Theor Appl Genet (1983) 67:87–96

Characterization of the multigene family coding for HMW glutenin subunits in wheat using cDNA clones

R. D. Thompson, D. Bartels, N. P. Harberd and R. B. Flavell
Plant Breeding Institute, Trumpington, Cambridge CB2 2LQ, UK

Received September 12, 1983
Communicated by D. von Wettstein

Summary. cDNA clones encoding wheat HMW glutenin subunits have been isolated from a cDNA bank made to poly A^+ RNA from developing wheat endosperm var. Chinese Spring. One such clone, pTag 1290, has enabled us to identify the HMW glutenin mRNA species. The DNA sequence of this clone has been partially determined and it contains several tandem DNA repeats. The sequence is discussed in relation to the generation of the HMW glutenin subunit gene family. Analysis of the organization of the HMW glutenin sequences in the wheat genome revealed that the genes encoding HMW glutenin subunits exist in low copy number and are located on the long arm of each of the homoeologous group 1 chromosomes.

Key words: HMW glutenin subunit genes – cDNA clones – Tandem DNA repeats – Chromosomal location – Gene copy number – Wheat – *Triticum aestivum*

Introduction

Cereal grains synthesize and accumulate large amounts of storage proteins which are deposited in protein bodies during the course of seed development. Gliadins and glutenins are the major storage proteins of wheat (*Triticum aestivum*) endosperm. Together they constitute over 80% of the total protein in the grain (Osborne 1907). Like other cereal storage proteins they are rich in glutamine and proline but poor in lysine, whilst the HMW glutenin subunits are further characterized by their high glycine content (12–20%) (Khan and Bushuk 1979; Field et al. 1982). The amino acid composition determines the unusual solubility properties of gliadins and glutenins. Gliadins are soluble in 70% ethanol or 50% propan-2-ol while the glutenins can only be extracted if a reducing agent and 1% acetic acid are included in these solvents (Miflin et al. 1980).

Gliadins are monomeric proteins, but glutenin is a large, heterogeneous protein aggregate of high molecular weight (up to several million). This aggregate consists of 15 or more different glutenin subunits varying in apparent molecular weight from approx. 11,500–150,000 (Bietz and Wall 1972; Payne et al. 1982b) cross-linked by disulphide bonds and/or secondary bonding forces such as hydrogen bonding, ionic bonding, and hydrophobic interactions (Kasarda et al. 1976). The major components can be grouped on the basis of their apparent molecular weight. The subunits with the highest molecular weight (90–150,000 appar. mol. wt.) are generally termed high molecular weight (HMW) glutenin subunits. The HMW subunits constitute a small proportion of the total glutenin aggregate (Payne and Corfield 1979). Each wheat cultivar possesses only three to five major HMW glutenins as determined by electrophoretic separation (Payne et al. 1982b). The glutenins play an important role in determining the baking quality of flour. They are thought to be solely responsible for the elastic properties of wheat dough (Kasarda et al. 1976). Baking quality, as measured by the SDS-sedimentation test (Axford et al. 1979), has been related to the presence of certain HMW glutenin subunits. Payne et al. (1981a) showed that the differences in baking quality between several wheat cultivars could be related to the presence or absence of specific HMW subunits, although the exact contribution made by individual components could not be directly assessed. Because of the importance of the HMW glutenin subunits their genetics has been studied extensively (for

a review, see Payne et al. 1982 a). The hexaploid wheat (*Triticum aestivum*) is composed of three genomes (AABBDD) and the contribution of each genome can be analysed by using nullisomic-tetrasomic and ditelocentric lines (Sears 1954, 1966). It was shown that all HMW glutenin subunits are controlled by loci named Glu-A1, Glu-B1 and Glu-D1 on the long arms of chromosomes 1A, 1B and 1D respectively (Payne et al. 1980; Payne et al. 1982 b). Each locus may code for more than one polypeptide, and a number of alleles of each locus have been identified in different wheat cultivars (Payne et al. 1981 b). Glu-D1 controls two HMW polypeptides, Glu-B1 either one or two HMW polypeptides and Glu-A1 one or none (Payne et al. 1980). In order to learn more about the structure of the HMW glutenin gene family we have isolated a cDNA clone encoding a HMW glutenin subunit and used it to assay the number of DNA sequences in the family and their chromosomal location, and also to determine the nucleotide sequence of a segment of one of the genes.

Materials and methods

Details of RNA preparation, cDNA cloning, filter hybridization of size fractionated RNA (Northern analysis), restriction mapping and DNA sequencing are described by Bartels and Thompson (1983).

Plant material

The hexaploid wheat variety 'Chinese Spring' and the various nullisomic-tetrasomic and ditelosomic lines derived from it by Dr. E. R. Sears (1966) were obtained from the wheat collection maintained at the Plant Breeding Institute (PBI), Cambridge. Developing endosperm tissue was harvested from plants 2–3 weeks after anthesis according to O'Dell and Thompson (1982).

In vitro translation and extraction of products

Poly(A) RNAs were translated in rabbit reticulocyte lysate, which was prepared and used according to Jackson and Hunt (1983). 20 µl translation assays were incubated for 2 h at 30 °C, using 5 µCi each L-[2,3,4,5-^{3}H]proline (10 µCi/mmol) and L-[4,5-^{3}H]leucine (130 Ci/mmol) as labelled amino acids. At the end of the incubation the samples were treated with pancreatic RNase (Boehringer, Mannheim) at a concentration of 0.1 mg/ml. 2 µl samples were withdrawn and assayed for incorporation while the remaining samples were extracted. The samples were made 50% (v/v) propan-1-ol, 2% (v/v) 2-β-mercaptoethanol, 1% acetic acid, and 10 µg of gliadins were added as carrier. The samples were extracted three times at 60 °C, with this solution, the extracts were pooled, neutralized with 1 M Tris and subsequently pyridethylated with 5-vinylpyridine (Koch Light). The extracted proteins were pooled and precipited with 2% LiCl overnight at 4 °C. The pellets were washed and taken up in 30 µl loading buffer (8 M urea, 4% sodium lauryl sulphate (SLS), 4% 2-β-mercaptoethanol, 0.25 M Tris pH 6.8) and 15 µl dye (0.04% bromophenol blue in 50% glycerol) and incubated at 70 °C immediately before loading on the gel. Polyacrylamide gel electrophoresis (PAGE) was done as described except that the gels were fixed in 10% trichloroacetic acid and 40% methanol at least 3 h before fluorography.

Hybrid-selected translation

The hybrid-arrested translation was performed as described elsewhere (Bartels and Thompson 1983), and the hybrid release translation was done according to Forde et al. (1981) with the following modifications: the plasmid DNA was cut with Eco RI, and the hybridization solution was supplemented with SLS to 0.1%.

Preparation of high molecular weight DNA from wheat

Method 1. Wheat shoots were frozen in liquid nitrogen and stored at −80 °C until use. They were then ground to fine powder in liquid nitrogen, resuspended in 0.2 M NaCl, 50 mM Tris-HCl pH 8.0, 10 mM disodium EDTA, 10 mM dithiothreitol (DTT), 0.2% SLS and 100 µg ml^{-1} Proteinase K and incubated at room temperature for 15–30 min. The mixture was then extracted with phenol: 8-hydroxyquinoline (80% phenol, 0.5% 8-hydroxyquinoline) and re-extracted with chloroform + 1% isoamyl alcohol. Following several further phenol: chloroform (1 : 1) extractions ethanol was layered onto the aqueous phase and the DNA spooled out using a glass rod. The DNA was resuspended in 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SLS, and further purified by CsCl gradient centrifugation and extensive dialysis against 10 mM Tris-HCl pH 8.0, 1 mM EDTA.

Method 2. Another procedure for the isolation of high molecular weight DNA was obtained from Thanh Huynh, Dept. of Biochemistry, Stanford University, California (personal communication) and subsequently modified slightly. Three or four etiolated wheat shoots were homogenized in 15% sucrose, 50 mM disodium EDTA, 0.25 M NaCl, 50 mM Tris-HCl pH 8.0 (ice-cold). Following a low-speed spin in a microcentrifuge the crude nuclear pellet was resuspended in 15% sucrose, 50 mM disodium EDTA, 50 mM Tris-HCl pH 8.0 (ice-cold). Diethylpyrocarbonate and SLS were added to 0.2% each and the suspension heated at 70 °C for 10 min. Following this potassium acetate was added to 0.5 M and the mixture left on ice for 30 min or longer. The K-SLS precipitate was then sedimented at 4 °C in a microcentrifuge. The DNA was ethanol precipitated and resuspended in 10 mM Tris-HCl pH 8.0, 1 mM disodium EDTA, 0.1% SLS. Following several extractions with phenol: chloroform (1 : 1), heat-treated RNAse A was added to 10 µg ml^{-1} and the preparation incubated for 30 min at 37 °C. The mixture was then extracted once more with phenol: chloroform (1 : 1), ethanol precipitated, and the pellet resuspended in 10 mM Tris-HCl pH 8.0, 1 mM disodium EDTA.

DNA of euploid wheat and of lines nullisomic-tetrasomic for the homoeologous group 1 chromosomes (Figs. 5 and 6) was prepared according to Method 1. DNA from the ditelosomic lines (Fig. 5) was obtained using Method 2.

Genomic "Southern" analysis

High molecular weight DNA was prepared from etiolated 5 day old shoots of euploid and aneuploid wheat lines by one of the methods described above.

DNA was digested to completion with Eco RI and separated by electrophoresis on 0.8% neutral agarose gels (5 µg DNA per track). Estimates of glutenin gene copy number were made by electrophoresis of known amounts of pTag 1290 linearized with Eco RI in tracks adjacent to wheat DNA samples. Lambda DNA fragments resulting from digestion with a variety of restriction endonucleases were used as size markers.

Following electrophoresis and staining the DNA was partially depurinated (Wahl et al. 1979) and transferred to a nitrocellulose filter (Southern 1975). The nitrocellulose was prehybridized for 3 h at 65 °C and then hybridized for 24 h at 65 °C in 0.6 M NaCl, 10 mM Pipes pH 6.8, 1 mM disodium EDTA, 0.1% SLS, 0.2% Gelatin, 0.2% Ficoll, 0.2% polyvinylpyrrolidone and 10 μg ml^{-1} sonicated calf thymus DNA.

The isolated cDNA insert of pTag 1290 was labelled to a specific activity of $5\times10^{7}-10^{8}$ cpm/μg with ^{32}P by nick-translation (Rigby et al. 1977). The concentration of probe used in the hybridization was approximately 10 ng ml^{-1}.

Following hybridization the filter was rinsed briefly in 2×SSC (1×SSC = 0.15 M NaCl and 0.015 M Na citrate), 0.1% SLS at room temperature, and then washed in three changes of 2×SSC 0.1% SLS at 65 °C (at least 15 min each change) with occasional agitation. Finally the filter was washed in two changes, 15 min each change, of 0.5×SSC, 0.1% SLS at 65 °C. Following washing the filter was dried at 55 °C and exposed to Kodak X-Omat AR X-ray film with a Dupont Cronex Intensifying Screen at −80 °C for 2–7 days. Autoradiographs were scanned using a Pye Unicam SP 1800 Ultraviolet Spectrophotometer with scanning attachment.

Results

Characterization of glutenin RNA in vitro translation products

The HMW glutenins of the variety Chinese Spring are resolved into four bands with apparent molecular weights between 92,000 and 120,000 by fractionation in a polyacrylamide gel (Payne et al. 1980). Using nullisomic-tetrasomic lines of Chinese Spring, it was shown that the bands with the fastest and slowest mobilities are controlled by Glu-D1 on chromosome 1D, whereas the other two are controlled by Glu-B1 on chromosome 1B (Payne et al. 1980) (see also Fig. 1 B).

The same approach was taken to identify the HMW glutenin proteins amongst the in vitro synthesized proteins. Total poly A^+ RNA was prepared from the developing endosperm of the variety Chinese Spring, its nullisomic-tetrasomic lines N1AT1D, N1BT1A and N1DT1A and of the variety Sicco. The RNAs were translated in the rabbit reticulocyte lysate system at their optimal concentrations. The proteins synthesized, which were reduced and extracted in 50% isopropanol and 1% acetic acid (Field et al. 1982), were separated in a denaturing 12.5% polyacrylamide gel containing DTT as a reducing agent (Fig. 1). The RNA from Chinese Spring directed the synthesis of four HMW polypeptides numbered 1–4 in Fig. 1 A.

The same pattern was obtained for N1AT1D RNA (Fig. 1 A) track e whereas the in vitro products from poly A^+ RNA from N1BT1A (Fig. 1 A) track f and N1DT1A (Fig. 1 A) track g each showed the absence of

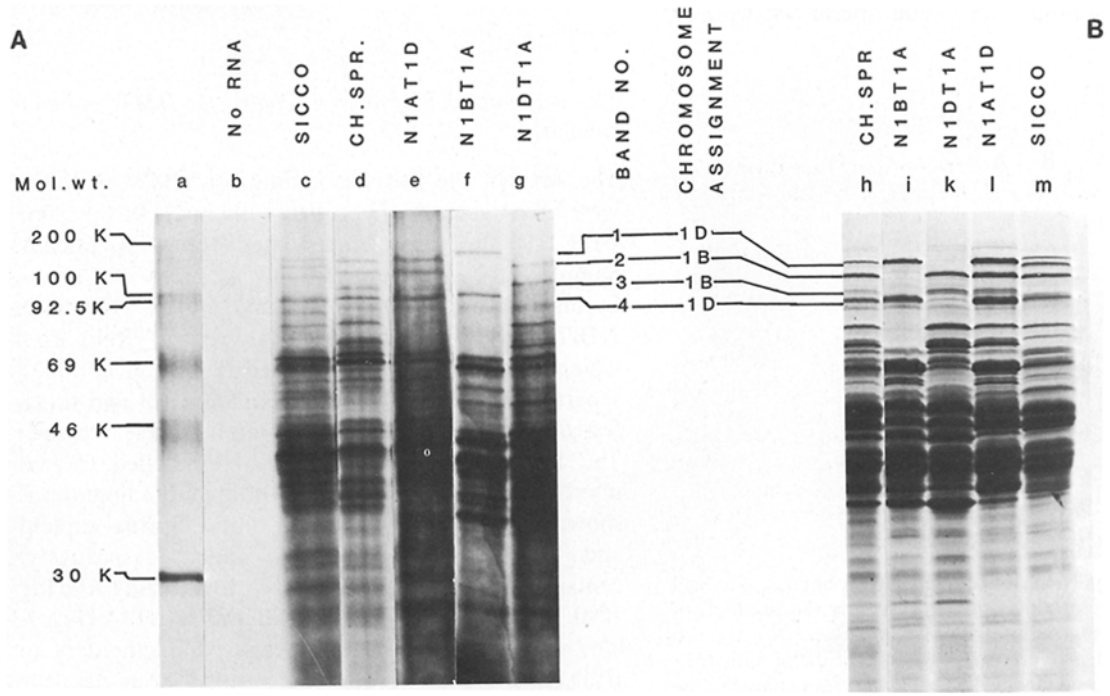

Fig. 1. A Poly A^+ RNA directed protein synthesis in a rabbit reticulocyte lysate. The in vitro translation products were separated on a 12.5% SDS-urea-polyacrylamide gel. *Lanes* (*c*) – (*g*) show the separated translation products of poly A^+ RNA from var. Sicco (*c*), var. Chinese Spring lines: euploid (*d*), N1AT1D (*e*), N1BT1A (*f*), N1DT1A (*g*). In *lane* (*b*) no RNA was used. ^{14}C-labelled proteins were separated in *lane* (*a*) as size markers. The HMW glutenin subunit bands of var. Chinese Spring are numbered and their chromosome assignment is given. **B** Fractionation of total endosperm proteins. Proteins extracted from (*h*) Chinese Spring (*i*) N1BT1A, (*k*) N1DT1A, (*l*) N1AT1D, (*m*) var. Sicco. The proteins were extracted as described by Payne and Corfield (1979) and subsequently separated on a 12.5% SDS-Urea-polyacrylamide gel

two high molecular weight proteins. This corresponds exactly to the in vivo protein pattern for these genotypes (Fig. 1 B), as first described by Payne et al. (1980). The results in Fig. 1 A show therefore that chromosome 1D controls the expression of the RNA coding for the fastest and slowest migrating proteins (Nos. 1 and 4, Fig. 1) and chromosome 1B directs the synthesis of proteins 2 and 3. The proteins synthesized in vitro using poly A^+ RNA from Sicco also reflect exactly the in vivo protein pattern (Fig. 1, lane c and m), as also observed for other varieties by Forde and Miflin (1983). Thus, on the basis of their mobility in one-dimensional PAGE, their solubility in 50% isopropanol, 1% acetic acid, the close resemblance between the patterns of in vivo and in vitro synthesized proteins and their chromosomal control, the protein bands (numbered 1–4 in Fig. 1 A) were identified as HMW glutenins.

Identification of cDNA clones coding for HMW glutenins

After identification of the bands in the in vitro translation pattern corresponding to HMW glutenins, cDNA clones could be screened by hybrid selected translation in order to identify cDNA clones encoding HMW glutenins. A bank of cDNA plasmid clones (pTag) had been constructed using as a template poly A^+ RNA from the developing wheat endosperm of Chinese Spring (Bartels and Thompson 1983). The majority of these cDNA clones, which hybridise strongly to wheat endosperm poly A^+ RNA, encode gliadin-like proteins (Bartels and Thompson 1983 and unpublished results). However, one small group of clones, defined by hybridization to the insert of the clone pTag 1290, did not cross hybridize to the gliadin cDNA clones but did hybridize to endosperm poly A^+ RNA from wheat, rye and barley (Bartels and Thompson 1983). The cDNA clone pTag 1290 was chosen as a representative of this group for hybrid selected translation experiments in order to examine its coding properties. pTag 1290 contains an insert of approximately 1,500 base pairs. Two different Hae III fragments of this insert were subcloned into the single-stranded phage M13. Both subclones (Hae 1 and Hae 2) were used for hybrid-arrested translation experiments (Bartels and Thompson 1983), the results of which are presented in Fig. 2. Both subclones hybridized to RNA coding for all four HMW glutenins. When the total poly A^+ RNA was translated after hybridization with pTag 1290, all the bands corresponding to HMW glutenins were absent (Fig. 2, lanes b and d) but were present when the DNA-RNA hybrids were melted prior to translation (Fig. 2, lanes c and e). In lane f of Fig. 2 are translation products from RNA hybridized with the vector M13 mp9 and in lane g the corresponding dissociated sample.

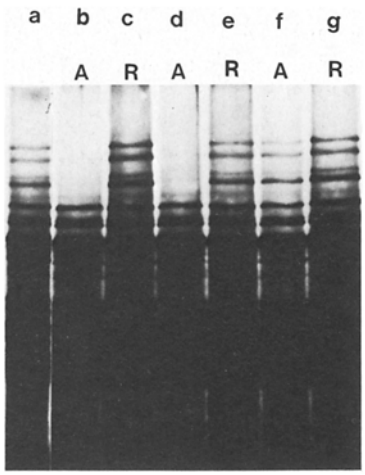

Fig. 2. SDS-Urea-polyacrylamide gel electrophoresis of hybrid arrested translation products. *Lane* (*a*) translation products of Chinese Spring endosperm poly(A) RNA; *lanes* (*b*) and (*d*) contain translation products from RNA hybridized with M13 Hae-1 and Hae-2 respectively; *lanes* (*c*) and (*e*) contain translation products from RNA hybridized with Hae-1 and Hae-2 and melted prior to translation. *Lanes* (*f*) and (*g*) show the same experiment for the M13 vector DNA on its own; in *lane* (*f*) the translation products after hybridisation of the vector DNA mp9 to the RNA, and in *lane* (*g*) translation products from the melted sample. *A* indicates arrested translation and *R* restored i.e., melted sample

Size determination of mRNAs coding for HMW glutenin subunits

The sizes of the mRNAs coding for HMW glutenins were investigated by hybridizing the insert of the clone pTag 1290 to size fractionated poly A^+ mRNA. Samples of 2 μg endosperm poly A^+ RNA of Chinese Spring, its nullisomic-tetrasomic lines N1AT1D, N1BT1A, N1DT1A and Sicco (as well as rRNA from Chinese Spring) were glyoxalated (Koller et al. 1982), separated on a denaturing urea-agarose gel and transferred to ultrablot paper (Collaborative Research Inc.) The filter was hybridized with ^{32}P-labelled excised insert from pTag 1290. The resulting autoradiograph is shown in Fig. 3. RNAs from Chinese Spring euploid and N1AT1D (Fig. 3, lanes b and c, respectively) contain four size classes of mRNA hybridizing to pTag 1290. In N1DT1A (Fig. 3, lane d) and N1BT1A (Fig. 3, lane e) two bands of mRNA each complementary to pTag 1290 are observed. The results show that the fastest and slowest migrating bands of glutenin mRNA are under the control of chromosome 1D, whereas 1B controls the synthesis of the other two. The insert of pTag 1290 hybridizes to the same four bands of mRNA in N1AT1D as in the euploid, suggesting that chromosome 1A does not direct synthesis of

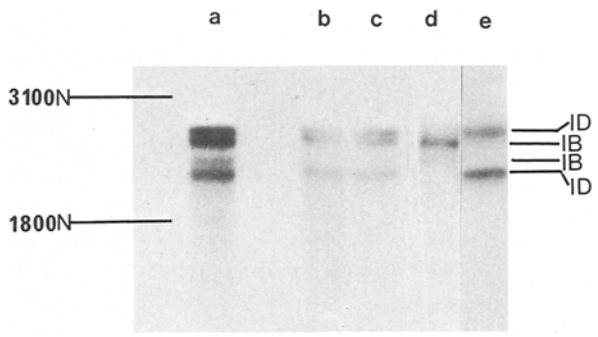

Fig. 3. Hybridizations of pTag 1290 to RNAs. Wheat endosperm poly(A) RNA was glyoxalated, fractionated on urea-agarose gels, and transferred to nitrocellulose (Thomas 1980). *Lane* (*a*) Sicco, (*b*) Chinese Spring euploid, (*c*) N1AT1D, (*d*) N1DT1A, (*e*) N1BT1A. The positions of the wheat cytoplasmic rRNAs are indicated on the *left hand side* and the chromosome assignment of the mRNA species on the *right hand side*

detectable amounts of mRNA homologous to pTag 1290. The endosperm poly A^+ mRNA of Sicco contains at least four different size classes of mRNA hybridizing to pTag 1290 (Fig. 3, lane a), which also correspond in relative sizes to the translation products of Sicco RNA.

In a subsequent experiment the same filter was hybridized with the insert of pTA 71 (a wheat ribosomal RNA gene clone, Gerlach and Bedbrook 1979) to estimate the sizes of the various mRNA sets hybridizing to pTag 1290. The position of the wheat ribosomal RNAs are indicated in Fig. 3 and their lengths relative to *E. coli* rRNA were calculated as 3100 and 1800 nucleotides (N). Taking the wheat rRNA lengths as standards, the sizes of the mRNAs in Chinese Spring homologous to pTag 1290 are estimated as 2700 N, 2400 N, 2250 N and 2100 N.

The sizes of the HMW glutenin subunits estimated from electrophoresis in SDS vary between 90,000 and 150,000 for Chinese Spring wheat. The largest mRNA detected in these genotypes (2700 N) could code for a protein of about 89,500 m.wt., assuming 200 N of non-coding RNA, and a mean residue m.wt. of 107.5 (Field et al. 1982), while the smallest RNA (2,200 N) could code for a protein of m.wt. 68,000. These figures are closer to those obtained by equilibrium centrifugation measurements on glutenin polypeptides (Field et al. 1982).

DNA sequence analysis of the clone pTag 1290

In order to confirm the identity of the clones and to gain information on the structure of a HMW glutenin protein, a portion of the insert was sequenced. One of the possible reading frames predicts a polypeptide fragment of similar amino acid composition to HMW glutenin (Field et al. 1982). The sequence shows extensive internal repetition (Fig. 4 A) and has been aligned to maximize residue homology in columns (Fig. 4 B). The sequence shows no homology with the recently determined amino-terminal protein sequence of individually isolated HMW glutenin subunits (Field et al. 1982).

The organisation of glutenin sequences in the wheat genome

a) Nullisomic-tetrasomic analysis. The insert from the cDNA clone pTag 1290 was used to investigate the organization of homologous sequences within the wheat genome. DNA from euploid wheat was digested to completion with EcoRI (an enzyme for which there is no site in pTag 1290 insert) separated on a gel, and transferred to nitrocellulose (Fig. 5 A). Five EcoRI fragment bands, of sizes 21.5 kb, 16.5 kb, 8.6 kb, 7.0 kb and 6.0 kb hybridized to the cDNA insert. Similar analysis of the DNA from lines which are nullisomic-tetrasomic for chromosomes within a homoeologous group reveals the chromosomal location of each fragment (Fig. 5 A). For example, EcoRI restricted DNA from plants nullisomic for chromosome 1B does not contain the 21.5 kb EcoRI fragment observed in euploid DNA. Hence this fragment must be derived from chromosome 1B, a conclusion supported by the observation that it is more intense in DNA obtained from plants tetrasomic for chromosome 1B (Fig. 5 A, tracks 2 and 7). Each of the EcoRI fragments hybridizing to the pTag 1290 insert can be assigned to a chromosome in this way, and all are derived from the members of homoeologous group 1 (chromosomes 1A, 1B and 1D).

b) Ditelosomic analysis. The genomic sequences homologous to the insert of pTag 1290 can be further localized to a chromosome arm by use of DNA from the group 1 ditelosomic lines in an experiment analogous to that described above (Fig. 5 B). For example, the 21.5 kb fragment is observed in EcoRI digested DNA from DT 1BL (short arm of chromosome 1B deleted) but not observed in DNA from DT 1BS (long arm of 1B deleted). Hence this fragment must be carried on the long arm of chromosome 1B (1BL). The 16.5 kb and 7.0 kb fragments can be similarly located to the long arm of chromosome 1A. Since the 8.6 kb and 6.0 kb fragments are absent in DNA from N1DT1A and N1DT1B (see above) but are present in DNA from DT1DL they must be derived from the long arm of chromosome 1D. The DT 1DS line was not available for study.

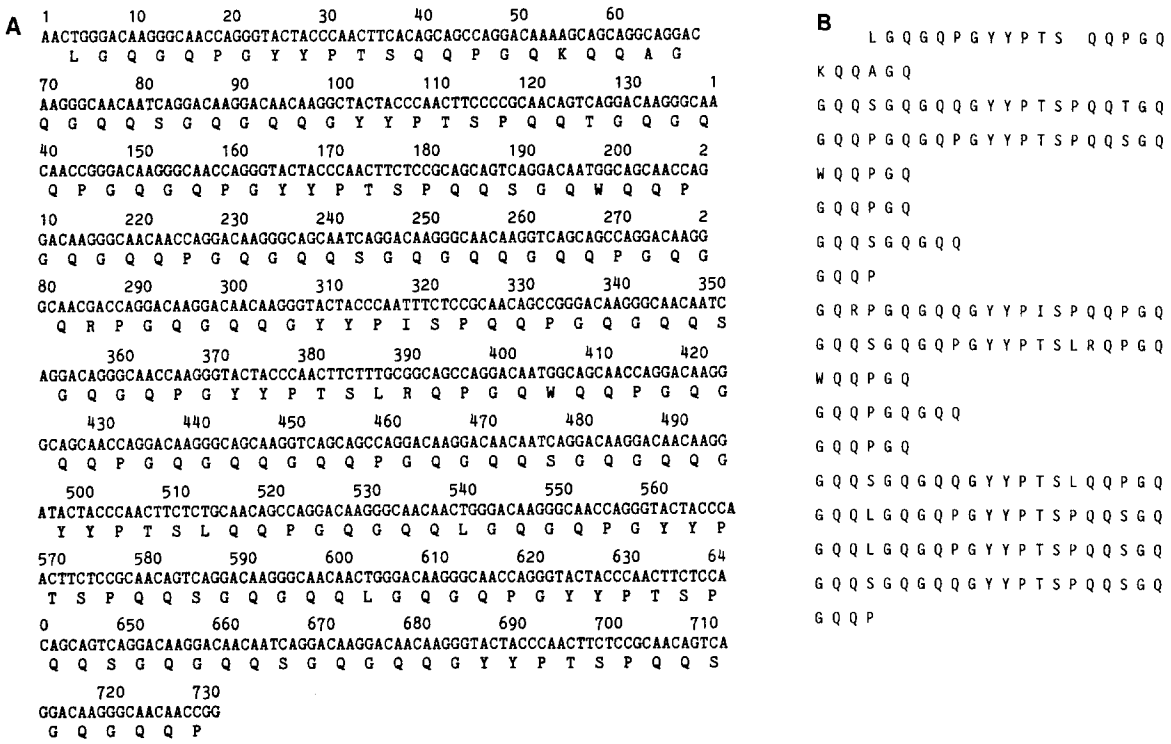

Fig. 4. **A** DNA sequence of first 730 nucleotides of pTag 1290. Below the DNA sequence is indicated the polypeptide sequence predicted by one reading frame. **B** Polypeptide sequence predicted by part of pTag 1290. The sequence is arranged to maximize homology between repeating units. Residues showing homology are aligned in vertical columns. Gaps have been introduced to maximize the alignment

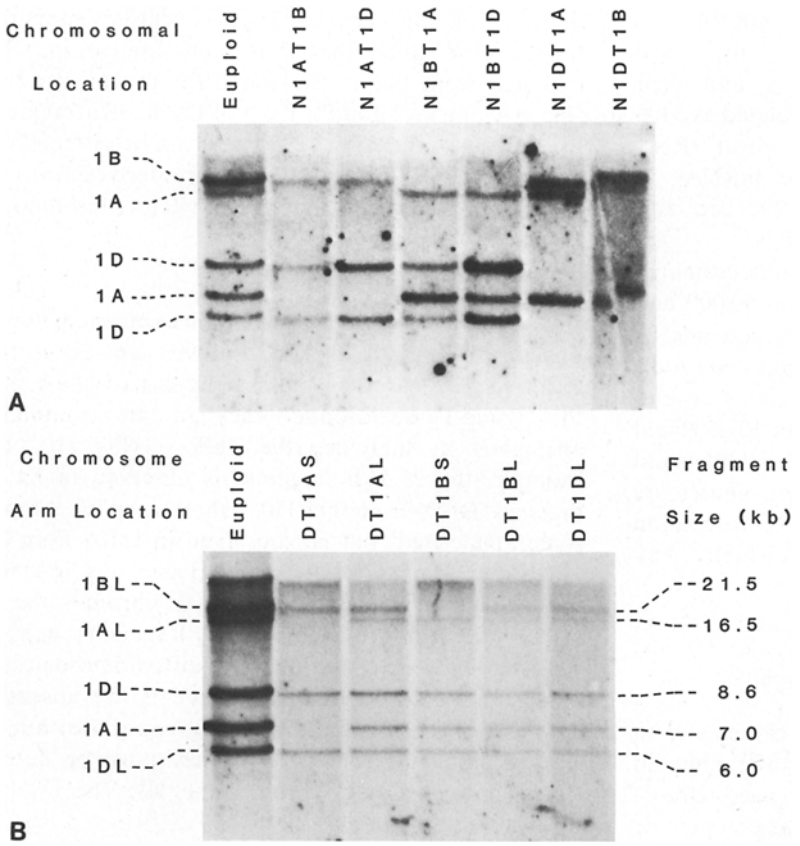

Fig. 5. Hybridization of the insert of pTag 1290 to EcoRI digested wheat DNA of (**A**) euploid and nullisomic-tetrasomic lines of Chinese Spring, (**B**) euploid and ditelosomic lines of Chinese Spring. Hybridization to DNA near the top of some tracks is considered to be an artifact due to the high concentration of DNA at the limiting mobility of the gel

The copy number of glutenin sequences in the wheat genome

In order to provide an estimate of the number of copies within the wheat genome of sequences closely related to the insert of pTag 1290, the insert was hybridized to known quantities of pTag 1290 DNA (linearized with EcoRI) on the same filter as known quantities of wheat DNA also cleaved with EcoRI. The results (Fig. 6) show that the number of wheat sequences in each EcoRI fragment is clearly less than five (compare track C with Track A, Fig. 6). As a rough estimate, each of the 1A and 1D EcoRI fragments appears to carry approximately one copy, and the 1B EcoRI fragment approximately two copies. This analysis assumes that the homoeologous loci are all equally homologous to the probe. However, because the intensity of hybridization of each of the DNA fragments to the probe is roughly comparable, it seems likely that sequence differences between homoeologues do not significantly affect the general conclusion that the HMW glutenin subunits in Chinese Spring are encoded by a small family of five to ten genes.

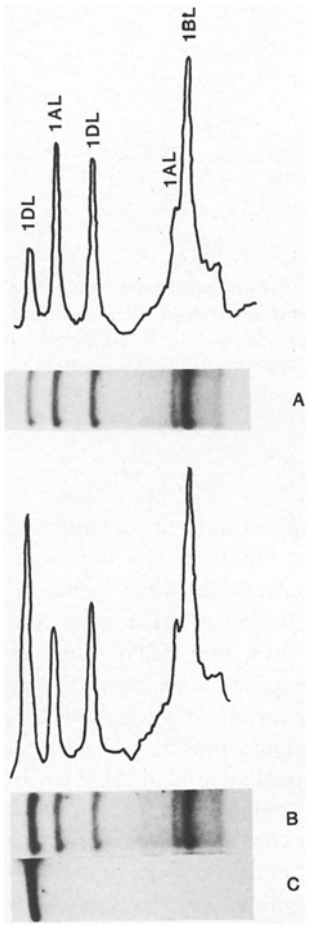

Fig. 6. Estimates of the number of glutenin sequences in wheat. Hybridization of the insert of pTag 1290 to (**A**) EcoRI digested euploid Chinese Spring DNA (shown with scan trace above), (**B**) EcoRI digested euploid Chinese Spring DNA together with EcoRI linearized pTag 1290 representing one copy of the insert sequence per wheat haploid genome (shown with scan trace above), (**C**) EcoRI linearized pTag 1290 representing 5 copies of insert sequence per haploid genome. 5 μg wheat DNA is the equivalent of 2.76×10^5 copies of the hexaploid wheat genome. 1.75×10^{-6} μg pTag 1290 DNA contain 2.76×10^5 copies (i.e. one copy per haploid genome) of the pTag 1290 insert. Linearized pTag 1290 comigrates with the smallest Chinese Spring (1DL) EcoRI restriction fragment hybridizing to the insert of pTag 1290

Discussion

The cDNA clone pTag 1290 derived from poly A^+ RNA from the developing endosperm of Chinese Spring has been identified as containing a sequence for a HMW glutenin polypeptide because it (i) hybridises to poly A^+ RNAs which translate in vitro into HMW glutenins, (ii) has a DNA sequence which gives an amino acid sequence characteristic of HMW glutenins (Field et al. 1982) and (iii) hybridises to DNA sequences found only on the long arms of chromosomes 1A, 1B and 1D, where the Glu-1 loci encoding HMW glutenins are known to reside, unlike genes for any of the other storage proteins (Payne et al. 1980). The pTag 1290 clone, and others related to it, have hybridisation properties which might be expected for a HMW glutenin clone in that they do not hybridise to clones for gliadin sequences (Bartels and Thompson 1983) which have different NH_2-terminal sequences and amino acid compositions, but do hybridise to endosperm poly(A) RNA from rye and barley, both of which have closely related HMW proteins (Field et al. 1982).

Genetic studies suggest that the Glu-1 loci on chromosomes 1B and 1D contain two tightly linked genes, but that there is only one gene at the Glu-1 locus on chromosome 1A. The finding that the cDNA clone hybridises to a similar extent to sequences on chromosomes 1A, 1B and 1D implies, as expected, that the glutenin genes contain similar sequences. This conclusion is also endorsed by the findings that both Hae III subclones of pTag 1290 hybridised to mRNAs specifying all four HMW glutenin subunits in Chinese Spring (Fig. 2) and that under more stringent hybridisation conditions pTag 1290 hybridised to mRNAs specifying all five HMW glutenin subunits in Sicco (results not illustrated). These results indicate that the coding regions for the two polypeptides (termed x and y subunits; Payne et al. 1981 b) produced by each 1B or 1D locus contain closely related DNA sequences despite showing different peptide fingerprint patterns (Payne et al., unpublished data).

Our estimates of the numbers of sequences in the wheat genome related to pTag 1290 are in good

agreement with these numbers of genes. The wheat glutenin gene family is therefore small unlike the zein families in maize (Hagen and Rubenstein 1981; Viotti et al. 1979; Burr et al. 1982; Pedersen et al. 1982) or those of most of the other wheat prolamin gene families (Harberd et al., unpublished) which contain many more members.

Gene structure, evolution and genetic variation

The part of the cDNA clone that has been sequenced represents between 30 and 38% of a complete HMW glutenin coding sequence. Throughout the sequenced region are many copies of the amino acid subrepeat unit XQQXGQ where X can be one of a number of amino acids. The subrepeat sequences can be compared easily when the amino acid sequence is aligned as illustrated in Fig. 4B. When codons corresponding to the vertical columns are compared it is found that 77% of them are identical. Of the variant codons, 57% are due to 3rd base substitutions which do not affect the amino acid subrepeat pattern. This method of alignment also reveals that 63 base pair (bp) subrepeat units also occur, specifying 21 amino acids. Each 63 bp subrepeat consists of two closely related 18 bp repeats separated by a 27 bp sequence, with partial homology to an 18 bp subrepeat.

This sequence arrangement suggests that part of the ancestral HMW glutenin gene evolved as illustrated in Fig. 7 by: (1) amplifications of an 18 bp repeat, specifying the amino acid sequence GQQPGQ or a DNA sequence closely related to it; (2) modification of repeating units by diverse mutations (deletions, insertions, base substitutions etc.); (3) amplification of a 63 bp segment of DNA consisting of two of the 18 bp repeats and a 27 bp sequence which evolved from an 18 bp repeat; (4) mutation (amplification, rearrangement, deletion, insertion, homogenisation, etc.) of the principal subrepeats. The 18 bp subrepeats not recognised today as part of 63 bp repeats could be (a) parts of the gene present before amplification of the 63 bp segments, the 63 bp repeats having become interspersed with the 18 bp repeats by recombination and/or (b) derived from 63 bp repeats by intrarepeat recombination, (recombination between the 18 bp repeats either side of a 27 bp component would lead to elimination of the 27 bp repeat), unequal crossing over or deletion.

Cycles of amplification of short sequences to form tandem arrays followed by the mutation, amplification and deletion of derivatives of the arrays as suggested here is well-known in the genomes of higher organisms. Several such compound arrays have been described in the wheat genome (Bedbrook et al. 1980a, b, 1981; Dennis et al. 1980).

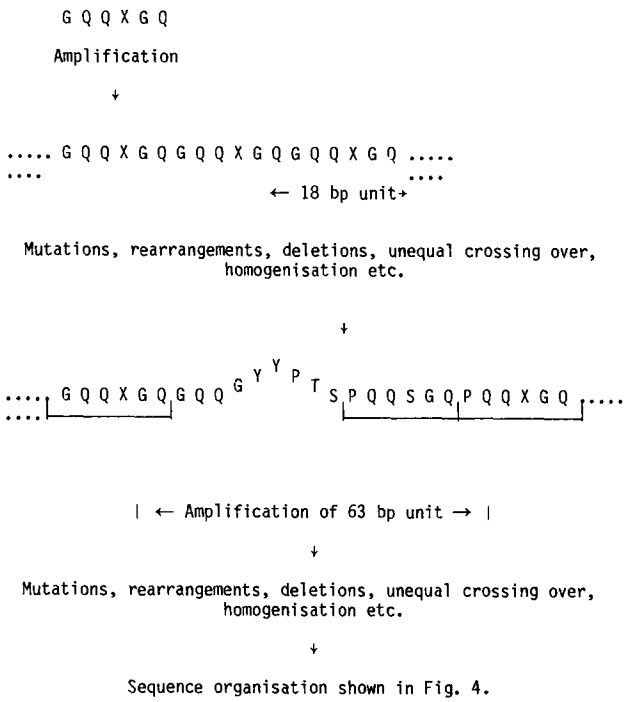

Fig. 7. Proposed scheme for the evolution of part of a HMW glutenin gene. The scheme is written in amino acid residues, rather than nucleotides, for convenience. The initial ancestral hexamer subrepeat has been shown as GQQPGQ but other amino acids specified by related codons could have been present as the first or fourth amino acids of the unit (Fig. 4B)

Direct repeats offer the opportunity for deletions to occur by intrastrand recombination. Furthermore, unequal crossing over can occur between genes at meiosis resulting in the inheritance of longer and shorter coding sequences. Thus the HMW glutenin gene structure suggests possible mechanism(s) for generation of polymorphisms for HMW glutenin size in wheat populations. Such polymorphisms are seen in diploid and hexaploid wheats (Payne et al. 1981b). In this paper analyses of the in vivo produced proteins (Fig. 1B) as well as the in vitro translated proteins, (Fig. 1A) of Chinese Spring and Sicco show that HMW glutenins have different mobilities on electrophoresis within and between varieties. This is also the case for their respective mRNAS. It will be interesting to see if future research substantiates the origin of the length variation as different numbers of the subrepeat recognised in Fig. 4B. We do not know to what extent the repeating sequence structure is the result of selection for a favourable storage protein structure in the endosperm. However it is worth noting that there is considerable variation in sequence and repeat unit length in the glutenin polypeptide studied here.

The presence of repeating units within storage protein sequences and their possible role in gene

evolution has been reported by Kasarda (1980) for gliadins and by Geraghty et al. (1981): Pedersen et al. (1982) and J. A Bietz (personal communication) for zeins.

The chromosome 1A of Chinese Spring does not give rise to a HMW glutenin protein (Payne et al. 1980, 1981b; Holt et al. 1981). No mRNA, complementary to pTag 1290 was detected from this chromosome on the basis that the hybridisation pattern with N1AT1D RNA was indistinguishable from that with euploid RNA. However, DNA sequences homologous to pTag 1290 reside on the long arm of chromosome 1A. It seems likely therefore that these sequences represent an incomplete or inactive HMW glutenin gene(s).

The clones described here will be used as probes to isolate corresponding chromosomal genes for HMW glutenins. Analysis of the complete deduced polypeptide sequences should reveal features characteristic of "good" and "poor" quality glutenin subunits, and hence help elucidate the role these proteins play in the baking process. The indication from our work that these genes are present in low copy number is of importance when the possibility of improving wheat breadmaking quality via transformation is considered.

Acknowledgements. We are indebted to Drs. Peter Payne and Colin Law for advice during the course of this work, to Dr. Richard Cuellar for the generous gift of DNA samples, and to Dr. Colin Lazarus for instructions in genomic southern analysis. D. B. is supported by an EEC grant and N. H. by an MRC Training Fellowship.

References

Axford DWE, McDermott EE, Redman DG (1979) Note on the sodium dodecyl sulfate test of breadmaking quality: comparison with Pelshenke and Zeleny tests. Cereal Chem 56:582–584

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

Bedbrook JR, Jones J, O'Dell M, Thompson R, Flavell RB (1980a) A molecular description of telomeric heterochromatin in *Secale* species. Cell 19:545–560

Bedbrook JR, O'Dell M, Flavell RB (1980b) Amplification of rearranged sequences in cereal plants. Nature 288:133–137

Bietz JA, Wall YS (1972) Wheat gluten subunits: molecular weights determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Cereal Chem 49:416–430

Burr B, Burr FA, St. John TP, Thomas M (1982) Zein storage protein gene family of maize. An assessment of heterogeneity with cloned messenger RNA sequences. J Mol Biol 154:33–49

Dennis ES, Gerlach WL, Peacock WJ (1980) Identical polypyrimidine-polypurine satellite DNAs in wheat and barley. Heredity 44:349–366

Field JM, Shewry PR, Miflin BJ (1982) The purification and characterization of homologous high-molecular-weight storage proteins from grain of wheat, rye and barley. Theor Appl Genet 62:329–336

Forde BG, Kreis M, Bahramian MB, Mathews JA, Miflin BJ, Thompson RD, Bartels D, Flavell RB (1981) Molecular cloning and analysis of cDNA sequences derived from poly A^+ RNA from barley endosperm: identification of B-hordein related clones. Nucleic Acids Res 9:6689–6707

Forde J, Miflin BJ (1983) Isolation and identification of mRNA for the high-molecular-weight storage proteins of wheat endosperm. Planta 157:567–576

Geraghty D, Peifer MA, Rubinstein I, Messing J (1981) The primary structure of a plant storage protein: zein. Nucleic Acids Res 9:5163–5174

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885

Hagen G, Rubenstein I (1981) Complex organization of zein genes in maize. Gene 13:239–249

Holt LM, Austin R, Payne PI (1981) Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin. 2. Relative isoelectric points determined by two-dimensional fractionation in polyacrylamide gels. Theor Appl Genet 60:237–243

Jackson RJ, Hunt T (1983) Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eucaryotic messenger RNA. In: Fleischer S, Fleischer B (eds) Methods in enzymology, Part J. Academic Press, London New York, pp 50–74

Kasarda DD, Bernardin JE, Nimmo CC (1976) Wheat proteins. In: Pomeranz Y (ed) Advances in cereal science and technology 1. Am Assoc Cereal Chem. St. Paul, Minnesota, pp 158–236

Kasarda DD (1980) Structure and properties of α-gliadins. Ann Technol Agric 29:151–173

Khan K, Bushuk W (1979) Studies in glutenin (XIII) gel filtration, isoelectric focusing, and amino acid composition studies. Cereal Chem 56:505–512

Koller B, Delius H, Dyer TA (1982) The organization of the chloroplast DNA in wheat and maize in the region containing the LS gene. Eur J Biochem 122:17–23

Miflin BJ, Byers M, Field JM, Faulks AJ (1980) The isolation and characterization of proteins extracted from whole milled grain, gluten and developing protein bodies of wheat. Ann Technol Agric 29:133–147

O'Dell M, Thompson RD (1982) The purification of immature barley endosperm. J Sci Food Agric 33:419–420

Osborne TB (1907) The proteins of wheat kernel. Carnegie Inst Wash Publ 84:1–119

Payne PI, Corfield KG (1979) Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. Planta 145:83–88

Payne PI, Law CN, Mudd EE (1980) Control by homoeologous group 1 chromosomes of the high-molecular-weigh subunits of glutenin, a major protein of wheat endosperm. Theor Appl Genet 58:113–120

Payne PI, Corfield KG, Holt LM, Blackman JA (1981a) Correlation between the inheritance of certain high-molecular-weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. J Sci Food Agric 32:51–60

Payne PI, Holt LM, Law CN (1981b) Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin. 1. Allelic variation in subunits amongst varieties of wheat (*Triticum aestivum*). Theor Appl Genet 60: 229–236

Payne PI, Holt LM, Lawrence GJ, Law CN (1982a) The genetics of gliadin and glutenin, the major storage proteins of the wheat endosperm. Qual Plant Plant Foods Hum Nutr 31:229–241

Payne PI, Holt LM, Worland AJ, Law CN (1982b) Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin. 3. Telocentric mapping of the subunit genes on the long arms of the homoeologous group 1 chromosoms. Theor Appl Genet 63: 129–138

Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29: 1015–1026

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Sears ER (1954) The aneuploids of common wheat. Mont Agric Exp Stn Bull 572: 1–58

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome manipulations and plant genetics, Suppl Heredity vol 20. Oliver and Boyd, Edinburgh, pp 29–45

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Thomas PS (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc Natl Acad Sci USA 77:5201–5202

Viotti A, Sala E, Marotta R, Alberi P, Balducci C, Soave C (1979) Genes and mRNAs coding for zein polypeptides in *Zea mays*. Eur J Biochem 102:211–222

Wahl GM, Stern M, Stark GR (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulphate. Proc Natl Acad Sci USA 76:3683–3687

Note added in proof

We have recently learned of another paper describing the isolation of HMW glutenin cDNA clones (Forde J et al., FEBS Lett, in press).

The Authors