ferent *sis* deletion mutants were inserted into this signal peptide vector so that the correct reading frame was maintained across the junction. These constructs were expected to encode an *env*(MuLV)-*sis*(SSV) fusion protein. Two of the deletion mutants, pDD154 and pDD156, regained full biological activity. The third, pDD160, was still unable to induce focus formation (Fig. 3), a result that was not surprising, as pDD160 contains a significant deletion of the PDGF-related region. The observation of full biological activity in pDD154 and pDD156, in which the PDGF-related region remained intact, confirms our hypothesis that a functional signal sequence is required for transformation by the v-*sis* gene product.

Our results were obtained by molecular manipulations of coding regions and subsequent correlations with biological activity. A more direct test of our model would be to examine the size and structure of the *sis*-related proteins synthesized in cells infected with some of the nontransforming mutants described above. To date, such experiments have yielded inconclusive results because of the lack of a highly reactive antibody to the *sis* gene product. We believe that our mutations are altering processing of the *sis* gene product, rather than transcription or translation for the following reasons: (i) A single base change in the coding region of the proposed signal sequence abolished transformation. This mutation would not be expected to affect transcription nor the initiation of translation. (ii) The deletion mutant pMH18 leaves all transcriptional signals and all six ATG codons of the *env-sis* coding region intact, yet is biologically nontransforming. (iii) RNA synthesized in vitro (*16*) from the wild-type *sis* gene and also from the nontransforming mutant pDD143 was successfully translated by means of a reticulocyte lysate system (*17*) to yield polypeptides of the anticipated molecular weights (*15*). Thus, the deleted *sis* gene in pDD143 can be transcribed to yield translatable RNA.

According to one model, transformation results from the interaction between the PDGF-related protein and the cell surface receptor for PDGF. The v-*sis* gene product is predicted to be either an integral membrane protein or a secreted protein, which would facilitate interaction with the PDGF receptor. Our experiments are consistent with the hypothesis that a signal sequence is required for transformation by v-*sis*. Further, these data indicate that the primary translation product of the *env-sis* coding region, in the absence of any proteolytic processing, would be either 33 or 30 kilodaltons depending on whether the first or second ATG codon is used. This is in contrast to other proposals that the primary translation product initiates at the third ATG codon and is about 28 kilodaltons (*4–8*).

Our results demonstrate a novel mechanism whereby viral encoded sequences may activate a cellular protooncogene. The activation of cellular sequences encoding a PDGF-related protein would occur as a result of viral signals which control transport of the viral *env* gene product. The point of integration of cellular sequences into the retroviral genome may be an important factor in the activation of those sequences into a transforming oncogene. Thus our data suggest that alteration of the cytological location of a growth factor can result in cellular transformation.

MARK HANNINK
DANIEL J. DONOGHUE*
Department of Chemistry, B-017,
University of California,
San Diego, La Jolla 92093

References and Notes

1. G. H. Theilen, D. Gould, M. Fowler, D. L. Dungworth, *J. Natl. Cancer Inst.* **47**, 881 (1971).
2. E. P. Gelman, F. Wong-Staal, R. A. Kramer, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3373 (1981).
3. K. C. Robbins, S. G. Devare, S. A. Aaronson, *ibid.*, p. 2918.
4. S. G. Devare, E. P. Reddy, J. D. Law, K. C. Robbins, S. A. Aaronson, *ibid.* **80**, 731 (1983).
5. S. F. Josephs, C. Guo, L. Ratner, F. Wong-Staal, *Science* **223**, 487 (1984).
6. R. F. Doolittle *et al.*, *ibid.* **221**, 275 (1983).
7. M. D. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983).
8. K. C. Robbins *et al.*, *ibid.* **305**, 605 (1983).
9. T. F. Deuel, J. S. Huang, S. S. Huang, P. Stroobant, M. D. Waterfield, *Science* **221**, 1348 (1983).
10. D. J. Donoghue, in preparation.
11. When inserted in the MuLV-derived expression vector, the *sis* gene should be expressed from a spliced messenger RNA very similar in structure to the *sis*-specific messenger RNA produced by SSV. In cells transformed by SSV, a spliced messenger RNA of 2.7 kilobases has been identified [A. Eva *et al.*, *Nature (London)* **295**, 116 (1982)] which presumably arises by splicing to the splice acceptor site of the SSAV *env* gene retained in the SSV genome. Although the 3′-splice acceptor site in SSV is not known, examination of the SSV nucleotide sequence reveals several potential splice acceptor sites upstream of the *env-sis* region.
12. F. L. Graham and A. J. van der Eb, *Virology* **52**, 456 (1973); P. Andersson, M. P. Goldfarb, R. A. Weinberg, *Cell* **16**, 63 (1979).
13. L. E. Henderson, T. D. Copeland, G. W. Smythers, H. Marquardt, S. Oroszlan, *Virology* **85**, 319 (1978).
14. T. M. Shinnick, R. A. Lerner, J. G. Sutcliffe, *Nature (London)* **293**, 543 (1981).
15. M. Hannink and D. J. Donoghue, data not shown.
16. M. R. Green, T. Maniatis, D. A. Melton, *Cell* **32**, 681 (1983).
17. H. R. B. Pelham and R. J. Jackson, *Eur. J. Biochem* **67**, 247 (1976).
18. The initial DNA sequence data suggested that v-*sis* was actually a *pol-env-sis* fusion gene in which the first two ATG codons shown in Fig. 1C were not present [S. G. Devare *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3179 (1982)]. Subsequent corrections to the SSV sequence (*4*) resulted in the structure of the *env-sis* coding region as shown in Fig. 1C.
19. A. Maxam and W. Gilbert, *ibid.* **74**, 560 (1977).
20. We thank Russell Doolittle, Tony Hunter, Immo Scheffler, Bart Sefton, and Suresh Subramani for advice and criticism and Richard Bold for excellent technical assistance. This work was supported by PHS grant CA34456 awarded by the National Cancer Institute, DHSS; the Chicago Community Trust/Searle Scholars Program (D.J.D.); and a Cell and Molecular Biology Training grant, GM07313 (M.H.).
* To whom reprint requests should be addressed.

14 May 1984; accepted 24 September 1984

Constitutive Fragile Sites and Cancer

Abstract. *Breaks were observed at 51 sites in homologous chromosomes in lymphocytes from ten humans and two great apes when cells were deprived of thymidine. The incidence of breaks was enhanced by caffeine, a substance that inhibits DNA repair in replicating cells. The locations of 20 sites were correlated with breakpoints that have been related to human malignancy.*

A few specific chromosome sites tend to be expressed as gaps or breaks during routine human metaphase chromosome preparations (320 bands per haploid set). With the exception of sites 3p14.2, 6q25.3, and 16q23.2, these sites have appeared infrequently and at low levels of expression, and therefore have been difficult to characterize (*1*). Deprivation of folic acid and thymidine results in the expression of 13 of the 16 known heritable fragile sites (h-fra), and in the appearance of "spontaneous" chromosome breaks (*2*), here termed constitutive fragile sites (c-fra). By examining elongated chromosomes [670 to 850 bands per haploid set (*3*)], we have found a large family of c-fra located at precise points in homologous chromosomes of human and primate genomes. The expression of these genomic weak points is enhanced when cells are exposed to 2.2 m*M* caffeine during the last 6 hours of culture.

Blood from five males and five females ranging in age from 10 to 65 years was studied. Eight of these individuals were normal; one was mentally retarded and had h-fra Xq28.1 (human 2 in Table 1); one had acute leukemia inv(16)(p13.1q22.1), and c-fra 16q22.1 (human 3 in Table 1). We also donated blood after having taken 5 mg of folinic acid a day orally for 3 days (humans 1 and 4 in Table 1). In addition, blood was also studied from one 24-year-old female chimpanzee (*Pan troglodytes*) and one 34-year-old female gorilla (*Gorilla gorilla*). We used blood lymphocytes which

Table 1. Expression of constitutive fragile sites (c-fra) as chromosome breaks per 100 mitoses in blood lymphocytes of humans and apes. Culture conditions: A, FTD media; B, FTD media plus caffeine; C, same as B after individual received 5 mg of folinic acid a day orally for 3 days; D, FdU; E, FdU plus caffeine; F, FdU plus caffeine plus thymidine. Values in columns A and D were based on 1000 cells; B and E were based on 200 cells; C and F were based on 500 cells.

c-fra	Human											Chimp	Gorilla
	1						2		3		4–10		
	A	B	C	D	E*	F	D	E	D	E	E†	E	E
1p31.2	0.1	2.5	0	0.6	5.5(3.5)	0	0	5	0.6	12	6.6(3-10)	12.5	13
p21.2	0.1	3.5	0	0.1	11(10)	0	0.6	11	0.3	15	8.3(4-11.5)	9	6
q21.3	1.2	1	1.2	1.8	4(3)	0.2	1.8	7.5	0.2	4.5	5.5(2-11)	2	3
q25.1	0	3	0.2	0.5	3.5(3.5)	0	0	9.5	0.2	4	4(1-5.5)	2	3
q44.1	0	5.5	0.8	0.5	18(14)	0.2	0.3	23	0.9	23	10.4(4-22)	8.5	18
2p24.2	0.1	3.5	0	0.3	9(8)	0	0	10	0.4	8.5	6.2(3-10)	9.5	2
p16.2	0	1.5	0	0.1	5(4)	0	0	6	0.2	6	3.5(1.5-5.5)	9	1
p13	0.1	1	0	0	3(3)	0	0	3.5	0.3	5	4.4(1.5-6)	1.5	2
q21.3	0.2	2	0.2	0.1	4.5(4.5)	0	0	7	0.1	5	5.3(2-12)	6	0
q32.12	0.4	4.5	0.2	0.1	14(13)	0	0	8	0.2	11	7.9(2-20)	12	4
q37.3	0.2	3	0	0.6	9(8)	0	0.3	16	0.4	11.5	5.6(2.5-10.5)	3	3
3p24.2	0	2	0.2	0.2	8.5(8)	0	0	4	0	4	4.6(2-7)	3.5	3.5
p14.2	2.9	37	0.6	9.3	45(31)	0.4	6.9	105	5	41	71.4(43-126)	41	23.5
q27	0.2	1	0	0	3.5(3.5)	0	1.2	5	0.4	5	4.1(2-6.5)	6	3.5
4p16.1	0.1	2	0.2	0	5(4)	0	0.3	3	0.2	4	4.4(2-7)	0	0.5
q31.1	0.3	2.5	0.2	0.6	9(7)	0.4	0	3.5	0.3	12.5	5.1(2-8.5)	9	2.5
5q31.1	0.4	2	0	0.8	5(4)	0.2	0.6	4.5	0.4	4	5.1(2.5-8)	3	3
6p25.1	0.3	6	0.4	0.6	14(11)	0	0.6	15	1	16	11.8(2.5-26)	6	21.5
p22.2	0.4	1	0	0.2	4.5(4)	0	0.6	4	0.4	4	4(2-7)	6	1.5
q25.3	0.2	4	0	1.1	8(6)	0.2	1.5	10	1.3	12	10.9(3.5-21)	3	8
7p22	0.1	3.5	0.2	0.2	4(4)	0	0	4	0.4	8	5.6(2-9)	11	3.5
p14.2	0.4	2.5	0	0.4	9(6)	0	0.3	13.5	0.5	12	6(3.5-9)	5	2
q21.2	0.6	1.5	0	0.5	4(4)	0	1.8	4	0.2	6	4.5(2.5-8)	10	9
q31.2	0	2	0	0.5	4.5(4.5)	0.4	0	8	0.4	9.5	7.5(4.5-12)	10	15.5
q32.3	0.2	7.5	0.2	1.0	26(21)	0	0.9	39	2	30	17.7(6-31)	11	7
8q22.1	0.4	2.5	0	0.7	11(9)	0	0.6	8	0.5	10	7.1(4-16)	1.5	9
q24.1	0	1.5	0.2	0.2	4(4)	0	0	5	0.3	1.5	3.9(2-5.5)	8	2
q24.3	0.2	1.5	0	0	5(5)	0	0.3	4	0.1	6	4.4(.5-8)	6	3
9q22.1	0	1.5	0	0.3	4(4)	0	0.9	5	0.1	2.5	3.1(1.5-4.5)	2.5	0
q32	0.1	3	0	0.4	6(5)	0	0.3	10	0.3	13	6.4(3-11)	2.5	0.5
10q22.1	0.2	2.5	0.2	0.4	2(2)	0	0.6	6.5	0.6	8	4(3-6)	5	0
q25.2	0	2	0	0.1	5.5(5)	0.2	0	4	0.2	4	3.7(2-5)	3	0
q26.13	0.1	2.5	0	0.3	5(4)	0	0.3	6	0	7	4.9(2-9)	3	0
11p15.1	0.1	1	0.2	0.1	4(3)	0	0	5	0.7	5.5	4.1(2.5-7)	10	8.5
p14.2	0.6	6.5	0.4	0.3	19(16)	0.6	0	17.5	0.5	18	8.3(3-23)	7.5	11.5
q14.2	0.2	1.5	0	0.3	10(7)	0.2	1.2	2	0.4	9	6.8(4-11)	9	6
q23.3	0	1.5	0	0.1	4(4)	0	0	2.5	0	0	4(2-7)	1.5	2
12q21.32	0.1	2	0	0.1	9(7)	0	0.3	6	0.7	8	3.8(0-8)	5	2.5
13q13.2	0	1.5	0.2	0.4	15(14)	0	0	13	0.4	11	4.1(1-8)	11	0.5
q21.2	0	2	0	0.5	5(5)	0	0	2.5	0.2	2	5(2-9)	5	1
14q13	0	2	0	0	4(3)	0	0	4.5	0	3	3.6(2-5)	3.5	0.5
q24.11	0.3	3.5	0	1.0	11(10)	0	0.9	15	0.5	10	6.3(3-9)	3.5	0.5
16q22.1	0.2	2.5	0	0.6	4.5(3.5)	0.4	0.6	2	2.2	21	4.9(3-7)	5	3.5
q23.2	1.2	36	0.8	4.1	41(36)	0.6	3.6	90.5	4.1	46.5	38.6(31-54)	24	10
17q23.1	0	3.5	0.4	0.3	5(4)	0	0	4	0.3	6	3.6(0.5-5)	2	2.5
18q12.2	0.1	2	0	0.2	6(6)	0.5	0	10	0.2	8	6.3(3-9)	3.5	6
q21.3	0	1.5	0	0	4(4)	0	0	4.5	0.2	4	4.1(1-8)	3	3
20p12.2	0.1	2	0	0	4.5(4)	0	0	5	0.1	5.5	4(3-6)	9	1
22q12.2	0	1	0	0.2	4(4)	0	0	5	0.1	5	4.7(2-7)	1.5	1
Xp22.31	0.3	17	0.4	2.1	48(40)	1	0	29	3.1	20	26.6(15-48.5)	39	26.5
q22.1	0.6	4.5	0	0.2	15(13)	0	0.6	15.5	0.9	11.5	8.6(3-15.5)	11	7
c-fra/100 cells	13.3	242	7.4	33	492(414)	5.5	27.9	611	33	519	415.3	385	267.5
Breaks/100 cells	19.8	282	15	45.6	601(519)	10.2	39.7	803	40.3	653	562.9	496	351
Percent c-fra	70	86	49	72	82(80)	54	70	76	82	80	74	78	76

*Numbers in parentheses represent percent of c-fra after correction for homozygous expression. †Numbers under humans 4–10 represent mean values and, in parentheses, ranges.

were stimulated with phytohemagglutinin M (4-day) cultured in (i) modified Eagle's medium (MEM) that was lacking folic acid and thymidine (FTD-media; Gibco), and supplemented with fetal bovine serum (5 percent) (*2*), or (ii) MEM media supplemented with fetal bovine serum (10 percent). In the latter case, cells were exposed during the final 24 hours to 0.1 m*M* fluorodeoxyuridine (FdU; Sigma) (*4*). After treatment, cells were exposed to Colcemid (0.05 μg/ml) for 20 minutes and 0.075*M* potassium chloride for 10 minutes, and fixed rapidly in a mixture of absolute methanol and acetic acid (3:1). Chromosome preparations were made by dropping the cell suspension from a height of 5 feet onto slides and then G-banding the chromosomes (*3*). When cells were cultured in MEM (in the presence of folic acid and thymidine), with or without caffeine, fewer than 1 percent of the chromosomes had visible breaks.

Fragile sites were mapped according to the nomenclature system for high-resolution human chromosomes (*3*). This system was also used for ape chromosomes as the nonheterochromatic bands of humans appears to correspond to those of great apes (*5*). Between 21 and 38 percent of the human lymphocytes showed one to three chromosome breaks per mitotic figure after being cultured in FTD media or exposed to FdU. In each culture, 1000 mitotic figures were analyzed in order to identify a relatively large number of breaks (approximately 450 breaks for cells exposed to FdU). Sites at bands 3p14.2, 16q23.2, and Xp22.31 were generally expressed at a level of 2 to 9 percent in the ten individuals tested. Breakage at the 51 sites accounted for 70 to 82 percent of all breaks scored in mitoses, the majority of which had 670 to 850 bands per haploid set (Table 1). While the same fragile sites were revealed by the two techniques, a twofold higher frequency of breaks occurred in the presence of FdU as compared to culturing in FTD medium (Table 1) (*6*).

When FdU-treated cultured human lymphocytes were exposed to 2.2 m*M* caffeine (Sigma) during the last 6 hours of culture, we observed a tenfold enhancement (at a minimum) in the expression of fragile sites relative to that seen in cultures treated with FdU alone. Between 11 and 63 breaks per mitotic figure appeared in approximately half of the cells (Fig. 1a) while one to ten breaks per metaphase were observed in the other half. Only 1 to 5 percent of the metaphases showed no chromosome breaks. It was possible to identify 1000 breaks from each individual by studying only 125 to 200 mitoses (Table 1). Approximately half of all sites were expressed six or more times and the other half three to five times per 100 cells (Table 1, combined average of ten humans, columns E) (*7*). Since c-fra sites were expressed in a homozygous state in some cells, the actual frequency of a given site per 100 cells was somewhat lower than that indicated above (human 1E in Table 1). Although homozygous expression for each c-fra was not found when 100 mitoses were examined (Table 1), each c-fra was observed to express homozygously in at least one mitosis when additional cells were analyzed in every individual.

Most of the breaks observed in the presence of caffeine were represented in the 51 c-fra sites (74 to 82 percent) and the rest were scattered among the chromosome bands (frequently in Giemsa-negative bands). The distribution of breaks among the 51 c-fra sites was similar in the presence and absence of caffeine, indicating that the expression of the c-fra sites is not elicited but only enhanced by caffeine. In addition to increasing the expression of c-fra sites, caffeine also increased the expression of the heritable fragile site Xq28.1 in a patient with a known h-fra Xq28.1 (human 2 in Table 1; showing 33 and 60 percent expression after exposure to FdU or FdU and caffeine, respectively), and of c-fra 16q22.1 in an individual whose bone marrow showed acute leukemia and inv(16)(p13.1q22.1) (human 3 in Table 1).

Chromosome breaks were visible at the 51 c-fra sites when lymphocytes from a chimpanzee and a gorilla were treated with FdU and caffeine (Table 1). In addition, three sites were revealed in the chimpanzee chromosomes at 1p32.1 (11.5 percent), 6p23 (10 percent), and 14q21.3 (5 percent), which were not expressed in human cells. The gorilla chromosomes showed a c-fra at 1p32.1 (8.5 percent) and a lower incidence of breakage at the 51 c-fra sites as compared to human chromosomes.

As previously reported for h-fra cells treated with FTD and FdU in the absence of caffeine (*2*), 92 to 99 percent of all constitutive breaks in humans and primates failed to be expressed when

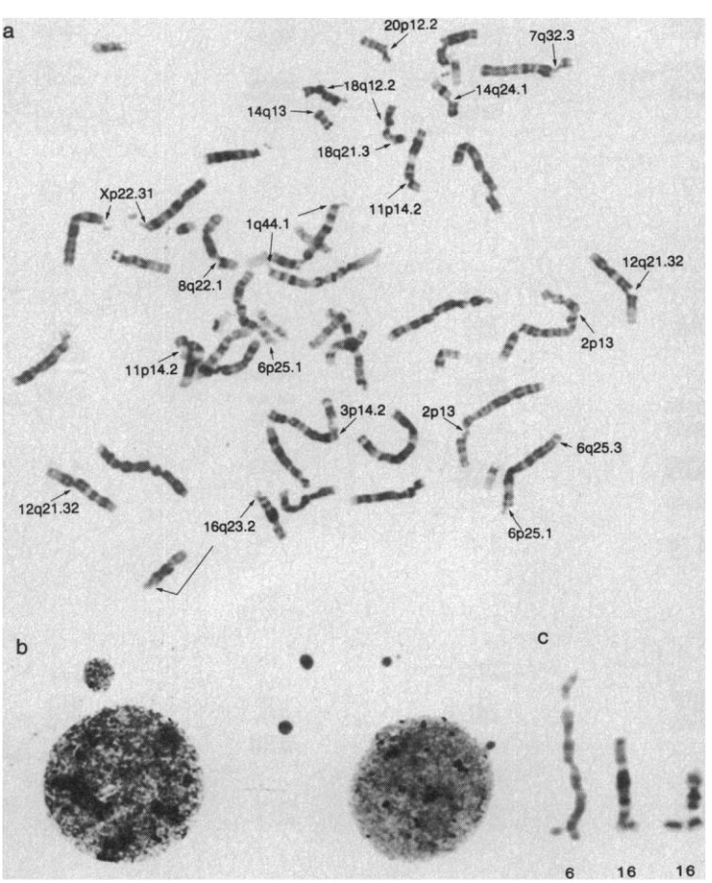
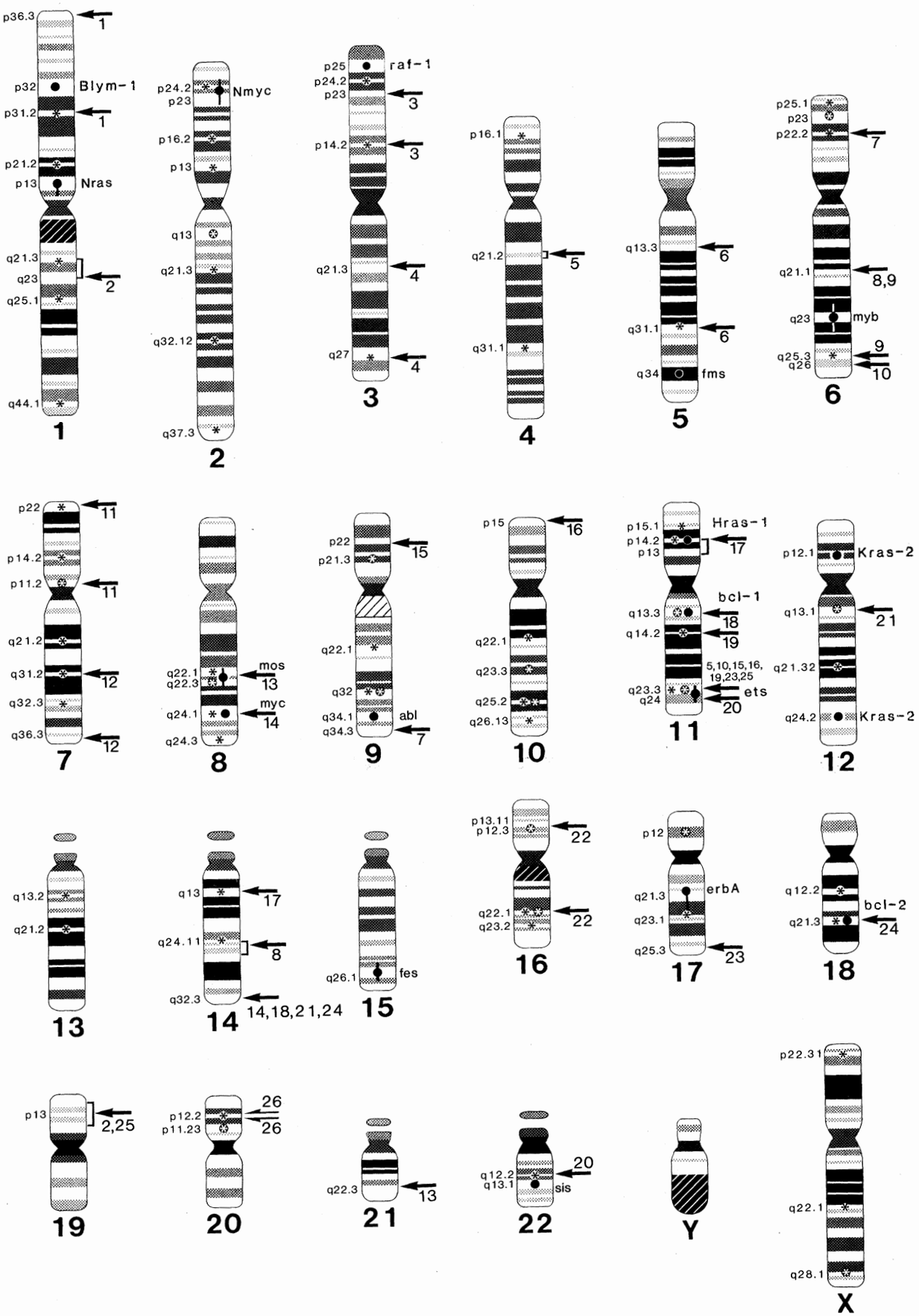


Fig. 1. (a) G-banded metaphase of a human cell treated with FdU and caffeine, showing approximately 50 chromosome breaks. Several c-fra expressed homozygously and c-fra that coincide with specific cancer chromosome breakpoints are indicated by arrows. (b) Micronuclei in human cells. (c) Triradii of human chromosomes. All figures were obtained from human 1.


p36.3
1
p32
Blym-1
p31.2
1
p21.2
p13
Nras
q21.3
q23
2
q25.1
q44.1
1
p24.2
p23
Nmyc
p16.2
p13
q13
q21.3
q32.12
q37.3
2
p25
raf-1
p24.2
p23
3
p14.2
3
q21.3
4
q27
4
3
p16.1
q21.2
5
q31.1
4
q13.3
6
q31.1
6
q34
fms
5
p25.1
p23
p22.2
7
q21.1
8,9
q23
myb
q25.3
q26
9
10
6
p22
11
p14.2
p11.2
11
q21.2
q31.2
12
q32.3
q36.3
12
7
q22.1
q22.3
mos
13
q24.1
myc
14
q24.3
8
p22
p21.3
15
q22.1
q32
q34.1
q34.3
abl
7
9
p15
16
q22.1
q23.3
q25.2
q26.13
10
p15.1
p14.2
p13
Hras-1
17
bcl-1
q13.3
18
q14.2
19
5,10,15,16,
19,23,25
q23.3
q24
ets
20
11
p12.1
Kras-2
q13.1
21
q21.32
q24.2
Kras-2
12
q13.2
q21.2
13
q13
17
q24.11
8
q32.3
14,18,21,24
14
q26.1
fes
15
p13.11
p12.3
22
q22.1
q23.2
22
16
p12
q21.3
erbA
q23.1
q25.3
23
17
q12.2
q21.3
bcl-2
24
18
p13
2,25
19
p12.2
p11.23
26
26
20
q22.3
13
21
q12.2
q13.1
20
sis
22
Y
p22.31
q22.1
q28.1
X

thymidine (5 mg/liter) was in the culture medium in the presence or absence of caffeine (human 1F in Table 1) (*7*). A similar suppression was observed with the FTD-caffeine test when humans were retested after receiving 5 mg of folinic acid (active formyl derivative of folic acid) a day orally for 3 days (human 1 in Table 1) (*7*).

Cultures exposed to caffeine and either FdU or FTD medium had a high percentage of micronuclei (1.1 to 5.3 percent in normal humans 1, 4, 5, 6, and 7; 5.1 percent in the chimpanzee; 2 percent in the gorilla). There were wide variations in the sizes of the micronuclei, which is in accord with their origin from diverse chromosome fragments (Fig. 1b). Triradial configurations, typically observed at h-fra sites (*2*) were also observed in humans at c-fra 3p14.2, 6q25.3, 8q24.1, 16q22.1, 16q23.2, and Xq22.1 (Fig. 1c) and in the chimpanzee at c-fra 3p14.2 and 16q23.2.

Four constitutive fragile sites map at the same band or subband as four heritable fragile sites (bands 9q32, 10q25.2, 11q23.3, and 16q22.1) (Fig. 2). These sites were expressed at very low frequencies when FdU alone was used (0 to 0.6 percent) and at a higher level when cells were treated with FdU and caffeine (2 to 7 percent) (Table 1). These observations suggest that an h-fra, which is expressed with a frequency of less than 0.2 percent in the general population and tends to be expressed heterozygously (more than 20 percent in FdU-treated cells) (*2, 4*), may represent a mutation of a c-fra. In addition, 7 of the 17 oncogenes that have been mapped to a specific chromosome band or region (*8, 9*) are located at or near c-fra bands (Fig. 2).

The detailed mechanism by which folic acid deficiency or FdU treatment induces the expression of fragile sites is not yet known. Their expression is related to partial inhibition of thymidylate synthetase (*2, 4*), resulting in thymidine deprivation and misincorporation of uracil in place of thymine (*10*). Heritable fragile sites may occur at DNA sequences whose methyl groups bind protein and are involved in chromosomal folding (*11*). If h-fra or c-fra remain demethylated after misincorporation of uracil, or if uracil is excised prior to mitosis (*11*), the chromosome structure may collapse at a specific point, yielding gaps and breaks.

The majority of fragile sites appear to be located either at the junction of Giemsa-negative and Giemsa-positive bands or in Giemsa-negative bands close to the junction (Fig. 2). We have previously found that the bulk of the structural genes in humans and primates are localized in Giemsa-negative bands and that Giemsa-positive bands are enriched in middle-repetitive AT (A, adenine; T, thymine)-rich DNA (*12*). It is possible that most fragile sites represent an evolutionarily conserved class of T-rich sequences that flanks protooncogenes and is particularly sensitive to thymidine deprivation. The mechanism by which caffeine enhances the expression of chromosomal lesions is probably related to its capacity to inhibit DNA repair in replicating cells (*13*).

Strikingly, 20 of the 51 c-fra and 6 of the 16 h-fra map at or close to breakpoints found in 26 of 31 specific structural chromosome defects known so far in leukemias, lymphomas, and malignant solid tumors (Fig. 2) (*8, 14*). Of the c-fra that may be involved, 17 are already known to be located at the same height within a band or subband as the breakpoints of specific chromosomal rearrangements in 22 malignancies. Even if a minimum of 17 coinciding c-fra breaks and only 335 (Giemsa-negative) of 670 total chromosome bands per haploid set are preferentially involved in breakage, the association is found to be highly significant (Yates correction for χ^2, $P < 0.0001$) (*7*).

In an initial study of cells (treated with FdU and caffeine) that had been isolated from patients with specific types of malignancies and chromosome defects we found: (i) an elevated expression (14 and 14.5 percent) of c-fra 7q31.2 in the normal blood cells of two patients with acute nonlymphocytic leukemia (ANLL), a history of heavy exposure to petroleum products, and deletion 7q31.2q36.3; (ii) an elevated c-fra 16q22.1 expression of 21 percent in the normal blood cells of a patient with ANLL and inv(16)(p13.11q22.1) (human 3 in Table 1); and (iii) an elevated c-fra 18q21.3 expression of 13 percent in the normal cells of a patient with follicular lymphoma and t(14;18)(q32.3;q21.3). If confirmed, these findings would suggest that some individuals may have a high expression of specific c-fra that predisposes them to certain types of malignancies.

At present, it is not clear if there is a general correlation between c-fra and retroviral protooncogenes that have been mapped to specific chromosome bands. Nevertheless, as there is a good association between c-fra and chromosomal rearrangements in cancer, it is possible that there may exist a larger class of nonretroviral protooncogenes that can be characterized through molecular cloning of chromosomal breakpoints, which we have accomplished for *bcl*-1 and *bcl*-2 (*15*). Our findings suggest that there are specific sites in the human and primate genome that are prone to chromosomal breakage and rearrangement under conditions of cellular thymidine deprivation. These sites are responsive to folinic acid intake, and to caffeine in vitro. If fragile sites prove to be closely linked but not functionally related to oncogene sites, their remarkable proximity would enable them to be used as valuable markers in the search for and analysis of protooncogenes. In addition, fragile sites may be used to locate DNA sequences that are involved in the single (*8*) or multiple (*16*) rearrangements of chromosomes that occur in neoplasia. The presence of a large number of homologous weakpoints in the genome may also provide the physical basis for somatic recombination, which appears to be crucial in the origin of some solid tumors (*17*).

Fig. 2. Human chromosomal map showing 670 Giemsa bands, 51 constitutive fragile sites (black asterisks), 16 heritable fragile sites (white asterisks), 17 protooncogenes (black circles), and two breakpoints for each specific chromosomal rearrangement found in 26 neoplasias (arrows). Pertinent bands are marked on the left of each chromosome and protooncogenes (*18*) are marked on the right. When a protooncogene is mapped to a small chromosomal region rather than a band, a black circle crossed by a bar is used. When there is uncertainty as to a precise cancer chromosome breakpoint, an arrow is preceded by a bracket. The rearrangements are numbered as follows: 1 = del(1)(p31.2p36.3) in disseminated neuroblastoma; 2 = t(1;19) (q21-23;p13) in pre-B-cell–type acute lymphocytic leukemia (pre-B-ALL); 3 = del(3) (p14.2p23) in lung cancer; 4 = inv(3) (q21q27) in acute nonlymphocytic leukemia (ANLL); 5 = t(4;11) (q21; q23.3) in ALL; 6 = del(5) (q13.3q31.1) in ANLL; 7 = t(6;9) (p22.2;q34.3) in ANLL; 8 = t(6;14) (q21;q24) in ovarian cancer; 9 = del(6) (q21q25.3) in non-Hodgkin's lymphoma (NHL); 10 = t(6;11) (q26;q23.3) in ANLL; 11 = del(7) (p11.2p22) in ANLL (*19*); 12 = del(7) (q31.2q36.3) in ANLL; 13 = t(8;21) (q22.1; q22.3) in ANLL; 14 = t(8;14) (q24.1; q32.3) in Burkitt's and related types of lymphoma and leukemia; 15 = t(9;11) (p22;q23.3) in ANLL; 16 = t(10;11) (p15;q23.3) in ANLL; 17 = t(11;14) (p13-14;q13) in T-ALL; 18 = t(11;14) (q13.3;q32.3) in NHL; 19 = del(11) (q14.2q23.3) in NHL; 20 = t(11;22) (q24;q12) in Ewing sarcoma and neuroepithelioma; 21 = t(12;14) (q13.1;q32.3) in T-cell lymphoma; 22 = inv(16) (p13.11q22.1) in ANLL; 23 = t(11;17) (q23.3;q25.3) in ANLL; 24 = t(14;18) (q32.3; q21.3) in follicular lymphomas; 25 = t(11;19) (q23.3;p13) in ANLL; 26 = del(20) (p12.2) in multiple endocrine cell carcinoma.

JORGE J. YUNIS*
A. LEE SORENG
Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis 55455

References and Notes

1. B. R. Reeves and S. D. Lawler, *Humangenetik* **8**, 295 (1970); A. Brøgger, *ibid.* **13**, 1 (1971); P. Aula and H. van Koskull, *Hum. Genet.* **32**, 143 (1976); M. G. Mattei, S. Ayme, J. F. Mattei, Y. Aurran, F. Giraud, *Cytogenet. Cell Genet.* **23**, 95 (1979); "Human Gene Mapping 7," in *ibid.* **37**, 1 (1984).
2. G. R. Sutherland, *Int. Rev. Cytol.* **81**, 107 (1983); G. R. Sutherland, P. B. Jacky, E. Baker, A. Manuel, *Am. J. Hum. Genet.* **35**, 432 (1983); P. B. Jacky, B. Beek, G. R. Sutherland, *Science* **220**, 69 (1983).
3. J. J. Yunis, *Hum. Pathol.* **12**, 549 (1981); *ibid.*, p. 540.
4. T. W. Glover, *Am. J. Hum. Genet.* **33**, 234 (1981).
5. J. J. Yunis and O. Prakash, *Science* **215**, 1525 (1982).
6. Human 1 was tested three times with FdU and FdU plus caffeine within a period of 6 months and showed similar results. Also, although data is only shown for human 1 in Table 1, FdU treatment gave superior results to those obtained with FTD media in five humans tested.
7. A c-fra was defined as a chromosome breakpoint that occurred at least four times per 100 cells in at least seven out of the ten humans tested. For these analyses, 200 cells per individual were examined. The probability of breakage occurring with this frequency by chance is $P < 0.001$, even when only Giemsa-negative bands, which appear to be preferentially but not exclusively involved in breakage, are considered (L. Sachs, *Applied Statistics*, Springer Verlag, New York, 1982). Most breaks at Giemsa-negative bands occurred at a background frequency of 0.5 to 1 percent. A complete suppression of c-fra was observed in humans 1 and 4 in FTD cultured cells after individuals had received 5 mg of folinic acid a day for 3 days and when thymidine was added to FdU cultures in the presence or absence of caffeine. Thymidine suppression of c-fra was also observed in the two primates.
8. J. J. Yunis, *Science* **221**, 227 (1983); *Prog. Med. Virol.*, in press.
9. M. Schwab *et al.*, *Nature (London)* **308**, 288 (1984); C. C. Morton *et al.*, *Science* **223**, 173 (1984); T. Bonner *et al.*, *ibid.*, p. 71; J. Groffen *et al.*, *Nucleic Acids Res.* **11**, 6331 (1983); M. Rabin *et al.*, *Cytogenet. Cell Genet.* **38**, 70 (1984); S. C. Jhanwar, R. S. K. Chaganti, C. M. Croce, *Somatic Cell. Mol. Gen.*, in press; C. de Taisne, A. Gegonne, D. Stehelin, A. Bernheim, R. Berger, *Nature (London)* **310**, 581 (1984).
10. M. Goulian, B. Bleile, B. Y. Tsent, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1956 (1980).
11. C. L. Krumdieck and P. N. Howard-Peeble, *Am. J. Med. Genet.* **16**, 23 (1983).
12. O. Sanchez and J. J. Yunis, *Chromosoma* **48**, 191 (1974); J. J. Yunis *et al.*, *ibid.* **61**, 335 (1977); J. J. Yunis and M. E. Chandler, *Cytogenet. Cell Genet.* **25**, 220 (1979).
13. C. C. Lau and A. B. Pardee, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2942 (1982); S. K. Das, C. C. Lau, A. B. Pardee, *Mutat. Res.* **131**, 71 (1984).
14. J. J. Yunis, *Cancer Genet. Cytogenet.* **11**, 125 (1984); *ibid.* **12**, 85 (1984); *ibid.* **13**, 17 (1984).
15. Y. Tsujimoto, J. Yunis, L. Onorato-Showe, J. Erikson, P. C. Nowell, C. M. Croce, *Science* **224**, 1403 (1984); Y. Tsujimoto, L. R. Finger, J. Yunis, P. C. Nowell, C. M. Croce, *ibid.* **226**, 1097 (1984).
16. A. A. Sandberg and N. Wake, in *Genes, Chromosomes, and Neoplasia*, F. E. Arrighi, P. N. Rao, E. Stubblefield, Eds. (Raven, New York, 1981), p. 297.
17. W. K. Cavenee *et al.*, *Nature (London)* **305**, 779 (1983); A. Koufos *et al.*, *ibid.* **309**, 170 (1984); S. L. Naylor, J. Minna, B. Johnson, A. Y. Sakaguchi, *Am. J. Hum. Genet.* **36**, 35S (1984).
18. Abbreviations denote the following protooncogenes: *Blym*, B-cell lymphoma; *Nmyc*, neuroblastoma *myc*; *raf1*, 3611 murine sarcoma; *fms*, McDonough feline sarcoma; *myb*, myeloblastosis; *mos*, Moloney sarcoma; *myc*, myelocytoma; *abl*, Abelson leukemia; *Hras*, Harvey sarcoma; *bcl*, B-cell lymphoma or leukemia; *ets*, E26 erythroleukemia; *Kras*, Kirsten sarcoma; *fes*, Snyder-Theilin feline sarcoma; *erb*, erythroblastosis; and *sis*, simian sarcoma.
19. J. J. Yunis, unpublished.

20. We thank D. Longrie-Kline for technical assistance, D. Aeppli for statistical analysis, H. McClure of the Yerkes Primate Center for chimpanzee and gorilla blood samples, and W. Hoffman for editorial assistance.

* To whom reprint requests should be addressed at Box 198, Mayo Memorial Building, 420 Delaware St., S.E., Minneapolis, Minn. 55455.

3 August 1984; accepted 28 September 1984.

The Long Terminal Repeat Sequences of a Novel Human Endogenous Retrovirus

Abstract. *The complete nucleotide sequence of both the 5′ and 3′ long terminal repeats (LTR's) has been determined for a human endogenous retroviral genome. These sequences are 593 and 590 nucleotides long and have diverged from one another by 8.8 percent. The LTR's resemble those of functional mammalian type C retroviruses in length and in the presence and location of eukaryotic promoter sequences. The 5′ LTR is followed by a presumptive primer binding site unlike that of any known mammalian type C retrovirus, exhibiting 17 out of 18 nucleotides complementary to arginine transfer RNA rather than proline transfer RNA.*

During replication of a retrovirus, the viral RNA is reverse-transcribed into DNA and integrated into the host genome. Because sequences specific to the 5′ and 3′ ends of the viral RNA (U5 and U3) are duplicated during this process, the integrated provirus is flanked by long terminal repeats (LTR's). The LTR's contain all the sequences necessary for transcription of the viral genome (*1, 2*). In addition to providing promoter functions for viral genes, the presence of the LTR sequences at both the 5′ and 3′ ends of the integrated provirus may lead to activation of host genes adjacent to the viral integration site (*3*).

We have isolated an endogenous retroviral sequence, ERV3, from a human recombinant DNA library by low-stringency hybridization to probes from two regions of the type C baboon endogenous virus (BaEV) genome (*4*). The ERV3 sequence appears to contain a full-length, integrated retroviral genome as revealed by DNA hybridization and sequencing studies. The sequence analysis of the LTR's enabled us to determine whether necessary signals for promotion of viral or cellular genes or both are present in these elements and to address the relationship of this retroviral sequence to other mammalian retroviruses.

Two Eco RI restriction enzyme fragments from the clone containing ERV3 hybridized to the BaEV LTR (Fig. 1) and were therefore subcloned into the plasmid vector pBR322 for sequence analysis (*5, 6*). A comparison of the sequences from the two LTR-hybridizing regions revealed a span of 593 nucleotides that is 91.2 percent homologous to a second 590-nucleotide sequence (Fig. 2). Two features common to proviruses further suggested that the ERV3 clone contains intact, full-length LTR elements: the presence of TG. . .CA termini (T, thymine; G, guanine; C, cytosine, A, adenine) surrounding each element; and the presence of duplicated host sequences at the junction of virus and host, a result of retroviral integration.

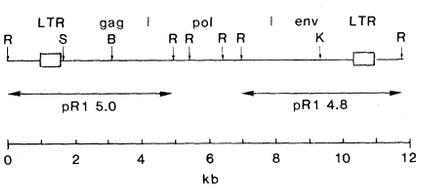


Fig. 1. A restriction map of the human endogenous retroviral locus ERV3. B is Bgl II, K is Kpn I, R is Eco RI, and S is Sma I. Boxed regions contain the sequenced LTR's. Two Eco RI subclones, pR1 5.0 and pR1 4.8, hybridized to the BaEV LTR; kb, kilobases.

Sequencing studies of many retroviral LTR's have shown that these elements end with inverted, complementary repeats of 2 to 16 nucleotides, characteristically beginning with the dinucleotide TG and ending with its inverted complement CA (*1*). The regions of homology between the two ERV3 sequences were bounded by TG. . .CA inverted, complementary repeats (Fig. 2). Further, the duplication of host sequences at the target site of retroviral integration for the ERV3 provirus was the flanking four-nucleotide direct repeat TATA (Fig. 2).

In addition to these two features of LTR boundaries, the viral sequences found adjacent to the ERV3 LTR's resemble recognized retroviral features. The tRNA's (transfer RNA's) used as primers in viral replication anneal to a nucleotide sequence within the viral RNA immediately adjacent to U5. This region, the primer binding site (PBS), is complementary to the 16 to 19 nucleotides at the 3′ terminus of a specific tRNA. The ERV3 proviral sequence in this region (adjacent to the U5 region of the 5′ LTR) was compared to all known tRNA sequences (*7*). It proved to be most closely related to a mouse arginine tRNA ($tRNA^{Arg}$), sharing 17 out of 18 complementary nucleotides (Fig. 3). Although the human equivalent of this $tRNA^{Arg}$ gene has not been sequenced, it may well be identical because tRNA's have been highly conserved in evolution. In contrast to this match of 17 out of 18 nucleotides with a $tRNA^{Arg}$, the putative PBS shared only 10 out of 18 nucleotides complementary to $tRNA^{Pro}$, the tRNA