Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 1642–1647, February 1996
Medical Sciences

Intergenic splicing of *MDS1* and *EVI1* occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family

SCOTT FEARS,* CAROL MATHIEU†, NANCY ZELEZNIK-LE†, SHI HUANG‡, JANET D. ROWLEY*†, AND GIUSEPPINA NUCIFORA†§

*Department of Molecular Genetics and Cellular Biology and †Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL 60637; and ‡La Jolla Cancer Research Foundation, La Jolla, CA 92037

Contributed by Janet D. Rowley, October 27, 1995

ABSTRACT **The *EVI1* gene, located at chromosome band 3q26, is overexpressed in some myeloid leukemia patients with breakpoints either 5′ of the gene in the t(3;3)(q21;q26) or 3′ of the gene in the inv(3)(q21q26). *EVI1* is also expressed as part of a fusion transcript with the transcription factor *AML1* in the t(3;21)(q26;q22), associated with myeloid leukemia. In cells with t(3;21), additional fusion transcripts are *AML1–MDS1* and *AML1–MDS1–EVI1*. *MDS1* is located at 3q26 170–400 kb upstream (telomeric) of *EVI1* in the chromosomal region in which some of the breakpoints 5′ of *EVI1* have been mapped. *MDS1* has been identified as a single gene as well as a previously unreported exon(s) of *EVI1*. We have analyzed the relationship between *MDS1* and *EVI1* to determine whether they are two separate genes. In this report, we present evidence indicating that *MDS1* exists in normal tissues both as a unique transcript and as a normal fusion transcript with *EVI1*, with an additional 188 codons at the 5′ end of the previously reported *EVI1* open reading frame. This additional region has about 40% homology at the amino acid level with the PR domain of the retinoblastoma-interacting zinc-finger protein RIZ. These results are important in view of the fact that *EVI1* and *MDS1* are involved in leukemia associated with chromosomal translocation breakpoints in the region between these genes.**

The protooncogene *EVI1* was initially identified and described in the mouse. It is activated in murine myeloid leukemia by proviral insertion in the *evi1* common integration site (1, 2). *EVI1* is not normally expressed in hematopoietic cells. In humans, the gene can be activated in myeloid leukemias and myelodysplastic diseases by chromosomal rearrangements either 5′ of the gene in the t(3;3)(q21;q26) or 3′ of the gene in the inv(3)(q21q26) by juxtaposition of the gene to enhancer elements of the ribophorin gene located at 3q21 (3–5). Activation of *EVI1* can also occur in the t(3;21)(q26;q22) as part of the fusion mRNA, *AML1–EVI1*, that is transcribed from the der(3) chromosome (6, 7). Abnormal expression of *EVI1* has also been detected in patients with myeloid leukemia and a cytogenetically normal karyotype (8), suggesting that inappropriate activation of this gene occurs through various mechanisms.

EVI1 is a nuclear protein containing a seven-zinc-finger domain at the N-terminal end, a three-finger domain in the central part of the molecule, and an acidic domain distal to the second group of zinc fingers (9). The human and mouse open reading frames are 91% homologous at the DNA level and 94% homologous at the amino acid level. The second exon of the gene, in frame although not translated, is highly conserved between the two species. The open reading frame starts in the third exon of the gene, where the first ATG is located (9). A putative promoter has been identified, by genomic sequencing and S1 protection analysis, immediately upstream of the first exon of the murine cDNA (10).

MDS1 was cloned as one of the partner genes of *AML1* in the t(3;21)(q26;q22), associated with therapy-related acute myeloid leukemia and myelodysplastic syndrome as well as with chronic myeloid leukemia in blast crisis (11, 12). In this translocation, *AML1*, located at 21q22, is fused to several genes, *EAP*, *MDS1*, and *EVI1*, all of which are located at 3q26 200–400 kb apart (7), and chimeric cDNAs have been isolated from cells with t(3;21) in which *AML1* is fused to *EAP*, to *MDS1*, to *EVI1*, or to *MDS1* and *EVI1* in the same transcript, producing, in the latter case, a very complex chimeric gene. Only the 3′ region of *MDS1* that is fused to *AML1* has been isolated and sequenced. The nature of *MDS1* is somewhat controversial; *MDS1* has been described as a unique gene (7), and also as one or more previously unreported exons of *EVI1* (6). *MDS1* has been mapped 170–400 kb upstream of *EVI1* (7); if it were part of *EVI1*, then *EVI1* would have two promoters separated by hundreds of kilobases. The existence of genes with multiple promoters is not unusual and has been documented in other cases (11, 12).

We have investigated the relationship between *MDS1* and *EVI1* by analyzing cDNA clones isolated from normal libraries. Here we present the complete cDNA sequence of *MDS1*¶ and show that *MDS1* and *EVI1* are expressed in normal tissues as a "fusion" gene containing most of *MDS1* spliced to the second exon of *EVI1*. The new part of *EVI1* encoded by the distal region of *MDS1* and *EVI1* second exon has 40% homology to the PR domain of the retinoblastoma-binding protein RIZ. These results are important in view of the role of *EVI1* in development and in leukemogenesis and of a recent report on the tumorigenicity of the *AML1–MDS1* fusion gene (13).

MATERIALS AND METHODS

Screening cDNA Libraries and cDNA Sequencing. The pancreas cDNA library was purchased from Stratagene and was used according to the manufacturer's instructions. The kidney cDNA library, prepared in λgt10, was the generous gift of Graeme Bell (University of Chicago, Chicago). Both libraries were screened with the PCR-amplified *MDS1* probe. The probe and the primers used for PCR amplification have been described (7). The cDNA inserts were sequenced with a Sequenase kit (United States Biochemical) according to Sanger's method.

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.

§To whom reprint requests should be addressed at: Section of Hematology/Oncology, Department of Medicine, 5841 South Maryland Avenue, MC 2115, Chicago, IL 60637.

¶The sequences reported in this paper have been deposited in the GenBank data base [accession nos. U43293 (*MDS1A*) and U43292 (*MDS1B*)].

Northern and Southern Blot Analysis, Preparation of Probes, PCR Analysis, and Primers' Sequence. A genomic DNA blot from different animal species (Bios, New Haven, CT) and human multiple tissue poly(A) RNA blots (Clone-Tech, Palo Alto, CA) were used according to the manufacturer's suggestions. The *MDS1* probe and the *EVI1* probe used in Southern and Northern blot analyses have been described (7). PCR was performed as described (7). The sequences of the primers used for synthesis of the RNA probe template are as follows: upstream primer, 5′-GATCGATCGGATCCCATATTCAAGAGCCATGCTCT-3′, downstream primer, 5′-GATCGGATCCTAATACGACTCACTATAGGGCGAATTAGGGTACCGAATACAACCAAGAGTGAACG-3′. The T7 promoter sequence was included as part of the downstream primer according to the manufacturer's suggestions. The primers contained a region of 15 bp at the 5′ end (for the upstream primer) and one of 18 bp (after the T7 promoter, for the downstream primer) that did not hybridize to *MDS1*. Thus, although the homology with *MDS1* was 300 bp, the final size of the transcribed RNA probe was 333 bp. PCR was carried out for 5 cycles at an annealing temperature of 54°C, followed by 25 cycles at an annealing temperature of 63°C. After PCR, the 333-bp fragment was gel-purified and quantitated.

The sequences of the oligomers used in the PCR analysis of the cDNA clones are as follows: primer 1, 5′-ATGCTTCACTGGATGTG-3′; primer 2, 5′-TGGGAGAGCAGAGGTCAA-3′; primer 3, 5′-CCTACGTCTGAGCTTCTC-3′; primer 4, 5′-AGTGAGGAGTACTGCAT-3′.

The sequence of primer 7 used in primer extension analysis is 5′-AGTGAGGAGTACTGCA-3′.

Primer Extension Analysis and RNase Protection Analysis. For *in vitro* transcription of the RNA probe, for the RNase protection assay, and for primer extension analysis, we used a MaxiScript kit and an RPAII kit (Ambion, Austin, TX) according to the manufacturer's instructions. The DNA template for *in vitro* transcription of the RNA probe was prepared by PCR amplification as described above. After *in vitro* transcription, the RNA probe was purified by electrophoresis and elution as suggested by the manufacturer. The RNase protection assay was carried out with 2 μg of human kidney or pancreas poly(A) RNA and 10 μg of yeast RNA according to the manufacturer's suggestions. The final samples were separated by electrophoresis in a denaturing 6% polyacrylamide gel and were detected by overnight exposure to x-ray film.

For the primer extension analysis, 2 μg of kidney mRNA was annealed to the primer end-labeled with ^{32}P, and the rection was carried out according to the manufacturer' instructions. The reaction products were separated on a denaturing 6% gel and exposed to film overnight.

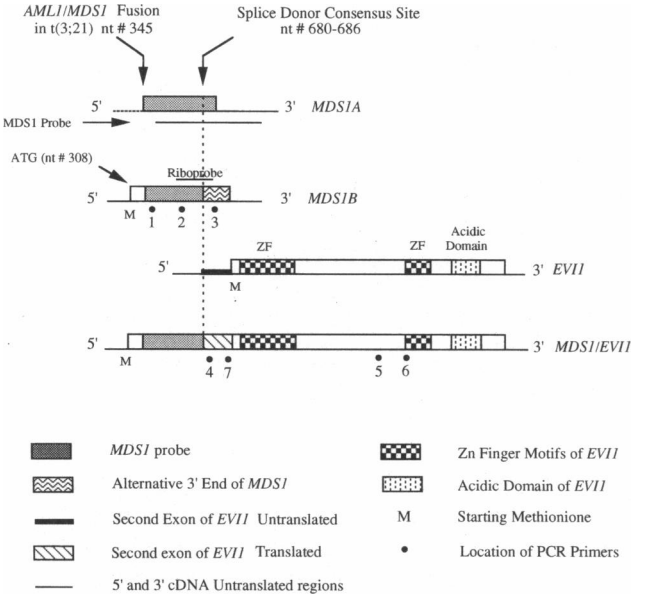

FIG. 1. Location of probes and relationship between the predicted forms of *MDS1* and *EVI1* cDNAs. The location of the *MDS1* probe, obtained by PCR amplification of *MDS1A*, and the location of the RNA probe are marked by lines. Reading frames are indicated by boxes, and 5′ or 3′ untranslated regions are indicated by thin lines. The top left arrow indicates the position of the fusion with *AML1* in the t(3;21). The sequence of *AML1* (not shown) is represented by a dotted line extending toward the 5′ end of the cDNA. The entire *MDS1A* (shaded box, top line) is contained in one exon. The top right arrow shows the position of the splice donor consensus site (nt 680–686) in *MDS1A*. The dotted vertical line indicates the splice junction site at nt 685 in the various cDNAs. The splice donor consensus (nt 680–686) is utilized for the alternative 3′ coding exon(s) of *MDS1B* corresponding to clones pHP5 and pHP6. The alternative 3′ exon(s) of *MDS1B* is shown by a box with a wavy pattern. The splice consensus site at nt 680–686 is also utilized for the second exon of *EVI1* corresponding to clones pHP2 and pHP7. The second exon of *EVI1* is shown as a thick black line. The diagram of *EVI1* is deduced from the sequence reported by Morishita *et al.* (9). The lower diagram shows the predicted *MDS1–EVI1*. The size of the cDNAs is not drawn to scale. The numbered dots correspond to the PCR primers used for the analysis of the cDNA clones. ZF, zinc finger domain.

RESULTS

Interspecies Conservation of *MDS1*. The *MDS1* probe (Fig. 1) strongly hybridized to the various species DNA (Fig. 2), indicating that the probe is highly conserved and is likely to be an exon(s). Only one band was detected by the probe in all cases, except for dog DNA (Fig. 2, lane 11). By analogy to the Southern blot patterns of other species, it is possible that the region that hybridizes to canine DNA is also one exon containing an *Eco*RI restriction site. PCR of genomic human DNA with primers flanking the probe amplified a fragment of 1.5 kb, confirming that, in humans, the probe is contained in one exon (data not shown).

Identification and Analysis of *MDS1* Clones from the two cDNA Libraries. *Analysis of the 5′ ends of the pancreas library clones.* The *MDS1* probe that we used to screen the libraries was obtained by PCR amplification of a chimeric cDNA

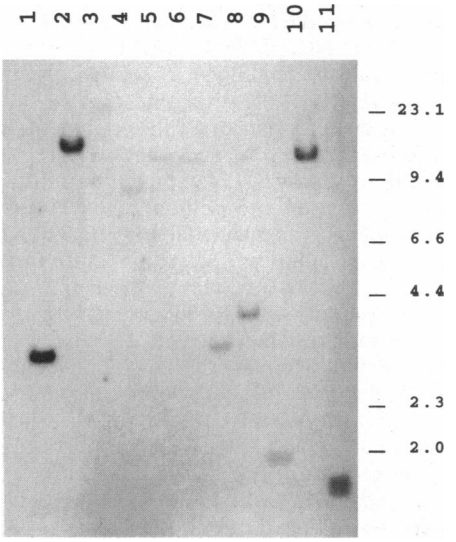

FIG. 2. Demonstration of evolutionary relatedness of *MDS1* by Southern blot analysis. Genomic DNA isolated from various species was digested with the restriction enzyme *Eco*RI, separated on an agarose gel, transferred to a nylon membrane, and hybridized to the *MDS1* probe. The origin of the DNAs is as follows. Lanes: 1, human; 2, marmoset; 3, pig; 4, hamster; 5, rat; 6, mouse; 7, sheep; 8, cow; 9, rabbit; 10, cat; 11, dog. Molecular size in kb is shown to the right.

isolated from a t(3;21) patient (7). The two genes involved in the chimeric transcript are *AML1* and *MDS1*. The 1.5-kb region downstream of the chimeric junction with *AML1* corresponds to the distal part of an *MDS1* transcript. The 5′ end of *MDS1A* had not been cloned. The 1.5-kb region represents a 3′ end of *MDS1*, and it is indicated as *MDS1A* in Fig. 1 and 3. We used the *MDS1* probe (Fig. 1) to screen 5×10^5 plaques from the λ Zap cDNA pancreas library and 5×10^5 plaques from the λgt10 cDNA kidney library. We isolated and purified seven cDNA clones from the pancreas library and 17 cDNA clones from the kidney library. Restriction digestion pattern of the pancreas clones, pHP1 to pHP7, showed that two of them, pHP4 and pHP6, were identical and that one of the clones, pHP3, contained multiple inserts (data not shown). Thus, five of the clones (pHP1, pHP2, pHP5, pHP6, and pHP7) were unique and were studied further and sequenced. They had novel DNA sequences extending at the 5′ end of *MDS1A* and ranging in size from 0.1 to 0.4 kb. Four of them were identical. This region (nt 1–240, Fig. 3) was very purine-rich and contained a stretch of 25 GA repeats. Downstream of the GA repeats, we identified an open reading frame starting with an ATG codon (nt 308, underlined in Fig. 3) located 37 nt upstream of the 5′ end of *MDS1A*. We concluded that this region represented the 5′ untranslated and the first translated exon(s) of the normal *MDS1* gene. The region is shown in Fig. 1, in which the 5′ untranslated exon(s) is indicated by a line, and the first translated exon(s) is indicated by an empty box. Two G→C changes (nt 225 and 237) were detected in the untranslated region, and they are indicated in Fig. 3. The changes could be the result of artifacts introduced during construction of the library. The remaining pHP clone, pHP1, had ≈100 bp of sequence upstream of the 5′ end of *MDS1A*. This sequence diverged from that of the other pHP clones (data not shown) and did not detect any band on multiple-tissue Northern blots. It could represent an incompletely processed *MDS1* RNA and it was not studied further.

Analysis of the 3′ ends of the pancreas library clones. Based on the size of the inserts, the pHP cDNA clones were divided into two groups, one with inserts between 1.3 and 1.4 kb (pHP5 and pHP6) and the second group with insert of 4–4.3 kb (pHP1, pHP2, and pHP7). All five pHP clones were completely sequenced. pHP5 and pHP6 were identical and diverged sharply from *MDS1A*, as well as from pHP1, pHP2, and pHP7 at nt 682. Fig. 3 shows the DNA sequence alignment of *MDS1A* and pHP6. After nt 682, pHP5 and pHP6 continued for ≈660 nt with no significant homology to sequences deposited in GenBank. The region of the divergence, between nt 680 and 686 of *MDS1A*, matches that of a splice donor site (indicated in boldface in Fig. 3). We concluded that pHP6 represented an alternative splice form of *MDS1* obtained by utilization of the splice donor consensus site with another exon(s) located farther downstream, and we called it *MDS1B*. The sequence of pHP5 was a few nucleotides shorter than that of pHP6 at the 5′ and 3′ ends. *MDS1B* encodes a 169-amino acid mostly hydrophilic polypeptide with a predicted mass of 19 kDa, with 26% acidic and basic residues, and with 20% proline and serine residues. Between nt 346 and 682, *MDS1B* is identical to 341 nt of the *MDS1A* (Fig. 3), which is defined at the 5′ end by the fusion junction with *AML1* detected in the *AML1–MDS1* (arrowhead in Fig. 3), and at the 3′ end by a splice consensus site (Fig. 3, nt 680–686 in bold face type). The sequence of *MDS1B* and *MDS1A* downstream of the splice consensus site diverged. *MDS1B* reading frame continued with 44 codons and *MDS1A* reading frame for 14 codons for before termination (stop codons underlined in Fig. 3). Downstream of the stop codons, the 3′ untranslated sequences of *MDS1B* and *MDS1A* were AT-rich and contained putative signals for selective destabilization in both of them (ATTTA, doubly underlined in Fig. 3), a consensus polyadenylylation signal was detected for *MDS1A* (AATAAA, doubly underlined in Fig. 3) and potential polyadenylylation signals were detected for *MDS1B* (CATAAA and TATAAA, doubly underlined, Fig. 3).

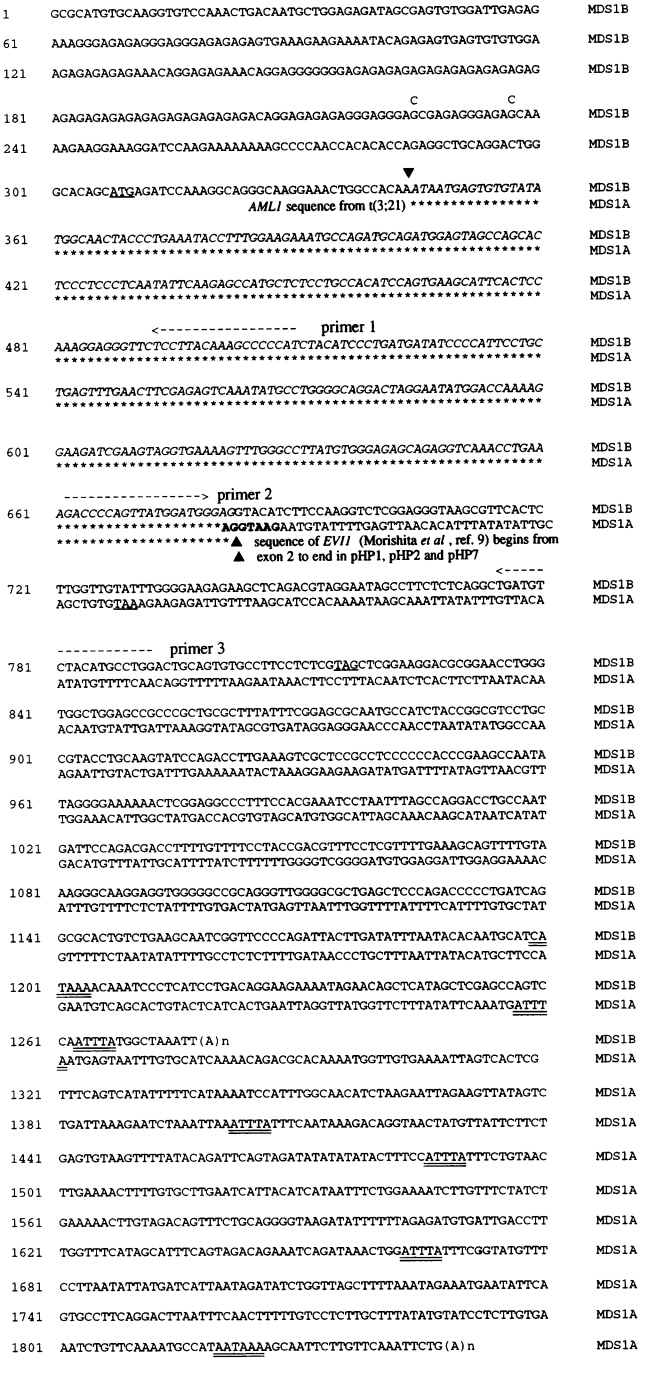

FIG. 3. Alignment of the nucleotide sequences of *MDS1A* and *MDS1B*. Asterisks indicate identities. The single arrowhead at nt 345 indicates the beginning of the *MDS1A* sequence and was obtained by sequencing of the 3′ end of the *AML1–MDS1* fusion gene (7). The *MDS1B* sequence was obtained by sequencing cDNA clones isolated from the pancreas cDNA libraries. The two arrowheads point to the splice junction at nt 680–686, in boldface type, that is used in *MDS1B*. The splice site is also utilized to join *MDS1* with *EVI1*. The sequence of *EVI1* after the splice junction starts with the second exon of *EVI1* as reported by Morishita *et al.* (9) and continues to the end of Morishita's *EVI1* cDNA. The region between the single and double arrowheads, indicated in italic type, is common to *MDS1A*, *MDS1B*, and *MDS1–EVI1*. The starting ATG of *MDS1B* and *MDS1–EVI1* at nt 308 is underlined. The two stop codons that terminate translation of *MDS1A* (nt 728) and *MDS1B* (nt 815) are also underlined. Several putative signals for selective destabilization, as well as consensus and potential signals for polyadenylylation, are doubly underlined.

The sequence of the longer clones (pHP1, pHP2, and pHP7) also diverged from *MDS1A* after nt 682 (splice donor site) and continued for ≈2.4 kb. After nt 682, *MDS1* was fused in-frame to the second exon of *EVI1* (Figs. 1 and 3). Downstream of the splice donor site, pHP2 and pHP7 were identical to the sequence of *EVI1* cDNA (9) except for nine base-pair differences, as was detected by Mitani *et al.* (6) in their *EVI1* sequence. The sequence of pHP2 and pHP7 encoded a polypeptide of 1240 residues. This alternative form of *EVI1* contains an additional 188 residues at the N-terminal end, 125 of which are encoded by *MDS1*, and 63 are encoded by the second and part of the third exons of *EVI1*. pHP1 had a deletion from nt 2194 to 2308 of the published sequences (data not shown). The deletion introduced a frame shift in the reading frame and a stop codon upstream of the second group of zinc fingers or the acidic domain. Interestingly, the 3′ end of the sequence of pHP1, pHP2, and pHP7, as well as the two *EVI1* cDNAs reported by Morishita *et al.* (9) and by Mitani *et al.* (6), terminated with the exactly same nucleotide. A diagram of the various cDNAs is shown in Fig. 1.

Analysis of the kidney library clones. PCR analysis of 10 of the 17 phage clones isolated from the human kidney library (HK1 to HK10) and one strand sequencing of the fragments PCR-amplified with primer 1 and a vector-specific primer showed that the 5′ end of the 10 clones contained the purine-rich region and GA repeats as seen for pHP6. The 3′ region of the phages was analyzed by PCR, with either primers 2 and 3, specific for *MDS1B*, or primers 2 and 4, which amplify the *MDS1–EVI1* junction (Figs. 1 and 3). The size of the separated reaction products were compared to those obtained by PCR with pHP2 and pHP6 used as controls. Three of the 10 phage clones corresponded to pHP6, whereas the 7 remaining clones corresponded to pHP2. To confirm that the 7 clones contained *EVI1* sequence and to determine whether they had a deletion at the 3′ end similar to that detected in pHP1, we analyzed them by PCR with primers 5 and 6, flanking the deletion observed in pHP1 (ref. 3 and Fig. 1). A band corresponding to the size of the normal non-deleted sequence was noted (results not shown). Thus, we concluded that 3 of the clones contained *MDS1* only, and the remaining 7 clones contained *MDS1–EVI1* (results not shown). We did not isolate any clone corresponding to *MDS1A* in either of the cDNA libraries that we screened, indicating that the spliced form *MDS1A* isolated from a leukemic patient with a t(3;21) is not the one preferentially expressed in kidney or pancreas. The kidney clones were not analyzed further.

Expression of *MDS1* and *EVI1* in Normal Human Tissues. We analyzed three commercially prepared multiple-human-tissue Northern blots with probes specific for the two genes. By using an *EVI1* probe, three major bands of 6.5, 5.8, and 5 kb were identified in two of the blots (Fig. 4*A*). This pattern and the size of bands has been reported (9) for *EVI1*. By using the *MDS1* probe, a major band of 6.5 kb and a much fainter band of ≈5.8 kb were also detected in both blots after a longer exposure (Fig. 4*B Upper*). We assigned the larger band identified by both probes to the *MDS1–EVI1* transcript. In addition, the *MDS1* probe identified more intense unique bands of approximately 2, 1.5, and 1 kb after shorter exposure time (Fig. 4*B Lower*). We assigned the 2-kb band to the *MDS1A* transcript and the 1.5-kb band to the *MDS1B* transcript. The smallest 1-kb band could represent an additional isoform of *MDS1*. The *EVI1* probe identified a unique transcript of ≈5 kb in addition to the 6.5-kb and 5.8-kb bands. These three bands have been previously seen by other groups. The results for the third blot differed from those of the other two blots: they showed only one band corresponding to a transcript of ≈6.5 kb after hybridization to the *EVI1* probe, and no smaller transcript with the *MDS1* probe. We have no explanation for the different results obtained with the third blot, other than that perhaps the quality of the mRNA was not satisfactory.

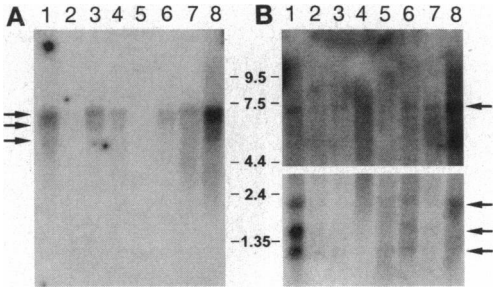

FIG. 4. Pattern of expression of *MDS1* and *EVI1* in various human tissues. Lanes: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas. The arrows indicate the position of the major bands. Numbers are size markers in kb. (*A*) *EVI1* probe. (*B*) *MDS1* probe.

RNase Protection Analysis and Primer Extension Analysis. The location of the RNA probe used for the protection assay is shown in Fig. 1. Two types of protected fragments are expected after treatment of the probe–mRNA duplex with RNase. If the probe hybridizes to *MDS1B* mRNA, the protected fragment will be 300 bp; alternatively, if the probe hybridizes to *MDS1–EVI1* mRNA, the protected fragment will be 250 bp. Although none of the cDNA clones that we analyzed from the kidney library contained the splice variant *MDS1A*, it is possible that this transcript exists at low levels in normal kidney, and the protected fragment would also be 250 bp long. The results of the assay are shown in Fig. 5. As expected, two bands of 300 and 250 bp were detected in the lane containing kidney mRNA (lane 3), but not in the lane containing yeast RNA (lane 2). Comparison of the intensity of the protected bands indicates that the transcripts are present in normal kidney approximately at the same level. Several identical minor bands smaller than 250 bp were detected in lanes 2 and 3, and they are probably due to nonspecific hybridization of the probe to the yeast RNA used as carrier in the assay. To confirm that different types of *EVI1* mRNA exist that may contain sequences 5′ to those previously reported, we hybridized kidney mRNA to primer 7 (third exon of *EVI1*) end-labeled with ^{32}P and extended *in vitro* with reverse transcriptase. The autoradiogram of the reaction products separated on a 6% sequencing gel is shown in Fig. 6. Four major bands of approximately 0.25, 0.45, 1, and 1.6 kb were detected. The relative intensity of the bands suggests that the most abundant transcript

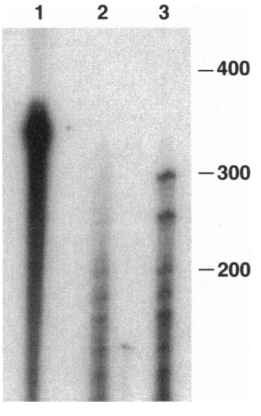

FIG. 5. RNA probe protection experiments. RNA probe spanning the region indicated in Fig. 1 was prepared, annealed to 10 μg of yeast RNA (lane 2) or 10 μg of yeast RNA plus 2 μg of human kidney mRNA (lane 3), and treated with RNase. Untreated probe was loaded in lane 1. Molecular size markers are shown at the right in nt. Two bands of 250 and 300 nt corresponding to protected mRNA fragments from *MDS1–EVI1* (250 nt) or *MDS1B* (300 nt) mRNA are shown in lane 3. Identical bands of smaller size observed in both lanes 2 and 3 are probably due to nonspecific hybridization of the probe to yeast RNA.

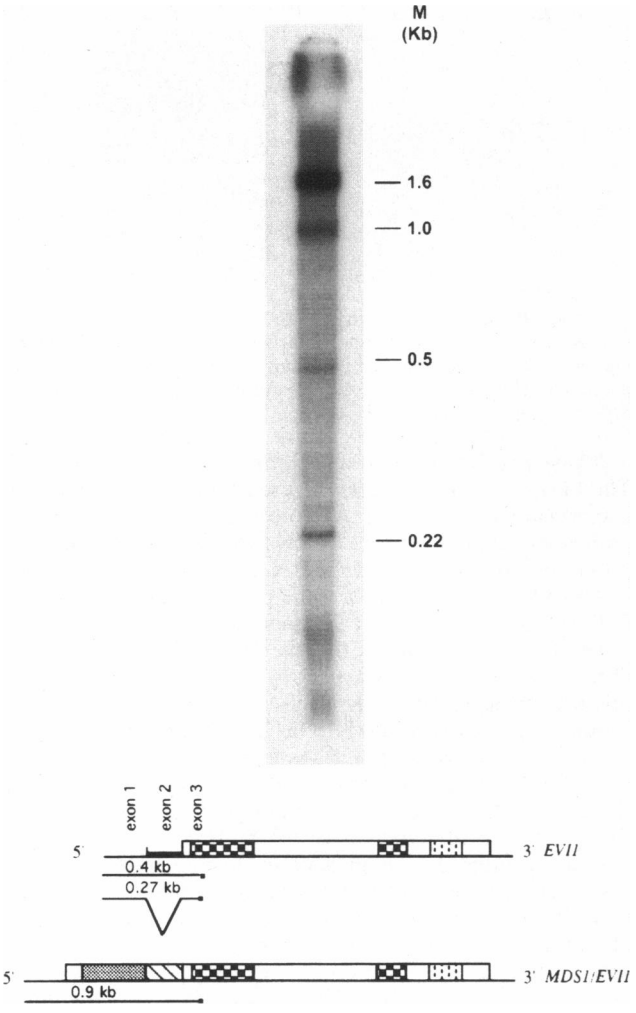

FIG. 6. Separation of primer extension products. (*Upper*) The labeled primer hybridized to exon 3 of *EVI1*. (*Lower*) Three expected products of the extension reaction. If the primer hybridizes to *EVI1* mRNA, two fragments are expected of 0.27 kb, in which exon 2 is not included, and of 0.4 kb, in which exon 2 is included. If the primer hybridizes to *MDS1–EVI1*, the size of the fragment would be 0.9 kb.

extending upstream from the third exon is the largest one of ≈1.6 kb. This band is a doublet, and its nature is still under investigation. We have assigned the three remaining bands to fragments shown in Fig. 6 *Lower*. Whereas the 0.27-, 0.9-, and 1.6-kb bands seem compact, the 0.4-kb band appears diffused, suggesting that there could be several minor sites where initiation of transcription occurs. Of the three assigned fragments, the one of 0.9 kb seems to be the most abundant, and this form of *EVI1* could be preferentially transcribed in the kidney.

DISCUSSION

In this report, we have described the sequence of two isoforms of a new gene, which we named *MDS1*. The two isoforms, *MDS1A* and *MDS1B*, differ in alternative 3′ translated and untranslated regions. Both 3′ untranslated regions include consensus and potential polyadenylylation signals and several putative signals for selective destabilization (doubly underlined in Fig. 3). We isolated *MDS1B* from pancreas and kidney libraries, whereas *MDS1A* was isolated only as the 3′ end of a chimeric message from the cDNA library of a patient with the t(3;21), but not as a normal cDNA from either the pancreas or kidney libraries. *MDS1A* is expressed as one exon. Within the open reading frame of *MDS1A*, we identified a splice donor site that yields *MDS1B*. As a consequence of splicing, *MDS1A* and *MDS1B* share 341 nt of the coding region (Fig. 3). Results of RNA probe protection assay and Northern blot analysis confirm that the *MDS1B* message exists in normal kidney (Figs. 4 and 5). Because of these results, i.e., identification of 3′ untranslated regions containing polyadenylylation and destabilization signals, and results of RNA probe protection experiments and Northern blot analysis, we believe that *MDS1* is a unique gene that is expressed in selected human tissues.

Sequencing of our longer cDNA clones shows that the open reading frame of *MDS1* can be expressed as a splice variant fused with the second exon of *EVI1*. Thus, *MDS1* and *EVI1* can be expressed as separate genes with no common coding exons or, alternatively, their open reading frame can be spliced together to form a single message encoding most of *MDS1* fused to the second exon of *EVI1*. Primer extension analysis, RNase protection assays, and Northern blot analysis confirm these results. The need to translate the conserved second exon of *EVI1* (considered by some groups as noncoding, ref. 6), in the fusion with *MDS1*, could explain the high sequence

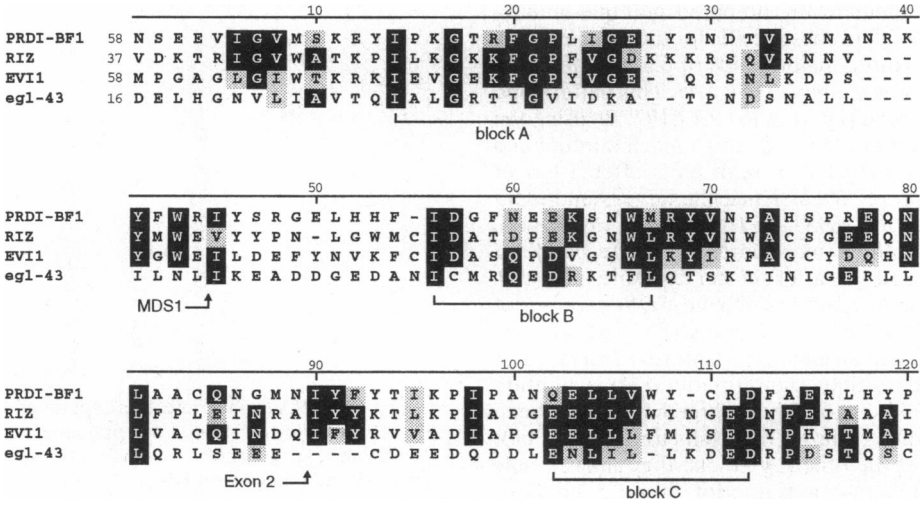

FIG. 7. Alignment of PR domain sequences. Only human sequences are presented. Identical (black background) or similar (shaded background) residues found at a position at least three times are shaded. Dashes indicate sequence gaps. Conserved blocks A, B, and C are underlined. The arrows mark the boundaries between the second exon of *EVI1* with *MDS1* and with the third exon of *EVI1*.

conservation of the second exon between human and murine cDNAs.

During revision of the manuscript, a GenBank homology search showed that the predicted translation product of the distal part of *MDS1* combined to the second exon of *EVI1* is 40% homologous to the amino end of the recently cloned retinoblastoma binding protein RIZ (14) and, to a lesser extent, to the PR domain of the transcription repressor PRDI-BF1/Blimp-1, that can drive B-cell differentiation and to the *Caenorhabditis elegans* zinc-finger protein egl-43. egl-43 has also extensive homology to EVI1 in the zinc-finger regions, suggesting that overall the two proteins are evolutionarily conserved. Analysis of the conserved residues in the PR domain (Fig. 7) showed that they can be divided into three blocks, A, B, and C. These blocks are of ≈12 amino acids and are evenly separated by a stretch of less conserved sequences of ≈30 amino acids. The fact that the homology region spans *MDS1* and the first part of *EVI1* supports our hypothesis that the two genes can also encode one single protein. It is of interest to note that in the t(3;3) and t(3;21), the chromosomal breakpoints occur between *MDS1* and *EVI1*, thus separating the PR domain from *EVI1*. The role of the three genes in normal cells are currently analyzed in our laboratory. Clearly much work needs to be done to determine how alterations in these two gene contribute to leukemia.

This work was supported by The G. Harold and Leila Y. Mathers Charitable Foundation (J.D.R.); by Department of Energy Grant DE-FG02-86ER60408 (J.D.R.); National Institutes of Health Grants CA 42557 (J.D.R.), CA 67189 (G.N.), and CA 57496 (S.H.); and the Pew Scholars Program in Biomedical Sciences (S.H.). G.N. is Special Fellow of the Leukemia Society of America. S.F. was supported by National Institutes of Health Graduate Training Grant HD-07009.

1. Morishita, K., Parker, D. S., Mucenski, M. L., Copeland, N. G. & Ihle, J. N. (1988) *Cell* **54,** 831–840.
2. Mucenski, M. L., Taylor, B. A., Ihle, J. N., Hartley, J. W., Morse, H. C., Jenkins, N. A. & Copeland, N. G. (1988) *Mol. Cell. Biol.* **8,** 301–308.
3. Morishita, K., Parganas, E., Willman, C. L., Whittaker, M. H., Drabkin, H. Oval, J., Taetle, R., Valentine, M. B. & Ihle, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3937–3941.
4. Fichelson, S., Dreyfus, F., Berger, R., Melle, J., Bastard, C., Miclea, J. & Gisselbrecht, S. (1992) *Leukemia* **6,** 93–99.
5. Suzukawa, K., Parganas, E., Gajjar, A., Abe, T., Takahashi, S., Tani, K., Asano, S., Asou, H., Kamada, N., Morishita, K. & Ihle, J. N. (1994) *Blood* **84,** 2681–2688.
6. Mitani, K., Ogawa, S., Tanaka, T., Miyoshi, H., Kurokawa, M. K., Mano, H., Yazaki, Y., Ohki, M. & Hirai, H. (1994) *EMBO J.* **13,** 504–510.
7. Nucifora, G., Begy, C. R., Kubayashi, H., Claxton, D., Pedersen-Bjergaard, J., Parganas, E., Ihle, J. & Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4004–4008.
8. Russel, M., List, A., Greenberg, P., Woodward, S., Glinsmann, B., Parganas, E., Ihle, J. N. & Taetle, R. (1994) *Blood* **84,** 1242–1248.
9. Morishita, K., Parganas, E., Douglass, E. C. & Ihle, J. N. (1990) *Oncogene* **5,** 963–971.
10. Bartholomew, C. & Ihle, J. N. (1991) *Mol. Cell. Biol.* **11,** 1820–1828.
11. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. & Leder, P. (1983) *Cell* **34,** 779–789.
12. Burnett, L. C., David, J. C., Harden, A. M., LeBeau, M. M., Rowley, J. D. & Diaz, M. O. (1991) *Genes Chromosomes Cancer* **3,** 461–467.
13. Zent, C., Kim, N., Hiebert, S., Zhang, D.-E., Tenen, D. G., Rowley, J. D. & Nucifora, G., *Top. Microbiol. Immunol.*, in press.
14. Buyse, I. M., Shao, G. & Huang, S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 4467–4471.