JOURNAL OF VIROLOGY, July 1991, p. 3864–3872
0022-538X/91/073864-09$02.00/0

Vol. 65, No. 7

Overlapping Retrovirus U5 Sequence Elements Are Required for Efficient Integration and Initiation of Reverse Transcription

DAVID COBRINIK,† ASHOK AIYAR, ZHENG GE, MICHAEL KATZMAN,‡ HAN HUANG,§ AND JONATHAN LEIS*

Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

Received 4 February 1991/Accepted 9 April 1991

A secondary structure in the 5′ noncoding region of avian retrovirus RNA, called the U5-leader stem, was shown previously to have a role in initiation of reverse transcription (D. Cobrinik, L. Soskey, and J. Leis, J. Virol. 62:3622–3630, 1988). We now show that an additional RNA secondary structure near the U5 terminus, called the U5-IR stem, is also important for reverse transcription. Mutations that disrupt the U5-IR stem cause a replication defect associated with both a decrease in synthesis of viral DNA in infected cells and a decrease in initiation of reverse transcription in melittin-permeabilized virions. Structure-compensating base substitutions in the U5-IR restore reverse transcription efficiency. In viral DNA, U5-IR sequences are included in the U5 terminal region that functions as a viral integration donor site. When base substitutions are introduced into these sequences, a reduced efficiency of integration in vitro and in vivo is observed. These observations indicate that U5-IR sequences have a structural role in reverse transcription of viral RNA and a sequence-specific role in the integration of viral DNA.

The early steps of retrovirus replication consist of reverse transcription of viral RNA and integration of viral DNA into cellular DNA. Reverse transcription is initiated by extension of a specific tRNA primer annealed to sequences near the 5′ end of the viral genome. A number of complex steps result in the production of double-stranded linear viral DNA containing long terminal repeat (LTR) sequences in the cytoplasm of infected cells as well as linear and circular viral DNA forms in cell nuclei (8). The linear form has recently been shown to be the primary substrate for integration. This reaction involves cleavage of the U5 and U3 LTR termini by the viral IN protein followed by ligation to cellular sequences that are also likely to be cleaved by IN (5, 12, 15, 21, 26).

Avian retrovirus particles are infectious as early as 10 min after budding (30), yet there is no detectable reverse transcription prior to cell infection (4, 13, 31). The mechanism that restricts reverse transcription in virions is not known. However, we have recently found that an RNA secondary structure called the U5-leader stem affects the efficiency of initiation and have proposed that this structure serves such a regulatory role (7).

The U5-leader stem is a potential RNA secondary structure consisting of sequences in the middle of U5 and in the leader region adjacent to the primer-binding site (7). Mutations which disrupt this structure impair reverse transcription in infected cells, while mutations which alter the sequence but retain the stem structure have no effect (7). Studies with permeabilized virions have localized the defect to the initiation of reverse transcription. Disruption of the U5-leader stem structure causes a decrease in incorporation of the first deoxynucleotides from the 3′ OH terminus of the $tRNA^{Trp}$ primer, while the amount of the $tRNA^{Trp}$ primer annealed to the RNA primer-binding site is equivalent to that found in wild-type virions (7). U5-leader stem structures also appear to be conserved among most retrovirus families (1, 7).

The U5-leader stem of avian retroviruses is in a region of RNA which can potentially form numerous secondary structures (14, 17). In this study we investigated whether one of these other structures, termed the U5-inverted repeat (U5-IR) stem (see Fig. 1A), could also affect initiation of reverse transcription. In viral RNA, this structure is composed of an inverted repeat sequence adjacent to the primer-binding site. In viral DNA, the U5-IR stem falls within the U5 terminal sequences that are recognized by the IN protein and functions as a viral integration site (21, 22). Thus, mutations that disrupt the U5-IR stem structure and alter its nucleotide sequence could also result in defects to integration. Our data indicate that U5-IR sequences are indeed required for both replication functions.

MATERIALS AND METHODS

Reagents. The Klenow fragment of DNA polymerase I (4 U/μl), T4 DNA ligase (2,000 U/μl), T4 polynucleotide kinase (10 U/μl), and various restriction enzymes were from Boehringer Mannheim Biochemicals or New England Biolabs, Inc. [α-^{32}P]dATP (800 Ci/mmol), [α-^{32}P]dCTP (3,000 Ci/mmol), and [γ-^{32}P]ATP (6,000 Ci/mmol) were from New England Nuclear Corp. pTZ18 DNA was obtained from Pharmacia. Thermostable DNA polymerase (Pyrostase; 4 U/μl) was from Molecular Genetics Resources.

Construction of U5-IR viral mutants. pRCAS-B and pRCAS-*neo* were generously provided by S. Hughes (National Cancer Institute-Frederick Basic Research Facility). pRCAS-B is the pRCAS vector (20), except that it has a central *Eco*RI fragment (spanning *gag* and *pol* genes) derived from a cloned Bryan high-titer Rous sarcoma virus (RSV) genome. pRCAS-*neo* contains the neomycin phosphotransferase gene inserted into the pRCAS *Cla*I cloning

* Corresponding author.
† Present address: Whitehead Institute for Biomedical Research, Cambridge, MA 02142.
‡ Present address: Department of Medicine, The Milton S. Hershey Medical Center, Pennsylvania State University College of Medicine, Hershey, PA 17033.
§ Present address: Department of Pathology and Immunology, Case Western Reserve University School of Medicine, Cleveland, OH 44106.

site. pDC101B was derived from pRCAS-B by deletion of the *Eco*RV fragment between nucleotide 3655 (28) and pBR322 nucleotide 189. pDC102 was constructed by digestion of RCAS-*neo* with *Sac*I, partial digestion with *Rsr*II, formation of blunt ends with the Klenow enzyme, and ligation to *Hin*dII-digested pTZ18. The *gag-pol Eco*RI fragment was then replaced with that of the Bryan high-titer clone to form pDC102B. U5 mutations were introduced into pDC101B by insertion of synthetic deoxynucleotides between the *Rsr*II and *Bst*EII sites. All clones were confirmed by sequencing.

Transfections were performed by the calcium phosphate precipitation technique (27) with ligation products derived from 7 μg of *Hpa*I-digested pDC102B and 5 μg of *Hpa*I-digested pDC101B (or its derivatives) per semiconfluent 60-mm-diameter dish of QT6 cells. The cells were fed 1 day after transfection and split on day 2, and selection was initiated on day 2. Virus production was monitored by an exogenous reverse transcriptase assay (7) with the following modifications: the final Nonidet P-40 concentration was 0.067%, and 15 μl of virus was assayed per 30-μl reaction mixture.

Analysis of reverse transcription products in permeabilized virions. Virus preparations and melittin reactions were as previously described (7), except that the reaction volumes were 30 μl. The amount of virus used in each reaction was that which incorporated 0.5 pmol of dCTP in a 30-min reverse transcriptase assay performed on a 1/10 dilution of the recovered virus and supplemented with excess dCTP. The final concentration and specific activity of [α-^{32}P]dATP were 2 μM and 330 Ci/mmol, respectively, when dATP was present as the only deoxynucleotide, and 25 μM and 40 Ci/mmol, respectively, when it was used in combination with 1 mM (each) dCTP, dGTP, and dTTP. The dATP was not rate limiting at these concentrations. The final melittin concentration was 7.5 μg/ml.

Analysis of integrated viral DNA. High-molecular-weight cellular DNA was prepared by the method of Hirt (19). The DNA was sheared by passing it through a 21-gauge needle several times, heated to 55°C, layered onto a gradient of 10 to 40% sucrose in 0.1 M NaCl–10 mM Tris · HCl (pH 8.0)–1 mM EDTA, and spun for 16 h at 25,000 rpm in an SW28 rotor at 4°C. Fractions containing DNA of >25 kb were combined, precipitated with 2 volumes of ethanol, and used in polymerase chain reactions (PCRs). Primers complementary to the *gag* gene (nucleotides 2251 to 2270) and to the *pol* gene (nucleotides 2667 to 2583) (28) were used in PCRs with the following cycling conditions: 92°C for 2 min, 55°C for 90 s, and then 72°C for 2 min. Amplification of pRCAS-B DNA, at concentrations between 0.01 to 1 pg, was linear up to 25 cycles. The cell-derived DNA being analyzed for viral DNA was thus amplified for 24 cycles as described elsewhere (23) by using Pyrostase and a Hybaid thermal cycler. Amplified DNA products were subjected to agarose gel electrophoresis and transferred to nylon membranes (27), and viral DNA sequences were detected by hybridization to [^{32}P]dCTP-labeled probes (2×10^9 cpm/μg). The probes were prepared by DNA polymerase I (Klenow fragment) extension of primers complementary to the *gag* and *pol* genes and DNA containing portions of these genes amplified from the viral clone pATV8 (6).

IN protein cleavage reactions. Reactions were performed as described previously (22).

Integrative recombination in vitro. Substrate oligodeoxynucleotides representing wild-type or mutant U5 sequences were prepared as described previously (22) and are as follows:

U5 wild-type plus strand	5′TGAAGCAGAAGGCTTCA3′
U5 wild-type minus strand	3′TCGTCTTCCGAAGTAA5′
U5 S2 plus strand	5′AGCAGAAGGC<u>AA</u>CA3′
U5 S2 minus strand	3′TCGTCTTCCG<u>TT</u>GTAA5′
U5 S4 plus strand	5′AGCAGAAGG<u>GAAG</u>A3′
U5 S4 minus strand	3′TCGTCTTCC<u>CTTC</u>TAA5′
U5 S4C plus strand	5′<u>TC</u>CAGAAGG<u>GAAG</u>A3′
U5 S4C minus strand	3′<u>AG</u>GTCTTCC<u>CTTC</u>TAA5′

The underlined sequences indicate base changes from the wild type for the S2, S4, and S4C mutations (see Fig. 1B and C). After the plus and minus strands are annealed, the resultant duplexes have a two-base 5′ minus-strand overhang.

Integrative recombination reactions were carried out in a 10-μl reaction volume containing 1 pmol of each of the duplexes shown above, 5′ ^{32}P labeled only in the plus strand (10^5 cpm/pmol); 25 mM Tris · HCl (pH 8.3); 10 mM 2-mercaptoethanol; 3 mM $MgCl_2$; and 1 pmol of IN. The reaction mixtures were incubated at 37°C for 90 min. The reactions were stopped by the addition of 10 μl of sequencing loading buffer (22), and the mixtures were heated to 90°C for 5 min and analyzed on a 20% polyacrylamide–7 M urea sequencing gel (22).

RESULTS

The RSV (Schmidt-Ruppin A) RNA sequence from nucleotides 55 to 131, which includes the primer-binding site and the previously studied U5-leader stem structure (7), is shown in Fig. 1A. This region also includes an additional secondary structure termed the U5-IR stem. This structure comprises the inverted repeat sequence between nucleotides 81 and 101 (U5-IR) and can potentially form a 7-bp stem with a 5-base loop. Analyses of the RSV 5′ noncoding region have indicated that the U5-IR stem is thermodynamically stable and is conserved among all members of the avian sarcoma and leukosis retrovirus family for which sequence information is available (1, 3, 7). This structural conservation suggests that the U5-IR stem may have an important role in viral replication. We sought to understand this role by analyzing the effects of mutations in the U5-IR sequence on viral replication.

Production of U5-IR mutant virus. The virus vector system used for these studies was derived from the pRCAS-*neo* vector described by Hughes et al. (20). pRCAS-*neo* is a Schmidt-Ruppin A strain of RSV in which the viral *src* gene is replaced by the neomycin phosphotransferase gene. Two plasmids were derived from pRCAS-*neo*: pDC101B, which contains the 5′ half of pRCAS-*neo*, and pDC102B, which contains the majority of the viral genome but lacks sequences 5′ of the leader region *Sac*I site and those 3′ of the U5 region *Rsr*II site in the downstream LTR (Fig. 2). Both of these plasmids contain Bryan high-titer viral sequences, which confer the high-titer growth phenotype, within portions of the *gag* and *pol* genes.

To obtain virus-producer cells, pDC102B and pDC101B were digested with *Hpa*I, ligated to assemble a full-length viral transcription unit, and transfected into quail QT-6 cells. Cells that expressed the virus-encoded *neo* gene were then selected in the presence of G418. The advantages of this

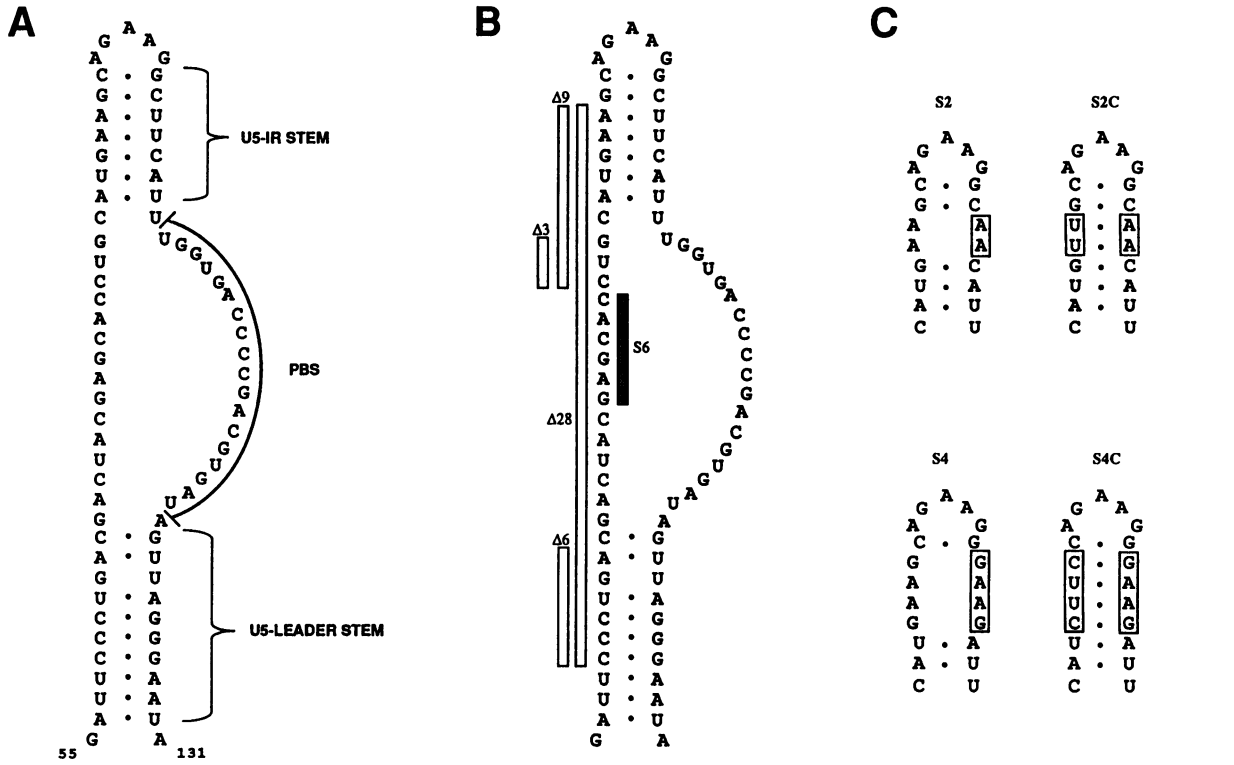

FIG. 1. Predicted RNA secondary structures for the 5′ noncoding region of wild-type and mutant RSV (Schmidt-Ruppin A) RNA near the primer-binding site (PBS). (A) Wild-type structure. Numbers indicate nucleotides from the viral RNA 5′ end. The PBS, U5-leader stem, and U5-IR stem are indicated. (B) U5-IR stem mutations Δ3 and Δ9. The Δ6, Δ28, and S6 mutations are described in the text. They are analyzed elsewhere (1, 6, 7). Sequence deletions are indicated by open bars; the sequence substitution is indicated by the stippled bar. (C) U5-IR stem substitution and compensating substitution mutations S2, S2C, S4, and S4C. Sequence substitutions are boxed, and the effects of these substitutions on the U5-IR stem structure are shown.

system are that U5 mutations can easily be cloned into pDC101B (see Material and Methods), that transfections can be performed without additional subcloning steps, and that large amounts of defective virus can be produced from QT-6 cells. In addition, truncation of the downstream LTR of pDC102B ensures that mutations introduced into the pDC101B U5 region cannot undergo reversion by recombination with wild-type U5 sequences.

Three types of mutations were introduced into the U5-IR stem: deletion mutations Δ3 and Δ9, substitution mutations S2 and S4, and structure-compensating substitution mutations S2C and S4C (Fig. 1B and C). The Δ3 mutation does not disrupt the U5-IR stem but decreases the spacing between the U5-IR and U5-leader stem structures. The Δ9, S2, and S4 mutations disrupt the U5-IR stem, while the S2C and S4C mutations retain the U5-IR stem structure but alter its sequence. We found that release of virus, as measured by reverse transcriptase activity, from pooled producer cells was roughly the same for the wild type and each of the mutant viruses described above and showed no correlation with viral growth rates (data not shown). This strongly suggests that late replication events such as transcription, splicing, translation, particle assembly, and budding were not affected by the U5-IR mutations.

Analysis of viral growth rates. Viruses with each of the mutations described above, obtained from QT-6 producer cells, were used to measure the rate of viral growth following infection of fresh QT-6 cells. Reverse transcriptase assays using exogenous template primers were performed directly on cell supernatants both to equalize the amount of virus used for infection and to monitor the appearance of progeny virus. Results represent the average of several infections which had nearly identical rates of viral growth.

The substitution mutations S2 and S4 resulted in impaired viral growth compared with the wild type (Fig. 3). The Δ9 and Δ3 deletion mutations also resulted in impaired and delayed growth of virus (data not shown). The viruses with structure-compensating mutations, S2C and S4C, displayed improved growth kinetics compared to their uncompensated S2 and S4 counterparts, but were still significantly defective in growth compared with the wild type (Fig. 3). The differences between S2 and S2C and between S4 and S4C were reproducibly observed in four separate kinetic virus growth experiments. This partial restoration of growth efficiency for S2C and S4C suggests that the U5-IR stem structure has a role in replication that is at least partially sequence independent. However, the delayed growth observed with viruses containing structure-compensating mutations indicates that there is a major sequence-dependent replication function encoded in the U5-IR.

Each of the virus mutants analyzed in this study is only partially defective. The eventual appearance of virus is unlikely to be the result of genetic reversions, since the progeny virus exhibited the same impaired-growth phenotype as that exhibited by virus derived from the original producer cells (data not shown).

Analysis of reverse transcription in permeabilized virions. We have previously shown that reverse transcription is

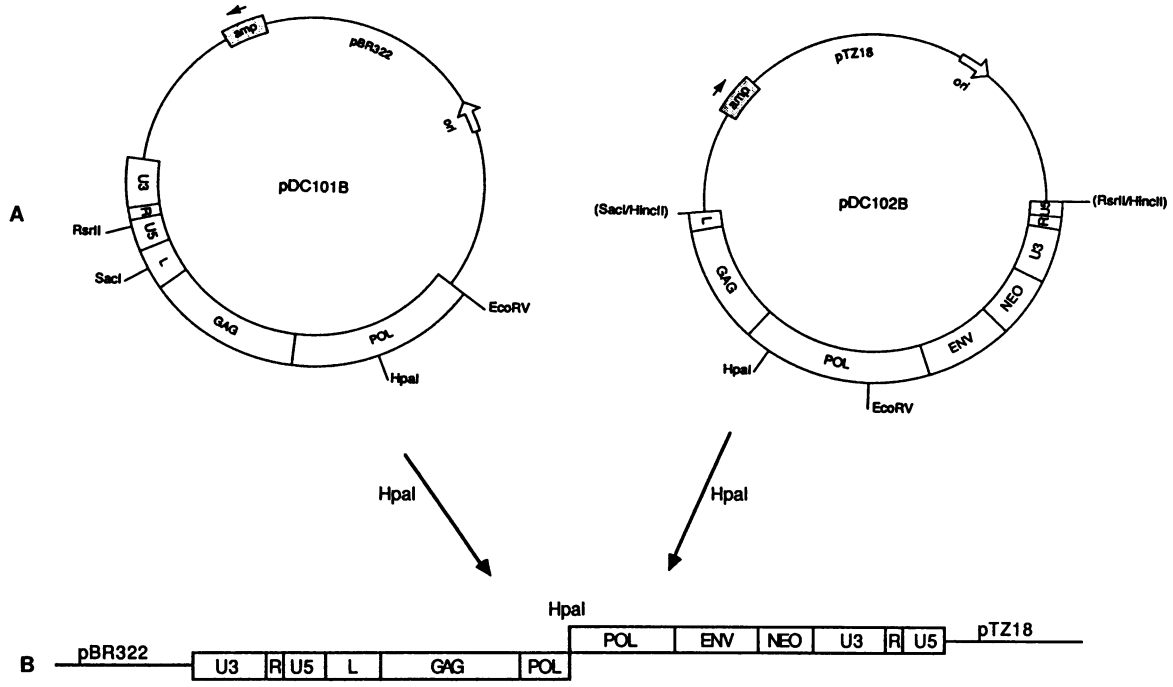

FIG. 2. Strategy for production of virus containing U5 mutations. pDC101B (or U5 mutant derivatives) and pDC102B are digested with *HpaI*; the products are ligated and transfected into QT-6 cells. (A) Structure of pDC101B and pDC102B. Wide bars indicate viral sequences. Thin lines indicate prokaryotic vector sequences. The U3, R, U5, and leader (L) noncoding regions and the *gag*, *pol*, *env*, and *neo* genes are indicated. The sizes of indicated regions of the viral genome are not to scale. ori, origin of replication. (B) Structure of the ligation product from which both the *neo* gene and full-length viral RNA can be expressed.

substantially less efficient when the U5-leader stem structure is disrupted (7). This was established by analyzing the synthesis of the initial intermediate in reverse transcription, a 101-nucleotide cDNA ($cDNA_{101}$) extending from the $tRNA^{Trp}$ primer to the 5′ terminus of viral RNA, in melittin-activated virions (Fig. 4E). In order to evaluate whether mutations in the U5-IR stem also affect reverse transcription, equal amounts of wild-type and mutant virus, determined by an exogenous reverse transcriptase assay, were incubated with melittin and [α-^{32}P]dATP in the presence of dCTP, dGTP, and dTTP. cDNAs were isolated from virions and treated with alkali to remove the covalently linked $tRNA^{Trp}$, and the amount of $cDNA_{101}$ was determined by polyacrylamide gel electrophoresis (Fig. 4B). The substitution mutations S2 and S4 resulted in decreased amounts of $cDNA_{101}$ synthesized, with a larger decrease being observed for the S2 virions. Structure-compensating mutations restored the reverse transcription efficiency, with that of S4C restored essentially to the wild-type level. Only a partial restoration was obtained with S2C (Fig. 4B). These differences are not due to a packaging defect in S2C and S4C (see below).

The deletion mutations Δ3 and Δ9 also resulted in reverse transcription defects. Because of the deletion of U5 sequences, the cDNA products in these cases were three and nine bases shorter than that of the wild type, respectively (Fig. 4C). The Δ9 mutation resulted in barely detectable amounts of $cDNA_{92}$, while the Δ3 mutation resulted in reduced levels of $cDNA_{98}$ compared with the wild type. These results correlated with the growth characteristics of Δ3 and Δ9.

The first two deoxynucleotides incorporated onto the 3′ OH terminus of the $tRNA^{Trp}$ primer are dAMPs (Fig. 4E). Thus, initiation of reverse transcription can be determined by monitoring the appearance of a dA-$tRNA^{Trp}$ formed in virions when [α-^{32}P]dATP is included in the melittin reaction mixtures as the only deoxynucleoside triphosphate. The relative amounts of dA-$tRNA^{Trp}$ produced with the wild type and each of the mutant viruses (Fig. 4A and D) were similar to the amounts of cDNA intermediate products obtained when all deoxynucleotides were present. These results indicate that disruption of the U5-IR stem impairs initiation of reverse transcription in melittin-activated virions.

Analysis of full-length double-stranded viral DNA in the cytoplasm of cells 12 h after infection with each of the U5 mutant viruses described above yielded results parallel to those observed in Fig. 4; the amount of viral DNA detected by Southern blotting correlated with the amounts of strong-stop cDNA and dA-$tRNA^{Trp}$ produced in melittin-activated virions (data not shown). These results cannot be explained by a RNA packaging defect, since equal amounts of wild-type and mutant virus (as judged by reverse transcriptase activity measured with exogenous template primers) contained equivalent amounts of viral RNA (as judged by a dot blot assay using a labeled viral cDNA probe [data not shown]). Taken together, these results indicate that the U5-IR stem structure, like the U5-leader stem structure, is necessary for efficient initiation of reverse transcription in infected cells.

Effect of U5 mutations on integration in vivo. The in vitro and in vivo analyses described above revealed that reverse transcription was not defective for the S4C mutant and was only partially defective for the S2C mutant. Since these

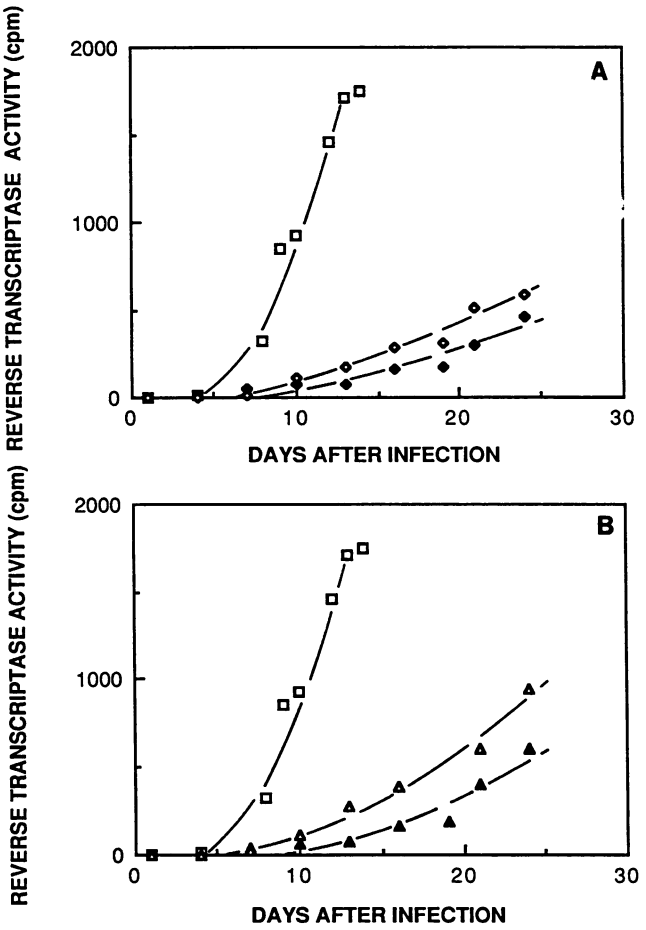

FIG. 3. Viral growth rates following infection of QT-6 cells with equal amounts of wild-type and U5-IR substitution mutant virus. The virus was detected by assay of reverse transcriptase activity in cell supernatants as described in Materials and Methods. □, wild-type virus. (A) ◆, S2 virus; ◇, S2C virus. (B) ▲, S4 virus; △, S4C virus.

viruses display a considerable reduction in growth efficiency (Fig. 3), replication steps other than reverse transcription seemed likely to be affected. We felt that integration could be impaired in such mutants, since the S2, S2C, S4, and S4C substitutions alter the U5 terminal sequence recognized by the viral IN protein (22). To evaluate this possibility, the in vivo integration efficiency for mutant virus was compared to that of the wild type.

Equal amounts of each mutant virus (measured by reverse transcriptase activity) were used to infect QT-6 cells, and the high-molecular-weight Hirt pellet DNA (19), which contains integrated provirus sequences, was extracted from cell nuclei after 12 h. In order to minimize the amount of contaminating unintegrated viral DNA, high-molecular-weight DNA (length of >25 kb) was isolated by sucrose density gradient centrifugation as described in Materials and Methods. Viral DNA was then detected by amplifying a 418-bp fragment from the *pol* region by the PCR technique and then hybridizing the fragment by Southern blotting to a ^{32}P-labeled viral DNA probe. Various concentrations of pRCAS-B DNA (0.01 to 1 pg) were amplified (Fig. 5A) in parallel reactions. This approach is similar to that previously described for quantitating cellular mRNA levels by PCR (27).

The amount of viral DNA detected after 24 cycles of amplification for each of the U5 mutants is shown in Fig. 5B. Comparison of the amount of amplified 418-bp fragment from the wild type (Fig. 5B, lane 4) with control DNA (Fig. 5A, lane 2) indicated that approximately 0.1 pg of the former was present before amplification. In the absence of virus infection, amplification of a 418-bp fragment was not detected (Fig. 5B, lane 3). The amounts of S2, S2C, S4, S4C, Δ9, and Δ3 proviral DNA were estimated by densitometry to be 18, 25, 29, 44, 7, and 20% of the amount of wild-type DNA, respectively (Fig. 5B).

A decrease in the amount of proviral DNA detected in cells under these conditions could result from either a direct effect upon integration or a decrease in the amount of the linear viral DNA precursor to integration synthesized during the early steps of viral replication. Since the S4C virus does not display a reverse transcription defect, we estimate that this mutation must cause more than a twofold decrease in the efficiency of integration of viral DNA. Since the S2, S4, and S2C mutants have defects in reverse transcription, the decrease in integrated provirus cannot be attributed solely to an integration defect. For example, the fivefold decrease in viral DNA detected for the S2C mutant might result from a three- to fourfold decrease in integration efficiency and a 1.5-fold decrease in reverse transcription efficiency (Fig. 4). Similar combined reverse transcription and integration defects are likely to occur for the S2 and S4 viruses. For the Δ3 and Δ9 viruses, the reduction in proviral DNA is likely to result solely from defects in reverse transcription, since the integration donor site sequences are intact in these mutants (21, 22).

Effect of U5 mutations on integration in vitro. To further define integration defects produced by U5-IR mutations, duplex oligodeoxynucleotides corresponding to the 16 terminal U5 bases of these mutants were evaluated for their ability to be nicked by the viral IN protein in vitro (Fig. 6A). Previous studies have shown that purified avian retrovirus IN protein nicks double-stranded DNA at the in vivo viral integration site, two bases from the U3 and U5 LTR termini (22). In this assay, cleavage of a 5′-^{32}P-labeled plus-strand 16-mer at a site two bases from the U5 terminus is monitored by conversion to a 14-mer.

In the presence of Mg^{2+}, the S2, S4, and S4C mutations nearly abolished the ability of the U5 terminal sequences to be cleaved (Fig. 6B, lanes b, e, h, and k). However, some 14-mer product could be detected for the S4C substrate after a longer exposure of the autoradiogram (Fig. 6B, lane n). In the presence of Mn^{2+}, the endonuclease activity of IN was 20 times higher than in the presence of Mg^{2+}, but cleavage occurred at many sites (lane c). When analyzed with Mn^{2+}, cleavage of the S2 substrate was substantially reduced (lane f). Cleavage of the S4 and S4C substrates was detected, although the total amount of cleavage was less than with the wild-type substrate (lanes f and i). These data indicate that base changes in the U5-IR stem decrease the efficiency of cleavage by IN and that the two-base substitution (S2) is more deleterious to cleavage than the four-base substitution (S4).

Oligodeoxynucleotides can also be used to demonstrate the joining reaction in vitro (12, 21). This is evidenced by the appearance of oligodeoxynucleotides longer than the starting substrate which arise by a self-integration reaction. The substrates used in these experiments represent LTR termini already nicked at the integration site. This allows one to examine primarily the joining reaction, although some DNA

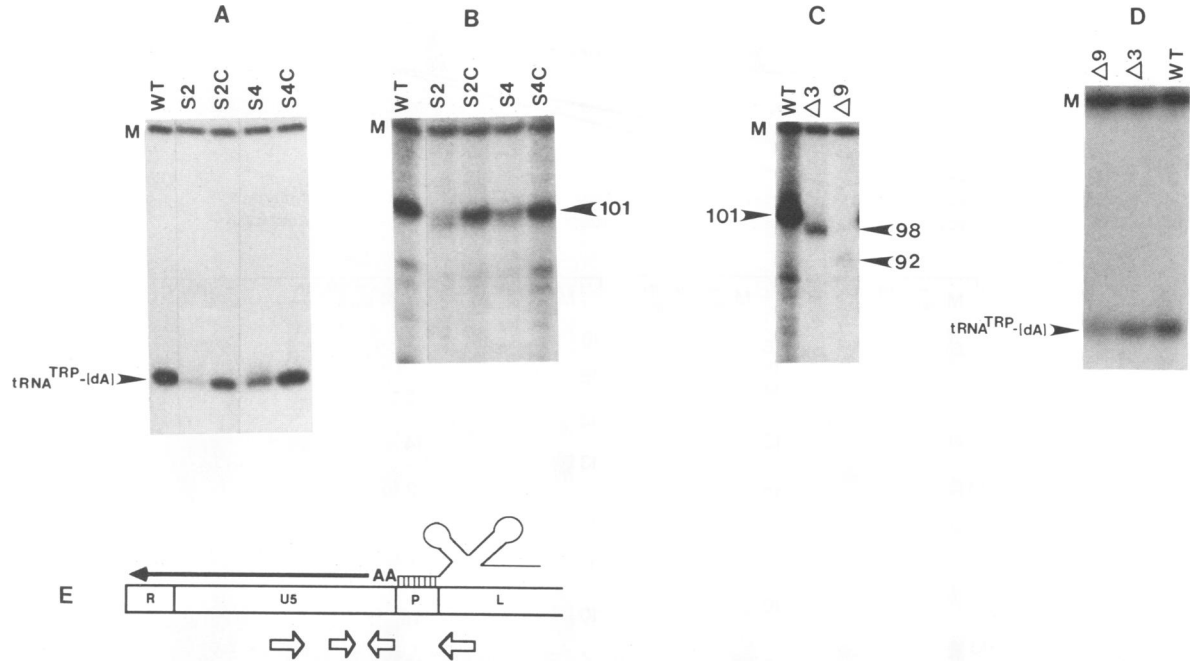

FIG. 4. Analysis of reverse transcription products for wild-type and U5-IR mutant virus in melittin-activated virions. Shown are $tRNA^{Trp}$ primer extension products formed after 30 min of incubation of virions with melittin and [^{32}P]dATP in the absence (A and D) or the presence (B and C) of the other three unlabeled deoxynucleotides. The viruses (wild type [WT] or U5 mutants Δ3, Δ9, S2, S2C, S4, and S4C) are indicated above the respective lanes. The migration distances of the dA-$tRNA^{Trp}$, $cDNA_{101}$, $cDNA_{98}$, and $cDNA_{92}$ products are as indicated. M denotes a ^{32}P-labeled 125-bp DNA fragment, added to each reaction upon termination, that serves as a control for the recovery and loading of reaction products. (E) Schematic representation of the initiation of reverse transcription. AA, initial dAMP residues incorporated onto the 3′ OH terminus of the $tRNA^{Trp}$ primer; closed arrow, $cDNA_{101}$ reverse transcription intermediate; open arrows, inverted repeat sequences; P, primer-binding site; L, leader region.

cleavage also occurs and results in oligodeoxynucleotides migrating faster than the starting substrate.

As shown in Fig. 6C, large integrative recombination products were detected in the presence of Mg^{2+} with the wild type (lane 8) but not with substrates containing the S2 mutation (lane 2). Such products, however, were detected with the S4 and S4C substrates (lanes 4 and 6), but in reduced amounts compared with the wild type. Densitometric tracing of the autoradiogram indicates that the amount of integrative recombinant product shown by the large arrowhead was reduced 42 and 63% for S4 and S4C, respectively. Similar results were obtained in the presence of Mn^{2+} (data not shown). Combining the in vitro results displayed in Fig. 6B and C, the S2 mutation is predicted to cause a greater integration defect than the S4 mutation because of its severely compromised nicking and joining reactions.

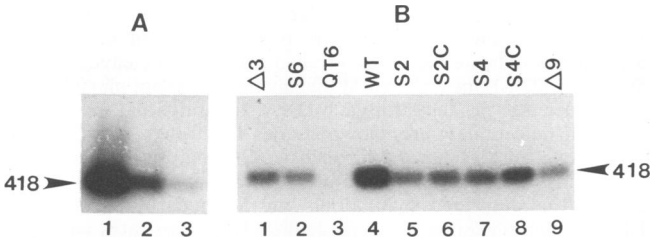

FIG. 5. Quantitation of integrated wild-type and U5 mutant provirus. (A) Amplification of a 418-bp fragment derived from the *pol* gene, as described in Materials and Methods, by using pRCAS-B DNA. Amplified DNA was analyzed on a 0.8% agarose gel and transferred to a nylon membrane, and viral sequences were detected by hybridization to randomly primed ^{32}P-labeled cDNA as described in Materials and Methods. The amounts of DNA amplified were as follows: lane 1, 1.0 pg; lane 2, 0.1 pg; and lane 3, 0.01 pg. (B) Chromosomal DNAs were isolated from cells 12 h after infection with wild-type or U5 mutant viruses and used as templates for amplification of the same 418-bp fragment. Equal amounts of amplified DNA were analyzed on a 0.8% agarose gel as described for panel A. The mutant viruses used for the initial infections of cells from which the DNA was isolated are indicated above the lanes. QT6 indicates DNA isolated from uninfected cells.

DISCUSSION

The experiments described above support the notion that U5-IR sequences have roles in both reverse transcription and integration. In viral RNA, U5-IR sequences form a secondary structure necessary for efficient initiation of reverse transcription, while in DNA these sequences are specifically recognized by IN and are required for integration. This represents an instance of a *cis*-acting element possessing unrelated replication functions in its RNA and DNA forms.

The U5-IR stem RNA secondary structure, like the U5-leader stem, is a primary determinant for efficiency of initiation of reverse transcription. There is a decrease in initiation in melittin-activated virions when the U5-IR structure is disrupted and an improved reverse transcription and growth efficiency when the U5-IR stem is restored by compensating mutations. We cannot rule out the possibility that the compensating mutations activated initiation independently of the inhibitory effect of the S2 or S4 substitution

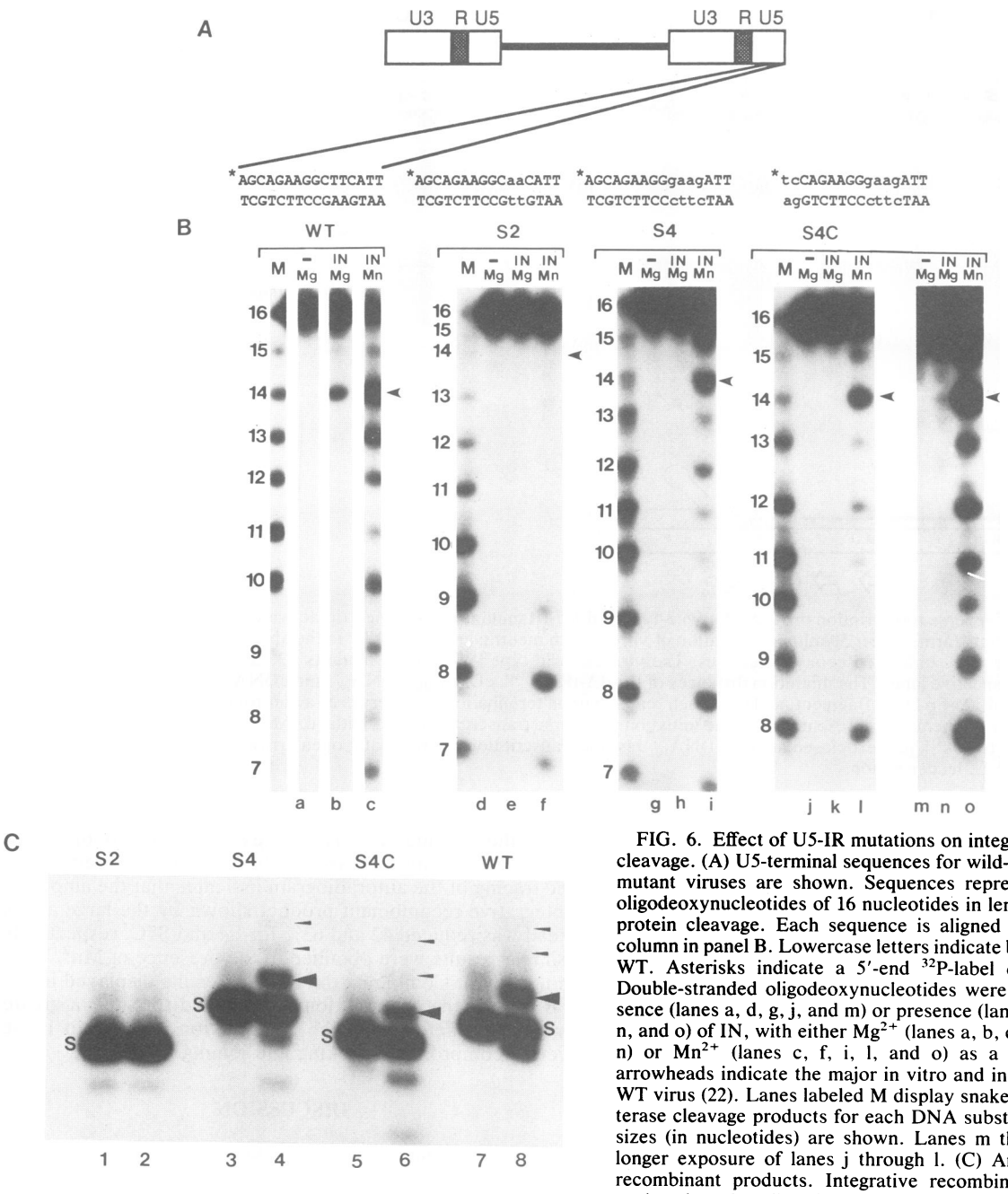

FIG. 6. Effect of U5-IR mutations on integration and IN protein cleavage. (A) U5-terminal sequences for wild-type (WT) and U5-IR mutant viruses are shown. Sequences represent double-stranded oligodeoxynucleotides of 16 nucleotides in length used to assay IN protein cleavage. Each sequence is aligned over the appropriate column in panel B. Lowercase letters indicate base changes from the WT. Asterisks indicate a 5′-end ^{32}P-label on plus strands. (B) Double-stranded oligodeoxynucleotides were incubated in the absence (lanes a, d, g, j, and m) or presence (lanes b, c, e, f, h, i, k, l, n, and o) of IN, with either Mg^{2+} (lanes a, b, d, e, g, h, i, k, m, and n) or Mn^{2+} (lanes c, f, i, l, and o) as a divalent cation. The arrowheads indicate the major in vitro and in vivo cleavage site of WT virus (22). Lanes labeled M display snake venom phosphodiesterase cleavage products for each DNA substrate, and the product sizes (in nucleotides) are shown. Lanes m through o represent a longer exposure of lanes j through l. (C) Analysis of integrative recombinant products. Integrative recombination products were analyzed as described in Materials and Methods by using 5′ ^{32}P-labeled duplex U5 oligodeoxynucleotide LTR substrates with two-base 5′ P overhangs representing the WT (lanes 7 and 8) and the S2 (lanes 1 and 2), S4 (lanes 3 and 4), and S4C (lanes 5 and 6) mutations. Double-stranded oligodeoxynucleotides were incubated in the absence (lanes 1, 3, 5, and 7) or presence (lanes 2, 4, 6, and 8) of IN. S denotes the respective starting substrates. DNA products larger than the starting substrates are indicated by the arrowheads.

mutation. This appears unlikely, however, since the Δ9 deletion disrupts the U5-IR stem structure without changing the bases involved in the S2 and S4 mutations and also results in an initiation defect. Additional evidence that supports a replication role for the U5-IR stem is the conservation of this structure in several retrovirus families (1).

Independent evidence supports a reverse transcription role for U5-IR sequences in murine leukemia virus (MuLV). Murphy and Goff (24) have shown that at least one mutation (*dl*Ava2) constructed in the comparable region of MuLV RNA causes a replication defect without affecting encapsidation of viral RNA. This mutation would disrupt the equivalent of the U5-IR stem. Another mutation, *dl*Ava3, which would also disrupt the U5-IR in MuLV, does not have any effect on replication. However, closer inspection of the sequence in *dl*Ava3 reveals that the potential to form an alternative but smaller U5-IR stem exists. Thus, MuLV stem structures may have similar roles to those demonstrated for avian sarcoma and leukosis virus.

In viral DNA, U5-IR sequences were found to be neces-

sary for efficient integration as well as for specific in vitro nicking at the viral integration site by purified IN protein (Fig. 5 and 6). These results were not unexpected, both because mutations near the U5 terminus of other retroviruses have been shown to cause integration defects (9, 10, 25, 26) and because mutations in this region were previously found to inhibit in vitro nicking by the avian retrovirus IN protein (22). Furthermore, Roth et al. (26) have shown that U5 mutations which block integration also block cleavage at the U5 terminus in vivo. The current study indicates that the U5-IR mutations which affect integration interfere with both IN-catalyzed cleavage and joining reactions in vitro, but to different extents. The S2 substitution appears to induce a stronger defect than the S4 substitution. In the case of the S2C virus, in which there is no defect in reverse transcription, we speculate that the ability of this virus to replicate and integrate in vivo (Fig. 3A and 5) might be due to cooperative interactions between the IN protein bound to the wild-type U3 LTR terminus and that bound to the mutated U5 LTR terminus. Alternatively, other viral or cellular proteins may contribute to the S2 integration efficiency in vivo, or integration mediated by oligomeric forms of viral DNA as described by Hagino-Yamagishi et al. (16) may occur.

We have found that the S2 mutation caused greater initiation and integration defects than the S4 mutation. Furthermore, S4C fully restored the reverse transcription efficiency, whereas S2C led to only a partial restoration. These findings were unexpected, since S2C has fewer nucleotide substitutions than S4C and forms a U5-IR stem of similar stability. Thus, the partial initiation defect of S2C suggests that both the U5-IR stem structure and sequence affect initiation. Apparently, the CTTC/GAAG U5-IR sequence is favored over GTTG/CAAC, which more closely resembles the wild-type sequence, GAAG/CTTC (Fig. 1C). We speculate that this could reflect the role of a protein that recognizes both the U5-IR stem structure and the GAAG sequence in either orientation in the stem. Although the same sequence preferences are displayed by the IN protein in in vitro cleavage and integration reactions, IN is unlikely to be involved in initiation since it can be mutated without affecting reverse transcription (29).

The biological consequences of the S2, S4, S2C, S4C, and Δ9 mutations can be explained on the basis of either disrupting the U5-IR stem structure, thereby affecting initiation of reverse transcription, or altering DNA sequences recognized by the IN protein for integration. The Δ3 mutation is more enigmatic. The sequences that are deleted by this mutation are not required by IN for specific cleavage of the U5 LTR terminus (22), nor do they appear to disturb the base pairing or sequence of the U5-IR stem. This suggests that features in addition to the U5-IR stem and U5-leader stem structures are required for maximal initiation efficiency. One intriguing possibility is that U5 sequences between the U5-leader and U5-IR stems, namely, those substituted in the S6 mutation (Fig. 1B), form base pairs with complementary sequences in the TψC loop of the $tRNA^{Trp}$ primer, as initially proposed by Haseltine et al. (17, 18) and supported by the experiments of Cordell et al. (11). This appears to be the case, since in vitro and in vivo studies indicate that this U5-primer interaction is important for initiation of reverse transcription (1). Thus, the Δ3 mutation might impair initiation of reverse transcription by altering the spacing between the U5-leader and U5-IR structures, thereby affecting the interactions between the TψC loop of the $tRNA^{Trp}$ primer and U5. This is discussed in more detail elsewhere (1).

Our previous analyses of a mutation that disrupts the U5-leader stem (Δ6 [Fig. 1B]) revealed a defect in the initiation of reverse transcription in infected cells and in reactions of melittin-permeabilized virions but not in reactions of purified viral RNA and reverse transcriptase (7). This raises the possibility that the U5-leader and possibly the U5-IR structures are important for signalling the initiation of reverse transcription following infection in the context of the highly structured virion core. The role of viral proteins in such signalling is unclear, although it may be relevant that initiation of poliovirus replication also is likely to require a RNA secondary structure near the initiation site. In this case, poliovirus genes whose products could interact with the secondary structure in *trans* have been identified by analysis of second-site revertants (2). We have isolated second-site revertants for several of the U5 mutations (5a), and a similar analysis might identify a retrovirus gene product that interacts with the U5-leader and U5-IR stem structures to regulate initiation of reverse transcription.

ACKNOWLEDGMENTS

We thank H.-J. Kung for critical comments on the manuscript.

This work was supported in part by Public Health Service grant CA 38046 from the National Cancer Institute and by Cancer Research Center grant P30 CA 43703. D.C. and H.H. are supported in part by Public Health Service training grant T32 GM 07250 from the National Institute of General Medical Sciences.

REFERENCES

1. **Aiyar, A., D. Cobrinik, Z. Ge, H.-J. Kung, and J. Leis.** Submitted for publication.
2. **Andino, R., G. E. Rieckhof, and D. Baltimore.** 1990. A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA. Cell **63**:369–380.
3. **Bilofosky, H. S., C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, C. D. Swindell, and C. S. Tung.** 1986. The GenBank genetic sequence data bank. Nucleic Acids Res. **14**:1–4.
4. **Boone, L. R., and A. M. Skalka.** 1981. Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with melittin. I. Kinetics of synthesis and size of minus- and plus-strand transcripts. J. Virol. **37**:109–116.
5. **Brown, B., B. Bowerman, H. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA *in vitro*. Cell **49**:347–356.

5a. **Cobrinik, D.** Unpublished data.

6. **Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis.** 1987. Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. J. Virol. **61**:1999–2008.
7. **Cobrinik, D., L. Soskey, and J. Leis.** 1988. A retroviral RNA secondary structure required for efficient initiation of reverse transcription. J. Virol. **62**:3622–3630.
8. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath, and B. Roizman (ed.), Virology, 2nd ed. Raven Press, Ltd., New York.
9. **Colicelli, J., and S. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell **42**:573–580.
10. **Colicelli, J., and S. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
11. **Cordell, B., R. Swanstrom, H. Goodman, and J. M. Bishop.** 1979. $tRNA^{Trp}$ as primer for RNA-directed DNA polymerase: structural determinants of function. J. Biol. Chem. **254**:1866–1874.
12. **Craigie, R., A.-T. Fujiwara, and F. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. Cell **64**:829–37.

13. **Dahlberg, J. E., R. C. Sawyer, J. M. Taylor, A. J. Faras, W. E. Levinson, H. M. Goodman, and J. M. Bishop.** 1974. Transcription of DNA from the 70*S* RNA of Rous sarcoma virus. I. Identification of a specific 4*S* RNA which serves as primer. J. Virol. **13**:1126–1133.
14. **Darlix, J.-L., M. Zuker, and P.-F. Spahr.** 1982. Structure-function relationship of Rous sarcoma virus leader RNA. Nucleic Acids Res. **10**:5183–5196.
15. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86**:3056–3069.
16. **Hagino-Yamagishi, K., L. A. Donehower, and H. E. Varmus.** 1987. Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. J. Virol. **61**:1964–1971.
17. **Haseltine, W. A., D. G. Kleid, A. Panet, E. Rothenberg, and D. Baltimore.** 1976. Ordered transcription of RNA tumor virus genomes. J. Mol. Biol. **106**:109–131.
18. **Haseltine, W. A., A. M. Maxam, and W. Gilbert.** 1977. Rous sarcoma virus genome is terminally redundant: the 5′ sequence. Proc. Natl. Acad. Sci. USA **74**:989–993.
19. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–371.
20. **Hughes, S. H., J. J. Greenhouse, C. J. Petropoulos, and P. Sutrave.** 1987. Adaptor plasmids simplify the insertion of foreign DNA into helper-independent retroviral vectors. J. Virol. **61**: 3004–3012.
21. **Katz, R., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. Cell **63**:87–95.
22. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63**:5319–5327.
23. **Mullis, K. B., and F. A. Faloona.** 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.
24. **Murphy, J. E., and S. Goff.** 1989. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. J. Virol. **63**:319–327.
25. **Panganiban, A., and H. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not for virus production. Nature (London) **306**:155–160.
26. **Roth, M., P. Schwartzberg, and S. Goff.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence of IN function and terminal DNA sequence. Cell **58**:47–54.
27. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **22**:853–869.
29. **Schwartzberg, P., J. Colicelli, and S. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
30. **Smith, R. E.** 1974. High specific infectivity avian RNA tumor viruses. Virology **60**:534–547.
31. **Temin, H. M., and D. Baltimore.** 1972. RNA-directed DNA synthesis and RNA tumor viruses. Adv. Virus Res. **17**:129–186.