Detection and cloning of new HTLV-related endogenous sequences in man

Andras Perl[1]*, Joseph D.Rosenblatt[3], Irvin S.Y.Chen[3], John P.DiVincenzo[1], Robert Bever[2], Bernard J.Poiesz[4] and George N.Abraham[1,2]

[1]Department of Medicine and [2]Department of Microbiology and Immunology, University of Rochester School of Medicine, Rochester, NY 14642, [3]Division of Hematology-Oncology, Department of Medicine, UCLA, Los Angeles, CA 90024 and [4]Department of Medicine, SUNY School of Medicine, Syracuse, NY 14202, USA

Received June 30, 1989; Revised and Accepted July 27, 1989

ABSTRACT

Human T-cell leukemia virus (HTLV) type I-related endogenous sequences (HRES) have been cloned from a human genomic library. HRES-1/1 is present in DNA of all normal donors examined. By nucleotide sequence analysis, HRES-1/1 contains two potential open reading frames capable of encoding a p25 and a p15. A 684 bp flanking region 5′ from the first ATG codon of p25 contains a TATA-box, a poly-adenylation signal, a putative tRNA primer binding site, and inverted repeats at locations which are typical of a retroviral long terminal repeat. Phylogenetic analysis suggests that HRES-1/1 entered the genome in primates, presumably as an exogenous retrovirus. From the deduced amino acid sequence of HRES-1/1 p25, residues 6−36 show a sequence homology of 32% and 39% to gag region segments of HTLV-I and HTLV-II, while residues 104−139 display a sequence homology of 33% and 28% to the gag regions of human immunodeficiency virus type 2 (HIV-2) and feline sarcoma virus (FSV), respectively. This suggests that the original exogenous virus infecting primate may be chimeric in structure. The HRES-1/1 genomic locus is transcriptionally active in lymphoid cells, melanoma cells, and embryonic tissues.

INTRODUCTION

Human cells contain a complex variety of endogenous retroviral sequences (ERSs). These human ERSs were isolated by low stringency hybridization to known mammalian ERSs (1−4), by hybridization to the 3′ terminus of tRNAs (5, 6), or during analyses of flanking regions of other genes (7, 8). While some ERSs are represented in a single copy per haploid genome (2, 9), others are highly repetitive and present at a frequency of 50 to 100 copies per haploid genome (1, 3−8). To date, the function of these ERSs is unknown and they have not been implicated in human disease.

In some species, such as the baboon (10) or the mouse (11, 12), ERSs are complete proviruses which can be expressed as infectious retroviral particles. In human cells, infectious endogenous retroviruses have not been detected.

To determine whether endogenous retroviral sequences with similarity to known human retroviruses exist in man, we undertook low stringency cloning with an HTLV-I specific DNA probe and demonstrate HTLV-I-related endogenous sequences (HRES) so far unknown in the human genome. The HRES-1/1 genomic locus is transcriptionally active and contains two large potential ORFs. Phylogenetic analysis suggests that the HRES-1/1 may have entered the genome at the developmental stage of primates, presumably as an exogenous retrovirus.

MATERIALS AND METHODS

Preparation and screening of genomic library

A genomic library was constructed in lambda DASH bacteriophage (Stratagene, San Diego, CA). DNA was extracted from cultured T lymphocytes of a patient, MA, with type II cryoglobulinemia, 0.4 μg partially digested with Sau3A and ligated with T4 DNA ligase to 1 μg of phage arms which had been prepared with BamHI digestion as described by Maniatis et al. (13). Ligated DNA was packaged with Gigapack Gold extracts (Stratagene) and recombinant phages were plated on *E. coli* P2392. Positive clones were identified by hybridization with a ^{32}P-labeled HTLV-I LTR and gag region-containing probe, pMAI. Hybridization was carried out in 6×SSC (1×SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 1×Denhardt's solution (13), 1 mM Na_2-EDTA, and 0.5% SDS. After overnight hybridizations, the filters were washed under low stringency conditions in 2×SSC and 0.1% SDS at 55°C for 2 h. Such screening would identify sequences containing at least 12 identical nucleotides in contiguity, assuming a 50% GC content (14).

Southern blot hybridizations

High molecular weight genomic DNA was isolated from peripheral blood lymphocytes and granulocytes of 33 normal donors and various cell lines, digested to completion with restriction endonucleases (BRL, Gaithersburg, MD), separated by electrophoresis in 0.7% agarose gels, denatured, and transferred to nylon membranes by Southern blotting as described earlier (15). The blots were hybridized as indicated above except for the addition of 100 μg/ml boiled salmon sperm DNA. Genomic blots were washed under high stringency conditions (16) in 0.1×SSC, 0.1% SDS for 90 min at 65°C.

Northern blot analysis

Poly A+ RNA was isolated by direct binding to oligo-dT cellulose in cell lysates (17), fractionated in 1% glyoxal gels, (13) and transferred to nylon membranes. Hybridization and washing were done under high stringency conditions.

Hybridization probes

The HTLV-I specific pMAI probe used in these experiments is a 1.5 kb EcoRI-ClaI fragment which contains the 5′ LTR and the entire gag gene of HTLV-I (SstI-SmaI fragment in Fig. 2) (18). As an HRES-1/1 specific probe, either the entire 5.5 kb HindIII fragment of HRES-1/1 (p1/1) or an EcoRI-SmaI fragment of p1/1 (1/1-ES) was used (Fig. 2). All DNA probes were purified from vector in preparative agarose gels and labeled with ^{32}P-dCTP by random oligomer priming (19). The 5.5 kb HindIII fragment of HRES-1/1 was cloned into the Bluescript SK vector which allowed the generation of strand specific ^{32}P-UTP labeled RNA probes using T3 or T7 RNA polymerase (Stratagene).

A human beta-actin cDNA probe, pFH5 (1.5 kb XhoI fragment) was used as a control for RNA expression (20). The integrity of mammalian genomic DNA samples was assured by hybridization to exon 3 (1.8 kb ClaI-EcoRI fragment) of the human c-myc locus (21).

DNA sequencing

The sequence of 2151 nucleotides of HRES-1/1 was determined by the chain termination method using the Kilobase Sequencing System (BRL). Suitable fragments of HRES-1/1 were subcloned into M13 mp18 and m19 vectors (BRL) to allow sequencing of the insert from both directions. The obtained sequence was analyzed with the University of Wisconsin Genetics Computer Group (UWGCG) software through the University of Rochester VAX computer system.

RESULTS

Restriction mapping of HTLV-related genomic sequences

8×10^5 recombinants of a genomic DNA library prepared from MA-T cells were screened under low stringency conditions (2×SSC, 55°C) with pMAI, a long terminal repeat (LTR) and gag region-containing HTLV-I probe. Such screening would detect sequences containing at least 12 identical nucleotides in contiguity (14), assuming a 50% GC content. From

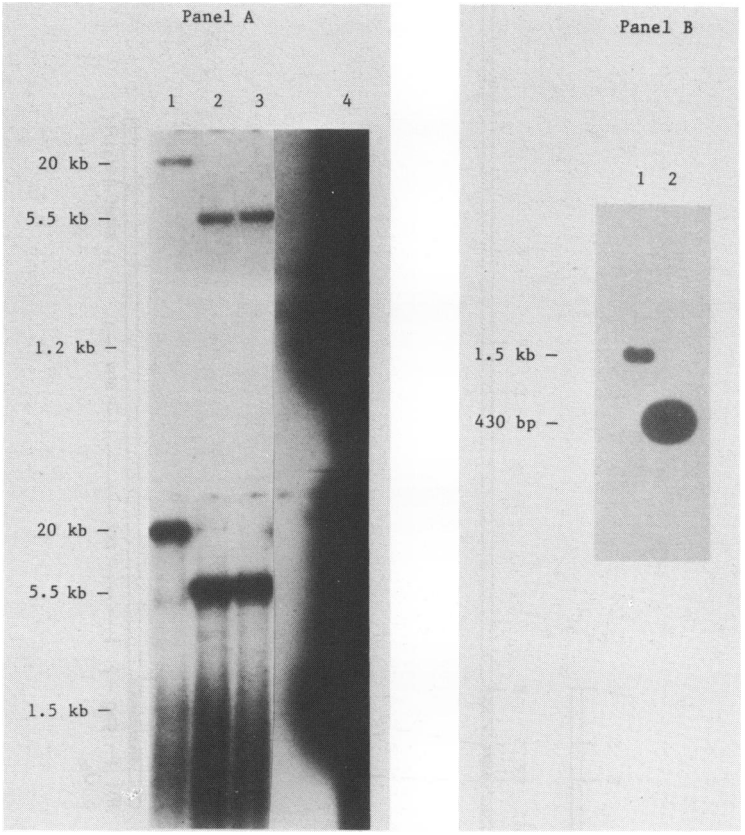
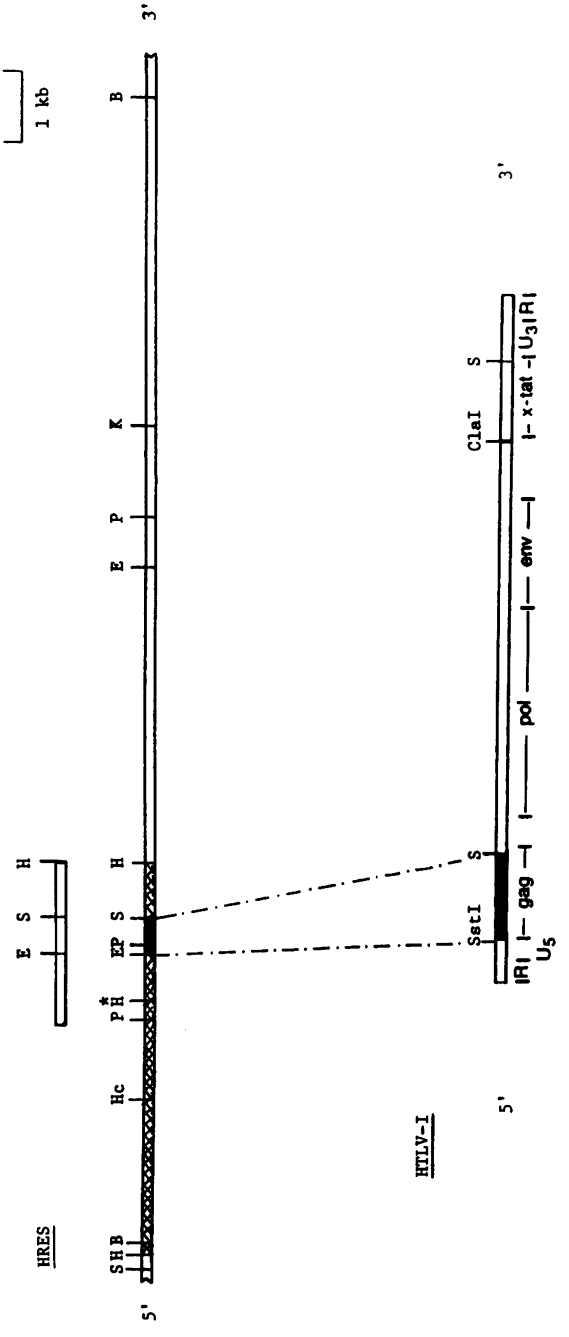


Fig. 1. *Panel A top gel:* the blot was hybridized with an HTLV-I X-tat and LTR-region containing probe (1.2 kb ClaI-HindIII (linker) fragment of pMAI). Samples shown are: lane 1, entire 20 kb genomic HRES insert cleaved from phage clone 1/1 with XbaI restriction endonuclease; lane 2, phage clone 1/1 digested with XbaI and HindIII; lane 3, phage clone 1/1 digested with HindIII; lane 4, 1.2 kb ClaI-HindIII (linker) fragment of HTLV-I containing X-tat and 3′ LTR (positive control); *Panel A bottom gel:* the blot was hybridized to a ^{32}P-labeled HTLV-I LTR and gag region-containing pMAI probe (1.5 kb EcoRI-ClaI fragment). Samples shown in lanes 1–3 are identical with those in the top gel. Lane 4 contained the 1.5 kb EcoRI-ClaI fragment of HTLV-I (positive control). Washing was done under low stringency conditions (2×SSC, 55°C). *Panel B*, Hybridization of 430 bp EcoRI-SmaI fragment of HRES p1/1 (1/1-ES) to the 1.5 kb EcoRI-ClaI fragment of HTLV-I probe, pMAI (lane 1) and to 1/1-ES (lane 2, self-hybridization). Washing was done under high stringency conditions (0.1×SSC at 65°C).


1 kb
3'
B
K
P
E
H
S
EP
PH
Hc
SHB
5'
HRES
E S H
ClaI S
x-tat
U3 R
pol
env
gag
SstI
R U5
5'
HTLV-I

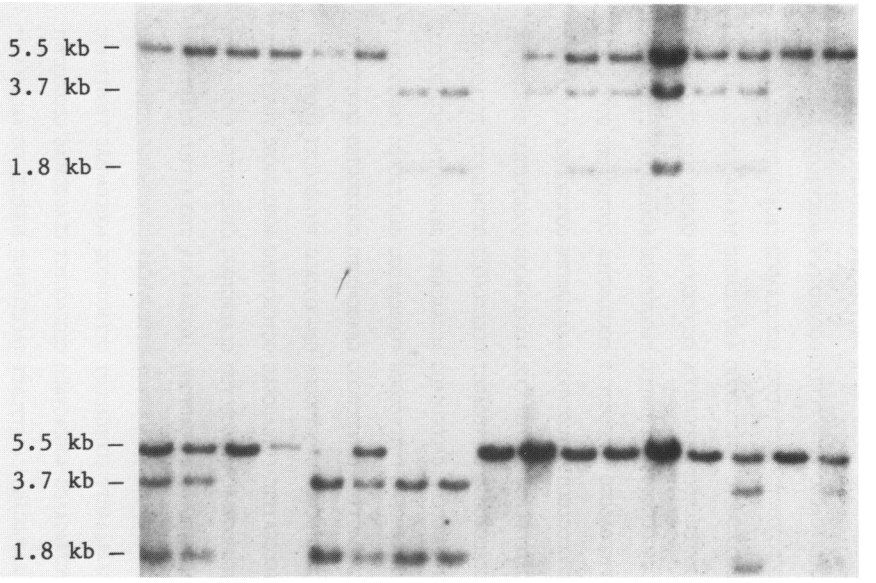


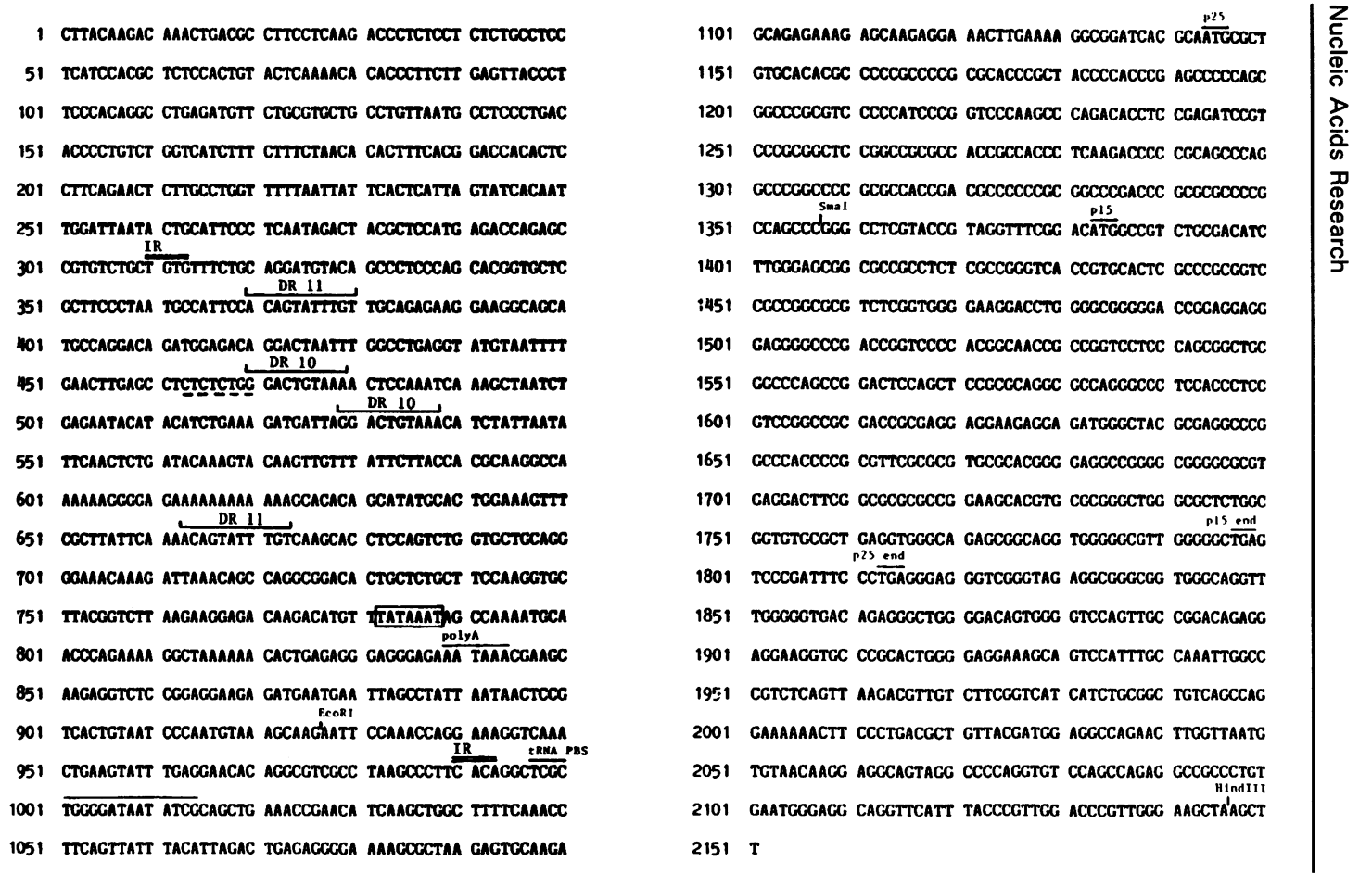
Fig. 3. Southern blot hybridization of the 5.5 kb HindIII fragment of HRES-1/1 (p1/1) to HindIII-digested genomic DNA isolated from 33 different normal human peripheral blood lymphocyte samples. The upper panel contained 16 samples while the lower contained 17 samples. Hybridization and washing was done under high stringency conditions (0.1×SSC, 65°C).

eight positive recombinants isolated from the genomic library, clone 1/1 was further analyzed in detail. The HTLV-I LTR- and gag-related region within bacteriophage clone 1/1 was localized to a 5.5 kb HindIII fragment (Fig. 1, Panel A). The 5.5 kb HindIII fragment (p1/1) was subcloned into pUC18. The HTLV-I LTR and gag-related region within p1/1 was further localized to an EcoRI-SmaI fragment (Fig. 1, Panel B). The precise length (430 bp) of the EcoRI-SmaI fragment of clone 1/1 (1/1-ES) was determined by dideoxy sequencing (see below). As shown in Fig. 1B, intensive cross-hybridization was also noted between the 1/1-ES and HTLV-I fragments under high stringency conditions (washing at 65°C with 0.1×SSC). This suggests a homology of up to 80 bp depending on the GC content of the hybridizing duplex. The location of p1/1 (5.5 kb HindIII fragment) and 1/1-ES (430 bp EcoRI-SmaI fragment) and the restriction map of the genomic locus is shown in Fig. 2.

Southern analysis of human genomic DNA samples

The entire 5.5 kb HindIII fragment of p1/1 showed strong (high stringency) hybridization to a large panel of genomic DNA samples isolated from peripheral blood lymphocytes (Fig. 3). This demonstrated that the 1/1 sequence is present in normal human genomic

Fig. 2. Restriction map of the HRES-1/1 locus. Recognition sites for EcoRI (E), HindIII (H), SmaI (S), BamHI (B), PvuII (P), KpnI (K) and HincII (Hc) restriction enzymes are indicated. The hatched area shows the 5.5 kb HindIII fragment (p1/1) subcloned for further analysis. The solid black areas mark the most related sections of HRES-1/1 and HTLV-I (18) as determined by cross-hybridizations and sequence analyses. 5′ to 3′ orientation is based on sequence analysis of a 2151 bp fragment of HRES-1/1 symbolized as an open bar above the restriction map.


1 CTTACAAGAC AAACTGACGC CTTCCTCAAG ACCCTCTCCT CTCTGCCTCC
51 TCATCCACGC TCTCCACTGT ACTCAAAACA CACCCTTCTT GAGTTACCCT
101 TCCCACAGGC CTGAGATGTT CTGCGTGCTG CCTGTTAATG CCTCCCTGAC
151 ACCCCTGTCT GGTCATCTTT CTTTCTAACA CACTTTCACG GACCACACTC
201 CTTCAGAACT CTTGCCTGGT TTTTAATTAT TCACTCATTA GTATCACAAT
251 TGGATTAATA CTGCATTCCC TCAATAGACT ACGCTCCATG AGACCAGAGC
IR
301 CGTGTCTGCT GTGTTTCTGC AGGATGTACA GCCCTCCCAG CACGGTGCTC
DR 11
351 GCTTCCCTAA TGCCATTCCA CAGTATTTGT TGCAGAGAAG GAAGGCAGCA
401 TGCCAGGACA GATGGAGACA GGACTAATTT GGCCTGAGGT ATGTAATTTT
DR 10
451 GAACTTGAGC CTCTCTCTGG GACTGTAAAA CTCCAAATCA AAGCTAATCT
DR 10
501 GAGAATACAT ACATCTGAAA GATGATTAGG ACTGTAAACA TCTATTAATA
551 TTCAACTCTG ATACAAAGTA CAAGTTGTTT ATTCTTACCA CGCAAGGCCA
601 AAAAAGGGGA GAAAAAAAAA AAAGCACACA GCATATGCAC TGGAAAGTTT
DR 11
651 CGCTTATTCA AAACAGTATT TGTCAAGCAC CTCCAGTCTG GTGCTGCAGG
701 GGAAACAAAG ATTAAACAGC CAGCCGGACA CTGCTCTGCT TCCAAGGTGC
751 TTACGGTCTT AAGAAGGAGA CAAGACATGT TTATAAATAG CCAAAATGCA
polyA
801 ACCCAGAAAA GGCTAAAAAA CACTGAGACG GAGGGAGAAA TAAACGAAGC
851 AAGAGGTCTC CGGAGGAAGA GATGAATGAA TTAGCCTATT AATAACTCCG
EcoRI
901 TCACTGTAAT CCCAATGTAA AGCAAGAATT CCAAACCAGG AAAGGTCAAA
IR
tRNA PBS
951 CTGAAGTATT TGAGGAACAC AGCCGTCGCC TAAGCCCTTC ACAGGCTCGC
1001 TGCGGATAAT ATCGCAGCTG AAACCGAACA TCAAGCTGGC TTTTCAAACC
1051 TTCAGTTATT TACATTAGAC TGAGAGGGGA AAAGCGCTAA GAGTGCAAGA
p25
1101 GCAGAGAAAG AGCAAGAGGA AACTTGAAAA GGCGGATCAC GCAATGCGCT
1151 GTGCACACGC CCCCGCCCCG CGCACCCGCT ACCCCACCCG AGCCCCCAGC
1201 GGCCCGCGTC CCCCATCCCG GTCCCAAGCC CAGACACCTC CGAGATCCGT
1251 CCCGCCGCTC CGGCCGCGCC ACCGCCACCC TCAAGACCCC CGCAGCCCAG
1301 GCCCGGCCCC GCGCCACCGA CGCCCCCCGC GGCCCGACCC GCGCGCCCCG
SmaI
p15
1351 CCAGCCCGGG CCTCGTACCG TAGGTTTCGG ACATGGCCGT CTGCGACATC
1401 TTGGGAGCGG CGCCGCCTCT CGCCGGGTCA CCGTGCACTC GCCCGCGGTC
1451 CGCCGGCGCG TCTCGGTGGG GAAGGACCTG GGGCGGGGGA CCGGAGGAGG
1501 GAGGGGCCCG ACCGGTCCCC ACGGCAACCG CCGGTCCTCC CAGCGGCTGC
1551 GGCCCAGCCG GACTCCAGCT CCGCGCAGGC GCCAGGGCCC TCCACCCTCC
1601 GTCCGGCCGC GACCGCGAGG AGGAAGAGGA GATGGGCTAC GCGAGGCCCG
1651 GCCCACCCCG CGTTCGCGCG TGCGCACGGG GAGGCCCGGG CGGGGCGCGT
1701 GAGGACTTCG GCGCCGCCCG GAAGCACGTG CGCGGGCTGG GCGCTCTGGC
p15 end
1751 GGTGTGCGCT GAGGTGGGCA GAGCGGCAGG TGGGGGCGTT GGGGGCTGAG
p25 end
1801 TCCCGATTTC CCTGAGGGAG GGTCGGGTAG AGGCGGGCGG TGGGCAGGTT
1851 TGGGGGTGAC AGAGGGCTGG GGACAGTGGG GTCCAGTTGC CGGACAGAGG
1901 AGGAAGGTGC CCGCACTGGG GAGGAAAGCA GTCCATTTGC CAAATTGGCC
1951 CGTCTCAGTT AAGACGTTGT CTTCGGTCAT CATCTGCGGC TGTCAGCCAG
2001 GAAAAAACTT CCCTGACGCT GTTACGATGG AGGCCAGAAC TTGGTTAATG
2051 TGTAACAAGG AGGCAGTAGG CCCCAGGTGT CCAGCCAGAG GCCGCCCTGT
HindIII
2101 GAATGGGAGG CAGGTTCATT TACCCGTTGG ACCCGTTGGG AAGCTAAGCT
2151 T

DNA and represents an HTLV-related *endogenous* sequence (HRES). While uniform hybridization patterns were noted in EcoRI, BamHI, PvuII, KpnI, and BglII digested genomic DNA samples, HindIII-digested DNA revealed a polymorphism of the HRES locus as shown in Fig. 3. The HRES-1/1 probe (5.5 kb) annealed to three polymorphic fragments (5.5 kb, 3.7 kb, and 1.8 kb), and three genotypes were differentiated: I, 5.5 kb fragment only; II, 3.7 kb and 1.8 kb fragments only; and III, all three polymorphic fragments. This hybridization pattern suggests that the presence or absence of a polymorphic HindIII site within the 5.5 kb fragment defines the two different allelic forms of the HRES-1/1 genomic locus. Thus, patterns I and II represent homozygous constellations while pattern III shows the heterozygous genotype. In normal donors, the frequency of genotypes is 17/33 (52%) for pattern I, 5/33 (15%) for pattern II, and 11/33 (33%) for pattern III. While not shown, hybridization of the 430 bp EcoRI-SmaI fragment of p1/1 (1/1-ES) to DNA samples from normal peripheral blood lymphocytes revealed only the 5.5 kb and/or the 1.8 kb HindIII fragments. Thus, the polymorphic HindIII site is located 1.8 kb upstream from the 3′ HindIII site of p1/1 (Fig 2). HindIII-digested genomic DNA from MA-T cells showed a heterozygous hybridization pattern with the p1/1 probe (not presented). Clone 1/1, which does not contain the polymorphic HindIII site, represents the genomic allele from MA-T cells carrying the 5.5-kb HindIII fragment.

Hybridizations with probes specific for single copy genes, e.g. the joining region of the immunoglobulin heavy chain gene (J_H, 22) and the constant region of the immunoglobulin kappa light chain gene (C_k, 23) showed bands of similar intensity as compared to those obtained with probes of HRES clone 1/1 under standard conditions. This suggests a single or very limited copy number of HRES-1/1 in the haploid genome.

Sequence analysis of HRES-1/1

Sequencing of the EcoRI-SmaI fragment of clone 1/1-1 was expected to determine the percent homology between the gag region of HTLV-I and clone 1/1, respectively. The EcoRI-SmaI fragment of clone HRES-1/1 as well as 5′ and 3′ flanking regions were subcloned into M13 mp18 and mp19 vectors which allowed sequencing of the inserts from both directions. Thus, the nucleotide sequence of 2151 bases upstream from the 3′ HindIII site of p1/1 was determined in both DNA strands (Fig. 4). The HindIII, SmaI, and EcoRI sites are indicated to orient the sequence in Fig. 4 with the restriction map of clone 1/1. (In Fig. 2 the open bar above the map of HRES-1/1 indicates the sequenced region of the genomic locus). As shown in Fig. 4, beginning at position 1144 the sequence contains a potential 669 base long open reading frame (ORF) which may encode a 25 kD protein (p25). An alternative reading frame from nucleotide position 1383 may encode a 15 kD protein. A putative tRNA primer binding site (PBS) was noted by sequence comparison of the 5′ flanking region of p25 with 3′ terminal 18-base long segments of all known tRNA sequences. The potential PBS, shown in Fig. 4, is 67% complementary to 18 nucleotides at the 3′ end of the chimpanzee histidine tRNA (24). Since the sequence of most tRNAs is still undefined, the indicated PBS may be more homologous to or identical with

Fig. 4. DNA sequence of 2151 nucleotides from HRES-1/1. Location of the sequenced region is shown as an open bar above the restriction map of HRES-1/1 in Fig. 2. The EcoRI, SmaI, and HindIII sites are indicated to orient the sequence with the restriction map. The TGTG...CACA inverted repeats (IR) indicate the boundaries of a putative LTR. The IRs, TATA sequence (boxed), polyadenylation signal (overlined), and the potential PBS at typical locations (26) are features of an LTR. DR10 and DR11 mark the location of two different 10 and 11 nucleotide long direct repeats. The position of the HIV-1 TAR signal is underlined by broken lines. The two alternative open reading frames are indicated by p25 and p15.

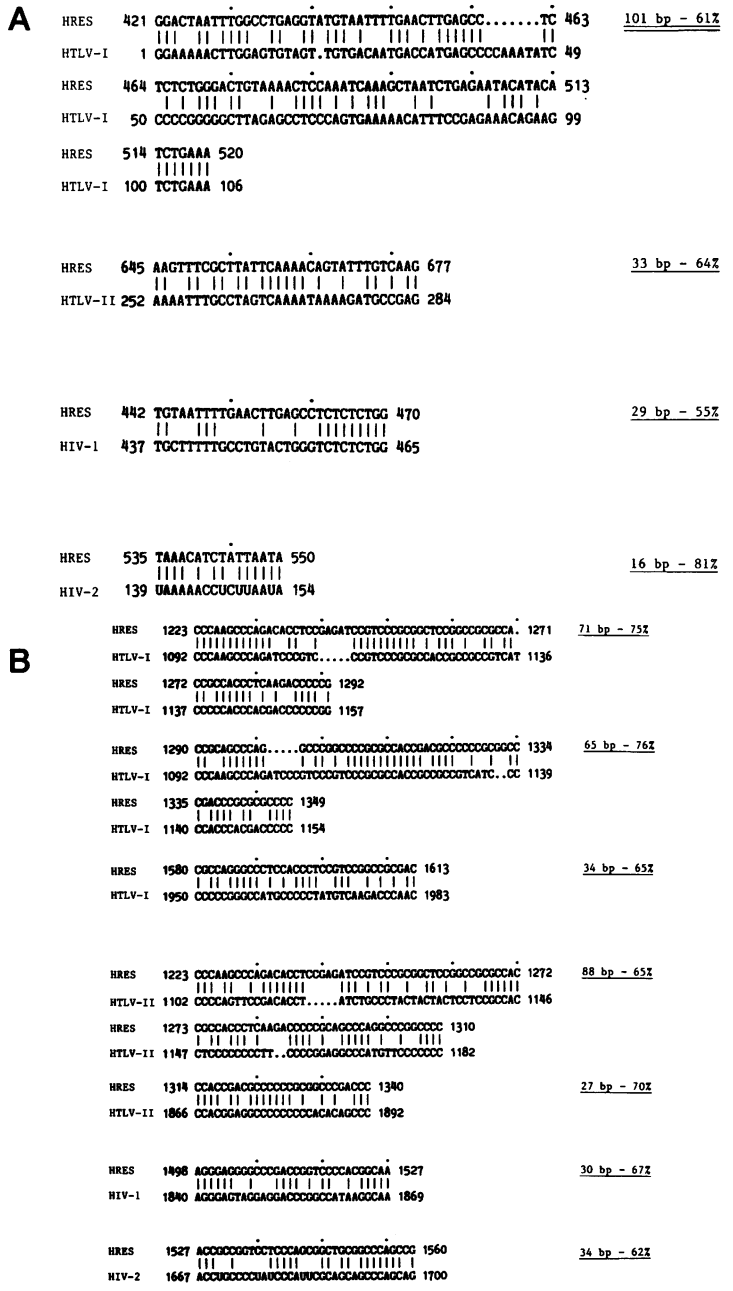


Fig. 5. Nucleotide sequence homologies detected with the Bestfit program of UWGCG. *A*, homology between the LTR-like region of HRES-1/1 (base positions 310−993) and the LTRs of HTLV-I (18), HTLV-II (28), HIV-1 (29), and HIV-2 (30). *B*, homology between the HRES-1/1 p25 ORF and gag regions of HTLV-I (18), HTLV-II (28), HIV-1 (29), and HIV-2 (30).

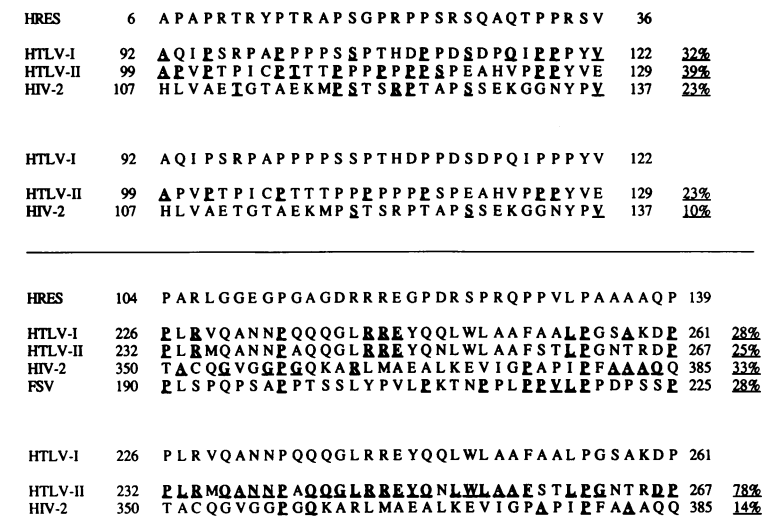


Fig. 6. Amino acid sequence identities detected with the Bestfit program of UWGCG between the HRES-1/1 p25 and gag proteins of HTLV-I (18), HTLV-II (28), HIV-2 (30), and FSV (31). The positions of homology are in bold face and underlined and the percent homology to HRES is indicated. Amino acid positions 6–36 (upper panel) correspond with base positions 1159–1252 while amino acid positions 104–139 (lower panel) correspond with base positions 1453–1560 in the p25 ORF of HRES-1/1. The regions of nucleotide sequence homology between the HRES-1/1 p25 ORF and gag of HTLV-I and HTLV-II (Fig. 5B) partially overlaps with the amino acid sequence alignments involving positions 6–36 of p25. The optimal DNA sequence alignment between HRES-1/1 p25 and HIV-2 gag (Fig. 5B) corresponds with the amino acid sequence homology between residues 129–138 of p25 and residues 375–384 of HIV-2 gag (30).

complementary sequences of the 3′ end of other tRNA(s). The PBS of HTLV-I/II is complementary to the 3′ end of the murine proline tRNA which does not match that of any known human tRNA (25). The LTRs of all integrated retroviruses start with TG and end with CA, which are part of a 3–16 bp inverted repeat (IR) at the ends of the LTR (26). In accordance with the TG...CA rule, the 684 bp LTR-like region of HRES is bounded by TGTG...CACA. The 3′ terminal dinucleotide CA is suitably located 3 base positions 5′ to the PBS. A TATA-box and a polyadenylation signal, which are additional characteristics of an LTR, were detected at appropriate locations 5′ to the PBS (26). Two different direct repeats (DR), 10 bp and 11 bp, were also noted. The LTR-like region of HRES contains a copy of the TAR regulatory sequence, CTCTCTGG, of the HIV-1 LTR (27).

The sequence of HRES-1/1 was compared to all sequences contained in the Genbank, NBRF and EMBL databases. These analyses showed that HRES-1/1 is different from any known human or viral sequence. However, nucleotide sequence comparisons between the LTR-like region of HRES-1/1 and the LTR regions of HTLV-I (18) and HTLV-II (28), as well as between the ORFs of HRES-1/1 and the gag regions of HTLV-I and HTLV-II, respectively, revealed significant levels of homology (Fig. 5A and B). As shown in Fig. 5A, a limited level of homology was found between the HRES LTR-like sequence and the LTRs of HTLV-I, HTLV-II, HIV-1 (29), and HIV-2 (30). The most homologous regions between HRES and HIV-1 LTR involve a complete copy of the TAR sequence.

The regions of highest homology between HRES and HTLV-I involve the p25 ORF

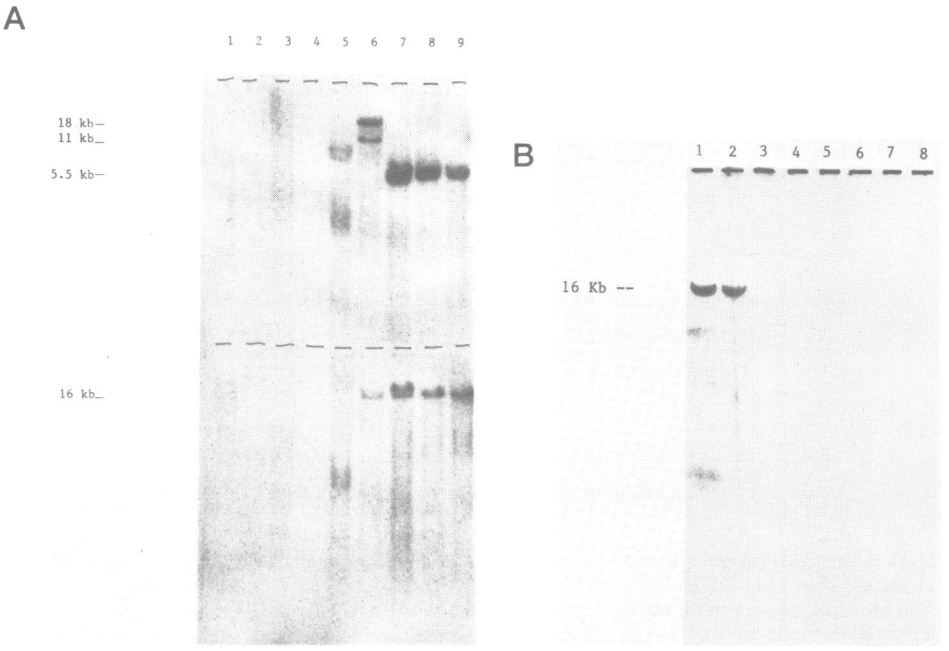


Fig. 7A. EcoRI (top panel) and BamHI (bottom panel) digested genomic DNA samples were hybridized to the 430 bp EcoRI-SmaI fragment of HRES-1/1 (1/1-ES) and washed under high stringency conditions. Lanes shown are: 1, E. coli DH1 DNA; 2, salmon sperm DNA; 3, EL-4 murine thymoma cell line DNA; 4, calf thymus DNA; 5, owl monkey lymphocyte DNA; 6, Rhesus macaque lymphocyte DNA; 7, MA-T; 8, SLB-I, HTLV-I infected human T-cell line; 9, human placenta DNA. Relatively faint owl monkey DNA fragments (6.6-kb and 8.6-kb EcoRI, as well as 3.1-kb and 3.7-kb BamHI) are not indicated in the figure. **Fig. 7B:** BamHI-digested DNA samples were hybridized to the 1/1-ES probe. Lane 1, human placental DNA; lane 2, Rhesus macaque lymphocyte DNA; lane 3, lamb lymphocyte DNA; lane 4, bat lymphocyte DNA; lane 5, canine lymphocyte DNA; lane 6, feline lymphocyte DNA; lane 7, equine lymphocyte DNA; and lane 8, bovine lymphoma DNA.

in HRES and the gag region in HTLV-I. The section of p25 most homologous to HTLV-I gag corresponds to the EcoRI-SmaI fragment of HRES clone 1/1 in accordance with the DNA cross-hybridization studies (Fig. 1). As shown in Fig. 5B, the homologous regions between HRES and HTLV-I involve at least 12 contiguous nucleotides explaining the cross-hybridization under low stringency conditions (2×SSC, 55°C). The facts that there are several areas of relatively high homology between HRES and HTLV-I and the 80−90% GC content in these areas also explain the hybridization under high stringency conditions (0.1×SSC, 65°C) between the HRES 1/1-ES probe and the HTLV-I gag probe.

To determine a possible functional relationship between HRES and the human exogenous retroviruses, the amino acid sequences were also analyzed. In the translated amino acid sequence of HRES p25 two clusters that showed significant homology to the gag proteins of human retroviruses were identified (Fig. 6). From amino acid 6 to 36 (base positions 1159−1252) HRES p25 shows a 32% identity with HTLV-I gag p19 and a 39% identity with HTLV-II gag p19. These areas of amino acid homology partially overlap with the

nucleotide sequence homologies shown in Fig. 5B. Within the same area there is a relatively low homology between HRES p25 and HIV-2, as well as, between HTLV-I and HTLV-II. From amino acid position 104 to 139 (base positions 1453−1560) HRES p25 shows 28% and 25% identity with HTLV-I gag p24 and HTLV-II gag p24, respectively. In the same region there is a relatively high 33% identity between HRES and HIV-2 gag which corresponds with the best nucleotide sequence alignment shown in Fig. 5B. Taking into account functional homology as well, ten contiguous amino acids in HRES p25 from position 129 to 138 may be fully homologous to a region of HIV-2 gag (from position 375 to 384). No appreciable homology was found between p25 and the gag region of HIV-1. A search of existing protein sequence databases revealed a 28% identity between the gag protein of the feline sarcoma virus (FSV, 31) and HRES p25 (Fig. 6). The same computer search demonstrated the significance of the amino acid sequence identities between HRES p25 and retroviral gag proteins. The human sequence N-myc (32), most homologous to HRES p25 showed a 25% amino acid homology within a range of 30 amino acids, while the level of homology with the next most 'related' human sequences (top 30 out of 8,000) was below 10%. The high (78%) amino acid homology between HTLV-I and HLTV-II as well as the low (14%) amino acid homology between HTLV-I and HIV-2 in the gag p24 regions are also shown to demonstrate the significance of amino acid homologies between the HRES p25 and the retroviral proteins. Translated amino acid sequence of p15 of HRES-1/1 showed no homology (over 10% identity involving at least 30 contiguous amino acids) to any known protein sequence.

Phylogenetic study of HRES-1/1

Previously isolated human ERSs have been shown to have an extensive nucleic acid sequence homology with various mammalian ERSs (1−4). Sequence analysis of the 2151 bp fragment of HRES-1/1 demonstrated neither identity nor considerable homology with any human sequence including known ERSs.

Hybridization of the HRES-1/1 to genomic DNA samples of selected phylogenetic stages demonstrates the presence of HRES in primates only (Fig. 7). Although EcoRI digestion produced different restriction fragments, BamHI cleaved an identical 16-kb fragment from both human and Rhesus macaque DNA samples. Hybridization of HRES-1/1 with owl monkey DNA, representing a more distant phylogenetic relation, showed faint hybridization and different restriction fragment sizes both with EcoRI and BamHI. No hybridization of HRES-1/1 was noted with other mammalian or vertebrate DNA samples (Fig. 7A and 7B), even under low stringency conditions. In control experiments the integrity of the DNA samples from these species was assured since the expected patterns were obtained when the blots were probed with the unrelated c-myc oncogene (data not shown).

Transcription of HRES-1/1 sequences

Infectious human endogenous retrovirus particles have never been isolated. While full-length transcripts from essential genes of human retrovirus-like sequences have not been found, selective transcription of the LTR- and env-related regions has been noted in colon carcinoma, breast carcinoma and placenta cells (33). Transcription of the genomic HRES-1/1 locus was studied by Northern blot analysis of poly A+ RNA samples. Single-stranded RNA probes were generated from HRES-1/1 cloned into the Bluescript phagemid (Stratagene, La Jolla, Ca). Due to the orientation of the insert, the T7 polymerase used the 'sense' strand, encoding p25 and p15, as template while the T3 polymerase used the 'antisense' strand as template. As shown in Fig. 8A, the HRES-1/1 specific 6-kb message was detected with the T7 probe but not the T3 probe in MA-T cells. The 6-kb transcript

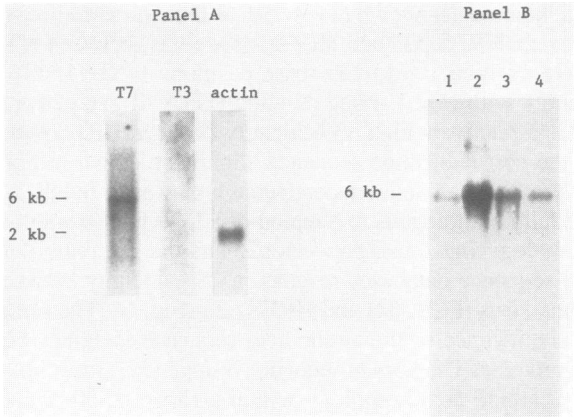


Fig. 8. Northern blot analysis of poly A+ RNA samples. Panel A: RNA from MA-T cells was hybridized to HRES-1/1 RNA probe generated with T7 polymerase (first lane), HRES-1/1 RNA probe generated with T3 polymerase (second lane), and human beta-actin cDNA probe, pHF5 (20) (third lane). Panel B: Hybridization of HRES-1/1 RNA probe generated with T7 polymerase to W7B melanoma cells (lane 1), W7A melanoma cells (lane 2), HL-60 promyelocytic leukemia cells (lane 3), and Molt4 T-cell leukemia cells (lane 4).

was also noted in W7B and W7A melanoma cells, HL-60 promyelocytic leukemia cells, and Molt-4 T-cell leukemia cells (Fig. 8B). HRES-1/1 transcripts were also noted in normal human placenta cells and EBV-transformed normal human peripheral blood B lymphocytes but not in K562 myeloid leukemia and U937 monocyte leukemia cell lines (not shown).

DISCUSSION

This report describes the detection and cloning of a novel human ERS, termed HRES-1/1, from a recombinant lymphocyte DNA library based on cross-hybridization with an HTLV-I LTR and gag region-containing probe. Nucleotide sequence analysis of a 2151 base long fragment of clone 1/1 revealed that the region most homologous to the gag gene of HTLV-I and HTLV-II contains two potential ORFs which may encode a p25 and/or a p15. A 684 bp flanking region 5′ to p25 contains a TATA-box, a polyadenylation signal, a putative PBS, and characteristic inverted repeat sequences at locations which are typical of a retroviral LTR (26). This region also shows a significant homology to the LTRs of HTLV-I and HTLV-II. In addition, the HRES LTR-like sequence contains direct repeats and a complete TAR sequence present in the HIV-1 LTR. The pMAI probe, which contains the 5′ LTR in addition to the gag region of HTLV-I or the p1/1 probe of HRES-1/1 hybridized only to the p1/1 fragment of the 20-kb recombinant clone 1/1. This suggests that clone 1/1 does not contain two LTRs and thus probably represents a deleted provirus.

By comparative sequence analysis two regions, one 31 and one 36 amino acid long, were noted in HRES p25 which showed identity in a range of 23% to 39% with gag protein sequences of HTLV-I, HTLV-II, HIV-2, and FSV. This level of identity is significant when compared to that between HTLV-I and HIV-2 (Fig. 6), as well as to that between p25 and all other known protein sequences. The level of amino acid homology between very closely related viruses such as HTLV-I/HTLV-II and HIV-1/HIV-2 is much higher, 78% and 54%, respectively. This suggests that the relationship of HRES p25 to human

retroviruses, particularly to HTLV-I/HTLV-II is more distant than that between HTLV-I and HTLV-II but closer than that between HTLV-I and HIV-2.

Hybridization analysis with genomic DNA samples of selected phylogenetic stages revealed that HRES-1/1 is apparently confined to the primate lineage. In this evolutionary aspect, the HRES-1/1 shows similarity with ERV1, an earlier described ERS (2). The cellular DNA regions flanking ERV1 in humans and chimpanzees have an identical restriction map. ERV1 is also a single-copy human proviral locus which contains a 3′ LTR but lacks a 5′ LTR (2). The data suggest that (i) the HRES-1/1 sequence was not phylogenetically inherited but entered the genome of primates as an exogenous element, and (ii) the human genomic HRES-1/1 sequence might have originated from an as yet unidentified exogenous retrovirus. From the deduced amino acid sequence of HRES-1/1 p25, residues 6–36 show a relatively high homology to gag regions of HTLV-I and HTLV-II, while residues 104–139 display a significant homology to gag regions of HIV-2 and FSV, suggesting that the original exogenous virus infecting primate may be chimeric in structure.

Unlike murine and baboon endogenous proviruses, human ERSs are not known to produce infectious retroviral particles or even retroviral proteins. Major (1.7 kb and 3 kb) transcripts and minor (2.2 kb and 3.6 kb) RNA species were detected by LTR- and env-region specific probes of a full length human ERS in human placenta, breast carcinoma, and colon carcinoma cells (33). Since sequences of cDNA clones corresponding to the 1.7 kb and 3 kb RNA species revealed in-frame termination codons, neither clone could encode full length env proteins. In contrast, HRES-1/1, which is transcribed in various human tissues and cell lines, does contain potential ORFs. This suggests that the HRES-1/1-encoded p25 and/or p15 may have biological significance. The present data show that HRES-1/1 is a transcriptionally active locus with homology to the LTR and gag regions of human retroviruses and open up new avenues to study the role of ERS in the human genome.

ACKNOWLEDGEMENTS

We thank Drs. Piero Balduzzi (University of Rochester) and Malcolm A. Martin (National Institutes of Health) for helpful discussions. This work was supported by grants AG-06350, AG-08177, AI-19658 and POI-AI-21288 from the US Public Health Service. A.P. was supported by the James P. Wilmot Cancer Research Foundation and American Cancer Society Institutional Grant IN-18-30.

*To whom correspondence should be addressed at Department of Molecular Medicine and Immunology, Roswell Park Memorial Institute, Buffalo, NY 14263, USA

REFERENCES

1. Martin, M.A., Bryan, T., Rasheed, S. and Kahn, A.S. (1981) Proc. Natl. Acad. Sci. USA 78, 4982–4986.
2. Bonner, T.I., O'Connell, C. and Cohen, M. (1982) Proc. Natl. Acad. Sci. USA 79, 4709–4713.
3. Callahan, R., Drohan, W., Tronick, S. and Schlom, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5503–5507.
4. Noda, M., Kurihara, M. and Takano, T. (1982) Nucl. Acids Res. 10, 2865–2879.
5. Kroger, B. and Horak I. (1987) J. Virol. 61, 2071–2075.
6. Harada, F., Tsukada, N. and Kato, N. (1987) Nucl. Acids Res. 15, 9153–9162.
7. Mager, D.L. and Henthorn, P.S. (1984) Proc. Natl. Acad. Sci. USA 81, 7510–7514.
8. Maeda, N. (1985) J. Biol. Chem. 260, 6698–6709.
9. O'Connell, C., O'Brien, S., Nash, W.G. and Cohen, M. (1984) Virology 138, 225–235.
10. Benveniste, R.E. and Todaro, G.J. (1974) Nature 252, 170–173.
11. Chattopadhyay, S.K., Lowy, D.R., Teich, N.M., Levine, A.S. and Rowe, W.P. (1974) Proc. Natl. Acad. Sci. USA 71, 167–171.

12. Risser, R. and Horowitz, J.M. (1983) Annu. Rev. Genet. 17, 85–121.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York.
14. Bolton, E.T. and McCarthy, B.J. (1982) Proc. Natl. Acad. Sci. USA 48, 1390–1397.
15. Perl, A., Wang, N., Williams, J.M., Hunt, M.J., Rosenfeld, S.I., Condemi, J.J., Packman, C.H. and Abraham, G.N. (1987) J. Immunol. 139,3512–3520.
16. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA. 76, 3683–3687.
17. Bradley, J.E., Bishop, G.A., St.John, T. and Frelinger, J.A. (1988) Biotechniques 6, 114–116.
18. Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA. 80, 3618–3622.
19. Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem. 132, 6–13.
20. Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. Kedes, L. (1983) Mol. Cell. Biol. 3, 787–795.
21. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983) Cell 34, 779–787.
22. Ravetch, J.V., Siebenlist, U., Korsmeyer, S.T., Waldmann, T.A. and Leder, P. (1981) Cell 27, 583–591.
23. Hieter, P.A., Max, E.E., Seidman, J.G. Maizel, J.V.,Jr, and Leder, P. (1980) Cell 22, 197–207.
24. Brown, W.M., Prager, E.M., Wang, A.M. and Wilson, A.C. (1982) J. Mol. Evol. 18, 225–239.
25. Weiss, R., Teich, N., Varmus, H. and Coffin, J. (1985) RNA tumor viruses. Ed. 2. Cold Spring Harbor Laboratory Press, New York.
26. Temin, H.M. (1981) Cell 27, 1–3.
27. Rosen, C.A., Sodroski, J.G., Goh, W.C., Dayton, A.I., Lippke, J., and Haseltine, W.A. (1986) Nature 319, 555–559.
28. Shimotono, K, Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 82, 3101–3105.
29. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Withehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R.Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985) Nature 313, 277–284.
30. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M. (1987) Nature 326, 662–669.
31. Hampe, A., Gobet, M., Sherr, C.J. and Galibert, F. (1984) Proc. Natl. Acad. Sci. USA 81, 85–89.
32. Kohl, N.E., Legouy, E., DePinho, R.A., Nisen, P.D., Smith, R.K., Gee, C.E. and Alt, F.W. (1986) Nature 319, 73–77.
33. Rabson, A.B., Hamagishi, Y., Steele, P.E., Tykocynski, M. and Martin, M.A. (1985) J. Virol. 56, 176–182.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.