

cDNA structure, tissue distribution, and chromosomal localization of rat PC7, a novel mammalian proprotein convertase closest to yeast kexin-like proteinases

(*in situ* hybridization/dibasic residues)

N. G. SEIDAH*[†], J. HAMELIN*[‡], M. MAMARBACHI*[‡], W. DONG*, H. TADROS[§], M. MBIKAY[§],
M. CHRETIEN[§], AND R. DAY*

J. A. DeSève Laboratories of [§]Molecular and *Biochemical Neuroendocrinology, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, PQ, Canada H2W 1R7

Communicated by Henry G. Friesen, Medical Research Council of Canada, Ottawa, Canada, November 29, 1995 (received for review September 20, 1995)

ABSTRACT By using reverse transcription-coupled PCR on rat anterior pituitary RNA, we isolated a 285-bp cDNA coding for a novel subtilisin/kexin-like protein convertase (PC), called rat (r) PC7. By screening rat spleen and PC12 cell λ gt11 cDNA libraries, we obtained a composite 3.5-kb full-length cDNA sequence of rPC7. The open reading frame codes for a prepro-PC with a 36-amino acid signal peptide, a 104-amino acid prosegment ending with a cleavable RAKR sequence, and a 747-amino acid type I membrane-bound glycoprotein, representing the mature form of this serine proteinase. Phylogenetic analysis suggests that PC7 represents the most divergent enzyme of the mammalian convertase family and that it is the closest member to the yeast convertases krp and kexin. Northern blot analyses demonstrated a widespread expression with the richest source of rPC7 mRNA being the colon and lymphoid-associated tissues. *In situ* hybridization revealed a distinctive tissue distribution that sometimes overlaps with that of furin, suggesting that PC7 has widespread proteolytic functions. The gene for PC7 (*Pcsk7*) was mapped to mouse chromosome 9 by linkage analysis of an interspecific backcross DNA panel.

The generation of bioactive proteins and peptides from inactive precursors involves a series of ordered enzyme-mediated posttranslational processing events that often occur along the intracellular secretory pathway and/or at the cell surface. The intracellular tissue-specific processing pattern of any given proprotein depends on the expression levels of the cognate processing enzyme(s). The elaboration of the protein structures of many precursors and their biosynthetic products revealed that cleavage at the C terminus of single basic residues and after pairs of multiple basic amino acids is a general recognition signal for precursor activation (1, 2). Since the identification of kexin as the subtilisin-like serine proteinase responsible for processing pro- α -mating factor in yeast *Saccharomyces cerevisiae* (3), six mammalian homologues were identified and named furin, PC1 (also called PC3), PC2, PC4, PACE4, and PC5 (also called PC6). Analyses of the biosynthesis of the convertases revealed that each member of the family is synthesized as a zymogen with an N-terminal prosegment that is processed at a preferred Arg-Xaa-(Lys/Arg)-Arg \downarrow motif (1, 2, 4–6). The proprotein convertases (PCs) are all capable of cleaving precursors at dibasic residues and some of them can also cleave after single Arg residues within the motif Arg-(Xaa) n -Arg \downarrow , where $n = 2, 4, \text{ or } 6$ (6–8).

The substrate specificities of each convertase were studied by using various precursors and in some cases the known PCs

were not effective in cleaving these proproteins. For example, proinsulin-like growth factor I, which contains a Lys-Xaa-Xaa-Arg \downarrow motif, is not cleaved by furin and is processed into the 70-amino acid (aa) mature insulin-like growth factor I in LoVo cells (9) devoid of furin activity (10). In addition, LoVo cells can process the AIDS virus surface glycoprotein gp160 into gp120/gp41 (11), suggesting that a convertase different from furin can also cleave this precursor and that the cognate convertase(s) must be widely expressed.

In this work, we report the cDNA sequence of a novel rat convertase named rat (r) PC7[¶] as it represents the seventh member of the subtilisin/kexin family. The structural data revealed that rPC7 is the closest member of the convertase family to the yeast kexin-like enzymes. We further present the tissue and cellular distribution of rPC7 obtained by Northern blot analyses and *in situ* hybridization histochemistry revealing a wide expression of PC7 in tissues and cells. Finally, the *Pcsk7* localization on mouse and human chromosomes showed that it is located on a distinct chromosome as compared to the other convertases.

MATERIALS AND METHODS

Reverse Transcription-Coupled PCR and Isolation of rPC7 λ gt11 Clones. Based on a partial human protein sequence reported (12) from BEN cells, and the alignment with the other PCs, we chose the sequences CAVGVAYG and GGE-HNDNCN and derived the following sense and antisense degenerate oligonucleotides: 5'-TG(C,T)GC(C,T)GT-(G,A)GG(C,T)GT(A,C)GC(A,T)TA(C,T)GG-3' and 5'-(A,G)CAGTT(A,G)CA(A,G)TT(G,A)TG(C,T)TC(G,C)CC(A,C)CC-3', respectively. The reverse transcription-coupled PCR protocol on mRNA isolated from rat anterior pituitary and human BEN cells, which gave 285-bp derived cDNAs, was similar to the one presented in ref. 13. The rat cDNA sequence (nt 982–1266, Fig. 1) was used as a probe in Northern blots of tissues and cells leading to the screening λ gt11 cDNA libraries (Clontech) from rat spleen and PC12 cells. From the spleen library, we isolated 29 clones one of which (λ 14C) contained nt 92–1436. The other clones were partially spliced forms of the 3' end of rPC7 cDNA. The composite full-length cDNA sequence of rPC7 was obtained from a clone (λ 10A) isolated from a rat PC12 cDNA λ gt11 library, representing nt 722–3468. All clones were completely sequenced in both directions by using specific 5'-end-labeled fluorescent primers and a Pharmacia ALF DNA sequencer.

Abbreviations: PC, proprotein convertase; r, rat.

[†]To whom reprint requests should be addressed.

[‡]The contributions of J.H. and M.M. were of equal scientific value.

[¶]The sequence reported in this paper has been deposited in the GenBank data base (accession no. U36580).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

→ → → → → → → →
 GAAGTGAAGTGGCTGTCAGTGGCCCTCCGGCTCCCGGGGAAGCTGAGAGTTCCGCGAGTGGCGGCTGCGCGCGCGCGCAGCGGTAG 90
 CAACTGCAACAGTAGCAACGAAGGCTGTGGTCTGCAGCTGTGTTCCAGTGCACATGTCTCACGGTGAAGGTGACATCTCCCTCCAGTT 180
 -36
 MetProLysGlyArgGln
 TCACAAAAAGAGTGGTATGGTTACGAAGACCAGGACCTAAACTTCACAGAGAGCCAGCCTGCTGTTCTGATGCCGAAAGGGAGACAG 270
 -30 -20 -10 -1↓
 LysValProArgLeuAspAlaArgLeuGlyLeuProIleCysLeuCysLeuGluLeuAlaIlePhePheLeuValProGlnValMetGly
 AAAGTCCCACGCTGGATGCCCGCTGGCCCTGCTATCTGCTCTGTCTGGAATTAGCCATCTCTTCTGTTTCCCAGGTCATGGGC 360
 1 10 20 30
 LeuThrGluAlaGlyGlyLeuAspThrLeuGlyAlaGlyGlyLeuSerTrpAlaValHisLeuAspSerLeuGlyGluArgLysGlu
 CTAACAGAGCAGGTGGTCTTGACACCTTGGGTGCAGGGGGCTAAGCTGGCTGTACATCTGGACAGCCCTAGAAGGTGAGAGGAAGAA 450
 40 50 60
 GluSerLeuIleGlnGlnAlaAsnAlaValAlaGlnAlaAlaGlyLeuValAsnAlaGlyArgIleGlyGluLeuGlnGlyHisTyrLeu
 GAGAGTCTGATACACAGGCAAAATGCTGTGGCCAGCAGCGGGCTGGTGAATGCTGGCGCATTGGAGAGCTCCAGGGGCACTACCTC 540
 70 80 90
 PheValGlnProAlaGlyHisGlyGlnAlaMetGluAlaGluAlaMetArgGlnGlnAlaGluAlaValLeuAlaLysHisGluAlaVal
 TTTGTCAGCCAGCTGGGCAGGGCAAGCCATGGAGGCGGAGCCATGCGGCAACAGGCAGAGGCTGTGTAGCCAAAGCATGAAGCTGTG 630
 100 110 120
 ArgTrpHisSerGluGlnArgLeuLeuLysArgAlaLysArgSerIleHisPheAsnAspProLysTyrProGlnGlnTrpHisLeuAsn
 CGTGGCACTCAGAGCAGAGGCTGCTGAAGAGGCCAAGCGCAGCATCCACTTCAATGATCCCAAGTATCCTCAACAGTGGCACCTGAAT 720
 ●●● 140 ◆◆
 AsnArgArgSerProGlyArgAspIleAsnValThrGlyValTrpGluArgAsnValThrGlyArgGlyValThrValValValValAsp
 AATCGCGGAGCCCGAAGAGACATCAATGTGACAGGTGTGGGAGAGAATGTAACATGGGCGAGGGGTGACGGTGGTAGTGGTGGAC 810
 160 170 180
 AspGlyValGluHisThrValGlnAspIleAlaProAsnTyrSerProGluGlySerTyrAspLeuAsnSerAsnAspProAspProMet
 GACGGATGGAGCACCCGTCAGGACATTCGACCCCACTATAGCCAGAGGAGGCTATGACCTCAACTTAATGACCCAGCTCCTATG 900
 190◆◆ 200 ●●● 210
 ProHisProAspGluGluAsnGlyAsnHisHisGlyThrArgCysAlaGlyGluIleAlaAlaValProAsnAsnSerPheCysAlaVal
 CCCCACCTGATGAGGAGAACGGTAACCCACATGGGACCCGGTGTGAGGAGAAATGACAGCTGTGCCAACACAGCTCTCTGTGCGATA 990
 220 230 240
 GlyValAlaTyrGlySerArgIleAlaGlyIleArgValLeuAspGlyProLeuThrAspSerMetGluAlaValAlaPheAsnLysHis
 GGTGTGGCTATGGGAGCCAAATAGCAGGTATCCGGGTGCTGGATGGACCCTCACAGACGATGGAAGCTGTGGCATTCAACAACAC 1080
 250 260 270
 TyrGlnIleAsnAspIleTyrSerCysSerTrpGlyProAspAspGlyLysThrValAspGlyProHisGlnLeuGlyLysAlaAla
 TATCAGATCAATGACATCTACAGCTGCAGCTGGGCGCCGATGATGATGGAAGACAGTGGATGGTCTCTCATCAGCTGGGAAGCTGCC 1170
 280 290◆◆ 300
 LeuGlnHisGlyValMetAlaGlyArgGlnGlyPheGlySerIlePheValValAlaSerGlyAsnGlyGlyGlnHisAsnAspAsnCys
 TTACAACATGGAGTATGGCTGGTCGCCAAGGCTTTGGAAGTATCTTTGTGGTTGCCAGTGGTAAATGGTGGCCAGCACATGACAACTGC 1260
 310 320 330
 AsnTyrAspGlyTyrAlaAsnSerIleTyrThrValThrIleGlyAlaValAspGluGluGlyArgMetProPheTyrAlaGluGluCys
 AACTATGATGGCTATGCCAATCCATCTACACTGTCCACATAGGAGCTGTGGATGAGGAGGACGGATGCCTTTTATGCGAGGAGTGT 1350
 340 350 360
 AlaSerMetLeuAlaValThrPheSerGlyGlyLysMetLeuArgSerIleValThrAspTrpAspLeuGlnLysGlyThrGly
 GCCTCATGTGCGAGTACCTCAGCTGGAGGAGACAAGATGCTGCGGAGCATTGTGACTACTGACTGGGACCTTCAGAAGGGCACTGGC 1440
 ◆◆370 380 390
 CysThrGluGlyHisThrGlyThrSerAlaAlaAlaProLeuAlaAlaGlyMetIleAlaLeuMetLeuGlnValArgProCysLeuThr
 TGCACTGAAGGCCACAGGAACTCAGCTGCAGCCCTCTAGCAGCTGGCATGATAGCTCTCATGCTGCGAGGTGCGGCCCTGCCTCAG 1530
 400 410 420
 TrpArgAspValGlnHisIleIleValPheThrAlaThrGlnTyrGluAspHisArgAlaAspTrpLeuThrAsnGluAlaGlyPheSer
 TGGCGGATGTCAGCACATCATTTGCTTCCAGCCACTCAGTATGAAGATCATCGTGCAGACTGGCTCACTAATGAGGCTGGATTCAGC 1620
 440 450
 HisSerHisGlnHisGlyPheGlyLeuLeuAsnAlaTrpArgLeuValAsnAlaAlaLysIleTrpThrSerValProTyrLeuAlaSer
 CACAGCCATCAGCATGTTTCCGCTCAACCGCTGGAGACTTGTCAATGCAGCCAAAGATCTGGACGCTGTGCCCTTACTTAGCTTCC 1710
 460 470 ●●● 480
 TyrValSerProMetLeuLysGluAsnLysAlaValProArgSerProHisSerLeuGluValLeuTrpAsnValSerArgThrAspLeu
 TATGTCAGCCCATGCTGAAAGAAAATAAGGCTGTCCACGGTCCCCCACTCTCTGGAGGTCCTATGGAATGTGACAGGACGGACCTG 1800
 490 500 (P) 510
 GluMetSerGlyLeuLysThrLeuHisValAlaValThrValSerIleThrHisProArgArgGlySerLeuGluLeuLysLeuPhe
 GAGATGTGGGGTGAAGACCTGGAACATGTGGCGGTGACAGTCTCCATCACTCACCCAGCAGTGGCAGCTTAGAACTGAAACTGTTT 1890
 520 530 540
 CysProSerGlyMetMetSerLeuIleGlyAlaProArgSerMetAspSerAspProAsnGlyPheAsnAspTrpThrPheSerThrVal
 TGTCCAGTGGCATGATGCTTTGATCGGCGCGCCCGCAGCATGGACTCGGACCTTAACGGCTTCAATGACTGGACATTTTCCACTGTG 1980
 550 560 570
 ArgCysTrpGlyGluArgAlaArgGlyValTyrArgLeuValIleArgAspValGlyAspGluProLeuGlnValGlyIleLeuGlnGln
 CGGTCTGGGGGAAAGAGCAAGAGGAGTCTACAGACTGGTTATCAGGATGTAGGAGATGAGCCGCTCCAGGTGGGCATCTCCAGCAG 2070
 580 590 600
 TrpGlnLeuThrLeuTyrGlySerThrTrpSerProValAspIleLysAspArgGlnSerLeuLeuGluSerAlaMetSerGlyLysTyr
 TGGCAGTGCAGCTGATGGCTCCAGCTGGAGTCCAGTAGACATCAAGGACAGACAAAGTCTCTAGAAAGTGTATGAGTGGAAAATC 2160
 610 620 (S) 630
 LeuHisAspAspPheThrLeuProCysProProGlyLeuLysIleProGluGluAspGlyTyrSerIleThrProAsnThrLeuLysThr
 CTGCATGATGACTTCACTGCTGCTGCCACCTGGCAAAATCTCCAGGAGGATGGTTACAGCATTACCCCTAACACACTCAAGACC 2250
 640 650 660
 LeuValLeuValGlyCysPheSerValPheTrpThrIleTyrTyrMetLeuGluValCysLeuSerGlnArgSerLysAlaSerThrHis
 CTTGTGCTGGTGGCTGCTCAGTGTCTTCTGACCAATTACTACATGCTAGAAGTGGCTGAGCCAGAGGAGCAAGGCTTCCACCCAT 2340
 670 680 (P) 690 (P)
 GlyCysArgArgGlyCysCysProTrpProProGlnSerGlnAsnSerLysGluValGlyThrAlaLeuGluSerMetProLeuCysSer
 GGCTGCAGGAGAGGATGCTGCCCTGGCCCCCAAAAGCCAAACTCAAGGAGTGGGACAGCAGTGAATCAATGCCACTGTGCGAGC 2430
 700 710 720
 SerLysAspLeuAspGlyValAspSerGluHisGlyAspCysThrThrAlaSerSerLeuLeuAlaProGluLeuLeuGlyGluAlaAsp
 AGCAAGACCTGGATGGATGGATTCAAGAGCAGGGACTGCACAACTGCCTTAGCTTGTGCGCCAGAGCTGTGGGGGAAAGCTGAC 2520
 (P) (P) 730 740 747
 TrpSerLeuSerGlnAsnSerLysSerAspLeuAspCysProProHisGlnProProAspLeuLysAspGlyGlnIleCys
 TGGAGTCTGTCCAGAAAGTAAAGTGAACCTGGATTTGCTCTCCCAACCCAGACCTGAAGGATGGACAGATCTGCTGACCCAG 2610
 AGCCAGCTTCTTCATGATCAACAGGCTCTTCTCAAACTTGGTTATGAGGCTTCAATCATGGATGCTCAGGAGAGAAATGGTCTGATA 2700
 GTGACATCTGCCACCAAGGCTCTGAAGCATTCTCAGCTTCTCAAGAGGTTGAGGGCCATCTTAATTAAGCCAGTGGTAAAGGACTG 2790
 GTGTTCAATGGCTCTCTAGCTGAGCTTTGCTGACAGGCTTTCGACAGGGGTTGGCGCCAGGCGCCAGGCGCCCTTGAATTCATCTCT 2880
 CGGGCTAGTCACTGCTCGGCCATCAGGACTCCACCCCTTCAAAAGTATGACAGCGGTGCCTGTATACCTAGSAACTCTCAACAG 2970
 AAGTCCGCTGTACAGGCGCCAGAGGAGGCTGCAAGAGGCTAGACATGCTACCCTGAAAGCAGCTGCTTCAATACCCCTTCTCT 3060
 TGTGGCTTCTTCAAGCCGCTCACTCACTACCACTCCCTCTGCTGAAAGAAATGTTGGCGGTGTCAGGAAACTGGGAGGACAG 3150
 GATTCAGCCAGCCACTTGTCTGCTCAGTGGGAGTGTCTGCTGGCTGCAGAGAGCTGAGACACAGTGTGTTCTGAAGATGACAT 3240
 GGGAGCCCTGCCTATGACCACTGCTTCTCTCTGCAAAAGTCTCAGGAAATGGCTCATTCTCCAGGCGCAGCTGTCCGCTGACTTT 3330
 TCTCCATTCTGGCTTCTCCCTTCCACTCTTTGGAGCTAATGTTAATGAATTTTTAATGCTTAAGATTGATTTTACTTTTCT 3420
 AAAGCAACATTTGTTGAATTTTTTCTGCACAGCTTCCAAATTAAGAGGAGT (AAAAAA) 3485

Fig. 1. cDNA and protein sequence of rat PC7. This figure emphasizes the active site (◆◆) Asp, His, and Ser, the important Asn (◆◆◆), the four potential N-glycosylation (●●●●), Tyr sulfation (S), and Ser/Thr phosphorylation (P) sites and the signal peptidase (↓) and zymogen cleavage (↓) sites. The predicted transmembrane sequence (aa 631-647), polyadenylation signal AATAAA, and alternative ATG/TGA start and TAA stop codons are underlined. The identified four exon-intron boundaries are shown by an arrowhead. The characterized mRNA transcription start sites in PC12 cells are also shown (→). The antisense oligonucleotides I and II used in the primer extension experiments are shown.

Northern Blot and *in Situ* Hybridization Analyses. Northern blot analysis was carried out by using an antisense 285-nt (nt 982–1266) ³²P-labeled rPC7 complementary RNA probe and 5 μg of total RNA obtained from the same rat tissues and 22 cell lines as described (5). *In situ* hybridization was carried out as reported (14) by using a 742-nt complementary RNA probe (nt 2170–2911) labeled with ³⁵S-labeled UTP.

Chromosomal Localization of the Mouse *Pcsk7* Gene. Linkage mapping of the gene coding for PC7 was conducted by PCR restriction fragment length polymorphism analysis of DNA from the (C57BL/6Ei × SPRET/Ei)F₁ × SPRET/Ei backcross panel established at The Jackson Laboratory (15). The following sense 5'-GAGCACACCGTCCAGGACATTG-3' (nt 820–841) and antisense 5'-TCCTGCACACCGGGTCCCATGGTG-3' (nt 951–928) oligonucleotides were used to amplify a 670-bp fragment in both C57BL/6 (B6) and *Mus spretus* (SPRET). Digestion of this fragment with *Alu* I generated a distinctive pattern of smaller fragments for B6 and SPRET. This *Alu* I restriction fragment length polymorphism was used to determine the segregation of *Pcsk7* alleles in the backcross mapping panel and to establish its linkage to loci already positioned by using the same panel. After the adopted nomenclature for this gene family (16), the locus coding for PC7 is called *Pcsk7*.

RESULTS

cDNA Sequence Analysis of rPC7. Based on the partial amino acid sequence reported for the human BEN cells' convertase, (12) we developed degenerate oligonucleotides that allowed us to isolate a cDNA from total RNA of BEN cells and of rat anterior pituitary. The deduced partial 95-aa human hPC7 represented the equivalent to aa 208–301 of rPC7 in Fig. 1 (data not shown). The 285-bp rPC7 cDNA [88% identity to human PC7] was used to perform Northern blot and *in situ* hybridization studies in rat tissues and cells. The data showed the widespread expression of this novel mRNA and its high levels in tissues such as colon and spleen and in the cell line PC12.

To obtain the full-length sequence of rPC7, we screened 1.5 × 10⁶ clones from λgt11 cDNA libraries obtained from rat spleen and from PC12 cells. From the spleen library, we isolated 29 clones, all but one (clone λ14C, nt 92–1436) of which represented partially spliced forms of rPC7 heteronuclear RNA. The arrowheads in Fig. 1 point toward the identified intron–exon junctions within our clones. All four sites are located close to the ones observed for PC1, PC2, furin, and PC4 genes (2, 4, 5, 17). These clones allowed the definition of most of the cDNA sequence of rPC7 from nt 92 to the 3' end, which was also confirmed by the sequence of nt 722–3468 deduced from a 2.8-kb cDNA obtained from a PC12 library. The composite cDNA and deduced protein sequence of rPC7 are shown in Fig. 1, where we also included the results of the primer extension of PC12 mRNAs with either of the oligonucleotides I or II (Fig. 1) followed by 5'-RACE-PCR analysis (17), defining the sequence of nt 1–91. This analysis allowed the identification of at least eight 5' ends, which, if confirmed by primer extension analysis, may represent transcription initiation sites in PC12 cells. All of these sites are clustered around the segment of nt 1–46 of the mRNA. This is consistent with the reports that furin, PC1, PC2, and PC4 mRNAs also exhibit more than one start site, possibly due to the absence of a TATA box within their proximal promoters (4, 5, 17).

The 3485-nt sequence of rPC7 contains an open reading frame from nt 253 to 2601, predicting a protein of 783 aa. We note two ATGs at nt 188, 199 followed by corresponding stop codons at nt 251 and 220. The sequence surrounding the proposed initiator methionine GCTGTTCTGATGC, although exhibiting a guanosine at positions –9 and –6 from the ATG (underlined) does not conform to the best sequence surround-

ing such a codon based on Kozak's rules, since we observe a cytidine at positions –3 and +4 (underlined), suggesting that translation of PC7 mRNA will be inefficient (18). The presence of a cytidine at position –3 was also observed for furin (19). A weak 36-aa signal peptide is predicted (Fig. 1), which upon cleavage would result in a 747-aa enzyme. It is noted that this segment represents an unusual signal sequence since it contains five basic residues at its N terminus and a Glu residue within the hydrophobic stretch. Alignment of the PC7 sequence with other PC sequences revealed the typical presence of a pro segment ending at the presumed zymogen activation site **Arg-Ala-Lys-Arg**¹⁰⁴↓, which is similar to the equivalent site demonstrated for the other PCs (6, 7). The mature type I membrane-anchored rPC7 enzyme, obtained after excision of the 104-aa pro segment would, therefore, contain 643 aa with a predicted 17-aa hydrophobic segment (aa 631–647) that, albeit short, may represent a transmembrane domain (based on the PC/GENE HELIXMEM algorithm). After this segment, we note the presence of a 100-aa cytosolic tail and a C-terminal Cys residue. The three active-site residues (Asp¹⁵⁰, His¹⁹¹, and Ser³⁶⁹) and the catalytically important Asn²⁹² are at similar relative positions to those found in the other PCs (6). We predict the presence of four N-glycosylation sites at Asn¹³⁰, Asn¹³⁸, Asn²⁰⁴, and Asn⁴⁷⁴ with Asn¹⁶³ expected not to be glycosylated (due to the presence of Pro in the sequence Asn-Tyr-Ser-Pro). By assuming ≈1.5 kDa for each of the four N-glycosylation sites, the predicted molecular masses of pro-PC7 and PC7 are about 88 and 76 kDa, respectively. By using the Wisconsin GCG package, we predict acidic isoelectric points of pH 5.6 for pro-rPC7 and pH 5.3 for rPC7. A notable difference between PC7 and the other PCs is the absence of the conserved **RRGDL** sequence (1, 6), which in rPC7 is **RRGSL** with Ser (underlined) in place of the usual Asp residue of the RGD sequence. Interestingly, this Ser⁵⁰⁴ residue could be phosphorylated by the physiological kinase recognizing the motif *Ser-Xaa-Glu*. Finally, a sulfation/phosphorylation site at Tyr⁶²¹ is predicted, as well as a number of phosphorylatable Ser (residues 690, 727, and 729) and Thr (residue 681) residues within the putative cytosolic tail of rPC7.

Using the phylogenetic tree alignment program of the GCG package and PILEUP to compare either the catalytic domain (residues 105–575) or the whole sequence of rPC7 with those of all the other PCs revealed that rPC7 is the most closely phylogenetically related member of the subtilisin/kexin family to the yeast enzymes kexin (20), krp (21), XPR6 (22), and kex1 (23) (Fig. 2). A similar result was also obtained when all the known PCs from all species were included in the calculations (data not shown), suggesting that PC7 diverged from the putative common ancestor of the convertases earlier than the others in the mammalian family. It remains to be seen whether this theoretical prediction of phylogenicity of rPC7 could be translated into a functional complementarity. A data bank search (GenBank, September 1995) revealed that rPC7 (nt 1941–2334) exhibits an 82% identity to residues 79–480 of a human 494-bp cDNA (GenBank accession no. H09374).

Tissue and Cellular Expression of rPC7 mRNA. Northern blot analyses of PC7 in rat, mouse, and human cell lines and across rat tissues are shown in Fig. 3. The estimated approximate size of PC7 mRNA is between 4.0 and 4.2 kb. Similar to furin, these data reveal that PC7 is widely expressed in cells with constitutive and regulated secretory pathways (5). In pituitary-derived cell lines, the expression levels of PC7 are highest in the gonadotrophs αT3–1, corticotroph AtT20, and somatomammotroph GH4C1 and very low in GH3 cells. High levels of PC7 mRNA are also found in the insulinoma cells Rin m5F and βTC-3 and in the adrenal cortex cell line Y1, the pheochromocytoma-derived PC12 cells, and the fibroblast cell line Ltk⁻. Note that PC7 is expressed in the furin-deficient (10) human colon carcinoma cell line LoVo. The tissue Northern blot data suggest that PC7 is most abundantly expressed in the

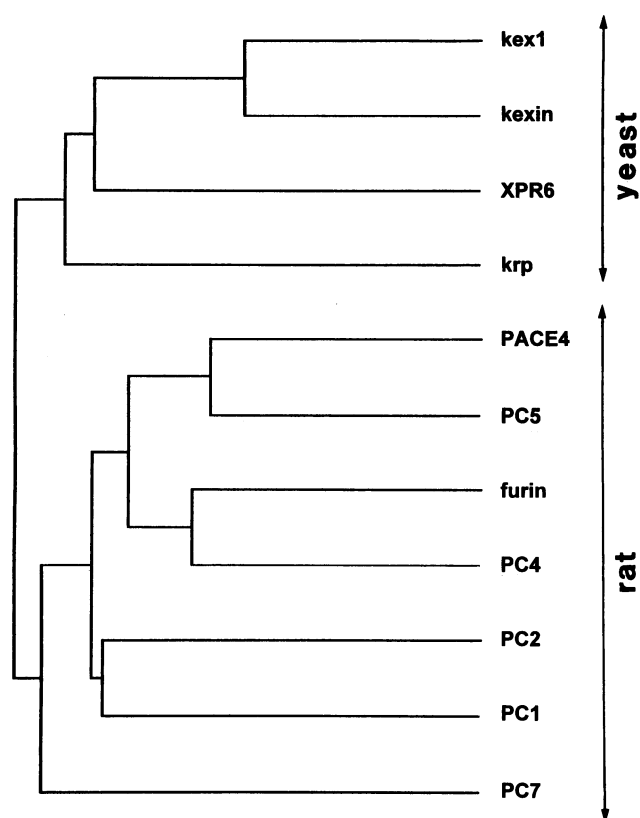


FIG. 2. Dendrogram representing the phylogenetic tree of the seven known rat convertases PC7, PC1, PC2, PC4, furin, PC5, and PACE4 as compared to the yeast kexin-like enzymes krp, XPR6, kexin, and kex1. The data were obtained from the GCG Wisconsin package using the PILEUP and phylogenetic tree programs.

colon and spleen (Fig. 3). Other tissues with a high abundance of PC7 include those of the gut (duodenum, ileum, and jejunum) and the reproductive tissues (testis, epididymus,

ovaries, and oviduct). PC7 is also expressed in the central nervous system and pituitary. Finally, the adrenal, kidney, liver, lung, pancreas, and thyroid also contain significant amounts of PC7 mRNA.

To gain more insight into the tissue distribution of PC7, we carried out a sampling of selected tissues by *in situ* hybridization histochemistry. Representative examples of PC7 in the brain, pituitary, thymus, cerebellum, testis, and spleen are shown in Fig. 4. In the brain, PC7 was widely expressed with the highest levels found in the dentate gyrus of the hippocampus. Also indicated is the expression of PC7 in pyramidal cells, habenula, arcuate nucleus, dorsomedial and ventromedial hypothalamic nuclei (Fig. 4A), and the granular layer of the cerebellum (Fig. 4D). In Fig. 4B, we observed PC7 mRNA in both the anterior and intermediate pituitary lobes. PC7 is also abundantly expressed in the testis (Fig. 4E) within tubules in a stage specific pattern (data not shown), suggesting its expression in germ cells in a very similar fashion to the testis-specific PC4 (24). Finally, we also could demonstrate the high expression of PC7 transcripts in both thymus (Fig. 4C) and spleen (Fig. 4F). The high abundance of PC7 in the thymic cortical region and in the white pulp of the spleen suggests the presence of PC7 in various lymphocyte subpopulations and its possible importance within the immune system.

Localization of *Pcsk7* Gene on Mouse Chromosomes. The gene for PC7 was mapped to mouse chromosome 9 by linkage analysis of a DNA panel from the progeny of an interspecific backcross. The *Pcsk7* locus did not segregate from the *D9Mit21* marker on this panel (Fig. 5).

DISCUSSION

The characterization of the PCs and the analysis of their tissue distribution have had a major impact on our understanding of the tissue-specific processing of precursors at either single or pairs of basic residues. Of all the known members of the PC family, only furin (19) and the PC5/6-B isoform (25) are type I membrane-associated proteins exhibiting both the presence of a C-terminal hydrophobic transmembrane sequence and a cytosolic tail. We now provide evidence that PC7 is the third

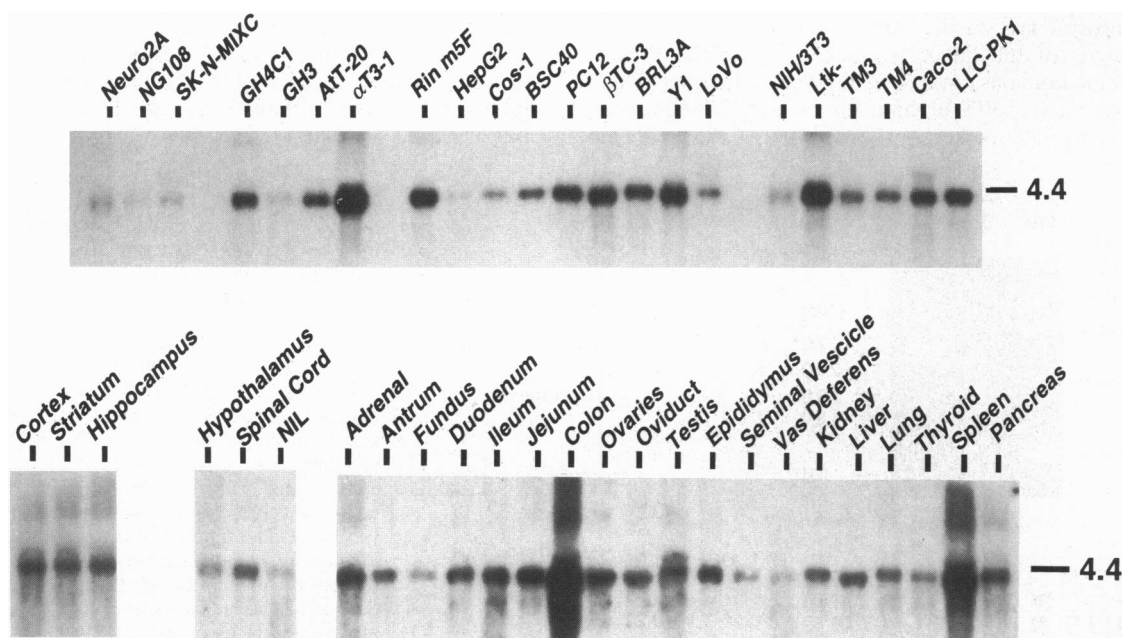


FIG. 3. Distribution of PC7 mRNA by Northern blot analysis in various cell lines (Upper) and rat tissues (Lower). Autoradiograms show the expression of PC7 in brain regions, pituitary, digestive system, reproductive tissues, and various peripheral organs. Total RNA (5 μ g) was loaded into each lane. The major form of PC7 observed is 4.2 kb. A band of lesser intensity is also sometime observed migrating at \approx 6.5 kb. The rPC7 complementary RNA probe used was 285 bp containing nt 982–1266. The x-ray films were exposed for 3 days.

respectively. This observation is relevant to the processing of the AIDS virus surface glycoprotein gp160 at the Arg-Glu-Lys-Arg ↓ site to produce gp120/gp41. Thus, although furin is a major candidate convertase of gp160 (28, 29), evidence exists that other intracellular convertase(s) expressed in the furin-deficient LoVo cells can also process this precursor (11). The expression of PC7 in LoVo cells (Fig. 3) suggests that it could also be a gp160 convertase.

The expression of PC7 deduced from both Northern blot and *in situ* hybridization analyses suggests widespread functions for this convertase that are both distinct and, in some cases, overlapping those of furin. For example, in the central nervous system, furin is very highly expressed in the choroid plexus while PC7 is not. However, similar to furin, PC7 exhibits high levels of expression in some discrete regions. In the case of furin, alternative use of promoters has been suggested as the cause of its differential expression, and its high level in some tissues is likely due to the use of promoters containing a TATA box (4). It may well be that a similar regulation of PC7 expression will be demonstrated. In this regard, we note that the mRNA sequence of PC7 exhibits both an AUG-burdened leader sequence and an AUG initiator codon with an unfavorable context (18). We also isolated at least four clones representing partially spliced mRNAs from spleen. These characteristics are similar to those found in many oncogenes, transcription factors, signal transduction components, and many proteins involved in the immune response (18) and could represent a ploy to modulate the protein levels of PC7, which could be harmful if overproduced. In the case of furin, the present body of evidence suggests that it is an important processing enzyme for precursors that negotiate the constitutive pathway of secretion. Although not definitive in proof, our present histochemical and structural data suggest that like furin, PC7 could cleave precursors within the constitutive secretory pathway. This raises the issue of the need for two furin-like molecules and their possible overlapping functions. Some of these questions may be answered by comparing the substrate specificities and intracellular compartmentalization of furin and PC7.

The sum of all the genetic linkage analyses on the convertases revealed that except for the pair furin/PACE4, the five other genes are found on different chromosomes (4, 5, 16). As shown in Fig. 5C, the *Pcsk7* locus is linked on mouse chromosome 9 to *D9Mit21*, which is a sequence tag derived from *Cyp1a2* the gene encoding cytochrome P1a2. This locus was used as anchor to position *Pcsk7* on the composite map of Imai and Kingsley (30). According to this map, there is syntheny between the *Cyp1a2*-carrying region of mouse chromosome 9 and either human chromosome 11q or 15q. Southern blot analysis of a panel of rodent-human somatic cell hybrids (16) mapped human *PCSK7* to chromosome 11 (data not shown). No mouse mutation near the *Pcsk7* locus is presently known that by its phenotype would make *Pcsk7* a candidate gene.

In conclusion, our data demonstrated the existence of a seventh member of the PC family, which seems to be closest to the yeast kexin-like enzymes. While the substrate specificity and specific physiological functions of PC7 are yet to be defined, this work raises the question of how many other PCs exist and to what extent their proteolytic functions are shared or divergent.

We acknowledge the assistance of J. Rochemont, X.-W. Yuan, and S. Emond. The gift of BEN cells from Dr. G. Hendy (Royal Victoria

Hospital, Montréal) was greatly appreciated. This work was supported by the Medical Research Council of Canada Grant PG-11474 for N.G.S., M.C., and M.M.; Grant MT-13196 for R.D.; Grant UI-12378 for N.G.S.; a Protein Engineering Network of Centres of Excellence (PENCE) grant; and by J. A. DeSève Succession. R.D. is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ).

- Seidah, N. G., Day, R. & Chrétien, M. (1993) *Biochem. Soc. Trans.* **21**, 685–691.
- Steiner, D. F., Smeekens, S. P., Ohagi, S. & Chan, S. J. (1992) *J. Biol. Chem.* **267**, 23435–23438.
- Fuller, R. S., Sterne, R. E. & Thorner, J. (1988) *Annu. Rev. Biochem.* **50**, 345–362.
- Van de Ven, W. J. M. & Roebroek, A. (1993) *Crit. Rev. Oncog.* **4**, 115–136.
- Seidah, N. G., Chrétien, M. & Day, R. (1994) *Biochimie* **76**, 197–209.
- Seidah, N. G. (1995) in *Intramolecular Chaperones and Protein Folding*, eds Shinde, U. & Inouye, M. (R. G. Landes Cie, Austin, TX), pp. 181–203.
- Bresnahan, P. A., Hayflick, J. S., Molloy, S. S. & Thomas, G. (1992) in *Mechanisms of Intracellular Trafficking and Processing of Proproteins*, ed. Peng Loh, Y. (CRC, Boca Raton, FL), pp. 225–250.
- Devi, L. (1991) *FEBS Lett.* **280**, 189–194.
- Duguay, S. J., Lai-Zhang, J. & Steiner, D. F. (1995) *J. Biol. Chem.* **270**, 17566–17574.
- Takahashi, S., Kasai, K., Hatsuzawa, K., Kitamura, N., Misumi, Y., Ikehara, Y., Murakami, K. & Nakayama, K. (1993) *Biochem. Biophys. Res. Commun.* **195**, 1019–1026.
- Ohnishi, Y., Shioda, T., Nakayama, K., Iwata, S., Gotoh, B., Hamaguchi, M. & Nagai, Y. (1994) *J. Virol.* **68**, 4075–4079.
- Gillespie, M. T., Bruzzaniti, A., Goodge, K., Mosely, J. M. & Martin, T. J. (1995) *J. Cell. Biochem. Suppl.* **19B**, 243 (abstr.).
- Seidah, N. G. (1995) *Methods Neurosci.* **23**, 3–15.
- Schäfer, M. K.-H. & Day, R. (1995) *Methods Neurosci.* **23**, 16–44.
- Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J. & Birkenmeier, E. H. (1994) *Mamm. Genome* **5**, 253–274.
- Mbikay, M., Seidah, N. G., Chrétien, M. & Simpson, E. M. (1995) *Genomics* **26**, 123–129.
- Ftouhi, N., Day, R., Mbikay, M., Chrétien, M. & Seidah, N. G. (1994) *DNA Cell Biol.* **13**, 395–407.
- Kozak, M. (1989) *J. Cell Biol.* **115**, 887–903.
- van den Ouweland, A. M. W., van Duijnhoven, H. L. P., Keizer, G. D., Dorssers, L. C. J. & van de Ven, W. J. M. (1990) *Nucleic Acids Res.* **18**, 664.
- Mizuno, K., Nakamura, T., Oshima, T., Tanaka, S. & Matsuo, H. (1988) *Biochem. Biophys. Res. Commun.* **156**, 246–254.
- Davey, J., Davis, K., Imai, Y., Yamamoto, M. & Matthews, G. (1994) *EMBO J.* **13**, 5910–5921.
- Enderlin, C. S. & Ogrzydzak, D. M. (1994) *Yeast* **10**, 67–69.
- Tanguy-Rougeau, C., Wesolowski-Louvel, M. & Fukuhara, H. (1988) *FEBS Lett.* **234**, 464–470.
- Seidah, N. G., Day, R., Hamelin, J., Gaspar, A., Collard, M. W. & Chrétien, M. (1992) *Mol. Endocrinol.* **6**, 1559–1570.
- Nakagawa, T., Murakami, K. & Nakayama, K. (1993) *FEBS Lett.* **327**, 165–171.
- Seizen, R. J., Creemers, J. W. M. & Van de Ven, W. J. M. (1994) *Eur. J. Biochem.* **222**, 255–266.
- Jean, F., Boudreault, A., Basak, A., Seidah, N. G. & Lazure, C. (1995) *J. Biol. Chem.* **270**, 19225–19231.
- Hallenberger, S., Bosch, V., Anglikler, H., Shaw, E., Klenk, H.-D. & Garten, W. (1992) *Nature (London)* **360**, 358–361.
- Decroly, E., Vandenbranden, M., Cogniaux, J., Ruyschaert, J. M., Howard, S. C., Jacob, G., Marshall, G., Kompelli, A., Basak, A., Jean, F., Lazure, C., Benjannet, S., Chrétien, M., Day, R. & Seidah, N. G. (1994) *J. Biol. Chem.* **269**, 12240–12247.
- Imai, K. & Kingsley, D. M. (1994) *Mamm. Genome* **5**, S139–S153.