Pancreatic triacylglycerol lipase in a hibernating mammal.
I. Novel genomic organization

Teresa L. Squire1 and Matthew T. Andrews2
1Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614; and 2Department of Biology, University of Minnesota Duluth, Duluth, Minnesota 55812

Submitted 22 November 2002; accepted in final form 24 September 2003

Squire, Teresa L., and Matthew T. Andrews. Pancreatic triacylglycerol lipase (PTL) is expressed in novel locations during hibernation in the thirteen-lined ground squirrel (Spermophilus tridecemlineatus). PTL cDNAs isolated from two of these locations, heart and white adipose tissue (WAT), contain divergent 5'-untranslated regions (5'-UTRs) suggesting alternative promoter usage or the possibility of multiple PTL genes in the ground squirrel genome. In addition, cDNAs isolated from WAT contain tracts of retroviral sequence in their 5'-UTRs. Our examination of PTL genomic clones isolated from a thirteen-lined ground squirrel genomic DNA library, coupled with genomic Southern blot analysis, enabled us to conclude that PTL mRNAs expressed in heart and WAT are the products of the same single-copy gene. The 5' portion of this gene spans 9.2 kb, is composed of 6 exons, and contains a full-length endogenous retroviral genome with conserved long terminal repeats (LTRs). Alignment of the ground squirrel PTL gene with the mouse, rat, and human PTL genes indicates that this retrovirus inserted into the ground squirrel genome ~200 bases upstream of the original PTL transcriptional start site. The insertion is a relatively recent event based on largely intact open-reading frames containing minimal frame-shift and nonsense mutations. The high-percentage identity (99.2%) shared between the open-reading frames containing minimal frame-shift and nonsense site. The insertion is a relatively recent event. The 5'-UTRs of the WAT PTL cDNAs contained distinct tracts of retroviral-like elements that were not found in the 5'-UTR of the heart PTL cDNA sequence. Bauer et al. (3) proposed two potential explanations for these differences at the genomic level, both of which required a retroviral insertion event. The first scenario involved the use of different transcriptional start sites and/or alternative splicing of the PTL transcript from a single gene. The second scenario suggested that the initial PTL gene was duplicated via nonhomologous recombination. Insertion of a retrovirus into the promoter of one of these two duplicate genes provided “novel” regulatory elements that allowed expression of this gene in WAT. Expression of this latter gene resulted in the chimeric mRNA that was found in WAT (3).

In this study, we seek to provide insight regarding the sequence organization and number of PTL genes that are present in the thirteen-lined ground squirrel genome. First, we examine PTL mRNA levels in several hibernating ground squirrel tissues using RT-PCR, a more sensitive method than Northern blot analysis, to determine how broadly this gene is expressed. Next, we isolate PTL cDNAs from a pancreas cDNA library and address whether the abundant PTL message found in this tissue resembles the form of PTL mRNA expressed in the heart or WAT. The structure and organization of the ground squirrel PTL gene(s) is then addressed using Southern blot analysis and sequencing of recombinant lambda clones from a thirteen-lined ground squirrel genomic library. Further analysis of the retroviral sequence, in combination with the Southern blot data, has led us to the conclusion that the sequences present in both WAT and heart PTL cDNAs are products of a single gene and that the retroviral insertion is a relatively recent event.

MATERIALS AND METHODS

Sections on Animals, Pancreas cDNA library construction and screening, Expression and isolation of recombinant PTL protein, and PTL assays (including Western blots) can be found in the MATERIALS AND METHODS of the accompanying paper (30).

RNA isolation. Total RNA was isolated from ground squirrel tissues as previously described in Andrews et al. (2) using a modifi-
cation of the method developed by Chomczynski and Sacchi (8). Before extraction, however, WAT homogenates from the abdominal WAT pad were centrifuged at 3,000 g for 10 min, and the lipid layer was removed. Extraction then proceeded on the WAT homogenate as described in Andrews et al. (2). RNA integrity was checked by separating total RNAs on 1.2% agarose gels containing 3% formaldehyde followed by staining with ethidium bromide.

RT-PCR. First-strand cDNA was generated from 2 μg total RNA using the SuperScript first-strand synthesis system for RT-PCR (Invitrogen-Life Technologies) with oligo(dT) as the primer. After synthesis, however, treatment with RNase H was not performed as recommended by the protocol. Products of the first-strand cDNA were diluted twofold with water before proceeding to PCR.

PCR was executed using 1 μl cDNA template from the diluted first-strand reaction and 2 μl Tag DNA polymerase (Invitrogen-Life Technologies) in the presence of 1.5 mM MgCl₂. Reactions using PTL and β-actin primers were done separately for 28 and 20 cycles, respectively, in HotStart storage and reaction tubes (Molecular Bio-Products). After the hot start, denaturation was performed at 94°C for 30 s, annealing at 55°C for 45 s, and extension at 72°C for 1 min. The primer pair used to measure relative PTL mRNA levels was 5’ CAGATGTCACACCCGCTT and 3’ and 5’ GTGCGCAATGACTGGAC 3’. This pair fell within the coding region of the message and spanned at least one intron based on alignment with the nucleotide sequence for the human PTL gene (accession no. AH003527; Ref. 29). The primer pair used to measure β-actin mRNA levels was 5’ GACAGATGACAGAGAGGAG 3’ and 5’ ACATCTGTCGAGAGGTGG 3’. β-actin primers served as a control for RNA integrity and amplification in the PCR reaction.

Two negative controls were also performed. For each RNA sample, an RT reaction was performed that replaced the SuperScript II RT enzyme with water. A portion of this reaction, as described above, was then used as template for PCR. In addition, a PCR reaction was carried out that substituted water for template. All control reactions were negative for DNA or other contamination (data not shown). PCR results were viewed on 5% acrylamide, 1X TBE gels.

**Genomic Southern blot analysis.** Frozen thirteen-lined ground squirrel liver was ground into fragments with a mortar and pestle and immersed in digestion buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% SDS, and 0.1 mg/ml proteinase K) for incubation overnight at 50°C with shaking. Genomic DNA was then isolated from this mixture using a cesium chloride (CsCl) gradient as described in Curtis and Haselkorn (9). Sixty micrograms of this genomic DNA was digested separately with each of five restriction enzymes: BamHI, EcoRI, HindIII, PstI, and XhoI. Triplicate Southern blots containing 20 μg of each digest were made according to standard methods (27) using Hybond- XL nylon membrane (Amersham). Probe hybridization and washes were performed according to the manufacturer’s suggestions for this membrane.

For detection of PTL gene fragments, the probes were generated via PCR using three different primer pairs and labeled by random priming. **Primer pair I** (5’ CCAATGATAGAGATGCGGC 3’ and 5’ GTTGGAAATGGTTGCTGGG 3’) was unique to the 5’-UTR of the WAT PTL cDNA clone 22A4, bases 319–603 (Fig. 3; accession no. AF177403; Ref. 3). **Primer pair II** (5’ ATGTGCTATAGAGAGCCG 3’ and 5’ ATGGCAGATCCGTCAGGC 3’) was unique to the 5’-UTR of the heart PTL cDNA clone 29H4, bases 1088–1852 of heart PTL sequences (27). The DNA isolated from two of the stocks was partially sequenced. These two DNA preparations were selected because they contained sequence found in both the heart and WAT 5’-UTRs, as well as sequence found at the start of the PTL coding region.

**DNA sequencing.** Positive cDNA and lambda library clones were sequenced using ABI Prism 377 automated cycle sequencers (PerkinElliott Biosystems). Multiple sequences for a particular cDNA clone were aligned and analyzed using MacVector and AssemblyLIGN software (Oxford Molecular Group). Multiple sequences for both lambda library clones were aligned and analyzed using SeqMan II, version 4.05 (DNASTAR). The consensus sequences that were generated using both programs were edited manually to resolve discrepancies. Finished sequences were compared with known sequences entered into the National Center for Biotechnology Information (NCBI) database using the BLAST tool (4).

**Long terminal repeat analysis.** Two programs, ModelInspector Release 4.7.4 (Genomatix Software, GEMS Laucher 3.1, accessed via http://www.genomatix.de/; Ref. 10) and MatInspector Release 5.2 (Genomatix Software; accessed via http://www.genomatix.de/; Ref. 23), were utilized to identify the putative long terminal repeats (LTRs) and the consensus elements within them. For each consensus element, the core sequence represents the four most highly conserved, contiguous bases in the defining matrix used by the MatInspector program.

**Comparative genomic analysis.** Human and rat PTL sequences were obtained from the UCSC Bioinformatics Site Genome Browser (http://genome.ucsc.edu/). For the human sequence, the human genome browser assembly was the Consolidated Sets Genome Browser release 20020402, one of the human genome assemblies that have been submitted to and approved by the Genome Assembly Database. The human PTL sequences were aligned against the rat PTL sequences using the BLAST tool (4).
fragment ID number was RNOR01013788, and the bicatiq was kaxw_ghoa. The mouse PTL sequence was obtained from the Ensembl web site (http://www.ensembl.org/) using mouse genome browser version 12.3.1 (March 3, 2003). The Ensembl gene ID was ENSMUSG00000042344. All nucleotide and amino acid alignments were performed with Clustal W (http://clustalw.genome.ad.jp).

Data deposition. The new sequences reported in this paper have been deposited in the GenBank database (accession nos. AF395870 and AY071823). RESULTS

PTL mRNA and protein in hibernating thirteen-lined ground squirrel tissues. Total RNA was isolated from 10 hibernating thirteen-lined ground squirrel tissues. First-strand cDNA was then generated from this total RNA, and PCR reactions were performed to test for the presence of PTL mRNA in these tissues. The primers chosen for this study represented sequence found within the coding region of the PTL message. Using this method, PTL mRNA was found in all 10 hibernating tissues examined (Fig. 1A). Previous observations with Northern blot analysis suggested that expression of the PTL message during hibernation was limited to WAT, heart, pancreas, and testes (3). The present experiment, coupled with the previous Northern blot analysis, demonstrates that these four tissues show the highest mRNA levels during hibernation. Furthermore, it also shows that PTL mRNA expression in this mammalian hibernator is more widespread than previously realized.

To determine whether the PTL mRNA was being translated into protein, we used Western blots containing protein prepared from ground squirrels during the hibernation season. Soluble protein extracts from the four tissues with the highest PTL mRNA levels (Fig. 1A and Ref. 3) were probed with anti-ground squirrel PTL antibody. A lane containing purified recombinant ground squirrel PTL, the protein against which the anti-PTL antibody was made, was utilized as the control. PTL protein of the same size as the recombinant protein was observed in each of the tissues (Fig. 1B). A larger band, ~65 kDa in size, was also observed in the pancreas lane. Although the functional significance of this larger-sized protein is not known, a protein of similar size was also observed when a lipolytic enzyme exhibiting activity similar to PTL was purified from rat brain, immunoblotted, and incubated with anti-rat PTL antibody (34). With the exception of this larger protein band seen in the pancreas lane, all other protein bands of a size different from the purified recombinant ground squirrel PTL were observed on the blot treated with preimmune serum and thus represent nonspecific binding (Fig. 1B).

Thirteen-lined ground squirrel PTL cDNA analysis. Earlier experiments using Northern blot analysis demonstrated that the size of the PTL mRNA in WAT was 2.3 kb, whereas the size of the PTL mRNA in heart, pancreas, and testes was 1.8 kb (3). Sequencing of two full-length PTL cDNAs isolated from a WAT cDNA library, 7G5 and 22A4 (accession nos. AF177402 and AF177403, respectively), revealed that the 5’-UTR of the message found in WAT differed from the 5’-UTR of the message found in heart (3). Starting just three bases upstream of the start codon, these two regions from heart and WAT shared no similarity. When the 5’-UTR sequences of the two WAT cDNA clones were compared with known sequences in the GenBank database using the BLAST tool (http://www.ncbi.nlm.nih.gov/blast/), it was found that this region contained segments of retroviral-like elements (3). Ultimately, this scenario led us to examine exactly how many PTL genes were present in the thirteen-lined ground squirrel genome. Before this question was addressed, the sequence of the PTL message expressed in its traditional location, the pancreas, was determined to learn whether it more closely resembled the mRNA in heart or those found in WAT.

Two full-length PTL cDNA clones were isolated from a thirteen-lined ground squirrel pancreas cDNA library that was created using poly(A)⁺ mRNA from hibernating and active ground squirrel pancreases. The consensus sequence obtained for these pancreatic clones (accession no. AF395870) showed only minor differences when compared with the heart PTL cDNA sequence (accession no. AF027293; Ref. 2). These differences included two base changes, one found in the 5’-UTR and the other found in the 3’-UTR, which denote
probable allelic or individual variations in the PTL message. The pancreatic PTL cDNAs were also missing the first eight bases found in the 5′-UTR of the heart cDNA clone. This difference, however, was likely due to the reverse transcription process used to generate the cDNAs and does not represent a true difference in length of the two messages. Thus, based on the size of the two messages seen on Northern blots (2, 3) and based on the sequences of the full-length cDNAs, it was concluded that the PTL mRNAs found in the pancreas and the heart were identical.

**Thirteen-lined ground squirrel PTL gene: Genomic library analysis.** The dissimilarity in PTL mRNAs from WAT (2.3 kb; 5′-UTR retroviral sequence) and that from heart and pancreas (1.8 kb; no detectable retroviral sequence) prompted us to investigate the possibility of multiple PTL genes. To address this question, our first approach was the construction and screening of a thirteen-lined ground squirrel genomic library. Approximately 325,000 plaques from the unamplified library were screened for the presence of *PTL* using an oligonucleotide probe whose sequence was complementary to the 5′-most bases of the PTL coding region. A secondary screen was then performed to isolate clones that contained sequence from the three regions of interest: 5′-UTR of WAT, 5′-UTR of heart/pancreas, and the 5′ portion of the PTL coding sequence. Two clones were isolated, labeled P2 and P3, that contained these three regions. The main difference between these two clones was that recombinant clone P3 contained the entire region studied, whereas the 5′ end of clone P2 ended in exon 2 (Fig. 2A). These genomic library clones were sequenced on both strands using primers generated for sequencing the heart and WAT PTL cDNAs. Additional primers, to fill gaps in the sequence, were created based on genomic clone sequencing runs. *PTL* gene sequencing results (accession no. AY071823) are summarized in Fig. 2A.

Analysis of these genomic library clones indicated that one *PTL* gene could account for the retroviral elements contained in the 5′-UTRs of the WAT cDNAs (7G5 and 22A4), as well as the 5′-UTR sequence found in heart and pancreatic cDNAs. Because the heart and pancreatic cDNAs were concluded to be identical, the heart and pancreatic cDNAs will be referred to by the nucleotide sequence present in the heart PTL cDNA. The genomic sequence found in each exon was nearly identical with its corresponding region in one or more of the PTL cDNAs (Fig. 2A). The sequence in exon 1, for example, shared 98.2% identity with the first 113 bases of the cDNA sequence in heart clone 29H4. Sequence of exon 2 shared 100% identity with bases 1–428 of the WAT cDNA clone 7G5. Exon 3 sequence was 99.8% identical to its corresponding cDNA region in WAT clone 22A4 bases 1–661. The last 135 bases of this exon were also 100% identical to bases 429–563 of clone 7G5. Exon 4 was 99.6% identical to the final 281 bases found in the 5′-UTRs of both WAT cDNA clones (bases 662–942 in 22A4 and bases 564–844 in 7G5). Exon 5 sequence was 99.5% identical to bases 1–366 in heart clone 29H4, on the other hand, was a complete match for the cDNA sequence found in this region. This exon marked the start of the PTL coding region in each of the cDNA clones. Overall, the genomic nucleotide sequence present in the exons was 99.7% identical to the equivalent cDNA sequences. Probable allelic variations in the ground squirrel genome could account for the less than 100% identity. All sequences present in the cDNAs were found in their original orientation in the genomic sequence with one exception. The first 52 bases in the 7G5 WAT cDNA were found in a reverse and complementary fashion in the genomic sequence. This type of inversion, present in the cDNA, was most probably the result of an error that occurred during first-strand cDNA synthesis of 7G5. Also, although not a change in orientation, the first portion of the heart PTL cDNA clone was found at two places in the genomic sequence. This cDNA sequence comprised all of exon 1 and was repeated again at the start of exon 5. This repeated sequence is part of a larger direct repeat of retroviral origin as indicated by the bold horizontal arrows in Fig. 2A. The significance of this larger direct repeat will be discussed shortly.

**Thirteen-lined ground squirrel PTL gene: Splice-junction analysis.** In the thirteen-lined ground squirrel genome, the presence of a single *PTL* gene requires that alternative splicing would occur to produce the two unique PTL cDNAs found in WAT. In addition, if alternative promoters are not used for regulation of this gene, then alternative splicing would also be required to produce the PTL cDNA form found in heart and pancreas. To examine the potential for alternative splicing of this ground squirrel gene, splice junction sequences for the first six exons were analyzed (Fig. 2B). The consensus sequence for the 5′ intron splice sites was 5′-NG | GUPuPuGtN46 [notation is based on that used in Goldstrohm et al. (11), where the arrow marks the exon-intron junction, and Pu denotes purine]. This sequence closely resembled the consensus sequence for mammalian 5′ intron splice sites, 5′-2AG \downarrow GUPuPuAGU46 (the underlined positions are the most highly conserved residues; reviewed in Ref. 11). The 3′ intron splice sites for this gene, however, were less well conserved. The consensus sequence

![Figure 2](https://example.com/image.png)

*Fig. 2.* The thirteen-lined ground squirrel *PTL* gene (accession no. AY071823). A: in this genomic schematic, exons are represented as boxes 1 through 6, while introns are represented as lines connecting these boxes. A vertical arrow (at "ATG") indicates the beginning of the coding region. Bold, horizontal arrows indicate a 396-bp direct repeat. A representation of the 5′-UTR of the WAT cDNA clone 22A4, is shown above the genomic schematic, while representations of the 5′-UTRs found in the heart PTL cDNA, 29H4, and in the WAT PTL cDNA, 7G5, are shown below it. The 5′-most 113 bases of the heart PTL cDNA, shown in light gray, are part of the direct repeat and are encoded in one of two possible locations in the *PTL* gene; either exon 1 or the first part of exon 5. The remaining 253 bases in the 5′-UTR of the heart PTL cDNA, shown in dark gray, comprise the second part of exon 5. The hatched (7G5) and checkered (22A4) regions in the two WAT cDNAs represent portions of these cDNAs that share no similarity to each other or to any sequences currently in GenBank. The stippled regions represent sequences in the WAT cDNAs that share similarity to portions of the porcine retroviral sequence (accession no. AF038600) encoding the Gag, Pol, and Env polyproteins. At the bottom, a restriction map of the genomic sequence is shown with the location of probes I and II used in Fig. 3B. B: BamHI; E: EcoRI; H: HindIII; P: PstI; B: intron-exon splice junctions in the 5′ portion of the thirteen-lined ground squirrel *PTL* gene. In each exon or intron, the number of bases omitted for space considerations is indicated. The start codon is highlighted in gray. The consensus sequences for the 5′ and 3′ intron splice sites are given below the genomic sequence. Abbreviations for the consensus sequence: r, purine; n, no base is conserved in this position. This figure was modeled after the one in Sims et al. (29).
derived from these sites was $^{4}$U(A/C)(C/A)(C/G) $\downarrow$ AU$^{+2}$ and differed somewhat from the mammalian consensus $^{−4}$NPvAG $\downarrow$ PuN$^{+2}$ (the notation is as previously described, Py denotes pyrimidine; reviewed in Ref. 11).

The lessened degree of conservation of the 3′ splice sites could enable alternative splicing of a primary PTL mRNA transcript and would address the issue of three unique PTL cDNAs. Exons 4 and 6, which contained sequence shared by both WAT clones, had 3′ splice sites that were identical to the mammalian consensus. Exon 2 and the first part of exon 3, on the other hand, were unique to specific WAT clones and had 3′ splice sites that deviated considerably from the consensus. Less efficient processing of these latter splice sites could explain the variation in cDNA products observed in WAT. Similarly, the 3′ splice site for exon 5 failed to conform to the mammalian consensus. If transcription in heart and pancreas started at exon 1 instead of exon 5, then this lessened degree of conservation could enable splicing of the primary transcript to produce the heart/pancreatic PTL cDNA.

**Fig. 3.** Southern blot analysis of the PTL gene(s) using genomic DNA isolated from the thirteen-lined ground squirrel. 

A: PTL cDNAs from heart (accession no. AF027293) and WAT clone 22A4 (accession no. AF177403). Thin lines represent the 5′- and 3′-UTRs, whereas thick lines represent the coding regions. The coding regions of the two PTL cDNAs are 100% identical at the nucleotide and amino acid levels. Horizontal arrows indicate the positions of the three probes (I–III) used in B. B: blots I, II, and III represent three identical Southern blots made using Spermophilus tridecemlineatus DNA. Each lane on the blot contains 20 μg of genomic DNA digested with one of five restriction endonucleases: BamHI (B), EcoRI (E), HindIII (H), PstI (P), and Xhol (X). Blot I was probed with probe I and was exposed to film for 3 days. Blot II was probed with probe II and was exposed to film for 7 days. Blot III was probed with probe III and was exposed to film for 7 days.

**Thirteen-lined ground squirrel PTL gene: Genomic Southern blot analysis.** A second approach was taken to investigate the possibility of multiple PTL genes present in the thirteen-lined ground squirrel genome. This approach involved the analysis of Southern blots containing ground squirrel genomic DNA. Three identical thirteen-lined ground squirrel genomic Southern blots were probed with $^{32}$P-labeled sequences complementary to the 5′-UTR of WAT PTL mRNA (probe I), the 5′-UTR of heart/pancreatic PTL mRNA (probe II), and the PTL coding region shared by all three PTL messages respectively (probe III; Fig. 3A). The results of this experiment are shown in Fig. 3B.

In blot I of Fig. 3B, multiple DNA fragments hybridized to the probe complementary to the 5′-UTR of the WAT PTL mRNA. This number of fragments was more than would be expected based on the restriction site analysis of the ground squirrel PTL gene (Fig. 2A). Sequence complementary to probe I fell entirely within exon 3 of the genomic sequence. Regardless of which enzyme was used to cut the DNA, only a single...
band would be expected to hybridize with this probe. The presence of multiple bands in each lane on the Southern blot (Fig. 3B) indicates that sequence complementary to this probe was present in the thirteen-lined ground squirrel genome at places other than simply upstream of the PTL gene. Because this probe sequence is retroviral in nature, the presence of multiple bands suggests that the retrovirus from which this probe sequence was derived had inserted itself at multiple sites in the ground squirrel genome. The ground squirrel PTL gene represents only one of these sites of insertion.

In blot II of Fig. 3B, a similar result was seen. Multiple DNA fragments also hybridized to the probe that was complementary to the 5′-UTR of heart/pancreatic PTL mRNA. A BlastN analysis of this probe sequence uncovered no similarity to any potentially repetitive DNA sequence. Thus each lane on the blot would be expected to contain only two bands based on the restriction site analysis of the ground squirrel PTL genomic sequence (Fig. 2A). Sequence complementary to probe II was present in exon 1 and again in exon 5. Furthermore, the first 113 base pairs (bp) of this sequence was part of a direct repeat (Fig. 2A). The nature of this direct repeat was later determined to be of retroviral origin (Fig. 5). Given this new context, the result obtained on Southern blot II (Fig. 3B) could be explained by retroviral insertion at multiple sites in the ground squirrel genome as described earlier for blot I results.

Blot III of Fig. 3B presented a picture that was different from the previous two blots. This blot was probed with sequence that was complementary to the PTL coding region. Each lane on this blot contained one prominent band and at most three or four bands in total. In general, this banding pattern was consistent with the restriction site pattern found within this region of the PTL cDNAs (Fig. 3A). Because HindIII, PstI, and XbaI were non-cutters within this region, a single band would be predicted in each of these lanes unless cut sites arose within introns. Although we did not sequence this portion of the ground squirrel gene, we were able to align probe III cDNA sequence with the human PTL gene (accession no. AH003527; Ref. 29). This alignment suggested that probe III spanned several exons (8 in humans), so the introduction of intronic cut sites would not be unexpected. Three bands in the XbaI lane, where only one was seen in each of the HindIII and PstI lanes, suggests that two XbaI sites were introduced in introns. BamHI and EcoRI sites, on the other hand, were present in the probe III cDNA sequence (Fig. 3A). As a result, at least two bands would be predicted for each of these lanes (blot III, Fig. 3B). Two bands were seen in the EcoRI lane, whereas only one band was seen in the BamHI lane. Because BamHI is not affected by mammalian CpG methylation, this mechanism could not be used to explain the presence of a single band in this lane. Revisiting the earlier alignment of the probe III cDNA sequence with the human PTL gene, however, does present one possible explanation. This cut site lies just five bases from an intron-exon splice junction. If this splice junction was not completely conserved between the two species, then the BamHI restriction site seen in the cDNA sequence could have been abolished by a splice site in the ground squirrel genomic DNA sequence. The faint bands migrating at the top of the BamHI and PstI lanes were not included in this analysis as they likely represent uncut genomic DNA.

In summary, Southern blot analysis indicates that while the 5′-UTR sequences from heart/pancreatic and WAT PTL cDNAs are found at multiple sites throughout the thirteen-lined ground squirrel genome, the PTL coding region sequence is much more limited in scope. In addition, although the existence of multiple PTL genes cannot be excluded on the basis of Southern blots alone, the limited banding patterns seen on the PTL coding region blot suggests the presence of a single PTL gene.

Thirteen-lined ground squirrel PTL gene: Retroviral sequence analysis. Previous sequence analysis of WAT PTL cDNAs revealed portions of retroviral sequence in their 5′-UTRs (3). These retroviral elements were present in a conserved linear order in the cDNAs (3) and in the PTL gene exons (Fig. 2A). This conservation of order suggested that the elements present in the cDNAs derived from a retrovirus that had integrated into the thirteen-lined ground squirrel genome upstream of the coding region in the PTL gene. To test the validity of this hypothesis, a BlastX analysis was performed on the ground squirrel genomic sequence. As part of this analysis, direct comparisons were made between the translated ground squirrel sequence and the Gag, Pol, and Env polyproteins encoded by four different full-length γ-retroviruses: porcine endogenous retrovirus (P-ERV; accession no. AF038600; Ref. 1), gibbon ape leukemia virus (GALV; accession no. U60065; Ref. 21), Mus dunni endogenous virus (MDEV; accession no. AF053745; Ref. 38), and Friend murine leukemia virus (FMLV; accession no. M93134; Ref. 24). To obtain a more accurate alignment of the translated ground squirrel sequence with each of these four γ-retroviruses, the low-complexity option for the BlastX analysis was deselected.

Overall, the length of the ground squirrel retroviral sequence (8,569 nt) was consistent with the length of complete retroviral genomes (Table 1). Additionally, alignment could be shown between the translated retroviral region of the ground squirrel sequence and, on average, ≥98% of the amino acids encoded by each of the three retroviral genes found in the four γ-retroviruses. Percentage amino acid identities for these alignments ranged from 39 to 64%, with the highest degree of similarity seen in the Pol polyprotein region and the lowest seen in part of the Env polyprotein.

Using the P-ERV for comparison, alignment with the Gag polyprotein began with base 1069 in exon 2 of the ground squirrel genomic sequence and continued through to base 2712 in intron 3 (Fig. 4). Within this region, one premature in-frame stop codon (denoted by an “X” in Fig. 4) was present near the start of intron 3. Continuing in this same +1 frame, alignment with the Pol polyprotein started another 154 bases downstream in intron 3. Deletion of a single nucleotide at base 3576 caused a frame-shift mutation (fs1) to occur. In the +3 frame, alignment with the Pol polyprotein was maintained and ended with base 6293 in intron 4. Within this region after fs1, two premature in-frame stop codons were present. Alignment with the Env polyprotein started with base 6205 in exon 4, which placed this ORF in the original +1 frame. Because the second half of the Pol polyprotein was encoded in the +3 frame, however, these two ORFs while overlapping were not in frame. This ORF arrangement for the Pol and Env polyproteins was consistent with that found in the four γ-retroviruses shown in Table 1. In the ground squirrel sequence, the Env ORF spanned the remainder of exon 4 and the majority of intron 4 and ended with base 8180. Deletion of a single nucleotide at base 7953 caused a frame-shift (fs2) in this ORF as well. This alignment
based on the BlastX analysis supports the hypothesis that an entire retroviral genome is present in the 5′ portion of the ground squirrel PTL gene.

*Thirteen-lined ground squirrel PTL gene: Identification of LTRs.* The process of reverse transcription that enables the original retroviral genome to integrate into the host genome creates LTRs that flank the coding regions of the gag, pol, and env genes (reviewed in Refs. 7 and 32). In the ground squirrel genomic sequence, a 396-bp direct repeat corresponding to bases 89–469 and bases 8262–8657 (bold horizontal arrows in Figs. 2A and 4) is located where one would expect to find the LTRs of an integrated provirus. The nucleotide sequences within this direct repeat are 99.2% identical. Two programs, ModelInspector (Genomatix Software; Ref. 10) and MatInspector (Genomatix Software; Ref. 23), were used to identify elements and structures common to mammalian C-type (γ-retrovirus) LTRs. Figure 5 shows these elements included a CCAAT box and a TATA box found in the U3 region of the LTR; a hairpin loop and a poly(A) signal found in the R region of the LTR; a poly(A) downstream element with the consensus GTGGT found in the U5 region of the LTR; and a terminal inverted repeat (10). Two additional consensus elements, an upstream element in the U3 region and a hairpin in the U5 region, however, were not present in the ground squirrel direct repeat.

The boundaries of each half of the direct repeat, which will now be referred to as the putative 5′- and 3′-LTRs respectively, were determined based on the locations of the tRNA primer-binding site (PBS) and the polypurine tract (PPT) (Fig. 5). The PBS lies immediately downstream of the 5′-LTR and acts as the priming site for minus-strand DNA synthesis of the retrovirus (reviewed in Ref. 32). This PBS is complementary to 18 bases at the 3′ end of a host-encoded tRNA. Although the tRNA for proline, and to a lesser extent glutamine, acts as the typical tRNA primer for mammalian C-type retroviruses (reviewed in Ref. 32), the PBS (bases 487–504) found in this ground squirrel genomic sequence shared 100% complementarity to the 3′ terminus of a human glycine tRNA (accession no. K00208; Ref. 12). The PPT, on the other hand, lies immediately upstream of the 3′-LTR and provides the priming site for plus-strand DNA synthesis of the retrovirus (reviewed in Ref. 32). In this ground squirrel sequence, a near perfect PPT (17/18 nucleotides) was present from bases 8244–8261. These sequences generally range from 7 to 18 bases in length (22). Last, bordering the integrated provirus was a 4-bp direct repeat (ATTCT). Formation of this direct repeat is a consequence of the viral DNA insertion event (as reviewed in Ref. 6).

Within the putative LTRs, boundaries for the U3, R, and U5 regions were determined by the definition of these regions. As reviewed in Vogt (35), the transcription start site establishes the boundary between the U3 and the R regions, and the polyadenylation site marks the boundary between the R and the
U5 regions. In the ground squirrel genomic sequence, the transcription start sites were located at the beginning of exons 1 and 5, 36 bases downstream of their respective TATA boxes (Fig. 5). Conversely, the polyadenylation signal spanned bases 409–414 and bases 8582–8587 in the 5′/H11032- and the 3′/H11032-LTRs, respectively. Relative to these locations, the R-U5 boundary was marked 21 bases downstream of this consensus element (Fig. 5). Two observations provided the foundation for this determination. The first, supplied by Chen and Barker (7), was that most R regions ended with the dinucleotide CA. An alignment with the R regions from other mammalian type C retroviruses was used to select the appropriate CA (7). The second, as reviewed in Petropoulos (22), was that the polyadenylation tract was commonly found 15–20 bases downstream of the polyadenylation signal. The proposed boundary met both of these guidelines.

Fig. 5. Sequence of the two putative long terminal repeats (LTRs) at both ends of the GS-ERV in the thirteen-lined ground squirrel PTL gene (accession no. AY071823). An asterisk denotes identical bases found in the two sequences. Conserved elements and structures are underlined with a single line in the 5′-LTR sequence, and the core sequence within each element according to MatInspector is set in bold text. The first and last bases of the 396-bp direct repeat are shaded light gray. The first base in exons 1 and 5 (an “A”) is boxed and is at the same relative location in both the 5′- and 3′-LTRs. The last base in exon 1 is the last base of the direct repeat. Exon 5 continues beyond the sequence shown. The 4-bp direct repeat (ATTTC), which flanks the integrated provirus, is shaded in dark gray. Numbers given refer to the base numbers in the PTL genomic sequence. PPT, polypurine tract; U3, unique 3′ region found in the retroviral genome; TIR, terminal inverted repeat; R, short repeat region at the ends of the retroviral genome; Poly(A) signal, polyadenylation signal; U5, unique 5′ region found in the retroviral genome; Poly(A) DSE, polyadenylation downstream element; and PBS, tRNA primer-binding site.

Thirteen-lined ground squirrel PTL gene: Comparative genomic analysis. To examine further the possibility that a retrovirus inserted into the promoter region of a single functional PTL gene, a comparison was made between the ground squirrel PTL gene, with the retroviral sequence removed, and the rat, mouse, and human PTL genes. As seen in Fig. 6, the 3′ end of exon 5 in the ground squirrel gene aligned with exon 1 in the rat, mouse, and human PTL genes. As seen in Fig. 6, the 3′ end of exon 5 in the ground squirrel gene aligned with exon 1 in the rat, mouse, and human PTL genes. Other areas of high sequence identity were also found within this region and could represent transcription factor binding sites important for the regulation of this gene in an intact promoter. With the retroviral sequence removed for this comparison, the upstream boundary for exon 5 in Fig. 6 is the retroviral
Fig. 6. Sequence alignment of the thirteen-lined ground squirrel (gsPTL) PTL gene with the proximal promoter regions and 5' portions of the rat (rPTL), mouse (mPTL), and human (hPTL) PTL genes. An asterisk denotes identical bases found in all four sequences. Bases 85–8657 in gsPTL correspond to the retroviral sequence and are not shown here. The retroviral insertion site (ATTC) in gsPTL is shaded in dark gray. At this site, conserved bases in the other three organisms are likewise shaded. Exons are shaded in light gray with the ground squirrel exon number noted in parentheses. The TATA box is underlined. A vertical arrow indicates the start of the coding region. Sequences for rPTL and hPTL were obtained through the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/). The sequence for mPTL was obtained from the Ensembl web site (http://www.ensembl.org/). More details are presented in the MATERIALS AND METHODS.
insertion site (ATTC). This insertion site fell within a region of low sequence identity among the four species. Rat PTL, however, did share three of the four bases in the insertion sequence with the ground squirrel gene. A look at the sequence downstream of exon 5 showed that sequence identity was high near splice site junctions, through portions of the intron, and throughout the next exon.

Because of the low sequence identity surrounding and upstream of the retroviral insertion site, two additional forms of analyses were performed on these sequences. First, individual alignments were made between the ground squirrel PTL gene sequence shown in Fig. 6 and the PTL gene sequences from rat, mouse, and human. Second, 5 kb of sequence located immediately upstream of exon 2 in each of the rat, mouse, and human PTL genes was analyzed with the program RepeatMasker2 (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker; A. F. A. Smit and P. Green, unpublished data) to look for endogenous retrovirus-like elements of the same class as those found in the ground squirrel gene. Overall, the promoter regions of the rat and mouse PTL genes showed the highest percent identities, 63% and 59%, respectively, with their aligned portions of the ground squirrel gene. Upstream of the insertion site, the rat PTL gene showed 37% identity with the ground squirrel gene, whereas mouse was 35% identical. Interestingly, although the human PTL promoter showed the lowest overall percent identity (54%) with the ground squirrel sequence, in the region upstream of the insertion site, percent identity was 43% between the two sequences.

The program RepeatMasker2 identified the retroviral sequence present in the ground squirrel PTL gene as an ERV_class I (data not shown). This classification was consistent with our earlier analysis which placed it in the mammalian type C, or γ-retrovirus, category (37). No ERV_class I elements were found in the 5 kb upstream of the rat, mouse, and human PTL coding regions.

DISCUSSION

In this paper we have demonstrated that a single PTL gene in the thirteen-lined ground squirrel encodes PTL mRNAs isolated from pancreas, heart (2), and WAT (3). Furthermore, we have demonstrated that an endogenous retrovirus of unknown origin, but with high similarity to mammalian type C retroviruses, or γ-retroviruses, is present at the 5′ end of this thirteen-lined ground squirrel PTL gene. We also show that PTL mRNA is detected in 10 different ground squirrel tissues during hibernation, thus indicating that expression of the PTL gene is broader than previously realized.

The presence of retroviral sequence proximal to the PTL gene provides a potential mechanism for directing seasonal expression of PTL in thirteen-lined ground squirrels. Novel expression of a pancreatic enzyme mediated by retroviral insertion has been observed previously (33). Parotid-specific expression of the human salivary amylase gene (AMY1C) is driven by a retroviral-like sequence present in its proximal promoter (33). Similarly, in mouse, androgen-responsiveness of the sex-limited promoter (Slp) gene was conferred by retroviral insertion upstream of its promoter (31). More recently, documentation of endogenous retroviral elements providing alternative promoters or enhancers for neighboring genes has been provided for the pleiotrophin gene (28), the endothelin B receptor and the apolipoprotein C-I genes (19), and the Midl gene in humans (15). It is likely that more such examples of retroviral-mediated gene regulation will be uncovered for the following reasons: 1) ~38.5% and 46% of the mouse and human genomes, respectively, are recognized as having been derived from the insertion of transposable elements, and 2) nearly 10% of mouse and 9% of human insertions are classified as LTR elements (36). For the human genome, if all classes of transposable elements are considered as a whole, then this number could total more than 1,000 genes (14).

Alignment of the ground squirrel PTL gene, after removal of its retroviral sequence, with the promoter regions of the rat, mouse, and human PTL genes illustrated that integration of the provirus occurred just over 200 bases upstream of the original transcription start site. Insertion of the retrovirus occurred in a region of low sequence identity, but disrupted normal promoter function as evidenced by inclusion of what appears to be the original TATA box in the 5′-UTR of the PTL cDNA sequence isolated from heart and pancreas. Although disruptive to normal promoter function, maintenance of the retroviral sequence at this location within the ground squirrel genome suggests that the inserted sequence was not deleterious to overall PTL gene function. Transposable element-derived sequences that are deleterious to gene function are likely to be removed from the genome by selection (14). Jordan et al. (14) presented this hypothesis based on their observation that the percentage of transposable element-derived sequences in human promoters increased as one moved farther upstream from the transcription start site. We propose that insertion of the retrovirus into the promoter region of this ground squirrel gene enabled novel expression of PTL mRNA in a broad range of tissues during hibernation (Fig. 1A). The product of this chimeric mRNA conferred a selective advantage to the organism in the form of low-temperature lipolysis during hibernation (2, 30) and, as a result, has enabled the retroviral sequence to be maintained in the ground squirrel lineage.

Analysis of the retroviral sequence contained in the thirteen-lined ground squirrel PTL gene suggests that the insertion event occurred in relatively recent history. The four bases (ATTC) that represented the original target site, and that were duplicated upon integration of the provirus, are perfectly conserved in the ground squirrel genomic sequence. The gag, pol, and env genes encoded ORFs that are also largely intact (Fig. 4). Along these lines, the longest, uninterrupted ORF was present in the env gene and represented over 500 amino acids. In addition, the LTRs were less than 1% divergent. Working under the assumption that the LTRs are identical at the time of insertion, several researchers have used the percent LTR divergence to date the time of insertion (13, 18, 25). To provide an estimate for the time of our ground squirrel retroviral insertion, we use the synonymous nucleotide substitution rate [0.013 substitutions per site per million years (Myr) or 1.3%/Myr] for the nuclear gene lecithin:cholesterol acyltransferase for the Marmota_Sciurus dichotomy (26). This rate returns an insertion date of ~300,000 years ago (0.8%/1.3%, divided by two for divergence from a common sequence) placing it well within the Spermophilus lineage (20).
ACKNOWLEDGMENTS

We thank L. B. Mamo and J. Thornbury for technical assistance, J. Guhaniyogi for creation of the pancreas cDNA library, R. Swanstrom for assistance in identifying the LTRs, and S. Curtis for the use of her laboratory. We also thank G. Wray, D. Mager, and J. Mercer for helpful discussions.

GRANTS

This work was supported by US Army Research Office Grant DAAD19-01-1-0014 and Augmentation Awards for Science and Engineering Training DAAS5-97-1-0175 and by North Carolina Biotechnology Center Grant 9805-ARG-0038.

REFERENCES


