 © Springer-Verlag 2001

O.D. Anderson · C.C. Hsia · A.E. Adalsteins
E.J.-L. Lew · D.D. Kasarda

Identification of several new classes of low-molecular-weight wheat gliadin-related proteins and genes

Received: 20 October 2000 / Accepted: 20 November 2000

Abstract During the initial phases of a wheat endosperm Expressed-Sequence-Tag (EST) project, several clones were determined to be related to wheat gliadin sequences, but not similar enough to be classified into any of the traditional gliadin families [α-, γ-, and ω-gliadins, low-molecular-weight (LMW) glutenins]. Complete sequences of these cDNA clones revealed four new classes of gliadin-related endosperm proteins, but lacking a prominent repeat domain which until now has been characteristic of the gliadins. Two of these classes are related to different minimally described groups of Triticeae endosperm proteins. One class of proteins, which has N-terminal amino-acid sequences matching members of a reported 25-kDa globulin family from wheat, is shown by amino-acid sequencing to match to a family of 25-kDa endosperm proteins, is encoded by a multigene family, and is most similar to the LMW-glutenins. A second new class shows N-terminal homologies to LMW secalins from rye, and has an amino-acid composition similar to wheat and barley LMW proteins with extraction properties similar to prolamins. The third class is most similar to α-gliadins, and the fourth class has no close association to previously described wheat endosperm proteins.

Keywords Wheat · Barley · Rye · Gliadin · Globulin

Communicated by J.S. Beckmann

O.D. Anderson (✉) · C.C. Hsia · A.E. Adalsteins · E.J.-L. Lew
D.D. Kasarda
Crop Improvement and Utilization Research Unit,
Western Regional Research Center,
Agricultural Research Service,
U. S. Department of Agriculture, 800 Buchanan Street,
Albany, CA 94710, USA
e-mail:oandersn@pw.usda.gov
Tel: 510-559-5773, Fax: 510-559-5777

Introduction

The seed storage proteins have been traditionally divided into specific fractions dependent upon the conditions of extraction and solubility (Osborne 1924): in water or dilute salts, albumins; in dilute salts, but insoluble in water, globulins; in alcohol-water mixtures, prolamins; and in dilute acid or alkali, glutelins. In wheat, the soluble prolamins comprise about 53%, the glutelins (largely insoluble prolamins, or glutenins) about 32%, and the globulins and albumins about 15% of the total endosperm protein (Wieser et al. 1998). The wheat prolamins appear to be distantly related to the amylase-trypsin inhibitor gene families (including CM proteins), and the 2S albumins from dicots (Shewry 1995). In addition, there have been reports of wheat and barley endosperm proteins with solubility properties similar to the gliadins (Aragoncillo et al. 1981; Prada et al. 1982), but with lower molecular weights than the major gliadin and hordein protein families. These smaller proteins are poorly characterized and have not yet been classified with respect to other endosperm proteins.

Although the classical prolamin polypeptides and genes have been extensively studied, less is known of the endosperm albumins and globulins. Cole et al. (1981) estimated that the albumins and globulins comprise 13% and 2%, respectively, of the total endosperm protein of a hard red winter wheat (cv Scout 66). Two types of globulin-like wheat seed proteins have been described, the 25-k globulins (Gomez et al. 1988, 1991) and the triticins (Singh and Shepherd 1985). These have been mapped to the *Glo-1* and *Tri-1* loci, respectively, on the short arms of the group-1 homoeologous chromosomes. This is the same chromosomal location as the γ- and ω-type gliadins (*Gli-1*), and low-molecular-weight (LMW) glutenins (*Glu-3*) (Payne 1987). Loci on the short arm of chromosome 1 from *Triticum monococcum* are in the relative order *Gli-A3* – *Glu-A3* – *Gli-A1* – *Glo-A1*, with the *Glo-A1* locus at the distal end of the short arm of chromosome 1A (Nieto-Taladriz et al. 1996). A partial coding sequence of a triticin has been reported

(Singh et al. 1993) and there has been only an N-terminal peptide sequence for a single 25-kDa globulin (Gomez et al. 1991).

We have wondered for some time if the current view of the defined sets of wheat endosperm proteins, focusing on the classical groupings because of their relative abundance, has missed additional groups with fewer gene copies and/or lesser amounts of expressed protein. In addition, a detailed characterization of the minor endosperm protein classes mentioned above is lacking. Clearly, there are still gaps in our understanding of the complete complement of cereal endosperm proteins.

ESTs (Expressed Sequence Tags) are DNA sequences (usually 200–500 bp) that ‘‘tag’’, or identify, specific genes expressed in a particular tissue under a given set of conditions. During a wheat endosperm EST project, a number of sequences were found to be related to gliadins and avenins (oat prolamins), but were not sufficiently close to classify them within those groups. These sequences were similar to the classical prolamins except that they lack the characteristic large repetitive domain. This DNA sequence data and the examination of a specific fraction of endosperm proteins led to the current report that wheat endosperm contains at least four, and probably more, subfamilies of LMW proteins closely related to the gliadins.

Materials and methods

DNA isolation, sequencing, and analysis

RNA was prepared from wheat cultivar Cheyenne endosperm 5–30 days after flowering (prepared by C. Tanaka, ARS-USDA), and a cDNA library was constructed in lambda-Zap II (Stratagene). Excision libraries were prepared according to the company's instructions for propagation as plasmids. Random cDNA colonies were picked, grown in 5 ml of Luria broth, and DNA prepared using Qiagen Qiaprep kits. Single-pass sequencing on a set of 800 clones was performed at several sites: Stanford DNA Sequencing Center (Nancy Federspiel), Montana State (Tom Blake), and in our laboratory. Clones corresponding to ESTs selected as representing interesting candidate genes were sequenced completely in both directions using an ABI310 DNA Sequencer and Big-Dye Terminator chemistry. Analysis of DNA sequences and phylogenetic comparisons were carried out using the Editseq, Mapseq, and Megalign modules of Lasergene software (DNAstar). Sequence matching to known sequences was done using web-based BLAST facilities of the National Center for Biotechnology Information. Depending on the context, we refer to both the original clone and the EST by the same name; i.e., EST 07h10 is the single-pass sequence and the clone 07h10 is the cDNA clone from which the complete sequence was determined.

Protein extraction, purification, and sequencing

A 50-g portion of flour (endosperm) from the wheat cv Yecora Rojo was extracted with 500 ml of water (magnetic stirring) for 15 min. The suspension was allowed to settle and the supernatant was discarded. The flour residue was then extracted similarly with 500 ml of a 40% ethanol-water solution, centrifuged at 10,000 *g* and the resulting flour residue retained. The residue remaining from the two prior extractions was then extracted with 200 ml of 0.1 M acetic acid for 1 h and the supernatant solution clarified by centrifugation at 10,000 *g*, and the clarified solution was lyophilized.

About 120-mg portions of the crude acetic acid extract were dissolved in 5 ml of 0.1 M acetic acid and fractionated on a Bio-Gel P-100 (BioRad) column (90 cm × 2.5 cm) equilibrated with 0.1 M acetic acid. The size fractions corresponding to masses of approximately 25 kDa were freeze-dried and screened by reversed-phase HPLC (RP-HPLC) with a Hewlett-Packard model 1090 apparatus fitted with a Vydac C_{18} semipreparative column equilibrated to 5°C. Samples from gel filtration were dissolved in 30% aqueous acetonitrile, 0.05% trifluoroacetic acid (TFA). The separation was achieved with a gradient ranging from 10% acetonitrile-water through 55% acetonitrile-water in 35 min (all solutions 0.05% in TFA). Elution was monitored by UV absorption at 210 nm. Peaks of interest were collected and analyzed by SDS-PAGE using Novex (Invitrogen) NuPAGE 4–12% Bis-Tris gels according to the procedure of Kasarda et al. (1998).

N-terminal amino-acid sequencing was carried out with an Applied Biosystems (PE-Applied Biosystems) model 477A pulsed liquid phase amino-acid sequencer and a Model 120A amino-acid analyzer as described by Lew et al. (1992).

Protein characterization

Purified protein was reduced in 50 mM of Tris-HCl buffer (pH 8) containing 6 M guanidinium hydrochloride, and free sulfhydryl groups were alkylated before proteolytic digestion or cyanogen bromide cleavage was performed. The solution was flushed with nitrogen, the reductant dithiothreitol was added at a ratio to cysteine of approximately 20:1, and the solution was again flushed with nitrogen, and reduction proceeded at room temperature for 2–4 h. Alkylation was then carried out by adding 4-vinypyridine at a ratio of about 2:1 relative to the molar amount of dithiothreitol, allowed to react in the dark for 2 h, and the reaction stopped by acidification with a small amount of glacial acetic acid. The alkylated protein was re-purified by RP-HPLC to remove reagents.

For trypsin digestion, reduced and alkylated protein was dissolved in 0.1 M Tris-HCl, pH 8, and trypsin (Sigma, Type XIII, TPCK treated), dissolved in 1 mM HCl, was added at a ratio of approximately 1:20 (enzyme:protein). Digestion was allowed to occur at room temperature (approximately 22°C) for 4 h, and then stopped by acidification.

For chymotrypsin digestion, reduced and alkylated protein was dissolved in 25 mM ammonium bicarbonate buffer (pH 7.9) and chymotrypsin (Boehringer Mannheim, stock number 103314), dissolved in 1 mM HCl, was added at a ratio of approximately 1:20 (enzyme:protein). Digestion was for 24 h at room temperature.

For endoproteinase Glu-C digestion, reduced and alkylated protein was dissolved in 25 mM ammonium bicarbonate buffer (pH 7.9) and endoproteinase Glu-C (Boehringer Mannheim, sequencing grade), dissolved in de-ionized water (50 mg in 50 ml), was added to yield a protein:enzyme ratio of about 10:1. Digestion was allowed to take place for 3 h at room temperature, and the reaction stopped by acidification.

For cyanogen bromide cleavage, an N-terminal peptide of the 25-k protein (recognized from its sequence) derived from the trypsin digestion and recovered from a single peak of the RP-HPLC separation was dissolved in 100 µl of 70% formic acid. The protein was then cleaved at methionine residues by adding 10 µl of 5 M cyanogen bromide dissolved in acetonitrile (Aldrich Chemical). The solution was tightly capped, mixed, and allowed to react overnight in the dark. Cleavage products were fractionated by RP-HPLC as above, yielding two peaks.

The resulting peptides from the above digestions and cleavages were fractionated by RP-HPLC and peaks collected for protein sequencing.

Results and discussion

Analysis of ESTs

In the initial phase of a wheat endosperm EST project, 800 randomly selected wheat endosperm cDNA clones were single-pass sequenced and the results analyzed for potential matches against known sequences. As expected, numerous high-probability matches to the known major wheat endosperm protein families were found; i.e., α-, γ- and ω-gliadins, and LMW- and HMW-glutenin subunits. The members of these known families are easily distinguished by BLAST probability scores. For example, BLASTN (nucleotide) searches with the α-gliadin CNN113 (Anderson et al. 1997) coding sequence found matches to GenBank α-gliadin entries with probabilities of matches being by chance from 0 to $2e^{-70}$, then an abrupt transition to e^{-3} and higher probabilities for matches to other wheat and non-wheat prolamins (data not shown). A BLASTX (amino-acid) search gave similar results, although the transition from α-gliadin matches to matches with the other prolamin types was not as abrupt – with probabilities of 0 to $2e^{-37}$ for α-gliadins, and then a transition to $2e^{-19}$ and higher for matches to several γ-gliadins and other prolamins.

In contrast, several ESTs showed a low to moderate level of matching to the known prolamin families, but not high enough to indicate membership in any specific known family. In our initial wheat EST project, cDNAs were not subject to any prior selection against known gene families. Therefore, we had reasoned that if there were additional prolamin families, or related families not yet identified, they would show up as BLAST results with lower or intermediate matches relative to known prolamins. In examining the BLAST results from the first 800 ESTs, several candidate ESTs were thus identified and are shown in Table 1. The best BLASTN and BLASTX matches of all these ESTs were to wheat prolamins and prolamins of other monocot cereals, particularly the avenins of oats. However, most of the BLASTN probabilities ($4e^{-6}$ to $6e^{-2}$) were well above what would usually be considered indicative of gene-family relatedness. All of the ESTs showed more significant BLASTX than BLASTN matches, the highest match for each ranging from $4e^{-13}$ to $5e^{-23}$, again with avenins being generally the most-significant matches.

EST sequences and encoded proteins

The five cDNAs corresponding to the ESTs shown in Table 1 were sequenced and the complete sequences are given in Fig. 1. Clones 07h10, 11dc7, and 09d3 contain complete open reading frames, while clones 07e6 and 12dc3 are missing the 5′ portion of their coding regions. The five cDNAs represent four distinct classes of sequence that share similarities with prolamins. Clones 07h10 and 07e6 are paired since they differ over their available DNA sequences only by 17 base changes, deletions of three and eight bases in 07e6 and an additional single base insertion in 07e6. The two derived amino-acid sequences differ at only five positions.

Table 1 EST sequences matched to GenBank sequences using BLAST. EST clone names are shown along with best nucleic-acid (BLASTN) and amino-acid (BLASTX) matches for sequences in Genbank and the probabilities of the those matches being significant. Lower scores indicate higher significance, with e^{-5} being typically considered the approximate higher limit for a meaningful match

EST	BLASTIN	Probability	BLASTX	Probability
07e6	α-Gliadin	$1.2e^{-2}$	Avenin	$6e^{-17}$
			γ-Gliadin	$3e^{-16}$
07h10	α-Gliadin	$4e^{-6}$	Avenin	$5e^{-23}$
	Avenin	$4e^{-6}$	γ-Gliadin	$4e^{-20}$
			γ-Hordein	$8e^{-19}$
09d3	Avenin	$5e^{-12}$	α-Gliadin	$5e^{-23}$
	LMW-glutenin	$2e^{-11}$	γ-Gliadin	$7e^{-18}$
	α-Gliadin	$7e^{-11}$	Avenin	$5e^{-17}$
11dc7	Avenin	$6e^{-2}$	Avenin	$4e^{-13}$
			Rice prolamine	$9e^{-11}$
			γ-Gliadin	$1e^{-10}$
12dc3	Avenin	$7e^{-3}$	Avenin	$4e^{-13}$
			γ-Gliadin	$8e^{-11}$
			γ-Secalin	$7e^{-8}$

Assuming the cleavage point of their signal peptides is as shown Fig. 1 (based on relations to previously determined N-terminal sequences as described below), the molecular weights of the mature polypeptides encoded by clones 07h10, 09d3 and 11dc7 are 19,734, 17,896 and 16,306 Daltons, respectively. These values would be similar to the molecular weights of the α- and γ-gliadins, and the LMW-glutenins without their repetitive domains (17,000–19,000 kDa).

The five new derived amino-acid sequences were compared to three known protein groups: prolamins including the wheat gliadin-type prolamins (α- and γ-gliadins, LMW-glutenin subunits) plus oat avenins; wheat puroindolines/grain softness proteins (GSP); and wheat α-amylase-trypsin inhibitors/CM proteins (Fig. 2). The five new sequences fall within or close to the branch containing the prolamins, with 07h10 and 07e6 most similar to LMW-glutenins, 09d3 most similar to α-gliadins, and 11dc7 and 12dc3 more distantly related to any of the known gliadins but still part of the gliadin branch in this analysis. The three major clusters of sequences in the tree shown in Fig. 2 remain associated with variations of parameters and using different sets of sequences, although the relative branching pattern of the three clusters changes with respect to each other in some analyses (data not shown). The clustal comparison gave similar, but not identical, relatedness within the prolamin branch when compared to the BLAST analysis (Table 1). The differences are relatively small and attributable to the different algorithms, and the close complex relationships within the gliadin family.

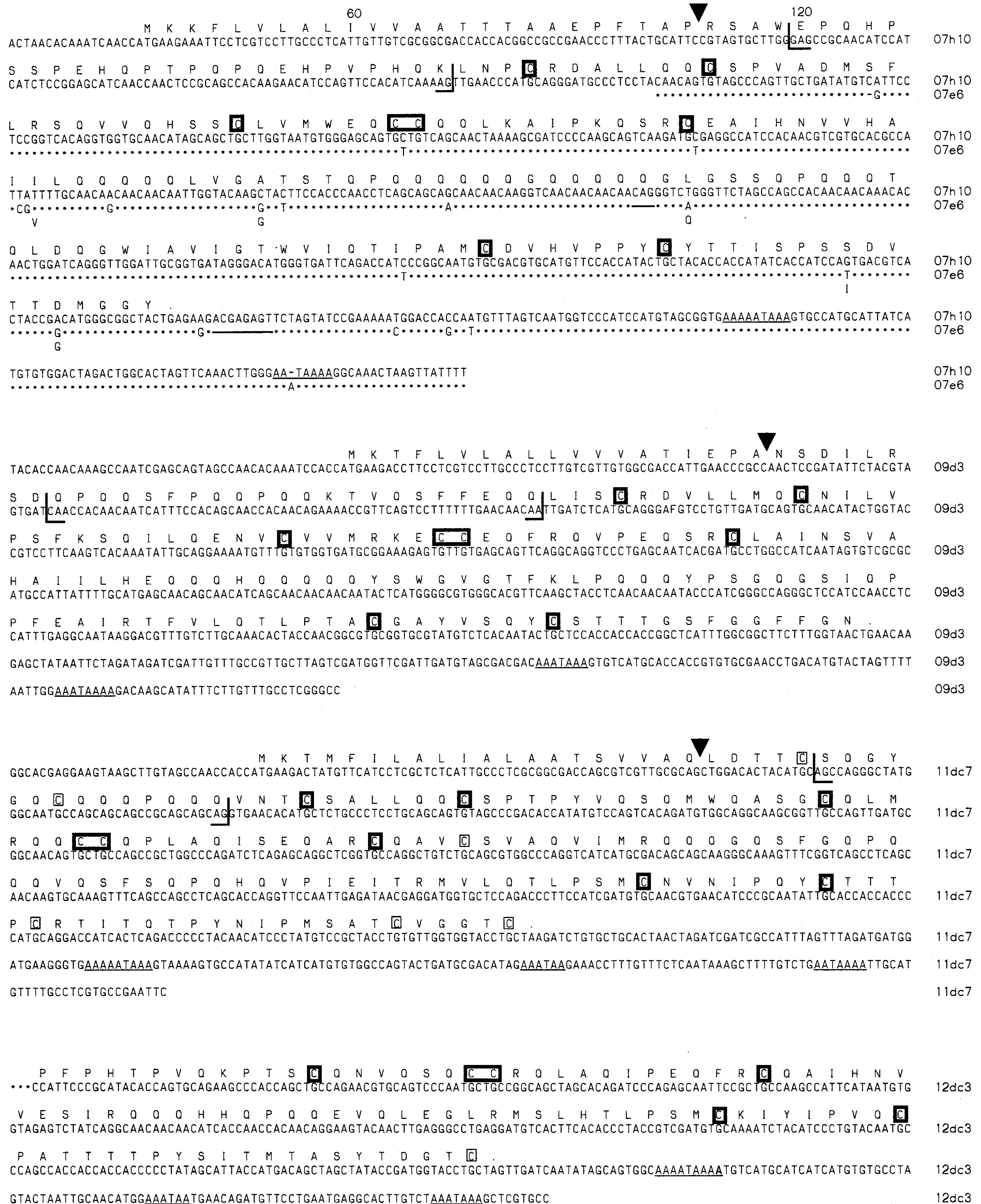

Fig. 1 DNA sequences of wheat endosperm gliadin-like cDNAs. The DNA and derived amino-acid sequences from five different gliadin-like DNA sequences are shown. Clones 07h10 and 07e6 are very similar and are aligned. The derived amino-acid sequence of 07h10 is shown above the DNA sequence and differences from the 07e6 amino-acid sequence are shown below the 07e6 DNA sequence. Cysteine residues are *boxed*: *thick boxes* are the eight cysteine residues conserved in several prolamin families, and *thin boxes* are non-conserved cysteine residues. *Arrowheads* indicate putative positions of the start of the mature polypeptides. *Half-brackets* below the DNA sequences indicate domains related to large gliadin repetitive domains. Polyadenylation elements are *underlined*

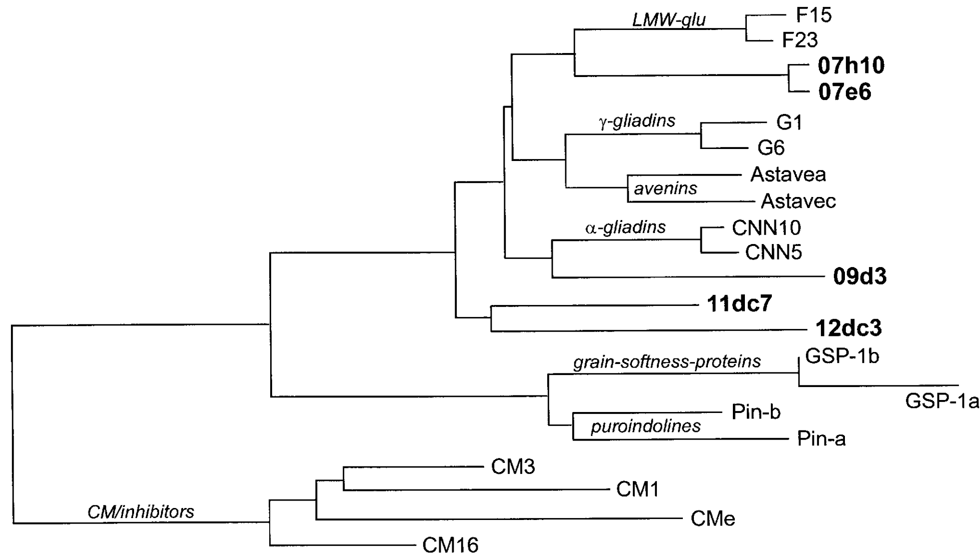

Fig. 2 Phylogenetic relationships among cDNA encoded proteins and related endosperm proteins. The hypervariable repetitive domains of the prolamins were removed before the analysis. Analysis was performed using the Megalign module of the Lasergene software package (DNAstar) and the Clustal method (Higgins and Sharp 1989). Sequence sources: LMW-glutenins (Cassidy et al. 1998), γ-gliadins (Anderson et al. 2001), avenins (Chesnut et al. 1989), α-gliadins (Anderson et al. 1997), grain-softness-proteins (GSP; Rahman et al. 1994), puroindolines (Gautier et al. 1994; Tranquilli et al. 1999), CM/inhibitors (Garcia-Maroto et al. 1990; Rodriguez-Palenzuela et al. 1989). The clone names are shown for the wheat prolamins, protein subclass names for the CM proteins, locus names for the puroindolines and GSPs, and Genbank names for the avenins

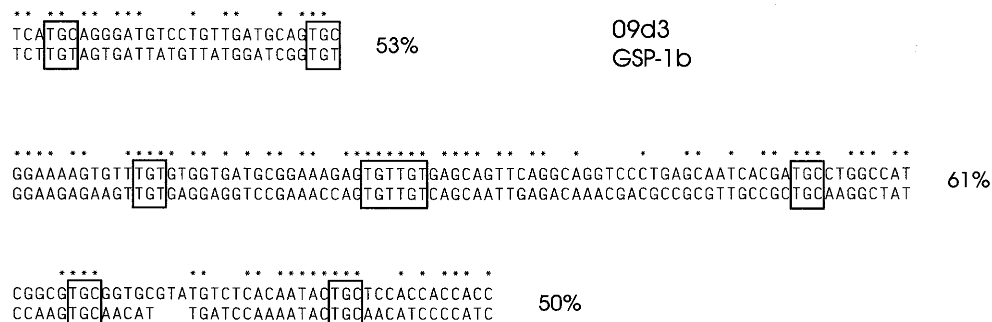

Fig. 3 Comparison of nucleotide sequences of clone 09d3 and *GSP-1b*. Conserved cysteine residue encoding regions are compared between the clone 09d3 DNA sequence and *GSP-1b* (Rahman et al. 1994). Asterisks above the sequences indicate conservation of bases. Cysteine codons are enclosed in *boxes*. The percentage of matching nucleotides is indicated to the right for each comparison: random sequences would be expected to match at 25% of base positions

The first three of the new classes have a conserved eight-cysteine residue pattern (boxed in bold in Fig. 1) common to the prolamins of the Triticeae. The fourth class, represented by 12dc3, contains the final six conserved cysteines in its partial sequence. Clone 12dc3 has at least one additional cysteine residue, whereas clone 11dc7 encodes five additional cysteine residues.

Although the puroindolines/GSP sequences are less commonly associated with the gliadins, the similar pattern and spacing of cysteine residues and the similarity in DNA sequence over the region encoding the conserved cysteines (Fig. 3) suggest a common origin for 09d3 and *GSP-1b*.

These new sequences lack the length and regular repeat motif structure of the repeating sequence domains found in traditional gliadins. However, the new (complete) sequences all encode peptide sequences near their N-termini that are reminiscent of the gliadin repeat domain (lower brackets in Fig. 1), and before the first of the conserved cysteine residues. In clone 09d3 this region contains ten glutamines out of 22 residues, and 11dc7 contains eight glutamines out of 14 residues. The 07h10 region contains fewer glutamine codons (5 out of 25), but the DNA sequence suggests that more glutamine codons may have once been present: 16 of the codons involve a single base change removed from a glutamine codon, and all are codons typically seen in gliadin repeat motifs where a single base change switches a glutamine codon (CAA, CAG) to another codon: proline (CCA, CCG), glutamate (GAA, GAG) and histidine (CAT). However, only the 09d3 sequence contains a sequence clearly characteristic of gliadins (QPQQSFPQQPQQ). The absence of any clear repeating-sequence domain may be due to a failure of the genes to evolve such a do-

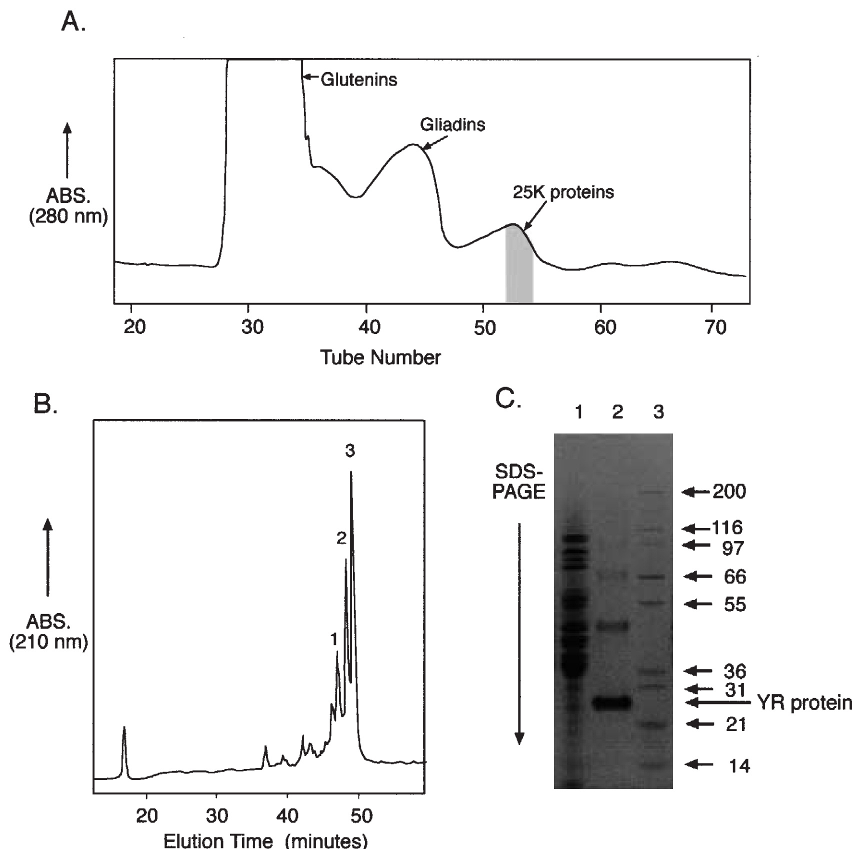

Fig. 4A–C Purification of LMW–proteins. **A** Chromatographic pattern of a 0.1 M acetic acid extract of flour proteins (see Materials and methods) from Yecora Rojo (YR) on BioGel P–100. Separation was carried out with 0.1 M acetic acid as an eluant. Fractions 52–54 contained the 25-k proteins. **B** RP–HPLC chromatographic pattern of fraction 54 from the BioGel separation of Fig. 4A. Major peaks (1,2,3) were collected for sequence analysis. **C** SDS–PAGE of flour proteins from Yecora Rojo from Fig. 4B. Proteins contained in fraction 3 from Fig. 4B are shown in *lane 2. Lane 1* shows the protein pattern of a total SDS extract (with reducing buffer) of Yecora Rojo flour. *Lane 3* illustrates the pattern of standard proteins used to calibrate the gel. Slower moving bands in Lane 2 are apparently polymers of the 25-kDa protein fraction, and are believed to be an occasionally seen artifact in some HPLC fractions of seed proteins (Kasarda, unpublished)

main, but might also be the consequence of a loss of most or all of the DNA corresponding to the repetitive domains by the crossing-over/slip-mismatching mechanisms involved in prolamin repetitive-domain evolution.

Wheat proteins that match the 07h10-derived protein

Independently of the EST sequencing results, we had been characterizing various fractions of the wheat endosperm proteins, one of which was found to match the encoded protein of DNA clone 07h10. Figure 4 summarizes the protein fractionation. The glutenin polymeric fraction eluted first from the gel filtration column at the void volume, followed by a monomeric gliadin peak, a LMW protein peak, and minor trailing fractions corresponding to relatively small proteins (Fig. 4A). The proteins collected from a portion of the LMW peak (fractions 52 and 54) were subjected to further fractionation by RP-HPLC as illustrated by the pattern in Fig. 4B for fraction 54, and N-terminal amino-acid sequencing was performed on protein from peaks 1, 2, and 3 plus two additional major peaks from fraction 52. The resulting five N-terminal sequences are compared in Fig. 5 with the N-terminal sequence derived from the DNA sequence of the 07h10 clone and the N-terminal sequence of the 25-k 'globulin' from *Triticum monococcum* described by Gomez et al. (1991). All appear to belong to a homologous group.

	1 ... 10 ... 15
Clone 07h10	R-S-A-W-E-P-Q-H-P-S-S-P-E-H-Q-
Peak 1	R-T-A-W-E-P-H-H-P-S-S-P-E-Q-Q-
Peak 2	R-I-A-S-E-P-Q-H-P-S-S-P-E-Q-
Peak 3	R-T-A-W-E-P-H-H-P-P-S-P-Q-Q-
Peak 4	R-S-A-W-E-P-Q-H-P-S-S-P-E-E-
Peak 5	R-S-A-W-E-P-Q-X-P-S-S-P-E-H-Q-
Seq. of Gomez et al.	S-T-A-W-E-P-C-W-P-S-S-X-X-Q-T-

Fig. 5 Comparison of N-terminal peptide sequences. The derived amino-acid sequence of clone 07h10 is compared to the N-terminal amino-acid sequences of individual RP-HPLC peaks (Fig. 4b) and the N-terminal '25-kda globulin' sequence from Gomez et al. (1991). Differences from the 07h10 derived sequence are enclosed in *boxes*. *X* = unidentified residue

Fraction 53 was used for SDS-PAGE analysis as shown in Fig. 4C and its mobility relative to the standards corresponds to a mass of approximately 25,000, which is the basis for referring to these polypeptides as 25-k proteins, and is the same size reported for the protein of Gomez et al. (1991). However, the MW of the derived polypeptides from clone 07h10 was 19,734. This difference may result from the failure of SDS-PAGE to yield an accurate mass as has been found for some classes of prolamins.

One of the 25-k proteins from Yecora Rojo (Peak 3) was characterized further. The amino-acid composition

Table 2 Amino-acid compositions. Derived amino-acid compositions of clones compared to wheat protein amino-acid compositions

Amino acid	Clone 07h10[a]	Protein peak-3[a]	Wheat globulin[b]	Clone 11dc7[a]	LMW gliadin[c]
Ala	5.6	5.3	5.5	5.3	6.6
Arg	2.5	2.7	3.2	3.8	3.3
Asp	3.7	5.0	3.7	0.7	1.7
Gly	4.3	5.8	8.2	5.3	6.6
Glu	26.5	27.8	28.3	28.2	28.4
His	4.9	5.3	3.2	0.7	0.5
Ile	5.6	4.2	5.5	5.3	5.4
Leu	6.2	6.6	5.5	5.3	4.2
Lys	1.8	2.5	1.8	0.0	0.2
Met	2.5	1.1	1.8	3.9	4.4
Phe	0.6	0.9	1.0	1.5	2.2
Pro	10.5	9.6	10.1	9.0	9.9
Ser	9.3	9.1	8.2	9.3	8.6
Thr	6.2	5.5	5.8	10.5	9.2
Tyr	1.8	1.7	1.8	3.0	4.2
Val	8.0	6.8	6.4	7.5	4.6
Total[d]	100.0	100.0	100.0	100.0	100.00

[a] This paper
[b] From Gomez et al. (1991); trp not determined
[c] From Prada et al. (1982); cys not determined
[d] Percent of total amino-acids determined; cys and trp not included

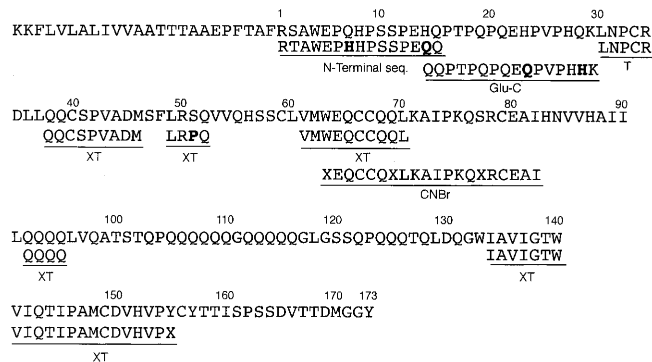

Fig. 6 Comparison of the derived amino-acid sequence of 07h10 with the peptide sequences of a 25-kDa protein from Yecora Rojo. Amino-acid sequences of peptides produced by digestion or cleavage of Peak 3 in Fig. 4A with trypsin (*T*), chymotrypsin (*XT*), endoproteinase Glu-C (*glu-C*), and cyanogen-bromide (*CNBr*). X = unidentified residues. The sequence is numbered beginning with the initial residue corresponding to the N-terminal protein sequence. Differences from the 07h10 sequence are in *bold*

of our Peak-3 component is compared in Table 2 with the 25-k globulin of Gomez et al. (1991) and the derived protein of clones 07h10. Their similarity in composition further supports the proteins encoded by clone 07h10 and the 25-k proteins from Yecora Rojo as being members of the same family described as a "low-molecular-weight globulin" fraction by Gomez et al. (1991).

Sequencing of peptides generated from Peak 3 yielded amino-acid sequences corresponding to 95 of the 173 (Fig. 6) residues in the sequence encoded by clone 07h10, and was identical except at five positions. The different N-terminal sequences from cultivar Yecora Rojo (Fig. 5), the partial amino-acid sequence from Peak 3 (Fig. 6), and the two different derived sequences from cv Cheyenne (Fig. 1) show that the original ESTs originate from a multigene family.

Relationship of clone 11dc7 to LMW gliadins and secalins

Clone 11dc7 is similar to the LMW prolamin fractions from both wheat and rye. Salcedo et al. (1979) described ten proteins with molecular weights less than 20,000 having similar alcohol-solubility and electrophoretic (pH 3.2) properties to the known gliadin groups and suggested they be designated as low-molecular-weight gliadins (LMW-gliadins). The genes for at least five of these components were assigned to the 4B, 7A and 7D chromosomes by aneuploid analysis (Salcedo et al. 1980), but not the group-1 chromosomes where most of the genes for the classical gliadins are found. Prada et al. (1982) studied the amino-acid compositions of several LMW protein fractions and reported that their compositions were similar to prolamins, although lower in glutamine and proline than the gliadins. The four components of Prada et al. (1982) had molecular weights ranging from approximately 16,000 to 19,000, and thus were significantly smaller than the main gliadin families which range in molecular weight from approximately 30,000 to 60,000. Ewart (1975) described a "gliadin" with a molecular weight of 18,000 and an amino-acid composition similar to the proteins of Salcedo et al. (1979). Similar proteins from barley have also been described (Aragoncillo et al. 1981; Salcedo et al. 1982). No sequence information has been reported, however, for any of these LMW proteins from wheat or barley. The final two columns in Table 2 compare the amino-acid composition of one of the proteins from Prada et al. (1982) with that derived from clone 11dc7. On the basis of these compositions, we suggest that the 11dc7 clone encodes a protein corresponding to one of the group of LMW-gliadins described by Salcedo et al. (1979) and Prada et al. (1982).

An extract of rye flour included two small proteins with masses of 15 and 18 kDa (rye-15 and rye-18) that showed weak immunoreactivity with antibodies in celiac serum (Rocher et al. 1996). Of the 34 positions of the

```
LDTTCSQGYGQCQQQPQQQVNTCSALLQQCSPTP        11dc7
LDTTXSQGYGQXQLQQQQMNYTXAAFLQQXXXTP      "Rye-15"
LDTTXSQGYGQXQLQQ                        "Rye-18"
```

Fig. 7 Similarity of gliadin-like ESTs to LMW-secalins. The derived N-terminal amino-acid sequence of clone 11dc7 is compared to the N-terminal sequences of two LMW-secalins from rye: rye-15 and rye-18 (Rocher et al. 1996). *X* = unidentified residues

rye-15 N-terminal amino-acid sequence, 21 of 28 match the 11dc7 derived sequence (Fig. 7). Five of the seven differences could result from a single base change in respective codons. Of the six residues not identified in the rye-15 N-terminal sequence, the 11dc7 sequence contains a cysteine codon corresponding to four of those unidentified positions (cysteines are broken down in peptide sequencing and usually must be chemically modified for identification, which apparently was not done by Rocher et al. 1996). Mass spectroscopy of the rye proteins gave molecular masses of 16,168 for rye-15 and 19,608 for rye-18, consistent with a mass of 17,896 Da calculated for the derived 11dc7 protein. These results support a close relationship between the protein encoded by clone 11dc7 and the LMW secalins from rye.

New classes of wheat endosperm proteins and genes

We report four new classes of wheat endosperm gene sequences, of which at least one class is a multigene family (represented by sequences 07h10 and 07e6). These four classes are similar enough to the gliadins to be included as members of the gliadin family of genes, and three of the classes are now represented by full-length DNA and derived amino-acid sequences for protein classes for which only the N-terminal sequence and/or amino-acid composition has previously been reported. These classes include the reported 25-k globulins (Gomez et al. 1991), a protein class related to LMW-secalins from rye (Rocher et al. 1996), and at least some of the smaller wheat endosperm proteins such as those reported for wheat by Prada et al. (1982) and for barley by Aragoncillo et al. (1981). The four new sequence types described in this paper cluster with the gliadin prolamins in a phylogenetic analysis, and are members of a gene superfamily including prolamins, CM proteins/inhibitors and puroindoline/GSPs (grain softness proteins), which Shewry (1995) has called the prolamin superfamily. It is likely that these newly described gliadin-like proteins serve as storage proteins like the other gliadin families. However, other possible roles cannot be ruled out without further evidence. We note the other, more distantly related, members of the gliadin super-family (CM proteins/inhibitors and puroindoline/GSPs) still have not had their functions clearly established. The use of transformation technology to the suppression and over-expression of such gene families may supply some clearer direction in understanding their exact biological roles.

The gliadin family now includes proteins with both a large repetitive domain and a conserved set of cysteine residues (α-gliadins, γ-gliadins, and LMW-glutenins), a class with a repetitive domain but no cysteines (ω-gliadins), and several examples from the current report with the conserved cysteine pattern but no, or relatively short, repetitive domains. Further study is needed, but the gliadin family is apparently much larger and more diverse than previously thought. It remains to be determined if the subdivisions of the gliadins are as discrete as thought, or are merely clusters of similar sequences along more of a sequence similarity continuum. We speculate that additional new classes of gliadin-related LMW proteins will be discovered and that most will belong to multigene families. Until a better understanding of the wheat endosperm proteins becomes available, we propose to refer to all of the gliadin-related LMW endosperm proteins as "LMW-gliadins".

Acknowledgments The authors thank Susan Altenbach, Ann Blechl, and Renato D'Ovidio for reading the manuscript and making suggestions and corrections, and William Vensel for guidance in carrying out the cyanogen bromide cleavage. All experiments comply with U.S. laws

References

Anderson OD, Litts JC, Greene FC (1997) The α-gliadin gene family. I. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. Theor Appl Genet 95:50–58

Anderson OD, Hsia C, Torres V (2001) The wheat γ-gliadin genes: characterization of ten new sequences and further understanding of γ-gliadin gene-family structure. Theor Appl Genet (in press)

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

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Chesnut RS, Shotwell MA, Boyer SK, Larkins BA (1989) Analysis of avenin proteins and the expression of their mRNAs in developing oat seeds. Plant Cell 1:913–924

Cole EW, Fullington JG, Kasarda DD (1981) Grain protein variability among species of *Triticum* and *Aegilops*: quantitative SDS-PAGE studies. Theor Appl Genet 60:17–30

Ewart JAD (1975) Isolation of a Capelle-Desprez gliadin. J Sci Field Agric 26:1021–1025

Garcia-Maroto F, Marana C, Mena M, Garcia-Olmedo F, Carbonero P (1990) Cloning of cDNA and chromosomal location of genes encoding the three types of subunits of the wheat tetrameric inhibitor of insect alpha-amylase. Plant Mol Biol 14:845–853

Gautier MR, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cysteine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25:43

Gomez L, Sanchez-Monge R, Salcedo G (1988) A family of endosperm globulins encoded by genes located in group-1 chromosomes of wheat and related species. Mol Gen Genet 214:541–546

Gomez L, Sanchez-Monge R, Lopez-Otin C, Salcedo G (1991) Amino-acid compositions and sequence analysis of the major low Mr globulins from *Triticum monococcum* L. endosperm. J Cereal Sci 14:117–123

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5:151–153
Kasarda DD, Woodard KM, Adalsteins AE (1998) Resolution of high molecular weight glutenin subunits by a new SDS-PAGE system incorporating a neutral pH buffer. Cereal Chem. 75:70–71
Lew EJ-L, Kuzmicky DD, Kasarda DD (1992) Characterization of low-molecular-weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino-acid sequencing. Cereal Chem 69:508–515
Nieto-Taladriz MT, Pernas M, Salcedo G, Carrillo JM (1996) Linkage mapping of '25-kDa globulin' genes on homoeologous group-1 chromosomes of bread and durum wheat. Theor Appl Genet 93:780–787
Osborne TB (1924) The vegetable proteins. Longmans, Green and Company, London
Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu Rev Plant Physiol 38:141–153
Prada J, Sanchez-Monge R, Salcedo G, Aragoncillo C (1982) Isolation of the major low-molecular-weight gliadins. Plant Sci Lett 25:281–289
Rahman S, Jolly CJ, Skerritt JH, Wallosheck A (1994) Cloning of a wheat 15-kDa grain softness protein (GSP). GSP is a mixture of puroindoline-like polypeptides. Eur J Biochem 223: 917–925
Rocher A, Calero M, Soriano F, Mendez E (1996) Identification of major rye secalins as coeliac immunoreactive proteins. Biochim Biophys Acta 1295:13–22
Rodriguez-Palenzuela P, Royo J, Gomez L, Sanchez-Monge R, Salcedo G, Molina-Cano JL, Garcia-Olmedo F, Carbonero P (1989) The gene for trypsin inhibitor Cme is regulated in trans by the *lys 3a* locus in the endosperm of barley (*Hordeum vulgare* L). Mol Gen Genet 219:474–479
Salcedo G, Prada J, Aragoncillo C (1979) Low MW gliadin-like proteins from wheat endosperm. Phytochemistry 18:725–727
Salcedo G, Prada J, Sanchez-Monge R, Aragoncillo C (1980) Aneuploid analysis of low-molecular-weight gliadins from wheat. Theor Appl Genet 56:65–69
Salcedo G, Sanchez-Monge R, Argamenteria A, Aragoncillo C (1982) Low-molecular-weight prolamins: purification of a component from barley endosperm. J Agric Food Chem 30: 1155–1157
Shewry PR (1995) Plant storage proteins. Biol Rev 70:375–426
Singh NK, Donovan GR, Carpenter HC, Skerritt JH, Langridge P (1993) Isolation and characterization of wheat triticin cDNA revealing a unique lysine-rich repetitive domain. Plant Mol Biol 22:227–237
Singh NK, Shepherd KW (1985) The structure and genetic control of a new class of disulphide-linked proteins in wheat endosperm. Theor Appl Genet 71:79–92
Tranquilli G, Lijavetzky D, Muzzi G, Dubcovsky J (1999) Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol Gen Genet 262:846–850
Wieser H, Antes S, Seilmeier W (1998) Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography. Cereal Chem 75: 644–650