Nucleic Acids Research, 1993, Vol. 21, No. 1 **135–143**

Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene

Anita E.Feuchter-Murthy+, J.Douglas Freeman and Dixie L.Mager*

Terry Fox Laboratory, B.C. Cancer Agency and Department of Medical Genetics, University of British Columbia, Vancouver, B.C. V5Z 1L3, Canada

Received September 1, 1992; Revised and Accepted November 5, 1992

ABSTRACT

As part of an investigation into the effects of endogenous retroviruses on adjacent genes, we have isolated a cDNA clone derived from the human teratocarcinoma cell line NTera2D1 representing a chimeric transcript in which an endogenous retrovirus-like element, RTVL-H, has been spliced to downstream cellular sequences. The 5′ terminus of this clone, termed AF-5, occurs one bp downstream of the predicted transcriptional start site in the RTVL-H long terminal repeat (LTR). AF-5 contains an open reading frame of 689 amino acids beginning within RTVL-H sequences that has two domains of homology with phospholipase A_2 (PLA_2). These domains, of approximately 120 amino acids each, are 30–38% identical to secreted PLA_2s and contain sequence features of both group I and II enzymes. The corresponding AF-5 transcript is 2.5 kb and is derived from a single copy novel gene termed PLA2L. Southern analysis indicates that the RTVL-H element is normally present in human DNA upstream of the PLA2L gene. RTVL-H/PLA2L chimeric transcripts were detected in two independent teratocarcinoma cell lines but not in several other cell lines or primary human tissues. Characterization of additional cDNA clones and PCR analysis indicates that multiple RTVL-H/PLA2L alternatively spliced transcripts are expressed. No evidence has been found for transcription from a non-LTR promoter. These findings strongly suggest that the endogenous LTR promotes expression of the human PLA2L gene in teratocarcinoma cells.

INTRODUCTION

Retrotransposons and endogenous retroviruses, like other transposable elements, are causative agents of genetic change residing within otherwise stable genomes. Their insertion into the genome can physically disrupt and inactivate cellular genes or can alter expression of adjacent genes due to effects imposed by the transcriptional control sequences carried by the LTRs of the element. The mutational capabilities of retrotransposons in organisms such as *Drosophila* and yeast are well documented (1,2) and numerous examples of inactivation or altered regulation of genes by murine retrotransposons or endogenous retroviruses have also been reported (reviewed in 3,4). Many of the murine examples have been found in somatic cells (i.e. cell lines or tumors) where insertion near a growth regulatory gene contributes to a measurable phenotype such as altered growth or malignant transformation (5,6). However, germ line events in which insertion of an element altered or abolished expression of an adjacent gene have also been reported in mice (7–9).

It has become evident in the last decade that the genomes of humans and other primates also harbor thousands of LTR containing elements that resemble retrotransposons or endogenous retroviruses (for reviews see 10,11). The abundance of these sequences raises the possibility that some are involved in adjacent gene expression, either because of recent rearrangements or because they have evolved to play that role. As one way of assessing the effects that endogenous retroviruses or related elements may have in humans, we have undertaken a study to identify and characterize heterologous cellular transcripts or genes that utilize an endogenous LTR as a transcriptional promoter. For this study we have focussed on the RTVL-H family of elements (12,13) for two primary reasons: First, RTVL-H elements and related solitary LTRs are dispersed and present in high copy numbers (approximately 1000 each) in the human genome (14). Second, RTVL-H LTRs can drive expression of a reporter gene (15) and promote high levels of endogenous RTVL-H transcripts in normal placenta (16,17) and in various transformed human cells, most notably teratocarcinoma cell lines (16). Thus there is a large pool of potentially functional RTVL-H LTRs in the genome that could act to control adjacent cellular gene expression.

In a previous report, we used a differential hybridization screening strategy to isolate several RTVL-H-cellular chimeric cDNA clones from an NTera2D1 cDNA library having a structure consistent with transcriptional initiation from within a solitary LTR or the 3′ LTR of an intact element (18). We have now modified this screening method to detect cDNAs derived

* To whom correspondence should be addressed at: Terry Fox Laboratory, B.C. Cancer Agency, 601 West 10th Avenue, Vancouver, B.C. V5Z 1L3, Canada

+ Present address: Molecular Neurogenetics Laboratory, Massachusetts General Hospital East, Charlestown, MA 02129, USA

from transcripts initiating within the 5′ LTR and splicing into a downstream cellular exon. Here we describe NTera2D1 cDNA clones representing LTR initiated transcripts that have spliced into a novel cellular gene with homology to phospholipase A_2. Results presented in this report suggest that the expression of this gene in teratocarcinoma cells is driven by the RTVL-H LTR.

MATERIALS AND METHODS

Library screening

A λgt10 cDNA library constructed from polyA$^+$ RNA from NTera2D1 cells (19) was provided by Drs. Maxine Singer and Ronald Thayer (20). The library was differentially screened using standard procedures (21) with the probes shown in figure 1a. The U3 and U5 specific probes were described previously (18). Probe 6, a 265 bp NcoI/OxaN1 fragment mapping 140 bp 3′ to the RTVL-H splice donor site, and probe 7, a 210 bp BglII/HindIII fragment mapping just 3′ to an observed cluster of RTVL-H splice acceptor sites, were isolated from RTVL- H2 (13). All probes were radioactively labelled by the random primer method (22). Hybridization conditions were: 65°C in 6×SSC (1×SSC is 0.15 M NaCl, 0.015 M Na citrate), 1×Denhardt's solution (1×Denhardt's solution is 0.02% ficoll, 0.02% bovine serum albumin, 0.02% polyvinyl pyrrolidone), 1% sodium dodecyl sulphate (SDS). Post hybridization washes were at 65°C in 3×SSC, 1% SDS. Phage clones of interest were plaque purified and cDNA inserts were subcloned into plasmid vectors.

Southern and Northern analysis

For Southern blot analysis, DNA was digested with restriction enzymes, electrophoresed in 0.8–1% agarose gels, and transferred to Zetaprobe (Biorad) membranes by standard methods (21). Genomic Southerns used 4 μg DNA per lane and were hybridized as described previously (18). For Northern analysis, total cellular RNA was isolated by guanidinium isothiocyanate preparation (23). RNA was electrophoresed in 1% agarose wick gels containing 0.66 M formaldehyde, using 2×MOPS buffer (0.04 M 3-[N-morpholino] propanesulphonic acid, 0.01M sodium acetate, 2mM EDTA pH 7.0) at 1V/cm and transferred to Zetaprobe membranes in 10×SSC. Hybridizations were at 65°C in 5×Denhardt's, 1.5×SSPE (0.27M NaCl, 15mM Na H2PO4, 1.5mM EDTA), 1% SDS in the presence of 500 μg/ml denatured salmon sperm DNA and denatured probe at 2–6×106 cpm/ml. Final post hybridization washes were at 50°C in 0.1×SSC, 0.5% SDS. Probes used are described in the appropriate figure legend.

DNA sequencing and computer analysis

Fragments were subcloned and sequenced using a modification of the dideoxy chain termination method (24). The sequence shown in figure 2 was determined on both DNA strands or at least twice on the same strand. Sequences were analyzed using software provided by the Genetics Computer Group (GCG) (25).

Polymerase chain reaction (PCR) analysis

PolyA$^+$ RNA from different sources was reversed transcribed using random hexamers as primers and Moloney-murine leukemia virus reverse transcriptase. Conditions for the 30 μl reaction volume were: 2 μg RNA, 100 ng random hexamers, 200 units reverse transcriptase, 30 units RNasin, 830 μM dNTPs in 50 mM Tris-HCl pH8, 60 mM KCl, 3mM $MgCl_2$, 10 mM dithiothreitol, 100 μg/ml bovine serum albumin. The reaction was carried out for 1 hour at 40°C, followed by an inactivation step of 5 min. at 95°C (26). One sixth of the cDNA reaction was used in PCR with various primers shown in Figures 2 and 5c. PCR conditions per 50 μl reaction volume were: 20 mM Tris pH 7.5, 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% bovine serum albumin, 250 μM dNTPs, 1.25 units Taq DNA polymerase, and 20 picomoles each primer. PCR cycle parameters were: 2 min, 95°C, 1×; 30 sec, 94°C, 30 sec, 50°C, 1 min, 72°C, 30×; 2 min, 72°C, 1× (27). Portions of the reactions were electrophoresed and gels were either blotted and hybridized or fragments were excised, purified using Geneclean (Bio 101) and subcloned into plasmid vectors for sequencing.

RESULTS AND DISCUSSION

Identification of a chimeric spliced RTVL-H/cellular cDNA clone

Our laboratory has previously identified a variety of spliced RTVL-H transcripts in NTera2D1 cells that utilize a single 5′ splice donor site located 150 bp downstream of the 5′ LTR and a cluster of splice acceptor sites found 1.8–2 kb 3′ to the donor site (16). The finding of these spliced transcripts suggests that chimeric transcripts promoted by an LTR in which the RTVL-H element has been spliced to an exon of a downstream gene may exist in teratocarcinoma cells (see figure 1a). Such splicing events could occur via deletion or mutation of the RTVL-H splice acceptor sequences, or conversely, by mutation involving sequences controlling splicing of the downstream gene. The only RTVL-H specific sequences that chimeric transcripts of this type should contain would be the LTR subregions R and U5, which by definition lie 3′ to the transcriptional initiation site (28), and the internal region 5′ to the splice donor site (see figure 1a). To identify such transcripts, we screened an NTera2D1 cDNA library (250,000 phage) for cDNA clones that hybridized with

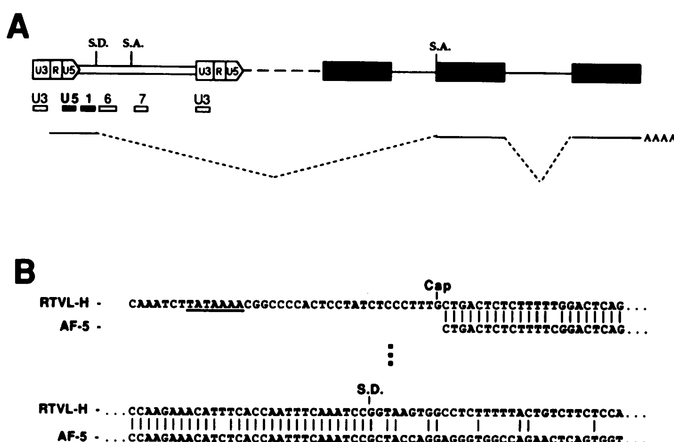

Figure 1. A) Strategy for isolating chimeric spliced cDNAs. An endogenous retroviral element, RTVL-H, with its 2 LTRs (subdivided into U3, R & U5) is shown 5′ to a hypothetical gene containing 3 exons (striped boxes). A possible transcript initiating in the 5′ LTR and splicing into the gene is illustrated. An NTera2D1 cDNA library was screened with the probes shown as white and black boxes and clones positive only for the U5 probe and internal probe 1 were chosen for further analysis. S.D., splice donor site; S.A., splice acceptor site. B) The 5′ terminal sequence of one clone, AF-5, compared to the consensus RTVL-H element. The clone begins 1 bp 3′ to the LTR CAP site and homology with RTVL-H extends as expected to the splice donor site. The 3 dots indicate 200 bp of homology with RTVL-H.

an RTVL-H internal probe 5′ to the splice donor site (probe 1 in figure 1a) and with a U5 specific probe, but which did not hybridize with a U3 specific probe or with probes 6 or 7, mapping downstream of the RTVL-H splice donor and splice acceptor sites respectively (figure 1a).

We identified several phage clones that met this criteria, however, most of these clones were subsequently shown to hybridize very weakly with probe 7 and were not characterized further. One 2.4 kb phage clone that hybridized only to probe 1 and the U5 probe was purified and sequenced. Figure 1b is a comparison of the 5′ region of this clone, termed AF-5, and relevant regions of an RTVL-H consensus sequence. This figure illustrates that the clone begins in an RTVL-H LTR one bp downstream of the expected transcriptional initiation site. The homology with RTVL-H extends to the 5′ splice donor site and terminates precisely at the G residue shown, as expected in a spliced transcript. This structure strongly suggests that AF-5 is, indeed, the product of a chimeric splicing event between an RTVL-H element and an unrelated exon.

Clone AF-5 is derived from a novel phospholipase A_2 related gene

The complete nucleotide sequence of clone AF-5 and the deduced amino acid sequence (689 amino acids) encoded by its open reading frame (ORF) are shown in figure 2. The ORF shown begins at an ATG within the LTR sequence (position 101) and continues in frame into the non-RTVL-H AF-5 sequence where it terminates at position 2167. The ATG within the LTR is not in a favorable context for translation initiation (29) but a more favorable one occurs at position 416 and is bracketed. Clone AF-5 contains a 253 bp 3′ noncoding region, and is not polyadenylated, suggesting that this clone is a 3′ truncation of a larger transcript. A search of the PIR database for proteins with homology to the predicted amino acid sequence of AF-5 revealed two PLA_2 like domains within the putative AF-5 protein (underlined in figure 2a). Further Genbank and PIR database searches revealed no other significant similarities to known sequences. Because the gene from which the AF_{-5} cDNA was derived has similarity to PLA2, we have elected to designate it PLA2L for $\underline{PLA_2}$ $\underline{l}$ike gene.

PLA_2s are a diverse family of enzymes that hydrolyze the sn-2 fatty acyl ester bond of phospholipids to produce free fatty acids and lysophospholipids (30). PLA_2s are abundant in mammalian pancreatic juices and in the venom of snakes and bees, where they play a digestive role (30). They also occur in trace amounts in many cell types and perform a variety of biological functions, such as controlling phospholipid metabolism, regulating membrane fluidity, and initiating arachidonic acid release during the onset of inflammation (31). Interestingly, there is some evidence that PLA_2 plays a significant role in the pathogenesis of chronic inflammatory disorders such as rheumatoid arthritis and psoriasis (32). Extracellular forms of PLA_2 from a variety of sources have

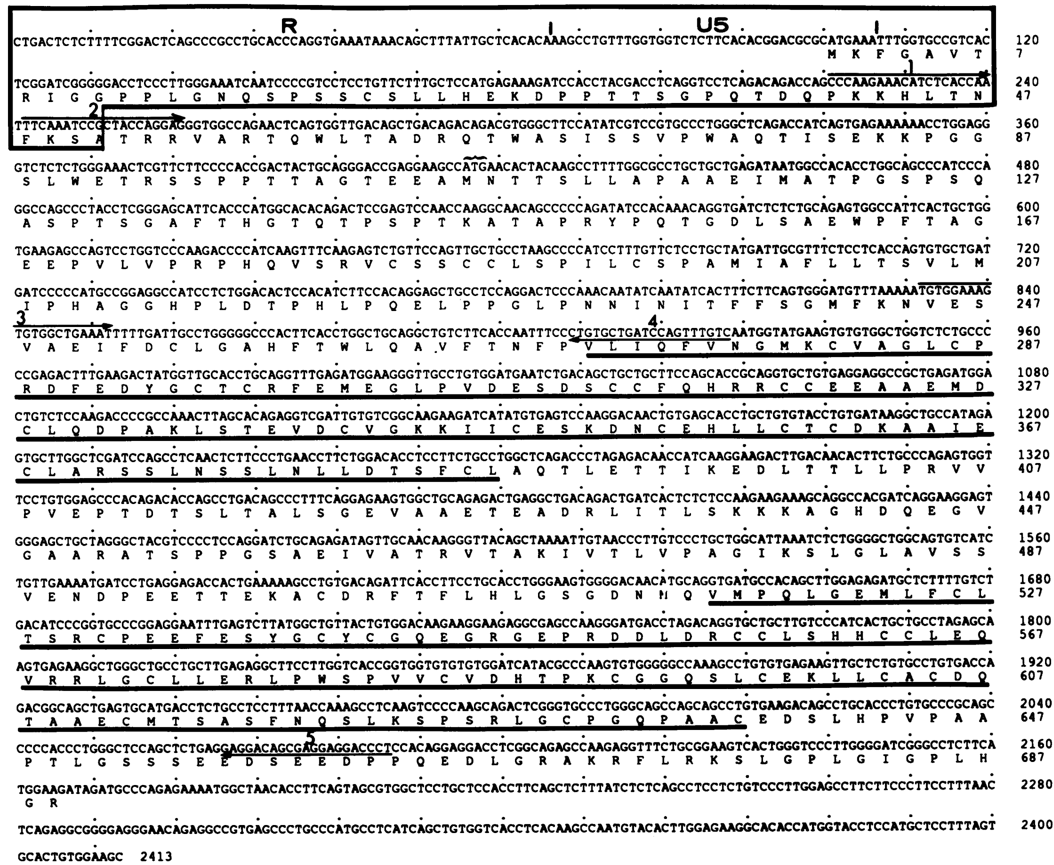

Figure 2. DNA and predicted amino acid sequence of cDNA clone AF-5. RTVL-H sequences are boxed with the R & U5 subregions of the LTR indicated. An ORF of 689 a.a. begins within the LTR at position 101 but an ATG in a more favorable context for translation initiation occurs at position 416 and is bracketed. The two domains of homology with PLA_2 are underlined. Numbered arrows shown above and below the sequence indicate the 5 PCR primers used in Figure 6. Primers 4 and 5 are the reverse complement of the sequence shown.

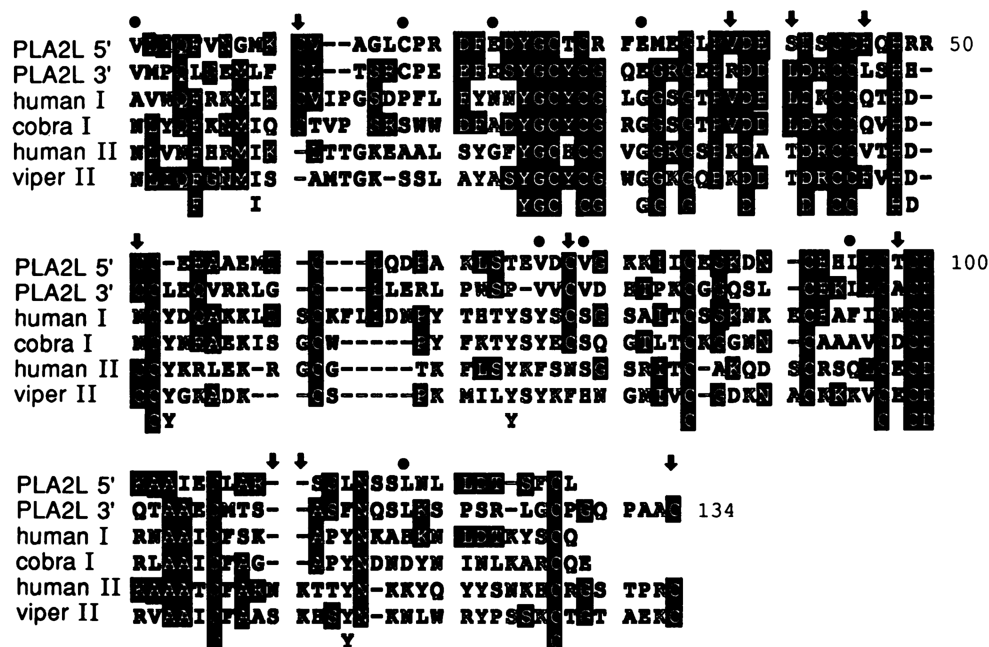

Figure 3. Amino acid comparison of the two PLA2L domains to other PLA_2 sequens. Human group I—pancreatic (50); cobra group I—*naja naja atra* (51); human group II—non pancreatic (52); viper group II—*bitis caudalis* (53). The 26 amino acids shown below are conserved in all active group I and II PLA_2s (39). Black boxes highlight identities of the PLA2L domains to other PLA_2 s. Dots show residues found in both PLA2L domains but not in other PLA_2s. Arrows indicate positions that distinguish group I and II.

been characterized and genes encoding several have been cloned and sequenced (33–36). However, none of the PLA_2 genes or cDNA clones examined have a repeated structure similar to PLA2L. Indeed, the structure of PLA2L appears as if two complete PLA_2 related genes are contained within it. Secreted PLA_2s are small (120–145 aa) rigid proteins in which the key active site residues and the alignment of cysteines are highly conserved. This alignment of C residues is strictly maintained in the PLA2L domains (see figure 3) suggesting that a PLA2L encoded protein would also be secreted. In addition a 23 amino acid sequence beginning with a potential initiator M at amino acid position 106 (bracketed in figure 2) fits predictions for a signal sequence reasonably well (37).

Secreted PLA_2s have been classified as belonging to group I or group II based on the positioning of certain C residues (38,39). Figure 3 is an amino acid sequence comparison of the two PLA2L repeats with human pancreatic PLA2 (group I), human nonpancreatic PLA_2 (group II) and two other randomly chosen PLA_2s from snake venom. While homology extends throughout, the most striking similarity of both repeats to the PLA_2s is in the calcium binding loop (residues 25–49) and in the absolute conservation of cysteines. Below the alignment is shown 26 conserved amino acids that have been found in all active group I and II PLA_2s (39). The PLA2L domains contain most of these amino acids but differ in certain positions, such as D^{49} which is thought to be a critical residue for enzymatic function (40). These domains have not yet been tested to determine if one or both encode PLA_2 activity. However, the maintenance of the PLA2L long open reading frame, despite the apparent ancient origin of the PLA_2 duplication (see below), suggests that it is functionally important.

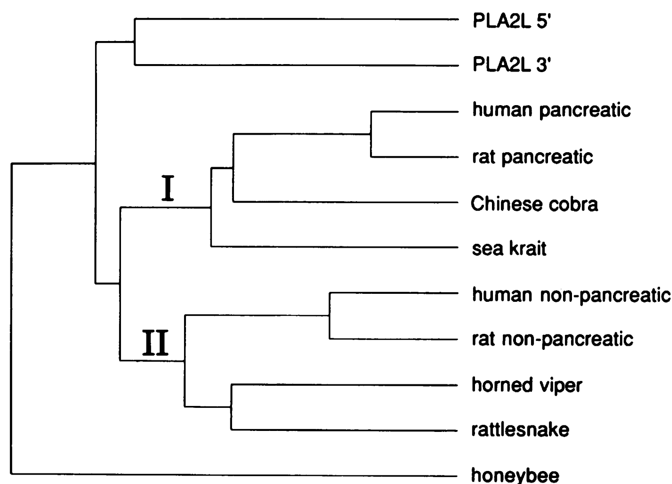

Figure 4. Dendogram of a multiple amino acid sequence comparison of PLA_2s. Lengths of the horizontal branches are proportional to the differences between sequences. Sequences included are those shown in figure 3 and the following: rat pancreatic (54); sea krait (55); rat non-pancreatic (56); rattlesnake (57); honeybee (58).

The PLA2L domains are distinct from group I and II PLA_2s

As mentioned above, secreted PLA_2s have been classified into group I or group II based on sequence differences, in particular on the positioning of certain C residues (38,39). In addition, from an alignment of 40 PLA_2 sequences (39), it is clear that there are several other amino acid positions that distinguish group I and II PLA_2s. These are marked by arrows in figure 3. It is

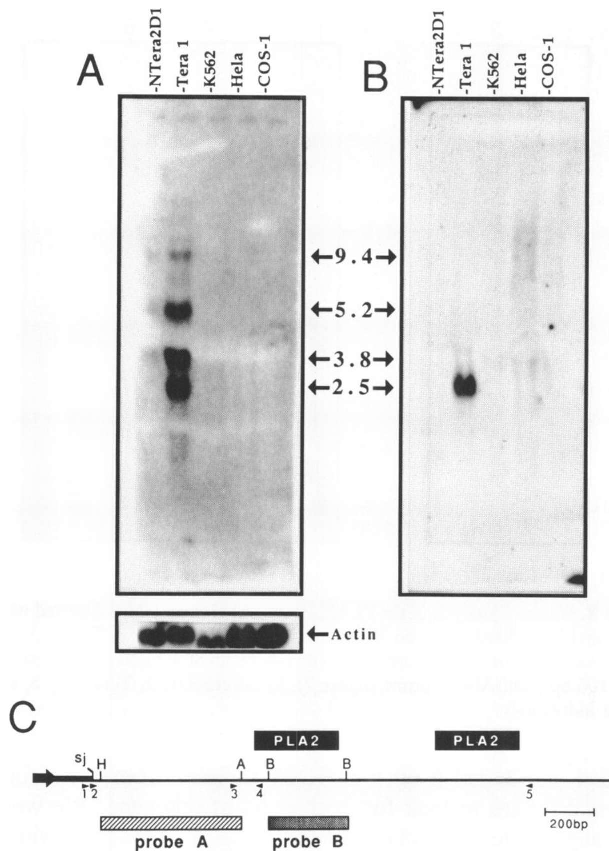

Figure 5. A) Northern analysis of total RNA from several cell lines using probe A. NTera2D1 and Tera1—human teratocarcinoma; K562—human leukemia; Hela—human cervical carcinoma; Cos1—monkey SV40 transformed fibroblast. B) Rehybridization to probe B. (Rehybridization to a β-actin probe is shown below panel A). C) Map of AF-5, showing the RTVL-H/cellular splice junction (sj), the PLA_2 homologous regions and restriction sites used for the isolation of probe A, a 587 bp HincII/ApaI fragment, and B, a 318 bp BbsI fragment. H, HincII; A, ApaI; B, BbsI. Arrows labeled 1–5 show locations of PCR primers used in Figure 6.

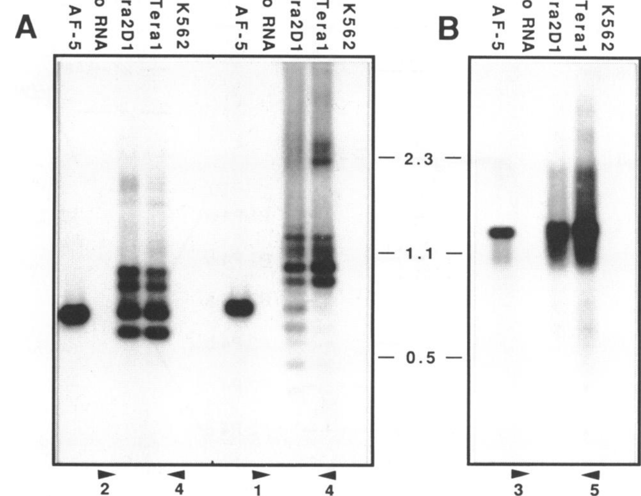

Figure 6. PCR of cDNA synthesized from $polyA^+$ RNA from NTera2D1, Tera1 or K562 cells. AF-5 cDNA is included as a positive control. Primers used for each reaction are shown below each panel. (See figures 2 & 5c for locations of primers.) Panel A was hybridized to probe A and panel B to probe B.

apparent that the two PLA2L domains have features characteristic of both group I and II (figure 3). Both domains have the group I specific C^{11} and C^{78} and a two amino acid gap at positions 110–111. However, both also have C^{51} which is group II specific. The 5′ domain has F^{46} characteristic of group II but V^{38} and a lack of C^{134} which are both features of group I. The 3′ domain has group I specific L^{41} and A^{98} but possesses C^{134} which is a hallmark of group II enzymes. Both domains also have a C at position 18 and several other residues in common (marked in figure 3) that do not occur in any other PLA_2 aligned by Davidson and Dennis (39). Thus the PLA2L domains cannot be definitively placed into either known group.

This conclusion is supported by the results of a global multiple sequence comparison of the PLA2L domains and several mammalian and snake PLA_2s (figure 4). The dendogram shown in figure 4 was generated by the GCG program pileup and is not a true phylogenetic tree. However, the lengths of the horizontal branches are proportional to the number of sequence differences. It can be seen that such an analysis places the PLA2L domains on a separate branch. Interestingly, assuming that the PLA2L domains were generated by a duplication event, they have diverged substantially from one another. At the protein level they are 37% identical and at the nucleotide level they are only 50% identical. There is no detectable homology at the nucleotide or protein level in the regions surrounding these domains indicating that, if the length of the original duplication extended beyond the PLA_2 homologous regions, sequence divergence has erased any record of it. The extreme age of the duplication and the fact that the two domains have features of both group I and II PLA_2s raises the possibility that the PLA2L gene may represent a precursor of the genes encoding the two groups.

Expression of PLA2L related transcripts

To analyze the expression of PLA2L transcripts, Northern analysis using total cellular RNA from NTera2D1, Tera1 (an unrelated human teratocarcinoma cell line), K562, Hela and Cos-1 cells was performed using probe A indicated in figure 5c. The results are shown in figure 5a. Of the lines tested, expression of transcripts hybridizing to this probe was detectable only in the teratocarcinoma cell lines, with much higher levels in Tera1 cells where transcripts of approximately 9.4, 5.2, 3.8, and 2.5 kb were observed. The Northern blot shown in the right panel of figure 5a was next rehybridized with probe B (see figure 5c) specific for the 5′ PLA_2-like domain. Figure 5b shows that this probe hybridizes only to the 2.5 kb transcript in Tera1 cells. These results suggest that the 2.4 kb AF-5 cDNA is a 3′ truncated clone of the 2.5 kb transcript detected by both probes. The structure of the larger transcripts that hybridize only to probe A is also of interest and will be addressed below.

To confirm that RTVL-H/cellular chimeric transcripts are expressed in teratocarcinoma cells, RNA-PCR analysis was performed using $polyA^+$ RNA from NTera2D1, Tera1 and K562 cells with primers shown in figures 2 and 5c. The position of primer 1 is within the RTVL-H sequence just 5′ to the splice donor site and primer 2 spans the RTVL-H-cellular junction. Each of these primers was used in conjunction with primer 4 (located at the 5′ end of the first PLA_2 like domain) to amplify the 5′

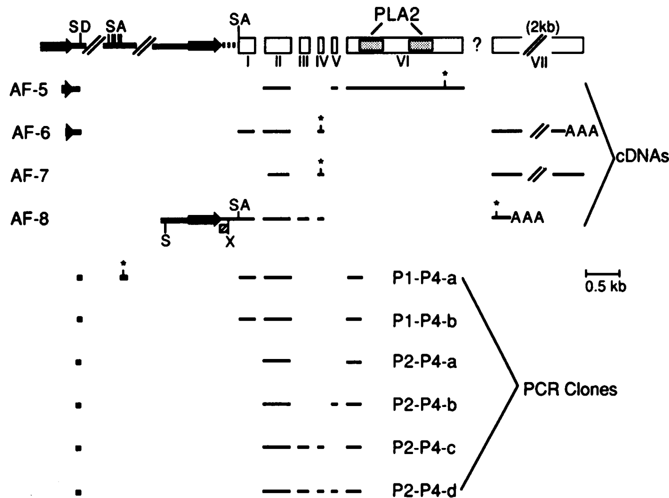

Figure 7. Structure of cDNA clones and PCR products. The top line shows the RTVL-H element with its LTRs and locations of known splice donor (SD) and splice acceptor (SA) sites. Boxes represent exons or groups of exons in the PLA2L gene. The ? denotes that the relative genomic order of the final 2 groups of exons is not known. Structures of the AF-5 cDNA and 3 other NTera2D1 cDNAs are shown as are structures of several PCR clones generated using primers 1 and 4 or 2 and 4. PCR clone P1-P4-a was derived from Tera1 while the rest were from NTera2D1. Thick lines show RTVL-H sequences and thin lines are non-RTVL-H sequences. Asterisks are stop codons. A potential splice acceptor site identified in the read-through cDNA, AF-8, is indicated. The 100 bp NdeI/AlwNI probe (striped box) and the StyI (S) and XmnI (X) restriction sites used for Southern analysis are shown below clone AF-8.

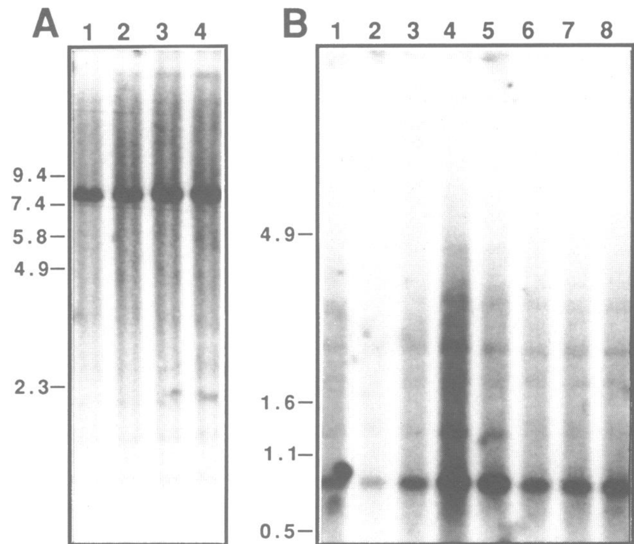

Figure 8. Southern analysis of the PLA2L locus. A) Human DNAs digested with EcoRI and hybridized to probe B (figure 5c). 1, NTera2D1; 2, Tera1; 3–4, two normal individuals. B) Human DNAs digested with StyI/XmnI and hybridized to the 100 bp NdeI/AlwNI probe (figure 7). 1, NTera2D1; 2, Tera1; 3–8, six normal individuals.

region of the transcript. Figure 6a shows the products of these amplification reactions hybridized to probe A. Both sets of primers amplify fragments that are the same size as in clone AF-5 in the two teratocarcinoma cell lines. Interestingly, however, several other hybridizing fragments of different sizes were also observed. No hybridizing PCR fragments were detectable from K562 poly A^+ RNA. The cluster of fragments seen in the NTera2D1 and Tera1 reactions suggest that a set of RNAs that differ slightly in size are expressed. These results also indicate that the patterns of expression in the two independent teratocarcinoma cell lines are very similar if not identical although the Northern results indicated that the overall level of expression is higher in Tera1.

PCR was also performed using primers 3 and 5 that bracket the PLA_2 related domains. In this case, a single hybridizing amplified fragment of the same size as in the cDNA AF-5 was observed in the teratocarcinoma cell lines (figure 6b). No equivalent fragment was amplified from K562 mRNA which agrees with the Northern analysis indicating that related transcripts are not expressed in these cells.

To determine if PLA2L is expressed in normal human tissues, we hybridized probe B to a Northern blot of polyA^+ RNAs from 8 primary adult tissues (pancreas, kidney, muscle, liver, lung, placenta, brain, heart) obtained from a commercial source (Clonetech). No hybridizing bands were observed indicating that expression of this gene is low or absent in these tissues. In previous work we have shown by Northern analysis that RTVL-H sequences in general are expressed in placental amnion and chorion while expression was not detectable in decidua or villus (16). Therefore, RNA-PCR was also performed with primer pairs 1 and 4 and 3 and 5 on total RNA samples from these four placental tissues to look for expression of this gene. We were generally unable to detect amplified transcripts with either primer set even after hybridization to the appropriate probe and autoradiography for several days. However, in two instances (primers 1 and 4 using villus cDNA and primers 3 and 5 using decidua cDNA) very faint hybridizing fragments of the expected size were observed (data not shown). It is therefore possible that the gene is expressed at very low levels in some placental tissues but since we did not detect transcripts in the same tissue using both primer pairs, the significance of this finding is as yet unclear. Thus, to date we have found definitive evidence for PLA2L gene expression only in the teratocarcinoma cell lines.

Alternative splicing of the PLA2L gene

One explanation for the amplification of different sized fragments as shown in figure 6a is that expression of this locus may involve alternative splicing. The observation that probe A detects transcripts on a Northern blot that are not detected by the probe within the 5′ PLA_2-like domain (figure 5a & b) also raises the possibility of alternative splicing. To investigate this possibility, we rescreened the NTera2D1 cDNA library with probe A and isolated three additional cDNA clones. Hybridization analysis of these clones revealed that two of them contained RTVL-H sequences and that none of the three contained the PLA_2-like domains. Sequencing of these clones and comparisons to the AF-5 cDNA sequence revealed evidence for alternative splicing as diagrammed in figure 7 and as will be discussed below. In addition to characterizing these cDNA clones, we also subcloned and sequenced several PCR generated fragments from the reactions shown in figure 6a. The structures of these fragments are compared to each other and to the cDNA clones in figure 7. This comparison also indicates that alternative splicing occurs and is the explanation for the heterogeneity in size of the PCR fragments.

The top line in figure 7 shows the genomic locus with the RTVL-H element and the PLA2L exons or blocks of exons that we can deduce from comparisons of the different cDNA sequences. All clones have a 390 bp region in common shown as box II in figure 7. This segment is likely a single exon or group of exons that is always found together. Downstream of this segment is a region a three small putative exons which are found in various combinations in the different sequences. Following the second of these exons (box IV), the three additional cDNA clones, AF-6, 7 and 8 differ completely from AF-5 and the PCR clones as diagrammed in the figure. They do not contain the PLA_2-like domains and instead splice into a different region. This region, labeled VII, is shown 3′ to region VI containing the PLA_2-like segments but we cannot rule out the possibility that the genomic order of regions VI and VII may actually be reversed. The 5′ terminus of clone AF-6 occurs within the RTVL-H 5′ LTR 87 bp downstream of the 5′ terminus of AF-5. The transcript corresponding to AF-6 utilized the RTVL-H splice donor site but spliced into an upstream exon (box I) that is not present in clone AF-5. The 5′ end of clone AF-7 terminates within the 390 bp segment (box II).

Clone AF-8 is of interest because its 5′ terminus occurs within the interior of the RTVL-H element as shown. Thus this clone most probably represents a 5′ truncated transcript that initiated within the 5′ LTR but then failed to use the RTVL-H splice donor site and was not polyadenylated in the 3′ LTR. Sequence analysis of this clone revealed a potential splice acceptor site CCTCCCTGAAG/G located 270 bp downstream of the 3′ LTR and corresponding to the beginning of the exon labeled I present in cDNA clone AF-6 and PCR clones P1-P4-a and b. The availability of clone AF-8 has allowed us to isolate a probe which should be adjacent to the RTVL-H element in genomic DNA (striped box in figure 7) that we have used to examine the genomic arrangement of the locus in different individuals (see below).

The termini of the PCR clones is of course fixed by the primers used to generate them but alternative splicing is again apparent. Figure 6a (left panel) shows that four primary fragments were amplified from teratocarcinoma RNA using the RTVL-H/cellular junction primer 2 and primer 4. Sequencing of PCR clones P2-P4-a-d which correspond to the four fragments revealed that their size differences are due to the variable presence of exons III, IV and V (figure 7). Use of primer 1 which lies entirely within RTVL-H gives even a more complex pattern (figure 6a, right panel). Thus far, two clones generated using primers 1 and 4 have been sequenced, P1-P4-a and b. They both include exon I, not found in AF-5. In addition, the transcript corresponding to clone P1-P4-a utilized known splice donor and splice acceptor sites within the RTVL- H element but then used a previously unknown cryptic splice donor site located 86 bp 3′ to the RTVL-H splice acceptor site. The inclusion of the short internal RTVL-H segment would add further size complexity to the resulting transcripts. Although we have not yet sequenced PCR clones representing all of the fragments shown in figure 6a, it is likely that they represent additional splicing variations localized to the 5′ region of the PLA2L gene. Interestingly, since primers 3 and 5 which flank the PLA_2 domains amplify a single fragment, it appears that variations due to splicing do not occur in this region of the gene. However, the Northern results (figure 5) and the structures of cDNA clones AF-6,7 and 8 indicate that alternative splicing does produce transcripts that lack the PLA_2 like domains altogether.

The PCR that generated the clones diagrammed in figure 7 was not quantitative so it is not known which spliced forms are most commonly produced in the teratocarcinoma cell lines. The Northern results using probe B (figure 5) indicate that one or more RNAs of approximately 2.5 kb are most abundant in Tera1 cells. Since exons I, III, IV and V are small in size (249, 154, 83 and 93 bp respectively), the 2.5 kb band could well be a mixture of different spliced forms. Interestingly, while PCR clones P1-P4-b and P2-P4-a-d all have an ORF throughout, P1-P4-a contains an in frame termination codon in its short internal RTVL-H segment (figure 7). This suggests that at least some of the alternatively spliced forms would be unable to encode a putative PLA2L protein. The structures of cDNAs AF-6, 7 and 8 are further evidence of abberant splicing. None of them contain the PLA_2 related regions and AF-6 and 7 include exon IV but not exon III resulting in the fusion of a different reading frame and a termination codon in exon IV. It is quite possible that the presence and use of the RTVL-H promoter and splicing signals induces downstream splicing variations that would not occur if a non-LTR 'cellular' promoter of the PLA2L gene was being used.

Southern analysis of the PLA2L locus

To determine if PLA2L homologous sequences are present in a single or multiple copies in the genome, Southern blot analysis of several human DNA samples was first performed with probe B indicated in figure 5c. Figure 8a shows that this probe hybridizes to a single EcoRI fragment of 8.0 kb suggesting that the PLA2L locus is unique. To ascertain if the RTVL-H element is normally present upstream of this locus, we next performed genomic Southern analysis using a 100 bp NdeI/AlwNI probe (striped box in figure 7) located between the 3′ LTR and the splice acceptor site in clone AF-8. As mentioned above, this fragment should be adjacent to the RTVL-H element in genomic DNA. This probe should hybridize to a 900 bp StyI/XmnI genomic fragment if the RTVL-H element is present (see figure 7). Figure 8b shows that a StyI/XmnI fragment of the expected size was detected in DNA from NTera2D1 and Tera1 and from all 6 unrelated normal human DNAs tested. This result indicates that the RTVL-H element is a normal part of the PLA2L gene locus in humans.

Involvement of the RTVL-H element in PLA2L gene expression

Several observations suggest that the RTVL-H LTR promotes expression of the PLA2L gene in teratocarcinoma cells. First, both of the cDNA clones that were isolated using a non-RTVL-H PLA2L region probe and that are sufficiently long to assess (clones AF-6 and 8) begin within RTVL-H sequences. Second, the 5′ termini of clones AF-5 and 6 are consistent with promotion from within the 5′ LTR and splicing into the PLA2L gene. Third, from the sequence of the 3′ LTR in clone AF-8, we have determined that the LTRs of this RTVL-H element belong to subtype I, the most common of the three LTR subtypes we have distinguished on the basis of sequence differences (41). Northern analysis using subtype specific probes has indicated that expression of type I LTRs occurs mainly in teratocarcinoma cell lines, with highest levels in Tera1 cells, whereas the other subtypes have somewhat different patterns of expression (D. Wilkinson and D. Mager, unpublished data). Therefore, the apparent restriction of PLA2L gene expression to the

teratocarcinoma cell lines, again with higher levels in Tera1 cells, is consistent with transcriptional regulation by the RTVL-H LTR. Fourth, although the structure of clone AF-8 suggests that the RTVL-H insertion occurred within an intron, we have found no evidence for utilization of a non-LTR upstream promoter. Such a promoter could well exist but its function may be suppressed by proximity of the LTR promoter and the latter's activation in teratocarcinoma cells. In this regard, it would be of interest to analyze the structure and expression of the PLA2L locus in a non-primate species since the RTVL-H family is specific to primates. In addition, while this RTVL-H element appears to be fixed in homo sapiens, it would be of interest to examine this locus in other primates to determine the evolutionary age of the insertion.

General role of human retrotransposons in genetic variation

We have shown here that an endogenous LTR appears to play a role in the transcriptional regulation of a novel PLA_2 related gene in humans. The possible biological significance of this occurance will be interesting to explore once more is known about the function of the PLA2L gene. While this study and our previous one (18) represent directed screens for transcripts promoted by LTRs, they are not the first reports that implicate human retrotransposons in the control of adjacent genes. In searching for genes specifically expressed in a cell line derived from a prostate metastasis, Liu and Abraham isolated a cDNA representing a spliced transcript between an RTVL-H element and a downstream cellular gene (42). In that case, splicing occurred from the same RTVL-H splice donor site as reported here to the second exon of the calbindin gene. Furthermore, there is evidence for protein expression from the RTVL-H/calbindin ORF initiating at the same ATG in the LTR that begins the RTVL-H/PLA2L ORF shown in figure 2. It has been suggested that RTVL-H induced deregulation of the calbindin gene may have contributed to the metastatic phenotype of the cell line (42). However, it is unknown whether the RTVL-H element is normally present upstream of the calbindin gene or whether it is a somatic insertion specific to the prostatic cell line. In another report, it has been shown that transcripts initiating in the 5′ LTR of the single copy human endogenous retrovirus ERV-3, (also termed HERV-R, 10), splice into a cellular Kruppel-related zinc finger gene (43,44). These spliced transcripts are abundant in placental chorionic villi but interestingly are not expressed in choriocarcinomas (45). In addition, recent evidence indicates that enhancer elements within an endogenous retrovirus are required for the normal tissue-specific expression of the human salivary amylase genes (46). Finally, instances of endogenous LTRs providing polyadenylation signals to cellular transcripts have also been reported (41,47,48). Thus it is evident that human retrotransposons or endogenous retroviruses, like those of other species, represent a large reservior of 'mobile mutagens' that can impact upon genes near which they reside. While insertions of retrotransposons can clearly be detrimental, their ability to alter gene regulation without destroying pre-existing gene function also provides a mechanism for the rapid acquisition of variation in evolution. Indeed, a recent study has found evidence that the presense of Ty retrotransposons in yeast is evolutionarily beneficial (49). It is tempting to speculate that the actions of retrotransposons in general may be an important avenue for evolutionary adaptive change.

ACKNOWLEDGEMENTS

We thank D.Wilkinson for providing RNA samples and Drs. R.Thayer and M.Singer for the NTera2D1 cDNA library. A.Feuchter-Murthy was supported by a Steve Fonyo Studentship from the National Cancer Institute of Canada. D.Mager is a Research Scholar of the Medical Research Council of Canada. This work was supported by a grant from the MRC of Canada.

REFERENCES

1. Echalier, G. (1989) *Adv. Virus Res.* 36, 33–105.
2. Boeke, J.D. (1989) In Berg, D.E. and Howe, M.M. (eds), Mobile DNA. American Society of Microbiology, Washington, DC, pp. 335–374.
3. Kuff, E.L. and Lueders, K.K. (1988) *Adv. Cancer Res.* 51, 183–276.
4. Keshet, E., Schiff, R. and Itin, A. (1991) *Adv. Cancer Res.* 56, 215–251.
5. Ymer, S., Tucker, W.Q.J., Sanderson, C.J., Hapel, A.J., Campbell, H.D. and Young, I.G. (1985) *Nature* 317, 255–258.
6. Heberlein, C., Kawai, M., Franz, M.J., Beck-Engeser, G., Daniel, C.P., Ostertag, W. and Stocking, C. (1990) *Oncogene* 5, 1799–1807.
7. Stavenhagen, J.B. and Robins, D.M. (1988) Cell 55, 247–254.
8. Chang-Yeh, A., Mold, D.E. and Huang, R.C.C. (1991) *Nucleic Acids Res.* 19, 3667–3672.
9. Steinmeyer, K., Klocke, R., Ortland, C., Gronemeier, M., Jockusch, H., Grunder, S. and Jentsch, T.J. (1991) *Nature* 354, 304–308.
10. Larsson, E., Kato, N. and Cohen, M. (1989) *Curr. Top. Microbiol. Immunol.* 148, 115–132.
11. Leib-Mosch, C., Brack-Werner, R., Werner, T., Backmann, M., Faff, O., Erfle, V. and Hehlmann, R. (1990) *Cancer Res.* 50 (Suppl 1), 5636s–5642s.
12. Mager, D.L. and Henthorn, P.S. (1984) *Proc. Natl. Acad. Sci. USA* 81, 7510–7514.
13. Mager, D.L. and Freeman, J.D. (1987) *J. Virol.* 61, 4060–4066.
14. Fraser, C., Humphries, R.K. and Mager, D.L. (1988) *Genomics* 2, 280–287.
15. Feuchter, A. and Mager, D. (1990) *Nucleic Acids Res.* 18, 1261–1270.
16. Wilkinson, D.A., Freeman, J.D., Goodchild, N.L., Kelleher, C.A. and Mager, D.L. (1990) *J. Virol.* 64, 2157–2167.
17. Johansen, T., Holm, T. and Bjorklid, E. (1989) *Gene* 79, 259–267.
18. Feuchter, A.E., Freeman, J.D. and Mager, D.L. (1992) *Genomics* 13, 1237–1246.
19. Andrews, P.W., Damjanov, I., Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C. and Fogh, J. (1984) *Lab. Invest.* 50, 147–162.
20. Skowronski, J., Fanning, T.G. and Singer, M. (1988) *Mol. Cell. Biol.* 8, 1385–1397.
21. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Spring Harbor, NY.
22. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6–13.
23. Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) Basic Methods in Molecular Biology. Elsevier, New York, New York.
24. Tabor, S. and Richardson, C.C. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4767–4771.
25. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387–395.
26. Brenner, C.A., Tam, A.W., Nelson, P.A., Engleman, E.G., Suzuki, N., Fry, K.E. and Larrick, J.W. (1989) *Biotechniques* 7, 1096.
27. Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. Eds. (1990) PCR protocols: A guide to methods and applications. Academic Press, Inc., New York.
28. Temin, H.M. (1982) *Cell* 28, 3–5.
29. Kozak, M. (1987) *Nucleic Acids Res.* 15, 8125–8132.
30. Dennis, E.A. (1983) In Boyer, P.D. (ed), The Enzymes: Lipid Enzymology, 3rd edition. Academic Press, New York, Vol XVI, pp. 307–353.
31. Dennis, E.A. (1987) *Bio/Technology* 5, 1294–1330.
32. Pruzanski, W. and Vadas, P. (1991) *Immunol. Today* 12, 143–146.
33. Seilhamer, J.J., Randall, T.L., Yamanaka, M. and Johnson, L.K. (1986) *DNA* 5, 519–527.
34. Seilhamer, J., Pruzanski, J.W., Vadas, P., Plant, S., Miller, J., Kloss, J. and Johnson, L.K. (1989) *J. Biol. Chem.* 264, 5335–5338.
35. Kramer, R.M., Hession, C., Johansen, B., Haynes, G., McGray, P., Chow, E.P., Tizard, R. and Pepinsky, R.B. (1989) *J. Biol. Chem.* 264, 5768–5775.
36. Kunsunoki, C., Satoh, S., Kobayashi, M. and Niwa, M. (1990) *Biochim. Biophys. Acta* 1087, 95–97.
37. von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683–4690.

38. Heinrikson, R.L., Kreuger, E.T. and Keim, P.S. (1977) *J. Biol. Chem.* 252, 4913–4921.
39. Davidson, F.F. and Dennis, E.A. (1990) *J. Mol. Evol.* 31, 228–238.
40. van den Bergh, C.J., Slotboom, A.J., Verheij, H.M. and de Haas, G.H. (1989) *J. Cell. Biochem.* 39, 379–390.
41. Mager, D.L. (1989) *Virology* 173, 591–599.
42. Liu, A.Y. and Abraham, B.A. (1991) *Cancer Res.* 51, 4107–4110.
43. Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larsson, E., Rydnert, J., Ohlsson, R. and Cohen, M. (1987) *J. Virol.* 61, 2182–2191.
44. Kato, N., Shimotohno, K., VanLeeuwen, D. and Cohen, M. (1990) *Mol. Cell. Biol.* 10, 4401–4405.
45. Kato, N., Larsson, E. and Cohen, M. (1988) *Int. J. Cancer* 41, 380–385.
46. Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C. and Meisler, M.H. (1992) *Genes Dev.* 6, 1457–1465.
47. Goodchild, N.L., Wilkinson, D.A. and Mager, D.L. (in press) *Gene* .
48. Paulson, K.E., Matera, A.G., Deka, N. and Schmid, C.W. (1987) *Nucleic Acids Res.* 15, 5199–5215.
49. Wilke, C.M. and Adams, J. (1992) *Genetics* 131, 31–42.
50. Verheij, H.M., Westerman, J., Sternby, B. and De Haas, G.H. (1983) *Biochim. Biophys. Acta* 747, 93–99.
51. Tsai, I.H., Wu, S.H. and Lo, T.B. (1981) *Toxicon* 19, 141–152.
52. Kanda, A., Ono, T., Yoshida, N., Tojo, H. and Okamoto, M. (1989) *Biochem. Biophys. Res. Commun.* 163, 42–48.
53. Viljoen, C.C., Botes, D.P. and Kruger, H. (1982) *Toxicon* 20, 715–737.
54. Ohara, O., Tamaki, M., Nakamura, E., Tsuruta, Y., Fujii, Y., Shin, M., Teraoka, H. and Okamoto, M. (1986) *J. Biochem* 99, 733–739.
55. Takasaki, C., Kuramochi, H., Shimazu, T. and Tamiya, N. (1988) *Toxicon* 26, 747–749.
56. Komada, M., Kudo, I., Mizushima, H., Kitamura, N. and Inoue, K. (1989) *J. Biochem.* 106, 545–547.
57. Randolph, A. and Heinrikson, R.L. (1982) *J. Biol. Chem.* 257, 2155–2161.
58. Kuchler, K., Gmachl, M., Sippl, M.J. and Kreil, G. (1989) *Eur. J. Biochem.* 184, 249–254.