J. Biochem. 122, 438-452 (1997)

Genomic Organization and Alternative Splicing of Human PACE4 (SPC4), Kexin-Like Processing Endoprotease[1]

Akihiko Tsuji, Chiemi Hine,[2] Yasuhiro Tamai, Kayoko Yonemoto, Kazushige Mori, Shigeru Yoshida, Miwa Bando, Eri Sakai, Kenji Mori, Tetsuya Akamatsu, and Yoshiko Matsuda[3]

Department of Biological Science and Technology, Faculty of Engineering, University of Tokushima, Minamijosanjima 2-1, Tokushima 770

Received for publication, April 14, 1997

PACE4 (paired basic amino acid cleaving enzyme) is a member of a family of the mammalian kexin-like proprotein convertases containing a subtilisin-like catalytic domain. Previously we reported seven isoform mRNAs of PACE4 that vary in size and 3′-coding sequence [A. Tsuji *et al.* (1994) *Biochem. Biophys. Res. Commun.* 200, 943-950; K. Mori *et al.* (1997) *J. Biochem.* 121, 941-948]. To determine the origin of these isoforms, the entire human PACE4 gene has been isolated as a set of overlapping genomic DNA fragments, and analyzed by restriction enzyme digestion and nucleotide sequence determination. The human PACE4 gene spans at least 250 kb and is distributed over 25 exons that range in size from 39 to 1,422 base pairs. Human PACE4 gene is the largest kexin-like proprotein convertase gene reported to date. The most striking feature of its genomic structure is the size of the introns and the number of exons, although the general organization of signal peptide, propeptide, and catalytic domains, which are conserved in this family, is very similar to that reported for other kexin-like protease genes. The structural analysis of PACE4 genomic DNA indicates that multiple PACE4 transcripts are produced as a consequence of alternative RNA splicing events, including exon skipping, and differences in the usage of the inner 5′-splicing donor and polyadenylation sites. A major transcriptional start site was detected 314 bp upstream from the ATG translational start site by primer extension analysis. Sequence analysis of the 5′-flanking region revealed that PACE4 gene lacks TATA and CCAAT boxes in the proximal upstream region of the start site, although potential binding sites for several transcription factors including SP1, AP1, AP2, PEA3, Ets-1, GHF (growth hormone factor)-1, CREB (cyclic AMP response element binding protein), and basic helix-loop-helix proteins, were present. An unusual sequence of six tandem repeats of a nonadecamer (GGCCTGGGGGTTCACCTGC) containing an E box is found in the 5′-flanking region. These results suggest that PACE4 is not a constitutive gene product and its expression is regulated by various transcription factors.

Key words: alternative splicing, isoform, kexin, PACE4, SPC4.

The post-translational processing of prohormones and proneuropeptides requires proteolytic cleavage at paired basic amino acids in order to release bioactive peptides (*1, 2*). Processing at monobasic sites also occurs for a minority of bioactive peptide. Besides peptide hormone, growth factor, growth factor receptor, plasma protein, and viral glycoprotein are all derived *via* processing at dibasic sites (*2-4*). Recently a novel family of proteases involved in propeptide processing at dibasic sites has been identified. The first known endoprotease of this family was kexin [EC 3.4.21.61], a *Kex2* gene product of the yeast *Saccharomyces cerevisae*; this enzyme is a Ca^{2+}-dependent serine protease which has a bacterial subtilisin-like catalytic domain and is involved in processing of pro-α-mating factor and pro-killer toxin (*5*). To date, seven kinds of mammalian kexin-like protease including furin (*6*), PC1/3 (*7*), PC2 (*8*), PACE4 (*9-12*), also known as PC7 (rat PACE4) (*13*), PC4 (*14*), PC5/6 (*15, 16*), and LPC/PC8 (*17-20*), have been isolated by cDNA cloning. Since their catalytic domains resemble that of subtilisin, these proteases are also called subtilisin-like proprotein convertases (SPC); SPC1 to SPC7, respectively. The structural organization of the mammalian kexin-family proteases appears to be highly homologous. All have a signal peptide, propeptide, subtilisin-like catalytic domain, and homoB domain.

[1] This work was supported in part by Grants-in-Aid for Scientific Research (C) and Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan. The nucleotide sequences reported in this paper will appear in the DDBJ/EMBL/GenBank DNA databases with accession numbers, AB001898-AB001914.

[2] Present address: National Cardiovascular Center Research Institute, Fujishirodai 5, Suita, Osaka 565.

[3] To whom correspondence should be addressed. Tel: +81-886-56-7523, Fax: +81-886-55-3161, E-mail: matsuda@bio.tokushima-u.ac.jp

Abbreviations: PACE, paired basic amino acid cleaving enzyme; SPC, subtilisin-like proprotein convertase; SP, signal peptide; SCD, subtilisin-like catalytic domain; CRR, cysteine-rich region; bHLH, basic helix-loop-helix.

The propeptide domain is considered to act as an intramolecular chaperone assisting the folding of the zymogen within the endoplasmic reticulum. The catalytic domain contains the active site residues, Asp, His, and Ser which are characteristic of serine-proteases. The homoB domain has been shown to be essential for proteolytic activity (*21*). However the carboxy terminal region varies greatly in length and sequence among kexin-family proteases. Moreover, coexpression experiments of these proteases with various propeptides such as proopiomelanocortin showed differing amino acid requirements around the cleavage site (*22*).

Briefly, mammalian kexin-family proteases are classified into three groups on the basis of their tissue and cell distributions. The first group including furin and LPC/PC8, which have a transmembrane domain at the carboxy-terminus, exhibits a ubiquitous distribution in tissues and is proposed to be responsible for processing of precursors for various proteins within the constitutive secretory pathway (*6, 17-19, 23*). By contrast, the second group including PC1, PC2, and PC4 shows a restricted distribution. For example, PC1 and PC2 are predominantly detected in neuroendocrine cells of the central nervous system and endocrine cells of the pancreas, thyroid, and adrenal gland, suggesting that these enzymes are responsible for processing of peptide hormones within regulated secretory pathways (*24-26*). PC4 is expressed only in testis spermatogenic cells (*14, 27*). On the other hand, members of the third group including PACE4 (PACE4A) and PC6, which have a long cysteine-rich region at the carboxy-terminus, exhibit a unique restricted distribution in both nonendocrine and endocrine cells (*23, 25, 28, 29*). Previously, a widespread distribution of human PACE4 mRNA was reported (*16*), so PACE4 had been grouped with furin until recently. However, rat PACE4 was shown to be expressed at high levels in neuroendocrine tissues such as anterior pituitary (*10, 13, 23, 28*). We found a high level of expression of PACE4 in mitral cells of rat olfactory bulb and developmental regulation (*30*). Moreover, Constam *et al.* reported that mouse PACE4 exhibit highly regulated expression patterns during embryogenesis (*20*). These results suggest unique physiological functions for PACE4. However the mechanisms by which differential expression of PACE4 is controlled are not known.

Previously we isolated two novel cDNAs encoding PACE4 isoforms (PACE4C and PACE4D) from a human placenta cDNA library (*11*). PACE4A and PACE4B were first identified as PACE4 and PACE4.1, respectively, by Kiefer *et al.* (*9*). The protein encoded by PACE4A contained a signal peptide (SP), propeptide, subtilisin-like catalytic domain (SCD), homoB domain, and cysteine-rich region (CRR). PACE4B lacks homoB and a cysteine-rich region. PACE4C lacks a cysteine-rich region. PACE4D lacks a signal peptide, propeptide and a cysteine-rich region. These PACE4 isoforms have isoform-specific amino acid sequences at the carboxy-terminus. We demonstrated tissue-specific expression of PACE4 isoforms by immunohistochemistry and RT-PCR, and suggested cell-specific functions of the isoforms (*13, 25, 30*). Recently we identified a novel PACE4 isoform (PACE4E) cDNA (*31*). PACE4E has a shorter cysteine-rich region than PACE4A and contains a specific amino acid sequence at the carboxy-terminus. PACE4A and PACE4E are further classified as type I and type II. Type II is missing 39 bp just in front of the cysteine-rich region (*13, 31*). Zhong *et al.* reported a new isoform, PACE4CS which is a C-terminally truncated form of PACE4C (*32*). Thus, to date, eight isoforms of human PACE4 have been reported. Figure 1 summarizes

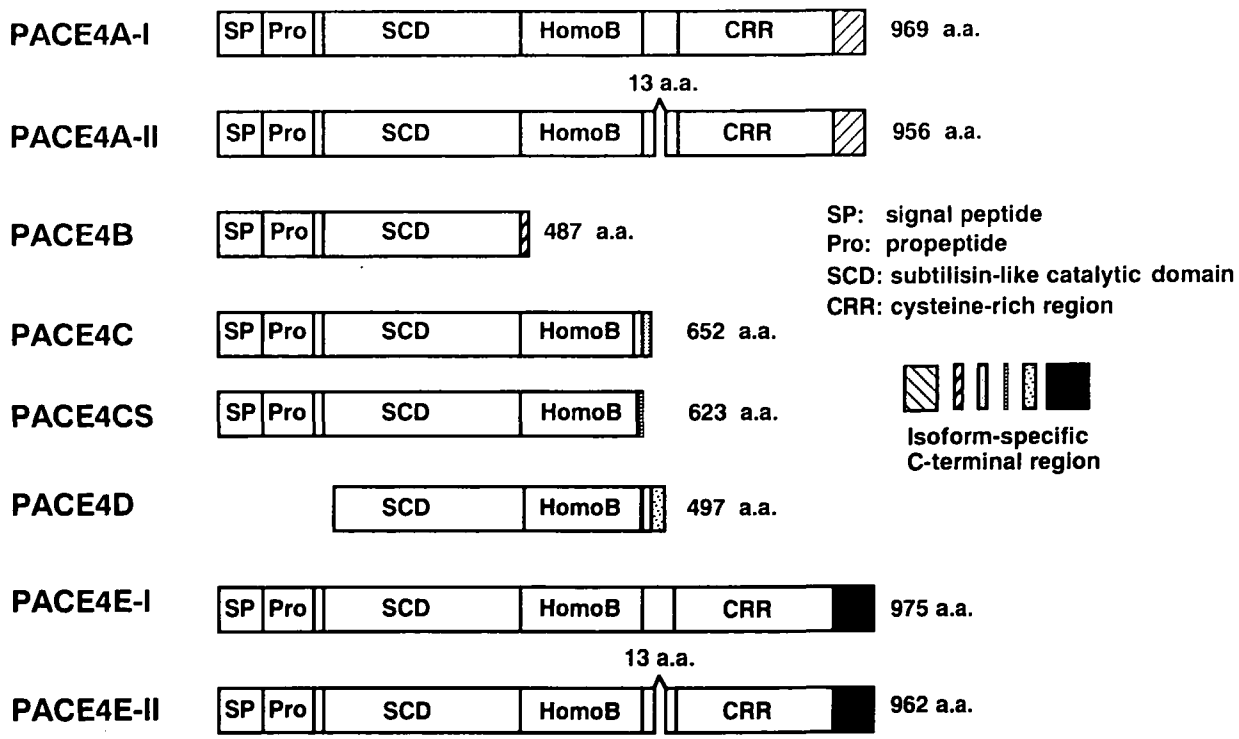

Fig. 1. Schematic domain structure of human PACE4 isoforms.

the predicted domain structures of these eight isoforms. It was presumed that the isoforms are products of alternative splicing of primary transcripts of a single PACE4 gene, though the mechanism involved has not been elucidated. In contrast, the carboxy-terminal structure appears to enhance delivery or storage of kexin-family proteases (*33, 34*). PACE4 isoforms have distinct carboxy terminal structures, which suggests that the intracellular localizations and functions of these isoforms differ.

As a first step in elucidating the mechanisms involved in the regulation of human PACE4 isoform expression, we have characterized the genomic organization of the human PACE4 gene and the sequence of its 5′-flanking region.

MATERIALS AND METHODS

Materials—Restriction endonucleases, DNA-modifying enzymes and reagents for dideoxy sequencing were purchased from Takara Shuzo (Kyoto). [α-^{32}P]dCTP and [γ-^{32}P]ATP were from Amersham (Buckinghamshire, UK). The other reagents used were of the highest grade available.

Southern Blot Analysis of Total Genomic DNA—Genomic DNA (10 μg) isolated from human peripheral blood leukocyte was digested with various restriction enzymes, resolved in a 0.8% agarose gel and transferred to a nylon membrane (GeneScreen Plus, Dupont/NEN, Boston, USA) by capillary blotting. Probes were labeled with [α-^{32}P]-dCTP using a *Bca*Best labeling kit (Takara, Kyoto). Prehybridization, hybridization with ^{32}P-labeled probe, and washing were carried out as described by the manufacturer of the Genescreen Plus membrane. The membrane was then exposed to X-ray film (Konica, Tokyo) at −80°C with two intensifying screens.

Mapping of the Gene Locus—Human chromosome mapping was performed using the BIOSMAP somatic cell hybrid panel (BIOS Laboratories, New Haven, CT, USA). The panel is composed of 20 hybrid cell lines containing human chromosomes: Cell line 010, 10; line 016, 16; line 212, Y; line 324, 18; line 423, 3; line 683, 1 (>30%), 5 (multiple deletion, MD), 12, 14, 19, 21, 22; line 734, 5, 9; line 750, 5, 13, 14, 15, 19; line 756, 5 (MD), 6, 7, 12 (5-30%), 13, 14, 19, 20, 21, Y; line 803, 4 (5-30%), 5, 8, 22, X; line 811, 8, 17, 18; line 852, 2; line 867, 1, 5, 13, 14, 18, 19; line 909, 5 (MD), 6, 8, 14, X; line 937, 1, 5, 14, 15 (5-30%), 17 (5-30%), 21; line 940, 5, 20; line 1006, 4, 5, 7, 13, 15, 19, 21, Y (5-30%); line 1049, 5, 11; line 1079, 3, 5; line 1099, 1, 5 (MD), 13, 19, 21, 22. Hamster/human hybrid DNAs (5 μg) were completely digested with restriction enzyme, size-separated on a 0.8% agarose gel and transferred onto nylon membrane. Then these membranes were hybridized with ^{32}P-labeled isoform-specific probes.

Probe Preparation and Genomic Library Screening—Eight kinds of human PACE4 cDNA fragment (5′-4D, SP, SCD, HomoB, CRR, A-3′, CD, and C-3′) were used as screening probes. 5′-4D was a 0.4 kb *Sal*I-*Dra*I restriction fragment (nucleotides −545 to −124 in PACE4D cDNA) containing the D-specific sequence. The SP probe was a 0.54 kb *Kpn*I restriction fragment (nucleotides −27 to 510 in PACE4 cDNA) containing signal peptide and propeptide. The SCD (nt 928 to 1344) and homoB (nt 1372 to 1836) probes were synthesized by polymerase chain reaction (PCR) amplification as described previously (*11, 13*). CRR probe (nt 1760 to 3162 in PACE4A cDNA) was a 1.4 kb *Eco*RI restriction fragment corresponding to the entire cysteine-rich region and part of the homoB and 3′-untranslated region. The A-3′ probe was a 1.1 kb *Eco*RI-*Sal*I restriction fragment (nt 3163 to 4234 in PACE4A cDNA) containing 3′-untranslated sequence. The CD probe was a 0.5 kb *Eco*RI fragment (nt 1760 to 2283 in PACE4C cDNA) containing 46 bp of CD-common sequence, 0.1 kb of homoB and 0.38 kb of C-specific sequences. C-3′ was a 0.7 kb *Eco*RI-*Sal*I restriction fragment of PACE4C (nt 2284-2950) containing 3′-untranslated C-specific sequence. To isolate the genomic DNA encoding 5′-untranslated region and signal peptide, 5′-labeled antisense oligonucleotide mixtures (5′-UTR, ATG and 144 oligonucleotide) were also used as screening probes. 5′-UTR, ATG and 144 oligoprobe corresponded to nucleotides −74 to −51 (5′-GAGTGCCG-CCTTTTAAAGCCGCTC-3′), −17 to 7 (5′-AGGCGGCGC-GCGCGGAGGCATAGCGGCGACAGGCT-3′), and 144 to 171 (5′-CAGCGCCAGCAGCAGCAGCCAGCGCCAG-3′) in PACE4 cDNA, respectively. The 5′-^{32}P-labeling of oligonucleotides was performed using polynucleotide kinase.

Human leukocyte genomic library cloned into the bacteriophage EMBL3 was purchased from CLONTECH Laboratories (Palo Alto, CA, USA) and screened with various ^{32}P-labeled PACE4 cDNA probes to obtain overlapping clones of the PACE4 gene. Approximately 10^6 clones were screened by plaque hybridization using Colony/Plaque Screen nylon membrane (Dupont/NEN) according to the manufacturer's protocol. The positive clones were purified by three rounds of hybridization and further hybridized with a variety of other probes to confirm the organization of the PACE4 gene. Phage DNAs purified from the positive clones were analyzed by restriction enzyme mapping and Southern blotting. Restriction fragments from these clones were subcloned into pUC 18 and sequenced by the dideoxy termination method using a *Bca*Best sequencing kit (Takara, Kyoto) and an ALF Red automated DNA sequencer (Pharmacia, Uppsala, Sweden).

A human PAC genomic library was screened by Genome Systems (St. Louis, MO, USA) with a PCR product generated from oligonucleotide primers designed from the sequence of λhPA271 clone, which was isolated with PACE4D-specific probe from an EMBL3 genomic library. The sequences of the sense and antisense primers are 5′-AGCCGTTTTGTCCACTGATCTATG-3′ and 5′-GCAT-TCAGTCTTCTAGCATTTGGA-3′, respectively. PCR consisted of 30 cycles (95°C 1 min, 60°C 1 min, 72°C 2 min). The amplified fragment (0.8 kb) contains the specific sequence for PACE4D cDNA located at the 5′-terminus. Plasmid DNA isolated from a positive PAC clone was analyzed by restriction enzyme mapping and sequencing.

PCR Amplification of Genomic DNA Encoding the 5′-Upstream Region—PCR was performed using oligonucleotide primers corresponding to nucleotides −892 to −874 (sense primer, 5′-TGACCTCATCCTCAAATCAGCCCG-3′) and nucleotides −74 to −51 (antisense primer, 5′-CA-GTGCCGCCTTTTAAAGCCGCTC-3′) of PACE4A cDNA. Human genomic DNA was amplified in 50 μl of mixture containing 0.2 μM primers, 0.5 μg of human genomic DNA, 10% dimethyl sulfoxide, 1×Ex *Taq* polymerase buffer, 1 mM each dNTP (7-deaza dGTP: dGTP 3:1), and 5 units of Ex *Taq* polymerase (Takara) for 30 cycles (1 min at 95°C, 1 min at 60°C, 2 min at 70°C). The amplified fragment (856

bp) was ligated into the pCRII™ vector (Invitrogen, San Diego, CA, USA) and sequenced.

Primer Extension Analysis—The transcription initiation site was mapped by primer extension using a 24 mer oligonucleotide (5′-GAGCGGCTTTAAAAGGCGGCACTC-3′) complementary to nucleotides −74 to −51 of PACE4A cDNA. This primer was labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. The labeled primer (1 pmol) was annealed to 25 μg of total HepG2 RNA by incubation overnight at 42°C in hybridization solution (80% formamide, 100 mM sodium citrate, pH 6.4, 300 mM sodium acetate, 1 mM EDTA). After precipitation of the annealed primer and template, the hybridized primer was extended by incubation at 42°C for 1 h with 10 units of avian myeloblastosis virus reverse transcriptase in 20 μl of extension buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM each dNTP, 2.8 mM spermine, and 2.8 mM sodium pyrophosphate). The reaction was terminated by the addition of 1 μl of 0.5 M EDTA and 1 μl of DNase-free pancreatic RNase for 30 min at 37°C. After phenol-chloroform extraction, the sample was precipitated by the addition of ethanol and denatured. The primer extension product was then analyzed on an 8% polyacrylamide-7 M urea gel in parallel with sequencing reactions using identical oligonucleotide primer and a 4.4 kb *Bam*HI genomic fragment from λhPA OL-16 containing the first exon and 5′-untranslated region. The gel was dried and exposed to X-ray film at −80°C with two intensifying screens for 9 days.

RESULTS

Mapping of Gene Locus—To obtain a chromosomal assignment for the genes encoding the novel isoforms, PACE4C and PACE4D, the isoform-specific cDNA probe was hybridized to the Southern blot of the panel of somatic cell hybrids. The panel of DNAs from hamster/human cell hybrids contained distinctive complements of human chromosomes. As shown in Fig. 2, human PACE4C and PACE-4D-specific bands were detected in cell line 750 (lane 08) and 1006 (lane 17), as was the case for PACE4A. The results showed complete concordance between human chromosome 15 and all types of PACE4. These were the only cell lines to contain human chromosome 15. Human PACE4 gene was localized at chromosome 15q26 by Kiefer *et al.* (*9*). Therefore our results indicate that the PACE4C and PACE4D genes reside on chromosome 15 like PACE-4A, suggesting that these isoform mRNAs are produced from the primary transcript by alternative splicing.

Isolation of Recombinant Bacteriophage Containing the Human PACE4 Gene—The gene structures of human furin [10 kb] (*35*), PC2 [>135 kb] (*36*), mouse PC1 [42 kb] (*37*), and PC4 [9.5 kb] (*38*) have been reported. The coding sequences of common domains (signal peptide, propeptide, catalytic, and homoB domains) are interrupted by introns at similar positions. The main differences among these genes are in intron structure and the 3′-terminal region encoding a specific domain. Preliminary Southern blot analysis of human genomic DNA using PACE4 cDNA probe indicated that the human PACE4 gene is relatively large by virtue of a long intron, in contrast to the compact structure of the furin gene. Moreover the structure of the PACE4 gene is expected to be very complicated if PACE4 isoform mRNAs are produced from a single gene by alternative splicing. Therefore we initially screened the genomic library (10^6 recombinant phages) using various PACE4 cDNA probes (5′-D, SP, SCD, HomoB, CRR, A-3′, C-3′, CD probes) separately to get overlapping clones. A total of 58 positive hybridization clones (one clone/5′-D probe, 3 clones/SP probe, 16 clones/SCD probe, 14 clones/HomoB probe, 15 clones/CD probe, 5 clones/CRR probe, 3 clones/A-3′ probe, and one clone/C-3′ probe) were isolated. The insert DNA of these clones was analyzed by restriction enzyme mapping and Southern blotting. Several clones (λhPA-271/5′-D probe, λhPA-7/SP probe, λhPA-44 and λhPA-50/SCD probe, λhPA-61 and λhPA-78/homoB probe, λhPA-6, λhPA-193 and λhPA-196/CRR probe, λhPA-19/C-3′ probe, λhPA-33 and λhPA-35/A-3′ probe, λhPA-164/CD probe) were selected for sequence analysis. These clones contained DNA sequences corresponding to a part of the propeptide, SCD, homoB domain, CRR and the isoform-specific region, as shown in Fig. 3. In the course of sequence analysis of λhPA-33 insert DNA, we found a PACE4E-specific sequence just behind the exon encoding nucleotides 2570 to 2699 of the CRR region.

However no clone containing the 5′-terminal sequence, including the 5′-untranslated region, signal peptide or a part of the propeptide, was obtained in the initial screening. Probably the highly GC-rich structure (89% GC content) of the sequence encoding the signal peptide impedes the reaction of *Bca*BEST DNA polymerase in the probe synthesis. Therefore we next attempted to isolate a clone containing the 5′-terminal sequence of the gene by using end-labeled oligonucleotide probes. Four positive clones (λhPA

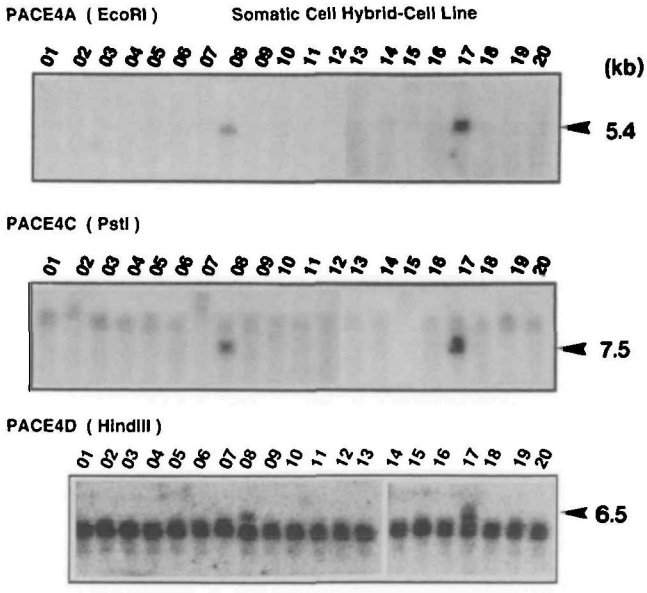

Fig. 2. **Mapping of the gene locus of PACE4A, 4C, and 4D.** DNAs from 20 individual hamster/human cell hybrids harboring distinctive complements of human chromosomes were analyzed by Southern blotting. The chromosomal complements of the hybrid lines are described under "MATERIALS and METHODS." Hybrid DNAs digested with *Eco*RI, *Pst*I, and *Hin*dIII were resolved in a 0.8% agarose gel, transferred to a nylon membrane and hybridized with radiolabeled A-3′ (PACE4A-specific probe), C-3′ (PACE4C-specific probe), and 5′-D (PACE4D-specific probe), respectively. The bands generated from human chromosome are indicated by arrows.

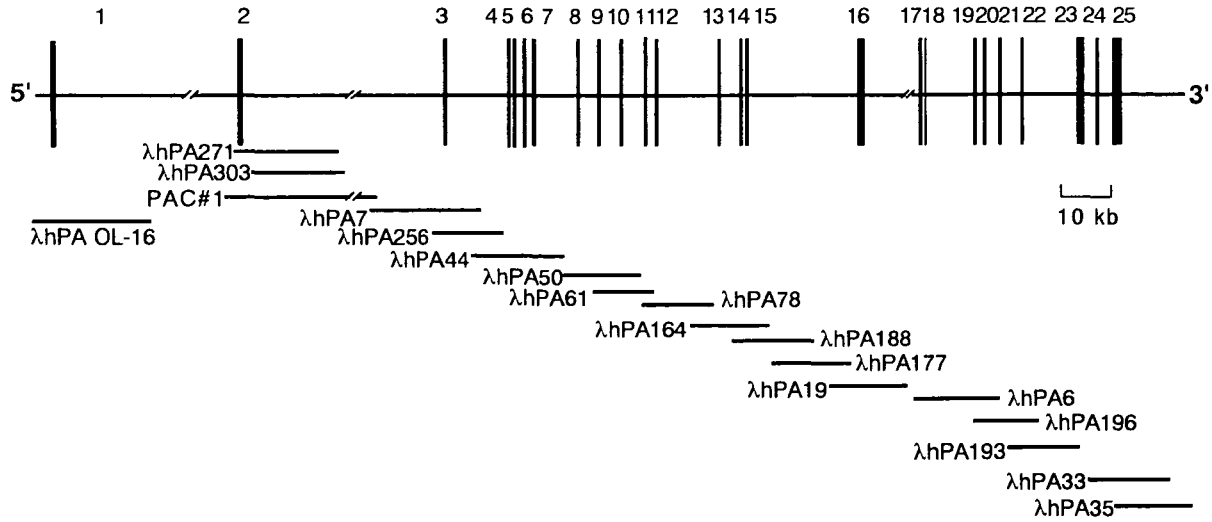

Fig. 3. **Organization of the human PACE4 gene.** The relative positions of nineteen genomic DNAs (18 from the EMBL genomic library and one from the PAC library) are shown in relation to the schematic structure of the PACE4 gene. Exons are shown by vertical bars and introns by horizontal lines. Exons are numbered from the 5′-end of the gene. The sizes of the intron between exons 1 and 2, exons 2 and 3, and exons 16 and 17 are unknown.

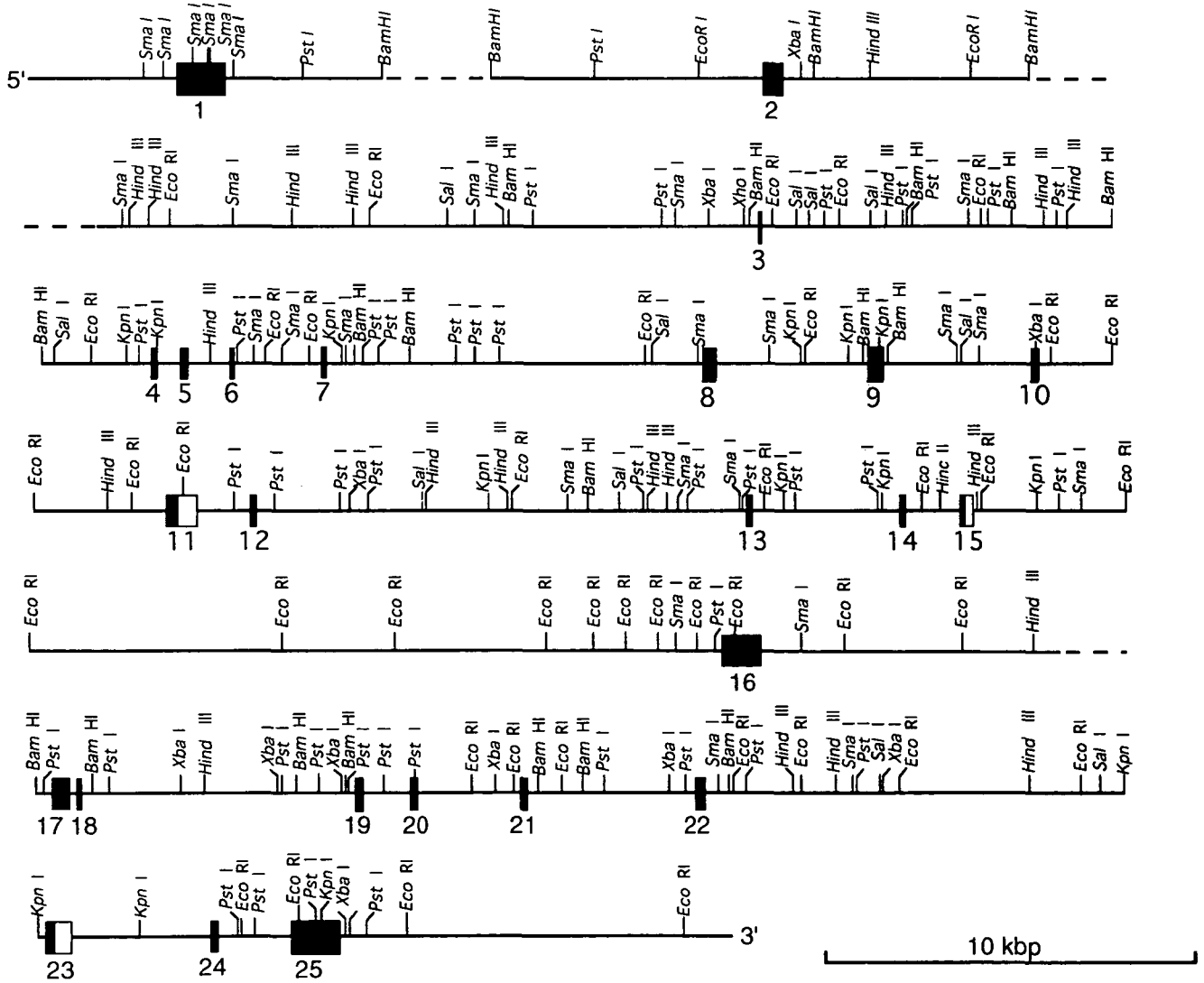

Fig. 4. **Restriction endonuclease map of the human PACE4 gene.** Exons are shown by boxes. Isoform-specific exons (PACE4B, 4D, and 4E) are shown by open boxes. Introns of unknown size are indicated by dashed lines.

TABLE I. **Exon-intron organization of the human PACE4 gene.** Exon sequences are in capital letters and intron sequences are in lower-case letters.

Exon Number	5'-Splice donor	Intron	3'-Splice acceptor	Exon Number
1 (611 bp)	Leu Gly Gln[99] TTG GGC CAG gtgagtgcg	(Intron A >20 kbp)	ttttttaaa TTTCAAACT	2
2 (418 bp)	TGAATAGAT gttcaattg	(Intron B >20 kbp)	[100]Ile Gly Asn gtcttgcag ATT GGA AAC	3
3 (105bp)	Asp Pro Gln[134] GAC CCC CAG gtacagagt	(Intron C 12.5 kbp)	[135]Val Lys Trp ttctggtag GTG AAA TGG	4
4 (111bp)	Trp Tyr Leu[171] TGG TAC CTG gtgagtagg	(Intron D 532bp)	[172]His Cys Gly tgcttacag CAT TGT GGC	5
5 (144bp)	Pro Asn Tyr[219] CCA AAT TAT gtaagtcaa	(Intron E 1261bp)	[220]Asp Ser Tyr atttcttag GAT TCC TAC	6
6 (77 bp)	Glu Asn Ly[245] GAA AAT AA gtacgtcac	(Intron F 2002bp)	[245]s His Gly ttgtttcag A CAC GGC	7
7 (89bp)	Ile Gly G[275] ATA GGA G gtaaggccg	(Intron G 7.5kbp)	[275]ly Ile Arg tgctttcag GC ATC CGC	8
8 (173bp)	Ile Lys Lys[332] ATT AAA AAG gtgtgagta	(Intron H 4.0kbp)	[333]Gly Arg Gln tccctccag GGC CGG CAG	9
9 (213bp)	Arg Lys Ile[403] CGA AAA ATC gtaagttct	(Intron I 3.8kbp)	[404]Val Thr Thr ctcccccag GTC ACC ACG	10
10 (101bp)	Glu Ala As[437] GAA GCA AA gtaagttcc	(Intron J 4.9kbp)	[437]n Ser Gln tcattgcag C AGC CAG	11
11common (104bp)	His Lys V[472] CAT AAA G gtgcggcag	(Intron K 2.0kbp)	[472]al Ser His tcttttcag TT AGC CAT	12
11common-11B(719bp)	TTGTGTGCC			
12 (118bp)	Arg Pro Ar[511] AGA CCC AG gtaaggctc	(Intron L 12kbp)	[511]g Ser Ile ctcccccag G AGC ATC	13
13 (189bp)	Ala Lys Ar[574] GCA AAG AG gtaaggcga	(Intron M 3.7kbp)	[574]g Leu Leu caattgcag G TTG CTG	14
14 (137bp)	Lys Gln G[620] AAG CAA G gtcagtggc	(Intron N 1151bp)	[620]ly Asp Leu acccattag GT GAT CTT	15
15CD(46bp)	Glu Glu Ar[635] GAA GAG AG gttcgtttc	(Intron O 22.0kbp)	[635]g Glu Pro gttttccag G GAA CCT	16
15CD-15D (168bp)	CGAGTCCTC			
16 (1049bp)	TAAACATTG	(Intron P >20kbp)	[620]ly Lys Leu gccttccag GG AAG TTG	17
17 (180bp)	Tyr Thr A[680] TAC ACA G gtaatgagc	(Intron Q 161bp)	[680]la Gln Ser ctcacacag CT CAA TCC	18
18 (39bp)	Gln Thr S[693] CAG ACC A gtaagtatg	(Intron R 2.0kbp)	[693]er Val Cys ttctttaag GT GTG TGC	19
19 (103bp)	Thr Ser Ar[727] ACC AGC AG gtaatgcat	(Intron S 1243bp)	[727]g Lys Cys ctcctacag G AAG TGC	20
20 (197bp)	Asp Glu Se[793] GAT GAA A gtaagtggc	(Intron T 2.3kbp)	[793]er Gln Lys tttctccag GT CAG AAA	21
21 (88bp)	Gly Phe Se[822] GGA TTC AG gtaaaaccc	(Intron U 4.3kbp)	[822]r Leu Ala tttcaacag C CTT GCA	22
22 (104bp)	Cys Val G[857] TGC GTG G gtgagttca	(Intron V 11kbp)	[857]ly Pro Gly gtctctcag GG CCA GGC	23AE
23AE (130bp)	Cys Arg Ar[900] TGT CGA AG gtacggtcc	(Intron W 4333bp)	[900]g Cys Asp ttcccacag G TGT GAC	24
24 (113bp)	Ser Asn A[938] AGC AAC G gtgagcagc	(Intron X 1822bp)	[938]la Asp Glu gggccccag CT GAC GAG	25
25 (1422bp)	TAACATCCC			

OL-3, 4, 9, and 16) which hybridized with three kinds of oligoprobes were isolated. A *Bam*HI-digested fragment (4.0 kb) liberated from λhPA OL-16 was subcloned and sequenced. This fragment contained the DNA sequence encoding the 5′-untranslated region, signal peptide, and a part of the propeptide.

There were gaps between λhPA-7 and λhPA-44, λhPA-44 and λhPA-50, λhPA-164 and λhPA-19, and λhPA-19 and λhPA-6, which we tried to fill with a second screening. By using *Sal*I-*Eco*RI fragment (1.6 kb) derived from the 5′-terminus of the insert DNA from λhPA-19 as screening probe, λhPA-177 and λhPA-188 clones were obtained and these clones overlapped with λhPA-164 and λhPA-19. The size of the intron between the exon encoding D-3′specific

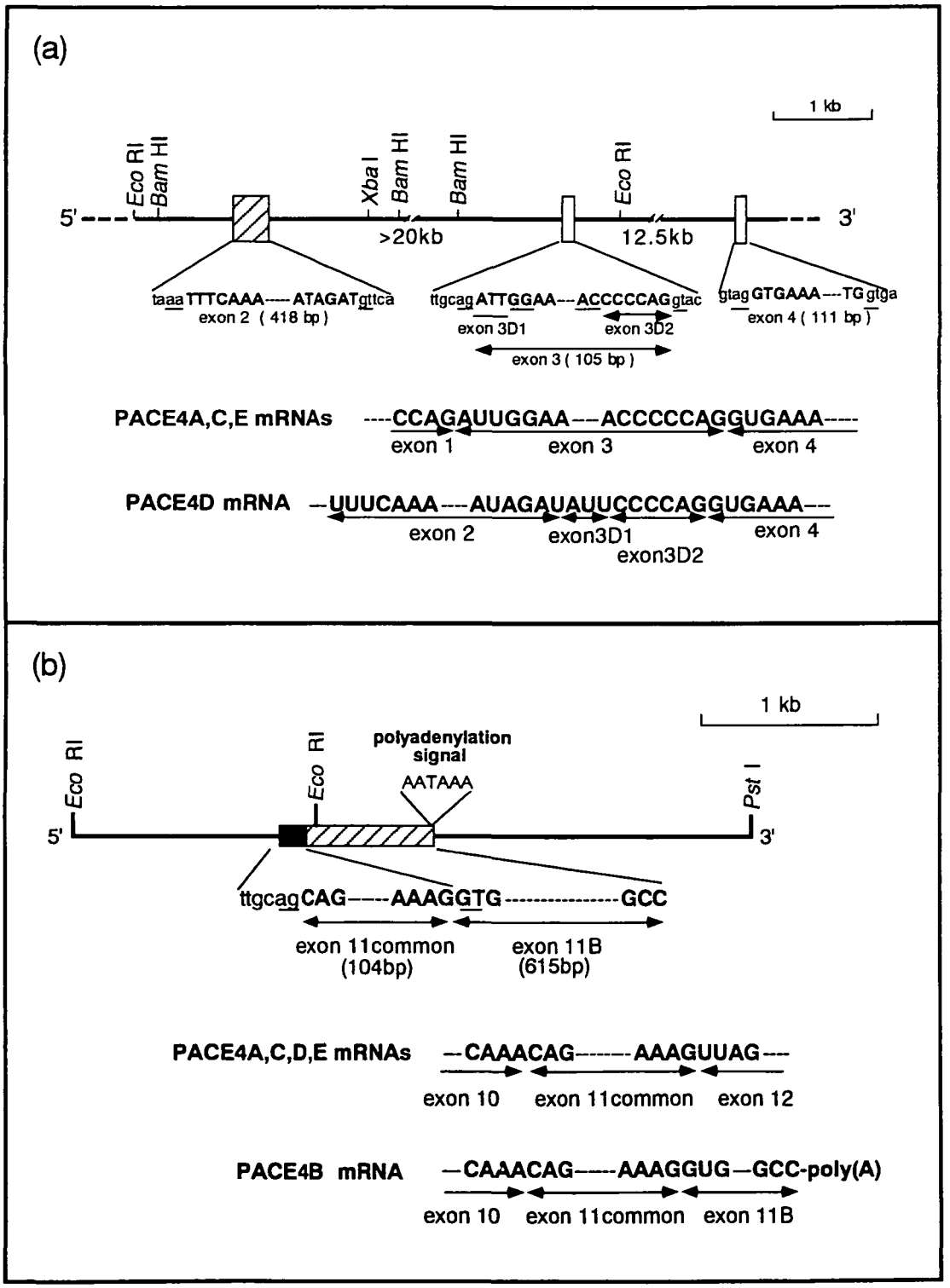

Fig. 5. Alternative splicing of exons 2 and 3 (a) and exon 11 (b).

sequence and the C-specific exon was estimated to be about 22 kb by restriction enzyme mapping and Southern blotting. λhPA-256 clone which overlapped with both λhPA-7 and λhPA-44 clones, was isolated by a second screening using *Pst*I fragment (0.5 kb) derived from the 5′-terminus of the DNA insert of λhPA-44 as a probe. To fill the gap between λhPA-19 and λhPA-6, a second screening using *Hin*dIII-*Sal*I fragment (2.1 kb) derived from the 3′-terminus of the λhPA-19 insert as a probe was performed. Three positive clones (λhPA-291, λhPA-292, and λhPA-293) were obtained, but none overlapped with λhPA-6.

Isolation of PAC Clone Containing the 5′-Terminal Sequence of PACE4D cDNA—To clarify the genomic organization at the 5′-terminal region of the PACE4 gene, we screened the PAC library by PCR, which amplified the sequence at the 5′-terminus of PACE4D cDNA. Two positive clones (hPAC 13355 and 13556) were obtained. The profiles of restriction enzyme digestion of the DNA of

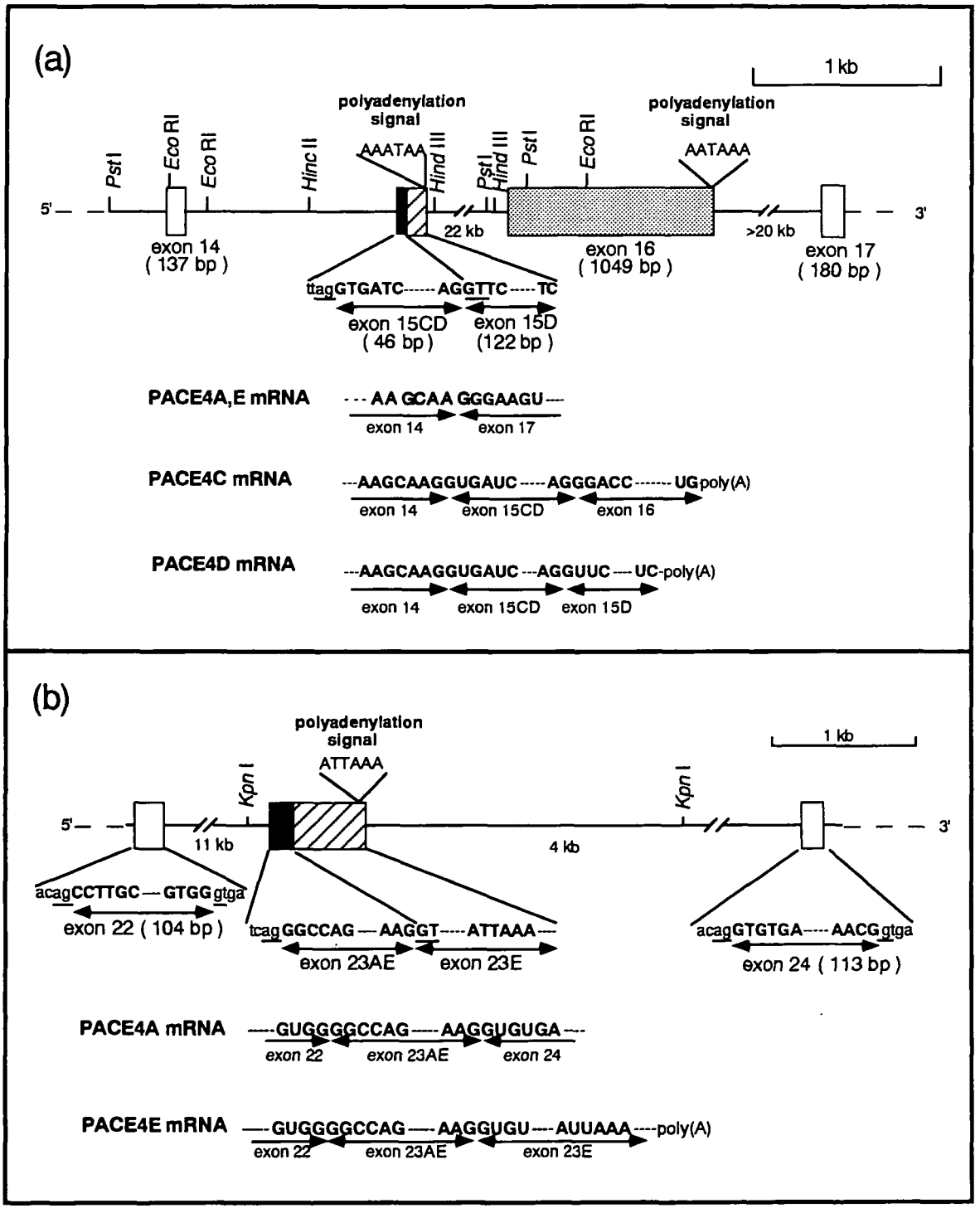

Fig. 6. **Alternative splicing of exons 15, 16, and 17 (a) and exons 23 and 24 (b).**

both plasmids were very similar, so we characterized the hPAC 13355 clone. Southern blot and sequence analyses showed that the insert DNA from this clone contained the 5′-terminal region of PACE4D cDNA and overlapped at the 3′-terminus with λhPA7. However it did not contain the sequence coding for PACE4 signal peptide.

Structural Organization of Human PACE4 Gene and Its Alternative Splicing—A map of the human PACE4 gene was constructed from 19 genomic DNA fragments (18 from the EMBL genomic library and one from the PAC library) as shown in Figs. 3 and 4. The gene consists of 25 exons and spans more than 250 kb. The four active site residues, Asp^{205}, His^{246}, Asn^{347}, and Ser^{420} are found on exons 5, 7, 9, and 10, respectively, like other kexin-family protease genes (*35-38*). The sequence of the exon-intron boundaries and the sizes of the exons and introns are summarized in Table I. The boundaries of the exon-intron junctions conform to the GT-AG rule except for the 5′-terminal portion of PACE4D (exon 2). Exon 1 encodes 97 amino acids which comprise the signal peptide and the first 36 amino acids of the propeptide region of PACE 4A, B, C, and E. Exon 2 encodes the 5′-D specific sequence (nucleotides −526 to −107 in PACE4D cDNA). Exons 3-10 consist of 105 bp (nucleotide 298-402 in PACE4A cDNA), 111 bp (nt 403-513), 144 bp (nt 514-657), 77 bp (nt 658-734), 89 bp (nt 735-823), 173 bp (nt 824-996), 213 bp (nt 997-1209), and 101 bp (nt 1210-1310), respectively. Thus, these exons encode the propeptide and catalytic domains. The hPAC clone 13355, which was isolated from the PAC library, contained exon 2 and overlapped with the 5′-terminus of λhPA-7. This clone does not contain exon 1. The average insert size of the PAC library is about 120 kb. Therefore these results suggested that exon 1 was located more than 100 kb upstream of exon 3. A comparison of the sequences of exons 2 and 3 with the 5′-terminal sequence of PACE4D mRNA showed that PACE4D mRNA is produced by unusual RNA splicing as shown in Fig. 5a. The sequence of exon 2 is specific for PACE4D mRNA and is excluded from other PACE4 isoform mRNAs. In the PACE4D mRNA, the two partial sequences at both termini of exon 3, 5′-ATT-3′ (exon 3D1) at the 5′-end and 5′-CCCCAG-3′ (exon 3D2) at the 3′-end, are included but the inner sequence (5′-GG-----AC-3′) is excluded. The sequence of the 3′-splice acceptor site of exon 2, 5′-aa-3′, is not consistent with the consensus sequence (5′-ag-3′). Moreover the sequences of the 5′-donor site of exon 3D1 and the 3′-acceptor site of exon 3D2 are also not consistent. Exons 12, 13, and 14 consist of 118 bp (nt 1415-1532), 189 bp (nt 1533-1721), and 137 bp (nt 1722-1858), respectively. These exons encode the homoB domain. Exon 16 encodes the specific 3′-terminal sequence of PACE4C mRNA (nt 1905-2953 in PACE4C cDNA). Exons 17-22, 24, and 25 consist of 180 bp (nt 1859-2038 in PACE4A cDNA), 39 bp (nt 2039-2077), 103 bp (nt 2078-2180), 197 bp (nt 2181-2377), 88 bp (nt 2378-2465) and 104 bp (nt 2466-2569), 113 bp (nt 2700-2812) and 1,422 bp (nt 2813-4234),

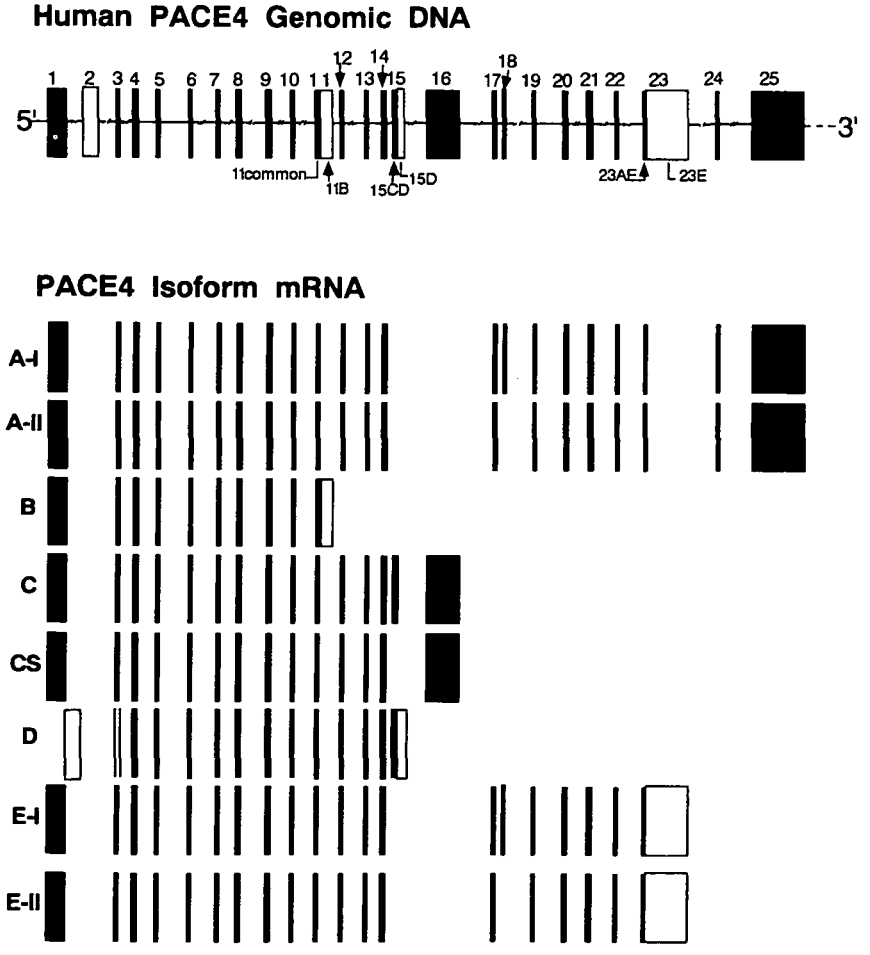

Fig. 7. **Schematic representation of exon composition of the alternatively spliced PACE4 messenger RNAs.** Exons are shown by boxes. Isoform-specific exons of PACE4B, C, D, and E are shown by open boxes. Exons 2 and 15D are specific for PACE4D. Exons 11B and 16 are specific for PACE4B and PACE4C, respectively. Exon 18 is contained in PACE4A-I and PACE4E-I. Exon 23E is specific for PACE4E.

respectively, and are utilized as single exons. However, the following three exons, 11, 15, and 23, contain internal donor sites and can be divided in two. The 5′-portion of exon 11, termed exon 11common, is composed of 104 bp and encodes the common sequence (nucleotides 1311-1414) for all PACE4 isoform mRNAs, while the 3′-portion of exon 11, termed exon 11B, consists of 615 bp and encodes the PACE4B mRNA-specific sequence as shown in Fig. 5b. The 5′-portion of exon 15, termed exon 15CD, encodes the common sequence for PACE4C and PACE4D mRNAs (nt 1859-1904 in PACE4C mRNA), while the 3′ portion of exon 15 (exon15D, 122 bp) encodes the PACE4D mRNA-specific sequence, as shown in Fig. 6a. Exon 23 consists of exon 23AE (5′-portion) and exon 23E (3′-portion) as shown in Fig. 6b. Exon 23AE encodes the common CRR sequence (nt 2570-2699) for PACE4A and PACE4E mRNAs. Exon 23E encodes the PACE4E mRNA-specific sequence. The polyadenylation signal (ATTAAA) is located 860 bp downstream from the 5′-end of exon 23E. As shown in Figs. 5b, 6a, and 6b, the common sequence 5′-AGGT-3′ is located at all the junction boundaries of exon 11common/11B, exon 15CD/D and exon 23AE/E. This sequence contains the consensus sequence (5′-GT-3′) for the donor site of RNA splicing. In the case of exon 23, PACE4A mRNA can be produced by splicing out the exon 23E sequence, and PACE4E mRNA can be generated by including the continuous exon 23E sequence, but not the separated exon 24 sequence; similar alternative RNA processing events can occur in exons 11 and 15. A schematic of the alternative splicings of PACE4 mRNA, based upon the exon-intron arrangements of human PACE4 gene together with the predicted primary structures from cDNAs of the isoforms, is shown in Fig. 7. Exon 1 encodes the 5′-untranslated region and the protein coding region for the signal peptide and part of the amino-terminal sequence of the propeptide. Exon 2 contains the D-specific sequence located at the 5′-terminus of PACE4D mRNA. This exon is spliced out in other PACE4 isoform mRNAs. Exon 3 encodes part of the carboxy-terminal sequence of the propeptide region. In the case of PACE4D mRNA, unusual splicing of exon 3 occurred, as mentioned above. The 7 exons (exons 4-10) located at the 5′-terminal region of the gene encode the subtilisin-like catalytic domain common for all PACE4 isoforms. Exon 11 consists of the common sequence (exon11common) for all PACE4 isoform cDNAs followed immediately by a unique PACE4B mRNA (exon11B) sequence containing the polyadenylation signal. The 3 exons which follow (exons 12-14) encode the homoB domain. Exon 15 consists of the sequence common to PACE4C and PACE4D (exon 15D) and a unique PACE4D mRNA sequence (exon 15D) containing polyadenylation signal. Exon 16 is located about 22 kb downstream from exon 15D and encodes the 3′-terminal unique sequence of PACE4C mRNA including the polyadenylation signal. Recently Zhong *et al.* (*32*) reported a novel isoform of PACE4, PACE4CS, a carboxy-terminally shortened version of PACE4C. Comparison of the 3′-terminal sequences of PACE4CS with genomic DNA indicates that exon 15CD is deleted in PACE4CS mRNA. The next 7 exons (exon 17-23AE) encode a sequence common to PACE4A and PACE4E mRNAs. Type I PACE4A and

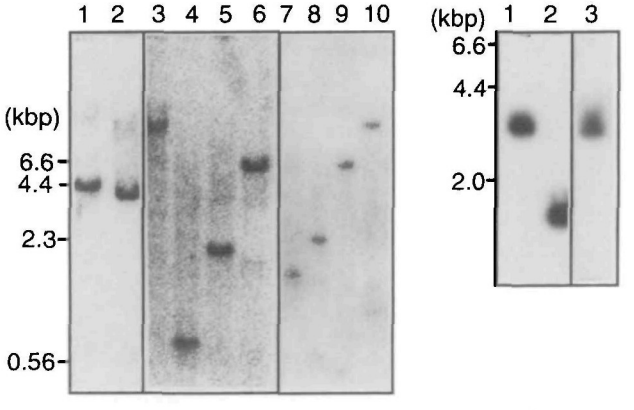

Fig. 8. **Southern blot analysis of the human PACE4 gene.** Human genomic DNA was digested with *Bam*HI [(1) lanes 1, 3, 7], *Hin*dIII [(1) lanes 2, 5, 9; (2) lane 3], *Pst*I [(1) lanes 4,8; (2) lane 2], and *Eco*RI [(1) lanes 6, 10; (2) lane 1]. The digests were separated on 0.8% agarose gel, blotted onto a nylon membrane and hybridized with radiolabeled probes. Marker DNAs (*Hin*dIII-digested λ DNA) were electrophoresed in parallel and the sizes of the DNA fragments are shown. (1) The filters were hybridized with a radiolabeled genomic fragment, *Sma*I-fragment (1.7 kb) from λhPA50 (left), *Eco*RI-*Pst*I fragment (0.6 kb) from λhPA76 (middle), and *Pst*I fragment (1.6 kb) from λhPA6 (right). (2) The filters were hybridized with a cDNA probe, A-3′ (left) or C-3′ (right).

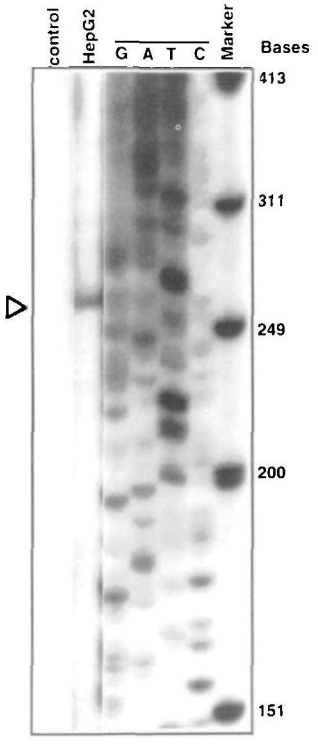

Fig. 9. **Determination of the human PACE4 transcription initiation site by primer extension analysis.** The radiolabeled 24-mer oligonucleotide was anealed to 25 μg of total RNA from Hep G2 cells and extended with reverse transcriptase as described under "MATERIALS and METHODS." The product was analyzed by 8% polyacrylamide gel electrophoresis in parallel with a sequencing reaction using the identical oligonucleotide primer and the genomic fragment cloned into the *Bam*HI site of PUC18, which is known to contain the first exon and 5′-untranslated region. Lane 1, control (no RNA); lane 2, the primer extension product. Molecular weight markers (end-labeled *Hin*fI-digested φX174 DNA) are shown on the right.

PACE4E mRNAs contain exon 18 while type II mRNAs do not. Because deletion of exon 18 does not result in a frame-shift, the carboxy-terminal amino acid sequences of Type I and II isoforms are the same (*19, 41*). Exon 23 consists of a sequence common to PACE4A and PACE4E mRNAs, and a unique 3′-terminal sequence of PACE4E mRNA containing a polyadenylation signal. We could not obtain PACE4E cDNA containing a poly (A) tail, so the location of the 3′-end of exon 23E was not identified. Exons 24 and 25 encode the 3′-terminal region of PACE4A mRNA.

Southern Blot Analysis of Human PACE4 Gene—Southern blot analysis of genomic DNA was performed using PACE4 gene fragments (1.7 kb *Sma*I fragment, 0.6 kb *Eco*RI-*Pst*I fragment and 1.6 kb *Pst*I fragment) and cDNA fragments (A-3′ and C-3′) as hybridization probes. The *Sma*I fragment (1.7 kb) derived from the λhPA-50 clone, and the *Eco*RI-*Pst*I fragment (0.6 kb) from the λhPA-76 clone contains the whole sequence of exons 8 and 13, respectively. The *Pst*I fragment (1.6 kb) from the λhPA-6 clone contains the whole sequences of exons 17 and 18. Exon 8 encodes a part of the catalytic domain. All PACE4 isoform cDNAs contain this sequence. PACE4A, C, D and E cDNAs contain the exon 13 sequence which encodes a part of the homoB region. The exon 17 and 18 sequences are included in both PACE4A and PACE4E (Type-I) cDNAs. As shown in Fig. 8, a single positive band was observed in the genomic DNA digests obtained with different restriction enzymes that do not cut the probe used. Thus the results of gene mapping and Southern blot analysis indicate that the multiple forms of PACE4 mRNA can be explained by alternative splicing of the primary transcript of a single gene.

Location of the Transcription Initiation Site and Nucleotide Sequence of the 5′-Flanking Region—To identify the transcription initiation sites, primer extension analysis was performed. For primer extension analysis, an oligonucleotide primer complementary to the 5′-untranslated sequence (nucleotides −74 to −51) was used to analyze total RNA isolated from Hep G2 cells known to express high levels of PACE4A mRNA. Primer extension resulted in a single band, indicating that a single site is used for the start of transcription, as shown in Fig. 9. This site is located 314 bp upstream from the translation initiation codon (ATG). A *Bam*HI fragment (4 kbp) from λhPA OL-16, known to contain the first exon encoding the signal peptide and a portion of the 5′-flanking region (see Fig. 4), was sequenced on both strands to characterize further the potential regu-

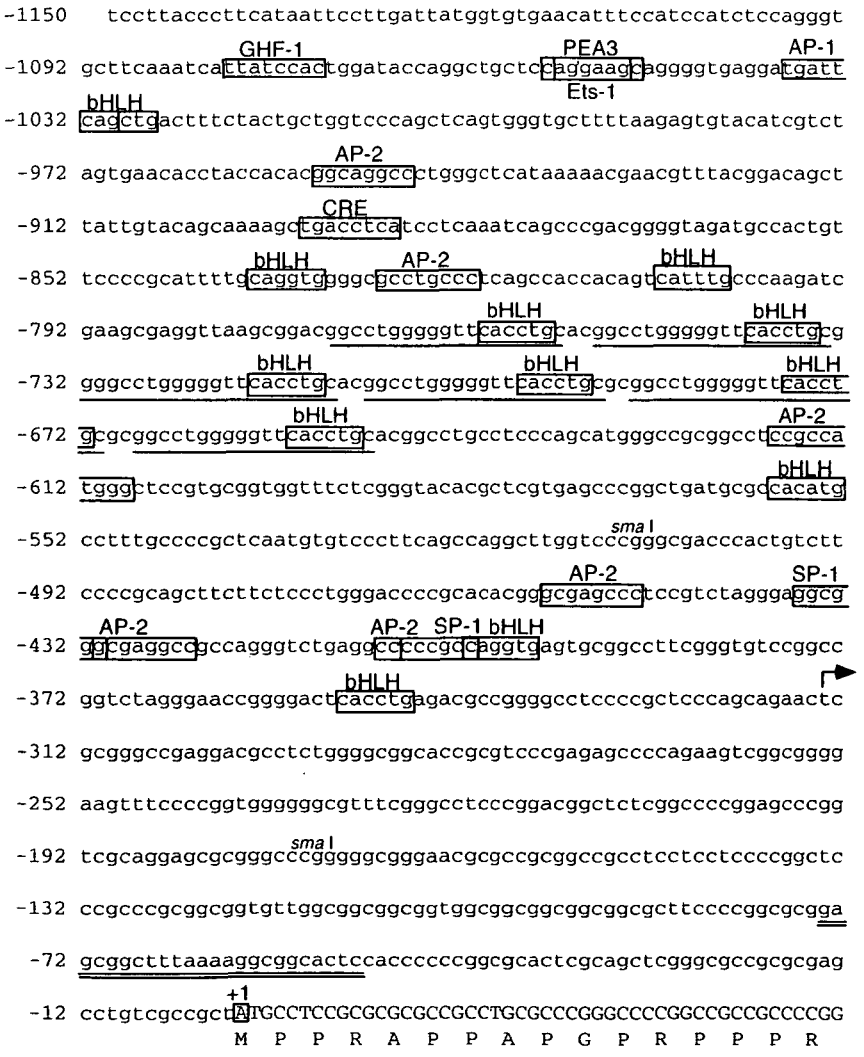

Fig. 10. **Nucleotide sequence of the 5′-flanking region of the PACE4 gene.** The numbering of nucleotides starts at the translation initiation site (+1). The transcription initiation site is indicated by a bent arrow. The putative regulatory elements are boxed. The repeated sequences (GGCCTGGGGGTTCACCTGC) containing the E box are underlined. The location of the primer used for primer extension analysis in Fig. 9 is double underlined. This sequence has been deposited in the DDBJ/EMBL/GenBank DNA databases and assigned the accession number AB001898.

latory sequences involved in PACE4 expression. The sequence 1.2 kb upstream of the ATG translation initiation codon is shown in Fig. 10. Furthermore, we directly amplified the 5′-upstream region by genomic PCR and confirmed the sequence of this region. Neither a TATA box nor CCAAT sequences are evident adjacent to the transcription initiation site. The most remarkable feature of the sequence of this region is the presence of 12 binding sites (E box, CANNTG) for basic helix-loop-helix proteins (*39*) at nucleotide positions −1032, −839, −807, −763, −741, −719, −698, −677, −656, −558, −402, and −352. In particular, GGCCTGGGGGTTCACCTGC sequences containing 6 repeats of the E box are located at nucleotide positions −775 to −650. Other possible binding sites for transcription factors such as SP1 (*40*), AP1 (*41*), AP2 (*42*), PEA3 (*43*), and Ets-1 (*44*) are also found in this region. Two SP1-like binding sequences are located at nucleotide positions −436 and −407/(inverted). SP1 is known to guide initiation in some TATA-less promoters (*45*). Potential AP1, PEA3, and Ets-1 binding sites are located at positions −1037, −1055, and −1056, respectively. Five other potential sites at positions −954, −829, −618, −456, and −431 share homology with the AP2 binding site, CCCC(A/G)(G/C)(G/C)C. Potential sites for cAMP response element binding protein (CREB) (*46*) and GHF-1 (*47*) were identified at −897 and −1083, respectively. Recently, Johnson *et al.* reported thyroid hormone-induced expression of rat pituitary PACE4 mRNA (*10*). We did not find the consensus thyroid hormone response element (*48*) in the 1.2 kb 5′- flanking region, although such an element may be present further upstream.

DISCUSSION

In this study we have characterized the human PACE4 gene. The exon-intron organization of this and other kexin-like protease genes is strikingly conserved within the amino-terminal region containing the signal peptide, propeptide and subtilisin-like catalytic domains. Its estimated size of 250 kb, makes the human PACE4 gene longer than the human furin (10 kb) (*35*), mouse PC1 (42 kb) (*37*), human PC2 (130 kb) (*36*), and mouse PC4 (9.5 kb) genes (*38*). The human PACE4 and furin genes are located in close proximity on chromosome 15, suggesting a local duplication during the evolution of the kexin-like protease family (*9*). Similarly both genes are localized on the chromosome 7 in mice (*49*). However, the overall gene structures of the two proteases differ. The furin gene consists of 15 exons and spans 10 kb (*35*), while the PACE4 gene is composed of 25 exons and spans more than 250 kb, being 25 times longer than the furin gene.

Alternative RNA splicing is a fundamental process in eukaryotes that contributes to tissue-specific and developmentally regulated patterns of gene expression. In this study, we elucidated the mechanism of production of multiple PACE4 transcripts by alternative splicing from a single copy of the gene. PACE4A consists of a signal peptide, propeptide, catalytic domain, HomoB domain, and cysteine-rich domain, and is expressed in various tissues and culture cells. However, its distribution is restricted in both endocrine and nonendocrine cells, in contrast to that of furin (*23*). PACE4B mRNA, which lacks a homoB domain and cysteine-rich region, is expressed only in human embryonic kidney cell line 293 (*9*). PACE4D mRNA, which lacks a signal peptide, propeptide and a cysteine-rich region, is expressed in various tissues, like PACE4A mRNA. However, PACE4B and PACE4D are supposed to be catalytically inactive, because propeptide and the homoB domain are essential for the correct folding of the enzyme. PACE4C, which is truncated at the carboxy-terminus compared to PACE4A, has a distinct sequence at the end of homoB domain, resulting in a shorter form in which the 32 amino acids following Gly^{620} are different. Exon 17 of human PACE4 gene, which is not contained in PACE4C mRNA (*11*), encodes an amino acid sequence, L-X-L-Y/H-G-T/S, which is highly conserved among the kexin-like protease family. Recent studies on the biosynthesis of PACE4C in transiently transfected COS-1 cells suggested that the PACE4C protein remains in endoplasmic reticulum in an inactive zymogen form, like PACE4CS (*31, 32*). PACE4C was not secreted into the culture medium at all. In contrast, a part of PACE4A was secreted as an active enzyme in transiently transfected cells (*31*). In general, abnormal (misfolded) or mislocalized proteins are degraded rapidly. However, accelerated proteolytic degradation of PACE4C protein was not observed in COS-1 cells. Moreover we observed specific expression of PACE4C in β cells of pancreatic islets as reported previously (*25*). These results suggested that PACE4C is not a product of abnormal splicing, but has unknown physiological functions. PACE4E is the most recent isoform to be identified by cDNA cloning (*31*). This novel cDNA encodes a 975 amino acid sequence and has a characteristic hydrophobic cluster at the carboxy-terminus. PACE4E was shown to be a functional enzyme by coexpression of PACE4E cDNA construct with precursor protein (pro-von Willebrand factor and pro-complement C3) cDNAs, like PACE4A. PACE4E could process these precursor proteins correctly to mature forms. However little PACE4E is secreted into culture medium unlike PACE4A, though its secretion was accelerated by removal of the hydrophobic region of the carboxy-terminus. These results indicated that the hydrophobic region causes retardation of PACE4E secretion, suggesting distinct intracellular localization of PACE4E compared with that of PACE4A. All introns of the human PACE4 gene contain polypyrimidine tract-like structures upstream of the 3′-splice site. The polypyrimidine tract is essential for efficient branch point utilization and 3′-splice site recognition. It probably also plays a significant role in alternative splicing. Many features of the pre-mRNA have been implicated in alternative splice site selection (*50*). These include the relative strength of the 5′- and 3′-splice sites, intron size, the location of branch points, multiple alternative branch points, branch point sequences, intron sequences between the 3′-splice site and upstream branch point, and exon sequences. Further analysis is necessary to clarify the mechanism of alternative splicing of PACE4 pre-mRNA.

Multiple transcripts have been detected for other kexin-like protease genes including furin (*51*), PC4 (*38*), and PC6 (*34, 52*). In the case of PC6, two transcripts (PC6A and PC6B) are produced and both are expressed in a tissue-specific manner. In the case of furin, three different transcripts, all with identical coding sequences but different 5′-untranslated sequences, are produced *via* alternative

promoter usage (*51*). However the functional role of the isoforms is still unclear.

The promoter region of human PACE4 gene was studied by sequence analysis. As in human PC2 and mouse PC4 genes (*36, 38*), there was no TATA box or CCAAT sequence adjacent to the transcription initiation site. It has been observed that the promoter region of some TATA-less house-keeping genes contains a GC-rich sequence. Similarly, human PACE4 gene has a GC-rich upstream region and two SP1 binding sites. Several other binding sites for transcription factors were identified. Among them, the E box (CANNTG) is especially interesting. E boxes are also present in the 5′-upstream region of two other kexin-like protease genes, PC1 (*37*) and PC4 (*38*). However, the characteristic 6 tandem repeats of the E box in the 5′-upstream region of the PACE4 gene is unique. The E box is a binding motif for various basic helix-loop-helix (bHLH) factors such as MASH (*53*), MASH-1 (*54*), MATH-2 (*55*), NeuroD (*56*), and MyoD family (*57*). These factors interact with the E box element in the promoters in collaboration with other bHLH proteins. bHLH factors play an essential role not only in myogenesis, but also in neurogenesis. MASH, MASH-1, MATH-2, and NeuroD are considered to be trans-acting factors involved in neurogenesis. They can interact with several E box sequences in collaboration with other bHLH such as E47 and are involved in the development and maintenance of the mammalian nervous system. In adult rat, PACE4 is expressed in the central nervous system, especially in the cerebellum, pituitary and spinal cord (*10, 13, 23, 30*). In contrast, Myo D family members are myogenic regulators which can directly activate skeletal muscle-specific genes and convert a wide range of cell types into skeletal muscle (*57*). However, PACE4 is hardly expressed in skeletal muscle. These results suggested that the MyoD family is not involved in the regulation of PACE4 expression. The AP-1 binding site is located at nucleotide −1039. The trans-acting factor AP-1, a heterodimer composed of *c-jun* and *c-fos* proto-oncogene products, influences basal transcription and is required for induction of transcription by phorbol esters (*43*). In addition to the AP-1 site, other phorbol ester response elements have been identified. Both PEA3 and Ets-1 motifs are present as overlapped sequences at positions −1055 and −1056, respectively, in close proximity to the AP-1 site. Both factors can act synergistically with AP-1 to achieve maximal induction of transcription by phorbol esters (*41, 44*). The regulatory region of this gene also contains six potential AP-2 sites. Human PC2 gene also contains two AP-2 binding sites (*36*), though mouse PC1 and PC4 genes do not have AP-2 binding sites (*37, 38*). AP-2 affects the action of both phorbol esters and cAMP (*46*). Moreover, a potential site, differing by 1 bp from the consensus sequence, for the classical cAMP-response element binding protein (CREB), TGACGT(A/C)A was identified at position −895. In contrast, a GHF-1 (growth hormone factor 1)-like element differing by 1 bp from the consensus sequence (A/T)TAT(C/T)CAT, was found at position −1081. GHF-1 is a pituitary-specific transcription factor that plays a critical role in cell-type specific expression of the growth hormone gene (*47*). In fact, rat PACE4 is expressed most strongly in the anterior pituitary, especially in growth hormone-producing cells (*10*). These findings suggest that PACE4 is not a constitutive gene product and this is consistent with its tissue and cellular distribution. Although further analysis is necessary to elucidate whether any of these binding sites are functionally important for the promoter activity, a comparison of the sequences of the 5′-flanking region suggests that PACE4 gene expression is regulated by a mechanism distinct from that of furin, PC1, PC2, or PC4. Recently, Mains *et al.* reported that the enzymatic activity of PACE4 differs from that of other kexin-family proteases in its sensitivity to various inhibitors (*58*). This result also supports unique physiological roles for PACE4.

In this report, we show that the PACE4 gene has the most complicated structure of all known kexin-like protease genes, and we have clarified the mechanism of production of its isoforms to be alternative splicing. The results support regulation of its transcription by various transcription factors. Additional studies on the function and regulation of PACE4 isoforms are under way to elucidate the physiological role of these proteins in the central nervous systems.

We are grateful to Yoshiko Hare, Taisuke Sato, Masaki Sakashita, Kazuya Murasato, and Yuki Kiyosue for excellent technical assistance. We also thank Yuka Sasaki, Tomomi Segawa, and Tokue Kuroda for secretarial assistance.

REFERENCES

1. Halban, P.A. and Irminger, J.-C. (1994) Sorting and processing of secretory proteins. *Biochem. J.* **299**, 1-18
2. Barr, P.J. (1991) Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell* **66**, 1-3
3. Sakaguchi, T., Matsuda, Y., Kiyokage, R., Kawahara, N., Kiyotani, K., Katunuma, N., Nagai, Y., and Yoshida, T. (1991) Identification of endoprotease activity in the *trans* Golgi membranes of rat liver cells that specifically processes *in vitro* the fusion glycoprotein precursor of virulent Newcastle disease virus. *Virology* **184**, 504-512
4. Walker, J.H., Sakaguchi, T., Matsuda, Y., Yoshida, T., and Kawaoka, Y. (1992) Location and character of the cellular enzyme that cleaves the hemagglutinin of a virulent avian influenza virus. *Virology* **190**, 278-287
5. Fuller, R.S., Brake, A., and Thorner, J. (1989) Yeast prohormone processing enzyme (KEX2 gene product) is a Ca^{2+} dependent serine protease. *Proc. Natl. Acad. Sci. USA* **86**, 1434-1438
6. Fuller, R.S., Brake, A.J., and Thorner, J. (1989) Intracellular targeting and structural conservation of a prohormone-processing endoprotease. *Science* **246**, 482-486
7. Smeekens, S.P., Avruch, A.S., LaMendola, J., Chan, S.J., and Steiner, D.F. (1991) Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. *Proc. Natl. Acad. Sci. USA* **88**, 340-344
8. Smeekens, S.P. and Steiner, D.F. (1990) Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease *Kex2*. *J. Biol. Chem.* **265**, 2997-3000
9. Kiefer, M.C., Tucker, J.E., Joh, R., Landsberg, K.E., Saltman, D., and Barr, P.J. (1991) Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15. *DNA Cell Biol.* **10**, 757-769
10. Johnson, R.C., Darlington, D.N., Hand, T.A., Bloomquist, B.T., and Mains, R.E. (1994) PACE4: A subtilisin-like endoprotease prevalent in the anterior pituitary and regulated by thyroid status. *Endocrinology* **135**, 1178-1185
11. Tsuji, A., Higashine, K., Hine, C., Mori, K., Tamai, Y., Nagamune, H., and Matsuda, Y. (1994) Identification of novel cDNAs encoding human kexin-like protease, PACE4 isoforms. *Biochem. Biophys. Res. Commun.* **200**, 943-950
12. Hosaka, M., Murakami, K., and Nakayama, K. (1994) PACE4A is a ubiquitous endoprotease that has similar but not identical

substrate specificity to other Kex2-like processing endoproteases. *Biomed. Res.* **15**, 383-390
13. Tsuji, A., Hine, C., Mori, K., Tamai, Y., Higashine, K., Nagamune, H., and Matsuda, Y. (1994) A novel member, PC7, of the mammalian kexin-like protease family: Homology to PACE4A, its brain specific expression and identification of isoforms. *Biochem. Biophys. Res. Commun.* **202**, 1452-1459
14. Nakayama, K., Kim, W.-S., Torii, S., Hosaka, M., Nakagawa, T., Ikemizu, J., Baba, T., and Murakami, K. (1992) Identification of the fourth member of the mammalian endoprotease family homologous to the yeast Kex2 protease: Its testis-specific expression. *J. Biol. Chem.* **267**, 5897-5900
15. Nakagawa, T., Hosaka, M., Torii, S., Watanabe, T., Murakami, K., and Nakayama, K. (1993) Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: Its striking structural similarity to PACE4. *J. Biochem.* **113**, 132-135
16. Lusson, J., Vieau, D., Hamelin, J., Day, R., Chrétien, M., and Seidah, N.G. (1993) cDNA structure of the mouse and rat subtilisin/kexin-like PC5: A candidate proprotein convertase expressed in endocrine and nonendocrine cells. *Proc. Natl. Acad. Sci. USA* **90**, 6691-6695
17. Bruzzaniti, A., Goodge, K., Jay, P., Taviaux, S.A., Lam, M.H.C., Berta, P., Martin, T.J., Moseley, J.M., and Gillespie, M.T. (1996) C8, a new member of the convertase family. *Biochem. J.* **314**, 727-731
18. Meerabux, J., Yaspo, M.L., Roebroek, A.J., Van de Ven, W.J., Lister, T.A., and Young, B.D. (1996) A new member of the proprotein convertase gene family (LPC) is located at a chromosome translocation breakpoint in lymphomas. *Cancer Res.* **56**, 448-451
19. Seidah, N.G., Hamelin, J., Mamarbachi, M., Dong, W., Tadros, H., Mbikay, M., Chrétien, M., and Day, R. (1996) cDNA structure, tissue distribution, and chromosomal localization of rat PC7, a novel mammalian proprotein convertase closest to yeast kexin-like proteinases. *Proc. Natl. Acad. Sci. USA* **93**, 3388-3393
20. Constam, D.B., Calfon, M., and Robertson, E.J. (1996) SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis. *J. Cell Biol.* **134**, 181-191
21. Hatsuzawa, K. Murakami, K., and Nakayama, K. (1992) Molecular and enzymatic properties of furin, a Kex2-like endoprotease involved in precursor cleavage at Arg-X-Lys/Arg-Arg sites. *J. Biochem.* **111**, 296-301
22. Zhou, A., Bloomquist, B.T., and Mains, R.E. (1993) The prohormone convertases PC1 and PC2 mediate distinct endoproteolytic cleavages in a strict temporal order during proopiomelanocortin biosynthetic processing. *J. Biol. Chem.* **268**, 1763-1769
23. Seidah, N.G., Chrétien, M., and Day, R. (1994) The family of subtilisin/kexin like pro-protein and pro-hormone convertases: Divergent or shared functions. *Biochimie* **76**, 197-209
24. Marcinkiewicz, M., Ramla, D., Seidah, N.G., and Chrétien, M. (1994) Developmental expression of the prohormone convertase PC1 and PC2 in mouse pancreatic islets. *Endocrinology* **135**, 1651-1660
25. Nagamune, H., Muramatsu, K., Akamatsu,T., Tamai, Y., Izumi, K., Tsuji, A., and Matsuda, Y. (1995) Distribution of the kexin family proteases in pancreatic islets: PACE4C is specifically expressed in B cells of pancreatic islets. *Endocrinology* **136**, 357-360
26. Nagamune, H., Muramatsu, K., Higashine, M., Akamatsu, T., Tamai, Y., Yonetomi, Y., Tsuji, A., Izumi, K., Sakaguchi, T., Yoshida, T., and Matsuda, Y. (1997) Cyclic AMP-inducible procalcitonin processing in thyroidal parafollicular cells is regulated by the kexin-family protease, PC1 in *Medical Aspects of Proteases and Protease Inhibitors* (Katunuma, N., Kido, H., and Travis, J., eds.) IOS Press, Amsterdam, in press
27. Torii, S., Yamagishi, T., Murakami, K., and Nakayama, K. (1993) Location of Kex2-like processing endoproteases, furin and PC4, within mouse testis by in situ hybridization. *FEBS Lett.* **316**, 12-16
28. Dong, W., Marcinkiewicz, M., Vieau, D., Chrétien, M., Seidah, N.G., and Day, R. (1995) Distinct mRNA expression of the highly homologous convertases PC5 and PACE4 in the rat brain and pituitary. *J. Neurosci.* **15**, 1778-1796
29. Beaubien, G., Schafer, M.K.H., Weihe, E., Dong, W., Chrétien, M., Seidah, N.G., and Day, R. (1995) The distinct gene expression of the pro-hormone convertases in the rat heart suggests potential substrates. *Cell Tissue Res.* **279**, 539-549
30. Akamatsu, T., Nagamune, H., Yoshida, S., Mori, K., Tsuji, A., Matsuda, Y., and Daikoku, K. (1997) Developmental expression of a novel Kexin-family protease, PACE4E, in the rat olfactory system. *Histochem. Cell. Biol.* in press
31. Mori, K., Kii, S., Tsuji, A., Nagahama, M., Imamaki, A., Hayashi, K., Akamatsu, T., Nagamune, H., and Matsuda, Y. (1997) A novel human PACE4 isoform, PACE4E is an active processing protease containing a hydrophobic cluster at the carboxy terminus. *J. Biochem.* **121**, 941-948
32. Zhong, M., Benjannet, S., Lazure, C., Munzer, S., and Seidah, N.G. (1996) Functional analysis of human PACE4-A and PACE4-C isoforms: Identification of a new PACE4-CS isoform. *FEBS Lett.* **396**, 31-36
33. Zhou, A., Paquet, L., and Mains, R.E. (1995) Structural elements that direct specific processing of different mammalian subtilisin-like prohormone convertases. *J. Biol. Chem.* **270**, 21509-21516
34. De Bie, I., Marcinkiewicz, M., Malide, D., Lazure, C., Nakayama, K., Bendayan, M., and Seidah, N.G. (1995) The isoforms of proprotein convertase PC5 are stored to different subcellular compartments. *J. Cell. Biol.* **135**, 1261-1275
35. Barr, P.J., Mason, O.B., Landsberg, K.E., Wang, P.A., Keifer, M.C., and Brake, A.J. (1991) cDNA and gene structure for a human subtilisin-like protease with cleavage specificity for paired basic amino acid residues. *DNA Cell Biol.* **10**, 319-328
36. Ohagi, S., LaMendola, J., LeBeau, M.M., Espinosa III, R., Takeda, J., Smeekens, S.P., Chan, S.J., and Steiner, D.F. (1992) Identification and analysis of the gene encoding human PC2, a prohormone convertase expressed in neuroendocrine tissues. *Proc. Natl. Acad. Sci. USA* **89**, 4947-4981
37. Ftouhi, N., Day, R., Mbikay, M., Chretien, M., and Seidah, N.G. (1994) Gene organization of the mouse pro-hormone and proprotein convertase PC1. *DNA Cell Biol.* **13**, 395-407
38. Mbikay, M., Raffin-Saison, M.-L., Tadros, H., Sirois, F., Seidah, N.G., and Chrétien, M. (1994) Structure of the gene for the testis-specific proprotein convertase 4 and of its alternate messenger RNA isoforms. *Genomics* **20**, 231-237
39. Garrell, J. and Campuzano, S. (1991) The helix-loop-helix domain: a common motif for bristles, muscles and sex. *BioEssays* **13**, 493-498
40. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079-1090
41. Gutman, A. and Wasylyk, B. (1990) The collagenase gene promoter contains a TPA and oncogene-responsive unit encompassing the PEA3 and AP-1 binding sites. *EMBO. J.* **9**, 2241-2246
42. Imagawa, M., Chiu, R., and Karin, M. (1987) Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell* **51**, 251-260
43. Wasylyk, C., Flores, P., Gutman, A., and Wasylyk, B. (1989) PEA3 is a nuclear target for transcription activation by non-nuclear oncogenes. *EMBO. J.* **8**, 3371-3378
44. Wasylyk, B., Wasylyk, C., Flores, P., Begue, A., Leprince, D., and Stehelin, D. (1990) The c-*ets* proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature* **346**, 191-193
45. Azizkhan, J.C., Jensen, D.E., Pierce, A.J., and Wade, M. (1993) Transcription from TATA less promoters: dihydrofolate reductase as a model. *Crit. Eukaryotic Gene Expression* **3**, 229-254
46. Roesler, W.J., Vandenbark, G.R., and Hanson, R.W. (1988) Cyclic AMP and the induction of eukaryotic gene transcription. *J. Biol. Chem.* **263**, 9063-9066
47. Bodner, M., Castrillo, J.L., Theill, L.E., Deerinck, T., Ellisman,

M., and Karin, M. (1988) The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **4**, 505-518
48. Umesono, K., Murakami, K.K., Thompson, C.C., and Evans, R.M. (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid and vitamin D3 receptors. *Cell* **65**, 1255-1266
49. Mbikay, M., Seidah, N.G., Chrétien, M., and Simpson, E.M. (1995) Chromosomal assignment of the genes for proprotein convertases PC4, PC5 and PACE4 in mouse and human. *Genomics* **26**, 123-129
50. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411-2414
51. Ayoubi, T.A.K., Creemers, J.W.M., Roebroek, A.J.M., and Van de Ven, W.J.M. (1994) Expression of the dibasic proprotein processing enzyme furin is directed by multiple promoters. *J. Biol. Chem.* **269**, 9298-9303
52. Nakagawa, T., Murakami, K., and Nakayama, K. (1993) Identification of an isoform with an extremely large Cys-rich region of PC6, a *Kex2*-like processing endoprotease. *FEBS Lett.* **327**, 165-171
53. Johnson, J.E., Birren, S.J., Saito, T., and Anderson, D.J. (1992) DNA binding and transcriptional regulatory activity of mammalian *achaete-scute* homologous (MASH) proteins revealed by interaction with a muscle-specific enhancer. *Proc. Natl. Acad. Sci. USA* **89**, 3596-3600
54. Guillemot, F., Lo, L.-C., Johnson, J.E., Auebach, A., Anderson, D.J., and Joyner, A.L. (1993) Mammalian *achaete-scute* homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**, 463-476
55. Shimizu, C., Akazawa, C., Nakanishi, S., and Kageyama, R. (1995) MATH-2, a mammalian helix-loop-helix factor structurally related to the product of *Drosophia* proneural gene *atonal*, is specifically expressed in the nervous system. *Eur. J. Biochem.* **229**, 239-248
56. Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., and Weintraub, H. (1995) Conversion of *Xenopus* ectoderm into neurons by Nruro D, a basic helix-loop-helix protein. *Science* **268**, 836-844
57. Edmondson, D.G. and Olson, E.N. (1993) Helix-loop-helix proteins as regulators of muscle-specific transcription. *J. Biol. Chem.* **268**, 755-758
58. Mains, R.E., Berard, C.A., Denault, J.-E., Zhou, A., Johnson, R.C., and Leduc, R. (1997) PACE4: a subtilisin-like endoprotease with unique properties. *Biochem. J.* **321**, 587-593