The FASEB Journal express article 10.1096/fj.01-0940fje. Published online May 8, 2002

Human proprotein convertase 2 homologue from a plant nematode: Cloning, characterization, and comparison with other species

Elena S. Kovaleva,* Alexander G. Yakovlev,† Edward P. Masler,* and David J. Chitwood*

*Nematology Laboratory, USDA, ARS/PSI, Beltsville, Maryland 20705; †Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007

Corresponding author: Elena S. Kovaleva, Nematology Laboratory, USDA, ARS/PSI, 10300 Baltimore Avenue, Bldg 011A, Beltsville, MD 20705-2350. E-mail: kovaleve@ba.ars.usda.gov

ABSTRACT

Proprotein convertases (PCs) are evolutionarily conserved enzymes responsible for processing the precursors of many bioactive peptides in mammals. The invertebrate homologues of PC2 play important roles during development that makes the enzyme a good target for practical applications in pest management. Screening of a plant nematode *Heterodera glycines* cDNA library resulted in isolation of a full-length clone encoding a PC2-like precursor. The deduced protein (74.2 kD) exhibits strong amino acid homology to all known PC2s, including human, and shares the main structural characteristics: signal peptide; prosegment; catalytic domain, with D/H/S catalytic triad, PC2-specific residues, and 7B2 binding sites; P domain (with RRGDT pentapeptide); and carboxyl terminus. Comparative analysis of PC2s from 15 species discloses the presence of an insert in the catalytic domain unique to nematodes. Expression of PC2-like mRNA found in eggs and juveniles was undetectable in adult stages of *H. glycines*. Nucleotide analysis reveals distinctive differences in base composition and codon usage between *H. glycines* and *Caenorhabditis elegans* PC2s. The *H. glycines* cDNA clone encoding PC2 is the first one isolated from plant-parasitic nematodes.

Key words: cDNA cloning • (G+C) content • plant-parasitic nematode • *Heterodera glycines*

The use of molecular biology techniques to study nematode genes important for neuroregulation of growth, molting, egg-laying, and hatching opens up a fascinating opportunity to manipulate the nematode life cycle. With respect to plant-parasitic nematodes, such as soybean cyst nematode *Heterodera glycines*, the major pathogen of soybean, numerous genes are involved in parasitism and therefore represent candidates for development of new strategies for controlling nematode-induced crop damage (1). Among these genes, proteinase genes have attracted particular attention (1, 2). Nearly all proteins regulating

physiological and developmental processes (hormones, neuropeptides, growth factors, other signal molecules) are synthesized as larger precursors subsequently activated by limited endoproteolysis. This catalytic function is often performed in part by proprotein convertases (PCs)--the kexin2/subtilisin (S8) family of serine proteinases. Mammalian PCs are grouped into seven types: PC1/3, PC2, furin/PACE, PC4, PACE4, PC5/6, and PC7/LPC/PC8 (3, 4, 5). Evolutionarily conserved homologues to mammalian PCs have been found in invertebrates. Among nematodes, available information on PCs is restricted to *Caenorhabditis elegans*, where four genes (*kpc-1*, *kpc-2*, *kpc-3*, and *kpc-4*) encoding kexin2/subtilisin PCs have been described (6). In addition, a *kpc-1* homologue (furin-like PC) has been identified in the animal-parasitic nematode *Dirofilaria immitis* (7) and a *kpc-4* homologue occurs in *Caenorhabditis briggsae* (8).

A complex family of PCs carries out common processing functions, and each enzyme has a specific biological role as a result of distinct substrate preference, subcellular localization, tissue distribution, and expression pattern. PC2 (E.C.3.4.21.94) is a convertase selectively expressed in neural and endocrine tissues and is responsible for proteolytic maturation of several hormones and neuropeptides in the regulated secretory pathway (9, 10). PC2 is unique among the convertases in that a neuroendocrine binding peptide 7B2 is required for its function. The expression of convertase PC2 and its specific binding peptide 7B2 are coinduced during neuronal differentiation, and enzyme function is spatially and temporally regulated by the dynamics of the interactions between enzyme and helper (4, 11).

Convertases responsible for the bioactivation of many prohormones and propeptides in invertebrates have important roles in ontogenesis. The PC2-like homologue in *C. elegans*, encoded by the *kpc-2* gene, is located mostly in neural cells, and the mutant phenotype has egg-laying deficiency and uncoordinated movement (6). A PC2-like homologue in *Drosophila melanogaster,* encoded by the *amontillado* gene, is expressed in early stages of embryogenesis and is crucial for hatching (12). In this study, we used informative nucleotide sequences from *C. elegans* and *D. melanogaster* PC2 homologues to screen a cDNA library from *H. glycines*, the soybean cyst nematode. We report here the isolation and characterization of mRNA encoding the PC2-like proteinase, the first convertase from a plant-parasitic nematode.

MATERIALS AND METHODS

Nematode cultures

H. glycines (NL1-RHp) were reared, collected, and staged as described previously (13). Four developmental stages of *H. glycines* were tested: eggs from yellow females, second-stage (J2) juveniles, white (young) females, and yellow (mature) females. A mixed population of *C. elegans* (strain N2) reared as previously described (14) was used as control. Total cellular RNA was isolated from homogenates of whole bodies of both species by the guanidinium thiocyanate method. After DNAse treatment, it was reverse transcribed with M-MLV RT (GIBCO, BRL**,** Gaithersburg, MD) and the resulting cDNA was used for polymer chain reaction (PCR) amplification.

Probe preparation and cDNA library screening

A *H. glycines* (J2 stage) cDNA library constructed in bacteriophage λ vector Uni-ZAP XR (Stratagene, La Jolla, CA) was provided by Dr. Eric Davis, North Carolina State University. This library was screened using a fragment of *H. glycines* PC2-like sequence obtained by PCR. The PCR primers were chosen on the basis of *C. elegans* and *D. melanogaster* PC2-like sequence alignment and corresponded to the conserved regions in the catalytic and P domains of the enzyme. The first pair consisted of 5'-GGNGTGGATTACATGCAT-3' (the sense primer) and 5'-GGCCCAAACAAAAATGGC-3' (the antisense primer); the second pair included 5'-TTTGTTTGGGCCAGTGG-3' (the sense primer) and 5'-AATGGCCANTTGGTGAATCC-3' (the antisense primer). Two PCR products of approximately the expected size (400 bp and 800 bp for the first and second pair of primers, respectivley) were obtained from *H. glycines* and *C. elegans* samples. The resulting PCR-amplified products for *H. glycines* cDNA were sequenced to confirm homology to known PC2-like sequences. These PCR products were then labeled with [α-^{32}P]dATP (3000 Ci/mmol, Amersham, Piscataway, NJ) by random priming (U.S. Biochemical, Cleveland, OH) according to the manufacturer's recommendations and used for screening of the cDNA library.

The first screening hybridization was performed at high stringency (50% formamide, 4 × SSPE at 42°C) and washing (three times for 2 min each at room temperature in 2 × SSC, 0.1% sodium dodecyl sulfate (SDS), two times for 30 min each at 65^{o}C in 1 × SSC, 0.1% SDS). Screening of ~2 × 10^6 λUni-ZAP XR plaques resulted in identification of six positive clones that were subjected to an additional round of plaque screening and isolation. Following in vivo excision of the pBluescript phagemid from the vector, using ExAssist/SOLR system (Stratagene), plasmid DNA was purified from isolated clones and sequenced.

PCR and sequencing

PCR was routinely performed in a buffer containing 10 mM Tris-HCl (pH 8.3), 5 mM $MgCl_2$, 0.2 mM deoxynucleotide mixture, 0.5 μM of amplification primers, and various amounts of template DNA. Cycling conditions were 30 cycles of 30 s at 94^{o}C, 15 s at 55^{o}C, and 1 min at 72^{o}C. PCR products were separated by electrophoresis in 1% agarose gels and visualized with ethidium bromide (0.5 ug/ml) in TBE electrophoretic buffer. Automated thermocycle sequencing was performed using AmpliTaq-FS kit (Perkin-Elmer, Boston, MA) and the upgraded Model 377 automated sequencer (Applied Biosystems, Foster City, CA).

Amplification of the 5'-end region

The 5′-end region of *H. glycines* PC2 cDNA was amplified from a cDNA library by using two rounds of nested PCR. According to predicted cDNA sequence orientation, M13REV and T3 sequencing oligonucleotides were used as sense PCR primers, and 5'-ATCCTGTGAGAGTTTGGC-3' and 5'-TTGTGTCCGACCGATCGA-3', derived from the informative sequence obtained from library screening, were used as antisense PCR primers. Two overlapping PCR products were purified and sequenced as described previously and the full-length *H. glycines* cDNA clone of 2.0 kb was then reconstructed (HglPC2).

Sequence analysis and comparison

For comparison with discovered HglPC2, we used sequence information available from the GenBank and literature about proconvertase genes of different species as well as fragments from expressed sequence tags (ESTs) database. To avoid redundant and nearly identical sequences, we used only one variant for each type and species. Codon usage tables and codon indices were calculated using the CononW program; for multiple alignments and phylogenetic tree construction, we used ClustalW (Baylor College of Medicine and European Bioinformatics Institute) and Phylip protdist programs (Neighbor-Joining/UPGMA method version 3.573c), made and distributed by J. Felsenstein (1995). Structural characteristics of the deduced protein were investigated using several freely available programs. Genome base composition parameters were taken from the Kazusa DNA Research Institute (KDRI) database (www.kazusa.org.jp).

RESULTS

Sequence analysis of the coding region demonstrated a single open reading frame of 2013 nucleotides that was predicted to encode a 671-amino acid residue protein with a MW of 74.2 kD (Fig. 1).

Expression of HglPC2 during different developmental stages

Primers used to amplify HglPC2 from a J2 *H. glycine*s cDNA library were used to detect HglPC2 during different developmental stages: eggs, J2 juveniles, young females (white), and mature females (yellow). The cDNA was prepared from total RNA for each stage of *H. glycine*s and for the mixed stages of *C. elegans*. Both pairs of PC2 primers provided positive PCR products from cDNA of immature stages (eggs or J2 juveniles), confirmed by sequencing. The same results were obtained for cDNA prepared from mixed stages of *C. elegans*, used as control. The amount of cDNA used in PCR amplification did not exceed 50 ng per incubation. The same PC2 primers did not yield any products from cDNA of adult stages (white or yellow females) of *H. glycine*s even for a 10-fold higher amount of cDNA. To check the quality of female cDNA, we performed simultaneous amplifications for two abundant proteins: actin and heat shock protein 70. Primers for these two control reactions were designed based on *H. glycines* mRNA sequences previously identified in our laboratory (unpublished data, accession numbers AF318603 and AF318605, respectively). The positive PCR products were obtained from cDNA of both mature stages of the *H. glycine*s for actin and for heat shock protein 70. The amount of cDNA required for amplification of these proteins was <20 ng per incubation. Thus, preliminary data showed that HglPC2 gene has a temporally restricted pattern of expression with much higher level of expression in early developmental stages (eggs and juveniles), whereas in mature stages, we were not able to detect HglPC2 using the PCR method.

Base composition and codon usage

The base composition of the HglPC2 coding region revealed an increased (G+C) content (56%) and even higher preference for G or C in the synonymous third codon position (62%). The relative amount of purines and pyrimidines in the silent position of HglPC2 was about equal, as were the ratios between A and T or C and G. The coding region of *C. elegans* PC2 in contrast, is AT-rich (GC = 47% and GC3s = 38%), and frequency of usage of C in the third position is higher than for G (0.315 and 0.155, respectively). The effective number of codons did not

indicate a strong overall bias for HglPC2 (N_c = 54), whereas for some amino acids, there were distinct favored codons. Comparison of codon usage tables for *H. glycines* and *C. elegans* PC2 depicted apparent differences in codon usage patterns for some amino acids. For example, *C. elegans* preferentially utilized CUU to code leucine, whereas *H. glycines* preferentially utilized CUG (relative synonymous codon usage RSCU = 2.3 and 2.1, respectively). Similar divergences were noted in the use of codons for valine, proline, threonine, aspartic acid, arginine, and glycine. Most differences are due to preferential usage of codons with U or A in the third position (*C. elegans*) or those with G or C (*H. glycines*) over other synonymous codons.

The diversity in base composition of PC2 in the absence of information about any other nematode full-length clones led us to investigate fragments of related sequences in other plant nematodes. The best matches found were a 639-bp-long fragment of cDNA from *Globodera rostochiensis* (AW506559), a member of the same subfamily as *H. glycines* (Heteroderinae) and a 525-bp-long fragment of cDNA from a more distant plant-parasitic nematode, *Meloidogyne incognita* (AW571345). Both fragments were isolated from the juvenile stage of nematodes and exhibited high homology to HglPC2 (84 and 78% amino acid identity, respectively). The *G. rostochiensis* fragment covered the main part of the catalytic domain of the enzyme, whereas the *M. incognita* fragment corresponded to the P domain and part of carboxyl terminus. Comparison of the nucleotide composition for both fragments and corresponding parts of HglPC2 revealed a strong resemblance between *G. rostochiensis* and *H. glycines* in (G+C) content (GC = 0.55 and 0.54, GC3s = 0.64 and 0.60, respectively). Meanwhile, the base composition of *M. incognita* differed from *H. glycines* markedly (GC = 0.40 and 0.58, GC3s = 0.19 and 0.65, respectively).

Structural analysis of putative HglPC2

Analysis of the amino acid sequence of HglPC2 (Fig. 1) suggested the presence of five domains: signal protein, prosegment, catalytic region (S8), P domain, and carboxyl terminus. The N-terminus contained a signal peptide, required for the transport of translated protein to the endoplasmic reticulum. The putative signal peptide of HglPC2 is 30 amino acid residues long, with the most probable cleavage site between LFA-LE; it is leucine-rich, having seven L's among the 30 residues. According to the SAPS, TMHMM and TMAP programs, the signal peptide of HglPC2 contained a hydrophobic transmembrane segment, with amino acid residues 7 through 29 most probably located within the membrane. The propeptide contained three consensus motifs for potential cleavage sites for the excision of the prosegment, which were found in positions KR-56, RTRR-86, and RLKR-114.

The catalytic domain (amino acid residues 135–488) had a classical triad of D/H/S active sites within the conserved patterns, and their positions are as follows: 177–188 (TAIMDDGVDYMH) for D; 222–232 (HGTRCAGEISA) for H, with thiol activation attributable to C-226; and 395–405 (GTSAAAPEAAG) for S, including the highly conserved G-395. HglPC2 exhibited a PC2-specific (but not other types of convertases) D-324 residue involved in oxyanion stabilization and a highly conserved G-325 immediately adjacent to the oxyanion hole. The tyrosine residue thought to be the primary recognition site for 7B2 neuropeptide within the catalytic domain was found at position Y-209, and the PC2-specific motif QP(Y/F)MTD(L/I) was located between the H and S active sites (QPYMTDL-263).

There are three putative N-glycosylation sites (N-174, N-297, and N-552), two of which are located within the catalytic domain, and a potential sulfation site (Y-186), typical for all PC2-like homologues. Another interesting region detected in HglPC2 is an ATP/GTP-binding site motif (P-loop) in position 284–291, GPTDDGKT. The catalytic domain of HglPC2 has a unique additional region (438–466), in which the pattern typical for the S8 domain is broken, and this "insert" divides the catalytic domain of HglPC2 into two subunits. The amino acid sequence of the insert was not similar to any known fingerprints or homologueical to any known proteins.

The P domain of HglPC2 (in position 500–640) contained a conserved RGD-motif observed within the middle of the P domain of most PCs (not only PC2s) (in position RGD-558), and the third putative N-glycosylation site (N-552). The carboxyl terminus of HglPC2 was 30 residues long, with a putative amphipathic α-helix in positions 655–669 preceded by charged residues, and contained neither a transmembrane region nor a histidine-rich area.

Comparison of HglPC2 with PC2s of other species

The structures of the predicted HglPC2 molecule and other PC2s are shown schematically in Figure 2. Distinct PC2-like enzymes are described from 14 eukaryotes, 5 of which are closely related mammals: For simplicity, the depicted comparison is with just one (*Homo sapiens*). The length, main proportions, and order of components are common for all PC2s. Moreover, each component of HglPC2 shares the essential features with some of the other members, but the combination of them is unique for HglPC2. Only LcuPC2 and BcaPC2 have a transmembrane segment within the signal peptide, similar to HglPC2. The N-glycosylation sites are similar for nematodes and arthropods, and the presence of the low-complexity region near the PC2-specific D amino acid residue is typical for nematodes, insects, and mollusks. Multiple alignment shows that the last two consensus motifs for potential cleavage sites for the excision of the prosegment are in similar positions for all members, although the first motif occurs at the same location as in HglPC2 (KR-56), only in CelPC2, OliPC2, and BcaPC2 (data not shown). The tyrosine residue, responsible for recognition of 7B2 (Y-209), is substituted by F in the PC2s of *C. elegans* and arthropods. The RRGDT pentapeptide sequence is identical only to CelPC2 and OliPC2, whereas other members have RRGDL and RRGDV variants, and in mollusk PC2s, the canonical integrin binding RGD motif is absent.

It is of interest that CelPC2, as well as the hypothetical protein C51E3.7B (identical to CelPC2 in the catalytic domain), has an insert similar to HglPC2 within the catalytic domain at position 431–456. The length and placement of these inserts are very similar, whereas the specific sequences for *C. elegans* and *H. glycines* differ. Figure 2 illustrates that this two-subunit S8 domain is found only in nematodes thus far. To investigate this hypothesis further, we analyzed nematode sequence fragments available from the EST database. The sequences of three animal-parasitic nematodes were identified as PC2-like fragments located within the catalytic domain. They were *Strongyloides stercoralis* (BG225849), *Parastrongyloides trichosuri* (BI743113), and *Ancylostoma caninum* (BI744250). All fragments have open reading frames, and to simplify the comparison, we converted their nucleotide sequences to amino acids.

Figure 3 presents a comparison of the predicted amino acid sequences of PC2s from the species discussed previously, and the three additional animal-parasitic nematodes, in the areas of

interest. Unfortunately, the fragment from *P. trichosuri* stopped at the beginning of the insert, but the available sequence in this region completely coincides with HglPC2. We marked the residues (shaded D and W) as boundary residues separating two areas with strong homology among all species (to the left of D and to the right of W) from the less conserved part (between D and W). This part is 4 amino acids long for vertebrates, 6 amino acids long for arthropods and mollusks, and 26–29 amino acids long for nematodes. The first 4 and the last 3 amino acid residues of the insert are identical for all nematodes, whereas the middle part is difficult to align. In general, the insert is proline/arginine-rich and leucine-rich, but it does not correspond to PRELP repeat or leucine zipper protein fingerprints. The BLAST program failed to reveal homology of the HglPC2 insert to any known proteins.

The phylogenetic analysis of PC2s

The amino acid sequence homology between HglPC2 and all known PC2s was estimated for the entire coding region and separately for the most conservative S8 domain and P domain (Table 1). The phylogenetic relatedness of *H. glycines* PC2 with those of related enzymes from various species was determined using a protein distance matrix, followed by the Neighbor-Joining/UPGMA method (Fig. 4). The similarity of HglPC2 to CelPC2 (75%) is distinctly higher than to any other nematode PC not of the PC2 type. For example, amino acid identity between HglPC2 and furin-like enzymes from *C. elegans* and *D. immitis* was 41 and 33%, respectively, which is lower than identity between HglPC2 and PC2-like enzymes from mammals (55%). These data support the theory that PC2 diverged evolutionarily from other members of the PC family well before the emergence of the nematode phylum.

The previously described differences in codon biases between nematode PC2s forced us to investigate nucleotide homology among species. The nucleotide relatedness of HglPC2 to CelPC2, DmePC2, and HsaPC2 (first, second, and last place in the order of amino acid homology) together with the *G. rostochiensis* fragment was estimated for 639 nucleotides within the catalytic domain. The highest homology was found for *G. rostochiensis* (27.4%), followed by DmePC2 (24.1%); CelPC2 and HsaPC2 were essentially equal in nucleotide homology (23.4 and 23.3%, respectively).

DISCUSSION

Base composition of genes and genomes of nematodes

Information on plant-parasitic nematode genes is very limited even compared with animal-parasitic species, let alone compared with the wealth of information on the model organism *C. elegans*, which has a completely sequenced genome. For instance, the KDRI database contains information on only 19 full-length coding sequences (CDSs) for *H. glycines*, 12 CDSs for *G. rostochiensis*, and 10 CDSs for *M. incognita*, the main nematode pests worldwide. This lack of information makes it inevitable that sequences from free-living and animal-parasitic nematodes are usually used to design molecular probes for identifying genes from plant nematodes. Analysis of the base composition of genes is very important for probe design in cases in which sequences from closely related species are not known. The differences between GC content in the model nematode *C. elegans* (or many other nematodes) and those from *H. glycines* are great

enough that data derived from *C. elegans* sequences are often not effective in primer design or predicting codon usage pattern. In our study, the sequences from GC-rich *D. melanogaster* PC2 and even human PC2 were very helpful in minimizing the degeneracy of PCR primers used for screening the *H. glycines* library.

Most nematode genes (with some exceptions) are AT-rich. This was shown for genomes of several free-living (15, 16) and animal-parasitic nematodes (17) and for a limited number of plant nematodes genes, primarily from *Meloidogyne* spp. (18). Our analysis of EST data showed the heterogeneity of plant nematodes in terms of base composition. Thus, on one hand, the PC2 of *H. glycines* and *G. rostochiensis* (Heteroderinae sub-family members) were GC-rich and, on the other hand, another plant nematode *M. incognita* (Meloidoderinae subfamily) was AT-rich in the same way as *C. elegans*. Overall, the trends observed for all other genes for these subfamilies of plant nematodes agree with the differences in base composition of PC2. Thus, GC3s for representatives of the Heteroderinae subfamily *H. glycines* and *H. schachtii* (9062 and 697 codons in the KDRI database, respectively) are 59.2 and 60.1%, respectively, and for *G. rostochiensis* and *G. pallida* (2973 and 1086 codons, respectively), 62.0 and 61.3%, respectively. Meanwhile, GC3s for representatives of the Meloidoderinae subfamily, *M. incognita* and *M. javanica* (3679 and 916 codons, respectively) are 27.6 and 26.8%, respectively. The difference in GC content between these two subfamilies of plant nematodes is too great to be ignored just because of the small numbers of codons available for estimation. Comparison of the base composition in PC2 with that of available genomes from various species revealed a slight bias of PC2: GC content in PC2 is slightly higher than genomic GC content in GC-rich species (e.g., GC3s = 0.711 for PC2 and 0.646 for the genome of *D. melanogaster*), and slightly lower in AT-rich species (e.g., GC3 = 0.288 for PC2 and 0.353 for the genome of *L. cuprina*). It is of interest that codon indices and codon frequencies for some amino acids are very similar for HglPC2 and human PC2, and GC content is identical up to the third digit (GC = 0.556).

To summarize the base composition analysis, we can conclude that only some plant nematodes, but not those from the Heteroderinae, corroborate the typical for nematode AT richness. This is in agreement with the recently proposed molecular evolutionary framework for the phylum Nematoda (19, 20), which suggested that plant-parasitic nematodes arose independently several times and that relationships between free-living, animal, and plant nematodes are complex. Application of base composition phylogenetics in addition to amino acid homology of proteins may help to clarify some of the inconsistencies and complexities in current nematode classification.

Developmental genes, the example of which is PC2, play especially important roles, and expression of it is often associated with specific developmental events. In our work, the clone encoding PC2 was isolated from a juvenile-stage library of *H. glycines*. Primers directly designed from the HglPC2 sequence were then used in PCR amplification of DNA samples from different stages, and PC2 mRNA was detected in eggs and larva, but not in adults. Consequently, we can suggest that HglPC2 expression level varies greatly during the development. This temporally restricted pattern of expression, required for differential processing of neuoropeptides that regulate switching between developmental stages, was demonstrated in invertebrates. In *D. melanogaster*, the *amontillado* gene peaks during late embryogenesis and the PC2-like proteinase regulates hatching (12); in *C. elegans*, PC2 is involved in egg laying through the

timed expression of specific neuropeptides (6, 24). In addition, all PC2-like fragments reported for plant and animal parasitic nematodes were obtained from larval stages, but not from adults. The data on plant nematode developmental genes contribute to genetic studies of nematode evolution. A comparative approach to the developmental genetics of nematodes should facilitate our understanding of how development evolves (21), which is important for basic and applied science.

Structural comparison of putative HglPC2 with other PC2s

Soybean cyst nematode PC2 is a new member of an evolutionarily conserved kex2/subtilisin-like PC family of enzymes with structural homology conserved from nematodes through mammals. The PC2 type (E.C.3.4.21.94), expressed selectively in neural and endocrine tissues of mammals, is a key enzyme for proteolytic maturation of several proteins in the regulated secretory pathway, generally by hydrolysis of Lys-Arg and Arg-Arg bonds. The involvement of PC2 in distinct metabolic pathways varies among species, whereas the main mechanisms of function are very similar. PC2 involvement in the processing of pro-glucagon, pro-insulin, and pro-IAPP in mammals is important for understanding the mechanisms of obesity and diabetes (4, 9, 22); recently, homologues of the diabetic autoantigens IA-2 and phogrin were discovered in *C. elegans* neurons (23). In invertebrates, PC2 and its potential substrates (neuropeptides and neurohormones) play important roles in regulating, for example, early embryogenesis, egg laying, hatching, and molting (24–31). These findings suggest that the PC2 of *H. glycines* qualifies as a good target enzyme for the development of novel antinematode compounds. HglPC2 sequence data should facilitate comprehensive studies on the direct role of the enzyme in nematode development and pathogenicity and should have implications in the search for other proteins (and their inhibitors) as molecular targets.

The architectural structure of HglPC2 has both typical and unique properties. The length of the catalytic S8 domain varies very little among all species (<1%); the distances between potential cleavage sites for the excision of the prosegment and active catalytic sites were determined very strictly, and their variations did not exceed 3%. The PC2-specific residue, D-324 in HglPC2, is characteristic of PC2s, whereas other kex2/subtilisin family members contain N. It is suggested that D substitution is related to the unique ability of PC2 to bind the chaperone neuropeptide 7B2 essential for proper PC2 functions (5, 32, 33). Site-directed mutagenesis showed that the tyrosine residue in mammals (Y-209 for HglPC2) might represent the primary recognition site for the 7B2 neuropeptide within the catalytic domain (32). The PC2s of *C. elegans* and arthropods have F-substitution in that position. The PC2-specific motif QP(Y/F)MTD(L/I), located in the catalytic domain between histidine and serine active sites (257–263 for HglPC2), contributes greatly to binding of the active 7B2 and is critical for PC2 inhibition by the 7B2 C-terminal peptide (34). The amino acid patterns for active catalytic sites were almost identical, and the ATP/GTP-binding site motif was 100% identical for all PC2s.

The P domain is conserved in the kex2/subtilisin family, and mutations in this region lead to failure of the catalytic centers to work properly. It is involved in rapid intramolecular cleavage of the prosegment of the enzyme in the endoplasmic reticulum (35) and also plays an important role in regulating the stability, as well as calcium and pH dependence of PCs (36). The percent identity of HglPC2 for P domain (Table 1) is almost the same for all species (48–56%) but *C.*

elegans (70%), unlike that for the catalytic domain, where similarity is much higher to nematode and arthropods than to other species. The HglPC2 homology to P domain of mollusks (48–49%) is lower than that of mammals (50–52%), which partially might be explained by the absence of the canonical RGD motif in mollusks. The carboxyl-terminal segment is the least conservative part of all PC2s; its length is maximal in BcaPC2.

Against the background of structural similarity of PC2s from different species, the presence of an insert in a nematode catalytic domain is striking. The comparative analysis of the area of interest for nematode PCs, other than the PC2 types, revealed no such insert in any enzyme from either free-living or parasitic nematodes. Thus, this feature appears to be not only nematode phylum-specific, but also unique to type 2 convertase. Although no homology to any known protein was found for the amino acid sequence of this region, EST database analysis revealed three interesting nucleotide homologies. The first is a 22-nucleotide-long region (1246–1267 for HglPC2) with homology to mRNA extracted from preparasitic J2s of *G. rostochiensis* (AW506055) with unknown function. The second is a region of a cDNA clone from tomato nutrient-deficient/stressed roots (BF097860) with 18 identical nucleotides (1232–1249 for HglPC2) and an adjoining region of similarity up to about 50 nucleotides. The third interesting relationship was found with four overlapping and almost identical clones from the soybean plant *Glycine max*, the main host of *H. glycines* (accession number for the longest clone, AW349498).

Eighteen nucleotides (1243–1260 for HglPC2) are completely identical to *G. max* sequences. For all four plant clones, the authors cite homology of the sequences to homeobox-leucine zipper protein. The proline/arginine- and leucine-richness of the HglPC2 insert is characteristic for the area beyond 18 identical nucleotides, further corroborating the homology. Even in the absence of overall similarity of genes from nematode and host plants, this observation is interesting because the family of plant proteins with a leucine zipper, nucleotide binding site, and leucine-rich repeat regions includes proteins required for resistance against viruses, bacteria, fungi, and nematodes (37). Although beyond the scope of this study, the observations are of interest in light of parasite-host relashionships and horizontal gene transfer (37, 38).

ACKNOWLEDGMENTS

We thank Sandra Sardanelli, University of Maryland, for generosity and expertise in growing soybean root-nematode cultures.

REFERENCES

1. Lilley, C. J., Urwin, P. E., and Atkinson, H. J. (1999) Characterization of plant nematode genes: identifying targets for a transgenic defence. *Parasitol*ogy **118**, S63–S72

2. Lilley, C. J., Devlin, P., Urwin, P. E., and Atkinson, H. J. (1999) Parasitic nematodes, proteinases and transgenic plants. *Parasitol. Today* **15**, 414–417

3. Rehemtulla, A., and Kaufman, R. J. (1992) Protein processing within the secretory pathway. *Curr. Opin. Biotechnol.* **3**, 560–565

4. Bergeron, F., Leduc, R., and Day, R. (2000) Subtilase-like pro-protein convertases: from molecular specificity to therapeutic applications. *J. Mol. Endocrinol.* **24**, 1–22

5. Seidah, N. G., and Chretien, M. (1999) Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res*. **848**, 45–62

6. Thacker, C., and Rose, A. M. (2000) A look at the *Caenorhabditis elegans* Kex2/Subtilisin-like proprotein convertase family. *BioEssays* **22**, 545–553

7. Jin, J., Poole, C.B., Slatko, B. E., and McReynolds, L. A. (1999) Alternative splicing creates sex-specific transcripts and truncated forms of the furin protease in the parasite *Dirofilaria immitis*. *Gene* **237**, 161–175

8. Thacker, C., Marra, M. A., Jones, A., Baillie, D. L., and Rose A. M. (1999) Functional genomics in *Caenorhabditis elegans*: an approach involving comparisons of sequences from related nematodes. *Genome Res.* **9**, 348–359

9. Rouille, Y., Duguay, S. J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A. A. Jr., Chan, S. J., and Steiner, D. F. (1995) Proteolytic processing mechanisms in the biosyntheses of neuroendocrine peptides: the subtilisin-like proprotein convertases. *Front. Neuroendocrinol.* **16**, 322–361

10. Creemers, J. W. M., Jackson, R. S., and Hutton, J. C. (1998) Molecular and cellular regulation of prohormone processing. *Semin. Cell Devel. Biol.* **9**, 3–10

11. Mbikay, M., Seidah, N. G., and Chretien, M. (2001) Neuroendocrine secretory protein 7B2: structure, expression and functions. *Biochem. J.* **357**, 329–342

12. Siekhaus, D. E., and Fuller, R. S. (1999) A role for amontillado, the *Drosophila* homologue of the neuropeptide precursor processing protease PC2, in triggering hatching behavior. *J. Neurosci*. **19**, 6942–6954

13. Masler, E. P., Kovaleva, E. S., and Sardanelli S. (1999) FMRFamide-like immunoactivity in *Heterodera glycines* (Nemata:Tylenchida). *J. Nematol*. **31**, 224–231

14. Chitwood, D. J., Lusby, W. R., Thompson, M. J., Kochansky, J. P., and Howarth, O. W. (1995) The glycosylceramides of the nematode *Caenorabditis elegans* contain an unusual, branched-chain sphingoid base. *Lipids* **30**, 567–573

15. Stenico, M., Lloyd, A. T., Sharp, P. M. (1994) Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res.* **22**, 2437–2446

16. Waterston, R. H., Sulston, J. E. and Coulson, A. R. (1997) The genome. In *C. elegans II* (Riddle, D. L., Blumenthal, T., Meyer, B. J., and Priess, J. R., eds) pp. 23–45, Cold Spring Harbor Laboratory Press, Plainview, NY

17. Fadiel, A., Lithwick, S., Wanas, M. Q., and Cuticchia, A. J. (2001) Influence of intercodon and base frequencies on codon usage in filarial parasites. Genomics **74**, 197–210

18. Hammond, M. P. and Bianco, A. E. (1992) Genes and genomes of parasitic nematodes. *Parasitol. Today* **8**, 299–305

19. Davis, E. L., Hussey, R. S., Baum, T. J., Bakker, J., Schots, A., Rosso, M. N., and Abad, P. (2000) Nematode parasitism genes. *Annu. Rev. Phytopathol.*, **38**, 365–396

20. Blaxter, M. L., De Ley, P., Garey, J.R., Liu, L. X , Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., and Thomas, W. K. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature*, **392**, 71–75

21. Goldstein, B. (2001) On the evolution of early development in the Nematoda. *Philos. Tran.s R. So.c Lond. B. Biol. Sci.* **356**,1521–1531

22. Wang ,J., Xu, J., Finnerty, J., Furuta, M., Steiner, D. F., and Verchere, C. B. (2001) The prohormone convertase enzyme 2 (PC2) is essential for processing pro-islet amyloid polypeptide at the NH2-terminal cleavage site. *Diabetes* **50**(3), 534–539

23. Zahn, T. R., Macmorris, M. A., Dong, W., Day, R., and Hutton, J. C. (2001) IDA-1, a *Caenorhabditis elegans* homologue of the diabetic autoantigens IA-2 and phogrin, is expressed in peptidergic neurons in the worm. *J. Comp. Neurol.* **429**, 127–143

24. Waggoner, L. E., Hardaker, L. A., Golik, S., and Schafer, W. R. (2000) Effect of a neuropeptide gene on behavioral states in *Caenorhabditis elegans* egg-laying. *Genetics* **154**, 1181–1192

25. Ishizaki, H., and Suzuki, A. (1994) The brain secretory peptides that control moulting and metamorphosis of the silkmoth*, Bombyx mori*. *Int. J. Dev. Biol.* **38**, 301–310

26. Klumperman, J., Spijker, S., van Minnen, J., Sharp-Baker, H., Smit, A. B., and Geraerts, W. P. M. (1996) Cell type-specific sorting of neuropeptides: a mechanism to modulate peptide composition of large dense-core vesicles. *J. Neurosci.* **16**, 7930–7940

27. Lacombe, C., Greve, P., Martin, G. (1999) Overview on the sub-grouping of the crustacean hyperglycemic hormone family. *Neuropeptides* **33**, 71–80

28. Li, C., Nelson, L. S., Kim, K., Nathoo, A., and Hart, A. C. (1999) Neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Ann. N.Y. Acad. Sci.* **897**, 239–252

29. Pryor, S. C., and Elizee, R. (2000) Evidence of opiates and opioid neuropeptides and their immune effects in parasitic invertebrates representing three different phyla: *Schistosoma mansoni*, *Theromyzon tessulatum*, *Trichinella spiralis*. *Acta Biol. Hung.* **51**, 331–341

30. Sithigorngul, P., Cowden, C., Stretton, A. O. (1996) Heterogeneity of cholecystokinin/gastrin-like immunoreactivity in the nervous system of the nematode *Ascaris suum*. *J. Comp. Neurol.* **370**, 427–442

31. Kimber, M. J., Fleming, C. C., Bjourson A. J., Halton, D. W., Maule, A. G. (2001) FMRFamide-related peptides in potato cyst nematodes. *Mol. Biochem. Parasitol.*, **116**, 199–208

32. Benjannet, S., Mamarmachi, A. M., Hamelin, J., Savaria, D., Munzer, J. S., Chretien, M., and Seidah, N. G. (1998) Residues unique to the pro-hormone convertase PC2 modulate its autoactivation, binding to 7B2 and enzymatic activity. *FEBS Lett*. **428**, 37–42

33. Muller, L., and Lindberg, I. (1999) The cell biology of the prohormone convertases PC1 and PC2. *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 69–108

34. Apletalina, E. V., Muller, L., and Lindberg, I. (2000) Mutations in the catalytic domain of prohormone convertase 2 result in decreased binding to 7B2 and loss of inhibition with 7B2 C-terminal peptide. *J. Biol. Chem.* **275**, 14667–14677

35. Glushankof, P., and Fuller, R. S. (1994) A C-terminal domain conserved in precursor processing proteases is required for intramolecular N-terminal maturation of pro-Kex2 protease. *EMBO J.* **13**, 2280–2288

36. Zhou, A., Martin, S., Lipkind, G., LaMendola, J., and Steiner, D. F. (1998) Regulatory roles of the P domain of the subtilisin-like prohormone convertases. *J. Biol. Chem.* **273**, 11107–11114

37. Milligan, S. B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., Williamson, V. M. (1998) The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* **10**, 1307–1319

38. Bird, D. M., and Koltai, H. (2000) Plant parasitic nematodes: habitats, hormones, and horizontally-acquired genes. *J.Plant Growth Regulation* **19**, 183–194

Received December 20, 2001; revised March 18, 2002.

Table 1

Amino acid sequence identity (%) between HglPC2 and other PC2s[a]

Genbank Accession No.	Abbreviation	Species	Total	S8 domain	P domain
AAA56868	Cel	*Caenorhabditis elegans*	75	88	70
AAF56615	Dme	*Drosophila melanogaster*	58	85	55
CAA70107	Lcu	*Lucilia cuprina*	59	85	55
AAK28328	Oli	*Orconectes limosus*	63	83	56
AAA03496	Aca	*Aplysia californica*	55	73	49
CAA48730	Lst	*Lymnaea stagnalis*	55	74	48
AAA87005	Bca	*Branchiostoma californiensis*	50	72	49
AAC62503	Rri	*Rana ridibunda*	55	75	53
CAA47118	Xla	*Xenopus laevis*	56	75	55
AAA60032	Hsa	*Homo sapiens*	55	74	52
AAA39376	Mmu	*Mus musculus*	55	74	51
AAA40946	Rno	*Rattus norvegicus*	55	74	51
CAA48593	Ssc	*Sus scrofa*	55	74	50
AAG17018	Bta	*Bos taurus*	55	74	52
AAB63525	KPC-1	*Caenorhabditis elegans*	--	48	--
CAB05501	KPC-3	*Caenorhabditis elegans*	--	32	--
AAB96753	KPC-4	*Caenorhabditis elegans*	--	41	--

[a]Similarity was estimated for catalytic S8 domain and P domain separately in addition to similarity of the whole molecule.

Fig. 1

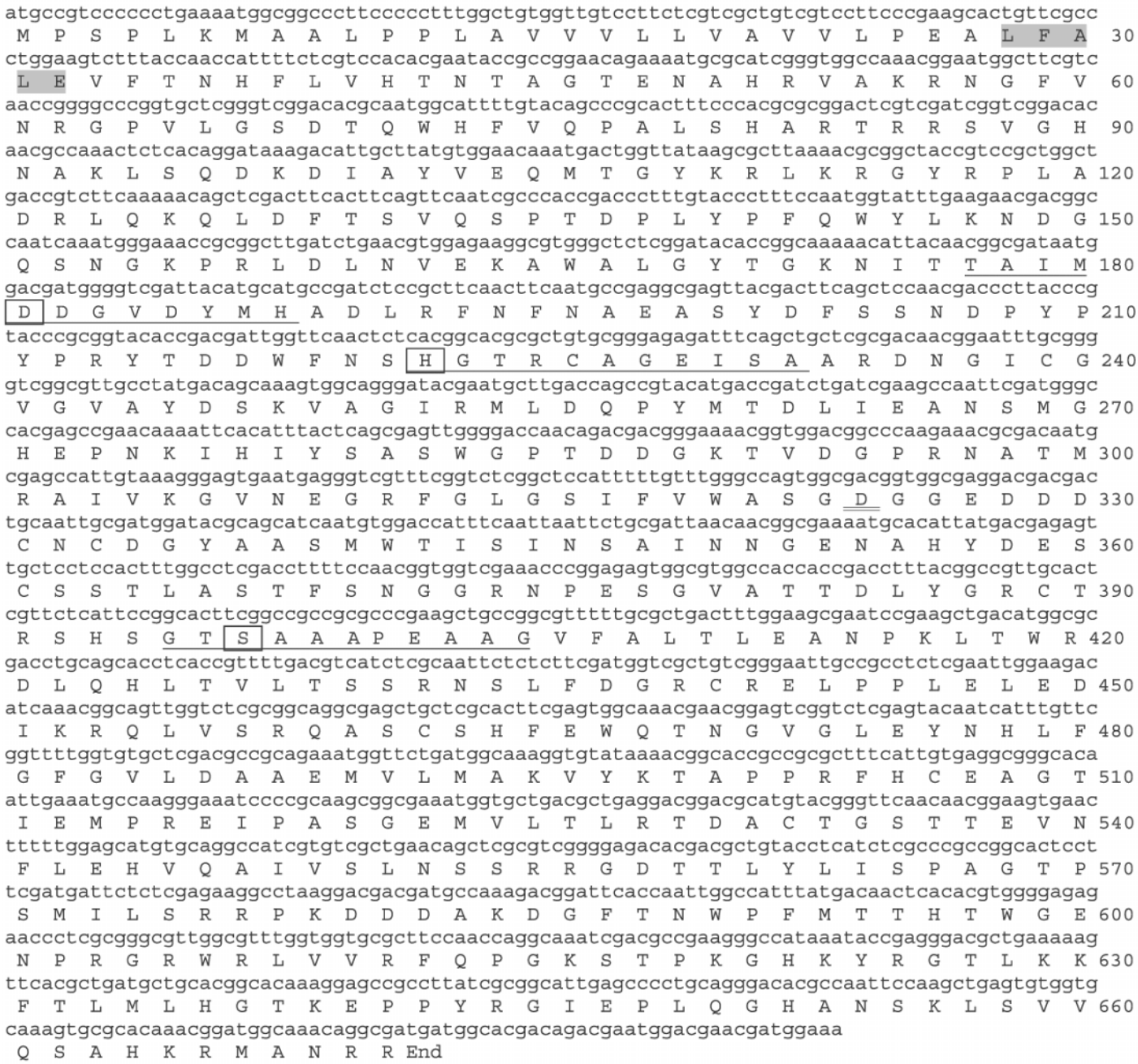

Figure 1. Nucleotide and deduced amino acid sequence of *Heterodera glycines* PC2. Amino acid numbers are indicated at the end of the line. The catalytic active patterns are underlined, and active residues are boxed. PC2-specific D residue (oxyanion hole) is double-underlined. Putative cleavage site of the signal peptide is shaded. The sequence data have been deposited in GenBank database under Accession No. AF386073.

Fig. 2

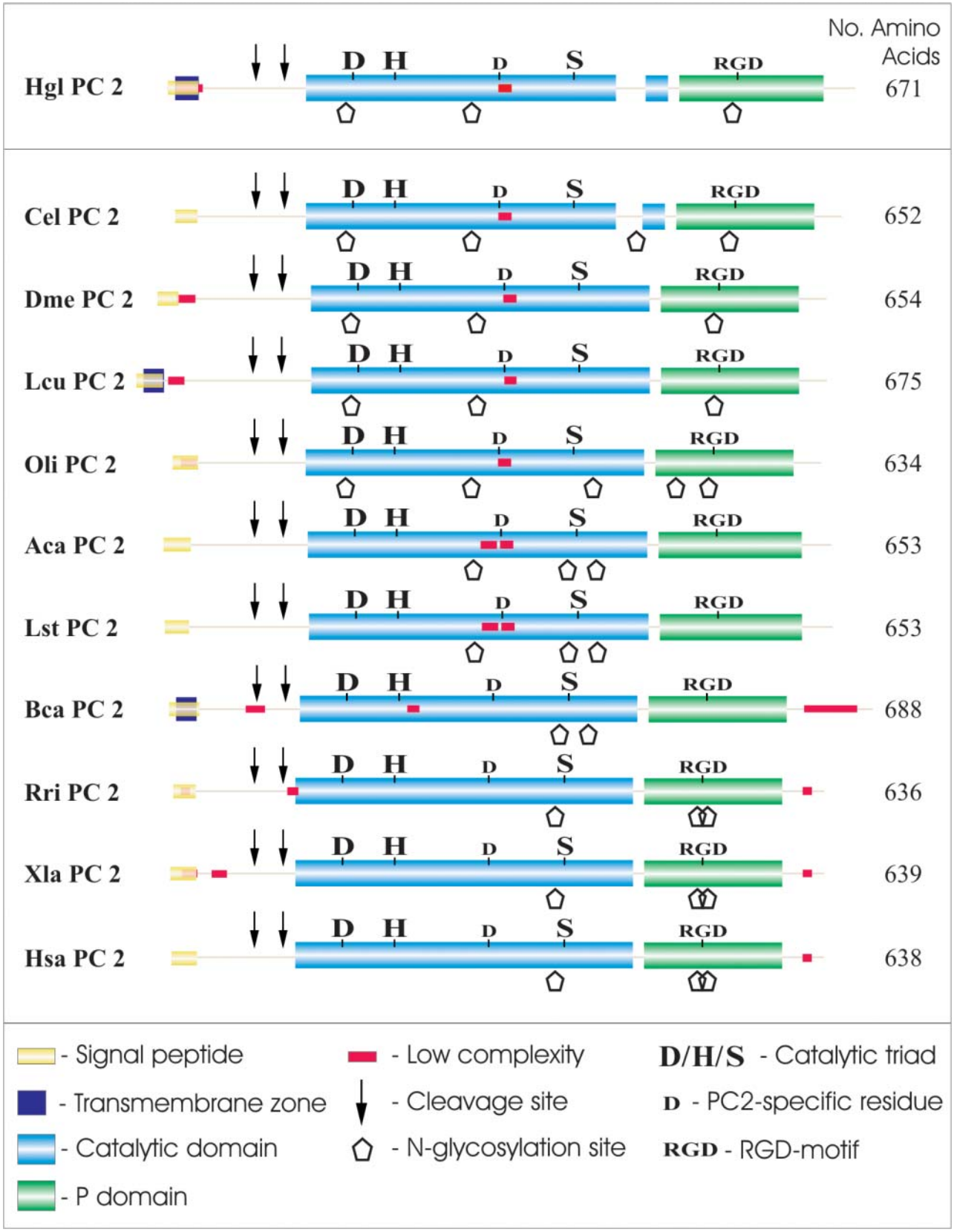

Figure 2. Diagrammatic structural comparison of the predicted HglPC2 and other PC2-like molecules. *Homo sapiens* PC2 was chosen from five known mammalian PC2s (see **Table 1** for the complete list of PC2s and species abbreviations).

Fig. 3

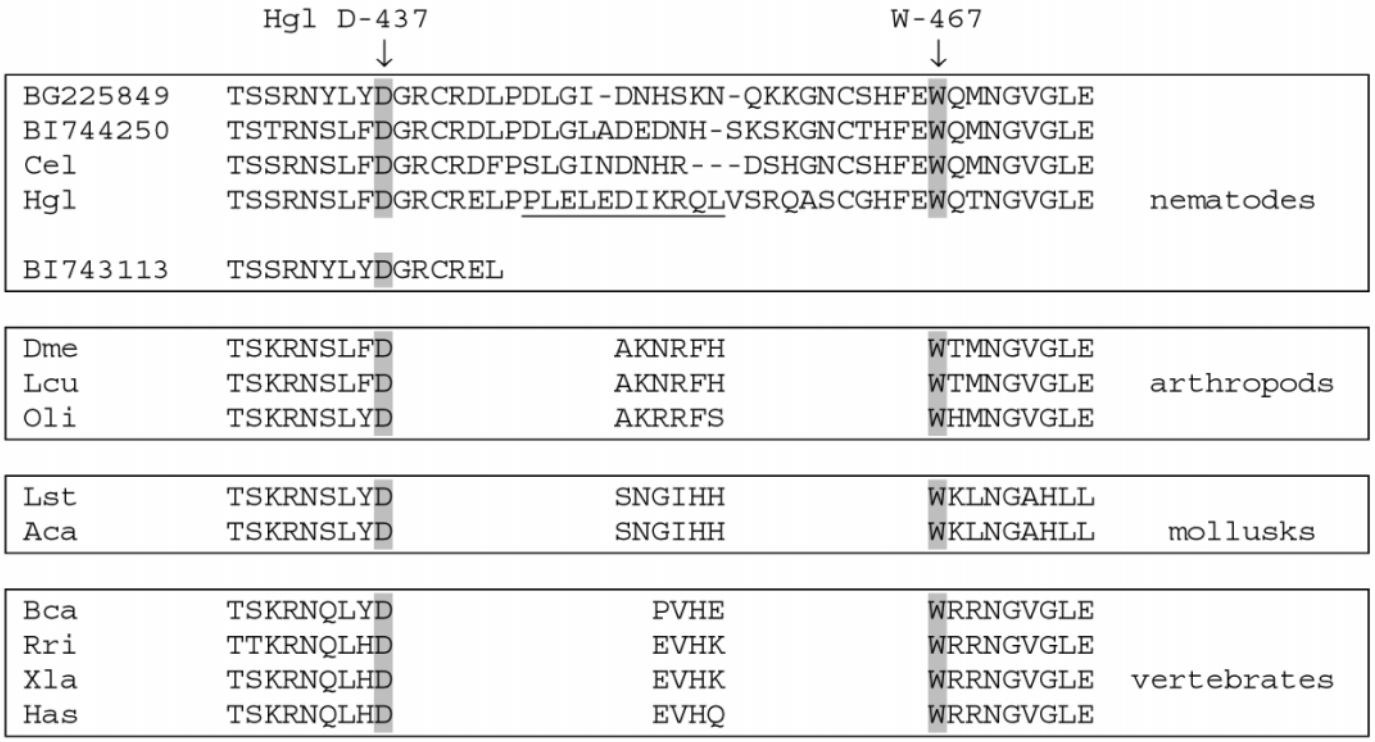

Figure 3. The insert in the catalytic domain (S8) of PC2. The shaded D (D-437 for HglPC2) and W residues (W-467 for HglPC2) are identical for all aligned sequences; amino acid residues between them are not conserved. Similarity of underlined residues in HglPC2 sequence is discussed in Discussion. BG225849, *Strongyloides stercoralis*; BI744250, *Ancylostoma caninum*; BI743113, *Parastrongyloides trichosuri*.

Fig. 4

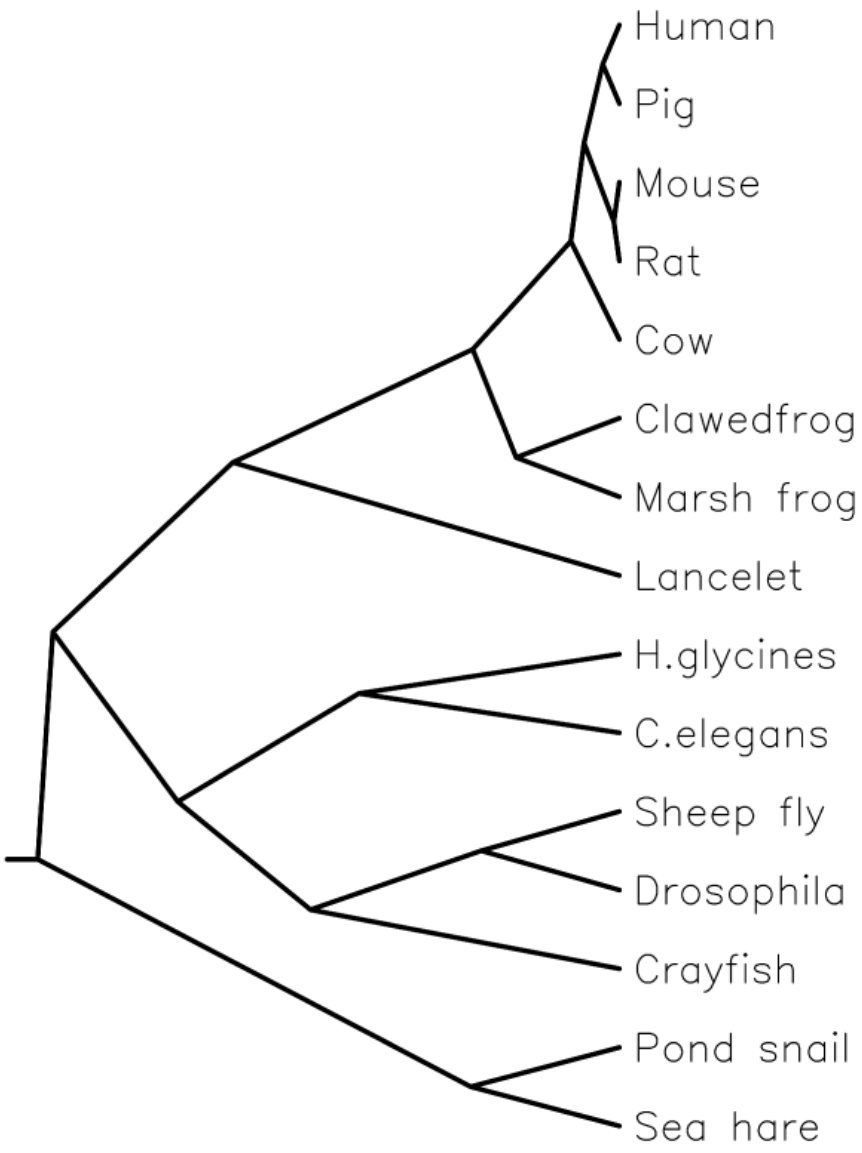

Figure 4. Phylogenetic tree of PC2s from various species based on distance matrix (Neighbor-Joining/UPGMA).