JOURNAL OF VIROLOGY, Apr. 1995, p. 2175–2186 Vol. 69, No. 4
0022-538X/95/$04.00+0

Secondary Structure Model of the Mason-Pfizer Monkey Virus 5′ Leader Sequence: Identification of a Structural Motif Common to a Variety of Retroviruses

GEOFFREY P. HARRISON,[1] ERIC HUNTER,[2] AND ANDREW M. L. LEVER[1]*

Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 2QQ, United Kingdom,[1] *and Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294*[2]

Received 14 October 1994/Accepted 10 January 1995

A stable secondary structure model is presented for the region 3′ of the primer-binding site to 130 bases into the *gag* sequence of the prototype type D retrovirus Mason-Pfizer monkey virus. Using biochemical probing of RNA from this region in association with free energy minimization, we have identified a stem-loop structure in the region, which from other studies has been shown to be important for genomic RNA encapsidation. The structure involves a highly stable stem of five G-C pairs terminating in a heptaloop. Comparison of the Mason-Pfizer monkey virus structure with one predicted for squirrel monkey retrovirus demonstrates an identical stem and a common ACC motif in the loop. Free energy studies of the secondary structure of the 5′ regions of eight other retroviruses predict stem loops which have similar GAYC motifs. We believe this may represent a common structural and sequence motif which among other functions may be involved in genomic RNA packaging in these viruses.

Newly transcribed retroviral mRNA carries not only viral protein-coding information but also a second series of *cis*-acting messages in the form of sequence and structural motifs which determine its intracellular localization and biochemical fate. Among these are signals determining transcription rate (48, 56), splicing (15), nuclear export (16) and, in the case of the genomic RNA, the ability to be encapsidated into a virion particle. For a number of retroviruses, encapsidation, or packaging, signals (Ψ) have been identified (36–38a, 51). In most cases this identification has been achieved by deletion mutagenesis leading to an RNA encapsidation defect. Vectors containing inserts of various sizes from the known packaging regions have been used to identify the minimum sequence sufficient for encapsidation of an RNA into a particular virion particle (1, 60).

Despite increased knowledge of RNA structures and the identification of packaging signal regions, the actual RNA-protein interactions involved in packaging are poorly understood (7, 8, 49). It is assumed that packaging involves the interaction of an RNA motif(s) with a component of the *gag*- or *gag/pol*-encoded polyproteins. There is no sequence conservation between identified packaging signals; therefore, the specificity of packaging is likely to involve physical recognition of unique RNA structures, formed because of intramolecular base pairing and further tertiary folding. The only sequence which has previously been shown to be common to Ψ's of different retroviruses is the GACG motif identified by Konings et al. (34), which is on a stem structure. For human immunodeficiency virus type 1 (HIV-1), we have published an RNA secondary structure model based on biochemical analysis, free energy minimization, and sequence comparison in a region known to be important for packaging (24). Other investigators have modelled the same leader region, generally using only biochemical and free energy minimization predictions, and have arrived at similar models, with some variations (5, 25, 49). The packaging region of Moloney murine leukemia virus (MoMulV) has been modelled by two groups (3, 58). A secondary structure which is essential for efficient encapsidation of avian sarcoma virus has been reported previously (33). Yang and Temin (63) showed that two hairpin structures are required for encapsidation of the spleen necrosis virus (SNV) genomic RNA and that a similar region in MoMuLV could substitute for these two hairpins in an SNV-based retroviral vector. Both stem loops in SNV contain the GACG sequence identified by Konings et al. (34).

Mason-Pfizer monkey virus (MPMV) is the prototype type D retrovirus (11, 30) which causes an immunodeficiency disease in newborn rhesus monkeys (9, 18). Related type D viruses which cause diseases similar to human AIDS in nonhuman primates (simian AIDS) have been isolated (12, 14, 27). Two isolates in particular, simian AIDS retrovirus type 1 (40) (SRV-1) and SRV-2 (39), have been directly linked to simian AIDS. MPMV has been referred to as SRV-3 (54). These three viruses have been completely sequenced; MPMV (55), SRV-1 (45), SRV-2 (57), and two variants of SRV-1 (26) have been partially sequenced. Squirrel monkey retrovirus (SMRV), another type D virus, has been partially sequenced (10). A type D virus isolated from a human lymphoblastoid line is very closely related to SMRV. It has been completely sequenced and is referred to as SMRV-H (43). The phylogenetic relationships between these viruses are represented in Fig. 1 on the basis of analysis of the 5′ ends of the genomes.

The Ψ of MPMV has been mapped to an area within 619 bp 3′ of the primer-binding site (PBS) (59). We have also studied this region and have further defined an important *cis*-acting Ψ to a short sequence between the PBS and splice donor (SD) (22).

In order to gain insight into the complex recognition process of encapsidation, we analyzed the RNA secondary structure of the Ψ of MPMV, using biochemical probing and computer modelling. We then compared the deduced structure with the predicted structure, and their sequences, in a group of closely and more distantly related retroviruses.

* Corresponding author. Mailing address: University of Cambridge, Department of Medicine, Addenbrooke's Hospital, Hills Rd., Cambridge CB2 2QQ, United Kingdom. Phone: 1223 336747. Fax: 1223 330103.

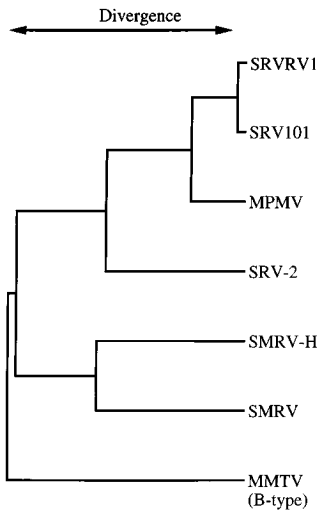

FIG. 1. Phylogenetic tree showing relationships between some type D retroviruses and mouse mammary tumor virus (MMTV), based on only the first 1 kb of their genomes.

MATERIALS AND METHODS

In vitro transcription. The 5′ leader region of MPMV was excised from the proviral clone pSHRM15 (47), by using the *Sph*I site at position 567 and the *Sac*I site at position 1561. The numbering is that of GenBank accession no. M12349. This fragment was ligated into the same sites in the expression vector pGem-4Z (Promega, Southampton, United Kingdom), and RNA was transcribed from the T7 promoter.

Biochemical and enzymatic probing. Biochemical analysis was carried out essentially as described by Harrison and Lever (24). Briefly, after in vitro transcription, RNA samples were digested with DNase 1, dissociated, and reannealed by heating to 65°C for 5 min and cooling to 25°C during a period of 20 min. The RNA samples were then divided into 2-μg aliquots and modified with one of the following agents: cobra venom RNase V_1 (Pharmacia), hydroxymethylpsoralen (psoralen) (Sigma), 2-keto-3-ethoxybutyraldehyde (kethoxal) (United States Biochemical), RNase T_1 (Boehringer Mannheim), and dimethyl sulfate (DMS) (Fluka).

Photoreactions with psoralen were carried out under germicidal UV light (wavelength, 366 nm). Samples were irradiated for 30 min at room temperature in Eppendorf tubes at a distance of 20 to 30 mm from a Sylvania G875 tube light.

cDNA synthesis. Modified RNA was extracted with phenol-chloroform, and 10 ng of a synthetic oligonucleotide primer was added in avian myeloblastoma virus reverse transcriptase (RT) buffer (Promega). The oligonucleotide primers used were 5′-TTG/CCC/CAT/ATC/CGA/GCG/C-3′ from bases 898 to 880, 5′-CCC/CGT/GTC/TTT/AAA/GCC/-3′ from bases 957 to 939, and 5′-TAT/GGT/TCC/CTC/TTG/CGG/-3′ from bases 1042 to 1025.

The RNA was dissociated by heating the solution to 70°C for 5 min, and the solutions were immediately stored on ice. It was found that the primer annealed effectively when the solutions were incubated with the extension mix. Extension analyses were carried out by adding 1 U of avian myeloblastoma virus RT (Promega); a final concentration of 1 mM (each) dATP, dTTP, and dGTP; and 1 μl of [α-^{32}P]dATP. These extension mixtures were incubated at 42°C for 1 h. cDNAs were precipitated under ethanol and redissolved in Tris-EDTA (TE) buffer. Formamide dye mix (1/10 volume) was added, and approximately equal amounts of radioactivity were loaded onto 6% polyacrylamide–7 M urea gels, along with dideoxy-sequencing ladders (50) generated by using the same oligonucleotide primer.

Free energy minimization studies. Free energy minimization studies were carried out on a UNIX system by folding lengths of sequences ranging from 100 to 300 bases at intervals of 30 bases from bases 700 to 1050. The program employed was Mfold, adapted for the Genetics Computer Group (University of Wisconsin) (29, 65), with a graphical presentation by using Squiggles (44) in the Genetics Computer Group program Plotfold. The Mfold program presents suboptimal foldings within 5 to 10% of the calculated minimum free energy. Bio-

TABLE 1. GenBank accession numbers of sequences referred to this study

Virus	GenBank accession no.
Baboon endogenous retrovirus	X05470
Bovine leukemia virus	K02120
Caprine arthritis-encephalitis virus	M33677
Feline leukemia virus	D00732
Gibbon ape leukemia virus	M26927
Mouse mammary tumor virus	M15122
Mason-Pfizer monkey virus	M12349
Moloney murine leukemia virus	J02255/6/7
Moloney murine sarcoma virus	J02266
Murine leukemia virus	J01998
Murine osteosarcoma virus	X03347
Spleen focus forming virus	K00021
Spleen necrosis virus	V01200
Squirrel monkey retrovirus	M26927
Variant of SMRV	M23385
Simian AIDS retrovirus type 1	
SRV101	M17561
SRV1LT	M17560
SRVRV1	M11841
Simian AIDS retrovirus type 2	M16605
Visna virus	M60609

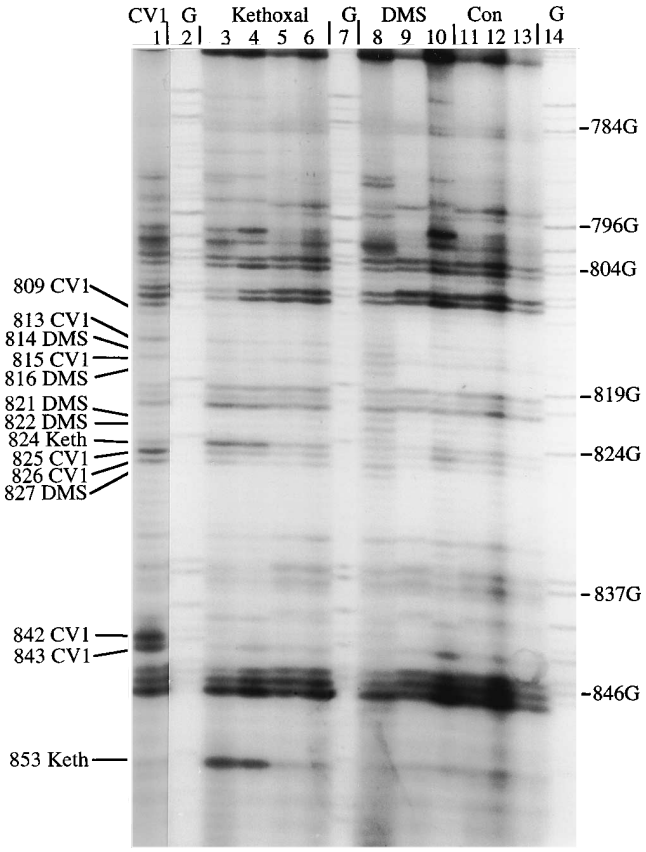

FIG. 2. Structural probing of MPMV 5′ leader sequence. An autoradiograph of a gel comparing RT RNA templates is shown. The sequence of the primer used in all lanes was 5′-TTGCCCCATATCCGAGCGC-3′ (nucleotides 898 to 880). The dideoxy-sequencing ladder was generated from the same sequence. Bands which are unique to lanes of modified RNA represent structure-specific modifications or cleavages; their positions can be read from the sequencing ladder. Lanes: 1, RNA digested with 0.5 U of RNase V_1 (CV1) at 0°C for 30 min; 2, 7, and 14, dideoxy G stops; 3 to 6, RNAs modified with kethoxal (Keth) (20% by volume in ethanol) at final concentrations of 1, 0.75, and 0.05%, respectively, for 30 min at 20°C; 8 to 10, RNA modified with DMS at final concentrations of 5, 1, and 3%, respectively, for 30 min at 20°C; 11 to 13, unmodified RNA (controls [Con]). The position numbers of the G stops are shown on the right. The positions of unique bands are shown on the left.

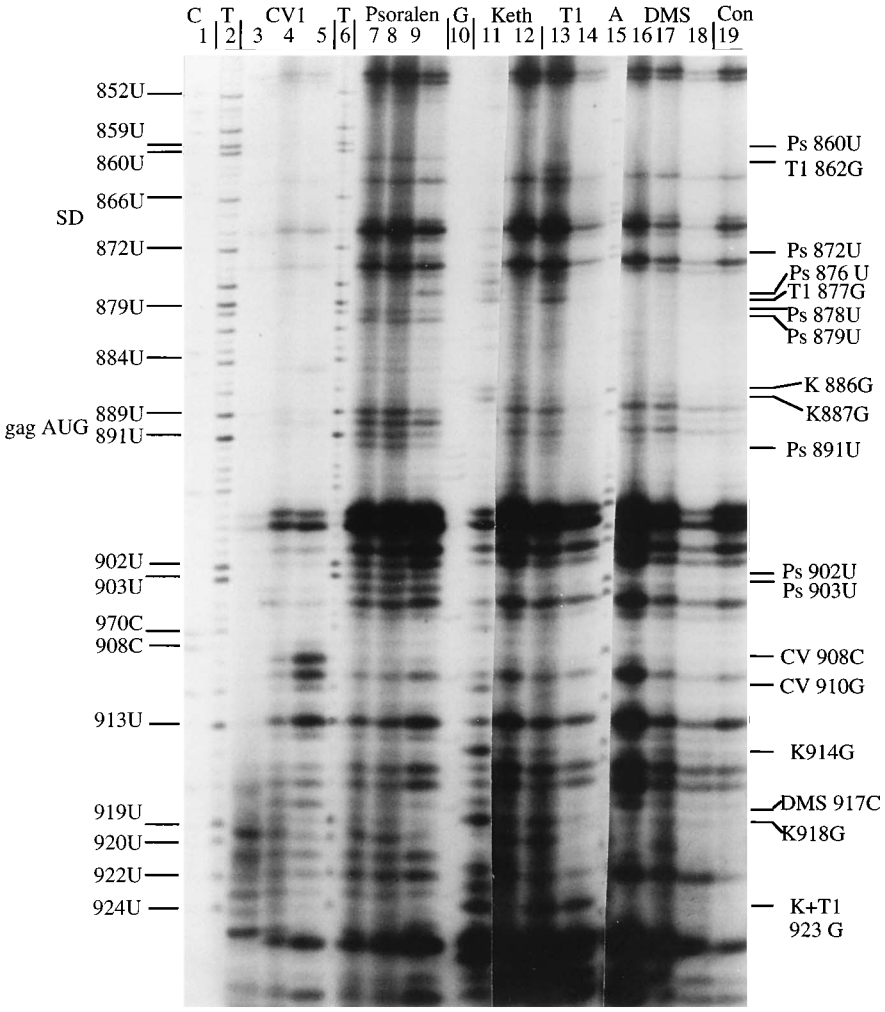

FIG. 3. Autoradiograph of gel showing comparison of RT RNA templates. The sequence of the primer used in all lanes was 5′-CCCCGTGTCTTTAAAGCC-3′ (nucleotides 957 to 939). The dideoxy-sequencing ladders were generated by using the same sequence. Lanes: 1, dideoxy C ladder; 2 and 6, dideoxy T ladder; 3 to 5, RNA digested with RNase V_1 (CV1 or CV) 1 U at 37°C, 0.75U at 20°C, and 0.5 U at 0°C, respectively, for 30 min; 7 to 9, RNA irradiated with UV light (wavelength, 366 nm) for 30 min at 20°C with psoralen (Ps) (dissolved in ethanol at 5 mg/ml) at final concentrations of 20, 40, and 80 μg/ml, respectively. 10, dideoxy G stops; 11 and 12, RNA modified with kethoxal (Keth) at final concentrations of 1 and 0.5%, respectively, for 30 min at 20°C; 13 and 14, RNA digested with RNase T_1 (T1) (10 and 30 U, respectively) at 0°C for 30 min; 15, dideoxy A stops; 16 to 18, RNA modified with DMS at final concentrations of 2, 1, and 0.5%, respectively, for 30 min at 0°C; 19, unmodified RNA template (control [Con]). The numbering of the dideoxy sequencing ladder is shown on the left, and the numbering of unique bands is on the right.

logically functional foldings may be found in suboptimal predictions, because the folding rules and energy parameters are not accurately known.

Nucleotide sequence accession numbers. GenBank accession codes of all sequences referred to in the text are shown in Table 1.

RESULTS

Free energy minimization. Extensive and comprehensive free energy predictions were made from the sequences of MPMV, SRV-1, SRV-2, SMRV-H, and mouse mammary tumor virus from the PBS to 300 nucleotides into the *gag* open reading frame. Also, the structures of this region in more than eight other distantly related retroviruses were examined. Free energy calculations with the MPMV sequence repeatedly predicted stem loops in the following regions: bases 778 to 792, 837 to 862, 878 to 898, and 902 to 946.

Probing for secondary structure. Examples of biochemical probing are presented in Fig. 2 to 4. In Fig. 2, RNase V_1 cleavages are particularly abundant in the population of RNA molecules which were loaded onto lane 1. These cleavages are evident as intense bands at positions 842 and 843. In Fig. 3 there are particularly intense psoralen-induced bands in lanes 7 to 9 at positions 902 and 903, both of which are uridines. A unique band at position G-923 on the same autoradiograph is evidence that this residue is unpaired in vitro. In Fig. 4 there is a band which is clearly unique to RNase V_1 (lane 1) at position U-966; U-990 is also clear. There is a unique psoralen band at U-967. We have chosen not to record the bands at position A-972 in lanes with RNase V_1, psoralen, and DMS, because psoralen usually only modifies pyrimidines (principally uridines). It is probable that this band is due to a spontaneous stop in RT. A band is visible in the same position in the control lanes in the original autoradiograph.

The complete set of biochemical probing data are presented in Fig. 5 and on the secondary structure model in Fig. 6. These data are taken from a large number of gels, of which the results shown in Fig. 2 to 4 are only examples.

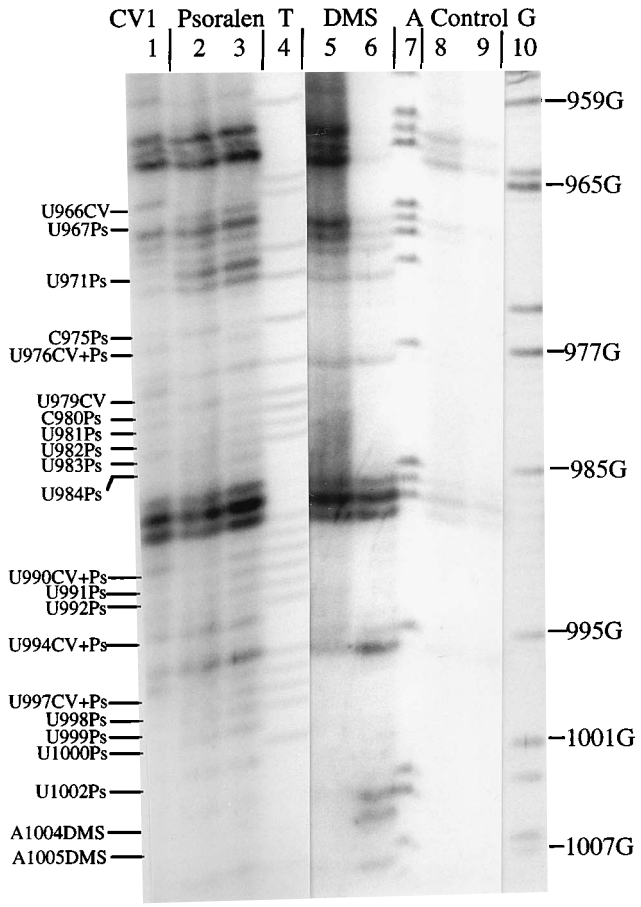

FIG. 4. Autoradiograph of gel showing comparison of RT RNA templates. The sequence of the primer used in all lanes was 5′-TATGGTTCCCTCTT GCGG-3′ (from nucleotides 1042 to 1025). Bands which are unique to lanes of modified RNA represent structure-specific modifications or cleavages. Their positions can be read from the sequencing ladder. Lanes: 1, RNA digested with 0.7 U of RNase V_1 (CV1 or CV) at 0°C for 30 min; 2 and 3, RNA irradiated with UV light (wavelength, 366 nm) for 30 min at 20°C with psoralen (Ps) at final concentrations of 80 and 40 μg/ml, respectively; 4, dideoxy T stops; 5 and 6, RNA modified with DMS at final concentrations of 1 and 2%, respectively, for 30 min at 37°C; 7, dideoxy A stops; 8 and 9, unmodified RNA (control). The sample in lane 9 was irradiated with 366 nm of UV light for 30 min at 20°C without any psoralen. Lane 10, dideoxy G stops.

Figures 6 and 7 show the results of three methods of analysis: free energy minimization studies, biochemical probing, and phylogenetic comparisons. In Fig. 7 the positions of nucleotide variations from all four SRV isolates are presented in lowercase.

Phylogenetic comparisons. Phylogenetic comparisons were initially made only between MPMV and SRV-1 and SRV-2. Sequence alignments were computed by using the Genetics Computer Group program Pileup (17). The alignment of MPMV with SRV-1 and SRV-2 is shown in Fig. 8. Sequence variations are presented in the secondary structure model in Fig. 7. It is assumed that regions which can be aligned will adopt a similar biologically relevant secondary structure (53). SMRV-H was found to have only 40% homology to MPMV in the region of 270 bases 3′ to the PBS and was too divergent from the aforementioned type D viruses for its sequence to be aligned with those in Fig. 8.

PBS. The PBS is presented as unpaired since in vivo it is associated with 18 bases of the $tRNA^{Lys,1-2}$ (46). If allowed to fold, the PBS is calculated to form a stem loop with free energy of −3.2 kcal (−13 kJ)/mol from base G-742 to C-754 with bases 747 to 749 as the loop.

Bases 763 to 799. Two possible structures for the region from bases 763 to 794 are presented in Fig. 9A and B. Naturally occurring compensatory sequence (covariant) changes were found between MPMV and SRV-2 at MPMV positions U-778 and C-779, which face nucleotides A-792 and G-791, respectively (panel A). A pair of compensatory substitutions adjacent to one another is very unlikely to be coincidental (19, 42). The conservation of the structure from bases 778 to 794 (Fig. 8A) between MPMV, SRV-1, and SRV-2 strongly suggests that it is functionally important. Although two of three sequence changes that occur in the structure shown in panel B are not disruptive, the structure lacks the compensatory mutations of the fold shown in panel A. The presence of the compensatory substitutions was considered to be more convincing than the free energy predictions, which suggest that the structure shown in panel B is more stable. The few biochemical data obtained for this region do not help to determine which of the structures exists in vitro. No similar structure was identified by free energy studies on the analogous regions of SMRV or mouse mammary tumor virus. There are two GACG sequences at positions 788 to 794 which will form the stem-loop structure represented in panel C. This bears some resemblance to the motif identified by Konings et al. (34). The biochemical data support the existence of this structure, but it is not conserved in SRV-2 and the free energy prediction is lower than those for the structures shown in panels A and B.

Bases 797 to 837. There are multiple insertions and deletions in the region from bases 797 to 837, which implies that this region does not have a function which is dependent on conserved structure or sequence in MPMV, SRV-1, and SRV-2. However, the large number of RNase V_1 cleavages and the presence of psoralen modifications in this region show that in MPMV the sequence is highly structured.

Bases 829 to 871. The stem loop between bases 838 and 863 is predicted repeatedly for MPMV across differently sized sequence windows by using the Genetics Computer Group program Mfold. Biochemical data point to the potential stem structure extending down from the terminal stem and loop to include the region from bases 829 to 871, although the Mfold predictions were not as consistent in the latter part. Thirteen of 14 bp in this stem are Watson-Crick interactions, 11 of which are G-C interactions and two are A-U. The other predicted pairing is a favorable G-U. These all indicate a highly stable structure. The five terminal G-C pairs (bases 842 to 858) are conserved in SRV-1 and SRV-2. Free energy calculations for SMRV-H predict a very similar structure with a very similar sequence, which is discussed below. The strong biochemical evidence of base pairing in MPMV at the base of this stem, with the notable feature of 100% of the pyrimidines interacting with psoralen, implies more-extensive double strandedness than can be predicted by using free energy minimization, which relies only on Watson-Crick interactions.

Bases 869 to 902. The *gag* initiation codon is predicted to be at the top of a short stem and loop, on the 3′ side of the loop (Fig. 10). Alternative structures for the region from bases 869 to 902 are presented in Fig. 10, which has the AUG in very similar positions in both structures. Structure A is predicted to have a free energy of −5.7 kcal (−24 kJ)/mol and is therefore included in the structure shown in Fig. 6 and 7. The structure shown in panel B is predicted to have a free energy of −3.8 kcal (−16 kJ)/mol. Naturally occurring sequence variations show no changes to suggest which is the more likely in vivo structure. The RNase V_1 digests support the presence of a helical region in the stem predicted by Mfold. The psoralen

Base	CV1	Ps	T1	Keth	DMS
C		++			+
780G				+	
U					
G					
A					
G					
785G					
A					
A					
G				+	
A					
790C					
G					
A	+				+
C					+
G				+	
795C	++	++			
G					
U					
U					
C					++
800G					
C	+				
C	+				
G	+				
G	+				
805C					
C					
G	++				
G					
C	+				
810G	+				
A					
U					
U	+	+			
A					+
815A	+				
A					+
A					
G					
U					
820G					
A					+
A					+
A					+
G				++	
825U	++	+++			
A	++				+
A					+
A					
C					
830U	+	+			
C		+			
U		+			
C	+	+			
U	+	+			
835U	+	+			
G					
G					
C		+			
C		+			+
840G					
C					+
C	++++	+			
G	++++				
C					
845G	+				
G	+				
G					
A					
A					
850C					
C					
U					
G			+	++++	
C					
855C					

Base	CV1	Ps	T1	Keth	DMS
856G	+				
C					
G	++++				
U					
860U		++++			
G					
G			++		
A					
C					
865C	+				
U	+	++			
G					
A					
A					
870A					
Splice G					
Donor U		+			
G					
A					
875G				+	
U		+			
G			++		
U		++			
U		++			
880G					
C					
G					
C	+				
U	+				
885C					
G	+			+	
G	+			+	
A					
U		+			
890A					
gag U		+			
G					
G					
G	+				
895G	+				
C					+
A					
A	+	+			
G					
900A					
A					
U		++			
U		++			
A					
905A					
G				+	
C					
C	++				
A					
910G	+			+	
C					
A					
U					
G				++	
915A					
A	+				
C					+
G				++++	
U					
920U		+			
A					
U					
G			++	+++	
U					
925A					+
G			+	+++	
A					
A					
C					
930A					
A					
U					

Base	CV1	Ps	T1	Keth	DMS
U					
G				+	
935A					+
A					+
G				+	
C					
A					
940G					
G					
C					
U					
U					
945U					
A					
A					
A					
G					
950A					
C		+			
A					+
C		+			
G				+	
955G					
G					
G					
A					
G					
960U					
A					
A					
A					
G					
965G					
U	+				
U	+	+			
A					
A					
970A					
U		++			
A					
U					
G				+	
975C		+			
U	+	+			
G				+	
A					+
U	+				
980C		+			
U					
U		+			
U	+	+			
U		+			
985G				++	
A					
A					
A					
U					
990U	+	+			
U		+			
U		+			
U					
U	++	++			
995G				+	
A					+
U	+	+			
U		+			
U		+			
1000U		+			
G				+	
U		+			
G				+	
A					++
1005A					++
G				+	
G				+	
A					+

FIG. 5. Results of biochemical and enzymatic probing of RNA secondary structure. Data from many gels, in addition to those presented in Fig. 2 to 4, are included. CV1, cobra venom RNase V_1; Ps, psoralen; T1, RNase T_1; Keth, kethoxal; + to ++++, approximate degree of intensity (from least to most) of the bands resulting from treatment. Intensities were determined visually.

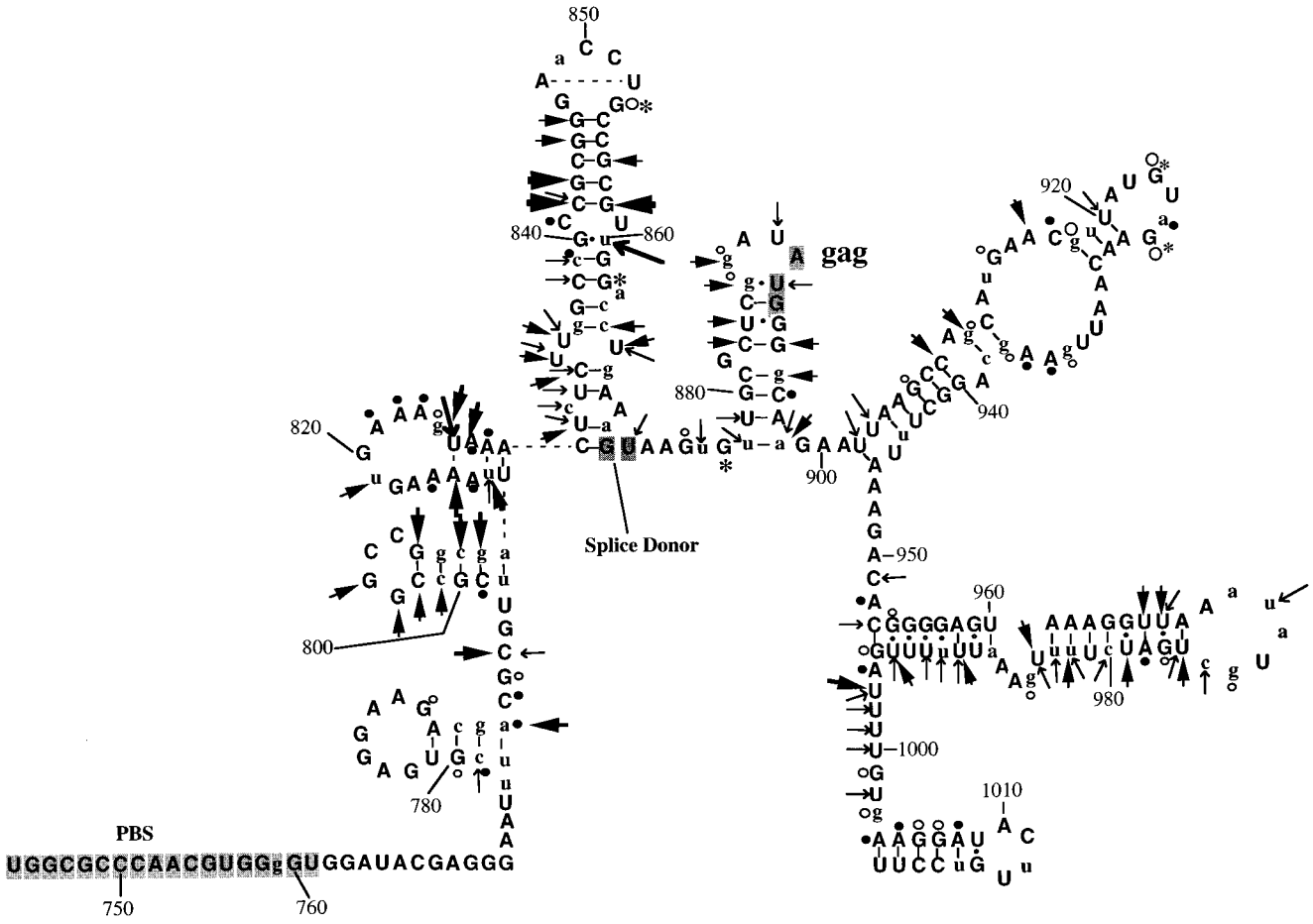

FIG. 6. Results of biochemical probing for structure, presented on a secondary structure model for MPMV, predicted by free energy minimization. The size of the symbol represents the approximate degree of intensity of the modification or cleavage. Bases presented in lowercase vary from MPMV in the SRV isolates. ⇾, cobra venom RNase V_1 site (double stranded); →, psoralen modification (double stranded); ●, DMS modification (single stranded); ○, kethoxal modification (single stranded); ❊, RNase T_1 site (single stranded).

modification at U-876 indicates that it could be paired to A-900, with two Gs facing one another next to them. U-872 of the SD signal could be paired to A-828. Similar structures are found in this region in other retroviruses, as discussed below.

Bases 904 to 943. The structure from bases 904 to 943 was repeatedly predicted in different sequence windows despite the fact that it has low free energy (−8.6 kcal [−36 kJ]/mol) and all the sequence changes in the paired regions are disruptive.

Pairing of bases A-915 to U-933 and A-916 to U-932 does not lower the free energy predicted. This structure was not found in free energy minimization studies of SMRV-H.

Bases 948 to 999. Three potential structures for the region from bases 948 to 999 were predicted, but only one is shown (Fig. 6). The free energy minimization predictions favor the structure presented in Fig. 6 and 7, with bases 969 to 975 forming the terminal loop. The strong psoralen modification at U-971 indicates that this stem loop is paired to another region. All three predictions for this region have double-stranded modifications in the terminal loop. It is possible that the terminal loop could form Watson-Crick-paired homodimers or that there could be long range intramolecular interactions.

Comparison of MPMV region from bases 840 to 860 with a similar stem loop in SMRV-H. The sequence of SMRV-H was studied by free energy minimization predictions. In SMRV-H the sequence from bases 564 to 583 is predicted to fold into a stem (Fig. 11), which is identical to MPMV bases 840 to 860 through five pairs of G-C bonds. The terminal loop of the MPMV structure has 7 unpaired bases, whereas the equivalent stem loop of SMRV-H has 9; however, a noncanonical A-A pair may occur at the base, making the loops identical in size. The loop contains an ACC motif (GCC in SRV-1 and SRV-2) consistent with a consensus sequence (discussed below). The lower half of the MPMV stem from bases 830 to 838, which has disruptive changes in SRV-1 and SRV-2, and an insertion at base 836 is not conserved in SMRV-H, although it can form a stem with bulges.

Conservation of structure at *gag* initiation codon. The *gag* initiation codon is predicted to be near the end of the stem on the 3′ side, in both models of the MPMV sequence presented in Fig. 10. The *gag* initiation codon is predicted to be at the same position in a similar stem in SMRV-H (Fig. 11). Base pairings which are conserved between the predicted structures for MPMV and SMRV-H are shaded in Fig. 11. The analogous regions of the following retroviruses are predicted by free energy minimization studies to form similar structures, having the *gag* AUG on the 3′ side of a stem loop: Moloney murine leukemia virus, bases 611 to 626, visna virus, bases 481 to 494, feline leukemia virus, bases 911 to 924; Gibbon ape leukemia virus, bases 615 to 637; baboon endogenous retrovirus, bases

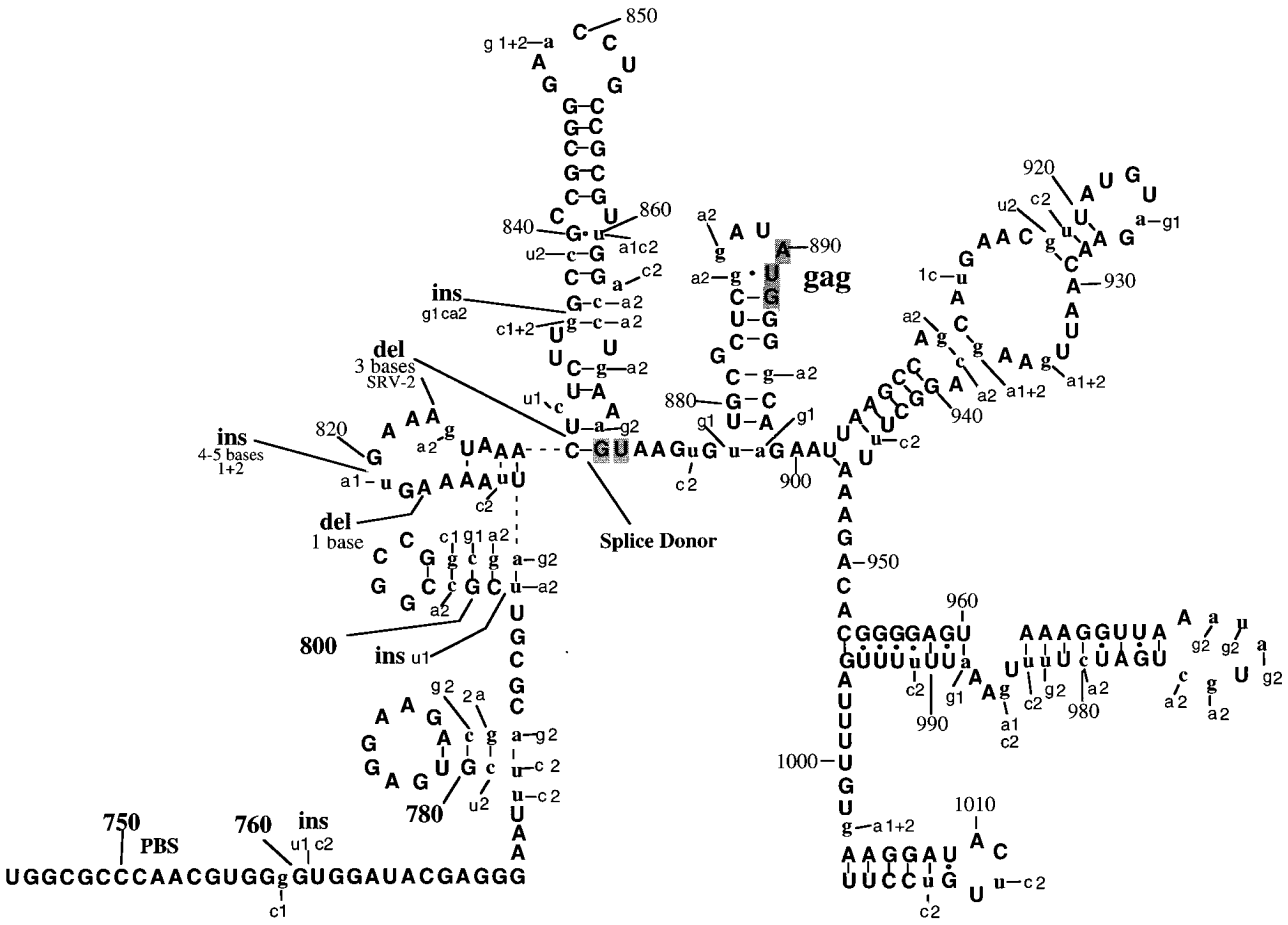

FIG. 7. Positions of sequence variations between MPMV and four SRV isolates. Bases which are not conserved between the three viruses are shown in lowercase. Sequence variations are indicated as follows: g1, base G in SRV-1; and u2, base U in SRV-2; etc.; ins, insert, del, deletion. The SD signal and the *gag* initiation codon are shaded. For full details of sequence variations, refer to Fig. 8.

547 to 569; and caprine arthritis-encephalitis virus, bases 493 to 517.

DISCUSSION

Biochemical probing and free energy calculations together are not sufficient to provide conclusive evidence of the nature of RNA secondary structures, but phylogenetic studies can be conclusive, as long as there are sufficient data to which to refer (53, 62). The use of phylogenetic analysis as well as the biochemical probing and free energy minimization data in this study provide convincing evidence of the nature of RNA secondary structures.

In a phylogenetic alignment, invariance of a residue indicates that it is important for the structure and/or function of the molecule. Residues which are complementary and whose compositions consistently covary must be functionally related, which almost always means that they are in physical contact. If two positions covary according to Watson-Crick rules, one can be fairly certain that they form a base pair. Such compensatory base substitution has proven very effective in determining the secondary structure of rRNAs (19, 23). Where multiple sequences exist, such as viral quasispecies of HIV-1, comparison and identification of conserved nucleotides can indicate bases which are likely to be important for a particular *cis*-acting function. This approach together with biochemical and computer modelling was used by us to identify the secondary structure in the 5′ leader sequence of HIV-1, which is stable and conserved and which involves regions known by deletion mutagenesis to be important for packaging (24).

To probe for the RNA secondary structure, in vitro-transcribed RNAs were either digested with structure specific enzymes or modified with structure-specific biochemicals. Cleavages and modifications induced by these agents cause RT to terminate or pause during DNA synthesis, which results in the generation of unique (or more-intense) bands on autoradiographs. The modified RNAs are primed with a short single-stranded DNA primer and are reverse transcribed with avian myeloblastoma virus RT. The position of these modifications can be determined by reading a dideoxy-sequencing ladder on the gel, which has been generated by using the same oligonucleotide primer (at 1 base 3′ to the RT termination or pause).

Pauses or stops in the cDNA synthesis give rise to bands 1 nucleotide shorter than the corresponding band in the dideoxy-sequencing ladder since cDNA synthesis stops at the nucleotide immediately preceding the modified position or cleavage site. This is because the cDNA cannot complement the modified or cleaved nucleotide. Bands in the lanes of the untreated RNA form a reproducible pattern of stops (52), where RT pauses or dissociates from the template for reasons which are not yet understood (6, 20). Unless a band is unique to the lane of modified RNA, it does not report structure. The exception

```
                750 PBS      760          770          780          790
MPMV  TGGCGC CCAACGTGGG G- -TGGATACG AGGGAATTTC GTGAGGAAGA CGACGCG-TT-C
SRVRV1 TGGCGC CCAACGTGG c Gt -TGGATACG AGGGAATTTC GTGAGGAAGA CGACGCGgTTT-
SRV101 TGGCGC CCAACGTGG c Gt -TGGATACG AGGGAATTTC GTGAGGAAGA CGACGCGgTTT-
SRV1LT TGGCGC CCAACGTGGG GC tTGGATACG AGGGAATTTC GTGAGGAAGA CGACGCG-TTT-
 SRV2 TGGCGC CCAACGTGGG GC-TGGATACG AGGGAAT c c t GTGAGGAAGA g a g CGCG-Ta aC

     800          810          820          830          840
MPMV GCCGGCCGGC GATTAAAAG T-----GAAAGTAAAC TcT CTTG-GCC GCCGCGGGAa
SRVRV1 GCCGGCC-c g GATTAAAAG a GAAAcGAAAGTAAAC TTTCTT c g GCC GCCGCGGGAG
SRV101 GCCGGCC-c g GATTAAA-G a GAAAcGAAAGTAAAC TTTCTT c g GCC GCCGCGGGAG
SRV1LT -CCGGCCGGC GATTAAA-G a aAAAcGAAAGTA AAC TTTCTTG-GCC GCCGCGGGAG
 SRV2 - aCGGCCGGC a gTcAAA -- TGAAA-GAAAa---AC ---CTT c c a GCt GCCGCGGGAG

     850          860          870 Splice Donor 880          890 gag      900
MPMV CCTGCCGCGT t GGACCTGAA AGTAAGTGTT GCGCTCGGAT ATGGGGCAAG AATTAAGCCA
SRVRV1 CCTGCCGCGT AGGACCTGAA AGTAAGTGgT GCGCTCGGAT ATGGGGCAg G AATTAAGCCA
SRV101 CCTGCCGCGT AGGACCTGAA -G.....
SRV1LT CCTGCCGCGT AGGACCTGAA -G.....
 SRV2 CCTGCCGCGT cGG c a aTa AA gGTAAG cGTT GCGCTC a a AT ATGGG aCAAG AATTAAG CCA

     910          920          930          940          950          960
MPMV GCATGAACGT TATGTAGAAC AATTgAA g CA GGCTTTAAAG ACACGGGGAG TAAAGGTTAA
SRVRV1 GCAc GAACGT TATGTg GAAC AATTAAAACA GGCTTTAAAG ACACGGGGAG TAAAGGTTAA
 SRV2 aCAT GAACt c TATGTAGAAC AATTAAAA aA GGCTcTAAAG ACACGGGGAG TAAAGGTTAA

     970          980          990          1000         1010         1020
MPMV ATATGCTGAT CTTTTGAAAT TTTTTGATTT TGTg AAGGAT ACTTGTCCTT GGTTTCCGCA
SRVRV1 ATATGCTGAT CTCTTa AA gT TTTTTGATTT TGTAAAGGAT ACTTGTCCTT GGTTTCCGCA
 SRV2 g g gT a aTGAT aTg cTc AAAT Tc TTTGATTT TGTAAAGGAT AC cTGc CCTT GGTTTCC aCA

     1030         1040
MPMV AGAGGGAACC ATAGATA
SRVRV1 AGAGGGAACC ATAGATA
 SRV2 AGAGGGAACa ATAGA cA
```

FIG. 8. Alignment of nucleotide sequences of MPMV, SRV-1, and SRV-2. Bases in lowercase are not conserved. PBS, SD, and *gag* AUG are underlined for each virus. The published sequences of SRV101 and SRV1LT do not extend into the *gag* open reading frame.

to this is when a band is much more intense than the equivalent bands in other lanes. When a modifying agent is sequence specific, bands can be significant only if they are present at nucleotides for which that agent is specific.

In any biochemical and enzymatic study of RNA secondary structure it should be noted that RNA in solution is in a dynamic state and that alternative structures of the same regions may exist simultaneously in a population of homologous molecules. Any individual RNA molecule may alternate between different conformations. This leads to some ambiguity in interpretation of biochemical data as single-stranded modifiers will be able to attack regions of RNA even if they are only transiently unpaired. A very stable double helix will, however, be much less vulnerable to single-stranded modification than a region which has no nearby complementary sequence.

Interpretation of RNase V_1 digests is complicated by the fact that although this enzyme digests RNA in regions of helical conformation, it does not cleave with any base specificity. It

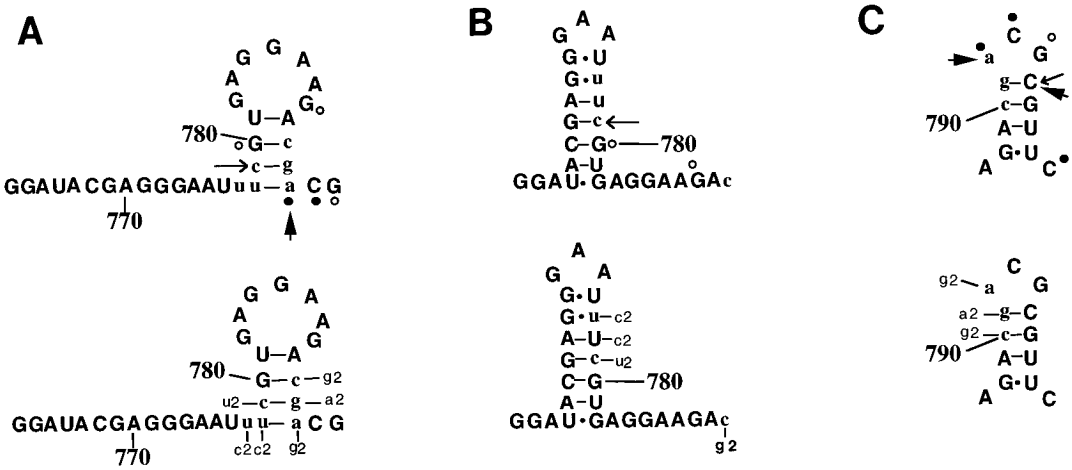

FIG. 9. Alternative folding patterns for region from bases 763 to 799. The predicted free energy values for the structures were −3.1, −4.7, and −2.2 kcal (−13, −20, and −9.2 kJ)/mol for panels A to C, respectively. See the legends to Fig. 6 and 7 for description of symbols.

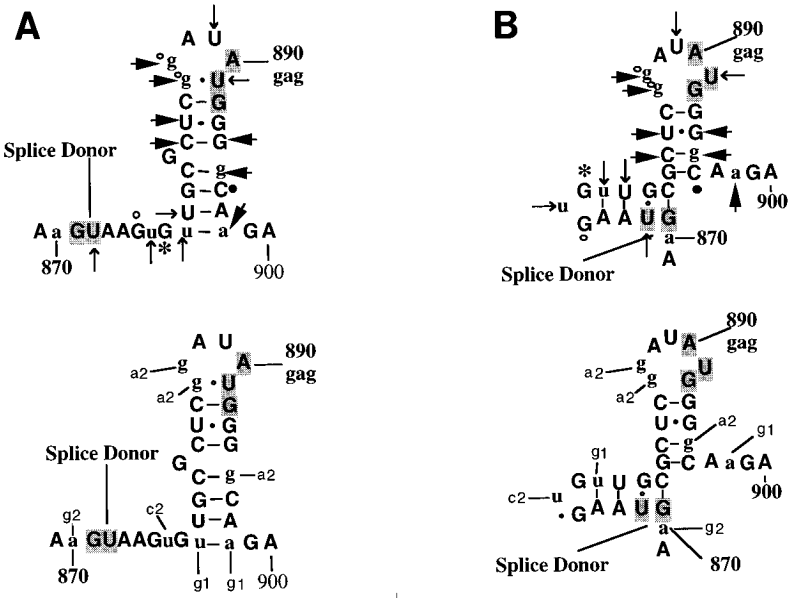

FIG. 10. Alternative folding models for region from bases 869 to 900. The free energy predictions were −5.7 and −3.8 kcal (−24 and −16 kJ)/mol for panels A and B, respectively. Symbols are defined in the legends to Fig. 6 and 7.

can also cleave 1 or 2 bases outside of a region which is helical (21, 31). Psoralen intercalates in base-paired regions and forms covalent adducts with pyrimidines, notably uridines, when irradiated with UV light at 366 nm (4, 64). Psoralen is found to intercalate particularly at the end of helices or in mismatches within helices (35). Kethoxal modifies unpaired G residues, and DMS modifies A's and C's at positions which interfere with RT. RNase T_1 cleaves single-stranded RNA at G residues. The size of structure-specific enzymes limits their effectiveness, because of steric hindrance. Interpretation is also complicated by the fact that, in RNA, base pairing is not limited to Watson-Crick bonds. Comparative analysis of rRNA has shown that noncanonical pairs occur frequently (62). Nuclear magnetic resonance analysis of the 5S RNA of a *Xenopus* sp. showed the existence of G-A, A-A, and U-U pairs (61). These noncanonical pairs are generally weaker than Watson-Crick pairs, but they are important in RNA structures. Another noncanonical RNA base pair which has been observed is G-G. Free energy minimization programs are not yet sufficiently complex to predict noncanonical pairings.

The conservation of the stem loop from base 838 to 863 between MPMV, SRV-1, and SRV-2, together with the existence of an identical stem in SMRV and SMRV-H, is strong evidence for its existence as a defined structure in vivo and for having an important function. Despite being upstream from the SD, and thus not unique to the genomic RNA, the site correlates with our deletion mutagenesis evidence (22) showing that this is an important part of a Ψ in MPMV. Sequences upstream of the SD have been found to affect packaging in Rous sarcoma virus (38) and HIV-1 (32).

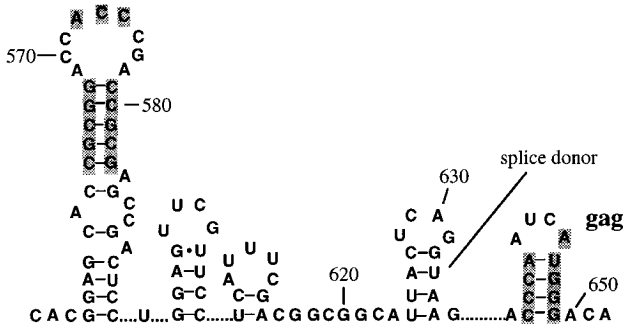

FIG. 11. SMRV-H 5′ leader sequence, a secondary structure model based on free energy minimization studies. Base pairings which are conserved between MPMV and SMRV-H are shaded (A-U to G-U changes are indicated as being conservative).

Sequence specificity of double-stranded regions is thought, in general, to be unimportant in RNA structures such as rRNA (62). However, in the SNV primary sequence, alterations of stem structures causing no change in free energy have been shown to affect packaging efficiency (63). The absolute conservation of the stem sequence in MPMV and SMRV-H may indicate similar constraints on primary sequences in these type D viruses.

In the loop from bases 847 to 853 of the MPMV structure, there is a triplet ACC (bases 849 to 851) which is conserved between MPMV and SMRV-H. In SRV-1 and SRV-2, the purine is conserved. Free energy predictions for type B viruses and a series of distantly related retroviruses, including bovine leukemia virus, caprine arthritis encaphalitis virus, AKR murine leukemia virus, FBJ murine osteosarcoma virus Friend spleen focus-forming virus, and FBR murine osteosarcoma virus, show similar motifs in stem-loop structures in analogous positions relative to the *gag* initiation codon which conform to the general pattern GAYC (Fig. 12). The structures for the murine retroviruses have been reported previously (58). The MPMV motif RCC is also seen in feline leukemia virus (data not shown) and SNV. G-C pairing of the top of the stem is conserved between feline leukemia virus, bovine leukemia virus, and baboon endogenous retrovirus and may have similar importance to the equivalent regions in MPMV, SRV-1, SRV-2, and SMRV-H.

The GACG-hairpin motifs identified by Konings et al. (34) occur in type C murine leukemia viruses and SNV (63). In the latter, linker-scanning mutagenesis showed that mutation of the GACG motif impairs packaging. Konings et al. (34) also

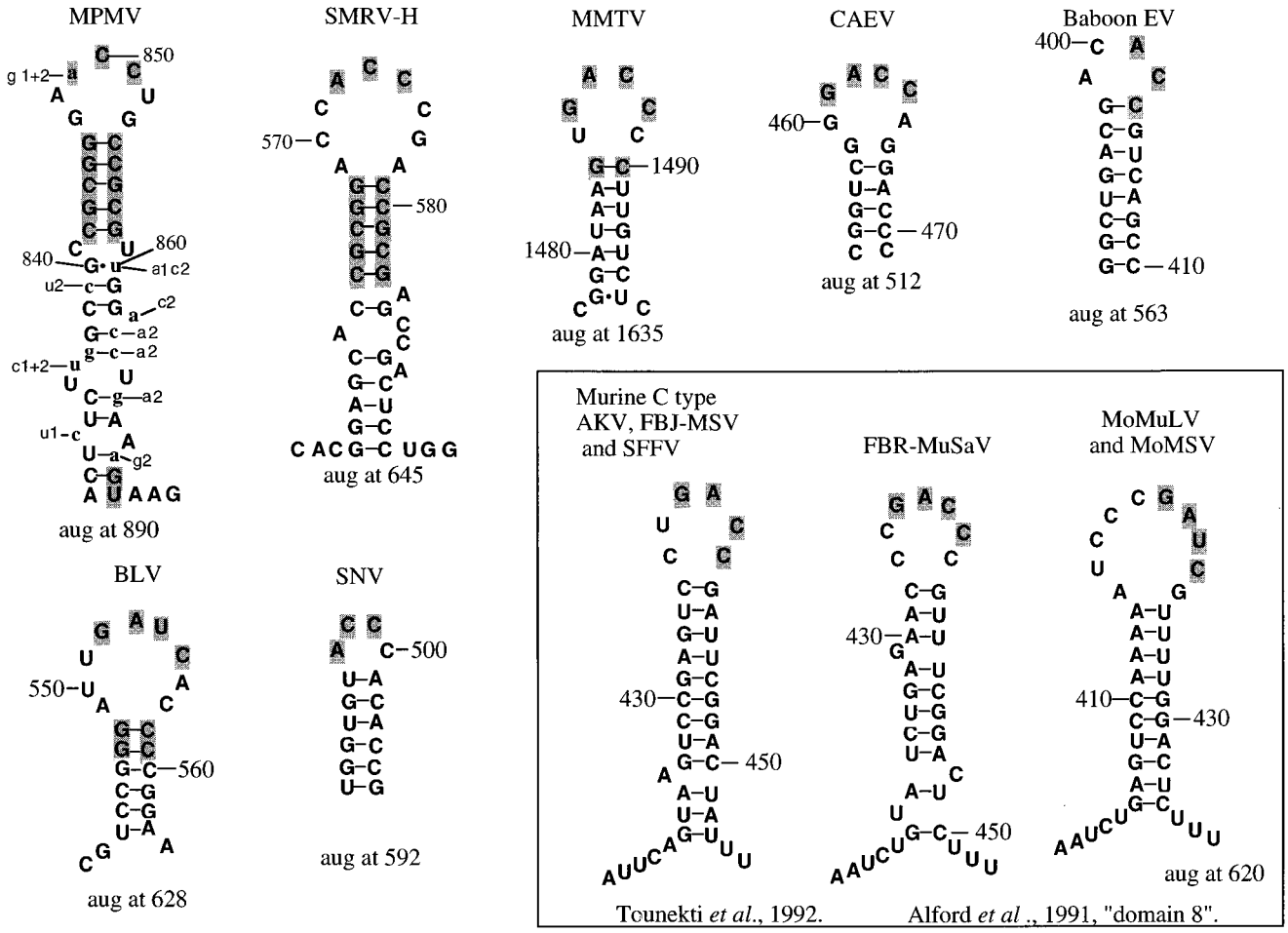

FIG. 12. Comparison of predicted stem loops in 5′ leader regions of 10 different retroviruses. Bases which are conserved between MPMV and SMRV-H are shaded, as are bases in analogous positions in the predicted structures of other viruses. The murine structures which are boxed were proposed by Tounekti et al. (58). The Moloney murine leukemia virus (MoMuLV) structure was first proposed by Alford et al. (3). Sequence variations for SRV-1 and SRV-2 are indicated on the MPMV structure as shown in Fig. 6. The positions of the *gag* AUG relative to the last C of the ACC motif are indicated for each virus. MMTV, mouse mammary tumor virus; CAEV, caprine arthritis-encephalitis virus; baboon EV, baboon endogenous virus; BLV, bovine leukemia virus; AKV, murine leukemia virus; FBJ-MSV, FBJ murine sarcoma virus; SFFV, spleen focus-forming virus; FBR-MuSaV, FBR murine sarcoma virus; MoMSV, Moloney murine sarcoma virus.

defined a region (bases 385 to 409) downstream of this double hairpin as being important for encapsidation, suggesting that there may be other RNA motifs involved in packaging. The sequence GACG is found in MPMV, twice between positions 788 and 794 in a region which is predicted to form a weak stem-loop structure (Fig. 9C), but this structure is not conserved in SRV-2. Covariant nucleotide changes in the same region strongly suggest that the structure shown in Fig. 9A has been conserved between MPMV and SRV-2. In SMRV-H the GACG motif can be found only within the PBS, where it is unlikely to be involved in packaging. In HIV-1 the same sequence is not predicted to be in a loop in any of the published models. While this motif has been demonstrated to affect packaging in SNV (63), it seems unlikely to be a major recognition signal in packaging of MPMV and closely related retroviruses, on the basis of the evidence presented here.

The stem loop from bases 878 to 898 containing the *gag* initiation codon is conserved between MPMV, SRV-1, SRV-2, and SMRV. Our free energy minimization studies predicted similar structures in this region for several other distantly related retroviruses. The structure may inhibit translation. We have demonstrated significant translational suppression by the HIV-1 Ψ structure (41). The structures at the *gag* AUG bear a striking resemblance to the group II GA coliphage (28, 53) and the fr coliphage (2) initiation codon regions. Mutagenesis in this region in coliphages (13) has shown that increasing the stability of this stem suppresses translation but disrupting it does not enhance translation.

We are able to present a stable secondary structure model for the 5′ leader region of MPMV, downstream of the PBS. We have identified a stem-loop structure in the region, which from other studies has been shown to be important for RNA encapsidation. This structure is predicted to involve a highly stable stem with a heptaloop. Comparison with a predicted structure for SMRV shows an identical stem and a common loop with an ACC motif. Computer predictions of the secondary structure of this region in eight other retroviruses point to the existence of stem loops with common GAYC motif. We believe that this may represent a common structural and sequence packaging motif in these viruses.

ACKNOWLEDGMENTS

This work was supported by the U.K. Medical Research Council AIDS Directed Programme, the Wellcome Trust U.K., and the Sykes Trust.

We thank F. M. Guesdon for providing unpublished data and plasmids and for helpful discussions.

REFERENCES

1. **Adam, M. A., and A. D. Miller.** 1988. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. J. Virol. **62:**3802–3806.
2. **Adhin, M. R., A. Avots, V. Berzin, G. P. Overbeek, and J. van Duin.** 1990. Complete nucleotide sequence of the group I RNA bacteriophage fr. Biochim. Biophys. Acta **1050:**104–109.
3. **Alford, R. L., S. Honda, C. B. Lawrence, and J. W. Belmont.** 1991. RNA secondary structure analysis of the packaging signal of Moloney murine leukemia virus. Virology **183:**611–619.
4. **Bachellerie, J.-P., J. F. Thompson, M. Wegnez, and J. F. Hearst.** 1981. Identification of the modified nucleotide produced by covalent photoaddition of hydroxymethyltrimethylpsoralen to RNA. Nucleic Acids Res. **9:**2207–2222.
5. **Baudin, F., R. Marquet, C. Isel, J.-L. Darlix, B. Ehresmann, and C. Ehresmann.** 1993. Functional sites, in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. J. Mol. Biol. **229:**382–397.
6. **Bebenek, K., J. Abbotts, J. D. Roberts, S. H. Wilson, and T. A. Kunkel.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J. Biol. Chem. **264:**16948–16956.
7. **Berkowitz, R. D., and S. P. Goff.** 1994. Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. Virology **202:**233–246.
8. **Berkowitz, R. D., J. Luban, and S. P. Goff.** 1993. Specific binding of human immunodeficiency virus type 1 *gag* polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. J. Virol. **67:**7190–7200.
9. **Bryant, M. L., M. B. Gardner, P. A. Marx, D. H. Maul, N. W. Lerche, K. J. Osborn, L. J. Lowenstine, A. Bogden, L. O. Arthur, and E. Hunter.** 1986. Immunodeficiency in rhesus monkeys associated with the original Mason-Pfizer monkey virus. J. Natl. Cancer Inst. **77:**957–965.
10. **Chiu, I.-M., and S. F. Skuntz.** 1986. Nucleotide sequence analysis of squirrel monkey retrovirus reveals a novel primer-binding site for $tRNA^{Lys}_{1,2}$. J. Virol. **58:**983–987.
11. **Chopra, H. C., and M. M. Mason.** 1970. A new virus in a spontaneous mammary tumor of a rhesus monkey. Cancer Res. **30:**2081–2086.
12. **Daniel, M. D., N. W. King, N. L. Letvin, R. D. Hunt, P. K. Sehgal, and R. C. Desrosiers.** 1984. A new type D retrovirus isolated from macaques with an immunodeficiency syndrome. Science **223:**602–605.
13. **De Smit, M., and A. van Duin.** 1990. Control of prokaryotic translational initiation by mRNA secondary structure. Prog. Nucleic Acid Res. Mol. Biol. **38:**1–35.
14. **Desrosiers, R. C., M. D. Daniel, C. V. Butler, D. K. Schmidt, N. L. Letvin, R. D. Hunt, N. W. King, C. S. Barker, and E. Hunter.** 1985. Retrovirus D/New England and its relation to Mason-Pfizer monkey virus. J. Virol. **54:**552–560.
15. **Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal.** 1986. HTLV III expression and production involve complex regulation at the levels of splicing and transduction of viral RNA. Cell **46:**807–817.
16. **Felber, B. K., and G. N. Pavlakis.** 1993. Molecular biology of HIV-1: positive and negative regulatory elements important for virus expression. AIDS **7**(Suppl.)**:**S51–S62.
17. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **35:**351–360.
18. **Fine, D. L., J. C. Landon, R. J. Pienta, M. T. Kubicek, M. J. Valerio, W. F. Loeb, and H. C. Chopra.** 1975. Responses of infant rhesus monkeys to inoculation with Mason-Pfizer monkey virus. J. Natl. Cancer Inst. **54:**651–658.
19. **Fox, G. E., and C. R. Woese.** 1975. 5S RNA secondary structure. Nature (London) **256:**505–507.
20. **Fry, M., and L. A. Loeb.** 1992. A DNA polymerase alpha pause site is a hot spot for nucleotide misinsertion. Proc. Natl. Acad. Sci. USA **89:**763–767.
21. **Guerrier-Takada, C., and S. Altman.** 1984. Structure in solution of M1 RNA, the catalytic subunit of ribonuclease P from *Escherichia coli*. Biochemistry **23:**6327–6334.
22. **Guesdon, F. M., S. S. Rhee, A. M. L. Lever, and E. Hunter.** Unpublished data.
23. **Gutell, R. R., B. Weiser, C. R. Woese, and H. F. Noller.** 1985. Comparative anatomy of 16S-like ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol. **32:**155–216.
24. **Harrison, G. P., and A. M. L. Lever.** 1992. The 5′ packaging signal and major splice donor region of human immunodeficiency virus type 1 have a conserved stable secondary structure. J. Virol. **66:**4144–4153.
25. **Hayashi, T., Y. Ueno, and T. Okamoto.** 1993. RNA packaging signal of human immunodeficiency virus type 1. FEBS Lett. **327:**213–218.
26. **Heidecker, G., N. W. Lerche, L. J. Lowenstine, A. Lackner, K. G. Osborn, M. B. Gardner, and P. A. Marx.** 1987. Induction of simian acquired immune deficiency syndrome (SAIDS) with a molecular clone of type D SAIDS retrovirus. J. Virol. **61:**3066–3071.
27. **Henrickson, R. V., D. H. Maul, K. G. Osborn, J. L. Sever, D. L. Madden, L. R. Ellingsworth, J. H. Anderson, L. J. Lowenstine, and M. B. Gardner.** 1983. Epidemic of acquired immunodeficiency in rhesus monkeys. Lancet **i:**388–390.
28. **Inokuchi, Y., R. Takahashi, T. Hirose, S. Inayama, A. B. Jacobson, and A. Hirashima.** 1986. The complete nucleotide sequence of the group II RNA coliphage GA. J. Biochem. **99:**1169–1180.
29. **Jaeger, J. A., D. H. Turner, and M. Zucker.** 1989. Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA **86:**7706–7710.
30. **Jensen, E. M., I. Zelljadt, H. C. Chopra, and M. M. Mason.** 1970. Isolation and propagation of virus from a spontaneous mammary carcinoma of a rhesus monkey. Cancer Res. **30:**2388–2393.
31. **Kean, J. M., and D. E. Draper.** 1985. Secondary structure of 345-base RNA fragment covering the S8/S15 protein binding domain of *Escherichia coli* 16S ribosomal RNA. Biochemistry **24:**5052–5061.
32. **Kim, H.-J., K. Lee, and J. J. O'Rear.** 1994. A short sequence upstream of the 5′ major splice site is important for encapsidation of HIV-1 genomic RNA. Virology **198:**336–340.
33. **Knight, J. B., Z. H. Si, and C. M. Stolzfus.** 1994. A base paired structure in the avian sarcoma virus 5′ leader is required for efficient encapsidation of RNA. J. Virol. **68:**4493–4502.
34. **Konings, D. A., M. A. Nash, J. V. Maizel, and R. B. Arlinghaus.** 1992. Novel GACG-hairpin motif in the 5′ untranslated region of type C retroviruses related to murine leukemia virus. J. Virol. **66:**632–640.
35. **Leffers, H., J. Egebjerg, A. Anderson, T. Christensen, and R. A. Garrett.** 1988. Domain VI of Escherichia coli 23S ribosomal RNA structure, assembly and function. J. Mol. Biol. **204:**507–522.
36. **Lever, A. M. L., H. Gottlinger, W. Haseltine, and J. Sodroski.** 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J. Virol. **63:**4085–4087.
37. **Lever, A. M. L., J. H. Richardson, and G. P. Harrison.** 1991. Retroviral RNA packaging. Biochem. Soc. Trans. **19:**963–966.
38. **Linial, M. L., and A. D. Miller.** 1990. Retroviral RNA packaging: sequence requirements and implications, p. 125–152. *In* R. Swanstrom and P. K. Vogt (ed.), Retroviruses: strategies of replication. Springer-Verlag, Berlin.

38a. **Luban, J., and S. Goff.** 1994. Mutational analysis of *cis*-acting packaging signals in human immunodeficiency virus type-1 RNA. J. Virol. **68:**3784–3793.

39. **Marx, P. A., M. L. Bryant, K. G. Osborn, D. H. Maul, N. W. Lerche, L. J. Lowenstine, J. D. Kluge, C. P. Zaiss, R. V. Henrickson, S. M. Shiigi, B. J. Wilson, A. Malley, L. C. Olsen, W. P. McNulty, L. O. Arthur, R. V. Gilden, C. S. Barker, E. Hunter, R. J. Munn, G. Heidecker, and M. B. Gardner.** 1985. Isolation of a new serotype of simian acquired immune deficiency syndrome type D retrovirus from Celebes black macaques (*Macaca nigra*) with immune deficiency and retroperitoneal fibromatosis. J. Virol. **56:**571–578.
40. **Marx, P. A., D. H. Maul, K. G. Osborn, N. W. Lerche, P. Moody, L. J. Lowenstine, R. V. Henrickson, L. O. Arthur, R. V. Gilden, M. Gravell, W. T. London, J. L. Sever, J. A. Levy, R. J. Munn, and M. B. Gardner.** 1984. Simian AIDS: isolation of a type D retrovirus and transmission of the disease. Science **223:**1083–1086.
41. **Miele, G., G. P. Harrison, and A. M. L. Lever.** Submitted for publication.
42. **Nishikawa, K., and S. Takemura.** 1974. Nucleotide sequence of a 5S RNA from *Torulopsis utilis*. FEBS Lett. **40:**106–109.
43. **Oda, T., S. Ikeda, S. Watanabe, M. Hatsushika, K. Akiyama, and F. Mitsunobu.** 1988. Molecular cloning, complete nucleotide sequence, and gene structure of the provirus genome of a retrovirus produced in a human lymphoblastoid cell line. Virology **167:**468–476.
44. **Osterburg, G., and R. Sommer.** 1981. Computer support of DNA sequence analysis. Comput. Programs Biomed. **13:**101–109.
45. **Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science **231:**1567–1572.
46. **Raba, M., K. Limberg, M. Burghagen, J. R. Katze, M. Simsek, and J. E. Heckman.** 1990. Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. **97:**305–318.
47. **Rhee, S. S., H. Hui, and E. Hunter.** 1990. Preassembled capsids of type D retroviruses contain a signal sufficient for targetting specifically to the plasma membrane. J. Virol. **64:**3844–3852.
48. **Rosen, C. A.** 1991. Regulation of HIV gene expression by RNA-protein interactions. Trends Genet. **7:**9–14.
49. **Sakaguchi, K., N. Zambrano, E. T. Baldwin, B. A. Shapiro, J. W. Erickson, J. G. Omichinski, G. M. Clore, A. M. Gronenborn, and E. Appella.** 1993. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. Proc. Natl. Acad. Sci. USA **90:**5219–5223.
50. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **79:**5463–5467.
51. **Schlesinger, S., S. Makino, and M. L. Linial.** 1994. *cis*-acting genomic elements and *trans*-acting proteins involved in the assembly of RNA viruses. Semin. Virol. **5:**39–49.
52. **Shelness, G. S., and D. L. Williams.** 1985. Secondary structure analysis of apolipoprotein II mRNA using enzymatic probes and reverse transcriptase. J. Biol. Chem. **260:**8637–8646.

53. **Skripkin, E., M. R. Adhin, M. H. de Smit, and J. van Duin.** 1990. Secondary structure of the central region of bacteriophage MS2 RNA. J. Mol. Biol. **211:**447–463.
54. **Sommerfelt, M. A., C. R. Roberts, and E. Hunter.** 1993. Expression of simian type D retroviral (Mason-Pfizer monkey virus) capsids in insect cells using recombinant baculovirus. Virology **192:**298–306.
55. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus, an immunosuppresive D-type retrovirus. Cell **45:**375–385.
56. **Steffy, K., and F. Wong Staal.** 1991. Genetic regulation of human immunodeficiency virus. Microbiol. Rev. **55:**193–205.
57. **Thayer, M. R., M. D. Power, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1987. Sequence relationships of type D retroviruses which cause simian acquired immunodeficiency syndrome. Virology **157:**317–329.
58. **Tounekti, N., M. Mougel, C. Roy, R. Marquet, J.-L. Darlix, J. Paoletti, B. Ehresmann, and C. Ehresmann.** 1992. Effect of dimerization on the conformation of the encapsidation *Psi* domain of Moloney murine leukemia virus RNA. J. Mol. Biol. **223:**205–220.
59. **Vile, R. G., M. Ali, E. Hunter, and M. O. McClure.** 1992. Identification of a generalised packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. Virology **189:**786–791.
60. **Watanabe, S., and H. M. Temin.** 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5′ long terminal repeat and the start of the *gag* gene. Proc. Natl. Acad. Sci. USA **79:**5986–5990.
61. **Wimberly, B., G. Varani, and I. Tinoco, Jr.** 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. Biochemistry **32:**1078–1087.
62. **Woese, C. R., and N. R. Pace.** 1993. Probing RNA structure, function and history by comparative analysis, p. 91–118. *In* R. F. Gesteland and J. J. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
63. **Yang, S., and H. M. Temin.** 1994. A double hairpin structure is necessary for the encapsidation of spleen necrosis virus retroviral RNA. EMBO J. **13:**713–726.
64. **Youvan, D. C., and J. E. Hearst.** 1982. Sequencing psoralen photochemically reactive sites in *Escherichia coli* 16S rRNA. Anal. Biochem. **119:**86–89.
65. **Zuker, M.** 1989. On finding all suboptimal foldings of an RNA. Science **244:**48–52.