fig. 4-15 Identification of QP1. (A) SDS gel analysis of the mutant oricoHERV-K113_GagProPol_F624D, with an exchange of Phenylalanine with Aspartic acid at the P1 site of NC. Free NC is diminished, but there is a rise in NC fusion protein. (B) MALDI-TOF MS analysis of trypsin digested NC and the new NC fusion protein. The Analysis of the fragments reveals a shift of the 1228.58 fragment, representing the C-terminal sequence GQPQAPQQTGAF of NC, to a 3705.17 fragment in the P1 mutant NC (HERV-K113 NC_F624D). This fragment correlates with the sequence "GQPQAPQQTGAD-PIQPFVPQGFQGQQPPLSQVFQG" and comprises the last tryptic fragment of the P1 mutant NC and the QP1 fragment, since this fragment could have only arisen by the N-terminal cleavage of trypsin and the C-terminal cleavage of the retroviral protease. (adapted from (George et al., 2011))

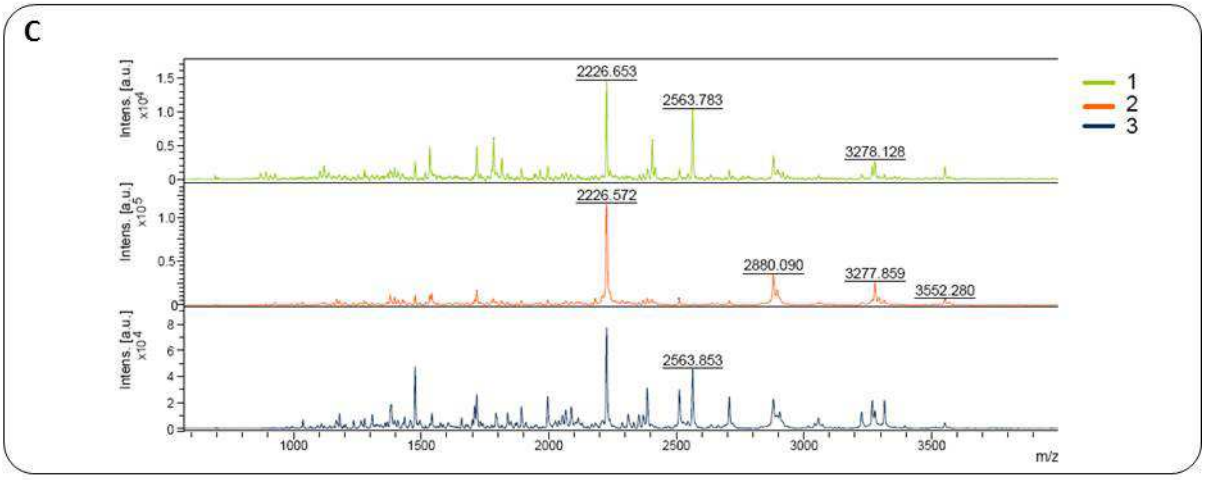
4.3.9 Verification of the last Gag subdomain QP2

After identifying QP1 one last C-terminal protein sequence of oricoHERV-K113 GPP remained elusive. Calculation determined it to be a protein with a size of 2,096.35 Da. Because it is rich in Glutamine and Proline as well, it was called QP2. Lacking an antibody against this sequence it was impossible to identify a RP-HPLC fraction of UC-pellets with this protein sequence. To overcome this obstacle a construct, called oricoGagCherry, carrying the oricoHERV-K113 Gag sequence linked with the sequence of the fluorescent protein mCherry was applied, which was provided by Anja Zimmermann. The advantage of this construct was that there were antibodies against mCherry (αCherry) readily available. Cells transfected with oricoGagCherry and oricoHERV-K113 GagProPol produced fluorescent VLPs. The correct ratio of oricoGagCherry and oricoHERV-K113_GagProPol for successful transfection for virus production in HEK 293T cells had been verified (information provided by Anja Zimmermann).

To generate VLPs the construct oricoGagCherry had to be transfected together with oricoHERV-K113 GagProPol by a ratio of 1:5. Proteins of UC virus pellets containing oricoGagCherry were separated by RP-HPLC and fractions analyzed on Western Blots with αCherry. Fractions in which a faint signal of Cherry was detected with αCherry eluted at approximately 42% acetonitrile, which also correlated to a new peak at approximately 42% acetonitrile or minute 49 (Fig 4-16 A). Two fractions from minute 48.5 and 49 were run on SDS gels, which were then silver nitrate stained. Three bands close to the expected size of mCherry as well as mCherry conjugated with the last amino acid sequence C-terminal of Gag were cut out and trypsin digested (Fig 4-16 B). MALDI MS analysis of the trypsin digests revealed that all 3 bands carried fragments of the mCherry sequence (Fig 4-16 C, D). Additionally, there was one fragment in the trypsin digest of the bands 1 and 3 (Fig 4-16 B, C) with the size of approximately 2,563Da. This fragment correlated with the sequence “ISQLPQYNNCPPPQAAVQQLVPR”, which carries the putative QP2 as well as the first four amino acids of the Cherry protein. This confirmed the existence of the QP2 protein.







D

1 ISQLPQYNNC PPPQAAVQQL VPRARDPPVA TMVSKGEEDN MAIIKEFMRF KVHMEGSVNG HEFEIEGEGE GRPYEGTQTA
81 KLKVTKGGPL PFAWDILSPQ FMYGSKAYVK HPADIPDYLK LSFPEGFKWE RVMNFEDGGV VTVTQDSSLQ DGEFIYKVKL
161 RGTNFPSDGP VMQKKTMGWE ASSERMYPED GALKGEIKQR LKLKDGGHYD AEVKTTYKAK KPVQLPGAYN VNIKLDITSH
241 NEDYTIVEQY ERAEGRHSTG GMDELYK

m/z (mi)	m/z (av)	Modifications	Start	End	Missed Cleavages	Sequence
1377.6467	1378.5560		162	174	0	(R)GTNFPSDGPVMQK(K)
1540.8846	1541.8451		221	234	0	(K)KPVQLPGAYNVNIK(L)
2211.0943	2212.5913		87	106	0	(K)GGPLPFAWDILSPQFMYGSK(A)
2225.0357	2226.3757		235	252	0	(K)LDITSHNEDYTIVEQYER(A)
2561.3293	2562.9718		1	23	0	(-)ISQLPQYNNCPPPQAAVQQLVPR(A)
2603.3399	2605.0093	1Acetyl	1	23	0	(-)ISQLPQYNNCPPPQAAVQQLVPR(A)
2878.3451	2880.1762		132	157	0	(R)VMNFEDGGVVTVTQDSSLQDGEFIYK(V)
3274.4706	3276.5006		52	81	0	(K)VHMEGSVNGHEFEIEGEGEGRPYEGTQTAK(L)

Fig. 4-16 Confirming QP2 by MALDI-MS. (A) 3D signal overlay of RP-HPLC run with UC Pellets of VLPs generated with oricoHERV-K113_Cherry/oricoHERV-K113_GagProPol (red) and oricoHERV-K113_GagProPol (green). There is a new peak after 49minutes eluting compared to the chromatogram of oricoHERV-K113_GagProPol. (B) Fractions from RP-HPLC run with viral proteins of oricoHERV-K113_Cherry/oricoHERV-K113_GagProPol from minutes 48.5 and 49 are separated on a SDS gel and silver nitrate stained. Dashed Boxes indicate the 3 protein bands which were cut out for MALDI-MS analysis. (C) MALDI MS spectra of the trypsinized proteins of the bands 1, 2 and 3 (see B). The size of the fragment of major peaks is stated above the peaks. (D) On top is the protein sequence of QP2 and mCherry which was used for in silico trypsin digestion on the website: http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest. Underneath is the list of trypsin generated fragments of the sequence with a size between 13500Da to 33000Da. Highlighted in yellow are the fragments, which have been identified in the MALDI-MS spectra (see C).

4.4 RT-activity and Electron microscopic pictures of Cleavage mutants

Mutations at the P1 position, the first amino acid upstream of the scissile bond, resulted in an increase of fused Gag subdomains on Western Blot (see 4.3.7). The exchange of Phenylalanine with Aspartic Acid at the P1 site of p15 and CA even nearly abolished the maturation process (Fig 4-14 B) and led to a rise of the Gag polyprotein. Interestingly none of the mutations though led to a visible reduction of viral proteins itself on Western Blot.

The enzyme Reverse Transcriptase synthesizes complementary DNA from an RNA template. It is encoded in the pol gene and integrated in the viral particle as a polyprotein together with Gag and Pro. After the egress of the virus from the cell also RT is set free by the protease during the maturation process.

To investigate the influence of the cleavage site mutations further the Reverse Transcriptase (RT) activity of the P1 mutant of the sites MA-SP1, SP1-p15, p15-CA, CA-NC, NC-QP1 and QP1-QP2 in the construct oricoHERV-K113_GagProPol were measured with a Cavidi Assay.

Therefore, cells were transiently transfected and the cell culture supernatant, containing released viral particles, applied according to the assay manual. Confirming the importance of the functional cleavage site between p15 and CA the P1 mutation at p15-CA also abolished RT activity. Mutating the P1 site between CA and NC led to a profound decrease in RT-activity. The mutation at the P1 site between MA and SP1, on the other hand, caused an increased RT activity by nearly 40%. Mutating the cleavage sites SP1-p15 and NC-QP1 only caused a slight decrease and a slight increase, respectively, of RT-activity. Interestingly, the

RT activity increased approximately 25%, when the Glycine at the P1 position of QP1-QP2 was exchanged with an Aspartic Acid.

As a reference for further discussion two additional constructs were used. The construct ΔQP1/2 was established by inserting a stop codon downstream of NC. It had been shown in the construct oriHERV-K113 that the deletion of QP1 and QP2 by this method would lead to an increased RT activity. This was also case with the ΔQP1/2 oricoHERV-K113_GagProPol construct. RT activity nearly doubled with this construct. In summary, there is no uniform effect on the RT activity when the P1 site is mutated.

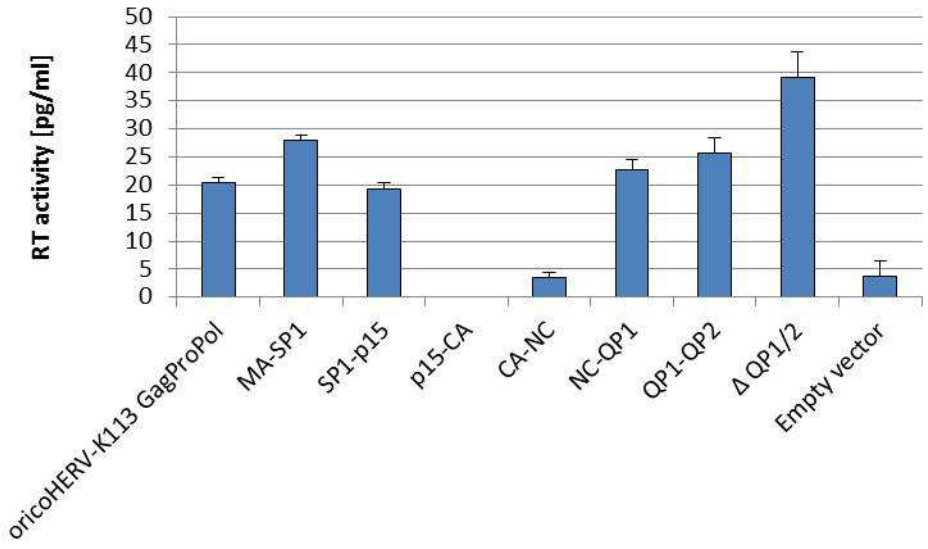


Fig. 4-17 RT activity of cleavage site P1 mutants measured by Cavidi assay. Mutating the P1 site at MA-SP1 (Y148A), SP1-p15 (D148A), p15-CA (F232D), CA-NC (G532D), NC-QP1 (F624D) and QP1-QP2 (G647D) led to different RT-activities. A stop codon inserted downstream of NC generated the deletion mutant ΔQP1/2. Empty vector control was the pcDNA3 vector.

Additionally to the RT activity, influence of the cleavage site mutations on the virion morphology was investigated by electron microscopy. Therefore, 293T cells were transfected with the cleavage mutant constructs and after two days the cells were fixated with Glutaraldehyde and prepared for electron microscopy.

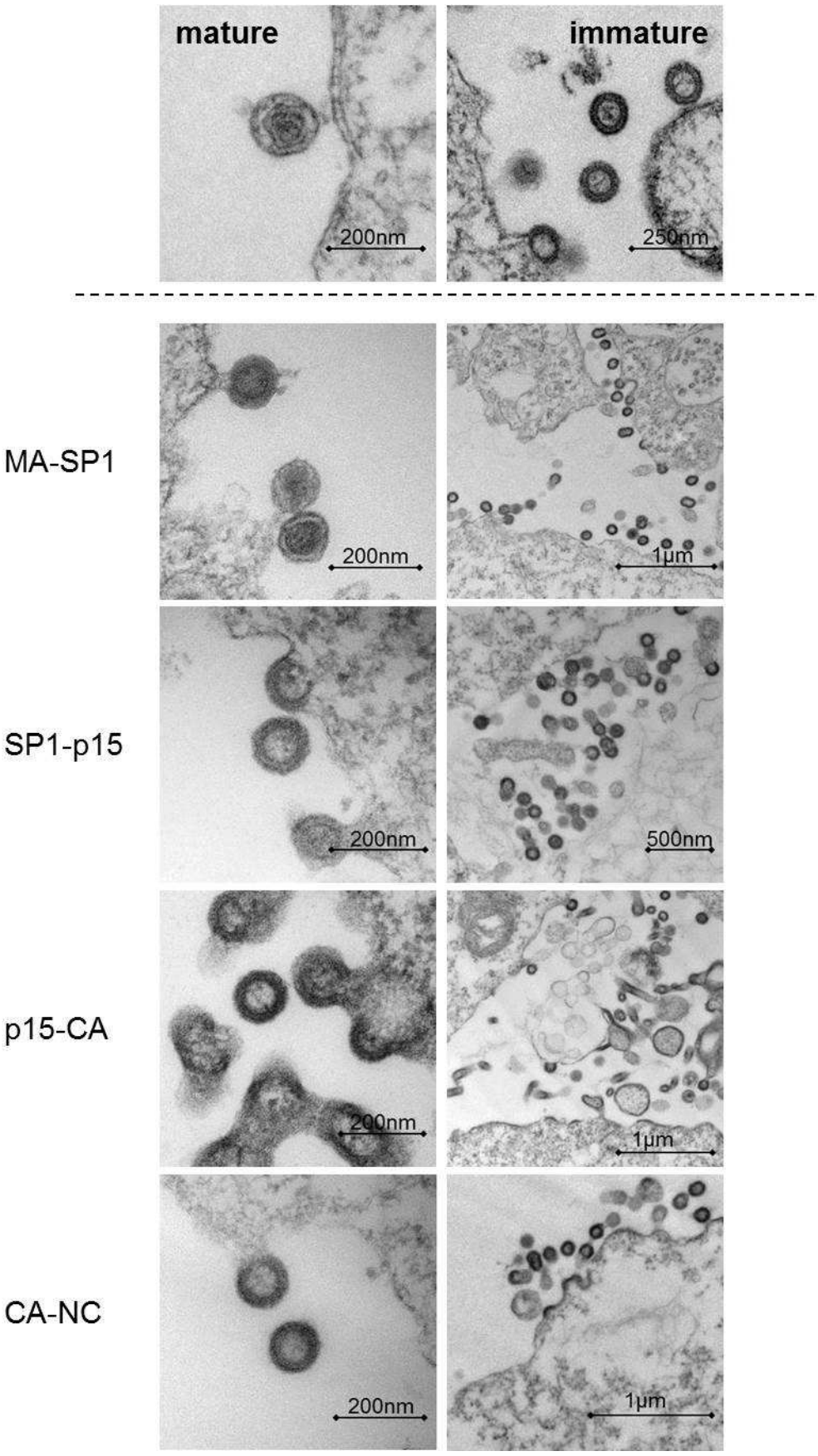

mature
immature
200nm
250nm
MA-SP1
200nm
1µm
SP1-p15
200nm
500nm
p15-CA
200nm
1µm
CA-NC
200nm
1µm

Fig. 4-18 Morphology of P1 cleavage site virus mutants. Top row: Electron microscopic pictures of a mature virion with an electron dense box shaped core and an immature virion with an electron dense outer ring and an electron lucent core. The mature virion was produced with the construct oricoHERV-K113_GagProPol and the immature virion was produced with the protease inactive construct oricoHERV-K113_GagPro^{-}Pol. Lower rows: Electron microscopic pictures of the P1 cleavage site mutants MA-SP1 (Y148A), SP1-p15 (D148A), p15-CA (F232D) and CA-NC (G532D)

EM pictures of cleavage mutant MA-SP1 showed mature and immature virus like particle. There were many immature particles visible close to the fixed cells, since the majority of mature VLP is released into the cell culture supernatant. As in the Western Blot analysis (data not shown) and in the Cavidi assay mutating the P1 site of SP1-P15 did not lead to any characteristic new phenotype. There were few mature among mostly immature VLPs. Mutations in the p15-CA P1 site led to the most severe phenotypic change. Budding defects occurred which resulted in chains of virus like particles and in long tubular structures. Additionally, cells expressing the mutant contained pleiomorphic accumulation of Gag protein at their plasma membranes, which were different from those seen with regular orioHERV-K113_GagProPol VLP producing cells. They did not curve in a radius characteristic to produce a spherical particle. The CA_NC P1 mutant exhibited only immature particles and no other phenotypic changes.

4.5 Phosphorylation of p15

Sequence alignment as well as previous data (Claudia Chudak) had shown that the subdomain P15 carries the L-domain amino acid sequence “PTAP” (Fig. 4-1). The L stands for “Late assembly” since mutations in these domains lead to typical defects during the assembly of the virions and their inability to separate from the plasma membrane. It had been shown for several L-domain containing proteins, such as murine leukemia virus p12 (Yueh and Goff, 2003), the pp16 protein of Mason-Pfizer monkey virus (Bradac and Hunter, 1984) or the HIV-1 p6 protein (Hemonnot et al., 2004), that they are phosphorylated.

Initial experiments with Pro Q diamond solution stained PP15 revealed that the p15 protein is phosphorylated (data not shown). We then performed ESI analysis on tryptic digested purified p15 protein. Phosphorylation is indicated when a tryptic fragment size is increased by 79.9Da. The digest contained non-phosphorylated as well as phosphorylated trypsin fragments. In summary five phosphorylated serine and three threonines were identified in the p15 sequence.

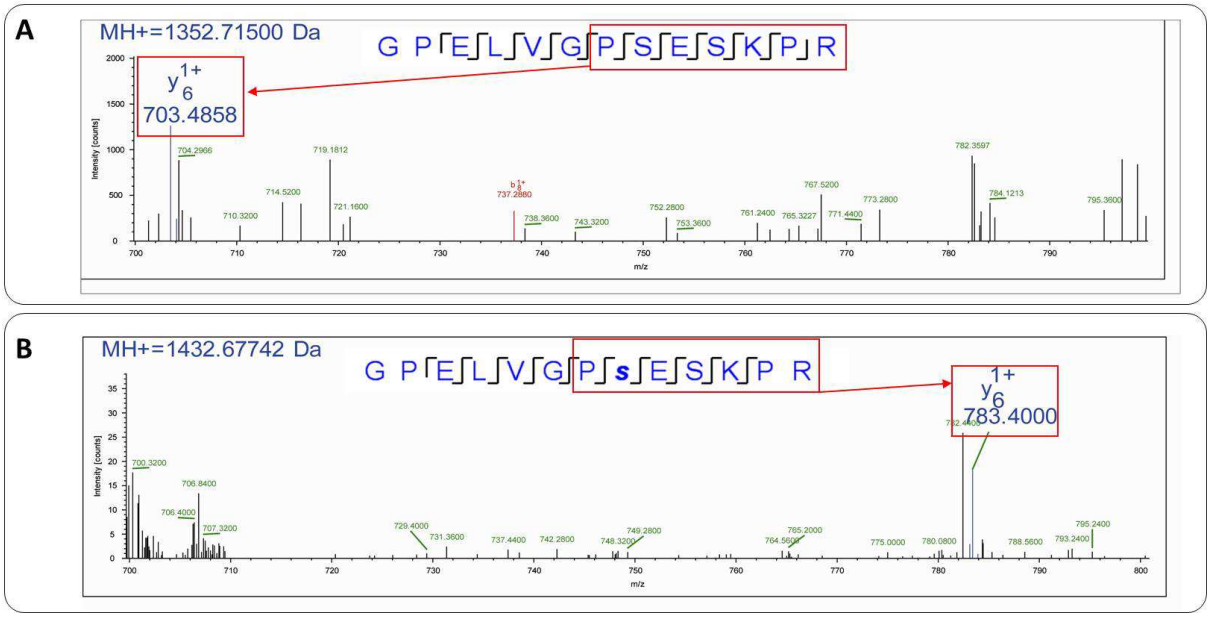


Fig. 4-19 Analysis of tryptic fragments of PP15. Representative MS/MS spectrum of the tryptic phosphopeptide „GPELVGPSESKPR“. (A) The unphosphorylated trypsin fragment with a MH+=1352Da is 79.9 Da smaller than (B) the phosphorylated trypsin fragment (MH+1432) indicating a phosphorylation site in this trypsin fragment. The shift of the Y61+ ion from 703.48 to 783.4 Da maps the phosphorylation to Ser-174 (indicated by the red boxes in A and B).

```
     MA(15.3 kDa)
   1 MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGKELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTEED
                                        SP1(1.5 kDa)  p15(15,0 kDa)
 101 SVSVSDAPGSCIIDCNEKTRKKSQKETESLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPSESKPRGTSPLPAGQVPVTLQPQKQVK
                                                                                    CA(27.7 kDa)
 201 ENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPPAPQGRAPYPQPPTRRLNPTAPPSRQGSELHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEP

 301 PTVEARYKSFSIKMLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQ

 401 NWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVIVELMAYENANPECQSAIKPLKGKVP
                                      NC(10,0 kDa)
 501 AGSDVISEYVKACDGIGGAMHKAMLMAQAITGVVLGGQVRTFGGKCYNCGQIGHLKKNCPVLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFD
                              QP1(2.5 kDa)            QP2(2.1 kDa)
 601 KNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQPPLSQVFQGISQLPQYNNCPPPQAAVQQ
```

Fig. 4-20 Amino acid sequence of Pr74Gag depicting all processing sites and the molecular masses of the subdomains, as well as phosphorylated amino acids in p15.

5. Discussion

The ability of retroviruses to insert their genomic information into the genome of a cell infers the possibility to influence the fundamental gene expression and thereby the protein expression as well as the whole characteristics of a cell. It has been shown to have positive effects and detrimental effects on the host cell. For example the expression of the HERV-W env protein during embryogenesis is vital for the development of the human embryo (Cheynet

et al., 2000). An example for detrimental effects is the development of cancer through HTLV (Jern and Coffin, 2008). It was inevitable that, during evolution retroviruses integrated into the germ cell of an organism and were then, if they were not too detrimental for the organism, passed on vertically, from one generation to the next. Today about 8% of the human genome consists of HERVs or HERV related sequences (Lander et al., 2001). Some HERV have been implicated in autoimmune diseases such as Opitz syndrome (Landry et al., 2002) or in the development of cancer (Hanke et al., 2013a; Ono et al., 1987; Wang-Johanning et al., 2003). Host cells developed defence mechanisms to combat the viral intruders such as the APOBEC family, tetherin and most likely other yet unknown factors (Lee and Bieniasz, 2007; Lee et al., 2008). Many HERVS acquired postinsertional mutations and deletions during their residency in the host cell, leaving them inactive.

The HERV-K clade consists of betaretrovirus-like endogenous retroviruses and includes the most recently active retroviruses (Macfarlane and Simmonds, 2004). One member of the HERV-K clade, HERV-K113 possesses open reading frames for all viral genes, but it acquired postinsertional mutations which render it replication incompetent (Lee et al., 2008; Subramanian et al., 2011).

Although HERVs have been implicated in the development of various cancer, leukaemia, schizophrenia and rheumatoid arthritis, fundamental properties of the virus have not been investigated yet. The aim of this study was to identify the exact cleavage sites in the Gag polyprotein and the number of subdomains of an HERV-K Gag. For the investigation the reconstituted gene sequence of the most recent integrated HERV-K113 was used.

5.1 Expression of oricoHERV-K113_GagProPol

To overcome the low virus production due to several postinsertional mutations the viral gene sequence of HERV-K113 was reconstituted according to an algorithm, which was previously successfully employed to reconstruct the original HERV-K113 envelope gene (Hanke et al., 2009). The aim of the reconstitution was to work with a protein sequence either identical or very closely related to the virus, which integrated approximately one million years ago. Therefore non-synonymous post-insertional mutations were identified by sequence comparison with 10 human specific well preserved HERV-K elements and reversed. With a similar method two infectious HERV-K genomes had been generated before, but in both cases

the most frequent nucleotide had been taken for each position (Dewannieux et al., 2006; Lee and Bieniasz, 2007). In contrast, for the oricoHERV-K113_GagProPol sequence only if none or one other sequence encoded the same amino acid it was considered a mutation and reversed to the most prevalent amino acid. If there were two or more sequences bearing the same nucleotide it was considered a variation. Additionally the gene sequence of Gag, Pro and Pol was codon-optimized and then cloned under control of a CMV-promoter, which lead to the production of high viral loads sufficient to analyse the correct Gag cleavage sites.

When analysed in thin section microscopy it was notable that oricoHERV-K113_GagProPol particles were assembled and released from the plasma membrane of transfected 293T cells. This is in contrast to the HERV-K113 related betaretroviruses such as Mouse Mammary tumour virus and Mason-Pfizer monkey virus, which assemble non-enveloped particles within the cytoplasm of infected cells (Menéndez-Arias et al., 1992; Ono et al., 1986; Rhee et al., 1990). Since HERV-K113 are related to beta-retroviruses based on the homology of the pol-gene it is another indication that viral morphology and site of assembly might be dependent on Gag (Dewannieux et al., 2006). In addition it had also been shown before that a single amino acid change in the Matrix region of MPMV Gag proteins leads to a switch from cytoplasmic to plasma membrane associated assembly (Rhee et al., 1990). In congruence with our findings previous publications showed that the two reconstituted HERV-K sequences, Phoenix and HERV-K_{CON}, as well as the few human cell lines, which express HERV-K, exhibit plasma membrane associated assembly of viral particles (Bieda et al., 2001; Boller et al., 2008; Dewannieux et al., 2006; Lee and Bieniasz, 2007).

Electron microscopic pictures of transiently transfected 293T cells showed mostly immature VLPs with an electron-dense outer ring and an electrolucent core located close to the cells. On electron microscopic pictures of pelleted virus, on the other hand, only mature virions with collapsed electron-dense core were visible. This indicated, that the proteases in the VLPs are active and that the majority of particles mature after being released from the cell. The maturation seems to be more delayed compared to other retroviruses, such as HIV. Maturation was also visible when virus pellets of protease competent oricoHERV-K113_GagProPol were compared to protease inactive VLP pellets on silver stained SDS gels. In the protease defective mutant only the 74kDa Gag precursor was visible, but in the protease active *gag-pro-pol* construct several smaller proteins were also visible. These smaller proteins were later

specifically detected on Western Blot with antibodies raised against HERV-K113 Gag subdomains.

5.2 Identification of cleavage sites and subdomains

Although the cleavage sites in the Gag proteins of many retroviruses have been identified in detail, it is still unclear, even for HIV, what the exact mechanism of cleavage site selection is. The retroviral PR acts fairly specific but it has not been possible to give a consensus substrate sequence (Bagossi et al., 2005). Selection depends on a mix of structural features influenced by the amino acid sequence of the site as well as of the PR. Therefore it was not possible to identify the HERV-K113 Gag cleavage site in silico. Comparison of the Gag sequence to the related beta retroviruses MPMV and MMTV revealed some subdomain specific elements. In the Matrix domain HERV-K113 shares a conserved region with MMTV and MPMV, which has been characterized in MPMV as a Cytoplasmic/retention signal spanning 18 amino acids. Introduction of these 18 amino acids into a similar region in Mo-MLV transferred the ability to assemble immature capsid in the cytoplasm (Choi et al., 1999). This region in MPMV includes at position 55 an Arg which determines in MPMV the Gag assembly site in the cytoplasm. Changing the Arg to a Trp led to the assembly of Gag proteins into immature particles at the plasma membrane of cells (Rhee et al., 1990). Similar to MPMV HERV-K113 encodes a sequence, which correlates to the consensus sequence PPxY of a late domain in the region between CA and MA. But other than in MPMV the main late domain has been shown as PTAP in HERV-K1113. The PTAP motive also lies in the domain between MA and CA. Mutation of this sequence led to typical budding defects (Chudak Diplomarbeit). The CA domain of HERV-K113 has the conserved polar residues Gln, Glu, Phe and Arg in the Major homology region with MPMV and MMTV in common. These four amino acids are highly conserved in a number of retroviruses and mutation of theses amino acids, e.g. through substitution with Ala, impaired viral replication and affected viral particle production (Mammano et al., 1994). The retroviral (CCHC) Zinc fingers are conserved in the nucleocapsid proteins and are necessary for efficient packaging of the viral RNA genome during assembly. Hence, mutating the Cis or His amino acids reduces the packaging of the viral RNA (Aldovini and Young, 1990; Gorelick et al., 1999; Kurth and Bannert, 2010).

The positions of the subdomains and the sequence comparison to the other beta retroviruses indicated where possibly the protease might cleave Gag. This led to the development of three antibodies against presumably the Matrix, Capsid and phosphoprotein (p15) domain. Viral subdomains of mature VLPs were purified by reverse phase HPLC and MA, CA and variant form of the p15 protein identified with the antibodies previously developed against them. For the separation of the subdomains the RP-HPLC protocol had to be adjusted to the specific subdomain. MA eluted very late in a 30 min RP-HPLC run, but by extending the elution gradient the separation of proteins was refined and MA eluted in distinct peaks after CA. Additional with MA eluted the MA-SP1-p15 which indicated that MA, SP1 as well as p15 are cleaved very late during the viral maturation process. The phosphoprotein p15 as well as the p15-SP1 intermediate proteins eluted in 2 distinct peaks in a 30 min RP-HPLC run. Cleavage at the N-terminus of SP1 was not as precise and we suggest a cryptic cleavage site three amino acids downstream of the first SP1 cleavage site. Similar cryptic cleavage sites have been discovered in HIV SP1 (Kräusslich et al., 1995). The elution of CA always resulted in the highest peak in the RP-HPLC chromatogram, probably because CA is the biggest subdomain in Gag. The analysis of these fractions on Western Blot showed no further proteins eluting at this time point. NC eluted at the beginning of the RP-HPLC run and was separated best from the other subdomains in a 60 min RP-HPLC run. Overall all subdomains eluted in order according to their size except of MA. This may be due to the general basic isoelectric point (pI) of 9.26. But since NC has a similar pI of 9.79 and is only 5 kDa smaller it is most probable that the conformation of MA leads to a stronger retention to the column.

N-terminal sequencing of purified MA, p15 and p15 intermediates, as well as of the purified NC then allowed the exact determination of the cleavage sites and confirmed the subdomains. Interestingly mass spectrometry analysis of the detected NC domain revealed a shorter amino acid sequence suggesting an additional C-terminal domain similar to the P4 protein in MPMV. To investigate this we introduced a cleavage block C-terminal of NC which revealed another cleavage site leading to two subdomains at the C-terminus of Gag. These two proteins, called QP1 and QP2, are glutamine and proline rich and consist of 23 and 19 amino acids respectively. Approximately half of the amino acids are either proline or glutamine although they do not come in stretches of more than three amino acids of the same type. The role of QP1 and QP2 remains unknown. QP2 harbors a PPPQ sequence similar to the canonical PPPY sequence of an L-domain which can interact directly with Nedd4-like ubiquitin ligases. But the introduction of a stop codon following the codon QP2 did not lead to a typical L-

domain defect (Chudak et al., 2013). It is rather unusual to have two proteins following NC, other retroviruses encode at the most one other protein at the C-terminus of Gag. The high number of proline restricts the structure of both proteins. Because the side-chain of Proline is cyclized back on the backbone amid position, the backbone confirmation itself is very limited. Also the bulkiness of the N-CH_2 group restricts the conformation of the residue preceding proline and proline cannot act as a hydrogen bond donor, because the amid proton is replaced by the CH_2 group (Williamson, 1994). Since the order of the subdomains in the Gag sequence determines also the arrangement inside the VLP it can be suggested that QP1 and QP2 stay close to the viral RNA sequence similar to NC. But NC also is a basic protein with an pI of 9.79 which suggests a close interaction with nucleoacids. The role of the QP proteins instead with a pI close to 6 is unpredictable. Maybe these two proteins had a function in the original host of the retrovirus before it accidentally integrated into the human genome. Blast analysis revealed that, other than HERVs, only parts of the Gag protein of Macaca fascicularis resemble the amino acid sequence of QP1 and 2. It was only possible to reveal the QP1 and QP2 proteins by using RP-HPLC and MALDI-TOF. Applying 2D gel electrophoresis to separate the mature viral proteins was not sensitive enough to detect these proteins (Kraus et al., 2011).

Additionally to the non-canonical QP-rich proteins there were two major forms of an phosphoprotein detected between MA and CA, with a size of 15 and 16.5 kDa. MALDI-TOF analysis and N-terminal sequencing confirmed that the two forms differ only in the N-terminus by a 14nt long peptide, which we named SP1. A similar spacer peptide, the p2, is found in RSV and regulates the maturation process (Cameron et al., 1992).

5.3 Mutational analysis of cleavage sites

Although the mechanism of site selection by retroviral PR remains poorly understood, it has been shown by Petit et al. that about 80% of all known PR cleavage sites can be classified into one of two categories when looking at 7 residues neighboring the cleavage site. These neighboring amino acids are numbered P4 to P1 upstream of the scissile bond and P1’ to P3’ downstream of the scissile bond. The type 1 cleavage sites are defined by an aromatic amino acid in P1 and a Pro at P1’. Type 2 cleavage sites on the other hand encode a hydrophobic residue, except Ile and Val, at the P1 position and favor a Val, Leu or Ala at P1’ (Eizert et al., 2008; Pettit et al., 2002, 1991). The cleavage site between MA and SP1 fit the type 2 cleavage

site category due to the aromatic Trp at the P1 position and the Val at P1'. Exchanging the Trp for an Ala led to an accumulation of a larger MA protein matching the size of MA-SP1 and an increase of MA-Sp1-p15 intermediate in Western Blot analysis. Electron microscopic pictures showed mature and immature particle in the cell culture in congruence with the fact, that no changes in the cleavage pattern of Capsid were visible on Western Blot and CA forms the virus core, which denotes the mature VLP. A similar effect had been shown before for MLV. Mutating the P2 and P1 sites upstream of the MA-p12 cleavage site of MLV did not lead to changes in VLP maturation when investigated in electron microscopy, although the MA cleavage from p12 was abolished (Oshima et al., 2004). Thus, similar to MLV, cleavage between MA and p15 is redundant for morphological maturation. RT activity for the MA-SP1 oricoHERV-K113 mutant increased slightly. A much conserved type 2 cleavage site in other retroviruses as well as in HERV-K113 is the site between CA and NC. Interestingly exchanging a Gly with an Asp at the P1 position of CA-NC did not lead not to a complete abrogation of cleavage at this site nor at the other sites. Although there were accumulations of CA-NC fusion proteins there was still a strong signal of CA detected on Western Blot. This indicates that an Arg at the P1 position can be tolerated by the HERV-K113 protease. Although there is fully cleaved CA visible on Western blot no mature VLPs were detected in Electron microscopic pictures. This fits well with the fact that RT activity measured by Cavidi assay reduced to approximately a fifth of the intensity measured in oricoHERV-K113_GagProPol constructs. It might be concluded from this, that the CA-NC fusion proteins inhibit the correct formation of the Capsid shell leading also to an inhibition of RT activity. The last type 2 cleavage site in oricoHERV-K113_Gag is situated between QP1 and QP2. Mutation of this site did not lead to significant changes in the RT activity or on Western Blots (data not shown). It had been investigated before that the introduction of an stop codon at the start of QP1 did not lead to any changes in the maturation of the VLPs (Chudak et al., 2013).

In congruence with all retroviruses the cleavage site at the N-terminus of CA is a type 1 cleavage site. Here the exchange of the Phe with an Asp inhibited the cleavage at other sites in Gag and led to the most severe phenotype of all mutants presented. On Western Blot a strong band corresponding to a fusion protein spanning from MA to at least CA was detected with α-MA, α-P15 and α-CA. There was no RT activity measured in supernatants of cells transfected with this mutant. The mutation led to a formation of aberrant particle structures. On electron microscopic pictures there were mostly tubular but sometimes also round shaped particles in cells visible. Also Gag accumulated in thick sheets under the cell membrane. This phenotype

has been observed similarly in MMTV when the n domain, which precedes the CA domain, is deleted (Zábranský et al., 2010). For RSV a similar effect was shown when a single amino acid in the C-terminus of p10, the domain N-terminal of CA, was exchanged with an Ala (Scheifele et al., 2007). The assembly into tubular structures has been shown before with purified intact RSV CA. It assembles in vitro into tubular structures and this effect can be reversed by the addition of the last 25 amino acids of p10 (Campbell and Vogt, 1997; Kingston et al., 2000). In conclusion it can be suggested, that similar to MMTV and RSV the aminoacids in p15 preceding CA are critical for correct core formation as well as overall virion structure. The other type 2 cleavage site is found between NC and QP1. Similar to the cleavage site between the QP domains mutating this cleavage site did not lead to specific phenotype differences.

The cleavage site between SP1 and p15 does not fulfill the requirements for a type 1 or type 2 cleavage site. Neither the hydrophilic Asp residue in P1 nor the Tyr in P1' fit the classic schema of a retroviral cleavage site. Mutating the Asp into an Ala did not lead to significant changes in the maturation process as visible in electron microscopic pictures and the measurement of the RT activity. In congruence with these results was the fact, that the HERV-K113 PR could process the type 2 CA-NC cleavage site even when Gly at the P1 site was exchanged with an Asp. On the other hand lead the introduction of an Asp at the type 1 cleavage site between p15 and CA to an abrogation of the maturation process. We conclude that an Asp at the P1 position can be tolerated by the viral Protease but there seem to be other factors influencing the cleavage effectiveness. It is also possible, that cleavage at this site is not mediated by the viral PR but through a cellular peptidase as a contaminant or as part of the virion. Cellular peptidases bound to the particle have been shown before for MLV and RSV (Pepinsky et al., 1996; Yoshinaka et al., 1980).

5.4 Phosphoprotein p15

Phosphoproteins are encoded in many alpha-, beta- and gammaretroviruses between MA and CA. Their sizes vary between 10 to 20 kDa and several contain one or more late assembly domains, but the overall functional role of these proteins has not been investigated in detail. ESI analysis on tryptic digested p15 proteins revealed five phosphorylated Ser and three phosphorylated Thr, including the Thr in the late domain sequence PTAP (Chudak 2009 diploma thesis). Exchange of the Thr in the PTAP sequence with an Ala led to severe budding

defects. It would be interesting to exchange the identified phosphorylated amino acids to elucidate their function in the HERV-K113 replication process. To investigate further the potential role of p15, we exchanged a sequence of 82 amino acids of p15 starting five amino acids downstream from the N-terminus of p15 and ending 14 amino acids upstream of PTAP. This sequence was exchanged with the sequence of mCherry (Williams, Diploma thesis). The exchange did not lead to budding defects. On electron microscopic pictures were immature and mature VLPs detected. The only defect was that on Western Blot no free MA was detected, which indicated that the MA-SP1-p15 cleavage sites were not processed. Further experiments need to be conducted to elucidate the function of this phosphoprotein.

5.5 Conclusion

The reconstituted Gag-Pro-Pol sequence enabled the identification of all Gag subdomains of HERV-K113. The separation of retroviral Gag subdomains through RP-HPLC and the subsequent analysis through MALDI was especially advantageous to reveal the presence of the small C-terminally encoded protein QP1 and QP2. Mutational analysis of the P1 site of the discovered cleavage sites showed similar effects as in analogous experiments in MPMV, MMTV or RSV and further proofed the results. Similar to other retroviruses (Oshima et al., 2004; Scheifele et al., 2007; Zábranský et al., 2010) the type 1 cleavage site between p15-NC is crucial for maturation. Mutations introduced at this site lead to the abrogation of the maturation process. Rather unusual is the cleavage site between SP1 and p15 and further experiments are necessary, also to reveal the potential role of SP1 and of the cryptic cleavage site inside of SP1. It has been shown for HIV-1 that deletion of the SP1 sequence abrogates particle production in vivo (Kräusslich et al., 1995) and similar effects have been shown in RSV as well (Keller et al., 2008).

In future experiments it would be of special interest to investigate the role of the QP proteins. The position at the C-terminus of Gag indicated that these proteins are like NC inside the capsid. Structural analysis of QP1 and QP2 might reveal if these proteins interact with the RNA or NC or both. Expression of QP proteins in cells would show if they have e.g. a cytopathic effect on cells. HERV have been implicated in a number of autoimmune diseases and it would be of interest if these proteins may play a role in disease development or severity.

6. References

Aldovini, a, Young, R. a, 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. J. Virol. 64, 1920–6.

Allain, B., Lapadat-Tapolsky, M., Berlioz, C., Darlix, J.L., 1994. Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome. EMBO J. 13, 973–81.

Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R., Summers, M.F., 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. J. Mol. Biol. 301, 491–511.

Bagossi, P., Sperka, T., Fehér, A., 2005. Amino acid preferences for a critical substrate binding subsite of retroviral proteases in type 1 cleavage sites. J. … 79, 4213–4218.

Baltimore, D., 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226, 1209–11.

Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J.-L., Ohlmann, T., 2007. Translational control of retroviruses. Nat. Rev. Microbiol. 5, 128–40.

Bannert, N., Kurth, R., 2004. Retroelements and the human genome: new perspectives on an old relation. Proc. Natl. Acad. Sci. U. S. A. 101 Suppl , 14572–9.

Bannert, N., Kurth, R., 2006. The evolutionary dynamics of human endogenous retroviral families. Annu. Rev. Genomics Hum. Genet. 7, 149–73.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868–71.

Beyer, U., Moll-rocek, J., Moll, U.M., Dobbelstein, M., 2011. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes.

Bieda, K., Hoffmann, a, Boller, K., 2001. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. J. Gen. Virol. 82, 591–6.

Boller, K., Schönfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R., Tönjes, R.R., 2008. Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. J. Gen. Virol. 89, 567–72.

Bowzard, J.B., Bennett, R.P., Krishna, N.K., Ernst, S.M., Rein, A., Wills, J.W., 1998. Importance of basic residues in the nucleocapsid sequence for retrovirus Gag assembly and complementation rescue. J. Virol. 72, 9034–44.

Bradac, J., Hunter, E., 1984. Polypeptides of Mason-Pfizer monkey virus. I. Synthesis and processing of the gag-gene products. Virology 138, 260–75.

Brandt, J., Schrauth, S., Veith, A.-M., Froschauer, A., Haneke, T., Schultheis, C., Gessler, M., Leimeister, C., Volff, J.-N., 2005. Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. Gene 345, 101–11.

Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M., 1987. Correct integration of retroviral DNA in vitro. Cell 49, 347–356.

Bukrinsky, M.I., Sharova, N., Dempsey, M.P., Stanwick, T.L., Bukrinskaya, A.G., Haggerty, S., Stevenson, M., 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc. Natl. Acad. Sci. 89, 6580–6584.

Büscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., Denner, J., 2005. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. Cancer Res. 65, 4172–80.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., Hoffmann, C., 2005. Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. 3, 848–58.

Callahan, R., Drohan, W., Tronick, S., Schlom, J., 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. U. S. A. 79, 5503–7.

Cameron, C.E., Grinde, B., Jacques, P., Jentoft, J., Leis, J., Wlodawer, a, Weber, I.T., 1993. Comparison of the substrate-binding pockets of the Rous sarcoma virus and human immunodeficiency virus type 1 proteases. J. Biol. Chem. 268, 11711–20.

Cameron, C.E., Grinde, B., Jentoft, J., Leis, J., Weber, I.T., Copeland, T.D., Wlodawer, a, 1992. Mechanism of inhibition of the retroviral protease by a Rous sarcoma virus peptide substrate representing the cleavage site between the gag p2 and p10 proteins. J. Biol. Chem. 267, 23735–41.

Campbell, S., Vogt, V.M., 1997. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. J. Virol. 71, 4425–35.

Cheynet, R.I.E., Bouton, O., Blond, J., Lavillette, D., 2000. An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus Receptor 74, 3321–3329.

Choi, G., Park, S., Choi, B., Hong, S., Lee, J., Hunter, E., Rhee, S.S., 1999. Identification of a cytoplasmic targeting/retention signal in a retroviral Gag polyprotein. J. Virol. 73, 5431–7.

Chudak, C., Beimforde, N., George, M., Zimmermann, A., Lausch, V., Hanke, K., Bannert, N., 2013. Identification of late assembly domains of the human endogenous retrovirus-K(HML-2). Retrovirology 10, 140.

Cimarelli, A., Sandin, S., Höglund, S., Luban, J., 2000. Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA. J. Virol. 74, 3046–57.

D'Souza, V., Summers, M.F., 2005. How retroviruses select their genomes. Nat. Rev. Microbiol. 3, 643–55.

Darlix, J., Lapadat-tapolsky, M., Rocquigny, H. De, Roques, B.P., Normale, E., 1995. First Glimpses at Structure-function Relationships of the Nucleocapsid Protein of Retroviruses 523–537.

De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., Summers, M.F., 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science 279, 384–8.

De Marco, A., Davey, N.E., Ulbrich, P., Phillips, J.M., Lux, V., Riches, J.D., Fuzik, T., Ruml, T., Kräusslich, H.-G., Vogt, V.M., Briggs, J. a G., 2010. Conserved and variable features of Gag structure and arrangement in immature retrovirus particles. J. Virol. 84, 11729–36.

Delchambre, M., Gheysen, D., Thines, D., Thiriart, C., Jacobs, E., Verdin, E., Horth, M., Burny, A., Bex, F., 1989. The GAG precursor of simian immunodeficiency virus assembles into virus-like particles. EMBO J. 8, 2653–60.

Demirov, D.G., Freed, E.O., 2004. Retrovirus budding. Virus Res. 106, 87–102.

Depil, S., Roche, C., Dussart, P., Prin, L., 2002. Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. Leukemia 16, 254–9.

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

Doxiadis, G.G.M., de Groot, N., Bontrop, R.E., 2008. Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. J. Virol. 82, 6667–77.

Eizert, H., Bander, P., Bagossi, P., Sperka, T., Miklóssy, G., Boross, P., Weber, I.T., Tözsér, J., 2008. Amino acid preferences of retroviral proteases for amino-terminal positions in a type 1 cleavage site. J. Virol. 82, 10111–7.

Ejthadi, H.D., Martin, J.H., Junying, J., Roden, D.A., Lahiri, M., Warren, P., Murray, P.G., Nelson, P.N., 2005. A novel multiplex RT-PCR system detects human endogenous retrovirus-K in breast cancer. Arch. Virol. 150, 177–84.

Ellermann V., B.O., 1908. Experimentelle Leukämie bei Hühnern. Zentralbl. Bakteriol. Parasitenkd. Infect. 46, 595–609.

Fields Virology, 2006. . Lippincott Williams & Wilkins.

Finzi, A., Orthwein, A., Mercier, J., Cohen, E.A., 2007. Productive human immunodeficiency virus type 1 assembly takes place at the plasma membrane. J. Virol. 81, 7476–90.

Gamble, T.R., 1997. Structure of the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein. Science (80-.). 278, 849–853.

Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I., 2008. The structural biology of HIV assembly. Curr. Opin. Struct. Biol. 18, 203–17.

George, M., Schwecke, T., Beimforde, N., Hohn, O., Chudak, C., Zimmermann, A., Kurth, R., Naumann, D., Bannert, N., 2011. Identification of the protease cleavage sites in a reconstituted Gag polyprotein of an HERV-K (HML-2) element. Retrovirology 8, 30.

Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., De Wilde, M., 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell 59, 103–12.

Gitti, R.K., Lee, B.M., Walker, J., Summers, M.F., Yoo, S., Sundquist, W.I., 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science 273, 231–5.

Goff, S., 2001. Retroviridae: The Retroviruses and Their Replication, in: D. Knipe et Al., Eds. Fields Virology. Lippincott Williams and Wilkins. pp. 1999–2069.

Goff, S.P., 2007. Host factors exploited by retroviruses. Nat. Rev. Microbiol. 5, 253–63.

Gorelick, R.J., Gagliardi, T.D., Bosche, W.J., Wiltrout, T. a, Coren, L. V, Chabot, D.J., Lifson, J.D., Henderson, L.E., Arthur, L.O., 1999. Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. Virology 256, 92–104.

Hanke, K., Chudak, C., Kurth, R., Bannert, N., 2013a. The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). Int. J. Cancer 132, 556–67.

Hanke, K., Hohn, O., Liedgens, L., Fiddeke, K., Wamara, J., Kurth, R., Bannert, N., 2013b. Staufen-1 Interacts with the HERV-K(HML-2) Rec and Gag Proteins and Increases Virion Production. J. Virol.

Hanke, K., Kramer, P., Seeher, S., Beimforde, N., Kurth, R., Bannert, N., 2009. Reconstitution of the ancestral glycoprotein of human endogenous retrovirus k and modulation of its functional activity by truncation of the cytoplasmic domain. J. Virol. 83, 12790–800.

Hemonnot, B., Cartier, C., Gay, B., Rebuffat, S., Bardy, M., Devaux, C., Boyer, V., Briant, L., 2004. The host cell MAP kinase ERK-2 regulates viral assembly and release by phosphorylating the p6gag protein of HIV-1. J. Biol. Chem. 279, 32426–34.

Henderson, L.E., Krutzsch, H.C., Oroszlan, S., 1983. Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification. Proc. Natl. Acad. Sci. U. S. A. 80, 339–43.

Henderson, L.E., Sowder, R., Copeland, T.D., Smythers, G., Oroszlan, S., 1984. Quantitative separation of murine leukemia virus proteins by reversed-phase high-pressure liquid chromatography reveals newly described gag and env cleavage products. J. Virol. 52, 492–500.

Hernandez, L.D., Hoffman, L.R., Wolfsberg, T.G., White, J.M., 1996. Virus-cell and cell-cell fusion. Annu. Rev. Cell Dev. Biol. 12, 627–61.

Hizi, a, Henderson, L.E., Copeland, T.D., Sowder, R.C., Krutzsch, H.C., Oroszlan, S., 1989. Analysis of gag proteins from mouse mammary tumor virus. J. Virol. 63, 2543–9.

Huebner, R.J., Todaro, G.J., 1969. ONCOGENES OF RNA TUMOR VIRUSES AS DETERMINANTS OF CANCER. Proc. Natl. Acad. Sci. 64, 1087–1094.

Ishida, T., Obata, Y., Ohara, N., Matsushita, H., Sato, S., Uenaka, A., Saika, T., Miyamura, T., Chayama, K., Nakamura, Y., Wada, H., Yamashita, T., Morishima, T., Old, L.J., Nakayama, E., 2008. Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity. Cancer Immun. 8, 15.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., Varmus, H.E., 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature 331, 280–3.

Jern, P., Coffin, J.M., 2008. Effects of retroviruses on host genome function. Annu. Rev. Genet. 42, 709–32.

Jha, A.R., Pillai, S.K., York, V. a, Sharp, E.R., Storm, E.C., Wachter, D.J., Martin, J.N., Deeks, S.G., Rosenberg, M.G., Nixon, D.F., Garrison, K.E., 2009. 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. 26, 2617–26.

Jouvenet, N., Neil, S.J.D., Bess, C., Johnson, M.C., Virgen, C.A., Simon, S.M., Bieniasz, P.D., 2006. Plasma membrane is the site of productive HIV-1 particle assembly. PLoS Biol. 4, e435.

Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6, e1001191.

Kaufmann, S., Sauter, M., Schmitt, M., Baumert, B., Best, B., Boese, A., Roemer, K., Mueller-Lantzsch, N., 2010. Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. J. Gen. Virol. 91, 1494–502.

Keller, P.W., Johnson, M.C., Vogt, V.M., 2008. Mutations in the spacer peptide and adjoining sequences in Rous sarcoma virus Gag lead to tubular budding. J. Virol. 82, 6788–97.

Kettmann, R., Deschamps, J., Cleuter, Y., Couez, D., Burny, A., Marbaix, G., 1982. Leukemogenesis by bovine leukemia virus: proviral DNA integration and lack of RNA expression of viral long terminal repeat and 3' proximate cellular sequences. Proc. Natl. Acad. Sci. 79, 2465–2469.

Kingston, R.L., Fitzon-Ostendorp, T., Eisenmesser, E.Z., Schatz, G.W., Vogt, V.M., Post, C.B., Rossmann, M.G., 2000. Structure and self-association of the Rous sarcoma virus capsid protein. Structure 8, 617–28.

Knipe, D.M., Howley, P., 2013. Fields Virology (Knipe, Fields Virology)-2 Volume Set. Lippincott Williams & Wilkins.

Kozak, C.A., Gromet, N.J., Ikeda, H., Buckler, C.E., 1984. associated with the Fv4 resistance gene 81, 834–837.

Kraus, B., Boller, K., Reuter, A., Schnierle, B.S., 2011. Characterization of the human endogenous retrovirus K Gag protein: identification of protease cleavage sites. Retrovirology 8, 21.

Kräusslich, H.G., Fäcke, M., Heuser, A.M., Konvalinka, J., Fa, M., Konvalinka, J.A.N., Zentgraf, H., 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity . The Spacer Peptide between Human Immunodeficiency Virus Capsid and Nucleocapsid Proteins Is Essential for Or.

Kuhelj, R., Rizzo, C.J., Chang, C.H., Jadhav, P.K., Towler, E.M., Korant, B.D., 2001. Inhibition of human endogenous retrovirus-K10 protease in cell-free and cell-based assays. J. Biol. Chem. 276, 16674–82.

Kurth, R., Bannert, N., 2010. Retroviruses: Molecular Biology, Genomics and Pathogenesis. Caister Academic Press.

Lander, E., Linton, L., Birren, B., Nusbaum, C., 2001. Initial sequencing and analysis of the human genome. Nature 409.

Landry, J.-R., Rouhi, A., Medstrand, P., Mager, D.L., 2002. The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. Mol. Biol. Evol. 19, 1934–42.

Lavie, L., Medstrand, P., Schempp, W., Meese, E., Mayer, J., 2004. Human Endogenous Retrovirus Family HERV-K (HML-5): Status , Evolution , and Reconstruction of an Ancient Betaretrovirus in the Human Genome † 78, 8788–8798.

Le Blanc, I., Rosenberg, A.R., Dokhélar, M.C., 1999. Multiple functions for the basic amino acids of the human T-cell leukemia virus type 1 matrix protein in viral transmission. J. Virol. 73, 1860–7.

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

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

Leis, J., Baltimore, D., Bishop, J.M., Coffin, J., Fleissner, E., Goff, S.S.P., Oroszlan, S., 1988. Standardized and Simplified Nomenclature for Proteins Common to 62, 1808–1809.

Li, X., Quan, Y., Arts, E.J., Li, Z., Preston, B.D., de Rocquigny, H., Roques, B.P., Darlix, J.L., Kleiman, L., Parniak, M.A., Wainberg, M.A., 1996. Human immunodeficiency virus Type 1 nucleocapsid protein (NCp7) directs specific initiation of minus-strand DNA synthesis primed by human tRNA(Lys3) in vitro: studies of viral RNA molecules mutated in regions that flank the primer binding site. J. Virol. 70, 4996–5004.

Löwer, R., Löwer, J., Frank, H., Harzmann, R., Kurth, R., 1984. Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses. J. Gen. Virol. 65 (Pt 5), 887–98.

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

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., Mandrand, B., 2004. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. U. S. A. 101, 1731–6.

Mammano, F., Ohagen, a, Höglund, S., Göttlinger, H.G., 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. J. Virol. 68, 4927–36.

Manrique, M.L., Celma, C.C., González, S.A., Affranchino, J.L., 2001. Mutational analysis of the feline immunodeficiency virus matrix protein. Virus Res. 76, 103–13.

Marsh, M., Helenius, A., 2006. Virus entry: open sesame. Cell 124, 729–40.

Medizin, B.K., Fakult, M., Saarlandes, D.E.S., Klatt, J., Universit, V., 2008. Verbreitung der humanen endogenen Retroviren HERV-K113 und HERV-K115 in der menschlichen Population.

Medstrand, P., Landry, J.R., Mager, D.L., 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J. Biol. Chem. 276, 1896–903.

Menéndez-Arias, L., Risco, C., Pinto da Silva, P., Oroszlan, S., 1992. Purification of immature cores of mouse mammary tumor virus and immunolocalization of protein domains. J. Virol. 66, 5615–20.

Meng, B., Lever, A.M., 2013. Wrapping up the bad news: HIV assembly and release. Retrovirology 10, 5.

Méric, C., Goff, S.P., 1989. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. J. Virol. 63, 1558–68.

Moore, R., Dixon, M., Smith, R., Peters, G., Dickson, C., 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of gag and pol. J. Virol. 61, 480–490.

Moyes, D.L., Martin, A., Sawcer, S., Temperton, N., Worthington, J., Griffiths, D.J., Venables, P.J., 2005. The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. Genomics 86, 337–41.

Müller, B., Patschinsky, T., Kräusslich, H.-G., 2002. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. J. Virol. 76, 1015–24.

Muriaux, D., Darlix, J.-L., 2010. Properties and functions of the nucleocapsid protein in virus assembly. RNA Biol. 7, 744–753.

Muriaux, D., Mirro, J., Harvin, D., Rein, A., 2001. RNA is a structural element in retrovirus particles. Proc. Natl. Acad. Sci. U. S. A. 98, 5246–51.

Murray, P.S., Li, Z., Wang, J., Tang, C.L., Honig, B., Murray, D., 2005. Retroviral Matrix Domains Share Electrostatic Homology : Models for Membrane Binding Function throughout the Viral Life Cycle Weill Medical College of Cornell 13, 1521–1531.

Narayan, S., Young, J.A.T., 2004. Reconstitution of retroviral fusion and uncoating in a cell-free system. Proc. Natl. Acad. Sci. U. S. A. 101, 7721–6.

Ono, M., Kawakami, M., Ushikubo, H., 1987. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell Stimulation of Expression of the Human Endogenous Retrovirus Genome by Female Steroid Hormones in Human Breast Cancer Cell Line T47D.

Ono, M., Yasunaga, T., Miyata, T., Ushikubo, H., 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. 60, 589–98.

Oshima, M., Muriaux, D., Mirro, J., Nagashima, K., Dryden, K., Yeager, M., Rein, A., 2004. Effects of Blocking Individual Maturation Cleavages in Murine Leukemia Virus Gag 78, 1411–1420.

Pal, R., Reitz, M.S., Tschachler, E., Gallo, R.C., Sarngadharan, M.G., Veronese, F.D., 1990. Myristoylation of gag proteins of HIV-1 plays an important role in virus assembly. AIDS Res. Hum. Retroviruses 6, 721–30.

Palmarini, M., Mura, M., Spencer, T.E., 2004. Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. J. Gen. Virol. 85, 1–13.

Pepinsky, R.B., Papayannopoulos, I.A., Campbell, S., 1996. Analysis of Rous Sarcoma Virus Gag Proteins by Mass Spectrometry Indicates Trimming by Host Exopeptidase 70, 3313–3318.

Pettit, S.C., Henderson, G.J., Schiffer, C.A., Swanstrom, R., 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. J. Virol. 76, 10226–33.

Pettit, S.C., Simsic, J., Loeb, D.D., Everitt, L., Hutchison, C. a, Swanstrom, R., 1991. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J. Biol. Chem. 266, 14539–47.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., Gallo, R.C., 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. 77, 7415–9.

Purdy, J.G., Flanagan, J.M., Ropson, I.J., Rennoll-Bankert, K.E., Craven, R.C., 2008. Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly. J. Virol. 82, 5951–61.

Rhee, S.S., Hui, H., Hunter, E., 1990. Preassembled Capsids of Type D Retroviruses Contain a Signal Sufficient for Targeting Specifically to the Plasma Membrane 64, 3844–3852.

Robbins, J.M., Gallahan, D., Hogg, E., Kozak, C., Callahan, R., 1986. An endogenous mouse mammary tumor virus genome common in inbred mouse strains is located on chromosome 6. J. Virol. 57, 709–713.

Rodríguez-Rodríguez, L., Tsuchihashi, Z., Fuentes, G.M., Bambara, R.A., Fay, P.J., 1995. Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. J. Biol. Chem. 270, 15005–11.

Roe, T., Reynolds, T.C., Yu, G., Brown, P.O., 1993. Integration of murine leukemia virus DNA depends on mitosis. EMBO J. 12, 2099–108.

Schechter, I., Berger, A., 2012. On the size of the active site in proteases. Biochem. Biophys. Res. Commun. 425, 497–502.

Scheifele, L.Z., Kenney, S.P., Cairns, T.M., Craven, R.C., Parent, L.J., 2007. Overlapping roles of the Rous sarcoma virus Gag p10 domain in nuclear export and virion core morphology. J. Virol. 81, 10718–28.

Schultz, A.M., Oroszlant, S., 1983. In vivo modification of retroviral gag gene-encoded polyproteins by myristic acid . In Vivo Modification of Retroviral gag Gene-Encoded Polyproteins by Myristic Acid 46.

Schultz, A.M., Rein, A., 1989. Unmyristylated Moloney murine leukemia virus Pr65gag is excluded from virus assembly and maturation events. J. Virol. 63, 2370–2373.

Skehel, J.J., Wiley, D.C., 1998. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell 95, 871–4.

Sommerfelt, M. a, 1999. Retrovirus receptors. J. Gen. Virol. 80 (Pt 12, 3049–64.

Subramanian, R.P., Wildschutte, J.H., Russo, C., Coffin, J.M., 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology 8, 90.

Tarlinton, R.E., Meers, J., Young, P.R., 2006. Retroviral invasion of the koala genome. Nature 442, 79–81.

Temin, H.M., Mizutani, S., 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226, 1211–3.

Tristem, M., 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J. Virol. 74, 3715–30.

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

Varki, A., Altheide, T.K., 2009. Comparing the human and chimpanzee genomes : Searching for needles in a haystack Comparing the human and chimpanzee genomes : Searching for needles in a haystack 1746–1758.

Vogt, P., 1997. Historical Introduction to the General Properties of Retroviruses.

Wang-Johanning, F., Frost, A.R., Jian, B., Epp, L., Lu, D.W., Johanning, G.L., 2003. Quantitation of HERV-K env gene expression and splicing in human breast cancer. Oncogene 22, 1528–35.

Wang-Johanning, F., Liu, J., Rycaj, K., Huang, M., Tsai, K., Rosen, D.G., Chen, D.-T., Lu, D.W., Barnhart, K.F., Johanning, G.L., 2007. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int. J. Cancer 120, 81–90.

Weiss, R. a, 2006. The discovery of endogenous retroviruses. Retrovirology 3, 67.

Williamson, M.P., 1994. The structure and function of proline-rich regions in proteins. Biochem. J. 297 (Pt 2, 249–60.

Wills, J.W., Craven, R.C., 1991. Form,_function,_and_use_of_retroviral_Gag_proteins.2.pdf.

Willsand, W., Craven, R.C., 1991. Form, function, and use of retroviral Gag profeins.

Yasuda, J., Hunter, E., 1998. A proline-rich motif (PPPY) in the Gag polyprotein of Mason-Pfizer monkey virus plays a maturation-independent role in virion release. J. Virol. 72, 4095–103.

Yoshinaka, Y., Kirk, C., Luftig, R.B., 1980. Aminopeptidase activity associated with purified murine leukaemia viruses. J. Gen. Virol. 46, 363–72.

Yueh, A., Goff, S.P., 2003. Phosphorylated Serine Residues and an Arginine-Rich Domain of the Moloney Murine Leukemia Virus p12 Protein Are Required for Early Events of Viral Infection 77, 1820–1829.

Zábranský, A., Hoboth, P., Hadravová, R., Stokrová, J., Sakalian, M., Pichová, I., 2010. The noncanonical Gag domains p8 and n are critical for assembly and release of mouse mammary tumor virus. J. Virol. 84, 11555–9.

Zhou, M., Deng, L., Kashanchi, F., Brady, J.N., Shatkin, A.J., Kumar, A., 2003. The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. Proc. Natl. Acad. Sci. U. S. A. 100, 12666–71.