ELSEVIER

FEBS Letters 343 (1994) 27–31

FEBS 13899

FEBS LETTERS

Structural characterization of a *Lymnaea* putative endoprotease related to human furin

August B. Smit[a,*], Sabine Spijker[a], Gregg T. Nagle[b], Susan L. Knock[c], Alexander Kurosky[c], Wijnand P.M. Geraerts[a]

[a]*Graduate School Neurosciences Amsterdam, Research Institute Neurosciences Vrije Universiteit, Faculty of Biology, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands*

[b]*Marine Biomedical Institute, Department of Anatomy and Neuroscience, and* [c]*Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, TX, USA*

Received 18 February 1994; revised version received 14 March 1994

Abstract

A number of peptides have been identified in the central nervous system of the freshwater snail, *Lymnaea stagnalis*, that function as hormones and neurotransmitters/neuromodulators. These peptides are typically proteolytically processed from larger prohormones mostly at sites composed of single or multiple basic amino acid residues. Previously we demonstrated a diversity of putative prohormone convertases that may be involved in prohormone processing in the *Lymnaea* brain. In the present report, we have characterized a cDNA clone encoding a putative endoprotease of 837 amino acids. The primary structure of the endoprotease (L*fur*2) was comparable to that of human furin and contained a putative catalytic domain, a Cys-rich domain, and a transmembrane region. The catalytic domain of L*fur*2 demonstrated about 70% residue identity when compared with human furin, PACE4 and *Drosophila* D*fur*1 and dKLIP-1. The L*fur*2 gene was expressed in the central nervous system as well as various peripheral tissues of *Lymnaea*.

Key words: Furin-related endoprotease; cDNA cloning; Polymerase chain reaction; Central nervous system; Mollusc; *Lymnaeastagnalis*

1. Introduction

The biosynthesis of many biologically active peptides and proteins requires endoproteolytic cleavage and further post-translational modifications of the corresponding precursor proteins. Recently, the coding sequences of a number of prohormone convertases (PCs) involved in the endoproteolytic processing of precursor proteins have been characterized by cDNA cloning. To date, furin [1,2], PACE4 [3], PC1/3 [4–6], PC2 [6–9], PC4 [10], PC5/6A [11,12], and PC6B [13] of vertebrates, D*fur*1, dKLIP-1, and D*fur*2 of the fruitfly *Drosophila melanogaster* [14–16], PC3 of the coelenterate *Hydra vulgaris*, [17] and LPC2 of the freshwater snail *Lymnaea stagnalis* [18] have been identified. These processing enzymes cleave the precursor proteins at basic amino acid residues, e.g. the neuroendocrine endoproteases PC2 and PC3 cleave at dibasic sites (mostly Arg-Arg, or Lys-Arg) [19,20], whereas furin cleaves at a consensus tetrabasic site (Arg-X-Lys/Arg-Arg) [21,22].

We previously [18] described three putative PCs in the central nervous system (CNS) of *Lymnaea* and one in *Aplysia* neuroendocrine bag cells [23]. The PCs were identified by cloning of a full-length cDNA encoding a *Lymnaea* PC2 (LPC2), and polymerase chain reaction (PCR) products encoding partial furin-related convertases, named *Lymnaea* furin1 (L*fur*1) and furin2 (L*fur*2), as well as *Aplysia* furin that have an amino acid sequence identity with furin of 80%, 68% and 71%, respectively. In the *Lymnaea* CNS, a number of prohormones have been described, some of which are exclusively cleaved at dibasic amino acid residues, e.g. the precursors of APGWamide [24] and molluscan insulin-related peptides [25,26]. In addition, others are cleaved at both monobasic and tetrabasic sites, e.g. the precursors of FMRFamide [27] and egg-laying hormone [28]. The identification of three structurally diverse putative endoproteases gives a strong indication that sequence- and/or prohormone-specific endoproteolysis occurs in the *Lymnaea* CNS. L*fur*1 and L*fur*2 primary structural information should prove valuable in understanding the functional aspects of the *Lymnaea* PCs.

In the present paper, we describe the cDNA cloning of L*fur*2. A full-length cDNA was obtained that encodes a *Lymnaea* PC that demonstrates a high degree of homology with human furin. The primary structural characteristics of furins include a Kex2-like catalytic domain, a Cys-rich domain and a transmembrane anchor. The L*fur*2 gene is widely expressed in the CNS as well as in various non-neuronal tissues.

*Corresponding author. Fax: (31) (20) 642 9202.

SSDI 0014-5793(94)00284-3

2. Materials and methods

2.1. Animals

Adult *L. stagnalis* (shell length, 28–34 mm) bred in the laboratory under standard conditions [29] were used.

2.2. PCR

The previously [18] generated PCR product encoding a partial L*fur*2 was cloned. Based on the sequence information of the cloned L*fur*2 PCR product, two nested sense oligonucleotides, OL5: 5′-ATAGGATCCGCTTTCATCAACGGCATAGAAC-3′ and OL6: 5′-GCTCTAGAGAACATGGCCGTAACGGTAAG-3′, were synthesized. In addition, two nested antisense primers were synthesized, OL7: 5′-CGCGGATCCTTACCGTTACGGCCATGTTC-3′, and OL8: 5′-ATAGGATCCGCCATGTTCTATGCCGTTGAT-3′. Amplification of cDNA was performed using OL5 and OL6, or OL7 and OL8, in a nested PCR, in combination with λZAP II primers T33 (5′-GCGCAATTAACCCTCACTAAAGG-3′) and T77 (5′-GCGTAATACGACTCACTATAGGGCGA-3′) for 40 cycles: 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min. Amplified cDNA was digested with *Xba*I and *Bam*HI, respectively, then cloned and sequenced.

2.3. Screening of the Lymnaea cDNA library of cerebral ganglia

Approximately 100,000 clones of an amplified λZAP II cDNA library of the cerebral ganglia of the CNS of *L. stagnalis* were plated at a density of 10^5 pfu/400 cm^2 and absorbed to charged Boehringer membranes (Boehringer-Mannheim, Germany). Clones were purified by screening at a lower plaque density. A PCR product, generated between OL6 and T33 on a single-stranded M13 clone containing L*fur*2 cDNA, was digested with *Bam*HI to remove polylinker sequences. This PCR product was used as a random primed labelled [α-^{32}P]dATP (specific activity >10^9 dpm/μg) probe. Membranes were hybridized in 6× SSC (1× SSC: 0.15 M NaCl and 0.015 M sodium citrate), 0.2% SDS, 5× Denhardt's solution and 10 μg/ml herring sperm DNA at 65°C. The filters were washed in 0.5× SSC, 0.2% SDS, at 65°C for 30 min, and autoradiographed.

2.4. Size determination of Lfur2 mRNA

Total RNA was isolated according to the method of Chomczynski and Sacchi [30], and mRNA was isolated from the total RNA using magnetic beads (Dynal A.S., Oslo, Norway). About 5 μg of mRNA from the CNS or from each of various other organs tested were glyoxylated, fractionated on a 1.5% agarose gel, transferred to a charged Boehringer membrane, and hybridized in 6× SSC, 0.2% SDS, 5× Denhardt's solution, and 10 μg/ml herring sperm DNA, at 65°C with the [α-^{32}P]dATP random primed labelled cDNA (specific activity >10^9 dpm/μg). Filters were washed in 0.5× SSC, 0.2% SDS at 65°C for 30 min, and autoradiographed.

2.5. Subcloning and nucleotide sequence analysis

The cDNA insert of the L*fur*2 clone was excised from the pBluescript II by digestion with *Bam*HI, and subsequently digested with *Sau*IIIA or *Rsa*I and randomly cloned into a *Bam*HI- or *Hin*dII-cut M13mp18, respectively. PCR products, L*fur*2 fragments subcloned in M13mp18, and pBluescript II L*fur*2 cDNA generated by in vivo excision, were sequenced in both orientations according to the dideoxy chain termination method [31] using T7 DNA polymerase. Following sequencing from universal primer sites present in the vectors, the sequence information was used to design new primers and sequencing was continued by primer walking. Sequence alignments were performed as described by Feng and Doolittle [39].

3. Results and discussion

We isolated a L*fur*2 clone from a λZAP II cerebral ganglia-specific cDNA library of *Lymnaea* by using a nested PCR strategy. Based on the sequence of a previously reported PCR product [18], two sense primers (OL5 and OL6) specific for L*fur*2 were designed and used in nested PCRs in combination with lZAP II universal primers (T33 and T77) on different fractions of the library. Analysis of the PCR mixtures by electrophoresis on an agarose gel revealed a single product of ~3.5 kb in the nested amplification with T33 of a single fraction. The cloned and sequenced product appeared to encode the 3′ part of the L*fur*2 sequence. Analysis of a second PCR on the fraction using two L*fur*2-specific antisense primers (OL7 and OL8) in combination with T77 showed a single product of 1,100 bp, which was also cloned and sequenced. Sequence alignment of this fragment with known PCs showed that it represented the 5′ part of the L*fur*2 cDNA. To identify the full-length sequence of the L*fur*2 clone, we used this 5′ cDNA fragment to screen 100,000 independent clones of the positive fraction of the cerebral ganglia library. Only one positive clone was isolated.

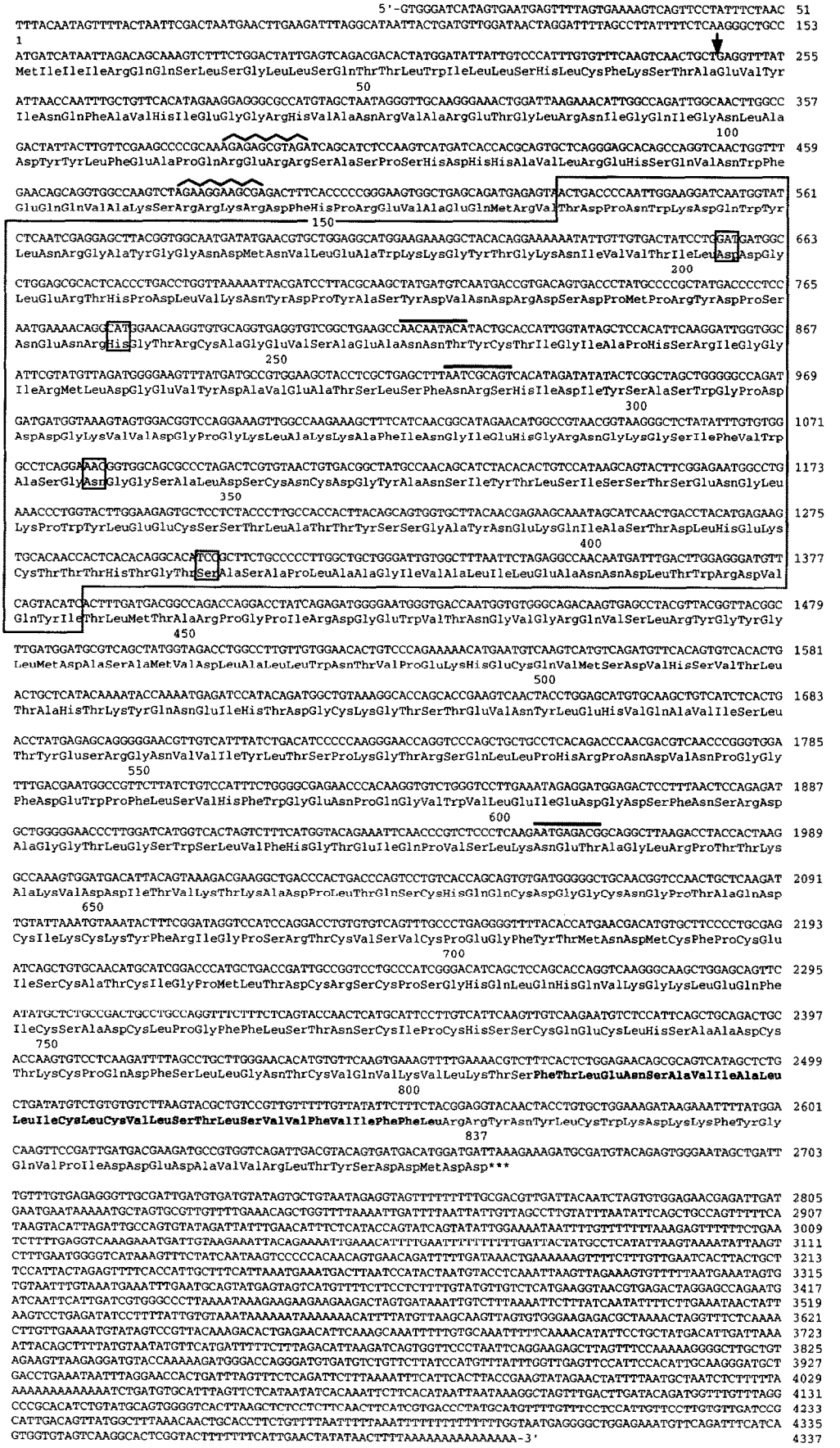
Sequencing of this clone revealed that it had a total length of 4,337 nucleotides (Fig. 1) and was organized as a single open reading frame of 2,511 nucleotides (837 amino acids) flanked by a 153 bp 5′ untranslated leader sequence and a 3′ untranslated region of 1673 bp. A short poly(A) stretch was found at the 3′ end of the cDNA clone; however, a consensus sequence for polyadenylation was absent. Taking into account that Northern blot analysis (see below) detected a transcript of ~6.5 kb, we concluded that the cloned cDNA of 4337 bp was not a full-length L*fur*2 transcript, and lacked ~2 kb of the 3′-untranslated cDNA.

Comparison (Fig. 2) of the sequences of known mammalian and invertebrate Kex2-like PCs with the deduced amino acid sequence of L*fur*2 indicates that the organization of preproL*fur*2 is structurally similar to preprofurin (Figs. 1 and 2), with a hydrophobic leader sequence, a propeptide with two consensus sequences for a tetrabasic cleavage site [21], a Kex2-like catalytic domain, and a C-terminus having a Cys-rich domain and a putative transmembrane domain. The signal peptide is most likely cleaved after residue Ala-31 [32], giving rise to a signal sequence of 31 residues. The proprotein may be cleaved either at the tetrabasic site at position 81 or at position 115 (Fig. 1). The important active site residues, Asp, Asn, His and Ser, are present in L*fur*2, as well as three putative sites for N-linked glycosylation, at Asn-222, Asn-258, and Asn-602 (Fig. 1).

Sequence alignment of L*fur*2 shows a particularly high

→

Fig. 1. Nucleotide sequence and predicted amino acid sequence of L*fur*2. The number of nucleotides is indicated at the end of each line; the number of amino acids is indicated above the sequence. The arrow indicates the predicted cleavage site of the signal peptide. The Kex2-like catalytic domain is boxed and the Asp, His, Asn and Ser residues proposed to be involved in the active site are indicated by small boxes. Solid bars above the sequence indicate consensus for N-linked glycosylation. Residues of the putative transmembrane domain are indicated in boldface. Wavy lines indicate the presence of putative tetrabasic cleavage sites of the proregion.



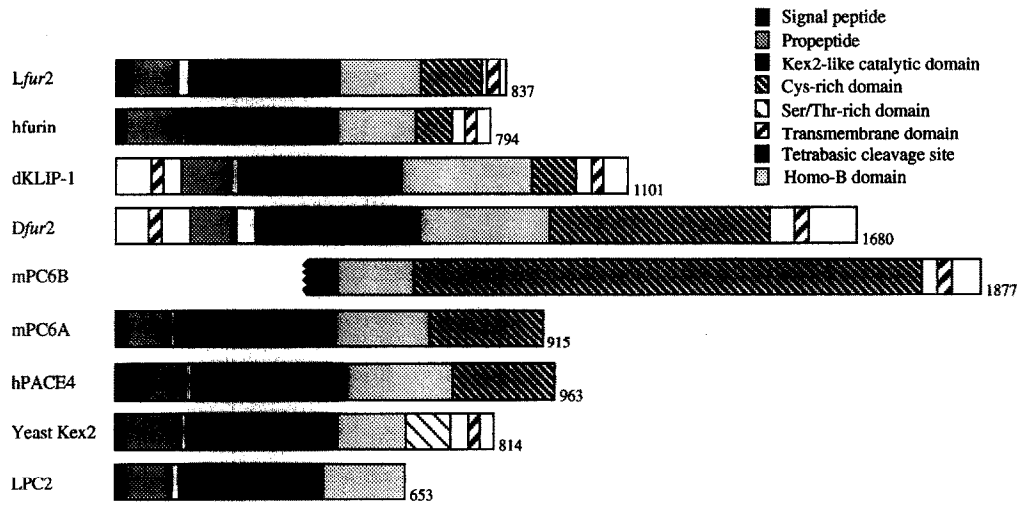


Fig. 2. Schematic representation of different protein domains of L*fur*2, furin and furin-related proteins and other prohormone convertases. Shown are: *Lymnaea* L*fur*2, human furin (hfurin) [4], *Drosophila* dKLIP-1 [15] and D*fur*2 [16], mouse PC6B [13] and PC6A [12], human PACE4 [3], yeast Kex2 [38] and *Lymnaea* PC2 [18]. The numbers indicate the amount of amino acid residues in each of the preproproteins. Only one of the two putative tetrabasic sites (cf. Fig. 1) is indicated.

amino acid sequence identity in the catalytic domain with other furins (67–70%) and related proteins (Table 1). However, in the C-terminal region the sequence identity with furin and related proteins is dramatically less. Also, the various domains in the C-terminal region differ remarkably in length. For example, the length of the Cys-rich domain varies from 89 amino acids in human furin to 1,115 amino acids in PC6B. Sequence alignment of the homo-B domain (Fig. 2) showed that only PC6B can be aligned correctly to the L*fur*2 sequence, having an amino acid sequence identity of 46%. The Cys-rich domain of L*fur*2 (131 amino acids) consists of a 3 times-repeated stretch of about 50 amino acid residues with a characteristic Cys motif that is found in several PCs (Fig. 3). The Cys-related motifs, which are separated from each other by 10 residues, are consistent with the consensus sequence, $CX_2CX_3CX_2CX_{5-7}CX_2CX_{10-15}CX_{3-5}C$ [13,16]. The Cys-rich domain is not essential for proteolytic activity of furin [22] but may play a role in the stabilization and/or intracellular targeting and localization of furin.

Table 1
Amino acid sequence identities (expressed as percentage identity) between the catalytic domain of L*fur*2 and those of PACE4 [3], dKLIP-1 [15], D*fur*1 [14], D*fur*2 [16], human furin [1], mouse PC6A [12], human PC1/3 [5], LPC2 [18], and Kex2 [38]

	Catalytic domain (%)	Species
PACE4	70	Human
dKLIP-1	69	*Drosophila*
D*fur*1	69	*Drosophila*
D*fur*2	67	*Drosophila*
Furin	68	Human
PC6A	68	Mouse
PC1/3	65	Human
LPC2	52	*Lymnaea*
Kex2	44	*Saccharomyces*

L*fur*2 contains a stretch of hydrophobic amino acids near the C-terminus of the protein (Fig. 1) which may serve as a transmembrane region, as seen in furin and related endoproteases (Fig. 2). The transmembrane domain of furin serves as a retention signal for the Golgi complex, thereby concentrating it in the trans-Golgi network [33,34], and therefore furin can function in both the constitutive and regulated secretory pathway. By contrast, PC2 and PC3, which lack a transmembrane domain, are targeted to the regulated secretory pathway and eventually become localized in dense-core vesicles [35–37]. Based on the differences in structural organization, in particular the presence or absence of a Cys-rich domain and a transmembrane domain, we can predict that L*fur*2 and the previously identified LPC2 [18] display different patterns of subcellular localization, i.e., L*fur*2 is localized in the trans-Golgi network and LPC2 is localized both in the trans-Golgi network and in dense-core vesicles. Therefore, we predict that L*fur*2 will be functional in the constitutive secretory pathway that is present in many cells, whereas LPC2 is restricted to cells specialized in regulated secretion of peptides and that are predominantly located in the CNS. To examine the possibility of a tissue-specific pattern of expression of the L*fur*2 and LPC2 encoding genes, we performed Northern blot analysis of mRNA of the CNS and various

618 -CHQQCDGGCNGCNGPTAQDCIKCKY-FRI-GPSRT------CVSVC-(10 a.a.) -675
676 -CFP-CQISCATCIGPMLTDCRSCPSGHQLQHQVKGKLEQFICSADC-(10 a.a.) -730
731 -CIP-CHSSCQECLHSAA-DCTKCPQDFSLLGNS--------C -762

Fig 3. Amino acid sequence alignment of the three Cys-rich repeats of L*fur*2. Cysteine residues are indicated in boldface. Gaps introduced into the alignment are indicated by hyphens. The start and end of each repeat is indicated at the beginning and the end of each line.

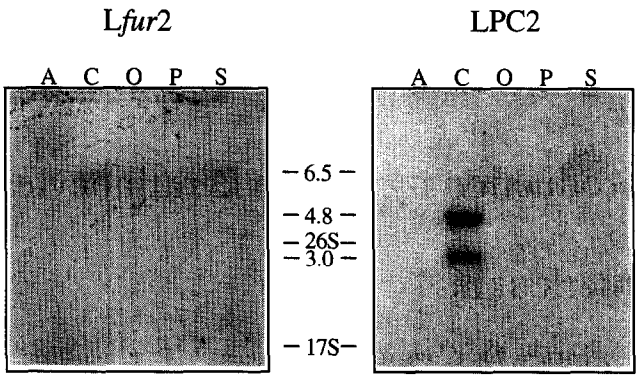


Fig. 4. Northern blotting of the CNS and peripheral organs of *Lymnaea*. Two identical Northern blots are shown of the albumen gland (A), CNS (C), ovotestis (O), prostate (P) and salivary (S) gland. In each lane 5 μg mRNA was applied and hybridized to radiolabelled cDNA probes specific for L*fur*2 (left panel) and LPC2 (right panel). 26 S (3,400 bases) and 17 S (1,800 bases) indicate the positions of yeast rRNAs (transcription size markers). The size of the L*fur*2 transcript is 6.5 kb; those of the LPC2 transcripts 4.8 and 3.0 kb.

peripheral organs (Fig. 4) using L*fur*2- and LPC2-specific cDNA probes. Indeed, as predicted, the expression of the LPC2 gene was exclusively in the CNS, whereas expression of the L*fur*2 gene was found in both the CNS and peripheral organs (except for the albumen gland).

In conclusion, we have identified the primary structure of a furin-related PC, L*fur*2, in the mollusc *Lymnaea*. Comparisons of the overall organization and the expression patterns of the L*fur*2 gene suggest that L*fur*2, like furin, is located in the Golgi apparatus and may function in both the constitutive and regulated secretory pathway; by contrast, LPC2 is very likely restricted to the neuronal-regulated secretory pathway.

References

[1] 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.
[2] 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.
[3] Kiefer, M.C., Tucker, J.E., Joh, R., Landsberg, K.E., Saltman, D. and Barr, P.J. (1991) DNA Cell Biol. 10, 757–769.
[4] Seidah, N.G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M.G., Lazure, C., Mbikay, M. and Chrétien, M. (1991) Mol. Endocrinol. 5, 111–122.
[5] Smeekens, S.P., Avruch, A.S., LaMendola, J., Chan, S.J. and Steiner, D.F. (1991) Proc. Natl. Acad. Sci. USA 88, 340–344.
[6] Seidah, N.G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. and Chrétien, M. (1990) DNA Cell. Biol. 9, 415–424.
[7] Smeekens, S.P. and Steiner, D.F. (1990) J. Biol. Chem. 265, 2997–3000.
[8] Robertus, J.D., Kraut, J., Alden, R. and Birktoft, J.J. (1972) Biochemistry 11, 4293–4303.
[9] 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.
[10] 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.
[11] Lusson, J., Vieau, D., Hamelin, J., Day, R., Chrétien, M. and Seidah, N.G. (1993) Proc. Natl. Acad. Sci. USA 90, 6691–6695.
[12] Nakagawa, T., Hosaka, M., Torii, S., Watanabe, T., Murakami, K. and Nakayama, K. (1993) J. Biochem. (Tokyo) 113, 132–135.
[13] Nakagawa, T., Murakami, K. and Nakayama, K. (1993) FEBS Lett. 327, 165–171.
[14] Roebroek, A.J.M., Pauli, I., Zhang, Y. and van de Ven, W.J.M. (1991) FEBS Lett. 289, 133–137.
[15] Hayflick, J.S., Wolfgang, W., Forte, M.A. and Thomas, G. (1992) J. Neurosci. 12, 705–717.
[16] Roebroek, A.J., Creemers, J.W.M., Pauli, I.G.L., Kurzik-Dumke, U., Rentrop, M., Gateff, E.A.F., Leunissen, J.A.M. and van de Ven, W.J.M. 1992) J. Biol. Chem. 267, 17208–17215.
[17] Chan, S.J., Oliva Jr., A.A., LaMendola, J., Grens, A., Bode, H. and Steiner, D.F. (1992) Proc. Natl. Acad. Sci. USA 89, 6678– 6682.
[18] Smit, A.B., Spijker, S. and Geraerts, W.P.M. (1992) FEBS Lett. 312, 213–218.
[19] Benjannet, S., Rondeau, N., Day, R., Chrétien, M. and Seidah, N.G. (1991) Proc. Natl. Acad. Sci. USA 88, 3564–3568.
[20] Thomas, L., Leduc, R., Thorne, B.A., Smeekens, S.P., Steiner, D.F. and Thomas, G. (1991) Proc. Natl. Acad. Sci. USA 88, 5297–5301.
[21] 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.
[22] Hatsuzawa, K., Murakami, K. and Nakayama, K. (1992) J. Biochem. (Tokyo) 111, 296–301.
[23] Nagle, G.T., van Heumen, W.R.A., Knock, S.L., Garcia, A.T., McCullough, D.A. and Kurosky, A. (1993) Comp. Biochem. Physiol. 105B, 345–348.
[24] Smit, A.B., Jiménez, C.R., Dirks, R.W. and Geraerts, W.P.M. (1992) J. Neurosci. 12, 1709–1715.
[25] Smit, A.B., Geraerts, W.P.M., Meester, I., van Heerikhuizen, H. and Joosse, J. (1991) Eur. J. Biochem. 199, 699–703.
[26] Smit, A.B., van Marle, A., van Elk, R., Bogerd, J., van Heerikhuizen, H. and Geraerts, W.P.M. (1993) J. Mol. Endocrinol. 11, 103–113.
[27] Saunders, S.E., Bright, K., Kellet, E., Benjamin, P.R. and Burke, J.F. (1991) J. Neurosci. 11, 740–745.
[28] 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.
[29] Van der Steen, W.J., van der Hoven, N.P. and Jager, J.C. (1969) Neth. J. Zool. 19, 131–139.
[30] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[31] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[32] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[33] 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.
[34] Molloy, S.S., Thomas, L., VanSlyke, J.K., Stenberg, P.E. and Thomas, G. (1994) EMBO J. 13, 18–33.
[35] Davidson, H.W., Rhodes, C.J. and Hutton, J.C. (1988) Nature 333, 93–96.
[36] Christie, D.L., Batchelor, D.C. and Palmer, D.J. (1991) J. Biol. Chem. 266, 15679–15683.
[37] Kirchmair, R., Egger, C., Gee, P., Hogue-Angeletti, R., Fischer-Colbrie, R., Laslop, A. and Winkler, H. (1992) Neurosci. Lett. 143, 143–145.
[38] Fuller, R.S., Brake, A.J. and Thorner, J. (1989) Science 246, 482–486.
[39] Feng, D.F. and Doolittle, R.F. (1990) Methods Enzymol. 183, 659–668.