Introduction

The study of cereal albumins and globulins has lagged somewhat behind that of the prolamins, which have been considered as typical reserve proteins. However, these protein fractions merit closer attention for a variety of reasons. The main individual albumins and globulins are at least as abundant as many prolamin components, and it can be speculated that in a tissue, such as the cereal endosperm, which is completely consumed during germination, all abundant proteins may play a reserve role. They have also a higher proportion of essential amino acids, as compared with the prolamins, and thus may be relevant in connection with the genetic alteration of overall grain composition. Finally, a high proportion of the main components of these protein fractions have inhibitory and even toxic properties, which may be related to the protection of this tissue during development and germination, and might influence the nutritional value of the cereal products. We report here the characterization in barley of cDNA clones encoding two major groups of proteins: the CM-proteins, a family that includes inhibitors of trypsin and a-amylase, and the thionins, a group of high-lysine toxic polypeptides.

The CM-Proteins: Trypsin and a-Amylase Inhibitors

The albumin and globulin fractions of cereal grains (endosperms) are made up of over 20 major and many minor components, as can be shown by two-dimensional electrophoretic methods or by high-performance liquid chromatography (see refs. 1,2). About one third of these proteins can be extracted with chloroform:methanol (2:1, v:v) and have been designated CM-proteins. A close relationship among the different CM-proteins (mol. wt. 12,000-16,000) was inferred from their amino acid compositions (compositional divergence indexes) and immunoch-
been obtained for many of these proteins in wheat and barley and the predicted homologies have been confirmed in all cases (2,8-10). Many individual CM-proteins have been shown to inhibit either trypsin or α-amylases in vitro, whereas for some members of this group no such activities have been demonstrated (2,9,10). Although the in vivo function of these proteins is unknown, it has been speculated, on the basis of their in vitro activities, that they may play a protection role in endosperm (11).

The genetic control of CM-proteins has been extensively studied both in wheat and in barley (1-4,10,12-16,22). The chromosomal locations of genes encoding these proteins have been established through the analysis of aneuploid, addition, and substitution lines. Genes for these proteins have been found to be dispersed in chromosomes of groups 3, 4, 6, and 7 of wheat (1,3,10,12,13,22) and in chromosomes 1 (homeologous to group 7 of wheat), 3, and 4 of barley (15). Both cis and trans regulatory effects on the expression of the CM genes have been characterized in wheat (4,14) and, recently, differential effects of high-lysine mutations on the accumulation of individual CM-proteins have been reported in barley (2).

**In vivo and in vitro synthesis of CM-proteins in barley**

Synthesis and deposition of CM-proteins during endosperm development have been investigated in barley (17). These proteins are synthesized by membrane-bound polysomes from about 10 d to about 30 d after anthesis (maximum 15-20 d). Precursors of higher apparent molecular weight (13,000-21,000) than the mature proteins (12,000-16,000) are obtained in vitro. It was assumed that these precursors were processed co-translationally, based on the failure to detect them in pulse-chase, in vivo experiments. The CM-proteins quantitatively appeared in the soluble fraction using very mild homogenization and subcellular fractionation procedures, indicating either a very labile association with membrane subcellular structures or a transference to the soluble fraction after processing (17).

Characterization of cDNA clones of the CM gene family

A cDNA library from developing barley endosperm, collected at 20 d after anthesis, was prepared by cloning into the Pst I site of pBR322, following the homopolymer tailing method (poly G-poly C), according to standard procedures. Screening for recombinants corresponding to the CM-proteins multi-gene family was carried out in two stages. A radioactive ss-cDNA probe, prepared from a poly (A)*RNA fraction enriched for mRNA's encoding CM-proteins, was used for the first screening. Recombinants from this preliminary screening were further
Fig. 1. Nucleotide and deduced amino acid sequences of clones pUP-13 and pUP-38. The beginning of the mature protein (↓) and the polyadenilation signals (poly A) are indicated. Marked base (*) uncertain.
tested by hybrid-selected translation in the in vitro wheat-germ system, using $^{35}$S-cysteine as labeled amino acid, and identification of translation products by immunoprecipitation with monospecific antibodies, electrophoresis and fluorography. Two clones (pUP-13 and pUP-38) with inserts of ~500 bp were selected for further characterization.

The inserts in clones pUP-13 and pUP-38 were sequenced according to Maxam and Gilbert (18). Their DNA sequences are presented in Fig. 1, together with the amino acid sequences deduced from their longest open-reading frames. In Fig. 2, the two deduced sequences have been aligned with that previously obtained by direct protein sequencing for a barley trypsin inhibitor (19). The sequence deduced from clone pUP-13 includes what has the characteristics of the hydrophobic core of a leader sequence. The most probable N-terminal amino acid of the mature protein is the Glu at position 14, according to the probability method of Heijne (20), which would mean that the first 13 amino acids belong to the leader sequence and would leave 7 amino acids preceding the first Cys in the mature protein. Five previously determined N-terminal sequences of barley CM-proteins show a highly variable stretch of 3 to 6 residues preceding the first Cys (8, 9). The following sequence of 24 residues, from the first to the third Cys, is close to 50% homologous to any of the five sequences, including the trypsin inhibitor CMe, the only CM-protein for which a complete direct sequence is available (19). The sequence deduced from clone pUP-38 does not include the N-terminus. Clones pUP-13 and pUP-38 share amino acids with protein CMe at 31 and 40 positions, respectively, out of the last 94 C-terminal residues. These proteins are also homologous to B-hordeins (21), to inhibitors isolated from other cereal species, and to the 2S storage proteins from castor bean and rape (8, 9).
**pTH-1**

(5') TCAAGGGTGTG

leader

met val cys leu leu ile leu gly leu val leu glu gln val gln val glu gly lys ser

ATG GTG TGT TTA CTT ATA CTC GGG TTC CTT CTC GAA CAG GTG CAA GAA GGC AAG AGT

cys cys arg ser thr leu gly arg asp cys tyr asn leu cys arg val arg gly als gln

TGC TGC AGG AGC ACC CTA GGA AGA AAC TGC TAC AAT TGC TGC GTC GGT GGT GCT CAG

lys leu cys ala gly val cys arg cys lys leu thr ser ser gly lys cys pro thr gly

AAG TTA TGG GCA GGC GTC TGT AGG TGT AAA CTC ACA AGT AGC GGA AAA TGC CCT ACA GGC

phe pro lys leu ala leu val ser asp glu pro asp thr val lys tyr cys asn

TTT CCC AAA TTG GCC CTT GTG TCC AAT TGA GAA GCA GCC ACC TCA GTG CAA GAC GAA GAA

leu gly cys arg ala ser met cys asp tyr met val asn ala ala ala asp glu leu

TGG GGG TGT AGG GCT TCC TGG TGG TGT AGG CTC AAT TGA AGT GAC TGC GCT GCT GCT GAC GAC GAA GAA

met lys tyr cys glu asp ala val asp gly asp ala ala cys ala val asp cys

TTG GGG TGT AGG GCT TCC TGG GAA AAT TGG CAA GAT GGC AAC TAT TCG CCA GCG

leu thr ser leu thr ser ter ter

GGC CTC ACA TCC CCT ACT GCC TAA TGA TGG TAT CCG AGT TTC CAA ACC TGG CTT CCG AGG TGT GTG CAT

poly A

ACC TTT GCT TCC TAT TAA AAA TGG GAT TCC CCA TCA GAG TAT TGG CAA ACC TGG CTT CCG AGG TGT GTG CAT

poly A

CAT CCT GTG GAT ATG AAA AAG CCG TCA TAA ATG AAT GAC AGT TGG CCT GCT GCA

(3')

**Fig. 3.** Nucleotide and deduced amino acid sequences of clone pTH-1. The beginning of the mature α-hordothionin (TH) and of the putative acidic protein (AP), as well as the polyadenylation signals (poly A) are indicated.

**The Thionins**

Thionins are polypeptides of mol. wt. ~5000, with high lysine and cystine contents, which belong to a homology group that also includes the visctoxtins from the mistletoes (Loranthaceae) and crambin from crambe (Cruciferae). This group of proteins has been thoroughly characterized at the structural level: many members of the group have been sequenced, detailed 1H NMR spectra at 600 MHz have been obtained, and the crystal structure of crambin has been determined at the highest resolution to date for any protein (for reviews see refs. 22-23). The thionins in cereal endosperm are of interest because they are lysine-rich proteins in a lysine-deficient tissue, and because, although orally inocuous, they are toxic to laboratory animals (intraperitoneally), cultured mammalian cells, yeasts, and phytopathogenic bacteria (see 22-23). Genetic and biochemical studies have shown that in hexaploid wheat and in rye there is one thionin variant per diploid genome, whereas in diploid barley, there is evidence for at
least a major and a minor component, respectively designated α- and β-hordothi-
nin. Genes encoding thionins have been located in the long arms of chromoso-
mes of group 1 (1A, 1B, 1D, of wheat; 1R of rye; see 22).

In vivo and in vitro synthesis of thionins in barley
Thionins are synthesized in developing barley endosperm from ∼8 days to ∼30
days after anthesis. Two types of thionin precursors, THP1 and THP2, have been
previously identified in vitro and in vivo, respectively, using monospecific
antibodies raised against the mature protein (24). The two precursors differed
in apparent mol. wt., as judged by SDS-electrophoresis, and in electrophoretic
mobility at acid pH, and were much larger than the final protein (24). The
mRNA encoding THP1 was obtained from membrane-bound polysomes. The co-transla-
tional conversion of THP1 into THP2 was proposed on the basis of our failure to
detect THP1 in vivo, whereas the conversion of THP2 into mature protein was
clearly shown in a pulse-chase experiment (24).

Characterization of a cDNA clone coding for α-hordothionin
The cDNA library prepared from developing barley endosperm described above was
screened for thionin recombinants following essentially the same strategy used
for the CM-proteins. Several clones were obtained and that with the largest
insert was selected for further characterization. The sequence of this insert
is presented together with the amino acid sequence deduced from the longest
reading frame in Fig. 3. The deduced sequence, which has 127 amino acids,
includes an internal sequence of 45 amino acids, which is identical to that
obtained by Ozaki et al. (25) for the major thionin (α-hordothionin) by direct
protein sequencing, and differs from that proposed by Mak (26) in three posi-
tions. The thionin sequence is preceded by a leader sequence of 18 residues
and followed by a sequence that corresponds to an acidic protein of 64 amino
acids. This structure is essentially in agreement with the previous evidence
indicating that thionin is synthesized as a much larger precursor which under-
goes two processing steps: the co-translational cleavage of a leader sequence
and the post-translational one of a larger peptide.

Acknowledgement

This research was supported by grant n° 1344 from the Comisión Asesora de Inves-
tigación Científica y Técnica.
References


