In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly

Nancy A. Jewell and Louis M. Mansky

1 Department of Molecular Virology, Immunology, and Medical Genetics, Center for Retrovirus Research, and Comprehensive Cancer Center, Ohio State University Medical Center, 2078 Graves Hall, 333 West 10th Ave, Columbus, OH 43210, USA
2 Molecular, Cellular, and Developmental Biology Graduate Program, Ohio State University, USA

Introduction

The RNA encapsidation process for retroviruses involves a recognition event of the genome-length viral RNA by the viral Gag polyprotein, which acts to initiate the assembly of virus particles. The mechanism underlying this recognition event is not entirely understood, but biochemical and genetic analyses have revealed that this event involves the interaction between stable RNA secondary structures at the 5’ end of the viral genome and, in many cases, amino acids in the nucleocapsid domain of the Gag protein. This brief review will focus on the recent research directed at elucidating the mechanism of RNA encapsidation for the complex retroviruses. The complex retroviruses represent three genera of the family Retroviridae (i.e. Lentivirus, HTLV-BLV and Spumavirus) and include the known retroviruses that infect humans (and have therefore received a great deal of recent attention): human immunodeficiency virus (HIV), human T-cell leukaemia virus (HTLV) and human foamy virus (HFV).

RNA–protein interactions involved in genome recognition

The genome recognition event is a crucial early step in the assembly process of retrovirus particles. This recognition event leads to the predominant encapsidation (packaging) of the genome-length RNA into assembling particles (Fig. 1). This discrimination process, which is primarily a viral RNA–protein interaction, is known to strongly favour the full-length viral RNA to that of spliced viral RNAs and cellular mRNAs. In general, the RNA sequences necessary and sufficient for the RNA encapsidation process are located in a region that includes the 5’-noncoding region along with the 5’-half of the gag gene. These sequences are referred to as the encapsidation signal (E) or packaging signal (ψ). The encapsidation signal can enhance the encapsidation of E-containing RNAs up to 200-fold that of non-E containing RNAs, indicating that RNA encapsidation is enhanced by the presence of E and is not absolute (Rein, 1994).

The Gag polyprotein for complex retroviruses, as is true for simple retroviruses, appears to be sufficient for the formation of particles, and Gag-only particles can encapsidate viral RNA. These observations imply that the genome recognition event involves full-length viral RNA and unprocessed Gag polyprotein, and indicate that the viral protein components needed for specific viral RNA encapsidation are in Gag. Gag includes three main domains (matrix, MA; capsid, CA; nucleocapsid, NC). Attempts to address the RNA encapsidation process have focused on identifying RNA-binding domains within unprocessed Gag (and the processed Gag products) and demonstrating specificity in binding of viral RNA to Gag either in cell-free reactions or in the uptake of viral RNA into virus particles.

RNA signals for genome recognition

Lentivirus

In most retroviruses the location of the encapsidation signal is in the untranslated leader region between the major subgenomic splice donor and the start of the gag codon; therefore, spliced messages would lack the entire encapsidation signal. In HIV-1 the primary encapsidation signal is located downstream of the primer-binding site (pbs), overlaps with the major splice donor, and extends into the gag gene (McBride & Panganiban, 1997). The stable RNA secondary structures in this region consist of four hairpin loops, referred to as stem–loops 1–4 (SL1, SL2, SL3 and SL4) (Fig. 2a) (McBride & Panganiban, 1997). SL1, SL3 and SL4 have been shown to contribute to the encapsidation of HIV-1, while SL2 does not appear to have a major role in encapsidation (Harrison et al., 1998; Luban & Goff, 1994; McBride & Panganiban, 1996, 1997). The primary HIV-1 encapsidation signal spans the major subgenomic splice donor site, which is located on SL2. Such an arrangement provides a mechanism for the distinction between full-length genomic viral RNA and spliced RNA, since the spliced RNA would lack part of the primary encapsidation signal (McBride & Panganiban, 1997). SL3 specifically interacts with the NC domain and only genomic viral RNA contains SL3. The two zinc-binding motifs of HIV-1 NC form hydrogen bonds with the exposed

Author for correspondence: Louis Mansky.
Fax +1 614 292 9805. e-mail mansky.3@osu.edu
Fig. 1. A simplified overview of complex retrovirus genome recognition and virus assembly. (1) Transcription of the integrated provirus produces a full-length viral RNA. (1a) A singly spliced viral RNA is made (for simplicity, multiply spliced RNAs are not shown). (2) Full-length and spliced viral RNAs are transported to the cytoplasm. (3) Spliced RNA is translated to produce the envelope glycoprotein and the unspliced viral RNA is translated to produce the Gag polyprotein precursor. (3a) Genome recognition between the encapsidation signal (E) of the full-length viral RNA and the Gag polyprotein precursor occurs. (4) Envelope glycoprotein is transported to the cell membrane; Gag polyprotein multimerizes using RNA (single or dimeric viral RNA?) as a scaffold, leading to RNA encapsidation; viral RNA–Gag complexes are transported to the cell membrane via actin. (5) Formation of immature particles at the cell membrane. (6) Budding of immature virus particles. (7) Virus particles are released; complete proteolysis of Gag polyprotein by the viral protease produces mature progeny virus; reverse transcription occurs to completion in spumavirus particles.

guanosines on the SL3 loop, providing the main interaction between the NC and the encapsidation signal (De Guzman et al., 1998) (discussed later in this review). More recent data indicate that the conformation of SL3 is altered upon Gag protein binding, and that SL3 has a purine-rich internal loop that can act as a site for Gag interaction (along with the SL3 loop) (Zeffman et al., 2000). SL1 is present on both spliced and full-length viral RNA, yet data have shown that SL1 contributes to the selective encapsidation of genomic viral RNA, likely in conjunction with SL3 and SL4 (McBride & Panganiban, 1997). It should be noted that sequences that can influence HIV-1 vector RNA encapsidation have been identified in the env gene (Kaye et al., 1995).

The creation of mutations that altered the secondary structure by abolishing base-pairing at the stems of SL1, SL3 and SL4 resulted in a threefold reduction in encapsidation efficiency as compared to wild-type (wt) (McBride & Panganiban, 1997). Second site mutations were constructed that restored the base-pairing of the stem–loops, and these compensatory mutants were able to encapsidate viral genomic RNA at a level slightly lower than wt virus. This indicates that the structure of the RNA as well as the sequence contributes to the encapsidation of viral genomic RNA (McBride & Panganiban, 1997). The importance of both RNA structure and primary sequence has also been observed for bovine leukaemia virus (BLV) (Mansky & Wisniewski, 1998) (see below). In addition to the stem mutations, the HIV-1 SL1 and/or SL3 were deleted and it was found that RNA encapsidation of these mutants was reduced by five- to tenfold, indicating that SL1 and SL3 are required for efficient encapsidation (McBride & Panganiban, 1997). The requirement for SL1 and SL3 in the encapsidation process is also position-dependent. An HIV-1 mutant, in which the positions of SL1 and SL3 were interchanged, produced encapsidation levels comparable to when one of the stem–loops was deleted (McBride & Panganiban, 1997). From these data it was hypothesized that SL1, SL3 and SL4 form a higher-order structure that is responsible for the discrimination of genomic viral RNA over spliced RNA.

HIV-1 SL1, SL2 and SL3 are likely not the only cis-acting elements within the genomic RNA involved in RNA encapsidation. Recent data suggest that the region upstream of the primary HIV-1 encapsidation signal may contribute to optimal encapsidation of the viral genomic RNA. Within this region is the TAR stem–loop, a stem–loop distal to the TAR element termed the r-U5, and a structure designated u5-l that surrounds the pbs (McBride et al., 1997). Deletion of these three secondary structures produced encapsidation efficiencies five- to tenfold
lower than wt, similar to those observed when the region of SL1–SL4 was mutated (McBride et al., 1997). HIV-1 TAR primarily functions to regulate virus transcription through interaction with the Tat protein. Mutations constructed at the bulge and loop regions of TAR, where Tat binds, had no effect on RNA encapsidation, indicating that the role of the bulge–loop region of TAR in encapsidation is independent of Tat (Helga-Maria et al., 1999; McBride et al., 1997). The lower portion of the TAR stem was mutated to disrupt the base-pairing, resulting in an encapsidation efficiency that was 10- to 25-fold lower than wt. When compensatory mutations were introduced that recreated the stem–loop, the encapsidation efficiency was fully restored (Helga-Maria et al., 1999). For HIV-1 encapsidation, it therefore appears that the formation of the TAR stem is the important factor and not the sequence of this region (Helga-Maria et al., 1999). For the other two regions, it seems that the structures are important factors also. The secondary structure of r-U5 was found not to be essential for encapsidation, but conservation among primate lentiviruses suggests that r-U5 may contribute to optimal encapsidation (McBride et al., 1997). Although the exact structure surrounding the pbs is not agreed upon, the top of the u5–1
stem–loop contains a secondary structure that is conserved among different retroviruses (McBride et al., 1997).

HIV-2 RNA encapsidation has been reported to be distinctly different to that of HIV-1 encapsidation. Four deletion mutations were introduced into the 5′ UTR of HIV-2, where two of the deletions removed sequences upstream of the major splice donor site and the other two deletions removed sequences downstream (McCann & Lever, 1997). The downstream mutations reduced encapsidation efficiency by twofold, whereas analogous mutations in HIV-1 produced a 10- to 100-fold reduction in encapsidation efficiency (McCann & Lever, 1997). Mutations upstream of the HIV-2 splice donor resulted in a significant deficiency, three- to sixfold, in encapsidation. The upstream deletions not only affected encapsidation, but replication as well. In HIV-2, the upstream deletion that removed the dimerization signal produced the most severe results (McCann & Lever, 1997). (The potential role of RNA dimerization in genome recognition and RNA encapsidation will be addressed later in this review.) The inverse occurs in HIV-1, where mutations downstream of the splice donor affect both encapsidation and replication. These data suggest that the HIV-2 encapsidation signal is positioned on both spliced and unspliced viral RNAs, though no evidence of efficient encapsidation of spliced RNA has been shown. It has been proposed that HIV-2 employs a novel mechanism for the selection of unspliced viral RNA, in which unspliced RNA is translated, producing the Gag polyprotein. This polyprotein then binds to the encapsidation signal on that same viral RNA, thus directing that RNA for encapsidation (Kaye & Lever, 1999). (The potential interplay of RNA encapsidation and translation is discussed later.) The initial binding of the Gag polyprotein to the viral RNA could lead to the attraction of other Gag polyproteins to the viral RNA. The Gag monomers could bind to each other via homology regions and the NC subdomains would bind the viral RNA through non-specific interactions (Kaye & Lever, 1999).

HTLV-BLV

The encapsidation signal region of BLV was initially mapped by deletion analysis (Mansky et al., 1995). The efficiency of RNA encapsidation revealed two important regions. The first region includes sequences downstream of the pbs and near the gag gene start codon, and was found to be essential for RNA encapsidation. The second region was a 132 nucleotide base sequence within the gag gene that facilitates efficient RNA encapsidation in the presence of the first region. These results led to the conclusion that the encapsidation signal necessary for efficient RNA packaging and virus production is discontinuous. Structure–function analysis (Mansky & Wisniewski, 1998) has provided genetic evidence that the primary encapsidation signal region of BLV consists of two stable RNA stem–loop structures located just downstream of the gag start codon in the MA domain that are required for RNA encapsidation and virus production (Fig. 2b). A secondary encapsidation signal was characterized in the CA domain of Gag that consists of one stable stem–loop structure. It was also found that the encapsidation signal region of either HTLV-1 or HTLV-2 can replace the BLV primary encapsidation signal region and lead to either efficient or a modest level of replication of the chimeric virus, respectively. HTLV-1 and HTLV-2 have similar SL1 and SL2 structures downstream of the Gag start codon (Fig. 2c), and the HTLV-1 SL1 and SL2 sequence has been shown to specifically replace the BLV SL1 and SL2 in a BLV vector (unpublished results). In comparison, the Moloney murine leukaemia virus (MoMLV) encapsidation signal includes four motifs (Fig. 2d) (Mougel et al., 1996). Motifs C and D are necessary for efficient encapsidation, and the presence of either motif C or D is crucial for encapsidation and virus replication. These observations indicate that simple and complex retroviruses may not be that easily separated simply on the basis of RNA encapsidation signal architecture.

Spumavirus

The cis-acting sequences that are likely required for spumavirus RNA encapsidation have been mapped in the HFV genome. Although spumaviruses contain viral DNA in their particles, the genome recognition event involves the unspliced viral RNA (pregenome) (Linial, 1999; Yu et al., 1999). Two regions in the HFV genome have been identified as being important for efficient replication of spumavirus-based vectors (Erlwein et al., 1998; Heinkelein et al., 1998; Wu et al., 1998). One region spans from the r region to the 5′ end of the gag gene, while the second region is located towards the 3′ end of the pol gene. The location of encapsidation sequences in the pol gene would be a unique observation. There are three sites at the 5′ end of the unspliced RNA that have been reported to represent the dimer linkage structure (DLS), and mutation of the upstream site diminished RNA dimerization and inhibited HFV vector transfer by wt helper HFV (Erlwein et al., 1997, 1998). Structure–function analyses of the RNA structures involved in HFV RNA encapsidation have not been reported to date.

Protein domains involved in specific RNA binding

Lentivirus

The HIV-1 Gag polypeptide is initially expressed as a 55 kDa precursor (Pr55Gag) (Fig. 3); however, during the later stages of budding it is cleaved by the HIV-1 viral protease. The mature viral proteins released are MA, CA, NC and p6, along with two spacer peptides p1 and p2 (Zhang et al., 1998). HIV-1 virions with a defective viral protease still contain viral RNA, providing evidence that contact occurs between Pr55Gag and viral RNA during encapsidation (Freed, 1998). Of the four major domains of HIV-1 Pr55Gag, NC has been shown to have
a significant role in the encapsidation of the viral RNA genome.

The NC proteins of all retroviruses share two characteristics: a high percentage of basic residues and, with the exception of spumaviruses, a zinc-binding motif composed of regularly spaced cysteine and histidine residues (Klug, 1999) (Fig. 4). The Cys-His motifs are present in one or two copies, depending upon the virus, and contain the sequence Cys-X2-Cys-X4-His-X4-Cys (CCHC), where X indicates a variable amino acid residue (Rein, 1994). The X residue is variable not only among different retroviruses, but also between the two motifs of a single NC. The four conserved residues chelate zinc ions and are similar to the zinc finger motifs found in many DNA-binding proteins. Several experiments have indicated that the zinc-binding domains of HIV-1 NC bind specifically to the viral genomic RNA via the encapsidation signal (Freed, 1998).

Mutations in the HIV-1 NC that abolished zinc binding produced non-infectious virions that lacked their genomes and increased the incorporation of spliced viral RNA and cellular mRNAs. HIV-1 NC mutants have been constructed in which the first two cysteines of the first Cys-His box were changed to tyrosines (Schwartz et al., 1997). The result was a fivefold decrease in the encapsidation of genomic viral RNA and twice the amount of subgenomic RNA encapsidated. The gain in subgenomic RNA is an actual increase in the encapsidation of subgenomic RNA and not just a decrease in the encapsidation efficiency of genomic viral RNA (Schwartz et al., 1997). When the second Cys-His box of HIV-1 contained the same mutations, the encapsidation efficiency of genomic viral RNA was only slightly affected, 73% of wt level. The combination of the two mutated Cys-His boxes reduced encapsidation of genomic RNA by tenfold and increased subgenomic encapsidation twofold (Schwartz et al., 1997). These results suggest that the first Cys-His box plays a more prominent role in encapsidation and displays some selectivity for genomic viral RNA to be encapsidated. Cotransfection experiments with wt NC protein and the Cys-His box NC mutant revealed that different ratios of wt to mutant (i.e. 1:1, 1:5, 1:10 and 1:20) all produced encapsidation levels near wt (Schwartz et al., 1997). It was further observed that as the ratio of wt to NC mutant (either the first or second Cys-His box) increased, the amount of subgenomic RNA associated with virus particles also increased. This indicates that as the amount of NC mutant increased, the ability of the heterogeneous virus particles to discriminate between genomic and subgenomic RNA was reduced (Schwartz et al., 1997).

Other experiments that altered the basic residues flanking the zinc finger motifs resulted in a general decrease in the ability of HIV-1 NC to bind RNA. Arginine and lysine residues of HIV-1 NC were mutated by alanine scanning mutagenesis to determine the role of these basic residues in HIV-1 encapsidation (Poon et al., 1996). Mutation of lysine residues 11 and 14, which are located at the amino terminus of Cys-His box 1, produced encapsidation efficiencies fourfold lower than wt (Poon et al., 1996). Substitution of arginine 7 with alanine reduced encapsidation of viral RNA by twofold (Poon et al., 1996). Additionally, altering arginine 7 to a neutral amino acid drastically reduces the binding affinity of NC for viral RNA, more than any other single amino acid mutation (Schmalzbauer et al., 1996). Mutation of arginine 32 and lysine 33, located between the two Cys-His boxes, also leads to a decrease in
binding ability (Schmalzbauer et al., 1996). RNA-binding proteins contain a consensus sequence of RXRR or RXRK (where X is any amino acid); arginines 7 and 29 each correspond to the beginning of this motif in the NC protein of HIV-1. This RNA-binding motif is also found in HIV-2, simian immunodeficiency virus, Rous sarcoma virus and MoMLV (Schmalzbauer et al., 1996).

Complex retroviruses, with the exception of the spuma-viruses, contain two zinc-binding motifs in the NC domain that are required for replication. Ablation of one of the Cys-His motifs results in non-infectious viruses and insufficient encapsidation. As eluded to above, the two Cys-His motifs are not interchangeable or functionally equivalent. The first zinc-binding motif plays a prominent role in RNA selection and encapsidation, as evidenced by experiments in which the two motifs were either switched or duplicated. Mutant HIV-1 viruses that contained two copies of the second Cys-His motif, or with the positions of the first and second motifs reversed, encapsidated viral RNA at less than 15% of wt virus (Gorelick et al., 1993). This low level of encapsidation is similar to that observed in mutants with altered zinc-binding ability. In contrast, HIV-1 mutants that contained two copies of the first Cys-His motif encapsidated viral RNA at 70% of the wt level (Gorelick et al., 1993). This suggests that the first zinc-binding motif needs to be in the primary position for effective encapsidation. Comparison of the amino acid sequence of the Cys-His motifs reveals that the first motif is more highly conserved among retroviruses, implying that these residues may assist with the recognition and encapsidation of viral RNA (Gorelick et al., 1993).

While it is known that the NC domain is necessary for RNA encapsidation, it is not certain whether NC actually confers the selective recognition of the viral genomic RNA, as it is known that NC possesses a non-specific RNA-binding activity (Berkowitz et al., 1995). In order to determine the specificity of HIV-1 NC for viral RNA, a chimera was constructed in which the entire HIV-1 NC domain of Gag was substituted with the mouse mammary tumour virus (MMTV) NC domain (Fig. 3). This HIV-1 chimera preferentially encapsidated HIV-1 genomic RNA when both HIV and MMTV genomes were present. In the reciprocal experiment, the MMTV NC domain of MMTV Gag was replaced with the HIV-1 NC domain and this MMTV chimeric Gag was found to encapsidate the MMTV genome when both the MMTV and HIV-1 genomes were present (Poon et al., 1998). These observations indicate that the NC domain and zinc-binding motifs are not solely responsible for specific HIV-1 RNA packaging, and support the hypothesis that not all the protein domains required for specificity of genome recognition and RNA encapsidation lie within the NC domain. To test this, an MA domain deletion mutant revealed that the MA domain was not required for RNA packaging (Poon et al., 1998). It is worthy to note here that 90% of the MA domain in HIV-1 has been found to be dispensable for virus replication in a cell line in which the
cytoplasmic domain of Env is not required (Reil et al., 1998).] In total, these results suggest that the CA domain or other viral proteins may contribute to HIV-1 RNA packaging.

In contrast to the results described above, an earlier study created a chimeric HIV-1 Gag in which the MoMLV NC domain (Fig. 3) was substituted for the HIV-1 NC domain, as well as the creation of a chimeric MoMLV that contained the HIV-1 NC in place of the MoMLV NC domain. The MoMLV mutant, with HIV-1 NC, preferentially encapsidated unspliced HIV-1 viral RNA over spliced HIV-1, while the HIV-1 chimeric mutant, with MoMLV NC, encapsidated RNA that contained the MoMLV encapsidation signal (Berkowitz et al., 1995). These results imply that the NC domain is solely responsible for genome recognition. The important consideration when comparing the results between these two reports is the number of Cys-His motifs in the NC domains. The MMTV has two zinc-binding domains (and is therefore similar to the HIV-1 NC), while the MoMLV has only one. As noted above, the number of motifs plays an important role in encapsidation. Therefore, taken together, the data from these two papers indicate that the two zinc-binding domains are necessary but not sufficient for specific HIV-1 RNA encapsidation (Poon et al., 1998).

Selective encapsidation of viral RNA may occur through the non-specific binding ability of NC coupled to the specific recognition of the viral genome via another region in the Gag polyprotein (Berkowitz et al., 1995). There has been one suggestion that p2, the spacer peptide between CA and NC, may be that region (Kaye & Lever, 1998). Through cross-packaging experiments, it was found that HIV-1 could encapsidate both HIV-1 and HIV-2 RNA, yet HIV-2 was unable to encapsidate HIV-1 RNA (Kaye & Lever, 1998). However, HIV-2 chimeras that contained the HIV-1 NC and p2 domains were able to encapsidate HIV-1 RNA. HIV-2 chimeras with only the HIV-1 NC domain also exhibited the ability to encapsidate HIV-1 RNA, although at a lower level than when both NC and p2 were present (Kaye & Lever, 1998). HIV-1 and HIV-2 NC proteins have a high level of amino acid sequence identity (60%), as well as many conservative substitutions elsewhere in the protein, while within the p2 domain they share only 35% amino acid identity, supporting the hypothesis that selective recognition may be found in other regions of the Gag precursor (Kaye & Lever, 1998). The identity between the NC domains of HIV-1 and HIV-2 is likely due to the mechanistic binding of the NC domain to the encapsidation signal.

The NMR solution structure of HIV-1 NC protein with SL3 has been determined and provides structural insight into the genome recognition event. The basic residues, arginine and lysine, at positions 3 and 10 of NC, form a 3_10 helix that interacts with the RNA major groove (De Guzman et al., 1998). The two zinc-binding motifs of HIV-1 NC form hydrogen bonds to the exposed guanosines on the SL3 loop, providing the main interaction between the NC and encapsidation signal (Fig. 5) (De Guzman et al., 1998). Structurally, the SL3 NC–RNA complex differs from other characterized protein–RNA complexes in which purine-purine base pairs open up the major groove to the insertion of alpha helices or beta sheets. In this complex, a kink in the RNA background is created, producing a widening of the major groove which allows penetration of the Lys-Arg helix (Fig. 5) (De Guzman et al., 1998). The association between the NC and SL3 is further stabilized by additional intra- and intermolecular interactions between the amino acids of the Cys-His motif and nucleic acids of the SL3 loop.

**HTLV-BLV**

The Gag precursor protein of HTLV-BLV group viruses is a 58 kDa polyprotein with MA (p19), CA (p24) and NC (p15 for HTLV, p12 for BLV) domains (Fig. 3). The NC protein of these viruses contains two zinc finger domains and has several conserved basic amino acid residues. While the NC domain is likely to have a major role in genome recognition and RNA encapsidation, evidence in the literature implicates the MA protein of BLV in these events. In particular, it has been previously determined, using an RNA gel mobility shift assay,
that the BLV MA protein specifically binds to RNAs representing the 5' end of the BLV RNA genome (Katoh et al., 1991, 1993). By cDNA mapping studies, it was found that the MA protein specifically bound to two different regions of the RNA. The first RNA region contains the DLS (Katoh et al., 1991, 1993). The second RNA region was at the 5' end of the gag gene, which is the location of the primary encapsidation signal for BLV (Mansky & Wisniewski, 1998). The BLV MA cleavage precursor p15 was found to have the determinants for specific RNA recognition, while further processing of p15 to the mature MA protein, p10, abrogates its ability to bind with specificity to RNAs containing the packaging signal (Katoh et al., 1991). The BLV NC protein was found to contain only non-specific RNA-binding activity (Katoh et al., 1991). These observations indicate that BLV MA in specific RNA binding of a region that overlaps with the primary BLV encapsidation signal. These data provide support for the hypothesis that the MA domain is involved in the BLV RNA encapsidation process.

**Spumavirus**

HFV encodes a 78 kDa Gag polyprotein that has MA, CA and NC domains (Linial, 1999; Pfrepper et al., 1999). Unique to the spumaviruses is the fact that the major homology region in the CA domain and the Cys-His boxes in the NC domain are not present in Gag. In the NC domain, there are three glycine-arginine-rich domains located at the carboxyl-terminal end of the polyprotein (Fig. 3) (Yu et al., 1996). The Gag protein does not appear to be efficiently cleaved and assays of extracellular proteins show two predominant forms of Gag, 78 and 74 kDa (Ensle et al., 1997; Linial, 1999; Pfrepper et al., 1999). This indicates that Gag is not cleaved by the viral protease into the MA, CA and NC polypeptides in mature virus particles. The cleavage of the 78 kDa protein to the 74 kDa protein requires the viral protease and is needed for virus infectivity (Ensle et al., 1997; Linial, 1999). The absence of a mature NC protein in foamy virus particles is notable, and may correlate with the observation of full-length viral DNA in particles. Thus, foamy viruses provide some clear evidence that full-length Gag is the molecule that recognizes the viral genomic RNA and initiates the assembly process. Once encapsidated, the RNA is reverse-transcribed to DNA and Gag is not processed into MA, CA and NC polypeptides. The NC domain of Gag is likely to be involved in foamy virus genome recognition, but this has not been studied in great detail. Alignment of the HFV NC domain with other complex retroviruses does not indicate significant homologies.

**Beyond genome recognition: complete encapsidation of the viral RNA and virus assembly**

A first step in virus assembly is the protein–protein interactions that occur between Gag polyprotein molecules (Fig. 1). The NC domain of Gag contains the sequences involved in these interactions (called the I domain). One explanation is that the basic amino acid residues in the NC domain (and perhaps the basic amino acid residues in the BLV MA) interact with the viral RNA, which acts as a scaffold for further binding and packing of Gag polyproteins. Recent analysis has indicated that mutating basic residues in NC to alanine impairs non-specific RNA-binding activity and negatively affects the ability of Gag to act as a polymerase and assemble particles (Cimarelli et al., 2000). When viral RNA is not available, cellular RNA can act as an RNA scaffold (Swanstrom & Wills, 1997). In this model, genome recognition would first occur, followed by subsequent interactions between Gag and the viral RNA to completely encapsidate the RNA. Gag–viral RNA interactions would therefore function in two ways for specific and efficient viral RNA encapsidation. First is the genome recognition step of the cis-acting encapsidation signal. Second is the internalization of the RNA molecules into the immature cores by non-specific RNA binding. The NC domain of the Gag polyprotein is the major RNA-binding domain, non-specifically, and in many cases, specifically.

The basic residues of the HIV-1 MA have been shown to contribute to Gag–Gag interaction in the presence of RNA and the absence of the NC domain (Burniston et al., 1999). This indicates that these residues have the potential to interact with RNA (Lochrie et al., 1997) and that MA–RNA interaction may play an important role in complete RNA encapsidation and virus particle assembly. Specific to HTLV-BLV group viruses, the MA domain of BLV Gag has been implicated in specific binding with BLV RNA (Katoh et al., 1991, 1993), and could therefore also play a specific role in the genome recognition step, or at least in subsequent RNA encapsidation. The presence of genomic RNA appears to enhance the multimerization of HIV-1 Gag, and it is likely that Gag multimerization is important in the RNA encapsidation process (Morikawa et al., 1999, 2000).

HIV-1 MA has been reported to have a nuclear export signal (located in the N-terminal end of MA) which could target both Gag and the genomic RNA to the cell membrane (Fig. 1) (Dupont et al., 1999). The cytoskeleton may play an important role in transporting the Gag–viral RNA complex to the cell membrane. In particular, the actin cytoskeleton has been found to interact with HIV-1 Gag through the NC domain (Liu et al., 1999; Wilk et al., 1999).

**Other issues related to genome recognition, RNA encapsidation and virus assembly**

RNA dimerization has been suggested to be required for encapsidation of viral genomic RNA, but a direct prerequisite of dimerization in order to have RNA encapsidation has not been experimentally verified (Harrison et al., 1998). Also, it has been suggested that RNA dimerization could interfere with translation, but again there is no evidence to support such a mechanism of regulation (Berkowitz et al., 1996). The MA
domain of Gag may play a role in the formation and/or stabilization of RNA dimers (Parent et al., 2000).

The mechanisms behind how viral genomic RNA is packaged, preferentially to that of spliced viral RNAs among the complex retroviruses, is not entirely clear. While for most retroviruses the primary encapsidation signal is found only in the unspliced genomic RNA, many of the spliced viral mRNAs contain secondary encapsidation signals. For HIV-1, spliced mRNAs can be encapsidated to the order of 10% of the levels of genomic RNA (Katz et al., 1986; Luban & Goff, 1994). Why these sequences are located in regions of the RNA that are present in the spliced RNA is not clear. The role of packaging the spliced viral RNAs in the retrovirus life-cycle is unknown.

Little is known regarding variables that determine whether newly synthesized viral unspliced RNA will be encapsidated or translated. It is possible that there is no regulation, and that all RNAs are competent for both translation and packaging. In this situation, it is likely that the viral RNAs would be initially translated until enough Gag polyprotein had been made to initiate the process of RNA encapsidation. However, it is plausible that there may be some discrimination between these functions of the unspliced viral RNA.

Antiviral drugs

The NC protein has been viewed as an attractive general target for the development of antiretroviral compounds (particularly against HIV) because of the roles of NC in both early and late stages of the virus life-cycle, including genome recognition and RNA encapsidation. The ability of NC protein to act as a nucleic acid chaperone also makes it an attractive antiviral drug target (Rein et al., 1998). The highly conserved nature of the CCHC motif provides even further rationale for targeting antiviral compounds to NC protein. The role of the NC zinc fingers in the interaction of NC with viral nucleic acid provides a unique viral RNA–protein interaction against which to target antiviral drugs. Cellular proteins typically contain CCHH or CCCC motifs, which suggests that drugs targeted against the CCHC motif would not affect the function of cellular proteins (Berg & Shi, 1996). Many of the compounds currently being studied interact with the zinc fingers, causing the zinc to be ejected and thereby inactivating NC function. The inactivation of NC is typically irreversible because of the formation of inter- and intramolecular disulfide cross-links between the Cys residues that form the zinc finger after ejection of the zinc molecule. One promising compound that inactivates virus infectivity by covalently modifying the zinc finger motifs is 2,2'-dithiodipyridine (Aldrithiol-2) (Arthur et al., 1998; Ott et al., 1998). This compound has been found to influence virus replication in vivo in mice and has been used for preparation and testing of whole killed-particle vaccines. Another compound that has received attention is azodicarbonamide (this drug is also targeted to the HIV-1 NC CCHC motifs), which was the first NC inhibitor to progress to human clinical trials (Rice et al., 1997). Azodicarbonamide has also been shown to be an immunosuppressive agent (Tassignon et al., 1999). Two zinc-ejecting compounds that have shown some promise as antivirals in cell culture for HIV-1 are cystamine and cysteamine (McDonnell et al., 1997). Structural information of the HIV NC–SL3 interaction should allow for further development of other compounds that are targeted to block viral RNA–HIV-1 NC interactions (De Guzman et al., 1998).

Summary

The initiation of complex retrovirus assembly involves a recognition event of the genome-length retroviral RNA by the viral Gag protein. The 5′ end of the viral genome and at least the NC domain of the Gag polyprotein are involved in this RNA–protein recognition event. Further, structure–function analysis of encapsidation signals and of the Gag polyprotein will help in elucidating the precise interactions. There are many similarities between the simple and complex retroviruses regarding these processes, and investigations of simple retroviruses will greatly aid in defining the underlying mechanisms of these steps with complex retroviruses. However, the complex nature of the relatively large genomic RNA and the Gag polyprotein predict that determination of the precise mechanisms involved in genome recognition, RNA encapsidation and the initiation of virus assembly will provide a formidable challenge to investigators studying these steps in the virus life-cycle.

We thank Marshall Williams and Kendra Norris for helpful comments and discussion of the manuscript. Research in our laboratory is supported by the Public Health Service (GM56615), the American Cancer Society, and the Ohio Cancer Research Associates.

References


