JOURNAL OF VIROLOGY, Aug. 1980, p. 436-443 Vol. 35, No. 2
0022-538X/80/08-0436/08$02.00/0

Molecular Mechanism for the Capture and Excision of the Transforming Gene of Avian Sarcoma Virus as Suggested by Analysis of Recombinant Clones

TADASHI YAMAMOTO, JAYA SIVASWAMI TYAGI, JOHN B. FAGAN, GILBERT JAY, BENOIT DECROMBRUGGHE, AND IRA PASTAN*

Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Structural analysis of two cDNA clones, derived from reverse transcripts of avian sarcoma virus 21S mRNA's, reveals unusual features in the organization and expression of the integrated avian sarcoma virus (ASV) proviral DNA and predicts a mechanism for recombination events that will lead to either the capture or the excision of the transforming gene of this virus. The latter is supported by our observation that there is an extensive homologous region on either side of the transforming gene that will allow site-specific deletion or integration to occur. Comparison of the clone derived from the *src*-specific 21S mRNA coding for the transforming gene product to that derived from the *env*-specific 21S mRNA coding for the envelope glycoprotein show that the common *c* region present at the 3′ terminus of the ASV genome is 326 bases long. Within this *c* region are nucleotide sequences that may play key roles in the life cycle of this virus. These regulatory sequences include (i) probable promoter sites for the initiation of transcription, (ii) a polyadenylation signal, and (iii) a sequence that is complementary to the 3′ termini of both the *env* and the *src* regions, which will allow the generation of transformation-defective deletions.

The RNA tumor viruses have a complex life cycle that involves synthesis of DNA from an RNA template, integration of that DNA into the host chromosome, and transcription and processing of viral mRNA's (4). To elucidate the biochemical basis of these processes it is of use to have recombinant clones containing various regions of the viral genome. Recently we have described the isolation and partial characterization of a cDNA clone (pSR1) derived from the avian sarcoma virus (ASV) 21S mRNA that contains the common *c* region of ASV and extra sequences at its 3′ end not previously found in genomic RNA (32). We believe these sequences arise by transcription of the integrated provirus, which is believed to contain terminal redundancies that are approximately 300 base pairs (bp) long (24, 12). These redundancies include the 20-bp repeats found in genomic RNA. Previously, part of the sequence of the cDNA cloned in pSR1 has been determined, and within this sequence, promoter-like regions and a polyadenylic acid [poly(A)] addition signal were identified (32).

Here we report the complete characterization of the recombinant plasmid pSR1 as well as characterization of another recombinant plasmid, pSR2, also derived from ASV 21S mRNA. The cDNA insert in pSR1 contains *src* sequences at its 5′ end, indicating that it is derived from *src-c* mRNA. In contrast, the cDNA insert in pSR2 contains *env* and *c* sequences, indicating that it is probably derived from *env-c* mRNA. Examination of the nucleotide sequence of the two inserts indicates that there are homologous spacer regions at the end of the *env* and *src* genes. These sequences may be the basis for the generation of transformation-defective viruses.

MATERIALS AND METHODS

Cells and viruses. Primary cultures of fibroblasts from chicken embryos (Truslow Farms, Chestertown, Md.) were grown in GM medium (27) at 39°C and were infected with frozen stocks of recently cloned viruses from the following sources: Schmidt-Ruppin strain of ASV, subgroup D (SR-ASV-D), and Prague strain of ASV, subgroup C (PR-ASV-C), were from the American Type Culture Collection; *td* SR-ASV-D mutants clones 14 and 22 were from P. K. Vogt. CEF transformed with the *env*-deficient Bryan strain of ASV clone N-2 (9) were prepared by H. Hanafusa using Sendai virus to enhance virus adsorption.

Recombinant plasmids. Two recombinant plasmids, pSR1 and pSR2, which were constructed and cloned in *Escherichia coli* C600 (32) were used. The parental plasmid was pBR322. The plasmid DNA was prepared from cells grown in super broth (32a) containing 25 μg of tetracycline per ml after amplification with chloramphenicol (250 μg/ml) in a P3 physical

containment facility.

Preparation of insert DNA and digestion of DNA with restriction enzymes. Purification of the DNA insert was achieved by digestion with *Pst*I followed by centrifugation on a 5 to 20% (wt/vol) sucrose gradient in 10 mM Tris-hydrochloride (pH 7.5), 5 mM NaCl, and 1 mM Na_2 EDTA. Digestions with restriction endonucleases (from New England Biolabs) were essentially as suggested by the manufacturer. The DNA fragments were analyzed by electrophoresis on a 7% polyacrylamide gel (19). The gel was either autoradiographed or stained with ethidium bromide.

Preparation and analysis of RNA. Total cellular RNA was extracted from both normal and ASV-transformed CEF by the guanidine-hydrochloride procedure (25). A 15-μg amount of total RNA was subjected to electrophoresis in a 1% agarose slab gel containing 6 mM methyl mercuric hydroxide (3). RNAs in the gel were transferred to diazobenzyloxymethyl (DBM) paper (Schleicher & Schuell Co., Keene, N.H.) as described by Alwine et al. (2).

Hybridization of cloned DNA fragments to DBM paper containing viral RNA. DBM filters on which RNAs were covalently bound were pretreated overnight and then hybridized with [^{32}P]DNA probes (1×10^5 to 5×10^5 cpm for 18 h) (1). After hybridization at 37°C, filters were washed four times for 15 min each with a solution of 1 × SSC (0.15 M NaCl plus 0.015 M sodium citrate), 0.2% SDS, and 2 mM Na_2 EDTA at 45°C and then autoradiographed at −70°C with intensifying screens (Kodak) with Kodak XR-2 film.

DNA sequencing and computer analysis. The DNA sequence of both inserts was determined as described by Maxam and Gilbert (18). Four base-specific cleavages were used (guanosine [G], adenosine plus guanosine [A + G], cytosine plus thymine [C + T] and cytosine [C]). The strategy of the sequencing is shown in Fig. 4B. A computer program (C. Queen and L. J. Korn, Methods Enzymol., in press) was used to analyze the nucleotide sequence of the pSR1 insert. Repetititions, dyad symmetries, and termination codons were detected by this program.

RESULTS AND DISCUSSION

Isolation and restriction analysis of cDNA clones. By using 21S mRNA from ASV-transformed CEF as template we constructed and isolated two recombinant plasmids, pSR1 and pSR2, containing ASV-specific sequences (32). Plasmid pSR1 has been previously studied and shown to contain DNA sequences corresponding to the *c* region of the ASV genome. In addition, there are extra sequences at the 3′ end of the insert which are not found contiguously within genomic RNA. We have now performed a more detailed restriction analysis of pSR1 (Fig. 1a). We have identified a *Pst*I site close to the 5′ end of the insert that was not previously observed because the small fragment generated ran at the front of the gel.

The restriction map of the cloned insert in pSR2 is shown in Fig. 1b and is based on analysis with restriction enzymes and the DNA sequence (see below). The length of the cloned insert was about 550 bp. It is evident that the 3′ end of the pSR2 insert contains *Hin*fI and *Pvu*I sites in the same relationship as in pSR1. However, the restriction map indicates that the 5′ end of the pSR2 insert is very different from that of pSR1.

Hybridization analysis with the complete pSR2 insert. Previously the pSR1 insert was shown to be derived from the *c* region of ASV and to hybridize with 39S, 35S, 28S, and 21S viral RNA species found in ASV-infected cells (32). We performed a similar analysis with pSR2. The cloned insert was excised, labeled with [α-

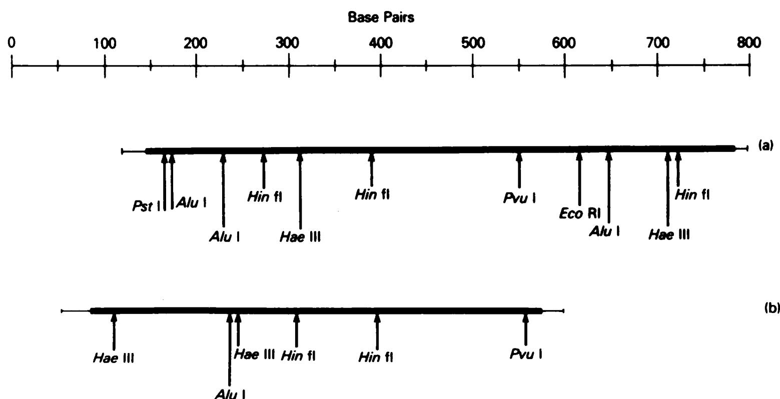

FIG. 1. *Restriction maps of the insert DNAs of pSR1 and pSR2. (a) Restriction map of the pSR1 insert, which was previously reported (32), has been extended. The digestion of pSR1 with PstI generated two DNA fragments of cDNA origin. The smaller fragment (50 bp) was located at the 5′ end of the larger fragment, judging from the position of the G-C tail in each fragment. Also, an AluI recognition site was found very close to the extreme 5′ end of the larger fragment. (b) Restriction map of the pSR2 insert. The boxed region indicates the nucleotide sequence corresponding to the viral RNA sequence and the line at both ends indicates the G-C tail.*

^{32}P]deoxyadenosine triphosphates and [α^{32}P]-deoxycytosine triphosphates by nick translation (17) and hybridized against filters containing RNA from cells infected with either nondefective (*nd*) SR-ASV-D, *nd* PR-ASV-C or transformation-defective (*td*) PR-ASV-C. The insert hybridized with all species of viral mRNA in *nd* SR-ASV-D- and *nd* PR-ASV-C-infected cells and with 35S and 21S RNA in *td* PR-ASV-C-infected cells (Fig. 2). Since restriction analysis showed that pSR2 contained some of the same sequences as pSR1 (Fig. 1b), we tentatively concluded that pSR2 also contained sequences from the *c* region.

Nucleotide sequence of pSR1 and pSR2 inserts. The entire DNA sequence of the pSR1 insert and the sequence of almost all of the pSR2 insert have been determined and are shown in Fig. 3A. The sequence of about 50 nucleotides at each end of the pSR2 insert has not been determined; more than half of these terminal sequences are made up of deoxycytosine:deoxyguanosine residues added during the cloning procedure (32).

As predicted by the restriction analysis, the nucleotide sequence of the cDNA inserts in pSR1 and pSR2 are identical over 199 bp. The identical sequence is located from position +1 to position +199 in the DNA insert of both pSR1 and pSR2. There is a single difference in position +138 which could be due to an error introduced either by reverse transcriptase or by DNA polymerase during the synthesis of the double-stranded cDNA. We assume that the insert of pSR2 is lacking sequences at its 3′ end that are found in pSR1 due to incomplete second-strand synthesis during preparation of the cDNA.

Hybridization analysis with restriction fragments from pSR1 and pSR2. The DNA inserts of recombinant plasmids pSR1 and pSR2 were derived from 21S viral mRNA that is composed of both *src-c* and *env-c* mRNA's (11, 31). The latter is derived from viruses that contain deletions in the *src* region; these deletions are probably generated soon after infection with *nd* SR-ASV-D, and such mutants are frequently found in stocks of ASV. Since the two plasmids have different restriction sites at their 5′ ends, we assumed that one was derived from *src-c* mRNA and the other was from *env-c* mRNA. To test this hypothesis we isolated a restriction fragment from the 5′ end of each plasmid, labeled the fragment with ^{32}P by polynucleotide kinase (18), and hybridized each to RNA from (i) uninfected CEF, (ii) CEF infected with *nd* SR-ASV-D, (iii) CEF infected with *td* SR-ASV-D containing either a small (1.47-kilobase [kb]: clone 14) or a large (1.87-kb: clone 22) deletion in the *src* gene, and (iv) CEF infected with Bryan strain of ASV containing a deletion of the *env* gene.

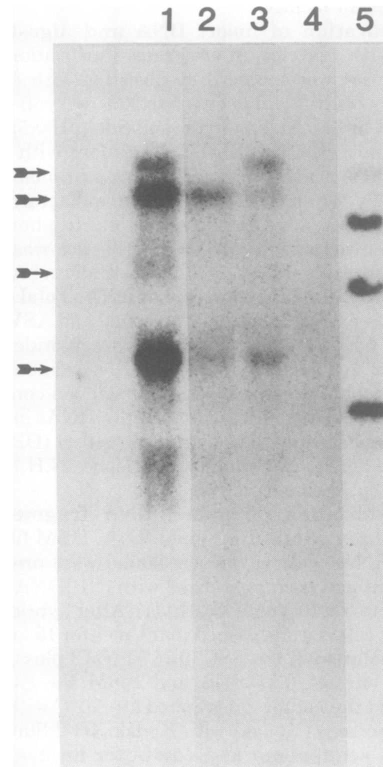

FIG. 2. *Hybridization of the pSR2 insert with ASV-specific RNA. Total cellular RNA (10 μg) extracted from CEF infected with various strains of ASV was electrophoresed on a 1% agarose-methyl mercuric hydroxide slab gel, and the fractionated RNA was subsequently transferred to a DBM paper as described in Materials and Methods. The entire 550-bp pSR2 DNA insert was labeled with ^{32}P by nick translation (specific activity, 5 × 10^7 cpm/μg), as described by Maniatis et al. (17), and was used as a probe for hybridization. The RNA used was prepared either from CEF infected with nd SR-ASV-D (lane 1), td PR-ASV-C (lane 2), or nd PR-ASV-C (lane 3) or from uninfected CEF (lane 4). HindIII restriction fragments of λ DNA were included as markers (lane 5); the molecular weights (× 10^6) of the fragments are: 15, 6.4, 4.3, 2.9, 1.6, and 1.4. The arrows indicate 39S, 35S, 28S, and 21S RNA.*

As shown in Fig. 4A, the 55-bp *Alu*I-*Alu*I fragment located between positions −183 and −129 at the 5′ end of the pSR1 insert hybridized with 39S RNA from *nd* SR-ASV-D, 37S RNA from a mutant containing a small *src* deletion, and 35S RNA from the Bryan strain of ASV mutant lacking *env*-specific sequences. It did not, however, hybridize to 35S RNA from a mutant containing a large *src* deletion. This result places the fragment within the *src* sequences, and indicated that the 5′ end of the pSR1 insert contains *src*-specific sequences that

are lost in the large *src* deletion (16). Consistent with this conclusion is the identification of a large RNase T_1-resistant oligonucleotide derived from the 3′ end of the *src* gene (21). This oligonucleotide has the composition (U_2, C_2, AAG, $A_{4-5}N$). This is almost identical to the composition of the sequence located between positions −139 and −129 in the pSR1 insert which is (U_2, C, AAG, AAAAU). That the *src-c* sequences in the pSR1 insert are contiguous as in viral RNA was shown both by R-loop analyses (32) and by S1 protection experiments (data not shown) where the pSR1 insert was shown to be colinear with viral RNA.

When the 180-bp *AluI* fragment located between position −131 and the extreme 5′ end of the pSR2 insert was examined, quite different results were obtained (Fig. 4B). The *AluI* fragment failed to hybridize with RNA from cells infected with the *env* deletion strain of the Bryan strain of ASV but did hybridize with RNA from cells infected with ASV containing either a large or a small *src* deletion as well as with *nd* SR-ASV-D. This result indicates that the 5′ end of the pSR2 insert has *env*-specific sequences. This is further supported by the observation that the sequence between positions −147 and −131 in the pSR2 insert has a composition of (U_2, C_3, AU_2, AG, AAC, AAU) which is almost identical to the RNase T_1-resistant oligonucleotide that is derived from the 3′ end of the *env* gene (U_3, C_4, AU_2, AG, AAC, AAU) (29).

To demonstrate that the pSR2 insert is colinear to *td* ASV RNA we carried out an R-loop analysis between the insert and a 39S and 35S viral RNA mixture. Most of the R-loop structures observed were about 550 bases long (data not shown).

Thus, pSR1 contains *src* and *c* sequences and pSR2 *env* and *c* sequences. Although these hybridization studies define the limits of the *src* and *env* genes, they do not provide any information on the location of the coding sequence within each of the two genes. The first termination codon closest to the 5′ end of the inserts, in each of the three reading frames, is found at positions −154, −150, and −110 for pSR1 and at positions −242, −163, and −96 for pSR2. This implies that if the coding region of the *src* or *env* gene does extend into the cloned insert, the polypeptide chain must terminate at one of these three termination codons and that there must be extensive noncoding sequences at the 3′ ends of both mRNA's.

Common region. The *c* region of ASV and avian leukosis virus (ALV) is the common region found at the 3′ end of both viral genomes (28). It terminates with a 20-bp sequence that is also found at the 5′ end of genomic RNA (10, 23). By heteroduplex mapping between *nd* and *td* ASV the *c* region has been determined to be about 400 to 600 bases long excluding the poly(A) tail (13, 16). If we assume that the insert in pSR2 is derived from mRNA of a defective virus in which the *src* gene has been deleted, then the region in which the sequences of pSR1 and pSR2 are identical should define the *c* region; this starts from position +1 in both pSR1 and pSR2. From position +1 in pSR1 to the end of the 20-bp repeat at position +326 is 326 bases. This value is close to the 400 to 600 bp estimated by heteroduplex mapping.

It has been previously suggested that the *c* region might code for a viral protein (21). Inspection of the nucleotide sequence of the *c* region reveals termination codons (TAA, TAG, and TGA) in every reading frame. The longest open frame starts at position +132 and ends at +344. This termination point lies beyond the 20-bp repeat that is found at the extreme 3′ terminus of genomic RNA and is within that region of the pSR1 insert that represents redundant sequences derived from the 5′ terminus of the viral genomic RNA. This open reading frame does not initiate with an AUG, but there is a GUG at position +177 which could initiate a polypeptide of 56 amino acids. A second open frame starts with an AUG at position +167 and ends at the UGA at +251. This sequence could code for a polypeptide of 28 amino acids.

Spacer region. It is important to note that the termination codon for *src* and that for *env* should be separated by at least 110 bases and 96 bases from the beginning of the *c* region in pSR1 and pSR2, respectively. Two important features characterize these sequences. First, the nucleotide sequences of these two regions are not identical but contain two clusters of very similar sequences. These clusters are identified by circles in Fig. 3A. Second, part of both these "spacer" regions can base-pair with elements at the 5′ end of the *c* region between positions +1 and +45 to form the stem-and-loop structures shown in Fig. 5. The arrow indicates the junction of the *c* region with the spacer region at the end of the *src* gene in pSR1 and the junction of the *c* region with the spacer region of the *env* gene in pSR2.

We suggest that pSR2 is derived from the mRNA of a transformation-defective variant where the *src* gene has been deleted. If so, it can be concluded that the spacer region lying to the 3′ side of the *env* gene as found in pSR2 is also located 5′ proximal to the *src* gene in the ASV genome and that the *src* gene is, therefore, bounded on both sides by similar spacer regions. Determination of the sequence of ASV genomic RNA will be required to confirm this suggestion.

A

(a)

-200

5′ (24 G tails)CACCCCTCGAGCTACCTGCAGGTTCAGCTG

-150 -100

CTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTG

-50 -1┐ ┌+1 +50

CTTCGCGATGTACGGGCCAGATATACGCGTATCTGAGGGGACTAGGGTGTGTTTAGGCGAAAAGCGGGGCTTCGGTTGTACGCGGTTAGGAGTCCCCTCAGGATATAGTAGTTTCGCTTT

+100 +150

TGCATAGGGAGGGGGAAATGTAGTCTTATGCAATACACTTGTAGTCTTGCAACATGGTAACGATGAGTTAGCAACATGCCTTACAAGGAGAGAAAAAGCACCGTGCATGCCGATTGGTGG

+200 +250 +300

AAGTAAGGTGGTACGATCGTGCCTTATTAGGAAGGCAACAGACAGGTCTGACATGGATTGGACGAACCACTGAATTCCGCATTGCAGAGATAATTGTATTTAAGTGCCTAGCTCGATACA

+350 +400

ATAAAC[GCCATTTGACCATTCACCAC]ATTGGTGTGCACCTGGGTTGATGGCCGGACCGTTGATTCCCTGACGACTACGAGCACCTGCATGAAGCAGAAGGCTTCAGGGAGTGCGTGCCAT

CTA(16 C tails) 3′

(b)

-250 -200

5′ (G tails)···· 20 bases ····GGCCGTTCATTTGCTGAAAGGACTGCTTTTGGGGCTTGTAGTTATTTTGTTGCTAGTAGTGTGCCTGCCTTGCCTT

-150 -100

TTGCAAATCGTATGCGGTAACATCAGAAAGATGATTAATAACTCCATCAGCTGCAAAAGGCCTGTGGCAGCCTGAAAGCAGAATAGTATAAGGCAGTACATGGGTGGTGGTATAGCGCT

-50 -1┐ ┌+1 +50

TGCGAGTCGGGTTGTAACGGGGCATGGCTTAACTAAGGGGACTATGGCATGTATAGGCGCAAAGCGGGGCTTCGGTTGTACGCGGTTAGGAGTCCCCTCAGGATATAGTAGTTTCGCTTT

+100 +150

TGCATAGGGAGGGGGAAATGTAGTCTTATGCAATACACTTGTAGTCTTGCAACATGGTAACGATGAGTTAGCAACATACCTTACAAGGAGAGAAAAAGCACCGTGCATGCCGATTGGTGG

AAGTAAGGTGGTACGATCG ··· 20 bases ··· (C tails) 3′

FIG. 3. *Nucleotide sequence of the common region and the neighboring regions of ASV. (A) The nucleotide sequence shown is the plus DNA strand of ASV. The extreme 5′ nucleotide of the region that is identical in the pSR1 and pSR2 inserts was defined as position +1. The triplets underlined with either dotted, broken, or solid lines represent termination codons in three different reading frames. The brackets in the pSR1 insert indicate the sequence that is terminally redundant in the viral genome. The circles indicate the homologous regions between pSR1 and pSR2. (a) Nucleotide sequence of the pSR1 insert. (b) Nucleotide sequence of the pSR2 insert. (B) Strategy employed for sequencing DNA of (a) the pSR1 insert and (b) the pSR2 insert. Although the nucleotide sequence of the 3′ region of the pSR1 insert has been reported elsewhere, a part of this portion was confirmed again. The DNA sequence was determined as described in Materials and Methods. After the kinase labeling at the 5′ end, the restriction fragments were either cleaved by another enzyme or subjected to strand separation before the chemical reactions were carried out. The boxed and open-boxed regions indicate where the nucleotide sequence has and has not been determined, respectively. The arrows indicate the region sequenced and the direction in which the sequence was read from a polyacrylamide gel.*

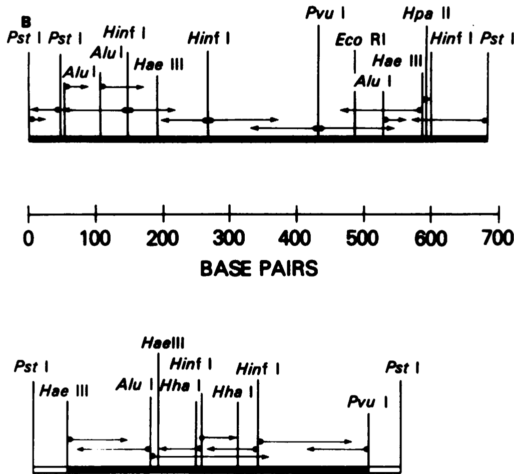

FIG. 3B

We recognize that other possibilities exist for the origin of the RNA that gave rise to pSR2.

Inspection of the nucleotide sequence of the spacer region reveals only short sequences with open reading frames. The spacer region in pSR1 contains an AUG codon at position −53 which is followed in the same reading frame by a termination codon at position +46 in the *c* region. Similarly, in pSR2, there is an AUG in the spacer region at position −80 which is followed in the same frame by a termination codon in the *c* region at position +65. However, we do not think the putative polypeptides, which would be coded by these sequences, exist, since the above open frames can derive from either *src-c* mRNA or *env-c* mRNA, indicating that they could not appear stoichiometrically in both wild-type and defective viruses in comparison to other viral proteins.

Deletion model. An unexplained observation with the ASV-ALV system is the generation of transformation-defective variants in culture through the deletion of the *src* gene from the ASV genome (6), and the reciprocal event involving the generation of transforming phenotypes (rASV) in vivo through the acquisition of the cellular *src* gene by ALV or *td* ASV (7, 8, 26, 29, 30). These integration-excision events apparently occur at high frequencies.

By reconstruction from the nucleotide sequences of pSR1 and pSR2, we propose that the *src* gene is flanked by two spacer regions of similar but not identical sequence each about 80 bases long. This repeat in sequence could be recognized by an enzyme system that is able to excise the *src* gene from the ASV genome and ligate the *c* region to the *env* gene, thereby generating *src* deletions. We believe that viability of the transformation-defective viruses which result from such deletions is determined by the conservation of a hairpin structure as in Fig. 5. Whereas such a site-specific recombination event can lead to the excision of the *src* gene by transforming viruses, the reciprocal event involving site-specific integration could explain the acquisition of the *src* gene by nontransforming variants.

Since the *c* region has multiple termination codons in all reading frames, it is unlikely that it codes for a specific protein. It is more likely that the *c* region serves various regulatory functions (Fig. 6). We have already described a possible promoter for transcription by RNA polymerase II between positions +133 and +158 (32). The A-T rich sequence, ATAATTGTATTTAAG, located between positions +270 and +284 could also be a promoter. Although we do not know yet which precise sequence could serve as a promoter, the sequence TATTTAAG is similar to the sequence (T)ATAAAAG, which

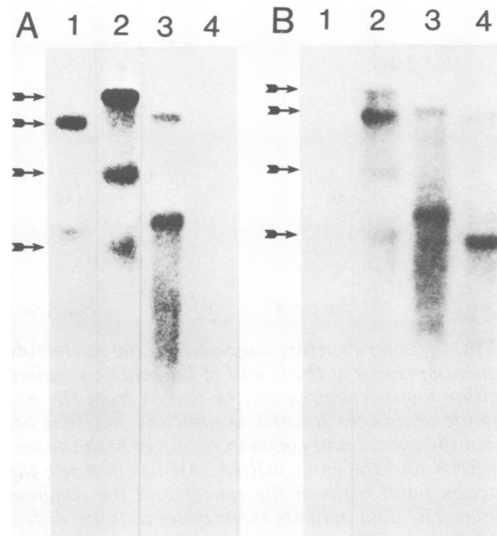

FIG. 4. Hybridization of the 5' DNA fragments from both pSR1 and pSR2 inserts with ASV-specific RNA. Total cellular RNA (15 μg), extracted from CEF infected with either the env-deficient Bryan strain of ASV (lane 1), nd SR-ASV-D (lane 2), td SR-ASV-D clone 14 (lane 3) or clone 22 (lane 4), was fractionated by electrophoresis on a 1% agarose-methyl mercuric hydroxide slab gel. The RNA was transferred to DBM paper as described in Materials and Methods. The DNA probe used for the hybridization was either the 55-bp AluI-AluI fragment from the 5' end of the pSR1 insert (A) or the 180-bp AluI fragment from the 5' end of the pSR2 insert (B). The probes were obtained by labeling the appropriate DNA fragment with ^{32}P by polynucleotide kinase (specific activity, 1.5 × 10^7 cpm/μg). Arrows indicate 39S, 35S, 28S, and 21S RNA.

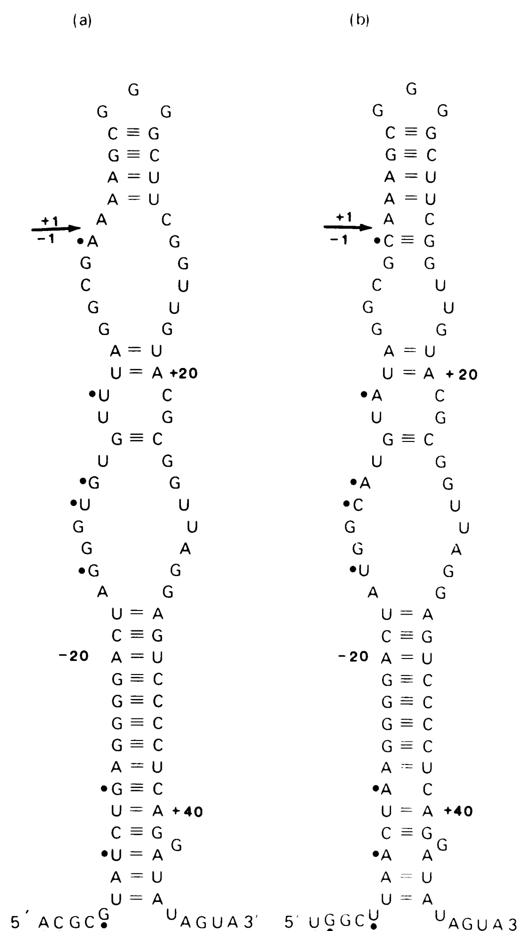

FIG. 5. *Loop structure suggested by the nucleotide sequence present at the 5′ end of the common region. Similar loop structures can be drawn from the nucleotide sequences located around the junction between the spacer and common region of both the src-c mRNA and the env-c mRNA. Arrows indicate the junction point between the spacer and the common region. The dots indicate those bases that are different between pSR1 and pSR2. (a) Loop structure from src-c mRNA. (b) Loop structure from env-c mRNA.*

may be a promoter for rabbit β globin (15) and for a major late transcription unit of adenovirus (33), and to the sequence TATTTATG, which may be a promoter for an early gene of SV40 (22). In addition, there is a poly(A) addition signal (20), AATAAA, at position +300. The 5′ end of the *c* region might participate in excision of the *src* region, resulting in the generation of transformation-defective deletions. The 20-bp redundant sequence located at the extreme 3′ part of the *c* region is essential for the synthesis of minus strand DNA (5, 10, 23). Another possible function for the *c* region could be in the

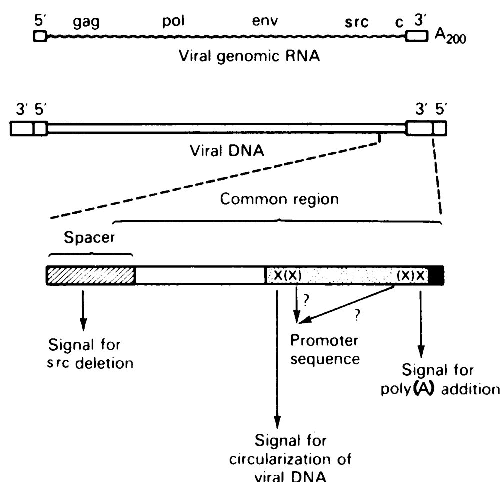

FIG. 6. *Schematic drawing of the regulatory functions of the common region. The dotted box in the bottom line represents the 200-bp sequence found at either end of the integrated proviral DNA, and the closed box represents the 20 bases of redundant sequence found at the end of the viral RNA genome.*

circularization of viral DNA. Since the 100-bp repeating sequence derived from the 5′ end of the viral genome is joined to the 200-bp repeating sequence from the *c* region (12), the nearly perfect direct repeat, AGCACCGCATG starting at position +157 and AGCACCTGCATG starting at +379, could be a signal for circularization.

ACKNOWLEDGMENTS

We thank H. Hanafusa for his gift of CEF infected with the *env*-deficient Bryan strain of ASV and B. Lovelace for her expert tissue culture assistance.

LITERATURE CITED

1. **Adams, S. L., J. C. Alwine, B. de Crombrugghe, and I. Pastan.** 1979. Identification of two collagen RNAs larger than pro-α1 and pro-α2 collagen messenger RNAs which are decreased in transformed fibroblasts. J. Biol. Chem. **254**:4935–4938.
2. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
3. **Bailey, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **67**:247–256.
4. **Bishop, J. M.** 1978. Retroviruses. Annu. Rev. Biochem. **47**:35–88.
5. **Collett, M. S., and A. J. Faras.** 1976. Evidence for circularization of the avian oncornavirus RNA genome during proviral DNA synthesis from studies of reverse transcription *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **73**: 1329–1332.
6. **Duesberg, P. H., and P. K. Vogt.** 1973. RNA species obtained from clonal lines of avian sarcoma and from avian leukosis virus. Virology **54**:207–219.

7. **Halpern, C. C., W. S. Hayward, and H. Hanafusa.** 1979. Characterization of some isolates of newly recovered avian sarcoma virus. J. Virol. **29**:91-101.
8. **Hanafusa, H., C. C. Halpern, D. L. Buchhagen, and S. Kawai.** 1977. Recovery of avian sarcoma virus from tumors induced by transformation defective mutants. J. Exp. Med. **146**:1735-1747.
9. **Hanafusa, T., T. Miyamoto, and H. Hanafusa.** 1970. A type of chick embryo cell that fails to support formation of infectious RSV. Virology **40**:55-64.
10. **Haseltine, W. A., A. M. Maxam, and W. Gilbert.** 1977. Rous sarcoma virus genome is terminally redundant: the 5' sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:989-993.
11. **Hayward, W. S.** 1977. Size and genetic content of viral RNAs in avian oncovirus-infected cells. J. Virol. **24**:47-63.
12. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma virus are terminally redundant, coextensive with unintegrated linear DNA and integrated at many sites. Cell **15**:1397-1410.
13. **Junghans, R. P., S. Hu, C. A. Knight, and N. Davidson.** 1977. Heteroduplex analysis of avian RNA tumor viruses. Proc. Natl. Acad. Sci. U.S.A. **74**:477-481.
14. **Kawai, S., P. H. Duesberg, and H. Hanafusa.** 1977. Transformation-defective mutants of Rous sarcoma virus with *src* gene deletions of varying length. J. Virol. **24**:910-914.
15. **Konkel, D., S. Tilghman, and P. Leder.** 1978. The sequence of the chromosomal mouse β-globin major gene: homologies in capping, splicing and poly(A) sites. Cell **15**:1125-1132.
16. **Lai, M. M. C., S. S. F. Hu, and P. K. Vogt.** 1977. Occurrence of partial deletion and substitution of the *src* gene in the RNA genome of avian sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **74**:4781-4785.
17. **Maniatis, T., A. Jeffrey, and D. Klein.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. U.S.A. **72**:1184-1188.
18. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560-564.
19. **Peacock, A. C., and C. W. Dingman.** 1968. Molecular weight estimation and separation of ribonucleic acid by electrophoresis in agarose-acrylamide composite gels. Biochemistry **7**:668-674.
20. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3' Noncoding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211-214.
21. **Purchio, A. F., E. Erikson, and R. L. Erikson.** 1977. Translation of 35S and of subgenomic regions of avian sarcoma virus RNA. Proc. Natl. Acad. Sci. U.S.A. **74**:4661-4665.
22. **Reddy, V. B., R. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman.** 1978. The genome of simian virus 40. Science **200**:494-502.
23. **Schwartz, D. E., P. C. Zamecnik, and H. L. Weith.** 1977. Rous sarcoma virus genome is terminally redundant: the 3' sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:994-998.
24. **Shank, P. R., S. H. Hughes, H.-J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated forms of avian sarcoma virus DNA: both termini of linear DNA bear a 300 nucleotide sequence present once or twice in two species of cellular DNA. Cell **15**:1383-1395.
25. **Strohman, R. C., P. S. Moss, J. Micou-Eastwood, D. Spector, A. Przybyla, and B. Paterson.** 1977. Messenger RNA for Myosin polypeptides: isolation from single myogenic cell cultures. Cell **10**:265-273.
26. **Vigne, R., M. L. Breitman, C. Moscovici, and P. K. Vogt.** 1979. Restitution of fibroblast-transforming ability in src deletion mutants of avian sarcoma virus during animal passage. Virology **93**:413-426.
27. **Vogt, P. K.** 1969. Focus assay of Rous sarcoma virus, p. 198-211. *In* K. Hobel and N. Salzman (ed.), Fundamental techniques in virology. Academic Press, Inc., New York.
28. **Wang, L.-H.** 1978. The gene order of avian RNA tumor viruses derived from biochemical analyses of deletion mutants and viral recombinants. Annu. Rev. Microbiol. **32**:561-592.
29. **Wang, L. H., C. C. Halpern, M. Nadel, and H. Hanafusa.** 1978. Recombination between viral and cellular sequences generates transforming sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **75**:5812-5816.
30. **Weiss, R., L. Crittenden, H. Purchase, and P. K. Vogt.** 1975. Genetic control and oncogenicity of endogenous and exogenous avian RNA tumor viruses, vol. 2, p. 248-253. *In* P. Bucalossi, W. Veronesi, and N. Cascinelli (ed.), Proceedings of the XI International Cancer Congress, Florence, 1974. American Elsevier, New York.
31. **Weiss, S. R., H. E. Varmus, and J. M. Bishop.** 1977. The size and genetic composition of virus-specific RNAs in the cytoplasm of cells producing avian sarcoma-leukosis viruses. Cell **12**:983-992.
32. **Yamamoto, T., G. Jay, and I. Pastan.** 1980a. Unusual features in the nucleotide sequence of a cDNA clone derived from the constant region of avian sarcoma virus messenger RNA. Proc. Natl. Acad. Sci. U.S.A. **77**:176-180.

32a. **Yamamoto. T., M. E. Sobel, S. L. Adams, V. E. Avredimento, R. DiLauro, I. Pastan, B. deCrombrugghe, A. Showalter, D. Pesciotta, P. Feitzek, and B. Olsen.** 1980. Construction of a recombinant bacterial plasmid containing pro-α1 (I) collagen DNA sequences. J. Biol. Chem. **255**:2612-2615.

33. **Ziff, E. B., and R. M. Evans.** 1978. Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit. Cell **15**:1463-1475.