Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6512–6516, October 1985
Biochemistry

Aleurain: A barley thiol protease closely related to mammalian cathepsin H

(gibberellic acid/abscisic acid/secretion)

JOHN C. ROGERS*, DUFF DEAN, AND GREGORY R. HECK

Division of Hematology-Oncology, Departments of Internal Medicine and Biology, Washington University School of Medicine, St. Louis, MO 63110

Communicated by J. E. Varner, June 13, 1985

ABSTRACT **We have isolated and sequenced a 1400-base-pair cDNA derived from gibberellic acid-treated aleurone cell mRNA. This sequence contains an open reading frame that would code for a protein of 361 amino acids. The carboxyl-terminal two-thirds of the predicted amino acid sequence is closely related to that of the rat lysosomal thiol protease cathepsin H; the initial 143 amino acids may code for a secretory peptide plus a prosegment. The expression of this aleurone thiol protease mRNA is unusual in that, in aleurone cells, its abundance is regulated by the plant hormones gibberellic acid and abscisic acid, but it is also expressed at high levels in leaf and root tissue. This protease may represent the equivalent of a plant lysosomal thiol protease.**

Barley aleurone cells form a tough layer surrounding the endosperm of the grain. In response to gibberellins secreted by the embryo, these cells secrete large quantities of several different hydrolases that digest the storage products in the endosperm for use by the seedling (1). These aleurone cells have been widely studied as models for mechanisms of hormone-mediated gene activation in plants (1), and α-amylase, the most abundant hydrolase produced, has received the most attention.

Less is known about other important enzymes secreted from aleurone cells. A gibberellic acid-induced endopeptidase was identified by Jacobsen and Varner (2), who demonstrated a time-course of induction and hormone-dose–response curve for this protease activity that were indistinguishable from those for α-amylase. Other workers identified constitutively expressed carboxypeptidase activities in barley aleurone layers (3, 4), but only the secretion of these enzymes was increased by gibberellic acid. Hammerton and Ho (5) further characterized the gibberellic acid-induced endopeptidase; this activity was destroyed by known thiol-protease inhibitors but not by inhibitors of serine proteases, and its estimated molecular mass was 37 kDa as assessed by polyacrylamide gel electrophoresis under denaturing conditions (5). They concluded that this enzyme represents the major endoprotease synthesized in response to gibberellic acid.

This communication provides more information regarding the primary structure and expression of a gibberellic acid-induced thiol protease. We have characterized a full-length cDNA clone from barley aleurone cells representing a mRNA that increases 7-fold in response to gibberellic acid. The sequence of its protein, as predicted from the nucleotide sequence, is very similar to that of the mammalian thiol protease cathepsin H (6) and less similar (although still related) to the sequences of the two plant thiol proteases papain (7) and actinidin (8). The predicted size of the mature protein is similar to that estimated for the thiol protease studied by Hammerton and Ho (5), and we suggest that our cDNA clone and that protease activity represent the same gene product.

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

MATERIALS AND METHODS

The purification of RNA and construction of the cDNA library from gibberellic acid-stimulated aleurone cells and its screening for abundant, gibberellic acid-induced clones has been described (9). Conditions for restriction enzyme mapping, DNA sequencing by the chemical-degradation method (10), nondenaturing and formaldehyde/agarose gel electrophoresis of RNA and restriction digests of genomic DNA, transfer to nitrocellulose, and hybridization have been described (11, 12). After hybridization, blots were washed with three changes of 15 mM NaCl/1.5 mM sodium citrate/0.1% $NaDodSO_4$ for a total of 1 hr at 55°C.

The predicted amino acid sequence derived from the only long open reading frame on the proper strand of the cDNA was used in a search of the sequences compiled in the National Biomedical Research Foundation Protein Sequence Databank carried in the Washington University Medical School DEC VAX-11 computer. Estimates of similarity and optimal alignments of related sequences were accomplished with the programs of Dayhoff *et al.* (13). The program of Zuker and Stiegler (14) was utilized to predict the structure of the mRNA at its 5′ end.

RESULTS

Sequence Analysis and Identification of the Thiol Protease cDNA. When the original cDNA library was screened (9), one clone was identified that (*i*) hybridized strongly to ^{32}P-labeled cDNA synthesized from gibberellic acid-stimulated aleurone cell mRNA, (*ii*) contained an insert of 1.5 kilobase pairs (kbp), and (*iii*) did not cross-hybridize to either type of amylase cDNA (ref. 9 and data not shown). This clone was characterized extensively because expression of the mRNA to which it corresponded was regulated in an interesting manner (see below) and because the cDNA insert appeared to be full- or nearly full-length when compared to the size of its hybridizable mRNA. Accordingly, the complete nucleotide sequences of both strands of the cDNA were determined, and all restriction sites used for end-labeling were bridged by separate sequencing reactions (Fig. 1).

The nucleotide sequence and the sequence of the protein predicted from the only long open reading frame beginning at the first ATG are presented in Fig. 2. The insert contains 1403 bp, excluding the poly(A) tail, and includes 36 bp of 5′ untranslated region. The predicted protein product contains 361 amino acids.

Initial homology searches of the Protein Sequence Databank utilized sequence data obtained 3′ from the second

Abbreviation: bp, base pair(s).

*To whom reprint requests should be addressed.

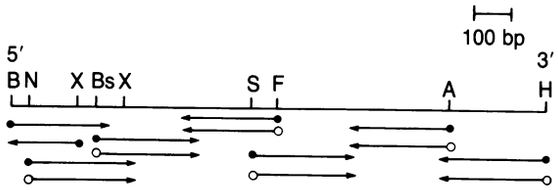

FIG. 1. Sequencing strategy. The cDNA insert was sequenced as described in *Materials and Methods*. Circles represent sites for end-labeling on 5' (●) or 3' (○) termini, respectively. Arrows indicate the direction and extent of each sequence determination. For sequencing from the *Nae* I (N) and *Stu* I (S) sites, subclones were constructed by inserting *Bam*HI linkers. Other sites used were *Bam*HI (B) and*Hin*dIII (H) in the vector polylinker and *Xho* I (X), *Bss*HII (Bs), *Hin*fI (F), and *Acc* I in the insert.

Xho I site and identified several thiol proteases. In Fig. 3, the amino acid sequences of these proteins are compared to that of the protein predicted from the barley cDNA. Alignments are derived from analyses using the program ALIGN (13) and gaps are assigned according to those results. It can be seen that the barley protein (designated aleurain, see below) contains 143 amino acids amino-terminal to the region homologous to the other proteases. Within the region of homology, rat liver cathepsin H and aleurain share 137 identical out of 218 (63%) possible matches, with only two single-residue gaps necessary for optimal alignment. Large continuous blocks of identical residues are encountered surrounding the cathepsin H active-site cysteine (circled and corresponding to aleurain residue 165) and histidine (circled, residue 307). The sequence corresponding to a known cathepsin H asparagine-glycosylation site (boxed, residue 256) is conserved; in addition, aleurain has another potential glycosylation signal, Asn-Ile-Ser (indicated by the line above residues 188–190). In contrast to the comparison with the mammalian protease, aleurain is substantially less similar to the two plant thiol proteases, with 81/205 matches (39%) requiring 10 gaps and 94/213 matches (44%) requiring 7 gaps for papain and actinidin, respectively.

The presence of 143 amino acids extra at the amino-terminal portion of the aleurone sequence is surprising. Since the gibberellic acid-induced thiol protease is a secreted protein (2, 5), we would expect to find a secretory peptide region that would be cleaved from the mature protein. The first 22 or 23 residues fit very well with the general characteristics of sequences found in such peptides, with a charged residue (arginine) at position 5 followed by 10 hydrophobic residues (15); the presumed cleavage site would be predicted to follow either alanine (position 22) or serine (position 23) (15). This leaves 120 residues unaccounted for. In the case of mammalian lysosomal proteases, it is clear that the mature protein is formed after acid-induced cleavage of a propeptide.

```
       10        20        30        40        50        60        70        80        90       100
AGTAAAATACTGCTACCCATCCGCCGGCGAAACGAAATGGCCCACGCCCGCGTCCTCCTCCTGGCGCTCGCCGTCCTGGCCACGGCCGCCGTCGCCGTCG
                                      M  A  H  A  R  V  L  L  L  A  L  A  V  L  A  T  A  A  V  A  V
                          Nae I
      110       120       130       140       150       160       170       180       190       200
CCTCCTCCTCCTCCTTCGCCGACTCCAACCCGATCCGGCCCGTCACCGACCGCGCCGCCTCGACGCTCGAGTCCGCCGTCCTCGGCGCGCTCGGCCGCAC
A  S  S  S  S  F  A  D  S  N  P  I  R  P  V  T  D  R  A  A  S  T  L  E  S  A  V  L  G  A  L  G  R  T
                                                                  Xho I
      210       220       230       240       250       260       270       280       290       300
CCGCCACGCCCTCCGCTTCGCGCGCTTCGCCGTCAGGTACGGCAAGAGCTACGAGAGCGCGGCGGAGGTGCGGAGGCGGTTCCGGATCTTCTCCGAGAGC
  R  H  A  L  R  F  A  R  F  A  V  R  Y  G  K  S  Y  E  S  A  A  E  V  R  R  R  F  R  I  F  S  E  S
                     BssH II
      310       320       330       340       350       360       370       380       390       400
CTCGAGGAGGTGCGCTCCACCAACCGGAAGGGCCTCCCCTACCGCCTCGGCATCAACCGCTTCTCGGACATGAGCTGGGAGGAGTTCCAGGCGACCCGTC
 L  E  E  V  R  S  T  N  R  K  G  L  P  Y  R  L  G  I  N  R  F  S  D  M  S  W  E  E  F  Q  A  T  R
Xho I
      410       420       430       440       450       460       470       480       490       500
TCGGCGCGGCGCAGACCTGCTCGGCGACGCTCGCCGGCAACCACCTCATGCGGGACGCCGCCGCGCTCCCGGAGACCAAAGACTGGAGGGAGGATGGGAT
L  G  A  A  Q  T  C  S  A  T  L  A  G  N  H  L  M  R  D  A  A  A  L  P  E  T  K  D  W  R  E  D  G  I

      510       520       530       540       550       560       570       580       590       600
CGTCAGCCCCGTCAAAAACCAGGCCCATTGCGGCTCCTGCTGGACGTTCAGCACTACTGGCGCACTTGAGGCAGCATATACTCAGGCCACTGGAAAGAAC
  V  S  P  V  K  N  Q  A  H  C  G  S  C  W  T  F  S  T  T  G  A  L  E  A  A  Y  T  Q  A  T  G  K  N

      610       620       630       640       650       660       670       680       690       700
ATCTCTCTTTCTGAGCAACAGCTGGTTGACTGTGCTGGTGGGTTCAATAACTTTGGATGCAACGGAGGCCTTCCATCCCAGGCGTTTGAGTACCAATACA
 I  S  L  S  E  Q  Q  L  V  D  C  A  G  G  F  N  N  F  G  C  N  G  G  L  P  S  Q  A  F  E  Y  Q  Y
                Pvu II                                                 Stu I
      710       720       730       740       750       760       770       780       790       800
ATGGCGGGATCGACACCGAGGAGTCCTACCCTTACAAGGGTGTCAATGGCGTCTGCCATTACAAGGCTGAAAATGCTGCGGTTCAGGTTTTGGACTCAGT
N  G  G  I  D  T  E  E  S  Y  P  Y  K  G  V  N  G  V  C  H  Y  K  A  E  N  A  A  V  Q  V  L  D  S  V
                   Hinf I
      810       820       830       840       850       860       870       880       890       900
AAACATCACACTGAATGCCGAGGATGAACTGAAGAACGCCGTCGGGTTGGTTCGCCCAGTGAGTGTGGCCTTTCAGGTGATTGACGGTTTCAGGCAGTAC
  N  I  T  L  N  A  E  D  E  L  K  N  A  V  G  L  V  R  P  V  S  V  A  F  Q  V  I  D  G  F  R  Q  Y

      910       920       930       940       950       960       970       980       990      1000
AAGAGCGGAGTTTACACCAGCGACCATTGTGGCACTACCCCTGATGACGTGAACCACGCCGTCTTGGCAGTCGGCTATGGTGTGGAGAACGGCGTTCCGT
 K  S  G  V  Y  T  S  D  H  C  G  T  T  P  D  D  V  N  H  A  V  L  A  V  G  Y  G  V  E  N  G  V  P

     1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
ATTGGCTCACCAAGAACTCGTGGGGCGCCGACTGGGGTGACAATGGTTACTTCAAGATGGAAATGGGCAAGAACATGTGTGCTATTGCTACCTGCGCCTC
Y  W  L  T  K  N  S  W  G  A  D  W  G  D  N  G  Y  F  K  M  E  M  G  K  N  M  C  A  I  A  T  C  A  S

     1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
CTACCCTGTCGTGGCGGCATGAAGAAGCGCCAAATCGGCGAGTACAGTCTCGATTCGCCAGACATGAATAAGGTGTTTTCAGCCCGAACAAGAACCTGAA
  Y  P  V  V  A  A

     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
GATTGGGGTATACAGGAACTGAAACTTTGCTGTGAATACAAACCACAAGGACTACGGTTGCAGTGAGCCCCGTGCGTGTACTCTTCCTCTGTATGGTTGT
       Acc I

     1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
ATGGCTATATATGTAGGTGATCCAGCAATTTATTAGGGTTTCTCATGTATGTCTACGACATTTTATTTCCGGGACAAATAAAATGGCCCATTCATAGTTCACC poly(A)
```

FIG. 2. Nucleotide and derived amino acid sequences of the cDNA insert. The amino acid sequence (standard one-letter abbreviations) begins with the first ATG codon in the only long open reading frame. Positions of restriction enzyme sites used for sequencing and subcloning are indicated.

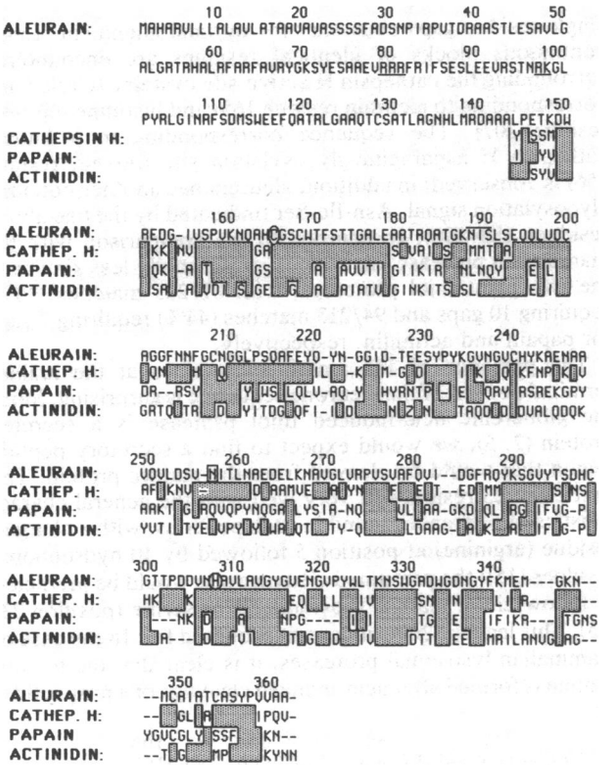

FIG. 3. Comparison of the predicted barley protein sequence and the sequences of three related thiol proteases. The barley protein (aleurain) sequence was aligned with each of the other sequences with the ALIGN program (13), using the Mutation Data Matrix plus a bias factor of 6 and a break penalty of 8. Residues that are identical to residues in the aleurain sequence are indicated by stippled boxes, and gaps are represented by dashes. Numbers refer to the aleurain sequence. The two putative active-site residues are circled. An asparagine residue in the aleurain sequence that corresponds to the position of the N-glycosylated asparagine residue in cathepsin H is boxed. A line over residues 188–190 designates another potential aleurain glycosylation site. Underlined are two locations where basic residues are adjacent.

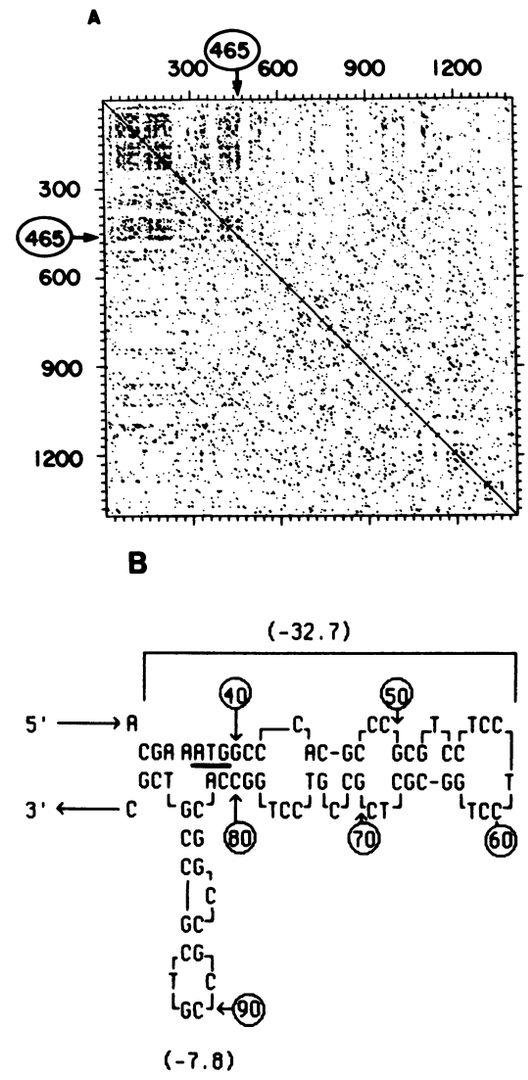

FIG. 4. (*A*) DOTMATRIX (13) analysis of the aleurain cDNA sequence. The cDNA nucleotide sequence was compared to itself, using a window size of 8 and a minimum score of 6. Numbers refer to the distance (in bp) from the 5′ end. The sequence encoding the region of the protein homologous to the thiol proteases begins at position 465 (circled). (*B*) Theoretical secondary structure at the 5′ end of the aleurain mRNA. RNA-folding analysis (14) was performed, using the first 150 nucleotides of the cDNA sequence; the sequence is presented in DNA form (i.e., T = U). Circled numbers refer to the distance from the 5′ end. The predicted free energy values (kcal/mol) for the parts of the hairpin structure are given in parentheses. The initiation codon is underlined.

For example, in the case of the human aspartyl protease cathepsin D, the secretory peptide comprises 20 residues, and the propeptide another 44 residues (16). However, the rat lysosomal thiol protease cathepsin B precursor is estimated to be 43 kDa, whereas the mature protein is about 31 kDa (17). These data suggest that cathepsin B might have a prosegment of about 120 amino acids. There is no information about such sequences in plant enzymes. This section of the aleurain sequence has two places where two or three basic amino acids are adjacent (Fig. 3, underlined); there is evidence that, in some mammalian secreted proteins, prosegments are cleaved on the carboxyl side of paired basic residues (18), although this is not found for the lysosomal aspartyl protease cathepsin D (16). That the estimated size of the gibberellic acid-induced aleurone thiol protease is about 37 kDa (5) suggests that much of this region between the cleavage of the probable signal peptide and the region homologous to thiol proteases must remain intact, if, in fact, our clone and that protease activity are products of the same gene.

Unusual Features of the Nucleotide Sequence. The nucleotide sequence of the aleurain cDNA was analyzed by use of the program DOTMATRIX (13) in order to screen for regions containing repetitive sequences. In this analysis, blocks of 8 nucleotides were compared and a dot was generated when 6 or more of the 8 matched. The results are presented in Fig. 4*A*. It can be seen that a section of the 5′ end of the sequence, up to about nucleotide 465 (circled), is enriched in short sequences that are repeated frequently. This result is primarily due to the fact that the base composition for that region is 72% G+C; in contrast, the base composition for the sequence 3′ to nucleotide 465 is only 52% G+C. It is interesting that this 5′ G+C-rich region encodes the "extra" 143 amino acids, and the transition from G+C-rich to balanced nucleotide representation occurs at the onset of the region homologous to thiol proteases. These findings suggest that perhaps the original thiol protease gene acquired another region by some recombinational mechanism, and this region proved useful to the organism. A precedent for such a fusion is provided by the gene encoding the chicken calcium-dependent thiol protease, the 5′ end of which encodes a papain-like thiol protease sequence, and the 3′ portion, a calcium-binding domain similar to those found in calmodulin-like proteins (19). This

speculation might be addressed by analyzing the organization of intervening sequences in the gene itself.

Results obtained from analysis of theoretical mRNA secondary structure (14) are consistent with the DOTMATRIX data. As presented in Fig. 4*B*, the predicted structure of the 5′ end includes a very stable hairpin-loop structure beginning with nucleotide 33. The free energy directing formation of the two component loops is indicated in parentheses (as kcal/mol; 1 kcal = 4184 joules), and these values, predicting very stable structures, are due to the high G+C content of the region. The initiation codon (underlined) is included in this structure. This pattern, where the initiation codon is locked in a stable hairpin loop, is also found in the three α-amylase mRNAs, for which some experimental data support the theoretical, computer-generated structures (9). We have speculated that these structures may affect translation efficiency and that gibberellic acid-mediated translational control (20, 21) might involve alteration of their stability (9).

Expression of Aleurain mRNA. The relative amounts of aleurain mRNA in different tissues were assessed by electrophoresing 10-μg samples of total RNA in a formaldehyde-containing agarose gel, transferring the RNA to nitrocellulose, and hybridizing the blot with nick-translated probe derived from the *Xho* I–*Acc* I cDNA fragment. The results are presented in Fig. 5 *Left*. It can be seen that detectable 1.6-kilobase mRNA is present in untreated aleurone cells (lane U); after incubation for 18 hr in the presence of 10^{-6} M gibberellic acid (lane G), the quantity of hybridizable RNA increases 7-fold (as assessed by densitometry). In contrast, incubation of the aleurone cells for 18 hr in the presence of 10^{-5} M abscisic acid (lane A) results in a substantial decrease in the amount of hybridizable RNA. This difference is not an artifact, because hybridization of identical blots with a probe for a mRNA that is not substantially affected by hormone treatments of the cells yields the expected pattern (data not shown). This pattern of RNA present in unstimulated cells, increasing with gibberellic acid and decreasing with abscisic acid treatments, is quantitatively similar to that found for the type A α-amylase mRNAs (9). In contrast to those gene products, however, RNA hybridizable to the aleurain probe and indistinguishable in size from that found in aleurone cells is present in shoot (Fig. 5, lane S) and root (lane R) tissue.

This observation, that aleurain mRNA appeared to be hormonally regulated in aleurone cells and, in contrast to α-amylase, expressed in shoot and root, raised the question of whether different genes might be expressed in different tissues. This was addressed by hybridization to nitrocellulose blots derived from restriction digests of genomic DNA. Initial results obtained with an intact cDNA probe (data not shown), as well as probes from the 5′ or 3′ (Fig. 5 *Right*, 5′ and 3′ probes, respectively) halves of the cDNA, identified numerous hybridizable bands, suggesting that these sequences were present in numerous copies in the genome. The issue was clarified by use of a probe derived from the *Xho* I–*Acc* I fragment encompassing the central part of the coding sequence (see Fig. 1). As demonstrated in Fig. 5 (internal probe), a single hybridizing band is detected in genomic DNA digested with either *Hin*dIII (lanes H) or *Eco*RI (lanes E).

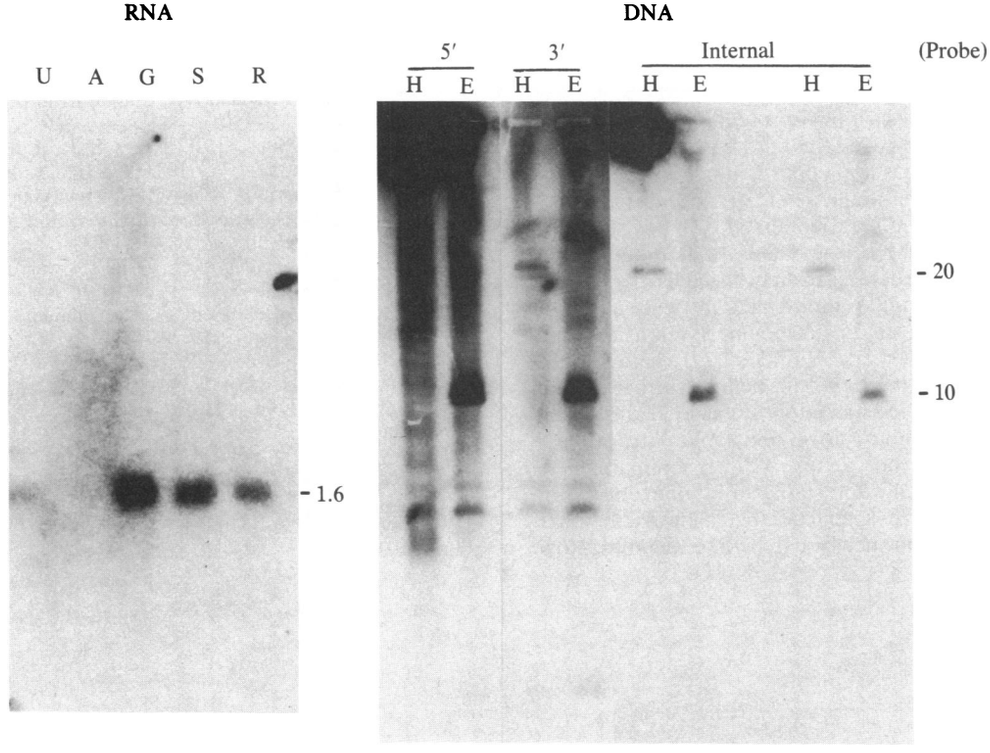

FIG. 5. Hybridization experiments. (*Left*) A nitrocellulose blot carrying RNA transferred from a formaldehyde-containing agarose gel in which total RNA samples (10 μg per lane) had been electrophoresed. Samples were from aleurone cells that had been untreated (lane U) or treated for 18 hr with 10^{-5} M abscisic acid (lane A) or 10^{-6} M gibberellic acid (lane G) or from shoot (lane S) or root (lane R) tissue. This blot was hybridized with the *Xho* I–*Acc* I fragment (nick-translated to a specific activity of 10^8 cpm/μg) and then washed (see *Materials and Methods*). The size (1.6 kilobases) of the single hybridizing species was estimated from simultaneous electrophoresis of denatured DNA markers. (*Right*) Aliquots (20 μg) of genomic DNA were digested with either *Hin*dIII (lanes H) or *Eco*RI (lanes E), electrophoresed in a 0.8% agarose gel, and transferred to nitrocellulose. Two identical blots were first hybridized with nick-translated probes (>10^8 cpm/μg) consisting of either the 5′ or the 3′ end of the cDNA insert. These were constructed by subcloning the two different fragments derived from a *Pvu* II digest (see Fig. 2). Hybridization and washes were as described in *Materials and Methods*. After autoradiographic exposure for 40 hr at room temperature, the blots were washed for 5 min at 100°C in 1 mM EDTA/0.1% $NaDodSO_4$. They then were rehybridized with the *Xho* I–*Acc* I probe (Internal), washed, and exposed for 20 hr at −70°C with an intensifying screen. The "internal" probe results therefore represent those obtained with the blot originally hybridized with the 5′ probe and the 3′ probe. Numbers refer to the size (in kilobases) of the two major bands.

This result suggests that only a single copy of the aleurain coding sequence is present in the barley genome, although two adjacent genes might give the same pattern. We have recovered an aleurain genomic clone from our library (12), in which there is only one gene present, centrally located in a 20-kbp insert (data not shown), but characterization of several different clones will be necessary before the issue can be resolved with certainty.

DISCUSSION

The amino acid sequence specified by the barley cDNA clone described here is surprisingly similar to the sequence of the rat thiol protease cathepsin H. In fact, the similarity between those two proteins, 63% homology with two single-residue gaps necessary for optimal alignment, compares favorably with the overall sequence similarity of the types A and B barley α-amylase isozymes, 71% and also requiring two gaps for optimal alignment. This sequence similarity allows us to predict that the barley protein is a thiol protease, and accordingly we have named it "aleurain," where "aleur-" indicates the cell from which its mRNA was isolated and "-ain" designates it a thiol protease. Cathepsin H is a protease probably present in all mammalian cell types and thought to function in general protein degradation in lysosomes (6, 22). Its pH optimum, about 5.5, is consistent with that assumed intracellular location. The degree of sequence conservation observed between these rat and barley proteins, from evolutionarily very distant species, indicates that the two proteins must have some highly specific function that will not tolerate more divergence. In this regard, it is interesting that thiol proteases are thought to be involved in processing a number of protein precursors *in vivo*, both intracellularly, as with proinsulin (23), and extracellularly, as with proapolipoprotein A-II (24).

The gibberellic acid-induced thiol protease activity secreted from aleurone cells has a pH optimum of about 4 (5), and secreted α-amylase has a similar acidic pH optimum. Little is known about enzyme compartmentalization in and secretion from plant cells, but there is evidence that α-amylase is packaged in vesicular structures in aleurone cells (25). Barley leaf mesophyll cells contain protease activity, with an acidic pH optimum, that is present in subcellular fractions containing vacuoles (26). Thayer and Huffaker (27) characterized two endoproteinase activities from vacuoles in barley leaf protoplasts; one, identified as endoproteinase 1, was inhibited by reagents known to block the active site of thiol proteases. The data do not permit a direct comparison to similar experiments characterizing the aleurone thiol protease activity (5). We speculate that aleurain in other tissues might be present in such lysosome-like vacuoles. Now that its complete sequence is known, it should be possible to make antibodies to defined regions by using synthetic peptides as antigens. Such antisera would be useful for *in vivo* localization and characterization of the protein.

This work was supported in part by Grant 83-CRCR-1-1332 from the United States Department of Agriculture.

1. Yomo, H. & Varner, J. E. (1971) in *Current Topics in Developmental Biology*, eds. Moscona, A. A. & Monroy, A. (Academic, New York), pp. 111–144.
2. Jacobsen, J. V. & Varner, J. E. (1967) *Plant Physiol.* **42,** 1596–1600.
3. Schroeder, R. L. & Burger, W. C. (1978) *Plant Physiol.* **62,** 458–462.
4. Mikola, L. (1983) *Biochim. Biophys. Acta* **747,** 241–252.
5. Hammerton, R. W. & Ho, T.-h. D. (1985) *Plant Physiol.*, in press.
6. Takio, K., Towatri, T., Katunuma, N., Teller, D. C. & Titani, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3666–3670.
7. Mitchel, R. E. J., Chaiken, I. M. & Smith, E. L. (1970) *J. Biol. Chem.* **245,** 3485–3492.
8. Carne, A. & Moore, C. H. (1978) *Biochem. J.* **173,** 73–83.
9. Rogers, J. C. (1985) *J. Biol. Chem.* **260,** 3731–3738.
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
11. Rogers, J. C. & Milliman, C. (1983) *J. Biol. Chem.* **258,** 8169–8174.
12. Rogers, J. C. & Milliman, C. (1984) *J. Biol. Chem.* **259,** 12234–12240.
13. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
14. Zuker, M. & Stiegler, P. (1981) *Nucleic Acids Res.* **9,** 133–148.
15. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12,** 5145–5164.
16. Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4910–4914.
17. Steiner, D. F., Docherty, K. & Carroll, R. (1984) *J. Cell. Biochem.* **24,** 121–130.
18. Gordon, J. I., Sims, H. F., Lentz, S. R., Edelstein, C., Scanu, A. M. & Strauss, A. W. (1983) *J. Biol. Chem.* **258,** 4037–4044.
19. Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. & Suzuki, K. (1984) *Nature (London)* **312,** 566–570.
20. Mozer, T. J. (1980) *Cell* **20,** 479–485.
21. Higgins, T. J. V., Jacobsen, J. V. & Zwar, J. A. (1982) *Plant Mol. Biol.* **71,** 191–215.
22. Takahashi, T., Dehdarani, A. H., Schmidt, P. G. & Tang, J. (1984) *J. Biol. Chem.* **259,** 9874–9882.
23. Docherty, K., Carroll, R. J. & Steiner, D. F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4613–4617.
24. Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. M. & Strauss, A. W. (1984) *J. Biol. Chem.* **259,** 15556–15563.
25. Locy, R. & Kende, H. (1978) *Planta* **143,** 89–99.
26. Heck, U., Martinoia, E. & Matile, P. (1981) *Planta* **151,** 198–200.
27. Thayer, S. S. & Huffaker, R. C. (1984) *Plant Physiol.* **75,** 70–73.