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.

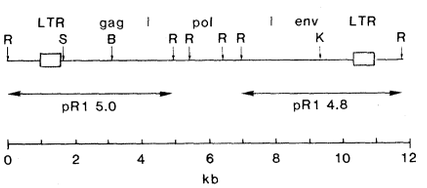

Fig. 1. A restriction map of the human endogenous retroviral locus ERV3. B is Bgl II, K is Kpn I, R is Eco RI, and S is Sma I. Boxed regions contain the sequenced LTR's. Two Eco RI subclones, pR1 5.0 and pR1 4.8, hybridized to the BaEV LTR; kb, kilobases.

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

used by all known mammalian type C retroviruses. The $tRNA^{Trp}$ used by the avian type C retroviruses and the $tRNA^{Lys}$ (*8*) used by the type B mouse mammary tumor virus were similarly divergent from the ERV3 PBS. Adjacent to the 3′ LTR, retroviruses contain the putative primer binding site for second strand complementary DNA synthesis, PB(+). This site consists of a short stretch of purine nucleotides (*8*). An 11-nucleotide stretch of purines precedes the ERV3 3′ LTR (Fig. 2).

A comparison of the 5′ and 3′ noncoding regions of eukaryotic genes revealed conservation of several oligonucleotide sequences, including a CCAAT box beginning 70 nucleotides upstream from the messenger RNA (mRNA) start site (−70), a TATA box from −25 to −30, and an AATAAA sequence at the 3′ end of the gene approximately 25 bases upstream from the polyadenylation site. These sequences are also found in viral genomes and have been implicated in the transcription of both viral and eukaryotic genes (*1*). In mammalian type C proviruses, these sequences are found within the LTR's in similar positions relative to the initiation site for eukaryotic mRNA synthesis. In addition, these transcriptional regulatory sequences are found in specific locations within the LTR, thus defining three regions. The U3 region of mammalian type C retroviral LTR's ranges from 342 to 480 nucleotides in length and contains two of the transcription signals described, the CCAAT and TATA sequences. The beginning of the R region (a sequence repeated at both 5′ and 3′ ends of the viral RNA) corresponds to the initiation site of mRNA synthesis in the provirus. In the mammalian type C provirus, it is located 23 nucleotides downstream from the TATA sequence and begins with the dinucleotide GC (*1*). Consistent with this, the human ERV3 LTR's contained a GC dinucleotide 23 base pairs downstream of the TATA sequence beginning at nucleotide 466. In addition, a CCAAA sequence, similar to CCAAT, began at nucleotide 433, 63 base pairs upstream of this putative mRNA start site. This defined a U3 region for ERV3 of 496 nucleotides, an appropriate length for a mammalian type C retrovirus. The R region is usually 60 to 70 nucleotides long and contains the polyadenylation signal AATAAA as well as a pyrimidine-rich region ending with CA, the site of polyadenylation in viral RNA. The ERV3 LTR's contained an R region of 63 nucleotides from position 496 to 558 with a polyadenylation signal beginning at nucleotide 539. The third LTR region, U5, has no demonstrated promoter sequences and is found to vary in length from 67 to 176 nucleotides in other mammalian retroviruses. This region in ERV3 was only 35 nucleotides long and ended in the dinucleotide CA of the inverted complementary repeat. Unlike the LTR's of ERV3 and other mammalian type C retroviruses, those of human T-cell leukemia virus (HTLV) and bovine leukemia virus (BLV) differ in the arrangement of these important signals (*9–11*).

Characterization of the SV40 (simian virus 40) promoter region led to the identification of 72–base pair direct repeats containing "enhancers," a single copy of which is necessary for high levels of viral expression. Similar long-direct repeats have been identified in retroviral LTR's and in some cases have been shown to possess enhancer activity. The

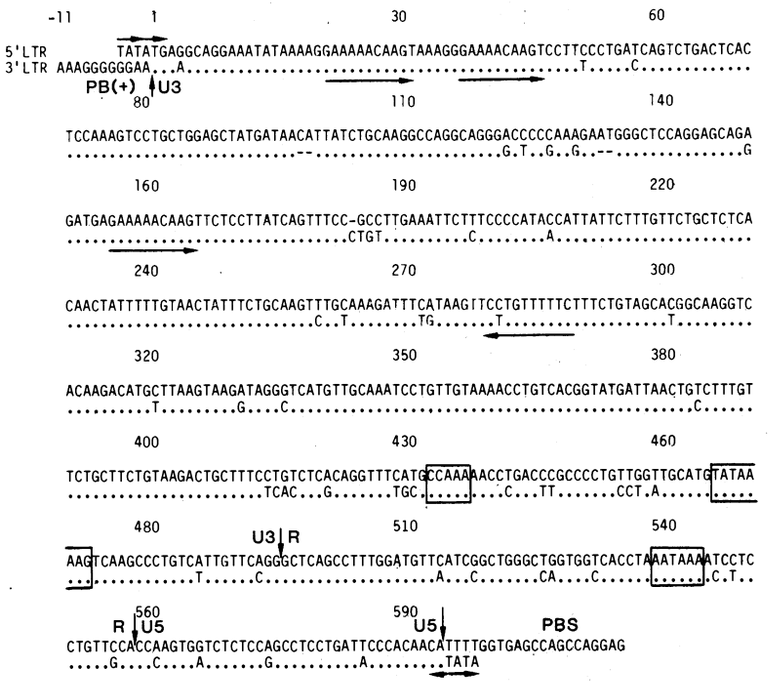

Fig. 2. The nucleotide sequence of the 5′ and 3′ LTR's and immediate flanking sequences. Dots represent identities between the two sequences. Vertical arrows separate LTR regions. Horizontal arrows indicate repeated sequences: TATA, duplication of human sequences upon provirus integration; TG-CA, inverted termini of LTR's; and -GAAAAACAAGT-, repeated sequences within U3. Boxed regions indicate transcriptional regulatory sequences. PBS, primer binding site for (−) strand synthesis. PB(+), primer binding site for second-strand synthesis.

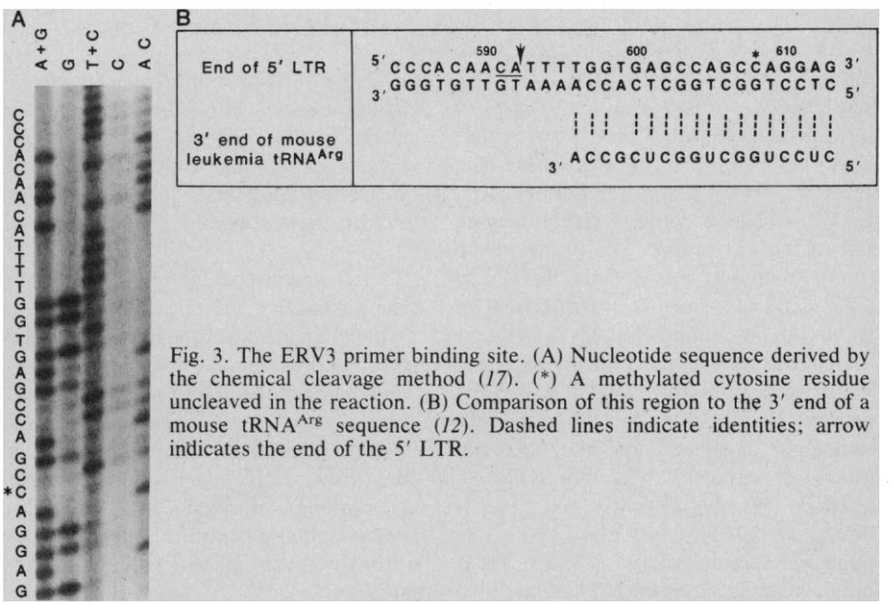

Fig. 3. The ERV3 primer binding site. (A) Nucleotide sequence derived by the chemical cleavage method (*17*). (*) A methylated cytosine residue uncleaved in the reaction. (B) Comparison of this region to the 3′ end of a mouse $tRNA^{Arg}$ sequence (*12*). Dashed lines indicate identities; arrow indicates the end of the 5′ LTR.

ERV3 LTR's did not appear to contain a long-direct repeat. However, the LTR's of two HTLV variants (*9, 10*) and an isolate of the related BLV (*11*) also lack these repeats and are certainly expressed. Four regions within the ERV3 LTR's displayed partial homology with the consensus sequence GTGG$^{AAA}_{TTT}$G that is present in many viral enhancers and is essential for the activity of the SV40 enhancers (*2*). These regions, beginning at LTR nucleotides 455, 505, and 563 in U5 and at nucleotide 597 within the putative primer binding site, each shared six out of eight nucleotide identities with the consensus sequence. The significance of these sequences is unclear, since retroviral enhancers characterized thus far are located in the U3 region of the LTR.

The LTR sequences of HTLV variants I and II are unrelated to each other except for the promoter sequences and a 21–base pair sequence repeated three times in the U3 region of each genome at comparable positions. These repeats are imputed to be equivalent to the enhancer sequences of other retroviruses (*9*). Similarly, the ERV3 LTR's contained a short repeated sequence (underlined in Fig. 2). This sequence, GAAAAACAAG, is also found in the U3 region of both BaEV (*12*) and HTLV-II (*9*) as a single copy. Further experiments are necessary to define whether these or other ERV3 LTR sequences function in enhancement of proviral expression.

Analysis of the nucleotide sequence of the ERV3 LTR's revealed a region of close nucleotide homology with the BaEV LTR that was responsible for their hybridization. Within a stretch of 41 nucleotides beginning at nucleotide 373 in the ERV3 3′ LTR there was only one nucleotide change with respect to the BaEV LTR (*12*) (two nucleotides were different in the ERV3 5′ LTR). This sequence was found in the U3 region of both ERV3 and BaEV LTR's but farther upstream in BaEV at U3 nucleotide 102. A human provirus characterized earlier, ERV1 (*13*), also contains this sequence but in the U5 region. We observed no other homology between the LTR's of ERV3 and those of other sequenced retroviruses, including HTLV (*9, 10*).

Because of the mechanism of retroviral replication, the two ERV3 LTR's were presumably identical when the provirus first integrated into the ancestral human genome (*1*). The two LTR sequences are not presently 100 percent homologous (Fig. 2). The 5′ LTR is 593 nucleotides long, whereas the 3′ LTR is only 590 nucleotides long. There are also 52 nucleotide substitutions spaced throughout the two LTR's, corresponding to an 8.8 percent divergence.

The possible use of $tRNA^{Arg}$ as a primer for replication of ERV3 suggests a separate lineage for this provirus. Because the ERV1 provirus is missing the 5′ LTR, the primer binding site cannot be identified. While this precludes classification of ERV3 and ERV1 in the same retroviral lineage by this criterion, comparisons of the DNA sequences from regions of the *gag* and *pol* genes indicate that these two proviruses are more closely related to each other than either is to BaEV or M-MuLV (Moloney murine leukemia virus) (*4*). It will be interesting to determine whether other BaEV LTR-hybridizing clones isolated from human DNA (*14*) also share this primer binding site and the 41-nucleotide homology.

In conclusion, the ERV3 LTR's contain sequences homologous to known transcriptional regulatory elements. Without knowing whether the ERV3 LTR's are, or were, used for expression, we were unable to assess the relevance of the nucleotide changes between them. It is unlikely that all the changes would have occurred within only one LTR even if the other was selected for by providing a necessary function, since every change would not be expected to be deleterious. Thus, the divergence of the two ERV3 LTR's by 8.8 percent does not necessarily preclude function by either of them. These sequences may therefore be capable of directing the expression of either ERV3 or nearby host genes. Consequently, the ERV3 provirus is a suitable probe with which to look for expression of endogenous retroviruses in both normal and tumor tissues of humans.

CATHERINE D. O'CONNELL
MAURICE COHEN
Laboratory of Molecular Virology and Carcinogenesis, Litton Bionetics Incorporated–Basic Research Program, National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland 21701

References and Notes

1. H. M. Temin, *Cell* **27**, 1 (1981).
2. Y. Gluzman and T. Shenk, *Enchancers and Eukaryotic Gene Expression* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983).
3. W. S. Hayward, B. G. Neel, S. M. Astrin, *Nature (London)* **290**, 475 (1981).
4. C. O'Connell, S. O'Brien, W. G. Nash, M. Cohen, *Virology* **138**, 225 (1984).
5. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
6. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
7. D. H. Gauss and M. Sprinzl, *Nucl. Acids Res.* **11**, 1 (1983).
8. H. R. Chen and W. C. Barker, *ibid.* **12**, 1767 (1984).
9. K. Shimotohno, D. W. Golde, M. Miwa, T. Sugimura, I. S. Y. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1079 (1984).
10. M. Seiki, S. Hattori, M. Yoshida, *ibid.* **80**, 3618 (1983).
11. D. Couez, J. Deschamps, R. Kettmann, R. M. Stephens, R. V. Gilden, A. Burny, *J. Virol.* **49**, 615 (1984).
12. T. Tamura, M. Noda, T. Takano, *Nucl. Acids Res.* **9**, 6615 (1981).
13. T. I. Bonner, C. O'Connell, M. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4709 (1982).
14. M. Noda, M. Kurihara, T. Takano, *Nucl. Acids Res.* **10**, 2865 (1982).
15. We thank E. Brownell for helpful discussions and J. Clarke for technical assistance. Sponsored by the National Cancer Institute, Department of Health and Human Services, under contract NO1-CO-23909 with Litton Bionetics, Inc.

21 May 1984; accepted 2 August 1984

Nucleotide Sequences of the Human and Mouse Atrial Natriuretic Factor Genes

Abstract. *Mouse and human atrial natriuretic factor (ANF) genes have been cloned and their nucleotide sequences determined. Each ANF gene consists of three coding blocks separated by two intervening sequences. The 5′ flanking sequences and those encoding proANF are highly conserved between the two species, while the intervening sequences and 3′ untranslated regions are not. The conserved sequences 5′ of the gene may play an important role in the regulation of ANF gene expression.*

Atrial natriuretic factor (ANF), a potent vasoactive peptide synthesized in mammalian atria, is thought to play a key role in cardiovascular homeostasis (*1–6*). Analysis of cloned DNA sequences complementary to AFN messenger RNA's (mRNA's) has defined a precursor molecule (preproANF) from which this cardiac hormone is derived (*7–11*). The ANF gene is actively transcribed; preproANF mRNA comprises 1 to 3 percent of atrial mRNA (*7, 12*). To study further the transcription, processing, and regulated expression of this cardiac hormone, we have cloned the genes encoding murine and human ANF and determined their nucleotide sequences.

The single ANF gene in rodents and humans (*7, 10*) hybridizes to a cloned rat ANF complementary DNA (cDNA) probe (*7*). Procedures for cloning single-copy eukaryotic genes from bacteriophage libraries are well established (*13*). Five bacteriophage clones contain-