

## Take two

Alan Rein

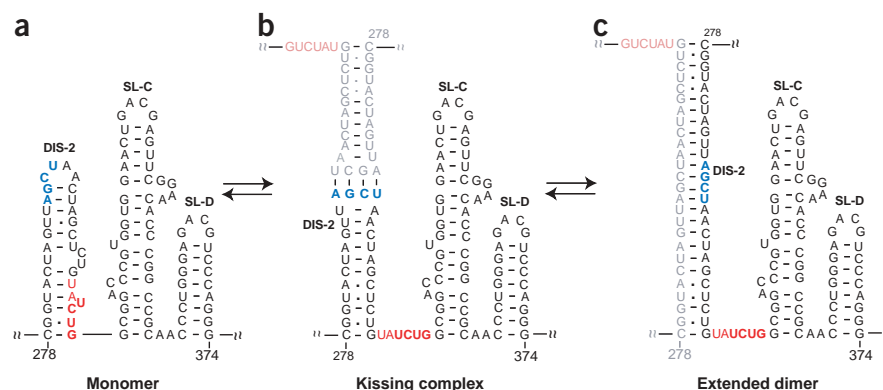
Retrovirus particles contain a dimeric form of their genomic RNA. A recent study suggests a mechanism that might underlie the selective incorporation of dimeric RNAs into assembling virus particles.

The genomic RNA of retroviruses, like that of many RNA viruses, leads a double life. On the one hand, the retroviral RNA must be incorporated into nascent virus particles in order to carry the viral genetic information to the new host cell. On the other hand, it serves as the mRNA for the Gag and Gag-Pol proteins within the virus-producing cell. How are these disparate roles balanced and coordinated? An article in a recent issue of *Nature*<sup>1</sup> describes structural studies on components of Moloney murine leukemia virus (MLV) that lead the authors to propose an elegant solution to this problem. The results show that a viral protein involved in particle formation binds selectively to a region of the RNA that is exposed in packaged RNA, but not in mRNA.

The assembly of a retrovirus particle is mediated by a single viral gene product, the Gag protein. Gag normally incorporates the genomic RNA into the nascent particle by recognizing a *cis*-acting signal, termed  $\psi$ , present in the RNA. The structure of  $\psi$  and the molecular details of this recognition are poorly understood at present, but  $\psi$  is generally believed to consist of a small number of stem-loop structures near the 5' end of the viral RNA<sup>2</sup>.

One striking feature of the packaged genomic RNA is that it is dimeric within the virus particle—that is, it is present in two identical copies joined by a limited number of Watson-Crick base pairs. The redundancy of the information provides insurance against breaks or other damage that would be fatal in a single copy. When the virus enters a new host cell and the RNA is copied into DNA by reverse transcriptase, the enzyme can jump from one copy to the other during DNA synthesis. This template-switching during reverse transcription can result in genetic recombination, an important source of genetic variation for retroviruses. In essence, retroviruses are the simplest diploid organisms.

The author is at the HIV Drug Resistance Program, National Cancer Institute-Frederick, P.O. Box B, 1050 Boyle Street, Building 535, Room 211, Frederick, Maryland 21702-1201, USA. e-mail: rein@ncifcrf.gov



**Figure 1** Stem-loops in MLV  $\psi$ . (a) The secondary structure of MLV bases 278–374 in monomeric form. (b) The kissing complex, in which DIS-2 has undergone a register shift, exposing the palindrome AGCU in the loop (blue). DIS-2 from two monomers is shown with intermolecular base-pairing between the loops. Note that the register shift has also exposed bases 304–309 (red). (c) The extended dimer, in which the stems, as well as the loops, of the DIS-2 elements of the two monomers are base-paired intermolecularly.

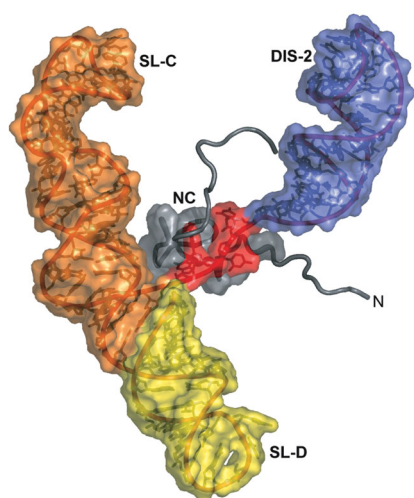
The exact structure of the linkage between the two RNAs in the dimer is not known. The situation is complicated by the fact that the RNA-RNA linkage is different in ‘immature’ particles, which have just been released from the cell, from that in ‘mature’ particles, in which the viral proteins have been cleaved by the viral protease<sup>3</sup>. Nevertheless, when partially denatured RNAs from mature particles were examined by electron microscope many years ago, they were found to be joined near their 5' ends; thus, at least in RNAs from mature virus particles, the most stable linkage is in this region of the RNA<sup>4,5</sup>. This region has sometimes been called the dimer linkage site (DLS).

A common feature of DLS regions is the presence of stem-loops (SLs) with four- or six-base palindromic sequences in the loops (‘kissing loops’), and there is extensive evidence that these structures contribute to the dimerization of short RNAs<sup>6–8</sup>. Thus, one simple, attractive scenario for dimerization is that the first step is base-pairing between the loops of two RNA monomers; this paired structure has frequently been referred to as a ‘kissing complex’ (Fig. 1). Such a complex might be converted during virus maturation to a more stable ‘extended dimer,’ in which bases in the stems, as well as those in the

loops, are paired intermolecularly rather than intramolecularly (Fig. 1c).

The fact that  $\psi$  and the DLS are in the same region of the viral RNA suggests the possibility that dimerization and packaging are intimately related to each other. In other words, perhaps dimerization of the RNA is a prerequisite for packaging. In fact, when packaging of viral RNA is reduced, either by actinomycin D treatment of the cells<sup>9</sup> or by changes in the structure of the RNA<sup>10</sup>, the genomic RNA in the released particles is still dimeric, even when the majority of particles do not contain genomic RNA; this is strong evidence that the packaged genomic RNA molecules enter the assembling particles in pairs, rather than singly. Furthermore, one striking exception to this rule is found in HIV-1 RNAs that have been engineered to contain two tandem copies of the  $\psi$  region: these RNAs are packaged into virus particles as monomers<sup>11</sup>. These results are all consistent with the idea that the real packaging signal is only present in dimeric RNAs (or in RNAs with a pseudo-dimeric three-dimensional structure<sup>11</sup>).

Extensive genetic analysis suggests that MLV  $\psi$  contains three stem-loops, designated SL-B (dimerization initiation site 2 or ‘DIS-2’ in the current work), SL-C and SL-D.



**Figure 2** Structure of the complex between MLV NC and an RNA representing MLV bases 278–374. NC (gray) binds to bases 306–309 (red), the flexible linker between DIS-2 and stem-loops C and D.

The loop in DIS-2 is palindromic, and this structure clearly plays an important role in the dimerization of short MLV-derived RNAs<sup>12</sup>. Some years ago, structures of short RNAs containing these three stem-loops were investigated by chemical structure-probing (in which the reactivity of individual bases toward various chemical agents was measured)<sup>13</sup>. It was found that the bases in the stem of DIS-2 shift register when the RNA dimerizes (Fig. 1a,b). This shift exposes the palindromic sequence, AGCU, which becomes the loop in the dimeric form; this conformation enables the two monomers to form the kissing complex (Fig. 1b). In addition, the shift moves the six bases UAUCUG (nucleotides 304–309 of MLV) from the 3' end of the stem, in which they are paired with nucleotides 278–282, to the region between DIS-2 and SL-C. In other words, bases 304–309 are paired in the monomeric RNA, but not in the dimer. This fact is a key element in the ideas now put forth by D'Souza and Summers<sup>1</sup>.

In a previous study, D'Souza and Summers and their colleagues determined the three-dimensional structure of a 101-base RNA molecule representing all three of these stem-loops by NMR<sup>14</sup>. Although the RNA molecule used in those studies was altered to prevent dimerization, it was designed to mimic the conformation of one monomer in a dimer (Fig. 1c). They found that SL-C and SL-D stack together as an extended helix; DIS-2 is connected to this helix by a flexible linker, consisting of the four bases UUCG (MLV nucleotides 306–309). Thus, the three-dimensional structure of this RNA completely confirms the unpaired nature of these bases in the dimer structure.

One of the cleavage products formed from Gag during virus maturation is nucleocapsid (NC), a very small, very basic protein with one or two zinc fingers that may mediate Gag interactions with nucleic acids. NC binds with high affinity and stoichiometrically to MLV  $\psi$  sequences only when all three stem-loops are present<sup>15</sup>. To gain insight into the specific interaction of MLV Gag with the  $\psi$  element in MLV genomic RNA, D'Souza and Summers have now determined the structure of the complex between MLV NC and the 101-base RNA previously described<sup>1</sup>. They found that NC binds specifically to the UUCG sequence linking DIS-2 with the SL-C–SL-D stack (Fig. 2). In fact, the binding of NC to the RNA could be completely ascribed to its affinity for the unpaired UUCG bases, as was confirmed first with a series of truncated RNAs and finally by isothermal titration calorimetry with the hexanucleotide rUAUCUG. The dissociation constants obtained from these titrations are in the range of 100 nM.

Because the UUCG is only unpaired and available to NC when the RNA is in the dimeric conformation, these observations suggest an elegant mechanism explaining the exclusive packaging of dimeric RNA. It will be of great interest to test, in biological systems, the possibility that these four bases are a critical element in  $\psi$ . However, the authors acknowledge the likelihood that  $\psi$  will not

turn out to be localized in a single site. Attempts to define  $\psi$  by genetic means have not been successful at pinpointing it to a small number of bases. In fact, the  $\psi$  region contains several UUCG sequences; perhaps they will be recognized in concert by several Gag molecules at the first stage of RNA packaging. Furthermore, an additional stem-loop ('DIS-1' in the present work) in the 5' region of the MLV genome seems analogous in many ways to DIS-2: it can also shift register to expose a palindromic sequence in the loop (there is evidence that DIS-1 can contribute to dimerization of short RNAs *in vitro* and of packaged MLV RNA<sup>16,17</sup>), and this register shift, like that in DIS-2, exposes the sequence UAUCUG.

It should be noted that a somewhat similar proposal has also been made for HIV-1 RNA<sup>18</sup>. Thus, it seems possible that a conformational switch between monomers and dimers, exposing a crucial sequence only in the dimeric RNA, will eventually be recognized as a general mechanism for the packaging of dimeric RNAs by retroviruses.

1. D'Souza, V. & Summers, M.F. *Nature* **431**, 586–590 (2004).
2. Berkowitz, R., Fisher, J. & Goff, S.P. *Curr. Top. Microbiol. Immunol.* **214**, 177–218 (1996).
3. Fu, W. & Rein, A. *J. Virol.* **67**, 5443–5449 (1993).
4. Bender, W. & Davidson, N. *Cell* **7**, 595–607 (1976).
5. Kung, H.J. *et al. Cell* **7**, 609–620 (1976).
6. Prats, A.C. *et al. J. Virol.* **64**, 774–783 (1990).
7. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B. & Ehresmann, C. *Proc. Natl. Acad. Sci. USA* **91**, 4945–4949 (1994).
8. Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. *Nat. Rev. Microbiol.* **2**, 461–472 (2004).
9. Levin, J.G. & Rosenak, M.J. *Proc. Natl. Acad. Sci. USA* **73**, 1154–1158 (1976).
10. Hibbert, C.S., Mirro, J. & Rein, A. *J. Virol.* **78**, 10927–10938 (2004).
11. Sakuragi, J.-I., Shioda, T. & Panganiban, A.T. *J. Virol.* **75**, 2557–2565 (2001).
12. Girard, P.M., Bonnet-Mathoniere, B., Muriaux, D. & Paoletti, J. *Biochemistry* **34**, 9785–9794 (1995).
13. Tounekti, N. *et al. J. Mol. Biol.* **223**, 205–220 (1992).
14. D'Souza, V., Dey, A., Habib, D. & Summers, M.F. *J. Mol. Biol.* **337**, 427–442 (2004).
15. D'Souza, V. *et al. J. Mol. Biol.* **314**, 217–232 (2001).
16. Ly, H. & Parslow, T.G. *J. Virol.* **76**, 3135–3144 (2002).
17. Oroudjev, E.M., Kang, P.C. & Kohlstaedt, L.A. *J. Mol. Biol.* **291**, 603–613 (1999).
18. Huthoff, H. & Berkhout, B. *RNA* **7**, 143–157 (2001).