Theor Appl Genet (1998) 96: 568–576

O. D. Anderson · F. A. Abraham-Pierce · A. Tam

Conservation in wheat high-molecular-weight glutenin gene promoter sequences: comparisons among loci and among alleles of the *GLU-B1-1* locus

Received: 5 August 1997 / Accepted: 6 November 1997

Abstract The high-molecular-weight glutenin (HMW) genes and encoded subunits are known to be critical for wheat quality characteristics and are among the best-studied cereal research subjects. Two lines of experiments were undertaken to further understand the structure and high expression levels of the HMW-glutenin gene promoters. Cross hybridizations of clones of the paralogous x-type and y-type HMW-glutenin genes to a complete set of six genes from a single cultivar showed that each type hybridizes best within that type. The extent of hybridization was relatively restricted to the coding and immediate flanking DNA sequences. Additional DNA sequences were determined for four published members of the HMW-glutenin gene family (encoding subunits Ax2*, Bx7, Dx5, and Dy10) and showed that the flanking DNA of the examined genes diverge at approximately −1200 bp 5′ to the start codon and 200–400 bp 3′ to the stop codon. These divergence sites may indicate the boundaries of sequences important in gene expression. In addition, promoter sequences were determined for alleles of the Bx gene (*Glu-B1-1*), a gene reported to show higher levels of expression than other HMW-glutenin genes and with variation among cultivars. The sequences of Bx promoters from three cultivars and one wild tetraploid wheat indicated that all Bx alleles had few differences and contained a duplicated portion of the promoter sequence "cereal-box" previously suspected as a factor in higher levels of expression. Thus, the "cereal-box" duplication preceeded the origin of hexaploid wheat, and provides no evidence to explain the variations in Bx subunit synthesis levels. One active Bx allele contained a 185-bp insertion that evidently resulted from a transposition event.

Key words: Wheat · High-molecular-weight glutenin · Promoter · Transposon · Genome evolution

Communicated by M. A. Saghai Maroof

O. D. Anderson (✉) · F. A. Abraham-Pierce[1] · A. Tam[2]
United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710, USA.

Present addresses:
[1] 1215 Santa Fe Avenue, Berkeley, CA 94706, USA
[2] Onyx Pharmaceuticals, 3031 Research Drive, Richmond, CA 94806, USA

Introduction

The high-molecular-weight (HMW) glutenins are a family of prominent polypeptides found in the seeds of wheats and related wild grasses. They are critical for dough gluten processing characteristics, and are a major topic in studies of wheat quality (Shewry et al. 1992, 1996) and wheat bioengineering (Vasil and Anderson 1997). The genes map to the *Glu-1* orthologous loci of each of the three genomes (A, B, and D) of *Triticum aestivum* L. The HMW-glutenin genes are divided into similar x-type and y-type paralogous genes which evolved by duplication. Each *Glu-1* compound locus contains one x-type and one y-type gene. Besides the need to understand the molecular basis of gluten visco-elastic properties, there is interest in determining the basis of the developmental specificity and the high levels of expression of the HMW-glutenin gene family in wheat endosperm: up to 2% of the seed protein per gene. In addition, individual HMW-glutenins accumulate to different levels; i.e., in some wheat cultivars the HMW-glutenin Bx polypeptide (encoded by the *Glu-B1-1* gene) accumulates to a higher level than the other HMW-glutenins (Marchylo et al. 1992; Kolster et al. 1993). Anderson and Greene (1989) suggested a connection between this higher expression and a 54-bp Bx promoter-sequence duplication that includes the "cereal-box" implicated in seed-specific expression (Forde et al. 1985a). This duplicated sequence is also

part of an 85-bp deletion in an inactive Cheyenne Ay (*Glu-A1-2*) gene (Forde et al. 1985 b).

Until recently, detailed promoter studies under native conditions were limited by the inability to transform monocots. This necessitated studying HMW-glutenin promoters in heterologous systems: transgenic tobacco (Halford et al. 1989; Robert et al. 1989; Thomas and Flavell 1990) and transient monocot cell assays (Blechl et al. 1994). These reports demonstrated that the first few hundred bp of the HMW-glutenin promoter were sufficient for gene expression and apparent seed specificity in tobacco. However, the accumulation of HMW-glutenin in tobacco was found to be less than 5%, or much lower, than the level in wheat endosperm (Roberts et al. 1989). The levels of activity of a HMW-glutenin promoter in transgenic tobacco varied over 2–3 orders of magnitude (Thomas and Flavell 1990). These results are in contrast with the stability of expression and the high levels of synthesis in the first report of HMW-glutenin promoter function in transgenic wheat (Blechl and Anderson 1996) where more extensive flanking DNA was used.

In order to further understand the control of HMW-glutenin gene expression, more distal sequence information is required both from individual examples of this gene family and from alleles that show differential expression. This information may indicate the presence of conserved sequence domains important in gene expression and would allow the polymerase chain reaction (PCR) to be used to isolate new promoters. Although PCR has been used for qualitative studies of the HMW-glutenins (D'Ovidio and Anderson 1994; D'Ovidio et al. 1994; Smith et al. 1994) the accurate amplification and cloning of intact specific HMW-glutenin promoters for use in promoter function studies and whole-plant transformations has not been demonstrated.

We have previously reported on the near-flanking DNA sequences for the complete set of HMW-glutenin genes from the wheat cultivar Cheyenne (Forde et al. 1985 b; Halford et al. 1987; Anderson and Greene 1989; Anderson et al. 1989). We now report additional flanking sequences for four of these genes. The new additional 5' promoter sequences were used to design primers for the specific amplification and sequencing of Bx promoter regions from different germplasms, both to test the accuracy of such amplifications from multigene families and to study the distribution of the Bx "cereal-box" duplication.

Materials and methods

HMW-glutenin clones

Most of the HMW-glutenin clones used for restriction mapping and sequencing were those already published: Ay (Forde et al. 1985 b); Ax2* and Bx7 (Anderson and Greene 1989); Dx5 and Dy10 (Anderson et al. 1989). The By9 clone was a different isolate of the same gene as reported in Halford et al. (1987). Table 1 lists the six different HMW-glutenin genes for hexaploid bread wheat, the specific HMW-glutenin subunits encoded by these genes in cultivar Cheyenne, and the clones for these genes used in the present report.

Table 1 Wheat cultivar Cheyenne HMW-glutenin genes and clones

HMW gene	HMW subunit	Clone
Glu-A1-1	Ax2*	pWHE3 (Ax2*)[a]
Glu-A1-2	Ay	λWHE4 (Ay)
Glu-B1-1	Bx7	λWHE5 (Bx7)[b]
Glu-B1-2	By9	λWHE6 (By9)
Glu-D1-1	Dx5	pWHE2 (Dx5)
Glu-D1-2	Dy10	pWHE1 (Dy10)

[a] Clone λ1Bi-2 in Anderson and Greene (1989)
[b] Clone λR23 in Anderson and Greene (1989)

DNA cloning and sequencing

Extensions of HMW-glutenin flanking sequences were obtained using the clones and methods described in Anderson and Greene (1989), and Anderson et al. (1989). DNA sequencing of both single-stranded M13 and plasmid DNAs was performed using dideoxynucleotide-chemistry sequencing kits (Pharmacia and Amersham), Sequenase DNA T7 polymerase (U.S. Biochemical), and the DNA polymerase I Klenow fragment (Pharmacia and Bethesda Research Laboratories). Compressions in sequencing were resolved using deaza-GTP and dITP kits from Pharmacia with both Sequenase and Klenow DNA polymerases.

Before cloning, the *Taq* polymerase was removed (Crowe et al. 1991), by making the PCR amplification reactions 40 µg/ml proteinase K, incubating for 30 min at 37°C, then extracting with phenol once and chloroform twice, and precipitating with ethanol. PCR products were cloned either by ligating directly into the *Sma*I site of M13 and pKS+ (Stratagene) or by restricting the PCR reaction products with *Bam*HI and ligating into the *Bam*HI site of the same vectors. Vectors containing HMW-glutenin fragments were identified with a HMW-glutenin hybridization probe prepared from an M13 clone (mB9G) which contains a 4.3-kb *Hin*dIII fragment of the cv Cheyenne Bx7 promoter and N-terminal coding DNA sequences subcloned from λWHE5(Bx7). Several clone isolates from each PCR had single-base differences from the other isolates and were assumed to result from PCR errors. The consensus sequences of the multiple isolates from each PCR are considered to be the true gene sequences.

Wheat DNA purification

The sources of seeds were: cultivars Chinese Spring and Cheyenne – Don Kasarda (USDA, Albany); cv Bidi – Peter Shewry (Long Ashton, UK); Chinese Spring nullisomic-tetrasomic genetic lines – Jan Dvorak (University of California, Davis); *Triticum turgidum* ssp. dicoccoides – Eviator Nevo (University of Haifa). DNA from the cultivars Cheyenne and Chinese Spring were prepared from embryo nuclei (Anderson et al. 1997). DNA from the wheat cultivar Bidi (durum) and wild tetraploid *T. turgidum* were prepared by compressing 100 mg of fresh, young leaves into a microfuge tube and covering with liquid nitrogen. Leaves were ground with a pre-cooled plastic pestle, keeping the leaf-tissue submerged. The nitrogen was immediately poured off, and 500 µl of hot (80°C) phenol/buffer

mixture was added (one part of neutralized phenol vortexed with one part of 100 mM NaCl, 100 mM Tris 8.0, 10 mM EDTA, 1.0% SDS). Samples were vortexed for about 60 s, then 250 µl of chloroform was added and the sample vortexed about 60 s. Phases were separated by centrifugation and the top phase was removed and a 1/20 vol of 3 M Na acetate (pH 7.0) and 2 vol of ethanol were added to this top phase. The samples were placed on dry ice for at least 15 min. The DNA was pelleted at 12 000 rpm for 10 min and then rinsed with cold 70% ethanol. The pellet was then dissolved in 50 µl 10 mM Tris 8.0, 1 mM EDTA.

Primer synthesis

The primer pair for amplification of the HMW-glutenin Bx sequences was: P1, GCGGATCCGAATTCCGGCAACAACTTG TGGGGTA; P2, GCAGGATCCACTACTGCCGCAAAGAG-GACCAGG. The first ten bases in the 5′ end of primer P1 contains ten bases providing the restriction sites *Bam*HI and *Eco*RI for the cloning of amplification products. Similarly, primer P2 contains nine additional bases that include a *Bam*HI site. The remainder of the sequences were complementary to the cv Cheyenne Bx7 DNA sequence. Primers P1 and P2 were also used in sequencing the ends of Bx clones.

Polymerase chain reaction conditions

PCR was carried out using a Perkin-Elmer Cetus DNA Thermal Cycler operated according to the manufacturers instructions. Basic cycling conditions were 30 cycles each with a 40-s denaturing step at 95°C, a 30 s annealing step at 53°C, and a 3-min extension step at 72°C. The final extension step was for 7 min, followed by a 7°C soak step. Reaction components were 100 ng of DNA in 1–5 µl, 5 µl of 10 × PCR buffer, 1 µl of 25 mM NTPs, 1 µM of each primer, 4 units of *Taq* polymerase (Perkin-Elmer), 3 µl of 25 mM $MgCl_2$, H_2O to a 50-µl volume, and the reaction was overlaid with 70 µl of mineral oil.

In some reactions the background of amplified DNA was high when preparing to clone PCR products, particularly with the DNA samples from cv Chinese Spring. To reduce the background, the first Bx amplification product was separated on a polyacrylamide gel. The appropriate band was cut out, crushed in H_2O, soaked overnight at 6°C, and spun for 10 min at 11 000 rpm in a microfuge. One microliter of the supernatant was used in a second round of amplification employing the same primers and conditions as in the first round.

Results and discussion

HMW-glutenin gene structure

Figure 1 A diagrams a generic HMW-glutenin gene structure. The coding sequence is composed of a central repetitive domain and two unique-sequence termini. Figure 1B shows the *Eco*RI fragments containing the six HMW-glutenin genes of wheat cultivar Cheyenne, and indicates the position of *Hin*dIII sites. The coding sequence is indicated by the arrow under each insert. All six genes contain a conserved *Hin*dIII site at the junction of the N-terminal domain and the central repetitive domain (Fig. 1 A), and a second conserved *Hin*dIII site near the polyadenylation site.

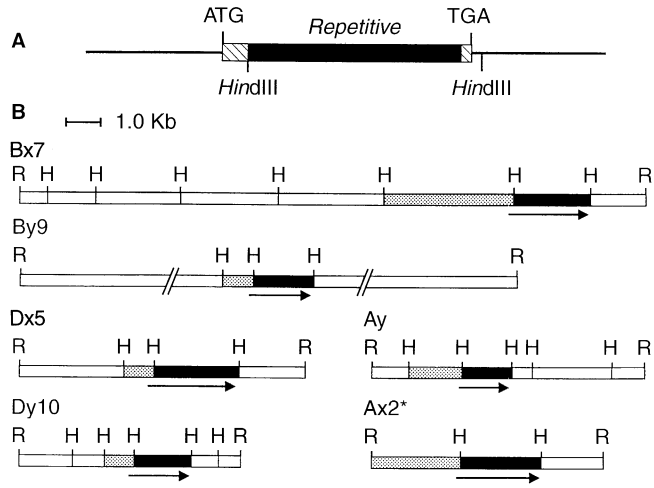

Fig. 1A, B Restriction maps of HMW-glutenin genes. **A** Generalized HMW-glutenin gene structure. The coding sequence is indicated by *bars*. The repetitive domain is *black* and the non-repetitive termini are *hatched*. The position is given for the two *Hin*dIII sites conserved in all known HMW-glutenin genes. ATG = start codon. TGA = stop codon. **B** *Eco*RI and *Hin*dIII sites are shown for all six *Eco*RI fragments containing the HMW-glutenin genes in the wheat cultivar Cheyenne. *Black bars* are the HMW-glutenin repetitive coding domains. *Shaded fragments* are the adjoining 5′ fragments to the repetitive domain and contain the N-terminal unique-coding region and portions of the promoter and 5′ flanking DNA. *R* = *Eco*RI. *H* = *Hin*dIII. The indicated By9 *Hin*dIII sites are only those known from DNA sequences. *Arrows* under the fragments indicate HMW-glutenin subunit coding regions

Cross-hybridization of HMW-glutenin gene sequences

To estimate the extent of sequence homology within *Eco*RI fragments containing the HMW-glutenin gene family, *Eco*RI + *Hin*dIII digests of these fragments were probed with entire paralogous x- and y-type clones. The autoradiograms of the resulting fragment hybridization patterns are shown for an Ax2* probe (Fig. 2 A) and a Dy10 probe (Fig. 2 B). As expected, the Ax2* probe hybridized more strongly to x-type genes (Ax2*, Bx7, Dx5) than to y-type (Ay, By9, Dy10). Similarly, the Dy10 probe hybridizes best to y-type genes. Both probes hybridized to the repetitive-domain fragments (filled dots in Fig. 2) and the immediate 5′ flanking fragments (open dots) containing the N-terminal coding domain plus promoter sequences of all six genes. However, there was more extensive hybridization with flanking DNA fragments with other members of the same type. For example, the Ax2* probe hybridizes only with the repetitive domain and the 5′ fragment of the y-type genes (Fig. 2 A). Similarly, the Dy10 probe hybridizes to more fragments, and more intensely, to other y-type genes than to x-type genes (Fig. 2 B). Thus, these hybridizations suggest that homology among the HMW-glutenin genes is best within the coding DNA and the immediate, promoter-containing, 5′ flanking

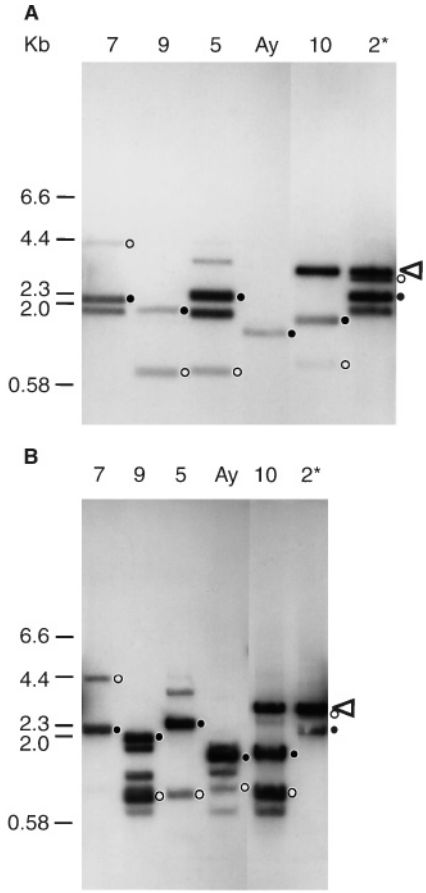

Fig. 2A, B Southern analysis of HMW-glutenin genes. Clones containing all six wheat cultivar Cheyenne HMW-glutenin genes within *Eco*RI insertions were restricted with *Eco*RI + *Hin*dIII and the fragments separated on agarose gels. Southern blots were probed with: **A** the complete Ax2* clone, or B the complete Dy10 clone. The *open arrowhead* indicates the plasmid band for the Ax2* and Dy10 clones (all other genes are cloned in lambda vectors which do not cross hybridize with the probe vector sequences). *Solid dots* indicate fragments containing the HMW-glutenin repetitive coding domain (same DNA as the black fragments in Fig. 1B). *Open dots* indicate fragments containing the N-terminal non-repetitive domain plus different lengths of the 5′ promoter and flanking DNA (the same DNA as the shaded fragments in Fig. 1B). Molecular-weight markers are given on the left

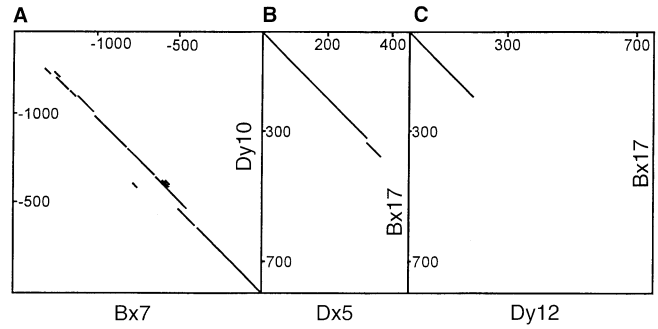

Fig. 3A–C Homology matrix alignments of HMW-glutenin gene flanking DNAs. The DNA sequences flanking the coding domains were compared for pairs of genes. The scoring criterion was a 90% match over a 20-base window. Larger windows and lower percent matches gave the same result but with more background. **A** The 5′ sequence of Bx7 from −1532 to +11 compared to the 5′ sequence of Dy10 from −1465 to +11. **B** The Bx17 3′ sequence from the stop codon to 748 bp downstream compared to the 3′ sequence of Dx5 from the stop codon to 431 bases downstream. **C** The Bx17 3′ sequence from the stop codon to 748 bp downstream compared to the 3′ sequence of Dy12 from the stop codon to 690 bp downstream

DNA. The hybridizations in Fig. 2 also suggest that the Ax2* sequence is less homologous to Bx7 than to Dx5, and that the Dy10 sequence is less homologous to Ay than to By9.

Comparison of HMW-glutenin gene flanking DNAs

We previously reported the locations of the DNA sequences for the wheat HMW-glutenin genes Ax2* (*Glu-A1-1c*; Anderson and Greene 1989), Bx7 (*Glu-B1-1b*; Anderson and Greene 1989), and Dy10 (*Glu-D1-2b*; Anderson et al. 1989) to positions −535, −681, and −480 bp upstream of the translation start codon, with the Dx5 gene (*Glu-D1-1d*; Anderson et al. 1989) to 126 bp downstream from the stop codon. We have now extended these four DNA sequences to positions −1334 (Ax2*), −1532 (Bx7), and −1465 (Dy10) for the 5' regions, and to 431 bp downstream from the stop codon for Dx5. The full sequences can be found under GenBank accessions M22208, X13927, X12929, and X12928, respectively. Analysis of the extended promoter sequences shows the three promoters contain a conserved DNA sequence for only approximately 1200 bp upstream of the coding region. The sequences further upstream have no detectable homology. This is shown in Fig. 3 A with a homology matrix of the Bx7 vs Dy10 5′ sequences (the other two pairwise plots give similar results). A similar divergence occurs in the HMW-glutenin 3′ flanking sequences. Figure 3 B shows that two x-type genes, Bx17 and Dx5, share homology for at least 400 bp from the stop codon, then diverge. Similarly, the divergence at approximately 200 bp for the x-type and y-type sequences is shown in Fig. 3 C for the Bx17 (Reddy and Appels 1993) and Dy12 genes (Thompson et al. 1985). Since the x-types are evolutionarily more related to each other than to the y-types, it may be that the loss of 3′ homology in non-critical sequences has proceeded less among similar-type genes then between x- and y-types.

Figure 4 gives the alignment of the three cv Cheyenne HMW-glutenin promoter sequences starting at the 5′ position where homology begins. The alignment shows mainly single base changes, plus short deletion/insertions. The single major difference in the sequence is the 54-bp duplication in the Bx promoter, which includes the "cereal-box," as previously reported (Anderson and Greene 1989).

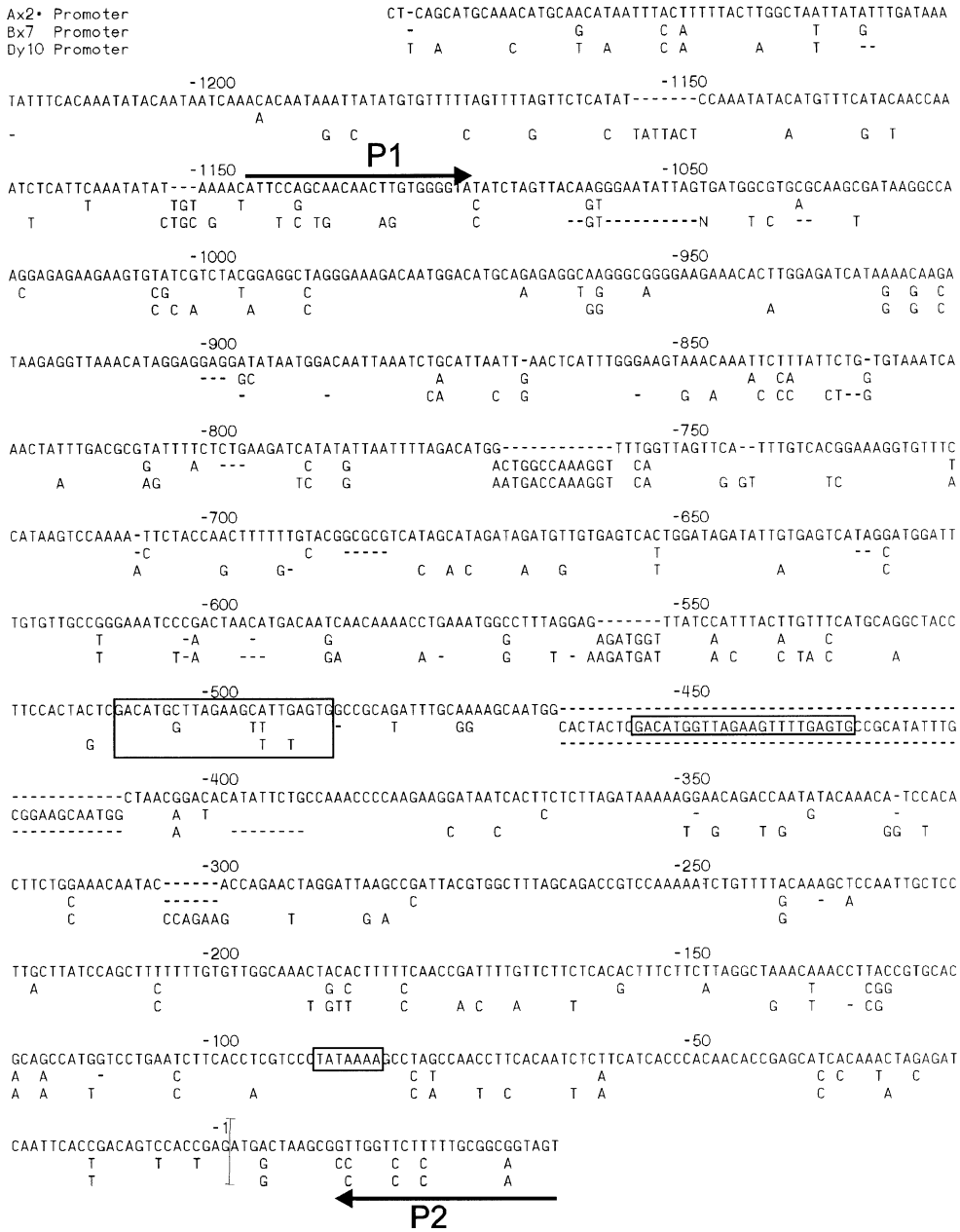

Fig. 4 HMW-glutenin promoter sequence comparison. The promoter sequences of three HMW-glutenin genes (Ax2*, Bx7, and Dy10) from cv Cheyenne are aligned. Positions are indicated as negative bp upstream of the start codon in the alignment and not absolute positions in any individual gene. The Ax2* sequence is used as the reference, with differences in the Bx7 and Dy10 genes shown as base changes and with dashes for deletions. The "cereal-box" element characteristic of cereal storage proteins is *boxed* both in the aligned portion of all three genes (−500) and in the duplicated segment of the Bx7 gene (−450). The TATA element is at −91 bp and is *boxed*. The promoter and coding DNA sequences are separated by a *vertical bar*. *Arrows* indicate the positions in the Bx7 sequence chosen for the forward (*P1*) and reverse (*P2*) PCR primers used to amplify Bx promoter sequences

Flanking-sequence divergences were also noted from extensive sequencing of the α-gliadin genes (Anderson et al. 1997). We speculate that the conserved regions contain the DNA sequences required to control gene expression. The divergence of the more-distal DNA sequence may also be significant in understanding Triticeae genome evolution. SanMiguel et al. (1996) have shown that an extended intergenic region in maize is composed of nested retrotransposons. The same intergenic region in sorghum shows no significant homology to the maize sequence (Avramova et al. 1996). In the present study we report flanking-sequence divergence among known homoeologous loci. A more extended sequence is necessary to see if the distal flanking DNAs of the HMW-glutenin genes are unrelated regions resulting from retrotransposon activity. However, there is evidence of transposon activity relative to the HMW-glutenin genes (see below).

Specific amplification of Bx promoters

The large size of the wheat genome and the technical challenge to construct and screen sufficiently large

genomic libraries would inhibit the routine study of additional HMW-glutenin promoter sequences. PCR can be an alternative to library construction, provided that appropriate primers can be developed to amplify a specific flanking sequence of a single member of a multiple-gene family from total wheat DNA. To test the feasibility of this approach, the Bx gene was chosen because of its apparent variation in expression levels among different alleles and cultivars.

The amplification, cloning and sequencing were carried out with the primers shown in Figs. 4 and 5. The forward primer sequence was chosen for three reasons. First, the 5′ end of the conserved portion of the Bx7 promoter region is a high AT region (standard protocols suggest approximately 50% CG for the best annealing); second, it allowed testing the limitations of primer design in attempting to distinguish a single gene from a multigene family; and third, it avoids the possibility of priming within a highly repeated retrotransposon element. Primers P1 (forward) and P2 (reverse) are shown in Fig. 5 and are compared to known sequences of all Cheyenne HMW-glutenin genes. The forward primer is similar to the Ax2* gene sequence and has a five-base mismatch to the Dy10 gene. The reverse primer was unique to the Bx sequence for at least one base at the 3' end and is thus predicted to allow the synthesis only of Bx DNA.

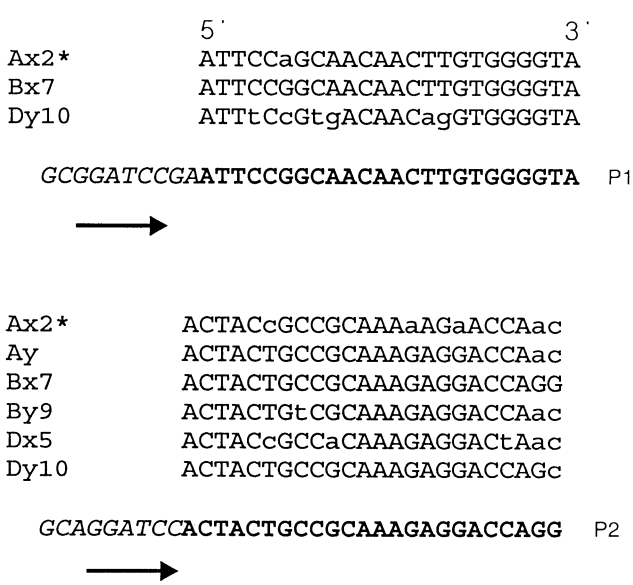

Fig. 5 Forward (*P1*) and reverse (*P2*) PCR primers. The DNA sequence at the same positions of P1 and P2 are shown for all six cv Cheyenne HMW-glutenins genes for the reverse priming site, and for the three genes for which the sequence is known at the forward priming site. Primers P1 and P2 contain an additional ten and nine bases, respectively, to create *Bam*HI sites (*arrows*) at the termini of all amplified DNAs. Primer P1 also includes an *Eco*RI site. The Bx7 sequence is the reference, and differences in sequence in the other genes are indicated by *lower-case letters*. Primer sequences in *bold* indicate Bx7 sequences, while *italicized sequences* are the extensions to generate restriction sites

To test the specificity of the amplification, a blind test used the DNA of several cultivars and genetic stocks. The results (data not shown) indicated that all lines expected to contain a Bx gene supported the amplification of a single band of the expected size (several hexaploid and tetraploid cultivars) whereas DNAs not possessing a Bx gene showed no amplification product:, namely Chinese Spring nullisomic-1B-tetrasomic-1D aneuploid, *T. monococcum* (an A-genome diploid), and rye. The exception to the 1100-bp expected product was from the tetraploid durum (A and B genomes) cultivar Bidi which amplified a 1300-bp product.

Comparison of DNA sequences of four Bx promoters

Bx genes for cloning and sequencing were chosen from four genotypes: cv Cheyenne (whose Bx7 gene was already sequenced and served as the control), Chinese Spring (whose Bx7 gene coding sequence had been shown to be different from Cheyenne; Anderson and Greene 1989), Bidi (a durum cultivar possessing the Bx20 gene), and a randomly chosen *T. turgidum* ssp. *dicoccoides* accession (a wild tetraploid ancestor of durum wheat) with an uncharacterized Bx gene. Multiple independent M13 isolates for all four products were sequenced: seven for Cheyenne, eight for Chinese Spring, five for Bidi, two for the *T. turgidum*. Clones were chosen to approximately represent equal numbers of the two orientations of the inserts, except for the *T. turgidum* inserts where the two sequenced isolates were of the same strand. All clones were confirmed as Bx sequences, establishing that the primers were specific for the Bx genes even in the presence of gene family members whose extended 5′ sequence is not available (Ay, By, Dx). Since the DNA sequence of any specific site was not different in more than one isolate from each gene, the consensus sequence was inferred to be that of the majority of clones. The sequences of both the *T. turgidum* clones were identical except for a single position where the original base cannot be determined between C and T.

The four consensus sequences are aligned in Fig. 6, and show a total of only 12 single-base substitutions, two single-base deletions, and one two-base insertion among the four sequences. The *T. turgidum* consensus sequence showed the most differences with six of the insertions, both deletions, and the 2-base insertion.The major difference among the sequences is a 185-bp insertion at position −1062 of the Bidi sequence (discussed below).

One of the original objectives of this work was to initiate an examination of the relationship of the Bx promoter "cereal-box" duplication to protein-synthesis levels. Since the 54-bp duplication occurs in all examined lines, including the two tetraploid lines (A and B genomes), the duplication occurred prior to the

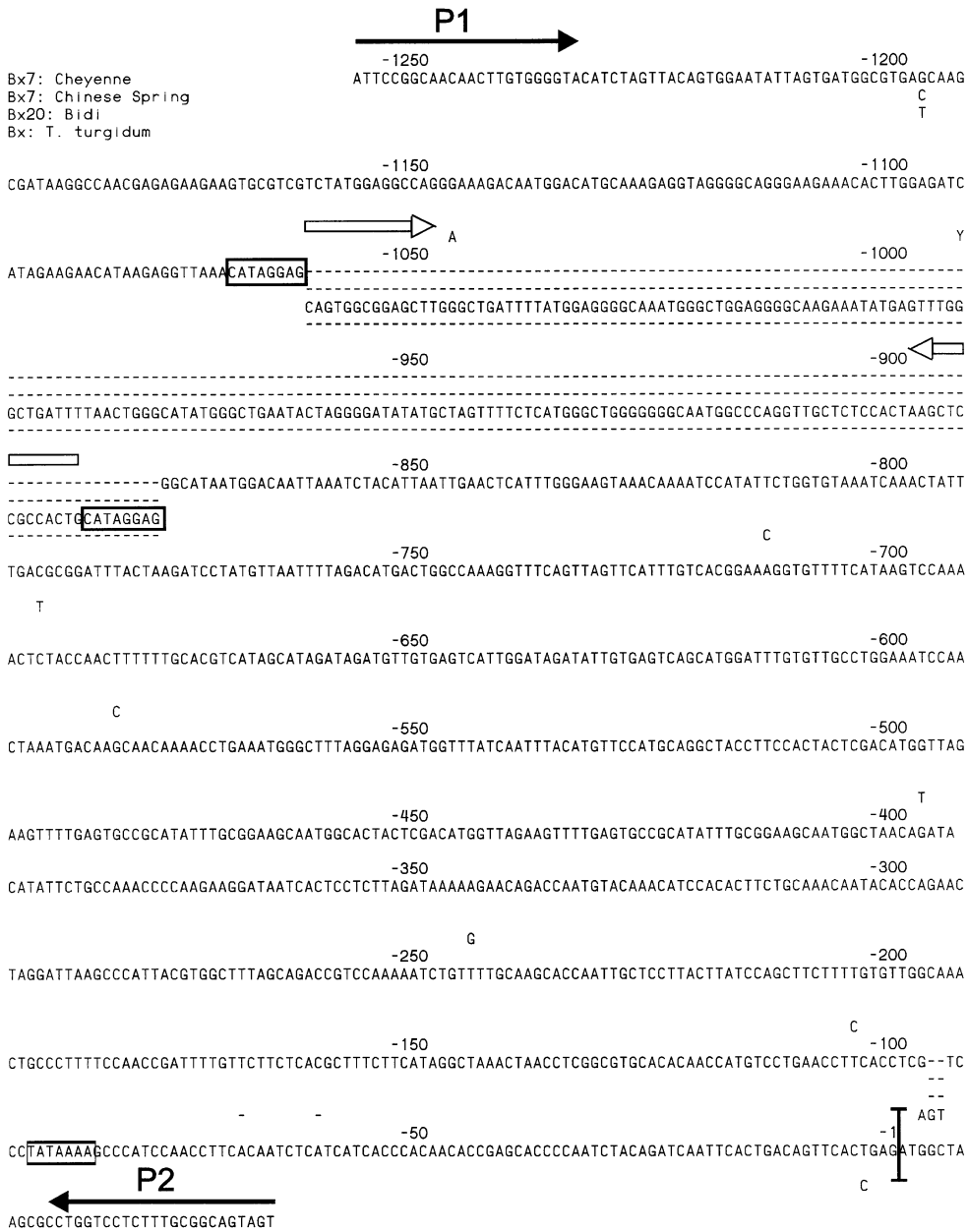

Fig. 6 Alignment of the consensus DNA sequences for the four Bx HMW-glutenin alleles. The Cheyenne Bx7 sequence is the reference, with differences in the other three indicated below the Cheyenne sequence. *Y* = the base in the *T. turgidum* sequence that is not resolved between C and T. The TATA element is boxed at −91 bp. The 5′ non-coding and coding sequences are separated by a *vertical bar*. Primer sequences for amplification are noted by *black arrows*. The Bidi sequence contains a 185-bp insertion at −1038. The 8-bp duplication of the DNA sequence just upstream of the insertion site *boxed*. *Open arrows mark* the inverted repeated ends of the insertion.

origin of hexaploid *T. aestivum* (A, B, and D genomes). We had previously speculated that the apparently higher expression of the cv Cheyenne Bx gene, when compared to the allelic cv Chinese Spring Bx gene, could be related to the cereal-box duplication in the Cheyenne gene (Anderson and Greene 1989). However, since all genes examined contain the same duplication in the promoter, a relationship between apparent levels of hexaploid cultivar Bx gene expression and the duplicated cereal-box is not supported. The cause of the apparent cultivar variation in Bx expression is still unknown.

Evidence for a transpositional event within the Bidi Bx promoter

Several features of the Bx20 (cv Bidi) insertion support a transpositional event. The insertion is flanked by an 8-bp duplication of the insertion site (boxes in Fig. 6), and the termini of the insertion are perfect 14-bp inverted repeats (open arrows) – two standard criteria for transposon identification (Federoff 1989). In addition, the interior of the insertion sequence contains variant repeats of a short motif, such as have been noted for the subterminal repetitive regions of the maize *Spm*

element (Federoff 1989) and the *Tourist* class of small elements associated with maize genes (Bureau and Wessler 1992). Finally, the core of the repeated motif (GGCTGA) is similar to the core of the *Ac* inverted repeat (YAGGGATGAAA; Pohlman et al. 1984). The sequence of the Bidi insertion is not homologous with reported transposable elements and thus may represent a new class of transposons. Searches for similar elements and their distribution in the wheat genome are needed to determine if the Bx20 insertion is a remnant of a transposition event of a larger element, such as *Ac*, *Spe*, or *Mutator* of maize, or if it is another of the small, highly repeated elements such as *Tourist* and *Stowaway* (Bureau and Wessler 1994).

The insertion into the Bx20 promoter has not disrupted the control of gene expression since the Bx20 HMW-glutenin subunit is still synthesized. This is in contrast to the inactivation of the Ay HMW-glutenin gene of cv Chinese Spring due to the 8-kb *WIS* 2-1*A* retrotransposon insertion into the coding region (Monte et al. 1995). Although we have shown the conservation of the HMW-glutenin promoter sequences out to about 1200 bp, there is no direct evidence as yet that this entire sequence is necessary for correct gene transcription. The Bx20 insertion may have either inserted upstream of critical control elements or may have inserted between such elements without effecting their function. Chasan (1994) has pointed out that most *Tourist* and *Stowaway* elements do not produce phenotypic effects and speculates that the structure of sites of insertion are inherently neutral to gene expression.

Since two examples of insertions into HMW-glutenin genes have been found incidentally (*WIS 2-1A* and the Bx20 insertion), such insertions are likely to be relatively recent and common events. We suggest that the HMW-glutenin genes could serve as a convenient trap for active wheat transposons, and a more systematic search among cultivated germplasm and wild wheat accessions (via protein gels and PCR) may yield numerous additional examples of recent transpositions.

Acknowledgements Thanks to Ann Blechl, Jorge Dubcovsky, and Vibha Srivastava for reading the manuscript and making helpful suggestions, and to Cheryl Hsia for assistance with the figures.

References

Anderson OD, Greene FC (1989) The characterization and comparative analysis of high-M_r glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Genet 77:689–700

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

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

Avramova Z, Tkihonov A, SanMiguel P, Jin Y-K, Liu C, Woo SS, Wing RA, Bennetzen JL (1996) Gene identification in a complex chromosomal continuum by local genomic cross-referencing. Plant J 10:1163–1168

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

Blechl AE, Lorens GF, Mackey BE, Greene FC, Anderson OD (1994) A transient assay for the promoter activity of wheat seed-protein genes and other genes expressed in developing endosperm. Plant Sci 102:69–80

Bureau T, Wessler SR (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294

Bureau T, Wessler SR (1994) Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. Proc Natl Acad Sci USA 91:1411–1415

Chasan R (1994) It's elemental. Plant Cell 6:773–775

Crowe JS, Cooper HJ, Smith MA, Sims MJ, Parker D, Gewert D (1991) Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acid Res 19:184

D'Ovidio R, Anderson OD (1994) PCR analysis to distinguish between alleles of a member of a multigene family correlated with wheat bread-making quality. Theor Appl Genet 88:759–763

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

Federoff NV (1989) About maize transposable elements and development. Cell 56:181–191

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

Forde J, Malpica J-M, Halford NG, Shewry PR, Anderson OD, Green FC (1985b) The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L). Nucleic Acids Res 13:6817–6832

Halford NG, Forde J, Anderson OD, Greene FC, Shewry PR (1987) The nucleotide and deduced amino-acid sequences of a 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

Halford NG, Ford, J, Shewry PR, Kreis M (1989) Functional analysis of the upstream regions of a silent and an expressed member of a family of wheat seed-protein genes in transgenic tobacco. Plant Sci 62:207–216

Kolster P, Krechting CF, van Gelde, WMJ (1993) Expression of individual HMW glutenin-subunit genes of wheat (*Triticum aestivum* L) in relation to differences in the number and type of homoeologous subunits and differences in genetic background. Theor Appl Genet 87:209–216

Marchylo BA, Lukow OM, Kruger JE (1992) Quantitative variation in high-molecular-weight glutenin subunit 7 in some Canadian wheats. J Cereal Sci 15:29–37

Monte JV, Flavell RB, Gustafson JP (1995) WIS 2-1A: an ancient retrotransposon in the *Triticeae* tribe. Theor Appl Genet 91:367–373

Pohlman RF, Federoff NV, Messing J (1984) Nucleotide sequence of the maize controlling element *Activator*. Cell 37:635–643

Reddy P, Appels R (1993) Analysis of a genomic DNA segment carrying the wheat high-molecular-weight (HMW) glutenin Bx17 subunit and its use as an RFLP marker. Theor Appl Genet 85:616–624

Robert LS, Thompson RD, Flavell RB (1989) Tissue-specific expression of a wheat high molecular-weight glutenin gene in transgenic tobacco. Plant Cell 1:569–578

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768

Smith RL, Schweder ME, Barnett RD (1994) Identification of glutenin alleles in wheat and Triticale using PCR-generated DNA markers. Crop Sci 34: 1373–1378

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

Shewry PR, Tatham AS, Barro F, Barcelo P, Lazzeri P (1996) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Bio/Technology 13: 1185–1190

Thomas MS, Flavell RB (1990) Identification of an enhancer element for the endosperm-specific expression of high-molecular-weight glutenin. Plant Cell 2: 1171–1180

Thompson RD, Bartels D, Harberd NP (1985) Nucleotide sequence of a gene from chromosome 1D of wheat encoding a HMW-glutenin subunit. Nucleic Acids Res 13: 6833–6846

Vasil, IK, Anderson OD (1997) Genetic engineering of wheat gluten. Trends Plant Sci 2: 292–297