Molecular Anatomy of a 5q Interstitial Deletion

By Lalitha Nagarajan, Beverly Lange, Linda Cannizzaro, Janet Finan, Peter C. Nowell, and Kay Huebner

A truncated granulocyte-macrophage colony-stimulating factor (GM-CSF) allele on a putative 5q- chromosome of HL-60 cells was cloned and, by comparison with counterpart normal sequences, analyzed for clues to molecular mechanisms facilitating rearrangement and deletion. Within the 17-kilobase (kb) pair locus surrounding the truncated GM-CSF gene remnant, there are no fewer than four rearranged genomic fragments that seemingly derive from chromosome 5 region q21→q23. Two of the fragments, which flank the truncated GM-CSF locus on the 5q-, are contiguous on the normal chromosome 5, centrometric to the normal GM-CSF allele, indicating at least one intrachromosomal insertion event, either preceded or followed by further deletion. Insertion and/or deletion was accompanied by juxtaposition of LINE sequences to the 5' side of the truncated GM-CSF locus within the inserted fragment. The entire rearranged locus is embedded in repetitive sequences, which may have mediated successive insertions or deletions. The extent of such stepwise deletions, resulting in loss of genes such as interleukin-3 (IL-3), IL-4, IL-5, and GM-CSF, whose gene products are critical to differentiation within the lineage of the affected hematopoietic stem cell, may be mirrored in the heterogeneity of symptoms and 5q- deletion sizes observed in myelodysplasias and acute leukemias carrying a 5q- chromosome. Perhaps most significantly, the sequences surrounding the insertion/deletion region are suggestive of recombination signals, including direct repeats and mirrored repeats. The site of insertion of the GM-CSF 3' region into an upstream (centromeric) locus is flanked by direct repeats; the upstream site into which it is inserted is also flanked by 12 base pair (bp) direct repeats. After insertion, one member of each pair of repeats is lost. The organization of this rearranged locus implies that the direct repeats had a role in the intrachromosomal recombination/deletion event.

MATERIALS AND METHODS

Genomic clones. Genomic libraries for HL-60 cells and cells from a cell line with normal chromosomes were prepared. DNAs were partially digested with endonuclease Sau3a, ligated to bacteriophage EMBL 3a arms, packaged and screened with appropriate probes as previously described. The HL-60 library was screened with the full-length 5.6 kb HindIII genomic subclone from the GM-CSF locus. Phages carrying the normal and rearranged GM-CSF loci from HL-60 were isolated and mapped. Repeat-free subclones in pUC vectors from the HL-60 rearranged GM-CSF locus were used to screen the second genomic library from the cell line with normal chromosomes to define the origin of the regions of the GM-CSF locus. All sequencing was performed with the dideoxy chain termination method as described previously.

Southern blot analysis. Total cellular DNAs from human and...
hybrid cell lines and primary tissues were isolated by phenol/chloroform extraction after lysis and proteinase K treatment.\textsuperscript{22,23} DNAs (~10 µg) were thoroughly digested with appropriate restriction enzymes, electrophoresed in 0.8% agarose, transferred to nylon filter, and hybridized to radiolabeled probes as previously described.\textsuperscript{9,22,23}

RESULTS

The partially deleted GM-CSF allele from HL-60 DNA has been molecularly cloned. Restriction enzyme analysis of this rearranged allele (Fig 1) confirmed that the locus has lost all GM-CSF coding information but carries a 3.1-kb genomic fragment of the 3’ flanking region of the GM-CSF locus. From the rearranged phage clone (Fig 1B) that was rich in human repetitive sequences, a repeat-free genomic probe, CPL 2.6, was subcloned from the region upstream of the truncated GM-CSF locus (Fig 1B) and localized to chromosome region 5q21 by somatic cell hybrid analysis and in situ chromosomal hybridization.\textsuperscript{26} The CPL 2.6 genomic probe detects a frequent SacI restriction fragment length polymorphism (RFLP) in the normal human population\textsuperscript{26} which detects deletion of one CPL 2.6 allele in DNA from tissue samples exhibiting the 5q- abnormality.\textsuperscript{27} DNAs shown in Fig 2 (lanes 2 and 9) were from two human cell lines heterozygous for this locus as evidenced by the presence of two distinct genomic fragments (3.5-kb allele A and 2.6-kb allele B; 1.1-kb fragment is a constant fragment); lanes 6 and 7 (Fig 2), respectively, contained DNA from BM white blood cells (WBCs) and peripheral blood lymphocytes (PBLs) of a normal homozygous individual; lanes 5 and 8 (Fig 2) contained DNA from BM WBCs of two 5q- cases that exhibited hemizygosity at this locus. The presence of both CPL 2.6 alleles in DNA from PBLs of a 5q- leukemia in remission (lane 10) and absence of allele A (3.5 kb) in BM WBCs (lane 11) and PB (lane 12) at relapse demonstrates the potential use of the CPL 2.6 RFLP in monitoring expansion of the 5q- clone in leukemias or myelodysplastic disorders with acquired interstitial deletions of chromosome 5. HL-60 DNA (Fig 2, lane 3) from which this probe was isolated, carries a normal allele A (3.5 kb) and a 7.0-kb rearranged allele.

Surprisingly, another repeat-free genomic probe, CPR 0.5, subcloned from the region downstream of the rearranged GM-CSF locus (Fig 1B) detected the same SacI polymorphism in normal human DNAs and a novel rearranged allele in HL-60 DNA (not shown). To understand the structural

---

**Fig 1.** Genomic maps of the normal and rearranged GM-CSF loci from HL-60 cells. Partially Sau3a-digested HL-60 DNA was ligated to bacteriophage (EMBL 3a) arms, packaged and screened with the full-length GM-CSF genomic clone\textsuperscript{8} as described previously.\textsuperscript{22,23} Restriction maps: Normal allele (A); rearranged allele (B). The truncated GM-CSF locus in the rearranged allele is denoted tGM-CSF. Two repeat-free probes isolated from the repeat-free regions (NCP) of the rearranged locus are designated CPL 2.6 and CPR 0.5. B, BamHI; C, SacI; H, HindIII, M, SmaI; and R, EcoRI.

**Fig 2.** CPL 2.6 probe detects a SacI polymorphism and deletion of one allele in 5q- cases. (A) DNA (10 µg per lane) from mouse cell line (lane 1), human cell line K562 (lane 2), HL-60 cell line (lane 3), mouse × HL-60 hybrid retaining normal chromosome 5 from HL-60 (lane 4), BM of a 5q- leukemia at diagnosis (lane 5), PBLs (lane 6) and BM (lane 7) from an individual with normal chromosomes 5, BM of a 5q- case (lane 8), and human leukemia cell line with uninvolved chromosomes 5 (lane 9). (B) DNA from patient with a 5q- leukemia (del 5q11-15q31-34); remission PB (lane 10), BM at relapse (lane 11), PBLs at relapse (lane 12). DNAs were thoroughly digested with restriction enzyme SacI, fractionated by electrophoresis on 0.8% agarose, transferred to nitrocellulose filters and hybridized to nick-translated radiolabeled probe CPL 2.6 as previously described.\textsuperscript{28} Sizes of CPL 2.6 specific fragments are shown on the left in kb. The CPL2.6 probe is a HindIII-HindIII fragment subcloned from the rearranged HL-60 locus shown in Fig 1: this probe has an internal SacI polymorphism and deletion of 5q- allele A (3.5 kb) from BM WBCs of two 5q- cases that exhibited hemizygosity at this locus. The presence of both CPL 2.6 alleles in DNA from PBLs of a 5q- leukemia in remission (lane 10) and absence of allele A (3.5 kb) in BM WBCs (lane 11) and PB (lane 12) at relapse demonstrates the potential use of the CPL 2.6 RFLP in monitoring expansion of the 5q- clone in leukemias or myelodysplastic disorders with acquired interstitial deletions of chromosome 5. HL-60 DNA (Fig 2, lane 3) from which this probe was isolated, carries a normal allele A (3.5 kb) and a 7.0-kb rearranged allele.

---
basis for this observation, the Sacl polymorphic locus (NCP) was isolated from a normal chromosome 5 and characterized (Fig 3). The NCP locus was rich in repetitive sequences except in the region of the 3.5-kb Sacl fragment, designated PLC 3.5 (Fig 3A). Analysis of the restriction maps of the normal Sacl polymorphic locus (Fig 3A) and the GM-CSF rearranged locus allowed several conclusions. The CPL 2.6 probe is identical to the repeat-free subclone NCP 2.6 from the normal chromosome 5 phage clone (Fig 3A), and the CPR 0.5 subclone is identical to sequences to the right of NCP (Fig 3A and B). Thus, the CPL 2.6 and CPR 0.5 fragments within the rearranged GM-CSF locus (Fig 3B) derived from a contiguous fragment of genomic DNA on human chromosome 5. An insertion of a 10-kb segment of genomic DNA has occurred in the middle of the 3.5-kb Sacl fragment (PLC 3.5, Fig 3A) of the normal Sacl polymorphic locus. The sequences to the left of CPL 2.6 on the HL-60 rearranged clone (Fig 3B), which contain repeats and thus are unsuitable as probes, are different from sequences on the left of the normal counterpart of CPL 2.6, the NCP fragment (Fig 3A), as shown by their entirely different restriction maps. In addition, the restriction map of the rearranged GM-CSF genomic clone matches, for each enzyme tested, the restriction map obtained from Southern blots of HL-60 genomic DNA has occurred in the middle of the 3.5-kb Sacl fragment, designated PLC 3.5, Fig 3A) of the normal Sacl polymorphic locus. The sequences to the left of CPL 2.6 on the HL-60 rearranged clone (Fig 3B), which contain repeats and thus are unsuitable as probes, are different from sequences on the left of the normal counterpart of CPL 2.6, the NCP fragment (Fig 3A), as shown by their entirely different restriction maps. In addition, the restriction map of the rearranged GM-CSF genomic clone matches, for each enzyme tested, the restriction map obtained from Southern blots of HL-60 genomic DNA.

Thus, the HL-60 truncated GM-CSF locus contains at least four distinct chromosome breaks and joinings or insertions. Possible molecular mechanisms underlying such complex recombinational events resulting in interstitial chromosomal deletion were investigated by nucleotide sequence analysis at the sites of insertion (Fig 4). Additional sequencing was performed to determine the nature of the inserted repetitive elements. The data showed that in the HL-60 5q- chromosome the NCP sequences around the site of insertion are AT rich (greater than 60%) and carry a number of mirrored and direct repeats at the site of insertion, as shown in Fig 4. The 12-bp direct repeat (TTCAGGACATGT) pair, preceded by a mirrored repeat, in the right half of the fragment-a sequence (Fig 4) is at least partially complemented at the extreme left of fragment a (CTGAA is the reverse complement of TTCAG and is also, in the reverse direction, preceded by a mirrored repeat). The region between the TTCAGGACATGT direct repeat pair is interrupted by insertion of the 3' end of the GM-CSF locus, and the left member of the repeat pair along with preceding sequences is lost on the 5q- chromosome. Analysis of sequences within the inserted fragment showed that a stretch of at least 200 nucleotides (Fig 4A, fragment d) was identical to the human LINE sequence published by Hattori et al28 (data not shown). The size of the inserted repetitive sequence (6 kb) is similar to that reported for a full-length LINE element, and its AT richness is similar to that reported by Furano et al29 for the rat LINE insertion. Together, these observations suggest that the entire repetitive element, between the rearranged CPL 2.6 and GM-CSF loci, is LINE derived as shown in Fig 4A. The presence of LINE sequences may suggest that because this region of the genome was undergoing recombination it captured a transposable element. Alternatively, the LINE sequences could have been involved in effecting the insertion.

Either scenario would result in a deleted chromosome 5 in a precursor cell which thus becomes haploid for gene loci in the critical deleted region. The HL-60 5q- chromosome, which was analyzed in detail in the absence of the normal chromosome 5 in somatic cell hybrids,27 has lost GM-CSF, IL-3, IL-4, and IL-5 loci, but retains27 FGFA, PDGFR, CSF1R, ADRBR, and CSF1 loci which are telomeric to the GM-CSF locus on normal chromosome 527 and the GRL locus which has recently been reassigned to a position telomeric to the GM-CSF locus.27 32 34

Direct evidence for involvement of a specific deleted or otherwise altered gene (GM-CSF, IL-3, or other) in the pathogenesis of the 5q- syndrome is lacking. Sequencing of the normal GM-CSF allele in HL-60 cells did not show point mutations (data not shown); in addition, HL-60 cells express a 1.1-kb GM-CSF transcript when induced to differentiate into granulocytes with TPA (K. Huebner and A. ar-Rushdi, unpublished observations), suggesting that the GM-CSF allele on the unaffected chromosome 5 is functional.

Repeat-free probes from the normal NCP and truncated GM-CSF loci did not detect short-range rearrangements in three other 5q- cases tested, but one NCP allele is consis-
MOLECULAR ANATOMY OF A 5q-

The genomic fragment used in the analysis of the appropriate NCP sequences is denoted a. The two rearranged regions sequenced from the HL-60 5q- allele are designated b and c. The normal 3' GM-CSF sequence was obtained by sequencing the corresponding region in the normal allele. A small fragment d was sequenced to identify the nature of the repetitive element present in the HL-60 rearranged locus. The entire inserted fragment is denoted e. (B) Direct comparison of NCP and 3' GM-CSF sequences with the HL-60 rearranged sequences. All sequencing was performed using the dideoxy chain termination method.22 The regions of sequence homology between NCP, 3' GM-CSF, and the HL-60 5q- b and c fragments (shown in A) is depicted. Mirrored repeats; direct repeats (→); inverted repeats (←); e represents the entire inserted fragment (10.0 kb) including fragment d (as shown in A). Search of the GENBANK database showed that the sequence of fragment d is identical to LINE sequences* between nucleotides 1496 to 1880 (data not shown).

DISCUSSION

Somatically acquired interstitial deletions in specific chromosomal regions are characteristic of many preneoplastic and neoplastic conditions.23 Understanding of the contributory role of characteristic deletions to etiology of specific tumors through reduction of specific loci to hemizygosity has been clarified by early theoretical considerations,66 and recent molecular analyses,28-41 but molecular mechanisms responsible for the predisposing deletions have not been studied.

Nucleotide sequences directly involved in deletion have been analyzed in specific cases, especially for some metabolic disorders, and possible genetic mechanisms involved in some of these deletions have been suggested,42,43 but for the generality of acquired large eukaryotic chromosomal deletions mechanisms of generation have not been studied because isolation of the deletion endpoints has not been possible. Elegant studies of deletions in prokaryotic genomes have suggested a mechanism of slipped strand mispairing of short-sequence homologies or repeated short sequences in generation of deletions,44,45, one of the repeat sequences involved was a pentanucleotide, CTCTT, which occurs in abbreviated form in our sequence just to the left of the inserted LINE element (Fig 4B) and in several other places in complementary form (GAGAA).

Perhaps the most striking sequence element involved in the HL-60 5q deletion, the 12-bp direct repeats (TTCAAGGACATGT), are similar (six matching, five complementary nucleotides, one inserted nucleotide) to the hepatitis B direct repeats (TTCACCTCTGC) in the cohesive ends of the viral genome.46 HBV genome integration often takes place in regions between the direct repeats,47 and in at least one case of HBV integration there was a cellular sequence 14 bp from the viral insertion which was the reverse complement of the HBV direct repeat that was the same distance from the viral–host junction.48 Variations on the theme TCTCC, or its complement, which occurs within the HBV repeat, appear in several places in the HL-60 rearranged locus, near a deletion in human mitochondrial DNA,49 on both sides of an immunoglobulin λ-bcl-2 translocation,50 and near regions involved in immunoglobulin isotype switching.51,52 Thus, although data are too scanty to suggest a detailed mechanism for the HL-60 5q insertion, the traces left by the insertion are tantalizingly suggestive of some intrachromosomal recombinatorial mechanism, perhaps similar to immunoglobulin isotype switching, which could have a physiologic function.

Isolation and analysis of the genomic fragments juxtaposed by the HL-60 5q deletion, although not yet providing probes to detect rearranged endpoints in other 5q- cases, leads to conclusions that might be relevant to this class of deletions. First, the deletion is not the result of a simple break and rejoicing; second, intrachromosomal insertion, possibly mediated by specific sequence elements near the point of insertion, can effect deletion. Certain implied conclusions

---

Fig 4. Sequence analysis of the rearranged GM-CSF and NCP loci in HL-60 cells. (A) Map positions of fragments used in the analysis. The genomic DNA fragment used in the analysis of the appropriate NCP sequences is denoted a. The two rearranged regions sequenced from the HL-60 5q- allele are designated b and c. The normal 3' GM-CSF sequence was obtained by sequencing the corresponding region in the normal allele. A small fragment d was sequenced to identify the nature of the repetitive element present in the HL-60 rearranged locus. The entire inserted fragment is denoted e. (B) Direct comparison of NCP and 3' GM-CSF sequences with the HL-60 rearranged sequences. All sequencing was performed using the dideoxy chain termination method.22 The regions of sequence homology between NCP, 3' GM-CSF, and the HL-60 5q- b and c fragments (shown in A) is depicted. Mirrored repeats; direct repeats (→); inverted repeats (←); e represents the entire inserted fragment (10.0 kb) including fragment d (as shown in A). Search of the GENBANK database showed that the sequence of fragment d is identical to LINE sequences* between nucleotides 1496 to 1880 (data not shown).
may also be generally applicable to interstitial deletions: Deletion is not the result of a single event but results instead from several stepwise rearrangements; deletions are not necessarily continuous (ie, remnants of loci internal to the deletion endpoints may remain); we know that several loci centromeric to CPL 2.6 have also been deleted on the HL-60 5q- chromosome.27 Thus, gene mapping based solely on determination of loci lost or retained in tissues exhibiting chromosomal deletions may lead to errors in gene localization and order.

In summary, we were able to analyze sequences involved in interstitial deletion on the HL-60 5q- chromosome because a previously cloned gene probe detected a nearby rearrangement. Many probes are now available for this region of chromosome 5, and longer range rearrangements might be detected by pulsed-field gel electrophoresis. When new deletion endpoints have been cloned and characterized, recombinatorial signals on deletion substrates may point to specific recombination mechanisms involved in generation of chromosomal interstitial deletions, a major class of genetic abnormality contributing to metabolic and neoplastic diseases of both inherited and sporadic forms. Finally, we propose, as an avenue for further study, that the large chromosomal interstitial deletions observed in myelodysplastic syndromes (eg, 5q- and 7q-) are the result of errors in a normal physiological process, possibly involved in generation of specific deletions necessary for expression of gene product(s) required in myeloid/erythroid terminal differentiation. This proposal is prompted by the recent description of the role of such chromosomal rearrangements in sporulating cells of Bacillus subtilis28; the mounting evidence that characteristic chromosomal translocations in T- and B-cell malignancies result from errors in physiologic immunoglobin superfamily gene rearrangements29; the observation that myelodysplasias arise from a multipotential hematopoietic stem cell with the potential for myeloid and lymphoid differentiation30 but the 5q- occurs only in the myeloid lineage; the occurrence of possible signal sequences in our analysis of the HL-60 5q- rearrangement. This proposal can be tested by a search for genomic rearrangements of 5q in terminally differentiated myeloid cells of normal individuals.

ACKNOWLEDGMENT
We thank Wendy Scattergood, Felicia Watson, and Esther Angert for expert technical assistance.

REFERENCES
13. Sutherland GR, Baker E, Callen DF, Campbell HD, Young IG, Sanderson CJ, Gason OM, Lopez AF, Vadass MA: Interleukin 3 is at 5q 33 and is deleted in the 5q-syndrome. Proc Natl Acad Sci USA 71:1150, 1988
19. Koblika BK, Dixon RAF, Frielle T, Dohman HG, Bol-
Committee on Structural Chromosome Changes in Neoplasia. Cytogenet.


Molecular anatomy of a 5q interstitial deletion

L Nagarajan, B Lange, L Cannizzaro, J Finan, PC Nowell and K Huebner

Updated information and services can be found at:
http://www.bloodjournal.org/content/75/1/82.full.html

Articles on similar topics can be found in the following Blood collections

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml