The Fish Biology Journal MEDAKA, 5: 23–28 (1993)

Gene structure of the hatching enzymes of the medaka and molecular basis for their action

Shigeki Yasumasu

Life Science Institute, Sophia University, Kioicho 7-1, Chiyoda-ku, Tokyo 102, Japan

Introduction

The hatching enzyme is defined as the enzyme which is secreted from the embryo of many animals and participates in the breakdown of the egg envelope at the time of hatching. The hatching enzyme system of the medaka, *Oryzias latipes*, is known to consist of two distinct proteolytic enzymes, high choriolytic enzyme (HCE) and low choriolytic enzyme (LCE) (Yasumasu *et al.*, 1988; Yamagami *et al.*, 1993). In early embryos, HCE and LCE are synthesized in the same hatching gland cells and packaged in the same secretory granules (Yasumasu *et al.*, 1992a). At the time of hatching, they are secreted together and digest the chorion cooperatively (Yamagami *et al.*, 1992). Recently, cDNAs for HCE and LCE were cloned and their primary structures were deduced (Yasumasu *et al.*, 1992b). These results will allow us to analyze the mechanism of enzyme action on the egg envelope in terms of the molecular structure of the hatching enzyme.

On the other hand, the hatching enzyme is synthesized only in the differentiated hatching gland cells in developing embryos. Therefore, the enzyme is a good probe for analyzing the mechanism behind the synthesis of embryo-specific proteins in terms of the expression of their genes and its regulation.

This article describes two types of approaches to the study of the hatching enzyme in medaka at the molecular level; the mechanism of egg envelope digestion by the hatching enzyme and the structure of the gene and its expression.

Cloning of cDNAs for HCE and LCE

Elucidation of the primary structure of the enzyme and the preparation of specific probes used in the search for the gene(s) were prerequisites for the study of the hatching enzyme at the molecular level.

cDNAs for HCE and LCE were cloned from a λ gt 11 library constructed from poly(A) $^{+}$RNA of day 3 embryos by immunological screening (Yasumasu *et al.*, 1992b). For HCE, cDNAs (HCE21 and HCE23) were cloned which were very similar, having a nucleotide sequence similarity of 93%. HCE21 and HCE23 were 940 and 910 bp long and contained open reading frames encoding 279 and 270 amino acids, respectively. HCE is synthesized in the form of a preproenzyme containing a signal peptide of 20 amino acids, a propeptide (activation peptide) of 50 (or 59) amino acids and a mature enzyme of 200 amino acids. This primary structure meets the criterion of a secretory protease. Two isoforms of HCE have been known to occur in the hatching liquid. The two cDNA clones seem to correspond to the isoforms. LCE cDNA was 936 bp long and contained an 813-bp open reading frame, which encoded a preproenzyme comprising a signal sequence of 20 amino acids, a propeptide of 51 amino acids and a mature enzyme of 200 amino acids. The similarity of deduced amino acid sequences of the mature enzyme portion of HCE to that of LCE was 55% and 6 cysteine residues were well conserved in each of them (Fig. 1). The amino acid sequence, HExxH motif, which is known to constitute an active site in some metalloproteases was also found in both HCE and LCE. The sea urchin hatching enzyme, although reported as a collagenase-like enzyme possessing this active site consensus (LePage and Gache 1990), showed no significant similarity to HCE and LCE in either sequence as a whole or in molecular size. Recently, the astacin family, a group of metalloproteases, has been reported (Dumermuth *et al.*, 1991). Both HCE and LCE contained the HExxHxxGFxxHE motif, an active center consensus characteristic of this family (Fig. 1). Moreover, their overall amino acid sequences indicated a considerable similarity (about 40%) to those of many members of this family. Thus, HCE and LCE are considered to belong to this family.

This paper is a summary of a special lecture given in the 26th Annual Meeting of Japanese Developmental Biologists, on the occasion of receiving the Prize in Memorial of Late Prof. Yoshihiro Kato, 1993.

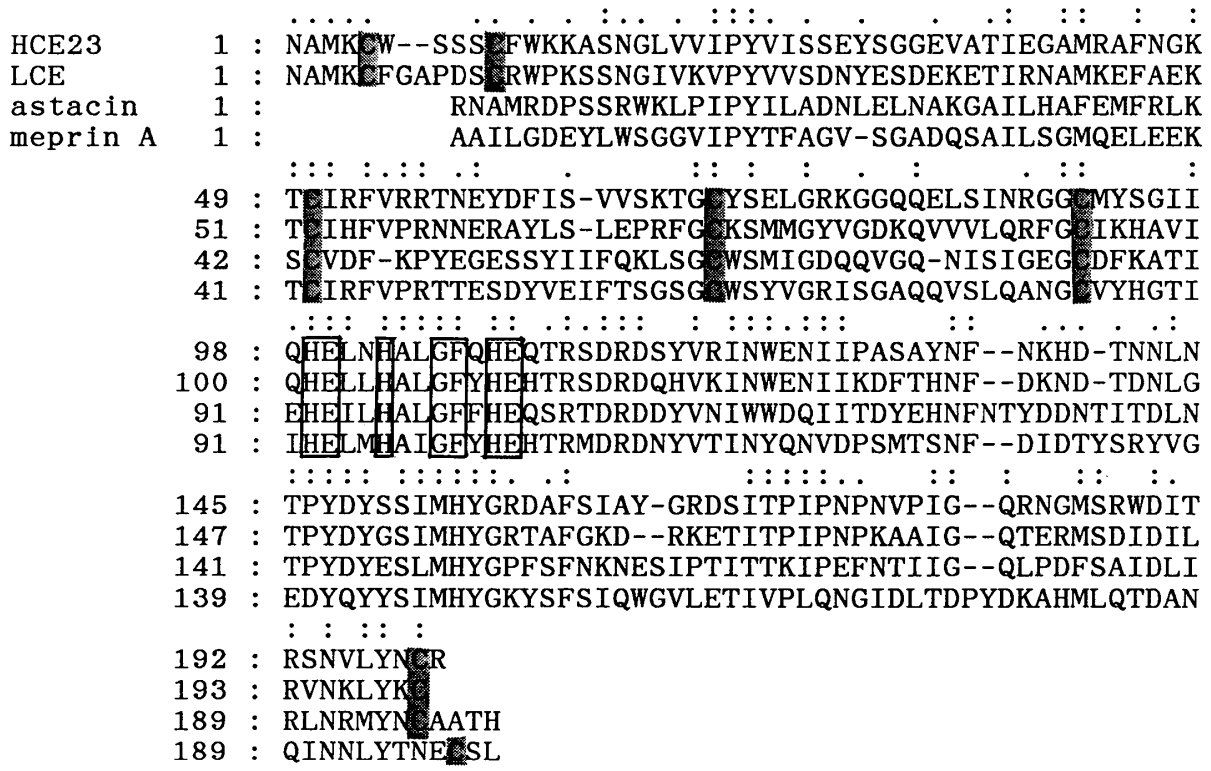

Fig. 1. Alignment of the amino acid sequences of four proteases of the astacin family including HCE23 and LCE.
Identical amino acid residues in both HCE23 and LCE are indicated by a single dot. Identical residues in at least three of the four sequences are indicated by two dots. Boxes indicate the amino acid residues of the active site consensus sequence of the astacin family. Cysteine residues are indicated by shadowed boxes. References for the reported sequences are as follows; astacin from *Astacus fluviatilis* (Tinati *et al.*, 1987), meprin A from the mouse kidney brush border (Dumermuth *et al.*, 1991).

Molecular mechanism of egg envelope digestion by the hatching enzymes

HCE and LCE were purified from the hatching liquid using gel filtration column chromatography and cation exchange column chromatography (Yasumasu *et al.*, 1989a, b). The purified HCE and LCE were very similar with respect to proteinase properties as well as protein chemical characteristics. The molecular weights of HCE and LCE estimated by SDS-PAGE were 24K and 25.5K, respectively. Both of them were basic proteins and their pI's were about 9.5–10.5. They had similar pH optima for proteolytic activity in a slightly alkaline range. They were found to be Zn-proteases from the results of metal analyses and the reactivation experiment of EDTA-denatured apoenzymes.

In spite of possessing similar characteristics as the proteolytic enzyme, their modes of action toward the egg envelope, chorion, were very different. HCE is a protease that has choriolytic activity, while LCE shows little choriolytic activity. When both enzymes were combined, however, they showed a marked synergistic choriolytic activity (Yasumasu *et al.*, 1988). When purified HCE acts upon isolated chorion, it swells the inner layer of chorion remarkably, releasing some low molecular weight peptides. On the other hand, purified LCE alone could not affect intact chorion but solubilized the HCE-swollen inner layer of chorion very efficiently. The combined action of HCE and LCE resulted in quick solubilization of the chorion without any significant swelling of the inner layer, as in the case of natural hatching. Thus, the synergistic choriolysis by both HCE and LCE is the result of the cooperative action of these two enzymes, i.e. a swelling of chorion by HCE, followed by a rapid solubilization of the swollen chorion by LCE.

Recently, it was found that HCE exhibited a unique proteolytic action toward chorion. HCE bound to the inner layer of chorion tightly when it exerted choriolytic action. The specificity of binding activity of HCE was estimated by using other insoluble substrates. HCE bound also to the egg chorion of rainbow trout but did not bind to the collagen filaments (Lee *et al.*, 1992). One of the monoclonal antibodies (MAB) against HCE gave us some interesting information about the binding activity of HCE (Yasumasu *et al.*, 1989c).

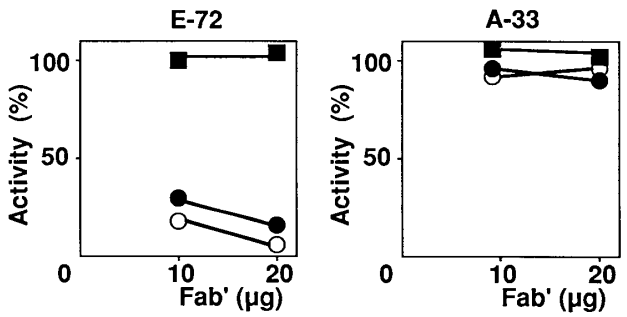

Fig. 2. The effects of MABs (E-72, A-33) on proteolytic activity, choriolytic activity and chorion binding activity of HCE. Three μg of HCE was incubated with an indicated amount of Fab' fragment of the MABs at 4°C for 5 hr prior to the assay of the activities. The values were expressed as a percent of activity in the absence of a Fab' fragment of MAB. ■: Proteolytic activity determined by using casein as a substrate, ○: choriolytic activity determined by turbidimetric method, ●: chorion binding activity determined by the immunological method.

Fig. 2 shows the effect of Fab's derived from two types of MABs, E-72 and A-33, on proteolytic activity, choriolytic activity and chorion binding activity of purified HCE. A-33 Fab' exhibited no effect on any of those activities. E-72 Fab' was found not to affect the proteolytic activity but did inhibit both the choriolytic activity and the chorion binding activity. An analysis using one type of MAB, E-72, revealed that the catalytic (choriolytic as well as proteolytic) activity could be separated from the chorion-binding activity. Moreover, binding of HCE to chorion is closely associated with its performance of choriolytic swelling. Considering that hydrolysis of both chorion and casein is performed probably through the same catalytic site of HCE molecule, it is suggested that a binding site and a catalytic site are separately present on an HCE molecule. When HCE attacks chorion, it binds to the chorion at the binding site and releases some peptides through the catalytic site. When HCE attacks some soluble substrates such as casein, however, it behaves like any other ordinary proteases by binding to and digesting the substrates at the catalytic site only. The existence of such a "binding site" is also suggested for other proteases, collagenase and the sea urchin hatching enzyme, which act on solid substrates (Murphy *et al.*, 1992; Nomura *et al.*, 1993). A high affinity to chorion is a unique character of this enzyme, and seems to be useful in the initiation of dissolution of the solid structure, chorion.

A comparison of amino acid sequences of the astacin family proteases shows that four cysteine residues are well conserved. They are possibly important for maintaining the three-dimensional conformation essential for functioning as a proteolytic enzyme. In addition to the four cysteine residue, two cysteine residues are in the N-terminal region of HCE and LCE. The N-terminal regions of HCE and LCE (amino acid numbers 1–14 for HCE and 1–16 for LCE) showed no homology to those of other members of the astacin family (Fig. 1). It may be expected that the unique sequences in the hatching enzymes play some important role in choriolysis, such as "binding".

Analysis of the mechanism of choriolytic action of these enzymes should be performed in parallel with that of the molecular structure of their substrate. Recently, we analyzed peptides that were released on swelling of the chorion by HCE and found that a peptide, named P-H2, consisted of 33 mol% of Pro and 25 mol% of Glx. Moreover, P-H2 contained most of γ-Glu ε-Lys crosslinks present in the chorion which were probably formed at the time of hardening of the chorion (Lee *et al.*, 1993). These results suggested that HCE selectively removed a restricted portion(s) of chorion which contained domains rich in Pro, Gln and γ-Glu ε-Lys. As a result of the swelling of the inner layer of chorion, LCE was able to digest it efficiently. The action of HCE probably reduces the hydrophobicity of the chorion making it accessible to LCE. Thus, the hatching enzyme study may be helpful also for elucidating the egg envelope structure.

Structure of the genes for HCE and LCE, and their expression

The genes encoding HCE and LCE were isolated from the EMBL-3 genomic libraries constructed from DNA of fish of the inbred drR strain, and their structure was determined (Yasumasu *et al.*, 1993). Fig. 3 shows the exon-intron structures of the LCE and HCE genes. The HCE gene was found to be 1 kbp long and to lack introns. Moreover, we found it to be a multigene, i.e., there were eight copies of HCE genes in a genome. The sequence of the coding region of their genes was almost identical. Seven of the eight HCE genes possessed entire coding sequences for HCE, while the remaining one lacked the initiation codon, i.e., one base substitution occurred in the initiation codon (ATG to AAG).

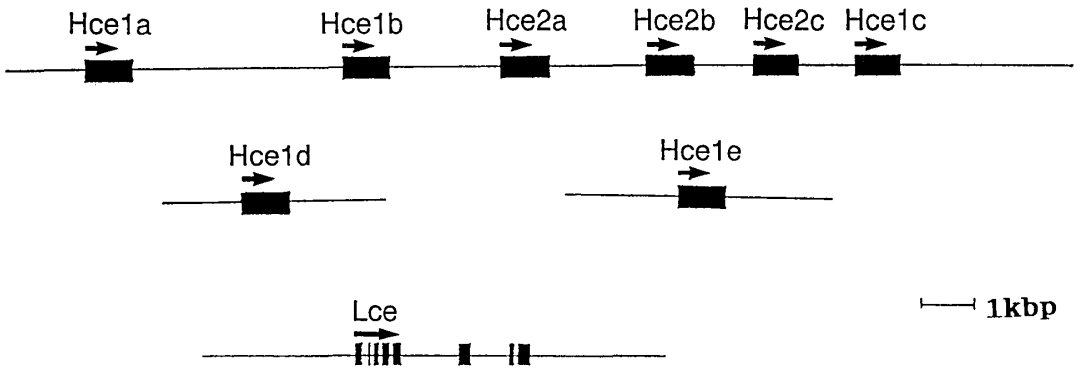

Fig. 3. Structure of the genes for HCE and LCE. The genes for HCE and the exons of an LCE gene are indicated by boxes. Solid lines indicate the introns and the flanking regions. The orientation of the coding region is shown by arrows.

Predicted amino acid sequences of their mature enzyme portions of all the HCE genes were very conservative although some substitutions and deletions were observed in the propeptide regions. The predicted amino acid sequences of mature enzyme portions of HCE23 and HCE21 were very similar with a similarity of 95% (only 9 amino acid substitutions in 200 amino acids of mature enzyme portion). Amino acid sequences of the substituted portions were found to be well conserved in the HCE genes, so that we could easily distinguish the genes for HCE23 from those for HCE21. We designated the genes encoding HCE23 and HCE21 as Hce1 and Hce2, respectively. Five copies of HCE23 gene (Hce1a-e) and three copies of HCE21 gene (Hce2a-c) were found in a genome. Within about 25-kb of genomic DNA, six of them, i.e., three copies of Hce1 and three copies of Hce2 gene, formed a cluster, while the remaining two, two copies of HCE23 gene (Hce1d and Hce1e), were located separately.

Unexpectedly, the structure of the LCE gene was quite different from that of the HCE gene. The LCE gene (Lce) was located within a 3.6 kbp stretch of genomic DNA and consisted of eight exons with seven introns. Sequences of the exon-intron boundaries complied well with the GT-AG splice junction rule. A Southern blot analysis revealed that the LCE gene is a single copy gene.

It was found that five- to ten-fold the amount of HCE relative to that of LCE was necessary to efficiently digest the egg envelope from the result of *in vitro* digestion of chorion by the purified hatching enzyme. In fact, a larger amount of HCE protein relative to LCE protein is found in zymogen granules of hatching gland cells (Yasumasu *et al.*, 1992a). Such a great difference in the copy number between the HCE gene and the LCE gene may be a cause of the difference in the amounts of both the enzymes.

The result of primer extension analysis of the LCE gene revealed that a TATA box consensus sequence was located at 28 bp upstream from the transcription-start site. TATA box consensus sequences were also found in the upstream regions of all the HCE genes. The nucleotide sequences of the 5'-flanking regions of all the HCE genes indicated an 80% to 95% similarity within the range of 200–400 bp. On the contrary, 5'-flanking region of the HCE genes did not show an overall similarity to that of the LCE gene within the range of 1.5 kb.

As described above, HCE and LCE are very similar in physicochemical property as proteins and belong to the same protease family. From this point of view, the genes for HCE and LCE are considered to have evolved from the same ancestral gene. In general, structures of the exon and intron are well conserved among the same families of proteins, and these genes appear to have evolved from an ancestral gene by duplication, followed by divergence of both the protein-coding regions and the sequences responsible for expression of the genes. In fact, it has been reported that the exon-intron organization is well conserved in many gene families (for instance, the genes of the collagenase family). Therefore, some dramatic differences in the structure of the HCE and LCE genes mentioned above seem hardly to be explicable by the ordinary evolutional process of one gene family alone. On the other hand, many reports have described intron-less genes that may have evolved from an intron-containing progenitor gene. These intron-less genes are considered to have arisen by reverse transcriptase-mediated processing of a transcript from an intron-contain-

ing ancestral gene, i.e. retroposons. As a result of the retroposition of a gene, both the intron-containing gene and intron-less gene coexist in the genome of the same species. The retroposons generally possess no promoter, so that many of these intron-less genes are pseudogenes (Vamin, 1985; Rogers, 1985). However, a functional intron-less gene which shows traces of a retroposon have been reported (McCarrey and Thomas, 1987; Fourel *et al.*, 1992). Taking into account the general idea described above, the HCE are possibly retroposed genes and may have evolved through two successively occurring processes, i.e. loss of the intron and gene duplication. Thus, the HCE and LCE genes may have passed through quite different evolutional processes from their common ancestral gene. At present, the gene structure of the other proteases of the astacin family has not yet been reported. We expect that the molecular evolution of the genes of the astacin family proteases can be discussed in the near future.

Provided that the HCE genes were retroposed genes, it is of great interest as to why two enzymes, HCE and LCE, are regulated to be expressed concurrently. In general, the 5'-upstream regions containing putative promoter regions in intron-less retroposed genes are different from those of intron-containing genes because the retroposition of the gene is considered to occur randomly. The expression of the intron-less genes and intron-containing genes is known to be regulated in a different spatio-temporal manner. Although the 5'-upstream regions of HCE genes and that of the LCE gene were also quite different, expression of both the HCE gene and the LCE genes was initiated and proceeded synchronously during development, as two constituents of the hatching enzyme system. To analyze this mechanism, it is important to identify the promoter regions of the HCE and LCE genes. It is hoped that a cis-regulating element(s) with a common sequence for both genes will be found. To understand concurrent expression of both genes, the study of some trans-regulation mechanisms, i.e., mechanisms through a trans-acting regulatory factor(s) on transcriptional activity, will also be necessary. The search for such a regulatory factor(s) should be done in the near future.

Elucidation of the relationship between expression of the hatching enzyme genes and differentiation of the hatching gland cells would also be very interesting. According to the results of *in situ* hybridization experiments employing an anti-sense RNA derived from a cDNA fragment for HCE as a probe, the genes of HCE were detected to be expressed first in a cell group located at the anterior end of the forebrain of the embryos (the so-called pillow or Polster) at the optic vesicle stage (Inohaya *et al.*, 1993). These results indicate that expression of the hatching enzyme genes is first detected in an area quite different from the final location of the hatching gland cells (pharyngeal cavity) and that expression of the genes continues in the cells migrating to their final destination during the last phase of morphogenesis. Although the hatching enzyme of medaka works at the last stage of embryonic development, expression of their genes is initiated at an early stage of embryogenesis. Terminal differentiation of the hatching gland cells in the early stage embryo may be controlled by the earlier expression of some trans-acting factor(s) concerned with the differentiation factor.

Conclusion

Two types of molecular approach to the hatching enzyme study were described. One is concerned with the function of the hatching enzyme and the other with its gene structure and expression.

Two constituent proteases, HCE and LCE, in the medaka hatching enzyme system are closely related to each other with regard to primary structure and physicochemical property. On the other hand, their modes of action on the substrate, the chorion, are very different; HCE swells the inner layer of chorion by binding to the chorion and then releasing some restricted peptides rich in Pro, Glx and γ-Glu ε-Lys, while LCE can not digest intact chorion but can digest completely the swollen chorion. To understand the difference in the mode of action at the molecular level, a protein engineering approach will be necessary.

The genes for HCE are considered to be retroposed intron-less genes, while the gene for LCE is an ordinary intron-containing gene. Although those genes are considered to have evolved through different processes, spatio-temporal regulation of their expression is established in the same way during differentiation of the hatching gland cells.

To understand this point, approaches from molecular biology and developmental biology will be required. In particular, some *cis*-regulating site and *trans*-regulatory factors should be researched in the near future.

Acknowledgments

The authors wish to thank Prof. K. Yamagami and Assistant Prof. I. Iuchi of Sophia University for their kind advice and supports throughout the present work. The author also wishes to express his cordial thanks to Prof. H. Shimada of Hiroshima University for his kind help in performing the cDNA cloning.

References

Dumermuth, E., E.E. Sterchi, W. Jiang, R.L. Wolz, J.S. Bond, A.V. Flannery and R.J. Beynon (1991) *J. Biol. Chem.*, **266**: 21381–21385.

Fourel, G., C. Transy, B.C. Tennant and M.A. Buendia (1992) *Mol. Cell. Biol.*, **12**: 5336–5344.

Inohaya, K., S. Yasumasu, M. Ishimaru, A. Ohyama, I. Iuchi and K. Yamagami (1993) *Zool. Sci.*, **10** Suppl.: 35.

Lee, K.S., S. Yasumasu, K. Nomura, I. Iuchi and K. Yamagami (1993) *Zool. Sci.*, **10** Suppl.: 35.

LePage, T. and C. Gache (1990) *EMBO J.*, **9**: 3003–3012.

McCarrey, J.R. and K. Thomas (1987) *Nature*, **326**: 501–505.

Murphy, G., J.A. Allan, F. Willenbrock, M.I. Cockett, J.P. O'Connell and A.J.P. Docherty (1992) *J. Biol. Chem.*, **267**: 9612–9618.

Nomura, K. and N. Suzuki (1993) *Zool. Sci.*, **10** Suppl.: 34.

Rogers, J.H. (1985) *Int. Rev. Cytol.*, **93**: 187–279.

Titani, K., H.-J. Torff, S. Hormel, S. Kumar, K.A. Walsh, J. Rodl, H. Neurath and R. Zwilling (1987) *Biochemistry*, **26**: 222–226.

Vamin, E.F. (1985) *Ann. Rev. Genet.*, **19**: 253–272.

Yamagami, K., T.S. Hamazaki, S. Yasumasu, K. Masuda and I. Iuchi (1992) *Int. Rev. Cytol.*, **136**: 51–92.

Yamagami, K., S. Yasumasu and I. Iuchi (1993) *In Physiological and Biochemical Aspects of Fish Development* (Eds. B.T. Walther and H.J. Fyhn). Univ. Bergen, Bergen, pp. 104–111.

Yasumasu, S., I. Iuchi and K. Yamagami (1988) *Zool. Sci.*, **5**: 191–195.

Yasumasu, S., I. Iuchi and K. Yamagami (1989a) *J. Biochem.*, **105**: 204–211.

Yasumasu, S., I. Iuchi and K. Yamagami (1989b) *J. Biochem.*, **105**: 212–218.

Yasumasu, S., S. Katow, Y. Umino, I. Iuchi and K. Yamagami (1989c) *Biochem. Biophys. Res. Commun.*, **162**: 58–63.

Yasumasu, S., S. Katow, T.S. Hamazaki, I. Iuchi and K. Yamagami (1992a) *Dev. Biol.*, **149**: 349–356.

Yasumasu, S., K. Yamada, K. Akasaka, K. Mitsunaga, I. Iuchi, H. Shimada and K. Yamagami (1992b) *Dev. Biol.*, **153**: 250–258.

Yasumasu, S., K. Inohaya, K. Yamada, I. Iuchi, H. Shimada and K. Yamagami (1992c) *Zool. Sci.*, **9**: 1157.

Yasumasu, S., K. Inohaya, K. Yamada, I. Iuchi, H. Shimada and K. Yamagami (1993) *Zool. Sci.*, **10** Suppl.: 35.