Theor Appl Genet (1986) 71:842–846

Characterization of cDNA clones of the family of trypsin/α-amylase inhibitors (CM-proteins) in barley (*Hordeum vulgare* L.)

J. Paz-Ares *, F. Ponz **, P. Rodríguez-Palenzuela, A. Lázaro, C. Hernández-Lucas, F. García-Olmedo and P. Carbonero

Departamento de Bioquímica, E.T.S. Ingenieros Agrónomos, UPM, Ciudad Universitaria, E-28040 Madrid, Spain

Received September 15, 1985; Accepted October 28, 1985
Communicated by R. Riley

Summary. Recombinants encoding members of the trypsin/α-amylase inhibitors family (also designated CM-proteins) were selected from a cDNA library prepared from developing barley endosperm. Inserts in two of the clones, pUP-13 and pUP-38, were sequenced and found to encode proteins which clearly belong to this family, as judged from the extensive homology of the deduced sequences with that of the barley trypsin inhibitor CMe, the only member of the group for which a complete amino acid sequence has been obtained by direct protein sequencing. These results, together with previously obtained N-terminal sequences of purified CM-proteins, imply that there are at least six different members of this dispersed gene family in barley. The relationship of this protein family to the B-3 hordein and to reserve prolamins from related species is discussed in terms of their genome structure and evolution.

Key words: Barley endosperm – Trypsin/α-amylase inhibitors – CM-proteins – cDNA clones

Introduction

The albumins and globulins of cereal endosperm are made up of over 20 major and many minor components, as can be shown by two-dimensional electrophoresis or by high-performance liquid chromatography (see García-Olmedo et al. 1984; Lázaro et al. 1985).

It is becoming evident that a substantial fraction of these proteins form a homologous group, encoded by a multi-gene family which is dispersed over several chromosomes (García-Olmedo and Carbonero 1970; Aragoncillo et al. 1975; Paz-Ares et al. 1983a; Salcedo et al. 1984; García-Olmedo et al. 1984; Shewry et al. 1984; Lázaro et al. 1985; Sanchez-Monge et al. 1986; Barber et al. 1986). This group of proteins, which can be extracted by chloroform/methanol mixtures (CM-proteins) and which is also soluble in aqueous alcohols, includes inhibitors of trypsin and α-amylases of diverse origin, as well as proteins with no known in vitro activity (Shewry et al. 1984; Barber et al. 1986). It has been speculated that this group of proteins could be involved in the protection of the endosperm tissue and might be relevant in relation to its nutritive value (Buonocore et al. 1977).

In barley (*Hordeum vulgare* L.), one α-amylase and two trypsin inhibitors, and two CM-proteins of undefined activity, have been shown to be homologous on the basis of their N-terminal sequences and overall amino acid compositions (Barber et al. 1986). We report here the characterization of cDNA clones that correspond to members of this multi-gene family.

Materials and methods

Plant material

Hordeum vulgare cv. 'Bomi' was the gift of H. Doll (Risø National Laboratory, Denmark). Developing endosperms were collected at approximately 20 days post-anthesis, by mechanical extrusion into liquid nitrogen using the method of O'Dell and Thompson (1982).

Reagents

Oligo (dT)-cellulose was obtained from Collaborative Research. AMV-reverse transcriptase was from Life Sciences Inc. Terminal transferase and S_1-nuclease were obtained from P-L Biochemicals. Restriction endonucleases and other enzymes were supplied by Boehringer, Amersham, New England Biolabs or Bethesda Research Laboratories. |^{35}S|-cysteine, and

* Present address: Max-Planck-Institut für Züchtungsforschung, 5000 Köln 30, Federal Republic of Germany
** Present address: Department of Plant Pathology, University of California, Davis, CA 95616, USA

$|\alpha\text{-}^{32}P|$ dATP and $|\gamma\text{-}^{32}P|$ ATP were from New England Nuclear and Amersham, respectively.

RNA extraction and in vitro protein synthesis

Total polysomal RNA and membrane-bound polysomal RNA were prepared from endosperm tissue ground to a fine powder under liquid nitrogen, as described by Paz-Ares et al. (1983 b). The poli(A)$^+$RNA fraction, isolated by chromatography on oligo-dT-cellulose, was translated in a wheat germ cell-free system prepared by us from commercial wheat germ (General Mills, USA).

Size fractionation of CH_3HgOH-denatured RNA was carried out by sucrose gradient centrifugation. Gradient fractions were analysed by in vitro translation, using $|^{35}S|$cysteine, followed by SDS-polyacrylamide gel electrophoresis of total translation products and fluorography, as previously reported (Paz-Ares et al. 1983 b).

Construction of a cDNA library

The construction of the cDNA library was essentially as described by Maniatis et al. (1982). Poli(A)$^+$RNA obtained from developing barley endosperm was used as a template for the synthesis of a complementary DNA strand (cDNA). Double stranded DNA (ds-cDNA) was synthesized with the Klenow fragment of DNA-polI. After S_1-nuclease treatment, the ds-cDNA was size-fractionated by preparative electrophoresis in low-melting agarose, and the fraction of molecular weight higher than 350 bp was purified by the method of Langridge et al. (1980). This fraction was used for cloning in the *PstI* site of plasmid pBR322 after homopolymeric tailing (poly G/poly C). Cells of *Escherichia coli* strain Mc1061 made competent by the $CaCl_2$ procedure, were transformed with the recombinant plasmid and selected for tetracycline resistance. About 10^4 colonies obtained from approximately 300 ng of ds-cDNA were sensitive to ampicillin and resistant to tetracycline ($Amp^S Tet^r$).

Screening of the cDNA library

The $Amp^S Tet^r$ bacterial colonies were further screened in two steps. In the first step, colony hybridization was carried out according to the Grunstein and Hogness (1975) procedure. Probes were prepared by synthesis of radioactive cDNA with the AMV reverse transcriptase, from appropriate RNA gradient fractions. In the second step, two rounds of hybrid-release translation were performed with the clones selected in the first step.

Hybrid-selected translation

Pools of five clones were grown in LB medium and their plasmids purified according to Birnboim and Doly (1979). Plasmid DNA was then covalently bound to diazobenzyloxymethyl (DBM) paper (Christophe et al. 1982). The hybridization buffer was: 50% formamide, 800 mM NaCl, 100 mM Tris-HCl, pH 7.5, 2 mM EDTA, 0.4% SDS, 2 mg/ml yeast tRNA. The filters were preincubated in this buffer for 60 min at 37 °C. After incubation with total polysomal RNA (100–500 µg/ml) in the same buffer for the same period of time, the filters were washed with 150 mM NaCl and then the selected mRNA was finally released with buffer without salt, heating for 5 min at 100 °C and rapidly cooling at 0 °C. In vitro protein synthesis was carried out as indicated above.

DNA sequencing

The method of Maxam and Gilbert (1980) was used for DNA sequencing of subclones obtained in plasmids pUC-12 or pUC-13 (Messing 1983). Restriction maps were carried out by standard techniques (Maniatis et al. 1982).

Results

A cDNA library from developing endosperm was constructed in the pBR322/*E. coli* Mc1061 vector system using poly(A)$^+$RNA from total polysomal RNA as starting template. A preliminary screening of this library for recombinants corresponding to the CM-proteins family was carried out with a radioactive cDNA probe. This probe was prepared from a gradient fraction that was enriched in mRNA for these proteins, as described by Paz-Ares et al. (1983 b). About 20% of the colonies gave both a positive signal with this probe and a negative one with a second probe which was similarly prepared from a pool of RNA fractions of higher sedimentation velocity (mRNA's for B- and C-hordeins and ribosomal RNA). Recombinants selected in this preliminary screening were further tested by hybrid-selected translation, electrophoresis and fluorography, first in pools of five clones and then as individual ones, as shown in Fig. 1. Clones pUP-13 and pUP-38 selected mRNA's that encoded proteins which were efficiently

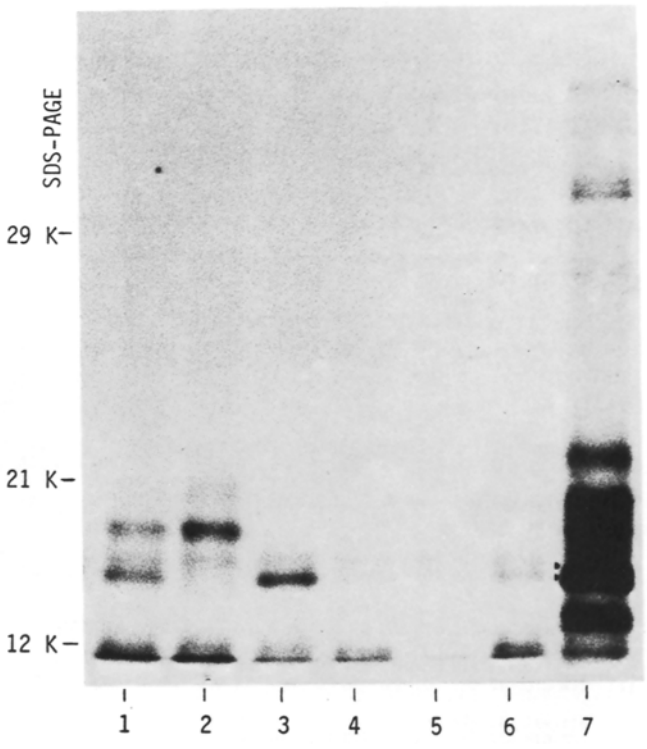

Fig. 1. Hybrid-release translation of messengers selected by a pool of five clones (1) and the individual clones from the same pool (2–6). *Lanes 2* and *3* correspond to clones pUP-38 and pUP-13, respectively *7*). Translation products from total poly (A)$^+$RNA. $|^{35}S|$ cysteine was used as precursor in all cases

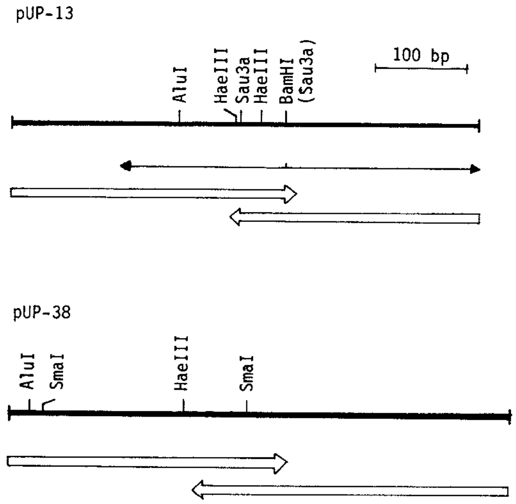

Fig. 2. Restriction maps and sequencing strategies for clones pUP-13 and pUP-38. *Double-lined arrows* indicate that both strands have been sequenced in pUC-12/pUC-13 subclones; *single-lined arrows* indicate that one strand was sequenced in the original clone

labeled with [^{35}S]cysteine, had apparent sizes within the range observed for the precursors of CM-proteins, and gave a weak antigenic reaction with monospecific antibodies raised against protein CMd (not shown). The inserts in these two clones were sequenced according to Maxam and Gilbert (1980), following the sequencing strategies outlined in Fig. 2. The nucleotide sequences are presented together with the amino acid sequences deduced from their longest reading frames in Fig. 3. The cDNA in clone pUP-13 encodes a sequence of 136 amino acids in which, following the criteria of Heijne (1983), the hydrophobic core and the cleavage site (after position 13) of a signal peptide can be discerned in the N-terminal domain, leaving a mature protein of 123 amino acids. This is in agreement with previous evidence from experiments of in vivo and in vitro synthesis, which indicated that CM-proteins were synthesized by membrane-bound polysomes as larger precursors that seemed to be co-translationally processed (Paz-Ares et al. 1983b). The cDNA in clone pUP-38 does not encode the N-terminal end of the corre-

```
pUP-13

     leu leu leu ala val leu thr thr val val ala thr ala▽glu arg asp tyr gly glu tyr   20
1    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   40
61   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   60
121  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   80
181  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  100
241  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  120
301  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  *               136
361  CCG GGT ACC CCT AGA GGG GAC TGC GTG CTG TTT GGT GGC GGC ATC AGT TAG TTAGCTCTA
                 poly A   1                   2          3       4     5
421  GGTAGTACTCAAATAAATGTTGCATGAGTCGATTGTGGTTGTGGTGCATGCATCCGTGGTATACAA
     poly A
487  AATAAAGGATGGAAAGTCT

pUP-38

     leu pro glu trp met thr ser ala glu leu asn tyr pro gly gln pro tyr leu ala lys   20
1    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 arg cys glu ala leu arg   40
61   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   60
121  ACT TCA ATG 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   80
181  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  100
241  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  *   *   *   *                                  108
301  GCC GTG GAA CAG CCG CTG TGG ATC TAG TGA TGA TAA AATCAGTCGTTCTGTAATAAGCATGC
        1                      2       3      4                  5
363  ATGTTGCGTACATAGGCGTAGGCGTGTGCGTGTGGTGTGCATGTATGCATATGTGAGCTCCGCACG
                                                  poly A
429  CTCAACATGTGTGGGCTATCTGCTATGAACGAGAATAAAGAGAACCATTTTGTGGTTCTTTAATTT

495  CA32
```

Fig. 3. Nucleotide sequences and deduced amino acid sequences of clones pUP-13 and pUP-38. The *vertical arrow* (▽) indicates the site of cleavage of the leader sequence deduced according to Heijne (1983). Segments numbered 1–5 indicate homologies in the non-coding 3′-ends. *Horizontal arrows* indicate repeats. Poly A indicates polyadenilation signals. The base marked with a dot (●) has not been corroborated in both strands. Stop signals are indicated with asterisks (*)

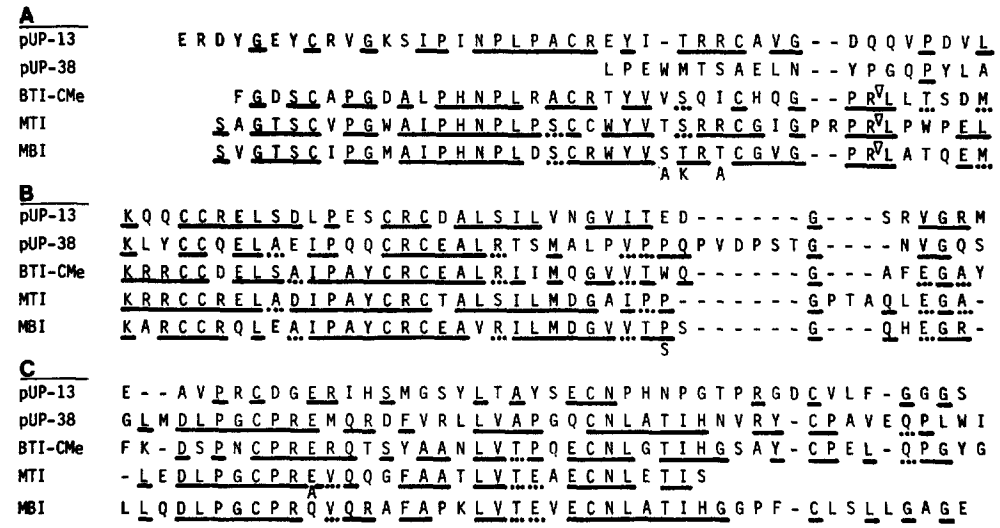

Fig. 4. Alignment of amino acid sequences deduced from clones pUP-13 and pUP-38 with those of barley trypsin inhibitor CMe (BTI-CMe; Odani et al. 1983), maize trypsin inhibitor (MTI; Mahoney et al. 1984) and millet bifunctional inhibitor (MBI; Campos and Richardson 1983). Gaps have been introduced to optimize alignment. Homologies have been underlined (—— or . . .). The sequences have been divided into the A, B, and C domains defined by Kreis et al. (1985)

sponding protein and the deduced amino acid sequence is different, though partially homologous, to that of clone pUP-13. As shown in Fig. 3, the non-coding 3′-ends of both clones share five short sequences, which appear in the same order but not at the same distance from each other, and two sets of similar, but not identical, tandem repeats. Other relevant features of this region are not shared by the two cDNAs, namely, two polyadenilation signals and one set of inverted repeats in pUP-13, and four stop signals in tandem and two overlapping sets of inverted repeats in pUP-38.

Discussion

In Fig. 4, the amino acid sequences deduced from clones pUP-13 and pUP-38 have been aligned with those obtained by direct protein sequencing for the barley trypsin inhibitor CMe (Odani et al. 1983), a maize trypsin inhibitor (Mahoney et al. 1984) and the α-amylase/trypsin bifunctional inhibitor from millet (Campos and Richardson 1984). This comparison clearly shows that the clones obtained correspond to the CM-proteins (or trypsin/α-amylase inhibitors) gene family. The sequences have been displayed in three parts (N-terminal, central, and C-terminal), which correspond approximately to the A, B, and C domains of the barley trypsin inhibitor CMe defined by Kreis et al. (1985). Five different N-terminal (A-domain) sequences of barley CM-proteins (including CMe) have been determined by direct protein sequencing (Shewry et al. 1984; Barber et al. 1986). The N-terminal deduced from pUP-13 is homologous but not identical to any of them, which implies that there are at least six different members of this dispersed multi-gene family in barley. The A domain of the pUP-13 protein is closer to those of the inhibitors of maize and millet than to those of the other barley CM-proteins. The anti-trypsin reactive site does not seem to be present in either of the deduced sequences, although, in the case of pUP-38, it can be speculated that most of the A domain is missing, and that the non-homologous terminal sequence of 15 amino acids is in fact an insertion between the A and B domains, which would be in line with the higher molecular weight of the protein selected by this clone (Fig. 1). The homologies of the pUP-13 and pUP-38 protein sequences with respect to each of the three inhibitors are very similar in the B region, whereas in the C region, the pUP-13 sequence shows significantly more divergence than that of pUP-38 (i.e., 11 versus 21 coincident positions with respect to CMe).

A weak homology has been reported by Kreis et al. (1985) between the A, B, and C domains of trypsin inhibitor CMe and α-gliadin and γ-secalin. It is of interest to note that the A and B domains of the pUP-13 protein and the B domain of that encoded by pUP-38 are closer than CMe to the above-mentioned reserve proteins. Furthermore, the A domain of pUP-13 is significantly closer to the corresponding region of the B-3 hordein reported by Forde et al. (1985) than to the same regions of α-gliadin and γ-secalin (sequences not shown). The homologies of the C domains of the proteins in Fig. 4 with respect to the above-mentioned reserve proteins are very weak in all cases. Genes encoding CM-proteins in barley have been found in chromosomes 1, 3, and 4 (Salcedo et al. 1984), so the relationship to the B-3 hordein (chromosome 5) implies that the ancestral sequence must have been dispersed over at least four of the seven barley chromosomes.

We have previously presented evidence of considerable divergence within this family at the levels of structure, in vitro activity and genetic regulation (Lázaro et al. 1985; Barber et al. 1985). A comparison of the 3′-end, non-coding regions of the two clones shows that although some homology has been conserved, sequence divergence is considerable and some relevant features, such as the invert repeats, are not shared.

Acknowledgements. We would like to thank Drs. C. Escarmís and A. González for useful advice with the DNA sequencing, and C. Rojas and L. Lamoneda for technical assistance. Financial support from the Comisión Asesora de Investigación Científica y Técnica (grant no. 1344) is also acknowledged.

References

Aragoncillo C, Rodriguez-Loperena MA, Carbonero P, García-Olmedo F (1975) Chromosome control of non-gliadin proteins from the 70% ethanol extract of wheat endosperm. Theor Appl Genet 45:322–326

Barber D, Sanchez-Monge R, Mendez E, Lázaro A, García-Olmedo F, Salcedo G (1986) New α-amylase and trypsin inhibitors among the CM-proteins of barley (*Hordeum vulgare* L.). Biochim Biophys Acta 869:115–118

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl Acids Res 7:1513–1519

Buonocore V, Petrucci T, Silano V (1977) Wheat protein inhibitors of α-amylase. Phytochemistry 16:811–820

Campos FAP, Richardson M (1984) The complete amino acid sequence of the α-amylase inhibitor I-2 from seeds of ragi (Indian finger millet, *Eleusine coracana* G.). FEBS Lett 167:221–223

Christophe D, Brocas H, Vassart G (1982) Improved synthesis of DBM paper. Anal Biochem 120:259–261

Forde BG, Kreis M, Williamson MS, Fry RP, Pywell J, Shewry PR, Bunce N, Miflin BJ (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

García-Olmedo F, Carbonero P (1970) Homeologous proteins synthesis controlled by homeologous chromosomes in wheat. Phytochemistry 9:1495–1497

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

Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain specific genes. Proc Natl Acad Sci USA 72:3961–3965

Heijne GV (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21

Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183:499–502

Langridge J, Langridge P, Bergquist PL (1980) Extraction of nucleic acids from agarose gels. Anal Biochem 103: 264–271

Lázaro A, Barber D, Salcedo G, Mendez E, García-Olmedo F (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

Mahoney WC, Hermodson MA, Jones B, Powers DD, Corfman RS, Reeck GR (1984) Amino acid sequence and secondary structural analysis of the corn inhibitor of trypsin and activated Hageman factor. J Biol Chem 259: 8412–8416

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base specific chemical cleavages. In: Grossman L, Moldave K (eds) Methods in enzymology, vol 65, pp 499–560

Messing J (1983) New M13 vectory for cloning. In: Wu R, Grossman L, Moldave K (eds) Methods in enzymology, vol 101, pp 20–78

O'Dell M, Thompson RD (1982) The purification of immature barley endosperm. J Sci Food Agric 33:419–420

Odani S, Koide T, Ono T (1983) The complete amino acid sequence of barley trypsin inhibitor. J Biol Chem 258: 7998–8003

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

Paz-Ares J, Ponz F, Aragoncillo C, Hernandez-Lucas C, Salcedo G, Carbonero P, García-Olmedo F (1983b) In vivo and in vitro synthesis of CM-proteins (A-hordeins) from barley (*Hordeum vulgare* L.). Planta 157:74–80

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

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

Sanchez-Monge R, Barber D, Mendez E, García-Olmedo F, Salcedo G (1986) Genes encoding α-amylase inhibitors are located in the short arms of chromosomes 3B, 3D and 6D of wheat (*Triticum aestivum* L.). Theor Appl Genet (in press)

Shewry PR, Lafiandra D, Salcedo G, Aragoncillo C, García-Olmedo F, Lew EJ-L, Dietler MD, Kasarda DD (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