Aberrant splicing events that are induced by proviral integration: Implications for *myb* oncogene activation

(oncogenes/plasmacytoid lymphosarcomas/RNA splicing)

DAN ROSSON*, DEBBIE DUGAN†, AND E. PREMKUMAR REDDY*‡

Department of Molecular Oncology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110

Communicated by Allan H. Conney, January 27, 1987 (received for review October 1, 1985)

ABSTRACT Activation of the mouse *c-myb* oncogene in Abelson virus-induced plasmacytoid lymphosarcomas was studied using cDNA cloning and nucleotide sequence analysis. The results presented here show that viral integration in the *myb* locus generates splicing errors at the 5' and 3' regions. Viral integration results in transcriptional initiation within the viral long terminal repeat and generation of a chimeric mRNA that lacks the first three coding exons. The alterations at the 3' end are caused by an aberrant splicing event in which additional splice-donor and -acceptor sequences within intronic sequences are used to splice an additional 363 nucleotides into the *myb* transcripts. The resulting insertion of 121 amino acids is in a region of the protein where other activated forms of the *myb* gene product have deletions. These results suggest that alterations in the 3' end of the *myb* gene play a crucial role in the activation of this gene.

A comparative study of retroviral oncogenes and their cellular homologues has given us invaluable information regarding the structural changes that protooncogenes often undergo to acquire transforming potential. A comparison of the *c-myb* cDNA sequence (1) with the *v-myb* sequence of avian myeloblastosis virus (AMV) (2, 3) and E26 leukemia virus (4) revealed that large stretches of coding sequences from the 5' and 3' ends have been deleted during the transduction of this gene into the viral genome. The AMV *myb* gene is derived from seven of the internal coding exons of the *c-myb* gene (3, 5) and has undergone deletions of 142 amino acids at its N-terminal end and 192 amino acids at its C-terminal end (1). Similar deletions have occurred in the E26 virus, in which *myb* is coded as part of a fusion polypeptide of three distinct elements, the viral *gag* sequences, the viral *myb*, and a third element derived from the protooncogene *ets* (4). Deletions of similar stretches of sequences in two independent viral isolates suggest that these deletions are important for the oncogenic activation of the *myb* gene.

Another model system that has allowed us to study the activation of the *myb* gene is the Abelson leukemia virus-induced plasmacytoid lymphosarcomas (ABPL tumors) of the mouse (6). Detailed analysis of these tumors revealed that they do not contain the integrated Abelson leukemia virus genome, but instead have undergone rearrangements in their *myb* locus as a result of Moloney murine leukemia virus (Mo-MuLV) integration (5, 7, 8).

To study the effect of the viral integration on the transcription of the rearranged *myb* locus in ABPL tumors, we have undertaken the cDNA cloning of *myb* mRNA from one of the ABPL tumors, ABPL-2, and determined its sequence. Our results show that the viral integration results in aberrant splicing events at both the 5' and 3' ends of the mRNA

transcripts producing alterations in similar positions to those seen in AMV and E26 virus.

MATERIALS AND METHODS

Cell Lines. Cell lines from ABPL-1, ABPL-2, and ABPL-4 were derived from tumors induced by the injection of pristane and a mixture of Abelson leukemia virus and Mo-MuLV (6, 9, 10). Cells were grown in suspension culture in Dulbecco's modified Eagle's medium supplemented with 10% (vol/vol) fetal calf serum.

cDNA Library Construction and Screening. The cDNA library was constructed from poly(A) RNA isolated from freshly excised mouse thymuses or ABPL tumor cells as described (11, 12). cDNA synthesis was as described (1). cDNA (50 ng) was ligated to 1 μ g of λ gt10 arms to obtain a phage library of 5×10^5 recombinants, which was screened by standard techniques (24). RNA gel blot analysis was as described by Lehrach *et al.* (13).

DNA Sequence Analysis. The nucleotide sequence was determined by the procedure of Maxam and Gilbert (14).

RESULTS

Transcription of *myb* Locus in ABPL Tumors. Earlier studies indicated that all ABPL tumors have undergone rearrangements in the *myb* locus as a result of integration of the Mo-MuLV proviral genome with the *myb* gene at a point immediately preceding the first exon that exhibits homology to the *v-myb* gene (5, 8). RNA isolated from the tumor tissues showed the presence of normal [3.8–4.0 kilobases (kb)] and abnormal (5.0 kb) transcripts that cross-hybridized to radio-labeled *v-myb* probes (7). Since the tumor cells grown *in vivo* are often contaminated with other lymphoid cells as well as granulocytes and macrophages, it was not clear whether the ABPL tumors synthesized normal *c-myb* transcripts or that these RNAs are derived from contaminating normal cells. To distinguish between these possibilities, we established three of the tumor lines in tissue culture and examined their RNAs by RNA gel blot analysis. These results (data not shown) indicated that the RNA derived from mouse thymus contains a single diffuse band in the range of 3.8–4.0 kb, whereas the three tumor RNAs contain transcripts >4.0 kb. There were no detectable 3.8- to 4.0-kb RNAs, although these cells retained one normal nonrearranged *c-myb* allele (7, 8).

Abbreviations: ABPL, Abelson virus-induced plasmacytoid lymphosarcomas; AMV, avian myeloblastosis virus; Mo-MuLV, Moloney murine leukemia virus; nt, nucleotide(s).

*Present address: The Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, PA 19104.

†Present address: Center for Neurobiology, Columbia University, 722 West 168th Street, New York, NY 10032.

‡To whom reprint requests should be addressed at: The Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, PA 19104.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

FIG. 1. Restriction map of ABPL-2 *c-myb* cDNA as derived from the two overlapping clones λ PL2-1 and λ PL2-6. Restriction sites are A, *Aat* II; Bg, *Bgl* II; E, *Eco*RI; N, *Nco* I; P, *Pst* I; Sc, *Sac* I; X, *Xho* I; Ss, *Ssp* I. The restriction map for normal *c-myb* was derived from Gonda *et al.* (15). The darkened portion of the bars indicates normal *c-myb* sequence. The open bar represents non-*myb* sequence. The position of the potential initiation (ATG) and terminator (TGA) codons are indicated.

Structure and Sequence of the Aberrant *c-myb* Transcripts in the ABPL-2 Tumor Cell Line. To determine the structure of the aberrant transcripts, we constructed a cDNA library from the poly(A) RNA of ABPL-2. Screening of the library with a radiolabeled *c-myb* probe showed the presence of a high percentage of positive clones; 24 of which were grown and further characterized by restriction enzyme cleavage mapping. A restriction map of the two clones used for detailed analysis is shown in Fig. 1. A comparison to the restriction map of the normal *myb* gene (15, 16) shows that the ABPL-2 cDNA differs from the normal *c-myb* cDNA at the 5' end as well as in the middle portion where insertion of ≈ 400 base pairs (bp) is seen. To delineate further the nature of the changes that have occurred in the formation of the aberrant *c-myb* mRNAs, we carried out complete nucleotide sequence analysis of two overlapping clones derived from this aberrant mRNA. The composite nucleotide sequence is shown in Fig. 2. The sequence analysis revealed that the 5' end of these cDNA clones contain sequences derived from the Mo-MuLV proviral genome (17). Thus, the first 1596 bases of the transcript appear to be derived from the helper viral genome that is followed by the *myb* coding sequences. The junction between viral and *myb* sequences occurs at a point within the viral genome that contains the *gag* coding sequence CAG-GCA-GGT-AGG-ACC in which the underlined sequence is homologous to the consensus splice donor sequence MAG $\downarrow$ GTRAGT (20), where the arrow indicates the splice point, M is either cytidine or adenosine, and R is either adenosine or guanosine. The splice-acceptor sequence is contained within the *c-myb* locus and is, in fact, the natural splice-acceptor sequence present at the junction of the first viral-related exon and its upstream intron. The splicing event between the viral and *myb* sequences occurs such that the *gag* sequences are fused to the *myb* sequences in a reading frame that is different from that utilized for the synthesis of *gag* protein. This reading frame contains two in-frame ATG codons beginning at positions 1509 and 1512 either of which could be used for the initiation of protein synthesis by the ABPL-2 mRNA.

The homology between the normal *c-myb* and the ABPL-2-derived *c-myb* sequences starts from position 1597 and continues for a stretch of 990 nucleotides (nt). At nt-2587 the *myb* sequence is interrupted by a stretch of 363 nt that bears no homology with the normal *c-myb* cDNA sequences. The interruption occurs at a point that constitutes the junction of the sixth and seventh *v-myb*-related exon. However, from

nt-2950, the homology with normal *c-myb* sequences resumes and is continued until the end of the cDNA clone. Since the insertion of this stretch of 363 nt does not change the reading frame of the 3' end of the transcript, the protein derived from such an aberrant mRNA would contain 715 amino acids and code for a protein that contains an N-terminal region derived from the Mo-MuLV genome and a C-terminal region derived from the *myb* gene that is interrupted by a stretch of 121 amino acids from position 360 through 481.

Origin of the Inserted Sequences Present in ABPL-2-Derived *c-myb* cDNA Clones. To determine the origin of the additional sequences present in the cDNA clones derived from the ABPL-2 tumor line, we initially prepared a DNA fragment specific for these sequences by cleavage of the cDNA clone with *Eco*RI and *Aat* II. A radioactive probe derived from this fragment was used to probe mouse genomic DNA. These hybridization experiments suggested that these sequences were present within the *myb* gene itself, presumably within the intronic sequences between viral-related exons 6 and 7. This was further confirmed by examination of mouse genomic clones (provided by J. Ihle, National Cancer Institute, Frederick Cancer Research Facility) and nucleotide sequence analysis of the intronic sequences present between viral-related exons 6 and 7. Sequence analysis of this intron revealed the presence of an identical stretch of sequence beginning ≈ 850 bases from the 3' end of the sixth exon (data not shown). The sequence of the mouse genome at this region (which normally constitutes an intron) was found to be GTTGTTCAG $\downarrow$ GATTCT=351=TCGCAG $\downarrow$ GTAGGGC-AG, where the arrows indicate splice points. The sequence on the 5' side of the insert corresponds well to the consensus splice-acceptor sequence (Y)₆NCAG $\downarrow$ GK (20), where Y is thymidine or cytidine, N is any nucleotide, and K is guanosine or thymidine. Similarly, the sequence immediately following the insertion sequences corresponds well to the consensus splice donor sequence MAG $\downarrow$ GTRAGT (20). Thus, in the generation of the ABPL-2 *myb* mRNA, an additional set of splice-donor and -acceptor sequences within intronic sequences at the 3' end of the gene are used. The splicing event that produces such an aberrant mRNA is illustrated in Fig. 3.

To determine whether the aberrant splicing observed constitutes the predominant transcriptional product or whether it represents a minor species that is produced by these tumors, we cloned and sequenced six additional cDNA

FIG. 2 (on opposite page). Nucleotide sequence of the ABPL-2 cDNA deduced from cDNA clones λ PL2-1 and λ PL2-6. The nucleotide sequence proceeds in the 5' to 3' direction. The deduced amino acid sequence from the open reading frame is shown below the nucleotide sequence. Numbering of the sequence is based on the sequence of the Mo-MuLV genome (17). Vertical arrows indicate *Eco*RI sites, boundaries for sequences homologous to the *v-myb* sequences of AMV and E26 virus, and the termination point where *c-myb* is disrupted in the NFS-60 cell line by integration of the Cas-Br virus (18, 19). The positions of those sequences coded for by the viral-related exons 1-7 and the insertion of intron sequences are also indicated. Termination codons inframe with the reading frame beginning at positions 1482 and 3655 are indicated by ***.

FIG. 2. (Legend appears at the bottom of the opposite page.)

FIG. 3. Comparison of the transcription events that generate the normal *myb* mRNA and the ABPL-2 mRNA. The seven viral-related *myb* exons are numbered and are represented by solid boxes (5). The open boxes represent nonviral-related exons. The integration site of Mo-MuLV is immediately before the first viral exon. In generation of the ABPL-2 mRNA, transcription is initiated within the Mo-MuLV genome. Splicing occurs normally throughout the processing of the mRNA except just after the sixth viral-related exon where an additional 363 nt are inserted into the message via an aberrant splicing event. The positions of the first three 5' exons as well as the seventh viral-related exon have been determined (5) and are as shown. The exons following the last viral-related exon have not been mapped and are illustrated for diagrammatical purposes. Hatched boxes, viral long terminal repeats.

clones containing the 3' end of the gene from the same library. All contained the identical sequence across the abnormal splicing site. Still other clones with this sequence have been identified through restriction enzyme analysis. We found no clones containing a normal 3' end. Therefore, we conclude that the aberrantly spliced *myb* mRNA is the vastly predominant *myb* mRNA in ABPL-2 cells. These conclusions are further confirmed by RNA gel blot analyses using the intron-specific probe as described below.

Aberrant Splicing of Intronic Sequences Occurs in Other ABPL Tumors. Most of the ABPL tumors showing aberrations in the *myb* locus have undergone proviral insertion within the same 1.5-kb segment of the intronic sequences preceding the *v-myb*-related sequences (5, 8). To verify whether splicing aberrations such as those seen in the ABPL tumor also occur in the other ABPL tumor RNAs, we prepared radioactive probes using the DNA fragment containing the intronic sequences (*EcoRI*-*Aat* II fragment) exclusively. As shown in Fig. 4, hybridization of this probe with the RNAs derived from ABPL-1, ABPL-2, and ABPL-4 tumors showed that the aberrant *myb* RNAs from the three tumors hybridize to this probe whereas little or no hybrid-

ization was observed with RNA from mouse thymus that contains high levels of normal *myb* mRNA.

DISCUSSION

To understand the mechanism of activation of the *myb* gene, we have compared the structure of the cellular and viral *myb*-coding sequences (1). These studies showed that during the generation of viral transforming genes, the *c-myb* sequences have undergone extensive deletions at both the N- and C-terminal ends. The deletion of similar stretches of N- and C-terminal sequences in two different viral isolates raised the possibility that such deletions in the coding regions could be the mechanism for conversion of the normal *c-myb* gene into its transforming counterpart.

In addition to the avian systems, which suggest that deletions are important in the activation of the *myb* gene, one other mouse model system provides additional support for this hypothesis. Cas-Br Mo-MuLV induces myeloid leukemias in NFS mice and one of the tumor lines, NFS-60, was found to contain a rearranged *myb* locus. This was found to be due to the integration of the proviral genome toward the 3' end of the *myb* gene resulting in the premature termination of *myb* gene transcription (18, 19). The *myb*-coding sequences in this aberrant mRNA terminate at a point near the end of the sixth viral-related exon that lies close to the site of deletion observed in AMV and E26 virus.

In the present communication, we have examined in detail the structure of the *c-myb*-encoded mRNAs present in the ABPL-2 tumor line. Earlier studies (7) had indicated that these tumors had undergone rearrangements in the *myb* locus, resulting in the synthesis of abnormal mRNAs. Molecular cloning and structural comparison of the normal and rearranged *c-myb* DNA sequences revealed that the rearrangements in all ABPL tumors were due to integration of the Mo-MuLV genome into 1.5 kb of cellular DNA immediately upstream from the *v-myb*-related sequence (5). A detailed characterization of the mouse *c-myb*-coding exons revealed that this integration occurred in intron sequences that precede the first viral-related exon. The results presented here demonstrate that the viral integration results in the initiation

FIG. 4. RNA gel blot analysis of ABPL-1, ABPL-2, and ABPL-4 mRNAs using an intron-specific probe. Lanes: 1, normal mouse thymus; 2, ABPL-2; 3, ABPL-1; 4, ABPL-4. The probe was a 300-bp *EcoRI*-*Aat* II fragment derived from the intron that appears in the cDNA sequence. Positions of 28S (5 kbp) and 18S (2 kbp) rRNA are as shown.

of transcription within the Mo-MuLV viral sequences and replacement of the *c-myb* N-terminal sequences by *gag* sequences. This results in the deletion of a minimum of three coding exons from the 5' end that is accompanied by the generation of aberrantly spliced mRNAs that contain additional sequences coded for by the intron that follows the sixth *v-myb*-related exon. This is particularly interesting since the deletions in the two viral transforming genes and in the NFS-60 cell line occur in this same region (1, 18). This suggests that the alterations (deletions or insertions) at the C-terminal end play a crucial role in the activation of the oncogenic potential of the ABPL-2 *myb* gene. The significance of the deletions at the N-terminal end are at present unclear. In the NFS-60 cell line, the 5' end appears to be unaffected, whereas the entire C-terminal end is deleted. This argues that the C-terminal deletions alone might be sufficient to activate this gene. However, it is possible that the deletions at the 5' end could enhance the transforming potential of the *c-myb* gene that already has an aberration at the 3' end. Fig. 5 summarizes the various alterations of the *myb* gene in these tumor model systems. With the availability of full-length cDNA clones of the *myb* gene, these questions can be addressed further by the construction of retroviruses with various mutations to test directly the affects of 5' and 3' deletions on the gene.

While this work was in progress, Shen-Ong *et al.* (19) published the nucleotide sequence of a *c-myb* cDNA clone derived from the ABPL-2 tumor line. While their results agree with ours regarding the sequences at the 5' end, they did not find the presence of intronic sequences following the sixth *v-myb*-related exon. Their results indicate normal splicing of the 3' end of the *c-myb* mRNA in this tumor line. For this reason, we have analyzed 12 different *c-myb* clones and found only the aberrantly spliced species. Furthermore, a probe derived from the intronic sequences hybridizes well with the aberrant mRNAs found in these cells. The discrepancy could be best explained by assuming that there is a small percentage of normally spliced mRNA in these cells and the clone that Shen-Ong *et al.* (19) used for sequence analysis is derived from such an mRNA. Regardless of the reason for the discrepancy, it is clear from the experimental data presented here that the ABPL-2 tumor cells contain abundant quantities of *c-myb* RNAs that have undergone alterations at both the 5' and 3' ends.

FIG. 5. Structural comparison of the *myb*-coding region of normal *c-myb* mRNA with activated forms of *myb* RNA in the AMV and the E26 virus and ABPL-2 and NFS-60 tumor cells. The hatched box in ABPL-2 represents the intronic sequences that disrupt the *myb* mRNA. Straight lines represent non-*myb* sequences. Open boxes, *myb* sequences. Dashed line, splicing event that occurs in the generation of the viral *myb* mRNA.

The results presented here also have important implications regarding the role of mRNA secondary structure in the processing of high molecular weight precursors to their final form. Potential splice-donor, (Y)₆NC ↓ AGGK, and -acceptor sites, MAG ↓ GTRAGT occur frequently along the mRNA molecule. Within the consensus donor and the acceptor sequences only the two underlined nucleotides are essentially invariant (20), whereas there is considerable fluctuation among the other residues. Therefore, the choice of which potential splicing events will predominate during mRNA formation must be governed by factors other than sequence alone. Studies using the technique of *in vitro* mutagenesis appear to suggest that the secondary structure of precursor mRNAs plays a crucial role in their splicing (refs. 21–23). Thus, deletion of upstream sequences from some of the mammalian genes has resulted in transcripts that are aberrantly spliced in regions that are distant from the site of deletion. Our results strongly support this hypothesis and provide, to our knowledge, the first *in vivo* example where the deletion of upstream sequences has resulted in splicing aberrations further downstream.

We are grateful to Drs. Tom Curran and Kay Huebner for a critical review of the manuscript and Dr. Jim Ihle for helpful suggestions. We also thank Jim Averbach for help in computer graphics.

1. Rosson, D. & Reddy, E. P. (1986) *Nature (London)* **319**, 604–606.
2. Rushlow, E. K., Lautenberger, J. A., Papas, T. S., Baluda, M. A., Perbal, B., Chirikjian, J. G. & Reddy, E. P. (1982) *Science* **216**, 1421–1423.
3. Klempnauer, K. H., Gonda, T. J. & Bishop, J. M. (1982) *Cell* **31**, 453–463.
4. Nunn, M. F., Seeburg, P. K., Moscovici, C. & Duesberg, P. H. (1983) *Nature (London)* **306**, 391–395.
5. Lavu, S. & Reddy, E. P. (1986) *Nucleic Acids Res.* **14**, 5309–5320.
6. Potter, M., Reddy, E. P. & Wivel, N. A. (1978) *Natl. Cancer Inst. Monogr.* **48**, 311.
7. Mushinski, J. F., Potter, M., Bauer, S. R. & Reddy, E. P. (1983) *Science* **220**, 795–798.
8. Shen-Ong, G., Potter, M., Mushinski, J., Lavu, S. & Reddy, E. P. (1984) *Science* **226**, 1077–1080.
9. Potter, M., Sklar, M. D. & Rowe, W. P. (1973) *Science* **182**, 592–594.
10. Premkumar, E., Potter, M., Singer, P. A. & Sklar, M. D. (1975) *Cell* **6**, 149–159.
11. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–308.
12. Mushinski, J. F., Bauer, S. R., Potter, M. & Reddy, E. P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1073–1077.
13. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4748.
14. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
15. Gonda, T. J., Gough, N. M., Dunn, A. R. & deBlaquiere, J. (1985) *EMBO J.* **4**, 2003–2008.
16. Bender, T. P. & Kuehl, W. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3204–3208.
17. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
18. Weinstein, Y., Ihle, J. N., Lavu, S. & Reddy, E. P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5010–5014.
19. Shen-Ong, G. L. C., Morse, H. C., Potter, M. & Mushinski, J. F. (1986) *Mol. Cell. Biol.* **6**, 380–392.
20. Cech, T. R. (1983) *Cell* **34**, 713–716.
21. Kihne, T., Wierenga, B., Reiser, J. & Weisman, C. (1983) *EMBO J.* **2**, 727–733.
22. Solnick, D. (1985) *Cell* **43**, 667–676.
23. Archibald, A. L., Thompson, N. A. & Kvist, S. (1986) *EMBO J.* **5**, 957–965.
24. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.