

Yongfang Wan · Zehong Yan · Kunfan Liu
Youliang Zheng · Renato D'Ovidio · Peter R. Shewry
Nigel G. Halford · Daowen Wang

Comparative analysis of the D genome-encoded high-molecular weight subunits of glutenin

Received: 5 April 2005 / Accepted: 12 July 2005 / Published online: 22 September 2005
© Springer-Verlag 2005

Abstract Four genes encoding novel 1Dx-type high-molecular weight (HMW) subunits were amplified by polymerase chain reaction, two each from *Aegilops tauschii* and bread wheat *Triticum aestivum*. The two subunits from *Ae. tauschii* (1Dx2.1¹ and 1Dx2¹) were both very similar in sequence to subunit 1Dx2 from bread wheat. In contrast, the two novel bread wheat subunits (1Dx2.2 and 1Dx2.2*) differed from subunit 1Dx2 in having different internally duplicated regions (of 132 and 186 amino acid, respectively) within their repetitive domains. These duplicated sequences were located adjacent to the regions from which they had been duplicated and had complete intact repeat motifs at each end. The implications of these results for HMW subunit evolution and wheat quality improvement are discussed.

Introduction

The high-molecular weight (HMW) subunits of glutenin are a group of storage proteins present in the seeds of wheat and related species of the Triticeae, including diploid and polyploid species of *Aegilops* (Payne 1987;

Shewry et al. 2003b). They are encoded by the *Glu-1* loci located on the long arms of the homoeologous group one chromosomes, with each locus comprising two genes encoding x- and y-type subunits. Consequently, three loci encoding six HMW subunits are present in hexaploid bread wheat (*Triticum aestivum*, $2n=6x=42$, AABBDD) which originated some 10,000 years ago from natural hybridisation between tetraploid wheat ($2n=4x=28$, AABB) and diploid *Ae. tauschii* ($2n=2x=14$, DD) (Feldman et al. 1995; Feldman et al. 2000). However, silencing of one or more of these genes results in the presence of only three to five individual HMW subunits in cultivars of bread wheat (Payne 1987).

The HMW subunits of wheat have been shown to play a crucial role in determining the processing properties of the grain (Shewry et al. 2003a, b). This has led to the isolation of a number of HMW subunit genes, providing complete amino acid sequences of a range of allelic and homeoallelic proteins from both cultivated hexaploid and tetraploid wheats and wild relatives. These studies have shown that the mature HMW subunit proteins have highly conserved structures, comprising short non-repetitive N-terminal domains (of 86–89 residues in x-type subunits and 104 residues in y-type subunits) and C-terminal domains (42 residues in all subunits) which flank a longer repetitive domain. The latter varies in length from about 630–830 residues in the subunits present most commonly in bread wheat accounting for the differences in mass of the whole subunits (Shewry et al. 2003b).

The N- and C-terminal domains of the subunits appear to have globular structures and contain cysteine residues which form inter- and intra-chain disulphide bonds, the former stabilising the glutenin polymers. In contrast, the repetitive domains of the subunits comprise highly repeated sequences, based on a combination of two or three short peptide motifs. Thus, the x-type subunits comprise hexapeptide motifs (consensus ProGlyGlnGlyGlnGln) which occur either in tandem arrays or adjacent to tripeptides (GlyGlnGln) or nonapeptides

Communicated by P. Langridge

Y. Wan · Z. Yan · K. Liu · D. Wang
Institute of Genetics and Developmental Biology,
Chinese Academy of Sciences, Beijing 100101, China

P. R. Shewry (✉) · N. G. Halford · Y. Wan
Rothamsted Research, Harpenden Hertfordshire, AL5 2JQ, UK
E-mail: peter.shewry@bbsrc.ac.uk
Tel.: +44-1582-763133
Fax: +44-1582-763010

R. D'Ovidio
University of Tuscia, Via SC De Lellis, 01100, Viterbo Italy

Y. Zheng · Z. Yan
Triticeae Research Institute, Sichuan Agricultural University,
Dujiangyan 611830, China

(GlyTyrTyrProThrSerProGlnGln), while the y-type subunits do not contain tripeptides and the interspersed nonapeptide motifs may differ in consensus sequence (either GlyTyrTyrProThrSerLeuGlnGln or GlyTyrTyrProThrSerProGlnGln). However, these motifs are not completely conserved but differ in substitutions and, to a limited extent, in the deletion or insertion of residues, both within and between subunits (Shewry et al. 2003b).

The D genome of bread wheat differs from the A and B genomes in that both the x- and y-type genes are usually expressed (Payne 1987). In addition, the proteins encoded by this genome vary considerably in their molecular masses based on their mobilities on SDS-PAGE, with some x-type subunits present in exotic germplasm (e.g. subunits 1Dx2.2, 1Dx2.2*) having much slower mobilities than those which commonly occur in commercial cultivars grown in most of the world (e.g. 1Dx2, 1Dx5). Subunits with similar slow mobilities have also been reported in *Aegilops* species containing the CC and DD genomes, but not in *Ae. tauschii* (the progenitor of the D genome of bread wheat), despite many accessions having been surveyed (Ladugah and Halloran 1988; William et al. 1993; Mackie et al. 1996; Wan et al. 2000).

The HMW subunits are unusual, perhaps unique, among plant proteins in exhibiting elastomeric properties. In contrast, elastomeric proteins occur more widely in animal systems, in mammalian muscle and connective tissue (collagen, elastin, spectrin, titin, fibrillin), insect tendons (resilin), the holdfasts and hinges of bivalve molluscs (abductin, byssus) and spider silks (Shewry et al. 2003c). Furthermore, most of these proteins resemble the HMW subunits by having repeated sequences (Tatham and Shewry 2003), particularly spider silks in which repetitive domains are interspersed by alanine-rich regions or non-repetitive spacer regions (Lewis 2003). Similarly, the fibroin protein of silk worm cocoons exists in allelic variants which differ in the size of their repetitive domains in a similar manner to the HMW subunits (Gage and Manning 1980; Manning and Gage 1980). It has been suggested that the genes encoding allelic variants of fibroin have evolved by partial duplication due to unequal crossing over between sister chromatids (Manning and Gage 1980), and due to misalignment of the repetitive DNA sequences during meiosis (Lewin 1997). A similar mechanism has also been invoked for the origin of the long 1Dx subunits 1Dx2.2 and 1Dx2.2* (Payne et al. 1983; D'Ovidio et al. 1994, 1996).

The D genome-encoded x-type subunits provide an attractive system to study the evolutionary mechanisms involved in HMW subunit evolution and the evolution of highly repetitive protein domains in general, because a wide range of forms exist in wheat and related wild species, including forms with high-molecular masses. We have therefore isolated and determined the sequences of four novel *Glu-D1x* genes, encoding two variant alleles from *Ae. tauschii* (1Dx2.1^t, 1Dx2^t) and two alleles with high-molecular masses from bread wheat (1Dx2.2,

1Dx2.2*) (Nakamura et al. 1990; Margiotta et al. 1993). Comparison of the amino acid sequences of these and other previously characterised 1Dx HMW subunits cast new light on the mechanisms of HMW subunit evolution.

Materials and methods

Plant materials and SDS-PAGE of seed proteins

Seeds of bread wheat lines MG315 (1Bx7+1By8, 1Dx2.2*+1Dy12) and MG7249 (1Ax2*, 1Bx7+1By8, 1Dx2.2+1Dy2) were obtained from the Germplasm Institute, National Research Council, Bari, Italy. Seeds of *Aegilops tauschii* accessions RM0198 (1Dx2^t+1Dy12) and AS2388 (1Dx2.1^t+1Dy12) were obtained, respectively, from the Institute of Crop Germplasm Resources of the Chinese Academy of Agricultural Sciences, Beijing, China and the Triticeae Research Institute, Sichuan Agricultural University, Dujiangyan, China. Bread wheat cv Avalon (1Ax1, 1Bx6+1By8, 1Dx2+1Dy12) was used as standard for comparison of electrophoretic mobility. For naming HMW glutenin subunits encoded by the *Glu-D1* locus, the superscript "t" was used to indicate that the subunit is derived from *Ae. tauschii*.

Total seed proteins were extracted in sample buffer containing 0.0625 M Tris-HCl (pH 6.8), 2% (w/v) SDS, 1.5% (w/v) DTT, 10% (v/v) glycerol and 0.0002% bromophenol blue. The extracts were heated for 5 min at 95°C and centrifuged for 10 min. The supernatant was loaded onto a 10% (w/v) SDS-PAGE gel using a Tris-borate buffer system as described by Shewry et al. (1995).

Cloning and sequencing of HMW subunit genes

The complete coding region sequences of the genes encoding subunits 1Dx2.2, 1Dx2^t and 1Dx2.1^t were amplified from genomic DNA. The PCR primers were designed according to the sequences of subunits 1Dx2 and 1Dx5 (Sugiyama et al. 1985; Anderson et al. 1989) as follows: (1) 5'-ATGGCTAAGCGGTTAGTCCT-3'. (2) 5'-CTGGCTGGCCGACAATGCGT-3'. A 50 µl reaction mix was used and the reaction performed at 94°C for 5 min, followed by 40 cycles at 94°C for 1 min, 68°C for 3 min then incubated at 68°C for 7 min. The PCR products were cloned into pGEM-T EASY Vector (Promega). The nucleotide sequences of the three genes 1Dx2.2, 1Dx2^t and 1Dx2.1^t were obtained by producing sets of overlapping subclones by nested deletion using exonuclease III and nuclease S1 (Halford et al. 1987). Nucleotide sequencing was performed by Oxford University Biochemistry Department sequencing service (Oxford, UK) or by TaKaRa (DaLian City, China).

The 1Dx2.2* gene was cloned into the pCR-XL-TOPO vector described by D'Ovidio et al. (1996). A

fragment of 1.9 kb was excised from the 1Dx2.2* gene by digestion with *Eco*R1 and *Nco*I and subcloned into a PUCBM20 vector. The nucleotide sequence was obtained by producing sets of overlapping subclones by nested deletions as above.

Expression of HMW subunit genes in *E. coli*

The three genes encoding subunits 1Dx2.2, 1Dx2^t and 1Dx2.1^t were amplified to remove the sequences encoding the signal peptides by PCR using the following primers:

(1) 5'-CTCTCATATGGAAGGTGAGGCCTCTGAG-GCCTCTGAGCAA-3' (*Nde*I restriction site added, underlined sequences).

(2) 5'-CTCTCTCGAGTTACTGGCTGGCCGACAA-TGCGTC-3' (*Xho*I restriction site added, underlined sequences).

The PCR products were cloned into the bacterial expression vector pET-30a (Novagen) and transformed into *E. coli* strain BL21(DE3)pLsS. HMW subunits were expressed in *E. coli* by induction with isopropyl β -D-thiogalactopyranoside (IPTG) (1 mM) for 3 h. The expressed proteins were extracted from 1 ml of bacterial cells for SDS-PAGE.

Comparison analysis of the HMW glutenin subunits

The protein sequences of subunits 1Dx2.2* (accession number AJ893508), 1Dx2.2 (AY159367), 1Dx2^t (AF480485), 1Dx2.1^t (AF480486), 1Dx2 (X03346), 1Dx5 (X12928), 1Dx2.1 (AY517724) and 1Dx1.5^t (AY594355) were aligned by eye to demonstrate the repeat structure of the central domains.

Results and discussion

Molecular cloning of four novel HMW subunit genes

Four 1Dx HMW subunit genes were selected for cloning and sequence analysis based on the mobilities of the encoded proteins. Subunits 1Dx2.2 and 1Dx2.2* have the slowest mobilities of all reported HMW subunits (Margiotta et al. 1993; Nakamura et al. 1990) and genes encoding these proteins were amplified from genomic DNA from bread wheat lines MG7249 and MG315, respectively (Fig. 1). A preliminary survey of accessions of *Ae. tauschii* (DD) identified lines containing subunits 1Dx2.1^t (AS2388) and 1Dx2^t (RM0198), which have mobilities similar but not identical to that of subunit 1Dx2 of bread wheat (Fig. 1) (Wan et al. 2000).

Non-degenerate oligonucleotides with identical sequences to those at the 5' and 3' ends of the coding regions of previously characterised genes encoding HMW subunits 1Dx2 (Sugiyama et al. 1985; EMBL accession number X03346) and 1Dx5 (Anderson et al. 1989; X12928) were used to prime PCR reactions to amplify the 1Dx HMW subunit DNA sequences from genomic DNA. The region amplified comprised the entire coding region of the gene. HMW subunit genes contain no introns, so genomic DNA can be used as a template for PCR amplification of the entire coding region with no intervening sequences.

A high-fidelity polymerase (Ex Taq, TaKaRa, Japan) was used instead of *Taq* polymerase in the amplification procedure to reduce the risk of introducing errors into the sequence. The different genes were amplified several times during the course of the study to synthesise more DNA for analysis, using genomic DNA as a template each time, and no evidence of either single base errors or changes in repeat structure was detected.

Fig. 1 SDS-PAGE of the HMW subunits present in seeds of cultivated bread wheat and *Ae. tauschii* lane a, *Ae. tauschii* AS2388; b, *Ae. tauschii* RM0198; c, wheat cv Avalon (HMW subunits 1Ax1, 1Bx6 + 1By8, 1Dx2 + 1Dy12); d, wheat MG7249 (1Ax2*, 1Bx7 + 1By8, 1Dx2.2 + 1Dy12) and e, wheat MG315 (1Bx7 + 1By8, 1Dx2.2* + 1Dy12). 1Dx subunits 2.1^t, 2^t, 2, 2.2 and 2.2* are indicated by arrows in tracks a, b, c, d and e, respectively

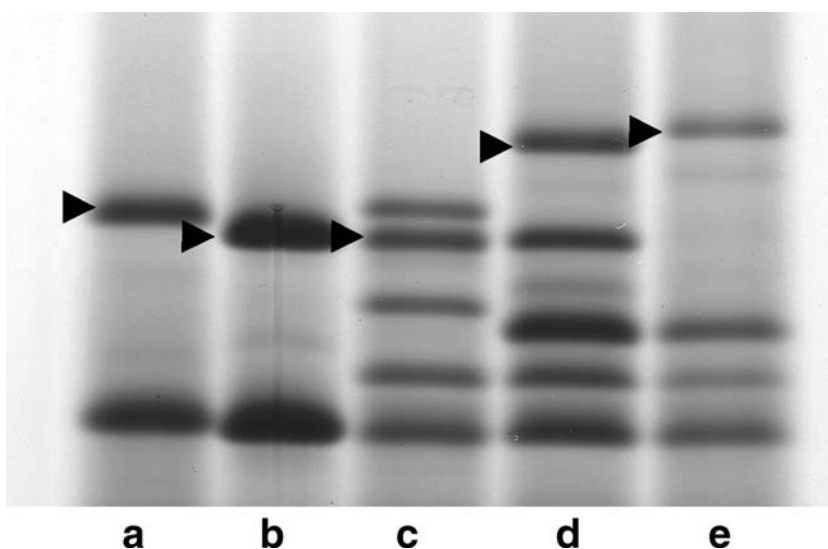
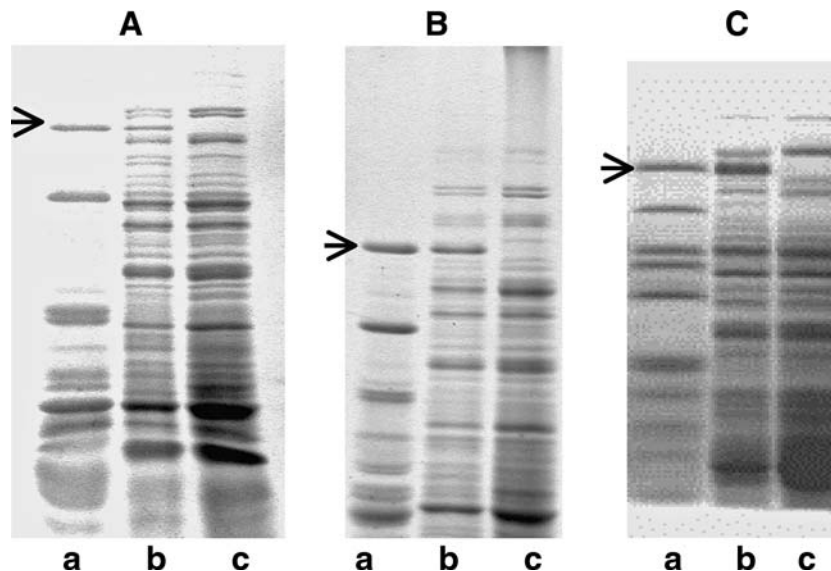


Fig. 2 Co-migration of HMW subunits expressed in *E. coli* with the corresponding proteins present in wheat grain. *A*, subunit 1Dx2^t; *B*, subunit 1Dx2.1^t; *C*, HMW subunit 1Dx2.2. Tracks *a* are total seed proteins from *Ae. tauschii* RMO198 (*A*), *Ae. tauschii* AS2388 (*B*) and *T. aestivum* MG7249 (*C*), with subunits 1Dx2^t, 1Dx2.1^t and 1Dx2.2, respectively, indicated by arrows. Tracks *b* and *c* are total protein extracts of *E. coli* cells expressing the corresponding genes with (*b*) and without (*c*) induction with IPTG



The nucleotide sequences of all four HMW subunit genes were determined by primer walking on overlapping nested deletions and deposited in the EMBL database under accession numbers AY159367 (1Dx2.2), AJ893508 (1Dx2.2*), AF480485 (1Dx2^t) and AF480486 (1Dx2.1^t). Each comprised a single open reading frame from beginning to end.

The preliminary characterisation of the 1Dx2.2* gene has been previously reported (D'Ovidio et al. 1996; Alvarez et al. 1997) while the authenticity of the other three genes was confirmed by expression of the coding regions in *E. coli* using the pET-30a expression vector, after removal of the sequence encoding the signal peptide (Fig. 2). This showed that the proteins expressed in *E. coli* co-migrated with the authentic subunits present in extracts of seed proteins.

Comparison of amino acid sequences

The characteristics of the mature proteins encoded by the four genes are compared with those of other 1Dx subunits in Table 1 and their amino acid sequences are aligned in Fig. 3. The amino acid sequences of four other 1Dx subunits are also included in this alignment:

1Dx2 (EMBL accession number X03346), 1Dx5 (X12928), 1Dx2.1 (AY517724), all from bread wheat, and 1Dx1.5^t (AY594355) from *Ae. tauschii*.

Although, the four novel 1Dx-type subunits differ considerably in size, from M_r of 87,000 for 1Dx2^t and 1Dx2.1^t to 101,000 for 1Dx2.2 and 107,000 for 1Dx2.2*, they all have similar structures to those of the 1Dx-type HMW subunits characterised previously. Each comprises a 21 amino acid signal peptide and short N- and C-terminal domains flanking a more extensive repetitive domain. The amino acid sequences of the N- and C-terminal domains are identical with the exception of a deletion of a single amino acid at position 60 of subunit 1Dx2 and an alanine to serine substitution at the penultimate residue of 1Dx1.5^t. The latter is the only change to distinguish these regions of the subunits of the cultivated species from those of its wild relative. The close similarity between the HMW subunits of cultivated wheat and its wild relatives, despite many years of selective breeding of cultivated wheat varieties, has also been demonstrated in a study by Wan et al. (2002) in which HMW subunit genes from *Ae. cylindrica* and *T. timopheevi* were cloned and characterised.

The structures of the repetitive domains of the 1Dx subunits from the two species are also essentially similar

Table 1 Comparison of the properties of the mature 1Dx2.2, 2.2*, 2^t and 2.1^t subunits with those of other 1Dx subunits

Subunit	Molecular mass	Number of amino acid residues				Number of cysteine residues			
		N-terminal domain	Repetitive domain	C-terminal domain	Total	N-terminal domain	Repetitive domain	C-terminal domain	Total
1Dx2	87,023	88	687	42	818	3	0	1	4
1Dx2 ^t	86,539	89	684	42	815	3	0	1	4
1Dx2.1	86,809	89	684	42	815	3	0	1	4
1Dx2.1 ^t	86,753	89	684	42	815	3	0	1	4
1Dx1.5 ^t	86,811	89	684	42	815	3	0	1	4
1Dx5	88,128	89	687	42	818	3	1	1	5
1Dx2.2	100,887	89	819	42	950	3	0	1	4
1Dx2.2*	106,600	89	872	42	1003	3	0	1	4

with those of subunits 1Dx2^t, 1Dx2.1^t and 1Dx1.5^t being particularly close to 1Dx2 in structure and size. In x-type subunits, the repetitive domain consists of three types of repeat motif (Halford et al. 1987; Shewry et al. 1989): a hexapeptide with the consensus sequence Pro Gly Gln Gly Gln Gln (boxed in red in Fig. 3), a nine-amino acid motif consisting of the hexapeptide plus a tripeptide with the consensus sequence Gly Gln Gln (boxed in yellow in Fig. 3), and a 15-amino acid repeat consisting of the hexapeptide plus a nonapeptide with the consensus sequence Gly Tyr Tyr Pro Thr Ser Pro Gln Gln (boxed in light blue in Fig. 3). In addition, 23 amino acid residues at the N-terminal ends of the repetitive domains of all

subunits show similarity with the other repeats but do not have the same structure, while truncated hexapeptide/nonapeptide repeats are present at the C-terminal ends of the repetitive domains (both boxed in dark blue in Fig. 3).

The variation in size of the subunits is due almost entirely to the insertion and duplication of blocks of these repeats. It is notable that there is more variation in this respect between different allelic 1Dx subunits within cultivated varieties than between those from cultivated wheat and wild *Ae. tauschii*. Although subunits 1Dx2.2 and 1Dx2.2* are both larger than subunit 1Dx2, comparisons of their amino acid sequences demonstrate that

Fig. 3 Multiple sequence alignment of the derived amino acid sequences of 1Dx-type HMW subunits. The alignments were assembled by eye to demonstrate the repeat structure of the central domains. Cysteine residues are highlighted in pink; hexapeptide repeat motifs are boxed in red, hexapeptide/tripeptide nine-amino acid repeat motifs in yellow and hexapeptide/nonapeptide 15 amino acid repeat motifs in light blue. Degenerate repeat sequences at the N- and C-terminal ends of the repetitive domain are boxed in dark blue. The sequences shown are those of the four HMW subunits described for the first time in this study, 1Dx2.2* (EMBL accession number AJ893508), 1Dx2.2 (AY159367), 1Dx2^t (AF480485) and 1Dx2.1^t (AF480486), plus those of 1Dx2 (X03346), 1Dx5 (X12928), 1Dx2.1 (AY517724) and 1Dx1.5^t (AY594355). Those denoted with a ^t are from *Ae. tauschii*, the others are from *T. aestivum*. The regions that have been duplicated in 1Dx2.2 and 1Dx2.2* are boxed in grey. The arrows indicate unusual amino acid residues that are present in blocks of repeats that have been duplicated in the evolution of subunits 1Dx2.2 and 1Dx2.2*

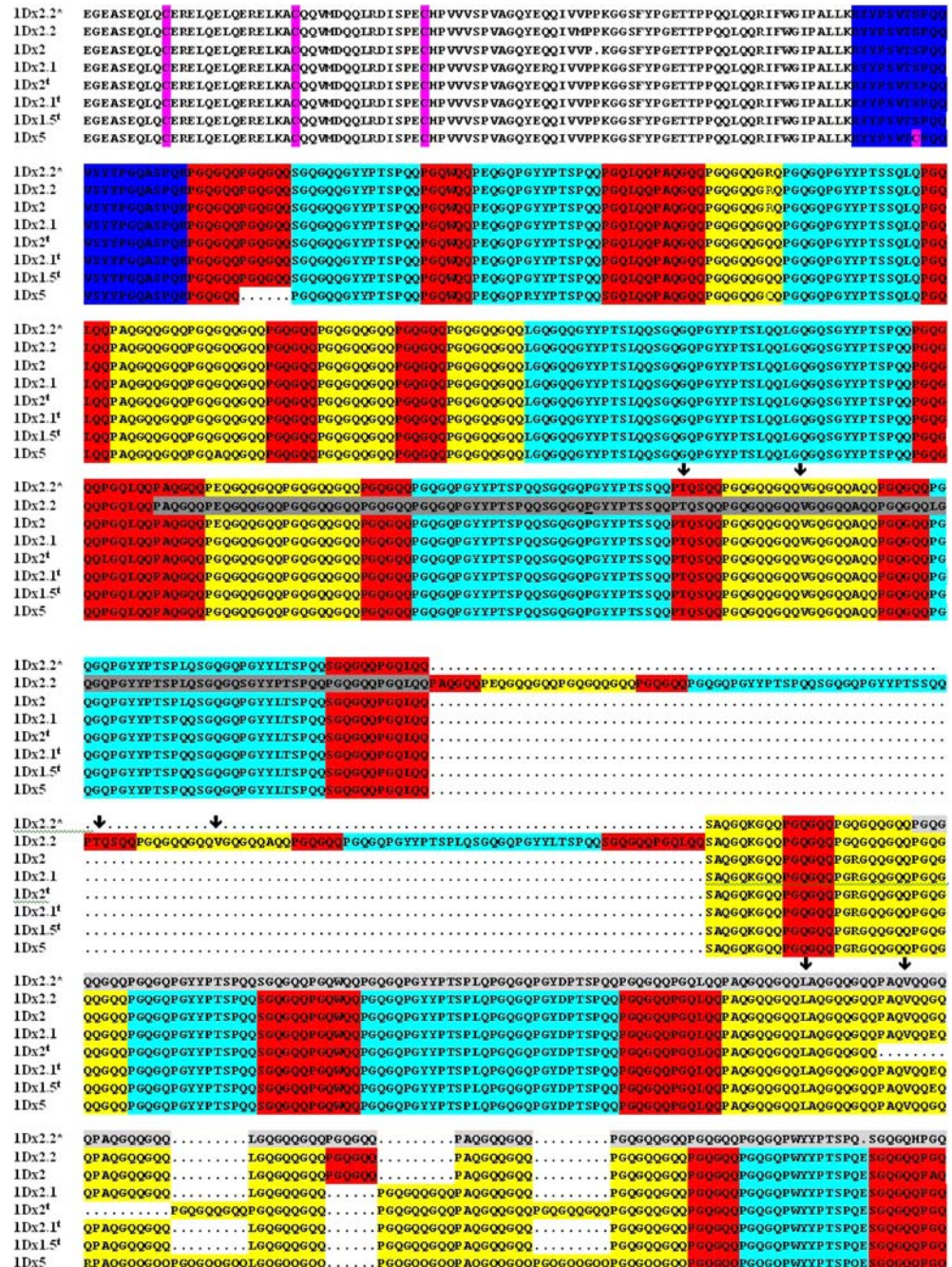
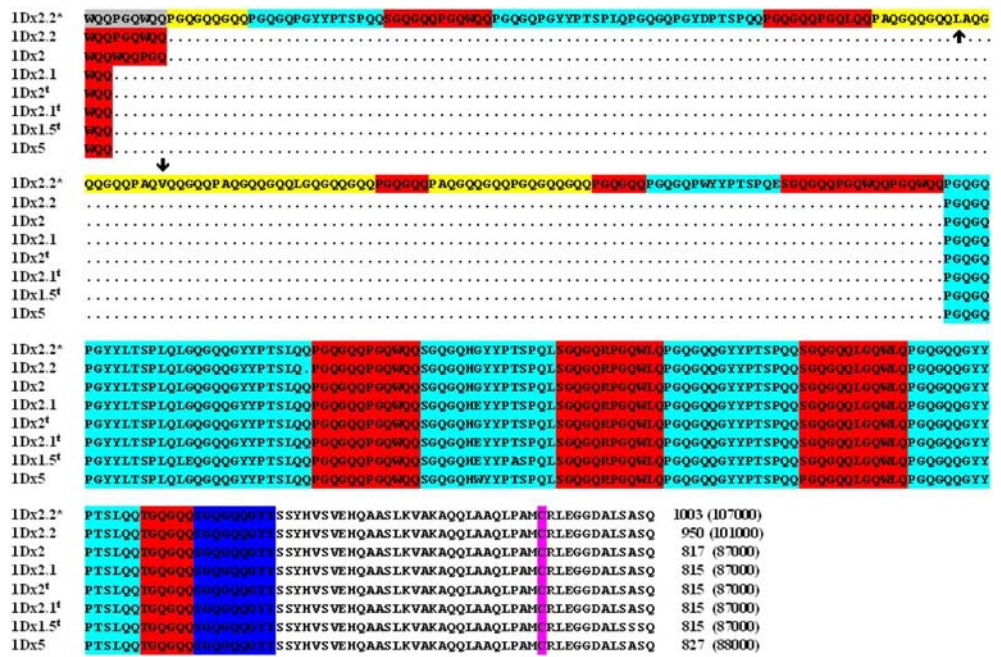


Fig. 3 Contd.

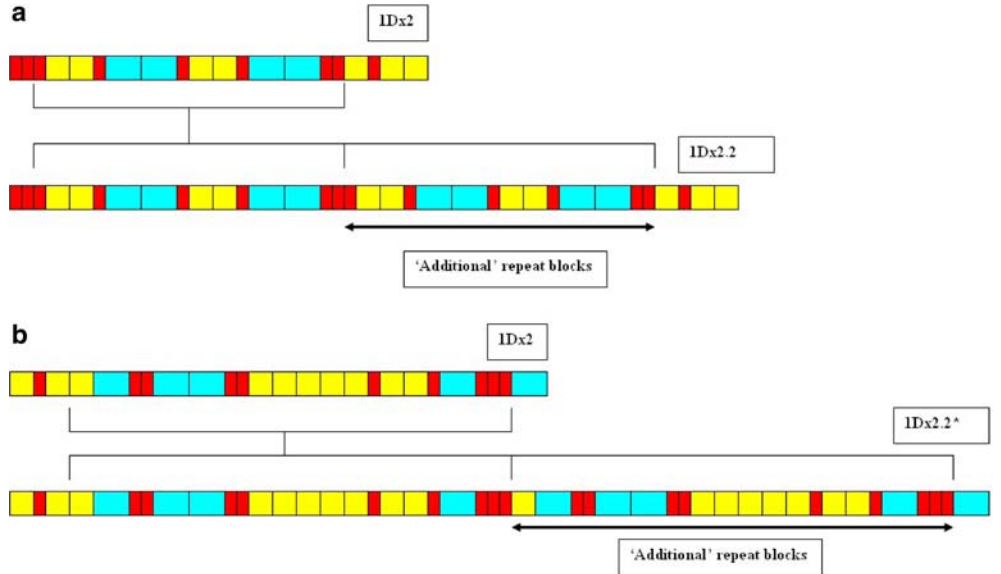


they have arisen from separate duplication events (as shown in Figs. 3, 4). The regions that have been duplicated can be identified based on the orders of the additional repeat blocks and the presence of unusual residues at the same relative positions in the duplicated regions (note for example, the residues indicated with an arrow in Fig. 3). Thus, subunit 1Dx2.2 arose from the duplication of a region comprising six hexapeptides, four nine-amino acid blocks and four 15-amino acid blocks, a total of 132 amino acid residues. This region is shaded in dark grey in Fig. 3 and shown diagrammatically in Fig. 4a. In 1Dx2.2* a different region has been duplicated, comprising nine hexapeptides, eight nine-amino acid blocks and four 15-amino acid blocks, a total of 186

amino acid residues. This region is shaded in light grey in Fig. 3 and shown diagrammatically in Fig. 4b.

Such duplications presumably arose from unequal crossing over during meiosis (D’Ovidio et al. 1996). It is notable that in both cases the ‘new’ region is immediately adjacent to that from which it was duplicated, and that complete, intact repeat blocks remain at both ends. The latter, at least, must also apply to the vast majority of similar events in HMW subunit evolution to account for the high degree of conservation of their repeat structures. In contrast, while the other major classes of wheat prolamins, the gliadins and LMW subunits, also contain repetitive domains, their repeat structures are much less well conserved (Shewry et al. 2003a).

Fig. 4 Diagrammatic representation of the major repeat block duplication events that have occurred in the evolution of 1Dx2.2 (a) and 1Dx2.2* (b), resulting in their unusually large size. Hexapeptide repeat motifs are represented in red, hexapeptide/tripeptide nine-amino acid repeat motifs in yellow and hexapeptide/nonapeptide 15-amino acid repeat motifs in light blue. Different regions of the same 1Dx2 protein are shown in parts a and b. The locations of the duplication events are shown in Fig. 3



As with other HMW subunits, there is considerable degeneracy within the repeat sequences, although the residues at certain positions (notably 3, 5 and 6 of the hexapeptide, 1–3 of the tripeptide and 1, 6 and 8 of the nonapeptide) are highly conserved (Shewry et al. 2003b; Feeney et al. 2003). Where mutations do occur they are often repeated in adjacent motifs, sometimes in combination with other mutations. One example of this is the presence of seven pairs of adjacent hexapeptides in the C-terminal half of the proteins in which the residue at position 1 of the first hexapeptide is a serine instead of a proline and the residue at position 4 of the second hexapeptide is a tryptophan instead of a glycine. This is consistent with the hypothesis that the repetitive domains have evolved through the duplication of blocks of repeat motifs and therefore any mutations that occurred were also duplicated. Alternatively, it is possible that mutations have spread through adjacent repeats through gene conversion (Shewry et al. 2003b).

Implications for wheat quality improvement

Two aspects of HMW subunit structure have been suggested to contribute to their role in determining dough strength. Firstly, their ability to form high-molecular mass polymers stabilised by inter-chain disulphide bonds. In this case the presence of an additional cysteine residue in subunit 1Dx5 (at position 97, at the N-terminal end of the repetitive domain) is considered to contribute to the formation of a higher proportion of large glutenin polymers in lines expressing subunits 1Dx5 + 1Dy10 than in lines expressing the allelic subunit pair 1Dx2 + 1Dy12 (Popineau et al. 1994). It is notable that none of the other 1Dx subunits discussed here contain this additional cysteine residue. The second aspect of subunit structure is the size and degree of conservation of the repetitive sequences. Békés et al. (1994, 1995) reported a positive relationship between the size of HMW subunits and their effect on dough strength when incorporated in dough using a 2 g Mixograph, with subunits 1Dx2.2 and 1Dx2.2* showing greater strength than subunit 1Dx2. This is consistent with a model in which the glutenin polymers interact via inter-chain hydrogen bonds formed between the subunit repetitive domains with longer subunits forming more stable interactions (Belton 1999; Feeney et al. 2003).

Finally, the apparently low level of variation in the HMW subunits of *Ae. tauschii* compared with that in the subunits encoded by chromosome 1D of bread wheat is surprising since the formation of hexaploid wheat is thought to have resulted in a genetic “bottleneck” with lower diversity present in the present D genome of bread wheat than in wild diploid populations. The existence of this bottleneck is supported by the studies of the puroindolines which are encoded by genes on chromosome 5D of bread wheat and *Ae. tauschii*. In particular, Massa et al. (2004) identified five new allelic forms of puro-

indoline a and three new allelic forms of puroindoline b in only 50 accessions of *Ae. tauschii*, which contrasts with bread wheat where a total of only eight puroindoline b alleles and two puroindoline a alleles (one a *null* form) have been reported in several hundred accessions (Morris 2002; Xia et al. 2005). We have no explanation why the HMW subunit genes encoded by chromosome 1D display a different pattern of variation between bread wheat and *Ae. tauschii* than the puroindolines encoded by the same chromosomes.

Acknowledgements Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. Yongfang Wan was supported by a Royal Society China-UK Joint Project Grant (Q813 “Aegilops species as a source of novel genes for wheat improvement”). Daowen Wang was supported by grants (2002CB111301, 2003AA222040) from the Ministry of Science and Technology of China.

References

- Alvarez JB, D’Ovidio R, Lafiandra D (1997) Comparison of allelic x-type genes present at the *Glu-D1* locus in bread wheat by PCR and sequence analysis of their N-terminal domain. *J Genet Breed* 51:61–166
- 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
- Békés F, Anderson OD, Gras PW, Gupta RB, Tam A, Wrigley CW, Appels R (1994) The contribution to mixing properties of 1D HMW glutenin subunits expressed in a bacterial system. In: Henry RJ, Ronalds JA (eds) Improvement of cereal quality by genetic engineering. Plenum Press, New York, pp 97–103
- Békés F, Gras PW, Gupta RB (1995) The effect of purified cereal polypeptides on the mixing properties of dough. In: Williams YA, Wrigley CW (eds) Proceedings of the 45th Australian Cereal Chemistry Conference. The Royal Australian Chemical Institute, Cereal Chemistry Division, pp 92–98
- Belton PS (1999) On the elasticity of wheat gluten. *J Cereal Sci* 29:103–107
- 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, Lafiandra D, Porceddu E (1996) Identification and molecular characterisation 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
- Feeney KA, Wellner N, Gilbert SM, Halford NG, Tatham AS, Shewry PR, Belton PS (2003) Molecular structures and interactions of repetitive peptides based on wheat glutenin subunits depend on chain length. *Biopolym Biospectrosc* 72:123–131
- Feldman M (2000) Origin of cultivated wheat. In: Bonjean AP, Angus WJ (eds) The world wheat book: a history of wheat breeding. Intercept Ltd, London, pp 3–56
- Feldman M, Lupton FGH, Miller TE (1995) Wheats. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman Group Ltd, London, pp 184–192
- Gage LP, Manning RF (1980) Internal structure of the silk fibroin gene of *Bombyx mori*. I. The fibroin gene consists of a homogeneous alternating array of repetitive crystalline and amorphous coding sequences. *J Biol Chem* 255:9444–9450
- Halford NG, Forde J, Anderson OD, Green 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 1A and 1D. *Theor Appl Genet* 75:117–126

- Lagudah ES, Halloran GM (1988) Phylogenetic relationships of *Triticum tauschii* the D genome donor to hexaploid wheat. I. Variation in HMW subunits of glutenin and gliadins. *Theor Appl Genet* 75:592–598
- Lewin B (1997) *Genes VI*. Oxford University Press, Oxford
- Lewis R (2003) Sequences, structures and properties of spider silks. In: Shewry PR, Tatham AS, Bailey AJ (eds) *Elastomeric proteins: structures, biomechanical properties and biological roles*. Cambridge University Press, Cambridge, pp 136–151
- Mackie AM, Lagudah ES, Sharp PJ, Lafiandra D (1996) Molecular and biochemical characterisation of HMW glutenin subunits from *T. tauschii* and the D genome of hexaploid wheat. *J Cereal Sci* 23:213–225
- Manning RF, Gage LP (1980) Internal structure of the silk fibroin gene of *Bombyx mori*. II. Remarkable polymorphism of the organisation of the crystalline and amorphous coding sequences. *J Biol Chem* 255:9451–9457
- Margiotta B, Colaprico G, D'Ovidio R, Lafiandra D (1993) Characterisation of high M_r subunits of glutenin by combined chromatography (RP-HPLC) and electrophoretic separations and restriction fragment length polymorphism (RFLP) analyses of their encoding genes. *J Cereal Sci* 17:221–236
- Massa AN, Morris CF, Gill BS (2004) Sequence diversity of puroindoline-a, puroindoline-b and the grain softness protein genes in *Aegilops tauschii* Coss. *Crop Sci* 44:1808–1816
- Morris CF (2002) Puroindolines: the molecular genetic basis of wheat grain hardness. *Plant Mol Biol* 48:633–647
- Nakamura H, Hirano H, Sasaki H, Yamashita A (1990) A high molecular weight subunit of wheat glutenin seed storage protein correlates with its flour quality. *Jpn J Breed* 40:485–494
- Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on break making quality. *Annu Rev Plant Physiol* 38:141–153
- 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
- Popineau Y, Cornec M, Lefebvre J, Marchylo B (1994) Influence of high M_r glutenin subunits on glutenin polymers and rheological properties of gluten and gluten subfractions of near-isogenic lines of wheat Sicco. *J Cereal Sci* 19:231–241
- 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. In: Miflin BJ (ed) *Oxford surveys of plant molecular and cell biology*, vol 6. Oxford University Press, Oxford, pp 163–219
- Shewry PR, Tatham AS, Fido RJ (1995) Separation of plant proteins by electrophoresis. In: Jones H (ed) *Methods in molecular biology—plant gene transfer and expression protocols*, vol 49. Humana Press, USA, pp 399–422
- Shewry PR, Halford NG, Lafiandra D (2003a) In: Hall JC, Dunlap JC, Friedman T (eds) *The genetics of wheat gluten proteins*. *Adv Genet*, vol 49. Academic Press, San Diego, pp 111–184
- Shewry PR, Halford NG, Tatham AS, Popineau Y, Lafiandra D, Belton PS (2003b) The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. *Adv Food Nutr Res* 45:221–302
- Shewry PR, Tatham AS, Bailey A (eds) (2003c) *Elastomeric proteins*. Cambridge University Press, Cambridge, pp 331–351
- Sugiyama T, Rafalski A, Peterson D, Sol D (1985) A wheat HMW glutenin subunit gene reveals a highly repeated structure. *Nucleic Acids Res* 13:8729–8737
- Tatham AS, Shewry PR (2003) Comparative structures and properties of elastic proteins. In: Shewry PR, Tatham AS, Bailey AJ (eds) *Elastomeric proteins: structures, biomechanical properties and biological roles*. Cambridge University Press, Cambridge, pp 338–351
- Wan Y, Liu K, Wang D, Shewry PR (2000) High-molecular-weight glutenin subunits in the *Cylindropyrum* and *Vertebrata* section of the *Aegilops* genus and identification of subunits related to those encoded by the Dx alleles of common wheat. *Theor Appl Genet* 101:879–884
- Wan Y, Wang D, Shewry PR, Halford NG (2002) Isolation and characterisation of five novel high molecular weight subunit of glutenin genes from *Triticum timopheevi* and *Aegilops cylindrica*. *Theor Appl Genet* 104:828–839
- William MDHM, Peña RJ, Mujeeb-Kazi A (1993) Seed protein and isozyme variations in *Triticum tauschii* (*Aegilops squarrosa*). *Theor Appl Genet* 87:257–263
- Xia L, Chen F, He Z, Chen X, Morris CF (2005) Occurrence of puroindoline alleles in Chinese winter wheats. *Cereal Chem* 82:38–43