Human *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase (human meprin): genomic structure of the α and β subunits

Dagmar HAHN*[1], Rainer ILLISSON†, Andres METSPALU† and Erwin E. STERCHI*

*Institute of Biochemistry and Molecular Biology and Department of Pediatrics, University of Berne, 3012 Berne, Switzerland, and †Institute of Molecular and Cell Biology, Tartu University, Estonian Biocentre, 23 Riia St., EE 2400 Tartu, Estonia

N-Benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase (PPH, human meprin), a zinc-metalloendopeptidase of the astacin family, consists of two similar subunits. As well as in small-intestinal epithelial cells, the enzyme is found in lamina propria leucocytes, human cancer cells and colorectal cancer tissue, making it a potential candidate for a role in tumour formation and cancer progression. To elucidate the mechanisms that control *PPH* gene expression and to gain more insights into the evolutionary relationship of the two subunits, we analysed the complete exon–intron organization and searched for putative regulatory elements in 3 kb of the upstream region of both genes. The human gene for the α subunit is approx. 35 kb in size and contains 14 exons. The gene for the β subunit is organized in 15 exons and spans approx. 27 kb. A comparison of both genes indicates strong structural similarities. The exons are almost identical in size, except exon 13 in PPHα, which codes for an additional I domain not present in PPHβ. The locations of the respective exon–intron junctions and the intron phases are almost identical; five of them contain conserved split codons. The main variation is in the intron lengths. It can be concluded that PPHα and PPHβ are derived from a common ancestor. Sequence analysis of the 5′ flanking DNA with a computer search for promoter elements and different promoter constructs transfected into Caco-2 cells revealed a number of potential regulatory motifs and suggests that each of the two genes is regulated independently.

Key words: astacin family, gene promoter, metalloprotease.

INTRODUCTION

N-Benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase (PPH, human meprin; EC 3.4.24.18), originally isolated from the microvillus membrane of small intestinal epithelial cells [1], belongs to the astacin protease family [MEROPS database (the Peptidase Database: http://www.bi.bbsrc.ac.uk/Merops/Merops.htm): family M12A] of metalloproteases [2,3]. These proteases are characterized by a conserved extended active site consensus sequence (HEXXHXXGFXHE) and a methionine turn [4]. They are involved in the formation of bone and cartilage [5], in the breakdown of the egg envelope during hatching [6,7], in dorsoventral patterning [8,9] and in early morphogenetic processes [10]. PPH hydrolyses a number of biologically active peptides such as angiotensins, bradykinin, substance P, the β chain of insulin, and parathyroid hormone [1] as well as extracellular matrix components such as laminin–nidogen, type IV collagen and fibronectin *in vitro* [11,12]. Structurally, PPH is a type I transmembrane protein composed of two disulphide-bridged multidomain subunits (α and β) [13,14]. Both subunits are encoded by single genes and have been mapped to human chromosomal regions 6p11–p12 (α subunit) and 18q12.2–18q12.3 (β subunit) [15]. Recent studies from our laboratory have shown that both PPH subunits are differentially expressed in epithelial cells of ileum and colon [16]. In ileum, both subunits are expressed mainly in a heterodimeric membrane-bound form, whereas in colon only secreted PPHα is found. The observation that Caco-2 cells, a colon carcinoma cell line, endogenously express only PPHα supports the finding of distinct regulation of both subunits. In colorectal cancer, PPHα accumulates in the tumour stroma. The ability to degrade extracellular matrix components suggests that PPH might therefore be associated with tumour cell migration, invasion and/or metastasis of carcinoma cells [17].

As a step towards elucidating the mechanisms responsible for PPH expression we report here the organization of the genes of both PPH subunits and the characterization of their 5′-regulatory sequences. These results extend and strengthen current concepts regarding the evolution of the astacin family of metalloproteases.

EXPERIMENTAL

Materials and general methods

Radioactive nucleotides [α-^{32}P]dCTP (3000 Ci/mmol), [γ-^{32}P]ATP (5000 Ci/mmol) and [α-^{35}S]dATP (1000 Ci/mmol) were purchased from NEN Life Science (Geneva, Switzerland). Ready-to-go DNA labelling kit and T7 sequencing kit were from Amersham Pharmacia Biotech (Dübendorf, Switzerland). All restriction enzymes, T4 DNA polymerase and T4 polynucleotide kinase were purchased from New England Biolabs (Schwalbach, Germany). DNAse I, Superscript reverse transcriptase and LipofectAMINE were from Life Technologies (Basel, Switzerland), as were cell culture media, penicillin, streptomycin and essential amino acids. T4 DNA-ligase, calf intestinal phosphatase, dNTPs and *Taq* polymerase were obtained from

Abbreviations used: AP1, activator protein 1; EGF, epidermal growth factor; HNF-3, hepatocyte nuclear factor 3; PEA3, polyoma-virus-enhancer-A-binding protein 3; PPH, *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase; RACE, rapid amplification of cDNA ends.

[1] To whom correspondence should be addressed (e-mail dagmar.hahn@mci.unibe.ch).

The nucleotide sequence data reported in this paper will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers AJ251058 for PPHα and AJ251059 for PPHβ.

Promega (Wallisellen, Switzerland). Oligodeoxynucleotides for sequencing and PCR were synthesized by Microsynth (Balgach, Switzerland) or Intron (Kaltbrunn, Switzerland). Colony and dot-blot hybridizations were performed with Hybond-C Extra nitrocellulose membrane; and Southern blot hybridization was done with Hybond™-N nuclear transfer membranes (Amersham Pharmacia). Hybridizations were performed at 42 °C in a solution containing 50 % (v/v) formamide, 6 × SSC (SSC is 0.15 M NaCl/0.015 M sodium citrate), 5 × Denhardt's (Denhardt's is 0.02 % Ficoll 400/0.02 % polyvinylpyrrolidone/0.002 % BSA), 0.5 % SDS and 20 μg/ml salmon-sperm DNA in the presence of 10^6 c.p.m./ml of labelled cDNA probe. The blots were washed at 55 °C successively in 2 × SSC for 15 min, in 2 × SSC/0.1 % SDS for 30 min and in 0.1 × SSC for 10 min (twice in each) before autoradiography. PCR used for the amplification of DNA fragments for subcloning or for the identification of clones containing coding PPHα or PPHβ sequences was performed under standard conditions. Amplification of introns was done on human genomic, cosmid or plasmid DNA as template with the Expand™ long-template PCR system (Roche Diagnostics, Rotkreuz, Switzerland), as described by the supplier.

Genomic library screening

Human chromosomes 6- and 18-specific cosmid library filters, subcloned in Lawrist 4 (obtained from the Reference Library Database, Max Planck Institute for Molecular Genetics, Berlin, Germany) [18] were prehybridized in a solution containing 6 × SSC, 10 × Denhardt's, 0.1 % SDS and 20 μg/ml salmon-sperm DNA at 65 °C for 6 h and then hybridized in the same solution containing at least 10^6 c.p.m./ml ^{32}P-labelled PPHα cDNA covering nt 1–2785 (*Bam*HI/*Bgl*II fragment) or PPHβ cDNA covering nt 558–2311 (*Eco*RI/*Eco*RI fragment) as well as ^{35}S-labelled plasmid DNA as the background control. PPHα and PPHβ cDNA [13,14] were labelled by random priming. Unincorporated nucleotides were removed by precipitation with ammonium acetate. The filters were washed for 20 min each at 65 °C successively in 2 × SSC, 2 × SSC/0.1 % SDS and 0.1 × SSC and then exposed to an autoradiographic film with an intensifying screen at −70 °C. Rescreening was performed by dot-blot hybridization. Cosmid DNA isolated by alkaline lysis [19] was digested with restriction enzymes specific for PPH cDNA species, separated in 0.8 % agarose, transferred to nylon membranes and hybridized with full-length as well as 5′- and 3′-specific probes for PPHα and PPHβ. The 5′-specific probe for PPHα covered nt 1–441 (*Bam*HI/*Bgl*II), and that for PPHβ nt 1–557 (*Eco*RI/*Eco*RI). The 3′-specific probe of PPHα consisted of nt 1854–2300 (*Stu*I/*Stu*I), and that for PPHβ nt 1628–2084 (*Stu*I/*Sac*I). For PPHα, clone 8c (109D2338Q) and clone 9b (109O046Q5) were identified as spanning the entire length of the PPHα gene; similarly, for PPHβ, clone 2f (111P1919Q3) and clone 7d (111D1338Q3) were identified as spanning the entire length of the PPHβ gene (numbers in parentheses are the original terminology from the Max Planck Institute).

Subcloning and construction of luciferase DNA plasmids

*Hin*dIII and *Pst*I fragments of PPHα genomic inserts and *Eco*RI fragments of PPHβ genomic inserts were selected for subcloning. The complete restriction digests were ligated into linearized dephosphorylated pBluescript KS− vector (Stratagene, Heidelberg, Germany) and transformed into competent XL1 Blue cells. Positive clones were identified by colony hybridization. Southern blot hybridization was performed to identify plasmids containing different PPHα and PPHβ fragments.

The PPHα promoter was isolated by PCR with the T7 primer and a specific oligonucleotide introducing a *Spe*I restriction site at the 3′ end, with the use of the genomic clone 9b8 as a template. After phosphorylation and treatment with T4 DNA polymerase, the approx. 3500 bp fragment was subcloned into the *Sma*I restriction site of pBluescript KS− and sequenced. The complete PPHα promoter −3426 to +10 (*Bam*HI/*Bam*HI), as well as the promoter fragments −2008 to −1357 (*Hin*cII/*Hin*cII) and −1356 to +10 (*Hin*cII/*Bam*HI) were subcloned into pGL3 basic (Promega). The PPHβ promoter was similarly amplified by PCR, subcloned into pBluescript KS and sequenced. The full-length promoter fragment from −3893 to +50 (*Kpn*I/*Xba*I), as well as a *Hin*dIII/*Spe*I fragment (−2358 to −243) and the *Spe*I/*Spe*I fragment (−242 to +50) were inserted into pGL3 basic. In case of incompatible restriction sites the 3′ recessed ends were filled in and inserted into the *Sma*I site of the pGL3 basic vector. The orientation of all inserts was determined by restriction analysis and sequencing.

RNA preparation and primer extension

Primer extension [20] was performed on human small-intestinal RNA isolated previously [14]. Oligonucleotide primers complementary to bases +52 to +72 of PPHα (5′-AATCGGTACAGCTGCTATGTC-3′) and to bases +30 to +50 of PPHβ (5′-ACGAGAAGAGCATCCAAGAAC-3′) cDNA were 5′ end-labelled with [γ-^{32}P]ATP. Heat-treated RNA (15 μg) was incubated with the labelled oligonucleotide at 65 °C for 90 min in a solution containing 150 mM KCl, 10 mM Tris/HCl, pH 8.0, and 1 mM EDTA. After being annealed, the primers were extended with 5 units of Superscript reverse transcriptase for 1 h at 42 °C in the presence of 150 μM dNTP, 150 ng/μl actinomycin D, 5.5 mM dithiothreitol, 50 mM KCl, 10 mM $MgCl_2$ and 23 mM Tris/HCl, pH 8.3. RNA was digested and the DNA was extracted with phenol/chloroform/3-methylbutan-1-ol (25:24:1, by vol.), precipitated with ethanol and separated in a 8 % (w/v) polyacrylamide/7 M urea gel, alongside sequencing ladders.

DNA sequencing

By using T3 and T7 primers the nucleotide sequences of plasmids containing genomic inserts were determined with the automated sequencing service from Microsynth. Other sequencing primers were designed on the basis of known cDNA or new sequences. Parts of the sequence were determined with the T7 sequencing kit by using [α-^{35}S]dATP as the labelled nucleotide. The overall organization of both genes was derived by a combination of fragment length analysis and DNA sequence analysis with the MacVector 6.5 program.

Cell culture and transfection

Caco-2 cells, grown in minimal essential medium/Earles Hepes medium without glutamine, supplemented with 20 % (v/v) fetal calf serum, 100 units/ml penicillin, 100 μg/ml streptomycin and 292 μg/ml glutamine, at 37 °C in a humidified air/CO_2 (19:1) environment were seeded out in 60 mm tissue culture dishes at 5×10^5 cell density and transfected 48 h later with LipofectAMINE. The transfection mixture contained 250 ng of pSV β-galactosidase control vector (Promega) and 0.37 pmol of reporter construct as well as 15 μl of LipofectAMINE in 300 μl of serum-free medium. At 72 h after transfection, the cells were lysed in 500 μl of reporter lysis buffer (Promega), frozen at −20 °C for 30 min and centrifuged briefly. The supernatants were stored in aliquots at −70 °C.

Determination of β-galactosidase and luciferase activity

β-Galactosidase activity was measured in 30 μl of cell extract as described by the supplier. The absorbance was read at 405 nm in a Mikro/Plate Reader 450 (Bio-Rad, Glattbrugg, Switzerland). For luciferase assays, 20 μl of cell extract was mixed with 100 μl of luciferase assay reagent (Promega) at room temperature. The light produced was measured for a period of 10 s, beginning 2 s after mixing, in a luminometer (TD-20/20; Promega). The luciferase activities were normalized to β-galactosidase activities to correct for transfection efficiencies and expressed as fold induction over the control reaction with pGL3 basic.

RESULTS

Isolation of human PPH genomic clones

Screening of genomic libraries prepared from human chromosomes 6 and 18 by using randomly labelled PPHα and PPHβ cDNA resulted in the identification of two positive clones for each subunit. In combination they spanned the entire PPHα and PPHβ genes. Appropriate restriction fragments containing 5′ and 3′ coding regions were subcloned into pBluescript KS−. T3 or T7 primers and oligonucleotide primers identical with the anti-sense strand at position +52 to +72 for PPHα and +30 to +52 for PPHβ were used to identify clones containing promoter and 5′ coding sequences by PCR. To identify clones containing continuous sequences, PCR was performed with T3 or T7 primers and oligodeoxynucleotide primers derived from known cDNA sequences or from genomic sequences derived during the sequencing project. The complete sequences of PPHα and PPHβ cDNA were determined, including exon/intron boundaries. For large introns, where the complete sequence was not established, PCR of human genomic DNA, plasmid DNA or cosmid DNA was performed with two oligodeoxynucleotide primers flanking these introns to estimate their approximate sizes.

Organization and comparison of PPHα and PPHβ genes

DNA sequencing of the genomic clones resulted in the identification of 14 exons for the PPHα gene and 15 exons for the PPHβ gene. In both genes the initiation codons and the termination codons are encoded within the first and last exons. The PPHα gene spans approx. 35 kb, the PPHβ gene 27 kb. Figure 1

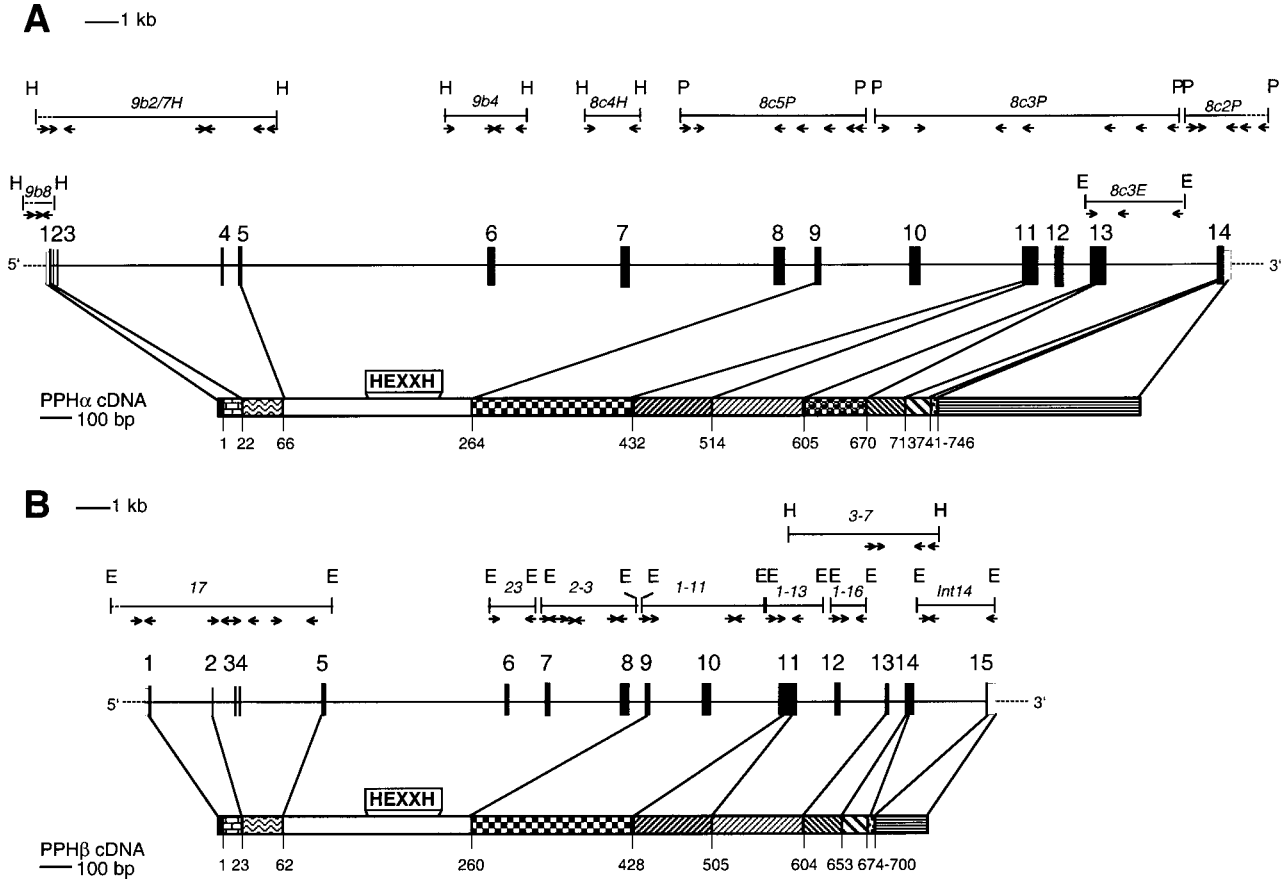

Figure 1 Organization of PPHα (A) and PPHβ (B) genes

The top line in each panel shows a schematic diagram of the respective gene. Exons are numbered and are represented as follows: black boxes, coding regions; open boxes, non-coding regions. Introns are presented as dashes. pKS clones containing genomic sequences are shown above the line. Arrows indicate the position and direction of primers used during the sequencing project. The bottom line in each panel demonstrates a schematic diagram of PPHα and PPHβ protein derived from the exons. The amino acid numbers at the beginning of each domain are also presented. Abbreviations: H, *Hin*dIII; E, *Eco*RI; P, *Pst*I. Regions of the protein are indicated as follows: ■, 5′ untranslated region; ▭, signal peptide; ▨, propeptide; □, protease; ▩, MAM (meprin subunit domain/A5 protein/receptor protein tyrosine phosphatase μ domain); ▨, MATH (meprin- and tumour-necrosis-factor-receptor-associated factors homology domain); ▨, intervening domain; ▩, I-domain (inserted domain); ▧, EGF-like domain; ◥, transmembrane domain; ▩, cytoplasmic domain; ▬, 3′ untranslated region.

Table 1 Intron–exon organization of the PPHα (a) and PPHβ (b) genes

The position and length of exons and introns were determined by sequencing. Asterisks indicate introns whose approximate sizes were detected by PCR. Amino acids are numbered from the initiation codon. The lengths of the 3′ non-coding region (NCR) of PPHα and PPHβ are based on the published cDNA sequences derived from 3′ RACE

(a)

Exon	Size (bp)	5′ splice donor	Intron	Size (kb)	3′ splice acceptor
1	5′ NCR(15) + 69	GTA CCG gtaagccg Val-Pro[23]	1	0.084	tttttcag ATT AAG CAT Ile[24]-Lys-His
2	25	GAA AAT G gtaagaat Glu-Asn[31]	2	0.093	tttaacag TA CAT GAT Val[32]-His-Asp
3	51	AAT TTA G gtgagttc Asn-Leu[48]	3	4.6*	acttccag CT GCA GGC Ala[49]-Ala-Gly
4	41	TTG CAG gtgagtac Leu-Gln[62]	4	0.4*	tttcacag AAA TCC AGA Lys[63]-Ser-Arg
5	76	AAT TTG G gtaatatt Asn-Leu[87]	5	7.5*	ttctgcag GG CTG AAT Gly[88]-Leu-Asn
6	118	TTT GAT GG gttggtat Phe-Asp-Gly[127]	6	3.0*	ctttaaag G TGC TGG Cys[128]-Trp
7	176	CTT TCA G gtgtgatt Leu-Ser[185]	7	4.5*	cctaacag GT TAC CAG Gly[186]-Tyr-Gln
8	222	AAT TGC A gtgagtat Asn-Cys[259]	8	0.862	atttgcag CC ACA ACT Thr[260]-Thr-Thr
9	150	TGC ACA G gtcagtga Cys-Thr[309]	9	2.9*	ttccatag GT GCC GGC Gly[310]-Ala-Gly
10	216	TTT CAA G gtacttag Phe-Gln[381]	10	3.5*	gtcctcag GA GAT GAT Gly[382]-Asp-Asp
11	465	TCT CCA G gtgggtgg Ser-Pro[536]	11	0.8*	ctttgcag CG ATA AAT Ala[537]-Ile-Asn
12	174	TTT GAA G gtactttt Phe-Glu[594]	12	0.498	tcctgaag AT ATC ACC Asp[595]-Ile-Thr
13	301	AGC TGC AG gtaggctc Ser-Cys-Arg[695]	13	3.3*	tccctcag G TGC ATC Cys[696]-Ile
14	157 + 3′ NCR(654)				

(b)

Exon	Size (bp)	5′ splice donor	Intron	Size (kb)	3′ splice acceptor
1	5′NCR(50) + 63	GGC TTG gtaaggaa Gly-Leu[21]	1	1.6*	tttcccag GCA ACT CCA Ala[22]-Thr-Pro
2	19	AAC TTT G gtgagtct Asn-Phe[27]	2	0.65*	aatcaaag AT GTA GAT Asp[28]-Val-Asp
3	45	AAT GAA G gtttgtgg Asn-Glu[42]	3	0.083	actcttag GT TTG GGA Gly[43]-Leu-Gly
4	44	GAT AGG gtgagttg Asp-Arg[57]	4	2.45*	tttgacag GCA CAA ATT Ala[58]-Gln-Ile
5	79	AGC TTG G gttagtat Ser-Leu[83]	5	5.9*	tccttcag AA ATG AAT Glu[84]-Met-Asn
6	118	GGC AGT GG gtaagttg Gly-Ser-Gly[123]	6	0.75*	cctgtaag C TGC TGG Cys[124]-Trp
7	179	CTG TCA G gtacattt Leu-Ser[182]	7	2.5*	tcctttag GC AGA GAG Gly[183]-Arg-Glu
8	219	AAC TGC T gtatgtga Asn-Cys[255]	8	0.5*	ttttgtag CC TCT TCC Ser[256]-Ser-Ser
9	153	TGC CAA G gtaacagg Cys-Gln[306]	9	1.9*	tttaacag GT TCT GGT Gly[307]-Ser-Gly
10	216	ATA AAA G gtacaat Ile-Lys[378]	10	2.2*	ttttgcag AA ATA CCC Glu[379]-Ile-Pro
11	444	ACC ACC G gttcgtaa Thr-Thr[526]	11	1.3*	tttttcag AT AAT GGA Asp[527]-Asn-Gly
12	180	GTG GAA G gtatgtca Val-Glu[586]	12	1.7*	cttcctag AC ATA TCT Asp[587]-Ile-Ser
13	127	GAG TGC AG gtaaggat Glu-Cys-Arg[629]	13	0.641	acatgcag G TGC CAG Cys[630]-Gln
14	205	CAA AAT gtaagttg Gln-Asn[697]	14	2.2*	tttagcag CAT GCT TTT His[698]-Ala-Phe
15	12 + 3′NCR(158)				

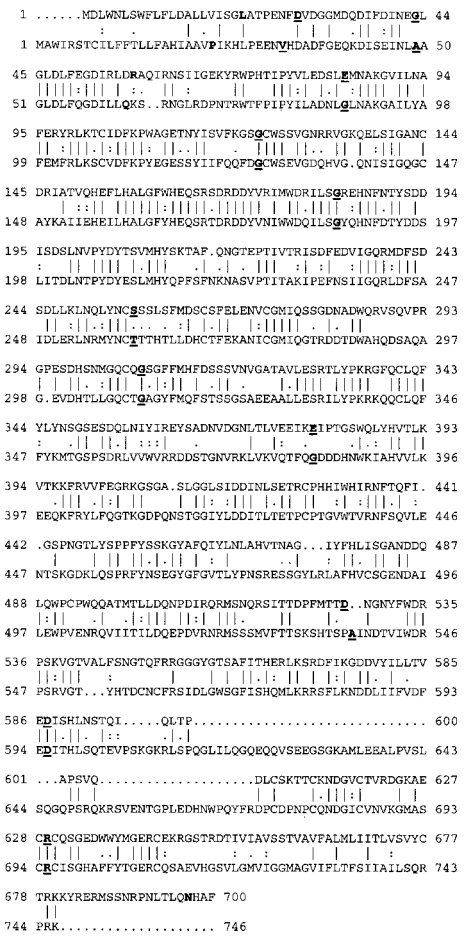

Figure 2 Optimal alignment of the amino acids of PPHβ (upper row) and PPHα (lower row)

Protein sequences were aligned by using the algorithm of Needleman and Wunsch (gap 12, length 3). Gaps (indicated by full points) were inserted for optimal alignment. A line is present between identical residues, two dots are present between chemically very similar residues, and one dot is present between similar residues. Amino acids whose codons are in front of phase 0 introns are indicated in bold; amino acids whose codons are interrupted by introns are also underlined.

shows a diagram of both genes. The exons of both genes mostly range in size from approx. 19 and 25 bp (exon 2 of PPHα and PPHβ respectively) to 465 and 444 bp (exon 11 of PPHα and PPHβ respectively). The exception is the last exon in PPHα, which is nearly 800 bp long and contains approx. 650 bp of non-coding 3′ sequences. Whereas the sizes of the equivalent exons in both genes are nearly identical, the intron sizes vary from 83 bp to approx. 7500 bp. Intron 1 of PPHα consists of only 84 bp, whereas the corresponding PPHβ intron has approx. 1500 bp. In contrast, intron 3 of PPHα is approx. 4600 bp in size, whereas the corresponding intron of PPHβ consists of only 83 bp. The sequences at the 5′ and 3′ ends of each intron are in agreement with the consensus sequence for intron/exon boundaries of other eukaryotic genes (Table 1) [21,22]. Although introns of phases 0, 1 and 2 are present, most of them are phase 1 introns. A cluster of exons is found at the 5′ ends of both genes; exons 1–7 are located within the first 500 bp of cDNA. Amino acid sequences and intron position alignments of both subunits (Figure 2) show a high conservation of intron positions, with the exception of introns 1 and 2 as well as intron 14 in PPHβ, which is not present in PPHα. Comparison between the gene structure and the domain organization of both subunits of PPH reveal that in both genes the first exon encodes the signal peptide sequence. The propeptide in PPHα, with the exception of the first two amino acids in exon 1, is encoded by the next four exons. In PPHβ exon 2 encodes for the last amino acid residue of the signal peptide and, together with exons 3, 4 and 5, also for the propeptide. In both subunits exons 5–9 encode the protease domain and exons 9–11 the MAM (meprin subunit domain/A5 protein/receptor protein tyrosine phosphatase μ) domain [23]. The MATH (meprin- and tumour-necrosis-factor-receptor-associated factors homology) domain [24] is completely encoded by exon 11, which together with exons 12 and 13 also encodes the intervening domain. In PPHα exon 13 also codes for the I domain, which is not present in PPHβ, and for the first 25 residues of the epidermal growth factor (EGF)-like domain. In PPHβ exon 13 is much smaller and codes for the last 16 residues of the intervening domain and 25 residues of the EGF-like domain. The last exon of PPHα encodes the rest of the EGF-like domain, the transmembrane domain and the short cytoplasmic tail of 6 residues. It also contains 654 bp of the non-coding region. In PPHβ the additional exon 15 encodes the last three residues of the cytoplasmic tail and contains the 3′ non-coding region.

Determination of transcriptional initiation sites

The initiation of transcription of both genes was determined by primer extension analysis and compared with sequencing reactions of two clones obtained by rapid amplification of 5′ cDNA ends (5′ RACE) with the same oligonucleotides [14]. For PPHα the primary start site seems to be an adenosine residue 15 nt upstream of the start codon (Figure 3, lane 3), a minor transcription start site seems to be 4 nt further upstream. For PPHβ, two transcription start sites were found (Figure 3, lane 2), one is 50 nt upstream of the start codon (A), the other 30 nt upstream of the ATG (also A). Whether the latter represents a major start site or a reverse-transcriptase pause is not known. These results are in agreement with the 5′ RACE data. The sequences around the adenosine at the primary start site in PPHα and the start site 50 nt in front of the ATG in PPHβ exactly match the consensus sequence $CA_{+1}NYYY$ calculated from the cap site of more than 500 eukaryotic promoters [25]. The nucleotides flanking the ATG start codons in PPHα and PPHβ at positions -3 and $+4$ are identical with the consensus sequence determined by Kozak [26] for the efficient initiation of translation. An alignment of both 5′ non-coding regions with an isoform of mouse meprinβ intestinal mRNA [27] is shown in Figure 4. The relatively short 5′ non-coding region of human PPHα shows no similarity to the 5′ non-coding regions of human or mouse β subunits. Alignment of the sequences of PPHβ and mouse meprinβ reveals that the transcription start site of mouse meprinβ corresponds to the adenosine at position -30 in human PPHβ, which was identified as a possible transcription starting point despite the sub-optimal fit of the surrounding nucleotides into the consensus cap sequence. The identity in this region between the two 5′ non-coding regions is 73%.

Nucleotide sequence of the 5′ flanking region of *PPH* genes

The 5′ upstream regions of both genes were searched for putative transcription-factor-binding sites with the Transcriptional Element Search System based on the TRANSFAC v. 3.3 database [28]. Figures 5(a) and 5(b) show putative transcription-factor-binding sites that might participate in gene expression. Sites with a quality of 1.0 within the first 700 nt in front of the ATG

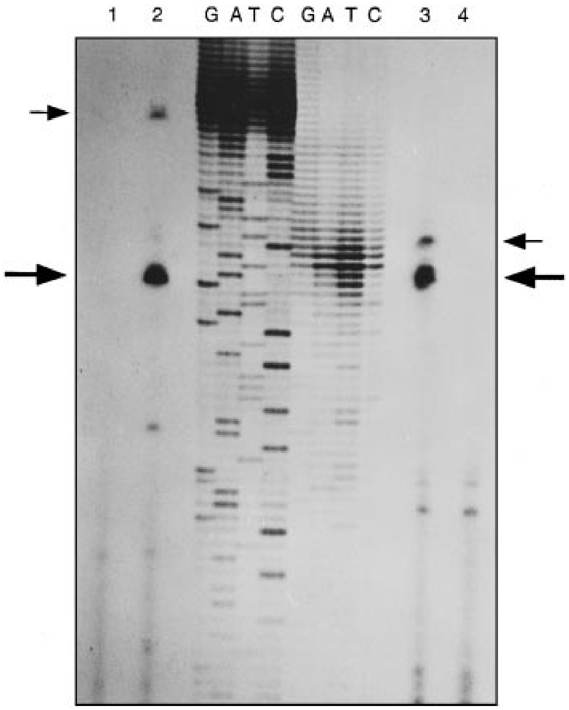

Figure 3 Mapping of the transcription initiation start site of PPHα and PPHβ genes

Total small-intestinal RNA was annealed to oligonucleotide primers identical with the anti-sense strand at position +52 to +72 for PPHα and +30 to +52 for PPHβ, then extended and analysed as described in the Experimental section. The products were separated in an 8% (w/v) polyacrylamide gel along with sequencing reactions of plasmids obtained by 5′ RACE with the same primers. Lanes 1 and 4, labelled oligonucleotide alone; lane 2, human intestinal RNA with PPHβ-specific oligonucleotide; lanes GATC from left to right, sequencing reactions of PPHβ-specific and PPHα-specific clones respectively obtained by 5′ RACE; lane 3, human intestinal RNA with PPHα-specific oligonucleotide. The large arrows point to the putative major transcription start site and the small arrows to the minor transcription start site in PPHα and PPHβ.

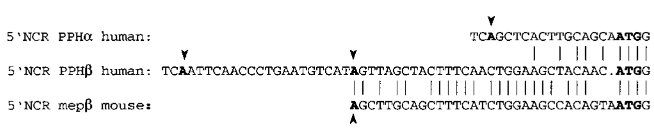

Figure 4 Sequence comparison of the 5′ non-coding regions (NCRs) of human PPHα (top row), human PPHβ (middle row) and mouse meprin β (bottom row) small-intestinal mRNA species

A gap (indicated by a full point) was introduced for optimal alignment. Vertical bars are located between identical nucleotides. The translation initiation codon is printed in bold, as well as the adenosine residues that together with arrowheads represent transcription initiation sites.

initiation codon are shown. At optimal positions from the transcription start sites [29], TATA-box and TATA-box-related elements can be identified at position −22 to −28 in PPHα and −27 to −32 in PPHβ respectively. Another regulatory element, the CAAT box, is located in reverse orientation at position −88 to −92 in PPHβ and at position −92 to −96 in PPHα. Half binding sites for the glucocorticoid receptor and progesterone receptor are present in both promoter regions analysed. Other binding sites identified include motifs that are recognized by the transcription factor polyoma-virus-enhancer-A-binding protein 3 (PEA3) [30], sites that are recognized by the astrocyte- and glioblastoma-specific octamer DNA-binding protein N-Oct3 [31] and signatures that are recognized by the liver- and intestine-specific factor hepatocyte nuclear factor 3 (HNF-3) [32,33]. In PPHα several GATA-binding sites are present that might be recognized by GATA-4, GATA-5 and GATA-6 [34,35]. PPHβ also contains half binding sites for the oestrogen receptor and for the activator protein 1, AP1. A schematic representation of the overall PPHα and PPHβ promoter regions, showing the location and widespread distribution of the transcription-factor-binding sites found, is shown in Figure 5(c).

Expression of promoter–reporter constructs in Caco-2 cells

The motifs mentioned above suggest that the expression of both PPH subunits can be regulated independently through interaction with different transcription factors. We have previously shown that Caco-2 cells endogenously express PPHα but not PPHβ, in a differentiation-dependent manner [17]. To test the presence of sequences essential for the expression of PPHα and sequences that negatively effect the expression of PPHβ in Caco-2 cells, we fused both promoters and different parts of the 5′ flanking region to the luciferase gene, transfected them into Caco-2 cells and estimated the luciferase activity 72 h after transfection. Results of the transfection experiments are shown in Table 2. The complete PPHβ promoter as well as nt −242 to +50 and nt −243 to −2358 were practically incapable of directing luciferase synthesis in Caco-2 cells. In contrast an approx. 200-fold induction over the promoterless pGL3 basic vector was observed with the PPHα promoter, exactly reflecting the situation in Caco-2 cells. Eliminating the PPHα promoter from nt −3426 to −1357 resulted in a decrease in luciferase activity to approx. 20% of the activity obtained with the full-length construct. Without the first 1356 nt no transcriptional activity was detectable.

DISCUSSION

In the present study we have characterized the structure of the genes coding for the PPH α and β subunits. Comparison of the genomic organization of the protease domains of mouse meprinα [36] with the two human subunits reveals a conserved match in the positions and phases of introns. Differences between these subunits are found only in intron size. Some of the introns are also conserved in other members of the astacin family such as tolloid [9], astacin [37], low choriolytic enzyme (LCE) [38] and human BMP1 [39]. Intron 5 in PPHα and PPHβ is present at the same position in meprinα and BMP1. In addition, intron 6 is also present in LCE, tolloid and astacin, and intron 7 is localized at the same position in all proteases mentioned except tolloid.

Jiang and Flannery [36] found a correlation of the exon–intron organization of the mouse meprinα protease domain with its secondary structure elements based on the astacin crystal structure [40]. With the determination of the genomic organization of the protease domain of human PPHα, these observations can now be extended.

The identity at the amino acid level between the same subunits across the three species (human, mouse and rat) is approx. 80%, whereas the similarity between the α and β subunits within a species is approx. 50%. This, added to the fact that intron positioning between the two subunits is largely conserved, suggests that distribution of both genes to different chromosomes by gene duplication had occurred before the divergence of the human and rodent lineage.

Comparison of the overall genomic structures permits speculations about the evolutionary relationship of the PPH subunits. We postulate that the ancient form of the α-subunit was the first

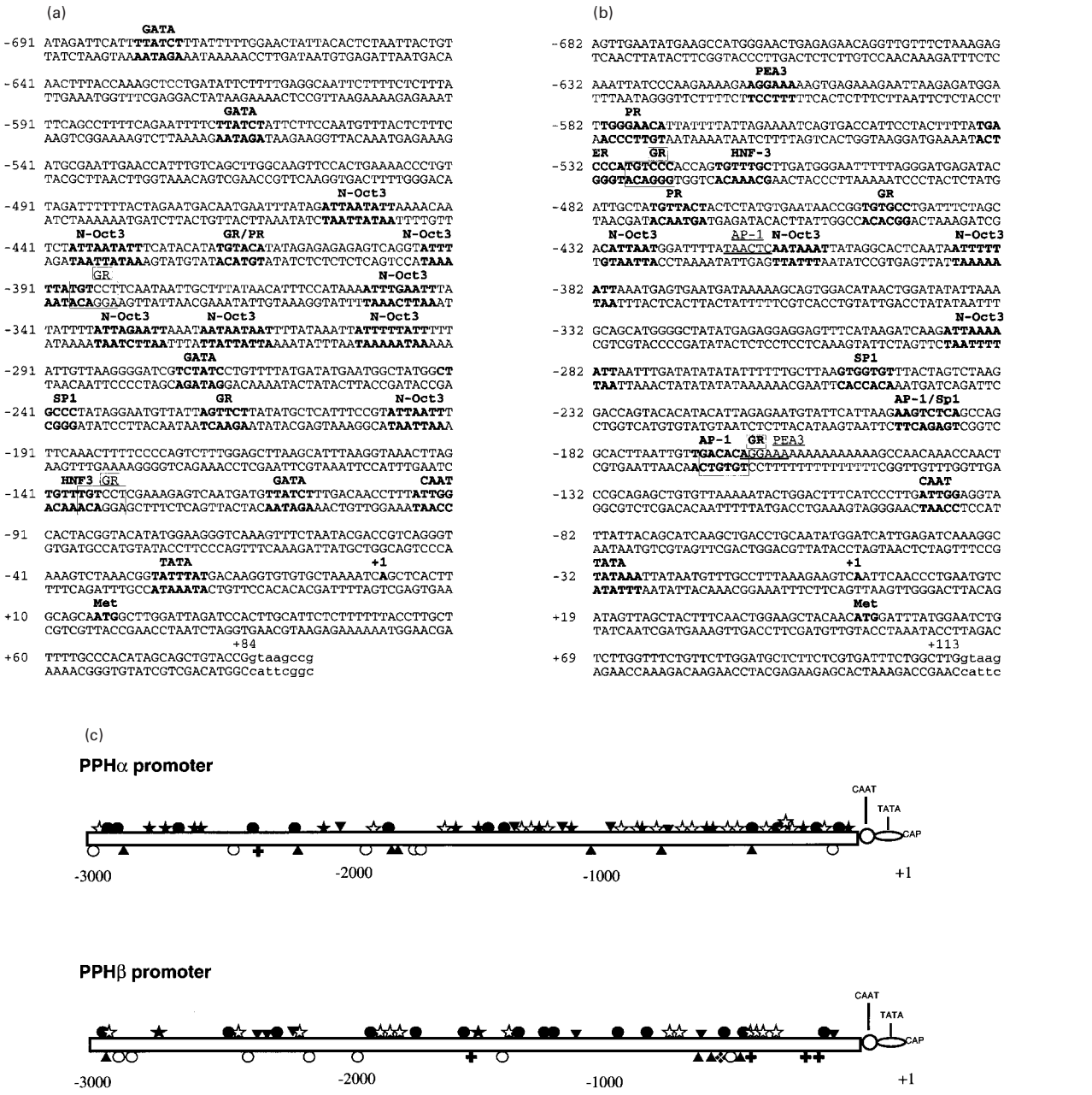

Figure 5 Nucleotide sequence of the 5′ flanking region of the PPHα (a) and PPHβ (b) genes and location of possible transcription-factor-binding sites spread over 3.0 kb of PPHα and PPHβ promoter (c)

(**a**, **b**) Position +1 denotes the putative transcription start site (cap site). The ATG translation initiation codon is shown in bold. Sequences with possible biological importance (see text) are also indicated in bold. The first exon/intron boundary located at +84 in PPHα and +113 in PPHβ is also shown. (**c**) Putative recognition sites for known transcription factors are represented by the following symbols in the overall promoter regions: ☆, N-Oct3 factor-binding site; ★, GATA-factor binding site; ●, glucocorticoid receptor (GR)-binding site; ○, HNF-binding-site; ✚, AP1-binding site; ▼, PEA3-binding site; ▲, progesterone receptor (PR)-binding site; ❖, oestrogen receptor (ER)-binding site.

of the mammalian meprins. It is the larger of the two, which is consistent with the finding in other gene families that the oldest members have accumulated the largest amount of intron DNA [41]. Despite its larger size, PPHα has one exon less than PPHβ owing to the possible loss of exon 15 during evolution. The specific I domain of PPHα is not encoded by a separate exon but is part of exon 13 and thus reflects the possibility of a loss of exon sequences in PPHβ rather than the insertion of these sequences in PPHα. Gene mutations, insertions and deletions, although keeping constant exon–intron positions, consequently led to the formation of two related but distinct protease subunits with divergent functions. The I domain, which is necessary for the proteolytic removal of the transmembrane domain in PPHα, was lost in PPHβ and resulted in the localization of a previously soluble protease to the cell membrane. In contrast, the possibility of regulating the proteolytic activity and/or secretion of PPHα by phosphorylation was lost by the truncation of the cytoplasmic tail [42].

Table 2 Luciferase activities of DNA constructs carrying 5′ upstream regions of the PPHα and PPHβ genes

Mixtures of 0.37 pmol of various reporter constructs and 250 ng of pSV β-galactosidase control vector were transfected into Caco-2 cells. At 72 h after transfection, cell extracts were prepared and subjected to luciferase assays. The luciferase activities are expressed as fold induction over activities obtained with promoterless vector pGL3 after normalization of the experimental values based on transfection efficiencies. Results are means $\pm$ S.D.

Promoter fragment	Relative luciferase activity (fold induction over pGL3)
PPHα -3426 to $+10$	230.5 ± 18.8
PPHα -2008 to -1357	1.3 ± 0.6
PPHα -1356 to $+10$	43.9 ± 5.9
PPHβ -3893 to $+50$	3.1 ± 1.6
PPHβ -2358 to -243	13.2 ± 3.9
PPHβ -242 to $+50$	2.1 ± 0.4

The basis of tissue-, tumour- and development-specific expression of PPHα and PPHβ is of special interest. Both subunits are expressed in enterocytes of small-intestinal epithelial cells [1,16]. We have sequenced approx. 3.0 kb upstream of both transcription initiation sites. Spread over this 5′ region there are a considerable number of putative binding sites that could interact with transcription factors expressed in different tissues, at different developmental stages and in different cell types. The inverted CAAT box found in both PPHα and PPHβ is also present in the promoter region of endopeptidase 24.11 [43], a peptidase of the gluzincin family of metallopeptidases [44]. Endopeptidase 24.11 is expressed in many cell types including intestinal epithelial cells and might be involved in the inactivation of biological active peptides [43]. It has also been reported that these inverted CAAT boxes are present in promoters of genes controlling the initiation of the cell cycle [45]. In PPHα a large number of optimal binding sites for GATA transcription factors exist. In contrast, optimal binding sites are only found far upstream of the transcription initiation site in PPHβ. GATA-4, GATA-5 and GATA-6 are expressed in the developing and adult intestine and are thought to function in regulating the terminal differentiation programme of intestinal epithelial cells [35]. Because other astacin metalloproteases are involved in proteolytic processes during early development [9,46,47], the interaction of GATA-binding proteins with the PPHα and perhaps the PPHβ promoter and the resulting influence of PPH on gut development is an interesting possibility. In this context it is conceivable that GATA factors, especially GATA-6, which has been shown to be expressed in Caco-2 cells [35], are involved in the tissue-specific expression of PPHα in Caco-2 cells. This is endorsed by our findings that (1) only the PPHα promoter is capable of directing luciferase expression in Caco-2 cells and (2) transcription-factor-binding sites in the far upstream region of the PPHα promoter (-1350 to -3430) are necessary for the increased enhancement of luciferase expression in Caco-2 cells. This rationale can also be extended to the observation that only PPHα is expressed in colon epithelial cells [17]. The identification, in both promoter regions, of binding sites for HNF-3, a transcription factor of the steroid receptor family that is expressed in liver, intestine and lung, suggests an involvement in the differentiation of cells [33]. Spencer-Dene et al. [48] have identified rat meprin in neuroepithelial cells of rat embryos. The presence of several binding sites for N-Oct3, a transcription factor that is expressed in the central nervous system during development and also in adult brain [31], and of sites for the transcription factor PEA3, which is abundantly found in brain [30], might imply a participitation of PPH in proteolytic processes during brain development and/or tumour growth. The involvement of PPHβ in tumour progression can also be inferred by the presence of AP1-binding sites. The ubiquitous factor AP1 regulates the transcription of genes involved in cell proliferation in response to external stimuli such as growth factors and tumour-promoting agents such as phorbol esters [49]. Matters and Bond have shown recently [50] that the treatment of HT29-18 C1 cells, a colon cancer cell line, with PMA significantly increases the transcription of human PPHβ' mRNA, an isoform of PPHβ mRNA.

Several putative binding sites for the glucocorticoid receptor are present in both promoter regions. The possibility that PPH genes might be regulated by steroid hormones is particularly interesting in the context of a possible overexpression of this protease in intestinal inflammation and its treatment with dexamethasone. We have recently demonstrated that PPH is expressed in leucocytes of the lamina propria [16] in addition to intestinal epithelial cells. This expression pattern is markedly increased in inflamed intestinal mucosa of patients with active coeliac disease (D. Lottaz, unpublished work), leading to an enhanced proteolytic potential in the subepithelial tissue.

Further studies investigating these important regulatory proteins and their interaction with PPHα and PPHβ promoter sequences under physiological and pathophysiological conditions will provide new insights into the participitation of PPH in gut development, cancer cell growth or gut inflammation.

We thank U. Luginbühl and E. Mandac for technical assistance, and Dr. T. Brunner and Dr. A. Stocker for helpful comments and computer analysis. This work was supported by grant 32-00-52736.97 from the Swiss National Science Foundation (E. E. S.).

REFERENCES

1 Sterchi, E. E., Naim, H. Y., Lentze, M. J., Hauri, H. P. and Fransen, J. A. (1988) Arch. Biochem. Biophys. **265**, 105–118
2 Sterchi, E. E. (1998) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.), pp. 1229–1232, Academic Press, London
3 Dumermuth, E., Sterchi, E. E., Jiang, W. P., Wolz, R. L., Bond, J. S., Flannery, A. V. and Beynon, R. J. (1991) J. Biol. Chem. **266**, 21381–21385
4 Bode, W., Gomis-Rüth, F. X. and Stöcker, W. (1993) FEBS Lett. **331**, 134–140
5 Wozney, J. M., Rosen, V., Celeste, A. J., Mitsock, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M. and Wang, E. A. (1988) Science **242**, 1528–1534
6 Yasumasu, S., Yamada, K., Akasaka, K., Mitsunaga, K., Iuchi, I., Shimada, H. and Yamagami, K. (1992) Dev. Biol. **153**, 250-8
7 Elaroussi, M. A. and DeLuca, H. F. (1994) Biochim. Biophys. Acta **1217**, 1–8
8 Shimell, M. J., Ferguson, E. L., Childs, S. R. and O'Connor, M. B. (1991) Cell **67**, 469–481
9 Finelli, A. L., Bossie, C. A., Xie, T. and Padgett, R. W. (1994) Development **120**, 861–870
10 Reynolds, S.D., Angerer, L. M., Palis, J., Nasir, A. and Angerer, R. C. (1992) Development **114**, 769–786
11 Kaushal, G. P., Walker, P. D. and Shah, S. V. (1994) J. Cell. Biol. **126**, 1319–1327
12 Walker, P. D., Kaushal, G. P. and Shah, S. V. (1998) Kidney Int. **53**, 1673–1680
13 Dumermuth, E., Eldering, J. A., Grünberg, J., Jiang, W. and Sterchi, E. E. (1993) FEBS Lett. **335**, 367–375
14 Eldering, J. A., Grünberg, J., Hahn, D., Croes, H. J., Fransen, J. A. and Sterchi, E. E. (1997) Eur. J. Biochem. **247**, 920–932
15 Bond, J. S., Rojas, K., Overhauser, J., Zoghbi, H. Y. and Jiang, W. (1995) Genomics **25**, 300–303
16 Lottaz, D., Hahn, D., Müller, S., Müller, C. and Sterchi, E. E. (1999) Eur. J. Biochem. **259**, 496–504
17 Lottaz, D., Maurer, C. A., Hahn, D., Büchler, M. W. and Sterchi, E. E. (1999) Cancer Res. **59**, 1127–1133
18 Lehrach, H. (1990) in Genome Analysis, vol. 1 (Davies, K. E. and Nitzetic, D., eds.), pp. 39–81, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

20 Triezenberg, S. J. (1995) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 4.8.1–4.8.5, John Wiley and Sons, New York
21 Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
22 Shapiro, M. B. and Senapathy, P. (1987) Nucleic Acids Res. **15**, 7155–7174
23 Beckmann, G. and Bork, P. (1993) Trends. Biochem. Sci. **18**, 40–41
24 Uren, A. G. and Vaux, D. L. (1996) Trends Biochem. Sci. **21**, 244–245
25 Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S. T. (1994) Mol. Cell. Biol. **14**, 116–127
26 Kozak, M. (1989) J. Cell. Biol. **108**, 229–241
27 Dietrich, J. M., Jiang, W. and Bond, J. S. (1996) J. Biol. Chem. **271**, 2271–2278
28 Schug, J. and Overton, G. C. (1997) Transcriptional Element Search System (TESS), Technical Report CBIL-TR-1997-1001-v 0.0, Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, Philadelphia, PA
29 Bucher, P. (1990) J. Mol. Biol. **212**, 563–578
30 Petersohn, D., Schoch, S., Brinkmann, D. R. and Thiel, G. (1995) J. Biol. Chem. **270**, 24361–24369
31 Atanasoski, S., Toldo, S. S., Malipiero, U., Schreiber, E., Fries, R. and Fontana, A. (1995) Genomics **26**, 272–280
32 Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E. and Costa, R. H. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 3948–3952
33 Lai, E., Prezioso, V. R., Tao, W. F., Chen, W. S. and Darnell, Jr., J. E. (1991) Genes Dev. **5**, 416–427
34 Laverriere, A. C., MacNeill, C., Mueller, C., Poelmann, R. E., Burch, J. B. and Evans, T. (1994) J. Biol. Chem. **269**, 23177–23184
35 Gao, X., Sedgwick, T., Shi, Y. B. and Evans, T. (1998) Mol. Cell. Biol. **18**, 2901–2911
36 Jiang, W. and Flannery, A. V. (1997) Gene **189**, 65–71
37 Geier, G., Jacob, E., Stöcker, W. and Zwilling, R. (1997) Arch. Biochem. Biophys. **337**, 300–307
38 Yasumasu, S., Shimada, H., Inohaya, K., Yamazaki, K., Iuchi, I., Yasumasu, I. and Yamagami, K. (1996) Eur. J. Biochem. **237**, 752–758
39 Takahara, K., Lee, S., Wood, S. and Greenspan, D. S. (1995) Genomics **29**, 9–15
40 Stöcker, W., Gomis-Rüth, F. X., Bode, W. and Zwilling, R. (1993) Eur. J. Biochem. **214**, 215–231
41 Cooke, N. E. and Baxter, J. D. (1982) Nature (London) **297**, 603–606
42 Kemp, B. E. and Pearson, R. B. (1990) Trends Biochem. Sci. **15**, 342–346
43 Crine, P., Dion, N. and Boileau, G. (1997) in Cell-Surface Peptidases in Health and Disease (Kenny, A. J. and Boustead, C. M., eds.), pp. 79–98, BIOS Scientific Publishers, Oxford
44 Hooper, N. M. (1994) FEBS Lett. **354**, 1–6
45 Knight, G. B., Gudas, J. M. and Pardee, A. B. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8350–8354
46 Finelli, A. L., Xie, T., Bossie, C. A., Blackman, R. K. and Padgett, R. W. (1995) Genetics (Princeton) **141**, 271–281
47 Lepage, T., Ghiglione, C. and Gache, C. (1992) Development **114**, 147–163
48 Spencer-Dene, B., Thorogood, P., Nair, S., Kenny, A. J., Harris, M. and Henderson, B. (1994) Development **120**, 3213–3226
49 Angel, P. and Karin, M. (1992) Matrix Suppl. **1**, 156–164
50 Matters, G. L. and Bond, J. S. (1999) APMIS **107**, 19–27

Received 18 August 1999/1 November 1999; accepted 25 November 1999