Ovochymase, a *Xenopus laevis* egg extracellular protease, is translated as part of an unusual polyprotease

**LeAnn L. Lindsay**, **Joy C. Yang**, and **Jerry L. Hedrick**

Section of Molecular and Cellular Biology, University of California, Davis, CA 95616

Communicated by John C. Gerhart, University of California, Berkeley, CA, August 2, 1999 (received for review April 5, 1999)

**ABSTRACT** Ovochymase, an extracellular *Xenopus laevis* egg serine active-site protease with chymotrypsin-like (Phe-X) substrate specificity, is released during egg activation. Molecular cloning results revealed that ovochymase is translated as part of an unusual polyprotein proenzyme. In addition to the ovochymase protease domain at the C terminus of the deduced amino acid sequence, two unrelated serine protease domains were present, each with apparent trypsin-like (Arg/Lys-X) substrate specificity, and thus, they were designated ovotrypsin1 (at the N terminus) and ovotrypsin2 (a mid domain). Also, a total of five CUB domains were interspersed between the protease domains. The presence of a hydrophobic signal sequence indicated that the polyprotein was secreted. Immunolocalization and Western blot studies of all three proteases showed that they are all present in the perivitelline space of unactivated eggs, apparently as proenzymes processed away from the original polyprotein. Western blot analysis also showed that the vast majority of the proteases in oocytes, eggs, and embryos were present as the proenzyme forms, suggesting that the functions of these proteases depend on very limited forms of activation.

Ovochymase, a 30-kDa *Xenopus laevis* egg serine active-site protease with chymotrypsin-like (Phe-X) substrate specificity, is released from dejellied eggs on activation, suggesting a role in fertilization or early development (1). Chymotrypsin-like protease activity has been found to be associated with fibers within the perivitelline space (extracellular matrix) of unactivated eggs treated briefly with the crude exudate from activated eggs, suggesting that ovochymase exists extracellularly as a proenzyme until activated and released (2). The use of trypsin inhibitors during egg activation prevents the release of ovochymase, suggesting that ovochymase is activated by a trypsin-like protease (1). A protease with a trypsin-like (Arg-X) substrate specificity of 45 kDa has been detected in activated egg exudate along with ovochymase, but its role has not been established (1).

In determining the structure and function of ovochymase, active ovochymase was purified from activated egg exudate, and the N-terminal amino acid sequence was determined (3). This information was used in the current study to obtain the complete ovochymase cDNA sequence. Ovochymase was found to be translated as part of a secreted polyprotein containing two trypsin-like proteases, a highly unusual situation in eukaryotic systems. Antibodies were generated toward each of the proteases for localization studies, revealing that all the proteases were secreted to the egg surface. Also, the majority of the protease proteins were present in the proenzyme forms during all stages of oocyte and embryo development. Thus, limited activation of the enzymes may be important to their function.

---

Abbreviation: RACE, rapid amplification of cDNA ends.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. U81290).

to whom reprint requests should be addressed. E-mail: Illindsay@ucdavis.edu.
multiple 5′ RACE products, an overlapping 3′ RACE procedure was performed.

DNA sequencing was performed by using a Sequenase kit (United States Biochemical) or through automated DNA sequencing performed by either the Division of Biological Sciences or the Advanced Plant Molecular Genetics DNA sequencing facilities on the University of California, Davis, campus. Sequences were analyzed with the Genetics Computer Group’s (Madison, WI) SEQUENCE ANALYSIS software by using default parameters.

RNA Analysis. Northern blots prepared by standard methods (9), containing mRNA from stage I–III oocytes or total RNA from oocytes representing each oogenesis stage (stages I–VI), were prehybridized at 42°C for 1–2 hr in 50% (vol/vol) formamide/6× SSC (0.15 M sodium chloride/0.015 M sodium citrate, pH 7)/5× Denhardt’s solution (0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% BSA)/0.5% SDS/100 mg/ml salmon sperm DNA. Hybridization was performed overnight under the same conditions by using [32P]dCTP-incorporated primary PCR ovochymase clone, followed by washing under high stringency conditions in 0.1× SSC/0.1% SDS at 65–70°C.

Reverse transcription–PCR was performed with a Titan One-Step Reverse Transcription–PCR kit from Roche Molecular Biochemicals by using primers to amplify the ovochymase transcript and primers to amplify β-tubulin as a control as described (10).

Ovarian tissue was prepared for in situ hybridization by the method of O’Keefe et al. (11), and probed for ovochymase mRNA by using a digoxigenin-labeled RNA probe as described by the manufacturer (Roche Molecular Biochemicals).

Immunolocalization and Western Blot Analysis. Polyclonal antibodies to the three proteases were generated toward protease-specific peptides based on the deduced amino acid sequences. These peptides (see Fig. L4), synthesized by Genosys (The Woodlands, TX), corresponded to amino acids 285–296 (CWGAKQIIRSQSG), 773–782 (CQNEKEQFSI), and 1,311–1,321 (IQNSKKRHYCG). The polypeptides were conjugated to maleimide-activated keyhole limpet hemocyanin (Pierce) and standard methods (12).

For immunolocalization studies, live or methanol-fixed (overnight at 4°C) dejellied eggs were first blocked for 10 min to 1 hr in fluorescently labeled secondary antibodies (Sigma). Antibodies were diluted in Tris-buffered saline/0.1% Tween 20/5% (vol/vol) nonfat dry milk, and wash solutions used the same buffer system without the milk. Peroxidase-reactive bands were visualized by enhanced chemiluminescence (Amersham Pharmacia).

RESULTS

Molecular Cloning. A 4,628-bp ovochymase cDNA was isolated with an ORF of 4,569 bp, a 5′ untranslated region of 13 bp, and a 3′ untranslated region including a polyadenylation signal sequence for secretion, with a predicted cleavage site after the five CUB domains. 100 AA, 100 amino acids.
the deduced amino acid sequence of the ovochymase cDNA, by Edman degradation (3), was found starting at Val-1,294 of the mature ovochymase protease, determined (Fig. 2B). Amplification of β-tubulin (tub) was performed as a control.

A single transcript of 5.3 kilobases was detected by Northern blotting analysis in mRNA isolated from early-stage oocytes (Fig. 2A), which correlated with the size of the cDNA sequence obtained by cloning, considering that only a small portion of the 5′ untranslated region was sequenced. A blot containing ovochymase mRNA, showing relative levels of ovochymase message in total RNA preparations from oocytes of different stages (I–VI), (C) Reverse transcription–PCR amplification of ovochymase message (ovo) with 1 μg each of total RNA from tissue, oviduct (PR), liver (L), and skeletal muscle (M). Amplification of β-tubulin (tub) was performed as a control. (Fig. 2A, B). Another Northern blot shows the size of the cDNA sequence (3), which is comparable to the measured molecular mass of 30 kDa as determined by SDS/PAGE (1). A database search showed that the ovochymase protease domain had the highest amino acid identity to other serine proteases with similar substrate specificity (bulky hydrophobic or aromatic residues) such as chymotrypsin and elastase, both of which have 41% identity to ovochymase (Table 1).

Ala-16. No C-terminal transmembrane region was present, nor was there a furin cleavage site indicative of constitutive secretion (16). With the signal sequence removed, the secreted protein would have a molecular mass of 165 kDa. The actual molecular mass may be larger, depending on the extent of glycosylation of the 11 potential N-linked glycosylation sites and 3 predicted (17) O-linked glycosylation sites (Fig. 1A). The protease domain extended to the end of the ORF.

Fig. 2. (A) Northern blot analysis of ovochymase mRNA, showing a single transcript at 5.3 kilobases. Positions of RNA size markers are shown to the left. (B) A Northern blot showing relative levels of ovochymase messages from oocytes of different stages (I–VI). (C) Reverse transcription–PCR amplification of ovochymase message (ovo) with 1 μg each of total RNA from tissue, oviduct (PR), liver (L), and skeletal muscle (M). Amplification of β-tubulin (tub) was performed as a control. (Fig. 2A, B). No C-terminal transmembrane region was present, nor was there a furin cleavage site indicative of constitutive secretion (16). With the signal sequence removed, the secreted protein would have a molecular mass of 165 kDa. The actual molecular mass may be larger, depending on the extent of glycosylation of the 11 potential N-linked glycosylation sites and 3 predicted (17) O-linked glycosylation sites (Fig. 1A).

A single transcript of 5.3 kilobases was detected by Northern blotting analysis in mRNA isolated from early-stage oocytes (Fig. 2A), which correlated with the size of the cDNA sequence obtained by cloning, considering that only a small portion of the 5′ untranslated region was sequenced. A blot containing total RNA from stage I–VI oocytes (Fig. 2B) showed that ovochymase message was first detectable at stage I of oogenesis, peaked at stage IV, and then disappeared at later stages. In situ hybridization confirmed that ovochymase mRNA was produced by the oocytes and not by follicle cells that were also present in oocyte preparations (data not shown). Analysis of various tissues by reverse transcription–PCR showed that the message for ovochymase was specific for ovary, with no message detected in pars recta oviduct, liver, or skeletal muscle (Fig. 2C).

Protease Domain Characterization. The N-terminal amino acid sequence of the mature ovochymase protease, determined by Edman degradation (3), was found starting at Val-1,294 of the deduced amino acid sequence of the ovochymase cDNA, and the protease domain extended to the end of the ORF. Thus, the ovochymase protease domain represented the C terminus of the molecule. An Arg was found immediately preceding Val-1,294, which is consistent with the hypothesis that a trypsin-like protease is responsible for ovochymase activation (1). Processing at this site would release a mature ovochymase protease with a calculated molecular mass of 25 kDa, which, considering that there is one potential N-glycosylation site within this domain (Fig. 1A), is comparable to the measured molecular mass of 30 kDa as determined by SDS/PAGE (1). A database search showed that the ovochymase protease domain had the highest amino acid identity to other serine proteases with similar substrate specificity (bulky hydrophobic or aromatic residues) such as chymotrypsin and elastase, both of which have 41% identity to ovochymase (Table 1).

This similarity was reinforced by the presence of a common S1 substrate binding site among these three proteases, a Ser residue (Fig. 3, arrowhead). Interestingly, the ovochymase protease domain also showed a high level of identity (40%) to the protease domain of Drosophila nudel, a large multidomain protein found in the perivitelline space of Drosophila eggs, which is involved in an extracellular protease cascade that generates a signal to set up dorsalventral polarity in the embryo (18). This high degree of similarity existed in spite of the fact that the substrate specificity of nudel (Arg/Lys-X) differs from that of ovochymase (Phe specific). The other known Drosophila perivitelline-space serine proteases, snake, easter, and gastrulation defective (19), all of which act downstream of nudel (18), showed lower degrees of similarity to ovochymase (Table 1).

A BLAST analysis of the remaining cDNA sequence revealed two additional serine protease domains N-terminal to the ovochymase protease domain (Fig. 1A and B). The protease domain alignment (Fig. 3) indicated that these two protease domains should be functional, because the His, Asp, and Ser of the serine protease catalytic triad were present, along with the conserved N-terminal hydrophobic amino acids and the Asp immediately preceding the active-site Ser necessary for proper folding of the catalytic site (20). The S1 substrate binding residue of Asp (Fig. 3, arrowhead) indicated a substrate specificity similar to trypsin, cleaving after basic amino acids. These characteristics led us to term these proteases ovotryptase1 and ovotryptase2, the order determined by their placement in the putative polyprotein. Both ovotryptase protease domains were preceded by Arg residues, indicating that the proteases would be activated by trypsin-like proteases. Results of database searches (Table 1) revealed that ovotryptase1 had the highest amino acid identity to Xenopus oviductin (47% identity), an oviductal trypsin-like protease that acts on the egg-envelope glycoprotein gp43 (21), whereas ovotryptase2 was most similar to Xenopus trypsin (44% identity).

Table 1. Percentages of amino acid identity between serine protease domains

<table>
<thead>
<tr>
<th>Serine protease</th>
<th>Ovochymase</th>
<th>Ovotryptase1</th>
<th>Ovotryptase2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovochymase (Xenopus)</td>
<td>—</td>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>Ovotryptase1 (Xenopus)</td>
<td>32</td>
<td>—</td>
<td>36</td>
</tr>
<tr>
<td>Ovotryptase2 (Xenopus)</td>
<td>39</td>
<td>36</td>
<td>—</td>
</tr>
<tr>
<td>Oviductin (Xenopus)</td>
<td>34</td>
<td>47</td>
<td>38</td>
</tr>
<tr>
<td>Trypsin (Xenopus)</td>
<td>35</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>Chymotrypsin b (bovine)</td>
<td>41</td>
<td>37</td>
<td>39</td>
</tr>
<tr>
<td>Elastase 2a (human)</td>
<td>41</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>Nudel (Drosophila)</td>
<td>40</td>
<td>36</td>
<td>37</td>
</tr>
<tr>
<td>Snake (Drosophila)</td>
<td>28</td>
<td>30</td>
<td>—</td>
</tr>
<tr>
<td>Easter (Drosophila)</td>
<td>37</td>
<td>31</td>
<td>36</td>
</tr>
<tr>
<td>Gastrulation defective (Drosophila)</td>
<td>27</td>
<td>24</td>
<td>23</td>
</tr>
</tbody>
</table>

Values of highest identity are underlined. GenBank accession numbers not given elsewhere are nudel (P98159), snake (X04513), easter (J03154); and gastrulation defective (AF056311).
tity). There did not seem to be any significant similarities between the ovotryptase domains and the ovochymase domain that would suggest a recent common ancestor, and there were no outstanding similarities between either of the ovotryptase domains and any of the four *Drosophila* egg perivitelline space proteases.

**CUB Domains.** The BLAST analysis of the ovochymase cDNA also revealed the presence of five CUB domains, which, along with the three protease domains, completed the domain structure of the translated protein (Fig. 1). Although the function of CUB domains is largely unknown, these domains are common to several types of extracellular proteases, such as the tolloid protein involved in dorsoventral-axis formation, bone morphogenetic protein (BMP), and the complement subcomponents Clr and Cls (22). An alignment of the polyprotein CUB domains with those from various other proteins is shown in Fig. 4. The most characteristic features are the conservation of the four Cys residues that form two disulfide bonds, and various hydrophobic and aromatic amino acids that participate in forming the antiparallel \( \beta \)-barrel topography of the molecule (22).

**The Proteases as Proteins.** A Western blot probed for ovochymase protein (Fig. 5A) showed the expected band at 30 kDa in samples of concentrated egg exudate or purified mature ovochymase protease, which represents the protease domain alone (3). In lysates of whole unactivated eggs, the antibodies recognized a band at \( \approx 50 \) kDa, most likely representing the ovochymase proenzyme as the protease domain with the fifth CUB domain attached (calculated total molecular mass without glycosylation \( \approx 546 \) kDa; see Fig. 1B). An ovary lysate containing early-stage oocytes also showed only a 50-kDa ovochymase protein, suggesting that the polypeptide is processed shortly after translation; no immunoreactive band at a higher molecular mass, which would represent the intact polyprotein, was observed, even after long exposure times. Identical results were observed under both reducing and nonreducing electrophoretic conditions (not shown). Western blots of activated dejellied egg lysates showed results that were identical to those of the whole unactivated egg lysates, indicating that only a small fraction of ovochymase was released on egg activation (data not shown). Calculations based on the intensity of the bands on the Western blots for concentrated egg exudate (representing exudate from 10,000 eggs in 5 ml) versus whole-egg lysates (4 eggs per ml) indicated that only about 0.01% of the ovochymase proenzyme present in an egg was activated and released as the mature form when dejellied eggs were activated. Egg lysates obtained at time points up to 30 hr after fertilization showed a constant level of the 50-kDa ovochymase protein and no 30-kDa protein.

Similar results were found by using antibodies toward the ovotryptase proteases (Fig. 5B and C). In unactivated eggs, a single 50-kDa protein was detected for ovotryptase1, which may correspond to the protease domain with the first CUB domain attached (calculated total molecular mass without glycosylation \( \approx 46 \) kDa); see Fig. 1B). An ovary lysate containing early-stage oocytes also showed only a 50-kDa ovotryptase protease, which is also processed shortly after translation; no immunoreactive band at a higher molecular mass, which would represent the intact polyprotein, was observed, even after long exposure times. Identical results were observed under both reducing and nonreducing electrophoretic conditions (not shown). Western blots of activated dejellied egg lysates showed results that were identical to those of the whole unactivated egg lysates, indicating that only a small fraction of ovotryptase was released on egg activation (data not shown). Calculations based on the intensity of the bands on the Western blots for concentrated egg exudate (representing exudate from 10,000 eggs in 5 ml) versus whole-egg lysates (4 eggs per ml) indicated that only about 0.01% of the ovotryptase protease present in an egg was activated and released as the mature form when dejellied eggs were activated. Egg lysates obtained at time points up to 30 hr after fertilization showed a constant level of the 50-kDa ovotryptase protease and no 30-kDa protein.

**Multiple alignment of the CUB domains in the Xenopus ovochymase proteosome (ovoCUBs) with those from *Drosophila* tolloid (P25723), mouse BMP-1 (P98063), pig spermadhesin AONQ-1 (P26322), Xenopus oviductin (U81291), and Xenopus hatching enzyme (D89632).**
with either the second and third or the third and fourth CUB domains attached (calculated total molecular masses without glycosylation = 59 kDa and 57 kDa, respectively). These protein bands were also present on Western blots of ovary lysates and persisted throughout the 30-hr postfertilization period (data not shown). No high molecular-mass bands were observed that would represent the intact polyprotein. Neither of the ovotryptase proteins were detected in activated egg exudate, indicating that these proteases were not released from dejellied eggs (data not shown). Therefore, it is unlikely that the trypsin-like protease purified from activated egg exudate is one of the ovotryptase proteases.

**Immunolocalization of Proteases.** The anti-protease antibodies were used to immunolocalize the three proteases by confocal microscopy by using live dejellied eggs to detect surface proteins or by using eggs fixed in methanol to examine binding to internal egg proteins. Fig. 6 A, C, and E shows that in live unactivated eggs, antibodies to all three proteases labeled the egg surfaces; a high magnification view of ovochymase labeling showed a wide area of labeling between the plasma membrane and egg envelope (Fig. 6 G and H). Methanol-fixed eggs had the same labeling patterns as live eggs (not shown). Activated dejellied eggs (Fig. 6 B, D, and F) showed reduced levels of labeling at the egg surface for ovochymase and ovotryptase1. This result is consistent for ovochymase in that some ovochymase protease activity is released at egg activation but inconsistent with the Western blot data that showed the vast majority of the protein remained associated with the egg. Perhaps, with live eggs, the antibody bound to a subpopulation of the ovochymase proteins, whereas the remaining proteins were inaccessible. It follows that for ovotryptase1, small amounts of protease may be released at egg activation but in amounts too small to be detected by Western blotting.

**DISCUSSION**

**Polyprotein Structure.** Unexpectedly, the cDNA cloning results revealed that ovochymase was translated as part of a large polyprotein containing two additional serine protease domains, as well as several CUB domains. The additional protease domains, designated ovotryptase1 and ovotryptase2, were predicted to be functional, with substrate specificities similar to trypsin, and apparently activated by a trypsin-like protease or proteases. The deduced amino acid sequence for ovochymase confirmed experimental evidence (1) that it is a secreted serine protease with specificity for amino acids with bulky hydrophobic/aromatic side chains, and it is activated by a trypsin-like protease. The ovotryptase domains did not have a significant level of similarity to each other or ovochymase, ruling out gene duplication as the mechanism of polyprotein formation. More likely, the polyprotein was formed through gene splicing.

Western blot analysis supported the conclusions that the cDNA sequence found for ovochymase does indeed represent a translated polyprotein and that this original protein is processed soon after translation. Antibodies generated toward peptides based on the deduced amino acid sequences of the three protease domains each recognized a single protein in egg lysates, each one smaller than that expected for the intact polyprotein. The molecular masses of these three proteins fit a logical model of polyprotein processing in which hydrolysis would occur at linkage points between protein domains, leaving each protease domain with at least one CUB domain attached.

CUB domains seem to be a common feature of *Xenopus* extracellular proteases involved in developmental processes, as
well as proteases in other organisms (22). *Xenopus* oviductin, a secreted oviductal serine protease that cleaves egg-envelope gp43 (ZPC) to gp41 during passage through the par recta oviduct to render the envelope penetrable to sperm (21), possesses two CUB domains C-terminal to the protease domain (10). *Xenopus* hatching enzyme, a secreted metalloprotease of the astacin family, which preferentially hydrolyzes egg-envelope gp120 to allow for embryo escape (23), also possesses two CUB domains C-terminal to the protease domain (24). We hypothesize that the CUB domains of these proteases function in extracellular-matrix binding. This hypothesis is supported by the fact that the mammalian sperm-madhesins, which are essentially one CUB domain, are secreted by the male reproductive tract to bind to the sperm membrane and have been shown to have egg-envelope binding activity (25). For mature *Xenopus* oviductin and the hatching enzyme, the CUB domains may bind to the envelope to concentrate the proteases in the region of their substrates. The CUB domains attached to ovochymase and the ovotryptases may serve to anchor the proenzymes to the egg surface. The hypothesis that CUB domains function in binding to the egg is testable, e.g., by using expressed CUB domains in egg-binding studies.

This example of a polyprotein containing multiple proteases is, to the extent of our knowledge, one of the first reported for a eukaryotic system. Viral DNA and RNA commonly code for polyproteins, and polyproteins have also been found representing prokaryotic proteins, plant proteins, fish antifreeze glycoproteins, and peptide hormones in a wide range of eukaryotic species. However, in higher animals, polyproteins represent relatively short polypeptides, not proteins, and in no reported case have the polyproteins contained protease domains. The only known exception is the *Xenopus* oviductal secreted protease oviductin, mentioned above, which was recently cloned by our laboratory (10). Oviductin is translated as the N-terminal portion of a protein containing another complete, albeit apparently inactive, serine protease domain; the two protease domains are cleaved apart before or shortly after secretion. Interestingly, ovochymase I had a very high amino acid identity (47%) to the oviductin protease domain, suggesting that these proteases may be evolutionarily related.

Why are ovochymase and the ovotryptases translated as part of a polyprotein? Cotranslation could be important in several ways, e.g., to ensure that the proteases are secreted together for proper localization on the egg surface/extracellular matrix or as a mechanism to produce the proper stoichiometry required for assembly of a multiprotein complex.

**Protease Function.** Previous studies with protease inhibitors suggested a role for ovochymase in the hydrolysis of egg-envelope component gp69/64 as part of the block to polyspermy (1). However, recent studies with purified ovochymase have indicated that it is not involved in the reaction (L.L.L. and J.L.H., unpublished observations). The results presented here, showing that ovochymase and the ovotryptases are present primarily in the proenzyme forms throughout fertilization and development, suggest that the functions involve limited activation of the proteases. This scenario is reminiscent of the blood-clotting proteases as well as the extracellular proteases involved in development, such as the *Drosophila* peritelline space proteases involved in embryonic axis formation mentioned above (18). In general, these proteases are uniformly present as proenzymes, and the activating signals are localized and brief, i.e., the proteases are deactivated by inhibitors or through the formation of complexes with other proteins or are immediately degraded (26). As for the timing of ovochymase/ovotryptase function, we know that, for at least ovochymase, a small fraction of active protease is released by dejellied eggs on activation, which suggests a role for the protease shortly after fertilization. However, it is also possible that this observed release is an artifact of the chemical dejellifying process, rendering a portion of the protease population susceptible to cleavage once the dejellied egg is activated. However, the fact that ovochymase and the ovotryptases are secreted to the egg surface during oogenesis argues for a role early in the developmental process (e.g., before fertilization, during the fertilization process, or soon thereafter), because extracellular proteases involved later in development tend to be expressed and secreted by the embryo after fertilization.

This work was supported in part by U.S. Public Health Service Grant HD-04906, National Science Foundation Grants MCB-95-07087 and IBM 9722667, and a University of California, Davis, Bridge Program grant to L.L.L.