Characterization of the Human Parathyroid Hormone-like Peptide Gene

FUNCTIONAL AND EVOLUTIONARY ASPECTS

(Received for publication, November 7, 1988)

Toshiyuki Yasuda†, Denis Banville‡, Geoffrey N. Hendy¶, and David Goltzman§

From the §Calcium Research Laboratory, Departments of Physiology and Medicine, McGill University and Royal Victoria Hospital, Montreal, Quebec H3A 1A1, Canada and the ¶Laboratory of Molecular Genetics, National Research Council Biotechnology Research Institute, Montreal, Quebec H4P 2R2, Canada.

The single-copy gene coding for the human parathyroid hormone-like peptide was isolated from a human placental genomic library. The gene spans 13 kilobases and contains seven exons. Exons I and II encode 5'-noncoding regions; each has its own transcription initiation site, and the two promoters are separated by over 1000 base pairs of genomic DNA. Exon III encodes the prepro-coding region, and exon IV encodes the mature peptide sequence. At the end of exon IV the splice site interrupts codon 139 of the mature peptide. Exon V, which is contiguous with exon IV, encodes a stop codon and a 3'-noncoding region. Exon VI encodes 34 additional amino acids, a stop codon, and a second 3'-noncoding region. Exon VII encodes two extra amino acids, a stop codon, and a third 3'-noncoding region. This genomic organization reveals how the multiple human parathyroid hormone-like peptide RNA transcripts, which have been observed, arise by both alternative splicing out of exons and use of multiple promoters. The mRNAs, which can potentially be formed from the primary transcript of this gene, could have one of three different carboxy-terminal coding regions. The use of different exons to encode the different functional domains, 5'-noncoding region, prepro-coding region, and mature peptide region is identical to the organization of the human parathyroid hormone gene. This strongly suggests a common evolutionary origin of the two genes.

Hypercalcemia is an important cause of morbidity of a wide variety of malignant neoplasms. Recently a peptide with an amino-terminal sequence homologous to that of parathyroid hormone (PTH) has been isolated (1-3) and is likely to be of significance in the pathogenesis of the abnormality in calcium homeostasis. Complementary DNAs encoding this human parathyroid hormone-like peptide (hPLP) have been prepared from a variety of tumors (4-7), and their use as hybridization probes for Northern blot analysis has revealed a pattern of multiple mRNA species in different tissues (4, 6, 7). Several cDNAs encoding variants of hPLP have now been characterized and have been found to share a common sequence for most of their coding regions but to differ either in the 5'-untranslated region, the carboxy-terminal coding region, or the 3'-untranslated region (4, 7). In contrast to this complexity at the mRNA level, Southern blot analysis of human genomic DNA suggests that only a single copy of the hPLP gene is present in the human genome (4, 6). In order to better understand the basis of this transcriptional heterogeneity and the evolutionary relationship of PLP to PTH, we have analyzed the structure and organization of the hPLP gene.

EXPERIMENTAL PROCEDURES

Construction and Screening of the Genomic Library—Human placental DNA was partially digested with Sau3AI, fractionated by electrophoresis on a 1% agarose gel, and restriction fragments, 15–23 kb in size, were ligated with BamHI-digested EMBL-3 and packaged in vitro. Approximately 450,000 plaques were screened by filter hybridization using polynucleotide kinase 35P-abeled oligomers, which were synthesized according to the hPLP cDNA sequence reported by Suva et al. (5). The oligonucleotides used were: a 30-mer 5' GTCATGGGAGGCTGTAGTCTCCACACAGCAC 3', complementary to nucleotides 109–138; a 41-mer 5' CCGTGAATCAGGCCTCGACGAGGATGTCAGACAG 3', complementary to nucleotides 485–525; a 20-mer 5' GCATTTTACAGAC 3', complementary to nucleotides 616–638 of the hPLP cDNA (5). Hybridization was performed in 0.9 M NaCl, 0.09 M Tris-HCl, pH 7.4, 0.006 M EDTA, 2 × Denhardt's solution, and 0.2% sodium dodecyl sulfate at 5 °C below the calculated Tm for each probe for 12-18 h. The filters were washed with several changes of 6 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7) initially at 4 °C, then at room temperature, and finally at 5–10 °C below the calculated Tm of the probe. The authenticity of each positive clone at the initial screening was tested by using two different probes at the second screening, and each positive clone was subjected to sequential plaque purification.

Restriction Enzyme Mapping and DNA Sequence Analysis—DNA was prepared from the purified recombinant phage, and restriction enzyme mapping was carried out by Southern blot hybridization using the oligomers as probes. Filter hybridization screening conditions were the same as those for the library screening. Restriction fragments identified by probes corresponding to the 5'-noncoding region, the prepro-coding region, the mature peptide coding region, and the 3'-noncoding region were subcloned into either pGEM-1 or KS(-) plasmid vectors for further restriction mapping and nucleotide sequence analysis, which was carried out by the dyeoxy chain termination method (8). In addition, in order to identify intron/exon
junctons, several restriction fragments containing intron DNA were subcloned, and the insert DNA was 32P-labeled by the random primer method and used for Southern blot analysis of the cloned DNA. Selected subclones labeled in the same manner were used in Southern blot analysis of restriction enzyme-digested human leukocyte DNA.

**Primer Extension Analysis of hPLP mRNA—**Total RNA was isolated from normal human keratinocytes and EC-GI cells (an esophageal cancer cell line (9)) by the guanidinium thiocyanate/cesium chloride method and used for Southern blot analysis of the cloned DNA. A 18-base oligonucleotide complementary to nucleotides 1876-1893 which hybridized to both oligomers. These were designated as terminal coding region oligomer as probe, 10 strongly positive terminal oligonucleotide identified two recombinant clones.

**RESULTS**

Identification and Restriction Mapping of Genomic Clones

At the initial screening of 450,000 plaques, using an amino-terminal coding region oligomer as probe, 10 strongly positive phase plaques were identified. Rescreening with a carboxyl-terminal oligonucleotide identified two recombinant clones which hybridized to both oligomers. These were designated as APLPg1 and APLPg7-2 (see Fig. 1) and contained 20 and 23 kb of human DNA, respectively. In addition, screening of the genomic library with an oligonucleotide corresponding to a part of the 3′-noncoding region identified recombinant clone λPLPg3 which contained 22 kb of human DNA (Fig. 1). In order to facilitate restriction mapping and DNA sequence analysis, a variety of restriction fragments was subcloned. A physical map of the hPLP gene and flanking regions is shown in Fig. 1.

Characterization of the 5′ End of the Gene by Primer Extension Analysis

**Promoter Region 1—** Primer extension of oligomer 1, with RNA from EC-GI cells, yielded two major bands at 220 and 145 nucleotides (Fig. 2). This would indicate that there are two transcription start sites in the first promoter region; one is at nucleotide 808 and the second is at nucleotide 883 (Figs. 2 and 4). The second of these sites lies within the cDNA clone isolated by Mangin et al. (4) indicating that this mRNA was transcribed from the first start site. Neither transcription start site is downstream of a typical TATA box sequence. There is a TCTTCTTC sequence 29 bp upstream of the start site is boxed.

**Promoter Region 2—** Primer extension of oligomer 2 yielded two bands at 89 and 91 nucleotides, and primer extension of oligomer 3 yielded two bands at 109 and 111 nucleotides with RNA from both EC-GI cells and keratinocytes (panel A) or EC-GI cells (panel B) were hybridized with an excess of polynucleotide kinase-labeled oligonucleotide 2 or 3 and the extension reaction performed as described under “Experimental Procedures.” The products were analyzed on an 8% acrylamide, 7 M urea gel (panel A, lane 1). Arrows indicate the two predominant extension products. Lane M contains size markers. Nucleotides 761-1032 from Fig. 4 are displayed in panel B. The antisense oligomer used is represented by (−) and arrowheads (↑) show the potential start sites indicated by the primer extension analysis.

![Physical map of human DNA encoding the PLP gene](image)

**FIG. 1. Physical map of human DNA encoding the PLP gene.** A human placental DNA library was prepared as described under “Experimental Procedures” and screened by filter hybridization using polynucleotide kinase 32P-labeled oligomers complementary to portions of hPLP cDNA. Three recombinant clones, APLPg1, APLPg3, and APLPg7-2, were isolated and characterized. These clones overlapped and spanned approximately 38.5 kb of genomic DNA. The restriction fragments which were subcloned for restriction mapping and nucleotide sequence analysis are shown (e.g., sh-1, sh-2, etc.). Restriction sites shown are: B, BamHI; E, EcoRI; H, HindIII; P, PvuII; S, SacI. *St* indicates two SacI sites approximately 50 bp apart.

![Primer extension analysis of hPLP mRNA in promoter region 1](image)

**FIG. 2. Primer extension analysis of hPLP mRNA in promoter region 1.** Twenty μg of total RNA from EC-GI cells were hybridized with an excess of polynucleotide kinase-labeled oligonucleotide 1 and the extension reaction performed as described under “Experimental Procedures.” The products were analyzed on an 8% acrylamide, 7 M urea gel (panel A, lane 1). Arrows indicate the two predominant extension products. Lane M contains size markers. Nucleotides 761−1032 from Fig. 4 are displayed in panel B. The antisense oligomer used is represented by (−) and arrowheads (↑) show the potential start sites indicated by the primer extension analysis.

![Primer extension analysis of hPLP mRNA in promoter region 2](image)

**FIG. 3. Primer extension analysis of hPLP mRNA in promoter region 2.** Twenty μg of total RNA from human keratinocytes (panel A) or EC-GI cells (panel B) were hybridized with an excess of polynucleotide kinase-labeled oligonucleotide 2 or 3 and the extension reaction performed as described under “Experimental Procedures.” The products were analyzed on an 8% acrylamide, 7 M urea gel (lane 2, oligomer 2; lane 3, oligomer 3). Arrows indicate the predominant extension products. Nucleotides 1741−1921 are displayed in panel C. Antisense oligomers used are represented by (−), and arrowheads (↑) show potential start sites indicated by the primer extension analysis. A TATAA sequence 30 bp upstream from the start site is boxed.
The nucleotide sequence of the human PLP gene

The nucleotide sequence of the human PLP gene was determined as described under "Experimental Procedures" from the subcloned DNAs shown in Fig. 1. Exons are shown in capital letters and intervening sequences and flanking DNA in lower-case letters. The protein coding and splicing signals are indicated by **boldface** and *italics*, respectively.

**DNA Sequence Analysis and Structural Organization of the hPLP Gene**

The nucleotide sequence of the human PLP gene is shown in Fig. 4. All exons were completely sequenced, as were all the intron/exon junctions. The small intervening sequences (I1) encode two different 5′ noncoding regions, and exon II encodes 22 nucleotides of the 5′ noncoding region and the prepro-coding region of the precursor peptide. Exon IV encodes the Lys-Arg prohormone cleavage site and the mature peptide coding region. Exons V, VI, and VII encode three alternative 3′ noncoding regions. Exons VI and VII also encode different additional carboxyl-terminal sequences of the mature peptide. All exon/intron junctions have the consensus GT-AG sequence, including the splice junction at the end of exon IV which contains a GT beginning in the Arg codon for amino acid 139 of the mature peptide (see Fig. 4). Southern
DNAs, using as probes subcloned DNA covering all exons, revealed a simple pattern of restriction fragments which was identical to that seen with the recombinant DNA (data not shown).

**Generation of Multiple mRNA Species from the Single-copy PLP Gene**

Fig. 5 illustrates how the multiple mRNA transcripts so far reported are derived from the human PLP gene.

**Comparison of the hPTH and hPLP Genes**

The hPLP gene shares important organizational features with the hPTH gene in that in each case the same functional domains are encoded by single exons (Fig. 6). In both genes the 5'-noncoding region is essentially encoded by a single exon which is spliced to a second exon encoding the prepro-coding region of the precursor peptide. This, in turn, is spliced to an exon encoding the Lys-Arg prohormone cleavage site and the mature peptide. In one case, when exon V is used to encode the 3'-noncoding region, the similarity of the hPLP and hPTH genes persists. However, when exons VI or VII are used, this no longer the case. The hPTH gene does not use a different exon to specify the 3'-noncoding region.

**DISCUSSION**

The hPLP gene spans 13 kb and contains seven exons. Both by the use of two different promoters and by the use of alternative splice sites to include or exclude different exons, heterogeneous hPLP mRNA species may be generated and account for the multiple mRNA species recently identified.

The 5'-untranslated region contained in exon I is unusually long. It is present in the hPLP cDNA characterized by Mangin et al. (4) which was judged to be a near full-length copy of a

**FIG. 5.** Generation of multiple mRNA species from the single-copy human prepro-PLP gene by alternative use of two promoters and alternative splicing of exons encoding the 5'-noncoding region, the 5'-untranslated region, and the 5'-noncoding region. Exons are shown as boxes designated by a Roman numeral. The hatched boxes denote 5'-noncoding regions; the black boxes denote coding regions; and the dotted boxes denote 3'-noncoding regions. Messenger RNA A, which was cloned from lung carcinoma and renal carcinoma cell lines (4, 6), is approximately 2.1 kb in length, and encodes a mature peptide of 141 amino acids. It is derived from exons I, III, IV, and VII. Messenger RNA B, which was cloned from a lung carcinoma cell line cDNA library (5), is approximately 1.4 kb in length and encodes a mature peptide of 141 amino acids. Messenger RNA C is from a renal carcinoma cell line cDNA library (6), is approximately 1.8 kb in length, and encodes a mature peptide of 139 amino acids. It is derived from exons II, III, IV, and VII. Messenger RNA D, which we predict to encode a peptide of 173 amino acids corresponds to the cDNA prepared from a renal carcinoma cell line by Mangin et al. (7) which was reported to encode a mature peptide of 166 amino acids. Therefore, examination of the cDNA with oligo(dT) primer, leading to the spurious result obtained. In the other case (7), the deduced amino acid sequence incorrectly suggested a mature peptide of 166 amino acids, possibly because of a similar cloning artifact induced by incorrect oligo(dT) priming. Therefore, examination of the hPLP gene structure has clarified these issues and predicts the use of this alternative exon VI to generate a cDNA

2.1-kb mRNA identified by Northern blot analysis. One of the start sites defined by our primer extension studies is upstream of the end of this clone, whereas the second is within the cDNA sequence. It cannot be absolutely excluded, therefore, that the downstream "start site" is an artifact caused by premature termination of the primer extension analysis. There is no sequence resembling a "TATA" box upstream of the putative transcriptional start sites. The 5'-untranslated region in exon II is present in all other hPLP cDNAs identified to date (2, 4, 5). In this region we identified two transcriptional start sites 2 base pairs apart. The data we have obtained by primer extension analysis of RNA from both normal human keratinocytes and an esophageal carcinoma cell line are fully consistent with that obtained previously by Thiede et al. (6) using RNA from a renal carcinoma cell line. There is a consensus "TATA" sequence approximately 30 bp and a sequence resembling a "CAAT" sequence 80 bp upstream of the putative transcription start site (6). Sequences similar to "CCCCACCCG," the consensus sequence for the transcription factor protein (AP-2) binding site (12, 13), are found in both promoter regions and represent potential regulatory elements for both cyclic AMP and phorbol esters as well as basal enhancers.

The prepro sequence of hPLP is 36 amino acids in length, but mature hPLP peptides of 139, 141, and 173 amino acids can be generated by alternative use of exons IV, VII, and VI, respectively; indeed corresponding cDNAs encoding peptides of 139 and 141 amino acids have been reported (4-6). Recent reports have identified cloned cDNAs corresponding to the transcript resulting from the use of exon VI as an acceptor of exon IV (6, 7). However, in one case (6), this sequence was interpreted as a "readthrough" of an intron sequence. Presumably, the sequence of 11 consecutive adenine residues, present normally within the coding region of exon VI, contributed to an internal start site for the reverse transcriptase in preparing the cDNA with oligo(dT) primer, leading to the spurious result obtained. In the other case (7), the deduced amino acid sequence incorrectly suggested a mature peptide of 166 amino acids, possibly because of a similar cloning artifact induced by incorrect oligo(dT) priming. Therefore, examination of the hPLP gene structure has clarified these issues and predicts the use of this alternative exon VI to generate a cDNA.
encoding a 173-residue peptide. Consequently, our analysis indicates that exons I, II, V, VI, and VII are combinatorial, whereas exons III and IV which encode the majority or all of the human prepro-PLP molecule are constitutive (14). Such selective use of specific exons provides a potential mechanism for developmental or tissue-specific gene regulation (15) and for production of hPLP peptides differing in the carboxy-terminal region.

Although amino-terminal peptides of hPLP have been shown to mimic a wide variety of the biological actions of PTH (16-21) the function of the midregion and carboxy terminus of the mature hPLP molecule remains enigmatic. The carboxy-terminal hPLP sequence 142-173 reported here is novel and predicts a peptide with a highly hydrophobic end of 15 amino acids in length terminating in a sequence of 4 leucine residues. Modification of the carboxy region of other proteins has been shown to markedly influence processing and/or secretion (22-25). Consequently, alterations in the carboxy end of the hPLP molecule could play a similar role in altering cellular handling of hPLP.

There is evidence that hPLP is expressed in normal tissues (4, 26, 27), although perhaps at lower levels than in malignancies. However, the potential physiological role of hPLP is unclear although it has been shown that PLP is expressed in lactating mammry tissue in the rat (27) and, therefore, may play a role in calcium mobilization in lactation. The promoter regions of the hPLP gene were found to be GC-rich. In particular, the promoter upstream of exon II was found to contain nine sequences of potential SP-1 binding sites (28, 29) of the type “CCGCC” and its inverted complement “GGCCG” (“GC” boxes). Because the promoters of genes, such as those encoding H-Ras I, the epidermal growth factor receptor, insulin, and insulin-like growth factor II, contain many sequence motifs of this type (and lack a typical TATA box sequence), it has been proposed that many of the genes involved in cellular growth will be characterized by multiple “GC” boxes in their promoters (30, 31). Whether this provides a clue to the physiological function of PLP remains to be clarified. In addition to these considerations, all the alternative 3′-untranslated regions were found to be AU-rich and to contain repeated sequence motifs of the type “AUUUUA.” We and others have recently cloned cDNA encoding rat PLP (26, 27) and found that these motifs are also present in the corresponding regions of rat PLP mRNA indicating interspecies conservation and implying functionality. Such motifs contribute to mRNA instability and are common in mRNA involved in cellular proliferation and/or differentiation, such as proto-oncogenes, lymphokines, and cytokines (32, 33). Consequently, both promoter and 3′-untranslated regions of the hPLP gene share structural features which may bear on a possible functional role for hPLP in cellular growth and proliferation.

The hPLP gene has been localized to the short arm of chromosome 12, whereas the hPTH gene is on the short arm of chromosome 11 (4, 34). Both chromosomes bear related genes (35) and are thought to have arisen from a single ancestral chromosome (36). We have found that the hPLP gene shares important organizational features with the hPTH gene (37) in that in each case the same functional domains are encoded by single exons. In both genes the 5′-untranslated region is essentially encoded by a single exon which is joined to a second exon encoding the prepro region of the precursor peptide. This, in turn, is spliced to an exon encoding the Lys-Arg prohormone cleavage site and the mature peptide. The acceptor splice site which interrupts the prohormone sequence is precisely conserved. When exons VI and VII are used to encode the 3′-untranslated region the analogy with the hPTH gene, which does not use a different exon to specify the 3′-untranslated region, diverges. However, when exon V is used to encode the 3′-untranslated region, resemblance of the hPLP and hPTH genes persists. Consequently taken together with considerations of sequence homology and chromosomal localization, the similarities in structural organization of the hPLP and hPTH genes determined by our studies provide compelling evidence for their derivation from a common ancestral gene.

Acknowledgments—We thank D. Allen, K. Kinachok, and P. Milne for secretarial assistance, Y. Boie and H. Bergeron for excellent technical advice and assistance, Dr. K. Sato for EC-G1 cells, J. Desjardins, R. Kremer, and W. Gulliver for the culture of keratinocytes and EC-G1 cells, and the Biotechnology Research Institute for accommodation to T. V.

REFERENCES


