Volume 312, number 2,3, 213–218 FEBS 11731 November 1992

Molluscan putative prohormone convertases: structural diversity in the central nervous system of *Lymnaea stagnalis*

August B. Smit, Sabine Spijker and Wijnand P.M. Geraerts

Department of Biology, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

Received 19 August 1992

In the cerebral ganglia of the central nervous system of the freshwater snail *Lymnaea stagnalis* many neuropeptides are proteolytically processed from larger prohormones at sites marked by both single and multiple basic amino acids. In the present study we identified cloned cDNA and PCR products corresponding to three putative endoproteases that may be involved in prohormone processing. The cDNA encodes a protein of 653 residues with an overall sequence identity of 60%, 41%, 35%, 40%, and 27% with the recently characterized endoproteases PC2, PC1/3, PC4, furin and Kex2, respectively. The *Lymnaea* preproconvertase has ~80% homology with the catalytic domain, and ~40% and ~50% with the N- and C-terminal part, respectively, of the vertebrate PC2. Two cloned PCR products, L*fur* 1 and L*fur* 2, show highest sequence identity to furin. Expression of the LPC2 gene is exclusively in the central nervous system, where two LPC2 transcripts of 3.0 and 4.8 kb were detected.

PC2-related convertase; Furin-related convertases; cDNA cloning; Polymerase chain reaction; Cerebral ganglia; *Lymnaea stagnalis*

1. INTRODUCTION

Proteolytic processing represents an ubiquitous cellular mechanism that plays a major role in the maturation of prohormones, neuropeptides and other proteins [1,2]. Recently, the coding sequences for a number of vertebrate processing enzymes were reported, i.e. furin [3–5], PC1/PC3 [6–9], PC2 [7,10,11], and PC4 [12]. Furin has been shown to correctly process several precursor proteins, e.g. β-nerve growth factor [13]. Furthermore, furin-like genes have been characterized from invertebrate species, i.e. from *Drosophila melanogaster* (D*fur* 1 and D*fur* 2; [14]) and *Caenorhabditis elegans* (bli-4; [15]). Mouse PC1/3 and PC2 and human PC2 are capable of cleaving the proopiomelanocortin precursor [16], each enzyme at distinct pairs of basic residues. The PC1/3 and PC2 endoproteases are uniquely expressed in endocrine and neural tissue. By contrast, mouse PC4 is exclusively expressed in the testis, and its processing capacity is as yet not established [12]. Furins and the PC1–4 are structurally and functionally related to the prohormone processing enzyme Kex2 [4] of the yeast *Saccharomyces cerevisiae* and to the subtilisins of bacteria [17]. These enzymes all share a well-conserved catalytic domain. Furin and Kex2 share, in addition a preproprotein region, a serine/threonine-rich region, a transmembrane anchor and a cytoplasmic tail, which are, however, not found in the PC1–4.

In the central nervous system (CNS) of the mollusc *Lymnaea stagnalis* various prohormones have been characterized, from which a large variety of neuropeptides can be derived by proteolytic cleavage [18]. Some of these prohormones are exclusively cleaved at dibasic residues, as in the precursors of APGWamide and molluscan insulin-related peptides [19–22]. Other prohormones are, in addition, cleaved at tribasic and tetrabasic sites and single basic residues, as in the precursors of caudodorsal cell hormones-I and -II [23] and FMRFamide [24]. In view of the observed diversity in prohormones and proteolytic cleavage sites, it can be assumed that in the *Lymnaea* CNS different processing enzymes occur that recognize distinct prohormones and/or cleavage sites. In order to address this question, we used polymerase chain reaction PCR and DNA cloning to identify diverse processing enzymes that are structurally related to the vertrebrate PC and furins. We report here the characterization of a cDNA clone encoding *Lymnaea* PC2 (LPC2), as well as PCR products that code for several putative convertases related to furin.

Correspondence address: A.B. Smit, Department of Biology, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands. Fax: (31) (20) 642 9202.

2. MATERIALS AND METHODS

2.1. *Animals*

Adult *L. stagnalis* (shell height, 28–34 mm) bred in the laboratory under standard conditions [25] were used.

2.2. *PCR*

Degenerate oligonucleotides were synthesized, based on amino acid sequences that are conserved among the PC1–4, Kex2, and furins, namely His-Gly-Thr-Arg-Cys-Ala-Gly and Val-Trp/Phe-Ala-Ser-Gly-Asn-Gly-Gly. The sequence of oligo (OL1), which is based on the former conserved amino acid sequence is: 5'-CGCGGATCCA(TC)-GG (GATC)AC(GATC)(AC)G(GATC)TG(TC)GC(GATC)GG-3'

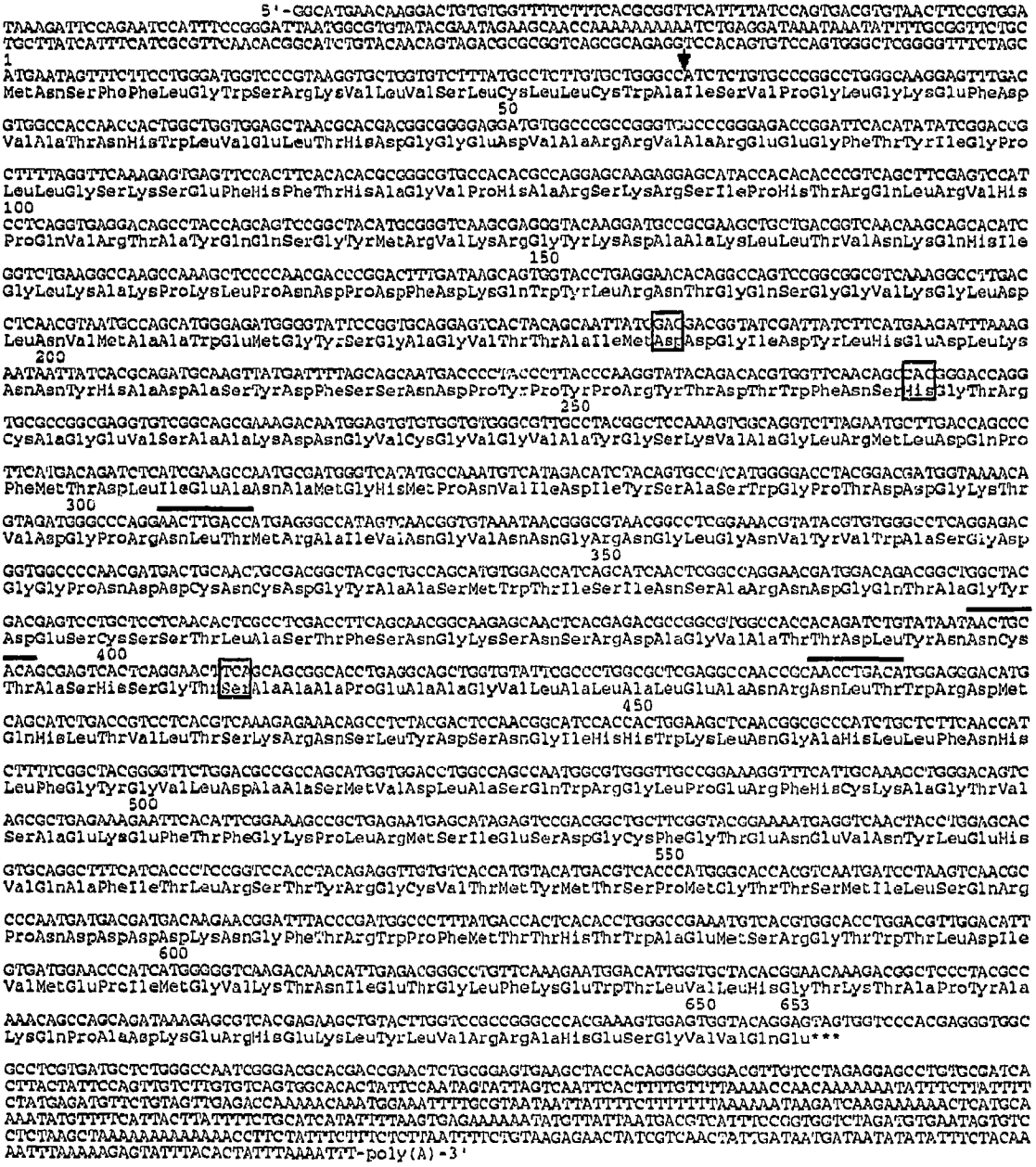

Fig. 1. Nucleotide and predicted amino acid sequence of LPC2. The number of nucleotides is indicated at the end of each line. The number of amino acids are indicated above the sequence. The arrow indicates the putative cleavage site of the signal peptide. The proposed active site involving Asp, His and Ser residues are indicated by boxes. Solid bars above sequencings indicate consensus sequence for *N*-linked glycosylation.

and the sequence of OL2, which is based on the latter amino acid sequence is: 5′-CGCGGATCC(GATC)CC(AG)TT(GATC)CC-(GATC)(GC)(GA) (GATC) GC (GAC) (AC) A (GATC) AC-3′. Two additional degenerate oligonucleotides, OL3 and OL4, were synthesized that are specific for respectively PC3- and furin-related enzymes; OL3: 5′-CGCGGATCCGGI(CTA)T(GATC)(AC)G(GATC)ATG-(CT)T(GATC)GA(TC)GG(GAT C)(AT)T-3′, and OL4: 5′-CGCG-GATTCGG(GATC)(GATC)TI(AC)GIATG(CT)T(GATC)GA(TC) GG(GATC)GA-3′. In the PCR with OL1–4, cDNA was amplified using 50 cycles: 94°C for 45 s, 53°C for 45 s, and 72°C for 1 min. Amplified cDNA was digested with *Bam*HI, subcloned, and sequenced.

2.3. *Screening of the* Lymnaea *cDNA library*

Approximately 80,000 clones of an amplified λZAP II cDNA library of the cerebral ganglia of the CNS of *L. stagnalis* were plated at a density of 20,000 pfu/120 mm^2 filter and absorbed to Hybond-N membranes (Amersham International Corp., UK). Clones were purified by rescreening at a lower plaque density. Hybridization of the membranes was performed with a Digoxigenin labeled LPC2 PCR product, generated between OL1 and OL2 on a single-stranded M13 clone containing LPC2 cDNA, according to the method of Lion and Haas [26]. Prehybridization and hybridization of the filters was performed in 6×SSC (1×SSC: 0.15 M NaCl and 0.015 M Na-citrate), 1% Boehringer Blocking Reagent (BBR; Boehringer Mannheim, FRG),

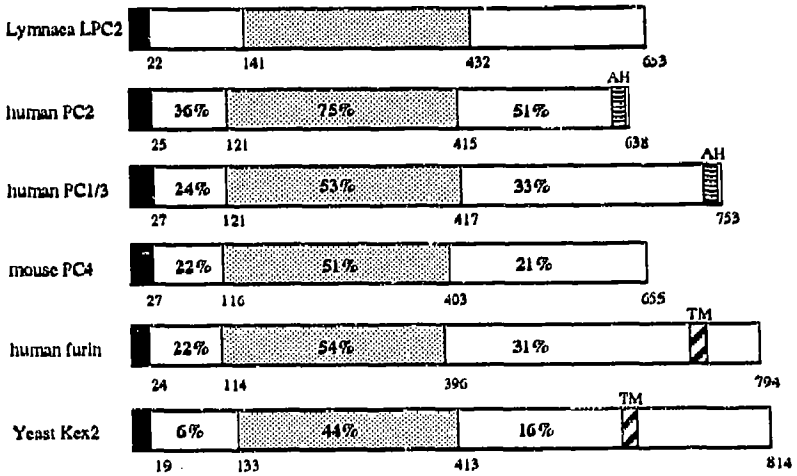

Fig. 2. Schematic representation of the preproproteins of *Lymnaea* LPC2, human PC1/3 [8], human PC2 [10], mouse PC4 [12], yeast Kex-2 [4], and human furin [1]. Numbers indicate amino acid residue from the start of the protein. Percentages indicate the amino acid sequence identity in regions of the vertebrate and yeast proteases with corresponding regions in LPC2. Indicated are the signal peptides (black boxed), the catalytic domains (dotted), transmembrane domains (TM), and putative amphiphatic α-helices (AH).

0.1% *N*-lauroylsarcosine, 0.02% SDS, at 68°C for 16 h. The filters were washed in 1×SSC at 68°C for 45 min, in 0.3% Tween-20/TBS (1×TBS: 0.1 M Tris-HCl and 0.015 M NaCl, pH 7.5) for 5 min, blocked in 1% BBR/TBS, incubated in $1:10^4$ diluted anti-Digoxigenin–alkaline phosphatase conjugate, in 1% BBR/TBS, and washed in 0.3% Tween-20/TBS for 3×10 min. The membranes were then equilibrated in AP-buffer (100 mM diethanolamine-actetate, 2 mM $MgCl_2$, 0.01 mM Zn-acetate, 0.02% thiomerosal) for 2 min, incubated in 0.24 mM 3-′-spiroadamantane)-4-methoxy-4-(3″-phosphorylox y)-phenyl-1,2,-dioxetane (Tropix, USA) in AP buffer in a sealed plastic bag for 2 min. After removal of the substrate the membrane was autoradiographed.

2.4. *Size determination of LPC2, Lfur 1 and Lfur 2 mRNA*

Total RNA was isolated according to the method of Chomczynski and Sacchi [27], and mRNA was isolated from the total RNA using magnetic beads (Dynal A.S., Oslo, Norway). About 4 μg of mRNA from the CNS and various other organs was glyoxylated, fractionated on a 1.6% agarose gel, transferred to a Hybond-N filter, and hybridized in 6×SSC, 0.2% SDS, 5×Denhardts, and 10 mg/ml tRNA and Herring sperm DNA, at 65°C for 16 h with the [γ-^{32}P]dATP 5′-end labeled oligonucleotide probes specific for LPC2 5′-ACATTTGGCA-TATGACCCATCGCATT-3′, *Lfur* 1 5′-CTGAGGGTTCAGCCC-TAGTGATTGG-3′, and *Lfur* 2 5′-ACTGCGATTAAAGCTCA-GCGAGGTA-3′ (specific activity >10^9 dpm/mg). Filters were washed in 1×SSC, 0.2% SDS at 65°C for 20 min, and autoradiographed. Glyoxylated yeast ribosomal RNAs, 26 S (3400 bases) and 17 S (1800 bases), were used as size markers.

2.5. *Nucleotide sequence analysis*

PCR products subcloned in M13mp19 and pBluescript II LPC2 cDNA generated by in vivo excision, were sequenced in both orientations according to the method of dideoxy chain termination [28], using sequenase. Following sequencing from universal primer sites present in the vectors, the sequence information was used to design new primers and the sequencing continued. Reactions were performed with standard nucleotide mixtures, or with dITP as a substitute for dGTP. Sequenase and reaction mixtures were obtained from USB.

3. RESULTS AND DISCUSSION

On the basis of the structural relationship of the catalytic domain of various vertebrate PC, degenerate oligonucleotide primers OL1 and OL2 were designed and used to amplify a PCR fragment with an expected length of 320 bp. Analysis of the PCR mixture by electrophoresis on an agarose gel revealed a single product, which was subcloned and sequenced. It appeared to encode a fragment with highest sequence identity to vertebrate PC2 (data not shown). To identify the full sequence of the *Lymnaea* PC2-like protein, we used this cDNA fragment to screen 80,000 independent clones of a cDNA library of the cerebral ganglia of *L. stagnalis*. Sixty clones were found that gave a hybridization signal and from 16 clones the insert lengths were determined and the longest cDNA of 2778 nucleotides was sequenced (Fig. 1).

The cDNA consists of a single open reading frame of 1959 nucleotides (653 amino acids), flanked by a 269-base-pair 5′ untranslated leader sequence and a 3′ untranslated region of 550 base-pairs. The deduced amino acid sequence reveals that it is organized similarly to the PC1–4 (Fig. 2), with a hydrophobic leader sequence, a subtilisin-like catalytic domain, and a C-terminus lacking a membrane anchor.

Northern blot hybridization of poly(A)$^+$ RNA from the *Lymnaea* CNS, using as a probe a primer directed to the LPC2 sequence, revealed two transcripts of 3 kb and 4.8 kb (see below). The cloned cDNA contains a poly(A) tail and very likely represents the transcript of 3 kb. Translation of the mRNA is probably initiated at the methionine residue at position 1. Cleavage of the signal peptide most likely occurs after residue Ala-22 [29], generating a signal sequence of 22 residues. The proprotein may be cleaved at two consensus motifs for precursor cleavage catalyzed by furin [30], at position 88 and/or 116 (Fig. 3).

Sequence alignment of the predicted protein with the

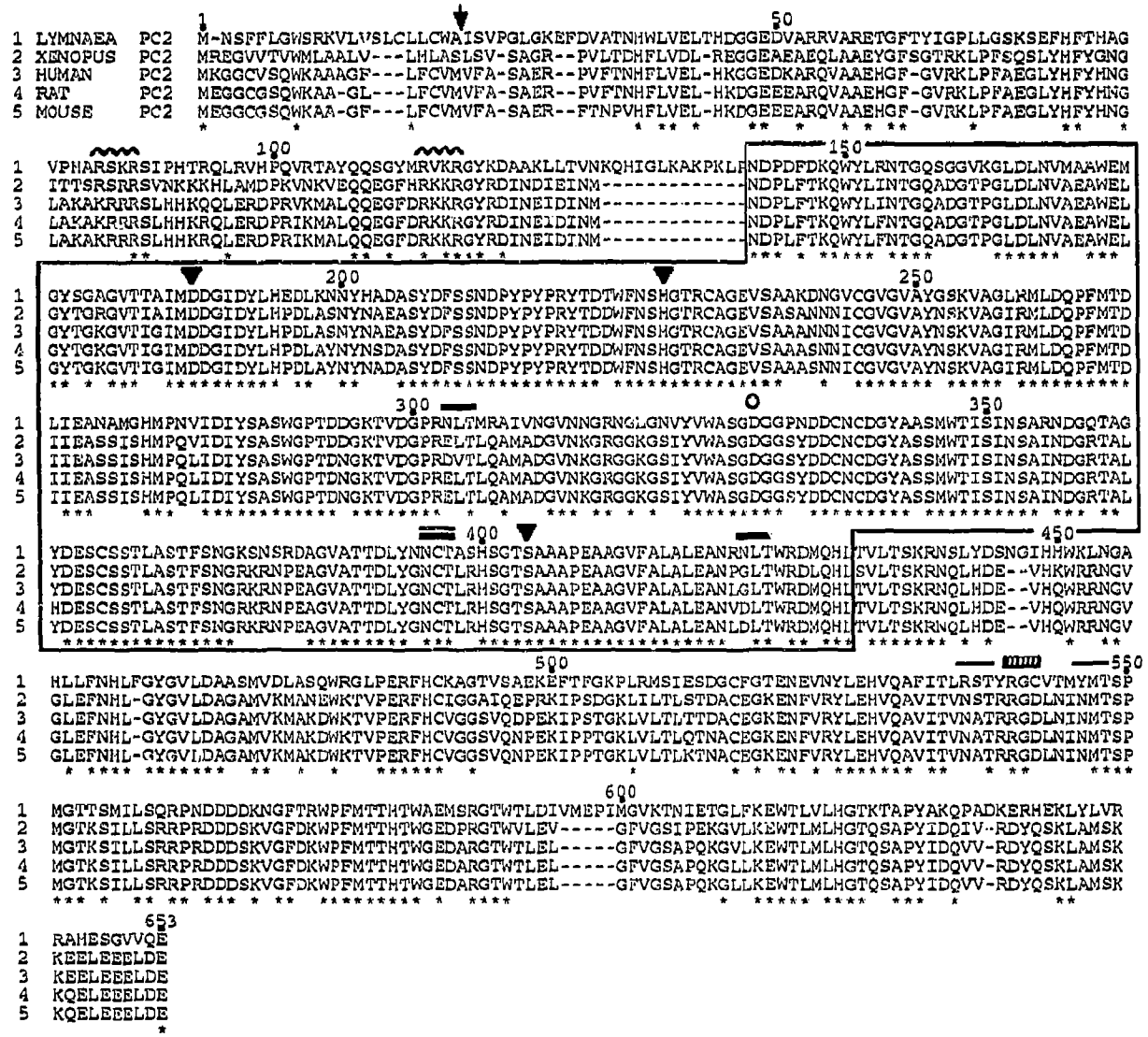

Fig. 3. Amino acid sequence alignments of PC2 of *Lymnaea stagnalis*, *Xenopus leavis* [32], human [8], rat [33], and mouse [7]. The subtilisin-like catalytic domains are boxed and the residues important for catalytic activity, Asp, His and Ser are indicated by a triangle. Residues identical to LPC2 and vertebrate PCs are indicated with an asterisk. Gaps introduced in the alignment are indicated by hyphens. Open circle indicates PC2-specific Asp residue (see text); hatched bar, position of a RGD motif in the vertebrate PC2; wavy lines, position of putative cleavage sites of the proprotein; thick black bars, *N*-glycosylation sites in vertebrate PC2; thin black bars, glycosylation sites in LPC2.

PC1–4, furins and Kex2 reveals that sequence identity is highest with PC2. Interestingly, PC2 specific residues, e.g. Pro-244 and Asp-310 [8], are also present in LPC2: Pro-264 and Asp-330. In all comparisons the catalytically important Asp, His and Ser residues align exactly (Fig. 3). The PC2 sequences characterized in vertebrates show a high degree of amino acid sequence identity, overall ranging from 80% to 90% and in the catalytic domain exceeding 95%. Based on both the sequence identity and specific structural features, LPC2 must be considered a member of the class of PC2 convertases, although the sequence identity of LPC2 and the vertebrate PC2 is significantly lower (~60% for the entire sequence and ~80% for the catalytic domain) than that among vertebrate PC2. From this we conclude that, apart from a few rigidly conserved structural features, the evolutionary constraint on the primary structure of the enzyme is less rigid than suggested by the data of the vertebrate PC2 alone. Interestingly, LPC2 contains sequences that are entirely absent from the vertebrate PC2, e.g. the 13 amino acid region immediately preceding the catalytic domain. Apart from the well-conserved sequences in the catalytic domain, LPC2 and the vertebrate PC2 share several conserved albeit functionally ill-defined C-terminally located domains. However, unlike in PC2 and PC1/3, a clear amphiphatic α-helix that

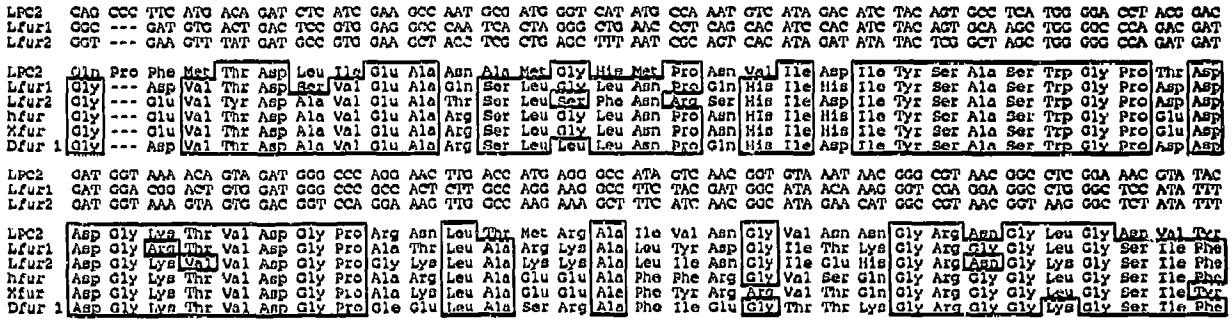

Fig. 4. Nucleotide sequence and amino acid sequence alignment of LPC2 cDNA, L*fur* 1 and L*fur* 2 PCR products, and selected furin and furin-like proteins; h*fur*, human furin [5]; X*fur*, *Xenopus* furin [35]; D*fur* 1, *Drosophila* furin [14]. Identical amino acid residues are boxed. The amino acid sequence depicted covers residue numbers 263–324 of the LPC2 protein.

may interact with the secretory granule membrane is absent in LPC2 (cf. Fig. 2).

In studies of the functioning of members of the subtilisin family, it has been shown that the Asn residue is of crucial importance in the catalytic mechanism [11,31]. It has been suggested that in PC2 the substitution of Asn to Asp considerably narrows the acidic pH range for optimal activity of the enzyme. Because both LPC2 and the vertebrate PC2 share an Asp residue (in LPC2, Asp-330) at a position analogous to that of the Asn residue in the subtilisins and in PC1/3 and PC4, we favor the view that it indeed may be important in the functioning of the enzyme.

All vertebrate PC2 sequences identified so far share three putative *N*-linked glycosylation sites that, if used, may be of importance for the functioning of the convertases. In LPCs, of the three putative *N*-linked glycosylation sites only the glycosylation at position 395 is shared with the vertebrate PC2. The remaining two putative glycosylation sites in LPCs are at unique positions (Asn-303 and Asn-423). This may indicate that only Asn-395 is used as a glycosylation site in PC of both invertebrates and vertebrates. All vertebrate PC2 share an Arg-Gly-Asp sequence which is thought to serve as a receptor recognition signal for cellular matrix proteins [34]. However, because the sequence is not present in LPC2, this suggests that it does not function as a recognition signal in the members of the PC family.

In order to study whether structurally different though related convertases occur in the cerebral ganglia of *Lymnaea*, a PCR approach was chosen in which the LPC2/PC2 specific residue Pro-264 was used to discriminate between LPC2 and novel LPC sequences. Degenerate primers OL3 and OL4 were designed to identify LPC that lack Pro-264 and have at position 265 either a (furin- and Kex2-like) charged residue Glu or Asp, or a (PC3-like) hydrophobic residue Leu, Met, or Ile. Analysis of the PCR mixtures resulting from the combinations OL3/OL2, and OL4/OL2 showed only PCR products of the size expected for the combination OL3/OL2. These products were cloned in M13mp19 and 20 clones were sequenced in which two different LPC encoding products were identified, each from two independent PCR (Fig. 4). Sequence alignment of these fragments indicates two LPC, which are clearly different from LPC2 and the vertebrate PC, but show highest amino acid identity with furin. Therefore, they are tentatively called L*fur* 1 and L*fur* 2. The PCR experiments do not exclude the existence of a PC3-like protein in *Lymnaea*, since such a protein, unlike PC2, may have residues different from Leu, Met or Ile at position 265.

To determine the tissue distribution of LPC2, L*fur* 1 and L*fur* 2, we carried out Northern Blot analysis of poly(A)$^+$ RNA isolated from the CNS and various peripheral organs (Fig. 5). Transcription of the LPC2 gene seems restricted to the CNS, because no LPC2 transcripts were detected in other organs. There were two LPC2 transcripts with lengths of respectively 3.0 and 4.8 kb. The LPC2 transcript of 4.8 kb may represent an

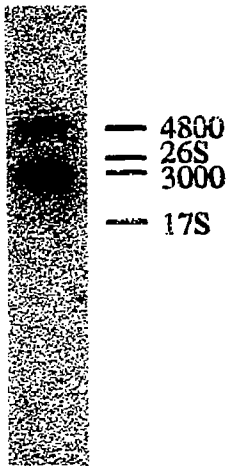

Fig. 5. Size determination of LPC2 mRNA as determined by Northern blotting. Poly(A)$^+$ RNA (4 μg), isolated from the CNS was size fractionated on a 1.6% agarose gel, blotted to Hybond-N and hybridized to ^{32}P-labeled oligonucleotide probes, specific for LPC2. 26 S (3400 bases) and 17 S (1800 bases) indicate the positions of yeast rRNAs (transcript size markers). The sizes of the LPC2 transcripts are 3.0 and 4.8 kb.

alternatively spliced mRNA or a related transcript from an as yet unidentified gene. (Interestingly, two PC2 transcripts of different sizes have also been found in a human insulinoma [10].) In the CNS, the L*fur* 1 and L*fur* 2 transcript size could not be determined due to low abundancy. PCR with L*fur* 1 and L*fur* 2 homologous primers demonstrated that in the cDNA library of the cerebral ganglia the L*fur* 1 and L*fur* 2 cDNAs were present at low frequency (1 out ~2×10^5 clones), which may explain the failure to identify the transcript using Northern blotting. It seems very likely, therefore, that L*fur* 1 and 2 are expressed in a limited set of neurons, or alternatively, at very low levels throughout the CNS.

In summary, the identification of LPC2, L*fur* 1 and L*fur* 2, makes it clear that in *Lymnaea* a class of related though different putative convertases exists, which are structurally related to vertebrate and *Drosophila* convertases. These *Lymnaea* convertases occur in the CNS, very likely in neurons of the cerebral ganglia that express a variety of prohormones and neuropeptides [18].

REFERENCES

[1] Docherty, K. and Steiner, D.F. (1982) Annu. Rev. Physiol. 44, 625–638.
[2] Andrews, P.C., Brayton, K. and Dixon, J.E. (1987) Experentia 43, 784–790.
[3] Roebroek, A.J.M., Schalken, J.A., Leunissen, J.A.M., Onnekink, C., Bloemers, H.P.J. and van de Ven, W.J.M. (1986) EMBO J. 5, 2197–2202.
[4] Fuller, R.S., Brake, A.J. and Thorner, J. (1989) Science 246, 482–486.
[5] Van den Ouweland, A.M.W., van Duijnhoven, J.L.P., Keizer, G.D., Dorssers, L.C.J. and van de Ven, W.J.M. (1990) Nucleic Acids Res. 18, 664.
[6] Seidah, N.G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M.G., Lazure, C., Mbikay, M. and Chretien, M. (1991) Mol. Endocrinol. 5, 111–122.
[7] Seidah, N.G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. and Chrtien, M. (1990) DNA Cell. Biol. 9, 415–424.
[8] Smeekens, S.P., Avruch, A.S., LaMendola, J., Chan, S.J. and Steiner, D.F. (1991) Proc. Natl. Acad. Sci. USA 88, 340–344.
[9] Nakayama, K., Hosaka, M., Hatsuwaza, K. and Murakami, K. (1991) J. Biochem. 109, 803–806.
[10] Smeekens, S.P. and Steiner, D.F. (1990) J. Biol. Chem. 265, 2997–3000.
[11] Robertus, J.D., Kraut, J., Alden, R. and Birktoft, J.J. (1972) Biochemistry 11, 4293–4303
[12] Nakayama, K., Kim, W-s., Torii, S., Hosaka, M., Nakagawa, T., Ikemizu, J., Baba, T. and Murakami, K. (1992) J. Biol. Chem. 267, 5897–5900.
[13] Bresnahan, P.A., Leduc, R., Thomas, L., Thorner, J., gibson, H.L., Brake, A.J., Barr, P.J. and Thomas, G. (1990) J. Cell Biol. 111, 1851–2859.
[14] Roebroek, A.J.M., Pauli, I., Zhang, Y. and van der Ven, W.J.M. (1991) FEBS Lett. 289, 133–137.
[15] Peters, K. and Rose, A. (1991) Worm Breed Gaz. 11, 28.
[16] Benjannet, S., Rondeau, N., Day, R., Chrtien, M. and Seidah, N.G. (1991) Biochemistry 88, 3564–3568.
[17] Siezen, R.J., de Vos, W.M., Leunissen, J.A.M. and Dijkstra, B.W. (1991) Prot. Engineering 7, 719–731.
[18] Geraerts, W.P.M. and Smit, A.B. (1991) in: Molluscan Neurobiology (Kits, K.S., Boer, H.H. and Joosse, J. eds.) pp. 300–308, Elsevier, Amsterdam.
[19] Smit, A.B., Vreugdenhil, E., Ebberink, R.H.M., Geraerts, W.P.M., Klootwijk, J. and Joosse, J. (1988) Nature 331, 535–538.
[20] Smit, A.B., Geraerts, W.P.M., Meester, I., van Heerikhuizen, H. and Joosse, J. (1991) Eur. J. Biochem. 199, 699–703.
[21] Smit, A.B., Thijsen, S.F.T., Geraerts, W.P.M., Meester, I., van Heerikhuizen, H. and Joosse, J. (1992) Mol. Brain Res. 14, 7–12.
[22] Smit, A.B., Jimnez, C.R., Dirks, R.W. and Geraerts, W.P.M. (1992) J. Neurosci. 12, 1709–1715.
[23] Vreugdenhil, E., Jackson, J.F., Bouwmeester, T., Smit, A.B., van Minnen, J., van Heerikhuizen, H., Klootwijk, J. and Joosse, J. (1988) J. Neurosci. 8, 4184–4191.
[24] Saunders, S.E., Bright, K., Kellet, E., Benjamin, P.R. and Burke, J.F. (1991) J. Neurosci. 11, 740–745.
[25] Van der Steen, W.J., van der Hoven, N.P. and Jager, J.C. (1969) Nether. J. Zool. 19, 131–139.
[26] Lion, T. and Haas, O.A. (1990) Anal. Biochem. 188, 335–337.
[27] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[28] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[29] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[30] Hosaka, M., Nagahama, M., Kim, W.-S., Watanabe, T., Hatsuzawa, K., Ikemizu, J., Murakami, K. and Nakayama, K. (1991) J. Biol. Chem. 266, 12127–12130.
[31] Wells, J.A., Cunningham, B.C., Graycar, T.P. and Estell, D.A. (1986) Phil. Trans R. Soc. London A 317, 415–423.
[32] Braks, J.A.M., Guldemond, K.W.C., van Riel, M.C.H.M., Coenen, A.J.M. and Martens, G.J.M. (1992) FEBS Lett. 305, 45–50.
[33] Hakes, D.J., Birch, N.P., Mezey, A. and Dixon, J.E. (1991) Endocrinology 129, 3053–3063.
[34] Rouslahti, E. (1988) Annu. Rev. Biochem. 275, 315–333.