CM-PROTEINS AND THIONINS IN CEREALS: CHARACTERIZATION AND CLONING OF cDNA

J. Paz-Ares, F. Ponz, P. Rodriguez-Palenzuela, A. Lazaro, C. Hernandez-Lucas, P. Carbonero, F. García-Olmedo
Departamento de Bioquímica, E.T.S.Ingenieros Agrónomos-UPM, 28040 Madrid, Spain

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 α-amylase, and the thionins, a group of high-lysine toxic polypeptides.

The CM-Proteins: Trypsin and α-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 immunochemical properties (3-7). More recently, N-terminal amino acid sequences have

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

pUP-13

leu leu leu ala val leu thr thr val val ala thr ala↓glu arg asp tyr gly glu tyr
TTG TTG CTC GCT GTC CTC ACC ACC GTC GTG GCA ACT GCG GAA CGG GAC TAC GGC GAG TAC

cys arg val gly lys ser ile pro ile asn pro leu pro ala cys arg glu tyr ile thr
TGC CGC GTG GGG AAG TCG ATT CCC ATC AAC CCT CTC CCC GCT TGC CGA GAG TAC ATC ACG

arg arg cys ala val gly asp gln gln val pro asp val leu lys gln gln cys cys arg
CGC CGG TGC GCC GTC GGA GAC CAG CAG GTG CCG GAT GTC CTC AAG CAG CAG TGC TGC CGG

glu leu ser asp leu pro glu ser cys arg cys asp ala leu ser ile leu val asn gly
GAG CTC AGC GAC CTG CCG GAA AGT TGC CGG TGC GAT GCC CTG AGC ATC CTA GTG AAC GGC

val ile thr glu asp gly ser arg val gly arg met glu ala val pro arg cys asp gly
GTG ATC ACG GAG GAC GGC TCC AGG GTC GGC CGG ATG GAG GCG GTG CCG CGG TGT GAC GGG

glu arg ile his ser met gly ser tyr leu thr ala tyr ser glu cys asn pro his asn
GAG AGG ATC CAT TCC ATG GGG TCG TAT CTC ACG GCG TAT AGT GAG TGC AAT CCG CAC AAT

pro gly thr pro arg gly asp cys val leu phe gly gly gly ile ser ter
CCG GGT ACC CCT AGA GGG GAC TGC GTG CTG TTT GGT GGC GGC ATC AGT TAG TTAGCTCTAGG

poly A (AATAAA) ... poly A (AATAAA)
TAGTACTCAAATAAATGTTGCATGAGTCGATTGTGGTTGTGGTGCATGCATCCGTGGTATACAAAATAAAGGATGGAAA

GTCT

pUP-38

leu pro glu trp met thr ser ala glu leu asn tyr pro gly gln pro tyr leu ala lys
TTA CCC GAA TGG ATG ACA TCC GCG GAG CTG AAC TAC CCC GGG CAG CCA TAC CTC GCC AAG

leu tyr cys cys gln glu leu ala glu ile pro gln gln cys arn cys glu ala leu arg
TTG TAT TGT TGC CAA GAG CTT GCA GAA ATT CCC CAG CAG TGC CGG TGC GAG GCG CTG CGC

thr ser met ala leu pro val pro pro gln pro val asp pro ser thr gly asn val gly
ACT TCA A*TG GCG TTG CCG GTA CCG CCT CAG CCC GTG GAC CCG AGC ACC GGC AAT GTT GGT

gln ser gly leu met asp leu pro gly cys pro arg glu met gln arg asp phe val arg
CAG AGC GGC CTC ATG GAC CTG CCC GGA TGC CCC AGG GAG ATG CAA CGG GAC TTC GTC AGA

leu leu val ala pro gly gln cys asn leu ala thr ile his asn val arg tyr cys pro
TTA CTC GTC GCC CCG GGG CAG TGC AAC TTG GCG ACC ATT CAC AAC GTT CGA TAC TGC CCC

ala val glu gln pro leu trp ile ter
GCC GTG GAA CAG CCG CTG TGG ATC TAG TGATGATAAAATCAGTCGTTCGTGAATAAGCATGCATGTTGCG

TACATAGGCGTAGGCGTGTGCGTGTGGTGTGCATGTATGCATATGTGAGCTCCGCACGCTCAACATGTGTGGGCTATCT

poly A (AATAAA)
GCTATGAACGAGAATAAAGAGAACCATTTTGTGGTTCTTTAATTTCA$_{32}$

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.

```
CMe       FGDSCAPGDALPHNPLRACRTYVVSQICHQGPRLLTSDMKRR
pUP-13  ERDYGEYCRVGKSIPINPLPACREYI-TRRCAVGDQQVPDVLKQQ
pUP-38                        LPEWMTSAELNYPGQPYLAKLY

CMe     CCDELSAIPAYCRCEALRIIMQGVVTWQ------GAFEGAYFK-D
pUP-13  CCRELSDLPESCRCDALSILVNGVITED------GSRVGRME--A
pUP-38  CCQELAEIPQQCRCEALRTSMALPVPPQPVDPSTG-NVGQSGLMD

CMe     SPNCPRERQTSYAANLVTPQECNLGTIHGSAY-CPEL-QPGYG
pUP-13  VPRCDGERIHSMGSYLTAYSECNPHNPGTPRGDCVLF-GGGIS
pUP-38  LPGCPREMQRDFVRLLVAPGQCNLATIHNVRY-CPAVEQPLWI
```

Fig.2. Alignment of amino acid sequences deduced from clones pUP-13 and pUP-38 with the sequence of barley trypsin inhibitor CMe (19). Homologies are boxed.

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                                                                       TH
 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 CTG GGG TTG GTT CTC GAA CAG GTG CAA GTA GAA GGC AAG AGT

 cys cys arg ser thr leu gly arg asn cys tyr asn leu cys arg val arg gly ala gln
 TGC TGC AGG AGC ACC CTA GGA AGA AAC TGC TAC AAC CTT TGC CGC GTC CGT 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 TGC GCA GGC GTC TGT AGG TGT AAA CTC ACA AGT AGC GGA AAA TGC CCT ACA GGC
                 AP
 phe pro lys leu ala leu val ser asn ser asp glu pro asp thr val lys tyr cys asn
 TTC CCC AAA TTG GCC CTT GTG TCC AAC TCA GAT GAA CCA GAC ACC GTC AAG TAT TGC AAC

 leu gly cys arg ala ser met cys asp tyr met val asn ala ala ala asp asp glu glu
 TTG GGG TGT AGG GCT TCC ATG TGT GAC TAC ATG GTC AAC GCA GCT GCT GAC GAC GAA GAA

 met lys leu tyr leu glu asn cys gly asp ala cys val asn phe cys asn gly asp ala
 ATG AAA CTC TAT TTG GAA AAT TGT GGT GAT GCT TGT GTC AAT TTC TGC AAC GGT GAT GCT

 gly leu thr ser leu thr ala ter ter
 GGC CTC ACA TCC CTT ACT GCC TAA TGA TGTGTATCCATGGTCTGAGATTTCACAGGGCAAGGTGTGTCTG
                 poly A
 ACCTTTGCGTTCAATAAAATTGGATCCCATCGAGAGTATCCAAACCAGTGTGTCAACCTGTTTATGTGTGTGTATTTT
                  poly A
 CATTCCTTGTTCGAATAAAAGCCGTCATAATGAATGCCATGTTGCTGCCA41 (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 polyadenilation 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 viscotoxins from the mistletoes (Loranthaceae) and crambin from crambe (Cruciferae). This group of proteins has been thouroughly characterized at the structural level: many members of the group have been sequenced, detailed ^{1}HNMR 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-defficient 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-hordothionin. Genes encoding thionins have been located in the long arms of chromosomes 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-translational 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 positions. 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 undergoes 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 Investigación Científica y Técnica.

References

1. García-Olmedo, F., P. Carbonero, G. Salcedo, C. Aragoncillo, C. Hernandez-Lucas, J. Paz-Ares and F. Ponz. 1984. Chromosomal location and expression of genes encoding low molecular weight proteins in wheat and related species. Kulturpflanze 32, 21-32.

2. Lazaro, A., D. Barber, G. Salcedo, E. Mendez and F. García-Olmedo. 1985. Differential effects of high-lysine mutations on the accumulation of individual members of a group of proteins encoded by a disperse multigene family in the endosperm of barley (Hordeum vulgare L.). Eur. J. Biochem. 149, 617-623.

3. García-Olmedo, F. and P. Carbonero. 1970. Homeologous protein synthesis controlled by homeologous chromosomes in wheat. Phytochemistry 9, 1495-1497.

4. Aragoncillo, C., M.A. Rodriguez-Loperena, G. Salcedo, P. Carbonero and F. García-Olmedo. 1978. Influence of homoeologous chromosomes on gene-dosage effects in allohexaploid wheat (Triticum aestivum L.). Proc. Natl. Acad. Sci. USA 75, 1446-1450.

5. Aragoncillo, C., R. Sanchez-Monge and G. Salcedo. 1981. Two groups of low molecular weight hydrophobic proteins from barley endosperm. J. Exp. Bot. 32, 1279-1286.

6. Salcedo, G., R. Sanchez-Monge and C. Aragoncillo. 1982. The isolation and characterization of low molecular weight hydrophobic salt-soluble proteins from barley. J. Exp. Bot. 33, 1325-1331.

7. Paz-Ares, J., C. Hernandez-Lucas, G. Salcedo, C. Aragoncillo, F. Ponz and F. García-Olmedo. 1983. The CM-proteins from cereal endosperm: Immunochemical relationships. J. Exp. Bot. 34, 388-395.

8. Shewry, P.R., D. Lafiandra, G. Salcedo, C. Aragoncillo, F. García-Olmedo, E.J-L. Lew, M.D. Dietler and D.D. Kasarda. 1984. N-terminal amino acid sequences of chloroform/methanol-soluble proteins and albumins from endosperms of wheat, barley and related species. Homology with inhibitors of α-amylase and trypsin and with 2S storage globulins. FEBS Lett. 175, 359-363.

9. Barber, D., R. Sanchez-Monge, E. Mendez, A. Lazaro, F. García-Olmedo and G. Salcedo. 1985. New α-amylase and trypsin inhibitors among the CM-proteins of barley (Hordeum vulgare) (submitted).

10. Sanchez-Monge, R., D. Barber, E. Mendez, F. García-Olmedo and G. Salcedo. 1985. Genes encoding α-amylase inhibitors are located in the short arms of chromosomes 3B, 3D and 6D of wheat (Triticum aestivum L.) (submitted).

11. Buonocore, V., T. Petrucci and V. Silano. 1977. Wheat protein inhibitors of α-amylase. Phytochemistry 16, 811-820.

12. Aragoncillo, C., M.A. Rodriguez-Loperena, P. Carbonero and F. García-Olmedo. 1975. Chromosomal control of non-gliadin proteins from the 70 % ethanol extract of wheat endosperm. Theor. Appl. Genet. 45, 322-326.

13. Rodriguez-Loperena, M.A., C. Aragoncillo, J.V. Torres, P. Carbonero and F. García-Olmedo. 1975. Biochemical evidence of chromosome homoeology among related plant genera. Plant Sci. Lett. 5, 387-393.

14. Salcedo, G., C. Aragoncillo, M.A. Rodriguez-Loperena, P. Carbonero and F. García-Olmedo. 1978. Differential allelic expression at a locus encoding an endosperm protein in tetraploid wheat (Triticum turgidum). Genetics 89, 147-156.

15. Salcedo, G., P. Fra-Mon, J.L. Molina-Cano, C. Aragoncillo and F. García-Olmedo. 1984. Genetics of CM-proteins (A-hordeins) in barley. Theor. Appl. Genet. 68, 53-59.

16. Fra-Mon, P., G. Salcedo, C. Aragoncillo and F. García-Olmedo. 1984. Chromosomal assignment of genes controlling salt-soluble proteins (albumins and globulins) in wheat and related species. Theor. Appl. Genet. 69, 167-172.

17. Paz-Ares, J., F. Ponz, C. Aragoncillo, C. Hernandez-Lucas, G. Salcedo, P. Carbonero and F. García-Olmedo. 1984. In vivo and in vitro synthesis of CM-proteins (A-hordeins) from barley (Hordeum vulgare L.). Planta 157, 74-80.

18. Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. In: Methods in Enzymology, Vol. 65 (Grossman, L. and K. Moldave, eds.). Academic Press. pp. 499-560.

19. Odani, S., T. Koide and T. Ono. 1983. The complete amino acid sequence of barley trypsin inhibitor. J. Biol. Chem. 258, 7998-8003.

20. Heijne, G.V. 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. 133, 17-21.

21. Forde, B.G., M. Kreis, M.S. Williamson, R.P. Fry, J.Pywell, P.R. Shewry, N. Bunce and B.J. Miflin. 1985. Short tandem repeats shared by B- and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes. EMBO J. 4, 9-15.

22. García-Olmedo, F., P. Carbonero and B.L. Jones. 1982. Chromosomal locations of genes that control wheat endosperm proteins. In: Advances in Cereal Science and Technology, Vol. V (Y. Pomeranz, ed.). AACC, Inc., St. Paul, Mn (USA). pp. 1-47.

23. Ponz, F., J. Paz-Ares, C. Hernandez-Lucas, F. García-Olmedo and P. Carbonero. 1985. Barley α-hordothionin: cloning of cDNA and gene expression in a high-lysine mutant (submitted).

24. Ponz, F., J. Paz-Ares, C. Hernandez-Lucas, P. Carbonero and F. García-Olmedo. 1983. Synthesis and processing of thionin precursors in developing endosperm from barley (Hordeum vulgare L.). EMBO J. 2, 1035-1040.

25. Ozaki, Y., K. Wada, T. Hase, H. Matsubara, T. Nakanishi and H. Yoshizumi. 1980. Amino acid sequence of a purothionin homologue from barley flour. J. Biochem. (Tokyo) 87, 549-555.

26. Mak, A.S. 1975. Ph. D. Thesis. Univ. of Manitoba, Winnipeg, Manitoba (Canada).