In vitro assembly of virus-like particles of a Gammaretrovirus, the Murine Leukemia Virus (XMRV)

Running title: MLV CANC in vitro assembly

Romana Hadravová1#, Alex de Marco2#, Pavel Ulbrich3, Jitka Štokrová1, Michal Doležal1,3, Iva Pichová1, Tomáš Rumí3, John A. G. Briggs2* and Michaela Rumlová1*

1 Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., IOCB & Gilead Research Center, Flemingovo nám. 2, 166 10 Prague, Czech Republic
2 Structural and Computational Unit, European Molecular Biology Laboratory, Heidelberg, Germany
3 Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague, Technická 5, 166 28, Prague, Czech Republic

# These authors contributed equally
*Corresponding authors

mailing addresses:
1 IOCB, Flemingovo nám. 2, 166 10, Prague, Czech Republic
e-mail: rumlová@uochb.cas.cz
phone number: +420220183538
fax number: +420220183556

Structural and Computational Unit, EMBL, Meyerhofstr. 1 69117 Heidelberg Germany

e-mail: briggs@embl.de

phone number: +49 6221 387 8482

Abstract word count: 178

Text word count: 4643
Abstract

Immature retroviral particles are assembled by self-association of the structural polyprotein precursor Gag. During maturation the Gag polyprotein is proteolytically cleaved yielding mature structural proteins: matrix (MA), capsid (CA) and nucleocapsid (NC) that re-assemble into a mature viral particle. Proteolytic cleavage causes the N-terminus of CA to fold back to form a β-hairpin, anchored by an internal salt bridge between the N-terminal proline and inner aspartate. Using an in vitro assembly system of capsid-nucleocapsid protein (CANC), we studied formation of virus-like particles (VLP) of a gammaretrovirus, the xenotropic murine leukemia virus (MLV)-related virus (XMRV). We show here that unlike for other retroviruses, XMRV CA and CANC do not assemble tubular particles characteristic of mature assembly. The prevention of the β-hairpin formation, either by the deletion of the N-terminal proline or ten initial amino acids enabled the assembly of ΔProCANC or Δ10CANC into immature-like, spherical particles. Detailed 3D structural analysis of these particles revealed that below a disordered N-terminal CA layer the C-terminus of CA assembles a typical immature lattice, which is linked by rod-like densities with the RNP.

Introduction

Assembly of fully infectious retroviruses comprises two different phases – immature and mature particle formation. Initially, the virus is formed in immature form by self-association of the major structural polyprotein Gag into spherical particles either at the plasma membrane (alpharetroviruses, gammaretroviruses or lentiviruses) or at a distinct site within the infected cell.
The particles are organized by lattices of Gag and Gag-derived polyproteins that associate with the plasma membrane that becomes outer layer of the immature particle released during budding. Upon release from the host cell, the Gag precursor is proteolytically processed into N-terminal membrane binding matrix protein (MA), two-domain capsid protein (CA) and RNA-binding nucleocapsid protein (NC). This cleavage leads to a dramatic morphological change and reorganization of the virus particle, from its immature form, into a mature, infectious virion.

Gag itself is sufficient for immature particle assembly. The interactions critical for the immature particle assembly are mediated by CA-CA and NC-RNA interfaces. CA consists of two subdomains: CA-NTD and CA-CTD, connected by a short sequence. Despite a low level of sequence homology, all available retroviral CA NTD structures show conserved structural motifs consisting of six to seven α-helices and an N-terminal β-hairpin anchored by a salt bridge between the N-terminal proline and an aspartate in helix 3 (10, 20, 29, 30, 42-44). CA-CTD contains a dimerization domain and is essential for the assembly of Gag. In the alpharetroviruses (e.g. Rous sarcoma virus, RSV) and lentiviruses (e.g. HIV-1), a short spacer peptide (SP, 12 or 14 amino acids, respectively) separates CA from NC. This SP domain plays a critical role in assembly, as mutations or deletions in SP influence immature spherical particle assembly (1, 28, 33, 37). Both HIV-1 and RSV SP sequences were also suggested to function as a molecular switch, because the deletion of SP1 of HIV-1 Gag led to a formation of tubular instead of spherical particles (23). Similarly, mutations within the SP sequence of RSV Gag change the spherical particle formation to the tubular one (28).
The appearance of all immature retroviral particles observed by transmission electron microscopy (TEM) is similar; they are roughly spherical with an inner, electron-dense ring. Based on cryo-electron microscopy (cEM) studies, the Gag polyprotein is arranged radially: associated with the plasma membrane through its MA domain and with the NC domain oriented towards the center of the particle \( \left( \text{65, 67} \right) \). Three-dimensional reconstructions of immature HIV-1 and RSV particles obtained by cryo-electron tomography have shown that the immature Gag lattice adopts a hexagonal arrangement \( \left( \text{5, 66} \right) \). To form a closed sphere from a hexagonal lattice, the incorporation of pentamers or irregular defects is required. Pentamers have not been observed in the immature Gag lattice, instead, the Gag hexameric lattice closes through the incorporation of heterogeneously shaped defects \( \left( \text{6, 35} \right) \). Upon maturation, NC with bound RNA condenses in the center of the particle, surrounded by a core shell assembled from released CA. The shapes of the mature cores differ according to retrovirus genus: in gamma-, delta-, and alpharetroviruses they are spherical, in betaretroviruses cylindrical and in lentiviruses conical. These assemblies are composed of a curved hexagonal lattice formed closed by incorporation of CA pentamers. The final shapes of retroviral cores are then determined by the location of the pentamers. The core of the most studied retrovirus, HIV-1, is fullerene-like cone composed of hexameric lattice with 12 pentamers, five at the narrow end of the cone and seven at the broad end of the cone \( \left( \text{18, 36, 47} \right) \). Although the CA domain drives both the mature and immature particle assembly, the lattices of immature particle differ from those of the mature ones \( \left( \text{5-7, 66} \right) \). The spacing of immature HIV-1 Gag lattice is smaller than that in the mature particle.
and the positions of the CA-NTD and CA-CTD also differ in the immature and mature lattices (6).

Gammaretroviruses, like other type C retroviruses, assemble their immature particles at the plasma membrane. Similarly to other retroviruses, their genome contains three open reading frames: gag, pol and env. One member of the gammaretroviral family, the Xenotropic murine leukemia virus-related virus (XMRV), was recently discovered in association with human prostate tumors (58) and chronic fatigue syndrome (CFS) (38). However, the association has not been confirmed in other studies (15–17, 25, 26, 51–53, 56, 60), thus the relevance of XMRV to human disease remains unclear. Recent publications provided strong evidence that XMRV does not cause CFS (32, 45, 49, 54). Its genome is most similar to that of exogenous Murine leukemia virus MLV DG-75, isolated from a human B-lymphoblastoid cell line (48), with which it shares 94% sequence identity (58). The XMRV Gag polyprotein consists of 536 amino acids and it is separated from Pro-Pol sequence with UAG stop codon. Based on the sequence similarity with the closely related MLV, it is expected that the XMRV Gag also comprises four structural proteins: MA, p12, CA and NC. The MLV capsid protein, similarly to all retroviral CAs, contains an N-terminal β-hairpin that is stabilized by Pro1-Asp54 interaction (43, 44). This hairpin is formed only upon maturation and its prevention markedly reduces viral infectivity (61, 62, 64).

Similarly to betaretroviruses, gammaretroviruses lack a spacer peptide between CA and NC. However, gammaretroviruses possess a unique feature that is the presence of a 41-amino acid stretch, rich in both positively and negatively charged residues at the C-terminus of CA.
(Fig.1 underlined sequence). This region, due to its possible α-helical structure, was named “charged assembly helix”, and is important for the assembly and production of infectious MLV (9, 63).

Studies on in vitro assembled virus-like particles (VLPs) have been proven to be a valuable tool for gaining basic information on retroviral assembly. A number of studies has mimicked the assembly of both immature virus particles and mature cores in vitro using purified recombinant Gag-derived capsid-nucleocapsid proteins (8, 21, 22, 31, 41, 57, 68). Although the in vitro assembly of retroviral genera including lentivirus HIV-1 (8, 12, 21, 22), betaretrovirus M-PMV (31, 57) and alpharetrovirus RSV (Rous sarcoma virus) (8, 40, 41, 68) has been intensively studied, there is a lack of information on the in vitro assembly of gammaretroviruses.

The in vitro assembly was achieved by using His-tagged MLV CA molecules anchored to Ni

chelating lipid nanotubes as an assembly substrate (24). Recently, a comparison of MLV and HIV-1 monomeric Gag molecules revealed their different behavior in solution. In contrast to HIV, the MLV Gag has a much weaker propensity for inter-protein interactions and has a rigid and extended structure, likely caused by a proline-rich region between MA and p12 (11). In this work, Datta et al., also demonstrated that MLV Gag could be assembled in vitro into fragile VLPs with an irregular appearance.

Here, we present the assembly of XMRV CANC protein into immature virus-like spherical VLPs by using an in vitro expression/assembly system. We prepared three N-terminal variants of XMRV CANC and compared their ability to assemble. We found that formation of the N-terminal β-hairpin must be prevented to obtain immature-like particles from XMRV.
CANC. Using electron microscopy and cryo-electron tomography we have characterized the structure of assembled immature XMRV VLPs and compared it with other retroviruses.

**Material and methods**

**Cloning.** All expression vectors were prepared by standard cloning techniques, propagated in *E. coli* DH5α and verified by sequencing. The fragment encoding XMRV CANC was obtained by PCR (using primers: 5´**CAT ATG** CCA CTC CGC ATG GGG, 3´**CTC GAG** CTA GTC ACC TAA GGT CAG G) of cDNA of human prostate carcinoma 22Rv1 cells infected with XMRV (kindly provided by Dr. Hajduch). The PCR fragment was digested with *Nde*I and *Xho*I and ligated into an expression vector pET-22b under the T7 promoter. The CANC derived constructs were prepared analogously, with the 5´ primers **CAT ATG** CAG TAC TGG CCG TTT TCC TC for Δ10CANC, **CAT ATG** CTC CGC ATG GGG GGA G for ΔProCANC.

**Bacterial expression:** Luria-Bertani medium containing ampicillin (100ug/ml) was inoculated with *E. coli* BL21 (DE3) cells carrying the appropriate construct. At the optical density at 590nm of ~ 0.8, the expression was induced by the addition of isopropyl-B-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4mM. The cells were harvested 4h post-induction.

**Protein purification:** Bacterial cells were resuspended in lysis buffer A (50mM Tri-HCl, 150mM NaCl, 1 mM EDTA, pH 8.0), disrupted with lysozyme, sonicated and incubated with sodium deoxycholate. After centrifugation at 10000 x g, proteins were solubilized from pellet by addition of 0.5% Triton X-100 and 1-1.5M NaCl in the lysis buffer A. The XMRV protein was precipitated from the supernatant by ammonium sulfate (final concentration 25% w/v). The pellet
after low speed centrifugation was resuspended in buffer E (20mM Tris-HCl, pH 8.0, 0.1M NaCl, 50μM ZnCl₂, 10mM DTT, 1mM PMSF), dialyzed overnight against the same buffer at 4°C and loaded on the top of DEAE cellulose column. The flow-through and washing fractions were pooled and loaded on the top of phosphocellulose column. The bound protein was eluted by NaCl gradient from 0.1M to 1M NaCl in the buffer E. The fractions containing desired proteins were dialyzed overnight against a storage buffer (buffer E with 0.5M NaCl), concentrated and loaded on the top of Sephadex G-100 column. The purified proteins were concentrated (CANC to 5 mg/ml, ΔProCANC to 2mg/ml, and Δ10CANC to 4 mg/ml), and stored at -70°C. The purity of each protein was analyzed by SDS-PAGE (Fig. 3) and the protein identities were confirmed by N-terminal sequencing.

**In vitro assembly**: An aliquot of 60μg of purified protein was mixed with 6μg of DNA (oligonucleotide 30-mer, 40-mer, λDNA or phage M13 ssDNA) or RNA (MS2) in a total volume of 100 μl of the storage buffer (20mM Tris-HCl, pH 8.0, 0.5M NaCl, 50μM ZnCl₂, 10mM DTT, 1mM PMSF). The mixture was dialyzed from 2 hours to overnight against the assembly buffer (50 mM Tris buffer pH 8.0, containing 100 mM NaCl and 1μM ZnCl₂) at 4°C. 100 μl of the assembled material (in the presence or absence of oligonucleotide) was loaded on the top of 20-65% (wt/wt) sucrose gradient and centrifuged in Beckman SW41Ti for 16 h at 4°C. Individual fractions were analyzed by Western blot analysis using rabbit anti-XMRV CANC antibodies. The sucrose density in the fractions was determined using refractometer.

**Electron microscopy**: Particles formed during assembly reaction were negatively stained with 4% sodium silicotungstate (pH 7.4) on carbon-coated grids and studied by transmission electron
microscopy (JEOL JEM-1200EX) with a microscope operated at 60kV. For thin-section electron microscopic analysis the bacterial pellets of the induced samples were fixed in 2.5% glutaraldehyde, postfixed in 1% osmium tetroxide, dehydrated by applying an ethanol series and embedded in Agar 100 epoxy resin. Ultrathin sections (70nm) were counterstained with saturated uranyl acetate and lead citrate.

Cryo-electron microscopy: *In vitro* assembled particles were deposited on C-flat holey carbon grids, and vitrified by plunge freezing in liquid ethane. Tilt series were collected on an FEI Tecnai F30 “Polara” transmission EM with Gatan GIF 2002 post column energy filter and 2kx2k Multiscan CCD camera. Data collection was performed at 300kV with an electron dose of approximately 30 e/Å². The nominal defocus 2.0 µm, with a magnification of 34,000X resulting in a pixel size at the specimen level of 4Å.

Cryo-electron tomography: *In vitro* assembled particles were mixed with 10nm gold beads, deposited on C-flat holey carbon grids, and vitrified by plunge freezing in liquid ethane. Tilt series were collected on an FEI Tecnai F30 “Polara” transmission EM with Gatan GIF 2002 post column energy filter and 2kx2k Multiscan CCD camera. Data collection was performed at 300kV using the SerialEM software package and tomograms were reconstructed using the IMOD software package (34). Tilt series were collected between -60 and +60 degrees with an angular increment of 3 degrees and a total electron dose of approximately 70e/Å². The defocus range was between 3.0 and 4.0µm, with a magnification of 34,000X resulting in a pixel size at the specimen level of 4Å.
2D Fourier analysis: 70 individual particles were boxed out using EMAN boxer (39) and padded to a final size of 1024x1024 pixels. A soft spherical mask was applied with a radius of 22nm for the rotationally averaged power spectrum, or 15nm for the rotationally aligned power spectrum. (The larger mask includes more data, giving stronger peaks necessary for the alignment. The narrower mask minimizes the smearing of the peak due to particle curvature.) Calculation of power spectra, their rotational average and rotational alignment were performed in IMAGIC (59).

RAF analysis: Radius-angle-frequency plots were generated as described previously (13). Briefly subtomograms of 384nm$^3$ were extracted from tomograms of 15 VLPs. The extraction was done along the surface of a sphere centered in the center of each VLP and with a radius equal to the mean radius at CA-CTD level. The subtomograms underwent an orthographic radial projection. The focus of the projection was determined as the geometric center of the particle from which the subtomogram was extracted. The resulting volumes corresponded to flattened subtomograms. The volumes were then padded to 1024x1024 pixels in XY and 2D power spectra were calculated at each radius. The rotation-autocorrelation function with a rotation range between 0° and 180° was calculated for each power spectrum. For each subtomogram this generates a 3D plot with radius on one axis, and the other two axes representing those of the rotational autocorrelation function of the power spectrum, namely angle and frequency. The presence of a peak at a particular point in radius, frequency and angle indicates that at that radius in the virus, the 2D power spectrum of the protein layer has peaks at that frequency which are arranged rotationally symmetrically repeating at that angle.
Subtomogram averaging: The extracted subtomograms were iteratively aligned and averaged using a six-dimensional search as described previously (6). The initial reference used for the alignment was the average of the subtomograms in the extraction position. 6-fold symmetry was applied to the average at each iteration. The threshold for the subtomograms that had to be averaged was set to the mean cross-correlation value of the all subtomograms to the reference.

Results

Bacterial and \textit{in vitro} assembly.

To obtain the XMRV expression vector, the CA and CA-NC regions were amplified using cDNA of human prostate carcinoma 22Rv1 cells infected with XMRV. DNA sequencing confirmed 100% identity of the CANC with the CANC of XMRV, clone VP62 (Fig. 1). To study the ability of the XMRV CA and CANC to assemble into virus-like particles (VLP), we used both bacterial and \textit{in vitro} assembly systems (50, 57). Transmission electron microscopy (TEM) of thin sectioned \textit{E. coli} BL21(DE3) expressing XMRV CA and CANC showed amorphous inclusion bodies instead of assembled mature-core-like structures (Fig. 2A, B). This finding was surprising because several studies have demonstrated that retroviral CANCs of different species (HIV-1, RSV), with Pro1 at their N-termini, assemble into tubular, mature-core-like structures (8, 21, 22, 41).

It is known that the prevention of the β-hairpin formation (either by the deletion or substitution of Pro1, or by N-terminal extension of CA) can shift the assembly mode to formation of immature-virus-like structures (21, 46, 50, 57, 61). We therefore decided to truncate the
XMRV CANC construct by the deletion of the first proline residue (ΔProCANC) with the aim of assembling immature-like structures. TEM analysis of the cells expressing ΔProCANC protein indicated its tendency to assemble into numerous curved structures (Fig. 2C) together with a small number of approximately spherical particles. In order to delete a substantial portion of the N-terminal β-hairpin we next generated a further N-terminal truncation of CANC, locating the start of the XMRV CA protein ten amino acids downstream from the Pro1, at Gln11 (Fig. 1). TEM of the thin sections of *E. coli* expressing the resulting Δ10CANC construct showed the presence of roughly spherical particles (Fig. 2D).

To further analyze the ability of XMRV CANC-derived proteins to assemble *in vitro*, all studied proteins were purified using procedures similar to those described earlier for M-PMV and HIV-1 CANC (8, 41, 57) (Fig. 3A). An *in vitro* assembly reaction was performed by overnight dialysis of a mixture of purified XMRV CANC-derived proteins with DNA (oligonucleotide 30-mer, 40-mer, λ DNA, phage M13 ssDNA) or RNA (MS2). To investigate the homogeneity of the assembled particles, the material was centrifuged through linear 20-65% sucrose gradient and individual fractions were analyzed by Western blotting using rabbit anti-XMRV CANC antibody. In the presence of the oligonucleotide (40-mer), the majority of the assembled particles were found in the fractions of sucrose density of 1.19-1.23 g/ml corresponding to the properly assembled spherical particles (Fig. 3B upper panel). In the absence of the oligonucleotides, bulk of unassembled material remained at the top of the sucrose gradient (Fig. 3B lower panel). The gradient profiles were similar for both the ΔProCANC and Δ10CANC proteins. The samples of the *in vitro* assembled particles were negatively stained with 4% sodium silicotungstate (pH 7.4).
and studied by TEM. In contrast to XMRV CANC protein, where only free or aggregated proteins were observed, the ΔProCANC and Δ10CANC proteins assembled into spherical particles (Fig. 3C). Particles of XMRV ΔProCANC and Δ10CANC were not detected if the *in vitro* assembly was carried out in the absence of nucleic acid or at the NaCl concentration higher than 0.5M (data not shown), suggesting that assembly must be initiated by the addition of nucleic acids under suitable conditions. Similar assembly requirements have been observed for the *in vitro* assembled particles of RSV CANC (8), HIV-1 CANC (8,21,22) and M-PMV ΔProCANC (57).

TEM images of negatively-stained *in vitro* assembled XMRV Δ10CANC particles were then compared to those of HIV-1 and M-PMV ΔProCANC (Fig 3D). The major difference between the particles was in the appearance of particle protein layers. While the HIV-1 and M-PMV ΔProCANC particles appeared as a single layer protein shell with visible rod-like projections (Fig. 3D), the XMRV Δ10CANC particle’s shell appeared as three smoother protein layers separated by distinct rings of lower electron density (Fig. 3D).

To study the fine structure of the *in vitro* assembled XMRV particles into more detail and to avoid artifacts associated with negative staining, we carried out cryo-electron microscopy (cEM) of particles made of Δ10CANC protein. The sample was prepared by the dialysis of Δ10CANC against pH 8.0 buffer in the presence of lambda dsDNA. The cEM images showed three density layers, with the outer two layers close together (Fig. 4A). The mean size of the particles was 82nm±11nm.
We compared our images of the *in vitro* assembled XMRV particles (Fig. 4A), with published images of PR− MLV (Fig. 3A and Fig. 5 from Yeager 1998 (67)). The number of layers and the appearance of the layers are extremely similar. For an optimal comparison, we applied the same image analysis procedure that was used for the PR− MLV. We calculated the rotationally averaged power spectra of the particle centers, which showed peaks corresponding to repeating elements in the protein lattice (Fig. 4B). Two peaks were seen, centered at 64 ± 9Å and 42 ± 3Å. If these are indexed as [1,0] and [1,1] reflections of a hexagonal lattice, they describe a unit spacing of 74Å or 73Å. The equivalent analysis for PR− MLV is shown in Fig. 6 from Yeager 1998 (67). The pattern of peaks is the same. For the PR− MLV, the peaks are centered at 67 ± 11Å and 45 ± 10Å, describing a unit cell spacing of 77Å and 78Å. We next carried out a rotational alignment and averaging of the power spectra, as described in (5), revealing that a hexagonal pattern is present in the XMRV particles (Fig. 4B), and confirming the assignment of the [1,0] and [1,1] reflections. Similar patterns were described for individual particles of PR− MLV (Fig. 3 in Yeager 1998 (67)).

To further understand the structural organization of the XMRV particles, we collected cryo-electron tomograms (Fig. 5A) and directly compared the radial density profile of XMRV Δ10CANC with those of HIV-1 (using datasets from de Marco et al. (12)). The radial density profile of XMRV Δ10CANC is similar to that of HIV-1, (Fig. 5B). By comparison with HIV-1, the two outer density layers correspond to the NTD and CTD of CA and the inner one to NC with bound nucleic acid.
In order to identify which layers show 6-fold symmetry we used the 3D information present in the cryo-electron tomograms and performed radius-angle-frequency (RAF) analysis (13) (Fig. 5C). Peaks in the RAF plot indicate the presence of symmetrically ordered density at a particular radius and with a particular spacing. We identified two clear peaks: one corresponds to the [1,0] reflection of the expected hexameric lattice in the region between NC and the CA-CTD, the other to the [1,1] reflection of the hexameric lattice in the CA-CTD region. In vitro assembled HIV particles show the same two peaks, as well as two further peaks in the region between CA-CTD and CA-NTD, and in the CA-NTD region. These peaks are missing in XMRV, indicating that, unlike in HIV, the CA-NTD region of the XMRV particles does not show any detectable hexameric lattice.

Knowing that the region between the RNP and the CA-CTD adopts a hexagonal lattice, we performed subtomogram averaging in order to get a 3D representation of the structure (Fig. 6A). Consistent with the RAF analysis, outermost in the structure is a poorly featured layer that corresponds to the CA-NTD (blue in Fig. 6B). In contrast, the CA-CTD domain (green) forms an arrangement similar to that seen in other retroviral immature lattices (12). As in HIV and RSV, the CA-CTD of XMRV appears to be connected to the RNP (gray in Fig. 6B) via rod like structures (red in Fig. 6B).

**Discussion**

The release of the N- and C-termini of CA, upon maturation, leads to local conformational transitions that influence the overall structure of the protein and thus creates novel protein-protein
interfaces (for review see (19)). The model for subunit arrangement in immature particles implies that the hexagonal lattice is held together by intra- and inter-hexamer interactions within SP and CA-CTD (66). The CA-NTD is not absolutely required for the immature particle assembly (2, 4). Cleavage of Gag by viral protease allows the formation of a β-hairpin at the N-terminus of CA-NTD and the separation of CA from the putative SP1 helical bundle. This induces structural rearrangement of CA and the formation of mature hexameric CA-NTD rings that hold together the mature lattice. CA-CTD interactions also mediate inter-hexamer interactions in the mature lattice.

Structural and biochemical data suggest that the ordered arrays of CA are regulated by the presence (β+) or absence (β-) of the β-hairpin at the N-terminus and the presence (SP+) or absence (SP-) of the SP sequence at the CA C-terminus, thus determining the particle morphology (23, 28). By this convention, within the immature virus CA is in the form β-/SP+, within the mature virus in the form β+/SP-. In vitro, the HIV-1 ΔMACANC and RSV ΔPRGag proteins where the β-hairpin formation is prevented by downstream N-terminal sequences of CA, but SP is present (β+/SP+), assembled into immature spherical particles (21, 27). In vitro and in vivo assembly studies on HIV CANC and RSV CANC show that the presence of both N- and C-terminal determinants (β+/SP+) of CA results in mature-like, tubular (and in the case of HIV-1 also conical) particle assembly (8, 18, 22). The HIV-1 ΔMACAΔSP1NC and RSV ΔPRGag proteins, which do not form the β-hairpin and SP is mutated or deleted (β-/SP-), assemble into tubular (and in the case of HIV-1 also conical) particles (23, 28). The mature CA proteins (β+/SP-) of HIV and RSV assemble in vitro into tubular structures, however, the conditions under
which are these structures formed (~1M NaCl) suggest different types of CA-CA interactions (14, 22, 30).

M-PMV, which lacks the SP sequence between CA and NC, appears to behave differently in vitro. In the absence of the β-hairpin and in the presence of downstream NC, M-PMV forms spherical particles. In the presence of the β-hairpin MPMV cannot be assembled in vitro. Here we show that XMRV behaves in the same way: it is unable to form mature-like structures in vitro from CA or CANC but can form immature-like ΔProCANC and Δ10CANC particles. In contrast to HIV-1 and RSV, XMRV and M-PMV do not have SP separating CA and NC domain. It begs the question, whether some part of XMRV CA or NC might substitute for the SP-like domain.

Based on the recent data showing that MLV, in contrast to other retroviruses, does not have strict structural requirements for the N-terminal portion of NC (55), it is likely that the SP-like domain occurs within the CA CTD. Here we show that the region between CA and NC forms a rod-like structure similar to that previously been described for HIV and RSV, and consistent with the presence of a structured region between the two folded domains. A possible candidate is the “charged assembly helix” at the C-terminus of the MLV CA (9). This region was suggested to form amphipathic α-helix and mutations within this region showed its necessity for MLV assembly (9). The SP-like domain in M-PMV was localized at the CANC junction (3), however, its precise length is currently being studied. It is possible that such sequences at the C-terminus of CA in XMRV and M-PMV are what prevents the in-vitro assembly of mature virus like assemblies.
Given the high degree of sequence similarity with MLV Gag, the arrangement in the immature-like in vitro assembled XMRV particles should be comparable to that in immature MLV particles, as described by Yeager et al. (67). Indeed, we found that both particles have a similar appearance and adopt a hexameric lattice with the same spacing. The CA and NC domains formed noticeable layers in both cryo-EM and TEM images of the in vitro assembled XMRV Δ10CANC, which correspond well to the features in the immature MLV particles shown by Yeager et al. (67).

3D structural analysis of the particles revealed that they form an ordered hexameric lattice with mostly similar structural features to those previously described for HIV, M-MPMV and RSV particles. Unexpectedly, in contrast to these other viruses, we found that the outer most layer in the XMRV Δ10CANC particles, corresponding to the CA-NTD region, does not form an ordered hexameric lattice. Two formal explanations can be considered: the CA-NTD may be laterally disordered in the in vitro assembled XMRV particles; or gammaretroviruses may adopt a different structural arrangement to other retroviruses, in which, for example, part of CA-NTD may fill the typical holes in the CA-NTD layer, giving a smooth appearance. Given the high degree of secondary structure similarity between the CA-NTD of MLV and those of other retroviruses, we consider the second possibility highly unlikely. It is sensible to conclude that the in vitro assembled XMRV Δ10CANC particles faithfully mimic the structure of the immature XMRV Gag lattice in the CA-CTD and immediately downstream regions, but shows some lateral disorder in the CA-NTD region. We hypothesize that the difference in the order of CA-NTD (as shown in the RAF plots and in the 3D structure, Fig.5C and 6B, respectively) could explain the
different appearance of the TEM images of negatively stained XMRV particles compared to HIV-1 and M-PMV (Fig. 3D). Although the cryo-EM radial densities of XMRV and HIV-1 are very similar (Fig. 5B), showing the same radial positions of the CA-NTD, CA-CTD, RNP domains, the cryo-electron tomography clearly shows that the degree of order in the XMRV CA-NTD (as shown in the RAF plots and in the 3D structure, Fig. 5C and 6B, respectively), is very different.

Many previous studies indicated the importance of in vitro assembly studies of different retroviruses to obtain valuable information on the principles of retroviral assembly. The system for the in vitro assembly of MLV was described for CA protein (24), recruited to lipid nanotubes as an assembly template. Another in vitro assembly was reported recently for MLV Gag (11). However in this system the MLV Gag VLPs were fragile and of “irregular appearance”. The protocol presented here for the efficient in vitro assembly of virus-like Gag-derived particles in solution provides the basis for the future studies of detailed protein packing and arrangement in gammaretroviral particle. Importantly, although the region downstream of XMRV CA does not contain a spacer peptide and has an unusual run of charged residues (9), we found that it assembles into a typical immature retroviral lattice with rod-like densities linking CA to the RNP. Further investigation of this system could therefore help to better understand the critical role of this region in retrovirus assembly.
Acknowledgement

We thank Dr. Hajduch for kindly providing us with the cDNA of 22Rv1 cells and Romana Cubinková for excellent technical assistance.

This work was supported by the Grant Agency of the Czech Republic grant 204/09/1388 and by the research projects 1M0508, 1M0520, Z40550506, and MSM 6046137305 from the Czech Ministry of Education and NIH grant CA 27834. JAGB was supported by a grant from the Deutsche Forschungsgemeinschaft within SPP 1175. This work was technically supported by EMBL IT Services.


Figure legends:

Figure 1
Amino acids alignment of XMRV and MLV DG-75, Moloney MLV and MLV (strain BM5) CANC proteins. The dots represent identical amino acid residues, the black arrows under the alignment show the start of XMRV ΔProCANC and Δ10CANC truncated proteins, underlined sequence shows the charged residues at the C-terminus of MLV and XMRV CA. Position of β-strands and α-helices from the solved MLV CA-NTD structure (pdb 3bp9) are shown above the alignment.

Figure 2
Electron micrographs of thin sections of E. coli cells expressing XMRV: CA (A), CANC (B), XMRV ΔProCANC (C), and Δ10CANC (D). The scale bar is 200 nm.

Figure 3
A) SDS-polyacrylamide analysis of purified XMRV proteins (Coomassie blue-stained gel): lane 1: CANC, lane 2: Δ10CANC, lane 3: Δ ProCANC. B) Western blot analysis of in vitro assembled XMRV particles in sucrose gradient fractions in the presence (upper panel) or in the absence (lower panel) of..
absence (lower panel) of oligonucleotides. Δ10CANC protein (60 µg) was mixed with 6 µg of oligonucleotide 40-mer, dialyzed overnight against the assembly buffer and centrifuged to equilibrium through a linear 20-65% sucrose gradient. Fractions were analyzed by Western blotting with rabbit anti-XMRV CANC antibodies. The sucrose densities in the fractions 8-12 were: 8: 1.19 g/ml, 9: 1.21 g/ml, 10: 1.22 g/ml, 11: 1.23 g/ml, 12: 1.25 g/ml. C) Transmission EM images of negatively stained \textit{in vitro} assembled material from XMRV ΔProCANC (left panel), and XMRV Δ10CANC (right panel). The scale bar is 200 nm. D) Transmission EM images of negatively stained VLPs assembled \textit{in vitro} from XMRV Δ10CANC, HIV-1 ΔProCANC and M-PMV ΔProCANC. The scale bar is 50 nm.

**Figure 4**

Cryo-electron microscopy of XMRV Δ10CANC particles. A) Cryo-electron micrographs of particles assembled \textit{in vitro} from XMRV Δ10CANC. Scale bar 50 nm. The white dashed circle shows the region used for Fourier analysis. B) Rotationally averaged power spectra of the particle centers, showing the positions of the [1,0] and [1,0] reflections from a hexameric lattice. The inset displays the sum of the rotational aligned power spectra, indicating the presence of a hexagonal lattice.

**Figure 5**

Cryo-electron tomography of XMRV Δ10CANC particles. A) Computational slice of 0.8 nm thickness through a cryo-electron tomogram of particles assembled \textit{in vitro} from XMRV
Δ10CANC. The white arrows point to the CA-NTD, the black arrows point to the CA-CTD, while the white arrowheads point to the NC region. The scale bar is 50 nm. B) Comparison between the radial density profiles of XMRV Δ10CANC (red) and HIV-1 (blue). The two profiles have been aligned on the CA-CTD, and the X-axis shows the radial distance from the CA-CTD region. Peaks corresponding to NC and to the NTD and CTD of CA are marked. C) The two panels show a slice through the Radius-Angle-Frequency (RAF) plot at 60° angle for XMRV Δ10CANC (top) and HIV-1 (bottom). The X-axes are aligned to the X-axis of the radial density plot (panel B). The expected positions of the [1,0] and [1,1] reflections for a hexameric lattice with a spacing of 7.4 nm at the CA-CTD are indicated by white dotted lines. Strong peaks indicating hexagonal order are seen for both XMRV Δ10CANC and HIV-1 in the CA-CTD and downstream regions. Only HIV shows strong peaks in the CA-NTD region. The radial position of the RNP and CA domains is indicated on the plots.

Figure 6

Structure of in vitro assembled particles of XMRV Δ10CANC. A) Orthoslices through the average structure. The right panels are radial sections, while the left panels are tangential sections. The dashed lines show the relative position of the orthoslices between the panels. Density is black. B) Surface rendering of the average structure. The color scheme follows an approximate radial division of the structure according to the different domains of Gag as previously assigned in (12). Gray is the NC-nucleic acid layer, red is the region between NC and the CA-CTD, green is the CA-CTD region, and blue is the CA-NTD. The dashed lines on the top
view represent the section cut out in the side view. The CA-NTD region does not show an ordered hexameric lattice.
Figure 1

Amino acid alignment of XMRV and MLV D-35, Moloney MLV and MLV (strain BM5) CA proteins. The dots represent identical amino acid residues, the black arrows under the alignment show the start of XMRV ΔEnvCA, and ΔCA, and truncated proteins, underlined sequence shows the charged residues at the C-terminus of MLV, and XMRV CA. Positions of β-strands and α-helices from the solved MLV CA NTD structure (pdb 3hp9) are shown above the alignment.
Figure 2
Electron micrographs of thin sections of E. coli cells expressing XMRV-CA (A), CANC (B), XMRV ΔProCANC (C), and Δ10CANC (D). The scale bar is 200nm.
Figure 4
Cryo-electron microscopy analysis of XMBV ΔICANC particles. A) Cryo-electron micrographs of particles assembled in vitro from XMBV ΔICANC. Scale bar 50 nm. The white dashed circle shows the region used for Fourier analysis. B) Rotational averaged power spectra of the particle centers, showing the positions of the [1,0] and [1,0] reflections from a hexameric lattice. The inset displays the sum of the rotational aligned power spectra, indicating the presence of a hexagonal lattice.
Figure 5
Cryo-electron tomography of XMRV Δ10CANC particles. A) Computational slice of 0.1nm thickness through a cryo-electron tomogram of particles assembled in vitro from XMRV Δ10CANC. The white arrows point to the CA-NTD, the black arrows point to the CA-CTD, while the white arrowheads point to the NC region. The scale bar is 50 nm. B) Comparison between the radial density profiles of XMRV Δ10CANC (red) and HIV-1 (blue). The two profiles have been aligned on the CA-CTD, and the X-axis shows the radial distance from the CA-CTD region. Peaks corresponding to NC and to the NTD and CTD of CA are marked. C) The two panels show a slice through the Radius-Angle-Frequency (RAF) plot at 60° angle for XMRV Δ10CANC (top) and HIV-1 (bottom). The X-axes are aligned to the X-axis of the radial density plot (panel B). The expected positions of the [1,0] and [1,1] reflections for a hexameric lattice with a spacing of 7.4 nm at the CA-CTD are indicated by white dotted lines. Strong peaks indicating hexagonal order are seen for both XMRV Δ10CANC and HIV-1 in the CA-CTD and downstream regions. Only HIV shows strong peaks in the CA-NTD region. The radial position of the KNP and CA domains is indicated on the plots.
Figure 6
Structure of in vitro assembled particles of XMRV ΔR(CANC). A) Orthodicies through the average structure. The right panels are radial sections, while the left panels are tangential sections. The dashed lines show the relative position of the orthodicies between the panels. Density is black. B) Surface rendering of the average structure. The color scheme follows an approximate radial division of the structure according to the different domains of Gag as previously assigned in (12): Gray is the NC-matrix acidic layer, red in the region between NC and the CA-CTD, green in the CA-CTD region, and blue in the CA-NBD. The dashed lines on the top view represent the section cut out in the side view. The CA-NBD region does not show an ordered hexagonal lattice.