Theor Appl Genet (1999) 98: 455–461

R. D'Ovidio · C. Marchitelli
L. Ercoli Cardelli · E. Porceddu

Sequence similarity between allelic *Glu-B3* genes related to quality properties of durum wheat

Received: 8 May 1998 / Accepted: 5 August 1998

Abstract Low-molecular-weight glutenin subunits (LMW-GSs) are wheat endosperm proteins mostly encoded by genes located at the *Glu-3* loci. These proteins are of particular interest in durum wheat because a correlation between LMW-GSs encoded by genes at the *Glu-B3* locus and the pasta-making quality of durum wheat semolina has been shown. We isolated and characterized two allelic *lmw-gs* genes located at the *Glu-B3* locus and present in durum wheat lines displaying different qualitative properties. The clones pLMW1CL and λLMW3.1 were found to contain allelic sequences encoding LMW-GSs belonging to the good and poor quality-related groups named LMW-2 and LMW-1, respectively. The LMW-GSs specified by these genes have very large repetitive domains which are composed of repeats regularly distributed along the domain. The main difference between these two proteins is an insertion of 13 amino acids within the repetitive domain which, by itself, seems insufficient to explain the qualitative differences between LMW-2 and LMW-1. These results further support the hypothesis that the greater amount of LMW-2, rather than sequence peculiarities, accounts for the better quality observed in durum wheat cultivars possessing these subunits. The characterization of the complete primary structure of these alleles, other than providing information for an understanding of the structure-function relationship among LMW-GSs and furnishing basic material for wheat engineering, should also assist in our understanding of the evolutionary relationship between the different *lmw-gs* genes.

Communicated by G. E. Hart

R. D'Ovidio (✉) · C. Marchitelli · L. E. Cardelli · E. Porceddu
Dipartimento di Agrobiologia e Agrochimica, Università della Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy
Fax: +39 761 357242
E-mail: dovidio@unitus.it

Key words *Triticum durum* · LMW glutenin gene · Allelic-specific primers · *Glu-B3* locus · Wheat quality

Introduction

Low-molecular-weight glutenin subunits (LMW-GSs) are a specific class of prolamins which, along with the high-molecular-weight glutenin subunits (HMW-GSs), form the backbone of the wheat endosperm glutenin polymer, whose molecular weight can be in the range of millions (Wrigley 1996). Molecular analysis of the genes encoding these proteins is of particular interest both for basic and applied research because they belong to highly polymorphic multigene families and their encoded proteins are those mainly responsible for the qualitative properties of durum wheat dough.

The structures of HMW-GSs and their encoding genes have been well characterized, and the nucleotide sequences of the complete set of HMW-GS genes present in the bread wheat cultivar 'Cheyenne' have been reported (reviewed in Shewry et al. 1992). Genetic and technological studies have also shown that specific HMW-GS allelic variants are associated with bread-making quality (Payne 1987).

LMW-GSs are represented by several components encoded by genes at the orthologous *Glu-3* loci. Noteworthy is the situation in durum wheat where the presence of specific alleles at the *Glu-B3* locus relates with differences in pasta-making quality. In particular, lines possessing a specific group of LMW-GSs, named LMW-2, have superior quality characteristics than lines possessing the allelic group, named LMW-1 (Pogna et al. 1990).

Despite their important effect on the qualitative properties of durum wheat and on the understanding of the molecular events underlying the evolution of gene families, only a limited number of *lmw-gs* genes have so far been characterized and for none of them are the

relative allelic forms or their direct influence on dough properties known.

In this paper we report the isolation and nucleotide sequences of two allelic *lmw-gs* genes which encode glutenin subunits correlated with opposite qualitative properties of durum wheat semolina.

Materials and methods

Plants

Durum wheat (*Triticum durum* Desf) genotypes possessing quality-related allelic variants at the *Glu-B3* locus have been used. The biotype γ-42 of cv 'Lira' possesses the poor quality-related LMW-GS, named LMW-1, whereas Line 21 possesses the good quality-related LMW-GS, named LMW-2. Two biotypes of cv 'Duramba' possessing LMW-1 or LMW-2 and their F_2 segregating population were also used in the present analysis.

SDS-PAGE analysis

Single seeds were crushed and gliadins were removed with 50% propanol. Glutenin subunits were extracted (1:10 w/v). with 125 m*M* TRIS-HCl, pH 6.8, buffer containing 2% SDS, 10% glycerol, 10% dimethylformamide and 1% DTT. Three-microliter aliquots were loaded on a Mini SDS-PAGE (T = 11%, C = 2.67%) and run according to the instruction manual (Bio-Rad, Richmond, Calif.). Gels were stained overnight with a 12% trichloroacetic acid solution containing 0.05% Coomassie Brilliant Blue R-250 in absolute ethanol (1% w/v) and destained in tap water.

DNA extraction

Genomic DNA was isolated from 5 g of leaves collected from single plants as reported in D'Ovidio et al. (1992a).

Polymerase chain reaction (PCR) amplification

PCR analyses with primers UTV7F/UTV7R and UTV7FEcoRI/UTV7RXbaI were performed following the protocol reported by D'Ovidio (1993). Primers UTV7F and UTV7R are specific for genes encoding LMW-GSs belonging to LMW1 and LMW-2, and correspond to the previously reported oligonucleotides named A and B, respectively (D'Ovidio 1993). The UTV7FEcoRI possesses the same sequence of UTV7F with the addition of an *Eco*RI restriction site at the 5′ end (5′-GCAGAATTCGTTGCGGCGACAAGTGCAA-3′), whereas the UTV7RXbaI possesses the same sequence of UTV7R with the addition of an *Xba*I restriction site at the 5′ end (5′-TAACTCTAGAGTAGGCACCAACTCCGGTGC-3′).

PCR analyses with primers UTV9F/UTV9R were carried out in a final reaction volume of 100 µl by using 100–300 ng of genomic DNA, 2.5 units of *Taq* DNA polymerase (Boeheringer Mannheim, Germany), 1 × *Taq* PCR buffer (Boeheringer Mannheim, Germany), 250 ng of each of the two primers and 200 µ*M* of each deoxyribonucleotide. Amplification conditions were 30 cycles at 94°C for 1 min and 30 s, 62°C for 1 min, and 72°C for 1 min. A final step at 72°C for 5 min was also performed. The oligonucleotides used have the following sequences: UTV9F) 5′ AGCCATATCCCTGGTTTGGAG 3′; UTV9R) 5′ TAAGTAGGCACCAACTCCGG 3′. Aliquots (10 µl) of the amplification products were fractionated on a 1.5% agarose gel in 1 × TBE buffer following standard procedures (Sambrook et al. 1989).

Digoxigenin labeling and hybridization experiments

Southern blot analyses were carried out following standard procedures (Sambrook et al. 1989). Digoxigenin-labeled probes were prepared using 300 ng of the pLMW21 clone (D'Ovidio et al. 1992b) and the 'Nick Translation Kit' (Boeheringer Mannheim, Germany) following the manufacturer's instructions.

Cloning, nucleotide sequencing and computer analysis

The amplification product of about 1.15 kbp obtained from Line 21 with primers UTV7F/UTV7R was purified from the agarose gel using the Gene Clean Kit (Bio101, La Jolla, Calif.) and used in PCR reactions using primers UTV9F/UTV9R. The new amplification product was purified and ligated into the *Eco*RV dephosphorylated site of the pGEM-T plasmid vector (Promega, Madison, Wis.) using standard procedures (Sambrook et al. 1989). The *Escherichia coli* strains NM522 (New England Biolabs), DH5α (Life Technologies) and SURE (Stratagene) were used for transformation experiments.

The amplification product of about 1.1 kbp obtained from biotype γ-42 with primers UTV7F/UTV7R was purified as above and used in PCR reactions using primers UTV7FEcoRI/UTV7RXbaI. Two micrograms of the amplification product obtained with this new pair of primers was double-digested with *Eco*RI and *Xba*I, fractionated on 1.5% agarose gel, purified and ligated into the λGEM-4 *Eco*RI-*Xba*I arms (Promega, Madison, Wis.) following standard procedures (Sambrook et al. 1989). The packaging of ligated DNA and titration of packaged phages were performed following the manufacturer' instructions (Promega, Madison, Wis.). The *Escherichia coli* strain LE392 was used for identifying the recombinant phage clones. Transfer of phages to nylon filters and plaque hybridization experiments were carried out following standard procedures (Sambrook et al. 1989). Nucleotide sequences were determined using the Thermo Sequenase™ radiolabeled terminator cycle sequencing kit (Amersham, UK.).

The PC/GENE computer program (IntelliGenetics, USA) was used to analyze the sequence data

Results

SDS-PAGE, PCR analysis and cloning

SDS-PAGE analysis of LMW-GSs of the biotype 'Lira' γ-42 and line 21 indicated that the former possesses the LMW-1 whereas the latter possesses the LMW-2 (Fig. 1A). Both groups of proteins are composed of a few polypeptides whose molecular weight ranges between about 38,000 and 42,000, but their precise number has not been determined due to extensive overlapping even in two-dimensional gel electrophoresis analyses. However, some polypeptides of the LMW-2 group possessed a stronger staining intensity than those present in the LMW-1 group (Fig. 1A). This difference was indicative of their larger amount, as demonstrated by Masci et al. (1995), and did not derive from differences in samples loading.

A PCR assay using primers UTV7F and UTV7R was performed on genomic DNA from the same genotypes analyzed by SDS-PAGE. Analyses on agarose gel showed the presence of amplification fragments of about 1.0 kb and 1.1 kbp in biotype 'Lira' γ-42 and of

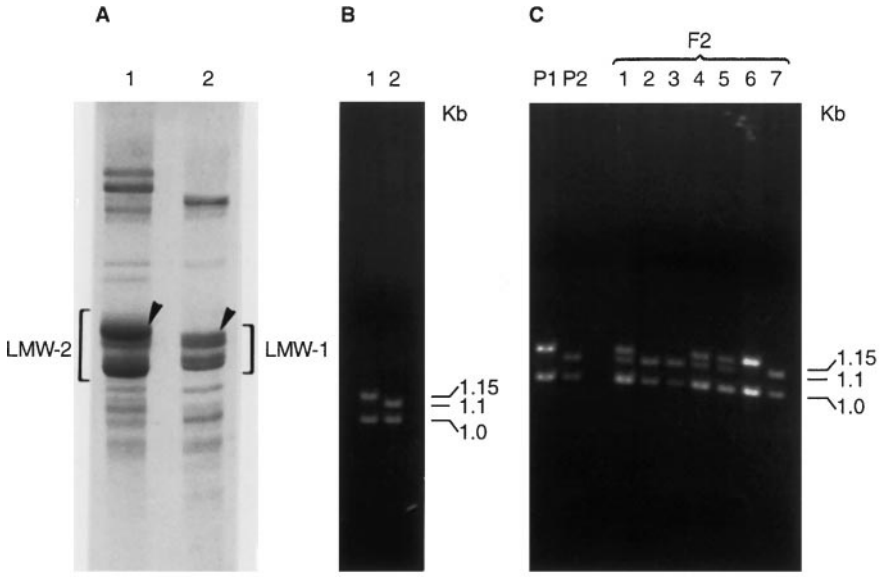

Fig. 1 **A** SDS-PAGE of glutenin subunits from durum wheat line 21 (*lane 1*) and biotype γ-42 of durum wheat cultivar 'Lira' (*lane 2*). The group of LMW-GSs designated LMW-1 and LMW-2 are indicated in *parentheses*. The *arrowheads* indicate the larger (slower moving) polypeptides in both LMW-1 and LMW-2. **B** Agarose gel (1.5%) of PCR products obtained with primers UTV7F/UTV7R specific for LMW-1 and LMW-2. *Lane 1* Durum wheat line 21, *lane 2* biotype γ-42 of durum wheat cultivar 'Lira'. **C** Agarose gel (1.5%) of PCR products obtained with primers UTV7F/UTV7R from durum wheat genotypes and their F_2 progeny. Pattern analysis of the F_2 population clearly showed that the 1.15-kbp and 1.1-kbp fragments are allelic, as they were present in the homozygous or heterozygous condition in the different genotypes of the population. *P1* Biotype of cv 'Duramba' expressing LMW-2, *P2* biotype of cv 'Duramba' expressing LMW-1, *lanes 1–7* genotypes of the F_2 progeny

1.0 kb and 1.15 kbp in line 21 (Fig. 1B). The same PCR assay was performed on genomic DNA of a F_2 segregating population of a cross between the two biotypes of the durum wheat cultivar 'Duramba' possessing LMW-1 or LMW-2. Pattern analyses of the F_2 population clearly showed that the 1.15-kbp and 1.1-kbp fragments are allelic (Fig. 1C).

The purified 1.15-kbp amplification product from line 21 was ligated into the *Eco*RV site of the pGEM-T plasmid vector. After transformation into the SURE *Escherichia coli* strain, the recombinant colonies were analyzed to detect the presence of the 1.15-kbp fragment. Several recombinant clones contained an insert, but none of these had the size expected. Restriction digestions and sequencing analyses of different recombinant clones showed that the sizes of the insert were about 50–200 bp shorter than that expected and that this was due to single deletions within the repetitive domain (data not shown). Similar results were also obtained using *E. coli* strains DH5α and NM522, and by other authors attempting to clone *lmw-gs* genes by RT-PCR (S. Altenbach, personal communication). We made an additional attempt to clone the 1.15-kbp PCR product using a pair of primers named UTV9F and UTV9R. UTV9F was developed on the basis of direct nucleotide sequencing of the 1.15-kbp fragment using the UTV7F primer. The position of UTV9F is 17 bp downstream from the 3′ end of the UTV7F and, due to the unclear result obtained from direct nucleotide sequencing, the sixth nucleotide of this primer might have been a C or a T. However, since the use of either C or T does not cause any change at the corresponding amino acid level, we have synthesized the UTV9F by including a T in this position. The UTV9R was developed on the basis of the available LMW-GS nucleotide sequences and is 3 bp shorter at the 3′ end and 3 bp longer at the 5′ end of the UTV7R. The UTV9F/UTV9R primers were used to amplify the 1.15-kbp PCR product obtained with primers UTV7F/UTV7R, and the new amplification product was ligated into the pGEM-T vector. The transformation result produced a few recombinant colonies that showed the presence of the expected PCR fragment and one of these, designated pLMW1CL, was sequenced.

The cloning of the 1.1-kbp PCR product from 'Lira' γ-42 was obtained following a different strategy. The 1.1-kbp amplification product was fractionated by agarose gel electrophoresis, purified from the gel and subjected to a second round of PCR reactions with primers UTV7FEcoRI/UTV7RXbaI. This new PCR product was double-digested with *Eco*RI and *Xba*I restriction enzymes and ligated into the λGEM-4 arms. The recombinant phages were identified by plaque hybridization using the digoxigenin-labeled pLMW21 clone. The size of the inserts contained in the recombinant phages was checked by PCR using UTV7F/UTV7R and by restriction digestions. Only the PCR analysis showed that several clones contained an insert of the correct size – all the restriction digestions with

*Eco*RI and *Xba*I failed to produce any digested product. Consequently, it was not possible to subclone the cloned insert into a plasmid vector. One of the recombinant phages, named λLMW3.1, was used for sequencing analysis.

Sequencing and comparison analyses

The nucleotide sequence of the pLMW1CL clone (patent no. MI96A002663) is 1,107 bp long and shows the same general structure of previously reported *lmw-gs* genes. This clone lacks the typical secretion signal because of the cloning strategy used. However, the presence of part of this signal was ascertained by direct sequencing of the 1.15 kb PCR product from which the pLMW1CL derives. Nucleotide sequence comparison showed that the signal sequence of the 1.15-kb PCR product was similar to that found in the 1.1-kb PCR product. The repetitive domain of the pLMW1CL clone is 558 bp long and is composed of 25 repeats, each ranging from 18 to 27 bp (Fig. 2A). The consensus sequence is a 24-bp fragment coding for the octapeptide PPFSQQQQ (Fig. 2C). The pLMW1CL contains an open reading frame encoding a polypeptide of 369 amino acids starting with the sequence SHIPGL- (Fig. 3). This sequence corresponds to the predominant

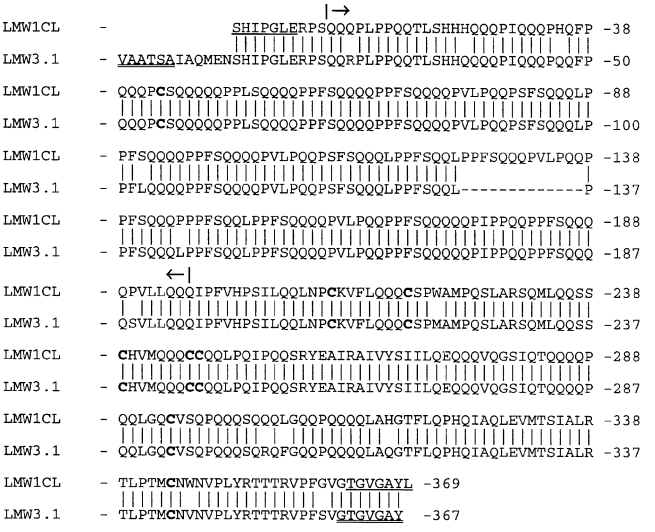

Fig. 3 Comparison between the deduced amino acid sequences of the *lmw-gs* genes contained in the pLMW1CL clone (EMBL Data library accession number AJ007746) and λLMW3.1 clone (EMBL Data library accession number Y18159). Conserved positions are indicated by a *vertical line*, and missing amino acids are indicated by *dashes*. The repetitive domain is indicated by *arrows*. Cysteine residues are indicated in *bold*. The amino acids corresponding to the UTV9F and UTV9R primers in the pLMW1CL clone are *underlined*, and the amino acids corresponding to the UTV7F and UTV7R primers in the λLMW3.1 clone are *double-underlined*

N-terminal amino acid sequence found in the larger polypeptide of the LMW-2 group (Masci et al. 1995; Masci et al. in preparation). Following a short N-terminal region (10 amino acids), the deduced protein contains a repetitive domain of 186 amino acids and a C-terminal domain of 173 amino acids. The polypeptide possesses a calculated molecular mass of 42,242, an isoelectric point of 8.32 and a glutamine and proline content of 34.6% and 16.8%, respectively. Its hydropathy profile (data not shown) revealed the hydrophilic

Fig. 2A–D Alignment of the nucleotides and amino acid repeats present in the repetitive domain of the pLMW1CL clone and in the λLMW3.1 clone. **A, B** Distribution of the repeats within the repetitive domain which lies between nucleotides 31 and 588 in the pLMW1CL clone (**A**) and between nucleotides 67 and 585 in the λLMW3.1 (**B**). The figure shows also the regular distribution of the repeats within the repetitive domain which lies between amino acids 11 and 196 in the pLMW1CL clone (**C**) and between amino acids 23 and 195 in the λLMW3.1 (**D**). Missing repeats are indicated by *dashes*. *Cons* Consensus sequence

A

```
                                CAG CAA
CAA CCA TTA CCA CCA CAA
CAA ACA TTA TCG CAC CAC CAC CAA CAA
CAA CCC ATC CAA CAA CAA CCA
CAC CAA TTT CCA CAA CAG
CAA CCA TGT TCA CAG CAA CAA CAA CAA
CCA CCA TTA TCG CAA CAA CAA CAA
CCA CCA TTT TCG CAG CAA CAA CAA
CCA CCA TTT TCA CAG CAA CAA
CAA CCA GTT CTA CCG CAA CAA
CCA TCA TTT TCG CAG CAA CAA CTA
CCA CCA TTT TCG CAG CAA CAA CAA
CCA CCA TTT TCA CAG CAA CAA
CAA CCA GTT CTA CCG CAA CAA
CCA TCA TTT TCG CAG CAA CAA CTA
CCA CCA TTT TCA CAG CAA CTA
CCA CCA TTT TCG CAG CAA CAA
CCA GTA CTA CCG CAA CAA
CCA CCA TTT TCG CAG CAA CAA CCA
CCA CCA TTT TCA CAG CAA CTA
CCA CCA TTT TCG CAG CAA CAA CAA
CCA GTA CTA CCG CAA CAA
CCA CCA TTT TCG CAA CAA CAA CAA
CAA CCA ATT CCA CCG CAA CAA
CCA CCA TTT TCG CAA CAA CAA
CAG CCA GTT CTA CTG CAA CAA CAA
```

B

```
                                CAG CAA
CGA CCA TTA CCA CCA CAA
CAA ACA TTA TCG CAC CAC CAA CAA CAA
CAA CCC ATT CAA CAA CAA CCA
CAA CAA TTT CCA CAA CAG
CAA CCA TGT TCA CAG CAA CAA CAA CAA
CCA CCA TTA TCG CAA CAA CAA CAA
CCA CCA TTT TCG CAG CAA CAA CAA
CCA CCA TTT TCA CAG CAA CAA
CAA CCA GTT CTA CCG CAA CAA
CCA TCA TTT TCG CAG CAA CAA CTA
CCA CCA TTT TTG CAG CAA CAA CAA
CCA CCA TTT TCA CAG CAA CAA
CAA CCA GTT CTA CCG CAA CAA
CCA TCA TTT TCG CAG CAA CAA CTA
CCA CCA TTT TCA CAG CAA CTA
--- --- --- --- --- --- ---
--- --- --- --- --- --- ---
CCA CCA TTT TCG CAG CAA CAA CTA
CCA CCA TTT TCA CAG CAA CTA
CCA CCA TTT TCG CAG CAA CAA CAA
CCA GTA CTA CCG CAA CAA
CCA CCA TTT TCG CAA CAA CAA CAA
CAA CCA ATT CCA CCG CAA CAA
CCA CCA TTT TCG CAA CAA CAA
CAG TCA GTT CTA CTG CAA CAA CAA
```

C

```
        QQ
1)  QPLPPQ
2)  QTLSHHQQ
3)  QPIQQQP
4)  HQFPQQ
5)  QPCSQQQQQ
6)  PPLSQQQQ
7)  PPFSQQQQ
8)  PPFSQQQQ
9)  PVLPQQ
10) PSFSQQQL
11) PPFSQQQQ
12) PPFSQQQQ
13) PVLPQQ
14) PSFSQQQL
15) PPFSQQL
16) PPFSQQQ
17) PVLPQQ
18) PPFSQQQP
19) PPFSQQL
20) PPFSQQQQ
21) PVLPQQ
22) PPFSQQQQ
23) QPIPPQQ
24) PPFSQQQQ
25) PVLLQQQ

Cons PPFSQQQQ
```

D

```
        QQ
1)  RPLPPQ
2)  QTLSHHQQQ
3)  QPIQQQP
4)  QQFPQQ
5)  QPCSQQQQQ
6)  PPLSQQQQ
7)  PPFSQQQQ
8)  PPFSQQQQ
9)  PVLPQQ
10) PSFSQQQL
11) PPFLQQQQ
12) PPFSQQQQ
13) PVLPQQ
14) PSFSQQQL
15) PPFSQQL
16) -
17) -
18) PPFSQQQL
19) PPFSQQL
20) PPFSQQQQ
21) PVLPQQ
22) PPFSQQQQ
23) QPIPPQQ
24) PPFSQQQQ
25) SVLLQQQ

Cons PPFSQQQQ
```

character of the repetitive domain and the general hydrophobic character of the N-terminal and C-terminal regions. The deduced LMW-GS also contains eight cysteine residues, seven of which are located in the C-terminal domain and one in the upstream region of the repetitive domain (Fig. 3).

The nucleotide sequence of λLMW3.1 is 1,102 bp long and corresponds to a *lmw-gs* gene. The deduced amino acid sequence is composed of 367 amino acids and includes the complete coding region and part of the signal peptide typical of a LMW-GS (Fig. 3). The repetitive domain is 519 bp long and is composed of 23 repeats ranging from 18 to 27 bp (Fig. 2B). The consensus sequence is a 24-bp fragment coding for an octapeptide whose sequence is PPFSQQQQ (Fig. 2D). The sequence SHIPGL was considered to be the N-terminal amino acid sequence of the mature protein, as it corresponds to the predominant N-terminal amino acid sequence found in the larger polypeptide composing the LMW-1 group (Masci et al. 1995). Based upon the above assumption, it was calculated that the λLMW3.1 clone codes for a mature LMW-GS composed of 355 amino acids, with a molecular mass of 40,608, and a pI of 8.71.

An amino acid sequence comparison between the proteins encoded by the pLMW1CL and the λLMW3.1 showed a homology of 92.9% with a few substitutions along the sequence and a deletion of 13 amino acids in the repetitive domain (Fig. 3). The deleted fragment comprises the heptapeptide repeat 16 and hexapeptide repeat 17 (Fig. 2). Both sequences showed the same number and position of cysteine residues; noteworthy is the conserved position of the cysteine at the beginning of the repetitive domain (residue 45) (Fig. 3) because it replaces the cysteine residue in the N-terminal region (residue 5) of most LMW-GSs deduced from the nucleotide sequences.

Nucleotide and amino acid sequence comparisons of the pLMW1CL and λLMW3.1 clones with the *lmw-gs* genes published so far (Bartels and Thompson 1983; Okita 1984; Okita et al. 1985; Pitts et al. 1988; Colot et al. 1989; Cassidy and Dvorak 1991; D'Ovidio et al. 1992b; 1997; Cassidy et al. 1998; Volckaert, GenBank X84960 and X84961) showed a high degree of homology (65–85%) along the entire sequence, with a strong sequence conservation in the C-terminal domain and high sequence variability in the repetitive domain. In particular, the reported allelic genes proved to be the largest ones among those characterized so far, and their larger size is almost entirely ascribable to their repetitive domain which was about 150–250 bp larger than that of the other *lmw-gs*.

Discussion

LMW-GSs encoded at the *Glu-B3* locus play a primary role in influencing the technological properties of durum wheat semolina. However, for reasons still unclear, genes located at this locus are recalcitrant to standard cloning procedures, and as a consequence of these difficulties, no *lmw-gs* genes known to be from this locus were characterized until very recently. An initial positive result was obtained with the development of primers specific for the coding region of *lmw-gs* genes located at the *Glu-B3* locus (D'Ovidio 1993; Van Campenhout 1995; D'Ovidio et al. 1997) and with the characterization of the first complete gene located at this locus (D'Ovidio et al. 1997). Moreover, correspondence between a 1B-coded LMW-GS and its encoding gene has been also demonstrated in a bread wheat cultivar (Masci et al., in preparation). On the basis of these positive results, we have continued our analysis and now we report the characterization of two additional *lmw-gs* genes located at this locus.

Both nucleotide sequences contain an open reading frame which encodes LMW-GSs with a very long repetitive domain, much larger than those encoded by the *lmw-gs* genes previously characterized. Their calculated molecular weights correspond to those obtained by SDS-PAGE for the larger glutenin subunits present in the allelic LMW-1 and LMW-2 groups (Masci et al. 1995; Carrillo et al. 1990). PCR analyses on a segregating F_2 population of durum wheat cultivars demonstrated the allelic nature of the reported *lmw-gs* genes, which was also confirmed using primers located in internal regions of the two *lmw-gs* genes (D'Ovidio and Porceddu 1996). This finding makes the results reported here particularly interesting because, apart from representing the first report on the identification of allelic *lmw-gs* genes, they concern the characterization of members encoding LMW-GSs related to the qualitative properties of durum wheat semolina.

One of the structural characteristics of these two *lmw-gs* alleles is the large size of their repetitive domains and the regularity of their repeats. In this regard, it is notable that a correlation between the length of the repeated sequence domain and dough properties has already been reported for the HMW-GS (Anderson et al. 1996). Although the repeats may vary in structure, especially with respect to the number of glutamine residues included, their consensus structure PPFSQQQQ is much more regular than that present in other LMW-GSs. This characteristic might exert a positive influence on gluten quality as measured by dough strength and elasticity.

The comparison of amino acid sequences between the deduced proteins encoded by the pLMW1CL and the λLMW3.1 clones showed the presence of eight cysteine residues and an almost perfect homology along the entire sequence, except in the repetitive domain which contained a 13 amino acid deletion.

Taking into consideration the molecular structure of the proteins encoded by the pLMW1CL and the λLMW3.1 clones, it seems very likely that both act as chain extenders of the growing glutenin polymer (Lew

et al. 1992). In fact, both genes possess eight cysteine codons located at corresponding positions. The first and seventh cysteines should form inter-molecular disulfide bonds, whereas the remaining cysteines should be involved in intra-molecular disulfide bonds (for review see Shewry and Tatham 1997). Moreover, the deletion of two hexapeptide repeats seems in itself insufficient to explain the different effects on quality observed between the two group of proteins.

If the intrinsic structure of the allelic LMW-GSs belonging to the LMW-1 and LMW-2 groups can not explain completely their different contribution to the viscoelastic properties of wheat dough, then the difference in their relative amounts can account for their opposite performance. In this respect, quantitative analyses of LMW-1 and LMW-2 demonstrated that the latter is present in a significantly greater amounts (Autran et al. 1987; Masci et al. 1995). Since transformation protocols are now available for transforming wheat with glutenin genes (Shewry et al. 1995; Blechl and Anderson 1996, Altpeer et al. 1996; Barro et al. 1997), it is possible to test the hypothesis that a higher quantity of specific LMW-GS, e.g. with a larger repetitive domain, could account for the superior performances of durum wheat cultivars.

The characterization of the primary structure of quality-related *lmw-gs* genes, other than providing information for understanding the structure-function relationship among LMW-GSs and the basic material for wheat transformation, may also help in identifying the molecular events underlying their origin and in clarifying the evolutionary relationship between LMW-GSs encoded at different loci both in cultivated and wild wheat relatives. Comparisons between coding regions of *lmw-gs* genes indicate that the repetitive domain is the most variable region and that the repeat motif is the major unit of deletions/duplications, leading to the divergence of repetitive domains composing the different *lmw-gs* genes. The occurrence of such mutations is particularly evident among genes located at the *Glu-B3* locus, whose repetitive domain is very regular and represented by a series of 18 (Volckaert, GeneBank X84960; D'Ovidio et al. 1997) to 25 copies of single units, all of which begin and end with three triplets coding for glutamine. The consensus sequence of the basic motif shows a general strong conservation along the entire repetitive domain with respect to the presence of various nucleotide substitions, some of which determine changes in corresponding amino acids. Noteworthy, due to its possible functional importance, is the occurrence of a TGT codon in the fifth repeat unit. This triplet may have arisen from a TTT triplet by a T → G transversion in the second nucleotide of the third triplet composing the original repeat. This single substitution causes a drastic change at the protein level, where the putative original phenylalanine is then substituted with a cysteine residue capable of forming disulfide bonds.

The sequencing data available do not enable the identification of the basic motif from which the repetitive domains of the present *lmw-gs* genes have evolved. However, a possible hypothesis is that they arose from a basic motif of 18–27 bp which underwent several mutational events such as deletions, duplications and substitutions. The sequence redundancy of the repetitive domain could have promoted deletions and duplications by molecular mechanisms such as unequal crossing-over or slippage during replication. The same molecular mechanisms have been also proposed for the evolution of the repetitive domain of HMW-GSs (Shewry et al. 1992), and a duplication of recent origin within this domain has been identified in a bread wheat genotype (D'Ovidio et al. 1996).

In conclusion, the results reported here represent an additional contribution to the elucidation of the structure and evolution of the *lmw-gs* gene family and to understanding the role that each specific member may contribute to the final gluten product.

Acknowledgements The research was supported in part by a grant for the 'Advanced Biotechnology' National Research Program issued from the Italian 'Ministero della Ricerca Scientifica e Tecnologica' to Enichem Spa, and in part by the 'Ministero delle Risorse Agricole, Alimentari e Forestali', National Research Project 'Plant Biotechnology'. The experiments comply with current Italian laws.

References

Anderson OD, Bekes F, Gras P, Kuhl JC, Tam A (1996) Use of a bacterial expression system to study wheat high-molecular-weight (HMW) glutenins and the construction of synthetic HMW-glutenin genes. In: Wrigley CW (ed) Gluten '96. Royal Australian Chemical Institute, Sydney, Australia, pp 195–198

Altpeer F, Vasil V, Srivastava V, Vasil IK (1996) Integration and expression of the high-molecular-weight glutenin subunit *1Ax1* gene into wheat. Nat Biotechnol 14:1155–1159

Autran JC, Laignelet B, Morel MH (1987) Characterization and quantification of low molecular weight glutenins in durum wheats. Biochimie 69:699–711

Barro F, Rooke L, Bekes F, Gras P, Tatham AS, Fido R, Lazzeri PA, Shewry PR, Barcelo P (1997) Transformation of wheat with high molecular weight subunit genes result in improved functional properties. Nat Biotechnol 15:1295–1299

Bartels D, Thompson RD (1983) The characterization of cDNA clones coding for wheat storage proteins. Nucleic Acids Res 11:2961–2977

Blechl AE, Anderson OD (1996) Expression of a novel high-molecular weight glutenin subunit gene in transgenic wheat. Nat Biotechnol 14:875–879

Carrillo JM, Vazquez JF, Orellana J (1990) Relationship between gluten strength and glutenin proteins in durum wheat cultivars. Plant Breed 104:325–333

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

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

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

D'Ovidio R (1993) Single-seed PCR of LMW glutenin genes to distinguish between durum wheat cultivars with good and poor technological properties Plant Mol Biol 22 : 1173–1176

D'Ovidio R, Porceddu E (1996) PCR-based assay for detecting 1B-genes for low-molecular-weight glutenin subunits related to gluten quality properties in durum wheat. Plant Breed 115 : 413–415

D'Ovidio R, Tanzarella OA, Porceddu E (1992a) Isolation of an alpha-type gliadin gene from *Triticum durum* Desf and genetic polymorphism at the *Gli-2* loci. J Genet Breed 46 : 41–48

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

D'Ovidio R, Lafiandra D, Porceddu E (1996) Identification and molecular characterization of a large insertion within the repetitive domain of a high-molecular-weight glutenin subunit gene from hexaploid wheat. Theor Appl Genet 93 : 1048–1053

D'Ovidio R, Simeone M, Masci S, Porceddu E (1997) Molecular characterization of a LMW-GS gene located on chromosome 1B and the development of primers specific for the *Glu-B3* complex locus in durum wheat. Theor Appl Genet 95 : 1119–1126

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

Masci S, Lew EJL, Lafiandra D, Porceddu E, Kasarda DD (1995) Characterization of low-molecular-weight glutenins Type 1 and Type 2 by RP-HPLC and N-Terminal sequencing. Cereal Chem 72 : 100–104

Okita TW (1984) Identification and DNA sequence analysis of a gamma-type gliadin cDNA plasmid from winter wheat. Plant Mol Biol 3 : 325–332

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of α/β type and γ-type gliadin DNA sequences. J Biol Chem 260 : 8203–8213

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

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

Pogna NE, Autran JC, Mellini F, Lafiandra D, Feillet P (1990) Chromosome 1B-encoded gliadins and glutenin subunits in durum wheat: genetics and relationship to gluten strength. J Cereal Sci 11 : 15–34

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York

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

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

Shewry PR, Thatam AS, Barro F, Barcelo P, Lazzeri P (1995) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Bio/Technology 13 : 1185–1190

Van Campenhout S, Vander Stappen J, Sagi L, Volckaert G (1995) Locus-specific primers for LMW glutenin genes on each of the group 1 chromosomes of hexaploid wheat. Theor Appl Genet 91 : 313–319

Wrigley CW (1996) Giant proteins with flour power. Nature 381 : 738–739