Molecular Cloning and Gibberellin-induced Expression of Multiple Cysteine Proteinases of Rice Seeds (Oryzains)*

Hirohito Watanabe, Kelko Abe, Yasufumi Emori†, Hiroshi Hosoyama, and Soichi Arai

From the Department of Agricultural Chemistry, Faculty of Agriculture and the †Department of Biophysics and Biochemistry, Faculty of Science, the University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

We screened a cDNA library of germinating rice seeds with a cDNA for aleurain (cysteine proteinase from barley) and obtained three distinct types of cDNA clones encoding three species of cysteine proteinases (oryzains α, β, and γ). The deduced amino acid sequences are distinct in part, but, on the whole, are similar to one another. The three sequences all contain the catalytic triad Cys-His-Asn (papain numbering). The three oryzains are similar to one another and also to other known cysteine proteinases such as papain and cathepsin H with respect to the sequences around the active site residues and the COOH-terminal Trp-rich region. Amino acid sequence comparison revealed that oryzains α and β are similar not only to each other (70% similarity) but also to actinidin and papain (about 50%), whereas oryzain γ was rather similar to aleurain (85%) and cathepsin H (60%).

Northern blot analysis revealed that the mRNAs for the three oryzains are expressed only in seeds, not in shoots or roots, and show different expression profiles during germination and when the seeds are treated with gibberellic acid. Oryzains α and γ are expressed continuously during germination with a maximum expression 5 days from the start of germination, but are present in neither ripening nor ripened seeds. On the other hand, oryzain β is expressed not only during germination, but also in ripened seeds before germination. It was noted that the expression of the three oryzain mRNAs is enhanced in different manners by gibberellic acid but is not enhanced by other plant hormones such as auxin. The induction of oryzain β mRNA is transient, reaching a maximum in 4 h from the addition of gibberellic acid and diminishing rapidly thereafter, while the induction of oryzain α and γ mRNAs continues over 5 days. Thus, multiple systems involving cysteine proteinases must be differentially involved in the germination process, probably under hormonal control.

Cysteine proteinases (EC 3.4.22.) are widely distributed in animals (1–6), plants (7–12) and microorganisms (13) where they are involved in many intracellular and extracellular processes of physiological importance. The activities of cysteine proteinases are controlled by protein inhibitors called cystatins. A number of cysteine proteinases and cystatins are already well characterized in animals (14), and refined enzymatic analyses have been carried out on multiple systems involving cysteine proteinases such as cathepsin H, L, and B, and cystatins such as stefin A, B, and cystatin β. In plants, however, aleurain which occurs in the aleurone of barley (12) is the only seed cysteine proteinase that has been well defined at the molecular level.

We have found a cysteine proteinase in germinating rice seeds whose expression reaches a maximum 5 to 8 days after the start of germination (15) and is also enhanced by gibberellic acid (GA3)1 (16). As for plant cystatins, we have found two types of proteinaceous cysteine proteinase inhibitors in rice seeds. These are oryzacystatin I (17–19) and oryzacystatin II (20), which are expressed during the ripening stage and stored in mature seeds. Oryzacystatin I and II exhibit potent inhibitory activities toward papain and cathepsin H, respectively. It is possible that the activities of the cysteine proteinases in rice seeds are inhibited by either or both oryzacystatins. To understand the physiology of cysteine proteinases and cystatins in plant seeds, their identification and characterization at the molecular level is of primary importance. We have also found that a major storage protein in rice seeds (glutelin) is efficiently degraded in vitro by a cysteine proteinase existing in germinating seeds (15). However, refined molecular analyses on this cysteine proteinase and other proteinases have not been performed.

Here we report on the isolation and characterization of cDNA clones for rice seed cysteine proteinases (oryzains) to define their molecular nature. We also report that the synthesis of mRNAs encoding oryzains is induced during germination by GA3.

EXPERIMENTAL PROCEDURES

Materials—DNA polymerase I from Escherichia coli, its Klenow fragment, bacterial alkaline phosphatase, T4 polynucleotide kinase, and T4 DNA ligase were purchased from Takara Shuzo Co. Restriction enzymes were the products of Takara Shuzo Co. and Toyobo Co. The oligo(dT)12 packed column was a product of Amersham. A multiprime DNA labeling kit (“Rapid Hybridization System-Multiprime”), [γ-32P]ATP (3000 Ci/mmol), and [α-32P]dCTP (3009 Ci/mm) were purchased from Amersham. The λgt10 phage vector arms and λ phage packaging kit (“Giga pack gold”) were products of Stratagene Cloning Systems. The cDNA synthesis kit was purchased from Pharmacia LKB Biotechnology Inc.

Rice Seeds—Cultivar Nipponbare of the rice Oryza sativa L. japonica was used. The seeds were harvested on days 3, 5, and 8 after germination and stored at −80 °C until used.

Determination of Partial Amino Acid Sequences—A cysteine proteinase was purified from rice seeds as described previously (15) and digested with lysylendopeptidase (Wako) (21). Peptides were fractionated on a Vydoc ODS column (Senshu Kagaku). Partial amino acid sequences were analyzed with a protein sequenator (Model 470A,

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: GA3, gibberellic acid; SDS, sodium dodecyl sulfate.
Multiple Cysteine Proteinases of Rice Seeds

Applied Biosystems) to give three sequences, Asp-Glu-Arg-Xaa-Asp-Val-Asn-Arg-Lys, Xaa-Gly-Xaa-Ile-Asp-Thr-Glu-Xaa-Asp-Tyr-Pro-Tyr-Lys, and Leu-Pro-Glu-Xaa-Asp-Arg-Xaa-Lys. We also analyzed the NH2-terminal sequence of the purified protein and obtained the sequence Leu-Pro-Glu-Xaa-Asp-Arg-Xaa-Lys, which corresponds to the third sequence described above.

Construction of cDNA Library and Isolation of cDNA Clones for Oryzains—Total RNA was extracted by the phenol-SDS method (22) from rice seeds harvested on days 3, 5, and 8 after germination. Poly(A)+ RNA was purified by oligo(dT)-cellulose column chromatography (23). Double-stranded cDNA was synthesized using a cDNA synthesis kit (Amersham) (24) and used for construction of a Agt10 cDNA library. Recombinant phages from the cDNA library described above were propagated on E. coli C600Hfti lawn. Plaques were transferred onto Nylon filters (HybondTM-N, Amersham), prehybridized for 16 h at 65°C as described previously (25), and hybridized for 24 h at 65°C with an aleurain cDNA fragment (12) labeled by the multiprime system (Amersham) as a probe. The filters were finally washed in 2× SSC containing 0.1% SDS at 65°C (26).

Nucleotide Sequencing—cDNA inserts excised by EcoRI digestion of recombinant phage DNA were subcloned into plasmid vector pUC18 (27) and sequenced by a modified dideoxy method (28).

RNA Blot Hybridization—Total RNA was extracted from rice seeds at various stages of germination as described above. The RNA sample (10–15 μg in total RNA) was denatured and electrophoresed in a formaldehyde-containing agarose gel (26). After electrophoresis, the RNA was transferred onto a Nylon membrane (HybondTM-N, Amersham) and hybridized with 32P-labeled cDNAs at 65°C in rapid hybridization solution (Amersham). The filter was finally washed in 0.1× SSC containing 0.1% SDS at 65°C.

RESULTS

Isolation and Identification of cDNA Clones for Cysteine Proteinases—Among 1.5×106 independent plaques from the Agt10 cDNA library, approximately 20 clones were hybridized with aleurain cDNA. These clones were divided into three types based on restriction mapping and cross-hybridization analyses (29). Three clones, λOZA022, λOZB102, and λOZC511, containing the longest inserts of these three types, were selected and sequenced by the strategy shown in Fig. 1. Since the three clones encoded different species of putative cysteine proteinase (to be described in detail later), the encoded proteins were termed oryzain α, β, and γ in correspondence with clones λOZA022, λOZB102, and λOZC511, respectively.

The nucleotide sequence of the λOZA022 cDNA insert encoding oryzain α contains a single, long open reading frame of 1374 nucleotides (458 amino acid residues) preceded by an in-frame termination codon (position -12 to -10 in Fig. 2A) and followed by a poly(T) tail. The deduced amino acid sequence contains all the partial amino acid sequences already found in peptides derived from the purified cysteine proteinase by partial digestion (see “Experimental Procedures”), indicating that oryzain α is the cysteine proteinase previously identified by protein chemistry (15). The nucleotide sequence of λOZB102, encoding oryzain β, contains a single open reading frame of 1413 nucleotides (471 amino acids) (Fig. 2B). That of λOZC511, encoding oryzain γ, contains a single open reading frame of 1086 nucleotides (362 amino acids) (Fig. 2C).

The initiation methionine codons for oryzains β and γ were assigned for the following reasons: 1) the nucleotide sequences around the putative initiation codons are consistent with those (AAAAAAGGCGT) proposed for plant mRNAs (30, 31); 2) hydropathy profiles (Fig. 3) and sequence similarity strongly suggest the existence of signal and pro-sequences which follow the putative initiation codons; and 3) the lengths of the mRNAs (Fig. 5) approximate the total length of the cDNA inserts.

Overall Structure and Sequence Homology of the Three Oryzains—The three oryzains are thought to be cleaved by post-translational modification. The signal sequences were estimated based on the general characteristic that a signal sequence contains a charged residue within the first five amino acids followed by a core of at least 9 hydrophobic residues (32). By this general rule, as well as from the hydropathy profiles (Fig. 3), we assigned possible signal sequences at Met1-Ala3, Met1-Ala3, and Met1-Ala4 for oryzains α, β, and γ, respectively.

The amino acid sequences of oryzains α, β, and γ were aligned for maximal homology (Fig. 4). The NH2 terminus of mature oryzain α was assigned as Leu126 because the NH2-terminal sequence analysis revealed the sequence Leu-Pro-Glu-Xaa-Val-Asp-Trp (see “Experimental Procedures”), which corresponds exactly to the deduced sequence of residues 129 to 135. The NH2 termini of oryzains β (Leu146) and γ (Leu143) were assigned from sequence similarity to the sequence around Leu129 of oryzain α. Thus, the prepropeptides of oryzains α, β, and γ are predicted to be the sequences from Met1-Ala126, from Met1 to Glu139, and from Met1 to Ala144, respectively. With respect to the homology between the prosequences, oryzains α and β show 53% similarity. However, the similarity between the prosequences of oryzains and other cysteine proteinase precursors is relatively low (20–42%). Oryzains α and β contain sequences similar to the COOH-terminal extension peptide sequence found in actinidin (11) and, therefore, are thought to have had telo-sequences; oryzain γ per se does not contain a telo-sequence. From sequence similarity to actinidin, the COOH termini of mature oryzains α and β should be Glu166 and Ala301, respectively. The molecular weight of the mature form of oryzain α is thus calculated to be 23,782, which agrees well with the value of 23,000 estimated from SDS-gel electrophoresis (15).

In the mature enzymes, oryzains α, β, and γ have catalytic triads at Cys168-His200-Asn209, Cys164-His301-Asn321, and Cys169-His306-Asn299, respectively, which are conserved among various cysteine proteinases (3, 4, 7, 11, 12). Oryzain α has 62% similarity to oryzain β, greater than oryzain γ (43%). Oryzains α and β are closer to actinidin and papain than to aleurain and cathepsin H (Table I), while oryzain γ has less similarity to actinidin and papain than to aleurain and cathepsin H. The sequences around the catalytic triads are also similar to those in other known cysteine proteinases (data not shown).

In the COOH-terminal regions of the mature forms, the three sequences contain Trp residues (for example, residues 305, 311, and 315 in oryzain α). This feature is common to other cysteine proteinases which contain 2 to 5 Trp residues in the corresponding region (33). Therefore, the mature oryzains would be expected to have proteolytic activity. In actuality,
we have found oryzain α to have potent proteolytic activity against casein and glutelin, a major storage protein of rice seeds (15).

Potential N-glycosylation sites exist at Asn445 in oryzain α, at Asn400 and Asn486 in oryzain β, and at Asn278 and Asn258 in oryzain γ. In the case of oryzain γ, one of the two potential glycosylation sites is of the form Asn-Ile-Thr, as is the case for cathepsin H (3) and aleurain (12).

The sequences of the mature forms of oryzains α, β, and γ are comparable to one another and to other known cysteine proteinases (Table I). The sequence similarity is significant for all comparisons and ranges from 32% to 85%. High similarity is observed in comparing oryzain γ and aleurain, suggesting that oryzain γ is a rice counterpart of barley aleurain. Relatively high sequence similarity is also observed between oryzain α and β, between oryzain γ and aleurain, between aleurain and cathepsin H, between oryzain γ and cathepsin H, etc. From these comparative studies, a putative pedigree for cysteine proteinases can be drawn as outlined in Fig. 7. This indicates that the seeds contain at least three cysteine proteinases extending over two subfamilies. Thus, in rice as a species of plant, multiple cysteine proteinases exist and...
probably have important functions. It is also suggested that divergence and evolution of cysteine proteinases occurred at an early stage in the history of evolution.

Expression of Oryzain mRNAs during the Germination of Rice Seeds—Each cDNA insert was used as a hybridization probe to evaluate the expression of oryzain mRNAs. Although the three cDNA sequences are similar, they did not cross-hybridize under high stringency conditions (data not shown). As shown in Fig. 5, single bands were detected at approximately 2.0, 2.0, and 1.5 kilobase for oryzain α, β, and γ mRNAs, respectively, at all stages of germination. However, the three mRNAs were not detected in shoots or roots (Fig. 5, lanes 9 and 10 for oryzain β; data not shown for oryzains α and γ). The mRNA for oryzain α was observed to peak at 5 days after germination and was not detected in mature seeds. This pattern is consistent with the time course profile observed for oryzain α activity (15), indicating that the activity of oryzain α may be controlled at the level of transcription. On the other hand, the mRNA for oryzain β is detected in mature seeds (Fig. 5, lane 5) and reaches a maximal level 9 days after germination. The profile for oryzain γ mRNA is similar to that of oryzain α; the mRNA reaches a maximal level 5 days after germination. These mRNAs are not expressed in the endosperm but in the aleurone or germ (data not shown).

Next, we examined whether the expression of oryzains is induced by GA, because previous studies on oryzain α (16) and aleurain (12) showed they are GA-inducible. The mRNAs for oryzains α and γ were induced 24 h after GA treatment (Fig. 6, I, lanes 1 and 11) and was constantly expressed for 5 days (lanes 1–4 and 11–14). On the other hand, expression of oryzain β mRNA was strongly induced 4 h after the addition of GA (Fig. 6, I, lane 6) and then returned to a nearly basal level in 24 h (lane 9). As shown in Fig. 6, II, other plant hormones had no significant effect on expression.

**DISCUSSION**

In the present study, we isolated three types of cDNA clones encoding cysteine proteinases, one of which, oryzain α, corresponds to the protein which we had previously purified from germinating rice seeds (15). The amino acid sequences of the three oryzains share the general characteristics common to cysteine proteinases of animal and plant origin. In addition, they have signal and pro-sequences suggesting that they are secretory enzymes and may exist as precursors. Our preliminary observation is that the mRNAs for oryzains exist in the aleurone or germ, not in the endosperm, although oryzain α exists in germinating endosperm (15). Thus, rice seeds possibly have a transport system for oryzains.

Physiological roles for multiple oryzains are suggested by the following two interesting observations. First, the three mRNAs for oryzains α, β, and γ all responded to the addition of GA, although the induction of oryzain β mRNA was transient (Fig. 6, I). Oryzains α and β, two closely related oryzains similar to actinidin and papain, are regulated differently by GA (Fig. 6). Along with the observation that oryzain β is expressed in both ripening and germinating rice seeds, oryzain β may be involved in rapid, but transient protein degradation induced by GA, and/or in housekeeping events. Oryzain α may be involved in relatively chronic events during the germination process. Based on its pattern of expression (Fig. 5), oryzain γ should be similar to oryzain α; however, since it lacks a telo-sequence (Fig. 4), differences in regulatory mechanisms of sorting and/or activation processes may reflect different physiological roles. Second, the action of oryzains α, β, and γ should also be considered with respect to their regulation by cysteine proteinase inhibitors (cystatins). Two cysteine proteinase inhibitors, oryzacystatin I (19) and II (20), exist in rice seeds before germination. These two endogenous cystatins have highly similar sequences (20), but show different specificities; oryzacystatin I inhibits papain more effectively than cathepsin H, whereas oryzacystatin II inhibits cathepsin H more effectively than papain (20). Considering this result in relation to the fact that oryzains α and β are highly similar to papain, it is possible that these two proteinases can be effectively inhibited by oryzacystatin I, whereas oryzain γ, which is similar to cathepsin H, may be inhibited by oryzacystatin II. Actually, we have observed that the hydrolytic activity of oryzain α for a major rice storage protein (glutelin) is stoichiometrically inhibited by oryzacystatin I in vitro (15).

---

**Table I**

| Oryzain α | Oryzain β | Oryzain γ | Actinidin | Papain | Aleurain | Cathepsin H | Cathepsin L | Cathepsin B |
|-----------|-----------|-----------|-----------|--------|----------|-------------|-------------|-------------|
| Oryzain α | 70        | 48        | 45        | 56     | 53       | 43          | 36          | 45          |
| Oryzain β | 48        | 48        | 40        | 54     | 52       | 44          | 40          | 49          |
| Oryzain γ | 45        | 40        | 47        | 40     | 32       | 40          | 35          | 47          |
| Actinidin | 40        | 40        | 40        | 40     | 32       | 40          | 35          | 47          |
| Papain    | 49        | 49        | 49        | 49     | 49       | 49          | 49          | 49          |
| Aleurain  | 43        | 44        | 45        | 44     | 44       | 44          | 44          | 44          |
| Cathepsin H| 36        | 41        | 41        | 41     | 41       | 41          | 41          | 41          |
| Cathepsin L| 36        | 40        | 40        | 40     | 40       | 40          | 40          | 40          |
| Cathepsin B| 45        | 45        | 45        | 45     | 45       | 45          | 45          | 45          |

---

H. Watanabe, K. Abe, and S. Arai, manuscript in preparation.

---

**Fig. 4. Comparison of the amino acid sequences of oryzain α, β, and γ.** The sequences are aligned for maximal matching. Identical amino acid residues are boxed. Arrows indicate cleavage sites in post-translational processing.
nation (20). Thus, oryzacystatins do not remain at the time when oryzains begin to appear. The existence of oryzacystatins before germination may be important in preventing the degradation of seed storage proteins by oryzains or by exogenous proteinases possibly brought by destructive insects. Since oryzacystatins disappear after germination, oryzains may catalyze the proteolysis of the storage proteins.

The physiological functions of oryzains will be disclosed by histochemical studies to define their compartmentalization in the seed cell. A similar histochemical study with oryzacystatins will also be necessary to elucidate the mechanism of oryzain-oryzacystatin interaction in vivo.

Acknowledgments—We thank Dr. J. C. Rogers, Washington University School of Medicine, for providing us with an aleurain cDNA clone and Drs. K. Suzuki and H. Kawasaki, the Tokyo Metropolitan Institute of Medical Science, for pertinent discussion.

REFERENCES

1. Chan, S. J., Segundo, B. S., McCormic, M. B., and Steiner, D. F. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 7721–7725
2. Takio, K., Towatari, T., Katsumina, N., Teller, D. C., and Titan, K. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3666–3670
3. Ishidoh, K., Imajoh, S., Emori, Y., Ohno, S., Kawasaki, H., Minami, Y., Kominami, E., Katunuma, N., and Suzuki, K. (1987) FEBS Lett. 226, 33–37
4. Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N., and Suzuki, K. (1987) FEBS Lett. 223, 69–73
5. Mason, R. W., Walker, J. E., and Northrop, F. D. (1986) Biochem. J. 240, 373–377
6. Ishidoh, K., Kominami, E., Suzuki, K., and Katunuma, N. (1989) FEBS Lett. 259, 71–74
7. Cohen, L. W., Coghan, V. M., and Dihel, L. C. (1986) Gene (Amst.) 48, 219–227
8. Jacquet, A., Kleinschmidt, T., Schnek, A. G., Loose, Y., and Braunitzer, G. (1989) Biol. Chem. Hoppe-Seyler 370, 425–434
9. Lindwall, G., and Cole, R. D. (1984) J. Biol. Chem. 259, 12241–12245
10. Carne, A., and Moore, X. (1978) Biochem. J. 173, 73–83
11. Podivinsky, E., Forster, R. L. S., and Gardner, R. C. (1989) Nucleic Acids Res. 17, 8363
12. Rogers, J. C., Dean, D., and Heck, G. R. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 6512–6516
13. Williams, J. G., North, M. J., and Mahbubani, H. (1985) EMBO J. 4, 999–1006
14. Katsumuna, N., and Kominami, E. (1986) in Cysteine Proteinases and Their Inhibitors (Turk., V., ed) pp. 219–227, Walter de Gruyter, Berlin
15. Abe, K., Kondo, H., and Arai, S. (1987) Agric. Biol. Chem. 51, 2763–2768
16. Arai, S., Hosoyama, H., and Abe, K. (1987) Agric. Biol. Chem. 52, 2957–2959
17. Abe, K., and Arai, S. (1985) Agric. Biol. Chem. 49, 3349–3350

Another important observation concerning oryzacystatin I and II is that their mRNAs are synthesized during the seed-ripening stage and large amounts of the cystatins accumulate in seeds before germination but degrade rapidly after germi-
Multiple Cysteine Proteinases of Rice Seeds

18. Abe, K., Kondo, H., and Arai, S. (1987) *Agric. Biol. Chem.* **51**, 1509–1514
19. Abe, K., Emori, Y., Kondo, H., Suzuki, K., and Arai, S. (1987) *J. Biol. Chem.* **262**, 16793–16797
20. Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y., and Arai, S. (1990) *J. Biol. Chem.* **265**, 15832–15837
21. Tsunasawa, S., Sugihara, A., Masaki, T., Sakiyama, F., Takeda, Y., Mawatani, T., and Narita, K. (1987) *J. Biochem. (Tokyo)* **101**, 111–121
22. Brawerman, G., Mandecki, J., and Lee, S. Y. (1972) *Biochemistry* **11**, 637–641
23. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1408–1418
24. Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25**, 263–269
25. Abe, K., Emori, Y., Kondo, H., Arai, S., and Suzuki, K. (1988) *J. Biol. Chem.* **263**, 7655–7659
26. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
27. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580
28. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238
29. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
30. Joshi, C. P. (1987) *Nucleic Acids Res.* **15**, 6643–6653
31. Kondo, H., Emori, Y., Abe, K., Suzuki, K., and Arai, S. (1989) *Gene (Amst.)* **81**, 259–265
32. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164
33. Suzuki, K., Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., and Kisaragi, M. (1985) in *The Biological Role of Proteinases and Their Inhibitors in Skin* (Ogawa, H., Lazarus, G. S., and Hope-Havu, V. K., eds) pp. 111–120, University of Tokyo Press, Tokyo