Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 3591–3595, April 1995
Evolution

Proprotein convertases in amphioxus: Predicted structure and expression of proteases SPC2 and SPC3

(precursor processing/subtilisin-related endoproteases/prohormone biosynthesis/protochordates)

ANTHONY A. OLIVA, JR.*, DONALD F. STEINER†, AND SHU JIN CHAN

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637

Contributed by Donald F. Steiner, December 21, 1994

ABSTRACT **SPC2 and SPC3 are two members of a family of subtilisin-related proteases which play essential roles in the processing of prohormones into their mature forms in the pancreatic B cell and many other neuroendocrine cells. To investigate the phylogenetic origins and evolutionary functions of SPC2 and SPC3 we have identified and cloned cDNAs encoding these enzymes from amphioxus (*Branchiostoma californiensis*), a primitive chordate. The amino acid sequence of preproSPC2 contains 689 aa and is 71% identical to human SPC2. In contrast, amphioxus preproSPC3 consists of 774 aa and exhibits 55% identity to human SPC3. These results suggest that the primary structure of SPC2 has been more highly conserved during evolution than that of SPC3. To further investigate the function(s) of SPC2 and SPC3 in amphioxus, we have determined the regional expression of these genes by using a reverse transcriptase-linked polymerase chain reaction (RT-PCR) assay. Whole amphioxus was dissected longitudinally into four equal-length segments and RNA was extracted. Using RT-PCR to simultaneously amplify SPC2 and SPC3 DNA fragments, we found that the cranial region (section 1) expressed equal amounts of SPC2 and SPC3 mRNAs, whereas in the caudal region (section 4) the SPC2-to-SPC3 ratio was 5:1. In the mid-body sections 2 and 3 the SPC2-to-SPC3 ratio was 1:5. By RT-PCR we also determined that amphioxus ILP, a homologue of mammalian insulin/insulin-like growth factor, was expressed predominately in section 3. These results suggest that the relative levels of SPC2 and SPC3 mRNAs are specifically regulated in various amphioxus tissues. Furthermore, the ubiquitous expression of these mRNAs in the organism indicates that they are involved in the processing of other precursor proteins in addition to proILP.**

In eukaryotic cells, proteins destined for export or the plasma membrane surface are often synthesized as larger precursor forms which are proteolytically processed during subsequent transport and maturation. The initial endoproteolytic cleavage is performed by proprotein convertases which recently have been identified as members of the subtilisin family of serine endoproteases. These enzymes recognize a motif of paired or multiple basic residues with a strong preference for arginine at the P1 site (1–3). In yeast a single subtilisin-related proprotein convertase, kexin, has been found (4). However, mammals express at least six members of this family, including SPC1 (furin) (5), SPC2 (PC2) (6, 7), SPC3 (PC3/PC1) (8, 9), SPC4 (PACE4) (10), SPC5 (PC4) (11), and SPC6 (PC5/PC6) (12, 13). These enzymes are related by a similar catalytic core domain but differ markedly in their carboxyl-terminal sequences.

We have investigated the structures of SPC2 and SPC3 and the functional roles of these enzymes in the processing of proinsulin to insulin. Studies have shown that both SPC2 and SPC3 are expressed in islet B cells (as well as in other neuroendocrine cells) (6–9), and *in vitro* transfection assays (12) have shown that these enzymes can cleave proinsulin appropriately. Here we report the cDNA sequences‡ and expression of SPC2 and SPC3 from the protochordate amphioxus. We have previously shown that amphioxus expresses an insulin-like peptide (ILP) mRNA (13). The putative precursor, proILP, is like proinsulin in that it contains a C-peptide which is flanked by paired basic residues, but it is also similar to the pro-insulin-like growth factors in that proILP also contains a carboxyl-terminal E domain which is probably removed during maturation. Our goals are to characterize the proprotein convertases which are expressed in protochordates and more specifically identify the convertases involved in processing proILP to its active form.

MATERIALS AND METHODS

Materials. Oligonucleotides were synthesized on an Applied Biosystems model 380B DNA synthesizer. Radioisotopes were purchased from Amersham. Plasmid pGEM4Z was obtained from Promega.

RNA and Reverse Transcriptase-Linked Polymerase Chain Reaction (RT-PCR) Assays. Live amphioxus (*Branchiostoma californiensis*) were purchased from Pacific Bio-Marine (Venice, CA). Total RNA was isolated by tissue homogenization in guanidine thiocyanate and centrifugation through a 5.7 M CsCl cushion as described (13). RNA was quantitated by absorbance at 260 nm (A_{260} = 1 at 40 μg/ml) and its integrity was verified by UV fluorescence on denaturing (0.66 M formaldehyde) 1% agarose gels stained with ethidium bromide at 1 μg/ml.

cDNA was synthesized with a Moloney murine leukemia virus (M-MLV) reverse transcriptase kit purchased from GIBCO/BRL. After 1-hr incubation at 37°C, the reverse transcriptase was inactivated by heating to 100°C for 3 min and an aliquot was removed for PCR. For the subtilisin-related proprotein convertases, cDNA fragments which contained the catalytic domain were amplified by using the following degenerate oligonucleotide primers: SQ-3 (sense), 5′-CAYGGNACNCGNTGYGC; SQ-4 (sense), 5′-CAYGGNACNAGRTGYGC; QS-1 (antisense), 5′-YTGCATRTCYCTCCANGT; and QS-2 (antisense), 5′-YTGCATRTCNCGCCANGT, in which R = G or A, Y = C or T, and N = G, A, T, or C. PCR was performed for 40 cycles each at 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min.

For RT-PCR to simultaneously amplify amphioxus SPC2 (480-bp) and SPC3 (334-bp) cDNA fragments, the following primers were used: SPC2-1 (sense), GGATGTTGGACCAAC-

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.

Abbreviations: ILP, insulin-like peptide; RT-PCR, reverse transcriptase-linked polymerase chain reaction; RACE, rapid amplification of cDNA ends.

*Present address: Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 75226.
†To whom reprint requests should be addressed.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U22051 and U22052).

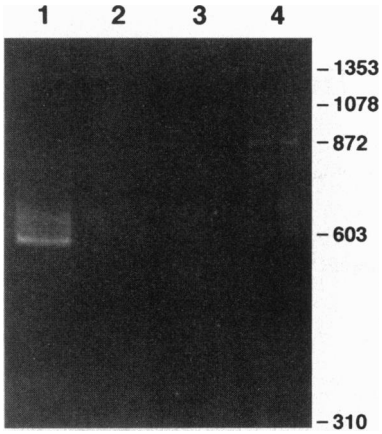

FIG. 1. Amplification of subtilisin-related proprotein convertase catalytic domains by PCR. Amphioxus cDNA reverse transcribed from 0.1 μg of poly(A)$^+$ RNA was subjected to 30 cycles of PCR using oligonucleotide primers SQ-3/QS-1 (lane 1), SQ-4/QS-1 (lane 2), SQ-3/QS-2 (lane 3), or SQ-4/QS-2 (lane 4). Primer sequences are listed in *Materials and Methods.* The reaction products were analyzed by electrophoresis in a 5% polyacrylamide gel and stained with ethidium bromide. Molecular weight standards (labeled in terms of bp) were φX174 DNA digested with *Hae* III.

CGTTC; SPC2-2 (antisense), CAGTGCCAGCGCGAAAACC; SPC3-1 (sense), TCCATCTTTGCCTGGGCGTC; and SPC3-2 (antisense), CGATCAGGTGCTGTACGTCTC. The template contained 100 ng of amphioxus RNA reverse transcribed into cDNA. PCR was performed for 30 cycles each at 94°C for 1 min, 56°C for 1 min, and 72°C for 1 min in 100 μl.

Isolation of poly(A)$^+$ RNA and Northern blot analysis were performed essentially as described (14).

cDNA Cloning. Cloned amphioxus SPC2 and SPC3 cDNAs were obtained using the RACE technique as described by Frohman *et al.* (15) and by screening a phage cDNA library prepared from whole amphioxus RNA (13). Other standard molecular biological techniques were as described by Sambrook *et al.* (16).

RESULTS AND DISCUSSION

To identify subtilisin-related proprotein convertase genes expressed in amphioxus we amplified whole amphioxus cDNA by using degenerate oligonucleotide primer sets which corresponded to highly conserved amino acid sequences in the catalytic domain of these proteases. When primer set SQ-3/QS-1 was used in PCR a 600-bp fragment was obtained, but no bands of the expected size were observed with the other primer combinations of SQ-3/QS-2, SQ-4/QS-1, or SQ-4/QS-2 (Fig. 1). However, restriction enzyme analysis showed that the 600-bp fragment was heterogeneous. For example, when the fragment was digested with *Bgl* I multiple bands in the range of 120–480 bp were observed (data not shown).

Table 1. Amino acid sequence identity between amphioxus and human (prepro)SPC2 and -SPC3

Domain	% sequence identity: Amphioxus SPC2 vs. human SPC2	% sequence identity: Amphioxus SPC3 vs. human SPC3
Pre	24	16
Pro	37	52
Catalytic	84	74
P	75	59
C terminus	61	21
Whole protein	71	55

The heterogeneous nature of the 600-bp fragment was confirmed by plasmid cloning into pGEM4Z. From individual clones we have obtained four distinct cDNA sequences, of which the deduced amino acid sequence from one clone showed the highest identity (80%) to human SPC2, while another clone exhibited the greatest identity (78%) to human SPC3. These sequences were expanded by using the rapid amplification of cDNA ends (RACE) technique (15) and by screening cDNA libraries (Fig. 2).

The complete coding sequences for amphioxus SPC2 and SPC3 contain 689 and 774 aa, respectively, whereas there are 638 and 753 aa in the human (prepro)SPC2 and -SPC3 sequences (Figs. 3 and 4). As summarized in Table 1, the amino acid sequence similarities were highest in the catalytic domains. In addition, amphioxus SPC2 contained an Asp-for-Asn substitution at position 317 (corresponding to residue 310 in human SPC2), as is the case in all other known SPC2 homologues (see below). In subtilisin the side chain of this Asn residue has been demonstrated to stabilize the transition state during catalysis (17). To form a similar hydrogen bond with the carbonyl oxygen next to the scissle bond, however, the Asp side chain in SPC2 would have to be protonated, and this is consistent with the probability that SPC2 in all species is active only in an acidic compartment such as the dense-core secretory granule.

A high percentage of identity for SPC2 and SPC3 is also found in the P domain, which consists of 190–192 aa located on the carboxyl side of the catalytic domain. Recently, transfection experiments have shown that when successively truncated forms of SPC1 (furin) were expressed, deletions within the P domain were correlated with loss of catalytic activity (18). Downstream from the P domain amphioxus SPC2 and

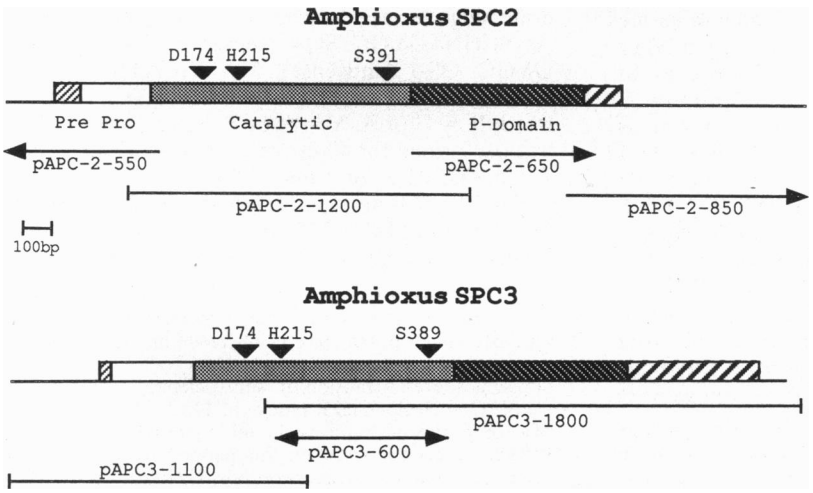

FIG. 2. Molecular cloning of amphioxus SPC2 and SPC3 cDNAs. The organization of amphioxus SPC2 and SPC3 cDNAs is shown schematically. Arrows indicate clones obtained by the RACE technique (15); brackets indicate clones isolated from a cDNA library.

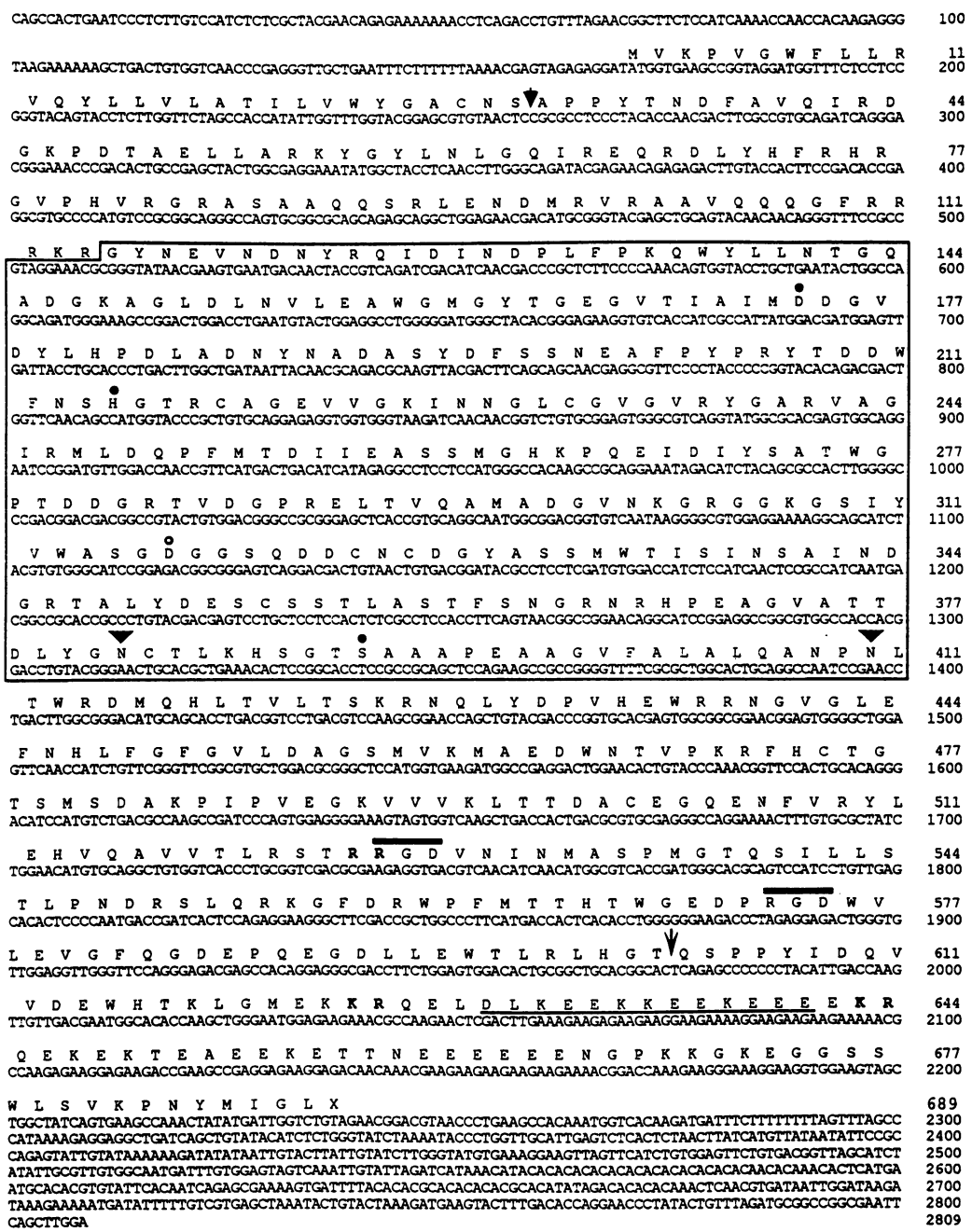

FIG. 3. Nucleotide and deduced amino acid sequence of amphioxus SPC2 cDNA. Putative cleavage site for the signal peptide is assigned to residue 30 (arrow). The subtilisin-related catalytic domain is boxed and dots indicate residues involved in the catalytic mechanism. Arrowheads identify two potential glycosylation sites, and two potential signals (RDG) for binding to extracellular matrix proteins are overlined. An arrow also identifies the end of the P domain at residue 602. A carboxyl-terminal glutamic acid-rich sequence, also found in human and rodent SPC2, is underlined.

SPC3 contained carboxyl-terminal sequences which are not highly conserved, although the carboxyl termini of both human and amphioxus SPC2 contained a region enriched in Glu residues which is predicted to form an amphipathic helix (1). Overall, amphioxus SPC2 exhibited 71% identity to human SPC2, whereas amphioxus SPC3 is 55% identical to human SPC3.

Very recently, SPC2 and SPC3 sequences have been reported from a number of phylogenetically diverse species, including frogs (19), fish (20), molluscs (21–23), and hydra (24). Fig. 5 shows the amino acid identities in the catalytic and P domains of SPC2 and SPC3 from these species compared to the human homologues and plotted against the estimated time of evolutionary divergence. The results clearly show that SPC2 has evolved more slowly than SPC3. It is not known whether this relatively slow substitution rate observed for SPC2 has a functional basis, but one possibility is that SPC2 may have a more ordered structure and/or restricted catalytic specificity.

Northern blot analysis revealed a single 2.7-kb band for amphioxus SPC3 mRNA, while two bands, a major band at 4.0 kb and a minor band at 7.0 kb, were observed for SPC2 mRNA (Fig. 6). Preliminary 3′RACE analysis indicate that the 7.0-kb SPC2 mRNA is due to extended transcription of the 3′ untranslated region (data not shown).

Although the SPC2 and SPC3 mRNAs could be easily detected on Northern blots, our attempts to perform *in situ* hybridization on thin (10- to 20-μm) saggital sections of whole mounted amphioxus were unsuccessful due to a high background. As an alternative, we have determined if there were

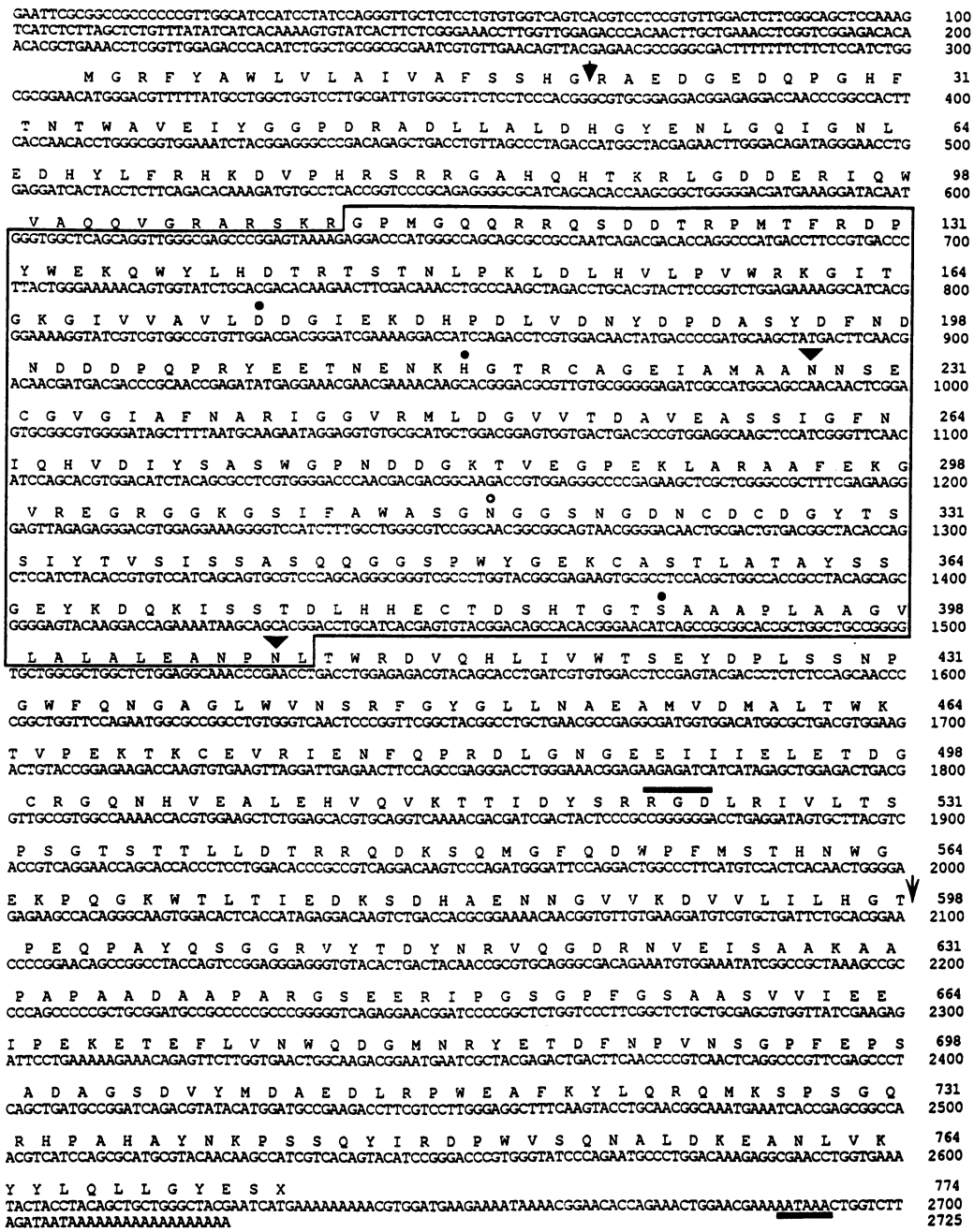

FIG. 4. Nucleotide and deduced amino acid sequence of amphioxus SPC3 cDNA. Putative cleavage site for the signal peptide is assigned to residue 19 (arrow). The subtilisin-related catalytic domain is boxed and dots indicate residues involved in the catalytic mechanism. Arrowheads identify two potential glycosylation sites and a potential signal (RDG) for binding to extracellular matrix proteins is overlined. An arrow also identifies the end of the P domain at residue 598.

any regional differences in the expression of SPC2 and SPC3. Whole amphioxus were sliced into equal-length sections and total RNA was extracted from each section. To obtain the ratio of SPC2 to SPC3 expression we performed multiplexed RT-PCR in which SPC2 and SPC3 DNA fragments were simultaneously amplified from the same RNA template. In calibration assays it was established that under the RT-PCR conditions, the amount of each DNA fragment obtained (as visualized by ethidium bromide staining) was roughly proportional to the amount of input RNA within the range of 20 to 2000 ng.

Fig. 7 shows the results of a representative experiment in which 100-ng amounts of amphioxus RNA were subjected to RT-PCR. With whole amphioxus RNA as template, equally intense ethidium bromide-stained SPC2 and SPC3 DNA fragments were obtained. This result indicated that whole amphioxus expressed approximately equal amounts of SPC2 and SPC3 mRNA. However, there were clear regional differences in the ratio of SPC2 to SPC3 expression. In the anterior (cranial) segment the SPC2-to-SPC3 ratio was 1:1, whereas in the posterior (caudal) segment the ratio was approximately 5:1 in favor of SPC2 (Fig. 7, lanes labeled 1 and 4). In contrast, the mid-body segments exhibited a reversed SPC2-to-SPC3 ratio of about 1:5. We also analyzed the expression of ILP mRNA in these segments by Northern blot and RT-PCR, and both analyses showed that ILP mRNA was predominately localized to segment 3 (data not shown).

The strong conservation of primary sequence between mammalian and amphioxus SPC2 and SPC3 strongly suggests

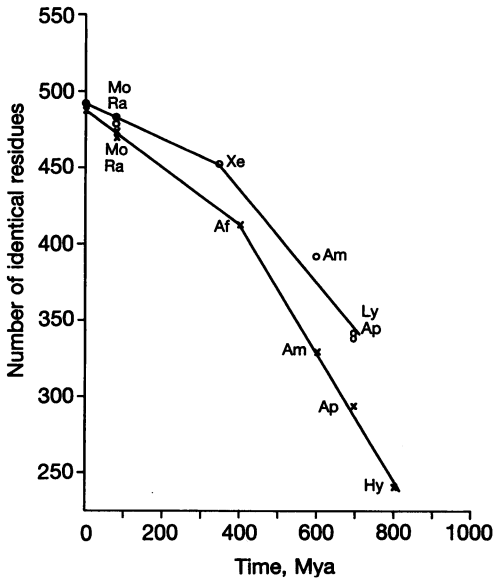

FIG. 5. Evolutionary divergence of SPC2 and SPC3 sequences. Identity of amino acid sequences in the catalytic and P domains of SPC2 (○) and SPC3 (×) from various species compared with human SPC2 and SPC3 is plotted against the estimated elapsed time (Mya, millions of years) since the species diverged (25, 26). Species are as follows: Mo, mouse (7, 8); Ra, rat (25); Xe, *Xenopus* (19); Af, anglerfish SPC3 (20); Am, amphioxus; Ly, *Lymnaea* (21); Ap, *Aplysia* (22, 23); and Hy, hydra SPC3-like (24).

that these enzymes are capable of cleaving the C-peptide domain from proILP as they do in processing proinsulin. However, it should be noted that we have not yet demonstrated that these enzymes and ILP are coexpressed in the same cell. Indeed, the broad regional distribution and variable ratio of SPC2 to SPC3 expression suggest that these enzymes probably act on additional substrates in amphioxus. Possible substrates include other pancreatic and/or neuroendocrine hormone precursors such as proglucagon, prosomatostatin, and pro-opiomelanocortin, although none of these hormones has yet been identified in amphioxus.

Whether SPC2 and SPC3 are necessary or sufficient to process proILP into a mature insulin-like peptide remains to be determined. An approach to the solution is to characterize the processing of proILP expressed in mammalian cells which contain SPC2 and SPC3. In particular, it is possible that the complete processing of proILP, particularly in the carboxyl-terminal extended sequence, will require an as-yet-unidentified proprotein convertase capable of cleaving at single basic residues. In our analyses of SPC DNA fragments amplified by PCR, we have observed that amphioxus expresses

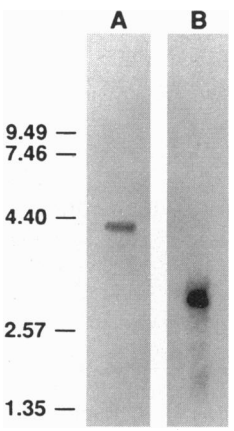

FIG. 6. Northern blot analysis of amphioxus SPC2 and SPC3 mRNAs. A 5-μg sample of amphioxus poly(A)$^+$ RNA was electrophoresed in a 1% agarose/0.66 M formaldehyde gel, blotted onto nitrocellulose, and hybridized with radiolabeled SPC2 cDNA (lane A) or SPC3 cDNA (lane B). RNA molecular weight markers (kb) were from BRL.

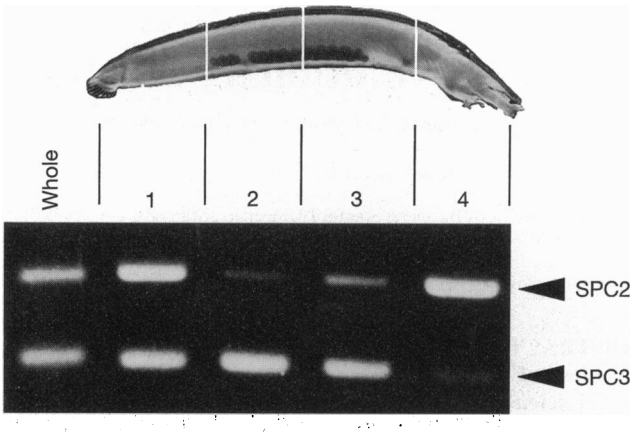

FIG. 7. Regional expression of SPC2 and SPC3 in amphioxus analyzed by RT-PCR. Whole amphioxus were sliced into four equal segments, as shown above (approximately life size); total RNA was isolated from each segment and subjected to RT-PCR amplification of SPC2 (480 bp) and SPC3 (334 bp) cDNA fragments. The reaction products were electrophoresed in a 1.5% agarose gel and stained with ethidium bromide.

at least two other subtilisin-related proprotein convertases in addition to SPC2 and SPC3.

We thank Sean Martin for assistance in preparing the figures and Florence Rozenfeld for assistance in preparation of the manuscript. This work was supported by National Institutes of Health Grants DK 13914 and DK20595 and by the Howard Hughes Medical Institute.

1. Steiner, D. F., Smeekens, S. P., Ohagi, S. & Chan, S. J. (1992) *J. Biol. Chem.* **267,** 23435–23438.
2. Smeekens, S. P. (1993) *Bio/Technology* **11,** 182–186.
3. Van de Ven, W. J. M., Roebroek, J. M. & Van Duijnhoven, H. L. P. (1993) *Crit. Rev. Oncogen.* **4,** 115–136.
4. Fuller, R. S., Sterne, R. E. & Thorner, J. (1988) *Annu. Rev. Physiol.* **50,** 345–362.
5. Roebroek, A. J. M., Schalken, J. A., Leunissen, J. A. M., Onnekink, C., Bloemers, H. P. J. & Van de Ven, W. J. M. (1986) *EMBO J.* **5,** 2197–2202.
6. Smeekens, S. P. & Steiner, D. F. (1990) *J. Biol. Chem.* **265,** 2997–3000.
7. Seidah, N. G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. & Chrétien, M. (1990) *DNA Cell Biol.* **9,** 415–424.
8. Smeekens, S. P., Avruch, A. S., LaMendola, J., Chan, S. J. & Steiner, D. F. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 340–344.
9. Seidah, N. G., Marcinkiewicz, M. & Benjannet, S. (1991) *Mol. Endocrinol.* **5,** 111–122.
10. Kiefer, M. C., Tucker, J. E., Joh, R., Landsberg, K. E., Saltman, D. & Barr, P. J. (1991) *DNA Cell Biol.* **10,** 757–769.
11. Nakayama, K., Kim, W. S., Torii, S., Hosaka, M., Nakagawa, T., Ikemizu, J., Baba, T. & Murakami, K. (1992) *J. Biol. Chem.* **267,** 5897–5900.
12. Smeekens, S. P., Albiges-Rizo, C., Carroll, R., Martin, S., Ohagi, S., Phillips, L. A., Benig, M., Gardner, P., Montag, A. G., Swift, H. H., Thomas, G. & Steiner, D. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8822–8826.
13. Chan, S. J., Cao, Q.-P. & Steiner, D. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9319–9323.
14. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York), pp. 143–146.
15. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
17. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. (1986) *Philos. Trans. R. Soc. London A* **317,** 415–423.
18. Hatsuzawa, K., Nagahama, M., Takahashi, S., Takada, K., Murakami, K. & Nakayama, K. (1992) *J. Biol. Chem.* **267,** 16094–16099.
19. Braks, J. A. M., Culdemond, K. C. W., van Riel, M. C. H. M. & Martens, G. J. M. (1992) *FEBS Lett.* **305,** 45–50.
20. Roth, W. W., Mackin, R. B. & Noe, B. D. (1993) *Endocrinol. J.* **1,** 131–140.
21. Smit, A. B., Spijker, S. & Geraerts, W. P. M. (1992) *FEBS Lett.* **312,** 213–218.
22. Ouimet, T., Mammarbachi, A., Cloutier, T., Seidah, N. G. & Castellucci, V. F. (1993) *FEBS Lett.* **330,** 343–346.
23. Chun, J. Y., Korner, J., Kreiner, T., Scheller, R. H. & Axel, R. (1994) *Neuron* **12,** 831–844.
24. Chan, S. J., Oliva, A. A., Jr., LaMendola, J., Grens, A., Bode, H. & Steiner, D. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6678–6682.
25. Colbert, E. H. & Morales, M. (1991) *Evolution of the Vertebrates: A History of the Backboned Animals Through Time* (Wiley/Liss, New York).
26. Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. & Raff, R. A. (1988) *Science* **239,** 748–753.