JOURNAL OF VIROLOGY, June 1987, p. 2038–2041
0022-538X/87/062038-04$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 61, No. 6

Splice Acceptor Site for the *env* Message of Moloney Murine Leukemia Virus

PEDRO A. LAZO,* VINAYAKA PRASAD, AND PHILIP N. TSICHLIS
Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received 4 November 1986/Accepted 2 March 1987

We report the isolation and sequence of a cDNA clone containing part of the *env* message of the Moloney murine leukemia virus (MoMuLV). This clone was derived from a rat thymic lymphoma induced by MoMuLV. The AG acceptor site employed in this message is located at position 5490 in the MoMuLV genome. This splice site is detectable at the cDNA level by the creation of a novel *Sac*I restriction site not present in the viral genome. In the −1 to −40 region, this AG acceptor site is preceded by four conserved heptanucleotides (PyXPyTPuAPy) that may function as acceptors for removal of the 5′ end of the intron.

Moloney murine leukemia virus (MoMuLV) is a murine retrovirus that lacks a transforming (*onc*) gene but that, when inoculated into rats or mice, induces either thymic lymphomas or splenic T-cell lymphomas (26). The MoMuLV virus has been cloned (2), and its complete nucleotide sequence has been determined (25). The availability of the sequence permitted the location of the boundaries of the known viral genes and a close approximation of the sequences of the corresponding mRNAs (29). However, information regarding the structure of these mRNAs is incomplete since it is based largely on indirect evidence (4, 10, 25, 26, 28). Because of the location of splice consensus sequences, for example, Shinnick et al. predicted that the splice acceptor site used for the generation of the envelope (*env*) message would be located at nucleotide 5501 (25). This prediction was later challenged in a report by Mann and Baltimore (15). While studying the effect of varying the position of the packaging-signal sequence within the viral genome, Mann and Baltimore observed the encapsidation of what appeared to be spliced viral mRNAs. Integrated copies of these putative messages were cloned and sequenced. These experiments showed that if indeed these integrated proviruses were derived from reverse-transcribed copies of spliced mRNA, then the most commonly used splice acceptor site for the *env* message was at position 5490.

In the course of our work on the induction of thymomas by MoMuLV (26, 27), we isolated a partial cDNA clone of a spliced subgenomic *env* message. This clone was analyzed for splice donor and acceptor site usage, and the data were interpreted in light of recent information on the structural requirements for splicing (5, 12), which involves the following steps: assembly of the splicing structure (3, 6, 7–9, 13, 22), hydrolysis and guanylation of the 5′ end of the intron, ligation of this 5′ end of the intron to a conserved heptanucleotide (11, 16–19, 21) near its 3′ end to form a branched RNA, and, finally, exon ligation (23).

The cDNA clone we describe here was isolated from a λgt11 cDNA library constructed from polyadenylated RNA (1, 14) isolated from one of the MoMuLV-induced rat thymic lymphomas (26, 27). This clone (designated p*env*1 in this report) was selected so that it hybridized to DNA probes derived from the U5 and *pol* regions but did not hybridize to a *gag* (*Pst*I-*Xho*I) probe. Restriction endonuclease mapping revealed a *Sac*I site which divided the cloned DNA into a 210-base-pair fragment hybridizing to the U5 probe and a 270-base-pair U5-negative fragment (Fig. 1A). This division was unexpected because there is only one *Sac*I site in the viral genome outside the long terminal repeat and that site is present in the *gag* region (25), for which the clone tested negative. Three other clones with the same phenotype as clone p*env*1 were isolated from this library. Because all the clones had an identical restriction map, including the location of the linker *Eco*RI sites, we concluded that they were the result of the library amplification and not independently originated cDNA clones. Hybridization of the 270-base-pair U5-negative fragment to polyadenylated tumor cell RNA revealed two RNA species comigrating with the MoMuLV genomic (8.3-kilobase) and *env* (3.0-kilobase) mRNA transcripts (Fig. 1B). This finding suggested that the cDNA clone p*env*1 may represent the 5′ end of the MoMuLV *env* mRNA transcript. To further characterize p*env*1, the *Eco*RI insert

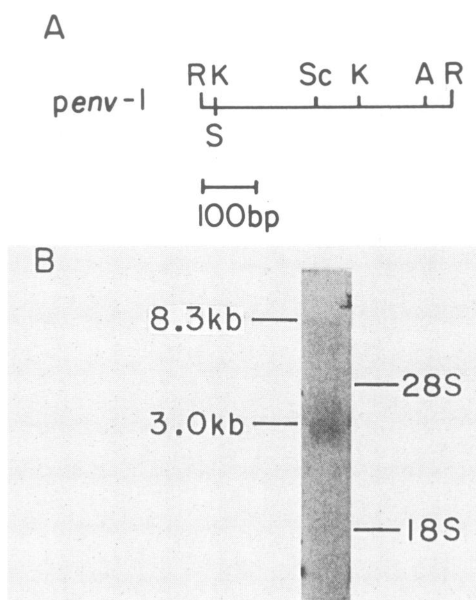

FIG. 1. (A) Restriction map of the insert present in the *env* cDNA clone p*env*1. R, *Eco*RI; A, *Acc*I; K, *Kpn*I; Sc, *Sac*I; and S, *Sma*I. (B) Northern blot analysis of polyadenylated RNA from rat thymoma analyzed with the 270-base-pair (bp) *Sac*I-*Eco*RI fragment from clone p*env*1. kb, Kilobases.

* Corresponding author.

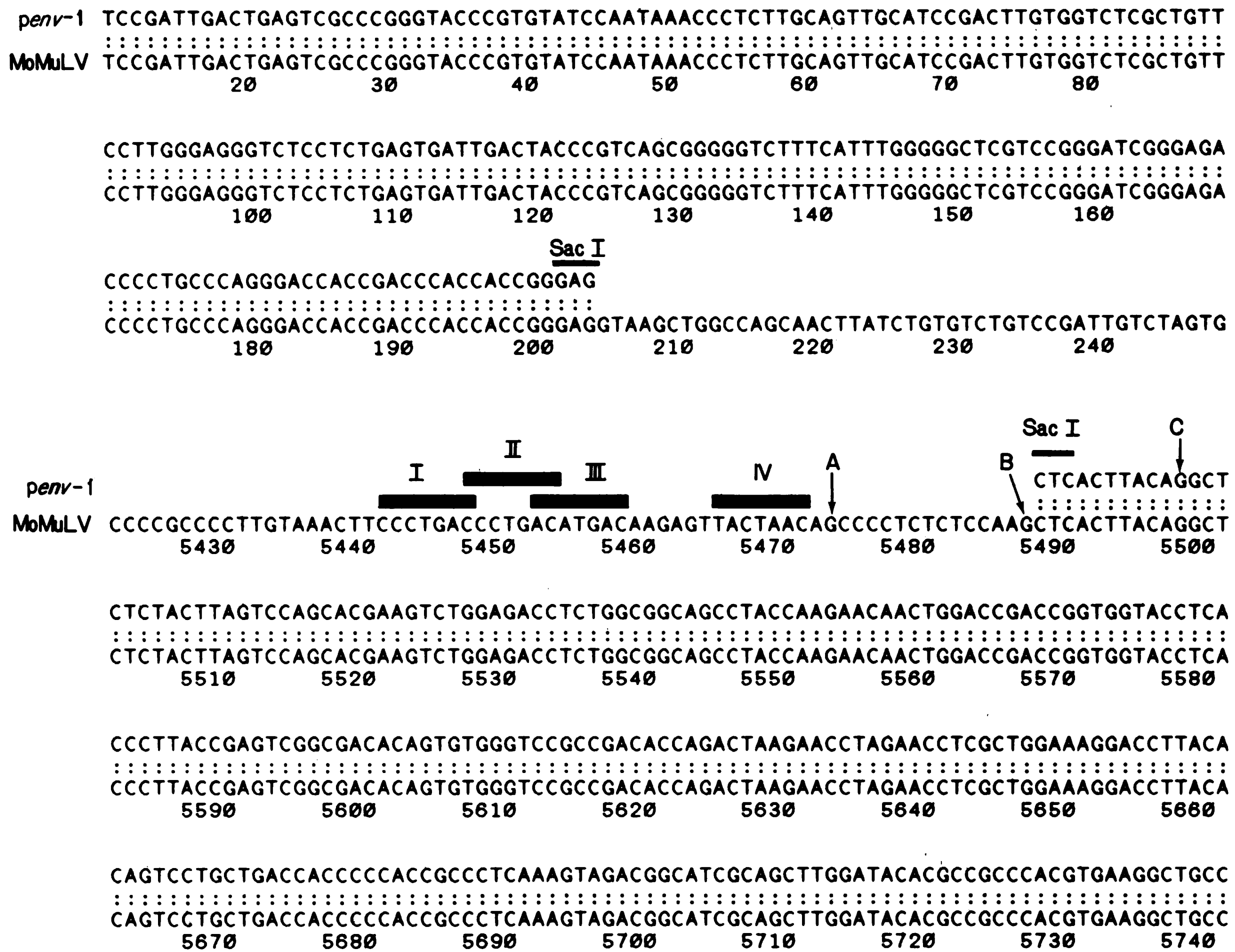

FIG. 2. Nucleotide sequence of the cDNA clone p*env*1 and comparison of it with the genomic sequence of MoMuLV (10, 25, 28). The AG dinucleotides that can serve as splice acceptor sites are labeled A, B, and C. The locations of the putative lariat heptanucleotides (I to IV) are indicated by bars. The location of a novel *Sac*I restriction site created as a result of the splicing reaction is also shown.

was subcloned in M13mp10 and its nucleotide sequence was determined in both directions by the dideoxy chain termination method (24).

The nucleotide sequence (Fig. 2) showed that the clone indeed represents the 5′ end of the MoMuLV *env* message and that the novel *Sac*I restriction site was created as a consequence of the splicing reaction. Interestingly enough, however, the sequencing of this clone located the AG of the splice acceptor site at position 5490. The splice donor and acceptor sites predicted by Shinnick et al. (25) on the basis of homologies with consensus sequences were located at positions 205 and 5501, respectively.

The region between nucleotides 5440 and 5501 of MoMuLV contains several interesting sequence features that could be relevant to splice acceptor site usage (Fig. 2). Three AG dinucleotides that can serve as acceptor sites are located at positions 5476, 5490, and 5501 (Fig. 2, arrows A, B, and C, respectively), as well as four heptanucleotides (bars I to IV in Fig. 2) which could work as acceptors for the removal of the 5′ end of the intron by forming a branched RNA structure or lariat (5, 12). These sequence features and a comparison of them to other messages are shown in Table 1.

The heptanucleotide usually lies in the −20 to −40 region with respect to the acceptor AG (19, 20), and the heptanucleotide sequence is conserved among species. Furthermore, in lower eucaryotes such as *Saccharomyces cerevisiae*, the function of the heptanucleotide (TACTAAC) is very sensitive to mutations (11, 16, 17). In mammals, however, there is more sequence variation (19, 21, 23), although the relative position with respect to acceptor AG and the final nucleotides, particularly the A, is conserved since the acceptor AG and nucleotides are required for branch formation in the lariat structure. The consensus sequence PyXPyTPuAPy is most similar to sequences I and II described in this report. Taking into account the sequence conservation and spacing, splice acceptor site B is the most likely to be used since it is compatible with all of the heptanucleotides except heptanucleotide IV. Splice acceptor sites A and C, although theoretically possible, have more restricted possibilities, since A can use only sequence I and C can use only sequence IV, which is identical to the yeast heptanucleotide.

In their analysis of apparently integrated copies of reverse-transcribed *env* mRNAs, Mann and Baltimore (15) observed that two of the copies were generated by using acceptor site B. However, a third copy was generated by using acceptor site A. The authors argued that acceptor site A represents a cryptic splice site used because of a mutation that may have altered the actual splice site. However, there

TABLE 1. Location of lariat heptanucleotides with respect to AG acceptors

Gene	Heptanucleotide sequence[a]	Nucleotide spacing	Acceptor AG
MoMuLV			
I	CCCTGAC	24	A
		40	B
		51	C
II	CCCTGAC	18	A
		34	B
		45	C
III	ACATGAC	13	A
		29	B
		40	C
IV	TACTAAC	01	A
		16	B
		27	C
Human β globin IVS-1	CACTGAC	37	
Human γ globin IVS-1	TTCTGAC	30	
Human ε globin IVS-1	CTCTAAT	31	
Human α globin IVS-1	CCCTCAC	19	
IVS-2	CACTGAC	18	
Rabbit β globin IVS-1	TGCTGAC	34	
Mouse β globin IVS-1	CACTAAC	36	
Rat insulin IVS-1	CCTCAAC	18	
Drosophila *ftz*	AGCTAAC	30	
Adenovirus 5 E1A	GTTTAAA	30	
Consensus sequence	PyXPyTPuAPy		

[a] The heptanucleotide sequences of eucaryotic genes other than the *env* gene of MoMuLV were obtained from Reed and Maniatis (19, 20) A in the consensus sequence is the nucleotide involved in the formation of a branched RNA molecule by formation of a 2′ phosphodiester bond with the 5′ end of the intron (5, 12).

was no evidence for such a mutation in the published sequence (15). The observation of four heptanucleotides and three AG acceptor sites is unusual, and we speculate that one reason for their concentration might be to drive the assembly of the splicing superstructure, the rate-limiting step in the splicing reaction (8, 9, 13). This organization may be important for the removal of large introns, like the 5,285-nucleotide intron removed for the generation of the *env* message.

From consideration of our observations regarding the sequence organization around the beginning of the *env* gene and the *env* message from MoMuLV-induced thymomas, as well as the data on the encapsidation of the putative spliced viral mRNAs (15), we concluded that the AG site at position 5490 is the main splice acceptor site used for the generation of the *env* message by MoMuLV.

The typing of the manuscript by Annmarie Shepherd is greatly appreciated.

P.A.L. is a Fellow of the Cancer Research Institute of New York. P.N.T. is a Scholar of the Leukemia Society of America. This work was supported by Public Health Service grants CA38047 from the National Cancer Institute to P.N.T. and CA06927 and RR05539 from the National Institutes of Health and by an appropriation from the Commonwealth of Pennsylvania to the Fox Chase Cancer Center.

LITERATURE CITED

1. **Aviv, M., and P. Leder.** 1972. Purification of biologically active globin mRNA by chromatography on oligothymidilic acid cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
2. **Berns, A. J. M., M. H. T. Lai, R. A. Bosselman, M. A. McKennett, L. T. Bacheler, H. Fan, E. C. R. Maandag, H. V. D. Putten, and I. M. Verma.** 1980. Molecular cloning of unintegrated and a portion of integrated Moloney murine leukemia viral DNA in bacteriophage lambda. J. Virol. **36**:254–263.
3. **Black, D. L., B. Chabot, and J. A. Steitz.** 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. Cell **42**:737–750.
4. **Bosselman, R. A., F. van Straaten, C. Van Beveren, I. M. Verma, and M. Vogt.** 1982. Analysis of the *env* gene of a molecularly cloned and biologically active Moloney mink cell focus-forming proviral DNA. J. Virol. **44**:19–31.
5. **Cech, T. R.** 1986. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell **44**:207–210.
6. **Chabot, B., D. L. Black, D. M. LeMaster, and J. A. Steitz.** 1985. The 3′-splice site of a premessenger RNA is recognized by a small nuclear ribonucleoprotein. Science **230**:1344–1349.
7. **Frendeway, D., and W. Keller.** 1985. The stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell **42**:355–367.
8. **Grabowski, P. J., S. R. Seiler, and P. A. Sharp.** 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell **42**:345–353.
9. **Grabowsky, P. J., and P. A. Sharp.** 1986. Affinity chromatography of splicing complexes: U2, U5, and U4 + U5 small nuclear ribonucleoprotein particles in the spliceosome. Science **233**: 1294–1299.
10. **Hughes, S.** 1982. Complete nucleotide sequences of three retroviral genomes and a cellular *onc* gene, p. 1349–1356. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **Jacquier, A. J., R. Rodriguez, and M. Rosbach.** 1985. A quantitative analysis of the effects of 5′ junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. Cell **43**:423–430.
12. **Keller, W.** 1984. The RNA lariat: a new ring to the splicing of mRNA precursors. Cell **39**:423–425.
13. **Kreiner, A. R., and T. Maniatis.** 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing *in vitro*. Cell **42**:725–736.
14. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Mann, R., and D. Baltimore.** 1985. Varying the position of a retrovirus-packaging sequence results in the encapsidation of both unspliced and spliced RNAs. J. Virol. **54**:401–407.
16. **Newman, A. J., R. Lim, S. Cheng, and J. Abelson.** 1985. Molecular consequences of specific intron mutations in yeast mRNA splicing *in vivo* and *in vitro*. Cell **42**:335–344.
17. **Parker, R., and C. Guthrie.** 1985. A point mutation in the conserved hexanucleotide at a yeast 5′ splice function encouples recognition, cleavage, and ligation. Cell **41**:107–118.
18. **Pikielny, C. W., and M. Rosbach.** 1985. mRNA splicing efficiency in yeast and the contribution of non-conserved sequences. Cell **41**:119–126.
19. **Reed, R., and T. Maniatis.** 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell **41**:95–105.
20. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46**:681–690.
21. **Ruskin, B., and M. Green.** 1985. Role of the 3′ splice site consensus sequence in mammalian pre-mRNA splicing. Nature (London) **317**:732–734.
22. **Ruskin, B., and M. Green.** 1985. Specific and stable intron-factor interactions are established early during *in vitro* pre-mRNA splicing. Cell **43**:131–142.
23. **Ruskin, B., A. R. Kraimer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell **38**:317–331.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
26. **Tsichlis, P. N., P. G. Strauss, and L. F. Hu.** 1983. A common region for proviral DNA integration in MoMuLV-induced rat

thymic lymphomas. Nature (London) **302**:445–449.

27. **Tsichlis, P. N., P. G. Strauss, and M. A. Lohse.** 1985. Concerted DNA rearrangements in Moloney murine leukemia virus-induced thymomas: a potential synergistic relationship in oncogenesis. J. Virol. **56**:258–267.
28. **Van Beveren, C., J. Coffin, and S. Hughes.** 1985. Appendix: Moloney murine leukemia virus (MoMLV) genome, p. 766–782. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Supplements and appendixes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Van Beveren, C., J. A. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma.** 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) **289**:258–262.