JOURNAL OF VIROLOGY, Feb. 1982, p. 535–541
0022-538X/82/020535-07$02.00/0

Vol. 41, No. 2

Nucleotide Sequence of the 5′ Noncoding Region and Part of the *gag* Gene of Rous Sarcoma Virus

RONALD SWANSTROM,* HAROLD E. VARMUS, AND J. MICHAEL BISHOP

Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Received 3 August 1981/Accepted 28 September 1981

Several functions of the retrovirus genome involve structural features in the vicinity of its 5′ terminus. In an effort to further elucidate the relationship between structure and function in retrovirus RNA, we have determined the sequence of the first 1,010 nucleotides at the 5′ end of the genome of Rous sarcoma virus by using the Maxam-Gilbert method to sequence suitable domains in cloned Rous sarcoma virus DNA. The results (i) locate the initiation codon for the *gag* gene of Rous sarcoma virus 372 nucleotides from the 5′ end of viral RNA; (ii) demonstrate that this codon is preceded by three methionine codons that are apparently not used in translation; (iii) sustain previous conclusions that the principal site to which ribosomes bind on the Rous sarcoma virus genome in vitro does not contain the initiation codon for *gag*; (iv) permit deduction of the amino acid sequence of a viral structural protein, p19; (v) confirm the amino-terminal sequence of $Pr76^{gag}$; and (vi) substantiate the identification of a splice donor site described in the accompanying manuscript (Hackett et al., J. Virol., 41:527–534, 1982).

The genomes of retroviruses assume multiple roles during the viral life cycle, serving as template for the synthesis of viral DNA by reverse transcriptase, messenger for the synthesis of several viral gene products, precursor for the genesis of subgenomic mRNA's, and vehicle for the transmission of viral genes from one host cell to another (1). Each of these roles involves structural features located in the vicinity of the 5′ terminus of the viral genome: the initiation site for reverse transcription (41); a nucleotide sequence (R) repeated at the 5′ and 3′ termini of viral RNA that is required for chain propagation by reverse transcriptase (3a); a ribosome binding site and AUG codon required to initiate translation (29, 45); a splice donor site used in the genesis of subgenomic mRNA's (5, 8, 9, 11, 25, 28, 30, 38, 46); and a nucleotide sequence of uncertain size that is apparently necessary for incorporation of viral RNA into maturing virions (23, 34). To facilitate the full elucidation of these functionally important features, we have determined the sequence of 1,010 nucleotides at the 5′ end of the genome of the subgroup A Schmidt-Ruppin strain of Rous sarcoma virus (RSV). Our experimental approach exploited the availability of molecularly cloned DNA derived from the genome of subgroup A Schmidt-Ruppin RSV (7). The results locate the initiation codon for the *gag* gene of RSV; demonstrate that this codon is preceded by three AUG codons that are apparently not used in translation, permit deduction of the amino acid sequence of a viral structural protein (p19), and identify a nucleotide sequence that is likely to be the splice donor site mapped in the accompanying manuscript (13).

MATERIALS AND METHODS

All analyses were performed on DNA derived from the molecular cloning of the genome of SR-A RSV. The viral genome was first cloned into the phage vector λgtWESλB (2, 21) as described previously (7). The SRA-2 clone of viral DNA was used in this study (7). Suitable restriction fragments were then subcloned into the plasmid vector pBR322 by conventional procedures (3). Two subclones were used for the present work (Fig. 1). One (pPvu DG) contained the *Pvu*II-D and -G fragments from RSV DNA, joined at the *Sac*I site to regenerate their normal orientation compared with the *Sac*I permuted SRA-2 clone (R. Parker, personal communication); the other plasmid used (pBam C) contained the 1.35-kilobase-pair *Bam*HI-C fragment of SRA-2 DNA, which extends from position 525 on the viral genome (the RNA capping site is used as position 1) to a point well within *gag* (7). Viral DNA in the subclones was sequenced by the procedure of Maxam and Gilbert (24). For this purpose, we prepared restriction fragments with 5′ overhangs that permit end labeling either by repair synthesis with reverse transcriptase (39) or by T4 polynucleotide kinase (24).

RESULTS

Strategy for determining the nucleotide sequence of viral DNA. The RNA genome of RSV is transcribed into DNA during the early hours of infection (42). The first stable product of viral

DNA synthesis is a linear duplex that is coextensive with the viral genome, but that has, in addition, terminal redundancies composed of three domains: U3, encoded at the 3′ end of the viral RNA; R, a small redundancy encoded at each end of the viral RNA; and U5, encoded at the 5′ end of the viral RNA (17, 32). These domains are arranged in the order 5′-U3-R-U5-3′ to give the long terminal repeat (LTR). The linear duplexes are then formed into closed circular molecules of two sorts—those containing one copy of the LTR sequence and those that contain two copies (17, 18, 32, 39). We have previously reported the isolation and molecular cloning of the circular DNAs of RSV (7). In this study we have used the SRA-2 clone of viral DNA which contains two copies of the LTR sequence (7, 39).

Two subclones of viral DNA in pBR322 were used for sequence determination (Fig. 1): (i) pPvu DG, a clone of the *Pvu*II-D and -G fragments that spans the fused 3′ and 5′ ends of the RSV genome and that includes ca. 250 nucleotides from the 3′ end of the RSV genome, both copies of the LTR sequence, and ca. 750 nucleotides from the 5′ end of the viral RNA (7, 39); and (ii) pBam C, which overlaps with the rightward domain of pPvu DG and extends from position 525 on the RSV genome to a point well within *gag* (7). Restriction mapping of these subclones revealed the cleavage sites illustrated in Fig. 1 and led to the sequencing strategy summarized there. To assure accuracy, each region of viral DNA was either sequenced independently from more than one restriction site or sequenced repeatedly from the same site.

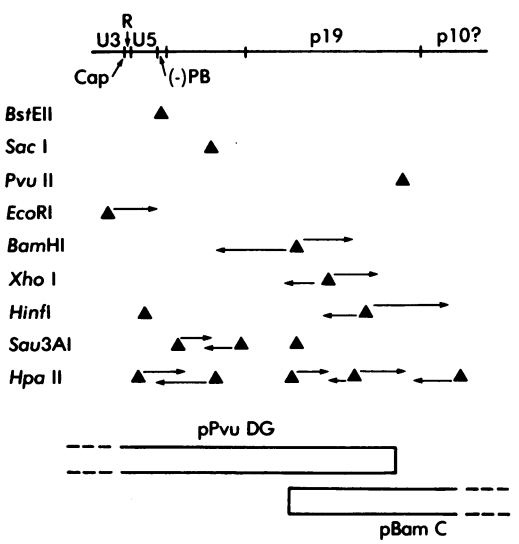

FIG. 1. Strategy for sequencing viral DNA. The molecular subclones pPvu DG and pBam C were mapped with restriction endonucleases to find cleavage sites useful in sequencing by the technique of Maxam and Gilbert (24). The two clones overlap by 325 nucleotides, permitting the construction of a continuous sequence. The domains of the viral genome represented in the sequenced DNA are depicted at the top of the drawing. The domains U3, R, and U5 are explained in the text. The 5′ end of the viral genome is denoted by the label Cap, the site of binding for $tRNA^{Trp}$ by (−)PB. The position of p19 has been approximated from the data reported in the present manuscript; the location of p10 is based largely on previous reports (35, 44) and unpublished data of E. Hunter (personal communication). Arrows indicate the direction of sequencing employed at individual restriction sites.

Nucleotide sequence of the DNA. Figure 2 illustrates the sequence of 1,010 nucleotides of viral DNA from the recombinant clones and identifies the location of all recognizable restriction sites. Figure 3 presents the same sequence as the plus strand of viral RNA. The sequence as illustrated in both figures begins at the 5′ terminus of the subgroup A Schmidt-Ruppin RSV genome and extends in the 3′ direction. Because the 3′ and 5′ ends of the viral genome are fused in the cloned DNA (see above), it was necessary to deduce the location of the 5′ terminus of the RNA. This was easily done by reference to previous analyses of other strains of RSV (12, 15, 19, 37, 40) and by our own studies using an adaptation of the chain-terminator sequencing technique of Sanger and colleagues (31, 40) to determine the sequence of the runoff DNA product synthesized from the 5′ terminus of subgroup A Schmidt-Ruppin RSV RNA (data not shown).

Functional aspects of the sequence. Several important landmarks could be located on the sequence. (i) The terminally redundant sequence known as R has been characterized previously and occupies positions 1 through 21 (3a). (ii) A sequence of 18 nucleotides (positions 102 through 119) is base-paired with the 3′ stem region of $tRNA^{Trp}$ to provide a primer for reverse transcriptase (4, 10, 39); as a consequence, viral DNA synthesis initiates at position 101 on the template (15, 37). (iii) Previous work has identified a ribosome binding site that includes the AUG at position 41 and neighboring nucleotides, positions 9 through 53 (6); paradoxically, this AUG apparently does not serve for the initiation of translation (see below). (iv) The beginning of the *gag* gene was identified by searching for an AUG followed by an extensive open reading frame (Fig. 4) and by reference to the previously determined amino acid sequence at the amino terminus of the *gag* gene product (27). A suitable AUG was found at position 372 (Fig. 4) and was followed by the predicted amino acid sequence (see below). (v) In the accompa-

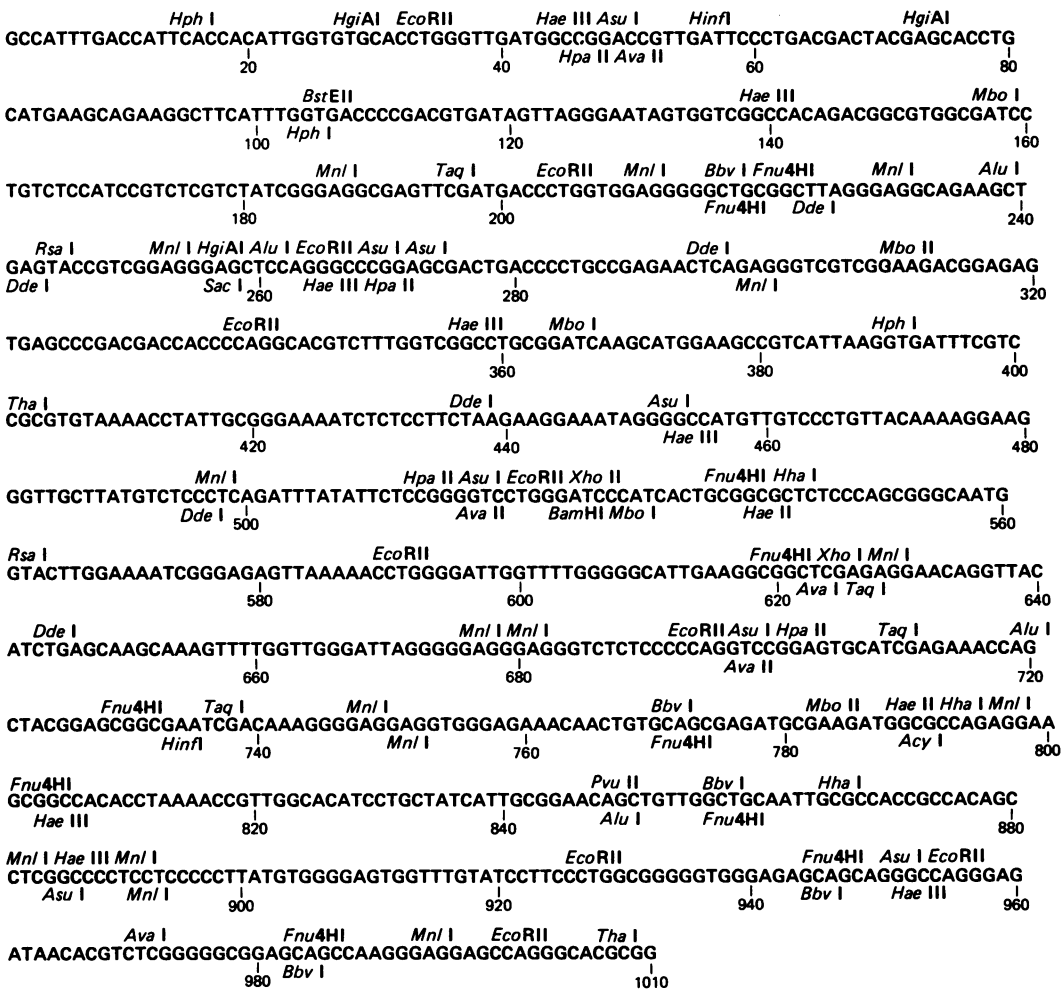

FIG. 2. Nucleotide sequence of cloned viral DNA. The sequence is numbered from the 5′ end of the viral genome and depicts the same polarity as the genome ("positive strand"). Identifiable sites of cleavage by restriction endonucleases were located by computer-assisted search. Identifiable regions within the sequence, as described in the text, are as follows: R, 1-21; U5, 22-101; (−)PB, 102-119; start of *gag* and p19, 372; end of p19, 902.

nying manuscript (13), a splice donor site has been mapped to the vicinity of position 390; an excellent facsimile of the canonical splice donor sequence (22) occupies positions 387 through 395 (see below; Fig. 5).

Although the sequence given here is more than ample to accommodate the entirety of p19, we cannot presently identify the carboxy terminus of the protein with certainty. Since p19 is but one of five proteins encoded by *gag* in a single polyprotein (with the probable order p19-p10-p27-p12-p15; see references 35 and 44; personal communication from E. Hunter), the carboxy terminus of p19 is not demarcated by a termination codon. On the basis of incomplete data describing amino acid sequences in p19 and other *gag* proteins, however, we suggest that p19 may terminate with amino acid residue 177 (tyrosine) in Fig. 3 (see below).

DISCUSSION

Authentication of the nucleotide sequence. We present here the sequence of 1,010 nucleotides, beginning at the 5′ terminus of the RSV genome and extending through the p19 domain of the *gag* gene. We have a number of reasons to believe that our identification and sequencing of the viral DNA have not been affected by any large errors. First, the subclone pPvu DG includes both copies of the LTR sequence and displays at least one of the functional activities of this domain—the ability to direct the initiation of

GCCAUUUGACCAUUCACCACAUUGGUGUGCACCUGGGUUGAUGGCCGGACCGUUGAUUCCCUGACGACUA 70

CGAGCACCUGCAUGAAGCAGAAGGCUUCAUUUGGUGACCCCGACGUGAUAGUUAGGGAAUAGUGGUCGGCCACAGACGGC 150

GUGGCGAUCCUGUCUCCAUCCGUCUCGUCUAUCGGGAGGCGAGUUCGAUGACCCUGGUGGAGGGGGCUGCGGCUUAGGGA 230

GGCAGAAGCUGAGUACCGUCGGAGGGAGCUCCAGGGCCCGGAGCGACUGACCCCUGCCGAGAACUCAGAGGGUCGUCGGA 310

1
met glu ala
AGACGGAGAGUGAGCCCGACGACCACCCCAGGCACGUCUUUGGUCGGCCUGCGGAUCAAGC AUG GAA GCC 380

10 20
val ile lys val ile ser ser ala cys lys thr tyr cys gly lys ile ser pro ser lys
GUC AUU AAG GUG AUU UCG UCC GCG UGU AAA ACC UAU UGC GGG AAA AUC UCU CCU UCU AAG 440

30 40
lys glu ile gly ala met leu ser leu leu gln lys glu gly leu leu met ser pro ser
AAG GAA AUA GGG GCC AUG UUG UCC CUG UUA CAA AAG GAA GGG UUG CUU AUG UCU CCC UCA 500

50 60
asp leu tyr ser pro gly ser trp asp pro ile thr ala ala leu ser gln arg ala met
GAU UUA UAU UCU CCG GGG UCC UGG GAU CCC AUC ACU GCG GCG CUC UCC CAG CGG GCA AUG 560

70 80
val leu gly lys ser gly glu leu lys thr trp gly leu val leu gly ala leu lys ala
GUA CUU GGA AAA UCG GGA GAG UUA AAA ACC UGG GGA UUG GUU UUG GGG GCA UUG AAG GCG 620

90 100
ala arg glu glu gln val thr ser glu gln ala lys phe trp leu gly leu gly gly gly
GCU CGA GAG GAA CAG GUU ACA UCU GAG CAA GCA AAG UUU UGG UUG GGA UUA GGG GGA GGG 680

110 120
arg val ser pro pro gly pro glu cys ile glu lys pro ala thr glu arg arg ile asp
AGG GUC UCU CCC CCA GGU CCG GAG UGC AUC GAG AAA CCA GCU ACG GAG CGG CGA AUC GAC 740

130 140
lys gly glu glu val gly glu thr thr val gln arg asp ala lys met ala pro glu glu
AAA GGG GAG GAG GUG GGA GAA ACA ACU GUG CAG CGA GAU GCG AAG AUG GCG CCA GAG GAA 800

150 160
ala ala thr pro lys thr val gly thr ser cys tyr his cys gly thr ala val gly cys
GCG GCC ACA CCU AAA ACC GUU GGC ACA UCC UGC UAU CAU UGC GGA ACA GCU GUU GGC UGC 860

170 180
asn cys ala thr ala thr ala ser ala pro pro pro pro tyr val gly ser gly leu tyr
AAU UGC GCC ACC GCC ACA GCC UCG GCC CCU CCU CCC CCU UAU GUG GGG AGU GGU UUG UAU 920

190 200
pro ser leu ala gly val gly glu gln gln gly gln gly asp asn thr ser arg gly arg
CCU UCC CUG GCG GGG GUG GGA GAG CAG CAG GGC CAG GGA GAU AAC ACG UCU CGG GGG CGG 980

210
ser ser gln gly arg ser gln gly thr arg
AGC AGC CAA GGG AGG AGC CAG GGC ACG CGG

FIG. 3. Proposed amino acid sequence for the p19 domain of RSV *gag*. The sequence presented in Fig. 2 is here rewritten as viral RNA, beginning at the 5′ terminus of the RSV genome. The underlining denotes the site of binding for $tRNA^{Trp}$. An open reading frame that may encode p19 was identified as described in the text and in the legend to Fig. 4 and used to deduce a proposed amino acid sequence for the p19 domain of *gag*.

RNA synthesis both in vitro (W. DeLorbe, personal communication) and in vivo (7; P. Luciw, personal communication). Second, both subclones have been mapped extensively with restriction endonucleases (7; Fig. 1). The sequence illustrated in Fig. 2 contains all of the sites predicted by these previous analyses. Third, our present sequence of the U5 domain is in agreement with previous results obtained from viral DNA synthesized in vitro (15, 37, 40). Fourth, the sequence correctly locates and represents the previously characterized binding site for $tRNA^{Trp}$ (4, 10). Fifth, the viral DNA was derived from a replication-competent virus, and the cloned DNA is infectious (7). Sixth, we can deduce from our sequence the previously determined amino acid sequence from the amino terminus of the *gag* gene (27); the sequence contains a single open reading frame in the *gag* region.

Locating *gag* proteins on the nucleotide sequence. The initial product of translation from *gag* appears to be $Pr76^{gag}$ (29, 45), a 76,000-dalton protein whose amino-terminal sequence has been determined to be: met-glu-ala-val-ile-lys-val-x-x-ala-x-lys (27). A virtually identical

Nucleotide Position	Reading Frame 1	2	3	Nucleotide Position	Reading Frame 1	2	3
7	OP			372			AUG
39			OP	386		OC	
41		AUG		391	OP		
54			OP	407		OC	
62		OP		437		OC	
82	AUG			448	AM		
83		OP		456			AUG
105			OP	489			AUG
116		OP		558			AUG
119		AM		583	OC		
123			AM	613	OP		
130	AM			644		OP	
198			AUG	670	AM		
199	OP			778	AUG		
225			AM	786			AUG
240			OP	812		OC	
278		OP		901	AUG		
321			OP	962		OC	

FIG. 4. Analysis of reading frames. All potential initiation and termination codons were located in each of the three reading frames; frame 1 begins with the first nucleotide at the 5′ terminus of the sequence (Fig. 3), frame 2 begins with the second, and frame 3 begins with the third. AUG, Methionine-potential initiation codon; AM, amber termination codon (UAG); OC, ochre termination codon (UAA); OP, opal termination codon (UGA). The AUG codon which we deduce initiates the *gag* gene in frame 3 is underlined.

sequence is encoded by nucleotide residues 372 to 410 in RSV RNA (Fig. 3), the only discrepancy being the presence of three amino acids between val and ala rather than two. Despite this unexplained discrepancy, we presume that the virtual identity suffices to locate the beginning of the *gag* gene on the nucleotide sequence.

The cleavage of $Pr76^{gag}$ gives rise to five viral proteins whose order within the polyprotein precursor is probably: p19-p10-p27-p12-p15 (35, 44; personal communication from E. Hunter). The carboxy terminus of p19 is not demarcated by a termination codon, but we have been able to deduce the approximate location of the terminus to be position 177 by using unpublished findings of E. Hunter (personal communication). (i) On the basis of carboxypeptidase treatment, p19 ends in tyrosine. Given the molecular weight of p19, the tyrosine residue at position 155, 177, or 183 might represent the carboxy terminus (Fig. 3). (ii) The tyrosine at residue 183 appears to lie within p10 (E. Hunter, personal communication). (iii) Carboxypeptidase cleaves tyrosine from p19 and then fails to progress further into the protein (E. Hunter, personal communication). The sequence before the tyrosine at residue 155 should be fully susceptible to hydrolysis. The proline residue that precedes the tyrosine at position 177 would allow only the release of the tyrosine (14), making this tyrosine the likely carboxy terminus of p19. Definitive identification of the carboxy terminus of p19 and the exact localization of p10 await further analysis of amino acid sequence in the isolated proteins.

Ribosome binding and the initiation of translation from *gag*. Translation from the mRNA's of eucaryotic organisms generally initiates only in the vicinity of the 5′ terminus of the RNA (M. Kozak, *in* A. Shatkin, ed., *Current Topics in Microbiology and Immunology*, in press). Sherman and Stewart (36) and Kozak (20) have proposed that ribosomes inevitably bind at or near the 5′ end of eucaryotic mRNA's and then "travel" to the first AUG downstream, where translation initiates. Previous findings with RSV were in accord with these views: translation of the intact RSV genome initiates only at *gag*, the gene located closest to the 5′ end of the genome (29, 45); and a strong ribosome binding site has been identified that includes the first AUG downstream from the 5′ end of RSV RNA (6). Paradoxically, however, the identified ribosome binding site does not represent the site of initiation for translation from *gag*. This paradox was apparent from a comparison of previous analyses of the nucleotide sequence in the U5 domain (15, 37) and the amino-terminal sequence of $Pr76^{gag}$ (27), and the paradox is fully manifest in our present data. The initiation codon for *gag* is preceded by three AUG codons, each of which is followed shortly and in frame by one or more termination codons (Fig. 4). It therefore appears that the site to which ribosomes initially bind in vitro at the 5′ end of the RSV genome does not of itself dictate the position at which translation starts. These findings add another example to the growing list of eucaryotic mRNA's which do not initiate translation at the first methionine codon downstream from the 5′ end of the RNA (26; Kozak, in press).

Kozak has recently compiled the nucleotide sequences that adjoin initiation codons in eucaryotic mRNA's and has found that adenosine usually occurs in the third position upstream from the AUG (83% of available examples), and guanosine occurs at the first position downstream from the AUG (63% of the available examples) (Kozak, in press). The significance of

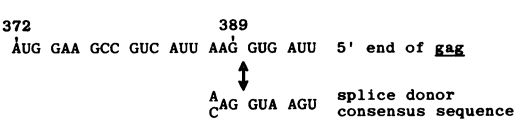

FIG. 5. Identification of a potential splice donor site in RSV RNA. Data described in the accompanying manuscript (13) locate a splice donor site in the vicinity of residue 390 of the RSV genome. The sequences adjoining this position are here aligned with the "consensus sequence" for splice donor sites, compiled from a large number of viral and cellular mRNA's (22).

these findings is not known, but the same features do occur in the sequences adjoining the initiation codon for *gag* of RSV (Fig. 3). One possible distinction between the AUG codons in the 5′ noncoding region of RSV and the initiation codon of *gag* is that the former are preceded by a U at position −3 from the AUG. Of the 56 cellular mRNA sequences reviewed by Kozak (in press), none has a pyrimidine at −3 from the initiation codon. Two of these RNAs have AUG codons in the 5′ noncoding region; in each case the AUG is preceded by a C at position −3 from the AUG. The presence of a purine or a pyrimidine at position −3 from an AUG codon may be one of the signals a cell normally uses to distinguish between correct and incorrect initiation sites. Further examples are needed to clarify this correlation.

Splicing in the genesis of RSV mRNA's. The subgenomic mRNA's of RSV are formed by splicing a short nucleotide sequence from the 5′ end of the viral genome to at least two positions within the genome (5, 20a, 25, 38, 46). In the accompanying manuscript (13), we have located the "donor" site for splicing approximately 18 nucleotides downstream from the beginning of the *gag* gene. This position is contained within a nucleotide sequence that displays close similarity to the previously enunciated "consensus sequence" for splice donor sites in eucaryotic mRNA's (Fig. 5). According to the consensus sequence, the position of the splice would be located between residues 389 and 390 of RSV RNA (Fig. 5), a deduction which is in exact accord with the results obtained by mapping the splice donor site with S1 nuclease (13) and which adds credence to those results.

If we have correctly located the splice donor site in RSV RNA, the initiation codon for *gag* is spliced onto the subgenomic mRNA's of the virus. Might this codon be used for initiation of translation in its transposed positions? The question cannot be answered with assurance as yet because the splice acceptor sites in the *env* and *src* subgenomic mRNA's have not been located, although indirect evidence suggests that the initiation codon for the *src* coding region lies outside of the spliced leader sequence (13). In contrast to RSV, the splice donor site for the *env* mRNA of mouse mammary tumor virus appears to lie upstream from the start of *gag*; there are no AUG codons in the leader sequence (J. Majors, personal communication).

Mapping 5′ noncoding regions of avian retroviruses. The sequence analysis of the 5′ terminus of RSV has provided a potentially useful battery of restriction enzyme sites that could be applied to rapidly mapping other avian retrovirus isolates. In particular, a *Bst*EII site (position 106) is present within the sequence of the $tRNA^{Trp}$ binding site [labeled (−)PB in Fig. 1], and a *Bam*HI site (position 525) and an *Xho*I site (position 625) are present within the coding region of p19 (Fig. 1 and 2). All three sites have been documented by nucleotide sequence analysis in avian erythroblastosis virus (M. Privalsky, personal communication), and they map to analogous positions in the endogenous viral locus, *ev*1 (16). The *Bam*HI and *Xho*I sites are present in the DNA of RSV Prague strain, RAV-O, and MC29 viruses (32, 33, 43). These three restriction enzyme sites will probably be reliable markers for quickly identifying the position of the $tRNA^{Trp}$ binding site and mapping the start of the *gag* sequences in other isolates.

ACKNOWLEDGMENTS

We thank Bill DeLorbe and Richard Parker for making cloned RSV DNAs available and John Majors, Martin Privalsky, Dennis Schwartz, and Eric Hunter for communicating results before publication. We also thank Bertha Cook for help in preparing the manuscript. The computer analysis of the sequence was done with the assistance of Peter Czernilofsky and Hugo Martinez.

R.S. was supported by Public Health Service Training grant 1T32 CA 09043 from the National Institutes of Health. The research was supported by Public Health Service grants CA 12705 and CA 19287 from the National Institutes of Health and by American Cancer Society grant MV48G to J.M.B. and H.E.V.

LITERATURE CITED

1. **Bishop, J. M.** 1978. Retroviruses. Annu. Rev. Biochem. **47**:35–88.
2. **Blattner, F. R., A. E. Blechl, K. Kenniston-Thompson, H. E. Faber, J. E. Richards, J. L. Sightom, P. W. Tucker, and O. Smithies.** 1978. Cloning human fetal globin and mouse α-type globin DNA: Preparation and screening of shotgun preparations. Science **202**:1279–1284.
3. **Bolivar, F., and K. Backman.** 1979. Plasmids of *Escherichia coli* as Cloning vectors. Methods Enzymol. **68**:245–266.

3a. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42**:1–26.

4. **Cordell, B., E. Stavnezer, R. Friedrich, J. M. Bishop, and H. M. Goodman.** 1976. The nucleotide sequence that binds primer for DNA synthesis to the avian sarcoma virus genome. J. Virol. **19**:548–558.
5. **Cordell, B., S. R. Weiss, H. E. Varmus, and J. M. Bishop.** 1978. At least 104 nucleotides are transposed from the 5′ terminus of the avian sarcoma virus genome to the 5′ termini of smaller viral mRNAs. Cell **15**:79–91.
6. **Darlix, J.-L., P.-F. Spahr, P. A. Bromley, and J.-C. Jaton.** 1979. In vitro, the major ribosome binding site on Rous sarcoma virus RNA does not contain the nucleotide sequence coding for the N-terminal amino acids of the *gag* gene product. J. Virol. **29**:597–611.
7. **DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
8. **Donoghue, D. J., E. Rothenberg, N. Hopkins, D. Baltimore, and P. A. Sharp.** 1978. Heteroduplex analysis of the nonhomology region between Moloney MuLV and the dual host range derivative HIX virus. Cell **14**:959–970.
9. **Donoghue, D. J., P. A. Sharp, and R. A. Weinberg.** 1979. An MSV-specific subgenomic mRNA in MSV transformed G8-124 cells. Cell **17**:53–64.
10. **Eiden, J. J., K. Quade, and J. L. Nichols.** 1976. Interaction

of tryptophan transfer RNA with Rous sarcoma virus 35S RNA. Nature (London) 259:245–247.
11. **Faller, D. V., J. Rommelaere, and N. Hopkins.** 1978. Large T1 oligonucleotides of Moloney leukemia virus missing in an *env* gene recombinant, HIX, are present on an intracellular 21S Moloney virus RNA species. Proc. Natl. Acad. Sci. U.S.A. 75:2964–2968.
12. **Furuichi, Y., A. J. Shatkin, E. Stavnezer, and J. M. Bishop.** 1975. Blocked, methylated 5′ terminal sequence in avian sarcoma virus RNA. Nature (London) 257:618–620.
13. **Hackett, P. H., R. Swanstrom, H. E. Varmus, and J. M. Bishop.** 1982. The leader sequence of the subgenomic mRNA's of Rous sarcoma virus is approximately 390 nucleotides. J. Virol. 41:527–534.
14. **Hartsuck, J. A., and W. N. Lipscome.** 1971. Carboxypeptidase A. *In* P. D. Boyer (ed.), The enzymes, vol. 3, p. 1–56. Academic Press, Inc., New York.
15. **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.
16. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus *evl*: integration creates as 6bp repeat of host DNA. Cell 23:155–164.
17. **Hsu, T. W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and J. M. Taylor.** 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. J. Virol. 28:810–818.
18. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell 22:376–386.
19. **Keith, J., and H. Fraenkel-Conrat.** 1975. Identification of the 5′ end of Rous sarcoma virus RNA. Proc. Natl. Acad. Sci. U.S.A. 72:3347–3350.
20. **Kozak, M.** 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 15:1109–1123.
20a. **Krzyzek, R. A., M. S. Collett, A. F. Lau, M. L. Perdue, J. P. Leis, and A. J. Faras.** 1978. Evidence for splicing of avian sarcoma virus 5′-terminal genomic sequences onto viral-specific RNA in infected cells. Proc. Natl. Acad. Sci. U.S.A. 75:1284–1288.
21. **Leder, P., D. Tiemeier, and L. Enquist.** 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. Science 196:175–177.
22. **Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz.** 1980. Are snRNP's involved in splicing? Nature (London) 283:220–224.
23. **Linial, M., E. Medeiros, and W. S. Hayward.** 1978. An avian oncovirus mutant (*SE*21Q1b) deficient in genomic RNA: biological and biochemical characterization. Cell 15:1371–1381.
24. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
25. **Mellon, P., and P. H. Duesberg.** 1977. Subgenomic, cellular Rous sarcoma virus RNAs contain oligonucleotides from the 3′ half and the 5′ terminus of virion RNA. Nature (London) 270:631–634.
26. **Mulligan, R. C., and P. Berg.** 1981. Factors governing expression of a bacterial gene in mammalian cells. Mol. Cell. Biol. 1:449–459.
27. **Palmiter, R. D., J. Gagnon, V. M. Vogt, S. Ripley, and R. N. Eisenman.** 1978. The NH_2-terminal sequence of the avian oncovirus *gag* precursor polyprotein ($Pr76^{gag}$). Virology 91:423–433.
28. **Panet, A., M. Gorecki, S. Bratosin, and Y. Aloni.** 1978. Electron microscopic evidence for splicing of Moloney murine leukemia virus RNAs. Nucleic Acids Res. 5:3219–3236.
29. **Pawson, T., G. S. Martin, and A. E. Smith.** 1976. Cell-free translation of virion RNA from nondefective and transformation-defective Rous sarcoma viruses. J. Virol. 19:950–957.
30. **Rothenberg, E., D. J. Donoghue, and D. Baltimore.** 1978. Analysis of a 5′ leader sequence on murine leukemia virus 21S RNA: Heteroduplex mapping with long reverse transcriptase products. Cell 13:435–451.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463–5467.
32. **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 avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell 15:1383–1395.
33. **Shank, P. R., S. H. Hughes, and H. E. Varmus.** 1981. Restriction endonuclease mapping of the DNA of Rous-associated virus O reveals extensive homology in structure and sequence with avian sarcoma virus DNA. Virology 108:177–188.
34. **Shank, P. R., and M. Linial.** 1980. Avian oncornavirus mutant (*SE*21Q1b) deficient in genomic RNA: characterization of a deletion in the provirus. J. Virol. 36:450–456.
35. **Shealy, D. J., A. G. Mosser, and R. R. Rueckert.** 1980. Novel p19-related protein in Rous-associated virus type 61: implications for avian *gag* gene order. J. Virol. 34:431–437.
36. **Sherman, F., and J. W. Stewart.** 1975. The use of iso-1-cytochrome c mutants of yeast for elucidating the nucleotide sequences that govern initiation of translation, p. 175–191. *In* G. Bernardi and F. Gros (ed.), Organization and expression of the eukaryotic genome: biochemical mechanism of differentiation in prokaryotes and eukaryotes, Proceedings of the 10th FEBS meeting. American Elsevier, New York.
37. **Shine, J., A. P. Czernilofsky, R. Friedrich, H. M. Goodman, and J. M. Bishop.** 1977. Nucleotide sequence at the 5′ terminus of the avian sarcoma virus genome. Proc. Natl. Acad. Sci. U.S.A. 74:1473–1477.
38. **Stoltzfus, C. M. and L. K. Kuhnert.** 1979. Evidence for the identity of shared 5′-terminal sequences between genome RNA and subgenomic mRNA's of B77 avian sarcoma virus. J. Virol. 32:536–545.
39. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Sequencing of cloned unintegrated avian sarcoma virus DNA: Viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. 78:124–128.
40. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1981. The terminal redundancy of the retrovirus genome facilitates chain elongation by reverse transcriptase. J. Biol. Chem. 256:1115–1121.
41. **Taylor, J. M.** 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. Biochim. Biophys. Acta 473:57–71.
42. **Varmus, H. E., S. Heasley, H.-J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop, and P. R. Shank.** 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. J. Mol. Biol. 120:55–82.
43. **Vennstrom, B., C. Moscovici, H. M. Goodman, and J. M. Bishop.** 1981. Molecular cloning of the avian myelocytomatosis virus genome, and recovery of infectious virus by transfection of chicken cells. J. Virol. 39:625–631.
44. **Vogt, V. M., R. Eisenman, and H. Diggelmann.** 1975. Generation of avian myeloblastosis virus structural proteins by proteolytic cleavage of a precursor polypeptide. J. Mol. Biol. 96:471–493.
45. **von der Helm, K., and P. H. Duesberg.** 1975. Translation of Rous sarcoma virus RNA in cell free systems from Ascites Krebs II cells. Proc. Natl. Acad. Sci. U.S.A. 72:614–618.
46. **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.