
Nucleotide sequence analysis of alpha-amylase and thiol protease genes that are hormonally regulated in barley aleurone cells

Robert F. Whittier, Duff A. Dean, and John C. Rogers*

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

Received December 9, 1986; Accepted February 24, 1987

Accession no. Y00107

ABSTRACT

We have determined the nucleotide sequences of Amy32b, a type A α -amylase gene, and of the gene for aleurain, a thiol protease closely related to mammalian cathepsin H. Both are expressed in barley aleurone cells under control of the plant hormones gibberellic acid and abscisic acid, but only aleurain is expressed at high levels in other barley tissues. Sequence analysis indicates that the 5' end of the aleurain gene, comprising 3 exons and 2 introns, may have become associated with the remainder of the gene, encoding the protease domain of the protein, by some sort of recombination event. This 5' domain of the gene is very G+C-rich and is flanked by inverted repetitive sequences. We found two different groups of homologous sequence elements. The first group consists of four blocks of sequences conserved in the same spatial arrangement in both genes; these are arranged at similar intervals upstream from the Amy32b TATA box and from a TATA box present in intron 3 of aleurain, outside of the 5' domain and upstream from the protease domain. A part of two of these conserved sequences is similar to the core sequence of certain enhancer elements characterized from mammalian cells. The second group of homologous elements is present in the upstream region of both genes. We speculate that these conserved sets of sequences may have some role in either the tissue specificity of expression of the genes or in some part of the hormonal regulation imposed on them.

INTRODUCTION

Barley aleurone cells provide an attractive system for the study of developmental and hormonal regulation of gene expression in plants. These cells surround the starchy endosperm of the seed. During germination, in response to the hormone gibberellic acid (GA) secreted by the embryo, aleurone cells are directed to synthesize and secrete a number of hydrolytic enzymes (1), including large amounts of α -amylase. Another hormone, abscisic acid (ABA), can prevent the GA-induced synthesis of α -amylase (2). The mechanisms of actions of these hormones are not understood, although the effects are mediated, at least in part, through changes in α -amylase mRNA abundance (3, 4).

As an approach towards understanding more about this complex system, we have cloned and sequenced full- or nearly full-length cDNAs representing each of the two different α -amylase isozymes (5, 6) and have used these probes to

study expression of their specific mRNAs (6, 7). We showed that there are two type A (or low pI) isozyme genes expressed in aleurone cells. One corresponds to our clone E cDNA (5) and the other to a genomic clone, Amy32b (7); these appear to be coordinately regulated (7). We also have cloned and characterized a cDNA that represents the gene for a thiol protease, aleurain, that is closely related to the mammalian lysosomal thiol protease, cathepsin H (8). The qualitative pattern of expression of aleurain mRNA in aleurone cells is similar to that for the two type A isozyme α -amylase mRNAs (7); it is present in untreated aleurone cells, increases in steady-state abundance about 10-fold with GA treatment, and decreases with ABA treatment of the cells (8). Quantitatively, aleurain steady-state mRNA abundance levels are at least 10-fold lower than α -amylase mRNA levels with any given hormonal treatment (9). Aleurain differs from the amylases, however, in that its mRNA is expressed in other plant tissues at levels comparable to those in GA-treated aleurone cells (8).

Our goal is to design systems with which we can test specific DNA sequences for their ability to respond to GA and ABA in a way that reflects those hormones' effects in aleurone cells, and thus learn more about the mechanisms with which the hormones may control expression of α -amylase and aleurain. Towards this goal, we have cloned the genes for one of the type A isozyme α -amylases (7) and for aleurain. Here we present the nucleotide sequences for both genes. We provide evidence based on analysis of those sequences that a portion of the aleurain gene, encoding 143 amino acids preceding the cathepsin H-like domain, may have become associated with the protease domain by some sort of recombination event. We identify conserved blocks of nucleotides in both genes and speculate that these conserved sets of sequences may have some role in either the tissue specificity of expression of the genes, or in some part of the hormonal regulation imposed on them.

MATERIALS AND METHODS

The general methods for agarose gel electrophoresis, restriction enzyme fragment analysis, blot transfer and hybridization to nitrocellulose filters, and preparation of RNA from barley aleurone layers have been previously described (6).

Isolation of barley genomic clones

Construction of a barley genomic library in λ MG14 and isolation and preliminary characterization of the type A α -amylase gene Amy32b has been described (7). The aleurain gene was isolated from the same library in an

identical fashion, except that the aleurain cDNA (8) served as the hybridization probe. Restriction fragments of the barley DNA insert that carried hybridizing sequences were subcloned into plasmids for restriction mapping and detailed analysis.

DNA sequence analysis

DNA sequencing was performed primarily by the dideoxy chain termination technique (10) as modified (11). Initial sequencing of both genes utilized an M13 deletion strategy (12). Gaps remaining in the sequence were resolved by subcloning specific restriction enzyme fragments into the bacteriophage vector, by using synthetic oligonucleotide primers complementary to adjacent sequences, or by the chemical degradation sequencing method (13).

Amy32b (Figure 1A) The portion of the clone presented was sequenced on both strands throughout its length, except for 150 bp near the 3' end; this area was subjected to 4 different sequencing runs in the same direction. Sequence was determined across all restriction enzyme sites used for subcloning or end-labeling.

Aleurain (Figure 1B) All sequence presented was determined in both directions with the following exceptions. No sequence data were collected in the downstream direction for nucleotides i) 598-674 in exon 3, ii) 1676-1767 in intron 4, and iii) 2138-2316 in intron 4. No sequence data were obtained in the upstream direction for nucleotides 2317 through 2400 in intron 4. Sequence data were obtained crossing all restriction enzyme recognition sites used in subcloning except for the Stu I recognition site at nucleotides 2314-2319 in intron 4. However, electrophoretic mobility of the Nde I to EcoRI fragment on a 1.5% agarose gel was in tight agreement with the predicted fragment size and ruled out the possibility of tandem Stu I recognition sites more than 30 bp apart (data not shown).

Preparation of Amy32b subclone

In order to obtain a probe useful for S1 experiments, a subclone of Amy32b was constructed from upstream sequences that terminated just in front of the ATG initiation codon. Briefly, this was accomplished by ligating a BamH I linker (8-mer, New England Biolabs, Beverly MA) onto the Bsh II site after filling in the 5' overhang with the large fragment of DNA polymerase I (14). The upstream sequences were then cloned into M13mp19 (15) as a Hind III-BamH I fragment.

Aleurain subclones

M13.mp19.XhoI-PstI, used in generating probe for S1 experiments, contains the 293 bp Pst I-Xho I fragment from the 5' end of the aleurain gene subcloned into the Pst I and Sal I polylinker sites of M13mp19 (15). The

single-stranded phage DNA therefore harbors this fragment in the sense orientation with respect to the aleurain gene. The 18 nucleotide oligomer primer used in conjunction with this subclone for preparing the probe was generously provided by Dr. Evan Sadler. Its sequence is CGTGGGCCATTCGTTTC and is complementary to nucleotides 51 through 68 as numbered in Figure 4. M13.mp18.aleurain5'down contains a 170 bp Nci I-Xho I fragment from the aleurain cDNA subcloned into the Sma I and Sal I polylinker sites of M13mp18. These 170 bp span the coding sequence from the middle of the second exon to within 6 bp of the end of the third exon. M13.mp19.aleurain3'up contains a 743 bp Hpa II-Acc I fragment from the aleurain cDNA subcloned into the Sma I polylinker site of M13mp19 in such an orientation that probes generated off of phage DNA proceed from the 3' toward the 5' end of the aleurain gene. These 743 bp span the coding sequence from within 7 bp of the 3' end of the third exon through almost the middle of the eighth exon.

Preparation of probes

The general method was that of Burke (16). The appropriate primer was annealed to its single-stranded phage DNA and extended with the large fragment of DNA polymerase I in the presence of 50 uCi of α [³²P]dCTP (3000 Ci/mmol) and the appropriate unlabeled deoxynucleotide triphosphates. After incubation at room temperature for 20 min, 8 nMol of unlabeled dCTP was added, and the incubation was allowed to continue for 20 more minutes. The double-stranded region was digested with Nco I (Amy32b), Pst I (M13.mp19.XhoI-PstI), EcoR I (M13.mp18.aleurain5'down) or Xba I (M13.mp19.aleurain3'up) to excise the labeled segment, and denatured in 0.3 M NaOH before electrophoresis in an agarose gel. The single stranded probe was localized in the gel by brief autoradiography, and the major radiolabeled band was eluted from DEAE paper (17).

S1 Mapping

The general method followed was that of Berk and Sharp (18) as modified (19). 200,000 cpm of the single stranded probe were co-precipitated with 20 μ g of RNA from aleurone layers, RNA derived from shoot tissue, or yeast tRNA. Each pellet was resuspended in 10 μ l of 0.4 M NaCl, 40 mM 1,4-Piperazinediethanesulfonic acid pH 6.5, 1mM EDTA, 80% formamide, then heated at 90°C for one minute, 52°C for 60 minutes, and 37°C for an additional 60 minutes. To each was added 300 μ l of 280 mM NaCl, 30 mM potassium acetate, pH 4.5, 4.5 mM zinc acetate containing 180 units of S1 nuclease (PL Biochemicals, Milwaukee, WI) in the case of Amy32b and 400 units in the case of aleurain. The samples were incubated at 37°C for 30 minutes. The reactions were then stopped by the addition of EDTA and ammonium acetate to

final concentrations of 10 mM and 2.5 M, respectively. The samples were precipitated with three volumes of ethanol in the presence of 5 µg of tRNA, and then analyzed by electrophoresis through a 5% acrylamide-urea sequencing gel.

Southern blots

Restricted DNA after gel electrophoresis was depurinated in 0.25 N HCl and transferred onto Zeta-Probe nylon membranes (Bio-Rad) by blotting in transfer solution (0.4 N NaOH, 0.6 M NaCl) overnight. The membranes were soaked in neutralizing solution (0.5 M Tris-HCl, pH 7.0, 1 M NaCl) for 15 min, dried and incubated 2 hrs at 80° C under vacuum. Membranes were wetted in a solution of 30 mM NaCl, 3 mM sodium citrate and 0.1% SDS then incubated

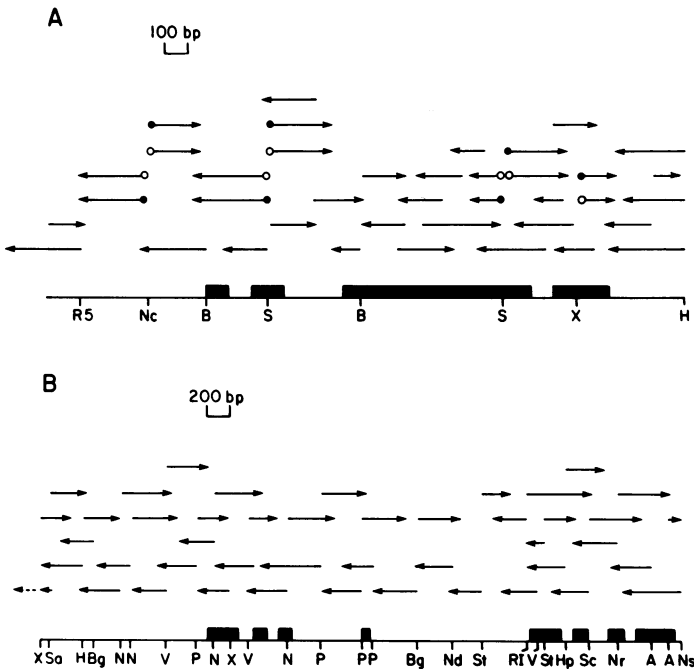


Fig. 1. Sequencing strategy for Amy32b and aleurain: The direction and extent of each individual DNA sequence determination is designated by the arrows. Arrows without attached symbols represent determinations with the dideoxy chain termination method, while determinations by the chemical degradation method with fragments labeled at the 3' end (closed circles) or 5' end (open circles) are designated by arrows with circles. For each gene, the exons are indicated by the enclosed boxes.

Abbreviations for restriction enzyme sites: A, Acc I; B, BssH II; Bg, Bgl II; H, Hind III; Hp, Hpa I; N, Nae I; Nc, Nco I; Nd, Nde I; Nr, Nar I; Ns, Nsi I; P, Pst I; R, EcoR I; R5, EcoR V; S, Sal I; Sa, Sac I; Sc, Sca I; St, Stu I; V, Pvu II; X, Xho I.

A

-606 CCTCCCCAG TTTATCTTCA ATTTGTCAA AAAATCATGT TCGGACCGTT TGACGGATCG ATACATGATC GCGTTGGATA AATTCCGTGG TCCGGACAGC

K

-506 ACGGTTCAA CCCTTACGGG AGGTTTGAGG GACGGCGTTG GAGATATCCT TAGCTGTTC CCGGTTGGTT TGTGAACCCA AAAGGTATAT CCTCGTAAAT

-406 ATTTCTGTTA CTGACCACA TTAAGAACAG TTTATATGGT CTGCATTGAA ATATCTCGGT GCAAGTGACA GGTGTACAGA TACATTTTGC ATCGAGTTCG

P P P E

-306 TTACCACGGA AACATGAAA ATCTTGCAA ATCTTGCAA ATCTTGCACT ATGTTCAATT GGAATAACAG TTAATCAAAA AAGTAAAGG CAACGCTGGG

E' A'

-206 TGATCCAGC TTGGATAGTG CTATCTTTC HcoI CCATGGAATT TGTGCCGGCC CGGATTGACT TGCCATCATC TGTTCACCT TTTCTCGTAA CAGAGTCTGG

B

-106 TATCCATGCA GTGCCCTCAA GCAACTCCT ACGGGACGT AGCTCGTGT AAATACCGCT GTGGCATCGA CTTCCTATAA ATACAAGCA CGTAGAACTC

BssHII
BssHII

-6 TTGTAACCAT CAATCACCAG TCTTTGAAAT CATTCAATCA CAGAACAAGA GTGCAGCGAA CAGTGTGAGA TCGACAGTAG CCGCGCGCAT GGGGAGAAC
H G K H

95 GGCAACTGTT GCTGCTTCTC TCTGCTGCTG CTCTCTCTCG CCGGGTTGGC GTCCGGCCAT CAAGTCTCTY TTCAGGTACA CATGTCGTG GTTTTTGATT
G N L C C F S L L L L L L L A G L A S G H Q V L F Q

195 TGTCCGGGTT GGCTCAGCCT GTTCTGTGA TCTCAGGAGC TTAATTAATC GTGGGAGCTG GAATTGATGT TGCAGGGGTT CAATCTGGAG TCGTGGAGC
G F H W E S W K

Sali

1295 AGAGCCGCGG GTGGTACAA ATGATGATGG GCAAGTCCGA CACATCCGCC GCCGCCGGG TCACGCAGCT CTGGCTCCCG CCGCCGTCGC ACTCCGCTC
Q S G G W Y H M H M G K V D D I A A A G V T H V W L P P P S H S V S

395 CAACGAAGGT CCTCTTAC ACTAAATCAT TCGTGTCTCA ACTGAACATC TAGATGCAAC CAAAAACGTG CTATGTTGGA TTGGGGTTGC AAACCTATTA
H E

495 ATTAATTCGT TGAGCAGAT GCTTGGACAG CTACCAGAAG CATTTCCTA CTCTATTTT CGATGGAAC GTACGGGCTG GGATGAAGAA CAAAACTGC

595 AAATACGATC AAACAAGTAT ACAGTATACT GTACAAACTA AAACGATGGA CTTGGGTGCA GGTACATGC CTGGCTGGCT CTACGACATC GACGCTCCA
G Y M P G R L Y D I D A S

BssHII

695 AGTACGGCAA CCGCGCGGAG CTCAAGTCCG TCACTCGGCG CTTCCACGGC AAGGGCGTGC AGGGATAGC CGACATCTGC ATCAACCACC GTTGGCGGA
K Y G H A A E L K S L I G A L H G K G V Q A I A D I V I H N P C A D

795 CTACAAGGAC AACCGCGGCA TCTACTGCAT CTTGAGGTT GGCACCTCCG ACGGCCGCTC CGACTGGGGC CCCCACATGA TCTGCGCGCA CGACACCAA
Y K D N R G I Y C I F E G G T S D G R L D W G P H M I C R D D T K

895 TACCGCGAGC GCACCGCCAA CCTTGACACC GGTGCCACT TCGCCGCCGC GCCTGACATC GACCACCTCA ACGACCGGCT CAGCGCGAG CTCAAGGAGT
Y A D G T A H L D T G A D F A A A P D I D H L N D R V Q R E L F E

995 GGCTCCTCTG GCTCAAGAGC GACCTCGGCT TCGACGCGTG GCGCCTCGAC TTCGCCAGGG GCTACTCGCC GGAGATGCC AAGGTGTACA TCGACGGCAC
W L L W L K S D L G F D A W R L D F A R G Y S P E M A K V Y I D G T

1095 GTCCCGAGT CTCGCGGTGG CCGAGGTGTG GATTAACATG GCCACCGGCG GCGACGGTAA GCCCACTAC GATCAGGAGC CGCACCGGCA GAACTGGTG
S P S L A V A E V W D N H A T G G D G K P H Y D Q D A H R Q N L V

1195 AACTGGGTGG ACAAGTCCG CCGCCCGGCA TCGCTGCGCA TGGTGTGCA CTTACAGACC AAGGGGATAT TGAACCGCGC CCGTGGAGGC GAGCTGTGGC
H W V D K V G G A A S A G H V F D F T T K G I L N A A V E G E L W

Sali

1295 GCTGATGGA CCCGACGGG AAGCCCCGGC ACATGATGG GTGGTGGCCA GCCAAGGCGG TCACTTCTGT CGACAACCAC GACACGGGCT CCACGACCC
R L I D P Q G K P P D V M G W W P A K A V T F V D N H D T G S T Q P

1395 CATGTGGCCA TTCCCTTCCG ACAAGGTGAT GCAAGGCTAC GCATACATCC TCACCCACCC AAGCACCCAT GCATCGTAA GCAAACCAAC TCTGGAGCTC
H W P F P S D K V H Q G Y A Y I L T H P G T P C I

1495 AGCTCGGCT TCACTCAGT ATGACTCCAA TCATTACGGA AACTMAAAT CTCTTGCTT GTCGGTTTGA GFTYACGA CCAATTCFTT AACTGGGGGT
F Y D H F F N W G

XhoI

1595 TCAAGGATGA GATCCGCGC CTGGTGGCCA TCAAGAACGC AAACGGCCTC ACGCGCAGCA GTGCGCTCGAG ATCCTCATG CACGAGAGG ACGCTACGT
F K D E I A A L V A I R K R N G I T A T S A L E I L M H E G D A Y V

```

1695  CCGCGAGATC GACGGCAAGG TGGTGGTGAA AATCGGGAAC AGGTACGACG TCGGGGCGGT GATCCCCGCTG GGTTCGCAA CGTCGGCGCA CGGCAAGGAC
      A E I D G K V V V K I G T R Y D V G A V I P A G F A T S A H G K D

1795  TAACCCCGTA TCGGAGAAAC AGCTCCGCGC GCGACACTAC AACGGAGCTG AAGCCTCCAC TCAATCCACAT TCAATCGAG CATGCATGAA TTTTCCAAA
      *

1895  TAATGATTCA CTTCGCTAT AGACACGCAG ATATATATGA TTAATFACGT AGTATATCTC TGCCCAAATG CTCACACGA CTGGAGTAAA TAAATFACGT
      *

1995  AGGATATGAA GAGGTATAG AGGCAATTAC GGCATTGCC ATGTTGCTGC GTGGTCCGT GGGTGTATAC TATGTGGA TGCTCCCGCA TAAACTGTT
      *

2095  CAGGAACAAG AAGCTTG
      HindIII

```

Fig. 2. Sequence of Amy32b: Numbering uses the cap site as +1. The location and extent of repeated elements is indicated with arrows; these are designated with letters so that similar sequences in both genes can be compared. Dots indicate identical nucleotides in the pairs of repeated elements. The TATA box in the promoter is underlined, as is the probable poly(A) addition signal at the 3' end. Asterisks are positioned over the first and last nucleotides that would be in the mRNA sequence, and the mRNA sequence is in bold-face type. Where that sequence is translated, the derived amino acid sequence of the protein product is presented in standard single letter code. Restriction enzyme sites relevant to sequencing, S1 mapping procedure or Southern blot analysis are identified.

at least 2 hrs at 42° C in prehybridization buffer (50 % formamide, 0.8 M NaCl, 50 mM sodium phosphate, pH 7.0, 5 mM EDTA, 0.1 % SDS, 1 x Denhardt's solution (14), 1 % glycine and 125 µg denatured salmon sperm DNA/ml). Prehybridization buffer was removed and blots were incubated in hybridization buffer (prehybridization buffer with 10 % dextran sulfate) at 65° C for 30 min prior to addition of radiolabeled probe. Hybridization proceeded for 12-16 hrs at 42° C. Subsequent washes were as described by Maniatis (14) except that the first wash solution included 1 % SDS.

Analysis of sequence homology

Alignment of sequence blocks during DNA sequence analysis, and comparison of gene regions in searches for homologies were assisted by the programs COMPARE and DOTMATRIX (Dr. Wayne Barnes, Biochemistry Department, Washington University School of Medicine) carried within the Washington University Medical School VAX-11 computer.

RESULTS

Sequence analysis

Southern blot analysis indicated that the structural integrity of both genes analyzed here had been accurately preserved during cloning (Figure 6; 7; unpublished data). The two genomic clones were therefore sequenced according to the strategy presented in Figure 1.

Amy32b gene structure

From previous hybridization studies and partial sequence analysis (7)

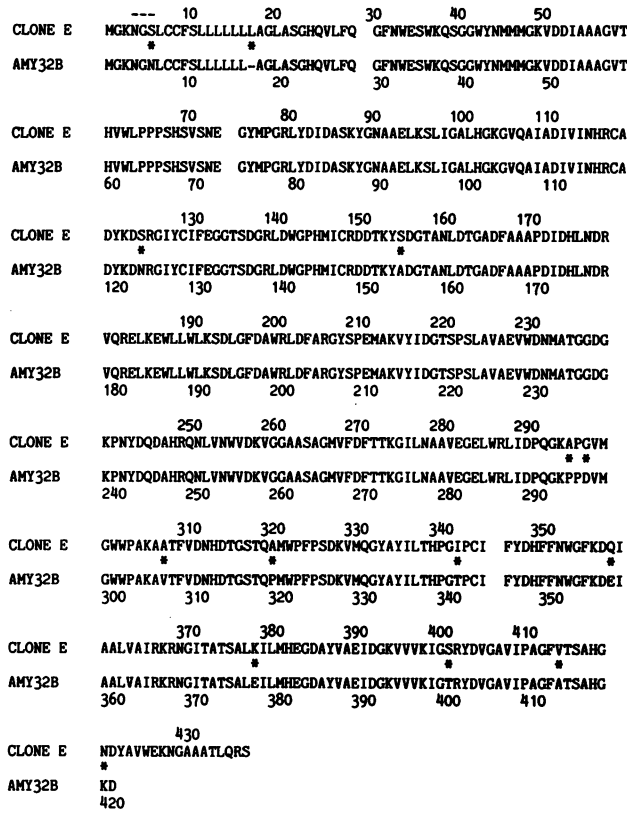


Fig. 3. Comparison of the sequences of the clone E and Amy32b type A (low pI) α -amylase isozymes: The derived protein sequences from clone E (5) and Amy32b are aligned; differences between the two are indicated with asterisks between the sequences. Open gaps in the sequences indicate the positions of the introns in Amy32b. The only potential signal for N-linked glycosylation is in the signal peptide region of clone E and is indicated with lines above the residues.

Amy32b was identified as one of two coordinately regulated type A isozyme genes. The other of the two genes is identified by our cDNA clone E (5). We presented evidence that the sequence of these two closely related genes was very similar except in the 5' untranslated regions (7). This close similarity enables the coding regions in the amy32b gene to be identified, thereby defining the protein sequence of that gene product. We aligned the clone E cDNA and Amy32b sequences to define the locations of the three introns identified in Figure 2. These are all bracketed by the canonical /GT...AG/ splice junction sequences (20). While, as previously noted (5),

the coding sequence is G+C rich (64%), the introns together have a nucleotide composition of 43% G+C. The 3' untranslated regions of the two clones also share highly conserved sequences; for example, of the 51 nucleotides at the 3' end of clone E, 46 are identical in Amy32b. We have not directly determined the poly(A) addition site for Amy32b but, from its similarity to clone E, tentatively assign that position to nucleotide 2014 (Figure 2).

The predicted protein sequence of the Amy32b gene product is compared to that of clone E in Figure 3. A single residue gap, in the signal peptide region, is necessary for optimal alignment. Amino acids that are not identical are indicated with asterisks. It can be seen that the Amy32b protein is predicted to be shorter than that of clone E by 17 residues, of which 16 are at the carboxyl-terminus. Excluding this difference, the two proteins are 97% identical. It is interesting that neither is predicted to have a signal for N-linked glycosylation (Asn-X-Ser or Thr) within the mature protein, and only the clone E protein has such a sequence within the signal peptide region (indicated by lines over residues 4-6).

Aleurain gene structure

The aleurain gene is interrupted by 7 introns. The onset of homology previously noted (8) between the amino acid sequence of cathepsin H and aleurain coincides almost precisely with the 5' boundary of exon 4. For this reason we will subsequently refer to exons 4-8 as the protease domain of the gene. We also noted previously that this boundary, in the cDNA, coincided with a substantial change in the nucleotide composition of the sequence; the first 465 bases contained 73% G+C residues, while the following sequence contained a more balanced 52%. This unusual nucleotide composition is maintained within the 5' portion of the gene, in that introns 1 and 2 also have a very high G+C content.

Exons in the protease domain appear individually to contain different functionally important units. The cysteine corresponding to the cathepsin H active site residue is in the small exon 4, while the other active site member, histidine, is in exon 7. The two potential sites for N-linked glycosylation are found together in exon 5.

When the nucleotide sequence of the gene was compared to that determined for the cDNA (8), three errors in the cDNA sequence were found and confirmed by reexamination of the cDNA sequencing autoradiograms. In two instances, these were single nucleotide changes of C to T at cDNA nucleotide 996, and T to C at cDNA nucleotide 1010. With respect to the derived protein sequence, the first change is silent but the second changes amino acid 325 from Thr to Ile. These numbers all refer to those used previously (8). The

SacI

-1399 ACAAGAACCA AGAAGTATGA TGCAAGGATG CAAATGGTTT GAGCTCTCAG CGAACGATAC GATCAAGCTA CTCACTTGAG AGCCCCCCCC CCCTTGATAG

-1299 AACACAATC TATCTTAACC AGAAAACCTA TCAAGGGCAA ACCTATACTT TGCACCTCGT CCTCTTGAGC TAGATGATGA TGATCTTGGC TTCTCTAAGA

-1199 TGGACCACCT TTCCTGATGT TGTGGCTTG ATGAAGACTA GTTGATTGCT CCCCCATACT CACTATGGGT GAGCCACTCT TCAGCATATC TTCACAAGTC

HindIII

-1099 CATTGCCACC ACAATGGACG GCAAGCTTCA AGCATGCTAT CTCCTGCTG ATCCACTTGA ACTTGACAT CGCAATTTTG ATGACGATCA CAACTTGGAG

BglII

-999 TCATACTCCA TGGGTTGTAT GAGATCTTCC CTTTGACGCA AGCCCATGGA AACACACCTA ACTCTCACAT AGAACTCTCA CGAAGACCAT GGGTTAGTAC

-899 ACAAACCGT AATGGACAAT GCTTACCATG AAGGTGGATG TCAACACCGT GTCGCCAAGG AAGAGGTGCT GGTTCGAGAA GATGCAGACC GACATGCTCA

HaeI

H

-799 AGTTCGACGA GGAGTGAATC ATGCGCGCGA GCGCCATCTT TTTTGGATGC CGGCAGGTGT GCTGGCATGA CCACGGGGGC CGCGATGGCC TGGTCAAACT

HaeI

H'

-699 CAAGTCCCAC CCTTTTATG TCTCTGGCATG TGTGTGCGCC GGCTCGCGAG CGTCATCTTT TGGCATGATG CTGACATGAG ACATGGCCGC TCGCATGATC

SalI

-599 GGTGGCGGGC CTTTTTTTAA ATGTTGATGT CGGACATAAA ATGGGTGCGAC GAATTGGACG CACTGCTGAC CTA AACCTGTA AACAGGGCGG ACGCCGGGGC

I A I'

-499 GACGGCCGAC CCAACAGACG AAAAAACGAA CAAATTCGCC GTCCGTTTGG ATCGTCCGTT GGACTTGCTC TAAGTGTCTT TTTCTTAATG AAACATGTCT

F1 F2 C D J

PvuII

-399 GATTTATAAC GAATGGGTTT CGATTTATAT GAAAACAGCG TGTTTGCAT GAATTTAAAC TACTGTTTGA AGAATTGTAT ACATGTAACG TTAGATGGTT

K

-299 AGATTTCGCA TCCGATATGT GGAGACTGCG AATTGTACCT TGTCAATGGA TAGATACAGG AGCAAAATTT AGGGCGTTGG AGATGGCCCTT AGTACATAGC

L L'

-199 CGCACATGTG TGCCACGCGT GGTATTTTGG CTCTTAGAAT ATTGAGACAA AGAATAATCC TCGCTAGAGT AAGTACGGAG TAGATTATAC TTACTGCGAA

G F3 G'

-99 TCACAAATAT CCTCTGCGTA CAGAGTCTCA AACAAGTGGT CAAGCGTGCC CCGGTGCGGC CCGCGATATA TAAAAACCCC GGCACCGGGA CGCGAGCTCC

M

HaeI

2 TCCCCGCCGC CTACCCACC- GAGTAAATA CTGCTACCCA TCCGCCGGCG AATCBAARS GGCACGCCCC GCGCTCTCTT CTTGAGGCTC GCGCTCTGCG

H A H A R V L L L A L A L A V L

KhoI

102 CCACGGCCGC CBTGCGGCTC GCTCTCTCTT CTTCTTCTCC CBACTCCABC CBAATCCGCG CCGTACACGA CCGCGCCGCG TCGAGGCTCG AGTCCGCCCT

A T T A V A V A S S S S F A D S B P I R P V T D R A A S T L E S A V

202 CTTCCGCGCG CTTGCGCCGA CCGCCACGCG CTTCCGCTTC GCGCGCTTCC CCGTCAAGTC CBTGCTACTG CTACGACCCC CCAATCTCC GCGCCAGCCT

L G A L G R T R H A L R F A R F A V R

N

PvuII

302 CGATTCTCCG TCTGTCCGCG CCGCGCCGCT CCTCATGGCA GCTGACGTCC CGGCGGTCCT GTTCTGATCC CTTTCCGAGG TACGCGAAGA CTTACGAGAG

T G K S T E S

N N'

KhoI

402 CCGCGCGGAG GTGCGAGGCG GGTTCGCGAT CTTTCCGCGA AGCTCTGAGG AGGCTGCTCC CACCAACCGG AAGGCGCTCC CTAACGCCCT CCGCATCAAC

A A E V R R R P R I F S E S L E E V R S T R R K G L P Y R L G I H

M'

HaeI

502 GGTACGACCC CCGACCATCC TCTAAFCGCC GCGCGCCGCT TCGGCGCCCG CCGGCTCTCC CTTTAAACGC TCGTGCCTTT TGTCTGCTC CCGCAGGCTT

R F

602 CTGCGACATC AGCTGGAGAG AGTTCGAGCG GACCCGCTCC GCGCGCGCGC ABACCTCTCC GCGCGCGCTC GCGCGCACCC ACTCTATGCG GAGCGCCGCG

S D H S W E E F Q A T R L G A A Q T C S A T L A G H L R D A A

702 GCGCTCCGCG AGACCTAAG TCCCAAGTIT TACTACTAC TCTCACCGCA GAACATGCCA TCAATTAAT ATGAAGCTGT GGTAGCGTAG CTGCGTAGA

A L P E T

802 ATGCTCCCTG TATTGTAAAG TTGATACGG TTGCTGAGGA CGAGCAGGCT CACATGGTCA GTTTCAGTGC ATGTCAATCC TCGGTTGAAA AGCCAGGATA

D' C' C' A'

PstI

902 AGTAATATAT TTGGATTCA AACCAAGTAT CTTGCGAGCT TCTTGGTTTG CCTGAATTGC TGCAGGATTT TGCCTGTCAC TGTATTTTCC ACGGTCCCGG

B

1002 CACATATGTT AGTAAAAGT TACTTCTACT GTATGTACAC CAGCCAATGA AAGAGCAGGA AATGTTAAT TCCCAAGTAG CATGTGGTA ATTAAGGAT

F4

1102 GCCGCACAC ATTTCTTCT ATTTATAACT AGTATGTTG TCGAAGACG AGTGTATATA TAGCATAGAC TATATCTGT TGTAGCCGTT GACGCTTAAT

PstI

1202 TTGCAGATAT TATGTCGCTT GATCCAGCAA GCAGCAGAGT TGGTGTACT ACCAATCGAC TCCATTAATG TGGGGAGTAA GTGATGTTT CATTTGCTCA

```

1302 GAABAACCTGG AGGGAGGATG GGATCGTCAG CCCGTCAAA ACCCAGGCC ATTGCGGCTC CTGCTGGAGC TTCAGTAAC TAATGAATT CTGCAGTTTT
    E D W R E E D G I V S F V K H Q A H C G S C W T F S
1402 TTTTTCCTCA CACGTGCCCC TGATCAGTA ATAACCAATG TGTGCCCAAT ACTGTGCCCT CTCATGGTGA AGCGTGCTCA AATCTGGTGT TTTTCAACTT
    EcoRV
1502 GACGTTCCATA CTATATGATA TCCGACAGGA AGTAGAATGA TCGATCCGAT CCAGTFAATT GTCTGGCCCA TAGAACAAAG AGACTTCAAT TCAATCATCG
1602 CCCTCAAAGT AATATGTGTG GACCGCCCTT TGAATCGTGC TGCTTCTAAT AAAGTACTAG TATTAATAAC TGATCTGTAT GGCCCAAGAT GGGGACCGAC
    BglII
1702 AGATCAATCC ATTGCCATG TTGAAAATG CTGGCTCTCA AGAGCGGTGG TTGAGTATAC TTGAGATCTT TTTCACTAGA CACAATCATG AATTATTGGG
1802 AAGTGTACAG CAATGGAGTT AAAAATCTGC GCTCGTGTGA CATGGGAGCA AATAAATAAA TAAATACTG AGCAACTAAT ATAATTATGT ATAGTACATT
1902 GATTTTCATG TCTTCTTCA GTTGGTGTTT GGTGACCAGT CGCTATCTTG TGGGCGCTTT CATGCCCTCT GTTCATTATT CCCACATCTA AATAAGAATGA
    HdeI
2002 ATATATCTTT CAATATCGT CAGGCGAATG CAATTCCAAA TGCACACAGG TTACAGTGCC ATATGAAGCT AGTAACATAA GCATCAGTGG AAGAATTTGC
2102 TTCACAACTT TTGCACAGT TTACGGTGA CATCCAAGGA ACGTTAATA TATACTGTAA AATGCACCCA ATTTTAA-CA ATTATTGAT CAAGTGGCAT
2202 ACAGATAAAT ATAGATCAG ATATTCAATG AGGTTACTTA CCTTAATTAC CC-A--ACTT TGTGATCAG GTAATACTCA GTAGTTACAT CACGAGTTCA
    StuI
2302 CTTCAGTCT GAAGCCCTAG ACTCACATGA AAACATAGAT TTAGATTTGT ATGTTATCAT GTTATCTGAA TTTGTGTGAT TGGGTACGAT TTAAGCATTT
2402 TAATAAATAA CTACCATGT AACTTTTCCC CATTTTCTAT ACAAAGATAT TCTCATTGGC ATCTACTGTA TTAAGTAATA CTGAAGTAGC CTAGCTACAA
2502 GTAATAATGAC AGATACTGAA ATCTGCCCCA TGCCATGTTT AGGCTCTATA TATAAATAAA ATATGGTCAT GCCATATTTT CAGCATTGGC CAGTTTATG
    EcoRI
2602 TGATTTGCAA TCGAGAAGAT AGCTCATTTT CTAGAACATT TCCTCAAATA GAAAAGAGT AAATGGGTTG TTCCACAAC AAAGCAATG AATTCATATA
    PvuII
2702 CTTATCTTGG CAGCACTACT GAGCCACTTG AGGCACATA TACTCAGGCC ACTGGAAGA ACATCTCTCT TTTCTGACCA CAGCTGGTGG ACTGTGCTGG
    T T G A L E A A T T Q A T G K H I S L S E Q Q L V D C A G
    StuI
2802 TGGGTTCAAT AACTTTGGAT GCAACGAGG CCTTCCATCC CAGGCGTTTG AGTACATCAA ATCAATGGC GGGATGACA CCGAGGATTC CTACCCCTAC
    G F H H F G C H G G L P S Q A F E Y I K Y H G G I D T E E S Y P Y
2902 AAGGTTGTCA ATGGCTCTG CCAATACAG GCTGAAAATG CTGCGGTTCA GGTTTTGGAC TCAATAAACA TCACACTGGT AAGGAAAAGG AACCCCTTCT
    K G V H G V C H Y K A E E A A V Q V L D S V H I T L
    HpaI
3002 AATTTGGGTT GATGTTAATC ATTAGATTC AGCTAACTA AGTTTTGGTT GATGCTTTTT TTTCTGGTA AGAATGCCBA GGATGAACTG AAGAACCAGC
    H A E D E L K H A
3102 TCGGGTGGT TCGCCAGT GAGTGGCTT TTCAGTGTAT TACCGGTTTC AGGCATACA AAGCCGAGT TTACACCAAC GACCATTTGG GCACATCCCC
    V G L V E F V S V A F Q V I D G F R Q Y K E G V Y T S D H C G T T P
    ScaI
3202 TGAATGAGT ACTACTAAC CGTAGCCCAT CAAAGTATCA AACCATGTAC CAGTTATCA TCAGCTATTT AAGTTCATC TTCAGTTAAT GTGCTTTCTT
    D
3302 CTGCTTTGTA TCTTACACT GTTCTGCAAT TTTGTTCTGA AAATGCAAGT TTTGCTGAAA CTACTAGGAC GTGAACCACG CCGTCTTGGC AGTCGGTAT
    D V H H A V L A V G Y
3402 GGTGTGAGA ACGGCTCCC GTATTGGCTC ATCAAGAATC CDTGGGCGCC CBACTGGGTT GACATGTTT ACTTCAGAT GGAATGCGC AAGAATATG
    G V E H G V P Y M L I K E S W G A D W G D H G Y P K H E H G K H H
3502 GGTGTGAGA CAAATCCGTT TGACATTTCA TCTTACGTTT GCTCTTCAGG ACACAGCGCA CTCATGAGGC GATACTGACT CGACTGTTT AAATGATGCA
    C
3602 GCTATTGCTA CTGCGCCTC CTACCTGTG GTGGCGGATG GAAGAAGCGC CAAATCGCGG AGTACAGCTC CGATTCGCCA GACATGAATA AGGTGTTTTT
    A I A T C A S Y P V V A A *
    AccI
3702 AGCCCCAACA AGAACCTGAA GATTGGGGTA TACAGAACT GAAACTTTGC TGTGAATACA AACCAAAAG ACTACGGTGG CAGTGAGCCC CGTGGCGTGA
    AccI
3802 CTCTTCTCTG GTATGGTGTG ATGGCTATAT ATGTAGGTGA TCCAGCAATT TATTAGGTTT TCTCATGTAT GCTCAGACA TTTTATTTCC GGGCAATAA
    EcoRV
3902 AAATGGCCCA TTCATAGTTC ACCCAGGCAT CATCTTACTG TCGATATCAG CAGGAAACTG GCGCCACAC ATGCAT
    NsiI

```

Fig. 4. Sequence of Aleurain: Conventions used are the same as in Figure 2.

other error involved 4 nucleotides. Correction results in an insertion of 3 bases and a change in another so that the sequence from what were identified as cDNA nucleotides 691-700 (previously TACCAATACA) now reads TACATCAATACA. This results in a change in protein sequence from what were amino acids 219-222 (previously YQYN) to YIKYN. Overall, these amino acid changes remove a gap in the previous alignment with cathepsin H (8) and result in another residue identity between the two sequences.

There are numerous direct and inverted repetitive sequences within the

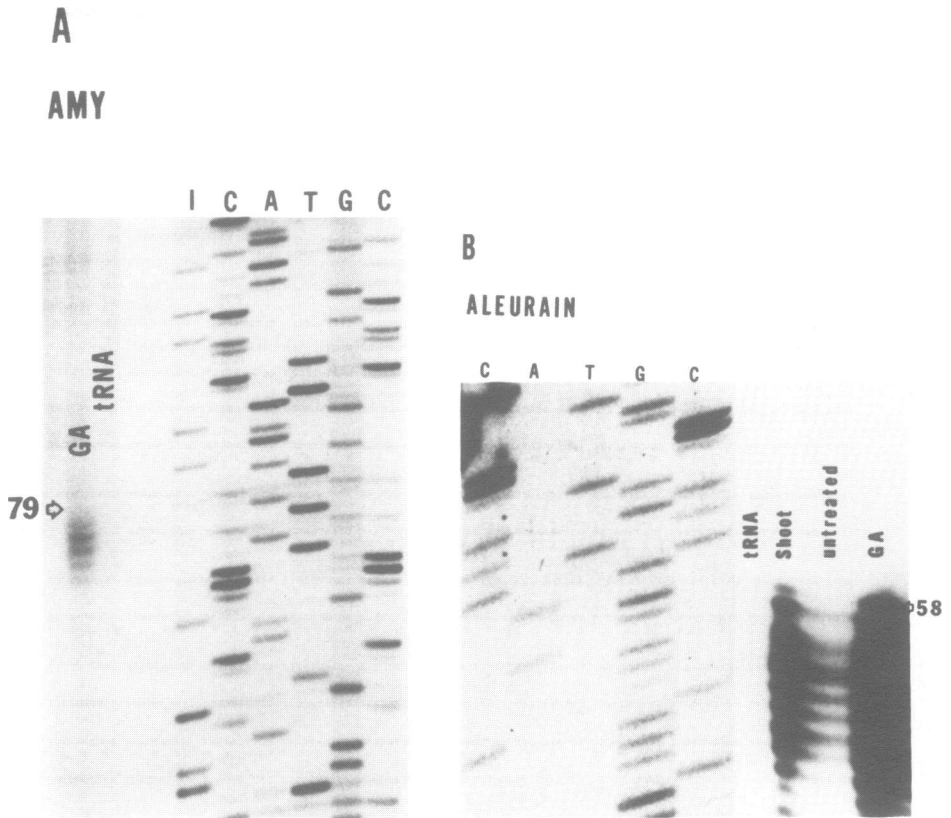


Fig. 5. S1 analysis of transcription initiation sites: Probes were synthesized as in Methods using as primers either the standard 15-mer sequencing primer (New England Biolabs, Beverly MA) for Amy32b or the special primer for aleurain (Methods). S1-resistant fragments were analyzed on standard DNA sequencing gels. Dideoxy sequencing reaction products were used as size markers; in the case of aleurain, the same DNA template and primer was used, while those for Amy32b were from an unrelated sequence. Sequencing lanes: dITP/ddGTP (I), dCTP/ddCTP (C), dATP/ddATP (A), dTTP/ddTTP (T), dGTP/ddGTP (G). The RNAs used in hybridizations are indicated on the figure in the appropriate lanes. The numbers on either side represent the size of the largest clearly identifiable protected fragments expressed as the distance in nucleotides that the fragments extend 5' from the ATG translation initiation codon.

5' half of the gene; these are designated with different letters in Figure 4. We will consider in detail those that may have some functional or evolutionary importance (Discussion).

Mapping of transcription initiation sites

The results of S1 analyses of the 5' ends of the two transcripts are

presented in Figure 5. In the case of Amy32b, previous data from primer extension indicated that initiation of transcription occurs at nucleotides that are 82 and 83 bp (base pairs) upstream from the ATG translation initiation codon (7). These results alone were not definitive, since premature termination of reverse transcriptase on the mRNA template could not be excluded. Therefore, confirmation of these results was sought by S1 analysis. The largest fragment of the probe that was protected from digestion by hybridization to mRNA has its 5' end at a position 79 nucleotides upstream from the translation start codon, ATG, (Figure 5A). S1 nuclease is known to digest away terminal nucleotides that are not single-stranded; since the probe was uniformly labeled, loss of nucleotides from either end in the hybridized region would give a ladder of smaller fragments similar to those seen in the figure. Given this problem with the technique, it is not surprising that the primer extension data indicated a slightly longer transcript. We interpret these results to be in close agreement and use the primer extension results to define the end of the transcript as 83 bp upstream from the translation start site. The fact that this site occurs 26 nucleotides downstream from a canonical TATA box (20) provides further support for our argument.

A probe complementary to, and extending beyond, the 5' end of the aleurain mRNA was hybridized to total RNA preparations from shoot tissue as well as from aleurone tissue incubated either in the absence of hormone (untreated) or in the presence of 1 μ M GA for 20 h. Analysis of the S1-resistant fragments is presented in Figure 5B. It can be seen that all three gave a similar-sized set of protected fragments, with the largest in each instance having its 5' end at a position 58 nucleotides upstream from the ATG initiation codon. The specificity of the result can be judged from the complete digestion of the probe in the presence of tRNA. This result is important because it supports our assertion (8) that only one aleurain gene is expressed in these different tissues. Primer extension data (not shown) also agreed, and we therefore assign the site of transcription initiation as indicated in Figure 4. The presence of a TATA box sequence beginning 33 bp upstream supports the conclusion that this is the transcription initiation site.

In this regard, it is interesting to note that the TATA sequence is bracketed by short G+C-rich blocks of sequences that form inverted repeats (designated as G and G' in Figures 4 and 7). The transcription initiation site is located just 3' to one arm of this palindrome.

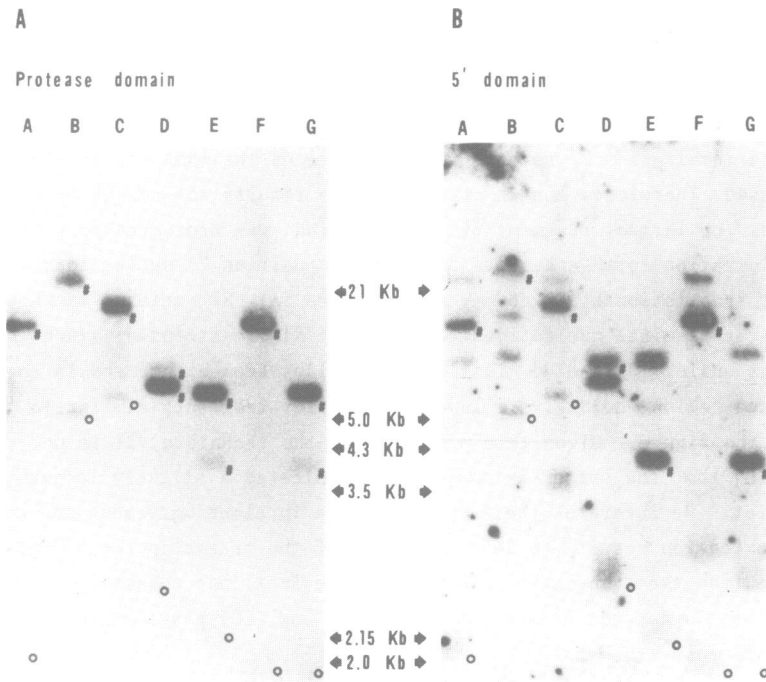


Fig. 6. Restriction pattern of barley sequences hybridizing with aleurain 5' domain specific and aleurain protease domain specific probes. Barley genomic DNA (10 μ g) was loaded onto each lane after digestion with the following restriction endonucleases: lane A: HindIII, B: EcoRI, C: BamHI, D: EcoRI + BamHI, E: EcoRI + HindIII, F: HindIII + BamHI, G: EcoRI + HindIII + BamHI. After transfer the blot was hybridized with radiolabeled probe generated off M13.mp19.aleurain3'up (6A). Bound probe was removed by washing in a 1 mM EDTA 0.1 % SDS solution for 30 min. at 100° C and the blot was rehybridized with a probe generated off M13.mp18.aleurain5'down (6B). Washes prior to autoradiography included 2.5 hr at 68° C in 15 mM NaCl, 1.5 mM sodium citrate, 0.5 % SDS.

Multiple genomic copies of a 5' aleurain sequence element

An intriguing set of sequence homologies between the upstream and intron 3 regions of the aleurain gene (see discussion) prompted a search for 5' domain sequences which might be present in the barley genome independent of the protease domain sequences. Genomic Southern blots (Figure 6) were analyzed with 5' domain specific and protease domain specific probes; these demonstrated restriction fragments which could be divided into three sets according to their pattern of hybridization. The first class (denoted by "#" in Figure 6) hybridized most strongly and could be explained by restriction sites known to be present in our genomic clone. For example, EcoRI cuts

between exons 4 and 5, generating two protease domain fragments. As expected, the more weakly hybridizing band is also detected by the 5' domain specific probe (lanes D, E and G). These fragments therefore correspond to our cloned gene. Reconstruction experiments indicated that these sequences were present in approximately one copy per haploid genome (data not shown).

A second class of restriction fragments (denoted by "o" in Figure 6) hybridized weakly to both 5' domain specific and protease domain specific probes. This class was present as one apparent restriction fragment in each digest. Although not apparent in some lanes of Figure 6A as photographed, these bands were visible in all lanes of the original autoradiograph. The significance of these sequences is unclear at present.

The third class of restriction fragments (all remaining bands in Figure 6B not identified by "#" or "o") hybridize solely with the 5' specific probe. The restriction sites delimiting these fragments do not correlate with those in the cloned aleurain gene. This class is present as multiple fragments; consistently two or three hybridizing bands can be identified in all digests except for the EcoR I BamH I double digest (Figure 6B lane D) in which the 2.6 kb (kilobase pairs) class II band appears to overlap a class III band (compare Figure 6B lane G). We interpret these results to mean that this final class are elements independent of the aleurain protease domain. The higher G+C content of the 5' hybridizing probe relative to the protease domain probe, however, could lead to detection of more distantly related sequences. Therefore, although the small number of bands detected indicates relatively specific hybridization, cloning of the genomic sequences giving rise to class III bands will be necessary to define their extent of homology with our probe. The existence of independent elements homologous to the 5' domain of aleurain is consistent with our speculation that the aleurain gene could have originated by recombination between one member of these elements and the protease domain (see discussion).

DISCUSSION

The genes analyzed here were cloned and sequenced as an initial step toward determining the sequences important for tissue specificity and hormonal control of plant gene expression in barley. Since these two genes are derived from separate families, similar sequences in their associated noncoding regions constitute candidate sites at which transcriptional regulation of these functions might occur. Indeed, two such sets of homologies were found, but in order to simplify their presentation it is

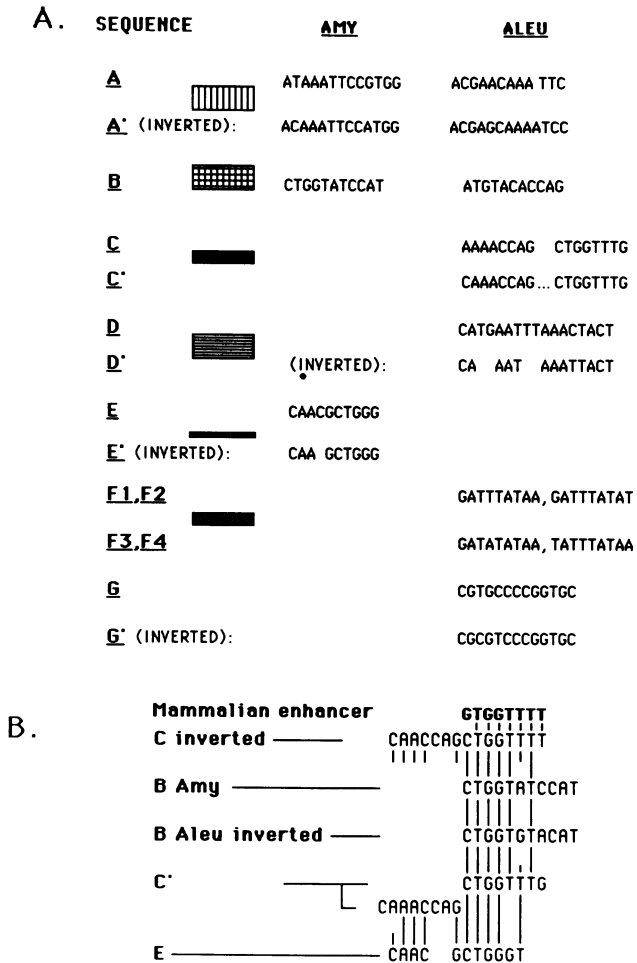


Fig. 8. Details of sequence elements identified in Figure 7: A. The letter designations and symbols for the sequence elements as used in Figure 7 are followed by the nucleotides (in 5' to 3' orientation) that comprise those elements in Amy32b (AMY) and aleurain (ALEU) respectively. B. The B, C, and E element sequences are aligned with the TGGGTTTC sequence to demonstrate similarities (indicated by vertical lines between identical nucleotides).

transposition event. First, homology with other known thiol proteases commences almost precisely with the amino acids coded for from the start of exon 4 and continues throughout the remainder of the coding sequence. Second, the G+C content of the 5' domain is 71 % with a fairly uniform distribution throughout its length while the protease domain contains a more

balanced 51 % G+C within its exons, or even an oppositely skewed 42 % G+C if introns as well as exons are tabulated. Third, as demonstrated in Figure 6, additional elements appear to be present in the barley genome with homology to the 5' domain but unassociated with protease domain sequences.

Considering the C sequences (Figures 7 and 8) as the outer borders of such a transposon, possible stem-and-loop structures are evident which would be reminiscent of structures proposed for several known plant transposable elements (21). Notably lacking from the outer borders of such a transposon is any target site duplication. However plant transposable elements have been reported lacking target site duplications (21) or with target site duplications as short as 3 bp (22) so that they might be obscured by random mutations even over a short evolutionary time frame.

Alternatively, if the C sequence duplications occurred fortuitously, and the D sequences constitute the outer edges of the element, a perfect 5 nucleotide GGGTT target site duplication is evident as illustrated in the enclosed box in Figure 7. This interpretation would invoke random mutations to account for the deterioration of homology at the outer edges of the D sequences. Such a possibility is not unreasonable considering that we have no measure of how long ago this putative transposition occurred. An analysis of the aleurain gene in other closely related monocots to determine if this 5' domain is present would help to resolve the question.

We have been unable to identify regions in published sequences of known plant transposons that are homologous to the inverted repeats just described. However, it is interesting that one transposon, Mu1 in maize (23), has a central core with a similar G+C-rich nucleotide composition. Rigorous proof that the identified domain in fact represents some sort of a transposable element, or its remnant, will require that at least one homologous element be shown to leave or invade a known genetic locus.

Sequence motifs shared between the aleurain gene and Amy32b

During sequence comparison we were surprised to find 3 sets of sequences in the Amy32b upstream region that appeared to be closely conserved in the aleurain gene, primarily within intron 3 outside the borders of the proposed transposable element (Figure 7). Eighty-nine bp upstream from an apparent TATA box in aleurain intron 3, and 63 bp upstream from the Amy32b TATA box are sequences, denoted "B" in Figures 7 and 8, inverted with respect to each other, but within which 9 of 11 nucleotides match (Figure 8). A further 55 bp upstream within the aleurain gene and 51 bp upstream in Amy32b is the right arm of a palindrome, denoted "A'", whose counterpart left arm, A, lies considerably farther upstream, in the case of aleurain, on the far

side of the putative transposed domain. Within larger palindromic arms, cores of 6 out of 7 nucleotides within the A' sequences match and 6 out of 6 nucleotides within the A sequences match (Figure 8). At a similar interval upstream from the A' elements are the C' inverted repeats in aleurain and the E-E' inverted repeats in Amy32b. As shown in Figure 8, 10 of 13 nucleotides within the arms of C' align with E; a similar relationship exists between C and E.

Another type of sequence, with a consensus of GTGGTAAA, is found 24 and 22 bp upstream from the authentic TATA boxes in aleurain and Amy32b respectively, as well as 31 bp upstream from the apparent TATA box in aleurain intron 3. This sequence is very similar to the GTGG(T/A)(T/A)(T/A)(T/A/G) sequences which have been noted for possible roles in the expression of the CAMV 35S promoter (24), chalcone synthase (25) and the pea *rbcS* gene family (26). In this regard it is of interest to note that these sequences as well as the sequences of the B, C and E elements all seem to have substantial similarity to the "core" sequence in enhancers from polyoma virus (27) and from the IgM heavy chain gene (28). In both of those examples, there are two variations, GTGGTTTG and GTGGTTTT, each of which seems to fit with one of the two C sequences (Figure 8).

Although no single comparison within these sets of apparently homologous sequences is likely to be significant, the very similar positioning and similar spacing of these elements considered as a group suggests that their occurrence is not due to chance and therefore of potential functional significance. This hypothesis predicts that sequences within intron 3 of the aleurain gene might have some effect on expression of the gene. Intron sequences have been shown to participate in hormonal regulation in another system (29).

The other outstanding sequence homology between the two genes, designated "K" in Figures 2 and 4, occurs in the upstream region of both genes and in the same orientation. Within this sequence 24 nucleotides can be aligned out of 26 in the aleurain gene and 30 in Amy32b. Dissecting this homology more precisely, a string of 8 nucleotide identities is terminated by 4 extra nucleotides in Amy32b. Directly following lies a further string of 18 nucleotides such that the first 10 match perfectly, the next 2 diverge and the final 6 again match perfectly. The K sequences are centered 220 bp and 468 bp upstream of the transcription initiation sites in the aleurain and Amy32b genes respectively. Disregarding the initial 8 nucleotide homology for lack of a rigorous statistical method to deal with a gap between the two sequences, the chances that two random 18 nucleotide sequences will match so

closely are ($4^{16} - 17$) or about one in 250 million. Considering the matrix of approximately 814,000 such comparisons possible between the sequenced upstream regions of the two genes, the probability of finding an 18 nucleotide homology of equally good fit between 2 sequences of similar length and random composition would be less than 1/300. The adjoining 8 nucleotide sequence homology may elevate the resemblance further into the realm of significance. For that reason the K sequences should be considered possible candidates for some regulatory function. The development of tissue specific, hormone responsive gene expression systems in barley will be required to test whether any of the sequences identified in this analysis do in fact play a role in determining the common regulatory aspects of the two genes.

ACKNOWLEDGEMENTS

This research was supported in part by grants No. 85-CRCR-1-1684 and 1692 from the U.S.D.A.

*To whom correspondence should be addressed at: Division of Hematology-Oncology, Box 8125, Washington University School of Medicine, 660 S. Euclid, St Louis, MO 63110, USA

REFERENCES

1. Varner, J.E., and Ho, T.-h.D. (1976) In Papaconstantinou, J. (Ed.), *The Molecular Biology of Hormone Action*, Academic Press, New York, pp.173-194.
2. Ho, T.-h.D., and Varner, J.E. (1976) *Plant Physiol.* 57, 175-178.
3. Mozer, T.J. (1980) *Cell* 20, 479-485.
4. Higgins, T.J.V., Jacobsen, J.V., and Zwar, J.A. (1982) *Plant Molecular Biology* 1, 191-215.
5. Rogers, J.C. and Milliman, C. (1983) *J. Biol. Chem.* 258, 8169-8174.
6. Rogers, J.C. (1985) *J. Biol. Chem.* 260, 3731-3738.
7. Rogers, J.C. and Milliman, C. (1984) *J. Biol. Chem.* 259, 12234-12240.
8. Rogers, J.C., Dean, D., and Heck, G.R. (1985) *Proc. Natl. Acad. Sci. USA* 82, 6512-6516
9. Mundy, J., and Rogers, J.C. (1986) *Planta* 169, 51-63
10. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 80, 3963-3965
11. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.
12. Dale, R.M.K., McClure, B.A., Houchins, J.P., (1985) *Plasmid* 13, 31-40.
13. Maxam, A.M, Gilbert, W., (1982) *Meth. Enzymol.* 65, 499-560.
14. Maniatis, T., Fritsch, E.F., Sambrook, J., (1982) *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
15. Norrander, J., Kempe, T., and Messing, J. (1983) *Gene* 26, 101-106.
16. Burke, J.F., (1984), *Gene* 30, 63-68.
17. Dreyzen, G., Bellard, M., Sassone-Corsi, P., Chambon, P., (1981) *Anal. Biochem.* 112, 295-298.
18. Berk, A.J., and Sharp, P.A. (1977) *Cell* 12, 721-732.
19. Favaloro, J.M., Treisman, R.H., and Kamen, R. (1980) *Meth. Enzymol.* 65, 718-749.

-
20. Mount, S.M. (1982) *Nucleic Acids Res.* **10**, 459-472.
 21. Nevers, P., Shepherd, N.S., and Saedler, H. (1986) In Callow, J.A. (ed.), *Advances in Botanical Research*, Academic Press, London, Vol. 12, pp. 103-203.
 22. Shepherd, N.S., Schwarz-Sommer, Z., Blumberg vel Spalve, J., Gupta, M., Wienand, U., and Saedler, H. (1984) *Nature* **307**, 185-187.
 23. Barker, R.F., Thompson, D.V., Talbot, D.R., Swanson, J., and Bennetzen, J.L. (1984) *Nucleic Acids Res.* **12**, 5955-5967.
 24. Odell, J.T., Nagy, F., and Chua, N.-H. (1985) *Nature* **313**, 810-812.
 25. Kaulen, H., Schell, J., and Kreuzaler, F. (1986) *EMBO J.* **5**, 1-8.
 26. Fluhr, R., Kuhlemeier, C., Nagy, F., and Chua, N.-H. (1986) *Science* **232**, 1106-1112.
 27. Griffin, B.E., Soeda, E., Barrell, B.G., and Staden, R. (1980) In Tooze, J. (ed), *DNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 831-896
 28. Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983) *Cell* **33**, 717-728.
 29. Moore, D.D., Marks, A.R., Buckley, D.I., Kapler, G., Payvar, F. and Goodman, H.M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 699-702