The *Sfpi-1* Proviral Integration Site of Friend Erythroleukemia Encodes the *ets*-Related Transcription Factor Pu.1

RALPH PAUL,¹ SCOTT SCHUETZE,¹ SUSAN L. KOZAK,¹ CHRISTINE A. KOZAK,² AND DAVID KABAT^{1*}

Department of Biochemistry and Molecular Biology, School of Medicine, Oregon Health Sciences University, Portland, Oregon 97201-3098,¹ and Laboratory of Viral Genetics, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892²

Received 19 June 1990/Accepted 26 September 1990

Previous studies identified a common site (*Sfpi-1*) for proviral integration in immortalized Friend erythroleukemias. cDNAs corresponding to a 1.5-kb *Sfpi-1* mRNA were isolated and sequenced. These were larger than an independently isolated *Sfpi-1* cDNA described by researchers from another laboratory, and they contained common differences from that sequence, including in the coding region four extra nucleotides that altered the reading frame. The properly translated protein is identical to Pu.1, a transcription activation factor that is related to the *ets* oncogene family. Genetic methods were used to map *Sfpi-1* with respect to other loci on mouse chromosome 2. Our results suggest that Pu.1 blocks erythroblast differentiation and thereby causes immortalization.

Previously, we (16, 21) and researchers from another laboratory (11, 12, 15) independently identified a common site for proviral integration (now designated *Sfpi-1* for SFFV proviral integration 1) in Friend erythroleukemias. Our evidence, which was based on studies using helper-free Friend spleen focus-forming virus (SFFV) that contained a small inserted tag of simian virus 40 DNA, suggested that proviral integration in *Sfpi-1* caused erythroblasts to lose their commitment to differentiate and to concomitantly become immortal (16, 21). We have now attempted to identify RNAs encoded by this region and to determine the role of the integrated provirus.

Figure 1 shows a map of the *Sfpi-1* genomic region, including the SFFV proviral integrations that we have identified in different erythroleukemias. Restriction endonuclease fragments from this entire region were used as hybridization probes for RNA blots. Only probes 1 through 3 hybridized under stringent conditions to discrete RNAs extracted from erythroleukemia cells. Probes 1 and 2 hybridized to 1.5-kb poly(A)-containing RNA that we previously described (16) and that was recently also analyzed by others (11, 12), whereas probe 3 identified an unusually small (ca. 200-nucleotide) RNA.

As described below and elsewhere (11, 12), these RNAs can be expressed in erythroleukemia cells by *Sfpi-1* alleles that lack an integrated provirus. Therefore, to ensure that we were analyzing RNAs transcribed from a leukemogenically relevant *Sfpi-1* allele, we used IP/IR erythroleukemia cells that contain an SFFV provirus in *Sfpi-1* but that lack the corresponding normal allele (16, 22). Accordingly, cytogenetic analysis indicated that IP/IR cells contain only one copy of mouse chromosome 2. Chromosome 2 contains *Sfpi-1* (12; also see below).

Poly(A)-containing RNA from IP/IR cells (2, 19) was used to construct a cDNA library of 2.5×10^6 phage in λ gt-10 (5). Among 4.5×10^5 plaques screened with probes 1 and 2, two positive phages were isolated. A third positive phage was isolated from a cDNA library constructed from F-745 erythroleukemia cells (3). Inserts were subcloned into the *EcoRI*

site of pGEM-3 (Promega Biotec, Madison, Wis.). Oligonucleotide primers (Midland Certified Reagent Co., Midland, Tex.) were used to sequence both strands of the cDNA clones by the dideoxynucleotide chain termination method (10, 20) with the Sequenase system (U.S. Biochemicals, Cleveland, Ohio). Regions of sequence compression were resolved by replacing dGTP with dITP in the reaction mixtures. The common sequence of these cDNAs and the sequence of the encoded protein are presented in Fig. 2. In general, the nucleotide sequence is similar to sequences recently reported by Moreau-Gachelin and coworkers (11, 12). However, our cDNAs are longer at both the 5' and 3' ends. Moreover, there are several differences between our sequences and theirs in the 5' and 3' untranslated regions. Most important, our cDNAs contain four additional nucleotides within the coding region at positions 789, 852, 857, and 941 that change the translational reading frame (Fig. 2).

Interestingly, the *Sfpi-1* protein encoded by our cDNAs is identical to the Pu.1 transcription activation factor that was recently described by Klemsz and coworkers (9). This identity also includes the 5' and 3' untranslated sequences of the cDNAs. However, our cDNA is 20 nucleotides longer at the 5' end than the reported Pu.1 cDNA clone. Klemsz et al. (9) also presented evidence that Pu.1 binds to a purine-rich DNA sequence, 5'-GAGGAA-3' (PU box), and that Pu.1 is related to the *ets* oncogene family of proteins. Structures and functions of other *ets*-related oncogenic proteins have also been described (1, 6, 9, 23). These proteins all contain PEST sequences (18) that cause them to turn over with short life spans, and they have been implicated in transcriptional control of mitosis and cellular differentiation (1, 6, 9, 23). Our results strongly suggest that Pu.1 is also an oncogenic protein. Moreover, since the Pu.1 in IP/IR erythroleukemia cells (Fig. 2) appears to be identical to that in normal cells (9), we suggest that oncogenesis does not require any change in the protein or mRNA sequences.

RNA blots that illustrate expression of the 1.5-kb *Sfpi-1*/Pu.1 RNA are shown in Fig. 3A and B. The RNA is synthesized in SFFV-infected erythroleukemia cells whether or not they contain a provirus in *Sfpi-1* (Fig. 3A). Moreover, this RNA occurs in a higher concentration in normal spleens than in erythropoietically active enlarged spleens from ane-

* Corresponding author.

FIG. 1. Map of the *Sfpi-1* genomic region showing the location and orientation of proviral integrations. The cleavage sites for *Bam*HI (B), *Eco*RI (E), *Hind*III (H), *Kpn*I (K), *Sal*I (S), and *Xba*I (X) are indicated. Position zero is arbitrarily defined as the site of SFFV proviral integration in erythroleukemia cell line IP/IR (represented as integration K beneath the genomic map). Symbols: ▲, sites of proviral integration in erythroleukemias examined; →, orientation of proviruses 5' to 3'. Erythroleukemias C, D, E, F, G, I, J, K, L, M, O, P, T, and U are labeled as previously described (21). Orientation of provirus P was not assignable because of an apparent partial deletion of the provirus. *Eco*RI-*Eco*RI fragments from throughout the genomic region were subcloned into pGEM-3 (Promega Biotec) and used as probes to detect unique transcripts. Single-copy probes 1 and 2, *Bam*HI-*Bam*HI fragments of approximately 800 and 500 bp, respectively, were found to hybridize to a unique 1.5-kb mRNA that was transcribed in an orientation opposite to that of the integrated proviruses. Probe 3 is a 2.2-kb *Sal*-*Eco*RI fragment that hybridizes to a small (ca. 200-nucleotide) RNA.

mic mice and from preleukemic mice infected either with Friend virus or with an erythropoietin-encoding retrovirus (Fig. 3B, compare lane 1 with lanes 2 through 4). These results imply that normal differentiating erythroid cells either lack *Sfpi-1* mRNA or contain much smaller amounts than normal spleen cells. The increased amounts of *Sfpi-1* RNA in immortalized erythroleukemia cell lines could be caused by proviral integration (e.g., in cell lines IP/IR, M1, and M4) or by alternative mechanisms (e.g., for cell lines M5 and M7). We also detected *Sfpi-1* RNA in normal bone marrow, brains, hearts, lungs, large intestines, thymic stromata, and a B-cell lymphoma and in a Leydig cell testicular tumor. This RNA was absent in stomachs, livers, kidneys, and thymocytes and in NIH 3T3 fibroblasts. High levels of Pu.1 RNA were previously found in macrophages and in B lymphocytes (9, 17).

To chromosomally map the murine *Sfpi-1* gene, we analyzed DNAs from the progeny of an intersubspecies backcross (8). DNAs from the inbred strains NFS/N and from *Mus musculus musculus* produce 33- and 40-kbp *Hpa*I fragments, respectively, that hybridize on Southern blots with probe 2. Of 53 progenies of the backcross [(NFS/N × m) F₁ × m], 30 contained the 33-kbp fragment, consistent with the expected 1:1 ratio for a single gene. The inheritance pattern for *Sfpi-1* was compared with that of other markers on chromosome 2, including *Hc* (the fifth component of complement) by using probe pMC104 to identify a *Pvu*II polymorphism and *Hdc* (histidine decarboxylase) by using a rat cDNA probe to identify an *Xba*I polymorphism as described previously (8). Of the 53 mice, 8 recombinants were identified between *Hc* and *Sfpi-1* (percent recombina-

FIG. 2. Sequence of the *Sfpi-1* cDNA. This sequence was determined from the IP/IR cDNA clone designated P15 and numbered from the most 5' nucleotide. A second IP/IR clone, P7, was also sequenced and extended from bp 28 to 1,333. A third clone, derived from an F-745 erythroleukemia cDNA library, had 100% homology with the P15 sequence throughout its length from bp 1 to 670. The predicted amino acid sequence is shown above the corresponding DNA sequence. The stop codon for the amino acid sequence shown is indicated (●). S1 and E1, Start and end sites, respectively, of the *Sfpi-1* cDNA sequence reported by Moreau-Gachelin et al. (12); S2 and E2, start and end sites of the Pu.1 sequence reported by Klemsz et al. (9); aa, amino acids. Nucleotides that were missing (boxed) from the *Sfpi-1* sequence reported by Moreau-Gachelin et al. (11, 12) and an additional T residue at bp 1,034 reported by Moreau-Gachelin et al. (11, 12) that was not observed in the clone P15, P7, or Pu.1 sequences (arrowhead) are shown.

tion = 15.1 ± 4.9 [mean $\pm$ standard error]) and 7 other mice were recombinant between *Hdc* and *Sfpi-1* (percent recombination = 13.2 ± 4.6). These data are compatible with previous evidence (12) and suggest the gene order centromere-*Hc*-*Sfpi-1*-*Hdc*. The *Sfpi-1* map location confirms that *Sfpi-1* is distinct from other known oncogenes.

Our results suggest that SFFV proviral integration in *Sfpi-1* contributes to leukemogenesis without altering the structure of the *Sfpi-1* mRNA or Pu.1 protein. Presumably, the provirus causes enhancer-mediated constitutive expression of this gene. Our previous results imply that this expression may block the cellular commitment to terminally differentiate, thereby causing indefinite proliferation of immortal erythroleukemia cells (16, 21).

After this paper had been submitted, Goebl (4) described the similarity between the previously reported murine *Sfpi-1*

FIG. 3. RNA blot analysis demonstrating active transcription in the *Sphi-1* locus. RNAs from mouse erythroleukemia cell lines (A) and from mouse tissues (B) were isolated (2, 19). Samples were electrophoresed through 1.2% agarose gels following denaturation with glyoxal and dimethyl sulfoxide (19) and transferred to a nytran membrane (Schleicher and Schuell, Keene, N.H.). The uniform transfer and integrity of RNA were visually determined by soaking the membrane in 0.3 M sodium acetate (pH 5.5)–0.02% methylene blue. The blots were hybridized at 42°C with a *Sphi-1* cDNA probe. The final wash was at 53°C with 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate. Erythroleukemia cell lines previously described by Spiro et al. (22) and Paul et al. (16) are identified here with the same nomenclature. (A) Five micrograms of poly(A)-containing RNA. Lanes: 1, the PA12 line of mouse NIH 3T3 fibroblasts; 2 through 6, IP/IR, M1, M4, M5, and M7 erythroleukemia cells, respectively. Cell lines M5 and M7 lack any provirus in *Sphi-1* (16). (B) Twenty micrograms of total RNA. Lanes: 1, normal spleens; 2, spleens from phenylhydrazine-treated mice; 3, enlarged spleens harvested 10 days after infection with an erythropoietin-producing virus (7); 4, preleukemic spleens harvested 14 days after infection with Friend virus; 5, IP/IR erythroleukemia cells; 6, F-745 erythroleukemia cells. A 0.24- to 9.5-kb RNA ladder (Bethesda Research Laboratories, Gaithersburg, Md.) was used as a size marker.

(11, 12) and Pu.1 (9) cDNA sequences, and he inferred that the *Sphi-1* sequences probably contained errors, including the four missing C residues in the coding region that we have observed (Fig. 2). In two letters (13, 14), Moreau-Gachelin et al. agreed that *Sphi-1* and Pu.1 are identical, although they acknowledged only one missed C residue in their sequence at position 789 (our numbering system). Clearly, however, adding only this one C residue to their sequence would leave major differences from the Pu.1 cDNA and derived protein sequences, specifically, a string of 28 incorrect amino acids due to frameshifting, and a protein shorter by 1 amino acid. The protein would be most altered in the region conserved in the *ets* oncogene family of proteins (23). A recently described human *Sphi-1* cDNA sequence (17) also appears to have missing nucleotides at three positions identical to and at one site proximal to those missing in the two murine cDNA sequences (11, 12). Our results agree with the suggestions of Goebel (4) and provide strong evidence that the *Sphi-1* and Pu.1 genes are the same.

Nucleotide sequence accession number. The nucleotide sequence accession number issued by GenBank for the sequence reported in this paper is M38252.

We thank our colleagues Maureen Hoatlin, Hao Wang, and Heather Beghtol for help and encouragement, John Adelman for

generous advice and help in constructing the IP/IR cDNA library, Vikram Patel (Northwestern University, Evanston, Ill.) for generously donating the F-745 cDNA library, Miles Wilkinson for generously donating several RNA blots and RNA samples, and Charles Van Beveran (La Jolla Cancer Research Foundation, La Jolla, Calif.) for helpful discussions.

This research was supported by grant CA25810 from the U.S. Public Health Service and by grants from the Alberta Heritage Foundation (to R.P.), The Medical Research Foundation of Oregon (to S.S.), and the Gerlinger Trust.

REFERENCES

- Bhat, N. K., C. B. Thompson, T. Lindsten, C. H. June, S. Fujiwara, S. Koizumi, R. J. Fisher, and T. S. Papas. 1990. Reciprocal expression of *ets 1* and *ets 2* genes during T-cell activation: regulatory role for the protooncogene *ets 1*. *Proc. Natl. Acad. Sci. USA* 87:3723–3727.
- Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294–5299.
- Friend, C., W. Scher, J. G. Holland, and T. Sato. 1971. Hemoglobin synthesis in murine virus-induced leukemic cells *in vitro*: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proc. Natl. Acad. Sci. USA* 68:378–382.
- Goebel, M. G. 1990. The PU.1 transcription factor is the product of the putative oncogene *Spi-1*. *Cell* 61:1165–1166.
- Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263–269.
- Gunther, C. V., J. A. Nye, R. S. Bryner, and B. J. Graves. 1990. Sequence-specific DNA binding of the proto-oncoprotein *ets-1* defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. *Genes Dev.* 4:667–679.
- Hoatlin, M., S. L. Kozak, F. Lilly, C. Kozak, and D. Kabat. *Proc. Natl. Acad. Sci. USA*, in press.
- Joseph, D. R., P. M. Sullivan, Y.-M. Wang, C. Kozak, D. A. Fenstermacher, M. E. Behrendsen, and C. A. Kahn. 1990. Characterization and expression of the complementary DNA encoding rat histidine decarboxylase. *Proc. Natl. Acad. Sci. USA* 87:733–737.
- Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveran, and R. A. Maki. 1990. The macrophage and B cell-specific transcription factor Pu.1 is related to the *ets* oncogene. *Cell* 61:113–124.
- Mierendorf, R. C., and D. Pfeffer. 1987. Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* 152:556–562.
- Moreau-Gachelin, F., D. Ray, N. J. de Both, M. J. M. van der Feltz, P. Tambourin, and A. Tavittian. 1990. *Spi-1* oncogene activation in Rauscher and Friend murine virus-induced acute erythroleukemias. *Leukemia* 4:20–23.
- Moreau-Gachelin, F., D. Ray, M.-G. Mattei, P. Tambourin, and A. Tavittian. 1989. The putative oncogene *Spi-1*: murine chromosomal localization and transcriptional activation in murine acute erythroleukemias. *Oncogene* 4:1449–1456.
- Moreau-Gachelin, F., D. Ray, P. Tambourin, and A. Tavittian. 1990. The putative oncogene *Spi-1* encodes a transcription factor. *Oncogene* 5:941.
- Moreau-Gachelin, F., D. Ray, P. Tambourin, A. Tavittian, M. Klemsz, S. R. McKercher, A. Celada, C. Van Beveran, and R. A. Maki. 1990. The PU.1 transcription factor is the product of the putative oncogene *Spi-1*. *Cell* 61:1166.
- Moreau-Gachelin, F., A. Tavittian, and P. Tambourin. 1988. *Spi-1* is a putative oncogene in virally induced murine erythroleukemias. *Nature (London)* 331:277–280.
- Paul, R., S. Schuetze, S. L. Kozak, and D. Kabat. 1989. A common site for immortalizing proviral integrations in Friend erythroleukemia: molecular cloning and characterization. *J. Virol.* 63:4958–4961.
- Ray, D., S. Culine, A. Tavittian, and F. Moreau-Gachelin. 1990. The human homologue of the putative proto-oncogene *Spi-1*: characterization and expression in tumors. *Oncogene* 5:663–668.

18. **Rogers, S., R. Wells, and M. Rechsteiner.** 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**:364–368.
19. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol. 1, p. 7.1–7.38. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Sanger, F., S. Nicklen, and R. A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
21. **Spiro, C., B. Gliniak, and D. Kabat.** 1988. A tagged helper-free Friend virus causes clonal erythroblast immortality by specific proviral integration in the cellular genome. *J. Virol.* **62**:4129–4135.
22. **Spiro, C., B. C. Gliniak, and D. Kabat.** 1989. Splenic accumulation of interleukin-3-dependent hematopoietic cells in Friend erythroleukemia. *J. Virol.* **63**:4434–4437.
23. **Watson, D. K., M. J. McWilliams, P. Lapis, J. A. Lautenberger, C. W. Schweinfest, and T. S. Papas.** 1988. Mammalian *ets-1* and *ets-2* genes encode highly conserved proteins. *Proc. Natl. Acad. Sci. USA* **85**:7862–7866.