JOURNAL OF VIROLOGY, Sept. 2002, p. 9046–9059 Vol. 76, No. 18
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.18.9046–9059.2002

Ahi-1, a Novel Gene Encoding a Modular Protein with WD40-Repeat and SH3 Domains, Is Targeted by the *Ahi-1* and *Mis-2* Provirus Integrations

Xiaoyan Jiang,[1] Zaher Hanna,[1,2] Mohammadi Kaouass,[1] Luc Girard,[1] and Paul Jolicoeur[1,3,4]*

Laboratory of Molecular Biology, Clinical Research Institute of Montreal, Montreal, H2W 1R7 Quebéc,[1] Département de Médecine[2] and Département de Microbiologie et d'Immunologie,[3] Université de Montréal, Montréal, H3C 3J7 Quebéc, and Department of Medicine, Division of Experimental Medicine, McGill University, Montreal, H3G 1A4 Quebéc,[4] Canada

Received 21 March 2002/Accepted 13 June 2002

The *Ahi-1* locus was initially identified as a common helper provirus integration site in Abelson pre-B-cell lymphomas and shown to be closely linked to the c-*myb* proto-oncogene. Since no significant alteration of c-*myb* expression was found in Abelson murine leukemia virus-induced pre-B-lymphomas harboring a provirus inserted within the *Ahi-1* locus, this suggested that it harbors another gene whose dysregulation is involved in tumor formation. Here we report the identification of a novel gene (*Ahi-1*) targeted by these provirus insertional mutations and the cloning of its cDNA. The Ahi-1 proviral insertions were found at the 3′ end of the gene, in an inverse transcriptional orientation, with most of them located around and downstream of the last exon, whereas another insertion was within intron 22. In addition, another previously identified provirus insertion site, Mis-2, was found to map within the 16th intron of the *Ahi-1* gene. The *Ahi-1* cDNA encodes a 1,047-amino-acid protein. The predicted Ahi-1 protein is a modular protein that contains one SH3 motif and seven WD40 repeats. The *Ahi-1* gene is conserved in mammals and encodes two major RNA species of 5 and 4.2 kb and several other shorter splicing variants. The *Ahi-1* gene is expressed in mouse embryos and in several organs of the mouse and rat, notably at high levels in the brain and testes. In tumor cells harboring insertional mutations in *Ahi-1*, truncated *Ahi-1*/viral fused transcripts were identified, including some splicing variants with deletion of the SH3 domain. Therefore, *Ahi-1* is a novel gene targeted by provirus insertion and encoding a protein that exhibits several features of a signaling molecule. Thus, *Ahi-1* may play an important role in signal transduction in normal cells and may be involved in tumor development, possibly in cooperation with other oncogenes (such as v-*abl* and c-*myc*) or with a tumor suppressor gene (*Nf1*), since Ahi-1 insertion sites were identified in tumors harboring v-*abl* defective retroviruses or a c-*myc* transgene or in tumors exhibiting deletion of *Nf1*.

The Abelson murine leukemia virus (A-MuLV) is a defective, replication-incompetent retrovirus. To replicate in vivo and in vitro, A-MuLV requires a nondefective helper MuLV (51). The Abelson virus complex is highly lymphomagenic and induces oligoclonal pre-B-cell lymphoma in susceptible mice (1, 3, 42, 45). However, the expression of the v-*abl* oncogene in target cells does not appear to be sufficient for tumor formation (17, 50) and additional genetic events are thought to be required. We have reported that the long terminal repeat (LTR) of the helper Moloney MuLV is involved in the development of lymphoma induced by A-MuLV (50). In fact, the role of the Moloney provirus as an insertional mutagen was demonstrated by the identification of a common helper provirus integration site, designated *Ahi-1* (for Abelson helper integration site 1), in 16% of A-MuLV-induced pre-B-cell lymphomas (41). All Moloney proviruses for which the precise integration site within this region could be mapped were found to be in the same orientation, suggesting a specific molecular requirement for this genetic event to be involved in malignant transformation in the presence of v-*abl*. *Ahi-1* was mapped to mouse chromosome 10 in close linkage to the c-*myb* proto-oncogene (41). By using long-range restriction mapping, we mapped the c-*myb* and *Ahi-1* regions within a 120-kbp DNA fragment, with the *Ahi-1* locus being located ca. 35 kbp downstream of c-*myb* (22). Interestingly, the *Mis-2* region, which represents another MuLV integration site, has also been mapped in the same region, ca. 160 kbp downstream of c-*myb* and ca. 120 kbp downstream of *Ahi-1* (62). We tested whether Moloney provirus integration in the *Ahi-1* locus enhances the expression of c-*myb* by a *cis*-acting mechanism, by examining c-*myb* gene expression in A-MuLV-induced pre-B-lymphomas (22). Our data did not reveal any clear evidence for such c-*myb* overexpression in the tumors which were tested.

In view of these findings, we have hypothesized that a novel gene within the Ahi-1 locus is targeted by provirus insertion and that its dysregulation contributes to tumor development. We report here the identification of that gene, *Ahi-1*, which is the target of these Ahi-1 and Mis-2 provirus insertions and which encodes a protein containing an SH3 motif and a WD40-repeat domain.

* Corresponding author. Mailing address: Laboratory of Molecular Biology, Clinical Research Institute of Montreal, 110 Pine Ave., West, Montreal, H2W 1R7 Quebec, Canada. Phone: (514) 987-5569. Fax: (514) 987-5794. E-mail: jolicop@ircm.qc.ca.

MATERIALS AND METHODS

Mice and viruses. MMTV^{D}/myc transgenic (Tg) mice have been described previously (36). They were bred on CD1 background. Newborn (<48 h) mice

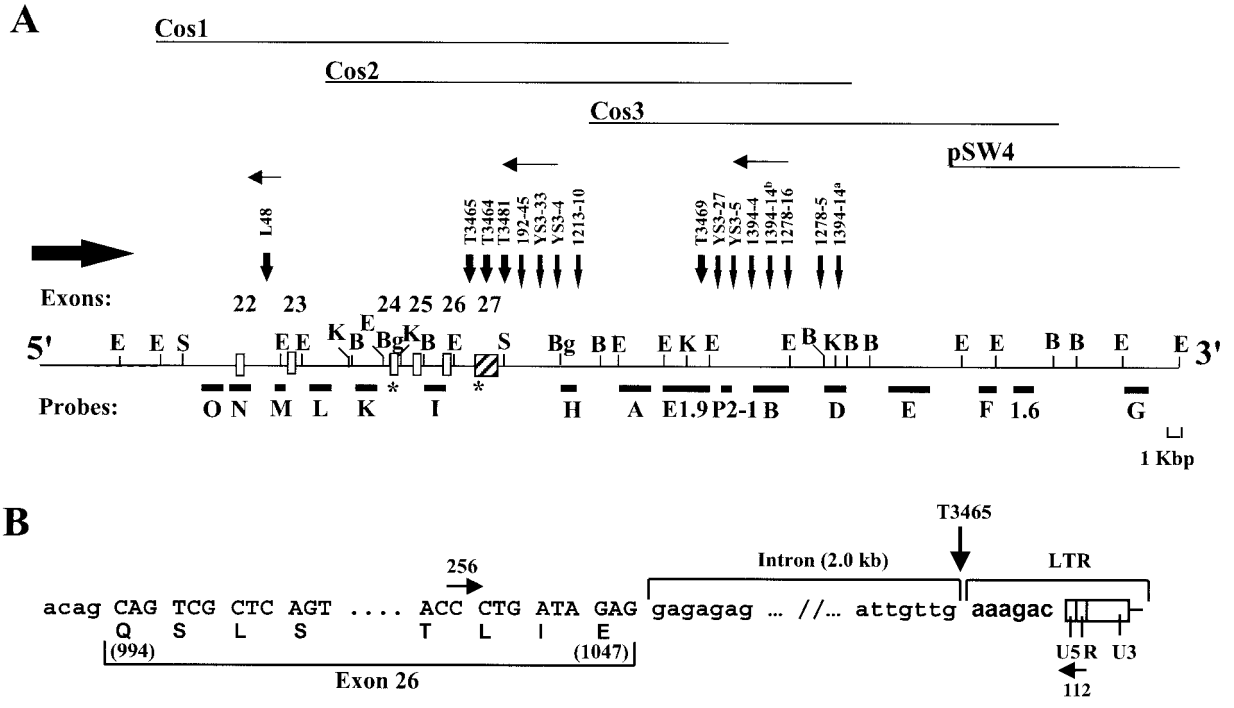

FIG. 1. Mapping of the MuLV proviruses within the *Ahi-1* locus. (A) Structural organization of the proviruses integrated at the 3′ end of the *Ahi-1* gene. The restriction endonuclease map of 80 kbp of the *Ahi-1* region was determined with phage (pSW4) and cosmid (Cos1, Cos2, and Cos3) clones and confirmed by Southern blot analysis of mouse DNA. The localization of the proviruses relative to the exons is shown. The intron-exon boundaries of exons 25 and 26 have been confirmed by sequencing genomic fragments. The provirus integration within the rat *Ahi-1* locus in tumor 1213-10 was detected with probe H but could not be mapped precisely because of the lack of restriction map in the rat. Symbols: open and strippled boxes, coding and noncoding exons, respectively; line, intronic sequences; closed horizontal boxes, single-copy fragments subcloned and used as probes; thin vertical arrows, sites of provirus integration in A-MuLV-induced pre-B-lymphomas; large vertical arrows, sites of provirus integration in T-cell thymomas of Mo-MuLV-infected $MMTV^D$/myc Tg mice; horizontal thin arrows, orientation of provirus (5′ to 3′); horizontal thick arrow, orientation of *Ahi-1* transcription; asterisks, stop codons. Restriction sites: B, *Bam*HI; Bg, BgII; E, *Eco*RI; K, *Kpn*I; S, *Sma*I. Not all sites are indicated. (B) Site of the provirus integration within the *Ahi-1* gene in tumor T3465. Sequences were obtained from a PCR product (~2.0 kbp) generated with primer 256 localized in exon 26 and primer 112 within the Moloney MuLV LTR. The sequences of the intron-exon boundaries are shown, respectively, in small and capital letters and the viral sequences in boldface. Encoded amino acids are shown in the lower line and the numbers refer to the amino acid position in the sequence shown in Fig. 3.

were inoculated with Moloney MuLV (10^5 PFU/ml) intraperitoneally, as described previously (41).

DNA extraction and restriction endonuclease digestion. DNA extraction, digestion with restriction endonucleases, separation of DNA fragments by agarose gel electrophoresis, and hybridization with ^{32}P-labeled probes by the method of Southern were performed as described previously (41).

Molecular cloning of mouse genomic DNA. The molecular cloning of the provirus with the adjacent *Ahi-1* sequences has been described elsewhere (41). To initiate chromosome walking, cosmid and phage libraries were constructed by partial digestion of C57BL/6 mouse liver DNA with *Sau*3A and ligation into the *Bam*HI site of the pWE15 and EMBL-3 vector, respectively, as described elsewhere (41, 48, 62). Molecular clones (Cos1, Cos2, Cos3, and pSW4) were isolated by screening these libraries with mouse *Ahi-1* probes E1.9 and G (Fig. 1A). To analyze these clones in more details and to derive additional probes necessary for a finer mapping of this region, a restriction map of the *Ahi-1* region was derived and additional single-copy fragments were isolated and subcloned (Fig. 1A). This map was confirmed by digesting normal mouse cellular DNA with various restriction endonucleases (P.L. Pharmacia, Montreal, Quebec, Canada) and hybridizing them with different probes derived from this region (data not shown), essentially as previously described (41, 48).

Exon amplification. Exon amplification to identify coding sequences in genomic DNA was performed as described by Buckler et al. (4). In brief, two cosmid clones (Cos1 and Cos3), representing 70 kbp of *Ahi-1* region around the integration sites, were digested with *Bam*HI and *Bgl*II and subcloned as pools into the vector pSPL1. In addition, another 14 individual *Bam*HI and *Bgl*II fragments derived from this region were cloned into pSPL1. Clones with inserts in sense and antisense orientations were selected and transfected into Cos-7 cells, and resulting RNAs were analyzed by reverse transcription-PCR (RT-PCR) with specific primers derived from the vector as described previously (4). Finally, amplified exons were purified by electrophoresis on 1% agarose gels and cloned into a dT-tailed (31) *Eco*RV-digested pBluescript II KS(+) plasmid vector (Stratagene). Several criteria were used to distinguish real exons from false-positive clones. These criteria included nucleotide sequencing and comparison to DNA sequences in the databases (that is, the National Center for Biotechnology Information [NCBI] and European Bioinformatics Institutes [EBI] databases), detection of RNA transcripts by Northern blot analysis, and screening of a library to identify positive cDNA clones.

Isolation of cDNA clones and sequencing. A rat brain λgt11 cDNA library was screened by using the RT-PCR fragment A (Fig. 2B). Two other mouse testis and thymus cDNA (the Uni-Zap XR library [Stratagene] libraries and the 5′ stretch library [Clontech]) were also screened with cDNA probe E1.3 in order to get a full-length cDNA. Furthermore, the missing 5′ end of the cDNA was cloned by 5′ RACE (rapid amplification of cDNA ends) as described previously (16) into a dT-tailed vector. In addition, mouse brain RNA was used to amplify mouse cDNA sequences by RT-PCR with specific primers (that is, 276 and 296) derived from the rat cDNA sequences (Fig. 2A). Several clones were isolated: seven overlapping fragments from the rat brain cDNA library, eight overlapping clones, and five full-length clones from the mouse testis and thymic cDNA libraries, respectively. These clones were subcloned into pGEM-3, analyzed with restriction endonucleases, and sequenced by the dideoxynucleotide Sanger's method (49) in both directions.

Probes. All probes containing cDNA or genomic fragments derived from the *Ahi-1* region were excised from the vector with appropriate restriction endonucleases, and the inserts were separated from the vector by agarose gel electrophoresis. The DNA fragments were recovered by electroelution and ethanol precipitation and used directly as hybridization probes. Probes were labeled by

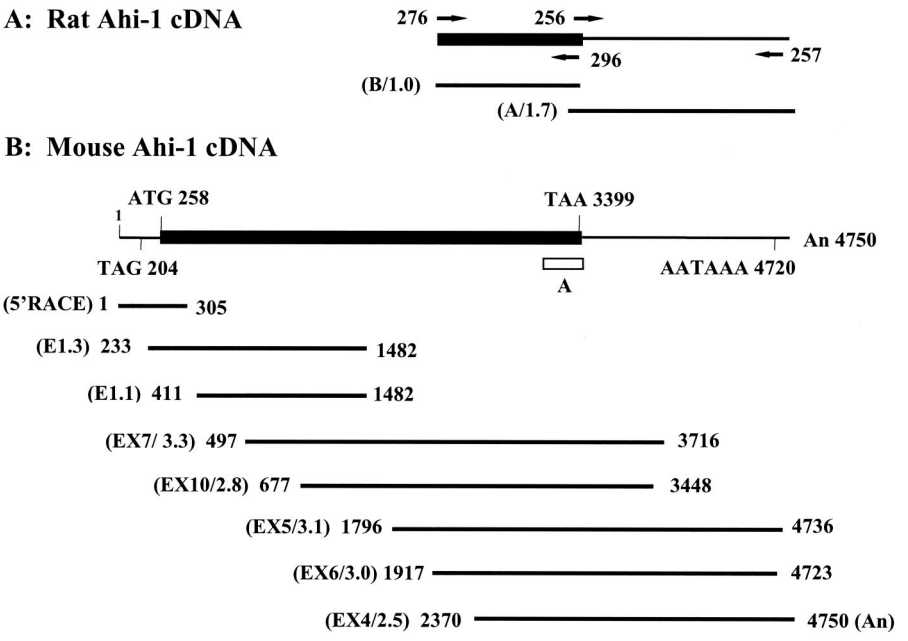

FIG. 2. Structure of the *Ahi-1* cDNA clones. (A) Two overlapping rat cDNA clones, obtained from a brain cDNA library, are shown with their positions from 5′ to 3′. Specific primers used for generating mouse cDNAs are indicated. (B) The complete *Ahi-1* cDNA sequence was derived with eight overlapping mouse cDNA clones from two testis cDNA libraries (E and EX) and by 5′ RACE. The methionine initiation codon, the termination codon, and the consensus polyadenylation signal are indicated. Symbols: thin line, *Ahi-1* untranslated region; closed box, *Ahi-1* coding sequences; horizontal lines, *Ahi-1* overlapping cDNA clones; open box, the location of the RT-PCR product-A. The numbers refer to the position of the nucleotides in Fig. 3. Numbers in parentheses represent the names of each clone (E or EX) with their respective lengths given in kilobase pairs.

incubating the denatured fragments in the presence of hexamers and the Klenow fragment of polymerase 1 (13).

RNA expression by Northern analysis. RNA was isolated from various tissues by the method of Chomczynski and Sacchi (6), separated on 1% formaldehyde-agarose gels, transferred onto nylon membranes, and hybridized as described above. Washings were performed for 15 min at room temperature in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), for 1 h at 65°C in 0.1× SSC–0.5% sodium dodecyl sulfate, and in 0.1× SSC three times for 2 min each time at room temperature. The oligonucleotide used to detect the 18S rRNA was 5′-ACGGTATCTGATCGTCTTCGAACC.

PCR and RT-PCR analysis. Total RNAs from the mouse tumor cells, from Jurkat cells or from mouse thymus (5 μg) were transcribed with Moloney MuLV reverse transcriptase as described previously (16). PCR was performed on the cDNAs obtained by using the following primers. For analysis of the different splicing forms of the 5′-untranslated human Ahi-1 sequences, the primers used were primer 910 (5′-GGGAGTTGATTTGCACTGCTC) and primer 899 (5′-ATCTCCTTTGCCATTTCTTCAG), which are located on exons 1 and 4, respectively (see Fig. 5C). For analysis of the novel sequences of exon 24 present at 3′ end of the AK000076 cDNA clone, we used primers 897 (5′-AGACCGACAGTCACTTTGCTG; exon 23/24) and 898 (5′-TCCCAATATTTTATGAGTTTCAAAG; exon 24). For analysis of the 3′ end Ahi-1/viral fused transcripts, in Mis-2 and Ahi-1 (including L48) rearranged tumors, the primer 828 within the U5 LTR (5′-AGTGATTGACTACCCGTCAGC), and a primer in exon 16 (primer 276, 5′-GAGAACAATGGTTGCGTTTTG), exon 17 (primer 277, 5′-CGAGTCTCTGTTTCACAAGC), exon 19 (primer 264, 5′-CACCAGGCCCTGCAAAGAAG), or exon 22 (primer 265, 5′-GTTAAAGGAGGGGACGCTC) were used. The PCR products obtained were purified on agarose gels, cloned, and sequenced. To confirm the presence of the novel sequences at the C terminus in the mouse EST clones BB615071 (exon 20) and BG29736 (exon 24), a primer in either exon 20 (primer 1311, 5′-CTGTCACAAATTCTCTGGTCAGG) or exon 24 (primer 1312, 5′-CATTCCTTGGTCCAGCAGCAG) and the primer 264 (exon 19) were used.

PCR analysis was used to map the provirus integration sites precisely in T-cell tumors arising in Moloney MuLV-infected MMTV^D/myc Tg mice. The Moloney MuLV R LTR primers 110 (5′-TAAGCTAGCTTGCCAAACCTACAGGTG) and 112 (5′-CCCGAGCTCAATAAAAGAGCCCACAAC) and the Ahi-1 specific primer 256 (5′-CCGCAAAGTCACCCTGATAGAG; exon 26) and primer 276 (exon 16) were used. Fragments of ~2.0 and ~2.2 kbp were amplified from T3465 and 1213-19 cell DNA, respectively, purified on agarose gels, cloned, and sequenced.

Sequence analysis. Homology searches were carried out by computer-assisted nucleotide and protein blasts on the NCBI and EBI Web sites by using the human and mouse genome databases. Search for known domains were done by using the conserved domain database at NCBI with the latest versions of SMART v.3.3, Pfam v6.6, PSPOR II, and ScanProsite programs.

Nucleotide sequence accession numbers. The nucleotide sequence accession number of the full-length mouse *Ahi-1* cDNA (Fig. 3A) is AY133241; the accession number of clone I of mouse *Ahi-1* cDNA (Fig. 5A) is AY133242; the accession number of the untranslated 5′-end of human *Ahi-1* cDNA (Fig 5B) is AY133243; and the accession numbers of fragments 1, 2, 4, and 5 of the rat *Ahi-1* genome (Fig. 6A) are AY133244, AY133245, AY133246, and AY133247, respectively.

RESULTS

Detection of proviruses integrated within the *Ahi-1* region in additional types of murine tumors. The *Ahi-1* locus was previously identified as a common provirus integration site in pre-B lymphoma induced by the Abelson virus complex (series YS3, 1278 and 1394, Fig. 1A) (41). To determine whether the *Ahi-1* locus was also targeted by provirus in MuLV-induced leukemias of other hematopoietic cell lineages, we studied T-cell tumors arising in MMTV^D/myc Tg mice infected with Moloney MuLV (16). *Eco*RV-digested tumor DNAs ($n = 28$) were screened by the Southern technique with *Ahi-1* probes D and I (Fig. 1A). Novel tumor-specific fragments, in addition to the normal germ line fragment (23 kbp), were detected in four tumors (T3464, T3465, T3481, and T3469). In addition, PCR

analysis confirmed and precisely mapped the integration in tumor T3465 within intron 26 (Fig. 1B), and RT-PCR analysis revealed the integration of a provirus between exons 22 and 23 in the L48 cell line (Fig. 1A) (see below, Fig. 7E), previously established from Moloney MuLV infected $MMTV^D$/myc Tg mice (16). The proviruses inserted at the *Ahi-1* locus in T-cell leukemia were found at a frequency of ~14% and to be integrated in the same orientation as those found in A-MuLV-induced lymphomas (Fig. 1A). Provirus integrations within Ahi-1 were also observed in 1 (i.e., 1213-10) of 20 distinct DNAs (5%) from Moloney MuLV-induced rat thymomas (series 1213, Fig. 1A). These insertions were mapped in the same clusters of provirus integrations as those found in Abelson virus-induced tumors (Fig. 1A). Therefore, rearrangements of the *Ahi-1* locus appear to be involved in the development of lymphomas from different cell lineages.

Identification of *Ahi-1* coding sequences by exon amplification. To identify a putative gene located within the *Ahi-1* locus, a bidirectional chromosome walking was initiated from the *Ahi-1* region first identified (41) by using cosmid and phage libraries. Sequences spanning almost 80 kbp of the *Ahi-1* region were cloned (Fig. 1A) and used to identify coding sequences within this region by using an exon amplification method (4). A total of five putative exons ranging from 58 to 300 bp were recovered by this technique. One of the RT-PCR products fragment A (160 bp; Fig. 2B), generated from the 4-kbp *Bgl*II fragment of Cos 1 clone, was found to hybridize to two major transcripts of 5 and 3.5 kb in mouse brain and testis RNA by Northern blot analysis (data not shown), indicating that this fragment contained potential exons(s). A summary of these results is shown in Fig. 1A and 2.

Molecular cloning and sequencing of mouse and rat *Ahi-1* cDNAs. To identify the *Ahi-1* gene and to map the provirus integration sites relative to this gene, we cloned the *Ahi-1* cDNA. Since the *Ahi-1* gene was highly expressed in brain and testis, a rat brain cDNA library was first screened. Several positive clones ranging from 0.6 to 2.5 kbp were obtained (Fig. 2). Northern blot analysis performed with both rat cDNA clone A and clone B showed the same hybridization pattern previously observed with the RT-PCR product A, indicating that part of the *Ahi-1* gene had been cloned. The full-length mouse *Ahi-1* cDNA was obtained by RT-PCR and 5′ RACE methods and by rescreening a mouse thymic cDNA library, as described in Materials and Methods. A schematic representation of these Ahi-1 cDNA clones is shown (Fig. 2).

The complete nucleotide and deduced amino acid sequence of the *Ahi-1* cDNA is shown in Fig. 3A. An open reading frame begins with a potential initiation methionine codon at nucleotide (nt) 258, based on Kozak consensus rules (25). An in-frame stop codon is located 204 nt upstream of this site. The predicted encoded protein is ~120 kDa long and contains 1,047 amino acids (aa). The in-frame termination codon (TAA) is followed by a long 3′-untranslated region (Fig. 2 and 3A). The sequence immediately upstream of the poly(A) tail contains a AATAAA polyadenylation signal (nt 4720).

The mouse *Ahi-1* gene product contains an SH3 motif and seven WD40 repeats. Computer-assisted analysis of the predicted amino acid sequence of the *Ahi-1* cDNA product revealed a number of significant structural features (Fig. 3B). In particular, the *Ahi-1* cDNA encodes a region of ca. 50 aa residues (aa 905 to 961) with significant sequence homology to the Src homology 3 (SH3) domain (25 to 35% identity [40 to 50% similarity]).

Several regions of the predicted Ahi-1 protein are also proline-rich (Fig. 3B) and contain nine potential SH3-binding sites (PxxP) at positions 9, 98, 182, 352, 803, 858, 976, 985, and 1001; the numbers here refer to the first proline residue in the PxxP consensus sequences present in high-affinity SH3 ligands (7, 43, 44, 56, 64).

Interestingly, another region of the Ahi-1 protein upstream of the SH3 motif and spanning ~320 aa residues (aa 448 to 769) exhibits significant sequence homology to the WD40 repeats (33). Seven WD40 repeats were identified in the predicted Ahi-1 polypeptide sequence at positions 448 to 490, 493 to 532, 537 to 576, 583 to 622, 641 to 678, 684 to 721, and 724 to 769. Each WD40 repeat consists of stretch of ca. 40 aa in which tryptophan and aspartic acid residues and certain other amino acids are highly conserved, a finding consistent with the range of 23 to 41 observed in other members of this class of proteins (33).

In addition, computer analysis with the ScanProsite tool revealed the presence of many potential phosphorylation sites: among them, 21 protein kinase C phosphorylation sites, 3 cyclic AMP (cAMP)- and -cGMP-dependent protein kinase phosphorylation sites (23), and 26 potential casein kinase II phosphorylation sites. Also, two potential tyrosine kinase phosphorylation sites are present (residue 71 [KLKEQLTY] and residue 941 [KDNEDWWY]) with sequences matching optimally the consensus sequence of tyrosine kinases [R/K-x-(2,3)D/E-x-(2,3)Y] (8) (Fig. 3A). Tyrosine phosphorylation has been shown to play an important role in several proteins involved in malignant transformation and signal transduction (57). The amino-terminal region of Ahi-1 contains an acid-rich domain (Fig. 3A). Such a domain is found in quite different proteins, such as nucleolines, calcium-binding proteins, and transcription factors (40). Finally, three potential PEST sequences (at residues 19 to 30, 100 to 118, and 417 to 432) (46), three potential nuclear localization signal (aa 62 to 94, 134 to 151, and 825 to 868) (10), and four potential glycosylation sites (at positions 310, 536, 821, and 851) were found in the Ahi-1 protein. A summary of the structural domains present in the Ahi-1 protein is depicted in Fig. 3B.

Conservation of the *Ahi-1* gene in evolution. To determine whether the *Ahi-1* gene has been conserved throughout evolution, we searched for homologous sequences in other species. Mouse, rat, hamster, cat, cow, and human DNA samples were digested with *Eco*RI and hybridized at high stringency with *Ahi-1* cDNA clone-B probe. Specific DNA fragments were detected in all of these species (Fig. 4), indicating a significant conservation of these sequences among mammals. Comparison of the mouse, rat, and human sequences now available (see below) confirmed the high conservation of this gene. However, no gene coding for a protein containing WD40 repeats and a SH3 domain was found in the sequences of *Drosophila melanogaster* or *Caenorhabditis elegans*. This suggests that the *Ahi-1* gene may have an important function in mammals.

Identification of different isoforms of mouse and human Ahi-1 cDNA. A homology search of the GenBank databases with our mouse and rat *Ahi-1* cDNA sequences identified three distinct human cDNA sequences (AL136797, AK024085, and

A

```
     GCGATCCCCAAGCTCGGTCCGCCGGCCAACGTGCCGCCCCCCTCGTCCTGCCTAGGCGCAGGGAACTTCGCGCCGCCGCCGGGCTGTCCCTCCTCTCTCCTGTTACCCTCAAAGTAACCT 120
     TCTGGCTACCTTAGAACTAAAATTGGCACTAAAAGCCTAGAAAAGTGAGCAGCACGTTGAAGATTCAGAGGACAATGTTAGTGTAGAGCCATACAAACAAGAAAAGACAAGTAAACACTA 240
     CCCAGATCGCACAGGAAATGGAGCCAGAAACTCCAGAGAAGGTGGATTCTGCACAAGAGAAAGTCAGAGGCAAGACCCCGACAGCAGATGATTCAGATGACAGCCGAGAAAAGACTGGCA 360
                      M  E  P  E  T  P  E  K  V  D  S  A  Q  E  K  V  R  G  K  T  P  T  A  D  D  S  D  D  S  R  E  K  T  G  I
     TAGAAGAGAAAGGAGAACTGACCGACGCCTATCAGCTGACGGTAGCTGAAGAAATGGCAAAGGAGATCAAGAAGAAAATAAGAAAGAAGCTAAAGGAACAGCTGACCTACTTTCCTCCAG 480
  36  E  E  K  G  E  L  T  D  A  Y  Q  L  T  V  A  E  E  M  A  K  E  I  K  K  K  I  R  K  K  L  K  E  Q  L  T  Y  F  P  P  D
     ACACTCTATTGCATGACGACAAGCTGGCCAGTGAAAAAAGAAAGAAGAAGAAGAAGAAAGTGCCAGTGCCCACTAAGCCTGAGTCAAGTCCCTCAGATGTCTGTGACAGTGCAGTTGAAG 600
  76  T  L  L  H  D  D  K  L  A  S  E  K  R  K  K  K  K  K  K  V  P  V  P  T  K  P  E  S  S  P  S  D  V  C  D  S  A  V  E  G
     GGGAACAAAAGAAAGAAGGTACCCCTGAGGACTCTCAGCACATGGAGGGAATCTGCTCGAGAGAGCAGGATGTGGATGCCACTGTGCCAGAGAACGCAAAGCCCAAACCCAAGAAGACAA 720
 116  E  Q  K  K  E  G  T  P  E  D  S  Q  H  M  E  G  I  C  S  R  E  Q  D  V  D  A  T  V  P  E  N  A  K  P  K  P  K  K  T  K
     AGAAGAAGACTAAAGCAGTTTCAAATGATAATGAAGACACTAATGGAGATGGTGTTCATGAGATAACAAGCCGAGACAGCCCAGTTCATCCCAAGTGCTTGCTGGATGATGACCTGCTCA 840
 156  K  K  T  K  A  V  S  N  D  N  E  D  T  N  G  D  G  V  H  E  I  T  S  R  D  S  P  V  H  P  K  C  L  L  D  D  D  L  V  M
     TGGGAGTCTACATTCACCGAACCGATAGACTTAAATCTGACTTTATGATTTCTCACCCAATGGTAAAGATCCATGTGGTTGATGAGCACACTGGCCAGTACGTCAAGAAAGATGACAGTG 960
 196  G  V  Y  I  H  R  T  D  R  L  K  S  D  F  M  I  S  H  P  M  V  K  I  H  V  V  D  E  H  T  G  Q  Y  V  K  K  D  D  S  E
     AACGTCCTGTTTCATCTTATTATGAGAAAGACAACGTGGACTATATTCTCCCTATTATGACACAGCCATATGACTTTAAAAAGCTAAAGTCAAGGCTTCCGGAGTGGGAGGAGCAAGTTA 1080
 236  R  P  V  S  S  Y  Y  E  K  D  N  V  D  Y  I  L  P  I  M  T  Q  P  Y  D  F  K  K  L  K  S  R  L  P  E  W  E  E  Q  V  I
     TTTTTAATGAAAATTTTCCCTATTTGCTTCGAGAGTTTGAAGAATGTCCAAAAGTCATCCTGTTCTTTGAGATTCTTGACTTTTTAAGCATGGATGAAATCAAGAATAACTCTGAGGTTC 1200
 276  F  N  E  N  F  P  Y  L  L  R  E  F  E  E  C  P  K  V  I  L  F  F  E  I  L  D  F  L  S  M  D  E  I  K  N  N  S  E  V  Q
     AAAACCAAGAGTGTGGCTTTCGGAAAATTGCCTGGGCGTTTCTTAAGCTTCTGGGAGCCAATGGGAATGCAAACATCAACTCAAAACTTCGCCTGCAGCTCTACTACCCACCGACTAAGC 1320
 316  N  Q  E  C  G  F  R  K  I  A  W  A  F  L  K  L  L  G  A  N  G  N  A  N  I  N  S  K  L  R  L  Q  L  Y  Y  P  P  T  K  P
     CTCGATCCCAGCTAAATGTCGTCGAGGTTTTTGAATGGTGGTCCAAGTGTCCAAGAAATCGTTATCCATCAACATTGTATGTAACCGTACGAGGATTGAAAGTCCCGGACTGTATAAAGC 1440
 356  R  S  Q  L  N  V  V  E  V  F  E  W  W  S  K  C  P  R  N  R  Y  P  S  T  L  Y  V  T  V  R  G  L  K  V  P  D  C  I  K  P
     CATCTTACCGCTCTATGATGGCTCTCCAGGAGGAAAGGGGTACACCAGTGTACTGTGAACGTCACCGTGAAACAAGTTCCGTGGACACAGAACCTGGACTAGAAGATTCAAAGGAGGAAG 1560
 396  S  Y  R  S  M  M  A  L  Q  E  E  R  G  T  P  V  Y  C  E  R  H  R  E  T  S  S  V  D  T  E  P  G  L  E  D  S  K  E  E  V
     TGAAGTGGAAACGTCTGCCTGGCCAGGCCTGTCGTATCCCAAACAAGCATCTCTTCTCACTGAATGCTGGAGAACGCGGCTGTTTCTGTCTTGATTTCTCTCACAATGGAAGGATATTAG 1680
 436  K  W  K  R  L  P  G  Q  A  C  R  I  P  N  K  H  L  F  S  L  N  A  G  E  R  G  C  F  C  L  D  F  S  H  N  G  R  I  L  A
     CAGCAGCCTGTGCCAGCCGAGACGGATATCCGATTATATTATATGAAATTCCTTCTGGACGTTTTATGAGAGAATTGTGTGGCCACCTCAATATCATTTATGATCTCGACTGGTCAAAAG 1800
 476  A  A  C  A  S  R  D  G  Y  P  I  I  L  Y  E  I  P  S  G  R  F  M  R  E  L  C  G  H  L  N  I  I  Y  D  L  D  W  S  K  D
     ACGATCGCTATCTCGTTACTTCATCCTCTGACGGCACTGCCAGAGTTTGGAAGAATGAAATCAATAGCACGAGCACGTTCAGAGTCTTACCTCACCCCTCCTTCGTCTACACGGCTAAAT 1920
 516  D  R  Y  L  V  T  S  S  S  D  G  T  A  R  V  W  K  N  E  I  N  S  T  S  T  F  R  V  L  P  H  P  S  F  V  Y  T  A  K  F
     TCCACCCAGCCACACGGGAGCTGGTGGTTACAGGATGCTACGACTCTATGATAAGGATTTGGAAAATTGATGCAAGGGAGGATGCTGCCATATTAGTGCGCCAGTTGGATGTGCACAAGA 2040
 556  H  P  A  T  R  E  L  V  V  T  G  C  Y  D  S  N  I  R  I  W  K  I  D  A  R  E  D  A  A  I  L  V  R  Q  L  D  V  H  K  S
     GCTTTGTCAACTCCATCTGTTTTGACGATGAAGGTCACCACATGTACTCAGGAGACTGCATCGGGGTCATCGTTGTCTGGGACACGTACGTCAAAGTGAATGATGTGCAAACCTCCGTGC 2160
 596  F  V  N  S  I  C  F  D  D  E  G  H  H  M  Y  S  G  D  C  I  G  V  I  V  V  W  D  T  Y  V  K  V  N  D  V  Q  T  S  V  R
     GCCACTGGACGATAAATAAGGAAATTAAGGAAACTGAATTCAGGGGAGTTCCAATAAGTTACCTGGAGGTTCATCCCAATGGAAAGCGTTTGCTAATCCACACCAAAGACAGTACTCTGA 2280
 636  H  W  T  I  N  K  E  I  K  E  T  E  F  R  G  V  P  I  S  Y  L  E  V  H  P  N  G  K  R  L  L  I  H  T  K  D  S  T  L  R
     GGATTATGGACCTGCGGATATTGGCAGCCAGGAAATTTGTGGGTGCAGCAAATTACCGTGAGAAGATCCATAGCACCTTGACGCCGTGCGGGACTCTGCTCTTTTCTGGGAGTGAGGATG 2400
 676  I  M  D  L  R  I  L  A  A  R  K  F  V  G  A  A  N  Y  R  E  K  I  H  S  T  L  T  P  C  G  T  L  L  F  S  G  S  E  D  G
     GGATAGTATACGTTTGGAACCCAGAGACAGGAGAACAAGTGGCAATGTACTCTGACCTGCCATTCAAGTCCACAATCCGAGACATCTCTTACCACCCCCTGGAGAATATGGTTGCCTTCT 2520
 716  I  V  Y  V  W  N  P  E  T  G  E  Q  V  A  M  Y  S  D  L  P  F  K  S  T  I  R  D  R  S  Y  H  P  L  E  N  M  V  A  F  C
     GTGCCTTTGGACAGAGCGAGCCGATTCTTCTTTATATTTATGATTTCCAGGTAGCCCAGCAGGAGGCTGAAATGTTGAAACGCTATAGTGGGACACTCCCGTTGCCTGGGATACATCAGA 2640
 756  A  F  G  Q  S  E  P  I  L  L  Y  I  Y  D  F  Q  V  A  Q  Q  E  A  E  M  L  K  R  Y  S  G  T  L  P  L  P  G  I  H  Q  S
     GTGAGGATGCTTTATGCACCTGCCCTAAGCTGCCCCAGCAGGGCTCTTTTCAGATTGATGAATTTGTCAACACTGAAAATAGTTCATCAAGAAAGATTCAACTTGTGAAACAGAGACTCG 2760
 796  E  D  A  L  C  T  C  P  K  L  P  Q  Q  G  S  F  Q  I  D  E  F  V  N  T  E  N  S  S  S  R  K  I  Q  L  V  K  Q  R  L  E
     AAACTGTCACAGAGGTGATACGATCCTGTGCTGCAAAAGTCAACAAAAATCTCTCAATGACTTCGCCACCACCAGGCCCTGCAAAGAAGCCCAGGGTAAAGCAGTCGTTCGTGCTGACCA 2880
 836  T  V  T  E  V  I  R  S  C  A  A  K  V  N  K  N  L  S  M  T  S  P  P  P  G  P  A  K  K  P  R  V  K  Q  S  F  V  L  T  T
     CTGATGAGATCATACATCAGTTTGGACTCCCCCAGACTGCATTTATCAGCATAGAGAGAGGGCCTTTTGTGCGCCATGTTGATCCACCACCAATGGTAGTGGCTCTTTATGACTACACAG 3000
 876  D  E  I  I  H  Q  F  G  L  P  Q  T  A  F  I  S  I  E  R  G  P  F  V  R  H  V  D  P  P  P  M  V  V  A  L  Y  D  Y  T  A
     CCAGCCGATCAGATGAACTAACCATCCATCGTGGAGACATTATCCGAGTGTATTTCAAAGATAATGAAGACTGGTGGTACGGCAGCGTAAGAAAGGGGCAGGAAGGGTTCTTTCCAGCTA 3120
 916  S  R  S  D  E  L  T  I  H  R  G  D  I  I  R  V  Y  F  K  D  N  E  D  W  W  Y  G  S  V  R  K  G  Q  E  G  F  F  P  A  N
     ATCATGTGGCCAGTGAAACACTGTATCGAGACTCCCCTCCGAAGGTAAAGGAGCGCTCCCCTCCTTTAACTCCCAAGGAGAAAACTAAACCAGAAAAGCCTCTGGCTTCTCAAAAGCAGT 3240
 956  H  V  A  S  E  T  L  Y  R  D  S  P  P  K  V  K  E  R  S  P  P  L  T  P  K  E  K  T  K  P  E  K  P  L  A  S  Q  K  Q  S
     CGCTCAGTAAGGGCAGACCCCTGGATCCCAGACTGGGCCCGCAGCCTGTGGGGCATTCTGAGAAGGGCAAAGATCAAAACGTGGAGGACCGAGGACACAAAGTAGATATGGAGACAAAGA 3360
 996  L  S  K  G  R  P  L  D  P  R  L  G  P  Q  P  V  G  H  S  E  K  G  K  D  Q  N  V  E  D  R  G  H  K  V  D  M  E  T  K  K
     AAAGCGAGCCGGTGGTCCGCAAAGTCACCCTGATAGAGTAAAAAGCTGGAGACGAAGCAGAAGCCGCCTGATGAGTGCACATGATGTGACGTCCACGCCAGACGGAGTTCTCGTTAGATT 3480
1036  S  E  P  V  V  R  K  V  T  L  I  E  *
     TTGGGTGAAAAGAAGCAGTATTTATTGACCTTAAAAATCTGAAGCAAACTTAGCTCTAAAAAGAAAAAAATCACTGTGGCCTTTGGGCTTAATTGTTATGAACCATTGTGAGTGTTACTG 3600
     GTCAAAGTATTGTGCTTGTGCTATTAGTAACTACATCAGAATTATATTGCTGCTGCTGGAAAAGCCAATAGAGAAAACCCCAGAACTGCTACTCAGTAAATGTGAATAACAGTGTGCGCA 3720
     GTGTTAGTCATTCACTAAGCCATCATCCCGTGTACTTGTGGCCATGATGTGGCTGTGCGTAGGCCGTCATCCCGTGTACTTGTGGCCATGATGCAGATGCTCTGCTTTTGAAGTTAGCCG 3840
     CGTGTAAAACCTAAGTAGCTGCTGGATATTGTGTCTTCAAACAGCCCCAGTAACAAACGCGGTAGAAAACGACACGTTTGCTTTTGAGAACTAAATGTTTGTGTTTCCCTTGAGTGAATG 3960
     TATGTCACTCAACACTCCTAGACTAAGAATGATTTTGGCTGGACTGACAACAGTACACCAAATTTATTTTTATTTTTTAGACAGAGTTTATATTTAACAGCATAAGTAATTTATGGATGT 4080
     AAACTCTGTAGTGGTTATAGTTTAGTCTGAATATTTGTTTTTAAATTATTACACAGCATTATTTATGTCTTTAGTTTCAGAATAACAGTGCCATTACCATATGTTTGTATGTTGCAGTTG 4200
     ATATAGCTGAGAACTCCTCCATCTTGGTTTTCTCTTCACTCATCCATTTCTTCAGCTGCATGATAAGTCTGTCAGTAAAGCATTTAACCCCAACAAGGTCAGACTATGGTAAAGTTAGAA 4320
     AAACATATTTACCGAAGTACACAATTGATCTCATTGAAATAAGAAGTGAGAACCGCTGTGTTGAGCACTGAGGTGTAGACTTGTCCAGATGCACAGCGTGAGTAGCTTCGAGTACGCATC 4440
     TGTTTGCACATGTTGTAGAAGCTTAAAATGGGAACATGACCCTGCTGTGGAGAACCTTGGTAGAATTGTTGTTAGTACTAAAGGAGATTTCCAGCGCATACAGGGTGGCACTGAAAACTG 4560
     CTGAAGCCCACAGCCCTCAGCCCCGAAGGCTCAGCGGCCAGTACCAGCGGCCTTGCTGTGGTTGTGTGTCTGCGTCCCTGGCACACTTGGGCAGGCAGCTCAGTCACTCTGAACTTGCTT 4680
     CTCTTCTGTCTTGTGAGTGTGCTGCTTGAAACAATGAAAAATAAAGCTGTGTTCCTACCCCTTCAAAAAA                                                   4750
```

B **Mouse Ahi-1**

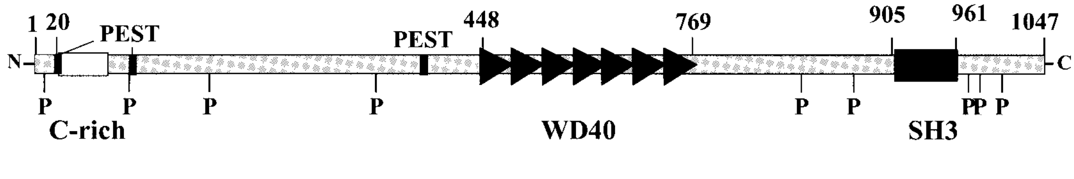

FIG. 3. Nucleotide sequence of mouse *Ahi-1* cDNA. (A) The nucleotide sequence of the *Ahi-1* cDNA and the deduced amino acid sequence of the Ahi-1 protein are shown. Amino acids are given in the single-letter code. The conserved SH3 domain is underlined with a dashed line. The WD40 repeats are underlined. (B) Structure of the putative Ahi-1 protein. A summary of structural motifs present in the Ahi-1 protein is shown, including: one SH3 domain (black box), seven WD40 repeats (▶), an acidic-rich domain ("C-rich") (□), nine proline-rich motifs (P), and three PEST sequences (■).

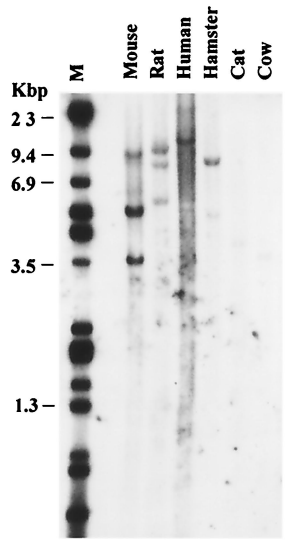

FIG. 4. Interspecies conservation of the *Ahi-1* gene. DNAs (20 μg) from various species were digested with *Eco*RI and analyzed by Southern blot with the rat *Ahi-1* cDNA fragment B as a probe. Hybridization and washing were done as described in Materials and Methods. Lanes:1, mouse; 2, rat; 3, human; 4, hamster; 5, cat; 6, cow; M, *Hin*dIII/*Eco*RI-digested λ DNA.

AK000076) and two mouse EST sequences (BG297436 and BB615071) that show high homology. Comparisons of these clones are schematically presented in Fig. 5A. The human Ahi-1 homolog clone AL136797 (5,277 bp) contains the entire coding region, whereas clones AK024085 (2,922 bp) and AK000076 (2,075 bp) appear to represent shorter isoforms. The alignment of the human and mouse *Ahi-1* cDNA sequences showed a high degree of homology and, interestingly, some major differences at both the 5′ and the 3′ ends. Two novel sequences are present in the AK000076 and AK024085 cDNA clones downstream of the nucleotides coding for the last WD40 repeat (exon 17) and for the SH3 motif (exon 22), respectively. These two novel sequences, missing in the mouse *Ahi-1* cDNA and in the human AL136797 clone, represent genuine exons and both contain an in-frame termination codon. We have confirmed the presence of the novel sequences in exon 24 present in the AK00076 clone by RT-PCR on RNA from Jurkat cells (data not shown). Moreover, these results are supported by the presence of the similar early termination codon in two mouse EST clones (BB615071 and BG 297436) found in GenBank. Furthermore, a similar sequence containing an early termination codon after the SH3 motif was found in one of our partial cDNA clones (clone i) isolated from the brain cDNA library.

In addition, at the 5′ end of the coding sequence of the human AL136797 cDNA, there are 420 additional nucleotides downstream of the ATG which showed no homology with the mouse *Ahi-1* cDNA. Furthermore, we identified and cloned novel extra untranslated 5′ sequences of human *Ahi-1*. Using RT-PCR on RNA from human Jurkat cells with the primers 910 and 899, we were able to amplify, clone, and sequence three PCR products (~0.8, 0.9, and 1.0 kb) corresponding to three transcribed exons, designated exons 1 (68 bp), 2 (62 bp), and 3 (86 bp) (Fig. 5C). These exonic sequences are not translated and are spliced differently, thus leading to three isoforms of human Ahi-1 5′-untranslated region (Fig. 5B and C). There is no homology between these human 5′-untranslated sequences and the mouse untranslated *Ahi-1* sequences.

The amino acid sequences predicted from these cDNA sequences were also analyzed. The human Ahi-1 protein contains an additional 140 aa at its N terminus that are not present in the mouse Ahi-1 protein. This region harbors a potential coiled-coil domain, as revealed by the SMART program. The two other shorter human isoform proteins contain an initiation codon but are likely to be missing some sequences at the N terminus. Two of the human AHI-1 proteins (AL136797 and AK024085) contain the SH3 and WD40 domains found in the mouse homolog. Within this WD40/SH3 region, the amino acid sequences of human and mouse Ahi-1 are highly conserved (>82% identity and 90% similarity). However, the sixth WD40 repeat of the human proteins was more divergent from that of the mouse. Furthermore, the AK000076 isoform lacks the SH3 domain. In addition, this protein contains a novel amino sequence at the C terminus not present in either the mouse or in the other human clones. This novel sequence is encoded by a single exon (exon 24) present in the human genome (see below).

Together, these results show that *Ahi-1* gene expression can be modulated by alternative splicing. Interestingly, the spliced variants lacking the SH3 domain (AK000076 and BB615071) would be expected to encode truncated forms of the protein with most likely a distinct function.

Mapping of the *Ahi-1* provirus insertion sites at the 3′ end of the gene. The orientation of the proviruses relative to that of the *Ahi-1* gene and the localization of the coding exons of the gene were performed by Southern blot, PCR, and sequence analyses (Fig. 1B). Moreover, the locations and sequences of these mouse exons were confirmed by data obtained from the database (see below and Fig. 6). These analyses showed that most of the proviruses had integrated at the 3′ end of the *Ahi-1* gene in two clusters and in an inverse transcriptional orientation relative to that of the *Ahi-1* gene. In the more proximal cluster, at least one provirus (T3465) was inserted within the last intron (Fig. 1B), whereas the others were integrated within the last noncoding exon or right downstream of this last exon. The second distal cluster of provirus insertion was ca. 16 kbp downstream of the last exon of the *Ahi-1* gene (Fig. 1A). Only in one tumor cell line (L48) did integration occur outside the main clusters in intron 22 (Fig. 1A).

Genomic organization of the mouse and human Ahi-1 genes. Based on the BLAST N and BLAST X homology programs, the organization of the mouse and human *Ahi-1* genes was deduced from the comparison of all of the mouse and human Ahi-1 cDNA isoform sequences available from the GenBank Databases. The results of these analyses are shown in Tables 1 and 2. This analysis revealed that the *Ahi-1* gene spans more than 200 kb in a region of the human chromosome 6 and a minimum of 100 kbp (distributed on at least nine contigs) on the mouse chromosome 10 (Fig. 6 and Tables 1 and 2). The mouse or human *Ahi-1* gene contains at least 27 and 33 exons, respectively. The introns separating exons are between 200 bp and 35 kbp long. The exons range in size from 26 to 560 bp, excluding exons 33 (human) and 27 (mouse), which constitute a large 3′-untranslated region. The exonic sequences are identical to the human and mouse Ahi-1 cDNA sequences with the

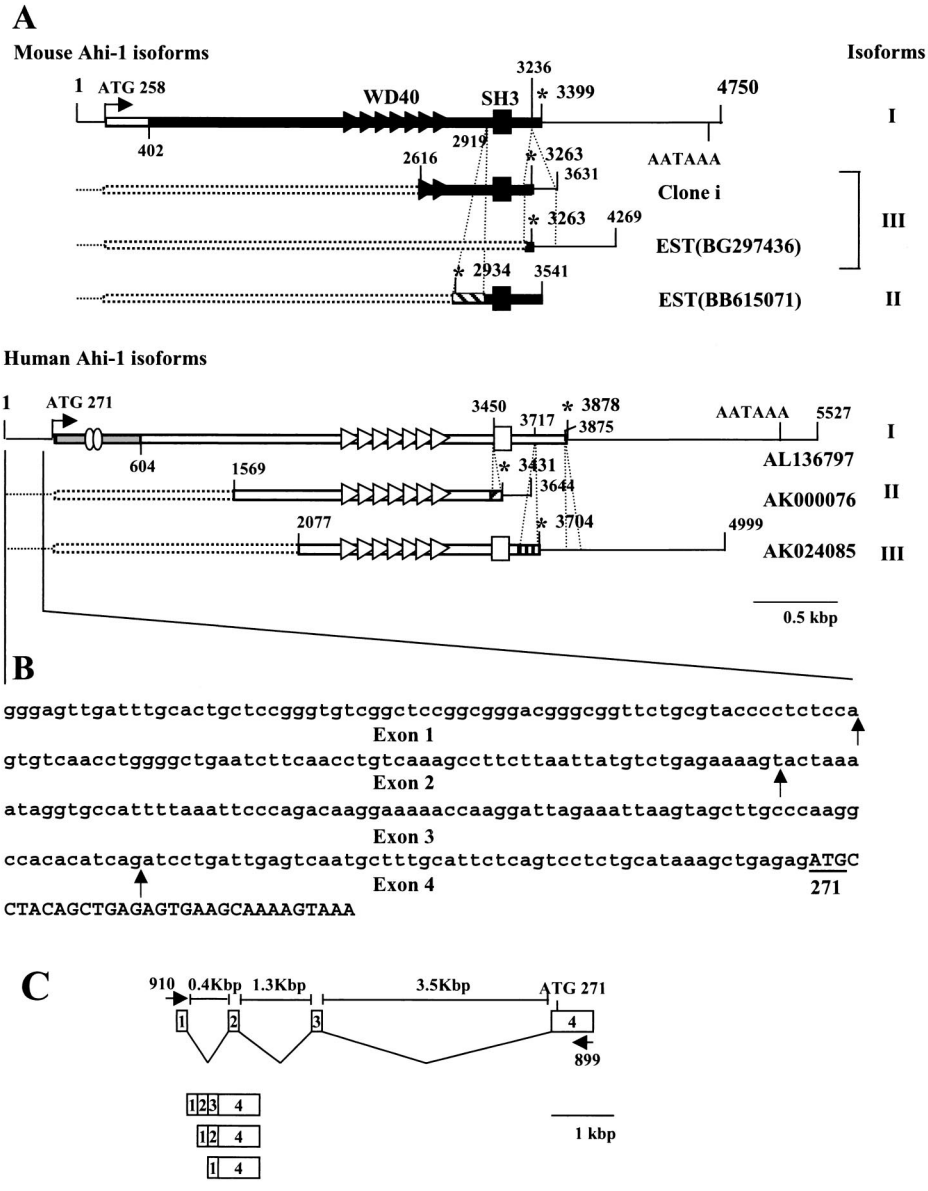

FIG. 5. Structure of the mouse and human *Ahi-1* cDNA isoforms. (A) Schematic alignment of mouse and human *Ahi-1* cDNA clones. The numbers refer to the positions of the nucleotides. The open reading frames are indicated by thick bars, boxes, and triangles, and the untranslated sequences are indicated by thin lines. In-frame stop codons are indicated by asterisks. The polyadenlyation signals are shown for some cDNAs. All strippled bars represent alternative exons: exons 24 and 20 in mouse and exons 24 and 28 in human *Ahi-1* cDNA. Note that the 5′ end (nt 1 to 402; □) and the 3′ end (nt 3190 to the end) sequences of the full-length mouse Ahi-1 are more divergent from the corresponding human Ahi-1 sequences. Also note that the 5′ end sequences (nt 271 to 604; aa 19 to 604; shaded bar) of human Ahi-1 are not present in mouse Ahi-1. The presence of exon 24 (containing a termination codon) in human Ahi-1 cDNA has been confirmed by RT-PCR analysis on a Jurkat cell RNA with primer 897 within the region spanning exons 23 and 24 and primer 898 in exon 24. Similarly, the presence of exon 20 (containing a termination codon) in the mouse *Ahi-1* gene was confirmed by RT-PCR on brain RNA with 5′primer 264 in exon 19 and 3′ primer 1312 in exon 24. Note that the 5′-untranslated sequences (nt 1 to 250) of the human Ahi-1 identified in our laboratory have been added to the nucleotides listed in the GenBank for each EST clone (AL136797, AK000076, and AK024085). (B) Nucleotide sequence of the 5′-untranslated human *Ahi-1*. RT-PCR from Jurkat cells was performed with the primers 910 and 899 located, respectively, in exons 1 and 4. The PCR products were cloned and sequenced. The putative ATG translation initiation codon is underlined. Vertical arrrows indicate intron-exon boundaries. Uppercase letters, coding sequences; lowercase letters, untranslated nucleotides. (C) Genomic organization of the 5′-untranslated region of human Ahi-1. Boxes indicate exons. The lines between the boxes represent introns. Translation initiation codon is indicated over exon 4.

exception of a few nucleotides. The seven WD40 repeats are distributed in seven exons: each of the putative WD40 repeats is encoded mainly by a single exon (exons 11 to 17 for mice and exons 14 to 20 for humans). Exon 26 (human) or 22 (mouse) encodes the SH3 domain.

Differential splicing of the mouse and human Ahi-1 exons. The availability of the genomic organization of the *Ahi-1* gene (Fig. 6) and of different forms of its encoded cDNA isoforms (Fig. 5) allowed an analysis of alternative exon usage. The full-length mouse *Ahi-1* cDNA (isoform I) that was cloned in

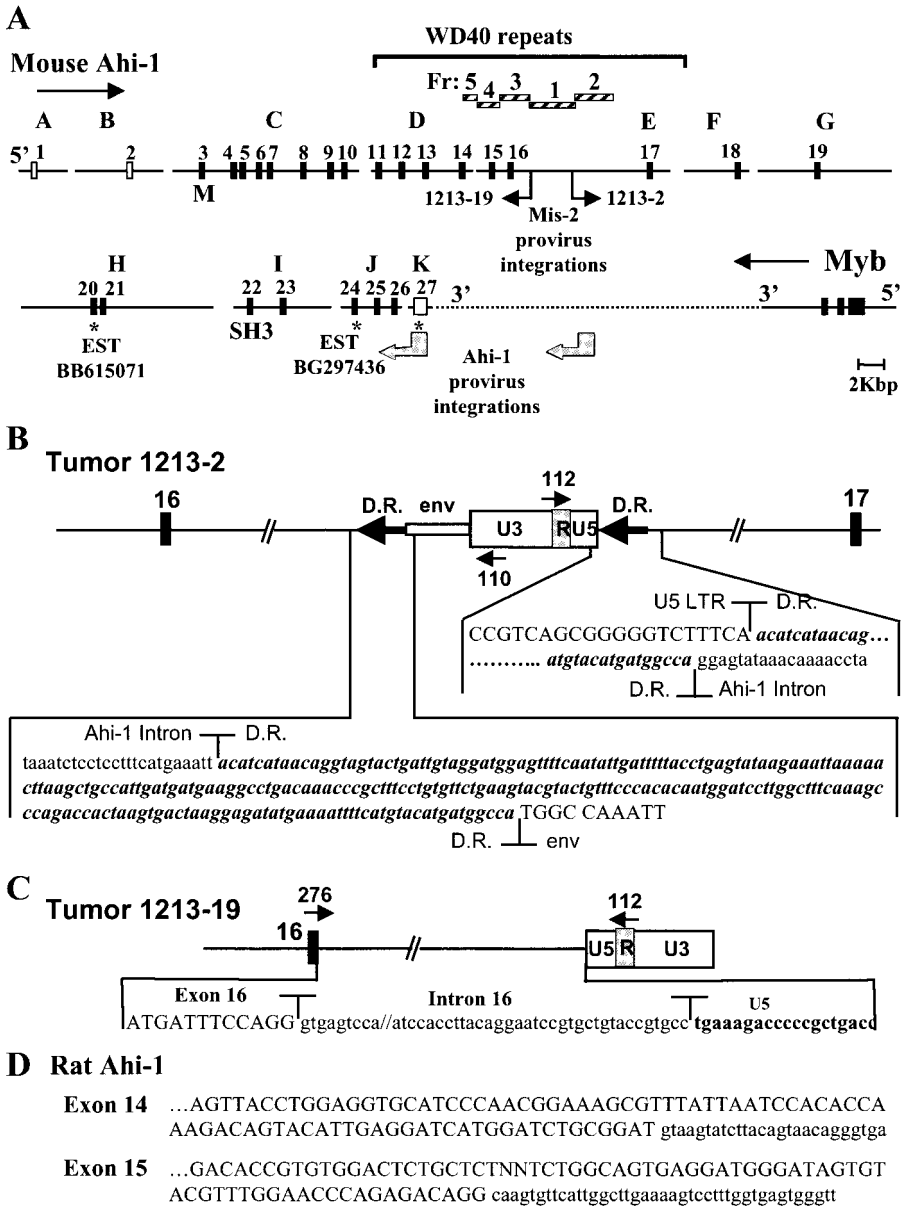

FIG. 6. Genomic organization of the mouse *Ahi-1* gene and mapping of Mis-2 integration sites. (A) Physical map and schematic representation of exons 1 to 27 of the mouse *Ahi-1* gene identified from 11 nonoverlapping combined contigs (A to K) drawn from the preliminary assembly of the mouse WGS reads based on an 15 October 2001 freeze at the Sanger Center and Whitehead Institute genome projects data bank. The contigs were as follows: B, c094702989 (7,308 bp); C, c072403314 (15,675 bp); D, c013001924 (9,046 bp); E, c052402941 (16,394 bp); F, 203795 (5,800 bp); G, c038402794 (11,198 bp); H, c047902506 (15,808 bp); I, c030002862 (8,235 bp); J, c114704732 (4,962 bp); and K, c098905234 (1,959 bp). The 5′ and 3′ end of *Ahi-1* and c-*myb* genes are indicated. The Ahi-1 and Mis-2 provirus integration sites are indicated. Black boxes represent coding exons, and the white box represents an untranslated exon. Strippled boxes are the rat DNA fragments (Fr 1 to 5) previously subcloned (62) around the Mis-2 provirus integrations. The exons coding for the SH3 and WD40 domains are also shown. Symbols: M, initiation codon; asterisks, termination codons. (B and C) Mapping and genomic organization of the proviruses integrated at the *Mis-2* locus in tumors 1213-2 (B) and 1213-19 (C). Note the large deletion of the provirus up to the end of the *env* gene, with the remaining 3′ end *env* and LTR in tumor 1213-2. Also note the perfect direct repeats (D.R.; 228 bp) flanking the deleted provirus. Black boxes represent exons 16 and 17, and the line represents the intron sequences. (D) Sequences of the intron-exon 14 and 15 junctions in rat DNA. Uppercase and lowercase letters: exon and intron sequences, respectively.

our laboratory consists of 25 exons (i.e., exons 1 to 27, excluding exons 20 and 24) with an in-frame termination codon present in exon 27. The EST clone BB615071 (isoform II) lacking the SH3 domain appears to contain 26 exons (i.e., exons 1 to 27, excluding exon 24) with a termination codon in exon 20. The EST clone BG297436 (isoform III) seems to contain 26 exons (i.e., exons 1 to 27, excluding exon 20), with a termination codon in exon 24.

The full-length human Ahi-1 clone AL136797 (isoform I) consists of 29 exons (i.e., exons 1 to 33, excluding exons 24, 28,

TABLE 1. Exon-intron boundaries of the mouse *Ahi-1* gene[a]

Exon[b]	cDNA position		Genomic position		Exon size (bp)	Intron size (bp)	Contig no.
	Start	End	Start	End			
1	1	20	ND	ND	20	ND	?
2	21	98	1924	2,005	78	>5,000	c0947002989
3	99	567	12990	13457	469	>4,000	c072403314
4	568	737	9836	10005	170	2,985	c072403314
5	738	954	9181	9397	217	439	c072403314
6	955	1150	7361	7556	196	1,632	c072403314
7	1151	1247	7048	7144	97	217	c072403314
8	1248	1433	3896	4081	186	2,967	c072403314
9	1434	1586	2068	2220	153	1,676	c072403314
10	1587	1719	913	1045	133	1,023	c072403314
11	1720	1841	8928	9049	122	>10,000	c073001924
12	1842	2073	6437	6668	232	2,260	c073001924
13	2074	2180	4617	4723	107	1,714	c073001924
14	2181	2299	1625	1742	119	2,875	c073001924
15	2300	2430	15414	15544	131	>2,500	c052402941
16	2431	2580	13831	13971	150	1,443	c052402941
17	2581	2772	1807	2008	192	11,823	c052402941
18	2773	2798	5262	5293	26	>7,000	203795
19	2799	2919	4833	4953	121	>10,000	2033793
20*	**2920**	**3164**	9870	10112	245	>11,000	c047902506
21	2920/**3165**	2975/**3220**	9691	9746	56	124	c047902506
22	2976/**3221**	3138/**3383**	7465	7627	163	>10,000	c030002862
23	3139/**3384**	3236/**3481**	4358	4455	98	3,010	c030002862
24*	3237	3631	ND	ND	395	>4,000	c114704732
25	3237/**3482**, 3632	3295/**3541**, 3691	2798	2856	60	1,999	c114704732
26	3296/3692	3398/3794	4630	4732	103	1,774	c114704732
27*	*3399/3795	4751/4269	500	1824/951	1,353/475		c098905234

[a] The full-length mouse *Ahi-1* cDNA (isoform I) consists of 25 exons, from 1 to 27, excluding exons 20 and 24; EST clone BB615071 (isoform II) appears to contain 24 exons, from 1 to 25, excluding exon 24; EST clone BG297436 and clone I (isoform III) may consist of 26 exons, from 1 to 27, excluding exon 20. The sizes of exons 20 and 24 are approximately determined. ND, not determined; ?, contigs containing exon 1 were not available in the Genbank database; boldfacing, cDNA position in isoform II; underlining, cDNA position in isoform III.

[b] *, Exons 20, 24, and 27 contain the termination codon in isoforms II, III, and I, respectively.

29, and 32). The first three exons were cloned in our laboratory by RT-PCR from Jurkat cells (Fig. 5). The termination codon is located in exon 33. The clone AK000076 (isoform II) lacking the SH3 domain appears to be made of 24 exons (i.e., exons 1 to 24) and contain an in-frame stop codon in exon 24. The clone AK024085 (isoform III) consists of 32 exons (i.e., exons 1 to 33, excluding exon 24) with a termination codon present in exon 28.

Mapping of the Mis-2 provirus insertion sites within the 16th intron of Ahi-1. We have previously identified Mis-2 as a common provirus insertion site in Moloney MuLV-induced rat thymomas (62) and found that it maps to mouse chromosome 10. Long-range restriction enzyme mapping analysis revealed that *Mis-2* mapped at a distance of ca. 160 kbp of the *Ahi-1* insertion site (62). Considering the length of the human and mouse *Ahi-1* genes, we predicted that the *Mis-2* integration sites would map within the *Ahi-1* gene itself. To test this hypothesis, we sequenced the rat genomic region adjacent to *Mis-2* integration sites in tumor 1213-2 and 1213-19 (Fr 1 to 5, Fig. 6A) (62). This analysis confirmed the site of the integration and revealed a high degree of homology of the rat *Ahi-1* sequences with those of the mouse *Ahi-1* around exons 14, 15, and 16 (contig D, Fig. 6). These results, combined with our previous mapping of the *Mis-2* provirus integration sites, indicated that the *Mis-2* locus corresponds to insertions within the intron 16 of the *Ahi-1* gene.

Expression of the *Ahi-1* gene in rodent tissues. Northern blot analysis was performed on RNA extracted from various tissues of adult rat and mouse. This analysis revealed the presence of Ahi-1 transcripts hybridizing with a probe derived from the 5′ end of mouse *Ahi-1* cDNA (probe E1.3) (Fig. 2B) in several tissues (Fig. 7A). Ahi-1 expression was much higher in the brain and testis than in other organs but was especially low in the liver. In addition, various Ahi-1 RNA species were present in different tissues, suggesting different splicing variants: a major hybridizing species of 5 kb in several mouse tissues tested and a less intensely hybridizing RNA of 3.5 kb in brain, 4.2 and 3.5 kb in testis, and 2 kb in embryos (Fig. 7A). In the rat thymus (Fig. 7B, lane 4) and in the brain and testis (data not shown), two major species of 5 and 2 kb could be detected with a 3′ end rat *Ahi-1* cDNA clone A (Fig. 2A) as a probe. In rat spleen, shorter (1.2 and 0.7 kb) and more abundant RNA species were detected (data not shown). These results indicated that the *Ahi-1* gene expresses several RNA species, a finding consistent with the various spliced variants cDNA identified (Fig. 5).

Expression of Ahi-1 in tumors with provirus insertions. To determine whether provirus integration at the *Mis-2* or the *Ahi-1* locus led to deregulation of the *Ahi-1* gene expression, we first examined *Ahi-1* expression in T- and B-cell tumors by Northern blot analysis with a 3′ end rat cDNA probe (probe B) or 5′ end mouse cDNA probe (probe E1.3) (Fig. 2), respec-

TABLE 2. Exon-Intron boundaries of the human *Ahi-1* gene[a]

Exon[b]	cDNA position		Genomic position		Exon size (bp)	Intron size (bp)
	Start	End	Start	End		
1	1	61	6014930	6014990	61	
2	62	130	6015330	6015391	69	340
3	131	218	6016682	6016766	88	1291
4	219	245	6020325	6020351	27	3559
5	246	370	6021832	6021956	125	1481
6	371	424	6044945	6044998	54	22989
7	425	984	6046206	6046765	560	1208
8	985	1166	6049273	6049454	182	2508
9	1167	1386	6054866	6055085	220	5412
10	1387	1579	6056653	6056845	193	1568
11	1580	1675	6059142	6059238	96	2297
12	1676	1861	6064104	6064289	186	4866
13	1862	2014	6065419	6065571	153	1130
14	2015	2147	6069865	6069997	133	4294
15	2148	2271	6074081	6074204	124	4084
16	2272	2501	6079323	6079552	230	5119
17	2502	2608	6081265	6081371	107	1713
18	2609	2727	6082579	6082697	119	1208
19	2728	2858	6083820	6083950	131	1123
20	2859	2999	6085272	6085412	141	1322
21	3000	3248	6101035	6101231	249	15623
22	3249	3275	6107602	6107628	27	6371
23	3276	3396	6117683	6117803	121	10055
24*	**3397**	**3645**	6124546	6124794	249	6743
25	3396	3451	6154392	6154447	56	29598
26	3452	3614	6189255	6189417	163	34808
27	3615	3712	6193963	6194060	98	4546
28*	3713	3818	6209762	6209867	106	15702
29	3819	3950	6211040	6211171	132	1173
30	3713/3951	3771/4009	6212021	6212079	59	850
31	3772/4010	3874/4112	6222054	6222156	103	9975
32	4113	4262	6226031	6226180	150	3875
33*	3875/4263	5535/5016	6226932	6228592	1661/754	752

[a] The full-length human *Ahi-1* (isoform I) consists of 29 exons, from 1 to 33, excluding exons 24, 28, 29, and 32; clone AK000076 (isoform II) seems to contain 24 exons from 1 to 24; clone AK02485 (isoform III) seems to contain 32 exons from 1 to 33, excluding exon 24. All exons are located on the same contig (NT_025741). Boldfacing, cDNA position in isoform II; underlining, cDNA position in isoform III.

[b] *, Exons 24, 28, and 33 contain the termination codon in isoforms II, III, and I, respectively.

tively. Many tumors showing rearrangement in a very small percentage of the tumor cells [as judged by the intensity of the rearranged *Ahi-1* or Mis-2 fragments relative to the germ line fragment (41, 62)] were not suitable for this analysis. Therefore, only tumors harboring a significant proportion of cells showing rearrangement were tested.

In the RNA of Moloney MuLV-induced rat tumors 1213-2 and 1213-19 rearranged in *Mis-2*, four major *Ahi-1* transcripts (5, 4.2, 2.5, and 1.2 kb) were detected. This pattern of expression was the same in rearranged and nonrearranged tumors, but it was distinct from that of the thymus where the 1.2- or 2.5-kb transcripts were, respectively, not detected or less abundant, whereas another 2-kb species was abundant (data not shown). In the rat T-cell thymoma (i.e., 1213-10) which had a provirus inserted in the *Ahi-1* locus, the levels of several RNA species (especially the 1.2-kb form) were higher than those found in unrearranged tumors (Fig. 7B). Analysis of *Ahi 1* expression in A-MuLV-induced pre-B cell tumors or in T-cell thymomas from Moloney MuLV-infected $MMTV^{D}$/myc Tg mice, which had a rearranged *Ahi-1* locus, showed a modest increase of *Ahi-1* 5-kb transcript in some, but not all, of these tumors, compared to levels detected in unrearranged tumors (data not shown). Interestingly, some of the tumors also showed higher levels of shorter Ahi-1 transcripts (~4.2, 4, and 3.5 kb) (Fig. 7C). The levels of the shorter 4.2-kb *Ahi-1* transcripts were especially high in the L48 cell line (Fig. 7D, lane 3). Therefore, Northern analysis revealed that the integration of a provirus at the 3′ end or in exon 16 of the *Ahi-1* gene appears to affect only modestly the levels of its transcription in some tumors but significantly in two of the tested tumors.

Since the provirus integration within intron 16 of *Ahi-1* (*Mis-2* locus) would be expected to cause truncation of the gene, as we previously reported for similar intronic provirus integrations within the *Notch-1* gene (16), RT-PCR analysis was also performed to detect such truncated transcripts. For identification of transcripts originating at the 5′ end of the gene and terminating at the site of provirus integration, we used primer 276 within exon 16 with primer 828 (i.e., 1213-9) or primer 110 (i.e., 1213-2) within the LTR. This analysis indeed revealed the presence of truncated *Ahi-1* transcripts deleted of their half 3′ end and fused to viral LTR sequences, with the generation of a translation termination codon in tumor 1213-19 RNA (Fig. 7E) but not in tumor 1213-2 RNA (data not shown). In tumor 1213-19 RNA, exon 16 is properly spliced and the provirus insertion has forced the utilization of a cryptic splice acceptor site in the Ahi-1 intron, 22 nt just

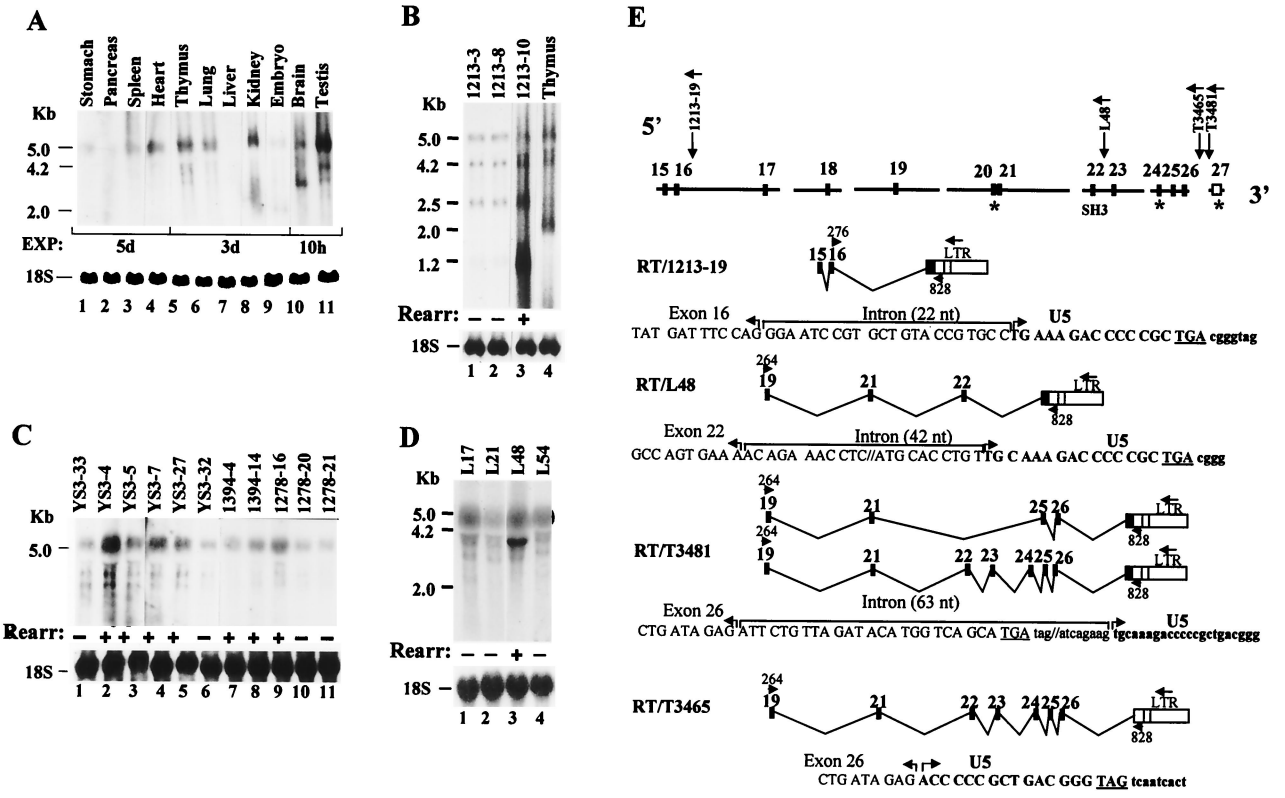

FIG. 7. Expression of the *Ahi-1* gene in normal tissues and tumors: (A) RNAs (20 μg) from various organs (lanes 1 to 11, as indicated) of a C57BL/6 mouse were screened by Northern blot analysis with *Ahi-1* cDNA clone-E1.3 as a probe. Exposure time (EXP) is indicated in hours (h) or days (d). (B) RNAs (20 μg) from Moloney MuLV-induced rat thymomas 1213-3, 1213-8, and 1213-10 (lanes 1 to 3) and rat thymus (lane 4) were hybridized with the ^{32}P-labeled rat cDNA probe-B. (C and D) Northern blot analysis of RNAs (20 μg) from A-MuLV-induced pre-B cell tumors (lanes 1 to 11) (C) or from cell lines established from thymomas of MMTV^{D}/myc Tg mice infected with Moloney MuLV (D). Hybridization was with the mouse *Ahi-1* cDNA probe E1.3. (E) Analysis of truncated *Ahi-1* transcripts. RT-PCR amplification was performed on tumor RNAs (5 μg) from various tumors showing *Ahi-1* rearrangement. RNAs were reversed transcribed in the presence of random hexamers. The cDNAs were then amplified by PCR with the indicated primers specific for the *Ahi-1* sequences and the viral LTR. The amplified fragments were separated on 1% agarose gels, purified, cloned, and sequenced. In panels A to D, the blots presented were rehybridized with a 18S ribosomal probe (bottom). In panels B to D, the presence (+) or absence (−) of proviral rearrangement (Rearr) in tumor DNAs is indicated.

upstream of the LTR. In tumor 1213-2 DNA, the LTR is inserted in the same transcriptional orientation as that of *Ahi-1* and may work as a promoter. RT-PCR amplification with a U5 LTR primer 878 and primers 277 and 265 in exon 17 or 22 did not reveal such 5′ viral/Ahi-1 fused transcripts originating in the LTR (data not shown), suggesting that this promoter is silent.

Similar analysis with RNA from tumors having provirus integration at the 3′ end of the gene (*Ahi-1* locus) was also carried out. In tumors T3465 and T3481, this experiment showed Ahi-1/LTR fused transcripts with the proper utilization of the exon 26 splice donor site and use of a cryptic splice acceptor site in the viral U5 region (T3465) or in intronic sequences upstream of the LTR (T3481), thus splicing out exon 27 (Fig. 7E). These transcripts would be expected to encode a full-length protein. In the same tumor T3481, another interesting truncated Ahi-1/LTR splice variant transcript species was generated by splicing out exons 20, 22, 23, 24, and 27. These transcripts would be expected to encode truncated Ahi-1 proteins deleted of most of their C terminus, notably of their SH3 domain. In tumor L48, this analysis with primer pairs in exon 19 and in LTR detected Ahi-1/LTR fused transcripts deleted of exons 23 to 27, with the use of a cryptic splice acceptor site in the sequences of intron 22. These transcripts would be expected to encode truncated Ahi-1 proteins deleted of their C terminus but still harboring the SH3 domain. Together, these results show that insertion of provirus within the *Ahi-1* gene leads to the generation of distinct species of truncated RNA.

DISCUSSION

The *Ahi-1* gene targeted by provirus insertional mutations encodes a modular protein. We have identified a novel gene (*Ahi-1*) mutated by provirus insertions in MuLV-induced tumors at two distinct sites, designated Ahi-1 and Mis-2. The 3′-end Ahi-1 provirus insertion site comprises two clusters: the more proximal cluster, with proviruses inserted within the last introns, the last noncoding exon, or within sequences just downstream of the last exon, and the distal cluster located ~16 kbp downstream of the last noncoding exon. Most of these Ahi-1 provirus insertions are located outside the coding region

of *Ahi-1*. Surprisingly, all of the mapped proviruses were found to be in an inverse transcriptional orientation relative to that of the *Ahi-1* gene itself. The second provirus insertion site, Mis-2, is intronic (within the 16th intron) and is located at a large distance (~120 kbp) from the Ahi-1 insertion site. At this Mis-2 site, the proviruses are integrated both in the sense and antisense transcriptional orientation relative to that of the *Ahi-1* gene. Therefore, it appears that *Ahi-1* is the gene targeted by these insertional mutational events.

The structure of the predicted protein encoded by the *Ahi-1* gene is particularly interesting, containing an SH3 domain, WD40 repeats and SH3-binding sites, all known to mediate specific protein-protein interactions. SH3 domains are known to bind to proline-rich motifs (24, 37, 52). Among the best-studied functionally important SH3 interactions is the complex formed between the Grb2/Sem5 SH3 domain and the GDP-GTP exchange protein Sos, which forms a link in a signal transduction pathway that functionally ties tyrosine kinase receptors to Ras (5, 11, 28, 30, 35, 54). Although the functional significance of the presence of a SH3 domain within Ahi-1 is not known at the present time, this protein clearly represents a novel member of the SH3-containing protein family.

The second most evident motif of the Ahi-1 protein is a WD40-repeat domain. Several proteins containing WD40 repeats have been identified (27, 33, 55, 58). These proteins are associated with multiple and diverse aspects of cellular metabolism, including signal transduction, regulation of cytoskeletal assembly, cell cycle regulation, RNA processing, programmed cell death, and gene regulation (33). In general, their functions appear to be regulatory in nature. Structural models suggest that the WD40-repeat-containing proteins are endowed with the potential for interactions with other proteins, a finding consistent with the fact that many such proteins are components of multiprotein complexes (33, 34). The Ahi-1 protein represents a novel member of this family. To our knowledge, Ahi-1 is the only protein reported to date harboring both WD40 repeats and a SH3 motif.

The third important series of motifs predicted in the Ahi-1 amino acid sequences are nine putative SH3-binding sites. Short (8 to 10 aa), proline-rich, and structurally conserved motifs were initially found to bind specifically to the SH3 domain of Abl (7, 43). These SH3-binding sites are present in a variety of proteins and appear to form two distinct classes: a class I motif containing a conserved N-terminal arginine and a class II motif containing a conserved C-terminal arginine (14, 29). The SH3-binding sites of Ahi-1 could mediate its interaction with other SH3-containing protein(s). One of these sites may even bind the SH3 domain of Ahi-1 itself, intramolecularly. Intramolecular interaction of the Src SH3 domain has been reported (12).

Finally, Ahi-1 harbors other motifs of interest, two potential tyrosine phosphorylation sites, one within the SH3 domain. Phosphotyrosine motifs are known to bind specifically to SH2 domains of proteins (24, 52). If the Ahi-1 protein can be phosphorylated at these sites by a tyrosine kinase, this phosphorylation may allow it to recruit and bind to other SH2-containing protein(s).

Therefore, Ahi-1 appears to be a modular protein containing several putative motifs that have been shown to be present in several signaling molecules and to mediate protein-protein interactions. It is tempting to suggest that Ahi-1 may play a role in signal transduction. Because of the high number of these putative motifs mediating protein-protein interactions, Ahi-1 may be a docking site or scaffold protein recruiting a number of other signaling molecules and modulating and integrating their action.

Ahi-1 provirus insertional mutations generate truncated forms of Ahi-1 RNA. The *Ahi-1* gene mutations that were analyzed lead to the generation of various forms of transcripts: some fused to viral sequences and some deleted of their 3′ end. This was expected with the provirus inserted within intron 16 (*Mis-2* locus), and such molecular rearrangement represents a mechanism very similar to the one we described previously with proviruses inserted within the *Notch1* gene (16, 20). However, the generation of similar truncated transcripts in tumors harboring provirus integrated at the 3′ end of the gene was more surprising. Some of these 3′-end insertional mutations seem to alter the normal splicing machinery and to force an alternative splicing within viral or intronic sequences close to them, thus deleting exons coding for the C terminus of Ahi-1. The molecular requirements for such unusual consequences of provirus integrations at the 3′ end of a gene remain obscure but may involve the opposite transcriptional orientation of the inserted proviruses relative to that of the *Ahi-1* gene. Indeed, all of the 27 proviruses inserted within the *Ahi-1* locus (also recently designated Epi1) are in the same direction (Fig. 1) (2, 41), and we show here that this is in opposite transcriptional orientation relative to that of *Ahi-1*. Future work on how such insertional mutants may affect the splicing machinery should be informative. Interestingly, some of the truncated transcripts generated in tumors harboring Mis-2 or Ahi-1 provirus insertions appear to have a similar coding capacity, both having the capacity to code for Ahi-1 proteins truncated of some C-terminal sequences, including of their SH3 domain.

How is the *Ahi-1* gene mutation involved in tumor development? Genetic evidence suggests that provirus insertional mutation of the *Ahi-1* gene contributes to tumor formation in different cell lineages. The *Ahi-1* locus was identified as a common provirus integration site initially in Abelson (v-*abl*-induced) pre-B lymphomas (41) and later in c-*myc* induced T-cell lymphomas of the $MMTV^{D}$/myc Tg mice (Fig. 1), suggesting that it may collaborate with these oncogenes. This locus has also been found by others to be rearranged by provirus insertion in other types of tumors, namely, in pre B-cell tumors of Eμ/myc Tg mice (18, 61) and in acute myeloid leukemia in *Nf1* heterozygous mice (2). Similarly, the *Mis-2* locus was identified as a common provirus insertion site in Moloney MuLV-induced rat T-cell lymphomas (thymomas) (62). The fact that these provirus insertional mutations are not random and were identified in a relatively high proportion of tumors (~3% for Mis-2, ~15% for Ahi-1 [Fig. 1] [41], and 44% for Epi1 [2]) suggests that they have been selected during the oncogenic process and that they may be involved in this process. This notion is reinforced by the several examples of common provirus insertional sites that have been found to activate protooncogenes in nondefective retrovirus-induced tumors (26, 39, 47, 60).

Molecular evidence also suggests that mutation of the *Ahi-1* gene may be involved in tumor formation. Our analysis of RNA in tumors harboring mutations in the *Ahi-1* gene suggests

that the mechanism by which these insertional mutations may enhance the oncogenic potential of the *Ahi-1* gene is through the generation of truncated forms of the *Ahi-1* RNA with the capacity to code for truncated Ahi-1 proteins. The levels of these truncated RNA are not very high in the *Ahi-1* rearranged tumors analyzed but are specific to these tumors and are not found in nonrearranged tumors. These are likely to be involved in the transformation process. The deletion of sequences coding for the C terminus, including for the SH3 domain in some truncated *Ahi-1* RNA found in tumors is likely to significantly affect the interaction of Ahi-1 with other proteins, possibly converting this molecule to a dominant-negative mutant. Alternatively, truncated Ahi-1 proteins may represent gain-of-function mutants, if the SH3 domain binds intramolecularly, as reported for other SH3-containing molecules (12), or bind to an inhibitor. In nonreceptor tyrosine kinases, such as the proto-oncogene c-*src* or c-*abl*, deletion or mutation of the SH3 domain generally leads to the oncogenic activation of their tyrosine kinase, suggesting that the SH3 domain binds to an inhibitor of the kinase activity (9, 15, 21, 32, 38, 53, 59, 63). Thus, the SH3 deletion in Ahi-1 may uncover some domains and allow novel protein interactions or may prevent the binding of an inhibitor. The fact that *Ahi-1* cDNA with the potential to encode very similar truncated proteins as the ones detected in tumors have been isolated in libraries from normal tissues, suggests that truncated Ahi-1 proteins may not be oncogenic in each tissue. Rather the inappropriate or higher expression of truncated Ahi-1 proteins in some specific cell types may contribute to the transformation process.

Because we were unable to detect truncated transcripts nor Ahi-1/viral fused transcripts in some *Ahi-1* rearranged tumors, it is also quite possible that the *Ahi-1* gene may be involved in oncogenesis by mechanism(s) not yet uncovered and distinct from the generation of truncated gene products. In addition, we cannot exclude the possibility that some proviruses inserted into the *Ahi-1* locus may act by other, as-yet-uncharacterized processes or may affect the expression of other genes surrounding Ahi-1 at a distance. This latter mechanism has recently been proposed for other provirus insertion sites upstream of c-*myb* (19).

It will be interesting to investigate how the *Ahi-1* gene, apparently targeted by these proviruses and coding for a protein that shows all of the characteristics of a modular signaling molecule involved in protein-protein interactions, participate in oncogenesis and what is its normal function.

ACKNOWLEDGMENTS

X.J, Z.H, and M.K. contributed equally to this study.

This work was supported by grants from the Medical Research Council of Canada and the National Cancer Institute of Canada to P.J. X.J. was the recipient of a fellowship from McGill University and from Hydro-Quebec.

We thank Alan J. Buckler for providing the pSPL1 vector. We thank Benoît Laganière and Patrick Couture for excellent technical assistance. We are also grateful to Rita Gingras for preparing the manuscript.

REFERENCES

1. **Abelson, H. T., and L. S. Rabstein.** 1970. Lymphosarcoma: virus-induced thymic-independent disease in mice. Cancer Res. **30:**2213–2222.
2. **Blaydes, S. M., S. C. Kogan, B. T. Truong, D. J. Gilbert, N. A. Jenkins, N. G. Copeland, D. A. Largaespada, and C. I. Brannan.** 2001. Retroviral integration at the Epi1 locus cooperates with Nf1 gene loss in the progression to acute myeloid leukemia. J. Virol. **75:**9427–9434.
3. **Boss, M., M. Greaves, and N. Teich.** 1979. Abelson virus-transformed haematopoietic cell lines with pre-B-cell characteristics. Nature **278:**551–553.
4. **Buckler, A. J., D. D. Chang, S. L. Graw, J. D. Brook, D. A. Haber, P. A. Sharp, and D. E. Housman.** 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA **88:**4005–4009.
5. **Chardin, P., J. H. Camonis, N. W. Gale, L. van Aelst, J. Schlessinger, M. H. Wigler, and D. Bar-Sagi.** 1993. Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2. Science **260:**1338–1343.
6. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
7. **Cicchetti, P., B. J. Mayer, G. Thiel, and D. Baltimore.** 1992. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. Science **257:**803–806.
8. **Cooper, J. A., F. S. Esch, S. S. Taylor, and T. Hunter.** 1984. Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. J. Biol. Chem. **259:**7835–7841.
9. **Dai, Z., and A. M. Pendergast.** 1995. Abi-2, a novel SH3-containing protein, interacts with the c-Abl tyrosine kinase and modulates c-Abl transforming activity. Genes Dev. **9:**2569–2582.
10. **Dingwall, C., and R. A. Laskey.** 1991. Nuclear targeting sequences: a consensus? Trends Biochem. Sci. **16:**478–481.
11. **Egan, S. E., B. W. Giddings, M. W. Brooks, L. Buday, A. M. Sizeland, and R. A. Weinberg.** 1993. Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation. Nature **363:**45–51.
12. **Erpel, T., G. Superti-Furga, and S. A. Courtneidge.** 1995. Mutational analysis of the Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. EMBO J. **14:**963–975.
13. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
14. **Feng, S., J. K. Chen, H. Yu, J. A. Simon, and S. L. Schreiber.** 1994. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science **266:**1241–1247.
15. **Franz, W. M., P. Berger, and J. Y. Wang.** 1989. Deletion of an N-terminal regulatory domain of the c-abl tyrosine kinase activates its oncogenic potential. EMBO J. **8:**137–147.
16. **Girard, L., Z. Hanna, N. Beaulieu, C. D. Hoemann, C. Simard, C. A. Kozak, and P. Jolicoeur.** 1996. Frequent provirus insertional mutagenesis of Notch1 in thymomas of $MMTV^D$/myc transgenic mice suggests a collaboration of c-myc and Notch1 for oncogenesis. Genes Dev. **10:**1930–1944.
17. **Green, P. L., D. A. Kaeler, and R. Risser.** 1987. Clonal dominance and progression in Abelson murine leukemia virus lymphomagenesis. J. Virol. **61:**2192–2197.
18. **Haupt, Y., W. S. Alexander, G. Barri, S. P. Klinken, and J. M. Adams.** 1991. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in Eμ-myc transgenic mice. Cell **65:**753–763.
19. **Haviernik, P.** 2002. Linkage on chromosome 10 of several murine retroviral integration loci associated with leukaemia. J. Gen. Virol. **83:**819–827.
20. **Hoemann, C. D., N. Beaulieu, L. Girard, N. Rebai, and P. Jolicoeur.** 2000. Two distinct *Notch1* mutant alleles are involved in the induction of T-cell leukemia in c-*myc* transgenic mice. Mol. Cell. Biol. **20:**3831–3842.
21. **Jackson, P., and D. Baltimore.** 1989. N-terminal mutations activate the leukemogenic potential of the myristoylated form of c-*abl*. EMBO J. **8:**449–456.
22. **Jiang, X., L. Villeneuve, C. Turmel, C. A. Kozak, and P. Jolicoeur.** 1994. The Myb and Ahi-1 genes are physically very closely linked on mouse chromosome 10. Mammalian Genome **5:**142–148.
23. **Kemp, B. E., and R. B. Pearson.** 1990. Protein kinase recognition sequence motifs. Trends Biochem. Sci. **15:**342–346.
24. **Koch, C. A., D. Anderson, M. F. Moran, C. Ellis, and T. Pawson.** 1991. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. Science **252:**668–674.
25. **Kozak, M.** 1986. Point mutations define a sequence flanking the aug initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:**283–292.
26. **Kung, H. J., C. Boerkoel, and T. H. Carter.** 1991. Retroviral mutagenesis of cellular oncogenes: a review with insights into the mechanisms of insertional activation. Curr. Top. Microbiol. Immunol. **171:**1–25.
27. **Li, D., and R. Roberts.** 2000. WD-repeat proteins:structure characteristics-,biological function, and their involvment in human diseases. Cell. Mol. Life Sci. **58:**2085–2097.
28. **Li, N., A. Batzer, R. Daly, V. Yajnik, E. Skolnik, P. Chardin, D. Bar-Sagi, B. Margolis, and J. Schlessinger.** 1993. Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signaling. Nature **363:**85–88.
29. **Lim, W. A., F. M. Richards, and R. O. Fox.** 1994. Structural determinants of

peptide-binding orientation and of sequence specificity in SH3 domains. Nature **372:**375–379.
30. **Lowenstein, E. J., R. J. Daly, A. G. Batzer, W. Li, B. Margolis, R. Lammers, A. Ullrich, E. Y. Skolnik, D. Bar-Sagi, and J. Schlessinger.** 1992. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. Cell **70:**431–442.
31. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154–1156.
32. **Mayer, B. J., and D. Baltimore.** 1994. Mutagenic analysis of the roles of SH2 and SH3 domains in regulation of the Abl tyrosine kinase. Mol. Cell. Biol. **14:**2883–2894.
33. **Neer, E. J., C. J. Schmidt, R. Nambudripad, and T. F. Smith.** 1994. The ancient regulatory-protein family of WD-repeat proteins. Nature **371:**297–300.
34. **Neer, E. J., and T. F. Smith.** 1996. G protein heterodimers: new structures propel new questions. Cell **84:**175–178.
35. **Olivier, J. P., T. Raabe, M. Henkemeyer, B. Dickson, G. Mbamalu, B. Margolis, J. Schlessinger, E. Hafen, and T. Pawson.** 1993. A *Drosophila* SH2-SH3 adaptor protein implicated in coupling the sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. Cell **73:**179–191.
36. **Paquette, Y., L. Doyon, A. Laperrière, Z. Hanna, J. Ball, R. P. Sekaly, and P. Jolicoeur.** 1992. A viral long terminal repeat expressed in $CD4^+$ $CD8^+$ precursors is downregulated in mature peripheral $CD4^-$ $CD8^+$ or $CD4^+$ $CD8^-$ T cells. Mol. Cell. Biol. **12:**3522–3530.
37. **Pawson, T., and G. D. Gish.** 1992. SH2 and SH3 domains: from structure to function. Cell **71:**359–362.
38. **Pendergast, A. M., A. J. Muller, M. H. Havlik, R. Clark, F. McCormick, and O. N. Witte.** 1991. Evidence for regulation of the human ABL tyrosine kinase by a cellular inhibitor. Proc. Natl. Acad. Sci. USA **88:**5927–5931.
39. **Peters, G.** 1990. Oncogenes at viral integration sites. Cell Growth Differ. **1:**503–510.
40. **Pierotti, A. R., A. Prat, V. Chesneau, F. Gaudoux, A. M. Leseney, T. Foulon, and P. Cohen.** 1994. N-arginine dibasic convertase, a metalloendopeptidase as a prototype of a class of processing enzymes. Proc. Natl. Acad. Sci. USA **91:**6078–6082.
41. **Poirier, Y., C. Kozak, and P. Jolicoeur.** 1988. Identification of a common helper provirus integration site in Abelson murine leukemia virus-induced lymphoma DNA. J. Virol. **62:**3985–3992.
42. **Potter, M., M. D. Sklar, and W. P. Rowe.** 1973. Rapid viral induction of plasmacytomas in pristane-primed BALB-c mice. Science **182:**592–594.
43. **Ren, R., B. J. Mayer, P. Cicchetti, and D. Baltimore.** 1993. Identification of a ten-amino acid proline-rich SH3 binding site. Science **259:**1157–1161.
44. **Rickles, R. J., M. C. Botfield, X. M. Zhou, P. A. Henry, J. S. Brugge, and M. J. Zoller.** 1995. Phage display selection of ligand residues important for *src* homology 3 domain binding specificity. Proc. Natl. Acad. Sci. USA **92:**10909–10913.
45. **Risser, R., M. Potter, and W. P. Rowe.** 1978. Abelson virus-induced lymphomagenesis in mice. J. Exp. Med. **148:**714–726.
46. **Rogers, S., R. Wells, and M. Lechsteiner.** 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science **234:**364–368.
47. **Rosenberg, N., and P. Jolicoeur.** 1997. Retroviral Pathogenesis, p. 475–585. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
48. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **Sanger, F., and A. R. Coulson.** 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **94:**441–448.
50. **Savard, P., L. DesGroseillers, E. Rassart, Y. Poirier, and P. Jolicoeur.** 1987. Important role of the long terminal repeat of the helper Moloney murine leukemia virus in Abelson virus-induced lymphoma. J. Virol. **61:**3266–3275.
51. **Scher, C. D., and R. Siegler.** 1975. Direct transformation of 3T3 cells by Abelson murine leukaemia virus. Nature **253:**729–731.
52. **Schlessinger, J.** 1994. SH2/SH3 signaling proteins. Curr. Opin. Genet. Dev. **4:**25–30.
53. **Seidel-Dugan, C., B. E. Meyer, S. M. Thomas, and J. S. Brugge.** 1992. Effects of SH2 and SH3 deletions on the functional activities of wild-type and transforming variants of c-Src. Mol. Cell. Biol. **12:**1835–1845.
54. **Simon, M. A., G. S. Dodson, and G. M. Rubin.** 1993. An SH3-SH2-SH3 protein is required for p21Ras1 activation and binds to sevenless and Sos proteins in vitro. Cell **73:**169–177.
55. **Smith, T. F., C. Gaitatzes, K. Saxena, and E. J. Neer.** 1999. The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. **24:**181–185.
56. **Sparks, A. B., J. E. Rider, N. G. Hoffman, D. M. Fowlkes, L. A. Quillam, and B. K. Kay.** 1996. Distinct ligand preferences of src homology 3 domains from src, yes, abl, cortactin, p53bp2, plcgamma, crk, and grb2. Proc. Natl. Acad. Sci. USA **93:**1540–1544.
57. **Ullrich, A., and J. Schlessinger.** 1990. Signal transduction by receptors with tyrosine kinase activity. Cell **61:**203–212.
58. **van der Voorn, L., and H. L. Ploegh.** 1992. The wd-40 repeat. FEBS Lett. **307:**131–134.
59. **Van Etten, R. A., P. Jackson, and D. Baltimore.** 1989. The mouse type IV c-*abl* gene product is a nuclear protein, and activation of transforming ability is associated with cytoplasmic localization. Cell **58:**669–678.
60. **van Lohuizen, M., and A. Berns.** 1990. Tumorigenesis by slow-transforming retroviruses: an update. Biochem. Biophys. Acta **1032:**213–235.
61. **van Lohuizen, M., S. Verbeek, B. Scheijen, E. Wientjens, H. van der Gulden, and A. Berns.** 1991. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. Cell **65:**737–752.
62. **Villeneuve, L., X. Jiang, C. Turmel, C. A. Kozak, and P. Jolicoeur.** 1993. Long-range mapping of Mis-2, a common provirus integration site identified in murine leukemia virus-induced thymomas and located 160 kilobase pairs downstream of Myb. J. Virol. **67:**5733–5739.
63. **Weng, Z., S. M. Thomas, R. J. Rickles, J. A. Taylor, A. W. Brauer, C. M. W. Seidel-Dugan, G. Dreyfuss, and J. S. Brugge.** 1994. Identification of Src, Fyn, and Lyn SH3-binding proteins: implications for a function of SH3 domains. Mol. Cell. Biol. **14:**4509–4521.
64. **Yu, H., J. K. Chen, S. Feng, D. C. Dalgarno, A. W. Brauer, and S. L. Schreiber.** 1994. Structural basis for the binding of proline-rich peptides to SH3 domains. Cell **76:**933–945.