Theor Appl Genet (1996) 93: 1048–1053

R. D'Ovidio · D. Lafiandra · E. Porceddu

Identification and molecular characterization of a large insertion within the repetitive domain of a high-molecular-weight glutenin subunit gene from hexaploid wheat

Received: 1 February 1996 / Accepted: 10 May 1996

Abstract High-molecular-weight (HMW) glutenin subunits are a particular class of wheat endosperm proteins containing a large repetitive domain flanked by two short N- and C-terminal non-repetitive regions. Deletions and insertions within the central repetitive domain has been suggested to be mainly responsible for the length variations observed for this class of proteins. Nucleotide sequence comparison of a number of HMW glutenin genes allowed the identification of small insertions or deletions within the repetitive domain. However, only indirect evidence has been produced which suggests the occurrence of substantial insertions or deletions within this region when a large variation in molecular size is present between different HMW glutenin subunits. This paper represents the first report on the molecular characterization of an unusually large insertion within the repetitive domain of a functional HMW glutenin gene. This gene is located at the *Glu-D1* locus of a hexaploid wheat genotype and contains an insertion of 561 base pairs that codes for 187 amino acids corresponding to the repetitive domain of a HMW glutenin subunit encoded at the same locus. The precise location of the insertion has been identified and the molecular processes underlying such mutational events are discussed.

Key words *Triticum aestivum* · HMW glutenin genes · Unequal crossing-over · PCR · *Glu-D1* locus

Introduction

High-molecular-weight glutenin subunits (HMW-GS) are wheat endosperm proteins coded by genes located at the *Glu-1* loci on the long arms of the group-1 chromosomes of hexaploid wheat (for review see Shewry et al. 1992). Their structures include large repetitive domains flanked by short N- and C-terminal non-repetitive domains. The repeating structure of the central domain is composed of tandem and interspersed repeats of 6-, 9- and 15-peptides motifs, all rich in proline and glutamine (Shewry et al. 1989; Kasarda et al. 1994). Secondary structural or molecular modelling analyses of this region suggest the occurrence of a β-spiral conformation (Tatham et al. 1984) or a spiral based on inverse γ turns (Kasarda et al. 1994), respectively. It has also been demonstrated that the repeating structure of the central domain contributes to the anomalous slow migration of HMW-GS, as compared to globular proteins, when analyzed by SDS-PAGE (D'Ovidio et al. 1996).

HMW-GS are under continuing study, both at the biochemical and the molecular level, because of their primary role in determining the viscoelastic properties of wheat flour (Payne 1987; Shewry et al. 1992). These studies have shown the occurrence of allelic variation at each *Glu-1* locus (Payne et al. 1981; Payne and Lawrence 1983) and have indicated that length polymorphism is mainly due to size variation of the repetitive domain (D'Ovidio et al. 1994, 1995). Molecular analysis has also shown that two different genes, identified as x and y, are present at each complex locus (Harberd et al. 1986).

Nucleotide-sequence comparisons of these genes have demonstrated the occurrence of small insertions and deletions within their repetitive domain, and unequal crossing-over or slippage during replication have been suggested as possible explanations for such mutational events (Shewry et al. 1989).

The detection, at the *Glu-D1* locus of the hexaploid wheat accession MG 315, of a 1Dx2.2* gene (Margiotta et al. 1993) coding for an HMW-GS having biochemical, molecular and conformational characteristics similar to those of the polypeptide encoded by the 1Dx2 gene (Margiotta et al. 1993; Buonocore et al. 1996), but with a larger molecular size of the repeating domain (D'Ovidio et al. 1994), provided suitable material for studying the molecular ba-

Communicated by G. E. Hart

R. D'Ovidio (✉) · D. Lafiandra · E. Porceddu
Dipartimento di Agrobiologia e Agrochimica,
Università della Tuscia, Via S. Camillo de Lellis,
01100 Viterbo, Italy

sis underlying the origin of the large length variation of the repetitive domain in HMW-GS. The results of this analysis are reported in the present paper.

Materials and methods

Materials

The analysis was carried out on the common wheat cultivars Cheyenne and Chinese Spring, and the hexaploid wheat accession MG 315, which were provided by the Germplasm Institute, National Research Council, Bari, Italy.

SDS-PAGE analysis

Total proteins were extracted from flour using a buffer containing 0.125 M Tris-HCl pH 6.8, 2.75% (w/v) sodium dodecyl sulphate (SDS), 10% (v/v) glycerol, 1% (w/v) dithiothreitol (DTT) in a 1:5 ratio (mg/μl). Protein extracts (5 μl) were analyzed on a Protean II Cell (Bio-Rad, Richmond, Calif., U.S.A.). Electrophoresis was carried out by SDS-PAGE (T=10%, C=1.28%) and the running buffer contained 0.2 M glycine, 0.05 M Tris pH 8.3, 0.1% SDS. Gels were stained overnight with 12%(w/v) trichloroacetic acid solution containing 0.05% Comassie Brilliant Blue R-250 in absolute ethanol (5% w/v) and de-stained in tap water.

Molecular weights were determined by using globular proteins as molecular-weight standards (myosin, β-galactosidase, phosphorylase B, serum albumin and ovalbumin).

DNA extraction

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

PCR amplification and restriction digestions

PCR reactions were performed in a final reaction volume of 100 μl using 100 ng of genomic DNA, 2.5 units of *Taq* DNA polymerase (Boehringer), 1 × *Taq* PCR buffer (Boehringer), 250 ng each of the two primers and 200 μM of each deoxyribonucleotide. Amplification conditions were those reported in D'Ovidio et al. (1994) using primers based on the nucleotide sequence of the 1Dx5 gene of cultivar Cheyenne (Anderson et al. 1989) and having the following sequences: (1) 5′ CCGAGATGGCTAAGCGGTTA 3′; (2) 5′ CTGGCTGGCCGACAATGCGT 3′.

Aliquots (10 μl) of the amplification products were digested with *Hin*dIII, *Nco*I and *Pst*I restriction enzymes using standard conditions (Sambrook et al. 1989). Undigested and digested PCR products were fractionated on a 1.5% agarose gel in 1 × TBE buffer following standard procedures (Sambrook et al. 1989).

Cloning and nucleotide sequence of the *Nco*I insertion

The *Nco*I fragment from the amplification product of MG315 was cloned into the *Nco*I site of de-phosphorylated pGEM 5zf(+) plasmid vector (Promega) and subjected to nucleotide sequence analysis using the chain-terminator method (Sanger et al. 1977). De-phosphorylation of the pGEM vector was carried out by using standard procedures (Sambrook et al. 1989). The PC/GENE computer program (IntelliGenetics, Inc., USA) was used to analyze the sequence data.

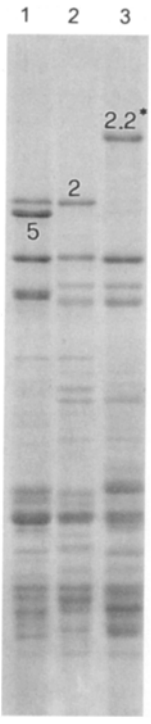

Fig. 1 SDS-PAGE of glutenin subunits from Cheyenne (*1*), Chinese Spring (*2*) and MG315 (*3*). *Numbers* on the gel designate allelic HMW-GS corresponding to the *Glu-D1* locus

Results

SDS-PAGE of endosperm proteins from the hexaploid wheat genotypes Cheyenne, Chinese Spring and MG 315 showed the presence of a very slow migrating Dx HMW-GS in MG 315 (Fig. 1). SDS-PAGE analysis, carried out in the presence of globular proteins as molecular-weight standards, made it possible to estimate that the size of the Dx2.2* subunit of MG 315 is about 25000 daltons larger than the allelic subunits 5 and 2 of Cheyenne and Chinese Spring, respectively (Fig. 1 and Table 1).

The gene coding for subunit Dx2.2* from MG 315 was amplified by PCR using oligonucleotide primers amplifying for the complete coding region of genes encoding Dx glutenin subunits (D'Ovidio et al. 1994), and the amplification product compared with those obtained from allelic genes of Cheyenne and Chinese Spring. PCR results were in agreement with those obtained on the proteins in that the 1Dx gene from MG 315 was about 560 bp larger than those from Cheyenne and Chinese Spring.

Table 1 Molecular weight (Mr) of the Dx HMW-GS of the analyzed genotypes as determined from SDS-PAGE and deduced amino-acid sequences. HMW-GS designation is according to Payne and Lawrence (1983) and Margiotta et al. (1993)

Genotype	Dx HMW-GS	Mr on SDS-PAGE	Mr from deduced amino-acid sequence
Cheyenne	Dx5	106.500	88.128
Chinese Spring	Dx2	109.500	87.009
MG315	Dx2.2*	133.500	106.826 (87.009 + 19.817)

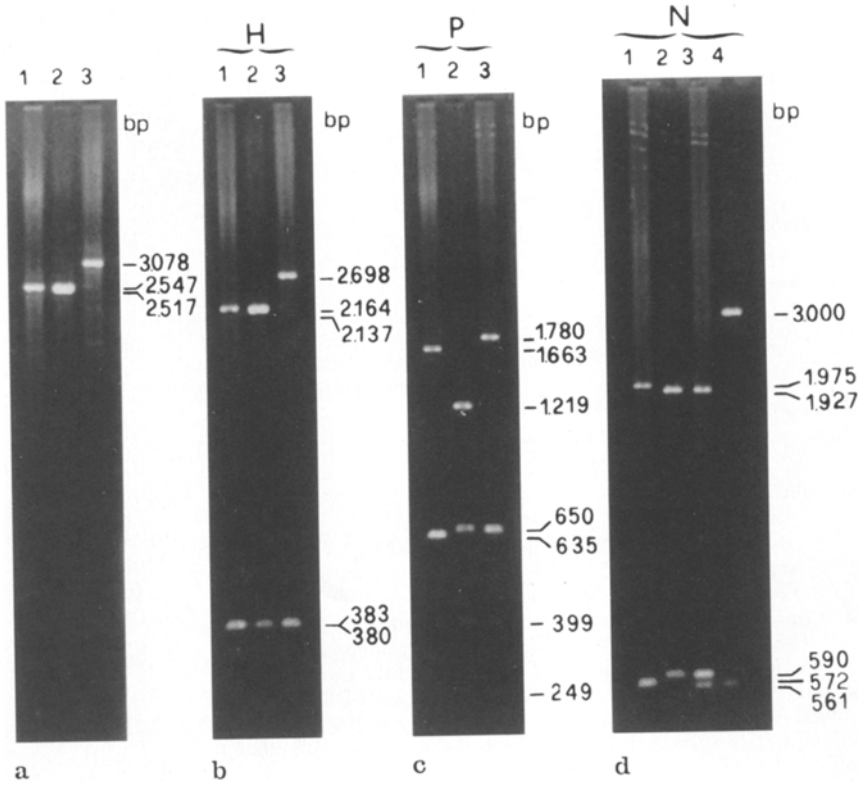

Fig. 2a–d 1.5% agarose gel of the amplification product corresponding to the 1Dx gene from Cheyenne (*1*), Chinese Spring (*2*) and MG315 (*3*) undigested (**a**) and subjected to restriction digestion with *Hin*dIII (**b**) *Pst*I (**c**) and *Nco*I (**d**). In panel **d** the *Nco*I fragment from the 1Dx2.2* gene cloned into the pGEM 5zf(+) vector (*4*) is also reported

Amplification products obtained from the three genotypes were subjected to restriction-digestion analysis in order to investigate the cause of the larger size of the Dx glutenin subunit from MG 315. Analyses carried out by using three different restriction enzymes, *Hin*dIII, *Pst*I and *Nco*I, whose positions are known for the 1Dx2 (Sugiyama et al. 1985) and 1Dx5 (Anderson et al. 1989) genes, showed different restriction patterns for the three genotypes (Fig. 2 and Table 2). However, the restriction patterns obtained from Chinese Spring and MG 315 showed some common fragments; in particular, *Nco*I digestion produced the same pattern in both genotypes except for the presence of an additional band of about 561 bp in MG 315 (Fig. 2 d, Table 2). This result made it possible to infer that the larger size of the complete 1Dx2.2* gene, as compared to 1Dx2, should be entirely due to the additional *Nco*I fragment.

Table 2 Molecular weights of undigested and digested PCR products corresponding to the Dx HMW-GS of the analyzed genotypes. The exact size of the PCR fragments were calculated on the basis of the nucleotide sequences of the 1Dx2 (Sugiyama et al. 1985) and 1Dx5 (Anderson et al. 1989) genes. The uncut PCR products were digested with *Hin*dIII, *Pst*I and *Nco*I restriction enzymes. The fragment length is reported in base pairs. The *Nco*I insertion is reported in *bold*

Genotype HMW-GS gene	Cheyenne 1Dx5	Chinese Spring 1Dx2	MG315 1Dx2.2*
Undigested	2547	2517	3078 (2517 + 561)
*Hin*dIII	2164	2137	2698
	383	380	380
*Pst*I	1663	1219	1780
	635	650	650
	249	399	399
		249	249
*Nco*I	1975	1927	1927
	572	590	590
			561

The cloning and sequencing of the *Nco*I fragment showed that its length is 561 bp and that it encodes 187 amino acids, equivalent to a molecular weight of 19.817 (Fig. 3). The calculated isoelectric point of this polypeptide is 7.77 and the hydropathicity value (GRAVY) is –18.42. Sequence comparison between the *Nco*I fragment from 1Dx2.2* and other HMW glutenin genes demonstrated the existence of an almost perfect similarity between the *Nco*I fragment and the central repetitive domain of the 1Dx2 gene (Sugiyama et al. 1985) (Fig. 4 a). It is worth noting that an almost perfect similarity between the 1Dx2.2* and 1Dx2 genes was also observed for the first 400 bp of their coding regions (data not shown).

On the basis of these results and those obtained from restriction-pattern analyses of PCR products from Chinese Spring and MG 315, which proved to be identical except for the presence of the additional band of 561 bp, it is possible to conclude that the *Nco*I fragment represents an insertion within the repetitive domain of a HMW gene related to the 1Dx2 gene.

By taking into consideration the sequence similarity between the 1Dx2.2* and 1Dx2 genes, and the distribution of the sequence identity between the *Nco*I fragment and

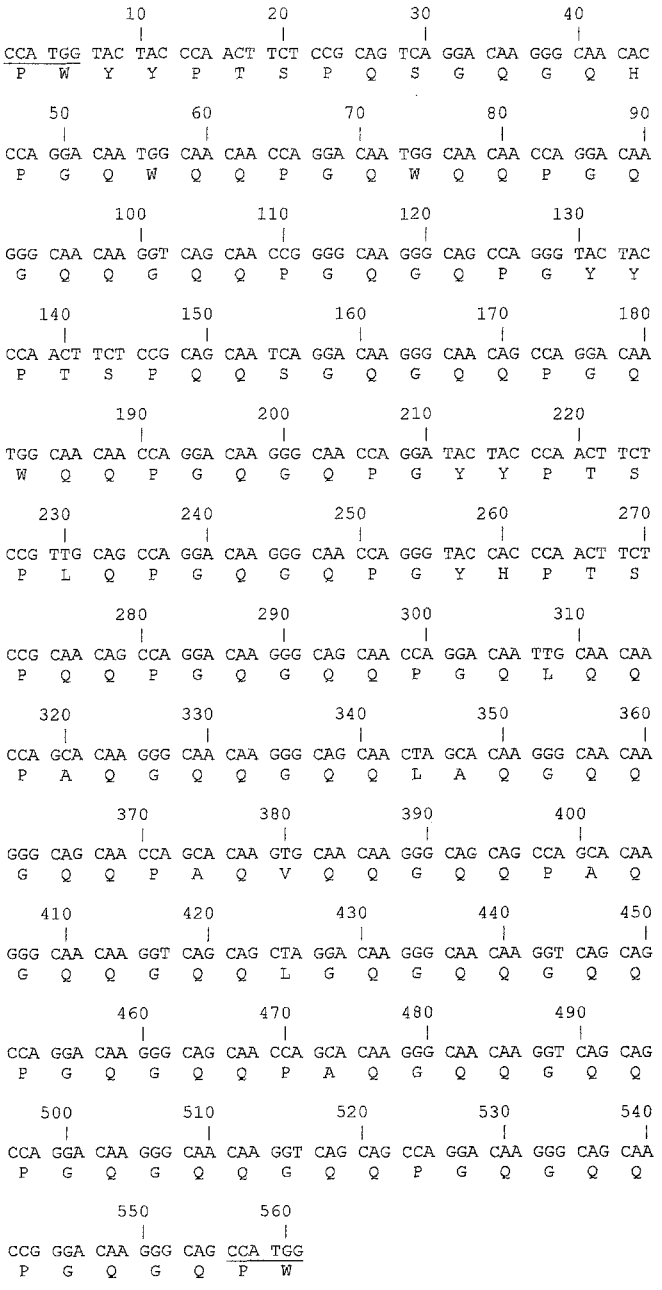

Fig. 3 Nucleotide and deduced amino-acid sequence of the *Nco*I fragment (clone pTUHMW2.2*) from the 1Dx2.2* gene. *Nco*I sites are *underlined*

1Dx2, it is possible to suggest a putative site of insertion which gave rise to 1Dx2.2* from a HMW gene related to 1Dx2 (Fig. 4). The insertion should have occurred within the region 2011–2025 bp and should have involved the repeating domain from 1453–1467 bp up to 2011–2025 bp. This inference was based on the observed identity between the first 96 bp of the *Nco*I fragment and region 1927–2025 of the 1Dx2 gene, and the rest of the *Nco*I fragment (97–561 bp) with a different region of the repeating domain of the 1Dx2 gene, which is composed of between 1453–1467 bp and 1926 bp (Fig. 4). Since the region between 82–96 bp of the *Nco*I fragment shows a perfect identity with regions 1453–1467 bp and 2011–2025 bp of the 1Dx2 gene (Fig. 4), it was not possible to establish the exact base position of the insertion involved.

Finally, on the basis of similarity between the 1Dx2.2* and 1Dx2 genes and of their encoded proteins, the molecular weight of the former was deduced by adding the molecular weight of the insertion to that of Dx2 (Tables 1 and 2).

Discussion

Several allelic HMW-GS encoded by genes at the *Glu-D1* locus have been described (Payne and Lawrence 1983) and, for some of them, the nucleotide sequences of corresponding genes has been reported. In particular, the complete nucleotide sequences of the genes coding for Dx5 and Dx2 HMW-GS have been determined (Anderson et al. 1989; Sugiyama et al. 1985). Sequence comparison between these two genes and their deduced amino-acid sequences showed a general high degree of homology along the sequence and the presence of short insertions and deletions within the repetitive domain.

The results of the reported analyses demonstrate that the large size of the 1Dx2.2* gene is due to a single duplication of 561 bp within the repetitive domain. On the basis of restriction analyses and sequence comparison it is also possible to suggest that this gene most probably derives from a HMW-GS gene related to 1Dx2.

Sequence analysis of the 561-bp fragment showed an almost perfect identity to part of the repetitive domain of the 1Dx2 gene, making it possible to indicate the putative site of the mutational event which caused the duplication at between 2011 bp and 2025 bp.

Two nucleotide differences and a deleted triplet found between the 561-bp fragment and the corresponding region of the 1Dx2 gene, caused two amino-acid changes and a deletion which, however, did not greatly affect the hydropathicity value of the deduced amino-acid sequence of the 561-bp fragment. In fact, this value (GRAVY=−18.42) proved to be very similar to that obtained for the entire repetitive domain of the 1Dx2 gene (GRAVY=−17.33). This result explains the similarity of the surface hydrophobicity values found between Dx2.2* and the allelic HMW-GS (Margiotta et al. 1993).

Several mechanisms, such as replication slippage, transposition, gene conversion, or unequal crossing over, can account for the gene-length modification found in the 1Dx2.2* gene. However, the presence of an almost perfect nucleotide-sequence duplication and the large size of the duplicated fragment, seems to suggest unequal crossing over as the most probable event responsible for generating the 1Dx2.2* gene (Fig. 5).

The very large size and regularity of motifs occurring within the repetitive domain of HMW-GS genes may greatly facilitate the occurrence of sequence mispairing, either between two homologous chromosomes at meiosis

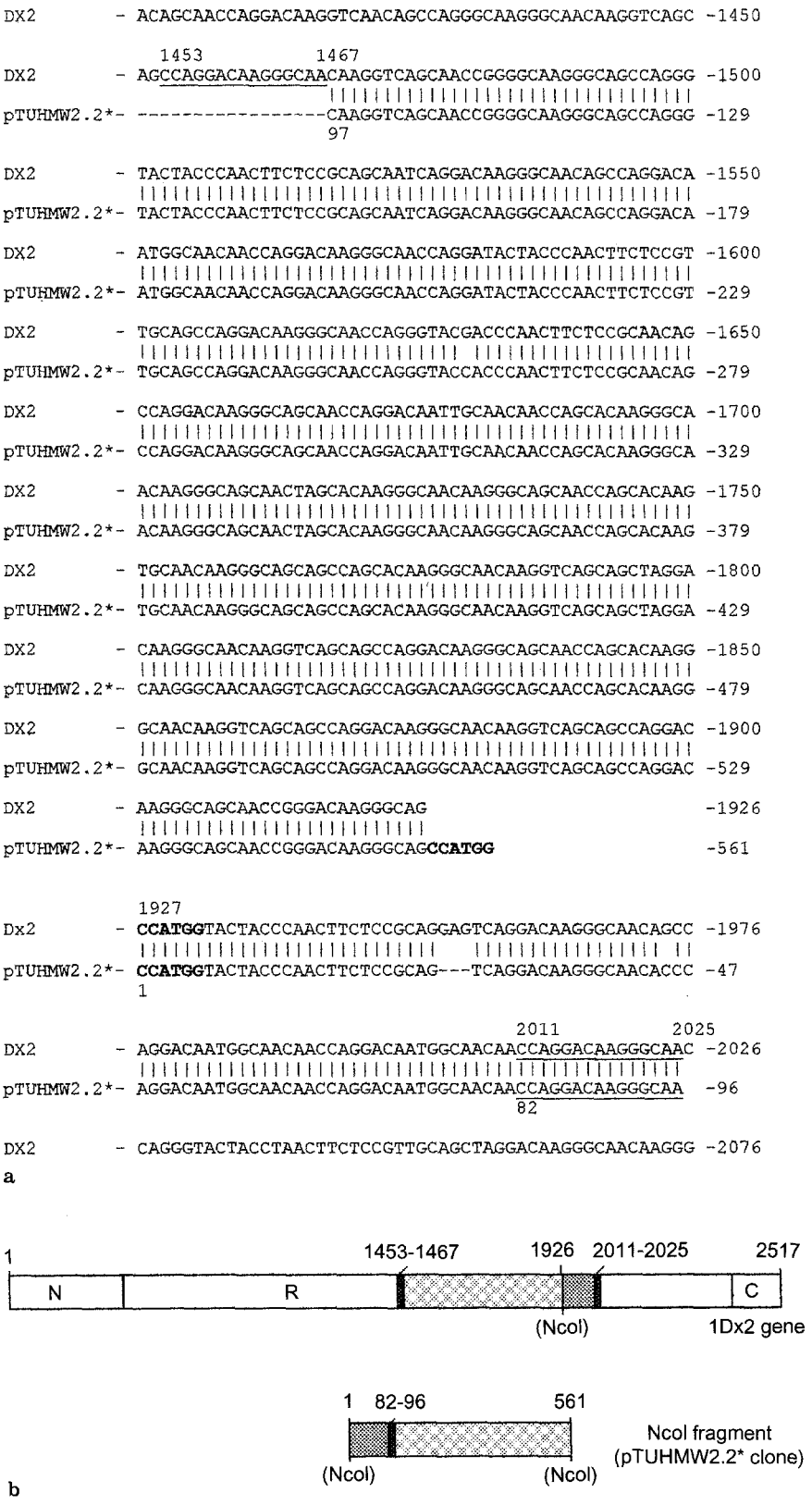

Fig. 4a, b Nucleotide comparison between the *Nco*I fragment (pTUHMW2.2* clone) and the 1Dx2 gene. Nucleotide identity is indicated by *vertical lines* and *Nco*I sites are in *bold*. Nucleotide sequences involved in unequal crossing over are *underlined* (**a**). Diagram showing the homology between the *Nco*I fragment from the 1Dx2.2* gene (pTUHMW2.2* clone) and a segment of the repetitive domain of the 1Dx2 gene. A similar filling-in pattern indicates corresponding sequence homology between the two sequences (**b**)

or between two sister chromatids of a chromosome during the mitosis of a germ-line cell. In both cases a reciprocal recombination process creates a sequence duplication in one chromatid and a corresponding deletion in the other.

The mutational event responsible for the large size of the 1Dx2.2* gene could have occurred in a diploid *Aegilops* species and then been transferred by rare outcrossing into common wheat, or else it could have arisen from a recent gene mutation within common wheat. This last hy-

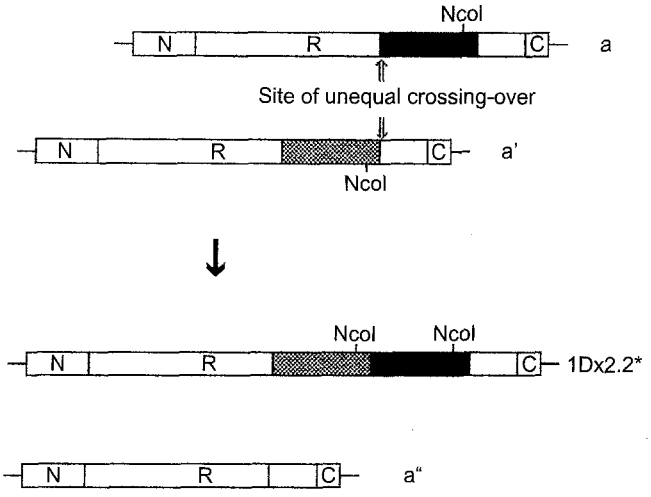

Fig. 5 Scheme of a possible unequal crossing over which could have given rise to the large repetitive domain of the 1Dx2.2* gene. *a*, *a'* represent two hypothetical HMW-GS genes related to 1Dx2. *a"* is a deleted gene resulting from unequal crossing over between *a* and *a'*

pothesis, previously suggested by Payne et al. (1983) for the HMW-GS Dx2.2, seems also to be supported by the absence of large Dx subunits, such as the Dx2.2*, in *Aegilops squarrosa* (Lagudah and Halloran 1988).

In conclusion, the duplication present in 1Dx2.2* was probably generated by an unequal crossing over event in hexaploid wheat. This result represents the first clear molecular evidence that insertions of blocks of repeated sequences play an important role in generating new allelic HMW-GS variants and, more generally, as previously suggested (Shewry et al. 1989; Shewry 1995), in the evolution of prolamins. In fact, the origin of HMW-GS proposed by these authors is based on the repeating structure of the central repetitive domain of these proteins, whereas the present data demonstrate the occurrence of a duplication of a large segment on the basis of a comparative analysis with an existing HMW-GS from which the new gene is probably derived.

Acknowledgements Research supported by the National Research Council of Italy, Special Project RAISA, Subproject 2; paper no. 2791.

References

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

Buonocore F, Hickman DR, Caporale C, Porceddu E, Lafiandra D, Tatham AS, Shewry (1996) Characterization of a novel high Mr glutenin subunit encoded by chromosome 1D of bread wheat. J Cereal Sci 23:55–60

D'Ovidio R, Tanzarella OA, Porceddu E (1992) 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, Porceddu E, Lafiandra D (1994) PCR analysis of genes encoding allelic variants of high-molecular-weight glutenin subunits at the *Glu-D1* locus. Theor Appl Genet 88:175–180

D'Ovidio R, Anderson OD, Masci S, Skerritt J, Porceddu E (1996) Construction of Dx5 genes modified in the repetitive domain and their expression in *Escherichia coli*. In: Schofield JD (ed) Wheat structure, biochemistry and functionality. The Royal Society of Chemistry, UK (in press)

D'Ovidio R, Masci S, Porceddu E (1995) Development of a set of oligonucleotide primers specific for genes at the *Glu-1* locus. Theor Appl Genet 91:189–194

Harberd NP, Bartels D, Thompson RD (1986) DNA restriction-fragment variation in the gene family encoding high-molecular-weight (HMW) glutenin subunits of wheat. Biochem Genet 24:579–595

Kasarda DD, King G, Kumosinski F (1994) Comparison of spiral structures in wheat high-molecular-weight glutenin subunits and elastin by molecular modeling. In: Kumosinski F, Liebman MN (eds) Molecular modeling: from virtual tools to real problems. American Chemical Society, pp 209–220

Lagudah ES, Halloran GM (1988) Phylogenetic relationships of *Triticum tauschii*, the D genome donor to hexaploid wheat. Theor Appl Genet 75:592–598

Margiotta B, Colaprico G, D'Ovidio R, Lafiandra D (1993) Characterization of high M_r subunits of glutenin by combined chromatographic (RP-HPLC) and electrophoresis separations and restriction fragment length polymorphism analyses of their encoding genes. J Cereal Sci 17:221–236

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

Payne PI, Lawrence GD (1983) Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1* and *Glu-D1*, which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11:29–3

Payne PI, Holt LM, Law CN (1981) Structural and genetic studies of the high-molecular-weight subunits amongst varieties of wheat (*Triticum aestivum* L.). Theor Appl Genet 60:229–241

Payne PI, Holt LM, Lawrence GD (1983) Detection of a novel high-molecular-weight subunit of glutenin in some Japanese hexaploid wheats. J Cereal Sci 1:3–8

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

Shewry PR (1995) Plant storage proteins. Biol Rev 70:375–426

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

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

Sugiyama T, Rafalski A, Peterson D, Soll D (1985) A wheat HMW glutenin subunit gene reveals a highly repeated structure. Nucleic Acids Res 13:8729–8737

Tatham AS, Shewry PS, Miflin BJ (1984) Wheat gluten elasticity: a similar molecular basis to elastin? FEBS Lett 177:205–208