Structure and Possible Mechanisms of TEL-AML1 Gene Fusions in Childhood Acute Lymphoblastic Leukemia

Joseph L. Wiemels and Mel Greaves

Leukaemia Research Fund Centre, Institute of Cancer Research, Chester Beatty Laboratories, London SW3 6JB, United Kingdom

ABSTRACT

TEL-AML1 gene fusion derived by chromosomal translocation is a common acquired genetic lesion in pediatric cancer that is present in ~25% of B-cell precursor acute lymphoblastic leukemias, and recent evidence suggests that this recombination event may initiate leukemogenesis prenatally during fetal hemopoiesis. Analysis of the DNA sequence and structure surrounding the breakpoints may reveal clues to their formation. A long-distance inverse PCR strategy was used to amplify TEL-AML1 genomic fusion sequences from diagnostic DNA from nine patients. Breakpoints were scattered within the 14 kb of intronic DNA between exons 5 and 6 of TEL and in two putative cluster regions within AML1 intron 1. Fusion sequences exhibited characteristic signs of nonhomologous end joining, including microhomologies at the end points, and small deletions and duplications. DNA sequences near the breakpoints did not reveal any consistent characteristic signal sequences of the V(D)J recombinase, topoisomerase II consensus sites, or other sequence motifs associated with recombination. However, several translocations occurred near a repeat region of TEL that was found to be highly polymorphic. This region was cloned and found in nuclease sensitivity assays to exhibit paranemic structures, which may have contributed to DNA breakage or illegitimate recombination. The data are compatible with the possibility that TEL-AML1 translocations occur by nonhomologous recombination involving imprecise, constitutive repair processes following DNA double-strand breaks.

INTRODUCTION

Reciprocal chromosomal translocations have provided both diagnostic markers and insight into the pathophysiology of leukemias by virtue of the biological effects of their fusion gene products (1, 2). However, the initial causes of the translocations and the molecular mechanisms involved in illegitimate recombination are largely unresolved. Primary genomic fusion sequences have in some cases revealed clues about the potential mechanisms. These include abundant V(D)J recombinase activity, topoisomerase II-mediated breakage followed by end joining, recombination mediated by Alu repeats or alternating purine/pyrimidine tracts of DNA, and sequences binding to specific DNA-binding proteins (3–12).

A hallmark of lymphoid development is the expression of the V(D)J activating genes RAG-1 and RAG-2, which facilitate immunoglobulin (IGH) or T cell receptor (TCR) gene rearrangement during the precursor cell stage (13). The V(D)J recombinase machinery has been implicated in the formation of translocations in lymphoid neoplasms between the IGH or TCR gene and partner genes that are not normally substrates for V(D)J recombinase activity (14, 15). A consistent feature of such translocations is the presence of heptamer/nonamer signal sequences on one or both sides, as well as the presence of nontemplate N nucleotides at the fusion junction. Such characteristics are also evident in the 40% of deletions in a marker gene, HPRT, in the T lymphocytes of newborn infants and young children, whereas only a few percent of adult deletions show such characteristics (16, 17). IGH-associated recombinants occur in mature B-cell lymphomas of germinal center origin, and it is curious and counterintuitive that there should appear to be a dearth of fusions involving the IGH loci in the B-cell precursor leukemias of childhood, a leukemia that corresponds to the developmental stage when the recombinate machinery is most active (13, 18). It has been suggested that recombinases might facilitate illegitimate recombination involving the non-IG/TCR genes when both partners, or even one partner only, possess appropriate sequences for recombination binding. This mechanism was proposed for the common MLL-AF4 fusion gene in infant ALL (1), although evidence that appropriate signal sequences straddling breakpoints are consistently present or functionally active is lacking. Alternative mechanisms of recombination involve homologous recombination via Alu repeats (7, 9), alternating purine pyrimidine tracts (10), or involvement of translin, a recombination hot spot binding protein (19). Rearrangements of the MLL gene in infant leukemia and therapy-associated secondary leukemia involve breaks close to DNease I-hypersensitive regions, a scaffold attachment region and a topoisomerase II site (4, 20, 21).

The common variant of childhood leukemia (cALL) is a B-cell lineage progenitor/stem cell cancer (18), and some 25% of cases have a karyotypically cryptic t(12;21) (22, 23) translocation producing a TEL(ETV-6)-AML1(CBFA2) fusion gene. Both TEL and AML1 are critical genes for hemopoiesis as revealed by homologous recombination, loss of function experiments in mice (24, 25). Both genes are found as fusion partners with a variety of other genes encoding kinases or transcription factors in both lymphoid and myeloid leukemias (26). The TEL-AML1 fusion appears to be unique to B-cell progenitor ALL. The basis of this selectivity (as with other fusion genes) is an important unresolved issue (27) but most likely reflects a selective impact of the chimeric protein on the proliferation and/or survival of B-cell precursors (26). Breakpoints in TEL and AML1 are constrained to particular introns (28), which probably reflects minimal functional requirements for transformation by the chimeric protein (29).

The timing and mechanism of the TEL-AML1 fusion event is clearly important for our understanding of the etiology and pathogenesis of childhood ALL. Studies on identical twins with ALL (30, 31) and retrospective analysis by PCR of neonatal blood spots or Guthrie cards (32) provide evidence that this genetic lesion may initiate leukemogenesis during fetal hemopoiesis in utero. The TEL-AML1 translocation has only been identified relatively recently, and few details of the genomic rearrangements have been reported. We have adapted a LDI-PCR method (33) for rapid amplification of the genomic fusion region. Sequence analysis and comparisons of fusion regions from a series of patients provide the opportunity to assess what molecular mechanism might play a role in producing this genetic recombinant. We report here that these sequences do not demonstrate the hallmarks of V(D)J or homologous recombination; instead, the
sequences display possible remnants of NHEJ activity typical of mammalian cells.

MATERIALS AND METHODS

Patients and Cell Lines. Diagnostic DNA samples were derived from five patients enrolled in the United Kingdom Childhood Cancer Study from the period 1992–1997. The REH cell line known to contain the TEL-AML1 translocation (34) was used as a positive control, and the MonoMac 6 cell line (which harbors a MLL gene rearrangement) was used as a negative control. Three ALL translocation sequences were derived from three pairs of identical twins with concordant ALL (30–32). All patients had common B-cell precursor ALL by immunophenotype. The age range was 2–14 years.

DNA Isolation and Southern Blots. DNA was isolated from mononuclear cells from peripheral blood or from bone marrow smear slides previously stored at −70°C, using conventional SDS/proteinase K/phenol methods. For Southern blots 5 µg of DNA were digested to completion with BamHI or EcoRI, separated on 0.7% agarose gels, capillary blotted to nylon membranes, and probed with various probes to TEL as shown in Fig. 1. Probes were made using PCR using the following primers: (a) probe 1, A (TGACCCATCTCGTAATC) and B (TGTCCTGGAAGACGATTTGC); (b) probe 2, A (ACTGGGCTCAGCCTGTAATC) and B (TGTCACAGCACATACATG); and (c) probe 3, A (TGACTCCGGGCTGGATCTCAG) and B (CGGAGATGAGACTACCTCGT).

LDI-PCR Analysis. For successful inverse PCR, a translocation must result in a rearranged restriction fragment that is shorter than the germ-line fragment. In practice, when rearrangements produced larger fragments, the unarranged TEL allele (or germ-line TEL from nonleukemic cells) will be preferentially amplified. In some cases, a rearranged band fragment shorter than the germ-line fragment was identified using Southern blots probing TEL intron 5. In other cases in which DNA was limited, LDI-PCR was performed: that is, without any Southern blot information. Both TEL-AML1 and its reciprocal, AML1-TEL, were routinely targeted in these blind cases. Multiple enzymes (BamHI, EcoRI, and HindIII) and primer sets were used to maximize the chances of amplifying rearranged TEL.

LDI-PCR was performed essentially as described previously (Ref. 33; Fig. 1A). Briefly, 0.5 µg of DNA was digested overnight with a restriction enzyme. The reaction was extracted with phenol and then with ethyl ether. The DNA was diluted up to 500 µl with 1× ligase buffer and treated with 5 units of T4 DNA ligase (Promega) overnight at 14°C. The dilute nature of the DNA favored intramolecular ligation (circumstantial) of the restriction fragments over intermolecular ligation. This mixture was purified and concentrated using Qiagen Qiaquick columns and eluted in a final volume of 40 µl of TE (10 mM Tris-HCl (pH 8.5) and 1 mM EDTA). One µl of this DNA solution was then used for subsequent PCR using inverse primers.

Two rounds of PCR were performed using nested sets of primers to ensure specificity and to yield enough product for sequencing and cloning. The primers (Oswel) used are shown diagrammatically in Fig. 1B and Table 1. For TEL-AML1 amplification and BamHI digests, primer A0 was used in combination with primer 6B, 1B, 4B, or 5B. For the reciprocal AML1-TEL, B0 was used with 7A, 5A, 4A, or 1A. For EcoRI digests, 1A and 1B are used together, and so on. In a single case, a HindIII digest was used with primers 3A and 3B. Nested primer sets are indicated by the suffix “-1” and contain NorI sites. PCR products from the nested reaction were sequenced directly with the same primers used in the PCR reaction and then resequenced by primer walking until the breakpoint was reached. In some cases, restriction analysis of the LDI-PCR product allowed for faster delineation of the breakpoint. If the sequencing of PCR products proved unsuccessful, the products were cloned using the NorI site in pBluescript, and the plasmid insert was sequenced. All sequencing was performed on an ABI Prism 377 automated sequencer.

Upon obtaining the breakpoint sequence, novel PCR primers spanning the breakpoint were synthesized and used to amplify the breakpoint sequence in the original diagnostic DNA, using other TEL-AML1 diagnostic DNAs and normal human DNAs as controls. Positive bands were then resequenced.

In most cases, an attempt was made to amplify the reciprocal translocation. Where the translocation occurred in AML1 in a location already sequenced by others (as reported to the GenBank database), primers targeting the flanking DNA of AML1 and TEL were directly synthesized, and patient samples were analyzed. For translocations in unsequenced regions of AML1, further LDI-PCR was performed to obtain the cryptic AML1 sequence and to attempt to amplify the reciprocal translocation. In this case, novel primers were synthesized based on the novel AML1 sequence close to the breakpoint. These primers were used to amplify AML1 from normal DNA prepared for inverse PCR. As the work presented in this study was completed, additional AML1

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv 0A</td>
<td>GGAGAGAAGAGACAGCTTCTAGCCAGT</td>
</tr>
<tr>
<td>Inv 0A-1</td>
<td>GGAGAGAAGAGACAGCTTCTAGCCAGT^a</td>
</tr>
<tr>
<td>Inv 0B</td>
<td>GGTGCTTCTCTGACAGCCGGTACAC</td>
</tr>
<tr>
<td>Inv 0B-1</td>
<td>GGTGCTTCTCTGACAGCCGGTACAC</td>
</tr>
<tr>
<td>Inv 1A</td>
<td>ATGGGCTCAGCCTGTAATTGC</td>
</tr>
<tr>
<td>Inv 1A-1</td>
<td>ATGGGCTCAGCCTGTAATTGC</td>
</tr>
<tr>
<td>Inv 1B</td>
<td>CATGTTGCTTGCAGCAGTACCTTGAAG</td>
</tr>
<tr>
<td>Inv 1B-1</td>
<td>CATGTTGCTTGCAGCAGTACCTTGAAG</td>
</tr>
<tr>
<td>Inv 2A</td>
<td>AGGAGACCAGGGTGATATT</td>
</tr>
<tr>
<td>Inv 3A</td>
<td>ACCCTCGGCTCAATGCTCATCAG</td>
</tr>
<tr>
<td>Inv 3A-1</td>
<td>ACCCTCGGCTCAATGCTCATCAG</td>
</tr>
<tr>
<td>Inv 3B</td>
<td>AGGACGACTCCTTGCTGATATT</td>
</tr>
<tr>
<td>Inv 3B-1</td>
<td>AGGACGACTCCTTGCTGATATT</td>
</tr>
<tr>
<td>Inv 4A</td>
<td>TGACTCGGCTGAGGCTTAGC</td>
</tr>
<tr>
<td>Inv 4A-1</td>
<td>TGACTCGGCTGAGGCTTAGC</td>
</tr>
<tr>
<td>Inv 4B</td>
<td>CCTCATTTGCGACTGCTGGATTC</td>
</tr>
<tr>
<td>Inv 4B-1</td>
<td>CCTCATTTGCGACTGCTGGATTC</td>
</tr>
<tr>
<td>Inv 5A</td>
<td>GTGACTCGGCTGAGGCTTAGC</td>
</tr>
<tr>
<td>Inv 5A-1</td>
<td>GTGACTCGGCTGAGGCTTAGC</td>
</tr>
<tr>
<td>Inv 5B</td>
<td>GAAACACTTGGCTGAGCAGTATG</td>
</tr>
<tr>
<td>Inv 5B-1</td>
<td>GAAACACTTGGCTGAGCAGTATG</td>
</tr>
<tr>
<td>Inv 6B</td>
<td>GAAACACTTGGCTGAGCAGTATG</td>
</tr>
<tr>
<td>Inv 6B-1</td>
<td>GAAACACTTGGCTGAGCAGTATG</td>
</tr>
</tbody>
</table>

^a primers contain the NorI site tail at the 5′ end (5′-TTTTATATTGCGGACC-3′).
intronic sequences were reported to the National Center for Biotechnology Information, allowing all cryptic AML1 sequences to be identified.

Analysis of Paranemic DNA Structure and Topoisomerase II Sites. Regions of TEL intron 6 were amplified by PCR and cloned into pBluescript. Plasmid maxipreps were performed (Qiagen maxiprep kit), and supercoiled plasmid DNA was analyzed for paranemic (single-stranded) character by nuclease sensitivity assays. A DNA structure is defined as paranemic if the participating strands are stably separated without mutual rotation of the opposite strands (35). Separated strands can be identified by single-strand DNA-specific endonucleases. Plasmid DNA (200 ng) was incubated with 1 unit of S1 nuclease (Promega) in a 12-μl reaction with buffer supplied by the manufacturer (pH 5.0) for 30 min at 23°C, 37°C, or 50°C. Reactions were stopped by a 5-min incubation at 70°C. The reaction buffer was adjusted, and the plasmids were cut with BamHI and electrophoresed in agarose. Alternatively, the S1-treated plasmids were treated with calf intestinal phosphatase, phenol-extracted and ethanol-precipitated, and end-labeled with 32P by T4 kinase. After heat inactivation, the plasmids were digested with BamHI and separated on a 6% polyacrylamide denaturing sequencing gels. Maxam and Gilbert sequencing ladders of the same plasmids were used as markers. As a second check, mung bean nuclease was also used to assay for paranemic structures at a pH value closer to physiological levels, pH 7.0.

Assays for topoisomerase II sites using purified topoisomerase II (Topogen) were performed as described previously (5). End-labeled fragments, which were also used for producing Maxam and Gilbert ladders, were subjected to increasing amounts of etoposide in the presence of topoisomerase II. Samples were subjected to electrophoresis in 6% polyacrylamide gels.

RESULTS

The LDI-PCR method was first attempted on the REH cell line (sample 4). REH cell DNA was digested with BamHI and digested separately with EcoRI, Southern blotted, and probed with TEL probes 1, 2, and 3. A 2.3-kb band fragment was seen in a EcoRI blot using probe 2, instead of the normal 6.4-kb band; therefore, this smaller band fragment was targeted with inverse PCR. Inverse 1A and 1B primers, followed by a second round with 1A-1 and 1B-1 primers, amplified the 2.3-kb fragment. The fragment was sequenced from the AML1 side by primer walking until TEL DNA was reached. A stretch of 10 cytosines and 10 adenines near the breakpoint (Fig. 3, sequence 4) obscured the sequencing beyond this point (due to Taq polymerase errors); therefore, the fragment was cloned, and a single clone was sequenced. A final PCR with primers on the TEL and AML1 sides was performed on REH DNA; a PCR product was only obtained using REH cell DNA (data not shown) and was not obtained from DNA from patients with TEL-AML1 translocations or normal DNA, thus indicating a unique clonal rearrangement.

Patients were identified as TEL-AML1 + by reverse transcription-PCR as part of a large-scale epidemiological study of childhood cancers (the United Kingdom National Case/Control Study of Childhood Cancer; translocations 1, 3, 7, 8, and 9 in Figs. 2–5) or as part of an international survey of leukemia in identical twins (translocations 2, 5, and 6). In many cases, DNA from these patients was limiting and precluded Southern blot analysis. A different strategy was then used, using BamHI digests first and various combinations of primers to amplify TEL-AML1 or AML1-TEL. Of a total of 16 samples treated as such, 5 translocations were amplified (translocations 1, 2, 7, 8, and 9). Of the remaining 11 samples, EcoRI digests followed by LDI-PCR allowed an additional sequence (translocation 3) to be identified. An additional sequence was obtained using a HindIII digest and inverse 3A and 3B primers (translocation 5) for a total of eight new sequences reported here. A sequence reported earlier in a pair of identical twins with ALL (30) is also included here (translocation 6) because we have refined its breakpoint and characterized its reciprocal. The twin pairs share an identical or clonotypic TEL-AML1 break-point sequence (30); therefore, only one sequence is reported here for each pair.

Breakpoints in TEL were distributed within intron 6, apparently skewed toward the 5' end of the intron (Fig. 2). There was some clustering because breakpoints 3–6 were located within a 384-bp stretch, and breakpoints 8 and 9 were 116 bp apart. There was no apparent clustering near Alu repeat sequences. Breakpoints in AML1 were more widely distributed, reflecting the extremely large (158 kb) intron 1. At least 300 bp of DNA from the non-TEL side of each inverse PCR product were sequenced to definitively assign each patient's translocation to the AML1 gene, which has now been completely sequenced (Fig. 2). All breakpoints occurred within this intron (translocations in other patients have been reported in intron 1 or 2; Ref. 36), except for patient 1, whose breakpoint preceded exon 1. This translocation, however, was the reciprocal AML1-TEL. We were not able to amplify TEL-AML1 for patient 1, which may have its breakpoint within intron 1 of AML1. This patient exhibited a fusion reverse transcription-PCR product, like all of the rest, which included exon 5 of TEL fused to exon 2 of AML1 (data not shown). Interestingly, three breakpoints were clustered near AML1 exon 1 (translocations 1, 2, and 9), and six were located in the 3' half of intron 1 (Fig. 2). Patient 1's breakpoint preceded AML1 exon 1 by 3534 bp, and patients 2 and 9 had breakpoints 754 and 2465 bp after exon 1, respectively. Patient 5's breakpoint was also close to an exon, 370 bp 5' of exon 2.

The reciprocal was identified for five of the nine translocations sequenced (Fig. 3). The cryptic AML1 sequence was discovered in the database for three of these translocations (translocations 2, 5 and 9), whereas inverse PCR was used to identify the cryptic AML1 sequence for two others (translocations 6 and 8). Of the remainder, we identified the AML1 cryptic sequence using inverse PCR for two (translocations 1 and 4), but we were unable to amplify the reciprocal. For these two, large deletions and/or additions may have precluded amplification of the reciprocal. For the last two (translocations 3 and 7), cryptic AML1 was identified from National Center for Biotechnology Information data, but amplification of the reciprocal was not attempted. Of the 14 breakpoints in which both original sequences are known, 9 had regions of microhomology at the breakpoint of between 1 and 9 bp (64%). Of the five breakpoints where both forward and reciprocal are known (10 breakpoints total), there were five duplications of 3–90 bp and five deletions of 48–208 bp. Two had duplications of both TEL and AML1, two had deletions of both, and one had a duplication of AML1 and deletion of TEL (Fig. 4).

Analysis of sequences did not demonstrate evidence of the involvement of site-specific or homologous recombination as general features of TEL-AML1 rearrangements; i.e., V(DJ) recombinase signal se-
quences or N region nucleotides, Alu repeats, topoisomerase II sites, translin consensus sequences, or alternating polypurine/polypyrimidine sites were not evident as recurring themes in visual screens of translocation sequences. Sequence searches for V(D)J, translin, and topoisomerase II motifs using the MacVector program yielded single- and double-mismatched motifs in some proximity to the translocation breakpoints, but not close enough to suggest causality. The possibility of topoisomerase II cutting sites near breakpoints was analyzed using

Fig. 3. Fusion sequences between TEL and AML1 and parent alleles. TEL-AML1 fusions and their reciprocals are shown for translocations 2, 5, 6, 8, and 9. AML1-TEL only is shown for translocation 1, and TEL-AML1 only is shown for translocations 3, 4, and 7. Microhomologies are indicated in bold letters. Hash marks indicate homology between the derivative chromosome and the parent alleles. Underlined sequences are duplicated from the parent allele.
STRUCTURE OF TEL-AML1 TRANSLOCATIONS IN LEUKEMIA

Fig. 4. Comparison of the sizes of deletions and duplications of parent DNA sequence in five TEL-AML1 translocations and their reciprocals.

Fig. 5. Sequences of RSL and RSS. The sequence of RSL is shown in full. Dash marks indicate homology with RSS, and dots are sequences absent in RSS. Other bp differences are indicated. Translocation sites are indicated with arrows. Sequences of RSS and RSL were derived from PCR products, which were subsequently cloned into pBluescript, resequenced, and used for nuclease sensitivity assays. Sensitive sites are indicated with disc objects, which refer to Fig. 7. Nuclease-sensitive sites in the top strand are shown as gray discs on the top strand. Sensitive sites on the bottom strand are shown inverted on the bottom; striped discs are sensitive sites at pH 5.0, and the dotted disc is the major sensitive site at pH 7.0. The location of the sensitive site in the bottom strand of the unique site in RSL (U in Figs. 6 and 7 and in the third line of the sequence data) was not mapped accurately and is omitted from this figure.

purified topoisomerase II and etoposide as an inhibitor on linearized plasmids containing the breakpoint sites for translocations 3–6. Sequencing gels did not demonstrate strong topoisomerase II sites at the locations of breakpoints (data not shown).

Individual translocations had some marked features. Translocations 4 and 6 both had a string of homologous adenes near the breakpoint that might have aided in lining up the DNA for ligation. The AML1 side of translocation 7 is a polypurine/polypyrimidine sequence made up entirely of adenine and guanine. Such sequences have been shown to be involved in paranemic structures (35) and also closely resemble the recombination hot spot sequence found near BCL-2 translocations (GG A/T GG A/T GG; Refs. 37 and 38). Most interestingly, four sequences (sequences 3–6) were found in a region of TEL with an Alu repeat followed by a highly repetitive and polymorphic sequence. Screening of 15 alleles from different individuals over this region by PCR (using primers inv 3A and CATAACGAGACCCCGTGTTG-GAT) and sequencing revealed two major polymorphic variants (13 alleles were RSS, and 2 alleles were RSL) differing in size by 121 bp (Fig. 5). Several bp substitutions and frameshifts differing from these variants were also apparent in some individuals. One of the translocations (translocation 5) with end points within the polymorphic region was derived from a RSS parent TEL allele. Most of the repeat region was deleted in the other translocation with end points in this region (translocation 6), so we cannot determine the allele origin. The region contained several in-line and inverted repeats, suggesting that the DNA structure of this region may have contributed to the formation of the translocations. We therefore assayed for paranemic regions.

S1 nuclease digests of intact plasmids showed that the RSS had one dominant S1-hypersensitive site, and the RSL had two (Fig. 6). DNA from a region of TEL not involved in any translocations (the cloned fragment of inv 4B and inv 5A primers) did not display paranemic structure unless at a high temperature (50°C), and then at random sites (Fig. 6). We were able to map the exact location of the hypersensitive sites on polyacrylamide gels (Fig. 7) and to determine whether the sites were sensitive at more relevant pH conditions. Plasmid DNA was first treated with nuclease enzymes and then labeled with ^32P at the 5' end at the site of breakage. The plasmids were then cut with a single-cutting restriction enzyme, resulting in fragments that could be sized on a sequencing gel. BamHI digests revealed the pattern of breakage on the top strand (Fig. 7A), and BglII digests revealed the pattern of breakage on the bottom strand (Fig. 7B). As in the agarose gel analysis (Fig. 6), two major sites of breakage were apparent in RSL, and one was apparent in RSS. The sensitive sites were 5–8 bp long, as would be seen with a looping out of DNA in conformations such as H-DNA or the nonpaired DNA at the end of a hairpin or cruciform structure. One site was common to both plasmids (site L). A faint cleavage site was also seen downstream of site L (Figs. 5 and 7B). Cleavage at this site was more complex on the reverse strand, with three apparent cleavage regions (Fig. 7B). These regions were centered at pyrimidine/purine short repeats (Fig. 5). Mung bean nuclease was also used so that sensitivity could be analyzed at a pH value closer to physiological pH values. Interestingly, the region remained sensitive on the shift from pH 5.0 to pH 7.0, but the paranemic structure of the site changed (Fig. 7B). A purine repeat (AGAGAGAG) sandwiched between two formerly sensitive regions became the dominant sensitive site (Figs. 5 and 7B). It is possible that protonation of cytosine residues was important for the lower pH structure; this may be indicative of a H-DNA structure (35). At the higher pH value, this was replaced by a simpler structure. Several breakpoints occurred in close proximity to these nuclease sensitive regions (translocations 4–6), indicating that the unique DNA structure may have been responsible for their formation.
DISCUSSION

TEL-AML1 is the most common gene fusion in pediatric leukemia and has a prenatal origin in at least some cases. The nature of the recombinational mechanism yielding these translocations is thus of great interest and may yield clues to the etiology of the disease. LDI-PCR proved to be an effective and relatively simple method for cloning TEL-AML1 translocations. Inverse PCR was chosen for two reasons: (a) sequence information was available on one side of the translocation (TEL intron 5) but absent on the other (AML1); and (b) the breakpoint cluster region of TEL was of manageable size. The addition of long-distance PCR enzymes allowed complete coverage of the TEL intron in a limited number of PCR reactions. We amplified translocations from as little as 100 ng of starting genomic DNA, without previous Southern blot information. The simplicity of the LDI-PCR method may make it highly applicable to the cloning of other translocations in cancer, especially including MLL gene translocations, because no information is needed on the partner gene. However it should be noted that LDI-PCR did not work in all cases, even when the DNA amount and quality were not limiting factors. In some cases, there may have been a limited amount of blast cell DNA because DNA used for cloning was generally isolated from blood rather than bone marrow.

We have characterized the breakpoint sequences of TEL-AML1 in nine clinical samples of childhood ALL. In each leukemia, the precise genomic breakpoint or fusion sequence was unique, except for that shared by leukemias in identical twin pairs with leukemia. The shared breakpoints in twin pairs (translocations 2, 5, and 6) indicate a prenatal, single clonal origin of leukemia (30–32). Breakpoints were clustered within TEL intron 5 and AML1 intron 1; thus, all leukemias express the same chimeric TEL-AML1 protein. Breakpoints were scattered along the entire 18-kb TEL intron 5, although four breakpoints (44%) were clustered within a HindIII fragment (translocations 3–6) that contains intriguing structural features, namely an Alu repeat and a polymorphic repeat region. A previous report (36) demonstrated 8 of 10 breakpoints (80%) within this fragment (analyzed by Southern blot), indicating that the actual proportion of translocations within this region may be somewhat higher than is indicated in our series. The AML1 side of TEL-AML1 breakpoints has not been mapped previously. We found three of nine breaks to be located near AML1 exon 1, and the remaining six were located some distance from each other (with no sequence overlaps) in the 3’ half of intron 1. The lack of translocations in the 87 kb of sequenced intronic DNA following breakpoint 9 suggests that two breakpoint cluster regions may exist in intron 1 of AML1. These nonrandom breaks may then be biologically or mechanistically relevant.

Translocations can be viewed as an accidental byproduct of DNA repair following damage that might otherwise result in cell death. Damage would take the form of DSBs that are then misrepaired. The repair of DSB in mammalian cells occurs by two mechanisms, homologous recombination and NHEJ (39, 40). These mechanisms have been studied in mammalian cells by transfecting cut DNA into cells and analyzing the rejoined ends (41, 42), transfecting bacterial restriction enzymes into cells and analyzing rejoined breakpoints in a marker gene (43), and the introduction of the 18-bp I-Sce-I recognition site into mammalian DNA to facilitate the creation of a single DSB (44). Analysis of joining sequences in nearly all reports indicates that mammalian cells and cell extracts are capable of ligating nonhomologous DNA ends with high efficiency, with little loss or gain of sequence at breakpoint ends. This NHEJ pathway is used in preference to homologous recombination when cells are engineered with substrates that will permit both forms of repair (44). Cells mutant in RAD54 (homologous pathway) and KU70 (NHEJ pathway) demonstrate a greater reduction in sensitivity to γ-rays than either single mutant, indicating that the repair pathways are complementary (39). Cells mutant in RAD51, a critical protein involved in the NHEJ pathway, lead to accumulation of DNA breaks and lethality (45, 46). Conversely, cells mutant in RAD52, an important protein in homologous recombination, show reduced homologous recombination but normal sensitivity to ionizing radiation (47).

NHEJ is accomplished by the Ku protein DNA-PK complex (40). Ku protein binds free DNA ends, recruiting the DNA-PK. On interaction with an opposing DSB-DNA-PK complex, transphosphorylation occurs between DNA-PKcs, leading to the activation of Ku-helicase, allowing microhomology base-pairing. Finally, unpaired DNA is resolved, and DNA ends are ligated. This results in charac-
teristic resolved junctions showing microhomologies of 1–6 bp and small deletions. Deletions may be caused by degradation of DNA before Ku binding or by processing before ligation. Duplications may also occur and can result from staggered breaks or slippage-mispair followed by primed DNA synthesis (7, 41, 48). We find these same characteristics in translocations between TEL and AML1, suggesting a common mechanism of formation. The translocations are reminiscent of those induced by DSB in an in vitro model, bleomycin-induced translocations in confluence-arrested Chinese hamster ovary cells (49). In this model, six reciprocal translocations showed eight microhomologies among 12 breakpoints and deletions of 1–24 bp at break-point ends. However, no duplications were seen. These translocations closely resemble the Ku-mediated NHEJ pathway demonstrated with cut plasmids and enzyme-mediated breaks in previous studies (41–43) and may reflect the dominance of the NHEJ pathway in the G2-M1 (39). The TEL-AML1 translocations reported here compare more closely with EWS-FLI-1 translocations in Ewing’s tumors (50). Of 36 sequenced translocations and their reciprocals, 11 were balanced (≤6 bp deleted or duplicated at the junction), and 21 showed large deletions (mean, 56 bp) or duplications (mean, 118 bp), or both (50). Microhomologies of <4 bp were not reported. An additional four translocations demonstrated locally derived inverted sequences of 6–34 bp. We did not see such inversions with TEL-AML1 translocations, but we cannot rule out their potential formation because we analyzed fewer sequences. Zucman-Rossi et al. (50) suggest that the EWS-FLI-1 translocations are formed by DSB on either side, with formation of a single-strand intermediate by exonuclease processing, further processing by DNA polymerase and endo-/exonucleases, and, finally, interchromosomal joining. TEL-AML1 translocations share the complexity seen in EWS-FLI-1 translocations, which may be indicative of formation during cell growth rather than cycle arrest, since activities of the DSB pathways vary in cell cycle (39). It should be noted that microhomologies are also common in another leukemogenic translocation, t(15;17) (PML/RARA), also implicating the NHEJ pathway of formation (51).

Four translocations were located in close proximity to each other on TEL and lay within a region containing an Alu sequence and a polymorphic repeat region (Figs. 2 and 5). We found that the repeat region exhibited paraneoplastic structure in supercoiled plasmid DNA, and we hypothesize that this may contribute to the formation of the translocation. It may be asked whether such a structure would form in vivo. Paraneoplastic structure requires large amounts of energy in the form of negative torsional stress on DNA. Human DNA is continuously under such stress, and local regions experience high levels of stress in the negative supercoiling produced in the wake of a transcribing polymerase (52). The human genome, like that of all eukaryotes, is abundant in cruciform and H-DNA forming sequences, indicating that these sequences serve a functional role (53). Paraneoplastic structures have long been recognized to form as a result of gene transcription (35, 54) and may in some cases be an absolute requirement for promoter activity (55). The TEL gene is critical for hemopoiesis in the bone marrow (24) and is therefore transcribed probably at very early stages. This supports the scenario in which paraneoplastic structures in TEL form at the appropriate window of time when translocations may occur.

Paraneoplastic structures have been found near other translocations as well. Breakpoints in the TCL-1 locus involved in t(10;14) translocations in T-cell leukemias occur near an S1-hypersensitive site (56). DNA sequences with alternating purine/pyrimidine have been suggested to be involved in the formation of Z-DNA, which might contribute to translocation formation (10, 56, 57). A region of DNA with single-stranded character may be particularly prone to breakage. Repeated breakage-rejoining events could account for the highly polymorphic nature of the repeat region in TEL. The nuclease-sensitive sites may also invite strand invasion or aberrant repair processes involving another segment of DNA undergoing repair—the paraneoplastic structure being a form of DNA “hangnail.” Ku protein binds well to single- to double-strand transitions of DNA, including double-strand DNA plasmids containing a “bubble” or hairpin loops (58); this binding might facilitate an inappropriate repair process resembling NHEJ.

Four AML1 breakpoints occurred within 3600 bp of an exon. The regions surrounding these exons also contain regulatory elements of AML1 expression (59) and are likely to be in contact with proteins involved in transcriptional processing of DNA. Such regions are commonly associated with paraneoplastic structures (35), but we did not test for this for AML1. Further characterization of TEL-AML1 sequences will help to clarify whether this is a common theme.

It should be noted that TEL-AML1 transcripts occasionally are missing exon 2, indicating that translocation occurs between exons 2 and 3 (36). Such translocations are rare in our United Kingdom Childhood Cancer Study database. Whether the intron 2 rearrangements produce a fusion protein with different properties than that generated by intron 1 breaks or are less common because of a 10X smaller intron target for rearrangement is not known.

Finally, the nature of the possible molecular mechanisms that may generate TEL-AML1 fusion by illegitimate recombination is relevant to the biologically important issue of cell type or lineage specificity of translocations (27). The type of mechanism suggested by our data, NHEJ, is unlikely to provide the basis for selective rearrangement in a unique cell type. More probably, the exclusivity of TEL-AML1 fusion genes to B-cell precursor leukemia resides in a highly selective, cell context-dependent impact of the chimeric protein itself.

ACKNOWLEDGMENTS

We are grateful to Drs. S. West, L. M. Wiedemann, and A. M. Ford for suggestions and criticisms; Dr. T. Willis for technical advice; Dr. H. King for DNA sequencing; and B. Deverson for help in preparation of the manuscript. Clinical specimens were part of the United Kingdom National Case/Control Study of Childhood Cancer and an international survey of leukemia in identical twins. We thank the Management Committee and the United Kingdom Childhood Cancer Group for facilitating access to samples, Dr. C. Price for initial molecular screening, and colleagues who provided clinical samples from twins with leukemia (these being co-authors of separate, detailed reports on those patients).

ADDITIONAL

A study by Thandla et al. (60) published at the time the present work was submitted reports the sequencing of four TEL-AML1 translocations. Like the present study, Thandla et al. (60) found little evidence for a V(D)J recombinase mechanism and found clustering of breakpoints near the repeat region on TEL intron 5. In addition, AML1 breakpoints were located mainly in the 40 kb immediately 5’ to exon 2 (60).

REFERENCES

4 C. Price and M. Greaves, unpublished observations.


Structure and Possible Mechanisms of TEL-AML1 Gene Fusions in Childhood Acute Lymphoblastic Leukemia

Joseph L. Wiemels and Mel Greaves


Updated version

Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/59/16/4075

Cited articles
This article cites 59 articles, 37 of which you can access for free at:
http://cancerres.aacrjournals.org/content/59/16/4075.full.html#ref-list-1

Citing articles
This article has been cited by 24 HighWire-hosted articles. Access the articles at:
/content/59/16/4075.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.