JOURNAL OF VIROLOGY, Mar. 1985, p. 969–972
0022-538X/85/030969-04$02.00/0

Vol. 53, No. 3

Cloning and Nucleotide Sequences of cDNAs Spanning the Splice Junctions of Rous Sarcoma Virus mRNAs

LUNG-JI CHANG AND C. MARTIN STOLTZFUS*

Department of Microbiology, University of Iowa, Iowa City, Iowa 52242

Received 21 September 1984/Accepted 20 November 1984

The cDNAs corresponding to the 5′ ends of the mRNAs coding for the envelope protein precursor (gPr92^{env}) of the B77 strain and the transforming protein (pp60^{src}) of the Prague B strain of Rous sarcoma virus were cloned into pBR322, and the nucleotide sequences surrounding the splice junctions were determined. Both mRNAs are products of single splicing events from a common donor splice site at nucleotide 398 from the 5′ end of the RNA to acceptor splice sites at nucleotides 5078 and 7054 for the *env* and *src* mRNAs, respectively. These results confirm and extend previous conclusions based on peptide mapping and single-strand nuclease mapping. Compared with the sequence of the Prague C genome RNA, the B77 strain contains a 6-nucleotide deletion in the sequence corresponding to the hydrophobic portion of the signal peptide of the envelope protein precursor.

The complete nucleotide sequence of the Prague C strain of Rous sarcoma virus (RSV) has been determined (16). Based on the sequence, it was proposed that the 3′ acceptor splice junction for the envelope (*env*) gene mRNA may occur at nucleotide 5078 from the 5′ end of the RNA in a 3′ consensus splice sequence (16). The general location of this acceptor site was confirmed by S1 nuclease mapping experiments, although the size of the S1-resistant product was 4 or 5 nucleotides longer than expected (5). The 5′ donor splice junction has also been determined by S1 mapping (7, 18) as well as by tryptic and chymotryptic peptide analysis (5) to lie within the *gag* gene coding region. Because the initiator AUG codon lies within the *gag* gene-containing leader region and is in phase with the *env* gene sequences, the primary *env* and *gag* gene products share a common 6-amino acid N-terminal sequence (5). From these data, however, we could not rule out the possibility that more than one splicing event occurred during the formation of the leader of the *env* gene mRNA, i.e., that there are multiple leader sequences, as is the case with spliced mRNAs arising from transcripts derived from the major late promoter of adenoviruses (1, 3). Similarly, the locations of the splice junctions of the mRNA for pp60^{src} have been inferred from single-strand nuclease mapping experiments (17); again, however, the possibility that multiple splicing events occur has not been ruled out. To test for these possibilities, as well as to construct cDNA clones which can then be used for independent expression of viral genes, we cloned the cDNAs corresponding to the 5′ termini of the *env* and *src* mRNAs and determined the nucleotide sequences across the splice junctions.

To isolate clones containing sequences spanning the *env* and *src* gene mRNA splice junctions, we purified restriction fragments downstream from the putative splice junctions and used these fragments as primers for the synthesis of cDNA. These restriction fragments were isolated from a genomic clone of the Prague A strain of RSV (pJD100), which was obtained from J. T. Parsons, University of Virginia. For the *env* sequences, we used a fragment extending from the *Xho*I site at nucleotide 5258 to the *Pst*I site at nucleotide 5564. This fragment was then denatured and annealed to 21S polyadenylated RNA from cells infected with a transformation-defective strain of B77 avian sarcoma virus. For the *src* gene mRNA, we used a primer extending from the *Ava*II site at nucleotide 7155 to an *Ava*II site at nucleotide 7283. This fragment was denatured and annealed to 21S polyadenylated RNA from cells infected with a nondefective Prague B strain of RSV. After the annealing reaction, the hybrids were selected by binding and elution twice to oligodeoxythymidylate-cellulose and used as template-primer for avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, Fla.). The second strand of cDNA was synthesized either by reverse transcriptase or by using the Klenow fragment of *Escherichia coli* DNA polymerase I (Bethesda Research Laboratories, Gaithersburg, Md.). After the second strand synthesis, the cDNA preparation was treated with mung bean nuclease (P-L Biochemicals, Inc., Milwaukee, Wis.) to remove the hairpin structure at the 5′ end of the second strand. The synthesized cDNA was cloned into the *Pst*I site of pBR322 by a homopolymeric deoxycytidine-deoxyguanosine tailing (19). *E. coli* HB101 cells were transformed by the recombinant plasmids, and transformants were selected for tetracycline resistance and ampicillin sensitivity. Colonies were lysed and screened by colony hybridization (6) with different ^{32}P-labeled nick-translated probes (15). For the selection of *env* mRNA clones which spanned the splice junction, we isolated a fragment from pJD100 extending from the *Kpn*I site at nucleotide 4995 to an *Xho*I site at nucleotide 5258. For the selection of *src* mRNA cDNA clones, we isolated a fragment extending from the *Ava*II site at nucleotide 7155 to the *Ava*II site at nucleotide 7283. For isolation of both cDNA clones, we used a probe extending from the *Eco*RI site at nucleotide −50 to the *Pst*I site at nucleotide 263, which contains sequences at the extreme 3′ and 5′ ends of the genomic RNA (3′-5′ probe). Colonies hybridizing to both 3′-5′ and gene-specific probes were suspected of being cDNA clones and were analyzed further.

A number of clones were obtained in these experiments which hybridized only to the 3′-5′ probe. These clones were also obtained when oligodeoxythymidylate was used to prime cDNA synthesis of viral RNA. Based on further sequence analysis (data not shown), the clones appeared to arise from initiations at the natural tryptophan tRNA primer-binding site of the viral RNA. This occurred in spite of the

* Corresponding author.

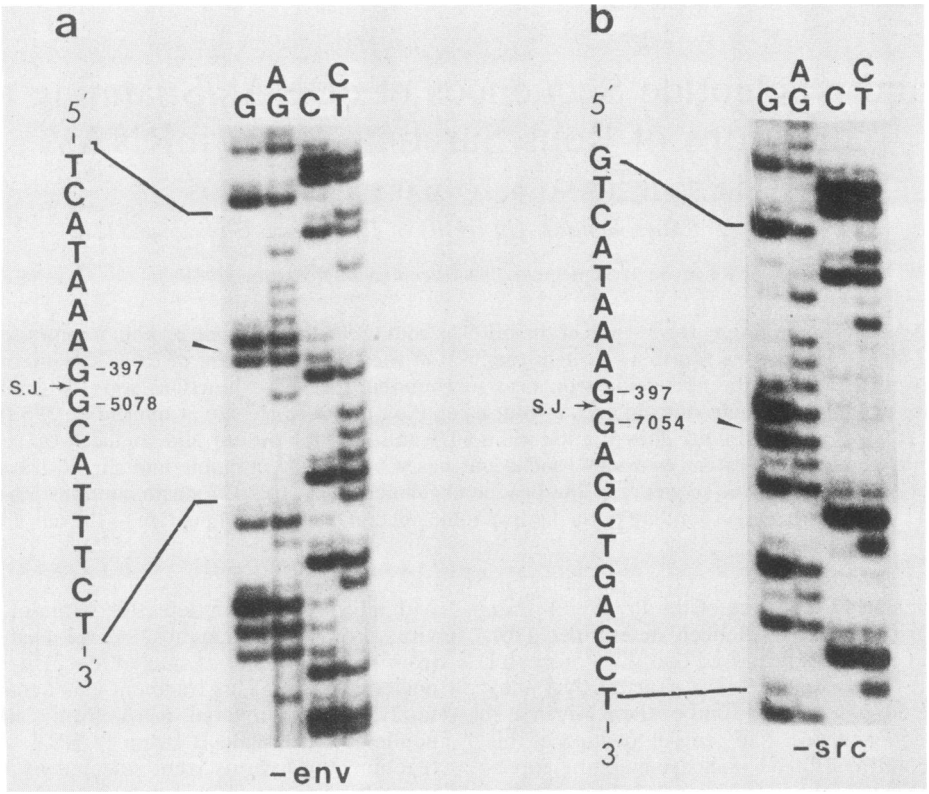

FIG. 1. Sequencing of the region around the splice sites of the *env* and *src* gene cDNA clones. Arrows indicate the locations of the splice junctions (S.J.) for the *env* (a) and *src* (b) gene mRNAs. The nucleotide numbers are derived by comparison with the published sequence for Prague C RSV genome RNA (16).

fact that the mRNA was prepared by complete denaturation of the RNA followed by glycerol gradient centrifugation of the denatured RNA, conditions which should have removed any tRNA primers bound to the RNA. It is possible, however, that small amounts of tRNA primer remain bound even after the denaturation procedure and that this accounts for the appearance of these clones. It is also possible that a fraction of the avian myeloblastosis virus reverse transcriptase, which is known to be associated with natural tRNA primers (13), is targeted to the primer-binding sites.

Two candidate B77 *env* gene mRNA cDNA clones were selected which hybridized to both the 3'-5' and *env* gene probes. The plasmid DNAs were purified and digested with restriction endonucleases *Xho*I and *Pst*I. One fragment of approximately 400 base pairs which hybridized to both the 3'-5' and the *env* gene probes was produced from both clones. This fragment was isolated, labeled at the *Xho*I end with [α-^{32}P]TTP (Amersham Corp., Arlington Heights, Ill.) by using the Klenow fragment of *E. coli* DNA polymerase I, and sequenced by the Maxam and Gilbert method (12). A portion of one of the sequencing gels showing the region included in the splice junction is given in Fig. 1A. The sequence obtained from both clones was identical and is shown (Fig. 2) along with the corresponding sequences of the Prague C genomic RNA, as determined by Schwartz et al. (16). It is clear from this result that the mRNA leader is derived by means of a single splicing event from nucleotide 398 to nucleotide 5078. This result confirms and extends our previous data on the location of the *env* gene mRNA splice junction and rules out the possibility of multiple splices within the 5' leader sequence. It is of interest that the B77 sequence contains a 6-nucleotide deletion within the sequences corresponding to the hydrophobic carboxy-terminal portion of the signal peptide of gPr92^{env}. This implies that the two amino acids coded for by the deleted sequences are not required for the glycoprotein to function. It is also of interest that there is a base change from a C to a T immediately 5' to the initiator AUG codon. This is in contrast to a number of other avian retroviruses in which the entire 9 nucleotides upstream from the AUG are conserved (2).

One Prague B RSV *src* gene mRNA cDNA clone was selected which hybridized to both the 3'-5' and *src* gene probes. Plasmid DNA from this clone was purified and digested with restriction enzymes *Ava*II and *Pst*I. A fragment of approximately 400 base pairs was obtained which hybridized to both probes. This fragment was isolated, labeled at the 3' end of the *Ava*II site with [α-^{32}P]dGTP (Amersham), and sequenced by the Maxam and Gilbert method. A portion of a sequencing gel in the region spanning the splice junction is shown in Fig. 1B, and the sequence together with the corresponding Prague C RSV genomic RNA sequence is shown in Fig. 3. Note that the *src* gene mRNA also arises as a result of a single splicing event from nucleotide 398 to nucleotide 7054. The location of the splice at this site implies, as previously pointed out by others (16), that the AUG sequence in the 5' leader which is used for the initiations of the *gag* and *env* gene products (the third AUG from the 5' end) cannot be used for the initiation of pp60^{src} synthesis since it is closely followed by in-phase terminator codon UGA at nucleotides 7063 through 7065. This is an interesting and unique situation for an mRNA; a functional AUG which is imbedded in a sequence which satisfies all of Kozak's rules for optimal initiation (i.e., an A at position -3

FIG. 2. cDNA sequence of the 5′ end and splice site of transformation-defective B77 RSV *env* gene mRNA. The sequence of the 5′-end cDNA spanning the splice junction is compared with the Prague C (PrC) genome sequence (16). Amino acids shown represent differences between these two strains. The hydrophobic region of the envelope protein signal peptide is marked. The location of the sequence coding for the N-terminus of the gp85 protein (9) is shown. ***, Initiation codon for *gag* and *env* gene products.

and a G at position +4 relative to the first A in the AUG [10]) is 5′ to the AUG actually used to initiate synthesis of pp60^{src}. This implies either that ribosomes can bind internally to the *src* gene AUG or that ribosomes initiating at the upstream third AUG remain attached to the mRNA after termination at the in-frame UGA codon and proceed to the fourth AUG. Evidence consistent with the latter possibility has recently been obtained (8, 11). Such an arrangement may be of

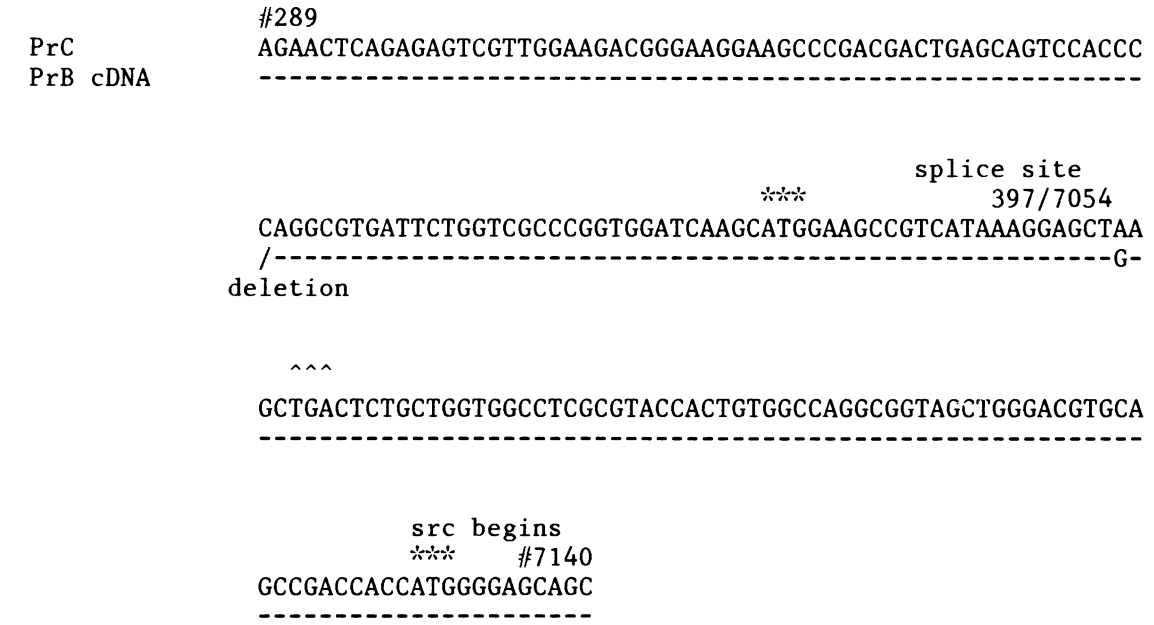

FIG. 3. cDNA sequence at the 5′ end and splice site of the *src* gene mRNA of Prague B (PrB) RSV. The sequence is compared with the Prague C (PrC) RSV genome sequence (16). The first indicated AUG is used to initiate both *gag* and *env* gene products, but it terminates upstream of the AUG used to initiate pp60^{src}. ***, Initiation codon; ^^^, termination codon.

significance in determining the translational efficiency of the *src* gene mRNA and therefore may play a role in regulating the expression of $pp60^{src}$. It has also been shown that the major ribosome-binding site on the RSV genome RNA occurs within the 5′-terminal 100 bases at the first AUG proximal to the 5′ cap (4, 14). This suggests the possibility that a ribosome may actually initiate and terminate protein synthesis several times before initiating at the appropriate AUG in the *src* gene mRNA.

The research was supported by Public Health Service grant CA28051 awarded by the National Cancer Institute.

LITERATURE CITED

1. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. U.S.A. **74**:3171–3175.
2. **Bizub, D., R. A. Katz, and A. M. Skalka.** 1984. Nucleotide sequence of noncoding regions in Rous-associated virus-2: comparisons delineate conserved regions important in replication and oncogenesis. J. Virol. **49**:557–565.
3. **Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts.** 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell **12**:1–8.
4. **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.
5. **Ficht, T. A., L.-J. Chang, and C. M. Stoltzfus.** 1984. Avian sarcoma virus *gag* and *env* gene structural protein precursors contain a common amino-terminal sequence. Proc. Natl. Acad. Sci. U.S.A. **81**:362–366.
6. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
7. **Hackett, P. B., R. Swanstrom, H. E. Varmus, and J. M. Bishop.** 1982. The leader sequence of the subgenomic mRNAs of Rous sarcoma virus is approximately 390 nucleotides. J. Virol. **41**:527–534.
8. **Hughes, S., K. Mellstrom, E. Kosik, F. Tamanoi, and J. Brugge.** 1984. Mutation of a termination codon affects *src* initiation. Mol. Cell. Biol. **4**:1738–1746.
9. **Hunter, E., E. Hill, M. Hardwick, A. Bhown, D. Schwartz, and R. Tizard.** 1983. Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its product. J. Virol. **46**:920–936.
10. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
11. **Liu, C.-C., C. C. Simonsen, and A. D. Levinson.** 1984. Initiation of translation at internal AUG codons in mammalian cells. Nature (London) **309**:82–85.
12. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
13. **Panet, A., W. A. Haseltine, D. Baltimore, G. Peters, F. Harada, and J. E. Dahlberg.** 1975. Specific binding of tryptophan transfer RNA to avian myeloblastosis virus RNA-dependent DNA polymerase (reverse transcriptase). Proc. Natl. Acad. Sci. U.S.A. **72**:2535–2539.
14. **Petersen, R. B., C. H. Hensel, and P. B. Hackett.** 1984. Isolation of a ribosome binding site for a leader peptide encoded by Rous sarcoma virus RNA. J. Virol. **51**:722–729.
15. **Rigby, P. M. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
16. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
17. **Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop.** 1983. Transduction of cellular oncogenes: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **80**:2519–2523.
18. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1982. Nucleotide sequence of the 5′ noncoding region and part of the *gag* gene of Rous sarcoma virus. J. Virol. **41**:535–541.
19. **Villa-Komaroff, L., A. Efstradiadis, S. Broome, P. Lemedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert.** 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. U.S.A. **75**:3727–3731.