Theor Appl Genet (1999) 98: 135–148

M. Ciaffi · Y. K. Lee · L. Tamas · R. Gupta
J. Skerritt · R. Appels

The low-molecular-weight glutenin subunit proteins of primitive wheats. III. The genes from D-genome species

Received: 24 November 1997 / Accepted: 18 August 1998

Abstract The isolation and characterisation by DNA sequencing of two different low molecular weight glutenin subunit (LMW-GS) genes from a genomic library derived from *Triticum tauschii* is described. These genes are similar (more than 90% similarity) but not identical to previously published LMW-GS gene sequences from cultivated wheats. A comparison of nucleotide sequence of the coding regions revealed the presence of insertions and deletions preferentially located in the region encoding the domains in the LMW-GS proteins rich in proline and glutamine and the middle part of the C-domain. The signal sequences, the amino-terminus and the remaining parts of the C-domain were conserved between all the LMW-GSs compared. The differences detected between the deduced amino-acid sequences in these three regions are only due to single nucleotide substitutions. The most important characteristic of all compared LMW-GS genes is the conservation of eight cysteine residues that could be involved in potential secondary or tertiary structure and disulphide-bond interactions. Comparisons between the 5′ and 3′ non-coding sequences of one of the isolated clones (LMW-16/10) with those of different prolamin genes from wheat, barley and rye led to the distinction of five different gene families, and confirmed the evolutionary relationships determined previously for these genes mainly on the basis of the coding region. In particular, the LMW-GS sequences are more closely related to the B-hordein sequences than to any other prolamin genes from wheat, barley and rye. Formal proof that the isolated genes coded for LMW-GSs, as defined by gel electrophoresis, was obtained by moving one of these genes (LMW-16/10) into a bacterial expression vector based on bacteriophage T7 RNA polymerase. The resulting plasmid directed the synthesis of large amounts of the mature form of the subunit in *Escherichia coli*. This protein exhibited solubility characteristics identical to those of the LMW-GSs and cross-reacted with antibodies reactive with these proteins.

Key words Seed storage proteins · LMW-GS genes · Molecular evolution · *Triticum tauschii* · Wheat quality · Bacterial expression

Communicated by B. S. Gill

M. Ciaffi[1] · Y. K. Lee · R. Gupta · J. Skerritt · R. Appels (✉)
CSIRO Plant Industry, PO Box 1600, Canberra ACT 2601, Australia

L. Tamas
Plant Science CRC, PO Box 1600, Canberra ACT 2601, Australia

Present address:
[1] DABAC, University of Tuscia, 01100 Viterbo, Italy

Introduction

Endosperm proteins of wheat consist predominantly of two classes of storage proteins termed gliadins and glutenins, so classified on the basis of their solubility in different solvents. The gliadins are readily soluble in aqueous alcohols and are monomeric proteins with intramolecular disulphide bonds or no cysteine residues. The glutenins, in contrast, are present in high Mr ($1–10 \times 10^6$) polymers stabilised by inter-chain disulphide bonds, and are only soluble in aqueous alcohols as reduced subunits (Kasarda 1989). In the absence of any known enzymatic and structural function, these proteins are presumed to act as a store of nitrogen, carbon and sulphur to be used by the seedling during germination. The proteins are therefore called storage proteins. Biochemical and genetical aspects of wheat storage proteins have received a good deal of attention in recent years due to their importance in determining the nutritional and technological properties of cultivated wheats. It is widely accepted that the glutenin fraction

of the gluten proteins is most important in producing the unique viscoelastic properties of wheat flour doughs and in determining variations in the bread- and pasta-making qualities of different cultivars (Kasarda 1989; MacRitchie 1992).

Glutenin protein subunits have been classified on the basis of their mobilities in SDS-PAGE under reducing conditions into two main groups: high and low glutenin subunits (HMW-GSs and LMW-GSs) . The molecular analysis of genes encoding these proteins is of considerable interest for studying the structure and properties of glutenin subunits and the distribution of disulphide bonds, two essential features to explain the physical properties of gluten in molecular terms (Kasarda 1989; Shewry et al. 1994). Complete amino-acid sequences are available for representatives of the main types of HMW-GSs present in cultivated wheats as a result of the sequencing of the genes coding for these proteins (see Shewry et al. 1989, 1992 for reviews), whereas the structure of genes encoding the LMW-GSs is based on only a limited number of genes (Okita et al. 1985; Colot et al. 1989; Cassidy and Dvorak 1991; D'Ovidio et al. 1992).

The LMW-GSs account for about 40% of the total gluten fraction and for 60–70% of the glutenins. They consist of a number of individual components which are classified into three groups (B, C and D) on the basis of their behaviour on SDS-PAGE and 2-D IEF/SDS-PAGE (Jackson et al. 1983). The major group, the B-subunits, are basic proteins, whereas the minor C-subunits have a wide range of isoelectric points, from slightly acidic to highly basic. The B group correspond to proteins ranging in molecular weight from about 40 000 to 50 000 while the C-subunits range from about 30 000 to 40 000. The remaining group, the D-subunits, have slower mobilities than the B- and C-subunits and form one of the most-acidic groups of proteins in the endosperm. Although the precise relationships of these three groups are still not established, preliminary studies on N-terminal sequences indicate that the C group consists mainly of proteins related to alpha- and gamma-type gliadins (Tao and Kasarda 1989; Lew et al. 1992), whereas the D group may be most closely related to omega-gliadins (Masci et al. 1993). In contrast, the B group appears to be structurally distinct, although related more closely to the alpha-type and gamma-type gliadins than to the HMW-GS and omega-gliadins (Lew et al. 1992).

The present study describes the isolation and characterisation by DNA sequencing of two different LMW-GS genes from a genomic library derived from the wild diploid species *Triticum tauschii*, the D-genome donor to the hexaploid cultivated wheat *Triticum aestivum*. The cloning of one of these genes in an *Escherichia coli* expression vector and the purification and characterisation of the corresponding protein is also reported.

Materials and methods

Genomic library construction and screening

A genomic library of diploid wild wheat *T. tauschii* was kindly provided by Dr. S. Rahman (CSIRO, Division of Plant Industry, Canberra) and its construction has been described in Rahman et al. (1997).

The phage library was screened by plaque hybridisation using as a probe the LMW-GS cDNA clone pTdUC1 (Cassidy and Dvorak 1991) and employing standard procedures (Sambrook et al. 1989).

Characterisation of the lambda DNA clones carrying LMW-GS genes

The isolated and purified lambda clones were digested with four restriction enzymes that cut inside the multiple cloning site of the lambda-GEM-12 vector: *Sac*I, *Bam*HI, *Eco*RI and *Not*I. After separation on 1% agarose gels the restriction fragments containing the LMW-GS genes from *T. tauschii* were located by Southern blotting using the same cDNA probe as employed for screening.

The oligonucleotides used as primers to amplify different regions of the LMW-GS genes from *T. tauschii* have been synthesised on the basis of a published sequence of a LMW-GS gene from *T. aestivum* cv Chinese Spring (Colot et al. 1989) and have the following sequences:

(a) 5′ ATGAAGACCTTCCTCGTCTT 3′,
(b) 5′ GTAGGCACCAACTCCGGTGC 3′,
(c) 5′ CGACAAGTGCAATTGCGCAGATGGA 3′,
(d) 5′ ACCTAGCAAGACGTTGTGGCATTGC 3′,
(e) 5′ AGATGCATCCCTGGTTTGGAG 3′,
(f) 5′ GAGGAATACCTTGCATGGGTT 3′,
(g) 5′ AACCCATGCAAGGTATTCCTC 3′,
(h) 5′ AATGGAAGTCATCACCTCAAG 3′.

The first set of sequence primers correspond to the 46–65-bp (a) and 967–948-bp (b) positions of the LMWG-1D1 genomic clone (Colot et al. 1989) and include all of the coding region with the signal peptide. The second set of primers used correspond to the 86–110-bp (c) and 554–530-bp (d) positions of the same clone, so that their PCR products should contain the whole N-terminus and repeat domain and a part of the C-terminal region. The third set of oligonucleotides correspond to the 115–135-bp (e) and 506–486-bp (f) positions and include part of the N- and C-terminal regions and the whole repeat domain, whereas the amplified products of the last set of primers (g and h) should contain the most variable region of the C-terminus. PCR reactions were carried out in a reaction volume of 10 µl with 1 unit of *Taq* DNA polymerase (Perkin Elmer Cetus, Norwall, Conn.), 1 × *Taq* PCR buffer (Perkin Elmer Cetus, Norwall, Conn.), 2 mM $MgCl_2$, 200 µM of each dNTP, 1 µM of each primer and 30 ng of lambda-clone DNA. The plasmid DNAs pARC7 and pUC19, containing the cDNA insert pTdUCD1 (Cassidy and Dvorak 1991), and the genomic clone LMWG-1D1, respectively, were used in these reactions as controls. DNA was subjected to an initial denaturation step at 94°C for 2 min, the amplification conditions were for 35 cycles at 92°C for 30 s, 60°C for 30 s and 70°C for 1 min, followed by a final incubation step at 70°C for 5 min before cooling to 25°C. Amplified products were analysed on 2% agarose gels. Bacteriophage SSP-1 DNA, digested with *Eco*RI, was used as a molecular-weight standard to estimate the size of the amplified products.

DNA analyses

After restriction-enzyme analyses and PCR characterisation of the isolated lambda clones, two of these were chosen for further

characterisation by DNA sequencing. A 3.6-kb *Sac*I-digested fragment from the lambda clone 16/10 that hybridised with the LMW-GS cDNA (pTdUC1) probe was subcloned into the *Sac*I site of the plasmid pGEM-7Zf and completely sequenced. The PCR-amplified product corresponding to the complete coding region of clone 14/1 was cloned into pGEM-T. Uni-directional deletions were generated using the Erase-a-base kit (Promega), following the manufacturers protocol. The dideoxy method (Sanger et al. 1977) was then used in a 373 Å automatic DNA sequencer (Applied Biosystems); sequencing reactions were performed with a *Taq* DyeDeoxy™ Terminator Cycle sequencing kit and thermal cycling according to Applied Biosystems. Both strands were entirely sequenced.

Nucleotide sequences were assembled and analysed using the Genetics Computer Group (GCG) package in ANGIS (Australian National Genomic Information Service). Comparisons of nucleotide sequences between *T. tauschii* clones and different prolamin genes were carried out using the computer program GAP (gap weight = 5.0, gap length = 0.1). Coefficients of identity or homology were defined as the percentage of identical nucleotides over the number of aligned positions, and gaps were ignored in that calculation. Optimal pairwise alignments of the delineated sequences were obtained using the computer programs GAP, FASTA, PILEUP combined with manual alignment.

Expression of a LMW-GS gene from *T. tauschii* in *E. coli*

The coding region, without the signal peptide, of the LMW-GS gene contained in the 3.6-kb genomic fragment from *T. tauschii* (LMW-16/10) was amplified by PCR using oligonucleotides with the sequences:

*Nde*I

oligonucleotide-1. 5′-GGAATTCCATA**TGGAGACTAGAT-**

GCATCCTGGTTT;

*Bam*HI

oligonucleotide-2. 5′-CGCGGATCC**TTATCAGTAGGCAC-**

CAACTCCGGTGC.

Oligonucleotide 1 corresponds to the sequence encoding the N-terminus of the mature proteins (first 25 nucleotides in bold) downstream from a *Nde*I restriction site (double underlined), including the ATG initiation codon, and seven extra nucleotides (underlined). Oligonucleotide 2 is complementary to a region that corresponds to the end of the coding region (in bold) including the two final stop codons. It also contains a *Bam*HI site (double underlined) and three additional nucleotides (underlined). The PCR fragment obtained by using these two set of primers was digested with *Nde*I and *Bam*HI and ligated into the pET11a expression vector (Novagen). The target gene was initially cloned using a host that did not contain the T7 RNA polymerase gene (*E. coli* strain, DH5-alpha) thus eliminating plasmid instability due to the production of proteins potentially toxic to the host cell; the construction pET-LMW-16/10 was confirmed by sequencing. The protein was expressed in *E. coli* JM109 (DE3) (Promega) after induction by 1 mM of IPTG (Studier et al. 1990).

Protein extraction and purification

Glutenins from *T. tauschii* accession CPI 10799 and *T. aestivum* cv Chinese Spring were prepared as described by Gupta and MacRitchie (1991) with the exception that alkylation was omitted.

For total bacterial-protein extraction, pelleted cells were solubilized in a sample buffer solution Tris HCl 0.125 M pH 6.8 containing 4% sodium dodecyl sulphate (SDS), 10% (v/v) glycerol and 0.3% dithiothreitol (DTT), or else in a sample buffer lacking DTT, for 1 h at 60°C.

For purification of the expressed proteins the methods of Tamas et al. (1994) and Lullien-Pellerin et al. (1994) were compared. For solubilisation of the inclusion bodies the pellet, after the washing step, was re-suspended in 70% aqueous ethanol containing 0.3 M DTT at 60°C and the expressed protein was precipitated by the addition of 2 vol of aqueous 1.5 M NaCl followed by standing for 24 h at 4°C. The pellet was dissolved in 4 M urea, precipitated by dialysis against distilled water, re-dissolved in 0.1 M acetic acid and freeze dried.

Gel electrophoresis

One-dimensional SDS-PAGE fractionation was carried out on 10% polyacrylamide gels as described by Gupta and MacRitchie (1991). Two-dimensional IEF/SDS-PAGE was performed according to the procedure of Kasarda et al. (1988), except for the substitution of Triton X100 in place of NP-40, and the use of 5% ampholytes (Pharmacia-LKB) with the pH ranges 5–7 (1.5%), 7–9 (2%), and 8–10.5 (1.5%).

For isoelectric-focusing (IEF) analyses, glutenins, after precipitation in 80% acetone at −20°C, and the purified expressed protein, were re-suspended in a solution containing 6 M urea, 1% (v/v) Triton X100, 0.3 M DTT and 10% (v/v) glycerol.

Detection of proteins

Gels were stained overnight with 0.025% Coomassie Blue R-250, following standard procedures. Immunological detection of proteins in immuno-blots was performed according to the procedure of Andrews and Skerritt (1996). SDS gels were electroblotted for 6 h at 250 mA in 25 mM Tris-HCl pH 8.3, 190 mM glycine, 10% (v/v) methanol, 0.02% SDS, using a semi-dry electroblotting apparatus. Nitrocellulose blots were pre-incubated overnight in a PBS (phosphate-buffered saline) solution containing 3% bovine serum albumin (BSA). Blots were then agitated for 2 h at room temperature in the solution described above but containing a monoclonal anti-prolamin antibody (80 818, IgM, subclone 20 F5, diluted 1/5000) reactive with the LMW-GS extracted from hexaploid wheat (J. Skerrit, CSIRO, Division of Plant Industry, Canberra, unpublished). After four washes in TBST buffer [10 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.05% (v/v) Tween 20], blots were incubated for 1 h in TBST buffer containing 2% BSA, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$ and anti-mouse antibody conjugated to alkaline-phosphatase (diluted 1/5000). For alkaline-phosphatase staining, blots were washed four times in TBST and then incubated in AP buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 55 mM $MgCl_2$) containing 0.03% nitroblue tetrazolium (NBT) and 0.0165% 5-bromo-4chloro-3indolyl phosphate (BCIP).

N-terminal amino-acid sequencing

For N-terminal amino-acid sequencing, cysteine residues of the expressed and purified protein were alkylated with 4-vinylpyridine prior to SDS-PAGE analysis as described by Gupta and MacRitchie (1991). The protein was transferred from the gel to a ProBlott™ polyvinylidene difluoride (PVDF) membrane by passive adsorption (D. Shaw, Australian National University, Canberra, unpublished). The N-terminal sequence was determined on a Perkin Elmer Applied Biosystems Procise™ 494 Protein Sequencer using a standard blot cycle, with identification of the PTH amino-acid derivatives by RP-HPLC (facility at the Australian National Uiversity, Canberra).

Results

Isolation and characterisation of LMW-GS clones

To define the LMW-GS gene family of *T. tauschii* prolamins, we isolated 14 lambda clones that hybridised under high-stringency conditions (0.1 × SSC, 0.1% SDS at 65°C) to the pTdUC1 clone but not to the gamma-gliadin cDNA pKa1a (Bartels et al. 1986). The cleavage sites for four restriction endonucleases (*Sac*I, *Bam*HI, *Eco*RI and *Not*I) were mapped on the inserts of the isolated clones. For all the restriction enzymes used, seven different patterns were observed with the most common being present in five clones (data not shown). The sizes of cloned inserts from *T. tauschii* DNA range from about 12 to 15 kb and each clone, regardless of the restriction endonuclease employed, contained only one subfragment that hybridised with the LMW-GS cDNA (pTdUC1) probe. These results suggested that only one LMW-GS gene resides within the 12–15-kb-long sequences of the isolated 14 clones.

The amplification of LMW-GS genes from the 14 isolated lambda clones allowed at least two different genes, based on the PCR products obtained with primers amplifying the entire gene (primers a and b, see Materials and methods), to be distinguished. Using PCR primers that flank the repetitive, proline/glutamine-rich region of the gene (primers c and d, e and f), two additional variant genes were identified; the total range of variants is shown in Fig. 1. The PCR fragments obtained from the first four clones (lanes 1–4, Fig. 1), including 16/10 (lane 1, Fig. 1), showed a uniform size for all the different sets of primers used indicating that most likely they contain the same gene.

Clone 1/1 (lane 5, Fig. 1) was unusual in having a larger C-domain (assayed using primers g–h), as well as a shorter repetitive domain (primers c–d and e–f). The amplified products corresponding to the complete coding region of the 5/1 and 14/1 clones (lanes 6 and 7, Fig. 1; primers a and b) were slightly larger than the PCR fragments of the other clones analysed. For clone 5/1 it is clear that the difference is mainly due to variation in the length of the repetitive, proline/

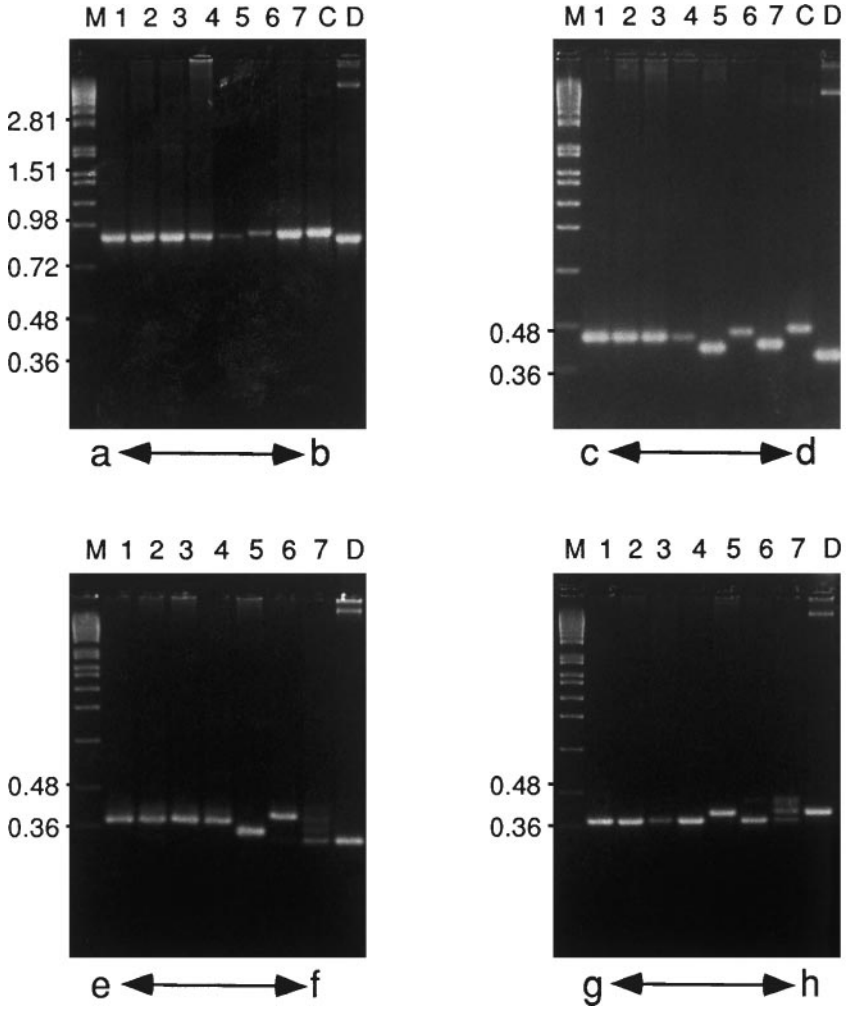

Fig. 1 2.0% agarose gels of PCR products from seven different lambda clones (*1–7*) and plasmid DNAs pARC7 and pUC19 containing the cDNA insert pTdUCD1 (*D*) and the genomic clone LMWG-1D1 (*C*), using four different sets of primers (*a-b*; *c-d*; *e-f*; *g-h*) that amplify different regions of the LMW-GS genes. The seven clones represented the different classes of lamdba clones identified on the basis of the restriction endonuclease digestions. Bacteriophage SSP-1 DNA, digested with *Eco*RI, (*M*), was used as a molecular-weight standard to estimate the size of the amplified products

glutamine-rich-region domain (lane 6, primers c and d), whereas the amplified product of clone 14/1 from the same region was shorter (lane 7, Fig. 1; primers c and d).

Nucleotide sequences of the LMW-16/10 and LMW-14/1 clones and their comparison with other prolamin genes

On the basis of the restriction-enzyme analysis and PCR characterisation of the above lambda clones, 14/1 and 16/10 were chosen for further characterisation by DNA sequencing.

A single large open reading frame (ORF) was found within the 3609-bp sub-fragment of clone 16/10 (LMW-16/10) with a coding capacity of 299 amino acids and an estimated molecular weight for the protein of 32 024. The 897-bp coding region was bounded by a 1191-bp upstream and 1521-bp downstream flanking sequence. The nucleotide sequence of the amplification product from the lambda clone 14/1 (LMW-14/1) was 909-bp long with a predicted molecular weight for the corresponding protein of 32 453.

The nucleotide sequences of the coding regions in LMW-16/10 and LMW-14/1 were compared to the nucleotide sequences of the entire coding regions of the LMW-GS genes from *T. aestivum* and *T. durum* reported to-date (Table 1). The two *T. tauschii* clones shared strong similarity with four of the five published LMW-GS genes ranging from 92.0% to 99.8%, whereas the genomic clone LP1211 had only 79.1% and 76.3% similarity to LMW-16/10 and LMW-14/1, respectively. LMW-16/10 is very similar to the clone LMWG-1D1, a Chinese Spring allele located on chromosome 1D (Colot et al. 1989), differing only by two single-nucleotide substitutions and two deletions located in the repeat domain and encoding for a dipeptide and a hexapeptide. A nucleotide comparison of LMW-14/1 with a cDNA clone isolated from the *T. aestivum* cv Cheyenne by Okita et al. (1985) showed the presence of only one insertion and one deletion together with five nucleotide changes out of 909 nucleotides compared. It is also interesting to note that the genes in LMW-14/1 and LMW-16/10 are closely related (92.0–93.9% similarity) to the two genes isolated from *T. durum* (Cassidy and Dvorak 1991; D'Ovidio et al. 1992); this species does not contain a D genome.

The nucleotide sequences of the coding regions in LMW-14/1 and LMW-16/10 were also compared to the nucleotide sequence of the coding region of other wheat and barley prolamin genes (data not shown). The LMW-GS gene sequences are more closely related to the B-hordein sequences than to any other prolamin genes from wheat, with a similarity ranging from 70 to 75% (see also Kreis et al. 1985; Colot et al. 1989). Over the total length of the coding regions, the alpha- and beta-gliadins, and the gamma-gliadins showed similarities of, on the average, 60% and 65%, respectively; the LMW-14/1 and LMW-16/10 gene sequences had on average only 48% similarity to the related C-hordein and omega-secalin sequences. Only very minor similarities were detected between the LMW-14/1 and LMW-16/10 sequences and any HMW-GS gene.

The amino-acid sequences were deduced from the nucleotide sequences of the following LMW-GS clones, LMWG-1D1, LMW-16/10, LMW-14/1, WHTGLIGBA, pTdUCD1 and pLMW21, and have been aligned in Fig. 2 for maximal similarity at the amino-acid level. The signal sequences and the amino-terminus regions appear to be conserved between all the LMW-GSs compared. The differences detected between the deduced amino-acid sequences in these two regions are only due to single nucleotide substitutions. Nucleotide comparisons of the six genes showed the presence of different insertions and deletions preferentially located in the repeat domain rich in proline and glutamine. The C-domain (region I3) was also variable, consistent with these regions being considered to be the most divergent regions and the principal sources of polypeptide-length polymorphism within and between different prolamin genes (Kreis et al. 1985). The nucleotide comparison of the LMW-16/10 and LMW-14/1 sequences revealed the presence of two insertions and one deletion as well as eight nucleotide changes in the repeat domain. The deletions are represented by two triplets coding for glycine, whereas the two insertions have different extensions and encode a heptapeptide and a tripeptide. The alignment of the deduced

Table 1 Comparison of LMW glutenin nucleotide sequences from *T. aestivum* and *T. durum* to *T. tauschii* clones LMW-16/10 and LMW-14/1

Designation	Accession	Origin	Type	Reference	Homology to	
					LMW-16/10	LMW-14/1
LMWG-1D1	X13306	*T. aestivum* cv Ch. Spring	Genomic	Colot et al. 1989	99.8	92.6
LP1211	X07747	*T. aestivum* cv Yamhil	Genomic	Pitts et al. 1988	79.1	76.3
WHTGLIGBA	M11077	*T. aestivum* cv Cheyenne	cDNA	Okita et al. 1985	92.4	99.4
pTdUCD1	X51759	*T. durum* cv Mexicali	cDNA	Cassidy and Dvorak 1991	92.0	93.4
pLMW21	X62588	*T. durum* cv Lira	PCR	D'Ovidio et al. 1992	92.3	93.9

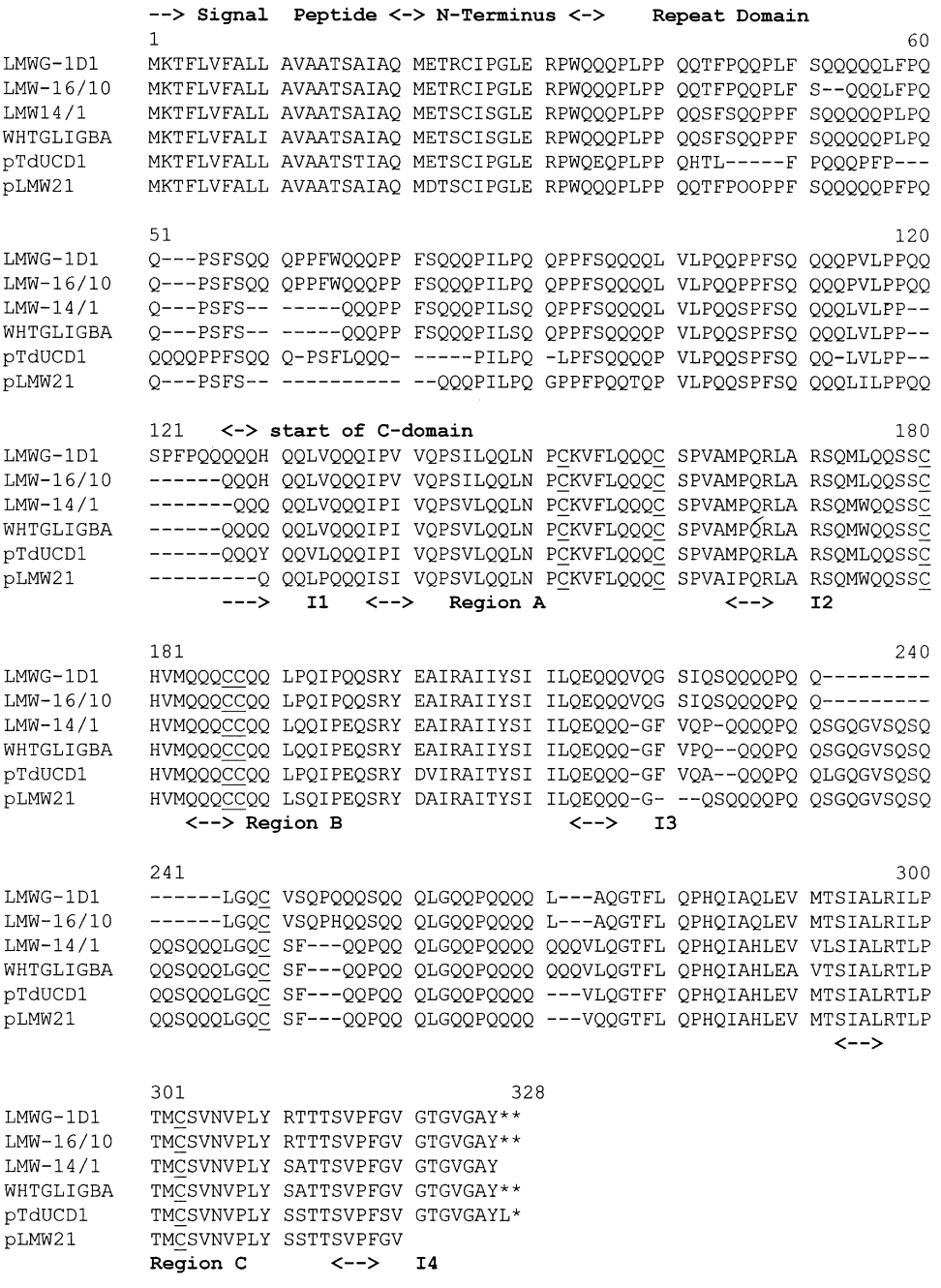

Fig. 2 Comparison of the deduced amino-acid sequences of the two *T. tauschii* clones with the previously published LMW-GS genes from *T. aestivum* and *T. durum*. The sequences of LMWG-1D1, LMW-16/10, LMW-14/1, WHTGLIGBA, pTdUCD1 and pLMW21 were aligned for maximal homology. Periods (*dashes*) indicate gaps inserted for maximal homology. The deduced amino- acid sequences have been divided into a signal peptide, an N-Terminus, a repeat domain rich in proline and glutamine, and C-domain. According to Kreis et al. (1985) seven regions are labelled below the sequences of the C-domain and delineated with *arrows*. The cysteine residues have been *underlined* and *stars* indicate stop codons

amino-acid sequences of the repeat domains in the genes from LMW-16/10 and LMWG-1D1 showed the presence of two deletions represented by a dipeptide and a hexapeptide, whereas in the same regions the clones LMW-14/1 and WHTGLIGBA differed only by a single amino acid change and insertion. One of the most relevant differences between the two *T. tauschii* clones is the insertion of 15 aminoacids in the variable region of the C-domain (region I3) of LMW-14/1.

The most important characteristic of all compared genes is the conservation of eight cysteine residues, which could be involved in potential secondary or tertiary structure and disulphide-bond interactions. Six of the cysteine residues are clustered in the middle of the protein, one cysteine is found very close to the N-terminus, at amino-acid position 5 in the mature protein sequence, and the other is 24–25 amino acids from the carboxyl terminus.

5′ Flanking region

The nucleotide sequence of the promoter region in LMW-16/10 was compared to those of the equivalent region of 16 different prolamin genes from wheat, barley and rye (Table 2). The LMW-GSs from wheat and

Table 2 Comparison of the promoter region of different prolamin genes to the *T. tauschii* clone LMW-16/10

Protein type	Designation	Accession	Length (bp) of 5′ sequence compared	Reference	Homology to LMW-16/10
Wheat LMW glutenin	LMWG-1D1	X13306	983	Colot et al. 1989	99.5
	LP1211	X07747	550	Pitts et al. 1988	93.5
Barley B hordein	HVB1HOR2	X87232	550	Brandt et al. 1985	83.3
	pBHR184	X03103	563	Forde et al. 1985	85.4
	pcr31	X53690	299	Vincente-Carbajosa et al. 1992	86.9
Wheat gamma-gliadin	L311	M13713	503	Rafalsky 1986	62.7
	pW1020	M36999	765	Scheets and Hedgcoth 1988	59.7
Barley gamma-hordein	lhorg-1	M36378	378	Cameron-Mills and Brandt 1988	59.7
Wheat alpha-beta gliadin	YAM-2	X01130	771	Anderson et al. 1984	52.6
	pW1215	K03074	593	Sumner-Smith et al. 1985	52.9
Barley C- hordein	lhor1-14	M36941	428	Entwistle 1988	52.9
	M564	S66938	783	Sainova et al. 1993	49.8
Rye omega-secalin	pSec2B	X60295	477	Hull et al. 1991	53.2
Wheat HMW-glutenin	Glu-1D-1b	X12928	480	Anderson et al. 1989	< 35.0
	Glu-1By9	X61026	750	Halford et al. 1987	< 35.0
Barley D-hordein	pHor3-1	X84368	434	Sorensen et al. 1996	< 35.0

the B-hordein are closely related groups that showed 99.5–83.3% nucleotide identity with the promoter region of the *T. tauschii* clone. The LMW-16/10 promoter region is very similar to the equivalent region in clone LMWG-1D1; comparisons of the promoter regions of these two clones revealed only three deletions of single nucleotides and two nucleotide changes out of the 983 nucleotides. The gamma-gliadin and gamma-hordein gene promoter regions (with more than 80% similarity between them) are slightly more closely related to the LMW-16/10 gene (59.7–62.7% similarity) than to the other prolamin genes. The promoter regions of the C-hordein and omega-secalin genes (more than 80% of identity), and the alpha- and beta-gliadins, show only 49–53% similarity to the equivalent region in LMW-16/10. Finally, only very minor similarities were detected between the promoter region of LMW-16/10 and equivalent regions of any of the HMW prolamin genes (wheat HMW-GSs and barley D-hordein).

On the basis of these comparisons, regardless of the species considered, we distinguished five families of sequences, within which each pairwise comparison gave a score of 80%: (1) LMW glutenin and B-hordein; (2) the gamma-families including gamma-gliadin and gamma-hordein; (3) alpha and beta-gliadins; (4) omega-secalins and the C-hordein; and (5) the HMW prolamins with the wheat HMW-GSs and the barley D-hordein.

Some 370 bp of 5′ flanking sequences from eight different prolamin genes representative of the first four families are compared by alignment similarity in Fig. 3. Despite considerable divergence between the 5′ flanking sequence of the four families there are, within the 370–350 bp compared, several short segments of the 5′ flanking non-coding region that show more than 80% homology. One of the most strongly conserved sequences among the four families is a 25-bp element located at a position about 300 bp upstream of the translation start site. This consensus sequence has been termed the endosperm box or prolamin element and is considered to confer endosperm-specific expression of the prolamin genes not only in wheat, barley and rye but also in maize (Forde et al. 1985; Kreis et al. 1985). In all prolamin genes of barley, wheat and rye, except for the HMW prolamins, this sequence consists of two smaller eight- and ten-bp motifs with a high degree of similarity to known cis-acting sequences. The 5′ conserved motif TGTAAAGT is termed the endosperm motif and resembles the viral SV40 enhancer core sequence TGTGGAAAGT (Mitchell et al. 1987). The 3′ conserved motif is very similar to the GCN4-binding sequence in yeast (Hill et al. 1986) and the binding site for the mammalian transcription factors jun (Ransone and Verma 1990) and AP1 (Piette et al. 1988) and is called the GCN4 motif. It is interesting to note that in some of the sequences compared in Fig. 3 (LMW-16/10, LMWG-1D1, pBHR184, pW1020, Yam-2) the 25 bp representing the −300 element is imperfectly repeated about 200–280 bp upstream and many other sequences similar to the sequence motifs of SV40 are scattered in both orientations along the promoter regions further upstream of the endosperm box (data not shown). For example, a close inspection of the sequences around the endosperm box in the promoter of LMW-16/10 shows that not only the endosperm motif (from −341 to −333) but also the GCN4 motif (from −315 to −306) is imperfectly duplicated.

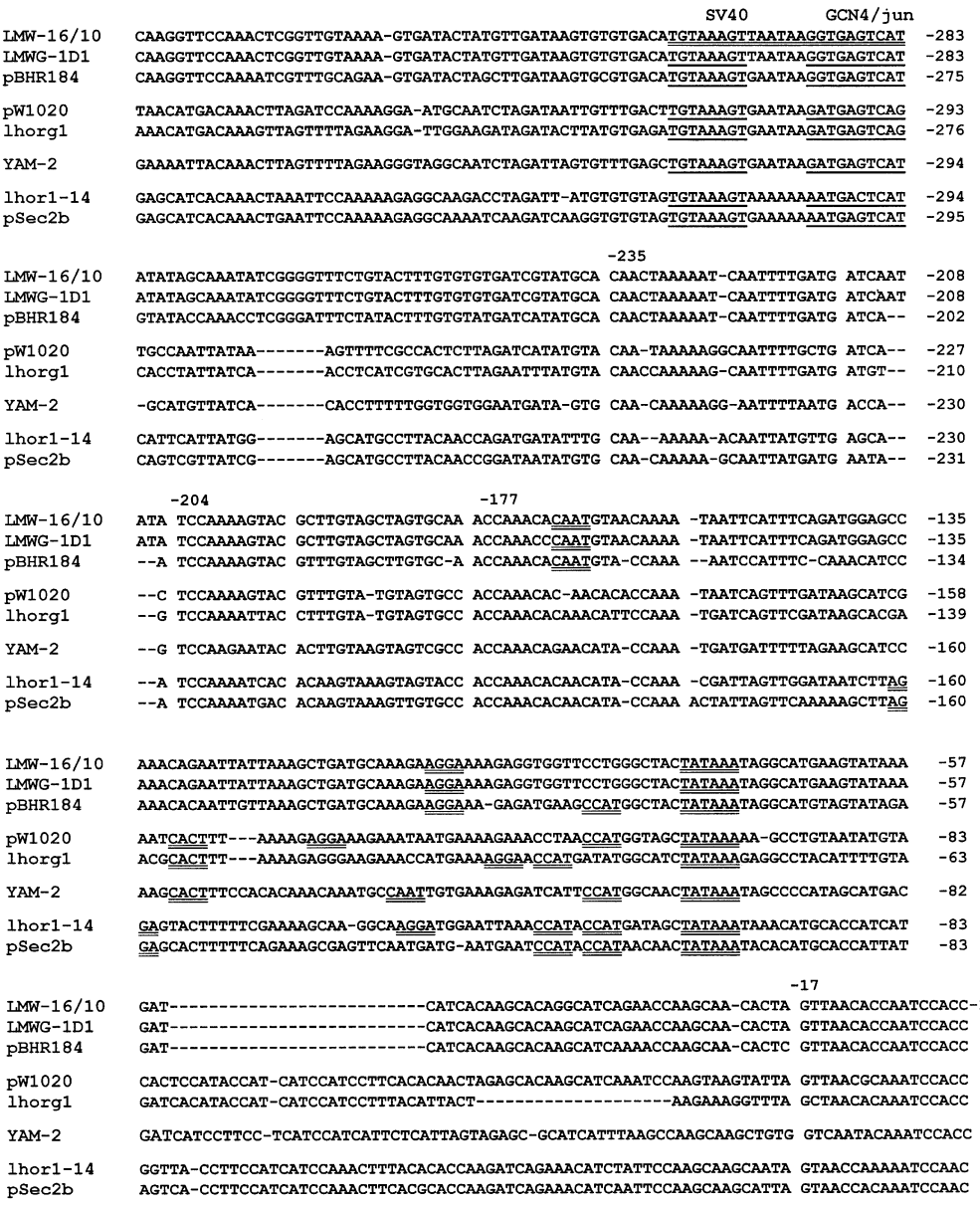

Fig. 3 Comparison of the 5′ flanking region of different prolamin genes from wheat, barley and rye (see Table 2 for details of prolamin genes) with the *T. tauschii* clone LMW-16/10. Periods (*dashes*) indicate gaps inserted for maximal homology. The TATA, AGGA boxes, and a sequence that may represent CAAT boxes are *double underlined* in addition to the endosperm box. The SV40 enhancer core-like, and the GCN4/jun, motifs are also indicated. *Numbering* starts at the first nucleotide 5′ upstream of the ATG start codon and the numbering is with respect to the LMW-16/10 sequence

3′ Flanking region

In common with the 3′ untranslated regions of most plant genes (Dean et al. 1986) those of the prolamins analysed have at least two copies of a sequence that conforms to the consensus for a polyadenylation signal sequence (AATAAA) (Fig. 4). The first motif is located 60–80 nucleotides 3′ of the termination codon, whereas the other is about 50 nucleotides further downstream, with the only exception being the C-hordein clone lhor1–14 where the second polyadenylation signal is not present. Only in the family of LMW-GSs and the B-hordein sequences is a third motif four very close to the second polyadenylation signal (four nucleotides further downstream). The comparison of the 3′ flanking regions from different prolamin genes (Fig. 4) revealed a strong conservation of the sequence around the two first AATAAA motifs but not elsewhere, and thus suggested that these two motifs are selected-for in evolutionary terms because they encode an important function.

Comparisons of the non-coding 3' flanking sequence of the LMW-16/10 clone with equivalent regions from the other prolamin genes (using the same source of data listed in Table 2), covering 200 bp, also led to the identification of the five families described in the previous section. Within each pairwise comparison a score over 70% similarity was obtained (data not shown). The LMW-GS and the B-hordein genes are a closely related group that show 99.9–77.4% nucleotide identity in their 3′ flanking regions. The 3′-untranslated regions of alpha-gliadin and gamma-gliadin, and of gamma-hordein show only 65.5–59.9% similarity to

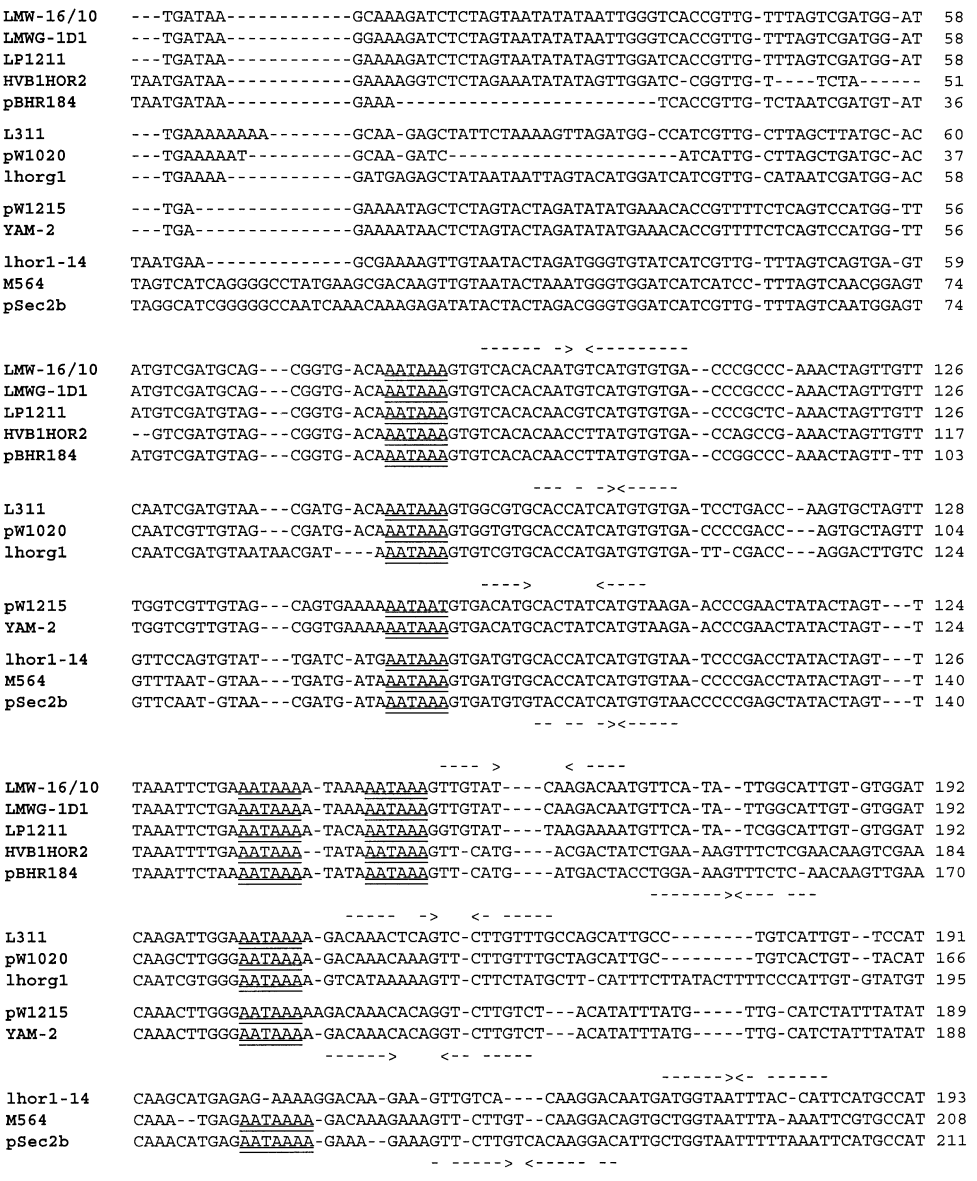

Fig. 4 Comparison of the 3′ flanking region of different prolamin genes from wheat, barley and rye (see Table 2 for details of different prolamin genes) with the *T. tauschii* clone LMW-16/10. *Dashes* correspond to gaps introduced to optimise the alignment. The polyadenylation signals AATAAA are *double underlined* and the sequences involved in potential secondary structures are indicated with *arrows*

the members of the first group; they are equally distant from each other. The LMW-GS and B-hordein sequences are as closely related to the 3′ flanking regions of C-hordein and omega-secalin (63.5–61.4% similarity) as they are to any of the other sulphur-rich sequences. The 3' flanking regions of C-hordein and omega-secalin are, in turn, almost equidistant in sequence similarity from all the sulphur-rich sequences. Finally, the HMW prolamin sequences show little similarity ($< 40\%$) to the 3′-untranslated sequences of any of the other prolamin genes.

Expression of the LMW-16/10 gene in *E. coli*

Formal proof that the LMW-GS genes in the literature encode proteins that correspond to the proteins identified as LMW-GS from polyacrylamide-gel analyses of seed storage proteins has been limited to N-terminal amino-acid analyses. To extend this proof of identity further, the LMW-16/10 gene was inserted into the expression vector pET11a, after deleting the signal peptide. Proteins extracted from bacteria containing either the control expression vector (pET11a) or the construct (pET-LMW-16/10) were analysed by SDS-PAGE under reducing conditions (Fig. 5). The bacteria with the pET-LMW-16/10 construct accumulated a protein that co-migrated with the C group of the LMW-GS extracted from wheat (lane C, Fig. 5) and *T. tauschii* (lane T, Fig. 5) after 2 h and 6 h of induction (lanes 5 and 6, Fig. 5). This protein was absent in un-induced cells (lane 4, Fig. 5) and in control bacteria (lanes 1–3, Fig. 5). Densitometer tracings of the protein gels indicated that, after 6 h of accumulation, the expressed subunit represented about 18% of the total *E. coli* proteins (data not shown).

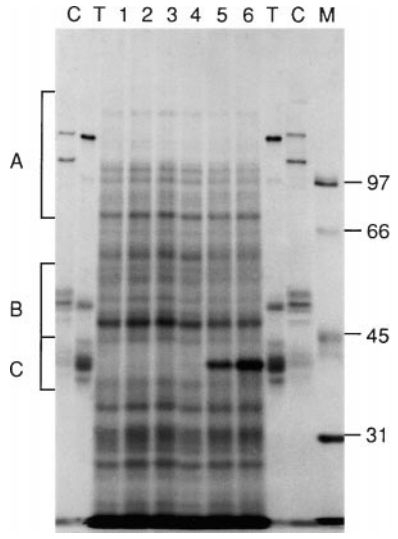

Fig. 5 SDS-PAGE of total reduced cell proteins from *E. coli* containing the control plasmid pET11a (*1–3*) or the recombinant expression vector pET-LMW-16/10 (*4–6*). 1 and 4 before induction; *2* and *5* after 2 h induction; *3* and *6* after 6 h induction. The glutenin subunits have been divided into three groups in the bread wheat cultivar Chinese Spring (*C*) and in *T. tauschii* (*T*): *A* = HMW-GS; *B* and *C* = LMW-GS. The molecular weights ($\times 10^{-3}$) of the protein markers (*M*) are indicated on the right side

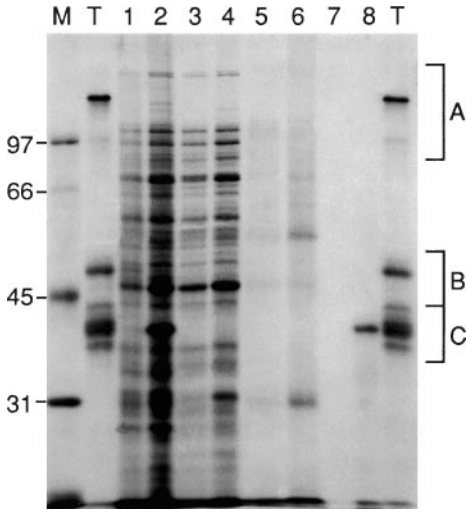

Fig. 6 Total reduced cell proteins from *E. coli* containing the recombinant expression vector pET-LMW-16/10, before and after 6 h induction, are reported in *lanes 1* and *2*, respectively. *Lanes 3* and *4* supernatants from centrifugation of *1* and *2*, respectively. *Lanes* 5 and *6* supernatants after washing with 0.5% Triton X100 and 10 mM EDTA of pellets from the centrifugation of *1* and *2*, respectively. *Lanes 7* and *8* pellets re-suspended in 70% ethanol containing DTT at 60°C after the washing step. The glutenin subunits have been divided into three groups in *T. tauschii* (*T*): *A* = HMW-GS; *B* and *C* = LMW-GS. The molecular weights ($\times 10^{-3}$) of the protein markers (*M*) are indicated on the left side

The apparent molecular weight of the newly synthesised glutenin (39–40 kDa) was larger than that calculated (32 kDa) on the basis of the nucleotide sequence of the cloned DNA. A similar discrepancy has been observed previously for different prolamins of the *Triticeae* (Bunce et al. 1985) and may be because these proteins do not migrate as globular molecules on SDS-gels due to their unusual amino-acid composition. In the absence of reducing agents (DTT) a new band, with an apparent molecular weight of 33–34 kDa, appeared only in bacteria carrying the pET-LMW-16/10 plasmid after 2 h and 6 h of induction (data not shown). This new band may represent the monomeric form of the expressed LMW-GS that is extractable in presence of SDS but without DTT; the change in migration relative to the reduced protein, would be due to the presence of intra-chain disulphide bonds that lead to a more compact conformation.

The observation that only small amounts of the subunit were extracted in the absence of DTT, suggested that the expressed protein was accumulating mainly as insoluble disulphide-bonded polymers in *E. coli* cells. Confirmation of this was obtained by preparing soluble cytoplasmic protein and insoluble protein extracts, by centrifugation, from induced bacteria carrying the pET-LMW-16/10 construct. Lysates of *E. coli* cells were obtained using two freeze/thaw cycles in the presence of lysozyme (Lullien-Pellerin et al. 1994). After centrifugation of the lysates, the supernatant and the resulting pellet were analysed by SDS-PAGE (Fig. 6). The expressed protein was found only in the pellet and was not detected in the supernatant (lane 4, Fig. 6), indicating that in *E. coli* the protein is localised in intracellular inclusion bodies. The insoluble nature of the LMW-GS protein synthesised in *E. coli* facilitated its purification. A single centrifugation of the bacterial lysate was used to isolate the inclusion bodies in which the expressed protein was only slightly contaminated. In order to eliminate these contaminating bacterial proteins that co-purify with the recombinant protein, the inclusion bodies were washed; the addition of 0.5% Triton X100 and 10 mM EDTA to the lysis buffer for the washing step (Lullien-Pellerin et al. 1994) was effective (lane 6, Fig. 6) and yielded high-purity material (lane 8, Fig. 6). After purification the total yield varied from 20 to 30 mg/l of culture.

To confirm the identity of the protein synthesised in the induced cells carrying the recombinant plasmid pET-LMW-16/10, the expressed protein was reacted in a Western blot with monoclonal anti-prolamin antibodies reactive with the LMW-GS extracted from hexaploid wheat. Figure 7 shows that these antibodies react positively with the induced proteins and in a specific way only with the LMW-GS of *T. tauschii*.

The recombinant protein was further characterised by N-terminal amino-acid sequence analysis. The first ten residues were determined and a single major sequence (METRCIPGLE) was found with no detectable

heterogeneity. It was identical to the deduced protein sequence of the LMW-16/10 clone isolated from *T. tauschii* (see Fig. 2 for comparison).

Two-dimensional analyses (IEF × SDS-PAGE) of the total glutenin from *T. tauschii*, of the expressed protein, and of a mixture of both, allowed the resolution of the complex pattern of the C-glutenin subunits from *T. tauschii* and the identification of the protein derived from the clone LMW-16/10 (Fig. 8). As in cultivated wheats, the B-subunits in *T. tauschii* are the most basic proteins, with their relative isoelectric point ranging from 7.5 to 8.5, whereas the C-subunits have a wide range of isoelectric points, from slightly acidic (about 6.5) to highly basic (8.5). These analyses showed that the expressed protein overlapped with a single subunit of the C group of the LMW-GS from *T. tauschii* with a very basic isoelectric point (about 8.5)

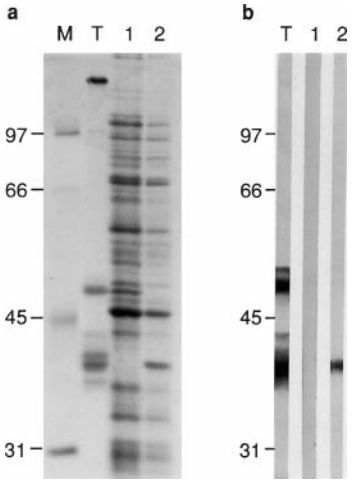

Fig. 7a, b Immunological reactivity of the LMW glutenin synthesised in *E. coli* and of the glutenin subunits extracted from *T. tauschii*. Total proteins were extracted from IPTG-induced bacteria containing either the control plasmid pET-11a (*lane* 1) or the recombinant expression vector pET-LMW-16/10 (*lane 2*) and were separated in SDS-PAGE together with the total glutenin subunits from *T. tauschii* (*T*). Gels were either stained with Coomassie blue (**a**) or reacted in Western blots with anti-prolamin antibodies reactive with LMW-GS purified from hexaploid wheats (**b**). The molecular weights ($\times 10^{-3}$) of the protein markers (*M*) are indicated on the left side

Discussion

Two different genomic clones from *T. tauschii*, the D-genome donor to the hexaploid cultivated wheats, were isolated and sequenced. The two clones, like other prolamin genes (Kreis et al. 1985), did not have their coding regions interrupted by introns and the proteins encoded by the genes possessed a proline- and a glutamine-rich domain in the form of irregular repeats, followed by a unique sequence (C-domain) often interspersed with several stretches of glutamine codons. The two isolated clones showed a high degree of homology with the previously published LMW-GS gene sequences from cultivated wheats, and one of these clones (LMW-16/10) contained an amino-terminus identical for the first 18 residues to that of a mature LMW-GS protein from hexaploid wheat (Lew et al. 1992). The three conserved regions of the C-terminal domain (A, B and C in Fig. 2) contain six of the eight cysteine residues in the LMW-GS protein, that are 100% conserved in the monomeric gamma-gliadins

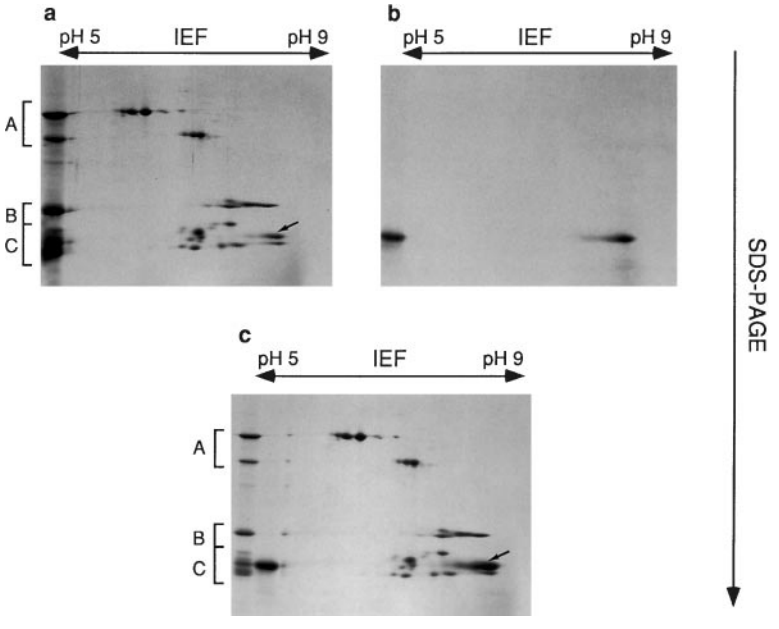

Fig. 8 Two-dimensional analyses (IEF × SDS-PAGE) of the total reduced glutenins from *T. tauschii* (**a**), of the expressed protein (**b**) and of a mixture of both (**c**). The glutenin subunits have been divided into three groups in *T. tauschii*: *A* = HMW-GS; *B and C* = LMW-GS. *Arrows* indicate the native *C* subunit from *T. tauschii* corresponding to the protein encoded by LMW-16/10 (**a**) and the same protein overlapping with the expressed protein (**c**)

and hence are considered to have the potential for forming three intra-chain disulphide bonds (Shewry and Tatham 1997). The other two cysteine residues, are separated by the repetitive domain, one cysteine very close to the N-terminus, at amino-acid position 5 in the mature protein sequence, and the other in the variable region of the C-domain. This indicates that the two additional cysteines present in the LMW-GS protein are available for the formation of interchain-disulphide bonds, as suggested by Thompson et al. (1993).

The high degree of sequence homology between all of the LMW-GS genes compared indicated that all the genes from cultivated and wild wheats, except the clone LP1211 (Pitts et al. 1988), are members of a distinct LMW-GS gene family that encodes for proteins with a characteristic N-terminal sequence starting with methionine and containing a cysteine residue at position 5. The data in the present study demonstrate that the genes also encode protein subunits that have electrophoretic mobility in the range of the C-subunits of glutenin and possess a very high basic isoelectric point. LP1211 encodes an unusual LMW-GS characterised by the absence of a typical amino-terminus and by the presence, after the signal peptide, of long polyglutamine stretches (Pitts et al. 1988). As discussed before by Cassidy and Dvorak (1991) these characteristics appear to set LP1211 apart from the other LMW-GS genes. However, the high degree of similarity detected between the promoter regions of LMW-16/10 and LP1211 suggests that LP1211 is a member of a large LMW-GS gene family.

The strong similarity between all the LMW-GS genes compared, except LP1211, may be due in part to the method used to isolate these genes. LMWG-1D1 (Colot et al. 1989) and pTdUC1 (Cassidy and Dvorak 1991) were both isolated using probes derived from sequence of the partial cDNA WHTGLG (Bartels and Thompson 1983). In the present study, the two *T. tauschii* clones (LMW-16/10 and LMW-14/1) were isolated using the cDNA clone pTdUC1 as a probe. This may have excluded the isolation of some members of the LMW-GS gene family. These might include those genes whose amino-terminus matches the protein sequence beginning with SHIPGLE determined by Tao and Kasarda (1989) and Lew et al. (1992), and the genomic clone, LP1211. However, WHTGLIGBA, a complete gene for LMW-GS, and the other partial related sequences WHTGLIGBB and WHTGLIGBC (Okita et al. 1985) were selected randomly from a cDNA library and probably represent the most abundantly expressed genes in cv Cheyenne.

On the assumption that sequences important in gene expression are likely to be conserved among a group of genes with the same pattern of expression, we have carried out a detailed comparison between the flanking non-coding sequences of the *T. tauschii* clone LMW-16/10 and those of 16 different prolamin genes from wheat, barley and rye (Table 2). The comparison led to the distinction of five different gene families and confirmed the evolutionary relationships determined previously for these genes mainly on the basis of the coding region (Shewry et al. 1994). The similarities and differences between and within the first four families, including all the sulphur-rich and sulphur-poor prolamins, suggest that they have a common origin and that at least three of the four families (the alpha- and beta-gliadins are present only in wheat) may have diverged before the separation of the ancestral diploid wheat, barley and rye species. HMW prolamins showed little or no homology to the 5′ and 3′ non-coding sequences of the other prolamin genes, indicating that they constitute a very distinct family that probably diverged independently from the other prolamin genes before the separation of the ancestral diploid wheat, barley and rye species. This considerable divergence, in particular for the 5′ flanking sequence, probably indicates that HMW prolamin gene expression might be under a regulatory mechanism different from those of the other prolamin genes, as was clearly demonstrated in barley for D-hordein with respect to the other hordeins (B and C) (Sorensen et al. 1996).

From an evolutionary point of view it is of interest to note the high degree of homology detected between the two *T. tauschii* clones and the previously published LMW-GS gene sequences from cultivated wheats. In particular a nucleotide sequence comparison of LMW-16/10 with the genomic clone LMWG-1D1 located on chromosome 1D of *T. aestivum* (Colot et al. 1989) revealed the presence of ten nucleotide changes and two deletions over the 3500 nucleotides where both were sequenced (data not shown). Considering that the wild wheat *T. tauschii* and the cultivated hexaploid wheat *T. aestivum* diverged at least 10 000 years ago, it is quite remarkable that except for a few point mutations and the two deletions these two genes are nearly identical. Similar results have been obtained by the comparison of two alpha/beta-gliadin genomic clones isolated from the hexaploid wheat Yamhill and from the diploid wheat *T. urartu*, the A-genome donor to the tetraploid and hexaploid cultivated wheats (Reeves and Okita 1987).

All the genes compared in Fig. 3 contain a typical TATA box. In addition, other motifs similar to known regulatory signals of eukaryotic gene expression, the CAAT and AGGA boxes (Messing et al. 1983), are present further 5′ to the gene sequences (see Fig. 3). The sequences around the initiator codon ATG, that are considered to be important in efficient translation initiation (Kozak 1986), are highly conserved among all the genes compared. In particular the bases just upstream from the translation start are very similar to the eukaryotic consensus CCACC, whereas the HMW prolamin genes possess three additional bases (GAG) between the CCACC and the initiator ATG (Anderson et al. 1989). Other conserved regions with no known clear function are found between the endosperm and

TATA boxes (Fig. 3). The −235 box contains an A + T-rich sequence. A + T-rich sequences are frequently found in the promoter regions of plant genes. Specific binding of the high-mobility group (HMG) proteins to these sequences has been suggested to contribute to the assembly of active transcription complexes. The second conserved region (box –177), mostly consisting of C and A residues, was first recognised in an alpha-gliadin promoter sequence by Reeves and Okita (1987) and was termed the 'CACA' box. Another short conserved region (box –204) has not been noted previously. It consists of 11 residues with a consensus sequence TCCAAAGTAC and in which the first five nucleotides are identical for all the prolamin genes compared.

A comparison of the LMW-16/10 3′ flanking sequence with that of related cDNAs indicated that polyadenylation occurs at a number of different sites. Similar observations have been made with other plant genes (Dean et al. 1986). It is interesting to note that all the genes compared in Fig. 4 contain some nearly perfect inverted repeats (indicated by arrows), near the AATAAA motifs. These sequences have the possibility of forming stabilised secondary stem-loop structures which may be important for the termination of transcription and/or polyadenylation of the transcript.

Functional and structural studies of LMW-GSs have always been limited by the difficulty of preparing adequate amounts of single homogeneous polypeptides. It is therefore particularly attractive to produce single components via bacterial expression of cDNA or genomic clones. We have obtained high-level expression of a LMW-GS using an inducible *E. coli* expression system. The LMW-GS was found in inclusion bodies and could be readily recovered by centrifugation of bacterial lysates. The levels of expression achieved in the present study (20–30 mg/l of culture) are appreciably higher than those reported for alpha- and the gamma-gliadins in yeast (up to 4 mg/l) (Pratt et al. 1991; Blechl et al. 1992) and comparable with those obtained for C-hordein and the wheat HMW-GS in *E. coli* with a similar vector to that used in the present study (Galili 1989; Tamas et al. 1994). High levels of expression (30–50 mg/l) were achieved, also, with a wheat LMW-GS using a baculovirus-based vector in cultured insect cells (Thompson et al. 1994). However, in this case the expressed protein had anomalous solubility properties. In contrast, the subunit expressed in the current study exhibited solubility characteristics identical to those of the native LMW-GS, being soluble in aqueous alcohol solutions or in 0.1 M acetic acid in the presence of reducing agents. Antibodies recognising a wide range of epitopes in native LMW-GS proteins also reacted well with the protein expressed in bacteria. Although further work is needed to compare the structure of the LMW-GS protein expressed in bacteria to the native form, the evidence to-date suggests that the protein from bacteria represents a valid source of LMW-GS protein for testing its technological properties. The experiments reported in Lee et al. (1998) provide the first data in this area and demonstrate that LMW-16/10 has a clear "activity" in changing the mixing characteristics of dough.

Acknowledgments The authors are grateful to Dr. M. Sissons for his help in carrying out the immunoblot analyses, and Dr. S. Rahman for his help in analysing lambda clones and many other valuable discussions.

References

Anderson OD, Litts JC, Gautier MF, Greene FC (1984) Nucleic acid sequence and chromosome assignment of a wheat storage protein gene. Nucleic Acids Res 12:8129–8144

Anderson OD, Greene FC, Yip RE, Halford NG, Shewry PR, Malpica-Romero JM (1989) Nucleotide sequences of the two high-molecular-weight glutenin genes from the D genome of a hexaploid bread wheat, *T. aestivum* L. cv Cheyenne. Nucleic Acids Res 17:461–462

Andrews JL, Skerritt JH (1996) Identification of wheat-rye translocation lines using antibody probes for *Gli-B1* and *Sec-1*. J Cereal Sci 23:61–72

Bartels D, Thompson RD (1983) The characterisation of cDNA clones coding for wheat storage proteins. Nucleic Acids Res 11:2961–2967

Bartels D, Altosaar I, Harberd NP, Barker RD, Thompson RD (1986) Molecular analysis of gamma-gliadin gene families at the complex *Gli-1* locus of bread wheat (*T. aestivum* L.). Theor Appl Genet 72:845–853

Blechl AE, Thrasher KS, Vensel WH, Greene FC (1992) Purification and characterisation of wheat alpha-gliadin synthesized in the yeast, *Saccharomyces cerevisiae*. Gene 116:119–127

Brandt A, Montembault A, Cameron-Mills V, Rasmussen S (1985) Primary structure of a B1 hordein gene from barley. Carlsberg Res Commun 50:333–345

Bunce NAC, White RP, Shewry PR (1985) Variation in estimates of molecular weights of cereal prolamins by SDS-PAGE. J Cereal Sci 3:131–142

Cameron-Mills V, Brandt A (1988) A gamma-hordein gene. Plant Mol Biol 11:449–461

Cassidy BG, Dvorak J (1991) Molecular characterization of a low-molecular-weight glutenin cDNA clone from *Triticum durum*. Theor Appl Genet 81:653–660

Colot V, Bartels D, Thompson R, Flavell R (1989) Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 216:81–90

D'Ovidio R., Tanzarella OA, Porceddu E (1992) Nucleotide sequence of a low-molecular-weight glutenin from *Triticum durum*. Plant Mol Biol 18:781–784

Dean C, Tamaky S, Dunsumuir P, Favreau M, Katayama C, Dooner H, Bedbrook J (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo. Nucleic Acids Res 14:2229–2240

Entwistle J (1988) Primary structure of a C-hordein gene from barley. Carlsberg Res Commun 53:247–258

Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in an endosperm storage protein gene from barley, wheat and maize. Nucleic Acids Res 13:7327–7339

Galili G (1989) Heterologous expression of a wheat high-molecular-weight glutenin gene in *Escherichia coli*. Proc Natl Acad Sci USA 86:7756–7760

Gupta RB, MacRitchie F (1991) A rapid one-step one-dimensional SDS-PAGE procedure for analysis of the subunit composition of glutenin in wheat. J Cereal Sci 14 : 105–109

Halford NG, Forde J, Anderson OD, Greene FC, Shewry PR (1987) The nucleotide and deduced amino-acid sequences of an HMW glutenin subunit gene from chromosome 1B of bread wheat (*Triticum aestivum* L.) and comparison with those of genes from chromosomes 1 A and 1D. Theor Appl Genet 75 : 117–126

Hill DE, Hope IA, Macke JP, Struhl K (1986) Saturation mutagenesis of the yeast *his*3 regulatory site: requirements for trancriptional induction and for binding by GCN4 activator protein. Science 234 : 451–457

Hull GA, Halford NG, Kreis M, Shewry PR (1991) Isolation and characterisation of genes encoding rye prolamins containing a highly repetitive sequence motif. Plant Mol Biol 17 : 1111–1115

Jackson EA, Holt LM, Payne PI (1983) Characterisation of high-molecular-weight gliadin and low molecular weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. Theor Appl Genet 66 : 29–37

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Pomeranz Y (ed) Wheat is unique. Am Assoc Cereal Chem, St. Paul, Minnesota, pp 277–302

Kasarda DD, Tao HP, Evans PK, Adalsteins AE, Yuen SW (1988) Sequencing of protein from a single spot of a 2-D gel pattern: N-terminal sequence of a major wheat LMW-glutenin subunit. J Exp Bot 39 : 899–906

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44 : 283–292

Kreis M, Shewry PR, Forde BG, Forde J, Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surveys Plant Mol Cell Biol 2 : 253–317

Lee Y-K, Bekes F, Gras P, Ciaffi M, Appels R., Morell M (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. IV. Functional properties of products from individual genes. Theor Appl Genet 98 : 149–155

Lew EJ, Kuzmicky DD, Kasarda DD (1992) Characterization of low-molecular-weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulphate-polyacrylamide gel electrophoresis, and N-terminal amino-acid sequencing. Cereal Chem 69 : 508–515

Lullien-Pellerin V, Gavalda S, Joudrier P, Gautier MF (1994) Expression of a cDNA encoding the wheat CM16 protein in *Escherichia coli*. Protein Express Purif 5 : 218–224

MacRitchie F (1992) Physicochemical properties of wheat proteins in relation to functionality. Adv Food Nutr Res 36 : 1–87

Masci S, Lafiandra D, Porceddu E, Lew EJ, Tao HP, Kasarda DD (1993) D-glutenin subunits: N-terminal sequences and evidence for the presence of cysteine. Cereal Chem 70 : 581–585

Messing J, Geraghty D, Heidecker G, Hu NT, Kridl J, Rubenstein I (1983) Plant gene structure. In: Kosuge T, Meredith CP, Holaender A (eds) Genetic engineering of plants: an agricultural perspective. Plenum Press, New York, pp 211–227

Mitchell PJ, Wang C, Tjian R (1987) Positive and negative regulation of trancription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell 50 : 847–861

Okita TW, Chessbrough V, Reeves CD (1985) Evolution and heterogeneity of the alpha/beta-type and gamma-type gliadin DNA sequences. J Biol Chem 260 : 8203–8213

Piette J, Hirai SI, Yaniv M (1988) Constitutive synthesis of activator protein-1 trancription factor after viral transformation of mouse fibroblasts. Proc Natl Acad Sci USA 85 : 3401–3405

Pitts EG, Rafalski JA, Hedgcoth C (1988) Nucleotide sequence and encoded amino-acid sequence of a genomic gene region for a low-molecular-weight glutenin. Nucleic Acids Res 16 : 11376–11376

Pratt KA, Madgwick PJ, Shewry PR (1991) Expression of a wheat gliadin protein in yeast (*Saccharomyces cerevisiae*). J Cereal Sci 14 : 223–229

Rafalski JA (1986) Structure of wheat gamma-gliadin genes. Gene 43 : 221–229

Rahman S, Abrahams S, Abbott D, Mukay Y, Samuel M, Morell M, Appels R (1997) A complex arrangement of genes at the starch branching enzyme-I locus in the D-genome donor of wheat. Genome 40 : 465–474

Ransone LJ, Verma MI (1990) Nuclear proto-oncogenes fos and jun Annu Rev Cell Biol 6 : 539–557

Reeves CD, Okita TW (1987) Analyses of alpha/beta-type gliadin genes from diploid and hexaploid wheat. Gene 52 : 257–266

Sainova OV, Mekhedov SL, Zhelnin LG, Khokhlova TA, Ananev EV (1993) Nucleotide sequence of a barley C-hordein gene. Genetika 29 : 1070–1079

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74 : 5463–5467

Scheets K, Hedgcoth C (1988) Nucleotide sequence of a gamma-gliadin gene: comparisons with other gamma-gliadin sequences show the structure of gamma-gliadin genes and the general primary structure of gamma-gliadins. Plant Sci 57 : 141–150

Shewry PR, Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25 : 207–227

Shewry PR, Halford NG, Tatham AS (1989) The high-molecular-weight subunits of wheat, barley and rye: genetics, molecular biology, chemistry and role of wheat gluten structure and functionality. Oxford Surveys Plant Mol Cell Biol 6 : 163–219

Shewry PR, Halford NG, Tatham AS (1992) The high-molecular-weight subunits of wheat glutenin. J Cereal Sci 15 : 105–120

Shewry PR, Mervyn JM, Tatham AS (1994) The prolamin storage proteins of wheat and related cereals. Prog Biophys Mol Biol 61 : 37–59

Sorensen MB, Muller M, Skerritt JH, Simpson D (1996) Hordein promoter methylation and trancriptional activity in wild-type and mutant barley endosperm. Mol Gen Genet 250 : 750–760

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185 : 60–89

Sumner-Smith M, Rafalski JA, Sugiyama T, Stoll M, Soll D (1985) Conservation and variability of wheat alpha/beta-gliadin genes. Nucleic Acids Res 13 : 3905–3916

Tamas L, Greenfield J, Halford NG, Tatham AS, Shewry PR (1994) A beta-turn-rich barley seed protein is correctly folded in *Escherichia coli*. Protein Express Purif 5 : 357–363

Tao PH, Kasarda DD (1989) Two-dimensional gel mapping and N-terminal sequencing of LMW-glutenin subunits. J Exp Bot 40 : 1015–1020

Thompson S, Bishop DHL, Tatham AS, Shewry PR (1993) Exploring disulphide bond formation in a low-molecular-weight subunit of glutenin using a baculovirus expression system. In: Proc 5th Int Wkshp on Wheat Gluten Proteins. Association of Cereal Research, Detmold, Germany, pp 345–355

Thompson S, Bishop DHL, Madgwick P, Tatham AS, Shewry PR (1994) High-level expression of a wheat LMW glutenin subunit using a baculovirus system. J Agric Food Chem 42 : 426–431

Vincente-Carbajosa J, Beritashvili DR, Kraev AS, Skryabin KG (1992) Conserved structure and organisation of B hordein genes in the Hor 2 locus of barley. Plant Mol Biol 18 : 453–458