The EMBO Journal vol.13 no.3 pp.713-726, 1994

A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA

Shiaolan Yang and Howard M.Temin[1]

McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, 1400 University Avenue, Madison, WI 53706, USA
[1]Corresponding author

Communicated by P.-F.Spahr

We conducted a mutational analysis within the previously defined encapsidation sequence (E) for spleen necrosis virus (SNV), an avian retrovirus. We found that two regions are necessary for efficient SNV replication. The first region is a double hairpin structure as proposed by Konings *et al.* (1992, *J. Virol.*, 66, 632-640); the second region is located downstream of the hairpins. We showed further that the double hairpin structure is required for efficient SNV RNA encapsidation. Our work is the first to demonstrate, via linker-scanning and site-directed mutagenesis, that a specific RNA secondary structure is required for the encapsidation of retroviral RNA. Analysis of a series of mutations within the E region indicates (i) that preserving the secondary structure of the two hairpins is important for efficient encapsidation and (ii) that the stem regions of the hairpins contain specific sequences critical for encapsidation. Within the hairpins, the presence of at least one of the two conserved GACG four-residue loops, but not the moderately conserved bulge sequence of the first hairpin, is crucial for function. The function of the hairpins is independent of the relative order of the two hairpins. However, the two hairpins are not redundant and are not functionally identical. Replacement of SNV double hairpin sequences with those of Moloney murine leukemia virus (M-MLV) has no detectable effect on the replication of SNV-based retrovirus vectors with reticuloendotheliosis virus strain A (REV-A) helper virus. Furthermore, replacement of the entire E sequence of SNV with that of Moloney murine sarcoma virus (M-MSV) and M-MLV results in retroviral vectors that replicate as well as SNV vectors with wild type SNV E. This result indicates that the encapsidation sequences of M-MSV/M-MLV and SNV are not virus specific and that, during packaging of SNV and MLV RNA with viral proteins from REV-A, the encapsidation sequences are recognized largely by their secondary or tertiary structures.

Key words: encapsidation sequence/four-residue loop/RNA packaging/secondary structure/spleen necrosis virus

Introduction

Retroviruses are a family of RNA viruses that replicate through a DNA intermediate. After virus entry into a target cell, virus-encoded reverse transcriptase copies the viral RNA genome into double-stranded DNA, which subsequently integrates into the host chromosomal DNA to form the provirus. Retroviruses then take advantage of the host cell RNA transcriptional machinery to synthesize the viral genomic RNA. Viral RNAs are specifically recognized by viral proteins and are then encapsidated/packaged to form virions. The signals that differentiate the retroviral RNA from cellular RNA are designated the encapsidation or packaging sequences and are abbreviated E or ψ.

For type C retroviruses like murine leukemia virus (MLV) and spleen necrosis virus (SNV, an MLV-like avian retrovirus), the minimal E or ψ is located within the leader sequences between the major splice donor site and the *gag* start codon, and the extended encapsidation sequence ψ^+ extends into the *gag* open reading frame (ORF) for MLV (Watanabe and Temin, 1982; Mann *et al.*, 1983; Mann and Baltimore, 1985; Armentano *et al.*, 1987; Bender *et al.*, 1987; Embretson and Temin, 1987; Adam and Miller, 1988). The E or ψ and ψ^+ are present only on the unspliced genomic RNA for these viruses. Rous sarcoma virus (RSV) has the major splice donor site just downstream of the *gag* start codon (Weiss *et al.*, 1985). Its ψ sequences are located 5′ to the viral coding sequence (Gallis *et al.*, 1979; Koyama *et al.*, 1984; Katz *et al.*, 1986; Aronoff and Linial, 1991; Aronoff *et al.*, 1993), possibly extending into the *gag* region (Pugatsch and Stacey, 1983), and also include a region near the 3′ end of the genomic RNA (Sorge *et al.*, 1983). For bovine leukemia virus (BLV), or the lentiviruses human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus (SIV), E or ψ is not so well defined. The encapsidation sequences for these retroviruses seem to include the viral sequences upstream of the major splice donor site (Rizvi and Panganiban, 1993; K.Boris-Lawrie and H.M.Temin, manuscript in preparation) and the leader region between the major splice donor site and the *gag* start codon (Lever *et al.*, 1989; Aldovini and Young, 1990; Clavel and Orenstein, 1990; Hayashi *et al.*, 1992), and they seem to extend into the *gag* coding sequences (Hayashi *et al.*, 1992; Richardson *et al.*, 1993; L.M.Mansky and H.M.Temin, manuscript in preparation). Sequences elsewhere in the viral genome may also be involved (Richardson *et al.*, 1993). This complexity may be the result of the complicated genomes of these retroviruses and the presence of various inhibitory elements within the genomes (Schwartz *et al.*, 1992).

A unique feature of retroviruses is that virions contain two identical copies of the RNA genome, non-covalently bound near their 5′ ends at the dimer linkage sequences (DLSs) (Bender *et al.*, 1978). The presence of dimeric RNA in retroviral virions allows recombination to occur between the two molecules. The recombination is believed to be beneficial for the virus to repair damage in its genome, to increase variation and to escape host immune response and anti-retroviral therapies (by assorting mutations) (Hu and Temin, 1990a,b; Howell *et al.*, 1991; Temin, 1991; Groenink *et al.*, 1992; Jones *et al.*, 1993). Retroviral RNAs containing the DLS are able to dimerize efficiently *in vitro*

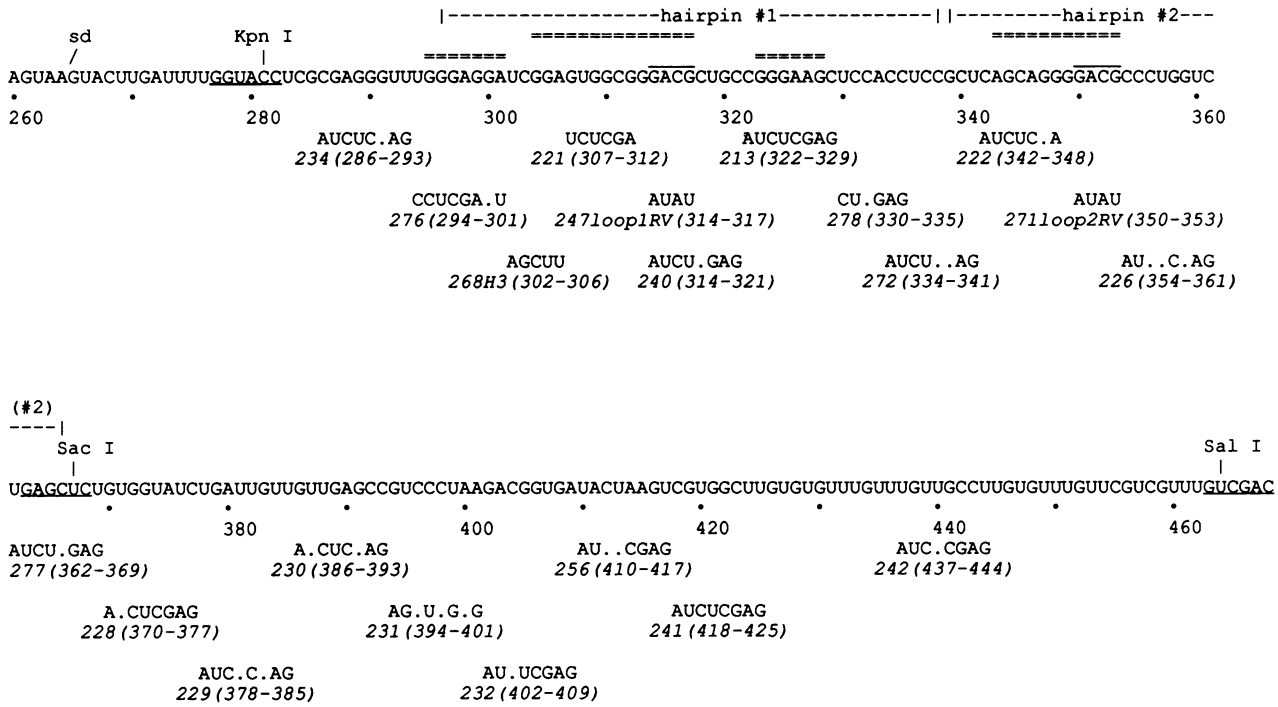

Fig. 1. Locations of the SNV E linker-scanning mutations. Numbering starts at the 5′ cap site of SNV genomic RNA (+1). The positions of recognition sites for the restriction enzymes *Kpn*I, *Sac*I and *Sal*I are underlined. The regions of the proposed hairpin-pair motif are indicated with dotted lines above the bases labeled hairpin #1 and #2. The GACG four-residue loop sequences are indicated with solid lines above the bases. The purine-rich sequences implicated in dimer formation *in vitro* are indicated with double dashed lines above the bases. 'sd' indicates the location of the splice donor site. The mutated bases in each individual mutant vector are presented below the wild type SNV sequence; bases that are different from the wild type bases are shown and non-mutated wild type bases are indicated by a dot (.). The names of the mutant vectors containing the linker-scanning mutations, as well as the coordinates of the mutations, are indicated below each vector sequence.

in the absence of added proteins. However, addition of retroviral nucleocapsid (NC) proteins seems to enhance dimerization (Prats *et al.*, 1990; Darlix *et al.*, 1992). For retroviruses (possibly excluding RSV), the DLSs have mapped to a region overlapping the E or ψ (Bieth *et al.*, 1990; Darlix *et al.*, 1990, 1992; Prats *et al.*, 1990; Katoh *et al.*, 1991; Marquet *et al.*, 1991; Sundquist and Heaphy, 1993). Encapsidation sequences also appear to include the sequences or structures important for binding to viral matrix-associated (MA) proteins, NC proteins or the NC domain of the Gag polyproteins *in vitro* (Katoh *et al.*, 1991; Luban and Goff, 1991; Sakaguchi *et al.*, 1993). Although it has long been postulated that dimerization of the retroviral genomic RNA and biochemical interactions between viral RNA and viral proteins are related to encapsidation, the exact nature of these interactions is not clear. So far certain structures have been proposed within several different retroviral E sequences (Prats *et al.*, 1990; Alford *et al.*, 1991; Harrison and Lever, 1992; Hayashi *et al.*, 1992; Konings *et al.*, 1992; Rizvi and Panganiban, 1993). A better understanding of the sequence- and structure-specific information within the E should help extend our knowledge about the encapsidation process.

For SNV, the E sequence was mapped by deletion analysis to a 180 nucleotide *Kpn*I–*Sal*I fragment located between the major splice donor site and the *gag* start codon, and both 5′ and 3′ halves of E are required (Watanabe and Temin, 1982; Embretson and Temin, 1987). The DLS of reticuloendotheliosis virus strain A (REV-A), a virus that is closely related to and shares 95% nucleotide sequence identity with SNV, maps to the same two halves of the *Kpn*I–*Sal*I fragment (Darlix *et al.*, 1992). Konings *et al.* (1992) proposed a common hairpin-pair motif in mammalian type C MLV-like retroviruses based on sequence analysis of the 5′ untranslated region (UTR) of these viruses. The motif has conserved four-residue loops, each consisting of the nucleotides GACG. In SNV, this motif is located within the E region, consistent with a role for this proposed structure in encapsidation and/or dimerization of SNV RNA.

We introduced a series of mutations in the SNV E to probe the sequence and structural information within the region. Our studies revealed a double hairpin structure, as proposed by Konings *et al.* (1992), and a downstream element required for efficient SNV replication. We found that preservation of the secondary structure of the hairpins is crucial for efficient SNV RNA encapsidation, and we present evidence for sequence-specific information within part of the stems. This is the first demonstration of the need for RNA secondary structure in the encapsidation of a retroviral RNA. We also studied various features of the hairpins. We found that the two hairpins are not redundant and are not functionally identical. In addition, we found that the ψ and ψ^+ sequences of M-MSV and M-MLV (a virus of the same genus as SNV), but not the E sequences of HIV-1 or BLV, can rescue an SNV E minus (E^-) retroviral vector for efficient replication.

Results

Assay of mutant vectors

We made a series of mutations within the previously defined SNV E region. The initial characterization involved a total of 21 linker-scanning mutations (Figure 1). Of the linker-scanning mutations, 18 were *Xho*I linker-scanning mutations

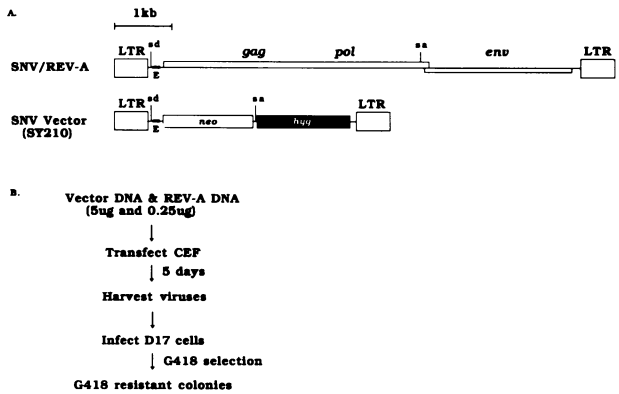

Fig. 2. (**A**) Diagrams of the proviral genomes of wild type SNV/REV-A and of the SNV vector used in this study. Open boxes labeled 'LTR' represent the long terminal repeats of SNV/REV-A. The vertical lines indicate the locations of the splice donor site (sd) and the splice acceptor site (sa). The long narrow open boxes represent the coding regions for the viral *gag*, *pol* and *env* genes, as indicated. The thin lines represent the non-coding viral sequences and the E region is represented by a thick line located between the splice donor site and the beginning of the *gag* open reading frame. The open rectangle labeled 'neo' is the *neo* gene, and the filled rectangle labeled 'hyg' is the *hyg* gene. (**B**) Flowchart of the protocol used to assay the E mutant vectors. The mutant/wild type vector plasmid DNA was cotransfected into CEF along with REV-A proviral plasmid DNA, at a ratio of 5:0.25 μg per 60 mm Petri dish. Virus was harvested 5 days later and was used to infect D17 target cells. The infected cells were selected with the drug G418, and the number of G418-resistant colonies was determined.

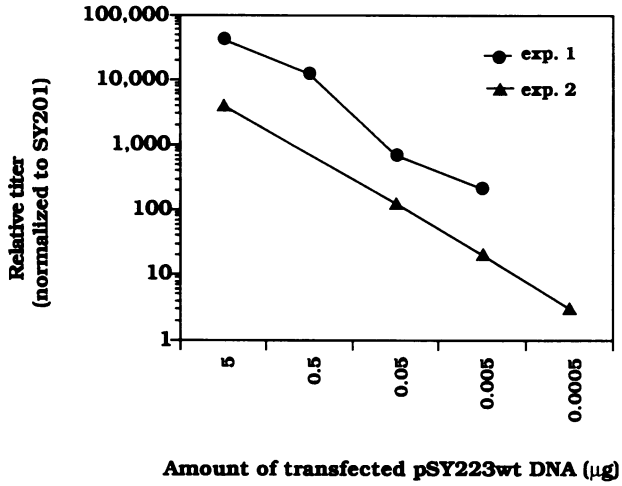

Fig. 3. Relative titers of SY223wt (normalized to that of SY201) as a function of the amount of plasmid DNA transfected. 10-fold serially diluted pSY223wt plasmid DNA was cotransfected into CEF along with 0.25 μg of the REV-A proviral DNA and 9 μg of carrier DNA per 60 mm Petri dish. Virus was harvested 5 days after transfection, and the titers of SY223wt virus were determined by assaying them on D17 target cells. The pSY201 plasmid DNA (5 μg, along with REV-A proviral DNA and carrier DNA) was tested in parallel as a control. Data from two independent experiments are shown.

that had five to eight base substitutions, one was a *Hin*dIII linker-scanning mutation that had five base substitutions, and the remaining two were *Eco*RV linker-scanning mutations in which one of the two conserved four-residue loop sequences of the proposed hairpin-pair motif was changed from GACG to AUAU.

All of the E mutations were cloned into an SNV retroviral vector, pSY210 (see Figure 2A). pSY210 is an SNV-based splicing vector and contains two selectable genes, *neo* and *hyg*, that allow the selection of resistance to the drugs G418 (an analog of neomycin) and hygromycin B in stably transfected or infected mammalian cells. The *neo* gene is expressed from the full-length vector viral RNA, and the *hyg* gene is expressed from the spliced subgenomic RNA. In the experiments described below, selection for the *neo* gene (i.e. G418 selection) was used whenever feasible. The hygromycin B selection was used only when the transfected or infected cells were helper cells that were already G418 resistant (Dougherty *et al.*, 1989).

To assay the E mutant vectors, the vector plasmid DNAs were cotransfected into chicken embryo fibroblasts (CEF) along with the replication-competent REV-A proviral plasmid DNA, at a ratio of 5:0.25 μg (Figure 2B). REV-A was used because it is capable of producing higher titer stocks of vector virus than SNV when used as a helper virus (Casella and Panganiban, 1993). Vector virus and helper virus production reached a peak level on the fifth day after transfection. At that time, virus was harvested from the cells and was used to infect D17 target cells. The D17 cells were placed under G418 selection 1 day after infection and, after selection, the titers of the vector virus were calculated from the number of G418-resistant colonies. For each assay, an SNV vector with a wild type E sequence, SY223wt, was included as a positive control, and an SNV vector with a deleted E sequence, SY201, was included as a negative control.

During the 5 days between the time of transfection and the time the virus was harvested, replication-competent REV-A helper virus was released from the transfected cells and spread in the culture, along with the vector virus packaged by the REV-A structural proteins. Replication of the vector virus depended upon the presence of a helper provirus in the same cell to provide the structural proteins *in trans*. Multiple rounds of infection of the vector virus occurred during the 5 day period. The effect of the base substitution mutations was magnified as a result of these multiple replication cycles. Furthermore, since vector virus can only be released from cells that were also infected by a REV-A helper virus, the wild type E sequence on the helper virus was in competition with the mutant E sequences on the vectors. Thus, the amount of the vector virus released from the cells reflected the packaging of mutant E vectors in the presence of competition from the wild type E sequence.

Because we used cotransfection in our assays, recombination between the two cotransfected DNAs could occur during transfection (Bandyopadhyay *et al.*, 1984), and certain recombination events could generate recombinant vectors with wild type E sequence. For this to happen with the mutant vector virus, one crossover event has to occur within a homologous region 24−175 nucleotides in length depending on the location of the mutations. We examined several G418-resistant D17 clones and found no evidence for such recombination events (see below).

To show that the system is sensitive and relatively quantitative, control experiments were performed in which 10-fold serial dilutions of pSY223wt DNA, which has the wild type SNV E sequence, were cotransfected with 0.25 μg of the REV-A plasmid DNA along with carrier DNA. As expected, the relative titers of the vector virus decreased correspondingly (Figure 3).

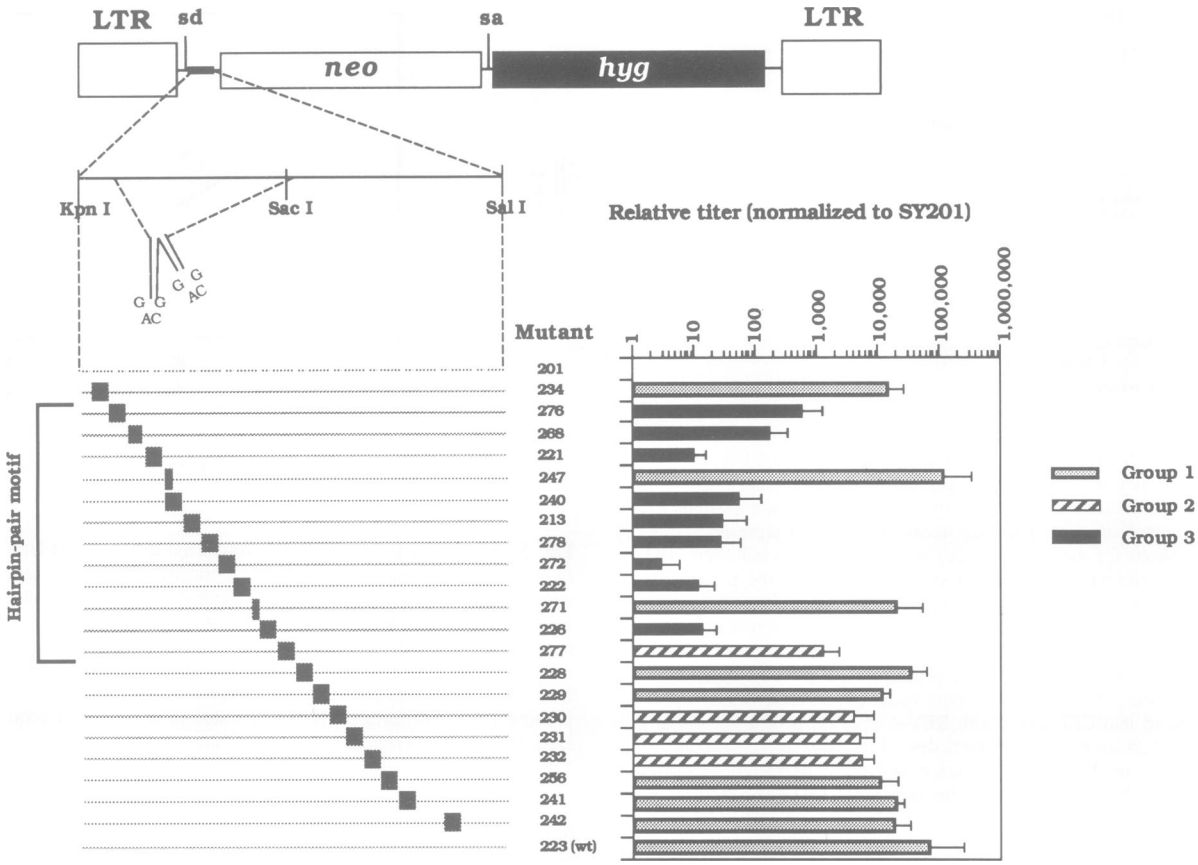

Fig. 4. Diagram of the location of linker-scanning mutations and the relative titers of each mutant vector. Shown on the top is the proviral genome of the SNV vector into which the various E mutations were cloned. The symbols are the same as those in Figure 2. The encapsidation sequence E is located within the *Kpn*I−*Sal*I fragment. The proposed hairpin-pair motif is located mainly within the 5′ half of E and overlaps the *Sac*I site. The shaded squares and rectangles indicate the locations of the linker-scanning mutations. The mutations that overlapped the proposed hairpin-pair motif are indicated with a labeled bracket. The mutant vectors were assayed as described above (Figure 2B), and the titer of each vector relative to that of the E^- vector SY201 is summarized to the right of the vectors. The relative titers of the mutant vectors containing the group 1 mutations and of SY223wt are depicted by light grey bars, those of the group 2 mutations by striped bars and those of the group 3 mutations by black bars. Each mutant was tested at least twice, and each test was performed in duplicate and in quadruplicate for SY201.

Regions of E important for retrovirus replication

The effects of various linker-scanning mutations on the replication of retroviral vectors are summarized as relative titers in Figure 4. The titer of each vector was normalized to that of E^- vector SY201 from the same assay. It should be noted that the titers of the mutants can be normalized to that of either the wild type E vector SY223wt or the E^- vector SY201. For ease of presentation, we chose the latter. To ensure reasonable precision in the normalization, we performed infections of SY201 in quadruplicate. Each mutant vector was tested at least twice, and each assay was performed in duplicate. The difference in titers between the wild type E vector SY223wt and the E^- vector SY201 ranged from ~1000 to >1 000 000, with a typical difference of ~10 000. The typical titer of SY201 was ~10−100 colony forming units (CFU) per ml.

Mutants exhibited a wide range of relative titers (Figure 4). In general, the mutations can be divided into three categories depending on their effects on relative virus titer. Group 1 consists of the mutation that had almost no effect on the replication of the vector virus. These mutant vectors had relative titers comparable with that of SY223wt. Group 2 mutations had a moderate effect on vector virus replication; they reduced the relative titers of the vector virus by a factor of 10−50 compared with that of the wild type E vector virus SY223wt. Group 3 mutations drastically reduced the relative titers of the vector virus; relative titers were at least a factor of 100 lower than that of the wild type E vector virus SY223wt. The boundary line between the group 1 and 2 mutations was arbitrarily chosen. The mutant vectors that contain group 1 or group 2 mutations had considerably higher relative titers than the mutant vectors containing group 3 mutations.

The group 3 mutations, which drastically reduced the relative titers of the mutant retroviral vectors, defined a region necessary for replication of the vector virus. The extent of reduction in relative titer ranged from a factor of 6000 (SY272) to 120 (SY276). The mutations overlapped the region of the hairpin-pair motif proposed by Konings *et al.* (1992; see Figures 1 and 4, mutants SY276−SY226). Interestingly, the only two mutant vectors within this region that did not have much reduction in relative titer of the vector virus were those that changed one of the two conserved GACG loop sequences of the proposed structure (SY247 and SY271). In addition, a downstream region was also found to be important for the replication of the SNV retroviral vectors, as shown by the three mutants containing the group 2 mutations SY230, SY231 and SY232.

To test whether or not the G418-resistant D17 colonies were the result of homologous recombination that could possibly occur between the cotransfected vector and helper sequences, two G418-resistant D17 colonies from one mutant

Table I. Relative titers of mutant vectors in helper cells[a]

Mutants	Relative titers (normalized to SY201)[b]	Description of mutation in E region
SY223wt	41 ± 10	wild type E
SY221	6 ± 5	six base substitutions of stem 1
SY222	7 ± 0.5	six base substitutions of stem 2
SY201[c]	1	E deleted

[a]The data in this table are taken from three independent assays.
[b]Relative titers are presented as average ± standard deviation.
[c]The average absolute titer of SY201 from the three assays presented here was 1400 CFU/ml.

vector and a pool of G418-resistant D17 colonies from another mutant vector containing linker-scanning mutations were analyzed by PCR, followed by restriction enzyme digestion. Because the original mutations were still present (data not shown), we do not believe that the G418-resistant D17 colonies in our assay were wild type E vectors generated from recombination.

Replication of the mutant E vectors in a single round of retrovirus replication

Mutants containing the group 3 mutations, which had a drastically reduced relative titer, were tested in CEF after multiple rounds of infection in the presence of REV-A helper virus. Two of these mutants, SY221 and SY222, were introduced into DAN and DSN SNV helper cells (Dougherty *et al.*, 1989; Hu and Temin, 1990a) to test the effect of these mutations on the replication of vector virus in a single cycle of replication in the absence of competing helper viral RNA. As controls, wild type E vector SY223wt and E^- vector SY201 were used.

We measured virus from one round of vector virus replication from the provirus in the helper cells to the provirus in the target cells. In the first approach, ~200–300 hygromycin B-resistant transfected DAN helper cells were pooled for each vector, and viruses produced from these cell pools were used to infect D17 target cells. In the second approach, ~150–200 hygromycin B-resistant DSN colonies were pooled for each vector after infection of these cells with viruses from transfected DAN helper cells; viruses produced from the infected DSN cell pools were used to infect D17 target cells (Table I). The identities of the mutant E vectors SY221 and SY222, E^- vector SY201 and the wild type E vector SY223wt in both the helper and the D17 target cells were confirmed by PCR analysis of the crude cell lysates, followed by restriction enzyme digestion and gel electrophoresis (data not shown). Both DAN and DSN helper cells lack the E sequence in their helper genome (Dougherty *et al.*, 1989); thus, in the helper cells, the E mutations in the retroviral vectors cannot revert back to wild type E through homologous recombination.

Similar results were obtained with either the DAN or the DSN helper cells, and the relative titers of the vector viruses from the two approaches are summarized in Table I. The average titer of the E^- vector SY201 produced from the helper cell pools was 1400 CFU/ml and was significantly higher than the average titer of SY201 from CEF. The ratio of average titers between SY223wt and SY201 was 40, smaller than that in CEF. The titers of mutants SY221 and SY222, each having a six base substitution in one of the two stems of the proposed hairpin-pair motif, were approximately a factor of six higher than that of E^- vector SY201 and approximately a factor of seven lower than wild type E vector SY223wt from the helper cell pools. Therefore, we conclude that the mutations we introduced did compromise the ability of vector virus to replicate after a single cycle in the absence of competing wild type virus.

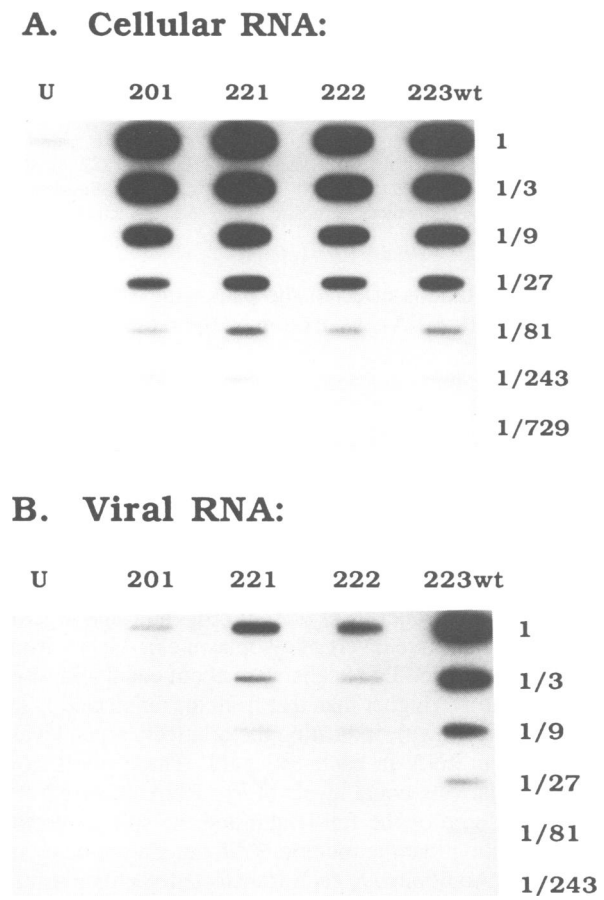

Fig. 5. Slot-blot analysis of relative cytoplasmic and viral RNA levels. (**A**) Cytoplasmic RNA was prepared from infected DSN helper cell pools as described in Materials and methods. RNA from one 100 mm Petri dish of infected or uninfected cells was 3-fold serially diluted and blotted onto nitrocellulose paper. The blot was probed with a randomly primed *hyg* probe. (**B**) Virus was harvested from infected DSN helper cell pools and viral RNA was prepared as described in Materials and methods. RNA from the equivalent of the supernatant medium from one-and-a-half 100 mm Petri dishes (7.5 ml in total) was 3-fold serially diluted and blotted. The probe is a randomly labeled *neo* probe.

Encapsidation of mutant vector retroviral RNAs

The linker-scanning mutations were located within the previously defined E region for SNV. To analyze whether

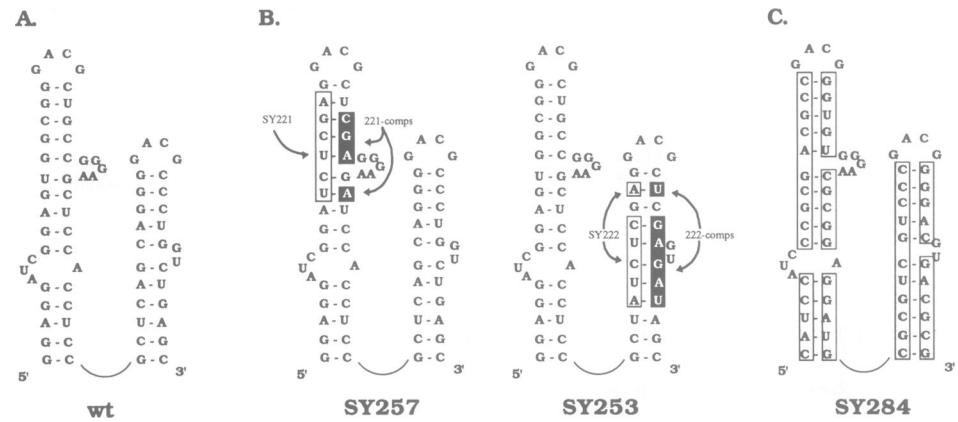

Fig. 6. Predicted RNA secondary structures of the SNV E region containing the double hairpin structure. Paired bases are indicated with a dash (-) between the bases. (**A**) The predicted structure of the wild type sequence is shown. (**B**) The predicted structures of the mutant vectors SY257 and SY253 containing the compensatory changes are shown. The bases that are different from the wild type sequences are boxed. The mutations that were present in the original mutant vectors SY221 and SY222 are shown in open boxes and are labeled SY221 and SY222, respectively. Compensatory mutations that restored the base pairing are shown in white within black boxes and are labeled 221-comps and 222-comps, respectively. (**C**) The predicted structure of the 'base switched' mutant vector SY284. The mutated bases are shown in open boxes.

or not the mutations affected the packaging efficiency of retroviral vector RNAs, we examined the relative levels of viral RNAs.

Viruses were harvested from subconfluent DSN helper cell pools containing SY201, SY221, SY222 or SY223wt proviruses, and viral RNA was extracted. At the same time, cytoplasmic RNA from the DSN helper cells was prepared. Both viral and cytoplasmic RNAs were quantified by slot blot analysis (Figure 5).

The infected DSN cells contain *neo* genes from both the plasmid pSV2neo, cotransfected during the construction of helper cells (Dougherty *et al.*, 1989), and the vector provirus. The relative levels of cytoplasmic *neo* RNA from each pool of infected DSN cells were about equal and were at least three times higher than that from the uninfected DSN cells (data not shown), indicating the relatively equal levels of vector *neo* RNA in each cell pool. These pools also contained relatively equal levels of *hyg* RNA (Figure 5A), which is the sum of the full-length and the spliced vector RNA, indicating that the mutations did not alter splicing of the vectors. Additionally, each transfected vector plasmid DNA (including pSY201, pSY221, pSY222 and pSY223wt) produced about equal numbers of hygromycin B-resistant colonies (data not shown). Because expression from the spliced RNA confers hygromycin B resistance, these results also indicate that the mutations did not interfere with the splicing event.

For the viral RNA from the virions, the relative levels of *neo* viral RNA were SY201:SY221:SY222:SY223wt = 1:9:6:81 (Figure 5B). The differences in relative viral RNA level among the vectors roughly correspond to the average differences in titer among the vectors. This result indicates that the mutations in SY201, SY221 and SY222 reduced the titer of the vectors through a reduction in the efficiency of packaging SNV vector RNA into virions.

A double hairpin structure necessary for retrovirus replication

The linker-scanning base substitution mutations that have a large effect on the encapsidation of SNV retroviral vector RNAs overlapped the region of the hairpin-pair motif proposed by Konings *et al.* [(1992); see Figures 1, 4 and 6A]. When the wild type E sequence was analyzed for minimal free-energy structures with the GCG FoldRNA program (Genetic Computer Group, Madison, WI), the same RNA secondary structure as described by Konings *et al.* (1992) was predicted, with two adjacent hairpins and GACG loop sequences (Figure 6A). The mutations present on SY272 were found to disrupt both hairpins, and the rest of the group 3 mutations were found to disrupt one of the hairpins by the same analysis. To test whether or not the proposed hairpin-pair motif is biologically significant, we made two mutant vectors that contain compensatory mutations to restore the base pairing within the stems, and the replication of the two vectors was examined in the CEF assay system described earlier.

SY221 contained six base substitutions within the left stem of proposed hairpin # 1. Four out of the six base changes could disrupt base pairing of the stem. Vector SY257 contained an additional four base substitution that restored base pairing (Figure 6B). SY222 had six base substitutions in the left stem of proposed hairpin # 2 that could disrupt the base pairing. Vector SY253 had six additional base changes that restored the proposed secondary structure (Figure 6B). Minimal free energy foldings of the mutant E sequences of SY257 and SY253 predicted the same local RNA secondary structures as the wild type.

The two vectors with the compensatory changes (SY257 and SY253) had relative titers comparable with that of wild type E vector virus SY223wt (Table II). This experiment was repeated several times, as indicated in Table II. The absolute value of the relative titers of the same mutants varied from one experiment to the next, probably because of the difference in transfection efficiency and the difference in cell growth of CEF cultures. However, vectors SY257 and SY253 consistently had relative titers comparable with those of SY223wt. Thus, when the base pairing was restored, the infectivity of the vector viruses was also restored. This result clearly confirms the secondary structure proposed by Konings *et al.* (1992) and demonstrates the biological significance of the structure. We designate such a structure as the double hairpin structure. Additional mutations were made to study the double hairpin structure (see below).

Table II. Relative titers of mutants containing compensatory mutations[a]

Mutants	Relative titers (normalized to SY201)	Description of mutation in E region
SY223wt	880	wild type E
SY221[c,d]	17	six base substitutions in stem 1
SY222[b,e]	11	six base substitutions in stem 2
SY257[b,d]	520	mutation in SY221 and compensatory mutation
SY253[c,e]	920	mutation in SY222 and compensatory mutation
SY201[f]	1	E deleted

[a]The data in this table are taken from one representative experiment.
[b]The mutants SY222 and SY257 were each tested independently twice more with similar results.
[c]The mutants SY221 and SY253 were each tested independently three more times with similar results.
[d]The mutants SY221 and SY257 were tested together in the same assay once more with similar results.
[e]The mutants SY222 and SY253 were tested together in the same assay once more with similar results.
[f]The absolute titer of SY201 in the assay presented here was 150 CFU/ml.

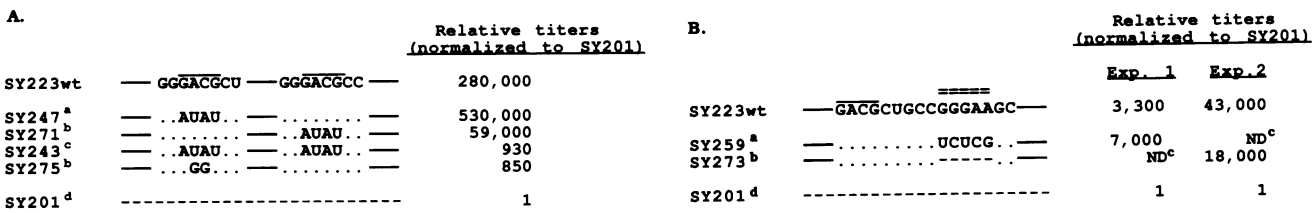

Fig. 7. (**A**) RNA sequences and relative titers of the mutant vectors containing mutations in the four-residue loops. Shown on the top is part of the wild type sequence of the double hairpin structure in SY223wt. The GACG four-residue loops are indicated with solid lines above the bases. The mutated bases in the mutant vectors are as indicated and the nucleotides that are identical to the wild type sequence are represented with dots (.). Solid lines represent intervening sequences. Deletions are represented with dashes (−). Relative titers of the vectors from one representative experiment are shown to the right of the sequences. [a]The mutant was tested independently four more times with similar results. [b]The mutants were tested together in the same assay two more times with similar results. [c]The mutant was tested independently two more times with similar results. [d]The absolute titer of SY201 in the experiment was 1 CFU/ml. (**B**) RNA sequences of the mutant vectors containing mutations at the bulge region and their relative titers. Shown on the top is part of the wild type sequence of the double hairpin structure in SY223wt. The bulge is indicated with double dashed lines above the bases. Symbols are as above. Relative titers of the vectors from two independent experiments are shown to the right of the sequences. [a]The mutant was tested independently three more times with similar results. [b]The mutant was tested once more with similar results. [c]Not determined. [d]The absolute titers of SY201 were 1 CFU/ml in experiment #1 and 20 CFU/ml in experiment #2.

Sequence elements within the double hairpin structure important for its function: four-residue loops and a bulge

The four-residue loop sequences of the double hairpin structure were found to be absolutely conserved among the type C MLV-like retroviral sequences examined (Konings *et al.*, 1992). We reasoned that the four-residue loop sequences are conserved for one or both of the following reasons: (i) they provide a recognition site at the primary sequence level for other factors, either an interacting protein or some other RNA sequence; and (ii) they are the sequences that the type C MLV-like retroviruses have evolved to form stable RNA helices, in analogy to the UNCG or GNRA tetraloops present in rRNA (N, any nucleotide; R, purines), where the hairpins form a compact structure with an unusual base pairing between the first and the last bases of the tetraloop and only two bases at the loop (Cheong *et al.*, 1990; Uhlenbeck, 1990; Woese *et al.*, 1990; Heus and Pardi, 1991).

We found that changing one of the two conserved four-residue loop sequences at a time did not have a large effect on the replication of the mutant vectors SY247 and SY271 (Figures 4 and 7A). Therefore, we made the vector SY243, which contained mutations within both four-residue loops from GACG to AUAU (Figure 7A). In addition, we made the vector SY275, which had the two middle bases of the first four-residue loop mutated from AC to two bulky bases GG (Figure 7A). If the four-residue loop sequences form a structure as proposed above, the two middle GG bases might be too bulky for the hairpin structure, which would in turn affect the stability of the hairpin. If the stability of the hairpins is affected, the vector virus titer is expected to decrease correspondingly. When the RNA sequences of the mutant vectors in Figure 7A were analyzed with the GCG program FoldRNA (which predicts the minimal free energy structure), all folded into a predicted structure that was the same as the predicted double hairpin structure of the wild type SNV sequence, with the desired mutations at the four-residue loops. In addition, the RNA sequence of the vector SY275 was examined with the GCG program MFold (which predicts the optimal and suboptimal structures), and the same structure was found to be optimal.

The mutant vectors containing various mutations at the four-residue loops were tested (Figure 7A). We found that mutating both of the four-residue loop sequences at once drastically reduced the relative titer of the mutant vector virus SY243 compared with the relative titer of the vectors with singly mutated four-residue loops. The factor of reduction is ~300 in the experiment shown in Figure 7A. In general, this vector replicated similarly to a mutant vector containing group 3 mutations. This result indicates that the presence of GACG in at least one of the two four-residue loops is sufficient for the recognition/interaction. When the two middle bases of the first four-residue loop were changed to GG, the mutant vector SY275 had a much reduced relative titer (a factor of ~350 lower than that of the wild type E

vector in Figure 7A), similar to a mutant vector containing group 3 mutations. Results with the vector SY275 are consistent with the idea that one of the major roles of the GACG four-residue loops is to provide increased thermodynamic stability to the double hairpin structure. Current energy rules are limited in their ability to estimate different contributions from different four-residue loop sequences (Jaeger *et al.*, 1989a). Our result with the mutant SY275 indicates that the two four-residue loop sequences GACG and GGGG are probably not equivalent energetically.

The bulge sequences at the right stem of hairpin #1 (Figure 6A) are moderately conserved among the type C MLV-like retroviruses examined (Konings *et al.*, 1992). Bulges are present at similar positions of hairpin #1 of these viruses and are composed of purine-rich sequences of three to five bases. The same region was implicated in the dimerization of REV-A retroviral RNA *in vitro* (Darlix *et al.*, 1992). Mutants SY259 and SY273 (Figure 7B) were made to test the function of the bulge sequences. Vector SY259 had a five base substitution mutation at the bulge, with four out of the five bases as pyrimidines rather than purines. These were the same bulge mutations present in the vector SY213 (containing a group 3 mutation; Figures 1 and 4), except that the linker-scanning mutation in SY213 also included three additional base changes in the stem of hairpin #1. Vector SY273 had the five base bulge sequences deleted. The two mutant vectors were tested independently and were found to replicate comparably with the wild type E vector SY223wt (Figure 7B). These results indicate that the bulge sequences are not required as part of the function of the double hairpin structure.

Changing the relative order of the two hairpins and duplication of hairpin #2

Since the bulge sequence of hairpin #1, which clearly differentiates the two hairpins, was not required for function, we were interested in testing whether or not the two hairpins are functionally similar. We therefore constructed vectors that had only one copy of hairpin #1 or hairpin #2 (SY251 and SY269, respectively), had the order of the two hairpins switched (SY270), or had two copies of hairpin #2 (SY285). In the vector SY285, the sequence of hairpin #1 was replaced by that of hairpin #2, and the sequence of hairpin #2 was replaced by that of hairpin #2 from the vector SY253, which contains compensatory mutations in hairpin #2 (Figure 6B). The sequence of hairpin #2 from SY253 is functionally equivalent to that of the wild type sequence, as judged from the comparable relative titers of the vector SY253 and the wild type vector SY223wt (Table II). This sequence was chosen so that the mutant vector SY285 would have only a limited number of direct repeats in the region (one six base direct repeat and two three base direct repeats), because short direct repeats are a hot spot for deletions during retrovirus replication (Pathak and Temin, 1990, 1992). Retroviral vectors that had only one of the two hairpins (mutant vectors SY251 and SY269) had similar relative titers, and the relative titers were at least a factor of 10−100 less than those of the vector containing wild type E SY223wt (Table IIIA). These relative titers were comparable with most of the mutant vectors containing group 3 linker-scanning mutations that disrupted the base pairing of either of the two hairpin stems (data not shown). Vector SY270, which had both hairpins present but in the reverse order, gave relative titers comparable with that of the wild type E vector SY223wt (Table IIIA). The vector containing two copies of hairpin #2 (SY285) gave relative titers higher than that of the vector with a single hairpin #2 (SY269), but lower than that of the wild type vector SY223wt by a factor of at least 20 (Table IIIB). The presence of a second copy of hairpin #2 can only partially restore the titer loss of the mutant vector with a single hairpin #2, indicating that the two hairpins share some functional similarity, but they are not identical.

Mutant vectors were also made that increased the spacing between the two hairpins. The vectors gave lower titers relative to that of the wild type E vector (data not shown). However, analysis with the GCG FoldRNA program indicates that the mutant vector RNA sequences would not fold into a structure with the same two hairpins. Thus, the presence of the additional bases between the two hairpins may have altered the local RNA secondary structure and destabilized the hairpins, and the real effect of separating the two hairpins (without destabilizing the structure) cannot be assessed from these vectors.

Sequence-specific information within the stems of the double hairpin structure

Because the two vectors containing compensatory mutations (SY257 and SY253) replicated well (Table II), the primary sequences changed within the stems of the double hairpin structure in these two mutants apparently are not required for the function of the hairpins. To address whether or not the entire stem regions contain any sequence-specific information, we constructed the 'base switched' mutant vector SY284. SY284 differs from SY223wt at every position within the stem regions, but retains the same double hairpin structure and the same sequences at the bulges and the loops (Figure 6C). The free energy of the structure in vector SY284 is similar to that of the wild type (−19.3 kcal/mol for the mutant versus −18.7 kcal/mol for the wild type). Analysis with the GCG Mfold program predicted the desired structure as optimal. This vector was tested in the CEF assay system described above. The relative titers of mutant SY284 were reduced relative to that of SY223wt by a factor of >100 (Table IIIC). This result indicates that the stem regions of the hairpins contain specific sequences critical for encapsidation. It is possible that the sequence substitution disrupted specific recognition of some *trans* factors, or some unknown tertiary structure was affected by the substitution.

Sequences downstream of the double hairpin structure required for encapsidation

The *Xho*I linker-scanning mutations also defined a region downstream of the double hairpin structure region as important for encapsidation, as marked by the mutant vectors SY230, SY231 and SY232 (Figure 4). The *in vitro* dimerization studies also mapped part of this region to be necessary for dimerization (Darlix *et al.*, 1992). A mutant vector (SY279) was constructed that contained only the encapsidation sequence up to the end of the double hairpin structure. This vector is different from the 3′ half E deletion vector studied by Embretson and Temin (1987) (JE129 5′), in that the latter lacks the last three bases of hairpin #2. The titer of the mutant vector SY279 was a factor of ~20 lower than that of SY223wt (Table IV). The magnitude of

Table III. Relative titers of hairpin mutants[a]

Mutants	Relative titers (normalized to SY201)	Description of mutation in E region
A		
SY223wt	15 000	wild type E
SY251[c,d]	190	one copy of hairpin #1
SY269[b,d,e]	260	one copy of hairpin #2
SY270[b,e]	3700	the order of the hairpins switched
SY201[f]	1	E deleted
B		
SY223wt	28 000	wild type E
SY269[g]	60	one copy of hairpin #2
SY285[c,g]	1200	two functional copies of hairpin #2
SY201[h]	1	E deleted
C		
SY223wt	2900	wild type E
SY284[c]	18	changes at every base in the stems
SY201[i]	1	E deleted

[a]The data in this table are taken from three separate representative experiments.
[b]The mutants SY269 and SY270 were each tested independently once more with similar results.
[c]The mutants SY251, SY284 and SY285 were each tested independently twice more with similar results.
[d]The mutants SY251 and SY269 were tested in the same assay once more with similar results.
[e]The mutants SY269 and SY270 were tested in the same assay once more with similar results.
[f]The absolute titer of SY201 in the assay presented here was 23 CFU/ml.
[g]The mutants SY269 and SY285 were tested in the same assay once more with similar results.
[h]The absolute titer of SY201 in the assay presented here was 25 CFU/ml.
[i]The absolute titer of SY201 in the assay presented here was 430 CFU/ml.

Table IV. Relative titers of the mutant SY279[a]

Mutants	Relative titers (normalized to SY201)[b]	Description of mutation in E region
SY223wt	3400 ± 820	wild type E
SY279	170 ± 4	sequences downstream of the hairpins deleted
SY201[c]	1	E deleted

[a]The data in this table were taken from two independent experiments.
[b]Relative titers are presented as average ± standard deviation.
[c]The average absolute titer of SY201 was 280 CFU/ml in the two assays presented here.

reduction in relative titer of this deletion vector is similar to that of the mutant vectors SY230, SY231 or SY232 containing linker-scanning mutations in this region (Figure 4), indicating that each linker-scanning mutation disrupted the signal in this region to about the same degree as the deletion vector. The exact nature of the signal within the region is not clear.

Replication of full-length SNV containing the E mutations

To assess the function of the double hairpin structure on the replication of full-length replication-competent SNV, we subcloned the linker-scanning mutations from SY221 and SY222 into a full-length SNV clone, resulting in the plasmids pSYSNV-221 and pSYSNV-222. The mutations each disrupted one of the two hairpins in the two full-length SNV mutants. As a control, we also used a full-length clone with the E sequence deleted (pSYSNVE$^-$). We tested replication of the full-length SNV clones by first transfecting serially-diluted plasmid DNAs into CEF, infecting fresh CEF with supernatants from the transfected CEF, and then assaying the supernatants of the infected CEF for reverse transcriptase (RT) activity. The end points of plasmid DNA dilutions that gave positive RT assay results were compared. Carrier DNA (8.8 μg/plate) was used for the transfection and we allowed 5 or 6 days before harvesting supernatants from transfected or infected CEF, respectively. When the E region was deleted, RT assays were negative even when 5 μg of pSYSNVE$^-$ plasmid DNA was used for the initial transfection (data not shown), indicating that E$^-$ SNV does not replicate. The end points of positive RT activity were 10 – 100 times higher for the full-length SNV DNA with six base substitutions within E than for the wild type SNV DNA (data not shown), indicating that the efficiency of replication of the mutants was crippled. In addition, the titers of the mutant SNV harvested from end point transfected CEF were at least three orders of magnitude lower than those of the wild type at the same dilution (data not shown). At the same time, crude cell lysates were made from a portion of the infected CEF and were subject to PCR, followed by *Xho*I endonuclease digestion. *Xho*I digestion revealed that the *Xho*I site was present in almost all of the PCR DNA (data not shown), confirming the presence of the introduced mutations. Viruses harvested from the infected CEF were also used for a second round of infection. The end points of infection were compared. The titers of the mutant SNV were lower by a

Table V. Relative titers of mutants with different MLV sequences[a]

Mutants	Relative titers (normalized to SY201)		Description of mutation in E region
	exp. 1	exp. 2	
SY223wt	800 000	4000	wild type E
SY254[c]	430 000	ND[d]	SNV hairpins replaced by those from M-MLV
SY280[b]	ND[d]	2600	SNV E replaced by ψ
SY281[b]	ND[d]	3200	SNV E replaced by ψ^+
SY201[e]	1	1	E deleted

[a]The data in this table were taken from two separate experiments.
[b]The mutants SY280 and SY281 were each tested independently once more with similar results.
[c]The mutant SY254 was tested independently twice more with similar results.
[d]Not done.
[e]The absolute titers of SY201 were 1 CFU/ml in experiment #1 and 450 CFU/ml in experiment #2.

factor of 10−1000 compared with the wild type SNV (data not shown). Cell lysates from the second round of infected CEF were made, amplified by PCR and digested with *Xho*I. Again, most of the PCR DNAs were completely digested by *Xho*I (data not shown). These results indicate that the base substitution mutations within the double hairpin structure, in the context of full-length SNV, compromised viral replication.

The specificity of the hairpins from SNV and M-MLV, as well as the specificity of the retroviral encapsidation sequences of SNV, M-MSV/M-MLV, HIV-1 and BLV

Despite the low level of similarity in the primary sequence of the double hairpin structures of SNV and M-MLV, they are extremely similar structurally (Konings *et al.*, 1992), raising the possibility that the two sequences are interchangeable. We replaced the double hairpin structure of SNV in our SNV-based vector with that of M-MLV, resulting in the mutant vector SY254. As a control for the experiment, SNV vectors were also constructed that had the entire E sequence of the SNV-based vector replaced by the ψ of M-MSV or the ψ^+ sequences of M-MSV and M-MLV from pLN (Miller and Rosman, 1989), resulting in the mutant vectors SY280 and SY281. Both ψ and ψ^+ included the sequences of the double hairpin structure from M-MSV (Miller and Rosman, 1989).

The mutant vectors SY254, SY280 and SY281 were tested in the CEF assay (Table V). Surprisingly all three vectors had relative titers comparable with that of SY223wt, which contains the wild type E sequence from SNV. These results clearly indicate that the E or ψ sequences for SNV and MLV are not virus specific. Since the sequences tested ranged from 200 to ~800 bases, we conclude that, during packaging of MLV and SNV retroviral vector RNAs with REV-A retroviral proteins, the encapsidation/packaging signals are recognized largely by the secondary or tertiary structures within the region.

We also replaced SNV E in our SNV-based vector with the putative encapsidation sequences from HIV-1 and BLV, resulting in the mutant vectors SY286, SY287 and SY288. The HIV-1 E sequence in SY286 starts immediately after the poly(A) signal in R and ends at the end of the p17^{gag} coding sequence. It contains most of the sequences so far identified as important for HIV-1 RNA encapsidation (Lever *et al.*, 1989; Aldovini and Young, 1990; Clavel and Orenstein, 1990; Harrison and Lever, 1992; Hayashi *et al.*, 1992; Rizvi and Panganiban, 1993). The two BLV fragments used to replace SNV E in vectors SY287 and SY288 were different only in the length of the R region included. They start either at the beginning of R or immediately after the major splice donor site in R, and extend 557 bases into the *gag* coding region. These two fragments contained the regions important for binding to BLV MA proteins and for dimerization of BLV RNA *in vitro* (Katoh *et al.*, 1991, 1993) and probably contain most of the sequences required for efficient packaging of BLV RNA (K.Boris-Lawrie and H.M.Temin, manuscript in preparation; L.M.Mansky and H.M.Temin, manuscript in preparation). These chimeric vectors were tested with the CEF assay described earlier. The presence of *gag* coding regions within the encapsidation sequences we inserted in our SNV vectors would potentially disrupt the proper *neo* gene expression; therefore we chose to select for hygromycin B resistance, which is expressed from the spliced viral RNA. As an assay for the *hyg* gene expression, we transfected DSN cells with equal amounts of vector plasmid DNAs and selected for hygromycin B-resistant colonies. The number of colonies were approximately a factor of three to eight times lower in the vectors containing HIV-1 or BLV E sequences than in those containing wild type SNV E (data not shown). Unlike the ψ and ψ^+ sequences of M-MSV and M-MLV, neither HIV-1 nor BLV E sequences rescued our E$^-$ SNV vector for efficient replication (data not shown).

Discussion

We report here mutational studies of regions within the previously defined E sequence of SNV. We identified two regions within the SNV E necessary for efficient replication of SNV-based retroviral vectors. The first region is a double hairpin structure as proposed by Konings *et al.* (1992); the second region is located downstream of the hairpins. Noteworthy is the fact that the two regions defined by the linker-scanning mutations are located in the 5′ and the 3′ halves of the E sequence, respectively; the same regions were discovered through deletion analysis to be necessary for SNV retroviral replication (Embretson and Temin, 1987). We found that preservation of the double hairpin structure is crucial for the function, and we present evidence for sequence-specific information within part of the stems of the hairpins. Our data indicate that: (i) the presence of the GACG

sequence in at least one of the four-residue loop sequences of the hairpins is required for function; (ii) the moderately conserved bulge sequence of hairpin # 1, as well as (iii) the relative order of the two hairpins, does not have a large effect on the function; and (iv) the two hairpins are not functionally identical. The double hairpin structure is crucial for the efficient replication of both SNV retroviral vectors and full-length SNV. In addition, we found that, in our system, the double hairpin structure of M-MLV is functionally equivalent to that of SNV, and the packaging signals of M-MSV and M-MLV (ψ and ψ^+), but not those of HIV-1 and BLV, can replace SNV E for efficient replication of SNV vectors.

We examined the relative levels of packaging of three mutant vectors, the E^- vector SY201 and the vectors SY221 and SY222 which were defective in replication, and the wild type E vector SY223wt. The effect of the mutations was at the level of packaging of SNV vector RNA. The majority of the other mutant vectors containing various linker-scanning mutations, especially those that drastically reduced the titers of the mutant vectors, were tested for expression by transfecting into D17 cells and DAN or DSN helper cells, followed by selection with G418 or hygromycin B, respectively. We obtained about equal numbers of G418-resistant or hygromycin B-resistant colonies for each vector (data not shown). Furthermore, when the plasmids SY201 and SY223wt were tested along with the mutant vectors, similar numbers of drug-resistant colonies were observed (data not shown). These results indicate that most of the mutations we introduced in the E region of SNV did not greatly affect the steps of transcription, translation or splicing. We think that they, like the mutant vectors SY201, SY221 and SY222, lowered the efficiency of encapsidation of vector SNV RNA.

Various RNA secondary structures have been proposed for the encapsidation/packaging sequences of several different retroviruses, including MLV, MLV-related viruses, RSV, HIV and SIV. The structures proposed are mostly based on chemical/enzymatic modification analysis of retroviral RNA, deletion analysis of retroviral RNA sequences, RNA secondary structure modeling based on comparative sequence analysis, or combinations of the above methods (Prats *et al.*, 1990; Alford *et al.*, 1991; Harrison and Lever, 1992; Hayashi *et al.*, 1992; Konings *et al.*, 1992; Tounekti *et al.*, 1992; Rizvi and Panganiban, 1993). Our study confirmed the secondary structure proposed by Konings *et al.* (1992) and is the first to demonstrate, via linker-scanning and site-directed mutagenesis, that a specific RNA secondary structure is required for the encapsidation of retroviral RNA.

We found that the double hairpin structures from SNV and M-MLV are functionally equivalent. The M-MLV double hairpin structure is consistent with previous reports of enzymatic and chemical probing of the M-MLV viral RNAs (Alford *et al.*, 1991; Tounekti *et al.*, 1992). The SNV E sequence, when placed in the reversed orientation, did not confer efficient encapsidation (Embretson and Temin, 1987; and data not shown); nor was dimerization efficient *in vitro* (Darlix *et al.*, 1992). One difference in the reversed sequence, even if a similar hairpin structure exists, is that neither of the four-residue loop sequences would be GACG. Our data indicate that GACG is required as at least one of the two four-residue loop sequences. Thus, the data agree with these previous observations. Furthermore, the results from mutant SY275 (which changed the first four-residue loop from GACG to GGGG) lead us to propose that, in addition to functioning possibly as the recognition site for some proteins and/or other RNA sequences, the GACG four-residue loop is probably involved directly in the stabilization of the RNA helices, in analogy to proposed functions of the tetraloops that are confined to certain consensus motifs in rRNA (Cheong *et al.*, 1990; Uhlenbeck, 1990; Woese *et al.*, 1990; Heus and Pardi, 1991). In fact, chemical analysis of the M-MLV retroviral RNA indicates that only the second base of the GACG four-residue loop sequence is marginally reactive to modification agents (Alford *et al.*, 1991; Tounekti *et al.*, 1992). These results are similar to the chemical modification data on rRNA (Moazed *et al.*, 1986), suggesting that the GACG four-residue loops have a function similar to that of the tetraloops.

The bulge sequence of hairpin # 1 of the SNV E is dispensable in the context of otherwise wild type E sequence. M-MLV has a bulge sequence of GGAA (Konings *et al.*, 1992). Alford *et al.* (1991) and Prats *et al.* (1990) examined a total of 16 MLV-like murine retrovirus sequences, and within hairpin # 1 the bulge region had the most variations. This finding appears to agree with a less stringent role for the bulge sequence. However, the bulge sequence is one of the four purine-rich regions implicated in REV-A dimer formation *in vitro* (Darlix *et al.*, 1992). It is possible that the bulge sequence contributes to dimer formation and/or some other functions, but the effect is not obvious in these experiments in the context of otherwise wild type surrounding sequences.

The 'base switched' mutant SY284 consistently gave titers lower than those of SY223wt by a factor of at least 100. This mutant was designed to have the same double hairpin structure as the wild type, the free energy of the structure is similar to that of the wild type, and the analysis with the GCG Mfold program predicted the desired structure as optimal. As compared with the Mfold folding results of the wild type E sequence, the mutant sequence potentially has more alternative suboptimal structures. We believe that the desired structure of the 'base switched' mutant SY284 does form based on the computer modeling, and the fact that it replicated poorly compared with the wild type indicates the presence of sequence-specific information within the stems of the hairpins. This result does not contradict the result that SNV and M-MLV hairpins are functionally similar, since SNV and MLV hairpins do share some identical bases at similar positions within the stems. It is possible that a few of these bases are necessary for the function of the hairpins. An example of the requirement for sequence conservation in a stem of a hairpin structure is the recognition of the *trans*-acting-responsive (TAR) element of HIV-1 by the HIV-1 Tat protein. The HIV-1 TAR forms a bulged hairpin structure (Muesing *et al.*, 1987). Within the TAR element, high-affinity binding with the viral Tat protein requires a U residue in the bulge and two specific adjacent base pairs (Weeks and Crothers, 1991).

The mechanisms through which the double hairpin structure modulates SNV retroviral RNA encapsidation are a matter of speculation. The same double hairpin structure region was found to be important for dimerization of REV-A RNA *in vitro* (Darlix *et al.*, 1992). The region encompassing a similar structure in M-MLV was shown to be important for *in vitro* dimerization of MLV RNA (Prats

et al., 1990). One possible mechanism is that the E mutant vectors are deficient in RNA dimerization. It will be interesting to examine the efficiency of *in vitro* dimerization of SNV mutant E RNAs. Such studies may lead to a better understanding of the link between encapsidation and dimerization. It is also possible that the E mutant vectors are defective in replication steps independent of, or in addition to, RNA dimerization.

Retroviruses can package heterologous retroviral RNA from within the same genus—for instance, MLV RNA by SNV and SIV RNA by HIV-1—although often with somewhat decreased efficiency (Embretson and Temin, 1987; Takeuchi *et al.*, 1992; Rizvi and Panganiban, 1993). Our finding that the replacement of SNV E by ψ or ψ^+ of MSV and MLV resulted in efficient replication of the SNV-based vectors extended the previous work. Our system is different from those mentioned above in that the fragments containing the encapsidation sequences from a heterologous virus were tested in an otherwise wild type SNV-based vector for packaging by REV-A proteins, as opposed to using an entirely different vector based on the heterologous virus. Thus, our system avoids possible variations in relative transcription levels of different virus-based vectors and any effects the different sequences outside the defined E region might have on encapsidation. MLV and SNV do not share apparent sequence homology in the E region, nor do HIV-1 and SIV. The facts that the E sequences extend over a stretch of a few hundred nucleotides and that apparently different sequences are functionally equivalent indicate that the packaging signals primarily reside within secondary and tertiary structures. HIV-1 and SIV E sequences were proposed to share a similar structural element with each other (Rizvi and Panganiban, 1993). Previously, HIV-1 vectors were shown not to be pseudotyped by MLV proteins (Delwart *et al.*, 1992) and our data showed that REV-A proteins failed to propagate SNV-based vectors containing HIV-1/BLV E sequences. One possible interpretation is that significant differences exist between the E sequences from viruses of different genera; however, other virus replication processes could also have been affected in these complicated chimeric vectors. In addition to the dimerization sequence, the encapsidation sequence region appears to include the recognition sequence/structure for MA proteins (Katoh *et al.*, 1991), NC proteins (Sakaguchi *et al.*, 1993) or the NC domain of the Gag polyproteins (Luban and Goff, 1991). Encapsidation of retroviral RNA most probably proceeds through specific protein–RNA recognition. The NC domain of the Gag polyprotein and the NC protein have been shown to be the major determinants in the packaging process *in vivo* and RNA dimerization *in vitro* (Méric and Spahr, 1986; Gorelick *et al.*, 1988; Méric *et al.*, 1988; Méric and Goff, 1989; Aldovini and Young, 1990; De Rocquigny *et al.*, 1992; Dupraz and Spahr, 1992; Housset *et al.*, 1993; Sakaguchi *et al.*, 1993; Sundquist and Heaphy, 1993). These specific interactions probably provide a basis for the specificity of RNA encapsidation. Since the NC protein of SNV is similar to that of MLV, efficient encapsidation of MLV E^+ RNA into REV-A virions, but not the HIV-1 or BLV E^+ RNA, is consistent with the above notion.

Materials and methods

Nomenclature

Names that start with a 'p', such as pSY223wt, designate plasmids, whereas those without a 'p', such as SY223wt, denote viruses derived from the plasmids.

Cells

CEF are primary cells highly permissive for infection by REV, including REV-A and SNV. Cells were from 11 day-old chick embryos (SPAFAS Inc., Norwich, CT) and had been through about five or six passages by the time they were used for transfection or infection. D17 cells are a dog osteosarcoma cell line, also permissive for REV infection. DSN and DAN cells are helper cell lines derived from D17 cells (Dougherty *et al.*, 1989) and contain the *neo* gene as the selectable marker (Southern and Berg, 1982).

CEF were maintained in Temin-modified Eagle's medium (Temin, 1968) supplemented with 20% tryptose phosphate broth (TP), 4% calf serum and 2% fetal bovine serum. D17 cells, as well as DSN and DAN helper cell lines, were maintained in Temin-modified Eagle's medium supplemented with 5–6% calf serum. All cells were maintained at 37°C with 6% CO_2. G418 and hygromycin B selections were performed at 600 and 80–120 μg/ml, respectively.

Infection and transfection

One day prior to infection or transfection, cells were plated at a density of 2×10^5 cells per 60 mm tissue culture Petri dish. Fresh virus stocks were made by clarifying the supernatant medium from CEF or appropriate virus-producing helper cells by centrifugation at 1600 *g* for 10 min. Infections were performed by incubating the cells with freshly harvested virus in the presence of polybrene (final concentration 7.5 μg/ml for CEF and 50 μg/ml for D17 cells or D17-derived helper cells), in a final volume of 0.4 ml per Petri dish at 37°C for 1 h, and then replacing the medium with fresh medium. Transfections were performed using the dimethyl sulfoxide-polybrene method (Kawai and Nishizawa, 1984). D17 cells or D17-derived helper cells were maintained in selective medium starting 1 day after infection or transfection.

Mutagenesis and plasmid construction

Construction of mutant E vectors. All E mutant vectors were made by direct cloning of PCR-amplified DNA. To introduce the desired mutations, a primary/combinational 2-step PCR protocol was used as described previously (Ito *et al.*, 1991) with one modification, unless otherwise noted below. We modified the design of the internal common primer so that it contained an additional six mismatched bases at its 5′ end, to increase the percentage of DNA containing the desired mutations in the final pool of combinational DNA amplified by PCR.

All PCRs were done with 14–20 cycles. The reactions were performed in a final concentration of 10 mM Tris–HCl (pH 8.3, 20°C), 1.5 mM $MgCl_2$, 50 mM KCl and 0.1 mg/ml (0.01%) gelatin. Taq DNA polymerase was obtained from Boehringer Mannheim (Indianapolis, IN). In the construction of mutant vectors, the final PCR DNA was digested with *Kpn*I and *Sal*I and was used to replace the *Kpn*I–*Sal*I E fragment of an SNV vector, pSY210. pSY210 is similar to the SNV splicing vector pJD216NeoHy (Dougherty and Temin, 1986), except that the polylinker sequences at the junction of the E sequence and the *neo* gene fragment contain a *Sal*I site. Standard cloning protocols were used (Sambrook *et al.*, 1989).

The template for the SNV E mutant vectors was pPB101, an infectious provirus clone of SNV (Bandyopadhyay and Temin, 1984). Noteworthy is the 'base switched' mutant vector pSY284, where the majority of the bases within the stems of the hairpins were changed to the bases complementary to the bases in the wild type. However, at 11 positions this was not possible for one of three reasons: (1) non-Watson–Crick base pairs were present in the wild type; (ii) we wanted to ensure that the overall free energy of the structure was comparable with that of the wild type; or (iii) we wanted to eliminate potential alternative structures as predicted by computer analysis. The final mutant sequence is predicted to have the best structure by the above criteria.

The template for vectors pSY280 and pSY281 was the MLV vector pLN (Miller and Rosman, 1989). A single round of PCR was performed with appropriate MLV primers and the PCR DNAs containing ψ or ψ^+ were cloned directly into pSY210. ψ is a sequence ~330 bases long, necessary for the encapsidation/packaging of M-MSV/M-MLV viral RNA. It starts shortly after the splice donor site of M-MSV and ends before the beginning of the *gag* open reading frame (Mann *et al.*, 1983). ψ^+ consists of ψ and an additional 418 bases into the *gag* gene of M-MLV. MLV retroviral RNAs containing ψ^+ are more efficiently packaged than those containing ψ (Armentano *et al.*, 1987; Bender *et al.*, 1987). The SNV E was replaced by either ψ or ψ^+ encapsidation/packaging sequences of M-MSV and M-MLV in the vectors.

The templates for vector pSY286 were pME337, an HIV_{BRU} isolate with a *Bgl*II deletion (nucleotides 6634–7214) in the *env* gene (kindly provided by Dr M.Emerman, Fred Hutchinson Cancer Research Center, Seattle, WA) and p37M1234, a partial HIV-1 clone containing various point mutations in the inhibitory elements present in the $p17^{gag}$ region (Schwartz *et al.*, 1992) (kindly provided by Dr B.Felber, NCI, Frederick, MD). Combinational PCR DNAs were made starting from the 3′ end of R,

immediately after the poly(A) signal, to the end of the p17^{gag} coding sequence. The PCR DNA contained the putative HIV-1 E region, including the entire U5, PBS, major splice donor site, leader sequence and 396 nucleotides of the *gag* coding sequences. Other than the mutations in the inhibitory elements, the entire region is wild type. The putative HIV E replaced the SNV E in this vector.

The template for vectors pSY287 and pSY288 was pBLV-SVNEO (Derse and Martarano, 1990). DNA from a single round of PCR with appropriate BLV primers was cloned directly into pSY210. The BLV putative E (nucleotides 212–1183, starting from the beginning of R and extending 557 bases into the *gag* coding region, or nucleotides 308–1183, starting from immediately after the major splice donor site in R and ending at the same position) replaced the SNV E in these vectors. [The poly(A) signal of BLV is present in U3 (Sagata *et al.*, 1985).]

The primer sequences for the PCR and the detailed PCR cloning protocols are described elsewhere (Yang, 1993).

Full-length SNV containing the E mutations. The full-length SNV E mutant proviral clones pSYSNV221 and pSYSNV222 were made by inserting PCR DNA containing the desired mutations into pSYSNV, a full-length infectious SNV clone with no extra duplications. The mutants contained the E mutations from the mutant vectors pSY221 and pSY222. The details of the multi-step cloning strategies are discussed elsewhere (Yang, 1993). pSYSNVE$^-$ has the *NruI–SalI* E fragment deleted from pSYSNV. (The *NruI* site is located immediately 3′ to the *KpnI* site of SNV E.)

Verification of the mutants. All the mutant vectors, except pSY280, pSY281, pSY286, pSY287, pSY288 and the full-length SNV E mutants, were sequenced throughout the region of DNA cloned from PCR. Some of the mutant vectors had a single base mismatch in the E region downstream of the double hairpin structure, but none had any secondary mutations in the region containing the double hairpin. The identities of the rest of the mutant vectors were verified by sequencing part of the cloned DNA. All the sequencing was performed on plasmid DNAs with the Sequenase kit 2.0 (US Biochemical, Cleveland, OH).

Verification of the proviruses. To verify the identity of the provirus in the cells, crude cell lysates were prepared as described (Higuchi, 1989). 5–10 μl of each lysate were amplified directly for 40 cycles. The PCR DNA was then digested with appropriate restriction endonucleases.

Preparation of cytoplasmic and viral RNA

Cytoplasmic RNAs were prepared as described in Berger and Birkenmeier (1979). Viral RNAs were prepared from freshly harvested virus. Virus stocks were clarified by centrifugation twice at 1600 *g* for 10 min to remove cell debris, followed by centrifugation at 25 000 *g* for 90 min through a 20% sucrose cushion to pellet virus. Pelleted virus was then resuspended in buffer containing 150 mM NaCl, 10 mM Tris (pH 7.4), 1 mM EDTA, 0.5% SDS and 100 μg of tRNA/ml. Proteinase K was immediately added to a final concentration of 50 μg/ml and the resuspended virus samples were allowed to incubate at 37°C for 15–20 min before being quickly frozen. The viral RNA samples were then purified with standard protocols of phenol/chloroform extraction and ethanol precipitation (Sambrook *et al.*, 1989).

RNA dot blotting, hybridization and RNA quantification

Cytoplasmic and viral RNAs were denatured in buffer containing a final concentration of 6 × SSC and 7.4% (w/v) formaldehyde at 66°C for 15 min. Samples were 3-fold serially diluted in 15 × SSC and were blotted onto nitrocellulose paper (Schleicher & Schuell, Keene, NH) with a slot blot manifold (Millipore, Bedford, MA). The nitrocellulose paper was then baked at 80°C for 2 h under vacuum to fix the RNA.

The *hyg* and *neo* gene probes used were randomly primed (Random Primed DNA Labeling Kit, Boehringer Mannheim, Indianapolis, IN) with DNA fragments embedded in a low melting point agarose gel. The *hyg* probe was the smallest fragment from an *Eco*RI and *Cla*I digestion of pJD216NeoHy. The *neo* probe was the entire *neo* gene digested from a pUC19-based *neo* subclone.

Hybridization was performed overnight at 42°C in buffer containing a final concentration of 50% formamide, 3 × SSC, 20 mM HEPES (pH 7.4), 0.4% SDS, 2 × Denhardt's solution, 100 μg/ml tRNA and 100 μg/ml sheared salmon sperm DNA. The blots were washed at room temperature twice with 2 × SSC/0.1% SDS for 5 min each, once with 0.5 × SSC/0.1% SDS for 15 min, and once with 0.1 × SSC/0.1% SDS for 15 min. The final wash was 0.1 × SSC/1% SDS at 42°C for 30 min.

The relative intensities of the bands on the blots were quantified with a Phosphorimager (Molecular Dynamics, Sunnyvale, CA).

RT assay

Cell supernatants were collected and subjected to reverse transcriptase assay as previously described (Temin and Kassner, 1974).

Minimal free energy folding of mutant vector E regions: optimal and suboptimal foldings

The GCG programs (version 8; ECG, Madison, WI) FoldRNA (Zuker and Stiegler, 1981) and MFold (Jaeger *et al.*, 1989a,b; Zuker, 1989) were used for the analysis of minimal free energy and suboptimal RNA secondary structures of the E sequences of the wild type and mutant vectors. The RNA structures predicted by FoldRNA and MFold were plotted with the GCG programs Squiggles and PlotFold, respectively. The calculation of minimal free energy was performed with the Turner energies (Freier *et al.*, 1986).

Acknowledgements

We thank D.Bettner, X.-J.Bi, A.Krueger, B.Pietz, J.Schoening and B.Seufzer for excellent technical assistance; K.Boris-Lawrie, J.Jones, L.Mansky, G.Pulsinelli and M.Wickens for helpful discussions; K.Boris-Lawrie, D.P.W.Burns, V.Hinshaw, J.Jones, L.Mansky, S.McBride, A.Panganiban, G.Pulsinelli and I.Riegel for critical reading of the manuscript; and N.Drinkwater for help with statistics. We also thank K.Boris-Lawrie and L.Mansky for discussing results prior to publication. We are grateful to A.Palmenberg for many discussions concerning RNA secondary structures and the use of various folding programs. This work was supported by grants CA22443 and CA07175 from the National Cancer Institute. H.M.T. is an American Cancer Society research professor.

References

Adam,M.A. and Miller,A.D. (1988) *J. Virol.*, **62**, 3802–3806.
Aldovini,A. and Young,R.A. (1990) *J. Virol.*, **64**, 1920–1926.
Alford,R.L., Honda,S., Lawrence,C.B. and Belmont,J.W. (1991) *Virology*, **183**, 611–619.
Armentano,D., Yu,S.-F., Kantoff,P.W., von Ruden,T., Anderson,W.F. and Gilboa,E. (1987) *J. Virol.*, **61**, 1647–1650.
Aronoff,R. and Linial,M. (1991) *J. Virol.*, **65**, 71–80.
Aronoff,R., Hajjar,A.M. and Linial,M.L. (1993) *J. Virol.*, **67**, 178–188.
Bandyopadhyay,P.K. and Temin,H.M. (1984) *Mol. Cell. Biol.*, **4**, 743–748.
Bandyopadhyay,P.K., Watanabe,S. and Temin,H.M. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3476–3480.
Bender,M.A., Palmer,T.D., Gelinas,R.E. and Miller,A.D. (1987) *J. Virol.*, **61**, 1639–1646.
Bender,W., Chien,Y.-H., Chattopadhyay,S., Vogt,P.K., Gardner,M.B. and Davidson,N. (1978) *J. Virol.*, **25**, 888–896.
Berger,S.L. and Birkenmeier,C.S. (1979) *Biochemistry*, **18**, 5143–5149.
Bieth,E., Gabus,C. and Darlix,J.-L. (1990) *Nucleic Acids Res.*, **18**, 119–127.
Casella,C.R. and Panganiban,A.T. (1993) *Virology*, **192**, 458–464.
Cheong,C., Varani,G. and Tinoco,I.,Jr (1990) *Nature*, **346**, 680–682.
Clavel,F. and Orenstein,J.M. (1990) *J. Virol.*, **64**, 5230–5234.
Darlix,J.-L., Gabus,C., Nugeyre,M.-T., Clavel,F. and Barré-Sinoussi,F. (1990) *J. Mol. Biol.*, **216**, 689–699.
Darlix,J.-L., Gabus,C. and Allain,B. (1992) *J. Virol.*, **66**, 7245–7252.
De Rocquigny,H., Gabus,C., Vincent,A., Fournié-Zaluski,M.-C., Roques,B. and Darlix,J.-L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.
Delwart,E.L., Buchschacher,G.L.,Jr, Freed,E.O. and Panganiban,A.T. (1992) *AIDS Res. Hum. Retroviruses*, **8**, 1669–1677.
Derse,D. and Martarano,L. (1990) *J. Virol.*, **64**, 401–405.
Dougherty,J.P. and Temin,H.M. (1986) *Mol. Cell. Biol.*, **68**, 4387–4395.
Dougherty,J.P., Wisniewski,R., Yang,S., Rhode,B.W. and Temin,H.M. (1989) *J. Virol.*, **63**, 3209–3212.
Dupraz,P. and Spahr,P.-F. (1992) *J. Virol.*, **66**, 4662–4670.
Embretson,J.E. and Temin,H.M. (1987) *J. Virol.*, **61**, 2675–2683.
Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimito,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.
Gallis,B., Linial,M. and Eisenman,R. (1979) *Virology*, **94**, 146–161.
Gorelick,R.J., Henderson,L.E., Hanser,J.P. and Rein,A. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 8420–8424.
Groenink,M., Andeweg,A.C., Fouchier,R.A.M., Broersen,S., van der Jagt,R.C.M., Schuitemaker,H., de Goede,R.E.Y., Bosch,M.L., Huisman,H.G. and Tersmette,M. (1992) *J. Virol.*, **66**, 6175–6180.
Harrison,G.P. and Lever,A.M.L. (1992) *J. Virol.*, **66**, 4144–4153.
Hayashi,T., Shisuo,T., Iwakura,Y. and Shibuta,H. (1992) *Virology*, **188**, 590–599.

Heus,H.A. and Pardi,A. (1991) *Science*, **253**, 191–194.
Higuchi, R. (1989) In Erlich,H.A. (ed.), *PCR Technology*. Stockton Press, New York, pp. 31–38.
Housset,V., De Rocquigny,H., Roques,B.P. and Darlix,J.-L. (1993) *J. Virol.*, **67**, 2537–2545.
Howell,R.M., Fitzgibbon,J.E., Noe,M., Ren,Z., Gocke,D.J., Schwartzer,T.A. and Dubin,D.T. (1991) *AIDS Res. Hum. Retroviruses*, **7**, 869–876.
Hu,W.-S. and Temin,H.M. (1990a) *Proc. Natl Acad. Sci. USA*, **87**, 1556–1560.
Hu,W.-S. and Temin,H.M. (1990b) *Science*, **250**, 1227–1233.
Ito,W., Ishiguro,H. and Kurosawa,Y. (1991) *Gene*, **102**, 67–70.
Jaeger,J.A., Turner,D.H. and Zuker,M. (1989a) *Proc. Natl Acad. Sci. USA*, **86**, 7706–7710.
Jaeger,J.A., Turner,D.H. and Zuker,M. (1989b) *Methods Enzymol.*, **183**, 281–306.
Jones,J.S., Allan,R.W. and Temin,H.M. (1993) *J. Virol.*, **67**, 3151–3158.
Katoh,I., Kyushiki,H., Sakamoto,Y., Ikawa,Y. and Yoshinaka,Y. (1991) *J. Virol.*, **65**, 6845–6855.
Katoh,I., Yasunaga,T. and Yoshinaka,Y. (1993) *J. Virol.*, **67**, 1830–1839.
Katz,R.A., Terry,R.W. and Skalka,A.M. (1986) *J. Virol.*, **59**, 163–167.
Kawai,S. and Nishizawa,M. (1984) *Mol. Cell. Biol.*, **4**, 1172–1174.
Konings,D.A., Nash,M.A., Maizel,J.V. and Arlinghaus,R.B. (1992) *J. Virol.*, **66**, 632–640.
Koyama,T., Harada,F. and Kawai,S. (1984) *J. Virol.*, **51**, 154–162.
Lever,A., Gottlinger,H., Haseltine,W. and Sodroski,J. (1989) *J. Virol.*, **63**, 4085–4087.
Luban,J. and Goff,S.P. (1991) *J. Virol.*, **65**, 3203–3212.
Mann,R. and Baltimore,D. (1985) *J. Virol.*, **54**, 401–407.
Mann,R., Mulligan,R.C. and Baltimore,D. (1983) *Cell*, **33**, 153–159.
Marquet,R., Baudin,F., Gabus,C., Darlix,J.-L., Mougel,M., Ehresmann,C. and Ehresmann,B. (1991) *Nucleic Acids Res.*, **19**, 2349–2357.
Méric,C. and Goff,S.P. (1989) *J. Virol.*, **63**, 1558–1568.
Méric,C. and Spahr,P.-F. (1986) *J. Virol.*, **60**, 450–459.
Méric,C., Gouilloud,E. and Spahr,P.-R. (1988) *J. Virol.*, **62**, 3328–3333.
Miller,A.D. and Rosman,G.J. (1989) *BioTechniques*, **7**, 980–990.
Moazed,D., Stern,S. and Noller,H.F. (1986) *J. Mol. Biol.*, **187**, 399–416.
Muesing,M.A., Smith,D.H. and Capon,D.J. (1987) *Cell*, **48**, 691–701.
Pathak,V.K. and Temin,H.M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6019–6023.
Pathak,V.K. and Temin,H.M. (1992) *J. Virol.*, **66**, 3093–3100.
Prats,A.-C., Roy,C., Wang,P., Erard,M., Housset,V., Gabus,C., Paoletti,C. and Darlix,J.-L. (1990) *J. Virol.*, **64**, 774–783.
Pugatsch,T. and Stacey,D.W. (1983) *Virology*, **128**, 505–511.
Richardson,J.H., Child,L.A. and Lever,A.M. (1993) *J. Virol.*, **67**, 3997–4005.
Rizvi,T.A. and Panganiban,A.T. (1993) *J. Virol.*, **67**, 2681–2688.
Sagata,N., Yasunaga,T., Tsuzuku-Kawamura,J., Ohishi,K. and Ogawa,Y. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 677–681.
Sakaguchi,K., Aambrano,N., Baldwin,E.T., Shapiro,B.A., Erickson,J.W., Omichiski,J.G., Clore,G.M., Gronenborn,A.M. and Appella,E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5219–5223.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: a Laboratory Manual*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schwartz,S., Campbell,M., Nasioulas,G., Harrison,J., Felber,B.K. and Pavlakis,G.N. (1992) *J. Virol.*, **66**, 7176–7182.
Sorge,J., Ricci,W. and Hughes,S.H. (1983) *J. Virol.*, **48**, 667–675.
Southern,P.J. and Berg,P. (1982) *J. Mol. Appl. Genet.*, **4**, 327–341.
Sundquist,W. and Heaphy,S. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3393–3397.
Takeuchi,Y., Simpson,G., Vile,R.G., Weiss,R.A. and Collins,M.K. (1992) *Virology*, **186**, 792–794.
Temin,H.M. (1968) *Int. J. Cancer*, **3**, 273–282.
Temin,H.M. (1991) *Trends Genet.*, **7**, 71–74.
Temin,H.M. and Kassner,V.K. (1974) *J. Virol.*, **13**, 291–297.
Tounekti,N., Mougel,M., Roy,C., Marquet,R., Darlix,J.-L., Paoletti,J., Ehresmann,B. and Ehresmann,C. (1992) *J. Mol. Biol.*, **223**, 205–220.
Uhlenbeck,O.C. (1990) *Nature*, **346**, 613–614.
Watanabe,S. and Temin,H.M. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 5986–5990.
Weeks,K.M. and Crothers,D.M. (1991) *Cell*, **66**, 577–588.
Weiss,R., Teich,N., Varmus,H. and Coffin,J. (1985) *RNA Tumor Viruses*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Woese,C.R., Winker,S. and Gutell,R.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
Yang,S. (1993) PhD thesis, University of Wisconsin-Madison, WI.
Zuker,M. (1989) *Science*, **244**, 48–52.
Zuker,M. and Stiegler,P. (1981) *Nucleic Acids Res.*, **9**, 133–148.

Received on September 29, 1993; revised on November 5, 1993