Linkage on chromosome 10 of several murine retroviral integration loci associated with leukaemia

Peter Haviernik,1 Stephen M. Festin,1† Rene Opavsky,1‡ Richard P. Koller,1 Nighean I. Barr,2 James C. Neil2 and Linda Wolff1

1 Leukemogenesis Section, Laboratory of Cellular Oncology, National Cancer Institute, NIH, Bethesda, MD 20892-4255, USA
2 Molecular Oncology Laboratory, University of Glasgow, Glasgow G61 1QH, UK

Mml loci have been identified as provirus integration sites among a subset of monocytic tumours induced by murine leukaemia virus (MuLV) infection of BALB/c and DBA/2 mice. These myeloid leukaemias contain a retrovirus integrated on chromosome 10 in proximity to the c-myb locus; however, c-myb expression was not altered. Detailed physical mapping enabled placement of the retroviral integration sites ~25 kb (Mml1), ~51 kb (Mml2), and ~70 kb (Mml3) upstream of the c-myb locus. Furthermore, the Fti1 (fit-1) locus, a common integration site in feline leukaemia virus-induced T cell lymphomas, was mapped upstream of Mml3. Sequence analysis of Mml1, Mml2 and Mml3 loci (39–6, 16–4 and 5–9 kb, respectively) in conjunction with the BLAST (basic local alignment search tool) homology searches against the expressed sequence tag (EST) database and the use of gene/exon prediction programs revealed potential coding sequences that were not confirmed by Northern analysis or RT–PCR. The sequences between c-myb and Fti1, which were shown to include two potential scaffold/matrix attachment regions (S/MARs), are most likely regulatory in nature. An extended search for transcribed sequences far upstream of Mml3 revealed five genes, four of which were expressed in multiple tissues in mice. These genes could not be linked to tumour formation by the virus but their homologous sequences were found on human chromosome 6, thus allowing extension of the syntenic region on mouse chromosome 10 to approximately 250 kb.

Introduction

During leukaemogenesis, haematopoietic cell growth and differentiation is altered by changes in expression or function of proto-oncogenes or tumour suppressors. One approach that has been utilized to identify genes that are involved in the oncogenic process is retroviral insertional mutagenesis (Jonkers & Berns, 1996; Kung et al., 1991). We have employed this approach to detect the genetic alterations involved in monocytic leukaemia through the development of a unique in vivo model. Intravenous inoculation of murine leukaemia virus (MuLV), either Moloney MuLV or amphotropic virus 4070A, in conjunction with pristane, results in insertional mutagenesis of haematopoietic progenitor cells and the ultimate outgrowth of leukaemic cells in the peritoneal cavity (Nason-Burchenal & Wolff, 1993; Wolff et al., 1991; Wolff et al., 1988). These tumours, called MML (murine leukaemia virus-induced myeloid leukaemia), are transplantable and can be cultivated in vitro as well. In the absence of virus, pristane-treated mice never develop these neoplasms.

Using this model, several loci of retrovirus integration have been identified. The most frequent target is c-myb (Mukhopadhyaya & Wolff, 1992; Nazarov & Wolff, 1995; Shen-Ong & Wolff, 1987; Wolff et al., 1991). Its locus is rearranged and its expression at the RNA level is constitutive due to the activity of the retrovirus promoter and/or enhancer. However, many of the leukaemias induced by the 4070A virus express normal c-myb and have integrated proviruses outside

---

Author for correspondence: Linda Wolff.
Fax +1 301 594 3996. e-mail lwolff@helix.nih.gov

† Present address: Department of Biology, Hamilton College, Clinton, NY 13323, USA.

‡ Present address: Division of Human Cancer Genetics, Ohio State University, Columbus, OH 43210, USA.
the c-myb locus. In some of these tumours insertion sites were found to be clustered on chromosome 10, approximately 25 kb upstream of c-myb in a locus named Mml1 (Koller et al., 1996).

Recently, we identified two other sites of integration, Mml2 and Mml3, in MML. Here we have reported their position 50–70 kb upstream of c-myb, as well as a physical map linking Mml1, Mml2, Mml3 and Fti1, and a comparison of this region to an analogous region on human chromosome 6. Interestingly, this chromosomal region seems to be a preferred area for retroviral integrations that contribute to leukemogenesis since Fti1, a common integration site in feline leukaemia virus (FeLV)-induced T cell lymphomas, maps to this region (Barr et al., 1999). Another integration site, Ahi1, in Abelson MuLV-induced lymphomas is located downstream of c-myb (Jiang et al., 1994). For this study, we set out to determine if, in a 250 kb region upstream of c-myb, there are any genes that might be altered in expression or function thus contributing to the neoplastic disease. We concluded that, although there are several genes more than 100 kb upstream of c-myb, none appeared to be related to the oncogenic process. The region of approximately 100 kb, surrounding Mml1, 2, 3 and Fti1, is devoid of genes and presumably contains sequences involved in chromosomal structure or regulation of gene expression at a distance.

**Methods**

- **Cell cultures.** The murine monoblastic cell line M1 (Liebermann & Hoffman-Liebermann, 1989) was maintained in RPMI 1640 medium supplemented with 10% heat-inactivated horse serum (Life Technologies). The leukaemic cell lines were cultured in Dulbecco’s modified Eagle’s medium with 10% heat-inactivated horse serum (Life Technologies). For cell differentiation experiments, cells were seeded at a density of 1–10^5 cells/ml in medium containing IL-6. IL-6 stocks were prepared as described previously (Schmidt et al., 2000).

- **Primers and probes.** The amplifiers used for PCR amplification of the probes are listed in Table 1. Other probes used in this study were a c-myb genomic BglII fragment [B probe in Mukhopadhyaya & Wolff (1992)], an Ahi1 0.8 kb genomic Pst–HindIII fragment [Ahi-D probe in Poirier et al. (1988)] and an 4070A env (RK107) 1.5 kb EcoRI–Nhel fragment.

- **RNA isolation, Northern analysis and RT–PCR.** Total RNA was prepared using TRIZol Reagent (Gibco–BRL) according to the manufacturer’s instructions. Samples containing 10 µg total RNA were resolved on a 1.2% agarose gel containing 0.25 M formaldehyde in a MOPS buffer system. After transfer onto nylon membrane (Nytran SuPerCharge, Schleicher & Schuell) and UV cross-linking the blot was hybridized with random-priming labelled probes (Ready-To-Go DNA Labelling Beads, Amersham Pharmacia Biotech).

The RT–PCR analysis was performed using the Titan One Tube RT–PCR Kit (Roche) according to the manufacturer’s instructions.

- **Mml2, Mml3 cloning and library screening.** The Mml2 and Mml3 integration sites were cloned from the 30-2-9 and the 30-2-7 tumours, respectively. In the case of 30-2-9, genomic DNA was purified from primary tumour tissue and for 30-2-7 from an established cell line. Genomic DNA was digested with EcoRI and size-selected on a 0.8% agarose gel. Fragments from a range of about 3–5 kb were purified and ligated to the EcoRI cut arms of the ZAP Express vector. These λ subgenomic libraries were screened with a 4070A env probe (RK107). An insert from the positive Mml2 recombinant phage (22.1.1) was recloned into phbluecriptII SK(–) and sequenced. The Mml3 recombinant phage DNA (22.2-2) was sequenced directly. The genomic sequences flanking the provirus were used to design amplimers for RK114 (Mml2) or MML-3 (Mml3) PCR probes.

The mouse genomic BAC library RPCI-22 (female spleen 129S6/SvEvTac) was screened with 212,000 clones, of 154 kb average insert size, and approximately 10-fold genomic coverage. Children’s Hospital Oakland Research Institute, Oakland, CA, USA (Osoegawa et al., 2000) was screened by Southern hybridization with either an Mml1 (RK29) or an Mml2 (RK114) probe.

A genomic region containing the wild-type Mml2 locus was cloned using a BAC DNA positive clone from RPCI-22 library (#401-G9) by digestion with BamHI. Size-selected and purified fragments of 15–20 kb were ligated into the zDASHII vector. The subgenomic λ library was screened using an Mml2 PCR probe (RK114) and the insert from a positive phage (ZII-2a) was subcloned into a plasmid. Similarly, the Mml3 locus was cloned from the same BAC (#401-G9) after BamHI digestion and size selection. Fragments of 4–8 kb were cloned directly into the pBluescriptII SK(–) vector and the library was screened with an Mml3 PCR probe (MML-3) to obtain a positive clone (pSK-MML3).

**Table 1. Amplifiers used for PCR amplification of probes**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Probe</th>
<th>Amplimer</th>
</tr>
</thead>
</table>
| Mml1  | RK29  | a212 (5’ GAACAGTGTGGTGCTGCAAGTGC 3’)  
|       |       | a222 (5’ GTCAACATAGAGATGATG 3’)      |
| Mml2  | RK114 | a310 (5’ GGAATTGTCTCACAAGGCTGCAAG 3’)  
|       |       | a313 (5’ GTTCAGCTCAGATGTCGCTAC 3’)      |
|       | ML2D-D| a419 (5’ CTTCAGCTCAGATGTCGCTAC 3’)      
|       |       | a420 (5’ AAGCCCTAGACTGATCCCCTG 3’)      |
| Mml3  | MML-3 | a387 (5’ GAATTCCATCTGAGACTATAAGAAG 3’)  
|       |       | a388 (5’ TGAGATAAAGGCTCCTCGAGATTCT 3’)  
| Fti1  | mu flt-1 | mu flt-T (5’ GTGTAATTTGCTTCTGACTGACCTG 3’)  
|       |       | mu flt-1r (5’ GCCAAAGGGAAAAAGGCAACATGAGTCAC 3’)  

820
Results

Cloning and mapping of novel integration sites Mml2 and Mml3 on chromosome 10 between c-myb and Fti1

In 4070A retrovirus-induced promonocytic leukaemias in DBA/2 mice, c-myb is often the target of retrovirus insertional mutagenesis (Wolff et al., 1991). However, many tumours have retroviral integration sites outside the immediate c-myb locus. The cloning of another novel integration site, Mml1, rearranged due to proviral integration in ten tumours, was reported (Koller et al., 1996) and we later identified three additional tumours with alterations at this site (data not shown). Recently, proviral integration sites other than c-myb or Mml1 were identified in MML by Southern blot analysis using a virus env-specific probe. Two of these, 30-2-7 and 30-2-9, contain a single provirus integrated in their genome (Fig. 1). Genomic DNA comprising proviral and flanking sequences was cloned and probes were designed that were specific for the new cellular proviral integration sites, Mml2 and Mml3. Rearrangements at these loci have not been detected in any other tumours so far, but subsequent mapping near Mml1 and c-myb indicated that the effects of all of these integrations might be similar.

Screening of a mouse BAC genomic library, RP11-22, using an Mml1 probe, RK29 yielded several positive clones (C6, G9, K9, D17, K14, L15). Interestingly, they were found to hybridize with probes Mml2 and Mml3 (Fig. 2). In addition, clones A7 and J8 were identified as being positive for Mml2 and Mml3, but not Mml1 or c-myb (Fig. 2). Fti1 had previously been shown to be in the same chromosomal region as Mml1, although the distance between the integration sites was not determined (Barr et al., 1999). A Fti1 probe hybridized to all of the BAC clones shown in Fig. 2. We therefore concluded that all of the proviral integration sites Mml1, 2, 3 and Fti1 are located in the same chromosomal region. Using endonuclease digestion with several rare cutting enzymes, PFGE, blotting, and hybridization with integration site-specific probes, we constructed a physical map of this region as shown in Fig. 2. According to our analysis, Mml1, Mml2, Mml3 and Fti1 are situated approximately 25, 51, 70 and 81–90 kb, respectively, upstream of the c-myb gene.

Effect of Mml1 integration on c-myb expression

Since Mml1 is located near the c-myb locus and c-myb is activated in a large majority of monocytic tumours (Wolff, 1996), it is possible that proviral insertion in Mml1 has a positive influence on c-myb expression at the mRNA level. To
our surprise, in some of the tumours (30A-2-6 and 30C-13) c-myb was not expressed, while in others (30-3-14, 30-4-18, 30C-19) c-myb was expressed (Fig. 3A, B). In the instances where c-myb was expressed, we wondered whether its expression could be modulated in response to normal signalling. It is known that in myeloid cells, c-myb is expressed during proliferation and down-regulated during differentiation (Coll et al., 1983; Duprey & Boettiger, 1985; Liebermann & Hoffman-Liebermann, 1989; Sheiness & Gardinier, 1984; Wolff et al., 1996). If proviral insertion resulted in constitutive expression, it would be expected to be refractory to differentiation-associated down-regulation. We analysed a c-myb-positive Mml1 cell line 30-4-18 following treatment with IL-6, an inducer of differentiation. Down-regulation of c-myb occurred in these cells in response to the differentiation signal as it did in the M1 monoblastic cell line and a phenotypically similar tumour with virus integration in the 3' end of the c-myb locus (45-16). (Fig. 3B). In this case, deregulated transcription of c-myb could not be an explanation for the sustained proliferation of leukaemia 30-4-18.

Sequence, informatics and expression analysis

Because a correlation between integration in Mml1 and deregulated c-myb expression was not established, we decided to search for genes that could be potential oncogenes or tumour suppressor(s) in the general vicinity of Mml1, Mml2, Mml3 and Fti1. Probes obtained in close proximity to the integration sites (RK29, ML2D-D, MML-3, mu fit-1) did not detect any gene expression in normal tissues including spleen, thymus, bone marrow, heart, liver, kidney or in tumour-derived cell lines, with or without integrations in these loci (data not shown). Due to an inability to detect any transcribed sequences close to the integration sites, we decided to sequence the areas surrounding the proviral integrations and search for genes using bioinformatic sequence analysis tools. Because of the possibility that retrovirus-induced transcriptional deregulation can occur from a long distance, large regions were sequenced (Jonkers & Berns, 1996).

To identify potential genes in the vicinity around c-myb, Mml1, Mml2, Mml3 and Fti1, 39-6 kb in the Mml1 region was sequenced including between Mml1 and the c-myb, and sequences around the Mml2 and Mml3 (16-4 kb and 5-9 kb, respectively). To sequence the region between Mml1 and c-myb we used lambda clones subcloned from a P1 clone (#5362, Genome Systems). For sequencing around Mml2 and Mml3, plasmid clones derived from a G9 BAC (RPCI-22, #401-G9) were used. After sequencing and prior to comparative sequence analysis, repetitive elements were filtered using the RepeatMasker program (A. F. A. Smit and P. Green, unpublished data, University of Washington Genome Center, http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). The extent
of masking varied. It was 27.4, 28 and 38.9% for Mml1, Mml2 and Mml3, respectively (Table 2).

Using the BLAST algorithm (Altschul et al., 1997) to search the expressed sequence tag (EST) database, four ESTs with coding capacity sequence were identified near the integration sites. Two ESTs were found in Mml1 and two in Mml2 sequences as shown in Table 3 and Fig. 4(A, B). In a second approach, gene/exon prediction programs were used to identify potential exons. Sequences were analysed by the RUMMAGE sequence annotation system (Taudien et al., 2000) (http://gen100.imb-jena.de/rummage), which includes several exon prediction programs (GenScan, Grail, mzzF and Xpound). The GenScan program predicts a complex gene structure comprising several exons (Burge & Karlin, 1997). The other programs search for individual exons of different reliability based on their probability score (Thomas & Skolnick, 1994; Xu et al., 1994; Zhang, 1997). No exon was predicted by all programs, so we analysed only those identified by at least two programs. Two were located in the Mml1 region and two in Mml2 (Fig. 4A, B). Next, RNAs from embryos, several normal tissues and tumours were analysed for transcription by Northern blot analysis using PCR-derived probes covering the EST regions or by RT–PCR with amplimers derived from predicted exon sequences. However, as in the case of probes that were proximal to integration sites, we again did not detect any transcripts. In support of this finding, we did not find any CpG islands (clusters over 400 bp with at least 55% GC content) in the region, which would have suggested the presence of a gene(s). Although no transcripts were detected, this result does not exclude the possibility that these regions are associated with genes that may be expressed at early stages of tumour formation.

Since retroviruses can activate genes from a long distance, the search was extended upstream of Mml3. Shotgun cloning (Deininger, 1983) and low redundancy sequencing (Bouck et al., 1998) of BAC clone J8 (RPCI-22, #471-J8) resulted in approximately 130 kb of sequence (Table 2). Within this sequence, regions of homology to several ESTs were found. The identified EST sequences were used to probe for expression by Northern blot analysis on mouse tissues and tumour cell lines and our characterization is summarized in Table 4. In addition, their presence is noted on the map in Fig. 2. Although probes to two of the ESTs (A2, A239) detected message in some non-haematopoietic tissues of the mouse, they did not detect message in the spleen or thymus or tumours with integrations in Mml1, Mml2 or Mml3. Probes to the other EST detected transcripts in all of the above types of tumours as well as in normal tissues; however, there was no alteration in size or abundance of RNA when comparing normal versus neoplastic cells. Therefore, we concluded that none of these ESTs appears to be involved in tumour development. One EST homologue, A115, also shows homology to the HBS1L/eRFS gene (Wallrapp et al., 1998). The protein encoded by this gene possesses GTP-binding activity and is related in primary structure to proteins involved in translational machinery. Human HBS1L spans 90 kb and is located approximately 130 kb upstream of c-myc on chromosome 6. Our genomic clone (A115) covers approximately 160 bp of mRNA sequence, which, from our comparison of the human mRNA (acc.# XM 051825) and genomic sequences (acc.# NT 027047, contig containing c-myc and HBS1L), appears to be homologous to the second exon of the human gene.

We compared our sequences using BLAST to a recently published draft of the human genome sequence (http://www.ncbi.nlm.nih.gov/genome/seq/) and identified regions homologous to Mml1 on human chromosome 6 upstream of human c-myc. This was not a surprise because this part of mouse chromosome 10 is considered to be syntenic with an analogous region on human chromosome 6. We were interested in the extent of homology based on collinearity of individual sequences. According to our data, the order of loci on mouse chromosome 10 is c-myc–Mml1–Mml2–Mml3–Fti1–A239–A182–A115/A2. We used the c-myc gene, which was localized on human chromosome 6 (acc.#, NT 025741) to position 2801411–2839158, as a reference. The order of the loci on the
**Table 2. Sequencing statistics**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Total (kb)*</th>
<th>Contiguous sequence total (kb)/number†</th>
<th>Non-contiguous sequence total (kb)/number‡</th>
<th>Masking level§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mml1</td>
<td>~ 39.6</td>
<td>NA</td>
<td>NA</td>
<td>27.4%</td>
</tr>
<tr>
<td>Mml2</td>
<td>~ 16.4</td>
<td>NA</td>
<td>NA</td>
<td>28.0%</td>
</tr>
<tr>
<td>Mml3</td>
<td>~ 5.9</td>
<td>NA</td>
<td>NA</td>
<td>38.9%</td>
</tr>
<tr>
<td>J8 – shotgun</td>
<td>~ 134</td>
<td>~ 88.1/84</td>
<td>~ 45.9/79</td>
<td>ND</td>
</tr>
</tbody>
</table>

* Size of sequenced regions of Mml loci or total unique sequence obtained by shotgun cloning and sequencing of 440 plasmid subclones of the J8 BAC.
† Overall size of assembled J8 sequence and the number of contigs obtained.
‡ Overall size of unique sequenced DNA in non-contiguous form and the number of individual clones not assembled into contigs.
§ Level of masking of repetitive elements as determined by Repeat Masker program.
NA, Not applicable; ND, not determined.

**Table 3. EST homologues in Mml**

<table>
<thead>
<tr>
<th>Locus</th>
<th>EST name</th>
<th>Distance from VIS/strand*</th>
<th>Accession †</th>
<th>Source</th>
<th>Length of homology</th>
<th>Level of homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mml1</td>
<td>M1est1</td>
<td>-1 kb/c</td>
<td>BB231324</td>
<td>M.m. neonate thymus</td>
<td>230 bp</td>
<td>95%</td>
</tr>
<tr>
<td>Mml1</td>
<td>M1est2</td>
<td>+12 kb/d</td>
<td>BB209075</td>
<td>M.m. neonate thymus</td>
<td>300 bp</td>
<td>96%</td>
</tr>
<tr>
<td>Mml2</td>
<td>M2est1</td>
<td>+1.5 kb/d</td>
<td>AU023577</td>
<td>M.m. unfertilized egg</td>
<td>570 bp</td>
<td>99%</td>
</tr>
<tr>
<td>Mml2</td>
<td>M2est2</td>
<td>-10 kb/c</td>
<td>Al684548</td>
<td>Hs. pooled library</td>
<td>470 bp</td>
<td>99%</td>
</tr>
</tbody>
</table>

* VIS, virus integration site.
† Homologous sequences were found by comparing the repeat masked Mml sequences with the EST database using the BLAST program.
‡ - , Upstream of the integration site; +, downstream of the integration site; d, direct strand; c, complementary strand.

human chromosome was determined to be c-myb–Mml1–Mml2–Mml3–A115–A274 (A274 was not mapped in the mouse BAC using Southern analysis), thus expanding the syntenic region to almost 250 kb.

As indicated above, our search for exons in close proximity to Mml1, Mml2 and Mml3 did not lead to direct identification of any genes influenced by proviral integration at these loci. The lack of detectable coding regions in the vicinity might suggest that these regions harbour regulatory or structural elements. Chromatin-organizing structures, such as scaffold/matrix attachment regions (S/MARs) are found in AT-rich areas. Although the overall AT content for the 39.6 kb Mml1 region, as shown in Fig. 4A, is 56.5%, we found several areas that have a relatively high AT content, between 61 and 65%.

The MAR-Wiz program (Singh et al., 1997) (www.future-soft.org/MAR-Wiz/) predicted the S/MAR around the position of 18.3 kb which is next to the provirus insertion site in one Mml1 tumour – V46–34 at a position around 17.9 kb. Since S/MARs often contain specific recognition sequences defined by a MAR/SAR recognition signature (MRS), we searched for these elements. The signature is made up of two degenerate sequences, an 8 bp sequence (AATAAYAA) and a 16 bp sequence (AWWRTAANNWWGNNNC), found within 200 bp of each other (van Drunen et al., 1999). Although we could identify several 8 bp sequences throughout the Mml1 sequence, we could not identify the 16 bp sequences. The Mml2 sequence of 16.4 kb (Fig. 4B) with AT content of 59.2% also contains several AT-rich areas of 66–72% AT. In one of them the MAR-Wiz predicted an S/MAR around 11.3 kb, but we did not find any complete MRS. We conclude from this search for S/MARs that there is some evidence for S/MARs. Although the data are only partially predictive, they warrant...
Mml integration loci on mouse chromosome 10

Fig. 4. Features of the Mml\(^1\) (A) and Mml\(^2\) (B) locus sequences. Several features including predicted exons, EST homologues and predicted S/MARs are depicted in relationship to virus integration sites. Virus integration sites are shown by vertical lines and the orientation of provirus is indicated by an arrow.

Table 4. EST homologues in J8 BAC

<table>
<thead>
<tr>
<th>Clone name</th>
<th>EST accession</th>
<th>Distance from Mml3</th>
<th>Origin</th>
<th>Expression (Northern)</th>
<th>Transcript size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Positive tissue</td>
<td>Negative tissue</td>
</tr>
<tr>
<td>A2</td>
<td>AI464150</td>
<td>&gt; 50 kb</td>
<td>M.m. Lv, Kd</td>
<td>L, Kd</td>
<td>Sp, Th</td>
</tr>
<tr>
<td>A115</td>
<td>AA108088</td>
<td>&gt; 50 kb</td>
<td>M.m. embryo, Ts</td>
<td>L, Kd, Sp, Th</td>
<td></td>
</tr>
<tr>
<td>A182</td>
<td>AI508562</td>
<td>&gt; 50 kb</td>
<td>M.m. LN, Th</td>
<td>Kd, Sp, Th</td>
<td>L, Kd</td>
</tr>
<tr>
<td>A239</td>
<td>AA915174</td>
<td>&lt; 50 kb</td>
<td>M.m. Ht, Ts</td>
<td>L, Kd, Sp, Th</td>
<td></td>
</tr>
<tr>
<td>A274</td>
<td>AI327317</td>
<td>ND</td>
<td>M.m. Th, Ts, MG</td>
<td>L, Kd, Sp, Th</td>
<td>+</td>
</tr>
</tbody>
</table>

* For each EST, expression was examined in representative tumours with integrations into Mml\(^1\), Mml\(^2\), Mml\(^3\) and c-myb.
† Lv, liver; Kd, kidney; Sp, spleen; Th, thymus; Ts, testis; Ht, heart; LN, lymph node; MG, mammary gland.
‡ ND, not determined.

Further experimentation to determine directly whether or not S/MARs are present.

Discussion

Our analysis of a set of monocytic leukaemias established a target region for retroviral integration that spans approximately 60 kb and is located between the Fli1 and c-myb loci. Although we found several genes transcribed from a region more than 80 kb upstream of c-myb, there was no evidence within this 80 kb area for the presence of transcribed sequences. The lack of genes in a region of this size is not unusual. Data obtained from sequencing of the human genome show that exons comprise only about 1–1.5% of the genome sequence and 20% of the genome sequences are contained in islands, larger than 500 kb, without a gene. However, gene density
varies between 43.5 kb per gene for the gene-rich chromosome 19 and 200 kb per gene for two gene-poor chromosomes, 13 and Y (Venter et al., 2001). With this in mind it would not be unusual to find an 80 kb plus segment of DNA that is devoid of genes.

Repeated integration of proviruses into this chromosome 10 region, in association with tumorigenesis, suggests an important function for these sequences, despite the absence of transcribed genes in this area, that might be altered in expression due to the provirus. The proviruses at these loci are like many others that appear to be involved in neoplastic disease but have not yet been assigned a function (Jonkers & Berns, 1996). The fact that the region is devoid of genes, and the MAR-Wiz program predicted two potential S/MARs, might suggest that at least part of the area contains structure-related regulatory sequences. Matrix attachment elements organize chromatin topologically, allowing spatial and temporal gene regulation and the nuclear matrix is implicated in organizing chromatin topologically, allowing spatial and temporal regulatory sequences. Matrix attachment elements (Berns, 1996). The fact that the region is devoid of genes, and disease but have not yet been assigned a function (Jonkers & Berns, 1996). The fact that the region is devoid of genes, and the MAR-Wiz program predicted two potential S/MARs, might suggest that at least part of the area contains structure-related regulatory sequences. Matrix attachment elements organize chromatin topologically, allowing spatial and temporal gene regulation and the nuclear matrix is implicated in the control of development and differentiation (Nepveu, 2001; Stein et al., 1995). Interestingly, alterations in nuclear architecture have been associated with the neoplastic condition (Deppert et al., 2000); there are transforming genes that have the ability to bind to nuclear matrix including SV40 Large T antigen and mutant p53, the latter of which was found to bind directly to the S/MAR (Deppert, 2000). In addition, S/MARs are AT rich and some proteins, including the mixed lineage leukaemia (MLL) protein, contain an AT hook domain, which binds to these sequences (Broeker et al., 1996; Caslini et al., 2000). Moreover, MLL fusion proteins have been implicated in myeloid leukaemia in man (Dimartino & Cleary, 1999). It is conceivable, therefore, that the virus could perturb the binding of DNA to matrix through the physical disruption or could introduce regulatory proteins such as those that bind to the LTR. Further studies will be required to determine the function of the sequences predicted to be S/MARs by the MAR-Wiz program.

One must still consider the possibility that regulatory regions associated with these integration sites could function to influence c-myb expression. Integrations have now been found in a wide region (approximately 150 kb) both upstream (Fti1, Mml3, Mml2, Mml1) of c-myb, as well as downstream (Ahi1). Until recently, downstream integrations were only found in lymphoid tumours. However, Blaydes et al. (2001) also discovered provirus insertions in this region in a series of BXH-2 myeloid leukaemias.

In the present study we considered the possibility that integrations affected Myb transcription by looking to see if, in established tumour cell lines, provirus integration in Mml1 caused increased expression of c-myb (Fig. 3A, B) and found that some cell lines did not express it. However, our inability to find increased expression in the leukaemias may not rule out the possibility that integration into this area influenced c-myb expression during an early phase of the disease prior to differentiation. Our previous data suggest that mature MMLs (with integrations in c-myb) develop from pre-leukaemic cells in the bone marrow or spleen that are immature (Nason-Burchenal & Wolff, 1993). If pre-leukaemic cells involving Mml1, 2, 3 are also immature cells, their mechanism of endogenous c-myb gene regulation could be very different from that of end-stage tumour cells, which are mature monocyctic cells. Integrations that affect c-myb could influence early stage progression, for example by providing increased proliferation potential or an anti-apoptotic role, since these are assigned functions of c-Myb. If this is the case, one would have to predict that, at later stages of progression, these functions are provided by additional oncogenic event(s). The proximity of this proto-oncogene to the Mml integration sites continues to raise the question of whether these integration sites affect c-myb.

Overall, the data presented here indicate that the 80 kb region upstream of c-myb, which contains three separate retroviral integration sites associated with myeloid tumours, may be devoid of genes, but could possess regulatory function. Since other integration sites associated with lymphoid and myeloid tumours have also been found in this chromosomal region, it will be important to determine how sequences in this area can influence the neoplastic state.

We thank Paul Jolicoeur for kindly providing the Ahi-1 probe. We are also grateful to Lawrence Kirschner for providing the PFGE apparatus for the physical mapping. We would like to thank Mark Miller for analysis of the sequencing reactions.

References


Received 15 October 2001; Accepted 7 December 2001

---

**Mml integration loci on mouse chromosome 10**