HOW RETROVIRUSES SELECT THEIR GENOMES

Victoria D'Souza and Michael F. Summers

Abstract | As retroviruses assemble in infected cells, two copies of their full-length, unspliced RNA genomes are selected for packaging from a cellular milieu that contains a substantial excess of non-viral and spliced viral RNAs. Understanding the molecular details of genome packaging is important for the development of new antiviral strategies and to enhance the efficacy of retroviral vectors used in human gene therapy. Recent studies of viral RNA structure *in vitro* and *in vivo* and high-resolution studies of RNA fragments and protein–RNA complexes are helping to unravel the mechanism of genome packaging and providing the first glimpses of the initial stages of retrovirus assembly.

GERM LINE
The cells that are in a direct mitotic line of descent from the zygote to its gametes, as distinct from somatic cells.

Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA.
Correspondence to M.F.S.
e-mail: summers@hhmi.umbc.edu
doi:10.1038/nrmicro1210

Retroviruses are a diverse family of RNA viruses that infect animals and cause various diseases, including leukaemia, tumours, neurological disorders and AIDS. They can incorporate modified cellular genes during replication that confer tumorigenicity, induce neoplastic transformations upon integration of the proviral DNA into or near important cellular genes, interfere with normal cellular functions or induce premature cell death[1]. Approximately 1% of the human genome consists of human endogenous retrovirus sequences that seem to have resulted from early infections and incorporation of viral sequences into the GERM LINE[2]. In addition, some retroviruses have been engineered to function as vectors for the delivery of corrective human genes, and vectors derived from the Moloney murine leukaemia virus (MoMuLV) have been used for the treatment of severe combined immunodeficiency.

All retroviruses, except the spumaretroviruses, contain two copies of their positive-strand RNA genomes, both of which are important for replication (FIG. 1). The genomes seem to exist as weak dimers in most immature particles, and are converted during proteolytic viral maturation to more stable, non-covalently linked dimers[3–5] that have a structural role in mature viruses[6]. During the early phase of replication, the diploid genome serves as a template for reverse transcription, allowing strand transfer to occur at otherwise deleterious breakpoints and promoting genetic diversity and the evolution of strains that are drug resistant and strains that can evade the immune response[7].

Retroviral genomes are selected for packaging from a cellular pool that contains a substantial excess of cellular and spliced viral mRNAs[8]. Extensive genetic and virological studies indicate that selection is mediated by interactions between the nucleocapsid (NC) domains of the assembling viral Gag polyproteins and segments of the viral genome, called Ψ-sites, which are located within the 5′-untranslated region (UTR) and/or upstream coding regions of the *gag* gene (for reviews, see REFS 8–13). Recently, information regarding RNA secondary structures and long-range base pairings that are important for packaging have been obtained from a combination of computational, phylogenetic and experimental studies, and high-resolution structures have been determined for several viral RNA fragments and NC–RNA complexes. The collective findings are providing new insights into the structural determinants and molecular mechanisms of retroviral genome packaging, which is the subject of this review.

The role of the Gag protein

The NC domain of Gag directs genome packaging. Chimeric human immunodeficiency virus type 1 (HIV-1) virions that contain the MoMuLV NC domain preferentially package MoMuLV RNA[14], and MoMuLV chimaeras that contain the HIV-1 NC domain preferentially package the HIV-1 genome[15]. Furthermore, the chimeric MoMuLV particles that contain HIV-1 NC preferentially package the full-length HIV-1 genome and ignore spliced HIV-1 mRNAs, indicating that NC is responsible for discriminating between spliced and unspliced viral RNAs. Similar results have been obtained for Rous sarcoma virus (RSV) and MoMuLV[16]. In some cases, retroviruses of a given genus can package each other's genomes, indicating that evolutionarily related retroviruses might use similar protein- and RNA-recognition elements[17–19]. In other cases, non-reciprocal packaging has been observed; for example, HIV-1 can efficiently encapsidate RNA that contains the HIV-2 packaging signal, but not vice versa[20]. A few studies have indicated that other elements in Gag might be important for genome selection[21,22], although with RSV, Gag constructs that contain deletions of the matrix, capsid and protease domains are all capable of efficiently packaging the viral genome[23].

Zinc knuckles of NC domains are essential for genome packaging. The NC domains of all retroviruses, except the spumaretroviruses, contain one or two conserved CCHC arrays ($C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$; C = Cys, H = His, X = variable amino acid)[24] that bind zinc with high affinity[25], are populated with zinc in intact virions[26] and are required for high-affinity NC–RNA interactions *in vitro*. Mutations that inhibit zinc binding[27–29], or treatment of viruses with agents that eject zinc[30], result in non-infectious virus particles that are

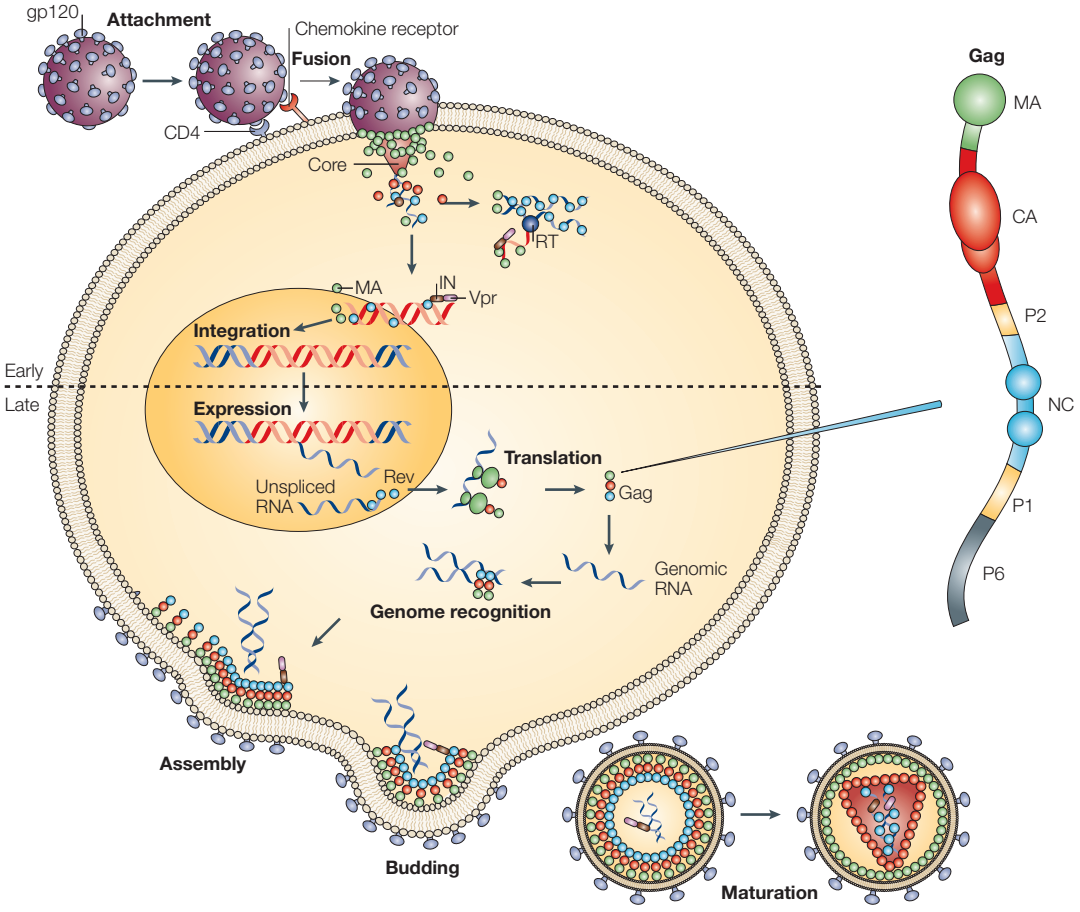

Figure 1 | **General features of the orthoretroviral replication cycle.** During the early phase of replication (upper portion), envelope glycoproteins on the surface of the virus bind to CD4 (blue) and chemokine (red) receptors on the cell surface, triggering fusion of the viral and cellular membranes. During or after uncoating of the core (red), the viral RNA genome is reverse-transcribed to proviral DNA, which is transported to the nucleus and integrated into the cellular DNA by the viral integrase enzyme. Transcription during the late phase produces the full-length viral RNA, which is capped at the 5′ end and polyadenylated at the 3′ end by the cellular machinery and transported to the ribosome. Translation of the unspliced RNA produces the viral Gag and, through an occasional read-through, Gag-Pol proteins, which assemble and bud to form immature particles. The actual site of initial genome recognition and binding (shown in the cytosol) is not known. Proteolytic cleavage of Gag by the viral protease produces the matrix (MA, green), capsid (CA, red) and nucleocapsid (NC, blue) proteins, which rearrange to form the mature particle. The domain structure of the HIV-1 Gag protein is also shown. Gag, group-specific antigen protein; IN, integrase; Pol, polymerase; Rev, regulator of viral protein expression; RT, reverse transcriptase; Vpr, viral propagation.

unable to replicate. MoMuLV particles containing mutant NC proteins that do not bind zinc are unable to efficiently package full-length genomes, and instead package detectable quantities of subgenomic mRNAs, indicating that the zinc knuckles help to discriminate between spliced and unspliced viral mRNAs[28,29].

Structures have been determined by NMR for either the zinc knuckles or the intact NC proteins (or both) of eight different retroviruses, and a representative structure of the HIV-1 NC protein[26,31] is shown in FIG. 2a. All CCHC zinc knuckles form a metal-coordinating reverse turn that is structurally similar to those of the iron-binding domains of rubredoxins and is stabilized by NH–S hydrogen bonds[32] (FIG. 2b). The N- and C-terminal tails are conformationally labile in HIV-1 NC[26] and in a larger Gag fragment that includes the C-terminal domain of the capsid protein[33], with zinc knuckles behaving like transiently interacting 'beads on a string'[34] (FIG. 2a). Most retroviral zinc knuckles are highly basic and contain a hydrophobic cleft on the surface of the miniglobular domain (FIG. 2c).

RNA elements that promote genome packaging

Regions of retroviral genomes that are important for packaging have been identified for several retroviruses, including RSV[35], MoMuLV[36], spleen necrosis virus[37], HIV-1 (REF. 38), bovine leukaemia virus[39], feline immunodeficiency virus[40], HIV-2 (REF. 41) and Mason–Pfizer monkey virus[42]. In some cases, relatively small fragments of the 5′-UTR are capable of directing RNA packaging[35,43–45]. However, the most efficient packaging is generally achieved for RNAs that contain relatively large portions of the 5′-UTR. A few studies have indicated that elements in downstream coding regions of viral genomes might promote packaging[46,47]. The 5′-UTR also contains U5 control regions that recruit cellular proteins required to initiate ribosomal translation, a specific tRNA primer-binding site (PBS) that is required for initiation of reverse transcription, and a regulatory region that helps control transcription. Although virions incorporate the appropriate primer during assembly, tRNA binding to the PBS does not seem to be required for genome packaging[48]. As indicated above, the 5′-UTR also generally contains the 5′ splice donor (SD) site, which participates in the generation of spliced viral transcripts that encode for accessory and/or envelope proteins. Except for RSV, the 5′-SD sites are located upstream of, or overlap with, elements that are required for packaging[33,35,49]. The packaging elements are therefore removed or eliminated upon splicing, providing a mechanism for specifically packaging the full-length genome and excluding spliced mRNA transcripts[36].

Retroviral packaging signals also generally overlap with elements that promote RNA dimerization[11,13,50,51], and there is now considerable evidence from studies of several retroviruses that genome dimerization and packaging are intimately coupled (for review, see REF. 12). Electron-microscopic images of RNAs that were isolated from virions indicate the presence of relatively stable dimer-linkage sites near the 5′ ends of the viral genomes[3]. Genomes isolated from rapidly harvested MoMuLV virions are dimeric, and the stability of the dimer seems to increase upon viral maturation[5]. Mutations that are designed to inhibit RNA dimerization generally result in a packaging defect[51–54]. Reduction of viral mRNA levels in infected cells by treatment with actinomycin D results in dramatic reductions in genomic RNA packaging, but the small amount of packaged RNA is nevertheless dimeric[55]. In addition, heterologous RNAs containing the MoMuLV Ψ-site that are encapsidated with poor efficiency are also dimeric[56], and particles containing mutations in NC that package the native genome with poor efficiency also exclusively package dimers[57]. Finally, Sakuragi and co-workers showed that single viral RNA molecules that contain two copies of the 5′-UTR are efficiently packaged as monomers, providing compelling evidence that HIV-1 genome packaging is promoted by an RNA structure formed by the association of two UTRs[58,59].

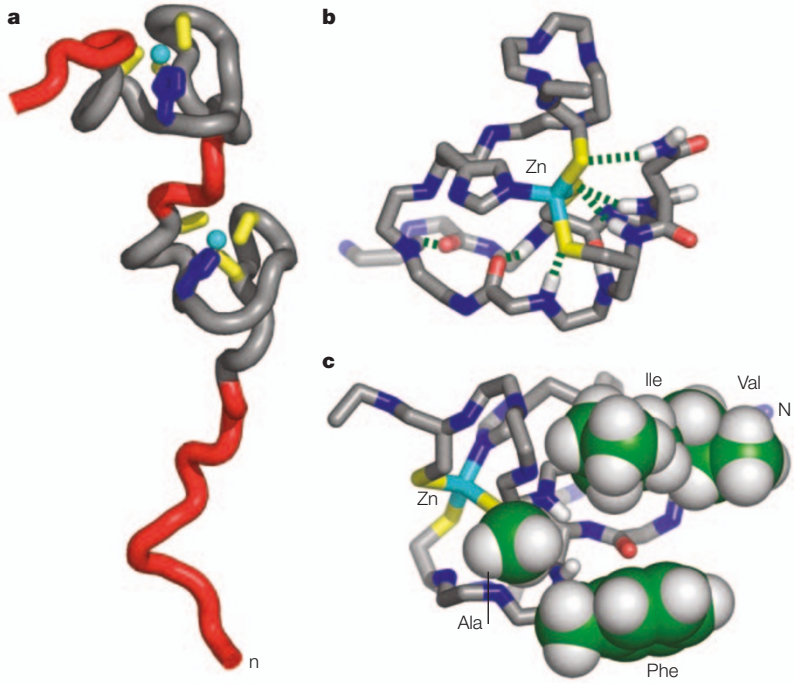

Figure 2 | **Structure of the HIV-1 nucleocapsid protein. a** | Representative NMR structure of the HIV-1 nucleocapsid protein, showing the two structurally independent zinc knuckle domains (backbone (grey), Cys (yellow) and His (blue) ligands, and the zinc atom (cyan)) and the flexible tails and linker (red). **b** | Structure of the N-terminal zinc knuckle, showing NH–S and some of the NH–O hydrogen bonds. **c** | Conserved hydrophobic residues (carbons in green, hydrogens in white) of the N-terminal HIV-1 zinc knuckle form a hydrophobic cleft. Backbone atoms at the base of the cleft that participate in hydrogen bonding with guanosine nucleotides are shown.

Genome packaging by MoMuLV

The MoMuLV Ψ-site. Nucleotides 215–565 of the MoMuLV 5′-UTR can function as an autonomous packaging signal, and this segment is referred to as the Ψ-site (FIG. 3a). Deletion of the Ψ-site abrogates packaging[60], and reinsertion of the Ψ-site near the 3′ end of the genome partially restores packaging[36].

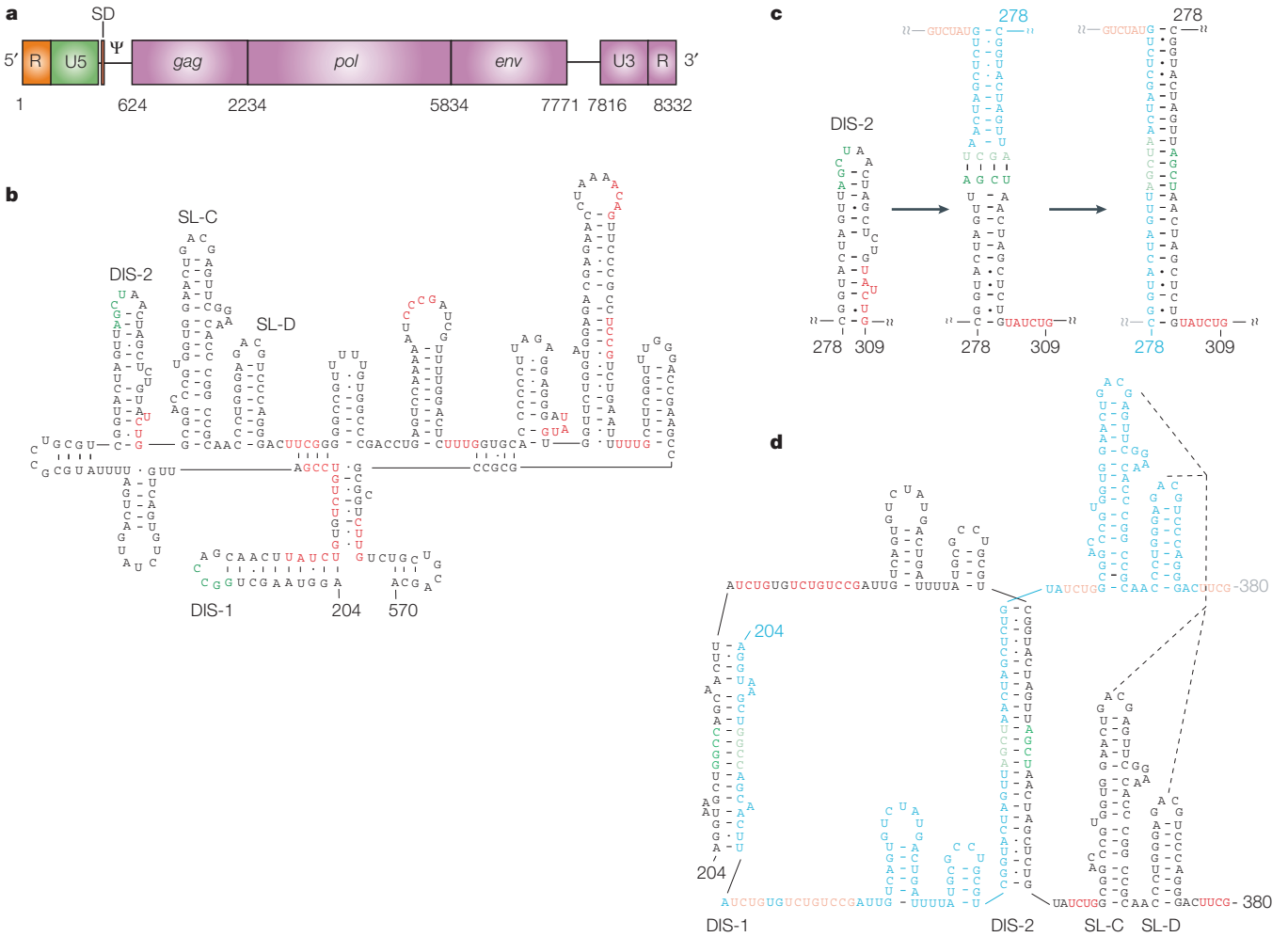

Figure 3 | **RNA secondary structures for Moloney murine leukaemia virus genome packaging. a** | Representation of the Moloney murine leukaemia virus genome, showing the relative locations of the coding and non-coding elements. **b** | Secondary structure predicted for residues 204–570 of the monomeric 5′-untranslated region (UTR). Base pairings for nucleotides 204–228 (DIS-1) and 231–555 are from REFS 69,70,81 and 64, respectively. Elements that are potentially capable of binding tightly to the nucleocapsid (NC) protein are shown in red. **c** | Potential dimerization mechanism for DIS-2 (278–309) (REFS 67,69,71). Changes in internal base pairing enable the palindromic AGCU segment (green) to form a potential kissing interaction with a second molecule (shown in lighter shading). The putative kissing intermediate then rearranges to form a duplex. Note that the NC-binding UAUCUG segment (red) is sequestered by base pairing in the hairpin species and exposed in the dimer. **d** | Potential secondary structure of residues 204–380 of the dimeric 5′-UTR, showing predicted duplex forms of DIS-1 and DIS-2 (REFS 64,69,70). UCUG and related elements that can potentially bind NC with high affinity are largely sequestered by local and long-range base pairing in the monomeric RNA and might become exposed in the dimer. DIS, dimerization initiation site; SD, splice donor site; SL, stem loop.

MoMuLV constructs that contain deletions that extend into the 5′ region of the Ψ-site are packaged with reasonable efficiency[61], and a shorter fragment that comprises ~100 nucleotides (referred to as a 'core-encapsidation signal', Ψ^{CES}) is sufficient to direct packaging of heterologous RNAs into virus-like particles[44,45]. However, packaging efficiency is influenced by residues upstream[62] and downstream[45,47,63] of the MoMuLV Ψ-site, with the most efficient packaging observed for Ψ+ RNAs that include the intact Ψ-site and ~470 additional downstream nucleotides[63].

In vitro chemical-accessibility mapping experiments involving the 5′ region of the MoMuLV genome, coupled with phylogenetic and free-energy calculations[64,65], indicate that the Ψ-site consists of a series of closely spaced stem loops (FIG. 3b). Differences in protection patterns with nuclease digestion were observed for large fragments of the MoMuLV 5′-UTR upon dimerization, indicating that dimerization-dependent changes in RNA conformation might be important for regulating splicing, reverse transcription, translation and/or genome packaging[64,65]. In particular, the monomeric Ψ-site contains two pseudo-palindromic stem loops, DIS-1 (dimerization initiation site 1) and DIS-2, both of which are important for RNA dimerization *in vitro* and encapsidation *in vivo*[66–70] (FIG. 3b). Both stem loops undergo changes in base pairing upon dimerization (FIG. 3c).

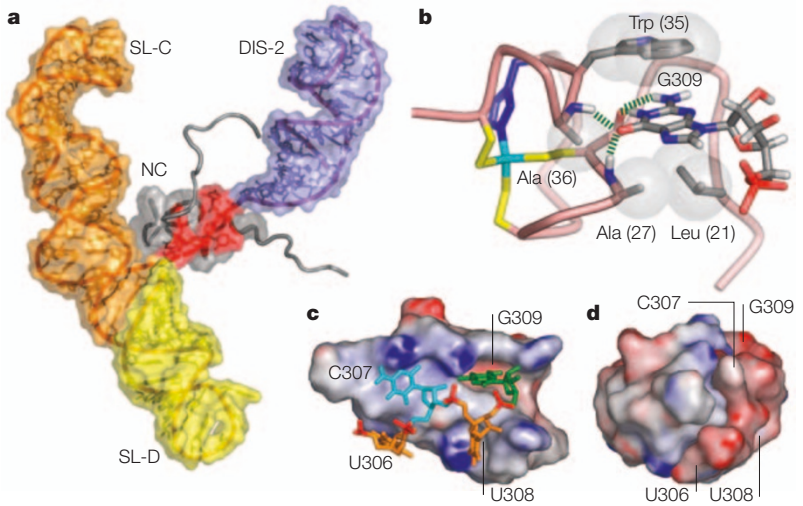

Figure 4 | **Structure of the Moloney murine leukaemia virus nucleocapsid protein bound to the core-encapsidation RNA element. a** | Structure of the Moloney murine leukaemia virus nucleocapsid (NC) protein (grey) bound to the 101-nucleotide core-encapsidation element of the Ψ-site. The RNA was mutated to remain monomeric but retain internal base pairings of the dimer (see REFS 76,81,166 for details). Stem loops DIS-2, SL-C and SL-D are shown in purple, orange and yellow, respectively. The NC zinc knuckle interacts exclusively with the UCUG element (red) that connects DIS-2 to SL-C. **b** | The exposed guanosine of the UCUG linker fits into a hydrophobic cleft and forms hydrogen bonds with backbone atoms of the zinc knuckle. **c**,**d** | The NC–RNA interface shows complementarity of both charge and shape. Positive and negative electrostatic potentials are shown in blue and red, respectively.

Dimerization has been proposed to proceed through multiple steps, including the rate-limiting conversion of a stable, monomeric hairpin to a less stable species, in which the central palindromic GGCC (DIS-1) or AGCU (DIS-2) segments are fully exposed; formation of an intermolecular 'kissing complex', in which the tetraloop residues form intermolecular base pairs; and conversion of the kissing complex to a more stable intermolecular duplex[67,69,71] (FIG. 3c).

Stem loops SL-C and SL-D, which are part of the core-encapsidation signal[44,45], also promote efficient genome packaging and infectivity[44,72,73]. Although MoMuLV RNAs containing SL-C, SL-D and a portion of SL-B can direct RNA packaging[44], constructs containing only SL-C and SL-D do not promote packaging[74] and do not bind NC with significant affinity[75,76]. These stem loops can form 'kissing dimers' that are mediated by intermolecular hydrogen bonds between G and C residues in their conserved GACG tetraloops[76,77], and it has been suggested that stem loops SL-C and SL-D might facilitate packaging indirectly by promoting RNA dimerization[56]. Although these stem loops enhance the rate of Ψ-RNA dimerization, they do not seem to influence the stability of the dimer[68]. Other RNA elements within the 5′-UTR and coding regions of the genome also probably either promote or participate in dimerization[64,78,79], and recent phylogenetic studies indicate that additional long-range interactions might occur between residues immediately downstream of the U5 region and those overlapping the *gag* initiation codon[80].

3D structural studies and packaging mechanism. The structure of the MoMuLV NC protein bound to a mutant form of the core-encapsidation signal that was engineered to remain monomeric but retain base pairing of the dimer ($m\Psi^{CES}$) has been determined by NMR[81] (FIG. 4a). The single zinc knuckle binds specifically to UCUG residues of the linker that connects DIS-2 to SL-C. Binding is promoted by interactions with a guanosine moiety that inserts into a hydrophobic pocket and forms hydrogen bonds with backbone NH groups and carbonyl groups of the zinc knuckle (FIG. 4b) and also by shape and charge complementarity (FIG. 4c,d). NC does not bind RNAs with secondary structures that are formed by the monomeric form of the native Ψ-site, in which the UCUG nucleotides are sequestered by base pairing[81]. Stem loop DIS-1, which also contains a UCUG element, shows essentially identical, dimerization-dependent structural changes and NC-binding behaviour[81]. These findings indicate that high-affinity NC binding to the MoMuLV Ψ-site is regulated by an RNA conformational switch, in which NC-binding UCUG elements are sequestered by base pairing in the monomeric RNA and become exposed to promote packaging upon dimerization (FIG. 5).

Thirteen UCUG sequences are present in the 418-nucleotide segment between the PBS and Gag start sites of the MoMuLV genome, accounting for more than 12% of the nucleotides in this region. By comparison, UCUG sequences occur at a lower frequency of one in 225 nucleotides in the coding and long-terminal-repeat regions of the MoMuLV genome. Five of the UCUG segments in the 5′-UTR are located between Ψ and the Gag initiation codon, within the Ψ+ segment that is known to enhance packaging efficiency[63]. By contrast, no UCUG sequences are present between the PBS and Gag initiation codon of HIV-1 (a representative lentivirus), and there are fewer than three UCUG sequences in representative α-retrovirus (RSV), β-retrovirus (mouse mammary tumour virus) and δ-retrovirus (human T-lymphotropic virus (HTLV)/bovine leukaemia virus) genomes[81]. The MoMuLV 5′-UTR contains additional segments with sequences similar to UCUG, including ACAG, UUUG and UCCG, which also seem to be sequestered by base pairing in the monomeric RNA, but are exposed in the dimer and are capable of binding NC with an affinity approaching that of UCUG[82].

The collective data indicate that MoMuLV genome packaging might proceed through the following mechanism. First, the initially synthesized genome might adopt a compact monomeric structure similar to that shown in FIG. 3b. This form of the genome would not be selected for packaging because most or all of the UCUG and related sequences are sequestered by base pairing and unable to bind NC with high affinity. Subsequent dimerization, possibly promoted by the chaperone activity[83] of the NC domain of Gag, would expose these sites for subsequent high-affinity binding. The high abundance and proximity of exposed UCUG and related elements within the dimeric 5′-UTR might also facilitate Gag–Gag interactions (FIG. 6).

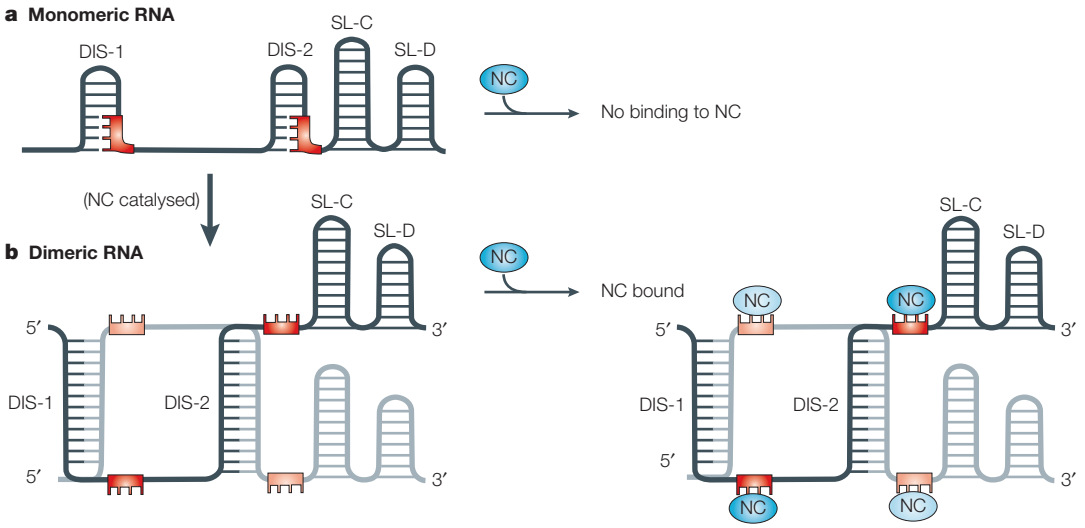

Figure 5 | **Proposed model for specific packaging of a dimeric genome.** Potential nucleocapsid (NC)-binding sites (red) are sequestered by base pairing in the monomeric RNA and do not bind NC (blue). Conversion to the dimer, induced by the chaperone activity of Gag or NC, exposes these elements for high-affinity NC binding. DIS, dimerization initiation site; Gag, group-specific antigen protein; SL, stem loop.

Genome packaging by HIV-1

The HIV-1 Ψ-site. Deletion mutagenesis studies have shown that the ~120 nucleotides located just upstream of the *gag* start codon in HIV-1 are required for efficient genome packaging[38,54,84–89], and this segment has often been referred to as the HIV-1 Ψ-site (FIG. 7a). However, more recent deletion studies indicate that nucleotides within the *gag*-coding[90–92], U5, PBS[93,94] and regulatory (which includes the *trans*-acting responsive element (TAR))[95] regions of the genome are also important for packaging. Efforts to identify a relatively short segment of the HIV-1 genome that can independently direct the packaging of heterologous RNAs have led to mixed and, in some cases, controvertible results. Hayashi and co-workers reported that a 46-nucleotide fragment of the HIV-1 5′-UTR can promote heterologous RNA packaging[96], but Goff and co-workers were unable to observe packaging for large RNA constructs beginning as early as nucleotide 20 (within TAR) and spanning up to nucleotide 506 (which includes 170 nucleotides of *gag*)[97]. On the other hand, a 1,015-nucleotide fragment of the HIV-1 genome that starts with the most 5′ nucleotide can efficiently direct heterologous RNA packaging[98]. It therefore seems that the HIV-1 packaging signal comprises the entire 5′-UTR and up to 300 nucleotides of *gag*[98].

3D structural studies of genome packaging. The secondary structure of the HIV-1 5′-UTR has been probed by chemical accessibility and mutagenesis experiments and analysed by phylogenetic and free-energy calculations[43,54,80,87,88,96,99–102], and is believed to consist of a series of closely spaced stem loops that have independent and, in some cases, possibly overlapping functions (FIG. 7b); although, as discussed below, alternate structures have also recently been proposed. The HIV-1 Gag and NC proteins can bind to fragments of the 5′-UTR *in vitro* with high affinity[96,101,103–107]. Stem loop SL1 contains the primary dimer-initiation site, a GC-rich loop required for efficient genome packaging that promotes dimerization through formation of a kissing intermediate[53,100,108–111]. NC does not bind with high affinity to SL1 (REF. 112), but can catalyse the conversion of SL1 from a kinetically trapped kissing dimer to a thermodynamically more stable duplex species[113,114]. The mechanism of this process seems to involve NC-mediated destabilization of the stem of SL1 (REF. 115), which accelerates strand exchange between residues of the stems without disrupting kissing interactions in the loops[115].

Stem loop SL2 contains the major splice-donor site and is capable of binding HIV-1 NC with high affinity[116]. This stem loop is destroyed upon splicing, indicating a possible mechanism for discriminating between spliced and unspliced viral mRNAs[116]. In the NMR structure of the NC–SL2 complex, the zinc knuckles of NC interact with exposed guanosines of the tetraloop in a manner similar to that observed in a zinc-knuckle–DNA complex[117] (FIG. 7c).

Stem loop SL3 also binds NC tightly[101,107], and in certain contexts seems capable of independently directing the packaging of heterologous RNAs[96]. On the one hand, deletion of SL3 from the HIV-1 Ψ-site, or mutations in SL3 that disrupt base pairing, considerably reduce genome packaging[54,102]. Packaging efficiency can be largely restored by substituting SL3 with an NC-binding RNA segment identified by SELEX (systematic evolution of ligands by exponential enrichment)[118]. On the other hand, an SL3 mutant containing a GCUA tetraloop is apparently packaged with near-wild-type efficiency[119]. The NMR structure of the NC–SL3 complex has shown that, as observed for NC–SL2, exposed guanosines of the GGAG tetraloop are inserted into the hydrophobic clefts of the zinc knuckles,

which themselves participate in extensive inter-knuckle contacts[106]. In addition, the N-terminal tail of NC forms a 3_{10} helix that binds to the minor groove of the SL3 stem (FIG. 7d). It is difficult to reconcile these findings with the observation that the SL3-tetraloop residues do not contribute significantly to genome packaging, especially as HIV-1 NC does not bind tightly to other stem-loop structures examined[112,120–122]. It is conceivable that SL2 and SL3 have redundant functions in genome recognition by interacting with different NC domains of assembling Gag molecules[54,87,88,102,116].

Oligoribonucleotides with sequences of SL4 contain unstable stems and a relatively stable GNRA-type (G–any base–purine–A) tetraloop[120]. NC binds less tightly to SL4 compared with SL2 and SL3[120,123], possibly owing to the unfavourable energetics that are required to disrupt the folding of the GNRA tetraloop. Although it was suggested that these essential residues might participate in long range RNA–RNA interactions, as is typical of GNRA tetraloops[120], recent studies of larger fragments of the 5′-UTR indicate that residues of SL4 might adopt an entirely different conformation (see below). Therefore, the biological relevance of the putative SL4 stem-loop structure is currently uncertain.

RNA conformational changes and implications for packaging. Fragments of the 5′ leader that include G1 (guanosine 1) through to G290 can adopt two distinct monomeric conformations that migrate at different rates on native polyacrylamide gels[124]. The faster migrating CONFORMER is obtained at low RNA concentrations and in the absence of Mg^{2+} or NC. Addition of Mg^{2+} shifts the equilibrium to the slowly migrating species. NC efficiently promotes dimerization of the slowly migrating species, but not the fast-migrating species, which led to the suggestion that this apparent RNA conformational switch might regulate a biological function[124]. Secondary-structure predictions and chemical-probing studies indicated that the fast-migrating species forms a rod-like structure (termed LDI, for long distance interaction), in which the residues of the DIS and poly(A) motifs are base paired. The slowly migrating species was proposed to adopt a secondary structure containing multiple hairpins, consistent with previous structure predictions, and is referred to as the 'branched multiple hairpin' (BMH) conformer[125]. The authors suggested that the BMH species forms initially upon transcription, but converts rapidly to the LDI conformer. Long-distance base pairing between the poly(A) and DIS segments in the putative LDI conformer were proposed to inhibit dimerization, and the binding of NC was proposed to revert the LDI structure back to the BMH conformation, allowing dimerization through the DIS site of stem loop SL1 (REF. 125). Mutations designed to disrupt base pairing in the putative LDI stem did not significantly affect the LDI/BMH equilibrium, and binding of the tRNA primer favoured the BMH conformer to some extent[126]. Predicted thermodynamic stabilities of LDI and BMH conformers formed by mutant RNAs were found to correlate with dimerization and packaging properties[127].

CONFORMER
One of many conformations that a protein or other molecule can adopt. For each molecule, the usual conformation that is adopted will be at energy minima.

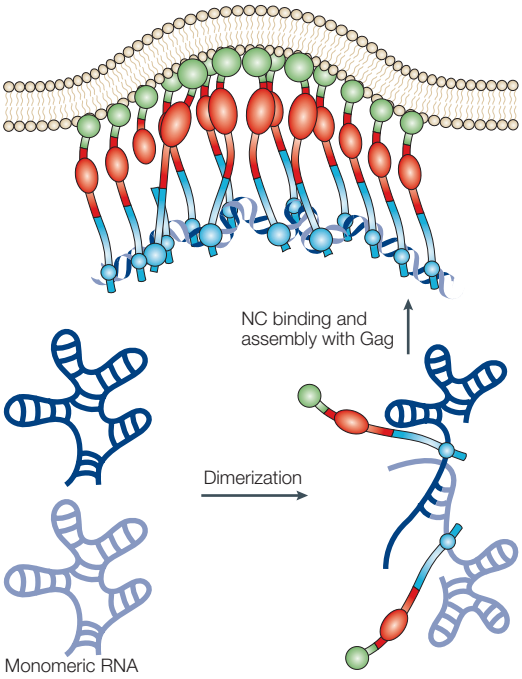

Figure 6 | **Model for dimeric genome recognition and the early stage of retrovirus assembly.** Nucleocapsid (NC)-binding sites that are sequestered by base pairing in the monomeric RNA become exposed upon dimerization (as observed for Moloney murine leukaemia virus; see also FIG. 5). The proximity of NC-binding sites might promote Gag–Gag interactions within the structurally defined Gag–RNA complex that could serve as a scaffold for efficient virus assembly. Gag, group-specific antigen protein.

Although the presence of fast- and slow-migrating monomeric 5′-UTR RNA conformers has been confirmed[80,124], evidence for the biological relevance of the putative LDI structure and the proposed LDI/BMH conformational switch remains to be obtained. The fact that the LDI structure is only obtained under dilute conditions and in the absence of Mg^{2+} raises questions about how such a structure might be generated within the ionic environment of the cell. The extended, rod-like secondary structure proposed for the LDI species seems intuitively inconsistent with a compactly folded tertiary structure, as implicated by the unusually fast migration rate on native polyacrylamide gels. In addition, recent phylogenetic analyses involving 20 divergent HIV-1 sequences are consistent with the BMH structure, but do not support the secondary structure proposed for the LDI conformer[80]. Chemical modification experiments involving HIV-1 RNA in infected cells and intact virions are generally consistent with the BMH conformer but not the LDI species[128]. If present, the LDI structure would either have to comprise less than 5–10% of the total RNA, or have an extremely short lifetime[128].

Ehresmann and co-workers obtained evidence that nucleotides just downstream of the poly(A) signal can interact with residues within the *gag*-coding region, forming a long-range pseudoknot structure[129]. In addition, Abbink and Berkhout obtained evidence

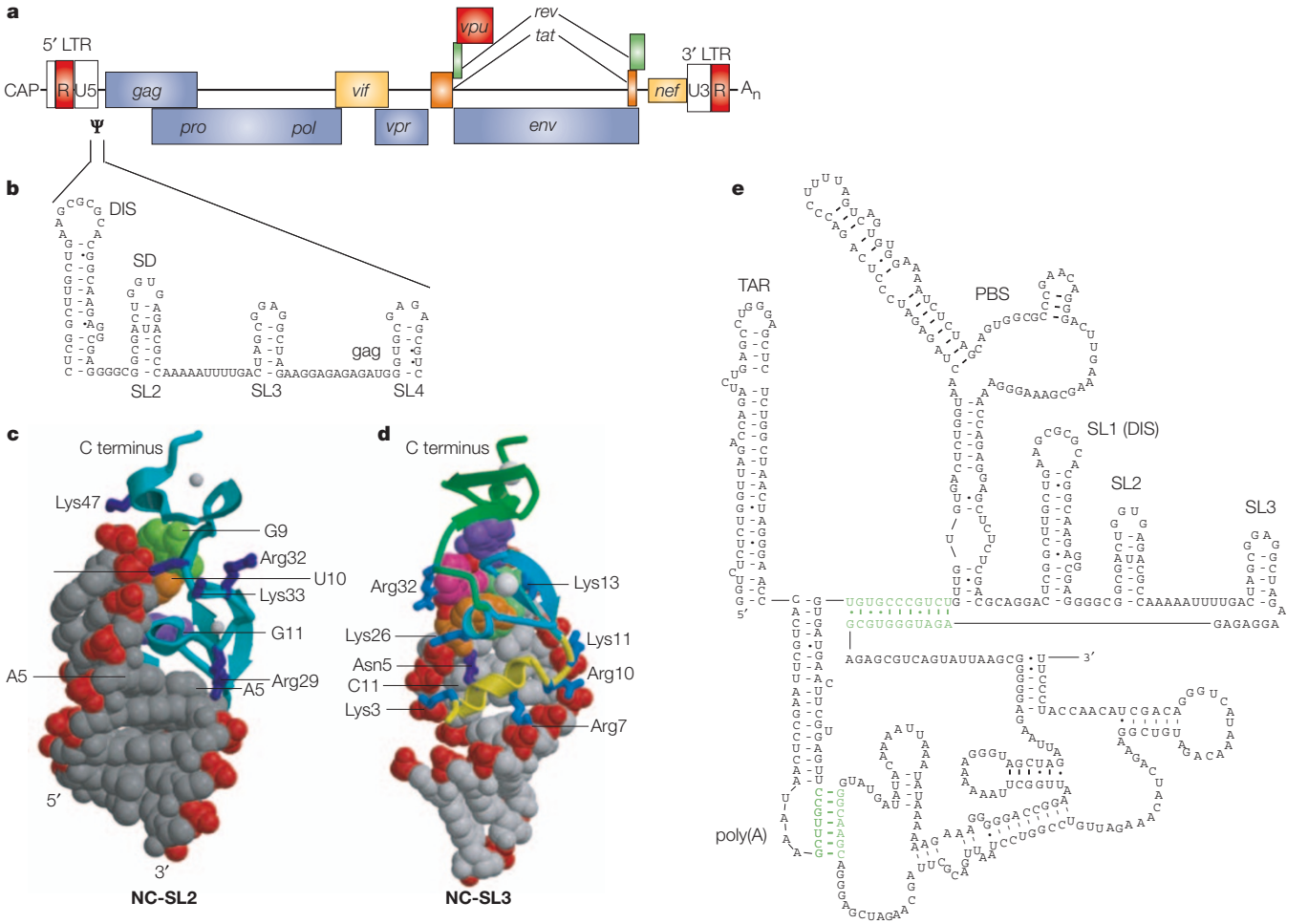

Figure 7 | **RNA elements and protein–RNA complexes associated with HIV-1 genome packaging.** **a** | Representation of the HIV-1 genome, showing gene structure, splicing patterns and location of the Ψ-site. **b** | Secondary structure predicted for the HIV-1 Ψ-site, showing the location of the dimerization initiation site (DIS), major splice-donor site (SD), and *gag* (group-specific antigen) initiation codon. **c**, **d** | NMR structures of the complexes formed between the HIV-1 nucleocapsid (NC) protein and stem loops SL2 (REF. 116) and SL3 (REF. 106), respectively. **e** | Secondary structure predicted for the intact HIV-1 5′ untranslated region, showing potential U5–AUG and poly(A)–*gag* long-range interactions (green) identified from *in vitro* assays and phylogenetic analyses[80,129,130]. In this more recent model, residues that had been predicted to form stem loop SL4 form an extended conformation and participate in long-range interactions. PBS, primer-binding site; CAP, 5′ cap; LTR, long terminal repeat; TAR, *trans*-acting responsive element.

that a segment just downstream of the U5 stem loop can form base pairs with a segment that overlaps with the *gag* initiation codon[130]. Both the U5-AUG and poly(A)-pseudoknot structures are supported by more recent phylogenetic analysis[80], which further indicates that the U5–*gag* interactions might be common to evolutionarily divergent retroviruses. A secondary structure map of the 5′-UTR of HIV-1 that shows these potential long-range interactions is shown in FIG. 7e.

Recent *in vivo* protection assays were unable to confirm the long range U5–*gag* or pseudoknot base pairings observed *in vitro*[128]. In addition, the chemical reactivity of the RNA in infected cells and virions was similar, which raises the possibility that most cellular RNA might already exist in the form that ultimately becomes packaged into virions. In fact, the reactivity pattern of SL1 strongly indicates that most genomic RNA in infected cells is dimeric[128]. Although evidence against the LDI conformer is strong (a 5–10% population of this conformer would be expected to result in chemical reactivity for several nucleotides, which was not observed), the reactivity data might not rule out the presence of a minor, monomeric BMH species. The reactivities of nucleotides that participate in the U5–*gag* and pseudoknot long-range base pairs of the BMH conformer should, in principle, be suppressed. However, the presence of a small population of this conformer would result in a small suppression of an otherwise large signal, which might be difficult to detect or quantify. Therefore, the evidence against the presence of the proposed LDI structure *in vivo* seems stronger than the evidence against the long-range interactions that are observed *in vitro* and predicted from phylogenetic studies for the BMH conformer.

It is tempting to speculate that the initially synthesized RNA might adopt a monomeric BMH structure with the long-range base pairings observed *in vitro*, and that dimerization disrupts these interactions. Although this model has similarities to that described above for MoMuLV genome packaging, it does not provide a clear explanation for how the packaging signal might be coupled with dimerization. The NC-binding sites of the MoMuLV Ψ-site are sequestered in the monomeric RNA and become exposed upon dimerization, whereas the NC-binding SL2 and SL3 stem loops of HIV-1 appear intact in both the monomeric BMH (*in vitro*) and dimeric (*in vitro* and *in vivo)* forms of the viral RNA (although SL2 is predicted to adopt an alternate conformation in the putative LDI structure). It is conceivable that the HIV-1 5′-UTR contains dimerization-dependent NC-binding sites that have yet to be identified. Therefore, the molecular mechanism that regulates the specific packaging of two copies of the HIV-1 genome remains a mystery.

Other retroviruses

Many of the underlying principles observed for HIV-1 and MoMuLV have been observed for other retroviruses. HIV-2 has a 5′-UTR that is substantially longer (>50%) than that of HIV-1, and some studies indicate that nucleotides downstream of the major splice-donor site do not contribute significantly to genome packaging[41,131,132]. If true, the Ψ-site would be contained in both spliced and unspliced transcripts, which would necessitate a selection mechanism different from that observed for HIV-1 and MoMuLV[41]. Evidence for co-translational genome packaging, which would not require such a selection mechanism, was reported[132]. However, several other studies have shown that residues between SD and the *gag* initiation codon are, in fact, crucial for packaging[133,134], and that the residues upstream of SD that promote packaging also promote dimerization[135,136] (as do residues in the PBS[136,137]). Therefore, the arrangement of elements used for HIV-2 genome packaging seems generally similar to that observed for HIV-1. Monomeric LDI and BMH conformers have also been proposed for HIV-2 (REFS 138,139). As indicated above, HIV-1 Gag is capable of packaging both HIV-1 and HIV-2 vector RNAs[20], and in this regard, the HIV-1 Gag and NC proteins bind preferentially to stem loop Ψ-3 of the HIV-2 leader, which has similarities to HIV-1 stem loop SL3 (REF. 140).

Unlike mammalian retroviruses, the packaging signals of the α-retroviruses seem to reside upstream of the 5′ major splice-donor site and are present in the full-length and subgenomic mRNAs[35,49]. At present, the mechanism for distinguishing between the spliced and unspliced α-retrovirus RNAs is unknown. Linial and co-workers identified a 160-nucleotide fragment of the RSV 5′-UTR, named MΨ, that is sufficient to direct packaging of heterologous RNAs into virus-like particles[141] with an efficiency that is only 2.6-fold less than that of the native, intact genome[142]. The results of *in vivo* mutagenesis studies are generally consistent with the predicted secondary structure of MΨ[143,144], except that the mutagenesis studies indicated that the L3 loop might not be essential. More recent studies confirmed that the L3 loop is not strictly required[143] and revealed that efficient genome packaging can be directed by an even smaller 82-nucleotide segment (named μΨ)[35]. There is some controversy over whether the α-retroviruses specifically package a dimeric genome. The L3 stem loop has been implicated in dimerization[142,145], as have pseudo-palindromic sequences upstream[145] and downstream[146,147] of MΨ. Interestingly, addition of a myristoylation signal to the N terminus of RSV Gag resulted in the production of particles that contain monomeric RNA[148], and additional studies indicated that the MA (matrix) protein might directly promote dimerization or indirectly influence dimerization by controlling the cellular location of RNA selection and/or virus assembly[149].

The feline immunodeficiency virus (FIV) 5′-UTR contains a relatively short nucleotide segment between the 5′ splice-donor site and *gag* open reading frame, and unlike the other mammalian retroviruses, these nucleotides are not required for encapsidation or replication[150]. Instead, efficient genome packaging requires two separate regions of the FIV genome, including the first ~100 nucleotides of the 5′-UTR and the first ~100 nucleotides of the *gag*-coding region[40,150]. Secondary structure analyses and chemical or enzymatic probing of these segments of the FIV genome have not been reported.

The encapsidation signal of bovine leukaemia virus (BLV), a member of the δ-retrovirus genus that includes HTLV-I and HTLV-II, also seems to be discontinuous. A ~150-nucleotide segment of the BLV genome that overlaps with the *gag* start codon is essential for encapsidation (called region 1), and a 132-nucleotide segment within the *gag*-coding region facilitates efficient encapsidation (region 2)[39]. The two stem loops of region 1 are not functionally equivalent[151]. Early studies revealed that the BLV-matrix precursor protein, MA(p15) (which includes MA and residues of the Gag precursor that link MA to CA), can bind tightly and specifically to the dimeric U5-5′-*gag* region of the BLV genome[152], and it was recently shown that basic residues in both NC and MA influence genome packaging[22], raising the possibility that MA might have a direct role in dimeric genome discrimination.

Mason–Pfizer monkey virus (MPMV) is a prototypical β-retrovirus that causes an immunodeficiency disease in newborn rhesus monkeys. A 624-nucleotide segment of the MPMV genome that resides immediately downstream of the PBS is sufficient to direct efficient packaging into MPMV-derived vectors[153]. This segment is predicted to consist of a series of conserved stem loops[154], and deletion[155] and mutagenesis[42] studies indicate that the sequence, but not the structure, of the putative stem loop immediately downstream of the PBS serves as the core packaging element for this retrovirus.

In summary, for all retroviruses examined so far, major determinants for genome recognition and packaging are located in the 5′-UTR and/or the 5′ region of the *gag* open reading frame. In nearly all cases, packaging elements either overlap with, or are adjacent to, elements that promote dimerization and participate in splicing, and all have predicted secondary structures that consist of a series of relatively short and closely spaced pseudopalindromic stem loops.

Overview and future directions

Results obtained over the past two decades have provided answers to several key questions related to the mechanism of retroviral genome packaging. How do retroviruses distinguish between viral and cellular RNAs? Gag assembly *in vitro* is promoted by nucleic acids[156,157], and *in vivo* assembly can occur in the absence of the native viral genome by incorporation of equivalent amounts of cellular RNAs[6,158–160]. Although retroviruses seem capable of packaging essentially any RNA, even ribosomes[158], RNAs that contain appropriate packaging signals are efficiently enriched in assembling virions. The NC domains of Gag and the cognate RNA-packaging signals serve as the primary determinants of this enrichment and can promote RNA packaging even when supplied in separate vectors or when the NC domains and Ψ-sites are incorporated into non-native Gag and RNA constructs, respectively. Genome selection might therefore be viewed as a competitive process, in which the degree of RNA enrichment is dependent on the competitive ability of Gag to bind viral RNA versus cellular RNAs.

In some cases, RNA packaging can be promoted by relatively small regions of the 5′-UTR. For example, segments comprising fewer than ~100 nucleotides of the MoMuLV and RSV 5′-UTRs can direct RNA packaging into virus-like particles. However, these shorter fragments, which contain only a small number of high-affinity NC-binding sites, are not packaged as efficiently as RNAs that contain larger fragments of the 5′-UTR, which presumably (and in some cases certainly) contain additional high-affinity NC-binding sites. Therefore, in some cases, packaging efficiency might simply be related to the number of high-affinity NC-binding sites that are presented by the RNA. For other retroviruses, genome selection requires larger fragments of the viral genome and might be mediated by a more complex mechanism. For example, efficient packaging of the HIV-1 genome requires the intact 5′-UTR and at least a portion of the 5′ end of the *gag* gene.

Other domains of the Gag polyprotein can also contribute to genome packaging, possibly through an indirect mechanism. In particular, mutations in the MA domain can lead to the production of genome-deficient particles[22,161–163]. HIV-1 MA contains a CRM1 (chromosome region maintenance 1 or exportin)-dependent nuclear export signal (NES), and mutations in MA that block the NES result in accumulation of both Gag and the viral genome in the nucleus and the production of RNA-deficient virions[161]. Mutations and exogenous agents that block the RSV MA nuclear-localization-signal activity also lead to the production of particles that are deficient in genomic RNA, and to the accumulation of Gag and viral RNA in the nucleus[162,163]. MoMuLV Gag proteins also seem to transiently access the nucleus[164]. A mechanism consistent with these observations has been proposed, in which the nuclear localization signal of MA initially targets Gag to the nucleus for genome binding, and the NES subsequently targets the Gag–RNA complex for nuclear export[162]. This proposal remains somewhat controversial[163]. Recent studies also indicate that viral genomes are recruited from the cytosol and transported in Gag-containing endosomes[165]. Future studies to identify the sites of initial genome recognition and the mechanism for co-localization of the viral Gag proteins and RNA at sites of virus assembly are clearly warranted.

How do retroviruses discriminate between their spliced and unspliced transcripts? Except for the RSV and avian leukosis virus, elements that are crucial for packaging overlap with, or reside downstream of, the major splice-donor site. These residues are removed during splicing and are not present in the unspliced viral transcripts, which explains why the spliced mRNAs are poorly packaged. The 5′ splice site of HIV-1 resides within a stem loop (SL2) that also serves as a high-affinity NC-binding site. Splicing destroys this stem loop and disrupts the 5′-UTR, all of which seem to be required for efficient packaging.

What is the mechanism for the specific packaging of two copies of the viral genome? For MoMuLV, the monomeric form of the full-length genome seems to adopt a three-dimensional structure that sequesters some or all of the high-affinity NC-binding sites. An RNA conformational switch exposes the NC-binding sites upon dimerization, providing an attractively simple mechanism for selectively packaging a dimeric genome. The LDI/BMH RNA-switch model proposed for HIV-1 and HIV-2 might also serve as a structural RNA switch for regulating genome dimerization and packaging. However, as indicated above, the physiological relevance of the LDI conformation has yet to be firmly established. It is conceivable that additional structural studies of the intact HIV-1 5′-UTR will reveal dimerization-dependent conformational changes that influence exposure of SL2, SL3 or other NC-binding sites, as observed for MoMuLV. However, it is also possible that HIV-1, and perhaps other retroviruses, uses different mechanisms to specifically package two copies of its genome. Alternatively, as most HIV-1 RNA in infected cells seems to exist as dimers, the need for a specialized dimer-selection mechanism might be questioned. Future studies of monomeric and dimeric forms of large RNA fragments that include the intact 5′-UTR and the application of modern biophysical approaches, such as mass spectrometry and fluorescence resonance energy transfer, should provide answers to these and other open questions and offer new macroscopic views of the structures that promote genome packaging during the early stages of retrovirus assembly.

1. Coffin, J. M., Hughes, S. H. & Varmus, H. E. *Retroviruses* (Cold Spring Harbor Laboratory Press, Plainview, 1997).
A general reference on the structure, function, epidemiology and history of retroviruses.
2. Nelson, P. N. *et al.* Demystified. Human endogenous retroviruses. *J. Clin. Pathol. Mol. Pathol.* **56**, 11–18 (2002).
3. Bender, W. *et al.* High-molecular-weight RNAs of AKR, NZB and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J. Virol.* **25**, 888–896 (1978).
4. Murti, K. G., Bondurant, M. & Tereba, A. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol.* **37**, 411–419 (1981).
5. Fu, W. & Rein, A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J. Virol.* **67**, 5443–5449 (1993).
6. Muriaux, D., Mirro, J., Harvin, D. & Rein, A. RNA is a structural element in retrovirus particles. *Proc. Natl Acad. Sci. USA* **98**, 5246–5251 (2001).
In addition to functioning as the genetic material for reverse transcription, the viral RNA has a structural role in mature retroviruses.
7. Hu, W. S. & Temin, H. M. Retroviral recombination and reverse transcription. *Science* **250**, 1227–1233 (1990).
8. Berkowitz, R., Fisher, J. & Goff, S. P. RNA packaging. *Curr. Top. Microbiol. Immun.* **214**, 177–218 (1996).
9. Rein, A. Retroviral RNA packaging: a review. *Arch. Virol.* **9**, 513–522 (1994).
10. Jewell, N. A. & Mansky, L. M. In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *J. Gen. Virol.* **81**, 1889–1899 (2000).
11. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nature Rev. Microbiol.* **2**, 461–472 (2004).
12. Russell, R. S., Liang, C. & Wainberg, M. A. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **1**, 23 (2004).
13. Greatorex, J. The retroviral RNA dimer linkage: different structures may reflect different roles. *Retrovirology* **1**, 22 (2004).
14. Berkowitz, R. D., Ohagen, A., Hoglund, S. & Goff, S. P. Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging *in vivo*. *J. Virol.* **69**, 6445–6456 (1995).
The chimeric HIV-1 Gag polyprotein containing the MoMuLV NC domain can specifically package the MoMuLV genome.
15. Zhang, Y. & Barklis, E. Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J. Virol.* **69**, 5716–5722 (1995).
16. Dupraz, P. & Spahr, P.-F. Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *J. Virol.* **66**, 4662–4670 (1992).
17. Rizvi, T. A. & Panganiban, A. T. Simian immunodeficiency virus RNA is efficiently encapsidate by human immunodeficiency virus type 1 particles. *J. Virol.* **67**, 2681–2688 (1993).
18. Yin, P. D. & Hu, W.-S. RNAs from genetically distinct retroviruses can copackage and exchange genetic information *in vivo*. *J. Virol.* **71**, 6237–6242 (1997).
19. Certo, J. L., Shook, B. F., Yin, P. D., Snider, J. T. & Hu, W.-S. Nonreciprocal pseudotyping: murine leukemia virus proteins cannot efficiently package spleen necrosis virus-based vector RNA. *J. Virol.* **72**, 5408–5413 (1998).
20. Kaye, J. F. & Lever, A. M. Nonreciprocal packaging of human immunodeficiency virus type 1 and type 2 RNA: a possible role for the p2 domain of *gag* in RNA encapsidation. *J. Virol.* **72**, 5877–5885 (1998).
21. Poon, D. T. K., Li, g. & Aldovini, A. Nucleocapsid and matrix protein contributions to selective Human Immunodeficiency Virus Type 1 genomic RNA packaging. *J. Virol.* **72**, 1983–1993 (1998).
22. Wang, H., Norris, K. M. & Mansky, L. M. Involvement of the matrix and nucleocapsid domains of the bovine leukemia virus Gag polyprotein precursor in viral RNA packaging. *J. Virol.* **77**, 9431–9438 (2003).
23. Lee, E.-G., Yeo, A., Kraemer, B., Wickens, M. & Linial, M. L. The Gag domains required for avian retroviral RNA encapsidation determined by using two independent assays. *J. Virol.* **73**, 6282–6292 (1999).
24. Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. W. & Oroslzan, S. Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J. Biol. Chem.* **256**, 8400–8406 (1981).
25. Green, L. M. & Berg, J. M. A retroviral Cys-Xaa2-Cys-Xaa4-His-Xaa4-Cys peptide binds metal ions: spectroscopic studies and a proposed three-dimensional structure. *Proc. Natl Acad. Sci. USA* **86**, 4047–4051 (1989).
26. Summers, M. F. *et al.* Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci.* **1**, 563–574 (1992).
Direct detection of zinc and its coordination environment in intact retroviruses provided the first direct evidence that retroviral NC proteins contain zinc knuckles.
27. Méric, C., Gouilloud, E. & Spahr, P.-F. Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His Boxes. *J. Virol.* **62**, 3328–3333 (1988).
28. Méric, C. & Goff, S. P. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J. Virol.* **63**, 1558–1568 (1989).
29. Gorelick, R. J., Henderson, L. E., Hanser, J. P. & Rein, A. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc. Natl Acad. Sci. USA* **85**, 8420–8424 (1988).
30. Rice, W. G. *et al.* Inhibition of HIV-1 infectivity by zinc-ejecting aromatic *C*-nitroso compounds. *Nature* **361**, 473–475 (1993).
Compounds that eject zinc from the NC zinc knuckles have potent antiviral activity.
31. Morellet, N. *et al.* Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ^{1}H NMR. *EMBO J.* **11**, 3059–3065 (1992).
32. Summers, M. F., South, T. L., Kim, B. & Hare, D. R. High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* **29**, 329–340 (1990).
33. Newman, J. L., Butcher, E. W., Patel, D. T., Makhaylenko, Y. & Summers, M. F. Flexibility in the P2 domain of the HIV-1 Gag polyprotein. *Protein Sci.* **13**, 2101–2107 (2004).
34. Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N. & Summers, M. F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **279**, 633–649 (1998).
35. Banks, J. D. & Linial, M. L. Secondary structure analysis of a minimal avian leukosis-sarcoma virus packaging signal. *J. Virol.* **74**, 456–464 (2000).
36. Mann, R. & Baltimore, D. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNA. *J. Virol.* **54**, 401–407 (1985).
The MoMuLV Ψ-site can direct packaging when relocated near the 3′ end of the genome, indicating that the residues function as an independent packaging domain.
37. Watanabe, S. & Temin, H. M. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5′ long terminal repeat and the start of the *gag* gene. *Proc. Natl Acad. Sci. USA* **79**, 5986–5990 (1982).
38. Lever, A. M. L., Göttlinger, H. G., Haseltine, W. A. & Sodroski, J. G. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**, 4085–4087 (1989).
39. Mansky, L. M., Krueger, A. E. & Temin, H. M. The bovine leukemia virus encapsidation signal is discontinuous and extends into the 5′ end of the *gag* gene. *J. Virol.* **69**, 3282–3289 (1995).
40. Browning, M. T., Mustafa, F., Schmidt, R. D., Lew, K. A. & Rizvi, T. A. Delineation of sequences important for efficient packaging of feline immunodeficiency virus RNA. *J. Gen. Virol.* **84**, 621–627 (2003).
41. Kaye, J. F. & Lever, A. M. Human innumodeficiency virus type-1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J. Virol.* **73**, 3023–3031 (1999).
42. Mustafa, F., Lew, K. A., Schmidt, R. D., Browning, M. T. & Rizvi, T. A. Mutational analysis of the predicted secondary RNA structure of the Mason–Pfizer monkey virus packaging signal. *Virus Res.* **99**, 35–46 (2004).
43. Hayashi, T., Ueno, Y. & Okamoto, T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett.* **327**, 213–218 (1993).
44. Mougel, M. & Barklis, E. A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions. *J. Virol.* **71**, 8061–8065 (1997).
45. Adam, M. A. & Miller, A. D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* **62**, 3802–3806 (1988).
46. Aschoff, J. M., Foster, D. & Coffin, J. M. Point mutations in the avian sarcoma/leukosis virus 3′ untranslated region result in a packaging defect. *J. Virol.* **73**, 7421–7429 (1999).
47. Yu, S. S., Kim, J.-M. & Kim, S. The 17 nucleotides downstream from the *env* gene stop codon are important for Murine Leukemia Virus packaging. *J. Virol.* **74**, 8775–8780 (2000).
48. Peters, G. & Hu, J. Reverse transcriptase as the major determinant for selective packaging of tRNAs into avian sarcoma virus particles. *J. Virol.* **36**, 692–700 (1980).
49. Katz, R. A., Terry, R. W. & Skalka, A. M. A conserved *cis*-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. *J. Virol.* **59**, 163–167 (1986).
50. Greatorex, J. & Lever, A. Retroviral RNA dimer linkage. *J. Gen. Virol.* **79**, 2877–2882 (1998).
51. Paillart, J.-C., Marquet, R., Skripkin, E., Ehresmann, C. & Ehresmann, B. Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* **78**, 639–653 (1996).
52. Berkhout, B. Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucl. Acid Res. Mol. Biol.* **54**, 1–34 (1996).
53. Laughrea, M. *et al.* Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **71**, 3397–3406 (1997).
54. McBride, M. S. & Panganiban, A. T. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation *in vivo*. *J. Virol.* **71**, 2050–2058 (1997).
55. Levin, J. G., Grimley, P. M., Ramseur, J. M. & Berezesky, I. K. Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D. *J. Virol.* **14**, 152–161 (1974).
MoMuLV particles contain two genomic RNA molecules, even when produced under conditions that inhibit mRNA production.
56. Hibbert, C. S., Mirro, J. & Rein, A. mRNA molecules containing MLV packaging signals are encapsidated as dimers. *J. Virol.* **78**, 10927–10938 (2004).
57. Housset, V., De Rocquigny, H., Roques, B. P. & Darlix, J.-L. Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *J. Virol.* **67**, 2537–2545 (1993).
58. Sakuragi, J.-I., Shioda, T. & Panganiban, A. T. Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions. *J. Virol.* **75**, 2557–2565 (2001).
Mutant HIV-1 RNAs that contains two copies of the 5′-UTR can be efficiently packaged as monomers, indicating that the native genome is specifically recognized for packaging as a dimer.
59. Sakuragi, J.-I., Iwamoto, A. & Shioda, T. Dissociation of genome dimerization from packaging functions and virion maturation of Human Immunodeficiency Virus Type 1. *J. Virol.* **76**, 959–967 (2002).
60. Mann, R., Mulligan, R. C. & Baltimore, D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**, 153–159 (1983).
61. Schwartzberg, P., Colicelli, J., Gordon, M. L. & Goff, S. Mutations in the *gag* gene of Moloney murine leukemia virus: effects on production of virions and reverse transcriptase. *J. Virol.* **49**, 918–924 (1984).
62. Murphy, J. E. & Goff, S. P. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: Effects on RNA packaging and reverse transcription. *J. Virol.* **63**, 319–327 (1989).
63. Bender, M. A., Palmer, T. D., Gelinas, R. E. & Miller, A. D. Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. *J. Virol.* **61**, 1639–1646 (1987).
64. Tounekti, N. *et al.* Effect of dimerization on the conformation of the encapsidation Ψ domain of Moloney murine leukemia virus RNA. *J. Mol. Biol.* **223**, 205–220 (1992).
Chemical probing of the MoMuLV 5′-UTR provides evidence for dimerization-dependent structural changes, which could have a regulatory function.
65. Mougel, M. *et al.* Conformational analysis of the 5′ leader and the *gag* initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucleic Acids Res.* **21**, 4677–4684 (1993).
66. Prats, A.-C. *et al. cis* elements and *trans*-acting factors involved in dimer formation of murine leukemia virus RNA. *J. Virol.* **64**, 774–783 (1990).
67. Girard, P.-M., Bonnet-Mathoniere, B., Mauriaux, D. & Paoletti, J. A short autocomplementary sequence in the 5′ leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry* **34**, 9785–9794 (1995).
68. De Tapia, M., Metzler, V., Mougel, M., Ehresmann, B. & Ehresmann, C. Dimerization of MoMuLV genomic RNA: redefinition of the role of the palindromic stem-loop H1 (278–303) and new roles for stem-loops H2 (310–352) and H3 (355–374). *Biochemistry* **37**, 6077–6085 (1998).
69. Ly, H. & Parslow, T. G. Bipartite signal for genomic RNA dimerization in the Moloney Murine Leukemia Virus. *J. Virol.* **76**, 3135–3144 (2002).

70. Oroudjev, E. M., Kang, P. C. E. & Kohlstaedt, L. A. An additional dimer linkage structure in Moloney Murine Leukemia Virus RNA. *J. Mol. Biol.* **291**, 603–613 (1999).
71. Rein, A. Take two. *Nature Struct. Mol. Biol.* **11**, 1034–1035 (2004).
72. Fisher, J. & Goff, S. P. Mutational analysis of stem-loops in the RNA packaging signal of the moloney murine leukemia virus. *Virology* **244**, 133–145 (1998).
73. Mougel, M., Zhang, Y. & Barklis, E. *cis*-active structural motifs involved in specific encapsidation of Moloney murine leukemia virus RNA. *J. Virol.* **70**, 5043–5050 (1996).
74. Beasley, B. E. & Hu, W. S. *cis*-acting elements imporant for retroviral RNA packaging specificity. *J. Virol.* **76**, 4950–4960 (2002).
75. Evans, M. J., Bacharach, E. & Goff, S. P. RNA sequences in the Moloney murine leukemia virus genome bound by the Gag precursor protein in the yeast three-hybrid system. *J. Virol.* **78**, 7677–7684 (2004).
76. D'Souza, V. *et al.* Identification of a high-affinity nucleocapsid protein binding site within the Moloney Murine Leukemia Virus Ψ-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **314**, 217–232 (2001).
77. Kim, C.-H. & Tinoco Jr., I. A retroviral RNA kissing complex containing only two G-C base pairs. *Proc. Natl Acad. Sci. USA* **97**, 9396–9401 (2000).
78. Oritz-Conde, B. A. & Hughes, S. H. Studies of the genomic RNA of leukosis viruses: implications for RNA dimerization. *J. Virol.* **73**, 7165–7174 (1999).
79. Tchenio, T. & Heidmann, T. The dimerization/packaging sequence is dispensable for both the formation of high-molecular-weight RNA complexes within retroviral particles and the synthesis of provirus of normal structure. *J. Virol.* **69**, 1079–1084 (1995).
80. Damgaard, C. K., Andersen, E. S., Knudsen, B., Gorodkin, J. & Kjems, J. RNA interactions in the 5′ region of the HIV-1 genome. *J. Mol. Biol.* **336**, 369–379 (2004).
81. D'Souza, V. & Summers, M. F. Structural basis for packaging the dimeric genome of Moloney Murine Leukaemia Virus. *Nature* **431**, 586–590 (2004).
NMR studies reveal that dimerization-dependent structural changes in the MoMuLV 5'-UTR can regulate high-affinity NC binding, indicating a mechanism for the selective packaging of a diploid genome.
82. Dey, A., York, D., Smalls-Mantey, A. & Summers, M. F. Position and sequence dependent binding of RNA to the nucleocapsid protein of Moloney Murine Leukemia Virus. *Biochemistry* (in the press).
83. Darlix, J.-L., Lapadat-Tapolsky, M., de Rocquigny, H. & Roques, B. P. First glimpses at structure–function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–537 (1995).
84. Aldovini, A. & Young, R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64**, 1920–1926 (1990).
85. Clavel, F. & Orenstein, J. M. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* **64**, 5230–5234 (1990).
86. Poznansky, M., Lever, A. M. L., Bergeron, L., Haseltine, W. & Sodroski, J. Gene transfer into human lymphocytes by a defective human immunodeficiency virus type 1 vector. *J. Virol.* **65**, 532–536 (1991).
87. McBride, M. S. & Panganiban, A. T. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* **70**, 2963–2973 (1996).
88. Harrison, G. P., Miele, G., Hunter, E. & Lever, A. M. L. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J. Virol.* **72**, 5886–5896 (1998).
89. Kim, H.-J., Lee, K. & O'Rear, J. J. A short sequence upstream of the 5′ major splice site is important for encapsidation of HIV-1 genomic RNA. *Virology* **198**, 336–340 (1994).
90. Luban, J. & Goff, S. P. Mutational analysis of *cis*-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* **68**, 3784–3793 (1994).
91. Buchschacher, J., G. L. & Panganiban, A. T. Human immunodeficiency virus vectors for inducible expresson of foreign genes. *J. Virol.* **66**, 2731–2739 (1992).
92. Parolin, C., Dorfman, t., Palu, G., Gottlinger, H. G. & Sodroski, J. Analysis in human immunodeficiency virus type 1 vectors of *cis*-acting sequences that affect gene transfer into human lymphocytes. *J. Virol.* **68**, 3888–3895 (1994).
93. Clever, J. L., Miranda, J., D. & Parslow, T. G. RNA structure and packaging signals in the 5′ leader region of the human immunodeficiency virus type 1 genome. *J. Virol.* **76**, 12381–12387 (2002).
94. Russell, R. S., Hu, J., Laughrea, M., Wainberg, M. A. & Liang, C. Deficient dimerization of human immunodeficiency virus type 1 RNA caused by mutations of the U5 RNA sequences. *Virology* **303**, 152–163 (2002).
95. Helga-Maria, C., Hammarskjold, M.-L. & Rekosh, D. An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type-1 genomic RNA. *J. Virol.* **73**, 4127–4135 (1999).
96. Hayashi, T., Shioda, T., Iwakura, Y. & Shibuta, H. RNA packaging signal of human immunodeficiency virus type 1. *Virology* **188**, 590–599 (1992).
97. Berkowitz, R. D., Hammarskjold, M.-L., Helga-Maria, C., Rekosh, D. & Goff, S. P. 5′ regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology* **212**, 718–723 (1995).
Unlike results obtained for MoMuLV, relatively large fragments of the HIV-1 5′-UTR are not sufficient to independently direct RNA packaging.
98. McBride, M. S., Schwartz, M. D. & Panganiban, A. T. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J. Virol.* **71**, 4544–4554 (1997).
The intact HIV-1 5′-UTR and residues of the *gag* open reading frame are capable of directing RNA packaging.
99. Baudin, F. *et al.* Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* **229**, 382–397 (1993).
100. Skripkin, E., Paillart, J. C. & Marquet, R., Ehresmann, B., Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl Acad. Sci. USA* **91**, 4945–4949 (1994).
101. Clever, J., Sassetti, C. & Parslow, T. G. RNA secondary structure and binding sites for *gag* gene products in the 5′ packaging signal of Human Immunodeficiency Virus Type 1. *J. Virol.* **69**, 2101–2109 (1995).
102. Clever, J. L. & Parslow, T. G. Mutant Human Immunodeficiency Virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414 (1997).
103. Sakaguchi, K. *et al.* Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc. Natl Acad. Sci. USA* **90**, 5219–5223 (1993).
104. Dannull, J., Surovoy, A., Jung, G. & Moelling, K. Specific binding of HIV-1 nucleocapsid protein to Ψ RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.* **13**, 1525–1533 (1994).
105. Berkowitz, R. D. & Goff, S. P. Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology* **202**, 233–246 (1994).
106. De Guzman, R. N. *et al.* Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science* **279**, 384–388 (1998).
107. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B. & Summers, M. F. NMR structure of stem loop SL2 of the HIV-1 Ψ-RNA packaging signal reveals a novel A-U-A base triple platform. *J. Mol. Biol.* **299**, 145–156 (2000).
108. Paillart, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. A loop–loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl Acad. Sci. USA* **93**, 5572–5577 (1996).
109. Clever, J. L., Wong, M. L. & Parslow, T. G. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J. Virol.* **70**, 5902–5908 (1996).
110. Laughrea, M. & Jette, L. A 19-nucleotide sequence upstream of the 5′ major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **33**, 13464–13474 (1994).
111. Takahashi, K.-I. *et al.* NMR analysis of intra- and inter-molecular stems in dimerization initiation site. *J. Biochem.* **127**, 681–686 (2000).
112. Lawrence, D. C., Stover, C. C., Noznitsky, J., Wu, Z.-R. & Summers, M. F. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem loop SL1. *J. Mol. Biol.* **326**, 529–542 (2003).
113. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. & Barre-Sinoussi, F. *cis* elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**, 689–699 (1990).
HIV-1 NC has chaperone activity that can catalyse dimerization of the viral RNA.
114. Feng, Y.-X. *et al.* HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA *in vitro*. *Proc. Natl Acad. Sci. USA* **93**, 7577–7581 (1996).
115. Mihailescu, M.-R. & Marino, J. P. A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc. Natl Acad. Sci. USA* **101**, 1189–1194 (2004).
116. Amarasinghe, G. K. *et al.* NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. *J. Mol. Biol.* **301**, 491–511 (2000).
117. South, T. L. & Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Ψ-site analog, dACGCC. *Protein Sci.* **2**, 3–19 (1993).
118. Clever, J. L., Taplitz, R. A., Lochrie, M. A., Polisky, B. & Parslow, T. G. A heterologous, high-affinity RNA ligand for human immunodeficiency virus Gag protein has RNA packaging activity. *J. Virol.* **74**, 541–546 (2000).
119. Russell, R. S. *et al.* Sequences downstream of the 5′ splice donor site are required for both packaging and dimerization of human immunodeficiency virus type-1 RNA. *J. Virol.* **77**, 84–96 (2003).
120. Amarasinghe, G. K. *et al.* Stem-loop SL4 of the HIV-1 Ψ-RNA packaging signal exhibits weak affinity for the nucleocapsid protein. Structural studies and implications for genome recognition. *J. Mol. Biol.* **314**, 961–969 (2001).
121. Yu, E. & Fabris, D. Direct probing of RNA structures and RNA–protein interactions in the HIV-1 packaging signal by chemical modification and electrospray ionization–Fourier transform mass spectrometry. *J. Mol. Biol.* **330**, 211–223 (2003).
122. Shubsda, M. F., Paoletti, A. C., Hudson, B. S. & Borer, P. N. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **41**, 5276–5282 (2002).
123. Hagan, N. & Fabris, D. Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 Ψ-recognition element. *Biochemistry* **42**, 10736–10745 (2003).
124. Berkhout, B. & Van Wamel, J. L. B. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA* **6**, 282–295 (2000).
125. Huthoff, H. & Berkhout, B. Two alternating structures of the HIV-1 leader RNA. *RNA* **7**, 143–157 (2001).
126. Berkhout, B. *et al. In vitro* evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. *J. Biol. Chem.* **277**, 19967–19975 (2002).
127. Ooms, M., Huthoff, H., Russell, R., Liang, C. & Berkhout, B. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* **78**, 10814–10819 (2004).
128. Paillart, J. C. *et al.* First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* **279**, 48397–48403 (2004).
129. Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. *In vitro* evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of the HIV-1 genomic RNA. *J. Biol. Chem.* **277**, 5995–6004 (2002).
130. Abbink, T. E. M. & Berkhout, B. A novel long distance base-pairing interaction in Human Immunodeficiency Virus Type 1 RNA occludes the Gag start codon. *J. Biol. Chem.* **278**, 11601–11611 (2003).
131. McCann, E. M. & Lever, A. M. Location of *cis*-acting signals important for RNA encapsidation in the leader sequence of human immunodeficiency virus type 2. *J. Virol.* **71**, 4544–4554 (1997).
132. Griffin, S. D. C., Allen, J. F. & Lever, A. M. The major human immunodeficiency virus type 2 (HIV-2) packaging signal is present on all HIV-2 RNA species: cotranslational RNA encapsidation and limitation of Gag protein confer specificity. *J. Virol.* **75**, 12058–12069 (2001).
133. Arya, S. K., M., Z. & Kundra, P. Human immunodeficiency virus type 2 lentivirus for gene transfer: expression and potential for helper virus-free packaging. *Hum. Gene Ther.* **9**, 1371–1380 (1998).
134. Poeschla, E. *et al.* Identification of a human immunodeficiency virus type 2 (HIV-2) encapsidation determinant and transduction of non-dividing cells by HIV-2-based lentivirus vectors. *J. Virol.* **72**, 6527–6536 (1998).
135. Dirac, A. M. G., Huthoff, H., Kjems, J. & Berkhout, B. The dimer initiation site hairpin mediates dimerization of the human immunodeficiency virus, type 2 RNA genome. *J. Biol. Chem.* **276**, 32345–32352 (2001).
136. Lanchy, J.-M. & Lodmell, J. S. Alternate usage of two dimerization initiation sites in HIV-2 RNA *in vitro*. *J. Mol. Biol.* **319**, 637–648 (2002).
137. Jossinet, F., Lodmell, J. S., Ehressmann, C., Ehresmann, B. & Marquet, R. Identification of the *in vitro* HIV-2/SIV RNA dimerization site reveals striking differences with HIV-1. *J. Biol. Chem.* **276**, 5598–5604 (2001).
138. Lanchy, J.-M., Rentz, C. A., Ivanovitch, J. D. & Lodmell, J. S. Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize *in vitro*. *Biochemistry* **42**, 2634–2642 (2003).

139. Dirac, A. M. G., Huthoff, H., Kjems, J. & Berkhout, B. Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucleic Acids Res.* **30**, 2647–2655 (2002).
140. Damgaard, C. K., Dyhr-Mikkelsen, H. & Kjems, J. Mapping the RNA binding sites for human immunodeficiency virus type-1 Gag and NC proteins within the complete HIV-1 and HIV-2 untranslated leader regions. *Nucleic Acids Res.* **26**, 3667–3676 (1998).
141. Banks, J. D., Yeo, A., Green, K., Cepeda, F. & Linial, M. L. A minimal avian retroviral packaging sequence has a complex structure. *J. Virol.* **72**, 6190–6194 (1998).
142. Banks, J. D., Kealoha, B. O. & Linial, M. L. An MΨ containing heterologous RNA, but not *env* mRNA, is efficiently packaged into avian retroviral particles. *J. Virol.* **73**, 8926–8933 (1999).
143. Doria-Rose, N. A. & Vogt, V. M. *In vivo* selection of Rous sarcoma virus mutants with randomized sequences in packaging signals. *J. Virol.* **72**, 8073–8082 (1998).
144. Lee, E.-G. & Linial, M. L. Yeast three-hybrid screening of Rous sarcoma virus mutants with randomly mutagenized minimal packaging signals reveals regions important for Gag interactions. *J. Virol.* **74**, 9167–9174 (2000).
145. Fosse, P. *et al.* A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry* **35**, 16601–16609 (1996).
146. Lear, A. L., Haddrick, M. & Heaply, S. A study of the dimerization of the Rous sarcoma virus RNA *in vitro* and *in vivo*. *Virology* **212**, 47–57 (1995).
147. Bieth, E., Gabus, C. & Darlix, J.-L. A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis *in vitro*. *Nucleic Acids Res.* **18**, 119–127 (1990).
148. Parent, L. J. *et al.* RNA dimerization defect in a Rous sarcoma virus matrix mutant. *J. Virol.* **74**, 164–172 (2000).
149. Garbitt, R. A., Albert, J. A., Kessler, M. D. & Parent, L. J. *trans*-acting inhibition of genomic RNA dimerization by Rous sarcoma virus matrix mutants. *J. Virol.* **75**, 260–268 (2001).
150. Kemler, I., Barraza, R. & Poeschla, E. Mapping of the encapsidation determinants of feline immunodeficiency virus. *J. Virol.* **76**, 11889–11903 (2002).
151. Mansky, L. M. & Gajary, L. C. The primary nucleotide sequence of the bovine leukemia virus RNA packaging signal can influence efficient RNA packaging and virus replication. *J. Virol.* **301**, 272–280 (2002).
152. Katoh, I., Kyushiki, H., Sakamoto, Y., Ikawa, Y. & Yoshinaka, Y. Bovine leukemia virus matrix-associated protein MA(p15): further processing and formation of a specific complex with the dimer of the 5′-terminal genomic RNA fragment. *J. Virol.* **65**, 6845–6855 (1991).
153. Vile, R. G., Ali, M., Hunter, e. & McClure, M. O. Identification of a generalized packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. *Virology* **189**, 786–791 (1992).
154. Harrison, G. P., Hunter, E. & Lever, A. M. L. Secondary structure model of the Mason–Pfizer monkey virus 5′ leader sequence: identification of a structural motif common to a variety of retroviruses. *J. Virol.* **69**, 2175–2186 (1995).
155. Schmidt, R. D., Mustafa, F., KLew, K. A., Browning, M. T. & Rizvi, T. A. Sequences both within the 5′ untranslated region and the *gag* gene are important for efficient encapsidation of Mason–Pfizer monkey virus RNA. *Virology* **309**, 166–178 (2003).
156. Ma, Y. M. & Vogt, V. M. Nucleic acid binding-induced Gag dimerization in the assembly of Rous sarcoma virus particles *in vitro*. *J. Virol.* **78**, 52–60 (2004).
157. Yu, F. *et al.* Characterization of Rous sarcoma virus Gag particles assembled *in vitro*. *J. Virol.* **75**, 2753–2764 (2001).
158. Muriaux, D., Mirro, J., Nagashima, K., Harvin, D. & Rein, A. Murine leukemia virus nucleocapsid mutant particles lacking viral RNA encapsidate ribosomes. *J. Virol.* **76**, 11405–11413 (2002).
159. Aronoff, R. & Linial, M. Specificity of retroviral RNA packaging. *J. Virol.* **65**, 71–80 (1991).
160. Linial, M., Medeiros, E. & Hayward, W. S. An avian oncovirus mutant (*SE* 21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* **15**, 1371–1381 (1978).
161. Dupont, S. *et al.* A novel nuclear export activity in HIV-1 matrix protein required for viral replication. *Nature* **402**, 681–685 (1999).
162. Scheifele, L. Z., Garbitt, R. A., Rhoads, J. D. & Parent, L. J. Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. *Proc. Natl Acad. Sci. USA* **99**, 3944–3949 (2002).
Mutations in RSV MA that inhibit nuclear export of Gag lead to accumulation of Gag and viral RNA in the nucleus and to the production of genome-deficient virus particles.
163. Callahan, E. M. & Wills, J. W. Link between genome packaging and rate of budding for Rous Sarcoma Virus. *Virology* **77**, 9388–9398 (2003).
164. Nash, M. A., Meyer, M. K., Decker, G. L. & Arlinghaus, R. B. A subset of Pr65gag is nucleus associated in Murine Leukemia Virus-infected cells. *J. Virol.* **67**, 1350–1356 (1993).
165. Basyuk, E. *et al.* Retroviral genomic RNAs are transported to the plasma membrane by endosomal vesicles. *Dev. Cell* **5**, 161–174 (2003).
166. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101 nucleotide core encapsidation signal of the Moloney Murine Leukemia Virus. *J. Mol. Biol.* **337**, 427–442 (2004).

Acknowledgements
Support from the NIH is gratefully acknowledged.

Competing interests statement
The authors declare no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez
avian leukosis virus | bovine leukaemia virus | feline immunodeficiency virus | HIV-1 | HIV-2 | Mason–Pfizer monkey virus | Moloney murine leukaemia virus | mouse mammary tumour virus | Rous sarcoma virus
CDC Infectious Disease Information:
http://www.cdc.gov/ncidod/diseases/index.htm
AIDS

FURTHER INFORMATION
Victoria D'Souza and Michael Summers' homepage:
http://www.hhmi.umbc.edu
Access to this interactive links box is free online.